

# Synthèse sur les modèles de représentation des relations spatiales dans les images symboliques

Valérie Gouet-Brunet\*, Maude Manouvrier\*, Marta Rukoz\*,\*\*

\*WISDOM (<http://wisdom.lip6.fr>) - Regroupement d'équipes de recherche dont CEDRIC/CNAM - 292, rue Saint-Martin - F75141 Paris Cedex 03  
LAMSADE - Université Paris-Dauphine - Pl. du Mar. Lattre de Tassigny - 75775 Paris Cedex 16  
[valerie.gouet@cnam.fr](mailto:valerie.gouet@cnam.fr), [manouvrier@lamsade.dauphine.fr](mailto:manouvrier@lamsade.dauphine.fr), [Marta.Rukoz@dauphine.fr](mailto:Marta.Rukoz@dauphine.fr)

\*\*Université Paris X Nanterre - 200, avenue de la République - 92001 Nanterre Cedex

**Résumé.** La description des relations spatiales entre objets dans les images fournit une sémantique forte qui vient enrichir les techniques bas niveau de représentation du contenu visuel des images, et qui se prête à de nombreux scénarios de recherche dans les bases d'images. Depuis les années 80 avec les travaux de Chang et al. (1987), un grand nombre d'approches ont été proposées pour décrire les relations spatiales dans les images dites symboliques, dans lesquelles les objets d'intérêt sont déjà extraits et identifiés. Cet article dresse un panorama des modèles existants. La typologie choisie sépare les approches dites implicites, qui produisent une représentation globale des relations spatiales existant entre tous les objets de l'image, des approches dites explicites, où la structure produite décrit directement toutes les relations spatiales entre objets. Toutes les approches présentées sont comparées selon plusieurs critères, notamment : type des relations spatiales décrites, volume de stockage, complexité de l'algorithme de comparaison d'images et scénarios applicatifs.

## 1 Introduction

La recherche d'images par le contenu (*Content-Based Image Retrieval*) permet de rechercher les images d'une base de données en fonction de leurs caractéristiques visuelles dites de bas-niveau ou de niveau pixel. Gouet-Brunet (2006) présente un état de l'art de ce domaine. Pour combler l'insuffisance sémantique des caractéristiques de bas-niveau, d'autres méthodes de description d'image sont apparues, permettant par exemple de prendre en compte la description des relations spatiales entre les objets contenus dans les images et fournissant de ce fait une sémantique forte enrichissant les techniques de bas niveau. Depuis les travaux de Chang et al. (1987), un grand nombre d'approches ont été proposées pour décrire les relations spatiales dans les images dites symboliques. Une image symbolique, également appelée logique ou en anglais *iconic* ou *pictorial image*, peut être vue comme une abstraction du niveau physique ou pixel de l'image originale. Alors qu'un objet est décrit par un ensemble de pixels au niveau physique, il est, après annotation manuelle, segmentation ou reconnaissance de forme,

identifié et décrit par un symbole ou une étiquette au niveau logique. Une image symbolique est donc représentée par un ensemble de symboles ou icônes représentant les objets d'intérêt identifiés dans l'image.

Cet article dresse un panorama des modèles de représentation des relations spatiales dans les images symboliques. Les approches modélisant les relations spatiales entre des objets décrits au niveau pixel, comme celles de Colliot et al. (2006) ou de Wawrzyniak et al. (2006), ne sont pas prises en compte ici. Nous avons choisi de classer les approches rencontrées en deux grandes catégories : les approches *implicites* et les approches *explicites*. Les premières produisent une représentation globale des relations spatiales entre objets contenus dans les images. Pour un type de relations spatiales donné, une inférence par rapport aux informations contenues dans le modèle de représentation est nécessaire pour exhiber la relation spatiale existant entre chaque couple d'objets. Les approches de la deuxième catégorie, quant à elles, produisent une structure décrivant toutes les relations spatiales d'un type donné entre objets. Elles sont explicites car toute l'information spatiale y est exprimée et aucune inférence n'est nécessaire pour obtenir la relation spatiale existant entre deux objets particuliers. La typologie choisie s'explique par plusieurs critères : l'intersection entre les deux catégories est nulle, et chaque catégorie a des caractéristiques et propriétés uniques qui la destine à des scénarios applicatifs dédiés, qui seront discutés dans l'article.

L'organisation de cet article est la suivante : la section 2 revisite les notions élémentaires pour la description des relations spatiales. Les sections 3 et 4 décrivent respectivement les approches implicites et les approches explicites avec les éventuelles structures d'index dédiées indispensables pour accélérer la recherche dans les grands volumes d'images. La section 5 synthétise et compare les approches présentées, puis conclut. Les principales approches décrites dans cet article sont illustrées à travers un exemple commun (représenté par la figure 1) et sont comparées selon plusieurs critères tels que les relations spatiales décrites, le volume de stockage produit, la complexité de l'algorithme de comparaison d'images et les scénarios applicatifs possibles.

## 2 Concepts de base

Cette section introduit les notions fondamentales et nécessaires à la compréhension des approches présentées dans les sections 3 et 4 : la représentation symbolique des objets (section 2.1), les types de relations spatiales traités (section 2.2), les mesures de similarité entre relations spatiales et images les plus communément rencontrées (section 2.3) ainsi que les paradigmes de recherche utilisés par ces méthodes (section 2.4).

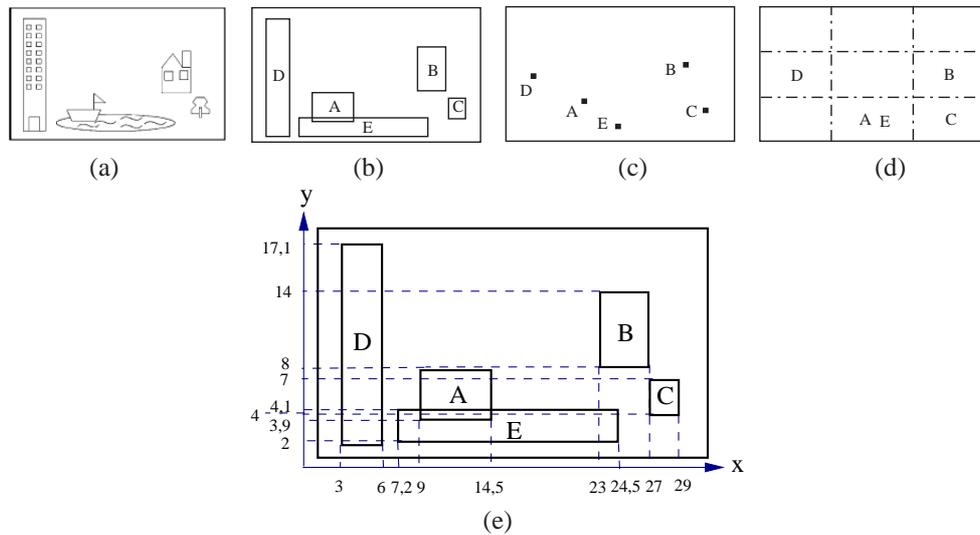
### 2.1 Représentation symbolique des objets

Chang (1991) définit formellement une *image symbolique* par une fonction  $X \times Y \rightarrow W$ , où  $(x, y)$  dénote une position bidimensionnelle,  $x \in X$  et  $y \in Y$ ,  $X = Y = \{1, 2, \dots, M\}$ ,  $M$  étant la taille de l'image, et  $W$  un sous-ensemble de  $V$  qui est un ensemble de symboles. Cette définition est communément reprise dans la littérature où les approches rencontrées sont toutes basées sur l'un des modèles suivants, illustrés par l'exemple de la figure 1 :

- Le centre de masse de l'objet (figure 1(c)) ;

- Une grille, avec la plupart du temps affectation de l'objet dans cette grille par son centre de masse (figure 1(d)) ;
- Le rectangle englobant minimum de l'objet, noté REM dans l'article (figure 1(b)) ;
- La surface exacte de l'objet (ou son approximation plus ou moins fine) pour les relations topologiques.

Il est clair que ces abstractions introduisent une perte d'information plus une moins grande pouvant nuire, dans certains cas, à la précision des relations spatiales décrites, mais les abstractions les plus simples ont le mérite d'être peu coûteuses en stockage et en temps de traitement.



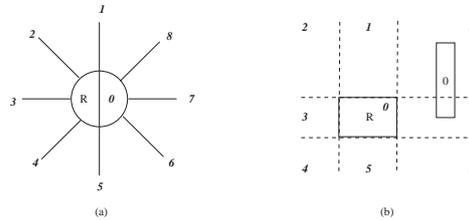
**FIG. 1** – (a) Image exemple (b) Représentation symbolique en REM (c) Représentation symbolique par centres de masse (d) Représentation symbolique par grille (avec affectation par les centres de masse) et (e) coordonnées des REM.

## 2.2 Les différents types de relations spatiales

Dans cette section, nous dressons un panorama des différents modèles de relations spatiales entre objets traités dans la littérature. Parmi les plus connues, nous pouvons mentionner les relations directionnelles, topologiques, géométriques, orthogonales et les projections symboliques. Les approches rencontrées utilisent, voire combinent, ces types.

**Relations directionnelles :** Chang (1991) propose la définition de relations spatiales selon 9 directions dans l'espace 2D, représentées sur la figure 2(a). Ces relations sont définies par un code entier entre un objet de référence ( $R$ ) et un autre objet, le code 0 indiquant que les deux objets symboliques sont au même endroit. Des variantes existent : alors que Chang (1991) se base sur le centre de masse des objets, Huang et Lee (2004) découpent l'espace en neuf zones délimitées par le prolongement des quatre cotés du REM de l'objet de référence, comme illustré par l'exemple de la figure 2(b).

## Les relations spatiales dans les images symboliques



**FIG. 2** – Codage des 9 relations directionnelles, (a) basé sur le centre de masse de l'objet de référence  $R$  pour Chang (1991) ou (b) sur les extrémités de son REM pour Huang et Lee (2004). Par exemple, "4" indique que l'objet d'intérêt se trouve au sud-ouest de  $R$ .

**Relations topologiques :** Egenhofer et Franzosa (1991) ont défini 8 relations topologiques fondamentales existant entre deux régions de taille non nulle : *disjoint*, *touch (meet)*, *inside*, *contains*, *overlap*, *covers*, *covered-by* et *equal*. Elles sont exclusives et s'obtiennent toutes à partir d'intersections entre les concepts de frontière et d'intérieur d'un objet :  $\partial A \cap \partial B$ ,  $A^\circ \cap B^\circ$ ,  $A^\circ \cap \partial B$  et  $\partial A \cap B^\circ$ , où  $\partial A$  et  $A^\circ$  sont respectivement la frontière et l'intérieur de l'objet  $A$ . Pour illustration, cinq de ces relations sont présentes dans la figure 7(b) en annexe.

**Projections symboliques et opérateurs spatiaux 1D :** Parmi les tous premiers modèles de représentation des relations spatiales existant, on trouve ceux issus de la projection des objets bidimensionnels : l'objet est projeté sur les axes  $x$  et  $y$  du repère image, et sur chaque axe, ce sont les relations entre intervalles obtenus qui sont décrites. Si  $p$  est le nombre de relations possibles sur chaque axe,  $p^2$  sera alors le nombre de représentations possibles pour décrire les relations spatiales en 2D.

Les premiers opérateurs spatiaux proposés sont ceux de Chang et al. (1987) :  $A < B$  ( $A$  est avant  $B$  ou  $A$  est au dessus de  $B$ ),  $A = B$  ( $A$  est à la même position 1D que  $B$ ) and  $A : B$  ( $A$  est au même endroit que  $B$ ). Ces seuls opérateurs ne sont pas suffisants pour donner une description complète des relations spatiales pouvant exister dans une image de complexité arbitraire. Très vite, Jungert (1988) et Lee et Hsu (1990) les ont étendus en proposant 6 nouveaux opérateurs "I", "%", "[", "]", "\", et "/" dont l'originalité était de prendre en compte le chevauchement entre objets. Ces opérateurs sont définis à partir des centres des projections des objets, de leur bornes extrêmes et de leur longueur. Par exemple, la relation " $A < B$ " est vraie quand " $\text{center}(A) < \text{center}(B)$ " est vérifié. Nous ne les détaillons pas davantage, car ils seront très vite remplacés dans la littérature par ceux de Lee et Hsu (1990) : reprenant le formalisme rigoureux de Allen (1983) sur la description des relations entre intervalles temporels, Lee et Hsu (1990) ont proposé d'utiliser 7 opérateurs similaires pour décrire les relations spatiales dans la représentation 2D *C-string*. Ces opérateurs 1D sont appliqués sur chaque axe tels que définis dans la table 1. Avec leur symétriques, ils permettent de caractériser un ensemble de  $13 \times 13 = 169$  relations spatiales 2D, répertoriées dans la figure 7(a) en annexe. Ces opérateurs décrivent sensiblement les mêmes relations spatiales que ceux de Jungert (1988), sauf qu'ils sont définis à partir des bornes extrêmes de la projection des objets sur chaque axe, comme dans Allen (1983) en 1D.

Classiquement, les 169 relations spatiales mises en jeu sont regroupées en cinq catégories :

Opérateur	Définition	Signification
$A < B$	$\text{end}(A) < \text{begin}(B)$	<i>A et B sont disjoints</i>
$A = B$	$\text{begin}(A) = \text{begin}(B) \text{ et } \text{end}(A) = \text{end}(B)$	<i>A et B sont identiques</i>
$A B$	$\text{end}(A) = \text{begin}(B)$	<i>A et B sont jointifs</i>
$A\%B$	$\text{begin}(A) < \text{begin}(B) \text{ et } \text{end}(A) > \text{end}(B)$	<i>A contient B (bornes différentes)</i>
$A[B$	$\text{begin}(A) = \text{begin}(B) \text{ et } \text{end}(A) > \text{end}(B)$	<i>A contient B (borne inf. identique)</i>
$A]B$	$\text{begin}(A) < \text{begin}(B) \text{ et } \text{end}(A) = \text{end}(B)$	<i>A contient B (borne sup. identique)</i>
$A/B$	$\text{begin}(A) < \text{begin}(B) < \text{end}(A) < \text{end}(B)$	<i>A et B se chevauchent partiellement</i>

**TAB. 1** – Les 7 opérateurs spatiaux 1D étendus par Lee et Hsu (1990). Les 6 opérateurs symétriques sont notés "<\*", "|\*", "[\*", "]\*", "%\*" et "/"\*".

*disjoint*, *join*, *part overlap*, *contain* et *belong*, représentées dans la figure 7(a). Cette classification permet notamment de mettre en place plusieurs niveaux de similarité entre images, selon que le raisonnement porte sur les opérateurs spatiaux ou seulement sur les catégories spatiales. Notons que certaines de ces relations portent le même nom que les relations topologiques mais elles ne portent pas la même information : deux objets peuvent vérifier la relation topologique *disjoint* alors que leurs REM ne vérifient pas la relation spatiale issue de projections symboliques correspondante.

**Relations géométriques :** Les approches rencontrées dans la littérature qui caractérisent de manière géométrique les relations spatiales sont spécifiques. Leurs points communs principaux sont de se baser sur une représentation des objets par leur centre de masse et d'exploiter des métriques entre ces points, de manière à caractériser les relations qui existent entre couples ou même triplets d'objets. Citons par exemple la structure *TSR* de Guru et Nagabhushan (2001) qui caractérise les relations entre triplets d'objets par des angles construits à partir du triangle formé par leurs centres de masse. De par leur nature et contrairement à la plupart des autres modèles, ces approches se prêtent bien à l'invariance à certaines transformations géométriques, notamment la rotation image.

**Relations orthogonales :** Les relations orthogonales visent à spécifier les relations spatiales entre objets en termes de points cardinaux : "nord", "sud", "est", "ouest" et même une combinaison des orientations intermédiaires, comme "sud-ouest". Ce sont des relations simples, mais populaires dans les scénarios d'interrogation de bases d'images. Dans la pratique, il n'existe pas d'opérateurs spatiaux ou de codage permettant de les représenter, c'est par raisonnement à partir des relations directionnelles, des projections symboliques ou de certaines relations géométriques que ces relations particulières sont déduites. En outre, telles que définies par Chang et Jungert (1986) et ensuite implémentées dans la plateforme IIDS de Chang et al. (1988), elles peuvent mettre en jeu un mécanisme de découpage des objets pour mieux traiter les configurations complexes, par exemple lorsqu'un objet en entoure un autre partiellement.

### 2.3 Les différentes mesures de similarité

Comparer les relations spatiales existant dans deux images dépend naturellement de l'approche de description choisie. Néanmoins, bon nombre de ces approches ont en commun leur

mesure de similarité ou leur algorithme de comparaison. Cette section présente les méthodes les plus communément rencontrées dans la littérature. Pour celles spécifiques à une approche, le lecteur est invité à consulter les paragraphes les décrivant dans les sections 3 et 4.

### Types de similarité entre relations spatiales

Lee et Hsu (1992) ont proposé trois niveaux de similarité pour les relations spatiales décrites par les projections symboliques 1D qu'ils ont définies dans Lee et Hsu (1990) pour la représentation *2D C-string*. La table 2 reprend ces trois types de similarité. Le couple  $(r_{A,B}^x, r_{A,B}^y)$  (resp.  $(r_{A,B}^{x'}, r_{A,B}^{y'})$ ) désigne les relations spatiales 1D selon  $x$  et  $y$  existant entre deux objets  $A$  et  $B$  présent dans une image  $I$  (resp.  $I'$ ). La catégorie mise en jeu dans la similarité de type-0 est celle définie pour les 169 relations spatiales décrite dans la section 2.2.

$$\left\{ \begin{array}{l} \text{type-0 : } \text{catégorie}(r_{A,B}^x, r_{A,B}^y) = \text{catégorie}(r_{A,B}^{x'}, r_{A,B}^{y'}) \\ \text{type-1 : } (\text{type-0}) \text{ et } (r_{A,B}^x = r_{A,B}^{x'} \text{ ou } r_{A,B}^y = r_{A,B}^{y'}) \\ \text{type-2 : } r_{A,B}^x = r_{A,B}^{x'} \text{ et } r_{A,B}^y = r_{A,B}^{y'} \end{array} \right.$$

TAB. 2 – Types de similarité de Lee et Hsu (1992).

Depuis, ces mesures ont été amplement reprises et adaptées par bon nombre de méthodes de représentation des relations spatiales, y compris par celles qui ne sont pas basées sur les projections symboliques. Des enrichissements de ces types de similarité ont aussi été proposés : dans Huang et Jean (1994) et plus tard dans Jean et Lo (2004), l'approche *2D C<sup>+</sup>-string* permet de définir de nouveaux niveaux de similarité plus nuancés grâce à l'ajout de contraintes sur les tailles et distances entre objets ; plus récemment, Yeh et Chang (2006) proposent 6 types de similarité dérivés de ces types qui permettent de combiner plusieurs types de relations spatiales.

Notons que pour les approches plus anciennes *2D-string* de Chang et al. (1987) et *2D G-string* de Chang et al. (1989), les mesures de similarité utilisées sont également de type-0, type-1 et type-2. Cependant, elles sont définies par rapport à la notion de rang d'un symbole dans une chaîne et sont donc différentes de celles de la table 2.

### Métriques pour les relations spatiales

Toutes les mesures de similarité basées sur les type- $i$  sont binaires, la similarité entre deux relations étant vérifiée ou ne l'étant pas. Pour certains types de relations spatiales, des métriques ont été proposées, en vue de permettre une recherche par similarité entre images plus nuancée. Dans tous les cas, ces métriques visent à établir une distance entre deux relations spatiales de la même famille. Ainsi, Egenhofer et Al-Taha (1992) ont enrichi les 8 relations topologiques qu'ils avaient proposées dans Egenhofer et Franzosa (1991), en définissant un graphe pondéré des relations topologiques voisines. Les sommets de ce graphe sont les relations et un arc relie deux relations voisines. Les relations topologiques étant définies à partir d'intersections de frontière et d'intérieur (cf. section 2.2), le poids ou coût de passage d'un arc dépend des différences entre ces intersections. En additionnant les poids mis en jeu, une distance peut ainsi être définie entre n'importe quel couple de relations. La métrique ainsi obtenue peut être résumée dans une table de  $8 \times 8 = 64$  distances. La même année, Freksa (1992) proposait une approche similaire pour les distances entre intervalles définis par Allen (1983). Le

graphe de voisinage des intervalles proposé met en jeu les 13 relations 1D possibles. Il existe ainsi pour cette famille une table de  $13 \times 13 = 169$  distances. La distance entre intervalles qui en est déduite peut être utilisée pour caractériser la similarité entre projections symboliques.

### Similarité entre images

Selon la définition de Eshera et Fu (1984), un graphe ARG (pour *Attributed Relational Graph*) est une triplet  $(V, E, A)$  où  $V$  est un ensemble de sommets,  $E$  un ensemble d'arcs reliant des sommets et  $A$  un ensemble d'attributs qui contient des attributs unaires associés aux sommets de  $V$  et des attributs binaires attachés aux arcs de  $E$ . Les graphes ARG se prêtent donc bien à la représentation de l'ensemble des relations spatiales existant entre objets d'intérêt dans les images. Que ce soit explicitement mentionné ou non par les auteurs, toutes les approches explicites présentées dans cet article décrivant les relations spatiales entre couples d'objets peuvent être vues comme une représentation efficace des graphes ARG. En adoptant ce formalisme, calculer la similarité entre deux images revient à faire de l'appariement de graphes ou de sous-graphes. Depuis les travaux de Barrow et Popplestone (1971), de nombreuses solutions ont été apportées pour l'appariement de graphes (se reporter par exemple à l'évaluation de ces méthodes par Petrakis (2002)). Nous ne les décrivons pas davantage ici, car dans la pratique, les approches de représentation des relations spatiales rencontrées procèdent à des simplifications de ces algorithmes qui sont souvent des problèmes NP-complet ou utilisent des algorithmes d'appariement *ad hoc*.

Les approches de description par chaînes de caractères utilisent le calcul de la plus longue sous-séquence commune (LCS pour *Longest Common Subsequence*) entre deux représentations. L'algorithme associé a longuement été étudié et optimisé, dans Hunt et Szymanski (1977) par exemple. Adapté en 2D aux mesures de similarité de type- $i$ , l'algorithme LCS proposé par Lee et al. (1989) permet de définir un degré de similarité de type- $i$  entre les images  $I$  et  $I'$  :  $c$  est le nombre d'objets en commun entre  $I$  et  $I'$  constituant des couples d'objets dont la relation spatiale vérifie la similarité de type- $i$ . Cette version bidimensionnelle de la LCS passe par la construction d'un graphe dont les sommets sont les objets en commun entre les deux images et les arêtes les relations spatiales binaires entre deux objets vérifiant une similarité de type- $i$  (bien que non spécifié par les auteurs, cette structure peut être vue comme un graphe ARG). Le degré de similarité de type- $i$  entre deux images est alors la cardinalité de la plus grande clique de ce graphe (Bron et Kerbosch (1973); Guan et al. (2000)).

Les approches plus récentes de description des relations spatiales entre images se sont attachées à mettre en place des degrés de similarité plus nuancés que ceux initialement proposés. En effet, l'explosion des applications autour de la problématique de la recherche d'images par contenu visuel a montré qu'il était fondamental de pouvoir effectuer une recherche par similarité plus souple dans les bases d'images, en permettant notamment le tri ou le seuillage des réponses retournées par le système. Ainsi, de nombreuses approches proposant une représentation des relations spatiales lui ont associé une mesure de similarité entre images qui permet au système de retourner non plus une seule réponse (la meilleure) mais un ensemble de réponses ordonnées selon leur similarité à la requête. Les approches  $\Theta\mathcal{R}$ -string de Gudivada (1998), *2D-Be-string* de Wang (2003), "Virtual Image" de Petraglia et al. (2001) et *9D-SPA* de Huang et Lee (2004) offrent cette possibilité.

## 2.4 Les différents paradigmes de recherche

A partir de la description des relations spatiales entre objets dans les images, le paradigme d'interrogation d'un ensemble d'images le plus communément rencontré est la recherche par similarité à partir d'une image exemple : une image contenant des objets labelisés est donnée en entrée du système et les images identiques ou similaires (selon le type ou la mesure de similarité disponible) en termes de relations spatiales sont cherchées et retournées. Selon la technique employée, la recherche peut être *exacte* si la mesure de similarité porte sur tous les objets de l'image et si on ne s'intéresse qu'aux images de la base contenant exclusivement ces objets (c'est par exemple le cas de l'approche Yeh et Chang (2006) pour des images au contenu générique), ou bien au contraire *partielle* si elle permet la recherche de sous-images (impliquant un sous-ensemble des objets de la requête) ou de sur-images (impliquant des images pouvant contenir d'autres objets). Les récentes implémentations de TSR, dans Guru et Punitha (2007), permettent de faire des recherches partielles (évaluées pour des images de visages). Certaines approches gèrent la présence de plusieurs occurrences du même objet dans l'image, comme TSR et les approches matricielles, d'autres non.

Les relations spatiales décrites peuvent aussi être vues comme une connaissance sur laquelle il est possible de raisonner à partir de lois ou de règles d'inférence qui viennent enrichir la recherche par similarité par du raisonnement spatial. Par exemple, plusieurs approches comme Petrakis et Orphanoudakis (1996), Chang et al. (2000), Lee et Hsu (1992) ou Huang et Lee (2004) définissent, en plus du mécanisme de description des relations spatiales, des règles pour déduire d'autres types de relations que celles inhérentes à la méthode, tels que les relations orthogonales (points cardinaux et inter-cardinaux) ou encore les cinq catégories spatiales. En général, les règles d'inférence associées aux approches explicites sont simples car les nouveaux types de relations sont déduits des relations directement disponibles entre deux objets, alors que celles des approches implicites requièrent des mécanismes plus sophistiqués, puisque pour déduire un nouveau type de relation entre deux objets, il est souvent nécessaire de déterminer au préalable la relation existant entre ces objets à partir d'autres objets.

Une variante de la recherche par l'exemple, présentée dans Chang et al. (2000), consiste à considérer un ou plusieurs objets d'intérêt comme requête : la requête peut être un objet et une relation spatiale donnée (dite *auxiliaire*) et ce sont les objets qui vérifient cette requête qui sont recherchés dans la base. Une autre variante, aussi présentée dans cet article, consiste à rechercher l'ensemble des relations spatiales existant entre deux objets donnés. Ces deux alternatives à la recherche par l'exemple classique supposent des règles d'inférence spécifiques aux types de relations spatiales mises en jeu.

## 3 Modèles implicites de représentation des relations spatiales

Cette section présente les modèles de représentation des relations spatiales que nous avons classés dans la catégorie "modèles implicites". Ces modèles produisent une représentation globale des relations spatiales des objets contenus dans les images sous forme de chaînes de caractères (section 3.1) ou sous forme d'arbres (section 3.2). Pour ces modèles, une inférence par rapport aux informations contenues dans le modèle de représentation est nécessaire pour déduire toutes les relations spatiales entre les objets. De plus, ces modèles ne sont pas dyna-

miques : aucune des approches rencontrées ne propose une solution pour éviter la reconstruction complète de la représentation en cas d'ajout ou de retrait d'un objet.

### 3.1 Représentation sous forme de chaînes de caractères

Une manière très répandue de représenter les relations spatiales consiste à les écrire sous forme de chaînes de caractères, les caractères mis en jeu étant les symboles des images et des opérateurs spatiaux. Cette section présente la première approche proposée par Chang et al. (1987), la *2D-string*, ainsi que ses variantes qui permettent soit d'améliorer le temps de recherche en la combinant avec des arbres quaternaires ou des arbres R, soit de prendre en compte un plus grand nombre de relations spatiales. Plusieurs autres travaux ont proposé d'étendre l'approche *2D-string* aux séquences d'images, permettant ainsi des requêtes sur la position relative des objets dans la séquence et notamment sur les changements de position dans le temps ; citons par exemple Arndt et Chang (1989), puis Liu et Chen (2002) qui proposent les *3D-strings*, représentées par la suite par des listes (*3D-lists*).

#### L'approche 2D-string

Soient  $V$  un ensemble de symboles (ou vocabulaire) et  $A$  un ensemble d'opérateurs spatiaux 1D. Une chaîne de caractères 2D (ou *2D-string*) sur  $V$  est définie telle que :  $(u, v) = (x_1 r_1^u x_2 r_2^u \dots r_{n-1}^u x_n, x_{p(1)} r_1^v x_{p(2)} r_2^v \dots r_{n-1}^v x_{p(n)})$ , où  $x_1 \dots x_n$  est une chaîne de caractères sur  $V$  ( $x_i \in V, 1 \leq i \leq n$ ),  $p : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  est une permutation sur  $\{1, \dots, n\}$  et  $r_1^u \dots r_{n-1}^u$  comme  $r_1^v \dots r_{n-1}^v$  sont des chaînes de caractères sur  $A$  ( $x_i^u \in A, x_i^v \in A, 1 \leq i \leq n$ ). Les opérateurs spatiaux définis et utilisés par Chang et al. (1987) sont les trois opérateurs simples  $A < B$ ,  $A = B$  and  $A : B$  définis dans la section 2.2. L'approche *2D-string* suppose une représentation symbolique de l'image sous forme de grille avec affectation des objets par les centres de masse (figure 1(d)). La chaîne est construite à partir d'un algorithme simple de parcours des éléments de la représentation symbolique et de plusieurs règles de réécriture permettant de simplifier la chaîne construite. Par exemple, la *2D-string* représentant l'image de la figure 1(d) est  $(D < A : E < B = C, B = D < A : E = C)$ . La sous-chaîne  $D < A : E < B = C$  correspond à un parcours de l'image selon l'axe  $x$  : l'objet  $D$  étant avant l'ensemble formé par les objets  $A$  et  $E$  (les étiquettes étant ordonnées par ordre alphabétique), lui-même situé avant l'ensemble formé par les objets  $B$  et  $C$ . Le raisonnement est le même pour la sous-chaîne représentant l'axe  $y$ . Cette représentation a le mérite de la simplicité. Cependant, les relations spatiales décrites sont sommaires et non exemptes d'ambiguïtés. Trois types de similarité (différents de ceux présentés dans la section 2.3) sont définis à partir de la notion de *rang* d'un symbole dans la chaîne (position de ce symbole par rapport à l'opérateur "<").

Niu et al. (1999) ont proposé une structure d'index, *2-D-S-tree*, basée sur l'arbre B+ de Comer (1979), pour indexer les *2D-strings*. Les expérimentations réalisées sur une base de 2000 à 15000 *2D-strings* montrent que les performances du *2-D-S-tree* sont plus ou moins constantes et ne se dégradent pas avec la taille de la base d'images. La structure est néanmoins dépendante de la distribution des données. Les auteurs indiquent en particulier qu'elle est efficace pour les bases contenant une grande variété d'images contenant elles-mêmes différents objets avec différentes relations spatiales. Les performances sont en revanche moins bonnes pour une base d'images contenant peu de symboles et peu de relations spatiales (ex. des images de visage).

Le *2-D-S-tree* est implémenté dans le système DISIMA de Oria et al. (2000).

### Variantes et extensions de l'approche 2D-string

Chang et Li (1988) ont proposé une combinaison entre la structure de 2D-string de Chang et al. (1987) et l'arbre quaternaire de Finkel et Bentley (1974), la *2D-H-string*, afin de permettre la représentation multi-résolution des images symboliques de taille  $2^p * 2^p$ ,  $p$  étant le nombre de pixels. Cette structure représente, sous forme de chaîne de caractères, la décomposition de l'image en quadrants et indique la localisation respective des objets de l'image dans les quadrants correspondants selon leurs relations orthogonales (voir section 2.2). La chaîne *2D-H-string* correspond à un parcours de l'arbre quaternaire en profondeur d'abord. Elle a été adaptée par Chang et Lin (1996) en *adaptive 2D-H-string* afin de supprimer une certaine forme de redondance pouvant apparaître dans la structure et de permettre la représentation des images de taille quelconque. Chang et Ann (1999) ont par la suite proposé d'ajouter la taille de l'image à la fin de la *2D-H-string* afin de supprimer des ambiguïtés pouvant apparaître lorsque des images différentes sont représentées par le même arbre quaternaire. Ces approches sont les seules, à notre connaissance, à combiner arbre quaternaire et *2D-string*. Néanmoins, il est évident que toutes les approches d'arbres quaternaires linéaires, telles que celles de Gargantini (1982); Lin (1997) et Yang et al. (2000), peuvent être considérées comme une sorte de *2D-string*, dans le sens où la position spatiale des objets de l'image est codée par une chaîne de caractères – voir synthèse détaillée dans Manouvrier et al. (2005). Il est à noter que les *2D-H-strings* et autres arbres quaternaires linéaires ont été proposés uniquement à des fins de stockage (compression/compactage pour du transfert d'images). Aucun mécanisme n'est envisagé par les auteurs pour permettre de faire des recherches d'images par similarité.

Petrakis et Orphanoudakis (1996) proposent une extension de l'approche *2D-string*, nommée *2D-string étendue*, afin de partitionner l'espace de recherche et d'associer une adresse unique à chaque image. Cette adresse est utilisée par la suite dans une structure de hachage permettant d'accélérer les recherches. Dans cette approche, une *2D-string* ( $u, v$ ) devient étendue lorsqu'elle est représentée par un triplet ( $z, r, s$ ) où  $z$  contient une relation d'ordre sur les symboles et les chaînes  $r$  et  $s$  sont construites en substituant chaque objet dans  $z$  par sa position dans  $u$  et  $v$  respectivement. Deux relations d'ordre sont proposées par les auteurs : dans la première, les noms des objets sont ordonnés selon leur projection horizontale. Lorsqu'ils ont la même projection horizontale, ils sont alors ordonnés par rapport à leur projection verticale. La deuxième relation d'ordre diffère de la première par le fait qu'elle permet que plusieurs objets aient le même rang. Dans Petrakis (2002), une extension de cette représentation prend en compte d'autres caractéristiques des objets comme leur surface et leur périmètre, entre autres.

D'autres extensions de l'approche *2D-string* ont été proposées afin de considérer des relations spatiales plus complexes ou plus précises, comme par exemple l'intersection entre deux objets, la symétrie entre deux images ou la distance entre les objets. L'approche *2D G-string* (pour *Generalized 2D-string*), décrite dans Chang et al. (1989) et Jungert et Chang (1989), propose de généraliser l'approche *2D-string* aux opérateurs spatiaux étendus par Jungert (1988). Cette représentation se distingue principalement par l'introduction d'un mécanisme de découpage des objets selon les axes par rapport aux coordonnées minimales et maximales des objets. A partir d'un tel découpage, les opérateurs spatiaux 1D mis en jeu pour décrire toutes les relations spatiales possibles entre parties d'objets se résument à "<", "=", et "|", ce dernier traduisant la relation "join". Deux opérateurs "(" et ")" sont également ajoutés pour regrouper localement

certaines structures, permettant une structuration hiérarchique de la représentation. Cette approche permet ainsi de traiter plus de relations spatiales que l'approche *2D-string* mais au prix de nombreux découpages d'objets si ceux-ci se chevauchent sur l'un des axes, induisant des chaînes longues. Dans l'approche *2D C-string* proposée par Lee et Hsu (1990, 1991), ce problème est atténué par l'utilisation des opérateurs spatiaux basés sur ceux de Allen (1983) et d'un mécanisme de découpage plus sophistiqué, réduisant ainsi notablement le nombre de parties d'objets. Ici, toutes les relations spatiales sont représentées par les opérateurs "<", "=", "|", "%", "[", et "]", qui ne nécessitent pas de découpage des objets. Lee et Hsu (1992) définissent alors de nouvelles mesures de similarité entre images et donnent les règles de *reconstruction de l'image*, assez coûteuses, nécessaires pour calculer ces mesures en permettant notamment de retrouver la relation spatiale existant entre deux objets à partir de la chaîne *2D C-string*.

Plusieurs variantes de l'approche *2D C-string* existent : Petraglia et al. (1993) tiennent compte des symétries et rotations pouvant survenir entre deux images. En particulier, ils proposent l'approche *2-R string*, qui consiste à utiliser un système de coordonnées polaires dont l'origine est choisie parmi les centres de masse des objets pour localiser les objets. L'algèbre et le mécanisme de découpage associés sont adaptés tout en restant dans le même esprit que ceux des *2D C-strings*. Ce changement de repère rend la chaîne obtenue invariante à la rotation image centrée en l'origine du nouveau repère, mais est dépendante de l'objet qui a servi de référence. Huang et Jean (1996) soulignent les faiblesses de cette solution, notamment l'absence dans la chaîne construite de l'objet de référence, réduisant ainsi l'information disponible lors de la recherche par similarité. Pour y remédier, ils proposent alors l'approche *RS-string* basée également les coordonnées polaires mais permettant d'intégrer l'objet centre du repère dans la chaîne. L'image est alors décrite par autant de *RS-strings* qu'il y a d'objets. La comparaison d'images passe donc par la comparaison des *RS-strings* associées aux objets en commun entre les deux images. Elle exploite les mêmes types de similarité que les *2D C-strings* mais est notablement plus coûteuse en temps CPU, du fait du nombre de chaînes à comparer. Dans Huang et Jean (1994), l'approche *2D C<sup>+</sup>-string* proposée rend encore plus précise la représentation des relations spatiales, en ajoutant dans la chaîne *2D C-string* une métrique associée aux objets et aux relations spatiales. Par exemple, dans la chaîne "A<sub>5</sub><<sub>8</sub>B<sub>2</sub>", "5" et "2" indiquent la taille des objets et "8" la distance qui les sépare. Ces informations supplémentaires permettent de définir six nouveaux types de similarité plus nuancés. Plus tard dans Jean et Lo (2004), ce sont au total onze types de similarité qui sont définis, en introduisant des contraintes sur les rapports de tailles et de distances entre paires d'objets. L'approche proposée par Lee et Chiu (2003), appelée *2D Z-string*, ramène la complexité de la chaîne à  $O(N)$  en supprimant complètement le mécanisme de découpage introduit pour les *2D G-strings*. Les sept opérateurs spatiaux sont donc directement utilisés, rendant ainsi inutile le mécanisme de découpage. Munis de métriques adéquates, les objets et opérateurs spatiaux mis en jeu dans la chaîne permettent de préserver la richesse de la description. L'algorithme de reconstruction et de comparaison des images est adapté à cette nouvelle représentation, et les types de similarité appliqués sont ceux développés pour les *2D C<sup>+</sup>-strings*. Par exemple, l'image de la figure 1 est représentée par la *2D Z-string* :  $(D_3 <_{1.2} (E_{17.3} \%_{1.8} A_{5.5}) /_{1.5} B_4 / C_2, D_{15.1} [(E_{2.1} /_{0.2} (A_{4.1} \%_{0.1} C_3) / B_6])$ . En effet, sur l'axe  $x$ , l'objet  $D$  de taille 3 est situé à une distance de 1.2 avant (<) l'ensemble formé par  $E$ , de taille 17.3 et  $A$  de taille 5.5,  $E$  contenant (%)  $A$  à partir d'une distance de 1.8, cet ensemble se chevauchant partiellement (/) avec  $B$  sur une distance de 1.5, lui-même ayant une frontière jointe (|) à la frontière de  $C$ . L'algorithme de reconstruction et de comparaison des

images est adapté à cette nouvelle représentation, et les types de similarité appliqués sont ceux développés pour les  $2D C^+$ -strings.

### Autres approches sous forme de chaînes de caractères

Les approches  $2D B$ -string,  $2D Be$ -string et  $RCOS$ -string, conceptuellement proches, utilisent une représentation symbolique des objets sensiblement différente de celle utilisée par les approches basées sur la  $2D$ -string : dans la chaîne, chaque symbole représente le début et la fin du REM de l'objet sur l'axe considéré, contrairement aux autres approches où le symbole traduit l'intervalle de projection de l'objet. L'approche  $2D B$ -string de Lee et al. (1992a) n'utilise aucun mécanisme de découpage et seul l'opérateur spatial "=" est exploité pour décrire les relations spatiales au sens de Lee et Hsu (1990). Sur chaque axe, la chaîne construite contient les symboles des objets par ordre croissant des bornes minimales et maximales de leur projection (chaque symbole est donc présent deux fois). Le rang  $R_{begin}$  (resp.  $R_{end}$ ) d'un objet est alors la position de la première (resp. seconde) occurrence de son symbole dans la chaîne, en tenant compte des éventuelles égalités. Par exemple, sur l'axe  $y$  de la figure 1,  $R_{begin}(D) = R_{begin}(E) = 1$  et  $R_{begin}(A) = 2$ . Les rangs des objets permettent de mettre en place trois types de similarité entre relations spatiales qui sont équivalents à ceux de Lee et Hsu (1992) exceptée la similarité de type-1 qui met ici en jeu les relations orthogonales. Cette représentation a les avantages d'être très simple tout en contenant l'information nécessaire à la détermination des 169 relations spatiales, d'éviter le découpage des objets et de générer des chaînes de longueur en  $O(N)$  et non  $O(N^2)$  comme la plupart des autres approches. Cependant, elle reporte le mécanisme d'inférence des relations spatiales complètes à la phase de recherche par similarité. Dans l'approche  $2D Be$ -string (*two dimensional begin-end boundary string*) de Wang (2003) aucun opérateur spatial n'est utilisé. Un objet virtuel  $e$ , appelé *dummy object*, est associé à chaque espace situé entre les projections des frontières des REM d'un même objet ou de deux objets sur l'axe des  $x$  ou l'axe des  $y$ . Une  $2D Be$ -string est de la forme  $(u, v) = (d_0x_1d_1x_2\dots d_{n-1}x_nd_n, d_0y_1d_1y_2\dots d_{n-1}y_nd_n)$  où  $d_i$  représente un objet virtuel  $e$  ou une chaîne vide,  $i = 0, 1, \dots, n$ , et  $x_i$  et  $y_i$  représentent la projection des frontières des REM des objets sur l'axe des  $x$  ou l'axe des  $y$ . Selon son emplacement dans la chaîne de caractère, l'objet  $e$  permet de représenter les 13 relations spatiales issues de opérateurs de Jungert (1988). Par exemple, l'image de la figure 1 est représentée par la  $2D Be$ -string suivante :  $(eD_b e D_e e E_b e A_b e A_e e B_b e E_e e B_e C_b e C_e e, e D_b E_b e A_b e C_b e E_e e C_e e A_e B_b e B_e e D_e e)$ .  $D_b$  représente la frontière de gauche (*begin*) du REM de l'objet  $D$  et  $D_e$  représente la frontière de droite (*end*) du REM de l'objet  $D$  sur l'axe des  $x$  ou des  $y$ . Sur l'axe des  $x$ ,  $d_0 = e$  car il existe un espace devant la frontière du REM de l'objet  $D$ . De même,  $d_{10} = e$  car il existe un espace après l'objet  $C$  sur l'axe des  $x$ . En revanche,  $d_8$  correspond à une chaîne vide car la frontière de fin du REM de l'objet  $B$  coïncide avec la frontière de début du REM de l'objet  $C$ . Le raisonnement est le même pour l'axe des  $y$ . Cette structure nécessite dans le pire des cas (si toutes les frontières sont distinctes)  $4N + 1$  symboles pour représenter une image de  $N$  objets, la complexité spatiale étant en  $O(N)$ . Wang (2003) propose un algorithme de similarité basé sur la comparaison LCS décrite dans la section 2.3. Cet algorithme permet, non seulement de retrouver les images similaires à une image requête, directement à partir de la  $2D Be$ -string, sans avoir à reconstruire l'image, mais permet aussi de calculer une mesure de similarité entre les images, qui correspond à la somme pondérée des similarités obtenues sur l'axe des  $x$  et sur l'axe des  $y$ , chacune étant calculée à partir de

la longueur des chaînes complètes et des LCS, avec ou sans objet virtuel, normalisée par le nombre maximum de symboles dans une chaîne.

La structure *RCOS-string* (pour *Relative Coordinates Oriented Symbolic (RCOS) string*) de Chang et Lee (1995) représente les images par une chaîne contenant la liste des noms des objets de l'image, suivies des coordonnées des points extrémité situés en bas à gauche et en haut à droite du REM de chacun de ces objets. Comparée à l'approche *2D-B-string*, cette structure occupe plus de place ( $5N$  octets, lorsque qu'un nom d'objet symbolique occupe 1 octet). Mais les auteurs indiquent que plus le nombre  $N$  d'objets augmente et plus le ratio de l'espace de stockage entre les deux approches diminue. Les expérimentations présentées dans l'article montrent également que le nombre de symboles dans la *RCOS-string* est plus faible que dans les approches *2D-G-string* ou *2D-C-string*. Elles montrent finalement que le nombre d'opérations nécessaires pour faire de la recherche d'images similaires est moindre pour la *RCOS-string* par rapport à la *2D-B-string*. Pour faire de la recherche par similarité, les auteurs définissent en particulier 2 arbres de décision, un pour chaque axe, permettant à partir des coordonnées de deux objets stockés dans la *RCOS-string*, de déduire leurs relations spatiales sur chaque axe. Chaque arbre représente les 13 relations spatiales 1D de Lee et Hsu (1990). Les deux arbres combinés permettant de représenter les 169 relations spatiales en 2D.

Dans Gudivada (1998), l'approche  $\Theta\mathcal{R}$ -string caractérise les relations géométriques existant entre objets (cf. section 2.2). Chaque objet, représenté par son centre de masse, est localisé dans l'image en coordonnées polaires, calculées dans un repère ayant pour origine le barycentre des centres de masse de tous les objets. Les objets sont alors ordonnés selon leur angle. La chaîne de caractères  $\Theta\mathcal{R}$ -string est construite à partir des symboles des objets ainsi triés. La version *augmentée* de cette structure apporte davantage d'informations puisque chaque objet intègre dans sa représentation les distances euclidiennes qui séparent son centre de masse de celui des deux objets qui lui sont voisins dans la chaîne. Le coût de construction d'une telle chaîne est en  $O(N \log N)$ . Notons que l'insertion d'un nouvel objet dans l'image nécessite le recalcul complet de la chaîne puisque l'origine du repère polaire en sera modifié. Pour comparer deux images, l'algorithme  $SIM_G$  proposé calcule un degré de similarité qui tient compte (a) des objets en commun dans les deux images (b) de la présence des mêmes voisins pour ces objets et (c) de la similarité de la distance qui sépare ces voisins. Il a les avantages d'être basé sur une représentation invariante à la translation et à la rotation image, de produire un degré de similarité (dans  $[0..1]$ ) entre images permettant le classement des réponses et d'être paramétrable selon trois poids qui permettent de favoriser plus ou moins les contraintes (a) (b) ou (c). Cependant, il a l'inconvénient majeur d'imposer la construction des  $\Theta\mathcal{R}$ -strings au moment de la comparaison, celles-ci étant construites uniquement à partir des objets en commun pour que la comparaison puisse être faite entre sous-images.

### 3.2 Représentation sous forme d'arbre

Au lieu de représenter une image symbolique par une chaîne de caractères, plusieurs auteurs proposent d'utiliser une représentation arborescente. Hsu et al. (1998a) indiquant que les *2D-C-string* sont efficaces pour la représentation des images mais le sont moins pour la recherche d'images par similarité, Hsu et al. (1999) ont proposé de représenter chaque axe d'une chaîne *2D-C-string* par un arbre, appelé *arbre 2D-C*. Les nœuds de cet arbre correspondent aux objets et les arcs sont étiquetés par les opérateurs spatiaux. Une distance entre arbres (ba-

sée sur la distance d'édition) est définie, des modifications étant appliquées aux *arbres 2D-C* représentant les images de la base pour les rendre comparables. Les auteurs utilisent cette distance dans un algorithme permettant de retrouver les images de la base contenant des sous-images de l'image requête. Hsu et al. (1998b) ont également étendu l'*arbre 2D-C* à la vidéo en représentant une séquence vidéo par un ensemble temporel ou ordonné d'*arbres 2D-C*.

Huang et al. (2007) proposent, quant à eux, une approche représentant les images par un arbre dont chaque nœud interne correspond à un objet dit critique, c'est-à-dire, selon les auteurs, un objet qui attire l'attention humaine (en général à cause de sa taille par rapport aux autres objets). Après identification d'un objet critique  $C$ , l'espace est décomposé en neuf zones à partir du prolongement des quatre frontières du REM de l'objet  $C$ , permettant de déduire les relations directionnelles des autres objets par rapport à  $C$ . À différence de l'approche de Chang (1991), seul l'objet critique est représenté par son REM, tous les autres étant placés dans l'espace par rapport au centre de masse de leur REM (voir figure 2). L'objet critique  $C$  devient racine d'un sous-arbre à neuf branches maximum. L'algorithme est récursif. Le premier objet critique choisi crée 9 zones dans lesquelles un autre objet critique peut à son tour être sélectionné. On obtient au final un arbre, nommée *Directional Division Tree (DDT)* représentant une hiérarchie visuelle des objets, la racine contenant l'objet le "plus critique" de l'image. Les sous-objets composant un autre objet sont représentés par des nœuds fils de l'arbre dont la racine correspond à l'objet composé. La similarité entre deux images est définie par une distance pondérée de distances topologiques et directionnelles entre chaque nœud et arc homologues des arbres DDT représentant les images à comparer.

## 4 Modèles explicites de représentation des relations spatiales

Cette section présente les modèles de représentation dites explicites, où toute l'information spatiale y est exprimée, aucune inférence n'étant nécessaire pour déduire les relations spatiales entre tous les objets. Nous avons classé ces modèles en quatre familles : les approches utilisant une représentation matricielle (section 4.1), celles représentant les images par des listes de  $n$ -uplets (section 4.2), celles les codant par des fichiers de signatures (section 4.3) et enfin par des arbres (section 4.4). A chaque fois, l'objectif est de minimiser l'espace de stockage, d'offrir des mécanismes de comparaison d'images ayant une complexité raisonnable, tout en permettant l'ajout ou le retrait d'objets sans avoir à reconstruire la représentation.

### 4.1 Représentation sous forme matricielle

Plusieurs approches représentent les relations spatiales des objets d'une image en utilisant une matrice carrée de taille  $N$ ,  $N$  étant le nombre d'objets dans l'image. A chaque fois, le choix d'une représentation matricielle est justifié par la mise en place d'un codage judicieux représentant la matrice avec un coût de stockage minimum tout en facilitant la comparaison entre deux matrices par des opérations simples efficaces. Les relations décrites par ces approches sont les relations directionnelles ou les relations spatiales par projections symboliques.

#### Codage matriciel des relations directionnelles

L'approche proposée par Chang (1991) représente les 9 relations directionnelles sous la forme d'un code entre 0 et 8, l'ensemble des codes obtenus étant stocké dans le triangle in-

férieur gauche d'une matrice  $M_{9DLT}$  (*9DLT Matrix* pour *nine direction lower-triangular matrix*). Le code  $M_{9DLT}(i, j)$  indique la relation directionnelle existant entre l'objet de la ligne  $i$  et l'objet de la colonne  $j$ . Par exemple, dans la figure 3, le code situé à la colonne  $A$ , ligne  $B$  indique que l'objet  $B$  est situé au nord est de l'objet  $A$ . Le coût de stockage de cette matrice est donc de  $\frac{N}{2}(N - 1)$  codes. Dans cette approche, deux images sont similaires spatialement si les objets qu'elles ont en commun vérifient les mêmes relations directionnelles, sinon elles ne le sont pas. Ce calcul de similarité spatiale exacte entre deux (sous-)images symboliques se fait par simples différences de matrices. Chang et Wu (1995) améliorent cette approche en permettant la comparaison d'une image par rapport à un ensemble d'images. Ici, chaque code  $M_{9DLT}(i, j)$  est représenté par un triplet  $(i, j, M_{9DLT}(i, j))$ , vu comme un point dans un espace tridimensionnel. La direction principale du nuage des  $\frac{N}{2}(N - 1)$  points 3D, exprimant toutes les relations existant dans une image symbolique, est ensuite extraite par ACP. Ce point unique, exploité comme index, permet de faire une recherche binaire dans l'ensemble des images. La méthode accélère considérablement la recherche dans un ensemble d'images des images ayant les mêmes relations spatiales, mais ne permet plus la comparaison de parties d'images, comme dans la méthode initiale.

$$M = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{ccccc} & A & B & C & D & E \\ A & - & - & - & - & - \\ B & 8 & - & - & - & - \\ C & 7 & 5 & - & - & - \\ D & 2 & 3 & 2 & - & - \\ E & 0 & 4 & 3 & 6 & - \end{array}$$

FIG. 3 – *9DLT Matrix* correspondant à l'image de la figure 1(d).

Chan et Chang (2001) représentent les relations directionnelles de la matrice 9DLT sous forme de chaîne de caractères. Il s'agit ici de décrire la relation directionnelle de code  $c$  entre deux objets  $A$  et  $B$  par la sous-chaîne " $ABC$ ". La chaîne complète décrivant l'image est donc composée de  $\frac{N}{2}(N - 1)$  sous-chaînes. Cette représentation est utilisée pour décrire des relations spatiales dans les vidéos. Afin de minimiser la redondance entre les chaînes obtenues entre trames successives et accélérer la recherche par similarité dans les bases de vidéos, les auteurs étendent la structure CCBT (*Common-Component Binary Tree*) de Chan et Chang (1998) à la *9DLT string*, la structure CCBT correspondant à un mécanisme de partage de parties communes entre arbres quaternaires représentant des images.

### Codage matriciel des relations spatiales par projections symboliques

Ce codage stocke pour chaque axe  $x$  et  $y$  toutes les relations spatiales existant entre  $N$  objets d'une image dans une matrice  $M$ . Soit  $A$  l'ensemble des codes décrivant les opérateurs spatiaux disponibles. La partie supérieure droite de  $M$  est renseignée par les opérateurs spatiaux sur l'axe  $y$  ( $r_{i,j}^y \in A, \forall i, j \ 1 \leq i < j \leq N$ ), alors que sa partie inférieure gauche est quant à elle renseignée par ceux de l'axe  $x$  ( $r_{i,j}^x \in A, \forall i, j \ 1 \leq j < i \leq N$ ). Plus formellement, on a :  $M(i, j) = r_{j,i}^x$  si  $i > j$ ;  $M(i, j) = r_{i,j}^y$  si  $i < j$  et  $M(i, j) = 0$  si  $i = j \ \forall r_{j,i}^x, r_{i,j}^y \in A, 1 \leq i, j \leq N$ .

L'ensemble  $A$  varie en fonction des approches décrites ci-dessous qui utilisent toutes une représentation des objets sous forme de REM. Les relations spatiales sont décrites à partir des

13 opérateurs spatiaux de Lee et Hsu (1990) et des 5 catégories qui en découlent (cf. section 2.2). Les mesures de similarité définies sont de type-0, type-1 et type-2 (cf. section 2.3).

La *PN Matrix* (pour *Prime-Number-based matrix*) de Chang et Yang (1997b) combine les avantages des approches *2D C-string* et *9DLT matrix*, permettant de décrire les 169 relations spatiales des *2D C-string* (et non présentes dans l’approche *9DLT matrix*), sans devoir procéder à leur complexe phase de découpage des objets. A partir de la classification des 169 relations possibles en 5 catégories spatiales, l’approche consiste à représenter dans la matrice chaque opérateur spatial par un nombre unique et judicieusement choisi comme le produit de nombres premiers. Cette représentation permet de coder la catégorie de la relation spatiale, tout en distinguant les opérateurs spatiaux mis en jeu au sein d’une même catégorie. Ensuite, 5 règles, impliquant des opérations numériques simples sur ces codes (modulo), permettent de retrouver efficacement les 5 catégories spatiales ; des opérations matricielles simples (différences entre matrices) permettent de déterminer la nature de la similarité (type-0, type-1 et type-2) entre les images comparées. Les nombres premiers à combiner sont choisis dans l’ensemble des 17 plus petits nombres premiers. Cette représentation requiert 20 bits par code, et donc en pratique  $3N^2$  octets pour la matrice entière. Ce codage est amélioré dans *GPN Matrix* (pour *Generalized Prime-Number-based matrix*) de Chang et al. (2001) en réduisant l’espace de stockage nécessaire, ici 14 bits seulement sont nécessaires pour stocker le codage des 13 opérateurs spatiaux à partir de 12 nombres premiers, Cette représentation permet aussi de réduire de moitié le temps de comparaison entre deux images, pour la similarité type-0. Une version *GPN\* Matrix* est aussi proposée : celle-ci requiert 11 bits par code, soit 3 bits de moins, mais au prix d’un calcul sensiblement plus complexe de la règle testant l’appartenance à la catégorie *join*, au moment de la comparaison entre images. Dans Chang et al. (2000) les 13 opérateurs spatiaux sont représentés par un code entre 1 et 13 stocké sur 4 bits dans une *Unique-ID-based Matrix* (matrice UID) de  $N^2$  octets au total. La figure 4 donne un exemple d’une telle matrice. Les trois types de similarité sont traités, par de simples différences de matrices (pour le type-0) et aussi par des produits matriciels particuliers (impliquant  $\frac{N}{2}(N - 1)$  multiplications de codes) pour les types 1 et 2. Cette nouvelle représentation diminue d’un facteur 10 le temps de comparaison par rapport à l’approche *2D C-string*.

$$M = \begin{matrix} & \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \begin{bmatrix} 0 & 3 & 9 & 13 & 6 \\ 1 & 0 & 2 & 13 & 2 \\ 1 & 3 & 0 & 13 & 6 \\ 2 & 2 & 2 & 0 & 8 \\ 13 & 6 & 2 & 1 & 0 \end{bmatrix} \end{matrix}$$

FIG. 4 – *UDI Matrix* correspondant à l’image de la figure 1(b).

Dans l’approche de Chang et al. (2003), les relations spatiales sont répertoriées dans une matrice *BP Matrix* (pour *Bit-Pattern-based matrix*), dans laquelle les opérateurs spatiaux sont représentés par un codage binaire sur 12 bits. Ici, les 5 catégories spatiales sont réparties sur les 9 premiers bits de poids faible (par exemple, le bit 0 représente la catégorie *disjoin* et les bits 1 et 2 la catégorie *join*), les 3 derniers permettant de distinguer les opérateurs spatiaux mis en jeu au sein d’une même catégorie. Des opérations logiques bit à bit et des opérations matricielles simples (identiques à celles de l’approche *PN Matrix*) permettent de déterminer la nature de

la similarité entre les images comparées. L'approche a été évaluée et comparée à l'approche *GPN Matrix* sur 2000 images symboliques contenant 20 objets au maximum : sur l'ensemble de la base, les temps de comparaison montrent que l'approche *BP matrix* est notablement plus rapide, quel que soit le type de similarité (type-0, type-1, type-2).

L'approche *Unique-Bit-Pattern Matrix* de Yeh et Chang (2008) utilise 16 bits pour représenter les 13 relations spatiales. La figure 5 présente un exemple d'une telle matrice. Le modèle tient compte de cinq transformations géométriques susceptibles d'intervenir entre deux images et pouvant donc perturber les relations spatiales mises en jeu : trois rotations orthogonales ( $90^\circ$ ,  $180^\circ$  et  $270^\circ$ ) et deux symétries (horizontale et verticale). En effet, le codage est choisi de manière à minimiser le coût de calcul faisant passer du code d'une relation spatiale à celui de cette relation après qu'elle ait subi une telle transformation. Le calcul met en jeu de simples décalages de bits et une transposition de matrice. Cependant, pour être invariante à ces transformations, la comparaison entre deux matrices  $Q$  et  $P$  nécessite aussi de comparer  $P$  aux cinq transformées de  $Q$  associées. Les auteurs présentent une évaluation de l'approche pour 100 objets et 500000 images qui montre notamment sa supériorité en termes de temps de calcul des matrices transformées, par rapport aux approches de Nabil et al. (1996) et de Petraglia et al. (2001) qui explorent également la description des relations spatiales face à ces transformations.

$$M = \begin{matrix} & \begin{matrix} A & B & C & D & E \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} & \left[ \begin{array}{ccccc} 0 & 0000\ 0000\ 0000\ 0100 & 0000\ 0011\ 0000\ 0000 & 0011\ 0000\ 0000\ 0000 & 0000\ 0000\ 0010\ 0000 \\ 0000\ 0000\ 0000\ 0001 & 0 & 0000\ 0000\ 0000\ 0010 & 0011\ 0000\ 0000\ 0000 & 0000\ 0000\ 0000\ 0010 \\ 0000\ 0000\ 0000\ 0001 & 0000\ 0000\ 0000\ 0100 & 0 & 0011\ 0000\ 0000\ 0000 & 0000\ 0000\ 0010\ 0000 \\ 0000\ 0000\ 0000\ 0010 & 0000\ 0000\ 0000\ 0010 & 0000\ 0000\ 0000\ 0010 & 0 & 0000\ 0000\ 1000\ 0000 \\ 0011\ 0000\ 0000\ 0000 & 0000\ 0000\ 0010\ 0000 & 0000\ 0000\ 0000\ 0010 & 0000\ 0000\ 0000\ 0001 & 0 \end{array} \right] \end{matrix}$$

FIG. 5 – *UBP Matrix* correspondant à l'image de la figure 1(b).

## 4.2 Représentation sous forme de liste de $n$ -uplets

Cette section présente les approches basées sur des listes de  $n$ -uplets. Les premières approches décrites caractérisent les relations existant entre *couples* d'objets. Quant à la dernière approche, elle caractérise les relations géométriques existant entre *triplets* d'objets.

### Codage par couples d'objets

Plusieurs approches représentent les images symboliques par des triplets  $(O_i, O_j, R_{ij})$ , où  $O_i$  et  $O_j$  correspondent à des objets et  $R_{ij}$  au codage de la relation spatiale entre ces objets. Une image est ainsi représentée par  $N(N - 1)/2$  triplets (la relation complémentaire  $R_{ji}$  n'étant pas codée). En général, ces approches indexent les triplets par des fonctions de hachage. Chang et Lee (1991) utilisent la relation directionnelle et l'algorithme de Cook et Oldehoeft (1982) pour la fonction de hachage prenant en entrée un triplet  $(O_i, O_j, R_{ij})$  et retournant une liste  $L_{ij}$  d'identifiants d'images dans lesquelles les objets  $O_i$  et  $O_j$  vérifient la relation directionnelle  $R_{ij}$ . Dans cette approche, si une image requête est représentée par  $n$  triplets  $(O_i, O_j, R_{ij})$ , alors la structure d'index est interrogée  $n$  fois, retournant  $n$  listes  $L_{ij}$ . Les images considérées comme similaires à l'image requête appartiennent à l'intersection des

listes  $L_{ij}$ . Cette approche a été implémentée sur une base exemple variant de 15000 images symboliques contenant environ une vingtaine d'objets différents. Les résultats expérimentaux comparant la structure de hachage à une comparaison séquentielle sont présentés dans Chang et Lee (1998b), montrant un gain de plus de 5000ms en moyenne pour 100 requêtes. La structure de hachage proposée par Chang et Lee (1991) étant statique, Sabharwal et Bratia (1997) ont proposé de relaxer certaines contraintes de cette structure (en stockant des informations complémentaires permettant de résoudre les éventuelles collisions) afin de permettre les ajouts et les suppressions dynamiques d'images dans la base. Chang et Wu (1992) et Wu et Chang (1994) ont également proposé une amélioration de l'approche de Chang et Lee (1991) en utilisant des algorithmes de hachage différents (l'algorithme de Ziegler (1977) pour les premiers et l'algorithme de Comer et O'Donnell (1982) pour les deuxièmes), permettant d'ajouter des opérateurs spatiaux sans avoir à reconstruire toute la structure de hachage. De plus, dans ces approches, la recherche des images similaires à une image requête ne se fait plus par intersection des listes  $L_{ij}$ , mais par recherche des identificateurs apparaissant dans le maximum de listes. Petrakis et Orphanoudakis (1996) présentent une extension de l'approche de Chang et Lee (1991) en regroupant les objets d'une image en sous ensembles (appelés *image subsets*), la taille de ces sous-ensembles variant de 2 à  $K_{Max}$ . Une structure de hachage est créée pour chaque sous-ensemble de taille  $K$ ,  $K \in [2, K_{Max}]$  permettant, selon les auteurs, de limiter le nombre de collisions par rapport à l'approche de Chang et Lee (1991). Les expérimentations sur une base de 1000 images montrent, pour certains requêtes, que leur approche offre de meilleures performances. Néanmoins, comme le souligne El-Kwae (2000), elle nécessite plus d'espace de stockage. Récemment, Mughal et al. (2007) ont proposé d'utiliser 3 fonctions de hachage différentes pour chacun des éléments des triplets indexés. Zhou et Ang (1997) proposent quand à eux leur propre algorithme de hachage pour indexer les triplets  $(O_i, O_j, R_{ij})$  avec  $R_{ij}$  prenant en compte, cette fois-ci, les 41 relations topo-directionnelles entre les objets (représentées à la figure 7(b) en annexe), en combinant les relations directionnelles avec 5 relations topologiques (*disjoint*, *tangent*, *partially overlap*, *contain* et *inside*). Le centre de masse et la surface des objets sont utilisés pour obtenir la relation directionnelle et déterminer la relation topologique ; par exemple, la relation spatiale entre les objets  $A$  et  $B$  de l'image de la figure 1(a) sera codée par  $(A, B, 2)$ , celle entre  $A$  et  $E$  par  $(A, E, 24)$ .

La méthode *9D-SPA* de Huang et Lee (2004) est basée sur les relations directionnelles et topologiques. A chaque objet  $O_i$  est assigné une valeur  $i$ ,  $1 \leq i \leq N$  où  $N$  est le nombre d'objets de la base. La relation spatiale entre deux objets  $O_i$  et  $O_j$  est représentée par un quadruplet  $(O_{ij}, D_{ij}, D_{ji}, T_{ij})$  où  $O_{ij}$  correspond à un codage de la paire d'objets  $(O_i, O_j)$ ,  $D_{ij}$  représente la relation directionnelle entre  $O_i$  et  $O_j$  en prenant  $O_i$  comme référence,  $D_{ji}$  la relation directionnelle entre  $O_i$  et  $O_j$  en prenant  $O_j$  comme référence et  $T_{ij}$  représente la relation topologique entre  $O_i$  et  $O_j$ . Par exemple, si les objets  $A, B$  de la figure 1(a) sont respectivement étiquetés par 1 et 2, la relation spatiale entre  $A$  et  $B$  est codée par le quadruplet  $(1, 2, 1, 32, 2, 0)$ . Une image est donc représentée par l'ensemble des quadruplets des relations spatiales entre tous les couples d'objets de l'image. Les auteurs définissent deux mesures de similarité entre les images, une similarité sur les composantes directionnelles et une similarité sur les composantes topologiques. Chacune de ces similarités est définie par une distance normalisée des distances entre les quadruplets homologues de deux images. Les auteurs définissent également une structure d'index à deux niveaux pour accélérer la recherche. Deux listes sont créées pour chaque paire d'objets  $O_{ij}$ . La première liste associe à chaque relation direc-

tionnelle la liste des images dans lesquelles  $O_{ij}$  satisfont cette relation. La deuxième indexe les relations topologiques du couple d'objets de la même manière.

L'approche *2D-PIR* (*2D Projection Interval Relationships*) de Nabil et al. (1995) caractérise les relations spatiales par les 13 opérateurs spatiaux 1D de Lee et Hsu (1990), enrichis par les opérateurs topologiques selon Egenhofer et Franzosa (1991). L'objectif est de réduire les ambiguïtés des opérateurs 1D, notamment la non intersection de deux objets bien que leurs REM soient en intersection. La relation spatiale entre deux objets est représentée par le triplet  $(\delta, \chi, \psi)$ , où  $\delta$  est le code de la relation topologique,  $\chi$  et  $\psi$  ceux des opérateurs spatiaux 1D sur chaque axe. Par exemple la relation spatiale entre les objets  $A$  et  $B$  de la figure 1(a) est codée par  $(dt, <, < *)_{AB}$ ,  $dt$  signifiant que  $A$  et  $B$  sont disjoints, et celle entre  $A$  et  $E$  par  $(ov, \%*, /)_{AE}$ ,  $ov$  signifiant que  $A$  chevauche (*overlaps*)  $B$ . Pour tous les couples d'objets de l'image, chaque triplet obtenu est inséré dans un graphe orienté, implémenté efficacement sous forme de listes chaînées. L'approche donne la possibilité de réaliser une recherche exacte de sur-images ou sous-images par comparaison des opérateurs spatiaux entre couples d'objets. La complexité de l'algorithme de comparaison est  $O(N^2)$ . Dans Nabil et al. (1996), les auteurs améliorent cette représentation en l'adaptant à certaines transformations de l'image (rotation et symétrie) et à la recherche par similarité en utilisant une mesure de similarité pour comparer deux relations qui correspond à une combinaison de la distance topologique définie par Egenhofer et Al-Taha (1992) avec la distance entre intervalles selon Freksa (1992).

L'approche "*Virtual Image*" de Petraglia et al. (2001) consiste à représenter les relations spatiales de Lee et Hsu (1990) par une liste de relations spatiales 1D atomiques pour chaque axe, de la forme  $O_i r O_j$ , où  $O_i$  et  $O_j$  sont deux objets et  $r$  l'opérateur spatial 1D selon l'axe des  $x$  ou  $y$ . Par exemple le codage des relations spatiales selon l'axe des  $x$  de l'image de la figure 1(a) est  $\{A < B, A < C, A < *D, A \% * E, \dots\}_x$ . L'originalité de l'approche réside dans la mesure de similarité entre images associée à cette représentation. Cette mesure intègre la distance entre intervalles 1D de Freksa (1992), déjà utilisée par Nabil et al. (1996) pour comparer deux relations spatiales représentées par les projections symboliques. Ici une image requête est définie par un ensemble d'objets obligatoires, i.e. qui doivent être retrouvés dans l'images cherchée, un ensemble d'objets optionnels, qui viendront augmenter la similarité s'ils sont aussi retrouvés, et des relations spatiales atomiques entre tous ces objets. Un score de comparaison, compris dans  $[0..1]$ , est calculé entre deux images : il est fonction du nombre d'objets optionnels retrouvés et des distances entre les relations spatiales associées à ces objets mais aussi associées aux objets obligatoires. Une image ne contenant pas les objets obligatoires aura un score nul. Les auteurs proposent également une amélioration de leur approche, appelée *R-virtual image*, qui consiste à rendre leur modèle invariant à la rotation image utilisant une technique identique à celle de l'approche *2-R string* de Petraglia et al. (1993). L'approche complète est expérimentée dans l'environnement IME (pour *Image Management Environment*), dédié à l'imagerie médicale.

### Codage par triplets d'objets

Contrairement aux autres approches directionnelles et topologiques de Zhou et Ang (1997) ou de Huang et Lee (2004) où la représentation des images symboliques est basée sur les relations spatiales entre deux objets, Guru et Nagabhusan (2001) proposent la méthode *TSR*, basée sur les relations géométriques entre trois objets non colinéaires d'une image. Le triangle déterminé par trois objets est représenté par un des six quadruplets  $(L_a, L_b, L_c, \theta)$  où  $L_a$ ,

$L_b$  et  $L_c$  sont les étiquettes de centres de masse des objets  $A$ ,  $B$  et  $C$ , et  $\theta$  est l'angle le plus petit entre la médiane issue du sommet  $A$  et le côté  $BC$  du triangle. Par exemple, si les objets  $A$ ,  $B$  et  $C$  de la figure 1(a) sont respectivement étiquetés par 1, 2 et 3, le triplet est codé par  $(3, 2, 1, 67.5)$ . Une image est ainsi représentée par l'ensemble de quadruplets  $S$  représentant les relations spatiales triangulaires entre tous les possibles triplets d'objets non colinéaires. Cette représentation est invariante à la rotation, à la translation et à l'échelle. Pour minimiser l'espace du stockage et diminuer le temps de recherche, les auteurs considèrent chaque quadruplet comme un point dans un espace  $4D$  et réalisent une ACP sur le nuage des points qui représentent les quadruplets. L'ensemble des directions principales obtenu après ACP est stocké dans une séquence ordonnée permettant ainsi de réaliser une recherche binaire. Cette approche ne peut être utilisée que pour une recherche globale exacte d'image. De plus, différents nuages des points peuvent produire la même direction principale après ACP, créant de l'ambiguïté. C'est pourquoi, Guru et al. (2003) proposent une amélioration dans laquelle chaque quadruplet est représenté par une clé unique stockée dans un arbre B, permettant de rechercher des sous-images, avec néanmoins un nombre des clés à stocker d'ordre  $O(N^3)$  où  $N$  est le nombre d'étiquettes de la base. Via cette nouvelle structure nommée *TRS+B-tree*, les auteurs indiquent passer d'un temps de recherche en  $O(NU_Q U_M)$  à un temps de recherche en  $O(U_Q \log_r N)$ , où  $U_Q$  correspond au nombre de clés distinctes dans l'image requête,  $U_M$  au nombre moyen de clés distinctes dans les images de la base,  $N$  est le nombre total d'images stockées dans la base et  $r$  est l'ordre de l'arbre B. Guru et Punitha (2007) montrent plusieurs résultats d'expérimentations sur des bases d'images synthétiques et des bases d'images réelles de visages. Pour réduire le volume de stockage, Punitha et Guru (2005) proposent une méthode où, pour chaque image, est stocké un triplet  $(K, \mu, \sigma)$  où  $K$  est le nombre des clés,  $\mu$  la valeur moyenne des valeurs des clés et  $\sigma$  l'écart-type. L'ensemble des triplets est stocké dans une séquence ordonnée. On revient à une méthode d'ordre  $O(\log I)$  où  $I$  est le nombre d'images de la base, qui permet seulement la recherche globale exacte des images.

### 4.3 Représentation par des fichiers de signatures

Plusieurs approches représentent les images symboliques par des fichiers de signatures, dans l'optique de faciliter les comparaisons d'images et la recherche par similarité. Les fichiers de signatures ont été utilisées à l'origine pour faire de la recherche documentaire. Une signature correspond à une chaîne binaire de taille fixe, produite par l'application d'une fonction de hachage sur une valeur. Une comparaison de signatures revient donc à un opération sur les bits de complexité  $O(1)$ . Les approches présentées dans cette section étendent plusieurs modèles de représentation des relations spatiales (2D-string ou matrice) présentées dans les sections précédentes, à la représentation sous forme de signatures.

Lee et al. (1992b) sont parmi les premiers auteurs à utiliser les fichiers de signatures pour indexer les images symboliques en étendant la représentation des 2D-B-string (voir section 3.1) en fichiers de signatures. Ils proposent 4 fichiers de signatures, un pour chaque type de requêtes : la recherche par objets et les recherches par similarité de types 0, 1 et 2. Chacun des fichiers de signature est généré à partir des 2D-B-strings représentant les images de la base. Pour permettre de faire de la recherche par objets, chaque image est associée à une signature calculée à partir des signatures des objets qu'elle contient. Pour permettre la recherche par similarité de type  $i$  ( $i \in \{0, 1, 2\}$ ), chaque image est associée à une signature combinant les signatures des couples d'objets appartenant à une des catégories utilisées pour calculer la si-

milarité de type  $i$ . Quatre fonctions de hachage sont utilisées afin qu'une même image aient des signatures différentes pour chaque type de recherche. Le tout est au final intégré dans une structure de signatures à deux niveaux, appelée *block signature* et basée sur l'approche de Sacks-Davis et al. (1987), dont les nœuds feuilles pointent vers les 2D-B-strings des images. Les signatures, en particulier celles se rapportant aux types de similarité, étant définies à partir des catégories des relations spatiales deux à deux, l'approche de Lee et al. (1992b) transforme l'approche des 2D-B-string, à l'origine implicite, en approche explicite.

El-Kwae et Kabuka (2000) proposent, quant à eux, une structure de signatures multi-niveaux (*Two Signature Multi-Level Signature File - 2SMLSF*) pour indexer des images représentées par des 2D-strings. Le nombre de niveaux de la structure dépend du nombre d'images dans la base. Au niveau des feuilles, la structure contient des signatures d'images calculées à partir des relations spatiales des objets deux à deux contenus dans chaque image. Les feuilles pointent vers les 2D-strings des images de la base. Les signatures des niveaux supérieurs sont uniquement calculées à partir des signatures des objets contenus dans les images. El-Kwae (2000) donne des résultats d'expérimentation sur une base de plus de 130000 images.

Contrairement à l'approche de Lee et al. (1992b) qui crée une signature pour chaque type de similarité, Chang et Yang (1997a) se limitent à des signatures pour la similarité de type 2. Pour calculer les signatures, les auteurs utilisent le codage matriciel (voir section 4.1) des relations spatiales déduites des 2D-C-strings représentant les images. La matrice des opérateurs étendus de Lee et Hsu (1990) est représentée par une matrice dite réduite (*Reduced Spatial Matrix - RSM*), similaire à la matrice UID. Puis la matrice RSM est représentée par deux chaînes spatiales, représentant les codes  $O_i O_j R_{ij}$  déduits de la matrice pour l'axe des  $x$  et l'axe des  $y$ . Ces chaînes sont ensuite codées par des signatures. Dans cette approche, chaque image est représentée par une signature composée de deux parties  $RS_1$  et  $RS_2$ .  $RS_1$  contient deux segments  $RS_1^x$  et  $RS_1^y$  de 13 bits chacun, correspondant à des drapeaux indiquant l'existence ou l'absence des 13 opérateurs spatiaux respectivement sur l'axe des  $x$  ou l'axe des  $y$ .  $RS_1$  contient également deux segments  $RS_2^x$  et  $RS_2^y$ , de 13 bits chacun également, chaque bit  $i$  correspondant à l'union des signatures des paires d'objets associés au même opérateur spatial  $i$ . Les auteurs proposent également une autre méthode de codage, remplaçant  $RS_2$  par 26 nombres premiers, le nombre de fausses alarmes étant selon les auteurs moins important avec les nombres premiers qu'avec les chaînes de bits. En effet, l'union de deux chaînes de bits peut être identique à une autre chaîne de bits existante, alors que la production de deux nombres premiers ne donnera jamais un nombre premier déjà utilisé.

Chang et al. (2002) proposent une extension de l'approche de la UID Matrix de Chang et al. (2000) (voir section 4.1) en codant la matrice UID par une signature de type de similarité 2. Ils utilisent un algorithme et un mode de représentation similaire à celui de Chang et Yang (1997a), précédemment cité. Les auteurs proposent de plus plusieurs algorithmes permettant de convertir les signatures de type 2 en signatures sur les objets ou sur les autres types de similarité (types 0 et 1). Ils présentent des résultats d'expérimentation montrant de meilleurs résultats de recherche et un espace de stockage moindre que l'approche de Lee et al. (1992b).

Récemment, Yeh et Chang (2006) ont proposé une approche de représentation des images par signature permettant de prendre en compte 289 relations spatiales. Les auteurs définissent pour cela 6 types de similarités. Cette approche combine les relations définies dans l'approche des 2D-C-strings de Lee et Hsu (1990, 1991), l'approche 9DLT de Chang (1991) et l'approche de Zhou et Ang (1997) (voir section 4.2) en représentant chaque image par une signature com-

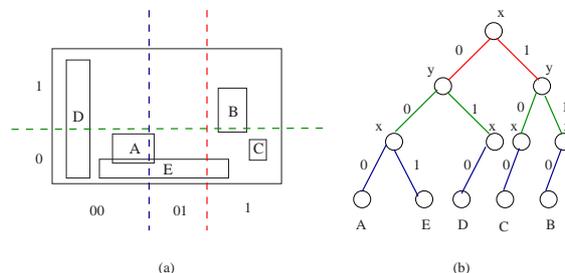
posée de quatre sous-signature. La première sous-signature, notée SCS (pour *Spatial Category string*) code les catégories spatiales. Les auteurs utilisent pour cela une approche similaire à celle de Chang et Yang (1997a), précédemment citée, en utilisant une matrice réduite. La deuxième sous-signature, notée DCS (pour *nine-Direction Code string*) code les relations directionnelles à partir de la matrice 9DLT. La troisième sous-signature, notée INS (pour *Identification Number String*) code des catégories spatio-directionnelle. Cette sous-signature complète les deux précédentes, les auteurs ayant analysé les différentes relations spatiales associées à un couple (catégorie spatiale, code directionnel), ce qui leur permet alors de déduire 289 relations spatiales auxquelles appartiennent les 169 relations définies par Lee et Hsu (1990). Finalement, la quatrième sous-signature, notée TSR (pour *Topological Relationship String*) code les catégories topologiques. Les auteurs comparent leur approche à celles étendant la 2D-B-string et la IUD-Matrix et indiquent obtenir de meilleures performances au niveau du nombre de réponses correctes à une requête de similarité.

#### 4.4 Représentation sous forme d'arbre

Li et Qu (1998) et Chang et Lee (1998a) représentent les images symboliques par un arbre binaire (*bin-tree*) équilibré correspondant à une division récursive de l'espace en 2 sous-espaces (ou *bins*) alternativement selon les axes  $x$  ou  $y$ , jusqu'à ce que les sous-espaces finaux (correspondant aux nœuds feuilles de l'arbre) ne contiennent qu'un seul objet (dans sa totalité ou en partie). Les nœuds internes sont de deux catégories, les nœuds  $x$  et les nœuds  $y$ , chaque niveau de l'arbre contenant des nœuds internes de même type. Les objets représentés par un sous-arbre gauche ayant pour racine un nœud  $x$  (resp.  $y$ ) sont situés, dans l'image, à gauche (resp. au-dessus) des objets représentés par le sous-arbre droit. La figure 6 donne un exemple d'arbre binaire représentant l'image de la figure 1(b). Une même image pouvant être représentée par différents arbres, Li et Qu (1998) basent leur algorithme de mise en correspondance de deux images non pas sur la comparaison des arbres mais sur un algorithme comparant la traversée des arbres, permettant de rechercher les images de la base contenant exactement l'image requête. Chang et Lee (1998a) proposent, quant à eux, un découpage de l'image en partitions de taille égale et un algorithme permettant de définir des similarités de type-0, 1 ou 2, telles que celles de Chang et al. (1987). Par rapport aux approches des *2D-strings*, Li et Qu (1998) et Chang et Lee (1998a) indiquent que la représentation des images par un arbre binaire offre une meilleure complexité en temps et en espace de stockage. La structure en arbre permet de plus d'ajouter des objets dans une image sans avoir à recréer entièrement l'arbre, alors qu'un même ajout dans une *2D-string* impose la reconstruction complète de la chaîne. L'arbre *db* (*Dynamic Bin-tree*) de Li et Qu (1998) est utilisé dans le système DISIMA de Oria et al. (2000).

Les structures d'accès spatial (*Spatial Access Method - SAM*)<sup>1</sup>, utilisées dans les systèmes d'information géographiques, peuvent être utilisées pour représenter les relations spatiales entre les objets d'une image. Parmi ses structures, nous pouvons citer les arbres *k-d* de Bentley (1975), tous les arbres quaternaires sur les points, comme le *Point quadtree* de Finkel et Bentley (1974), sur les rectangles, comme l'arbre quaternaire *MX-CIF* de Kedem (1982), l'arbre *R* de Guttman (1984) et ses variantes, comme le *Packed-R-tree* de Roussopoulos et Leifker (1985), etc. – voir Samet (1984) pour les arbres quaternaires et Papadopoulos (2005) pour une synthèse sur les arbres *R* et leurs variantes, ou Samet (2006) donne une synthèse globale.

<sup>1</sup>Une démonstration de ces structures est présentée à l'adresse <http://donar.umiacs.umd.edu/quadtree/index.html>



**FIG. 6** – L'image de la figure 1 récursivement découpée en sous-espaces alternativement selon l'axe des  $x$  et des  $y$  (a) et représentée par un Dynamic Bin-tree (b).

Chaque image de la base est représentée par une structure d'accès spatiale. Il y a donc autant de structures que d'images. Ces structures permettent de représenter certaines relations spatiales. Papadias et al. (1994) présentent par exemple comment faire de la recherche d'images par relations directionnelles en utilisant l'arbre R. Papadias et al. (1998) présentent plusieurs algorithmes permettant de faire des requêtes spatiales sur les relations topologiques (*inside*, *overlap*), directionnelles ou celles dites basées sur une distance qualitative (*near*, *far*) en utilisant un arbre R. Récemment, Osborn et Barker (2006) ont proposé une structure, le *2DR-tree*, permettant de représenter les relations directionnelles entre les REM des objets d'une image.

L'arbre R peut également être utilisé pour indexer les images et accélérer la recherche. Petrakis et Faloutsos (1997) et Petrakis et al. (2002) utilisent par exemple un arbre R pour indexer des images représentées par des graphes ARG, chaque graphe ARG étant représenté par un vecteur à  $f$ -dimensions. Shan et Lee (1998) proposent quant à eux un mécanisme de filtrage pour accélérer la recherche d'images similaires représentées par des 2D-strings. Dans cette approche, chaque image est représentée par trois rectangles multidimensionnels, un pour chaque type de similarité, chaque rectangle regroupant les intervalles des codes déduits de la 2D-string représentant l'image pour un type de similarité donné. Un arbre R indexe au final les rectangles multidimensionnels pour accélérer les recherches d'images similaires.

## 5 Synthèse comparative et conclusions

Dans cet article, nous avons rappelé les fondements et fait le point sur les approches permettant de décrire les relations spatiales entre objets dans les images, et pour certaines dans les vidéos. Nous avons choisi d'investiguer les images dites symboliques, pour lesquelles les objets d'intérêt sont déjà détectés, segmentés et identifiés (labelisés). Nous avons répertorié les approches rencontrées selon leur nature implicite ou explicite. D'autres typologies auraient pu être choisies, mais celle-ci a les avantages de produire une partition de ces approches et de mettre en lumière certaines propriétés fondamentales : les approches explicites requièrent un minimum de raisonnement spatial au moment de la recherche par similarité, sont dynamiques (un objet peut être ajouté dans l'image sans nécessiter la reconstruction de la représentation), au prix d'un volume de stockage plus important puisque toutes les relations entre couples d'objets (ou triplets d'objets pour l'approche *TSR*) doivent être stockées. Au contraire, les approches implicites sont plus compactes, mais requièrent des règles d'inférence au moment de la re-

cherche pour déduire certaines des relations et supposent de reconstruire la représentation si un objet est ajouté ou retiré. L'état de l'art relatif à la problématique des relations spatiales dans les images symboliques est très important (plus d'une centaine de références sont répertoriées dans cet article), justifiant cette étude. Cette section synthétise les différents modèles présentés précédemment et présente plusieurs tableaux comparatifs. Les critères de comparaison des approches ont été choisis parmi les critères considérés comme importants pour les modèles de recherche d'images par similarité.

### 5.1 Types de relations spatiales et opérateurs spatiaux associés

Toutes les solutions rencontrées permettent de décrire un ou plusieurs types de relations spatiales, parmi les cinq suivants : relations directionnelles, topologiques, géométriques, orthogonales et projections symboliques. La table 3 répertorie les principales caractéristiques des types de relations spatiales décrites dans la section 2.2, en termes de degrés de liberté ou nombre de relations possibles, et de codage ou opérateurs mis en jeu. Par exemple, il existe 8 relations topologiques et 169 relations spatiales issues des projections symboliques. En ce qui concerne les relations géométriques, le codage employé dépend fortement du modèle utilisé. Nous renvoyons le lecteur aux approches géométriques correspondantes, par exemple l'approche sous forme de string  $\Theta\mathcal{R}$ -string de la section 3.1 et l'approche *TSR* de la section 4.2. Les relations orthogonales sont quant à elles déduites de règles à partir des approches directionnelles, symboliques ou géométriques (se reporter à la section 2.2). En combinant plusieurs de ces types, certaines approches, notamment celles de Zhou et Ang (1997) et de Yeh et Chang (2006), permettent d'enrichir le nombre de relations possibles et désambigüisent certaines d'entre elles.

Type	d.d.l.	Codage / Opérateurs spatiaux
Relations directionnelles ( $R_D$ )	9	{1, 2, 3, 4, 5, 6, 7, 8, 9}
Relations topologiques ( $R_T$ )	8	{dt, to, in, ct, ov, co, cb, eq} <sup>1</sup>
Projections symboliques ( $R_S$ )	169	{=, <, /, [, ], %, /, <*, /*, [*, ]*, %*, /* }
Relations géométriques ( $R_G$ )	-	Divers codages
Relations orthogonales ( $R_O$ )	4	( $R_D$ ou $R_S$ ou $R_G$ ) + règles
2D-PIR - Nabil et al. (1995)	< 1352	$R_S + R_T$
Zhou et Ang (1997)	41	$R_T + R_D$
9D-SPA - Huang et Lee (2004)	< 72	$R_T + R_D$
Yeh et Chang (2006)	289	$R_T + R_D + R_S$

**TAB. 3** – Récapitulatif des types de relations spatiales 2D existantes (notées  $R_D$ ,  $R_T$ ,  $R_S$ ,  $R_G$  et  $R_O$ ) et des approches combinant plusieurs de ces types.

La deuxième partie du tableau récapitule les approches qui combinent plusieurs de ces types de relations. Dans les approches de Zhou et Ang (1997) et de Yeh et Chang (2006), les auteurs comptabilisent le nombre minimum de relations spatiales engendrées par de telles combinaisons (resp. 41 et 289), certaines combinaisons étant en effet redondantes. Par exemple, si la catégorie *disjoin* des projections symboliques est vérifiée pour deux REM d'objets, alors

<sup>1</sup>Pour les relations topologiques, le codage utilisé représente respectivement : *disjoint*, *touch*, *inside*, *contains*, *overlap*, *covers*, *covered-by* et *equal* (notations de Nabil et al. (1995)).

la relation topologique *disjoin* entre ces objets l'est aussi nécessairement. Ces informations n'étant pas données pour les deux autres approches *2D-PIR* et *9D-SPA*, nous faisons figurer ici le nombre théorique de relations spatiales engendrées par ces combinaisons, certaines de ces relations n'étant pas pertinentes.

Les représentations spatiales mises en jeu dans ces approches les rendent invariantes à la translation, puisqu'aucune position absolue d'objet n'est stockée. La plupart d'entre elles sont invariantes au changement d'échelle, exceptées les approches *2D C<sup>+</sup>-string* et *2D Z-string* qui intègrent une métrique dans leur représentation. Les approches impliquant les relations directionnelles, orthogonales, ou les projections symboliques ne sont bien sûr pas invariantes à la rotation image ou à la symétrie. Cependant, des améliorations ont été réalisées de manière à traiter ces transformations, notamment dans les approches suivantes : *2-R string*, *RS-strings*, *ΘR-string*, *UBP Matrix* (rotations simples), "Virtual Image" améliorée, *2D-PIR* améliorée et *TSR*. Cette robustesse est souvent obtenue au prix de mesures de similarité plus complexes.

## 5.2 Complexité de stockage

La table 4 synthétise la complexité de stockage des grandes familles d'approches. Il est à noter que la complexité est ici exprimée par rapport au nombre d'objets de l'image. Il est clair qu'en terme de nombre d'octets ou de bits occupés, le coût de stockage est propre à chaque approche et peut être plus ou moins important en fonction du modèle de représentation.

Type de modèle	Complexité de stockage	Section correspondante
2D-string avec découpage	$O(N^2)$	3.1 ( <i>2D-C</i> , <i>2D-G</i> , <i>2D C<sup>+</sup></i> , <i>2R</i> et <i>RS-string</i> )
2D-string sans découpage	$O(N)$	3.1 (toutes les autres approches <i>string</i> )
Approches matricielles	$O(N^2)$	4.1
Listes de n-uplets	$O(N(N - 1)/2)$	4.2
Fichiers de signature	$O(1)$	4.3
Arbre	$O(1)$	3.2 et 4.4
TSR	$O(N^3)$ ou $O(1)$	4.2

TABLE 4 – Complexité de stockage des approches, en fonction du nombre  $N$  d'objets de l'image.

## 5.3 Similarité entre images

La table 5 compare les représentations symboliques et les similarités entre images pour les différents modèles présentés dans cet article.

En résumé, pour évaluer la similarité entre images, plusieurs méthodes peuvent être utilisées :

- Les approches les plus simples, telles que celle de Chang et Lee (1991), proposent une comparaison booléenne, généralement notée similarité de type-2, pour laquelle la réponse à la comparaison de deux images est "oui/non ces deux images sont similaires" ;
- D'autres approches proposent une comparaison booléenne associée à des similarités de type- $i$  plus souples ou plus évoluées. Par exemple, la similarité de type-0 selon Lee et Hsu (1992) (voir table 2) permet de comparer les catégories spatiales issues des projections symboliques. L'approche *2D C<sup>+</sup>-string* de Huang et Jean (1994) met en jeu 9

## Les relations spatiales dans les images symboliques

Modèle	Section	Représentation symbolique	Similarité entre images traitée
2D-string originale	3.1	grille sur les centres de masse	type-0, type-1, type-2
2D H-string	3.1	centres de masse	Aucune
2D G-string	3.1	REM	type-0, type-1, type-2
2D C-string	3.1	REM	type-0, type-1, type-2
2D C+-string	3.1	REM	de type-00 à type-32 (11 types)
2D Z-string	3.1	REM	de type-00 à type-22 (9 types)
2D-B-string	3.1	REM	type-0, type-1, type-2
2D-Be-string	3.1	REM	distance normalisée
RCOS-string	3.1	REM	type-0, type-1, type-2
$\Theta$ -R-string	3.1	centres de masse	degré de similarité $SIM_G$
Arbre DDT	3.2	REM	distance pondérée
Matrice 9DLT	4.1	centres de masse	type-2
Matrices PN, UID, GPN, BP, UBP	4.1	REM	type-0, type-1, type-2
Zhou et Ang (1997)	4.2	REM	type-2
TSR	4.2	centres de masse	distance normalisée
2D-PIR	4.2	REM	dist. topologique et entre intervalles
9D-SPA	4.2	REM	dist. topologique et directionnelle
Virtual Image	4.2	REM	distance normalisée
Fichiers de signature (sauf approche de Yeh et Chang (2006))	4.3	REM	type-0, type-1, type-2
Yeh et Chang (2006)	4.3	REM et centres de masse	6 types de similarité
Arbre binaire	4.4	centres de masse	type-0, type-1, type-2

TAB. 5 – Résumé des méthodes de comparaison entre deux images.

types de similarité, qui tiennent compte de contraintes sur les tailles des objets et distances entre paires d'objets. Yeh et Chang (2006) définissent quant à eux 6 types de similarité permettant de représenter 289 relations spatiales combinant relations spatiales de Lee et Hsu (1990), les relations directionnelles et les relations topologiques.

- Certaines approches, telles que celles de Petraglia et al. (2001), de Punitha et Guru (2005) ou de Huang et al. (2007), utilisent des mécanismes de comparaison produisant un degré de similarité entre deux relations spatiales (avec notamment les distances topologique et par intervalles 1D) et entre deux images. Ce sont principalement les approches les plus récentes qui proposent ces mesures plus fines, dans l'objectif de permettre la recherche par similarité dans les bases d'images. Ainsi, les systèmes associés peuvent fournir plusieurs réponses, triées par ordre de similarité, ou seuillées.

La table 6 résume les méthodes de comparaison entre deux images utilisées par grande famille de modèles. Il est à noter que les approches ne figurant pas dans cette table sont spécifiques soit parce qu'elles utilisent des structures d'arbres soit parce qu'elles correspondent à des approches géométriques.

Type de modèle	Section	Méthode de comparaison
2D-string (sauf $\Theta$ -R-string)	3.1	LCS 2D
Approches matricielles	4.1	Opérations matricielles simples
Fichiers de signature	4.3	Comparaisons bit à bit
TSR (2001) + 9DLT (95)	4.2 et 4.1	Comparaison de 2 vecteurs après ACP

TAB. 6 – Résumé des méthodes de comparaison entre deux images.

## 5.4 Scénarios applicatifs

Un critère important dans la classification des approches est le type de recherche que la méthode autorise. La plupart des approches rencontrées ont été développées dans le souci de fournir un scénario de recherche par similarité à partir d'un exemple, qui peut être une image (recherche exacte ou partielle) ou bien un ou plusieurs objets d'intérêt (cf. section 2.4). La table 7 dresse la liste des scénarios possibles selon l'approche et indique si l'approche gère la présence répétée du même objet dans l'image.

Modèle	Scénario de recherche par similarité			Multi-occ. d'objets
	image exacte	image partielle	objet	
2D-string et variantes	D	RS	RS	non
Approches matricielles	D	D	P	oui
Zhou et Ang (1997)	D	P	P	non
TSR	D	D	D	oui
2D-PIR	D	D	P	non
9D-SPA	D	D	P	non
Virtual Image	D	D	D	oui
El-Kwae et Kabuka (2000)	D	non	non	non
Lee et al. (1992b)	D	D	D	non
Chang et Yang (1997a) et Chang et al. (2002)	D	RS	RS	non
Yeh et Chang (2006)	D	P	P	non
Arbre binaire	D	RS	RS	non

TAB. 7 – Scénarios applicatifs traités. "D" indique que l'approche a été pensée pour ce type de scénario et le traite directement, "RS" indique que l'approche nécessite du raisonnement spatial pour permettre ce scénario, "P" indique que la nature du modèle de représentation autorise le scénario mais que la mesure de similarité présentée ne le permet pas.

L'analyse de cette table montre que les approches implicites et explicites ont des caractéristiques et propriétés uniques qui les destinent à des scénarios applicatifs dédiés. La compacité des approches implicites induit un espace de stockage moindre, qui les rend facilement manipulables au moment d'une recherche par similarité exacte portant sur l'intégralité de l'image ; les approches à base de chaînes de caractères se prêtent très bien à ce type de scénario. Cependant, elles apparaissent moins efficaces voire inadéquates lors d'une recherche par similarité

plus fine réalisée sur des objets particuliers (sélectionnés a priori ou bien dynamiquement au moment de la recherche) pour les deux raisons suivantes : d'une part, des inférences induisant un coût supplémentaire (ou encore la reconstruction de l'image pour les approches *2D-strings*, phase coûteuse) sont nécessaires pour déduire les relations spatiales existant entre ces objets ; d'autre part, une éventuelle structure d'indexation construite sur l'ensemble des représentations globales des relations par image ne sera pas efficace pour ce type de scénario de recherche. En outre, la base de connaissance induite n'est pas dynamique puisque l'ajout ou le retrait d'un objet dans l'image impose un mécanisme de reconstruction complet de la représentation, aucun des articles cités ne faisant part d'un algorithme en permettant une reconstruction partielle. Les approches explicites, quant à elles, requièrent un plus grand espace de stockage, mais l'accès direct à toutes les relations spatiales d'un type donné, associé à une mesure de similarité et à une structure d'indexation dédiées, impliquent de fait un coût moindre lors de scénarios de recherche fins sur des parties d'images ou sur des objets particuliers ; c'est par exemple le cas de l'approche *TSR*, mais ce n'est pas le cas de l'approche Yeh et Chang (2006) qui est explicite mais est associée dans cette version de l'article à une mesure de similarité globale qui n'autorise que la recherche d'images exacte. En outre, ces approches sont entièrement dynamiques, puisque l'ajout ou le retrait d'un objet n'a pas d'influence sur l'information déjà stockée. On notera également que bon nombre des approches existantes ont l'inconvénient de ne pas gérer la multi-occurrence d'objets dans l'image ; c'est pourtant une configuration très probable dans les images au contenu générique.

Les domaines d'application des approches rencontrées dépendent principalement du type de relation spatiale pris en compte et des scénarios de recherche possibles. Bien que la majeure partie d'entre elles n'ont pas été expérimentées pour un domaine d'application particulier, la classification suivante nous semble pertinente : étant prédisposées à la recherche d'images exacte, les approches implicites sont bien adaptées à des applications spécifiques où la recherche d'une configuration spatiale d'objets particuliers est requise, comme la reconnaissance de visages (après une phase de détection du visage) ou encore l'imagerie médicale. Lorsqu'une description fine des relations spatiales est requise, il faut probablement se tourner vers les approches géométriques qui abordent aussi l'aspect de l'invariance de la description à certaines transformations de l'image. D'un autre côté, les approches explicites se prêtent bien à des scénarios plus dynamiques où la requête peut être définie par l'utilisateur au moment de la recherche et où la recherche doit être plus souple en portant sur des parties d'images ou même simplement des objets d'intérêt. Elles apparaissent donc bien adaptées pour la recherche interactive dans des collections d'images au contenu hétérogène où les scénarios d'usages varient et où les contenus sont amenés à être enrichis souvent (images du web, albums de photos personnelles, etc). Pour ce type d'applications orientées grand public, les approches décrivant les relations spatiales sous forme de relations orthogonales, topologiques et de projections symboliques apportent une précision suffisante en même temps qu'une sémantique appropriée.

## 5.5 Intégration dans les systèmes de recherche d'images

La réalisation des systèmes de recherche d'images qui prennent en compte les relations spatiales entre les objets n'est pas nouvelle. En effet, Chang et al. (1988) ont construit le premier système de recherche d'images par le contenu (IIDS) avec des relations spatiales qui utilisent leur approche de *2D-string* et certaines règles qui permettent de déterminer les relations orthogonales entre les objets. Depuis IIDS, le nombre de systèmes pour la recherche

d'images par le contenu en utilisant des relations spatiales entre les objets a beaucoup augmenté, citons quelques exemples. Le système VisualSEEK de Smith et Chang (1996) utilise les centres de masse des REM pour calculer la distance entre deux objets avec une indexation en arbre quaternaire et une représentation *2D-string* avec des règles pour en déduire les relations orthogonales (adjacence, proximité, chevauchement et de contour). Le système Kmed proposé par Hsu et al. (1996), pour l'imagerie médicale, utilise des  $n$ -uplets de largeur variable avec des caractéristiques spatiales qui permettent de déduire des relations topologiques entre deux objets. Le système ImageMap, proposé par Petrakis et al. (2002), utilisent une projection des graphes ARG avec certaines relations topologiques (adjacence et chevauchement) dans un espace multidimensionnel. Dans la plateforme DISIMA de Oria et al. (2000), l'algèbre de Allen (1983) est utilisé avec les projections sur les axes  $x$  et  $y$  des REM, pour représenter des relations topologiques, directionnelles et de distance entre deux objets. Ce système implémente également l'*arbre 2-D-S* et l'*arbre db*.

## 5.6 Evaluation

Il est très difficile de dresser un palmarès des meilleures approches pour une application donnée, même sur des critères simples tels que le temps de recherche dans une base d'images, en fonction du nombre d'objets par exemple. En effet, dans les articles rencontrés, les approches ne sont pas souvent comparées à celles de la même famille, et il n'existe pas de base d'évaluation communément répandue. Parmi les évaluations les plus rigoureuses, citons l'évaluation de Petrakis (2002) qui compare l'approche *2D-string étendue* aux graphes ARG ; les approches matricielles de Chang (matrices *UID*, *PN*, *GPN* et *BP*) qui ont été évaluées sur la même base vérité-terrain (2000 images contenant 20 objets symboliques) ; ou encore l'approche  $\Theta\mathcal{R}$ -string, évaluée sur la base d'images TESSA de Gudivada (1994) qui contient 75 objets répartis sur 10 images originales, et 15 variantes de ces images après avoir appliqué aux objets des transformations géométriques (translation, rotation, changement d'échelle et leurs combinaisons). De cette constatation apparaît la nécessité de mettre en place une plateforme permettant d'évaluer la pertinence de toutes ces approches sur une base d'images réaliste suffisamment grande. Par exemple, pour une application de recherche dans une base d'images au contenu générique, la base ouverte LabelMe<sup>2</sup> pourrait être pertinente, puisqu'elle contient actuellement plus de 40.000 images dans lesquelles les objets ont été segmentés et annotés.

## Références

- Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Commun. ACM* 26(11), 832–843.
- Arndt, T. et S.-K. Chang (1989). Image sequence compression by iconic indexing. In *IEEE Workshop Visual Languages*, Rome, Italy, pp. 177–182.
- Barrow, H. et R. Popplestone (1971). Relational descriptions in picture processing. *Machine Intelligence VI*, 377–396.
- Bentley, J. (1975). Multidimensional binary search trees used for associative searching. *Commun. ACM (CACM)* 18(9), 509–517.
- Bron, C. et J. Kerbosch (1973). Algorithm 457 : finding all cliques of an undirected graph. *Communications of the ACM* 16(9), 575–577.

<sup>2</sup>LabelMe, the open annotation tool : <http://labelme.csail.mit.edu/>.

## Les relations spatiales dans les images symboliques

- Chan, Y.-K. et C.-C. Chang (1998). An efficient data structure for storing similar binary images. In *Proc. of the 5th Int. Conf. on Foundations of Data Organization (FODO'98)*, Kobe (Japan), pp. 268–275.
- Chan, Y.-K. et C.-C. Chang (2001). Spatial similarity retrieval in video databases. *Journal of Visual Communication and Image Representation* 12(2), 107–122.
- Chang, C. et S. Lee (1991). Retrieval of similar pictures on pictorial databases. *Pattern Recognition* 24(7), 675–680.
- Chang, C. et D. Lin (1996). A spatial data representation : an adaptive 2d-h string. *Pattern Recognition Letters* 17(2), 175–185.
- Chang, C.-C. (1991). Spatial match retrieval of symbolic pictures. *Journal of Information Science and Engineering* 7(3), 405–422.
- Chang, C.-C. et C.-F. Lee (1995). Relative coordinates oriented symbolic string for spatial relationship retrieval. *Pattern Recognition* 28(4), 563–570.
- Chang, C.-C. et C.-F. Lee (1998a). A bin-tree oriented iconic indexing scheme for retrieving symbolic pictures. *Data Knowledge Engineering* 26(2), 121–133.
- Chang, C.-C. et C.-F. Lee (1998b). A spatial match retrieval mechanism for symbolic pictures. *Journal of Systems and Software* 44(1), 73–83.
- Chang, C.-C. et T.-C. Wu (1992). Retrieving the most similar symbolic pictures from pictorial databases. *Information Processing & Management* 28(5), 581–588.
- Chang, C.-C. et T.-C. Wu (1995). An exact match retrieval scheme based upon principal component analysis. *Pattern Recognition Letters* 16(5), 465–470.
- Chang, S. et Y. Li (1988). Representation of multi-resolution symbolic and binary pictures using 2D-H strings. In *Proc. of the IEEE Workshop on Languages for Automata*, Maryland, pp. 190–195.
- Chang, S.-K. et E. Jungert (1986). A spatial knowledge structure for image information systems using symbolic projections. In *Proc. of 1986 ACM Fall Joint Computer Conference*, Los Alamitos, CA, USA, pp. 79–86. IEEE Computer Society Press.
- Chang, S.-K., E. Jungert, et Y. Li (1989). Representation and retrieval of symbolic pictures using generalized 2D strings. In *SPIE Proc. of Visual Communications and Image Processing Conference*, Philadelphia, PA, pp. 1360–1372.
- Chang, S.-K., Q.-Y. Shi, et C.-W. Yan (1987). Iconic indexing by 2-D strings. *IEEE Trans. Pattern Analysis and Machine Intelligence* 9(3), 413–428.
- Chang, S.-K., C.-W. Yan, D.-C. Dimitroff, et T. Arndt (1988). An intelligent image database system. *IEEE Transactions on Software Engineering* 14(5), 681–688.
- Chang, Y.-I. et H.-Y. Ann (1999). A note on adaptive 2d-h strings. *Pattern Recognition Letters* 20(1), 15–20.
- Chang, Y.-I., H.-Y. Ann, et W.-H. Yeh (2000). A unique-ID-based matrix strategy for efficient iconic indexing of symbolic pictures. *Pattern Recognition* 33(8), 1263–1276.
- Chang, Y.-I., H.-Y. Ann, et W.-H. Yeh (2002). An Efficient Signature File Strategy for Similarity Retrieval from Large Iconic Image Databases. *Journal of Visual Languages and Computing* 13(2), 117–147.
- Chang, Y.-I. et B.-Y. Yang (1997a). Efficient access methods for image databases. *Information Processing Letters* 64(2), 95–105.
- Chang, Y.-I. et B.-Y. Yang (1997b). A prime-number-based matrix strategy for efficient iconic indexing of symbolic pictures. *Pattern Recognition* 30(10), 1745–1757.
- Chang, Y.-I., B.-Y. Yang, et W.-H. Yeh (2001). A generalized prime-number-based matrix strategy for efficient iconic indexing of symbolic pictures. *Pattern Recognition Letters* 22(6-7), 657–666.

- Chang, Y.-I., B.-Y. Yang, et W.-H. Yeh (2003). A bit-pattern-based matrix strategy for efficient iconic indexing of symbolic pictures. *Pattern Recognition Letters* 24, 537–545.
- Colliot, O., O. Camara, et I. Bloch (2006). Integration of fuzzy spatial relations in deformable models-application to brain MRI segmentation. *Pattern Recognition* 39(8), 1401–1414.
- Comer, D. (1979). The Ubiquitous B-Tree. *Computing Surveys* 11(2).
- Comer, D. et J. O'Donnell (1982). Geometric Problems with Application to Hashing. *SICOMP* 11(2), 217–226.
- Cook, C. et R. Oldehoeft (1982). A letter oriented minimal perfect hashing function. *ACM SIGPLAN Notices* 17(9), 18–27.
- Egenhofer, M. J. et K. K. Al-Taha (1992). Reasoning about gradual changes of topological relationships. In *Proc. of the International Conference GIS - From Space to Territory : Theories and Methods of Spatio-Temporal Reasoning on Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, London, UK, pp. 196–219. Springer-Verlag.
- Egenhofer, M. J. et R. D. Franzosa (1991). Point set topological relations. *International Journal of Geographical Information Systems* 5, 161–174.
- El-Kwae, E. (2000). Signature-Based Indexing for Retrieval by Spatial Content in Large 2D-String Image Databases. In *12th Int. Symposium on Foundations of Intelligent Systems*, London (UK), pp. 97–108.
- El-Kwae, E. et M. Kabuka (2000). Efficient content-based indexing of large image databases. *ACM Trans. Inf. Syst.* 18(2), 171–210.
- Eshera, M. et K. Fu (1984). A graph distance measure for image analysis. *IEEE Transactions on Systems, Man and Cybernetics* 14(3), 398–408.
- Finkel, R. et J. Bentley (1974). Quad-Trees : A Data Structure for Retrieval on Composite Keys. *Acta Informatica* 4, 1–9.
- Freksa, C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence* 54(1), 199–227.
- Gargantini, I. (1982). an Effective Way to Represent Quadrees. *Communications of ACM* 25(12), 905–910.
- Gouet-Brunet, V. (2006). *Encyclopédie de l'Informatique et des Systèmes d'Information*, Chapter Recherche par contenu visuel dans les grandes collections d'images. J. Akoka, I. Comyn-Wattiau (eds.), Collectif Vuibert.
- Guan, D. J., C.-Y. Chou, et C.-W. Chen (2000). Computational complexity of similarity retrieval in a pictorial database. *Information Processing Letters* 75(3), 113–117.
- Gudivada, V. N. (1994). TESSA-An image testbed for evaluating 2-D spatial similarity algorithms. *ACM SIGIR Forum* 28(2), 17–36.
- Gudivada, V. N. (1998). ThetaR-String : A geometry-based representation for efficient and effective retrieval of images by spatial similarity. *IEEE Transactions on Knowledge and Data Engineering* 10(3), 504–512.
- Guru, D. et P. Nagabhushan (2001). Triangular spatial relationship : a new approach for spatial knowledge representation. *Pattern Recognition Letters* 22(9), 999–1006.
- Guru, D. et P. Punitha (2007). Symbolic image indexing and retrieval by spatial similarity : An approach based on B-tree. *Pattern Recognition*. (in Press).
- Guru, D., P. Punitha, et P. Nagabhushan (2003). Archival and retrieval of symbolic images : An invariant scheme based on triangular spatial relationship. *Pattern Recognition Letters* 24(14), 2397–2408.
- Guttman, A. (1984). R-trees : A Dynamic Index Structure for Spatial Searching. In *Proc. of ACM SIGMOD Int. Conf. on the Management of Data - SIGMOD Record* 14(2), pp. 45–57.

## Les relations spatiales dans les images symboliques

- Hsu, C.-C., W.-W. Chu, et K. Taira (1996). A Knowledge-Based Approach for Retrieving images by content. *IEEE Transaction on Knowledge and Data Engineering* 8(4), 522–532.
- Hsu, F.-J., S.-Y. Lee, et B.-S. Lin (1998a). Similarity Retrieval by 2D C-Trees Matching in Image Databases. *Journal of Visual Com. and Image Repres.* 9(1), 87–100.
- Hsu, F.-J., S.-Y. Lee, et B.-S. Lin (1998b). Video Data Indexing by 2D C-Trees. *Journal of Visual Languages and Comp.* 9(4), 375–397.
- Hsu, F.-J., S.-Y. Lee, et B.-S. Lin (1999). 2D C-Tree Spatial Representation for Iconic Image. *Journal of Visual Languages and Comp.* 10(2), 147–164.
- Huang, G., W. Zhang, et L. Wenyin (2007). A Discriminative Representation for Symbolic Image Similarity Evaluation. In *7th IAPR Int. Workshop on Graphics Recognition*, Curitiba (Brazil).
- Huang, P. et C. Lee (2004). Image Database Design Based on 9D-SPA Representation for Spatial Relations. *IEEE Transaction on Knowledge and Data Engineering* 16(12), 1486–1496.
- Huang, P.-W. et Y.-R. Jean (1994). Using 2D C+-string as spatial knowledge representation for image database systems. *Pattern Recognition* 27(9), 1249–1257.
- Huang, P.-W. et Y. R. Jean (1996). Spatial reasoning and similarity retrieval for image database systems based on RS-strings. *Pattern Recognition* 29(12), 2103–2114.
- Hunt, J. W. et T. G. Szymanski (1977). A fast algorithm for computing longest common subsequences. *Communications of the ACM* 20(5), 350–353.
- Jean, Y.-R. et H.-Y. Lo (2004). An improved similarity measure for image database based on 2D C+-string. In *Int. Computer Symposium*, Taipei, Taiwan, pp. 547–552.
- Jungert, E. (1988). Extended symbolic projection used in a knowledge structure for spatial reasoning. *Pattern Recognition*, 343–351.
- Jungert, E. et S.-K. Chang (1989). An algebra for symbolic image manipulation and transformation. In *Visual Database Systems*, pp. 301–317.
- Kedem, G. (1982). The quad-CIF tree : A data structure for hierarchical on-line algorithms. In *Proc. of the 19th Conf. on Design automation*, Piscataway (USA), pp. 352–357.
- Lee, A. J. T. et H.-P. Chiu (2003). 2D Z-string : a new spatial knowledge representation for image databases. *Pattern Recognition Letters* 24(16), 3015–3026.
- Lee, S.-Y. et F.-J. Hsu (1990). 2D C-string : a new spatial knowledge representation for image database systems. *Pattern Recognition* 23(10), 1077–1088.
- Lee, S.-Y. et F.-J. Hsu (1991). Picture algebra for spatial reasoning of iconic images represented in 2D C-string. *Pattern Recognition Letters* 12(7), 425–435.
- Lee, S.-Y. et F.-J. Hsu (1992). Spatial reasoning and similarity retrieval of images using 2D C-string knowledge representation. *Pattern Recognition* 25(3), 305–318.
- Lee, S.-Y., M.-K. Shan, et W.-P. Yang (1989). Similarity retrieval of iconic image database. *Pattern Recognition* 22(6), 675–682.
- Lee, S.-Y., M.-C. Yang, et J.-W. Chen (1992a). 2D B-String : a spatial knowledge representation for image database systems. In *Proc. of ICSC'92 Second Int. Computer Science Conf.*, pp. 609–615.
- Lee, S.-Y., M.-C. Yang, et J.-W. Chen (1992b). Signature file as a spatial filter for iconic image database. *Journal of Visual Languages and Computing* 3, 373–397.
- Li, X. et X. Qu (1998). Matching spatial relations using db-tree for image retrieval. In *Proc. of the 14th Int. Conf. on Pattern Recognition - Volume 2*, Brisbane (Australia), pp. 1230–1234.
- Lin, T.-W. (1997). Compressed quadtree representations for storing similar images. *Image and Vision Computing* 15(11), 833–843.

- Liu, C. et A. Chen (2002). 3D-List : A Data Structure for Efficient Video Query Processing. *IEEE Transaction on Knowledge and Data Engineering* 14(1), 106–122.
- Manouvrier, M., M. Rukoz, et G. Jomier (2005). *Spatial Databases : Technologies, Techniques and Trend*, Chapter IV - Quadtree-Based Image Representation and Retrieval. Y. Manolopoulos, A. Papadopoulos and M. Vassilakopoulos (Eds), IDEA Group Publ., Information Science Publishing and IRM Press.
- Mughal, M., M. Nawaz, F. Ahmad, S. Shahzad, A. Bhatti, et S. Mohsin (2007). A 3D-Hash Function for Fast Image Indexing and Retrieval. In *4th Int. Conf. on Computer Graphics, Imaging and Visualization (CGIV 2007)*, Bangkok (Thailand), pp. 341–348.
- Nabil, M., A. H. H. Ngu, et J. Shepherd (1996). Picture similarity retrieval using the 2D projection interval representation. *IEEE Transactions on Knowledge and Data Engineering* 8(4), 533–539.
- Nabil, M., J. Shepherd, et A. H. H. Ngu (1995). 2D projection interval relationships : A symbolic representation of spatial relationships. In *Symposium on Large Spatial Databases*, pp. 292–309.
- Niu, Y., M. Özsu, et X. Li (1999). 2-D-S Tree : An Index Structure for Content-Based Retrieval of Images. In *Proc. Multimedia Computing and Networking*, San Jose (California USA), pp. 110–121.
- Oria, V., M. Tamer Özsu, P. Iglinski, S. Lin, et B. Yao (2000). DISIMA : a distributed and interoperable image database system. In *ACM SIGMOD Int. Conf. on Management of data*, Dallas, USA, pp. 600.
- Osborn, W. et K. Barker (2006). Searching through spatial relationships using the 2DR-tree. In *Proc. of the 10th Int. Conf. on Internet Multimedia and Systems Applications (IMSA 2006)*, Honolulu (USA).
- Papadias, D., N. Mamoulis, et V. Delis (1998). Algorithms for Querying by Spatial Structure. In *Proc. of the 24rd Int. Conf. on Very Large Data Bases*, San Francisco (USA), pp. 546–557.
- Papadias, D., Y. Theodoridis, et T. Sellis (1994). The Retrieval of Direction Relations using R-trees. In *DEXA '94 : Proc. of the 5th Int. Conf. on Database and Expert Systems Applications*, London (UK), pp. 173–182.
- Papadopoulos, A. (2005). *Nearest Neighbor Search - A Database Perspective*, Chapter The R-Tree and Variations, pp. 13–21. Springer - Series in Computer Science.
- Petraglia, G., M. Sebillio, M. Tucci, et G. Tortora (1993). Towards normalized iconic indexing. In *Proc. of the 1993 IEEE Workshop on Visual Languages*, Bergen, Norway, pp. 392–394.
- Petraglia, G., M. Sebillio, M. Tucci, et G. Tortora (2001). Virtual images for similarity retrieval in image databases. *IEEE Transactions on Knowledge and Data Engineering* 13(6), 951–967.
- Petrakis, E. (2002). Design and Evaluation of Spatial Similarity Approaches for Image Retrieval. *Image and Vision Computing* 20(1), 59–76.
- Petrakis, E., C. Faloutsos, , et K.-I. D. Lin (2002). ImageMap : An Image Indexing Method Based on Spatial Similarity. *IEEE Transaction on Knowledge and Data Engineering* 14(5), 979–987.
- Petrakis, E. et C. Faloutsos (1997). Similarity Searching in Medical Image Databases. *IEEE Transaction on Knowledge and Data Engineering* 9(3), 435–447.
- Petrakis, E. et S. Orphanoudakis (1996). A Generalized Approach for Image Indexing and Retrieval Based on 2-D Strings. In *Intelligent Image Database Systems*, pp. 197–218. World Scientific Pub. Co.
- Punitha, P. et D. Guru (2005). An invariant scheme for exact match retrieval of symbolic images : Triangular spatial relationship based approach. *Pattern Recognition Letters* 26(7), 893–907.
- Roussopoulos, N. et D. Leifker (1985). Direct spatial search on pictorial databases using packed R-trees. In *Proc. of the 1985 ACM SIGMOD Int. Conf. on Management of data*, Austin, USA, pp. 17–31.
- Sabharwal, C. et S. Bratia (1997). Image databases and near-perfect hash table. *Pattern Recognition* 30(11), 1867–1876.

- Sacks-Davis, R., A. Kent, et K. Ramamohanarao (1987). Multikey access methods based on superimposed coding techniques. *ACM Transactions on Database Systems (TODS)* 12(4), 655–696.
- Samet, H. (1984). The Quadtree and Related Hierarchical Structures. *ACM Computing Surveys* 16(2), 187–260.
- Samet, H. (2006). *Foundations of Multidimensional And Metric Data Structures*. Morgan Kaufmann Pub.
- Shan, M.-K. et S.-Y. Lee (1998). Multidimensional interval filter : a new indexing method for subpicture query of image retrieval. *Pattern Recognition Letters* 19(13), 1241–1255.
- Smith, J. et S. Chang (1996). Visualseek : A fully automated content-based image query system. In *ACM Multimedia*, pp. 87–98.
- Wang, Y.-H. (2003). Image indexing and similarity retrieval based on spatial relationship model. *Information Sciences* 154(1-2), 39–58.
- Wawrzyniak, L., D. Nikitenko, et P. Matsakis (2006). Speaking with Spatial Relations. *Int. J. of Intelligent Systems Technologies and Applications - Special Issue on Intelligent Image and Video Processing and Applications : The Role of Uncertainty* 1(3/4), 280–300.
- Wu, T.-C. et C.-C. Chang (1994). Application of geometric hashing to iconic database retrieval. *Pattern Recognition Letters* 15(9), 871–876.
- Yang, Y.-H., K.-L. Chung, et Y.-H. Tsai (2000). A compact improved quadtree representation with image manipulations. *Image and Vision Computing* 18(3), 223–231.
- Yeh, W.-H. et Y.-I. Chang (2006). An efficient signature extraction method for image similarity retrieval. *Journal of Information Science and Engineering* 22(1), 63–94.
- Yeh, W.-H. et Y.-I. Chang (2008). An efficient iconic indexing strategy for image rotation and reflection in image database s. *Journal of Systems and Software* 81(7), 1184–1195.
- Zhou, X. M. et C. H. Ang (1997). Retrieving similar pictures from a pictorial database by an improved hashing table. *Pattern Recognition Letters* 18(8), 751–758.
- Ziegler, S. (1977). Smaller faster table driven parser. Madison Academic Computing Center, University of Wisconsin, Madison, Wisconsin.

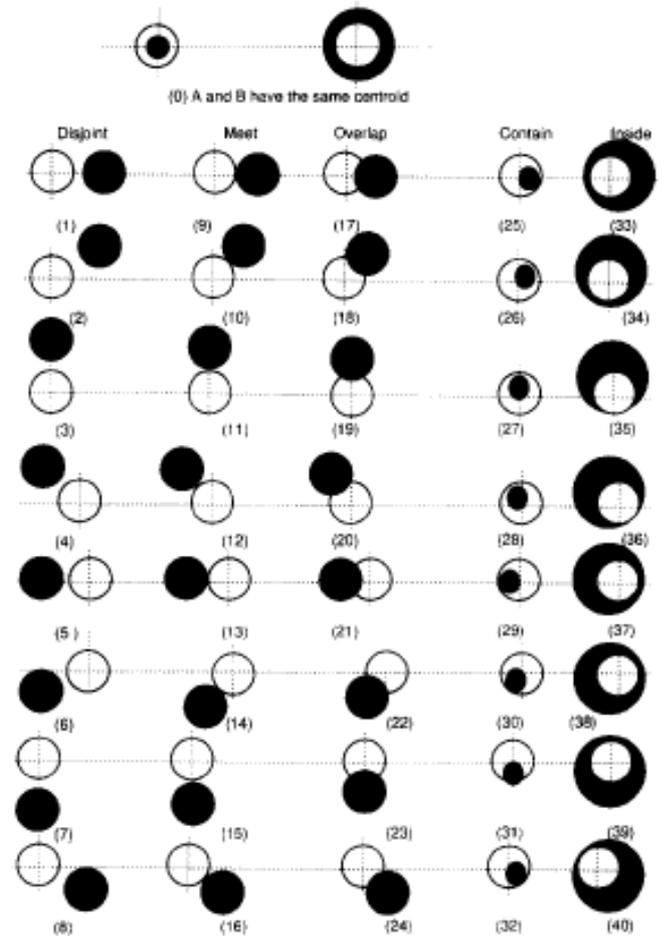
## Annexe

### Summary

The description of the spatial relationships between objects in an image provides a strong semantics to enrich the low level representation techniques of the visual image content. After Chang et al. (1987), a great number of approaches had been proposed, describing spatial relationships in symbolic images, in which the objects of interest have been already extracted and identified. This article draws up a panorama of the existing models. We use a typology that distinguishes two kinds of approaches: the implicit ones, which produce a total representation of the spatial relationships of the image, and the explicit ones, which produce a structure directly describing all the spatial relationships between objects. All the listed approaches are compared according to several criteria, in particular: type of spatial relationships used, volume of storage, complexity of the images comparison algorithm and scenarios of use.

DISJOINT (48)			JOIN (40)			PART_OVLP (50)			CONTAIN (16)	BELONG (16)

(a)



(b)

FIG. 7 – (a) Les 169 relations spatiales obtenues à partir des 7 opérateurs spatiaux 1D de Lee et Hsu (1990). Pour chaque configuration, les deux opérateurs spatiaux (selon les axes  $x$  et  $y$ ) sont indiqués sous le dessin. La figure est extraite de l'article de Lee et Hsu (1992). (b) Les 41 relations combinant relations directionnelles et topologiques définies dans l'article Zhou et Ang (1997).