

Une adaptation du descripteur CSS pour la recherche de parties communes entre objets

B.Tellez

A. Baskurt

LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'information)

Université Claude Bernard, Lyon 1

Bât. Nautibus, 8 boulevard Niels Bohr, 69622 Villeurbanne cedex – France

{btellez, abaskurt}@ligim.univ-lyon1.fr

Résumé

Cet article traite du problème de la recherche de formes 2D dans des bases d'images. Actuellement et malgré l'explosion de l'information mise à disposition sur le Web, il est souvent difficile de retrouver des images autrement que par l'utilisation de mots clés. Pour pallier cet insuffisance, une solution consisterait à disposer d'outils qui puissent identifier les objets présents dans l'image et ainsi permettraient l'indexation et l'interrogation de ces images. Dans ce contexte, la forme de l'objet est souvent caractéristique pour son identification. Nous avons choisi de travailler avec le descripteur CSS proposé par F.Mokhtarian et retenu pour la norme MPEG7. Nous proposons de l'étendre afin qu'il nous permette de détecter non seulement des objets similaires mais surtout qu'il nous permette d'identifier un objet lorsque celui-ci constitue une sous-partie d'un autre objet. Quelques résultats sur la base MPEG7 sont présentés et commentés.

1. Introduction

Alors que l'information disponible sur le Web est chaque jour plus volumineuse et plus inaccessible, il devient urgent de proposer des méthodes et des outils qui permettent d'organiser et d'indexer rapidement les documents multimédias. Dans ce contexte, la norme MPEG7 propose de décrire le contenu de ces documents multimédias par un ensemble de descripteurs normalisés. Parmi ces documents, nous nous intéressons plus spécifiquement à la description des images fixes ainsi qu'aux méthodes permettant de les retrouver dans une collection. Généralement, les critères employés pour décrire une image sont des descripteurs de bas-niveau tels que la couleur et la texture. La forme tient une place à part parmi ces descripteurs dans la mesure où elle apporte une information moins ambiguë sur l'objet car plus proche de sa signification. En effet, l'objectif final reste évidemment l'identification et l'indexation des objets contenus dans une image pour faciliter la recherche d'images sur des critères de plus haut niveau tels que la sémantique.

Dans cet article, nous présentons une méthode de recherche d'objets basée sur la forme. Cette méthode est une extension du descripteur CSS (*Curvature Scale Space*) proposé pour la norme MPEG7. Le CSS est un descripteur multi-échelle proposée par F.Mokhtarian [1]. Le CSS est basé sur la description des parties concaves d'un objet. Il est multi-échelle, invariant à l'échelle, à la translation et à la rotation. Si ce descripteur fournit de très bons résultats pour comparer des objets un à un (et répondre à la question : "*Est-ce que ces deux objets sont similaires ?*"), le processus de normalisation employé ne permet pas d'établir la similarité lorsque l'objet représente une partie d'un autre objet. En effet, lorsqu'il s'agit de traiter des images, une phase de segmentation est souvent nécessaire afin d'extraire des objets graphiques qui seront par la suite interprétés. Dans ce cadre, il est rare que cette phase nous fournisse les objets exacts à comparer aux modèles présents dans notre base. De plus, les objets présents dans une image peuvent être partiellement occultés par d'autres objets.

Après avoir fait un bref rappel sur la mise en œuvre du descripteur CSS, nous présenterons la méthodologie que nous avons mis en place pour faciliter l'utilisation du CSS dans un cadre objet/sous-objet. Elle repose sur une dénormalisation du descripteur (échantillonnage et longueur variant en fonction de la taille de l'objet et de son échelle). Cette dénormalisation permet d'obtenir des descripteurs qui possèdent des caractéristiques communes donc plus faciles à comparer lorsqu'ils traitent d'objets possédant une partie similaire. Finalement, nous présentons quelques résultats et commentaires sur l'utilisation de ce descripteur dans un contexte réel d'application.

2. Bref rappel sur le descripteur CSS

La forme de l'objet est une propriété qui le caractérise de manière intrinsèque. S'il est très facile de décrire un objet simple ou de donner une description approximative d'un objet complexe, il est difficile de caractériser et de comparer la forme d'objets réels (ceux apparaissant dans une image fixe). La nature des attributs d'une forme peut

être globale (surface, élongation, orientation,...) ou bien locale (sommets, coins, courbure...). Ces caractéristiques doivent conduire à une représentation qui présentent des propriétés d'invariance et d'unicité. Cette dernière est nécessaire pour pouvoir introduire une distance qui permette d'établir la similarité ou la non similarité de deux formes. De nombreuses méthodes ont été présentées dans la littérature[2]. Nous ne les détaillerons pas dans cet article mais on retiendra les critères qui permettent d'évaluer la qualité de ces méthodes : efficacité (temps de calcul et espace mémoire nécessaire), portée (types de formes traitées), unicité (une forme = un descripteur unique), stabilité (sensibilité au bruit).

Dans la norme MPEG7, deux méthodes ont été retenues pour décrire les formes : la méthode ART (*Angular Radial Transform*) est une méthode orienté région alors que la méthode CSS (*Curvature Scale Space*) est plutôt une méthode orienté contour. Nous avons choisi de travailler sur l'approche contour et nous rappelons brièvement ici les propriétés du descripteur CSS.

Le descripteur CSS a été proposé initialement par Mokhtarian, Abassi et Kitler [1]. Son principe de base repose sur la description des concavités d'une courbe et son filtrage successif afin d'y suivre l'évolution de ces concavités. Le rôle du filtrage est de lisser la courbe et donc de faire disparaître progressivement ses concavités.

De manière plus précise, la forme est décrite par une courbe fermée représentant son contour. Chaque point est repéré par son abscisse curviligne normalisée sur cette courbe $\Gamma = \{x(u), y(u) | u \in [0, 1]\}$. La fonction Γ subit alors un filtrage par un noyau gaussien [3]. A ce stade, on recherche les points d'inflexion de la courbe afin d'en caractériser les contours. Ces points d'inflexion sont obtenus lorsque la courbure K vaut 0^1 .

$$K = \frac{\dot{x}(u)\ddot{y}(u) - \ddot{x}(u)\dot{y}(u)}{(\dot{x}^2(u) + \dot{y}^2(u))^{3/2}}$$

Ces points servent à construire itérativement une image (appelée **image CSS**) par un empilement successif de couches de points où chaque couche représente les points d'inflexion trouvés à un nombre de filtres donné. Ainsi, l'image qui est obtenue représente, en abscisse, l'abscisse curviligne et en ordonnée, le nombre de filtres. La figure 1 montre l'image CSS obtenue pour décrire le contour d'une carte de l'Europe.

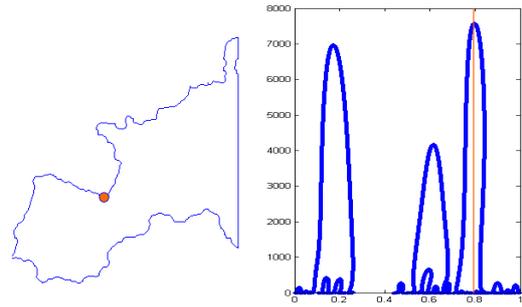


Figure 1. Carte d'Europe et son descripteur

Chaque pic représente une concavité du contour initial. La hauteur du pic est proportionnelle à la taille du détail associé (voir cercle identifié sur la figure). L'abscisse curviligne permet d'assurer l'invariance en translation et en échelle. L'invariance en rotation est elle aussi assurée par un décalage latéral des pics dans l'image CSS. Dès lors, la comparaison de deux formes consistera à mesurer la distance entre les deux ensembles respectifs de pics de chaque image CSS. Pour des raisons d'efficacité et de compacité du descripteur, la norme MPEG7 préconise de ne conserver qu'un certain nombre de pics et de les remplacer par leur amplitude. La comparaison s'effectue de manière récursive en essayant l'ensemble des décalages possibles d'un descripteur par rapport à l'autre et en évaluant à chacune de ces étapes, le coût nécessaire pour comparer chacun des pics.

Le descripteur CSS présente des avantages incontestables d'invariance, de compacité, d'efficacité et de stabilité. En effet, l'échantillonnage et la normalisation employés pour décrire le contour permettent de comparer deux formes de manière sûre. Si les pics ne peuvent pas être mis en correspondance alors on peut en déduire que les objets sont différents. Malheureusement, lorsque l'étude doit être faite sur des objets différents mais où ce qui est recherché est une portion commune, les processus d'échantillonnage et de normalisation ne sont plus adaptés. Ainsi, nous allons voir que choisir un nombre identique de points d'échantillonnage n'est pas une solution adaptée à ce contexte et que la normalisation du contour ne facilite pas la comparaison de portions de courbe dont la longueur est différente.

3. Généralisation du descripteur CSS

Cette partie détaille l'adaptation du CSS pour qu'il permette la détection de parties communes entre deux objets, quelque soit leur forme complète et leur échelle. Les solutions consistent à dénormaliser le descripteur pour que la comparaison des objets puissent se faire en partie et pas seulement sur leur totalité. Cette dénormalisation peut se faire **a priori** notamment en choisissant un échantillonnage adapté au contexte objet/sous-objet ou bien **a posteriori** en essayant de recalculer les descripteurs

¹ Plus précisément, on identifie les passages par zéros grâce aux changements de signe (négatif/positif ou positif/négatif) qui sont distingués et utilisés.

CSS. A posteriori, les temps de calcul sont moins importants qu'a priori mais au dépend d'une qualité moindre. Nous avons choisi de présenter dans cette article une méthode de dénormalisation a priori mais nous montrerons en conclusion que l'analyse du descripteur a posteriori peut malgré tout servir de méthode heuristique pour réduire les temps de calcul induits par la "dénormalisation".

L'objectif de notre démonstration consiste à rechercher une description similaire de la partie commune de deux objets lorsque l'un est un sous-objet de l'autre. Dans ce contexte, la propriété multi-échelle du CSS ne peut plus s'appliquer. La solution pour s'assurer que la partie commune soit représentée de la même manière dans les deux descripteurs est de s'assurer que le nombre de points d'échantillonnage qui la décrivent sont équivalents. Ainsi, si le nombre de points d'échantillonnage est le même, il est évident que le nombre d'itérations nécessaire à son filtrage sera le même. Ainsi, on s'assure que l'amplitude des pics sera la même. Cette nécessité de disposer d'un échantillonnage équivalent peut justifier à lui seul le choix de la dénormalisation a priori. En effet, si l'échantillonnage était vraiment différent car la partie non commune d'un des objets était largement plus grande que la partie commune alors il se pourrait que celle-ci disparaisse rapidement dans les phases de filtrage, le résultat étant qu'il n'y aurait plus suffisamment d'information à comparer. Concernant la longueur de la partie commune (abscisse curviligne) dans le descripteur CSS, la dénormalisation est plus simple et repose sur le rapport à la longueur totale pour chaque objet. La partie suivante décrit comment obtenir l'échantillonnage adéquat.

3.1. Coefficient d'échantillonnage

Pour obtenir une description similaire d'une partie commune à deux objets, il est nécessaire de disposer d'un échantillonnage adéquat. Dans la méthode classique (multi-échelle) où l'on cherche à savoir si des objets sont similaires, un échantillonnage identique pour les objets, quelque soit leur taille, est suffisant. Pour la situation où un sous-objet est recherché, nous montrons ici comment l'échantillonnage doit être choisi en fonction des tailles respectives des deux objets.

Cet échantillonnage tient compte de deux aspects. Le premier concerne le pas d'échantillonnage : celui-ci doit être équivalent pour les deux courbes ainsi la partie commune sera traitée de la même manière lors du calcul du CSS et notamment lors des filtrages successifs. Le second concerne l'échelle des objets². En effet, comme nous ne nous situons pas dans la recherche de deux

formes similaires, la propriété multi-échelle du CSS ne s'applique plus de manière classique.

Le coefficient C est obtenu à partir de l'équation qui décrit l'échantillonnage des différents objets. Dans un premier temps, nous considérerons que les objets (un sous-objet d'un objet de référence) sont à la même échelle. Si l'objet O_q est un sous-objet de l'objet O_r , notre objectif est d'obtenir un échantillonnage identique pour la partie commune de ces objets. Ainsi leur représentation sur l'image CSS sera identique. Pour que l'échantillonnage soit identique, il faut que le nombre de points qui découpent la partie commune soit égal pour les deux objets. Si N_r est le nombre de points retenu pour décrire l'objet de référence O_r alors le nombre de points qui décrit la partie commune dans O_r vaut

$$N_r^c = \frac{L_r^c}{L_r^T} \times N_r$$

où L^c représente la longueur de la partie commune dans O_r et L^T la longueur totale de l'objet. Même chose pour l'objet recherché

$$N_q^c = \frac{L_q^c}{L_q^T} \times N_q$$

Dans cette dernière équation, la valeur N_q est la valeur recherchée à savoir le nombre de points nécessaire pour échantillonner O_q . Elle doit permettre d'assurer que le nombre de points "donné" pour échantillonner la partie commune dans O_q soit le même que celui donné pour la partie commune dans O_r . Si l'échantillonnage doit être le même pour les deux objets alors N_q^c et N_r^c doivent être égaux. On a donc

$$\frac{L_q^c}{L_q^T} \times N_q = \frac{L_r^c}{L_r^T} \times N_r \quad (1)$$

et on en déduit facilement que

$$N_q = \frac{L_q^T}{L_r^T} \times N_r$$

car, si on suppose que les deux objets sont à la même échelle, la longueur de la partie commune dans ces deux objets est égale ($L_q^c = L_r^c$). Il est intéressant de voir que lorsque si un objet est un sous-objet d'un autre objet présent dans la base alors il est possible en influant sur le nombre de points d'échantillonnage du contour d'obtenir un échantillonnage identique de leur partie commune et donc une comparaison plus aisée. Cette propriété est importante car elle est fréquente dans le cas de la segmentation. Il est rare que les objets qui en sont issus correspondent réellement aux objets à identifier. Il faut souvent les combiner entre eux pour reconstruire l'objet de référence. Malheureusement, il est tout aussi rare que l'objet de référence soit décrit dans la base à la même échelle que celui identifié dans l'image. Nous devons donc compléter notre coefficient d'échantillonnage pour qu'il intègre ces notions d'échelle.

² Plus précisément, il s'agit de l'échelle de la partie commune représentée dans les deux objets.

Lorsque les échelles sont différentes, la seule différence avec la démonstration précédente concerne l'assertion indiquant que la longueur des "parties communes" sont équivalentes. Pour avoir cette échelle relative entre les deux objets, nous supposons que l'objet présent dans la base est une représentation d'un objet réel avec une échelle indéterminée e_r et que l'objet requête est un sous-objet de cet objet de référence dont l'échelle par rapport à l'objet réel est e_q . En s'appuyant sur la partie commune L^c -celle apparaissant dans l'objet réel O_r -, on sait qu'elle vaut pour les objets O_r et O_q

$$L_r^c = L_r \times e_r \text{ et } L_q^c = L_q \times e_q$$

Par conséquent,

$$L_q^c = L_r^c \times \frac{e_q}{e_r}$$

et de (1), nous déduisons que le coefficient N_q doit valoir

$$N_q = \frac{L_q^c}{L_r^c} \times \frac{e_r}{e_q} \times N_r$$

La figure 2.a confirme l'équation précédente dans le cas classique à savoir la reconnaissance d'objets identiques mais avec des échelles différentes. Un coefficient d'échantillonnage égal est donc bien adapté. Dans la figure 2.b, le coefficient est adapté à la situation de recherche de l'objet O_q comme requête sur l'objet O_r avec un nombre de points d'échantillonnage différent pour permettre un échantillonnage identique de la partie commune.

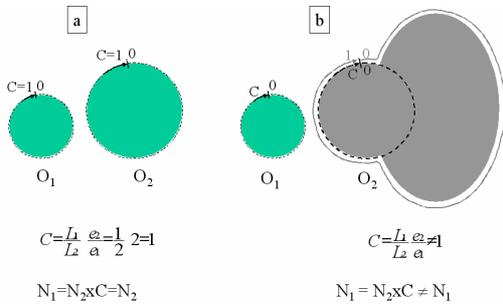


Figure 2. Normalisation des pas d'échantillonnage

Au final, l'utilisation d'un tel coefficient permet d'obtenir des CSS qui présentent des similarités très fortes en terme de nombre de points mais également d'amplitude lorsqu'ils concernent une même partie des objets. Le problème majeur dans l'utilisation d'un tel échantillonnage est qu'il est très difficile de connaître a priori l'échelle respective des objets. Il faudra donc les essayer toutes (ce qui peut diminuer l'efficacité) ou employer des heuristiques qui réduiront l'espace de recherche. Par exemple, il est possible d'utiliser le descripteur a posteriori (réduit) pour déterminer les candidats potentiels et ensuite, utiliser la description complète fourni avec notre coefficient d'échantillonnage pour préciser les résultats.

3.2 Application du CSS généralisé

Dans notre approche, le calcul du descripteur CSS ne présente pas de spécificité propre si ce n'est le nombre de points d'échantillonnage qui varie en fonction des objets à comparer. Nous en rappelons brièvement les étapes importantes et leur adaptation.

La première partie concerne l'échantillonnage et la normalisation du contour (décrite sous forme d'abscisse curviligne). Nous utilisons notre coefficient d'échantillonnage pour adapter la description au contexte objet/sous-objet. Cette étape nous assure que l'amplitude des pics soit similaire.

La seconde partie consistant à effectuer de manière récursive le filtrage de la courbe ainsi que l'extraction des points d'inflexion reste inchangée. Ce processus s'arrête lorsque le contour ne présente plus de concavité.

Finalement, pour compléter la recherche de la partie commune, il est nécessaire de "dénormaliser" l'abscisse des points d'inflexion trouvés sur chacune des images CSS. La solution consiste à ramener les représentations de la partie commune à la même longueur en multipliant la représentation de l'objet requête par C. La figure 3 montre un exemple d'application du coefficient qui généralise les images CSS. Le premier schéma montre la superposition des deux descripteurs CSS : en bleu, le CSS de la carte d'Europe de la figure 1, en rouge le descripteur de la carte d'Espagne à la même échelle que la carte d'Europe. Le second schéma montre un agrandissement (rectangle noir) du premier schéma et qui correspond à la partie commune des deux images.

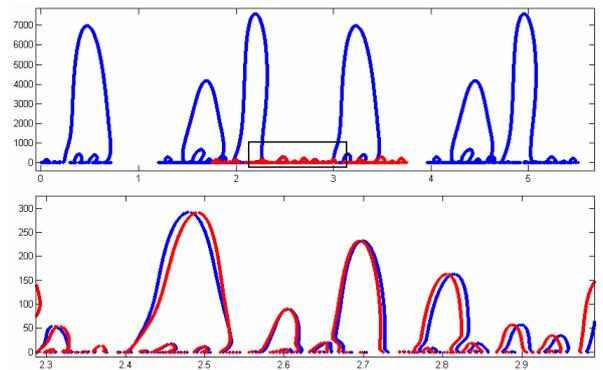


Figure 3. Superposition de 2 descripteurs après échantillonnage et dénormalisation de la longueur

Il est à noter la précision de la superposition des pics en amplitude et en abscisse. De plus, l'échantillonnage assure également une quasi-similarité du nombre de points constituant les pics. Cette similarité permet d'améliorer la qualité de la comparaison ce que ne permettrait pas l'échantillonnage a posteriori évoqué en introduction.

3.3. Similarité entre descripteurs CSS

Dans la procédure de comparaison, la méthode proposée par MPEG7 fournit une réponse quantitative en terme de coût sur la similarité des deux descripteurs et donc des deux formes. Le processus de comparaison préconisé s'appuie non pas sur les pics complets (points d'inflexion à chaque itération) mais sur un ensemble de valeurs représentant l'amplitude maximale de chaque pic sachant qu'un nombre fini de pics sont retenus pour entrer dans le processus de comparaison. Ce nombre de pics étant choisi petit (13 pour la norme MPEG7), il est alors possible d'évaluer, de manière exhaustive en essayant tous les décalages possibles, le coût de l'appariement entre les deux descripteurs. Ce type de comparaison est suffisant pour détecter si deux objets sont semblables et dans quelle mesure. Il est possible ainsi de classer les images d'une base de la plus ressemblante à la moins ressemblante.

Pour identifier la partie commune de deux objets, le processus doit être adapté pour tenir compte des parties non communes des objets. En effet, si nous nous en tenions à un choix arbitraire d'un nombre de pics, le risque serait d'éliminer les pics intéressants (ceux de la partie commune) si l'objet référence était beaucoup plus grand que l'objet requête. Les pics conservés pourraient décrire une partie non commune de l'objet de référence. Pour limiter le risque, nous proposons deux solutions : la première consiste à conserver tous les pics de l'objet référence dont l'amplitude est supérieure au minimum des pics conservés pour l'objet requête. La deuxième ne fait aucune sélection et conserve les pics dans leur intégralité (comme dans la figure 3). Même si cette solution se révèle évidemment beaucoup plus consommatrice en temps de calcul et qu'elle nécessite une mesure entre pics adaptée (distance géodésique préconisée dans [4]), elle nous permet d'exploiter pleinement et de valider l'échantillonnage adapté. De plus, le choix de prendre comme descripteur du pic son amplitude maximale ne semble pas toujours adapté notamment pour décrire la disposition des pics [5] et donc la topologie de la courbe.

Pour déterminer la partie commune entre deux objets, la procédure de comparaison doit donc rechercher dans chacune des images CSS l'ensemble de pics qui vérifie deux propriétés : chacun des pics est à une distance minimale du pic correspondant et la longueur de la courbe délimitée par les pics est maximale. La méthode employée actuellement est calquée sur la recherche de sous-chaînes communes dans des chaînes de caractère. Elle reste relativement efficace dans la mesure où le nombre de pics à comparer est généralement peu élevé (environ 50 pour l'Europe). Les premiers résultats sont satisfaisants dans la mesure où ils permettent de détecter, de manière assez sûre, une partie commune. Cette partie commune n'est pas encore optimale et les raisons seront détaillées dans la partie suivante.

4. Résultats et discussion

Pour terminer, nous présentons et commentons des résultats dans des situations où la détection de la partie commune entre deux objets s'avère nécessaire. La première concerne l'identification d'un objet comme étant un sous-objet d'un autre objet. La deuxième est plus spécifique à des processus de traitement d'images car l'objet à identifier est occulté par un autre objet.

4.1. Détection d'un sous-objet

L'exemple traité est celui de l'Europe et de l'Espagne déjà présenté dans l'article. La partie commune trouvée avec notre méthode est en noir dans la figure 4 et montre que cette partie commune est parfaitement identifiée.

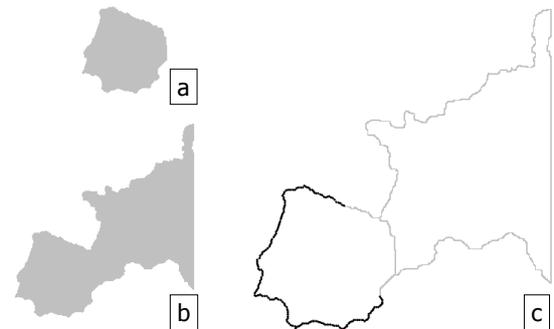


Figure 4. Détection de la sous-chaîne commune (en noir) de deux objets : Espagne (a), Europe (b)

La partie commune n'est pas maximale car l'endroit où le sous-objet s'insère dans l'objet référence peut conduire à de nouvelles concavités dont l'amplitude déterminera jusqu'où l'identification de la partie commune peut être établie. Des solutions basées sur la topologie et la disposition relative des pics (inclusion notamment) permettent de limiter les effets de telles distorsions [5].

4.2. Objet occulté

L'utilisation du CSS pour traiter les parties occultées a été traitée par Mokhtarian [6]. L'approche choisie –que l'on qualifiera de descendante– consiste à découper les différents objets qui apparaissent en s'appuyant sur la description CSS. Ce découpage produit des segments qui sont ensuite appareiller et regrouper pour correspondre à ceux du modèle.

Notre approche emploie la recherche d'une partie commune maximale entre la partie extraite de l'image et les objets présents dans la base. Les poissons utilisés ici pour la démonstration sont issus de la base MPEG7. L'objectif est de repérer le poisson noir caché partiellement par le poisson vert (fig.5.a). La comparaison et la mise en relation permettent de reconnaître et donc de remplacer la partie identifiée (fig.5.b) par le modèle présent dans la base (fig.5.c en rouge). De plus, elle nous permet

de quantifier la meilleure correspondance avec les objets référence sans avoir à rajouter une étape supplémentaire de décomposition des objets présents de la base.



Figure 5. identification et reconstruction d'une image à partir de son modèle

5. Conclusion

Nous avons présenté dans cet article, une méthodologie qui permet d'étendre le descripteur à la reconnaissance de partie d'objets. Notre approche a pour objectif d'introduire un échantillonnage qui s'adapte au couple d'objets à comparer. L'échantillonnage que nous avons proposé est conforme à l'utilisation classique du descripteur (objets similaires mais qui peuvent être d'échelle différente) et il le précise pour les autres situations où seule une partie des objets doit être comparée et mise en relation. (objet et sous-objet à des échelles différentes). Ce travail de reconnaissance des sous-chaînes communes a pour vocation de faciliter l'indexation et l'interprétation des images. Il permet de déterminer assez précisément de quel modèle de la base, un élément de l'image se rapproche. Même si cette dernière ne représente qu'une part du modèle. Ces travaux vont être intégrés dans un outil d'indexation à vocation sémantique. Ils faciliteront la tâche de l'opérateur en lui fournissant les objets les plus probables dans une classe donnée.

Références

- [1] F.Mokhtarian, S.Abbasi, and J.Kittler, "Efficient and robust retrieval by shape content through curvature scale spac", *Proc. International Workshop on Image Databases and MultiMedia Search*, Amsterdam, The Netherlands, pp 35-42, 1996.
- [2] R. C. Velkamp, M. Hagedoorn. "State-of-the-art in shape matching". *Technical Report UU-CS-1999-27*, Utrecht University, the Netherlands, 1999.
- [3] J.Babaud, A.Witkin, M.Baudin, et R.Duda. "Uniqueness of the Gaussian Kernel for scale-space filtering", *IEEE Trans. on PAMI*, 8, pp 26-33,1986.
- [4] S. Mahmoudi, M. Daoudi, "Indexation par la forme utilisant les M-tree". *In Proc. CORESA'2001*, pp 41-44, 2001.
- [5] F. Dachicourt, "Description de Formes pour l'Indexation d'Images", *Mémoire de DEA*, INSA Lyon p43 2001.
- [6] F. Mokhtarian, "Silhouette-Based Occluded Object Recognition through Curvature Scale Space", *Machine Vision and Applications*, vol. 10, number 3, pp. 87-97, 1997.