MajecSTIC 2009
Avignon, France, du 16 au 18 novembre 2009

Δ-TSR : une approche de description des relations spatiales entre objets pour la recherche d'images

Nguyen Vu HOANG[1,2], V. GOUET-BRUNET[2,*], M. MANOUVRIER [1,*], M. RUKOZ [1,*]

1 : Lab. LAMSADE, Université Paris-Dauphine, 75775 Paris Cedex France.
2 : Lab. CEDRIC, CNAM, 75141 Paris Cedex France.
* : Encadrants, M. Rukoz (directrice de thèse).

Contact : nguyenvu.hoang@dauphine.fr

Résumé
Cet article présente une nouvelle approche, Δ-TSR, pour la recherche par similarité dans les bases d'images, où les images sont décrites par les relations spatiales entre leurs objets. Cette approche paramétrable offre différentes descriptions d'image basées sur les co-occurences de triplets d'objets dont les relations géométriques sont codées en utilisant les angles du triangle formé par les objets. Une description semi-locale est également proposée, tenant compte du voisinage des objets, afin d'être robuste aux changements de point de vue. Toutes ces descriptions sont invariantes à la rotation en 2D, à la translation ou au changement d'échelle de l'image. Δ-TSR peut être appliquée aussi bien aux images symboliques (où les objets sont représentés par des étiquettes ou des icônes), qu'aux images représentées par des régions saillantes (par exemple les points d'intérêt représentant les zones de forte variabilité dans l'image). L'approche a été expérimentée avec différents paramètres. Les résultats obtenus montrent que Δ-TSR améliore deux approches apparentées de la littérature (en terme de qualité de recherche et de temps d'exécution) et prouvent son passage à l'échelle.

Abstract
This article presents Δ-TSR, a new image content representation exploiting the spatial relationships existing between its objects of interest. Δ-TSR allows different image descriptions based on co-occurences of object triplets whose geometric relationships are coded with triangle angles. A semi-local representation of the relationships is also proposed, making the description robust to viewpoint changes, if required by the application. The approach is invariant to translation, 2D rotation and scale and it can be applied not only to symbolic images (where objects are represented by labels or icons) but also to contents represented by low-level visual features such as interest points (representing strong variability zones in image). We show that Δ-TSR improves two state-of-the-art approaches, in terms of quality of retrieval as well as of execution time. The experiments also highlight its effectiveness and scalability against large image databases.

Mots-clés : Recherche par similarité, relations spatiales, base d'images

Keywords: Similarity retrieval, spatial relationships, image database

---

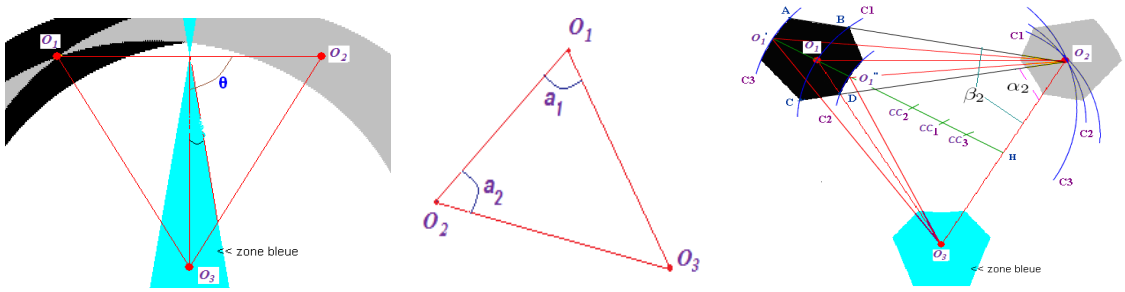
1. Introduction

La recherche d'images par contenu visuel dans les collections d'images (CBIR, pour *Content Based Image Retrieval*) est un domaine très actif depuis une dizaine d'années. Il existe de nombreuses solutions pour décrire le contenu visuel des images. Parmi les approches les plus connues, citons les approches de description globale de couleur, texture et forme dont certaines sont incluses dans le standard MPEG-7 [6]. D'autres types d'approches bas-niveau existent également, et reposent sur une description locale du contenu, représenté par un ensemble de régions ou encore

de points d'intérêt. Ces approches de caractérisation sont de bas-niveau (i.e. sont directement extraites de l'analyse des pixels, sans intégrer de connaissances extérieures de plus haut niveau), mais peuvent être améliorées par la représentation spatiale des objets contenus dans les images. Plusieurs approches ont été proposées pour décrire ces relations. Les relations directionnelles, topologiques, géométriques et orthogonales (voir par exemple de la synthèse [1]) sont les catégories les plus connues. En particulier, les relations géométriques permettent l'invariance à certaines transformations géométriques de l'image, comme la rotation. Dans ces approches, les objets sont représentés par leur centroïde, et des mesures de similarité sont définies afin de comparer les relations existant entre couples ou triplets d'objets.

L'objectif de ce travail est de proposer une représentation efficace et performante de la disposition spatiale des objets dans l'image. Les principes de notre approche, appelée Δ-TSR, sont présentés dans la section 2. La pertinence de Δ-TSR en termes de qualité et de temps de recherche par rapport à deux approches apparentées de la littérature est évaluée dans la section 3. Finalement, la section 4 conclut.

2. Présentation de l'approche Δ-TSR



(a) (b) (c)

FIGURE 1 – (a) Zones de variation dans TSR (b) Relation triangulaire de 3 objets O_1, O_2, O_3 dans Δ-TSR (c) Zones de variation dans Δ-TSR et sa démonstration géométrique.

La section 2.1 présente la modèle description des relations spatiales de notre approche. Cette approche peut être enrichie en fonction de plusieurs stratégies décrites dans la section 2.2. La mesure de similarité est présentée dans la section 2.3. Finalement, la section 2.4 présente notre proposition d'indexation pour accélérer l'accès aux données.

2.1. Description des relations spatiales

Δ-TSR est inspirée de l'approche géométrique TSR (*Triangular Spatial Relationship*) [2, 8] qui caractérise les relations spatiales triangulaires entre objets représentés par leurs centroïdes. Soit O_i un objet représenté par les coordonnées image de son centroïde et par une étiquette numérique L_i ; deux objets de coordonnées différentes peuvent avoir la même étiquette. Il existe différentes manières d'affecter des étiquettes. Dans cet article nous avons utilisé une méthode basée sur les *Bag-of-Features* [12] et de description Harris couleur [7] qui est détaillée dans la section 3.1. Dans TSR, le triangle formé par trois objets O_1, O_2 et O_3 est représenté par un quadruplet (L_1, L_2, L_3, θ) où les objets sont ordonnés selon leur étiquette et où θ est le plus petit angle entre le côté O_1O_2 et la médiane passant par O_3 (voir la figure 1(a)). Dans TSR, une image est donc représentée par l'ensemble des triangles de tous les triplets d'objets possibles. Pour diminuer l'espace de stockage et accélérer la recherche, l'intervalle $[0°..90°]$ auquel appartient θ est divisé en D_θ classes et chaque quadruplet est représenté par une clé unique définie par l'équation 1 :

$$K = D_\theta(L_1 - 1) \times (N_L)^2 + D_\theta(L_2 - 1) \times N_L + D_\theta(L_3 - 1) + (C_\theta - 1) \quad (1)$$

avec $C_\theta \in [1..D_\theta]$ le numéro de la classe à laquelle θ appartient, N_L le nombre d'étiquettes dans la base d'images et $(L_1, L_2, L_3) \in [1..N_L]^3 \subset \mathbb{N}^3$ avec $L_1 \geqslant L_2 \geqslant L_3$. Deux triangles associés à la même clé K sont considérés comme similaires. Cette approche présente deux inconvénients.

D'une part, la valeur de D_θ est incluse dans le codage de la clé 1. Par conséquent, si l'utilisateur veut changer cette valeur, toutes les clés de toutes les images doivent être recalculées. D'autre part, à cause de la définition de θ, une petite variation de θ peut produire une grande zone de variation pour les triangles similaires, comme le montre la figure 1(a). Cette figure représente la zone de variation d'un triangle équilatéral $O_1O_2O_3$ tolérant une variation de θ de $10°$. Les zones noire, grise et bleue représentent les variations possibles de O_1, O_2 et O_3 respectivement : par exemple pour O_2 et O_3 fixés et O_1 variant dans la zone noire, tous les triangles associés ont la même clé K.

Reprenant l'idée de TSR [8], nous proposons une nouvelle modèle description d'image appelée Δ-TSR_{3D}, applicable aussi bien aux objets symboliques représentés par un point (par exemple, leur centroïde) qu'aux descripteurs visuels de bas-niveau tels que les points d'intérêt [4, 7]. Chaque image I de la base est représentée par un ensemble $\Delta\text{-TSR}_{3D}(I)$ contenant les signatures de toutes les relations triangulaires entre ses objets (classés selon leur étiquette) tel que :

$$\Delta\text{-TSR}_{3D}(I) = \{S^a(O_i, O_j, O_k)/O_i, O_j, O_k \in I; i, j, k \in [1..N_I]; L_i \geq L_j \geq L_k\} \quad (2)$$

avec N_I le nombre d'objets dans l'image I, les autres variables étant les mêmes que celles de l'équation 1. La relation triangulaire entre trois objets O_i, O_j et O_k est représentée par une signature à trois dimensions :

$$S^a(O_i, O_j, O_k) = (K_1, K_2, K_3)\,;\, \forall i, j, k \in [1..N_I] \quad (3)$$

$$\text{avec} \begin{cases} K_1 = (L_i - 1)(N_L)^2 + (L_j - 1)N_L + (L_k - 1) \\ K_2 = a_i;\ K_2 \in [0°..180°] \\ K_3 = a_j;\ K_3 \in [0°..180°] \end{cases} \quad (4)$$

$$\text{et} \begin{cases} (a_i, a_j \in \mathbb{N}) \wedge (a_i, a_j \in [0°..180°]) \\ L_i = L_j \Longrightarrow a_i \geqslant a_j \\ L_j = L_k \Longrightarrow a_i \geqslant 180° - a_i - a_j \end{cases} \quad (5)$$

où a_i et a_j [1] sont les angles des sommets O_i, O_j respectivement (voir la figure 1(b)). Alors que dans l'approche TSR [8], une même clé K est associée à différents triangles en fixant un intervalle de tolérance (codé par D_θ et C_θ dans l'équation 1), dans notre approche, chaque signature est associée à un seul triangle et son symétrique. De plus, au lieu d'avoir une clé qui dépend de l'intervalle de tolérance, la signature S^a est indépendante de cet intervalle, que nous appelons ici δ_a. Ainsi, modifier δ_a n'a aucun impact sur la description de l'image et sur l'indexation de la base. En retour, δ_a est utilisé pour définir la similarité entre les triangles ; nous considérons que les triangles similaires au triangle T_Q sont les triangles T_I dont les angles vérifient les contraintes de tolérance définies par :

$$\begin{cases} \alpha_1 = max(a_i(T_Q) - \delta_a, 0°) \leqslant a_i(T_I) \leqslant \beta_1 = min(a_i(T_Q) + \delta_a, 180°) \\ \alpha_2 = max(a_j(T_Q) - \delta_a, 0°) \leqslant a_j(T_I) \leqslant \beta_2 = min(a_j(T_Q) + \delta_a, 180°) \\ a_k(T) = 180° - a_j(T) - a_i(T)\,;\ \ \forall T = T_Q, T_I \\ \alpha_3 = max(a_k(T_Q) - \delta_a, 0°) \leqslant a_k(T_I) \leqslant \beta_3 = min(a_k(T_Q) + \delta_a, 180°) \end{cases} \quad (6)$$

Ces contraintes définissent les zones de variation acceptables pour chaque sommet du triangle, comme l'illustre la figure 1(c).

Si la géométrie des triangles (leurs angles) n'est pas prise en compte dans la signature, on obtient la signature réduite appelée $\Delta\text{-TSR}_{1D}$, telle que :

$$\Delta\text{-TSR}_{1D}(I) = \{S^\ell(O_i, O_j, O_k)/O_i, O_j, O_k \in I; i, j, k \in [1..N_I]; L_i \geq L_j \geq L_k\} \quad (7)$$

avec : $S^\ell(O_i, O_j, O_k) = (K_1)\ ;\ \forall i, j, k \in [1..N_I]$. Cette signature caractérise les co-occurrences de triplets d'objets. Elle est similaire à celles proposées dans [10,11] qui caractérisent quant à elles les co-occurrences de doublons d'objets.

2.2. Stratégies de sélection des triplets d'objets

Dans la mesure où les signatures $\Delta\text{-TSR}_{1D}$ et $\Delta\text{-TSR}_{3D}$ caractérisent tous les triplets d'objets de l'image, il est probable que certains de ces triplets impliquent des objets situés loin les uns des

1. Pour des gains de temps et d'espace, nous choisissons a_i et a_j dans $\mathbb{N}$ au lieu de $\mathbb{R}$. Notons que nos expériementations avec $\mathbb{R}$ n'ont donné aucune véritable amélioration par rapport à celles avec $\mathbb{N}$.

autres spatialement. Une telle représentation semble adéquate pour une description globale du contenu de l'image, mais pas pour la description de parties d'images ou d'objets d'intérêt, qui est utile pour la recherche ou reconnaissance d'objets. Ici, une description *semi-locale*, qui privilégie les plus petits triangles, est largement suffisante et plus robuste à ce type de scénario : elle permet notamment d'améliorer la robustesse aux changements de point de vue des objets d'intérêt représentés avec plusieurs triangles. Comme description semi-locale, nous considérons les triangles appartenant à un voisinage semi-local de rayon r. Par défaut, r est fixé pour tous les objets O_i, mais il peut être adapté selon l'échelle de l'objet si elle est disponible, comme c'est le cas avec les points d'intérêt SIFT [4] extraits à des échelles spécifiques. Dans les expérimentations de la section 3, nous avons fait varier r et sélectionné la valeur de r fournissant les résultats optimaux. Nous présentons ici plusieurs stratégies de sélection des triangles dans un voisinage semi-local, toutes appliquées à chaque objet O_i d'une image I :

Stratégie SL : Élagage semi-local des triangles
1. Chercher O_j dans le voisinage de O_i tel que $d_{L_2}(O_i, O_j) \leqslant r$ où d_{L_2} est la distance Euclidienne ;
2. Construire toutes les relations triangulaires de la liste des objets $\{O_j\}$ trouvés.

Une conséquence intéressante de cette stratégie est que le nombre de triangles ($C^3_{N_I}$ par défaut, où N_I est le nombre d'objets dans I) est nettement réduit à $N_I \times C^3_{\bar{n}}$ triplets en moyenne, où $\bar{n}$ est le nombre moyen d'objets dans chaque voisinage. Avec cette stratégie, toutes les relations triangulaires possibles sont construites dans un voisinage semi-local.
Dans chaque voisinage, nous pouvons encore réduire la complexité de la description par l'ajout d'autres stratégies d'élagage, comme les deux suivantes :

Stratégie SL_{sem} : Élagage basé sur la sémantique Cette stratégie est également semi-locale et produit une triangulation qui est déduite des étiquettes et non de la géométrie des objets, d'où la nomination de sémantique, les étiquettes bien que numériques représentant une certaine information sémantique sur les objets de l'image. Cette triangulation est définie comme suit :
1. Chercher tous les objets O_j dans le voisinage de O_i, comme avec la stratégie SL mais supprimer O_i de $\{O_j\}$;
2. Ordonner la liste $\{O_j\}$ par ordre décroissant de leurs étiquettes L_j ;
3. A partir de cette liste de taille $|\{O_j\}|$, construire toutes les relations triangulaires (O_i,O^l_j,O^m_j) où O^l_j est le l-ème objet de $\{O_j\}$, $l = 1, .., |\{O_j\}|$, $m = 1$ si $l = |\{O_j\}|$ et $m = l + 1$ sinon.

Avec cette stratégie la taille de signature dans l'image est réduite à une moyenne de $N_I \times (\bar{n}-1)$ triangles. Dans chaque voisinage, on obtient un ensemble minimal de triangles, non nécessairement disjoints, qui relient chaque objet O_j à trois autres objets au moins.

Stratégie SL_{geo} : Élagage basé sur la géométrie
Différemment de la stratégie SL_{sem}, cette stratégie fournit une triangulation des objets dans un voisinage semi-local, directement déduite de leur géométrie, comme suit :
1. Chercher tous les objets O_j dans le voisinage de O_i, comme avec la stratégie SL ;
2. Sélectionner les relations triangulaires qui vérifient la triangulation de Delaunay.

Une triangulation de Delaunay est choisie pour cette stratégie, car elle maximise l'angle minimal de triangles construits, de manière à préserver une certaine localité, réduisant ainsi leur étirement, et améliorant donc la robustesse de la description aux changements de point de vue. Cette triangulation assure une taille de signature de $N_I \times [2(\bar{n}-1)-\bar{e}]$ en moyenne, où $\bar{e}$ est le nombre moyen d'objets dans l'enveloppe convexe. Dans chaque voisinage, on obtient un ensemble minimal de triangles disjoints, qui forment une partition de l'enveloppe convexe associée aux objets $\{O_j\}$.

2.3. Mesure de similarité

Avec Δ-TSR, la similarité entre deux images peut être vue comme le ratio de leurs triangles similaires. Soient T_Q un triangle composé des sommets O_1,O_2,O_3 de l'image requête Q et T_I un triangle composé des sommets O'_1,O'_2,O'_3 d'une image I, tels que les objets O_i et O'_i ont la même étiquette L_i ($i \in [1..N_I]$). Chaque image de la base étant représentée par une collection de signatures $S^a(T_I)$, le problème de la recherche d'images similaires est un problème d'adéquation entre les signatures $S^a(T_Q)$ et $S^a(T_I)$ telles que $K_1(T_Q) = K_1(T_I)$ en tenant compte de l'intervalle de tolérance δ_a. Nous proposons une mesure de similarité entre les images, notée SIM, basée sur la mesure de similarité entre les signatures de triangle, notée sim^a. Ces mesures varient dans l'in-

tervalle $[0, 1]$ et augmentent avec la similarité.

Mesure de similarité entre triangles : la similarité entre $S^a(T_Q)$ et $S^a(T_I)$ est définie par :

$$sim^a(S^a(T_Q), S^a(T_I)) = \begin{cases} 1 \text{ si } \delta_a = 0 \\ \frac{1}{2}(\sum_{i=2}^{3}(1 - \frac{|K_i(T_Q) - K_i(T_I)|}{\delta_a})) \text{ si } \delta_a \neq 0 \\ 0 \text{ si } S^a(T_I) \text{ ne vérifie pas les contraintes de tolérance de l'éq. 6} \end{cases} \quad (8)$$

Mesure de similarité entre images : Soient Δ-TSR(I) et Δ-TSR(Q) les signatures associées aux images Q et I respectivement, et SP(Q,I), l'ensemble des couples $(S^a(T_Q), S^a(T_I))$ des triangles de I et de Q les plus similaires, tel que :

$$SP(Q, I) = \left\{ \begin{array}{l} (S^a(T_Q), S^a(T_I)) \ / \\ S^a(T_Q) \in \Delta\text{-TSR(Q)} \ \wedge S^a(T_I) \in \Delta\text{-TSR(I)} \ \wedge sim(S^a(T_Q), S^a(T_I)) \neq 0 \wedge \\ sim(S^a(T_Q), S^a(T_I)) = \max_{\forall S^a(T_I') \in \Delta\text{-TSR(I)}} (sim(S^a(T_Q), S^a(T_I')) \wedge \\ sim(S^a(T_Q), S^a(T_I)) = \max_{\forall S^a(T_Q') \in \Delta\text{-TSR(Q)}} (sim(S^a(T_Q'), S^a(T_I)) \end{array} \right\} \quad (9)$$

La similarité entre les images Q et I est définie comme suit :

$$SIM(Q, I) = \frac{\sum_{k=1}^{card(SP(Q,I))} sim(SP_k(Q, I))}{card(\Delta\text{-TSR(Q)})} \quad (10)$$

où $SP_k(Q, I)$ est le $k^{ème}$ élément de $SP(Q, I)$. La formule représente la proportion de triangles de Q qui ont trouvé un appariement dans I ; cette proportion est pondérée par la mesure sim entre ces triangles. Les images résultat sont ordonnées par ordre croissant de SIM.

2.4. Méthode d'accès associée

Comme indiqué dans la section 2.3, la recherche par similarité des images nécessite la comparaison des signatures de l'image requête avec les signatures de chaque image stockée dans la base pour le calcul de leur mesure de similarité. Comme dans TSR [8], nous proposons d'utiliser une structure d'index pour accélérer la recherche. Pour trouver les signatures similaires à une signature $S^a(T_Q) = (K_1(T_Q), K_2(T_Q), K_3(T_Q))$, le processus de recherche est le suivant :

1. Sélectionner toutes les signatures $S^a(T_I)$ telles que $K_1(T_I) = K_1(T_Q)$;
2. Dans cet ensemble, choisir celles qui vérifient l'intervalle de tolérance δ_a ;
3. Calculer $sim^a(S^a(T_Q), S^a(T_I))$.

Si l'ordonnancement des signatures multidimensionnelles est tel que $S^a(T_I) > S^a(T_Q)$ si et seulement si $\exists i \ / \ K_i(T_I) > K_i(T_Q) \wedge \forall j < i \ K_j(T_I) = K_j(T_Q)$, alors le processus de recherche devient la recherche de l'ensemble des signatures S_I^a dans l'intervalle $[BI_i, BI_f]$ où $BI_i = (K_1, K_2 - \delta_a, K_3 - \delta_a)$ et $BI_f = (K_1, K_2 + \delta_a, K_3 + \delta_a)$. Par conséquent, il est optimal d'utiliser un arbre B à clés composites [9] pour indexer les composantes de la signature S^a. De cette façon, la complexité de recherche devient $O(N_{MT} log_b N_T)$ où N_{MT} est le nombre moyen de triangles dans l'image, N_T le nombre total de triangles dans la base et b est le degré de l'arbre B.

Comme S^a est une signature multidimensionnelle, nous avons aussi expérimenté une structure d'index multidimensionnelle classique, l'arbre R. Cependant, cette structure n'apporte aucune amélioration par rapport à l'arbre B, comme le montrent les résultats de la section 3.3.

3. Evaluation de l'approche Δ-TSR

Cette section est consacrée à l'évaluation de Δ-TSR pour la recherche d'objet ou de sous-images dans une collection d'images, en comparant ses performances à deux approches apparentées de la littérature : TSR [8] et BoF [12].

3.1. Cadre d'évaluation

Matériel. Toutes les approches ont été développées en Java. Les tests ont été effectués sur un PC ayant un processeur Intel Core 2 2.17GHz et 4 Go de RAM, sous Windows XP.

Bases d'images. Nous avons utilisé deux bases d'images : l'une de 6000 images, notée DB_{6000}, et l'autre de 600 images, notée DB_{600}, qui est un sous-ensemble de DB_{6000}. Chaque image contient

FIGURE 2 – Exemples de deux objets avec différents arrière-plans et poses 3D/rotations 2D.

un objet de la classique base d'images *COIL-100* [2], synthétiquement inséré sur une photo faisant office d'arrière-plan (images de 352 × 288 pixels au contenu hétérogène et téléchargées à partir d'Internet). Dans DB_{6000}, nous considérons 6000 arrière-plans différents et 100 objets avec 6 poses 3D différentes combinées à une rotation 2D sur 10 arrière-plans. Il y a donc 60 images par objet et 10 images par rotation 2D/pose 3D, comme illustré sur la figure 2. DB_{600} est obtenue en prenant au hasard 20 objets sous 3 poses 3D/rotations 2D différentes et avec 10 arrière-plans, ce qui conduit à 30 images par objet.

Description du contenu visuel par sacs de mots visuels. Parmi les approches qui décrivent le contenu d'une image avec des caractéristiques locales de bas-niveau, l'approche *Bag-of-Features* (BoF), initialement proposée dans [12], est très répandue pour la reconnaissance d'objets. Elle repose sur une description dite *par sacs de mots visuels*. Inspirée du principe de l'indexation textuelle, chaque point d'intérêt extrait automatiquement de l'image est associé à un étiquette, communément appelé "mot visuel", défini par analyse du signal bidimensionnel présent dans un voisinage local autour du point, et classé par utilisation d'une technique de regroupement des données de type k-means [5]. Le contenu visuel de chaque image est représenté par un vecteur de taille fixe comptabilisant la fréquence d'apparition de chaque mot visuel dans l'image. Cette représentation a l'avantage de décrire localement le contenu de l'image, de manière compacte. Nous avons choisi d'évaluer Δ-TSR sur ce type d'approche, car cette dernière a le défaut de ne pas intégrer une représentation spatiale des mots visuels. Les points sont extraits avec le détecteur de Harris couleur et leur voisinage local est décrit par un ensemble d'invariants différentiels en couleur [7]. Dans DB_{6000}, le nombre de points extraits varie entre 300 et 600 points par image. Le vocabulaire visuel est obtenu à l'aide de l'algorithme k-means ; il est paramétré pour produire une taille (N_L) de vocabulaire de 1000 mots. La figure 2 montre des exemples de points d'intérêt extraits ; les couleurs des points représentent les mots visuels et donc les étiquettes qui les caractérisent.

Implémentation et techniques de comparaison. Nous avons implémenté de manière optimale deux approches de la littérature : TSR [8] implémentée en semi-local (voir la section 2.2) et BoF [12], pour les comparer à Δ-TSR. Dans TSR, les clés K sont indexées par un arbre B. L'indexation utilisée pour BoF est un fichier inversé [3], optimal pour ce type de description.

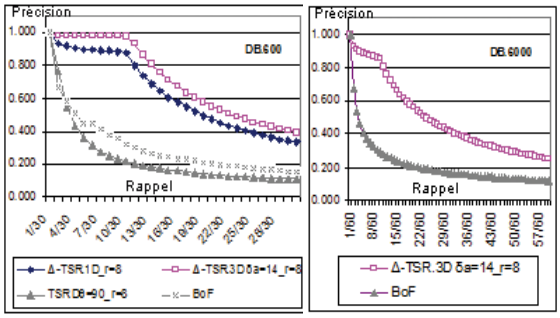
Critères d'évaluation. Toutes les approches ont été évaluées tant en termes (1) de qualité des résultats, en calculant les courbes précision/rappel (P/R), que (2) de temps d'exécution d'une requête, en mesurant le temps CPU et IO. Ces mesures correspondent à la moyenne des résultats obtenus en prenant chaque image de la base comme image requête.

3.2. Evaluation de qualité

Cette section compare $\Delta\text{-TSR}_{1D}$, $\Delta\text{-TSR}_{3D}$ à TSR et à BoF en terme de qualité de résultats. Parce que l'évaluation est effectuée pour la recherche de sous-image/objet sur les bases d'images contenant des objets 3D (voir la section 3.1), nous évaluons TSR, $\Delta\text{-TSR}_{1D}$, $\Delta\text{-TSR}_{3D}$ avec leur représentation semi-locale (stratégie SL présentée dans la section 2.2). Des expérimentations préliminaires ont été réalisées pour fixer les paramètres associés à ces approches : rayon de voisinage r, seuils de tolérance δ_a pour $\Delta\text{-TSR}_{3D}$ et D_θ pour TSR.
Quelle que soit l'approche, les meilleurs résultats sont obtenus avec $r = 8$ et avec $\delta_a = 14°$ pour $\Delta\text{-TSR}_{3D}$ et $D_\theta = 1$ pour TSR. La précision baisse lorsque le rayon augmente démontrant ainsi la pertinence de la représentation semi-locale. Des triangles de plus grande taille conduisent en effet à une représentation moins efficace car (i) moins résistante aux changements de point de vue

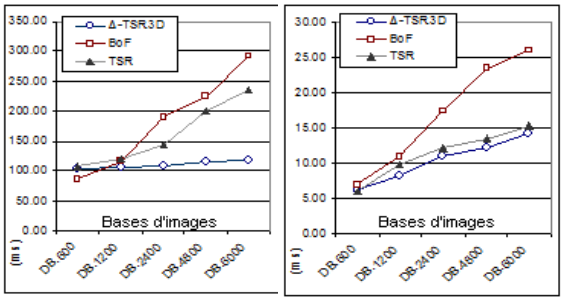
2. http ://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php.

et (ii) impliquant plus de triplets de points d'intérêt à la fois de l'objet et du fond et qui sont par conséquent moins reproductibles à travers les images. Une conséquence intéressante de ce résultat est que, avec un petit rayon, la complexité de Δ-TSR$_{3D}$ ainsi que celle de TSR est réduite : pour $r = 8$ sur DB$_{600}$, nous obtenons 301091 triangles, alors qu'il en a 11300207 pour $r = 32$. Avec TSR, la valeur est optimale avec $D_\theta = 1$, ce qui signifie qu'il n'y a qu'une seule classe d'angles pour θ. Ce résultat indique que la meilleure performance est obtenue sans aucune contrainte d'angle, indiquant que cette approche n'est pas adaptée à la recherche d'objet en utilisant des descripteurs locaux. La figure 3(a) montre que la qualité des résultats obtenus avec Δ-TSR$_{1D}$ est supérieure à



(a) (b)

FIGURE 3 – Comparaison des meilleures configurations des approches sur (a) DB$_{600}$ et (b) DB$_{6000}$.



(a) Temps CPU (b) Temps IO

FIGURE 4 – Temps d'exécution en faisant varier la taille de la base d'images.

celle de BoF (cela avait été déjà remarqué dans [10, 11] pour des couples de mots visuels). La caractérisation de la géométrie des triplets d'objets de Δ-TSR$_{3D}$ apporte une information pertinente qui augmente encore plus la précision aussi bien sur DB$_{600}$ que sur DB$_{6000}$ (voir figure 3(b)).

Finalement, nous pouvons remarquer que Δ-TSR est sensible aux changements de point de vue 3D : à partir du rappel 11/30 où le point de vue est modifié, la précision diminue. En effet, la pose 3D conduit à la disparition / apparition de points d'intérêt, ce qui implique la modification de l'ensemble des triplets de points à travers les images.

3.3. Evaluation du temps d'exécution

Afin d'améliorer le temps d'exécution pour chaque approche, nous avons développé plusieurs structures d'index. La meilleure technique d'indexation pour BoF est le fichier inversé. Dans l'approche TSR, comme les clés sont mono-dimensionnelles, les auteurs utilisent un arbre B. Dans notre approche, bien que les signatures soient multidimensionnelles, elles peuvent être ordonnées et être indexées par un arbre B à clés de recherche composites. Parce que les signatures sont multidimensionnelles, nous avons également expérimenté une indexation par un arbre R [3].

Le tableau 1 montre le temps CPU et IO des différentes approches. Sur la petite base DB$_{600}$, BoF avec le fichier inversé est l'approche la plus rapide. Sur la même base d'images, les temps obtenus avec TSR et Δ-TSR sont équivalents lorsque notre approche est indexée par un arbre B. Avec l'arbre R, nous avons obtenu de plus mauvais temps. En effet, comme expliqué dans la section 2.3, le processus de recherche de Δ-TSR est basé sur un ordonnancement. Pour ce type de recherche particulier, l'arbre R n'est pas la meilleure structure d'index. Alors que, sur la plus grande base DB$_{6000}$, le temps CPU est multiplié par deux ou trois pour TSR et BoF, il est presque constant avec Δ-TSR$_{3D}$ avec un arbre B. Le temps IO est quant à lui multiplié par deux pour TSR et Δ-TSR$_{3D}$, et il augmente plus pour BoF. D'autre part, le temps d'exécution de notre approche est amélioré par les stratégies d'élagage (voir le tableau 2). La stratégie SL$_{sem}$ permet un bon rapport temps/qualité : elle devient même meilleure que BoF à la fois en termes de temps d'exécution et de qualité des réponses.

3. Dont le code source a été repris de `http ://www.rtreeportal.org`.

	DB_{600}		DB_{6000}	
Approche / Index	CPU	IO	CPU	IO
BoF / Fich. inver.	85,532	6,916	292,653	26,152
TSR/ B-tree	107,218	5,982	235,402	15,319
Δ-TSR_{3D} / B-tree	103,360	6,182	118,618	14,193
Δ-TSR_{3D} / R-tree	281,721	18,383	328,655	24,343

TABLE 1 – Temps CPU/IO (en ms) des différentes approches sur les 2 bases DB_{600} et DB_{6000}.

Stratégie	CPU	IO	P.M.
SL	103,360	6,182	0,716
SL_{sem}	70,153	4,643	0,675
SL_{geo}	76,165	4,762	0,548
BoF	85,532	6,916	0,309

TABLE 2 – Temps CPU/IO moyen (en ms) pour les stratégies d'élagage de Δ-TSR_{3D} sur DB_{600} et leur précision moyenne (P.M.).

Afin d'évaluer le passage à l'échelle de Δ-TSR_{3D}, nous avons fait varier la taille de la base, de 600 jusqu'à 6000 images. Cette évaluation a été réalisée avec 1000 étiquettes. Le temps d'exécution est influencé par l'algorithme de recherche (dépendant de la structure d'index utilisée), par le temps de calcul de la mesure de similarité et par le temps d'ordonnancement des images résultat. La figure 4 présente le temps d'exécution (CPU et IO) obtenu sur les base d'images de taille différente pour BoF, TSR et Δ-TSR_{3D}. Ces expérimentations montrent que la taille de la base d'images a une faible influence sur Δ-TSR_{3D}. Pour information, le temps de préparation (extraction des points d'intérêt, classification et indexation) sur DB_{6000} est d'environ 24h.

4. Conclusions et perspectives

Dans ce document, nous avons proposé l'approche Δ-TSR basée sur les relations triangulaires entre objets permettant une description des images invariante à la rotation 2D, à la translation et aux changements d'échelle. Cette approche est paramétrable : Δ-TSR_{1D} décrit des co-occurences de triplets d'objets et Δ-TSR_{3D} intègre la géométrie entre ces triplets. Nos expériences montrent que Δ-TSR améliore non seulement la qualité de recherche par similarité mais aussi le temps d'exécution par rapport à deux approches apparentées de la littérature, TSR [8] et BoF [12].
Δ-TSR peut être encore améliorée en introduisant par exemple d'autres relations spatiales ou l'orientation des objets. L'étiquetage des objets peut également avoir un impact, c'est pourquoi nous étudions actuellement d'autres approches pour étiqueter les objets. D'autre part, dans cet article, les expérimentations ont été réalisées sur les caractéristiques visuelles de bas-niveau. Une des perspectives de ce travail consistera à examiner d'autres types objets symboliques de plus haut-niveau.

Bibliographie

1. V. Gouet-Brunet, M. Manouvrier, et M. Rukoz. Synthèse sur les modèles de représentation des relations spatiales dans les images symboliques. *Revue des Nouvelles Technologies de l'Information*, (RNTI-E-14) :19–54, novembre 2008.
2. D. Guru et P. Nagabhushan. Triangular spatial relationship : a new approach for spatial knowledge representation. *Pattern Recogn. Lett.*, 22(9) :999–1006, 2001.
3. T. Bell. Managing Gigabytes I. H. Witten, A. Moffat. Compressing and indexing documents and images. In *Morgan Kaufmann Publishers, ISBN :1558605703*, 1999.
4. David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2) :91–110, 2004.
5. J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press*, page 281–297, 1967.
6. B. S. Manjunath, Philippe Salembier, et Thomas Sikora. *Introduction to MPEG-7 : Multimedia Content Description Interface*. Wiley & Sons, avril 2002.
7. P. Montesinos, V. Gouet, R. Deriche, et D. Pelé. Matching color uncalibrated images using differential invariants. *IVC Journal*, 18(9) :659–672, juin 2000.
8. P. Punitha et D. S. Guru. Symbolic image indexing and retrieval by spatial similarity : An approach based on B-tree. *Pattern Recogn.*, 41(6) :2068–2085, 2008.
9. E. McCreight R. Bayer. Organization and maintenance of large ordered indices. In *Acta Informatica, 1(3)*, pages 173–189, 1972.
10. S. Savarese, J. Winn, et A. Criminisi. Discriminative Object Class Models of Appearance and Shape by Correlatons. In *Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2033–2040, 2006.
11. J. Sivic, B. Russell, A. Efros, A. Zisserman, et W. Freeman. Discovering objects and their location in images. In *International Conference on Computer Vision*, pages 370– 377, 2005.
12. J. Sivic et A. Zisserman. Video Google : A text retrieval approach to object matching in videos. In *International Conference on Computer Vision*, pages 1470–1477, octobre 2003.

Ce travail a été réalisé dans le cadre du PPF WISDOM et de l'ANR MDCO DISCO