



HAL
open science

Proposition d'un descripteur de formes et du modèle vectoriel pour la recherche de symboles

Thi Oanh Nguyen, Salvatore Tabbone, Oriol Ramos Terrades

► To cite this version:

Thi Oanh Nguyen, Salvatore Tabbone, Oriol Ramos Terrades. Proposition d'un descripteur de formes et du modèle vectoriel pour la recherche de symboles. Colloque International Francophone sur l'Écrit et le Document - CIFED 08, Oct 2008, Rouen, France. pp.79-84. hal-00335044

HAL Id: hal-00335044

<https://hal.science/hal-00335044>

Submitted on 28 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proposition d'un descripteur de formes et du modèle vectoriel pour la recherche de symboles

T.-O. Nguyen^{1,2} – S. Tabbone¹ – O. Ramos Terrades¹

¹ LORIA - Université Nancy 2
Campus scientifique - BP 239 - 54506 Vandoeuvre-les-Nancy, France

² Institut de la Francophonie pour l'Informatique
42 Ta Quang Buu, Hanoi, Vietnam

{thi-oanh.nguyen, tabbone, ramos}@loria.fr

Résumé : Dans cet article, nous présentons une méthode alternative pour la recherche de symboles graphiques en adaptant le concept de contexte de forme¹ aux points d'intérêt d'un symbole pour le décrire et en utilisant le modèle vectoriel pour le rechercher. Le descripteur que nous avons adapté est invariant à la rotation, à la translation et au changement d'échelle. Un vocabulaire visuel construit sur des descripteurs nous permet de réduire la complexité d'appariement d'un symbole avec un grand ensemble de candidats en utilisant le modèle vectoriel utilisé en recherche d'information. Les résultats expérimentaux sur des symboles complets et incomplets sont très prometteurs.

Mots-clés : descripteur de symboles, symbole graphique, mot visuel, recherche par le contenu, modèle vectoriel.

1 Introduction

La reconnaissance de formes est toujours un défi suscitant beaucoup de recherches [JAI 00]. La représentation des formes pour la reconnaissance des symboles est aussi un sujet bien étudié et il existe beaucoup de synthèse sur la reconnaissance des formes [TAB 07]. D'habitude, le choix d'une représentation particulière est généralement conduit par des exigences liées à l'applications telles que : la robustesse aux bruits, la stabilité par rapport aux distorsions, l'invariance aux transformations géométriques ou la tolérance aux occlusions.

En général, les descripteurs de formes peuvent être divisés en deux grandes classes : les uns s'appuient sur la forme entière et les autres sur le contour. Les descripteurs basés sur le contour incluent les descripteurs de Fourier [RUI 98, ZHA 05], qui ont été largement utilisés. Les descripteurs basés sur la région prennent en compte tous les pixels de la forme et les méthodes les plus courantes sont basées, par exemple, sur la théorie des moments [PRO 92].

Pour rechercher des formes similaires, des mesures de similarité entre des descripteurs sont définies pour déterminer la distance entre eux. Cependant, la comparaison n'est pas toujours simple et elle exige parfois des transformations. De plus, la dimension du descripteur est souvent élevée, ce

qui augmente la complexité quand on cherche des objets similaires dans une grande collection.

Cet article se situe dans cette problématique de la reconnaissance et la recherche de symboles. Il est basé sur des travaux précédents sur la reconnaissance de formes et la recherche d'images [BEL 02, SIV 06]. Le schéma général de l'approche est présenté dans la Fig. 1. Dans un premier temps, nous introduisons le principe du *Contexte de Forme* pour des Points d'Intérêt (CFPI). C'est une adaptation du descripteur défini dans l'article de S. Belongie et al. [BEL 02] aux symboles graphiques. Plus précisément, des points d'intérêt de chaque symbole sont extraits et des informations dans la région au voisinage de ces points sont représentées par des descripteurs de forme. Le choix du descripteur est motivé par la nature des documents graphiques qui sont, en général, en niveaux de gris ou binaires. Un descripteur de forme s'adapte donc bien à ce document pour capturer des informations. Nous avons choisi le *contexte de forme* du fait de sa performance pour des objets visibles partiellement [BEL 02] et les symboles extraits d'un document graphique sont très souvent partiellement occultés à cause des erreurs de segmentation.

Ensuite, les descripteurs sont regroupés en classes pour construire un vocabulaire visuel qui est un processus d'abstraction. Chaque classe est considérée comme un mot visuel et les *contextes de formes* appartenant à la même classe partagent l'information de forme similaire quel que soit le symbole où le point d'intérêt est extrait. Finalement, chaque symbole est décrit par des mots visuels et apparié au symbole requête. L'approche est similaire à celle de [SIV 06] où l'efficacité de la recherche est atteint grâce aux fichiers inverses et au modèle vectoriel.

Cet article est organisé comme suit. Dans le paragraphe 2, nous présentons en détail une méthode pour la représentation des symboles graphiques en se basant sur le *contexte de forme*. Le système de recherche des symboles utilisant le modèle vectoriel est abordé dans le troisième paragraphe. Ensuite, nous présentons des résultats expérimentaux et l'évaluation de l'adaptabilité du système aux symboles graphiques (§4). Enfin, des conclusions et des perspectives sont données dans le paragraphe 5.

¹Traduit de Shape Context.

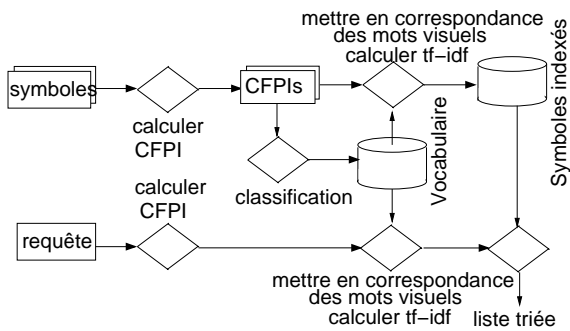


FIG. 1 – Système de recherche de symboles.

2 Notion de *contexte de forme* pour les points d'intérêt

Dans cette partie, nous présentons une solution adaptative pour décrire un symbole. L'approche est basée sur le descripteur proposé par Belongie et al. [BEL 02] qui est invariant à la rotation et au changement d'échelle mais exige de nombreuses descriptions par symbole. Comme le *contexte de forme* contient de l'information abondante et redondante de la configuration locale de l'objet, nous proposons une solution pour diminuer la quantité et la redondance d'information en attachant le contexte uniquement aux points d'intérêts.

2.1 Rappel du *contexte de forme*

Le *contexte de forme* d'un point de contour p_i d'une forme est déterminé par la distribution des points de contour dans la région au voisinage de p_i [BEL 02]. C'est un histogramme de coordonnées relatives des points de contour par rapport à des points p_i , les *points de référence*.

Une forme est représentée par l'ensemble de points échantillonnés des contours externes et internes, $\mathcal{C} = \{p_1, p_2, \dots, p_n\}$, $p_i \in \mathbb{R}^2$ où n est le nombre de points de contour. Pour un point p_i , les coordonnées relatives des $n-1$ autres points sont déterminées. Les coordonnées relatives sont les coordonnées du point dans un système de coordonnées log-polaire utilisant p_i comme origine :

$$q = (\log(r_q), \theta_q), \forall q \neq p_i \wedge q \in \mathcal{C} \quad (1)$$

où r_q est la distance entre q et p_i , θ_q est l'angle entre le vecteur $\vec{p_i q}$ et l'axe horizontal. Le *contexte de forme* h_i du point p_i est défini par :

$$h_i(l) = \#\{q \neq p_i : (q - p_i) \in \text{bin}(l)\}, l = \overline{1, L} \quad (2)$$

où $h_i(l)$ est le nombre de points de contour appartenant à la $l^{\text{ème}}$ classe de l'histogramme et $\text{bin}(l) = \{(r_q, \theta_q) : r_q \in [r_l, r_l + \Delta_{r_l}] \wedge \theta_q \in [\theta_l, \theta_l + \Delta_{\theta_l}]\}$. Une forme de l'objet \mathcal{O} est ainsi décrite comme l'ensemble des *contextes de forme* des points de contour :

$$\mathcal{O} \equiv \{h_i | p_i \in \mathcal{C}\} \quad (3)$$

Cependant, le *contexte de forme* décrit ci-dessus n'est pas invariant à la rotation et au changement d'échelle. Pour obtenir l'invariance au changement d'échelle, les distances radiales sont normalisées par la distance moyenne α des n^2

paires de points de la forme [BEL 02]. Les auteurs ont également proposé d'utiliser le vecteur tangent associé à chaque point au lieu de l'axe absolu horizontal pour que le *contexte de forme* soit aussi invariant à la rotation.

2.2 Contexte de forme pour les points d'intérêt

Comme le *contexte de forme* est un descripteur qui contient des informations de la configuration locale associée à chaque point de contour, il semble bien adapté aux symboles graphiques auxquels nous nous intéressons. Cependant, une représentation d'un symbole en utilisant des *contextes de forme* de presque tous les points de contour engendre une grande complexité de calcul de descripteurs et aussi d'appariement au niveau des symboles au moment de la recherche.

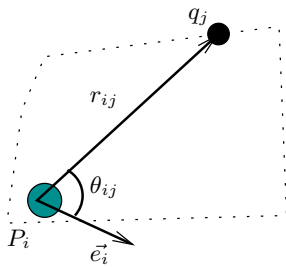
Il existe de nombreuses recherches qui ont montré qu'un objet peut être efficacement localisé à partir de ses points d'intérêt [LOW 04, SIV 06, AGA 04, BOS 06]. Ainsi, dans notre contexte, nous proposons de retenir seulement des informations au niveau des points caractéristiques appelés points clefs ou points d'intérêts pour la description d'un symbole.

Détection des points d'intérêt : Il y a beaucoup de méthodes proposées pour détecter des points d'intérêts d'une image [SCH 98, MIK 05, TAB 05]. Nous avons choisi le détecteur DoG (Difference-of-Gaussian) présenté dans [LOW 04] pour nos expériences. Des points d'intérêts sont considérés comme des extrema existant dans une pyramide des échelles construite par la convolution d'une image I avec des filtres de différences de Gaussienne (voir (4)) de tailles différentes. D'après l'évaluation de Mikolajczyk et al. [MIK 05], le descripteur SIFT (Scale Invariant Feature Transform [LOW 04]), aux points détectés par DoG, donne une très bonne performance. En outre, comme l'opérateur de DoG est une proche approximation de la fonction de Laplacien de Gaussienne, la plupart des points détectés sont à proximité des jonctions du modèle de l'objet [TAB 05] qui jouent un rôle important pour distinguer un modèle d'un autre.

$$D(x, y, \delta) = (G(x, y, k\delta) - G(x, y, \delta)) * I(x, y) \quad (4)$$

Descripteur CFPI : Supposons $\mathcal{IP} = \{P_1, P_2, \dots, P_N\}$ l'ensemble de points d'intérêt et $\mathcal{C} = \{q_1, q_2, \dots, q_n\}$ l'ensemble des points de contour de l'objet. Chaque point dans \mathcal{IP} est considéré comme un point de référence pour calculer le *contexte de forme* correspondant. Pour que l'objet soit bien représenté quelles que soient son orientation et sa taille, le descripteur doit être invariant à la rotation et au changement d'échelle, les coordonnées relatives des points de contour doivent ainsi être normalisées. Cependant, des points d'intérêt ne sont pas toujours des points de contour [TAB 05], c'est-à-dire $\mathcal{IP} \not\subseteq \mathcal{C}$, le vecteur tangent (comme proposé dans [BEL 02]) n'est plus applicable pour la normalisation. Au lieu d'utiliser ce vecteur, nous prenons l'orientation dominante du point d'intérêt comme la direction positive de l'axe x . Alors, chaque point d'intérêt P_i est représenté par ses coordonnées et son orientation dominante :

$$P_i = \{x_i, y_i, \vec{e}_i\} \quad (5)$$


 FIG. 2 – Coordonnées relatives de q_j

Les coordonnées log-polaires relatives d'un point de contour $q_j \in \mathcal{C}$ dans l'équation (1) sont réécrites comme suit :

$$q_j^{P_i} = (\log(r_{ij}), \theta_{ij}) \quad (6)$$

où P_i est le point de référence, r_{ij} est la distance normalisée de q_j à P_i et $\theta_{ij} = \angle \overrightarrow{P_i q_j}, \vec{e}_i >$ (voir Fig. 2).

Le *contexte de forme* associé au point P_i est défini suivant l'équation (2). C'est un histogramme de L classes (L bins). Si 5 intervalles sont choisis pour $\log(r)$ avec r entre 0.125α et 2α et 12 intervalles équivalents pour la coordonnée angulaire θ , comme dans [BEL 02], alors L vaut 60. Un objet \mathcal{O} est maintenant décrit par l'ensemble des *contextes de forme* des points d'intérêt P_i .

$$\mathcal{O} \equiv \{h_i | P_i \in \mathcal{IP}\} \quad (7)$$

3 Recherche des symboles

Selon [BEL 02], la distance entre deux formes est mesurée comme la somme symétrique des coûts de meilleurs appariements des *contextes de forme*. Cela cause un problème de complexité quand on cherche des formes (objets) similaires parmi de nombreux candidats. Dans ce paragraphe, nous présentons une adaptation d'une technique de recherche textuelle pour l'indexation et la recherche des symboles. L'objectif est de réduire la complexité de calcul à l'étape d'appariement, calculée *en-ligne*, grâce aux informations prédéterminées à l'étape *hors-ligne*. Cette approche est semblable à celle de [SIV 06] où le modèle vectoriel de la recherche textuelle est exploité afin d'avoir une recherche efficace en se basant sur des mots visuels.

3.1 Construction d'un vocabulaire visuel

Tout d'abord, des descripteurs CFPI de chaque symbole existant dans la base sont déterminés comme indiqué en §2.2. Chaque descripteur est un vecteur de dimension L . Ensuite, des classes contenant des descripteurs similaires sont créées en utilisant une technique de classification. Chaque classe est considérée comme un mot visuel qui est identifié par le centre de la classe. Tous les points dont le descripteur appartient à cette classe sont aussi considérés comme des occurrences du mot visuel dans la base. Pour faciliter le problème de classification, nous avons utilisé la méthode des k-means avec la distance "cosinus" (l'angle entre deux vecteurs) pour nos tests.

Un symbole est maintenant représenté par des mots visuels et peut être traité comme un document textuel. La Fig. 3

montre un exemple de trois classes définies sur une base de symboles et correspondant à trois mots visuels.

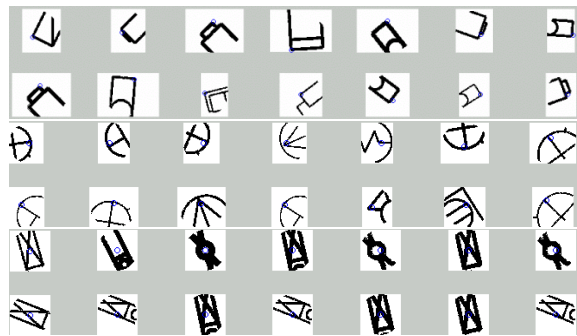


FIG. 3 – Exemple de trois classes correspondant à trois mots visuels.

3.2 Modèle vectoriel pour la recherche des symboles

Le modèle vectoriel est le modèle le plus utilisé en recherche d'informations. Il est plus performant que d'autres modèles classiques pour des collections générales [BAE 99]. Avec ce modèle, un document est représenté comme un vecteur de fréquences d'apparition des mots. Il est souvent décrit par un vecteur de fréquences pondérées dont chaque composante apporte un compromis entre deux facteurs : la fréquence d'apparition des termes (mots) dans le document (facteur tf) et l'estimation du degré d'importance de ce mot pour distinguer un document pertinent des non pertinents de la base (facteur idf).

Chaque descripteur CFPI d'un symbole j est apparié avec une classe la plus proche. Ce descripteur est considéré comme un mot visuel apparaissant dans le symbole j , qui est considéré comme un document j . Dans la suite, les termes "document" et "symbole" sont utilisés indifféremment.

Maintenant, nous pouvons appliquer ce modèle pour l'indexation des symboles. Un symbole correspond à un document contenant des mots visuels. Un symbole j est représenté par un vecteur de $tf-idf$ \vec{s}_j :

$$\vec{s}_j = \{w_{1,j}, w_{2,j}, \dots, w_{K,j}\} \quad (8)$$

où K est la taille du vocabulaire et $w_{i,j}$ est la fréquence pondérée du mot i dans le document j :

$$w_{i,j} = tf_{i,j} * idf_i, i = \overline{1, K}$$

$$tf_{i,j} = \frac{freq_{i,j}}{\max_l freq_{l,j}}; idf_i = \log \frac{N}{n_i}$$

où $freq_{i,j}$ est la fréquence d'apparition du mot i dans le document j , N est le nombre total des documents dans la base et n_i est le nombre de documents dans lesquels le mot i apparaît. $tf_{i,j}$ est la fréquence normalisée du terme.

Recherche : Le vecteur $tf-idf$ du symbole de requête \vec{s}_q est calculé de la même façon : déterminer les descripteurs CFPI, puis, mettre en correspondance ces descripteurs avec des mots visuels et enfin, calculer des fréquences pondérées des mots existants dans le symbole requête. Le degré de similarité entre la requête et un symbole j dans la base de données

est quantifié par leur corrélation mesurée par l'angle entre deux vecteurs \vec{s}_j et \vec{s}_q .

$$\text{sim}(s_q, s_j) = \frac{\vec{s}_j \cdot \vec{s}_q}{|\vec{s}_j| \times |\vec{s}_q|} \quad (9)$$

Les degrés de similarité entre la requête et les symboles dans la base sont triés afin d'obtenir une liste de symboles triés par ordre de pertinence.

4 Résultats expérimentaux

Comme l'objectif de notre système est aussi celui de n'importe quel système de recherche, c'est-à-dire de retrouver des symboles similaires à une requête, classés selon leur degré de similarité, nous utilisons une mesure commune pour évaluer la performance du système : la courbe de *précision/rappel* [BAE 99, SMI 98, GEV 04, MIK 05].

Le **rappel** est défini comme le nombre de documents pertinents retrouvés au regard du nombre de documents pertinents dans toute la base.

$$\text{Rappel} = |Ra|/|R| \quad (10)$$

La **précision** est la fraction de documents pertinents parmi des documents récupérés.

$$\text{Precision} = |Ra|/|X| \quad (11)$$

où X est l'ensemble de documents récupérés, Ra est l'ensemble de documents pertinents trouvés, $Ra \subseteq X$ et R est l'ensemble de documents pertinents qui sont contenus dans la base de données.

Pour les données de test, nous avons choisi la collection de test GREC². Cette collection est divisé en deux ensembles. L'un contient 50 symboles différents correspondant à 50 classes (l'ensemble A), l'autre (l'ensemble B) 250 occurrences de 50 classes de symboles, obtenues par des transformations linéaires (changement d'échelles et de rotation) de chaque élément dans A. Le nombre d'occurrences de chaque classe n'est pas égal, il varie de 1 à 10. L'ensemble B est choisi comme une base d'apprentissage. Les descripteurs déterminés sur cet ensemble servent à la construction d'un vocabulaire visuel et chaque symbole dans B sera indexé comme un document dans la base. Le nombre de classes est fixé expérimentalement à $K = 200$ pour le k-means et le vocabulaire contient donc 200 mots visuels.

L'ensemble A joue le rôle de requêtes test. Comme le nombre maximum des symboles pour une classe est égal à 10, nous ne nous intéressons qu'aux 10 premiers documents. Les valeurs de la précision et du rappel sont calculées pour chaque valeur $r = \overline{1, 10}$ qui correspond aux r premiers documents récupérés. Les 50 requêtes dans A sont effectuées, les valeurs de précision/rappel sont déterminées par la moyenne de toutes les requêtes à chaque valeur de r . Les deux courbes les plus hautes dans la Fig. 4 sont les *précision/rappel* moyens de 50 requêtes correspondant à deux valeurs différentes de dimension de CFPI, $L = 36$ et 60.

Comme indiqué dans la Fig. 4, la courbe atteint une précision élevée (80%) avec un rappel de 70%. Le degré de précision le moins bon (44%) est obtenu avec $r = 10$. En effet, les valeurs de précision et de rappel sont calculées à partir de dix premiers symboles récupérés mais le nombre de symboles pertinents dans la base pour chaque requête n'est pas toujours à 10. Donc, les requêtes dont le nombre total de symboles pertinents dans la base est inférieur au nombre des documents récupérés (r) causent un effet négatif sur la moyenne des précisions. Quelques exemples de résultats sont également montrés sur la Fig. 5.

Nous nous intéressons ici au comportement de notre descripteur aux informations vis à vis des configurations locales. Nous avons fait des expériences avec la dimension du descripteur plus petite $L = 36$, c'est-à-dire, des informations représentées par CFPI $L = 36$ sont "plus locales" que celles représentées par CFPI $L = 60$. L'objectif est de savoir si des informations locales aux points d'intérêt sont suffisantes pour décrire un symbole. Dans ce contexte, au lieu de diviser $[0.125\alpha, 2\alpha]$ en 5 intervalles pour classer les distances radiales r_{ij} , nous ne prenons en compte que les points dont la distance radiale est dans $[0.125\alpha, \alpha]$ divisé en 3 intervalles. Nous utilisons toujours 12 intervalles pour les coordonnées angulaires θ_{ij} , L est donc égal à 36. La Fig. 4 montre le résultat de notre test. Comme remarqué sur la figure, il n'y a pas de différence significative entre $L = 36$ et $L = 60$. Cela prouve que même si l'information représentée par un descripteur est réduite, l'ensemble des descripteurs aux points d'intérêt assure une bonne représentation du symbole et ainsi de la performance de recherche.

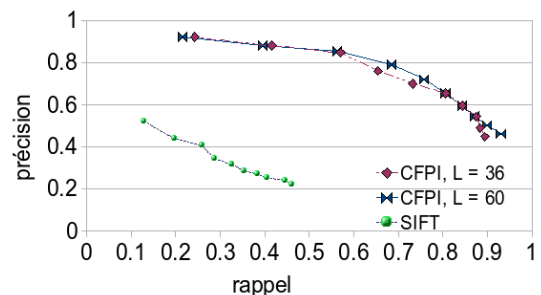


FIG. 4 – Précision/rappel avec descripteurs CFPI (dimension $L = 36, 60$) et SIFT

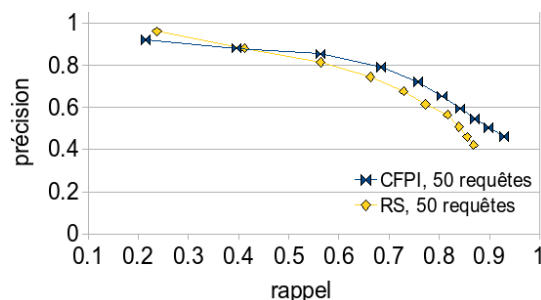


FIG. 6 – Précision/rappel avec CFPI et \mathcal{R} -signature

Comme évalué par [MIK 05], SIFT est le meilleur descripteur dans le cas général. Ainsi, nous avons vérifié l'ef-

²<http://www.cvc.uab.es/grec2003/SymRecContest/>











	0.8665	0.8497	0.7783	0.75833	0.74844	0.73637	0.66721	0.63876	0.47859	0.45994
	0.87801	0.86293	0.84001	0.75694	0.74643	0.73205	0.72836	0.72331	0.68417	0.66483
	0.89175	0.72103	0.70424	0.68103	0.63804	0.60888	0.605	0.56833	0.5272	0.52397
	0.74506	0.69319	0.64009	0.53133	0.53131	0.52822	0.52651	0.52157	0.51939	0.51814
	0.82715	0.75145	0.74055	0.65338	0.63722	0.5217	0.50573	0.49987	0.49039	0.47762
	0.74309	0.73371	0.66902	0.63268	0.61847	0.61282	0.61057	0.60441	0.59287	0.58834
	0.65469	0.62397	0.61916	0.57349	0.55615	0.52488	0.51466	0.50872	0.49731	0.49369
	0.9076	0.90589	0.89023	0.88761	0.85385	0.64745	0.62052	0.59799	0.51314	0.41218
	0.66398	0.62824	0.5996	0.59825	0.58326	0.57343	0.56617	0.51135	0.47655	0.47314
	0.80455	0.78785	0.75291	0.62738	0.54045	0.52316	0.49704	0.46458	0.45425	0.42255

FIG. 5 – Exemples de résultats. La première colonne contient des requêtes. Les autres sont des symboles les plus similaires à la requête. Le nombre de symboles pertinents pour chaque requête est égal respectivement à 7, 4, 3, 5, 4, 1, 4, 5, 7, 3.

ficacité, en vue de rechercher des symboles, des descripteurs CFPI, SIFT (vecteur de dimension 128) et un autre descripteur de forme \mathcal{R} -signature [TAB 04] qui donne des résultats également intéressants pour la reconnaissance des formes. Pour le test du descripteur SIFT, nous avons utilisé le même processus décrit dans cet article en remplaçant CFPI par SIFT. Pour la \mathcal{R} -signature, comme un symbole est représenté par un descripteur, nous faisons juste la comparaison entre le descripteur de la requête et les autres descripteurs correspondant aux symboles de la base et déterminons les symboles les plus proches. Notre but n'est pas de faire une évaluation des descripteurs, mais de montrer que le descripteur SIFT n'est pas bien adapté pour des symboles graphiques. Des résultats obtenus avec CFPI montrent une courbe de précision/rappel meilleure pour notre descripteur que SIFT (cf Fig. 4). Les résultats obtenus par la \mathcal{R} -signature sont quasi semblables (cf Fig. 6). Cependant, pour la requête avec des symboles incomplets, nous allons montrer ci-dessous que CFPI fournit des meilleurs résultats.

Afin de vérifier l'adaptabilité des descripteurs et du modèle vectoriel pour rechercher des symboles incomplets, nous avons aussi fait des tests avec notre système et avec la \mathcal{R} -signature en prenant des symboles incomplets comme requêtes (voir Fig. 7). Ce test montre que le système fournit une possibilité de retrouver des symboles similaires même si ce n'est pas un symbole complet. Cette capacité de prendre en compte des symboles partiellement définis est un avantage pour construire, dans l'avenir, un système de recherche de symboles dans des grands documents graphiques.

5 Conclusions et perspectives

Nous avons proposé une adaptation du contexte de forme pour décrire des symboles graphiques. Le descripteur CFPI est simple et invariant à la rotation et au changement d'échelles. Il fournit une bonne représentation de la configuration locale correspondant à chaque point d'intérêt et nous permet de réduire la complexité de description d'un symbole. De plus, en utilisant une technique de recherche textuelle à partir du modèle vectoriel appliqué à des mots visuels, la complexité d'appariement est aussi réduite.

Les résultats expérimentaux sont prometteurs mais ils doivent être considérés comme préliminaires. Quand un symbole est représenté par des CFPIs, si le nombre de points d'intérêt du symbole est trop petit, le vecteur représentant le symbole n'assure pas une description efficace du "contenu" du symbole car le nombre de mots visuels est faible. Donc, une de nos perspectives est de chercher une solution pour ajouter "automatiquement" des points clés quand le symbole a un faible nombre de points d'intérêt. De plus, notre travail présenté ici n'est que la première partie d'un système qui ne permettra pas seulement de trouver des symboles isolés mais aussi de récupérer des symboles localisés dans des documents graphiques. Notre objectif est de définir une méthode pour indexer et détecter des symboles dans une grande collection de documents graphiques.

Références

[AGA 04] AGARWAL S., AWAN A., ROTH D., Learning to detect objects in images via a sparse, part-based represen-

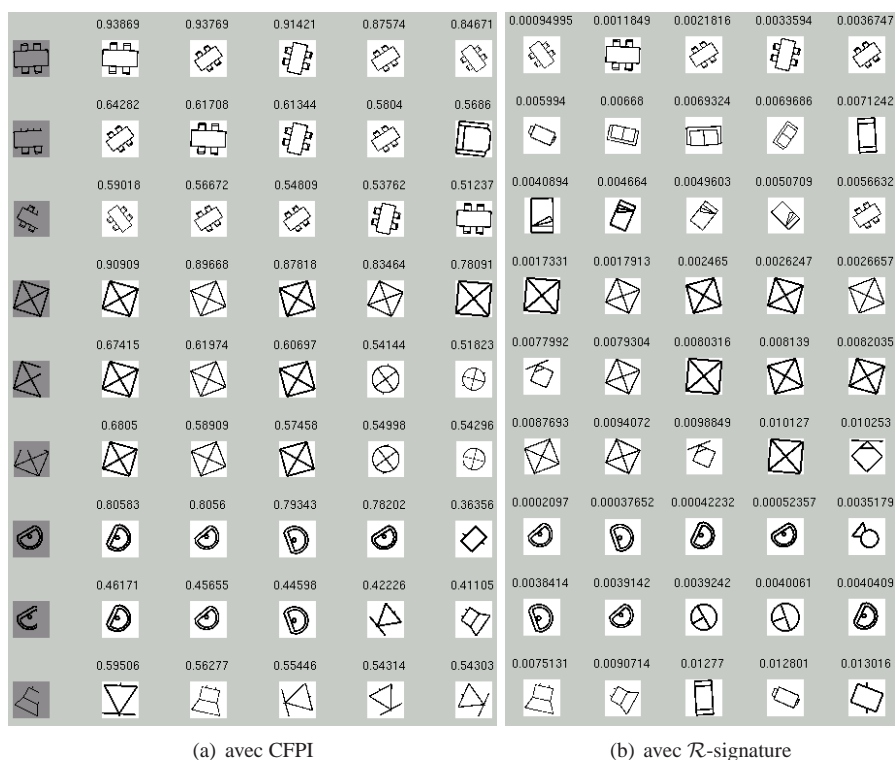


FIG. 7 – Résultats obtenus pour des requêtes avec des symboles incomplets. La première colonne contient les requêtes. Les 5 plus proches symboles récupérés avec CFPI (a) et avec \mathcal{R} -signature (b) sont présentés aux colonnes de 2 à 10.

tation, *PAMI*, vol. 26, n° 11, 2004, pp. 1475-1490.

[BAE 99] BAEZA-YATES R., RIBEIRO-NETO B., *Modern Information Retrieval*, ACM Press / Addison-Wesley, New York, 1999.

[BEL 02] BELONGIE S., MALIK J., PUZICHA J., Shape Matching and Object Recognition Using Shape Contexts, *PAMI*, vol. 24, n° 4, 2002, pp. 509-522.

[BOS 06] BOSCH A., ZISSERMAN A., MUNOZ X., Scene Classification via pLSA, *Computer Vision, ECCV 2006*, vol. 3954/2006, pp. 517-530, Springer Berlin / Heidelberg, May 2006.

[GEV 04] GEVERS T., SMEULDERS A., *Emerging Topics in Computer Vision*, Chapitre Content-Based Image Retrieval : An Overview, Addison-Wesley / Prentice Hall, 2004.

[JAI 00] JAIN A., DUIN R., MAO J., Statistical pattern recognition : a review, *PAMI*, vol. 22, n° 1, 2000, pp. 4-37.

[LOW 04] LOWE D. G., Distinctive image features from scale-invariant keypoints, *IJCV*, vol. 60, n° 2, 2004, pp. 91-110.

[MIK 05] MIKOLAJCZYK K., SCHMID C., A performance evaluation of local descriptors, *PAMI*, vol. 27, n° 10, 2005, pp. 1615-1630.

[PRO 92] PROKOP R., REEVES A., A survey of moment-based techniques for unoccluded object representation and recognition, *CVGIP : Graphical Models and Image Processing*, vol. 54, n° 5, 1992, pp. 438-460.

[RUI 98] RUI Y., SHE A., HUANG T., A modified Fourier descriptor for shape matching in MARS, *Workshop on Image Databases and Multi Media Search*, vol. 8, 1998, pp. 165-180.

[SCH 98] SCHMID C., MOHR R., BAUCKHAGE C., Comparing and evaluating interest points, *ICCV, Bombay, India*, 1998, pp. 230-235.

[SIV 06] SIVIC J., ZISSERMAN A., Video Google : Efficient Visual Search of Videos, *Toward Category-Level Object Recognition*, vol. 4170/2006, pp. 127-144, Springer Berlin / Heidelberg, 2006.

[SMI 98] SMITH J. R., Image Retrieval Evaluation, *IEEE Workshop on Content-based Access to Image and Video Databases*, IEEE Computer Society Washington, DC, USA, June 1998, page 112.

[TAB 04] TABBONE S., WENDLING L., Recognition of symbols in grey level line drawings from an adaptation of the Radon transform, *Proceedings of 17th ICPR, Cambridge (UK)*, vol. 2, 2004, pp. 570-573.

[TAB 05] TABBONE S., ALONSO L., ZIOU D., Behavior of the Laplacian of Gaussian Extrema, *Journal of Mathematical Imaging and Vision*, vol. 23, n° 1, 2005, pp. 107-128.

[TAB 07] TABBONE S., LLADOS L., A propos de la reconnaissance de documents graphiques : synthèse et perspectives, *TAIMA, Hammamet*, 2007.

[ZHA 05] ZHANG D., LU G., Study and evaluation of different Fourier methods for image retrieval, *Image and Vision Computing*, vol. 23, 2005, pp. 33-49.