

# Coopération de descripteurs locaux-globaux pour l'indexation d'images à niveaux de gris

F. Ghorbel<sup>(1,2)</sup>, M. Chakchouk<sup>(1)</sup> et S. Derrode<sup>(3)</sup>

(1) Projet Imagerie Numérique et Applications (PINA)

Centre d'Etudes et de Recherche des Télécommunications (CERT)

42, rue Asdrubal 1002 Lafayette Tunis, Tunisie

(2) Groupe de Recherche en Images et Formes de Tunisie (GRIFT) Lab. CRISTAL

Ecole Nationale des Sciences de l'Informatique (ENSI)

Campus Universitaire La Manouba, Tunisie

(3) Département Image et Traitement de l'Information (ITI)

Ecole Nationale Supérieure des Télécommunications Bretagne (ENSTB)

BP 835, 29285 Brest Cedex, France

## **Emails :**

faouzi.ghorbel@cert.mincom.tn,

chakchouk.moez@cert.mincom.tn et stephane.derrode@enst-bretagne.fr

## 1 Introduction

Il est bien connu que l'indexation de base d'images ou de vidéos est un thème de recherche qui présente un enjeu important pour les opérateurs de télécommunications. Récemment, plusieurs groupes de recherche ont mené des travaux sur les techniques de hachage, la classification des signatures, la gestion de bases de données, les interfaces de recherche d'images et l'adaptation des descriptions d'images dans ce cadre. Dans la plupart des méthodes de description proposées, la propriété d'invariance par rapport à certains types de transformation (euclidienne, affine, projective, ou locale etc.), est souvent souhaitée. Ces descripteurs permettent de décrire, de manière locale ou globale des caractéristiques de l'image (les contours, la couleur, la texture, la forme 3D, le mouvement, etc.).

Dans cette contribution, nous proposons d'enrichir nos méthodes de description pour améliorer les résultats de recherche d'images à niveaux de gris. Pour cela, nous proposons de réaliser une coopération des descripteurs globaux et des descripteurs locaux qui présentent respectivement des propriétés d'invariance à certaines transformations globales (similitudes planes) et locales (occultation,

changement d'illumination, etc.). Pour la description globale, nous utilisons une famille invariante et complète construite à partir des moments géométriques complexes [1]. Pour la description locale, nous nous sommes inspirés de la méthode introduite dans [2] qui est basée sur des descripteurs différentiels. Nous proposons d'améliorer les performances de ces derniers descripteurs (sensibilité au bruit, nombre de points d'intérêts, etc.) [3] en calculant des invariants obtenus à partir des moments géométriques sur un voisinage de chaque point d'intérêt. Cette méthode de description locale-globale est invariante par rapport aux transformations locales (déplacements par morceaux) et par conséquent, elle l'est également par rapport aux transformations globales (déplacements globaux).

Dans ce travail, la deuxième section est consacrée à la présentation du cadre théorique des méthodes de description utilisées. Ainsi, nous commençons par rappeler l'expression de la famille complète de descripteurs invariants par rapport aux similitudes. Celle-ci est basée sur les moments géométriques. Nous rappelons ensuite la méthode d'extraction des primitives locales (les points d'intérêts) utilisés. Dans la troisième section, nous formulons le problème de coopération de descripteurs par l'utilisation de quelques résultats mathématiques sur les distances. Dans la quatrième section, nous montrons l'avantage de la coopération entre les descripteurs globaux et locaux pour la recherche d'images à niveaux de gris. Pour cela, nous présentons les résultats expérimentaux de recherche d'images réalisés sur la base d'objets de l'université de Columbia.

## 2 Formulation de la description

### 2.1 Descripteurs globaux

Différents types de descripteurs globaux pour les images à niveaux de gris vérifiant la propriété de complétude ont été récemment introduits [4][1][5]. Ces descripteurs peuvent être calculés soit à partir des moments complexes, des moments géométriques ou bien à partir de la transformée de Fourier-Mellin. Nous avons montré dans [1] que la coopération de différents descripteurs globaux n'améliore pas sensiblement les résultats d'indexation. C'est pour ces raisons que nous proposons de ne conserver qu'un type de descripteur global. Pour des raisons de clarté de l'exposé, nous proposons dans ce papier d'étudier les descripteurs globaux complets calculés à partir des moments géométriques. Nous rappelons que les moments géométriques centrés d'ordre  $p$  d'une image plane  $f$  considérée comme une fonction bornée de  $\mathbb{R}^2$  vers  $\mathbb{R}$ , est donné par l'expression suivante :

$$\forall k = 0, \dots, p, m_{p-k,k} = \int \int_{\mathbb{R}^2} (x - x_G)^{p-k} (y - y_G)^k f(x, y) dx dy, \quad (1)$$

où  $(x_G, y_G)$  sont les coordonnées du centre de gravité de l'image. Les moments centrés réduits obtenus en normalisant les moments centrés  $m_{p-k,k}$  par rapport

à  $m_{0,0}$  permettent de définir les polynômes suivants :

$$\forall (p, r) \in \mathbb{N}^2, 0 \leq r \leq p, \mathcal{P}_f(p-r, r) = \frac{1}{r! C_p^k} \sum_{k=0}^{p-r} Q_{p-r,k}^r(f), \quad (2)$$

$$\text{où } \begin{cases} Q_{p,k}^0(f) = (-i)^k C_p^k \cdot \frac{m_{p-k,k}}{m_{0,0}} \\ Q_{p-r,k}^r(f) = (p-k-r+1)Q_{p-r+1,k}^{r-1}(f) - (k+1)Q_{p-r+1,k+1}^{r-1}(f) \end{cases} \cdot (3)$$

Les descripteurs invariants complets à base de moments géométriques proposés dans [1] sont alors donnés par la formule suivante :

$$\mathcal{J}_f(p-r, r) = |P_f(2, 1)|^{-\frac{p}{3}} \cdot P_f(p-r, r) \cdot e^{i(p-2r) \arg(P_f(0,1))}. \quad (4)$$

## 2.2 Primitives locales

Les points d'intérêts sont connus par leur efficacité pour la description locale des images à niveau de gris. Ils correspondent à une variation importante du niveau de gris suivant plusieurs directions. Ces points d'intérêts sont connus pour être stables et invariants par rapport à différentes transformations planes locales ou globales [6]. Dans ce papier, nous utilisons la méthode introduite dans [7] pour détecter les points d'intérêts. Après leur détection, une description locale est ensuite réalisée en calculant les descripteurs des moments géométriques sur un voisinage donné de chaque point d'intérêt.

## 3 Coopération local-global : Distances

Un bon nombre de résultats mathématiques sur les distances donnent une solution élégante et convainquante au problème de comparaison des formes. En effet, le choix du type de distance entre descripteurs est une question fondamentale et loin d'être triviale. Cette question a souvent retenu l'attention des chercheurs dans le domaine de la reconnaissance de formes. Dans le contexte du système d'indexation que nous proposons, trois questions sur le choix des distances à utiliser sont bien posées, à savoir :

- Une première problématique plus simple concerne le cas où la dimension des descripteurs est la même quelque soit l'image. Cette propriété se trouve chez les descripteurs globaux. Cela se ramène souvent à chercher la meilleure parmi les distances usuelles définies sur les espaces vectorielles (euclidienne, infinie, etc.) en cherchant l'espace correspondant des descripteurs [8].
- La seconde, plus délicate, correspond à la situation où la dimension des primitives n'est pas la même. C'est le cas des invariants locaux. L'utilisation de la distance de Hausdorff paraît une solution intéressante et qui donne satisfaction dans la plupart des configurations.

- La coopération entre les deux types de descripteurs nécessite l’association de deux distances de natures différentes. La normalisation des deux distances est une opération préalablement indispensable afin de ne pas donner plus de contribution à un type de descripteur qu’à un autre.

Dans ce qui suit, nous explicitons les expressions des différentes distances utilisées. La distance entre les descripteurs globaux de deux objets  $\mathbf{F}$  et  $\mathbf{G}$ , est choisie comme la norme de la différence des descripteurs :

$$d_p(\mathbf{F}, \mathbf{G}) = \left( \sum_{k=1}^N |J_k(\mathbf{F}) - J_k(\mathbf{G})|^p \right)^{\frac{1}{p}} \quad (5)$$

où  $J_k(\mathbf{F})_{k=1,N}$  est la suite de descripteurs de l’objet  $\mathbf{F}$  tronquée jusqu’à l’ordre  $N$  et  $p$  est un nombre réel supérieur ou égal à l’unité.

Le nombre de points d’intérêts dépend de la complexité de l’image. Notons par  $P_i(\mathbf{F})$ ,  $i = 1..N_1$ , les points d’intérêts d’une image  $\mathbf{F}$  et par  $J_k(P_i, \mathbf{F})$  la suite des descripteurs invariants calculés en  $P_i(\mathbf{F})$  qui forment un nuage de points dans l’espace des invariants. La distance de Hausdorff, entre les deux nuages de points correspondants à deux images différentes  $\mathbf{F}$  et  $\mathbf{G}$ , peut s’écrire comme :

$$\Delta(\mathbf{F}, \mathbf{G}) = \max(\rho(\mathbf{F}, \mathbf{G}), \rho(\mathbf{G}, \mathbf{F})), \quad (6)$$

$$\begin{aligned} \text{où } \rho(\mathbf{F}, \mathbf{G}) &= \max_{1 \leq i \leq N_1} \min_{1 \leq j \leq N_2} \left( \sum_{k=1}^D |J_k(P_i, \mathbf{F}) - J_k(P_j, \mathbf{G})|^p \right)^{\frac{1}{p}} \\ \rho(\mathbf{G}, \mathbf{F}) &= \max_{1 \leq j \leq N_2} \min_{1 \leq i \leq N_1} \left( \sum_{k=1}^D |J_k(P_i, \mathbf{F}) - J_k(P_j, \mathbf{G})|^p \right)^{\frac{1}{p}} \end{aligned} \quad (7)$$

et  $D$  est l’ordre de troncature de la suite des invariants des moments calculés sur les points d’intérêt.

Il est bien connu en mathématique [9] que la distance normalisée  $\hat{d}$  d’une métrique  $d$  donnée par :

$$\hat{d} = \frac{d}{1 + d} \quad (8)$$

est topologiquement équivalente à celle-ci. Afin de comparer les formes en se basant conjointement sur les invariants locaux et les invariants globaux, tout en maîtrisant l’importance donnée à chaque type de descripteur, nous proposons de calculer la métrique suivante :

$$\delta(\mathbf{F}, \mathbf{G}) = \alpha_{global} \cdot \hat{d}_p(\mathbf{F}, \mathbf{G}) + \alpha_{local} \cdot \hat{\Delta}(\mathbf{F}, \mathbf{G}) \quad (9)$$

où  $\alpha_{local}$  et  $\alpha_{global}$  sont des réels positifs tels que  $\alpha_{local} + \alpha_{global} = 1$

## 4 Résultats en indexation

Les expériences présentés ci-dessous sont réalisés sur la base d’objets de l’université de Columbia. Cette base contient 1440 images représentant 20 objets pris dans 72 positions différentes. La taille des images est de  $128 \times 128$  pixels.

Au vu des résultats de recherche que nous avons obtenu, nous tentons de donner dans le tableau ci-dessous une comparaison des performances de chaque type de descripteurs.

<b>Descripteurs</b>	<b>Discrimination</b>	<b>Temps de calcul</b>	<b>Résultat aberrant</b>
<b>Locaux</b>	-	-	+
<b>Globaux</b>	+	+	-

Il est important de noter que les invariants locaux donnent un meilleur score pour les images présentant une symétrie relativement à l'image requête (voir figure 1). Toutefois, ces invariants globaux peuvent être modifiés dans l'objectif d'être muni de la propriété d'invariance par rapport aux transformations de ce type. La figure 2 confirme que les invariants locaux ont un meilleur comportement vis à vis des problèmes d'occlusion.

Comme décrit dans la section 3, la coopération consiste à la normalisation des deux distances, puis au calcul de la combinaison linéaire de ces distances normalisées (8). Dans les résultats présentés, nous ne privilégions pas un type de descripteur par rapport à l'autre. Cela se traduit par  $\alpha_{local} = \alpha_{global} = \frac{1}{2}$ . Un point essentiel qui se dégage de ces expérimentations se résume dans ce qui suit : la coopération semble apporter une certaine stabilisation des résultats de recherche. Cela signifie que dans le cas où l'un des deux est meilleur, la coopération fournit un résultat intermédiaire. Cela évite le plus souvent les résultats de recherche aberrants. Par conséquent, les performances en moyenne tendent à s'améliorer, au sens de la discrimination et de la stabilité.

## 5 Conclusion et perspectives

Dans ce papier, nous avons comparé puis réalisé une coopération entre deux descripteurs de nature différente. Un premier basé sur les moments géométriques de l'image et muni de la propriété de complétude est de nature globale. Un second basé sur des invariants des moments calculés sur un voisinage autour de points d'intérêts est de type local. Les résultats qui sont souvent complémentaires, nous ont permis de proposer un moteur de recherche pour les bases d'images à niveaux de gris qui soit stable. C'est à dire un moteur qui, en moyenne, donne des résultats meilleurs que les deux pris séparément. Cette étude a aussi permis de dégager un certain nombre de perspectives pour améliorer les performances de ce moteur, à savoir :

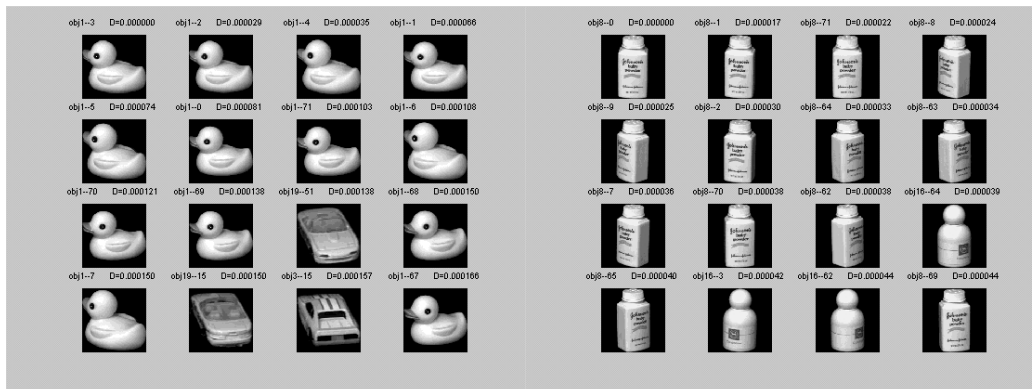
- L'utilisation d'autres méthodes pour l'extraction des points d'intérêts ou de la notion de petites régions d'intérêts qu'on pourrait obtenir par une segmentation statistique ou morphologique.

- Dans nos travaux précédents [10], nous avons proposé un moteur de recherche local-global utilisant une description des contours des images. La coopération avec ce type de primitive pourrait amener à améliorer les performances globales de manière sensible.

**Remerciements :** *Nous tenons à remercier M. Gérard Eude, Chef du Service Programmation et Action Scientifique de France Télécom R&D, pour l'intérêt qu'il porte à nos travaux sur l'indexation.*

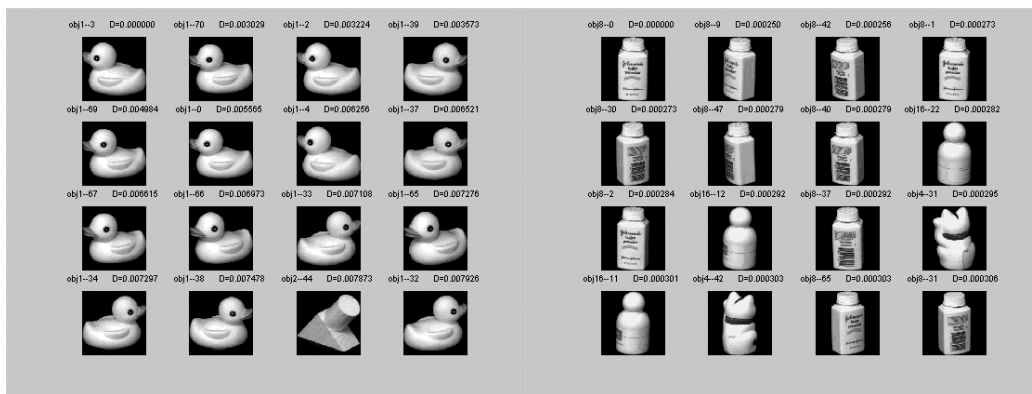
## Références

- [1] S. Derrode, R. Mezhoud, and F. Ghorbel. Comparison between two complete sets of shape descriptors for 2D grey-level object content-based retrieval. *Annales des Télécommunications*, pages 77–86, 2000.
- [2] C. Schmid and R. Mohr. Local gray-value invariants for image retrieval. *IEEE Transactions on Pattern Analysis Machine Intelligence*, pages 530–535, 1997.
- [3] J.M. Jolion and S. Bres. Influence du codage jpeg sur des descripteurs d'images. *Traitement du signal*, pages 309–320, 1999.
- [4] S. Derrode. *Représentation de formes planes à niveaux de gris par différents approximations de Fourier-Mellin analytique en vue d'indexation de bases d'images*. PhD thesis, Université de Rennes I, 1999.
- [5] F. Ghorbel. A complete invariant description for gray-level images by the harmonic analysis approach. *Pattern Recognition Letters*, pages 1043–1051, 1994.
- [6] R. Deriche and G. Giraudon. A computational approach for corner and vertex detection. *Int. Journal of Computer Vision*, 10(2) : pages 101–124, 1993.
- [7] C. Harris and M. Stephens. A combined corner and edge detector. *Alvey Vision Conference*, pages 147–151, 1988.
- [8] F. Ghorbel. Distance et invariance pour la classification d'objets dans les images. *International Meeting on Distance Analysis, DISTANCIA'92*, pages 130–134, 1992.
- [9] J. Dieudonné. *Eléments d'analyse, Cahiers scientifiques*, volume 2. Gauthier-Villars, Paris, 1983.
- [10] M. Kamoun, M. Chakchouk, N. Triki, and F. Ghorbel. Indexation de bases d'images par le contenu en faisant coopérer deux types de descripteurs de formes. *CORESA'99, Sophia Antipolis France*, pages 319–326, 1999.



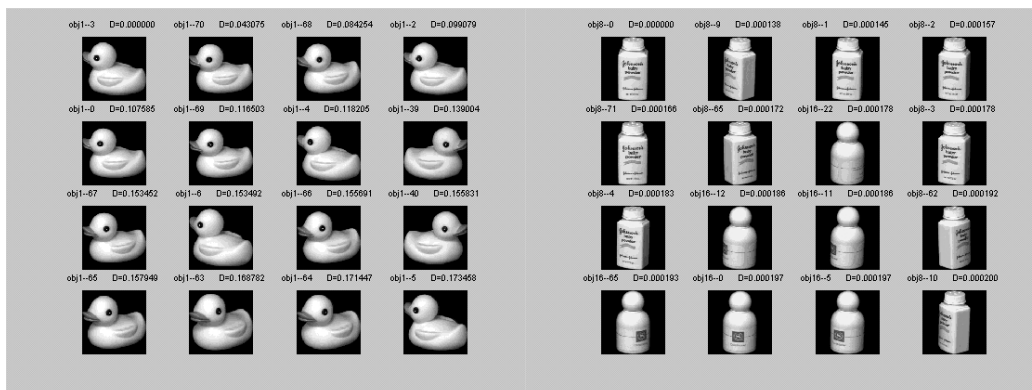
(a)

(b)



(c)

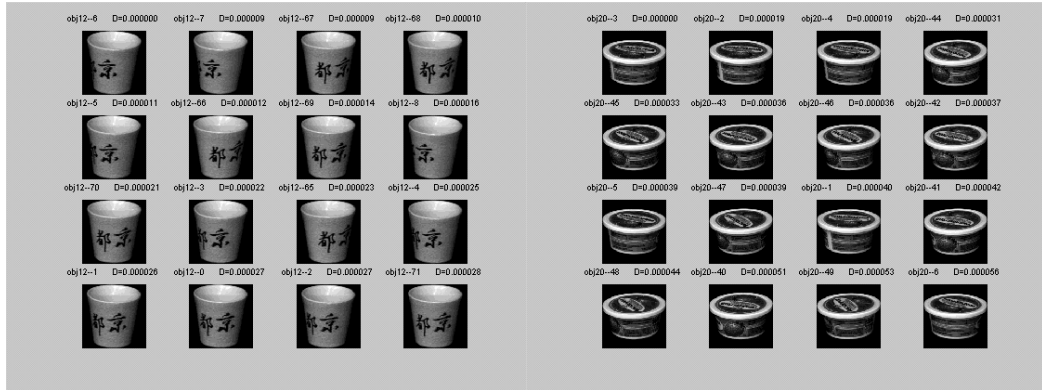
(d)



(e)

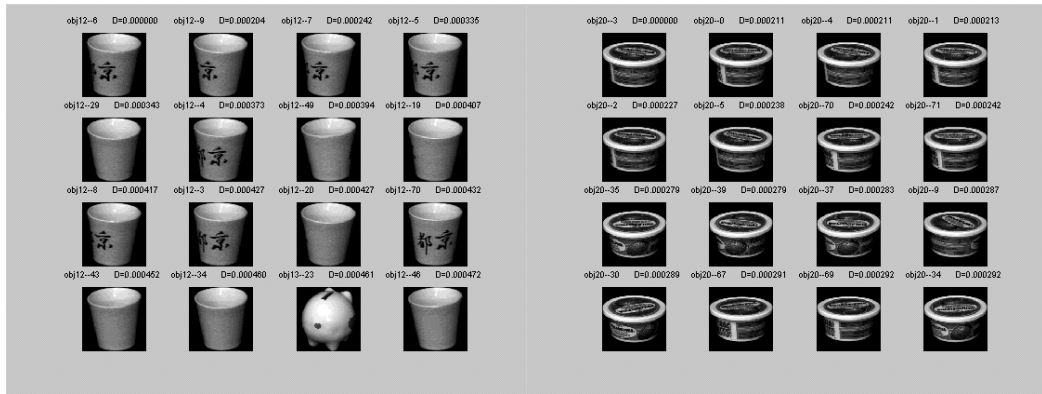
(f)

FIG. 1 – Résultat de recherche d'images ; Images requêtes : Obj1\_3 (à droite) et Obj8\_0 (à gauche) en utilisant les descripteurs globaux (a),(b) ; les descripteurs locaux (c),(d) ; la coopération des deux descripteurs (e),(f) .



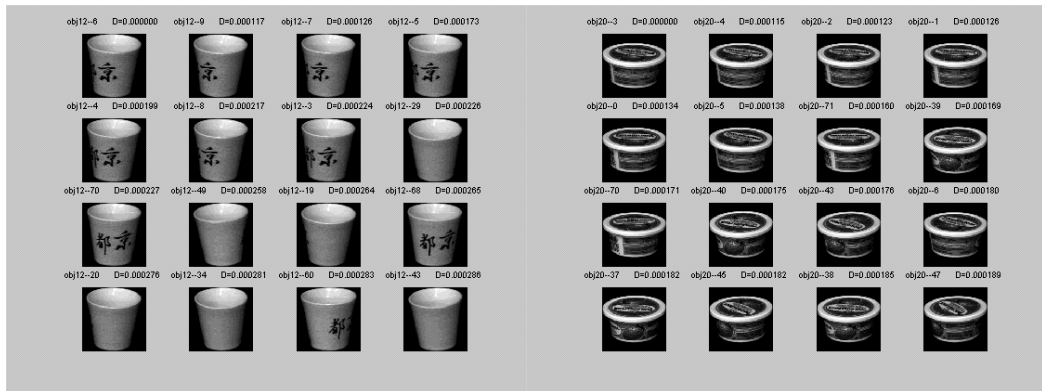
(a)

(b)



(c)

(d)



(e)

(f)

FIG. 2 – Résultat de recherche d'images ; Images requêtes : Obj12\_6 (à droite) et Obj20\_3 (à gauche) en utilisant les descripteurs globaux (a),(b) ; les descripteurs locaux (c),(d) ; la coopération des deux descripteurs (e),(f).