

Comparaison de deux familles complètes de descripteurs de formes pour l'indexation de bases d'objets 2D à niveaux de gris

Stéphane DERRODE*
Rim MEZHOUD**
Faouzi GHORBEL**

Résumé

Cet article présente deux familles de descripteurs de formes pour l'indexation et la recherche d'objets à niveaux de gris dans des bases d'images. Ces familles sont invariantes par similitudes planes et complètes. La propriété de complétude garantit qu'un objet est identifié de manière unique à une similitude près et assure une discrimination totale des formes, ce qui se révèle intéressant lorsque les collections d'images sont soit fortement homogènes, soit très volumineuses. La première famille complète de descripteurs invariants est issue du prolongement analytique de la transformée de Fourier-Mellin et la seconde est extraite des moments complexes de l'images. La correspondance entre les formes est estimée grâce à des mesures de similarité induites par les deux représentations invariantes. Ces familles ont été testées et évaluées ces familles sur deux bases d'objets à niveaux de gris. Les descripteurs se montrent discriminants et robustes aux faibles déformations de formes et les résultats de recherche sont encourageants. La propriété de complétude apparaît comme une propriété intéressante qui offre une représentation riche et flexible des formes.

Mots-clés : Indexation, transformée de Fourier-Mellin, moments complexes, descripteurs invariants, similitudes, robustesse, complétude, similarité, distance entre formes.

COMPARISON BETWEEN TWO COMPLETE SETS OF SHAPE DESCRIPTORS FOR 2D GREY-LEVEL OBJECT CONTENT-BASED RETRIEVAL

Abstract

In this work, we propose the use of two sets of 2D shape descriptors for grey-level object content-based retrieval and image indexing. These families are invariant under planar similarities and complete. The com-

pleteness property ensures that an object is identified in a unique way up to a similarity transformation and guarantees a perfect discrimination between shapes. This reveals to be interesting when image databases are either homogeneous or large. The first set of invariant descriptors is extracted from the analytical Fourier-Mellin transform, while the second one comes from the complex moment image representation. The similarity between shapes is estimated according to the distances induced by the two invariant descriptions. We tested and compared these families on two real grey-level object databases for content-based image retrieval. Retrieval results show that the invariant descriptors produce a reliable shape description, discriminating and stable to small shape variations. The completeness is an interesting feature that offers an extensive and flexible representation of shapes.

Keywords : Indexation, Fourier-Mellin transform, complex moments, invariants, stability, completeness, similarity, distance between shapes.

Sommaire

- I. Introduction
 - II. Descripteurs de formes d'objets plans à niveaux de gris
 - III. Deux familles complètes de descripteurs de formes
 - IV. Résultats expérimentaux de recherche par le contenu
 - V. Conclusions et perspectives
- Bibliographie (25 réf.)

I. INTRODUCTION

L'essor des dispositifs d'acquisition (caméras numériques, scanner, IRM, ...), des capacités de stockage (mémoire de masse), des systèmes de transmission et de

* Département Image et Traitement de l'Information (ENST Bretagne), BP 832, F 29285 Brest cedex, France. Stephane.Derrode@enst-bretagne.fr

** Groupe de Recherche Images et Formes de Tunisie (ENSI) & Centre des Études et Recherches des Télécommunications, 42 Rue Asdrubal, 1002 Tunis, Tunisie. Rim.Mezhoud@ensi.rnu.tn, Faouzi.Ghorbel@cert.mincom.tn

diffusion (réseau Internet, réseaux spécialisés à hauts débits) a favorisé la création et le développement de bases d'images volumineuses et de natures très diverses. Ceci pose le problème de l'archivage, de l'exploitation et de la consultation de ces bases. Les techniques traditionnelles d'indexation manuelle de ces bases ont rapidement montré leurs limites et l'accès à l'information par le contenu visuel des images est devenu, depuis une dizaine d'années, un sujet de recherche particulièrement dynamique. Ainsi, de nombreuses techniques ont été développées et plusieurs états de l'art ont été présentés [AIG96, RUI99].

Un des principaux problèmes de l'indexation d'images par le contenu est de choisir une représentation pertinente des images donnant naissance à des primitives visuelles significatives et fiables qui traduisent le contenu sémantique de la base. En général, ces primitives sont regroupées en 3 classes :

- les descripteurs liés à la couleur : histogrammes et moments,
- les descripteurs de texture : matrice de cooccurrence, direction principale et rugosité, filtres de Gabor et ondelettes,
- les descripteurs de formes : descripteurs de Fourier et des moments, segments, courbure et points caractéristiques.

L'extraction de primitives pertinentes est un problème difficile qui ne connaît pas de solution dans le cas général. Les descripteurs dépendent de la nature de la base, qu'elle soit hétérogène ou homogène comme une base de visages ou une base d'empreintes digitales, et de l'application qui en est faite. En effet, il semble qu'il n'existe pas d'attribut unique qui puisse modéliser une base d'images selon tous les points de vues pour des raisons multiples qui peuvent être illustrées par la subjectivité de la requête de l'utilisateur. Les moteurs de recherche tentent donc de combiner des primitives variées pour améliorer la discrimination et la classification. Par conséquent, le nombre de descripteurs de formes est de plus en plus élevé et l'un des problèmes est alors de déterminer une mesure de similarité qui tienne compte du poids affecté à chaque type de descripteurs. La plupart des moteurs font également intervenir l'utilisateur par le moyen d'une interface interactive qui l'associe à l'interprétation des résultats en permettant, par exemple, la sélection de plusieurs requêtes simultanées (QBIC [FLI95], Virage [BAC96]). Actuellement, pour s'adapter à la subjectivité de la requête de l'utilisateur, les efforts portent sur la conception de systèmes d'apprentissage dynamique qui modifient les poids affectés aux descripteurs au fur et à mesure des requêtes lancées successivement par l'utilisateur (FourEye et Photobook [PEN96, MIN98], Mars [RUI98]).

Dans ce travail, nous mettons l'accent sur l'apport de quelques méthodes de représentation des images et de reconnaissance de formes (descripteurs et classification) pour l'indexation de bases d'objets à niveaux de gris¹ et la recherche par le contenu. Dans la section II, nous pré-

sentons un bref rappel des descripteurs issus de méthodes de reconnaissance de formes ainsi que leur utilisation dans les moteurs de recherche par le contenu. Nous y présentons notamment la propriété de complétude qui semble souhaitable pour de tels systèmes et apparaît utile dans l'esprit de la future norme MPEG-7. Dans la section III, deux familles complètes de descripteurs invariants par similitudes sont extraites à partir de la représentation des images par la transformée de Fourier-Mellin et les moments complexes. Dans la section IV, ces deux familles sont comparées pour leur application en recherche d'images par le contenu avec les sept invariants de Hu qui ne possèdent pas les propriétés énoncées. Les résultats sont illustrés par des expériences menées sur deux bases d'objets réels. Enfin, nous concluons sur quelques perspectives permettant d'enrichir le système expérimental proposé.

II. DESCRIPTEURS DE FORMES D'OBJETS PLANS À NIVEAUX DE GRIS

La plupart des descripteurs d'objets que l'on rencontre dans les systèmes d'indexation ont été introduits dans le cadre d'applications en reconnaissance/discrimination de formes. Le problème fondamental est de déterminer dans quelle mesure deux objets sont similaires, indépendamment de leur pose dans l'image. Ceci est généralement réalisé grâce au calcul de descripteurs de formes qui doivent satisfaire certains critères parmi lesquels :

- l'invariance à un ensemble de transformations géométriques modélisant le point de vue ou le mouvement apparent des objets,
- la robustesse vis à vis des faibles variations de formes et des approximations numériques.
- la simplicité pour un calcul en « temps réel »,

Il est important de distinguer les descripteurs des objets-contours et les descripteurs des objets à niveaux de gris ainsi que le caractère local ou global de ces descripteurs.

Dans le cas des objets contours, de nombreux descripteurs ont été proposés. Dans [GAR92], l'auteur utilise une approximation polygonale des contours, alors qu'un ensemble de primitives telles que le périmètre, l'élongation, le nombre de trous, ... a été utilisé dans [EGG92] pour la recherche d'objets par le contenu. Dans [FLI95], les auteurs ont, pour leur part, exploité des combinaisons de primitives géométriques telles que la surface, la circularité, l'excentricité, l'orientation des axes principaux et des moments algébriques. Ces descriptions restent rudimentaires et ne permettent de discriminer des objets qu'en faible nombre.

Dans [MOK96], les auteurs proposent une approche basée sur une représentation multiéchelle des points d'inflexion du contour, ce qui a permis d'extraire des in-

1. Les méthodes de segmentation sous-jacentes à l'extraction des objets ne sont pas abordées dans ce travail (hypothèses MPEG-4 et MPEG-7).

variants locaux par transformations affines. La similarité entre les contours est alors évaluée par des techniques d'appariement et le calcul d'une erreur quadratique.

Une autre solution bien connue consiste à utiliser les coefficients de Fourier. Dans [CRI82], T. Crimmins propose une famille complète de descripteurs globaux invariants par similitude. La propriété de complétude garantit qu'un objet est identifié de manière unique à une transformation près et assure une discrimination totale des formes. Par extension de ces travaux, une nouvelle famille complète et stable a été introduite dans [GHO92]; la propriété de stabilité garantissant de plus qu'une faible variation de la forme des objets n'induit pas de changements notables dans la valeur des descripteurs. Cette dernière famille a été testée en indexation de bases d'images [KAM99] où une vraie distance entre formes est utilisée pour mesurer et trier les objets selon leur similarité.

Il est parfois nécessaire de calculer des descripteurs directement depuis les objets à niveaux de gris lorsque le profil externe des objets n'est pas suffisamment discriminant. Ce problème est plus complexe que le précédent car la description doit prendre en compte tant l'information sur le contour que sur la texture interne des objets.

Dans [SCH97], les auteurs proposent d'utiliser une description locale des objets à niveaux de gris. Des descripteurs invariants par rotation sont calculés à partir de certains points d'intérêts directement extraits de l'image à niveaux de gris. D'autres méthodes utilisent les moments de l'image [FLI95] ou bien encore, plus récemment, la transformée de Fourier-Mellin [MIL99]. Bien qu'invariants par similitudes, les ensembles de descripteurs ne sont pas complets et ne représentent donc qu'une signature partielle de la forme des objets. Dans la suite, nous définissons deux familles complètes de descripteurs invariants extraits depuis les moments complexes de l'image et de la transformée de Fourier-Mellin analytique.

III. DEUX FAMILLES COMPLÈTES DE DESCRIPTEURS DE FORMES

Dans cette section, nous présentons deux familles complètes de descripteurs invariants par similitudes utilisées pour la recherche et la classification des objets à niveaux de gris. La première famille provient de la transformée de Fourier-Mellin analytique et la seconde est issue des moments complexes de l'image.

III.1. Famille issue de la transformée de Fourier-Mellin

La transformée de Fourier-Mellin (TFM) est un outil bien connu en reconnaissance de formes. Celle-ci permet d'extraire des descripteurs de formes invariants par simi-

litudes planes. Les ensembles de descripteurs obtenus ne tiennent généralement compte que de l'amplitude de la TFM, l'information contenue dans la phase étant rejetée [SHE86, GRA91]. De tels invariants ne représentent que partiellement les objets, ce qui apparaît insuffisant et parfois critique pour discriminer des objets en nombre important comme c'est le cas des bases d'images utilisées en indexation. Pour résoudre ce problème, nous présentons un procédé de normalisation de la transformée de Fourier-Mellin analytique (TFMA) permettant d'extraire des familles complètes de descripteurs invariants par similitude.

La TFMA a été introduite pour pallier le problème de divergence des intégrales lorsqu'elles sont calculées sur des images à niveaux de gris [GHO94]. Pour une fonction f décrite en coordonnées polaires par rapport à son centre d'inertie, celle-ci s'écrit :

$$\forall (k, \nu) \in \mathbf{Z} \times \mathbf{R}, M_{f_\sigma}(k, \nu) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^\infty f(r, \theta) r^{-\sigma - i\nu} e^{-ik\theta} d\theta \frac{dr}{r} \quad (1)$$

avec σ un nombre réel strictement positif et $(r, \theta) \in \mathbf{R}_+^* \times \mathbf{S}^1$, \mathbf{S}^1 désignant le cercle unité. La TFMA fournit une représentation unique des objets et il est possible de reconstruire une image depuis son spectre en utilisant la transformée inverse :

$$\forall (r, \theta) \in \mathbf{R}_+^* \times \mathbf{S}^1, f(r, \theta) = \int_{-\infty}^{+\infty} \sum_{k \in \mathbf{Z}} M_{f_\sigma}(k, \nu) r^{-\sigma + i\nu} e^{ik\theta} d\nu \quad (2)$$

Soit maintenant l'objet g obtenu en appliquant une similitude $S_{\alpha, \beta}$, avec $\alpha \in \mathbf{R}_+^*$ le facteur d'échelle et $\beta \in \mathbf{S}^1$ l'angle de rotation, à un objet f quelconque : $g(r, \theta) = f(\alpha r, \theta + \beta)$. En utilisant (1) et par simple changement de variable, nous obtenons l'action d'une similitude sur la TFMA :

$$\forall (k, \nu) \in \mathbf{Z} \times \mathbf{R}, M_{g_\sigma}(k, \nu) = \alpha^{-\sigma + i\nu} e^{ik\beta} M_{f_\sigma}(k, \nu) \quad (3)$$

L'équation (3) est appelée théorème du retard (par analogie au résultat classique sur la transformée de Fourier) et suggère que la TFMA est bien adaptée au calcul de descripteurs invariants à la pose, l'orientation et la taille des objets. Cependant, les descripteurs extraits depuis le module de la TFM ne sont plus valides à cause du terme $\alpha^{-\sigma}$ introduit par le prolongement analytique.

Pour remédier à ce problème, une famille complète et sommable de descripteurs invariants par similitude issue de la TFMA a été proposée [GHO94]. Pour n'importe quelle valeur de σ strictement positive, celle-ci peut s'écrire de la manière suivante :

$$\forall (k, \nu) \in \mathbf{Z} \times \mathbf{R}, I_{f_\sigma}(k, \nu) = M_{f_\sigma}(0, 0) \frac{-\sigma + i\nu}{\sigma} e^{-ik \arg(M_{f_\sigma}(1, 0))} M_{f_\sigma}(k, \nu) \quad (4)$$

Chaque descripteur invariant $I_{f_\sigma}(k, \nu)$ est construit en normalisant la phase et l'amplitude de l'harmonique de Fourier-Mellin correspondante $M_{f_\sigma}(k, \nu)$. La normalisation est réalisée grâce aux deux facteurs suivants :

$$\begin{cases} \Gamma_f = |M_{f_\sigma}(0, 0)|^{\frac{1}{\sigma}} \\ \Theta_f = -\arg(M_{f_\sigma}(1, 0)) \end{cases} \quad (5)$$

L'invariance numérique effective des descripteurs dépend essentiellement de ces deux paramètres de normalisation. D'autres facteurs combinant plusieurs harmoniques ont été récemment introduits pour améliorer la robustesse numérique des familles [DER99].

L'unicité de la famille de descripteurs se vérifie en inversant l'expression de l'équation (4) :

$$\begin{aligned} \forall (k, \nu) \in \mathbf{Z} \times \mathbf{R}, M_{f_{\sigma}}(k, \nu) \\ = M_{f_{\sigma}}(0, 0) \frac{-\sigma + i\nu}{\sigma} e^{-ik \arg(M_{f_{\sigma}}(1, 0))} I_{f_{\sigma}}(k, \nu) \end{aligned} \quad (6)$$

Ainsi, la TFMA d'un objet peut être reconstruite à partir de l'ensemble de ses descripteurs invariants et des deux paramètres de normalisation. On peut alors reconstruire l'image originale en utilisant la TFMA inverse (2). Des expériences en ce sens sont proposées dans la section III.3.

On montre que la famille est également sommable (au sens L^2), ce qui implique l'existence d'une vraie distance sur l'espace des formes [GHO98] :

$$d_2(I_{f_{\sigma}}, I_{g_{\sigma}}) = \left(\int_{-\infty}^{+\infty} \sum_{k \in \mathbf{Z}} |I_{f_{\sigma}}(k, \nu) - I_{g_{\sigma}}(k, \nu)|^2 d\nu \right)^{\frac{1}{2}} \quad (7)$$

Cette distance est nulle pour des objets de même forme et est utilisée par la suite pour mesurer le degré de similarité des objets indépendamment de leur pose, leur orientation et leur taille.

III.2. Famille issue des moments complexes

Depuis la théorie des moments algébriques, H. Ming-Kuel [MK62] a extrait un ensemble de sept descripteurs invariants par similitudes planes (appelés invariants de Hu). Ceux-ci sont construits comme des combinaisons non linéaires des moments centrés et réduits de l'image. Il est clair que cette famille n'est pas complète car plusieurs objets peuvent avoir des descripteurs identiques sans pour autant être de même forme. Pour pallier cet inconvénient, le procédé de normalisation présenté ci-dessus peut être utilisé pour extraire une famille complète de descripteurs invariants partir de la description des images par les moments complexes [MEZ99].

L'expression des moments complexes (MC) d'une image f , décrite en coordonnées polaires, est donnée par [ABU85] :

$$\begin{aligned} \forall (p, q) \in \mathbf{N} \times \mathbf{N}, c_f(p, q) \\ = \int_0^{\infty} \int_0^{2\pi} f(r, \theta) r^{p+q+1} e^{-i(p-q)\theta} d\theta dr \end{aligned} \quad (8)$$

Les moment complexes fournissent une représentation unique des images. De manière analogue au paragraphe précédent, l'action d'une similitude $S_{\alpha, \beta}$ se traduit par la relation suivante :

$$\forall (p, q) \in \mathbf{N} \times \mathbf{N}, c_g(p, q) = \alpha^{-\gamma} e^{-i(p-q)\beta} c_f(p, q) \quad (9)$$

avec $\gamma = p + q + 2$.

En appliquant le procédé de normalisation, on construit une famille complète de descripteurs invariants par similitudes planes de la manière suivante :

$$\begin{aligned} \forall (p, q) \in \mathbf{N} \times \mathbf{N}, C_f(p, q) \\ = |c(0, 0)|^{-\frac{\gamma}{2}} e^{-i(p-q)\arg(c_f(1, 0))} c_f(p, q) \end{aligned} \quad (10)$$

Pour comparer deux objets indépendamment des rotations et facteurs d'échelle qu'ils ont subis, nous utiliserons une erreur quadratique entre les descripteurs invariants des objets :

$$e_2(C_f, C_g) = \left(\sum_{p \in \mathbf{Z}} \sum_{q \in \mathbf{Z}} |C_f(p, q) - C_g(p, q)|^2 \right)^{\frac{1}{2}} \quad (11)$$

III.3. Illustration de la complétude pour la description des formes

Dans le cas de ces deux représentations, la propriété de complétude se traduit par un nombre infini de descripteurs invariants, ce qui apparaît comme contradictoire avec les contraintes numériques. Cependant, l'expression des familles permet d'extraire un nombre de descripteurs de forme aussi important que souhaité. En élargissant l'ensemble, nous approchons la complétude et la description numérique des images devient de plus en plus précise.

Pour illustrer ceci, la figure 1 présente les résultats de la reconstruction de quatre images de papillons à partir

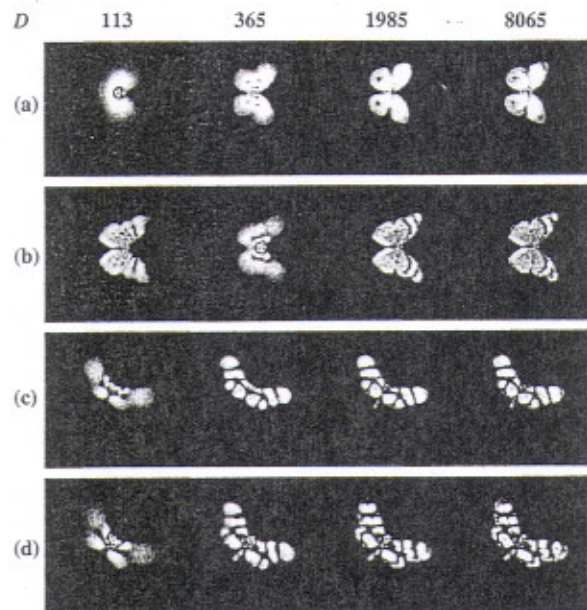


FIG. 1. — Reconstruction des formes de quatre images de papillons en fonction du nombre de descripteurs invariants utilisés pour la reconstruction (TFMA).

Shape reconstruction of four butterfly images according to the number of invariant descriptors used in the reconstruction process (AFMT).

d'ensembles de descripteurs invariants issus de la TFMA de différentes tailles. Les méthodes d'approximation utilisées pour estimer la TFMA et son inverse sont décrites dans [DER99]. Nous avons utilisé ici l'approximation log-polaire qui présente l'avantage d'être rapide et particulièrement robuste.

Lorsque le nombre de descripteurs est faible ($D=7$, $D=13$), la représentation des images est grossière et ne capture qu'une information rudimentaire sur la forme des objets. Lorsque les images présentent des différences visuelles notables, comme c'est le cas des images (a), (b) et (c), un nombre de descripteurs invariants relativement faible est suffisant. Par contre, il n'est possible de distinguer les papillons (c) et (d) que lorsque le nombre de descripteurs invariants est relativement important ($D=31$, $D=63$). Le nombre des descripteurs invariants utilisés pour caractériser une base d'images doit donc être adapté en fonction de sa taille et de son homogénéité.

Il convient de noter que les moments complexes, bien que complets et inversibles, n'ont pas permis de reconstruire une image à niveaux de gris avec une qualité satisfaisante. Ceci est dû aux problèmes d'approximation numérique des moments géométriques largement discutés dans la littérature [TEH88, LIA96].

IV. RÉSULTATS EXPÉRIMENTAUX DE RECHERCHE PAR LE CONTENU

Dans cette section, nous présentons une comparaison succincte des invariants de Hu et des deux familles complètes de descripteurs de formes pour l'indexation et la recherche par le contenu à partir de deux bases d'objets réels.

Nous avons organisé les résultats de recherche sous la forme d'un tableau : les colonnes représentent les trois ensembles de descripteurs invariants testés (invariants de Hu, des moments complexes et de la TFMA), et les lignes correspondent au nombre de descripteurs invariants D utilisés pour estimer la similarité avec l'objet requête. Chaque résultat de recherche se présente sous la forme d'une image rassemblant les neuf objets les plus similaires à l'image requête, classés par ordre de similarité de gauche à droite et de haut en bas. Dans les exemples que nous proposons, l'image requête appartient à la base de modèles. La requête est donc toujours la première image retrouvée avec une mesure de similarité exactement nulle.

IV.1. Principe du système de recherche

Pour tester les familles de descripteurs de formes dans le cadre d'une application d'indexation, nous avons développé une plate-forme de test de type client-serveur.

Une interface graphique accessible depuis l'Internet permet à un utilisateur d'interroger la base de modèles à partir d'une image requête.

Le principe de l'algorithme de recherche est relativement simple. La phase d'indexation consiste à calculer et à stocker un grand nombre de descripteurs invariants pour tous les modèles de la base de référence. Pendant la phase de recherche, les distances et sont utilisées pour mesurer la similarité entre une requête et les objets de la base, indépendamment de leur position, de leur orientation et de leur taille. Ainsi, la recherche d'un objet inconnu g parmi un ensemble de N objets f_i constituant la base de modèles est réalisée à travers ces trois étapes :

1. on extrait les descripteurs invariants de g ,
2. on estime la distance entre les invariants de g et les invariants de tous les modèles f_p
3. on sélectionne et on trie les M modèles les plus similaires à la requête, c'est à dire ceux qui présentent la distance la plus faible avec g .

La recherche dans la base de modèles est exhaustive. La mise en place de structures arborescentes permettant d'accélérer la sélection des modèles les plus similaires à la requête reste une perspective importante pour le développement de cette plate-forme de test, comme nous le signalerons en conclusion. L'avantage d'indexer directement les images par leurs descripteurs invariants est d'éviter une phase préalable d'apprentissage sur la base de modèles. Ainsi, on peut ajouter ou supprimer des modèles sans avoir à reconstruire tous les index des images préalablement stockées.

IV.2. Base d'images de papillons

La première base d'objets est constituée de $N=98$ images représentant 70 papillons de formes différentes. Les papillons montrent une forte homogénéité et leur principale différence se situe dans la texture des ailes. Des descripteurs basés uniquement sur le contour extérieur des papillons seraient sans doute insuffisants pour les discriminer. Cette base est donc particulièrement bien adaptée pour tester l'invariance et la discrimination des invariants (Fig. 2).

La figure 2 présente les résultats de recherche obtenus avec l'image Pap15_1 pour les trois ensembles de descripteurs invariants. Ce papillon est représenté par quatre images similaires dans la base de modèles. Dans le cas des invariants de Hu, les quatre images sont bien présentes mais ne sont pas exactement aux quatre premières positions. Certains papillons de formes légèrement différentes sont mieux classés, ce qui peut être directement imputé aux problèmes bien connus d'approximation des moments géométriques, ainsi qu'à la manière dont les invariants sont construits. Par contre, les descripteurs de formes issus des MC permettent de retrouver toutes les images similaires de la base, quel que soit

2. Base COIL (Columbia Object Image Library), <http://www.cs.columbia.edu/CAVE/>.

D	Invariants de Hu (10^{-13})	Invariants des MC (10^{-6})	Invariants de la TFMA
7	 pap15_1 (0,00)	 pap15_1 (0,00)	 pap15_1 (0,00)
	 pap57_1 (0,87)	 pap15_4 (1,06)	 pap15_2 (0,03)
	 pap15_2 (7,53)	 pap15_3 (4,85)	 pap15_3 (0,04)
 pap15_3 (8,90)	 pap28_4 (1,03)	 pap1_2 (0,11)	
 pap12_2 (11,9)	 pap14_6 (10,9)	 pap58_1 (0,11)	
 pap15_4 (12,0)	 pap1_2 (12,3)	 pap59_1 (0,14)	
 pap12_2 (13,2)	 pap57_1 (15,5)	 pap28_4 (0,17)	
 pap12_1 (18,9)	 pap58_1 (16,7)	 pap28_5 (0,18)	
 pap1_2 (32,2)		 pap28_1 (0,18)	
40		 pap15_1 (0,00)	 pap15_1 (0,00)
		 pap15_4 (1,51)	 pap15_2 (0,19)
		 pap15_3 (6,89)	 pap15_3 (0,24)
	 pap15_2 (8,13)	 pap28_4 (14,6)	 pap15_4 (0,42)
	 pap14_6 (16,1)	 pap1_2 (0,64)	 pap58_1 (0,64)
	 pap1_2 (18,8)	 pap59_1 (0,65)	 pap56_1 (0,76)
	 pap57_1 (23,3)	 pap50_1 (0,77)	 pap50_1 (0,77)
	 pap28_3 (24,1)		
180		 Pap15_1 (0,00)	 pap15_1 (0,00)
		 Pap15_4 (1,51)	 pap15_2 (0,34)
		 Pap15_3 (6,89)	 pap15_3 (0,45)
	 Pap15_2 (8,13)	 pap15_4 (0,38)	 pap59_1 (0,96)
	 Pap28_4 (14,6)	 pap59_1 (0,96)	 pap1_2 (0,98)
	 Pap14_6 (16,1)	 pap50_1 (1,02)	 pap1_2 (0,98)
	 Pap1_2 (18,8)	 pap56_1 (1,07)	 pap58_1 (1,07)
	 Pap57_1 (23,3)		
	 Pap28_3 (24,1)		

FIG. 2. — Résultats de recherche obtenus à partir de l'objet pap15_1 avec environ 7, 40 et 180 invariants pour les trois familles de descripteurs de formes.

Retrieval results obtained from object pap15_1 with about 7, 40 and 180 invariants for the three sets of shape descriptors.

le nombre d'invariants utilisés. Enfin, pour les invariants issus de la TFMA, nous retrouvons trois images pour $D=7$ et les quatre images pour des ensembles de descripteurs plus importants ($D=40$ et $D=180$). Ainsi un nombre insuffisant de descripteurs invariants peut être préjudiciable à la robustesse du système de recherche.

IV.3. Base de d'images de l'université de Columbia

La seconde base utilisée pour nos expériences provient de l'université de Columbia². Elle est constituée de $N=1440$ images de 20 objets 3D prient dans 72 positions différentes (par rotation de 5° autour d'un axe vertical). La principale difficulté de cette base réside dans le fait que le mouvement apparent des objets fausse sensiblement le modèle de transformation géométrique que nous considérons ici, laissant apparaître des occlusions et des apparitions dans les images. Cette base est donc bien adaptée pour tester la stabilité et la robustesse des descripteurs aux faibles déformations des formes.

Les figures 3 et 4 illustrent les résultats de recherche obtenus pour les images obj5_64 et obj8_1. Pour ces deux exemples, les invariants de Hu ne permettent de retrouver aucun objet similaire à l'image requête, ce qui confirme leur instabilité aux faibles variations de formes. Les résultats obtenus par les invariants des MC montrent entre 2 et 4 images similaires dans le premier exemple, et entre 4 et 5 dans le second exemple. Les descripteurs issus de la TFMA permettent, quant à eux, de retrouver entre 7 et 9 images similaires dans le premier exemple, et entre 7 et 8 dans le second exemple.

Ces deux exemples de recherche sont parmi les plus défavorables que nous ayons rencontrés lors de nos expériences. Ceux-ci illustrent l'importance de la propriété de complétude et permettent de comparer la robustesse relative de chacun des ensembles de descripteurs sur des exemples difficiles pour notre système. Des expériences complémentaires, non reproduites ici, ont montré que les résultats de recherche sont perceptiblement satisfaisants et que les mesures de similarité retranscrivent une certaine proximité visuelle.

V. CONCLUSIONS ET PERSPECTIVES

Dans cet article, nous avons illustré l'importance de la propriété de complétude pour le choix de descripteurs en vue d'application en indexation et recherche d'images par le contenu. Deux familles complètes de descripteurs invariants par rapport aux similitudes planes ont été testées et comparées à l'ensemble non complet des descripteurs invariants de Hu. Celles-ci donnent des résultats de recherche sensiblement supérieurs aux descripteurs de

Hu. Toutes deux présentent de très bonnes qualités d'invariance numérique aux similitudes et les mesures de similarité induites par les représentations invariantes traduisent correctement la proximité visuelle entre les formes.

L'hypothèse principale de nos deux méthodes est basée sur un modèle de mouvement rigide de type similitude plane. Néanmoins, ces descripteurs, et notamment ceux obtenus par la transformée de Fourier-Mellin analytique, s'avèrent robustes pour des mouvements laissant apparaître de légères apparitions / occlusions (base de l'université de Columbia). Les expériences ont également montré que l'accroissement du nombre d'invariants peut améliorer le taux de recherche, notamment lorsque la base est fortement homogène. La propriété de complétude des descripteurs apparaît comme une propriété intéressante qui offre une représentation riche et flexible des formes. Pour toutes ces raisons, nous pensons que ces familles s'inscrivent bien dans le cadre de la future norme MPEG-7.

Différentes perspectives sont envisageables pour compléter et enrichir le système de recherche proposé. L'indexation directe d'un objet à partir de ses invariants permet de compléter une base de modèles sans qu'il y ait lieu de modifier les descripteurs déjà stockés. Cependant, cette technique impose une confrontation exhaustive de l'objet requête avec tous les éléments de la base, ce qui peut engendrer des temps de recherche relativement importants. Une des voies les plus intéressantes à explorer concerne donc la phase d'indexation des descripteurs et l'accélération de la recherche. Pour éviter de scruter l'ensemble des modèles, une solution consiste à construire une structure arborescente par quantification vectorielle des descripteurs invariants. L'identification précise de l'espace des formes nous permet d'envisager un apprentissage adapté tenant compte du support des descripteurs invariants. Dans ce cas, il serait également intéressant de tester la robustesse descripteurs invariants aux erreurs de quantification, et notamment son influence sur la reconstruction des images.

Manuscrit reçu le 2 décembre 1999

accepté le 24 février 2000

BIBLIOGRAPHIE

- [ABU85] ABU-MOSTAFA (Y.), PSALTIS (D.), Image normalisation by complex moments, *IEEE trans. on Pattern Analysis and Machine Intelligence*, (1985), 7, n° 1, pp. 46-55.
- [AIG96] AIGRAIN (P.) et al, Content-based representation and retrieval of visual media a state-of-the-art review, *Multimedia Tools and Applications*, (1996), 3, n° 3, pp 179-202.
- [BAC96] BACH (J.R.) et al, The Virage image search engine : An open framework for image management, dans *Symposium on Electronic Imaging: Science and Technology - Storage & Retrieval for still image and video databases IV*, (février 1996), 2670, pp. 76-87.
- [CRI82] CRIMMINS (T.R.), A complete set of Fourier descriptors for two-dimensional shapes, *IEEE trans. on Systems, Man, and Cybernetics*, (1982), 12, n° 6, pp. 848-855.








D	Invariants de Hu (10^{-12})	Invariants des MC (10^{-6})	Invariants de la TFMA
7	 <p>Obj5_64 (0,00) Obj9_22 (1,43) Obj9_6 (1,35)</p> <p>Obj9_3 (2,29) Obj10_4 (2,29) Obj10_43 (2,29)</p> <p>Obj10_41 (2,29) Obj10_42 (2,29) Obj10_43 (2,29)</p>	 <p>Obj5_64 (0,00) Obj5_27 (1,66) Obj9_27 (1,20)</p> <p>Obj9_63 (2,21) Obj9_22 (2,37) Obj9_21 (2,40)</p> <p>Obj9_28 (2,64) Obj6_21 (2,82) Obj9_27 (2,84)</p>	 <p>obj5_64 (0,00) obj5_4 (0,29) obj5_63 (0,33)</p> <p>obj5_3 (0,34) obj9_61 (0,33) obj9_62 (0,36)</p> <p>obj5_62 (0,37) obj5_70 (0,37) obj5_61 (0,38)</p>
40		 <p>Obj5_64 (0,00) Obj5_27 (3,06) Obj9_63 (4,49)</p> <p>Obj5_65 (5,09) Obj19_9 (5,29) Obj9_64 (5,51)</p> <p>Obj5_28 (5,70) Obj6_21 (5,71) Obj9_27 (5,80)</p>	 <p>obj5_64 (0,00) obj5_63 (0,72) obj5_3 (0,77)</p> <p>obj5_61 (0,78) obj5_34 (0,92) obj5_35 (0,93)</p> <p>obj5_33 (0,95) obj5_36 (0,95) obj5_60 (0,95)</p>
180		 <p>Obj5_64 (0,00) Obj5_27 (3,06) Obj9_63 (4,49)</p> <p>Obj5_65 (5,09) Obj19_9 (5,29) Obj9_64 (5,51)</p> <p>Obj5_28 (5,70) Obj6_21 (5,71) Obj9_27 (5,80)</p>	 <p>obj5_64 (0,00) obj5_63 (0,98) obj5_3 (1,00)</p> <p>obj5_61 (1,02) obj5_60 (1,13) obj19_28 (1,16)</p> <p>obj19_29 (1,17) obj5_34 (1,18) obj5_27 (1,19)</p>

FIG. 3. — Résultats de recherche obtenus à partir de l'objet obj5_64 avec environ 7, 40 et 180 invariants pour les trois familles de descripteurs de formes.

Retrieval results obtained from object obj5_64 with about 7, 40 and 180 invariants for the three sets of shape descriptors.















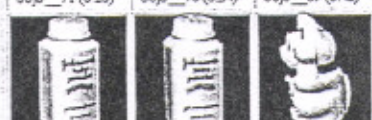



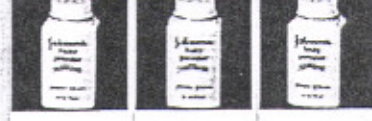


D	Invariants de Hu (10^{-12})	Invariants des MC (10^{-6})	Invariants de la TFMA
7	 Obj16_1 (0,00) Obj16_3 (7,53) Obj16_30 (7,53)  Obj16_31 (7,53) Obj16_32 (7,53) Obj16_33 (7,53)  Obj16_34 (7,53) Obj16_35 (7,53) Obj16_36 (7,53)	 Obj8_1 (0,00) Obj8_2 (0,27) Obj9_38 (0,67)  Obj9_34 (0,80) Obj8_70 (0,80) Obj4_38 (1,40)  Obj8_3 (1,50) Obj5_35 (1,55) Obj8_69 (1,57)	 obj8_1 (0,00) obj8_0 (0,04) obj8_2 (0,04)  obj8_71 (0,07) obj8_70 (0,07) obj2_69 (0,10)  obj2_70 (0,10) obj8_3 (0,11) obj2_68 (0,13)
40		 Obj8_1 (0,00) Obj8_2 (0,60) Obj9_34 (2,31)  Obj8_3 (2,60) Obj8_0 (2,70) Obj4_31 (2,16)  Obj9_34 (3,27) Obj4_63 (3,37) Obj4_39 (3,45)	 obj8_1 (0,00) obj8_0 (0,13) obj8_2 (0,23)  obj8_71 (0,28) obj8_70 (0,34) obj8_69 (0,42)  obj8_66 (0,52) obj8_68 (0,55) obj4_16 (0,57)
180		 Obj8_1 (0,00) Obj8_2 (0,60) Obj9_34 (2,31)  Obj8_3 (2,60) Obj8_0 (2,60) Obj4_31 (2,16)  Obj9_34 (3,27) Obj4_63 (3,37) Obj4_39 (3,45)	 obj8_1 (0,00) obj8_0 (0,23) obj8_2 (0,33)  obj8_71 (0,43) obj8_70 (0,40) obj8_69 (0,40)  obj8_66 (0,71) obj7_42 (0,75) obj8_68 (0,76)

FIG. 4. — Résultats de recherche obtenus à partir de l'objet obj8_1 avec environ 7, 40 et 180 invariants pour les trois familles de descripteurs de formes.

Retrieval results obtained from object obj8_1 with about 7, 40 and 180 invariants for the three sets of shape descriptors.

- [DER99] DERRODE (S.), Représentation de formes planes à niveaux de gris par différentes approximations de Fourier-Mellin analytique en vue d'indexation de bases d'images, *Thèse de l'université de Rennes*, décembre 1999.
- [EGG92] EGGLESTON (P.), Content-based feature indexing and retrieval for image databases, in *SPIE*, (1992), 1819.
- [FLI95] FLICKNER (M.) et al, Query by image and video content : the QBIC system, *IEEE trans. on Computers*, (1995), 28, n° 9, pp. 23-32.
- [GAR92] GARY (E.), MEHROTRA (R.), Shape similarity-based retrieval in image databases, in *SPIE*, (1992), 1662.
- [GHO92] GHORBEL (F.), Stability of invariant Fourier descriptors and its inference in the shape classification, dans *Proc. Of the IEEE Int. Conf. on Pattern Recognition*, 20 août 1992, The Hague (NL), pp. 130-134.
- [GHO94] GHORBEL (F.), A complete invariant description for gray-level images by the harmonic analysis approach, *Pattern Recognition Letters*, (1994), 15, pp. 1043-1051.
- [GHO98] GHORBEL (F.), Toward a unitary formulation for invariant pattern recognition; application to image coding, *Ann. Telecommunic.*, (1998), 53, n° 5-6, pp. 242-260.
- [GRA91] GRACE (A.E.) et SPANN (M.), A comparison between Fourier-Mellin descriptors and moment based features for invariant object recognition using neural networks, *Pattern Recognition Letters*, 1991, Vol. 12, pp. 635-643.
- [KAM99] KAMOUN (M.), CHAKCHOUK (M.), BOULILA TRIKI (N.) et GHORBEL (F.), Indexation des bases d'images par le contenu en faisant coopérer deux types de descripteurs de formes, *COmpression et REprésentation des Signaux Audiovisuels (CORESA'99)*, 14-15 juin 1999, Sophia-Antipolis (Fr.).
- [LIA96] LIAO (S.X.) et PAWLAK (M.), On image analysis by moments, *IEEE trans. on Pattern Analysis and Machine Intelligence*, 1996, Vol. 18(3), pp. 254-266.
- [MEZ99] MEZHOUD (R.), BANNOUR (T.), DERRODE (S.) et GHORBEL (F.), A complete and invariant description of gray level images from geometric moments, dans *1^{er} atelier de Traitement et d'Analyse d'Images : Méthodes et Applications (TAIMA'99)*, Hammamet (Tn), 1999.
- [MIL99] MILANESE (R.) et CHERBULIEZ (M.), A rotation-, translation-, and scale-invariant approach to content-based image retrieval, *Journal of Visual Communication and Image Representation*, 1999, Vol. 10, pp 186-196.
- [MIN98] MINKA (T.) et PICARD (R.), Interactive learning using a society of models, *Pattern Recognition*, 1997, Vol. 30(4), pp. 565-681.
- [MK62] MING-KUOEL (H.), Visual pattern recognition by moment invariants, *IRE trans. on Information Theory*, 1962, Vol. 8, pp. 179-187.
- [MOK96] MOKHTARIAN (F.), ABBASI (F.) et KITTLER (J.), Robust and efficient shape indexing through curvature scale space, dans *Proc. of the 1996 British Machine and Vision Conference BMVC'96*, septembre 1996, Edinburgh, Scotland.
- [PEN96] PENTLAND (A.), PICARD (R.) et SCLAROFF (S.), Photo-book: Tools for content-based manipulation of image databases, *International Journal of Computer Vision*, 1996, Vol. 18(3), pp. 233-254.
- [RUI98] RUI (Y.), HUANG (T.S.), ORTEGA (M.) et MEHROTRA (S.), Relevance feedback: a power tool in interactive content-based image retrieval, *IEEE trans. on Circuits and Systems for Video Technology, special issue on Interactive Multimedia Systems for the Internet*, 1998, Vol. 8(5), pp. 644-655.
- [RUI99] RUI (Y.), HUANG (T.S.) et CHANG (S.), Image retrieval: current techniques, promising directions, and open issues, *Journal of Visual Communication and Image Representation*, 1999, Vol. 10, pp. 39-62.
- [SCH97] SCHMID (C.) et MOHR (R.), Local grey-value invariants for image retrieval, *IEEE trans. on Pattern Analysis and Machine Intelligence*, 1997, Vol. 19(5), pp. 530-535.
- [SHE86] SHENG (Y.) et DUVERNOY (J.), Circular-Fourier radial-Mellin transform descriptors for pattern recognition, *Journal of the Optical Society of America A*, 1986, Vol. 3(6), pp. 885-888.
- [TEH88] TEH (C.H.) et CHIN (R.T.), On image analysis by the method of moments, *IEEE trans. Pattern Analysis Machine Intelligence*, 1988, Vol. 10(4), pp. 496-513.