

Reconnaissance de formes par invariants complets et convergents ; Application à l'indexation de bases d'objets à niveaux de gris

Stéphane DERRODE¹, Rim MEZHOU² et Faouzi GHORBEL²

¹Département Image et Traitement de l'Information (Enst-Bretagne),
BP 835, 29285 Brest cedex, France.

²Groupe de Recherche Images et Formes de Tunisie (ENSI)
& Direction des Etudes et Recherches (CERT), 42 Rue Asdrubal, 1002 Tunis, Tunisie.

Stephane.Derrode@enst-bretagne.fr, Rim.Mezhoud@ensi.rnu.tn,
Faouzi.Ghorbel@cert.mincom.tn.

Résumé – La recherche d'images par le contenu est un domaine de recherche émergent et particulièrement dynamique du fait de l'essor de grandes bases d'images. Elle requiert un ensemble de descripteurs, une méthode robuste pour les extraire et, enfin, une mesure de similarité rapide qui traduit la ressemblance visuelle entre les formes. Dans ce travail, nous présentons et comparons deux familles complètes et convergentes de descripteurs globaux pour la recherche et l'archivage d'objets à niveaux de gris. Notre approche est invariante aux similitudes planes. La robustesse et la performance des descripteurs sont illustrées par plusieurs expériences numériques sur deux bases.

Abstract – Content-based image retrieval is an important and emerging research area since the development of large databases. The retrieval requires the choice of a suitable set of features, a robust method to extract them and a similarity measure which corresponds to the visual similarity between shapes. This paper presents two complete and convergent sets of global features for grey-level object retrieval and storage. Our approach is invariant to planar similarities. The robustness and the discrimination performance of the features are illustrated by experiments on two databases.

1. Introduction

L'essor des systèmes d'acquisition (caméra, scanner, IRM), de transmission et de diffusion (réseau Internet, réseaux spécialisés) et de stockage (mémoire de masse) a favorisé la création et le développement de grandes bases d'images. Ceci pose le problème de la représentation et de l'accès aux données. L'utilisation d'attributs textuels pour indexer les images a montré ses limites et l'accès à l'information par le contenu est devenu un sujet de recherche particulièrement dynamique. Ainsi de nombreuses techniques, essentiellement issues de recherches effectuées en reconnaissance de formes, ont été développées et plusieurs états de l'art ont été présentés ([1], [2]).

Un des problèmes centraux 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 (histogramme), les descripteurs de texture (matrice de cooccurrence, indices de direction principale et de rugosité, filtres de Gabor et ondelettes) et les descripteurs de formes (descripteurs de Fourier et des moments, points caractéristiques). L'extraction de primitives pertinentes est un problème qui ne connaît pas de solution dans le cas générale car il semble qu'il n'existe pas d'attributs qui puissent modéliser une base selon tous les

points de vues (pour des raisons multiples qui peuvent être illustrés par la subjectivité de la requête de l'utilisateur). Les systèmes actuels tentent de combiner des primitives variées pour améliorer la discrimination et la classification (QBIC [3], Virage [4]). Certains font intervenir l'utilisateur par le moyen de systèmes interactifs (Photobook [5], Surfimage) pour affiner la recherche.

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'archivage, l'indexation et la recherche de bases d'objets à niveaux de gris. Dans la section 2, nous présentons un bref rappel sur les descripteurs globaux d'objets à niveaux de gris ainsi que leur utilisation dans les systèmes d'indexation et de recherche par le contenu. Nous y présentons notamment les propriétés de complétude et de convergence des descripteurs qui nous semble souhaitable pour de tels systèmes et apparaissent utiles dans l'esprit de la future norme MPEG-7. Dans la section 3, nous présentons *deux familles complètes et convergentes de descripteurs invariants par similitudes* qui ont été récemment proposées. La première famille provient de la transformée de Fourier-Mellin [6] et la seconde est issue des moments géométriques de l'image [7]. Dans la section 4, nous les comparons pour leur application en archivage et recherche d'images par le contenu avec les 7 invariants de Hu [8] qui ne possèdent pas les propriétés énoncées. Les résultats sont illustrés par des expériences menées sur deux bases

TAB. 1 : Synthèse des propriétés des descripteurs utilisés.

	Invariants complets issus de la TFM	Invariants complets issus des moments
Représentation des objets à niveaux de gris ❶	TFM analytique ($k \in \mathbb{Z}$, $v \in \mathbb{R}$ et $s > 0$) $M_f(k, v) = \frac{1}{2p} \cdot \int_0^\infty \int_0^{2p} f(r, q) r^{s-iv} e^{-ikq} \frac{dr}{r} dq$	Polynômes des moments ($p \in \mathbb{N}$ et $0 \leq l \leq p$) $P_f(p-l, l) = \frac{1}{l! C_p^l} \sum_{k=0}^{p-l} Q_{p-l, k}^l(f) \text{ avec}$ $\begin{cases} Q_{p, k}^0(f) = (-i)^k C_p^k m(p-k, k) \\ Q_{p-l, k}^l(f) = (p-k-l+1) Q_{p-l+1, k}^{l-1}(f) - (k+1) Q_{p-l+1, k+1}^{l-1}(f) \end{cases}$
Théorème du retard / similitudes ❷	$M_g(k, v) = a^{-s+iv} e^{ikb} M_f(k, v)$	$P_g(p-l, l) = a^{p+1} e^{i(p-2l)q} P_f(p-l, l)$
Famille complète d'invariants ❸	$I_f(k, v) = M_f(0, 0)^{-s+iv} e^{ik \text{Arg}(M_f(1, 0))} M_f(k, v)$	$J_f(p-l, l) = P_f(2, 1) ^{p/3} e^{i(p-2l) \text{Arg}(P_f(0, 1))} P_f(p-l, l)$
Mesure de similarité ❹	$d_2(I_f, I_g) = \left(\int_{-\infty}^{+\infty} \sum_{k \in \mathbb{Z}} I_f(k, v) - I_g(k, v) ^2 dv \right)^{\frac{1}{2}}$	$e_2(J_f, J_g) = \left(\sum_{p \in \mathbb{Z}} \sum_{l=0}^p J_f(p-l, l) - J_g(p-l, l) ^2 \right)^{\frac{1}{2}}$

d'objets réels (base originale de papillons et base de Columbia). Enfin, nous concluons sur quelques perspectives permettant d'enrichir le système expérimental proposé.

2. Descripteurs globaux d'objets

La plupart des descripteurs de formes ont été introduits dans le cadre d'applications de reconnaissance de formes. Le problème fondamental est de déterminer si deux objets sont similaires, indépendamment de leur pose dans l'image. Nous considérons ici l'invariance aux *similitudes planes* (translation suivi d'une rotation et d'une homothétie). Il est bien connu que ces descripteurs invariants doivent satisfaire certains critères parmi lesquels : la *robustesse* vis à vis des faibles variations de formes et des approximations numériques, et la *simplicité* pour le calcul en 'temps réel'. Nous pensons que deux propriétés supplémentaires sont également souhaitables :

- la *complétude* garantie qu'un objet est identifié de manière unique par ses descripteurs, à une transformation près. C'est une propriété importante pour la discrimination totale des formes et pour la reconstruction d'un objet depuis l'ensemble de ses descripteurs invariants (archivage).
- la *convergence* des descripteurs assure une certaine concentration de l'énergie à l'origine. Du point de vue numérique, un nombre limité de descripteurs convergents capture l'essentiel de la forme des objets (reconstruction numérique).

Les descripteurs extraits à partir des coefficients de Fourier et des moments sont largement utilisés dans les systèmes actuels d'indexation et de recherche. Pour un objet requête, les modèles de référence sont triés en fonction d'une mesure de similarité (erreur quadratique). Des expériences récentes ont montré que la combinaison de ces deux représentations améliorent les résultats obtenus depuis une seule d'entre elles [9]. L'utilisation de la transformée de Fourier-Mellin n'a été étudiée que très récemment dans le cadre d'applications de

recherche et d'indexation [10]. Les descripteurs ne sont ni complets ni convergents. Nous présentons deux familles de descripteurs invariants d'objets à niveaux de gris possédant ces deux propriétés importantes.

3. Reconnaissance de formes

Les deux familles que nous proposons sont présentées dans le tableau 1. Deux objets f et g ont *même forme* si et seulement s'il existe une similitude de paramètres $a > 0$ et $b \hat{\mathbb{I}} [0; 2p[$ qui permet de transformer l'un en l'autre.

La première famille est issue du prolongement analytique de la *transformée de Fourier-Mellin* (TFM) qui a été introduit pour remédier au problème numérique de divergence de l'intégrale [6]. La seconde famille provient de la théorie des moments algébriques. Celle-ci nous a permis de construire les *polynômes des moments* (PM) qui s'expriment comme une combinaison linéaire des moments centrés (μ) de l'image [7] (❶).

Ces deux représentations sont *inversibles* et strictement identiques à la représentation originale des images à niveaux de gris. Actuellement, seule la transformée de Fourier-Mellin nous a permis de reconstruire un objet à niveaux de gris avec une bonne qualité [11]. Le *théorème du retard* (❷) suggère que la TFM et les PM sont bien adaptés pour le calcul de descripteurs globaux invariants à la pose, l'orientation et la taille des objets. Ainsi, deux familles I et J (❸) ont été construites par un procédé de normalisation qui supprime l'information purement géométrique tout en conservant l'intégrité sur la forme des objets (complétude). Deux objets sont donc de même forme si et seulement si l'ensemble de leurs descripteurs invariants sont identiques. Nous pouvons alors affirmer que les *mesures de similarité* (❹) seront nulles pour des objets de même forme. Celle-ci nous servira, dans le paragraphe suivant, pour discriminer et classer les objets selon leur similarité.

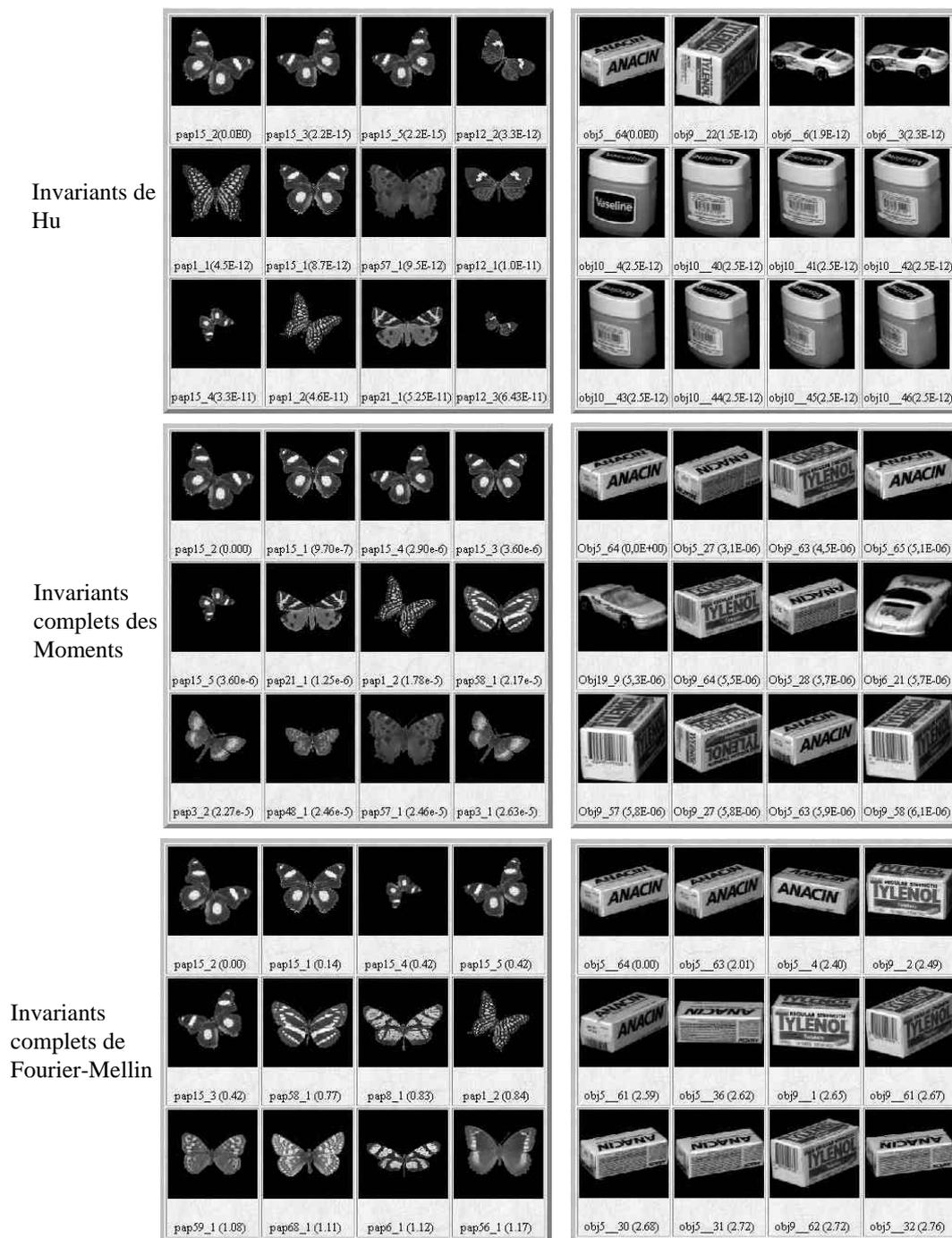


Fig. 1 : Un exemple de recherche pour la famille de Hu, des polynômes des moments et de la transformée de Fourier-Mellin, pour les deux bases d'objets.

4. Résultats expérimentaux de recherche

Le principe de la recherche d'une image inconnue dans une base d'objets de référence est le suivant :

- *Phase de pré-traitement (off-line)* : on calcule et on stocke les descripteurs invariants des objets de la base de modèles.
- *Phase de confrontation (on-line)* : on calcule les invariants de l'image requête. Puis, on estime la mesure de similarité entre les invariants de l'objet inconnu et ceux des différents modèles.
- *Phase de sélection (on-line)* : on trie les modèles par ordre croissant des mesures de similarité obtenues et on

sélectionne ceux qui offrent la plus petite distance avec l'objet requête.

Les représentations complètes des objets contiennent une infinité de descripteurs invariants, ce qui peut apparaître comme contradictoire avec les contraintes de traitements numériques. Néanmoins, les deux familles permettent d'extraire automatiquement autant de descripteurs que souhaité. En élargissant l'ensemble d'invariants, on capture progressivement les détails de l'image, assurant ainsi une description et une discrimination de plus en plus fine des formes. En s'appuyant sur la propriété de convergence des descripteurs, une stratégie systématique d'extraction consiste

à sélectionner les descripteurs centraux et à élargir l'ensemble ainsi formé en augmentant l'ordre. Le nombre d'invariants peut être ajusté en fonction de l'homogénéité de la base de référence. De plus, des expériences récentes sur la TFM ont montré qu'un nombre suffisamment large (mais limité) de descripteurs s'avère suffisant pour reconstruire un objet à niveaux de gris avec une très bonne qualité [11].

La figure 1 illustre les résultats de l'algorithme de recherche pour les invariants de Hu [8], de la TFM et des polynômes des moments. Nous avons testé deux bases de modèles (base de papillon et base de Columbia). Les 12 premières images les plus ressemblantes à l'image requête (coin supérieur gauche) sont affichées.

4.1 Base de papillons

La première base est constituée de 90 images de papillons et présente une forte homogénéité des formes. Les deux familles complètes identifient en priorité tous les objets de même forme (5 papillons). Ce n'est pas le cas des invariants de Hu qui classent les papillons 15_1 en 6^{ème} position et 15_4 en 9^{ème} position. Les premières images de papillons différents ne sont pas identiques et la notion de proximité des formes est différente pour les 2 méthodes. Dans le cas des polynômes des moments, des expériences non reproduites ici, ont montré que l'on atteint 100% de réussite et ceci quel que soit le nombre de descripteurs utilisés. Les descripteurs de faible ordre sont donc prépondérants dans la classification des formes. Les invariants issus des PM sont également plus discriminants que ceux de la TFM. Dans le cas de la TFM, un nombre trop faible de descripteurs peut engendrer des confusions de formes.

4.2 Base de Columbia

La seconde base contient 1440 images représentant 20 objets dans des conditions de prise de vue différentes. La principale difficulté de cette dernière base réside dans le fait que le mouvement apparent des objets ne correspond pas au modèle de transformation géométrique que nous considérons (similitude plane). Nous avons choisi un exemple défavorable pour les 3 méthodes. D'autres requêtes montrent des résultats beaucoup plus satisfaisants. Dans le cas des invariants de Hu, aucun modèle convenable n'est retrouvé. Dès la 3^{ème} image, les images présentent d'importantes différences visuelles avec l'objet requête. Par contre, dans le cas des invariants complets, les résultats de recherche sont plus cohérents et perceptiblement satisfaisants. Les mesures de similarité retranscrivent une certaine ressemblance visuelles (boîtes similaires). Des expériences plus approfondies ont montrées que, d'une manière générale, les invariants issus de la TFM permettent de retrouver un plus grand nombre d'images similaires à l'objet requête.

5. Conclusions

Dans ce papier, nous avons illustré l'importance des propriétés de complétude et de convergence pour le choix de descripteurs en vue d'application d'archivage et de recherche d'images par le contenu. Deux familles complètes et

convergentes de descripteurs invariants par rapport aux similitudes planes ont été testées et comparées par rapport à la famille d'invariants de Hu. Celles-ci donnent des résultats de recherche sensiblement supérieurs aux descripteurs de Hu. Toutes deux présentent également de très bonnes qualités d'invariance numérique aux similitudes. Les mesures de similarité employées traduisent bien la proximité visuelle entre les formes.

L'hypothèse principale de nos 2 méthodes est basée sur un modèle de mouvement rigide de type similitude plane. Néanmoins, ces descripteurs, et notamment ceux obtenus par la TFM, s'avèrent très satisfaisants pour des mouvements laissant apparaître de légères apparitions / occlusions (base de Columbia). La complétude des descripteurs est une propriété sémantique intéressante qui offre une représentation riche et flexible des formes.

Nous comptons développer ce système expérimental en intégrant des descripteurs locaux, mieux à même de décrire les déformations locales de texture (visages) ainsi que les occlusions. Une autre perspective consiste à tester la robustesse des 2 familles à la quantifications des descripteurs pour développer un recherche hiérarchique arborescente.

Références

- [1] P. Aigrain et al. *Content-based representation and retrieval of visual media: a state-of-the-art review*. Multimedia Tools and Applications, 3(3), 1996.
- [2] Y. Rui et al. *Image retrieval: current techniques, promising directions, and open issues*, JVCIR, 10, 39-62, 1999.
- [3] M. Flickner et al. *Query by image and video content: the QBIC system*. IEEE Computer, 28(9), 23-32, 1995.
- [4] J.R. Bach et al. *Virage image search engine: An open framework for image management*, Storage & Retrieval for image and video Databases IV, IS&T/SPIE, Feb. 96.
- [5] A. Pentland et al. *Photobook: Tools for content-based manipulation of image databases*. IJCV18, 233-254, 96.
- [6] F. Ghorbel. *A complete invariant description for gray-level images by the harmonic analysis approach*. Pattern Recognition Lett. 15, 1043-1051, 1994.
- [7] F. Ghorbelet al. *A complete and invariant description of gray level images from geometric moments*. Soumis à Pattern Recognition Lett, 1999.
- [8] H. Ming-Kuel. *Visual pattern recognition by moment invariants*. IRE trans. Inf. Theory IT(8), 179-187, 1962.
- [9] B.M. Mehtre, et al. *Shape measures for content based image retrieval: A comparison*. Information Processing & Management 33(3), 1997.
- [10] R. Milanese et al. *Invariant content based image retrieval using the Fourier-Mellin transform*. IAPR'98, Plymouth (UK), Nov. 1998.
- [11] S. Derrode et F. Ghorbel. *Robust and efficient Fourier-Mellin transform approximations for invariant gray-level image description and reconstruction*. Soumis à Pattern Recognition, 1998.