

# Une évaluation des descripteurs visuels MPEG-7 pour la recherche d'images par le contenu

Amel ZNAIDIA, Titus ZAHARIA et Françoise PRETEUX

Institut Télécom ; Télécom SudParis ; Département ARTEMIS  
9, Rue Charles Fourier, 91011 Évry France

{Amel.Znaidia, Titus.Zaharia, Françoise.Preteux}@it-sudparis.eu

**Résumé :** Le monde de la (post)-production génère chaque année des milliards de données visuelles à forte valeur ajoutée. Un enjeu économique majeur est de pouvoir les réutiliser de façon optimale. Au-delà des systèmes d'annotation (semi-) automatiques dont les limitations sont bien connues, cela nécessite aujourd'hui de maîtriser les nouveaux modes d'accès et de requêtes par le contenu. C'est l'un des objectifs du projet structurant HD3D-IIO du pôle de compétitivité Cap Digital. Cet article présente la première phase de l'approche technologique développée dans HD3D-IIO. Elle porte sur une évaluation expérimentale des descripteurs visuels MPEG-7 pour la recherche par le contenu d'images couleurs ou texturées. Le corpus d'analyse est composé de 2414 images naturelles et de 4426 images issues de séries de dessins animés. La vérité terrain indispensable pour une évaluation objective mesurable des performances des descripteurs, a été construite avec les experts et artistes. L'ensemble des descripteurs visuels MPEG-7 de couleur et de texture a été considéré pour évaluation selon différents critères. Les résultats expérimentaux obtenus sont discutés et démontrent la pertinence des approches d'indexation automatique proposées par la norme. Les meilleurs scores sont obtenus par les descripteurs *couleur structure* et *couleurs dominantes*.

**Mots clés :** Norme MPEG-7, indexation, recherche d'images par le contenu, requête par similarité, descripteur de couleur, descripteur de texture.

## 1 Contexte et objectifs

Des milliers de nouvelles images sont aujourd'hui produites ou auto-produites quotidiennement, dans des contextes aussi bien grand public qu'industriels. Cela conduit à la constitution de masses considérables de données multimédias, qui lancent à la communauté scientifique d'importants défis en termes de méthodologies et d'outils de structuration de bases de données, de description de contenus et de recherche d'information. Pour une utilisation efficace et optimale de ces collections multimédias, il est nécessaire de disposer d'outils adéquats et opérationnels d'accès et de description de l'information.

Identifiant le besoin croissant de normaliser la description des contenus audiovisuels, le groupe L'ISO/WG11 (MPEG- Moving Picture Expert Group) a développé la norme de description des contenus dit "Multimedia Content Description Interface", connue sous le nom de MPEG-7. La norme MPEG-7 regroupe un riche ensemble de descripteurs aussi bien par le contenu (audio et visuel) que sémantiques. Elle propose également des schémas de description qui combinent descripteurs et schémas de description. Le langage de description est fondé sur XML Schema [1]. Parmi les différentes parties de la norme, la partie 3, MPEG-7 Visual (ISO/IEC 15938/3) [2, 9], propose un éventail de descripteurs adaptés et optimisés pour des applications de recherche par le contenu selon différentes caractéristiques visuelles comme la forme, le mouvement, la texture ou la couleur.

L'objectif de cet article est d'investiguer dans quelle mesure les technologies proposées par MPEG-7 peuvent aujourd'hui répondre aux besoins des industriels de la production des contenus numériques. Le cadre de cette analyse est celui du projet structurant HD3D-IIO du pôle de compétitivité Cap Digital. HD3D-IIO concerne la création et la fabrication des contenus numériques pour les industries techniques de l'audiovisuel et du cinéma. Son ambition est de doter le secteur de

moyens technologiques nouveaux, dans un environnement de travail en mutation permanente et en confrontation avec les exigences de la compétitivité à l'échelle mondiale. Ces moyens doivent être conçus collectivement, dans la perspective d'une industrie ouverte du point de vue des échanges numériques. Le consortium HD3D-IIO regroupe des acteurs professionnels majeurs de la production audio-visuelle (publicité, dessin animés 2D/3D, films de cinéma, effets spéciaux 2D/3D...) en Ile de France, comme les sociétés Attitude Studio, Duboi Duran, Eclair, LTC, Mac Guff, Mikros Images, Teamto, ou 2 Minutes.

Un des objectifs de HD3D-IIO concerne la réutilisation des contenus images en environnement de production. Ce besoin fait écho à la constatation suivante : chaque production (*e.g.* dessin animé, clip vidéo ou film) conduit à des dizaines de milliers d'images. Or, de nombreux éléments (*e.g.* parties de décor d'un film, personnages/accessoires d'un dessin animé...), peuvent être ré-exploités lors d'une nouvelle création, à condition toutefois de pouvoir les retrouver facilement dans les collections d'images précédemment produites. Aujourd'hui, aucun outil efficace n'est disponible pour accéder à de telles bases d'images et pour faciliter/accélérer les modes de production multimédia en capitalisant sur des créations antérieures. Le seul recours possible est d'exploiter le savoir-faire et l'expérience des professionnels et artistes impliqués dans la chaîne de production. Cela est en revanche consommateur en ressources humaines et a un impact sur le temps et donc le coût de production.

L'indexation automatique et par le contenu d'images pourrait apporter des éléments de réponse à cette problématique. Deux contraintes doivent être néanmoins respectées pour garantir une solution viable. La première concerne l'efficacité des requêtes (*i.e.* capacité de retrouver l'information pertinente dans des grandes bases d'images). La seconde est liée au temps de calcul associé qui doit être compatible avec des réponses quasi instantanées à des requêtes interactives.

Cet article se propose d'établir, à l'aide d'une étude expérimentale sur un corpus de données réelles du monde audio-visuel, comment les technologies MPEG-7 de description des contenus visuels peuvent répondre à ces différentes contraintes. Les descripteurs MPEG-7 considérés pour évaluation sont brièvement rappelés dans la section 2. L'étude expérimentale comparée est ensuite présentée en section 3, ainsi que la discussion détaillée sur les performances obtenues. Enfin, la section 4 conclut l'article et ouvre quelques perspectives de développement futur.

## 2 Descripteurs visuels MPEG-7 considérés

Les descripteurs visuels MPEG-7 [2, 9] prennent en compte l'ensemble des attributs de couleur, texture, forme et mouvement classiquement utilisés dans le domaine de l'indexation par le contenu. Pour l'application HD3D-IIO, seuls les attributs de couleur et de texture ont été jugés pertinents dans la phase actuelle du projet.

Le standard MPEG-7 propose quatre descripteurs de couleur et trois descripteurs de texture. Présentons tout d'abord de manière concise chacune de ces représentations, en énonçant le principe, les principales caractéristiques ainsi que les mesures de similarité associées.

### 2.1 Descripteurs de Couleur

Attribut largement utilisé dans le domaine de l'indexation d'images et la recherche par le contenu, la couleur a été également considérée dans le cadre de MPEG-7. Quatre descripteurs de couleur adaptés aux objectifs de requête par similarité sont actuellement normalisés.

#### 2.1.1 Histogramme Scalable par Transformée de Haar

Ce descripteur propose une représentation multi-résolution, par transformée de Haar [3], d'un histogramme de couleur spécifié dans l'espace HSV (Hue-Saturation-Value). Le caractère multi-résolution de la transformée assure un premier type de scalabilité, un même histogramme admettant des représentations à différents niveaux de détails (*i.e.* différents nombre de bins).

Les coefficients de Haar sont quantifiés de manière non-linéaire en un nombre de plans de bits (à partir des bits les plus significatifs), d'une façon optimisée pour l'espace de couleurs HSV. Cela conduit à un deuxième type de scalabilité, allant des approximations binaires des coefficients jusqu'aux représentations en pleine résolution. Les mesures de similarités recommandées par MPEG-7 s'appuient aussi bien sur la distance  $L_1$  dans l'espace des coefficients de Haar, que sur une distance  $L_2$  dans l'espace des histogrammes.

Par définition, l'histogramme scalable de couleur est invariant par rapport aux transformations de similarité (rotations et homothéties).

Comme tout histogramme de couleur, le descripteur MPEG-7 n'intègre aucune information de localisation spatiale, ce qui limite son pouvoir de discrimination. Le descripteur couleur-structure, présenté par la suite s'affranchit de cette limitation.

### 2.1.2 Histogramme couleur-structure(CS)

Le descripteur couleur-structure (CS) étend et enrichit la notion d'histogramme en introduisant dans la représentation un minimum d'information spatiale locale. Un élément structurant, représenté par un masque binaire, est translaté en chaque pixel de l'image. Tous les intervalles de l'histogramme correspondant aux couleurs présentes à l'intérieur du masque sont alors incrémentés. Ainsi, l'histogramme CS représente-t-il la fréquence relative des éléments structurants contenant une couleur donnée.

Les mesures de similarités adoptées pour l'histogramme CS sont les distances  $L_1$  et  $L_2$ .

Pour assurer une certaine invariance par rapport aux homothéties, les images sont normalisées à une taille fixe avant l'extraction du descripteur.

### 2.1.3 Descripteur par couleurs dominantes

Ce descripteur fournit une description compacte des couleurs représentatives dans une image. Le descripteur est défini par :

$$F = \{(c_i, p_i, v_i, s)\}, (i = 1, 2, \dots, N) \quad (1)$$

où  $N$  désigne le nombre de couleurs dominantes (au maximum 8 dans MPEG-7),  $c_i$  la valeur (dans un espace de couleur donné) de la  $i^{\text{ème}}$  couleur dominante,  $p_i$  un poids exprimant sa fréquence relative d'apparition dans l'image et  $v_i$  sa variance. La cohérence spatiale  $s$  représente l'homogénéité spatiale des couleurs dominantes dans l'image.

Par définition, le descripteur par couleurs dominantes est intrinsèquement invariant aux transformations de similarité.

Considérons les deux descripteurs ci-dessous :

$$\begin{aligned} F_1 &= \{c_{1i}, p_{1i}, v_{1i}, s_1\}, (i = 1, 2, \dots, N_1) \\ F_2 &= \{c_{2i}, p_{2i}, v_{2i}, s_2\}, (i = 1, 2, \dots, N_2) \end{aligned} \quad (2)$$

Le degré de similarité entre deux couleurs  $c_{1k}$  et  $c_{2l}$ , noté  $a_{1k,2l}$ , est défini par:

$$a_{1k,2l} = \begin{cases} 1 - \frac{\delta(c_{1k}, c_{2l})}{\delta_{\max}} & \text{si } \delta(c_{1k}, c_{2l}) \leq T_d \\ 0 & \text{sinon} \end{cases} \quad (3)$$

où  $\delta(c_{1k}, c_{2l})$  désigne la distance euclidienne entre deux couleurs  $c_{1k}$  et  $c_{2l}$ , et  $T_d$  est un seuil prédéfini.

La mesure de similarité entre deux descripteurs  $F_1$  et  $F_2$  est alors définie par :

$$D^2(F_1, F_2) = \sum_{i=1}^{N_1} p_{1i}^2 + \sum_{j=1}^{N_2} p_{2j}^2 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2 * a_{i,2j} p_{1i} p_{2j} \quad (4)$$

Cette mesure correspond à une version explicitée de la forme quadratique proposée dans [4] sous l'hypothèse que la distance entre deux couleurs dominantes d'une description est supérieure au seuil  $T_d$ . Cette hypothèse doit être assurée par l'algorithme d'extraction des couleurs dominantes. A titre informatif (les algorithmes d'extraction des descripteurs ainsi que les mesures de similarité restent non-normatifs), le standard recommande l'utilisation de l'algorithme LBG (Lindo-Buzo-Gray) [5].

Notons la complexité de calcul de cette mesure de similarité, qui est quadratique avec le nombre de couleurs dominantes. Toutefois, comme le nombre maximal de couleurs dominantes étant limité à 8, cela ne pénalise pas les temps de réponse aux requêtes (cf. Section 3).

#### 2.1.4 Distribution spatiale de couleur

Ce descripteur vise à capturer la disposition spatiale (*color layout*) des couleurs dans une image. L'image est tout d'abord divisée en 64 ( $8 \times 8$ ) blocs rectangulaires. Chaque rectangle est représenté par sa couleur dominante, qui est par définition la moyenne des couleurs des pixels constituant le rectangle considéré. On construit ainsi trois matrices ( $8 \times 8$ ), une par composante de couleur. Les coefficients quantifiés des transformées en cosinus discrète 2D (DCT - *Discrete Cosine Transform*) de chacune de ces matrices sont ensuite calculés. Ces coefficients sont parcourus en zigzag, à partir des fréquences les plus basses.

La mesure de similarité recommandée pour ce descripteur est une distance  $L_2$  pondérée entre les coefficients DCT ainsi déterminés.

Notons que le descripteur de distribution spatiale de couleur n'est pas invariant aux rotations. Son utilisation est donc pertinente uniquement pour des applications concernant des requêtes globales, où le positionnement des différents couleurs dans l'image doit être pris en compte.

Cela conclut la revue des descripteurs MPEG-7 de couleur. Analysons à présent comment l'information de texture est prise en compte par la norme.

## 2.2 Descripteurs de texture

La texture de l'image représente une primitive visuelle de première importance pour la navigation et la recherche par similarité dans de larges collections d'images. Dans le cadre du standard MPEG-7, trois descripteurs de texture sont proposés, dont seulement deux, rappelés ci-dessous, sont adaptés pour des requêtes par similarité.

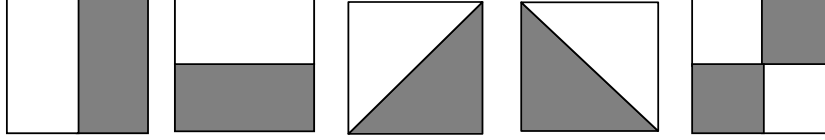
### 2.2.1 Histogramme d'orientation des contours

Ce descripteur représente l'histogramme des orientations des contours (gradient) de l'image. Les orientations sont grossièrement classées en cinq catégories : horizontale, verticale, diagonale à  $45^\circ$ , diagonale à  $135^\circ$  et non-directionnelle. Pour leur détection, on utilise un banc de cinq filtres linéaires adaptés à chacune de ces classes (Figure 1).

Notons que le descripteur est calculé à partir de l'image de luminance associée à une image couleur quelconque.

1	-1	1	1	$\sqrt{2}$	0	0	$\sqrt{2}$	2	-2
1	-1	-1	-1	0	$-\sqrt{2}$	$-\sqrt{2}$	0	-2	2

a. Filtres directionnels.



b. Types d'arêtes associées.

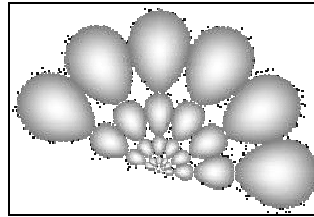
**Figure 1 :** Détection des orientations de contour par filtres directionnels adaptés.

Des versions semi-globales du descripteur d'orientation des contours sont également proposées.

Dans tous les cas, la distance  $L_1$  entre les divers histogrammes obtenus, est suggérée comme mesure de similarité.

### 2.2.2 Descripteur de texture homogène

Ce descripteur est fondé sur des mesures énergétiques dans des sous-bandes du domaine spectral de Fourier correspondant à un banc de filtres de Gabor [6, 7]. Dans MPEG-7, cinq échelles et six orientations (Figure 2) sont considérées, ce qui correspond à un total de 30 sous-bandes spectrales.



**Figure 2 :** Réponses fréquentielles des filtres de Gabor.

Soit  $S(\omega, \theta)$  la transformée de Fourier de l'image à analyser. Numérotons les 30 sous-bandes selon la règle suivante :

$$\forall s \in \{0, 1, \dots, 4\}, \forall r \in \{0, 1, \dots, 5\} \quad i = 6 * s + r + 1 \quad (5)$$

où  $s$  est le paramètre d'échelle et  $r$  celui d'orientation.

En notant par  $G_i(\omega, \theta)$  la fonction de transfert du filtre de Gabor associé à la sous-bande  $i$ , l'énergie moyenne  $e_i$ , et sa dispersion  $\sigma_i$ , sont définies par les équations suivantes :

$$e_i = \int \int_{\omega \theta} |G_i(\omega, \theta) S(\omega, \theta)|^2 |\omega| d\theta d\omega \quad (6)$$

$$\sigma_i = \int \int_{\omega \theta} \left[ |G_i(\omega, \theta) S(\omega, \theta)|^2 - e_i \right]^2 |\omega| d\theta d\omega \quad (7)$$

Le descripteur de texture homogène, noté  $T$ , est alors défini par le vecteur de tous ces éléments :

$$T = (\log(1 + e_i), \dots, \log(1 + \sigma_i)) \quad (8)$$

Une distance  $L_1$  pondérée entre les descripteurs est recommandée comme mesure de similarité.

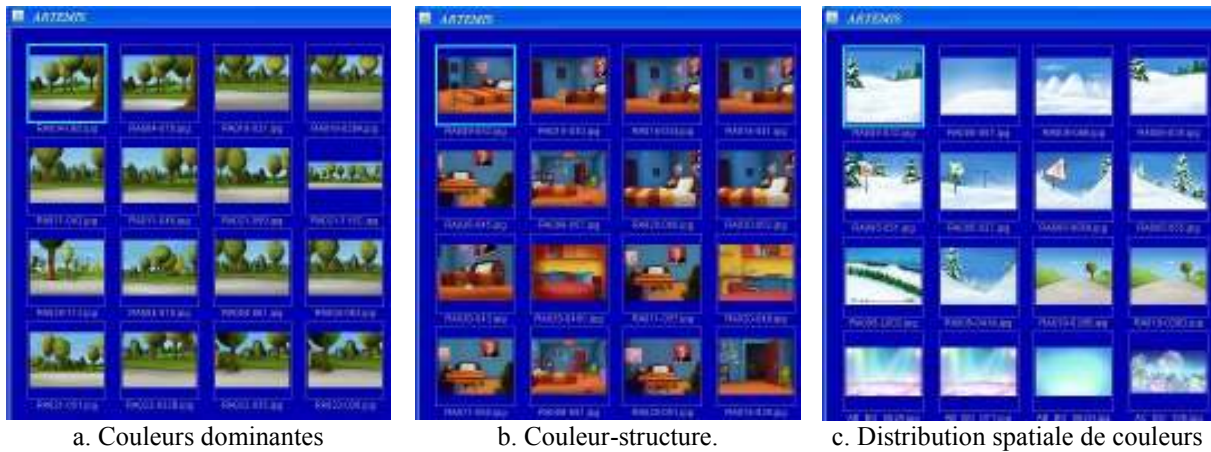
En ce qui concerne les critères d'invariance, aucun des deux descripteurs de texture n'est invariant aux transformations de similarité. Toutefois, l'invariance peut être acquise en modifiant les mesures de



- *Bull-Eye* (BE), défini comme le pourcentage d'images pertinentes retrouvées dans les  $(2Q)$  premiers résultats retournés.



Figure 4 : Requête par similarité avec la plate-forme d'indexation MPEG-7 AMIS.



a. Couleurs dominantes      b. Couleur-structure.      c. Distribution spatiale de couleurs

Figure 5 : Résultats des requêtes pas similarité sur la base 2minutes.



a. Histogramme scalable.      b. Histogramme d'orientations.      c. Distribution spatiale de couleurs

Figure 6 : Résultats des requêtes pas similarité sur la base Mikros.

Les Figures 5 et 6 présentent respectivement des exemples de résultats de requêtes pour quelques images de test des bases 2minutes et Mikros. Dans tous les cas, des images similaires sont obtenues sur les 16 premiers résultats retrouvés.

Les scores FT et BE moyens (sur l'ensemble des images catégorisées) associés aux descripteurs MPEG-7, ainsi que les temps de réponse aux requêtes sur les deux bases considérées sont indiqués dans les Tableaux 1 et 2.

Sur la base Mikros, nous observons que le descripteur couleur-structure présente les meilleurs résultats de requête (avec des scores FT et BE de 63% et 71%, respectivement), pour un temps de réponse aux requêtes raisonnable (2 secondes). Le descripteur par histogramme de couleur scalable offre des performances légèrement inférieures (scores FT et BE de 59% et 68%, respectivement). Cela montre la pertinence d'injecter dans la description un minimum d'information spatiale.

Pour ces images hautement texturées, le descripteur par couleur dominante est pénalisé par le nombre réduit de couleurs dominantes supporté par MPEG-7 (maximum 8) et n'arrive pas à prendre en compte la diversité des couleurs présentes dans ces images. Comment décrire en effet ces images naturelles de scènes urbaines fortement texturées par seulement 8 couleurs dominantes ? Les scores FT et BE associés tombent à 51% et 63%, respectivement. Des performances similaires sont obtenues par le descripteur de distribution spatiale de couleur.

**Tableau 1 :** Performances des requêtes par similarité sur la base de test Mikros.

Descripteurs	Score FT	Score BE	Temps d'extraction	Temps de calcul de la mesure de similarité
Dominant color	51.53 %	63.86 %	42.12 s	0.12 s
Color Structure	<b>63.21 %</b>	<b>71.84 %</b>	1.92 s	0.06 s
Scalable Color	59.24 %	68.69 %	1.87 s	0.07 s
Color Layout	52.79 %	64.21 %	0.76 s	0.06 s
Edge Histogram	48.58 %	58.43 %	0.32 s	0.09 s
Homogenous Texture	31.49 %	39.87 %	0.2 s	0.2 s

**Tableau 2 :** Performances des requêtes par similarité sur la base de test 2minutes.

Descripteurs	Score FT	Score BE	Temps d'extraction	Temps de calcul de la mesure de similarité
Dominant color	<b>57.21 %</b>	<b>72.58 %</b>	10.26s	0.35s
Color Structure	54.72 %	67.36 %	0.51s	0.23s
Scalable Color	56.30 %	65.11 %	0.42s	0.25s
Color Layout	40.62 %	53.41 %	0.19s	0.18s
Edge Histogram	28.93 %	32.09 %	0.07s	0.25s
Homogenous Texture	20.52 %	25.29 %	0.18s	0.48s

En ce qui concerne les descripteurs de texture, leurs performances sont nettement inférieures à celles des descripteurs couleur. Le descripteur d'orientations des contours n'est pas invariant aux rotations, ce qui en pénalise les performances (score FT de 48%). Quant au descripteur de texture homogène, il reste bien trop élémentaire pour décrire ces images complexes (score BE de 39%), où plusieurs textures différentes co-existent dans une même image. Cela est sans surprise en raison de la complexité conceptuelle de la notion de texture, dont le rôle dans le monde de l'indexation d'images naturelles reste encore une question ouverte.

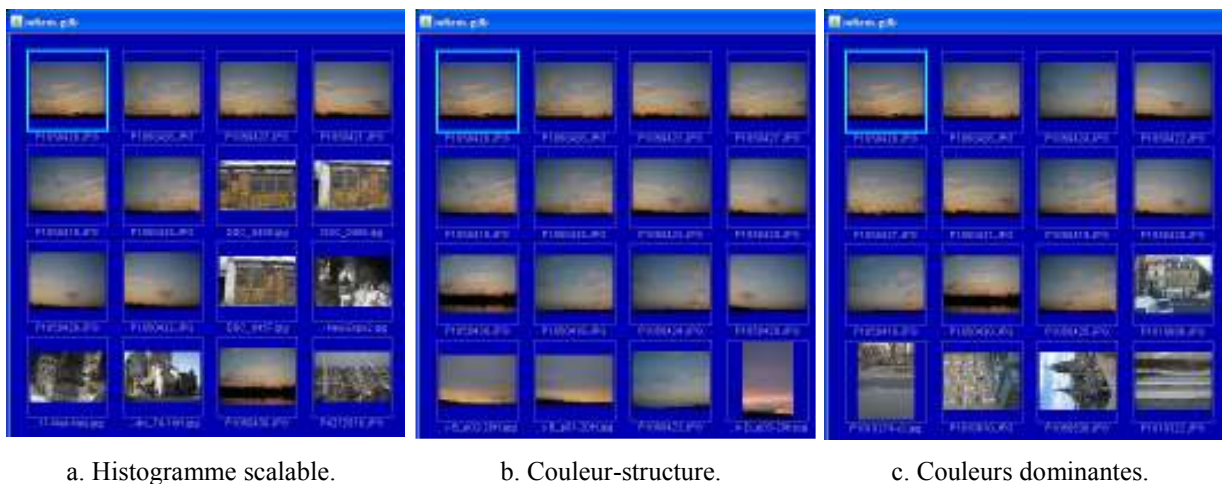
Examinons à présent les performances des requêtes sur la base d'images synthétiques 2minutes.

Les meilleurs résultats sont obtenus par les descripteurs de couleurs dominantes (score FT de 57% et BE de 72%) et d'histogramme scalable (score FT de 56% et BE de 65%). Le descripteur couleur-structure n'apporte rien dans le cas de ces images aux couleurs relativement homogènes (score FT de 54% et BE de 67%). Comme dans le cas de la base Mikros, ici encore le descripteur de distribution spatiale offre des performances limitées.

En ce qui concerne les descripteurs de texture, les performances associées chutent drastiquement (scores FT de 28% et de 20% respectivement pour les descripteurs d'orientation des contours et de texture homogène). Cela est sans surprise pour ces images qui présentent des textures plus ou moins uniformes.

D'une manière générale, les seuls descripteurs qui offrent sur les deux bases considérées des scores BE supérieurs à 65% et donc acceptables pour l'application considérée sont l'histogramme de couleur scalable, le descripteur par couleurs dominantes et le descripteur couleur-structure. Le descripteur par couleurs dominantes est de loin le plus adapté pour les images synthétiques de la base 2minutes. Ici, en raison du caractère quasi-homogène des couleurs, ce descripteur trouve toute sa pertinence. En revanche, pour les images naturelles, ses performances se dégradent significativement.

La Figure 7 présente un exemple de requête par similarité selon une même image requête (représentant un coucher de soleil) avec les descripteurs histogramme scalable, couleur-structure et par couleurs dominantes.



**Figure 7 :** Requetes pas similarité sur la base Mikros selon une même image requête.

Les meilleurs résultats dans ce cas sont obtenus par le descripteur couleur-structure, qui retrouve des images similaires sur les 16 premières positions. Pour les descripteurs d'histogramme scalable et par couleurs dominantes, 7 et 5 résultats faux positifs sont respectivement retrouvés dans les 16 premières images retrouvées. Cela montre encore une fois l'intérêt d'intégrer dans la description un minimum d'information spatiale.

Les temps de réponse aux requêtes sont équivalents pour l'ensemble des descripteurs considérés, à l'exception du descripteur par couleurs dominantes, qui est de loin le plus complexe. Le calcul de la mesure de similarité est effectué en moins d'une seconde pour l'ensemble des descripteurs. Vu le volume des deux bases considérées (2414 et 4426 images), cela est tout à fait acceptable.

En ce qui concerne le temps d'extraction du descripteur, il est de l'ordre de 0.1-2 secondes pour l'ensemble des descripteurs sauf pour le descripteur de couleurs dominantes (42s), qui implique une procédure de segmentation complexe. Des optimisations de cette procédure ou bien l'utilisation d'autres algorithmes de *clustering*/segmentation (comme celui proposé dans [12]) seraient nécessaires afin de ne pas pénaliser ce descripteur.

Ces résultats démontrent que les technologies MPEG-7 restent aujourd'hui encore pertinentes pour des objectifs de requêtes par similarité. Des études plus récentes ré-considèrent les approches MPEG-7 tout en les améliorant. C'est le cas notamment du descripteur par couleurs dominantes, qui a fait l'objet d'études récentes [13, 14]. Ainsi, dans [13] les auteurs introduisent un nouvel algorithme de quantification de couleurs appelé *Linear Block Algorithm* (LBA) qui permet de diminuer la complexité de calcul de la procédure d'extraction du descripteur. Une nouvelle mesure de similarité, plus proche de la perception humaine est également proposée. Enfin, mentionnons également l'approche proposée dans [14], appelé *Dominant Color Structure Descriptor* qui hybride les avantages des descripteurs couleur-structure et par couleurs dominantes pour une représentation encore plus discriminante.

Notons que procédure d'extraction, mesure de similarité ainsi que modalités de combinaison des descripteurs restent non-normatives, ce qui permet aux technologies MPEG-7 de bénéficier pleinement de nouvelles avancées méthodologiques.

#### 4 Conclusions et perspectives

Cet article présente une évaluation expérimentale des descripteurs visuels MPEG-7 pour la recherche d'images par le contenu. Les performances ont été évaluées objectivement dans le cadre du projet Cap Digital HD3D-IIO à partir de deux bases de test avec vérité terrain. Les scores First Tier et Bull Eye pour l'histogramme scalable, le descripteur par couleurs dominantes ou encore le descripteur couleur-structure, sont supérieurs à 60-70%. Pour des applications de type capitalisation/réutilisation de contenus, ces résultats démontrent la pertinence des outils MPEG-7.

Dans le futur, nous proposons d'améliorer les performances du descripteur par couleurs dominantes en raison de son fort potentiel en termes de pouvoir de discrimination. Il s'agira d'optimiser la procédure d'extraction qui pénalise actuellement sévèrement cette approche en raison de sa complexité de calcul. Ensuite, nous étudierons l'impact du nombre de couleurs dominantes considéré sur les performances des requêtes. Enfin, à plus long terme, nos recherches porteront sur l'analyse des mécanismes de combinaison/fusion des attributs de couleur et de texture dans un modèle unifié et adapté aux requêtes par similarité.

#### Références

- [1] <http://www.w3.org/XML/Schema>.
- [2] L. Cieplinski, W.-Y. Kim, J.-R. Ohm, M. Pickering, A. Yamada, "Text of ISO/IEC 15938-3/FDIS Information technology - Multimedia content description interface - Part 3 Visual", ISO/IEC JTC1/SC29/WG11, MPEG01/N4358, Juillet 2001.
- [3] A. Poularakis, The transforms and applications handbook, Electrical Engineering Handbook Series, Boca Raton, 2000.
- [4] J. Hafner, S. Sawhney, W. Equitz, M. Flickner, W. Niblack, "Efficient color histogram indexing for quadratic form distance functions", IEEE Trans. on PAMI, Vol. 17, No. 7, Juillet 1995.
- [5] Y.Linde, A. Buzo, R. M. Gray, An Algorithm for vector Quantizer Design, IEEE Trans, Commun., Vol. COM-28, pages 84-95, January 1980.
- [6] B. S. Manjunath and P. Wu and S. Newsam and H. D. Shin, "A Texture Descriptor for Browsing and Similarity Retrieval", Journal of Signal Processing: Image Communication, Vol. 16, pp. 33-43, 2000.
- [7] C. K. Chui, Introduction to wavelets, Academic Press, Boston, MA, 1992.
- [8] T. Zaharia, F. Prêteux, AMIS: A Java-based MPEG-7 compliant indexing system, Proceedings IAPR International Conference on Image and Signal Processing (ICISP'2003), Agadir, Morocco, Vol. 1, pages 104-111, 2003.
- [9] B.S. Manjunath, P. Salembier, T. Sikora (Editors), "Introduction to MPEG-7", John Wiley & Sons, ISBN 0471486787, Royaume-Uni, 2002.
- [10] R. Osada, T. Funkhouser., B. Chazelle., D.P. Dobkin., "Matching 3D models with shape distributions", *Proc. International Conference on Shape Modeling and Applications*, Mai 2001.
- [11] T. Zaharia, F. Prêteux, "3D versus 2D/3D shape descriptors: A comparative study", Proc. SPIE Conference on Image Processing: Algorithms and Systems III - IS&T / SPIE Symposium on Electronic Imaging, Science and Technology '04, San Jose, CA, Vol. 5298, January 2004, pp. 47-58.
- [12] D. Comaniciu, P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", IEEE Trans. Pattern Analysis Machine Intell, Vol. 24, No. 5, pp. 603-619, 2002.
- [13] Nai-Chung Yang, Wei-Han Chang, Chung-Ming Kuo, Tsia-Hsing Li, A Fast MPEG-7 dominant color extraction with new similarity measure for image retrieval, Journal of Visual Communication and Image Representation, vol. 19(2), pp. 92-105, 2008.
- [14] Ka-Man Wong, Lai-Man Po, Kwok-Wai Cheung, Dominant Color Structure Descriptor for Image Retrieval, *IEEE International Conference on ImageProcessing*, vol. 6, pp. 365-368, 2007.