
La plateforme INVENIO

Une approche MPEG-7 pour la réutilisation des contenus multimédias

RÉSUMÉ. Dans cet article, nous proposons une nouvelle plateforme d'indexation de contenus multimédias, dite INVENIO (INDEXING VISUAL ENVIRONMENT FOR MULTIMEDIA ITEMS AND OBJECTS). Fondée entièrement sur la norme ISO/MPEG-7, la plateforme INVENIO offre, dans un système intégré, moteurs d'extraction de métadonnées MPEG-7, outils d'annotation, moteur de requête, outils de gestion de bases de données multimédias, ainsi que des interfaces utilisateurs appropriées et ergonomiques. Pour valider la plateforme INVENIO, nous avons considéré une application industrielle d'indexation et de reconnaissances d'images couleur/texture fixes et animés, liée au projet CapDigital HD3D-IIO. Il s'agit notamment d'une problématique de réutilisation de contenus numériques au sein d'une chaîne de production audiovisuelle. Les expérimentations réalisées concernent différentes chaînes de production audiovisuelle, incluant des contenus aussi bien naturels que de synthèse (i.e. dessins animés). Les solutions d'indexation et de reconnaissance d'images proposées dans de cet article démontrent notamment que l'exploitation des technologies MPEG-7 à travers la plateforme d'indexation INVENIO permet de réaliser une économie significative de temps de travail, ainsi qu'une réutilisation optimale des contenus numériques en cours de production.

ABSTRACT. This paper proposes a new multimedia indexing platform, so-called INVENIO (INDEXING VISUAL ENVIRONMENT FOR MULTIMEDIA ITEMS AND OBJECTS). Based on the ISO/MPEG-7 standard INVENIO integrates within a unified platform feature extraction engines, annotation tools, database management utilities and search engine, with user-friendly and ergonomic user interfaces. For validation of the INVENIO system, we have considered an industrial application related to the French CapDigital HD3D-IIO structuring project and concerning the re-use of content within the audio-visual production chain. Experiments relate to different audio-visual production chains, including both natural and synthetic content (i.e. cartoons). The proposed indexing solutions show that the exploitation of the MPEG-7 content-based indexing technologies within the INVENIO system makes it possible to achieve significant gains in production time as well as an optimal reutilisation of digital content under production.

MOTS-CLÉS : Plateforme d'indexation d'images, norme MPEG-7, indexation par le contenu, moteur de recherche, interfaces, ergonomie, couleur, texture.

KEYWORDS: Image indexing platform, MPEG-7 standard, content-based image indexation, search engine, interfaces, ergonomics, texture, color.

1. Introduction

L'Internet, omniprésent dans le quotidien d'un public de plus en plus large à travers le monde, s'est imposé depuis maintenant des années comme le média incontournable, qui rassemble dans des immenses bases de données distribuées, le plus vaste volume d'information disponible aujourd'hui. Les contenus numériques de l'Internet deviennent de plus en plus divers, autoproduits et personnalisés, incluant images, vidéos, données graphiques 3D... Cette immense masse de données distribuées nécessite en revanche de disposer de solutions efficaces d'indexation et de recherche, permettant à l'utilisateur d'accéder à l'information. Les solutions commerciales existantes s'appuient aujourd'hui exclusivement sur des moteurs de recherche par mots-clés (*e.g.* Google, Yahoo, Exalead...). Dans le cas spécifique de contenus audiovisuels, les limitations d'une telle approche sont multiples : polysémie et complexité des contenus, barrières linguistiques, subjectivité des annotations, difficulté d'annotation manuelle d'un grand volume de données...

Aujourd'hui les méthodologies d'indexation par le contenu (Smeulders *et al.*, 2000) visent à proposer des solutions alternatives pour l'indexation des documents multimédias. Le principe consiste à associer aux contenus des métadonnées non plus textuelles, mais liées intrinsèquement au contenu audiovisuel lui-même. Cela est possible puisque les représentations mathématiques permettent la description de manière automatique (ou semi-automatique) et discriminante du flux audiovisuel. Ce sont des caractéristiques perceptuelles associées classiquement aux attributs visuels comme la forme, la couleur, la texture ou encore le mouvement, qui rendent cette indexation possible. Le principe de la recherche de contenus dans une base de données change alors de façon fondamentale. L'utilisateur présente à l'entrée de son moteur de recherche non plus des mots-clés mais un exemple (*e.g.* une image/vidéo ou une partie d'une image/vidéo), ou une ébauche (*e.g.* un dessin réalisé à la main). Les mesures de similarités associées aux descripteurs images permettent alors de réaliser des requêtes automatiques et de retrouver des images similaires.

Le domaine de l'indexation d'images par le contenu a connu une effervescence spectaculaire depuis le milieu des années 1990 et jusqu'à présent, comme en témoignent l'impressionnant volume de méthodes et techniques proposées dans la vaste littérature scientifique (voir Smeulders *et al.*, 2000, Datta *et al.*, 2005 pour des synthèses critiques) consacrée à ce sujet. Dans ce cadre, une étape marquante a été la sortie officielle au début des années 2000 de la norme ISO MPEG-7 (ISO/IEC 15938-3, ISO/IEC 15938-5). Officiellement appelée « *Multimedia Content Description Interface* », la norme MPEG-7 propose un large éventail de technologies de description de documents multimédia, intégrant des approches aussi bien textuelles que par le contenu.

En particulier, la spécification ISO MPEG-7 propose aujourd'hui un riche ensemble de descripteurs visuels (Manjunath *et al.*, 2002, Auteur *et al.*, 2004) et de schémas de description, exprimés dans un langage de description de données fondé sur une approche XML Schema (XML Schema). Toutefois, la pertinence de la

norme pour des applications industrielles grandeur nature reste à démontrer. Pour cela, il est indispensable d'élaborer et de mettre en œuvre des outils efficaces et ergonomiques permettant de faciliter les processus d'annotation, de consultation, de navigation dans des bases de données et de requête de l'information.

La plateforme d'indexation INVENIO (*INDEXING VISUAL ENVIRONMENT FOR MULTIMEDIA ITEMS AND OBJECTS*) proposée dans cet article propose une solution unifiée à cette problématique. Fondée entièrement sur la norme MPEG-7, la plateforme INVENIO offre, dans un système intégré, moteurs d'extraction de métadonnées MPEG-7, outils d'annotations, moteur de requête, outils de gestion de bases de données, ainsi que des interfaces utilisateurs appropriées et ergonomiques.

Le domaine d'application ciblé par INVENIO est, de manière générale, celui de l'indexation d'images par le contenu. Toutefois, pour valider la plateforme, nous avons développé une application industrielle d'indexation d'images couleur/texture pour la production de films d'animation. L'objectif est de favoriser la réutilisation de contenus numériques au sein d'une chaîne de production audiovisuelle. Ce développement est lié au projet structurant HD3D-IIO du pôle de compétitivité CapDigital (www.capdigital.com).

La suite de cet article est structurée comme suit. Tout d'abord nous énonçons les objectifs généraux du projet HD3D-IIO en mettant notamment l'accent sur la problématique de réutilisation de contenus au sein d'une chaîne de production audiovisuelle (Paragraphe 2). L'approche par indexation d'images proposée est ensuite décrite, en précisant notamment les éléments de la norme MPEG-7 retenus pour nos développements, les aspects d'interface utilisateur et d'ergonomie ainsi que les fonctionnalités supportées (Paragraphe 3). Une évaluation expérimentale conduite sur les bases de données grandeur nature constituées au sein du projet HD3D-IIO est présentée dans le Paragraphe 4. Enfin, le Paragraphe 5 conclut l'article et ouvre quelques perspectives de travail futur.

2. Contexte HD3D-IIO : réutilisation de contenus

Le projet 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 entre diverses entreprises du secteur. Le consortium HD3D-IIO regroupe des acteurs professionnels majeurs de la production audiovisuelle (publicité, dessin animés 2D/3D, films de cinéma, effets spéciaux 2D/3D...) en Ile de France, comme les sociétés Attitude Studio, Duran Duboi, Eclair, LTC, Mac Guff, Mikros Images, Teamto, ou 2 Minutes.

À long terme, l'ambition de HD3D-IIO est d'augmenter la compétitivité de l'industrie francilienne du film numérique face à des concurrents anglais ou

américains. Pour cela, de nombreux verrous technologiques nécessitent d'être levés. Cela concerne la spécification de formats d'échanges de contenus 2D et 3D, l'élaboration de plateformes de production collaborative avec une maîtrise des développements réalisés sur des sites distants, la spécification de méthodes de réutilisation de contenus ou encore la prise en compte des aspects de protection et de traçabilité des contenus, nécessaires pour leur transmission sécurisée.

Dans ce cadre, un des objectifs majeurs 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 2D ou 3D...), 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 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 consommateur en ressources humaines et à un impact sur le temps et donc sur le coût de production.

L'indexation automatique de texture et donc des contenus des bases de données d'images manipulées lors d'une production peut représenter un gain de temps significatif lors de la production d'un dessin animé. En effet, un moteur d'indexation automatique permet la recherche et l'extraction d'images afin d'optimiser la réutilisation de contenus numériques. Aujourd'hui le partage et la réutilisation des contenus dans le cadre d'une production est difficile, les solutions existantes présentent de nombreuses limitations :

- l'indexation manuelle est fastidieuse et peu utilisée à cause des contraintes de temps pour l'indexer,
- la construction et le classement des éléments nécessaires à la production de dessins animés (décors, textures, objets, personnages,...) dans des fichiers à l'intérieur d'une arborescence au sein d'un disque dur partagé sur l'intranet est utile mais peu commode puisque les utilisateurs doivent maîtriser la logique de sa construction (classement, arborescence).
- la base de données avec une recherche par mots clés ou par nom ou par extension de fichiers est fastidieuse. Souvent, les mots clés sont absents. De plus, même lorsqu'ils existent, ils sont trop limitatifs. Ainsi, les noms de fichiers suivent la logique définie en début de production et la nomenclature varie fortement d'une production à une autre. Quand à l'extension une recherche sur des images JPEG ou PSD (extension de fichier du logiciel Photoshop) peut conduire à un résultat de plusieurs milliers d'images.

L'indexation automatique par le contenu d'images pourrait apporter des éléments de réponse à cette problématique. Au minimum trois 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. Enfin, la troisième concerne l'élaboration de moteurs de recherche adaptés et aisés d'utilisation, supportant un large éventail de formats d'images et intégrables dans les chaînes de production existantes.

La plateforme d'indexation INVENIO proposée dans cet article, décrite dans la section suivante, se propose notamment de répondre à ces différentes contraintes pour assurer une réutilisation optimale des contenus numériques en cours de production.

3. La plateforme INVENIO

L'approche d'indexation d'images par le contenu proposé dans la plateforme INVENIO s'appuie entièrement sur les technologies de description de contenus multimédias proposées par la norme ISO MPEG-7.

La section suivante décrit notamment les éléments MPEG-7 retenus pour intégration dans la plateforme en raison de leur pertinence pour les objectifs de réutilisation et d'indexation automatique d'images.

3.1 Cadre normatif MPEG-7 : descripteurs retenus

La norme MPEG-7 regroupe un riche ensemble de descripteurs aussi bien par le contenu (audio et visuel) que sémantiques/textuels. Elle propose également une architecture qui combine descripteurs et schémas de description pour décrire des concepts liés à la structure de l'image et de plus haut niveau sémantique. Le langage de description est fondé sur XML Schema.

Parmi les différentes parties de la norme, la partie 3, dite *MPEG-7 Visual* (ISO/IEC 15938/3, 2002), 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.

Pour l'application HD3D-IIO, seuls les attributs de couleur et de texture ont été jugés pertinents dans la phase actuelle du projet. C'est notamment cette partie qui a été retenue pour intégration dans la plateforme INVENIO.

Le standard MPEG-7 propose actuellement quatre descripteurs de couleur et trois descripteurs de texture, présentés succinctement dans le tableau 1.

Descripteurs de couleur	
<i>Nom</i>	<i>Principe</i>
Espace de couleur	Spécification de l'espace de couleur de représentation des couleurs (<i>e.g.</i> RGB, HSV, Luv, HMMD...)
Quantification de couleur	Spécification d'une quantification uniforme de l'espace de couleur (composante par composante)
Histogramme de couleur scalable	Histogramme de couleurs indexé et représenté de manière multi-résolution par transformée de Haar
Descripteur par couleurs dominantes	Représentation d'une image par un nombre relativement réduit (maximum 8) de couleurs dites dominantes, obtenues par algorithmes de <i>clustering</i>
Descripteur couleur-structure	Histogramme de couleur enrichi d'information spatiale à l'aide d'un élément structurant
Distribution spatiale de couleur	Description de la distribution globale des couleurs dans l'image par transformée en Cosinus discrète
Descripteurs de texture	
<i>Nom</i>	<i>Principe</i>
Histogramme des orientations des contours	Classification grossière des orientations des contours en 5 catégories et construction d'un histogramme par rapport à ces classes
Texture homogène	Représentation multi-résolution par transformée de Gabor avec 6 orientations et 5 échelles

Tableau 1. *Descripteurs MPEG-7 retenus pour intégration dans INVENIO.*

L'ensemble des descripteurs retenus sont formalisés à l'aide de représentations aussi bien textuelles (à base de XML Schema) que binaires (pour des raisons de compacité et de transmission). En outre, ils sont munis de mesures de similarité adaptées, à base de distance dans les espaces de représentation associés. Cela est indispensable pour pouvoir effectuer des requêtes par similarité.

Pour une description plus approfondie des descripteurs, avec définition mathématique, mesures de similarité associées et propriétés, le lecteur est invité à se reporter à (Auteur *et al.*, 2004, Auteur *et al.*, 2007).

3.2. Fonctionnalités supportées

Etant essentiellement destiné à des graphistes, l'intégration dans un environnement graphique et ergonomique est déterminante pour son adoption par les utilisateurs. Afin de répondre à ces objectifs, INVENIO propose les fonctionnalités suivantes :

– la requête est faite sous la forme de la recherche par l'exemple. Une image entière ou une région d'une image est tout d'abord spécifiée. La requête en fonction des critères de recherche de couleur/texture spécifiés par l'utilisateur est ensuite effectuée.

– un deuxième mode de requête, cette fois par mots clés sélectionnés à partir d'une liste représentative des textures :

- Matières : métal, bois, pierre, verre, ciel, nuage, liquide, végétale

- Motifs : damier, brique, hexagonal, radial, cubique, cellulaire, graduée, léopard, oignon, matelassé, aplat

- Régularité : ordonnée, désordonnée

- Formes : ronde, ovale, rectangle, carré, autres

- Fréquences : radiale, ondulation, vaguelette

- Couleurs : vert, bleu, jaune, orange, violet

- Types : animale, végétale, humain, objet, paysage

Cette liste permet de constituer une ontologie de type « production », liée à un environnement métier. Les utilisateurs indexent eux-mêmes les images au fur et à mesure de leur recherche. Ainsi cette indexation est réutilisable par l'ensemble de la production et garantit une indexation textuelle à minima pour le futur.

– un système de description personnelle « tags » que l'on peut ajouter permet de représenter l'image de façon personnelle. Lors d'une production, les tags en « langage utilisateur » (*i.e.* annotation en texte libre) permettent d'indexer une image en vue de créer un « dictionnaire terminologique » lié à un projet.

Outre ces éléments de requêtes et d'annotation par descripteurs individuels, le système permet également d'hybrider multiples critères de recherche et de description. La combinaison des descripteurs permet notamment de prendre en compte la polysémie de l'image en croisant un mode recherche lié à la structure de l'image (descripteur de couleurs) et un autre lié à un niveau de représentation visuelle (mots-clés et tags). Ainsi ce formalisme descriptif combine différents niveaux de signification de l'image, sa représentation mathématique et sa représentation visuelle. Cette méthode permet donc d'établir des relations entre les différentes composantes d'une description de l'image, cette sémantique donne du sens à partir d'un ensemble de données a priori variés. L'interface déployée offre à l'utilisateur une relation efficace entre une indexation en langage naturel, et un langage de requête structurel afin d'organiser une représentation des modes d'indexation à partir de critères multiformes.

3.1 Interface et ergonomie

La figure 1 illustre la présentation graphique adoptée par INVENIO.

Les différents éléments constitutifs de l'interface, énumérés de 1 à 7 dans la Figure 1 sont explicités ci-dessous :

1 : La représentation graphique de la base de données est symbolisée par une spirale 3D déroulante d'images qui donne à l'utilisateur l'impression de profondeur

et qui permet la navigation dans la base. À chaque requête, les images de la base sont triées et présentées par ordre décroissant de similarité avec l'image requête.

2 : Cadre de sélection de l'image sur laquelle on effectue la requête.

3 : Cadre d'affichage/mémorisation de l'image sélectionnée

4 : Barre de navigation dans la spirale des résultats

5 : Liste des critères de recherche structurels : descripteurs de couleurs et de texture (cf. 3.1). Le menu utilisateur permet d'indexer l'image à partir de la liste des mots clés : matières, motifs, régularité, formes, fréquences, couleurs, types (cf. 3.2), et/ou des tags.

6 : Menu de sélections des serveurs de calcul : les requêtes peuvent être lancées sur plusieurs serveurs en parallèle afin d'optimiser le temps de réponse aux requêtes. La première requête est prioritaire, la deuxième s'effectue en tâche de fond.

7 : Données relatives aux formats du fichier image utilisé et métadonnées textuelles.

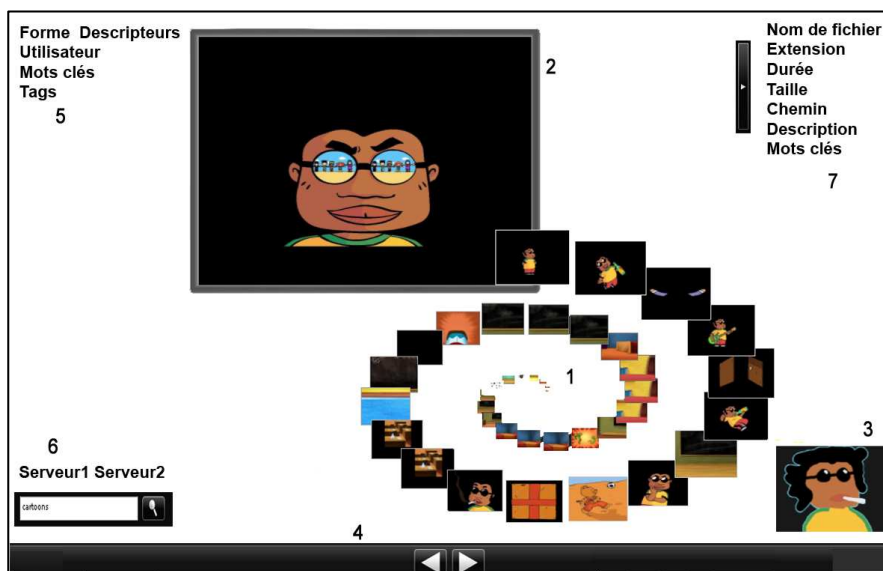


Figure 1. L'interface utilisateur de la plateforme INVENIO

Notons enfin que, pour augmenter davantage la sensation d'immersion 3D dans une base d'images, et une interaction plus directe l'utilisation d'un écran tactile est envisagée.

4. Evaluation expérimentale objective

Pour évaluer notre approche, nous avons exploité un corpus de deux bases de données. La première contient 2414 images naturelles produites par la société Mikros Image lors du tournage du film Faubourg 36. La taille de ces images est de

3008 x 2000 pixels. Elles représentent dans leur intégralité des scènes urbaines, avec prises de vue de nuit et de jour, aériennes ou non. La seconde base provient de dessins animés et inclue 4426 images de synthèse (de taille 1024 x 819 pixels pour 90% d'entre elles) produites par la société 2minutes pour différents films d'animation (Yakari, Atomic Betty, Cajou, Step by Step).

Pour évaluer objectivement les descripteurs visuels pour la recherche par le contenu nous avons constitué une vérité-terrain. Pour cela, nous avons construit pour chaque base de données un ensemble de 50 images de test, pour lesquelles nous avons identifié manuellement, avec la participation des experts et artistes, l'ensemble d'images similaires existant dans la base.

Les critères d'évaluation des requêtes par similarités sont les scores FT (*First Tier*) et BE (*Bull Eye ou ST - Second Tier*) (Osada *et. al.*, 2000), déjà utilisés pour l'évaluation des descripteurs MPEG-7 lors du développement de la norme. Les scores FT et BE obtenus, ainsi que les temps de calcul associés (calcul des descripteurs et mesure de similarité) sont résumés dans le Tableau 2.

Descripteurs	Score FT Mikros/2minutes	Score BE Mikros/2minutes	Temps d'extraction Mikros/2minutes (s)	Temps de calcul de la mesure de similarité (s)
Couleur dominante	51.53 / 57.21	63.86 / 72.58	42.12 / 10.26	0.12 / 0.35
Couleur-structure	63.21 / 54.72	71.84 / 67.63	1.92 / 0.51	0.06 / 0.23
Couleur scalable	59.24 / 56.30	68.69 / 65.11	1.87 / 0.42	0.07 / 0.25
Distribution spatiale	52.79 / 40.62	64.21 / 53.41	0.76 / 0.19	0.06 / 0.18
Histogramme d'orientations	48.58 / 28.93	58.43 / 32.09	0.32 / 0.07	0.09 / 0.25
Texture homogène	31.49 / 20.52	39.87 / 25.29	0.2 / 0.18	0.2 / 0.48

Tableau 2. Performances des requêtes par similarité (bases Mikros et 2minutes).

Ces résultats montrent que les descripteurs les plus prometteurs en termes de pertinence des requêtes sont ceux par couleur dominante (avec le désavantage d'un temps d'extraction élevé) et celui de couleur-structure, qui offre les meilleurs résultats sur les deux bases. Quant au temps de calcul des mesures de similarité, déterminant pour le temps de réponse aux requêtes, ils sont tout à fait acceptables (moins de 0.5 s sur quelques milliers d'images) pour l'ensemble des descripteurs. Validé par les artistes et les experts graphiques dans le cadre du projet HD3D-IIO, INVENIO offre ainsi un outil performant pour retrouver et réutiliser des images au sein de la chaîne de production.

6. Conclusion et perspectives

La plateforme INVENIO intègre, dans une conception graphique innovante, outils d'annotation et de recherche d'images à partir de critères de description d'ordre textuelle, sémantique et structurelle.

L'ensemble des outils de description structurelle est fondé sur la norme ISO MPEG-7, dont INVENIO propose une implantation efficace.

Pour valider cette plateforme, nous avons développé une application de réutilisation des contenus au sein d'une chaîne de production audiovisuelle, liée au projet CapDigital HD3D-IIO. Les expérimentations réalisées concernent différentes étapes de production, incluant des contenus aussi bien naturels que de synthèse (i.e. dessins animés). Les solutions proposées dans de cet article démontrent notamment que l'utilisation de la norme MPEG-7 à travers la plateforme d'indexation INVENIO est tout à fait pertinente pour favoriser et accélérer la réutilisation des contenus au sein d'une chaîne de production audiovisuelle.

Les perspectives de travail futur concernent tout d'abord la fusion des critères de recherche textuels et les critères de structure du contenu. Cela permettrait de combler le décalage sémantique caractéristique des approches d'indexation par descripteurs visuels, en exploitant une connaissance métier.

À plus long terme, il serait intéressant d'étendre la plateforme, pour le moment dédiée exclusivement à l'indexation d'images fixes, à d'autres types de médias, comme les vidéos et les données graphiques 3D.

Bibliographie

Auteur *et al.*, 2004

Auteur *et al.*, 2007

Manjunath B. S., Salembier P., Sikora T., *Introduction to MPEG-7 : Multimedia Content Description Interface*, John Wiley & Sons, 2002.

Gupta A., Jain R., Santini S., Smeulders A. W. M., Worring M., Content-Based Image Retrieval at the End of the Early Years, IEEE Trans. on PAMI, Volume 22, Issue 12, Décembre 2000.

Datta R., Li J., Wang J., Wang, J.Z., Content-based image retrieval: approaches and trends of the new age Proc. 7th ACM SIGMM international workshop on Multimedia information retrieval, p. 253 – 262, 2005.

XML Schema Spécification disponible sur <http://www.w3.org/XML/Schema>.

Chazelle B., Dobkin D.P., Funkhouser T., Osada R., « Matching 3D models with shape distributions », *Proc. International Conference on Shape Modeling and Applications*, Mai 2001.

ISO/IEC 15938-5, *Information technology - Multimedia Content Description Interface - Part 5: Multimedia Description Schemes*. 2003

ISO/IEC 15938-3 : 2002, MPEG-7-Visual, *Information Technology – Multimedia content description interface – Part 3: Visual*, 2002.