

Modélisation des connaissances hétérogènes des documents multimédia

Françoise Deloule*, Daniel Beauchêne*, Bogdan Ionescu**, Patrick Lambert**

*Equipe IC-Condillac, LISTIC,
Université de Savoie, 73 376 le Bourget du Lac Cedex
Francoise.Deloule@univ-savoie.fr
Daniel.Beauchene@univ-savoie.fr

**LISTIC, Université de Savoie,
Polytech'Savoie, BP 80439 74944 Annecy-le-Vieux Cedex
Bogdan.ionescu@univ-savoie.fr
Patrick.Lambert@univ-savoie.fr

Résumé. Les objets multimédia sont des supports de la connaissance qui se développent rapidement. Le défi à relever consiste à proposer des outils performants pour atteindre le contenu de ces objets et fournir une approche sémantique de ces très vastes bases de connaissances. Nous proposons une approche bipolaire. D'un point de vue externe sur ces objets multimédia, nous utilisons les ontologies terminologiques pour repérer les éléments sémantiques pertinents dans les métadonnées et péritextes. D'un point de vue interne, nous construisons des indicateurs de niveaux sémantiques croissants à partir des images. Enfin, nous proposons une fusion de ces données hétérogènes pour essayer de fournir des outils performants de recherche de contenus dans les bases d'objets multimédia. Nous illustrons nos propos par un corpus de films d'animation.

1 Introduction

Si on en juge par leur développement actuel, les objets multimédia et la vidéo en particulier sont sans aucun doute les supports du futur pour la circulation, le partage et l'élaboration des connaissances. Ce mouvement ne touche pas seulement le domaine du divertissement, mais tous les domaines : la culture, l'éducation, l'enseignement..., tant du point de vue des professionnels que des particuliers. La profusion de telles sources d'informations et de connaissances nécessite la création d'outils performants, adaptés à ces nouvelles caractéristiques, pour gérer et rendre accessible au plus grand nombre le contenu de ces bases.

Mais les attentes des utilisateurs ne se résument pas à un accès à plus d'informations et de connaissances, sous les diverses formes disponibles aujourd'hui. Elles portent surtout sur une demande de services plus élaborés, pour une véritable interaction avec le contenu véhiculé par ces documents (filtrage d'information, sélection ciblée, réponse appropriée à une demande contextualisée, résumé automatique, ...).

Connaissances hétérogènes des documents multimédia

Les travaux développés en gestion des connaissances (Dieng et al. 2005) (Teulier 2005) et en sciences cognitives (Desclés 1990), (Petitot 1990), (Ganascia 2006) apportent une réponse partielle à ces attentes lorsqu'il s'agit de documents numérisés. Ainsi nous savons modéliser une partie des connaissances issues des textes et les travaux menés en traitement d'image ou en traitement du son permettent de trouver un certain nombre d'informations dans ces documents multimédia. Cependant, si nous savons tirer des textes des informations de haut niveau sémantique, les techniques d'extraction de caractéristiques des images et du son sont loin de nous fournir des éléments d'une qualité sémantique identique.

Il nous faut donc proposer des moyens d'enrichir la sémantique des indicateurs fournis par le son et l'image mais aussi définir, enrichir et exploiter les coopérations et corrélations entre ces connaissances d'origines diversifiées. Nous proposons d'aborder ce problème selon deux points de vue différents. A partir d'une vision "externe" de nos objets, des informations textuelles disponibles, des méta données et des péritextes, nous travaillerons sur une modélisation symbolique aboutissant à la construction d'une ontologie terminologique. A partir d'une vision "interne" de nos objets, des informations issues des indicateurs obtenus par les analyses des images (et du son), nous nous intéresserons à une modélisation qualitative des films, pour définir des éléments comme une ambiance ou un rythme à l'aide de descripteurs ayant un contenu sémantique. Enfin, nous proposons de combiner ces deux approches pour voir quels sont les points de convergence et les enrichissements mutuels possibles.

Pour illustrer nos propos, nous disposons d'un corpus fourni par CITIA, la cité de l'image en mouvement, organisatrice du Festival International du Cinéma d'Animation d'Annecy, et la société de production Folimage. Après avoir précisé le contenu du corpus (partie 2), nous présentons les processus de modélisation retenus (parties 3 et 4) et ce que nous pouvons établir comme corrélation entre ces connaissances hétérogènes (partie 5). Si certains résultats sont propres au cinéma d'animation, la démarche et la méthode sont génériques et peuvent s'appliquer aux divers objets multimédia.

2 Un corpus varié

2.1 Les bases de données

CITIA organise le Festival International du Film d'Animation et le Marché International du Film d'Animation chaque année à Annecy. CITIA est confronté à deux problématiques proches de nos préoccupations. D'une part, il possède un grand ensemble de films qu'il doit, dans une visée patrimoniale, conserver et mettre à disposition de divers professionnels: la constitution d'une base numérisée permet d'obtenir un support de sauvegarde fiable et un accès facile, des outils informatiques doivent alors être mis en place pour permettre la recherche ou la navigation dans la base constituée. D'autre part, CITIA a un objectif d'exploitation et de diffusion de ses connaissances et de ses compétences: l'utilisation et l'exploitation de cette nouvelle base multimédia par la construction d'outils de caractérisation et d'analyse des films présentent un intérêt fondamental tant pour les professionnels, les cinémathèques, les enseignants que pour le grand public.

Actuellement, CITIA peut mettre à notre disposition une importante base de données ainsi que les films eux-mêmes, sur des supports divers. La base de données contient des informations sur près de 24 000 films d'animation et 55 000 professionnels du secteur, complétées par des milliers de photographies, des dossiers documentaires, livres et magazines. Elle

référence 15 000 films sur divers supports et s'enrichit chaque année à l'occasion du festival de près de 1 500 nouveaux titres. Ces films sont en principe décrits par :

- le titre de l'œuvre (dans la langue originale, en français et en anglais),
- la nationalité et l'année de production,
- les noms des « auteurs » : scénario, graphismes, sons, etc...
- des indications sur les techniques utilisées, la durée, le support,
- la catégorie de compétition choisie pour ce festival,
- le genre du film, l'âge du public visé,
- un synopsis en français et en anglais...

Ces données sont assez complètes pour les films qui ont été présentés en compétition à l'un des festivals, beaucoup moins pour d'autres films. Les indications sur les techniques utilisées et les synopsis ont été le plus souvent fournies par ceux qui ont inscrit le film au festival, qui ne sont pas toujours les auteurs. Et parfois ce sont même les personnes qui doivent saisir les données pour remplir la base qui ont "improvisé" pour remplir les champs vides: par exemple, le texte du champ *synopsis* peut selon les cas, contenir un résumé du film, une accroche publicitaire ou une critique. La qualité des traductions aussi est variable selon les périodes de saisie. Il faut préciser qu'un grand nombre de fiches est traité en un court laps de temps au moment du festival, par plusieurs personnes d'horizons culturels différents, parfois simples bénévoles. CITIA dispose de peu de personnel permanent. Une autre particularité de ce domaine est que les auteurs sont là pour revendiquer leur créativité et ont parfois du mal à se satisfaire des catégories proposées lors de l'inscription.

Dans l'état actuel de la base, concernant les films présentés au festival, certaines fiches descriptives sont accessibles sur le site du CITIA (www.annecy.org) à la rubrique Animaquid. La présence d'une image "représentative" sur une fiche nous indique que le film physique est disponible à CITIA. Ces données partielles sont complétées par des éléments textuels différemment accessibles, comme les dossiers documentaires, les livres ou magazines du domaine, les critiques de journaux, les commentaires ou légendes de photos, ou encore des images extraites du film.

Concernant les films eux-mêmes, CITIA constitue une base numérique à partir des différents supports, notamment dans un but de sauvegarde indispensable. Ce transfert technique est relativement long et coûteux à effectuer. Les films sont transférés au format mpeg2. Les images ont un format de 704x326 pixels, chaque pixel étant codé sur 24 bits. Ces films, bien qu'encore peu nombreux par rapport à l'existant (environ 1500 pour le moment), couvrent une large palette de techniques et de genres. Ils serviront de supports à nos travaux sur les images. Cette base se complète des films plus récents produits directement en format numérique.

Nous précisons ici que nous ne cherchons pas à extraire les textes ou discours inclus dans ces documents, mais des éléments complémentaires qu'apportent les média audio et visuels. Les informations que l'on peut chercher à extraire des flux vidéo se situent à deux niveaux : au niveau des images et au niveau du son, de manière complémentaire, synchronisée ou non. Pour des raisons de moyens disponibles, nous nous sommes actuellement concentrés sur l'analyse des images. Le son, qui devra être exploité ultérieurement, permettrait d'identifier des instants particuliers tels que applaudissements, explosion, cri, présence de parole et /ou de musique, d'acteurs, caractéristiques du ton ou du style d'un dialogue... La problématique

Connaissances hétérogènes des documents multimédia

du son nous semble assez proche de celle soulevée par les images, à savoir le faible niveau de la sémantique des informations et des résultats des analyses directement disponibles.

La particularité de ce corpus réside dans le fait que les films d'animation présentent des spécificités liées aux techniques d'animation utilisées, aux palettes couleurs employées ou à l'organisation de l'enchaînement des plans... Nous pensons donc que les caractéristiques extraites des images vont nous fournir des éléments intéressants pour une recherche basée sur les contenus dans la base des films et documents ainsi constituée.

2.2 Intervention des experts

A côté de ce corpus important, nous disposons d'accès à la bibliothèque spécialisée du CITIA et à d'autres bibliothèques. Mais surtout, nous avons une ressource indispensable: des experts. En effet, la réussite d'un tel projet repose sur la collaboration de spécialistes de différents domaines: le domaine cible, le cinéma d'animation, les domaines ressources que sont les domaines de la gestion des connaissances (terminologie, ontologie, ingénierie des connaissances), de l'analyse d'image et de la gestion de projet collaboratifs. Dans la suite du texte, le terme expert fait référence aux experts en cinéma d'animation.

Ces experts interviennent à tous les niveaux: représentativité des échantillons du corpus retenus, définitions des éléments textuels et usages de ces définitions, choix des termes appropriés, connaissances des techniques, des créateurs et des auteurs, des sensibilités, des documents de référence, validation de l'ontologie, suggestions d'utilisations futures Leurs interventions sont indispensables à la construction d'une base de connaissance et d'outils d'exploitation efficaces. Ils apportent les connaissances du métier et leurs compétences, ils pourront légitimement valider nos propositions.

3 Modélisation des connaissances issues des textes

Nos objectifs ici sont de faciliter la mémorisation et l'exploitation de ces grandes bases multimédia en perpétuelle évolution, partiellement inexploitable car les données sont parfois incomplètes et souvent de qualité hétérogène. Certaines informations, comme les mots-clés, n'ont parfois été créées que dans une fin d'indexation, sans rapport direct avec le contenu réel du document, parfois empreintes d'une forte subjectivité. Les recherches basées sur des mots clés prédéfinis ou sur leurs combinaisons sont souvent peu satisfaisantes, fournissant trop ou trop peu de résultats et fréquemment des résultats inadaptés et décevants. D'une façon générale, les outils associés aux bases de données s'avèrent largement insuffisants pour faire face à l'explosion du nombre de documents disponibles.

En abordant la gestion des objets multimédia d'un point de vue externe, en exploitant seulement les données textuelles et les métadonnées *sur* ces objets, nous pouvons extraire bon nombre de connaissances pertinentes sur le domaine. Nous pourrions organiser ces connaissances et les confronter aux connaissances extraites des objets eux-mêmes au travers de l'analyse des images d'un film.

3.1 Extraction de connaissances et terminologie

L'information textuelle restant la plus abondante et la plus outillée, il est donc primordial de l'exploiter au mieux pour en extraire des connaissances. Dans cette partie, nous nous appuyons sur le contenu des textes *externes* aux films (péritextes) nous donnant des informations sur leurs caractéristiques internes et sur toutes les métadonnées disponibles. Bien que ces données soient hétérogènes en terme de présence et de qualité, d'interprétation, de richesse de la langue utilisée, de longueur de description, d'originalité sur les désignations, nous cherchons à exprimer la sémantique associée à ces écrits. Pour cela, nous choisissons de combiner une approche terminologique (Slodzian 1995), (Bourigault, Slodzian 1999) et une approche ontologique (Gruber 1995), (Bachimont 2000) en nous appuyant sur les travaux réalisés dans ce domaine.

Afin de comprendre les métiers concernés et le domaine du cinéma d'animation, il nous faut clarifier et "normaliser" le vocabulaire rencontré dans tous ces écrits, et pour cela, définir un **langage métier** (Rey 1992), (Depecker 2002). Nous voulons construire une terminologie métier consensuelle, cohérente au sens de la consistance logique de l'ensemble des significations, partageable et réutilisable. Or différents intervenants ont déjà créé une certaine hétérogénéité dans les informations et connaissances exprimées, utilisant des vocabulaires différemment interprétés ou interprétables. Donc, pour atteindre cet objectif, dans l'ensemble des textes dont nous disposons, nous cherchons à identifier le vocabulaire utilisé spécifiquement par les métiers concernés. Nous pouvons l'identifier par diverses méthodes statistiques ou d'informatique linguistique. Nous ne recherchons pas seulement des termes isolés, mais aussi les mots composés ou les expressions propres au métier, dont la pertinence pourra être signalée ou validée par les experts. Nous obtenons ce que nous nommons les candidats-termes. Puis à l'aide des connaissances complémentaires (définitions, compréhension dans un contexte, illustration par des images, bibliographie...) et l'intervention des experts, nous allons pouvoir les regrouper en catégories (termes synonymes, terme non spécifique, terme reflétant un concept ou une propriété du domaine...) et confronter les usages en fonction des profils des différents acteurs (étudiants, festivaliers, intervenants CITIA ...).

Si les termes extraits par les différents outils informatiques des données textuelles apportent une aide significative à la construction de la terminologie métier, la pertinence de celle-ci repose essentiellement sur l'interaction experts - termes.

3.2 Conceptualisation et ontologie

Au delà des mots, nous souhaitons faire émerger des concepts pour modéliser et exploiter ces connaissances, et étudier les relations qui peuvent les associer. Un concept représente une pluralité d'objets ayant des propriétés et un comportement communs. Pour être utilisé, il doit bien sûr être nommé, désigné par un terme. Par contre, il peut être défini en dehors du contexte d'utilisation de chacun des objets qu'il représente. Ainsi, nous définissons des bases de connaissances terminologiques (Meyer et al. 1992) qui sont constituées :

- des textes, sources d'informations qui caractérisent le sens des termes, fournissant des significations en contexte,
- des termes, mots ou expressions du langage du métier,
- des concepts dénotés par ces termes, ayant leur propre signification, signification hors contexte, donc indépendante de l'expert.

Connaissances hétérogènes des documents multimédia

Dans le cadre d'une terminologie ontologique (Roche 2005), la signification d'un mot est définie comme étant la connaissance décontextualisée qu'il représente, alors que le sens est une signification actualisée en fonction du contexte et de la situation. Cette connaissance dépend directement de considérations épistémologiques du modèle ontologique, qui lui n'appartient plus au domaine linguistique. Nous avons retenu le modèle ontologique *OK* pour Ontological Knowledge (Roche 2001). Ce choix repose notamment sur le fait que la construction de cette ontologie nous amène à compléter les définitions des termes, à éliciter les connaissances profondes du domaine, à conforter la définition des concepts et à les organiser de façon rigoureuse, assurant ainsi une exploitation ultérieure pertinente basée sur la sémantique. En contrepartie de ces avantages, sa construction finale n'est pas aisée et repose sur une organisation des concepts favorisant certaines relations.

Le modèle *OK* s'appuie sur des principes pluridisciplinaires comme l'épistémologie (distinction de l'essence des choses de ce qui les décrit), la linguistique (le signe est arbitraire, la langue est un système qui structure les mots), l'intelligence artificielle (impact du langage de représentation sur la modélisation) et de la logique (sémantique logique et ensembliste des connaissances, formalisation *a posteriori*).

La construction d'une ontologie *OK* repose sur les principes suivants :

Un domaine est décrit par plusieurs catégories qui peuvent ensuite être combinées pour les différentes utilisations du système. Par exemple pour le cinéma d'animation les techniques, les supports, les outils ou les publics forment autant de catégories distinctes.

Une catégorie regroupe les termes dont les définitions sont liées sémantiquement. Par exemple dans la catégorie des techniques nous trouverons les termes découpage, ombres chinoises, dessin sur cellulo, ...

Dans chaque catégorie, les termes sont répartis selon leur nature, ce qui permet de définir des concepts, des différences, des attributs et de prendre en compte certaines relations conceptuelles. Chaque concept d'une catégorie est défini par rapport au précédent en terme de *sorte-de*, relation de subsumption (par exemple dans la catégorie des techniques, les <ombres chinoises> sont une *sorte de* <découpage>), puis d'un « complément » précisant ce qui le différencie du précédent. Ce "complément" constitue l'atome de signification élémentaire à partir duquel se construisent et se divisent les concepts, il porte sur l'**essence** des choses. De ce fait un concept possède ou ne possède pas une caractéristique essentielle (séparation binaire), et prend place dans une arborescence, sur laquelle nous pourrions appliquer des éléments de logique (formalisation *a posteriori*, sémantique logique et ensembliste des connaissances). Par exemple, les <ombres chinoises> se différencient des <éléments découpés> par le mode d'éclairage. L'attribut enfin permet de décrire l'état du concept et non son essence. Par exemple le matériau permettant de construire les ombres chinoises est un attribut qui prendra les valeurs *papier*, *être vivant*, *pâte à modeler*, Nous regroupons l'ensemble des indications de ces caractéristiques essentielles et nous regardons, pour chaque concept, si une caractéristique a un sens et s'il la possède. Nous pouvons ainsi affiner les définitions des concepts par la mise en lumière de différences essentielles puis définir les domaines de valeur des attributs. Enfin, nous pouvons construire l'ontologie *OK* sous sa forme sous-jacente d'arbre de Porphyre, un arbre binaire permutable... La terminologie ontologique indique ainsi le(s) rôle(s) joué(s) par chaque terme dans le métier considéré.

Cette ontologie, support et résultat de la conceptualisation du domaine, est soumise aux experts pour validation. Dans le cadre de l'étude menée à CITIA, des retours vers les bases d'informations et parfois vers les films ont été nécessaires pour définir les termes dénotant les concepts, repérer les synonymes et obtenir un consensus.

Enfin, l'ontologie construite à partir d'un corpus représentatif mais restreint doit être validée sur l'ensemble du corpus, et recueillir le consensus des experts, en prenant en compte les diverses utilisations possibles. Lors de la prise en compte du corpus élargi, certains éléments "nouveaux" doivent être étudiés, pour déterminer s'il y a lieu de créer un nouveau concept et/ou de nouvelles différences spécifiques, ou s'il s'agit d'un même concept ou d'une même différence désigné par un terme synonyme. La terminologie (NFISO 2001) et l'ontologie sont alors mutuellement complétées, les relations de synonymie ou de traduction peuvent être enrichies. L'ontologie complète est constituée des différents arbres de Porphyre obtenu pour chaque catégorie de concepts. La figure 1 nous permet d'illustrer nos propos avec un petit extrait de la catégorie des techniques¹. La catégorie des techniques permet de décrire les ombres chinoises comme étant une technique générique, construite image par image (et non une technique numérique), utilisant le dessin, c.à.d. une technique à plat et non en volume (ces éléments sont donnés par la partie gauche de l'ontologie non visible sur la figure). Cette technique de dessin est utilisée en direct-sous-la-caméra. Elle nécessite du découpage qui doit être éclairé par-dessous. Cette technique d'ombres chinoises se différencie en "fin" de définition de l'utilisation des poudres par le fait que l'exécution en direct-sous-la-caméra est une composition nécessitant des particules. L'attribut de "poudre" permet de préciser s'il s'agit de sable ou de café. La technique des écrans d'épingles est aussi décrite par un départ commun, et se différencie de la technique des poudres par l'utilisation d'éléments qui ne sont pas des particules et nécessite la manipulation d'épingles. Le concept *Anonymous3* est un concept non manipulé dans l'application et sa dénomination n'a alors ni nécessité ni sens ici.

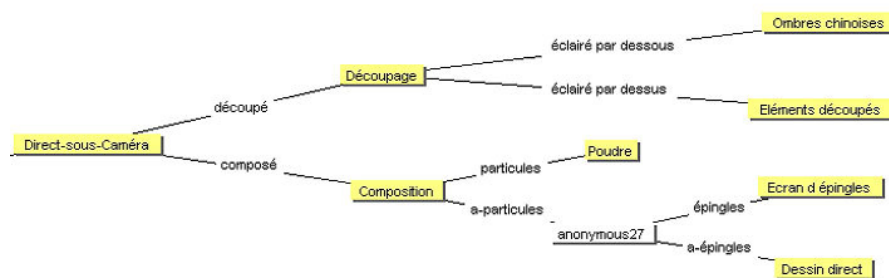


Fig 1: Un extrait d'ontologie, catégorie des techniques.

Cet extrait décrit les techniques d'animation réalisées directement sous la caméra. Les éléments utilisés peuvent être découpés (concept de découpage) ou composés à partir d'éléments qui peuvent être des poudres, des épingles, voire un dessin réalisé directement sous la caméra.

¹ Ontologie et terminologie sont construites avec les outils OCW et TCW de Ontologos Corp (www.ontologos-corp.com).

4 Modélisation des données "images"

Le corpus disponible est aussi constitué de l'ensemble des objets multimédia eux mêmes. Nous avons déjà utilisé partiellement cet aspect en visualisant certains films pour valider ou compléter l'ontologie proposée. Une partie de ces films a été numérisée et nous voulons compléter les connaissances acquises précédemment en exploitant les autres média, par une approche "interne" de nos objets.

Lorsque nous regardons un film, nous percevons et analysons naturellement un grand nombre d'informations. Un des objectif des objets multimédia, au travers des perceptions auditives et visuelles, est de faire appel à nos capacités sensorielles et sentimentales. Un document multi média n'est pas une succession innocente de mots, d'images et de sons, mais un "savant" agencement de tous ces éléments pour capter l'attention du destinataire, du spectateur et provoquer ses émotions. En général, ce récepteur reconnaît des "formes", un style, peut déterminer une "ambiance", un ressenti sur un film... Actuellement, ces éléments perçus ne sont pris en compte que s'ils sont décrits dans les péritextes. Alors que l'analyse intuitive, lorsqu'elle est guidée par des questions a priori, permet notamment de valider les concepts décrits précédemment, d'identifier les techniques utilisées, de nommer le genre du film, de classer un film dans l'ontologie précédemment définie, de compléter les éléments de catégorisation...

Cependant, perception et analyse resteront partiellement décelables pour nos systèmes, et leur expression risque d'être empreinte de la subjectivité de l'auteur. Sans vouloir aller jusqu'à l'étude des sentiments, ce qui relèverait d'autres disciplines, nous voulons exploiter les éléments auditifs ou visuels qui contribuent plus objectivement à créer ces émotions. Ainsi, nous souhaitons étudier des éléments qualitatifs dont le rôle peut apporter des critères intéressants pour l'exploitation des bases de données multimédia.

Pour cela, il faut plutôt se tourner vers les théories des formes, les théories de la singularité et des bifurcations (Petitot 1990). Elles nous permettent de faire des associations ou des disjonctions sur des éléments "réels" et non plus sur des symboles, et au travers de modélisations qualitatives (Chui et Kuipers 1992), (Deloule et al. 1990), (Tomasena et al. 1990) , de faire émerger des connaissances et des concepts différents. Pour prendre en compte l'imprécision inhérente aux informations manipulées, nous utilisons des représentations floues (Zadeh 1965). Dans l'état actuel des choses, le traitement d'images peut fournir un nombre important d'indicateurs numériques, de faible niveau sémantique. Ces indicateurs sont très utilisés pour l'indexation des images fixes. Pour ce qui concerne les films, et donc une grande quantité d'images en mouvement, la masse des résultats fournis devient difficilement exploitable. Ceci nous donne un autre argument pour nous orienter vers une analyse qualitative des images et pour définir des descripteurs d'un niveau sémantique plus riche.

Les éléments recherchés dans les images (ou le son), excluant volontairement les éléments textuels, sont donc clairement explicités. Actuellement, nous centrons nos travaux sur la caractérisation de l'ambiance des films, au travers d'éléments tels que le rythme ou la distribution des couleurs. Ces travaux sont dirigés par le fait que nous souhaitons établir par la suite des corrélations avec l'ontologie construite à partir des textes, notamment étudier les corrélations éventuelles entre ambiance et technique ou entre ambiance et genre d'un film.

4.1 Indicateurs issus des images

Dans le domaine de l'analyse des images, la thématique appelée « indexation » mobilise fortement les chercheurs du domaine depuis bientôt plus de quinze ans (Del Bimbo 1999), (Santini 2001), (Smeulders et al. 2001). L'objectif est d'arriver à extraire de chaque image un ensemble de descripteurs (ou index) permettant la caractérisation de l'image. Ces descripteurs sont ensuite utilisés pour mettre en place des mécanismes de recherche ou de navigation dans de grandes bases d'images. Le passage de l'image fixe à la vidéo est plus récent (Snoek et al. 2005), (Huet et al. 2003), et rendu plus complexe par l'augmentation considérable de la masse de données à traiter et par la prise en compte de l'information de mouvement. La norme MPEG7 (Mpeg7 2004) définit une syntaxe de description et de codage des objets multimédia. Mais la création des descriptions, c'est-à-dire l'extraction des caractéristiques des documents et l'indexation, ainsi que l'utilisation des descriptions (moteur de recherche, navigation,...) sont exclues du champ de la normalisation. Or c'est justement là l'un des objectifs que nous poursuivons.

Ici, nous souhaitons enrichir ou spécialiser les outils de gestion de nos bases de connaissances, construits à partir des données textuelles, par les informations contenues dans les images des films. Cependant, nous devons faire face à deux difficultés. D'une part, les techniques d'analyses d'images fournissent des caractéristiques de « bas niveau », souvent numériques, qu'il est difficile de traduire en descripteurs de nature sémantique. On parle alors de « fossé sémantique ». Ainsi, on sent bien que la connaissance de la distribution des couleurs utilisées dans un film présente un lien avec l'atmosphère se dégageant de ce film, mais ce lien est difficile à formaliser. D'autre part, les descripteurs issus des images sont en général définis sur une autre base terminologique que celle issue de l'analyse des textes. L'analyse des images pourra fournir des résultats du type : film à dominante rouge, ou présentant une grande variété de couleurs, etc., caractérisations qu'il n'est pas simple de relier au genre « comédie » par exemple.

Pour atteindre nos objectifs, nous procédons par étapes. Les deux premières étapes sont directement liées aux traitements des images, pour en retirer des informations et des indicateurs spécifiques, avec l'approche définie précédemment. La troisième étape se constitue d'une exploitation des résultats précédents pour caractériser les films.

La première étape consiste à exploiter les éléments structurels d'un film, notamment à retrouver le découpage en plans successifs, par segmentation temporelle. La technique utilisée repose sur la recherche des changements de plan. Il existe différents types de changements de plan :

- les changements de plan simples (*cuts*). Le passage d'un plan à l'autre se fait sur deux images. On utilise une mesure de différence entre images voisines pour détecter ces changements. Ce principe est très classique : un des axes du projet TREC (<http://trec.nist.gov/>) a pour objectif la comparaison des techniques de segmentation temporelle. Du fait de la particularité des films d'animation, nous avons développé une approche spécifique basée sur la différence entre histogrammes couleurs calculés sur une palette réduite. La différence est ensuite comparée à un seuil, calculé automatiquement (Ionescu et al. 2006). Pour ne pas être trop sensible aux déplacements à l'intérieur des images, quatre histogrammes, correspondant aux quatre cadrans de l'image, sont utilisés. Une décision majoritaire est alors prise.

Connaissances hétérogènes des documents multimédia

- les changements de plan par effets spéciaux. Dans les films d'animation, on distingue principalement 3 types d'effets spéciaux :
 - les transitions lentes de type *fade*. Le passage d'un plan à l'autre se fait par transformation progressive de l'image vers une image uniforme (noire ou constante) – *fade out* – puis réapparition progressive de l'image suivante par le mécanisme inverse – *fade in* -. Ce type de transition est détecté en recherchant des décroissances (ou des croissances) au cours du temps de la moyenne ou de la variance de l'intensité ou de la chrominance des images.
 - Les changements brefs de couleurs. Spécifiques aux films d'animation, c'est un effet visuel qui se caractérise par un brusque changement de couleur à l'intérieur d'un même plan, sur une durée très courte (quelques images). La recherche de changements de plan simples détecte deux transitions, la similarité des images situées avant et après ces deux transitions permet de détecter cet effet.
 - Les fondus enchaînés. Ces transitions sont les plus difficiles à détecter. La technique utilisée repose sur l'analyse de l'amplitude des contours (Lienhart 2001). Un fondu enchaîné se traduit par une décroissance lente de l'amplitude des contours de la dernière image du plan suivi d'une croissance lente de l'amplitude des contours de la première image du plan suivant.

Les résultats obtenus, pour lesquels nous avons aussi fait un travail manuel de confirmation, nous donne entre 80 et 90% de bons résultats. Ce pourcentage de détection des changements de plan est satisfaisant pour poursuivre nos travaux.

La deuxième étape consiste à construire une représentation compacte de la distribution des couleurs de chaque film. La stratégie employée comporte trois phases :

- sélection de $p\%$ images de chacun des plans fournis par la segmentation temporelle. Cette étape permet un gain de temps en réduisant le nombre d'images analysées. Expérimentalement, nous avons pu montrer qu'une valeur de $p\%$ de l'ordre de 15% permet une réduction importante sans modification sensible du résultat de la caractérisation des couleurs.
- réduction des couleurs. Le codage initial des couleurs est effectué sur plus de 16 millions de niveaux. La réduction des couleurs reste indispensable pour permettre des traitements raisonnables. Elle joue un rôle très important pour le reste de l'analyse. L'objectif est dans un premier temps de définir une palette réduite suffisamment représentative pour permettre une caractérisation pertinente. La réduction peut être adaptée à chaque film ou fixe et commune à tous les films. La solution adaptée à chaque film donne en général des résultats de meilleure qualité, mais complique les tâches de comparaison des films entre eux. Nous avons opté pour une réduction fixe en utilisant la palette « webmaster » composée de 216 couleurs (Visibone 2006). Cette palette présente l'avantage de spécifier chacune des 216 couleurs par des noms caractérisant l'intensité, la teinte et la saturation des couleurs. Dans un deuxième temps, chaque couleur de chaque pixel est remplacée par la couleur de la palette la plus proche, au sens de la distance euclidienne, dans l'espace CIELab. Un mécanisme de diffusion d'erreur (Wurd 2004) complète cette phase qui fournit des

images perceptuellement très proches des images initiales, les différences étant quasi imperceptibles à l'oeil.

- la dernière phase consiste à calculer deux histogrammes : un histogramme global et un histogramme des couleurs élémentaires. L'histogramme global donne la distribution, pour les $p\%$ images sélectionnées, des couleurs du film sur la palette « web-master », et l'histogramme des couleurs élémentaires, basé sur la roue des couleurs d'Itten (Itten, 1961), donne la distribution des couleurs projetée sur 15 couleurs élémentaires ("Orange", "Red", "Pink", "Magenta", "Violet", "Blue", "Azure", "Cyan", "Teal", "Green", "Spring", "Yellow", "Gray", "White", "Black").

La figure suivante donne les deux histogrammes obtenus sur le film « le trop petit prince ».

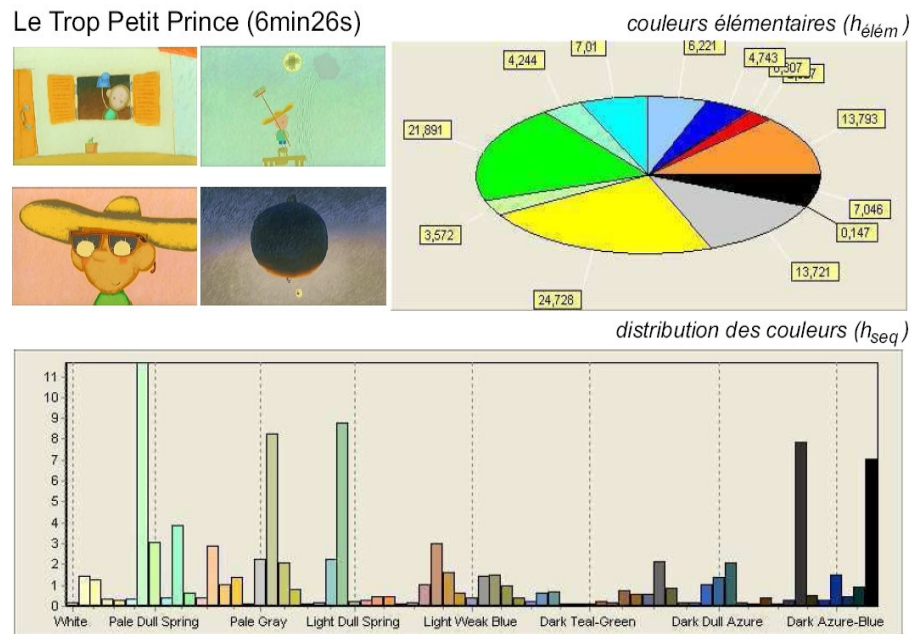


Fig. 2 : Film « Le Trop Petit Prince »

Histogramme des couleurs élémentaires ($h_{\text{élé}}^m$) où l'on note une prédominance du jaune, du vert, du gris et de l'orange – Histogramme global des couleurs où ne sont présentées que les couleurs dont la fréquence est supérieure à 4%.

La troisième étape consiste à exploiter les résultats des analyses d'images précédentes pour créer des caractérisations des films étudiés. Nous cherchons à définir des descripteurs plus riches sémantiquement que ces valeurs numériques obtenues, afin de réduire le "fossé sémantique" séparant ces indicateurs des concepts issus des pértextes.

4.2 La caractérisation des films

A partir des résultats et des nouvelles possibilités d'analyses des deux étapes précédentes, nous proposons de bâtir une caractérisation des films dans trois directions :

- des caractérisations liées à la segmentation temporelle du film (RYTHME), reprenant la répartition des changements de plan dans le film,
- des caractérisations liées aux couleurs (COULEURS) étudiant la distribution des principales couleurs utilisées dans le film, et enfin
- des caractérisations liées aux effets spéciaux (EFFETS) notamment obtenu par la texture, repérée par un traitement de type « image statique », ou le fondu-enchaîné.

Ces pistes sont exploratoires et donc non exhaustives quant aux caractéristiques à définir et aux éléments intéressants. Nous chercherons d'une part à augmenter le nombre de descripteurs, d'autre part à augmenter leur sémantique dans les travaux en cours et futurs.

Nous proposons de désigner ces caractérisations par des **termes** choisis, et de les qualifier par une échelle de **valeurs** associées. Pour une lecture plus simple, nous gardons dans ce paragraphe les formes du texte associées à ces définitions. Les **termes** et **échelles de valeurs** pourront être modifiés en fonction des consensus obtenus. Les termes peuvent avoir des synonymes et les échelles de valeurs des correspondances, afin de prendre en compte la richesse des résultats proposés par les différents outils existants.

La segmentation du film, donnant la densité et les positions relatives des transitions, nous fournit des indicateurs simples qui donnent une bonne idée du rythme d'une séquence d'images. Ces valeurs doivent bien sûr être prises relativement à la durée du film. Nous pouvons définir des indicateurs numériques du rythme de la séquence tels que n_t , nombre de transitions à l'instant t et δ_t , écart moyen entre transitions à l'instant t , ces deux valeurs étant calculées sur une fenêtre temporelle d'une minute centrée sur l'instant t , ainsi que N , moyenne des n_t et Δ = moyenne des δ_t .

Ces résultats numériques nous permettent de définir d'une part une **rapidité d'évolution** et de déterminer qualitativement ses valeurs : *lente, moyenne, alerte*, et d'autre part un rapport d'action nommé **activité** (*faible, moyen, fort*) existant dans le film.

La présence de changements rapides de couleurs (*short color changes*) peut être exploitée pour déceler une notion d'**explosivité** (valeurs: *oui, non*), notamment dans les dessins animés. Cette indication peut aussi révéler un rythme plus ou moins soutenu.

La détection de l'utilisation des effets spéciaux et notamment des fondus enchaînés ou de transitions plus ou moins lentes contribuent à donner à un film une ambiance de **mystère**, part de mystère qui peut être *inexistante* ou *faible, moyenne, élevée*. Elle peut être aussi indicative du rythme des actions, dans le registre inverse de l'explosivité.

L'étude des couleurs, par une interprétation qualitative, nous permet de définir des caractéristiques décrivant une ambiance liée aux couleurs ou teintes dominantes. A partir des histogrammes présentés précédemment, on peut définir l'**intensité**, c'est à dire la répartition des couleurs claires et foncées (*claire, contrastée, foncée*), la **saturation** précisant la répartition des couleurs faibles et fortes ou pures et atténuées (*faible, contrastée, forte, x*), la **chaleur** des couleurs dominantes (*froide, contrastée, chaude, x*), la **diversité** des couleurs élémentaires utilisées, définie sur les 15 couleurs principales, (*réduite, moyenne, élevée*), la

variété (richesse) définie sur 216 couleurs (**réduite, moyenne, élevée**), et le **contraste** utilisant la pondération relative des couleurs adjacentes et des couleurs complémentaires (**oui, non**). La valeur x permet de ne pas prendre en compte ou de ne pas déterminer cette valeur.

Enfin, une étude sur image fixe peut nous fournir des éléments intéressants pour étudier la texture des matériaux ou du rendu sur l'image obtenue. Il est possible d'isoler quelques images représentatives (par exemple le milieu des plans les plus importants) et, en utilisant les techniques d'analyses d'images statiques, d'identifier certaines techniques de réalisation, comme par exemple les techniques à base de poudre, les découpages ou les pâtes à modeler, ou de préciser la **texture** avec des notions de transparence, netteté et épaisseur des contours.... Dans certain cas, on peut identifier la nature de l'éclairage et sa position (identification pour les ombres chinoises par exemple, lumière indirecte et éclairage par le fond).

Nous avons, au travers des résultats obtenus par ces différentes techniques, identifié un certain nombre de caractéristiques que nous avons désignées par des termes. Nous proposons d'organiser ces descripteurs pour des utilisations futures. Il ne s'agit pas ici de construire une catégorie de notre ontologie, mais de regrouper les notions recouvertes par ces termes et d'augmenter la sémantique de ces descripteurs. Nous regardons si, au travers de relations qualitatives, nous pouvons obtenir des descripteurs plus performants. Nous avons cité plus haut les termes de couleur, rythme et texture: la COULEUR peut être définie par une combinaison des descripteurs Intensité, Contraste, Saturation, Chaleur et **Tonalité** dominante pour certains films (la tonalité étant donnée par une combinaison des descripteurs diversité et variété), le RYTHME peut être caractérisé par une combinaison des descripteurs Activité, Rapidité et Explosivité. Enfin les effets spéciaux comme ceux transmis par la texture ou la notion de mystère peuvent interférer avec les couleurs et le rythme. Une étude des corrélations entre ces trois familles de descripteurs nous permettra sans doute de caractériser partiellement l'AMBIANCE d'un film. Et nous pourrons ainsi nous intéresser aux rapports existants entre une ambiance et un genre, ou entre les techniques utilisées et une ambiance.

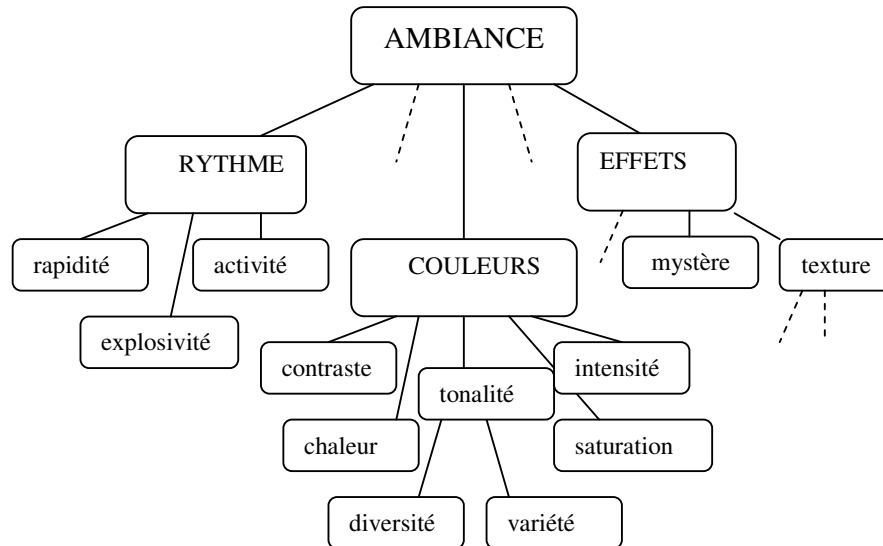


Fig. 3 – Descripteurs issus des images

5 Fusion de connaissances hétérogènes

Nous avons, par une approche externe des objets multimédia, étudié les textes qui les décrivent ou les jugent et les métadonnées des bases de données qui les recensent. Nous avons défini une ontologie terminologique et de ce fait une conceptualisation d'une partie des connaissances. Par l'analyse des images, donc une approche interne, nous obtenons des descripteurs de nos films permettant de définir des caractéristiques floues comme l'ambiance et de mettre en jeu d'autres types de connaissances. Notre but est de voir comment ces deux approches peuvent se compléter pour obtenir des outils de gestion des objets multimédia plus performants.

Nous proposons d'étudier l'interaction de ces connaissances hétérogènes et nous en attendons l'émergence de nouveaux concepts permettant de caractériser, par exemple, l'atmosphère qui se dégage d'un film. Ces travaux sont encore en cours de développement, mais on peut déjà annoncer les directions vers lesquelles il semble possible de chercher pour établir des liens entre les caractérisations issues des images et les concepts issus des textes.

5.1 Texte et images : complémentarité réciproque

Comme nous l'avons déjà indiqué, la difficulté est de donner un niveau sémantique aux descripteurs "image" compatible avec le niveau sémantique des concepts issus des textes. Pour atteindre cet objectif, une solution peut consister à établir une classification supervisée des images (Borges, 1998), la supervision étant apportée par les caractéristiques textuelles. La procédure se décompose alors en deux phases, une phase d'apprentissage précédant une phase d'exploitation. La phase d'apprentissage se décompose en cinq étapes:

- catégorisation d'un corpus de films à l'aide de la terminologie mise en place,
- extraction des caractéristiques image sur ce même corpus,
- construction de l'espace des caractéristiques image et positionnement des caractéristiques extraites du corpus,
- étiquetage des points de cet espace par les termes issus de l'approche textuelle et
- recherche de classes homogènes (Kaufman et Rousseeuw, 1990).

A ce niveau, les tests faits sur un échantillon de film ne peuvent être généralisés. Il est probable que la plupart des caractéristiques, sur l'ensemble du corpus, ne se groupent pas naturellement en classes ou que ces classes ne puissent pas être nommées de façon simple et utilisable. La solution sera dans la recherche de nouvelles caractéristiques ou dans l'utilisation de techniques de classification adaptées. Dans la phase d'exploitation, la classification obtenue en phase d'apprentissage sera utilisée pour catégoriser d'autres films en exploitant uniquement les données images (situation où le film n'a que peu ou pas de descriptif textuel). On peut également envisager que cette classification vienne en soutien de l'analyse textuelle pour confirmer certains choix.

Selon ce schéma, la terminologie utilisée est celle apportée par le texte, l'image venant renforcer ou suppléer le texte. Bien sûr, l'application de cette démarche demande un choix des descripteurs image qui soit en lien plus ou moins serré avec cette terminologie. Il nous semble que les descripteurs attachés au rythme ou à la couleur pourraient être associés à la terminologie sur les genres. Par exemple, un film du genre "drame" aura un rythme sans doute lent et une dominance des couleurs sombres, alors qu'une comédie aura un rythme plus alerte et des couleurs plus éclatantes et plus claires.

Inversement, on peut essayer de définir les caractéristiques visuelles (audio) associées à une technique de réalisation. Par exemple, l'utilisation de poudre provoque une texture particulière et un rythme de modification des images spécifique. Ces caractéristiques sont implicitement utilisées par les experts pour identifier un certain nombre de techniques de réalisation dans les films visionnés, mais les associations de ces caractéristiques et des autres connaissances de l'expert ne sont, pour l'instant, ni explicites, ni exprimées en fonctions des techniques d'analyse d'image.

5.2 L'analyse conjointe des descripteurs

Dans cette deuxième approche, le texte et l'image sont d'abord exploités indépendamment de manière à fournir leurs propres caractéristiques. Ceci suppose que l'analyse des images parvienne à fournir des caractéristiques dont la nature sémantique soit suffisante. On peut pour cela s'aider des transcriptions numériques / symboliques proposées par la logique floue (Bouchon-Meunier et al. 1990).

Ensuite, en se plaçant dans l'espace de dimension N des caractéristiques (N est le nombre total de caractéristiques de type texte et image), on procède à une analyse factorielle dans le but de faire émerger des axes principaux caractéristiques. L'intervention des experts est alors indispensable pour donner du "sens" à ces axes factoriels.

A titre d'exemple, nous montrons un résultat issu des tentatives de classification à partir des indicateurs images, sur des films dont nous connaissons les caractéristiques sur l'ontologie. Nous avons testé expérimentalement un certain nombre de méthodes de classification non supervisée, répétant les expériences en faisant varier les paramètres, pour trouver des

Connaissances hétérogènes des documents multimédia

résultats exploitables et stables. La méthode utilisée finalement est la méthode des nuées dynamiques pour laquelle nous avons cherché la distance la plus appropriée. Après des tests notamment avec les distances euclidiennes, cityblock, cosinus et corrélation, nous avons retenu la distance euclidienne. Elle nous fournit les classes les plus appropriées, que nous chercherons à nommer et caractériser de façon aussi fine que possible, avec l'aide des experts. Dans chaque classe, nous avons des films qui ont des "caractéristiques proches", caractéristiques que nous allons tenter d'exprimer avec l'aide des experts.

La figure suivante (figure 4) donne un résultat obtenu sur la classification en fonction des techniques de couleurs, ce qui correspond à 18 paramètres pour chaque film. Les tests ont été faits pour 2, 3 et 4 classes. Pour la visualisation ci-dessous, chaque film est représenté par les trois premières composantes principales du vecteur de caractéristiques et le numéro de la classe est encadré. Certains films représentatifs des classes sont résumés par quelques images et leur histogramme global.

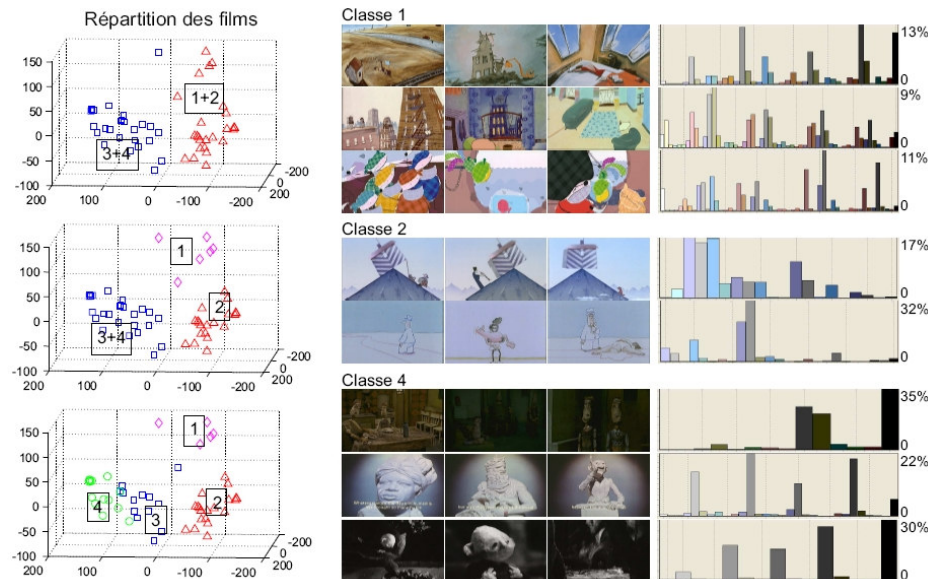


Fig.4 : *exemple de répartition des films en classes en fonction des techniques de couleurs.*

Nous pouvons observer que les films se retrouvent bien groupés selon la similarité des techniques de couleurs utilisées :

- pour $N = 2$ les films sont divisés en deux groupes : les films ayant un contenu riche en couleurs (classe 1+2) et les films utilisant une palette réduite de couleurs (classe 3+4).
- en augmentant le nombre de classes à $N = 3$, seule la classe 1+2 se subdivise : la classe 2 contient les films ayant une diversité moyenne et des couleurs adjacentes, la classe 1 contient les films ayant une variété/diversité élevée des couleurs,

- pour $N = 4$, les classe1 et classe2 restent inchangées. Seule la classe 3+4 se divise en une classe 3 contenant quelques films à palette couleur réduite mais sans autres caractéristiques communes et une classe4 contenant la plupart des films ayant une palette très réduite et ayant comme couleurs prédominantes des couleurs sombres.

6 Conclusion

L'étude des apports réciproques des textes et des images n'en est qu'à ces débuts. Dans le domaine de l'image, l'ensemble des descripteurs doit être enrichi et les relations qu'ils entretiennent doivent être étudiées de façon plus formelle. Nous nous sommes limités à des descripteurs "d'ambiance", une facette de ce que nous appellerons "atmosphère", et qu'il conviendra peut-être de compléter, pour décrire le contenu des documents multimédia. L'extraction des caractéristiques de mouvement, la mise en place d'opérateurs spécifiques (recherche de visages, suivi de personnages, etc.) donneront une meilleure pertinence aux informations sémantiques issues de cette approche. Dans le domaine du texte, l'utilisation des critiques de films et d'autres sources de documents, la gestion des synonymes et la prise en compte de l'évolution des termes dans le temps permettront d'affiner la base de connaissances terminologiques.

Nous avons surtout proposé ici une méthodologie de construction d'une base de connaissances à partir de données hétérogènes d'un point de vue sémantique et de part la diversité de leur nature. Par une approche bipolaire, nous pouvons exploiter les caractéristiques de réalisation (visuelle pour le moment, audio dans un futur proche) au travers des images tandis que l'étude des méta données disponibles, des descriptions ou commentaires, sous une forme textuelle, nous permet de définir des concepts organisés dans une ontologie OK associée à une terminologie. Les contenus sémantiques associés aux termes issus de ces deux approches sont évidemment encore éloignés mais réduire ce fossé sémantique reste un enjeu important, pour établir des relations entre les apports des différents média en vue de l'organisation et de l'exploitation des documents audio visuels.

Enfin, les techniques de fusion envisagées doivent être affinées et validées par des expérimentations sur une large base de films, en accord avec les experts du domaine. En effet, la construction d'outils de gestion et d'exploitation des grandes bases d'objets multimédia efficaces et pertinents, basés sur les contenus et non seulement sur la structure ou les éléments descriptifs sera favorisée par la fusion de ces connaissances hétérogènes.

Références

- Bachimont B., (2000). *Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances*. In J. Charlet, M. Zacklad, G. Kassel & D. Bourigault (Eds.), *Ingénierie des connaissances, évolutions récentes et nouveaux défis*. Paris: Eyrolles
- Bouchon-Meunier B., Desprès S., Dubois D., Gascuel O., Genoeche A., Prade H., (1990). *Interface entre symbolique et numérique*, Actes des 3^{ème} journées nationales du PRC-IA, Hermès, pp. 89-138, Paris.
- Bourigault D., Slodzian M., (1999). *Pour une terminologie textuelle*, Revue Terminologies Nouvelles n°19 déc1998 – juin.

Connaissances hétérogènes des documents multimédia

- Burges, C.J.C. (1998). *A tutorial on Support Vector Machines for Pattern Recognition*, Data Mining and Knowledge Discovery, Kluwer Academic Publishers, vol. 2, n°. 2, pp. 12-16, 1998.
- Chiu C. & Kuipers B. J., (1992). *Comparative analysis and qualitative integral representations*, In Boi Faltings and Peter Struss (Eds.), Recent Advances in Qualitative Physics, MIT Press, 1992.
- Del Bimbo, A. (1999). *Visual Information Retrieval*, Morgan Kaufmann Publishers Inc. San Francisco California, USA.
- Deloule F., Tomasena M., Vescovi M. (1990). *State of the art in qualitative modelling and recommendations for the QMBS techniques to be used*, rapport interne, 30 p, fév.1990.
- Depecker, L., (2002). *Entre signe et concept: éléments de terminologie générale*, Presses Sorbonne Nouvelle
- Desclés, J-P., (1990). *Langages applicatifs, langues naturelles et cognition*, Paris, Hermès.
- Dieng R., Corby O., Gandon F., Giboin A., Golebiowska J. Matta N., Ribière M., (2005). *Knowledge management, methodes et outils pour la gestion des connaissances*, Dunod.
- Ganascia J.G., (2006). *Sciences cognitives*, Eyrolles
- Gruber T. (1995). *Towards principles for the design of ontologies used for knowledge sharing*, Int. J. Human Comput. Stud. 43
- Huet B., Yahiaoui I. and Merialdo B., (2003). *Comparison of multi-episode video summarization algorithms*, EURASIP Journal on applied signal processing Special issue on multimedia signal processing, 2003(1), January.
- Ionescu B., Coquin D., Lambert P., Buzuloiu V. (2006). *Semantic Characterization of Animation Movies Based on Fuzzy Action and Color Information*, 4th Int. Workshop on Adaptive Multimedia Retrieval, CD-ROM, Geneva, Switzerland, July 2006
- Itten, J. (1961). *The Art of Color: the Subjective Experience and Objective Rational of Color*. New York: Reinhold.
- Kaufman, L., P.J. Rousseeuw P.J.(1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Lienhart R., (2001). *Reliable Transition Detection In Videos: A Survey and Practitioner's Guide*, International Journal of Image and Graphics (IJIG), Vol. 1, No. 3, pp. 469-486.
- Meyer I., Skuce D., Bowker L., Eck K., (1992). "Towards a new generation of terminological resources: an experiment in building a Terminological Knowledge Base", Proceedings of the 14th International Conference on Computational Linguistics (COLING 92), pp 956-960
- Mpeg7 (2004). norme ISO <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- NFISO (2001). [NF ISO 704 2001] *AFNOR*, avril 2001 ISSN 0335-3931
- Petitot, J., (1990). *Le Physique, le Morphologique, le Symbolique. Remarques sur la Vision*, Revue de Synthèse, 1-2, 139-183.

- Rey, A., (1992). *La terminologie, noms et notions*, PUF, Que sais-je?, n°178, 2ème édition corrigée
- Roche C., (2001). *From Information Society to Knowledge Society: the Ontological Issue*, CASYS'2001, Liège, Belgium, 13-18-Août 2001
- Roche C., (2005). *Terminologie et ontologie*, revue Langage n°157, mars 2005, ed. Larousse
- Santini S., (2001). *Exploratory Image Databases: Content-Based Retrieval*, Academic Press, 2001.
- Slodzian M., (1995) *Comment revisiter la doctrine terminologique aujourd'hui ?*, Banque des mots, n° spécial 7/1995
- Smeulders A., Worring M., Santini S., Gupta A., and Jain R., (2001). *Image databases at the end of the early years*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22(12), 2001, pp. 1349-1380.
- Snoek C.G.M., Worring M., (2005). *Multimodal Video Indexing : A Review of the State-of-the-art*, Multimedia Tool and Applications, vol. 25, no. 1, pages 5–35, 2005.
- Teulier R., Charlet J., Tchounikine P. (2005). *Ingénierie des connaissances*, R. Teulier, J. Charlet, P. Tchounikine, Communication Medias, ed. L'harmattan.
- Tomasena M., Deloule F., Vescovi M. (1990) *Selection of a qualitative model-based system*, Deliverable Report à l'intention de la Commission des Communautés Européennes, Projet Esprit 2671, 28 pages, février 1990
- Visibone, (2006). Webmaster Palette. <http://www.visibone.com/colorlab>, 2006.
- Wurd (2004). "Worldnet User's Reference Desk", (www.wurd.com/pwp_color.php)
- Zadeh L.A. (1965). *Fuzzy sets*. Information and Control. **8** (1965) 338–353.

Summary

The multimedia objects are new supports of knowledge which develop quickly. The challenge to take up consists in proposing powerful tools to reach the contents of these objects and to provide a semantic approach of these very large knowledge bases. We propose a bipolar approach. From an external point of view on these multimedia objects, we use terminological ontologies to locate the relevant semantic elements in the metadata and in the texts which exist on or around the multimedia objects. From an internal point of view, we build increasing semantic level indicators coming from the images. Lastly, we propose a fusion of these heterogeneous data to try to provide powerful tools of search based on the contents in the bases of multimedia objects. We illustrate our matters by a cartoon film corpus.