



HAL
open science

Structuration de collections d'images par apprentissage actif crédibiliste

Hervé Goëau

► **To cite this version:**

Hervé Goëau. Structuration de collections d'images par apprentissage actif crédibiliste. Interface homme-machine [cs.HC]. Université Joseph-Fourier - Grenoble I, 2009. Français. NNT: . tel-00410380

HAL Id: tel-00410380

<https://theses.hal.science/tel-00410380v1>

Submitted on 20 Aug 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structuration de collections d'images par apprentissage actif crédibiliste

THÈSE

Soutenue publiquement le 25 mai 2009 pour l'obtention du titre de

Docteur de l'Université Joseph Fourier de Grenoble

spécialité

Signal, Images, Parole, Télécoms

par

Hervé GOËAU

Rapporteurs :	Jenny BENOIS-PINEAU	Université Bordeaux 1
	Olivier COLOT	Université des Sciences et Technologies de Lille
Examineurs :	Michèle ROMBAUT	Université Joseph Fourier Grenoble
	Sylvie PHILIPP-FOLIGUET	ENSEA Cergy-Pontoise
	Jean-Michel RENDERS	Xerox Research Centre Europe
Directeur de thèse :	Denis PELLERIN	Université Joseph Fourier Grenoble
Encadrement industriel :	Olivier BUISSON	Institut National de l'Audiovisuel
	Marie-Luce VIAUD	Institut National de l'Audiovisuel

Laboratoire Grenoble Images Parole
Signal Automatique (GIPSA-lab)



Institut National de l'Audiovisuel



REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de thèse Denis Pellerin pour son encadrement scientifique, ses conseils avisés et son soutien permanent. Je suis très reconnaissant pour sa disponibilité et pour avoir pris en charge certaines démarches administratives auprès de l'école doctorale. Je remercie également Michèle Rombaut d'avoir accepté, à l'invitation de Denis, de participer activement à cet encadrement, alors que rien ne l'y obligeait. Je suis très reconnaissant pour toute son expertise sur le Modèle des Croyances Transférables qu'elle a pu amener dans ce travail, et pour toutes ses recommandations pertinentes sur la construction et la rédaction du manuscrit. A eux deux, Michèle et Denis m'ont transmis une méthodologie de travail et la rigueur scientifique nécessaire pour valoriser mes travaux de thèse. Je leur dois en grande partie la réussite et la qualité pédagogique de ce travail.

Je remercie sincèrement les membres du jury de ma thèse pour avoir accepté de rapporter et examiner mes travaux : Jenny Benois-Pineau, Olivier Colot, Sylvie Philipp-Foliguet et Jean-Michèle Renders. Je suis très honoré qu'ils m'aient accordé une partie de leur temps si précieux.

Je remercie infiniment mes encadrants de l'Ina, Marie-Luce Viaud et Olivier Buisson, pour leur soutien, dès mon premier stage dans leur équipe (5 ans déjà !) et leurs encouragements pour mon orientation vers la recherche, malgré mon parcours atypique. Merci à Marie-Luce de m'avoir permis de travailler sur des sujets de recherche passionnants. Merci à Olivier pour m'avoir accompagné de très près dans mes travaux. Ses contributions à la réussite de ce travail sont immenses grâce à nos discussions scientifiques, ses recherches bibliographiques, ses conseils, son optimisme et son dynamisme.

Un grand merci également à toute l'équipe : Jérôme pour ses conseils et pour m'avoir permis d'acquérir des compétences en programmation objet, Agnès pour sa disponibilité et son aide précieuse pour l'évaluation avec Véronique, la personne responsable de la Photothèque de l'Ina que je remercie également.

Un salut amical à la " bande " de l'Ina : Félicien, Thomas, Jeremy, Quentin, Sébastien P., Ludovic, Sébastien G., sans oublier les collègues des premières heures : Jean-Philippe, Julien, Sorin, Marc, Jean-Pascal.

Je remercie également les personnes qui ont su rendre mon séjour à l'Ina agréable : Jean-Etienne, Fabrice, Louis L., Alain, Chantal, Valérie, Laurence, Louis C., Pascal, Steffen. . .

Une mention spéciale pour mon cousin Cédric qui m'a accueilli de nombreuses fois à Grenoble pendant ces trois années, avec qui j'ai pu passer des moments de détente dans des moments de travail intenses.

Je remercie tendrement mes parents, mon frère et ma famille qui m'ont accompagné et cru en moi tout au long de ces années.

Enfin, un immense merci à mon épouse Laëtitia sans qui rien ne serait possible. Merci pour ta patience et ton soutien dans les moments difficiles. J'espère que nous pourrons concrétiser rapidement tous nos rêves...

RÉSUMÉ

L'indexation des images est une étape indispensable pour valoriser un fond d'archive professionnel ou des collections d'images personnelles. Le "documentaliste" se doit de décrire précisément chaque document collecté dans la perspective de le retrouver. La difficulté est alors d'interpréter les contenus visuels et de les associer entre eux afin de couvrir différentes catégories qui peuvent être souvent très subjectives. Dans ce travail, nous nous inspirons du principe de l'apprentissage actif pour aider un utilisateur dans cette tâche de structuration de collections d'images. A partir de l'analyse des contenus visuels des images, différentes stratégies de sélection active sont développées afin d'aider un utilisateur à identifier et cerner des catégories pertinentes selon son point de vue. Nous proposons d'exprimer ce problème de classification d'images avec apprentissage actif dans le cadre du Modèle des Croyances Transférables (MCT). Ce formalisme facilite la combinaison, la révision et la représentation des connaissances que l'on peut extraire des images et des classes existantes à un moment donné. La méthode proposée dans ce cadre permet ainsi une représentation détaillée de la connaissance, notamment en représentant explicitement les cas d'appartenances à aucune ou à de multiples catégories, tout en quantifiant l'incertitude (liée entre autre au fossé sémantique) et le conflit entraîné par l'analyse des images selon différentes modalités (couleurs, orientations). Une interface homme-machine a été développée afin de valider notre approche sur des jeux de tests de référence, des collections d'images personnelles et des photographies professionnelles issues de l'Institut National de l'Audiovisuel. Une évaluation a été conduite auprès d'utilisateurs professionnels et a montré des résultats très positifs en termes d'utilité, d'utilisabilité et de satisfaction.

Mots-clés : Classification supervisée de collection d'images, apprentissage actif, Modèle des Croyances Transférables, Fusion d'informations, multi-étiquetage

Abstract

Image annotation is an essential task in professional archives exploitation. Archivists must describe every image in order to make easier future retrieval tasks. The main difficulties are how to interpret the visual contents, how to bring together images which can be associated in same categories, and how to deal with the user's subjectivity. In this thesis, we use the principle of active learning in order to help a user who wants organize with accuracy image collections. From the visual content analysis, complementary active learning strategies are proposed to the user to help him to identify and put together images in relevant categories according to his opinion. We choose to express this image classification problem with active learning by using the Transferable Belief Model (TBM), an elaboration on the Dempster-Shafer theory of evidence. The TBM allows the combination, the revision and the representation of the knowledge which can be extracted from the visual contents and the previously identified categories. Our method proposed in this theoretical framework gives a detailed modeling of the knowledge by representing explicitly cases of multi-labeling, while quantifying uncertainty (related to the semantic gap) and conflict induced by the analysis of the visual content in different modalities (colors, textures). A human-machine interface was developed in order to validate our approach on reference tests, personal images collections and professional photos from the National Audiovisual Institute. An evaluation was driven with professional users and showed very positive results in terms of utility, of usability and satisfaction.

Keywords: image collection structuration, supervised classification, active learning, Transferable Belief Model, multi-labeling

TABLE DES MATIÈRES

1	Introduction	1
1.	Contexte et motivations	1
1.1.	Catégorisation, classification, structuration	3
1.2.	L'Institut National de l'Audiovisuel	6
2.	Objectifs de la thèse	8
2.1.	Structuration de collection d'images	8
2.2.	Systèmes existants pour l'aide à l'organisation de collection d'images	8
2.3.	Discussion	10
3.	Contributions et organisation du mémoire	11
2	Système de structuration de collections d'images	15
1.	Contexte applicatif	16
1.1.	Homogénéité visuelle de collections d'images	16
1.2.	Hétérogénéité visuelle de collections d'images	16
1.3.	Diversité des objectifs de structuration de collections d'images	18
1.4.	Besoin du multi-étiquetage	18
2.	Description générale du système proposé	19
2.1.	Modélisation et synthèse de la connaissance	20
2.2.	Sélection active d'images	21
2.3.	Interface homme-machine	22
2.4.	Le retour de l'expertise	22
3	Modélisation et synthèse de la connaissance	25

TABLE DES MATIÈRES

1.	Introduction	26
1.1.	Position du problème	26
1.2.	Contraintes théoriques	26
1.3.	Méthodes de classification automatique	28
1.4.	Choix d'une approche théorique	35
2.	Extraction d'information, description de contenu visuel	36
2.1.	Descripteur global de couleurs	36
2.2.	Descripteur global des orientations	37
3.	Manipulation de descriptions visuelles pour le classement d'images	38
4.	Modèle des Croyances Transférables	41
4.1.	Représentation de la connaissance	42
4.2.	Combinaison	45
4.3.	Prise de décision	46
5.	Modélisation de la connaissance	46
5.1.	Règles de représentation de la connaissance à partir d'une distance	46
5.2.	Conversion numérique - symbolique	48
6.	Fusion de témoignages au sein d'une classe	51
6.1.	Combinaison de voisins	51
6.2.	Réévaluation de distributions de masses	53
7.	Fusion de témoignages pour l'ensemble des classes	55
7.1.	Espaces de discernement	55
7.2.	Fusion multi-classe	58
7.3.	Fusion multi-descripteur	60
8.	Conclusion	63
4	Sélection active d'images	65
1.	Introduction	65

1.1.	Problématique	65
1.2.	Présentation de l'apprentissage actif	66
2.	Utilisation de l'apprentissage actif dans notre contexte applicatif	68
2.1.	Particularités du contexte applicatif	68
2.2.	Stratégies de sélection d'images	69
2.3.	Mise à jour de la connaissance	71
2.4.	Traitement de la connaissance avec le Modèle des Croyances Transférables	71
3.	Stratégies orientées "image"	75
3.1.	Stratégies basées sur les hypothèses du cadre de discernement	76
3.2.	Stratégies basées sur l'analyse des distributions de masses	82
3.3.	Bilan sur les stratégies orientées "image"	86
4.	Stratégie orientée "classe"	87
5.	Mise à jour de la connaissance après étiquetage des images	89
5.1.	Définition de zones de connaissances	89
5.2.	Méthode d'estimation	91
6.	Conclusion	93
5	Interface pour la structuration de collection d'images	95
1.	Introduction	95
2.	Propositions automatiques d'étiquettes	96
2.1.	Schémas d'étiquetages	96
2.2.	Etiquetage simple	98
2.3.	Etiquetage multiple	101
2.4.	Limitation de l'espace de décision	102
3.	Interface	104
3.1.	Description globale de l'interface	104
3.2.	Interactions avec l'utilisateur	105

TABLE DES MATIÈRES

4.	Conclusion	109
6	Performances, expérimentations, évaluations	115
1.	Introduction	116
2.	Performances de classification automatique	117
2.1.	Métriques et descripteurs	118
2.2.	Fusion de descripteurs	125
2.3.	Influence du paramètre f dans le cas de la fusion tardive	128
2.4.	Optimisation du calcul des distributions de masses	128
2.5.	Caractérisations et mesures sur une base de données multi-étiquetée	131
3.	Caractérisations des stratégies de sélection active d'images	135
3.1.	Evaluation classique des stratégies de sélection active	135
3.2.	Caractérisation des stratégies dans le contexte applicatif	138
3.3.	Influence du paramètre f	141
3.4.	Sélection active dans le cas du multi-étiquetage	143
4.	Evaluation avec une documentaliste de l'INA	148
5.	Extensions	153
5.1.	Intégration d'informations métadonnées partielles ou inexactes	153
5.2.	Structuration de vidéos	156
6.	Conclusion	162
	Bilan et perspectives	165
	Annexes	169
7.	Communications	170
	Liste des figures	180
	Liste des tableaux	183
	Bibliographie	192

INTRODUCTION

1. Contexte et motivations

Les bibliothèques sont apparues progressivement durant l'antiquité pour répondre au besoin d'organiser la conservation des écrits, garants d'une mémoire collective. La mythique bibliothèque d'Alexandrie au VII^e siècle avant JC est la première à pratiquer une forme de dépôt légal, c'est-à-dire l'obligation de fournir un exemplaire ou l'original d'un document afin de s'assurer de la conservation de ce patrimoine. A la renaissance, François I^{er} fut l'initiateur du dépôt légal en France par l'édit du 28 décembre 1537. Depuis, au fil du temps, cette pratique s'est étendue à travers le monde et à différents médias (cinéma, musique, télévision, ...).

Au delà du prestige et de la notoriété que peuvent apporter une bibliothèque aux pouvoirs politiques, religieux, ou privés, les objectifs d'une bibliothèque sont de conserver les documents pour les générations futures afin de constituer une mémoire collective et d'offrir un lieu de recherche dans les archives.

Traditionnellement, les techniques bibliothécaires telles que la classification consistent à définir les contours de thématiques et sous thématiques, participant ainsi à l'émergence de nouveaux savoirs. Avec l'apparition du numérique, les méthodes classiques bibliothécaires se sont transposées naturellement pour créer des bases de données relatives aux documents multimédia accessibles alors par thématique et mots-clés ou attributs (date, auteurs, lieux...). La préservation et la conservation des documents audiovisuels est une pratique répandue et indispensable pour valoriser les fonds propres des agences de presse et de photographies professionnelles (AFP, Belga, Capa, Reuters, Sipa,...), des archives audiovisuelles (BBC, INA, RAI, ...), ou encore dans les milieux culturels et artistiques.

Toutefois, le travail d'organisation et d'annotation, traditionnellement confié à des experts devient de plus en plus difficile au regard de la multiplication des contenus.

De même, avec la démocratisation des équipements numériques, le grand public est à son tour de plus en plus confronté, à une moindre échelle, aux problèmes de gestion de collections de données personnelles. Un exemple significatif est celui de la photographie : l'acquisition d'instantanés est devenue un geste banal et il est de plus en plus difficile d'organiser et de retrouver des contenus parmi les quantités de données collectées au fil des années.

Pour faire face aux problèmes d'accès à l'information pertinente parmi des volumes importants, les moteurs de recherche tels que Google ou Yahoo offrent une réponse adaptée notamment à la recherche de documents textes. Grâce à ces outils performants, Internet s'installe progressivement dans nos habitudes

de consultation de documents textuels, mais aussi multimédias, notamment depuis la généralisation des portails vidéo tels que Youtube et DailyMotion.

Or il est important de distinguer deux différences fondamentales qui ne sont pas toujours perceptibles pour le grand public entre Internet et les bibliothèques numériques. La fédération des bibliothèques numériques [Fed] a reprecisé la fonction et le rôle des bibliothèques numériques : *"Organisations qui fournissent des ressources, dont du personnel spécialisé, pour sélectionner, structurer, rendre accessible intellectuellement, interpréter, distribuer, préserver l'intégrité et la persistance dans le temps de collections d'œuvres digitales afin de les rendre lisibles et économiquement accessibles pour une communauté ou un ensemble de communautés d'utilisateurs"*.

D'une part, la bibliothèque est une organisation dont le rôle est de contrôler, de valider et de préserver ses ressources selon des critères explicites, en général spécifiés dans un cahier des charges. D'autre part, cette organisation possède un devoir d'accès auprès de communautés, incluant la notion d'assistance par du personnel spécialisé. Par ces contraintes, la bibliothèque numérique assure à ses utilisateurs une qualité et une fiabilité d'usage, tant au niveau de ses ressources que de leur accès.

Sur la toile, il n'y a qu'une organisation partielle, voire inexistante. Une recherche sur internet privilégie un accès précis grâce à l'indexation automatique des documents textuels par le contenu, notamment si la requête comporte de nombreux mots comme par exemple la recherche d'une citation. Cependant, la recherche de concepts est difficilement formulable. De plus, le décompte de liens (le fameux page rank de Google) privilégie les documents qui ont les moyens de se fabriquer une audience. Le risque est alors de ne favoriser qu'une partie de l'information, au détriment d'œuvres plus critiques. En cela, les réponses renvoyées par un moteur de recherches ne peuvent garantir une qualité maximale.

La garantie de qualité d'une bibliothèque numérique réside dans le paradigme suivant : *toute œuvre inaccessible est une œuvre perdue*. Deux phénomènes peuvent rendre une œuvre inaccessible. Si un document pertinent à une requête se trouve mêlé à un trop grand nombre de documents de faible pertinence, l'utilisateur ne peut plus analyser les résultats et atteindre sa cible : il est confronté au bruit documentaire. A l'opposé, si aucun résultat ne répond à une requête correctement formulée, l'utilisateur est confronté au silence documentaire.

Un bon système documentaire doit donc minimiser le bruit et le risque de silence documentaires. La structuration sémantique et catégorielle des contenus constitue la réponse "métier" des bibliothèques à cette problématique.

Dans cette thèse, nous abordons la documentation de collections d'images dans une perspective de valorisation et de conservation des documents. Nous souhaitons proposer un outil assistant des utilisateurs dans des tâches d'organisation et d'annotation de collections d'images. Cet outil doit aider les utilisateurs à documenter précisément et sans erreurs toutes les images des collections. Cette qualité de traitement permettra alors de minimiser le bruit et le risque de silence documentaire lors de la recherche de contenus.

1.1. Catégorisation, classification, structuration

Une des notions clés dans ce travail est la structuration de collection d'images. La structuration est un outil puissant pour analyser et manipuler des images dans différents contextes applicatifs. Par exemple, elle permet d'identifier des groupes d'images à annoter par lot au sein d'une banque de photographies et peut apporter un gain de productivité significatif dans une chaîne de documentation. La structuration peut aider également à résoudre les problématiques de navigation au sein de vidéos : identifier les liens entre différents groupes de séquences permet alors de manipuler le contenu et de concevoir de nouveaux modes d'exploration et de navigation dans les vidéos.

Il est nécessaire de distinguer cette notion des termes proches que sont la catégorisation et la classification. En général, catégorisation et classification sont des termes synonymes qui renvoient à un processus par lequel différents objets sont perçus comme similaires.

Pour la théorie classique dite "aristotélicienne", les catégories sont des entités discrètes qui se définissent par un ensemble de caractéristiques communes aux éléments qui les constituent. Cette description assez rigide représente mal la complexité du monde. Elle est relativisée par exemple par la théorie des prototypes [Ros73] qui affirme qu'une catégorisation n'est jamais idéalement réalisée mais s'approche graduellement d'un prototype ou d'un modèle abstrait.

Pour les sciences cognitives, le terme de catégorisation renvoie à un processus fondamental dans la perception et la compréhension de concepts et d'objets, dans la prise de décision et dans toutes les formes d'interaction avec l'environnement [JR99]. La catégorisation permet à l'individu d'organiser et de réduire la complexité de l'environnement, en le découpant en objets qu'il regroupe et attribue à différentes catégories.

Le terme de "catégorisation" relève donc du domaine de la psychologie. Le terme de "classification" quand à lui correspond aux processus, aux théories mathématiques et techniques employés permettant la catégorisation.

Une ambiguïté réside lorsque le terme de "classification" est associé à la notion de "classement", comme dans les sciences de l'information ou les sciences bibliothécaires, où "classification" représente un système organisé et hiérarchisé de classement des connaissances. Un exemple typique est la classification scientifique des espèces vivantes, dite "systématique", où les caractéristiques morphologiques permettent de diviser le monde vivant en de nombreuses catégories emboîtées.

Le terme de "structuration" est un terme générique dépendant du contexte dans lequel il est employé. En effet, les définitions dans les dictionnaires sont relativement abstraites : "manière dont les parties d'un tout sont arrangées entre elles", ou "action de doter d'une structure ou fait de l'acquérir" ou encore "organisation des parties d'un système, qui lui donne sa cohérence et en est la caractéristique permanente".

En indexation multimédia, le terme de "structuration" est rarement employé dans le contexte de collections d'images fixes. Par contre, il est utilisé pour les vidéos, par exemple pour désigner une analyse permettant l'établissement d'une table des matières afin de naviguer dans une vidéo [KGOG06], ou bien

pour aligner une grille de programmation sur un flux télévisuel [Po107], [NG08].

Dans ce travail présenté ici, nous souhaitons étendre le concept de structuration aux collections d'images fixes. Dans ce cas, nous définissons la structuration comme une extension plus libre de la catégorisation.

La catégorisation d'images fixes sous-entend une description relativement rigide du monde car elle est souvent liée soit à la nature des objets contenus dans une image (une personne, un véhicule, une maison, ...) soit à la nature d'une scène (intérieur-extérieur, jour-nuit, urbain-naturel, ... [HB07]), [VSWB06]. Ce type de problématique est plus souvent nommé catégorisation d'objets ou de scènes : les images d'une même catégorie sont des illustrations d'un même concept clairement identifiable et pouvant être exprimé explicitement par un ou plusieurs mots-clés.

Cependant, même si la catégorisation de scènes et d'objets est très contrainte par la nature des éléments composant l'image, il est possible qu'une part de subjectivité réside d'un utilisateur à l'autre. En d'autres termes, rien ne garantit qu'il existe une organisation universelle valable pour tout utilisateur.

De plus, si la catégorisation d'images en scènes et-ou objets est un traitement utile, il est possible qu'un utilisateur, dans certains cadres applicatifs, ne soit pas intéressé par ce type de catégorisation. Par exemple, certains utilisateurs peuvent souhaiter effectuer de la catégorisation d'images complètement subjective. Dans ce cas, un utilisateur est libre d'organiser les contenus comme il l'entend sans forcément respecter la catégorisation classique en objets et scènes.

Nous souhaitons employer la notion plus générale de structuration de collection d'images pour englober différents cadres applicatifs (illustrés figure 1.1) jouant sur le degré de subjectivité autorisé : de la catégorisation d'images en scènes ou objets, à la catégorisation personnelle la plus subjective possible.

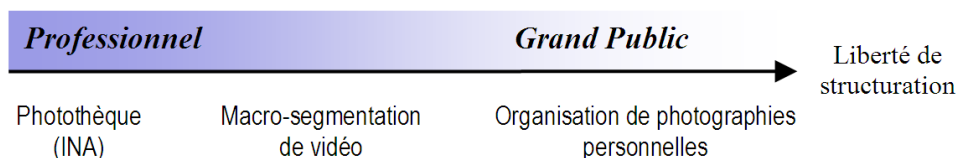


FIGURE 1.1 : La liberté d'expression des catégories par un utilisateur durant la structuration d'une collection d'images ou d'une vidéo, dépend du contexte applicatif.

Dans un cadre professionnel, l'interprétation de l'utilisateur est soumise à des règles, à un protocole de documentation. Toute annotation d'image doit utiliser des mots-clés appartenant à un vocabulaire prédéterminé dans un Thésaurus, afin de garder une cohérence entre le travail de description des différents documentalistes et afin de garantir la pérennité du fond. Toute interprétation personnelle est alors canalisée, ce qui constitue en partie le savoir-faire du métier de documentaliste.

La figure 1.2 illustre un cas réaliste pouvant être rencontré au sein de la Photothèque de l'Institut National de l'Audiovisuel. Dans ce reportage photographique, la collection d'images contient des photographies réalisées lors d'un tournage d'une émission. Le documentaliste peut regrouper toutes les images contenant une caméra. Dans ce cas, une catégorie "camera" est rattachée à un concept concret et explicite, valable notamment pour tout autre documentaliste.



FIGURE 1.2 : *Exemple d'une classe d'images contenant une caméra issues d'un reportage photographique de la photothèque de l'INA.*

En relâchant la contrainte d'expression des catégories, nous pouvons trouver une autre utilisation dans un contexte de macro-segmentation de vidéo. Cette tâche consiste à segmenter une vidéo en séquences temporellement homogènes et à répertorier les différents groupes narratifs. Il paraît alors raisonnable d'orienter cette structuration sur les éléments revenant régulièrement comme par exemple les personnages, les décors ou encore la musique. Mais malgré ce cadre relativement rigide, deux utilisateurs ne porteront peut-être pas leur attention sur les mêmes parties narratives. Par exemple, dans le cas d'une fiction, un spectateur peut considérer comme élément structurant l'action autour d'un seul personnage, alors qu'un second spectateur peut être plus attentif aux différents lieux et décors où se situent les actions (voir figure 1.3).



FIGURE 1.3 : *Deux points de vue narratifs différents selon deux utilisateurs distincts. Le premier utilisateur porte son attention principalement sur le personnage de gauche, alors que le second est plus attentif à l'intrigue se déroulant dans un même décor.*

Relâcher complètement la contrainte d'expression des catégories est adapté au tri de photographies personnelles. Dans ce type d'application grand public, la structuration peut être totalement arbitraire, propre au raisonnement et ainsi qu'à l'humeur du moment de l'utilisateur. Même si de grandes catégories ap-

paraîtront probablement comme "portraits" ou "paysages", d'autres beaucoup personnelles pourront être définies. Deux personnes n'organiseront pas forcément de la même manière une collection car elle est le reflet d'une période vécue de leur histoire. Le plus souvent, elles ne donneront pas le même nombre de groupes et ne subdiviseront pas de la même manière. Nous pouvons même supposer qu'un utilisateur unique ne reproduira pas la même organisation s'il effectue l'opération plusieurs fois à des instants différents.

La figure 1.4 illustre ce cas : lors d'un séjour à l'étranger, des photographies sont prises pendant les trajets en voiture. Le voyageur souhaite regrouper ces images sous une même catégorie ne contenant pas seulement des images de routes, mais des portraits, des paysages variés photographiés sur plusieurs jours, à chaque fois qu'il se déplace. Il ne peut utiliser l'information de date de prise de cliché, puisqu'il s'est déplacé de manière aléatoire, à différents instants de la journée. De plus, les contenus visuels peuvent être très différents, alors que les photographies sont liées entre elles selon le raisonnement de l'utilisateur.



FIGURE 1.4 : Exemple d'une classe d'images reliées par un concept relativement abstrait et personnel de photographies prises sur la route lors d'un déplacement pendant un séjour.

1.2. L'Institut National de l'Audiovisuel

L'Institut National de l'Audiovisuel (INA), le partenaire industriel du contrat CIFRE de cette thèse, a pour mission la conservation, l'exploitation, la mise à disposition du patrimoine audiovisuel et radiophonique auprès des professionnels de l'audiovisuel, ainsi que l'accompagnement des évolutions du secteur audiovisuel au travers de ses activités de recherche, de formation et de production. L'INA est un Établissement Public à caractère Industriel et Commercial (EPIC) créé après l'éclatement de l'ORTF en 1974 et la réforme de l'audiovisuel mise en place le 6 janvier 1975. Au sein de l'INA, l'Inathèque de France est chargée, aux termes de la loi du 20 juin 1992 relative au dépôt légal et du décret d'application du 31 décembre 1993, de collecter, de conserver et de rendre accessible les documents sonores et audiovisuels

radiodiffusés ou télédiffusés.

Les documentalistes de l'INA sont chargées de réaliser l'ensemble des fonctions documentaires concourant à la production, la diffusion, la conservation d'œuvres, de programmes ou de documents sur quelque support que ce soit. Ils ont la difficile et lourde tâche d'assurer l'annotation manuelle des flux audiovisuels arrivant en continu et l'enrichissement *a posteriori* des contenus en adoptant un point de vue rédactionnel.

L'offre grand public de l'INA tente de s'adapter aux nouveaux modes de consommation du public, en proposant les œuvres de moins en moins dans leur intégralité, mais plus sous la forme d'extraits et de sélections. Cette nouvelle tendance d'exploitation implique une augmentation considérable des traitements documentaires, notamment une description et un découpage plus fins des vidéos en séquences et une synchronisation entre les fiches documentaires et les contenus visuels. Par ailleurs, le public attend de plus en plus de la part de l'INA un recul sur le patrimoine audiovisuel français à travers une meilleure connaissance et une réelle analyse de l'évolution des programmes au cours du temps¹.

Ces évolutions du cœur de métier de l'INA s'inscrivent dans un contexte de recherche de gain de productivité, notamment en documentant des traitements de flux plus importants avec un effectif constant de documentalistes. Un des challenges pour atteindre cet objectif est de faire évoluer les outils disponibles pour traiter l'information entrante.

Dans le contexte de l'INA, à l'heure où une grande part des consultations commence à s'effectuer en ligne, un enjeu de taille est d'assurer la continuité sémantique entre les processus d'indexation et de recherche. La structuration du fond est donc une tâche fondamentale. Par exemple, la Photothèque de l'INA contenant un fond d'environ 1,4 millions photographies est en cours de numérisation. Actuellement, 220 000 photographies sont référencées dans une base documentaire et sont répertoriées en "reportages". Un reportage peut contenir quelques dizaines à plusieurs centaines d'images et correspond à une série de photographies prises autour d'un même thème : tournage d'émissions ou de fictions, des images de studio reflétant l'activité durant une période, des équipements techniques, des reportages d'actualité entre 1961 et 1974,...

Cependant, il n'y a pas de description documentaire au niveau des photographies. Le traitement documentaire de ces reportages implique une description plus fine de chaque photographie afin de valoriser et d'identifier les contenus en vue de prochaines exploitations. Un documentaliste annoté et insère manuellement dans la base de données de l'INA environ 55 photographies pour une durée de travail réglementaire de 7h12 par jour. A ce rythme, la numérisation et le traitement documentaire intégral du fond se compte en dizaines d'années pour un effectif de moins d'une dizaine de documentalistes. Certaines images peuvent poser également des problèmes d'ouverture de droits (droits à l'image des personnes apparaissant dans les photographies, droits d'auteurs pour les mises en scène des réalisateurs...), et ajoutent un traitement documentaire supplémentaire.

1. Quelques exemples : l'observatoire des médias et l'outil éducatif Jalon :
<http://www.ina.fr/observatoire-medias/ina-stat/>
<http://www.ina.fr/archives-tele-radio/professionnels/jalons.html>

Il peut être intéressant de proposer aux documentalistes un outil complémentaire pour regrouper des photographies pouvant partager certains champs d'annotation dans une perspective de gain de productivité.

2. Objectifs de la thèse

2.1. Structuration de collection d'images

L'objectif de cette thèse est d'aider un utilisateur devant organiser une collection d'images en lui proposant un outil l'assistant dans cette tâche, à partir de caractéristiques extraites automatiquement sur les images. Or plusieurs enjeux peuvent être identifiés lors de la structuration de collection d'images.

Premièrement, les collections d'images à traiter peuvent être très diverses. Il peut s'agir de photographies personnelles (vacances, événements familiaux, ...), de fonds d'archives professionnels (illustrations de presse, la photothèque de l'INA, ...), ou encore d'images extraites de vidéos.

En second lieu, les contenus peuvent être très variés au sein même d'une collection qui peut contenir des images en noirs et blancs, en couleurs, des portraits, des paysages, des peintures, des schémas, des logos,...

De plus, l'utilisateur a rarement une visibilité *a priori* des contenus qu'il doit traiter, ou relativement vague dans le cas de photographies personnelles. Généralement, un utilisateur ne sait pas à l'avance comment il va structurer la collection : combien de groupes d'images il va devoir identifier, la taille de ces groupes ou si certaines images peuvent être présentes dans plusieurs groupes...

Généralement, seul le contenu visuel des images peut être exploité. Des techniques de traitement d'images permettent d'extraire des caractéristiques visuelles (couleurs, textures, formes,...) et d'établir des rapports de proximité entre les contenus avec des mesures de similarité. Un enjeu est alors de voir comment ces mesures de proximités visuelles peuvent être exploitées pour constituer des groupes pertinents pour l'utilisateur. Parfois, des informations complémentaires peuvent être disponibles comme dans le cas de photographies personnelles prises avec des appareils numériques récents où des métadonnées rapportent la date de prise de cliché et certaines caractéristiques comme la focale, le temps de pose ou l'ouverture, ... Mais dans le cas des photographies plus anciennes venant de supports argentiques, ces informations sont absentes.

2.2. Systèmes existants pour l'aide à l'organisation de collection d'images

2.2.1. Systèmes manuels

Il existe plusieurs solutions libres ou commerciales pour assister le tri manuel de photographies comme Picajet [Pic], Imatch [Wes], XnView [Gou]. Généralement, l'utilisateur dispose de catégories prédéfinies

qu'il peut associer aux images : des personnes (membres de la famille, amis, . . .), des événements (anniversaires, mariages. . .), des animaux, des objets, des paysages. . . L'utilisateur peut également définir de nouveaux mots-clés lui permettant d'exprimer des catégories personnelles. Les images peuvent être ainsi soit "taguées" directement dans les métadonnées du fichier [JEI], [IPT], ou soit gérées dans une base de données externe.

Du point de vue ergonomique, la plupart de ces outils permettent d'associer les mots-clés des catégories soit en cochant des cases dans une interface, et-ou en utilisant des actions de type glissé-déposé sur des zones symbolisant les différentes catégories disponibles. L'application d'organisation de photographies personnelles de Photoshop Element [Incb] utilise les mêmes principes en y ajoutant un détecteur de visage pour aider l'utilisateur à détailler plus précisément les catégories décrivant des personnes.

L'outil Picasa [Inca] met en avant l'information de date de prise de cliché présente dans les métadonnées comme première organisation de base et propose une interface de navigation ergonomique dans la collection par événements.

Dans l'outil ACDSsee Photo Manager [sys] l'intégralité des informations métadonnées est accessible pour trier les photographies par caractéristiques techniques comme la focale, l'ouverture, l'utilisation de flash, . . . Cela peut se révéler très utile pour repérer rapidement un certain type d'images pour un public maîtrisant les techniques photographiques.

La plupart de ces outils exploitent très rarement les informations sur le contenu visuel (il est par exemple possible de parcourir les images par couleur dominante dans Picasa, et de s'appuyer sur une détection de visage dans Photoshop pour trier les photographies).

2.2.2. Systèmes automatiques et expérimentaux

S'il existe plusieurs solutions logicielles pour l'annotation et la catégorisation manuelles d'images, les systèmes d'aide à la structuration automatique de collection d'images ne sont pas encore diffusés sur le marché, mais sont étudiés dans différents travaux de recherches. Une première approche s'appuie sur le regroupement automatique de photographies par événements temporels grâce aux dates de cliché des métadonnées, en la combinant parfois avec des informations sur les contenus visuels [ACD⁺03], [CFGW05]. D'autres travaux proposent une structuration spatio-temporelle grâce à la géolocalisation de photographies disponibles sur certains modèles d'acquisition [NSPGM04], [PG04].

Certains travaux privilégient la reconnaissance de visages comme l'outil Facebubble [XZ08] ou [CWX⁺07] notamment pour les collections de photographies familiales.

Un grand nombre de travaux se focalisent depuis plusieurs années sur la catégorisation automatique de scènes naturelles et d'objets. Les catégories concernées peuvent être intérieur-extérieur, forêt-montagne-coucher de soleil-villes-plages, dessins ou photographies. . . Généralement, le principe est d'entraîner des systèmes de classification sur des images annotées, puis de prédire les annotations possibles sur de nouvelles images. Différentes campagnes d'évaluation telles que TrecVid [SOK06] ou ImageClef

[NDH09] visent à stimuler et à favoriser les avancées scientifiques dans le domaine de la recherche et d'indexation par le contenu de documents visuels. Le lecteur intéressé peut trouver une comparaison des techniques les plus significatives de ces dernières années dans [BMnM07].

2.2.3. Systèmes hybrides automatique-manuel

L'utilisateur peut être mis à contribution pour une collaboration homme-machine. Cela se justifie en partie par le fait que des regroupements automatiques d'images reflètent difficilement l'interprétation personnelle d'un utilisateur. L'utilisateur est alors mis à contribution pour donner des exemples d'images convenant aux regroupements que la personne désire. Dans cet esprit, beaucoup de travaux concernent la recherche d'images par le contenu avec un système de bouclage de pertinence [CFB04], [LHZ⁺00], [LMZ99]. L'approche classique est alors de demander de manière itérative à l'utilisateur d'identifier des images qui répondent positivement et/ou négativement à sa requête, afin que le système puisse affiner la qualité des résultats et se rapprocher progressivement d'images pertinentes pour l'utilisateur.

Dans ce type d'application, l'objectif n'est pas de structurer entièrement une collection d'images, mais de répondre au besoin de rechercher des images de la même manière qu'un utilisateur le ferait sur internet. L'outil se focalise donc généralement à un problème de classification binaire (appartenance ou non au type d'image désiré). Certains travaux soulignent la difficulté d'exprimer une requête à travers cette classification binaire. Ils proposent de déporter l'attention de l'utilisateur, habituellement focalisée sur la formulation de la requête, sur l'organisation des images réponses en arbres hiérarchiques [RMD⁺07] ou en différents groupes que l'utilisateur peut créer à volonté [UM06]. Ces approches permettent une meilleure expressivité de la requête et facilitent la capture des intentions de l'utilisateur.

2.3. Discussion

Parmi toutes ces approches, plusieurs points peuvent être retenus pour concevoir un outil de structuration de collection d'images. Les outils de gestion manuelle de photographies permettent d'atteindre une très bonne qualité de description des contenus puisque que l'utilisateur traite et valide chaque image. L'association de plusieurs mots-clés sur une même image et la création de nouvelles catégories sont aussi possibles. L'utilisateur est libre d'exprimer ses catégories de la manière la plus personnelle qu'il soit. Ces solutions logicielles comportent un travail de conception sur l'ergonomie et les modalités d'interaction pour faciliter le travail de l'utilisateur.

Les approches totalement automatiques sont adaptées à la gestion de très grandes bases d'images, impossibles à traiter pour un utilisateur seul. Si des informations métadonnées sont présentes, ces techniques peuvent être efficaces pour regrouper les images en événements temporels, voir spatio-temporels. L'analyse des contenus peut aider un utilisateur à dégager des groupes par similarités visuelles qui peuvent potentiellement être associés à des catégories.

Les travaux de recherche s'appuyant sur une collaboration entre la machine et l'humain utilisent une ap-

proche intéressante puisqu'ils combinent le traitement automatique des contenus visuels, tout en laissant la possibilité à un utilisateur d'exprimer ses propres catégories.

Cependant, toutes ces approches ne répondent pas complètement aux objectifs fixés. Il n'est pas envisageable de proposer une énième solution de traitement manuel de collection d'images : le travail est trop répétitif et pose problème si les collections sont nombreuses et de taille importante.

L'analyse automatique peut être très efficace si des métadonnées sont présentes notamment grâce à l'utilisation d'appareils d'acquisition numériques récents (moins d'une décade). Or toutes les images n'ont pas ces informations. En effet, dans le cas de l'INA ou d'autres fonds d'archives de presse ou de photographie, la majorité des images sont scannées à partir de supports argentiques. De plus, l'organisation temporelle n'est pas toujours pertinente : un utilisateur peut avoir l'objectif de rapprocher des images éloignées dans le temps. La géolocalisation quand à elle est une information encore plus rare, fournie par quelques modules GPS complémentaires vendus pour certains appareils photographiques professionnels.

Un bon paradigme est de supposer que les utilisateurs sont intéressés surtout par les visages présents dans les photographies pour organiser des collections d'images. Pourtant si la détection de visage est de plus en plus maîtrisée [VJ01], la reconnaissance de visage reste un problème difficile. Cette approche peut s'envisager sur un nombre limité de visages (par exemple les membres de sa famille, des amis, . . .), mais pas sur des collections d'images généralistes où les personnes photographiées se comptent par milliers.

Les approches traitant automatiquement les images par le contenu s'appuient sur les informations visuelles présentes dans tous les cas. La catégorisation d'images en scènes et en objets est un véritable challenge pour les applications futures d'analyse d'images par le contenu. Cependant, les objectifs de ces travaux ne concernent pas l'expression de catégories personnelles, mais *a contrario* la recherche d'invariants visuels pour les lier à des catégories sémantiques communes à tout utilisateur. La difficulté est de faire correspondre des informations de type signal portées par les pixels avec ce qui est représenté et perçu dans les images par un utilisateur. Le risque est alors de mal associer les images entre elles de manière automatique, et d'imposer de nombreuses corrections à l'utilisateur *a posteriori*.

Nous pouvons nous inspirer de travaux traitant d'une collaboration entre la machine et l'humain. La plupart de ces travaux traitent de la recherche d'images par le contenu dans de grandes bases de données. Le but est alors de retrouver des illustrations d'un concept à un moment donné, sans proposer l'organisation et la structuration de collections complètes d'images. Toutefois, si notre but applicatif est différent, nous pouvons nous inspirer du principe d'expression d'une classe lors d'une recherche d'images, en l'étendant au travail sur plusieurs classes simultanément.

3. Contributions et organisation du mémoire

Nous souhaitons proposer un système d'assistance hybride automatique-manuel pour le tri et l'organisation de collections d'images. Pour élaborer notre système nous allons développer les points suivants :

Un système semi-automatique ne décidant pas seul

Une partie automatique et une partie manuelle doivent être conçues conjointement pour établir une collaboration entre la machine et l'utilisateur. Nous pouvons prendre ainsi les avantages des deux approches. Premièrement, une analyse des contenus doit se faire de manière automatique. Le système doit pouvoir déterminer un état de connaissance pour faire des propositions de classement sur une sous-partie des images disponibles. L'utilisateur contrôle ces propositions, et le système peut alors réévaluer la connaissance en fonction des classements validés par l'utilisateur. Ce bouclage doit faire progresser le deux acteurs : d'une part le système automatique doit s'adapter en fonction d'une sémantique exprimée par l'utilisateur, et d'autre part, les propositions de classements effectuées par le système doivent également aider l'utilisateur à progresser dans sa réflexion sur la structuration qu'il est en train de réaliser.

Une sélection d'images utiles pour l'utilisateur

Le système ne peut soumettre des propositions de classements pour l'intégralité des images à l'utilisateur, car ce dernier ne saurait gérer toutes ces informations à la fois. La partie automatique doit donc être capable de sélectionner certaines images en priorité sur lesquelles elle doit faire des propositions de classement. Il peut exister plusieurs critères de sélections intéressant un utilisateur comme par exemple, les plus "faciles" ou les plus "difficiles" à classer. Nous pourrions élaborer la partie analyse de la connaissance et la partie sélection d'images de manière indépendante, mais nous avons choisi de concevoir ces deux parties dans un même formalisme pour plus de cohérence et pour éviter tout problème d'échelle.

Une interface pour aider à progresser plus vite

Une interface graphique est naturellement nécessaire pour la communication entre la machine et l'utilisateur, car le but de ces travaux est de proposer un prototype qui puisse être testé en situation réelle. De plus, une bonne conception des modalités d'interaction et la manière de représenter les données analysées peuvent aider l'utilisateur à progresser plus efficacement et rapidement dans sa tâche de structuration.

Le plan du manuscrit est le suivant :

Chapitre 2 : Ce chapitre détaille plus précisément les besoins et les concepts que nous devons développer pour mettre en place un système de structuration à partir de quelques exemples d'organisation de différentes collections d'images. Nous présentons alors l'architecture générale du système se divisant en 3 modules chacun développé dans les chapitres suivants.

Chapitre 3 : Il concerne la modélisation et la synthèse automatique de la connaissance à partir d'extractions de paramètres sur les contenus visuels pour associer des images dans des groupes. En parcourant différentes méthodes de classification d'images par le contenu, nous identifions quels types d'approche peuvent respecter des contraintes fortes sur la manière de modéliser la connaissance à partir de distances entre descripteurs visuels. Puis, nous présentons notre méthode pour modéliser cette connaissance dans le cadre du formalisme de fusion d'informations du Modèle des Croyances Transférables.

Chapitre 4 : Ce chapitre est consacré à la sélection active d'images. Nous partons de l'hypothèse que, lorsqu'une collection possède beaucoup d'images et qu'il est impossible pour l'utilisateur de les traiter toutes à la fois, il est intéressant de classer certaines images en priorité. En classant ces images, il est alors possible d'améliorer la modélisation de la connaissance sur les classes d'une part, et d'aider l'utilisateur à mieux se représenter les groupes d'images qu'il est en train de constituer d'autre part. Nous nous inspirons de techniques d'apprentissage actif pour proposer nos propres méthodes de sélection active, toujours avec le même formalisme de fusion d'informations.

Chapitre 5 : Ce chapitre décrit comment le système dialogue avec l'utilisateur. Nous présentons théoriquement la manière dont le système peut proposer des classements automatiques des images à l'utilisateur. Puis, nous présentons l'interface homme-machine développée au cours de cette thèse. Elle permet à l'utilisateur de contrôler complètement la manière dont la connaissance doit être modélisée pour satisfaire différentes contraintes de classements, et comment concrètement il peut gérer les propositions de classements automatiques.

Chapitre 6 : Ce dernier chapitre est dédié aux expérimentations et évaluations. Dans un premier temps, la partie automatique est évaluée seule sur des critères de performances classiques de classification. Puis, dans un second temps, les différentes méthodes de sélection d'images proposées sont expérimentées et analysées sur différentes collections d'images. Une évaluation qualitative avec une documentaliste de l'INA est ensuite présentée. La dernière partie concerne différents cadres applicatifs dont le tri de collections de photographies personnelles et une exploitation de l'outil pour la structuration et la navigation dans les vidéos.

SYSTÈME DE STRUCTURATION DE COLLECTIONS D'IMAGES

Description du contenu

1. Contexte applicatif	16
1.1. Homogénéité visuelle de collections d'images	16
1.2. Hétérogénéité visuelle de collections d'images	16
1.3. Diversité des objectifs de structuration de collections d'images	18
1.4. Besoin du multi-étiquetage	18
2. Description générale du système proposé	19
2.1. Modélisation et synthèse de la connaissance	20
2.2. Sélection active d'images	21
2.3. Interface homme-machine	22
2.4. Le retour de l'expertise	22

Nous avons défini précédemment la structuration de collection d'images comme une extension générale de la catégorisation englobant la catégorisation classique de scènes ou d'objets en indexation multimédia, et ainsi que la catégorisation subjective et personnelle. Dans le premier cas, une catégorie est illustrée par des images contenant le même type d'objet (véhicule, maison, animaux...) ou le même type de scène (intérieur/extérieur, paysage, portrait...), reconnaissable par n'importe quel utilisateur. Dans le cas de la catégorisation subjective, une catégorie définie par un utilisateur devient alors plus difficile à percevoir pour un autre utilisateur.

Pour désigner un groupe d'images cohérent sémantiquement pour un utilisateur, nous utiliserons le terme de "classe" au lieu de catégorie, puisque ce sont les aspects théoriques qui vont être présentés. De plus, le terme "classe" est plus générique et permet d'éviter toute confusion avec le cadre rigide que sous-entend la catégorisation de scènes ou d'objets.

Par ailleurs, une étiquette désigne en général un identifiant de classe dans une méthode de classification. Par abus de langage, nous utiliserons indifféremment "étiquette" et "classe" pour nommer un groupe d'images perçu comme étant homogène par un utilisateur. A la fin de la structuration d'une collection d'images, l'utilisateur peut valoriser ses classes, en transformant par exemple les étiquettes en un ensemble de mots-clés permettant par la suite de retrouver plus facilement les images.

1. Contexte applicatif

Les collections d'images peuvent avoir des contenus visuels très variés et les objectifs applicatifs peuvent être différents. Selon ces différents contextes, les collections d'images sont plus ou moins faciles à organiser. Différents types de traitements de collections d'images sont présentés ci-dessous.

1.1. *Homogénéité visuelle de collections d'images*

Un premier type de collection d'images, pouvant être qualifié d'homogène, se prête particulièrement bien à la catégorisation d'images par scènes ou objets. En effet, dans ces collections, les membres d'une même classe présentent à la fois un sens commun et des caractéristiques visuelles communes, ce qui donne une certaine homogénéité des contenus.

Dans ces collections, la composition de l'image est rarement prise en compte : soit le sujet photographié domine le sens associé à l'image, soit c'est l'ambiance globale d'une image qui attire l'attention de l'utilisateur. Généralement, l'interprétation des images n'est pas particulièrement ambiguë : tout utilisateur perçoit immédiatement dans quelles classes peuvent être associées les images.

Un exemple de collection d'images visuellement homogène est présenté dans le tableau 2.1. Cette collection est tirée de la base Corel souvent utilisée pour comparer l'efficacité de méthodes de classifications par la catégorisation d'objets et de scènes [VSWB06], [BMnM07]. Dans cet exemple, il est fort probable que la majorité des utilisateurs distinguent trois classes pouvant être explicitement rattachées respectivement aux mots-clés "bus", "chevaux" et "fleurs".

1.2. *Hétérogénéité visuelle de collections d'images*

L'homogénéité visuelle d'une collection d'images n'est pas garantie. Il est possible que des petits groupes d'images soient très similaires visuellement et sémantiquement pour un utilisateur. Mais une même classe peut contenir des groupes d'images visuellement très différents entre eux, bien que toutes les images soient interprétées similaires par un utilisateur.

Le tableau 2.2 illustre une structuration d'un extrait d'une collection d'images de la Photothèque de l'INA. Dans cet exemple, un utilisateur a identifié trois classes représentant respectivement des images d'un présentateur vedette photographié seul, des images du plateau en cours de tournage et des images de préparatifs des scènes à tourner. Certaines images sont prises sous le même point de vue et représentent donc la même scène ou les mêmes objets photographiés. Pourtant, dans cet exemple, l'utilisateur a préféré orienter la structuration sur la présence ou non du présentateur, le plateau en cours de tournage et les préparations des scènes, indépendamment du fait que les caractéristiques visuelles soient communes ou non.










Etiquette	Annotation	Exemples		
		1	2	3
A	Bus			
B	Chevaux			
C	Fleurs			

TABLE 2.1 : Extrait d'une collection d'images "homogène" : les images peuvent être facilement regroupées en trois classes visuellement et sémantiquement homogènes pour tout utilisateur. Chaque image est clairement une illustration sans ambiguïté d'une classe.












Etiquette	Annotation	Exemples			
		1	2	3	4
A	Gilbert Bécaud				
B	Plateau				
C	Préparations				

TABLE 2.2 : Organisation d'une collection d'images visuellement hétérogène. Les classes possèdent chacune des images ayant des contenus visuels dissemblables. De plus, certaines images aux contenus visuels proches ne sont pas associées dans les mêmes classes par l'utilisateur, comme par exemple les images A3 et C2 ou les images B2 et C4, ou encore A1 et C1. Cette structuration est justifiée par le choix de l'utilisateur de distinguer les photographies avec Gilbert Bécaud, et celles pendant et hors tournage des scènes.

1.3. Diversité des objectifs de structuration de collections d'images

La structuration d'une collection d'images peut être subjective. Il est alors possible que les images d'une même classe soient visuellement très différentes, et qu'à l'inverse des images membres de classes distinctes soient visuellement proches.

Cependant, la subjectivité n'est pas la seule cause conduisant à cette hétérogénéité visuelle des classes. Les objectifs applicatifs peuvent également amener à cette hétérogénéité. En effet, l'utilisateur peut avoir un cahier des charges à respecter et il doit alors contraindre son interprétation personnelle.

Par exemple, une deuxième structuration est effectuée sur la collection d'images précédente (tableau 2.3) dans un contexte applicatif différent. Dans ce deuxième cas, la structuration est clairement orientée par des objectifs juridiques (droits d'auteurs, droits à l'image). L'utilisateur a donc identifié les photographies qui contiennent des éléments qui peuvent éventuellement nécessiter l'ouverture de droits par des services juridiques pour de prochaines exploitations.

1.4. Besoin du multi-étiquetage

Certaines collections d'images se prêtent facilement à de l'étiquetage "simple". Par exemple, dans l'extrait précédent de la base Corel, les photographies mettent en avant un objet photographié : un cheval, un bus ou une fleur. Un utilisateur aura certainement tendance à vouloir associer une image à une seule de ces étiquettes.

Cependant, de nombreuses images peuvent souvent être "multi-étiquetées" : les photographies sont la plupart du temps composées de différentes régions (ciel, herbe, mer, route...), ou d'objets (voiture, table, maison, ...) ou encore de personnes. Les images peuvent alors être potentiellement associées à plusieurs étiquettes.

De plus, nous pouvons souligner que tous les utilisateurs n'abordent pas le multi-étiquetage de la même manière : certains utilisateurs préféreront plutôt l'approche d'étiquetage simple, c'est-à-dire qu'ils prêtent attention surtout à l'objet photographié. Ils supposent que, une fois la collection organisée, pour retrouver une image, ils doivent se focaliser avant tout sur le sujet ou l'ambiance photographiée. D'autres utilisateurs préféreront l'approche multi-étiquette en pensant que les différentes régions des images méritent d'être associées chacune à leur propre étiquette.

Les tableaux 2.4 et 2.5 illustrent comment deux utilisateurs peuvent structurer une même collection d'images avec une approche d'étiquetage simple et multiple. Dans le premier tableau 2.4, l'utilisateur prête attention uniquement au type de scène (plages, villes-monuments, portraits) qu'il utilise pour structurer la collection. Dans le deuxième tableau 2.5, un second utilisateur prête plutôt attention à la composition des photographies pour attribuer parfois plusieurs étiquettes à certaines images. Par exemple, la photographie A3 et C2 représentant un portrait sur un fond de plage est associée aux deux classes correspondantes.



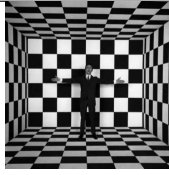
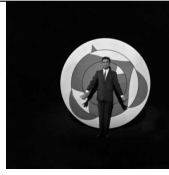


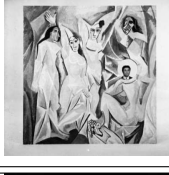




Etiquette	Annotation	Exemples			
		1	2	3	4
A	Gilbert Bécaud				
B	Peintures				
C	Gilbert Bécaud et peintures				
D	Autres				

TABLE 2.3 : Une structuration d'une collection d'images visuellement hétérogène selon un deuxième utilisateur sur le même extrait de collection d'images de la figure 2.2. L'utilisateur est contraint d'identifier les photographies nécessitant l'ouverture de droits. Il a regroupé toutes les photographies contenant des représentations de peintures célèbres et Gilbert Bécaud. L'image C2 contient à la fois une reproduction des *Demoiselles d'Avignon* de Picasso avec une incrustation du visage de Gilbert Bécaud dans le tableau. La dernière classe regroupe toutes les images ne posant pas de problème de droits a priori. Les contenus d'une même classe sont visuellement hétérogènes, et certaines images similaires ne se retrouvent pas dans les mêmes classes comme par exemple les images A4 et B2.

2. Description générale du système proposé

Nous avons décrit la difficulté qu'implique la variété des collections d'images pouvant être traitées, et comment des utilisateurs peuvent appréhender de différentes manières l'association d'images aux classes. Nous proposons un système avec une architecture souple permettant de traiter ces différents cas de figure.

Le système comporte deux grandes parties :

1. Un cœur complètement automatique permet de modéliser de la connaissance à partir de caractéristiques extraites des contenus visuels (couleur, texture, ...). L'objectif de cette partie est de quantifier les appartenances des images aux classes à partir de ces extractions automatiques. La difficulté est alors de pouvoir distinguer des mesures d'appartenances sur une et/ou plusieurs classes à la fois.
2. Une partie semi-automatique permet la collaboration entre l'utilisateur et la machine. L'utilisateur



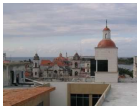








Etiquette	Annotation	Exemples				
		1	2	3	4	5
A	Mer, Plage					
B	Villes, Monuments					
C	Portraits					

TABLE 2.4 : Exemple d'étiquetage simple sur un extrait d'une collection de photographies personnelles. L'utilisateur considère qu'une image possède un sens dominant. Pour lui, une image est une illustration d'une seule étiquette. Par exemple, la photographie C2 est identifiée comme un portrait, indépendamment des autres éléments qui composent l'image comme le fond de mer.

peut indiquer à la partie automatique de lui fournir un certain type d'images à étiqueter en priorité. La partie automatique propose alors des étiquettes sur cette sélection d'images que l'utilisateur valide ou non.

Nous proposons de diviser le système en 3 modules (figure 2.1) avec une boucle de retour :

1. une modélisation et synthèse de la connaissance,
2. une sélection active d'images,
3. et une interface homme-machine.

2.1. Modélisation et synthèse de la connaissance

Ce premier module consiste à tirer un maximum d'informations sur les images. Le module s'appuie sur les images qui ont été précédemment étiquetées et "rangées" dans des classes par l'utilisateur. A partir de descripteurs sur les contenus visuels des images, le module établit un état de connaissance, c'est-à-dire un ensemble de degrés d'appartenance des images non étiquetées aux différentes classes. La difficulté est alors de quantifier des degrés d'appartenance cohérents avec des étiquetages pouvant convenir à l'utilisateur. Or, il est nécessaire de gérer les imperfections des informations fournies par les extracteurs et les mesures de dissimilarités employées entre ces descripteurs. La modélisation de la connaissance peut être complexe et délicate. C'est une étape théorique et cruciale car une modélisation maladroite peut amener à des performances de classification peu satisfaisantes.

Étiquette	Annotation	Exemples				
		1	2	3	4	5
A	Mer, Plage					
B	Villes, Monuments					
C	Portraits					

TABLE 2.5 : Exemple de multi-étiquetage sur le même extrait de la collection précédente (tableau 2.4). Un second utilisateur est plus attentif à la composition des images (sujet photographié, fond, éléments de décors. . .). Les images peuvent alors posséder une ou plusieurs étiquettes. Par exemple, l'image A1 est associée à une seule étiquette car le contenu visuel et sémantique est relativement homogène. Par contre les images A3 et C2 représentent la même photographie associées à deux classes distinctes, prenant en compte le sujet photographié (une personne) et le fond (la mer). De même la photographie représentant les quais d'une ville (A4 et B3) est associée à la fois à une classe "ville" et "mer". Enfin, l'images en B4 et C5 est associée également aux deux étiquettes "ville" et "portrait".

2.2. Sélection active d'images

Un utilisateur ne peut pas contrôler simultanément des centaines de proposition de classements d'images à la fois. Le but de ce module de sélection active d'image est de faciliter la tâche de l'utilisateur, en lui proposant une sous-partie des images à étiqueter. L'utilisateur doit pouvoir disposer alors de stratégies pouvant l'aider de différentes manières. Par exemple l'utilisateur peut demander au module une sélection des images :

- "faciles" à traiter car les contenus sont proches d'images déjà classées,
- "difficiles" car trop d'étiquettes sont proposées à la fois,
- ou "difficiles" car aucune étiquette n'est proposée.

L'utilisateur peut changer à n'importe quel moment de type de stratégie. Le module fournit en sortie une liste globale des images non étiquetées triées par le critère de sélection de la stratégie courante. Par exemple, l'utilisateur peut demander au module de lui fournir la liste des images ordonnées de la plus facile à la plus difficile à étiqueter. Il peut alors traiter en priorité les premières images de cette liste.

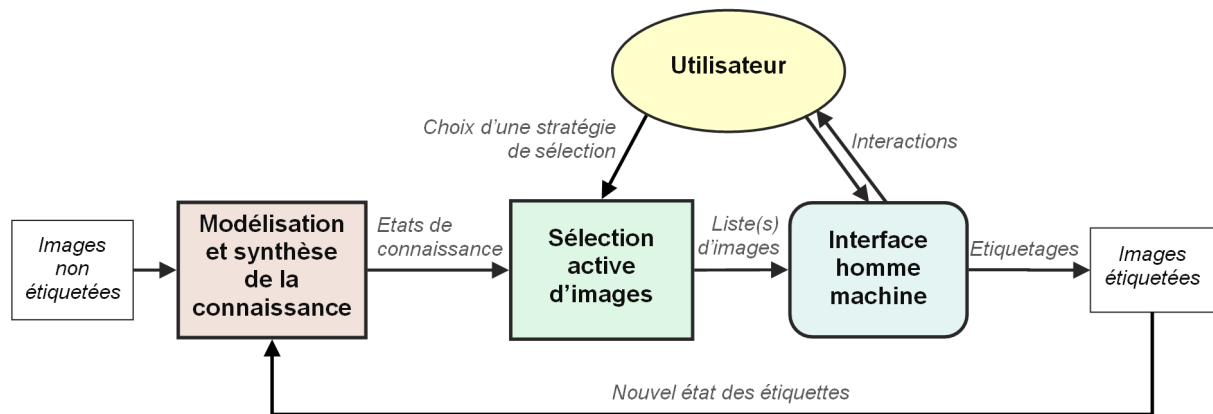


FIGURE 2.1 : *Vue générale du système proposé. Une collection d'images est constituée d'images non étiquetées. Le module de modélisation de la connaissance convertit des descriptions sur les contenus visuels en différentes mesures d'appartenance aux classes. Puis cette analyse est utilisée dans un module de sélection active d'images. L'utilisateur peut demander à ce module différents types d'images à sélectionner en priorité. Le module fait alors des propositions de classement d'images que l'utilisateur valide à travers une interface homme-machine. Chaque nouvelle image étiquetée est alors prise en compte par le module de modélisation et synthèse de la connaissance pour raffiner la connaissance des images encore non étiquetées, améliorant ainsi la sélection active d'images et les propositions de classements automatiques.*

2.3. Interface homme-machine

Le dernier module concerne le dialogue entre l'utilisateur et le système. Dans un premier temps, la manière dont les propositions automatiques d'étiquettes sur des images sont effectuées est présentée théoriquement. Puis dans un second temps, tous les aspects interfaces homme machine sont présentés :

- Divers contrôles permettant à l'utilisateur d'autoriser les cas de multi-étiquetage, d'activer des options de rejets pour mettre de côté certaines images, de lancer une animation classant automatiquement les images. . .
- Une représentation graphique intelligible et épurée est proposée afin d'aider l'utilisateur à percevoir facilement si une étiquette proposée automatiquement sur une image lui convient.
- Cette représentation est conçue pour autoriser un maximum d'interactions directement dans la vue en prenant en compte des critères ergonomiques.

2.4. Le retour de l'expertise

A chaque fois qu'une image est étiquetée dans une ou plusieurs classes, l'état de connaissance sur la constitution visuelle des classes peut être affiné puisqu'une illustration supplémentaire est ajoutée. Ce nouvel état de connaissance est réinjecté dans le premier module de modélisation et de synthèse de la connaissance.

L'étiquetage d'une seule image peut fortement modifier l'état de connaissance. Par exemple, des images non étiquetées considérées précédemment comme étant "difficiles" à classer peuvent devenir "faciles" si

la nouvelle image étiquetée est représentative d'un voisinage visuellement dense. A l'opposé, des images non étiquetées précédemment jugées "faciles" peuvent devenir "difficiles" si des images du voisinage ont été étiquetées dans différentes classes.

Nous avons présenté les différents modules constituant l'architecture du système proposé. Chaque module est détaillé dans les prochains chapitres. Nous abordons dans un premier temps la partie automatique de modélisation et la synthèse de la connaissance.

MODÉLISATION ET SYNTHÈSE DE LA CONNAISSANCE

Description du contenu

1. Introduction	26
1.1. Position du problème	26
1.2. Contraintes théoriques	26
1.3. Méthodes de classification automatique	28
1.3.1. Gestion de structures de données potentiellement complexes	29
1.3.2. Méthodes de classification multi-étiquette	31
1.3.3. Gestion de nouveaux contenus visuels	34
1.4. Choix d'une approche théorique	35
2. Extraction d'information, description de contenu visuel	36
2.1. Descripteur global de couleurs	36
2.2. Descripteur global des orientations	37
3. Manipulation de descriptions visuelles pour le classement d'images	38
4. Modèle des Croyances Transférables	41
4.1. Représentation de la connaissance	42
4.2. Combinaison	45
4.3. Prise de décision	46
5. Modélisation de la connaissance	46
5.1. Règles de représentation de la connaissance à partir d'une distance	46
5.2. Conversion numérique - symbolique	48
6. Fusion de témoignages au sein d'une classe	51
6.1. Combinaison de voisins	51
6.2. Réévaluation de distributions de masses	53
7. Fusion de témoignages pour l'ensemble des classes	55
7.1. Espaces de discernement	55
7.2. Fusion multi-classe	58
7.3. Fusion multi-descripteur	60
8. Conclusion	63

1. Introduction

1.1. Position du problème

Dans ce chapitre nous nous focalisons sur la première étape de modélisation et de synthèse de la connaissance (voir le cadre rouge dans la figure 3.1). Le problème est le suivant : un utilisateur a manuellement regroupé et étiqueté quelques images en différentes classes cohérentes selon son expertise. Le but de ce module est alors de modéliser un état de connaissances sur toutes les images encore non étiquetées en établissant un ensemble de degrés d'appartenance aux différentes classes. En fonction des contraintes théoriques définies ci-dessous, les degrés d'appartenance doivent fournir une analyse détaillée sur les états potentiels des images.

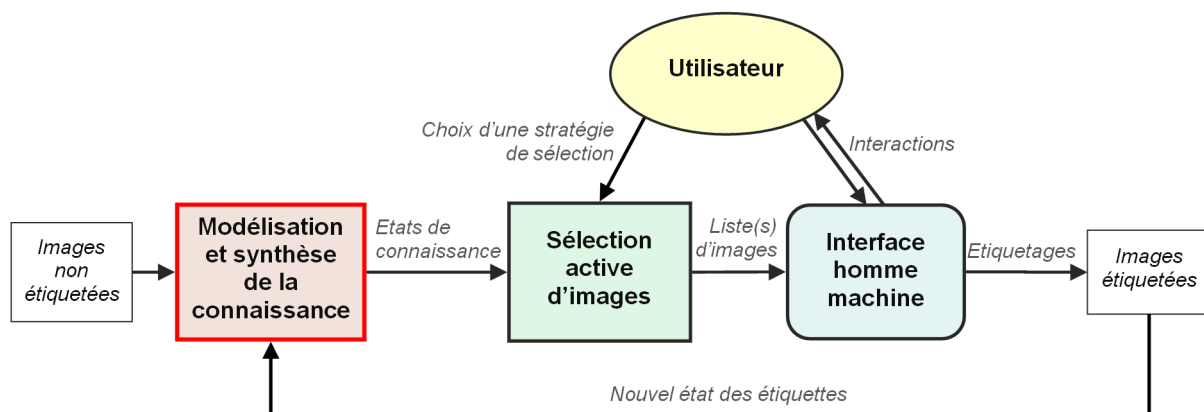


FIGURE 3.1 : Schéma général de l'architecture du système de structuration de collection d'images proposé. Le module de modélisation et de synthèse de la connaissance analyse les images non étiquetées et celles qui ont été précédemment étiquetées par l'utilisateur. Cette analyse permet d'établir un état de la connaissance quand à l'appartenance des images non étiquetées aux classes. Cette étape prépare les données pour le prochain module de sélection active d'images et aux propositions automatiques d'étiquettes dans l'interface homme-machine.

1.2. Contraintes théoriques

Les contextes applicatifs exposés précédemment ont explicité les différents traitements auxquelles sont soumises les collections d'images, notamment en fonction de l'hétérogénéité visuelle potentielle des classes et des besoins éventuels de multi-étiquetage. Ces contraintes ont un impact sur le choix de l'approche théorique à développer pour établir la modélisation et synthèse de la connaissance.

Traiter des collections d'images sans a priori sur l'homogénéité visuelle des contenus

En indexation multimédia, la gestion de documents par le contenu se fait généralement à partir d'extractions bas niveau sur diverses informations (couleurs, textures, formes, ...) mises sous forme vectorielle. Des mesures de dissimilarité entre ces vecteurs descripteurs, comme par exemple la distance

euclidienne, permettent de quantifier la proximité entre les contenus visuels. Cependant, les rapprochements des contenus ne sont pas obligatoirement en adéquation avec la sémantique perçue et exprimée par un utilisateur. Ce problème de *fossé sémantique* est bien identifié dans la communauté [SWS⁺00], [DV03]. Deux images proches en termes de descriptions visuelles peuvent être perçues comme étant très éloignées sémantiquement. Inversement, deux descriptions visuelles éloignées peuvent correspondre à une situation où les contenus sont proches au sens d'un utilisateur.

D'autre part, selon ses objectifs applicatifs, un utilisateur peut souhaiter regrouper de manière subjective les images. Or, ces regroupements peuvent être difficiles à percevoir par un autre utilisateur si ce dernier ne connaît pas les objectifs applicatifs. Cette subjectivité vient ajouter une difficulté supplémentaire au problème de fossé sémantique. Le résultat de ces deux contraintes est que, potentiellement, les membres d'une même classe définie par un utilisateur peuvent être très éparpillés dans l'espace des descripteurs. Il est alors risqué d'avoir un *a priori* sur l'homogénéité visuelle des classes dans une méthode de classification.

Exprimer l'appartenance à plusieurs classes à la fois

Le module de modélisation de la connaissance doit pouvoir exprimer aussi bien des situations d'étiquetages simples et de multi étiquetages. L'étiquetage simple correspond au cas où un utilisateur souhaite classer chaque image de manière exclusive dans une seule classe. Le module doit donc fournir pour toute image non étiquetée un ensemble de mesures d'appartenance pour toutes les classes, soit une par classe. Ainsi, l'état de connaissance doit renseigner sur le fait qu'une classe se distingue des autres pour y classer une image.

Cependant, l'utilisateur peut vouloir exprimer des cas où une même image peut appartenir de manière non exclusive à plusieurs classes. Dans ce cas, il est nécessaire d'exprimer des cas multi-étiquette et la difficulté est alors de formaliser l'expression de mesures d'appartenance sur une ou plusieurs classes à la fois. D'un point de vue théorique, il faut notamment considérer s'il est préférable de s'appuyer uniquement sur des mesures représentant l'appartenance à chacune des classes prises individuellement pour en déduire l'appartenance multiple à plusieurs classes, ou s'il faut formaliser des mesures spécifiques pour ces cas.

Mesurer la non appartenance à toutes les classes

Certaines images non étiquetées peuvent avoir des descriptions visuelles très éloignées de toutes les images étiquetées. Pour avoir une modélisation complète de la connaissance, il peut être intéressant d'exprimer le cas où aucune classe ne semble pertinente pour une image non étiquetée. En effet, ce type d'information peut être utile pour identifier des contenus visuels différents des membres des classes connues. Ces images peuvent être alors utilisées pour définir de nouvelles classes, ou bien pour être associées à des classes existantes afin de compléter les différents aspects visuels d'une même classe.

Ces différentes contraintes théoriques (hétérogénéité visuelle des classes, multi-étiquetage et non appar-

tenance aux classes), vont nous guider dans nos choix d'approches théoriques pour mettre en place une méthode de modélisation et d'analyse de la connaissance, en s'inspirant des techniques de classification.

1.3. Méthodes de classification automatique

Les objectifs du module de modélisation de la connaissance s'apparentent à un problème de classification automatique, dans le sens où le but est de répartir les images en différentes classes. Cependant, la phase de décision, où une image est réellement étiquetée, sera accomplie dans le dernier module d'interface homme-machine. L'unique objectif du module décrit dans ce chapitre est la modélisation et de synthèse de la connaissance sur toutes les images encore non étiquetées en établissant un ensemble de degrés d'appartenance aux différentes classes.

La bibliographie sur les méthodes de classification automatique est très vaste et il existe de nombreuses techniques pour traiter ce problème [Bis06], [DHS00] : les fonctions discriminantes linéaires, les réseaux de neurones, les méthodes basées sur les k plus proches voisins, les machines à vecteur supports...

Le choix d'une approche vis-à-vis d'une autre doit prendre en compte avant tout le type de problème de classification considéré, (voir figure 3.2) car ils ne soulèvent pas les mêmes difficultés théoriques. La classification bi-classe est de loin le problème le plus étudié car beaucoup de situations peuvent se réduire à un problème dichotomique d'appartenance ou non à une classe. Dans le cadre de classification d'images, cette approche est bien adaptée par exemple à la recherche d'images par le contenu où l'objectif est de déterminer automatiquement si une image correspond positivement ou non à une requête.

La classification multi-classe généralise le cas bi-classe. Intuitivement, on peut supposer que plus le nombre de classes est élevé, plus il est risqué de mal classer les données. Si certaines méthodes de classifications sont particulièrement adaptées au cas bi-classe comme les Machines à Vecteurs Support (SVM), leur extension théorique au cas multi-classe peut être difficile. Par exemple, toujours pour le cas des SVM, des travaux proposent une extension au cas multi-classe [DK05], [TC01], mais le passage d'une approche binaire vers la discrimination à catégories multiples n'est pas trivial car leur emploi est difficile d'un point de vue théorique et les performances ne se distinguent pas significativement des méthodes de décomposition impliquant des SVM bi-classe [Gue07].

En effet, plusieurs travaux abordent le problème multi-classe sous la forme de multiples problèmes de classification bi-classe. Il existe plusieurs méthodes de décomposition comme par exemple l'approche *one against all* consistant à construire autant de classifieurs binaires du même type qu'il existe de classes indépendantes, ou encore l'approche *one against one* consistant plutôt à confronter 2 par 2 des classifieurs binaires associés à chacune des classes. Le débat concerne alors l'intérêt d'une approche vis-à-vis de l'autre. L'approche *one against all* serait la plus précise [RK04] alors que la stratégie *one against one* permettrait un apprentissage plus rapide et la gestion d'un nombre plus important de classes [MCS06].

Ces approches par décomposition soulignent la difficulté de formaliser un modèle de classification réellement multi-classe. Ce problème de formalisation se retrouve également pour le cas de la classification

multi-étiquette, moins fréquemment étudiés. Au problème initial multi-classe, s'ajoutent des hypothèses d'appartenance multiple de certains éléments dans plusieurs classes à la fois. Les classes ne peuvent être alors supposées comme étant mutuellement exclusives, ce qui complexifie les difficultés théoriques.

Le module de modélisation de la connaissance doit permettre de faire face à ces différentes situations. Un utilisateur doit pouvoir traiter 2 ou plusieurs classes simultanément et il doit pouvoir étiqueter des images dans plusieurs classes s'il le souhaite.

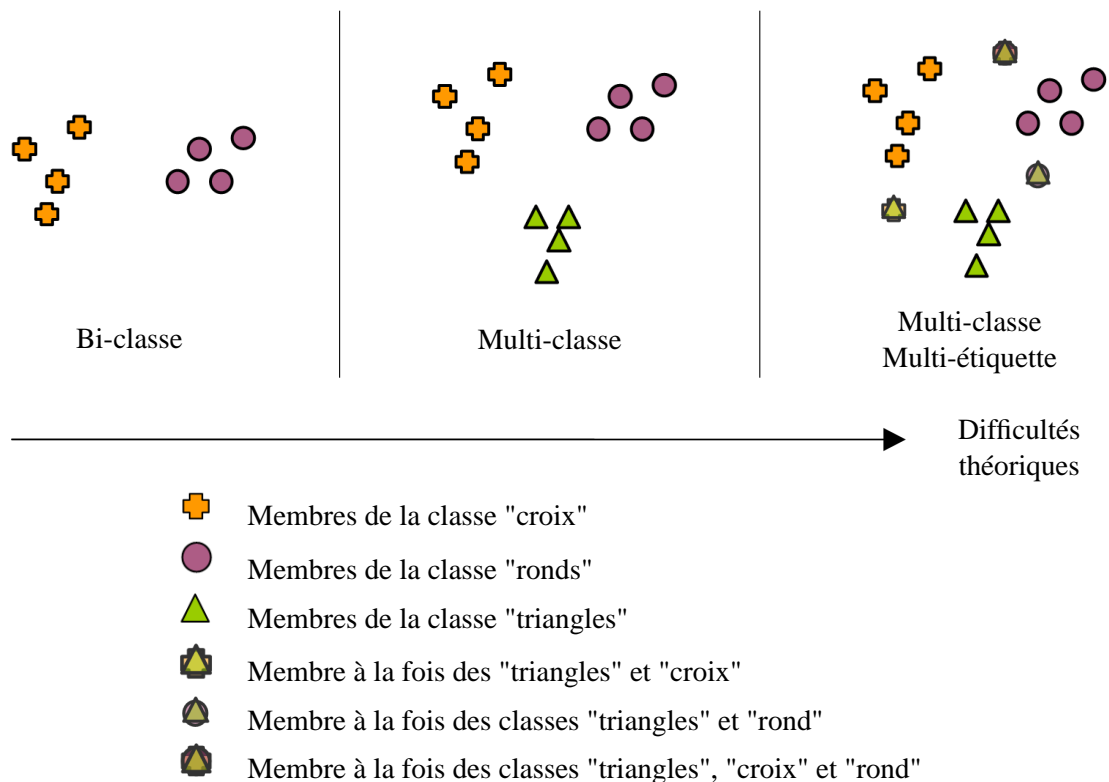


FIGURE 3.2 : Différents problèmes de classification, du plus simple ou plus difficile théoriquement. Les membres des classes sont représentés dans un espace à 2 dimensions. Dans le cas de la classification bi-classe, l'objectif est de retrouver tous les éléments qui appartiennent à une classe ou non. Le cas multi-classe généralise le problème bi-classe à plusieurs classes. La classification multi-classe et multi-étiquette complexifie le problème en autorisant l'appartenance de certains échantillons à plusieurs classes simultanément.

1.3.1. Gestion de structures de données potentiellement complexes

Le schéma figure 3.2 donne des représentations simplistes de classes pour différents problèmes de classification. Dans ces exemples, les classes sont facilement distinguables les unes des autres. En effet, les membres d'une même classe sont proches entre eux, alors que les membres de deux classes sont éloignés. Ces structures de données homogènes favorisent et facilitent la classification automatique.

Cependant, cette structure homogène des données n'est pas garantie sur des données réelles comme

l'illustre la figure 3.3 reprenant les trois types de classification dans des situations plus réalistes. Cette propriété des données s'explique particulièrement dans le cas de la classification des images avec les effets du fossé sémantique. Concrètement, dans un espace de description des images sur lequel travaille un algorithme de classification, la proximité entre 2 images ne garantit pas intégralement le fait qu'elles appartiennent à la même classe. De manière opposée, 2 images éloignées peuvent appartenir à une même classe.

Il est préférable de ne pas avoir d'*a priori* sur les structures des classes. Il est possible que, au sein d'un même jeu de données, certaines classes soient à la fois homogènes visuellement et sémantiquement, et que d'autres classes soient complètement éparpillées dans l'espace des descripteurs.

Pour gérer des structures de données potentiellement complexes, il est possible de faire appel à des techniques d'apprentissage automatique pour permettre au système de s'adapter en fonction des intentions de l'utilisateur. Il existe principalement les approches non, semi et totalement supervisées. Les approches non supervisées [Gha04], [DHS00] utilisent exclusivement des données non étiquetées pour les séparer en différents groupes homogènes. Ce type d'approche n'est pas adapté à notre problème car l'utilisateur est indispensable pour identifier des classes parfois extrêmement complexes. Les approches semi-supervisées [Zhu05] sont intéressantes car elles exploitent à la fois des données *a priori* fiables, les échantillons étiquetés par l'utilisateur expert, et les données non étiquetées. Cette approche bénéficie ainsi de nombreuses données pour établir de manière statistique l'appartenance des échantillons aux classes. Cependant, il peut être risqué de donner des étiquettes *a priori* aux images non étiquetées pour estimer des fonctions d'appartenance.

L'approche supervisée est la moins risquée puisque l'utilisateur expert ou "oracle" permet de s'appuyer sur des données *a priori* fiables pour pouvoir estimer l'appartenance aux classes des échantillons non étiquetés. Les méthodes de classification supervisées les plus courantes [DHS00] utilisent l'estimation bayésienne ou les SVM ou encore les k plus proches voisins (knn). L'estimation bayésienne peut être très efficace si les classes sont homogènes, mais elle est plutôt mal adaptée aux structures de données trop complexes. L'emploi des SVM donne de bonnes performances et permet de s'adapter aux structures de données complexes. Mais les SVM peuvent entraîner un coût de calcul élevé notamment si la structure est trop complexe : dans le cas extrême, il peut être nécessaire de définir autant de vecteurs supports qu'il y a d'échantillons. L'approche des k plus proches voisins, malgré sa simplicité théorique, permet d'atteindre généralement de bonnes performances et est bien adaptée aux structures complexes. En revanche, cette approche implique une certaine lourdeur puisque que les classes sont représentées par tous leurs membres.

La question est maintenant de savoir si ces techniques d'apprentissage supervisé peuvent facilement se généraliser au cas multi-étiquette.

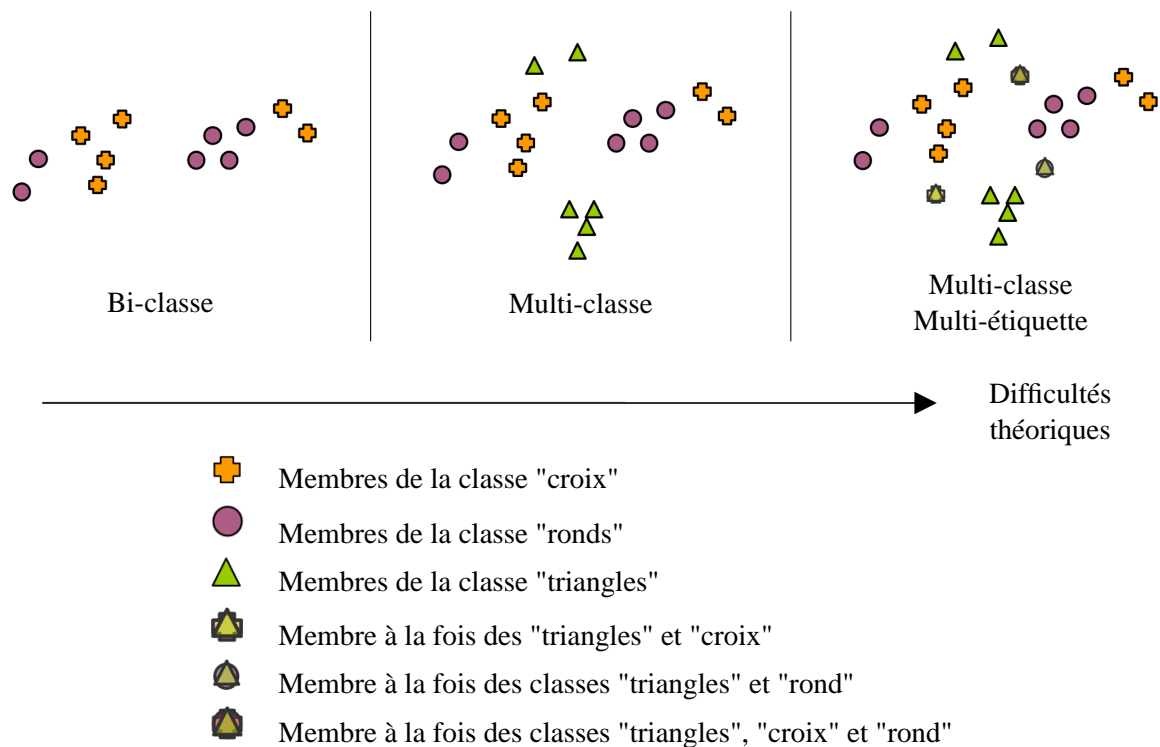


FIGURE 3.3 : Les différents problèmes de classification (voir figure 3.2) sont reconsidérés dans des situations plus réalistes : les membres d'une même classe peut être éclaté en plusieurs groupes éloignés. En conséquence, il n'y a pas d'homogénéité des données et les différentes classes peuvent éventuellement se superposer.

1.3.2. Méthodes de classification multi-étiquette

Les travaux sur la classification multi-étiquette d'images sont relativement récents [TKV08]. Pourtant, le multi-étiquetage est assez naturel en classification d'images, notamment pour la catégorisation de scène. Par exemple, il est facile de concevoir qu'une image de plage avec des bâtiments puisse être étiquetée dans 2 classes à la fois (plage et scène urbaine).

Dans la littérature, il est souligné que toutes les bases de données n'ont pas la même proportion d'échantillons multi-étiquetés. Dans le cas de collections d'images, un grand nombre de documents peut avoir une seule étiquette, tandis que la proportion d'images partageant plusieurs étiquettes peut être faible (voir l'exemple tableau 3.1). La classification multi-étiquette est donc d'autant plus difficile qu'il y a peu de données multi-étiquetées, car il y a peu d'information disponible pour établir des statistiques d'appartenance multiple à plusieurs classes.

Dans [TK07], les auteurs identifient différentes méthodes de classification multi-étiquette en distinguant le classifieur mis en œuvre (knn, SVM, bayésien, ...) d'une opération de transformation sur les étiquettes considérées. Cette opération de transformation revient à exprimer le problème de classification multi-classe et multi-étiquette en un problème multi-classe seul. Ils énoncent notamment 3 opérations de transformation d'étiquettes couramment utilisées :

étiquettes	nombre d'images
urbain	135
plage	13
personnes	49
urbain+plage	6
urbain+personnes	22
plage+personnes	5
urbain+plage+personne	1
total	231
nombre d'étiquettes à retrouver	266

TABLE 3.1 : Exemple de répartition des étiquettes multiples dans une collection d'images de 231 photographies avec 3 étiquettes de bases. Les répartitions sont déséquilibrées : les images possédant plusieurs étiquettes sont beaucoup moins nombreuses que celles possédant une seule étiquette. L'enjeu du problème de classification est alors d'arriver à retrouver 266 étiquettes pour 231 images.

Echantillon / Etiquette	A	B	C	D
1	X			
2	X	X		
3		X	X	
4	X			X

TABLE 3.2 : Exemple d'un jeu de données multi-étiquette : les échantillons peuvent avoir une étiquette comme l'échantillon 1, ou plusieurs comme pour les trois autres échantillons.

- "PT3" : le problème est transformé en un problème de classification multi-classe en considérant les étiquettes simples et les unions d'étiquettes (voir les tableaux 3.2 et 3.3) comme des étiquettes simples.
- "PT4" : le problème est transformé en un ensemble de problèmes de classification bi-classe où sont considérés uniquement les étiquettes positives et négatives de chaque classes indépendamment des autres (voir le tableau 3.4).
- "PT6" : le problème est transformé en considérant une série de classifieurs bi-classe, à la manière des techniques de boosting, notamment Adaboost [SS00], (voir le tableau 3.5).

Leurs expérimentations sur différents jeux multi-étiquetés, avec des densités d'étiquettes par images de différents ordres, ont mis en évidence que la transformation "PT3" donne globalement les meilleurs résultats en termes de précision (le rapport entre le nombre total d'étiquettes correctes et le nombre total d'étiquettes proposées) et de rappel (le rapport entre le nombre total d'étiquettes correctes et le nombre total d'étiquettes devant être retrouvées). Cette transformation "PT3" possède notamment l'avantage vis-à-vis des autres transformations d'être performante indépendamment du choix de la technique de classification utilisée (les meilleurs résultats sont obtenus avec les classifieurs knn et SVM). Cette étude est très intéressante pour la communauté car, dans la littérature sur le multi-étiquetage, la transformation "PT4" est souvent préférée (cela peut s'expliquer par le fait que beaucoup de travaux se basent sur des techniques de classification bi-classe que l'on tente de généraliser au cas multi-étiquette).

Echantillon / Etiquette	A	A+B	B+C	A+D
1	X			
2		X		
3			X	
4				X

TABLE 3.3 : Transformation du jeu de données 3.2 avec l'opération de transformation "PT3" : les unions des étiquettes de base sont considérées comme des étiquettes à part entière pour pouvoir traiter un problème de classification multi-classe avec étiquetage simple.

Ech/Et	A+	A-	Ech/Et	B+	B-	Ech/Et	C+	C-	Ech/Et	D+	D-
1	X		1		X	1		X	1		X
2	X		2	X		2		X	2		X
3		X	3	X		3	X		3		X
4	X		4		X	4		X	4	X	

TABLE 3.4 : Transformation du jeu de données 3.2 avec l'opération de transformation "PT4" : le problème est traité par un ensemble de classifieurs binaires, un par étiquette. C'est l'approche la plus souvent utilisée.

Echantillon	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
Étiquette	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Poids	+1	-1	-1	-1	+1	+1	-1	-1	-1	+1	+1	-1	+1	-1	-1	+1

TABLE 3.5 : Transformation du jeu de données 3.2 avec l'opération de transformation "PT6" : le problème est traité conjointement par un ensemble de classifieurs binaires "faibles", un pour chaque possibilité d'étiquetage de chaque échantillon exemple. L'étiquetage d'un nouvel échantillon peut alors se faire à partir du signe ou d'un score donné par l'ensemble des sorties de tous les classifieurs.

1.3.3. Gestion de nouveaux contenus visuels

Une image non étiquetée peut posséder un contenu visuel très différent des classes existantes. Cette image peut être considérée de différentes manières (voir figure 3.4) :

- L'image est isolée car elle possède un contenu visuel "exotique". Elle peut être considérée comme du bruit comme le cas des images uni-couleur par exemple.
- L'image n'est pas isolée et elle est représentative d'un voisinage de contenus visuels encore non explorés.

Dans notre cadre applicatif, un utilisateur découvre progressivement le contenu visuel d'une collection qu'il doit organiser. Il est donc nécessaire de permettre l'ajout d'une nouvelle classe si une image ne semble adéquate pour aucune des classes existantes.

De plus, une même classe peut être très hétérogène visuellement, mais cohérent pour un utilisateur pour les raisons évoquées plus haut de fossé sémantique et de subjectivité. Une image représentant un nouveau contenu visuel, n'est pas obligatoirement utilisée pour définir une nouvelle classe : elle peut contribuer à enrichir les différents aspects visuels d'une même classe.

Cette gestion de la nouveauté doit pouvoir se formaliser explicitement dans le module de modélisation et d'analyse de la connaissance de notre système. Par rapport aux besoins multi-classe et multi-étiquette, les techniques de classification basées sur les SVM ou les knn semblent adaptées. Pour gérer la nouveauté, l'emploi d'une technique type SVM est moins facile. En effet, la nature discriminante des SVM impose de considérer des classes prédéfinies. Les techniques à base de SVM se focalisent alors sur les échantillons situés aux frontières entre 2 classes, la marge. Dans sa conception la plus basique, un échantillon situé loin de la marge entre 2 classes, est associé dans la classe du côté où il se trouve.

Or, en considérant la gestion de nouveauté visuelle, il est préférable d'utiliser une approche générative afin de distinguer les frontières "intérieures" et "extérieures" (voir la figure 3.5). En effet, en définissant entièrement les limites des classes, il est alors plus facile d'identifier les échantillons situés loin de toutes les classes.

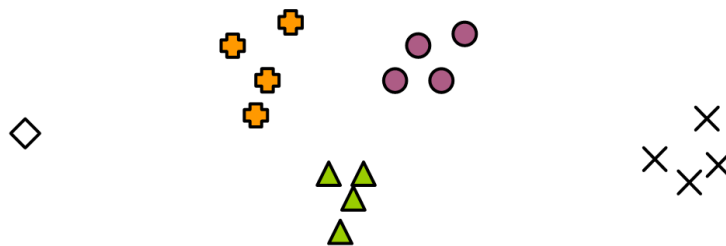


FIGURE 3.4 : Images non étiquetées loin de toutes les images étiquetées : 3 classes de 4 images chacune sont représentées par des croix pleines, des ronds et des triangles dans un espace de description théorique à 2 dimensions. Cinq images encore non étiquetées sont représentées par 4 croix et un losange. L'image non étiquetée représentée par le losange est isolée : il est probable que le contenu visuel ne soit pas représentatif d'une classe. A l'inverse, les images non étiquetées représentées par des croix sont certainement représentatives d'un contenu visuel localement homogène, et peuvent être utilisées pour définir une nouvelle classe ou bien être associées à une des classes existantes.

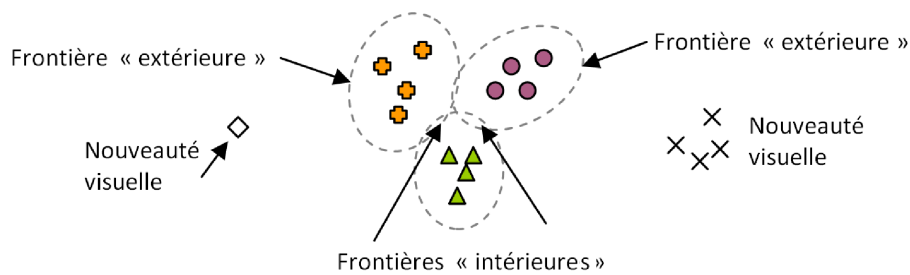


FIGURE 3.5 : La distinction entre frontières "extérieures" et "intérieures" doit être prise en compte pour pouvoir gérer la nouveauté de contenus visuels. Les techniques de classification discriminatives ne permettent pas d'identifier explicitement les frontières extérieures. Pourtant, il est important de pouvoir identifier des échantillons loin des frontières extérieures des classes pour éventuellement définir de nouvelles classes ou compléter les différents aspects visuels d'une classe existante.

1.4. Choix d'une approche théorique

Nous avons associé la modélisation et la synthèse de la connaissance à un problème de classification multi-classe, multi-étiquette et devant prendre en compte la gestion de nouveauté. Les méthodes basées sur les plus proches voisins - knn - semblent être adaptées à nos besoins. En effet, elles permettent une grande souplesse et peuvent se décliner au cas multi-classe. En s'appuyant sur des mesures de dissimilarité entre échantillons, elles permettent d'identifier facilement la nouveauté. De plus, elles peuvent s'employer dans le cas multi-étiquette, tout en étant compétitive avec des techniques plus avancées [TK07].

Cependant, l'emploi des knn doit nous permettre d'établir un état de connaissance riche décrivant aussi bien des cas d'appartenance d'une image à une seule classe, que l'appartenance à plusieurs classes ou l'appartenance à aucune classe. Il est alors préférable d'avoir un seul cadre unifié permettant d'exprimer toutes ces situations.

Le choix d'un cadre formel vis-à-vis d'un autre peut être motivé par la gestion de l'incertitude et de l'imprécision. En effet, les méthodes de type knn s'appuient sur des mesures de dissimilarité entre les échantillons étiquetés et ceux encore non étiquetés. Dans le cas d'images, les descripteurs visuels sont utilisés pour quantifier ces proximités entre images. Or, un descripteur visuel ne peut capturer parfaitement l'intégralité du contenu visuel d'une image. Il est le plus souvent spécialisé pour décrire un aspect du contenu comme les couleurs ou les textures. De plus, les proximités entre les descripteurs images peuvent difficilement répondre aux contraintes de fossé sémantique, et à la subjectivité des regroupements d'images qui seraient effectués par différents utilisateurs sur une même collection d'images.

Notre choix s'est porté sur la théorie des fonctions de croyance, en particulier le Modèle des Croyances Transférables (MCT ou TBM en anglais pour *Transferable Belief Model*). En effet, ce cadre formel propose des outils performants pour gérer l'incertitude et l'imprécision. Le MCT permet une modélisation de la connaissance complète, notamment en modélisant le doute entre des états (l'appartenance à une ou plusieurs classes dans notre cas). De plus, cette modélisation intègre intrinsèquement une place aux états "non prévus", ce qui permet de formaliser l'identification nouveaux contenus visuels pour définir par exemple de nouvelles classes.

Dans la suite de ce chapitre, nous allons d'abord présenter des descripteurs visuels qui ont été mis en œuvre pour établir des rapports de proximité entre des images. Puis, nous discuterons du lien que nous pouvons faire entre ces proximités et l'établissement de classes d'images cohérentes pour un utilisateur. Les principaux outils du Modèle des Croyances Transférable sont ensuite présentés. La méthode est alors détaillée en trois grandes étapes, correspondant à 3 échelles d'analyse de la connaissance.

1. La première échelle d'analyse "locale" concerne l'interprétation des proximités calculées entre les descripteurs de 2 images en des fonctions de croyance. Les croyances sont ensuite fusionnées pour combiner des témoignages de plusieurs images membres d'une même classe vis-à-vis d'une image non étiquetée. A la fin de cette étape, un ensemble de fonctions de croyances décrit l'appartenance ou non de cette image à la classe.
2. La seconde échelle "multi-classe" aborde la fusion des fonctions de croyance relatives à l'appartenance ou non d'une image pour un ensemble de plusieurs classes.
3. La dernière échelle "multi-descripteur" aborde la fusion des fonctions de croyance relatives à l'appartenance ou non d'une image pour un ensemble de plusieurs classes selon plusieurs types de descripteurs visuels, en adoptant ainsi un schéma de fusion tardive des informations apportées par les différents descripteurs.

2. Extraction d'information, description de contenu visuel

Les collections d'images que nous souhaitons traiter peuvent être très diverses (reportages photographiques, vidéos, photographies personnelles,...). Il est donc difficile de décrire des descripteurs visuels sur un seul type de collection d'images, et nous utiliserons donc des descripteurs visuels standards. Nous sommes conscients qu'ils pourront être remplacés ou complétés plus tard par des descripteurs plus informatifs. Nous avons privilégié dans ce chapitre le travail sur la formalisation, dans le cadre du Modèle des Croyances Transférables, d'une méthode de modélisation de la connaissance en présence de sources d'informations imparfaites. En effet, un descripteur visuel résume souvent l'information selon une seule modalité (couleurs, textures, formes, mouvements pour les vidéos,...) et ne rapporte qu'une partie de l'information. L'utilisation du MCT est particulièrement adaptée à la modélisation de la connaissance et cohérente pour gérer les imperfections des informations apportées par les descripteurs visuels.

2.1. Descripteur global de couleurs

Une description classique et globale d'une image consiste à recenser la fréquence d'apparition des couleurs des pixels. Un histogramme résume ainsi la distribution des couleurs et possède les propriétés intéressantes d'invariance aux rotations et translations. La qualité de description d'un histogramme dépend alors du choix d'un espace de représentation des couleurs et de la quantification retenue.

Il existe diverses espaces de représentation de la couleur : Rouge-Vert-Bleu ($\{R,V,B\}$ ou $\{R,G,B\}$),

Teinte-Saturation-Valeur ($\{T,S,V\}$ ou $\{H,S,V\}$), (voir figure 3.6)... L'espace $\{R,V,B\}$ peut être directement utilisé malgré des défauts bien identifiés comme l'impossibilité de reproduire certaines couleurs en synthèse additive ou son incompatibilité avec le système visuel humain quand à la perception d'écarts de couleurs. En effet, le système visuel humain ne discrimine pas de la même manière toutes les couleurs, car la sensibilité aux écarts de couleurs n'est pas reliée de manière linéaire au comportement des photorécepteurs de l'œil.

Pour faire face à cette non-uniformité, les espaces de couleur $\{L,a,b\}$ et $\{L,u,v\}$ ont été proposés et normalisés par la Commission Internationale de l'Éclairage (CIE) en 1976 [dl86]. Ces systèmes visent à uniformiser la perception des différences de couleurs. Les relations non-linéaires entre les composantes $\{L,a,b\}$ d'une part et $\{L,u,v\}$ d'autre part, ont pour but d'imiter la réponse logarithmique de l'œil. Dans le cas $\{L,a,b\}$, la composante "L" est la "clarté", qui va de 0 (noir) à 100 (blanc), la composante "a" représente la gamme de l'axe rouge-vert, et la composante "b" celle de l'axe jaune-bleu. L'espace $\{L,u,v\}$ est assez proche de l'espace $\{L,a,b\}$ en utilisant d'autres relations non-linéaires entre les composantes u et v . En particulier, pour comparer deux couleurs quelconques (L_1, u_1, v_1) et (L_2, u_2, v_2) dans l'espace $\{L, u, v\}$, la CIE recommande l'utilisation de la distance euclidienne :

$$\Delta E = \sqrt{(L_1 - L_2)^2 + (u_1 - u_2)^2 + (v_1 - v_2)^2} \quad (3.1)$$

où ΔE modélise l'écart de couleurs perçu par l'œil humain.

Le choix de l'espace pour représenter l'information de couleur d'une image est souvent arbitraire. L'espace $\{H,S,V\}$ facile à se représenter intuitivement, est souvent utilisé et une étude dans [LL03] montre que cet espace serait le plus performant pour une tâche de recherche d'images par le contenu.

De plus, le choix de l'espace de couleurs peut dépendre fortement de la métrique de comparaison des histogrammes. Nous avons testé, dans le dernier chapitre sur les expérimentations, différentes métriques classiques [TFMB04] (distance de Manhattan, distance euclidienne, test du χ^2 et la distance de Bhattacharyya) pour comparer des histogrammes couleurs dans les espaces ($\{L,u,v\}$, $\{L,a,b\}$, $\{R,G,B\}$ et $\{H,S,V\}$). Nous avons constaté que l'espace de couleurs permettant les meilleures performances de classification d'images n'était pas le même d'une collection d'images à l'autre, selon la métrique et les paramètres internes de l'algorithme de classification.

2.2. Descripteur global des orientations

La description seule de la couleur n'apporte pas d'information spatiale. Pour pallier ce manque, il est courant de compléter l'analyse visuelle avec des descripteurs sur le contenu structurel sous l'angle des textures, des orientations ou des formes. Ces types de descripteurs peuvent présenter les propriétés intéressantes d'invariance aux conditions de prise de vue telles que les changements d'illumination, de dynamique, et de dérive colorimétrique (voir la figure 3.7 pour une illustration).

Il existe de nombreuses méthodes pour décrire les contenus structurels des images comme l'analyse

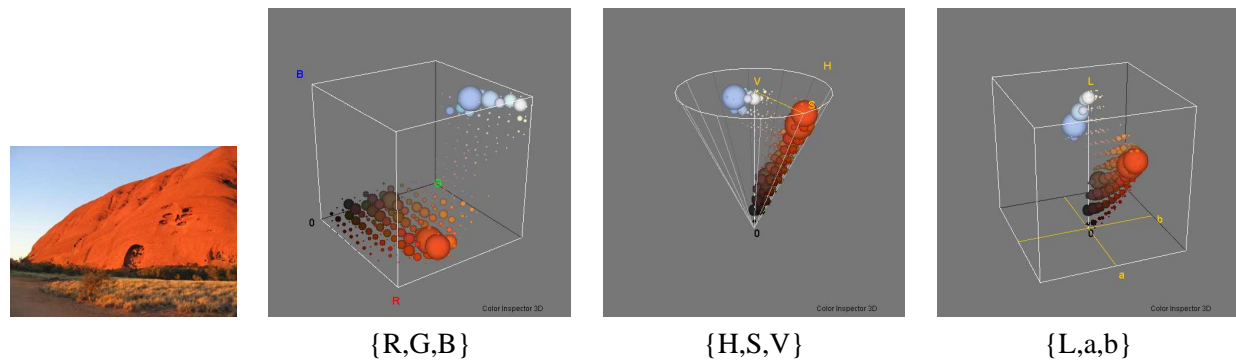


FIGURE 3.6 : Exemples de représentations d'histogrammes de 256 classes couleurs dans les espaces $\{R,G,B\}$, $\{H,S,V\}$, $\{L,a,b\}$. L'espace couleur est découpé en 256 "cubes" de tailles identiques et chaque "bulle" représente alors une classe de couleurs dont la taille est proportionnelle au nombre de pixels de cette couleur. Nous pouvons remarquer que l'histogramme dans l'espace $\{R,G,B\}$ a tendance à détailler en beaucoup de classes les couleurs "foncées", là où le système visuel humain a du mal à distinguer les couleurs entre elles.

statistique de matrices de co-occurrence des niveaux de gris [Har79] ou les méthodes spectrales basées sur des bancs de filtres de Gabor [Lee96]. Par la nature très locale des descriptions de textures, les applications sont souvent très ciblées comme la détection et le suivi d'objets ou de visages [VJ01], [BR05].

Les descripteurs de forme cherchent à décrire de manière précise les contours des objets dans une image (angles, courbures, jonction entre contours). Mais ce type de description est plutôt réservé aux représentations 2D (caractères d'écriture, logos, schémas...) [ZL01], [IV96]. Dans le cas des images naturelles, la description des formes est plutôt adaptée aux photographies contenant un seul objet mis en évidence, sur un fond uni comme le propose [YHB08] dans l'étude de feuilles de différentes espèces de plantes.

L'étude des textures est plus locale, et l'analyse des formes est plutôt apte à décrire des objets isolés. Or, les collections d'images sur lesquelles nous désirons travailler ne représentent pas spécifiquement des objets bien distincts, et nous voulons pouvoir les comparer dans leur globalité. Nous avons retenu dans un premier temps comme descripteur l'histogramme classique des orientations basé sur les gradients des contours [JV96], un descripteur standard permettant d'avoir la distribution globale des orientations des pixels en fonction de leur amplitude. L'histogramme polaire est réglé de manière à considérer 8 orientations entre 0 et 180 degrés et 8 amplitudes. La figure 3.8 donne quelques illustrations représentant ces descripteurs sur des photographies d'une même collection.

3. Manipulation de descriptions visuelles pour le classement d'images

Le rôle du module décrit dans ce chapitre est de modéliser un état des connaissances quand à l'appartenance potentielle des images à une ou plusieurs classes à partir d'informations visuelles données par des



FIGURE 3.7 : Exemple de dérive couleurs au sein d'une même collection d'images : la même scène est photographiée par plusieurs modèles d'appareils photographiques numériques différents. Les conditions d'illumination liées à l'angle de vue, ainsi que la qualité et les réglages internes des appareils produisent une collection d'images avec des écarts de couleurs parfois très importants. Il peut être alors intéressant de s'appuyer sur des descriptions sur le contenu structurelle des images pour pouvoir les comparer entre elles par exemple avec des descriptions de textures ou des orientations.

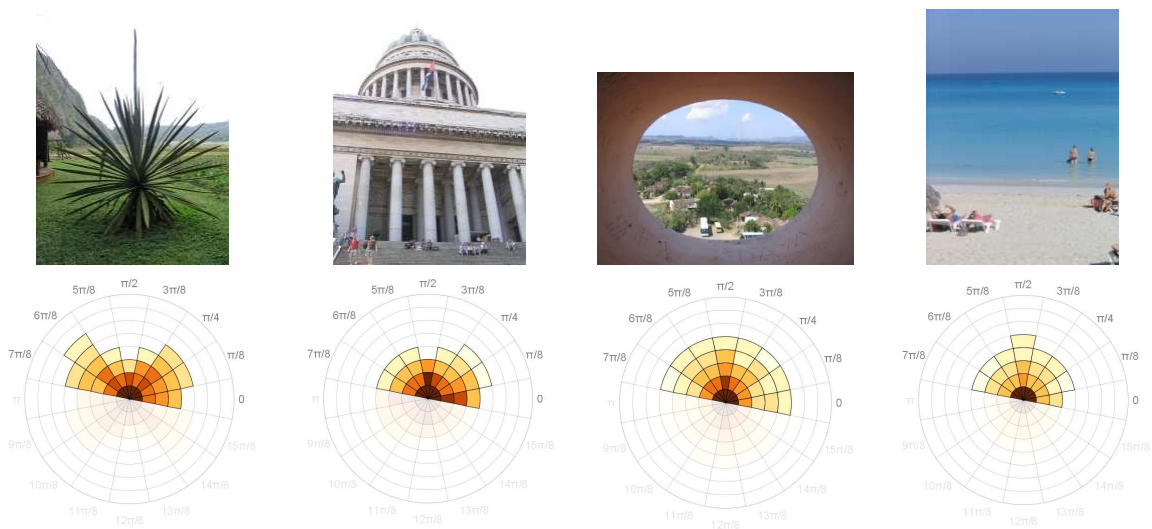


FIGURE 3.8 : Exemples de représentations d'histogrammes polaires classiques des orientations des gradients. Les histogrammes polaires permettent de recenser la distribution des gradients des pixels selon 8 orientations et 8 amplitudes différentes. L'espace est ainsi découpé en 64 "arcs pleins" dont l'intensité de la couleur est proportionnelle au nombre de pixels représentant le même type d'orientation. Par exemple, la première image est marquée par des orientations intenses sur les directions "diagonales". La seconde image est plutôt marquée par de nombreuses petites orientations verticales et horizontales. La troisième image possède la particularité de recouvrir toutes les orientations dans toutes les directions, ce qui est cohérent avec la scène photographiée. Enfin la dernière image possède une majorité d'orientations verticales.

vecteurs descripteurs des couleurs et des orientations. L'utilisation d'une distance entre ces descriptions visuelles permet d'établir un premier rapport de proximité entre les images. Cependant, ces rapprochements risquent d'être insuffisants pour identifier des groupes d'images cohérents selon l'interprétation d'un utilisateur.

Par exemple, la figure 3.9 représente un cas idéal où une distance employée sur des descripteurs visuels permet d'isoler facilement un groupe d'images similaires. Dans cet exemple, 3 ensembles de 50 images unicouleurs sont générés artificiellement selon des lois normales à 3 dimensions dans l'espace couleur $\{R,G,B\}$ avec des matrices de variance-covariance identiques et avec pour centres 3 couleurs

différentes "violet", "vert" et "cyan", de telle manière que les 3 groupes soient facilement séparables. Un histogramme des distances de 50 intervalles se focalisant sur le groupe des images "violette" permet d'observer les distances intra-classes entre toutes les images violettes, et inter-classes entre les images violettes et les celles des 2 autres groupes. Dans ce cas idéal, si un utilisateur conçoit comme un groupe d'images de couleurs similaires comme étant les membres d'une même classe, la sémantique exprimée est en adéquation avec la distribution des distances, et un simple seuillage permet d'isoler la classe "violet".

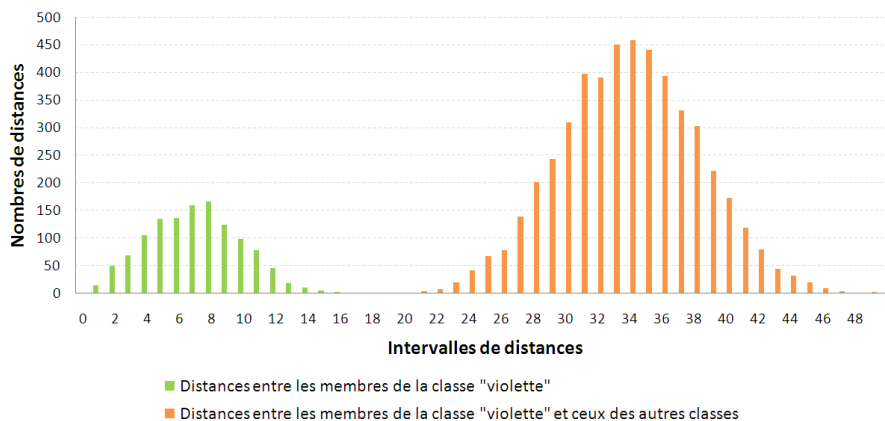


FIGURE 3.9 : Distribution des distances au sein d'une collection contenant 3 classes d'images uni-couleur de 50 images chacune générées artificiellement par des lois normales tri-dimensionnelles. L'histogramme représenté ici se focalise sur une seule classe "violette" et l'on peut observer les distances "intra-classes" entre les membres de cette classe, et les distances "inter-classes" entre les membres de la classe et ceux des autres classes. Dans ce cas idéal, un simple seuillage sur les distances permettrait d'isoler facilement la classe des images "violette".

La figure 3.10 illustre un cas plus réaliste sur une collection d'images naturelles. Un utilisateur a identifié un groupe de photographies de montagnes, parmi d'autres représentant des fleurs, des bus, des animaux, ... L'histogramme représenté s'appuie sur le calcul de distances euclidiennes L2 sur les descripteurs couleurs basés sur l'espace couleur $\{L,a,b\}$. L'histogramme représente ici la distribution des distances intra-classe, entre les membres de la classe "montagnes" et des distances inter-classes entre les membres de la classe "montagnes" et les autres images. Dans cet exemple, il est difficile d'isoler les membres de la classe "montagnes" des autres images si l'on s'appuie directement sur les distances. Le témoignage apporté par le descripteur couleur ne satisfait pas entièrement l'hypothèse que deux images proches appartiennent à une même classe.

A l'extrême, on peut avoir des situations où un utilisateur a regroupé au sein d'une même classe des images très éloignées dans l'espace des descripteurs. Utiliser directement ces distances pour faire de la classification automatique d'images risque de provoquer de nombreux étiquetages incorrects. Il est nécessaire de prendre en compte les imprécisions et imperfections liées à l'utilisation des distances entre descripteurs et tenter de quantifier la confiance sur les témoignages apportés par ces distances.

Tout système traitant des données réelles est confronté à des mesures présentant une part d'incertitude et d'imprécision. De nombreux formalismes tentent de gérer ces deux notions dans leur processus de modé-

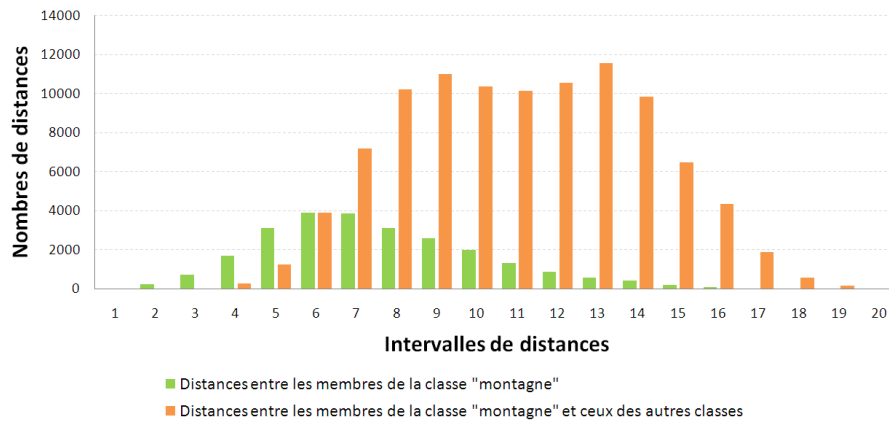


FIGURE 3.10 : Distribution des distances entre les descripteurs couleurs basées sur l'espace couleur $\{L,a,b\}$ intra et inter-classe pour la classe "montagnes" d'une collection d'images Corel. Dans cet exemple plus réaliste, il est difficile d'isoler les membres de la classe "montagnes" des autres images si l'on s'appuie directement sur les distances.

lisation de la connaissance. L'approche la plus courante est incontestablement la théorie des probabilités. Dans ce cadre, on modélise la notion de doute à travers l'équiprobabilité entre différents états considérés. Or, cet *a priori* d'équiprobabilité est rarement vérifié : ce n'est pas parce qu'il manque de l'information sur certains états, que tous les états peuvent être vérifiés à la même fréquence. Il est possible de mieux représenter le doute comme par exemple avec la théorie des possibilités [DP88], en distinguant incertitude et imprécision.

Nous avons choisi d'utiliser la théorie des fonctions de croyance, plus particulièrement le Modèle des Croyances Transférables [Sme94]. Cette approche se distingue des probabilités par une modélisation de la non connaissance, au lieu de se focaliser sur des *a priori* qui peuvent se révéler erronés. Cette théorie relativement récente offre des outils puissants de gestion de l'incertitude, de l'imprécision et du conflit sur les données.

Le but de la prochaine étape est alors de modéliser la connaissance apportée par des distances entre descripteurs visuels en termes de fonctions de croyance. Mais avant de discuter de cette étape de modélisation, les principes fondamentaux du Modèle des Croyances Transférables et ainsi que ses principaux outils doivent être présentés.

4. Modèle des Croyances Transférables

Le Modèle des Croyances Transférables (MCT ou en anglais TBM pour *Transferable Belief Model*) est un cadre formel développé par Smets [Sme94] au début des années 90 pour représenter et combiner la connaissance fournie par différentes sources d'information.

Tout capteur mesurant des grandeurs physiques (lumière, sons, pression. . .) ou numériques (descripteurs visuels, détecteurs d'objets dans les images. . .), garde généralement une part d'imprécision et d'incertitude liées à différents facteurs (environnement, limitations, défaut de conception. . .).

Le MCT est une solution parmi d'autres, comme la théorie des probabilités ou la théorie des possibilités, proposant des outils performants pour la gestion de l'imprécision et de l'incertitude afin d'améliorer une éventuelle prise de décision.

Une particularité propre au MCT vis-à-vis des autres théories, est la formalisation explicite du doute, ce qui peut s'avérer utile pour représenter une hésitation entre plusieurs classes par exemple dans un problème de reconnaissance de formes.

Une seconde particularité du MCT est la capacité à évaluer la discordance entre plusieurs sources d'information lors de leur combinaison. Une mesure de conflit peut souligner alors un manque de cohérence du modèle de fusion. Elle peut être également considérée comme une véritable source d'information par exemple pour détecter des changements d'états dans une vidéo [RRP07].

Enfin, une dernière différence notable par rapport aux autres théories est le mécanisme de raisonnement en deux niveaux :

1. le niveau *credal*, venant du latin *credo* "je crois", où la connaissance est formalisée par des fonctions de croyance, et révisée par des règles de combinaison,
2. le niveau *pignistique* du latin *pignus* voulant dire "pari", dans lesquels les fonctions de croyance sont transformées en fonctions de probabilités pour effectuer une prise de décision.

L'avantage du MCT est de disposer d'outils dédiés à la modélisation de la connaissance au niveau *credal*, alors que, dans la théorie des probabilités la connaissance est "introduite" directement au niveau même de la décision, notamment par l'utilisation d'*a priori*.

Dans la suite du chapitre, nous nous focalisons sur le niveau *credal* et utilisons une partie des outils disponibles du MCT pour modéliser l'état de connaissances sur les images quand à l'appartenance à des classes. Le niveau *pignistique* est abordé dans les chapitres suivants. Par ailleurs, le lecteur intéressé pourra se référer au chapitre 2 dans [Ram07] pour une présentation complète de l'ensemble des outils offerts par le MCT.

4.1. Représentation de la connaissance

Au niveau *credal*, la première étape de la modélisation de la connaissance dans le cadre du MCT consiste à identifier tous les états possibles, ou *hypothèses*, afin de définir un *cadre de discernement* Ω :

$$\Omega = \{H_1, H_2, \dots, H_K\} \quad (3.2)$$

où Ω est un ensemble fini représentant tous les états possibles identifiés H_i (avec $i \in \{1, 2, \dots, K\}$), ou *hypothèses*, que peut prendre un système. Les hypothèses sont considérées comme étant exclusives et

le cadre de discernement Ω est sensé être exhaustif (même si par la suite nous verrons qu'il est possible d'identifier des hypothèses non prévues dans le cadre de discernement).

Une fonction de masse est la représentation la plus communément utilisée dans le cadre du MCT pour exprimer une fonction de croyance. Elle est définie par :

$$m^\Omega : 2^\Omega \rightarrow [0, 1] \quad (3.3)$$

$$B \mapsto m^\Omega(B) \quad (3.4)$$

où 2^Ω est l'espace puissance rassemblant tous les sous-ensembles possibles formés des hypothèses et unions d'hypothèses de Ω . Chaque sous-ensemble $B \subseteq \Omega$, est appelé généralement *proposition*.

Une masse $m^\Omega(B)$ représente ainsi le degré de croyance attribué à la proposition B qui, compte tenu de la connaissance à un moment donné, n'a pas pu être affectée à un sous-ensemble plus spécifique.

L'ensemble des masses sur toutes les propositions de 2^Ω constitue une distribution de masses m^Ω , appelée également BBA (pour *Basic Belief Assignment*) et doit vérifier :

$$\sum_{B \subseteq \Omega} m^\Omega(B) = 1 \quad (3.5)$$

Une distribution de masses peut éventuellement contenir de nombreuses propositions ayant une masse nulle. Les *éléments focaux* sont *a contrario* les propositions possédant de la masse et pointant la présence d'information.

Cette modélisation se distingue fortement de l'approche classique des probabilités. Prenons l'exemple classique de la recherche d'un coupable d'un meurtre cité dans l'article fondateur de Smets [Sme94]. Des enquêteurs disposent de trois suspects H_1 , H_2 et H_3 . Si nous raisonnons dans les MCT nous définissons alors un cadre de discernement $\Omega = \{H_1, H_2, H_3\}$ représentant trois hypothèses envisageables : "le suspect H_n est le coupable".

En considérant qu'il n'y a pas de complicité entre les suspects, les hypothèses sont alors supposées exclusives, ce qui signifie que deux hypothèses ne peuvent être vraies simultanément.

Si à un instant donné, les enquêteurs ne disposent d'aucun témoignage, il n'y a aucune connaissance. Dans la théorie des probabilités, cette situation se traduit par un cas d'équiprobabilité où les probabilités sont affectées uniquement sur des hypothèses singleton H_1 , H_2 et H_3 :

$$p(H_1) = p(H_2) = p(H_3) = \frac{1}{3}$$

Dans cette même situation d'incertitude maximale, le MCT permet une représentation plus fine de la

connaissance avec une distribution de masses définie sur $2^3 = 8$ propositions :

$$\begin{aligned}
 m^\Omega(\emptyset) &= 0 \\
 m^\Omega(\{H_1\}) &= m^\Omega(\{H_2\}) = m^\Omega(\{H_3\}) = 0 \\
 m^\Omega(\{H_1, H_2\}) &= m^\Omega(\{H_2, H_3\}) = m^\Omega(\{H_1, H_3\}) = 0 \\
 m^\Omega(\{H_1, H_2, H_3\}) &= 1
 \end{aligned}$$

Dans le cadre du MCT, la totalité de la masse de croyance, est affectée à la proposition $\{H_1, H_2, H_3\}$ représentant le *doute* formé de l'union de toutes les hypothèses (cette proposition de doute peut être également notée par Ω puisque le cadre de discernement décrit l'ensemble des hypothèses disponibles). La distribution de masses comporte alors un seul élément focal pointant sur une ignorance totale.

Imaginons maintenant qu'un premier témoignage rapporte avec certitude que le coupable est chauve, ce qui est le cas des deux premiers suspects H_1 et H_2 . La distribution de masse peut être alors affinée en transférant la masse sur la proposition plus spécifique $\{H_1, H_2\}$. La nouvelle distribution de masses comporte alors un seul élément focal réduisant l'ignorance aux deux hypothèses H_1 et H_2 , au lieu des trois disponibles :

$$\begin{aligned}
 m^\Omega(\emptyset) &= 0 \\
 m^\Omega(\{H_1\}) &= m^\Omega(\{H_2\}) = m^\Omega(\{H_3\}) = 0 \\
 m^\Omega(\{H_1, H_2\}) &= 1 \\
 m^\Omega(\{H_2, H_3\}) &= m^\Omega(\{H_1, H_3\}) = 0 \\
 m^\Omega(\{H_1, H_2, H_3\}) &= 0
 \end{aligned}$$

Dans cet exemple, la masse sur l'ensemble vide $m^\Omega(\emptyset)$ est nulle. Il convient alors de parler de *monde fermé*, ce qui suppose que toutes les hypothèses envisageables ont été identifiées, et donc que le cadre de discernement Ω est exhaustif. Dans les MCT, il est possible de supposer une situation de *monde ouvert* exprimé par $m^\Omega(\emptyset) > 0$. Dans l'exemple précédent, cela supposerait qu'un quatrième suspect H_4 encore non identifié pourrait être le coupable.

Une des principales difficultés consiste à concevoir un cadre de discernement et une distribution de masses de la manière la plus fidèle possible à la connaissance disponible. Par contre, l'avantage est de pouvoir s'abstenir d'informations *a priori*.

Remarque : pour alléger les notations, nous supprimerons les accolades des propositions, tout en sachant qu'il s'agit bien d'ensembles.

4.2. Combinaison

Considérons deux sources d'information pour lesquelles ont été définies deux distributions de masses m_1^Ω et m_2^Ω sur un même cadre de discernement Ω . Il s'agit maintenant de les fusionner pour synthétiser les informations de ces 2 sources en définissant une nouvelle distribution de masses. De nombreux opérateurs de fusion de distribution de masses ont été développés dans le cadre du MCT. La *règle de combinaison conjonctive* est sans aucun doute la plus emblématique de la théorie des fonctions de croyance, notée :

$$m_1^\Omega \otimes_2(D) = (m_1^\Omega \otimes m_2^\Omega)(D) = \sum_{B \cap C = D} m_1^\Omega(B) \cdot m_2^\Omega(C) \quad (3.6)$$

avec B, C, D des propositions de 2^Ω . Cette règle de fusion possède la propriété intéressante de *spécialisation* en transférant la masse sur des sous-ensembles de cardinalité plus faible.

Exemple : combinaison de 2 distributions de masses dans un cadre de discernement contenant 2 hypothèses

Considérons deux distributions de masses m_1^Ω et m_2^Ω définies sur un même cadre de discernement $\Omega = \{H_1, H_2\}$. Les distributions de masses possèdent des masses pointant sur les singletons H_1, H_2 et le doute entre les deux hypothèses $\{H_1, H_2\} = \Omega$. La combinaison des différentes propositions avec la règle conjonctive est représentée par le tableau 3.6. La distribution de masses résultante $m_{1,2}^\Omega$ est alors :

$$\begin{aligned} m_{1,2}^\Omega(H_1) &= m_1^\Omega(H_1) \cdot m_2^\Omega(H_1) + m_1^\Omega(H_1) \cdot m_2^\Omega(\Omega) + m_1^\Omega(\Omega) \cdot m_2^\Omega(H_1) \\ m_{1,2}^\Omega(H_2) &= m_1^\Omega(H_2) \cdot m_2^\Omega(H_2) + m_1^\Omega(H_2) \cdot m_2^\Omega(\Omega) + m_1^\Omega(\Omega) \cdot m_2^\Omega(H_2) \\ m_{1,2}^\Omega(\Omega) &= m_1^\Omega(\Omega) \cdot m_2^\Omega(\Omega) \\ m_{1,2}^\Omega(\emptyset) &= m_1^\Omega(H_1) \cdot m_2^\Omega(H_2) + m_1^\Omega(H_2) \cdot m_2^\Omega(H_1) \end{aligned}$$

Après combinaison, les masses étant compris entre 0 et 1, les propositions de cardinalité plus faible, les singletons H_1, H_2 en l'occurrence, ont une masse qui augmente. La masse sur le doute Ω quant à elle diminue, illustrant ainsi l'effet de spécialisation de la règle de combinaison conjonctive.

		m_1^Ω		
		H_1	H_2	Ω
m_2^Ω	H_1	H_1	\emptyset	H_1
	H_2	\emptyset	H_2	H_2
	Ω	H_1	H_2	Ω

TABLE 3.6 : Intersections des propositions de deux distributions de masses pour la règle de combinaison conjonctive.

Cette règle de combinaison conjonctive peut faire apparaître de la masse sur l'ensemble vide représentant explicitement le *conflict* entre les deux sources. Le conflit peut être interprété de plusieurs manières. Il peut souligner une mauvaise modélisation de la connaissance, par exemple si le cadre de discernement

n'est pas exhaustif (hypothèse de monde ouvert), ou bien si les deux hypothèses ne sont pas réellement exclusives. Un deuxième cas d'apparition du conflit peut venir du fait que les deux sources d'information ne concernent pas la même grandeur (par exemple un capteur mesure la pression et une autre la température) et n'apportent donc pas les mêmes types de témoignages.

4.3. *Prise de décision*

Le niveau *pignistique* a pour rôle de déterminer l'hypothèse la plus probable à partir de la connaissance modélisée au niveau *credal*. Cette phase de décision peut s'appuyer sur la *distribution de probabilités pignistiques* [Sme05] notée $BetP\{m^\Omega\}$ obtenue à partir d'une distribution de masses m^Ω . La transformée pignistique consiste à répartir de manière équiprobable la masse d'une proposition B sur les hypothèses H_i contenues dans B . La probabilité pignistique d'une hypothèse H_i du cadre de discernement Ω est définie par :

$$BetP\{m^\Omega\}(H_i) = \frac{1}{1 - m^\Omega(\emptyset)} \sum_{B \subseteq \Omega, H_i \in B} \frac{m^\Omega(B)}{|B|} \quad (3.7)$$

avec $|B|$ le cardinal de la proposition B . La décision est généralement prise en retenant l'hypothèse H_i possédant la plus grande probabilité pignistique $BetP\{m^\Omega\}(H_i)$:

$$\omega_0 = \operatorname{argmax}_{H_i \in \Omega} BetP\{m^\Omega\}(H_i) \quad (3.8)$$

5. Modélisation de la connaissance

Dans le cadre du MCT, le niveau *credal* consiste à lier un ou plusieurs paramètres numériques extraits d'un phénomène observé aux hypothèses d'un cadre de discernement. Cette conversion numérique - symbolique est cruciale et elle peut poser des problèmes de performances si le phénomène étudié est mal modélisé. Cependant, cette difficulté est également valable pour les autres théories (probabilités, théorie des possibilités. . .). Dans la méthode proposée, le système dispose de vecteurs descripteurs décrivant le contenu visuel d'images. Le but est alors de calculer des distributions de masses à partir de distances entre ces descripteurs pour décrire des croyances sur les hypothèses d'appartenance ou non d'une image dans une classe.

5.1. *Règles de représentation de la connaissance à partir d'une distance*

La première étape d'extraction d'information fournit un ensemble de descripteurs visuels pour chaque image sous une forme vectorielle. En calculant une distance entre les descripteurs de deux images, un

premier rapport de proximité quantifie la dissimilarité des contenus visuels. La distance entre les descripteurs de 2 images donne une information sur le fait que ces 2 images appartiennent à la même classe C_q ou non. Nous pouvons définir un cadre de discernement Ω_q décrivant ces états :

$$\Omega_q = \{H_q, \overline{H}_q\} \quad (3.9)$$

avec H_q l'hypothèse affirmant que les 2 images appartiennent à une même classe C_q (c'est-à-dire qu'elles ont la même étiquette q), et \overline{H}_q l'hypothèse complémentaire modélisant le cas où les 2 images n'appartiennent pas à la même classe C_q .

L'objectif est maintenant d'interpréter les relations de proximité entre 2 images par des règles expertes. On propose 4 approches représentées par les règles ci-dessous. Un premier raisonnement correspondant à la règle 1 pouvant être qualifié de "Bayésien" typique des probabilités exprime deux *a priori* :

Règle 1 :

1. *Si deux images sont proches alors elles appartiennent à la même classe.*
2. *Si deux images sont éloignées, elles n'appartiennent pas à la même classe.*

Cette règle est efficace dans le cas où les classes d'images sont facilement séparables, c'est-à-dire si les images au sein d'une même classe sont très proches, et les images appartenant à deux classes distinctes sont très éloignées.

Il peut y avoir des cas où les classes sont difficilement séparables, si par exemple des images au sein d'une même classe sont très éloignées. Un second avis expert, exprimé par la règle 2, peut alors s'appuyer sur un seul *a priori* considérant que deux images proches appartiennent à une même classe. Par contre, aucun *a priori* n'est exprimé lorsque deux images sont éloignées, ce qui se traduit par un doute élevé.

Règle 2 :

1. *Si deux images sont proches alors elles appartiennent à la même classe.*
2. *Si deux images sont éloignées, alors il n'y a pas d'information apportée par la distance : le doute est élevé.*

En effet, le fossé sémantique exprime qu'il est possible que deux images éloignées dans l'espace des descripteurs peuvent être proches pour un utilisateur. En plaçant un doute maximal sur le cas où les images sont très éloignées, on ne peut affirmer avec certitude que les images n'appartiennent pas à une même classe. C'est donc une règle optimiste pour les images proches avec le risque de les classer dans une même classe, et prudente pour les images éloignées.

Cependant, il est également possible que des classes soient très homogènes, avec des distances intra-classe faibles, mais se chevauchant entre elles. Dans ce cas, un *a priori* peut considérer avec certitude que deux images éloignées n'appartiennent pas à la même classes, alors que si elles sont proches, elle peuvent aussi bien appartenir à deux classes différentes qu'être dans la même classe. Cette interprétation est décrite par la règle 3 :

Règle 3 :

1. *Si deux images sont proches, alors il n'y a pas d'information portée par la distance : le doute est élevé.*
2. *Si deux images sont éloignées, les images n'appartiennent pas à la même classe.*

Cette règle est duale de la règle 2. Il est même possible d'envisager que les classes ne sont pas exclusives : une image située entre deux classes peut être éventuellement associée aux 2 classes à la fois.

Ces trois premières règles sont toutes des interprétations expertes acceptables d'une distance. Cependant, elles ne sont peut être pas applicables dans un même contexte. La règle 1 est adaptée au cas où les classes sont homogènes et facilement séparables. Les règles 2 et 3 sont duales, car elles reviennent à transférer la totalité du doute sur l'une des deux hypothèses de base. Il est possible de trouver un compromis entre les *a priori* catégoriques de la règle Bayésienne, tout en introduisant une part de doute, ce qui est exprimé par la règle 4.

Règle 4 :

1. *Si deux images sont proches alors elles appartiennent à la même classe.*
2. *Si deux images sont éloignées, alors elles n'appartiennent pas à la même classe.*
3. *Dans tous les cas, il existe une part de doute, et si les images sont à une distance intermédiaire, le doute est maximum.*

La contrainte de fossé sémantique est ainsi respectée, car la présence de doute modélise les éventualités où les *a priori* ne sont pas valables, même si les images sont très proches ou très éloignées dans l'espace des descripteurs. Le tableau 3.7 récapitule plus loin les différentes règles qui viennent d'être décrites.

5.2. Conversion numérique - symbolique

A partir des règles qui viennent d'être définies, il s'agit de déterminer la croyance qu'une image appartient à une classe C_q , sachant que cette classe contient une ou plusieurs images exemple.

Soit une image l_q^0 étiquetée dans une classe C_q , et soit une deuxième image u sans étiquette¹. Le problème est de savoir dans quelle mesure l'image non étiquetée u peut être classée dans la classe C_q selon le témoignage apporté par l'image étiquetée l_q^0 .

L'opération de conversion numérique - symbolique (voir figure 3.11) utilise la distance $d(u, l_q^0)$ entre les descripteurs des images u et l_q^0 pour établir la croyance en l'appartenance ou non de l'image non étiquetée u à la classe C_q , et déterminer ainsi une distribution de masses $m_{l_q^0}^{\Omega_q}$ définie sur le cadre de discernement Ω_q .

La question est maintenant de savoir comment établir concrètement la conversion numérique - symbolique aboutissant à une distribution de masses $m_{l_q^0}^{\Omega_q}$. Dans notre problème de classification, les méthodes non-paramétriques semblent adaptées car les observations sont peu nombreuses, la structure des données

1. les notations l et u renvoient aux termes "labeled" et "unlabeled" en anglais pour "étiqueté" et "non étiqueté".

Règle	Éléments focaux	Modélisation
Règle 1	H_q, \overline{H}_q	De type "bayésienne", pas de doute, deux <i>a priori</i> : "Si les 2 images sont proches alors la masse sur H_q est maximale, et celle sur \overline{H}_q est minimale." "Si les 2 images sont loin alors la masse sur H_q est minimale, et celle sur \overline{H}_q est maximale."
Règle 2	H_q, Ω_q	"Si les 2 images sont proches alors la masse sur H_q est maximale, et celle sur le doute Ω_q est minimale." "Si les 2 images sont loin alors la masse sur H_q est minimale, et celle sur le doute Ω_q est maximale."
Règle 3	\overline{H}_q, Ω_q	"Si les 2 images sont proches alors la masse sur le doute Ω_q est maximale, et celle sur \overline{H}_q est minimale." "Si les 2 images sont loin alors la masse sur \overline{H}_q est maximale, et celle sur le doute Ω_q est minimale."
Règle 4	$H_q, \overline{H}_q, \Omega_q$	"Si les 2 images sont TRES proches alors la masse sur H_q est maximale, celle sur le doute Ω_q est minimale, et celle sur \overline{H}_q est nulle." "Si les 2 images sont TRES loin alors la masse sur H_q est nulle, celle sur le doute Ω_q est minimale et celle sur \overline{H}_q est maximale." "Si les 2 images sont à distance intermédiaires alors les masses sur H_q et \overline{H}_q sont soit nulles ou tendent vers zéro, et celle sur le doute Ω_q est maximale."

TABLE 3.7 : Les quatre règles expertes pour établir une distribution de masses m_q^Ω dans le cadre de discernement Ω_q à partir d'une distance d mesurée entre deux vecteurs descripteurs d'images.

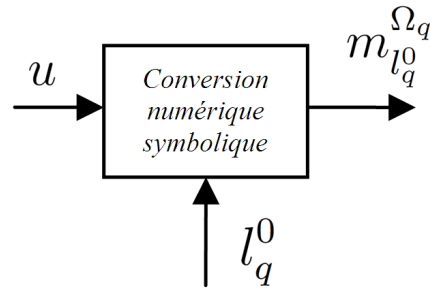


FIGURE 3.11 : Conversion numérique symbolique : une distance $d(u, l_q^0)$ entre les descripteurs d'une image non étiquetée u , et d'une image étiquetée l_q^0 appartenant à une classe C_q permet d'établir une distribution de masses m^{Ω_q} . Cette distribution décrit les croyances sur l'appartenance ou non de l'image non étiquetée u à la classe C_q .

n'est pas connue à l'avance et peut être potentiellement très complexe. En effet, il est possible qu'un utilisateur veuille regrouper un nombre limité d'images visuellement très différentes en une même classe. On ne peut pas modéliser l'information donnée par ces quelques images étiquetées avec des méthodes statistiques car elles ne sont pas suffisamment représentatives. Pour classer une image non étiquetée u , il est préférable de raisonner sur les plus proches voisins. Nous nous sommes inspiré de la méthode des k plus proches voisins évidentiels proposée par Denœux [Den08] (parfois noté dans la littérature KnnEv ou K-EV). Dans cette méthode, les k plus proches voisins d'une classe C_q permettent d'évaluer la croyance que l'image u appartient à cette classe C_q .

La première étape consiste à calculer une distribution de masses à partir d'une distance $d(u, l_q^i)$ entre un échantillon u non étiqueté et un voisin l_q^i portant l'étiquette d'une classe C_q . La modélisation de cette connaissance utilise la règle 2 identifiée dans le tableau 3.7. L'auteur considère, premièrement, que même si un échantillon l_q^i est étiqueté correctement, il n'apporte pas un témoignage certain à 100%. Autrement dit, lorsque que la distance $d(u, l_q^i)$ tend vers 0, un facteur d'affaiblissement dans l'expression de la fonction de croyance doit laisser une part de doute.

Deuxièmement, l'auteur observe qu'un échantillon étiqueté l_q^i n'apporte que de l'information sur l'appartenance à une classe C_q . En effet, comme il l'a été exprimé dans la règle 2 précédente (tableau 3.7), l'échantillon étiqueté n'apporte aucune information si la distance $d(u, l_q^i)$ est grande, notamment sur l'appartenance à d'autres classes. Par conséquent, le reste de la masse disponible est attribuée au doute total Ω_q , l'ensemble des propositions disponibles du cadre de discernement Ω_q .

Ces remarques permettent de convertir la distance $d(u, l_q^i)$ en une distribution de masses m^{Ω_q} contenant uniquement les deux éléments focaux $m^{\Omega_q}(H_q)$ et $m^{\Omega_q}(\Omega_q)$:

$$m^{\Omega_q}(H_q) = \alpha(u, l_q^i) \quad (3.10)$$

$$m^{\Omega_q}(\Omega_q) = 1 - \alpha(u, l_q^i) \quad (3.11)$$

$$m^{\Omega_q}(\overline{H_q}) = 0 \quad (3.12)$$

$$m^{\Omega_q}(\emptyset) = 0 \quad (3.13)$$

avec $\alpha(u, l_q^i)$ une fonction strictement monotone et décroissante, comprise entre 0 et 1, en fonction de la distance $d(u, l_q^i)$. La fonction proposée par Dencœux est :

$$\alpha(u, l_q^i) = \alpha_0 \cdot e^{-\left(\frac{d(u, l_q^i)}{\sigma_q}\right)^\beta} \quad (3.14)$$

avec α_0 un facteur d'affaiblissement fixé de l'ordre de 0.9, et $\beta \in \{1, 2, \dots\}$. σ_q peut être assimilé à un paramètre d'étalement de la zone de connaissance des échantillons étiquetés l_q^i d'une même classe C_q . Un paramètre σ_q est estimé pour chaque classe C_q en fonction par exemple de la distance moyenne d_q des échantillons $\{l_q^0, l_q^1, l_q^2, \dots\}$ de la classe :

$$\sigma_q = \frac{1}{d_q^\beta} \quad (3.15)$$

A eux trois, les paramètres α_0 , β et σ_q modélisent la connaissance $\alpha(u, l_q^i)$ autour de chaque échantillon étiqueté l_q^i dans l'espace de description considéré.

Cette règle est exprimée pour la seule comparaison entre une image non étiquetée u et une image étiquetée l_q^0 appartenant à une classe C_q . Or, le seul témoignage par classe peut être trop pauvre en information pour prendre une décision de classement. Si plusieurs images étiquetées sont disponibles, il peut être intéressant d'en considérer d'avantage pour modéliser la connaissance. Ces distances entre l'image non étiquetée u et les images étiquetées $\{l_q^0, l_q^1, l_q^2, \dots\}$ peuvent être ainsi converties en distributions de masses, puis fusionnées pour affiner la connaissance.

6. Fusion de témoignages au sein d'une classe

La fusion de témoignages au sein d'une même classe C_q consiste à fusionner les témoignages apportés par les k plus proches images étiquetées $\{l_q^0, l_q^1, \dots, l_q^k\}$ de la classe C_q concernant une image non étiquetée u . Comme le décrit la figure 3.12, la distribution de masses résultante m^{Ω_q} subit ensuite une opération de transfert de masses détaillée plus loin.

6.1. Combinaison de voisins

La combinaison de voisins fusionne les témoignages apportés par les k plus proches échantillons $\{l_q^0, l_q^1, l_q^2, \dots\}$ d'une classe C_q vis-à-vis de l'échantillon non étiqueté u en utilisant la règle de combinaison conjonctive. En considérant uniquement les deux plus proches voisins l_q^0 et l_q^1 , la combinaison donne une nouvelle distribution de masses $m_{0,1}^{\Omega_q}$. La masse sur le doute Ω_q est donnée par :

$$m_{0,1}^{\Omega_q}(\Omega_q) = (1 - \alpha(u, l_q^0))(1 - \alpha(u, l_q^1)) \quad (3.16)$$

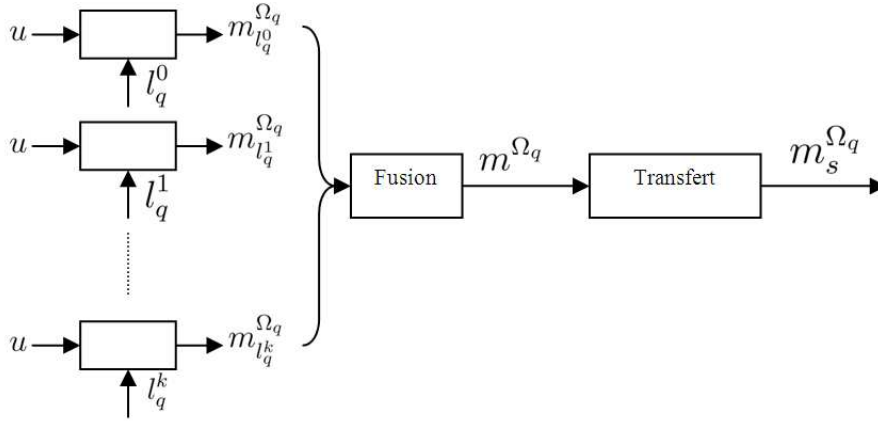


FIGURE 3.12 : Fusion des distributions de masses apportées par les différents témoignages des membres d'une même classe pour une seule image non étiquetée u .

La somme des masses d'une distribution étant égale à 1, la masse $m_{0,1}^{\Omega_q}(H_q)$ sur la proposition H_q est alors :

$$m_{0,1}^{\Omega_q}(H_q) = 1 - (1 - \alpha(u, l_q^0))(1 - \alpha(u, l_q^1)) \quad (3.17)$$

En généralisant les expressions à une combinaison des k plus proches voisins d'une même classe C_q , la distribution de masses résultant m^{Ω_q} est :

$$m^{\Omega_q}(\Omega_q) = \prod_{i=0}^{k-1} (1 - \alpha(u, l_q^i)) \quad (3.18)$$

$$m^{\Omega_q}(H_q) = 1 - \prod_{i=0}^{k-1} (1 - \alpha(u, l_q^i)) \quad (3.19)$$

Il est intéressant de noter comment s'interprète la nouvelle distribution de masse m^{Ω_q} après combinaison avec la règle conjonctive, notamment sur le doute Ω_q :

- Si les k plus proches voisins $\{l_q^0, l_q^1, l_q^2, \dots\}$ fournissent tous des distributions de masses $m_i^{\Omega_q}$ avec une masse $m_i^{\Omega_q}(\Omega)$ sur le doute élevée, alors la distribution de masses résultant m^{Ω_q} possède également une masse $m^{\Omega_q}(\Omega_q)$ élevée sur le doute Ω_q .
- Si les k plus proches voisins fournissent tous des distributions de masses avec un doute faible, alors la distribution de masses résultante possède également un doute faible.
- Si un seul des voisins fournit une distribution de masses avec un doute faible, alors que tous les autres voisins fournissent des distributions de masses avec une masse sur le doute élevée, alors la distribution de masses résultante possède une masse sur le doute faible.

Autrement dit, un seul voisin l_q^i proche de l'échantillon non étiqueté u permet de diminuer fortement le doute et d'obtenir une croyance élevée sur la proposition H_q , même si tous les autres voisins sont loin de u , comme l'illustre la figure figure 3.13.

Remarque : cette règle évite l'apparition de conflit, ce qui est préférable car il est important de considérer une hypothèse de *monde fermé* ($m^{\Omega_q}(\emptyset) = 0$). En effet, la distribution de masses m^{Ω_q} est spécialisée pour accepter ou douter sur le fait qu'une image non étiquetée u soit dans la classe C_q . Il n'y a point d'autre situation envisageable.

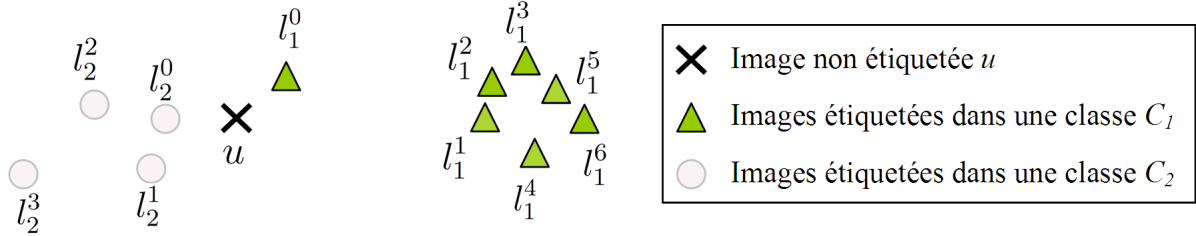


FIGURE 3.13 : Situation où un seul voisin est réellement proche de l'échantillon u à classer vis-à-vis des autres voisins d'une classe C_1 . La modélisation de la connaissance avec la règle conjonctive permet de maintenir une croyance élevée sur l'appartenance à la classe dans la distribution de masses m^{Ω_1} . En filigrane sont représentés des images étiquetées dans une autre classe C_2 , rappelant que même s'ils sont proches de l'image non étiquetée u , ils ne sont pas pris en compte pour la distribution de masse m^{Ω_1} relative à la classe C_1

6.2. Réévaluation de distributions de masses

D'après les équations 3.19 la distribution de masses m^{Ω_q} combinant les témoignages des k plus proches voisins de l'image non étiquetée u met de la masse uniquement sur la proposition H_q et le doute Ω_q . Or, cette distribution de masses m^{Ω_q} est alors incomplète car il est impossible d'avoir de la croyance sur la non appartenance d'une image dans une classe C_q .

En particulier, dans le cas où tous les voisins sont éloignés de l'image non étiquetée u , la masse résultante sur le doute $m^{\Omega_q}(\Omega_q)$ est élevée. Or il serait souhaitable que dans ce cas de figure, si tous les voisins sont éloignés, que l'on puisse affirmer avec certitude que la masse sur la proposition \overline{H}_q soit élevée.

D'après le tableau 3.7 page 49 énonçant différents avis expert pour convertir une distance en distribution de masses, la règle 4 est celle qui donne la représentation la plus complète de la connaissance. Elle permet de prendre en partie compte du problème de fossé sémantique car le doute est toujours présent et quantifié malgré deux *a priori* sur les proximités.

Pour tendre vers une modélisation de la connaissance se rapprochant de la règle 4, nous proposons de transférer une partie de la croyance en la proposition H_q vers le doute Ω_q , et une partie du doute Ω_q vers la proposition \overline{H}_q . Par exemple, dans le cas où tous les voisins $\{l_q^0, l_q^1, \dots\}$ sont éloignés de l'image non étiquetée u , cette situation peut être réinterprétée en posant un nouvel *a priori* affirmant que si tous les voisins ont tendance à être éloignés, alors la masse sur la proposition \overline{H}_q doit augmenter en même temps que celle sur le doute Ω_q doit diminuer.

Il est important que cette opération de transfert de masses évite d'autre part l'apparition de conflit lors de combinaison de témoignages. Pour cela, nous posons une condition de consonance [Ram07]. La

consonance autorise de la masse sur plusieurs propositions uniquement si elles s'emboîtent. Autrement dit, il n'est possible d'avoir de la masse simultanément sur les propositions H_q et $\{H_q, \overline{H}_q\}$ d'une part, et \overline{H}_q et $\{H_q, \overline{H}_q\}$ d'autre part. Avec la consonance, il n'est donc pas possible d'avoir de la masse à la fois sur les propositions H_q et \overline{H}_q , ce qui se traduit bien par un conflit nul.

L'opération de transfert de masses redéfinit la distribution de masses m^{Ω_q} en une nouvelle distribution de masses notée $m_s^{\Omega_q}$. Nous pouvons effectuer cette opération avec un jeu de trois fonctions de type triangle (voir figure 3.14) respectant la consonance :

$$m^{\Omega_q}(\emptyset) = 0 \quad (3.20)$$

$$m^{\Omega_q}(H_q) + m^{\Omega_q}(H_q, \overline{H}_q) = 1 \quad \text{si } m^{\Omega_q}(H_q) > m_0 \quad (3.21)$$

$$m^{\Omega_q}(\overline{H}_q) + m^{\Omega_q}(H_q, \overline{H}_q) = 1 \quad \text{si } m^{\Omega_q}(H_q) \leq m_0 \quad (3.22)$$

avec m_0 un seuil fixé arbitrairement à 0.5, ne favorisant aucune des deux hypothèses H_q et \overline{H}_q . Nous

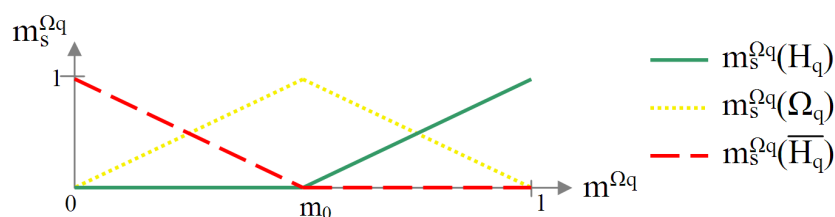


FIGURE 3.14 : Fonctions triangles pour calculer la nouvelle distribution de masses $m_s^{\Omega_q}$: en vert la fonction pour calculer la masse sur la proposition H_q , en rouge celle pour \overline{H}_q , et en jaune celle associée au doute (H_q, \overline{H}_q) .

pouvons exprimer cette opération de transfert des masses avec les notations matricielles suivantes :

Si $m^{\Omega_q}(H_q) \geq m_0$:

$$\begin{bmatrix} m_s^{\Omega_q}(H_q) \\ m_s^{\Omega_q}(\overline{H}_q) \\ m_s^{\Omega_q}(\Omega_q) \end{bmatrix} = \begin{bmatrix} \frac{1}{1-m_0} & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{1}{1-m_0} & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} m^{\Omega_q}(H_q) \\ m^{\Omega_q}(\overline{H}_q) \\ m^{\Omega_q}(\Omega_q) \end{bmatrix} + \begin{bmatrix} -\frac{m_0}{1-m_0} \\ 0 \\ \frac{1}{1-m_0} \end{bmatrix} \quad (3.23)$$

et si $m^{\Omega_q}(H_q) < m_0$:

$$\begin{bmatrix} m_s^{\Omega_q}(H_q) \\ m_s^{\Omega_q}(\overline{H}_q) \\ m_s^{\Omega_q}(\Omega_q) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -\frac{1}{m_0} & 0 & 0 \\ \frac{1}{m_0} & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} m^{\Omega_q}(H_q) \\ m^{\Omega_q}(\overline{H}_q) \\ m^{\Omega_q}(\Omega_q) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad (3.24)$$

Soit en prenant le cas particulier où $m_0 = \frac{1}{2}$: si $m^{\Omega_q}(H_q) \geq \frac{1}{2}$:

$$\begin{bmatrix} m_s^{\Omega_q}(H_q) \\ m_s^{\Omega_q}(\overline{H_q}) \\ m_s^{\Omega_q}(\Omega_q) \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} m^{\Omega_q}(H_q) \\ m^{\Omega_q}(\overline{H_q}) \\ m^{\Omega_q}(\Omega_q) \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \quad (3.25)$$

si $m^{\Omega_q}(H_q) < \frac{1}{2}$:

$$\begin{bmatrix} m_s^{\Omega_q}(H_q) \\ m_s^{\Omega_q}(\overline{H_q}) \\ m_s^{\Omega_q}(\Omega_q) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} m^{\Omega_q}(H_q) \\ m^{\Omega_q}(\overline{H_q}) \\ m^{\Omega_q}(\Omega_q) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad (3.26)$$

7. Fusion de témoignages pour l'ensemble des classes

Après avoir modélisé la connaissance concernant l'appartenance d'une image à une classe C_q , dans cette partie, nous combinons toutes les croyances sur l'appartenance pour une seule image non étiquetée u pour plusieurs classes.

7.1. Espaces de discernement

Dans la méthode originale des KnnEv, les hypothèses de type H_q d'appartenance à une classe C_q sont considérées de manière exclusive. L'exclusivité implique une dépendance entre les différentes hypothèses du type H_q . Dans ce cadre, si une image peut être étiquetée dans une classe C_q , elle ne peut pas l'être dans une autre. Par déduction, dans la méthode des KnnEv, l'hypothèse $\overline{H_q}$ peut s'exprimer par :

$$\overline{H_q} = \{H_i / H_i \in \Omega, H_i \neq H_q\} \quad (3.27)$$

Cependant, cette modélisation ne répond pas entièrement aux objectifs et aux contraintes que nous nous sommes fixés.

En effet, avec cette modélisation, la fusion des témoignages concurrent par plusieurs classes peut entraîner du conflit. La difficulté est alors de gérer ce conflit. Or, nous ne souhaitons pas que les classes entre en compétition. Nous souhaitons exprimer explicitement des cas d'ambiguïté entre deux ou plusieurs classes. Dans le cas du MCT, les hypothèses ambiguës peuvent être exprimées directement dans un cadre de discernement. L'avantage de ces hypothèses ambiguës est qu'elles peuvent servir à localiser des images non étiquetées dans des zones de doute entre plusieurs classes. L'utilisateur peut alors être libre de choisir une seule ou plusieurs étiquettes pointées par l'expression de l'hypothèse. De plus, cette approche ouvre la perspective de faire des classements de manière non exclusive et donc d'appliquer des cas de multi-étiquetage.

Le cadre de discernement doit être modifié. Nous proposons de traiter cette modélisation de la connaissance en deux temps comme l'indique la figure 3.15 :

1. Chaque classe C_q est traitée de manière indépendante, sans prendre en compte l'existence des autres classes et fournit ainsi chacune une distribution de masses $m_s^{\Omega_q}$ sur son propre cadre de discernement Ω_q .
2. Toutes les connaissances apportées par les différents cadres de discernement sont ensuite fusionnées dans un même cadre de discernement Ω .

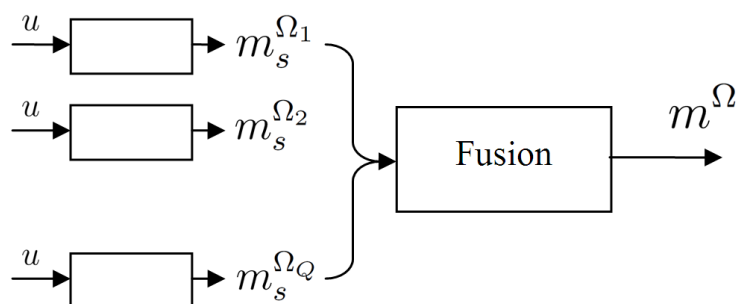


FIGURE 3.15 : Chaque classe (respectivement C_1, C_2, \dots, C_Q) fournit une distribution de masses (respectivement $m_s^{\Omega_1}, m_s^{\Omega_2}, \dots, m_s^{\Omega_Q}$) à partir des k plus proches voisins respectifs vis-à-vis d'une image non étiquetée u . Ensuite une étape de fusion consiste à combiner ces distributions de masses calculées sur les Q cadres de discernement indépendants. Le but est de synthétiser la connaissance dans un cadre de discernement global Ω , afin de préparer les données pour le niveau pignistique.

L'indépendance entre les classes implique que les croyances fournies par les témoignages d'une classe n'influence pas celles données par les témoignages des autres classes. En effet, un échantillon étiqueté l_q^i dans la classe C_q ne donne aucune information sur le fait qu'une image non étiquetée u appartienne ou non à une classe C_r différente de C_q puisque l'exclusivité des classes n'est pas considérée. A chaque classe C_q correspond donc son propre cadre de discernements Ω_q décrivant deux états possibles :

$$\Omega_q = \{H_q, \overline{H}_q\} \quad (3.28)$$

avec H_q l'hypothèse décrivant l'état "l'image appartient à la classe C_q ", et \overline{H}_q l'hypothèse décrivant l'état opposé "l'image n'appartient pas à la classe C_q ". La distribution de masses comporte alors quatre masses sur les quatre propositions de l'espace puissance 2^{Ω_q} , respectivement, le conflit, l'appartenance, la non appartenance et le doute :

$$2^{\Omega_q} = \{\emptyset, H_q, \overline{H}_q, \Omega_q\} \quad (3.29)$$

Rappelons que la consonance est respectée ce qui implique que la masse sur le conflit est toujours nulle, puisqu'il ne peut pas y avoir de la masse simultanément sur les propositions H_q et \overline{H}_q .

Dans un second temps, les cadres de discernement liés aux différentes classes sont intégrés dans un seul

cadre de discernement Ω défini comme l'espace produit de ces Q cadres de discernement locaux Ω_q :

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_Q \quad (3.30)$$

Pour comprendre l'intérêt d'utiliser un cadre de discernement différent de celui de la méthode des KnnEv, il est intéressant d'interpréter la proposition combinant les hypothèses H_1, H_2, \dots, H_Q .

Considérons un cas à 3 classes C_1, C_2 et C_3 . Dans la méthode des KnnEv, la masse $m^\Omega(\{H_1, H_2, H_3\})$ représente la masse sur le doute $m^\Omega(\Omega)$. Avoir de la masse élevée sur cette proposition informe sur le fait qu'aucune des trois classes ne se distingue des autres pour classer de manière exclusive une image dans une des trois classes. Par contre, dans la nouvelle modélisation proposée, avoir de la masse sur l'hypothèse (H_1, H_2, H_3) s'interprète différemment : elle implique une croyance élevée sur le fait que l'image appartienne aux trois classes simultanément. Le tableau 3.8 récapitule cet exemple.

Proposition	Modélisation KnnEv	Modélisation proposée
Appartenance à la classe C_1	$m^\Omega(H_1)$ "u appartient à C_1 ET n'appartient pas à C_2 , ni à C_3 "	$m^\Omega(H_1, \overline{H_2}, \overline{H_3})$ "u appartient à C_1 ET n'appartient pas à C_2 , ni à C_3 "
Proposition composée de H_1, H_2 et H_3	$m^\Omega(\{H_1, H_2, H_3\})$ "doute sur l'appartenance à l'une des classes"	$m^\Omega(H_1, H_2, H_3)$ "u appartient aux 3 classes"

TABLE 3.8 : Interprétations de deux propositions dans le cadre de discernement de la méthode des knn évidentiel et dans le cadre de discernement proposé, dans un cas à 3 classes C_1, C_2 et C_3

Dans le cadre de la théorie du MCT, chaque cadre de discernement Ω_q associé à chacune des classes peut être vu comme un *raffinement* du cadre de discernement Ω . Le nouveau cadre de discernement Ω contient alors 2^Q hypothèses. Chaque hypothèse ω_i de ce cadre de discernement Ω est une combinaison d'hypothèses "positives" du type H_q et "négative" \overline{H}_q issues des cadres de discernement $\Omega_1, \Omega_2, \dots, \Omega_Q$. L'expression générale d'une hypothèse ω_i du cadre de discernement global Ω est exprimée par :

$$\omega_i = (H_{p_1}, H_{p_2}, \dots, H_{p_P}, \overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}) \quad (3.31)$$

avec

- $\{H_{p_1}, H_{p_2}, \dots, H_{p_P}\}$ un ensemble de P hypothèses "positives" du type H_q ,
- $\{\overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}\}$ un ensemble de N hypothèses "négatives" du type \overline{H}_q ,
- de telle manière que les indices p_1, p_2, \dots, p_P et n_1, n_2, \dots, n_N désignent chacun un cadre de discernement local différent, et $P + N = Q$.

Une hypothèse ω_i est donc Q-uplet d'hypothèses du type H_q et \overline{H}_q . Mais pour le cadre de discernement Ω , une hypothèse ω_i est vue comme un singleton.

Exemple : Raffinement dans le cas à 3 classes.

Trois cadres de discernement indépendants Ω_1 , Ω_2 et Ω_3 sont chacun 3 raffinements distincts d'un nouveau cadre de discernement global Ω , comme l'illustre la figure 3.16. A chaque hypothèse ω_i de Ω correspond alors une combinaison des hypothèses des 3 sous-espaces locaux Ω_1 , Ω_2 et Ω_3 , ce qui donne $2^3 = 8$ hypothèses dans l'espace produit Ω .

$$\begin{aligned} \Omega &= \Omega_1 \times \Omega_2 \times \Omega_3 \\ &= \{(H_1, H_2, H_3), (H_1, H_2, \overline{H}_3), (H_1, \overline{H}_2, H_3), (H_1, \overline{H}_2, \overline{H}_3), \\ &\quad (\overline{H}_1, H_2, H_3), (\overline{H}_1, H_2, \overline{H}_3), (\overline{H}_1, \overline{H}_2, H_3), (\overline{H}_1, \overline{H}_2, \overline{H}_3)\} \end{aligned} \quad (3.32)$$

Ces différentes hypothèses peuvent être représentées de manière ensembliste comme dans la figure 3.17.

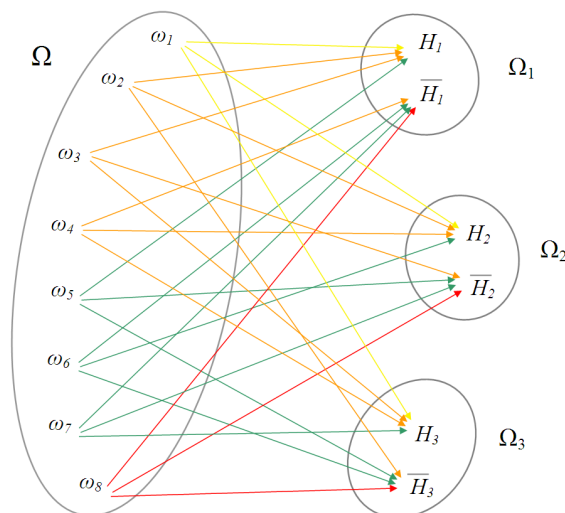


FIGURE 3.16 : Les cadres de discernement Ω_1 , Ω_2 et Ω_3 sont vus comme des raffinements d'un nouvel espace produit Ω .

7.2. Fusion multi-classe

Il s'agit maintenant de combiner les différentes distributions de masses $\{m^{\Omega_1}, m^{\Omega_2}, \dots, m^{\Omega_Q}\}$ fournies par les différents cadres de discernement "locaux" $\{\Omega_1, \Omega_2, \dots, \Omega_Q\}$ associés chacun à une classe. La nouvelle distribution de masses m^Ω est définie grâce à l'opération d'extension vide [Dia78], [Sme93]. L'opération d'extension vide est un processus conservatif de réallocation des masses de croyances issues de cadres de discernement disjoints. Chaque proposition est une combinaison de Q propositions "locales" parmi les 3 disponibles : H_q ou \overline{H}_q ou le doute $\Omega_q = \{H_q, \overline{H}_q\}$. Une proposition peut se noter (B_1, B_2, \dots, B_Q) où chaque élément B_q représente soit la proposition H_q ou \overline{H}_q ou le doute Ω_q d'un

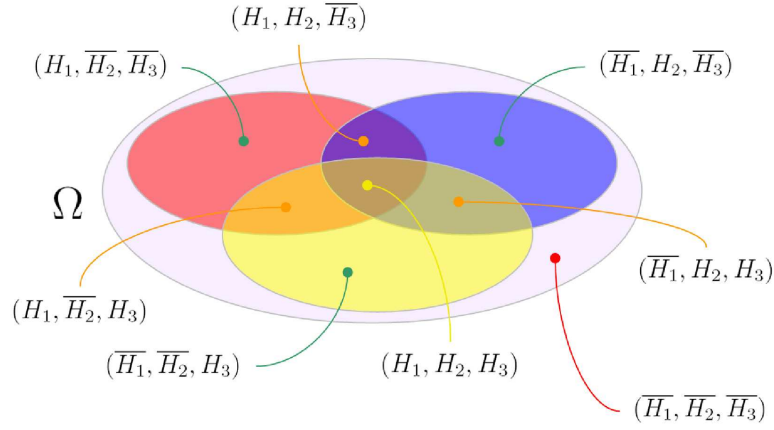


FIGURE 3.17 : Représentation ensembliste des hypothèses du cadre de discernement global Ω . Chacune de ses hypothèses est une intersection des hypothèses de base $H_1, \overline{H_1}, H_2, \overline{H_2}, H_3$ et $\overline{H_3}$ issues des trois cadres de discernement Ω_1, Ω_2 et Ω_3 .

cadre de discernement Ω_q . Avec cette opération, la nouvelle distribution de masses m^Ω contient alors 3^Q masses sur 3^Q propositions correspondant à toutes les combinaisons possibles.

Pour calculer une masse sur une proposition (B_1, B_2, \dots, B_Q) de la distribution de masses m^Ω , l'opération d'extension vide applique l'expression suivante :

$$m^\Omega(B_1, B_2, \dots, B_Q) = \prod_{B_q \in 2^{\Omega_q}} m^{\Omega_q}(B_q) \quad q \in [1, Q] \quad (3.33)$$

avec chaque $B_q \in 2^{\Omega_q}$ étant une des propositions d'un cadre de discernement "local" Ω_q .

Exemple : cas à 3 classes

Trois cadres de discernement disjoints Ω_1, Ω_2 et Ω_3 sont combinés et définissent un nouveau cadre de discernement Ω . L'opération d'extension vide permet de définir une distribution de masses sur contenant $3^3 = 27$ propositions formulées dans la figure 3.18.

Les cadres de discernement Ω_1, Ω_2 et Ω_3 fournissent 3 distributions de masses $m^{\Omega_1}, m^{\Omega_2}$ et m^{Ω_3} :

$$\begin{aligned} &\{m^{\Omega_1}(H_1), m^{\Omega_1}(\overline{H_1}), m^{\Omega_1}(\Omega_1)\} \\ &\{m^{\Omega_2}(H_2), m^{\Omega_2}(\overline{H_2}), m^{\Omega_2}(\Omega_2)\} \\ &\{m^{\Omega_3}(H_3), m^{\Omega_3}(\overline{H_3}), m^{\Omega_3}(\Omega_3)\} \end{aligned} \quad (3.34)$$

L'opérateur d'extension vide permet de calculer une distribution contenant des masses sur 27 propositions données figure 3.18 Le tableau 3.9 donne 6 exemples d'expressions (parmi 27 disponibles) des propositions de la nouvelle distribution de masses m^Ω . Remarquons que la condition de consonance impose que les distributions de masses "locales" $m^{\Omega_1}, m^{\Omega_2}$ et m^{Ω_3} , ne possèdent chacune que 2 éléments focaux (il n'y a de la masse que sur 2 propositions). En conséquence la distribution de masses finale m^Ω

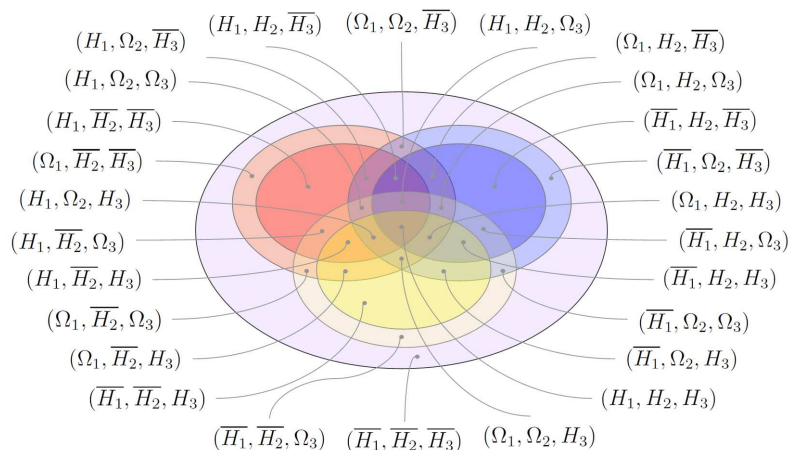


FIGURE 3.18 : Représentation ensembliste des propositions de la distribution de masses de Ω (remarque : la proposition $(\Omega_1, \Omega_2, \Omega_3)$ n'est pas représentée).

ne possède finalement que 8 éléments focaux.

Masse	Valeur
$m^\Omega(H_1, H_2, H_3)$	$m^{\Omega_1}(H_1).m^{\Omega_2}(H_2).m^{\Omega_3}(H_3)$
$m^\Omega(\overline{H_1}, H_2, H_3)$	$m^{\Omega_1}(\overline{H_1}).m^{\Omega_2}(H_2).m^{\Omega_3}(H_3)$
$m^\Omega(\Omega_1, H_2, H_3)$	$m^{\Omega_1}(\Omega_1).m^{\Omega_2}(H_2).m^{\Omega_3}(H_3)$
$m^\Omega(\Omega_1, \overline{H_2}, H_3)$	$m^{\Omega_1}(\Omega_1).m^{\Omega_2}(\overline{H_2}).m^{\Omega_3}(H_3)$
$m^\Omega(\Omega_1, \Omega_2, \overline{H_3})$	$m^{\Omega_1}(\Omega_1).m^{\Omega_2}(\Omega_2).m^{\Omega_3}(\overline{H_3})$
$m^\Omega(\Omega_1, \Omega_2, \Omega_3)$	$m^{\Omega_1}(\Omega_1).m^{\Omega_2}(\Omega_2).m^{\Omega_3}(\Omega_3)$

TABLE 3.9 : Quelques exemples de calcul de masses sur 6 propositions de la distribution de masses du cadre de discernement Ω dans le cas à 3 classes.

7.3. Fusion multi-descripteur

Jusqu'à présent, nous avons décrit toutes les étapes de fusion en ne considérant des distances entre images que selon un type de descripteur. Or, nous disposons de 2 types de descripteurs (couleurs et orientations), et nous pourrions par la suite en combiner davantage ou les remplacer par des plus informatifs.

Une première solution courante en indexation multimédia, consiste à concaténer les descripteurs en un seul même grand vecteur par image. Cette fusion "précoce" de descripteurs est facile à mettre en œuvre et l'on peut alors utiliser directement des mesures de dissimilarités sur ces vecteurs concaténés. Toutefois, un premier inconvénient est de se retrouver avec des vecteurs de très grande dimension, et il est connu que plus le nombre de dimensions augmente, plus les vecteurs ont tendance à être équidistants [SJRU99].

Un second inconvénient de cette fusion précoce, est de ne pas pouvoir distinguer facilement les apports

des différents descripteurs. Or, chaque type de descripteur n'établit pas les mêmes rapports de proximité entre les images. Les témoignages peuvent aller dans le même sens et être complémentaires (par exemple deux images ont des contenus similaires en orientation et en couleur), ou bien être contradictoires (par exemple deux images sont proches en contenu couleur, mais sont très différentes en orientation). En conséquence, il peut être intéressant de considérer dans un premier temps chaque descripteur indépendamment des autres, puis de les combiner pour avoir une modélisation de la connaissance plus complète. Il a été démontré que ce schéma de fusion désigné dans la littérature comme "tardive" donne globalement de meilleurs résultats que le schéma de fusion "précoce" dans les applications d'indexation multimédia [SWS05]. En contrepartie, ce schéma de fusion tardive nécessite de mettre en place plusieurs classifieurs, ce qui est plus coûteux en terme de temps de calcul, et pose également des problèmes de formalisation.

Dans le cadre de notre modélisation, le schéma de fusion "tardive" est donné figure 3.19. Considérons une image non étiquetée u , un ensemble de classes $\{C_1, C_2, \dots, C_Q\}$ et un ensemble de type de descripteurs distincts $\{d_1, d_2, D\}$. Le but est d'obtenir une distribution de masses m^Ω fusionnant toutes les sources d'information, afin de modéliser la connaissance quand à l'appartenance de l'image u aux classes.

Chaque type de descripteur (couleurs, orientations) est considéré indépendamment des autres. Par exemple, dans la figure le descripteur d_1 peut représenter l'histogramme de couleur dans l'espace $\{L, a, b\}$, d_2 l'histogramme des orientations. Les fusions décrites précédemment (fusions des knn puis fusions multi-classe) sont effectuées ainsi pour chaque type de descripteurs, ce qui permet de calculer les distributions de masses $m_{d_1}^\Omega, m_{d_2}^\Omega, \dots, m_D^\Omega$, décrivant les états de connaissance apportés pour chaque type de descripteurs.

Notons, par ailleurs que les k plus proches voisins de l'image non étiquetée u au sein d'une classe C_q ne sont probablement pas les mêmes selon les différents type de descripteurs.

Enfin la dernière étape consiste à fusionner toutes les distributions de masses apportées par les différents types de descripteurs $m_{d_1}^\Omega, m_{d_2}^\Omega, \dots, m_D^\Omega$ en utilisant la règle de combinaison conjonctive.

Exemple : trois classes C_1, C_2 et C_3 définissent un cadre de discernement Ω :

$$\Omega = \{(H_1, H_2, H_3), (H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (H_1, \overline{H_2}, \overline{H_3}), \quad (3.35)$$

$$(\overline{H_1}, H_2, H_3), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3), (\overline{H_1}, \overline{H_2}, \overline{H_3})\} \quad (3.36)$$

On dispose de deux types de descripteurs différents (couleurs et orientations) et les étapes de fusion décrites précédemment ont permis de calculer 2 distributions de masses m_c^Ω et m_o^Ω sur le même cadre de discernement Ω pour une image non étiquetée u en prenant en compte toutes les classes. Chaque distribution de masses contient 27 masses sur 27 propositions de 2^Ω (sachant que en pratique la condition de consonance réduit le nombre d'éléments focaux à 8 masses). Pour calculer la nouvelle distribution de masses m_{c+o}^Ω , la règle conjonctive est utilisée pour combiner les 2 distributions de masses m_c^Ω et m_o^Ω .

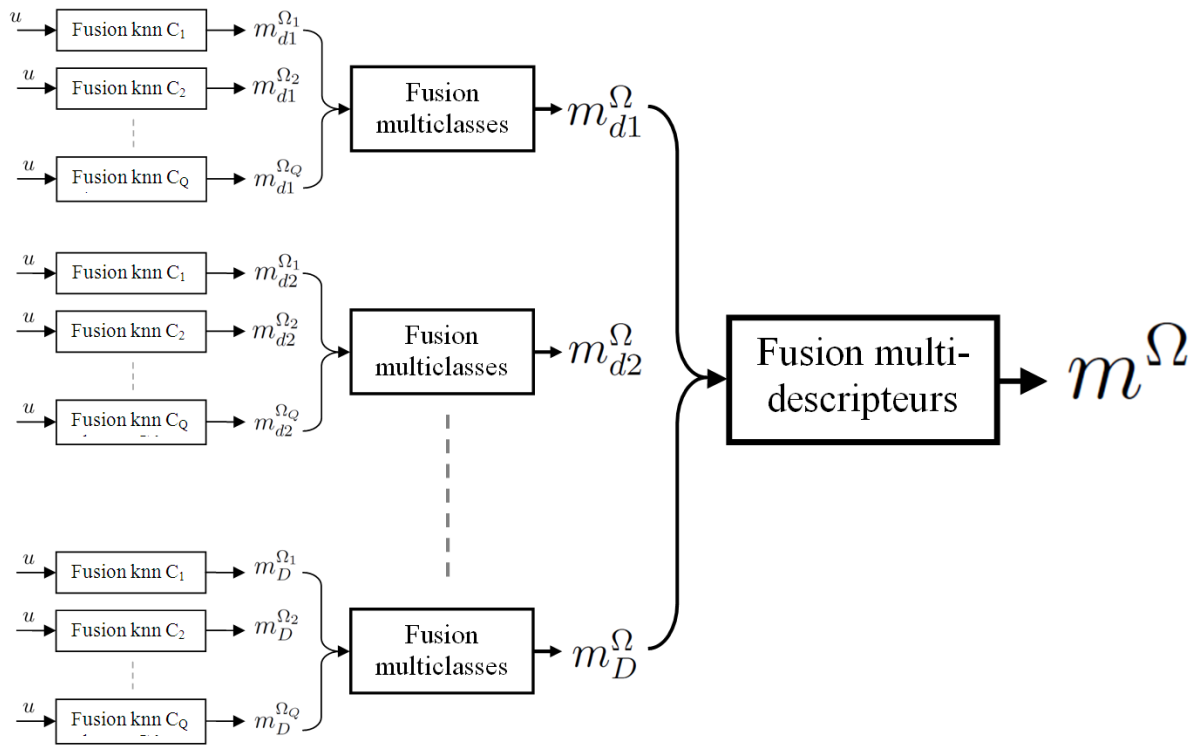


FIGURE 3.19 : Fusion des distributions de masses apportées par différents descripteurs. On retrouve les 3 échelles d'analyse décrites dans ce chapitre : des distributions de masses sont données localement pour chaque knn pour chaque classe et pour chaque type de descripteur. Puis elles sont fusionnées pour chaque type de descripteur pour considérer l'ensemble des classes. Enfin, ces fusions multi-classe sont combinées entre elles pour établir une fusion tardive de tous les témoignages apportés selon les différents descripteurs.

Par exemple, la proposition $(H_1, \Omega_2, \overline{H_3})$ possède une masse $m_{c+o}^{\Omega}(H_1, \Omega_2, \overline{H_3})$ dont l'expression est :

$$\begin{aligned}
 m_{c+o}^{\Omega}(H_1, \Omega_2, \overline{H_3}) &= m_c^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_o^{\Omega}(H_1, \Omega_2, \overline{H_3}) \\
 &+ m_c^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_o^{\Omega}(\Omega_1, \Omega_2, \overline{H_3}) \\
 &+ m_c^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_o^{\Omega}(H_1, \Omega_2, \Omega_3) \\
 &+ m_c^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_o^{\Omega}(\Omega_1, \Omega_2, \Omega_3) \\
 &+ m_o^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_c^{\Omega}(\Omega_1, \Omega_2, \overline{H_3}) \\
 &+ m_o^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_c^{\Omega}(H_1, \Omega_2, \Omega_3) \\
 &+ m_o^{\Omega}(H_1, \Omega_2, \overline{H_3}) \cdot m_c^{\Omega}(\Omega_1, \Omega_2, \Omega_3)
 \end{aligned}$$

La fusion des distributions de masses issues de plusieurs descripteurs peut entraîner l'apparition de conflit. Cela peut se produire lorsque que l'on combine les masses de 2 propositions B et C ayant une intersection vide. Concrètement, la masse sur le conflit $m^{\Omega}(\emptyset)$ correspond à la somme de toutes les

masses sur les combinaisons de propositions d'intersection vide :

$$m^\Omega(\emptyset) = \sum_{B \cap C = \emptyset} m_{d1}^\Omega(B).m_{d2}^\Omega(C) \quad (3.37)$$

Ce conflit reflète alors un désaccord entre les témoignages apportés par les différents descripteurs et apporte une information supplémentaire à considérer dans le niveau *pignistique*.

Exemple : cas à 3 classes La fusion des deux distributions de masses m_c^Ω et m_o^Ω avec la règle de combinaison conjonctive utilise $27 * 27 = 729$ combinaisons de masses pour calculer intégralement la nouvelle distribution de masses m_{c+o}^Ω . Après fusion, on retrouve les 27 propositions initiales plus 1 masse sur le conflit $m_{c+o}^\Omega(\emptyset)$. La masse sur le conflit $m_{c+o}^\Omega(\emptyset)$ correspond à la somme de toutes les masses sur les combinaisons de propositions sans d'intersection. Par exemple, l'intersection entre les propositions $(H_1, \Omega_2, \overline{H_3})$ et $(\overline{H_1}, \overline{H_2}, \overline{H_3})$ est vide. Nous pouvons écrire en effet :

$$\{(H_1, \Omega_2, \overline{H_3})\} \cap \{(\overline{H_1}, \overline{H_2}, \overline{H_3})\} = \{(\emptyset, \overline{H_2}, \overline{H_3})\} = \emptyset$$

En conséquence, la combinaison des masses suivantes contribuent toutes les deux à mettre de la masse sur le conflit :

$$m_c^\Omega(H_1, \Omega_2, \overline{H_3}).m_o^\Omega(\overline{H_1}, \overline{H_2}, \overline{H_3})$$

et

$$m_o^\Omega(H_1, \Omega_2, \overline{H_3}).m_c^\Omega(\overline{H_1}, \overline{H_2}, \overline{H_3})$$

8. Conclusion

Nous avons décrit toutes les étapes pour établir une modélisation de la connaissance quand à l'appartenance d'une image non étiquetée à des classes. Nous avons identifié à travers différentes méthodes de classification des besoins spécifiques tels que le cas multi-classe, le multi-étiquetage et l'identification de nouveauté visuelle. Nous avons alors choisi d'exprimer dans le seul même cadre formel du Modèle des Croyances Transférables la modélisation de la connaissance pour satisfaire ces besoins.

La modélisation proposée permet de décomposer la fusion d'informations en 3 échelles d'analyses. Premièrement, nous nous sommes inspirés de travaux sur une méthode de classification de type knn évidentiel "KnnEv" et nous l'avons adapté afin d'interpréter les rapports de proximité entre les knn au sein d'une même classe d'une image non étiquetée à partir de mesures imparfaites et imprécises issues de la comparaison de descripteurs visuels. Nous avons notamment modifié les expressions des fonctions de croyance de la méthode originale des "KnnEv", pour établir une distribution de masses répondant mieux à nos besoins.

Puis, nous avons utilisé une deuxième échelle d'analyse pour considérer la fusion multi-classe. Nous avons en particulier redéfini un cadre de discernement permettant une description plus en adéquation avec nos objectifs de multi-étiquetage des images. Cette étape de fusion synthétise toutes les observations entre les k plus proches voisins de toutes les classes considérées selon un type de descripteur.

Enfin, nous avons posé comme dernière échelle d'analyse un schéma de fusion tardif de toutes les sources d'information apportées par les différents descripteurs. Des distributions de masses à l'échelle multi-classe sont calculées pour chaque type de descripteurs à considérer, puis sont fusionnées. Au final, nous obtenons une distribution de masses modélisant toute la connaissance que nous avons pu extraire à partir des images membres des classes et selon plusieurs types de descripteurs, et ce, pour une seule image non étiquetée.

Dans les chapitres suivants, nous décrivons comment exploiter les informations pour un ensemble d'images non étiquetées pour lesquelles nous avons calculé une distribution de masses pour chacune d'elles.

SÉLECTION ACTIVE D'IMAGES

Description du contenu

1. Introduction	65
1.1. Problématique	65
1.2. Présentation de l'apprentissage actif	66
2. Utilisation de l'apprentissage actif dans notre contexte applicatif	68
2.1. Particularités du contexte applicatif	68
2.2. Stratégies de sélection d'images	69
2.3. Mise à jour de la connaissance	71
2.4. Traitement de la connaissance avec le Modèle des Croyances Transférables	71
3. Stratégies orientées "image"	75
3.1. Stratégies basées sur les hypothèses du cadre de discernement	76
3.1.1. Décomposition et interprétation des hypothèses du cadre de discernement	76
3.1.2. Stratégie de l'échantillon le plus positif ou "Most Positif" (MP)	78
3.1.3. Stratégie du plus rejeté ou "Most Rejected" (MR)	80
3.1.4. Stratégie du plus globalement ambigu ou "Most Global Ambiguous" (MGA)	81
3.1.5. Stratégies du plus localement ambigu ou "Most Local Ambiguous" (MLA_n)	82
3.2. Stratégies basées sur l'analyse des distributions de masses	82
3.2.1. Stratégie du plus incertain ou "Most Uncertain" (MU)	83
3.2.2. Stratégie du plus en conflit ou "Most Conflicted" (MC)	85
3.3. Bilan sur les stratégies orientées "image"	86
4. Stratégie orientée "classe"	87
5. Mise à jour de la connaissance après étiquetage des images	89
5.1. Définition de zones de connaissances	89
5.2. Méthode d'estimation	91
6. Conclusion	93

1. Introduction

1.1. Problématique

La taille des collections d'images peut être très variable, de quelques dizaines à plusieurs centaines d'images. Or, un utilisateur ne peut analyser les contenus visuels de toutes ces images en simultanément. Il

est préférable de lui proposer de traiter une image par image ou par petits lots d'images sur lesquelles il puisse porter toute son attention. Ce chapitre propose donc d'aborder le problème de sélection active d'images afin d'assister l'utilisateur dans sa tâche d'étiquetage des images.

L'idée la plus simple serait de sélectionner les images, telles qu'elles se présentent, dans l'ordre des fichiers, ou bien dans un ordre aléatoire. Cependant, cette succession de contenus visuels peut être perçue comme étant sans suite logique pour l'utilisateur, nécessitant de s'adapter en permanence.

Il peut être intéressant de contrôler la manière dont le système sélectionne des images à étiqueter. Face à la diversité des contenus visuels, un utilisateur peut appréhender de différentes manières les images non encore étiquetées. Certaines images peuvent lui sembler faciles à étiqueter si par exemple elles ressemblent visuellement à des images déjà étiquetées : l'utilisateur percevra alors rapidement si ces images peuvent être associées ou non à des étiquettes existantes. D'autres images peuvent lui paraître plus difficiles à étiqueter pour diverses raisons : par exemple, l'utilisateur peut hésiter entre plusieurs étiquettes disponibles, ou bien il ne peut trouver aucune étiquette adéquate. . .

Le problème est alors d'arriver à qualifier et quantifier des degrés de difficultés d'étiquetage sur les images afin de les exploiter pour proposer plusieurs **stratégies de sélection**. En effet, au cours de son travail de classement des images, l'utilisateur peut vouloir changer de stratégie selon le type de problème qu'il veut résoudre. Par exemple, au début, lorsque très peu d'images ont été étiquetées, l'utilisateur peut demander au système de lui fournir des images visuellement très différentes de celles étiquetées pour couvrir toute la diversité visuelle d'une même classe, ou bien, pour pouvoir identifier de nouvelles classes. A un autre moment, il peut vouloir que le système lui fournisse des images similaires aux membres des classes afin de confirmer la cohérence visuelle des classes. . .

Le précédent module de modélisation et de synthèse de la connaissance (figure 4.1) fournit en sortie un état de connaissance sur toutes les images non étiquetées. L'objectif du module de sélection active d'images est alors d'analyser cette connaissance pour organiser les images en une ou plusieurs listes de sélection en fonction de la stratégie que l'utilisateur a choisi. La première image de la liste (ou les premières images des listes) est ensuite présentée à l'utilisateur pour qu'il la classe.

A chaque fois qu'une ou plusieurs images sont étiquetées par l'utilisateur, le module réévalue les sélections pour les images restant à étiqueter. Nous proposons de nous inspirer des méthodes d'apprentissages actifs pour traiter les images de cette manière.

1.2. Présentation de l'apprentissage actif

Dans [BL07], les auteurs rappellent les origines du concept d'apprentissage actif issu des sciences pédagogiques [CE91]. L'apprentissage actif correspond à un ensemble de méthodes cherchant une participation active des élèves à leur propre formation, s'opposant ainsi à la pédagogie traditionnelle. Les centres d'intérêt de l'apprenant sont mis en avant et le rôle de l'apprentissage actif est de susciter l'esprit d'exploration et de coopération. Le rôle du professeur est alors de choisir judicieusement les mises en

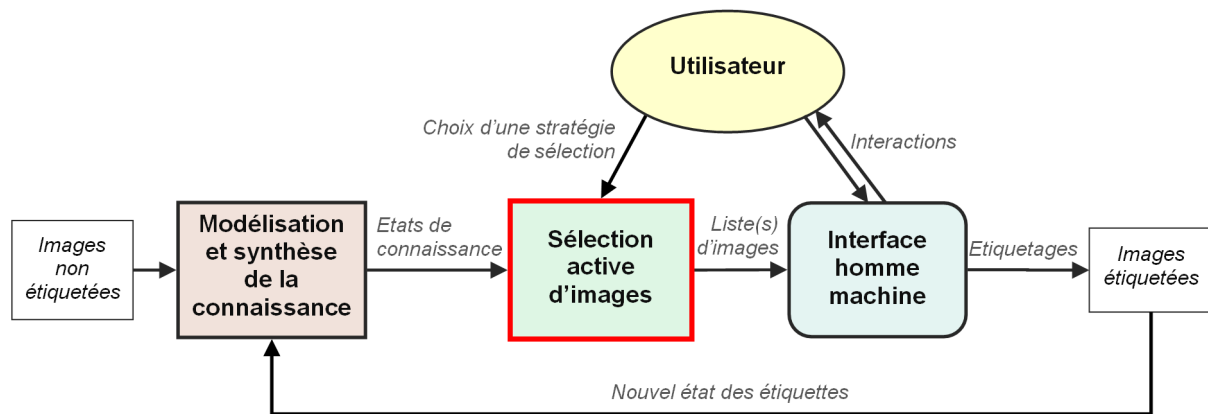


FIGURE 4.1 : Ce chapitre se focalise sur le module de sélection active des images. Ce module s'appuie sur les fonctions de croyance calculées précédemment afin d'élaborer différentes stratégies de sélections d'images non étiquetées. Les images sont ainsi ordonnées dans une ou plusieurs listes pour être présentées à l'utilisateur dans l'interface graphique.

situation pour atteindre l'objectif pédagogique le plus rapidement possible.

Par analogie, l'apprentissage actif en sciences informatiques correspond à cette recherche de situations pouvant faire progresser plus rapidement un système adaptatif. Les méthodes d'apprentissage actif ont attiré l'attention des milieux académiques et industriels ces dernières années car elles semblent prometteuses pour résoudre le problème du fossé sémantique en indexation multimédia. Par exemple, les systèmes de recherche d'images par le contenu avec boucles de retour de pertinence peuvent utiliser un composant d'apprentissage actif pour converger plus rapidement vers des réponses satisfaisant l'utilisateur [KSLC07].

Une technique d'apprentissage actif évalue les échantillons par leur utilité pour qu'un algorithme d'apprentissage s'adapte plus rapidement. En effet, il est difficile pour un système d'apprentissage de généraliser à partir de peu d'exemples. Sélectionner en priorité certains échantillons de manière judicieuse peut aider le système à généraliser plus vite. L'apprentissage actif a donc un objectif d'économie de traitements et permet d'éviter une exploration exhaustive ou aléatoire des données d'apprentissage.

Le mécanisme typique d'une méthode d'apprentissage actif dans un contexte de classification de données peut se décrire selon les étapes suivantes :

1. Considérant un nombre total de N échantillons à étiqueter.
2. Considérant $P < N$ échantillons à utiliser pour effectuer un apprentissage.
3. Initialisation en étiquetant quelques échantillons.
4. Répétition de :
 - (a) Entraînement d'un classifieur en s'appuyant sur les échantillons étiquetés.
 - (b) Sélection d'un échantillon non étiqueté selon un critère à définir, et attribution d'une étiquette (généralement par un expert humain).
5. Fin de l'entraînement après P itérations. Puis, étiquetage automatique des $N - P$ échantillons restants.

A chaque itération, tous les échantillons non étiquetés sont réévalués par un critère définissant une **stratégie** de sélection. Le traitement des données non étiquetées peut être vu comme une liste qui est triée à chaque itération, des échantillons les plus au moins représentatifs de la stratégie. Le premier échantillon de la liste est retiré et étiqueté par l'utilisateur, puis l'ordonnement de la liste est réévalué après le ré-entraînement du classifieur.

Une stratégie de sélection bien conçue doit garantir après la fin de l'entraînement de meilleures performances de classification qu'un apprentissage traditionnel qui aurait été effectué sur un même nombre d'échantillons P sélectionnés aléatoirement. Idéalement, pour des performances de classification équivalentes, une méthode basée sur l'apprentissage actif doit utiliser, de manière significative, moins d'échantillons qu'une méthode sans sélection active.

2. Utilisation de l'apprentissage actif dans notre contexte applicatif

2.1. Particularités du contexte applicatif

L'utilisation d'une méthode d'apprentissage actif est motivée par une économie de traitements en cherchant à réduire le nombre d'échantillons nécessaires à l'entraînement d'un classifieur. Généralement, un critère d'arrêt correspond à un nombre P d'étiquetages fixé à l'avance est utilisé : lorsque que P échantillons ont été étiquetés, la sélection active s'arrête. Dans notre cadre applicatif, l'objectif est différent car l'utilisateur doit traiter l'intégralité d'une collection d'images, en garantissant que toutes les étiquettes soient correctes. L'entraînement du classifieur n'est donc plus une étape préliminaire à la classification d'une collection d'images : il est permanent.

La priorité est d'aider l'utilisateur à analyser les contenus visuels pour bien "penser" et construire les classes. Cette observation peut remettre en cause en partie l'objectif souvent souhaitable de généralisation d'un classifieur [CAL94]. Si une méthode de sélection active recherche les échantillons les plus "utiles" pour optimiser l'apprentissage d'un algorithme de classification, il n'est pas garanti que cette stratégie soit "utile" pour aider l'utilisateur à construire les classes et à classer correctement les images. Par exemple, si sélectionner les images les plus ambiguës aux frontières de 2 classes est utile pour l'entraînement du classifieur, cette succession des images risque de fatiguer l'utilisateur car les images sélectionnées demandent beaucoup d'attention et de réflexion.

Nous souhaitons plus de souplesse en proposant à l'utilisateur différentes stratégies. Ainsi, il devient maître du type de sélection d'images. Certaines stratégies peuvent l'aider à étiqueter des images visuellement similaires aux contenus déjà étiquetés afin de confirmer la composition des classes. D'autres stratégies peuvent l'aider à "désambiguïser" des classes qui ont des contenus visuels qui se recouvrent. D'autres stratégies encore peuvent aider l'utilisateur à trouver des contenus visuels très différents de ceux

des classes.

2.2. Stratégies de sélection d'images

Le problème central de l'apprentissage actif est la définition d'une stratégie de sélection. Les travaux dans la littérature [BL07], [LC07] se focalisent donc sur la définition d'un critère de sélection s'appuyant souvent sur des heuristiques pour prédire l'utilité d'un échantillon.

La sélection des échantillons les plus positifs

Une première approche propose de sélectionner l'échantillon le plus probable ou "positif" (parfois appelé également le plus pertinent - "relevant" - dans le cadre de recherche d'images par exemple [CFB04]). Ainsi, à chaque itération, l'échantillon sélectionné est celui qui possède la mesure d'appartenance la plus élevée à l'une des classes. Cette approche permet de sélectionner des échantillons *a priori* situés dans le voisinage des données déjà étiquetées dans l'espace des descripteurs. Les données positives étant souvent plus rares que les données négatives, ce type de stratégie permet d'augmenter rapidement le nombre d'images dans les classes. Les échantillons positifs peuvent être considérés comme étant "faciles" à étiqueter pour l'utilisateur : en effet, le contenu visuel étant connu, l'utilisateur peut avoir moins de difficulté à trouver l'étiquette qui convient à l'échantillon sélectionné.

Un inconvénient de cette approche est qu'elle ne favorise pas qu'une généralisation rapide du classifieur. En effet, les échantillons étiquetés étant globalement similaires, la stratégie ne permet pas de couvrir toute la diversité des contenus visuels. Pour généraliser plus rapidement, il est préférable de sélectionner des échantillons plus difficiles. Plusieurs approches vont dans ce sens, sachant qu'intuitivement un échantillon peut être jugé difficile parce qu'il est ambigu, ou visuellement très différent des membres des classes, ou encore parce qu'il provoque un désaccord entre plusieurs classifieurs,...

Sélection des échantillons les plus incertains ou-et les plus ambigus

Les approches par incertitude [LC94], [TM92] consistent à prendre en compte la confiance que possède un classifieur sur les mesures d'appartenance qu'il prédit pour les données non étiquetées. Une mesure d'incertitude peut se baser directement sur les probabilités d'appartenances des échantillons aux classes : classiquement, l'échantillon possédant une distribution de probabilités proche de l'équiprobabilité, est le plus incertain. Ces approches se focalisent sur la frontière entre 2 classes [WKBD06], [TC01].

Ce type d'approche est intuitif et facile à mettre en œuvre. Il favorise la généralisation rapide dans le cas de classification binaire, car il permet de désambiguïser les frontières entre 2 classes. Une limite de ces méthodes est qu'elles se focalisent sur des zones locales entre 2 classes, sans en explorer d'autres qui pourraient être plus utiles pour le classifieur. Dans [NS04], les auteurs proposent une amélioration par un pré-clustering afin de sélectionner des échantillons incertains dissemblables entre eux.

Remarque : dans la suite du manuscrit, nous distinguerons ambiguïté et incertitude. En effet, ces 2 notions peuvent prêter à confusion dans la littérature, car de nombreux travaux abordent l'apprentissage

actif dans le cadre de problèmes bi-classe. Dans le cas précis de la classification binaire, l'échantillon le plus ambigu étant situé à la frontière des 2 classes est souvent désigné comme étant, à juste titre le plus incertain. Or, dans notre problème, nous considérons le cas multi-classe ainsi qu'éventuellement l'étiquetage multiple. Dans ce cadre, les frontières peuvent être bien plus complexes et certains échantillons peuvent posséder plusieurs étiquettes. Nous désignerons alors l'échantillon le plus ambigu celui situé le plus proche d'une frontière entre 2 ou plusieurs classes. Potentiellement, sa position intermédiaire entre plusieurs classes peut être exploitée pour prédire plusieurs étiquettes à la fois. L'échantillon le plus incertain quand à lui est l'échantillon dont la confiance sur les prédictions d'étiquettes est la plus faible, *a priori* indépendamment du fait qu'il soit ambigu ou non.

Sélection des échantillons les plus éloignés des données étiquetées

D'autres approches se basent sur les échantillons les plus éloignés de ceux précédemment étiquetés [Bri03], [WKBD06]. Dans le cadre de la classification d'images, ce type de stratégie permet d'explorer la diversité visuelle. Les échantillons sélectionnés étant très différents de ceux qui sont déjà étiquetés permettent au système d'apprentissage de généraliser, car chaque nouvel étiquetage permet de couvrir une nouvelle portion de l'espace des descripteurs.

Cependant, cette sélection peut souffrir d'un manque de représentativité des échantillons si ces derniers sont isolés (par exemple des images noires ou uni-couleurs n'intéressant pas l'utilisateur dans le cadre de classification d'images). L'ajout d'une mesure de représentativité ou densité peut aider à sélectionner des échantillons plus représentatifs d'un voisinage [XAZ07].

Sélection des échantillons les plus en désaccord

Un autre type de stratégie sélectionne en priorité les échantillons difficiles à étiqueter parce qu'ils sont conflictuels, au sens de plusieurs modèles de classification. Le but est de traiter en priorité les échantillons qui maximisent une mesure de désaccord de prédiction d'étiquettes entre différents classifieurs [SOS92], [FST97]. L'idée est de trouver rapidement le meilleur modèle de classification qui s'adapte le mieux aux données à traiter [MM04].

Un des avantages est que le vote par majorité est en général robuste. Par contre, ce type de stratégie peut être difficile à mettre en œuvre car il faut choisir les classifieurs implémentés, si possible de manière à ce qu'ils soient complémentaires.

Autres approches

La plupart des heuristiques exprimées à travers ces différentes stratégies correspondent à un choix entre une approche exploratrice, cherchant à trouver des échantillons dans des zones non connues comme dans le cas de la diversité, ou une approche cherchant à raffiner la connaissance sur des zones connues difficiles à discerner comme dans le cas des plus ambigus.

Plusieurs approches tentent un compromis en combinant plusieurs critères de sélection complémentaires [GC04], [CFB04]. Dans [WKBD06] par exemple, les auteurs pondèrent les critères du plus incertain, du plus représentatif au sens de la densité, et du plus dissemblable. L'avantage est de pouvoir exprimer

des critères de sélection plus riches, mais la difficulté est alors de trouver une pondération efficace des critères. Plusieurs stratégies peuvent être également considérées en simultanément : dans [OKS05] les auteurs cherchent à retenir la stratégie la plus utile à chaque itération, ou dans [BEYL04] où l'échantillon sélectionné est celui qui provoque le consensus selon différentes stratégies.

Il existe également d'autres approches comme l'échantillonnage par réduction de l'erreur de généralisation [RM01] nécessitant de définir une fonction d'estimation et qui est exhaustive (donc plus coûteuse) car toutes les valeurs d'étiquettes sont envisagées pour tous les échantillons.

L'apprentissage actif appliqué au multi-étiquetage d'images est à ce jour assez peu abordé. Dans [QHR+08] les auteurs développent une approche originale prenant en compte la corrélation entre les classes pour sélectionner les échantillons.

2.3. Mise à jour de la connaissance

Une méthode d'apprentissage actif classique repose sur deux étapes en boucle, l'étiquetage par un utilisateur d'un échantillon sélectionné par une stratégie, puis l'entraînement d'un classifieur sur les données étiquetées. Il faut bien remarquer que l'ajout de ces nouveaux échantillons étiquetés peut potentiellement fortement modifier la connaissance des classes à chaque itération.

La figure 4.2 donne une illustration dans un cas simple se focalisant sur une seule classe. Dans cette figure, les échantillons sont disposés dans le plan en accord avec les distances de descripteurs. A la première itération, 3 échantillons ont été étiquetés dans une même classe. Cinq autres échantillons non étiquetés sont évalués par une stratégie de type "le plus éloigné" des échantillons précédemment étiquetés. En conséquence, à l'itération 1, l'échantillon le plus éloigné est numéroté "1", et le moins éloigné est numéroté "5". Entre les 2 itérations, il est demandé à l'utilisateur d'étiqueter l'échantillon le plus éloigné numéroté "1". L'utilisateur décide de classer cet échantillon dans la même classe que les 3 premiers échantillons. La connaissance est alors réévaluée et cette nouvelle configuration des données étiquetées modifie complètement le jugement sur le reste des échantillons non étiquetés en tant qu'échantillons éloignés. Ainsi, à la deuxième itération, l'échantillon le plus éloigné, noté maintenant "1" était à l'itération 1 le moins éloigné, et inversement l'échantillon le moins éloigné noté maintenant "4", était le second plus éloigné l'itération précédente.

2.4. Traitement de la connaissance avec le Modèle des Croyances Transférables

Dans les structures classiques, un composant d'apprentissage actif est conçu indépendamment des classifieurs utilisés en amont. Dans notre cas, nous formalisons toutes les stratégies de sélection proposées avec le Modèle des Croyances Transférable (MCT) déjà utilisé par le module précédent de modélisation et synthèse de la connaissance. Ce choix permet d'éviter tout problème éventuel d'échelle des infor-

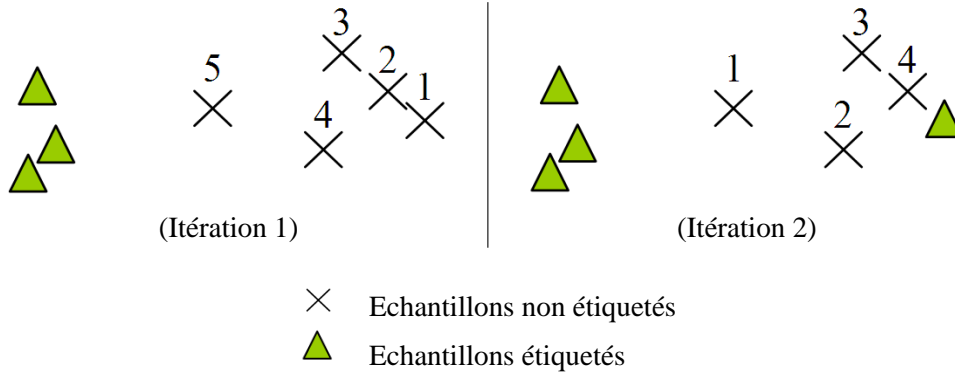


FIGURE 4.2 : Illustration de la mise à jour de la connaissance : dans ce cas très simple, les échantillons non étiquetés (croix) sont triés par une mesure d'éloignement des échantillons étiquetés. A l'itération 1, le plus éloigné est noté "1" et le moins éloigné est noté "5". L'utilisateur étiquette l'échantillon le plus éloigné. A l'itération 2, la connaissance a changé : par exemple, l'échantillon le plus éloigné, noté maintenant "1" était à l'itération précédente celui qui était le moins éloigné.

mations à fusionner : les données sont représentées par des fonctions de croyance qui sont directement analysées et manipulées pour établir des critères de sélection.

Ce parti pris est cohérent avec le MCT car nous avons pris le soin de bien séparer tout le processus en 2 étapes. Le module précédent représente le **niveau crédal** (du latin *credo* "je crois") et a permis de formaliser la connaissance. Le module de sélection d'images nécessite d'émettre des avis sur les images non étiquetées pour pouvoir les comparer, ce qui est typique du **niveau pignistique** (du latin *pignus* "pari").

Avant de décliner les différentes stratégies, il convient de rappeler brièvement comment la connaissance est représentée en entrée du module de sélection d'images.

Considérant un ensemble d'images non étiquetées $U = \{u_1, u_2, \dots, u_N\}$, le module de modélisation et de synthèse de la connaissance fournit un état de connaissance pour chaque image en fonction de toutes les images étiquetées disponibles et selon plusieurs descripteurs.

Chaque étiquette q disponible est associée à une classe C_q et un cadre de discernement local Ω_q :

$$\Omega_q = \{H_q, \overline{H}_q\} \tag{4.1}$$

avec H_q l'hypothèse décrivant l'état "l'image non étiquetée u appartient à la classe C_q " et \overline{H}_q l'hypothèse décrivant l'état "l'image non étiquetée u n'appartient pas à la classe C_q ".

Un cadre de discernement global Ω est défini comme l'espace produit de Q cadres de discernements locaux Ω_q associés aux Q classes :

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_Q \tag{4.2}$$

Des distributions de masses $m_{u,d}^\Omega$ sont calculées sur 2^Ω , pour chaque couple "image non étiquetée u "

et "descripteur d ". Ces nouvelles distributions contiennent chacune des masses sur 3^Q propositions globales.

Ainsi, chaque image non étiquetée u est associée à autant de distributions de masses que de descripteurs sont considérés. La dernière étape de fusion, consiste à combiner avec la règle conjonctive les distributions de masses calculées pour chaque descripteur. A la fin, chaque image non étiquetée u possède sa distribution de masses m_u^Ω , le résultat de toutes les fusions de sources d'informations, où $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_Q$ et $\Omega_q = \{H_q, \overline{H}_q\}$.

Cette connaissance peut être exploitée pour établir des stratégies de sélection d'images. Dans le cadre du MCT, la *distribution de probabilités pignistiques* [Sme05] notée $BetP\{m^\Omega\}$ est utilisée pour estimer l'hypothèse d'un cadre du discernement Ω la plus probable. Formellement, la probabilité pignistique d'une hypothèse ω_i d'un cadre de discernement Ω est :

$$BetP\{m^\Omega\}(\omega_i) = \frac{1}{1 - m^\Omega(\emptyset)} \sum_{B \subseteq \Omega, \omega \in B} \frac{m^\Omega(B)}{|B|} \quad (4.3)$$

où \emptyset représente le conflit, et B une proposition de la distribution de masse m^Ω sur laquelle de la croyance a été affectée. Au regard de l'expression, la transformée pignistique consiste à répartir de manière équiprobable la masse d'une proposition B sur les hypothèses contenues dans B .

Exemple : cas à 3 classes

Une distribution de masses m_u^Ω a été calculée pour 1 image non étiquetée u dans un cadre de discernement global Ω , issu de la combinaison de 3 cadres de discernement locaux indépendants Ω_1 , Ω_2 et Ω_3 :

$$\begin{aligned} \Omega &= \Omega_1 \times \Omega_2 \times \Omega_3 \\ &= \{(H_1, H_2, H_3), (H_1, H_2, \overline{H}_3), (H_1, \overline{H}_2, H_3), (H_1, \overline{H}_2, \overline{H}_3), \\ &\quad (\overline{H}_1, H_2, H_3), (\overline{H}_1, H_2, \overline{H}_3), (\overline{H}_1, \overline{H}_2, H_3), (\overline{H}_1, \overline{H}_2, \overline{H}_3)\} \end{aligned} \quad (4.4)$$

La distribution de masses m_u^Ω contient au total 28 propositions (voir figure 4.3). Ces masses sont utilisées pour calculer la distribution de probabilités pignistiques sur les 8 hypothèses du cadre de discernement Ω . La masse d'une proposition est prise en compte dans le calcul d'une probabilité pignistique d'une hypothèse uniquement si cette proposition contient l'hypothèse. Par exemple, la proposition $(\Omega_1, \overline{H}_2, H_3)$ peut se réécrire avec les deux hypothèses du cadre de discernement Ω :

$$(\Omega_1, \overline{H}_2, H_3) = ((H_1, \overline{H}_2, H_3), (\overline{H}_1, \overline{H}_2, H_3)) \quad (4.5)$$

En conséquence, la proposition $(\Omega_1, \overline{H}_2, H_3)$ participe uniquement aux calculs des 2 probabilités pignistiques des hypothèses $(H_1, \overline{H}_2, H_3)$ et $(\overline{H}_1, \overline{H}_2, H_3)$. Cette proposition possède donc une cardinalité $|(\Omega_1, \overline{H}_2, H_3)| = 2$, c'est-à-dire le nombre d'hypothèses du cadre de discernement Ω qu'elle contient (d'autres exemples de cardinalité de proposition sont donnés dans le tableau 4.1). Nous pouvons calculer

par exemple la probabilité pignistique de l'hypothèse $(H_1, \overline{H_2}, H_3)$:

$$\begin{aligned}
 \text{Bet}P\{m_u^\Omega\}(H_1, \overline{H_2}, H_3) = & \frac{1}{1 - m_u^\Omega(\emptyset)} \left(m_u^\Omega(H_1, \overline{H_2}, H_3) \right. \\
 & + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, H_3)}{2} + \frac{m_u^\Omega(H_1, \Omega_2, H_3)}{2} + \frac{m_u^\Omega(H_1, \overline{H_2}, \Omega_3)}{2} \\
 & + \frac{m_u^\Omega(\Omega_1, \Omega_2, H_3)}{4} + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, \Omega_3)}{4} + \frac{m_u^\Omega(H_1, \Omega_2, \Omega_3)}{4} \\
 & \left. + \frac{m_u^\Omega(\Omega_1, \Omega_2, \Omega_3)}{8} \right) \quad (4.6)
 \end{aligned}$$

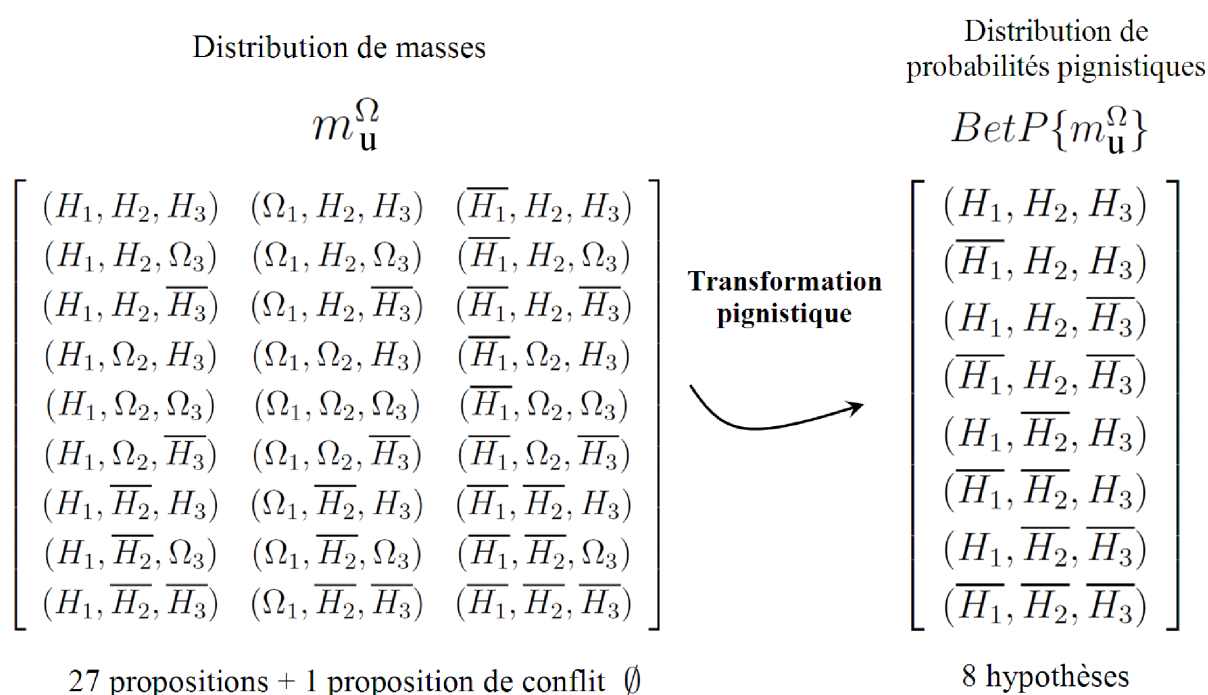


FIGURE 4.3 : Transformation pignistique dans le cas à 3 classes. Toutes les propositions de la distribution de masses sont utilisées pour calculer la distribution de probabilités pignistiques des hypothèses du cadre de discernement Ω .

A un instant t , on dispose d'un ensemble d'images classées et d'un ensemble d'images à classer. Le problème est alors de définir la "meilleure" image à classer selon un critère de sélection. A partir des fonctions de croyance sont déclinées différentes stratégies de sélection d'images proposées à l'utilisateur. Chaque image non étiquetée $u \in U$ est associée à sa propre distribution de probabilités pignistiques $\text{Bet}P\{m_u^\Omega\}$, et ces probabilités sont exploitées directement pour comparer les images non étiquetées entre elles en fonction des hypothèses du cadre de discernement Ω choisi par l'utilisateur.

Dans la suite de ce chapitre, nous allons distinguer deux types de stratégies :

1. **Les stratégies orientées "image" :** correspondant à l'approche classique de l'apprentissage actif.

Proposition	Notation complète	Cardinal	Masse
$(H_1, H_2, \overline{H_3})$	$(H_1, H_2, \overline{H_3})$	1	$m^{\Omega_1}(H_1).m^{\Omega_2}(H_2).m^{\Omega_3}(\overline{H_3})$
$(H_1, \Omega_2, \overline{H_3})$	$\left((H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, \overline{H_3}) \right)$	2	$m^{\Omega_1}(H_1).m^{\Omega_2}(\Omega_2).m^{\Omega_3}(\overline{H_3})$
$(\Omega_1, \Omega_2, \overline{H_3})$	$\left((H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, \overline{H_3}), \right.$ $\left. (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, \overline{H_3}) \right)$	4	$m^{\Omega_1}(\Omega_1).m^{\Omega_2}(\Omega_2).m^{\Omega_3}(\overline{H_3})$
$(\Omega_1, \Omega_2, \Omega_3)$	$\left((H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, \overline{H_3}), \right.$ $\left. (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, \overline{H_3}), \right.$ $\left. (H_1, H_2, H_3), (H_1, \overline{H_2}, H_3), \right.$ $\left. (\overline{H_1}, H_2, H_3), (\overline{H_1}, \overline{H_2}, H_3) \right)$	8	$m^{\Omega_1}(\Omega_1).m^{\Omega_2}(\Omega_2).m^{\Omega_3}(\overline{\Omega_3})$

TABLE 4.1 : Exemple de propositions de la distribution de masses m^Ω . La notation complète indique explicitement les hypothèses de Ω contenant la proposition.

A chaque itération, toutes les images non étiquetées sont évaluées par un critère propre à la stratégie choisie par l'utilisateur afin de trouver quelle image doit être étiquetée en priorité. Ce type de stratégie convient parfaitement pour un mode de traitement "image par image", ou pas à pas.

2. **Les stratégies orientées "classe"** : où le but est de proposer à l'utilisateur des lots d'images cohérents, pouvant être potentiellement regroupés sous la même étiquette. Cette approche plus globale favorise l'étiquetage par lot d'images et permet un gain de productivité, sans réévaluer toute la connaissance à chaque image étiquetée.

3. Stratégies orientées "image"

Les stratégies orientées "image" requièrent que le module de sélection active d'images ordonne toutes les images non étiquetées par un critère lié à la stratégie choisie par l'utilisateur et l'état de connaissance. Ces critères de sélection doivent être intuitifs pour que l'utilisateur puisse maîtriser quelle stratégie qui lui convient le mieux selon ses besoins à un instant donné. Or, la formalisation avec le MCT possède l'avantage d'être basée sur des représentations symboliques relativement faciles à interpréter par l'utilisateur, notamment à travers les hypothèses du cadre de discernement Ω .

3.1. Stratégies basées sur les hypothèses du cadre de discernement

3.1.1. Décomposition et interprétation des hypothèses du cadre de discernement

Le cadre de discernement $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_Q$ (Q étant le nombre de classes) décrit toutes les hypothèses identifiées qui peuvent être associées à une image non étiquetée u . L'expression générale d'une hypothèse ω_i du cadre de discernement global Ω est exprimée par :

$$\omega_i = (H_{p_1}, H_{p_2}, \dots, H_{p_N}, \overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}) \quad (4.7)$$

avec

- $\{H_{p_1}, H_{p_2}, \dots, H_{p_P}\}$ un ensemble de P hypothèses "positives" du type H_q ,
- $\{\overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}\}$ un ensemble de N hypothèses "négatives" du type \overline{H}_q ,
- de telle manière que les indices p_1, p_2, \dots, p_P et n_1, n_2, \dots, n_N désignent chacun un cadre de discernement local différent, et $P + N = Q$.

Le cadre de discernement Ω peut être subdivisé en $Q + 1$ sous-ensembles d'hypothèses. Chaque sous-ensemble représente un ensemble de combinaisons du même nombre de P d'hypothèses locales "positives" H_q et du même nombre de N d'hypothèses locales "négatives" \overline{H}_q . La taille de chaque sous-ensemble est donnée par le coefficient binomial C_Q^P .

Exemple : cas à 3 classes.

Un cadre de discernement global Ω , issu de la combinaison de 3 cadres de discernement locaux indépendants Ω_1, Ω_2 et Ω_3 , contient alors $2^3 = 8$ hypothèses :

$$\begin{aligned} \Omega &= \Omega_1 \times \Omega_2 \times \Omega_3 \\ &= \{(H_1, H_2, H_3), (H_1, H_2, \overline{H}_3), (H_1, \overline{H}_2, H_3), (H_1, \overline{H}_2, \overline{H}_3), \\ &\quad (\overline{H}_1, H_2, H_3), (\overline{H}_1, H_2, \overline{H}_3), (\overline{H}_1, \overline{H}_2, H_3), (\overline{H}_1, \overline{H}_2, \overline{H}_3)\} \end{aligned} \quad (4.8)$$

Ce cadre de discernement Ω est décomposé en 3+1 sous-ensembles d'hypothèses :

- $C_3^0 = 1$ hypothèse globale sans hypothèse locale positive : $(\overline{H}_1, \overline{H}_2, \overline{H}_3)$,
- $C_3^1 = 3$ hypothèses globales avec 1 hypothèse locale positive : $(H_1, \overline{H}_2, \overline{H}_3)$, $(\overline{H}_1, H_2, \overline{H}_3)$ et $(\overline{H}_1, \overline{H}_2, H_3)$
- $C_3^2 = 3$ hypothèses globales avec 2 hypothèses locales positives : $(H_1, H_2, \overline{H}_3)$, $(H_1, \overline{H}_2, H_3)$, et $(\overline{H}_1, H_2, H_3)$,
- $C_3^3 = 1$ hypothèse globale avec 3 hypothèses locales positives (H_1, H_2, H_3) .

Cette décomposition Ω peut être interprétée de manière plus intuitive. Ces sous-ensembles peuvent être en effet regroupés en 3 entités : une hypothèse de *rejet*, un ensemble d'hypothèses *positives*, et un ensemble d'hypothèses *ambiguës*.

Hypothèse de rejet

L'hypothèse de rejet ω_r est constituée exclusivement d'hypothèses locales "négatives" du type \overline{H}_q :

$$\omega_r = (\overline{H}_1, \overline{H}_2, \dots, \overline{H}_Q) \quad (4.9)$$

Cette hypothèse de rejet ω_r est utilisée pour exprimer un état de connaissance affirmant qu'aucune des Q classes disponibles ne convient pour une image non étiquetée u .

Exemple : cas à 3 classes. Dans l'exemple précédent à 3 classes, l'hypothèse de rejet est notée par : $(\overline{H}_1, \overline{H}_2, \overline{H}_3)$.

Hypothèses positives

Une hypothèse positive ω_p^q se compose d'une seule hypothèse locale positive du type H_q , le reste étant des hypothèses locales négatives du type \overline{H}_q :

$$\omega_p^q = (H_q, \overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}) \quad (4.10)$$

avec q identifiant le cadre de discernement local fournissant l'hypothèse positive H_q , et, n_1, n_2, \dots, n_N les identifiants des cadres de discernement locaux fournissant les hypothèses locales négatives $\overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}$. Considérant Q cadres de discernement associés à Q classes, il y a $C_Q^1 = Q$ hypothèses positives, soit une par classe. Une hypothèse positive ω_p^q exprime un état de connaissance affirmant qu'une image non étiquetée u peut être classée uniquement dans la seule classe C_q . Ce type d'hypothèse convient parfaitement au classement de manière exclusive dans une seule classe.

Exemple : cas à 3 classes. Dans l'exemple précédent à 3 classes, le cadre de discernement Ω comporte 3 hypothèses positives exprimant chacun l'appartenance d'une image non étiquetée u à l'une des 3 classes disponibles :

- $(H_1, \overline{H}_2, \overline{H}_3)$: l'appartenance à la classe C_1 ,
- $(\overline{H}_1, H_2, \overline{H}_3)$: l'appartenance à la classe C_2 ,
- $(\overline{H}_1, \overline{H}_2, H_3)$: l'appartenance à la classe C_3 .

Hypothèses ambiguës

L'ensemble des hypothèses restantes du cadre de discernement Ω expriment des cas d'ambiguïté. Ces hypothèses sont constituées de P (avec $P \geq 2$) hypothèses locales positives du type H_q et de N (avec $N \leq Q - 2$ et $P + N = Q$) hypothèses locales négatives \overline{H}_q . La forme générale d'une hypothèse ambiguë est donnée par :

$$\omega_a^P = (H_{p_1}, H_{p_2}, \dots, H_{p_P}, \overline{H}_{n_1}, \overline{H}_{n_2}, \dots, \overline{H}_{n_N}) \quad (4.11)$$

avec p_1, p_2, \dots, p_P les identifiants des cadres de discernement locaux fournissant des hypothèses positives $H_{p_1}, H_{p_2}, \dots, H_{p_P}$, et, avec n_1, n_2, \dots, n_N les identifiants des cadres de discernement locaux fournissant des hypothèses négatives $\overline{H_{n_1}}, \overline{H_{n_2}}, \dots, \overline{H_{n_N}}$.

Une hypothèse ambiguë ω_a^P possède un degré d'ambiguïté P désignant le nombre d'hypothèses locales positives. Une hypothèse ambiguë ω_a^P exprime donc qu'une image non étiquetée u peut être classée dans P classes $\{C_{p_1}, C_{p_2}, \dots, C_{p_P}\}$, et qu'elle ne peut pas l'être dans les autres classes $\{\overline{C_{n_1}}, \overline{C_{n_2}}, \dots, \overline{C_{n_N}}\}$ disponibles. Ce type d'hypothèse permet donc d'éliminer des classes non pertinentes pour y classer une image non étiquetée u .

Parmi toutes hypothèses ambiguës, l'hypothèse d'ambiguïté globale est à noter :

$$\omega_{ga} = (H_1, H_2, \dots, H_Q) \quad (4.12)$$

Cette hypothèse modélise le cas où une image non étiquetée u peut potentiellement appartenir à toutes les classes disponibles. En général, cette hypothèse n'est pas très pertinente pour de la classification d'images, car il est peut probable qu'une image puisse être associée à toutes les classes disponibles. Cependant, cette hypothèse peut souligner un manque de précision des informations apportées par les sources.

Exemple : cas à 3 classes. Dans l'exemple précédent à 3 classes, le cadre de discernement Ω comporte 4 hypothèses ambiguës pouvant s'interpréter de la manière suivante pour une image non étiquetée u :

- $(H_1, H_2, \overline{H_3})$: l'hypothèse d'appartenance aux classes C_1 et C_2 , et la non appartenance à la classe C_3 ,
- $(H_1, \overline{H_2}, H_3)$: l'hypothèse d'appartenance aux classes C_1 et C_3 , et la non appartenance à la classe C_2 ,
- $(\overline{H_1}, H_2, H_3)$: l'hypothèse d'appartenance aux classes C_2 et C_3 , et la non appartenance à la classe C_1 ,
- (H_1, H_2, H_3) : l'hypothèse d'appartenance à toutes les classes à la fois C_1, C_2 et C_3 .

Trois hypothèses décrivent donc des ambiguïtés locales entre 2 classes, et une dernière hypothèse représente l'ambiguïté globale.

3.1.2. Stratégie de l'échantillon le plus positif ou "Most Positif" (MP)

Cette stratégie s'intéresse à estimer quelle image non étiquetée u de U est la plus positive au sens des probabilités pignistiques. Dans un premier temps, pour chaque image non étiquetée $u \in U$, le maximum de probabilité pignistique $PP(u)$ est calculé sur Ω_P le sous-ensemble contenant uniquement les hypothèses positives du cadre de discernement Ω .

$$PP(u) = BetP_{max}^{\Omega_P}(u) = \max_{w_i \in \Omega_P} BetP\{m_u^\Omega\}(w_i) \quad (4.13)$$

Par exemple, dans le cas à 3 classes, chaque image $u \in U$ est associée à la valeur $PP(u)$ correspond au maximum de probabilité pignistique sur le sous-ensemble :

$$\Omega_P = \{(H_1, \overline{H}_2, \overline{H}_3), (\overline{H}_1, H_2, \overline{H}_3), (\overline{H}_1, \overline{H}_2, H_3)\} \quad (4.14)$$

Puis, dans un second temps, les images de U sont comparées grâce à leur maximum de probabilités pignistiques sur les hypothèses positives pour sélectionner l'image u_{mp} la plus positive :

$$u_{mp} = \operatorname{argmax}_{u \in U} PP(u) \quad (4.15)$$

La figure 4.4 illustre le comportement attendu par cette stratégie sur un exemple de 11 images. Dans ce cas simple, les images non étiquetées les plus proches de celles déjà étiquetées sont les plus positives. Elles sont certainement plus faciles à classer pour l'utilisateur car le contenu visuel des images non étiquetées est similaire à celui des images étiquetées.

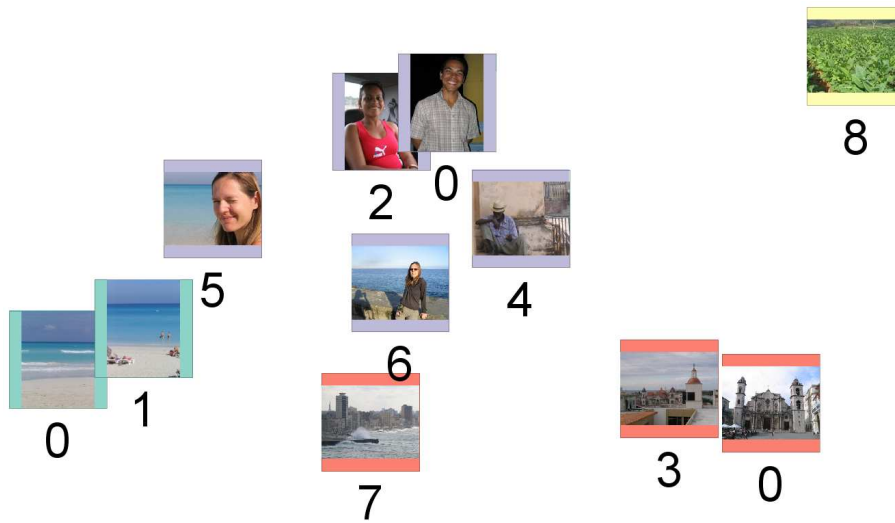


FIGURE 4.4 : Illustration de la sélection d'échantillons par une stratégie du plus positif (MP). Le placement des images est corrélé avec des mesures de similarités visuelles. Les images exemples numérotées "0" ont été étiquetées par l'utilisateur pour initialiser 3 classes distinctes : "mer", "ville" et "portrait". La couleur des cadres des images quant à elle correspond à une vérité terrain établie par un utilisateur. Les images non étiquetées sont numérotées dans l'ordre croissant de la plus à la moins positive. Les images non étiquetées les plus proches des images exemples des classes sont les plus positives. L'image la plus éloignée (numérotée 8) est la plus visuellement différente des images exemples et sera certainement sélectionnée en dernier.

3.1.3. Stratégie du plus rejeté ou "Most Rejected" (MR)

La stratégie du plus rejeté consiste à sélectionner l'image non étiquetée u_{mr} possédant la probabilité pignistique $BetP_r(u)$ la plus élevée sur l'hypothèse de rejet ω_r du cadre de discernement Ω :

$$u_{mr} = \operatorname{argmax}_{u \in U} BetP_r(u) \quad (4.16)$$

avec pour chaque image $u \in U$, $BetP_r(u)$ la probabilité pignistique calculée sur l'hypothèse de rejet ω_r :

$$BetP_r(u) = BetP\{m_u^\Omega\}(\omega_r) \quad (4.17)$$

Cette stratégie revient à sélectionner à chaque itération l'image la plus éloignée de toutes les images étiquetées. La figure 4.5 illustre le comportement attendu par cette stratégie sur les mêmes données de la figure 4.4 précédente. L'image la plus rejetée permet d'identifier un nouveau type de contenu visuel. L'utilisateur devra alors choisir de classer cette image dans une des classes existantes, ou bien de définir une nouvelle classe à partir de ce premier exemple de contenu visuel.

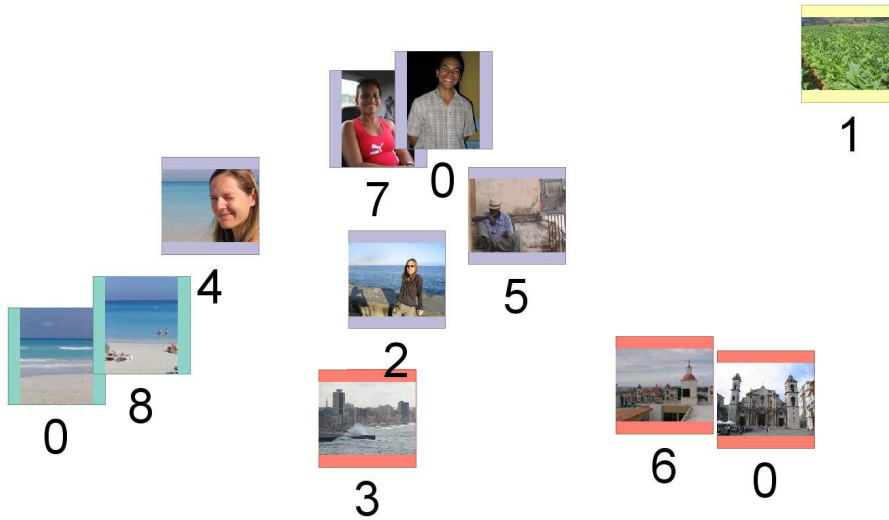


FIGURE 4.5 : Illustration de la sélection d'échantillons par une stratégie du plus rejeté (MR). A la première itération, l'image sélectionnée la première, notée "1", est la plus éloignée de toutes les images membres des classes notées "0". A l'inverse les images les moins rejetées sont celles qui sont les plus proches des images exemples des classes (notées "6", "7" et "8").

3.1.4. Stratégie du plus globalement ambigu ou "Most Global Ambiguous" (MGA)

L'échantillon le plus globalement ambigu u_{mga} possède la probabilité pignistique la plus élevée sur l'hypothèse d'ambiguïté globale ω_{ga} :

$$u_{mga} = \operatorname{argmax}_{u \in U} \operatorname{Bet}P_{ga}(u) \quad (4.18)$$

avec pour chaque image $u \in U$, $\operatorname{Bet}P_{ga}(u)$ la probabilité pignistique calculée sur l'hypothèse d'ambiguïté globale ω_{ga} :

$$\operatorname{Bet}P_{ga}(u) = \operatorname{Bet}P\{m_u^\Omega\}(\omega_{ga}) \quad (4.19)$$

La figure 4.6 illustre le comportement attendu par cette stratégie sur les mêmes données de la figure 4.4. Dans cet exemple, l'image la plus globalement ambiguë est celle qui est la plus proche du centre de gravité théorique des 3 images étiquetées. Cette image est représentative d'un manque d'information apporté par les descripteurs et les images membres de classe. D'après la connaissance modélisée à travers la distribution de masses, l'image la plus globalement ambiguë n'est pas la plus rejetée, mais il est difficile de cibler précisément quelle est l'étiquette la plus pertinente pour cette image.

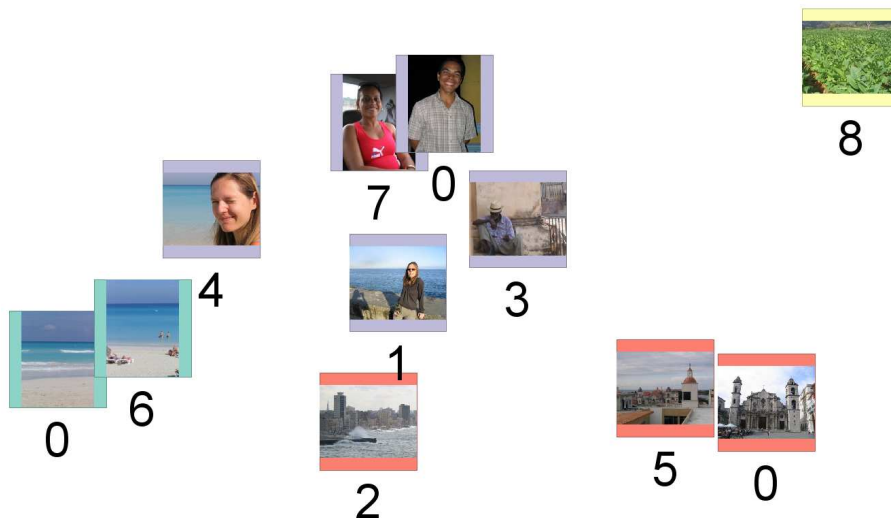


FIGURE 4.6 : Illustration de la sélection d'échantillons par une stratégie du plus globalement ambigu. L'image la plus centrale, située près du centre de gravité du triangle formé par les 3 images étiquetées "0", est celle qui est sélectionnée la première par la stratégie MGA.

3.1.5. Stratégies du plus localement ambigu ou "Most Local Ambiguous" (MLA_n)

Ce type de stratégie s'intéresse à estimer quelle est l'image non étiquetée u de U la plus ambiguë selon un degré n correspondant au nombre de classes en jeu à considérer. Dans un premier temps, pour chaque image non étiquetée $u \in U$, le maximum de probabilité pignistique $PLA_n(u)$ est calculé sur Ω_{LA_n} le sous-ensemble contenant uniquement les hypothèses d'ambiguïté locale entre n classes du cadre de discernement Ω .

$$PLA_n(u) = BetP_{max}^{\Omega_{LA_n}}(u) = \max_{w_i \in \Omega_{LA_n}} BetP\{m_u^\Omega\}(w_i) \quad (4.20)$$

Puis, dans un second temps, les images de U sont comparées par leur maximum de probabilités pignistiques $PLA_n(u)$ sur les hypothèses d'ambiguïté locale de degré n pour sélectionner l'image u_{la_n} la plus localement ambiguë à n classes :

$$u_{la_n} = \operatorname{argmax}_{u \in U} PLA_n(u) \quad (4.21)$$

Nous nous focaliserons plus particulièrement au cas $n = 2$ afin de considérer les frontières locales entre 2 classes. Par exemple, dans le cas à 3 classes, nous pouvons considérer l'image la plus localement ambiguë de degré 2, c'est-à-dire entre 2 classes. Chaque image $u \in U$ est alors associée à sa propre mesure $PLA_2(u)$ correspondant au maximum de probabilité pignistique sur le sous-ensemble :

$$\Omega_{LA_2} = \{(H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (\overline{H_1}, H_2, H_3)\} \quad (4.22)$$

Ainsi, il est possible de disposer de deux cas extrêmes d'ambiguïté, la plus globale et la plus locale.

La figure 4.7 illustre le comportement attendu par cette stratégie sur les mêmes données de la figure 4.4. Dans cet exemple limité à 1 image par classe, les images les plus localement ambiguës sont à distance intermédiaire entre 2 classes deux à deux. Ces images peuvent être difficiles à classer pour l'utilisateur et elles permettent de se focaliser sur la désambiguïsation des contenus visuels entre 2 classes.

3.2. Stratégies basées sur l'analyse des distributions de masses

Ce type de stratégies exploite les outils propres au MCT pour retrouver des approches couramment employées dans les techniques d'apprentissage actif :

- une approche par incertitude maximum basée sur la mesure de non-spécificité issue du MCT,
- une approche rappelant les techniques d'apprentissage actif par comité de modèles basée sur une mesure de désaccord correspondant à la masse sur le conflit.



FIGURE 4.7 : Illustration de la sélection d'échantillons par une stratégie du plus localement ambigu entre 2 classes (MLA_2). Les images situées entre 2 membres de 2 classes distinctes sont sélectionnées en priorité.

L'état de connaissance modélisé par les distributions de masses est directement exploité pour établir ces stratégies, sans utiliser les distributions de probabilités pignistiques.

3.2.1. Stratégie du plus incertain ou "Most Uncertain" (MU)

Le principe du minimum d'information (PMI) est couramment utilisé dans le cadre du MCT lorsque l'on doit choisir une distribution de masses parmi d'autres. Le PMI impose de choisir celle qui est la moins engagée et renvoie ainsi à la notion de maximum d'entropie.

Un échantillon provoquant une incertitude élevée possède une distribution de masses telle qu'il est difficile d'avoir un maximum de probabilité pignistique significatif sur une hypothèse en particulier (la distribution de probabilités pignistiques tend vers le cas d'équiprobabilité). Cette incertitude peut être liée à une absence de connaissance, un manque de sources d'informations par exemple.

Pour mesurer cette incertitude, la mesure la plus répandue dans le cadre du MCT est la *non-spécificité* $N(m^\Omega)$ sur une distribution de masses m^Ω basée sur la cardinalité des propositions B pondérées par la valeur des masses [DP86] :

$$N(m^\Omega) = \sum_{\emptyset \neq B \subseteq \Omega} m(B) \log_2(|B|) \quad (4.23)$$

Dans notre cadre, considérant un ensemble de distributions de masses m_u^Ω définies dans le cadre de discernement Ω , l'idée est de faire étiqueter par l'utilisateur l'image u_{mu} qui est la moins informative, au sens de la moins engagée pour une des hypothèses du cadre de discernement Ω . Formellement, l'image

non étiquetée u_{mu} de U la plus incertaine est :

$$u_{mu} = MU(U) = \operatorname{argmax}_{u \in U} N(m_u^\Omega) \quad (4.24)$$

avec pour chaque image $u \in U$, avec $N(m_u^\Omega)$ la mesure de non-spécificité sur sa distribution de masses m_u^Ω .

Remarque : d'après la formalisation de la connaissance effectuée avec le MCT, il est important de noter que, en général, l'échantillon le plus incertain ne correspond pas à celui le plus globalement ambiguë (sauf dans le cas particulier de classification bi-classe). Si un échantillon possède une probabilité pignistique élevée sur l'hypothèse d'ambiguïté globale, cela indique avec certitude qu'il est ambigu. Par contre, dans le cas d'une incertitude élevée, la proposition de doute global $(\Omega_1, \Omega_2, \dots, \Omega_Q)$ dans une distribution de masses absorbe une grande partie de la masse totale. En conséquence, si l'on observe l'expression de la transformée pignistique, cette situation tend vers le cas d'équiprobabilité, car toutes les hypothèses du cadre du discernement Ω font intervenir la proposition de doute global dans le calcul de leur probabilité pignistique.

Exemple à 3 classes : L'expression de la probabilité pignistique sur l'hypothèse d'ambiguïté globale est donnée par :

$$\begin{aligned} \operatorname{Bet}P\{m_u^\Omega\}(H_1, H_2, H_3) = & \frac{1}{1 - m_u^\Omega(\emptyset)} \left(m_u^\Omega(H_1, H_2, H_3) \right. \\ & + \frac{m_u^\Omega(\Omega_1, H_2, H_3)}{2} + \frac{m_u^\Omega(H_1, \Omega_2, H_3)}{2} + \frac{m_u^\Omega(H_1, H_2, \Omega_3)}{2} \\ & + \frac{m_u^\Omega(\Omega_1, \Omega_2, H_3)}{4} + \frac{m_u^\Omega(\Omega_1, H_2, \Omega_3)}{4} + \frac{m_u^\Omega(H_1, \Omega_2, \Omega_3)}{4} \\ & \left. + \frac{m_u^\Omega(\Omega_1, \Omega_2, \Omega_3)}{8} \right) \end{aligned} \quad (4.25)$$

Si dans la distribution de masses m_u^Ω de l'image u , la masse est concentrée sur la proposition de doute global $(\Omega_1, \Omega_2, \Omega_3)$, la probabilité pignistique de l'hypothèse (H_1, H_2, H_3) tend vers $1/8$ (et de même pour chacune des 7 autres hypothèses du cadre de discernement Ω).

La figure 4.8 illustre le comportement produit par cette stratégie sur les mêmes données de la figure 4.4. Le critère d'incertitude est peut être moins intuitif pour l'utilisateur, car il est difficile d'appréhender quelles images sont *a priori* les plus incertaines. Nous pouvons néanmoins noter dans cet exemple que l'image identifiée précédemment comme étant la plus rejetée (numérotée "8") est clairement engagée vers cette hypothèse de rejet avec certitude.

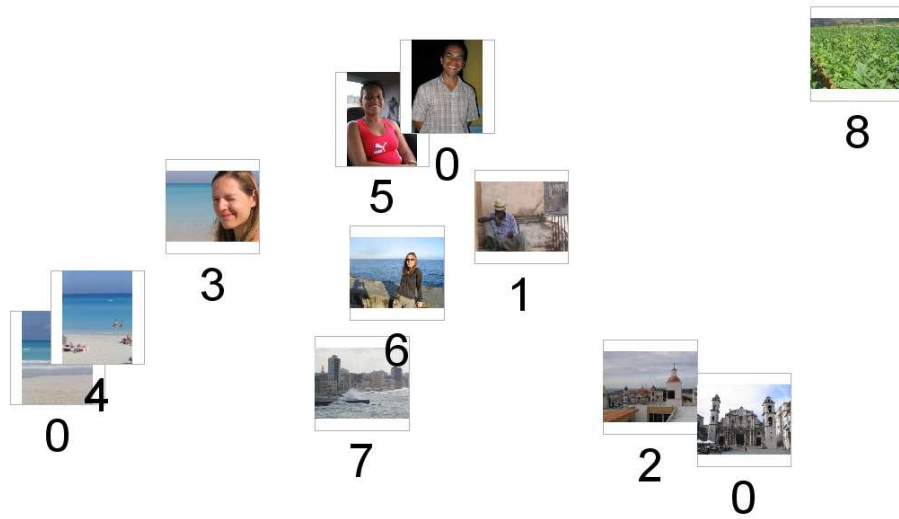


FIGURE 4.8 : Illustration de la sélection d'échantillons par une stratégie du plus incertain avec la mesure de non-spécificité. Nous pouvons remarquer que l'image numérotée 8 est la plus engagée.

3.2.2. Stratégie du plus en conflit ou "Most Conflicted" (MC)

L'incertitude peut être également liée à une situation de connaissance trop informative, dans le sens où toutes les sources d'information fusionnées ne vont pas dans le même sens. Dans notre cadre, la fusion des informations apportées par des descripteurs visuels de nature distinctes (couleur, orientation) peuvent, pour certaines images, entraîner du conflit concernant l'hypothèse la plus probable au sens de la distribution de probabilité pignistique.

La stratégie du plus en conflit consiste alors à étiqueter en priorité les images pour lesquelles les descripteurs sont le plus en désaccord. Cette information de conflit est directement disponible dans la distribution de masses m_u^Ω d'une image u .

Formellement, l'image non étiquetée u_{mc} la plus en conflit est celle qui possède une masse sur la proposition de conflit $m_u^\Omega(\emptyset)$ la plus élevée :

$$u_{mc} = \operatorname{argmax}_{u \in U} c(u) \quad (4.26)$$

avec $c(u)$ la mesure de conflit :

$$c(u) = m_u^\Omega(\emptyset) \quad (4.27)$$

La figure 4.9 tente d'illustrer comment le conflit se manifeste à la première itération de cette stratégie (toujours sur le même exemple de la figure 4.4). Les deux illustrations du haut représentent, pour la couleur d'une part et pour les orientations d'autre part, et pour chaque image non étiquetée, quelle type d'hypothèse du cadre de discernement Ω possède le maximum de probabilité pignistique : R pour l'hypothèse de rejet, A2 pour une hypothèse du type ambiguïté locale à 2 classes, A3 pour l'hypothèse

d'ambiguïté globale et P une hypothèse de type positif. En comparant les deux illustrations, on ne retrouve pas le même type d'hypothèse dans les deux cas. Par exemple, l'image en haut à droite, de couleur dominante verte est indiquée comme étant rejetée selon les informations de couleur. Or elle est indiquée comme étant globalement ambiguë selon l'information orientation. En conséquence, après fusion des 2 descripteurs, l'image est fortement en conflit comme l'indique la figure du bas.

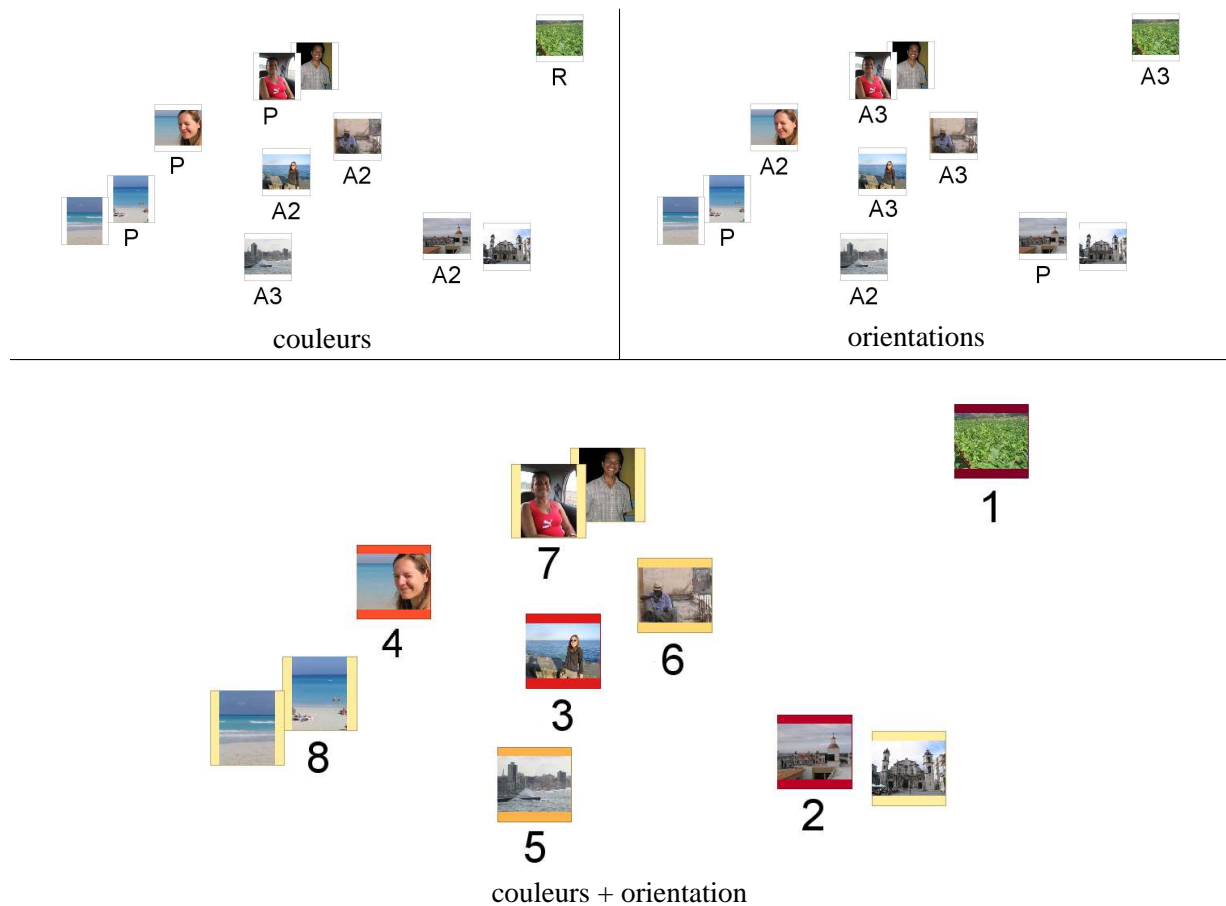


FIGURE 4.9 : Illustration de la sélection d'échantillons par une stratégie du plus en conflit selon la fusion de descripteurs visuels distincts. Les figures du haut "couleur" et "orientations" indiquent pour chaque image non étiquetée le type d'hypothèse donnant le maximum de probabilités pignistiques : P pour une hypothèse positive, R pour l'hypothèse de rejet, et A₃ pour l'hypothèse d'ambiguïté globale et A₂ pour les hypothèses d'ambiguïtés locales. Dans la figure, après fusion des informations apportées par les descripteurs, les images sont numérotées par ordre croissant, de la plus à la moins en conflit. Les images les plus en conflit peuvent être des images qui étaient pourtant identifiées par les autres stratégies comme étant indifféremment la plus positive ou la plus rejetée, ou la plus ambiguë.

3.3. Bilan sur les stratégies orientées "image"

Les stratégies "orientées images" utilisent le mécanisme classique de l'apprentissage actif en sélectionnant de manière itérative une seule image pour être étiquetée par l'utilisateur. L'utilisateur dispose de plusieurs stratégies qui peuvent l'aider à sélectionner un certain type d'images à instant t . Cette panoplie

de stratégies peut l'aider de différentes manières pour bien cerner la composition visuelle des classes : soit en confirmant les aspects visuels des classes avec les images les plus positives, soit en désambiguïsant des contenus proches entre plusieurs classes, ou encore en explorant des nouveaux aspects visuels avec les plus rejetés, ...

Ces stratégies "orientées images" amènent à réviser la connaissance, à chaque fois qu'une image est traitée par l'utilisateur. Le coût de calcul global, pour traiter une collection d'images dans son intégralité, est important puisque qu'à chaque itération, les distributions de masses de toutes les images non étiquetées restantes sont recalculées, et ceci pour finalement n'étiqueter qu'une seule image. Du point de vue de l'implémentation, ce coût peut être optimisé en ne calculant par exemple que les masses strictement nécessaires aux probabilités pignistiques des hypothèses utilisées par la stratégie en cours.

Cependant, pour certaines stratégies, il n'est pas garanti que la connaissance soit tout le temps fortement modifiée entre 2 itérations. Par exemple, dans le cas de la stratégie du plus positif, si l'image sélectionnée est classée par l'utilisateur dans la classe indiquée par l'hypothèse positive, l'information apportée l'étiquetage de cette image est minime. Il est probable que l'image qui était la deuxième plus positive à cette itération devienne l'image la plus positive à la prochaine.

De plus, les stratégies orientées "image" ne permettent pas de cibler précisément une classe d'intérêt pour l'utilisateur. Par exemple, dans le cas de la stratégie des plus positifs, les images sélectionnées peuvent être successivement positives pour une classe, puis pour une autre... Il est possible que les contenus visuels soient très différents d'une image à l'autre puisqu'elles peuvent être probables pour différentes classes, ce qui peut entraîner une certaine fatigue pour l'utilisateur.

Partant de ces observations, nous souhaitons proposer une stratégie alternative à l'utilisateur, dans un souci d'économie de calculs et de gain de productivité en permettant l'étiquetage simultané de lots d'images tout en ciblant les classes.

4. Stratégie orientée "classe"

L'idée générale de la stratégie orientée "classe" est de pouvoir faire étiqueter par l'utilisateur plusieurs images en une seule itération. Chaque classe est alors associée à une liste contenant les images non étiquetées triées de la plus à la moins probable. L'utilisateur peut alors se concentrer sur une liste en particulier, en focalisant son attention sur les contenus visuels d'une seule classe.

La stratégie orientée "classe" s'appuie sur un bilan complet de la connaissance. Un ensemble L de listes d'images non étiquetées est défini à partir du cadre de discernement Ω , où une liste est associée à chacune des hypothèses $\omega \in \Omega$. Chaque image non étiquetée u est alors insérée dans la liste correspondant à l'hypothèse ω_u possédant le maximum de probabilité pignistique :

$$\omega_u = \operatorname{argmax}_{\omega \in \Omega} \operatorname{Bet}P\{m^\Omega\}(\omega) \quad (4.28)$$

L'ensemble des images non étiquetées U est au final éclaté et réparti sur les différentes listes.

Exemple : cas à 3 classes

Le cadre de discernement $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ permet de définir un ensemble de 8 listes :

- l_0 la liste des images possédant le maximum de probabilité pignistique sur l'hypothèse de rejet $(\overline{H_1}, \overline{H_2}, \overline{H_3})$,
- l_1 la liste des images possédant le maximum de probabilité pignistique sur l'hypothèse positive $(H_1, \overline{H_2}, \overline{H_3})$,
- de même, l_2 la liste pour l'hypothèse positive $(\overline{H_1}, H_2, \overline{H_3})$,
- l_3 pour l'hypothèse positive $(\overline{H_1}, \overline{H_2}, H_3)$,
- $l_{1,2}$ pour l'hypothèse d'ambiguïté locale $(H_1, H_2, \overline{H_3})$,
- $l_{1,3}$ pour l'hypothèse d'ambiguïté locale $(H_1, \overline{H_2}, H_3)$,
- $l_{2,3}$ pour l'hypothèse d'ambiguïté locale $(\overline{H_1}, H_2, H_3)$,
- et $l_{1,2,3}$ pour l'hypothèse d'ambiguïté globale (H_1, H_2, H_3) ,

Chaque liste est ordonnée par le maximum de probabilité pignistique des images qu'elle contient. L'utilisateur peut ainsi avoir un aperçu global de la connaissance. Il peut par exemple se concentrer sur une liste d'images positives pour une classe C_q en particulier. Les premières images étant les plus probables, l'utilisateur peut décider de les classer ensemble par petits lots de quelques images. Au fur et à mesure, les images sont de moins en moins probables et l'utilisateur est libre d'arrêter de classer les images si elles lui paraissent inadéquates. Quand l'utilisateur n'est plus satisfait de la composition des listes, il peut demander au système de réévaluer la connaissance en fonction des nouvelles images qu'il vient de classer.

L'utilisateur peut également choisir de se focaliser sur une liste d'images localement ambiguës en ciblant précisément les 2 classes. Ainsi, l'utilisateur peut désambiguïser des classes distinctes possédant pourtant des contenus visuels similaires, en se focalisant sur des images proches de la frontière de ces classes. Enfin, l'utilisateur a accès à toutes des images visuellement très différentes des images précédemment étiquetées, grâce à la liste des images rejetées l_0 . Il peut alors y sélectionner des images aux contenus visuels qui l'intéressent pour compléter les différents aspects visuels des classes existantes, ou bien créer des nouvelles classes.

Exemple : cas à 3 classes

7 images non étiquetées possèdent les distributions de probabilités pignistiques suivantes :

Im. / Hyp.	(H_1, H_2, H_3)	$(H_1, H_2, \overline{H_3})$	$(H_1, \overline{H_2}, H_3)$	$(\overline{H_1}, H_2, H_3)$	$(H_1, \overline{H_2}, \overline{H_3})$	$(\overline{H_1}, H_2, \overline{H_3})$	$(\overline{H_1}, \overline{H_2}, H_3)$	$(\overline{H_1}, \overline{H_2}, \overline{H_3})$
u_1	0,01	0,05	0,04	0,1	0,6	0,1	0,05	0,05
u_2	0,7	0,05	0,05	0,1	0,0	0,0	0,1	0,0
u_3	0,0	0,03	0,07	0,03	0,04	0,02	0,01	0,8
u_4	0,01	0,05	0,04	0,3	0,7	0	0	0
u_5	0,0	0,0	0,0	0,1	0,0	0,8	0,1	0
u_6	0,0	0,1	0,0	0,0	0,4	0,2	0,2	0,1
u_7	0,0	0,8	0,01	0,09	0,05	0,04	0,01	0,0

Les images sont séparées dans les listes suivantes :

- u_2 dans la liste $l_{1,2,3}$
- u_7 dans la liste $l_{1,2}$
- aucune dans la liste $l_{1,3}$
- aucune dans la liste $l_{2,3}$
- u_4 puis u_1 puis u_6 dans la liste l_1
- u_5 dans la liste l_2
- aucune dans la liste l_3
- u_3 dans la liste l_0

L'utilisateur dispose ainsi d'un bilan de la connaissance sur ces 7 images. Il peut se concentrer sur une des listes, par exemple la liste l_1 concernant l'hypothèse $(H_1, \overline{H_2}, \overline{H_3})$. Il est probable que l'utilisateur décide de classer les images $u_4, u_1, et u_6$ ensemble dans la classe C_1 correspondante.

Par rapport aux stratégies orientées "images", on a déplacé le coût de calcul :

- Une stratégie orientée "image" demande de faire des calculs intensifs sur toutes les images non étiquetées, pour 1 seule image étiquetée par l'utilisateur à chaque itération. Ce coût est optimisable car la plupart des stratégies ne nécessitent pas de calculer l'intégralité des distributions de masses et des distributions de probabilités pignistiques.
- La stratégie orientée "classes" est beaucoup moins intensive car l'utilisateur peut étiqueter plusieurs images en une itération. Par contre, elle nécessite de calculer l'intégralité des distributions de masses et des distributions de probabilités pignistiques, ce qui peut représenter un coût de calcul plus élevé que celui d'une itération d'une stratégie orientée "image".

Scénario d'utilisation des 2 types de stratégies :

Il est recommandé à l'utilisateur d'utiliser la stratégie orientée "classe" ponctuellement, en interrompant de temps en temps une stratégie orientée "image" en cours. En particulier, lorsque trop peu d'images ont été étiquetées, la connaissance des classes est trop partielle. Il est préférable d'utiliser une stratégie de type "image" : l'utilisateur peut par exemple commencer avec une stratégie du plus rejeté pour identifier de nouveaux contenus visuels, puis confirmer les aspects visuels des classes avec la stratégie du plus positif. Quand l'utilisateur se représente mieux les contenus des classes, il peut faire appel à la stratégie orientée "classe" pour étiqueter plus rapidement par lots les images candidates aux différentes classes.

5. Mise à jour de la connaissance après étiquetage des images

5.1. Définition de zones de connaissances

A chaque itération du processus d'apprentissage actif, une ou plusieurs images sélectionnées sont étiquetées par l'utilisateur. Ces images ajoutent de l'information sur la constitution visuelle des classes qui doit être prise en compte pour réévaluer la connaissance. En effet, la modélisation de la connaissance se base sur une approche de type k plus proches voisins formalisée avec le Modèle des Croyances Transférables.

Or, ces images étiquetées peuvent devenir à leur tour des voisins des images restantes non étiquetées et modifier ainsi les fonctions de croyance. Nous précisons dans cette partie comment les fonctions de croyances sont modifiées en fonction des nouveaux étiquetages.

Dans le module de modélisation et synthèse de la connaissance, la toute première étape consiste à convertir une distance $d(u, l_q^i)$ dans un espace de description entre une image non étiquetée u et d'une image membre l_q^i d'une classe C_q , en des fonctions de croyance. Une classe C_q est associée à son propre cadre de discernement local $\Omega_q = \{H_q, \overline{H}_q\}$, et la définition des fonctions de masses repose sur les expressions :

$$m_u^{\Omega_q}(H_q) = \alpha_0 \cdot e^{-\left(\frac{d(u, l_q^i)}{\sigma_q}\right)^\beta} \quad (4.29)$$

$$m_u^{\Omega_q}(\Omega_q) = 1 - m_u^{\Omega_q}(H_q) \quad (4.30)$$

où la masse $m_u^{\Omega_q}(H_q)$ représente la croyance sur la proposition H_q "l'image non étiquetée u appartient à la classe C_q ", et $m_u^{\Omega_q}(\Omega_q)$ la masse sur le doute.

Les paramètres α_0 , β et σ_q permettent de contrôler une zone de connaissance autour de l'échantillon étiqueté l_q^i dans l'espace de description considéré. Les paramètres α_0 et β sont fixés arbitrairement et sont les mêmes pour toute image étiquetée ($\alpha_0 = 0,9$ et $\beta = 2$). Le paramètre σ_q permet alors de contrôler la zone de connaissance autour d'une image membre d'une classe C_q . Nous pouvons voir l'influence de ce paramètre σ_q en prenant 2 cas extrêmes :

σ_q tend vers 0 :

Si σ_q est faible, la masse $m_u^{\Omega_q}(H_q)$ tend vers un Dirac. Toute image non étiquetée u comparée avec un voisin membre d'une classe C_q est alors associée à une distribution de masses n'ayant que de la masse sur le doute $m_u^{\Omega_q}(\Omega_q) = 1$. Après combinaison des k plus proches voisins, la distribution de masses résultant contient également uniquement de la masse sur le doute. L'opération de transfert de masses (voir chapitre 3 page 53), répartit de manière égale la masse sur les propositions \overline{H}_q et Ω_q :

$$\begin{aligned} m_u^{\Omega_q}(\Omega_q) &= 0,5 \\ m_u^{\Omega_q}(\overline{H}_q) &= 0,5 \end{aligned} \quad (4.31)$$

En conséquence, en considérant le niveau supérieur de fusion multi-classe (page 58), avec l'opération d'extension vide, la distribution de masses résultant n'a que de la masse sur des propositions combinant des propositions locales de doute Ω_q et des hypothèses négatives \overline{H}_q . Les distributions de probabilités pignistiques sont alors complètement déséquilibrées. Par exemple, l'hypothèse d'ambiguïté globale ne se calcule qu'à partir de la proposition de doute globale, tandis que l'hypothèse de rejet a une probabilité pignistique nettement supérieure à toutes les autres hypothèses. En conséquence, les images non étiquetées possèdent toutes la même distribution de probabilités pignistiques et il ne sera pas possible de comparer les images non étiquetées pour en sélectionner une en particulier.

σ_q est très grand :

En suivant le même raisonnement, plus le paramètre σ_q est grand, plus la masse $m_u^{\Omega_q}(H_q)$ tend vers α_0 . Dans ce second cas, toute image u de l'espace est associée à une masse maximale sur la proposition H_q . Après les différentes étapes de fusions, toute image non étiquetée u aura la même distribution de probabilités pignistique déséquilibrée, où cette fois, l'hypothèse d'ambiguïté globale aura une probabilité pignistique nettement supérieure à toutes les autres.

Il est nécessaire d'estimer les paramètres σ_q pour définir les zones de connaissance propre à chaque image étiquetée dans chaque espace de description, en garantissant un compromis entre doute et croyance en la proposition H_q afin d'obtenir des distributions de masses et de probabilités pignistiques plus riches et nuancées entre elles.

Un paramètre σ_q pourrait être estimé par étiquette de classe C_q . Cette estimation est cohérente si la classe considérée est très homogène visuellement. Cependant, les collections d'images traitées dans notre système, les classes peuvent être très hétérogènes visuellement et peuvent potentiellement se recouvrir entre elles. Il faut donc avoir une adaptation locale des fonctions de croyance en fonction des étiquettes et des espaces de descriptions considérés pour chaque image étiquetées.

Nous proposons donc d'associer à chaque image membre l_q^i d'une classe C_q son propre paramètre $\sigma_{i,q}^s$, pour chaque espace de description s .

5.2. Méthode d'estimation

La fonction de masse sur la proposition H_q étant de forme exponentielle, toutes les zones de connaissances des classes se chevauchent à différents degrés selon les proximités entre les images étiquetées. Nous proposons de prendre en compte ce taux de "chevauchement" f entre ces classes pour estimer le paramètre $\sigma_{i,q}^s$ d'une image étiquetée l_q^i selon un espace de description s avec la méthode suivante :

- Pour chaque espace de description s
 - Pour chaque classe C_q
 - Pour chaque membre de la classe l_q^i
 1. Recherche des k' plus proches voisins "négatifs", les k' images étiquetées $\{l_r^0, l_r^1, \dots, l_r^{k'}\}$ avec $r \neq q$ les plus proches appartenant à des classes autres que C_q
 2. Estimation du paramètre $\sigma_{i,q}^s$ en fonction de la distance moyenne d' des k' négatifs avec l_q^i

L'estimation du paramètre $\sigma_{i,q}^s$ se base le taux de chevauchement f des zones de connaissances, la valeur de la masse $m_u^{\Omega_q}(H_q)$ à la distance intermédiaire moyenne $\frac{d'}{2}$ des k' négatifs :

$$d' = \frac{\sum_{j=1}^{k'} d(l_r^j, l_q^i)}{k'} \quad (4.32)$$

$$f = \alpha_0 \cdot e^{-\left(\frac{d'}{2\sigma_{i,q}^s}\right)^\beta} \quad (4.33)$$

L'estimation de $\sigma_{i,q}^s$ se calcule alors en inversant l'expression de la fonction de masse $m_u^{\Omega_q}(H_q)$ (pour $\beta = 2$) :

$$\sigma_{i,q}^s = \frac{d'}{2} \sqrt{\frac{1}{\ln \alpha_0 - \ln f}} \quad (4.34)$$

avec $f \in]0; \alpha_0[$

Ce nouveau paramètre f est global à toutes les images étiquetées, et contrôle le chevauchement des connaissances aux frontières des classes. Autrement dit, f gère le taux d'ambiguïté **aux frontières** des classes.

Le paramètre f contrôle en réalité la capacité de généralisation des modèles de classe C_q . En estimant un paramètre $\sigma_{i,q}^s$ en fonction des k' plus proches voisins négatifs de l'image étiquetée l_q^i , on espère maximiser la zone de connaissance couverte dans l'espace des descripteurs. Ainsi, on limite l'ambiguïté aux frontières "intérieures" entre les classes tout en tentant d'étendre au maximum la zone de connaissance aux frontières extérieures.

La figure 4.10 donne une illustration des paramètres de $\sigma_{i,q}^s$ sur une collection d'images réelles représentée dans un espace de descripteur couleur. Cette visualisation utilise une projection dans le plan basée sur une méthode de réduction de dimensions [Thi06] en s'appuyant sur les vecteurs descripteurs couleur dans l'espace $\{L,u,v\}$, c'est-à-dire que les images sont placées de telle manière à respecter le mieux possible les proximités réelles dans l'espace de description. La taille des images est corrélée avec leur propre paramètre $\sigma_{i,q}^s$, tout comme la forme des fonctions de croyance représentées par des "halos" colorés. La couleur des "halos" correspond aux étiquettes attribuées par un utilisateur sur ces images.

Nous pouvons voir que les classes ne sont pas visuellement homogènes : certaines images portant la même étiquette peuvent être très éloignées, comme pour certains membres de la classe des images étiquetées sous la couleur "verte". Nous pouvons voir surtout que certaines images sont isolées et encerclées d'images portant d'autres étiquettes. Elles ont tendance à posséder une zone de connaissance limitée autour d'elles car elles sont situées dans des zones d'ambiguïtés entre les classes. Si une nouvelle image non étiquetées se trouve dans cette portion de l'espace des descripteurs, la distribution de masses aura certainement beaucoup de masse sur des propositions contenant du doute.

A l'inverse, les images situées en périphérie sont plutôt éloignées de l'ensemble des images étiquetées, aux limites des contenus visuels connus. En conséquence, elles sont associées à de grandes zones de

connaissance comme pour les images du haut et du bord gauche. Ainsi, si une nouvelle image non étiquetée est décrite dans une portion de l'espace des descripteurs non connus, sa distribution de masses sera fortement influencée par les grandes zones de connaissance des images étiquetées périphériques.

Remarque : dans le cas particulier de l'étiquetage multiple, il est plus intéressant d'augmenter le paramètre f . Ainsi, l'ambiguïté aux frontières des classes est plus marquée et l'on favorise les probabilités pignistiques sur les hypothèses ambiguës, ce qui permet de pointer les images qui peuvent potentiellement porter plusieurs étiquettes.

6. Conclusion

Nous avons décrit dans ce chapitre l'intérêt de sélectionner des images non étiquetées en priorité pour aider un utilisateur dans sa tâche d'organisation d'une collection.

Nous nous sommes inspiré des méthodes d'apprentissage actif et nous avons exprimé dans le même cadre formel du Modèle des Croyances Transférables les stratégies de sélection les plus courantes. L'utilisateur dispose ainsi d'un module pouvant l'aider à "construire" les classes en choisissant la stratégie de sélection qui lui convient à un moment donné. Il peut par exemple demander des images représentant de nouveaux contenus visuels, ou bien demander des images très ressemblantes à une des classes. Il peut également demander des images ambiguës entre plusieurs, ou bien des images difficiles à classer parce que les témoignages issus de différents descripteurs divergent, ou bien encore des images difficiles sur lesquelles le système n'arrive pas à émettre un jugement en particulier.

Le prochain chapitre tente de proposer une aide supplémentaire à l'utilisateur en lui suggérant de manière automatique des classements sur les images sélectionnées à travers l'interface développée au cours de cette thèse.

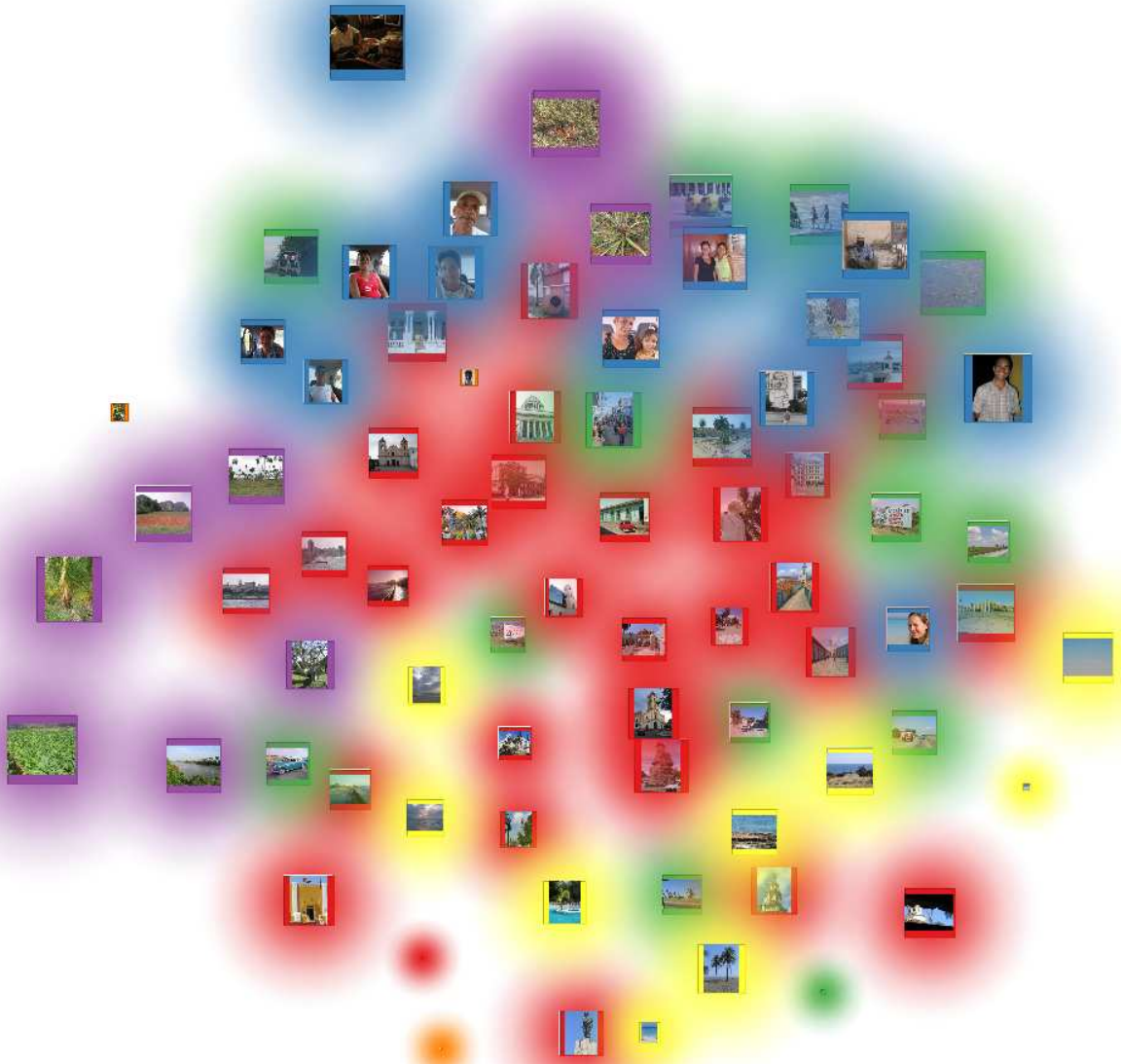


FIGURE 4.10 : Exemple d'adaptation locale des zones de connaissances autour des images étiquetées : cette figure représente une collection d'images étiquetées par 6 étiquettes. Les images sont placées dans l'espace 2D à partir d'une méthode de réduction de dimension s'appuyant sur les vecteurs descripteurs couleur dans l'espace $\{L,u,v\}$, c'est-à-dire que les images sont placées de telle manière à respecter le mieux possible les proximités réelles dans l'espace de description. La taille des images est corrélée avec leur propre paramètre $\sigma_{i,q}^s$, tout comme la forme des fonctions de croyance (les "halos" colorés). Chaque couleur d'un "halo" correspond à une étiquette.

INTERFACE POUR LA STRUCTURATION DE COLLECTION D'IMAGES

Description du contenu

1. Introduction	95
2. Propositions automatiques d'étiquettes	96
2.1. Schémas d'étiquetages	96
2.2. Etiquetage simple	98
2.2.1. Etiquetage simple sans rejet (S)	98
2.2.2. Etiquetage simple avec options de rejets	99
2.3. Etiquetage multiple	101
2.3.1. Etiquetage multiple sans rejet en distance (M)	101
2.3.2. Etiquetage multiple avec rejet en distance autorisé (M+D)	102
2.4. Limitation de l'espace de décision	102
3. Interface	104
3.1. Description globale de l'interface	104
3.1.1. La vue interactive	104
3.1.2. Le panneau de contrôle	105
3.2. Interactions avec l'utilisateur	105
3.2.1. Interactions dans la vue globale	105
3.2.2. Contrôle de l'animation	106
3.2.3. Type de classification	107
3.2.4. Stratégies	107
4. Conclusion	109

1. Introduction

Ce chapitre présente concrètement comment l'utilisateur peut manipuler une collection d'images pour l'organiser et la structurer à travers une interface (voir figure 5.1). La problématique de la conception d'une interface homme-machine est complexe et est un domaine de recherche à part entière. L'objectif est d'arriver à proposer une interface intuitive simple et épurée avec un minimum de contrôles.

Il est difficile de trouver des références bibliographiques répondant exactement à notre problème : si les interfaces sont omniprésentes et indispensables pour élaborer des produits finis dans l'industrie, c'est

une discipline relativement ouverte, nécessitant souvent de développer des composants graphiques spécifiques adaptés au fonctionnement du système réalisé.

Nous proposons d'aborder la conception de notre interface sous deux angles pour assister l'utilisateur :

1. La proposition automatique d'étiquettes : l'utilisateur peut être assisté pour étiqueter une image. L'idée est de suggérer automatiquement des étiquettes sur les images sélectionnées afin de l'accompagner dans sa réflexion sur des choix qui peuvent être parfois très difficiles. Par le fait que la connaissance se raffine progressivement à chaque nouvelle image étiquetée, la pertinence des étiquettes proposées s'améliora au fur et à mesure et l'utilisateur aura de moins en moins de corrections à faire.
2. La conception d'une représentation graphique interactive intelligible comme support de réflexion : le but est de rendre compte, visuellement à l'utilisateur, l'analyse automatique de la connaissance sur les classes et de voir explicitement les propositions de classements sur les images non étiquetées. En même temps, nous essayons de minimiser le nombre de manipulations que l'utilisateur doit effectuer pour organiser une collection d'images, notamment avec le concept de "validation passive" lié à l'utilisation d'animation.

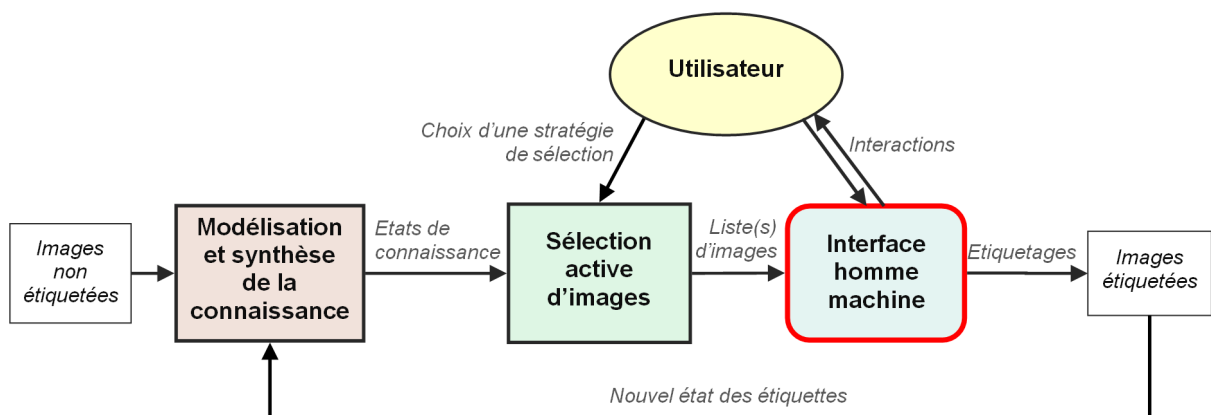


FIGURE 5.1 : Ce chapitre en cours se focalise sur le dernier module concernant les échanges entre l'utilisateur et le système à travers une interface homme-machine. Ce module s'appuie sur les fonctions de croyance afin de proposer des étiquettes sur des images sélectionnées par le module précédent. L'interface permet à l'utilisateur de contraindre ces propositions en autorisant ou non les cas de multi-étiquetage, et des cas de rejet en distance ou en ambiguïté. La vue principale donne un aperçu de l'état de connaissance des classes et est complètement interactive de telle manière que l'utilisateur peut définir et redéfinir les classes si besoin.

2. Propositions automatiques d'étiquettes

2.1. Schémas d'étiquetages

Cette partie présente la façon dont les étiquettes automatiques sont faites sur une image non étiquetée sélectionnée par une stratégie. Ces propositions s'appuient sur les probabilités pignistiques utilisées

précédemment dans le module de sélection active d'images.

Considérons une image u non étiquetée sélectionnée et présentée à l'utilisateur pour que ce dernier l'étiquette. L'interface homme-machine est alors le lieu du "pari", où se jouent les décisions à partir de la connaissance extraite sur toutes les images non étiquetées selon les différents descripteurs visuels et les classes disponibles.

Le cadre de discernement Ω est exhaustif car tous les étiquetages possibles que peut envisager un utilisateur sur une image non étiquetée u sont représentés. La difficulté est de déterminer automatiquement l'hypothèse qui convient le mieux à l'utilisateur. Cependant, un utilisateur peut vouloir organiser une collection d'images de différentes manières, ce qui n'engage pas alors toutes les hypothèses du cadre de discernement Ω . Par exemple, un utilisateur peut vouloir classer les images d'une collection de manière exclusive, c'est-à-dire qu'il n'envisage pas qu'une image puisse appartenir à plusieurs classes. Il est alors inutile de s'appuyer sur les hypothèses ambiguës pour lui proposer de classer l'image dans plusieurs classes à la fois.

L'utilisateur peut contrôler le type de propositions d'étiquettes que le système lui fait en choisissant parmi les 3 options :

- l'étiquetage simple ou multiple, selon que l'utilisateur souhaite classer les images de manière exclusive ou non,
- l'activation de rejet en distance si l'utilisateur veut mettre de côté des images trop différents des contenus des classes pour les traiter plus tard,
- l'activation de rejet en ambiguïté si l'utilisateur veut mettre de côté des images ambiguës entre plusieurs classes pour les traiter plus tard.

En choisissant ces options l'utilisateur contraint les propositions d'étiquettes et limite ainsi les propositions inadéquates. La combinaison de ces options induit 6 schémas de décisions distincts (voir tableau 5.1). Ces 6 schémas consistent alors à retenir uniquement les hypothèses du cadre de discernement nécessaires pour trouver l'étiquette la plus probables à proposer à l'utilisateur.

Schéma de décision	multi-étiquetage	rejet en ambiguïté	rejet en distance
S			
S+R			✓
S+A		✓	
S+R+A		✓	✓
M	✓		
M+R	✓	✓	✓

TABLE 5.1 : Dénominations des 6 schémas d'étiquetages. L'utilisateur peut choisir entre étiquetage simple (S) ou multi-étiquette (M) en activant éventuellement des options de rejet en ambiguïté (A) ou de rejet en distance (D).

Exemple : cas à 3 classes

Trois classes C_1 , C_2 , et C_3 sont représentées à travers le cadre de discernement Ω contenant 8 hypo-

thèses :

$$\begin{aligned}\Omega &= \Omega_1 \times \Omega_2 \times \Omega_3 \\ &= \{(H_1, H_2, H_3), (H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (H_1, \overline{H_2}, \overline{H_3}), \\ &\quad (\overline{H_1}, H_2, H_3), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3), (\overline{H_1}, \overline{H_2}, \overline{H_3})\}\end{aligned}$$

L'utilisateur doit choisir pour une image non étiquetée u , une des 8 étiquettes possibles :

- l'hypothèse de rejet $(\overline{H_1}, \overline{H_2}, \overline{H_3})$: l'utilisateur décide de mettre de côté cette image s'il ne trouve pas de classe adéquate pour cette image,
- une des hypothèses positives $(H_1, \overline{H_2}, \overline{H_3})$ ou $(\overline{H_1}, H_2, \overline{H_3})$ ou $(\overline{H_1}, \overline{H_2}, H_3)$ si l'utilisateur décide de classer l'image non étiquetée u de manière exclusive dans la classe (respectivement) C_1 ou C_2 ou C_3 ,
- une des hypothèses ambiguës $(H_1, H_2, \overline{H_3})$, $(H_1, \overline{H_2}, H_3)$ ou $(\overline{H_1}, H_2, H_3)$ ou (H_1, H_2, H_3) si l'utilisateur décide de classer l'image non étiquetée u de manière non exclusive dans les classes (respectivement) C_1 et C_2 , ou C_1 et C_3 , ou C_2 et C_3 , ou enfin C_1 et C_2 et C_3 .

Le but est alors de proposer automatiquement à l'utilisateur la meilleure hypothèse parmi les 8 disponibles.

2.2. Etiquetage simple

L'utilisateur peut choisir en premier lieu si le système doit lui faire des propositions de classement de manière exclusive ou non. Si l'utilisateur choisit l'étiquetage simple, il impose au système de faire des propositions de classement d'image de manière exclusive, c'est-à-dire qu'une image non étiquetée ne peut appartenir qu'à une seule classe à la fois.

2.2.1. Etiquetage simple sans rejet (S)

Faire des propositions d'étiquettes simple sans autoriser le rejet correspond au schéma de classification le plus basique où l'utilisateur souhaite classer les images de manière exclusive en supposant que toutes les classes nécessaires pour organiser sa collection d'images ont été identifiées. Cela correspond donc à un modèle de classification avec des *a priori* contraignant le nombre de propositions possibles.

Dans le cadre de la formalisation de notre problème avec le MCT, un espace de décision Ω_S est défini à partir du cadre de discernement Ω en retenant uniquement les hypothèses positives. L'hypothèse ω_s positive possédant la plus grande probabilité pignistique indique alors quelle classe proposer à l'utilisateur pour étiqueter une image u :

$$\omega_s = \operatorname{argmax}_{\omega_i \in \Omega_S} \operatorname{BetP}\{m^{\Omega_S}\}(\omega_i) \quad (5.1)$$

avec Ω_S contenant exclusivement les hypothèses positives du cadre de discernement Ω .

Exemple : cas à 3 classes Le cadre de discernement $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ permet de définir l'espace de décision Ω_S , le sous-ensemble de Ω contenant uniquement les hypothèses positives :

$$\Omega_S = \{(H_1, \overline{H_2}, \overline{H_3}), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3)\}$$

Cependant, la sélection d'une image et la proposition d'une étiquette s'effectuent de manière indépendante. Par exemple, une image peut être sélectionnée par une stratégie du plus rejeté. Il est alors probable que, en l'état de la connaissance calculée, cette image ne soit recommandée pour aucune des classes. Mais l'utilisateur peut vouloir forcer le système de proposer malgré tout une étiquette. Ainsi, le système est capable de lui indiquer à la fois que l'image n'est pas facilement associable à une des classes, tout en proposant malgré tout une étiquette.

2.2.2. Etiquetage simple avec options de rejets

L'utilisateur peut également autoriser des options de rejets en distance ou-et en ambiguïté. A un moment donné, en fonction de la connaissance des classes, il est possible que certaines images non étiquetées soient difficiles à classer, si elles ne ressemblent pas au contenu des classes, ou si elles sont trop proches de plusieurs classes à la fois. Le système doit être capable alors d'indiquer qu'il ne possède pas toutes les informations nécessaires pour faire une proposition de classement satisfaisant l'utilisateur.

Etiquetage simple avec option de rejet en distance (S+D)

Ce paradigme de fonctionnement correspond à ce qui est désigné souvent dans le cadre du MCT comme l'hypothèse de **monde fermé** ou de **monde ouvert**. Concrètement, si l'utilisateur choisit de fonctionner en monde fermé, c'est-à-dire sans rejet, il considère que toutes les classes définies sont nécessaires et suffisantes pour classer toutes les images. A l'inverse, si l'utilisateur autorise le rejet, en distance il suppose une hypothèse de monde ouvert, c'est-à-dire, qu'il autorise le système à lui indiquer éventuellement qu'une image ne peut pas être classée dans une des classes connues. Cela peut indiquer notamment à l'utilisateur que cette image rejetée peut être utilisée pour définir une nouvelle classe, ou un nouvel aspect visuel d'une classe existante. Dans ce cas, si la proposition est validée par l'utilisateur, l'image est mise de côté afin d'être traitée plus tard.

L'hypothèse de rejet $\omega_r = (\overline{H_1}, \overline{H_2}, \dots, \overline{H_Q})$ est alors ajoutée aux hypothèses positives dans l'espace de décision Ω_{S+R} . La meilleure proposition d'étiquette pour une image u est soumise automatiquement à l'utilisateur en retenant l'hypothèse ω_{s+r} possédant la plus grande probabilité pignistique :

$$\omega_{s+r} = \operatorname{argmax}_{\omega_i \in \Omega_{S+R}} \operatorname{Bet}P\{m^{\Omega_{S+R}}\}(\omega_i) \quad (5.2)$$

avec Ω_{S+R} contenant les hypothèses positives et l'hypothèse de rejet du cadre de discernement Ω . Si l'hypothèse ω_{s+r} correspond à l'hypothèse de rejet ω_r , l'image est mise de côté.

Exemple : cas à 3 classes

L'espace de décision Ω_{S+R} est défini par :

$$\Omega_{S+R} = \{(H_1, \overline{H_2}, \overline{H_3}), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3), (\overline{H_1}, \overline{H_2}, \overline{H_3})\}$$

Étiquetage simple avec rejet en ambiguïté autorisé (S+A)

L'utilisateur peut considérer que toutes les classes ont été identifiées. Cependant, certaines images non étiquetées peuvent avoir des contenus visuels proches aux images appartenant à différentes classes. Dans ce cas, si l'option de rejet en ambiguïté n'est pas activée, le risque est de forcer une étiquette pointant une classe qui ne sera pas retenue par l'utilisateur. L'option de rejet en ambiguïté permet alors de mettre de côté les images trop ambiguës, et d'attendre que la connaissance sur les classes soit plus précise pour les traiter plus tard.

Dans ce cas, un espace de décision ω_{s+a} est défini à partir du cadre de discernement Ω comme le sous-ensemble contenant les hypothèses positives et ambiguës. La meilleure proposition d'étiquette pour une image u est soumise à l'utilisateur automatiquement en retenant l'hypothèse ω_{s+a} possédant la plus grande probabilité pignistique :

$$\omega_{s+a} = \operatorname{argmax}_{\omega_i \in \Omega_{S+A}} \text{BetP}\{m^{\Omega_{S+A}}\}(\omega_i) \quad (5.3)$$

avec Ω_{S+A} contenant les hypothèses positives et les hypothèses ambiguës du cadre de discernement Ω . Si l'hypothèse ω_{s+a} correspond à l'une des hypothèses d'ambiguïté, l'image est mise de côté. Sinon, l'hypothèse ω_{s+a} est positive et indique la classe à conseiller à l'utilisateur.

Exemple : cas à 3 classes

L'espace de décision Ω_{S+A} est défini par :

$$\Omega_{S+A} = \{(H_1, \overline{H_2}, \overline{H_3}), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3), \\ (H_1, H_2, H_3), (H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (\overline{H_1}, H_2, H_3)\}$$

Étiquetage simple avec rejets en distance et en ambiguïté autorisés (S+D+A)

Dans ce dernier cas, l'utilisateur souhaite classer les images non étiquetées de manière non exclusive, tout en laissant le système lui indiquer si une image n'est pas classable dans une des classes disponibles ou si elle est trop ambiguë.

Dans ce cas, l'espace de décision à considérer est le cadre de discernement Ω dans son intégralité et l'hypothèse ω_{s+a+r} possédant la plus grande probabilité pignistique indique quel traitement effectuer

sur l'image non étiquetée en jeu :

$$\omega_{s+r+a} = \operatorname{argmax}_{\omega_i \in \Omega} \operatorname{Bet}P\{m^\Omega\}(\omega_i) \quad (5.4)$$

Si l'hypothèse ω_{s+r+a} est une hypothèse positive, le système fait une proposition de classe (celle indiquée par l'hypothèse positive). Par contre si l'hypothèse ω_{s+r+a} est l'hypothèse de rejet ou une des hypothèses ambiguës, l'image est mise de côté.

2.3. Etiquetage multiple

L'étiquetage multiple peut être pertinent pour certaines collections d'images (voir chapitre 2). En effet, un utilisateur peut vouloir définir les classes d'une collection par différents aspects sémantiques des contenus visuels. Par exemple, il peut avoir défini une classe "personne" et une classe "paysage de nature". Dans ce cas, l'utilisateur peut apprécier que le système soit capable de distinguer par exemple les images contenant uniquement un portrait, de celles contenant uniquement un paysage, et de celles contenant à la fois un portrait sur un fond de paysage nature.

Autoriser l'étiquetage multiple n'impose pas de classer absolument toutes les images dans plusieurs classes : le challenge est plutôt de distinguer les images qui peuvent potentiellement être classées dans plusieurs classes de celles qui sont plutôt aptes à être classées dans une seule.

Nous pouvons reprendre le même principe de rejet que dans le cas de la classification de manière exclusive, toutefois, en ne prenant pas en compte le rejet en ambiguïté, puisque cette fois-ci, une hypothèse ambiguë du cadre de discernement Ω permet d'indiquer à l'utilisateur les classes potentielles pour y classer une image.

2.3.1. Etiquetage multiple sans rejet en distance (M)

Dans le cadre du MCT, un espace de décision Ω_M est défini en retenant les hypothèses positives et ambiguës du cadre de discernement Ω . Ces hypothèses ω^p sont exprimées par l'expression générique :

$$\omega^p = (H_{p_1}, H_{p_2}, \dots, H_{p_P}, \overline{H_{n_1}}, \overline{H_{n_2}}, \dots, \overline{H_{n_N}}) \quad (5.5)$$

avec $1 < p < Q$, Q étant le nombre de classes définies par l'utilisateur, p_1, p_2, \dots, p_P les identifiants des classes indiquées comme étant pertinentes, et, avec n_1, n_2, \dots, n_N les identifiants des classes indiquées comme n'étant pas pertinentes.

L'espace de décision Ω_M correspond donc au cadre de discernement Ω auquel on a soustrait l'hypothèse de rejet $\omega_r = (\overline{H_1}, \overline{H_2}, \dots, \overline{H_Q})$.

La proposition d'étiquette soumise à l'utilisateur est effectuée en retenant l'hypothèse ω_m possédant la

plus grande probabilité pignistique parmi ces hypothèses :

$$\omega_m = \operatorname{argmax}_{\omega_i \in \Omega_M} \operatorname{Bet}P\{m^{\Omega_M}\}(\omega_i) \quad (5.6)$$

L'hypothèse ω_m peut donc pointer précisément sur la ou les classes les plus pertinentes pour une image non étiquetée u .

Exemple : cas à 3 classes

L'espace de décision Ω_M est défini par :

$$\Omega_M = \{(H_1, H_2, H_3), (H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (H_1, \overline{H_2}, \overline{H_3}), (\overline{H_1}, H_2, H_3), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3)\} \quad (5.7)$$

2.3.2. Etiquetage multiple avec rejet en distance autorisé (M+D)

Le dernier schéma d'étiquetage que peut désirer l'utilisateur exploite l'intégralité du cadre de discernement Ω . L'utilisateur souhaite que le système soit capable de distinguer les propositions d'étiquetages simples, d'étiquetages multiples tout en considérant l'éventualité de rejet.

Considérant une image u , la proposition d'étiquette automatique soumise à l'utilisateur est effectuée en retenant l'hypothèse ω_{m+r} possédant la plus grande probabilité pignistique du cadre du discernement :

$$\omega_{m+r} = \operatorname{argmax}_{\omega_i \in \Omega} \operatorname{Bet}P\{m^{\Omega}\}(\omega_i) \quad (5.8)$$

2.4. Limitation de l'espace de décision

Les schémas de décision n'impliquent pas tous le même nombre d'hypothèses. Plus un espace de décision contient d'hypothèses, plus le nombre de possibilités d'étiquetages est important, et en conséquence, plus il est difficile de trouver automatiquement la meilleure hypothèse satisfaisant l'utilisateur. De même, plus le nombre de classes Q est important, plus l'espace de discernement Ω contient d'hypothèses. En conséquence, l'espace de décision contient également plus d'hypothèses et il est alors difficile de trouver automatiquement la meilleure proposition d'étiquette (voir tableau 5.2).

Notamment, dans le cas de l'étiquetage multiple, le nombre de possibilités de classement augmente de manière exponentielle avec le nombre de classes considérées. Cependant, il est raisonnable de se demander si une image non étiquetée u peut potentiellement être classée dans un grand nombre de classes simultanément. En effet, dans le cadre de la classification d'images, nous pouvons supposer qu'un utilisateur envisagera rarement, voir jamais, de classer une image dans plus de 3 classes.

Cette hypothèse réaliste permet de limiter fortement les espaces de décision en éliminant des espaces de décisions les hypothèses ambiguës possédant un degré d'ambiguïté supérieur à 3. Pour avoir un ordre de grandeur, la figure 5.2 compare le nombre d'hypothèses considérées lors de la détermination d'une proposition d'étiquette en fonction du nombre de classes.

Remarque : cette limitation des espaces de décisions en adéquation avec les besoins de l'utilisateur permet d'alléger fortement le coût de calcul des distributions de probabilités pignistiques.

Schéma d'étiquetage	Nombre d'étiquettes possibles
Etiquetage simple sans rejet (S)	Q
Etiquetage simple avec rejet en distance autorisé (S+R)	$Q + 1$
Etiquetage simple avec rejet en ambiguïté (S+A)	$2^Q - 1$
Etiquetage simple avec les 2 rejets autorisés (S+D+A)	2^Q
Etiquetage multiple sans rejet (M)	$2^Q - 1$
Etiquetage multiple avec rejet en distance autorisé (M+R)	2^Q

TABLE 5.2 : Nombre de classement possibles pour une image non étiquetée u , en fonction du nombre de classes Q disponibles pour les 6 schémas de décision. Plus le nombre de classes est élevé, et plus le risque de mauvaise proposition automatique d'étiquette est élevé. L'autorisation des rejets augmente le risque de proposition d'étiquetage non pertinente pour l'utilisateur.

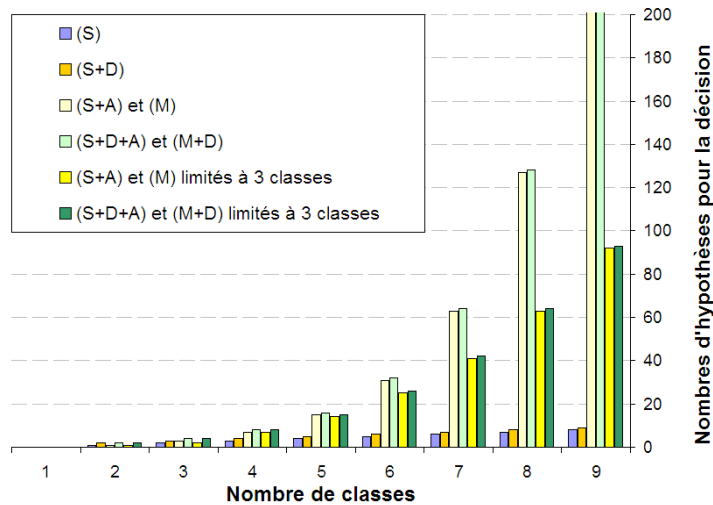


FIGURE 5.2 : Comparaison du nombre d'hypothèses considérées pour choisir une proposition d'étiquetage d'une image selon les 6 schémas de décision. La limitation à 3 classes ambiguës contraint l'explosion combinatoire du nombre d'hypothèses de décision.

A cette étape le système a sélectionné une image suivant une stratégie choisie par l'utilisateur, et il a déterminé l'hypothèse de Ω la plus probable suivant les contraintes choisies par l'utilisateur. Il s'agit maintenant de proposer une interface permettant à l'utilisateur de valider les propositions d'étiquettes faites automatiquement.

3. Interface

Le but de cette partie est de décrire brièvement l'interface et les différentes vues qui ont été développées au cours de cette thèse, afin de faciliter les manipulations de l'utilisateur sur une collection d'images.

3.1. Description globale de l'interface

La figure 5.3 présente l'interface contenant 2 zones :

- à droite, une vue interactive représentant les images et les classes que l'utilisateur peut manipuler,
- à gauche, un panneau permettant à l'utilisateur de contrôler différents paramètres sur le fonctionnement du cœur du système.

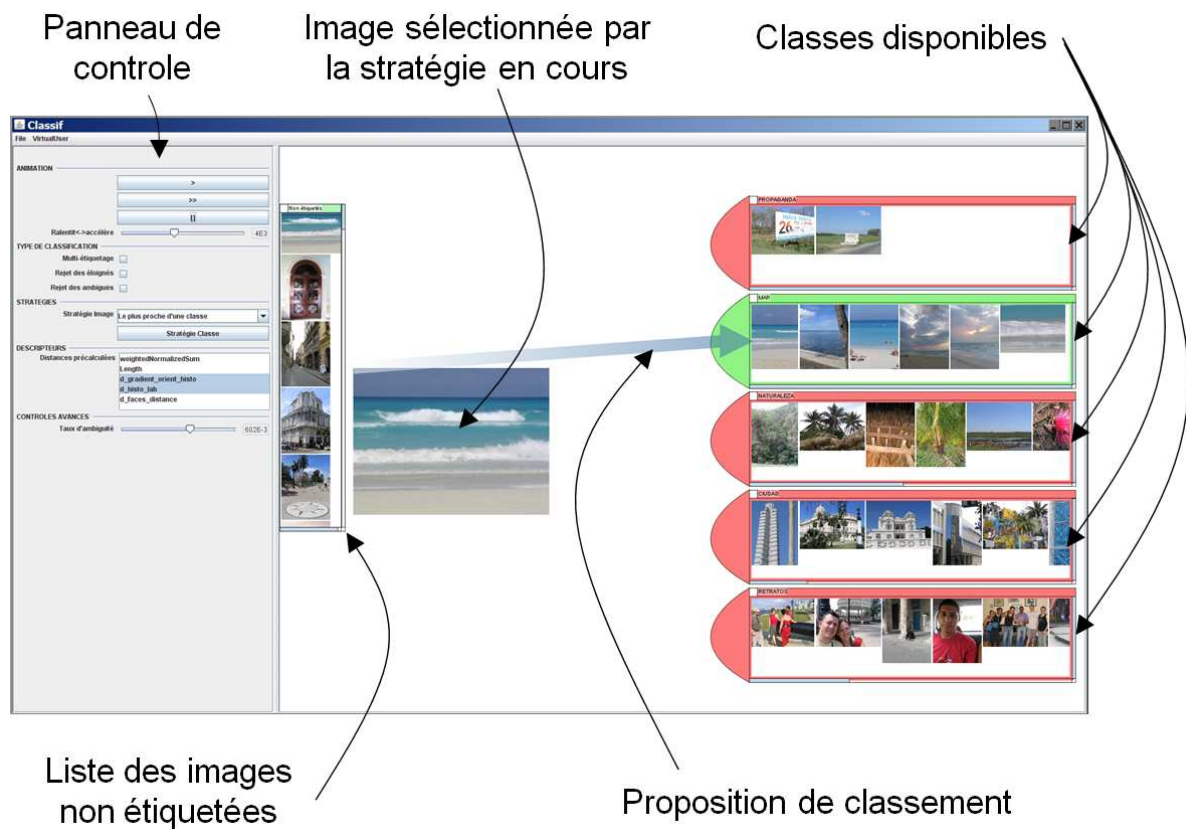


FIGURE 5.3 : Vue globale de l'interface.

3.1.1. La vue interactive

La vue interactive (voir figure 5.3) a pour rôle de présenter à l'utilisateur l'organisation actualisée de la collection d'images et quelle proposition de classement lui faire pour une ou plusieurs images sélectionnées.

A gauche de la vue interactive, une liste verticale contient les images qui n'ont pas été encore étiquetées. Cette liste est triée de haut en bas, de l'image la plus à la moins représentative suivant la stratégie de sélection d'images en cours. Afin de garantir une meilleure visibilité, la liste n'affiche que les premières images les plus représentatives de la stratégie.

A droite de la vue interactive, les images qui ont été précédemment étiquetées sont rangées dans des listes horizontales représentant chacune une classe. Pour les mêmes raisons de lisibilité, seule une petite partie des images membres des classes est affichée.

La première image de la liste des images non étiquetées, la plus représentative de la stratégie de sélection en cours, est dupliquée et agrandie au centre de la vue. Les affichages des listes des classes se mettent alors à jours : pour chaque liste, les k plus proches voisins de l'image non étiquetée dans l'espace des descripteurs sont ainsi représentés de gauche à droite. Une proposition de classement est alors indiquée à l'utilisateur par une flèche pointant la classe conseillée (colorée en vert, les classes exclues étant colorées en rouge).

3.1.2. Le panneau de contrôle

Le panneau de contrôle (voir figure 5.4) reprend les différents paramètres que peut contrôler l'utilisateur sur le fonctionnement du système :

- ANIMATION : le contrôle d'une animation jouant automatiquement en boucle les actions successives de sélection, proposition et validation d'étiquette pour une image.
- TYPE DE CLASSIFICATION : correspond aux options définissant le schéma de décision (voir page 96) imposé par l'utilisateur (classification de manière exclusive ou non, avec rejet ou non).
- STRATEGIES : le choix d'une stratégie et le contrôle du taux d'ambiguïté entre les classes (le paramètre f voir page 89).

3.2. Interactions avec l'utilisateur

3.2.1. Interactions dans la vue globale

L'utilisateur peut valider ou non la proposition d'étiquetage faite par le système sur l'image non étiquetée sélectionnée. L'interaction de base est le glissé-déposé : l'utilisateur peut déplacer l'image dans la classe qui lui convient.

De plus, l'utilisateur peut déplacer les images à volonté si certains membres d'une classe lui paraissent moins pertinents qu'auparavant, soit en les déplaçant dans une autre classe ou soit en les remettant dans la liste des non étiquetées s'il ne sait plus quelle classe choisir.

L'utilisateur peut également créer, par un menu contextuel, une nouvelle liste vide pour représenter une nouvelle classe. Il doit alors y déplacer au moins une image pour avoir un minimum de connaissance

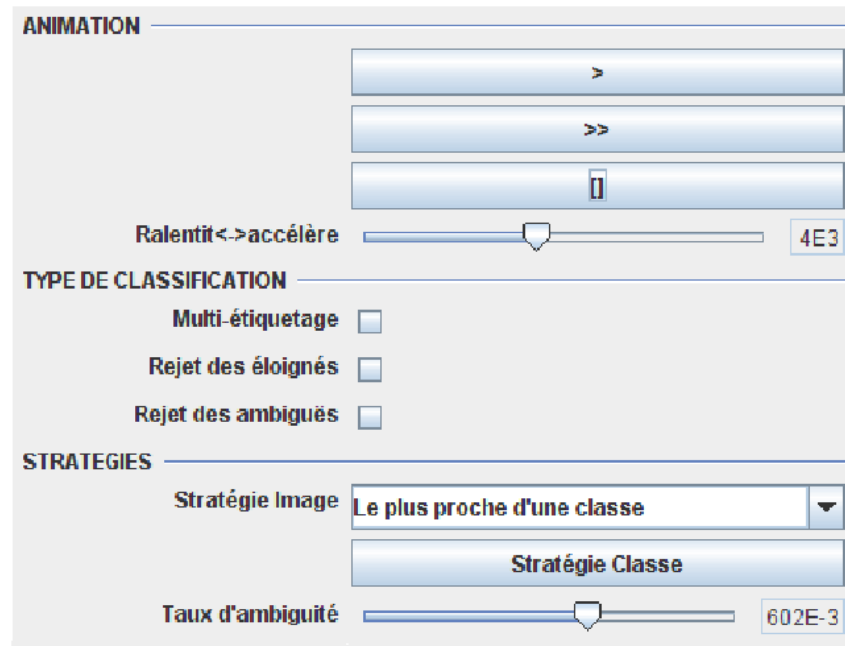


FIGURE 5.4 : Le panneau de contrôle

sur cette nouvelle classe, afin quelle soit prise en compte pour les prochaines sélections d'images et les prochaines propositions de classement.

3.2.2. Contrôle de l'animation

Dans le fonctionnement le plus simple, l'utilisateur doit classer manuellement une par une les images successivement sélectionnées. Il en résulte une multitude d'actions de type "glissé-déposé" qui peut être rapidement fatiguant pour l'utilisateur. Nous proposons le contrôle d'une animation pour faciliter le travail de l'utilisateur. L'objectif est de considérer une validation "passive" et une correction "active" :

- La validation est "passive" car si l'utilisateur est satisfait de la proposition de classement faite sur l'image sélectionnée, une animation simule le glissé-déposé. L'utilisateur n'a donc pas d'action manuelle à effectuer.
- La correction est "active" car si l'utilisateur n'est pas satisfait de la proposition de classement, il interrompt l'animation en déplaçant l'image ou la flèche dans la classe qui lui convient. Une fois l'image étiquetée, l'animation se poursuit sur le reste des images non étiquetées. Une correction est donc plus fatigante pour l'utilisateur car il doit faire une action manuelle.

L'animation possède 4 boutons de contrôles :

- "Play" pour lancer une animation répétant en boucle les étapes suivantes : sélection de l'image la plus représentative de la stratégie en cours, affichage d'une proposition de classement et déplacement de l'image dans la classe proposée.
- "Avance rapide" pour classer directement les images dans les classes sans afficher les déplacements.
- "Pause" pour interrompre l'animation et avoir le temps d'analyser les contenus si l'animation est trop

rapide.

- Un réglage de la durée d'une boucle d'animation, correspondant au classement d'une image, proportionnelle avec la valeur de la probabilité pignistique de l'hypothèse représentant la classe proposée. Par défaut la durée d'une boucle est de 5 secondes.

3.2.3. Type de classification

Cette section permet à l'utilisateur de contraindre le type de propositions de classements sur les images non étiquetées sélectionnées (voir la première partie de ce chapitre 5) :

- Le choix entre des étiquetages simples pour classer les images de manière exclusive (voir figure 5.3), et entre l'étiquetage multiple pour classer les images de manière non exclusive dans les classes (voir figure 5.5).
- L'activation ou non des rejets en distance ou-et en ambiguïté : en activant ces options, des listes de rejets apparaissent pour mettre de côté les images (voir figure 5.6).

Remarque : il n'est pas possible d'activer le rejet en ambiguïté et l'étiquetage multiple en même temps puisque l'ambiguïté est éventuellement utilisée pour proposer les images dans plusieurs classes.

3.2.4. Stratégies

L'utilisateur peut choisir une des stratégies orientées "image" dans une boîte de sélection, ou bien demander la stratégie orientée "classe" sur l'ensemble des images non étiquetées (voir les stratégies présentées au chapitre 4). L'utilisateur peut changer à tout instant de stratégie.

Les stratégies orientées "images" permettent de sélectionner les images une par une en prenant celle qui à chaque itération est la plus représentative de la stratégie choisie par l'utilisateur. À chaque fois que l'image est classée, la connaissance est réévaluée et la liste verticale à gauche des images non étiquetées est réordonnée par la stratégie en cours.

Lorsque que la connaissance sur les classes se confirme, à force d'y classer des images, l'utilisateur peut passer à **la stratégie orientée "classe"**. La figure 5.7 présente la vue après un appel de cette stratégie orientée "classe". Plusieurs "listes de propositions" sont créées en face de chaque liste de classe. Les images à l'intérieur de ces listes de propositions sont triées en interne de la plus à la moins probable (au sens des probabilités pignistiques), de droite à gauche. L'utilisateur peut alors cliquer sur les triangles verts pour valider ces propositions de classement soit image par image, ou par petit lots de 5 images (les 5 images affichées). De même, 2 listes d'images en haut affichent les images qui sont rejetées en distance, et celles qui le sont par ambiguïté.

Lorsque que l'utilisateur n'est plus satisfait des propositions de classement (les dernières images des listes étant de moins en moins probables), il peut relancer la stratégie classe pour réévaluer la connaissance en fonction de toutes les images qu'il vient d'étiqueter.

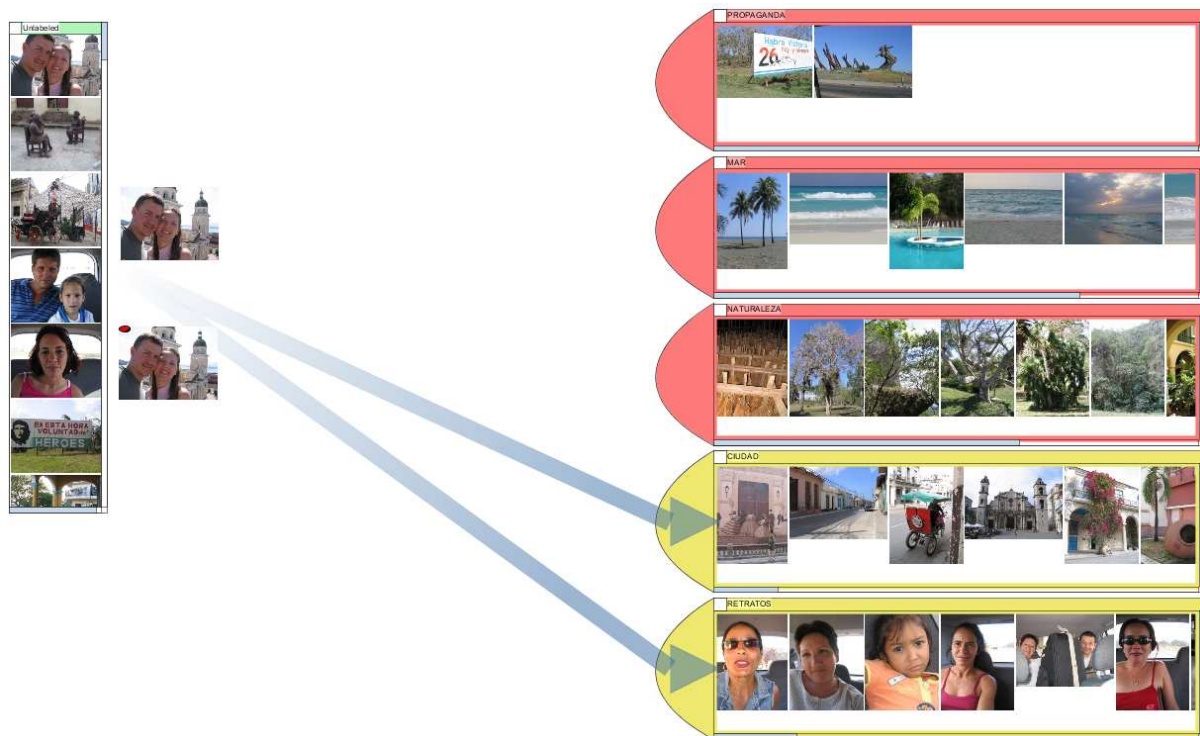


FIGURE 5.5 : *Etiquetage multiple : l'image sélectionnée est dupliquée pour pouvoir être intégrée dans les classes pointées par les flèches. Dans ce cas une image représentant un portrait sur un fond de monument historique est recommandée pour la classe contenant des portraits et la classe contenant des images de ville. Si cette proposition de classement ne satisfait pas l'utilisateur il peut soit déplacer les flèches vers les classes qu'il pense être plus pertinentes, soit ajouter une flèche supplémentaire vers une troisième classe en cliquant sur la zone colorée devant la classe, ou bien encore soit supprimer une des 2 flèches en déplaçant le bout de la flèche dans une zone blanche.*

L'utilisateur peut contrôler le taux d'ambiguïté entre les classes (voir les explications sur le paramètre f exposé en fin de chapitre 4 page 89). Si ce paramètre est trop faible, les images auront tendance à être toutes rejetées, et à l'inverse si le paramètre f est trop élevé, les images auront tendance à être toutes globalement ambiguës.

Par défaut, ce paramètre est réglé sur la valeur $f = 0,75$. La stratégie orientée "classe" permet d'étalonner manuellement ce paramètre. Un bon réglage consiste alors à répartir les images dans les différentes listes de propositions de classement et dans les listes des rejetés et des ambiguës, en évitant que les images soient toutes rejetées ou ambiguës.

Remarque : proposition d'une représentation de la connaissance avec les ambiguïtés

La stratégie orientée classe permet de se focaliser sur les images les plus positives par classes, c'est-à-dire les plus probables (au sens des probabilités pignistiques), en affichant des listes de propositions en face de chaque liste de classe (figure 5.7). Cette vue convient bien pour de la classification d'images de manière exclusive et permet un gain de productivité sachant que les images peuvent être étiquetées par petits lots simultanément.

Cependant, cette vue ne permet pas de situer précisément les images ambiguës. En effet, ces images sont mises dans une seule liste à part, en haut de la figure 5.7, ce qui ne permet pas de voir entre quelles classes les images sont ambiguës. Nous pouvons proposer une seconde vue plus adaptée à la lecture des ambiguïtés (voir figure 5.8).

Cette vue tente de positionner dans l'espace 2D toutes les images non étiquetées en fonction de leur état (positif, rejeté, ambigu). La connaissance extraite sur les images est représentée plus finement, notamment pour les images ambiguës où des arêtes pointent vers les classes possibles. Ces liens représentent donc vers quelles classes les images peuvent être potentiellement étiquetées.

Dans cette vue :

- Chaque classe est représentée par une seule image membre, et toutes ces images sont disposées régulièrement sur un cercle.
- La liste des images rejetées est disposée sur un second cercle extérieur.
- Les images positives proposées pour chaque classe sont disposées autour des images "classe", en colimaçon, de la plus probable à la moins probable.
- Les images ambiguës sont disposées entre les "classes" en s'appuyant sur le barycentre des probabilités pignistiques sur les hypothèses positives.

Pour conclure, cette seconde vue est une proposition alternative qui exigerait des développements supplémentaires et un travail de réflexion sur l'ergonomie, notamment sur les actions à effectuer pour étiqueter les images par lots de taille importante. Pour l'instant, l'utilisateur peut étiqueter les images par lots en les sélectionnant avec une boîte englobante. Mais nous pourrions envisager des composants graphiques simplifiant cette tâche, par exemple avec des zones réactives au passage de la souris pour sélectionner plusieurs images à étiqueter en même temps.

4. Conclusion

Ce chapitre a montré la façon dont on tire profit de la connaissance sur la base d'images pour aider l'utilisateur à les classer.

L'utilisateur peut imposer au système le type d'étiquetage avec lequel il veut travailler : étiquetage simple ou multiple, avec rejet(s) ou non.

L'interface proposée a pour but de limiter le nombre d'actions nécessaires à l'utilisateur pour étiqueter les images. En effet, le système est capable de faire des propositions d'étiquettes et de "jouer" automatiquement le classement des images à travers une animation. Cependant l'utilisateur reste maître du système et il peut interrompre à tout instant les étiquetages qui ne lui conviennent pas.

Un enjeu majeur de notre système est alors d'arriver à faire les propositions pertinentes pour l'utilisateur. Plus les propositions satisferont l'utilisateur, moins il lui sera nécessaire d'effectuer des corrections et moins l'organisation de la collection d'images sera fatigante.

Or, les différentes stratégies de sélections d'images permettent d'améliorer de différentes manières la connaissance sur les classes. Le but du prochain chapitre consacré aux expérimentations est de voir dans quelles mesures ces différentes stratégies peuvent aider un utilisateur à classer des images, et indirectement voir la fatigue qu'elles peuvent entraîner sur l'utilisateur.

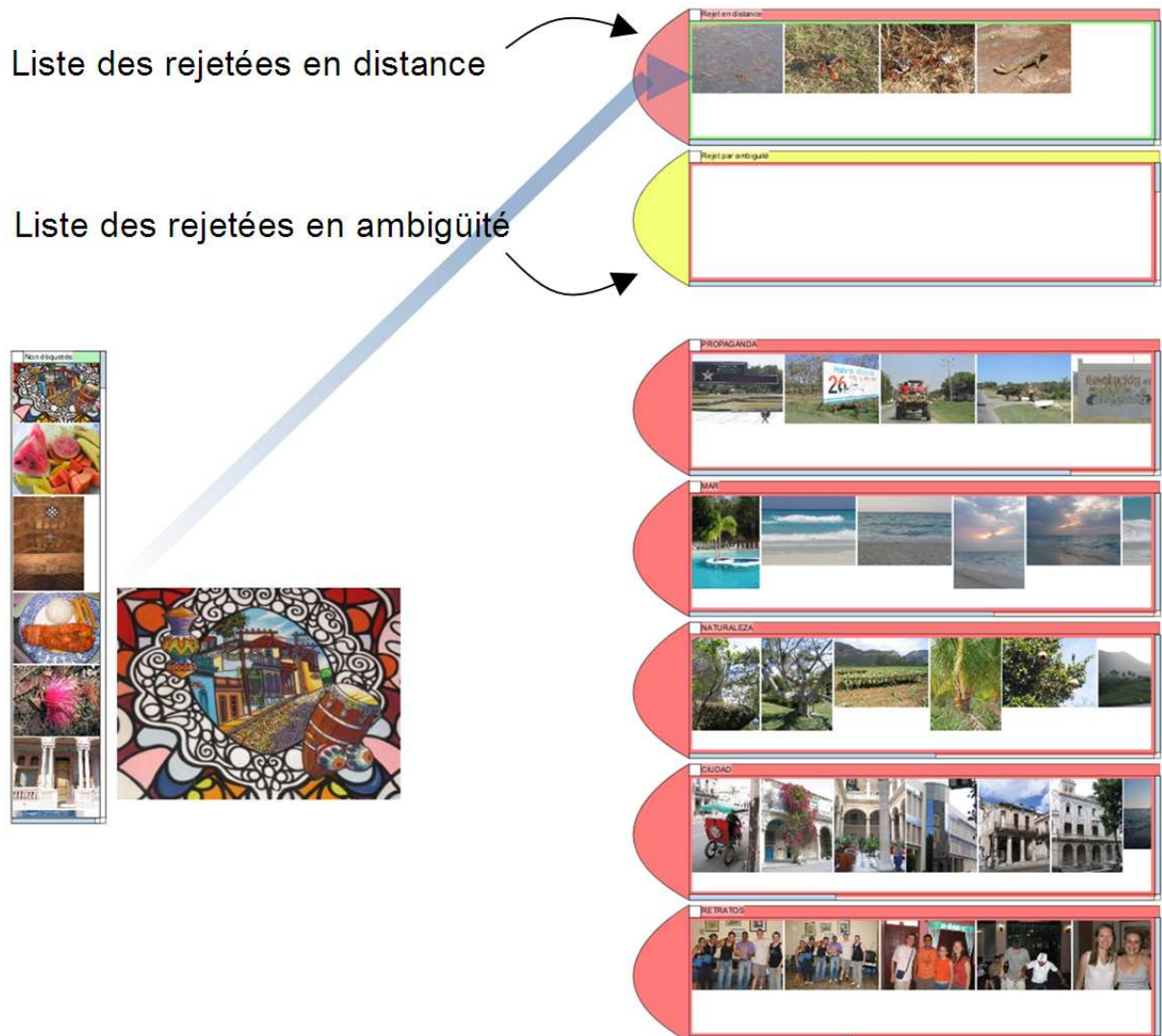


FIGURE 5.6 : Proposition de rejet en distance sur une image sélectionnée : selon l'état de la connaissance l'image au centre de la figure ne correspond à aucune des classes proposées. L'utilisateur a activé les options de rejets en distance et en ambiguïté. Le système propose alors à l'utilisateur de mettre l'image de côté dans la liste des images rejetées.



FIGURE 5.7 : Stratégie orientée "classe" : plusieurs listes de propositions sont créées en face de chaque liste de classe. Les images à l'intérieur de ces listes sont triées en interne de la plus à la moins probable. L'utilisateur peut cliquer sur les triangles verts pour valider ces propositions de classement image par image, ou par petit lots de 5 images. Deux listes d'images en haut affichent les images qui sont rejetées en distance, et celles qui le sont par ambiguïté.

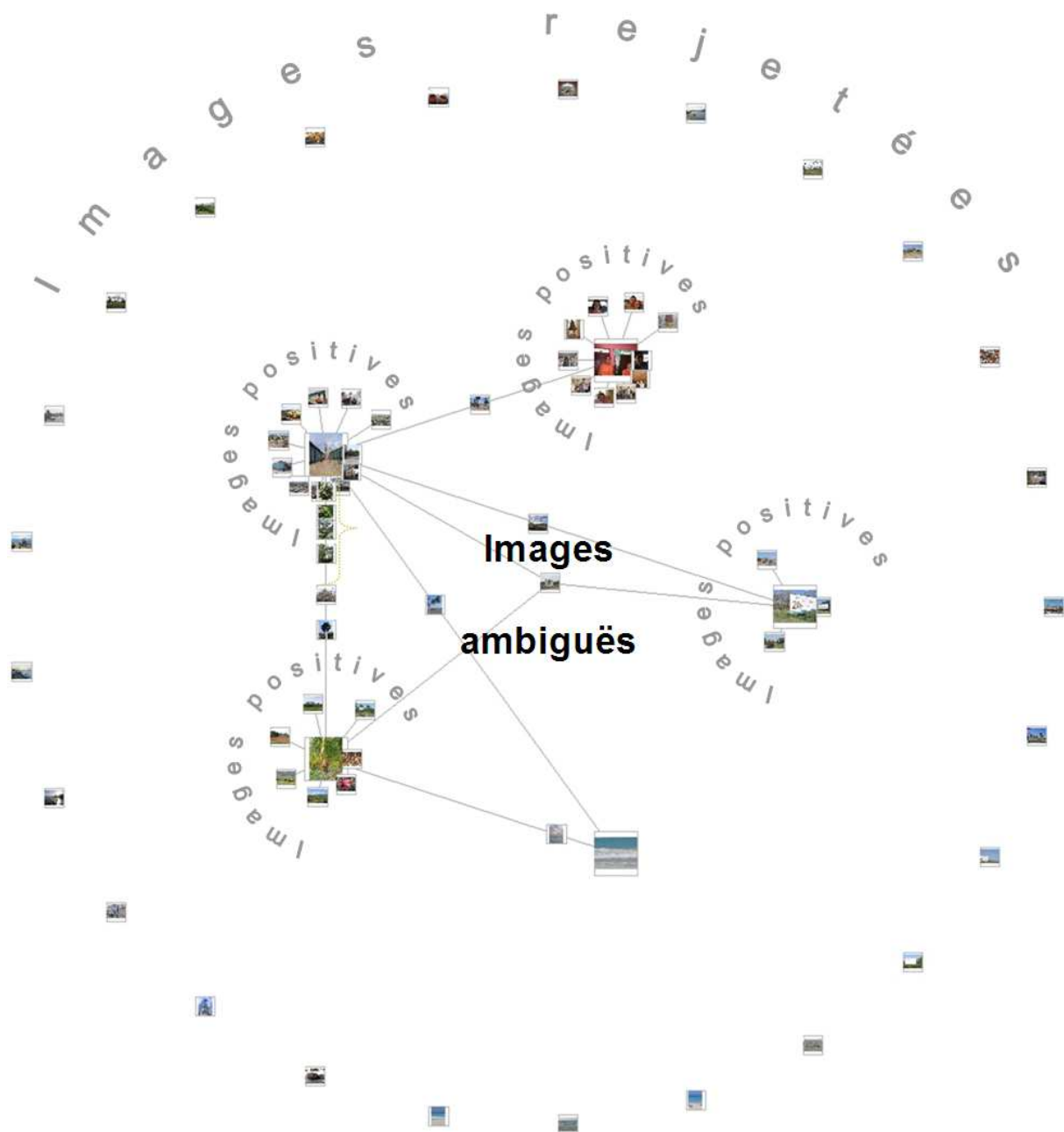


FIGURE 5.8 : Représentation de la connaissance pour afficher explicitement les ambiguïtés : 5 classes sont représentées par une seule image chacune (de grande taille). Les images membres sont disposées sur un premier cercle, et les images rejetées sont disposées sur un deuxième cercle extérieur. Les images ambiguës sont situées à l'intérieur du cercle et leur placement est déduit en fonction du barycentre des probabilités pignistiques sur les hypothèses positives. Ainsi, l'utilisateur peut percevoir dans quelles mesures une image non étiquetée est ambiguë et avec quelles classes.

PERFORMANCES, EXPÉRIMENTATIONS, ÉVALUATIONS

Description du contenu

1. Introduction	116
2. Performances de classification automatique	117
2.1. Métriques et descripteurs	118
2.1.1. Comparaisons dans les mêmes conditions d'expérimentation	119
2.1.2. Comparaisons en fonction du paramètre f	120
2.1.3. Influence du paramètre k	123
2.2. Fusion de descripteurs	125
2.2.1. Résultats avec l'approche tardive	125
2.2.2. Comparaison entre approches précoce et tardive	126
2.3. Influence du paramètre f dans le cas de la fusion tardive	128
2.4. Optimisation du calcul des distributions de masses	128
2.5. Caractérisations et mesures sur une base de données multi-étiquetée	131
3. Caractérisations des stratégies de sélection active d'images	135
3.1. Evaluation classique des stratégies de sélection active	135
3.2. Caractérisation des stratégies dans le contexte applicatif	138
3.2.1. Courbes d'évaluation	138
3.2.2. Liens avec l'effort de l'utilisateur	140
3.3. Influence du paramètre f	141
3.4. Sélection active dans le cas du multi-étiquetage	143
4. Evaluation avec une documentaliste de l'INA	148
5. Extensions	153
5.1. Intégration d'informations métadonnées partielles ou inexactes	153
5.2. Structuration de vidéos	156
5.2.1. Journaux télévisés	157
5.2.2. Episode d'une série	160
6. Conclusion	162

1. Introduction

Ce dernier chapitre présente des évaluations sur différents aspects du système proposé pour aider un utilisateur à structurer une collection d'images par le contenu. Les évaluations se décomposent en 3 axes d'études reprenant les différentes étapes du système que nous avons décrit aux précédents chapitres (voir figure 6.1) :

1. **Performances de classification automatique** : des expérimentations permettent d'analyser dans quelles mesures le choix de mesures de dissimilarités, des descripteurs, du schéma de fusion (précoce ou tardif) et des réglages internes du système modifient les fonctions de croyance, et en conséquence, le pourcentage d'images correctement classées automatiquement.
2. **Caractérisation des stratégies** : des expérimentations mettent en évidence l'impact des différentes stratégies de sélection active sur les performances de classification des images. Chaque stratégie est analysée dans la perspective de son exploitation par un utilisateur.
3. **Evaluation utilisateur** : cette partie concerne une évaluation qualitative du système avec une documentaliste fortement impliquée dans les activités de la Photothèque de l'INA.

Une dernière partie présente des extensions envisageables dans les cadres applicatifs de l'organisation personnelle de photographies avec intégration d'informations métadonnées, et de la structuration de vidéos.

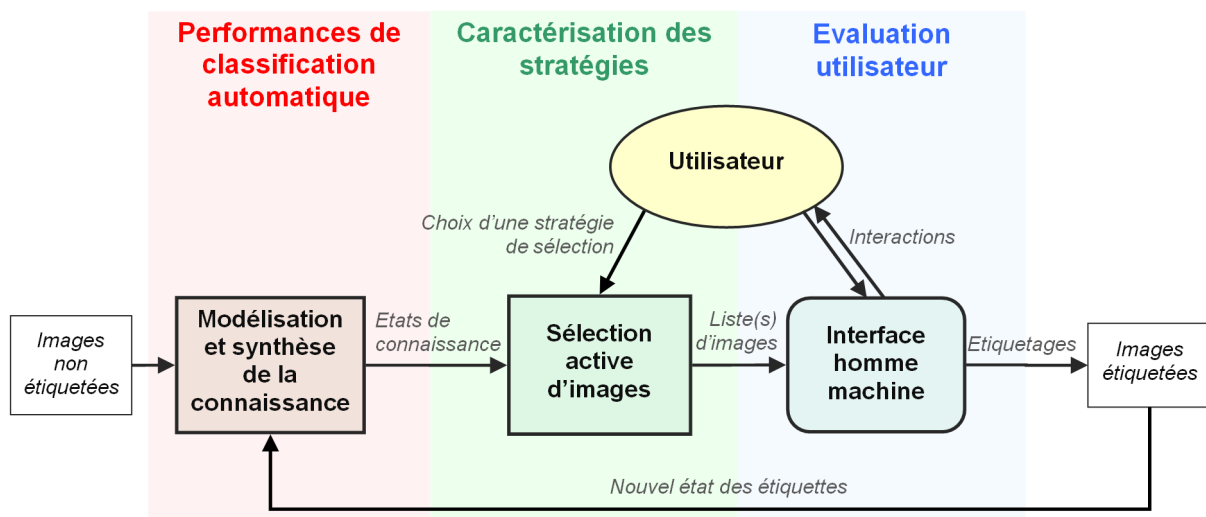


FIGURE 6.1 : Les trois axes d'évaluations étudiés dans ce chapitre : les performances de classification automatique, la caractérisation des stratégies de sélections actives d'images, et enfin une évaluation utilisateur.

2. Performances de classification automatique

Nous avons présenté en grande partie notre travail sous l'angle d'une aide que nous pouvons apporter à un utilisateur pour l'assister dans l'organisation de ses collections d'images. Cependant avant d'étudier les aspects impliqués par la présence d'un utilisateur, il est important de pouvoir apprécier les performances de classification automatique du coeur du système sur des bases d'images. Les résultats de ces analyses peuvent indiquer comment paramétrer le système par défaut.

Les performances du système dépendent en grande partie de la formalisation des fonctions de croyance. En particulier les premières étapes de la modélisation de la connaissance reposent sur les fonctions de croyance (voir page 51) :

$$m^{\Omega_q}(\Omega_q) = \prod_{i=0}^{k-1} (1 - \alpha(u, l_q^i)) \quad (6.1)$$

$$m^{\Omega_q}(H_q) = 1 - \prod_{i=0}^{k-1} (1 - \alpha(u, l_q^i)) \quad (6.2)$$

où $\Omega_q = \{H_q, \overline{H_q}\}$ est le cadre de discernement associé à une classe C_q , u une image non étiquetée, l_q^i un ième voisin appartenant à C_q et $\alpha(u, l_q^i)$ la fonction de masse :

$$\alpha(u, l_q^i) = \alpha_0 \cdot e^{-\left(\frac{d(u, l_q^i)}{\sigma_{i,q}^s}\right)^\beta} \quad (6.3)$$

D'après ces expressions, les paramètres qui peuvent fortement influencer la modélisation de la connaissance, et en conséquence les performances de classification, sont :

1. le type de distance d et de l'espace de description s où sont représentées les images u et l_q^i ,
2. le paramètre global f contrôlant le taux de chevauchement entre les classes en modifiant simultanément les valeurs de tous les paramètres $\sigma_{i,q}^s$ de chaque image étiquetée l_q^i ,
3. k le nombre de plus proches voisins à combiner.

Dans un premier temps, il est donc important d'étudier quelles mesures de dissimilarité sont pertinentes, pour quel descripteur, et avec quelles valeurs de f et de k .

Dans un deuxième temps, nous pouvons voir dans quelles mesures les combinaisons de différents descripteurs améliorent les performances de classification.

Ces fusions de descripteurs peuvent entraîner un coût de calculs élevé. Nous montrons plus loin qu'il est possible d'optimiser les coûts de calculs des distributions de masses en limitant leur nombre de propositions.

Enfin, nous étudions pour clore cette partie, le cas de classification automatique multi-étiquette en comparant les performances de notre méthode avec d'autres approches couramment employées pour ce type de problème.

2.1. Métriques et descripteurs

Conditions d'expérimentation

Jeu de données :

Une première série de tests est effectuée sur le jeu de données que nous nommerons *gt1000*, une base d'images Corel de 1000 images proposée dans [WLW01]. Ce jeu de données est souvent utilisé pour évaluer des systèmes d'indexation multimédia et se prête bien à des problèmes de classification multi-classe. La base contient 10 classes contenant 100 images chacune liées à des catégories sans réelle ambiguïté sémantique, c'est-à-dire que les images représentent chacune explicitement une seule illustration d'un des concepts suivants : "fleurs", "montagnes", "bus", "dinosauriens", "nourriture", "peuples", "monuments", "plage", "chevaux" et "éléphants".

Chaque classe est divisée en 2 parties afin de disposer d'un jeu d'apprentissage de 10x50 images exemples, et d'un jeu de 10x50 images à tester. Il est alors possible de mesurer le pourcentage d'images correctement classées parmi les 500 images de tests, après l'entraînement du système sur les 500 images d'apprentissage, pour différents choix de mesures de dissimilarité et de descripteurs visuels.

Formalisation :

Dans le cadre de notre méthode, cette base de données implique de définir un cadre de discernement Ω contenant théoriquement de 2^{10} hypothèses, soit toutes les combinaisons des hypothèses de base "positives" de H_q et "négatives" \overline{H}_q décrivant les états d'appartenance des images non étiquetées à une des classes C_q ($q \in \{1, 2, \dots, 10\}$).

Les expérimentations sur cette base d'images *gt1000* sont effectuées dans le cadre d'un problème de classification multi-classe, où les images doivent être classées automatiquement de manière exclusive. De plus, toutes les classes sont connues et une image non étiquetée appartient obligatoirement à une des classes d'après la vérité terrain. Il n'est donc pas nécessaire de prendre en compte l'hypothèse de rejet pour classer automatiquement une image. Nous opérons donc avec le schéma "S" d'étiquetage simple sans rejet (voir page 98), utilisant l'espace de décision Ω_S , l'ensemble des 10 hypothèses positives du cadre de discernement Ω , afin de décider quelle étiquette associer aux images.

L'entraînement du système sur les images de la base d'apprentissage, consiste à estimer automatiquement les paramètres $\sigma_{l_i}^s$ associés aux images étiquetées de la base d'apprentissage en fonction de son étiquette q , de l'espace de description s considéré, de la distance utilisée et du paramètre de chevauchement f (voir chapitre 4 page 89).

Descripteurs :

Nous utilisons ici des descripteurs visuels standards : un histogramme des orientations selon 8 directions et 8 amplitudes, et des histogrammes recensant les couleurs selon les espaces couleurs $\{H,s,v\}$, $\{L,a,b\}$, $\{L,u,v\}$ et $\{R,g,b\}$, pour 8 intervalles par composantes, ce qui donne des vecteurs descripteurs de 512

dimensions.

Ces descripteurs standards peuvent être remplacés par des descripteurs plus proches de l'état de l'art. Mais il est toutefois intéressant de pouvoir déterminer l'espace de couleurs le plus pertinent. D'après la littérature, les espaces $\{H,s,v\}$ et $\{L,a,b\}$ sont régulièrement utilisés (sachant que $\{L,u,v\}$ est préféré parfois à $\{L,a,b\}$ pour des raisons de coût de calculs [BLSB04]).

Métrie :

Les fonctions de croyance reposent sur une conversion symbolique-numérique d'une distance entre des descripteurs d'images en une masse (voir chapitre 3 page 48). Les descripteurs utilisés étant des histogrammes, nous avons comparé quatre métriques classiquement utilisées en indexation multimédia [TFMB04] : la distance de Manhattan d_{L_1} , la distance Euclidienne d_{L_2} , le test du χ^2 et la distance de Bhattacharyya d_{Bhat} :

$$d_{L_1}(h_1, h_2) = \sum_{i=1}^n |h_1(i) - h_2(i)| \quad (6.4)$$

$$d_{L_2}(h_1, h_2) = \sqrt{\sum_{i=1}^n (h_1(i) - h_2(i))^2} \quad (6.5)$$

$$d_{\chi^2}(h_1, h_2) = \sum_{i=1}^n \frac{|h_1(i) - h_2(i)|}{|h_1(i) + h_2(i)|} \quad (6.6)$$

$$d_{Bhat}(h_1, h_2) = -\log \left(\sum_{i=1}^n \sqrt{\frac{h_1(i) \cdot h_2(i)}{|h_1| \cdot |h_2|}} \right) \quad (6.7)$$

avec h_1 et h_2 deux histogrammes de même dimensions n , et $|h|$ la norme L_1 d'un histogramme h .

2.1.1. Comparaisons dans les mêmes conditions d'expérimentation

La figure 6.2 donne les résultats pour la comparaison des 4 métriques sur les descripteurs des orientations, et pour les mêmes paramètres internes $f = 0,7$ et $k = 5$.

La distance de Bhattacharyya, fournit le meilleur résultat : elle permet d'obtenir un écart de 8,8%, soit 44 images correctement classées de plus (sur un total de 500), vis-à-vis de la distance L_2 , la moins performante dans ce test. La distance L_1 et le test du χ^2 donnent des résultats équivalents autour de 62% des images correctement classées automatiquement.

La figure 6.3 fournit les résultats par métrie pour les différents descripteurs couleurs avec pour paramètres internes $f = 0,7$ et $k = 5$. Les distances de Bhattacharyya et L_1 donnent globalement de meilleurs résultats vis-à-vis des distances L_2 et surtout du χ^2 . En particulier, les distances de Bhattacharyya et L_1 donnent de bons résultats autour de 80% d'images correctement classées sur les trois espaces de couleurs $\{H,s,v\}$, $\{L,u,v\}$, et $\{R,g,b\}$. L'espace $\{L,a,b\}$ donne généralement les moins bonnes performances de classification. Le meilleur résultat est obtenu pour la distance L_1 sur l'espace $\{H,s,v\}$

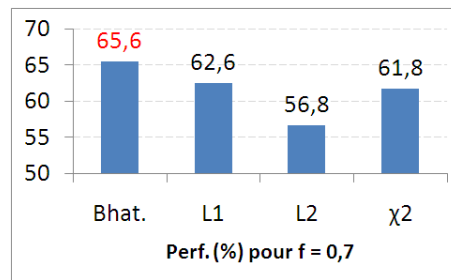


FIGURE 6.2 : Pourcentage des images correctement classées automatiquement sur le jeu de test gt1000 en utilisant différentes métriques comme mesure de dissimilarité entre les descripteurs des orientations des images, et avec pour paramètres internes $f = 0,7$ et $k = 5$. La distance de Bhattacharyya permet d'obtenir le meilleur pourcentage avec 65,6% des images correctement classées.

avec 81,4%, soit 407 images correctement classées.

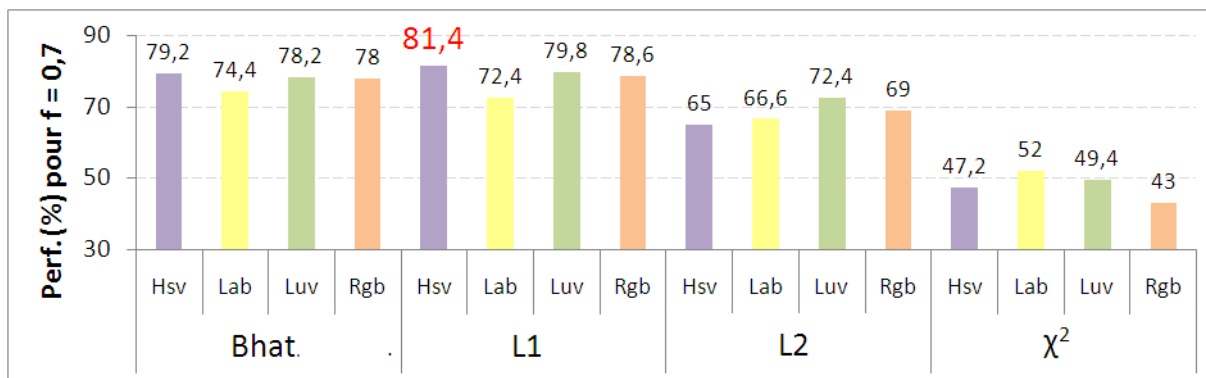


FIGURE 6.3 : Pourcentage des images correctement classées automatiquement sur le jeu de test gt1000 en utilisant différentes métriques sur différents descripteurs couleurs, avec pour paramètres internes $f = 0,7$ et $k = 5$. Les distances L_1 et de Bhattacharyya permettent d'obtenir des meilleures performances de classification automatique du même ordre avec les espaces couleurs $\{H,s,v\}$, $\{L,u,v\}$, et $\{R,g,b\}$.

2.1.2. Comparaisons en fonction du paramètre f

Au regard de ces résultats, nous pourrions conclure que la combinaison la plus intéressante est le descripteur couleur $\{H,s,v\}$ avec la distance L_1 . Cependant, les fonctions de masses sont très influencées par le paramètre f contrôlant le chevauchement des classes.

Nous pouvons refaire les tests avec en passant de $f = 0,7$ à $f = 0,8$, c'est-à-dire en élargissant les zones de connaissance autour des images étiquetées et en provoquant ainsi plus d'ambiguïté entre les classes (voir figure 6.4).

Nous pouvons constater que la distance L_1 appliquée aux descripteurs dans l'espace couleurs $\{H,s,v\}$, donnant 81,4% de bonnes classifications avec $f = 0,7$ passe à 60,8% dans le cas où $f = 0,8$, ce qui correspond à une perte importante de 103 images, soit environ un cinquième en moins du nombre total d'images à classer. Les 3 distances L_1 , L_2 et χ^2 tendent quasiment toutes à perdre en performances de classification pour tous les descripteurs couleurs. Par contre la distance de Bhattacharyya permet de

maintenir des performances équivalentes et même un certain gain par rapport au test précédent avec $f = 0,7$.

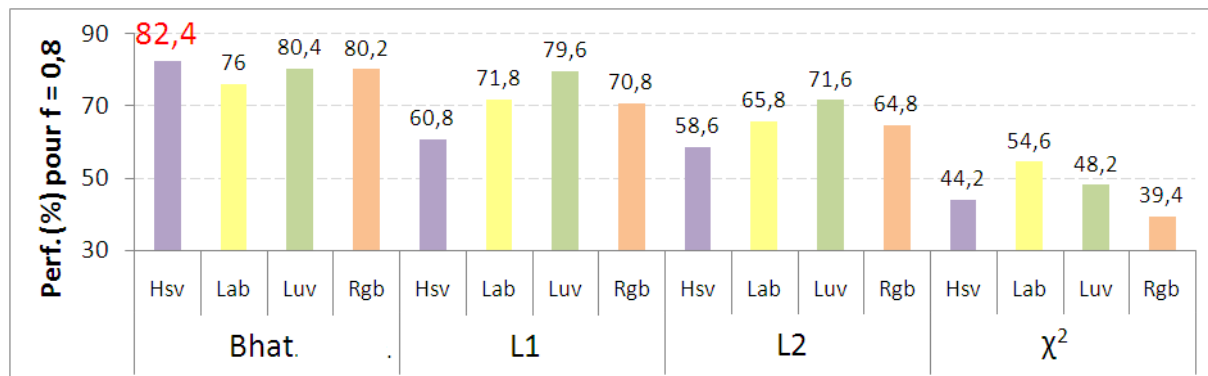


FIGURE 6.4 : Pourcentage des images correctement classées automatiquement sur le jeu de test *gt1000* en utilisant différentes métriques sur différents descripteurs couleurs, et avec pour paramètres internes $f = 0,8$ et $k = 5$. La distance de Bhattacharyya permet de maintenir des performances élevées tandis que les autres mesures voient leurs performances pratiquement toutes se détériorer, vis-à-vis du cas où $f = 0,7$.

Pour comprendre d'où viennent ces différences de performances, nous pouvons analyser les mesures de non-spécificité moyenne des distributions de masses des 500 images de tests. La mesure de non spécificité (voir page 83) nous renseigne sur la finesse de la modélisation de la connaissance. Plus la non spécificité d'une distribution de masses est élevée, plus la masse est répartie sur des propositions de cardinalité élevée, celles qui sont composées de doutes de type Ω_q . En conséquence la distribution de probabilités pignistiques tend vers le cas d'équiprobabilité des hypothèses du cadre de discernement. A l'inverse, si la non-spécificité d'une distribution de masses d'une image non étiquetée est faible, cela signifie qu'une des hypothèses du cadre de discernement Ω , se distingue des autres désignant ainsi une classe pour y associer l'image non étiquetée.

La figure 6.5 rapporte les mesures de non-spécificité moyenne des distributions de masses de toutes les images non étiquetées, pour les précédentes mesures figure 6.3 dans le cas où $f = 0,7$. Les non-spécificités moyennes avec la distance χ^2 sont nettement plus élevées que dans le cas des autres distances. La modélisation de la connaissance est donc plutôt grossière. En conséquence, les valeurs des probabilités pignistiques tendent à être proches, ce qui pourrait expliquer l'apparition d'erreurs de classification plus fréquentes.

La figure 6.6 donne les mesures de non-spécificité moyennes dans le second cas où $f = 0,8$. En augmentant f , on accroît le chevauchement des zones de connaissances des images étiquetées appartenant à différentes classes et l'on augmente ainsi l'ambiguïté entre les classes. Il en résulte alors une augmentation générale des non spécificités moyennes pour toutes les distances pour chaque descripteur.

Cependant, la distance de Bhattacharyya permet de conserver plus de spécificité vis-vis des distances L_1 et L_2 où la perte peut être extrêmement importante. En particulier, nous pouvons voir que la perte de spécificité moyenne des distributions de masses est l'une des plus importante pour la distance L_1 appliquée à l'histogramme $\{H,s,v\}$. Cette observation explique pourquoi cette combinaison est la plus performante

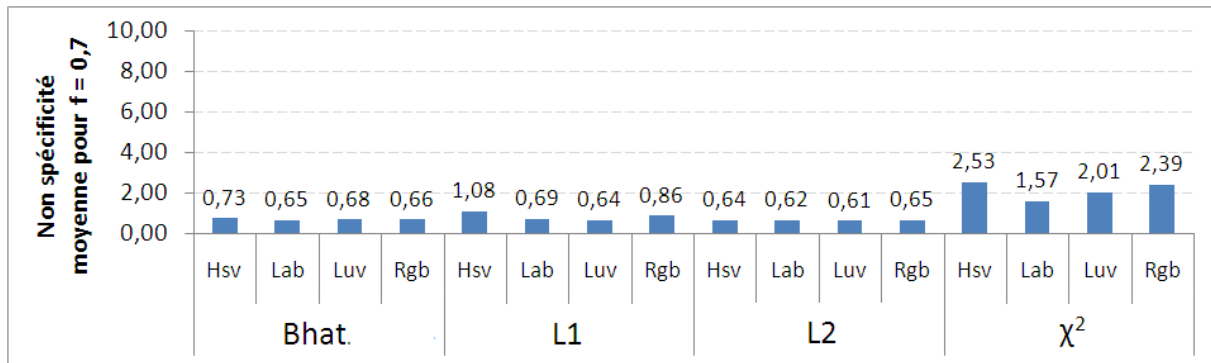


FIGURE 6.5 : Non-spécificité moyenne des distributions de masses sur la base gt1000, avec $f = 0.7$ sur les mêmes combinaisons de distances-descripteurs que la figure 6.3

pour $f = 0,7$ avec 81,4% de bonnes classifications et l'une des moins performantes avec 62% d'images classées correctement pour $f = 0,8$. A l'inverse, sur ces tests la distance de Bhattacharyya est la plus intéressante et permet de limiter la perte de spécificité tout en conservant globalement les meilleures performances avec la plupart des descripteurs. La distance L_1 sur le descripteur couleur $\{L,u,v\}$ permet également d'atteindre une performance de classification correcte de 79,6%.

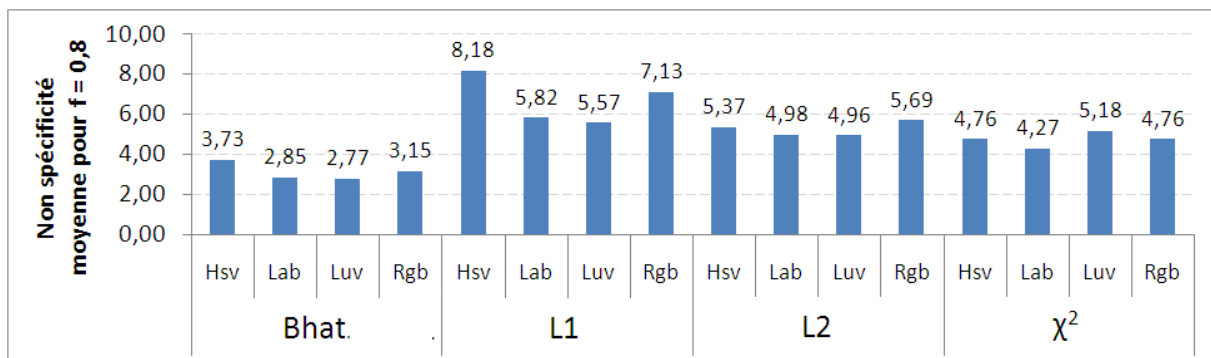


FIGURE 6.6 : Non-spécificité moyenne des distributions de masses sur la base gt1000, avec $f = 0.7$ sur les mêmes combinaisons de distances-descripteurs que la figure 6.4

Nous pouvons raisonnablement supposer que chaque couple métrique-descripteur ne permet pas d'obtenir les meilleures performances de classifications pour les mêmes valeurs de f . La figure 6.7 compare des courbes représentant l'évolution des performances de classification pour quelques valeurs de f . Nous avons retenu dans ce graphique pour chacune des 4 métriques le descripteur permettant d'atteindre le pourcentage le plus élevé de bonnes classifications. Le descripteur $\{H,s,v\}$ est le descripteur le plus adéquat aux distances L_1 et d_{bhat} . La distance L_2 permet des performances relativement bonnes avec le descripteur $\{L,u,v\}$. Le descripteur $\{L,a,b\}$ est le plus pertinent pour le test du χ^2 mais donne des résultats nettement inférieurs aux autres combinaisons descripteur-métrique. Cependant, il faut bien noter la tendance des courbes car certaines sont plus ou moins robustes aux variations de f . En particulier, la distance de d_{Bhat} permet de maintenir des performances de classifications élevées sur une plus large plage de valeurs de f . Nous avons par ailleurs observé ce type d'évolution pour tous les autres descripteurs.

L'analyse de ces tendances peut motiver le choix d'une métrique vis-à-vis d'une autre. Les distances L_1

et d_{Bhat} permettent d'obtenir des performances élevées de classification. La distance de Bhattacharyya semblent moins sensible au paramètre f et moins dépendant du choix du descripteur couleur.

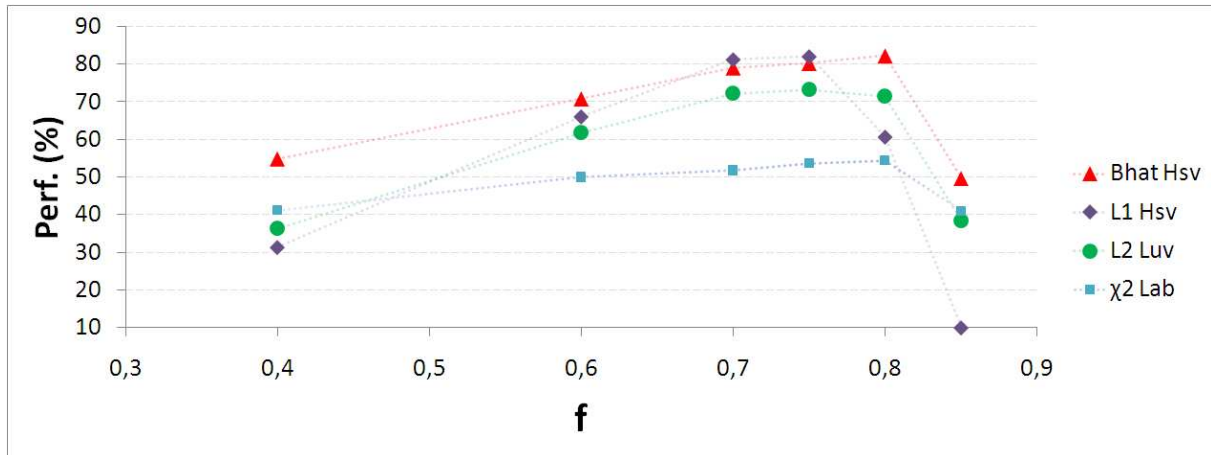


FIGURE 6.7 : Comparaison des performances de classification en fonction de f pour différentes métrique avec leurs descripteurs permettant d'atteindre le pourcentage le plus élevé.

2.1.3. Influence du paramètre k

Dans les premières étapes de la modélisation de la connaissance, on calcule une distribution de masses pour une image non étiquetée vis-à-vis d'une classe. Pour ce faire, les distributions de masses calculées pour k images membres d'une même classes et voisines de l'image non étiquetée dans l'espace de description considéré sont fusionnées pour établir une nouvelle distribution de masse.

Il est donc important de déterminer dans quelle mesure le nombre de k de plus proches voisins peut influencer les performances de classification automatique.

La figure 6.8 compare les performances de classification automatique dans les mêmes conditions d'expérimentations sur la base *gt1000* (distance de Bhattacharyya, descripteur couleur {H,s,v} et $f = 0.8$). Ce test montre qu'il existe une valeur de k optimale autour de $k = 5$ permettant d'atteindre des performances de classification automatique au dessus de 80%. Si k est réglé à une valeur trop grande au delà de 50, les performances de classification saturent à 72,8%, ce qui fait une perte de l'ordre de 10% si k est mal réglé.

Bilan

Nous pouvons retenir de ces expérimentations que le choix d'une mesure de dissimilarité entre les descripteurs d'images est primordiale. La distance classique L_2 n'apparaît pas comme étant la plus pertinente. Le test du χ^2 , pourtant réputé comme étant bien adapté aux comparaisons d'histogrammes, n'est pas adapté aux descripteurs que nous utilisons. Les distances L_1 et de Bhattacharyya permettent les performances les plus élevées. Cependant, la distance L_1 paraît plus sensible au choix du descripteur et à

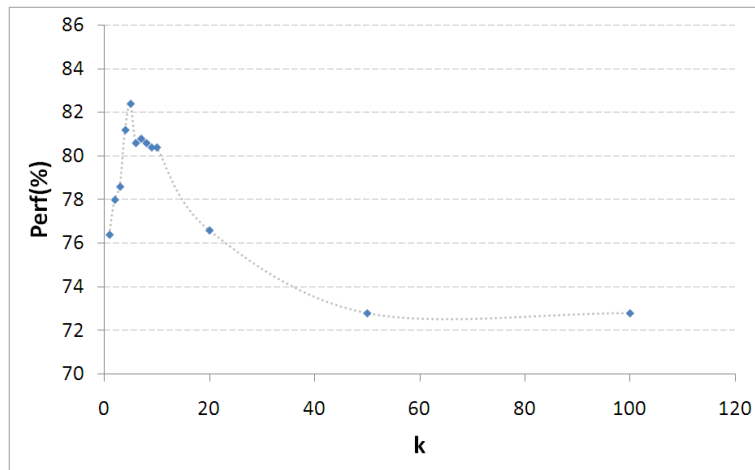


FIGURE 6.8 : Comparaison des performances de classification en fonction de k dans le cas où la distance de Bhattacharyya est appliquée sur le descripteur couleur $\{H,s,v\}$ et avec $f = 0.8$ sur la base $gt1000$. Il existe une valeur de k optimale autour de $k = 5$ permettant d'atteindre des performances de classification automatique au dessus de 80%.

la valeur du paramètre f contrôlant l'ambiguïté entre les classes. La distance de Bhattacharyya apparaît comme étant la plus intéressante, et offre une certaine robustesse au choix des descripteurs couleurs et au paramètre f .

Par contre, il est difficile de conclure sur l'usage d'un espace de couleurs vis-à-vis d'un autre. Si l'espace $\{L,u,v\}$ semble être adapté aux distances L_1 et L_2 , les différences de performances entre les espaces de couleurs pour la distance de Bhattacharyya ne sont pas significatives. De plus, il n'est pas certain que l'on puisse généraliser à partir de ces tests sur la seule base $gt1000$. En effet, nous avons constaté sur d'autres collections d'images que la distance de Bhattacharyya appliquée aux descripteurs couleurs $\{L,a,b\}$ offre les meilleures performances, ce qui n'est pas le cas ici.

Nous pouvons voir également que les performances de classification sont intimement liées à la non-spécificité moyenne des distributions de masses des images de test. Une perspective intéressante serait de sélectionner automatiquement la meilleure métrique en se basant sur l'analyse des non-spécificités moyennes des distributions de masses.

Enfin, le descripteur des orientations ne donne pas de performances aussi bonnes que pour les descripteurs couleurs. Il peut donc être intéressant de voir si la combinaison avec un descripteur orientation permet d'accroître les performances de classification.

2.2. Fusion de descripteurs

Conditions d'expérimentation

La combinaison de descripteurs est étudiée en effectuant des tests sur la même base *gt1000* précédente. Nous disposons d'un descripteur des orientations et de plusieurs descripteurs couleurs. Ces types d'information sont complémentaires puisqu'elles analysent des caractéristiques visuelles de différentes natures. Il est alors raisonnable de penser que la combinaison de ces descripteurs peuvent améliorer les performances de classification automatique.

2.2.1. Résultats avec l'approche tardive

La méthode que nous proposons, utilise un schéma de fusion tardif (voir chapitre 3 page 60). Chaque espace de description est considéré indépendamment des autres. Ainsi, une image non étiquetée est associée à autant de distributions de masses que d'espaces de descriptions sont considérés. Puis, les distributions de masses sont fusionnées, en attribuant éventuellement de la masse sur la proposition de conflit.

La figure 6.9 donne les résultats en termes de performances de classification pour les différentes combinaisons de descripteurs couleurs avec le descripteur des orientations pour les 4 métriques précédentes, avec $f = 0,7$ le taux d'ambiguïté entre les classes, et $k = 5$ le nombre de voisins.

Par rapport au test précédent (figure 6.3), la fusion permet dans la plupart des cas de gagner en performances de classification. Dans les cas où les performances de classification sur les descripteurs seuls sont faibles, le gain peut être très important, notamment pour la distance du χ^2 . Par exemple, la fusion des descripteurs orientations et $\{H,s,v\}$ permet de passer de 47,2% à 64%, soit 84 images supplémentaires correctement classées. Cependant, cette distance du χ^2 , donne toujours les performances les plus faibles.

Globalement, la distance de Bhattacharyya donne une fois de plus les meilleurs résultats, notamment avec le descripteur $\{R,g,b\}$ qui permet de gagner 4,2%, soit 21 images de plus, vis-à-vis du descripteur $\{R,g,b\}$ seul.

Il faut noter également que dans ce test la distance L_1 ne permet pas de gagner d'images supplémentaires avec le descripteur $\{L,u,v\}$, et provoque même une baisse significative de 7% pour le descripteur $\{H,s,v\}$. Cette baisse peut s'expliquer en mesurant la non-spécificité moyenne et le conflit moyen des distributions de masses dans ce cas. Par exemple, dans les mêmes conditions d'expérimentation, la non-spécificité moyenne des distributions de masses est de 0,61 avec le descripteur des orientations et de 1,08 pour le descripteur $\{H,s,v\}$. La non spécificité moyenne des distributions de masses après fusion des deux descripteurs est nettement réduite à 0,14. De plus, le conflit moyen mesuré sur ces mêmes distributions de masses est très faible et vaut 0,02 (par comparaison le conflit moyen vaut 0,1 pour la fusion $\{R,g,b\}$ et orientations avec la distance de Bhattacharyya). Cela signifie que les masses se concentre sur des propositions de cardinalité faible et qu'il y a peu de doute, et que les informations apportées par les

descripteurs sont peu conflictuels. En conséquence, les distributions de masses permettent un engagement très fort pour une des hypothèses du cadre de discernement, ce qui peut se traduire parfois par des erreurs de classifications.

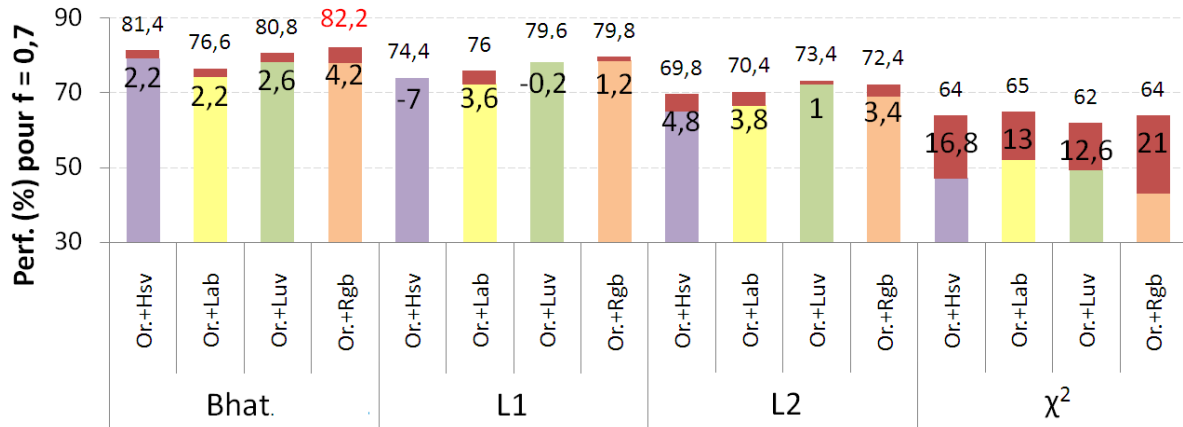


FIGURE 6.9 : Pourcentage d'images classées correctement selon différentes combinaisons du descripteur des orientations avec les descripteurs couleurs, et avec différentes métriques. Sur chaque barre d'histogramme est ajouté en rouge le pourcentage gagné par rapport aux tests précédent figure 6.3 sur les descripteurs couleurs seuls effectués dans les mêmes condition ($f = 0,7$).

2.2.2. Comparaison entre approches précoce et tardive

Il est intéressant de voir si la stratégie de fusion tardive est avantageuse vis-à-vis d'une approche précoce. En effet, l'approche par fusion précoce est courante en indexation multimédia car elle est pratique à mettre en oeuvre. Généralement, la fusion précoce consiste à concaténer en un seul vecteur descripteur les différents descripteurs à combiner. Un nouvel espace de descripteur est considéré et en conséquent les k plus proches voisins ne sont pas les mêmes que dans le cas de l'approche tardive.

Il est possible d'adopter cette stratégie de fusion dans notre méthode. La figure 6.9 compare les performances de classification pour les fusions précoces et tardives sur différents descripteurs, dans les mêmes conditions d'expérimentation ($f = 0,7$ et $k = 5$), toujours sur la même base *gt1000*, en se focalisant uniquement sur la distance de Bhattacharyya qui a donné jusqu'à présent les meilleures performances pour les précédents tests.

La fusion précoce donne globalement de bonnes performances de classification automatique, légèrement inférieures à l'approche tardive, sauf pour le cas de la combinaison des descripteurs {L,a,b} et orientations.

Mais en mesurant les non-spécificités moyennes des distributions de masses, nous pouvons voir que les distributions de masses après fusion tardive sont nettement plus spécifiques. Cela signifie que la modélisation de la connaissance est beaucoup plus précise, tout en garantissant des performances de classification supérieures.

De plus, la fusion tardive permet de bénéficier d'une information supplémentaire grâce au conflit. Dans ce test, le conflit moyen des distributions de masses est quantifiable, et l'on peut ainsi voir dans quelles mesures les sources d'informations sont en désaccord (voir figure 6.11). Nous pouvons par exemple, observer que la combinaison de descripteurs occasionnant le moins de conflit n'est pas forcément celle qui offre les meilleures performances dans ce test.

Or, avec la fusion précoce, dans sa version la plus basique, ce degré d'analyse n'est pas disponible, et même si les performances peuvent être élevées, on ne peut pas savoir directement, sans formalisation supplémentaire, quelle est la contribution des descripteurs et si ils sont en accord.

Enfin, soulignons un dernier problème lorsque l'on souhaite concaténer des descripteurs de dimensions différentes. Un descripteur fournissant un vecteur de dimensions très réduites risque de contribuer très peu au calcul de proximité entre les images, notamment s'il est concaténé avec un second vecteur de dimension importante. A l'extrême, une source d'informations peut juste fournir un scalaire. Avec la fusion tardive, il est possible de gérer ces cas sans problème d'échelle, comme nous le verrons plus loin lors de fusion de descripteurs visuels avec une information de temps.

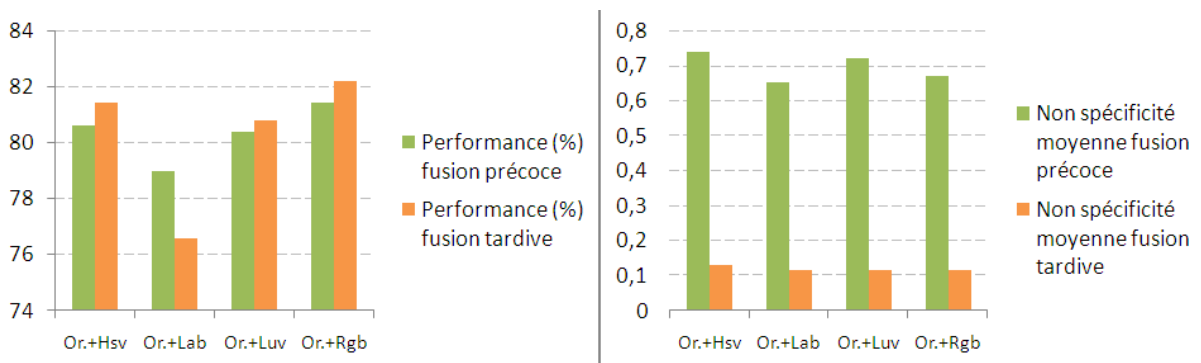


FIGURE 6.10 : Comparaison des approches de fusion tardive et précoce : à gauche sont données les performances de classification avec la distance de Bhattacharyya pour $f = 0,7$ et $k = 5$ sur le jeu de test *gt1000* pour toutes les combinaisons des descripteurs couleurs avec le descripteur des orientations. A droite sont indiquées les non-spécificités moyennes des distributions de masses associées.

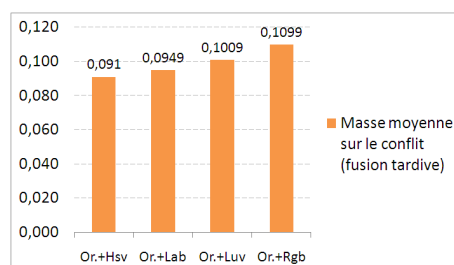


FIGURE 6.11 : Mesure du conflit moyen des distributions de masses avec la fusion tardive sur les mêmes tests que la figure 6.10. Il n'est pas possible d'avoir ce type d'information dans le cas de la fusion précoce.

2.3. Influence du paramètre f dans le cas de la fusion tardive

Les précédents tests ont révélé que le paramètre f peut influencer fortement les performances de classification, puisqu'il contrôle indirectement la non-spécificité des distributions de masses. C'est donc un paramètre critique dans notre approche et il est important de voir s'il existe une plage de réglage, de préférence sans minimum locaux, permettant d'optimiser les performances de classification.

La figure 6.12 présente les performances de classification en fonction de f pour les 2 meilleures métriques testées pour les fusions tardives des descripteurs orientations et $\{R,g,b\}$, toujours sur la base $gt1000$.

Sur le premier graphique, nous pouvons voir qu'il existe bien une plage de valeurs de f permettant des performances de classification optimales pour les 2 métriques comme nous l'avons vue précédemment dans le cas de descripteurs pris individuellement. Les performances de classification sont globalement plus élevées pour la distance de Bhattacharyya. De plus, l'utilisation de cette distance permet d'avoir une plage de valeurs de f optimale entre 0,6 et 0,8 garantissant plus de 80% des images correctement classées. Si ce paramètre doit être réglé manuellement, il est donc moins risqué d'avoir de mauvaises performances de classification avec la distance de Bhattacharyya, qu'avec la distance L_1 .

Le deuxième graphique mesure le conflit moyen des distributions de masses des images en fonction de f . Nous pouvons voir que les courbes de conflit suivent la tendance des courbes de performances : plus le pourcentage de classification correcte est élevé, plus le conflit moyen l'est également. Le conflit moyen atteint son maximum approximativement où f est optimale pour les performances de classification. Une étude théorique doit être réalisée pour analyser ce phénomène et pour voir s'il peut être exploité pour proposer une adaptation automatique du paramètre f .

2.4. Optimisation du calcul des distributions de masses

A terme, l'outil doit pouvoir fonctionner sur des machines standards. Or, l'approche proposée, de par son lien avec le formalisme des Croyances Transférables, possède un caractère combinatoire, notamment pour les distributions de masses, ce qui peut poser d'importants problèmes de temps de calculs. Nous proposons ici de trouver un compromis entre le temps de calculs et la finesse de la connaissance modélisée à travers une limitation de la taille des distributions de masses.

La limitation des calculs peut se faire au niveau des distributions des masses. Une distribution de masses m^Ω d'une image non étiquetées décrivant tous les états de connaissances contient théoriquement 3^Q propositions, dû à l'utilisation de l'opération d'extension vide (voir chapitre 3 page 58), plus une proposition de conflit dans le cas où plusieurs descripteurs sont fusionnés. En pratique, il y a beaucoup moins d'éléments focaux, i.e. de propositions avec de la masse non nulle, à cause de la contrainte de consonance imposée lors de la construction de la fonction de croyance. Nous pouvons encore limiter ce nombre d'éléments focaux en reportant toutes les masses sur des propositions de cardinalité élevée sur

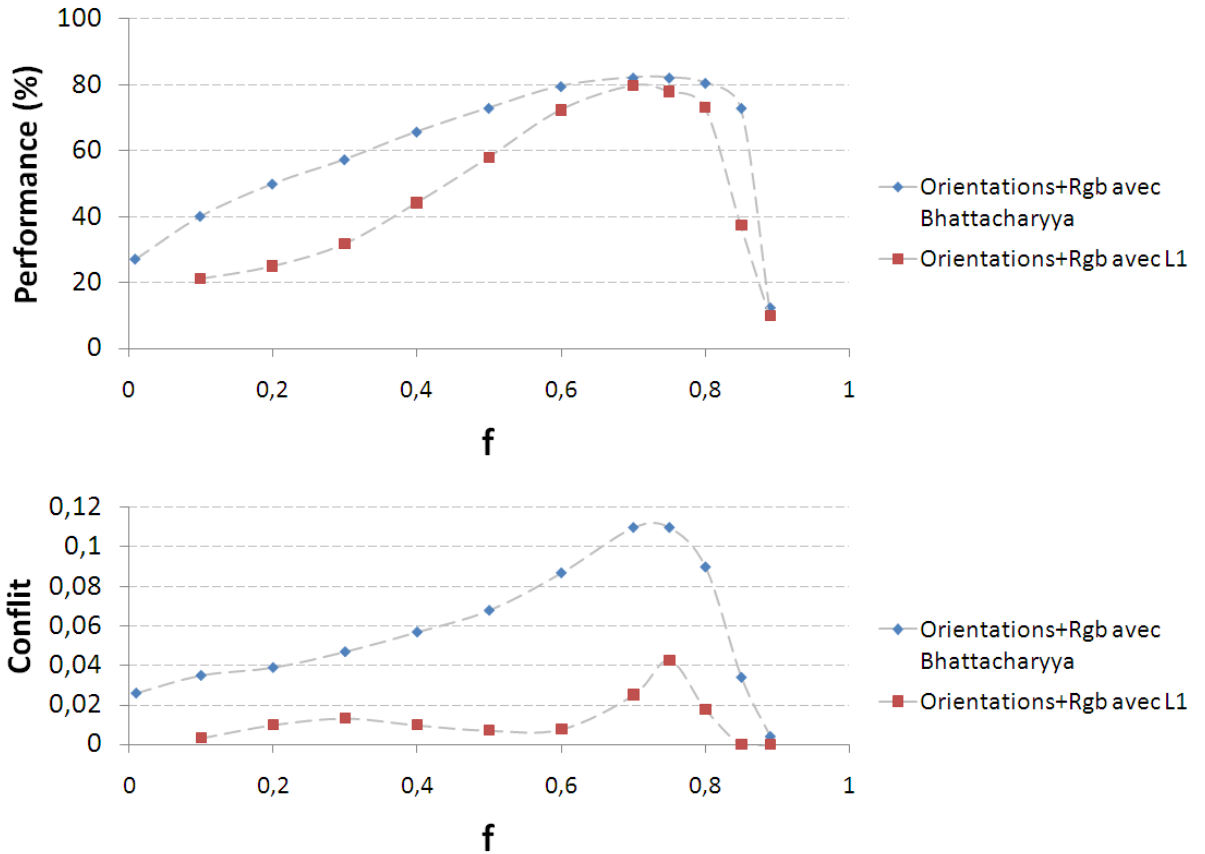


FIGURE 6.12 : Performances de classification et conflit moyen en fonction du paramètre f pour les distances de Bhattacharyya et L_1 sur les mêmes fusions de descripteurs orientations et $\{R,g,b\}$ sur la base $gt1000$.

la proposition de doute globale.

En effet, la prédiction d'étiquette d'une image repose sur la distribution de probabilités pignistiques. Or le calcul d'une probabilité pignistique d'une hypothèse du cadre de discernement Ω consiste à faire la somme de toutes les masses des propositions qui contiennent cette hypothèse en les pondérant par leur cardinalité. En conséquence plus la cardinalité d'une proposition est élevée, moins sa contribution à une probabilité pignistique est importante. Nous pouvons alors éviter le calcul des masses des propositions de cardinalité élevée en transférant leurs masses sur la proposition de doute global, ce qui permet de conserver la masse totale.

Exemple à 3 classes

Trois cadres de discernement disjoints Ω_1 , Ω_2 et Ω_3 sont combinés et définissent un nouveau cadre de discernement Ω :

$$\begin{aligned}
 \Omega &= \Omega_1 \times \Omega_2 \times \Omega_3 \\
 &= \{(H_1, H_2, H_3), (H_1, H_2, \overline{H_3}), (H_1, \overline{H_2}, H_3), (H_1, \overline{H_2}, \overline{H_3}), \\
 &\quad (\overline{H_1}, H_2, H_3), (\overline{H_1}, H_2, \overline{H_3}), (\overline{H_1}, \overline{H_2}, H_3), (\overline{H_1}, \overline{H_2}, \overline{H_3})\}
 \end{aligned} \tag{6.8}$$

Par exemple, à partir de la distribution de masses m^Ω d'une image non étiquetée u , la probabilité pignistique de l'hypothèse $(H_1, \overline{H_2}, H_3)$ se calcul grâce à l'expression :

$$\begin{aligned}
 \text{Bet}P\{m_u^\Omega\}(H_1, \overline{H_2}, H_3) = & \frac{1}{1 - m_u^\Omega(\emptyset)} \left(m_u^\Omega(H_1, \overline{H_2}, H_3) \right. \\
 & + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, H_3)}{2} + \frac{m_u^\Omega(H_1, \Omega_2, H_3)}{2} + \frac{m_u^\Omega(H_1, \overline{H_2}, \Omega_3)}{2} \\
 & + \frac{m_u^\Omega(\Omega_1, \Omega_2, H_3)}{4} + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, \Omega_3)}{4} + \frac{m_u^\Omega(H_1, \Omega_2, \Omega_3)}{4} \\
 & \left. + \frac{m_u^\Omega(\Omega_1, \Omega_2, \Omega_3)}{8} \right) \quad (6.9)
 \end{aligned}$$

Il est possible de limiter le nombre de masses calculées dans la distribution de masses m^Ω en ne prenant pas en compte les propositions de cardinalité élevée. Le tableau 6.1 donne pour 3 valeurs de cardinalité maximale autorisée l'expression de la probabilité pignistique de l'hypothèse $(H_1, \overline{H_2}, H_3)$. Cette contrainte sur les cardinalités des propositions permet de contrôler la finesse de la modélisation de la connaissance. Dans ce exemple à 3 classes, l'expression de la probabilité pignistique prend en compte des propositions de cardinalité 1, 2, 4 et 8 pour le doute global. Si la cardinalité maximum des propositions est limité à 4 (hors doute global) on retrouve l'expression entière de la probabilité pignistique. Si on limite cette cardinalité à 2, les masses sur les propositions $(\Omega_1, \Omega_2, H_3)$, $(\Omega_1, \overline{H_2}, \Omega_3)$ et $(H_1, \Omega_2, \Omega_3)$ ne sont pas calculées et le doute global $(\Omega_1, \Omega_2, \Omega_3)$ absorbe leurs masses. Dans le cas extrême, si la cardinalité maximum est limitée à 1, toute la distribution de masses est calculée exclusivement pour les propositions de cardinalité 1, et tout le reste de la masse est transféré sur le doute global. En conséquence, une probabilité pignistique est calculée uniquement à partir de la masse de la proposition de cardinalité contenant l'hypothèse et à partir du doute global.

Cardinalité maximale (hors doute global)	Nb masses m_u^Ω	$\text{Bet}P\{m_u^\Omega\}(H_1, \overline{H_2}, H_3)$
4	27	$ \frac{1}{1 - m_u^\Omega(\emptyset)} \left(m_u^\Omega(H_1, \overline{H_2}, H_3) + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, H_3)}{2} + \frac{m_u^\Omega(H_1, \Omega_2, H_3)}{2} + \frac{m_u^\Omega(H_1, \overline{H_2}, \Omega_3)}{2} + \frac{m_u^\Omega(\Omega_1, \Omega_2, H_3)}{4} + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, \Omega_3)}{4} + \frac{m_u^\Omega(H_1, \Omega_2, \Omega_3)}{4} + \frac{m_u^\Omega(\Omega_1, \Omega_2, \Omega_3)}{8} \right) $
2	21	$ \frac{1}{1 - m_u^\Omega(\emptyset)} \left(m_u^\Omega(H_1, \overline{H_2}, H_3) + \frac{m_u^\Omega(\Omega_1, \overline{H_2}, H_3)}{2} + \frac{m_u^\Omega(H_1, \Omega_2, H_3)}{2} + \frac{m_u^\Omega(H_1, \overline{H_2}, \Omega_3)}{2} + \frac{m_u^\Omega(\Omega_1, \Omega_2, \Omega_3)}{8} \right) $
1	9	$ \frac{1}{1 - m_u^\Omega(\emptyset)} \left(m_u^\Omega(H_1, \overline{H_2}, H_3) + \frac{m_u^\Omega(\Omega_1, \Omega_2, \Omega_3)}{8} \right) $

TABLE 6.1 : Limitation du nombre de propositions d'une distribution de masses en fonction de la cardinalité maximum autorisée.

Nous pouvons voir l'influence de cette limitation sur les cardinalités maximum des distributions de masses sur le jeu de test *gt1000* sur les graphiques de la figure 6.13. Ces tests ont été effectués en se

basant sur la distance L_1 entre histogrammes couleurs dans l'espace $\{L,u,v\}$. Les paramètres internes du système sont $k = 5$ plus proches voisins considérés par classe, et $f = 0,75$ le facteur d'étalement des fonctions de croyance. Sachant qu'il y a 10 classes, une distribution de masses complète peut contenir des propositions ayant une cardinalité jusqu'à $2^{10} = 1024$ pour la proposition de doute global.

Les tests sont effectués en limitant la cardinalité maximum à $2^0, 2^1, 2^2, 2^3, 2^4$, soit 1, 2, 4, 8 et 16. Nous pouvons voir que les performances de classification automatique saturent rapidement à environ 80% d'étiquettes correctement prédites dès une cardinalité maximum de 4. Ce résultat est d'autant plus important que le gain en temps de calculs et en place mémoire en termes de masses calculées est très élevé : pour des performances équivalentes vis-à-vis d'une limitation de cardinalité maximum de 16, on diminue par 2,6 le nombre de masses à calculer, et par 5 les temps de calculs.

De plus, le dernier graphique donne la non-spécificité moyenne des distributions de masses des 500 images de tests. Moins la cardinalité maximum autorisée est élevée, plus on transfère de la masse sur la proposition de doute global. En conséquence, plus la non spécificité est élevée, plus la distribution de probabilités pignistiques tend vers le cas d'équiprobabilité des hypothèses du cadre de discernement, ce qui implique qu'il est plus difficile d'avoir un engagement pour une des hypothèses en particulier.

Les performances de classifications sont intimement liées à la non-spécificité moyenne des distributions de masses des images de test : plus on autorise les propositions de cardinalités élevés, plus les distributions de masses sont spécifiques, en tendant rapidement vers une valeur de non spécificité moyenne autour de 1,4. Pour une cardinalité maximum de 4, la non spécificité moyenne vaut 1,8, ce qui est suffisant pour atteindre des performances de classification équivalente à des distributions de masses plus spécifiques.

2.5. Caractérisations et mesures sur une base de données multi-étiquetée

La formalisation avec le Modèle des Croyances Transférables nous a permis de mettre en place un système générique permettant d'envisager différents problèmes de classification :

- la classification bi-classe (classements de manière exclusive en 2 classes),
- la classification multi-classe (classements de manière exclusive en plus de 2 classes),
- et la classification multi-classe et multi-étiquette (classements de manière non exclusive en plus de 2 classes).

Ce paragraphe se focalise sur le dernier problème de classification multi-classe et multi-étiquette. Ce type de classification peut être considéré comme étant le plus difficile, puisqu'il s'agit d'arriver à proposer pour chaque image non étiquetée un ensemble d'étiquettes pertinentes. Le problème est alors d'arriver à limiter à la fois le nombre d'étiquettes proposées en trop car fausses et limiter le nombre d'étiquettes manquantes.

Conditions d'expérimentation :

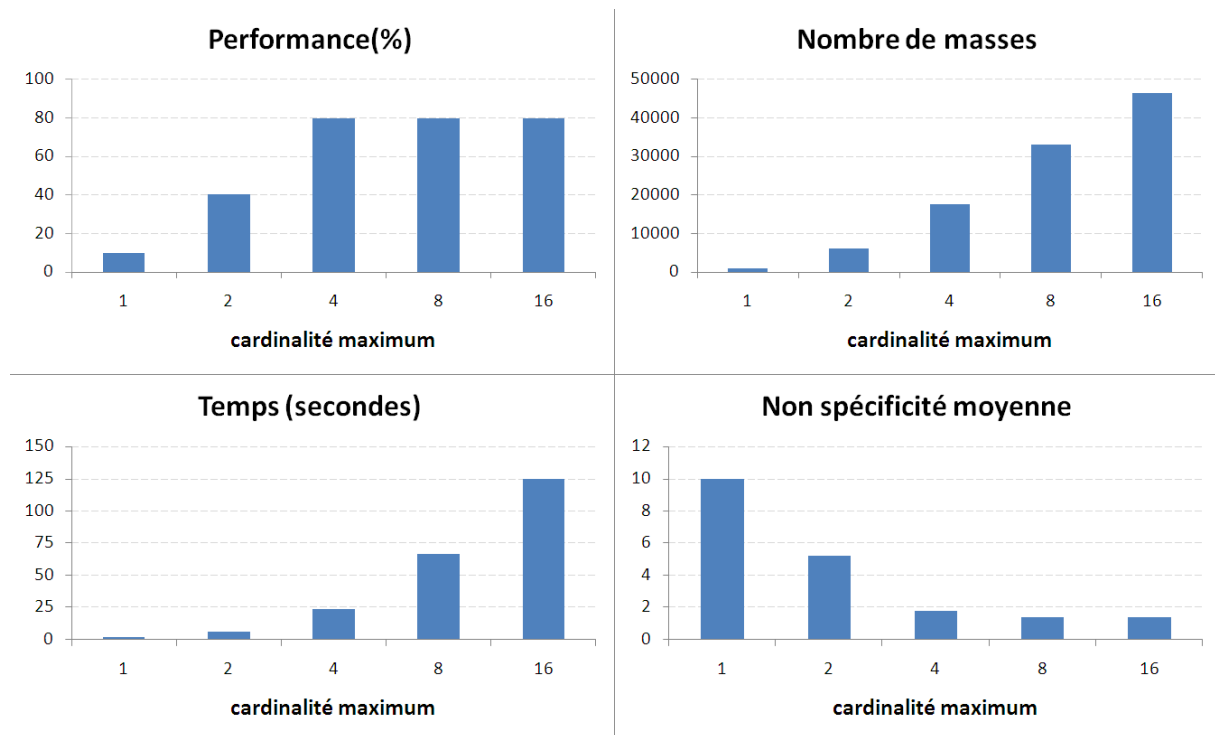


FIGURE 6.13 : Influence de la limitation du nombre de propositions d'une distribution de masses par la cardinalité maximum autorisée, en termes de performances de classification automatique et de temps de calculs. A titre indicatif, sont représentés le nombre de propositions dans la distribution de masses, et la non-spécificité moyenne de toutes les distributions de masses de toutes les images non étiquetées.

Le jeu de données "scene-classification", disponible sur le site officiel de LibSVM [CL01], est utilisé pour cette évaluation. Ce jeu de test a été proposé par Boutell dans [BLSB04] pour comparer les performances des systèmes de classification multi-étiquette.

Les images ne sont pas fournies. D'après le papier de référence [BLSB04], les images sont issues de la banque d'images Corel où l'auteur a identifié 6 étiquettes correspondant aux concepts ("urban", "sunset", "fall foliage", "field", "mountain" et "beach"). Les images sont associées en moyenne 1,08 étiquettes.

Un fichier texte décrit 2407 images associées chacune à une ligne désignant des étiquettes de classes (entre 1 et 6), et un vecteur descripteur. Un vecteur descripteur correspond à la concaténation de moyennes et de variances d'histogrammes locaux de $7 \times 7 = 49$ blocs réguliers, sur les 3 composantes couleurs $\{L,u,v\}$, soit $2 \times 3 \times 49 = 294$ dimensions.

Le test consiste alors à entraîner le système sur 1211 images exemples, puis à prédire les étiquettes des 1196 images restantes.

Formalisation :

Un cadre de discernement Ω est défini et contient de 2^6 hypothèses, toutes les combinaisons d'hypothèses de bases H_q et \overline{H}_q décrivant les états d'appartenance des images non étiquetées à une classe C_q ($q \in$

$\{1, 2, \dots, 6\}$). Nous ne savons pas combien d'étiquettes possède une image au maximum et il n'est donc pas possible de limiter le nombre d'hypothèses ambiguës dans le cadre de discernement Ω .

Mesures utilisées :

L'expérimentation consiste à mesurer les performances en termes d'étiquettes correctement prédites. Il s'agit ici d'un problème multi-étiquette et toutes les classes sont connues à l'avance. Pour pouvoir prédire une étiquette sur une image l'espace de décision Ω_M dédié à la classification multi-étiquette sans option de rejet en distance est donc utilisé (voir page 96). Cet espace de décision contient donc $2^6 - 1$ hypothèses, toutes les hypothèses du cadre de discernement Ω moins l'hypothèse de rejet $(\overline{H_1}, \overline{H_2}, \overline{H_3}, \overline{H_4}, \overline{H_5}, \overline{H_6})$.

Des mesures de rappel et de précision dans le cas spécifique du multi-étiquetage sont utilisées pour mesurer les performances de classification :

- Considérant D un jeu de test contenant $|D|$ échantillons multi-étiquetés $(x_i, Y_i), i = 1 \dots |D|, Y_i \subseteq L, L$ étant l'ensemble des d'étiquettes disponibles.
- Considérant H un classifieur multi-étiquette prédisant des ensembles d'étiquettes $Z_i = H(x_i)$ pour chaque échantillon à tester x_i .

Rappel : rapport entre le nombre d'étiquettes correctement prédites sur le nombre d'étiquettes à retrouver pour tous les échantillons testés :

$$Rappel(H, D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Y_i|} \quad (6.10)$$

Précision : rapport entre le nombre d'étiquettes correctement prédites sur le nombre d'étiquettes prédites au total pour tous les échantillons testés :

$$Precision(H, D) = \frac{1}{|D|} \sum_{i=1}^{|D|} \frac{|Y_i \cap Z_i|}{|Z_i|} \quad (6.11)$$

Résultats :

Dans nos expérimentations, le meilleur compromis entre rappel et précision est obtenu pour pour un nombre $k = 5$ de plus proches voisins, et une valeur $f = 0,78$ avec la distance L_1 :

$$\begin{aligned} Rappel &= 0,745 \\ Precision &= 0,714 \end{aligned} \quad (6.12)$$

La figure 6.14 compare nos résultats obtenus avec différents tests issus de travaux dans [TK07]. Dans ce travail les auteurs donnent les résultats pour 12 méthodes de classifications sur le même jeu de données. Plus exactement, les auteurs comparent 4 algorithmes de classification (kNN [AKA91], C4.5 [Qui93], Bayes naïf [JL95] et SMO (à base de SVM [Pla99]) pour 3 schémas de transformations d'étiquettes notés

PT3, PT4 et PT6 (voir chapitre 3 page 31). Notre approche est similaire au schéma de transformation d'étiquettes PT3 car nous considérons les unions d'étiquettes simples à travers les hypothèses ambiguës du cadre de discernement.

Ces résultats placent la méthode que nous proposons comme étant l'une des plus performantes avec la méthode "PT3+SMO" : notre approche permet d'obtenir plus de précision, alors que la méthode "PT3+SMO" permet plus de rappel (l'approche "PT3+SMO" donne une précision de 0,713 et un rappel de 0,737). Ce résultat valide la pertinence de notre approche sur un problème difficile de classification.

Il est également intéressant de noter, selon ces évaluations, que les performances des approches à base de knn sont plutôt performantes indépendamment du choix de transformations d'étiquettes. Par ailleurs les auteurs font remarquer dans [TK07] que l'approche PT3 est relativement peu utilisée vis-à-vis de l'approche classique PT4, malgré sa pertinence.

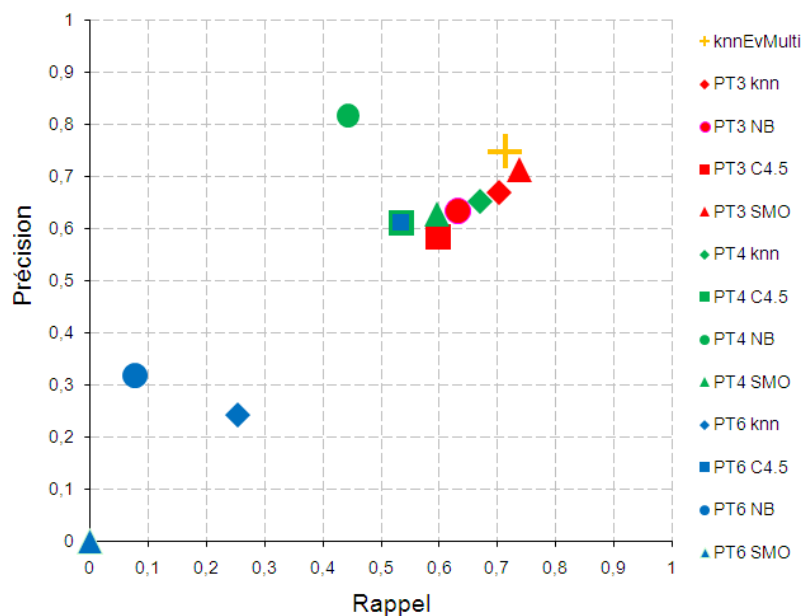


FIGURE 6.14 : Comparaison des différentes méthodes de classification multi-classe et multi-étiquette évaluées dans [TK07]. Notre méthode notée "KnnEvMulti" est représentée par la croix orange.

Bilan des expérimentations sur les performances de classification

Dans cette partie, des performances de classification automatique ont été évaluées. Il a été constaté que les performances dépendent en grande partie du choix de la métrique et du paramètre f contrôlant l'ambiguïté entre les classes. Le paramètre f permet des performances de classification élevée sur une plage de valeurs relativement large. Le choix du descripteur couleur s'est révélé moins important, notamment pour la distance de Bhattacharyya. Le paramètre k peut être fixé à une valeur de l'ordre de 5 plus proches voisins. Il est possible également d'optimiser le coût de calculs en limitant la taille des distributions de masses tout en gardant des performances élevées. Par la partie suivante, ces observations donnent des

repères pour paramétrer notre système par défaut, dans le cadre de la sélection active.

3. Caractérisations des stratégies de sélection active d'images

Dans cette partie, nous étudions l'impact des stratégies de sélection active des images que nous avons proposées dans le chapitre 4. Le but est de caractériser l'apport ou non des stratégies en termes de performances de classification. En effet, nous avons formalisé dans le même cadre les stratégies de sélection les plus courantes en apprentissage actif. Nous pouvons alors voir quelle stratégie est la plus intéressante non seulement en termes de performances de classification pure, mais également vis-à-vis d'une perspective d'une aide pour l'utilisateur.

Dans un premier temps, nous effectuons une évaluation classique de l'apprentissage actif, puis nous proposons de comparer les stratégies en se plaçant dans le cas où une collection d'images est quasiment vierge de toute étiquette.

3.1. *Evaluation classique des stratégies de sélection active*

Le but initial d'une stratégie de sélection active d'échantillons est de faire étiqueter par un utilisateur les échantillons les plus "utiles" pour améliorer un algorithme de classification.

Nous avons formalisé 6 stratégies de sélection active d'échantillons (voir chapitre 4) : le plus positif (MP), le plus rejeté (MR), le plus globalement ambigu (MGA), le plus localement ambigu à 2 classes (MLA2), le plus conflictuel (MC) et le plus incertain (MU). Pour comparer ces stratégies, une évaluation classique en apprentissage actif consiste à comparer les performances de classification par une stratégie de sélection, avant, puis après l'ajout d'échantillons supplémentaires.

Conditions d'expérimentation :

Le même jeu de test *gt1000* qui a servi dans la première partie de ce chapitre sur les performances de classification automatique est utilisé ici. On reprend le même découpage en 2 parties et on dispose ainsi d'un jeu d'apprentissage de 10x50 images exemples, et d'un jeu de test 10x50 images.

L'apprentissage se déroule en plusieurs étapes :

1. On apprend de façon automatique les 500 images du jeu d'apprentissage, de la même manière que dans la partie 1 de ce chapitre.
2. On mesure les performances de cette classification automatique sur les 500 images de tests. On dispose ainsi un pourcentage de référence. Puis, on détermine dans quelles mesures l'ajout de nouvelles images étiquetées dans la base d'apprentissage améliore ou non les performances de classification.
3. On sélectionne suivant la stratégie choisie 50 images issues du jeu de test et l'on simule l'éti-

quetage que ferait un utilisateur grâce à la vérité terrain. Pour rappel, les 50 images ne sont pas sélectionnées en une seule fois, mais par un critère propre à la stratégie étudiée une après l'autre, en révisant toute la connaissance après chaque étiquetage d'une image.

4. Il reste donc $500 - 50 = 450$ images dans le jeu de test, et la base d'apprentissage en contient maintenant 550. On mesure à nouveau les performances de classification automatique sur ces 450 images de tests, que l'on peut comparer avec les performances initiales, pour constater si la stratégie de sélection active a permis de faire progresser l'algorithme de classification.

Formalisation :

Nous considérons le cadre de discernement Ω constitué en théorie de 2^{10} hypothèses, c'est-à-dire toutes les combinaisons d'hypothèses de bases H_q et \overline{H}_q décrivant les états d'appartenance des images non étiquetées à une classe C_q ($q \in \{1, 2, \dots, 10\}$).

Il s'agit ici d'un problème multi-classe avec étiquetage simple, c'est-à-dire qu'une image ne peut appartenir à plusieurs classes à la fois. De plus, toutes les classes sont connues à l'avance. L'espace de décision Ω_S dédié à la classification d'images de manière exclusive, sans option de rejet en distance et en ambiguïté est donc utilisé (voir 96).

Choix des descripteurs et de la métrique :

Dans la première partie de ce chapitre, il a été observé que certaines métriques appliquées sur certains descripteurs fournissent des performances de classification automatique différentes. Nous souhaitons mettre en avant le mécanisme de sélection active. Si nous choisissons la combinaison de "métrique-descripteurs" les plus performants, les différences de performances sans et après apprentissage actif risquent de ne pas être très significatifs si les performances sont déjà bonnes. Il peut être plus intéressant pour cette expérimentation de choisir une combinaison "métrique-descripteurs" donnant des performances de classification intermédiaires afin de constater l'apport d'une stratégie vis-à-vis d'une autre. Les tests sont effectués en utilisant la distance L_2 pour les fusions des descripteurs des orientations et de couleurs dans l'espace $\{L,a,b\}$, avec $f = 0,8$ et $k = 5$.

Résultats :

Après l'entraînement sur les 500 images de la base d'apprentissage, la mesure de performance de classification de 72,4%, c'est-à-dire qu'en l'état de connaissance, le système est capable de classer correctement 72,4% des 500 autres images du jeu de test.

La figure 6.15 compare les résultats pour les 6 stratégies de sélection sur 50 images. Un pourcentage moyen d'images correctement classées est également donné pour une quelques tests avec une stratégie de sélection aléatoire d'échantillons.

Ces résultats illustrent bien le fait que les différentes stratégies font progresser plus ou moins efficacement les performances de classification. Les différents tests sur les stratégies aléatoires servent de référence : l'ajout de 50 échantillons sélectionnés aléatoirement n'améliore pas significativement les performances de classification. Malgré l'ajout de ces échantillons, le pourcentage de bonnes étiquettes reste à peu près

identique.

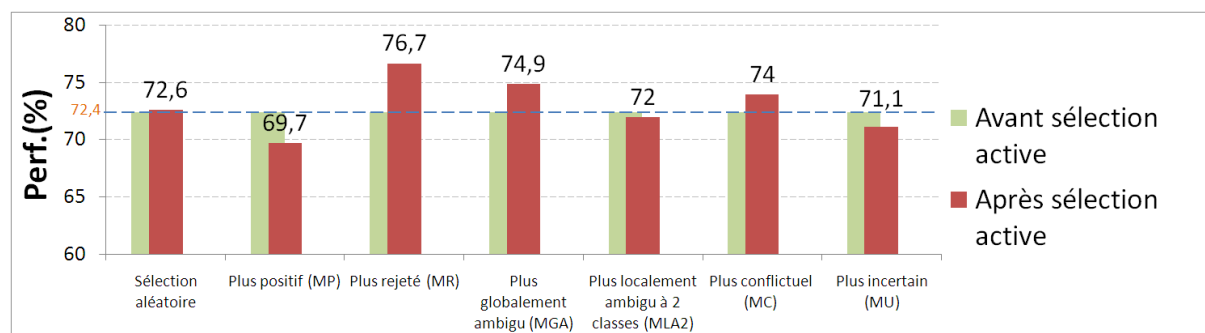


FIGURE 6.15 : Comparaison des performances de classification avant et après apprentissage actif sur le jeu de données gt1000. Avant la sélection active (en vert), 72,4% des images peuvent être classées correctement. Après la sélection active de 50 images supplémentaires (en rouge) les performances de classification sont modifiées : la stratégie aléatoire n'apporte pas de gain, les stratégies MP, MLA2 et détériore les performances, tandis que les autres l'augmentent. La stratégie la plus "rentable" est celle du plus rejeté MR.

La stratégie du plus positif (MP) détériore quant à elle les performances de classification. Ce résultat est attendu : en effet, ajouter les images les plus positives revient à sélectionner les plus probables pour une classes. Ce sont celles qui sont dans des portions des espaces de description connues, dans des zones où beaucoup d'images ont été déjà étiquetées. En sélectionnant les 50 plus positives, on laisse le classement des images les plus difficiles à classer. A la fin de l'apprentissage actif, le pourcentage d'images difficiles à classer est plus important ce qui explique le taux d'erreur.

Les autres stratégies sélectionnent les plus difficiles à classer mais selon leur propre critère. La stratégie de sélection des 50 plus rejetées entraîne le gain le plus important : après l'apprentissage actif avec cette stratégie, le système est capable de classer correctement 76,7% des 450 images restantes. En recherchant les images les plus rejetées, on explore des nouvelles portions des espaces de descriptions, où aucune image n'a été encore étiquetée. En étiquetant une image rejetée, les images non étiquetées de son voisinage ont tendance à devenir positives, ce qui augmente globalement le pourcentage de bonnes classifications.

La stratégie du plus localement ambiguë entre 2 classes (MLA2) n'apporte pas de gain significatif et détériore même légèrement le pourcentage initial de bonnes classifications. Ce résultat peut s'expliquer par le fait que les classes se recouvrent dans les espaces de description. Les frontières sont alors très complexes, et une image encore non étiquetée très localement ambiguë est peut être seule dans une zone isolée. En l'étiquetant, on risque de ne résoudre qu'un cas particulier et il n'est pas garanti qu'elle aide à mieux classer par la suite les images non étiquetées restantes comme tente de l'illustrer la figure 6.16.

A l'inverse, l'utilisation de la stratégie du plus globalement ambigu (MGA) permet de passer d'un pourcentage de bonnes classifications de 72,4% à 74,9%. Ce gain peut s'expliquer par le fait que ces échantillons sont plutôt situés dans des zones "encerclées" d'images membres de différentes classes. Elles ne sont pas suffisamment proches des membres des classes pour être positives, mais pas suffisamment éloignées pour être rejetées.

La stratégie du plus conflictuel permet également un gain sur la prédiction des étiquettes restantes. Il est important de remarquer que cette performance est liée au paramètre f . Si f est important, comme c'est le cas ici, nous avons vu dans la première partie de ce chapitre, que le conflit augmente. A l'inverse si f est faible, la masse sur le conflit dans les distributions de classes tend vers 0. Il est alors peu probable que la sélection active soit efficace dans ce cas, puisque les comparaisons d'images non étiquetées pour en sélectionner une s'appuieront sur des valeurs peu significatives.

Enfin, la stratégie du plus incertain détériore les performances. Tout comme pour le conflit, cette stratégie s'appuie sur une mesure des distributions de masses. Dans le cas présent, de fusion de descripteurs, les distributions de masses tendent à être très spécifiques. En conséquence les mesures de non-spécificité tendent à être très faibles. Or la stratégie du plus incertain se base sur la sélection de l'image ayant la distribution de masses de la plus grande non spécificité. La sélection active n'est donc pas efficace, peut être parce que la sélection s'appuie sur des valeurs trop faibles, ne permettant pas de distinguer entre elles les classes potentielles d'une image.

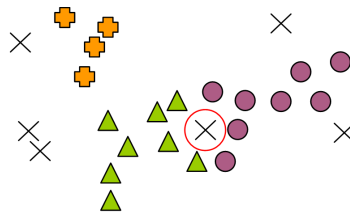


FIGURE 6.16 : Cas d'une sélection d'un échantillon très localement ambiguë entre 2 classes (croix entourée en rouge). Cet échantillon est sélectionné pour être étiqueté par l'utilisateur, soit dans la classe "rond" ou la classe "triangle" (voir même la classe "croix pleine"). Or, cet échantillon est isolé et n'apporte que très peu d'information "utile" pour classer les autres échantillons non étiquetés (représenté également par des croix).

Ce type d'expérimentations met en avant la façon dont les stratégies peuvent faire progresser ou non une méthode de classification. Cependant, ce type d'évaluation ne répond pas entièrement à notre cahier des charges, car ici nous partons du fait que 500 images sont déjà étiquetées. Autrement dit, la connaissance des classes est déjà bien établie car elles possèdent de nombreux exemples. Or, un utilisateur souhaitant organiser une collection ne possède pas, dans la majorité, des cas toute cette connaissance des classes, et nous devons nous mettre dans une situation où très peu voire aucune images n'ont encore été étiquetées.

3.2. Caractérisation des stratégies dans le contexte applicatif

3.2.1. Courbes d'évaluation

Dans cette partie, les stratégies sont évaluées en se rapprochant des conditions dans lesquelles un utilisateur travaille. Le système a été conçu pour pouvoir aider un utilisateur à structurer une collection d'images depuis le début, c'est-à-dire vierge de toute étiquette. On ne dispose donc pas d'une base d'ap-

prentissage. Les stratégies de sélection ont pour objectif de sélectionner les images à faire étiqueter par l'utilisateur et compléter la base d'apprentissage à la volée.

Conditions d'expérimentation :

Un jeu de test *gt500* est utilisé pour faire ces tests. Celui-ci contient 5 classes de 100 images, chacune issues du jeu de test *gt1000* utilisé dans la première partie de ce chapitre. Les 5 classes ("fleurs", "chevaux", "dinosaures", "bus" et "plage") sont volontairement choisies comme étant les plus homogènes visuellement afin de bien mettre en évidence le comportement de sélection des stratégies.

Pour caractériser les stratégies, on suppose que les 5 classes ont déjà été identifiées, c'est-à-dire que chaque classe est initialisée par une seule image exemple. Il reste alors 495 images à classer une par une en utilisant une stratégie.

Formalisation :

Nous considérons le cadre de discernement Ω constitué de 2^5 hypothèses, c'est-à-dire toutes les combinaisons d'hypothèses de bases H_q et \overline{H}_q décrivant les états d'appartenance des images non étiquetées à une classe C_q ($q \in \{1, 2, 3, 4, 5\}$).

Il s'agit ici d'un problème multi-classe avec étiquetage simple, c'est-à-dire qu'une image ne peut appartenir à plusieurs classes à la fois. De plus, toutes les classes sont connues à l'avance. L'espace de décision Ω_S dédié à la classification d'images de manière exclusive, sans option de rejet en distance et en ambiguïté est donc utilisé (voir page 96).

Courbes d'évaluation :

Les stratégies sont toutes testées individuellement : de manière itérative, en fonction de l'état de connaissance des classes, une image est sélectionnée en fonction de la stratégie choisie, puis une proposition d'étiquette est effectuée sur cette image. Si cette proposition ne correspond pas à la vérité terrain simulant le choix d'un utilisateur, on comptabilise le nombre d'étiquettes incorrectes accumulées, puis on ajoute cette image dans la base d'apprentissage avec la bonne étiquette.

Cette évaluation permet de tracer des courbes comme dans la figure 6.17 comparant la stratégie de sélection aléatoire avec celle du plus positif MP dans les mêmes conditions d'expérimentation ($k = 5$, $f = 0,6$ avec les descripteurs des orientations et $\{R,g,b\}$ et l'utilisation de la distance de Bhattacharyya). L'axe horizontal représente la chronologie des sélections successives d'images, et l'axe vertical correspond à l'accumulation des étiquettes incorrectes proposées par le système de façon automatique.

Ce type de courbe permet 2 échelles d'analyses complémentaires pour étudier les stratégies de sélection active, en termes de performances de classification. Cependant, il est aussi important de faire le lien avec l'effort de l'utilisateur que peut entraîner les stratégies et comment il peut les mettre à profit.

Les courbes peuvent se lire de 2 manières :

1. **Performance finale** : le dernier point de la courbe permet de se focaliser avant tout sur la per-

formance finale de classification qu’implique l’utilisation d’une stratégie. Dans cet exemple, les stratégies "aléatoire" et MP donnent au final respectivement un nombre de 37 et de 30 propositions d’étiquettes incorrectes. Cela veut dire que 92,52% des images ont été proposées dans les bonnes classes avec la stratégie aléatoire, contre 93,94% pour la stratégie du plus positif.

2. **Performance globale :** en observant les courbes dans leur intégralité, il est possible de voir à quel moment les mauvaises propositions d’étiquettes se produisent. Les 2 stratégies ne concentrent pas les mauvaises propositions aux mêmes moments. La stratégie de sélection aléatoire provoque des propositions d’étiquettes incorrectes régulièrement, alors que la stratégie MP en fait très peu sur les 3 premiers quarts. Par contre, la stratégie MP provoque de plus en plus de mauvaises propositions. Cette accélération finale peut s’expliquer par le fait que la stratégie du plus positif MP "retarde" le classement des images les moins probables, au sens des probabilités pignistiques, d’appartenir à une classe.

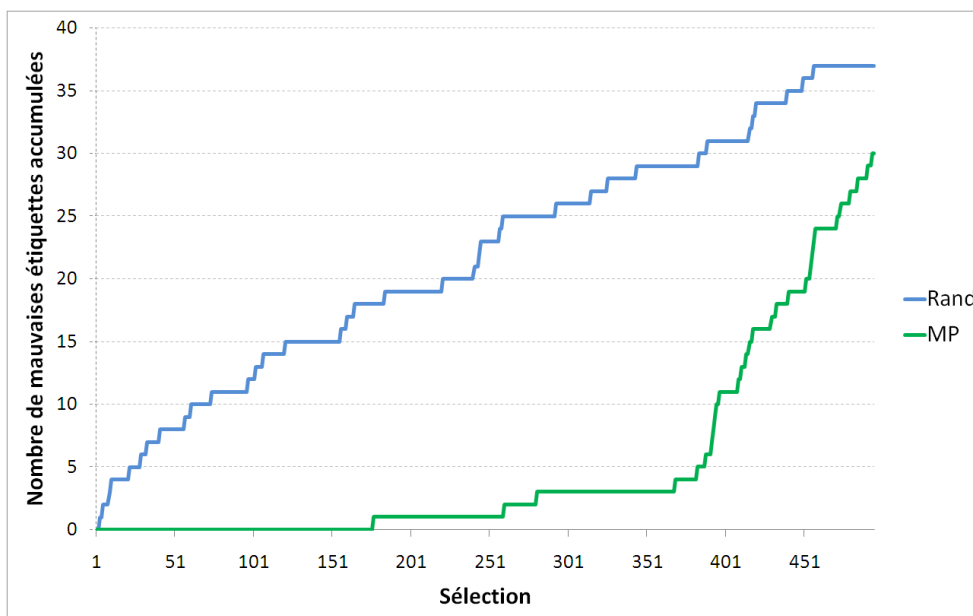


FIGURE 6.17 : Courbes d’évaluations comparant la sélection aléatoire (Rand) avec la stratégie du plus positif (MP). Les 2 stratégies ne provoquent pas les mauvaises propositions d’étiquettes aux mêmes instants de sélection et ne donnent pas le même nombre final d’images proposées de manière incorrectes dans les classes.

3.2.2. Liens avec l’effort de l’utilisateur

Nous pouvons compléter la figure précédente, en traçant l’ensemble des courbes pour toutes les stratégies (voir figure 6.18) dans les mêmes conditions d’expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l’espace $\{R,g,b\}$, $f = 0,6$ et $k = 5$). Ces courbes nous renseignent sur le nombre d’actions qu’aurait fait un utilisateur en utilisant la même stratégie pour classer les images, et à quels moments il les aurait faites. Les stratégies ne demandent donc pas le même type d’effort d’étiquetage au même instant pour l’utilisateur. Dans cette expérience, les stratégies peuvent être regroupées en 2

groupes :

- Les stratégies du plus positif MP, du plus conflictuel MC et du plus incertain MU permettent de retarder les mauvaises propositions d'étiquettes. En particulier, la stratégie MP ne provoque aucune mauvaise proposition sur le premier tiers des images sélectionnées. Ce type de comportement permet de mettre en confiance l'utilisateur sur la capacité du système à classer correctement les images dans les bonnes classes. Cependant, ces stratégies peuvent à la longue provoquer une certaine monotonie lassant l'utilisateur si la suite des images possède des contenus visuels très similaires. D'autre part, à la fin du classement des images, le nombre de mauvaises propositions augmente ce qui peut surprendre l'utilisateur et raviver son attention au moment où il est le plus fatigué.
- Les stratégies du plus rejeté MR, du plus globalement et localement ambiguë MGA et MLA2 permettent de concentrer un maximum de mauvaises propositions d'étiquettes dès le début du classement d'images. En particulier la stratégie MR donne le plus faible nombre de mauvaises propositions de classes. Ce résultat final est d'autant plus intéressant, que la stratégie MR permet de ne plus faire d'erreurs de classification après le premier tiers des sélections. L'apprentissage est complètement réalisé dès la 170^{ième} sélection (sur 495 en tout). Par contre, cette stratégie MR demande beaucoup d'attention à l'utilisateur dès les premières sélections sur des images difficiles. Les stratégies en ambiguïté reproduisent la même tendance que la stratégie MR, mais avec moins d'efficacité. En particulier, il subsiste quelques dernières erreurs de classification parmi les dernières sélections.

Remarque : ces courbes renseignent sur les actions faites par un utilisateur dans le temps. Mais elles ne reproduisent pas fidèlement le temps. Par exemple, si la stratégie MR provoque en peu de sélections la fin de l'apprentissage, il n'est pas garanti que l'utilisateur traite ces premières images rapidement. En effet, ces images rejetées sont difficiles à classer puisqu'elles ont des contenus visuels très éloignés des images classées. En conséquence, ces images demanderont certainement un temps de réflexion plus important.

Ces courbes donnent les premières tendances que l'on peut attendre des stratégies. Or, nous avons vu dans la première partie de ce chapitre que les métriques et le paramètre f influencent fortement les performances de classifications automatiques. Il est donc raisonnable de penser qu'ils peuvent influencer également le comportement des stratégies.

3.3. Influence du paramètre f

Nous souhaitons voir ici dans quelles mesures les stratégies reproduisent ou non les mêmes tendances observées dans le paragraphe précédent en changeant le paramètre f . En effet, nous avons vu dans la première partie de ce chapitre que ce paramètre f , contrôlant l'ambiguïté entre les classes, peut influencer fortement les performances de classification automatique.

Le tableau 6.2 reprend les comparaisons de stratégies sur le même jeu de données *gt500* qu'au paragraphe précédent, pour 4 valeurs distinctes de f ($f = 0,4$, $f = 0,6$, $f = 0,7$ et $f = 0,8$), avec les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace {R,g,b} et

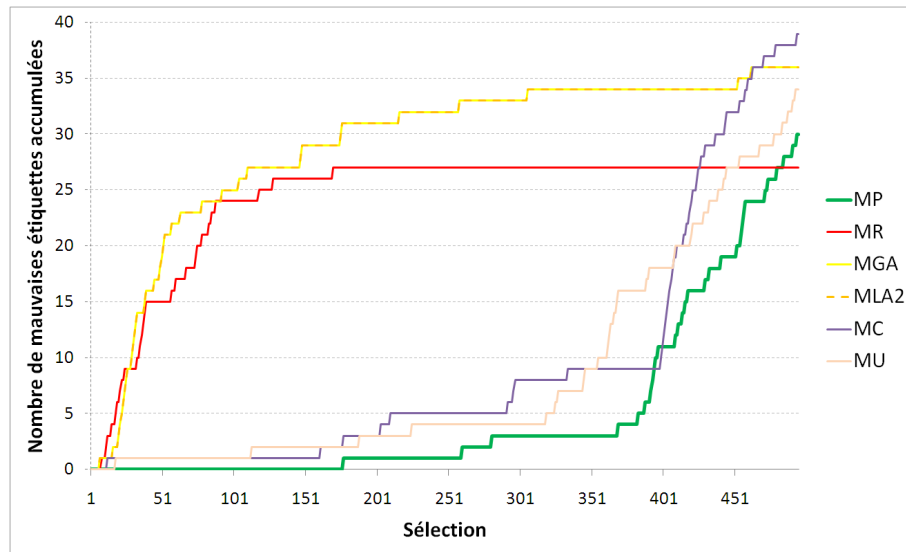


FIGURE 6.18 : Comparaisons de toutes les stratégies sur le jeu de données *gt500* dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace $\{R,g,b\}$, $f = 0, 6$ et $k = 5$). Les différentes stratégies ne donnent pas toutes le même nombre de mauvaises propositions d'étiquettes au final, et ne provoquent pas ces propositions incorrectes aux mêmes moments.

$k = 5$).

Tendances :

Nous pouvons observer que les stratégies du plus positif MP, du plus rejeté MR et du plus globalement ambigu MGA ont tendance à reproduire les mêmes comportements, malgré de grandes variations des valeurs de f . La stratégie du plus positif MP conserve bien la caractéristique de retarder les sélections d'images sur lesquelles de mauvaises propositions d'étiquetage sont plus probables. Les stratégies du plus rejeté MR et du plus globalement ambigu MGA gardent également leur capacité de généralisation : elles permettent de sélectionner en priorité des images les plus difficiles à classer, et permettent de finir l'apprentissage rapidement.

Les stratégies du plus localement ambigu MLA2, du plus conflictuel MC et du plus incertain sont plus dépendant des valeurs de f . La stratégie du plus conflictuel MC est l'illustration la plus représentative de ce comportement. En effet, pour les valeurs $f = 0, 6$ et dans une moindre mesure $f = 0, 7$ la stratégie MC tend à reproduire la même tendance que la stratégie MP, en sélectionnant des images où le système risque de faire des mauvaises propositions d'étiquette à la fin. Par contre, pour $f = 0, 8$, la stratégie MC reproduit la même tendance que les stratégies MR et MGA en produisant un maximum de mauvaises propositions sur le début des sélections. Pour $f = 0, 4$ la stratégie MC à tendance à reproduire la même tendance que la stratégie aléatoire. En effet, les classes ne se chevauchent pas dans le cas où f est faible et le conflit est très faible. La sélection par la stratégie MC s'appuie donc sur des valeurs tendant vers 0, ce qui revient à sélectionner aléatoirement les échantillons.

Synthèse :

Certaines stratégies reproduisent les mêmes tendances, tandis que d'autres peuvent complètement changer. Il semble exister une valeur optimale de f pour chaque stratégie permettant d'atteindre un nombre minimal de mauvaises propositions d'étiquettes.

Une valeur faible de f privilégie la stratégie du plus positif MP en permettant de bien sélectionner les images les plus difficiles à classer sur les dernières sélections et en donnant un nombre minimal de mauvaises propositions.

La stratégie du plus rejeté MR permet à l'inverse de sélectionner les images les plus difficiles à classer dans les premières sélections pour la plupart des valeurs de f . Il semble y avoir une valeur optimale autour de $f = 0,7$ permettant de minimiser le nombre final de mauvaises propositions.

La stratégie du plus globalement ambigu MGA est plus intéressante sur des valeurs faibles de f , où il est possible de généraliser plus vite. Par contre, elle provoque plus de mauvaises propositions que la stratégie MR pour une même valeur de f .

Les stratégies du plus localement ambigu MLA2 et du plus conflictuel MC peuvent provoquer des comportements assez différents selon les valeurs de f . La stratégie MLA2 peut être éventuellement intéressante dans le cas où f est faible en reproduisant la même tendance que la stratégie du plus globalement ambigu. A l'inverse, la stratégie du plus conflictuel est plus intéressant pour une valeur élevée de f où l'on peut reproduire le même type de tendance qu'une stratégie MR.

La stratégie du plus incertain MU est intéressante pour des valeurs faibles de f , où elle reproduit la même tendance de progression lente que la stratégie du plus positif MP.

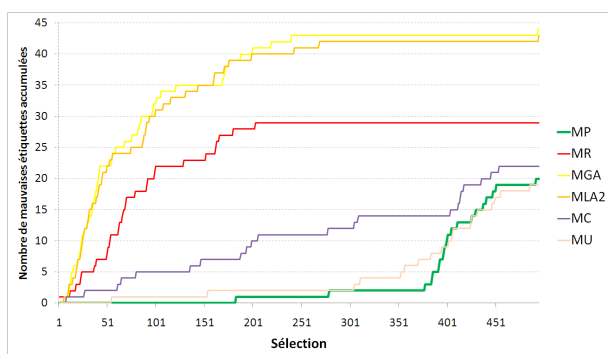
3.4. Sélection active dans le cas du multi-étiquetage

Une première évaluation a été réalisée précédemment dans le cas de classification automatique multi-étiquette, montrant que notre méthode est compétitive sur le benchmark de LibSvm d'après les travaux dans [TK07]). Nous souhaitons voir ici, comment se comportent les stratégies de sélection active dans le cadre du multi-étiquetage.

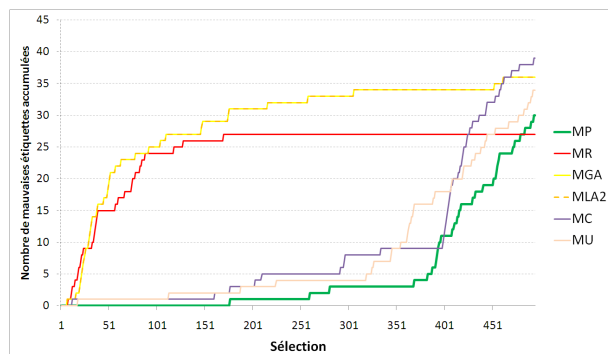
Conditions d'expérimentation :

Un utilisateur souhaite structurer une collection de photographies personnelles prises durant un séjour à l'étranger. Il a utilisé l'outil pour établir une vérité terrain d'images multi-étiquette $gt772$. Le résultat de son organisation a permis de faire émerger 7 étiquettes pouvant être associées aux mots-clés suivants : "intérieurs", "bâtiments", "révolution", "portraits", "paysage nature et végétation", "mer, plage et repos", "culture". La collection d'images contient 722 photographies, et au final l'utilisateur a identifié 835 étiquettes, ce qui donne une moyenne de 1,16 étiquettes par images.

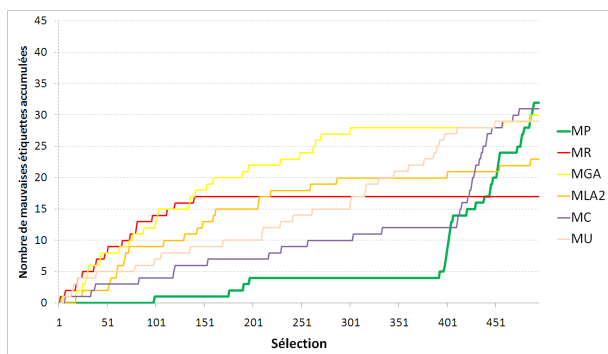
Ce jeu de donné est très difficile car les contenus peuvent être très divers au sein d'une même classe. Par exemple, la classe "culture" contient des images de danseurs, de masques, d'instruments de musique. ...



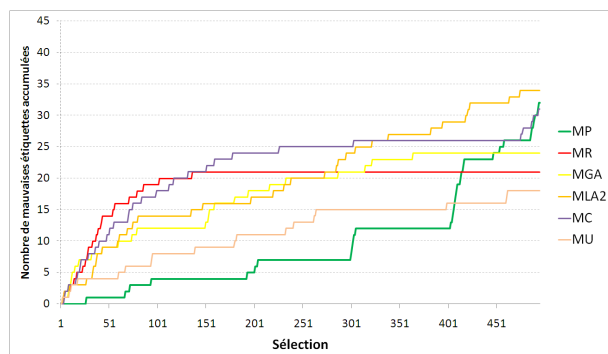
$f = 0,4$



$f = 0,6$



$f = 0,7$



$f = 0,8$

TABLE 6.2 : Comparaisons pour 4 valeurs distinctes de f ($f = 0,4$, $f = 0,6$, $f = 0,7$ et $f = 0,8$) de toutes les stratégies sur le jeu de données gt500 dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace $\{R,g,b\}$ et $k = 5$). Certaines stratégies reproduisent les mêmes tendances, tandis que d'autres peuvent complètement changer.

Les images de la classe "portrait" sont fortement multi-étiquetées car le fond de l'image peut être associé très souvent à l'une des 6 autres étiquettes.

Formalisation :

Nous considérons le cadre de discernement Ω constitué de 2^7 hypothèses, c'est-à-dire toutes les combinaisons d'hypothèses de bases H_q et \overline{H}_q décrivant les états d'appartenance des images non étiquetées à une classe C_q ($q \in \{1, 2, 3, 4, 5, 6, 7\}$).

Il s'agit ici d'un problème multi-classe et multi-étiquette et toutes les classes sont connues à l'avance. L'espace de décision Ω_M dédié à la classification d'images de manière non exclusive, sans option de rejet en distance et en ambiguïté est donc utilisé (voir page 96).

Enjeux et paramétrage :

Dans le cas du multi-étiquetage, nous ne pouvons pas reproduire les mêmes courbes que précédemment pour illustrer l'impact des stratégies car deux types de mauvaises propositions d'étiquetage doivent être distingués. En effet, lorsque le système fait une proposition d'étiquetage, il est possible qu'une partie des étiquettes proposées soient correctes, qu'une partie soit non pertinente, et qu'une autre partie ne soit pas proposée. Pour analyser l'impact d'une stratégie en termes de propositions incorrectes, nous pouvons distinguer l'accumulation des étiquettes non proposées durant les sélections et l'accumulation des étiquettes proposées en trop.

Généralement, dans un système de classification, on essaie de trouver le meilleur compromis entre rappel et précision. Nous pouvons adopter ce principe en réglant le système pour qu'il y ait globalement autant d'étiquettes proposées en trop, car incorrectes, que d'étiquettes manquantes car non proposées.

Cependant, il faut souligner que les utilisateurs ne préféreront pas forcément adopter ce point de vue. Certains peuvent préférer que le système propose globalement plus d'étiquettes qu'il n'en faut quitte à supprimer celles qui ne sont pas pertinentes. A l'inverse, d'autres utilisateurs préféreront que le système fasse moins de propositions en privilégiant la précision des étiquettes proposées. Notons également que la préférence pour l'une ou l'autre approche peut être motivée par l'ergonomie de l'interface : si par exemple, il est plus rapide et moins coûteux en termes d'interactions d'éliminer des mauvaises étiquettes que d'en ajouter de nouvelles, des utilisateurs peuvent préférer que le système propose globalement plus d'étiquettes.

Avant de faire les tests pour chaque stratégie, nous avons tout d'abord analysé quelle métrique choisir avec quel descripteur des couleurs à combiner avec celui des orientations, pour plusieurs valeurs de f et pour $k = 5$. Des tests de classification automatique en divisant la base en 2, de la même manière que nous l'avons fait dans la première partie de ce chapitre, ont révélé que les meilleures performances sont obtenues avec la distance de Bhattacharyya sur les descripteurs couleurs $\{L,a,b\}$ combinés avec les descripteurs des orientations pour plusieurs valeurs de f autour de 0,8.

Dans un second temps, nous avons testé arbitrairement une stratégie (la stratégie du plus rejeté MR) pour plusieurs valeurs de f afin de trouver le paramétrage permettant d'obtenir le meilleur compromis entre

le nombre final d'étiquettes manquantes et celui des étiquettes proposées en trop car non pertinentes. Ce compromis est obtenu avec $f = 0,825$ avec environ 387 étiquettes manquantes et 384 étiquettes en trop sur 828 étiquettes à retrouver en 715 sélections d'images.

Résultats :

Les figures 6.19 et 6.20 donnent les accumulations des étiquettes manquantes d'une part, et des étiquettes proposées en trop d'autre part, en fonction des sélections.

Nous pouvons premièrement remarquer que les tendances des courbes sont nettement moins marquées que sur les jeux de données Corel. Les stratégies du plus positif MP et du plus rejeté MR "bornent" l'ensemble des stratégies, mais la stratégie MP "retarde" moins les mauvaises propositions, tandis que la stratégie MR généralise moins rapidement. La plupart des autres stratégies ne se distingue pas par une tendance particulière : elles ont tendance à effectuer des propositions d'étiquettes incorrectes régulièrement comme le ferait une stratégie de sélection aléatoire.

Dans ces conditions d'expérimentations, la stratégie du plus positif MP et dans une moindre mesure, la stratégie du plus localement ambigu MLA2, permettent d'obtenir les meilleurs résultats avec 340 étiquettes manquées et 263 étiquettes proposées en trop pour la stratégie MP, et 353 étiquettes manquées et 279 étiquettes en trop pour la stratégie MLA2. Ces stratégies donnent donc au final un couple de rappel précision de (0,574 ;0,630) pour MP et (0,586 ;0,646) pour MLA2. Autrement dit, avec ces stratégies il est possible de d'atteindre jusqu'à 58,6% des 828 étiquettes à retrouver, et lorsque que le système propose des étiquettes jusqu'à 64,6% sont correctes.

Ces performances peuvent paraître moins bonnes qu'avec les jeux de tests de la base Corel, mais il faut bien prendre en compte la difficulté de la base *gt772* où certaines classes contiennent beaucoup d'images multi-étiquetées et où les contenus visuels sont beaucoup plus diversifiés que dans le cas des classes Corel.

Il est intéressant de noter que la stratégie du plus localement ambigu MLA2 semble pertinente pour cette tâche de multi-étiquetage, alors qu'elle ne semblait pas l'être pour les tâche de classification de manière exclusive. En effet, la grande majorité des images multi-étiquetées de ce jeu de test *gt772* possèdent 2 étiquettes et rarement au delà. Cette stratégie sélectionne donc en priorité les plus probables pouvant être multi-étiquetées dans 2 classes, et elle peut être vue comme l'équivalent d'une stratégie du plus positif à 2 classes.

Nous pouvons constater que la stratégie du plus rejeté MR très pertinente dans le cas d'une classification de manière exclusive, n'est pas performante. En effet, la grande valeur de $f = 0,825$, nécessaire pour faire des propositions d'étiquettes multiples a pour conséquence de faire tendre les valeurs des probabilités pignistiques sur l'hypothèse de rejet vers 0. Les sélections de la stratégie MR se basent sur des valeurs trop faibles pour être significatives.

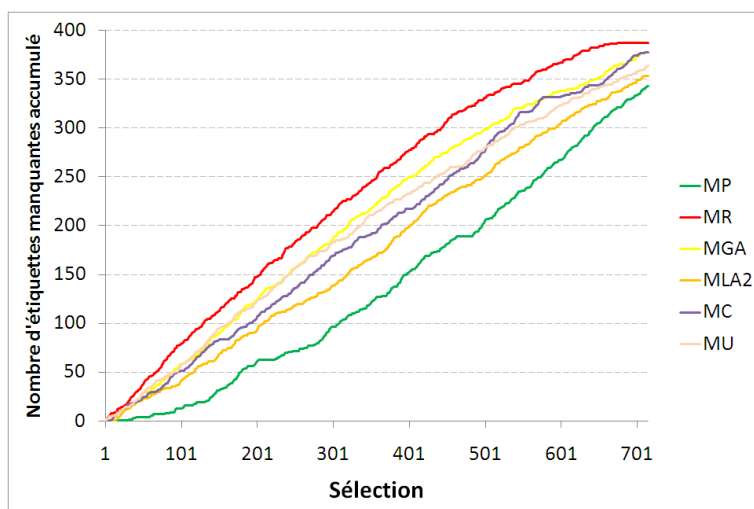


FIGURE 6.19 : Comparaisons de toutes les stratégies en nombre d'étiquettes non proposées manquantes sur le jeu de données gt772 dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace $\{L,a,b\}$, $f = 0,825$ et $k = 5$).

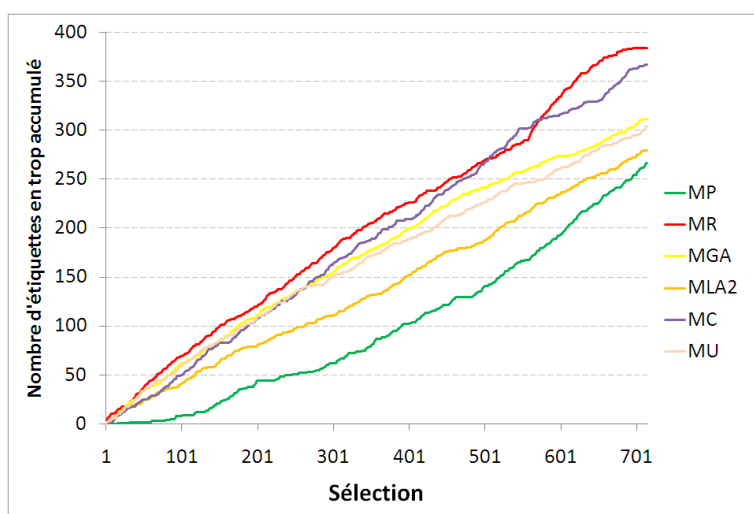


FIGURE 6.20 : Comparaisons de toutes les stratégies en nombre d'étiquettes proposées en trop car incorrectes sur le jeu de données gt772 dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace $\{L,a,b\}$, $f = 0,825$ et $k = 5$).

Nous avons jusqu'à présent dans ce chapitre, décrit d'un point de vue quantitatif les performances de classifications automatiques et caractérisé les stratégies de sélection. La prochaine partie se focalise sur des aspects qualitatifs du système grâce à une évaluation avec un utilisateur potentiel de l'outil au sein de l'INA.

4. Evaluation avec une documentaliste de l'INA

Le travail présenté dans cette thèse est orienté autour de l'aide que l'on peut apporter à un utilisateur pour qu'il structure des collections d'images. Il est donc important de présenter l'outil développé à des utilisateurs potentiels afin de constater dans quelles mesures il peut réellement être utilisé en pratique.

D'après la littérature [Nie93], [Sha91], [DM96], [TPSC⁺03], [Sen90], quatre critères sont couramment utilisés pour évaluer un système interactif :

- Utilité : l'adéquation avec les besoins utilisateurs qui vérifie si toutes les fonctions nécessaires sont disponibles.
- Utilisabilité : étudiant la facilité de manipulation du système par un utilisateur (proche des notions d'ergonomie).
- Acceptabilité : prenant en compte des notions d'acceptation sociale ou de comptabilité techniques.
- Apprenabilité : facilité avec laquelle on apprend à s'en servir.

Ce dernier critère ne sera pas pris en compte car il dépasse le cadre de cette thèse. Le critère d'utilisabilité est le plus délicat à évaluer. Il est étudié selon la norme 9241-11 à partir de :

- l'efficacité : vérifiant si le produit permet aux utilisateurs d'atteindre le résultat prévu,
- l'efficience : mesurant les ressources nécessaires pour atteindre le résultat (effort moindre ou temps minimal),
- la satisfaction : déterminant le confort d'utilisation.

Méthodologie "penser à haute voix"

Pour évaluer le système avec un utilisateur, nous proposons de retenir la méthodologie du "thinking aloud" [Nie92]. Cette méthode consiste à demander à l'utilisateur de "penser à haute voix" lors de la réalisation des tâches que l'expérimentateur lui demande d'effectuer. L'utilisateur doit dérouler un scénario prédéfini en commentant oralement toutes ses actions ainsi que la qualité des résultats. L'évaluateur doit aider l'utilisateur à exprimer ses impressions en lui posant des questions sur les actions qu'il effectue.

Cette méthode fournit de nombreuses données qualitatives et éventuellement quantitatives si les conditions de test le permettent. Dans notre cas particulier, étant donné que le prototype testé apporte des fonctionnalités nouvelles, il est difficile d'étudier l'efficience, car les systèmes de tris de d'images sont manuels et il n'existe pas de système pouvant servir de référence pour une comparaison.

Cette technique de protocole verbal a surtout l'avantage de donner accès aux parcours cognitifs et aux stratégies mises en place par l'utilisateur pendant la tâche. L'observation de l'utilisateur est importante pour noter toutes les gênes ou réactions non exprimées oralement. Il s'agit de noter tous les éléments retraçant la facilité de manipulation de l'interface par les utilisateurs : noter si l'interface est facile à comprendre, facile à apprendre, facile à utiliser, compter les remarques négatives et positives, les hésitations, doutes, les problèmes de manipulations, les erreurs avec leurs origines. . .

Cette méthode s'accompagne généralement à la fin d'un questionnaire de satisfaction concernant le sys-

tème évalué. Le questionnaire utilisé aborde trois volets : les fonctionnalités, l'ergonomie et la visualisation. Certaines questions sont inspirées de [Lew95]).

Utilisateur

Nous avons demandé à la personne responsable du service de la Photothèque à l'INA d'être notre utilisateur. Cette personne est très impliquée et représentative des futurs utilisateurs. En particulier, elle a identifié tous les besoins utilisateurs depuis la mise en place de ce service et elle peut donc être potentiellement intéressée par l'utilisation d'un outil facilitant l'organisation de collections d'images.

La collection

La collection contient environ 900 photographies autour du tournage d'un téléfilm "Borgia" des années 70 prises par plusieurs photographes employés à l'Ortf (voir figure 6.21). L'utilisation de cette collection permet une parfaite adéquation entre l'évaluation et les besoins utilisateurs : les documentalistes de la Photothèque effectuent des regroupements sur ce type de collection en identifiant les plans de tournage (lorsque que le photographe prend place à proximité de la caméra), les scènes de direction d'acteurs, des portraits d'acteurs, différentes ambiances autour du tournage (restauration, préparation des décors, mise en place du matériel technique,...), mais aussi des thèmes spécifiques à la collection (chevaux, baisers, danse,...). Environ un quart des photographies sont en noir et blanc, sachant que les photographies en couleurs présentent parfois avec des tons ternes, peu saturés et sombres, où il n'est pas facile à percevoir rapidement le contenu.

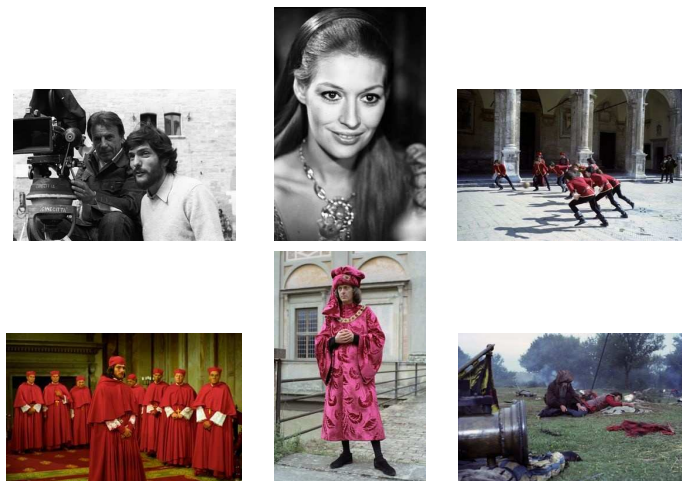


FIGURE 6.21 : Exemple d'images de la collection Borgia.

Déroulement

La première étape de l'évaluation est la présentation par le concepteur du fonctionnement de l'outil sur une base de photographie personnelle. Les principales fonctionnalités sont présentées progressivement, en même temps, que des images sont organisées pas-à-pas. La vue principale et les contrôles disponibles sont présentés : création de classes, déplacement des images à la volée, proposition automatique

d'étiquettes, animation. ...

Les points très importants qui ont été soulignés pendant cette présentation sont qu'il est possible de :

- mettre une image de côté dans une liste de rejet, si l'on ne souhaite pas l'étiqueter immédiatement.
- retirer une image d'une classe vers une autre si sa présence dans la classe ne satisfait plus l'utilisateur,
- associer des contenus visuels très variés dans une même classe,
- multi-étiqueter les images.

Pour finir les stratégies du plus proche MP, du plus rejeté MR, et du plus globalement ambigu MGA sont présentés pour montrer que le système peut sélectionner des images "faciles" et "difficiles".

Dans un second temps, la personne évaluant le système prend en main l'outil, sur le corpus *Borgia* de la photothèque. Elle démarre donc de rien et seule la liste verticale d'images non étiquetées est affichée.

Son premier réflexe est de parcourir rapidement la liste afin de trouver une image l'intéressant pour créer une première classe. Elle lance l'animation avec la stratégie MP. Elle remarque une monotonie des contenus visuels successifs des images similaires. Elle choisit alors la stratégie du plus rejeté pour créer de nouvelles classes.

L'utilisateur se sert beaucoup de la liste de rejet comme classe "tampon" pour y déposer temporairement les images qu'elle ne peut pas classer, mais elle les reprend assez rapidement pour les classer.

La documentaliste était un peu "déroutée" lors de la présentation du système en observant que des images très différentes visuellement étaient mises dans une même classe. Cependant, lors de cette évaluation, l'utilisateur perçoit rapidement l'intérêt de cette approche et finit procéder de la même manière pour ses propres classes.

Par ailleurs, l'utilisateur n'hésite pas à créer de nouvelles classes en retriant des images déjà classées.

Les classes créées correspondent à des groupes de photographies de "baisers", "Vatican", "portrait", "nuit", "cavalier", "danse", ... (voir figure 6.22 pour avoir un instantané de l'organisation créée par l'utilisateur).

L'utilisateur est assez gêné par la taille des images, malgré une grande surface d'affichage : l'utilisateur déplace les classes au plus près possible de la liste des images encore non étiquetée.

Synthèse des observations, des commentaires et du questionnaire

L'utilisateur est enthousiaste et souhaiterait utiliser ce système tant pour l'usage au sein de la photothèque que pour son usage personnel. En effet, ce système présente des fonctionnalités nouvelles qui apportent de nombreux avantages :

- Des fonctionnalités nouvelles et puissantes : notamment grâce à la possibilité de créer des classes contenant des images qui correspondent à un même concept mais peuvent être visuellement différentes.
- Possibilité de créer des classes "conceptuelle" et de réorganiser les classes en court de traitement.
- Correction des affectations des images juste avant leur envoi ou après leur classement.

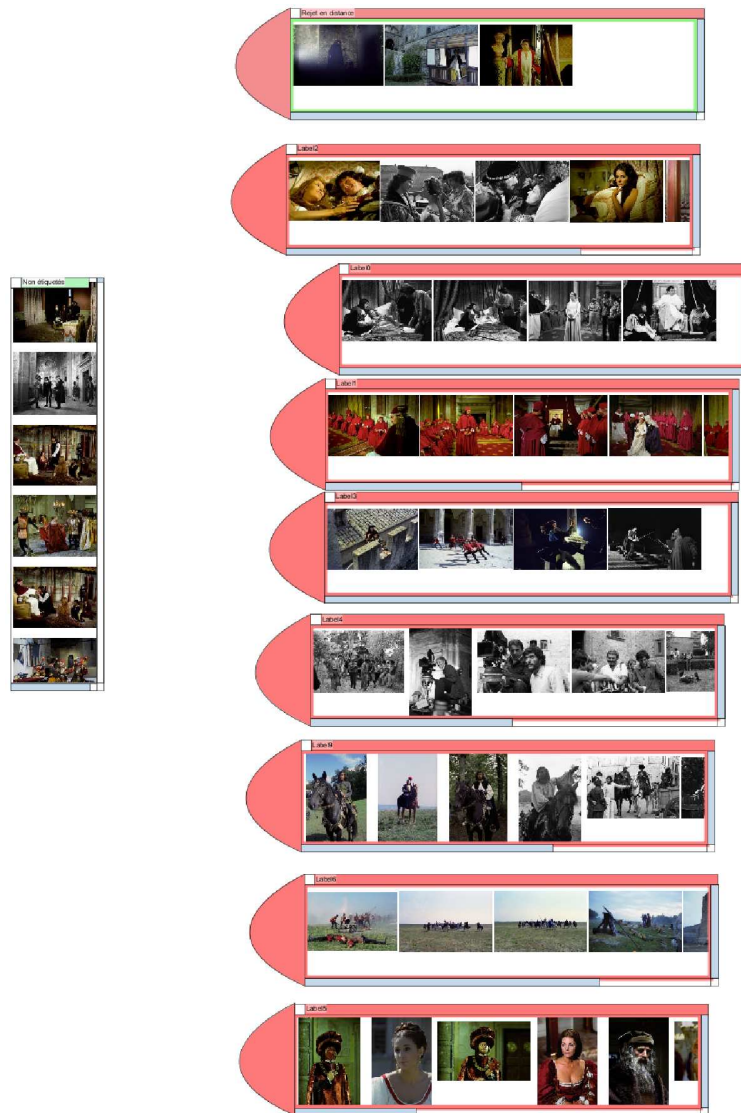


FIGURE 6.22 : *Impression d'écran de l'interface durant l'évaluation de l'outil sur le corpus Borgia. La documentaliste a utilisé régulièrement la liste de rejet (première liste horizontale du haut). Elle a organisé la collection avec les thèmes suivants : "baiser", "direction d'acteurs", "vatican", "combats", "tournage", "chevaux", "champs de bataille", "portraits".*

- Possibilité d'utiliser une classe rejet.
- Possibilité de changer de stratégies de l'ordre d'affichage des photos à classer.
- Possibilité de classer les images une par une ou par petits groupes.
- Ce système permet une initialisation très fine avec un contrôle utilisateur fort puis un classement plus automatisé dans un second temps.
- Possibilité de trier des images de plusieurs reportages en même temps.

Utilité :

L'utilisateur a souligné une adéquation entre le travail des documentalistes dans leurs tâches quotidiennes, notamment pour la création de "chutiers", des regroupements de photographies qui peuvent

hériter des mêmes mots-clés de descriptions.

Utilisabilité - efficacité :

Le critère d'efficacité est positif car cet outil permet d'obtenir des classes d'images tout à fait pertinentes au regard des spécialistes de la documentation. En effet ces classes regroupent les images selon des critères définis par les documentalistes eux-mêmes et sont très riches en termes de diversité visuelle. Une dizaine de classes reste facilement supervisable par un humain, mais au-delà il serait souhaitable en revanche de pouvoir créer des sur-classes ou des sous-classes. L'utilisateur a été très surpris par la qualité des propositions de classements, notamment pour une classe constituées d'images d'ambiance de tournage. Le temps de réponse du système n'ont pas occasionné de gêne et son compatible avec les contraintes d'interactivité.

Utilisabilité - efficacité :

L'efficacité n'a pas été évaluée car nous ne possédions pas de valeurs de référence pour comparer avec des résultats quantitatifs reflétant les ressources dépensées. Néanmoins, on peut supposer qu'une bonne efficacité est liée à une bonne gestion des options. Par exemple, la stratégie du plus rejeté est préférable au début définir pour rapidement un maximum de classes. La stratégie "orientée classe" pour traiter les images par petits lots est plutôt à utiliser en fin un classement.

Utilisabilité - satisfaction :

Le questionnaire, tout comme l'observation utilisateur, permet d'affirmer que le critère de satisfaction est très positif. Les réactions font preuve d'un grand enthousiasme et le fait que l'utilisateur testeur demande à disposer rapidement de ce système est sans doute le meilleur indicateur de satisfaction.

Utilisabilité - apprentissage :

La prise en main de l'outil est rapide et intuitive pour la création de classes tout comme le contrôle du classement des images. L'utilisateur qui n'avait jamais manipulé cet outil avant les tests, a pu mener sa tâche sans difficulté. En revanche, seulement deux stratégies de sélection d'image (la plus positive MP et la plus rejetée MR) ont été rapidement assimilées. Les autres stratégies demandent un apprentissage un peu plus long afin de pouvoir en tirer les avantages.

Recommandations

Un certain nombre de recommandations a été établi à partir des remarques utilisateur afin de pouvoir améliorer le système dans une optique d'intégration dans un système opérationnel.

Recommandations sur de nouvelles fonctionnalités :

- Avoir accès à une vue d'ensemble des images pour obtenir un aperçu général du contenu au départ.
- Possibilités de créer des sur-classes et des sous-classes.

Recommandations sur l'ergonomie et les interactions :

- Trouver une solution pour optimiser l'affichage des images, par exemple avec un zoom automatique

- au passage de la souris, ou en minimisant/agrandissant automatiquement la taille des classes.
- Pouvoir précipiter le mouvement d'une image lorsque l'utilisateur est d'accord avec la proposition de classement.
- Ajout des fonctions standards : la sélection multiple d'images et de raccourcis clavier.

5. Extensions

Nous proposons ici d'aborder ici deux cadres applicatifs où peut être exploité notre outil :

1. L'intégration d'informations issues de métadonnées pour une application "grand public" de tri de photographies personnelles.
2. Une application de macro-segmentation supervisée de vidéo en vue de proposer un outil de structuration semi-automatique.

5.1. Intégration d'informations métadonnées partielles ou inexactes

Objectif :

Dans cette expérimentation, nous nous plaçons dans le cadre d'une application "grand public" de tri de photographies personnelles.

Du point de vue théorique, nous voulons montrer que, grâce au formalisme du MCT, nous pouvons ajouter facilement des informations complémentaires, hors contenu visuel, telle que la date de prise de cliché même si cette dernière est partielle, imprécise ou inexacte.

Contexte :

Un utilisateur possède une collection de taille relativement importante de 1821 photographies personnelles prises à l'occasion d'un mariage. Les photographies, proviennent de 24 modèles d'appareils photographiques différents. Cette collection peut donc être vue comme un reportage photographique "multi-caméras", où de nombreuses scènes correspondent à des temps forts concentrés principalement sur 2 jours.

Le but est d'organiser la collection en 6 "temps" forts, sachant qu'une petite partie des photographies a été prise pendant toute l'année précédant le mariage. Nous pouvons nommer ces groupes d'images "préparation", "séance de photographies", "mairie", "cérémonie et vin d'honneur", "soirée" et "repas du lendemain". En organisant les photographies ainsi, l'utilisateur espère valoriser sa collection d'images pour pouvoir retrouver plus facilement certaines images plus tard.

Description de la collection :

Les contenus des photographies peuvent être très liés sémantiquement et présentent souvent des conte-

nus visuels similaires, des scènes sous différents angles. Cependant, chaque appareil photographique possède ses propres caractéristiques techniques n'entraînant pas les mêmes qualités d'images. En effet, chaque appareil ne provoque pas les mêmes dérives colorimétriques, notamment pour les scènes intérieures. De plus, bien que les photographies tournent autour du même thème, les images peuvent avoir des contenus parfois très divers : certains photographes ont choisi de produire des photographies en noir et blanc, d'autres ont utilisé leur téléphone portable pour prendre des clichés de faible résolution, d'autres photographies proviennent d'appareils argentiques entraînant une qualité visuelle assez médiocre après numérisation,...

Enfin, les temps forts se déroulant parfois sur plusieurs heures, certaines images peuvent avoir des contenus visuels très différents au sein d'un même groupe (changement de luminosité, intérieurs/extérieurs, ...).

Le but étant d'organiser la collection en fonction du temps, nous pourrions utiliser directement l'information de date de prise de cliché disponible sur la plupart des appareils photographiques numériques. Or, les appareils n'ont pas été synchronisés entre eux. En conséquence, les séries de photographies issues des différents appareils peuvent être décalées entre elles de quelques minutes, voire d'une heure (à cause de mauvais réglages d'heures d'été-hiver), éventuellement quelques heures, voire même quelques années. De plus, environ un quart de ces photographies ne possèdent pas d'information temporelle. Nous ne pouvons donc pas profiter de manière directe de l'information temporelle pour organiser les images.

Formalisation :

Il s'agit d'un problème de classification de manière exclusive avec 6 classes, c'est-à-dire qu'une image ne peut appartenir à plusieurs classes à la fois. Nous considérons donc un cadre de discernement Ω constitué en de 2^6 hypothèses, c'est-à-dire toutes les combinaisons d'hypothèses de bases H_q et \overline{H}_q décrivant les états d'appartenance des images non étiquetées à une classe C_q ($q \in \{1, 2, \dots, 6\}$). L'espace de décision Ω_S dédié à la classification d'images de manière exclusive, sans option de rejet en distance et en ambiguïté, est utilisé (voir page 96).

Nous souhaitons combiner les descripteurs visuels avec l'information de temps. Or, cette information est partielle, imprécise ou voire inexacte. L'information de temps, si elle est présente, est représentée par un scalaire. Rappelons que les premières étapes de la modélisation de la connaissance reposent sur les fonctions de croyance convertissant en une distribution de masses, une distance entre une image non étiquetée u avec un voisin l_q^i membre d'une classe C_q . Nous pouvons calculer ces distributions de masses relative à cette description temporelle de la même manière que nous le faisons pour les contenus visuels, en mesurant la distance temporelle d_t entre les images. Pour une image non étiquetée u et un i ème voisin

l_q^i membre d'une classe C_q , la distribution de masses est alors :

$$m^{\Omega_q}(H_q) = \alpha_0 \cdot e^{-\left(\frac{d_t(u, l_q^i)}{\sigma_{i_q}^t}\right)^\beta} \quad (6.13)$$

$$m^{\Omega_q}(\Omega_q) = 1 - m^{\Omega_q}(H_q) \quad (6.14)$$

$$m^{\Omega_q}(\overline{H_q}) = 0 \quad (6.15)$$

Cependant, lorsque l'information temporelle est manquante pour l'une des deux images, nous ne pouvons pas extraire d'information. En conséquence, toute la masse est sur le doute Ω_q et la distribution de masses fournie est alors :

$$m^{\Omega_q}(H_q) = 0 \quad (6.16)$$

$$m^{\Omega_q}(\Omega_q) = 1 \quad (6.17)$$

$$m^{\Omega_q}(\overline{H_q}) = 0 \quad (6.18)$$

Ce formalisme permet de modéliser le manque de connaissance par un doute maximum. Lors de la combinaison des témoignages, ce type de distributions de masses concentrées sur le doute se comporte comme un élément neutre ne perturbant pas les autres distributions de masses.

Résultat :

Les figures 6.23 et 6.24 illustrent le gain obtenu par la fusion des descripteurs visuels et temporels pour les stratégies de sélection du plus positif MP et du plus rejeté MR dans les mêmes conditions d'expérimentation (utilisation de la distance L_1 , $f = 0,75$ et $k = 5$). Chaque classe a été initialisée par une image. Les courbes "Contenu Visuel" correspondent à une première fusion entre les descripteurs d'orientation et de couleur. La complémentarité des informations de contenus visuels et de temps permet un gain très important. Dans le cas de la stratégie du plus positif MP, l'utilisation des descripteurs visuels permet de faire au final 70,89% propositions d'étiquettes correctes, tandis que le temps seul en permet 73,75%. La combinaison de toutes les informations permet de diviser quasiment par 2 le nombre de mauvaises propositions pour atteindre un pourcentage de 84,79% de classements corrects.

Dans le cas de la stratégie du plus rejeté, la combinaison permet d'atteindre un gain du même ordre que la stratégie MP. De plus, la combinaison des 3 sources d'information permet une généralisation rapide : environ 80% des mauvaises propositions de classements sont réalisées dans le premier tiers des images sélectionnées.

Une bonne utilisation de l'outil pourrait consister alors à démarrer avec la stratégie du plus rejeté pour les premières centaines d'images. Puis, la connaissance s'affinant progressivement, il pourrait être intéressant de basculer sur la stratégie "orientée classe" (voir page 107) permettant d'étiqueter par petits lots, les images les plus positives pour chaque classe, afin d'obtenir un gain de productivité.

Dans notre méthode, il est possible de fusionner des informations de natures différentes des contenus

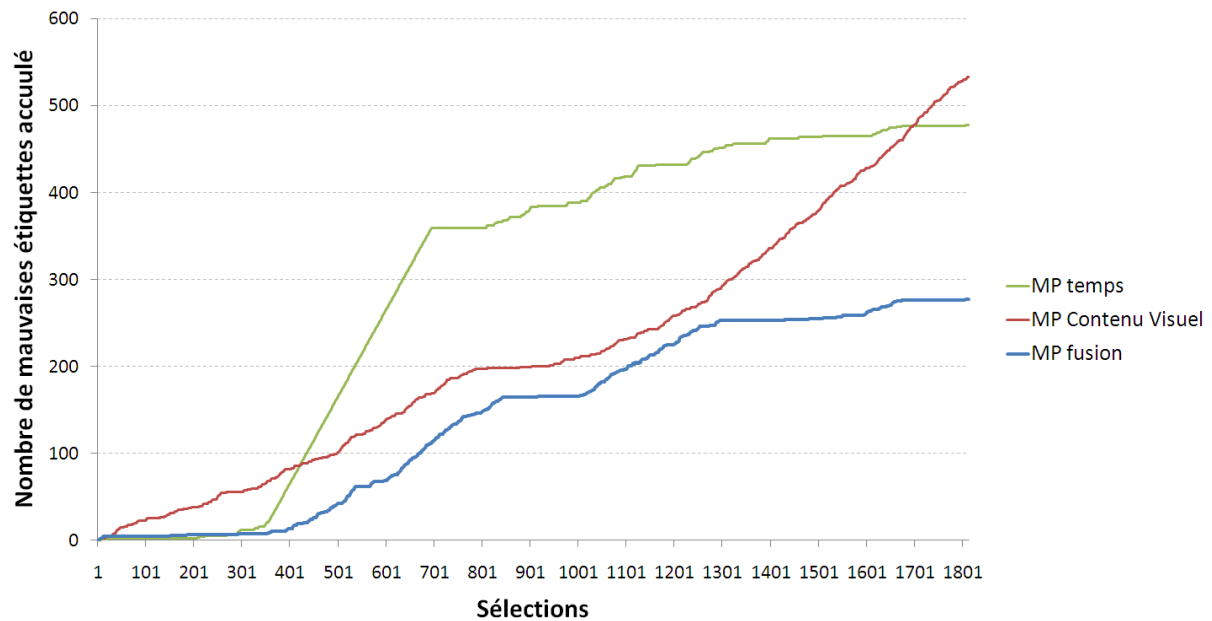


FIGURE 6.23 : Stratégie du plus positif MP sur les 3 combinaisons d'information suivante : "Contenu Visuel" la combinaison des descripteurs couleurs et orientations, "temps" l'information de date de prise de cliché EXIF, "fusion" la combinaison des 3 informations de temps, de couleurs et des orientations. La complémentarité des informations permet un gain très important. L'utilisation des descripteurs visuels permet de faire au final 70,89% propositions d'étiquettes correctes, tandis que le temps seul en permet 73,75%. La combinaison de toutes les informations permet de diviser quasiment par 2 le nombre de mauvaises propositions pour atteindre un pourcentage de 84,79% de classements corrects.

visuels pour accroître la qualité des propositions d'étiquette, même si ces informations sont incertaines et imprécises. Par la suite, nous pourrions expérimenter l'ajout d'autres informations comme les temps d'exposition, la longueur des focales ou encore les boites englobant des détections de visages sur certains modèles d'appareils photographiques.

5.2. Structuration de vidéos

De nombreuses collections de programmes télévisuels suivent une construction narrative se développant autour d'un fil conducteur représenté par un présentateur, des inserts graphiques, ou éventuellement des jingles. . . Le but est de faire émerger les événements redondants de la vidéo présentant un repère ou un intérêt narratif. Nous les nommerons par la suite les éléments structurant d'une vidéo.

L'objectif de cette partie est de développer un outil de structuration interactive de vidéos basé sur les processus de classification semi-automatique proposé dans les chapitres précédents.

Cet outil comporte deux parties :

- une partie permettant de retrouver les éléments structurant d'un programme
- une partie permettant de visualiser les contenus de manière à en refléter la structure narrative.

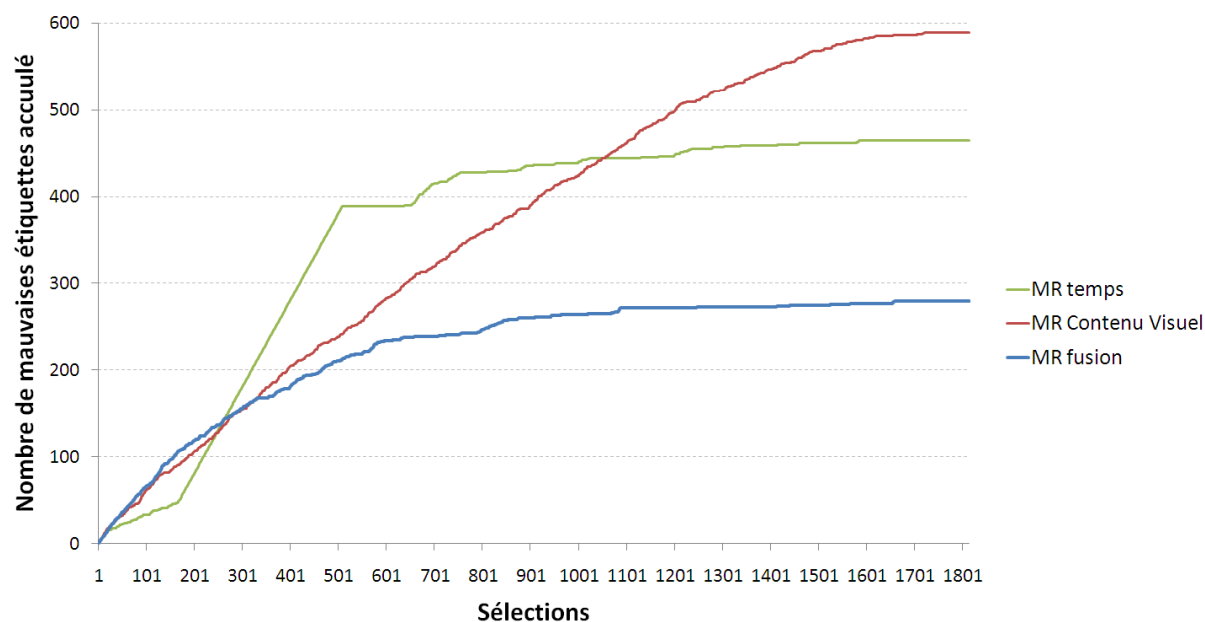


FIGURE 6.24 : Stratégie du plus rejeté MR dans les mêmes conditions que l'expérience que la figure précédente 6.23. La combinaison permet d'atteindre un gain du même ordre que la stratégie MP. De plus, la combinaison des 3 sources d'information permet une généralisation rapide : environ 80% des mauvaises propositions de classements sont réalisées dans le premier tiers des images sélectionnées.

Le cas le plus simple concerne les journaux télévisés qui sont naturellement structurés grâce aux alternances entre les images revenant régulièrement dans le temps (plateau, présentateur, habillage graphique. . .) et celles des reportages. Sur ce type de programmes, l'outil de classification que nous avons développé permet de discerner les séquences de "plateau / non plateau" ou "présentateur/ non présentateur". Ce cas est relativement simple car les images présentateur/plateau présentent une homogénéité visuelle extrêmement élevée : fond caractéristique, peu de changements de prises de vues.

Cependant, d'autres types de programmes ont des structures narratives moins évidentes, notamment dans le cas des fictions. Il est alors plus difficile de retrouver une adéquation entre le contenu visuel et le fil narratif de la vidéo. De plus, comme pour les collections d'images, il existe une part de subjectivité venant des utilisateurs : deux utilisateurs ne distingueront peut être pas les mêmes groupes d'images comme éléments structurant une vidéo. L'outil que nous proposons peut alors aider un utilisateur à identifier les parties narratives de la vidéo afin d'orienter la visualisation sur l'axe narratif qu'il désire.

5.2.1. Journaux télévisés

Formalisation :

Il s'agit d'un problème de classification bi-classe. Nous considérons donc un cadre de discernement Ω

constitué en de 4 hypothèses :

$$\Omega = \{(\overline{H_1}, \overline{H_2}), (H_1, \overline{H_2}), (\overline{H_1}, H_2), (H_1, H_2)\} \quad (6.19)$$

où H_1 représentent les états "l'image appartient à la classe présentateur", et $\overline{H_1}$ l'état complémentaire. L'espace de décision considéré est alors :

$$\Omega_s = \{(H_1, \overline{H_2}), (\overline{H_1}, H_2)\} \quad (6.20)$$

Méthode générale :

1. Des imagenttes de la vidéo sont extraites régulièrement à une fréquence arbitraire fixée par l'utilisateur. Par exemple, pour un journal télévisé "20h" de France 2, il est raisonnable d'échantillonner la vidéo toutes les 10 secondes car l'on suppose qu'un plan reste fixe sur le présentateur pour une durée d'au moins 10 secondes au minimum.
2. Deux classes "présentateur" et "non présentateur" d'images sont initialisées par une image chacune.
3. L'outil est utilisé pour classer le reste des images et identifier ainsi les 2 groupes "narratifs".
4. La visualisation est créée (la méthode de création de la visualisation est détaillée plus loin).

Exemple du journal télévisé de France2 (le 20h du 13 juin 1997) :

La figure 6.25 présente les deux premières imagenttes utilisées pour initialiser 2 classes "présentateur" et "non présentateur". L'outil est ensuite utilisé pour segmenter la vidéo en ces 2 classes.

La figure 6.26 présente les résultats obtenus en utilisant 4 stratégies de sélection active en utilisant comme paramètre interne la distance de Bhattacharyya sur le descripteur couleur {L,a,b} uniquement, avec $f = 0,8$ et $k = 5$.

Les courbes comparent les stratégies de sélection du plus positif MP, du plus rejeté MR, du plus globalement ambigu MGA, et du plus incertain MU. L'homogénéité des contenus visuels permet d'atteindre de très bonnes performances de classification. Dans le cas le plus défavorable, la stratégie MR produit 4 mauvaises propositions de classement sur 167 images au total. Dans le meilleure des cas, la stratégie MU permet de segmenter la vidéo sans aucune erreur.

Ces résultats permettent d'envisager, une automatisation complète de la segmentation pour ce type de programme. En effet, ces journaux s'intègrent dans une collection de programmes visuellement homogène sur une période de temps signifiante. Le modèle de classes généré pour un journal pourrait alors s'étendre à l'ensemble de la collection.

Une fois la segmentation réalisée et validée par l'utilisateur, une visualisation est créée (voir la figure 6.27) en s'appuyant sur une approche graphe [Thi06] selon la méthode suivante :



FIGURE 6.25 : Les deux premières imagettes dans le temps du journal télévisé de France2 du 20h du 13 juin 1997 servant à initialiser deux classes : "présentateur" et "non présentateur".

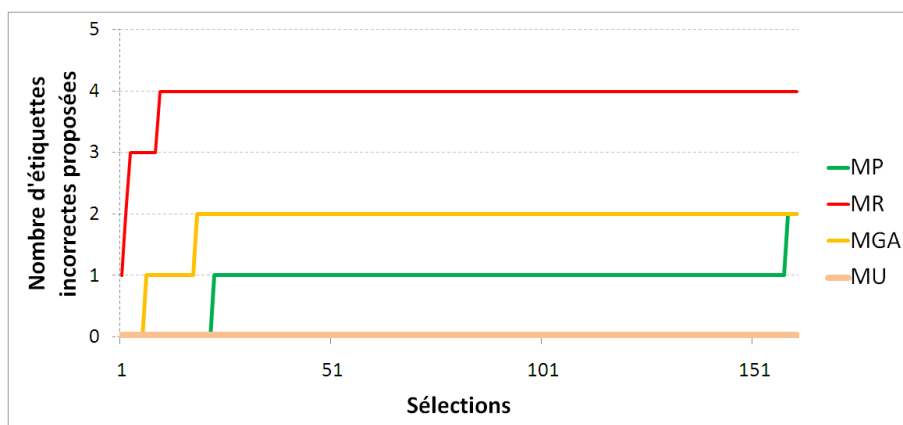


FIGURE 6.26 : Courbes d'évaluations comparant les stratégies de sélection du plus positif MP, du plus rejeté MR, du plus globalement ambigu MGA, et du plus incertain MU. L'homogénéité des contenus visuels permet d'atteindre de très bonnes performances de classification : en particulier la stratégie MU permet de segmenter la vidéo sans aucune erreur.

1. Un nœud représente une imagette de la vidéo.
2. Des arrêtes temporelles sont créés entre deux imagettes successives : on obtient ainsi un graphe représentant uniquement la ligne temporelle de la vidéo.
3. Des arrêtes temporelles sont créés entre deux imagettes successives au sein de la classe "présentateur". On obtient ainsi une deuxième ligne temporelle croisant la première à chaque alternance "présentateur / non présentateur".
4. Le graphe est positionné grâce à un modèle de force basé sur [FR91].

On obtient ainsi une représentation globale de la vidéo reflétant la structure narrative. Sur la figure 6.27 nous pouvons voir de gauche à droite les imagettes du présentateur. Chaque boucle représente un reportage ou une interruption avec le présentateur (interview, "inserts" graphiques...). Les images étant échantillonnées régulièrement dans le temps, on conserve l'information de durée des reportages, contrairement aux approches se focalisant sur des images clés [VBTGV07]. Par exemple, dans cette vidéo, nous pouvons voir qu'il y a une longue interruption à la quatrième boucle. Cette interruption révèle un événement particulier lié à l'actualité qui correspond à un décrochage en direct du salon du Bourget. La couleur renseigne également sur le type de contenu. Par exemple, l'aspect bleuté autour de la neuvième boucle indique une interview plateau. L'aspect "orangé" des deux dernières boucles indique des sujets

culturels.

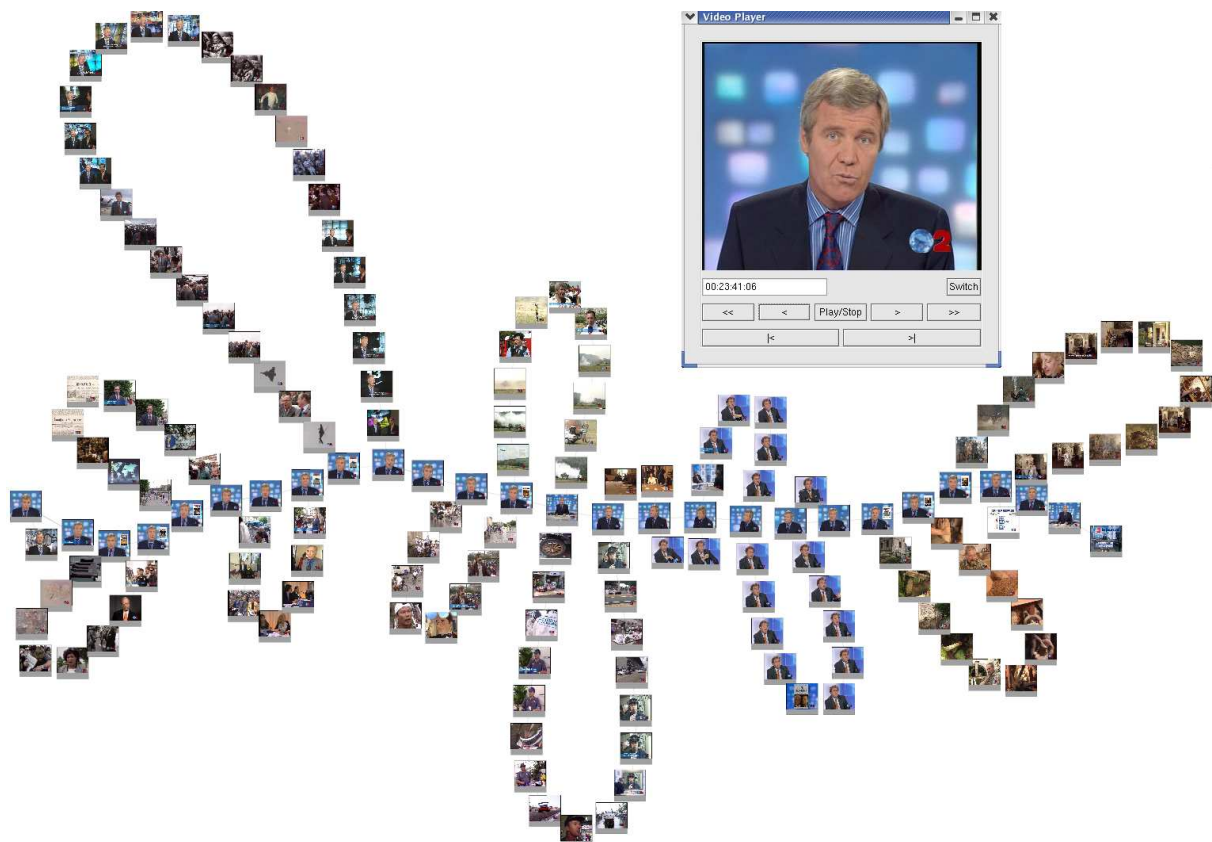


FIGURE 6.27 : *Visualisation du journal télévisé de France2 du 20h du 13 juin 1997 reflétant la structure narrative. Les imagerie du présentateur seul sont positionnées dans l'ordre chronologique de gauche à droite. Chaque boucle représente un reportage ou une autre interruption avec le présentateur (interview, graphiques...). Les images étant échantillonnées régulièrement dans le temps, la longueur renseigne sur la durée des reportages et des interruptions. Le lecteur en haut à droite illustre que dans une version intégrée on peut naviguer dans le contenu par le biais des timecodes des imagerie.*

Dans le cas des journaux télévisés, il est possible d'envisager une version complètement automatisée pour produire ce type de visualisation. Il faudrait des investigations supplémentaires pour décliner notre système vers une approche complètement non supervisée, sans intervention de l'utilisateur.

5.2.2. Episode d'une série

Dans les fictions, il est très fréquent que plusieurs fils narratifs s'entremêlent. Dans les séries, notamment le "montage alterné" est extrêmement employé : ce procédé permet de rythmer la narration en montrant de manière alternée plusieurs actions simultanées.

Dans ce cas, il est bien plus difficile de trouver automatiquement la structure narrative du programme et la présence de l'utilisateur est indispensable pour guider le système à identifier les contenus visuels permettant de regrouper les séquences d'une même histoire.

La figure 6.28 donne les résultats d'une macro-segmentation d'un épisode d'une série en trois classes identifiées par l'utilisateur : "générique", "histoire1" et "histoire2". La vidéo est échantillonnée à une fréquence de 3 secondes car les alternances de plans sont plus rapides que pour un JT. Les descripteurs couleurs et orientations sont combinés avec l'information de timecode des images. L'apprentissage des classes avec les stratégies MR et MC est accompli après l'étiquetage d'environ 10% des images. Cette performance s'explique par la structure des données : il existe une forte redondance des contenus visuels donnant des groupes d'images très denses dans l'espace des descripteurs. L'exploration des contenus avec la stratégie du plus rejeté MR permet d'identifier rapidement les différents angles de vues. Une fois que les images les plus représentatives de la diversité visuelle sont étiquetées, le reste des images peut s'étiqueter automatiquement sans erreurs.

La figure 6.29 reprend le même procédé de visualisation précédent mais en considérant 2 fils conducteurs cette fois-ci. Nous obtenons ainsi une représentation de la structure de la vidéo. En particulier, nous pouvons remarquer que les deux fils narratifs s'entremêlent dans les premières boucles, ce qui révèle bien l'utilisation du montage alterné dans ce programme. La longueur des boucles permet d'indiquer la durée des actions.

Ce premier travail sur des programmes moins structurés s'oriente au niveau applicatif sur des outils d'analyse filmique dont une cible potentielle pourrait être les chercheurs en sciences humaines. En effet, par l'Inatèque de France, l'INA met à disposition pour des buts de recherche sociologique des corpus spécifiques. Par exemple, les émissions politiques constituent des corpus intéressants pour les chercheurs qui espèrent évaluer les temps de paroles et d'apparition des hommes politiques et leur propos. Un tel outil leur permettrait d'analyser les programmes du point de vue d'un ou de plusieurs intervenants.

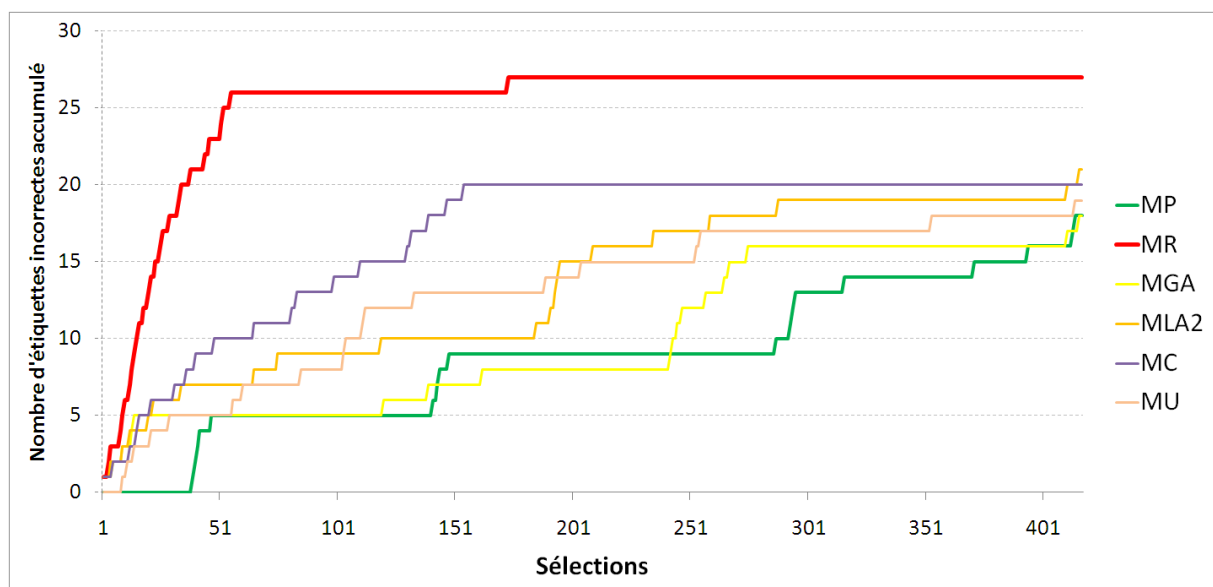


FIGURE 6.28 : Courbes comparant les stratégies de sélection du plus positif MP, du plus rejeté MR, du plus globalement ambigu MGA, et du plus incertain MU. L'homogénéité des contenus visuels permet d'atteindre de très bonnes performances de classification : en particulier la stratégie MU permet de segmenter la vidéo sans aucune erreur.

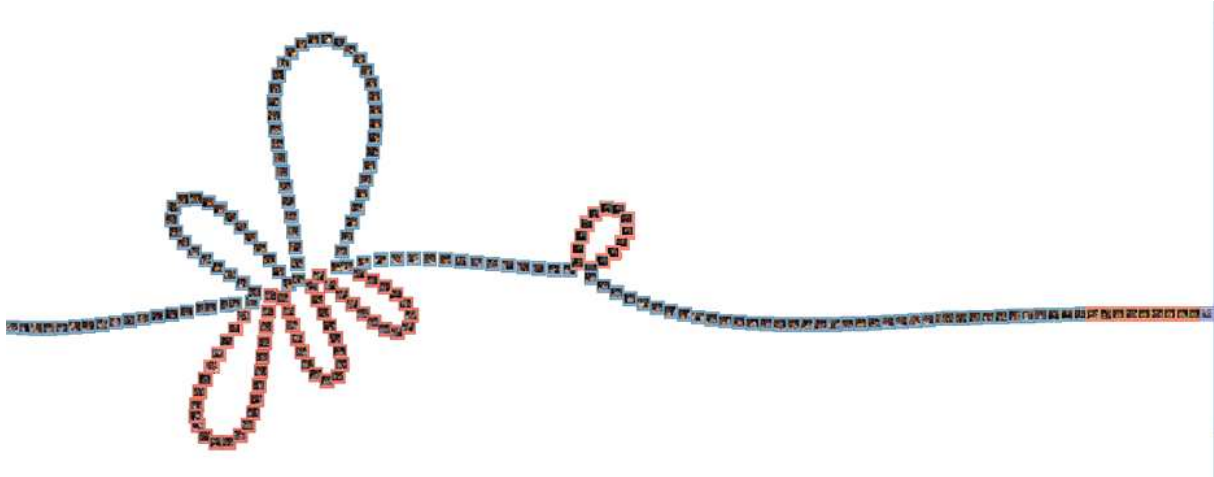


FIGURE 6.29 : Visualisation d'un épisode d'une série télévisée "Friends" contenant 2 fils narratifs.

6. Conclusion

En ce qui concerne la modélisation de la connaissance, nous pouvons conclure que les paramètres qui présentent le plus d'impacts en termes de performances sont le choix de la distance entre descripteurs et la zone d'influence f des fonctions de masses.

La fusion tardive possède plusieurs avantages sur l'approche de fusion précoce de descripteurs. En effet, d'une part, les performances de la fusion tardive sont globalement supérieures et permettent la combinaison des vecteurs de dimensions différentes sans problème d'échelle. D'autre part, cette approche offre la possibilité d'analyser et de quantifier (grâce à l'utilisation du MCT) les contributions de chaque descripteur à travers le conflit.

Nous avons proposé et expérimenté 6 stratégies dans le cadre de notre formalisme. Dans le cas d'une classification exclusive et en fonction des usages envisagés, les stratégies du plus positif et du plus rejeté sont généralement les plus intéressantes et les plus robustes aux réglages de paramètres. La stratégie MR permet de limiter le nombre de mauvaises propositions de classements tout en généralisant rapidement. La stratégie donne en général un nombre de mauvaises propositions d'étiquettes plus important que la stratégie MP, mais est quand même intéressante du point de vue de l'usage si un utilisateur veut passer par des phases sans mauvaises propositions d'étiquettes.

L'intérêt d'avoir développé plusieurs stratégies réside dans l'apport potentiel de la combinaison de stratégies. Par exemple, un critère de sélection d'image par ambiguïté n'est pas pertinent au début de la classification des images, car les classes ne sont pas encore bien identifiées et pauvres en images. Une stratégie de sélection par ambiguïté présente de l'intérêt dès lors que deux classes en viennent à se chevaucher dans l'espace de description. D'un point de vue des usages, l'alternance de stratégies apparaît comme une piste de recherche intéressante.

Notre approche nous a permis de valider la classification automatique multi-étiquette sur un corpus

d'images de référence avec d'excellentes performances. Par contre, il est plus difficile de comparer les stratégies de sélections dans le cas multi-étiquette, car les propositions d'étiquettes sont partiellement incorrectes (il y a des étiquettes manquantes, en trop et en commun). Les stratégies du plus positif et du plus localement ambiguë semblent être les plus intéressantes.

L'évaluation de l'outil par la responsable de la Photothèque de l'INA a permis de confirmer l'adéquation avec leurs besoins. Elle a montré aussi que l'interface est de prise en main rapide et intuitive. La satisfaction et l'utilisabilité sont excellente pour les tâches à réaliser. Sur la demande de la responsable de la Photothèque, le département de la Recherche mettra ce prototype à disposition des documentalistes.

La possibilité de prendre en compte de métadonnées partielles, peu fiables et complémentaires au contenu visuel, montre la puissance du formalisme du MCT pour la modélisation des connaissances. D'un point de vue applicatif, ces possibilités ouvrent des champs d'investigation importants : en effet, dans le cadre de la Photothèque, le fond est composé de reportage périodiques, comme par exemple le Tour de France. Ces reportages sont partiellement annotés et contiennent les mêmes types de scènes (podium, lignes d'arrivée, images de peloton. . .). L'enjeu dans ce cadre est d'effectuer des classifications sur des images par les contenus visuels et des annotations partielles. La dernière phase du processus serait alors d'étendre les annotations aux images non annotées.

D'un point de vue application professionnelle, cet outil présente une première approche de classification assistée. Pour les usages de la Photothèque de l'INA, les documents étant déjà regroupés au sein de reportages contenant de 40 à 3000 images, il apporte déjà une solution intéressante pour les documentalistes. Néanmoins, les volumes pris en compte en terme de nombre d'images et de nombre de classes restent relativement modestes pour des temps de calcul compatibles avec l'interactions. La prise en compte d'un nombre d'images supérieur nécessiterait des optimisations supplémentaires. De plus, il peut être difficile pour un utilisateur de considérer au même niveau de nombreuses classes à la fois. En effet, les travaux sur la mémoire [Mil56] montrent qu'au delà de 7 classes en parallèle, le discernement de l'utilisateur n'est plus maximal. La prise en compte d'un nombre de classes supérieur nécessiterait la mise en œuvre d'une hiérarchie de classes.

Par ailleurs, l'évaluation a montré l'intérêt de l'utilisateur pour une visualisation globale (page 113) alternative à notre visualisation locale proposée dans notre interface. Nous avons anticipé ce travail et il existe ce type de visualisation globale dans notre système. Mais cette possibilité reste à explorer et ouvre sur des problématiques différentes de visualisation d'information.

BILAN ET PERSPECTIVES

Bilan

Les collections d'images ne sont généralement pas toutes organisées de la même manière en fonction des contenus visuels, du contexte applicatif et des objectifs des utilisateurs. Pour faire face à cette diversité de traitement des collections d'images, le système qui a été proposé dans ce travail de thèse s'appuie sur une collaboration entre une partie automatique pouvant classer des images par leurs contenus et un utilisateur pour assister ce dernier dans sa tâche d'organisation d'une collection.

La partie automatique s'appuie sur une modélisation structurant les connaissances extraites sur les contenus des images et se remettant à jour en fonction des actions de l'utilisateur. Pour ce faire, la théorie du Modèle des Croyances Transférables (MCT) a montré sa parfaite adéquation pour décrire l'appartenance d'une image à une ou plusieurs classes en identifiant différents états (positifs, ambigus, rejetés, en conflit, incertains). En particulier, elle permet de combiner des descripteurs de tailles très différentes sans poser de problème d'échelle de valeurs et de prendre en compte des connaissances complémentaires telles que celles apportées par les métadonnées des photographies même si ces dernières sont incomplètes ou imprécises.

Cette formalisation avec le MCT permet ainsi d'envisager différents traitements des collections d'images avec le même système pour effectuer de la classification bi-classe, multi-classe et multi-étiquette. Nous avons validé ces différents schémas de classification sur des collections d'images variées. Nous avons montré en particulier, que sur une tâche de classification automatique multi-étiquette, réputée comme étant la plus difficile, notre système est performant par rapport à d'autres techniques de classification.

Un utilisateur ne pouvant pas traiter toutes les images non étiquetées à la fois, le système sélectionne successivement des images à classer en proposant des étiquettes en fonction de la connaissance des classes. Une originalité de ce travail de thèse réside en la description de stratégies de sélection inspirées par les méthodes d'apprentissage actif. La plupart des stratégies classiques sont ainsi exprimées dans le même cadre formel du MCT utilisé pour la modélisation de la connaissance. Ces stratégies sont en adéquation avec des besoins utilisateurs organisant une collection d'images. Elles permettent à l'utilisateur de traiter en priorité les images non étiquetées qu'il juge les plus utiles, à un instant donné, pour construire les classes. Par exemple, la stratégie du plus positif lui permet par exemple d'identifier les images non étiquetées aux contenus les plus similaires aux classes. Celle du plus rejeté permet de trouver de nouveaux contenus visuels. Celle encore du plus ambigu permet à l'utilisateur de désambigüiser des images ayant des contenus visuels communs à plusieurs classes ou de trouver des images pouvant être potentiellement

multi-étiquetées. . .

Ce travail ayant été développé dans le cadre d'une collaboration avec l'INA, nous avons réalisé un prototype complet écrit en langage Java fonctionnant sur des ordinateurs standards. Ce prototype nous a servi de plateforme d'expérimentation et nous a permis d'évaluer les différentes étapes du système. Le prototype a été évalué par un utilisateur professionnel et les résultats sont très positifs. Le département de la Recherche envisage de mettre à disposition cet outil au sein d'un service opérationnel : la Photothèque de l'INA.

La complexité de la tâche d'organisation d'une collection d'images est telle le système proposé ne peut être qu'une première étape qui ouvre sur de nouvelles perspectives.

Perspectives

Modélisation de la connaissance

La modélisation de la connaissance extraite des images vers des concepts de plus haut niveau sémantique est un challenge qui dépasse de loin le cadre de cette thèse. En effet, les performances du système en termes de propositions automatiques d'étiquettes faites à l'utilisateur sont dépendantes de la manière dont la connaissance est modélisée à travers le choix des descripteurs, d'une mesure de dissimilarité et du classifieur de base (les k plus proches voisins dans notre cas).

L'avantage du formalisme qui a été mis en place permet d'envisager l'ajout de descripteurs complémentaires qui permettraient à terme d'accroître la connaissance. En particulier, nous pourrions étudier la prise en compte des descriptions plus locales (blocs, blobs, points d'intérêts. . .). Cette modélisation étant basée sur une comparaison entre descripteurs d'images, il est clair que le choix d'une mesure de dissimilarité entre ces descripteurs se révèle comme étant très important. Nous pourrions envisager de concevoir des mesures de dissimilarités adaptatives, en s'inspirant des techniques de "feature selection" afin de rechercher les composantes des descripteurs les plus représentatives par classe.

Nous nous sommes inspirés d'une méthode de classification de type k plus proches voisins pour combiner des témoignages apportés par les différents membres des classes. La généralité du système permet d'envisager d'autres classifieurs de bases (SVM, Bayes, . . .), dès lors qu'ils fournissent une sortie évaluée. Ces valeurs en sortie des classifieurs peuvent être alors interprétées par des règles expertes afin de les transformer en distributions de masses.

Au delà de toutes ces améliorations potentielles à expérimenter, il faut souligner que la formalisation avec le MCT offre des outils d'analyse puissants, tels que la non-spécificité et le conflit, pour choisir un type de classifieur. En effet, il n'est pas garanti qu'il existe un classifieur idéal offrant toujours des performances de classification optimales pour toutes les collections d'images. Il semble que les performances de classification soient corrélées aux mesures de non spécificité moyenne et du conflit moyen des distributions de masses. La mise en place un module d'analyse des distributions de masses permettrait

de choisir automatiquement les classifieurs et les paramètres les plus adaptés à la collection d'images à traiter.

Etude des combinaisons de stratégies

Dans le système proposé dans le cadre de ces travaux, l'utilisateur choisit la stratégie de sélection des images. Or, ces stratégies ont des effets extrêmement différents en termes d'enrichissement de la connaissance sur la collection et d'actions correctives de l'utilisateur. Elles ne conduisent ni au même nombre final de mauvaises propositions d'étiquettes, ni aux mêmes "moments" et ni même sur les mêmes images. Dans le système actuel, l'utilisateur peut changer de stratégie quand cela lui semble opportun.

On peut alors se demander s'il n'existe pas un enchaînement de stratégies qui permettrait de diminuer le nombre de mauvaises propositions d'étiquettes. Nous avons par exemple fait une expérience préliminaire où l'on alterne successivement une sélection d'un plus rejeté avec des sélections de plus positifs, tant que l'utilisateur valide les propositions d'étiquettes. Cette alternance permet de diminuer sensiblement le nombre de mauvaises propositions d'étiquettes.

Il faudrait alors être capable de savoir à tout moment quelle stratégie est la stratégie la plus pertinente. L'étude de combinaisons de stratégies pose cependant deux problèmes importants : d'une part comment évaluer les performances de ces combinaisons, et d'autre part, dans quelle mesure un utilisateur accepterait ce fonctionnement où il n'est plus maître du choix de stratégie.

Organisations hiérarchiques

Il est naturel pour les utilisateurs de vouloir organiser les classes dès que le nombre de classes et d'images est important. En effet, il est difficile pour un utilisateur de considérer plus d'une dizaine de classes à la fois, et une extension de l'outil vers une approche hiérarchique semble adaptée. Notre méthode pourrait se décliner dans une version hiérarchique de la manière suivante : à chaque niveau de hiérarchie correspondrait un cadre de discernement. Le premier niveau hiérarchique correspondrait à la méthode décrite dans ce manuscrit. Puis chaque classe, pourrait être considérée à son tour comme un ensemble d'images non étiquetées pour le niveau hiérarchique inférieur permettant de définir ainsi un nouveau cadre de discernement.

La formalisation de cette nouvelle fonctionnalité pose également des verrous théoriques concernant le multi-étiquetage dans le cadre d'organisations hiérarchiques.

Apprentissage semi-supervisé

Nous pourrions aborder d'autres cadres applicatifs travaillant sur un nombre d'images beaucoup plus élevés, tels que la structuration de vidéos ou la recherche d'images par le contenu. D'autre part, le personnel de la Photothèque de l'INA a exprimé l'intérêt d'organiser un plus grand nombre d'images notamment pour croiser plusieurs collections relatant les mêmes types d'événements, comme par exemple les reportages photographiques du Tour de France sur plusieurs années.

Pour répondre à ce besoin de traiter un grand nombre d'images, nous pourrions décliner notre méthode

dans une version semi-supervisée. Nous pourrions estimer les paramètres de classes en nous appuyant conjointement sur les données non étiquetées et celles étiquetées. Le but serait alors de réaliser une première organisation grossière à plus grande échelle afin d'identifier les principales classes de façon non supervisée. Cette démarche pourrait s'inscrire également dans une approche hiérarchique comme exprimée plus haut.

ANNEXES

7. Communications

Publications

H. Goëau, J. Thièvre, M-L. Viaud, and D. Pellerin. Interactive visualization tool with graphic table of video contents. In IEEE International Conference on Multimedia & Expo (ICME 2007), July 2007

H. Goëau, J. Thièvre, M-L. Viaud, and D. Pellerin. Navigation par le contenu dans les vidéos avec des tables graphiques des matières. In 21e Colloque GRETSI, Troyes, France, Septembre 2007

H.Goëau, O. Buisson, ML Viaud. Image Collection Structuring Based On Evidential Active Learner. In Sixth International Workshop on Content-Based Multimedia Indexing, CBMI 2008, London, UK, june 2008.

ML Viaud, J. Thièvre, H. Goëau, A. Saulnier, and O. Buisson. Interactive components for visual exploration of multimedia archives. In ACM International Conference on Image and Video Retrieval (CIVR 2008), Niagara Falls, Canada, July 2008

Rapports

Report on the state-of-the-art on advanced visualisation methods (D7.2) - September 2007 Report of Vitalas project (Video and image Indexing and reTrievAl in the LARge Scale)

Communications

H. Goëau, Interactive Visualization Tool with Graphic Table of Video Contents, MCVC'08 Cannes (MUSCLE Conference joint with VITALAS Conference), fev 2008

TABLE DES FIGURES

1.1	La liberté d'expression des catégories par un utilisateur durant la structuration d'une collection d'images ou d'une vidéo, dépend du contexte applicatif.	4
1.2	Exemple d'une classe d'images contenant une caméra issues d'un reportage photographique de la photothèque de l'INA.	5
1.3	Deux points de vue narratifs différents selon deux utilisateurs distincts. Le premier utilisateur porte son attention principalement sur le personnage de gauche, alors que le second est plus attentif à l'intrigue se déroulant dans un même décors.	5
1.4	Exemple d'une classe d'images reliées par un concept relativement abstrait et personnel de photographies prises sur la route lors d'un déplacement pendant un séjour.	6
2.1	Vue générale du système proposé. Une collection d'images est constituée d'images non étiquetées. Le module de modélisation de la connaissance convertit des descriptions sur les contenus visuels en différentes mesures d'appartenance aux classes. Puis cette analyse est utilisée dans un module de sélection active d'images. L'utilisateur peut demander à ce module différents types d'images à sélectionner en priorité. Le module fait alors des propositions de classement d'images que l'utilisateur valide à travers une interface homme-machine. Chaque nouvelle image étiquetée est alors prise en compte par le module de modélisation et synthèse de la connaissance pour raffiner la connaissance des images encore non étiquetées, améliorant ainsi la sélection active d'images et les propositions de classements automatiques.	22
3.1	Schéma général de l'architecture du système de structuration de collection d'images proposé. Le module de modélisation et de synthèse de la connaissance analyse les images non étiquetées et celles qui ont été précédemment étiquetées par l'utilisateur. Cette analyse permet d'établir un état de la connaissance quand à l'appartenance des images non étiquetées aux classes. Cette étape prépare les données pour le prochain module de sélection active d'images et aux propositions automatiques d'étiquettes dans l'interface homme-machine.	26

3.2	Différents problèmes de classification, du plus simple ou plus difficile théoriquement. Les membres des classes sont représentés dans un espace à 2 dimensions. Dans le cas de la classification bi-classe, l'objectif est de retrouver tous les éléments qui appartiennent à une classe ou non. Le cas multi-classe généralise le problème bi-classe à plusieurs classes. La classification multi-classe et multi-étiquette complexifie le problème en autorisant l'appartenance de certains échantillons à plusieurs classes simultanément.	29
3.3	Les différents problèmes de classification (voir figure 3.2) sont reconsidérés dans des situations plus réalistes : les membres d'une même classe peut être éclaté en plusieurs groupes éloignés. En conséquence, il n'y a pas d'homogénéité des données et les différentes classes peuvent éventuellement se superposer.	31
3.4	Images non étiquetées loin de toutes les images étiquetées : 3 classes de 4 images chacune sont représentées par des croix pleines, des ronds et des triangles dans un espace de description théorique à 2 dimensions. Cinq images encore non étiquetées sont représentées par 4 croix et un losange. L'image non étiquetée représentée par le losange est isolée : il est probable que le contenu visuel ne soit pas représentatif d'une classe. A l'inverse, les images non étiquetées représentées par des croix sont certainement représentatives d'un contenu visuel localement homogène, et peuvent être utilisées pour définir une nouvelle classe ou bien être associées à une des classes existantes.	34
3.5	La distinction entre frontières "extérieures" et "intérieures" doit être prise en compte pour pouvoir gérer la nouveauté de contenus visuels. Les techniques de classification discriminatives ne permettent pas d'identifier explicitement les frontières extérieures. Pourtant, il important de pouvoir identifier des échantillons loin des frontières extérieures des classes pour éventuellement définir de nouvelles classes ou compléter les différents aspects visuels d'une classe existante.	35
3.6	Exemples de représentations d'histogrammes de 256 classes couleurs dans les espace {R,G,B}, {H,S,V}, {L,a,b}. L'espace couleur est découpé en 256 "cubes" de tailles identiques et chaque "bulle" représente alors une classe de couleurs dont la taille est proportionnelle au nombre de pixels de cette couleur. Nous pouvons remarquer que l'histogramme dans l'espace {R,G,B} a tendance à détailler en beaucoup de classes les couleurs "foncées", là où le système visuel humain à du mal à distinguer les couleurs entre elles.	38
3.7	Exemple de dérive couleurs au sein d'une même collection d'images : la même scène est photographiée par plusieurs modèles d'appareils photographiques numériques différents. Les conditions d'illumination liées à l'angle de vue, ainsi que la qualité et les réglages internes des appareils produisent une collection d'images avec des écarts de couleurs parfois très importants. Il peut être alors intéressant de s'appuyer sur des descriptions sur le contenu structurelle des images pour pouvoir les comparer entre elles par exemple avec des descriptions de textures ou des orientations.	39

3.8 Exemples de représentations d'histogrammes polaires classiques des orientations des gradients. Les histogrammes polaires permettent de recenser la distribution des gradients des pixels selon 8 orientations et 8 amplitudes différentes. L'espace est ainsi découpé en 64 "arcs pleins" dont l'intensité de la couleur est proportionnelle au nombre de pixels représentant le même type d'orientation. Par exemple, la première image est marquée par des orientations intenses sur les directions "diagonales". La seconde image est plutôt marquée par de nombreuses petites orientations verticales et horizontales. La troisième image possède la particularité de recouvrir toutes les orientations dans toutes les directions, ce qui est cohérent avec la scène photographiée. Enfin la dernière image possède une majorité d'orientations verticales. 39

3.9 Distribution des distances au sein d'une collection contenant 3 classes d'images uni-couleur de 50 images chacune générées artificiellement par des lois normales tri-dimensionnelles. L'histogramme représenté ici se focalise sur une seule classe "violette" et l'on peut observer les distances "intra-classes" entre les membres de cette classe, et les distances "inter-classes" entre les membres de la classe et ceux des autres classes. Dans ce cas idéal, un simple seuillage sur les distances permettrait d'isoler facilement la classe des images "violette". 40

3.10 Distribution des distances entre les descripteurs couleurs basées sur l'espace couleur {L,a,b} intra et inter-classe pour la classe "montagnes" d'une collection d'images Corel. Dans cet exemple plus réaliste, il est difficile d'isoler les membres de la classe "montagnes" des autres images si l'on s'appuie directement sur les distances. 41

3.11 Conversion numérique symbolique : une distance $d(u, l_q^0)$ entre les descripteurs d'une image non étiquetée u , et d'une image étiquetée l_q^0 appartenant à une classe C_q permet d'établir une distribution de masses m^{Ω_q} . Cette distribution décrit les croyances sur l'appartenance ou non de l'image non étiquetée u à la classe C_q 50

3.12 Fusion des distributions de masses apportées par les différents témoignages des membres d'une même classe pour une seule image non étiquetée u 52

3.13 Situation où un seul voisin est réellement proche de l'échantillon u à classer vis-à-vis des autres voisins d'une classe C_1 . La modélisation de la connaissance avec la règle conjonctive permet de maintenir une croyance élevée sur l'appartenance à la classe dans la distribution de masses m^{Ω_1} . En filigrane sont représentés des images étiquetées dans une autre classe C_2 , rappelant que même s'ils sont proches de l'image non étiquetée u , ils ne sont pas pris en compte pour la distribution de masse m^{Ω_1} relative à la classe C_1 53

3.14 Fonctions triangles pour calculer la nouvelle distribution de masses $m_s^{\Omega_q}$: en vert la fonction pour calculer la masse sur la proposition H_q , en rouge celle pour \overline{H}_q , et en jaune celle associée au doute (H_q, \overline{H}_q) 54

3.15 Chaque classe (respectivement C_1, C_2, \dots, C_Q) fournit une distribution de masses (respectivement $m^{\Omega_1}, m^{\Omega_2}, \dots, m^{\Omega_Q}$) à partir des ses k plus proches voisins respectifs vis-à-vis d'une image non étiquetée u . Ensuite une étape de fusion consiste à combiner ces distributions de masses calculées sur les Q cadres de discernement indépendants. Le but est de synthétiser la connaissance dans un cadre de discernement global Ω , afin de préparer les données pour le niveau *pignistique*. 56

3.16 Les cadres de discernement Ω_1, Ω_2 et Ω_3 sont vus comme des raffinements d'un nouvel espace produit Ω 58

3.17 Représentation ensembliste des hypothèses du cadre de discernement global Ω . Chacune de ses hypothèses est une intersection des hypothèses de base $H_1, \overline{H_1}, H_2, \overline{H_2}, H_3$ et $\overline{H_3}$ issues des trois cadres de discernement Ω_1, Ω_2 et Ω_3 59

3.18 Représentation ensembliste des propositions de la distribution de masses de Ω (remarque : la proposition $(\Omega_1, \Omega_2, \Omega_3)$ n'est pas représentée). 60

3.19 Fusion des distributions de masses apportées par différents descripteurs. On retrouve les 3 échelles d'analyse décrites dans ce chapitre : des distributions de masses sont données localement pour chaque knn pour chaque classe et pour chaque type de descripteur. Puis elles sont fusionnées pour chaque type de descripteur pour considérer l'ensemble des classes. Enfin, ces fusions multi-classe sont combinées entre elles pour établir une fusion tardive de tous les témoignages apportés selon les différents descripteurs. 62

4.1 Ce chapitre se focalise sur le module de sélection active des images. Ce module s'appuie sur les fonctions de croyance calculées précédemment afin d'élaborer différentes stratégies de sélections d'images non étiquetées. Les images sont ainsi ordonnées dans une ou plusieurs listes pour être présentées à l'utilisateur dans l'interface graphique. 67

4.2 Illustration de la mise à jour de la connaissance : dans ce cas très simple, les échantillons non étiquetés (croix) sont triés par une mesure d'éloignement des échantillons étiquetés. A l'itération 1, le plus éloigné est noté "1" et le moins éloigné est noté "5". L'utilisateur étiquette l'échantillon le plus éloigné. A l'itération 2, la connaissance a changé : par exemple, l'échantillon le plus éloigné, noté maintenant "1" était à l'itération précédente celui qui était le moins éloigné. 72

4.3 Transformation pignistique dans le cas à 3 classes. Toutes les propositions de la distribution de masses sont utilisées pour calculer la distribution de probabilités pignistiques des hypothèses du cadre de discernement Ω 74

4.4 Illustration de la sélection d'échantillons par une stratégie du plus positif (MP). Le placement des images est corrélé avec des mesures de similarités visuelles. Les images exemples numérotées "0" ont été étiquetées par l'utilisateur pour initialiser 3 classes distinctes : "mer", "ville" et "portrait". La couleur des cadres des images quant à elle correspond à une vérité terrain établie par un utilisateur. Les images non étiquetées sont numérotées dans l'ordre croissant de la plus à la moins positive. Les images non étiquetées les plus proches des images exemples des classes sont les plus positives. L'image la plus éloignée (numérotée 8) est la plus visuellement différente des images exemples et sera certainement sélectionnée en dernier. 79

4.5 Illustration de la sélection d'échantillons par une stratégie du plus rejeté (MR). A la première itération, l'image sélectionnée la première, notée "1", est la plus éloignée de toutes les images membres des classes notées "0". A l'inverse les images les moins rejetées sont celles qui sont les plus proches des images exemples des classes (notées "6", "7" et "8"). 80

4.6 Illustration de la sélection d'échantillons par une stratégie du plus globalement ambigu. L'image la plus centrale, située près du centre de gravité du triangle formé par les 3 images étiquetées "0", est celle qui est sélectionnée la première par la stratégie MGA. . . 81

4.7 Illustration de la sélection d'échantillons par une stratégie du plus localement ambigu entre 2 classes (MLA_2). Les images situées entre 2 membres de 2 classes distinctes sont sélectionnées en priorité. 83

4.8 Illustration de la sélection d'échantillons par une stratégie du plus incertain avec la mesure de non-spécificité. Nous pouvons remarquer que l'image numérotée 8 est la plus engagée. 85

4.9 Illustration de la sélection d'échantillons par une stratégie du plus en conflit selon la fusion de descripteurs visuels distincts. Les figures du haut "couleur" et "orientations" indiquent pour chaque image non étiquetée le type d'hypothèse donnant le maximum de probabilités pignistiques : P pour une hypothèse positive, R pour l'hypothèse de rejet, et A_3 pour l'hypothèse d'ambiguïté globale et A_2 pour les hypothèses d'ambiguïtés locales. Dans la figure, après fusion des informations apportées par les descripteurs, les images sont numérotées par ordre croissant, de la plus à la moins en conflit. Les images les plus en conflit peuvent être des images qui étaient pourtant identifiées par les autres stratégies comme étant indifféremment la plus positive ou la plus rejetée, ou la plus ambiguë. . . . 86

4.10	Exemple d'adaptation locale des zones de connaissances autour des images étiquetées : cette figure représente une collection d'images étiquetées par 6 étiquettes. Les images sont placées dans l'espace 2D à partir d'une méthode de réduction de dimension s'appuyant sur les vecteurs descripteurs couleur dans l'espace $\{L,u,v\}$, c'est-à-dire que les images sont placées de telle manière à respecter le mieux possible les proximités réelles dans l'espace de description. La taille des images est corrélée avec leur propre paramètre $\sigma_{i,q}^s$, tout comme la forme des fonctions de croyance (les "halos" colorés). Chaque couleur d'un "halo" correspond à une étiquette.	94
5.1	Ce chapitre en cours se focalise sur le dernier module concernant les échanges entre l'utilisateur et le système à travers une interface homme-machine. Ce module s'appuie sur les fonctions de croyance afin de proposer des étiquettes sur des images sélectionnées par le module précédent. L'interface permet à l'utilisateur de contraindre ces propositions en autorisant ou non les cas de multi-étiquetage, et des cas de rejet en distance ou en ambiguïté. La vue principale donne un aperçu de l'état de connaissance des classes et est complètement interactive de telle manière que l'utilisateur peut définir et redéfinir les classes si besoin.	96
5.2	Comparaison du nombre d'hypothèses considérées pour choisir une proposition d'étiquetage d'une image selon les 6 schémas de décision. La limitation à 3 classes ambiguës contraint l'explosion combinatoire du nombre d'hypothèses de décision.	103
5.3	Vue globale de l'interface.	104
5.4	Le panneau de contrôle	106
5.5	Etiquetage multiple : l'image sélectionnée est dupliquée pour pouvoir être intégrée dans les classes pointées par les flèches. Dans ce cas une image représentant un portrait sur un fond de monument historique est recommandée pour la classe contenant des portraits et la classe contenant des images de ville. Si cette proposition de classement ne satisfait pas l'utilisateur il peut soit déplacer les flèches vers les classes qu'il pense être plus pertinentes, soit ajouter une flèche supplémentaire vers une troisième classe en cliquant sur la zone colorée devant la classe, ou bien encore soit supprimer une des 2 flèches en déplaçant le bout de la flèche dans une zone blanche.	108
5.6	Proposition de rejet en distance sur une image sélectionnée : selon l'état de la connaissance l'image au centre de la figure ne correspond à aucune des classes proposées. L'utilisateur a activé les options de rejets en distance et en ambiguïté. Le système propose alors à l'utilisateur de mettre l'image de côté dans la liste des images rejetées.	111

5.7 Stratégie orientée "classe" : plusieurs listes de propositions sont créées en face de chaque liste de classe. Les images à l'intérieur de ces listes sont triées en interne de la plus à la moins probable. L'utilisateur peut cliquer sur les triangles verts pour valider ces propositions de classement image par image, ou par petit lots de 5 images. Deux listes d'images en haut affichent les images qui sont rejetées en distance, et celles qui le sont par ambiguïté. 112

5.8 Représentation de la connaissance pour afficher explicitement les ambiguïtés : 5 classes sont représentées par une seule image chacune (de grande taille). Les images membres sont disposées sur un premier cercle, et les images rejetées sont disposées sur un deuxième cercle extérieur. Les images ambiguës sont situées à l'intérieur du cercle et leur placement est déduit en fonction du barycentre des probabilités pignistiques sur les hypothèses positives. Ainsi, l'utilisateur peut percevoir dans quelles mesures une image non étiquetée est ambiguë et avec quelles classes. 113

6.1 Les trois axes d'évaluations étudiés dans ce chapitre : les performances de classification automatique, la caractérisation des stratégies de sélections actives d'images, et enfin une évaluation utilisateur. 116

6.2 Pourcentage des images correctement classées automatiquement sur le jeu de test *gt1000* en utilisant différentes métriques comme mesure de dissimilarité entre les descripteurs des orientations des images, et avec pour paramètres internes $f = 0,7$ et $k = 5$. La distance de Bhattacharyya permet d'obtenir le meilleur pourcentage avec 65,6% des images correctement classées. 120

6.3 Pourcentage des images correctement classées automatiquement sur le jeu de test *gt1000* en utilisant différentes métriques sur différents descripteurs couleurs, avec pour paramètres internes $f = 0,7$ et $k = 5$. Les distances L_1 et de Bhattacharyya permettent d'obtenir des meilleurs performances de classification automatique du même ordre avec les espaces couleurs $\{H,s,v\}$, $\{L,u,v\}$, et $\{R,g,b\}$ 120

6.4 Pourcentage des images correctement classées automatiquement sur le jeu de test *gt1000* en utilisant différentes métriques sur différents descripteurs couleurs, et avec pour paramètres internes $f = 0,8$ et $k = 5$. La distance de Bhattacharyya permet de maintenir des performances élevées tandis que les autres mesures voient leurs performances pratiquement toutes se détériorer, vis-à-vis du cas où $f = 0,7$ 121

6.5 Non-spécificité moyenne des distributions de masses sur la base *gt1000*, avec $f = 0.7$ sur les mêmes combinaisons de distances-descripteurs que la figure 6.3 122

6.6 Non-spécificité moyenne des distributions de masses sur la base *gt1000*, avec $f = 0.7$ sur les mêmes combinaisons de distances-descripteurs que la figure 6.4 122

6.7	Comparaison des performances de classification en fonction de f pour différentes métrique avec leurs descripteurs permettant d'atteindre le pourcentage le plus élevé.	123
6.8	Comparaison des performances de classification en fonction de k dans le cas où la distance de Bhattacharyya est appliquée sur le descripteur couleur {H,s,v} et avec $f = 0.8$ sur la base <i>gt1000</i> . Il existe une valeur de k optimale autour de $k = 5$ permettant d'atteindre des performances de classification automatique au dessus de 80%.	124
6.9	Pourcentage d'images classées correctement selon différentes combinaisons du descripteur des orientations avec les descripteurs couleurs, et avec différentes métriques. Sur chaque barre d'histogramme est ajouté en rouge le pourcentage gagné par rapport aux tests précédent figure 6.3 sur les descripteurs couleurs seuls effectués dans les mêmes condition ($f = 0, 7$).	126
6.10	Comparaison des approches de fusion tardive et précoce : à gauche sont données les performances de classification avec la distance de Bhattacharyya pour $f = 0, 7$ et $k = 5$ sur le jeu de test <i>gt1000</i> pour toutes les combinaisons des descripteurs couleurs avec le descripteur des orientations. A droite sont indiquées les non-spécificités moyennes des distributions de masses associées.	127
6.11	Mesure du conflit moyen des distributions de masses avec la fusion tardive sur les mêmes tests que la figure 6.10. Il n'est pas possible d'avoir ce type d'information dans le cas de la fusion précoce.	127
6.12	Performances de classification et conflit moyen en fonction du paramètre f pour les distances de Bhattacharyya et L_1 sur les mêmes fusions de descripteurs orientations et {R,g,b} sur la base <i>gt1000</i>	129
6.13	Influence de la limitation du nombre de propositions d'une distribution de masses par la cardinalité maximum autorisée, en termes de performances de classification automatique et de temps de calculs. A titre indicatif, sont représentés le nombre de propositions dans la distribution de masses, et la non-spécificité moyenne de toutes les distributions de masses de toutes les images non étiquetées.	132
6.14	Comparaison des différentes méthodes de classification multi-classe et multi-étiquette évaluées dans [TK07]. Notre méthode notée "KnnEvMulti" est représentée par la croix orange.	134
6.15	Comparaison des performances de classification avant et après apprentissage actif sur le jeu de données <i>gt1000</i> . Avant la sélection active (en vert), 72,4% des images peuvent être classées correctement. Après la sélection active de 50 images supplémentaires (en rouge) les performances de classification sont modifiées : la stratégie aléatoire n'apporte pas de gain, les stratégies MP, MLA2 et détériore les performances, tandis que les autres l'augmentent. La stratégie la plus "rentable" est celle du plus rejeté MR.	137

6.16 Cas d'une sélection d'un échantillon très localement ambiguë entre 2 classes (croix entourée en rouge). Cet échantillon est sélectionné pour être étiqueté par l'utilisateur, soit dans la classe "rond" ou la classe "triangle" (voir même la classe "croix pleine"). Or, cet échantillon est isolé et n'apporte que très peu d'information "utile" pour classer les autres échantillons non étiquetés (représenté également par des croix). 138

6.17 Courbes d'évaluations comparant la sélection aléatoire (Rand) avec la stratégie du plus positif (MP). Les 2 stratégies ne provoquent pas les mauvaises propositions d'étiquettes aux mêmes instants de sélection et ne donnent pas le même nombre final d'images proposées de manière incorrectes dans les classes. 140

6.18 Comparaisons de toutes les stratégies sur le jeu de données *gt500* dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace {R,g,b}, $f = 0,6$ et $k = 5$). Les différentes stratégies ne donnent pas toutes le même nombre de mauvaises propositions d'étiquettes au final, et ne provoquent pas ces propositions incorrectes aux mêmes moments. 142

6.19 Comparaisons de toutes les stratégies en nombre d'étiquettes non proposées manquantes sur le jeu de données *gt772* dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace {L,a,b}, $f = 0,825$ et $k = 5$). 147

6.20 Comparaisons de toutes les stratégies en nombre d'étiquettes proposées en trop car incorrectes sur le jeu de données *gt772* dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace {L,a,b}, $f = 0,825$ et $k = 5$). 147

6.21 Exemple d'images de la collection *Borgia*. 149

6.22 Impression d'écran de l'interface durant l'évaluation de l'outil sur le corpus *Borgia*. La documentaliste a utilisé régulièrement la liste de rejet (première liste horizontale du haut). Elle a organisé la collection avec les thèmes suivants : "baiser", "direction d'acteurs", "vatican", "combats", "tournage", "chevaux", "champs de bataille", "portraits". 151

6.23 Stratégie du plus positif MP sur les 3 combinaisons d'information suivante : "Contenu Visuel" la combinaison des descripteurs couleurs et orientations, "temps" l'information de date de prise de cliché EXIF, "fusion" la combinaison des 3 informations de temps, de couleurs et des orientations. La complémentarité des informations permet un gain très important. L'utilisation des descripteurs visuels permet de faire au final 70,89% propositions d'étiquettes correctes, tandis que le temps seul en permet 73,75%. La combinaison de toutes les informations permet de diviser quasiment par 2 le nombre de mauvaises propositions pour atteindre un pourcentage de 84,79% de classements corrects. 156

6.24	Stratégie du plus rejeté MR dans les mêmes conditions que l'expérience que la figure précédente 6.23. La combinaison permet d'atteindre un gain du même ordre que la stratégie MP. De plus, la combinaison des 3 sources d'information permet une généralisation rapide : environ 80% des mauvaises propositions de classements sont réalisées dans le premier tiers des images sélectionnées.	157
6.25	Les deux premières imagettes dans le temps du journal télévisé de France2 du 20h du 13 juin 1997 servant à initialiser deux classes : "présentateur" et "non présentateur".	159
6.26	Courbes d'évaluations comparant les stratégies de sélection du plus positif MP, du plus rejeté MR, du plus globalement ambigu MGA, et du plus incertain MU. L'homogénéité des contenus visuels permet d'atteindre de très bonnes performances de classification : en particulier la stratégie MU permet de segmenter la vidéo sans aucune erreur.	159
6.27	Visualisation du journal télévisé de France2 du 20h du 13 juin 1997 reflétant la structure narrative. Les imagettes du présentateur seul sont positionnées dans l'ordre chronologique de gauche à droite. Chaque boucle représente un reportage ou une autre interruption avec le présentateur (interview, graphiques. . .). Les images étant échantillonnées régulièrement dans le temps, la longueur renseigne sur la durée des reportages et des interruptions. Le lecteur en haut à droite illustre que dans une version intégrée on peut naviguer dans le contenu par le biais des timecodes des imagettes.	160
6.28	Courbes comparant les stratégies de sélection du plus positif MP, du plus rejeté MR, du plus globalement ambigu MGA, et du plus incertain MU. L'homogénéité des contenus visuels permet d'atteindre de très bonnes performances de classification : en particulier la stratégie MU permet de segmenter la vidéo sans aucune erreur.	161
6.29	Visualisation d'un épisode d'une série télévisée "Friends" contenant 2 fils narratifs.	162

LISTE DES TABLEAUX

- 2.1 Extrait d'une collection d'images "homogène" : les images peuvent être facilement regroupées en trois classes visuellement et sémantiquement homogènes pour tout utilisateur. Chaque image est clairement une illustration sans ambiguïté d'une classe. 17
- 2.2 Organisation d'une collection d'images visuellement hétérogène. Les classes possèdent chacune des images ayant des contenus visuels dissemblables. De plus, certaines images aux contenus visuels proches ne sont pas associées dans les mêmes classes par l'utilisateur, comme par exemple les images A3 et C2 ou les images B2 et C4, ou encore A1 et C1. Cette structuration est justifiée par le choix de l'utilisateur de distinguer les photographies avec Gilbert Bécaux, et celles pendant et hors tournage des scènes. 17
- 2.3 Une structuration d'une collection d'images visuellement hétérogène selon un deuxième utilisateur sur le même extrait de collection d'images de la figure 2.2. L'utilisateur est contraint d'identifier les photographies nécessitant l'ouverture de droits. Il a regroupé toutes les photographies contenant des représentations de peintures célèbres et Gilbert Bécaud. L'image C2 contient à la fois une reproduction des Demoiselles d'Avignon de Picasso avec une incrustation du visage de Gilbert Bécaux dans le tableau. La dernière classe regroupe toutes les images ne posant pas de problème de droits *a priori*. Les contenus d'une même classe sont visuellement hétérogènes, et certaines images similaires ne se retrouvent pas dans les mêmes classes comme par exemple les images A4 et B2. 19
- 2.4 Exemple d'étiquetage simple sur un extrait d'une collection de photographies personnelles. L'utilisateur considère qu'une image possède un sens dominant. Pour lui, une image est une illustration d'une seule étiquette. Par exemple, la photographie C2 est identifiée comme un portrait, indépendamment des autres éléments qui composent l'image comme le fond de mer. 20

2.5	Exemple de multi-étiquetage sur le même extrait de la collection précédente (tableau 2.4). Un second utilisateur est plus attentif à la composition des images (sujet photographié, fond, éléments de décors...). Les images peuvent alors posséder une ou plusieurs étiquettes. Par exemple, l'image A1 est associée à une seule étiquette car le contenu visuel et sémantique est relativement homogène. Par contre les images A3 et C2 représentent la même photographie associées à deux classes distinctes, prenant en compte le sujet photographié (une personne) et le fond (la mer). De même la photographie représentant les quais d'une ville (A4 et B3) est associée à la fois à une classe "ville" et "mer". Enfin, l'images en B4 et C5 est associée également aux deux étiquettes "ville" et "portrait". . . .	21
3.1	Exemple de répartition des étiquettes multiples dans une collection d'images de 231 photographies avec 3 étiquettes de bases. Les répartitions sont déséquilibrées : les images possédant plusieurs étiquettes sont beaucoup moins nombreuses que celles possédant une seule étiquette. L'enjeu du problème de classification est alors d'arriver à retrouver 266 étiquettes pour 231 images.	32
3.2	Exemple d'un jeu de données multi-étiquette : les échantillons peuvent avoir une étiquette comme l'échantillon 1, ou plusieurs comme pour les trois autres échantillons. . . .	32
3.3	Transformation du jeu de données 3.2 avec l'opération de transformation "PT3" : les unions des étiquettes de base sont considérées comme des étiquettes à part entière pour pouvoir traiter un problème de classification multi-classe avec étiquetage simple.	33
3.4	Transformation du jeu de données 3.2 avec l'opération de transformation "PT4" : le problème est traité par un ensemble de classifieurs binaires, un par étiquette. C'est l'approche la plus souvent utilisée.	33
3.5	Transformation du jeu de données 3.2 avec l'opération de transformation "PT6" : le problème est traité conjointement par un ensemble de classifieurs binaires "faibles", un pour chaque possibilité d'étiquetage de chaque échantillon exemple. L'étiquetage d'un nouvel échantillon peut alors se faire à partir du signe ou d'un score donné par l'ensemble des sorties de tous les classifieurs.	33
3.6	Intersections des propositions de deux distributions de masses pour la règle de combinaison conjonctive.	45
3.7	Les quatre règles expertes pour établir une distribution de masses m_q^Ω dans le cadre de discernement Ω_q à partir d'une distance d mesurée entre deux vecteurs descripteurs d'images.	49
3.8	Interprétations de deux propositions dans le cadre de discernement de la méthode des knn évidentiel et dans le cadre de discernement proposé, dans un cas à 3 classes C_1 , C_2 et C_3	57

3.9	Quelques exemples de calcul de masses sur 6 propositions de la distribution de masses du cadre de discernement Ω dans le cas à 3 classes.	60
4.1	Exemple de propositions de la distribution de masses m^Ω . La notation complète indique explicitement les hypothèses de Ω contenant la proposition.	75
5.1	Dénominations des 6 schémas d'étiquetages. L'utilisateur peut choisir entre étiquetage simple (S) ou multi-étiquette (M) en activant éventuellement des options de rejet en ambiguïté (A) ou de rejet en distance (D).	97
5.2	Nombre de classement possibles pour une image non étiquetée u , en fonction du nombre de classes Q disponibles pour les 6 schémas de décision. Plus le nombre de classes est élevé, et plus le risque de mauvaise proposition automatique d'étiquette est élevé. L'autorisation des rejets augmente le risque de proposition d'étiquetage non pertinente pour l'utilisateur.	103
6.1	Limitation du nombre de propositions d'une distribution de masses en fonction de la cardinalité maximum autorisée.	130
6.2	Comparaisons pour 4 valeurs distinctes de f ($f = 0, 4, f = 0, 6, f = 0, 7$ et $f = 0, 8$) de toutes les stratégies sur le jeu de données <i>gt500</i> dans les mêmes conditions d'expérimentations (distance d_{Bhat} , descripteurs orientations et couleurs dans l'espace {R,g,b} et $k = 5$). Certaines stratégies reproduisent les mêmes tendances, tandis que d'autres peuvent complètement changer.	144

BIBLIOGRAPHIE

- [ACD⁺03] John ADCOCK, Matthew COOPER, John DOHERTY, Jonathan FOOTE, Andreas GIRGENSOHN et Lynn WILCOX : Managing digital memories with the fxpal photo application. *In MULTIMEDIA '03 : Proceedings of the eleventh ACM international conference on Multimedia*, pages 598–599, New York, NY, USA, 2003. ACM.
- [AKA91] David W. AHA, Dennis KIBLER et Marc K. ALBERT : Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.
- [BEYL04] Yoram BARAM, Ran EL-YANIV et Kobi LUZ : Online choice of active learning algorithms. *The Journal of Machine Learning Research*, 5:255–291, 2004.
- [Bis06] Christopher M. BISHOP : *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, August 2006.
- [BL07] A. BONDU et V. LEMAIRE : Etat de l'art sur les méthodes statistiques d'apprentissage actif. *Revue des Nouvelles Technologies de l'Information, Numéro spécial sur l'apprentissage et la fouille de données*, 2007.
- [BLSB04] M. R. BOUTELL, J. LUO, X. SHEN et C. M. BROWN : Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771, 2004.
- [BMnM07] Anna BOSCH, Xavier MUÑOZ et Robert MARTÍ : Which is the best way to organize/classify images by content ? *Image and Vision Computing*, 25(6):778–791, June 2007.
- [BR05] Stanley T. BIRCHFIELD et Sriram RANGARAJAN : Spatiograms versus histograms for region-based tracking. *In CVPR '05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 1158–1163, Washington, DC, USA, 2005. IEEE Computer Society.
- [Bri03] Klaus BRINKER : Incorporating diversity in active learning with support vector machines. *In Proceedings of the 20th International Conference on Machine Learning*, pages 59–66. AAAI Press, 2003.
- [CAL94] David COHN, Les ATLAS et Richard LADNER : Improving generalization with active learning. *Mach. Learn.*, 15(2):201–221, 1994.
- [CE91] Bonwel C. et J. EISON : Active learning : Creating excitement in the classroom. AEHE-ERIC Higher Education N°1, 1991.
- [CFB04] Michel CRUCIANU, Marin FERECATU et Nozha BOUJEMAA : Relevance feedback for image retrieval : a short survey. Report of the DELOS2 European Network of Excellence (FP6), 2004.

- [CFGW05] Matthew COOPER, Jonathan FOOTE, Andreas GIRGENSOHN et Lynn WILCOX : Temporal event clustering for digital photo collections. *ACM Trans. Multimedia Comput. Commun. Appl.*, 1(3):269–288, 2005.
- [CL01] Chih-Chung CHANG et Chih-Jen LIN : *LIBSVM : a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [CWX⁺07] Jingyu CUI, Fang WEN, Rong XIAO, Yuandong TIAN et Xiaoou TANG : Easyalbum : an interactive photo annotation system based on face clustering and re-ranking. *In CHI '07 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 367–376, New York, NY, USA, 2007. ACM.
- [Den08] Thierry DENÈUX : A k -nearest neighbor classification rule based on dempster-shafer theory. *In* Ronald R. YAGER et Liping LIU, éditeurs : *Classic Works of the Dempster-Shafer Theory of Belief Functions*, volume 219 de *Studies in Fuzziness and Soft Computing*, pages 737–760. Springer, 2008.
- [DHS00] Richard O. DUDA, Peter E. HART et David G. STORK : *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.
- [Dia78] Persi DIACONIS : [a mathematical theory of evidence. (glenn shafer)]. *Journal of the American Statistical Association*, 73(363):677–678, 1978.
- [DK05] Kai-Bo DUAN et Sathiya S. KEERTHI : Which is the best multiclass svm method ? an empirical study. pages 278–285. 2005.
- [dl86] Commission Internationale de L'ECLAIRAGE : Colorimetry. Technical Report, Bureau Central de la CIE, 2nd edition, 1986.
- [DM96] A. DILLON et M.MORRIS : User acceptance of information technology : theories and models. *Annual review of information science and technology*, 1996.
- [DP86] D. DUBOIS et H. PRADE : A set-theoretic view of belief functions : Logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12:193–226, 1986.
- [DP88] D. DUBOIS et H. PRADE : *Théorie des possibilités, Application à la représentation des connaissances en informatique*. Edition Masson, 1988.
- [DV03] Chitra DORAI et Svetha VENKATESH : Guest editors' introduction : Bridging the semantic gap with computational media aesthetics. *IEEE MultiMedia*, 10(2):15–17, 2003.
- [Fed] Digital Library FEDERATION : Providing leadership for libraries. <http://www.diglib.org/>.
- [FR91] T. M. J. FRUCHTERMAN et E. M. RHEINGOLD : Graph drawing by forcedirected placement. *In Software - Practice and Experience*, volume 21, pages 1129–1164, 1991.
- [FST97] Yoav FREUND, Eli SHAMIR et Naftali TISHBY : Selective sampling using the query by committee algorithm. *In Machine Learning*, pages 133–168, 1997.

- [GC04] Philippe H. GOSSELIN et Matthieu CORD : A comparison of active classification methods for content-based image retrieval. In *CVDB '04 : Proceedings of the 1st international workshop on Computer vision meets databases*, pages 51–58, New York, NY, USA, 2004. ACM.
- [Gha04] Zoubin GHAHRAMANI : Unsupervised learning. *Advanced Lectures in Machine Learning Lecture Notes in Computer Science*, pages 72–112, 2004.
- [Gou] Pierre GOUGELET : Xnview. <http://www.xnview.com/>.
- [Gue07] Y. GUERMEUR : *SVM Multiclasses, Théorie et Applications*. Hdr, Université Nancy 1, 2007.
- [Har79] R. M. HARALICK : Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.
- [HB07] Nicolas HERVÉ et Nozha BOUJEMAA : Image annotation : which approach for realistic databases ? In *CIVR '07 : Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 170–177, New York, NY, USA, 2007. ACM.
- [Inca] Google INC. : Picasa. <http://picasa.google.fr/>.
- [Incb] Adobe Systems INCORPORATED : Adobe photoshop elements 7. <http://www.adobe.com/fr/products/photoshopelwin/>.
- [IPT] IPTC : The iptc-naa standards. <http://www.controlledvocabulary.com/>.
- [IV96] Jukka IIVARINEN et Ari VISA : Shape recognition of irregular objects. In *Intelligent Robots and Computer Vision XV : Algorithms, Techniques, Active Vision, and Materials Handling, Proc. SPIE 2904*, pages 25–32, 1996.
- [JEI] JEITA : Exif image format. <http://www.exif.org/>.
- [JL95] George H. JOHN et Pat LANGLEY : Estimating continuous distributions in bayesian classifiers. pages 338–345, 1995.
- [JR99] M. I. JORDAN et S. RUSSEL : Categorization. *The MIT Encyclopedia of the Cognitive Sciences*, pages 104–106, 1999.
- [JV96] Anil K. JAIN et Aditya VAILAYA : Image retrieval using color and shape. *Pattern Recognition*, 29:1233–1244, 1996.
- [KGOG06] E. KIJAK, G. GRAVIER, L. OISEL et P. GROS : Audiovisual integration for tennis broadcast structuring. *Multimedia Tools and Applications*, 30:289–311, 2006.
- [KSLC07] Deok-Hwan KIM, Jae-Won SONG, Ju-Hong LEE et Bum-Ghi CHOI : Support vector machine learning for region-based image retrieval with relevance feedback. *ETRI Journal*, 29(5):700–702, 2007.
- [LC94] David D. LEWIS et Jason CATLETT : Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of the Eleventh International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann, 1994.

- [LC07] Loïc LECERF et Boris CHIDLOVSKII : Apprentissage actif pour l'annotation de documents. In *Quatrième Conférence francophone en Recherche d'Information et Applications - CORIA07*, Saint Etienne, 2007.
- [Lee96] Tai Sing LEE : Image representation using 2d gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–971, 1996.
- [Lew95] J. R. LEWIS : Ibm computer usability satisfaction questionnaires : Psychometric evaluation and instructions for use. *Int. J. Hum.-Comput. Interact.*, 7(1):57–78, 1995.
- [LHZ⁺00] Ye LU, Chunhui HU, Xingquan ZHU, HongJiang ZHANG et Qiang YANG : A unified framework for semantics and feature based relevance feedback in image retrieval systems. In *MULTIMEDIA '00 : Proceedings of the eighth ACM international conference on Multimedia*, pages 31–37, New York, NY, USA, 2000. ACM.
- [LL03] Suryani LIM et Guojun LU : Effectiveness and efficiency of six colour spaces for content based image retrieval. In *International Workshop on Content-Based Multimedia Indexing, CBMI*, 2003.
- [LMZ99] C. S. LEE, W.-Y. MA et H. ZHANG : Information embedding based on user's relevance feedback for image retrieval. In S. PANCHANATHAN, S.-F. CHANG et C.-C. J. KUO, éditeurs : *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 3846 de *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 294–304, août 1999.
- [MCS06] Jonathan MILGRAM, Mohamed CHERIET et Robert SABOURIN : One against one or one against all : which one is better for handwriting recognition with svms. *tenth International Workshop on Frontiers in Handwriting Recognition*, 2006.
- [Mil56] G. MILLER : The magical number seven, plus or minus two : Some limits on our capacity for processing information. *The Psychological Review*, 63, 1956.
- [MM04] Prem MELVILLE et Raymond J. MOONEY : Diverse ensembles for active learning. In *ICML '04 : Proceedings of the twenty-first international conference on Machine learning*, page 74, New York, NY, USA, 2004. ACM.
- [NDH09] Stefanie NOWAK, Peter DUNKER et Mark HUISKES : Imageclef large scale visual concept detection and annotation task, 2009.
- [NG08] Xavier NATUREL et Patrick GROS : Detecting repeats for video structuring. *Multimedia Tools Appl.*, 38(2):233–252, 2008.
- [Nie92] Jakob NIELSEN : Evaluating the thinking-aloud technique for use by computer scientists. *Advances in human-computer interaction (vol. 3)*, pages 69–82, 1992.
- [Nie93] J. NIELSEN : Usability engineering. Academic Press, Boston, 1993.
- [NS04] Hieu T. NGUYEN et Arnold SMEULDERS : Active learning using pre-clustering. In *ICML '04 : Proceedings of the twenty-first international conference on Machine learning*, page 79, New York, NY, USA, 2004. ACM.

- [NSPGM04] Mor NAAMAN, Yee Jiun SONG, Andreas PAEPCKE et Hector GARCIA-MOLINA : Automatic organization for digital photographs with geographic coordinates. In *JCDL '04 : Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, pages 53–62, New York, NY, USA, 2004. ACM.
- [OKS05] Thomas OSUGI, Deng KUN et Stephen SCOTT : Balancing exploration and exploitation : A new algorithm for active machine learning. In *ICDM '05 : Proceedings of the Fifth IEEE International Conference on Data Mining*, pages 330–337, Washington, DC, USA, 2005. IEEE Computer Society.
- [PG04] A. PIGEAU et M. GELGON : Organisation statistique spatio-temporelle d'une collection d'images acquises d'un terminal mobile géolocalisé. In *Congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'2004)*, pages 76–84, January 2004.
- [Pic] PICAJET.COM : Picajet. <http://www.picajet.com/>.
- [Pla99] John C. PLATT : Fast training of support vector machines using sequential minimal optimization. pages 185–208, 1999.
- [Pol07] Jean-Philippe POLI : *Structuration automatique de flux télévisuels*. Thèse de doctorat, Université Paul Cézanne - Aix-Marseille III, 2007.
- [QHR⁺08] Guo-Jun QI, Xian-Sheng HUA, Yong RUI, Jinhui TANG et Hong-Jiang ZHANG : Two-dimensional multi-label active learning with an efficient online adaptation model for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(1), 2008.
- [Qui93] J. Ross QUINLAN : *C4.5 : programs for machine learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- [Ram07] Emmanuel RAMASSO : *Reconnaissance de séquence d'états par le Modèle des Croyances Transférable - Application à l'analyse de vidéos d'athlétisme*. Thèse de doctorat, Université Joseph Fourier de Grenoble, 2007.
- [RK04] Ryan RIFKIN et Aldebaro KLAUTAU : In defense of one-vs-all classification. *J. Mach. Learn. Res.*, 5:101–141, 2004.
- [RM01] Nicholas ROY et Andrew MCCALLUM : Toward optimal active learning through sampling estimation of error reduction. In *ICML '01 : Proceedings of the Eighteenth International Conference on Machine Learning*, pages 441–448, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- [RMD⁺07] REGE, MANJEET, DONG, MING, FOTOUHI et FARSHAD : Building a user-centered semantic hierarchy in image databases. *Multimedia Systems*, 12(4-5):325–338, March 2007.
- [Ros73] Eleanor ROSCH : On the internal structure of perceptual and semantic categories. *T.E. Moore (Ed.), Cognitive Development and the Acquisition of Language*, 1973.
- [RRP07] E. RAMASSO, M. ROMBAUT et D. PELLERIN : State filtering and change detection using tbm conflict - application to human action recognition in athletics videos. *IEEE Transaction on Circuits and Systems for Video Technology*, 17(7):944–949, 2007.

- [Sen90] B. SENACH : Evaluation ergonomique des interfaces homme-machine : une revue de la littérature. Research report, INRIA, 1990.
- [Sha91] B. SHACKEL : Usability - context, framework, design and evaluation. Human Factors for Informatics Usability, Shackel, B. and Richardson, S, Cambridge University Press, Cambridge, 1991.
- [SJRU99] Beyer Kevin S., Goldstein JONATHAN, Ramakrishnan RAGHU et Shaft URI : When is "nearest neighbor" meaningful ? *In ICDT '99 : Proceedings of the 7th International Conference on Database Theory*, pages 217–235, London, UK, 1999. Springer-Verlag.
- [Sme93] Philippe SMETS : Belief functions : The disjunctive rule of combination and the generalized bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35, August 1993.
- [Sme94] Philippe SMETS : The transferable belief model. *Artif. Intell.*, 66(2):191–234, 1994.
- [Sme05] Philippe SMETS : Decision making in the tbm : the necessity of the pignistic transformation. *Journal of Approximate Reasoning*, 2005.
- [SOK06] Alan F. SMEATON, Paul OVER et Wessel KRAAIJ : Evaluation campaigns and trecvid. *In MIR '06 : Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 321–330, New York, NY, USA, 2006. ACM.
- [SOS92] H. S. SEUNG, M. OPPER et H. SOMPOLINSKY : Query by committee. *In COLT '92 : Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294, New York, NY, USA, 1992. ACM.
- [SS00] Robert E. SCHAPIRE et Yoram SINGER : Boostexter : A boosting-based system for text categorization. *Machine Learning*, 39(2/3):135–168, 2000.
- [SWS⁺00] Arnold W. M. SMEULDERS, Marcel WORRING, Simone SANTINI, Amarnath GUPTA et Ramesh JAIN : Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000.
- [SWS05] Cees G. M. SNOEK, Marcel WORRING et Arnold W. M. SMEULDERS : Early versus late fusion in semantic video analysis. *In MULTIMEDIA '05 : Proceedings of the 13th annual ACM international conference on Multimedia*, pages 399–402, New York, NY, USA, 2005. ACM.
- [sys] ACD SYSTEMS : Acdsee. <http://www.acdsee.com/>.
- [TC01] Simon TONG et Edward CHANG : Support vector machine active learning for image retrieval. *In ACM Multimedia*, pages 107–118, New York, NY, USA, 2001. ACM Press.
- [TFMB04] A. TRÉMEAU, C. FERNANDEZ-MALOIGNE et P. BONTON : *Image numérique couleur. De l'acquisition au traitement*. Dunod, 2004. Parmi les auteurs et contributeurs : S. Philipp-Foliguet, M. Cord.
- [Thi06] Jérôme THIÈVRE : *Cartographies pour la Recherche et l'Exploration de données Documentaires*. Thèse de doctorat, Université Montpellier II, Sciences et Techniques du Languedoc, 2006.

- [TK07] Grigorios TSOUMAKAS et Ioannis KATAKIS : Katakis i : Multi-label classification : An overview. *International Journal of Data Warehousing and Mining*, pages 1–13, 2007.
- [TKV08] Grigorios TSOUMAKAS, Ioannis KATAKIS et Ioannis VLAHAVAS : Multi-label classification bibliography, August 2008.
- [TM92] Sebastian THRUN et K. MOELLER : Active exploration in dynamic environments. *In Advances in Neural Information Processing Systems 4*. Morgan Kaufmann, 1992.
- [TPSC⁺03] A. TRICOT, F. PLEGAT-SOUTJIS, J-F. CAMPS, A.AMIEL, G.LUTZ et A. MORCILLO : Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des eiah. C. Desmoulins, P. Marquet et D.Bouhineau (dir.). Environnements informatiques pour l'apprentissage humain, Paris : ATIEF - INRP, 2003.
- [UM06] Jana URBAN et Joemon M.JOSE : Evaluating a workspace's usefulness for image retrieval. 2006.
- [VBTGV07] Anne VERROUST-BLONDET, Jérôme THIÈVRE, Hervé GOËAU et Marie-Luce VIAUD : Report on the state-of-the-art on advanced visualisation methods. (D7.2) Report of Vitalas project (Video and image Indexing and reTrievAl in the LArge Scale), 2007.
- [VJ01] Paul VIOLA et Michael JONES : Rapid object detection using a boosted cascade of simple features. *In IEEE Computer Vision and Pattern Recognition, CVPR 2001*, pages 511–518, 2001.
- [VSWB06] Julia VOGEL, Adrian SCHWANINGER, Christian WALLRAVEN et Heinrich H. BÜLTHOFF : Categorization of natural scenes : local vs. global information. *In APGV '06 : Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 33–40, New York, NY, USA, 2006. ACM.
- [Wes] Mario M. WESTPHAL : Imatch. <http://www.photools.com/>.
- [WKBD06] Yi WU, Igor KOZINTSEV, Jean-Yves BOUGUET et Carole DULONG : Sampling strategies for active learning in personal photo retrieval. *In Proceedings of the 2006 IEEE International Conference on Multimedia and Expo, ICME 2006*, pages 529–532, Toronto, Ontario, Canada, July 2006.
- [WLW01] J. Z. WANG, J. LI et G. WIEDERHOLD : SIMPLicity : semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:947–963, 2001.
- [XAZ07] Zuobing XU, Ram AKELLA et Yi ZHANG : Incorporating diversity and density in active learning for relevance feedback. pages 246–257. 2007.
- [XZ08] Jun XIAO et Tong ZHANG : Face bubble : photo browsing by faces. *In AVI '08 : Proceedings of the working conference on Advanced visual interfaces*, pages 343–346, New York, NY, USA, 2008. ACM.
- [YHB08] Itheri YAHIAOUI, Nicolas HERVÉ et Nozha BOUJEMAA : Shape-based image retrieval in botanical collections. 4261, 2008.

BIBLIOGRAPHIE

- [Zhu05] Xiaojin ZHU : Semi-supervised learning literature survey. Rapport technique 1530, Computer Sciences, University of Wisconsin-Madison, 2005.
- [ZL01] Dengsheng ZHANG et Guojun LU : Shape retrieval using fourier descriptors. *In In Proceedings of 2nd IEEE Pacific Rim Conference on Multimedia*, pages 1–9. Springer, 2001.