UNIVERSITE D'EVRY-VAL D'ESSONNE

Laboratoire d'Informatique, Biologie Intégrative et Systèmes Complexes



THÈSE

Présentée par

Rostom KACHOURI

Pour l'obtention du titre de

Docteur de l'Université d'Evry-Val d'Essonne

Spécialité: Sciences pour l'ingénieur - Traitement des IMages et du SIgnal

Classification multi-modèles des images dans les bases hétérogènes

Thèse soutenue le 29 juin 2010

Devant le jury composé de:

S. Ruan	: Professeur, Université de Reims,	Rapporteur
W. Puech	: Professeur, Université de Montpellier II,	Rapporteur
K. Hamrouni	: MCF, ENIT-Université de Tunis,	Rapporteur
A. Ben Hamida	: Professeur, ENIS-Université de Sfax,	Examinateur
D. S-Masmoudi	: MCF, ENIS-Université de Sfax,	Examinateur
K. Djemal	: MCF, Université d'Evry Val d'Essonne,	Encadrant
N. Derbel	: Professeur, ENIS-Université de Sfax,	Co-directeur
H. Maaref	: Professeur, Université d'Evry Val d'Essonne,	Directeur

Dédicaces

À l'âme de mon père Sadok
À ma mère Monia
Auxquels, je dois ce que je suis

À Imen
À notre fille Rahma
À notre futur garçon Rayan

Remerciements

J'aimerais avant tout exprimer ma gratitude et mes remerciements à mon encadrant scientifique de thèse, M. Khalifa Djemal, maître de conférences de l'université d'Evry. J'ai énormément appris à ses côtés. Il a su m'initier à la recherche, me pousser à toujours faire mieux et il m'a souvent aidé à surmonter les difficultés de ce cheminement qu'est le doctorat. Encore plus que ses grandes qualités scientifiques, j'ai beaucoup apprécié ses qualités humaines, en particulier l'écoute, le partage et la compréhension, qui m'en ont fait un ami.

J'adresse également à mon directeur de thèse, M. Hichem Maaref, professeur de l'université d'Evry, mes plus sincères remerciements. C'est un homme d'une grande patience, très à l'écoute et très compréhensif et qui est directement responsable du bon déroulement de mes travaux. Je le remercie chaleureusement pour ses attentions et les nombreuses discussions professionnelles et personnelles que nous avons eues.

Je veux absolument remercier M. Nabil Derbel, professeur de l'université de Sfax et Mme Dorra Sellami-Masmoudi, maître de conférences à l'École nationale des ingénieurs de Sfax et habilitée à diriger les recherches, sans qui cette collaboration de thèse en cotutelle ne serait pas ce qu'elle est aujourd'hui.

J'exprime ma gratitude à M. William Puech, professeur de l'université de Montpellier II ainsi qu'à Mme Su Ruan, professeur de l'université de Reims, pour avoir accepté de juger mes travaux en tant que rapporteurs. Je remercie M. AHMED BEN HMIDA, professeur de l'université de Sfax, qui a accepté de faire partie de ce jury. Je tiens également à exprimer ma reconnaissance à M. Kamel Hamrouni, maître de conférences à l'École nationale des ingénieurs de Tunis et habilité à diriger les recherches, pour avoir accepté de juger mes travaux et présider ce jury.

Une thèse est un travail assez personnel qui s'inscrit toutefois dans une équipe. Je remercie donc tous mes collègues doctorants, tous les membres temporaires ou permanents que j'ai croisé au cours de cette expérience, je pense en particulier à Imen, Ahmed, Mohamed, Slim, Amin, Imen, Anwar, Anis, Asma, Salima, Iman, Mouna, christophe, Jean-yves, christophe, Pierre, Hamdi, Abdelkader, Samia, Nader, Majid, yassine, Khalid (ils sont trop nombreux pour pouvoir les citer tous). Ils m'ont beaucoup apporté pendant ces dernières années: joie, fous rires, complicité, soirées, voyages, ouverture d'esprit.

Comme il y a aussi une vie en dehors du travail (si tout de même un petit peu) j'ai eu la chance de rencontrer hors du labo des personnes qui sont aujourd'hui des amis: Tarek, Hedi, Mehdi, Hamdi, Sabeur, Khalil, Sebastien, Nader et tant d'autres! Auxquels je dois aussi de bons moments de sincérité, rigolade et complicité.

Enfin *−last but not least−* j'exprime toute ma gratitude à ma mère Monia, ma femme Imen et ma petite fille Rahma qui ont supporté avec moi les meilleurs et les pires moments et qui m'ont donné la force d'aller au bout de cette aventure. Bref, pour ce bout de chemin fait ensemble.

Résumé

La reconnaissance d'images est un domaine de recherche qui a été largement étudié par la communauté scientifique. Les travaux proposés dans ce cadre s'adressent principalement aux diverses applications des systèmes de vision par ordinateur et à la catégorisation des images issues de plusieurs sources. Dans cette thèse, on s'intéresse particulièrement aux systèmes de reconnaissance d'images par le contenu dans les bases hétérogènes. Les images dans ce type de bases appartiennent à différents concepts et représentent un contenu hétérogène. Pour ce faire, une large description permettant d'assurer une représentation fiable est souvent requise. Cependant, les caractéristiques extraites ne sont pas nécessairement toutes appropriées pour la discrimination des différentes classes d'images qui existent dans une base donnée d'images. D'où, la nécessité de sélection des caractéristiques pertinentes selon le contenu de chaque base. Dans ce travail, une méthode originale de sélection adaptative est proposée. Cette méthode permet de considérer uniquement les caractéristiques qui sont jugées comme les mieux adaptées au contenu de la base d'images utilisée. Par ailleurs, les caractéristiques sélectionnées ne disposent pas généralement des mêmes performances. En conséquence, l'utilisation d'un algorithme de classification, qui s'adapte aux pouvoirs discriminants des différentes caractéristiques sélectionnées par rapport au contenu de la base d'images utilisée, est vivement recommandée. Dans ce contexte, l'approche d'apprentissage par noyaux multiples est étudiée et une amélioration des méthodes de pondération des noyaux est présentée. Cette approche s'avère incapable de décrire les relations non-linéaires des différents types de description. Ainsi, nous proposons une nouvelle méthode de classification hiérarchique multi-modèles permettant d'assurer une combinaison plus flexible des caractéristiques multiples. D'après les expérimentations réalisées, cette nouvelle méthode de classification assure des taux de reconnaissance très intéressants. Enfin, les performances de la méthode proposée sont mises en évidence à travers une comparaison avec un ensemble d'approches cité dans la littérature récente du domaine.

Mots-clefs: Reconnaissance d'images par le contenu, Bases d'images hétérogènes, Extraction d'attributs, Sélection adaptative des caractéristiques les plus pertinentes, Apprentissage par noyaux multiples, Classification multi-modèles.

Abstract

Image recognition is widely studied by the scientific community. The proposed research in this field is addressed to various applications of computer vision systems and multiple source image categorization. This PhD dissertation deals particularly with content based image recognition systems in heterogeneous databases. Images in this kind of databases belong to different concepts and represent a heterogeneous content. In this case and to ensure a reliable representation, a broad description is often required. However, the extracted features are not necessarily always suitable for the considered image database. Hence, the need of selecting relevant features based on the content of each database. In this work, an adaptive selection method is proposed. It considers only the most adapted features according to the used image database content. Moreover, selected features do not have generally the same performance degrees. Consequently, a specific classification algorithm which considers the discrimination powers of the different selected features is strongly recommended. In this context, the multiple kernel learning approach is studied and an improved kernel weighting method is presented. It proved that this approach is unable to describe the nonlinear relationships of different description kinds. Thus, we propose a new hierarchical multi-model classification method able to ensure a more flexible combination of multiple features. Experimental results confirm the effectiveness and the robustness of this new classification approach. In addition, the proposed method is very competitive in comparison with a set of approaches cited in the recent literature.

Key words: Content based image recognition, Heterogeneous image databases, Feature extraction, Adaptive feature selection, Multiple kernel learning, Multi-model classification.

Table des matières

Dédicaces

Remerciements

Résumé

Abstract

Index des symboles **1**

Introduction générale **6**

1 Extraction d'attributs et classification d'images dans les bases hétérogènes 10

1.1 Introduction 11

1.2 Reconnaissance d'images par le contenu 11

1.2.1 Types d'attributs 12

1.2.1.1 La couleur 13

1.2.1.2 La texture 14

1.2.1.3 La forme 14

1.2.2 Catégories d'attributs 15

1.2.2.1 Attributs globaux 15

1.2.2.2 Attributs locaux 16

1.2.3 Méthodes de classification et évaluation des performances 17

1.2.3.1 Classification 17

1.2.3.2 Rétroaction et bouclage de pertinence 17

1.2.3.3 Performances de classification d'images 18

1.3 Bases d'images hétérogènes utilisées 18

1.3.1 La base d'images COLUMBIA 19

1.3.2 La base d'images COREL 19

1.3.3 La base d'images CALTECH 20

1.4 Techniques de classification d'images 22

1.4.1 Méthodes de classification non-supervisée 22

1.4.2 Méthodes de classification supervisée 24
1.5 Caractéristiques utilisées pour la description d'images dans les bases hétérogènes 26
1.5.1 Introduction 26
1.5.2 Les couleurs moyennes 27
1.5.3 Les histogrammes couleurs 28
1.5.4 La matrice corrélogramme 28
1.5.5 Les statistiques de premier ordre 29
1.5.6 La matrice de co-occurrences 29
1.5.7 Les transformées en ondelettes de Daubeshie 31
1.5.8 Le descripteur GIST 31
1.5.9 La matrice norme du gradient 31
1.5.10 Les moments de Hu 33
1.5.11 Optimisation 33
1.5.12 Conclusion 34
1.6 Évaluation des caractéristiques 35
1.7 Conclusion 38

2 Sélection Adaptative des caractéristiques les plus pertinentes 40
2.1 Introduction 41
2.2 Méthodes de sélection 41
2.2.1 Les méthodes Filtres 41
2.2.2 Les méthodes Enveloppantes 42
2.3 Travaux connexes 42
2.3.1 L'Analyse en Composantes Principales 42
2.3.2 La sélection basée sur les Algorithmes Génétiques 44
2.4 La Sélection Adaptative des Caractéristiques pertinentes 45
2.4.1 Introduction 45
2.4.2 Apprentissage multiple 45
2.4.3 Discrimination linéaire de Fisher 46
2.4.4 Conclusion 49
2.5 Évaluation de la sélection des caractéristiques par ACP, AG et SAC . . . 49
2.5.1 ACP 49
2.5.2 AG 52
2.5.3 SAC 55
2.5.4 Résultats de classification 57
2.6 Conclusion 59

3 Machines à Vecteurs Supports à noyaux multiples 61
3.1 Introduction 62
3.2 Machines à Vecteurs Supports 62
3.2.1 Introduction 62

3.2.2 Notions de base . . . 63
3.2.3 Cas linéairement séparable . . . 63
3.2.4 Cas non-linéairement séparable . . . 64
3.2.5 Extension aux problèmes multi-classes . . . 65
3.2.5.1 Algorithme SVM: Un contre tous . . . 66
3.2.5.2 Algorithme SVM: Un contre un . . . 66
3.2.5.3 Algorithme DAG-SVM . . . 67
3.2.6 Conclusion . . . 67
3.3 Approche d'apprentissage par noyaux multiples . . . 67
3.3.1 Principe . . . 67
3.3.2 Techniques de pondération de noyaux . . . 68
3.4 Pondération de noyaux basée sur les taux d'apprentissage SVM . . . 70
3.4.1 Introduction . . . 70
3.4.2 Principe de MK_SVM . . . 70
3.4.3 Apprentissage multiple . . . 71
3.4.4 Calcul des différents poids de noyaux . . . 71
3.4.5 Conclusion . . . 73
3.5 Évaluation de la méthode MK_SVM . . . 73
3.5.1 Résultats de classification . . . 74
3.5.2 Temps d'exécution . . . 75
3.6 Conclusion . . . 76

4 Méthode de classification hiérarchique multi-modèles 78
4.1 Introduction . . . 79
4.2 Classification hiérarchique multi-modèles . . . 79
4.2.1 Introduction . . . 79
4.2.2 Principe de l'approche multi-modèles . . . 79
4.2.3 Résolution de conflit de décision par NCC . . . 80
4.2.4 Classification hiérarchique . . . 81
4.2.5 Conclusion . . . 83
4.3 Évaluation et comparaison . . . 83
4.3.1 Évaluation de MC-MM à travers différentes méthodes de sélection 83
4.3.2 Performances de MC-MM utilisant la méthode de sélection SAC . 84
4.3.3 Comparaison des résultats MC-MM basée SAC avec des méthodes de classification connues . . . 86
4.3.3.1 Méthode simultanée classique . . . 86
4.3.3.2 Approche des *Multiple Instance Learning* . . . 88
4.3.3.3 Approche des sacs de mots-visuels . . . 89
4.3.3.4 Approche d'apprentissage par noyaux multiples . . . 92
4.4 Conclusion . . . 96

Conclusion générale et perspectives **98**

Liste des publications **102**

Bibliographie **105**

Table des figures

1.1 Architecture d'un système de reconnaissance d'images par le contenu basé sur la classification. 11
1.2 Image couleur *RVB*: association de trois sous-images monochromes, correspondant respectivement au différentes composantes Rouge, Vert et Bleu. 13
1.3 Différents modèles de texture. 14
1.4 Différentes transformations géométriques que peut subir une image. . . 15
1.5 Échantillons d'images segmentées de la base COREL. 16
1.6 Graphe typique de précision-rappel. 18
1.7 Échantillons des 100 objets de la base d'images COIL-100 de COLUMBIA. 19
1.8 Échantillons d'images de différentes classes de la base COREL utilisée. 20
1.9 Échantillons d'images de différentes classes de la base CALTECH-256 utilisée. 20
1.10 Approche de classification supervisée: type de modélisation a) Intrinsèque et b) Discriminante. 25
1.11 Échantillons des caractéristiques extraites respectivement des images Dinosaure et Rose: a) Histogramme, b) Matrice des normes du gradient, et c) GIST. 27
1.12 Coefficients d'ondelettes pour deux niveaux de décomposition. 31
1.13 Échantillons des descripteurs GIST extraits pour différentes images de la base COREL. 32
1.14 Nouvelle forme de la caractéristique des normes du gradient des image Dinosaure et Rose: a) la projection selon x et b) la projection selon y. . 35
1.15 Taux de classification des modèles individuels associés à l'ensemble des caractéristiques employées pour différentes sous-bases d'images COREL (à gauche) et CALTECH-256 (à droite). 38

2.1 Les deux premières directions principales (b_1 et b_2) d'un nuage de N_{Im} observations dans $\mathbb{R}^d$. 43
2.2 Diagramme standard d'un algorithme génétique. 44
2.3 Diagramme de la méthode de Sélection Adaptative des Caractéristiques les plus pertinentes qui correspondent aux modèles sélectionnés dans $\mathcal{M}_s$. 48

2.4 Allure des valeurs propres rangées par ordre décroissant respectivement pour les sous-bases COREL avec : a) 3 classes, b) 5 classes, c) 7 classes et d) 10 classes. 50
2.5 Allure des valeurs propres rangées par ordre décroissant respectivement pour les sous-bases COLUMBIA avec : a) 3 classes, b) 5 classes, c) 7 classes et d) 10 classes. 51
2.6 Comparaison des précisions moyennes de classification respectivement basées sur les sélections ACP (90%), AG_2 et SAC pour différentes sous-bases d'images COREL. 58

3.1 Exemple de classification SVM linéairement séparable. 64
3.2 Exemple de classification SVM non-linéairement séparable. 65
3.3 Exemple de classification SVM multi-classes ($N_C = 3$). 66
3.4 Diagramme de la méthode de classification MK_SVM utilisant la nouvelle pondération KW-BTR. 72
3.5 Performances de classification moyenne des deux algorithmes MK_SVM et *SimpleMKL* pour différentes sous-bases d'images COREL. 75
3.6 Courbes précision-rappel de a) MK_SVM et b) *SimpleMKL* pour différentes sous-bases d'images COREL. 75

4.1 Règle de décision du classifieur NCC: classification de l'image requête I_q dans l'espace d'attributs $\mathcal{F}_m$, à travers les deux classes évaluées C_i et C_j. 80
4.2 Diagramme illustrant l'architecture de la méthode de classification hiérarchique multi-modèles MC-MM. 82
4.3 a) Les images correctement classifiées et b) celles mal classifiées de la classe *African peoples and villages* pour la sous-base d'images COREL utilisée (10 classes). 86
4.4 Comparaison des précisions moyennes de MC-MM et de la méthode de classification classique pour différentes sous-bases d'images COREL. . . 87
4.5 Les dix histogrammes de mots-visuels qui représentent l'ensemble des classes de la sous-base d'images COREL avec 10 classes. Chaque histogramme illustre le nombre d'occurrence des différents mots-visuels extraits à partir des images de la classe correspondante. 89
4.6 a) Sac de mots-visuels qui représente l'ensemble des images appartenant aux 10 classes de la sous-base COREL utilisée, b) Distances entre les caractéristiques histogrammes de mots-visuels des différentes paires d'images de la sous-bases COREL utilisée (10 classes). 90
4.7 Graphes des deux matrices de confusion respectives de MC-MM (à gauche) et BoF (à droite) pour la sous-base d'images COREL avec 10 classes. 91
4.8 Courbes précision-rappel de MC-MM et BoF pour la sous-base d'images COREL avec 10 classes. 91
4.9 Comparaison des précisions moyennes de MC-MM et *SimpleMKL* pour différentes sous-bases d'images a) COREL et b) CALTECH-256. 93

4.10 Courbes précision-rappel de a) MC-MM et b) *SimpleMKL* pour différentes sous-bases d'images COREL. 93
4.11 Courbes précision-rappel de a) MC-MM et b) *SimpleMKL* pour différentes sous-bases d'images CALTECH-256. 93
4.12 Les images requêtes de la classe *Sunset scenes* de la base COREL (à gauche) et les images mal-classifiées (à droite) avec respectivement MC-MM et *SimpleMKL*. Toutes les images requêtes restantes (non illustrées à droite) sont correctement classifiées à travers chaque méthode. 94
4.13 Les images requêtes de la classe *Chimp* de la base CALTECH-256 (à gauche) et les images mal-classifiées (à droite) avec respectivement MC-MM et *SimpleMKL*. Toutes les images requêtes restantes (non illustrées à droite) sont correctement classifiées à travers chaque méthode. 94

Liste des tableaux

1.1 Noms et identifiants des 23 classes utilisées à partir des deux bases d'images COREL et CALTECH-256. 21
1.2 Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques couleurs pour différentes sous-bases COREL. 36
1.3 Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques textures et formes pour différentes sous-bases COREL. 36
1.4 Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques couleurs pour différentes sous-bases COLUMBIA. . . 37
1.5 Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques textures et formes pour différentes sous-bases COLUMBIA. 37

2.1 Variation de la dimension des composantes principales sélectionnées avec la méthode de sélection ACP en fonction de différentes sous-bases d'images COREL. 50
2.2 Variation de la dimension des composantes principales sélectionnées avec la méthode de sélection ACP en fonction de différentes sous-bases d'images COLUMBIA. 51
2.3 Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 3 classes. 53
2.4 Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 5 classes. 53
2.5 Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 7 classes. 53
2.6 Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 10 classes. 53
2.7 Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 3 classes. 54
2.8 Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 5 classes. 54
2.9 Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 7 classes. 54
2.10 Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 10 classes. 55

2.11 Variation des paramètres SAC pour différentes sous-bases COREL. . . . 56
2.12 Variation des paramètres SAC pour différentes sous-bases COLUMBIA. . 56
2.13 Résultats de classification SVM basée sur la sélection ACP pour différentes sous-bases d'images COREL. 57
2.14 Résultats de classification SVM basée sur la sélection ACP pour différentes sous-bases d'images COLUMBIA. 57
2.15 Résultats de classification SVM basée ACP (90%), AG_2 et SAC pour différentes sous-bases d'images COREL. 58

3.1 Exemple de pondération de noyaux KW-BTR pour la sous-base COREL avec 5 classes. 74
3.2 Comparaison des résultats de classification MK_SVM et *SimpleMKL* pour différentes sous-bases d'images COREL. 74
3.3 Comparaison des temps d'apprentissage et de pondération (en Minutes) des deux algorithmes MK_SVM et *SimpleMKL* pour différentes sous-bases COREL. 76

4.1 Classification MC-MM à ravers différentes méthodes de sélection pour quatre sous-bases d'images COREL. 84
4.2 Évaluation de la généralisation MC-MM. Étude comparative avec les deux procédures de généralisation *décroissante* et *max_modèles* pour différentes sous-bases COREL. 85
4.3 Matrice de confusion de la classification MC-MM pour la sous-base COREL avec 10 classes. 85
4.4 Évaluation des deux méthodes de classification MC-MM et classique pour différentes sous-bases d'images COREL. 87
4.5 Comparaison de la la précision moyenne de classification de la méthode MC-MM avec celles de différentes méthodes *MIL* de l'état de l'art pour la sous-base COREL avec 10 classes. 88
4.6 Étude comparative des deux méthodes MC-MM et *SimpleMKL* pour différentes sous-bases d'images COREL et CALTECH-256. 92
4.7 Comparaison des temps d'apprentissage (en Minutes) de *MC-MM* et *SimpleMKL* pour différentes sous-bases COREL et CALTECH-256. 95

Index des symboles

Symboles	**Significations**
P	Nombre total des images pertinentes dans une base.
TP	Taux des vrai positifs, i.e. nombre d'images pertinentes retrou vées.
FP	Taux des faux positifs, i.e. nombre d'images non-pertinentes retrouvées.
$\mathcal{I}$	Image couleur.
R	Matrice couleur Rouge de $\mathcal{I}$ dans l'espace couleur RVB.
V	Matrice couleur Verte de $\mathcal{I}$ dans l'espace couleur RVB.
B	Matrice couleur Bleue de $\mathcal{I}$ dans l'espace couleur RVB.
p	Pixel de l'image $\mathcal{I}$.
N	Nombre de pixels p dans l'image $\mathcal{I}$.
$c = \mathcal{I}(p)$	Couleur d'un pixel p dans l'image $\mathcal{I}$.
cl	Nombre de couleurs dans l'image $\mathcal{I}$.
RVB	Caractéristique couleur moyenne de l'image $\mathcal{I}$, extraite à partir de l'espace couleur RVB.
HSV	Caractéristique couleur moyenne de l'image $\mathcal{I}$, extraite à partir de l'espace couleur HSV.
HMMD	Caractéristique couleur moyenne de l'image $\mathcal{I}$, extraite à partir de l'espace couleur $HMMD$.
YCrCb	Caractéristique couleur moyenne de l'image $\mathcal{I}$, extraite à partir de l'espace couleur $YCrCb$.
Hist	Caractéristique histogrammes couleurs de l'image $\mathcal{I}$, extraite à partir de l'espace couleur RVB.
L_∞	Norme infinie.
$Dist$	Distance inter-pixel dans une image.
Cor	Matrice corrélogramme de l'image $\mathcal{I}$.
A-Cor	Caractéristique auto-correlogramme de l'image $\mathcal{I}$.
NG	Image niveau de gris correspondante à l'image $\mathcal{I}$.
ng	Nombre de niveaux de gris dans l'image NG.
$\bar{M}_{oy}$	Moyenne de la distribution des niveaux de gris de l'image NG.
$\bar{E}_{type}$	Écart type de la distribution des niveaux de gris de l'image NG.
$\bar{V}_{ar}$	Variance de la distribution des niveaux de gris de l'image NG.
FOS	Caractéristique statistiques de premiers ordres de l'image $\mathcal{I}$.
Dir	Direction inter-pixel dans une image.
$SGLD$	Matrice de co-occurences de l'image $\mathcal{I}$.
SGLD	Ensemble des attributs extraits de la matrice de co-occurence

	SGLD de l'image $\mathcal{I}$: {*Moyenne*, *Energie*, *Variance*, *Contraste*, *Entropie*, *Homogeneite* & *Correlation*}.
c_{ij}^{l}	Coefficients d'ondelettes de Daubeshie de l'image $\mathcal{I}$ pour un niveau de décomposition l, i,$j \in$ [0,1].
h_i; h_j	Filtres de Daubeshie selon les lignes et les colones.
h_{LAP}	Matrice de convolution du filtre Laplacien.
e_{ij}^{l}	Coefficient condensé de la transformation d'ondelette de Daubeshie de l'image $\mathcal{I}$ pour un niveau de décomposition l, i,$j \in$ [0,1].
Daub	Caractéristique coefficients condensés de la transformation d'ondelette de Daubeshie de l'image $\mathcal{I}$ à deux niveaux de décomposition.
GIST	Descripteur GIST de l'image $\mathcal{I}$.
G_X; G_Y	Gradients directionnels selon x et y (approximations des dérivées horizontale et verticale) de l'image NG.
h_X; h_Y	Masques de *Sobel*, i.e. matrices de convolution horizontale et verticale.
G	Matrice normes du gradient correspondante à G_X et G_Y.
P_X	Projection de G selon X.
P_Y	Projection de G selon Y.
G-norm	Caractéristique projection de la matrice normes du gradient de l'image $\mathcal{I}$ selon X et Y.
$\phi_{i=1,2,...,7}$	Moments invariants de Hu, calculés à partir de l'image $\mathcal{I}$.
Hu	Caractéristique des moments de Hu de l'image $\mathcal{I}$.
$\mathcal{B}_{app}$	Base d'apprentissage.
N_{Im}	Nombre d'images d'apprentissage dans la base $\mathcal{B}_{app}$.
x^i	Vecteur caractéristique de dimension d, i = [1,2,...,n].
n	Nombre de caractéristiques x^i, i = [1,2,...,n].
$\mathcal{F}_i = \mathbb{R}^d$	Espace d'attributs de la caractéristique x^i, i = [1,2,...,n].
$X_i = [x_1^i, x_2^i, ..., x_{N_{Im}}^i]$	Nuage des N_{Im} points, dits aussi individus ou observations dans l'espace d'attribut $\mathcal{F}_i$, i = [1,2,...,n].
$\bar{X}$	Caractéristique moyenne d'un nuage d'observations X.
$\mathbf{C}$	Matrice de covariance d'un nuage d'observations X.
d_{x_j}	Déformation principales de l'observation x_j par rapport à la forme moyenne $\bar{X}$.
Ψ	Base modale, correspond à la diagonalisation de la matrice de covariance $\mathbf{C}$.
λ_i	Vecteur propre de la matrice de covariance $\mathbf{C}$, i = [1,2,...,d].
Ψ_p	Sous-matrice de Ψ contenant les p premiers vecteurs propres de la matrice de covariance $\mathbf{C}$.
b_i	Composante principale d'un nuage d'observations X, i = [1,2, ...,p].
$\mathbb{B}$	Vecteur des amplitudes modales de deformations, représente une forme dans l'espace p-dimensionnel défini par les composantes principales retenues.
λ_T	Inertie totale correspondante à la matrice de covariance $\mathbf{C}$.
τ	Pourcentage de variance relatif aux p premiers modes, i.e. le taux d'inertie total correspondant.
N_C	Nombre de classes contenues dans $\mathcal{B}_{App}$.

$\mathcal{C} = \{C_1,C_2,...,C_{N_C}\}$	Ensemble des N_C classes contenues dans $\mathcal{B}_{App}$.
$\mathcal{M} = \{M_1,M_2,...,M_n\}$	Espace des modèles correspondants aux caractéristiques extraites à partir de la base d'images $\mathcal{B}_{app}$, i.e. $x^{i=1,2,...,n}$.
$N_{Err}(M_i)$	Nombre d'images mal classifiées (ou erronées) selon le modèle M_i, $i = [1,2,...,n]$.
$N_{Corr}(M_i)$	Nombre d'images correctement classifiées selon le modèle M_i, $i = [1,2,...,n]$.
$E(M_i)$	Taux d'erreur de classification du modèle M_i, $i = [1,2,...,n]$.
$Per(M_i)$	Taux de classification en phase d'apprentissage (performance d'apprentissage) du modèle M_i, $i = [1,2,...,n]$.
$\mathcal{T} = \{Per(M_{s1}),Per(M_{s2}),...,Per(M_{sn})\}$	Performances d'apprentissage des différents modèles dans $\mathcal{M}$, organisées par ordre de pertinence décroissant.
$m_1(i)$	Valeur moyenne des performances d'apprentissage situées à gauche de la performance d'apprentissage $Per(M_{si})$ dans la liste $\mathcal{T}$, $i = [1,2,...,n]$.
$m_2(i)$	Valeur moyenne des performances d'apprentissage situées à droite de la performance d'apprentissage $Per(M_{si})$ dans la liste $\mathcal{T}$, $i = [1,2,...,n]$.
$v_1^2(i)$; $v_2^2(i)$	Variances respectives à $m_1(i)$; $m_2(i)$, $i = [1,2,...,n]$.
$P(i)$	Discriminant de Fisher.
F_S	Seuil adaptatif de sélection des modèles pertinents.
$\mathcal{T}_s = \{Per(M_{s1}),Per(M_{s2}),...,Per(M_{sk})\}$	Liste des performances d'apprentissage correspondantes aux modèles sélectionnés.
$\mathcal{M}_s = \{M_{s1},M_{s2},...,M_{sk}\}$	Sous-ensemble des modèles sélectionnés correspondants aux caractéristiques les plus pertinentes.
$Y = [y_1,y_2,...,y_i,...,y_{N_{Im}}]$	Ensemble des étiquettes correspondantes aux images d'apprentissage; pour un problème de classification binaire, $y_i \in \{-1,1\}$.
$f(x)$	Fonction de décision SVM.
α_i; b	Hyper-paramètres d'apprentissage SVM, $i = [1,2,...,N_{Im}]$.
x	Vecteur support SVM.
Φ	Transformation non-linéaire de $\mathcal{F}$ dans un espace de Hilbert $\mathcal{H}$; $Card(\mathcal{H}) > Card(\mathcal{F})$.
$K(x,x_i) = \langle\Phi(x),\Phi(x_i)\rangle_{\mathcal{H}}$	Fonction noyau, $i = [1,2,...,N_{Im}]$.
ct, u, θ et σ	Paramètres de noyaux.
M_{od_1}	Modèle permettant de séparer les deux classes C_1 et C_2 dans l'algorithme *DAG-SVM*.
M	Nombre de noyaux dans l'approche MKL.
β_m	Poids du noyau k_m, $m = [1,2,...,M]$.
ξ_i	Variables molles (slack variables en anglais).
C	Terme de régularisation de marge SVM.
(L_2, L_1)	Normes mixtes.
L_2	Norme quadratique.
$S_m(\boldsymbol{\alpha})$	Fonction différentiable de (Sonnenburg et al., 2006).
δ	Valeur optimale d'un sous-ensemble de contraintes selon le programme d'optimisation de (Sonnenburg et al., 2006).
$\mathcal{K} = [k_1,k_2,...,k_M]$	Ensemble des M noyaux de base.
$N_{Corr}(k_m)$	Nombre d'images correctement classifiées avec le noyau k_m, $m = [1,2,...,M]$.
TR_{k_m}	Taux d'apprentissage correspondant au noyau k_m, $m = [1,2,$

	...,M].
$\mathcal{TR}$ = [TR_{k_1},TR_{k_2},..., TR_{k_M}]	Ensemble des taux d'apprentissage TR_{k_m}, m = [1,2,...,M].
$\mathcal{K}^{'}$ = [$k_1^{'}$,$k_2^{'}$,...,$k_M^{'}$]	Organisation des noyaux de l'ensemble $\mathcal{K}$ selon l'ordre des taux d'apprentissage décroissant.
$\boldsymbol{\beta}$ = [β_1,β_2,...,β_M]	Ensemble des poids correspondants aux noyaux $k_m^{'}$, m = [1,2, ...,M].
ζ_{C_i}	Centres de la classe C_i dans un espace d'attributs $\mathcal{F}$, i = [1,2, ...,N_C].
Niv_l	Niveau hiérarchique l de la méthode MC-MM, l = [1,2,...,k-1]; k est le nombre de modèles sélectionnés par la méthode SAC.
I_q	Image requête.
$\mathcal{F}_m(I_q)$	Projection de I_q dans l'espace $\mathcal{F}_m$.
d_i	Distance Euclidienne entre ζ_{C_i} et $\mathcal{F}(I_q)$, i = [1,2,...,N_C].
$\mathcal{B}_{Test}$	Base d'images de test.
$C^{M_{sk}}(I_q) \in \mathcal{C}$	Classification de l'image I_q avec le modèle M_{sk}.
$C^{M_{sk},M_{sk-1}}(I_q) \in \mathcal{C}$	Classification de l'image I_q avec les modèles M_{sk}, M_{sk-1}.
$C(I_q)$ = $C^{M_{sk},M_{sk-1},...,M_{s1}}(I_q) \in \mathcal{C}$	Classification de l'image I_q selon la méthode MC-MM.

Introduction générale

Contexte et motivations

La classification d'images dans les bases hétérogènes suscite de plus en plus d'intérêt dans le domaine de reconnaissance d'images. En effet, de nombreuses applications des systèmes de vision par ordinateur, comme les interactions Homme-Machine, la détection des événements, le contrôle des processus et d'autres applications spécifiques à un domaine donné, ont fait naître, conjointement au développement croissant des images numériques issues du Web, un grand besoin pour mettre au point des techniques de classification d'images plus efficaces. Toutefois, les performances de classification des images dans les bases hétérogènes dépendent principalement de la qualité de description de ces images.

Au début, les techniques de description utilisées étaient, généralement, textuelles. Dans ce cas, un ensemble de mots-clés est associé manuellement à chaque image. Ce procédé de description utilise beaucoup de sémantique de par la description des images par un humain. Une expertise humaine, permettant d'assurer une certaine corrélation entre le texte et les images, est exigée. Il est cependant subjectif à cause du très grand nombre d'interprétations pouvant être attribuées au contenu d'une même image. Ainsi, le problème majeur est que les résultats de reconnaissance d'images basée sur une description par mots-clés peuvent être complètement hors sujet. D'autant plus, lorsqu'il s'agit de base hétérogènes, où le contenu des images est assez varié et difficile à décrire. Par ailleurs, l'association de mots-clés à l'image est une démarche réaliste pour des petites bases d'images, mais elle est complètement impensable et impraticable pour des grandes bases. En raison de ces limitations inhérentes aux techniques de description textuelle, leur utilisation devient de plus en plus difficile dans les bases hétérogènes, gérant une très grande collection d'images.

De nouvelles techniques de description basées sur le contenu même des images sont apparues et présentent de nos jours un intérêt croissant. Contrairement aux méthodes d'annotation textuelle, ces nouvelles techniques permettent l'accès aux images via leurs propriétés visuelles: informations qui peuvent être tirées de l'image elle-même telles que la couleur, la texture, la forme, etc... Ces propriétés visuelles sont extraites sous formes d'ensembles d'attributs appelés souvent caractéristiques. Par conséquent, l'architecture type d'un système de reconnaissance d'images par le contenu, désigné généralement par l'acronyme anglais CBIR (*Content Based Image Recognition*), repose principalement sur deux étapes primordiales: l'*Extraction d'attributs visuels* pour la description des images et la *Classification* qui mène à la reconnaissance de ces images. Dans ce cadre, plusieurs approches ont été proposées pour développer soit des techniques d'extraction d'attributs soit des méthodes de classification supervisées ou non supervisées.

Dans le cas des bases hétérogènes, dites aussi généralistes, les images sont de divers types et leur contenu est assez varié. De ce fait, les caractéristiques pertinentes pour la description d'une catégorie d'images ne sont pas nécessairement efficaces pour d'autres.

Généralement face à ce problème, des caractéristiques de différents types (dites aussi hétérogènes) sont employées pour assurer une description robuste des images. Cette large description couvre, généralement, la majorité des caractéristiques visuelles des images dans les bases hétérogènes. Cependant, elle souffre essentiellement d'une grande dimensionnalité. De plus, selon la base d'image considérée, une partie de cette description est souvent non adaptée au contenu de la base et par conséquent inefficace lors de la phase de reconnaissance. D'où, la nécessité de sélection des caractéristiques pertinentes selon le contenu de la base d'images considérée. En outre, généralement les caractéristiques sélectionnées ne disposent pas des mêmes performances de discrimination. Considérant l'effet négatif des caractéristiques les moins performantes, une simple concaténation de toutes les caractéristiques sélectionnées ne mène pas toujours à des résultats de classification optimaux. En conséquence, l'utilisation d'un algorithme de classification, qui s'adapte aux pouvoirs discriminants des différentes caractéristiques sélectionnées ainsi que le contenu de la base d'images utilisée, est bien recommandée. Cela permet d'améliorer davantage les taux de reconnaissance finaux.

Dans ce travail nous nous intéressons à la reconnaissance d'images par le contenu dans les bases d'images hétérogènes. La phase de reconnaissance est totalement automatique et adaptée au contenu de la base considérée. Aucune rétroaction de l'utilisateur pour indiquer un bouclage de pertinence n'est envisagée. Le système étudié se base ainsi sur une construction hors ligne d'une large description d'images, qui repose sur une variété de caractéristiques hétérogènes. Ce sujet de recherche était et reste un problème ouvert et difficile à cause de la grande difficulté de description des images dans ce type de bases et la complexité de combinaison des caractéristiques hétérogènes. Dans cette thèse, nous contribuons à la résolution de quelques problématiques liées aux systèmes de reconnaissance d'images hétérogènes par le contenu. Une méthode de sélection adaptative des caractéristiques les plus pertinentes et une méthode de classification hiérarchique multi-modèles sont proposées.

Contribution et plan du manuscrit

Les travaux effectués dans le cadre de cette thèse portent sur l'extraction d'attributs hétérogènes, l'optimisation de la description et/ou la dimension des caractéristiques, la sélection des caractéristiques les plus pertinentes et le développement d'une nouvelle approche de classification d'images hétérogènes. L'originalité de notre travail réside principalement dans la sélection adaptative des caractéristiques les plus pertinentes et la classification hiérarchique multi-modèles. Le manuscrit de la thèse est ainsi organisé en quatre chapitres, de la manière suivante :

Dans le premier chapitre, les éléments fondamentaux d'un système de reconnaissance d'images par le contenu sont présentés. Nous discutons ainsi les bases d'images et leurs contenus, les types d'attributs existants en littérature, les deux catégories de classification et les mesures de pertinence dédiées aux systèmes CBIR. En particulier, les bases d'images hétérogènes employées dans cette thèse sont exposées. Suite à une présentation de l'état-de-l'art des différentes techniques de classification qui existent dans la littérature, l'ensemble des différentes caractéristiques hétérogènes retenues est détaillé. Pour améliorer les performances de reconnaissance basée sur cet ensemble de caractéristiques, une op-

timisation de certaines d'entres elles est également introduite. Une large évaluation de ces caractéristiques, après optimisation, est fournie à la fin de ce chapitre. Cette étude montre que le pouvoir discriminant de chaque caractéristique dépend du contenu de la base d'images considérée.

Le deuxième chapitre est consacrée à la sélection des caractéristiques les plus pertinentes. Initialement, les propriétés des deux catégories de sélection: *Filtres* et *Enveloppantes* sont rappelées. Les méthodes de sélection les plus utilisées, en particulier la technique linéaire d'analyse en composantes principales et la sélection basée sur les algorithmes génétiques sont décrites. Une étude singulière des fonctions de sélectivité employées par ces algorithmes est aussi effectuée. Ensuite, une nouvelle méthode de sélection adaptative des caractéristiques les plus pertinentes est proposée. L'évaluation des performances de cette méthode de sélection est ensuite présentée. Une comparaison avec les différentes méthodes de sélection traitées, dans ce chapitre, est également fournie.

Dans le troisième chapitre, l'approche d'apprentissage par noyaux multiples connue par l'acronyme anglais MKL (*Multiple Kernel Learning*) appliquée à la reconnaissance des images hétérogènes par le contenu est abordée. Dans cette thèse, l'approche MKL est considérée en particulier pour les Machines à Vecteurs Supports (SVM). Suite à une brève exposition des notions de bases des SVMs, l'approche MKL est présentée. En tenant compte de l'importance de l'étape de calcul des poids de noyaux dans cette approche, une nouvelle méthode de pondération des noyaux est proposée. Une évaluation de la classification SVM à noyaux multiples basée sur la nouvelle pondération, ainsi qu'une comparaison avec un algorithme récent, sont fournies à la fin de ce chapitre.

Le quatrième chapitre est consacré à la proposition d'une nouvelle méthode de classification hiérarchique multi-modèles. Une évaluation de cette nouvelle méthode de classification en se basant sur la sélection adaptative des caractéristiques, proposée dans le chapitre 2 est effectuée. Des comparaisons avec la méthode de classification classique (employant un modèle unique), l'approche des *Multiple Instance Learning* ainsi que l'approche des sacs de mots-visuels et l'approche MKL sont fournies. Les performances de notre proposition en terme de temps de calcul sont aussi mises en évidence.

Enfin, la conclusion générale présente une synthèse des contributions apportées ainsi que les pistes définissant des perspectives possibles pour des travaux futurs.

Chapitre 1

Extraction d'attributs et classification d'images dans les bases hétérogènes

Les éléments fondamentaux d'un système de reconnaissance d'images par le contenu sont présentés dans ce chapitre, une attention particulière est réservée au cas des bases hétérogènes. L'ensemble des caractéristiques hétérogènes utilisées dans cette thèse est discuté suite à une présentation d'un état-de-l'art des techniques de classification d'images. Une large évaluation des performances de reconnaissance de cet ensemble de caractéristiques est également fournie.

1.1 Introduction

La recherche d'images a été basée, initialement, sur l'annotation textuelle des images (Jeon et al., 2003), (Ben-Ismail et al., 2008), sachant qu'une forte corrélation doit exister entre le contenu de ces images et les étiquettes textuelles associées. Le coût de génération, ainsi que le *fossé linguistique* associés à ces annotations (Hare et al., 2006), expliquent le grand intérêt accordé, aujourd'hui, aux techniques de recherche d'images basées sur le contenu (Gevers and Smeulders, 2004). Dans cet objectif et depuis une dizaine d'années, de nouvelles méthodes ont permis la mise en place de différents systèmes de recherche et de reconnaissance d'images par le contenu souvent désignés par l'acronyme anglais CBIR (*Content Based Image Retrieval*).

Dans une première partie de ce chapitre, les éléments fondamentaux d'un système de reconnaissance d'images par le contenu sont étudiés. Nous détaillons, en particulier, les bases d'images et leurs contenus, les types de description d'images, les méthodes de classification et les mesures de pertinence dédiées aux systèmes CBIR dans la section 1.2. Les bases d'images hétérogènes utilisées sont présentées dans la section 1.3. Dans une deuxième partie, une synthèse des techniques de classification d'images est exposée dans la section 1.4. Ensuite, l'ensemble des caractéristiques retenues pour la description des images hétérogènes est décrit dans la section 1.5. Une large évaluation des performances de reconnaissance de cet ensemble de caractéristiques est également fournie à la fin du chapitre.

1.2 Reconnaissance d'images par le contenu

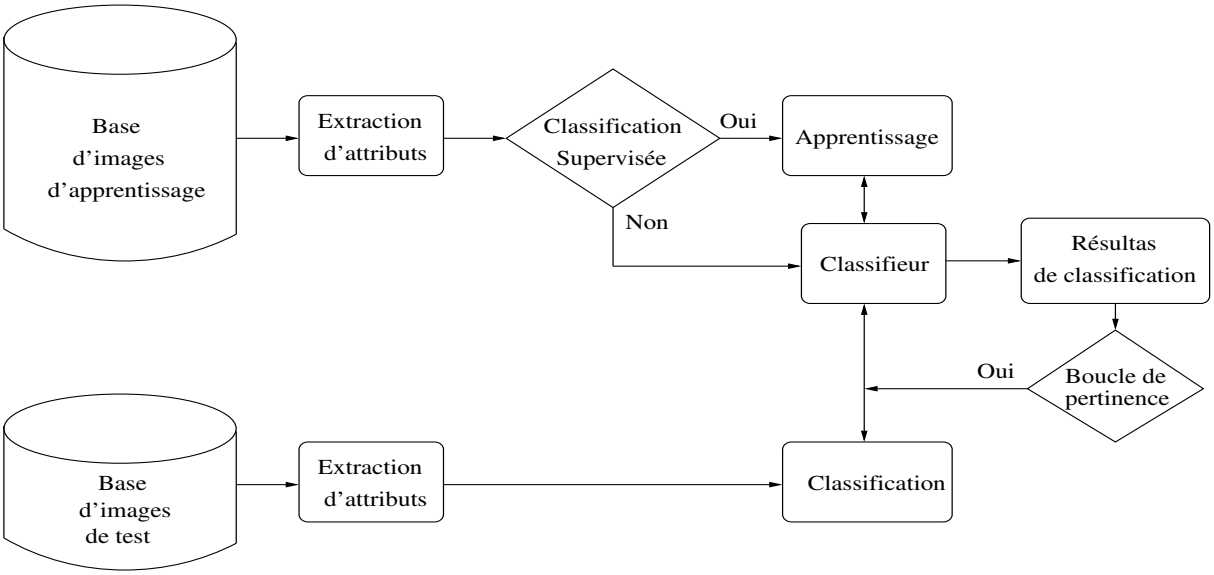


FIG. 1.1 – *Architecture d'un système de reconnaissance d'images par le contenu basé sur la classification.*

L'architecture type des systèmes CBIR est illustrée par la figure 1.1. Cette architecture est composée principalement par deux étapes importantes, à savoir l'extraction d'attributs visuels pour la description des images et la classification qui mène à la reconnaissance de ces images. Dans ce cadre, plusieurs approches ont été proposées pour développer soit des techniques d'extraction d'attributs, soit des méthodes de classification. Par ailleurs, les

performances d'un système CBIR reposent sur un ensemble d'éléments fondamentaux tels que le type de la base d'image considérée, le type de la description d'images employée ainsi que la méthode de classification utilisée. Dans cette section, nous discutons l'ensemble de ces éléments. L'analyse présentée, comporte les définitions principales de chaque partie, ce qui permet de mieux comprendre son influence sur le système de reconnaissance d'images.

Un système de reconnaissance d'images est souvent conçu pour un type bien déterminé de base d'images. Les bases d'images se classent, généralement, en deux grandes catégories:

- **Les bases spécialisées** sont des bases dans lesquelles nous trouvons des images d'un domaine particulier (images médicales, images satellitaires, images architecturales, images biométriques telles que les images de visages ou d'empreintes, images artistiques telles que les tableaux de musée, etc...).

- **Les bases hétérogènes désignées aussi par généralistes** sont des bases d'images de sujets très variés comprenant des catégories d'images très différentes (par exemple couchers de soleil, montagne, plage, animaux, véhicules, bâtiments, etc...).

Les images dans le cas des bases spécialisées ont le plus souvent un contenu sémantique homogène et présentent plusieurs caractéristiques en commun. Les applications visées sont généralement professionnelles. La reconnaissance d'images dans de telles bases est en fait un problème spécifique à un domaine bien particulier. On cite par exemple, l'authentification des individus dans les bases d'images biométriques telles que les bases d'iris (Jlassi and Hamrouni, 2010) et de visages (Rowshan et al., 2008). Ou encore la détection de pathologies dans le cas des bases d'images médicales, tel que le dépistage du cancer du sein dans les bases mammographiques (Cheikhrouhou et al., 2009) et la segmentation des tumeurs de cerveau dans les bases d'images IRM (Chen and Ruan, 2010). Une intégration de la vérité terrain définie par les connaissances d'un expert du domaine est alors nécessaire pour développer des algorithmes spécifiques. L'objectif est d'optimiser l'efficacité du système de reconnaissance d'images dans ces bases et de répondre aussi aux attentes de l'expert. Par contre, la deuxième catégorie de bases d'images généralistes regroupe des images appartenant à des concepts hétérogènes. Les images sont alors de divers types et elles ont un contenu assez complexe. Par conséquent, la difficulté de reconnaissance d'images est accrue dans ce type de bases à cause des limitations de description des images. En outre, ces bases requièrent l'identification préalable de concepts spécifiques pour la création d'une vérité-terrain.

1.2.1 Types d'attributs

Les images sont des objets numériques très riches en terme d'informations. En plus de l'espace mémoire gigantesque exigé, la manipulation directe de ces images dans un système de reconnaissance d'images par le contenu ne permet pas d'obtenir des temps de réponse réalistes. Il convient donc d'utiliser une représentation de dimension réduite pour caractériser le contenu de ces images. L'objectif principal de l'extraction d'attributs est de déterminer pour chaque image, une représentation (signature) qui soit, d'une part compacte, pour être rapidement accessible et facilement comparable, et d'autre part suffisamment complète pour bien caractériser l'image. Il est recommandé d'employer des attributs invariants aux transformations colorimétriques telle que le changement d'éclairage et aux transformations géométriques telle que le changement d'échelle. Ceci permet de palier aux différentes transformations que peut subir une image. Cependant,

la caractérisation robuste et discriminante des images reste un grand défi en traitement d'images. Généralement, les attributs dits de bas niveau sont souvent les plus utilisés pour la description d'images par le contenu. Ces attributs décrivent les principales caractéristiques visuelles existant dans une image, à savoir la couleur, la texture et la forme.

1.2.1.1 La couleur

Le physicien James Clerk Maxwell (Pascale, 2003) a prouvé que toute couleur, étant initialement une sensation provoquée par la lumière avec l'œil, le cerveau et notre expérience, est une synthèse de trois couleurs seulement: le rouge, le vert et le bleu. À partir de cette découverte en 1865, la colorimétrie (la science qui étudie la couleur) a vu le jour.

Les attributs couleurs sont les premiers utilisés dans les systèmes de reconnaissance d'images par le contenu et ils sont toujours les plus utilisés grâce à leur simplicité d'extraction, leur richesse de description et leur efficacité de reconnaissance. Les auteurs dans (Bimbo, 2001) ont fourni une étude étendue des différentes méthodes employées pour l'extraction d'attributs couleurs. Un ensemble d'attributs couleurs a été testé pour être inclus dans le standard MPEG-7 (Manjunath et al., 2001). Ces attributs couleurs dépendent directement de l'espace couleur utilisé pour la représentation couleur de l'image. Dans la littérature, plusieurs espaces couleurs sont étudiés.

Défini en 1931 par la compagnie internationale de l'éclairage (C.I.E.), l'espace couleur *RVB* (Rouge Vert et Bleu) représente les couleurs par synthèse soustractive. Chaque couleur est représentée par trois composantes: rouge, vert et bleu. L'espace *RVB* est généralement l'espace couleur de base dans lequel sont capturées les images. La figure 1.2 illustre la décomposition d'une image couleur en trois sous-images monochromes selon l'espace couleur *RVB*.

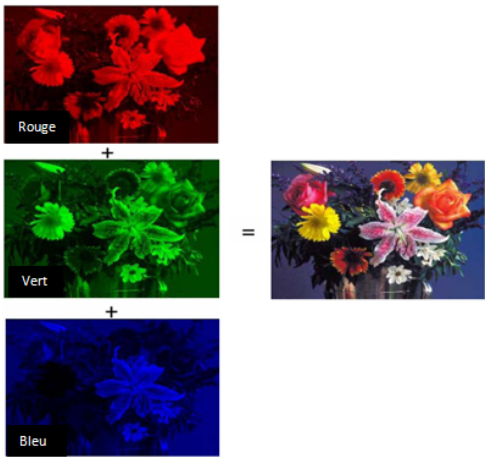


FIG. 1.2 – *Image couleur* RVB*: association de trois sous-images monochromes, correspondant respectivement au différentes composantes Rouge, Vert et Bleu.*

Les autres espaces couleurs représentent des transformations directes de l'espace *RVB*. Ces transformations ont pour objectif d'améliorer la représentativité colorimétrique de l'image et d'assurer des mesure de distance plus linéaires.

Les espaces couleurs *HSI*, *HSV*, *HSB* et *HLS* (appelés communément *HSx*) sont plus proches de la perception humaine de couleur que l'espace couleur *RVB*, mais ils ne sont pas encore perceptuellement uniformes. Nous citons aussi les espaces couleurs *YUV* et *YIQ* où la composante Y représente la luminance du pixel et les pairs de composantes UV et IQ

représentent les composantes chromatiques. Nous pouvons trouver aussi l'espace couleur *XYZ*, où les composantes X et Z sont les composantes chromatiques et la composante Y est la somme pondérée des valeurs R, V et B. Comme il n'est pas perceptuellement uniforme, l'uniformisation de l'espace *XYZ* se fait par une transformation projective qui débouche sur l'espace couleur *LUV*. D'autres espaces couleurs existent aussi tels que *HMMD*, $L^*a^*b^*$, *LST* et *YCrCb*. Pour plus de détails concernant les espaces couleurs, nous recommandions le lecteur intéressé de consulter la référence suivante (Pascale, 2003).

1.2.1.2 La texture

La texture est liée aux apparences de surface des images ce qui est d'une grande importance dans tout domaine relatif à la perception visuelle. Dans la littérature, nous trouvons plusieurs définitions de la texture. Nous citons parmi lesquelles: « la texture est l'attribut qui représente la distribution spatiale des niveaux de gris des pixels d'une région » (IEEE-Press, 1990). Ou encore: « la texture est l'ensemble des propriétés de voisinage locales des niveaux de gris d'une région de l'image » (Livens et al., 1997). Bien qu'il n'y ait pas jusqu'à présent une définition précise de la texture, toutes les définitions possibles affirment que la texture est l'information générée par les variations des niveaux de gris entre des pixels voisins au sein d'une image. Différents modèles de texture sont illustrés dans la figure 1.3.

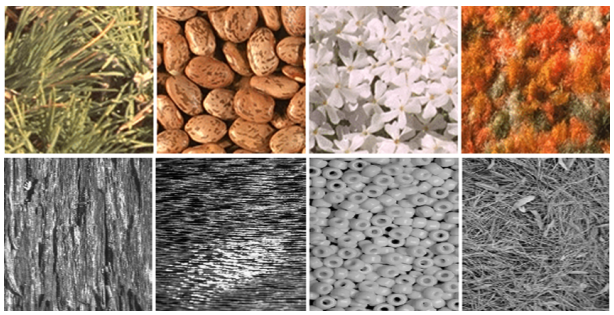


FIG. 1.3 – *Différents modèles de texture.*

Les attributs textures sont de plus en plus utilisés dans la description d'images car ils atténuent certains problèmes liés à la description couleur. En effet, la description texture est très efficace, surtout dans le cas d'une distribution couleur très proche. Les attributs textures sont divisés en deux catégories : La première est déterministe et fait référence à une répétition spatiale d'un motif de base dans différentes directions. Cette approche structurelle correspond à une vision macroscopique des textures. La deuxième approche, dite microscopique, est probabiliste et cherche à caractériser l'aspect chaotique qui ne comprend ni motif localisable, ni fréquence de répétition principale.

1.2.1.3 La forme

Contrairement aux attributs couleurs et textures qui s'intéressent à la description du contenu général de l'image, les attributs formes sont capables de caractériser les différents objets contenus dans l'image.

Généralement, ce type d'attribut indique l'aspect général d'un objet, comme son contour, donc une segmentation sous forme de traitement préliminaire de l'image est souvent nécessaire. Deux catégories d'attributs formes peuvent être extraites: La première

catégorie est basée sur la géométrie des régions de l'image. La deuxième est basée sur les statistiques des intensités de pixels des différentes régions dans l'image.

Une image peut subir des transformations géométriques comme la rotation, la translation et le changement d'échelle. Pour assurer une description assez robuste et efficace, les attributs formes couvrent, généralement, toutes les échelles de représentation (allant des détails jusqu' à la forme grossière) que comporte un objet. De plus, ils sont souvent insensibles aux différentes variations causées par des transformations géométriques. La figure 1.4 présente un exemple des transformations géométriques que peut subir une image.

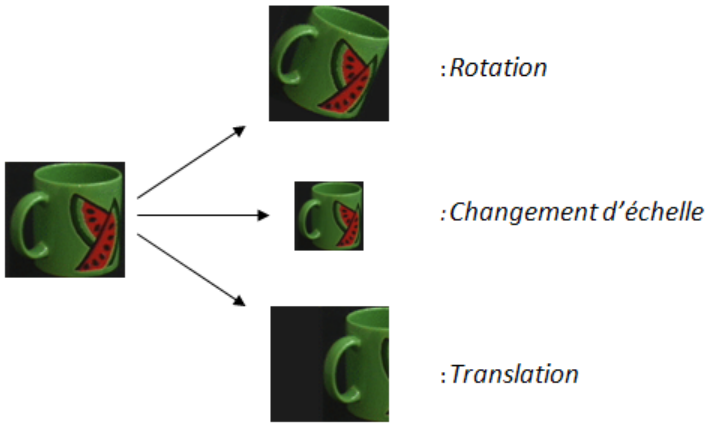


FIG. 1.4 – *Différentes transformations géométriques que peut subir une image.*

1.2.2 Catégories d'attributs

Les attributs visuels sont extraits à partir de l'image selon trois niveaux d'abstractions: le niveau global, le niveau régions et le niveau points d'intérêt. Le premier niveau reflète une description globale des images, tandis que les deux derniers permettent d'assurer une description locale. De ce fait, les différents types d'attributs, décrits dans la section 1.2.1, peuvent être considérés pour l'image entière ou pour une partie spécifique de l'image. Dans la littérature, il est d'usage de distinguer deux catégories d'attributs: les attributs globaux et les attributs locaux (Aly et al., 2009).

1.2.2.1 Attributs globaux

L'approche globale de reconnaissance d'images par le contenu est initialement développée dans les systèmes Photo-Book (Pentland et al., 1994), QBic (Flickner et al., 1995) ou Virage (Bach et al., 1996), et plus récemment par ImageRover (Sclaroff et al., 1997), PicToSeek (Gevers and Smeulders, 1999), Compass (Brunelli and Mich, 2000) et Ikona (Boujemaa et al., 2001). Cette approche globale permet la reconnaissance des images en se basant sur des similarités visuelles mesurées sur les images entières. Une image décrite globalement est représentée par un seul vecteur d'attributs, ce qui n'est pas le cas avec les attributs locaux. Par conséquent, la phase de classification et de reconnaissance devient plus rapide. Généralement, ce type d'attributs est robuste et peu affecté par le bruit qui peut s'ajouter au contenu de l'image. Les histogrammes de couleur ou de niveaux de gris et les corrélogrammes sont des exemples typiques de ce type de description (Stricker and

Swain, 1994), (Huang et al., 1997). L'ensemble des caractéristiques globales, employées dans ce travail de thèse, est présenté dans la section 1.5.

1.2.2.2 Attributs locaux

L'approche locale de reconnaissance d'images se base essentiellement sur la description de parties spécifiques dans l'image, représentées souvent par des sous-régions homogènes. Plusieurs travaux ont été proposés pour la localisation de ces sous-régions. Les subdivisions systématiques d'images en carrés sont simples et rapides mais imprécises (Moghaddam et al., 1999), (Malki et al., 1999). Cependant, une subdivision manuelle assure des segmentations en régions plus proches des attentes de l'utilisateur mais elle n'est pas adaptée pour des bases d'images de grande taille (Bimbo and Vicario, 1998). Généralement, les méthodes de segmentation non-supervisées sont les plus utilisées. Ces méthodes permettent de détecter automatiquement les sous-régions homogènes dans l'image (Carson et al., 1999), (Ma and Manjunath, 1999) en se basant sur les répartitions de couleur et de texture. Des échantillons d'images segmentées par région de la base COREL sont illustrées dans la figure 1.5. Les sous régions détectées sont représentées par leurs couleurs dominantes (Chen and Wang, 2004).



FIG. 1.5 – *Échantillons d'images segmentées de la base COREL.*

La description d'image est construite dans ce cas par un ensemble de vecteurs d'attributs. Les attributs utilisés sont, généralement, extraits pour chaque sous régions dans l'image. Une technique récente impliquant la mise en correspondance de régions pour la recherche d'images est proposée par (Wang et al., 2001) dans le système SIMPLIcity et plus tard par (Sridhar et al., 2002) et (Stehling et al., 2002). En réalité, la similarité entre deux images est mesurée comme une combinaison des similarités entre les régions constituantes de chaque image. Le système effectue, dans ce cas, une recherche globale car toutes les primitives de l'image sont prises en compte.

Il est possible également de caractériser le contenu des images par le biais de la description des points d'intérêt. Plusieurs méthodes de détection de points d'intérêt dans les images sont proposées tout en soulignant différents aspects d'invariance avec différentes propriétés et différentes tailles d'échantillons. On cite, par exemple, la Différence de Gaussienne (DoG pour *Difference of Gaussian*) (Lowe, 2004), Harris Laplace (Mikolajczyk and Schmid, 2004) et Hessien Affine (Mikolajczyk et al., 2005). Les approches locales basées sur des descriptions de points d'intérêt permettent d'effectuer des recherches précises mais au prix d'un temps de calcul très élevé (Schmid et al., 1998), (Gouet and Boujemaa, 2001). Effectivement, le nombre élevé des points d'intérêt détectés par image (quelques centaines) et l'aspect dimensionnelle des descripteurs (128 dimensions pour un vecteur unique SIFT *Scale-Invariant Feature Transform* (Lowe, 2004)) rendent la mise en correspondance entre les images très coûteuse en temps de calcul. Des travaux récents ont contribué à la compensation des limites de ces descripteurs en utilisant une nouvelle approche appelée *sac des descripteurs* ou *sac de mots-visuels* (Csurka et al., 2004b), (Willamowski et al., 2004), (Zhang et al., 2007). L'idée est de construire pour chaque image une caractéristique globale qui détermine les différentes occurrences d'un ensemble de catégories de descripteurs

locaux, appelés aussi mots-visuels. Les performances d'une telle approche exigent une majorité de points d'intérêt en commun des images appartenant à un même concept.

Par conséquent, les travaux employant une description locale se situent, généralement, dans des contextes précis et les approches utilisées s'appuient sur certaines propriétés connues de ces contextes pour fonctionner. Les systèmes basés sur une description locales manquent donc de généralité par rapport aux images qu'ils peuvent traiter. Effectivement, les attributs locaux sont surtout utilisés dans le cas des bases d'images dites spécifiques où les images appartenant à une catégorie bien déterminée représentent une répartition assez homogène. Les méthodes développées sont souvent dédiées aux corpus utilisés et parfois peu adaptables à d'autres images. Ainsi, une description locale est beaucoup moins performante dans le cas des images hétérogènes, où les images d'une même classe ne présentent pas forcément une majorité de sous régions ou de points d'intérêts en commun.

1.2.3 Méthodes de classification et évaluation des performances

1.2.3.1 Classification

On distingue principalement deux types de classification, supervisée et non-supervisée, suivant que l'on dispose ou non d'une expertise des images à classifier (figure 1.1). La classification supervisée se base sur une phase d'apprentissage où l'on cherche à produire automatiquement des règles de décision à partir d'images déjà classifiées. Le choix de la technique d'apprentissage à utiliser est très important dans un tel processus. La classification de nouvelles images de test est réalisée dans une seconde phase, appelée généralisation. À partir des images d'apprentissage, le classifieur utilisé doit bien généraliser les règles de décision apprises sur de nouvelles images requêtes appartenant à une base dite de test. L'étape d'apprentissage est réalisée hors ligne, alors que celle de généralisation est effectuée en ligne. La classification non-supervisée se ramène quand à elle au regroupement des images de la façon la plus naturelle possible sans aucune connaissance a priori du contenu de la base. Cette volonté de regrouper naturellement les images est assez ambiguë. Elle est formalisée, le plus souvent, par l'objectif de réduire la variation du contenu des images intra-classe et de mieux caractériser la discrimination des images inter-classe. Un état-de-l'art des différentes techniques de classification, qui existent dans la littérature, est présenté dans la section 1.4.

1.2.3.2 Rétroaction et bouclage de pertinence

Généralement, les résultats de classification sont obtenus d'une façon automatique selon la description préalable du principe de base des différents types de classification. Toutefois, il est possible de se servir d'une boucle de pertinence où l'utilisateur raffine progressivement l'efficacité de reconnaissance. Dans ce cas, l'utilisateur marque les images résultantes de la classe candidate comme pertinentes, non pertinentes ou neutres vis à vis l'image requête. Le processus de reconnaissance est ainsi répété à chaque fois avec ces nouvelles informations jusqu'à l'obtention du résultat désiré. Cette interaction n'est pas toujours efficace, sachant qu'elle suppose un ensemble de connaissance chez l'utilisateur. On distingue généralement deux types d'utilisateurs: les experts et les non-spécialistes. Uniquement, les experts peuvent interagir efficacement avec le système pour améliorer les résultats de reconnaissance. Cependant, pour les non-spécialistes, aucune interaction ne peut être considérée.

1.2.3.3 Performances de classification d'images

L'évaluation des performances d'un système de reconnaissance d'images concerne à la fois sa capacité de reconnaissance et son efficacité en terme de rapidité d'exécution. Cette dernière est facile à évaluer, il suffit de mesurer le temps de réponse moyen du système sur plusieurs requêtes. L'évaluation des capacités de reconnaissance est effectuée en comparant le résultat de classification obtenu à un résultat de référence, appelé souvent vérité terrain. En effet, la connaissance de la vérité-terrain permet de réaliser une évaluation automatique et de fournir des performances fiables. Les critères de qualité les plus utilisés sont des mesures de performance standard (Bimbo, 2001), il s'agit principalement de la précision et du rappel. Soit pour une image requête donnée, TP le taux des vrai positifs, FP le taux des faux positifs et P le nombre total des images pertinentes dans la base. Notons que les taux TP et FP indiquent, respectivement, le nombre d'images pertinentes et non-pertinentes retrouvées. La précision et le rappel sont définis comme suit:

$$\text{Précision} = \frac{TP}{TP + FP} \ , \qquad \text{Rappel} = \frac{TP}{P} \tag{1.1}$$

En variant le nombre d'images requêtes, un graphe de précision-rappel peut être obtenu. La précision est souvent une fonction décroissante du rappel. Le graphe typique de précision-rappel est illustré par la figure 1.6.

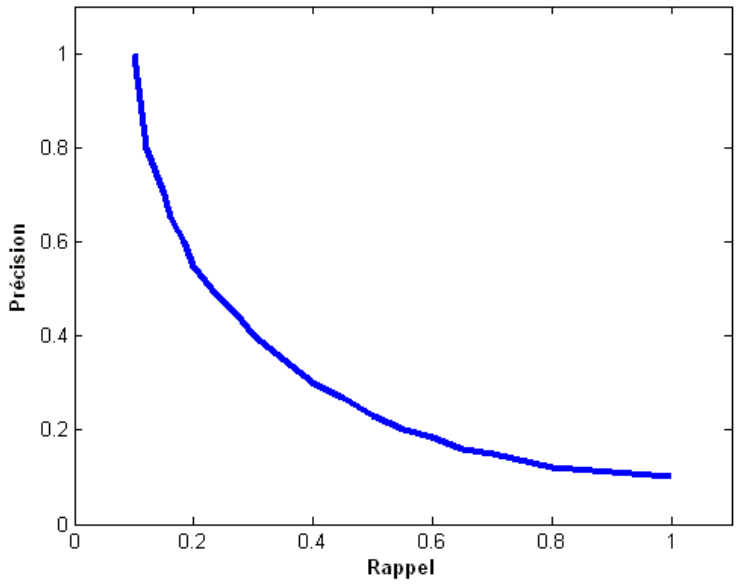


FIG. 1.6 – *Graphe typique de précision-rappel.*

Vu que les valeurs scalaires de la précision et du rappel sont comprises toutes les deux entre 0 et 1, le graphe de précision-rappel est dit idéal si la précision est égale à un pour toute valeur de rappel.

1.3 Bases d'images hétérogènes utilisées

Pour évaluer et valider les différentes approches étudiées et développées dans ce travail, trois bases d'images hétérogènes sont utilisées, à savoir, la base COLUMBIA, la base

COREL et la base CALTECH-256. Ces bases constituent des concepts bien définies où chaque image ne peut appartenir qu'à une seule classe. Dans ce qui suit, une description des trois bases d'images utilisées est fournie.

1.3.1 La base d'images COLUMBIA

La base d'images COLUMBIA contient des images tirées d'objets 3D pris en différentes positions. Dans cette thèse, la collection utilisée de cette base est COIL-100[1]. La collection COIL-100 est utilisée surtout dans le domaine de reconnaissance d'objets. Elle est constituée par des images en couleurs prises à partir de 100 objets différents. L'ensemble de ces objets est affiché dans la figure 1.7.



FIG. 1.7 – *Échantillons des* 100 *objets de la base d'images COIL*-100 *de COLUMBIA.*

À l'aide d'un plateau motorisé (tourne-disques) et un fond noir, chaque objet est capturé selon différentes positions pour constituer une classe de la base. Le plateau est tourné de 360° pour changer la position de l'objet par rapport à une caméra couleur fixe. Une image de l'objet est prise à chaque intervalle de 5°. Ceci correspond à 72 positions différentes pour chaque objet. COIL-100 contient donc 7200 images de taille 128×128 pixels.

1.3.2 La base d'images COREL

La base COREL[2] est tirée de différents CD-ROMs publiées, strictement pour des raisons de recherche, par COREL corporation. Cette base comprend 2300 images catégorisées par le concepteur en une vérité-terrain de 23 classes thématiques (personnes, animaux, fleurs, montagnes, etc.), contenant 100 images chacune. Chaque image est de taille 384×256 pixels ou 256×384 pixels. Cette base est considérée comme une référence pour les spécialistes travaillant sur la reconnaissance d'images dans les bases hétérogènes. Des échantillons de différentes classes d'images contenues dans la base COREL utilisée sont représentés dans la figure 1.8.

1. La base COIL-100 (*Columbia Object Image Library*) est accessible en ligne à l'adresse suivante: http://www.cs.columbia.edu/CAVE/databases.

2. La base COREL peut être téléchargée du site web de James Z. Wang's Research Group, disponible en ligne à l'adresse suivante: http://wang.ist.psu.edu/docs/home.shtml.

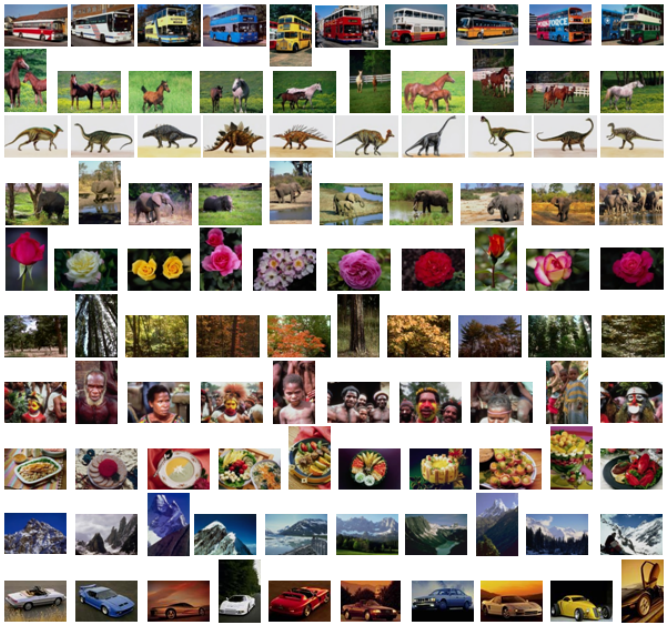


FIG. 1.8 – *Échantillons d'images de différentes classes de la base COREL utilisée.*

1.3.3 La base d'images CALTECH

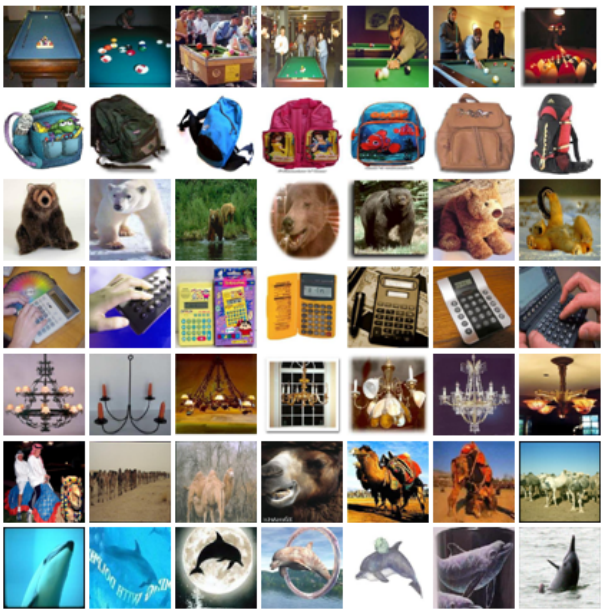


FIG. 1.9 – *Échantillons d'images de différentes classes de la base CALTECH-256 utilisée.*

La base CALTECH-256[3] (rassemblée par (Griffin et al., 2007)) est composée d'images tirées de 256 catégories différentes, c'est une extension de la base CALTECH-101[4]. Elle contient de 80 à 827 images par catégorie. Le nombre total des images contenues dans

3. La base CALTECH-256 contient moins d'alignement entre les images d'une même classes que dans le cas de la base CALTECH-101. Elle est disponible en ligne à l'adresse suivante: http://www.vision.caltech.edu/Image_Datasets.

4. La base CALTECH-101 (rassemblée par (Fei-Fei et al., 2006)) est composée d'images tirées de 101 catégories d'object, elle contient de 31 à 800 images par catégorie. Cette base d'images est connue par sa grande variabilité inter-classe.

cette base est 30608 images de taille 300 × 300 pixels. Cette base est connue par une grande variabilité inter et intra-classes. Dans ce travail, nous utilisons un ensemble de 23 classes de la base CALTECH-256, contenant 100 images chacune. La figure 1.9 illustre des échantillons de différentes classes d'images contenues dans le sous-ensemble utilisé.

Les noms ainsi que les identifiants (IDs) correspondants aux 23 classes utilisées à partir des deux bases d'images COREL et CALTECH-256 sont présentés dans le tableau 1.1 (on maintient l'appellation anglaise).

TAB. 1.1 – *Noms et identifiants des* 23 *classes utilisées à partir des deux bases d'images COREL et CALTECH-256.*

IDs	Noms de classes	
	COREL	CALTECH-256
0	Buses	Backpack
1	Horses	Baseball-bat
2	Dinosaurs	Bathtub
3	Elephants	Bear
4	Flowers	Billiards
5	Forests	Binoculars
6	African peoples and villages	Bowling-ball
7	Food	Bread-maker
8	Mountains and glaciers	Calculator
9	Cars	Camel
10	Dogs	Chandelier
11	Lizards	Chess-board
12	Fashion models	Chimp
13	Sunset scenes	Coin
14	Beach	Computer monitor
15	Waterfalls	Cormorant
16	Antique furnitures	Diamond-ring
17	Battle Ships	Dolphin
18	Skiing	Elk
19	Desserts	Fern
20	Butterfly	Fire-truck
21	Historical buildings	Harp
22	Night scenes	Ketch

Nos expérimentations sont réalisées en utilisant, pour chaque base d'images, neuf sous-bases différentes. Ces sous-bases contiennent un nombre variable de classes thématiques allant de 3 jusqu'à 23 classes. Dans les bases hétérogènes, certaines images de classes différentes peuvent se révéler d'un aspect très proches, alors que des images de la même classe peuvent avoir des structures dissemblables. De ce fait, la classification supervisée est utilisée. Grâce à l'étape d'apprentissage, ce type de classification assure une meilleure séparation des différentes classes dans les bases d'images hétérogènes. Une synthèse des techniques de classification, existant dans la littérature, est présentée dans la section 1.4. Par ailleurs, vu la nature hétérogène et le manque de connaissance à priori sur le contenu des bases utilisées, les images appartenant sémantiquement au même concept ne présentent pas forcément une majorité de sous-régions ou de points d'intérêt communs. Par conséquent, l'utilisation d'une description globale des images est adoptée. D'autre part, les résultats de reconnaissance obtenus restent fortement dépendants de

la description d'images utilisée. Pour couvrir les différentes caractéristiques visuelles des images hétérogènes, une large description est généralement recommandée (Manjunath et al., 2001), (Androutsos et al., 2005). Dans notre cas, un ensemble de caractéristiques de différents types est extrait. Plus de détails sont fournis dans la section 1.5. Ainsi, le système CBIR étudié dans notre contexte de travail, se base essentiellement sur une construction hors ligne d'une large description globale des images. L'étape de reconnaissance est basée sur une technique de classification supervisée. Cette étape est totalement automatique et adaptée au contenu de la base considérée. Aucune rétroaction de l'utilisateur, pour l'évaluation des résultats à travers un bouclage de pertinence, n'est envisagée. Dans la section suivante, un état-de-l'art des techniques de classification d'images est présenté.

1.4 Techniques de classification d'images

Les performances d'un système de reconnaissance d'images par le contenu reposent essentiellement sur les caractéristiques ainsi que la catégorie d'attributs utilisées pour la description des images. Cependant, le bon usage de cette description d'images, à travers un algorithme de classification, qui s'adapte au processus de reconnaissance considéré, est d'une extrême importance. Ainsi, la méthode de classification utilisée influe, conjointement avec la nature de description d'images employée, sur les résultats de reconnaissance obtenus. Le critère principal de toute méthode de classification se base sur le fait que les images appartenant au même concept doivent avoir au moins une ou plusieurs propriétés communes (Jurgen, 2004). Le processus de classification général comporte deux étapes: l'étape d'apprentissage et celle de généralisation. L'étape d'apprentissage sert à construire un modèle à partir des images de la base considérée. L'étape de généralisation représente la classification proprement dite. Particulièrement, le modèle construit en phase d'apprentissage est utilisé en généralisation pour estimer les classes des différentes images requêtes (Han and Kamber, 2006).

Une grande variété de systèmes/approches et techniques de classification se trouve dans la littérature. Cependant, aucune approche ne prétend être la plus performante pour tous les problèmes de classification (Stork et al., 2001). Dans ce cadre, un bref état-de-l'art des différentes techniques de classification est présenté dans cette section. Nous rappelons que ces techniques sont principalement catégorisées en deux familles: la classification non-supervisée et la classification supervisée (section 1.2.3.1).

1.4.1 Méthodes de classification non-supervisée

Dans l'approche de classification non-supervisée, les classes qui constituent la base d'images ne sont pas identifiées. Généralement, le nombre de classes est également inconnu. Ainsi, la classification non-supervisée, connue en anglais par *clustering* consiste à déterminer les différentes classes naturellement sans aucune connaissance préalable. L'objectif, dans ce cas, est d'identifier une structure dans les images de la base en se basant sur leurs contenus. Tout en caractérisant la discrimination des images inter-classe et réduisant la variation du contenu des images intra-classe, les images sont attribuées aux différentes classes estimées selon la similarité de leurs caractéristiques. Par conséquent, le choix d'une mesure de distance appropriée est une étape importante dans tout algorithme de *clustering*. Cette distance spécifie la façon avec laquelle la similitude de deux images

est calculée. Cela est d'une grande influence sur la forme finale des classes. Effectivement, certaines images peuvent être proches l'une de l'autre en fonction d'une certaine distance et plus éloignées selon une autre (Dengsheng and Guojun, 2003). La tâche de regroupement d'images peut être éventuellement simplifiée si le nombre de classes dans la base est connu, mais la classification reste toutefois non-supervisée puisque le classifieur n'est toujours pas guidé par des images déjà étiquetées. De ce fait, l'analyse de la répartition des images, dites aussi observations, dans l'espace des attributs reste le seul moyen qui permet de caractériser la signification des véritables classes. Dans la suite, nous effectuons un aperçu plus détaillé des méthodes de classification non-supervisée qui existent dans la littérature.

Il existe plusieurs approches pour la catégorisation des méthodes de classification non-supervisée. Souvent, ces méthodes sont regroupées en deux grandes catégories: paramétriques et non-paramétriques (Fukunaga, 1972). Les méthodes paramétriques sont basées sur une pré-analyse d'une structure de données globale. Les performances de ce type de méthodes dépendent essentiellement des hypothèses effectuées suite à l'approximation de leurs paramètres (le nombre de classes par exemple). Le critère de classification le plus intuitif et fréquemment utilisé est le critère de l'erreur quadratique. Il est parfaitement adapté au cas des classes compactes et isolées (Jain et al., 1999). Parmi les méthodes paramétriques qui se basent sur ce critère, la méthode la plus communément utilisée est celle de l'algorithme *K-means* (McQueen, 1967). Plusieurs variantes de cet algorithme ont été proposées pour gérer la sensibilité à la répartition initiale (Anderberg, 1973). Dans le cas des méthodes non-paramétriques, aucune hypothèse sur le(s) paramètre(s) caractérisant(s) n'est effectuée. Le nombre de classes est particulièrement non défini. Les images sont regroupées selon une fonction de densité. Par conséquent, la performance des méthodes non-paramétriques est, naturellement, très sensible à la distribution d'images dans la base. L'exemple classique d'approches non-paramétriques est l'algorithme appelé CAST pour *Cluster Affinity Search Technique*. C'est une approche itérative (Portnoy et al., 2002) qui examine efficacement les observations aberrantes (Shamir and Sharan, 2002).

Une deuxième structuration des méthodes de classification non-supervisée est couramment utilisée dans la littérature. Cette structuration inclut trois catégories:

- **Les méthodes de classification hiérarchiques:** constituent la famille de techniques qui génèrent des suites de partitions emboîtées les unes dans les autres, et allant depuis la partition triviale où une seule classe contient toutes les observations jusqu'à la partition triviale où chaque observation constitue une classe à part. La méthode de classification hiérarchique doit choisir une partition assez adaptée entre ces deux extrêmes. La taxonomie communément utilisée des méthodes hiérarchiques, comprend cinq sous-catégories (Jain et al., 1999): Les approches « Ascendantes / Descendantes », « Monothétiques / Polythétiques », « Déterministes / Floues », « Déterministes / Stochastiques » et « Incrémentales / Non-incrémentales ».
- **Les méthodes de classification par partitionnement:** permettent principalement d'organiser les images en K classes. Comme K est un paramètre d'entrée, une certaine connaissance du domaine est nécessaire, ce qui n'est pas toujours disponible en pratique (Ester et al., 1996). Un algorithme de partitionnement commence généralement par une partition initiale, puis utilise une stratégie de contrôle itérative pour optimiser une certaine fonction objective. Chaque classe est représentée par son centre de gravité (*K-centroid*) ou par une des images de la classe située à proximité de son centre (*K-medoid*). Un ensemble de méthodes appartenant à cette catégorie,

telles que le modèle de mélange de gaussienne, l'algorithme EM pour *Expectation-Maximization* et le modèle non-supervisé de Bayes existent dans la littérature. Par ailleurs, l'algorithme de partitionnement le plus populaire est l'algorithme *K-means* (Berry, 2003).

– **Les méthodes de classification par deux voies de *clustering*, connues par *co-clustering* ou *bi-clustering*:** sont des algorithmes où non seulement les images sont groupées, mais aussi leurs caractéristiques, c'est à dire, si les images sont représentées dans une matrice de données, les lignes et les colonnes sont regroupées en même temps (Kim and Cipolla, 2008).

L'aperçu présenté des différentes approches de *clustering* démontre une large gamme de méthodes de classification non-supervisée. Le choix d'une méthode particulière dépend fortement de l'application et du type de sortie étant souhaité. Dans la section suivante, les méthodes de classification supervisée sont exposées.

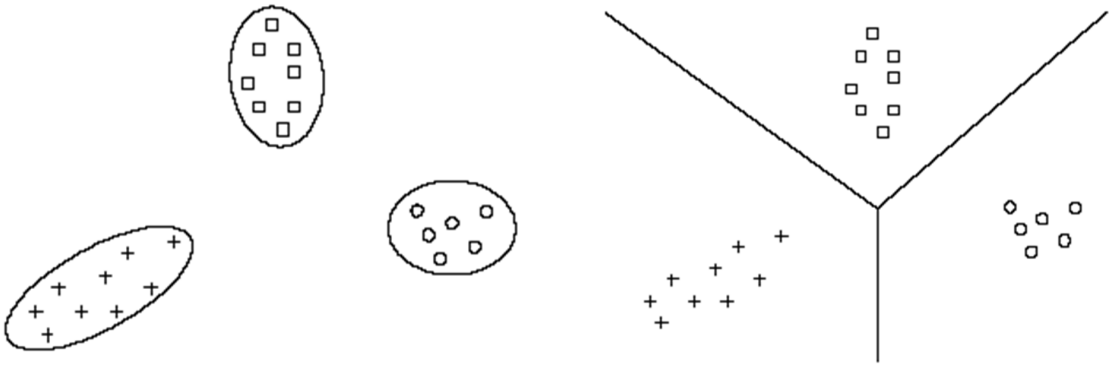
1.4.2 Méthodes de classification supervisée

Dans cette approche de classification, une connaissance de la signification de chaque classe est exigée. Évidemment le nombre de classes est fixé et les images disponibles sont déjà attribuées aux différentes classes de la base. L'objectif est ainsi d'affecter de nouvelles images aux classes appropriées suivant la modélisation effectuée en phase d'apprentissage. En général deux types de modélisation peuvent être distinguées: la modélisation intrinsèque et la modélisation discriminante.

– **Modélisation intrinsèque:** permet d'expliciter les classes par la définition de leurs propriétés intrinsèques. Dans ce cas, l'objectif est de déterminer un modèle qui caractérise au mieux chaque classe dans l'espace d'attributs. La classification de nouvelles images de test repose sur les mesures de similarité entre ces images et les différents modèles de classes.
– **Modélisation discriminante:** assure une description implicite des classes par la définition des frontières de décision permettant de discriminer ces classes. Ainsi, le but est de chercher des hyper-plans, dans l'espace d'attributs correspondant. Ces hyper-plans doivent séparer au mieux les différentes classes de la base. En généralisation, il suffit de trouver de quel côté de ces hyper-plans se trouve chaque image requête.

Une illustration de ces deux types de modélisation est présentée par la figure 1.10. En se basant sur l'une de ces modélisation, plusieurs méthodes de classification supervisées existent dans la littérature. Nous décrivons succinctement dans la suite les méthodes les plus utilisées pour la classification d'images.

Comme méthodes intrinsèques, la classification bayésienne détermine la probabilité d'appartenance de chaque image aux différentes classes de la base (Langley and Sage, 1999). Cette méthode est dite *Naïve* ou *Simple* car elles supposent que les images d'apprentissage sont toujours indépendantes. Si les images ne suivent pas une loi de distribution connue, les conditions de probabilité employées seront erronées et la règle de Bayes ne peut être appliquée. Les arbres de décision, quand à elles, représentent des processus récursifs (Quinlan and Rivest, 1989), (Quinlan, 1993). Les images requêtes sont examinées par l'arbre de décision de chaque classe dans la base. Concrètement, chaque noeud dans l'arbre de décision contient un test (de type « Si ... Alors ») correspondant à la valeur



a) b)

FIG. 1.10 – *Approche de classification supervisée: type de modélisation a) Intrinsèque et b) Discriminante.*

d'un attribut donné. Les feuilles de l'arbre représentent les décisions tirées à partir des résultats de ces tests. Les arbres de décision fonctionnent mieux si le nombre de valeurs possible pour chaque attribut est faible. Par contre, pour un grand jeu de données, l'arbre devient illisible. Dans le même contexte, les réseaux de neurones artificiels, connus par l'acronyme anglais ANN (*Artificial Neural Networks*), reposent sur la théorie des perceptrons (Maren, 1990). Un ANN est composé de plusieurs neurones répartis sur une couche d'entrée, une couche de sortie et un nombre de couches cachées. Par ailleurs, il n'existe pas de règle pour choisir le nombre de couches du réseau. Les MLP pour *Multi Layer Perceptron* sont constitués par une ou plusieurs couches cachées (Freeman and Skapura, 1991). Par contre, les réseaux à fonctions radiales de base, connus par l'acronyme anglais RBF (*Radial Basis Functions*), sont constitués par une seule couche cachée (Oyang et al., 2005). Généralement, les ANNs sont robustes aux erreurs et sont donc bien adaptés aux problèmes d'apprentissage d'images bruitées. Cependant, ils ont une pauvre possibilité d'interprétation et un choix de paramètres très délicats.

Par ailleurs, les méthodes discriminantes représentent souvent des techniques de classification linéaires. Deux types de classification linéaire sont utilisés: les fonctions discriminantes et les machines à vecteurs supports. Les fonctions discriminantes sont des méthodes de classification simples qui séparent les images appartenant aux différentes classes en se basant sur une analyse linéaire. Ces fonctions peuvent être performantes dans le cas où les différentes classes de la base sont linéairement séparables. Cependant, dans la plus part des cas, elles ne permettent pas d'atteindre des précisions considérables. Afin d'améliorer les performances de ces fonctions, de nouvelles variantes sont présentées dans la littérature (Kimura et al., 1997). Face aux fonctions linéaires, les machines à vecteurs supports, connues par l'acronyme anglais SVMs (*Support Vector Machines*), sont initialement conçues pour les problèmes de classification binaires. Elles permettent de séparer linéairement les exemples positifs des exemples négatifs dans l'ensemble des images d'apprentissage par un hyper-plan qui garantisse un maximum de marge (Vapnik, 1999). L'efficacité des SVMs est souvent supérieure à celle de toutes les autres méthodes de classification supervisées. Pour les problèmes de non-séparabilité, les SVMs permettent d'effectuer une transformation non-linéaire des observations d'entrée dans un espace de dimension supérieure pour se ramener au cas linéaire. En outre, les SVMs peuvent s'adres-

ser également aux problèmes de classification multi-classes (Weston and Watkins, 1999), (Xu and Chan, 2003), (Platt et al., 2000).

En revanche, il existe des méthodes de clasification supervisées qui ne nécessitent pas de modélisation. En effet, au contraire des méthodes de classification supervisée traditionnelles, aucun modèle n'est construit à partir des images d'apprentissage. C'est le cas du raisonnement à partir d'exemples telle que la méthode des K-plus proches voisins (Kppv), connue en anglais par *K-nearest neighbor* (Weiss, 1991), ou encore *Memory Based Reasoning* (Stanfill, 1987). La règle de décision de la méthode Kppv consiste principalement à attribuer une image requête à la classe représentée en majorité dans son voisinage. Ce voisinage est défini par les K plus proches observations dans la base d'apprentissage par rapport à l'image requête considérée. Deux paramètres sont généralement requis, à savoir le nombre de voisins K à tenir en compte et la fonction de similarité à utiliser. L'efficacité des Kppv est généralement sensible à la densité des différentes régions de l'espace de représentation des images d'apprentissage. De plus, la méthode Kppv nécessite souvent un espace mémoire très important du fait que la totalité de la base d'apprentissage doit être stockée.

À la fin de cette synthèse, nous distinguons principalement deux méthodes de classification: supervisée et non-supervisée. Les méthodes non-supervisées déterminent automatiquement les différentes classes de la base d'images, mais présentent souvent des capacités de précision assez limitées. Les méthodes supervisées, quand à elles, nécessitent un ensemble de connaissances reflétant la vérité terrain de la base d'images d'apprentissage. D'où elles aboutissent, généralement, à de meilleures performances de classification. Dans ce travail de thèse, le contexte supervisé est adopté, en particulier, les classifieurs SVMs sont utilisés. Un choix justifié par la qualité de ces classifieurs en terme de performance de classification. Plus de détails concernant les SVMs sont fournis dans le chapitre 3. Dans la section suivante, l'ensemble des caractéristiques, utilisées pour la description des images, est détaillé.

1.5 Caractéristiques utilisées pour la description d'images dans les bases hétérogènes

1.5.1 Introduction

Comme il a été déjà signalé précédemment et face au contenu assez variable des images hétérogènes, la description adoptée dans ce travail est basée sur l'extraction d'attributs globaux. Une image décrite globalement, contrairement à l'utilisation des caractéristiques locales, est représentée généralement par un seul vecteur d'attributs pour chaque caractéristique extraite. Ceci permet de rendre plus rapide la reconnaissance des images. Les caractéristiques globales couramment utilisées reflètent principalement des propriétés visuelles de bas niveau dans l'image telles que la couleur, la texture, la forme, etc... Une littérature abondante constitue l'état de l'art de ce type de description. Plusieurs contributions récentes sont aussi proposées (Datta et al., 2008), (He and Cercone, 2009). Afin d'assurer une riche description nécessaire pour la nature hétérogène des images traitées, un ensemble de douze caractéristiques globales de différents types (dites aussi hétérogènes) est utilisé. Nous choisissons essentiellement des caractéristiques classiques et éprouvées pour une meilleure évaluation des différentes méthodes proposées dans cette thèse. Tou-

tefois, il est toujours possible d'employer d'autres caractéristiques pour la description des images.

Les caractéristiques couleurs moyennes (Faloutsos et al., 1994) calculées dans quatre espaces couleurs différents, à savoir les espaces *RVB*, *HSV*, *HMMD* et *YCrCb*, la caractéristique histogramme couleur (Hafner et al., 1995) ainsi que la caractéristique corrélogramme (Huang et al., 1997) sont utilisées pour la description couleur. Comme caractéristiques de texture, les statistiques de premier ordre (Press et al., 1987) calculées à partir de l'histogramme niveaux de gris des images, ainsi que les attributs extraits de la matrice de co-occurrence (Haralick et al., 1973), appelée souvent par l'acronyme anglais SGLD (*Spatial Gray Level Dependency*), sont aussi utilisés. Nous calculons en plus la matrice contenant les normes du gradient (Delingette and Montagnat, 2001) de tous les pixels dans l'image et les transformées en ondelettes de Daubeshie. Le descripteur GIST (Oliva and Torralba, 2001) est aussi maintenu comme une caractéristique représentant la couleur et la texture. Enfin, les moments de Hu (Hu, 1962) sont employés comme caractéristiques de forme. La figure 1.11 présente quelques échantillons des caractéristiques extraites à partir des images Dinosaure et Rose appartenant à deux classes différentes de la base COREL, décrite et présentée dans la section 1.3.2.

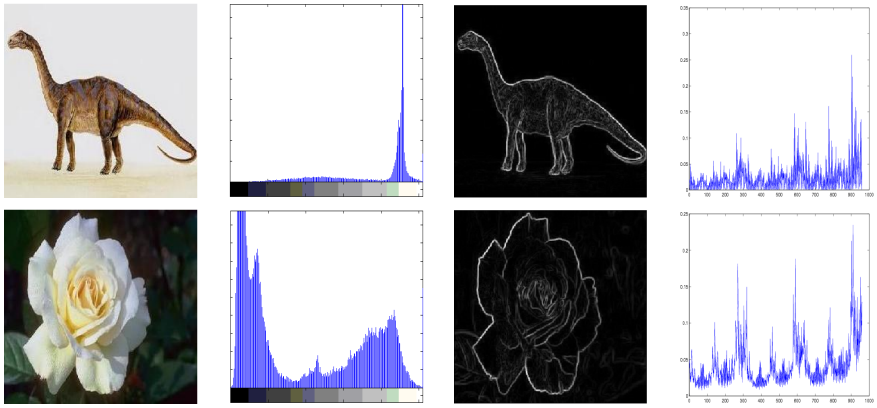


FIG. 1.11 – *Échantillons des caractéristiques extraites respectivement des images Dinosaure et Rose: a) Histogramme, b) Matrice des normes du gradient, et c) GIST.*

Soit $\mathcal{I}$ une image couleur quantifiée sur cl couleurs $\{c_1,c_2,...,c_{cl}\}$. Chaque pixel p_i, $i = [1,2,...,N]$, de coordonnés (x,y) est doté d'une couleur $\mathcal{I}(p_i)$: $p_i \in \mathcal{I}_{c_j} \Leftrightarrow \mathcal{I}(p_i) = c_j$, $\forall j \in \{1,2,...,cl\}$. Dans l'espace couleur RVB, les trois matrices couleur (Rouge, Vert et Bleu) de $\mathcal{I}$ sont notées respectivement par R, V et B. Les différentes caractéristiques utilisées dans ce travail sont présentées ci après.

1.5.2 Les couleurs moyennes

La caractéristique couleur moyenne (Faloutsos et al., 1994), sert à déterminer les différentes distributions moyennes de couleur dans une image. Elle peut être calculée pour n'importe quel espace couleur. Le nombre d'attributs constituant cette caractéristique dépend alors du nombres d'axes dans l'espace couleur utilisé pour la représentation de l'image. Par exemple, la caractéristique couleur moyenne de l'image $\mathcal{I}$ dans l'espace couleur RVB est formulée comme suit:

$$\bar{x} = (\bar{R}_{moy},\ \bar{V}_{moy},\ \bar{B}_{moy})^t \tag{1.2}$$

Sachant que:

$$\bar{R}_{moy} = \frac{1}{N}\sum_{p=1}^{N} R(p), \tag{1.3}$$

$$\bar{V}_{moy} = \frac{1}{N}\sum_{p=1}^{N} V(p), \tag{1.4}$$

$$\bar{B}_{moy} = \frac{1}{N}\sum_{p=1}^{N} B(p). \tag{1.5}$$

1.5.3 Les histogrammes couleurs

Les histogrammes couleur (Hafner et al., 1995) sont des indicateurs de répartition de niveaux de couleurs dans une image. À partir d'un espace discret de couleur défini par les axes couleurs correspondants (par exemple Rouge, Vert et Bleu), l'histogramme couleur est déterminé par la discrétisation de l'image couleur et le calcul du nombre d'occurrence de chaque couleur dans cette image. Les histogrammes couleur dans l'espace RVB sont calculés comme suit, $\forall i \in \{1,2,...,cl\}$:

$$hist_R(c_i) = \frac{1}{N}\sum_{p=1}^{N} R(p) == c_i, \tag{1.6}$$

$$hist_V(c_i) = \frac{1}{N}\sum_{p=1}^{N} V(p) == c_i, \tag{1.7}$$

$$hist_B(c_i) = \frac{1}{N}\sum_{p=1}^{N} B(p) == c_i. \tag{1.8}$$

1.5.4 La matrice corrélogramme

Les caractéristiques précédentes (la couleur moyenne et l'histogramme couleur) ne tiennent pas en considération la distribution spatiale de couleur dans l'image. Dans ce cadre et pour palier à ces limitations de description couleur, la caractéristique corrélogramme a été proposé (Huang et al., 1997). Un corrélogramme présente les corrélations spatiales entre couleurs en fonction d'une distance inter-pixel, prise au sens de la norme L_∞:

$$|p_1 - p_2| = max\{|x_1 - x_2|,|y_1 - y_2|\}. \tag{1.9}$$

Pour une distance $Dist \in \{1,2,...,N\}$ fixée a priori, le correlogramme de la couleur c_i par rapport à la couleur c_j est défini par:

$$Cor_{c_i,c_j}^{Dist} = \underset{p_1 \in \mathcal{I}_{c_i}}{prob}[p_2 \in \mathcal{I}_{c_j} \quad / |p_1 - p_2| = Dist]. \tag{1.10}$$

Pour tout pixel de couleur c_i dans l'image $\mathcal{I}$, Cor_{c_i,c_j}^{Dist} retourne la probabilité[5] de trouver un pixel de couleur c_j à une distance $Dist$.

1.5.5 Les statistiques de premier ordre

Soit NG l'image niveau de gris correspondante à l'image $\mathcal{I}$. L'histogramme des ng niveaux de gris de l'image NG est calculé comme suit, $\forall i \in \{1,2,...,ng\}$:

$$hist(i) = \frac{1}{N}\sum_{p=1}^{N} NG(p) == i \tag{1.11}$$

Étant donné qu'une image peut être considérée comme une distribution statistique de niveaux de gris. Les statistiques de premier ordre (Press et al., 1987), calculées à partir de l'histogramme de cette image, permettent d'effectuer une étude des textures y contenues (analyse texturelle). Les attributs extraits sont respectivement la moyenne, l'écart type et la variance de la distribution des niveaux de gris dans l'image. La caractéristique correspondante est alors:

$$\bar{x} = (\bar{M}_{oy},\ \bar{E}_{type},\ \bar{V}_{ar})^t \tag{1.12}$$

Sachant que:

$$\bar{M}_{oy} = \frac{1}{ng}\sum_{i=1}^{ng} i\ hist(i), \tag{1.13}$$

$$\bar{E}_{type} = \sqrt{\frac{1}{ng}\sum_{i=1}^{ng} (i - Moy)^2\ hist(i)}, \tag{1.14}$$

$$\bar{V}_{ar} = \frac{1}{N}\sum_{p=1}^{N} (NG(p) - Moy)^2. \tag{1.15}$$

1.5.6 La matrice de co-occurrences

La matrice de co-occurrences (Haralick et al., 1973), calculée à partir de l'image niveau de gris NG, représente une étude statistique du second ordre de l'image. Elle ressemble à la matrice correlogramme, présentée dans la section 1.5.4, sauf que la notion de voisinage ici est définie par une distance $Dist$ suivant une direction Dir. La matrice de co-ocurrences établit la probabilité de retrouver dans ce voisinage le niveau de gris j en partant d'un niveau de gris i:

$$SGLD_{i,j}^{(Dist,Dir)} = \underset{p_1 \in NG_i}{prob}\ [p_2 \in NG_j \quad / \ |p_1 - p_2| = Dist, \llcorner(p_1,p_2) = Dir]. \tag{1.16}$$

Il est possible de calculer sur cette matrice des attributs statistiques classiques comme la moyenne et la variance, ainsi que d'autres attributs telles que l'entropie, l'inertie, la

5. Cette probabilité correspond au nombre d'occurrences des pixels p_2 de couleur c_j dans le voisinage du pixel p_1 ($p_1 \in \mathcal{I}_{c_i}$). Le voisinage considéré est fixé par la distance $Dist$.

corrélation, etc... Selon Haralick, treize attributs de texture peuvent être calculés (Haralick et al., 1973). Uniquement les sept les plus pertinents et plus utilisés sont retenus:

$$Moyenne = \frac{1}{ng^2} \sum_{i=1}^{ng} \sum_{j=1}^{ng} SGLD_{i,j}^{(Dist,Dir)} \tag{1.17}$$

$$Energie = \sum_{i=1}^{ng} \sum_{j=1}^{ng} [SGLD_{i,j}^{(Dist,Dir)}]^2 \tag{1.18}$$

$$Variance = \sum_{i=1}^{ng} \sum_{j=1}^{ng} (i - Moyenne)^2 SGLD_{i,j}^{(Dist,Dir)} \tag{1.19}$$

$$Contraste = \sum_{i=1}^{ng} \sum_{j=1}^{ng} (i-j)^2 SGLD_{i,j}^{(Dist,Dir)} \tag{1.20}$$

$$Entropie = -\sum_{i=1}^{ng} \sum_{j=1}^{ng} SGLD_{i,j}^{(Dist,Dir)} Log[SGLD_{i,j}^{(Dist,Dir)}] \tag{1.21}$$

$$Homogeneite = \sum_{i=1}^{ng} \sum_{j=1}^{ng} \frac{1}{1+(i-j)^2} SGLD_{i,j}^{(Dist,Dir)} \tag{1.22}$$

$$Correlation = \left(\sum_{i=1}^{ng} \sum_{j=1}^{ng} (i-\mu_x)(j-\mu_y) SGLD_{i,j}^{(Dist,Dir)} \right) \ / \ (\sigma_x \sigma_y) \tag{1.23}$$

avec μ_x, μ_y et σ_x, σ_y sont les moyennes et les variances respectives des deux distributions marginales $p_x(i)$ et $p_y(j)$:

$$\mu_x = \sum_{i=1}^{ng} i \ p_x(i), \ \ \sigma_x^2 = \sum_{i=1}^{ng} (i-\mu_x)^2 p_x(i) \tag{1.24}$$

$$\mu_y = \sum_{j=1}^{ng} j \ p_y(j), \ \ \sigma_y^2 = \sum_{j=1}^{ng} (j-\mu_y)^2 p_y(j) \tag{1.25}$$

$$\tag{1.26}$$

sachant que $p_x(i)$ et $p_y(j)$ sont calculées comme suit:

$$p_x(i) = \sum_{j=1}^{ng} SGLD_{i,j}^{(Dist,Dir)}, \ \ p_y(j) = \sum_{i=1}^{ng} SGLD_{i,j}^{(Dist,Dir)} \tag{1.27}$$

1.5.7 Les transformées en ondelettes de Daubeshie

Pour un niveau de décomposition l, les coefficients d'ondelettes de Daubeshie sont calculés comme suit, i,$j \in [0,1]$:

$$c_{ij}^{l}(x,y) = \begin{cases} \sum_{\alpha}\sum_{\beta} NG(\alpha,\ \beta)\ h_i(2x-\alpha)\ h_j(2y-\beta), & l=1, \\ \sum_{\alpha}\sum_{\beta} c_{00}^{l-1}(\alpha,\ \beta)\ h_i(2x-\alpha)\ h_j(2y-\beta), & l \geq 2, \end{cases} \tag{1.28}$$

sachant que $h_i(k)$ et $h_j(k)$ représentent les filtres de Daubeshie respectivement selon les lignes et les colonnes.

Les attributs textures considérés, dans notre cas, sont les coefficients d'ondelettes de Daubechie correspondants à deux niveaux de décomposition. La figure 1.12 illustre le principe de base de la répartition de ces coefficients pour chaque image.

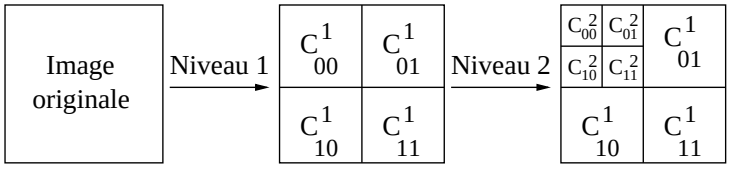


FIG. 1.12 – *Coefficients d'ondelettes pour deux niveaux de décomposition.*

1.5.8 Le descripteur GIST

Le descripteur GIST est une caractéristique globale qui tente d'imiter le système de perception humain, (Potter, 1975), (Biederman, 1982), (Tversky and Hemenway, 1983), pour identifier rapidement les principales régions de couleur et de texture dans une image. Il a été proposé initialement pour développer des représentations de scènes à faible dimension (Oliva and Torralba, 2001). GIST a montré récemment de très bon résultats pour la recherche et la reconnaissance d'images (Li et al., 2008).

Plusieurs modèles GIST utilisant différents types de caractéristiques de bas niveau ont été récemment proposés (Oliva and Torralba, 2001), (Renniger and Malik, 2004), (Siagian and Itti, 2007), (Torralba et al., 2003). Le modèle d'Oliva et Torralba (Oliva and Torralba, 2001) est adopté pour l'extraction du descripteur GIST. Initialement, l'image est pré-filtrée, ensuite elle est segmentée systématiquement en carrés selon une grille 4×4. Enfin, les histogrammes des orientations du gradient sont extraits en se basant sur les transformées de Fourrier calculées à partir de plusieurs échelles de l'image. L'implémentation GIST[6] utilisée, considère à l'entrée les images de taille carrée fixe et produit un vecteur de dimension 960 attributs, respectivement au trois chrominances R, V et B de l'image. Un ensemble d'échantillons des descripteurs GIST extraits de différentes images de la base COREL est présenté dans la figure 1.13.

1.5.9 La matrice norme du gradient

La norme du gradient (Delingette and Montagnat, 2001) présente le degré de différence d'intensité d'un pixel par rapport à ces voisins. Il s'agit de traiter le contour de l'image ce

6. L'implémentation MatLab utilisée du descripteur GIST est disponible sur la page web de Antonio Torralba à l'adresse suivante: http://people.csail.mit.edu/torralba/code/spatialenvelope.

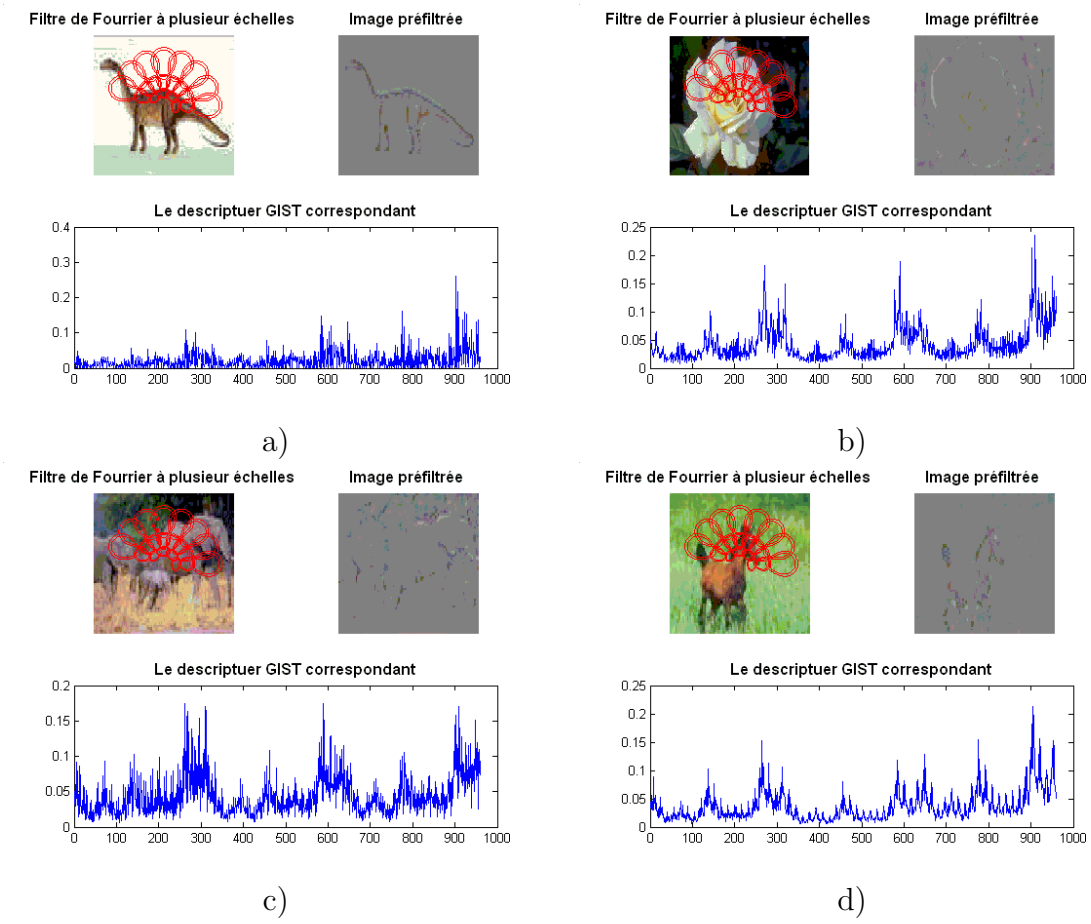


FIG. 1.13 – *Échantillons des descripteurs GIST extraits pour différentes images de la base COREL.*

qui permet de distinguer les différentes régions et les différents objets dans cette image. La norme du gradient d'une image peut être calculée à l'aide de différents filtres, à savoir le filtre de *Sobel*, le filtre de *Prewitt*, le filtre *Laplacien*, etc... Le filtre de *Sobel*, considéré dans notre cas, suppose que le contenu des différentes sous régions de l'image est homogène, ainsi le contour peut être détecté en se basant sur les discontinuités des niveaux de gris dans l'image. Initialement, l'image niveau de gris NG subit une convolution avec les masques de *Sobel*, h_X et h_Y, pour le calcul des gradients directionnels selon x et y (approximations des dérivées horizontale et verticale). Deux images, appelées G_X et G_Y, sont obtenues suite à ces convolutions. G_X et G_Y contiennent, respectivement, en chaque point des approximations de la dérivée horizontale et verticale de chaque pixel de l'image. Ces images sont calculées comme suit:

$$G_X = h_X \otimes NG, \qquad G_Y = h_Y \otimes NG \tag{1.29}$$

avec h_X et h_Y sont des matrices de convolution de taille 3×3:

$$h_X = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \qquad h_Y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

La matrice norme du gradient correspondante combine, en chaque pixel de l'image, les approximations des gradients horizontaux et verticaux pour obtenir une approximation globale de la norme du gradient:

$$G = \sqrt{G_X^2 + G_Y^2} \tag{1.30}$$

1.5.10 Les moments de Hu

Les Moments sont des attributs de forme très puissants. Un moment est une somme pondérée de tous les pixels en fonction de leurs positions dans l'image (plus de détails sont fournis dans (Teague, 1980)). En 1962, les sept moments de Hu ont été proposés (Hu, 1962):

$$\begin{cases} \phi_1 = \mu_{20} + \mu_{02} \\ \phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\ \phi_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\ \phi_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\ \phi_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03}) \\ \qquad (\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ \phi_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\ \phi_7 = (3\mu_{21} - \mu_{03})(\mu_{3,0} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{0,3})^2] - (\mu_{3,0} - 3\mu_{12}) \\ \qquad (\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{cases} \tag{1.31}$$

Les moments de Hu sont invariants aux différentes transformations géométriques décrites dans la section 1.2.1.3.

1.5.11 Optimisation

Suite à l'extraction de cette variété de caractéristiques, quelques traitements sont apportés à certaines d'entre elles (Kachouri et al., 2008b). Ces traitements permettent principalement de réduire la dimension de ces caractéristiques, tout en préservant leur pertinence de description voire même en l'améliorant. Dans ce contexte, l'optimisation des caractéristiques suivantes est effectuée:

- **La matrice correlogramme (section 1.5.4):** La caractéristique auto-correlogramme (Huang et al., 1999) est utilisée à la place de la matrice correlogramme. En fait, l'auto-correlogramme est défini comme une sous-caractéristique du correlogramme, il s'agit de sa diagonale. Elle capture seulement la corrélation spatiale entre les niveaux de couleur identiques dans l'image.
- **Les ondelettes de Daubeshie (section 1.5.7):** Deux niveaux de décomposition en ondelettes rendent sept ensembles de coefficients, tel qu'il est indiqué par la figure 1.12. Une réduction du nombre de ces coefficients est effectuée par le calcul d'une seule valeur résultante pour chaque ensemble. Il s'agit de l'énergie sous-bande de tous les coefficients d'ondelettes appartenant à l'ensemble considéré. Vu que les coefficients bas niveau de l'approximation c_{00}^2 ne sont pas adaptés à une analyse de

texture, une mesure des directions indépendantes du signal haute fréquence correspondant est réalisée par un filtrage Laplacien:

$$c^2_{00}(x,y) = \sum_{\alpha}\sum_{\beta} c^2_{00}(\alpha,\, \beta)\; h_{LAP}(x-\alpha,\, y-\beta), \tag{1.32}$$

avec h_{LAP} est la matrice de convolution (3×3) du filtre Laplacien:

$$h_{LAP} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Les attributs textures extraits des différents détails ainsi que du filtrage Laplacian de l'approximation sont calculés comme suit:

$$e^l_{ij} = \frac{1}{KZ}\sum_{k=1}^{K}\sum_{z=1}^{Z} |c^l_{ij}(k,z)|^2, \qquad i,j \in [0,1]\ et\ l \in [1,2], \tag{1.33}$$

avec K et Z sont les deux dimensions respectives de l'ensemble des coefficients d'ondelettes c^l_{ij} (plus de détails sont fournis dans (Serrano et al., 2004)).

– **La matrice norme du gradient (section 1.5.9):** Généralement, les images sont caractérisées à l'aide d'un histogramme de la norme du gradient. Il est également possible d'utiliser des histobins de 8, 16 ou 32, ... bins. Un histobin est une forme réduite de l'histogramme, où chaque bin représente la somme d'un ensemble d'éléments voisins de l'histogramme. Le nombre de voisins est déterminé par le nombre de bin dans l'histobin. Ces méthodes décrivent l'information texture dans l'image. Pour mettre en valeur l'aspect contour, la description norme du gradient est exploitée autrement dans notre travail. Une projection selon les deux axes de cette matrice est effectuée:

$$P_X(x) = \frac{1}{max\ G}\sum_{y} G(x,y), \quad et \quad P_Y(y) = \frac{1}{max\ G}\sum_{x} G(x,y) \tag{1.34}$$

Les deux vecteurs normalisés P_X et P_Y représentent une description forme de l'image à partir de sa matrice norme du gradient correspondante. Ils reflètent respectivement la variation verticale et horizontale du contour. Cette méthode permet de réduire la dimension de la matrice norme du gradient, tout en préservant sa qualité de description. La figure 1.14 montre la capacité de discrimination de cette nouvelle caractéristique des normes de gradient entre deux images de classes différentes de la base COREL.

– **Les moments de Hu (section 1.5.10):** Afin de considérer le contour plus que toutes autres informations dans l'image, les moments de Hu sont calculés à partir de la matrice norme du gradient au lieu de l'image initiale. Ceci permet d'enrichir davantage la description forme de cette caractéristique.

1.5.12 Conclusion

Dans cette section, l'ensemble des caractéristiques de différents types, extrait pour la description d'images dans les bases hétérogènes, a été présenté. Une évaluation de cet ensemble est effectuée (Kachouri et al., 2008a). Les résultats de cette évaluation, pour différentes bases d'images hétérogènes, sont illustrés et discutés dans la section suivante.

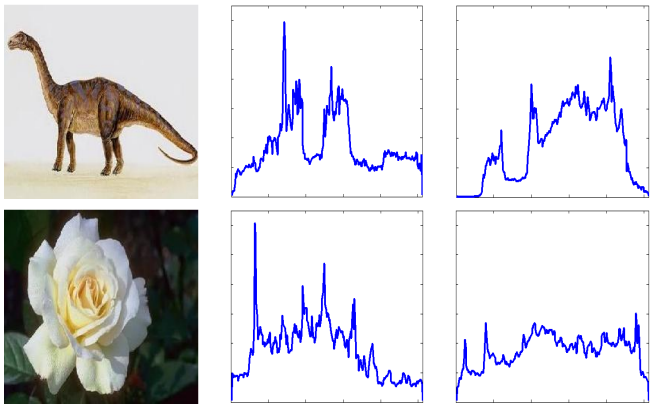


FIG. 1.14 – *Nouvelle forme de la caractéristique des normes du gradient des image Dinosaure et Rose: a) la projection selon x et b) la projection selon y.*

Dans le reste de ce manuscrit, les différentes caractéristiques couleur moyennes sont notées, respectivement aux espaces couleurs correspondants, par RVB, HSV, HMMD et YCrCb, l'histogramme couleur par Hist, l'auto-correlo-gramme par A-Cor, les statistiques de premiers ordres par FOS, l'ensemble des attributs extraits de la matrice de co-occurence par SGLD, les coefficients condensés de la transformation d'ondelette de Daubeshie à deux niveaux de décomposition par Daub, le descripteur GIST par GIST, les deux projections de la matrice normes du gradient par G-norm et les moments invariants par Hu.

1.6 Évaluation des caractéristiques

Nous avons implémenté sous MatLab les différentes caractéristiques décrites dans la section 1.5. Ensuite, afin d'évaluer le pouvoir discriminant individuel de chacune de ces caractéristiques, nous avons entraîné l'ensemble des modèles SVMs correspondants sur les bases d'images hétérogènes utilisées (section 1.3). Dans un premier temps, quatre sous-bases d'images, constituées respectivement par 3, 5, 7 et 10 classes sont utilisées à partir de la base COREL et la base COLUMBIA. Les performances de classification correspondantes à chaque modèle, pour les différentes sous-bases employées, sont mesurées à travers les deux critères de qualité: précision et rappel (plus de détails ont été fournis dans la section 1.2.3.3).

Les valeurs moyennes de la précision (P_{moy}) et du rappel (R_{moy}) des modèles évalués, pour les quatre sous-bases COREL, sont présentées par les tableaux 1.2 et 1.3. Les meilleures et les pires performances dans ces tableaux sont affichées en gras pour les différentes sous-bases d'images utilisées. Similairement, les tableaux 1.4 et 1.5 présentent les performances de classification des différents modèles SVMs correspondants à l'ensemble des caractéristiques extrait, pour les quatre sous-bases COLUMBIA. Idem, les meilleures et les pires performances dans ces tableaux sont affichées en gras.

Les tableaux 1.2, 1.3, 1.4 et 1.5 montrent que les performances de classification illustrées varient bien évidement en fonction des différentes caractéristiques employées. Les performances de classification d'un modèle d'une caractéristique donnée varient encore en fonction des différentes sous-bases d'images utilisées. Ce qui prouve que le pouvoir discriminant de chaque caractéristique n'est pas absolu, mais il varie sensiblement en fonction du contenu de la base d'image considérée.

TAB. 1.2 – *Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques couleurs pour différentes sous-bases COREL.*

Sous-bases COREL		RVB	HSV	HMMD	YCrCb	Hist	A-Cor
3 classes	P_{moy}	0.8320	0.9071	0.9338	**0.8129**	**0.9643**	**0.9643**
	R_{moy}	0.8000	0.9067	0.9333	**0.7733**	**0.9600**	**0.9600**
5 classes	P_{moy}	**0.6735**	0.6889	0.7384	0.6897	0.8928	**0.9338**
	R_{moy}	**0.6640**	0.6880	0.7280	**0.6640**	0.8800	**0.9280**
7 classes	P_{moy}	0.6236	0.5684	**0.5657**	0.6444	**0.8593**	0.8549
	R_{moy}	0.6057	**0.5429**	0.5543	0.6114	**0.8400**	0.8343
10 classes	P_{moy}	0.5655	**0.4585**	0.4830	0.5750	**0.8055**	0.7996
	R_{moy}	0.5560	**0.4480**	0.4760	0.5720	0.7800	**0.7840**

TAB. 1.3 – *Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques textures et formes pour différentes sous-bases COREL.*

Sous-bases COREL		FOS	SGLD	Daub	GIST	G-norm	Hu
3 classes	P_{moy}	**0.6801**	**0.8981**	0.7690	0.8919	0.8648	0.7351
	R_{moy}	**0.6800**	**0.8800**	0.7733	0.8400	0.8267	0.7333
5 classes	P_{moy}	0.5858	0.7439	**0.3903**	**0.8457**	0.7016	0.6451
	R_{moy}	0.5680	**0.7200**	**0.4640**	0.5440	0.5840	0.6480
7 classes	P_{moy}	0.5087	0.6844	**0.2881**	**0.8231**	0.7288	0.5147
	R_{moy}	0.5143	**0.6743**	0.5314	**0.3714**	0.4857	0.5200
10 classes	P_{moy}	0.4228	0.6387	**0.2104**	**0.7089**	0.6776	0.4336
	R_{moy}	0.4520	**0.6160**	**0.2880**	0.4760	0.4760	0.4520

D'après le tableau 1.2, la caractéristique Hist est la plus performante parmi les différentes caractéristiques couleurs employées, pour la classification des sous-bases COREL avec 3, 7 et 10 classes. Cependant, la caractéristique A-Cor est plus pertinente pour la classification de la sous-base COREL avec 5 classes. De même, les caractéristiques couleurs les moins performantes pour la classifications des sous-bases COREL avec 3, 5, 7 et 10 classes sont respectivement YCrCb, RVB, HMMD et HSV. D'après le tableau 1.3, parmi les différentes caractéristiques textures et formes employées, GIST aboutit aux meilleures performances de classification, pour les sous-bases COREL avec 5, 7 et 10 classes. Pour la classification de la sous-base 3 classes, c'est la caractéristique SGLD qui s'avère la plus pertinente. D'après les tableaux 1.4 et 1.5, les performances de classification des différentes caractéristiques sont nettement meilleures. Ceci est dû essentiellement au contenu moins complexe des images de la base COLUMBIA par rapport au contenu des images de la base COREL (plus de détails sont fournis dans la section 1.3). Cependant, le pouvoir discriminant de chaque caractéristique dépend toujours de la base d'image considérée. On constate, par exemple, que la caractéristique YCrCb est la plus pertinente, conjointement avec la caractéristique RVB, pour la classification des sous-bases COLUMBIA avec 3, 5 et 7 classes. Mais elle ne l'est pas pour la classification de la sous-base COLUMBIA 10 classes (tableau 1.4). De plus, les caractéristiques textures et formes les moins performantes pour la classification des sous-bases COLUMBIA sont FOS pour la classification de la sous-base 3 classes, Hu pour la classification de la sous-base 7 classes, et Daub pour la classification des sous-bases 5 et 10 classes (tableau 1.5).

TAB. 1.4 – *Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques couleurs pour différentes sous-bases COLUMBIA.*

Sous-bases COLUMBIA		RVB	HSV	HMMD	YCrCb	Hist	A-Cor
3 classes	P_{moy}	**1.0000**	**1.0000**	**0.7444**	**1.0000**	0.9733	0.9506
	R_{moy}	**1.0000**	**1.0000**	**0.7391**	**1.0000**	0.9710	0.9420
5 classes	P_{moy}	**1.0000**	0.9840	**0.8109**	**1.0000**	0.9840	0.9704
	R_{moy}	**1.0000**	0.9826	**0.8000**	**1.0000**	0.9826	0.9652
7 classes	P_{moy}	**0.9835**	0.9330	**0.7012**	**0.9835**	0.8919	0.8664
	R_{moy}	**0.9814**	0.9317	**0.6832**	**0.9814**	0.8820	0.8571
10 classes	P_{moy}	**0.9582**	0.9454	**0.7888**	0.9439	0.9075	0.8623
	R_{moy}	**0.9435**	**0.9435**	**0.7783**	0.9217	0.9000	0.8565

TAB. 1.5 – *Évaluation de l'efficacité de classification des modèles individuels associés aux caractéristiques textures et formes pour différentes sous-bases COLUMBIA.*

Sous-bases COLUMBIA		FOS	SGLD	Daub	GIST	G-norm	Hu
3 classes	P_{moy}	**0.8275**	**1.0000**	0.8537	0.9310	0.9367	0.8632
	R_{moy}	0.7971	**1.0000**	**0.7391**	0.9130	0.9275	0.7681
5 classes	P_{moy}	0.8882	**0.9437**	**0.6902**	0.9314	0.9007	0.8766
	R_{moy}	0.8696	**0.9217**	**0.7565**	0.8957	0.8783	0.8087
7 classes	P_{moy}	0.8333	0.8478	0.5813	**0.9069**	0.8163	**0.5214**
	R_{moy}	**0.7950**	**0.7950**	0.5528	0.7329	0.7702	**0.5466**
10 classes	P_{moy}	0.6896	0.8376	**0.4691**	**0.9271**	0.8277	0.4971
	R_{moy}	0.7435	**0.8000**	0.5783	0.7304	0.7391	**0.5304**

Une évaluation plus large des différentes caractéristiques a été effectuée sur 23 classes des bases COREL et CALTECH-256. La figure 1.15 illustre respectivement les taux de classification correspondants à l'ensemble des caractéristiques employées pour différentes sous-bases d'images COREL et CALTECH-256. Les sous-bases utilisées sont respectivement constituées avec 3, 5, 7, 10, 13, 15, 17, 20 et 23 classes pour chacune des deux bases d'images.

D'après cette évaluation, nous pouvons constater que parmi l'ensemble des caractéristiques extraites, Hist, A-Cor, GIST et G-norme sont souvent les plus pertinentes et la caractéristique Daub est pratiquement celle la moins performante. Cependant, les courbes affichées par la figure 1.15 prouvent davantage que les performances de classification d'un modèle de caractéristique donné dépendent essentiellement du contenu de la base d'images considérée. Nous constatons par exemple que HMMD s'avère plus performante que Hu pour la classification de toutes les sous-bases d'images COREL. Pour la classification de la base CALTECH-256, Hu devient légèrement plus efficace que HMMD. Par ailleurs, pour la même base d'images COREL, les taux de classification des caractéristiques FOS et SGLD varient en fonction des différentes sous-bases utilisées. Effectivement, pour les sous-bases 3, 5, 7 et 10 classes, SGLD est plus pertinente que FOS. Cependant, pour le reste des sous-bases COREL, les taux de classification de la caractéristique FOS deviennent meilleurs que ceux de la caractéristique SGLD.

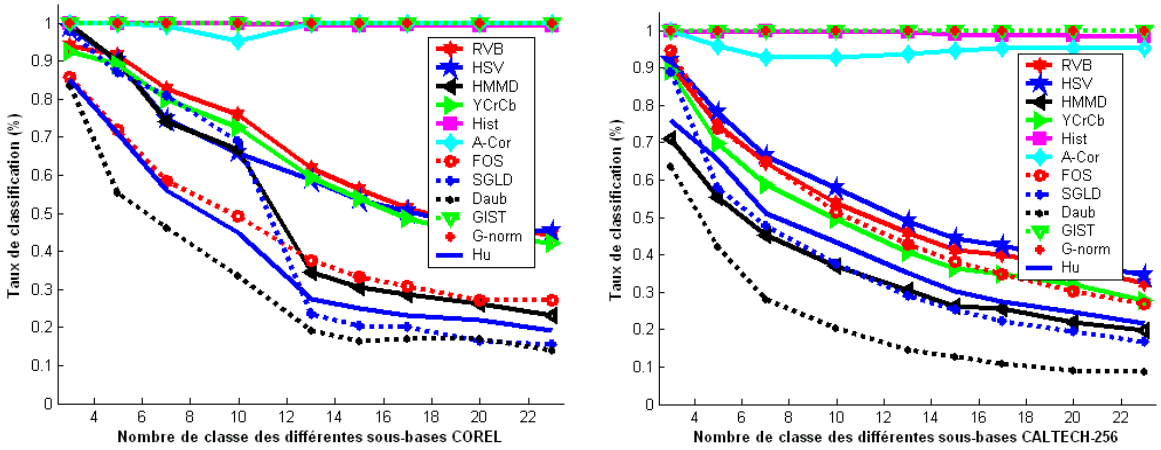


FIG. 1.15 – *Taux de classification des modèles individuels associés à l'ensemble des caractéristiques employées pour différentes sous-bases d'images COREL (à gauche) et CALTECH-256 (à droite).*

1.7 Conclusion

Dans ce chapitre, les éléments fondamentaux d'un système de reconnaissance d'images par le contenu ont été présentés. Une grande attention a été allouée à la classification et à la description d'images dans les bases hétérogènes. Un état-de-l'art des techniques de classification d'images a été exposé. Ensuite, un ensemble de caractéristiques bas niveau de différents types, ainsi qu'une variété d'optimisation de certaines caractéristiques ont été décrits. Une large évaluation de la pertinence de chacune des caractéristiques employées a été effectuée à travers différentes sous-bases appartenant à trois bases d'images hétérogènes de référence. Les résultats de cette évaluation ont prouvé que le pouvoir discriminant de chaque caractéristique n'est guère absolu et que les performances de classification des différentes caractéristiques dépendent fortement du contenu de la base d'image considérée.

C'est dans ce cadre que nous proposons, dans le chapitre suivant, la sélection adaptative des caractéristiques les plus pertinentes pour la reconnaissance d'images dans les bases hétérogènes.

Chapitre 2

Sélection Adaptative des caractéristiques les plus pertinentes

Nous traitons dans ce chapitre la sélection des caractéristiques pertinentes pour la reconnaissance d'images dans les bases hétérogènes. Après une étude des deux approches de sélection existantes (*Filtres* et *Enveloppantes*), une méthode originale de sélection adaptative des caractéristiques est proposée. Cette sélection est totalement adaptée à l'ensemble des caractéristiques employées et au contenu de la base d'images hétérogène utilisée.

2.1 Introduction

Le nombre ainsi que l'hétérogénéité des caractéristiques extraites pour la description du contenu d'images dans les bases hétérogènes, limitent généralement les performances de classification. En effet, ces caractéristiques ne sont pas nécessairement toutes pertinentes et utiles pour la classification des images en phase de reconnaissance. Le problème rencontré dans ce cas est le choix des caractéristiques pertinentes. Dans cet objectif, des méthodes de sélection des caractéristiques s'avèrent indispensables. Plusieurs méthodes de sélection ont été proposées et publiées (Weston et al., 2000), (Tieu and Viola, 2004), (Muller et al., 2004), (Kudo and Sklansky, 2000). Toutefois, il n'y a toujours pas une méthode générique qui assure automatiquement la sélection des caractéristiques pertinentes en fonction du contenu de la base d'images considérée. Dans cet objectif, nous proposons une nouvelle méthode de sélection. Cette méthode permet d'augmenter les performances de la description par une sélection adaptative d'un sous-ensemble de caractéristiques pertinentes en fonction du contenu de la base d'images et du classifieur utilisés.

Dans ce chapitre, les deux catégories de méthodes de sélection existantes dans la littérature, à savoir les méthodes *filtres* et les méthodes *enveloppantes* sont décrites dans la section 2.2. Ensuite, deux méthodes de sélection assez connues sont présentées dans la section 2.3. La proposition de notre méthode de sélection adaptative est détaillée dans la section 2.4. Le chapitre est enfin clôturé par la section 2.5, où on présente une évaluation des différentes méthodes de sélection discutées.

2.2 Méthodes de sélection

La sélection d'un sous-ensemble de caractéristiques qui sera considéré comme le plus pertinent est un domaine assez étudié et les algorithmes proposés sur ce sujet font l'objet, depuis plusieurs années, d'une littérature abondante (Kohavi and John, 1997), (Guyon and Elisseeff, 2003). Les méthodes de sélection proposées s'avèrent assez diverses et variées, cependant, il n'existe pas de cadre théorique unifacteur pour toutes ces méthodes. Deux branches principales peuvent être distinguées en littérature, les méthodes filtres et les méthodes enveloppantes.

2.2.1 Les méthodes Filtres

Les méthodes filtres, connues en anglais par *Filters*, sont très rapides. Elles reposent sur des considérations théoriques, qui permettent généralement de mieux comprendre les relations de dépendance entre les attributs (Yu and Liu, 2004). Le principe est de sélectionner un sous-ensemble d'attributs selon un certain critère, comme la corrélation par exemple. Le sous-ensemble permettant d'optimiser ce critère sera ainsi retenu (Kohavi and John, 1997). On cite parmi les méthodes de sélection appartenant à ce type l'Analyse en Composantes Principales (ACP) (Jolliffe, 1986) et l'Analyse Linéaire Discriminante (ALD)(Fukunaga, 1990). Cependant, comme elles ne prennent pas en compte le classifieur utilisé en phase de généralisation, les méthodes filtres restent souvent incapables de garantir des taux de reconnaissance considérables (Addison and Macintyre, 2003).

2.2.2 Les méthodes Enveloppantes

Bien que conceptuellement plus simples que les filtres, les méthodes de sélection enveloppantes, connues en anglais par *Wrappers*, ont été introduites plus récemment (John et al., 2000). Ce type de sélection fait appel à un algorithme de classification. Il s'agit, en fait, d'une approche qui teste différents sous-ensembles de jeux de caractéristiques et qui choisit le sous-ensemble donnant les meilleures performances. Ainsi, le sous-ensemble de caractéristiques sélectionné est bien adapté au classifieur utilisé. Les taux de reconnaissance sont généralement élevés car la sélection prend en compte le biais intrinsèque de l'algorithme de classification (Bi et al., 2003), (Zhu et al., 2004), (Chen et al., 2006). L'inconvénient majeur de cette technique de sélection est son besoin coûteux en temps de calcul, surtout quand le nombre d'images croît. Généralement, le système doit être relancé plusieurs fois avec différentes combinaisons de caractéristiques pour conserver au final la meilleure solution (Kohavi and John, 1997).

2.3 Travaux connexes

Plusieurs comparaisons entre les différentes méthodes de sélection, qui existent dans la littérature, ont été effectuées. Un récapitulatif des méthodes les plus populaires est présenté dans (Kudo and Sklansky, 2000). Dans cette section, une évaluation de deux méthodes de sélection assez connues et très utilisées est présentée. Ces deux méthodes sont l'ACP comme technique filtre et la sélection basée sur les Algorithmes Génétiques (AGs) comme technique enveloppante.

2.3.1 L'Analyse en Composantes Principales

L'analyse en composantes principales (ACP) est sans doute celle la plus connue et la plus utilisée dans le domaine de l'analyse de données. Introduite par les travaux de Hotelling et Loève (Hotelling, 1933), (Loève, 1948), l'ACP est une méthode factorielle d'analyse de données multi-dimensionnelles. Elle détermine une décomposition d'un vecteur d'attribut en composantes décorrélées et orthogonales tout en ajustant au mieux sa distribution. Les composantes énumérées sont dites principales et sont ordonnées par ordre décroissant selon leur degré d'ajustement. Ces nouvelles variables (les composantes principales, qu'on note en ce qui suit CPs) sont en fait des combinaisons linéaires des attributs de départ. Elles correspondent essentiellement aux directions de la plus grande variance (figure 2.1). L'ACP a été exposée en détails dans plusieurs travaux, nous en rappelons ici les points fondamentaux.

Soit N_{Im} le nombre d'images d'apprentissage, chaque image est représentée par une caractéristique x de dimension d, dans un espace d'attributs $\mathcal{F} = \mathbb{R}^d$. Le nuage correspondant des N_{Im} points, dits aussi individus ou observations, est noté par $X = [x_1,x_2,...,x_{N_{Im}}]$. L'objectif de l'ACP est de déterminer la base orthogonale ajustant au mieux ce nuage de points selon un critère géométrique. Elle détermine pour cela les directions successives de variance maximale qui correspondent également à cet optimum (voir figure 2.1, où la première composante correspond à la droite d'allongement maximum b_1). L'ordre fourni sur les CPs obtenues est utile d'une part à la caractérisation et menant d'autre part à une approximation modale permettant une représentation compacte de la distribution des observations. L'ACP caractérise ainsi la variabilité inhérente des attributs d'apprentissage en révélant leurs déformations principales relativement à une forme de référence,

classiquement, la forme moyenne. Par ailleurs, l'ACP est une méthode de second ordre car elle se base uniquement sur l'étude de la matrice de covariance des attributs.

Soient $\bar{X}$ la caractéristique moyenne des N_{Im} images d'apprentissage, et $\mathbf{C}$ la matrice de covariance, données par:

$$\bar{X} = \frac{1}{N_{Im}} \sum_{j=1}^{N_{Im}} x_j, \tag{2.1}$$

$$\mathbf{C} = \frac{1}{N_{Im}} \sum_{j=1}^{N_{Im}} d_{x_j} d_{x_j}^t, \qquad \text{avec } d_{x_j} = x_j - \bar{X},\ j = [1,2,...,N_{Im}]. \tag{2.2}$$

La diagonalisation de la matrice de covariance $\mathbf{C}$ produit la base modale Ψ:

$$\mathbf{C} = \Psi \Lambda \Psi, \qquad \text{où } \Lambda = diag(\lambda_1,\lambda_2,...,\lambda_d), \text{ avec } \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d. \tag{2.3}$$

Les composantes principales sélectionnées sont en fait les p premiers vecteurs propres λ_i, $i = [1,2,...,p]$, avec $p < d$. Une approximation modale peut être réalisée en conservant uniquement les p premieres composantes. Une forme approchée s'écrit alors:

$$X = \bar{X} + \Psi_p b_p, \qquad \text{avec } b_p = \Psi_p^t (X - \bar{X}), \tag{2.4}$$

où Ψ_p est une sous-matrice de Ψ contenant les p premiers vecteurs propres de $\mathbf{C}$ et définissant la base d'approximation modale. Le vecteur des amplitudes modales de déformations $\mathbb{B} = (b_1,b_2,...,b_p)^t$ représente une forme dans l'espace p-dimensionnel défini par les composantes principales retenues. Il fournit une représentation compacte de la distribution initiale des observations ($p < d$ voire $p \ll d$). La qualité de la représentation obtenue sur la base tronquée peut être évaluée par le pourcentage de variance relatif aux p premiers modes, c'est-à-dire par le pourcentage d'inertie total τ correspondant:

$$\tau = \frac{\sum_{i=1}^{p} \lambda_i}{\lambda_T} \times 100, \qquad \text{où } \lambda_T = \sum_{i=1}^{d} \lambda_i. \tag{2.5}$$

La figure 2.1 présente le principe géométrique de l'ACP.

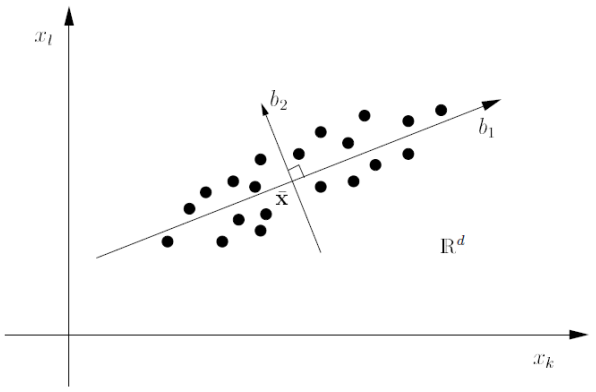


FIG. 2.1 – *Les deux premières directions principales (b_1 et b_2) d'un nuage de N_{Im} observations dans $\mathbb{R}^d$.*

2.3.2 La sélection basée sur les Algorithmes Génétiques

Les algorithmes génétiques (AGs), ont été développés dans les années 1970, comme une méthode d'optimisation efficace. Il existe un grand nombre de variétés d'AG, mais les principes de base restent inchangés. Dans un AG (Kudo and Sklansky, 2000), on a une population de base qui est souvent composée de chaînes de caractères correspondant chacune à un chromosome. Souvent chaque chromosome est une chaîne binaire de taille n. Les mécanismes d'un AG de base sont assez simples (Goldberg, 1991), il s'agit de faire des copies de chaînes et des échanges de morceaux de chaînes. Essentiellement, un AG est composé de trois opérateurs génétiques: la sélection, le croisement et la mutation. La sélection correspond à la copie de chaque chaîne en fonction des valeurs d'une fonction de sélectivité à optimiser. Après la sélection, le croisement est appliqué. Il correspond à un croisement des différentes chaînes tirées lors de la sélection. Le croisement se fait en deux étapes. Initialement, les nouveaux éléments sélectionnés sont appariés au hasard, puis chaque paire de chaînes subit un croisement. C'est-à-dire que l'on échange aléatoirement k éléments entre les différentes paires de chaînes. Pour finir la dernière étape est la mutation. Elle permet de modifier aléatoirement un élément de chaîne. C'est une modification qui n'apparaît qu'occasionnellement. Ces étapes sont effectuées jusqu'à ce que l'on obtienne des résultats qui n'évoluent plus, (Rebaine, 2005). La figure 2.2 présente les différentes étapes d'un AG.

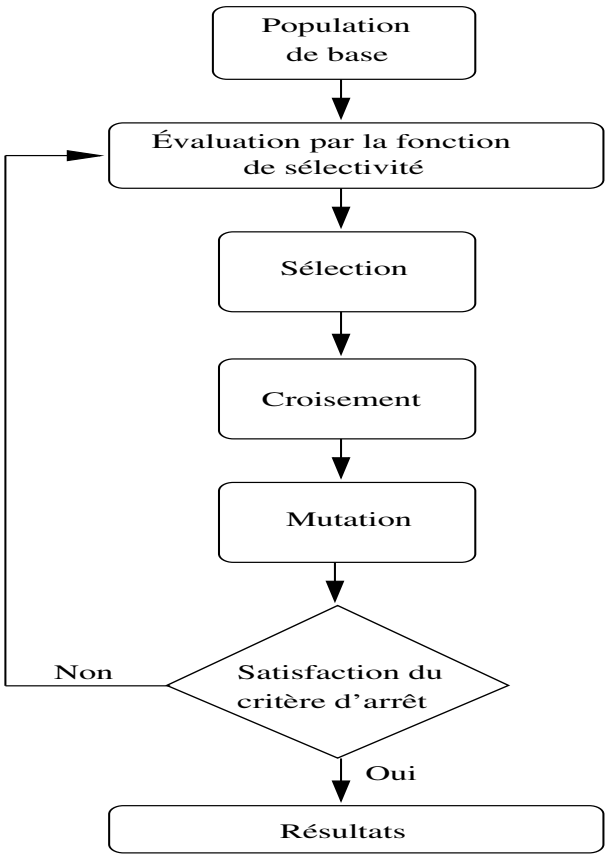


FIG. 2.2 – *Diagramme standard d'un algorithme génétique.*

En outre, l'AG est un algorithme efficace pour trouver des solutions optimales dans les problèmes d'optimisation. Plusieurs études ont été faites sur l'utilisation des AGs pour la sélection des caractéristiques (Vriesenga, 1995), (Siedlecki and Sklansky, 1989). Dans ce contexte, nous présentons brièvement la sélection basée sur les AGs.

L'AG est essentiellement un algorithme évolutionnaire qui cherche à résoudre des problèmes d'optimisation en utilisant les méthodes de l'évolution (la sélection naturelle dite aussi darwinienne et la génétique dans les systèmes biologiques). Il vise à sélectionner, à chaque génération (itération), de nouvelle population (sous-ensemble de caractéristiques dans notre cas) jusqu'à la satisfaction d'un certain critère d'arrêt. L'efficacité d'un AG repose essentiellement sur la pertinence de la fonction de sélectivité employée. La fonction de sélectivité autorise aux AGs d'explorer l'espace de population initiale (l'ensemble initial des caractéristiques extraites) et d'évaluer la qualité de la solution obtenue. Cette fonction doit être bien adaptée au problème, vu qu'elle a généralement des effets destructifs aussi bien que constructifs. Les AGs peuvent avoir tendance à converger vers des optimums locaux ou même des points arbitraires plutôt que l'optimum global du problème. Dans le contexte de sélection des caractéristiques, l'approche de sélection enveloppante basée sur les AGs intègre principalement les taux de classification correspondants aux différentes caractéristiques en tant que critères pour la conception de la fonction de sélectivité (Zhuo et al., 2008).

2.4 La Sélection Adaptative des Caractéristiques pertinentes

2.4.1 Introduction

De nouvelles méthodes qui combinent les techniques filtres et enveloppantes sont récemment proposées (Guyon and Elisseeff, 2003). Elles permettent de profiter à la fois de la rapidité des méthodes filtres et de l'adaptation des méthodes enveloppantes entre les caractéristiques sélectionnées et le classifieur utilisé. Dans ce même contexte, la méthode de Sélection Adaptative des Caractéristiques pertinentes est présentée dans cette section. Dans la suite du document, nous désignons cette méthode par l'acronyme SAC. Pour la partie enveloppante de la méthode, les SVMs (Vapnik, 1999) sont employés. Ces classifieurs sont présentés dans le chapitre suivant. Les SVMs sont utilisés pour l'évaluation de l'ensemble des caractéristiques extraites en phase d'apprentissage. Par conséquent, les mesures de pertinence effectuées pour les différentes caractéristiques sont basées sur les images d'apprentissage. D'où l'adaptation du pouvoir discriminant de chaque caractéristique évaluée avec le contenu de la base d'images considérée. Le Discriminant Linéaire de Fisher, noté souvent par l'acronyme anglais FLD (*Fisher Linear Discriminent*) (Fisher, 1936), est ensuite utilisé pour l'aspect filtre de la méthode. FLD assure la sélection des caractéristiques les plus pertinentes en se basant sur la distribution linéaire des évaluations SVMs effectuées. Le principe de la méthode de sélection proposée, est détaillé ci après.

2.4.2 Apprentissage multiple

Face à une large description constituées par n caractéristiques différentes, les performances d'un apprentissage unique, considérant à la fois toutes les n caractéristiques, ne peuvent pas représenter le pouvoir discriminant de chacune des caractéristiques utilisées. De ce fait, un apprentissage multiple, permettant d'évaluer séparément les différentes caractéristiques, est adopté. Pour chaque apprentissage, les images de la base sont classifiées

en se basant sur la description d'une seule caractéristique. Les performances d'apprentissage résultantes déterminent le pouvoir discriminant de la caractéristique considérée.

Soit $\mathcal{B}_{App}$ une base d'apprentissage. On rappelle que N_{Im} est le nombre total des images hétérogènes de cette base et que $\mathcal{F}_i$ est l'espace d'attributs correspondant à la caractéristique x^i, $i = [1,2,...,n]$. Le nuage des N_{Im} observations dans $\mathcal{F}_i$ est noté par $X_i = [x_1^i,x_2^i,...,x_{N_{Im}}^i]$, $i = [1,2,...,n]$. Soit $\mathcal{C} = \{C_1,C_2,...,C_{N_C}\}$ l'ensemble des N_C classes contenues dans $\mathcal{B}_{App}$. Pour l'évaluation de l'ensemble des caractéristiques en fonction du contenu de $\mathcal{B}_{App}$, on entraîne le classificateur SVM en considérant chacune des caractéristiques candidates. C'est à dire on effectue, pour les n caractéristiques extraites, n apprentissages différents où l'on utilise à chaque fois une caractéristique unique. On note par $\mathcal{M} = \{M_1,M_2,...,M_n\}$ l'espace des modèles obtenus. Les n modèles appartenant à $\mathcal{M}$ sont les différentes hyper-surfaces qui correspondent aux n caractéristiques extraites. Ces hyper-surfaces représentent les n frontières de séparation entre les différentes classes dans $\mathcal{C}$. Pour chaque modèle de cet espace, les N_{Im} images de la base d'apprentissage sont divisées en deux groupes: groupe des images mal classifiées (ou erronées) et groupe des images correctement classifiées. On note dans ce qui suit, le nombre d'images dans ces deux groupes respectivement par $N_{Err}(M_i)$ et $N_{Corr}(M_i)$, $i = [1,2,...,n]$. Pour chaque modèle dans $\mathcal{M}$, il en résulte le taux d'erreur et le taux de classification en phase d'apprentissage (performance d'apprentissage), qu'on note respectivement $\forall i \in \{1,2,...,n\}$, $E(M_i)$ et $Per(M_i)$:

$$E(M_i) = \frac{N_{Err}(M_i)}{N_{Im}}, \qquad Per(M_i) = \frac{N_{Corr}(M_i)}{N_{Im}},\ i = [1,2,...,n] \tag{2.6}$$

Comme chaque caractéristique est associée à un modèle, on propose de réaliser la sélection des caractéristiques pertinentes à travers la sélection des modèles les plus performants dans $\mathcal{M}$. L'avantage de cette sélection est qu'elle assure une réduction du nombre de modèles, avec une faible augmentation dans la probabilité d'erreur de classification. Elle permet de déterminer un sous-ensemble de modèles pertinents et d'exclure le reste des modèles qui sont considérés comme non discriminants pour la classification de la base d'images considérée. Ceci permet non seulement d'alléger le traitement de généralisation en ligne, mais aussi d'éviter l'effet négatif des modèles non pertinents sur le résultat final de reconnaissance. Pour cela, une discrimination linéaire de Fisher est appliquée.

2.4.3 Discrimination linéaire de Fisher

L'objectif de la sélection des caractéristiques par la discrimination linéaire de Fisher est principalement l'identification d'une direction de l'espace des attributs sur laquelle les projections des caractéristiques sont particulièrement bien séparées. La sélection des caractéristiques les plus pertinentes devient donc une classification binaire de l'ensemble original des caractéristiques selon cette direction de Fisher. Deux classes de caractéristiques en résultent: la classe des caractéristiques *très pertinentes* et la classe des caractéristiques *peu pertinentes*. La direction de projection est souvent obtenue comme une combinaison linéaire des variables originales. Cependant, le discriminant de Fisher correspondant ne peut pas prétendre être toujours le meilleur sous-espace linéaire de projection pour une discrimination optimale de la pertinence des différentes caractéristiques. Dans la méthode de sélection SAC, nous proposons de considérer les performances d'apprentissage des modèles correspondants à ces caractéristiques pour l'identification d'une meilleure discrimination de Fisher.

Soit $\mathcal{T} = \{Per(M_{s1}),Per(M_{s2}),...,Per(M_{sn})\}$ la liste des performances d'apprentissage des différents modèles dans $\mathcal{M}$, organisées par ordre de pertinence décroissant ($Per(M_{s1}) \geq Per(M_{s2}) \geq ... \geq Per(M_{sn})$). Cette organisation est très importante puisqu'elle est bien adaptée à la fois aux caractéristiques employées et au contenu même de la base d'images considérée. Elle change dès que l'ensemble des caractéristiques ou le contenu de la base d'images change. À partir de la liste de performance d'apprentissage $\mathcal{T}$ pré-organisée, on détermine selon le critère discriminant de Fisher une valeur optimale F_S qui représente le seuil adaptatif de sélection des modèles les plus pertinents. F_S sépare ainsi la catégorie des performances d'apprentissage correspondantes aux modèles *très pertinents* de celles correspondantes aux modèles *peu pertinents* dans la liste $\mathcal{T}$. Pour cela, on commence par le calcul de $m_1(i)$ et $m_2(i)$, $\forall i \in [1,2,...,n]$. Ces deux valeurs représentent les deux moyennes des performances d'apprentissage situées, respectivement, de part et d'autre de la performance d'apprentissage $Per(M_{si})$ dans la liste $\mathcal{T}$:

$$m_1(i) = \frac{1}{i}\sum_{j=1}^{i} Per(M_{sj}),$$

$$m_2(i) = \frac{1}{n-i}\sum_{j=i+1}^{n} Per(M_{sj}), \quad i = [1,2,...,n] \tag{2.7}$$

En second lieu, en fonction de ces deux valeurs moyennes $m_1(i)$ et $m_2(i)$, on calcule les deux variances respectives $v_1^2(i)$ et $v_2^2(i)$:

$$v_1^2(i) = \sum_{j=1}^{i} |Per(M_{sj}) - m_1(i)|^2,$$

$$v_2^2(i) = \sum_{j=i+1}^{n} |Per(M_{sj}) - m_2(i)|^2, \quad i = [1,2,...,n] \tag{2.8}$$

Maximiser la marge de séparation entre les modèles pertinents et non pertinents revient essentiellement à maximiser le discriminant de Fisher $P(i)$:

$$P(i) = \frac{|m_1(i) - m_2(i)|}{v_1^2(i) + v_2^2(i)}, \quad i = [1,2,...,n] \tag{2.9}$$

Ainsi, le seuil adaptatif de sélection des modèles pertinents F_S est déterminé comme suit:

$$1 \leq F_S \leq n = \arg\max_i P(i) \tag{2.10}$$

$$1 \leq F_S \leq n : \begin{cases} F_S = n & \text{SAC préserve tous les modèles calculés} \\ F_S < n & \text{SAC élimine les modèles jugés non pertinents} \end{cases}$$

F_S représente en fait l'indice k de la performance d'apprentissage à partir de laquelle le reste des performances dans la liste $\mathcal{T}$ correspondent à des modèles considérés comme non assez pertinents. La liste des performances d'apprentissage des modèles sélectionnés est $\mathcal{T}_s = \{Per(M_{s1}),Per(M_{s2}),...,Per(M_{sk})\}$, avec $Per(M_{si})$ est la performance d'apprentissage du modèle sélectionné M_{si}, $i = [1,2,...,k]$. Les caractéristiques les plus pertinentes sont

celles correspondantes aux différents modèles sélectionnés M_{si}, $i = [1,2,...,k]$. En phase de généralisation de la classification hiérarchique multi-modèles, présentée dans le chapitre 4, uniquement le sous-ensemble des modèles sélectionnés $\mathcal{M}_s = \{M_{s1},M_{s2},...,M_{sk}\}$ est employé. Plus de détails sont fournis dans la section 4.2.4.

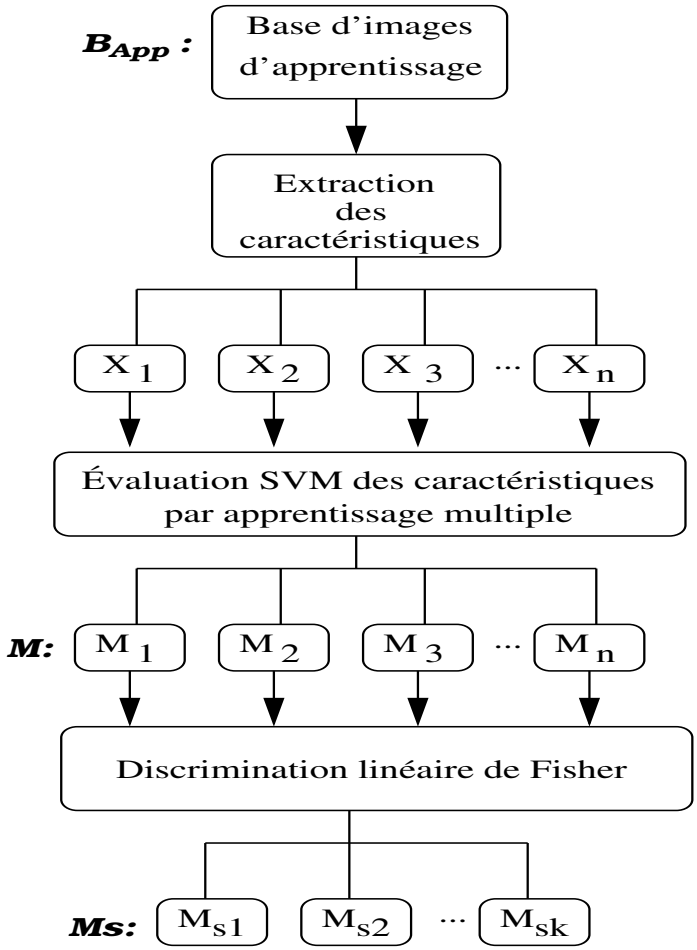


FIG. 2.3 – *Diagramme de la méthode de Sélection Adaptative des Caractéristiques les plus pertinentes qui correspondent aux modèles sélectionnés dans* $\mathcal{M}_s$.

La figure 2.3, illustre le diagramme de la méthode de sélection adaptative des caractéristiques les plus pertinentes. L'algorithme SAC est fourni dans ce qui suit:

Algorithme SAC

SAC $(\mathcal{B}_{app},\mathcal{F}_i)$, $i = 1,2,...,n$

1: **Pour** chaque caractéristiques x^i, $i = 1,2,...,n$ **Faire**
2: $M_i = Apprentissage\ SVM\ (X_i)$; Eq. 3.3
3: $Per(M_i) = \frac{N_{Corr}(M_i)}{N_{Im}}$; Eq. 2.6
4: **Fin Pour**
5: $\mathcal{T}$ = *Organisation des performances d'apprentissages des différents modèles* $(Per(M_i))$; $\forall i \in \{1,2,...,n\}$
6: $k = FLD\ (\mathcal{T})$; Eq. 2.10
7: $\mathcal{M}_s = \{M_{s1},M_{s2},...,M_{sk}\}$

2.4.4 Conclusion

Dans cette section, la méthode de sélection SAC a été présentée. Cette sélection permet d'identifier un sous-ensemble des caractéristiques les plus pertinentes, en se basant principalement sur le pouvoir discriminant individuel de ces caractéristiques. Après un apprentissage multiple, les caractéristiques sont sélectionnées en se basant sur une discrimination Fisher de leurs taux d'apprentissage. Cette sélection est totalement adaptée à la base d'images, les caractéristiques et le classifieur utilisés. Une évaluation de la méthode de sélection proposée est illustrée et discutée dans la section suivante.

2.5 Évaluation de la sélection des caractéristiques par ACP, AG et SAC

Dans cette section, nous présentons les paramètres de sélection obtenus, pour différentes sous-bases d'images COREL et COLUMBIA, avec les méthodes de sélection ACP, AG et SAC, décrites précédemment. L'ensemble des douze caractéristiques présentées dans le chapitre 1 est employé pour la réalisation de cette évaluation.

Comme technique de réduction de dimension, l'évaluation de la méthode ACP est effectuée selon une mesure du nombre des composantes principales sélectionnées à partir du vecteur d'attributs initial. La dimension de ce dernier correspond à la concaténation de la totalité des caractéristiques extraites, 3295 attributs dans notre cas. Cependant, les méthodes de sélection enveloppantes AG et SAC, employant conjointement le classifieur SVM pour l'évaluation de caractéristiques, sont expertisées selon la pertinence des caractéristiques sélectionnées. Pour cela, le seuil de sélection (c'est-à-dire le nombre de caractéristiques sélectionnées), ainsi que la liste de ces caractéristiques et leurs taux d'apprentissage correspondants sont fournis.

2.5.1 ACP

Pour les différentes sous-bases COREL utilisées, l'allure des valeurs propres, rangées par ordre décroissant en fonction des vecteurs propres correspondants, est illustré par la figure 2.4. On constate, dans les quatre cas, que les valeurs propres ne sont significatives que pour les premiers vecteurs propres. Effectivement, elles se dégradent rapidement quand le nombre de vecteurs propres croit. D'où, le principe de réduction de dimension de la méthode ACP. Seulement, les premiers vecteurs propres sont considérés. Ceci permet de retenir uniquement les composantes dites principales à partir de la distribution initiale des images d'apprentissage. Le nombre de composantes principales sélectionnées (CPs) dépend d'un taux d'inértie fixé par l'utilisateur. Bien que sa détermination reste empirique, la valeur de ce taux est d'une extrême importance. Le tableau 2.1 montre que selon une faible variation des taux d'inertie maintenus, le nombre des vecteurs propres reflétant le nombre de CPs retenues diminue sensiblement. D'où la forte réduction de dimensionnalité qui peut être apportée par la méthode ACP.

De même, la figure 2.5 et le tableau 2.2 présentent respectivement l'allure des valeurs propres, rangées par ordre décroissant en fonction des vecteurs propres correspondants, et la variation du nombre de composantes principales à travers les différentes sous-bases COLUMBIA utilisées.

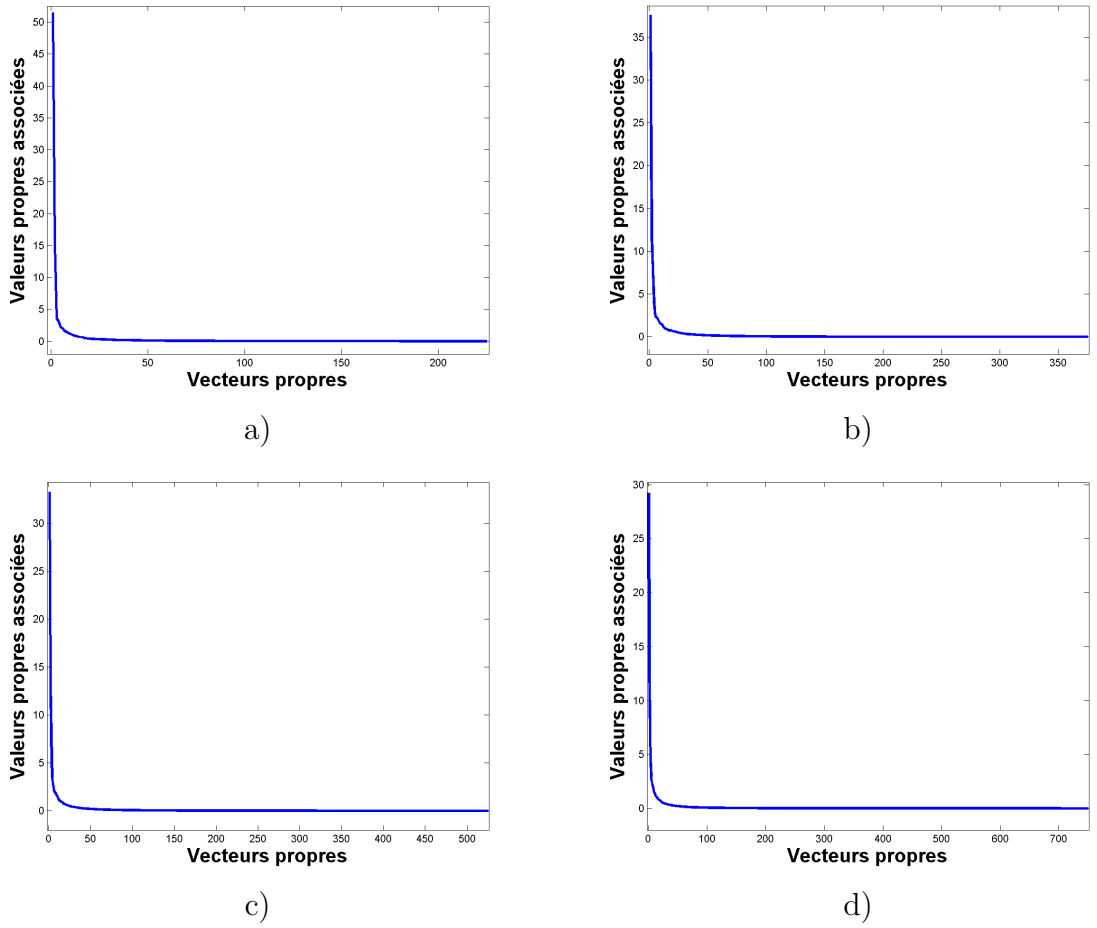


FIG. 2.4 – *Allure des valeurs propres rangées par ordre décroissant respectivement pour les sous-bases COREL avec : a)* 3 *classes, b)* 5 *classes, c)* 7 *classes et d)* 10 *classes.*

TAB. 2.1 – *Variation de la dimension des composantes principales sélectionnées avec la méthode de sélection ACP en fonction de différentes sous-bases d'images COREL.*

Sous-bases COREL	Dimensions des composantes principales sélectionnées selon différents taux d'inertie τ		
	100 %	95 %	90 %
3 classes	225	45	22
5 classes	375	76	41
7 classes	525	98	52
10 classes	750	116	60

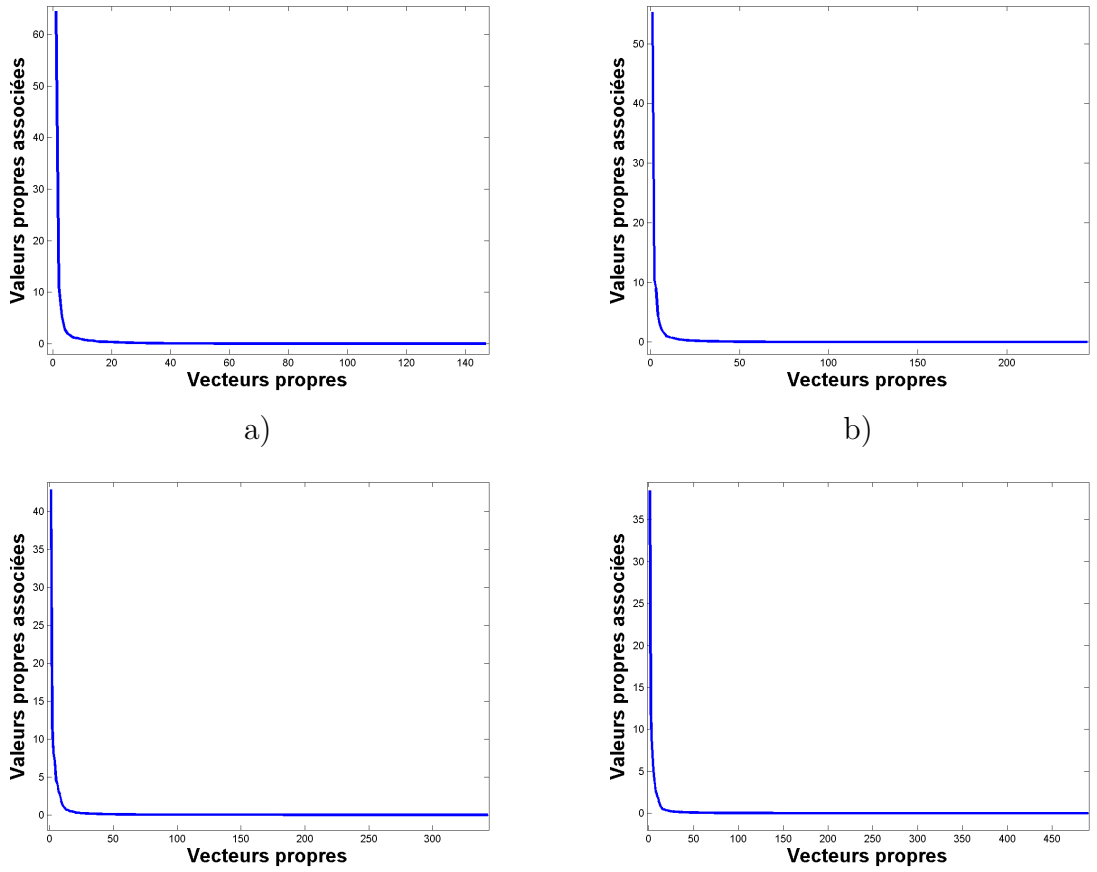


FIG. 2.5 – *Allure des valeurs propres rangées par ordre décroissant respectivement pour les sous-bases COLUMBIA avec : a)* 3 *classes, b)* 5 *classes, c)* 7 *classes et d)* 10 *classes.*

TAB. 2.2 – *Variation de la dimension des composantes principales sélectionnées avec la méthode de sélection ACP en fonction de différentes sous-bases d'images COLUMBIA.*

Sous-bases COLUMBIA	Dimensions des composantes principales sélectionnées selon différents taux d'inertie τ		
	100 %	95 %	90 %
3 classes	147	18	9
5 classes	245	23	11
7 classes	343	26	13
10 classes	490	31	14

Nous observons pour les différentes sous-bases d'images COREL et COLUMBIA employées que le nombre de CPs retenues dans chaque cas est très vulnérable par rapport au taux d'inertie choisi. Un choix qui reste flou et assez variable d'une distribution d'images à une autre. Par conséquent, les performances de sélection de la méthode ACP ne peuvent être évaluées qu'à partir des résultats de classification qui en résultent. L'évaluation des différentes méthodes de sélection étudiées à travers leurs performances de classification est fournie dans la section 2.5.4.

2.5.2 AG

Une sélection enveloppante des caractéristiques pertinentes est effectuée en se basant sur les AGs[1]. Les caractéristiques, ayant des taux de classification élevés en phase d'apprentissage, auront des valeurs de sélectivité meilleures, et donc une forte probabilité d'être sélectionnées dans les prochaines générations de l'AG. Les taux de classification examinés sont notés τ_i, $i = [1,2,...,n]$. Dans ce contexte, nous proposons trois fonctions de sélectivité:

- **Fonction de sélectivité 1:** La fonction de sélectivité 1 est conçue grâce aux taux de classification pondérés par des poids reflétant l'ordre de pertinence des différentes caractéristiques. On note ces poids par w_i, $i = [1,2,...,n]$. La formule est alors comme suit:

$$\textit{Sélectivité}_1 = w_i \times \tau_i, \;\; i = [1,2,...,n] \tag{2.11}$$

- **Fonction de sélectivité 2:** Une fonction de sélectivité objective et unique, qui favorise à la fois les meilleurs taux de classification et les faibles dimensions des caractéristiques, est conçue. La fonction de sélectivité 2 permet ainsi de résoudre un problème de sélection à critères multiples. Soit Dim_i, $i = [1,2,...,n]$ la dimension de la caractéristique i, la fonction de sélectivité est alors donnée par:

$$\textit{Sélectivité}_2 = \tau_i + \frac{1}{Log(Dim_i)}, \;\; i = [1,2,...,n] \tag{2.12}$$

- **Fonction de sélectivité 3:** La fonction de sélectivité 3 permet de calibrer le taux de classification de chaque caractéristique par rapport aux taux minimal et maximal, respectivement τ_{min} et τ_{max}, de toutes les caractéristiques employées. La fonction de sélectivité est calculée comme suit:

$$\textit{Sélectivité}_3 = \frac{\tau_i - \tau_{min}}{\tau_{max} - \tau_i + 1}, \;\; i = [1,2,...,n] \tag{2.13}$$

Dans le reste de ce manuscrit, les différentes méthodes enveloppantes de sélection des caractéristiques basées sur les AGs sont désignées comme suit: AG_1, AG_2 et AG_3, respectivement par rapport à l'utilisation des fonctions de sélectivité: $\textit{Sélectivité}_1$, $\textit{Sélectivité}_2$ et $\textit{Sélectivité}_3$.

Les tableaux 2.3, 2.4, 2.5 et 2.6 illustrent respectivement les résultats des trois méthodes de sélection des caractéristiques basées sur les trois AGs étudiés pour quatre sous-bases

1. La boite à outils d'Optimisation des Algorithmes Génétiques (GAOT) conçue pour MatLab 5 est utilisée. GAOT implémente des évolutions simulées dans l'environnement MatLab tout en permettant des représentations à la fois binaires et réel.

TAB. 2.3 – *Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 3 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, G-norm }	{ 0.9422, 1 }
AG_2	7	{ HMMD, YCrCb, Hist, A-Cor, FOS, SGLD, G-norm }	{ 0.9955, 0.9244, 1, 1, 0.8577, 0.9777, 1 }
AG_3	2	{ RVB, Hist }	{ 0.9422, 1 }

TAB. 2.4 – *Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 5 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, G-norm }	{ 0.9173, 1 }
AG_2	8	{ RVB, HSV, HMMD, Hist, A-Cor, FOS, SGLD, G-norm }	{ 0.9173, 0.9013, 0.9013, 1, 1, 0.7200, 0.8693, 1 }
AG_3	2	{ RVB, Hist }	{ 0.9173, 1 }

TAB. 2.5 – *Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 7 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, G-norm }	{ 0.8285, 1 }
AG_2	9	{ RVB, HSV, HMMD, YCrCb, A-Cor, FOS, SGLD, Daub, Hu }	{ 0.8285, 0.7485, 0.7409, 0.8000, 0.9942, 0.5847, 0.8095, 0.4609, 0.5600 }
AG_3	2	{ RVB, Hist }	{ 0.8285, 1 }

TAB. 2.6 – *Variation des paramètres des différentes sélections AGs pour la sous-base COREL avec 10 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, G-norm }	{ 0.7600, 1 }
AG_2	9	{ RVB, HSV, HMMD, YCrCb, A-Cor, FOS, SGLD, G-norm, Hu }	{ 0.760, 0.660, 0.6666, 0.7253, 0.9530, 0.4920, 0.6906, 1, 0.4493 }
AG_3	2	{ RVB, Hist }	{ 0.7600, 1 }

d'images COREL. D'après les caractéristiques sélectionnées dans chaque cas, il est clair que AG_1 et AG_3 ne sélectionnent pas des caractéristiques assez adaptées à la base d'images utilisée. En fait, ces deux sélections AGs convergent vers des optimums locaux pour toutes les sous-bases d'images COREL employées. Effectivement, AG_1 et AG_3 sélectionnent respectivement {RVB, G-norm} et {RVB, Hist} dans tous les cas. Par contre, la méthode de sélection AG_2 est beaucoup meilleure. Les caractéristiques sélectionnées dans chaque cas changent en fonction de la sous-base d'images considérée. De plus, selon leurs taux de classification en phase d'apprentissage, ces caractéristiques sont souvent pertinentes.

De même, les tableaux 2.7, 2.8, 2.9 et 2.10 illustrent respectivement les résultats des trois méthodes de sélection des caractéristiques AG_1, AG_2 et AG_3 pour les quatre sous-bases d'images COLUMBIA employées. Pour cette base d'images, on constate que les

TAB. 2.7 – *Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 3 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, HSV }	{ 1, 1 }
AG_2	2	{ RVB, HSV }	{ 1, 1 }
AG_3	1	{ RVB }	{ 1 }

TAB. 2.8 – *Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 5 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, YCrCb }	{ 1, 1 }
AG_2	2	{ RVB, YCrCb }	{ 1, 1 }
AG_3	1	{ RVB }	{ 1 }

TAB. 2.9 – *Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec 7 classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	2	{ RVB, YCrCb }	{ 1, 1 }
AG_2	2	{ RVB, YCrCb }	{ 1, 1 }
AG_3	1	{ RVB }	{ 1 }

trois méthodes de sélection AG ne sont pas efficaces. Effectivement, elles sélectionnent presque les même caractéristiques indépendamment de la sous-base d'images utilisée. En fait, AG_1, AG_2 et AG_3 convergent quasiment vers les mêmes optimums locaux, pour les trois premières sous-bases d'images COLUMBIA avec 3, 5 et 7 classes. Par exemple, AG_1

et AG_2 sélectionnent d'une façon similaire {RVB, HSV}, pour la sous-base COLUMBIA avec 3 classes, et {RVB, YCrCb}, pour les sous-bases COLUMBIA avec 5 et 7 classes. Par ailleurs, la méthode AG_3 sélectionne toujours la même caractéristique {RVB}, sans considérer la variation du contenu des différentes sous-bases d'images.

Cette dégradation des performances de sélection des différentes méthodes AGs, peut s'expliquer par le contenu des images traitées. Contrairement aux images de la base COREL, celles appartenant à différentes classes de la base COLUMBIA ont généralement des contenus assez proches. Ceci est dû essentiellement à l'arrière plan commun de toutes les images de la base (fond noir). De plus, les images d'une même classe contiennent toutes une représentation d'un même objet effectuées selon différentes prises de vue. Plus de détails concernant la base COLUMBIA sont fournis dans la section 1.3.1. Les caractéristiques extraites ont souvent des pouvoirs de discrimination comparables pour cette base. Par conséquent, la sélection des caractéristiques les plus pertinentes est généralement difficile. Dans le reste du manuscrit, les résultats de sélections des différentes méthodes AGs ne sont considérées que pour la base d'images COREL.

TAB. 2.10 – *Variation des paramètres des différentes sélections AGs pour la sous-base COLUMBIA avec* 10 *classes.*

	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
AG_1	1	{ RVB }	{ 1 }
AG_2	5	{ HSV, YCrCb, Hist, SGLD, Daub }	{ 0.9367, 0.9979, 1, 0.9020, 0.5755 }
AG_3	1	{ RVB }	{ 1 }

Par analogie avec le cas des sous-bases COREL, AG_2 semble être plus efficace que AG_1 et AG_3 pour la sélection des caractéristiques les plus pertinentes dans la sous-base COLUMBIA avec 10 classes.

2.5.3 SAC

Les résultats de sélection des caractéristiques les plus pertinentes, obtenus avec la méthode SAC pour différentes sous-bases d'images COREL et COLUMBIA, sont présentés dans les tableaux 2.11 et 2.12.

D'après le tableau 2.11, on observe que neuf caractéristiques sont sélectionnées pour les trois premières sous-bases COREL avec 3, 5 et 7 classes et seulement quatre pour la dernière sous-base avec 10 classes. Ceci résulte de l'accroissement de la complexité du contenu des images en fonction du nombre de classes d'images dans la base. Absolument, l'augmentation du nombre de classes implique plus de complexité dans le contenu de la base d'images, ce qui restreint de manière évidente le nombre de caractéristiques pertinentes qui permettent d'assurer des performances d'apprentissage considérables. Par ailleurs, les caractéristiques pertinentes sélectionnées dans chaque cas ne dépendent pas uniquement du nombre de classes dans la sous-base considérée, mais aussi du contenu même de cette base. Effectivement, bien que les neuf caractéristiques sélectionnées sont les mêmes pour les trois premières sous-bases (3, 5 et 7 classes), l'ordre de pertinence de ces caractéristiques diffèrent d'une sous-base à l'autre (tableau 2.11). Ce qui prouve

TAB. 2.11 – *Variation des paramètres SAC pour différentes sous-bases COREL.*

Sous-bases COREL	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
3 classes	9	{Hist, A-Cor, G-norm, GIST, HMMD, HSV, SGLD, RVB, YCrCb}	{1, 1, 1, 1, 0.9955, 0.9866, 0.9777, 0.9422, 0.9244}
5 classes	9	{Hist, A-Cor, G-norm, GIST, RVB, HSV, HMMD, YCrCb, SGLD}	{1, 1, 1, 1, 0.9173, 0.9013, 0.9013, 0.8933, 0.8693}
7 classes	9	{Hist, G-norm, GIST, A-Cor, RVB, SGLD, YCrCb, HSV, HMMD}	{1, 1, 1, 0.9942, 0.8285, 0.8095, 0.80, 0.7485, 0.7409}
10 classes	4	{Hist, G-norm, GIST, A-Cor}	{1, 1, 1, 0.9533}

l'adaptation du pouvoir discriminant de chaque caractéristique par rapport au contenu de la base d'image utilisée par la méthode SAC.

Le tableau 2.12 prouve également que l'ensemble des caractéristiques sélectionnées ainsi que leurs degrés de pertinence dépendent de la base d'images considérée. Par exemple, FOS est sélectionnée pour les quatre sous-bases COLUMBIA. De même, Hu est sélectionnée pour les sous-bases COLUMBIA avec 3 et 5 classes. Par ailleurs, ces deux caractéristiques ne sont sélectionnées pour aucune sous-base COREL. De plus, la caractéristique HMMD, sélectionnée pour la sous-base COREL avec 3 classes, ne figure pas dans la liste des caractéristiques sélectionnées pour la sous-base COLUMBIA avec le même nombre de classes. D'où l'aspect adaptatif de la méthode de sélection SAC.

TAB. 2.12 – *Variation des paramètres SAC pour différentes sous-bases COLUMBIA.*

Sous-bases COLUMBIA	Paramètres de sélection		
	Seuil de sélection	Caractéristiques sélectionnées	Taux d'apprentissage correspondants
3 classes	10	{RVB, HSV, YCrCb, Hist, A-Cor, FOS, SGLD, G-norm, GIST, Hu}	{1, 1, 1, 1, 1, 1, 1, 1, 0.9659}
5 classes	11	{RVB, YCrCb, Hist, A-Cor, FOS, G-norm, GIST, SGLD, Hu, HMMD, HSV}	{1, 1, 1, 1, 1, 1, 1, 0.9591, 0.9551, 0.9306, 0.9020}
7 classes	10	{RVB, YCrCb, Hist, A-Cor, G-norm, GIST, FOS, HSV, SGLD, HMMD	{1, 1, 1, 1, 1, 1, 0.9358, 0.9271, 0.9037, 0.8921}
10 classes	8	{A-Cor, G-norm, GIST, YCrCb, HSV, HMMD, FOS, SGLD}	{1, 1, 1, 0.9979, 0.9367, 0.9224, 0.9122, 0.9020}

Dans cette section, une évaluation des trois méthodes de sélection ACP, AG et SAC a été présentée. Nous constatons que l'ACP représente une forte technique de réduction de dimension. Cependant, elle est très sensible à la valeur du taux d'inertie appliqué.

La sélection AG, quand à elle, permet de sélectionner des ensembles de caractéristiques pertinentes, mais le choix de la fonction de sélectivité appropriée pour chaque base reste assez délicat. Contrairement aux méthodes ACP et AG, la méthode SAC est totalement automatique. Elle assure la sélection d'un sous-ensemble des caractéristiques les plus pertinentes indépendamment de toute sorte de réglage ou de calibration de paramètres. De plus, cette sélection est caractérisée principalement par son aspect adaptatif. En effet, la sélection des caractéristiques est parfaitement adaptée au contenu de la base d'images considérée. Dans la section suivante, les résultats de classification obtenus en se basant sur les différentes méthodes de sélection étudiées sont présentés.

2.5.4 Résultats de classification

Les tableaux 2.13 et 2.14 illustrent respectivement les résultats de classification SVM obtenus en se basant sur la sélection ACP pour les différentes sous-bases d'images COREL et COLUMBIA. Les meilleures et les pires performances dans ces tableaux sont affichées en gras.

TAB. 2.13 – *Résultats de classification SVM basée sur la sélection ACP pour différentes sous-bases d'images COREL.*

Sous-bases COREL		Taux d'inertie utilisés		
		100 %	95 %	90 %
3 classes	P_{moy}	**0.4656**	0.8155	**0.8301**
	R_{moy}	**0.4933**	0.5867	**0.6533**
5 classes	P_{moy}	**0.0400**	0.6481	**0.8258**
	R_{moy}	**0.2000**	0.3680	**0.4400**
7 classes	P_{moy}	**0.1644**	0.3083	**0.7405**
	R_{moy}	**0.1943**	0.243	**0.3657**
10 classes	P_{moy}	**0.2103**	0.3197	**0.7621**
	R_{moy}	**0.1320**	0.1680	**0.2640**

TAB. 2.14 – *Résultats de classification SVM basée sur la sélection ACP pour différentes sous-bases d'images COLUMBIA.*

Sous-bases COLUMBIA		Taux d'inertie utilisés		
		100 %	95 %	90 %
3 classes	P_{moy}	**0.8922**	**1.0000**	0.9861
	R_{moy}	**0.7733**	**0.9200**	0.9067
5 classes	P_{moy}	**0.8958**	0.9314	**0.9437**
	R_{moy}	**0.7200**	0.8240	**0.8480**
7 classes	P_{moy}	**0.7489**	0.8982	**0.9015**
	R_{moy}	**0.5086**	0.5943	**0.6286**
10 classes	P_{moy}	**0.8170**	0.9204	**0.9221**
	R_{moy}	**0.4720**	0.5600	**0.5960**

La classification basée sur la sélection ACP s'avère meilleure dans la base COLUMBIA. Ceci est dû principalement à la complexité du contenu de la base COREL (plus de

détails concernant les différentes bases d'images utilisées sont fournis dans le chapitre 1). Cependant, les résultats de classification des deux bases d'images montrent que l'efficacité de la sélection ACP varie sensiblement en fonction des taux d'inertie utilisés. Dans le cas des différentes sous-bases d'images COREL et COLUMBIA employées, le taux 90% semble être le meilleur, mais le taux 95% permet d'obtenir une classification plus pertinente pour la sous-base COLUMBIA avec 3 classes. Bien que des travaux de recherche existent (Thomas, 2000), la détermination du taux d'inertie optimal reste généralement empirique et aucune méthode théorique n'est encore approuvée.

Pour comparer la qualité de sélection des trois méthodes étudiées, la base COREL qui semble être beaucoup plus compétitive est retenue. En se basant sur les résultats de sélection AG[2] et les performances de classification ACP[3], la sélection AG_2 ainsi que la méthode ACP correspondante à un taux d'inertie de 90%, que l'on note dans la suite ACP (90%), sont considérées pour une comparaison avec la méthode SAC. Le tableau 2.15[4] présente les résultats de classification SVM basés ACP (90%), AG_2 et SAC. Les meilleures et les pires performances dans ce tableau sont affichées en gras.

TAB. 2.15 – *Résultats de classification SVM basée ACP (90%), AG_2 et SAC pour différentes sous-bases d'images COREL.*

Sous-bases COREL		Résultats de classification basée		
		ACP (90 %)	AG_2	SAC
3 classes	P_{moy}	**0.8301**	0.9540	**0.9879**
5 classes	P_{moy}	**0.8258**	0.9230	**0.9750**
7 classes	P_{moy}	**0.7405**	0.8554	**0.9123**
10 classes	P_{moy}	**0.7621**	0.8207	**0.8371**

Pour plus de lisibilité, les performances de classification présentées dans le tableau 2.15 sont illustrées à travers la figure 2.6 pour les quatre sous-bases d'images COREL.

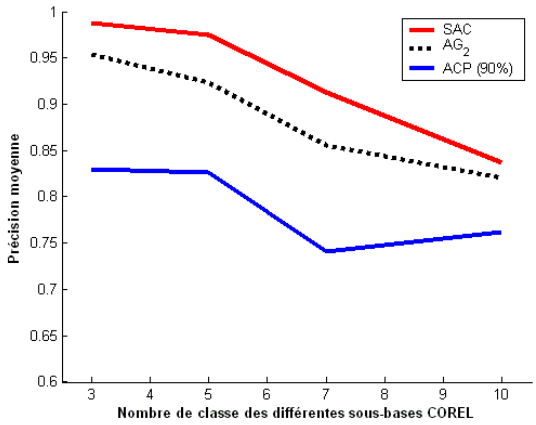


FIG. 2.6 – *Comparaison des précisions moyennes de classification respectivement basées sur les sélections ACP (90%), AG_2 et SAC pour différentes sous-bases d'images COREL.*

2. Les résultats de sélection AG sont fournis dans les tableaux 2.3, 2.4, 2.5 et 2.6.

3. Les performances de classification ACP sont fournis dans le tableau 2.13.

4. Les résultats de classification basés sur les sélections AG_1 et AG_3 sont également fournis dans le chapitre 4 (tableau 4.1).

D'une part, il est clair d'après la figure 2.6 que les deux méthodes de sélection AG_2 et SAC sont plus performantes que la méthode ACP. Évidemment, en tant que techniques enveloppantes, AG_2 et SAC permettent d'assurer une sélection des caractéristiques beaucoup mieux adaptée au classifieur SVM employé en phase de généralisation. D'autre part, on observe que les caractéristiques sélectionnées par la méthode SAC mènent toujours à des résultats de classification meilleurs par rapport aux résultats de celles sélectionnées par la méthode AG_2. Ce qui prouve davantage la robustesse et l'efficacité de la méthode de sélection proposée.

2.6 Conclusion

Nous avons présenté dans ce chapitre une nouvelle méthode de Sélection Adaptative des Caractéristiques les plus pertinentes, intitulée SAC. Cette méthode est essentiellement destinée pour les systèmes de reconnaissance d'images hétérogènes par le contenu. Suite à un apprentissage multiples des classifieurs SVMs, la méthode de sélection proposée détermine un sous-ensemble des caractéristiques les plus pertinentes. Les caractéristiques sont sélectionnées en se basant sur une discrimination Fisher de leurs taux d'apprentissage individuels. Cette sélection est parfaitement adaptée au contenu de la base d'images utilisée et aux caractéristiques employées. Une large évaluation de la méthode SAC, ainsi que deux autres méthodes de sélections assez populaires a été effectuée à travers deux bases d'images de référence. Cette évaluation confirme l'aspect adaptatif de notre proposition. De plus, il a été bien vérifié que la méthode SAC ne nécessite aucune intervention de la part de l'utilisateur pour un réglage de paramètres. Effectivement, contrairement aux méthodes ACP et AG, les caractéristiques les plus pertinentes sont sélectionnées par la méthode SAC d'une manière automatique.

Une comparaison des performances de classification basée sur les différentes méthodes de sélection, prouve la pertinence et la supériorité de la sélection SAC.

Chapitre 3

Machines à Vecteurs Supports à noyaux multiples

L'approche d'apprentissage par noyaux multiples, appelée souvent par l'acronyme anglais MKL (*Multiple Kernel Learning*) est présentée dans ce chapitre. Récemment proposée, MKL est destinée à améliorer les performances des méthodes de classification à noyaux, en particulier les machines à vecteurs supports (SVM). La pertinence des résultats de classification MKL dépend essentiellement de la qualité des poids affectés à chaque noyau. Après avoir présenté les différentes méthodes de pondération utilisées dans la littérature, une nouvelle pondération destinée pour l'apprentissage SVM à noyaux multiples est proposée.

3.1 Introduction

Inspirée de la théorie statistique de l'apprentissage, SVM fut introduite par Vladimir Vapnik comme méthode de classification binaire par apprentissage supervisé (Vapnik, 1999). Grâce à ses performances, cette méthode de classification est devenue en une courte période un outil standard dans l'état de l'art de plusieurs problèmes de reconnaissance, en particulier celui des images hétérogènes par le contenu. À l'origine, SVM repose sur l'existence d'un classifieur linéaire qui sépare deux classes dans un espace approprié. Son extension aux problèmes multi-classes est rapidement mise en évidence (Weston and Watkins, 1999), (Vapnik, 1999), (Friedman, 1997), (Xu and Chan, 2003), (Platt et al., 2000). De plus, pour permettre une séparation optimale dans les cas non-linéairement séparables, SVM utilise des fonctions noyaux (kernels). Les développements récents des SVMs, ont prouvé que l'utilisation des noyaux multiples génére plus de flexibilité et permet d'améliorer l'interprétabilité de ces méthodes (Lanckriet et al., 2004a). Le noyau est ainsi défini comme une combinaison linéaire convexe de plusieurs noyaux. Par conséquent, le problème de choix de noyau est transféré vers le calcul des poids optimaux de différents noyaux. En général, les algorithmes MKL assurent l'apprentissage des hyper-paramètres SVM et des poids de noyaux dans un même problème d'optimisation.

Plusieurs méthodes ont appliqué l'approche MKL dans le contexte de reconnaissance d'images (Suard et al., 2007), (Varma and Ray, 2007), (Gehler and Nowozin, 2009). Ces méthodes sont basées souvent sur un ensemble d'optimisation pour le calcul des poids de noyaux (Rakotomamonjy et al., 2007). Par ailleurs, elles souffrent généralement d'un temps de calcul très important. Nous proposons ainsi une nouvelle méthode de pondération, en se basant sur le degré de pertinence de chaque noyaux. La nouvelle pondération permet donc de mieux considérer les noyaux les plus pertinents en un temps d'exécution beaucoup plus rapide que les méthode classiques. Cette méthode est destinée principalement pour l'apprentissage des SVMs à noyaux multiples.

Dans ce chapitre, le principe de fonctionnement général des SVMs est rappelé dans la section 3.2. L'approche d'apprentissage par noyaux multiples est ensuite présentée dans la section 3.3. Particulièrement, les méthodes de pondération de noyaux utilisées dans la littérature sont exposées dans la même section. Enfin, la nouvelle méthode de pondération est proposée dans la section 3.4. L'évaluation de cette méthode ainsi qu'une comparaison avec un algorithme MKL récent sont discutées dans la section 3.5.

3.2 Machines à Vecteurs Supports

3.2.1 Introduction

Depuis les années quatre vingt dix, les SVMs n'ont cessé de susciter l'intérêt de plusieurs communautés de chercheurs de différents domaines d'expertise. La performance de ces classifieurs dépasse celle des méthodes supervisées traditionnelles dans plusieurs applications (Bi et al., 2003), (Zhu et al., 2004), (Ritendra et al., 2006), (Chen et al., 2006). Dans cette section, le principe de fonctionnement général des SVMs est présenté. La linéarité et la non-linéarité ainsi que l'extension aux problèmes multi-classes de la méthode SVM sont également exposées.

3.2.2 Notions de base

Pour deux classes d'images données, le but des SVMs est de trouver un classifieur linéaire qui sépare les images tout en maximisant la distance entre ces deux classes. Il s'agit d'un hyper-plan de séparation. Les images les plus proches de cet hyper-plan, c'est à dire les plus difficiles à classifier, sont appelées vecteurs supports. Les SVMs assurent une modélisation discriminante qui s'appuie sur la détermination des différents vecteurs supports dans les deux classes. Il est évident qu'il existe une multitude d'hyper-plans valides mais la propriété remarquable des SVMs est que cet hyper-plan doit être optimal. Formellement, cela revient à chercher un hyper-plan dont la distance minimale aux différents vecteurs supports est maximale. Ainsi, les SVMs sont souvent appelés des *Séparateurs à Vaste Marge*. Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classifie de nouvelles images en phase de généralisation. En général, la classification d'une image requête est donnée par sa position par rapport à l'hyper-plan optimal trouvé en phase d'apprentissage.

Il existe deux cas de modèles SVMs: les cas linéairement séparables et les cas non-linéairement séparables. Les premiers sont les plus simples, car ils permettent de trouver facilement le classifieur linéaire (figure 3.1). Dans les cas non-linéairement séparables, le classifieur de marge maximale ne peut pas être utilisé car ceci ne fonctionne que dans les cas linéairement séparables. Un système de marges élastiques est alors mis en place pour tolérer des erreurs de classification et donc tolérer des vecteurs supports à l'intérieur de la marge (figure 3.2). Ce compromis entre erreurs et taille de la marge est contrôlé par un hyper-paramètre d'apprentissage, que l'on note par C. Dans la plupart des problèmes de classification réels, il n'y a pas de séparation linéaire possible entre les données. Particulièrement, pour la reconnaissance d'images par le contenu dans les bases hétérogènes, les images appartenant aux différentes classes sont souvent non-linéairement séparables. Nous rappelons ici les deux cas de séparation linéaire et non-linéaire.

3.2.3 Cas linéairement séparable

Un problème de classification binaire est linéairement séparable, si les images d'apprentissage appartenant aux deux classes peuvent êtres complètement séparables. La classification par SVM est ainsi dite linéaire et il s'agit, dans ce cas, d'une comparaison entre les images à travers un simple produit scalaire. Soient N_{Im} le nombre d'images d'apprentissage, $X = [x_1,x_2,...,x_i,...,x_{N_{Im}}]$ le nuage d'observations dans l'espace d'attributs correspondant et $Y = [y_1,y_2,...,y_i,...,y_{N_{Im}}]$ l'ensemble de leurs étiquettes. Notons que toutes les étiquettes y_i appartiennent à l'ensemble {-1,1} (classification binaire). À partir de l'examination des couples d'entrée-sortie $\{(x_i, y_i);\ \ i = 1,2,...,N_{Im}\}$, le but est de trouver une fonction $f : X \rightarrow Y$, telle que la probabilité $P(f(X) \neq Y)$ soit minimale. Dans le cas linéaire, la fonction de décision SVM est la suivante:

$$f(x) = \sum_{i=1}^{N_{Im}} \alpha_i y_i x.x_i + b \tag{3.1}$$

avec, x est un vecteur support et les x_i sont les observations des différentes images d'apprentissage. Les multiplicateurs de Lagrange α_i et le biais b sont des hyper-paramètres d'apprentissage déterminés par la résolution du Lagragien du problème d'optimisation initial. La figure 3.1 illustre un exemple de classification binaire linéairement séparable.

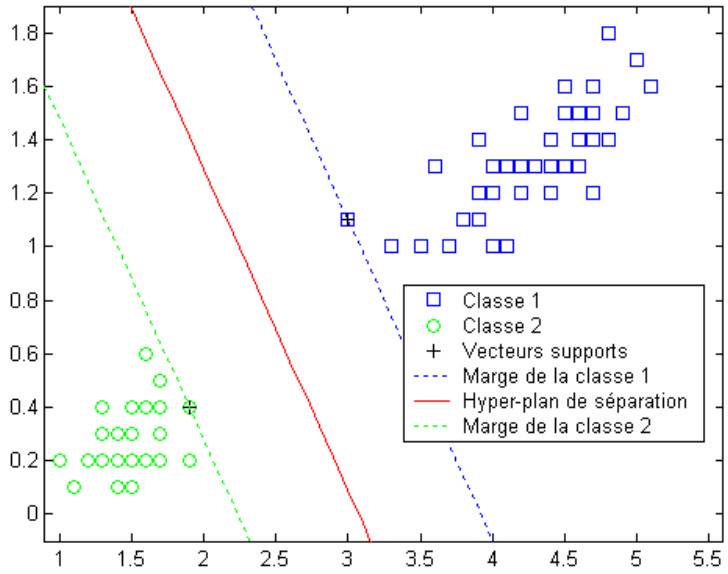


FIG. 3.1 – *Exemple de classification SVM linéairement séparable.*

L'équation (3.1) permet de déterminer un hyper-plan de séparation optimal dans le cas linéaire. Cependant, cette équation ne convient plus lorsque les images d'apprentissage sont non-linéairement séparables. En conséquence, une classification SVM non-linéaire, qui se base sur une fonction noyau, est nécessaire.

3.2.4 Cas non-linéairement séparable

On considère que les images d'apprentissage sont initialement représentées dans un espace d'attributs $\mathcal{F} = \mathbb{R}^d$. Pour résoudre le problème de non-séparabilité des observations correspondantes dans cet espace, les SVMs consistent à effectuer une transformation non-linéaire Φ de $\mathcal{F}$ dans un espace de Hilbert $\mathcal{H}$ de dimension supérieure, éventuellement infinie ($Card(\mathcal{H}) > d$). Cette transformation non-linéaire des vecteurs caractéristiques x_i, $i = [1,2,...,N_{Im}]$ permet une séparation linéaire des images dans un nouvel espace appelé, généralement, espace de re-description. Plus la dimension de l'espace de re-description est importante, plus la probabilité de pouvoir trouver un hyper-plan séparateur entre les images est élevée. On a donc une transformation d'un problème de séparation non-linéaire dans l'espace de représentation initial en un problème de séparation linéaire dans un espace de re-description de plus grande dimension. En pratique, cette transformation est réalisée à l'aide d'une fonction noyau, souvent appelée K et décrite comme suit:

$$K(x,x_i) = \langle \Phi(x), \Phi(x_i) \rangle_{\mathcal{H}}, \qquad i = [1,2,...,N_{Im}]. \tag{3.2}$$

$$\text{avec} \begin{cases} \Phi : \mathcal{F} \to \mathcal{H} ; & Card(\mathcal{H}) > Card(\mathcal{F}) \\ \quad x \mapsto \Phi(x) \end{cases}$$

La fonction de décision SVM devient donc:

$$f(x) = \sum_{i=1}^{N_{Im}} \alpha_i y_i K(x,x_i) + b \tag{3.3}$$

Les fonctions noyaux peuvent avoir plusieurs formes. Quelques familles de fonctions noyaux paramétrables sont connues et il revient à l'utilisateur de les tester pour déterminer celle qui convient le mieux pour son application. Pour un vecteur support x et une observation $x^{'}$, les noyaux combinant x et $x^{'}$ les plus utilisés sont les suivant:

$$\text{Polynômial:} \qquad K(x,x^{'}) = (x.x^{'} + ct)^{u} \tag{3.4}$$

$$\text{Sigmoîdal:} \qquad K(x,x^{'}) = tanh(x.x^{'} + \theta) \tag{3.5}$$

$$\text{Laplacien:} \qquad K(x,x^{'}) = e^{-\frac{\|x-x^{'}\|}{\sigma}} \tag{3.6}$$

$$\text{Gaussien:} \qquad K(x,x^{'}) = e^{-\frac{\|x-x^{'}\|^2}{2\sigma^2}} \tag{3.7}$$

avec, ct, u, θ et σ sont des paramètres de noyaux. Un exemple de classification non-linéairement séparable est illustré par la figure 3.2. Dans ce cas, l'utilisation d'une fonction noyau s'avère nécessaire pour se ramener, en passant par une dimension plus grande, au cas linéairement séparable (figure 3.1).

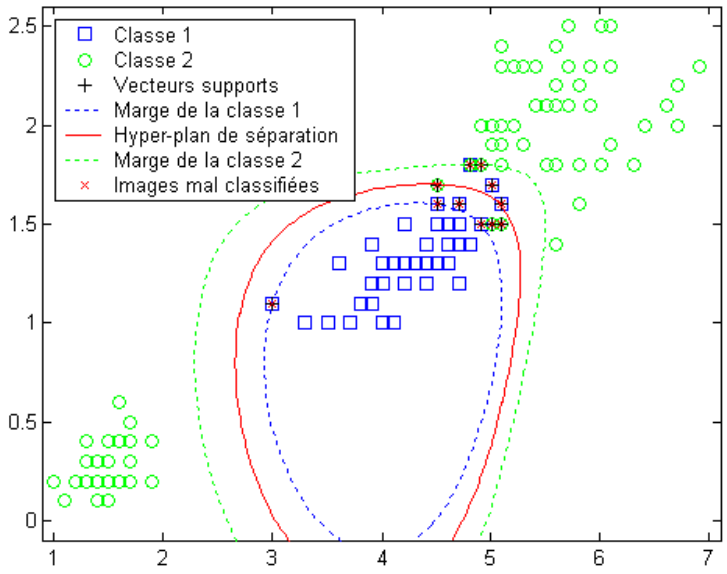


FIG. 3.2 – *Exemple de classification SVM non-linéairement séparable.*

3.2.5 Extension aux problèmes multi-classes

Bien qu'ils soient conçus comme des classifieurs binaires, les SVMs peuvent s'adresser également aux problèmes de classification multi-classes. Un exemple de classification SVM à trois classes ($N_C = 3$) est illustré par la figure 3.3.

Des approches permettant de considérer la classification multi-classes comme un seul problème d'optimisation existent en littérature (Weston and Watkins, 1999). L'inconvénient de ces méthodes est qu'elles sont beaucoup plus gourmandes en terme de calcul que la résolution de plusieurs problèmes de classification binaires. Dans la littérature, il est

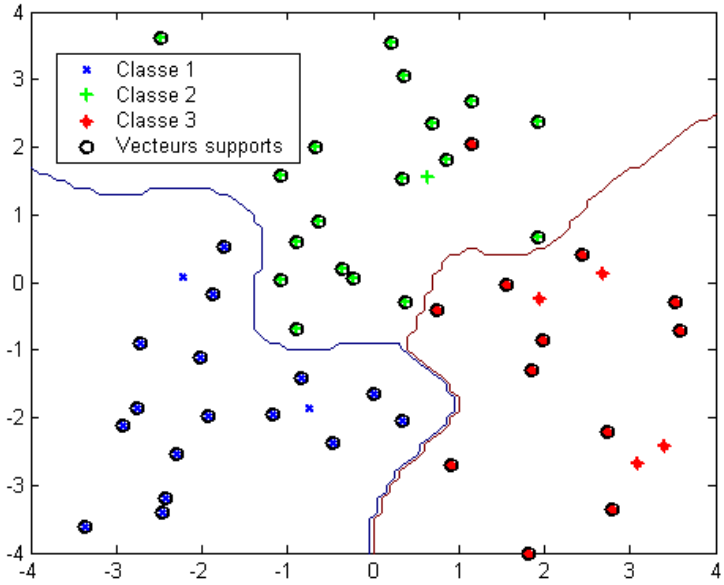


FIG. 3.3 – *Exemple de classification SVM multi-classes ($N_C = 3$).*

d'usage de distinguer une variété de techniques de décomposition des problèmes multi-classes en un ensemble de classification binaire. Ci après, nous rappelons les techniques les plus utilisées pour l'extension des SVMs aux problèmes multi-classes.

3.2.5.1 Algorithme SVM: Un contre tous

Un problème de classification de N_C classes ($N_C > 2$) est décomposé selon la technique *un contre tous* en N_C classifications binaires. Pour cela, N_C classifieurs SVMs binaires sont construits (Vapnik, 1999). L'apprentissage du $i^{\text{ème}}$ classifieur SVM est réalisé de façon que les images de la $i^{\text{ème}}$ classe sont considérées positives, alors que toutes les images des autres classes sont considérées négatives. En phase de généralisation, une image requête est présentée à l'ensemble des N_C classifieurs SVM. Elle est étiquetée ainsi en fonction du rendement maximum parmi les N_C classifieurs. L'inconvénient de cette méthode réside dans sa complexité d'apprentissage, vu que le nombre d'échantillons d'apprentissage devient très important. En effet, chacun des N_C classifieurs SVMs est entrainé en utilisant tous les échantillons disponibles.

3.2.5.2 Algorithme SVM: Un contre un

La technique *un contre un*, quand à elle, construit $N_C(N_C - 1)/2$ classifieurs SVMs binaires, en utilisant toutes les paires de combinaisons binaires des N_C classes. L'apprentissage de chaque classifieur SVM est réalisé en considérant les images de la première classe comme des exemples positifs et les images de la seconde classe comme des exemples négatifs. Pour combiner les différents classifieurs SVM, l'algorithme de vote majoritaire pour *max-wins voting* est adopté. Cet algorithme permet de trouver la classe résultante en choisissant la classe votée par la majorité des classifieurs (Friedman, 1997). Le nombre d'images utilisées pour l'apprentissage de chaque classifieur SVM est très réduit, car seuls les images appartenant à deux classes parmi les N_C classes sont prises en consideration. D'où, un temps d'apprentissage plus court. L'inconvénient de cette méthode réside dans le fait que chaque image requête doit être présentée à un grand nombre de classifieurs

$(N_C(N_C-1)/2)$. Il en résulte un ralentissement de la phase de généralisation, en particulier lorsque le nombre de classes est considérable (Xu and Chan, 2003).

3.2.5.3 Algorithme DAG-SVM

Similairement à la technique *un contre un*, l'algorithme *DAG-SVM* consiste en un apprentissage de $N_C(N_C-1)/2$ classifieurs SVM binaires. Comme son nom l'indique, cette méthode s'appuie, dans la phase de généralisation, sur un graphe de décision orienté acyclique (DAG pour *Directed Acyclic Graph*) (Platt et al., 2000). *DAG-SVM* crée un modèle pour chaque paire de classes. Soit le modèle M_{od_1} permettant de séparer les deux classes C_1 et C_2. La particularité de cet algorithme est qu'il considère que M_{od_1} ne vote pas réellement pour la classe C_1 s'il classifie une image requête dans C_1. *DAG-SVM* suppose plutôt, dans ce cas, que M_{od_1} vote contre la classe C_2. Ceci est évident, puisque cette requête doit se trouver de l'autre côté de l'hyper-plan séparateur que la plupart des images appartenant à la classe C_2. Par conséquent, l'algorithme ne tient plus compte de tous les modèles impliquant la classe C_2. Ainsi, après chaque classification avec un des modèles binaires, une des classes candidates est rejetée. De ce fait, une seule classe est retenue, suite à N_C-1 étapes uniquement. L'image requête considérée est ainsi attribuée à cette classe. Ceci permet d'obtenir des taux de reconnaissance similaire à ceux de la technique *un contre un*, mais le temps de réponse est beaucoup plus réduit. Dans notre travail de thèse, les SVMs [1] multi-classes utilisés sont basés sur le principe de l'algorithme *DAG-SVM*.

3.2.6 Conclusion

Dans cette section, nous avons présenté le principe de fonctionnement des SVMs. C'est une méthode de classification très performante, qui s'adresse à la fois aux cas linéaire et non-linéaire ainsi qu'aux problèmes de classification binaires et multi-classes. Cependant, l'inconvénient majeur des SVMs est le choix du noyau approprié et la détermination empirique de ses paramètres en fonction de la distribution des images d'apprentissage (Duan et al., 2003). En conséquence, des modèles plus flexibles sont fortement recherchés. Dans ce cadre, l'approche d'apprentissage par noyaux multiples est initialement proposée en 2004 (Lanckriet et al., 2004a). Cette proposition est basée principalement sur l'utilisation de noyaux multiples dans le même processus d'apprentissage ce qui s'avère une solution très intéressante pour palier aux limites des fonctions de décision à noyau unique.

3.3 Approche d'apprentissage par noyaux multiples

3.3.1 Principe

Pour améliorer les performances de la fonction de décision SVM et mettre en valeur son interprétabilité, les auteurs dans (Lanckriet et al., 2004a) ont proposé l'approche d'apprentissage par noyaux multiples désignée généralement par l'acronyme anglais MKL (*Multiple kernel Learning*). Dans cette approche, le noyau $K(x,x^{'})$ est défini comme une

1. La boite à outils MatLab de *Object-Oriented SVM* utilisée dans notre travail de thèse est disponible en ligne à l'adresse suivante: http://theoval.sys.uea.ac.uk/∼gcc/svm/toolbox.

combinaison de plusieurs noyaux de base. La formulation correspondante de cette proposition est la suivante:

$$K(x,x^{'}) = \sum_{m=1}^{M} \beta_m k_m(x,x^{'}) \tag{3.8}$$

avec $\beta_m \geq 0$, $\forall m \in \{1,2,...,M\}$ et $\sum_{m=1}^{M} \beta_m = 1$, où M est le nombre de noyaux de base utilisés, m est l'indice de noyaux et β_m est le poids du noyau k_m.

Chaque noyaux k_m, $m = [1,2,...,M]$ peut être calculé sur différents sous-ensemble de la base d'apprentissage, par différentes caractéristiques ou selon différentes formulations et différents paramètres (Lanckriet et al., 2004a). La fonction de décision SVM à noyaux multiples est donc la suivante:

$$f(x) = \sum_{i=1}^{N_{Im}} \alpha_i y_i \left(\sum_{m=1}^{M} \beta_m k_m(x,x_i) \right) + b \tag{3.9}$$

Initialement, la performance d'apprentissage SVM dépendait fortement de la représentation des images dans l'espace d'attributs correspondant. Ceci est implicitement considérée par le choix du noyau employé. L'utilisation de plusieurs noyaux dans le même processus d'apprentissage améliore les performances de classification et permet de surmonter les limites des méthodes de classification basées sur un noyau unique. Cependant, les noyaux utilisés n'ont pas un même degré de pertinence. Ensuite, le problème de représentation d'images à travers le choix des noyaux est transféré vers le calcul des poids optimums des différents noyaux utilisés dans cette approche d'apprentissage par noyaux multiples. Plusieurs méthodes de pondération ont été proposées dans ce contexte (Bach et al., 2004), (Sonnenburg et al., 2006), (Rakotomamonjy et al., 2008). Les méthodes de pondération de noyaux les plus connues sont discutées et présentées dans la section suivante.

3.3.2 Techniques de pondération de noyaux

Une façon simple pour respecter les critères de pondération MKL, présentés à la suite de l'équation (3.8), est d'utiliser un poids fixe et commun pour tous les M noyaux:

$$\beta_m = \frac{1}{M}, \forall m \in \{1,2,...,M\} \tag{3.10}$$

Vu qu'elle ne considère pas la variation de pertinence des différents noyaux employés, cette pondération altère généralement les résultats de classification finaux et elle est souvent non efficace. De ce fait, les poids des différents noyaux doivent être bien adaptés à l'efficacité de ces noyaux à travers le problème de classification considéré.

Des travaux récents sont concentrés sur des méthodes plus efficaces. En général, les poids des noyaux sont tirés dans le même problème d'optimisation que les hyperparamètres d'apprentissage. α_i, b et les valeurs des coefficients β_m sont ainsi obtenus en résolvant le problème d'optimisation dual suivant:

$$\min_{\alpha,b,\beta,\xi} \frac{1}{2} \left(\sum_{m=1}^{M} \beta_m \sum_{i=1}^{N_{Im}} \alpha_i y_i k_m(x,x_i) + b \right)^2 + C \sum_{i=1}^{N_{Im}} \xi_i \tag{3.11}$$

$$\text{sachant que } \begin{cases} y_i f(x_i) \geq 1 - \xi_i, \, \forall i \in \{1,2,...,N_{Im}\} \\ \xi_i \geq 0, \, \forall i \in \{1,2,...,N_{Im}\} \\ \beta_m \geq 0, \, \forall m \in \{1,2,...,M\} \\ \text{et } \sum_{m=1}^{M} \beta_m = 1 \end{cases}$$

où les différentes variables ξ_i sont appelées des variables molles (*slack variables* en anglais), elles sont introduites pour considérer les erreurs de classification qui peuvent être associées à chaque point d'apprentissage. L'hyper-paramètre d'optimisation C, appelé aussi terme de régularisation de marge, permet de pondérer ces erreurs de classification pour optimiser la fonction séparatrice finale. Cette méthode permet de déterminer itérativement les hyper-paramètres d'apprentissage ainsi que les coefficients de pondération MKL en même temps. Initialement, les auteurs dans (Lanckriet et al., 2004b) ont introduit l'approche des noyaux multiples pour les classifications binaires. L'inconvénient majeur de cette proposition est qu'elle est basée sur un problème de programmation quadratiquement contraint. Éventuellement, elle devient rapidement intraitable dès que le nombre d'images d'apprentissage ou le nombre de noyaux est grand. De plus, le problème de minimisation dans l'approche de (Lanckriet et al., 2004b) est réellement un problème dual dans lequel les coefficients de pondération sont régularisés selon les deux normes mixtes (L_2, L_1) à la place de la norme quadratique L_2. En conséquence, la formulation MKL de Lanckriet et al. n'est pas un problème de minimisation régulier. Bach et al. ont proposé une version plus régulière de ce problème pour traiter des problèmes à moyenne échelle (Bach et al., 2004). Afin de pouvoir supporter les problèmes à grande échelle, les auteurs dans (Sonnenburg et al., 2005), (Sonnenburg et al., 2006) ont adressé l'approche MKL en résolvant un problème de programmation linéaire semi-infinie, connu par l'acronyme anglais SILP (*Semi-Infinite Linear Program*):

$$\begin{cases} \max\limits_{\delta,\beta} \delta \\ \text{tel que } \sum\limits_{m=1}^{M} \beta_m = 1, \, \beta_m \geq 0, \, \forall m \in \{1,2,...,M\} \\ \text{et } \sum\limits_{m=1}^{M} \beta_m S_m(\boldsymbol{\alpha}) \geq \delta \\ \text{avec } 0 \leq \alpha_i \leq C, \text{ et } \sum\limits_{i=1}^{N} \alpha_i y_i = 0, \, \forall i \in \{1,2,...,N_{Im}\} \end{cases} \tag{3.12}$$

Sachant que la formulation de la fonction différentiable $S_m(\boldsymbol{\alpha})$ est la suivante:

$$S_m(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{i=1}^{N_{Im}} \sum_{j=1}^{N_{Im}} \alpha_i \alpha_j y_i y_j k_m(x_i,x_j) - \sum_{i=1}^{N_{Im}} \alpha_i \tag{3.13}$$

Sonnenburg et al. utilisent l'algorithme appelé *Column Generation Technique* qui consiste à chercher, pour un sous-ensemble de contraintes, les valeurs optimales de δ et β_m, $\forall m \in \{1,2,...,M\}$, puis à déterminer si $\boldsymbol{\alpha}$ satisfait la contrainte: $\sum_{m=1}^{M} \beta_m S_m(\boldsymbol{\alpha}) \geq \delta$. Dans ce cas, la solution peut être optimale, sinon des contraintes sont rajoutées à l'ensemble de départ et ce processus est itéré jusqu'à l'obtention de la convergence des différentes valeurs des poids β_m, $m \in [1,2,...,M]$. L'avantage de cette dernière formulation (équation (3.12)) est que l'approche MKL est considérée par la résolution itérative

d'un problème SVM classique à noyau unique. Cependant, l'algorithme itératif de Sonnenburg et al. nécessite un grand nombre d'itérations avant de pouvoir converger vers des solutions raisonnables. Effectivement, les problèmes SILP peuvent avoir un nombre infini de contraintes (Glashoff and Gustafson, 1983).

Récemment, Rakotomamonjy et al. ont proposé une optimisation de l'algorithme de Sonnenburg et al. (Rakotomamonjy et al., 2008). Les deux algorithmes minimisent la même fonction objective, mais ils diffèrent au niveau de l'efficacité de calcul. En fait, le nouvel algorithme, appelé *SimpleMKL*, optimise les coefficients de pondération par une méthode de descente du gradient. Ce qui lui permet de surmonter les limites des problèmes SILP. Cet algorithme est utilisé pour la comparaison des résultats obtenus dans la partie expérimentale de ce chapitre.

Dans la section suivante, nous proposons une méthode originale de pondération des noyaux. La méthode proposée met en évidence les noyaux les plus pertinents dans la combinaison linéaire de l'approche MKL.

3.4 Pondération de noyaux basée sur les taux d'apprentissage SVM

3.4.1 Introduction

Les techniques de pondération proposées pour l'approche MKL sont souvent efficaces et conduisent généralement à des résultats de classification satisfaisants. Toutefois, des méthodes de pondération plus rapides, permettant en particulier d'améliorer les résultats de classification dans les problèmes de grande échelle, sont toujours recherchées. En outre, la complexité du contenu des bases d'images hétérogènes rend la détermination des poids optimaux très difficile (Cao et al., 2009). Dans ce cadre, une nouvelle pondération de noyaux est décrite dans cette section. Cette pondération est destinée aux apprentissages SVM à noyaux multiples, que l'on appelle MK_SVM, acronyme anglais de *Multiple Kernel based SVM*.

3.4.2 Principe de MK_SVM

Afin d'améliorer l'efficacité de pondération des noyaux, nous exploitons des méthodes similaires dans d'autres applications spécifiques, en particulier les noyaux dits *String Kernels*. Ce type de noyaux se révèle significativement efficace pour le regroupement des données de type chaîne de caractères, souvent employées en biologie. Plusieurs méthodes sont proposées en littérature, nous citons à titre d'exemple le *locality-improved kernel* (Zien et al., 2000), le *string subsequence kernel* (Lodhi et al., 2001), le *spectrum kernel* (Leslie et al., 2002) et le *Weighted-Degree kernel* (WD) (Rätsch et al., 2005). D'autres méthodes sont citées dans le travail de (Kilho, 2007). Dans notre travail, nous nous intéressons particulièrement à la méthode WD (Rätsch et al., 2005). Cette méthode calcule efficacement les similitudes entre les séquences tout en considérant la position de la longueur des chaînes de caractères. Les auteurs dans (Rätsch et al., 2005) présument que les poids doivent satisfaire le critère suivant: $\beta_m > \beta_{m+1}$, $\forall m \in \{1,2,...,M-1\}$. Cette condition est basée sur le fait que la mise en correspondance des sous-chaînes est pondérée avec un score dépendant de la longueur de chacune de ces chaînes et que les chaînes de caractères à pondérer sont déjà organisées selon leur longueur.

Par analogie avec la méthode WD, les noyaux utilisés dans notre contexte, comme les chaînes de caractères, peuvent être pondérés linéairement sans l'utilisation d'aucune méthode d'optimisation. Effectivement, en s'inspirant du travail de (Rätsch et al., 2005), les poids des noyaux sont calculés en considérant la pertinence de chaque noyaux par rapport à l'ensemble du reste des noyaux à pondérer. La méthode MK_SVM adapte ainsi la pondération de noyaux au principe de la méthode WD. Ceci nécessite l'évaluation du degré de pertinence de chaque noyau selon une certaine mesure de qualité. En effet, les noyaux sont pondérés avec des scores dépendant de leurs pertinences par rapport à la base d'images considérée.

3.4.3 Apprentissage multiple

Rappelons l'apprentissage multiple de la méthode de sélection SAC qui a été présenté dans le chapitre 2 (section 2.4.2). Cette approche évalue séparément les différentes caractéristiques. Ce qui permet de déterminer le pouvoir discriminant de chacune d'entre elles. D'une manière similaire, nous présentons dans cette section l'utilisation d'un apprentissage multiple pour la mesure du degré de pertinence de chaque noyau.

Soit $\mathcal{K} = [k_1,k_2,...,k_M]$ l'ensemble des M noyaux à utiliser. Nous commençons par un apprentissage multiple de la base $\mathcal{B}_{App}$, selon les différents noyaux dans $\mathcal{K}$. En considérant les images correctement classifiées conformément à chaque noyau k_m, $m \in [1,2,...,M]$. nous calculons les taux d'apprentissage correspondants:

$$TR_{k_m} = \frac{N_{Corr}(k_m)}{N_{Im}}, \forall m \in \{1,2,...,M\} \tag{3.14}$$

où $N_{Corr}(k_m)$ est le nombre d'images correctement classifiées avec le noyau k_m. Soit $\mathcal{TR} = [TR_{k_1},TR_{k_2},...,TR_{k_M}]$ l'ensemble des taux d'apprentissage obtenus, avec TR_{k_m} le taux correspondant au noyau k_m, $\forall m \in \{1,2,...,M\}$. L'ensemble $\mathcal{TR}$ permet de déterminer l'ordre de pertinence des différents noyaux. Dans ce contexte, un nouvel ensemble de noyaux $\mathcal{K}' = [k'_1,k'_2,...,k'_M]$ est défini par l'organisation des noyaux de l'ensemble $\mathcal{K}$ selon l'ordre des taux d'apprentissage décroissant. Le premier noyau k'_1 est ainsi le noyau le plus pertinent, tandis que le dernier noyau dans $\mathcal{K}'$ est le moins pertinent (k'_M).

3.4.4 Calcul des différents poids de noyaux

Afin de mettre en valeur les noyaux les plus pertinents, nous réalisons une pondération selon l'organisation des noyaux effectuée dans $\mathcal{K}'$:

$$\beta_m = 2\,\frac{M-m+1}{M(M+1)}, \quad \forall m \in \{1,2,...,M\} \tag{3.15}$$

$$\text{tel que } TR_{k'_m} \;\geq\; TR_{k'_{m+1}}, \forall m \in \{1,2,...,M-1\}$$

où $\forall m \in \{1,2,...,M\}$, $\beta_m \geq 0$, $\sum_{m=1}^{M} \beta_m = 1$ et $\forall m \in \{1,2,...,M-1\}$, $\beta_m > \beta_{m+1}$.

Soit $\boldsymbol{\beta} = [\beta_1,\beta_2,...,\beta_M]$ l'ensemble des poids correspondants aux différents noyaux k'_m, $m = [1,2,...,M]$. Selon cette nouvelle pondération (équation (3.15)), le noyau le plus pertinent, k'_1, est mieux considéré avec le coefficient de pondération le plus élevé β_1. Ainsi

de suite, les valeurs des coefficients β_m décroissent harmonieusement avec les taux d'apprentissage $TR_{k'_m}$, $m = [1,2,...,M]$. Compte tenu de cette nouvelle pondération, que l'on appelle par l'acronyme anglais KW-BTR (Kernel Weighting Based on Training Rates), la combinaison linéaire des différents noyaux employés est alors donnée par:

$$K(x,x') = \sum_{m=1}^{M} 2 \ \frac{M-m+1}{M(M+1)} \ k'_m(x,x') \tag{3.16}$$

$$\text{tel que } TR_{k'_m} \ \geq \ TR_{k'_{m+1}}, \forall m \in \{1,2,...,M-1\}$$

Selon l'équation (3.16), l'approche MKL est adressée par la résolution d'un problème SVM standard à noyau unique. La fonction de décision MK_SVM est formulée par l'équation suivante:

$$f(x) = \sum_{i=1}^{N_{Im}} \alpha_i y_i \left(\sum_{m=1}^{M} 2 \ \frac{M-m+1}{M(M+1)} \ k'_m(x,x_i) \right) + b \tag{3.17}$$

avec $TR_{k'_m} \ \geq \ TR_{k'_{m+1}}, \forall m \in \{1,2,...,M-1\}$.

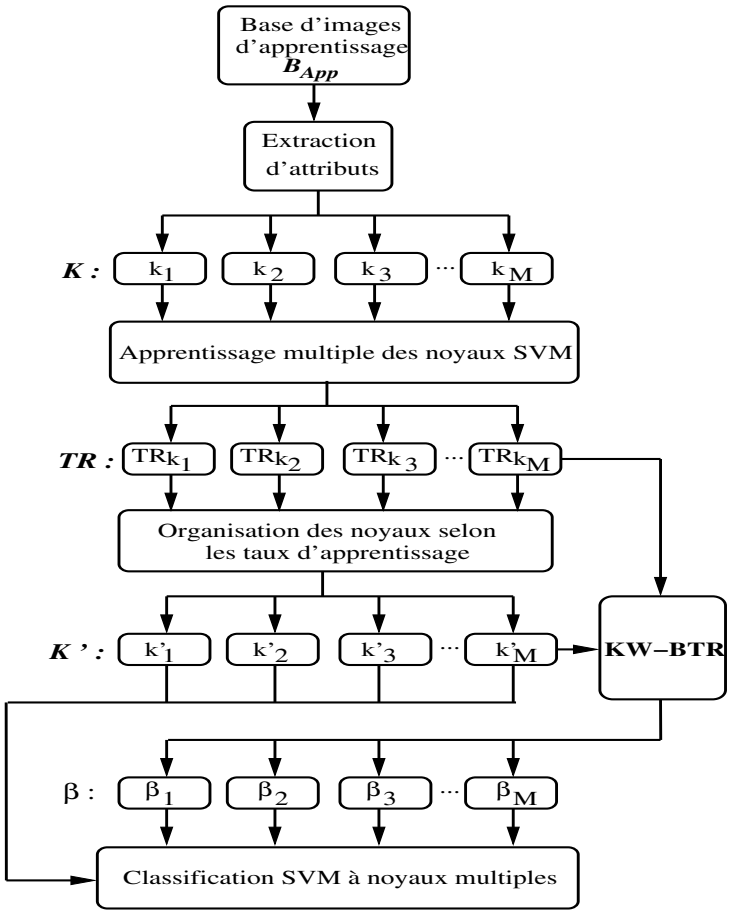


FIG. 3.4 – *Diagramme de la méthode de classification MK_SVM utilisant la nouvelle pondération KW-BTR.*

La figure 3.4 illustre le diagramme de la méthode MK_SVM. Ce diagramme présente l'organisation des noyaux utilisés en fonction de leurs taux d'apprentissage SVM. En se

basant sur cette organisation des noyaux, les poids sont calculés à l'aide de la pondération KW-BTR. L'algorithme MK_SVM est présenté dans la suite:

Algorithme MK_SVM

MK_SVM ($\mathcal{B}_{app}$,$\mathcal{K}$)

Pour chaque noyau $k_m \in \mathcal{K}$, $m = 1,2,...,M$ **Faire**
    Apprentissage SVM (X,k_m); Eq. 3.3
    $TR_{k_m} = \frac{N_{Corr}(k_m)}{N_{Im}}$; Eq. 3.14
Fin Pour
$\mathcal{K}' =$ *Organisation des taux d'apprentissages des différents noyaux* ($\mathcal{K}$,$\mathcal{TR}$)
Pour chaque noyau $k'_m \in \mathcal{K}'$, $m = 1,2,...,M$ **Faire**
    $\beta_m = KW\text{-}BTR\ (k'_m, TR_{k'm})$; Eq. 3.15
Fin Pour
Calcul du noyau multiple ($\mathcal{K}'$,$\boldsymbol{\beta}$); Eq. 3.16
Apprentissage SVM à base de noyaux multiples (X,K); Eq. 3.17

3.4.5 Conclusion

Dans cette section, La méthode MK_SVM a été présentée. Particulièrement, l'originalité de cette méthode réside dans sa nouvelle technique de pondération, que l'on a appelée KW-BTR. En se basant sur le degré de pertinence de chaque noyau, KW-BTR permet de mieux considérer les noyaux les plus pertinents. Au lieu d'être tirés du même problème d'optimisation que les hyper-paramètres SVM, les poids de noyaux sont calculés linéairement en fonction des taux d'apprentissage correspondants. Ainsi, les temps de calcul sont significativement réduits. Une évaluation de la méthode MK_SVM est illustrée et discutée dans la section suivante.

3.5 Évaluation de la méthode MK_SVM

L'ensemble $\mathcal{K}$ des noyaux utilisés pour les expérimentations est le suivant:

- **Le noyau Linéaire**[2].
- **Le noyau Polynômial** (équation (3.4)), avec les paramètres $\{ct = 1,u = 5\}$.
- **Le noyau Sigmoîdal** (équation (3.5)), avec le paramètre $\{\theta = 1\}$.
- **Le noyau Gaussien** (équation (3.7)), avec le paramètre $\{\sigma = 10.5\}$.

Les paramètres des noyaux, ainsi que le terme de régularisation de marge C[3] sont déterminés empiriquement pour la base d'images COREL utilisée.

2. Le noyau de type produit scalaire euclidien correspondant à ne pas modifier l'espace d'attributs initial $\mathcal{F}$.

3. La valeur du terme de régularisation de marge est choisie de façon à mieux contrôler le phénomène de sur-apprentissage.

3.5.1 Résultats de classification

En utilisant une concaténation des douze caractéristiques présentées dans le chapitre 1, un apprentissage multiple des différents noyaux $k_m \in \mathcal{K}$, $\forall m \in \{1,2,3,4\}$ est effectué pour quatre sous-bases COREL. Les poids des différents noyaux sont alors calculés selon la pondération KW-BTR proposée (équation (3.15)). Un exemple des taux d'apprentissage obtenus ainsi que les poids correspondants des différents noyaux utilisés pour la sous-base COREL avec 5 classes est fourni dans le tableau 3.1. En se basant sur cette pondération des noyaux, une classification MK_SVM est réalisée pour différentes sous-bases COREL.

TAB. 3.1 – *Exemple de pondération de noyaux KW-BTR pour la sous-base COREL avec 5 classes.*

k'_m	Gaussien	Polynômial	Linéaire	Sigmoîdal
$TR_{k'_m}$	1	1	0.9573	0.8964
β_m	0.4	0.3	0.2	0.1

Dans le même contexte, une classification des mêmes sous-bases COREL est également effectuée à l'aide de l'algorithme *SimpleMKL*[4] pour la comparaison des résultats. Le tableau 3.2 présente les performances de la classification MK_SVM et *SimpleMKL* pour quatre sous-bases COREL. Les meilleures performances sont affichées en gras.

TAB. 3.2 – *Comparaison des résultats de classification MK_SVM et* SimpleMKL *pour différentes sous-bases d'images COREL.*

Sous-bases COREL		*SimpleMKL*	**MK_SVM**
3 classes	P_{moy}	0.9753	**1.0000**
	R_{moy}	0.9733	**1.0000**
5 classes	P_{moy}	0.9028	**0.9467**
	R_{moy}	**0.9801**	0.9440
7 classes	P_{moy}	0.8640	**0.8901**
	R_{moy}	0.8514	**0.8857**
10 classes	P_{moy}	0.8118	**0.8510**
	R_{moy}	**0.8360**	0.8080

D'après le tableau 3.2, MK_SVM classifie correctement toutes les images requêtes avec une précision moyenne égale à 100%, pour la sous-base COREL avec 3 classes. Alors que la précision moyenne de la classification *SimpleMKL* pour la même sous-base est 97.53%. De même, pour les sous-bases COREL avec 5, 7 et 10 classes, MK_SVM est toujours plus efficace que *SimpleMKL* en terme de précision. Les courbes de précisions moyennes correspondantes aux deux méthodes comparées sont présentées dans la figure 3.5.

(figure 4.3 (a))

Pour les quatre sous-bases COREL employées, les courbes précision-rappel de MK_SVM et *SimpleMKL* sont illustrées respectivement dans les figures 3.6 (a) et (b). Bien que la différence ne soit pas significative, nous observons que MK_SVM est toujours plus performante que *SimpleMKL*.

4. Une boite à outil MatLab de l'algorithme *SimpleMKL*, développée par Alain Rakotomamonjy, est disponible à l'adresse suivante: http://asi.insa-rouen.fr/enseignants/~arakotom/code/mklindex.html.

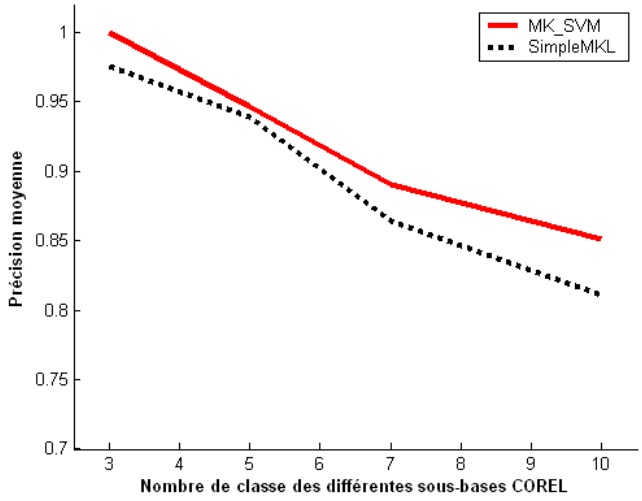


FIG. 3.5 – *Performances de classification moyenne des deux algorithmes MK_SVM et* SimpleMKL *pour différentes sous-bases d'images COREL.*

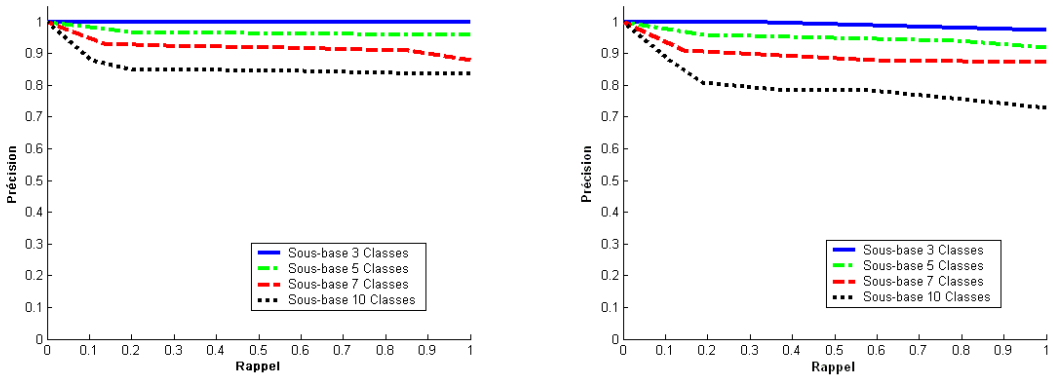


FIG. 3.6 – *Courbes précision-rappel de a) MK_SVM et b)* SimpleMKL *pour différentes sous-bases d'images COREL.*

Pour la sous-base COREL avec 3 classes, la courbe précision-rappel de MK_SVM est parfaite[5], mais celle de l'algorithme *SimpleMKL* décroît légèrement. De même, pour les trois autres sous-bases COREL, les courbes précision-rappel de MK_SVM sont meilleures que celles de *SimpleMKL*. Par exemple, pour la sous-base COREL avec 7 classes, les précisions de MK_SVM et *SimpleMKL* sont respectivement autour de 92% et 85%, pour une valeur de rappel de 60%. Aussi, pour la sous-base COREL avec 10 classes, à 40% de rappel, leurs précisions respectives sont autour de 88% et 78%. Les temps de calcul des deux méthodes sont discutées dans la section suivante.

3.5.2 Temps d'exécution

Le tableau 3.3 présente les temps d'apprentissage et de pondération des noyaux des deux algorithmes pour différentes sous-bases COREL. Le temps d'exécution le plus rapide pour chaque sous-base est affiché en gras. Nous pouvons observer clairement que

5. Précision égale à 1.000 pour toutes valeurs de rappel.

l'algorithme MK_SVM, utilisant la nouvelle pondération de noyaux KW-BTR, nécessite significativement moins de temps de calcul que l'algorithme *SimpleMKL*. L'algorithme MK_SVM assure l'apprentissage et la pondération des noyaux, pour la sous-base COREL avec 3 classes, en moins d'une minute, cependant *SimpleMKL* nécessite 2,7084 *mn*. De même, pour les sous-bases COREL avec 5, 7 et 10 classes, MK_SVM est plus rapide que *SimpleMKL*. Effectivement, par exemple MK_SVM nécessite seulement 12.8102 *mn* pour l'apprentissage et la pondération des noyaux pour la sous-base COREL avec 10 classes, alors que *SimpleMKL* requière dans ce cas 37.8110 *mn*. Ceci s'explique par le fait que MK_SVM, à la différence de *SimpleMKL*, ne résout aucun problème d'optimisation. La pondération des noyaux dans MK_SVM s'effectue une fois pour toute en se basant sur la pertinence des taux d'apprentissage correspondants. Par contre, *SimpleMKL* détermine les coefficients de pondération des différents noyaux itérativement selon une optimisation de la descente du gradient. D'où ses besoins en temps de calcul.

TAB. 3.3 – *Comparaison des temps d'apprentissage et de pondération (en Minutes) des deux algorithmes MK_SVM et* SimpleMKL *pour différentes sous-bases COREL.*

Sous-bases COREL	Temps d'exécution	
	SimpleMKL	**MK_SVM**
3 classes	2.7084	**0.8619**
5 classes	7.7222	**3.8038**
7 classes	28.4550	**8.5301**
10 classes	37.8110	**12.8102**

3.6 Conclusion

Nous avons présenté dans ce chapitre un algorithme de classification SVM efficace à noyaux multiples (MK_SVM). Une pondération originale des noyaux a été proposée pour cette nouvelle classification. Les poids des noyaux sont calculés en fonction des taux d'apprentissage correspondants. En plus de l'amélioration des résultats de classification prouvée par les expérimentations réalisées, MK_SVM assure significativement un temps d'exécution inférieur aux algorithmes MKL existants. Une comparaison avec un algorithme récent est également fournie.

En considérant les mêmes images et la même formulation de noyaux, l'apprentissage MKL peut être effectué selon des caractéristiques différentes. L'optimisation des poids des différents noyaux représente ainsi une sorte de sélection des caractéristiques. Cependant une telle pondération linéaire des caractéristiques présente certaines limites. En particulier, elle s'avère incapable de décrire les relations non-linéaires des différents types de description. De plus, la pondération calculée des noyaux est maintenue fixe sur tous l'ensemble des images considérées. Nous proposons, dans le chapitre suivant, une nouvelle méthode de classification multi-modèles permettant d'assurer une combinaison de caractéristiques plus flexible. Les caractéristiques sélectionnées sont employées hiérarchiquement en se basant sur leur pouvoir discriminant dans la base d'images considérée.

Chapitre 4

Méthode de classification hiérarchique multi-modèles

Dans ce chapitre, une Méthode originale de Classification hiérarchique Multi-Modèles que l'on appelle MC-MM est proposée. Cette classification permet d'assurer la reconnaissance de chaque image requête itérativement selon un ordre hiérarchique des différentes caractéristiques sélectionnées.

4.1 Introduction

Face aux limites rencontrées dans les approches MKL, de nouvelles méthodes sont proposées pour palier aux inconvénients de la combinaison linéaire des caractéristiques de différents types (Cao et al., 2009). Ces méthodes utilisent des poids variables qui conduisent à une fusion non-linéaire et localement adaptative des caractéristiques multiples. Dans le même contexte, nous proposons dans ce chapitre, une nouvelle Méthode de Classification hiérarchique Multi-Modèles que l'on appelle MC-MM (section 4.2). Une évaluation de MC-MM à travers différentes méthodes de sélection ainsi qu'une comparaison avec d'autres méthodes de classification sont présentées dans la section 4.3. Les résultats expérimentaux obtenus montrent que la méthode de classification proposée assure toujours de très bonne performances de classification.

4.2 Classification hiérarchique multi-modèles

4.2.1 Introduction

MC-MM permet de reconnaître, itérativement, les images de test selon un ordre hiérarchique reflétant le pouvoir discriminant des différentes caractéristiques considérées. Elle utilise en particulier un bloc approprié pour lever l'ambiguïté de classification dans chaque niveau hiérarchique. Ainsi, selon l'image requête traitée, MC-MM permet d'ajuster l'organisation des caractéristiques en tant qu'une sorte de pondération dynamique. En conséquence, une grande souplesse dans la combinaison des caractéristiques est assurée.

À l'opposé de certaines méthodes qui évaluent la pertinence des différentes images contenues dans la base d'apprentissage (Cao et al., 2009), MC-MM tient compte de toutes les images pour la construction du modèle complet de chaque caractéristique. Absolument, les images considérées comme aberrantes appartiennent réellement à la base de données d'apprentissage. Elles représentent ainsi un complément d'information permettant de modéliser les différentes classes de la base hétérogène utilisée.

4.2.2 Principe de l'approche multi-modèles

Prenant en compte l'hétérogénéité du contenu des images, le système de reconnaissance étudié peut être considéré comme un système complexe. De tels systèmes peuvent être toujours modéliser par plusieurs modèles plus simples en affectant chaque modèle à un sous-ensemble du système global. Dans ce contexte, nous présentons une nouvelle méthode de classification (MC-MM) destinée pour les systèmes de reconnaissance d'images hétérogènes par le contenu. MC-MM combine hiérarchiquement les résultats de classification des différents modèles SVMs sélectionnés, d'où son aspect hiérarchique multi-modèles. L'ordre hiérarchique utilisé est basé sur les degrés de pertinence des modèles sélectionnés, c'est celui de leurs performances d'apprentissage croissantes. Ceci permet d'assurer une amélioration progressive de la décision de classification et par la suite une qualité de reconnaissance ascendante.

Selon la taxonomie utilisée pour les méthodes hiérarchiques (Jain et al., 1999), MC-MM peut être considérée comme une approche monothétique puisqu'elle utilise les différentes caractéristiques d'une manière séquentielle. Par conséquent, chaque image requête peut être classifiée différemment au cours du processus de reconnaissance (figure 4.2). Par

ailleurs, MC-MM est aussi une approche déterministe (Jain et al., 1999). En effet, chaque image requête n'est attribuée, à la fin du processus de classification, qu'à une seule classe. Pour cela, si une image requête est classifiée différemment selon deux niveaux hiérarchiques consécutifs le classifieur du plus proche centre de classe, que l'on appelle NCC pour *Nearest Cluster Center*, est employé.

4.2.3 Résolution de conflit de décision par NCC

Rappelons que $\mathcal{C} = \{C_1, C_2, ..., C_{N_C}\}$ est l'ensemble des N_C classes contenues dans une base d'images d'apprentissage $\mathcal{B}_{App}$. Le classifieur NCC représente un processus simple permettant d'attribuer une image requête à la classe la plus proche dans l'ensemble $\mathcal{C}$ selon un espace d'attributs donné.

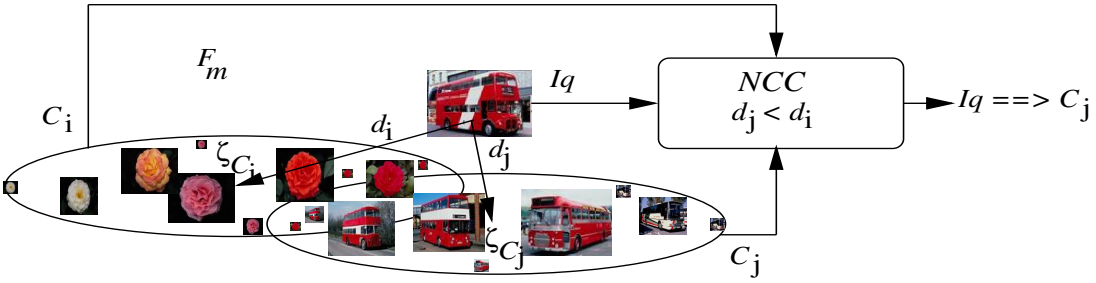


FIG. 4.1 – *Règle de décision du classifieur NCC: classification de l'image requête I_q dans l'espace d'attributs $\mathcal{F}_m$, à travers les deux classes évaluées C_i et C_j.*

Soit une image requête I_q attribuée différemment aux classes C_i et C_j, $1 \leq i \neq j \leq N_C$, avec les modèles considérés par deux niveaux hiérarchiques consécutifs. On note ces deux niveaux par Niv_l et Niv_m. Supposant que le modèle considéré par Niv_m est plus pertinent que celui utilisé dans le niveau Niv_l et que $\mathcal{F}_m$ est son espace d'attributs correspondant. Initialement, NCC calcule dans l'espace $\mathcal{F}_m$ les centres ζ_{C_i} et ζ_{C_j} respectivement des deux classes C_i et C_j. Les centres de classes sont représentés par la caractéristique moyenne des images d'apprentissage appartenant aux classes correspondantes. En fait, chaque centre calculé est fictif et peut ne correspondre à aucune image dans la classe considérée. On note par $\mathcal{F}_m(I_q)$ la projection de l'image I_q dans l'espace $\mathcal{F}_m$. Comme le montre la figure 4.1, la règle de décision du classifieur NCC consiste à attribuer l'image I_q à la classe dont le centre est le plus proche. Ceci se base principalement sur les deux distances d_i et d_j, qui correspondent aux mesures de similarité calculées entre $\langle \mathcal{F}_m(I_q), \zeta_{C_i} \rangle$ et $\langle \mathcal{F}_m(I_q), \zeta_{C_j} \rangle$. Pour cela, nous employons la distance Euclidienne (Wang et al., 2005). Cette distance est la plus utilisée pour les mesures de similarité dans les espaces d'attributs. Toutefois, il est toujours possible d'utiliser d'autres métriques de l'état de l'art (Dengsheng and Guojun, 2003).

Il est à signaler que l'apport du classifieur NCC n'est requis qu'en cas de conflit entre les décisions SVMs des niveaux hiérarchiques consécutifs. Effectivement, les SVMs, agissant par séparation, exploitent les relations de chaque classe avec les autres. Par contre, le classifieur NCC modélise chaque classe séparément. Dans ce qui suit, on détaille le principe de la méthode de classification MC-MM.

4.2.4 Classification hiérarchique

Vu que les caractéristiques extraites pour la description d'images dans les bases hétérogènes, ne sont pas nécessairement toutes pertinentes et utiles pour la reconnaissance, notre nouvelle méthode de classification hiérarchique envisage une pré-sélection des caractéristiques les plus pertinentes. Ce sujet est abordé dans le chapitre 2. Dans cette section, nous considérons les résultats de la méthode de sélection adaptative, que l'on a appelée SAC. À partir des n caractéristiques, extraites initialement, la méthode SAC évalue la pertinence des n modèles correspondants ($\mathcal{M} = \{M_1,M_2,...,M_n\}$). Une analyse discriminante des performances des différents modèles dans $\mathcal{M}$, permet de sélectionner un sous-ensemble de k modèles ($\mathcal{M}_s = \{M_{s1},M_{s2},...,M_{sk}\}$, $k \leq n$). Ces modèles sont considérés comme les plus pertinents. En phase de généralisation, seulement les caractéristiques correspondantes aux modèles contenus dans $\mathcal{M}_s$ sont extraites à partir d'une base d'images de test $\mathcal{B}_{Test}$. Cela permet d'accélérer davantage cette phase, puisqu'on évite d'extraire les caractéristiques correspondantes aux modèles jugés non pertinents.

Pour une image requête I_q dans $\mathcal{B}_{Test}$, la nouvelle méthode de classification hiérarchique multi-modèles MC-MM considère initialement le modèle ayant la plus faible performance d'apprentissage dans $\mathcal{M}_s$ (M_{sk}). Notons le résultat de classification de I_q avec ce modèle par $C^{M_{sk}}(I_q) \in \mathcal{C}$. Cette classification est affinée progressivement à travers les différents niveaux hiérarchiques utilisés. En fait, à chaque niveau dans l'architecture MC-MM, I_q est classifiée avec un modèle subséquent dans $\mathcal{M}_s$, jusqu'à atteindre au dernier niveau le modèle le plus performant M_{s1} (figure 4.2). La classification de I_q selon chaque niveau est toujours comparée à celle obtenue avec le niveau précédent dans l'hiérarchie de MC-MM. Cette classification est principalement validée si les deux niveaux comparés présentent une classification similaire de I_q. Lorsqu'il y a conflit entre les décisions prises par deux niveaux consécutifs, l'ambiguïté est levée grâce à l'utilisation du classifieur NCC présenté dans la section 4.2.3.

Au premier niveau hiérarchique, le résultat de classification de l'image I_q avec le modèle M_{sk-1}, que l'on note par $C^{M_{sk-1}}(I_q) \in \mathcal{C}$, est comparé avec $C^{M_{sk}}(I_q)$. Si I_q est classifiée d'une façon similaire avec les deux modèles M_{sk} et M_{sk-1}, la classification de I_q dans ce premier niveau est directement validée:

$$C^{M_{sk},M_{sk-1}}(I_q) = C^{M_{sk}}(I_q) = C^{M_{sk-1}}(I_q)$$

Dans le cas contraire ($C^{M_{sk}}(I_q) \neq C^{M_{sk-1}}(I_q)$), I_q est attribuée à la classe validée par le classifieur NCC. La distance de la projection de I_q dans l'espace d'attributs $\mathcal{F}_1$ [1] par rapport aux deux classes $C^{M_{sk}}(I_q)$ et $C^{M_{sk-1}}(I_q)$ est examinée. La classe $C^{M_{sk},M_{sk-1}}(I_q)$ est donc celle ayant le centre le plus proche.

De la même façon dans le deuxième niveau, les résultats de classification $C^{M_{sk},M_{sk-1}}(I_q)$ et $C^{M_{sk-2}}(I_q)$ sont comparés. Ainsi de suite jusqu'à atteindre le dernier niveau hiérarchique considérant le modèle le plus performant M_{s1}. Notons par $C(I_q) = C^{M_{sk},M_{sk-1},...,M_{s1}}(I_q)$ la classification de I_q avec MC-MM. On peut formuler cette classification comme suit:

$$C(I_q) = [\ [\ [C^{M_{sk}}(I_q) \otimes C^{M_{sk-1}}(I_q)]... \otimes C^{M_{s2}}(I_q)] \otimes C^{M_{s1}}(I_q)], \tag{4.1}$$

1. $\mathcal{F}_1$ correspond au premier niveau Niv_1 dans l'architecture hiérarchique MC-MM.

avec

$$C^{M_{si}}(I_q) \otimes C^{M_{sj}}(I_q) = \begin{cases} C^{M_{sj}}(I_q), & \text{si } C^{M_{si}}(I_q) = C^{M_{sj}}(I_q) \\ NCC(I_q, C^{M_{si}}(I_q), C^{M_{sj}}(I_q)), & \text{sinon} \end{cases}$$

Bien que basée sur deux classifieurs (SVM et NCC), la classification MC-MM proposée est principalement discriminante via la classification des SVMs. Les mesures de similarité effectuées en cas d'utilisation du classifieur NCC sont employées pour situer les images candidates exclusivement par rapport aux deux classes proposées par les modèles SVMs considérés. La figure 4.2, illustre l'architecture de la classification multi-modèles MC-MM.

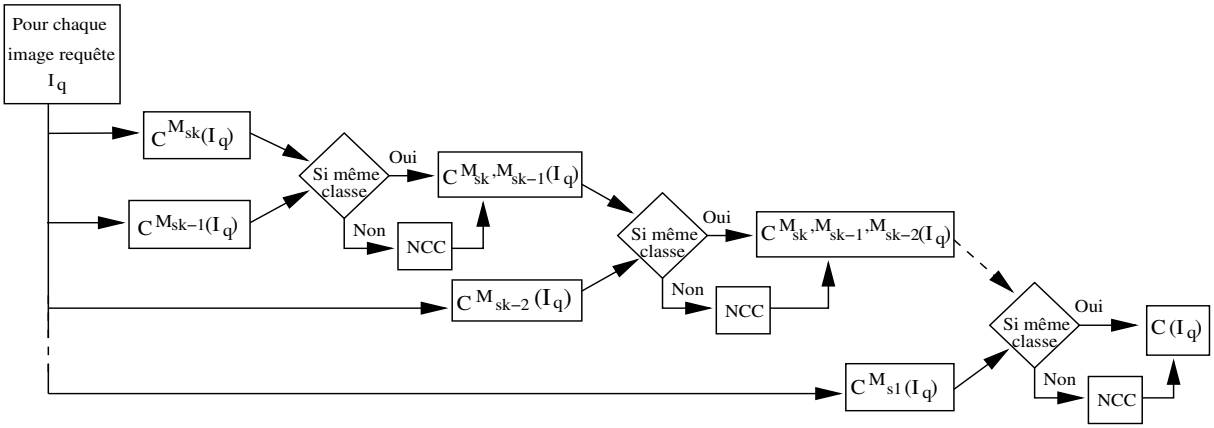


FIG. 4.2 – *Diagramme illustrant l'architecture de la méthode de classification hiérarchique multi-modèles MC-MM.*

Pour résumer la discussion précédente en pseudo code, les algorithmes MC-MM et NCC sont fournis dans ce qui suit:

Algorithme MC-MM

MC-MM ($\mathcal{B}_{Test}$,$\mathcal{M}_s$); Eq. 4.1

Pour chaque image test $I_q \in \mathcal{B}_{Test}$ **Faire**
    $C(I_q) = C^{M_{sk}}(I_q)$
    Pour chaque modèle $M_{sj} \in \mathcal{M}_s$, $j = k$-1,...,2,1 **Faire**
        Si $C(I_q) \neq C^{M_{sj}}(I_q)$ **Alors**
            $C(I_q) = NCC(I_q, C(I_q), C^{M_{sj}}(I_q))$
        Fin si
    Fin pour
Fin pour

Algorithme NCC

NCC (I_q,$C^{M_{si}}$,$C^{M_{sj}}$)

    $\mathcal{F}(I_q) = \mathcal{F}_{M_{si}}(I_q)$
Si $Niv_{M_{si}} < Niv_{M_{sj}}$ **Alors**
    $\mathcal{F}(I_q) = \mathcal{F}_{M_{sj}}(I_q)$

4: **Fin si**
5: $d_i = |\mathcal{F}(I_q) - \zeta_{C^{M_{si}}}|$; $d_j = |\mathcal{F}(I_q) - \zeta_{C^{M_{sj}}}|$
6: **Si** $d_i < d_j$ **Alors**
7: $C(I_q) = C^{M_{si}}$
8: **Sinon**
9: $C(I_q) = C^{M_{sj}}$
10: **Fin si**

4.2.5 Conclusion

Dans cette section, la méthode MC-MM a été présentée. Il s'agit d'une classification hiérarchique qui ne nécessite aucune rétroaction de l'utilisateur ou bouclage de pertinence. Les résultats de classification sont affinés progressivement à travers les différents niveaux hiérarchiques utilisés. En se basant sur la méthode de sélection SAC, présentée dans le chapitre 2, MC-MM assure des performances de classification très compétitives grâce à l'utilisation hiérarchique de modèles bien adaptés au contenu de la base d'images considérée. De plus, l'ordre d'utilisation des différents modèles sélectionnés peut être ajuster selon une complémentarité de décision de chaque deux niveaux consécutifs. Les performances de classification MC-MM sont présentées, discutées et comparées dans la section suivante.

4.3 Évaluation et comparaison

Dans cette section, une évaluation des performances de la nouvelle méthode de classification hiérarchique MC-MM est présentée. La section 4.3.1 est réservée à l'évaluation de MC-MM à travers différentes méthodes de sélection. Les performances de MC-MM utilisant particulièrement la méthode de sélection SAC sont étudiées dans la section 4.3.2. Des comparaisons des résultats de classification de MC-MM basée SAC avec les résultats d'autres méthodes connues dans la littérature sont enfin discutées dans la section 4.3.3.

4.3.1 Évaluation de MC-MM à travers différentes méthodes de sélection

Nous discutons dans cette section les résultats de classification MC-MM à travers la méthode SAC et les différentes méthodes de sélection AG présentées dans le chapitre 2. Le même ensemble de caractéristiques est employé (pour plus de détails voir chapitre 1). Une étude comparative entre les trois méthodes de sélection AG (AG_1, AG_2 et AG_3) et la méthode SAC est effectuée. La base COREL avec 10 classes est utilisée pour les expérimentations.

Le tableau 4.1 illustre les résultats de classification MC-MM à travers les différentes méthodes de sélection évaluées. Les résultats obtenus montrent que la méthode de sélection AG_2 est la plus efficace parmi les trois algorithmes génétiques de sélection utilisés. Ceci confirme que AG_1 et AG_3 convergent vers des optimums locaux comme il était déjà évoqué dans le chapitre 2. En conséquence, ces deux méthodes de sélection ne permettent pas d'aboutir à des performances de classification assez élevées. Cependant, bien qu'elle assure

TAB. 4.1 – *Classification MC-MM à ravers différentes méthodes de sélection pour quatre sous-bases d'images COREL.*

Sous-bases COREL		Méthodes de sélection			
		AG_1	AG_2	AG_3	SAC
3 classes	P_{moy}	0.9293	0.9540	0.8811	**0.9879**
	R_{moy}	0.9200	**0.9467**	0.8533	0.9333
5 classes	P_{moy}	0.8467	0.9230	0.8235	**0.9750**
	R_{moy}	0.8240	**0.9120**	0.8080	**0.9120**
7 classes	P_{moy}	0.7827	0.8554	0.7889	**0.9123**
	R_{moy}	0.6914	**0.8229**	0.7657	**0.8229**
10 classes	P_{moy}	0.6877	0.8207	0.7286	**0.8371**
	R_{moy}	0.6200	0.7880	0.7120	**0.8160**

avec MC-MM des résultats de classification compétitifs, la méthode AG_2 s'avère moins performante que la méthode SAC. En effet, pour les différentes sous-bases COREL utilisées les résultats de classification MC-MM basée SAC sont toujours les plus performants. Ce qui prouve l'efficacité de la combinaison de la sélection adaptative avec la classification hiérarchique multi-modèles. Dans le reste de ce chapitre les résultats présentés sont tous basés sur le couple ⟨SAC, MC-MM⟩.

4.3.2 Performances de MC-MM utilisant la méthode de sélection SAC

En se basant sur la sélection adaptative, on discute dans cette section les résultats de classification MC-MM. Comme il est détaillé dans la section 4.2.4, la généralisation MC-MM procède hiérarchiquement à partir des modèles les moins performants vers les plus performants. Une évaluation de cette procédure de généralisation est proposée. Dans ce cadre, nous comparons les performances de la généralisation MC-MM avec deux autres procédures de généralisation possibles. La première adopte le même principe proposé sauf qu'elle procède de façon opposée, c'est à dire des modèles les plus performants vers les moins performants. La deuxième assigne les images requêtes à la classe considérée par la majorité des modèles sélectionnés. Dans ce qui suit, ces deux généralisations sont indiquées, respectivement, par *généralisation décroissante* et *généralisation max_modèles*. Le tableau 4.2 illustre l'étude comparative des performances des différentes généralisations étudiées pour quatre sous-bases d'images COREL. Les meilleures et les pires performances dans ce tableau sont affichées en gras.

D'après le tableau 4.2, on remarque que la procédure de généralisation MC-MM est la plus performante parmi les trois procédures comparées. Pour différentes sous-base COREL, MC-MM assure toujours des taux de classifications prépondérants que les deux autres procédures de généralisation. Ceci est justifié par la qualité ascendante de la généralisation MC-MM. Les images requêtes sont initialement classifiées selon le modèle le moins performant et tout au long du processus de généralisation MC-MM, la classification de ces images est affinée davantage avec des modèles plus pertinents. D'où, une amélioration progressive de la classification des images à travers les différents niveaux hiérarchiques employés. La *généralisation décroissante* qui procède de façon opposée est la moins performante parmi les trois procédures évaluées. Le principe même de cette

TAB. 4.2 – *Évaluation de la généralisation MC-MM. Étude comparative avec les deux procédures de généralisation* décroissante *et* max_modèles *pour différentes sous-bases COREL.*

Sous-bases COREL	Performances de classification des trois procédures de généralisation		
	décroissante	max_modèles	**MC-MM**
3 classes	**0.8248**	0.9803	**0.9879**
5 classes	**0.7555**	0.8425	**0.9750**
7 classes	**0.7512**	0.7914	**0.9123**
10 classes	**0.7096**	0.7638	**0.8371**

généralisation permet aux modèles les moins performants de dégrader les résultats de classification des images. En fait, l'ordre d'utilisation des modèles dans les deux procédures de généralisation MC-MM et *décroissante* constitue une sorte de pondération de ces modèles. Les premiers modèles utilisés sont les moins influents vu que la classification des images avec ces modèles risque d'être modifiée par les modèles qui suivent. Par conséquent, il est évident que les modèles les moins performants doivent être employés en premiers lieu dans une telle architecture. Ce qui explique la supériorité des taux de classifications MC-MM par rapport à ceux de la *généralisation décroissante*. La *généralisation max_modèles* assure aussi des résultats meilleurs que ceux de la *généralisation décroissante*, mais bien qu'elle soit parfois assez compétitive (sous-base COREL avec 3 classes par exemple), MC-MM reste toujours la plus performante. De plus, la *généralisation max_modèles* ne prend pas en compte les degrés de pertinence des différents modèles sélectionnés. Elle peut classifier ainsi les images requêtes selon des modèles moins performants.

TAB. 4.3 – *Matrice de confusion de la classification MC-MM pour la sous-base COREL avec* 10 *classes.*

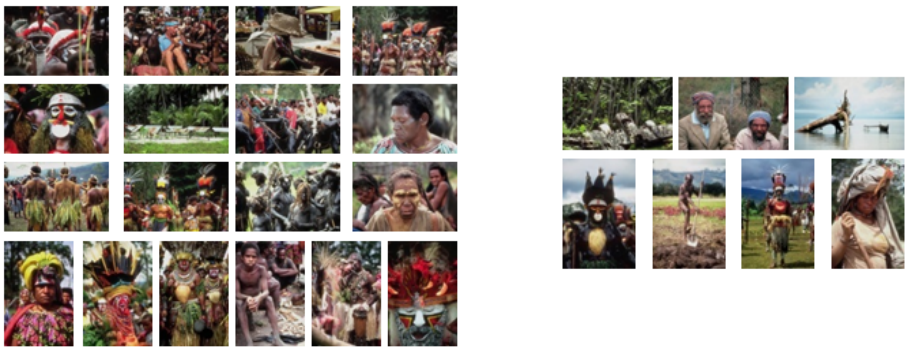
	Cl. 0	Cl. 1	Cl. 2	Cl. 3	Cl. 4	Cl. 5	Cl. 6	Cl. 7	Cl. 8	Cl. 9
Cl. 0	**84**	0	0	0	0	0	8	8	0	0
Cl. 1	0	**96**	0	0	0	0	0	0	0	4
Cl. 2	0	0	**100**	0	0	0	0	0	0	0
Cl. 3	0	0	0	**73**	0	0	0	0	27	0
Cl. 4	9	12	0	0	**71**	4	0	4	0	0
Cl. 5	8	0	0	0	0	**92**	0	0	0	0
Cl. 6	4	0	0	12	0	0	**72**	0	12	0
Cl. 7	8	0	0	8	0	0	0	**76**	8	0
Cl. 8	6	0	0	17	0	0	4	0	**73**	0
Cl. 9	0	0	0	0	0	0	0	0	0	**100**

On présente dans le tableau 4.3 la matrice de confusion obtenue suite à la classification MC-MM de la base COREL avec 10 classes[2]. Les colonnes de cette matrice représentent les classes estimées, tandis que les lignes représentent les classes réelles (ou de référence). Ainsi, le pourcentage des images correctement classifiées est illustré par la diagonale de cette matrice, ces valeurs sont affichés en gras dans le tableau 4.3. En effet, un système de

2. Les valeurs affichées sont en pourcentage.

classification est d'autant meilleur que sa matrice de confusion s'approche d'une matrice diagonale. Un des intérêts de cette matrice est qu'elle montre rapidement si le système parvient à classifier correctement. D'après le tableau 4.3, on peut voir que la classification est réussie à 100% pour les classes d'identifiants 2 et 9, correspondant respectivement aux classes *Dinosaurs* et *Cars*[3]. Pour le reste des classes les taux de classification varient entre 71% et 96%. Ce qui prouve l'efficacité de la classification MC-MM qui s'avère aussi performante face à la complexité du contenu de la base d'images hétérogène COREL.

À titre indicatif, on présente dans la figure 4.3 un exemple d'images-requêtes de la classe *African peoples and villages*[4], les images-candidates qui ont été correctement classifiées et celles mal classifiées.



a) b)

FIG. 4.3 – *a) Les images correctement classifiées et b) celles mal classifiées de la classe* African peoples and villages *pour la sous-base d'images COREL utilisée (*10 *classes).*

On constate que 18 images-requêtes parmi 25 sont correctement classifiées (figure 4.3 (a)) et seulement 7 sont mal classifiées (figure 4.3 (b)). Nous tenons à signaler que les caractéristiques sélectionnées pour la base COREL avec 10 classes s'avèrent pertinentes pour la classification de 72% des images-requêtes de la classe *African peoples and villages*. Cependant, nous précisons que la faiblesse de description des images mal classifiées est due principalement au contenu même de ces images. D'après la figure 4.3, nous pouvons bien remarqué qu'au contraire des images correctement classifiées, la plupart des images mal classifiées contiennent une majorité de régions communes avec d'autres classes dans la base. Ces régions (d'herbes ou de ciel) semblent être caractérisantes pour certaines classes autre que la classe *African peoples and villages*. Ceci explique bien les inévitables erreurs de reconnaissance présentées, par exemple, par la classification erronée de six parmi les sept images illustré par la figure 4.3 (b) dans les classes *Dinosaurs* et *Mountains and glaciers*, d'identifiants respectives 3 et 8 (tableaux 1.1 et 4.3).

4.3.3 Comparaison des résultats MC-MM basée SAC avec des méthodes de classification connues

4.3.3.1 Méthode simultanée classique

Le tableau 4.4 présente les performances de classification de la méthode proposée pour différentes sous-bases COREL. Nous fournissons également dans ce tableau les per-

3. Se référer au tableau 1.1 pour consulter les noms des différentes classes de la base COREL utilisée.
4. La classe *African peoples and villages* correspond à l'identifiant 6.

formances de classification de la méthode classique, dite aussi méthode simultanée (Kachouri et al., 2008b). La méthode de classification classique utilisée est basée sur un modèle SVM unique où l'ensemble des caractéristiques extraites est employé simultanément. Les meilleures performances sont affichées en gras dans le tableau 4.4.

TAB. 4.4 – *Évaluation des deux méthodes de classification MC-MM et classique pour différentes sous-bases d'images COREL.*

Sous-bases COREL		Performances de classification	
		Classique	**MC-MM**
3 classes	P_{moy}	0.7222	**0.9879**
	R_{moy}	0.4933	**0.9333**
5 classes	P_{moy}	0.7760	**0.9750**
	R_{moy}	0.3920	**0.9120**
7 classes	P_{moy}	0.7711	**0.9123**
	R_{moy}	0.3771	**0.8229**
10 classes	P_{moy}	0.7065	**0.8371**
	R_{moy}	0.3960	**0.8160**

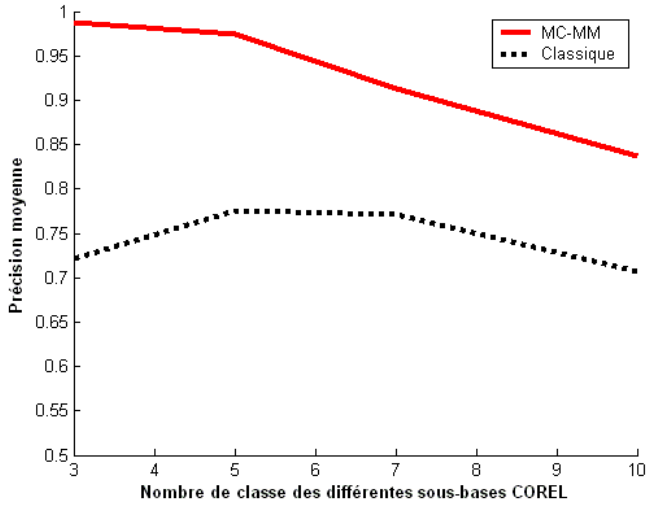


FIG. 4.4 – *Comparaison des précisions moyennes de MC-MM et de la méthode de classification classique pour différentes sous-bases d'images COREL.*

D'après les résultats présentés, il est clair que la classification MC-MM est beaucoup plus performante. Cette évaluation montre bien l'apport de la sélection adaptative des modèles pertinents ainsi que celui de l'emploi hiérarchique des modèles sélectionnés à travers la méthode MC-MM. Les précisions moyennes des deux méthodes de classification pour différentes sous-bases COREL sont présentées dans la figure 4.4. La performance de reconnaissance MC-MM décroît quand le nombre de classes augmente. On obtient par exemple des précision moyennes de 0.9879 pour 3 classes et 0.8371 pour 10 classes. Ceci est bien évident, en tenant compte de l'accroissement de la complexité du contenu des bases utilisées en fonction du nombre de classes. Par ailleurs, on remarque bien que l'utilisation hiérarchique des modèles SVMs ainsi que leur complémentarité avec les décisions du clasifieur NCC aboutissent toujours à des précisions de classification supérieurs à ceux

obtenus avec chaque modèle utilisé séparément (tableaux 4.4, 1.2 et 1.3). Par conséquent, l'utilisation hiérarchique des modèles sélectionnés par le classifieur MC-MM permet toujours d'améliorer la caractérisation des images et d'assurer par la suite une meilleure reconnaissance.

4.3.3.2 Approche des *Multiple Instance Learning*

Les méthodes connues par l'acronyme anglais MIL (*Multiple Instance Learning*) permettent d'apprendre un concept à partir d'une série d'ensemble d'exemples positifs et négatifs. Chaque ensemble peut contenir de nombreux cas, mais un ensemble est désigné positif si au moins l'un des exemples y contenu se situe dans le concept envisagé. Un ensemble est considéré comme négatif si tous ses exemples sont négatifs. L'approche MIL a été très bien adaptée aux contexte de la classification d'images. Dans ce cas, les ensembles sont les images de la base considérée et les exemples sont les différentes sous-régions de ces images.

Nous comparons les performances de classification MC-MM avec celles de différentes méthodes MIL présentes dans la littérature, à savoir les méthodes MILES[5] (Chen et al., 2006), DD-SVM[6] (Chen and Wang, 2004), MI-SVM (Andrews et al., 2003), et *K-means*-SVM (Csurka et al., 2004a). La méthode *K-means*-SVM construit des groupes de régions d'images en utilisant l'algorithme *K-means*. Dans cette méthode, les SVMs sont construites en utilisant comme attributs le nombre de régions désignant chaque classe. La méthode DD-SVM entraîne dans l'espace des attributs un classifieur SVM construit à partir d'une projection définie par les maximums et les minimums locaux de la fonction de Densité Diverse (DD). Sachant que la fonction DD mesure les co-occurrences des régions similaires dans les différentes images de même classe (Maron and pérez, 1998). La méthode MILES a été proposée comme une amélioration de DD-SVM, par l'emploi d'une sélection enveloppante des attributs performants. Effectivement, les SVMs sont appliqués simultanément pour la sélection ainsi que pour la construction du classifieur dans la méthode MILES. La méthode MI-SVM est basée sur la sélection des régions considérées comme positives pour la classification des images. Elle utilise le même ensemble d'attributs régions que DD-SVM. Le noyau gaussien est utilisé pour les SVMs dans toutes les méthodes comparées.

TAB. 4.5 – *Comparaison de la la précision moyenne de classification de la méthode MC-MM avec celles de différentes méthodes* MIL *de l'état de l'art pour la sous-base COREL avec* 10 *classes.*

Méthodes comparées	Performances
MC-MM	**83.7**
MILES (Chen et al., 2006)	82.6
DD-SVM (Chen and Wang, 2004)	81.5
MI-SVM (Andrews et al., 2003)	74.7
K-means-SVM (Csurka et al., 2004a)	**69.8**

5. Une implémentation MatLab de la méthode MILES est disponible en ligne à l'adresse suivante: http://www.cs.olemiss.edu/∼ychen/MILES.html.

6. Une implémentation MatLab de la méthode DD-SVM est disponible en ligne à l'adresse suivante: http://www.cs.uno.edu/∼yixin/ddsvm.html.

Les performances de classification de ces méthodes pour la base COREL avec 10 classes sont fournies dans le tableau 4.5[7]. La meilleure et la pire performances dans ce tableau sont affichées en gras. Cette analyse met en évidence la supériorité de la performance de notre méthode par rapport à celles des différentes méthodes comparées. MC-MM fournit une précision moyenne de classification de l'ordre de 83.7%. Par contre, la meilleure performance obtenue, parmi toutes les autres méthodes évaluées, est celle de la méthode MILES qui aboutit seulement à 82.6% comme précision moyenne de classification.

4.3.3.3 Approche des sacs de mots-visuels

Une comparaison de MC-MM avec la méthode sac des descripteurs (Csurka et al., 2004b), (Willamowski et al., 2004), (Zhang et al., 2007) désignée souvent par l'acronyme BoF (*Bag of Features*) est aussi effectuée. L'approche BoF consiste à évaluer un jeu de descripteurs visuels locaux sur un ensemble de points d'intérêt extraits de l'image. Elle permet ensuite de caractériser les statistiques de ces descripteurs locaux par le biais de leur histogramme de valeurs quantifiées (figure 4.5). Ceci permet de caractériser globalement le contenu de l'image.

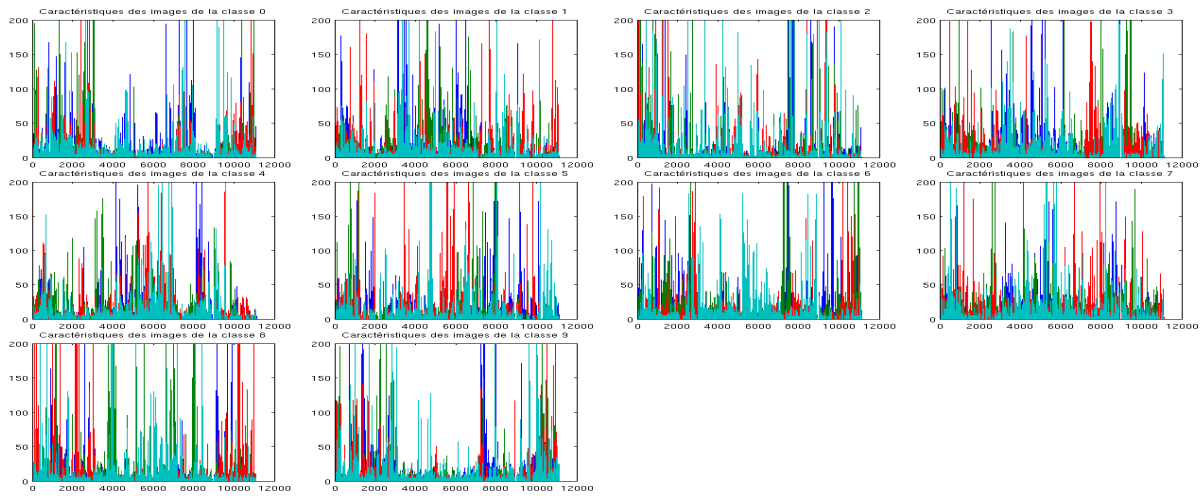


FIG. 4.5 – *Les dix histogrammes de mots-visuels qui représentent l'ensemble des classes de la sous-base d'images COREL avec* 10 *classes. Chaque histogramme illustre le nombre d'occurrence des différents mots-visuels extraits à partir des images de la classe correspondante.*

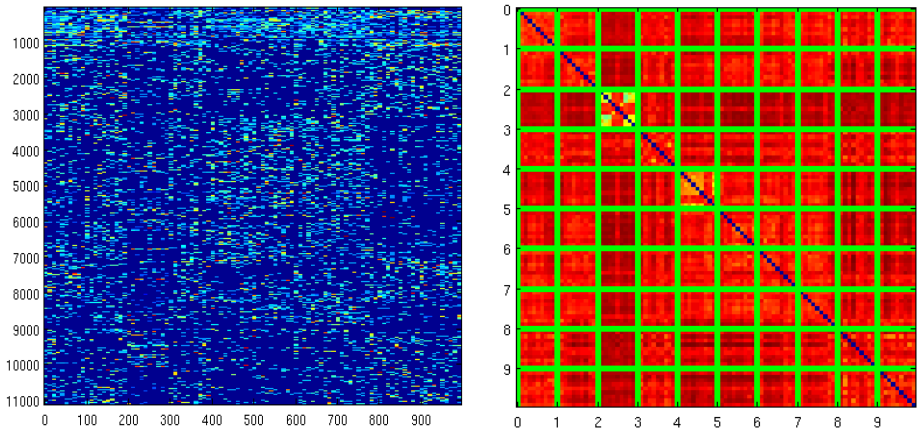
Une large évaluation dans l'état de l'art a bien prouvé les performances de la méthode sac des descripteurs pour la classification d'images par le contenu (Zhang et al., 2007). Le processus de la méthode BoF[8] employée dans ce travail est constitué essentiellement par trois étapes. La première est la détection et la description des points d'intérêt dans l'ensemble des images de la base via l'utilisation du descripteur SIFT (Lowe, 2004). La deuxième est le regroupement de tous les descripteurs extraits en différentes catégories appelées mots-visuels à l'aide de l'algorithme k-means. D'où, il est également possible d'appeler la méthode sac des descripteurs par sac de mots-visuels. Chaque catégorie trouvée représente une région typique des images. Lors de notre experimentation, le nombre totale de mots-visuels extraits de la sous-base COREL avec 10 classes est autour de 11000

7. Les valeurs affichées sont les précisions moyennes de classification (en pourcentage).

8. Une implémentation MatLab de la méthode sac des descripteurs (Bag of Features) est disponible en ligne à l'adresse suivante: http://www.vlfeat.org/ vedaldi/code/bag/bag.html.

mots-visuels. La figure 4.5 illustre les dix histogrammes de mots-visuels qui représentent l'ensemble des différentes classes de la base d'images COREL utilisée. Les images peuvent être alors décrites dans une troisième étape comme un sac de mots-visuels. Les occurrences de chaque mot-visuel dans une image représente la caractéristique de cette image.

Nous affichons dans la figure 4.6 (a) les caractéristiques histogramme de mots-visuels de toutes les images de la sous-base COREL utilisée. Une évaluation de la robustesse de description de ces caractéristiques est fournie dans la figure 4.6 (b). Cette figure illustre la distribution de distances par paires d'images qui présente l'organisation des distances entre les histogrammes de mots-visuels des différentes images dans la base. Le pouvoir discriminant de ces caractéristiques s'avère assez performant pour caractériser une forte similarité entre chaque image et elle-même (voir la diagonale de la matrice de distance illustrée par la figure 4.6 (b)). Le degré de similarité entre les images appartenant à une même classe est beaucoup moins significatif.



a) b)

FIG. 4.6 – *a) Sac de mots-visuels qui représente l'ensemble des images appartenant aux* 10 *classes de la sous-base COREL utilisée, b) Distances entre les caractéristiques histogrammes de mots-visuels des différentes paires d'images de la sous-bases COREL utilisée (*10 *classes).*

En utilisant cette description d'images basée sur les sacs de mots-visuels, on obtient une précision moyenne de classification de l'ordre de 74.0% avec la méthode BoF pour la sous-base COREL avec 10 classes. Par conséquent, en se basant sur la sélection adaptative des caractéristiques globales les plus pertinentes, la méthode de classification MC-MM proposée dans ce chapitre est beaucoup plus performante (précision moyenne de classification de l'ordre de 83.7% pour la même base).

Pour une meilleure illustration de la comparaison MC-MM et BoF, nous présentons dans la figure 4.7 les graphes des matrices de confusions correspondantes à ces deux méthodes. Il est clair d'après le graphe situé dans la partie gauche[9] de cette figure que la méthode de classification MC-MM est aussi performante pour toute les classes de la sous-base d'images COREL avec 10 classes. Alors que le graphe situé dans la partie droite de la même figure, décrivant la matrice de confusion de la méthode BoF, montre que cette méthode ne permet pas d'assurer des taux de classification homogènes pour les différentes classes de la sous-base. En effet, bien qu'elle se montre assez compétitive pour

9. La matrice de confusion correspondante au graphe situé dans la partie gauche de la figure 4.7 est fournie dans le tableau 4.3.

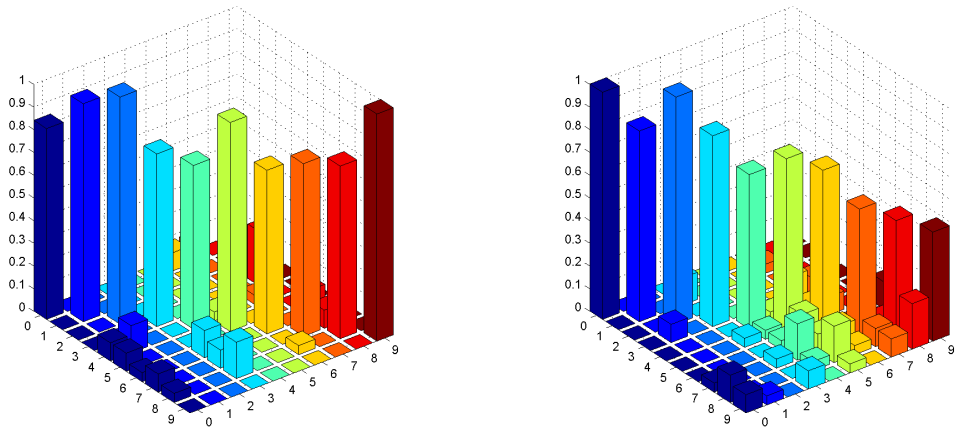


FIG. 4.7 – *Graphes des deux matrices de confusion respectives de MC-MM (à gauche) et BoF (à droite) pour la sous-base d'images COREL avec* 10 *classes.*

la classification des images appartenant aux sept premières classes de la sous-base COREL utilisées, la méthode BoF fournit des taux de classification nettement dégradés pour les trois classes d'identifiants 7, 8 et 9. Ceci est évident vu que les images hétérogènes appartenant sémantiquement au même concept ne présentent pas forcément une majorité de points d'intérêt communs. On affiche dans la figure 4.8 les courbes précision-rappel de MC-MM et BoF pour la même sous-base d'images COREL (10-classes).

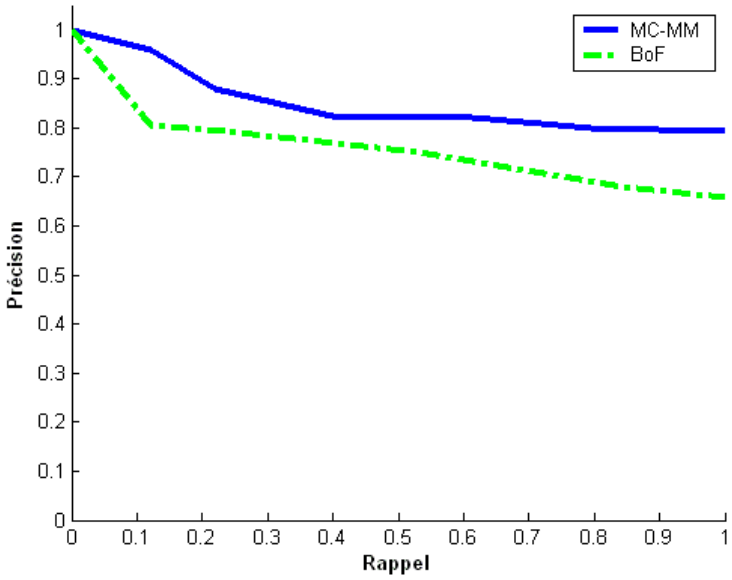


FIG. 4.8 – *Courbes précision-rappel de MC-MM et BoF pour la sous-base d'images COREL avec* 10 *classes.*

La figure 4.8 montre que la courbe précision-rappel de la méthode MC-MM est meilleure que celle de la méthode BoF. Les précisions de MC-MM sont toujours supérieures pour toutes les valeurs de rappel. On cite par exemple que pour un rappel de 100%, la précision de la méthode MC-MM est de l'ordre de 80%, alors que celle de BoF est de l'ordre de 67%.

4.3.3.4 Approche d'apprentissage par noyaux multiples

Nous avons également réalisé une comparaison des performances de la méthode MC-MM avec l'approche d'apprentissage par noyaux multiples (MKL). Dans ce contexte, MC-MM est comparée avec l'algorithme *SimpleMKL*[10] (Rakotomamonjy et al., 2008). Une évaluation des deux méthodes est effectuée pour neuf sous-bases d'images avec différents nombres de classes (3, 5, 7, 10, 13, 15, 17, 20 et 23) tirées des deux bases standards COREL et CALTECH-256. Plus de détails concernant ces deux bases d'images sont fournis dans le chapitre 1.

TAB. 4.6 – *Étude comparative des deux méthodes MC-MM et* SimpleMKL *pour différentes sous-bases d'images COREL et CALTECH-256.*

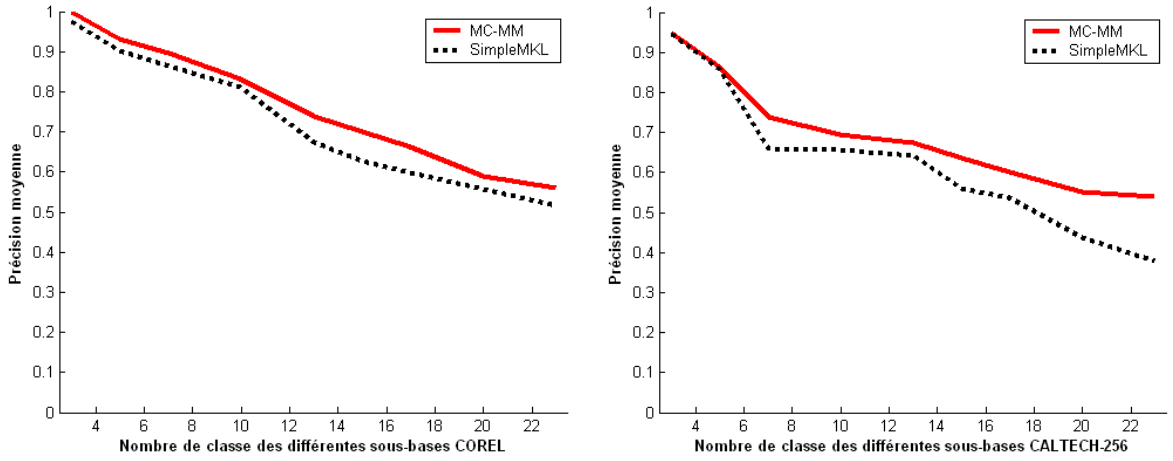
Sous-bases	COREL		CALTECH-256	
	SimpleMKL	**MC-MM**	*SimpleMKL*	**MC-MM**
3 classes	0.9753	**0.9879**	0.9467	**0.9487**
5 classes	0.9028	**0.9750**	0.8596	**0.8648**
7 classes	0.8640	**0.9123**	0.6600	**0.7370**
10 classes	0.8118	**0.8371**	0.6557	**0.6934**
13 classes	0.6750	**0.7380**	0.6423	**0.6743**
15 classes	0.6258	**0.7007**	0.5605	**0.6360**
17 classes	0.5974	**0.6629**	0.5360	**0.6018**
20 classes	0.5573	**0.5906**	0.4367	**0.5525**
23 classes	0.5157	**0.5590**	0.3773	**0.5392**

Le tableau 4.6 illustre les performances des deux méthodes comparées pour les différentes sous-bases d'images employées. Les meilleures performances dans ce tableau sont affichées en gras. D'après les résultats présentés, nous prouvons que la méthode de classification MC-MM proposée est toujours plus performante que *SimpleMKL* pour toutes les sous-bases d'images COREL et CALTECH-256 utilisées. Les courbes de précisions moyennes correspondantes sont présentées dans les figures 4.9 (a) et (b).

Pour les différentes sous-bases COREL utilisées, les courbes précision-rappel de MC-MM et *SimpleMKL* sont illustrées respectivement dans les figures 4.10 (a) et (b). Bien qu'elles aient presque les mêmes allures pour les sous-bases avec 3, 5, 7 et 10 classes, les courbes de précision-rappel de MC-MM pour ces sous-bases restent meilleures que celles de *SimpleMKL*. Pour les sous-bases restantes qui correspondent à 13, 15, 17, 20 et 23 classes, nous pouvons remarquer que la classification MC-MM est nettement plus pertinente que *SimpleMKL*.

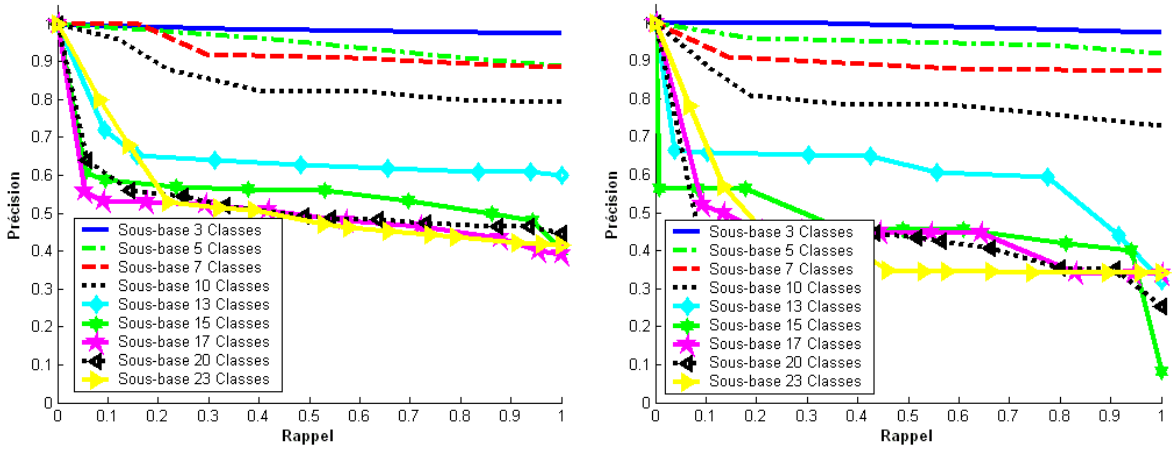
D'une manière similaire, les figures 4.11(a) et (b) illustrent respectivement les courbes précision-rappel de MC-MM et *SimpleMKL* pour différentes sous-bases d'images CALTECH-256. Il est clair que la qualité de classification des deux méthodes se dégrade proportionnellement par rapport au cas des sous-bases COREL. Ceci est principalement dû à la variation de la complexité du contenu des deux bases. Cependant, MC-MM est toujours plus performante que *SimpleMKL* pour toutes les sous-bases CALTECH-256 employées.

10. L'algorithme *SimpleMKL* est déjà présenté dans le chapitre 3.



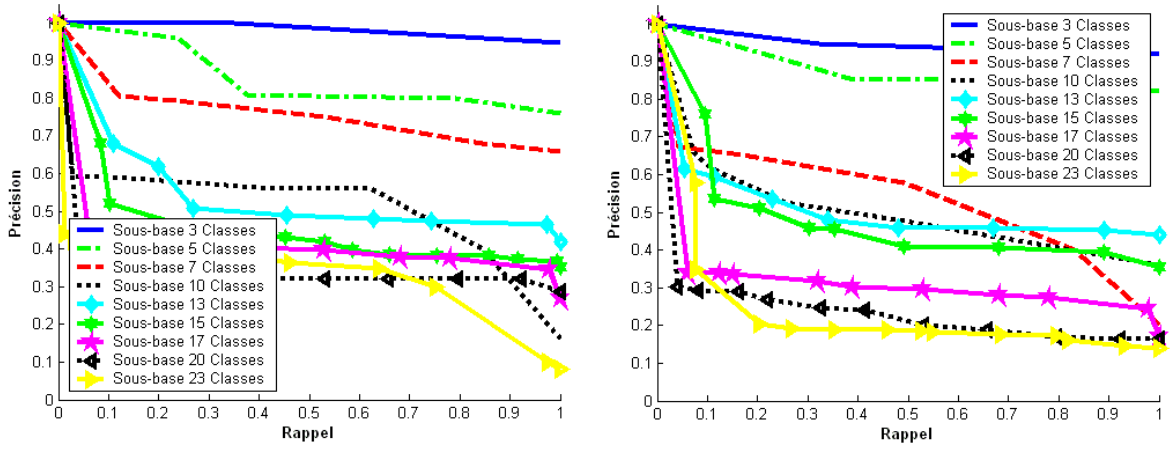
a) b)

FIG. 4.9 – *Comparaison des précisions moyennes de MC-MM et* SimpleMKL *pour différentes sous-bases d'images a) COREL et b) CALTECH-256.*



a) b)

FIG. 4.10 – *Courbes précision-rappel de a) MC-MM et b)* SimpleMKL *pour différentes sous-bases d'images COREL.*



a) b)

FIG. 4.11 – *Courbes précision-rappel de a) MC-MM et b)* SimpleMKL *pour différentes sous-bases d'images CALTECH-256.*

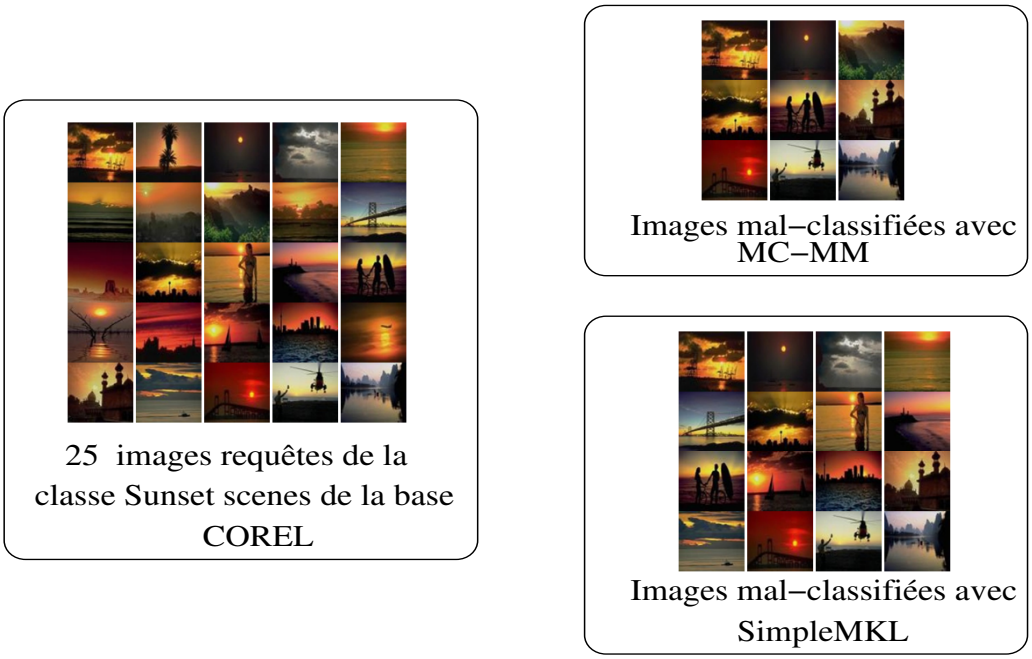


FIG. 4.12 – *Les images requêtes de la classe* Sunset scenes *de la base COREL (à gauche) et les images mal-classifiées (à droite) avec respectivement MC-MM et* SimpleMKL. *Toutes les images requêtes restantes (non illustrées à droite) sont correctement classifiées à travers chaque méthode.*



FIG. 4.13 – *Les images requêtes de la classe* Chimp *de la base CALTECH-256 (à gauche) et les images mal-classifiées (à droite) avec respectivement MC-MM et* SimpleMKL. *Toutes les images requêtes restantes (non illustrées à droite) sont correctement classifiées à travers chaque méthode.*

Deux exemples d'images requêtes mal-classifiées et correctement classifiées avec les deux méthodes MC-MM et *SimpleMKL* sont présentés dans les figures 4.12 et 4.13. Pour chaque figure, les 25 images requêtes utilisées sont affichées à gauche. Les images illustrées à droite sont les images mal-classifiées respectivement avec MC-MM et *SimpleMKL*. Dans chaque cas, toutes les images restantes, parmi les 25 requêtes de départ, sont correctement

classifiées à travers la méthode correspondante. La figure 4.12 montre que *SimpleMKL* n'arrive pas à classifier correctement 16 images à partir des 25 requêtes de la classe *Sunset scenes* de la base COREL (23 classes). Cependant, les images mal-classifiées de la même classe avec MC-MM sont uniquement 9. De la même façon, la figure 4.13 montre que les images mal-classifiées de la classe *Chimp* de la base CALTECH-256 sont respectivement 8 et 13 à travers les deux méthodes MC-MM et *SimpleMKL*.

TAB. 4.7 – *Comparaison des temps d'apprentissage (en Minutes) de* MC-MM *et* SimpleMKL *pour différentes sous-bases COREL et CALTECH-256.*

Sous-bases	COREL		CALTECH-256	
	SimpleMKL	**MC-MM**	*SimpleMKL*	**MC-MM**
3 classes	2.7084	**0.1985**	0.4350	**0.2819**
5 classes	7.7222	**0.7825**	0.7891	**0.7276**
7 classes	28.4550	**1.6156**	**1.1431**	1.7754
10 classes	37.8110	**3.7262**	5.0934	**4.0073**
13 classes	11.5164	**8.8968**	16.1382	**8.1480**
15 classes	26.1216	**12.1298**	125.4888	**11.7506**
17 classes	22.7951	**15.1858**	487.9302	**15.8950**
20 classes	37.1055	**22.7309**	612.0108	**24.2861**
23 classes	47.8590	**31.5707**	828.9810	**32.5798**

Par ailleurs, la méthode MC-MM est aussi plus efficace que *SimpleMKL* en terme de temps d'exécution. Le tableau 4.7 illustre les temps de calcul correspondants des apprentissages MC-MM et *SimpleMKL* pour les différentes sous-bases COREL et CALTECH-256 utilisées. Les meilleures performances sont affichées en gras dans ce tableau. On constate que le processus d'apprentissage MC-MM est beaucoup plus rapide que celui de *SimpleMKL*. De plus, pour les deux bases d'images COREL et CALTECH-256, le temps d'apprentissage MC-MM est linéairement proportionnel aux nombres d'images dans la base d'apprentissage. En effet, le temps d'apprentissage MC-MM est autour de 0.2 *mn* dans le cas des sous-bases avec 3 classes et autour de 32 *mn* dans le cas des sous-bases avec 23 classes. Cependant, le temps d'apprentissage *SimpleMKL* est aléatoire. En fait pour la même base COREL, le temps de calcul nécessaire pour l'apprentissage de la sous-base avec 10 classes est plus élevé que celui nécessaire pour l'apprentissage de la sous-base avec 17 classes. Dans le même contexte, *SimpleMKL* assure l'apprentissage de la sous-base CALTECH-256 avec 15 classes pendant 125.4888 *mn*. Par contre, il nécessite 26.1216 *mn* seulement pour l'apprentissage de la sous-base COREL avec 15 classes. En outre, *SimpleMKL* est basée sur un programme d'optimisation pour déterminer les poids de chaque noyaux. Par conséquent, il souffre généralement d'un temps de calcul important pour converger vers des coefficients optimaux. Ainsi, vu la complexité du traitement et de la classification des images dans les bases hétérogènes, il est généralement difficile de définir des contraintes adaptées permettant de déterminer les différents poids de chaque noyau. Effectivement, il s'avère que, pour les sous-bases CALTECH-256 avec 15, 17, 20 et 23 classes, *SimpleMKL* nécessite un grand nombre d'itérations pour converger vers des solutions raisonnables. Il a besoin, par exemple, de plus de treize heures pour l'apprentissage de la sous-base CALTECH-256 avec 23 classes.

4.4 Conclusion

Une nouvelle Méthode de Classification hiérarchique Multi-Modèles, intitulée MC-MM, est présentée dans ce chapitre. Cette méthode est essentiellement destinée pour les systèmes de reconnaissance d'images par le contenu dans les bases hétérogènes. MC-MM considère particulièrement les modèles correspondants aux caractéristiques sélectionnées par la méthode de sélection adaptative SAC, présentée dans le chapitre 2. Ces modèles sont employés hiérarchiquement selon l'ordre de leurs performances d'apprentissage d'une façon automatique et sans aucune rétroaction de l'utilisateur ou bouclage de pertinence. Pour toute image requête, le résultat de classification est affiné progressivement tout au long des niveaux hiérarchiques de la méthode MC-MM. Des résultats de classification similaires apportent une confirmation de décision des niveaux consécutifs. Par contre, l'ambiguïté de tout conflit de décision est levée grâce à l'utilisation du clasifieur NCC.

Une large évaluation de la méthode de classification MC-MM est assurée grâce aux experimentations effectuées pour les deux bases d'images de référence COREL et CALTECH-256. En comparaison avec des méthodes existantes dans la littérature, MC-MM présente de très bonnes performances de classification et génère des résultats meilleurs pour la classification de 23 classes thématiques.

Conclusion générale et perspectives

Dans ce travail de thèse, nous nous sommes intéressés à la reconnaissance d'images par le contenu dans les bases hétérogènes. Nous avons tout d'abord étudié de manière approfondie les éléments fondamentaux des systèmes CBIR, en réservant une attention particulière aux spécificités des bases d'images hétérogènes. La description des images, appartenant à ces bases, par des caractéristiques multiples est principalement prise en compte. Ainsi, une variété de caractéristiques bas niveaux de différents types a été utilisée. Pour améliorer les performances de ces caractéristiques, quelques optimisations ont été apportées. Nous avons prouvé, à l'aide d'une large évaluation pour des bases de référence, que le pouvoir discriminant de chaque caractéristique n'est guère absolu et que les performances de classification des différentes caractéristiques dépendent fortement du contenu de la base d'images considérée.

En se basant sur cette étude, nous avons décidé de nous placer dans le contexte de classification supervisée. Ceci permet d'assurer le contrôle des caractéristiques employées à travers la phase d'apprentissage associée. Dans ce travail, les classifieurs SVMs ont été choisis pour leur efficacité assez éprouvée. Cependant, nous nous sommes fixés principalement quatre critères que doit satisfaire l'approche de reconnaissance d'images hétérogènes développée. Ces critères sont les suivantes:

- Les caractéristiques retenues doivent être bien adaptées au contenu de la base d'images considérée.
- L'étape de reconnaissance doit être totalement automatique et n'inclure aucune rétroaction ou bouclage de pertinence.
- La méthode de classification doit assurer une meilleure souplesse dans la combinaison des caractéristiques hétérogènes.
- Le temps de réponse de la méthode doit être fortement réduit.

La nouvelle approche de reconnaissance d'images hétérogènes par le contenu, présentée dans ce travail de thèse, satisfait l'ensemble de ces critères. Elle permet de sélectionner un sous-ensemble des caractéristiques les plus pertinentes pour une base d'images donnée. Pendant la phase de généralisation, ces caractéristiques sont exploitées individuellement d'une façon hiérarchique et assez complémentaire. Les résultats de la classification finale des images ne se basent sur aucune rétroaction de l'utilisateur ou bouclage de pertinence. Dans ce cadre, deux originalités ont été proposées, à savoir une méthode de Sélection Adaptative des Caractéristiques les plus pertinentes, que l'on a appelée SAC et une Méthode de Classification hiérarchique Multi-Modèles, intitulée MC-MM.

En se basant sur le pouvoir discriminant de chaque caractéristique par rapport à une base base d'images donnée, une méthode originale et efficace pour la sélection adaptative des caractéristiques est développée. Suite à un apprentissage multiples basé sur l'ensemble des caractéristiques hétérogènes, la méthode de sélection SAC détermine un sous-ensemble des caractéristiques les plus pertinentes. Les caractéristiques sont sélectionnées

en se basant sur une discrimination Fisher de leurs taux d'apprentissage individuels. Cette sélection est totalement adaptée à la base d'images, les caractéristiques et le classifieur utilisés. D'où, elle permet d'améliorer considérablement les performances de description.

Les méthodes d'apprentissage par noyaux multiples, désignées par MKL, ont été appliquées avec sucés à la reconnaissance d'images. Cependant, elles souffrent généralement d'un temps de calcul très important. Dans ce contexte, nous avons proposé une nouvelle méthode de pondération (KW-BTR) destinée pour la classification SVM à noyaux multiples (MK_SVM). En se basant sur le degré de pertinence de chaque noyau, la nouvelle pondération permet de mieux considérer les noyaux les plus pertinents. Outre que l'amélioration des résultats de classification prouvée par les expérimentations réalisées, MK_SVM basée KW-BTR assure significativement des temps d'exécution inférieurs aux algorithmes MKL existants. Par ailleurs, face aux limites des méthodes MKL à travers la combinaison linéaire des caractéristiques hétérogènes, nous avons proposé la méthode de classification MC-MM. Avec cette classification, uniquement les modèles des caractéristiques sélectionnées via la méthode SAC sont considérés. Ces modèles sont employés hiérarchiquement suivant un ordre reflétant les performances d'apprentissage des caractéristiques correspondantes. Pour toute image requête le résultat de classification est affiné progressivement tout au long des différents niveaux hiérarchiques de la méthode. L'ordre d'utilisation des caractéristiques peut être ajusté selon une complémen-tarité de décision entre chaque deux niveaux consécutifs.

Pour évaluer l'approche de reconnaissance proposée, une large étude expérimentale de la méthode de sélection SAC ainsi que de la classification MC-MM est effectuée pour des base d'images de référence. Une étude comparative avec un ensemble de méthodes existantes dans la littérature est également réalisée. Cette étude a montré une très bonne fiabilité de notre approche. Effectivement, MC-MM basée SAC permet d'aboutir à des résultats de reconnaissance très efficaces. De plus, elle est beaucoup plus performante que les méthodes comparées. En terme de temps d'exécution, notre approche présente un très bon compromis entre la précision et la rapidité. Elle permet par exemple d'obtenir, pour différentes sous-bases CALTECH-256, des facteurs d'accélération allant de 1.5 jusqu'à 25 par rapport à l'algorithme *SimpleMKL* tout en assurant des précisions meilleures.

Il convient à noter que MC-MM ne peut pas être employée avec des systèmes mono-modèle basés sur la description d'une caractéristique unique. Ceci est dû essentiellement à l'hypothèse principale de la méthode (classification multi-modèles). En réalité, MC-MM est destinée aux systèmes basés sur une large description construite par un ensemble de caractéristiques multiples.

Les perspectives que nous envisageons dans le prolongement de ce travail de thèse s'articulent autour des points suivants.

– Description des images:

  Dans ce travail, nous avons pris en compte les propriétés visuelles de bas niveau pour la description des images. Comme nous l'avons déjà évoqué, le contenu assez complexe des images dans les bases hétérogènes engendre souvent des structures dissemblables pour des images du même concept. Comme perspective, nous pouvons intégrer de nouvelles caractéristiques afin d'enrichir davantage la description des images. La conception d'une nouvelle caractéristique assurant une description adaptée aux contenu des images hétérogènes pourrait être aussi envisageable. Ceci

permet d'améliorer encore les performances de la classification et de la reconnaissance.
Par ailleurs, pour enrichir davantage cette description bas niveau dans les bases hétéro-gènes, nous pensons à une stratégie à double niveaux de caractérisation des images. Ainsi, on peut améliorer les performances du système de reconnaissance proposé en utilisant une deuxième description d'aspect sémantique. Également, l'utilisation d'une combinaison de descriptions locales et globales peut être envisagée.

– Évaluation des caractéristiques:

Dans l'approche proposée, le pouvoir discriminant constituant le critère d'évaluation des différentes caractéristiques employées est calculé à travers le processus d'apprentissage. En effet, la sélection adaptative proposée ainsi que l'hiérarchie du classifieur MC-MM dépendent essentiellement des taux d'apprentissage de chaque caractéristique. Nous pensons que les performances calculées pour les différentes caractéristiques reflètent certainement une précision de la qualité de description de ces caractéristiques, mais elles comprennent aussi un certain effet du classifieur utilisé. L'idée est alors de reconsidérer l'évaluation des caractéristiques tout en réduisant la spécificité du classifieur voire même en l'éliminant. Une solution pour réduire l'impact du classifieur consiste à effectuer plusieurs apprentissage pour l'évaluation de chaque caractéristique. Différents classifieurs seront employés pour l'apprentissage (l'évaluation). Le pouvoir discriminant de chaque caractéristique sera la moyenne de l'ensemble des différentes performances d'apprentissage calculées. Toutefois, la détermination d'une nouvelle méthode permettant de mesurer la qualité de la description des différentes caractéristiques indépendamment des méthodes de classification reste une piste à explorer.

– Passage à l'échelle:

Les expérimentations effectuées dans ce travail ont montré l'efficacité de notre approche de reconnaissance d'images hétérogènes par le contenu. Cependant faute de ressources mémoire, ces expérimentation ont été réalisées sur des bases d'images de taille relativement modeste (jusqu'à 2300 images) comparativement aux bases réelles dont le nombre d'images se chiffre à plusieurs millions. Nous envisageons, par conséquent, d'étendre ces expérimentations à des bases d'images de plus grande taille. Dans un tel contexte, de nouveaux problèmes en termes de reconnaissance et de temps d'exécution risquent d'apparaître.

Liste des publications

Articles dans des revues internationales avec comités de lecture (2)

- Rostom Kachouri, Khalifa Djemal et Hichem Maaref, '*Multi-Model Classification Method in Heterogeneous Image Databases*', Accepté pour publication au journal international Pattern Recognition, ELSEVIER, 2010.

- Rostom Kachouri, Khalifa Djemal et Hichem Maaref, '*Multiple Kernel based SVM for heterogeneous image recognition system*', Soumis au journal international Signal and Imaging Systems Engineering, INDERSCIENCE ENTERPRISES.

Article dans des revues nationales avec comités de lecture (1)

- Rostom Kachouri, Khalifa Djemal et Hichem Maaref, '*Sélection adaptative de caractéristiques pertinentes et classification hiérarchique multi-modèles pour la reconnaissance d'images dans les bases hétérogènes*', En cours de révision à la revue scientifique francophone TRAITEMENT DU SIGNAL.

Chapitre de livre (1)

- Khalifa Djemal, Hichem Maaref et Rostom Kachouri (2009). Image Retrieval System in Heterogeneous Database, AUTOMATION CONTROL-Theory and Practice, A D Rodic (Ed.), ISBN: 978-953-307-039-1, INTECH, disponible à l'adresse suivante: http://sciyo.com/articles/show/title/image-retrieval-system-in-heterogeneous-database.

Communications internationales avec actes (7)

- Rostom Kachouri, Khalifa Djemal et Hichem Maaref, '*Adaptive Feature Selection for Heterogeneous Image Databases*', dans le Proceedings du Second IEEE International Conference on Image Processing Theory, Tools & Applications, IPTA'10, 07-10 July 2010, Paris, France .

- Rostom Kachouri, Khalifa Djemal, Hichem Maaref, Dorra Sellami Masmoudi et Nabil Derbel, '*Feature extraction and relevance evaluation for heterogeneous image database recognition*', dans le Proceedings du First IEEE International Workshops on Image Processing Theory, Tools & Applications, IPTA'08, 24-26 November, 2008, Sousse, Tunisia.

- Rostom Kachouri, Khalifa Djemal, Dorra Sellami Masmoudi et Nabil Derbel, '*Content based Image Recognition based on QUIP-tree Model*', dans le Proceedings du Fifthh IEEE International Multi-Conference on Systems, Signals & Devices, SSD'O8, 20-23 July 2008, Amman, Jordan.

- Rostom Kachouri, Khalifa Djemal, Hichem Maaref, Dorra Sellami Masmoudi et Nabil Derbel, '*Heterogeneous image retrieval system based on features extraction and SVM classifier*', dans le Proceedings du Fifthh IEEE International Conference on Informatics in Control, Automation and Robotics, ICINCO'08, 11-15 May 2008, Funchal, Madeira, Portugal.

- Rostom Kachouri, Khalifa Djemal, Hichem Maaref, Dorra Sellami Masmoudi et Nabil Derbel, '*Content description and classification for Image recognition system*', dans le Proceedings du Third IEEE International Conference on Information and Communication Technologies: from theory to applications, ICTTA'08, 07-11 April 2008, Damascus, Syria.

- Rostom Kachouri, Khalifa Djemal, Dorra Sellami Masmoudi, Hichem Maaref et Nabil Derbel, '*On the heterogeneous image retrieval with QUIP-tree*', dans le Proceedings du Fourth IEEE International Multi-Conference on Systems, Signals & Devices, SSD'07, 19-22 March 2007, Hammamet, Tunisia.

- Imène Cheikhrouhou, Rostom Kachouri, Khalifa Djemal, Dorra Sellami Masmoudi, Emna Daoud, Zeineb Mnif, Hichem Maaref et Nabil Derbel, '*MLP Neural Network Classifier for breast cancer diagnostic*', dans le Proceedings du Fourth IEEE International Multi-Conference on Systems, Signals & Devices, SSD'07, 19-22 March 2007, Hammamet, Tunisia.

Bibliographie

Addison, J. and Macintyre, J. (2003). Intelligent techniques: A review. *Springer Verlag*.

Aly, M., Welinder, P., Munich, M., and Perona, P. (2009). Automatic discovery of image families: Global vs. local features. In *International Conference on Image Processing ICIP'09*.

Anderberg, M. (1973). *Cluster Analysis for Applications (Probability & Mathematical Statistics Monograph)*. 0120576503. Academic Press Inc., New York.

Andrews, S., Tsochantaridis, I., and Hofmann, T. (2003). Support vector machines for multiple-instance learning. In *Advances in Neural Information Processing Systems*, pages 561–568.

Androutsos, P., Kushki, A., Plataniotis, K., and Venetsanopoulos, A. (2005). Aggregation of color and shape features for hybrid query generation in content based visual information retrieval. In *Signal Processing*, volume 7, pages 385–393.

Bach, F., Lanckriet, G., and Jordan, M. (2004). Multiple kernel learning, conic duality, and the smo algorithm. In Brodley, C., editor, *Twenty-first International Conference on Machine learning, ACM*.

Bach, J., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., and Shu, C. (1996). *The Virage image search engine : an open framework for image management*. Number 2670. Storage and Retrieval for Image and Video Databases.

Ben-Ismail, M., Frigui, H., and Caudill, J. (2008). Empirical comparison of automatic image annotation systems. In *First International Workshops on Image Processing Theory, Tools & Applications IPTA'08*.

Berry, M. (2003). *Survey of Text Mining: Clustering, Classification, and Retrieval*. 0387955631. Springer-Verlag, New York.

Bi, J., Bennett, K., Embrechts, M., Breneman, C., and Song, M. (2003). Dimensionality reduction via sparse support vector machines. *Journal of Machine Learning Research*, 3:1229–1243.

Biederman, I. (1982). Do background depth gradients facilitate object identification? *Perception*, 10:573–578.

Bimbo, A. (2001). Visual information retrieval. *Morgan Kaufmann Publishers, San Francisco, USA*.

Bimbo, A. and Vicario, E. (1998). Using weighted spatial relationships in retrieval by visual contents. In *IEEE Workshop on Image and Video Libraries*.

Boujemaa, N., Fauqueur, J., Ferecatu, M., Fleuret, F., Gouet, V., Saux, B., and Sahbi, H. (2001). Ikona : Interactive generic and specific image retrieval. In *International Workshop on Multimedia Content-Based Indexing and Retrieval MMCBIR'01*.

Brunelli, R. and Mich, O. (2000). Compass : an image retrieval system for distributed databases. In *IEEE International Conference on Multimedia and Expo ICME'00*.

Cao, L., Luo, J., Liang, F., and Huang, T. (2009). Heterogeneous feature machines for visual recognition. In *IEEE International Conference on Computer Vision ICCV'09*.

Carson, C., Thomas, M., Belongie, S., Hellerstein, J., and Malik, J. (1999). Blobworld : A system for region-based image indexing and retrieval. In *International Conference on Visual Information System VIS'99*.

Cheikhrouhou, I., Djemal, K., Masmoudi, D., Maaref, H., and Derbel, N. (2009). Empirical descriptors evaluation for mass malignity recognition. In *The First International Workshop on Medical Image Analysis and Description for Diagnosis Systems MIAD'09*.

Chen, V. and Ruan, S. (2010). Graph cut based segmentation of brain tumor from mri images. *International Journal on Sciences and Techniques of Automatic control & computer engineering, to appear*.

Chen, Y., Bi, J., and Wang, J. (2006). Miles: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1931–1947.

Chen, Y. and Wang, J. (2004). Image categorization by learning and reasoning with regions. *Journal of Machine Learning Research*, 5:913–939.

Csurka, G., Bray, C., Dance, C., and Fan, L. (2004a). Visual categorization with bags of keypoints. In *Proceedings Workshop Statistical Learning in Computer Vision ECCV'04*, pages 59–74.

Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004b). Visual categorization with bags of keypoints. In *ECCV Workshop on Statistical Learning in Computer Vision*.

Datta, R., Joshi, D., Li, J., and Wang, J. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 2(40).

Delingette, H. and Montagnat, J. (2001). Shape and topology constraints on parametric active contours. *Computer Vision and Image Understanding*, 83(2):140–171.

Dengsheng, Z. and Guojun, L. (2003). Evaluation of similarity measurement for image retrieval. *Neural Networks and Signal Processing*, 2:928–931.

Duan, K., Sathiya, S., and Poo, A. (2003). Evaluation of simple performance measures for tuning the svm hyperparameters. *Neurocomputing*, 51:41–59.

Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Han, J., editor, *Second International Conference on Knowledge Discovery and Data Mining*.

Faloutsos, C., Equitz, W., Flickner, M., Niblack, W., Petkovic, D., and Barber, R. (1994). Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(4):231–262.

Fei-Fei, L., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:594–611.

Fisher, R. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and video content : the qbic system. *IEEE Computer*, 28(9):23–32.

Freeman, J. and Skapura, D. (1991). Neural networks algorithms, applications, and programming techniques. *Reading, MA: Addison-Wesley*.

Friedman, J. (1997). Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University.

Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. 0122698509. USSR, Russian, Nauka, Moscow.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. second ed. Academic Press.

Gehler, P. and Nowozin, S. (2009). On feature combination for multiclass object classification. In *Twelfth IEEE International Conference on Computer Vision*, pages 1–8.

Gevers, T. and Smeulders, A. (1999). The pictoseek www image search system. In *International Conference on Mathematics and Computer Science*, volume 1, pages 264–269.

Gevers, T. and Smeulders, A. (2004). Content-based image retrieval: An overview. In Medioni, G. and Kang, S., editors, *Emerging Topics in Computer Vision, Prentice Hall*.

Glashoff, K. and Gustafson, S. (1983). Linear optimization and approximation. *Springer-Verlag*.

Goldberg, D. (1991). *Algorithmes Génétiques*. Addison-Wesley.

Gouet, V. and Boujemaa, N. (2001). Object-based queries using color points of interest. In *IEEE Workshop on Content-Based Access of Image and Video Libraries CBAIVL'01*.

Griffin, G., Holub, A., and Perona, P. (2007). Caltech 256 object category dataset. Technical Report UCB/CSD 04-1366, California Institute of Technology.

Guyon, I. and Elisseeff, A. (2003). An introduction to feature and variable selection. *Journal of Machine Learning Research*, 3:1157–1182.

Hafner, J., Sawhney, H., Equitz, W., Flickner, M., and Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(17):729–736.

Han, J. and Kamber, M. (2006). *Data Mining: Concepts and Techniques*. 1558609016. Morgan Kaufmann, second revised edition.

Haralick, R., Dinstein, I., and Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621.

Hare, J., Sinclair, P., Lewis, P., Martinez, K., Enser, P., and Sandom, C. (2006). Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In *Proceedings of Mastering the Gap: From Information Extraction to Semantic Representation, 3rd European Semantic Web Conference*.

He, D. and Cercone, N. (2009). Local triplet pattern for content-based image retrieval. *Springer Verlag*, (5627):229–238.

Hotelling, H. (1933). Analysis of complex statistical variables into principal components. *Journal of Educational Psychologiy*, 24:417–441.

Hu, M. (1962). Visual pattern recognition by moment invariants. *IEEE Transactions on information Theory*, 8:179–187.

Huang, J., Kumar, S., Mitra, M., and Zhu, W. (1999). Spatial color indexing and applications. In *International Conference on Computer Vision*, volume 3, pages 245–268.

Huang, J., Kumar, S., Mitra, M., Zhu, W., and Zabih, R. (1997). Image indexing using color correlograms. In *Computer Vision and Pattern Recognition*, pages 762–768.

IEEE-Press, editor (1990). *IEEE Standard Glossary of image Processing and Pattern Recognition Terminology*, New York.

Jain, A., Murty, M., and Flynn, P. (1999). Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323.

Jeon, J., Lavrenko, V., and Manmatha, R. (2003). Automatic image annotation and retrieval using crossmedia relevance models. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR'03*.

Jlassi, H. and Hamrouni, K. (2010). Detection of blood vessels in retinal images. *International Journal on Image Graphics*, 10(1):57–72.

John, G., Kohavi, R., and Pfleger, K. (2000). Irrelevant features and the subset selection problem. In *International Conference on Machine Learning*.

Jolliffe, I. (1986). Principal component analysis. *Springer Verlag*.

Jurgen, F. (2004). *Spatial Modeling in Natural Sciences and Engineering: Software Development and Implementation*. 3540208771. Springer-Verlag Berlin and Heidelberg GmbH. and Co. K., Germany, first edition.

Kachouri, R., Djemal, K., Maaref, H., Masmoudi, D., and Derbel, N. (2008a). Feature extraction and relevance evaluation for heterogeneous image database recognition. In *First International Workshops on Image Processing Theory, Tools & Applications IPTA'08*.

Kachouri, R., Djemal, K., Maaref, H., Masmoudi, D., and Derbel, N. (2008b). Heterogeneous image retrieval system based on features extraction and svm classifier. In *International Conference on Informatics in Control, Automation and Robotics ICINCO'08*, pages 137–142.

Kilho, S. (2007). *Position-Aware String Kernels with Weighted Shifts and a General Framework to Apply String Kernels to Other Structured Data*, chapter Intelligent Data Engineering and Automated Learning, pages 316–325. Springer-Verlag Berlin Heidelberg.

Kim, T. and Cipolla, R. (2008). Mcboost: Multiple classifier boosting for perceptual co-clustering of images and visual features. In *Twenty-Second Annual Conference on Neural Information Processing Systems NIPS'08*, pages 841–856.

Kimura, F., Wakabayashi, T., Tsuruoka, S., and Miyake, Y. (1997). Improvement of handwritten japanese character recognition using weighted direction code histogram. *Journal of the Pattern Recognition Society*.

Kohavi, R. and John, G. (1997). Wrappers for features subset selection. *Artificial Intelligence*, 97:273–324.

Kudo, M. and Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25–41.

Lanckriet, G., Bie, T., Cristianini, N., Jordan, M., and Noble, W. (2004a). A statistical framework for genomic data fusion. *Bioinformatics*, 20:2626–2635.

Lanckriet, G., Cristianini, N., Bartlett, P., El-Ghaoui, L., and Jordan, M. (2004b). Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5:27–72.

Langley, P. and Sage, S. (1999). Tractable average-case analusis of naive bayesian classifiers. In Kaufman, M., editor, *Sixteenth International Conference on Machine Learning*, pages 220–228.

Leslie, C., Eskin, E., and Noble, W. (2002). The spectrum kernel: a string kernel for svm protein classification. In *7th Pacific Symposium of Biocomputing*.

Li, X., Wu, C., Zach, C., Lazebnik, S., and Frahm, J. (2008). Modeling and recognition of landmark image collections using iconic scene graphs. In *European Conference on Computer Vision ECCV'08*.

Livens, S., Scheunders, P., Van-Wouwer, G., and Van-Dyck, D. (1997). Wavelets for texture analysis, an overview. In *Image Processing and Its Applications*, volume 2, pages 581–585.

Lodhi, H., Shawe-Taylor, J., Cristianini, N., and Watkins, C. (2001). Text classification using string kernels. In *Advances in Neural Information Processing Systems NIPS'01*, volume 13.

Loève, M. (1948). *Fonctions aléatoires du second ordre*. Supplement to P. Levy.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2(60):91–110.

Ma, W. and Manjunath, B. (1999). Netra : A toolbox for navigating large image databases. In *Multimedia Systems*, volume 7, pages 184–198.

Malki, J., Boujemaa, N., Nastar, C., and Winter, A. (1999). Region queries without segmentation for image retrieval by content. In *International Conference on Visual Information System VIS'99*, pages 115–122.

Manjunath, B., Ohm, J., Vasudevan, V., and Yamada, A. (2001). Color and texture descriptors. In *IEEE Transactions on circuits and systems for video technology*, volume 11, pages 703–715.

Maren, A. (1990). *Handbook of Neural Computing Applications*. 0125460902. Academic Press, Inc.

Maron, O. and pérez, T. (1998). A framework for multiple-instance learning. In *Advances in Neural Information Processing Systems NIPS'98*, volume 10, pages 570–576.

McQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Cam, L. and Neyman, J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*.

Mikolajczyk, K. and Schmid, C. (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, (60):63–86.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Gool, L. (2005). A comparison of affinene region detectors. *International Journal of Computer Vision*, 2(65):43–72.

Moghaddam, B., Biermann, H., and Margaritis, D. (1999). Defining image content with multiple regions of interest. In *IEEE Workshop on Content-Based Access of Image and Video Libraries CBAIVL'99*.

Muller, H., Michoux, N., Bandon, D., and Geissabuhler, A. (2004). A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23.

Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175.

Oyang, Y., Hwang, S., Ou, Y., Chen, C., and Chen, Z. (2005). Data classification with radial basis function networks based on a novel kernel density estimation algorithm. *IEEE Transactions on Neural Networks*, 16:225–236.

Pascale, D. (2003). A review of rgb color spaces...from xyy to r'g'b'. *The BabelColor company*.

Pentland, A., Picard, R., and Sclaroff, S. (1994). *Photobook: Content-based Manipulation of Image Databases*, volume 2. Storage and Retrieval for Image and Video Databases.

Platt, J., Cristianini, N., and Shawe-Taylor, J. (2000). Large margin dagsvm's for multi-class classification. In *Advances in Neural Information Processing System NIPS'00*, volume 12, pages 547–553.

Portnoy, D., Bellaachia, A., Chen, Y., and Elkhahloun, A. (2002). E-cast: A data mining algorithm for gene expression data. In Zaki, M., Wang, J., and Toivonen, H., editors, *Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'02, Workshop on Data Mining in Bioinformatics (BIOKDD)*.

Potter, M. (1975). Meaning in visual search. *Science*, 187(4180):965–966.

Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1987). *Numerical Recipes*. The Art of Scientific Computing.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Quinlan, J. and Rivest, R. (1989). Inferring decision trees using the minimum description length principle. *Inform. Comput.*, 80(3):227–248. An early version appeared as MIT LCS Technical report MIT/LCS/TM-339 (September 1987).

Rakotomamonjy, A., Bach, F., Canu, S., and Grandvalet, Y. (2007). More efficiency in multiple kernel learning. In *International Conference on Machine Learning*, volume 27, pages 775–782.

Rakotomamonjy, A., Bach, F., Canu, S., and Grandvalet, Y. (2008). Simplemkl. *Journal of Machine Learning Research*, 2:2491–2521.

Rätsch, G., Sonnenburg, S., and Schölkopf, B. (2005). Rase: recognition of alternatively spliced exons in c.elegans. In *Bioinformatics*.

Rebaine, D. (2005). *Methode de branch and bound*. Université du Quebec à Chicoutimi.

Renniger, L. and Malik, J. (2004). When is scene identication just texture recognition? *Journal of Vision Research*, 44:2301–2311.

Ritendra, D., Dhiraj, J., Jia, L., and Wang, J. (2006). Studying aesthetics in photographic images using a computational approach. In *Lecture Notes in Computer Science, Proceedings of the European Conference on Computer Vision*, volume 3, pages 288–301.

Rowshan, B., Bin-Khalid, M., and Yusof, R. (2008). Multi-level fuzzy score fusion for client specific linear discriminant analysis based face authentication system. In *IEEE International Conference Signal Image Technology and Internet Based Systems, SITIS'08*, pages 213–220.

Schmid, C., Mohr, R., and Bauckhage, C. (1998). Comparing and evaluating interest points. In *International Conference on Computer Vision ICCV'98*.

Sclaroff, S., Taycher, L., and Cascia, M. (1997). Imagerover : A content-based image browser for the world wide web. In *IEEE Workshop on Content-based Access of Image and Video Libraries CBAIVL'97*.

Serrano, N., Savakisb, A., and Luoc, J. (2004). Improved scene classification using efficient low-level features and semantic cues. *Pattern Recognition*, 37:1773–1784.

Shamir, R. and Sharan, R. (2002). Algorithmic approaches to clustering gene expression data. *Current Topics in Computational Molecular Biology MIT Press*.

Siagian, C. and Itti, L. (2007). Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):300–312.

Siedlecki, W. and Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, pages 335–347.

Sonnenburg, S., Rätsch, G., and Schäfer, C. (2005). A general and efficient algorithm for multiple kernel learning. In *Advances in Neural Information Processing Systems NIPS'05*, volume 17, pages 1–8.

Sonnenburg, S., Rätsch, G., Schäfer, C., and Schölkopf, B. (2006). Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7:1531–1565.

Sridhar, V., Nascimento, M., and Li, X. (2002). Region-based image retrieval using multiple-features. In *International Conference on Visual Information System VIS'02*.

Stanfill, C. (1987). Memory-based reasoning applied to english pronunciation. *Association for the Advancement of Artificial Intelligence AAAI*, pages 577–581.

Stehling, R., Nascimento, M., and Falcao, A. (2002). Microm : A metric distance to compare segmented images. In *International Conference on Visual Information System VIS'02*.

Stork, D., Duda, R., and Hart, P. (2001). *Pattern Classification*. 9755031030. John Wiley and Sons Inc., New York, second edition.

Stricker, M. and Swain, M. (1994). The capacity of color histogram indexing. In *Conference on Computer Vision and Pattern Recognition*.

Suard, F., Rakotomamonjy, A., and Bensrhair, A. (2007). Model selection in pedestrian detection using multiple kernel learning. In *Intelligent Vehicle Symposium*, pages 13–14.

Teague, M. (1980). Image analysis via the general theory of moments. *Optical Society of America*, 70:920–930.

Thomas, P. (2000). Automatic choice of dimensionality for pca. Technical report, Advances in Neural Information Processing System NIPS'00.

Tieu, K. and Viola, P. (2004). Boosting image retrieval. *International Journal on Computer Vision*, 56(2):17–36.

Torralba, A., Murphy, K., Freeman, W., and Rubin, M. (2003). Context-based vision system for place and object recognition. In *IEEE International Conference on Computer Vision ICCV'03*, pages 1023–1029.

Tversky, B. and Hemenway, K. (1983). Categories of the environmental scenes. *Cognitive Psychology*, 15:121–149.

Vapnik, V. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10:988–999.

Varma, M. and Ray, D. (2007). Learning the discriminative powerinvariance trade-off. In *IEEE International Conference on Computer Vision*.

Vriesenga, M. (1995). *Genetic Selection and Neureal Modeling for Designing Pattern Classifier*. PhD thesis, University of California, Irvine.

Wang, J., Li, J., and Wiederhold, G. (2001). Simplicity : Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Wang, L., Zhang, Y., and Feng, J. (2005). On the euclidean distance of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1334–1339.

Weiss, S. (1991). Small sample error rate estimation for k-nn classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):285–289.

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., and Vapnik, V. (2000). Feature selection for svms. In *Advances in Neural Information Processing Systems NIPS'00*.

Weston, J. and Watkins, C. (1999). Multi-class support vector machines. In *European Symposium on Artificial Neural Networks ESANN'99*.

Willamowski, J., Arregui, D., Csurka, G., Dance, C., and Fan, L. (2004). Categorizing nine visual classes using local appearance descriptors. In *ICPR Workshop on Learning for Adaptable Visual Systems*.

Xu, P. and Chan, A. (2003). Support vector machine for multi-class signal classification with unbalanced samples. In *International Joint Conference on Neural Networks*, pages 1116–1119.

Yu, L. and Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, 5:1205–1224.

Zhang, J., Marszalek, M., Lazebnik, S., and Schmid, C. (2007). Local features and kernels for classification of texture and object categories: A comprehensive study. *International Journal of Computer Vision*.

Zhu, J., Rosset, S., Hastie, T., and Tibshirani, R. (2004). 1-norm support vector machines. In *Advances in Neural Information Processing Systems NIPS'04*, volume 16, pages 49–56.

Zhuo, L., Zheng, J., Wang, F., Li, X., Ai, B., and Qian, J. (2008). A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII(B7):397–402.

Zien, A., Rätsch, G., Mika, S., Schölkopf, B., Lengauer, T., and Müller, K. (2000). Learning the kernel matrix with semidefinite programming. *Engineering support vector machine kernels that recognize translation initiation sites*, 16:799–807.