Université de Montréal

Consensus ou fusion de segmentation pour quelques applications de détection ou de classification en imagerie

par

Aymen Khlif

Département d'informatique et de recherche opérationnelle
Faculté des arts et des sciences

Thèse présentée à la Faculté des arts et des sciences
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en Informatique

Mai, 2018

RÉSUMÉ

Récemment, des vraies mesures de distances, au sens d'un certain critère (et possédant de bonnes propriétés asymptotiques) ont été introduites entre des résultats de partitionnement (*clustering*) de donnés, quelquefois indexées spatialement comme le sont les images segmentées. À partir de ces métriques, le principe de segmentation moyenne (ou *consensus*) a été proposée en traitement d'images, comme étant la solution d'un problème d'optimisation et une façon simple et efficace d'améliorer le résultat final de segmentation ou de classification obtenues en moyennant (ou fusionnant) différentes segmentations de la même scène estimée grossièrement à partir de plusieurs algorithmes de segmentation simples (ou identiques mais utilisant différents paramètres internes). Ce principe qui peut se concevoir comme un débruitage de données d'abstraction élevée, s'est avéré récemment une alternative efficace et très parallélisable, comparativement aux méthodes utilisant des modèles de segmentation toujours plus complexes et plus coûteux en temps de calcul.

Le principe de distance entre segmentations et de moyennage ou fusion de segmentations peut être exploité, directement ou facilement adapté, par tous les algorithmes ou les méthodes utilisées en imagerie numérique où les données peuvent en fait se substituer à des images segmentées. Cette thèse a pour but de démontrer cette assertion et de présenter différentes applications originales dans des domaines comme la visualisation et l'indexation dans les grandes bases d'images au sens du contenu segmenté de chaque image, et non plus au sens habituel de la couleur et de la texture, le traitement d'images pour améliorer sensiblement et facilement la performance des méthodes de détection du mouvement dans une séquence d'images ou finalement en analyse et classification d'images médicales avec une application permettant la détection automatique et la quantification de la maladie d'Alzheimer à partir d'images par résonance magnétique du cerveau.

Mots clefs : visualisation des bases d'images, indexation, détection de mouvements, classification/quantification de la maladie d'Alzheimer, fusion/moyennage/consensus de segmentations.

ABSTRACT

Recently, some true metrics in a criterion sense (with good asymptotic properties) were introduced between data partitions (or clusterings) even for data spatially ordered such as image segmentations. From these metrics, the notion of average clustering (or *consensus* segmentation) was then proposed in image processing as the solution of an optimization problem and a simple and effective way to improve the final result of segmentation or classification obtained by averaging (or fusing) different segmentations of the same scene which are roughly estimated from several simple segmentation models (or obtained with the same model but with different internal parameters). This principle, which can be conceived as a denoising of high abstraction data, has recently proved to be an effective and very parallelizable alternative, compared to methods using ever more complex and time-consuming segmentation models.

The principle of distance between segmentations, and averaging of segmentations, in a criterion sense, can be exploited, directly or easily adapted, by all the algorithms or methods used in digital imaging where data can in fact be substituted to segmented images. This thesis proposal aims at demonstrating this assertion and to present different original applications in various fields in digital imagery such as the visualization and the indexation in the image databases, in the sense of the segmented contents of each image, and no longer in the common color and texture sense, or in image processing in order to sensibly and easily improve the detection of movement in the image sequence or finally in analysis and classification in medical imaging with an application allowing the automatic detection and quantification of Alzheimer's disease.

Keywords : visualization of image databases, indexing, movement detection, classification/quantification of the Alzheimer disease, fusion/averaging/consensus of segmentations.

TABLE DES MATIÈRES

RÉSUMÉ . . . . . **ii**

TABLE DES MATIÈRES . . . . . **iv**

LISTE DES TABLEAUX . . . . . **ix**

LISTE DES FIGURES . . . . . **xi**

LISTE DES APPENDICES . . . . . **xviii**

LISTE DES SIGLES . . . . . **xix**

DÉDICACE . . . . . **xxiii**

REMERCIEMENTS . . . . . **xxiv**

CHAPITRE 1 : INTRODUCTION **1**

1.1 Contexte de recherche 1

1.2 La segmentation d'images 4

1.2.1 Définition d'une partition 4

1.2.2 Stratégies de segmentation d'images 6

1.3 Moyennage consensuel de segmentation d'images 7

1.4 Consensus de segmentations en imagerie 11

1.5 État de l'art sur les techniques de visualisation/navigation/recherche dans les grandes base de données 15

1.5.1 Recherche d'images par le contenu visuel 15

1.5.1.1 Introduction 15

1.5.1.2 Traitements de bas niveau : extraction des descripteurs 16

1.5.1.2.1 Descripteurs de couleurs 17

1.5.1.2.2 Descripteurs de texture 19

1.5.1.2.3 Descripteurs de formes 20
1.5.1.2.4 Détecteurs et descripteurs de points d'intérêt 23
1.5.2 Visualisation 24
1.5.3 Conclusion 25
1.6 État de l'art sur la détection/segmentation du mouvement 26
1.6.1 Méthodes basées sur la modélisation de l'arrière-plan 27
1.6.2 Méthodes basées sur la mesure du mouvement 28
1.6.3 Amélioration des méthodes de détection du mouvement 29
1.7 Détection et classification en imagerie médicale 29
1.7.1 Théorie de l'imagerie par résonance magnétique 30
1.8 Structure du document 31
1.8.1 Plan de la thèse 31
1.9 Publications 33

I Consensus de segmentation dans le cadre de la visualisation, navigation et la recherche dans les grandes bases de données d'images 35

CHAPITRE 2 : SEGMENTATION DATA VISUALIZING AND CLUSTERING 36
2.1 Introduction 37
2.2 Distance-Based Clustering Components 40
2.2.1 Used Distance 40
2.2.2 Cluster Center Estimation 42
2.3 Examples of Applications 45
2.3.1 Visualization of Image Databases 45
2.3.2 Segmentation-Based Clustered Visualization 51
2.3.3 Query-by-drawing search 57
2.3.4 Hierarchical Clustering 57
2.3.5 Discussion 62

2.4 Conclusion 62

II Consensus de segmentation pour améliorer la précision des algorithmes de détection de mouvement 65

CHAPITRE 3 : CHANGE DETECTION ALGORITHM IMPROVEMENT BY ADAPTIVE FUSION OF RESULTS OBTAINED FROM DIFFERENT COLOR SPACES 66

3.1 Introduction 67
3.2 Related Work in Combination of Color Spaces 71
3.3 Single Threshold BGS Methods in Different Color Spaces 74
3.3.1 Introduction 74
3.3.2 Color Models 75
3.3.3 BGS Algorithm Performance in Different Color Spaces 77
3.4 Methods 78
3.4.1 Proposed Model 78
3.4.2 Estimation of the optimal combination of color models 82
3.5 Results and Discussion 83
3.6 Conclusion 89

III Consensus de segmentation pour la classification d'imagerie médicale : application au diagnostic de la maladie d'Alzheimer 91

CHAPITRE 4 : CLASSIFICATION OF ALZHEIMER'S DISEASE SUBJECTS FROM MRI USING THE PRINCIPLE OF CONSENSUS SEGMENTATION 92

4.1 Introduction 93
4.2 Experimental study 95
4.2.1 Dataset description 95
4.2.2 MRI data preprocessing 96

4.2.3 Prototypes NC and AD 97
4.2.4 Hybrid classification 99
4.3 Experimental results 101
4.4 Conclusion and future works 101

CHAPITRE 5 : CLASSIFICATION OF ALZHEIMER'S DISEASE SUBJECTS FROM MRI USING THE PRINCIPLE OF CONSENSUS SEGMENTATION BASED ATLAS 104

5.1 Introduction 105
5.2 Consensus of segmentations in imaging 108
5.2.1 Introduction 108
5.2.2 Consensus segmentation model principle 110
5.2.3 Criteria or distance used in consensus segmentation field 112
5.2.3.1 Global Consistency Error (GCE) 112
5.2.3.2 Variation of Information (VoI) 113
5.2.3.3 Probabilistic Rand Index (PRI) 113
5.2.3.4 The F Measure 114
5.3 Proposed classification model 114
5.3.1 Data description 114
5.3.2 MRI data preprocessing 115
5.3.3 Prototypes NC and AD 116
5.3.4 Two-step classification 118
5.4 Experiments and results 120
5.4.1 Evaluation criteria 121
5.4.2 Performance measures & comparison with state-of-the-art methods 121
5.4.2.1 Classification results into two classes (NC *vs.* AD) . . 121
5.4.2.2 Classification results into four classes (NC, very mild AD, mild AD and moderate AD) 124
5.4.3 Visualization of image databases 125
5.5 Conclusion 127

CHAPITRE 6 : CONCLUSION 128
6.1 Sommaire des contributions 129
6.2 Perspectives générales 130

CHAPITRE 7 : ORIENTATIONS FUTURES DE LA RECHERCHE . . . 131
7.1 Reconstruction 3D d'objets à partir d'images multivues 131
7.2 Débruitage des segmentations en mouvement par moyennes non-locales 135
7.3 Fusion de plusieurs points de vue pour améliorer la détection du mouvement 137
7.4 Analyse en Composante Principale/Indépendante de segmentations . . . 142
7.5 Autres applications possibles 142

BIBLIOGRAPHIE 143

LISTE DES TABLEAUX

3.1 F-measure result, on CDnet 2014 dataset, for four pixel-level single-threshold BGS methods ; namely the BGS based on a simple difference, median filter [1], a mixture of Gaussians [2] or an eigen-decomposition [3], operating in different color spaces (the optimal threshold is indicated in parenthesis next to the F-measure value). 77

3.2 Mean Phi coefficient obtained on some pairs of color models, on the training sequences of the CDnet, for the Zivkovitch's BGS model. . . . 83

3.3 F-measure results obtained for some different pairs or triplets of color spaces for the three variants (supervised, semi-supervised, unsupervised) of our fusion procedure and for respectively the Zivkovitch's [2], eigen [3] and median [1] BGS model. 84

3.4 Detailed evaluation results, in terms of F-measure, of the [HSV-LAB] Zivkovitch's BGS fusion algorithm for each category of the CDnet 2014 dataset. 89

3.5 Real time operation : $CSF_{(GMM\text{-}ZIVKOVIC)}$ (unsupervised) 89

4.1 Demographic information of the subjects in the two classification classes 96

4.2 Classification results comparison between our method and the morphometric approaches proposed in [4] and several other approaches (using a leave-one-out cross validation test and the same dataset). 102

5.1 Demographic information of the subjects in the two classification classes. 115

5.2 Classification results comparison between our method and the morphometric approaches proposed in [4] and several other approaches (using a leave-one-out cross validation test and the same dataset). 122

5.3 Comparison of the classification results of our approach with different consensus criteria for the creation of prototypes (98 subjects). 124

5.4 Comparison of the classification results of our approach with different consensus criteria for the creation of prototypes (for the entire database : 416 subjects). 124

I.1 Échéancier de la thèse i

LISTE DES FIGURES

1.1 Résultats de segmentation obtenus par l'algorithme VOIBFM [5] sur les cinq premières images de l'ensemble de données Weizmann (première et deuxième lignes : ensemble de données d'objet unique, troisième et quatrième lignes : ensemble de données à deux objets). 5

1.2 Exemples d'ensemble de segmentation et de résultat de fusion en utilisant plusieurs critères, (a) cerveau normal et (b) cerveau malade. 12

1.3 Exemples d'ensemble de segmentation et de résultat de fusion au sens de l'inertie ou de la variance intra-cluster (pour l'ensemble des valeurs d'histogramme d'étiquettes locales fournies par chaque segmentation [6]) 13

1.4 Principe général de la recherche d'images par le contenu et les caractéristiques visuelles de bas niveau (couleur, forme, texture) d'une image requête donnée par l'utilisateur. 16

1.5 Répartition des couleurs au sein des principaux espaces couleurs. 17

1.6 Les images (a) et (b) possèdent des histogrammes de couleurs similaires. Les images (b) et (c) possèdent des histogrammes de couleurs très différents. 18

1.7 Exemples de descripteurs géométriques [7]. 21

1.8 Une carte de visualisation MDS du BSDS300 basée sur la distance VoI entre les segmentations [8]. A gauche : voir la carte des (300) images en couleur mappées en fonction de leur similarité en termes de résultat de segmentation en fonction de la région. Droite : voir la carte des résultats de la segmentation en fonction de leur similarité dans le sens de la distance VoI. 25

1.9 Définition des plans de coupe en imagerie médicale 30

1.10 Principe de fonctionnement de l'IRM 31

2.1 A MDS visualization map of the BSDS300 based on the VoI distance between segmentations. Left : view map of the (300) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense. 45
2.2 Magnified details of Fig. 2.1 showing the grouping of images of the BSDS300 based on the similar layout and arrangement of the objects detected and segmented in the scene. In lexicographic order, a group of pictures mostly showing one or two people at the center of the image (top left) or showing that the three images of bears of the BSDS300 are close (top right) and finally a group of pictures showing that the six birds of the BSDS300, perched on a branch are relatively close or showing, for the bottom or left part of the image, an animal on a mountainside or located in the middle of the image. 49
2.3 6 nearest neighbors of a given (the leftmost) image belonging to the BSDS300. From top to bottom, group of images mainly showing 1-] a small and elongated animal or object on the grass/sand or in the sky/water. 2-] an animal or a group of animals. 3-] pyramidal or (highly) elongated structures (pyramid or mountain). 4-] wild mammals. 5-] group of men. 6-] small pyramidal structures in land/sea-scape. 50
2.4 From left to right, the closest and the farthest image, in the mean VoI distance sense, from the center of the BSDS300 with their associated segmentation. The segmentation map which is the exact center of the BSDS300, as estimated by the Algorithm (VoI-Based Mean Segmentation Estimation), is the predictable blank image or one region segmentation (which is not included in the BSDS300 but not too far from the segmentation of the leftmost image). 51
2.5 Fisher's distance (or cluster separability measure) for the two clustering strategies as a function of the number of clusters. 54

2.6 Some prototype (segmentation) centers obtained with the VoI-based K-means for 5, 6, 7 and 8 clusters. 55
2.7 Images from the BSD300 belonging to the cluster corresponding to the cluster prototype, in term of segmentation map, shown on the leftmost image. Images exhibiting elongated structures such as persons or statues or ears of corn (or reeds) have been clustered and retrieved from the BSDS300. 56
2.8 Examples of query-by-drawing search (three nearest neighbors) in the BSDS300 in the VoI distance sense. 58
2.9 A dendrogram or hierarchical agglomerative clustering based on the VoI distance between segmentations on the first 50 images of the BSDS300, showing the 50 agglomerations and the average segmentation between each similar groups of images, in terms of their region-based descriptive content. 60
2.10 Images from the BSDS300 related to the first agglomerations of the HAC, or most similar images, in terms of their region-based descriptive content. 61
2.11 Top : Image 101085 from the BSDS300. Bottom : Top 12 most similar images returned by Google image search. A Google search online returned several images of statues with the same color and texture between them (and enough similar to the query image) but none of them show three elongated vertical structures with possibly a different color or texture as our method can give (cf. Fig 2.7). 64

3.1 RGB, HSV and CIE-LAB color spaces. From left to right, (**a**) the RGB color space, (**b**) the HSV (hue, saturation or chrome, value) color space and (**c**) the CIE-LAB color space in which, L= 0 yields black and L= 100 indicates diffuse white, A indicates its position between red/magenta and green (negative values for A indicate green while positive values indicate magenta) and B indicates its position between yellow and blue (negative values for B indicate blue and positive values indicate yellow). 76

3.2 The distribution of the difference (green curve) between the current frame and the background model is composed of two distributions ; the distribution corresponding to the *background* class (blue) and the one corresponding to the *foreground* label (red). For a given threshold τ_1, the number of false positives and false negatives are proportional respectively to the blue and read area. By increasing τ_1, we decrease the number of false positives, but increase the number of false negatives and conversely. . . 79

3.3 Examples of binary motion segmentations obtained from the Zivkovitch's BGS operating in the RGB, HSV, LAB color space and the fusion result achieved with our supervised fusion model in different pairs of color spaces. 86

3.4 Examples of binary motion segmentations obtained from the Zivkovitch's BGS operating in the RGB, HSV, LAB color space and the fusion result achieved with our unsupervised fusion model in different pairs of color spaces. 87

3.5 Comparisons of results (on different CDnet 2014 images) from different algorithms : from top to bottom respectively ; CSF (*i.e.* our [HSV-LAB] supervised fusion approach using the Zivkovitch's BGS), SuBSENSE [9], FTSG [10], CwisarDH [11], Spectral-360 [12], AMBER [13], KNN [2], SCSOBS [14], KDE-ElGammal [15], RMoG [16] algorithm. 88

4.1 Snapshot of a specific subject. (a) One original scan. (b) Atlas-registered image. (c) Brain masked version of Figure 1(b). (d) The GM/WM/CSF segmentation image. 97
4.2 Methodology for the creation of NC and AD prototypes 98
4.3 The methodology of the proposed method. 99
4.4 $|D\_NC - D\_AD|$ large (case 1 or 2) or very low (case 3) 100
4.5 Classification results for the different methods. 102

5.1 Snapshot of a specific subject. (a) One original scan. (b) Atlas-registered image. (c) Brain masked version of Fig. 5.1 (b). (d) The GM/WM/CSF segmentation image. 116
5.2 Methodology for the NC and AD prototype creation. 117
5.3 Prototype AD and NC according to different criteria of merging (GCE [17], FM [18], LSQ [6], VOI [5], PRI [19]). 118
5.4 $|D_{NC} - D_{AD}|$ large (case 1 or 2) or low (case 3). 119
5.5 The methodology of the proposed two-step classification method. . . . 120
5.6 Performance comparison between our method and the morphometric approaches proposed in [4] for AD *vs.* NC. 123
5.7 From the lexicographic older : prototypes using the GCE distance and built from (a) : CDR=0.5 (very mild AD), (b) CDR=1 (mild AD), (c) CDR=2 (moderate AD). 125
5.8 A MDS visualization map of the considered subset of the OASIS database (49 healthy and 49 unhealthy subjects) based on the GCE distance between (pair of) segmentations according to their similarity. MRIs framed in blue are NC and MRIs framed in red are AD. the large red and blue frame is respectively the AD and NC prototypes. 126

7.1 Exemple de reconstruction 3D dense d'objet à partir d'images prises selon des points de vue différents, par un modèle traditionnel de reconstruction stéréo multivue (MVS) [20] qui se heurte ici surtout au manque de texture de l'objet à reconstruire et aussi un peu aux problèmes de spécularités, de réflexions et de transparence. On peut remarquer que la reconstruction obtenue est très bruitée et présente à quelques endroits des trous et de nombreux artéfacts. 132
7.2 Principe de la méthode de reconstruction 3D proposée. Des sous-échantillonnages du nuage de points donné par la méthode MVS permettent d'obtenir une série de reconstructions 3Ds bruitées qui seront ensuite fusionnées au sens d'un critère par un modèle de moyennage consensuel régularisé. . 133
7.3 Illustration du principe des méthodes de débruitage par (*patchs*) comme celui du filtrage par moyennes non locales proposé par Buades *et al.* [21]. Une imagette (ou *patch*) peut se retrouver presque à l'identique à plusieurs endroits de l'image. Dans cet exemple, dans le voisinage spatial (fenêtre carrée bleue), deux voisinages carrés (ou *patchs*), associés au pixel $q1$ et $q2$, sont très semblables au voisinage du pixel p (contrairement au voisinage du pixel $q3$). 136
7.4 Segmentation en mouvement d'une séquence vidéo par l'algorithme de mélange gaussien adaptatif [22] sur la séquence *railway*. 137
7.5 Voisinage de recherche spatial (taille 27×27 pixels) et temporel (de taille 5) illustrée par la fenêtre jaune et patch p (de taille 9×9 pixels) à débruiter illustrée par la fenêtre rouge. 137
7.6 Cube vidéo ou vue compacte de la séquence vidéo obtenue en empilant toutes les images 2D le long de la dimension temporelle (téléchargée depuis le site web : changedetection.net). 138
7.7 Séquence vidéo de 1700 images vue comme une succession d'images 2Ds (ou de coupes dans le plan (x,y) pour différents $t = n\Delta t$ (les images $n = 155$, 266, 827, 1641, 1686 et 1700 sont représentées). 139

7.8 **(a)** Coupe sagittale dans le plan (x,t) (vue 2) et **(b)** transversale dans le plan (y,t) (vue 3) pratiquée sur le cube d'images de la Fig. 7.7. 139
7.9 étapes d'une stratégie de fusion multivue. 141

II.1 A MDS visualization map of the Weizmann200 based on the VoI distance between segmentations. Left : view map of the (200) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense. ii
II.2 Magnified details of Fig. II.1 showing the grouping of images of the Weizmann200 based on the similar layout and arrangement of the objects detected and segmented in the scene. iii
II.3 A MDS visualization map of the Stanford715 based on the VoI distance between segmentations. Left : view map of the (715) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense. iv
II.4 Magnified details of Fig. II.3 showing the grouping of images of the Stanford715 based on the similar layout and arrangement of the objects detected and segmented in the scene. v
II.5 A MDS visualization map of the Msrc590 based on the VoI distance between segmentations. Left : view map of the (590) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense. vi
II.6 Magnified details of Fig. II.5 showing the grouping of images of the Msrc590 based on the similar layout and arrangement of the objects detected and segmented in the scene. vii

LISTE DES APPENDICES

Annexe I : **Échéancier de la thèse** . . . . . **i**

Annexe II : **Suppléments de résultats expérimentaux à la partie 1** . . . **ii**

LISTE DES SIGLES

AD	Alzheimer's Disease
BM	Background model
BGS	Background Subtraction
BSDS300	Berkeley Segmentation Dataset 300
CS	Color Space
CSF	Cerebro-Spinal Fluid
CDR	Clinical Dementia Rate
CPU	Central Processing Unit
CDnet	ChandeDetection.net
CD	Change detection
CDP	Color Difference Processing
CSF	Color space fusion
CSFV	Cerebrospinal fluid volume
CAD	Computer-Aided Diagnosis
CBIR	Content Based Image Retrieval
CBVIR	Content Based Visual Information Retrieval
DFT	Discrete Fourier Transform
DS	Dempster-Shafer
EM	Expectation Maximisation
FTSG	Flux tensor with split gaussian mdoels
FPS	Frames Per Second
FBM	Features Based Morphometry

FN	False Negatives
FP	False Positives
GCE	Global Consistency Error
GMM	Gaussian Mixture Model
GGMRF	Generalized Gaussian Markov Random Field
GM	Mixture of Gaussians or Grey Matter
HAC	Hierarchical Agglomerative Clustering
HV	Hippocampal Volume
HSL	Hue saturation lightness
HSV	Hue saturation value
HSI	Hue saturation Intensity
IUTIS	In unity there is strength
ICA	Independent Component Analysis
KNN	K-Nearest Neighbors
LOOCV	Leave-One-Out Cross Validation
LSQ	Least Square
LVV	Lateral Ventricles Volume
LDA	Linear Discriminant Analysis
LRE	Local Refinement Error
MVS	Multi-View Stereo
MRF	Markov Random Fields
MDS	Multidimensional Scaling

MDC	Minimal Distance Classifier
MD	Minimal Distance
MAP	Maximum A Posteriori
MBS	Multimode Background Subtraction
MCS	Multi-classifier systems
MRI	Magnetic Resonance Imaging
NL-Means	Non-Local Means
NC	Normal Control
NN	Neural Network
OpenMP	Open Multi-Processing
OBM	Object Based Morphometry
OASIS	Open Access Series of Imaging Studies
PBAS	Pixel-Based Adaptive Segmenter
PRI	Probabilistic Rand Index
PCA	Principal Component Analysis
PSD	Positive and Semi-Definite
RSB	Rapport Signal sur Bruit
ROI	Region Of Interest
RGB	Red Green Blue
SVM	Support Vector Machine
SOM	Self Organized Map
SURF	Speeded up robust features
SuBSENSE	Self-Balanced sensitivity segmenter

sMRI	structural Magnetic Resonance Imaging
SIFT	Scale Invariant Features Transform
SPM	Statistical Parametric Mapping
TBM	Tensor Based Morphometry
TP	True Positives
TN	True Negatives
VoI	Variation of Information
VOIBFM	VoI-Based Fusion Model
VBM	Voxel Based Morphometry
WM	White Matter

Je dédie cette thèse à :

Ma mère ***Awatef***
Pour son soutien incroyable durant mes études.
Je n'oublie pas ses énormes sacrifices.

L'âme de mon cher père ***Abdellatif***
Source de mon inspiration dans la vie.

À Mon cher frère ***Hamadi*** pour son encouragement indéfectible.

À Ma sæur ***Marwa*** pour leur soutien moral tout au long de ma formation.

À tous ceux qui me sont chers.

REMERCIEMENTS

À l'occasion du présent travail de doctorat je désire remercier le ministère de l'enseignement supérieur tunisien et l'université de Montréal pour avoir co-financé ce travail de recherche à travers plusieurs bourses d'excellence.

Je ne saurais, réellement, trouver les expressions éloquentes que mérite mon directeur de recherche Mr. **Max Mignotte**, afin de le remercier pour sa sympathie, ses encouragements, ses conseils, son dévouement pour le travail et sa totale présence tout au long du déroulement de cette thèse.

Je veux remercier également les membres du jury qui ont bien voulu me faire l'honneur de juger cette thèse.

Mes remerciements par ailleurs s'adressent à toute ma famille pour son encouragement et son soutien tout au long de mes études.

Finalement, Je tiens à exprimer toute ma reconnaissance, ma gratitude et mon respect à tous les membres du laboratoire de traitement d'images (DIRO).

CHAPITRE 1

INTRODUCTION

1.1 Contexte de recherche

En imagerie numérique, un grand nombre d'algorithmes, que ce soit en traitement d'images (pour la recherche d'informations de bas niveau) ou en vision par ordinateurs (pour l'estimation d'information de plus haut niveau) utilisent des procédures ou des méthodes provenant soit de l'analyse et de l'estimation statistique (possiblement multivariées) des données ou encore de la théorie de l'apprentissage.

Toutes ces méthodes qui se ramènent quelquefois à un problème d'étude de similarités ou de dissimilarités ou encore de partitionnement (segmentation) des données, de réduction de dimensionnalité, de débruitage, de décorrélation (blanchiment du signal) ou de régularisation de problèmes inverses mal posés, utilisent en fait une notion de distance entre les données éventuellement bruitées et souvent aussi la notion de valeur moyenne entre des données appartenant à des groupements (*clusters*) ou des classes identiques.

Ces données en imagerie peuvent être, par exemple, la valeur de couleur ou le niveau de gris d'un pixel ou sa position (coordonnée), un vecteur de signature spectrale dans le cas d'une image multispectrale, ou une liste de points 2Ds pour caractériser un contour dans le cas du traitement d'images. Dans le cas de la vision par ordinateur, ces données peuvent être une valeur de profondeur, un vecteur de mouvement ou un point ou un nuage de points (ou de voxels) pour caractériser une forme dans un espace 3D, etc.

Quelquefois, les données que l'on aimerait analyser, traiter, débruiter, ou classer sont à un niveau d'abstraction plus élevé et sont déjà le résultat d'un traitement classique. C'est le cas par exemple des données provenant d'un partitionnement ou du résultat d'une classification ou d'un *clustering* (possiblement non supervisée). Lorsque ces données représentent des niveaux de gris ou des niveaux de couleurs et sont agencées spatialement, comme le sont les pixels d'une image, ces résultats de partitionnement sont

communément appelés des cartes de segmentation en régions.

Ces dernières années, de vraies mesures de distances au sens d'un certain critère ont été introduites pour ces données segmentées. C'est le cas de la distance VoI (variation d'entropie) permettant de quantifier la distance existant entre deux cartes de segmentation au sens de l'entropie de Shannon et qui correspond à une vraie métrique (distance positive, symétrique et obéissant à la propriété de l'inégalité triangulaire) dans l'espace des *clusters* (cf. Sec 2.2.1). À partir de cette distance ou de ce critère VoI, la notion de segmentation moyenne a été ensuite proposée en [5] comme étant la solution d'un problème d'optimisation dont une solution approchée (mais presque optimale) est estimée avec une complexité linéaire en fonction du nombre de pixels de l'image et sans la restriction que toutes les segmentations, destinées à être fusionnées ou moyennées, contiennent le même nombre de régions. D'autres distances, basées sur le Rand Index [19], le GCE (Global Consistency Error) [23] ou la F-mesure [18] qui utilisent respectivement, la notion de paires d'étiquettes communément bien classifiées, la nature intrinsèque multi-échelle d'une segmentation (en terme de degré de raffinement possible) ou la notion de précision-rappel utilisée en classification, ont permis aussi de quantifier efficacement la distance existant entre deux groupements de données ou cartes de segmentation et définir ainsi une segmentation moyenne (ou consensuelle) associée à un critère de distance spécifique.

Ce moyennage de segmentations permet conceptuellement de débruiter un ensemble de cartes de segmentation (de la même scène) estimées grossièrement (et rapidement) à partir d'un modèle de segmentation simple. Au final, ce moyennage (ou cette fusion) permet d'obtenir une segmentation finale plus précise, c'est-à-dire plus proche de la vérité-terrain (au sens d'un critère de distance donné) que ne le sont chaque segmentation individuelle. Cette stratégie qui peut se concevoir comme un débruitage (ou une technique de réduction de bruit) de données d'abstraction élevée, par moyennage s'est avérée récemment une alternative efficace et très parallélisable, comparativement aux méthodes utilisant des modèles de segmentation toujours plus complexes et plus coûteux en temps de calcul. Ce moyennage de segmentations permet aussi de représenter un

groupement d'images (et leurs segmentations) en un prototype (de segmentation), qui résumera l'aspect structurel des différents objets géométriques communs ou des différentes structures communes que contient ce groupement d'images.

Plus généralement, le principe de distance entre segmentations (au sens d'un certain critère) et de moyennage de segmentations (d'une même scène) peut être exploité, directement ou facilement adapté, par tous les algorithmes ou les méthodes utilisées en imagerie numérique où les données peuvent en fait se substituer à des images segmentées.

L'objectif de cette proposition de thèse est de démontrer cette assertion et de présenter différentes applications originales du moyennage ou fusion de segmentations dans des domaines comme :

- L'indexation ou le multimédia (cf. Chap 2) où ce principe est exploité dans le cadre de la visualisation, l'exploration, la navigation et la recherche dans des grandes bases d'images au sens, non plus des couleurs ou de la texture (présentes dans l'image) comme il est communément proposé dans la littérature, mais en fonction de la façon structurelle dont sont disposés les différents objets ou éléments présents (et préalablement segmentés) dans ces images.

- Le traitement d'images vidéo (cf. Chap 3) et plus spécifiquement, la détection du mouvement dans une séquence d'images, où ce principe de fusion de segmentations est exploité pour combiner efficacement ou débruiter de multiples résultats de carte de détection de mouvement obtenus par le même modèle de soustraction de fond, opérant dans des espaces de couleurs différentes et complémentaires.

- La classification des données ou l'analyse d'images médicales (cf. Chap 4 & Chap 5) et plus précisément pour la classification et la quantification de la maladie d'Alzheimer. Dans cette application, le principe de moyennage consensuel au sens de certains critères est utilisé afin de construire deux prototypes de segmentation anatomique (ou atlas segmenté) du cerveau humain (sain / malsain) ou encore plusieurs prototypes de segmentation du cerveau malsain associés à différents degrés

de sévérité de la maladie d'Alzheimer et en termes des structures anatomiques telles que matière grise, matière blanche et liquide céphalo-rachidien. Ces prototypes basés sur une segmentation par consensus ont plusieurs caractéristiques attrayantes pour notre tâche de classification qui seront décrites ultérieurement.

1.2 La segmentation d'images

1.2.1 Définition d'une partition

La segmentation d'image est une étape de prétraitement fréquente en analyse d'images dont le but est de simplifier la représentation d'une scène (ou d'une image) par un ensemble de régions significatives et spatialement cohérentes (aussi connus, en anglais, sous le nom de *segments* ou *super-pixels*) ayant des attributs similaires (tels que les parties cohérentes des différents objets de la scène ou celles de l'arrière-plan) (cf. Fig 1.1).

Plus formellement, la segmentation consiste à partitionner une image I en n sous ensembles R_i, appelés régions, tels que $\forall i, j \in \{1, .., n\}$:

(i) $I = \bigcup_{i=1}^{n} R_i$

(ii) $R_i \neq \emptyset$

(iii) $i \neq j \Rightarrow R_i \cap R_j = \emptyset$

où $\bigcup$ représente une union d'ensembles disjoints et $\cap$ représente l'intersection ensembliste. Ces sous-ensembles R_i constituent les différentes régions de l'image. Ces régions obtenues doivent respecter les propriétés d' homogénéité où (a) chaque pixel appartient à une région et une seule, (b) tous les pixels doivent être traités, (c) toute région doit être connexe (4-Connexité ou 8-Connexité par exemple) [24].

Cette tâche de vision de bas-niveau, qui change la représentation d'une image en quelque chose qui devient plus facile à analyser, est souvent l'étape préliminaire et aussi une étape cruciale dans le développement de plusieurs algorithmes de compréhension d'images et les systèmes de vision par ordinateur nécessitant une reconstruction 3D [25] ou une localisation/ reconnaissance d'objet 3D [26, 27].

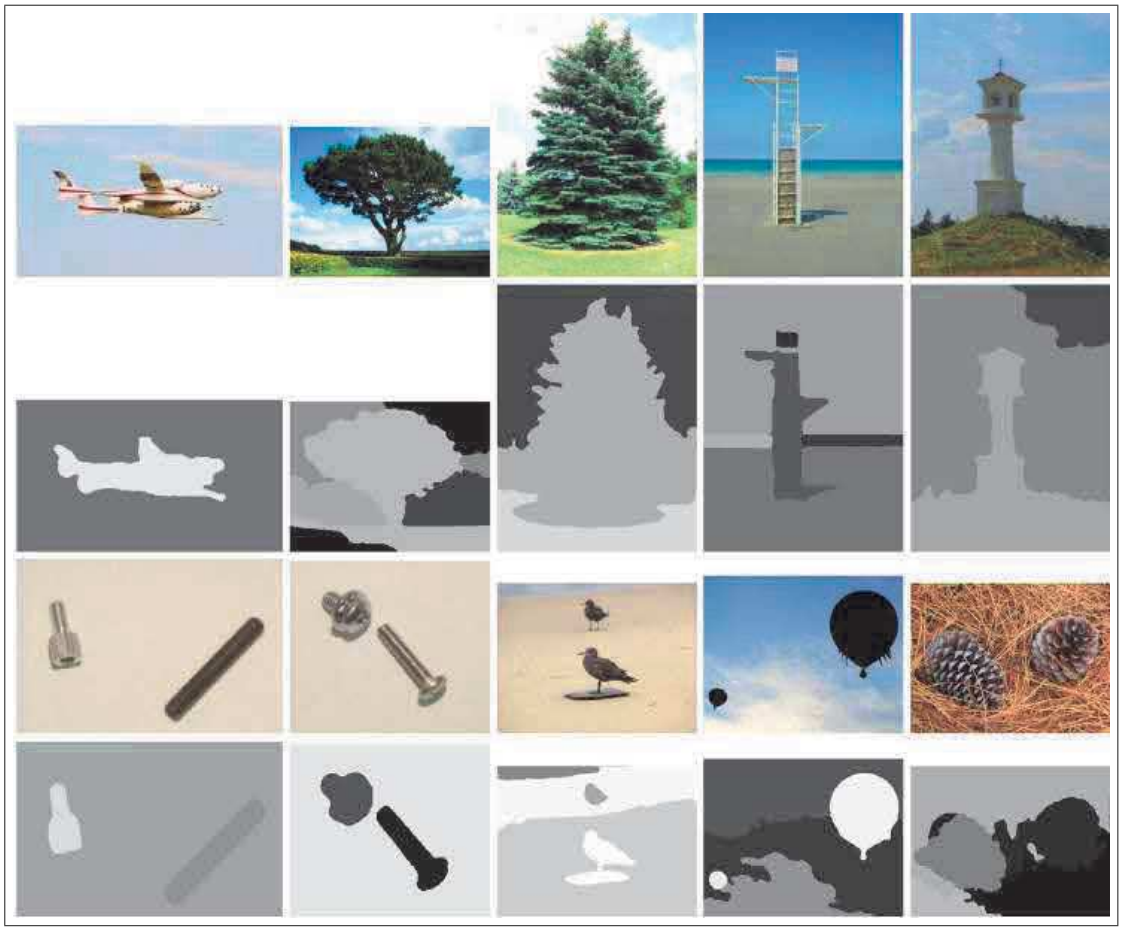


FIGURE 1.1 : Résultats de segmentation obtenus par l'algorithme VOIBFM [5] sur les cinq premières images de l'ensemble de données Weizmann (première et deuxième lignes : ensemble de données d'objet unique, troisième et quatrième lignes : ensemble de données à deux objets).

1.2.2 Stratégies de segmentation d'images

Jusqu'à présent, une pléthore de méthodes de segmentation en régions ont été proposées dans la littérature pour résoudre le problème de la segmentation non supervisée d'images naturelles texturées. La majorité de ces méthodes exploitent d'abord une étape d'extraction des caractéristiques de textures (dont le but est de caractériser chaque région significative texturée destinée à être segmenté), suivie d'une technique de regroupement, tentant de grouper spatialement (avec différents critères ou stratégies) des régions spatialement cohérentes partageant des caractéristiques similaires. Des années de recherche en segmentation ont ainsi porté sur la recherche de caractéristiques (locales) d'image plus sophistiquées et/ou des techniques de regroupement plus élaborées et des améliorations significatives dans les résultats de la segmentation finale ont été obtenues, généralement au détriment d'une augmentation de la complexité du modèle de segmentation et/ou de sa complexité de calcul.

Ces méthodes comprennent des modèles de segmentation exploitant directement des systèmes de regroupement (ou *clustering*) [6, 28–30] exploitant une modélisation par mélange de gaussiennes, [31] une classification floue [32–34] (*fuzzy clustering*) ou basée sur la théorie des ensembles flous [35] (après une possible approche de dé-texturation de l'image [30, 36, 37]), basés sur l'algorithme du *mean-shift* ou plus généralement sur des algorithmes recherchant les modes d'une distribution (d'un ensemble de niveaux de couleurs) [38–40], des procédures simulant le remplissage des bassins (*watershed algorithm*) [41, 42] ou utilisant la stratégie de croissance des régions [43–45], du codage avec pertes et des modèles de compression [42, 46], des transformées en ondelettes [47], des modèles Markoviens (*Markov Random Field*) [48–55], ou plus généralement Bayésiens [56] basés sur la notion de texton [57] ou sur celle des graphes [58–60], et/ou utilisant des méthodes variationnelles [57, 61–64], de surfaces déformables [65], de contours actifs [66], avec des graphes de partitionnement [67], des courbes [68], des techniques de seuillage itératif non supervisé [69, 70], basées sur les algorithmes génétiques [71], les cartes d'auto-organisatrices [72] (*Kohonen*), diverses techniques d'apprentissage [73], la topologie [74], les objets symboliques [75] et finalement le regroupement spectral [76]

pour n'en nommer que quelques-unes.

Une alternative récente et efficace à ces différentes approches de segmentation susmentionnées consiste à combiner (ou fusionner) plusieurs cartes de segmentation (de la même scène) estimée rapidement et grossièrement à partir d'un modèle de segmentation simple pour obtenir, *in fine*, une segmentation finale améliorée très robuste[1]. Dans cette stratégie, au lieu de chercher le meilleur algorithme de segmentation (avec ses paramètres internes optimaux) ce qui est difficilement possible en tenant compte des différents types d'images existantes, on préfère chercher le meilleur modèle de fusion de segmentation, ou plus précisément le critère le plus adapté (*i.e.,* efficace) pour la fusion de plusieurs cartes de segmentations.

1.3 Moyennage consensuel de segmentation d'images

Combiner (ou fusionner) plusieurs cartes de segmentation d'image peut être considéré comme un cas particulier du problème dit de "clustering collaboratif" permettant la combinaison de plusieurs méthodes de clustering (appelé aussi classification par méthode ensembliste et *cluster ensemble problem* ou *collaborative clustering* en anglais). Ces méthodes dites d'*ensemble clustering* ont pour concept de combiner plusieurs résultats de regroupements (d'un même ensemble de données) dans le but ultime d'améliorer le résultat de clustering ou de regroupement final. Ce problème du clustering par ensemble fut initialement exploré récemment dans le domaine de l'apprentissage machine [77–79]. Comme il a été déjà mentionné, la combinaison de plusieurs segmentations est un cas particulier du clustering collaboratif car un aspect distinctif des données image est leur agencement (ou cohérence) spatial ; une segmentation d'image est donc un processus de groupement de données (caractéristiques texturales pour chaque pixel)

[1] Les cartes de segmentation destinées à être fusionnées peuvent être données soit par différents algorithmes (idéalement complémentaires) ou bien par le même algorithme avec différentes valeurs des paramètres internes ou états (graines ou *seeds*) du générateur de nombre pseudo-aléatoire (pour les méthodes de segmentation stochastiques), ou en utilisant des caractéristiques texturales différentes, possiblement appliquées à une image de départ éventuellement exprimée dans différents espaces de couleurs ou transformations géométriques (par exemple ; facteur d'échelle, inclinaison, transformation log-polaire) ou encore par d'autres moyens.

qui sont indexées spatialement. Par conséquent, dans la segmentation d'image, le regroupement des pixels dans les différents clusters doit prendre en compte non seulement la similitude des données dans l'espace des caractéristiques (texturales), mais aussi leur cohérence spatiale.

Il est intéressant de noter aussi que ce problème de clustering par ensemble peut également être considéré comme étant un cas particulier d'un problème de débruitage dans lequel chaque segmentation (appartenant à l'ensemble des segmentations à fusionner) représente en fait une solution ou une observation bruitée. L'objectif final de la combinaison de segmentations consiste à trouver une solution de segmentation débruitée qui serait, en fait, un consensus ou un compromis (en matière de niveau de détails, localisation des contours, nombre de régions, etc.) exhibé par chaque segmentation. Dans un certain sens, la segmentation finale, résultat de cette fusion, représente aussi la moyenne de toutes les segmentations individuelles selon un certain critère.

Il est important de rappeler tout d'abord, qu'il existe un cas tout particulier d'images (et/ou d'images segmentées) pour lequel une fusion ou moyennage de segmentation peut être facilement estimé et implémenté. Il s'agit (du nombre restreint) d'images dont chacune des régions spatiales segmentées (une région est un ensemble de pixels connectés spatialement et appartenant à une même classe ou un même *cluster*) peuvent être ensuite, dans une seconde étape, associée, sans ambiguité, à une seule et unique étiquette sémantique. C'est le cas par exemple d'une segmentation en deux classes ou deux *clusters* (détection binaire) d'une image, issue d'une séquence vidéo, en région mobile et immobile. Dans ce cas, l'étiquette sémantique mobile est donnée à la région dont la mobilité est en moyenne (*i.e.,* sur l'ensemble des pixels de la région considérée) la plus mobile (des deux). On peut aussi citer l'exemple des images MRI segmentées en trois classes anatomique du cerveau. Les trois classes ou *cluster* de cette image peuvent être associé, sans ambiguité, à la classe "liquide céphalo-rachidien", "matière grise" et "matière blanche". En effet, la classe céphalo-rachidien possède toujours une moyenne de niveau de gris inférieur à la classe matière grise qui elle possède toujours une moyenne (de niveaux de gris) plus faible que la classe matière blanche. Dans ces deux cas, une fois l'étiquette

sémantique identifiée pour chacune des régions de l'image segmentée, un seul et unique critère de fusion peut être facilement et directement utilisé pour fusionner cet ensemble d'images segmentées, il s'agit de la fusion consensuelle (localement pour chaque pixel) par vote majoritaire, communément appelé fusion (ou critère de la) médiane ou encore filtre médian.

Néanmoins, pour tous les autres types d'images où la vraie étiquette sémantique ne peut être asssociée à chaque *cluster*, on ne peut utiliser le principe simple et direct de la fusion médiane. Dans ce cas, conceptuellement, on doit associer ou définir pour chacun des critères de fusion considérées une métrique exprimant une notion de distance entre deux cartes de segmentation (ou de façon équivalente une mesure de similarité entre deux segmentations ou regroupements). À partir de cette distance, on considérera que la segmentation moyenne ou consensuelle, que l'on cherche à estimer, à partir d'un ensemble de carte de segmentation, est la segmentation qui minimise la somme des distances existantes entre elle et toute les autres segmentations issues de l'ensemble de segmentation. Cette modélisation conduit ensuite à un problème d'optimisation plus ou moins complexe selon le critère considéré. Finalement, dans le cas des images de détection binaire de mouvement et de segmentation en trois classes des régions anatomiques du cerveau qui sont deux des trois parties considérées dans cette thèse, on pourra indifféremment choisir entre le critère simple de la médiane[2] et d'autres critères plus élaborés comme le critère VoI [5], GCE [23], PRI [80] et F-mesure [18] (cf. Fig 1.2) et quelquefois même combiner les résultats de segmentation consensuelle obtenus par différents critères complémentaires pour obtenir de meilleurs résultats.

Initialement, cette stratégie du moyennage consensuel, à l'exception du cas particulier et direct du filtre médian, a été initialement proposée dans [81] [82] avec la restriction que toutes les segmentations, destinées à être fusionnées, devaient contenir le même nombre de régions, puis un peu plus tard, sans cette restriction (donc avec un nombre

[2] En fait, la carte de segmentation consensuelle créée par le principe de la fusion médiane est celle aussi qui minimise une certaine distance. Cette distance est en fait tout simplement l'erreur locale de classification sémantique par rapport à l'ensemble des cartes de segmentation à fusionner ou plus simplement la somme ou la proportion d'étiquettes sémantiques qui sont différentes par rapport à celles estmées par chaque segmentation de l'ensemble de segmentation considéré.

arbitraire de régions) dans [6,83]. Depuis ces travaux pionniers, cette fusion de plusieurs segmentations (en région)[3] de la même scène (afin d'obtenir un résultat de segmentation plus fiable) est maintenant effectuée selon plusieurs stratégies et/ou des critères bien définis.

Par exemple, nous pouvons citer le modèle de fusion proposé dans [6] qui fusionne un ensemble de segmentations au sens de l'inertie ou de la variance intra-cluster (pour l'ensemble des valeurs d'histogramme d'étiquettes locales fournies par chaque segmentation) puisque le résultat de la segmentation consensuelle finale est estimé en appliquant le schéma itératif bien connu de l'algorithme des K- moyennes (cf. Fig 1.3). En faisant cela, l'auteur suppose en fait implicitement un modèle de fusion basé sur un mélange de distributions dans lequel les étiquettes attribuées aux différentes régions (données par chaque segmentation destinée à être fusionnée, sont modélisées comme des variables aléatoires distribuées selon K groupements sphériques d'égal volume (ou selon des distributions Gaussiennes avec une matrice de variance-covariance identique) pouvant être regroupées efficacement avec l'algorithme des K moyennes. Selon le même principe, on peut également citer le modèle de fusion proposé dans [84] qui utilise la même stratégie, mais pour l'ensemble des étiquettes *floues* d'appartenance à chaque région et pour lesquelles la procédure de fusion est ainsi réalisée par l'algorithme des K-moyennes flou (donc au sens de l'inertie intra-classe pondérée). Cette fusion de cartes de segmentation (en régions) peut également être réalisée au sens de l'indice probabiliste RAND [80] (ou PRI), avec une fonction de consensus encodant l'ensemble des contraintes, en termes de paires d'étiquettes de pixel (identique ou non), fournies par chacune des segmentations destinées à être fusionnées. Le critère du PRI peut être optimisé soit par une méthode d'optimisation algébrique [85] ou bien par un algorithme de marche aléatoire [83] (et combiné avec un estimateur au sens de l'information mutuelle pour estimer le nombre optimal de régions), ou par un algorithme d'*Expectation Maximisation* (EM) [86] (appli-

[3] Cette stratégie peut également être exploitée, plus généralement, pour divers autres problèmes portant sur les champs d'étiquettes autres que des cartes de segmentation spatiales (par exemple, détection de mouvement ou d'estimation, segmentation 3D, reconstruction 3D, champ estimation de la profondeur, etc.).

qué sur des super-pixels, préliminairement obtenus par une sur-segmentation) ou enfin au sens du PRI pénalisé par une contrainte globale sur le processus de fusion [19] (limitant le nombre et la taille des régions) avec une approche markovienne et une méthode d'optimisation analytique et itérative. Cette fusion de cartes de segmentation peut également être effectuée au sens de l'accumulation des évidences [78] (en utilisant un modèle de regroupement hiérarchique), ou au sens de la variation d'information (VoI) [5] (en utilisant un modèle à base d'énergie optimisé par une stratégie de minimisation locale itérative combinée avec une contrainte de connectivité) ou au sens de la précision et du rappel (ou de la F-mesure) [18] (en utilisant une méthode d'optimisation hiérarchique exploitant les différents super-pixels appartenant à l'ensemble de segmentation). On peut citer également de mentionner le modèle de fusion proposé dans [87] au sens du critère du SVM (machine à vecteurs de support), *i.e.,* au sens de l'hyperplan de marge maximale (entre les classes) et dans lequel une image hyperspectrale est segmentée par une technique de fusion de décision combinant plusieurs classifieurs SVMs entraînés à partir de l'image hyperspectrale initiale exprimée dans différents sous-espaces de caractéristiques différentes. Mentionnons finalement aussi la procédure de fusion Bayésienne pour la segmentation d'image satellitaire proposée dans [88] où les étiquettes de classe obtenues à partir des cartes de segmentation différentes (obtenues à partir de capteurs différents) sont fusionnées par un modèle qui estime chaque étiquette de classe finale (une fois fusionnée) (pour chaque pixel) au sens du maximum *a posteriori* de la distribution Logit. Finalement, récemment, la fusion de cartes de segmentation peut également être effectuée au sens de la combinaison (ou la fusion) de plusieurs critères [89, 90].

1.4 Consensus de segmentations en imagerie

En théorie de l'estimation, un théorème bien connu montre que lorsqu'on a des données qui sont dégradées par du bruit et que ce bruit est non corrélé (*i.e.,* imprévisible, imprédictible comme le bruit gaussien blanc (caractéristique du bruit entachant les appareils électroniques), le moyennage de ses données (bruitées) permet d'obtenir des données qui sont moins bruitées. Plus précisément le rapport signal sur bruit (RSB), en

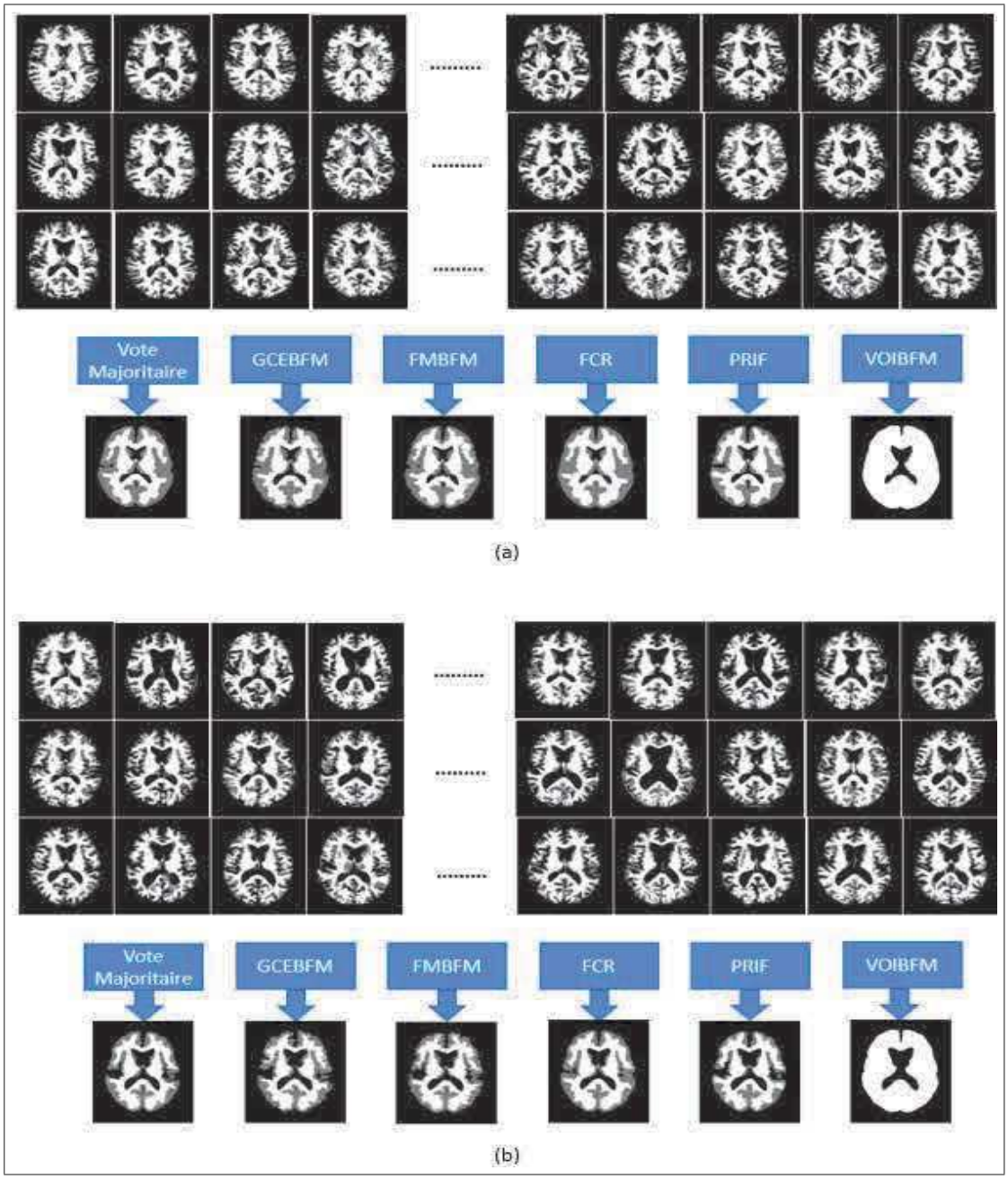


FIGURE 1.2 : Exemples d'ensemble de segmentation et de résultat de fusion en utilisant plusieurs critères, (a) cerveau normal et (b) cerveau malade.

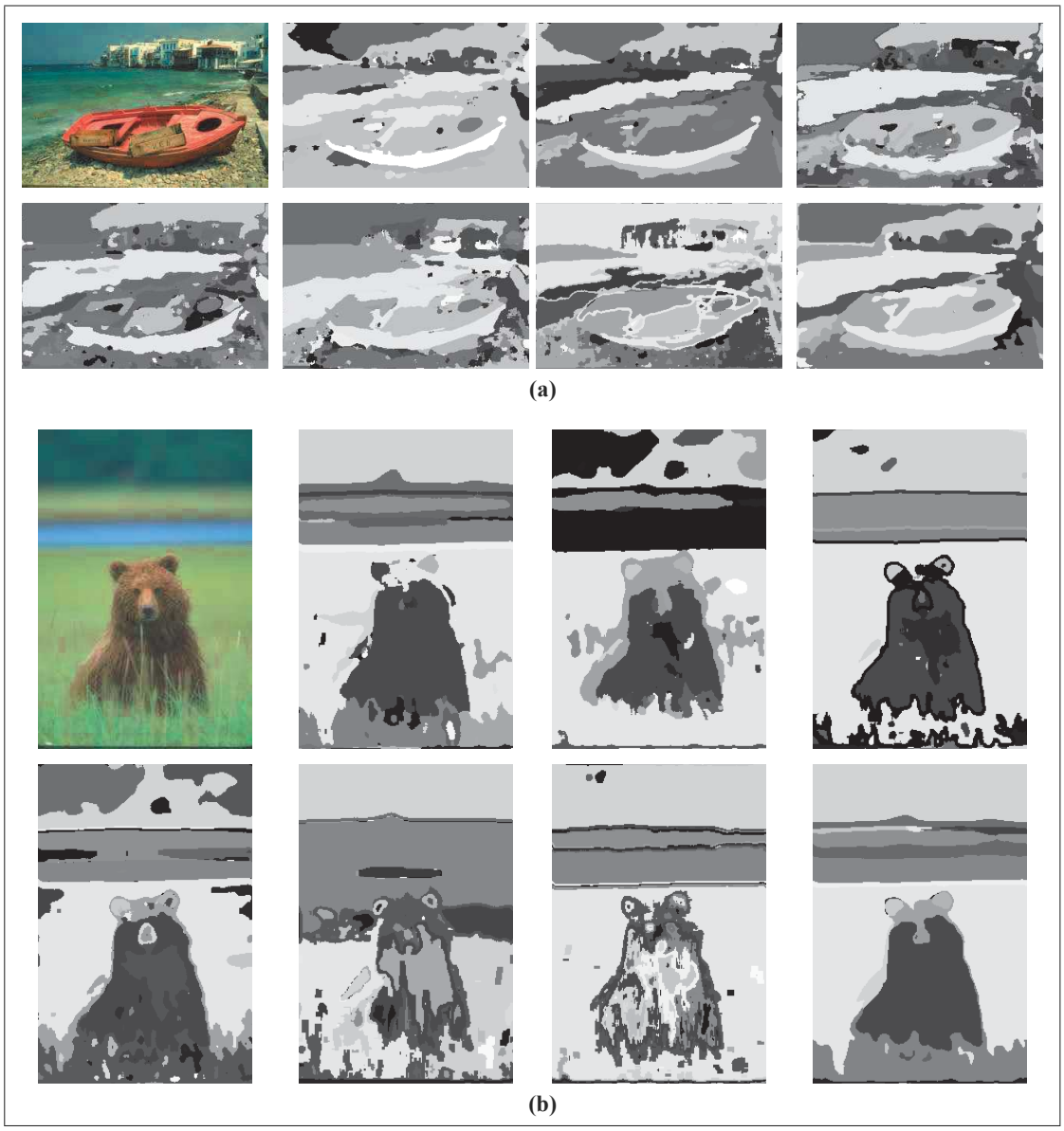


FIGURE 1.3 : Exemples d'ensemble de segmentation et de résultat de fusion au sens de l'inertie ou de la variance intra-cluster (pour l'ensemble des valeurs d'histogramme d'étiquettes locales fournies par chaque segmentation [6])

.

amplitude, des données moyennées augmente de $\sqrt{N}$ ou N est le nombre de données qui seront moyennées.

Dans notre cas, si on généralise la propriété précédente, et si on prend comme données bruitées, des données multidimensionnelles comme des images segmentées bruitées, leurs moyennages à toutes les chances de donner une image segmentée avec un bruit (de segmentation) beaucoup moins important si le bruit entachant ces images segmentées à la propriété d'être non corrélé. Dans le cas d'une image segmentée, ce bruit représente en fait des erreurs de classification et ces erreurs ou ce bruit est non corrélé si les différentes cartes de segmentation destinées à être fusionnées sont complémentaires, c'est-à-dire générant des erreurs de classification différentes. On peut retrouver aussi ce principe en raisonnant par l'absurde. En effet, il est logique de penser que la fusion de carte de segmentation générerait des erreurs assez semblables (sur chaque segmentation), c'est-à-dire très corrélées (ou très prévisibles d'une segmentation à l'autre), et que le moyennage de ces cartes de segmentation ne permettrait donc pas d'obtenir une segmentation finale très différente de chacune des segmentations de l'ensemble (de segmentation) et donc une segmentation très débruitée. C'est le principe même de la fusion de décision qui stipule que combiner ou fusionner des informations fournies par différents experts ou méthodes peut améliorer la solution proposée à un problème à partir du moment où ces informations ont la propriété d'être complémentaires.

Le principe de fusion ou moyennage consensuel de cartes de segmentations 2D constituera le lien commun des algorithmes d'indexation/visualisation, de détection de mouvement et de classification que nous présenterons dans cette proposition de thèse.

1.5 État de l'art sur les techniques de visualisation/navigation/recherche dans les grandes base de données

1.5.1 Recherche d'images par le contenu visuel

1.5.1.1 Introduction

Avec les technologies de l'électronique et de l'informatique, les images numériques se sont multipliées à une vitesse vertigineuse. Pour gérer et utiliser efficacement ces images, un système d'indexation et de recherche d'images est vite devenu nécessaire. Les premiers systèmes de recherche d'images, qui furent développés, utilisaient un ensemble de mots-clés associés à chaque image pour les caractériser. Malgré le grand succès de cette technique, celle-ci fut vite confrontée à plusieurs problèmes. En plus d'être coûteuse pour les bases d'images volumineuses, cette technique ne peut que difficilement décrire parfaitement le contenu d'une image d'une manière exhaustive et aussi nécessite inévitablement l'intervention d'un humain. Pour pallier cet inconvénient, une deuxième approche d'indexation a été proposée dans la littérature. Celle-ci est la recherche par le contenu. Appelée en anglais CBIR (Content Based Image Retrieval), la Fig. 1.4 résume son principe. En résumé, cette technique vise à effectuer une recherche d'images portant sur les caractéristiques visuelles de bas niveau (descripteurs de couleur, forme, texture) d'une image requête donnée par l'utilisateur. Le système mesure en fait une distance de similarité entre les caractéristiques (couleur, forme, texture) de cette image requête et celles des images issues de la base de données. Les images candidates, c'est-à-dire celles considérées comme étant similaires à l'image requête sont celles pour lesquelles cette distance de similarité est faible.

La mesure de similarité utilisée est généralement une vraie distance qui respecte les propriétés mathématiques à savoir (i) l'inégalité triangulaire, (ii) la symétrie et (iii) la réflexivité.

(i) $dist(v1,v2) + dist(v2,v3) >= dist(v1,v3)$

(ii) $dist(v1,v2) = dist(v2,v1)$

(iii) $dist(v1, v2) = 0 \Leftrightarrow v1 = v2$

Dans ce cadre de recherche, plusieurs mesures de similarités ont été proposées dans la littérature et définies pour des valeurs scalaires, ensemblistes, vectorielles, etc. On peut citer, entre autres, la différence absolue, cosinus, Harman, Dice, Jacquard, degré d'inclusion, produit scalaire, Manhattan, Hamming, Euclidienne, Hausdorff [91].

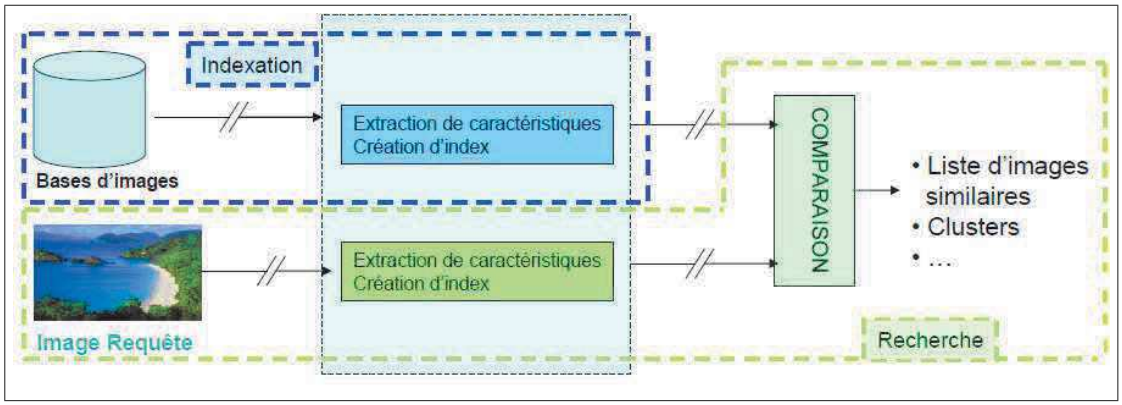


FIGURE 1.4 : Principe général de la recherche d'images par le contenu et les caractéristiques visuelles de bas niveau (couleur, forme, texture) d'une image requête donnée par l'utilisateur.

1.5.1.2 Traitements de bas niveau : extraction des descripteurs

Il existe dans la littérature un nombre très important de descripteurs de bas niveau en vue de caractériser ou résumer le contenu visuel d'une image. Les descripteurs permettent de représenter généralement une petite imagette, centrée en un pixel donné, par un vecteur de caractéristiques de moins grande taille que le nombre de pixels contenus dans cette imagette et possédant des propriétés intéressantes d'invariance (au bruit, changement d'illumination, déformation, occlusions, etc.) de la scène considérée. Ci-dessous, nous présentons les principaux types de descripteurs et les méthodes utilisées pour les extraire.

1.5.1.2.1 Descripteurs de couleurs

La caractéristique de couleur est largement utilisée pour la recherche d'images par le contenu. Il existe plusieurs espaces colorimétriques pour représenter la couleur telle que RGB, HSV, LAB, XYZ, YCbCr, etc. (cf. Fig 1.5). Il n'y a pas d'espace couleur idéal pour l'indexation, mais chaque espace couleur peut intégrer des propriétés d'invariance, relativement à l'ombre, au changement d'illumination ou au bruit, qui peut s'avérer intéressante pour une indexation réussie. Une analyse détaillée des différents espaces de couleurs est donnée dans le chapitre 3.

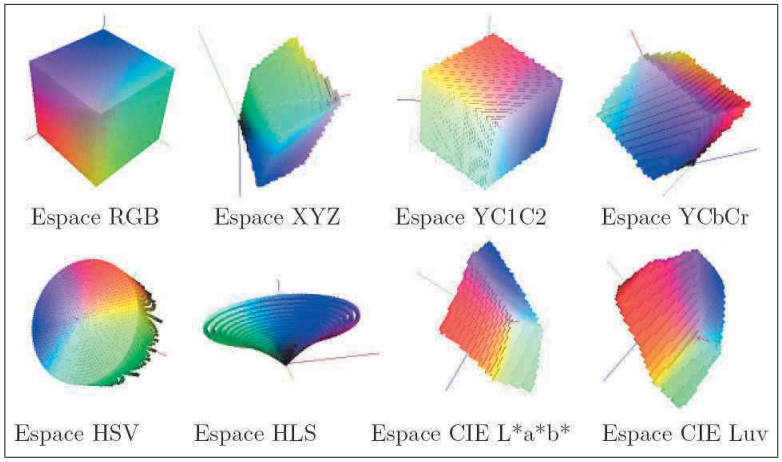


FIGURE 1.5 : Répartition des couleurs au sein des principaux espaces couleurs.

Les premières signatures couleur d'une scène ou imagette texturée ont été réalisées par des histogrammes de couleurs (représentant la fréquence d'apparition des différentes couleurs) requantifiés en un certain nombre de bins réduit. Le succès de l'histogramme couleur requantifié résulte de sa simplicité de calcul et d'utilisation, sa capacité à caractériser une texture par un mélange de couleur et enfin sa robustesse à la rotation, aux changements d'échelle et aux occlusions de la scène. Cependant, l'utilisation de l'histogramme couleur introduit deux limitations importantes. Le premier concerne la perte de l'information spatiale et sémantique de la scène, caractéristiques importantes pour décrire le contenu global des images. Le second inconvénient est lié à la quantification de l'espace des attributs [92]. Afin de bien comprendre ce problème, prenons l'exemple de la Fig. 1.6. Les images (a) et (b) possèdent des histogrammes de couleurs très proches

alors qu'elles n'ont aucun rapport sémantique. Alors que, les images (b) et (c), qui sont sémantiquement très similaires, possèdent des histogrammes de couleurs très différents. En résumé, deux textures différentes peuvent avoir le même histogramme couleur et deux objets identiques, sémantiquement parlant, peuvent avoir deux histogrammes couleur très différents.

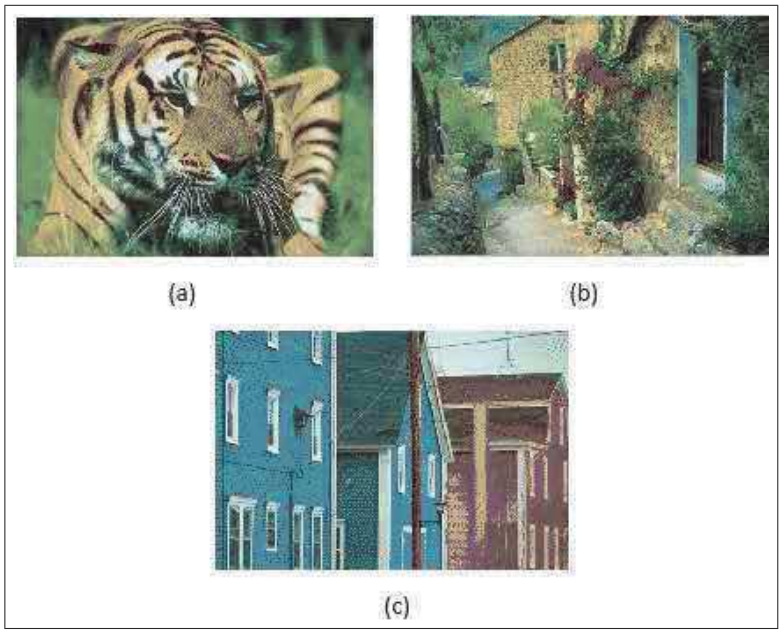


FIGURE 1.6 : Les images (a) et (b) possèdent des histogrammes de couleurs similaires. Les images (b) et (c) possèdent des histogrammes de couleurs très différents.

Dans un premier temps, afin de limiter la perte de l'information spatiale de la technique de l'histogramme couleur, les auteurs dans [93] ont proposé de diviser l'image en un ensemble de blocs égaux, puis de les comparer à tous les blocs des images de la base de données afin de retrouver les images similaires. Dans [94], les auteurs ont proposé d'utiliser les corrélogrammes des couleurs. Un corrélogramme des couleurs est un histogramme à trois dimensions, dont la première et la seconde sont les distributions des paires de pixels et la troisième dimension est leur distance spatiale [95]. Pour leur part, [93] ont proposé les vecteurs de cohérence de couleurs. Ces derniers classifient chaque pixel en cohérent ou non cohérent selon l'appartenance de ce pixel à une région homogène de couleur. D'autres approches ont été proposées pour représenter l'attribut

couleur tel que les moments de couleurs (la moyenne, la variance et l'asymétrie) [96] et les ensembles de couleur (color sets) [97].

1.5.1.2.2 Descripteurs de texture

Caractériser la texture de l'image fournit une bonne information sur l'arrangement structurel des surfaces dans l'image. La définition littéraire de la texture est la "*répétition spatiale d'un même motif dans différentes directions de l'espace*". Plus précisément, la texture peut être vue comme un ensemble de pixels spatialement agencé selon un certain nombre de relations spatiales, ainsi créant une région homogène. Plusieurs techniques ont été développées pour l'analyser, ces techniques peuvent être divisées en quatre classes suivantes : [98, 99].

(a) **Les méthodes statistiques :** analysent la distribution spatiale des valeurs de niveaux de gris. Ces méthodes incluent les matrices de co-occurrence des niveaux de gris [100], les fonctions d'autocorrélation et les variogrammes [101].

(b) **Les méthodes fréquentielles :** souvent utilisées en traitement du signal, analysent les fréquences de l'image. Parmi elles, on retrouve les filtres de Gabor [102], la transformée de Fourier [103], les décompositions en ondelettes [104], etc.

(c) **Les méthodes dites géométriques ou structurelles :** ces méthodes permettent de caractériser les propriétés géométriques des textures et trouver les règles qui régissent leur organisation spatiale.

(d) **Les méthodes basées sur un modèle :** ces méthodes décrivent la texture par un modèle probabiliste. Celle-ci est alors caractérisée par les paramètres de ce modèle. Parmi les modèles les plus utilisés, nous retrouvons les champs de Markov (MRF), les fractales [105], les modèles de différentiation fractionnaire et les modèles autorégressifs (AR) [99].

1.5.1.2.3 Descripteurs de formes

La forme est l'un des attributs de bas niveau le plus utilisé pour décrire la structure géométrique générique du contenu visuel. Elle est intéressante pour retrouver certains concepts qui ne peuvent l'être autrement. Prenons par exemple, le cas d'un *ballon*. Il n'y a pas de couleurs spécifiques qui puissent le caractériser, chacun à sa couleur. Un autre exemple est celui d'une *panthère*, il y a des panthères de couleur unie et des panthères au pelage tacheté, par contre, ces deux exemples (ballon et panthère) ont une forme très caractéristique. La forme est donc une information discriminante qui peut être utile pour réduire le fossé sémantique.

Les caractéristiques de la forme sont extraites à partir des régions qui ont été segmentées préalablement. Une analyse très détaillée des différentes méthodes de segmentation est donnée dans la section 1.2. En général, les représentations de la forme peuvent être divisées en deux catégories : basées sur les contours ou basées sur les régions. Pour la première approche, seule la frontière externe entre objets ou régions est exploitée, tandis que pour la seconde, la forme de toute la région est prise en considération [106]. Nous présentons dans ce qui suit quelques méthodes de description des formes.

A) Les descripteurs de forme simples

Il existe de nombreux descripteurs qui sont très facilement calculables à partir du contour de l'image. On peut citer : l'aire (nombre de pixels dans l'objet incluant le contour) , le périmètre (la longueur en pixel de la circonférence de l'objet), longueur de l'axe majeur (la ligne qui connecte les deux points extrêmes), longueur de l'axe mineur (Perpendiculaire à l'axe majeur et les 4 points formant un rectangle qui contient la frontière), l'excentricité (le ratio de la longueur de l'axe majeur à celle de l'axe mineur de la forme), la rectangularité (le rapport de l'aire de la forme à l'aire du plus petit rectangle englobant), la variance circulaire et elliptique, la circularité, la compacité, la courbure, la distance au centroïde, l'énergie de flexion (la somme des carrées des courbures à chaque point de contour), etc. (cf. Fig. 1.7) [7, 106].

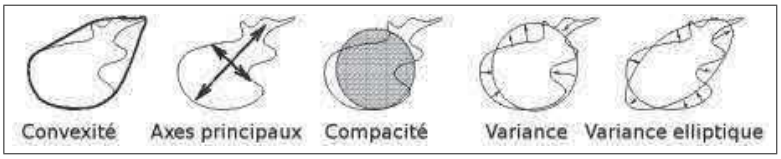


FIGURE 1.7 : Exemples de descripteurs géométriques [7].

B) Les moments invariants

Les moments invariants ont été utilisés comme descripteurs de forme 2D pour caractériser la forme de la région extraite d'une image 2D. On rencontre dans la littérature plusieurs types d'invariants, cependant les plus utilisés sont les invariants d'Hu aussi appelés moments géométriques et les moments de Zernike.

- **Les moments géométriques invariants :** Hu en 1962 [107] a proposé sept moments invariants qui sont indépendants à la déformation géométrique (échelle, translation, rotation). Ces descripteurs sont depuis largement utilisés en indexation [108–110].

Soit deux entiers $n, m \in \mathbb{Z}$ et une image en niveau de gris *I(x, y)*, les moments géométriques se calculent de la manière suivante :

$$m_{nm} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} I(x,y) x^n y^m d_x d_y \tag{1.1}$$

Il faut noter qu'au niveau de l'équation (1.1) les moments géométriques obtenus n'assurent pas les différentes variations géométriques (translation, l'échelle et rotation). Les moments invariants seront obtenus en utilisant la notion des moments centraux μ_{nm} exprimés par rapport au centre de gravité de l'image de coordonnée $(\bar{x}, \bar{y})$, avec $\bar{x} = m_{10}/m_{00}$ et $\bar{y} = m_{01}/m_{00}$.

$$\mu_{nm} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x-\bar{x})^n (y-\bar{y})^m I(x,y) d_x d_y \tag{1.2}$$

L'invariance à l'échelle s'obtient en normalisant ces moments. On obtient alors les moments centraux normalisés η_{nm} définis par :

$$\eta_{nm} = \mu_{nm} / \mu_{00}^{\lambda} \tag{1.3}$$

avec $\lambda = (n+m+1)/2$ et $n, m = 0, 1, 2...$

À partir des moments centrés normalisés, nous pouvons calculer un ensemble de sept paramètres invariants. Ces sept moments invariants sont :

$$\phi_1 = \eta_{20} + \eta_{02} \tag{1.4}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \tag{1.5}$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \tag{1.6}$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \tag{1.7}$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} - \eta_{03})[3(\eta_{30} - \eta_{12})^2 - (\eta_{21} - \eta_{03})^2]\end{aligned} \tag{1.8}$$

$$\begin{aligned}\phi_6 = {} & (\eta_{20} - \eta_{02})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} - \eta_{12})(\eta_{21} - \eta_{03})\end{aligned} \tag{1.9}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned} \tag{1.10}$$

- **Les moments invariants de Zernike :** Introduit par Teague [111], cette approche consiste à projeter l'image dans un espace vectoriel en utilisant un ensemble de polynômes orthogonaux (les polynômes de Zernike) [112]. L'invariance par rapport à la translation est obtenue en prenant le centre de gravité comme origine des coordonnées. Le premier moment régulier est normalisé pour obtenir l'invariance par rapport à l'échelle [7, 113].

C) Les descripteurs de Fourier

La transformée de Fourier discrète (TFD) est particulièrement intéressante pour l'in-

dexation car elle permet de décrire la forme de l'objet à différents niveaux de détails. Les descripteurs de Fourier sont calculés à partir du contour des objets. Si on considère les n points du contour consécutifs comme une fonction discrète $x(n) = (x_1(n), x_2(n))$, on peut définir une fonction discrète complexe $u(n) = x_1(n) + jx_2(n)$. Cette fonction peut être transformée dans le domaine fréquentiel par une TFD. Les coefficients de cette transformation, appelés les descripteurs de Fourier sont quelquefois utilisés pour représenter le ou les formes présentes dans l'image [106].

1.5.1.2.4 Détecteurs et descripteurs de points d'intérêt

Certaines méthodes d'indexation combinent aussi les descripteurs précédents de bas-niveau avec des descripteurs de moyen-niveau qui utilisent une mesure de similarité sur la distribution des points particuliers ou d'intérêts préalablement détectés sur l'image requête et les images candidates ou alors sur une ressemblance locale entre descripteurs qui sera alors calculé sur une région autour du point d'intérêt [114].

L'idée d'utiliser ces points remonte aux travaux de Hans Moravec en 1981 [115] sur la recherche de correspondance entre images stéréoscopiques. Ce détecteur souffre de nombreuses limitations ce qui lui vaut de fonctionner dans un contexte limité. Harris et Stephen ont identifié certaines de ces limitations et, en les corrigeant, ont proposé en 1988 un détecteur de coins connu sous le nom de détecteur d' Harris [116]. Il se base sur une fonction d'autocorrélation du signal c'est-à-dire sur les changements du signal dans plusieurs directions [117]. Néanmoins leur méthode n'était pas invariante à l'échelle et aux transformations affines. Par ailleurs, David Lowe a popularisé le concept de points d'intérêt en développant le détecteur et descripteur appelé Scale Invariant and Feature Transform (SIFT) [118], est une méthode permettant de décrire un point à partir des orientations locales du gradient [119]. Plus précisément, Un descripteur SIFT est un vecteur d'entiers de taille 128 qui décrit une zone $16x16$ autour d'un point d'intérêt, l'ensemble de ces descripteurs établissent ainsi une véritable signature numérique du contenu de l'image. Les vecteurs de description construits par cette méthode présentent l'avantage d'être robustes aux déformations géométriques usuelles et aux varia-

tions d'intensité. Un peu plus récemment en 2006, une variante appelée SURF (Speeded Up Robust Features) [120] est apparue. Cette méthode s'inspire de SIFT et utilise des approximations pour augmenter la rapidité des traitements, notamment grâce à l'utilisation d'images intégrales. Une autre variante est celui de PCA-SIFT [121], dans cet algorithme les auteurs utilisent une réduction de dimensions par analyse en composantes principales pour réduire la grille des gradients locaux à un vecteur de dimension beaucoup plus petite [119].

1.5.2 Visualisation

Dans le domaine de la recherche d'images, la problématique de la visualisation des grandes collections ou bases d'images représente un des challenges très importants, car elle doit permettre d'utiliser efficacement un espace d'affichage limité et donner à l'utilisateur une vue d'ensemble, tout en ayant la possibilité d'accéder à certains détails. Il s'agit en effet de visualiser les résultats (images) pour les utilisateurs sur une interface ou *mapping* (principalement 2D ou 3D) après une phase de recherche [95]. La façon la plus populaire de présenter des images est un *mapping* 2D matrice rectangulaire de vignettes (cf. Fig. 1.8). Cela permet la navigation de telle manière que l'utilisateur puisse comprendre rapidement et intuitivement quels types d'images sont contenus dans la base de données et leur distribution pour une analyse plus poussée. Dans une base d'images médicales de cerveaux sains et malsains (pour une pathologie donnée), cette distribution d'images devrait permettre d' analyser et comprendre les différents modes de pathologie et de degré de sévérité d'une maladie donnée (Alzheimer, schizophrenia, etc.). Généralement, cette distribution est typiquement réalisée par une technique basée sur la cartographie telle que l'analyse en composantes principales (ACP) ou la mise à l'échelle multidimensionnelle (MDS) à l'aide d'une métrique entre les caractéristiques de couleur ou de texture de bas niveau et calculée pour chaque paire d'images de la base de données. Plus précisément, MDS est avant tout une technique de réduction de dimensionnalité non linéaire qui tente de trouver un nouveau mapping de dimension peu élevée (2D ou 3D) tel que les distances, initialement définies entre chaque pair d'images, se conservent le mieux possible dans l'espace réduit. Cette conservation des distances se fait au sens des

mondres-carrées pour le MDS. Les idées fondamentales de MDS furent d'abord proposées par Young et Householder [122] puis développées par Torgerson [123] où son algorithme original (appelé MDS classique) calcule la "meilleure" projection de l'ensemble d'objets dans un sous-espace de dimension donnée.

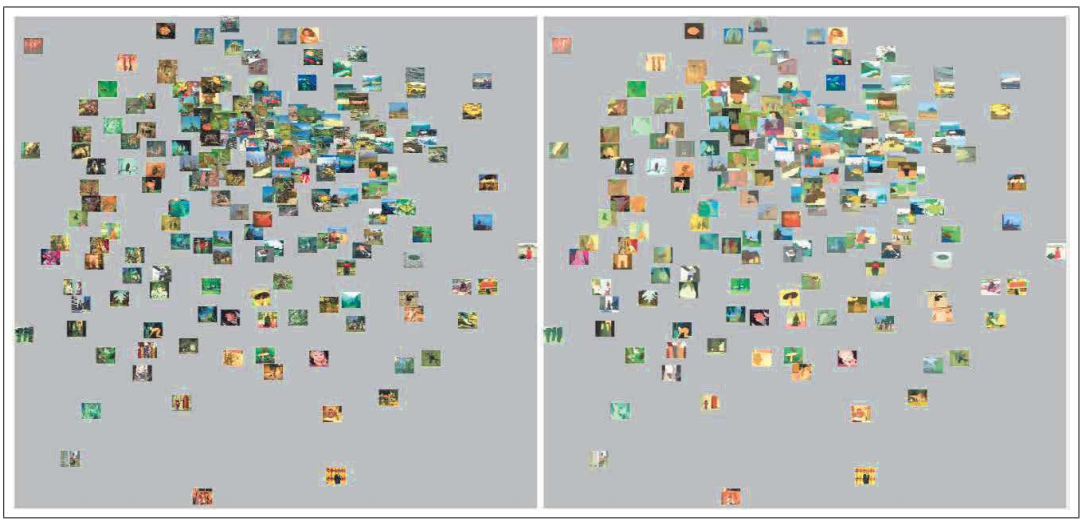


FIGURE 1.8 : Une carte de visualisation MDS du BSDS300 basée sur la distance VoI entre les segmentations [8]. A gauche : voir la carte des (300) images en couleur mappées en fonction de leur similarité en termes de résultat de segmentation en fonction de la région. Droite : voir la carte des résultats de la segmentation en fonction de leur similarité dans le sens de la distance VoI.

1.5.3 Conclusion

La recherche d'images par le contenu est un domaine qui attire de plus en plus l'attention des chercheurs, car cette technique permet aussi de naviguer ou de structurer une grande base de données d'une certaine façon afin d'en faciliter une tâche subséquente de classification ou de compréhension (de l'image, ou d'un groupe d'images) et ainsi possiblement d'inférer un ensemble de relations ou d'hypothèse de plus haut niveau d'abstraction. La problématique est de bien comprendre l'intention des utilisateurs et de trouver des descripteurs efficaces pour bien exprimer le contenu de l'image. Jusqu'à présent, les descripteurs utilisés pour la recherche d'image par le contenu ont été majoritairement basés sur la prise en compte directe du contenu visuel (couleur, forme, texture) de bas niveau d'abstraction et de ce fait, ont des problèmes pour prendre en

considération des informations de plus niveau d'abstraction, tel que la sémantique de l'image. Cette lacune est à l'origine du problème de *fossé sémantique*, défini comme le fossé entre la représentation bas niveau d'une image et l'interprétation haut niveau que les humains en font. Le défi actuel consiste donc à améliorer les descripteurs visuels afin de réduire ce fossé sémantique et de passer à l'utilisation de caractéristiques de plus hauts niveaux d'abstraction (possiblement en complémentarité avec les descripteurs visuels de bas niveau communément utilisés). Une possibilité est d'interpréter une image donnée en fonction des régions préalablement segmentées, leur forme et leur distribution dans l'image et plus précisément de définir une vraie distance, sémantiquement intelligente, entre deux images segmentées, prenant en compte l'interaction spatiale de ces formes préalablement segmentées pour l'indexation et la navigation dans une grande base de données. Cette étude constituera la première contribution de notre thèse (cf. Chap 2).

1.6 État de l'art sur la détection/segmentation du mouvement

La détection de mouvement a pour but de décider quelles parties des images (d'une séquence ou d'une vidéo) correspondent à des objets mobiles.

C'est une étape indispensable pour de nombreuses applications de vision par ordinateur, car ces zones de mouvement correspondent souvent à des évènements sur lesquels un système de vision doit se focaliser. Parmi ces diverses applications, on peut citer la vidéo surveillance, les applications propres à la robotique (détection d'obstacles, égo-mouvement, etc.), le suivi d'objets, la reconstruction 3D, la détection de chute, etc. C'est pour cette raison que ce problème a fait l'objet de nombreux travaux depuis les années 80. Ainsi, un grand nombre d'algorithmes de détection du mouvement a déjà été proposé.

Si on considère le problème de détection de mouvement comme un problème de segmentation consistant à séparer ou classer les pixels en 2 classes distinctes, l'arrière-plan (background ou zone immobile) et l'avant-plan (foreground ou zone mobile). De nombreuses méthodes ont été proposées dans la littérature. Parmi celles-ci, on peut distinguer les méthodes de détection basées sur la soustraction de fond, celles basées sur

des mesures des zones de mouvement et enfin celles qui s'appuient sur un calcul du flot optique. Toutes ces différentes techniques ont leurs propres avantages et leurs limites.

1.6.1 Méthodes basées sur la modélisation de l'arrière-plan

Plusieurs méthodes de détection de mouvement par soustraction de l'arrière-plan ont été proposées, la détection d'objets mobiles par ces méthodes se fait ensuite par une opération de soustraction entre l'image courante $I(x,y,t)$ et un modèle d'arrière-plan $B(x,y)$ préalablement estimé contenant les éléments statiques de la scène.

Si $\mid I(x,y,t) - B(x,y) \mid \geq h$ alors le pixel de coordonnées (x,y) à l'instant t est en mouvement, sinon ce pixel appartient à l'arrière-plan, ou h est le seuil sélectionné. Parmi les méthodes de soustraction de fond les plus connues, on peut citer :

Méthode de la moyenne temporelle : Dans cette méthode le modèle de l'arrière-plan est estimé pour chaque pixel $I(x,y,t)$ de l'image I à un instant t. On considère que chaque image de la séquence contribue à la mise à jour de l'arrière-plan. La valeur du pixel de cet arrière-plan est donnée par l'équation suivante :

$$B(x,y,t) = \frac{(t-1)B(x,y,t-1) + I(x,y,t)}{t} \tag{1.11}$$

L'avantage de cette approche est qu'elle est simple et rapide. Parmi les problèmes majeurs de cette technique est le nombre d'images nécessaires que l'on doit stocker dans une mémoire tampon afin d'estimer l'arrière-plan, la variation de luminosité, mouvement de la caméra et les ombres portées des objets.

Méthode de mixture de distributions gaussiennes : Stauffer *et al.* [124] ont proposé une méthode adaptative basée sur un modèle défini par un mélange de gaussiennes. La méthode est constituée de deux étapes principales : la première étape cherche à modéliser la distribution de niveaux de gris (ou de couleur) de chaque pixel le long de la séquence d'images par un mélange de K distributions gaussiennes (généralement compris entre 3 et 5) afin de tenir compte, à la fois, du bruit gaussien, du changement graduel de la

luminosité et de la présence de petits mouvements (par exemple : feuillage d'un arbre). La seconde étape vise à décider de la classe de chaque pixel, soit comme un pixel en mouvement (si son niveaux de gris est en dehors de ce mélange de gaussienne) ou bien comme étant un pixel de l'arrière-plan. L'un des problèmes majeurs de cette technique est l'initialisation manuelle du nombre de gaussiennes.

Modèle de mélange gaussien adaptatif : Dans la méthode de mélange de gaussiennes proposée par Stauffer *et al.* il faut fixer dès le début le paramètre K, qui est le nombre de composantes du mélange. Dans [22], Zivkovic à proposé une estimation automatique de K à chaque image en fonction de la dynamique de la scène (les scènes dynamiques sont modélisées à l'aide de plusieurs gaussiennes alors que les scènes statiques avec un nombre réduit de gaussiennes).

1.6.2 Méthodes basées sur la mesure du mouvement

Plusieurs méthodes de détection d'objets mobiles par mesure du mouvement ont été proposées dans la littérature. Parmi ces méthodes, on peut citer :

Méthodes différentielles : Les approches basées sur la différence entre images ne nécessitent pas de modèles d'arrière-plan. En effet, la présence de mouvements dans une image peut se définir comme un changement de la valeur de l'intensité d'au moins un pixel entre deux images consécutives dans une séquence vidéo. L'approche la plus simple consiste donc à observer la différence pixel par pixel de deux trames consécutives dans un flux vidéo.

Si $\mid I(x,y,t) - I(x,y,t-1) \mid \geq h$ alors le pixel de coordonnées (x,y) à l'instant t est en mouvement, sinon ce pixel appartient à l'arrière-plan, ou h est un seuil sélectionné. L'avantage de cette approche est qu'elle est simple et rapide. L'un des inconvénients de cette technique est qu'elle ne permet pas de détecter le mouvement dans les zones uniformes intérieures à l'objet. Pour pallier ces problèmes, certains auteurs suggèrent l'utilisation d'une double différence à partir de trois images consécutives.

Méthodes basées sur l'estimation du flot optique : Les méthodes de détection de mou-

vement par flot optique ont pour but de caractériser le mouvement dans une scène. Ils permettent de déterminer pour chaque pixel un vecteur de déplacement entre deux images consécutives. Le principe des méthodes de flot optique est le suivant, entre deux images successives, la luminance I d'un objet doit rester la même. Cela revient à estimer v tel que :

$$I(x,t) - I(x+vdt, t+dt) = 0 \tag{1.12}$$

1.6.3 Amélioration des méthodes de détection du mouvement

Dans le but d'améliorer les méthodes de détection du mouvement, nous verrons dans cette thèse qu'il est très avantageux d'utiliser le principe de fusion de segmentation expliqué à la Section 1.3 pour combiner soit efficacement de multiples résultats de cartes de détection de mouvement obtenus par différents modèles de détection du mouvement, mais surtout, plus simplement, les résultats obtenus par le même modèle de soustraction de fond, opérant dans des espaces de couleurs différentes et complémentaires. Nous quantifierons les gains obtenus par cette stratégie et détaillerons la façon d'estimer la paire d'espace couleur la plus appropriée pour une méthode de soustraction de fond donnée que l'on désire améliorer.

1.7 Détection et classification en imagerie médicale

Les banques d'images issues de l'imagerie médicale sont de plus en plus volumineuses, créant ainsi l'opportunité d'effectuer des analyses sur les variations anatomiques et l'évolution de certaines maladies (par ex. l'Alzheimer, cancer du sein et l'épilepsie, etc.). La technique de segmentation consensuelle (cf. Sec 1.3) nous permettra de traiter ces données d'une toute nouvelle façon afin de pouvoir nous fournir un meilleure détection et diagnostic de ces maladies.

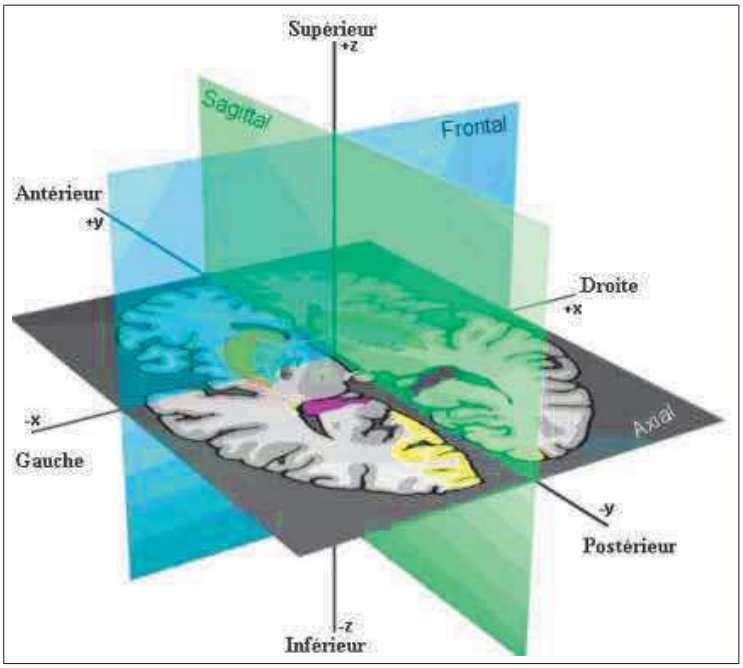


FIGURE 1.9 : Définition des plans de coupe en imagerie médicale

1.7.1 Théorie de l'imagerie par résonance magnétique

L'imagerie par résonance magnétique a été principalement développée vers 1980. L'IRM est un examen qui permet d'obtenir des vues en deux ou trois dimensions (frontal, sagittal et axial (cf. Fig 1.9)) de l'intérieur du corps de façon non invasive avec une résolution en contraste relativement élevée. Le corps humain contient un grand nombre d'atomes d'hydrogène dont le noyau est composé d'un unique proton (chargé positivement) qui est en mouvement. Tous ces protons peuvent être considérés comme des petits aimants. En utilisant des aimants très puissants, ces protons sont stimulés simultanément, ce qui a pour effet d'orienter tous ces atomes dans la même direction. Une fois cette première étape accomplie, les atomes sont mis en résonance en leur faisant subir une excitation par un champ magnétique (radiofréquence). Cette mise en résonance produit une rotation des petits aimants du patient. Le retour à l'état normal produit une variation du champ magnétique. Dans une antenne correctement placée, cette variation du champ magnétique induit la formation du courant électrique : c'est le signal IRM. Lors d'un examen IRM, c'est l'analyse de ce signal par un ordinateur qui permet d'ob-

tenir les images des différentes parties du corps humain (cf. Fig 1.10).

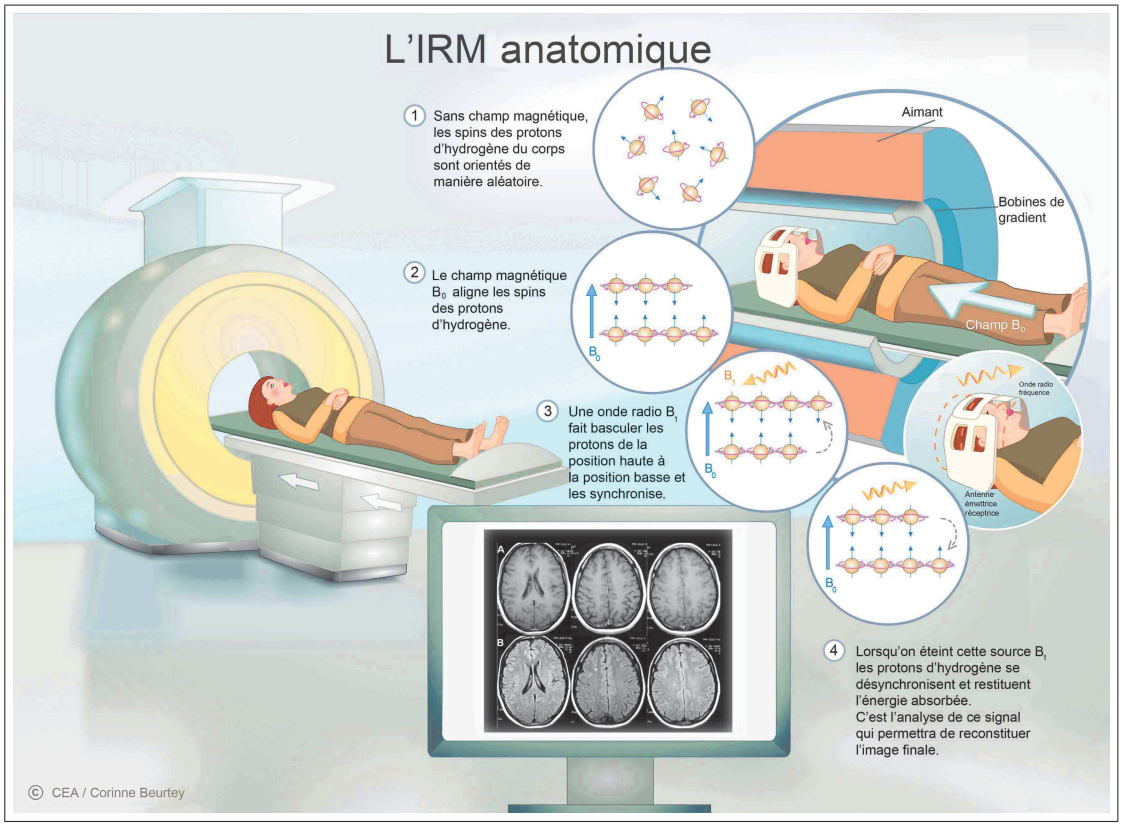


FIGURE 1.10 : Principe de fonctionnement de l'IRM

1.8 Structure du document

1.8.1 Plan de la thèse

Dans cette thèse par articles, les contributions sont organisées en trois parties :

Partie 1 : Consensus de segmentation dans le cadre de la visualisation, navigation et la recherche dans les grandes bases de données d'images

Le chapitre 2 présente notre première contribution avec un article et propose d'utiliser le principe de moyennage consensuel de segmentation au sens du VoI dans le cadre de la navigation et la recherche dans des grandes bases de données d'images. Cette navigation se fait grâce à un nouveau critère puisqu'il permet aux utilisateurs de naviguer

efficacement et de rechercher en fonction, non plus de la répartition des couleurs ou de la similarité entre textures existant dans les images, comme il est communément fait dans la littérature mais en fonction du contenu descriptif (*i.e.* la segmentation) de chaque image et plus précisément, sur la base de la disposition géométrique des différentes formes d'objets préalablement détectés et segmentés dans la scène. Ces informations descriptives, fournies à un niveau d'abstraction plus élevé, peuvent être une information importante et complémentaire qui permet à l'utilisateur de chercher et naviguer dans la collection d'images de manière plus intuitive et efficace. En outre, nous étudions et analysons différents algorithmes et méthodes utilisant la distance et le moyennage au sens du VoI permettant de regrouper, classer ou trouver rapidement une image ou un (sous-)ensemble (ou classe) d'images en fonction de son contenu descriptif.

Partie 2 : Consensus de segmentation pour améliorer la précision des algorithmes de détection de mouvement

Le chapitre 3 présente notre deuxième contribution avec un article portant sur la détection des objets mobiles dans une séquence d'images. Nous proposons un nouveau système qui consiste à améliorer facilement n'importe quelle méthode de détection de mouvement existante. Le principe consiste à fusionner différentes cartes de segmentations issues du même algorithme de détection de mouvement, fonctionnant dans des espaces de couleurs différentes. Nous avons développé une méthode permettant d'estimer la paire (ou le sous-ensemble) d'espace couleur la plus complémentaire et optimal pour le modèle de détection considéré dont la fusion des résultats de détection (obtenue dans les différents espaces couleur préalablement estimés) permet d'obtenir une segmentation plus fiable et plus précise. L'idée sous-jacente derrière cette fusion de données est de tirer avantage de la complémentarité (au bruit, changements d'éclairage, informations sur les couleurs, highlights, etc.) des différents résultats de détection du mouvement lorsque ceux-ci sont exprimés par différents espaces couleurs.

Partie 3 : Consensus de segmentation pour la classification d'imagerie médicale : application au diagnostic de la maladie d'Alzheimer

Les chapitres 4 et 5 présentent notre troisième contribution avec deux articles portant

sur la classification des images cérébrales IRM par le contenu pour l'aide au diagnostic de la maladie d'Alzheimer. À cette fin, nous proposons un nouveau système de diagnostic assisté par ordinateur (CAO) qui utilise le concept de consensus de segmentation pour construire deux prototypes 2D du cerveau (prototype sain et prototype malsain) en termes de la substance grise, substance blanche et fluide cérébro-spinale. D'abord, la segmentation en trois régions a le mérite de réduire efficacement le contenu d'information d'une image cérébrale et de supprimer le bruit et les artefacts qui ne sont pas pertinents pour la détection de l'Alzheimer et qui peuvent dégrader la performance de la classification. Deuxièmement, en tant que modèle de consensus, cela nous permet également de supprimer les composants indésirables de l'image cérébrale tels que la variabilité anatomique existant entre les individus (du même groupe) qui ne sont pas pertinents pour la détection et la quantification de cette maladie. Le processus de classification fait intervenir ces prototypes ou atlas segmentés et la distance entre segmentations et un mélange de deux classifieurs ; le premier basé sur celui de la distance minimale et le second basé sur les K-plus proche voisins.

1.9 Publications

Les principales communications dans des conférences et journaux internationaux reliées à nos travaux sont les suivantes :

- **Travaux sur le consensus de segmentation dans le cadre de la visualisation, navigation et la recherche dans les grandes bases de données d'images**
  - A. Khlif, M. Mignotte. Segmentation data visualizing and clustering. *Multimedia Tools and Applications (MTA)*, 76 (1) :1531-1552, January 2017.
- **Travaux sur le consensus de segmentation pour améliorer la précision des algorithmes de détection de mouvement**
  - A. Khlif, M. Mignotte. Change Detection Algorithm Improvement By Adaptive Fusion of Results Obtained From Different Color Spaces. Submitted to *International Journal of Image and Data Fusion*, August 2017.

- **Travaux sur le consensus de segmentation pour la classification d'imagerie médicale : application au diagnostic de la maladie d'Alzheimer**

  - A. Khlif, M. Mignotte. Classification of Alzheimer's disease subjects from MRI using the principle of consensus segmentation. *19th Irish Machine Vision and Image Processing conference, IMVIP'17*, published by the Irish Pattern Recognition & Classification Society, pages 123-130, Maynooth University, Co. Kildare, Ireland, September 2017.

  - A. Khlif, M. Mignotte. Classification of Alzheimer's disease from MRI using the principle of consensus segmentation based atlas. *Computerized Medical Imaging and Graphics* (Elsevier), February 2018.

Première partie

Consensus de segmentation dans le cadre de la visualisation, navigation et la recherche dans les grandes bases de données d'images

CHAPITRE 2

SEGMENTATION DATA VISUALIZING AND CLUSTERING

Cet article a été pubilé dans le journal *Multimedia Tools and Applications (MTA)* comme l'indique la référence bibliographique.

A. Khlif, M. Mignotte. Segmentation data visualizing and clustering, *Multimedia Tools and Applications (MTA)*, 76 (1) :1531-1552, January 2017.

Cet article est presenté içi dans sa version originale.

Abstract

Browsing, searching and retrieving images from large databases based on low level color or texture visual features have been widely studied in recent years but are also often limited in terms of usefulness. In this paper, we propose a new framework that allows users to effectively browse and search in large image database based on their segmentation-based descriptive content and, more precisely, based on the geometrical layout and shapes of the different objects detected and segmented in the scene. This descriptive information, provided at a higher level of abstraction, can be a significant and complementary information which helps the user to browse through the collection in an intuitive and efficient manner. In addition, we study and discuss various ways and tools for efficiently clustering or for retrieving a specific subset or class of images in terms of segmentation-based descriptive content which can also be used to efficiently summarize the content of the image database. Experiments conducted on the Berkeley Segmentation Datasets show that this new framework can be effective in supporting image browsing and retrieval tasks.

Keywords : Berkeley dataset, clustering algorithm, entropy, database browsing and retrieving images, hierarchical clustering, K-means, multidimensional visualization, query-by-drawing, segmentation data clustering, descriptive content based image classification, variation of information, visualization of image databases.

2.1 Introduction

Clustering is the task of grouping together, in a feature space, data samples in the same group (or cluster) that are similar in the sense of a given distance measure. It is the main task of data exploration and analysis which is then useful for looking for patterns or structures (and/or correlations) in the data that are of interest.

Image segmentation is a special case of clustering problem where the grouping of image data (or pixels) into clusters must take into account not only their similarity in the feature space but also the requirement of their spatial coherence, since an essential feature of image data is the spatial ordering of its pixels. In the image segmentation case, structures of interest are spatially coherent regions such as consistent parts of objects or of the background sharing similar attributes. The detection and localization of these regions, with a spatial clustering, allows to change and simplify the representation of the image into something that is both compact, and also easier to analyze since this economic description, in terms of detecting homogeneous regions, efficiently describes the geometric content of the input image.

In these two latter cases, the clustering performance and its subsequent usefulness closely depend on the used distance measure. In this sense, depending on this similarity metric, the application herein proposed, consisting of achieving an unsupervised clustering of segmentation results (estimated, for example, from an image database) could either be meaningless or have a certain interest. Indeed, in such application, it is crucial to consider an appropriate distance, across the lattice of possible spatial clusterings, for efficiently comparing two segmentation results in a objective, reliable, comprehensible and perceptual criterion sense which should be also capable of taking into account the inherent variability of each possible perceptually consistent interpretation/segmentation of an input image (possibly segmented at different detail levels by different human segmenter). Nevertheless, it is not trivial to find a true (in the mathematical sense) and meaningful distance between two segmentation maps. Indeed, the two segmentations might have a different number of segments and the correspondence between segments are not known and this is especially true in the case of segmentations obtained from

two different images. If such a distance exists, an unsupervised clustering of segmentation results could be efficiently used for obtaining an overview or for browsing and/or retrieving images from large image collections or database or for retrieving a specific subset or class of images in terms of segmentation-based descriptive content (such as the layout and similar arrangement of the different objects segmented in the image) and not in terms of low-level features such as color and/or texture as it is commonly used in image database browsing applications. Moreover, a distance-based clustering can also be exploited, among other things, in order to build an image database with a substantial amount of diversity in the dataset. This can be easily achieved by automatically removing images with to much similar content in terms of region-based descriptive content or to quantify this amount of diversity with a dispersion measure such as the average of all distances obtained for each segmentation pair.

In addition, an automatic clustering procedure often requires the estimation of the center of each cluster. It is the case of the K-means clustering procedure which is the most commonly used clustering algorithm so far proposed in the literature [125] or more generally, also the case of all class of (distance-based) clustering algorithms known as iterative refinement algorithms. To this end, the estimation of cluster centers, in a non-Euclidean and non-standard distance sense, when the underlying data are segmentation results, is far from being trivial. Fortunately, this problem has been recently solved in the image processing community in order to provide an interesting alternative to the existing complex and computationally costly segmentation models. Indeed, an effective, simple and commonly accepted segmentation strategy consists in averaging (*i.e.*, efficiently combining or fusing) multiple quickly estimated segmentation results (of the same scene) obtained from some simple segmentation models (or by the same segmentation algorithm with different values of its internal parameters) to achieve a final improved segmentation result. This strategy has initially been introduced in [81, 82] with the restriction that all input segmentations (to be averaged or fused) should contain the same number of regions and then a little later without this restriction, with an arbitrary number of regions [5,6,19,83,85] in different criteria senses with algebraic, analytic or stochastic

procedures. In the case of unsupervised clustering of segmentation results, these cluster centers or prototypes could be efficiently used to quickly visualize the region-based descriptive content of each class of segmentation map ensemble associated to a given image database. Also, these cluster centers could estimate a *clustering meaningfulness* distance allowing to quantify the diversity in the dataset or the performance of each clustering algorithm, among others things. It is true that the K-means procedure and more generally, all iterative refinement clustering algorithms have also a major shortcoming in the fact that the number of clusters must be specified in advance.

Finally, it is worth mentioning that, there is no work, reported in the literature (to our knowledge), which proposes and exploits, for several visualizing or browsing applications, an automatic clustering result of segmentation maps estimated from an image database. In the context of browsing large image collections, which is a non trivial task, mainly due to the semantic gap existing between the user subjective notion of similarity and the one according to which a browsing system organizes the images, we can cite several existing techniques. A possible strategy consists of the use of a PCA or MDS mapping-based visualization technique for grouping similar (in terms of color and/or texture-based features) images together on a plane [126], eventually projected onto a sphere [127], on a hierarchical (browsing) tree which can be customized according to user preferences [128], or other graphs (a good survey of existing strategies can be found in [129]). Higher level strategies are proposed for example in [130] where a browsing model integrates high level semantic concepts which allows to help users to narrow a search domain rather than to browse the whole collection [130] or in [131] in which the underlying idea is to mine and interpret the information from the user's interaction in order to understand the user's needs. Based on the Dempster-Shafer theory of evidence and the combination of color and text features, the system's interpretation is used for suggesting new relevant images to the user.

The remainder of this paper is divided into the following sections : First, the proposed distance measure, defined on the space of clusterings, for efficiently comparing two segmentation results, is presented in Section 2.2.1. The averaging procedure of seg-

mentation results used for estimating the cluster centers based on this aforementioned distance is recalled in Section 2.2.2. Section 2.3 shows a variety of applications. Finally Section 2.4 concludes this paper.

2.2 Distance-Based Clustering Components

2.2.1 Used Distance

The variation of information (VoI) metric [132, 133] is a recent information theory based measure for comparing the similarity of two segmentation results (or clusterings). This metric quantifies the information shared between two segmentations by, more precisely, measuring the amount of information that is lost or gained in changing from one segmentation to another. Equivalently (and conceptually), it also represents roughly the amount of randomness in one segmentation which cannot be explained by the other [132]. The VoI is a true metric on the space of clusterings which is non-negative, symmetric and obeys the triangle inequality [133]. This VoI metric is currently exploited as a clustering or segmentation quality metric that measures the agreement of the segmentation result with a given ground truth. To this end, it was recently used in image segmentation [29, 30, 36, 46, 54, 134] as a quantitative and perceptually interesting measure to compare automatic segmentation of an image to a ground truth segmentation (e.g., a manually hand-segmented image given by an expert) and/or to objectively evaluate the efficiency of several unsupervised segmentation methods.

Let $S^{\mathrm{A}} = \{C_1^{\mathrm{A}}, C_2^{\mathrm{A}}, \ldots, C_{R^{\mathrm{A}}}^{\mathrm{A}}\}$ and $S^{\mathrm{B}} = \{C_1^{\mathrm{B}}, C_2^{\mathrm{B}}, \ldots, C_{R^{\mathrm{B}}}^{\mathrm{B}}\}$ be respectively the first and second segmentation (or the segmentation test result to be evaluated and the ground truth segmentation) between which the VoI distance has to be estimated and R^{A} being the number of regions[1] in S^{A} and R^{B} the number of regions in S^{B}. The VoI between S^{A} and S^{B} is defined as :

$$\mathrm{VoI}(S^{\mathrm{A}}, S^{\mathrm{B}}) = H(S^{\mathrm{A}}) + H(S^{\mathrm{B}}) - 2 \cdot I(S^{\mathrm{A}}, S^{\mathrm{B}}) \tag{2.1}$$

[1] A region is a set of 8-connected pixels belonging to the same class and a class, a set of pixels possessing similar textural characteristics.

where $H(S^{A})$ and $H(S^{B})$ denote respectively the classical entropy associated with the segmentation S^{A} and S^{B} and $I(S^{A},S^{B})$ the mutual information between these two segmentations. Let n be the number of pixels within the image, n_i^{A} the number of pixels in the i-th cluster of the segmentation S^{A}, n_j^{B} the number of pixels in the j-th cluster of the segmentation S^{B} and finally n_{ij} the number of pixels which are together in the i-th cluster (or region) of the segmentation S^{A} and in the j-th cluster of the segmentation S^{B}. The entropy is always non-negative (it takes a value of 0 only when there is no uncertainty, namely when there is only one cluster) and is defined as :

$$\begin{aligned} H(S^{A}) &= -\sum_{i=1}^{R^{A}} P(i)\log P(i) = -\sum_{i=1}^{R^{A}} \frac{n_i^{A}}{n}\log\frac{n_i^{A}}{n} \\ H(S^{B}) &= -\sum_{i=1}^{R^{B}} P(j)\log P(j) = -\sum_{i=1}^{R^{B}} \frac{n_j^{B}}{n}\log\frac{n_j^{B}}{n} \end{aligned} \tag{2.2}$$

with $P(i) = n_i^{A}/n$ being the probability that a pixel belongs to cluster S^{A} (respectively $P(j) = n_j/n$ being the probability that a pixel belongs to cluster S^{B}) in the case where i and j represent two discrete random variables taking respectively R^{A} and R^{B} values and uniquely associated to the partition S^{A} and S^{B}. Let now $P(i,j) = n_{ij}/n$ represents the probability that a pixel belongs to C_i^{A} and to C_j^{B}, the mutual information $I(.)$ between the partitions S^{A} and S^{B} is equal to the mutual information between the random variables i and j and is expressed in the following way :

$$I(S^{A},S^{B}) = \sum_{i}^{R^{A}}\sum_{j}^{R^{B}} P(i,j)\log\frac{P(i,j)}{P(i)\,P(j)} \tag{2.3}$$

The VoI is a true metric across the lattice of possible clusterings (taking a value of 0 when two clusterings are identical and positive otherwise) and is bounded by $\log n$. However, if S^{A} and S^{B} have at most $R^{\max}$ clusters (i.e., regions), it is bounded by $2\log R^{\max}$ [133]. Let us also note that if we have several possible ground truth segmentations for a same scene (which could be possibly segmented at different levels of details by different human segmenters), this measure is able to take into account the inherent variability of possible interpretations between each human observer and more precisely the inherent variability

of each possible perceptually consistent interpretation/segmentation of an input image by a simple averaging technique [29]. This variability is also due to the fact that the image segmentation problem is inherently ill-posed (and consequently, this problem has multiple solutions notably for the different possible values of the number of classes not known *a priori*) [5].

2.2.2 Cluster Center Estimation

Let $\{S_k\}_{k\leq L}$ be a finite ensemble of L segmentations $\{S_k\}_{k\leq L} = \{S_1, S_2, \ldots, S_L\}$ existing in a given cluster. The estimation of the center of these L segmentations (also called the cluster prototype or the consensus segmentation) can be efficiently achieved in the VoI distance sense (cf. Sec 2.2.1) by the solution of the following optimization (or so-called *median partition* [79]) problem :

$$\begin{aligned} \hat{S}_{\overline{\mathrm{VoI}}} &= \arg\min_{S\in\mathcal{S}_n} \overline{\mathrm{VoI}}\,(S, \{S_k\}_{k\leq L}) \\ &= \arg\min_{S\in\mathcal{S}_n} \frac{1}{L}\sum_{k=1}^{L} \mathrm{VoI}\,(S, S_k) \end{aligned} \tag{2.4}$$

with $\mathcal{S}_n$ is the set of all possible segmentations using n pixels. Herein, each estimation of the center of a given cluster (i.e., the best compromise segmentation solution resulting in a consensus in terms of contour accuracy or detail level displayed by each segmentations in $\{S_k\}_{k\leq L}$), thus appears as the segmentation solution which minimizes the average pairwise VoI distance between all elements of the cluster. Equivalently, this partition solution can be expressed as the result of a minimization problem on a consensus function (using the $\overline{\mathrm{VoI}}$ distance) which can be solved with a steepest local energy descent procedure [5]. In this iterative minimization procedure, a new label x is assigned to pixel s (initially with label l_s), if this pixel is connected to the x-th region and if the local decrease in the energy function $\overline{\mathrm{VoI}}(.)_{s:l_s\to x}$ is positive with :

$$
\begin{aligned}
\Delta\overline{\mathrm{VoI}}\left(\hat{S}^{[p]}_{\overline{\mathrm{VoI}}},\{S_k\}_{k\leq L}\right)_{s:m\to x} = \\
L\cdot\Bigg\{-\tfrac{n_m}{n}\log\left(\tfrac{n_m}{n}\right)-\tfrac{n_x}{n}\log\left(\tfrac{n_x}{n}\right) \\
+\tfrac{(n_m-1)}{n}\log\left(\tfrac{n_m-1}{n}\right)+\tfrac{(n_x+1)}{n}\log\left(\tfrac{n_x+1}{n}\right)\Bigg\} \\
-2\cdot\textstyle\sum_{l=1}^{L}\Bigg\{\tfrac{n_{m,\mathcal{L}_s^l}}{n}\log\left(\tfrac{n_{m\mathcal{L}_s^l}}{n}\cdot\tfrac{n}{n_m}\cdot\tfrac{n}{n_{\mathcal{L}_s^l}}\right) \\
+\tfrac{n_{x\mathcal{L}_s^l}}{n}\log\left(\tfrac{n_{x\mathcal{L}_s^l}}{n}\cdot\tfrac{n}{n_x}\cdot\tfrac{n}{n_{\mathcal{L}_s^l}}\right) \\
-\tfrac{(n_{m\mathcal{L}_s^l}-1)}{n}\log\left(\tfrac{(n_{m\mathcal{L}_s^l}-1)}{n}\cdot\tfrac{n}{(n_m-1)}\cdot\tfrac{n}{n_{\mathcal{L}_s^l}}\right) \\
-\tfrac{(n_{x\mathcal{L}_s^l}+1)}{n}\log\left(\tfrac{(n_{x\mathcal{L}_s^l}+1)}{n}\cdot\tfrac{n}{(n_x+1)}\cdot\tfrac{n}{n_{\mathcal{L}_s^l}}\right)\Bigg\}
\end{aligned}
\tag{2.5}
$$

where $\mathcal{L}_s^l$ denotes the label at site s of the l-th segmentations ($l \leq L$) of the segmentation ensemble $\{S_k\}_{k\leq L}$ and we recall that $n_{m\mathcal{L}_s^l}$ designates the number of pixels which are together in the m-th cluster (or region) of the segmentation S and in the $\mathcal{L}_s^l$-th cluster of the segmentation $S_l \in \{S_k\}_{k\leq L}$ (see Algorithm : VoI-Based Mean Segmentation Estimation).

As initialization of this steepest gradient descent, we can start from the segmentation result (among the L segmentation results to be averaged), ensuring the minimal consensus energy in the $\overline{\mathrm{VoI}}$ sense [5]. Another strategy consists in initializing the gradient procedure from a synthetic image spatially divided by K horizontal or vertical rectangles with K different labels and to take, at convergence, the segmentation solution ensuring the minimal consensus energy. In order to improve the convergence, we also use a multiresolution approach by considering the optimization problem at a lower resolution level (with the downsampling of $\{S_k\}_{k\leq L}$ by a scale factor c). After convergence of the gradient descent, the result obtained at this lower resolution level is interpolated and then used as initialization for the gradient procedure at the full resolution level. This strategy drastically reduces the complexity and computational effort and provides an accelerated convergence toward an improved estimate (as noticed, for example, in other energy-based. models [134–136]).

Algorithm 1
VoI-Based Mean Segmentation Estimation

$\overline{\text{VoI}}$	Mean VoI (cf. Eq 2.4)
$\{S_k\}_{k\leq L}$	Set of L segmentations to be averaged
$T_{\max}$	Maximal number of iterations

1. Initialization

$$\hat{S}^{[0]}_{\overline{\text{VoI}}} = \arg\min_{S\in\{S_k\}_{k\leq L}} \overline{\text{VoI}}(S, \{S_k\}_{k\leq L})$$

or

$\hat{S}^{[0]}_{\overline{\text{VoI}}} \leftarrow$ image divided into $K(=6) \neq$ rectang. regions

2. Steepest Local Energy Descent

while $p < T_{max}$ **do**

for *each pixel with label* l_s *at site* s **do**

- Let $\mathcal{E}$ the set of labels $\neq l_s$ contained in the local (squared) fixed-size ($N_w = 7$) neighborhood of s
- Draw a new label x according to the uniform distribution in the set $\mathcal{E}$
- if $x = \emptyset$ **then continue ;**
- if pixel s with label x is not 4-connected with the x-th region in $\hat{S}^{[p]}_{\overline{\text{VoI}}}$ **then continue ;**
- Compute $\Delta\overline{\text{VoI}}\left(\hat{S}^{[p]}_{\overline{\text{VoI}}}, \{S_k\}_{k\leq L}\right)_{s:l_s\to x}$ (cf. Eq 2.5)
- **If** $\Delta\overline{\text{VoI}}\left(\hat{S}^{[p]}_{\overline{\text{VoI}}}, \{S_k\}_{k\leq L}\right)_{s:l_s\to x} > 0$ Then replace label l_s by label x at site s

$p \leftarrow p+1$

2.3 Examples of Applications

2.3.1 Visualization of Image Databases

FIGURE 2.1 : A MDS visualization map of the BSDS300 based on the VoI distance between segmentations. Left : view map of the (300) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense.

Due to the huge number of the images in a collection or a database and the limited size of a computer monitor, it may be interesting to find a strategy to provide a quick overview of these images. Generally, it will ensure that these images are displayed as thumbnails and correctly arranged in such a way that the user can quickly and intuitively understand what types of images are contained in the database and their distribution for further analysis. Generally this is typically achieved by a mapping-based technique such as principal component analysis or multi-dimensional scaling (MDS) along with a metric between low-level color or texture features and computed for each pair of images of the database. More precisely, MDS is before above all, a nonlinear dimensionality reduction technique that attempts to find an embedding from the initial feature vectors in the high dimensional space such that distances (or conceptually, the original relation-

ships of these images in term of a given distance) are preserved in a low dimensional space. The foundational ideas behind MDS were first proposed by Young and Householder [137] and then further developed by Torgerson [138] in which the original algorithm (called classic MDS) exploits a spectral method which consists of finding embedding coordinates by computing the top eigenvectors of a *double-centered* transformation of the distance matrix (called a Gramian matrix) sorted by decreasing eigenvalue.

Instead of considering low-level color or texture features to arrange these image thumbnails, as it is commonly used in image database browsing or navigation systems, it may be interesting to arrange them according to their descriptive content extracted by a (region-based) segmentation or more precisely, based on the spatial arrangement of the different objects detected or segmented in the image regardless of their own color or texture. This mapping-based visualization technique, made at a higher level of abstraction, is herein possible since the VoI distance (cf. Sec 2.2.1) is a true metric and also a quantitative and perceptual measure to compare two segmentations, which also inherently takes into account the variability of each possible perceptually consistent interpretation/segmentation of an input image which could be possibly segmented at different detail levels.

The VoI-based true metric allows us to estimate a distance matrix (describing the dissimilarities between each existing pair of segmentations) from which the classic MDS then computes a Gramian matrix having the same properties as the one obtained with a classical Euclidean distance based distance matrix (*i.e.*, positive and semi-definite (PSD)) which then ensures (positive eigenvalues and) a good convergence and accuracy of the MDS algorithm (and more generally of all MDS methods based on eigendecomposition). As explained in [139], the use of a distance which is not a true metric is somewhat equivalent to considering a noise corrupted version of the Gramian or distance matrix and consequently an inaccurate and unreliable (visualization) mapping.

We first present an MDS image overview of the Berkeley Dataset (BSDS300) [140], based on the VoI distance (cf. Fig 2.1). The BSDS300 consists of 300 color images of size 481×321. For each color image of, a set of ground truth segmentations, provided by

human observers (between 4 and 7), is also available. In our application, we can either exploit the result of a segmentation algorithm or use the ground truth segmentations when these one are, of course, available[2].

In our case, the ground truth segmentation map with the median value of the number of regions estimated by the set of human observers (amongst 4 and 7) for each image of the BSDS300 has been here exploited[3]. The images are also squared by stretching them in order to get some invariance in the spatial arrangement of the different parts and object shape segmented in the image. The visualization 2D map of the image thumbnails of the images and their associated ground truth segmentation based on the VoI distance computed for each pair of segmentations is shown in Fig. 2.1. Images (to the left) with a similar layout and arrangement of the different objects detected or segmented in the image regardless of their own color or texture are placed close to each other while images with dissimilarity arrangement are far from each other. In this example, the target output dimension of the MDS-based mapping technique equals to 2 dimensions which is the size of the output mapping. A mapping on a cube would have been also possible with a target output equals to 3 dimensions. It is worth mentioning that we can also efficiently evaluate the reliability of the MDS-based mapping technique (as a function of the target output dimension) by using the correlation-based metric which is simply the correlation of the VoI distance between pairwise segmentations and their corresponding (pairwise) 2D Euclidean distances in the target space [141]. In the absence of perceptual error and thus, with no loss of information in this dimensionality reduction problem, the ideal correlation metric is one. For a 2D MDS visualization map of the BSDS300 based on the VoI distance between segmentations, we obtain a correlation metric $\rho = 0.856$ which means that there are only 14.4 percent of pairs of images whose 2D Euclidean distance between them does not preserve the monotonicity of the initial VoI distance (*i.e.*, or 14.4 loss of information of this 2D MDS mapping according to the correlation metric).

[2] Our approach is tolerant to different types of image degradation (e.g., noise, blur, distortions) insofar as the segmentation method is able to give a segmentation map which is robust enough for these types of degradation.

[3] It is worth mentioning that we could also exploit all the set of ground truth segmentations for each image by computing an average VoI distance computed across all the existing ground truth segmentations.

Fig. 2.2 and 2.3 show respectively some magnified details of Fig. 2.1 and the 6 nearest neighbors of a given image in the database illustrating well the efficiency of our image mapping method based on either the similar layout and spatial arrangement or (to some extend) the similarities in the geometrical shapes of the objects detected and segmented in the scene (and despite the fact that the BSDS300 exhibit a great diversity of images).

Let us note that we can easily quantify the amount of diversity, in term of our main criterion (layout and similar arrangement of the different objects detected or segmented in the image regardless of their own color or texture) by computing the average over all the distances obtained for each segmentation pair :

$$\mathcal{D} = \frac{\sum_{k,l \;\; k \neq l} \mathrm{VoI}(S_k, S_l)}{N(N-1)} \tag{2.6}$$

where N is the number of images or segmentations and $N(N-1)$ the number of pair of segmentations. This diversity measure $\mathcal{D}$ will be even closer to 1 that the diversity will be better in the image database, according to our visualization criterion. For the BSDS300, we obtain a diversity measure metric $\mathcal{D} = 3.05$. In order to increase the diversity in the image database, in our criterion sense, it consists in searching the two most similar segmentations and removing one of them. To attain this goal, a hierarchical clustering (cf. Sec 2.3.4) can be exploited.

Fig. 2.4 shows respectively the two images of the BSDS300 which are respectively the closest and the farthest, in the segmentation-based mean VoI distance sense, from the center of the 300 pictures of the BSDS300 with their associated ground-truth segmentation. It is interesting to note that the estimation of the center of the BSDS300, in term of segmentation map (see Algorithm : VoI-Based Mean Segmentation Estimation), namely ; $\arg\min_{S \in \mathcal{S}_n} \overline{\mathrm{VoI}}\,(S, \{S_k\}_{k \leq L})$), is the blank image (exhibiting one segment/region for the entire image) which is not too far from $\arg\min_{S \in \{S_k\}_{k \leq L}} \overline{\mathrm{VoI}}\,(S, \{S_k\}_{k \leq L})$, the segmentation associated to the center of the BSDS300 (cf. Fig 2.4.(a) at right). This estimation result is also comprehensible, since, in the VoI distance sense, any existing segmentations are also a refinement of this one region segmentation [29].

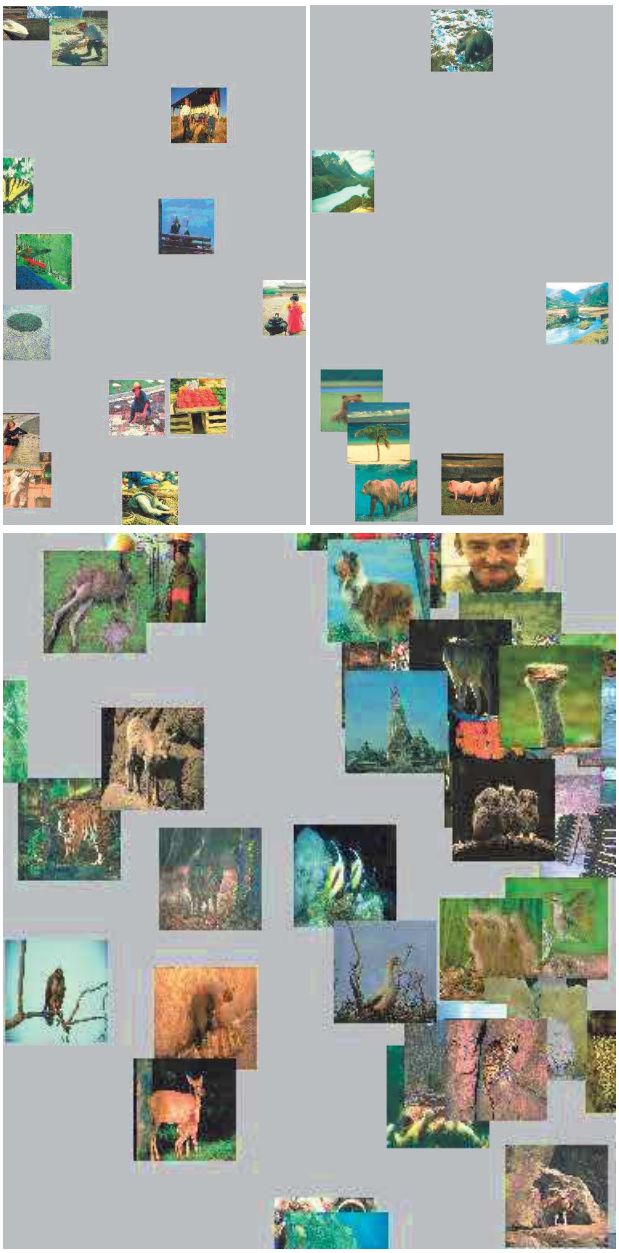


FIGURE 2.2 : Magnified details of Fig. 2.1 showing the grouping of images of the BSDS300 based on the similar layout and arrangement of the objects detected and segmented in the scene. In lexicographic order, a group of pictures mostly showing one or two people at the center of the image (top left) or showing that the three images of bears of the BSDS300 are close (top right) and finally a group of pictures showing that the six birds of the BSDS300, perched on a branch are relatively close or showing, for the bottom or left part of the image, an animal on a mountainside or located in the middle of the image.

FIGURE 2.3 : 6 nearest neighbors of a given (the leftmost) image belonging to the BSDS300. From top to bottom, group of images mainly showing 1-] a small and elongated animal or object on the grass/sand or in the sky/water. 2-] an animal or a group of animals. 3-] pyramidal or (highly) elongated structures (pyramid or mountain). 4-] wild mammals. 5-] group of men. 6-] small pyramidal structures in land/sea-scape.

(a) **(b)**

FIGURE 2.4 : From left to right, the closest and the farthest image, in the mean VoI distance sense, from the center of the BSDS300 with their associated segmentation. The segmentation map which is the exact center of the BSDS300, as estimated by the Algorithm (VoI-Based Mean Segmentation Estimation), is the predictable blank image or one region segmentation (which is not included in the BSDS300 but not too far from the segmentation of the leftmost image).

2.3.2 Segmentation-Based Clustered Visualization

With the VoI distance defined in Section 2.2.1, and the estimation procedure of the center of each cluster or ensemble of segmentations presented in Section 2.2.2, an unsupervised clustering of segmentation results can be efficiently designed for reducing the number of images that are required to be displayed by grouping images with a similar layout and spatial arrangement of the different objects detected or segmented in the image regardless of their own color or texture or for retrieving a specific subset or class of images in terms of (segmentation-based) descriptive content. To this end, several strategies (that we will also evaluate the reliability later) are possible.

The first strategy (Algorithm : VoI-Based MDS+K-means) consists first of using the MDS based dimensionality reduction technique presented in Section 2.3.1 and then exploiting the reduced data of this mapping in a classical K-means algorithm. At the convergence of the K-means algorithm, each new cluster center, in term of segmentation map, is estimated by Algorithm (VoI-Based Mean Segmentation Estimation). The second strategy (Algorithm : VoI-Based K-means) consists in directly using the non-reduced data, *i.e.*, the segmentation maps along with, at each iteration of the K-means algorithm, the estimation of the center (or prototype) of each cluster (or consensus segmentation) until convergence is achieved.

Algorithm 2
VoI-Based MDS+K-means

VoI	VoI distance (cf. Eq 2.1)
$\{S_k\}_{k\leq N}$	Set of N segmentations to be clustered
D_{out}	Output dimension
K	Number of classes
$T_{\max}$	Maximal number of iterations

Initialization

i. Compute the distance matrix M_{ij} describing the dissimilarities between each pair of segmentations with the VoI distance

ii. Based on M_{ij}, use the MDS-based dimensionality reduction to estimate a D_{out}-dim. mapping of $\{S_k\}_{k\leq N}$. Let $\{s_k\}_{k\leq N}$ be this mapping

iii. Choose K initial cluster centers $c_1^{[1]},\ldots,c_K^{[1]}$ among the L D_{out}-dim. vectors $\{s_k\}_{k\leq N}$

1. while $c_i^{[p+1]} \neq c_i^{[p]}\ \forall i$ **or** *iter. number* $< T_{max}$ **do**

1. At the p^{th} step, assign segmentation S_k ($k \leq N$) to cluster i
if : $\| s_k - c_i^{[p]} \| < \| s_k - c_j^{[p]} \| \quad \forall j \neq i$

2. Determine new cluster centers by :

$$c_i^{[p+1]} = \frac{1}{N_i} \sum_{s_m \in Cl_i^{[p]}} s_m$$

where N_i is the number of samples in $Cl_i^{[p]}$

2. Determine each new cluster centers, in term of segmentation map (prototype), by Algorithm (VoI-Based Mean Segmentation Estimation)

Algorithm 3
VoI-Based K-means

VoI	VoI distance (cf. Eq 2.1)
$\{S_k\}_{k\leq N}$	Set of N segmentations to be clustered
K	Number of classes
$T_{\max}$	Maximal number of iterations

Initialization Choose K initial cluster centers $c_1^{[1]},\ldots,c_K^{[1]}$ among the N segmentations

while $c_i^{[p+1]} \neq c_i^{[p]}\ \forall i$ **or** *iteration number* $< T_{max}$ **do**

1. At the p^{th} step, assign S_k ($k \leq N$), to cluster i

$$\text{if}: \quad \text{VoI}\,(S_k, c_i^{[p]}) < \text{VoI}\,(S_k, c_j^{[p]}) \quad \forall j \neq i$$

2. Determine new cluster centers by Algorithm (VoI-Based Mean Segmentation Estimation)

In order to evaluate and compare the reliability of these two above-described clustering strategies, we can estimate the class separability or the Fisher's distance [142] on each clustering result. This distance is simply the within-class inertia divided by the between-class inertia. In our case, this distance is meaningful if this one is computed in the VoI metric sense such as :

$$\mathcal{F} = \frac{\sum_{k=0}^{K}\sum_{S_m\in Cl_k}\text{VoI}(S_m,c_k)}{\sum_{k=0}^{K}\sum_{l=0\ l\neq k}^{K}\text{VoI}(c_k,c_l)} \tag{2.7}$$

where Cl_k is the k-th cluster, $\{S_k\}$ the set of segmentations to be clustered and $\{c_k\}$ the set of prototype centers. This number $\mathcal{D}$ can be also viewed as a *clustering meaningfulness* metric since it clearly measures just that and we should expect that this value be close to zero for good clustering results.

Fig. 2.5 illustrates the Fisher's distance obtained by the two different above-mentioned clustering strategies as a function of K, the number of clusters. We can easily notice that the reliability of the two strategies are comparable, in term of cluster separability. More precisely, slightly better for the VoI-Based K-means (Algorithm 3) for $K > 8$ and slightly

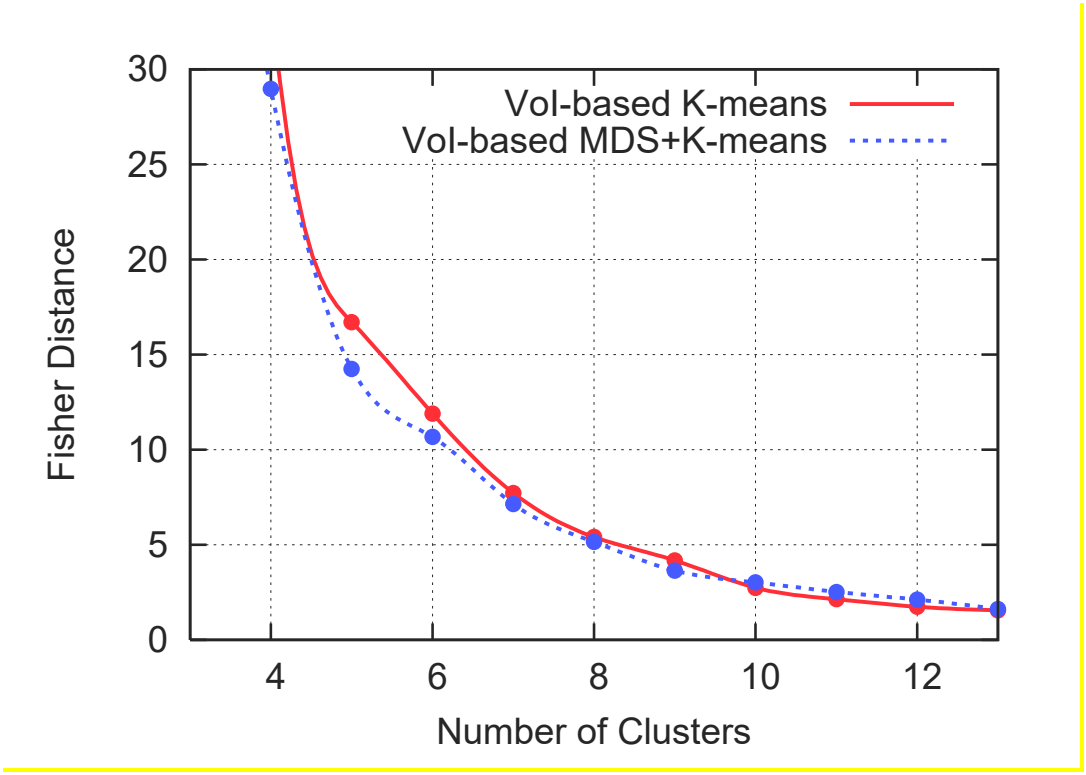


FIGURE 2.5 : Fisher's distance (or cluster separability measure) for the two clustering strategies as a function of the number of clusters.

better for the MDS based mapping (Algorithm 2) for $K < 8$. Computationally speaking, the two algorithms require approximately one minute in order to estimate a partition into a specified number of clusters or classes.

In our case, the K cluster prototypes (centers), (or more simply the K segmentation maps which are the closest of these K prototypes) can be exploited to efficiently summarize the content of the image database and to check, according to this set of cluster prototype segmentations if the database is correctly diversified. Fig. 2.6 shows some cluster prototypes/centers, in term of segmentation maps obtained with the VoI-based K-means for different numbers of clusters. Let us note that the optimal number of clusters can be empirically defined, in our application, by checking if the cluster prototypes are (visually or in the VoI distance sense) pairwise different or as soon as a cluster is just composed by one component only. This condition is fulfilled from 10 classes (for the two different

clustering strategies and for the BSDS300) from 10 classes (and this upper bound can be an interesting cue to quantify the diversity of an image-base). Fig. 2.7 shows us all the selected images from the BSDS300 assigned to the 4th cluster of the VoI-based K-means procedure (cf. Fig 2.6 for $K = 8$) whose cluster prototype, in term of segmentation map, is recalled on the leftmost image. On this segmentation map, we can see several (four or five) elongated structures (with a sort of head above them) which have been automatically clustered and retrieved from the BSDS300. Indeed, among them, we can see images exhibiting between two or four elongated structures such as persons or statues or ears of corn (or reeds). All the other clustering results can be consulted at the web page of the author's website[4].

FIGURE 2.6 : Some prototype (segmentation) centers obtained with the VoI-based K-means for 5, 6, 7 and 8 clusters.

[4] Source code (in C++ language) of our algorithm with the set of clustering and mapping results for each clustering strategy are publicly available at the following http address `http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/#SCVOI`

FIGURE 2.7 : Images from the BSD300 belonging to the cluster corresponding to the cluster prototype, in term of segmentation map, shown on the leftmost image. Images exhibiting elongated structures such as persons or statues or ears of corn (or reeds) have been clustered and retrieved from the BSDS300.

2.3.3 Query-by-drawing search

The proposed framework allows us to easily design and perform a query-by-drawing or query-by-sketch search procedure which would allow a user to formulate a query by simply (and coarsely) drawing a desired configuration or layout of the different geometric shapes of the objects he wants to search and to automatically retrieve in the database. Fig 2.8 shows some examples of a schematic drawing showing one or several geometric shapes and the three nearest neighbor images, in the segmentation-based VoI distance sense, retrieved in the BSDS300. It is interesting to note that the image which is at the center of the BSDS300 (cf. Fig 2.4.(a)) appears in all the test examples. This result is comprehensible, since we recall that, in the VoI distance sense, any existing segmentations are a refinement of this (almost similar) one region segmentation [29]. In addition, this latter segmentation which is very close to the one-region segmentation is also very close of a segmentation map exhibiting one or two simple geometrical segmented shapes as the tested drawing examples. This is also true for the snake and eagle images of the BSDS300.

2.3.4 Hierarchical Clustering

Another grouping strategy consists in using the VoI based distance between two segmentation maps (defined in Section 2.2) along with a Hierarchical Agglomerative Clustering (HAC) based visualization approach. The HAC-based visualization method produces an informative nested hierarchy of similar groups of object or clusters and iteratively builds the hierarchy from the individual elements by progressively merging clusters. It outputs a dendrogram showing all N levels of agglomerations where N is the number of images in the image database, in terms of their region-based descriptive content (and without requiring any parameters such as the number of clusters as the K-means procedure). The first agglomeration corresponds to the most similar pair of images in the database and also define the $(N-1)$ clusters existing in the image-base. The last agglomeration allows us to define the two main clusters existing in the image database. Between the first and last iteration, one can also easily search in the dendrogram a data partitioning or segmentation with a specified number of clusters. Algorithm 4 outlines

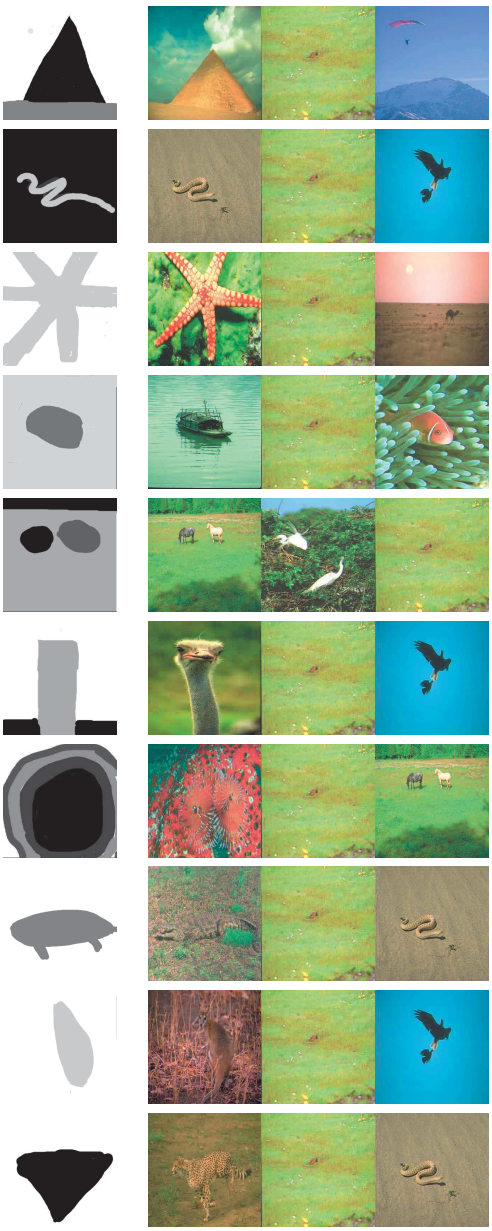


FIGURE 2.8 : Examples of query-by-drawing search (three nearest neighbors) in the BSDS300 in the VoI distance sense.

Algorithm 4
VoI-Based HAC

VoI	VoI distance (cf. Eq 2.1)
$\{S_k\}_{k \leq N}$	Set of N segmentations to be clustered

Initialization. Compute the distance matrix M_{ij} describing the dissimilarities between each pair of segmentations with the VoI distance

while *all images are in one cluster* **do**

1. Find the two most similar (center of) clusters or images (by searching through M_{ij}

2. Join the two clusters or images to produce a new cluster

3. Determine new cluster centers by Algorithm (VoI-Based Mean Segmentation Estimation)

4. Update M_{ij} by computing the VoI distance between the center of this new luster and all other clusters or images

the VoI distance-based HAC algorithm. It is worth mentioning that the HAC algorithm does not make implicit assumptions on cluster shapes, contrary to the K-means based clustering Algorithms (see Algorithm 2 and 3) which *a priori* assume (sometimes wrongly) that the considered clusters are spherical with equal volumes (or the presence of Gaussian distributions with identical covariance matrix) [142]. Computationally speaking, the HAC algorithm requires approximately two hours on the BSDS300 in order to compute the nested hierarchy of clusters, integrating together the different partitioning of the BSDS300 into different numbers of clusters. Fig. 2.9 shows us a dendrogram on the first 50 images of the BSDS300. The full dendrogram estimated for the entire database is available on the author's website. Fig. 2.10 presents us the images related to the first agglomerations, *i.e.*, the set of the most similar images from the BSDS300, in terms of their region-based descriptive content.

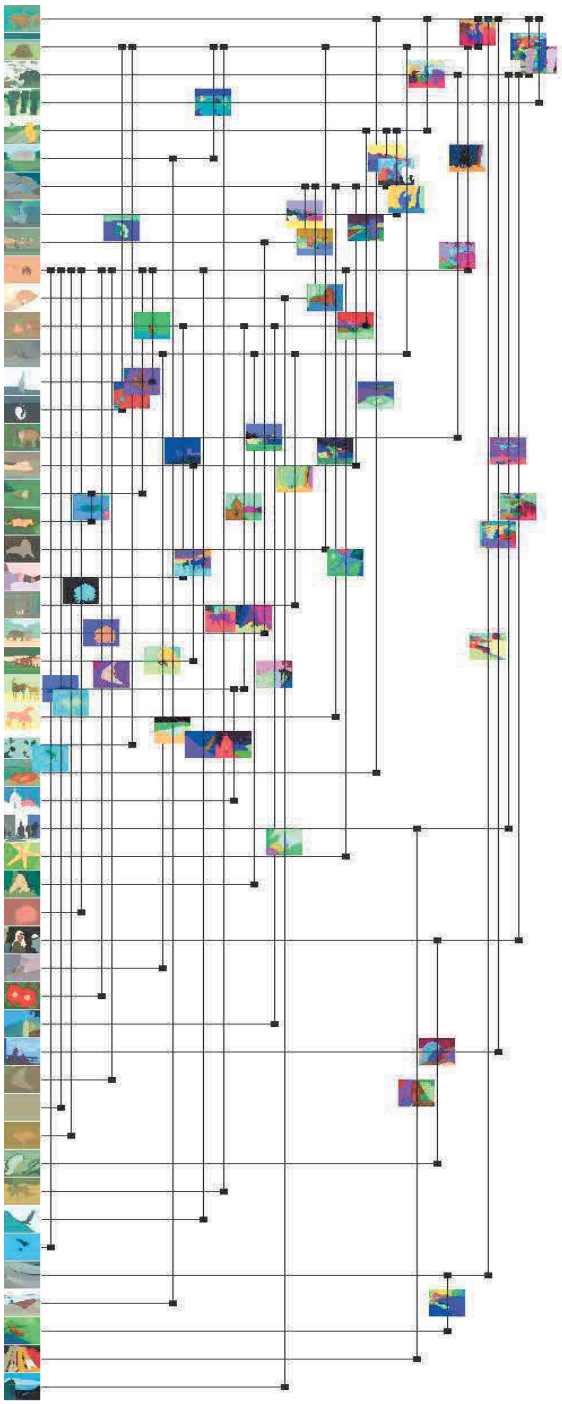


FIGURE 2.9 : A dendrogram or hierarchical agglomerative clustering based on the VoI distance between segmentations on the first 50 images of the BSDS300, showing the 50 agglomerations and the average segmentation between each similar groups of images, in terms of their region-based descriptive content.

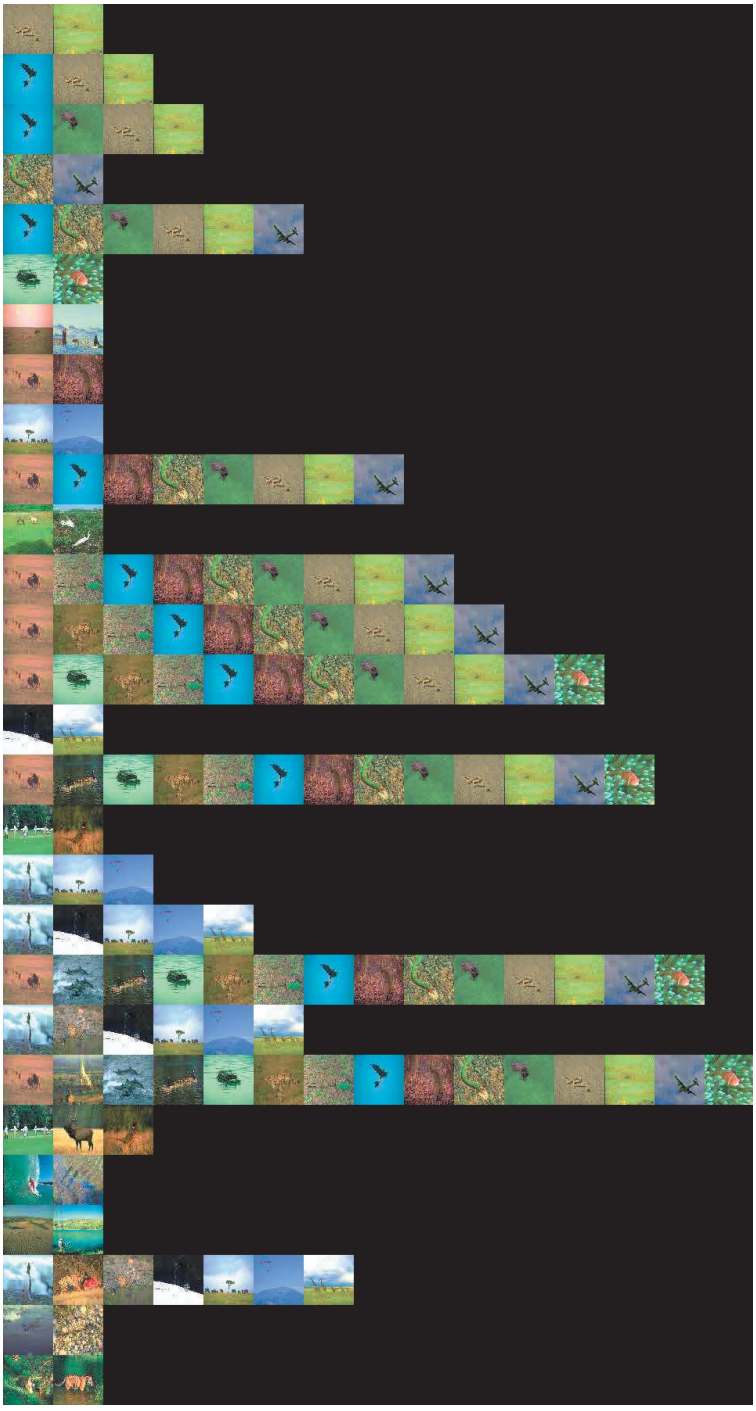


FIGURE 2.10 : Images from the BSDS300 related to the first agglomerations of the HAC, or most similar images, in terms of their region-based descriptive content.

2.3.5 Discussion

In order to better understand the behavior and properties of our search algorithm and the differences of our indexing strategy with previous approaches, we have asked Google image[5] to search similar images to a particular image of the BSDS300 in which we can see three stone totems (*i.e.*, three vertical elongated stone sculptures) with a specific texture and color (*i.e.*, the image number 101085) which was one of the image contained in the cluster shown in Fig. 2.7 (more precisely, this image was assigned to the 4^{th} cluster of our VoI-based K-means procedure, see Section 2.3.2). A Google search online returned several images of statues (cf. Fig 2.11) with the same color and texture (than the query image) but among the top 12 retrieved images, there is no retrieved images with three vertical structures (four of them exhibit two elongated vertical structures) contrary to our approach (cf. Fig 2.7) which have naturally grouped images with a similar layout and spatial arrangement (all the three vertical structures existing in the BSDS300) for the different objects detected in the image regardless of their own color or texture.

2.4 Conclusion

In this paper, we have proposed a new MDS based visualization map[6], only based on the geometrical layout and shapes of the different objects detected and segmented in the scene which provides promising image overviews for large image database. Besides, we have also presented various ways and tools for efficiently clustering or for retrieving a specific subset or class of images in terms of their segmentation-based descriptive content. This descriptive segmentation-based information, provided at a higher level of abstraction, can be a significant and complementary information which can be combined with the commonly used color or texture cues or SIFT features or text information [143, 144] in order to further help the user to browse through large collections in a more

[5] At this stage, it is important to recall that Google search image will not seek in a specific database (with 300 images as the BSDS300), but on the whole web image database and therefore it will have more choice to refine its search process which will be more accurate and efficient.

efficient and intuitive manner.

It is worth recalling that the VoI-based distance and average segmentation estimation process can be exploited by all procedures requiring (iteratively or not) the *mean* observations of a sample of (possibly partitioned) data, such as, the mean shift or (more generally) mode seeking based procedures which does not require prior knowledge of the number of clusters. It is also the case of statistical procedures such as PCA (principal component analysis) which could be also generalized in order to study the variability existing in a medical image segmentation database of specific segmented anatomical structures.

In addition, a similar approach could be applied to image data containing a temporal dimension such as video image sequence and the 3D-generalized proposed method could be useful and exploited, in the same way, for video structure analysis (in terms of their segmentation-based descriptive content) or for video indexing problems and retrieval. It could also include query interfaces and video clustering with, for this 3D temporally coherent data, the estimation of a visual (spatio-temporal) prototype (center) model for each cluster which could be subsequently exploited for video classification.

[6] This map has also been estimated on 3 other image databases, namely ;
1) the Weizmann database (1 & 2 objects) (200 images),
2) the Microsoft Research Cambridge Object Recognition Image Database (MSRC) (591 images),
3) The Stanford Background Dataset (DAGS) (715 images).
The references of these image databases and the obtained visualization maps are publicly available (with the source code of our algorithm) at the following http address :
`http://www.iro.umontreal.ca/~mignotte/ResearchMaterial/#SCVOI`

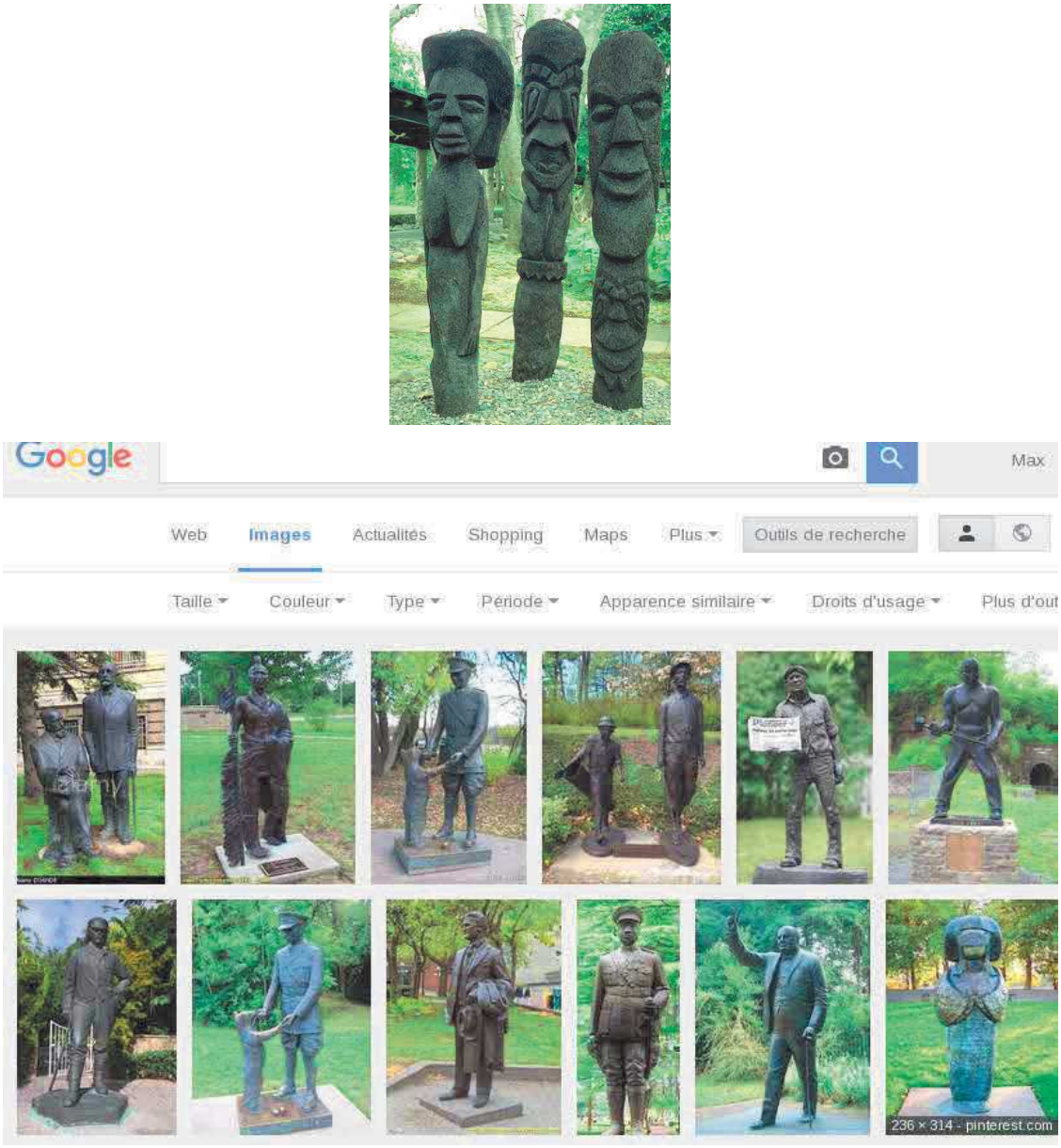


FIGURE 2.11 : Top : Image 101085 from the BSDS300. Bottom : Top 12 most similar images returned by Google image search. A Google search online returned several images of statues with the same color and texture between them (and enough similar to the query image) but none of them show three elongated vertical structures with possibly a different color or texture as our method can give (cf. Fig 2.7).

Deuxième partie

Consensus de segmentation pour améliorer la précision des algorithmes de détection de mouvement

CHAPITRE 3

CHANGE DETECTION ALGORITHM IMPROVEMENT BY ADAPTIVE FUSION OF RESULTS OBTAINED FROM DIFFERENT COLOR SPACES

Cet article a été accepté avec modifications mineures au *International Journal of Image and Data Fusion* comme l'indique la référence bibliographique.

A. Khlif, M. Mignotte. Change Detection Algorithm Improvement By Adaptive Fusion of Results Obtained From Different Color Spaces. *International Journal of Image and Data Fusion*, August 2017.

Cet article est presenté içi dans sa version originale.

Abstract

In this paper, we study the idea of combining multiple change detection results in a video sequence obtained by the same background subtraction model (BGS), operating in different and complementary space colors. We show that this strategy is efficient if this one is associated, as proposed in this work, with an adaptive fusion procedure, able to first estimate (for each considered BGS algorithm) the specific color space models to be combined and afterwards that continuously takes into account, at best and online, the complementarity of the different motion detection results in the fusion process. Experiments performed on the ChangeDetection.net (CDnet 2014) dataset, have shown that the proposed fusion strategy can easily and significantly improve, in terms of overall F-Measure, the efficiency of a number of BGS models especially pixel-based BGS algorithms using a single global threshold.

Keywords : CDnet 2014, Background subtraction model, Change detection, Decision fusion or combination, Foreground background binary segmentation, Hill climbing method, Majority-vote rule, Pearson correlation, Space colors, HSV, LAB, RGB

3.1 Introduction

Motion detection is a fundamental pre-processing step in image processing whose goal is to identify coherent motion regions of interest in video sequences [145]. This low-level vision task which consists, in fact, in automatically segmenting each frame of the video sequence into two significant and informative regions (namely the foreground and background) is often the preliminary and also crucial step in the development of many video understanding algorithms or multimedia and video processing systems [146, 147] such as human computer interaction [148], object tracking [149, 150], action recognition [151], people counting, traffic video analysis, video compressing and annotation to name a few [152]. In general, change detection technique can be divided into two main categories : unsupervised and supervised change detection methods [153].

Recently, Wang *et al.* have introduced the `www.ChangeDetection.net` (CDnet 2014) dataset [154] (as the latest and expanded release of CDnet 2012), in which a wide variety of realistic (indoor and outdoor) videos[1] have been captured (under different scenarios and cameras or camera settings) and manually segmented (at the pixel level) by a human expert. This important and valuable initiative has allowed the image processing community, first, to objectively rank the methods that are most suitable for a specific problem as well as to identify the change detection (CD) algorithms that are reliable overall, with a relatively good confidence level. The simplest strategy to segment the foreground regions in a video sequence consists in subtracting the pixels in the current frame from those in a previous or reference one playing the role of a background model (BM). From this basic background modeling, prone to many errors (mainly due to the inherent dynamic nature of real-world scenes), several improvements have been proposed, validated and then quantified, in terms of "classification gain", on CDnet. Among these, we can mention the use of a simple adaptive filter to construct a reliable background model (*e.g.*, an average [155], median [1], Kalman [156] or Wiener filter [157]) or the

[1] More precisely, CDnet 2014 includes 53 video sequences divided in 11 different categories (associated with a specific change detection problem such as dynamic background, shadows, night time, camera jitters, intermittent object motion, challenging weather, low frame-rate, air turbulence, etc.) for a total of more than 160000 video frames carefully manually segmented.

histogram analysis over time [158] or an eigen-decomposition [3] or, more formally, also the use of a statistical framework to construct a parametric model for the color distribution of each pixel based on the past observations (such as a single Gaussian [159], a mixture of Gaussians (GM) [160]) or *via* a non parametric modeling strategy [15] or by the generalization of these algorithms and the taking into account of the spatial relationship between pixels by modeling regions as mixture distributions rather than individual pixels [16].

Up to now, the CDnet initiative has also allowed us to quantify the relative importance of an appropriate distance metric between the Background model and the pixel to be classified [161] (*e.g.*, Euclidean versus Mahalanobis), or to estimate the classification gain [2, 22] obtained by automatically estimating the optimal number of Gaussians per pixel in the GM-BGS aforementioned model proposed by [160], or by online estimating the GM model with Expectation Maximization algorithm [162] or to better understand the relative interest of a background modeling approach based on a neural network [11, 14, 163], a self organized map (SOM) [164], a physically based [12], a graph-cut energy-based BGS model [165] or an edge-based foreground background estimation (and interior classification) approach [166, 167].

In addition, the CDnet benchmark dataset brought to light a number of interesting findings such as the potential to exploit the spatial relationship or spatial-temporal correlation across different or neighbouring pixels [16, 163, 168, 169] (*via* possibly local binary similarity [170] or local context descriptors [171]) and allowing to detect more easily camouflaged foreground objects while the effects of illumination variation are rightly ignored. It also allows to highlight the benefits of using both a background, but also a foreground model [154, 165, 169] and the idea of consensus approach as classification rule [9, 170] and also the appealing concept of sharable models [169], conceptually establishing many-to-one relationships between pixels and models and allowing, in fact, each pixel to dynamically search the best matched model in its current spatial neighbourhood (which turned out to be robust to camera jitters, dynamic back- ground, etc.) and the appropriateness of a a pixel-level loop (or feedback mechanism) [9, 170] to dy-

namically adjust the internal parameters of the model and to dynamically control the algorithm sensitivity and identify complex motion patterns.

Finally, the CDnet's benchmark has also confirmed that the concept of information fusion (possibly of different models or combination of decisions), *i.e.*, the so-called mixture of experts or committee machines in machine learning, was particularly interesting and well-suited for this binary classification problem. Indeed, the most efficient change detection algorithms, validated up to now on CDnet, also exploit (among other things) this approach. It is, for example the case of the FTSG model [10] which is a multi-level system that efficiently combines information from a motion computation method based on spatio-temporal tensor formulation, and the results from a split Gaussian mixture model. It is also the case of the MBS (Multimode Background Subtraction) model [172] which considers a bank (or multiple) pixel-wise single Gaussian BG models (unlike conventional CD methods which have a single BG model per pixel) of the scene for each cluster obtained by a K-means algorithm (with a correlation distance) on the set of training images. In the change detection context, let us stress out that the fusion of a spatial segmentation map and a temporal segmentation has also been studied [173, 174] (as in [175], in order to improve the updating efficiency of the BM). The hierarchical approach, such as the one combining a block-based and a pixel-based refinement model [176, 177] can also be viewed, in this context, as a particular *model fusion* technique.

Finally, in addition to these above-mentioned model fusion approaches, a decision fusion strategy can be adopted, either by exploiting an adaptive classifier cascade for labelling each pixel [177] or with an ensemble of BGS classifiers [178] such as the GM-BGS jointly considering 13 mixtures of Gaussian (using intensity value in RGB color channel and texture information derived from the gradient and Haar features) whose binary decision is then combined with the average fusion rule. In the same spirit, [179] combines two BGS classifiers, one classifier based on the depth feature and the other one based on the color features. Finally, let us also mention [180], with the IUTIS algorithm, which tries to combine or fuse multiple efficient change detection algorithms (and more precisely their outputs) with a genetic algorithm. In their case, this evolutionary algo-

rithm allows to automatically and adaptively select the best subset of inputs algorithms to be combined along with the best way to combine their detection solutions to get the best (trade-off) motion detection result. This fusion of decision model[2] turned out to be well suited for the motion detection problem and more efficient than a simple majority vote scheme [154] and a probabilistic rand index-based fusion algorithm [19].

In this work, we propose an original unsupervised fusion strategy, both simple and easy to implement which can significantly improve (in terms of classification error), the efficiency of a number of background subtraction (BGS) algorithms, especially those based (or mainly based) on a single (main) threshold. More precisely, this fusion procedure exploits the idea of combining, with a simple majority vote scheme, two (or several) motion detection results, obtained by the same change detection algorithm on a video sequence expressed in two (or several) different (complementary) color spaces. In the proposed procedure, the optimal adaptive threshold, associated to each used color space, is automatically adjusted in order to take into account, at best, the complementarity of the two motion detection results in the fusion process.

The remainder of this paper is organized as follows : Section 3.2 describes the related research work which has been conducted in the field of fusion, combination or cooperation of multiple color spaces, in order to achieve a better result. Section 3.3 provides the rational and motivations behind our decision fusion approach based on different and complementary color spaces and discusses the behavior and performance of well-known pixel-level BGS models when the video sequence is coded in various color spaces. Section 3.4 presents the proposed fusion model and the related adaptive estimation procedure. Finally, Section 3.5 quantifies the gain, in terms of overall F-Measure, obtained by our fusion strategy on CDnet 2014 and presents a set of comparisons with existing change detection methods. Finally Section 3.6 concludes the paper.

[2] Regarding this fact, it is quite normal since the fitness function of each candidate solution of this genetic algorithm-based decision fusion method exploits the average rank (across video sequences) and performance measures obtained (on the CDnet dataset) by each motion detection method (to be fused). *i.e.*, some information which itself exploits the ground truths and thus strongly over-fits and supervises the IUTIS fusion procedure.

3.2 Related Work in Combination of Color Spaces

Up to now, relatively few research works have been proposed in the literature on the idea (and the interest) of combining or fusing processing results obtained (possibly by the same algorithm), from input images expressed in different color spaces, in the goal to achieve a better result.

We can, however, mention the problem of natural image segmentation for which a recent alternative to the existing computationally costly and complex segmentation models consists in simply combining (or fusing) multiple low-cost and weak (*i.e.*, rough) image segmentation results of the same scene, associated with a simpler segmentation model, to provide a final improved segmentation (also called a consensus solution given by a mixture of experts or a classifier ensemble). In this strategy, the difficulty lies in generating an appropriate segmentation ensemble which should include, a lot of complementary and supplementary information. To this end, the authors in [5,6,19,181] propose to fuse, according a (respectively) least square, weighted least square, rand index or entropy criterion, a set of segmentations obtained, by a simple K-means algorithm, from the same image expressed in different (between 6 and 10 in these applications) color spaces. According to the author in [6, 19], this process is somewhat equivalent to observing the image scene with several sensors or cameras with different characteristics but it is also and above all, a necessary condition to ensure the diversity required to build a reliable consensus segmentation solution, outperforming all the individual ensemble members, (*i.e.*, the input segmentation to be fused). A similar strategy is used in [182] with a set of segmentations (to be fused) obtained by a fuzzy K-means algorithm on several pre-selected color channels (components) belonging to several different color spaces or in [183], where the input segmentations are achieved by a Support Vector Machine (SVM) classifier, using the image expressed in 5 different color spaces, and are combined according to the Dempster-Shafer (DS) theory or in [84] in which multi-level thresholding results, applied on an input image expressed by different color spaces, are fused together by the DS evidence theory using as input features, the mass functions (*i.e.*, a membership degree) associated with each pixel label of each segmentation to be

fused. In the same vein, the authors in [184] have proposed a method of color image segmentation based on fuzzy homogeneity and data fusion techniques. In the first step, an initial segmentation is obtained by using the histogram thresholding. In the second step, the evidence theory is employed to merge the information coming from the three images (R, G, and B).

In addition to the segmentation problem, the benefit of combining processing or classification results obtained in different color spaces was also exploited for different detection and classification problems and more precisely for the concealed weapon detection [185]. In this work, an image and its infrared version are fused (with a wavelet-based technique) to provide a fused image showing a detailed description of the people in the scene and any hidden weapons. Also, it can be used (with a multi-scale decomposition of the image) to enhance the background scenery and details from the visible image [186] or for person re-identification in [187] where the combination of features expressed in (6) different color spaces have been used to build a random forest to learn the similarity function of pairs of person images. This approach is claimed to improve the final classification result but especially to alleviate the over-fitting problem when there are limited training data. Chinadro [188] proposed a texture classification system based on a set of independent classifiers each assigned to a different color space. In order to classify an input textured image, the author proposes to fuse the classification decision of each classifier using the product rule and this strategy significantly improves the classification rate ($+20\%$) comparatively to an approach based on a single color space. The same strategy is used in [189] for the ear classification with the sum rule of several 1-nearest neighbor classifiers trained by Gabor-based features extracted from different color spaces. To further improve the face detection and localization, the author in [190] combines, by a defuzzification-based fusion scheme, several detection results obtained on different binary segmentations of an input image formed by an individual color component of different pre-selected color spaces. Peer *et al.* [191] also improve the face detection by first detecting the skin by a set of heuristic and geometric constraints in different color spaces as [192] (or [193] with only two color spaces), and this strategy turns out to be

robust especially to illumination conditions and changes in face expressions. Let us note that skin disease detection is also a domain for which the strategy based on the combination of color spaces are well suited. Indeed, a fully automated computer aided diagnosis system of melanoma skin cancer detection was recently proposed in [194]. This system comprises a feature extraction (of the gray level co-occurrence matrix features for the image converted in different color spaces) followed by a selection step to keep the most prominent features describing melanoma, which are finally used to feed a set of classifiers.

Another application for which the combination of color spaces is interesting is the Color Difference Processing (CDP) which may be useful in printing, textile, paint, ink, and some other industries. For example, it is difficult but often needed to determine the subtle color change of the dye. To this end, Zhou *et al.*, in [195], propose a multi-color space feature fusion/selection model in which the best color features having the greatest color difference are first elected and then fused by using a wavelet-based combination scheme. Another interest is the detection of low-contrast targets such as the detection of foreign fiber in the wool, white hair, or plastic in cotton. With a somewhat similar strategy, a technique allowing the automatic detection of the light-colored cotton impurities was proposed in [196] which analyzes the different color features of the targets in different color spaces to select several optimal features and to fuse them using also a wavelet transform in order to finally obtain a single image with clearly defined features.

Another application for which the combination of color spaces is interesting is the fire detection [197–199]. For example, a fire detection method was proposed in [197], in which they define a set of three rules using a combination of the RGB and the HSI color spaces.

Finally, the strategy of taking into account multiple complementary color spaces have been recently exploited to strengthen the invariance properties and the distinctiveness of color descriptors. This invariance can be obtained in [200, 201] by simple concatenation of color-based descriptors computed independently for several color spaces. As a result and thanks to this color invariance property, the obtained large SIFT or SURF descriptors

produce very good matching quality and are among the methods with the best overall performance in the evaluation [200, 201].

Each research work we have mentioned above, relies on a specific application-related fusion technique which is different than the one proposed here and, most importantly, combine processing results obtained from images coded in different and not fully mathematically justified color spaces. In addition to our original fusion procedure, related to the motion detection task, we also propose, as the second novelty of this work, a reliable statistical estimation of the specific color space models to be combined for each considered BGS algorithm.

3.3 Single Threshold BGS Methods in Different Color Spaces

3.3.1 Introduction

As already said in Section introduction, CDnet has allowed us to evaluate many different BGS methods ; from the traditional, easy to implement, real-time, single-threshold pixel-based BGS model, detecting the moving objects from the simple thresholded difference between the current frame and, either a reference (or background) image [1, 155, 158] or a statistical [15, 159, 160], cluster or region [16], neural [11, 14, 163, 164], spectral background model [157], (which is generally continuously updated at each video frame), to name a few, and the recent and more sophisticated BGS models, often built upon these classical algorithms, exploiting several tricks such as prior knowledge about the problem (*via* distribution, constraints, algorithmic conditions, feedback structure, (fine-tuned) hidden or internal parameters to further improve the classification accuracy obtained on the CDnet benchmark dataset. Nevertheless, let us note that with a traditional, simple, single threshold based BGS model, there is less chance that the model was fitted accidentally by errors in data which is not always the case of more complex models. In any case, one should include model complexity when evaluating a model and more precisely, the obtained gain in terms of classification error should be also objectively evaluated with respect to the increase of complexity of the resulting proposed model.

In this work, we propose an original and low complexity technique to significantly and easily improve any single-threshold (either pixel based or region based) BGS model. This improvement is not based on additional constraints, prior knowledge or informative high level cues about the video data structure but simply based on the complementarity existing between the video data information represented by two different (nonlinearly related) color spaces.

3.3.2 Color Models

In fact, let us stress out that a color model is just an abstract mathematical model describing the way colors can be represented as tuples of numbers (*e.g.*, triples in RGB or quadruples in CMYK color space) [202] and the choice of the Red Green and Blue primary colors, in image processing or vision application, is mainly related to the human perception of colors[3] and more precisely to the physiology of the human eye for which these three primary colors are also stimuli that maximize the difference between the responses of the cone cells of the human retina to light of different wavelengths. In fact, a given image, expressed in different color spaces, is somewhat equivalent to observing the image scene with several sensors or cameras or different experts with different characteristics [6, 19] (in this sense, our fusion procedure can be viewed as a special type of multi-classifier systems (MCS) [203] or multi-sensor data fusion model [204]). More formally or mathematically, the non-linear conversion formulas between color spaces significantly modify the shape of the color distributions [205], which are on-line estimated for BGS with parametric or non parametric statistical background models or indirectly estimated for other BGS models. Similarly, in a BGS model, a (non-linear) conversion into another color space can be also viewed as a non linearly transformed distance metric (which may not be mathematically expressible) between the current frame and the background model expressed in another space or as different (nonlinearly related) feature cues computed for the same color space.

[3] In fact, before the electronic age, the RGB color model already had a solid theory behind it [202].

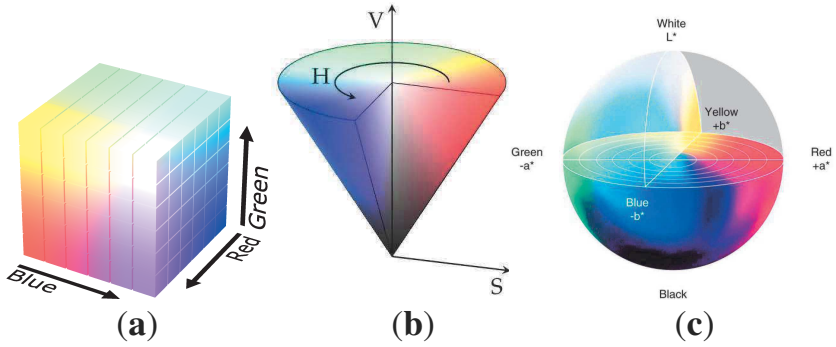


FIGURE 3.1 : RGB, HSV and CIE-LAB color spaces. From left to right, (**a**) the RGB color space, (**b**) the HSV (hue, saturation or chrome, value) color space and (**c**) the CIE-LAB color space in which, L= 0 yields black and L= 100 indicates diffuse white, A indicates its position between red/magenta and green (negative values for A indicate green while positive values indicate magenta) and B indicates its position between yellow and blue (negative values for B indicate blue and positive values indicate yellow).

Amongst the different existing color spaces, we have experimentally found (see the two following Subsections) that (more particularly) three of them have the appealing property to bring complementary information (about the scene) which can be advantageously exploited by our fusion-based change detection procedure. In addition to the RGB space, it is the HSV and CIE-LAB color spaces (cf. Fig 3.1). The HSV (hue, saturation or chrome, value) is a color space (CS) which may be depicted as a cone and which has the property to separate the luminance or the image intensity, from the color information (chromacity). It describes colors (hue or tint) in terms of their shade (saturation or amount of gray) and their brightness (value or luminance). As a method of describing colors, HSV offers a more intuitive representation of the relationship between colors and color progressions (especially for graphic artists and designers) and approximate the way humans describe and interpret color. Contrary to the RGB CS, which approximates the way human vision encodes images (by using three additive primary color channels ; Red, Green, and Blue) and which is plainly based on the acquisition capabilities of cone cells in retina (which are able to react to different wavelengths), and contrary to the HSV CS, which approximates the way humans describe and interpret color, the CIE-LAB color space has been constructed to be perceptual uniform. Its main property consists in pre-

senting color data equidistantly on the color space as perceived in reality (*i.e.*, uniform changes of components in the this color space aim to correspond to uniform changes in perceived color).

3.3.3 BGS Algorithm Performance in Different Color Spaces

TABLE 3.1 : F-measure result, on CDnet 2014 dataset, for four pixel-level single-threshold BGS methods ; namely the BGS based on a simple difference, median filter [1], a mixture of Gaussians [2] or an eigen-decomposition [3], operating in different color spaces (the optimal threshold is indicated in parenthesis next to the F-measure value).

BGS Algorithms	RGB	HSV	LAB	YCbCr	XYZ	H123	I123	TSL	YIQ
Simple difference	$0.27_{(16)}$	$0.20_{(20)}$	$0.27_{(8)}$	$0.26_{(16)}$	$0.27_{(20)}$	$0.27_{(12)}$	$0.26_{(16)}$	$0.26_{(20)}$	$0.27_{(12)}$
Median	$0.51_{(61)}$	$0.42_{(53)}$	$0.52_{(24)}$	$0.52_{(37)}$	$0.50_{(69)}$	$0.52_{(37)}$	$0.50_{(45)}$	$0.45_{(69)}$	$0.51_{(45)}$
GMM	$0.55_{(25)}$	$0.50_{(25)}$	$0.56_{(16)}$	$0.53_{(16)}$	$0.54_{(22)}$	$0.55_{(22)}$	$0.56_{(22)}$	$0.51_{(22)}$	$0.55_{(22)}$
Eigen	$0.29_{(15)}$	$0.19_{(24)}$	$0.30_{(14)}$	$0.29_{(14)}$	$0.26_{(14)}$	$0.29_{(14)}$	$0.29_{(14)}$	$0.18_{(14)}$	$0.26_{(14)}$

Table 3.1 summarizes the achieved performance result (in terms of F-measure), on CDnet 2014 dataset, for a video sequence coded in different color spaces, obtained by the four basic and most popular pixel-level single-threshold BGS methods ; namely the BGS algorithm using as reference image, the previous one, or a background model, respectively obtained by a median filter [1], a mixture of Gaussians [2, 22] or an eigen-decomposition [3]. In these four methods, the main internal parameter corresponds to the distance threshold between the tested pixel (in the current frame) and the background model. This threshold value is the supervised optimal and fixed value determined by trial and error, for each considered color space. We can notice that this optimal threshold is obviously different for each BGS method but also can be different for each color space and can nearly vary from simple to double from one color space to another. We can also note that some of color spaces are not overall well suited for the BGS method used in this test, such as the HSL or the HSV color space. Finally we can also notice that the RGB and LAB color space are the two color models, which give, in average across different BGS model, the best results. Let us also note that the average F-score-based classification rate obtained by each individual color space, give us no useful information about the potential overall complementarity and combination of some of them. This potential of fusion between several color spaces, and more precisely between the motion

detection results obtained in several color spaces will be empirically investigated in the next Section.

3.4 Methods

3.4.1 Proposed Model

The fusion procedure, proposed in this work, is simply based on the idea of combining, with a simple majority vote scheme, two (or several) binary (foreground-background) segmentation results, obtained by the same change detection algorithm on an image sequence expressed in two (or several) different (complementary) color spaces. In our application, the majority vote is achieved with a spatial window of size $N_w \times N_w$ pixels, centered on the pixel s to be classified, and collecting the binary class labels of the N_c (two or several) segmentation results obtained by each color space conversion, and which finally assigns to that central pixel s, the class label that has the majority vote (in this set of $N_w \times N_w \times N_c$ labels). This strategy ensures both the spatial regularization of the final fused (detection) map result and also a reliable decision fusion between results obtained in different color spaces.

In fact, if we consider a fusion scheme between two color spaces and that the main (single) threshold of the first BGS algorithm using the first color space is either kept constant and fixed once-for-all, for all videos of CDnet or adaptively estimated for each frame, the only difficulty of our approach lies in the unsupervised and adaptative estimation of the main (single threshold) of the second BGS algorithm, operating in the second color space domain, in order to ensure, at best, the complementarity of the two motion detection results during the majority voting-based fusion process. In our strategy, this second parameter is simply dynamically adjusted in order to keep the same number of pixels classified as foreground (or *mobile* class) between the first and second algorithm operating in the two color space domains. To this end, we start (for the first frame) by setting the second threshold τ_2 to the same value as the first threshold τ_1 and, for each frame, the second threshold is dynamically adjusted with a hill climbing method with fixed step size using the following procedure ; if the number of pixels, classified as

background, in the detection map of the first algorithm, is greater than for the second algorithm; then $\tau_2 = \tau_2 - \delta$ else $\tau_2 = \tau_2 + \delta$ with δ, a fixed step size. In our application, we take δ, a small value (*e.g.*, $0.1 < \delta < 0.01$) in order to get a smooth, robust to noise, prediction of τ_2. This prediction model rightly assumes that this parameter may only slightly changed in consecutive frames and allows us to also get an estimation which is robust against the effect of a possible poor detection map given either by the first or second algorithm at a given time. In the case of a fusion strategy between three color spaces, the third threshold τ_3 has to be adjusted comparatively to the first algorithm, in the same manner.

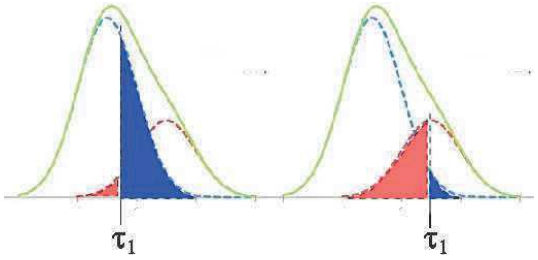


FIGURE 3.2 : The distribution of the difference (green curve) between the current frame and the background model is composed of two distributions; the distribution corresponding to the *background* class (blue) and the one corresponding to the *foreground* label (red). For a given threshold τ_1, the number of false positives and false negatives are proportional respectively to the blue and read area. By increasing τ_1, we decrease the number of false positives, but increase the number of false negatives and conversely.

The above-mentioned procedure leads to a *semi-supervised* algorithm since we consider that the threshold of the first BGS algorithm is fixed, once for all to the optimal value giving the best classification score on the CDnet benchmark, as it is generally the case for simple pixel-based BGS methods such as those given in Subsection 3.3.3 and currently referenced and benchmarked on the website hosting of CDnet. Nevertheless, in order to obtain a fully *unsupervised* algorithm, we can propose and evaluate the following simple estimation procedure for the first parameter. This procedure relies on the fact that we concretely search to estimate a threshold which ensures the best possible denoised detection result, at each frame of the video sequence. Nevertheless, a too low

threshold τ_1 gives raise to false positives while a too high threshold may result in a lot of false-negatives in the final binary detection map. Knowing that, in our case, false negative are prone to be less numerous than false positives (since they are usually within the mobile regions, and also more difficult to estimate) and knowing that false positives are often isolated (or quasi-)isolated pixels; we can estimate their proportion by counting the number of small regions classified as foreground or by counting the number of pixels classified foreground which are not surrounded in a squared window by more than 30 percent of pixels also classified as foreground. If this proportion, within the whole image, is below a threshold ρ, then $\tau_1 = \tau_1 - \delta$ else $\tau_1 = \tau_1 + \delta$.
This procedure is somewhat equivalent to an unsupervised thresholding algorithm which will search to adaptively estimate, over the video frames, the optimal threshold of a bi-modal histogram (composed of a mixture of two different uni-modal distributions) ensuring not too many false positives (cf. Fig 3.2). In addition τ_1 is constrained to stay within the range $[20 \pm 15]$ in order to avoid some trivial cases (background or foreground for the whole image) and consequently a bad estimation for τ_1.

Algorithm 1 summarizes in pseudo-code the proposed *semi-supervised* or *unsupervised* fusion procedure (called CSF for Color Space Fusion) and more precisely the updating strategy of τ_1 and τ_2 for each frame of CDnet. We recall that, in this algorithm, if τ_1 is fixed as a constant optimal value, this leads to the *semi-supervised* fusion procedure. If τ_1 and τ_2 are constant and fixed to their optimal value, this leads to the *supervised* fusion procedure.

Since the majority vote-based fusion procedure is low time consuming, it is worth mentioning that our fusion procedure can be done with almost no CPU time overload if we use a multi-core processor and if each algorithm (operating in each color space) is run in parallel[4] simultaneously on different cores.

[4] To this end, the application programming interface OPENMP (Open Multi-Processing) provides a simple and efficient means of doing so.

Algorithm 1
Unsupervised Fusion Procedure

I	Frame to be binary segmented (**Input**)
S	Fused segmentation result (**Output**)

N_w	Majority vote window size
ρ	Maximal proportion of isolated false positives (for the adaptive update of τ_1)

S_{c1}, S_{c2}	Segmentation given by the algorithm operating in the $1st_{(c1)}$ or $2nd_{(c2)}$ color space
τ_1, τ_2	Main threshold of the algorithm operating in the 1st or 2nd color space
δ	Step size of the gradient (*e.g.*, $\delta = 0.1$)

1. Initialization :

$\triangleright$ Convert I into the 1st and 2nd color space

$S_{c1} \leftarrow$ SEGMENTATION ALG. $\left(I_{c1}\right)$ WITH THRESHOLD τ_1

$S_{c2} \leftarrow$ SEGMENTATION ALG. $\left(I_{c2}\right)$ WITH THRESHOLD τ_2

2. Threshold Updating Scheme for τ_1

$N_p \leftarrow$ Proportion of pixels, not surrounded in a second-order neighbourhood by more than 2 pixels also classified as foreground.

For τ_1 (algorithmically) constrained to be within the interval $[20 \pm 15]$:

if $(N_p \leq \rho)$ **then** $\tau_1 -= \delta$ **else** $\tau_1 += \delta$

3. Threshold Updating Scheme for τ_2

$N_{S_{c1}} \leftarrow$ Number of pixels classified *foreground* in S_{c1}

$N_{S_{c2}} \leftarrow$ Number of pixels classified *foreground* in S_{c2}

if $\left(N_{S_{c1}} \geq N_{S_{c2}}\right)$ **then** $\tau_2 -= \delta$ **else** $\tau_2 += \delta$

4. Majority Vote-Based Fusion

$S \leftarrow$ MAJORITY VOTE(S_{c1}, S_{c2}) with a $(N_w \times N_w)$ window

3.4.2 Estimation of the optimal combination of color models

This Section presents an original procedure for estimating the optimal combination of color spaces for each considered BGS, *i.e.*, for finding the set of color models achieving the best complementarity of the different motion detection results required to finally achieve the best fusion result.

To this end, it is worth recalling that the complementarity/diversity between motion detection results (to be combined) is essential and recommended to ensure a reliable fusion result. Indeed, the more varied the set of initial segmentations is, the more information is available for the fusion process and conversely, it is also logic to think that a combination of similar segmentation solutions is unable to give an improved segmentation map that outperforms the initial individual detections to be fused [5]. In our application, each motion detection result is a two-class or binary segmentation map and can be represented as a binary vector. In order to statistically measure the similarity existing between two binary maps or vectors, we can rely on the generalized version of the correlation coefficient for binary variables, *i.e.*, the so-called Phi coefficient (or Pearson correlation coefficient introduced by K. Pearson in [206] whose square is also related to the chi-squared test between two binary vectors) :

$$\phi = \frac{n_{11}\, n_{00} - n_{10}\, n_{01}}{\sqrt{n_{1\bullet}\, n_{0\bullet}\, n_{\bullet 0}\, n_{\bullet 1}}} \tag{3.1}$$

where n_{ij} is the count of number of (binary) observations equals to i in the first segmentation and equals to j in the second observation and $n_{1\bullet}$ and $n_{\bullet 0}$ are respectively the total number of "1" in the first vector and "0" in the second vector. In our application, The mean ϕ estimation, previously obtained across a training sequence or during the first frames of the video sequence, is a good indicator of similarity or non-complementarity between two CD results. Indeed, for the specific Zivkovitch's BGS model and on the set of the training sequences of the CDnet database, and for the first eight (most widely used in computer vision) color models, the correlation metrics summarized in Table 3.2 indicate that the pair of color model HSV-LAB (followed closely by RGB-HSV) shows the

most interesting complementarity, in term of detection result which will be confirmed later (in Table 3.3, Fig. 3.3 and 3.4). Also, for the specific Eigen or Median BGS, the pair RGB-LAB turns out the most interesting, in term of complementarity.

TABLE 3.2 : Mean Phi coefficient obtained on some pairs of color models, on the training sequences of the CDnet, for the Zivkovitch's BGS model.

	HSV	**LAB**	**XYZ**	**HSL**	**I123**	**H123**	**YIQ**
RGB	0.56	0.67	0.87	0.72	0.76	0.75	0.76
HSV		0.55	0.60	0.57	0.58	0.56	0.55
LAB			0.71	0.64	0.75	0.77	0.73
XYZ				0.76	0.81	0.79	0.82
HSL					0.67	0.70	0.70
I123						0.92	0.94
H123							0.93

3.5 Results and Discussion

In this Section we quantify the gain obtained by our *supervised*, *semi-supervised* and unsupervised fusion strategy, in terms of F-measure, on CDnet 2014 dataset.
In all the experiments, the size of the spatial window, used in the majority vote-based fusion rule, is $N_w = 9$. In the case of our *supervised* fusion strategy, τ_1 and τ_2 are kept constant (and optimal) for the entire CDnet 2014 dataset. For the *semi-supervised* fusion strategy, only τ_2 is unsupervised and estimated (step 3. of Algorithm 1) and for the *unsupervised* version of our fusion procedure, τ_1 is dynamically updated (step 2. of Algorithm 1) for each frame of CDnet by considering $\rho = 8\%$.

Table 3.3 summarizes the F-measure results obtained for different pairs or triplets of color spaces (CS) for the three variants of our CSF fusion procedure. One can notice several interesting points ; first the more interesting, in the complementarity sense, pair and triplet of CSs are respectively [HSV-LAB] and [RGB-HSV-LAB] (and this, despite the fact that the HSV CS does not seem particularly suited to our problem when used alone). Second, a triplet of CSs such as [RGB-HSV-LAB] does not bring more information to the fusion process compared to the simple pair [HSV-LAB], third the semi-supervised

TABLE 3.3 : F-measure results obtained for some different pairs or triplets of color spaces for the three variants (supervised, semi-supervised, unsupervised) of our fusion procedure and for respectively the Zivkovitch's [2], eigen [3] and median [1] BGS model.

Zivkovitch's BGS	Sup.	Semi-Sup.	Unsup.
[RGB-LAB]	0.628	0.612	0.614
[RGB-HSV]	0.622	0.608	0.620
[RGB-I123]	0.617	0.608	0.615
[RGB-YCbCr]	0.617	0.611	0.615
[RGB-YIQ]	0.616	0.608	0.615
[RGB-H123]	0.616	0.608	0.615
[RGB-XYZ]	0.612	0.603	0.611
[RGB-HSL]	0.607	0.601	0.610
[HSV-LAB]	**0.635**	**0.616**	**0.624**
[HSV-I123]	0.626	0.617	0.624
[HSV-YIQ]	0.625	0.617	0.625
[HSV-H123]	0.624	0.616	0.624
[RGB-HSV-LAB]	**0.636**	**0.616**	**0.621**
[RGB-HSV-I123]	0.624	0.617	0.621
[RGB-HSV-YIQ]	0.623	0.617	0.621
[RGB-HSV-H123]	0.623	0.617	0.622

Eigen BGS	Sup.	Semi-Sup.	Unsup.
[RGB-LAB]	0.392	0.366	0.322
[RGB-HSV]	**0.433**	**0.421**	**0.444**
[RGB-I123]	0.360	0.345	0.366
[RGB-YCbCr]	0.362	0.345	0.368
[RGB-YIQ]	0.362	0.346	0.369
[RGB-H123]	0.361	0.344	0.367
[RGB-XYZ]	0.333	0.336	0.339
[RGB-HSL]	0.309	0.302	0.326
[HSV-LAB]	0.412	0.402	0.428

Median BGS	Sup.	Semi-Sup.	Unsup.
[RGB-LAB]	0.561	0.550	0.535
[RGB-HSV]	**0.557**	**0.552**	**0.556**
[RGB-I123]	0.558	0.546	0.540
[RGB-YCbCr]	0.558	0.545	0.540
[RGB-YIQ]	0.558	0.545	0.540
[RGB-H123]	0.558	0.546	0.540
[RGB-XYZ]	0.547	0.548	0.537
[RGB-HSL]	0.542	0.541	0.538

or unsupervised roughly gives the same performance than the supervised version[5] of our procedure without improvement and even a slight deterioration in classification rate but with the undeniable advantage of not being dependent on a supervised threshold whose the optimal value is given, in the case of the supervised fusion procedure, by a computationally expensive exhaustive search algorithm using the ground truth labels and finally four, in the case of the Zivkovitch's BGS, a bi-color space fusion allows us to easily improve the F-measure of 16% (from 0.55 to 0.64) which is significant since the best BGS up to date[6] [207] is 0.74 (only $\approx$ 15% better) and [180] with 0.77. In the case of

[5] It is important to mention that this threshold is optimal for the CDnet 2014 dataset, but may not be valid and suitable for another dataset and that the exhaustive search required to find this optimal threshold is computationally demanding and uses the ground truth labels. Let us also note that, generally, the change detection algorithms, which are currently referenced on CDnet 2014, rely mainly on a main threshold which is optimal (or supervised) for this dataset.

[6] Let us also mention the recent interactive method proposed in [208] which reaches 0.92 with a pure machine learning approach.

the eigen-BGS [3], this improvement is around 55% (from $F = 0.290$ to $F = 0.444$).

Figures 3.3 and 3.4 show us some examples of binary motion segmentations obtained from the Zivkovitch's BGS algorithm operating in only one CS ; namely the RGB, LAB or HSV CS and the fusion result achieved with our supervised and unsupervised fusion model in different pairs of CS. One can notice the complementarity between the segmentation maps to be fused obtained in these three color spaces. In these three different color spaces, the Zivkovitch's BGS estimation model reacts differently and complementarily to color information, textures, noise, illumination changes, highlights, shadows, turbulence, etc. The same phenomenon happens with our semi or unsupervised fusion model which tends to get the same number of estimated *foreground* detections between the two segmentations to be fused (since τ_2 is adjusted in this way, see Algorithm 1. step 3.) but the set of *foreground* responses are a bit different between the two binary maps (while being equally valid for each of them ; as would a processing on data coming from multiple sensors) and this difference is exploited, in a complementary way, in our fusion model.

Detailed evaluation results, in terms of F-measure, of the supervised version of the [HSV-LAB] fusion algorithm for each category of the CDnet 2014 dataset is reported in Table 3.4. Fig. 3.5 shows a comparison of the results from different BGS approaches referenced on CDnet (and on the same subset of images tested and shown in [180], for a fair comparison) such as the SuBSENSE [9], FTSG [10], CwisarDH [11], Spectral-360 [12], AMBER [13], KNN [2], SCSOBS [14], KDE-ElGammal [15], RMoG [16] algorithm and finally our CSF algorithm ([HSV-LAB] supervised fusion approach using the Zivkovitch's BGS.

The proposed algorithm is currently implemented in C++ and with default settings, it is able to achieve real time operation with 23 FPS (Frames Per Second) (Average processing time per frame : 43 ms) for images on an Intel core i7 PC with 16GB RAM (cf. Tab 3.5). Nevertheless, as already said, the computation time could be reduced by half if we use a multi-core processor and if each algorithm (operating in each color space) is run in parallel simultaneously on different cores.

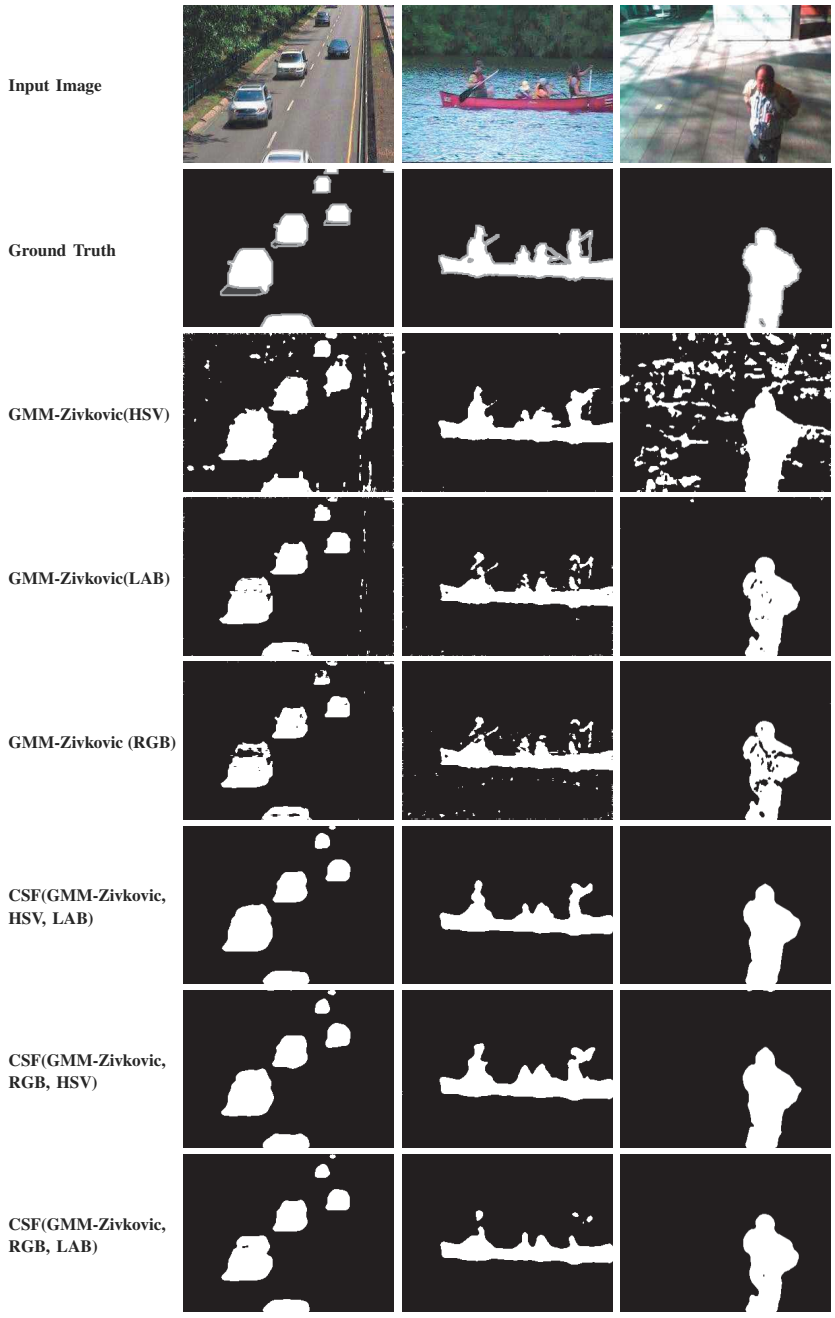


FIGURE 3.3 : Examples of binary motion segmentations obtained from the Zivkovitch's BGS operating in the RGB, HSV, LAB color space and the fusion result achieved with our supervised fusion model in different pairs of color spaces.

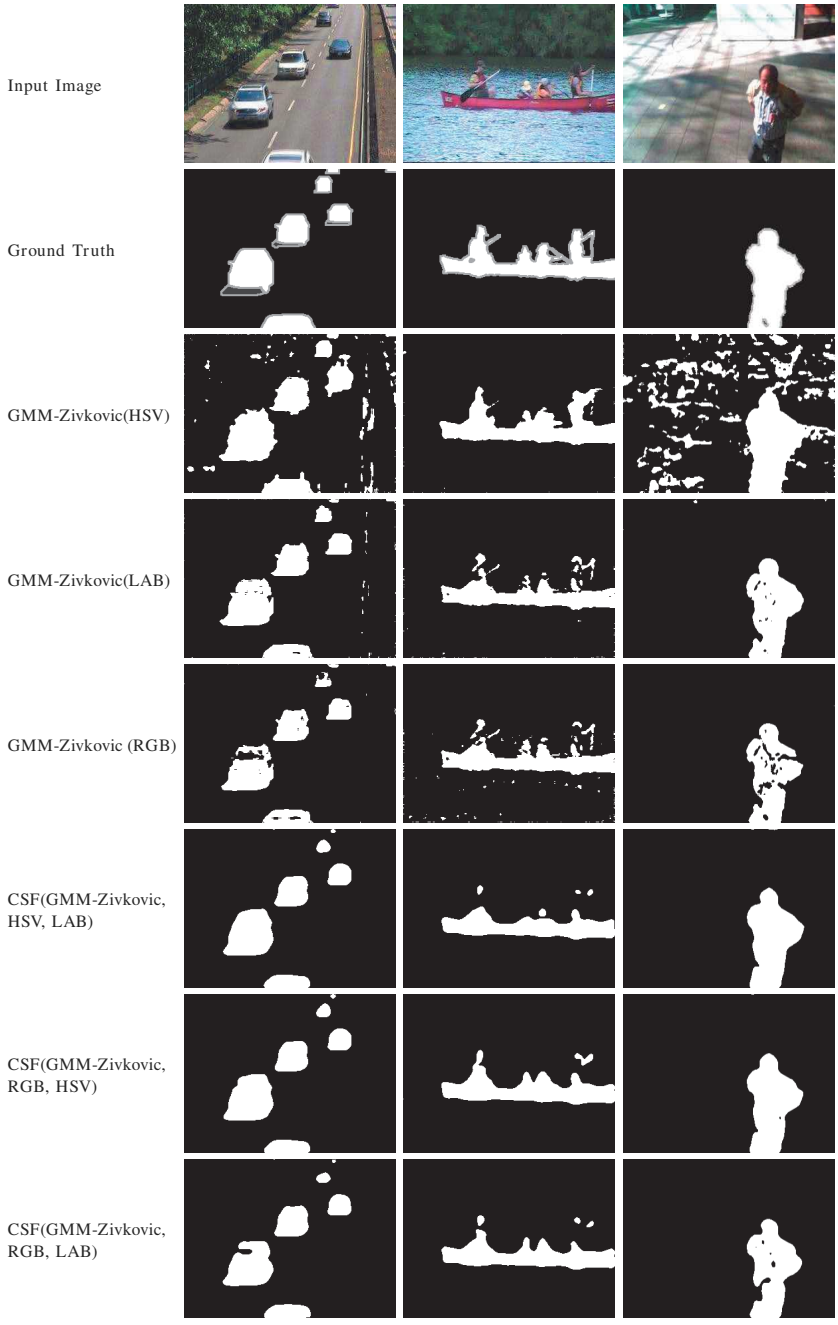


FIGURE 3.4 : Examples of binary motion segmentations obtained from the Zivkovitch's BGS operating in the RGB, HSV, LAB color space and the fusion result achieved with our unsupervised fusion model in different pairs of color spaces.

FIGURE 3.5 : Comparisons of results (on different CDnet 2014 images) from different algorithms : from top to bottom respectively ; CSF (*i.e.* our [HSV-LAB] supervised fusion approach using the Zivkovitch's BGS), SuBSENSE [9], FTSG [10], CwisarDH [11], Spectral-360 [12], AMBER [13], KNN [2], SCSOBS [14], KDE-ElGammal [15], RMoG [16] algorithm.

TABLE 3.4 : Detailed evaluation results, in terms of F-measure, of the [HSV-LAB] Zivkovitch's BGS fusion algorithm for each category of the CDnet 2014 dataset.

Category	Supervised	Semi-supervised	Unsupervised
Baseline	0.871	0.866	0.871
Dyn. Backg.	0.742	0.741	0.752
Camera Jitter	0.763	0.746	0.738
Inter. Obj. Mot.	0.521	0.466	0.458
Shadow	0.814	0.812	0.813
Thermal	0.629	0.569	0.620
Bad Weather	0.756	0.753	0.784
Low Frame-rate	0.625	0.600	0.604
Night Videos	0.365	0.364	0.369
PTZ	0.227	0.230	0.228
Turbulence	0.676	0.631	0.644
Overall	0.635	0.616	0.625

TABLE 3.5 : Real time operation : CSF(GMM-ZIVKOVIC) (unsupervised)

	Baseline (1700 frames)	**Dynamic-Background-Canoe (1189 frames)**	**Shadow-People-In-Shade (1199 frames)**
	Average processing time per frame (ms)		
CSF(GMM-ZIVKOVIC) HSV-LAB	54 ms	72 ms	89 ms
CSF(GMM-ZIVKOVIC) RGB-HSV	27 ms	53 ms	54 ms
CSF(GMM-ZIVKOVIC) RGB-LAB	50 ms	77 ms	85 ms

3.6 Conclusion

In this paper, we have presented an original and adaptive fusion procedure which is able to noticeably improve, in terms of classification rate, the efficiency of any single-threshold BGS approaches. The basic and underlying idea behind this model is to take advantageously into account the complementarity (to noise, illumination changes, highlights, color information, etc.) of the detection results generated in different color spaces. This procedure has the merit to be both low complex in terms of implementation (i.e., simple to implement) or processing time (due to its potential strong multi-core scalability) or memory usage. In addition, the proposed strategy is also low-complex in term

of model. This means that the proposed approach does not rely on complex additional constraints, feedback structures or *a priori* knowledge about the statistical properties of data in order to improve the result. This ensures that the proposed procedure is not prone to over-fit the data and that the obtained gain is reliable. This procedure can also be easily combined with most of the model improvements recently proposed in the change detection domain.

Troisième partie

Consensus de segmentation pour la classification d'imagerie médicale : application au diagnostic de la maladie d'Alzheimer

CHAPITRE 4

CLASSIFICATION OF ALZHEIMER'S DISEASE SUBJECTS FROM MRI USING THE PRINCIPLE OF CONSENSUS SEGMENTATION

Cet article a été publié dans la conférence *19th Irish Machine Vision and Image Processing conference (IMVIP'17)* comme l'indique la référence bibliographique.

A. Khlif, M. Mignotte. Classification of Alzheimer's disease subjects from MRI using the principle of consensus segmentation. *19th Irish Machine Vision and Image Processing conference, IMVIP'17*, published by the Irish Pattern Recognition & Classification Society, pages 123-130, Maynooth University, Co. Kildare, Ireland, September 2017.

Cet article est presenté içi dans sa version originale.

Abstract

In this paper, we develop an original and reliable detection and classification framework for Alzheimer's Disease (AD) in structural Magnetic Resonance Images (MRI). This work exploits recent advances made in segmentation and multimedia indexing and classification for Content Based Visual Information Retrieval (CBVIR). More precisely, It exploits the concept of consensus segmentation to define two segmentation prototypes (Normal Control and Alzheimer's disease) of the brain in terms of cerebral Grey Matter (GM), White Matter (WM) and Cerebro-Spinal Fluid (CSF). The classification is then performed by computing the Pott distance of each image w.r.t the two prototypes. Based on a threshold on the computed distance, brain images are classified using either Minimal Distance (MD) or K-Nearest-Neighbors (KNN) classifier. Our approach has been evaluated on the baseline MR images of 98 subjects from the Open Access Series of Imaging Studies (OASIS) database. The used subset consists of 49 subjects who have been diagnosed with very mild to mild AD and 49 non-demented individuals. The experimental results show that our classification of patients with AD versus NC

subjects achieves accuracy of 86%. Results demonstrate very promising classification performance while being simple compared to the state-of-the-art classical volumetric AD diagnosis methods.

Keywords : Consensus segmentation, Alzheimer's disease, Normal control, Prototype, KNN, OASIS database, MRI, Image classification, Machine learning techniques, Clinical dementia rate.

4.1 Introduction

Alzheimer's disease (AD) is an irreversible neurodegenerative dementia that occurs most frequently in older adults and that gradually destroys regions of the brain that are responsible for memory, learning, thinking, and behavior [4]. Current estimates indicate that 5.3 million Americans of all ages are afflicted with this illness and this number is expected to increase to 16 million people by 2050, unless a cure is found. The socio-economic consequences of this increase are cumbersome and makes early diagnosis of AD a public health emergency.

Medical information from structural Magnetic Resonance Imaging (sMRI) has long time been the most used neuroimaging modality to detect brain atrophy in AD studies. In fact, two main families of methods can be distinguished to extract features from MRI for AD classification. Several studies report the use of sophisticated measurement techniques that assess anatomical changes in areas compromised by AD such as the Hippocampal Volume (HV), the Lateral Ventricles Volume (LVV), CSF Volume (CSFV), etc., These (so-called volumetric) methods are only based on form, size and/or shape derived features extracted from the brain structures.

Aside from volumetric approaches, morphometric methods have gained great interest among which we can distinguish : Voxel Based Morphometry (VBM) [209] which is a widely used whole-brain analysis method, which allows an exploration of the differences in local concentrations of grey matter and white matter. Tensor Based Morphometry (TBM) [210] was proposed to identify local structural changes from the gradients of deformations fields. Object Based Morphometry (OBM) [211] was introduced to per-

form shape analysis of anatomical structures and recently, Features Based Morphometry (FBM) [212] was proposed as a method for relevant brain features comparison using a probabilistic model on local image features in scale-space [213]. Voxel based methods work directly on the voxel grid and are computationally very efficient. An advantage of these approaches, compared to the ROI-based volumetric methods, is the fact that they do not require *a priori* assumptions about the location, the size or number of ROIs to be analyzed, since they provide voxel wise measures determined in the entire brain. More, there is no evidence that other regions (except hippocampus and entorhinal cortex) did not provide any information for AD and NC [214]. Recent studies on AD diagnosis found that the quantity of CSF is a biomarker of AD [215]. Indeed, smaller hippocampal volume is associated with greater CSF amount [216].

With the features estimated, either by volumetric or voxel-based methods, on training cases, a classifier can be trained and applied to predict the diagnosis of a testing case, whose features are extracted in the same way. Among the most popular classifiers : Linear Discriminant Analysis (LDA) [217], Neural Network (NN) [218], Support Vector Machines (SVM) either with linear or non-linear [219–222], K-Nearest-Neighbors (KNN) [223] and Bayes classifier.

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men to the extent that it is possible to discriminate the gender *via* MRI analysis. Moreover, it has been shown that VBM is sensitive to the gender differences [224]. For these reasons, we have been very cautious in this study ; therefore, as proposed in [4, 218, 225, 226], we have selected a set of 98 MRI women's brain volumes. It must be noted that this is a large number of subjects compared with the other studies referred above as it has also been mentioned in [226].

In this work we propose an automatic content analysis based framework for recognition of AD using MRI scans. Image content analysis and classification methodologies are now more and more used for medical information mining and retrieval [227, 228] with the aim of Computer-Aided Diagnosis (CAD). The proposed method exploits recent advances made in segmentation and multimedia indexing and classification for Content Ba-

sed Visual Information Retrieval (CBVIR) and, more precisely, the concept of consensus segmentation to define two segmentation prototypes (prototype_AD and prototype_NC) of the brain in terms of cerebral grey matter, white matter, and cerebro-spinal fluid. This two prototypes are then cleverly combined with the KNN algorithm to increase the classification results.

4.2 Experimental study

In order to investigate the detection performance of the proposed MDC-KNN (Minimum Consensus Distance-KNN) classifier, a set of appropriate experiments were conducted. For the experimental purposes, specific software was developed in C++. All experiments were executed in an Intel i7 3.3 GHz PC with 16 GB RAM.

4.2.1 Dataset description

The data analyzed in this paper was obtained from the OASIS database[1]. As proposed in [4, 218, 225, 226], a subset of the OASIS database [229] was selected in order to evaluate the detection performance of the proposed classifier. This two-class dataset has across-sectional collection of 416 subjects covering the adult lifespan aged $(18-96)$ including individuals with early-stage Alzheimer's disease. A subset of the original OASIS dataset including 98 right-handed women (aged $65-96$ years) is considered herein. More precisely, the used subset consists of 49 subjects who have been diagnosed with very mild to mild AD (class 1) and 49 non-demented (class 2). The designation of demented and non-demented is based on the Clinical Dementia Rate (CDR). A CDR of 0 corresponds to the normal cognitive state, CDR > 0 to dement. The demographic information of these subjects is summarized in Table 4.1.

[1] `http://www.oasis-brains.org`

TABLE 4.1 : Demographic information of the subjects in the two classification classes

	Very mild to mild AD	Normal Controls
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5/1/2)	31/17/1	0
MMSE	24 (15-30)	28.96 (26-30)

4.2.2 MRI data preprocessing

Before any further analysis, the images must be pre-processed to eliminate variations due to the different MR acquisitions. The preprocessing steps are summarized in Figure 4.1 for the T1 image of one subject.

All (axial section) MRI images were : (a) corrected for inter-scan head movement [229] and rigidly aligned to the Talairach and Tournoux space [230], (b) transformed to a template with a 12-parameter affine registration an merged into a 1-mm isotrpic image [229], (c) skull-stripped and corrected for intensity inhomogenity [231], and finally (d) segmented into $K = 3$ classes corresponding to the three existing cerebral tissues, respectively ; "cerebrospinal fluid", "white matter" and "grey matter". This 3-class segmentation is obtained by applying firstly, 10 times the K-Means clustering algorithm with different seeds and different number of neighbors of the pixel to be classified and secondly by combining them with a simple majority vote scheme[2]. In our application, the majority vote is achieved with a spatial window (of size 3×3 pixels and centered on the pixel s to be classified) that collects the class labels of the 10 segmentation results obtained by each K-mean clustering and by finally assigning to that central pixel s, the class label that has the majority vote. This strategy ensures both an efficient spatial regularization of the final segmentation result and also a reliable decision fusion between results obtained by these K-mean clusterings. In this segmentation, the "CSF", the "white matter" and the "grey matter" are represented by a dark, a grey, and a white

region respectively, in order to visually express the activity level of the blood flow.

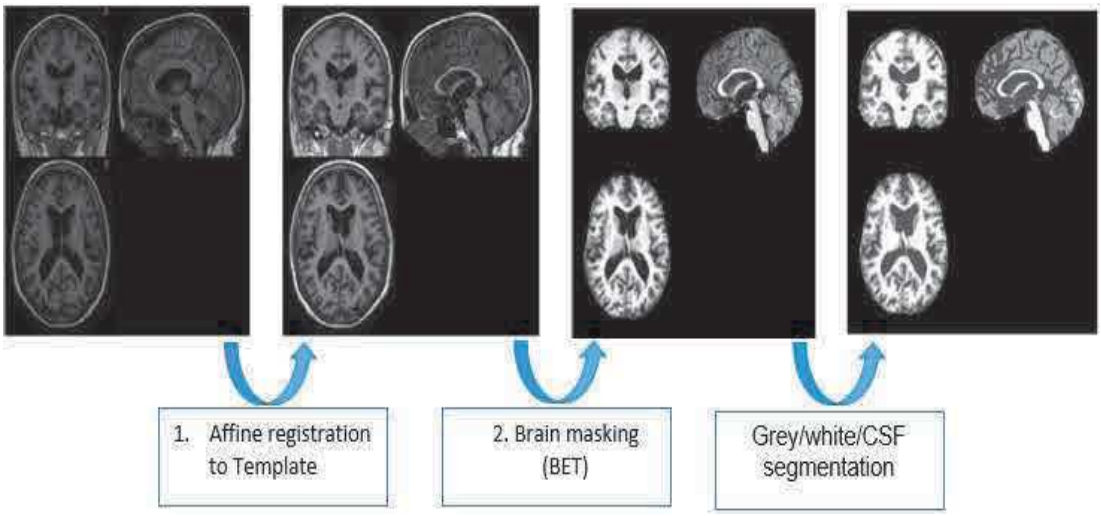


FIGURE 4.1 : Snapshot of a specific subject. (a) One original scan. (b) Atlas-registered image. (c) Brain masked version of Figure 1(b). (d) The GM/WM/CSF segmentation image.

4.2.3 Prototypes NC and AD

From the above-mentioned segmentations (416 subjects), two segmentation prototypes (prototype_AD and prototype_NC), in terms of cerebral GM, WM and CSF are built by combining all segmentations related to subjects who have been diagnosed with very mild to mild AD (class 1 which contains 100 subjects) and those related to the non-demented (class 2 which contains 316 subjects) with the majority vote scheme. The processing is illustrated in Figure 4.2, whereas the overall procedure is explained in the next section.

[2] We have noticed that these segmentation technique provides better classification results than a SPM software (Statistical Parametric Mapping) based segmentation technique (the classification accuracy is better by 2%).

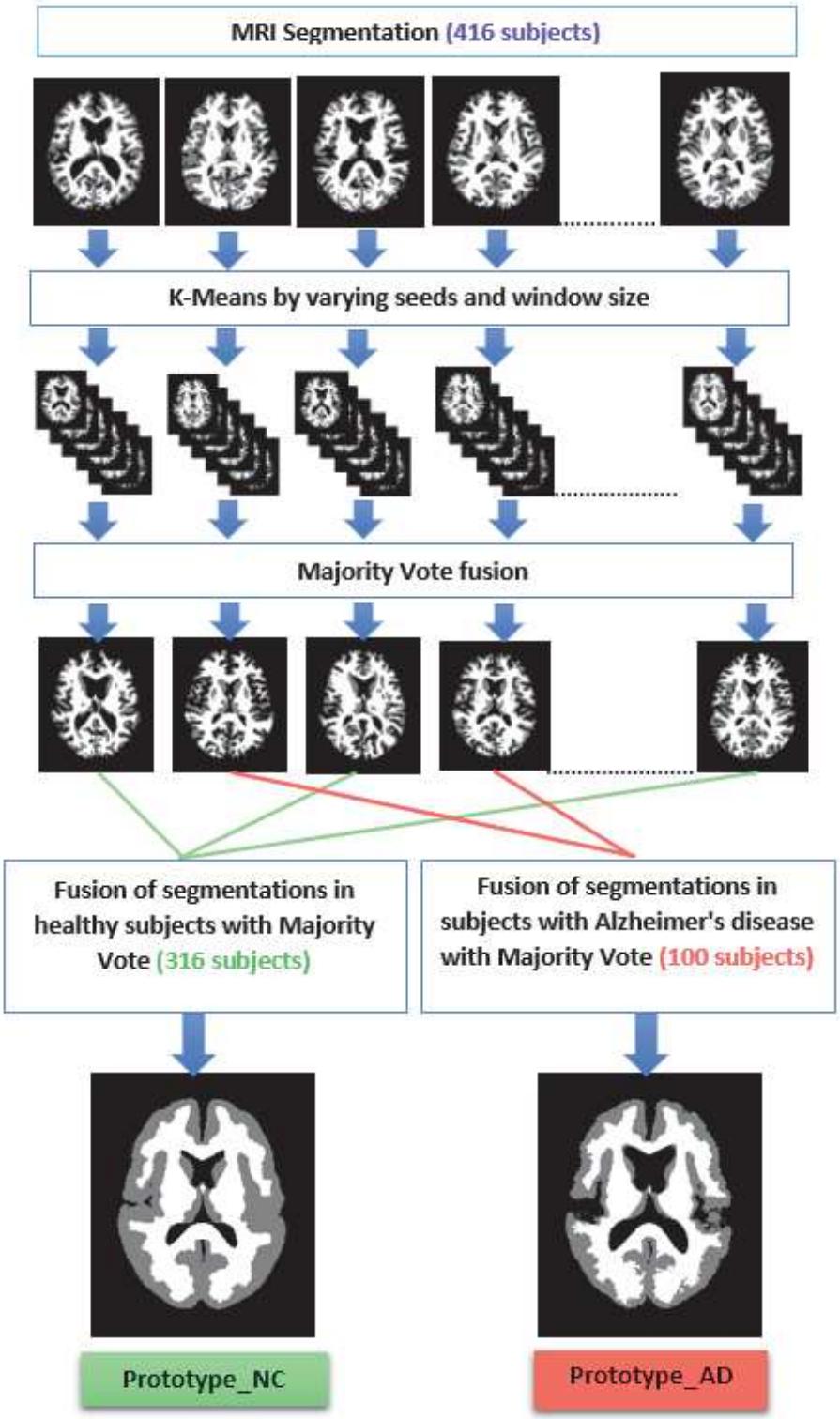


FIGURE 4.2 : Methodology for the creation of NC and AD prototypes

4.2.4 Hybrid classification

The proposed hybrid classification technique named Minimum Consensus Distance-KNN (MDC-KNN), combining the previously estimated prototypes (Prototype_NC and Prototype_AD) and the KNN algorithm (weighted as proposed in [232] with the Pott distance in our case) consists of three stages :

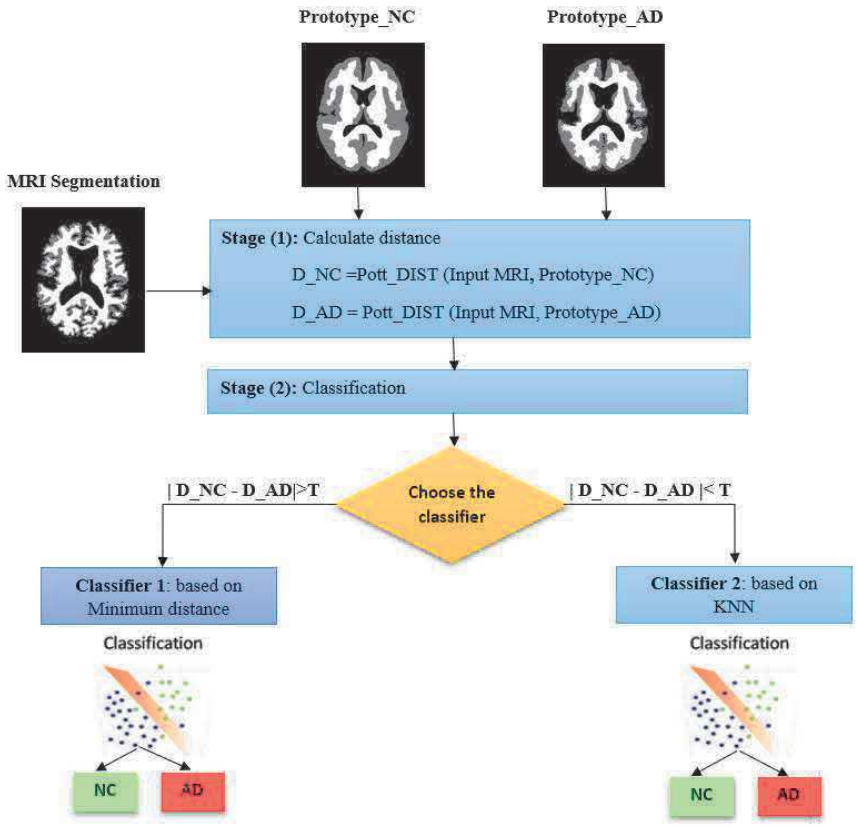


FIGURE 4.3 : The methodology of the proposed method.

- Calculate the two Pott distances (the normalized number of differences in percentage[3]) between the segmentation related to an input MRI image and respectively the Prototype_NC (D_NC) and the Prototype_AD (D_AD).

- Choose the classifier KNN or (prototype-based) Minimal distance (MD). The choice of classifier (KNN or MD) depends on the difference between D_NC and D_AD. If the difference is large (greater than a threshold T which was set to 1.5%[4]) then,

in this case (case 1 or 2 of Figure 4.4), we choose the MD classifier since in this case we are certain that the segmentation is very close to one of the two prototypes and consequently we can thus rely on this classification procedure. Otherwise, we apply the KNN classifier (case 3 of Figure 4.4) in with the Pott distance between the segmentation related to an input MRI image and the 3-class segmentations obtained for each image of the test set.

- Classification.

The schematic diagram for proposed methodology is shown in Figure 4.3.

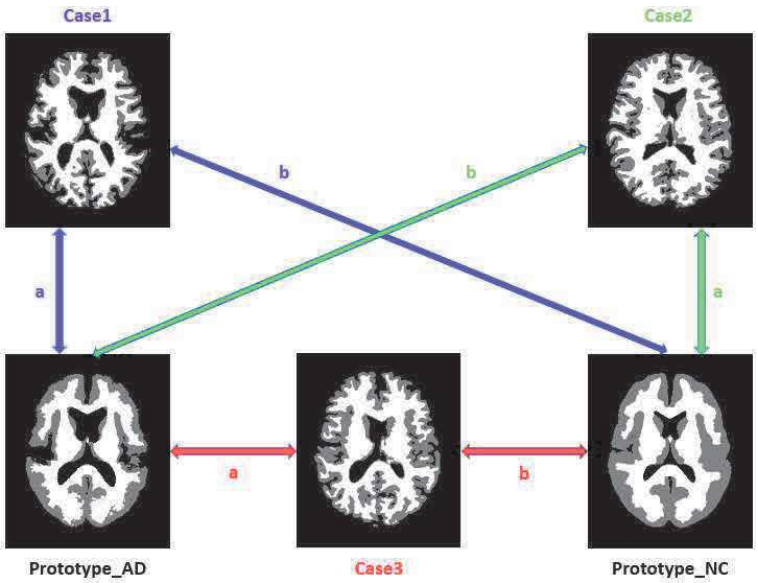


FIGURE 4.4 : $|D\_NC - D\_AD|$ large (case 1 or 2) or very low (case 3)

[3] $Pott\_Dist(S1,S2) = \frac{\Sigma_{i=1}^{N} 1-\delta_{x_i^1 x_i^2}}{N}$ Where δ is the Kronecker delta function and the summation is over all the N pixels of the segmented image and x_i^1, x_i^2 is respectively the label of the first and second segmentation at pixel i.

[4] We have computed than the distance (in percentage of pixel difference) between Prototype_NC & Prototype_AD is 15%. Based on this, we have set this threshold T to an order of magnitude ($\div 10$) lower (i.e., $T = 1.5\%$).

4.3 Experimental results

We evaluate the performance of the proposed method in terms of sensitivity=TP/(TP + FN), specificity=TN/(TN + FP) and accuracy=(TN + TP)/(TN+TP+FN+FP). Where True Positives (TP) are AD patients correctly identified as AD, True Negatives (TN) are controls correctly classified as controls, False Negatives (FN) are AD patients incorrectly identified as controls and False Positives (FP) are controls incorrectly identified as AD. Sensitivity is the proportion of AD subjects correctly classified, and the specificity is the proportion of correctly classified controls.

The actual feature datasets (98 subjects : 49 AD and 49 NC) have been used in several works in the literature, hence results obtained with a variety of classifier models are publicly available for comparison. The AD detection statistics, accuracy, sensitivity and specificity, of the proposed MDC-KNN classifier by performing a leave-one-out cross validation test, in comparison with the state-of-the-art models [4, 225, 233] are summarized in Table 4.2.

By examining the results of Table 2, it follows that the proposed classifier MDC-KNN with $K = 3$ (This number was empirically tested) demonstrated superior performance than some conventional classifiers such as RBF, MLP-BP, PNN, Linear SVM (8% higher accuracy) and some advanced classification models (Indep-linear-SVM, Indep-rbf-SVM, linear-AB-SVM, rbf-AB-SVM, Kernel-LICA-DC, LVQ1, LVQ2, rbf-DAB-SVM). Let us note that the single KNN classifier (witout being combined with our prototype based MD classifier) reaches a classification value of 73% (See in Table 4.2 KNN-Pott-MV).

4.4 Conclusion and future works

In this paper, we presented a novel classification framework, based on prototype defined using consensus segmentation, is proposed for Alzheimer's disease. The proposed framework learns prototypes for Normal and AD class by first segmenting each image into CSF, WM, and GM (tissue classes) using K-means with 10 different seeds/number-of-neighbors, followed by a majority vote. All the segmented images (for both AD and

TABLE 4.2 : Classification results comparison between our method and the morphometric approaches proposed in [4] and several other approaches (using a leave-one-out cross validation test and the same dataset).

Classifier type	Accuracy	Sensitivity	Specificity
MDC-KNN (our approach)	0.86	0.85	0.87
rbf-DAB-SVM	0.85	0.78	0.92
LVQ2	0.83	0.74	0.92
LVQ1	0.81	0.72	0.90
LC-KNN	0.80	0.80	0.79
rbf-AB-SVM	0.79	0.78	0.80
MLP-BP	0.78	0.69	0.88
PNN	0.78	0.62	0.94
Linear SVM	0.78	0.72	0.88
Indep-rbf-SVM	0.75	0.56	0.95
Kernel-LICA-DC	0.74	0.96	0.52
Indep-Linear-SVM	0.74	0.51	0.97
KNN-Pott-MV	0.73	0.61	0.85
Linear-AB-SVM	0.71	0.54	0.88
RBF	0.66	0.65	0.68

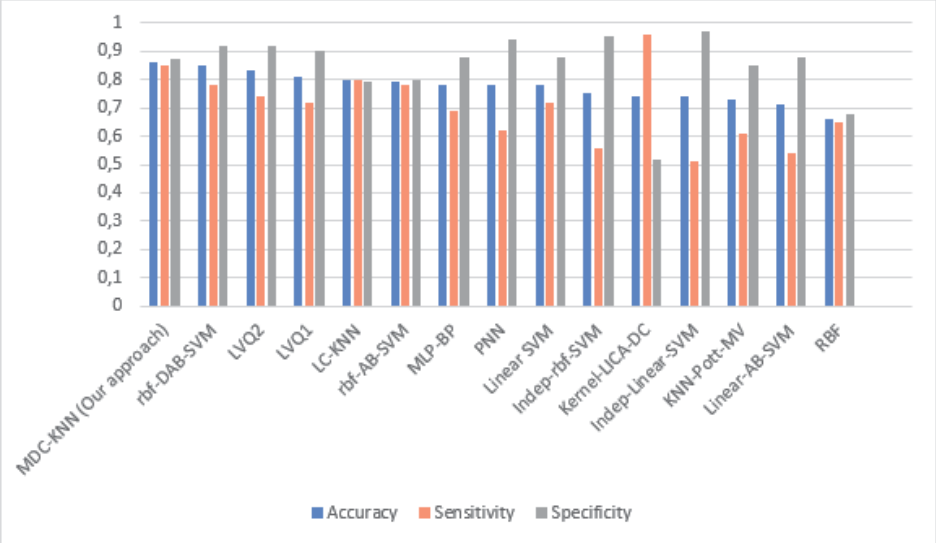


FIGURE 4.5 : Classification results for the different methods.

NC) are then combined using majority vote to generate the prototypes. The classification is then performed by computing the Pott distance of each image w.r.t the two prototypes. Based on a threshold on this distance, brain images are then classified using either the MD or KNN classifier. Experiments are conducted using 98 subjects and compared with various classifiers.The aforementioned experimental results, in particular through the high performance of the proposed classifier in terms of accuracy, have demonstrated the remarkable potential of prototypes (prototype_AD and prototype_NC) in classification applications. A possible extensions to the current work is the use of another more advanced fusion algorithm for the creation of prototypes. The strength of the proposed work consists in the following :

- Our approach is less time-consuming compared to the state of the arts Volumetric methods, computer-based and does not require the intervention of an expert during the classification/retrieval phase.
- It is extensible to other diseases that can be diagnosed by brain MRI such as Schizophrenia.
- The method could be extended by combining axial, coronal, and sagittal MRI data for improving the classification accuracy.

CHAPITRE 5

CLASSIFICATION OF ALZHEIMER'S DISEASE SUBJECTS FROM MRI USING THE PRINCIPLE OF CONSENSUS SEGMENTATION BASED ATLAS

Cet article a été accepté avec modifications mineures au journal *Computerized Medical Imaging and Graphics* comme l'indique la référence bibliographique.

A. Khlif, M. Mignotte. Classification of Alzheimer's disease from MRI using the principle of consensus segmentation based atlas. *Computerized Medical Imaging and Graphics* (Elsevier), February 2018.

Cet article est presenté içi dans sa version originale.

Abstract

In this paper, we develop an original and reliable detection and classification framework for Alzheimer's Disease (AD) in structural Magnetic Resonance Imaging (MRI). This work exploits recent advances made in segmentation and multimedia indexing and classification for Content Based Visual Information Retrieval (CBVIR). More precisely, our strategy exploits the concept of consensus segmentation to build two segmentation prototypes (Prototype Normal Control (Prototype_NC) and Prototype Alzheimer's Disease (Prototype_AD)) of the brain in terms of cerebral Grey Matter (GM), White Matter (WM) and Cerebro-Spinal Fluid (CSF) regions. A first classification is then performed by computing a specific distance between the three-class segmentation of each MR brain image w.r.t the two prototypes. Based on a threshold on this computed distance, brain images are then classified using either Minimal Distance (MD) or K-Nearest Neighbors (KNN) classifier. Our approach has been evaluated on the baseline data of MR images of 98 subjects from the Open Access Series of Imaging Studies (OASIS) database, which contains a large number of subjects compared to current reported studies. The used subsets consist of 49 subjects who have been diagnosed with very mild to mild AD and 49 non-demented individuals. The experimental results show that our classification of

patients with AD versus NC subjects achieves accuracy of 90% by performing a Leave-One-Out Cross Validation (LOOCV). Results demonstrate the validity of the proposed method and especially its simplicity and high accuracy compared to the other state-of-the-art AD diagnosis approaches proposed in the literature.

Keywords : Alzheimer's disease, classification, consensual segmentation, Global Consistency Error (GCE), K-Nearest Neighbors (KNN), Minimal Distance Classifier (MDC), Magnetic Resonance Image (MRI).

5.1 Introduction

Alzheimer's disease is an irreversible neurodegenerative dementia that occurs most frequently in older adults and that gradually destroys regions of the brain that are responsible for memory, learning, thinking, and behavior [4]. Current estimates indicate that 5.3 million Americans of all ages are afflicted with this illness and this number is expected to increase to 16 million people by 2050, unless a cure is found. The socio-economic consequences of this increase are cumbersome and makes early diagnosis of AD a public health emergency.

Medical information from structural Magnetic Resonance Imaging (sMRI) has long time been the most used neuroimaging modality to detect brain atrophy in AD studies [234, 235]. In fact, two main families of methods can be distinguished to extract features from MRI for AD classification which are the morphometric methods and volumetric methods. Several studies report the use of sophisticated measurement techniques that assess anatomical changes in areas compromised by AD such as the Hippocampal Volume (HV), the Lateral Ventricles Volume (LVV), CSF Volume (CSFV) [216, 221, 236]. These (so-called volumetric) methods are only based on form, size and/or shape derived features extracted from the brain structures. Indeed, such volumetric measurements, require the segmentation of these ROIs from the MR images, most often manually. Furthermore, a priori assumptions about the expectedly affected brain structures is needed to select the appropriate ROI [237]. Aside from volumetric approaches, morphometric methods have gained great interest among which we can distinguish : Voxel Based Mor-

phometry (VBM) [209] which is a widely used whole-brain analysis method, which allows an exploration of the differences in local concentrations of grey matter and white matter. Tensor Based Morphometry (TBM) [210] was proposed to identify local structural changes from the gradients of deformations fields. Object Based Morphometry (OBM) [211] was introduced to perform shape analysis of anatomical structures and recently, Features Based Morphometry (FBM) [212] was proposed as a method for relevant brain features comparison using a probabilistic model on local image features in scale-space [213, 238].

Voxel based methods work directly on the voxel grid and are computationally very efficient. An advantage of these approaches, compared to the ROI-based volumetric methods, is the fact that they do not require *a priori* assumptions about the location, the size or number of ROIs to be analyzed, since they provide voxel wise measures determined in the entire brain [237]. More, there is no evidence that other regions (except hippocampus and entorhinal cortex) did not provide any information for AD and NC [214]. Recent studies on AD diagnosis found that the quantity of CSF is a bio-marker of AD [215]. Indeed, smaller hippocampal volume is associated with greater CSF amount [216].

Nevertheless, these methods are less accurate due to the limited resolution of the voxel grid and less robust to noise, mainly because of the inherent inter subject anatomical variability and the effects of a brain pathology [237]. Indeed, some pathologies may affect not only a single anatomical structure or interconnected regions, but specific structures localized far away from each other. This kind of patterns are difficult to find and analyze with the standard morphometrical techniques [237]. To cope with this issue, we have developed an automatic content-based framework for recognition of Alzheimer's disease subjects using MRI segmented scans with the concept of consensus segmentation. In our case, the principle of *consensus* segmentation allows us to build two reliable segmentation-based prototypes, one corresponding to healthy individuals and the second one corresponding to unhealthy subjects (with AD) from a training database of previously segmented brain MR images. First, let us stress us that the segmentation is a representation of an image into something that is easier to analyze than the MR image itself. In addition, these two consensus segmentation-based prototypes have also several

appealing advantages whose the first one is to suppress or remove undesired components in the brain image (to be classified) such as the anatomical variability existing between individuals which are not relevant for the detection and quantification of AD. The second advantage is that the segmentation-based prototypes allows us to not to take into account the noise of the MR imaging modality into the classification procedure. In addition and finally, the difference between these two prototypes also allows us to localize the regions in the brain that are affected by the AD or to identify ROIs with a specific textural pattern in the segmentation which could represent an interesting early bio-marker of this disease (cf. Fig 5.3). These two segmentation-based prototypes (NC and AD), built from a training segmented database can be viewed and understood as being a mean and "denoised atlas" of an healthy or unhealthy subjects (with AD). Indeed, in estimation theory, an estimator based on the average (or the weighted average) operation generally yields to an optimal denoised solution (when the noise is uncorrelated). In the consensus theory, this average (or consensus or compromise) can be achieved according to different defined criteria (cf. Sec 5.2.3).

With the features estimated, either by volumetric or voxel-based methods, on training cases, a classifier can be trained and applied to predict the diagnosis of a testing case, whose features are extracted in the same way. Among the most popular classifiers : Linear Discriminant Analysis (LDA) [217], Neural Network [218], linear or non-linear Support Vector Machines (SVM) [219–222] and KNN [223].

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men to the extent that it is possible to discriminate the gender *via* MRI analysis. Moreover, it has been shown that VBM is sensitive to the gender differences [224]. For these reasons, we have been very cautious in this study ; therefore, as proposed in [4, 218, 225, 226, 233, 239–244], we have selected a set of 98 MRI[1] women's brain volumes. This subsample from OASIS contains the same number of AD patients and controls, therefore is a well class balanced sample [233]. It must be noted that this is a large number of subjects compared with the other studies referred above as it has also

been mentioned in [226].

In this work we propose an automatic content analysis based framework for recognition of AD using MRI scans. Image content analysis and classification methodologies are now more and more used for medical information mining and retrieval [227, 228] with the aim of Computer-Aided Diagnosis (CAD). The proposed method exploits recent advances made in segmentation and multimedia indexing and classification for content based visual information retrieval and, more precisely, the concept of consensus segmentation to define two segmentation prototypes (Prototype_NC and Prototype_AD) of the brain in terms of cerebral GM, WM, and CSF. These two prototypes are then used with a Minimal Distance Classifier (MDC) and cleverly combined with the KNN algorithm to increase the classification accuracy.

The rest of the paper is organized as follows : Section 5.2 presents the related work on the consensus of segmentation in imaging, describes the combination model and the consensus model, according to different criteria (or different metrics between two segmentations), to built two segmentation prototypes Prototype_NC and Prototype_AD. Section 5.3 presents the proposed MDC-KNN classifier. Section 5.4 presents experimental results, and some comparisons with several state-of-the-art methods and finally, Section 5.5 concludes this work and outlines its further perspectives.

5.2 Consensus of segmentations in imaging

5.2.1 Introduction

Image segmentation is a frequent pre-processing step in computer vision whose goal is to simplify and transform the representation of an image with a set of coherent significant and spatial regions with similar attributes [245]. This low-level vision task, which changes the representation of an image into something that becomes easier to analyze, is often the preliminary and crucial step in the development of many image-understanding algorithms and computer vision systems. In our case, the segmentation of a brain MR

[1] http://www.ehu.es/ccwintco/index.php/GIC-experimental-databases

image into three kinds of regions (classes), each one associated with a specific brain anatomical tissue (CSF, WM and GM) has the merit to efficiently reduce the information content of a brain MR image and to suppress undesired components such as noise or artifacts which are not relevant for the detection and classification of AD.

Recently, the notion of *consensus* segmentation has recently been proposed in the image processing community. A *consensus* segmentation is conceptually the compromise (in terms of level of details, contour accuracy, number of regions, etc.) exhibited by each segmentation map (or spatial clustering) belonging to a set of segmentations. In some sense, the *consensus* segmentation is the average of all the individual segmentations belonging to a (so-called) segmentation ensemble, according to a defined criterion.

This concept of *consensus* clustering has been first exploited in image processing as an interesting alternative for segmenting complex scenes. Indeed, in this case, a possible and reliable segmentation approach consists in combining (fusing or merging) multiple low-cost and rough image segmentation results of the same scene associated with simpler segmentation model[2] to provide a *consensus* solution which is a final improved segmentation map. In fact, this strategy can be viewed and easily understood as a (special type of) denoising problem in which each segmentation (to be merged) is in fact a noisy segmentation solution (or observation). In estimation theory, an estimator based on the average operation generally yields to an optimal denoised solution (when the noise is uncorrelated). In the same way, in the consensus theory, an interesting denoised segmentation solution also turns to be the average (in a statistical criterion sense) of all the individual segmentations to be combined.

Let us mention that when the different anatomical classes can be easily identified from the spatial clustering result, as it is the case of a three-class segmentation of a brain MR image (since the CSF, grey and white matter exhibit regions with increasing mean value of grey level), the simplest *consensus* segmentation model can be based on the local majority vote criterion. This technique simply consists firstly in collecting, for each

[2] by varying the internal parameters of a given segmentation algorithm (or seeds for stochastic algorithms) and/or by using different features for an input image or simply by using different segmentation algorithms.

pixel the class labels of the different segmentation maps of the segmentation ensemble and secondly by assigning to that central pixel, the class label that has the majority vote. A spatially regularized variant of this method consists in gathering the class labels contained in a 3D window centered on this pixel. Nevertheless, many other consensus criteria in the segmentation field exist and which have also the appealing advantage of not requiring the number of classes and to previously identify the semantic class of each segmented region. For example, we can mention the consensus model proposed in [6] that averages or merges a set of segmentation maps in the sense of inertia or intra-cluster variance criterion (or its variant [84], in the weighted intra-class inertia sense), in the Probability Rand Index [80] (PRI) sense with different optimizers or strategies [19, 78, 83, 85, 86], in the Variation of Information (VoI) [5, 246], Global Consistency Error (GCE) [17], F-measure [18], in the sense of the Support Vector Machine (SVM) criterion [87] (*i.e.,* in the maximum margin hyper-plane sense between the classes) or finally in the sense of the Maximum A Posteriori (MAP) of the Logit distribution [88] to name a few.

It is also worth mentioning that the notion of *consensus* segmentation has been also recently exploited in CBVIR [8] in order to help a user to browse through a large database in an intuitive and efficient manner. For example, it is also used for searching a specific subset or class of images in terms of their segmentation-based descriptive content and, more precisely, according to the geometrical layout and shapes of the different objects detected and segmented within the image (and not based on low level color or texture visual features as it has been commonly and widely proposed in this domain and which is limited in usefulness).

5.2.2 Consensus segmentation model principle

Let us consider that we have at our disposal, a set of L segmentations $\left\{S_k\right\}_{k\leq L} = \left\{S_1, S_2, ..., S_L\right\}$ of size N pixels to be averaged (or combined) in order to obtain either a final improved segmentation result relatively to each member of $\left\{S_k\right\}_{k\leq L}$ if each segmentation map S_k results from the same image or either to build a prototype if each segmentation is related to an anatomical structure affected by a same disease. In this latter

case, which is herein considered, it allows us to provide a kind of segmented anatomical atlas that captures the segmented anatomy of unhealthy anatomical structure without taking into account the inherent anatomical shape variability existing between individuals which are not relevant for the detection and quantification of Alzheimer disease.

Conceptually, most of the consensus criteria used to fuse or average a set of segmentations, is based upon the definition of a distance or metric between two segmentation maps (measuring the similarity between the two clusterings). This is straightforwardly the case for the VoI [5], GCE [17], PRI [19] F-measure [18] criterion and this distance can be easily found for the other criteria (*e.g.*, for the majority vote criterion this distance could be simply the sum (for each pixel) of the proportion of labels that are different to the label associated with the final majority vote).

Based on this distance, the estimation of the consensus or prototype of these L segmentations can be conceptually defined as the segmentation solution $\hat{S}$ which is at the center of the segmentation ensemble $\left\{S_k\right\}_{k\leq L}$ or equivalently as being the segmentation $\hat{S}$ that minimizes the considered average pairwise distance between the consensus segmentation $\hat{S}$ and all other segmentation S_k of the segmentation ensemble. Formally, this optimization problem called the *median partition* [79]) problem can be expressed as :

$$\hat{S}_{\text{Dist.}} = \arg\min_{S\in\mathcal{S}_n}\left\{\frac{1}{L}\sum_{k=1}^{L}\text{Dist.}(S,S_k)\right\} \tag{5.1}$$

with $\mathcal{S}_n$ is the set of all possible segmentations of size N pixels and $\sum_{k=1}^{L}\text{Dist.}(S,S_k)$ is commonly called the consensus or cost function of this energy-based consensus model for the considered distance or criterion. This optimization problem can then be solved either with deterministic gradient procedures, algebraic optimization methods, dynamic programming or either with stochastic search or simulation procedures or machine Learning based optimization schemes [5, 17–19, 78, 79, 83, 85–88].

5.2.3 Criteria or distance used in consensus segmentation field

5.2.3.1 Global Consistency Error (GCE)

The Global Consistency Error (GCE) [17] measures the extent to which one segmentation map can be viewed as a refinement of another segmentation. Segmentations which are related in this manner are considered to be consistent, since they could represent the same image segmented at different scales.

Let $S_1 = \{C_1^1, C_2^1, \ldots, C_{R^1}^1\}$ and $S_2 = \{C_1^2, C_2^2, \ldots, C_{R^2}^2\}$ be respectively the first and second segmentation and R^1 and R^2 being respectively the number of clusters C (or segments[3]) in S_1 and S_2. For a given pixel p_i, let $C_{<p_i>}^1$ and $C_{<p_i>}^2$ be the cluster that contain that pixel. The local refinement error (LRE) is then defined at pixel p_i as :

$$\mathrm{LRE}(S_1, S_2, p_i) = \frac{|C_{<p_i>}^1 \setminus C_{<p_i>}^2|}{|C_{<p_i>}^1|} \tag{5.2}$$

where $\setminus$ denotes the set differencing operator and $|C|$ the cardinality of the set of pixels C. As noticed in [17], this clustering (or segmentation) error measure is not symmetric and encodes a measure of refinement in one direction only. A possible and natural way to combine the LRE at each pixel into a measure which is symmetric is to consider :

$$\mathrm{GCE}_{(S_1,S_2)} = \frac{1}{2N}\left\{\sum_{i=1}^{N}\mathrm{LRE}_{(S_1,S_2,p_i)} + \sum_{i=1}^{N}\mathrm{LRE}_{(S_2,S_1,p_i)}\right\} \tag{5.3}$$

This segmentation error, based on the GCE, is a distance whose values lie in the range $[0,1]$. A value of 0 indicates that the two segmentations are identical (perfect match) and a value of 1 indicates maximum deviation between the two segmentations being compared.

[3] In fact, two different measures can be defined, one applying on clusters or class, another applying on regions or segments. For an image, a region is a set of connected pixels belonging to the same class and a class, a set of pixels possessing similar textural characteristics.

5.2.3.2 Variation of Information (VoI)

The variation of information (VoI) metric is a recent information theory based measure for comparing the similarity of two segmentation results. This metric quantifies the information shared between two segmentations by, more precisely, measuring the amount of information that is lost or gained in changing from one segmentation to another [8].

$$VoI(S_1,S_1) = H(S_1) + H(S_2) - 2I(S_1,S_2) \tag{5.4}$$

Where H and I are respectively the entropies of and mutual information between segmentation S_1 and S_2. The VoI is a true distance on the space of clusterings which is positive, symmetric and obeys the triangle inequality. It takes a value of 0 when two clusterings are identical and positive otherwise. It is bounded by $\log N$ (N is the number of pixels), and if S_1 and S_2 have at most $R^{\max}$ clusters, it is bounded by $2\log R^{\max}$ [133].

5.2.3.3 Probabilistic Rand Index (PRI)

The Rand index [80] simply computes the proportion of pairs of pixels with compatible region (or cluster[3]) label relationships between the two segmentations to be compared. if $l_i^{S_1}$ designates the region or cluster labels associated to the segmentation maps S_1 at pixel location i and if $\mathcal{I}$ is the identity function, the Rand index is given by :

$$\begin{aligned}\mathrm{PRand}(S_1,S_2) = \frac{1}{N(N-1)}\sum_{i,j\ i\neq j}\Big\{\ &\mathcal{I}(l_i^{S_1}=l_j^{S_1} \text{ and } l_i^{S_2}=l_j^{S_2}) \ + \\ &\mathcal{I}(l_i^{S_1}\neq l_j^{S_1} \text{ and } l_i^{S_2}\neq l_j^{S_2})\ \Big\}\end{aligned} \tag{5.5}$$

A value of 1 indicates that the two segmentations are identical and a value of 0 indicates that the two segmentations do not agree on any pair of points (e.g., when all the pixels are gathered in a single region in one segmentation whereas the other segmentation assigns each pixel to an individual region). When the number of labels in S_1 and S_2 are much smaller than the number of pixels N, a computationally inexpensive estimator of the RI can be found in [80]. Since the PRI is a measure of similarity lying in $[0,1]$, a distance

based on the PRI criterion can be $[1 - \text{PRand}(S_1, S_2)]$.

5.2.3.4 The F Measure

The global F measure (or harmonic mean of precision-recall measure) [247] provides a performance score, evaluating the agreement between region boundaries between two segmentations. This latter measure is, in fact, deduced from the well-known precision/recall values. Qualitatively, the precision measure (P) is defined as the fraction of detections that are true boundaries; this measure is low when there is significant over-segmentation, or when a large number of boundary pixels have poor localization. The Recall (R) measure gives the fraction of the true boundaries detected [18]. A particular (vision) application can define a relative cost α related to these two quantities, which controls a harmony between P and R [247] (or equivalently focuses attention at a specific point on the precision-recall curve). The F measure is defined as :

$$F_\alpha(S_1, S_2) = \frac{PR}{\alpha R + (1-\alpha) P} \tag{5.6}$$

The best F measure, for a given α reflects the optimal compromise between how much true signal is required and how much false alarms can be tolerated. Since the F measure of similarity lying in $[0, 1]$, a distance based on the F criterion can be $[1 - F_\alpha(S_1, S_2)]$.

5.3 Proposed classification model

5.3.1 Data description

The data analyzed in this paper was obtained from the OASIS database[4]. The data set consists of a cross-sectional collection of 416 subjects aged 18 to 96 years, including 218 subjects aged 18 to 59 years and 198 subjects aged 60 to 96 years. For the older subjects, 98 subjects NC and 100 subjects who were diagnosed with verry mild to moderate AD. The detailed statistics of the data set was described in the literature [229]. As proposed in [4, 218, 225, 226, 233, 239–244], a subset of the OASIS database, *i.e.*, 98 right-handed women (aged $65 - 96$ years) is considered herein and was selected in order to evaluate

the detection performance of the proposed classifier. More precisely, the used subset consists of 49 subjects who have been diagnosed with very mild to mild AD (class 1) and 49 non-demented (class 2). The designation of demented and non-demented is based on the Clinical Dementia Rate (CDR). A CDR of 0 corresponds to the normal cognitive state, CDR > 0 to dement. The demographic information of these subjects is summarized in Table 5.1.

TABLE 5.1 : Demographic information of the subjects in the two classification classes.

	Very mild to mild AD	**Normal Control**
No. of subjects	49	49
Age	78.08 (66-96)	77.77 (65-94)
Education	2.63 (1-5)	2.87 (1-5)
Socioeconomic status	2.94 (1-5)	2.88 (1-5)
CDR (0.5/1/2)	31/17/1	0
MMSE	24 (15-30)	28.96 (26-30)

5.3.2 MRI data preprocessing

In this work, we used T1-weighted magnetic resonance images of the brain, that were captured with a T Vision 1.5 scanner. As described in [229], all (axial section) MRI images were : (a) corrected for inter-scan head movement and rigidly aligned to the Talairach and Tournoux space [230], (b) transformed to a template with a 12-parameter affine registration and merged into a 1-mm isotropic image [229], (c) skull stripping with the Brain Extraction Tool (BET[5]) and corrected for intensity inhomogeneity [231], and finally (d) segmented into $K = 3$ classes corresponding to the three existing cerebral tissues (CSF, WM and GM). This 3-class segmentation is obtained by applying firstly, 10 times the K-means clustering algorithm with different seeds and different number of neighbors of the pixel to be classified and secondly by combining them with a simple majority vote scheme[6].

[4] `http://www.oasis-brains.org`

[5] `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BEText`

[6] We have noticed that this simple segmentation technique provides better classification results than a SPM software (Statistical Parametric Mapping) based segmentation technique (the classification accuracy

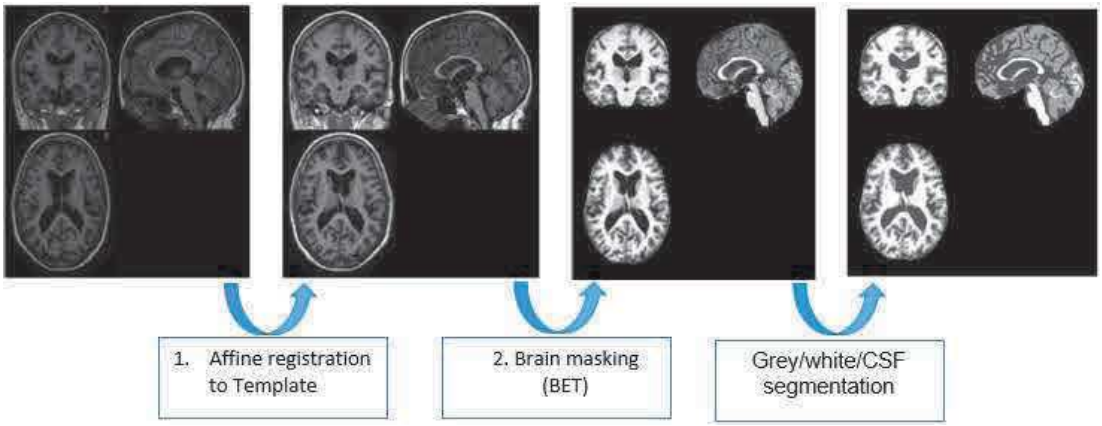


FIGURE 5.1 : Snapshot of a specific subject. (a) One original scan. (b) Atlas-registered image. (c) Brain masked version of Fig. 5.1 (b). (d) The GM/WM/CSF segmentation image.

In our application, the majority vote is achieved with a spatial window (of size 3×3 pixels and centered on the pixel s to be classified) that collects the class labels of the 10 segmentation results obtained by each K-mean clustering and by finally assigning to that central pixel s, the class label that has the majority vote. This strategy ensures both an efficient spatial regularization of the final segmentation result and also a reliable decision fusion between results obtained by these K-mean clusterings. In this segmentation, the *CSF*, the *white matter* and the *grey matter* are represented by a dark, a grey, and a white region respectively, in order to visually express the activity level of the blood flow. The preprocessing steps are summarized in Fig. 5.1 for the T1 image of one subject.

5.3.3 Prototypes NC and AD

Our classification model is based on a two-step classification procedure whose the first step relies on a minimum distance based classifier using two consensus segmentation based prototypes of the brain, in terms of CSF, WM and GM regions and built from the entire OASIS basis (cf. Sec 5.2.2 & 5.2). More precisely, one brain prototype corresponds to unhealthy individuals (with AD or class 1) and is obtained by combining

is better by 2%).

all segmentations related to subjects who have been diagnosed with very mild to mild AD (from 100 subjects) and the second one (class 2), corresponding to healthy subjects, is built from the set of non-demented subjects contained in the OASIS database (316 healthy subjects).

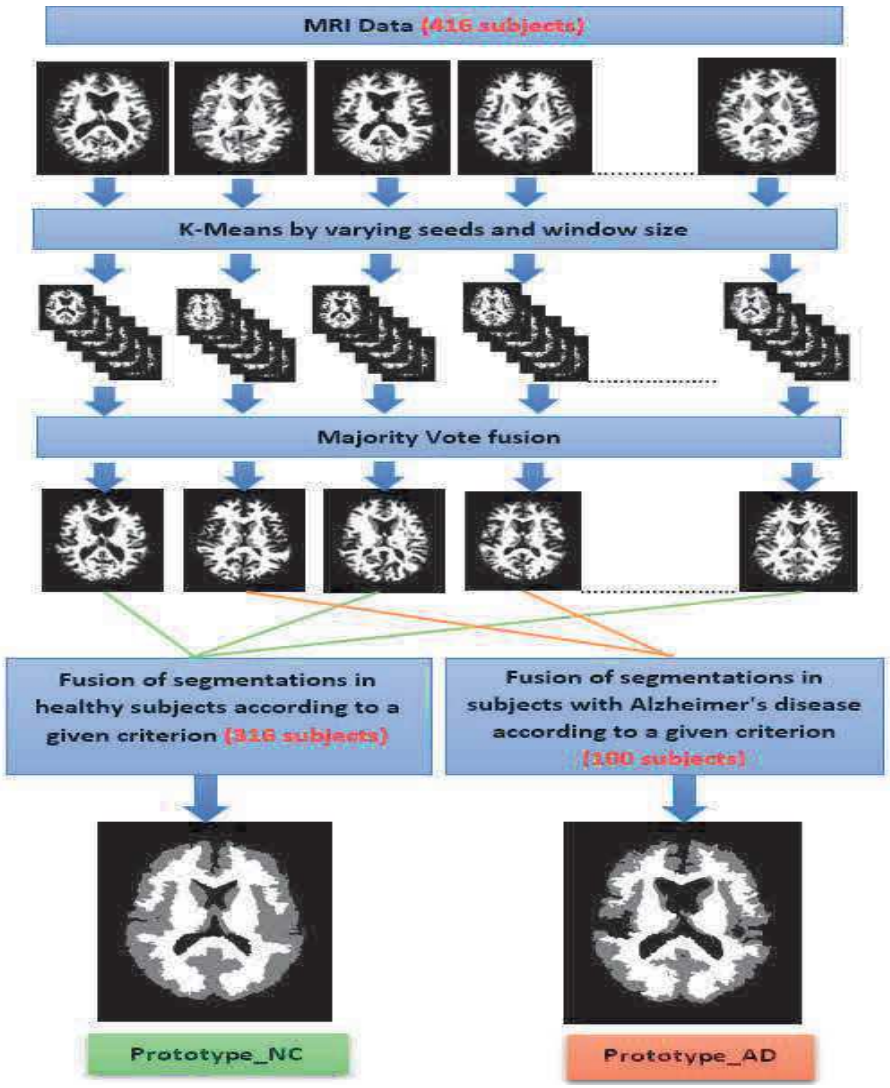


FIGURE 5.2 : Methodology for the NC and AD prototype creation.

As already said (cf. Sec 5.2.1), these two consensus segmentation based prototypes can be achieved according to different criteria (each criterion being, in fact, conceptually

based on the definition of a specific distance between two segmentation maps). In this work, we will test and compare the efficiency of different consensus criteria for the two prototypes used in our classification scheme. The methodology for the creation of the two consensus segmentation based prototypes is illustrated in Fig. 5.2.

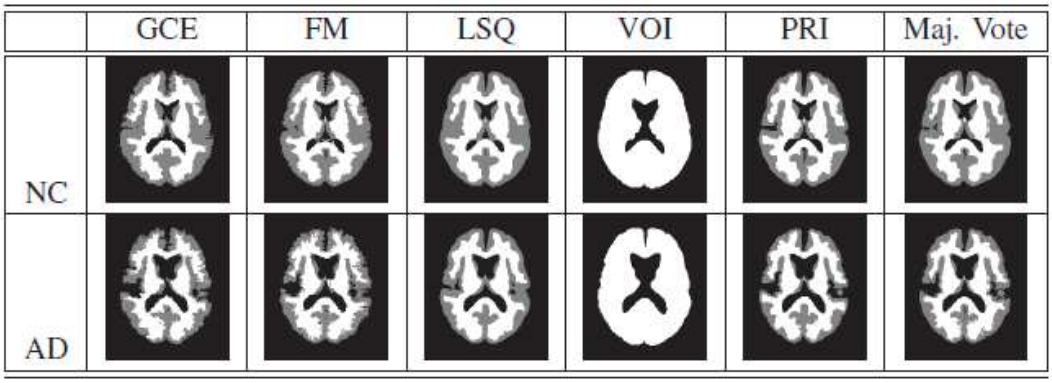


FIGURE 5.3 : Prototype AD and NC according to different criteria of merging (GCE [17], FM [18], LSQ [6], VOI [5], PRI [19]).

5.3.4 Two-step classification

The proposed hybrid classification technique named MDC-KNN (Minimum Consensus Distance-KNN), combining the previously estimated prototypes (Prototype_NC and Prototype_AD) and the KNN (weighted KNN as proposed in [232]) algorithm consists of three stages :

- Calculate the two Consensus distances between the segmentation related to an input MRI image and respectively the Prototype Prototype_NC (D_NC) and the Prototype Prototype_AD (D_AD).

- Choose the classifier KNN or (prototype-based) Minimal distance (MD). The choice of classifier (KNN or MD) depends on the difference between D_NC and D_AD. If the difference is large (greater than a threshold T which was set to 1.5%[7]), then, in this case (case 1 or 2 of Fig. 5.4), we choose the MD classifier since in this case we are certain that the segmentation is very close to one of the two prototypes and consequently we can thus rely on this classification procedure. Otherwise, we

apply the weighted KNN classifier (case 3 of Fig. 5.4) in with the GCE distance (cf. Sec 5.2.3.1) between the segmentation related to an input MRI image and the 3-class segmentations obtained for each image of the test set.

- Classification.

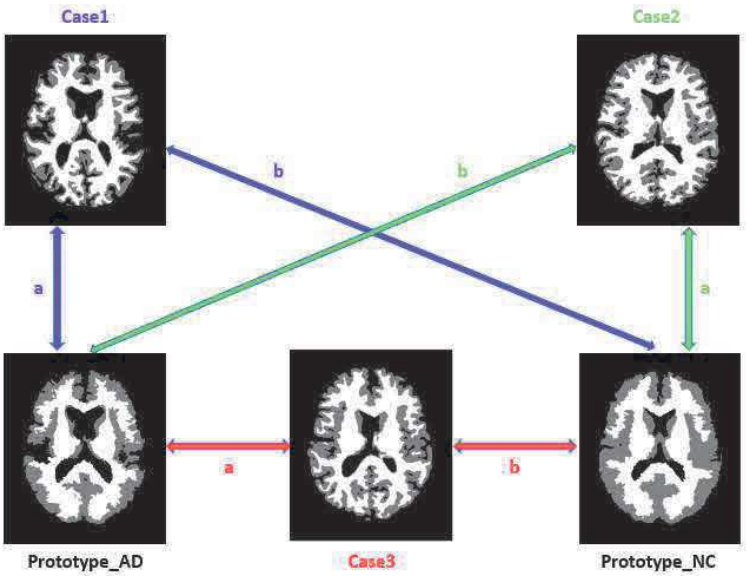


FIGURE 5.4 : $|D_{NC} - D_{AD}|$ large (case 1 or 2) or low (case 3).

Let us note that our two-step classification procedure which combines, in the first step, the MD classifier based on two consensus segmentation based prototypes followed by the second step, which uses a classical weighted KNN classifier, can be also viewed as a single hierarchical KNN classifier. Indeed the MD classifier based on our two prototypes can be seen as a simple 1-nearest neighbour in which the prototype summarizes the set of segmented brains belonging to a same pathology class (healthy or AD). The schematic diagram for the proposed two-step classification methodology is outlined in Fig. 5.5

We have computed than the distance (in percentage of pixel difference) between Prototype_NC & Prototype_AD is 15%. Based on this, we have set this threshold T to an order of magnitude ($\div 10$) lower (*i.e.*, $T = 1.5\%$).

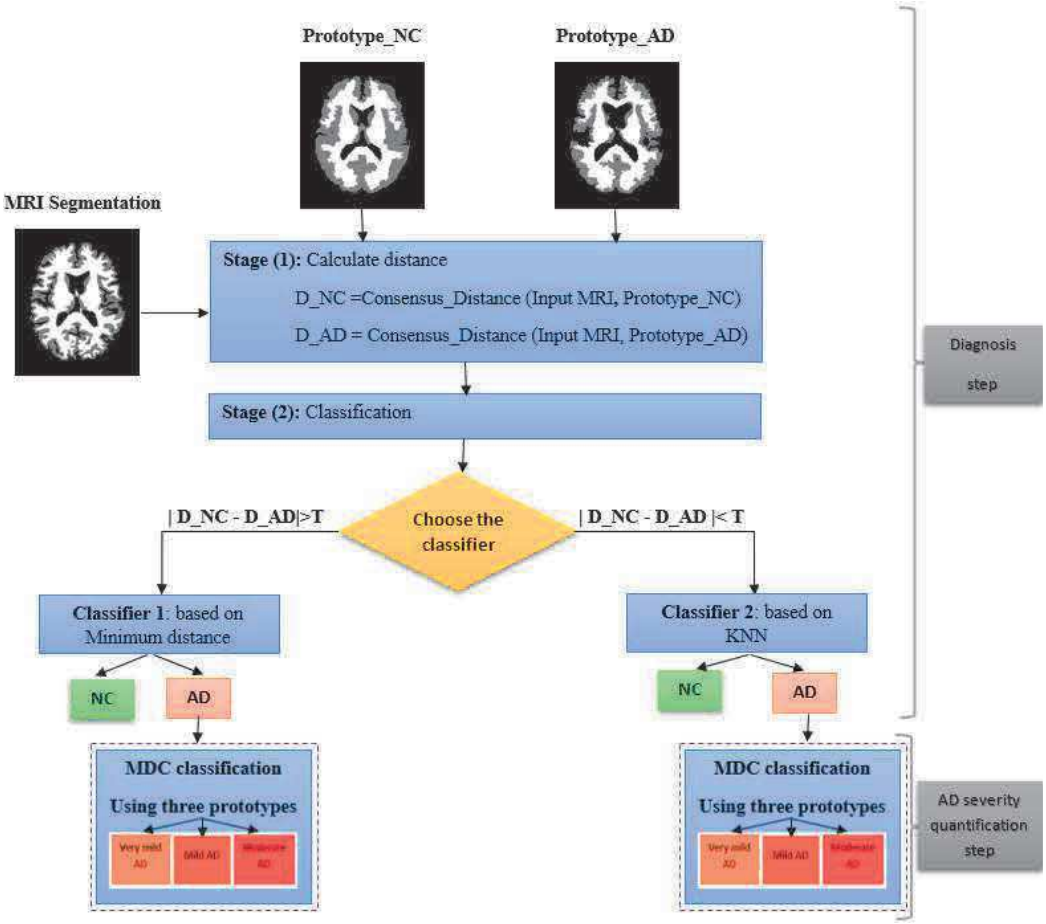


FIGURE 5.5 : The methodology of the proposed two-step classification method.

5.4 Experiments and results

In this section we describe our experiments and report the results of the proposed method. In order to investigate the detection performance of the proposed MDC-KNN classifier, a set of appropriate experiments were conducted. For the experimental purposes, specific software was developed in C++. All experiments were executed in an Intel i7 3.3 GHz PC with 16 GB RAM.

5.4.1 Evaluation criteria

We evaluate the performance of the proposed classification method, against recent leading classification methods, in terms of sensitivity=TP/(TP + FN), specificity=TN/(TN + FP) and accuracy=(TN + TP)/(TN+TP+FN+FP) where True Positives (TP) are AD patients correctly identified as AD, True Negatives (TN) are controls correctly classified as controls, False Negatives (FN) are AD patients incorrectly identified as controls and False Positives (FP) are controls incorrectly identified as AD. Sensitivity is the proportion of AD subjects correctly classified, and the specificity is the proportion of correctly classified controls [248, 249] (cf. Tab 5.2).

In addition, we also compare the performance of our classification method with different consensus criteria, for the creation of the two prototypes (cf. Sec 5.3.3 and Figures 5.2, 5.3 and 5.5), namely ; GCE, FM, Least Square (LSQ), VoI, PRI and Majority vote (cf. Tab 5.3 and 5.4).

5.4.2 Performance measures & comparison with state-of-the-art methods

5.4.2.1 Classification results into two classes (NC *vs.* AD)

The actual feature dataset (49 AD and 49 NC) has been used in several works in the literature, hence results obtained with a variety of classifier models are publicly available for comparison, for example, [4] applied a KNN classifier defined in the LC (Lattice Computing) context, on the 98 female subjects. He obtained 80% accuracy, 79% specificity and 80% sensitivity, also, authors in [225] studied the feature-extraction process with VBM analysis and achieved the best results with 85% accuracy for the rbf-DAB-SVM classifier (cf. Tab 5.2). The AD detection statistics, accuracy, sensitivity and specificity, of the proposed MDC-KNN classifier by performing a leave-one-out cross validation[8] test, in comparison with the state-of-the-art models [4, 225, 233, 242, 243] are summarized in Table 5.2 regarding an OASIS subset of 98 subjects (cf. Sec 5.3.1).

[8] This procedure iteratively leaves out the information on each subject and trains the model on the remaining subjects for subsequent class assignation of the person that was not included in the training procedure.

TABLE 5.2 : Classification results comparison between our method and the morphometric approaches proposed in [4] and several other approaches (using a leave-one-out cross validation test and the same dataset).

Classifier type	Accuracy	Sensitivity	Specificity
MDC-KNN (GCE meas.)	**0.90**	**0.90**	**0.91**
rbf-DAB-SVM	0.85	0.78	0.92
PCA+Multi-Kernel SVM	0.84	0.85	0.86
PCA+Linear SVM	0.83	0.83	0.83
LVQ2	0.83	0.74	0.92
LVQ1	0.81	0.72	0.90
MDC (GCE meas.)	**0.80**	**0.81**	**0.79**
LC-KNN	0.80	0.80	0.79
rbf-AB-SVM	0.79	0.78	0.80
PCA+SRAN	0.79	0.79	0.79
MLP-BP	0.78	0.69	0.88
PNN	0.78	0.62	0.94
Linear SVM	0.78	0.72	0.88
Indep-rbf-SVM	0.75	0.56	0.95
Kernel-LICA-DC	0.74	0.96	0.52
Indep-Linear-SVM	0.74	0.51	0.97
KNN (GCE meas.)	**0.73**	**0.71**	**0.75**
Linear-AB-SVM	0.71	0.54	0.88
RBF	0.66	0.65	0.68

By examining the results of Table 5.2 & Fig. 5.6, it follows that the proposed classifier MDC-KNN with GCE measure (for the MD classifier) and $K = 3$ (this number was empirically tested) demonstrated superior performance than some conventional classifiers such as RBF, MLP-BP, PNN, Linear SVM (12% higher accuracy) and some advanced classification models (Indep-linear-SVM, Indep-rbf-SVM, linear-AB-SVM, rbf-AB-SVM, Kernel-LICA-DC, LVQ1, LVQ2, rbf-DAB-SVM, PCA+Linear SVM, PCA+SRAN, PCA+Multi-Kernel SVM). Let us note that the single KNN classifier (without being combined with our prototype based MD classifier) with the GCE distance reaches a classification value of only 73%. Conversely, the single MD classifier (without being combined with our KNN classifier) with the GCE distance reaches a classification value

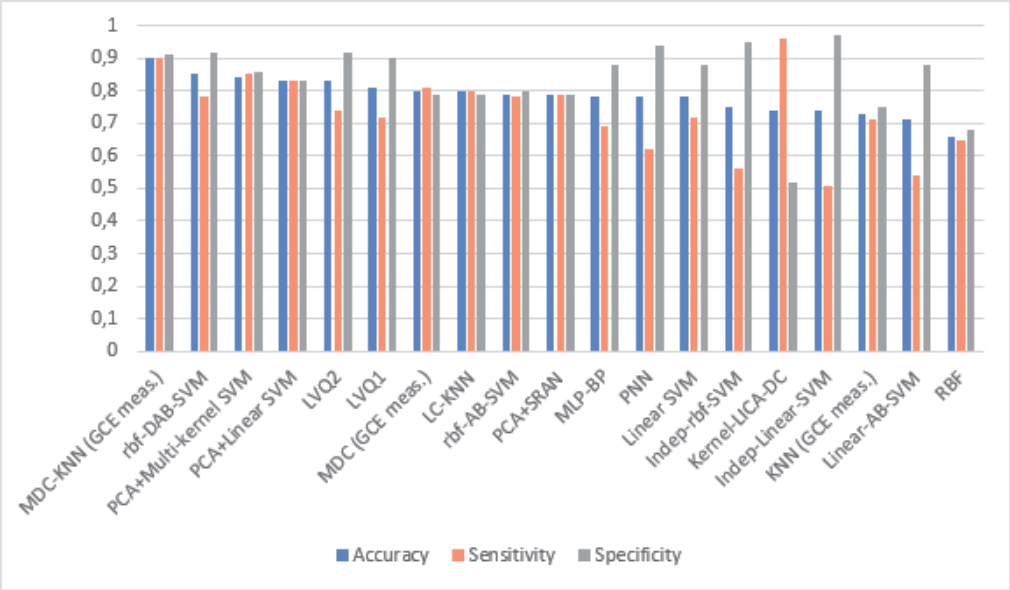


FIGURE 5.6 : Performance comparison between our method and the morphometric approaches proposed in [4] for AD *vs.* NC.

of only 80% (cf. Tab 5.2). In addition, we have processed a total of 416 MR images from 100 AD and 316 control subjects and noticed that our classification results is stable over the entire database and thus reliable (cf. Tab 5.4).

Let us note that our method (MDC-KNN) achieves the best result with the GCE measure or criterion and the second best result with the Majority_Vote criterion (cf. Tab 5.3 & 5.4). This indicates that the GCE is an interesting distance measure between segmentations with which useful consensus-based segmentations can be efficiently exploited in a classification framework.

5.4.2.2 Classification results into four classes (NC, very mild AD, mild AD and moderate AD)

Previous classification aims to distinguish between two separated classes (AD *vs.* NC). Nevertheless, when we include the subjects with CDR = 0.5, CDR=1 and CDR=2, we have 4 classes of the disease (NC, very mild AD, mild AD and moderate AD), and

TABLE 5.3 : Comparison of the classification results of our approach with different consensus criteria for the creation of prototypes (98 subjects).

Criterion	**Accuracy**	**Sensitivity**	**Specificity**
GCE	0.90	0.90	0.91
FM	0.88	0.81	0.95
Least Square	0.88	0.86	0.90
VoI	0.79	0.61	0.97
PRI	0.85	0.85	0.86
Majority Vote	0.89	0.88	0.90

TABLE 5.4 : Comparison of the classification results of our approach with different consensus criteria for the creation of prototypes (for the entire database : 416 subjects).

Criterion	**Accuracy**	**Sensitivity**	**Specificity**
GCE	0.90	0.88	0.91
FM	0.89	0.88	0.90
Least Square	0.88	0.86	0.89
VOI	0.81	0.62	0.88
PRI	0.86	0.85	0.87
Majority Vote	0.90	0.87	0.91

it is interesting to distinguish between these different classes. To this end, we can divide the task into two stages :

- Step I : in this step, all the subjects are taken and classified as NC or AD as described in Fig. 5.5, the AD class includes patients with very mild to moderate AD, this first step can be viewed as a diagnosis step.

- Step II : in this step, we use the MDC classifier based on three new prototypes, *i.e.*, respectively built from brain segmented images having (in the training base) a CDR=0.5, CDR=1 and CDR=2 (Fig. 5.7) to classify them into the very mild AD, mild AD and moderate AD classes, this second step allows us to quantify the severity of the disease.

The classification performance with the second experimental group achieved 86% with the GCE measure for the entire database (416 subjects) and 87% for the subset of database (98 subjects).

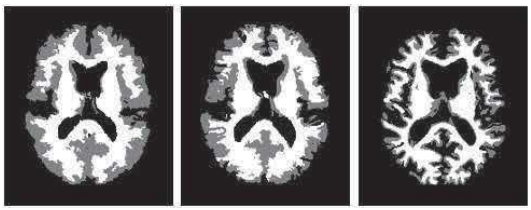


FIGURE 5.7 : From the lexicographic older : prototypes using the GCE distance and built from (a) : CDR=0.5 (very mild AD), (b) CDR=1 (mild AD), (c) CDR=2 (moderate AD).

5.4.3 Visualization of image databases

In order to appreciate all the relevancy of the GCE distance, used in our classification procedure, it may be interesting to find a strategy to provide a quick overview of how are distributed the 49 healthy and 49 unhealthy brains of our subset of the OASIS database with our two prototypes according to the GCE distance. This can be done with the Multi-Dimensional Scaling (MDS) and the technique described in [8] with, in our application the GCE distance between pair of segmentations (instead of the VoI distance as applied in [8]). To this end, it consists in computing the distance matrix describing the dissimilarities between each existing pair of segmentations, in term of GCE distance, and used this distance matrix with a technique which will attempt to find an embedding, in a (for example) 2D space, such that pairwise distances between these segmentations are preserved as much as possible. This embedding method, aiming at preserving the original relationships of these images (in term of a given distance), is done with the MDS visualization technique in the least square sense. It is now used in navigation systems, browsing or image database in which it may be interesting to arrange the images of the database according to their descriptive content extracted by a segmentation process, or otherwise said, based on the spatial arrangement of the different objects detected or segmented in the image (instead of simply their own color or texture) [8].

This 2D MDS visualization map of the MRIs based on the GCE distance is shown

in Fig. 5.8. The MDS estimates a (possible) 2D mapping with 17% of error (or loss of information if we consider the MDS technique as a dimensionality reduction scheme). In spite of this error, we can see two clusters ; the blue brains are related to the healthy brains and are in a cluster located at the bottom and left of the image and conversely the red brains, corresponding to unhealthy brains are in a cluster located at the top and right of the image.

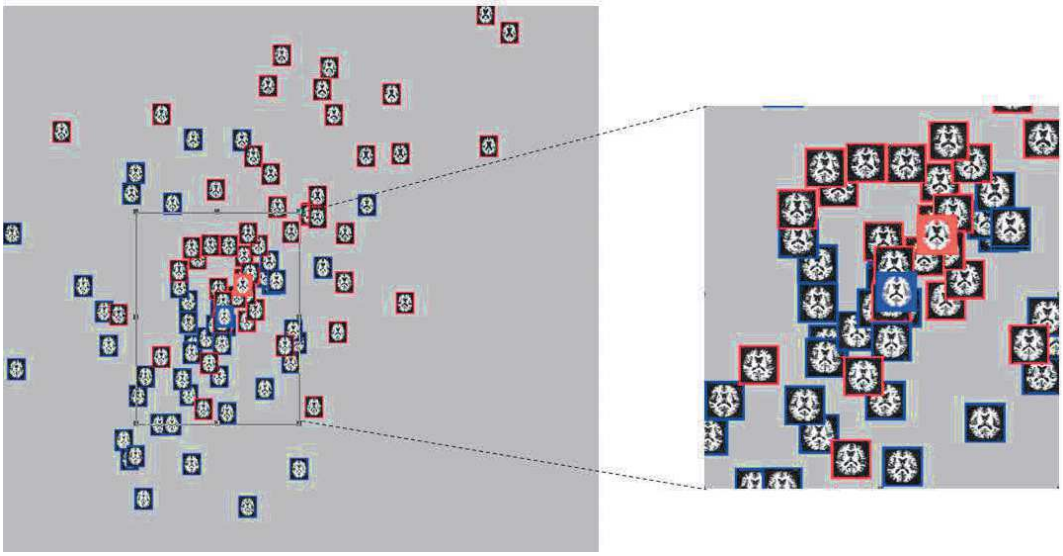


FIGURE 5.8 : A MDS visualization map of the considered subset of the OASIS database (49 healthy and 49 unhealthy subjects) based on the GCE distance between (pair of) segmentations according to their similarity. MRIs framed in blue are NC and MRIs framed in red are AD. the large red and blue frame is respectively the AD and NC prototypes.

5.5 Conclusion

In this paper, we have presented a novel and reliable prototype-based classification framework, in structural MRI, for the early detection and classification of the Alzheimer's disease. The proposed framework relies on two prototypes, based on the recent concept of consensus segmentation, to define two average models of segmented brain, corresponding to healthy subjects or unhealthy individuals affected by dementia. These

two consensus segmentation-based prototypes have several appealing characteristics for our classification task. First, since these two consensus models are built from segmentation maps, it allows us to reduce the information content of a brain MR image and to suppress undesired components such as noise which may degrade the classification performance. Second, as a consensus model, it also allows us to suppress undesired components in the brain image such as the anatomical variability existing between individuals (of the same group) which are not relevant for the detection and quantification of AD. In this study, different consensus criteria of segmentations have been tested and it turns out that the Global Consistency Error seems the most relevant criterion for modeling a healthy or unhealthy brains consensus which could be subsequently used for the early detection of AD in structural MRI. This can be due to the inherent nature of this criterion which can take into account the inherent multiscale nature of an image segmentation by measuring the level of refinement existing between two segmentations. Experiments, conducted on 98 subjects show the validity of the proposed method and especially its simplicity and high accuracy compared to the other state-of-the-art AD diagnosis approaches proposed in the literature. In addition,

- Our approach is less time-consuming compared to the state of the arts Volumetric methods, computer-based and does not require the intervention of an expert during the classification/retrieval phase.
- It is extensible to other diseases that can be diagnosed by brain MRI such as Schizophrenia.
- The method could be extended by combining axial, coronal, and sagittal MRI data for improving the classification accuracy.

CHAPITRE 6

CONCLUSION

Dans cette thèse, nous avons montré qu'il était intéressant et utile d'exploiter et d'adapter le principe de distance entre segmentations et de moyennage consensuel de segmentations, au sens d'un certain critère, dans quelques algorithmes classiques utilisés en visualisation, traitement d'images ou en classification dans le cas où les données à analyser, traiter ou utiliser sont présentées à un niveau d'abstraction plus élevé comme des segmentations d'images.

Conceptuellement, la plupart des critères consensuels utilisés, pour fusionner ou faire la moyenne d'un ensemble de segmentations, sont basés sur la définition d'une distance ou d'une métrique entre deux cartes de segmentation (mesurant aisi la similarité entre les deux regroupements spatiaux). C'est directement le cas pour le critère VoI, GCE, PRI, F-mesure et à partir de celui-ci, on définit la segmentation de consensus comme étant la solution de segmentation qui minimise la somme des distances la séparant de toutes les autres cartes de segmentation de l'ensemble de segmentation.

Pour le critère de vote majoritaire, qui ne peut s'appliquer que lorsque chaque classe ou cluster ou région peut être associée, sans ambiguité, à une seule et unique étiquette sémantique, cette distance est simplement la proportion moyenne d'étiquettes (sémantiques) différentes par rapport au vote majoritaire final. Néanmoins, pour d'autres critères, cette distance existe mais n'est pas explicitement paramétriquement exprimable [6, 86–88].

Cela nous a permis de présenter la faisabilité de différentes applications originales dans des domaines aussi variés que la visualisation ou la recherche d'informations dans de grandes bases d'images, ou encore la détection ou finalement la classification permettant d'exploiter ces données de segmentations, déjà exprimées à un plus haut niveau d'abstraction que de simples quantités scalaires ou de grandeurs vectorielles (comme les niveaux de gris ou de couleurs d'une image), à travers de plusieurs prototypes de seg-

mentation pour chaque *cluster* ou regroupement de segmentations ou d'un résultat de consensus pour améliorer une carte de détection de mouvement.

Dans tous les cas, la condition indispensable est de compter sur un ensemble de cartes de segmentations qui soient complémentaires, diverses et, pour quelques applications, représentatives (par exemple pour représenter toutes les deformations pathologiques d'une maladie ou d'un degré de sévérité d'une maladie).

6.1 Sommaire des contributions

Dans le cadre de cette thèse, nous avons étudié la faisabilité du moyennage consensuel de segmentations dans les applications suivantes :

- La présentation des données où ce principe est exploité plus précisément dans le cadre de la visualisation, l'exploration, la navigation et la recherche dans des grandes bases d'images à un niveau d'abstraction plus élevé de ce qui se fait habituellement (basé sur les couleurs/textures de chaque image), *i.e.*, en fonction de la façon dont sont disposés et agencés les différents objets présents dans ces images. Cette méthode peut aussi être très complémentaire avec les techniques de visualisation/ exploration/ navigation/ recherche utilisant les indices locaux de bas niveau du type couleur/texture.

- Le traitement d'images et plus spécifiquement, la détection du mouvement dans une vidéo, où ce principe est exploité pour fusionner différentes cartes de segmentations issues du même algorithme de détection de mouvement, fonctionnant dans des espaces de couleurs différentes complémentaires préalablement estimés. L'idée sous-jacente derrière ce modèle est de prendre avantage de la complémentarité (au bruit, changements d'éclairage, disposition des couleurs, gestion des ombres, reflets, etc.) de la détection obtenue dans chaque espace de couleur par le modèle considéré pour finalement obtenir une segmentation finale, au sens du mouvement, débruitée, plus fiable et plus précise. Les tests que nous avons réalisés montrent des gains d'améliorations significatives pour une méthode proposée

qui reste d'une part facilement implémentable mais aussi, du fait de la structure parallèle des processeurs multi-core actuels, sans accroissement notable du temps de calcul.

- La classification en imagerie médicale et plus précisément, la classification des images cérébrales IRM du cerveau par le contenu pour l'aide au diagnostic de la maladie d'Alzheimer. Dans notre application, on utilise ce principe pour construire, à partir d'une base d'apprentissage, tout d'abord deux prototypes (sain et malsain) de segmentation du cerveau (en termes de la substance grise, substance blanche et fluide cérébro-spinal) puis ensuite plusieurs prototypes de segmentation selon la sévérité de la maladie. Ces prototypes sont utilisés comme référence dans un mélange de classifieurs pour augmenter leur efficacité. À notre connaissance, le principe de consensus de segmentation n'a jamais été utilisé dans le domaine de la classification en imagerie médicale ou autre. La méthode proposée est automatique (ne nécessite pas l'intervention d'un clinicien) et les résultats obtenus apportent une amélioration notable par rapport aux méthodes volumétriques et morphométriques avancées et proposées dans la littérature, en termes de précision de classification et de temps de traitement. De même, soulignons, que notre méthode, par sa nature très différente des méthodes volumétriques ou morphométriques pourrait s'avérer particulièrement intéressante à combiner avec les méthodes classiques proposées dans la littérature.

6.2 Perspectives générales

En fait, on le rappelle, le principe de distance entre segmentations (au sens d'un certain critère) et de moyennage de segmentations peut être exploité, directement ou facilement adapté, plus généralement, par tous les algorithmes ou les méthodes (utilisées en imagerie numérique ou plus généralement en analyse des données ou apprentissage machine) où les données peuvent en fait se substituer à des images segmentées. De ce fait, les applications sont indénombrables. Le chapitre suivant cite en détail quelques applications originales qui peuvent tirer profit de ce principe.

CHAPITRE 7

ORIENTATIONS FUTURES DE LA RECHERCHE

Dans ce chapitre, nous citons plusieurs applications originales, qui peuvent tirer profit du principe de moyennage consensuel de segmentations et qui pourraient être une extension possible à nos travaux de recherche.

7.1 Reconstruction 3D d'objets à partir d'images multivues

Une des applications possibles est lié au problème de la reconstruction 3D dense d'objets à partir d'images multivues et plus précisément à partir de photographies prises selon des points de vue différents et par différentes caméras non calibrées (*i.e.*, dont on ne connaît pas les caractéristiques techniques : focale, position, orientation, etc.) et possiblement de résolutions différentes.

Récemment, de nombreuses techniques de reconstruction dense 3D ont été proposées dans la littérature avec des systèmes stéréos multivues (MVS) capables de reconstruire des monuments entiers [20, 250]. Les résultats obtenus par ces techniques MVS donnent de très bons résultats de reconstruction en présence de conditions idéales ; *i.e.,* lorsque la reconstruction se fait à partir de nombreuses images, en haute résolution, sous des points de vue très différents et variés et surtout lorsque l'objet à reconstruire est très texturé et sans spécularité ou de réflexion et de transparence apparaissant sur les différentes surfaces de l'objet à reconstruire (hypothèse que la scène ne contient que des objets lambertiens). Sous ces conditions idéales, les résultats de reconstruction sont très bons mais lorsque celles-ci ne sont pas réunies, ces techniques permettent seulement d'obtenir des reconstructions très bruitées avec de nombreux artéfacts ou de trous (cf. Fig 7.1).

En fait, les techniques de reconstruction stéréo multivue (MVS) peuvent se regrouper en trois catégories ; respectivement, les méthodes basées sur l'estimation et la combinaison de cartes de profondeur (*depth-map based*) [251], celles basées sur des méthodes volumétriques (*volumetric methods*) [252] qui cherchent un partitionnement 3D et plus

précisément à classer chaque voxel dans la classe *"intérieur"* ou *"extérieur"* de l'objet et enfin celles utilisant le principe de croissance d'imagettes (*patch-growing*) [20] combinant des étapes successives de filtrage et d'expansion pour maximiser l'idée de consistance photométrique entre les images. Ces trois classes de méthodes possèdent les mêmes limites lorsque les conditions idéales précédemment énumérées (présence dans la scène d'objets lambertiens) ne sont pas remplies.

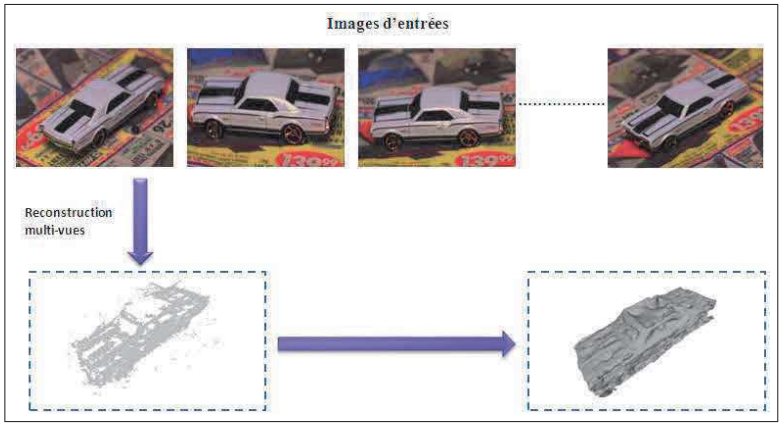


FIGURE 7.1 : Exemple de reconstruction 3D dense d'objet à partir d'images prises selon des points de vue différents, par un modèle traditionnel de reconstruction stéréo multivue (MVS) [20] qui se heurte ici surtout au manque de texture de l'objet à reconstruire et aussi un peu aux problèmes de spécularités, de réflexions et de transparence. On peut remarquer que la reconstruction obtenue est très bruitée et présente à quelques endroits des trous et de nombreux artéfacts.

Face à ces problèmes nous pourrions proposer une méthode de reconstruction 3D qui combinerait d'une part -1] le principe de la méthode MVS ou une autre semblable, -2] une technique de sous échantillonnage des points caractéristiques invariants[1] (préalablement détectés dans ces images multivues et utilisée dans ces méthodes de reconstruction décrites précédemment, cf. Fig 7.1) avec une reconstruction 3D possible pour

[1] Ces méthodes MVS utilisent dans chaque image une détection de points caractéristiques (avec leur orientation) correspondant aux points particuliers de l'objet à reconstruire (coins, points d'inflexion, de courbure localement maximale ou encore d'intersection de plusieurs contours comme les jonctions). Une mise en correspondance de ces points d'intérê t est ensuite utilisée pour estimer les paramètres de la relation géométrique (homographie) qui peut lier chaque paire d'images ce qui permet d'obtenir au final, après filtrage et élimination de points sans correspondance, un nuage dense de points caractéristiques dans un espace 3D qui est ensuite utilisé pour la construction d'un maillage géométrique de l'objet à reconstruire.

chaque sous-ensemble de points et enfin -3] le principe de moyennage consensuel de segmentation au sens d'un certain critère, mais ici généralisé en 3D ; une reconstruction 3D voxélique étant vue ici comme une segmentation 3D en deux classes. Ce principe permettrait d'obtenir une reconstruction 3D (ou segmentation 3D volumique) débruitée dans un premier temps mais aussi possiblement d'inclure dans ce modèle de moyennage (ou de fusion) basé sur l'optimisation d'une fonction d'énergie, un terme *a priori* sur la répartition (*a priori* désirée) des points du volume à reconstruire pour obtenir une reconstruction 3D régularisée et encore plus précise.

L'approche proposée consisterait à considérer plusieurs sous-échantillonnages du nuage de points donné par la méthode MVS et utiliser la série de reconstructions volumiques 3Ds (ou segmentations 3Ds) grossières et bruitées obtenue pour chacun de ses sous-échantillonnages dans une procédure de fusion ou de moyennage consensuel au sens d'un critère spécifique pour obtenir, on l'espère, une reconstruction finale débruitée (cf. Fig 7.2). Dans un premier temps, cette stratégie permettrait, on l'espère, d'atténuer sensiblement le bruit ou les artefacts que l'on obtiendrait avec la méthode MVS brute dans le cas d'images présentant un objet à reconstruire qui présenterait un manque de texture ou des problèmes de spécularités, de réflexions et de transparence.

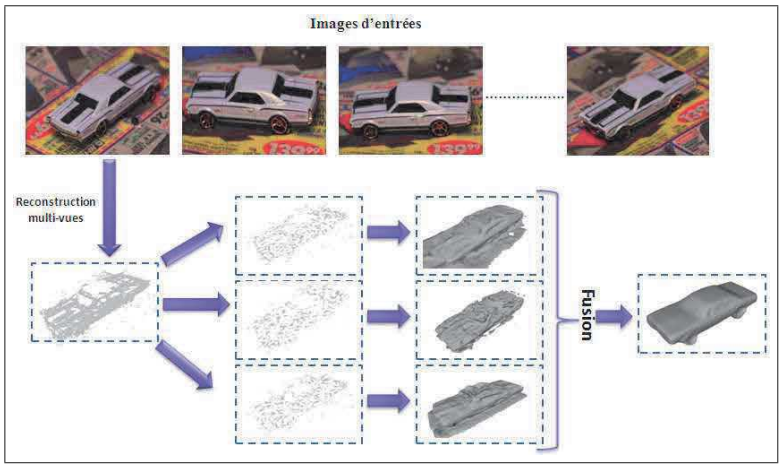


FIGURE 7.2 : Principe de la méthode de reconstruction 3D proposée. Des sous-échantillonnages du nuage de points donné par la méthode MVS permettent d'obtenir une série de reconstructions 3Ds bruitées qui seront ensuite fusionnées au sens d'un critère par un modèle de moyennage consensuel régularisé.

De plus, puisque le modèle de fusion ou de moyennage consensuel s'exprime comme étant la solution d'un problème d'optimisation (d'une fonction d'énergie), on peut inclure dans celui-ci un terme *a priori* qui pénaliserait une répartition trop bruitée des (coordonnés des) points (autour de la surface de l'objet) dans notre problème d'estimation de la reconstruction fusionnée finale afin d'obtenir une reconstruction 3D régularisée et plus précise. À ce titre, le modèle *a priori* GGMRF de Bouman [253], initialement utilisé pour la régularisation d'images (en restauration) et généralisé ici en 3D serait idéale. Avec cette stratégie, et selon le critère VoI, notre modèle de fusion régularisé serait exprimé comme étant le minimum de la fonction d'énergie suivante :

$$\begin{aligned} \hat{S}_{\overline{\text{VoI}}} &= \arg\min_{S\in\mathcal{S}_n}\left\{\overline{\text{VoI}}\,(S,\{S_k\}_{k\leq L}) + \beta\cdot\underbrace{\sum_{s\in S_{\text{ext}}}\sum_{<s,t>}|x_s-x_t|^q}_{\Omega(S)}\right\} \\ &\text{avec :}\quad \overline{\text{VoI}}\,(S,\{S_k\}_{k\leq L}) = \frac{1}{L}\sum_{k=1}^{L}\text{VoI}\,(S,S_k) \end{aligned} \tag{7.1}$$

Dans cette notation, $\hat{S}$, $\mathcal{S}_n$ et $\{S_k\}_{k\leq L}$ désignent respectivement la segmentation (ou reconstruction volumique) 3D fusionnée que l'on cherche à estimer, l'ensemble des segmentations 3Ds binaires (en deux classes) possibles et l'ensemble des L reconstructions volumiques 3Ds (grossières et bruitées) obtenues pour chacun des sous-échantillonnages. Pour le deuxième terme ($\Omega(S)$) (terme de régularisation) $<s,t>$ désigne l'ensemble des plus proches voisins d'un voxel s situé sur la surface extérieure de l'objet avec x_s, la coordonnée (x,y,z) du voxel s. q (avec $1\leq q\leq 2$) et un paramètre contrôlant si le lissage des positions des voxels extérieurs est fait selon la norme L_1 et assez similairement au terme de régularisation L_1 (variation totale) proposé par Rudin et al. proposé en [254]. Finalement, β est le paramètre contrôlant le degré de lissage de la reconstruction volumique $\hat{S}$ obtenue après fusion.

Il reste à trouver un bon optimiseur qui estimera correctement la reconstruction volumique $\hat{S}$, exprimée comme étant le minimum de la fonction d'énergie donnée par l'équation (7.1).

7.2 Débruitage des segmentations en mouvement par moyennes non-locales

Une autre application du moyennage consensuel de segmentation, dans le cadre de la détection de mouvement, serait de combiner ce principe avec l'algorithme de débruitage de Buades [21]. Cela permettrait de réaliser un post-traitement qui pourrait améliorer sensiblement les résultats d'un algorithme de détection de mouvement. Plus précisément, on pourrait considérer la séquence d'images segmentées (donnée par un algorithme de détection de mouvement donné) comme des données bruitées et debruiter ces images segmentées en généralisant l'algorithme de Buades pour le débruitage d'images segmentées, au sens du mouvement, *i.e.*, avec une étiquette binaire prenant deux valeurs possibles (mobile/immobile). Le bruit serait ici représenter par l'ensemble des mauvaises classifications. Les algorithmes classiques de débruitage sur des images en niveaux de gris ou couleurs, qui ont été proposés dans la littérature, ne peuvent être ici directement utilisés dans ces images segmentées en deux classes, mise à part, bien sur, le classique filtre spatial médian qui constitue généralement le post-traitement le plus utilisé pour débruiter rapidement (et plus ou moins efficacement) une carte de segmentation binaire de détection de mouvement.

Le modèle de débruitage introduit par Buades *et al.* [21] est l'une des plus efficaces des méthodes de débruitage existantes (sur des images naturelles en niveaux de gris ou couleurs). Il est simple et séduisant à la fois, car il repose sur la propriété que, dans les images naturelles, une petite imagette (ou *patch*) peut se retrouver presque à l'identique à plusieurs endroits de l'image (mais plus généralement dans un voisinage spatial local). Le cadre mathématique naturellement adapté à cette méthode, mais aussi à toutes ses variantes qui exploitent la présence inhérente d'autosimilarités spatiales dans une image naturelle, est donc celui de l'estimation statistique : à partir de plusieurs versions dégradées d'un même *patch*, on cherche à estimer le *patch* non bruité sous-jacent (cf. Fig 7.3).

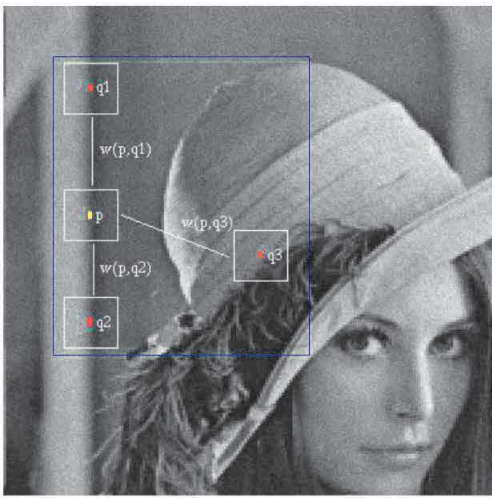


FIGURE 7.3 : Illustration du principe des méthodes de débruitage par (*patchs*) comme celui du filtrage par moyennes non locales proposé par Buades *et al.* [21]. Une imagette (ou *patch*) peut se retrouver presque à l'identique à plusieurs endroits de l'image. Dans cet exemple, dans le voisinage spatial (fenêtre carrée bleue), deux voisinages carrés (ou *patchs*), associés au pixel $q1$ et $q2$, sont très semblables au voisinage du pixel p (contrairement au voisinage du pixel $q3$).

Dans notre cas (débruitage d'une séquence d'images segmentées binaires), cette propriété de redondance ou de présence inhérente d'autosimilarités existent ; faiblement pour un voisinage uniquement spatial mais assez fortement pour un voisinage spatio-temporel. Cette propriété vient du fait que dans une séquence d'images (acquises par une caméra fixe) les objets en mouvement ne se déforment que relativement peu d'une image à l'autre (cf. Fig 7.4). De ce fait les images segmentées présentent aussi, localement dans le temps, des autosimilarités spatiales, c'est-à -dire des *patchs* binaires auto-similaires qui peuvent être donc recherchées dans une fenê tre de voisinage spatial et temporel et que l'on peut utiliser dans une méthode de débruitage par *patchs* comme celle proposée par Buades *et al.* Ainsi, de la même façon que l'algorithme de Buades, mais appliqué cette fois sur des images segmentées (ou *patchs* binaires), un moyennage (pondéré) entre les *patchs* peut être utilisé pour trouver le *patch* non bruité optimal. Dans notre cas, cette recherche peut se faire au sens d'un certain critère et le moyennage de *patchs* binaires peut être efficacement (et rapidement) fait selon le principe de moyennage consensuel de segmentations au sens de ce critère (cf. Fig 7.5).

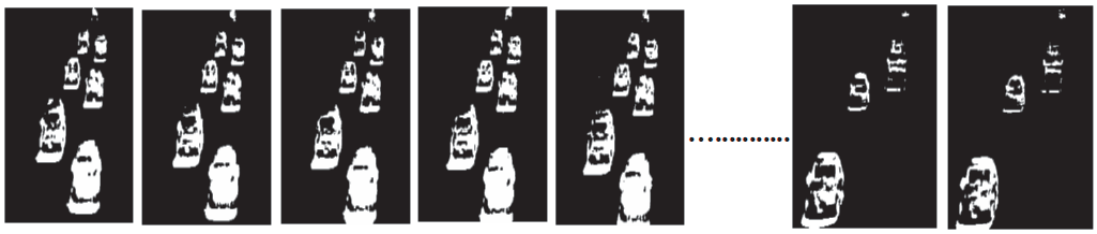


FIGURE 7.4 : Segmentation en mouvement d'une séquence vidéo par l'algorithme de mélange gaussien adaptatif [22] sur la séquence *railway*.

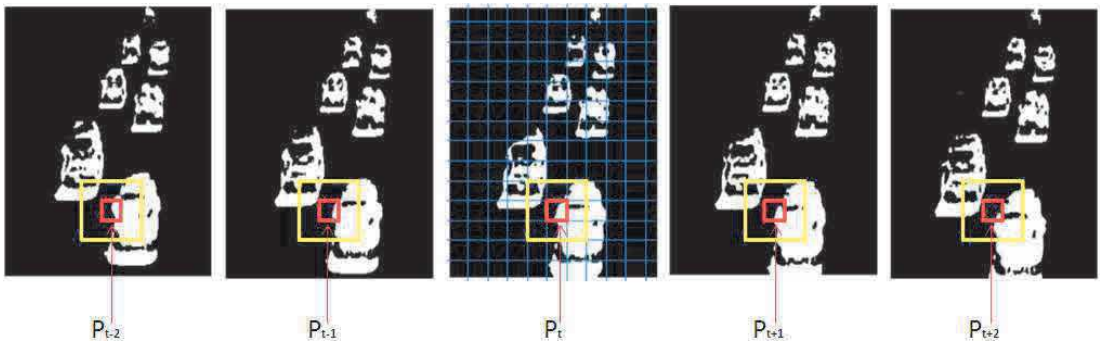


FIGURE 7.5 : Voisinage de recherche spatial (taille 27×27 pixels) et temporel (de taille 5) illustrée par la fenêtre jaune et patch p (de taille 9×9 pixels) à débruiter illustrée par la fenêtre rouge.

7.3 Fusion de plusieurs points de vue pour améliorer la détection du mouvement

Dans le but d'améliorer la partie détection de mouvement et plus précisément n'importe quelle méthode de détection de mouvement, on pourrait aussi fusionner plusieurs cartes de segmentation, obtenues à partir de cet algorithme, mais appliqué selon trois points de vue différents ou selon trois lectures différentes de la séquence d'images vidéo. En effet, la représentation des données vidéo en dimension 2D+temps nous permet en fait de lire le cube 3D de données (puis de détecter le mouvement) de trois façons différentes qui sont complémentaires. Plus précisement ;

- Vue 1 : Lecture traditionnelle ou temporelle selon des coupes frontales (coupes dans le plan (x,y) pour différents t)
- Vue 2 : Lecture du cube d'images verticalement ou selon des coupes sagittales

(coupes dans le plan (y,t) pour différents x)

- Vue 3 : Découpage du cube d'images horizontalement ou selon des coupes transversales (coupes dans le plan (x,t) pour différents y)

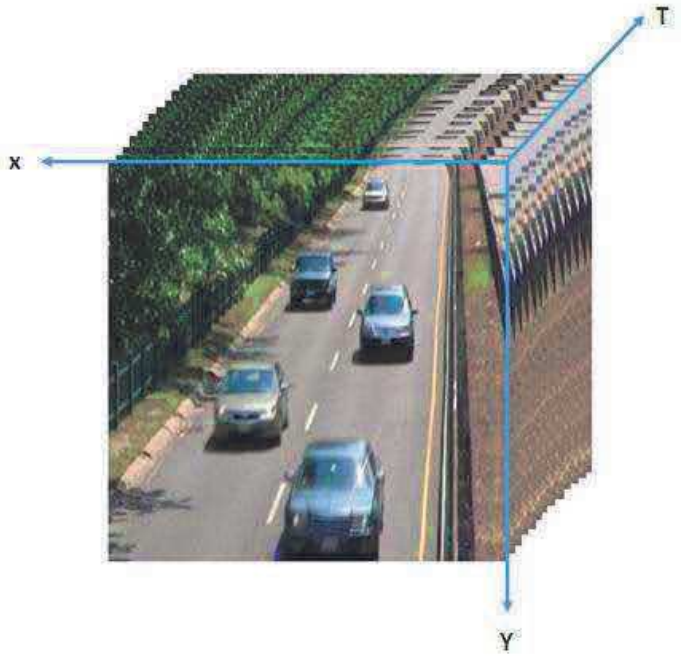


FIGURE 7.6 : Cube vidéo ou vue compacte de la séquence vidéo obtenue en empilant toutes les images 2D le long de la dimension temporelle (téléchargée depuis le site web : changedetection.net).

Notons que la lecture puis le traitement de la séquence vidéo selon ces trois points de vue ont déja été utilisées dans [255] dans la cadre de la reconnaissance automatique des gestes humains dans une séquence d'images mais pas, à notre connaissance, pour améliorer une carte de détection de mouvement d'un algorithme de segmentation.

Afin d'étudier la façon dont on peut se représenter une séquence vidéo en dimension 2D + t, reprenons l'exemple du cube vidéo de la Figure 7.6 représentant une scène obtenue à partir d'une caméra vidéo statique fixée sur le devant du pont d'une autoroute. Cette séquence vidéo est constituée d'une succession d'images numériques (qui peuvent être représentées par une fonction définie dans un espace à trois dimensions $I(x,y,t)$)

prise à des instants successifs avec un intervalle de temps assez court (30 images par seconde).

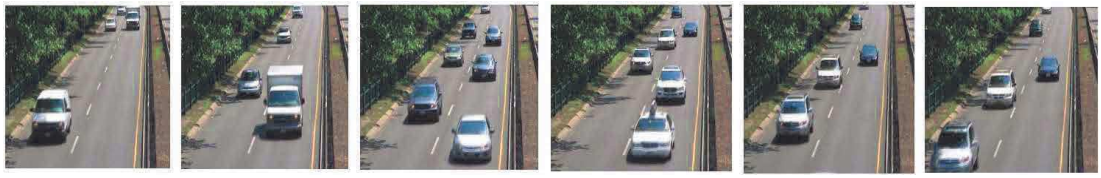
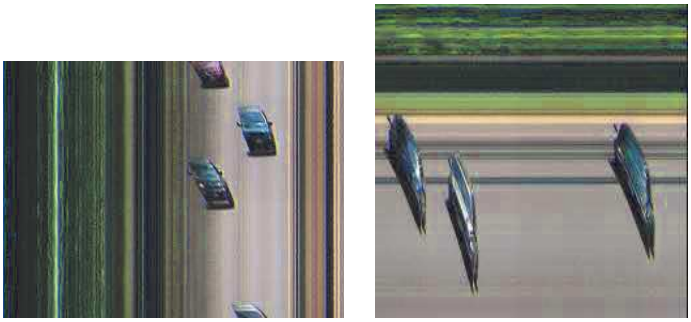


FIGURE 7.7 : Séquence vidéo de 1700 images vue comme une succession d'images 2Ds (ou de coupes dans le plan (x,y) pour différents $t = n\Delta t$ (les images $n = 155$, 266, 827, 1641, 1686 et 1700 sont représentées).



(a) **(b)**

FIGURE 7.8 : **(a)** Coupe sagittale dans le plan (x,t) (vue 2) et **(b)** transversale dans le plan (y,t) (vue 3) pratiquée sur le cube d'images de la Fig. 7.7.

La Figure 7.8(a) représente une coupe dans le plan (x,t) c'est-à -dire une coupe horizontale du cube d'images. On observe, de gauche à droite, des bandes verticales de couleur uniforme représentant les arbres lorsque ces derniers sont vues à partir d'une lecture transversale (vue d'en haut) du cube vidéo, puis les voitures selon cette même coupe et évoluant du haut vers le bas de l'image $I_y(x,t)$ et finalement le bord droit de la route. La Figure 7.8(b) représente une coupe dans le plan (y,t) c'est-à -dire selon une coupe verticale du cube d'images avec, au centre, des motifs spatio-temporels représentant les voitures (vues à partir d'une coupe sagittales) se déplaçant, cette fois-ci de gauche à droite. Ces trois vues sont complémentaires car elles présentent les objets en mouvement (comme les voitures dans cet exemple) de trois façons différentes qui donne-

ront nécessairement trois résultats de segmentation différents une fois le cube de données segmenté 3D reconstitué. Cela se comprend assez bien lorsque cette stratégie sera associée à des algorithmes de détection de mouvement qui ne sont pas uniquement basés sur le calcul d'un simple gradient (*i.e.,* basé sur une simple différence locale d'images) mais comme ceux, par exemple, basées sur la modélisation puis la soustraction de l'arrière-plan estimé à partir des images précédentes de la séquence vidéo ou encore ceux basé sur une estimation du mouvement faîte à partir d'un voisinage contextuel.

Il est cependant important de comprendre que l'application d'un algorithme de détection de mouvement selon la vue 2 (coupe transversale) ou la vue 3 (coupe sagittale) ne pourra être directement utilisé pour une fusion combinant ces trois vues. En effet, ces algorithmes vont détecter directement ou indirectement (indirectement *via*, par exemple, une modélisation puis une soustraction d'un modèle d'arrière-plan) tous changements (ou gradient) d'intensité existant dans la direction des y pour la vue 2 (et respectivement dans la direction des x pour la vue 3). La détection des pixels de fort gradient d'intensité dans la vue 2 inclura donc nécessairement deux classes de pixels ; ceux appartenant aux bords (*edges*) dans la direction des x des objets mobiles (voitures) et ceux appartenant aux bords (dans cette même directions) des objets immobiles (comme les arbres ou les différents marquages signaletiques de la route). Comme, on désire seulement détecter les bords appartenant aux objets mobiles, un filtrage sera donc nécessaire, pour éliminer les bords en y (pour la vue 2) ou en x (pour la vue 3) associés aux objets immobiles. Ce filtrage peut se faire simplement, dans notre cas, une fois le cube de données segmenté 3D reconstitué, en considérant toutes les détections mobiles restant fixes longtemps (dans la séquence d'images) comme étant des artéfacts et appartenant donc nécessairement à des objets immobiles que l'on ne désire pas détecter.

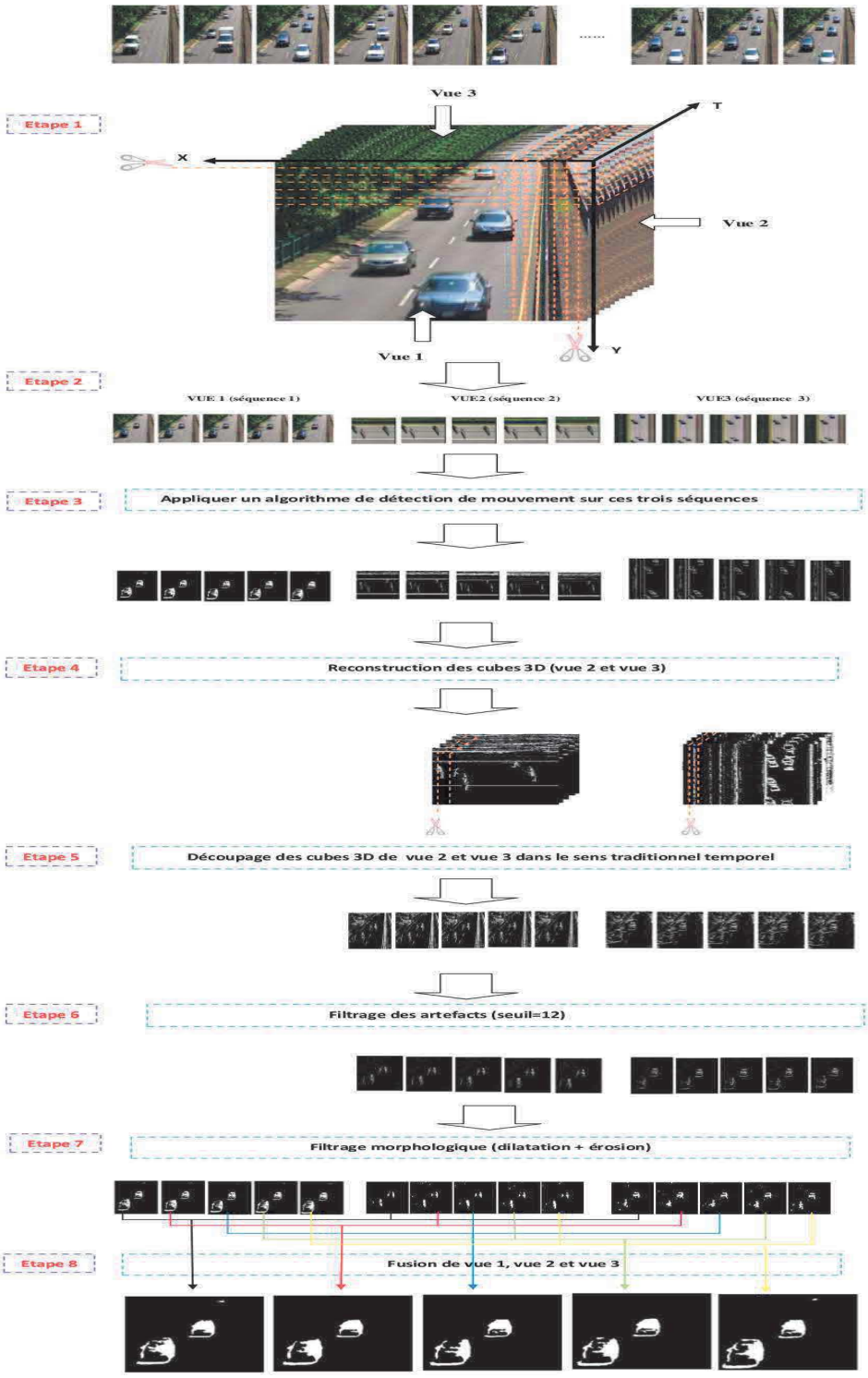


FIGURE 7.9 : étapes d'une stratégie de fusion multivue.

7.4 Analyse en Composante Principale/Indépendante de segmentations

Une application intéressante du principe de distance entre segmentation et moyennage consensuel de segmentations consisterait à adapter l'algorithme d'analyse en composante principale (ACP) sur un ensemble d'images segmentées. Cette application permettrait d'étudier la variabilité ou les modes de déformation d'une structure anatomique (préalablement segmentée) en termes des différentes régions (segmentées) qui la compose et au sens d'un certain critère (*i.e.*, distance entre segmentations). On pourrait ensuite utiliser cette information de variabilité (de segmentation), à travers un modèle paramétrique concis, pour régulariser efficacement un modèle de segmentation ou de reconstructrion 3D de cette même structure anatomique dans une image de rayon X ou échographique (ou une paire d'images) d'un autre patient. Ce même exemple pourrait être bien sur généralisé en 3D pour la reconstruction 3D d'objets. On pourrait citer aussi l'analyse en composante indépendante (ACI) d'images sur une base d'images segmentées qui permettrait de décomposer chaque image segmentée comme la somme de composante individuelle pour mieux comprendre sémantiquement le rôle et la classification future non supervisée de chaque région détectée dans cette segmentation, etc.

7.5 Autres applications possibles

Similairement au principe de navigation dans les grandes bases de données image (en fonction de la façon dont sont disposés et agencés les différents objets présents dans ces images) proposé dans le chapitre 2 et celui de la définition de prototypes de segmentation d'une structure anatomique selon la présence ou non ou la sévérité d'une maladie, on pourrait résumer toute base de données d'images médicales associées à une maladie par un mélange de K prototypes de segmentation pour différentes valeurs de K. La visualisation de ces différents prototypes de segmentation pour différentes valeurs de K, représenté par un graphe de distance, nous permettrait de résumer cette base de données par les différents modes de pathologie et de degré de sévérité d'une maladie donnée.

BIBLIOGRAPHIE

[1] N. J. B. McFarlane and C. P. Schofield. Segmentation and tracking of piglets in images. *Mach. Vis. Appl.*, 8 :187–193, 1995. doi:10.1007/bf01215814.

[2] Z. Zivkovic and F. vander Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7) :773–780, 2006. doi:10.1016/j.patrec.2005.11.005.

[3] N.M. Oliver, B. Rosario, and A.P. Pentland. A bayesian computer vision system for modeling human interactions, cvpr'07. *Proceedings of IEEE International Conference on Pattern Analysis and Machine Intelligence*, 22(8) :831 – 843, Aug 2000. doi:10.1109/34.868684.

[4] G. Papakostas, A. Savio, M. Graña, and V.G. Kaburlasos. A lattice computing approach to alzheimer's disease computer assisted diagnosis based on MRI data. *Neurocomputing*, 150 :35–42, 2015. doi:10.1016/j.neucom.2014.02.076.

[5] M. Mignotte. A label field fusion model with a variation of information estimator for image segmentation. *Information Fusion*, 20 :7–20, 2014. doi:10.1016/j.inffus.2013.10.012.

[6] M. Mignotte. Segmentation by fusion of histogram-based k-means clusters in different color spaces. *IEEE Transactions on Image Processing*, 17(5) :780–787, May 2008. doi:10.1109/tip.2008.920761.

[7] D. Zuwala. *Reconnaissance de symboles sans connaissance a priori*. PhD thesis, PhD thesis, Institut National Polytechnique de Lorraine, 2006. Available at : http://docnum.univ-lorraine.fr/2006_ZUWALA_D.pdf.

[8] A. Khlif and M. Mignotte. Segmentation data visualizing and clustering. *Multimedia Tools and Applications*, 76(1) :1531–1552, 2017. doi:10.1007/s11042-015-3148-6.

[9] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. Subsense : A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, pages 359–373, 2015. doi:10.1109/tip.2014.2378053.

[10] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan. Static and moving object detection using flux tensor with split gaussian models. In *Proceedings of IEEE Workshop on Change Detection, CVPRW'14*, June 2014. doi:10.1109/cvprw.2014.68.

[11] M. D. Gregorio and M. Giordano. Change detection with weightless neural networks. In *Proceedings of IEEE Workshop on Change Detection, CVPRW'14*, pages 409–413, June 2014. doi:10.1109/cvprw.2014.66.

[12] M. Sedky, M. Moniri, and C. C. Chibelushi. Spectral-360 : A physics-based technique for change detection. In *Proceedings of IEEE Workshop on Change Detection, CVPRW'14*, pages 405–408, June 2014. doi:10.1109/cvprw.2014.65.

[13] B. Wang and P. Dudek. A fast self-tuning background subtraction algorithm. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2014. doi:10.1109/cvprw.2014.64.

[14] L. Maddalena and A. Petrosino. The SOBS algorithm : what are the limits ?. In *Proceedings of IEEE Workshop on Change Detection, CVPRW'2012*, pages 21–26. IEEE, 2012. doi:10.1109/cvprw.2012.6238922.

[15] A. Elgammal, D. Harwood, and L. Davis. Non-parametric model for background subtraction. In *Proceedings of the Sixth European Conference on Computer Vision, ECCV'00*, pages 751–767, 2000. doi:10.1007/3-540-45053-x_48.

[16] S. Varadarajan, P. Miller, and H. Zhou. Spatial mixture of gaussians for dynamic background modelling. In *Proceedings of the tenth IEEE International Confe-*

rence on Advanced Video and Signal Based Surveillance, AVSS'13, pages 63–68, 2013. doi:10.1109/avss.2013.6636617.

[17] L. Khelifi and M. Mignotte. A novel fusion approach based on the global consistency criterion to fusing multiple segmentations. *IEEE Transactions on Systems, Man, and Cybernetics : Systems*, 2016. doi:10.1109/tsmc.2016.2531645.

[18] C. Helou and M. Mignotte. A precision-recall criterion based consensus model for fusing multiple segmentations. *International Journal of Signal and Image Processing, IJSIP*, 7 :61–82, 2014. doi:10.14257/ijsip.2014.7.3.07.

[19] M. Mignotte. A label field fusion Bayesian model and its penalized maximum Rand estimator for image segmentation. *IEEE Trans. Image Processing*, 19(6) :1610–1624, June 2010. doi:10.1109/tip.2010.2044965.

[20] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8) :1362–1376, 2010. doi:10.1109/tpami.2009.161.

[21] A. Buades, B.Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. *SIAM Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, 4(2) :490–530, 2005. doi:10.1137/040616024.

[22] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of IEEE International Conference on Pattern Recognition, ICPR'04*, pages 28–31. IEEE, August 2004. doi:10.1109/icpr.2004.1333992.

[23] L. Khelifi and M. Mignotte. GCE-based model for the fusion of multiples color image segmentations. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2574–2578. IEEE, 2016. doi:10.1109/icip.2016.7532824.

[24] R. Hedjam. Segmentation non-supervisée d'images couleur par sur-segmentation markovienne en régions et procédure de regroupement de régions par graphes pondérés. Master's thesis, University of Montreal, 2008. Available at : http://www.iro.umontreal.ca/RachidHedjamMSc.pdf.

[25] S. Benameur, M. Mignotte, F. Destrempes, and J.A. De Guise. Three-dimensional biplanar reconstruction of scoliotic rib cage using the estimation of a mixture of probabilistic prior models. *IEEE trans. on Biomedical Engineering*, 52(10) :2041–2057, 2005. doi:10.1109/tbme.2005.855717.

[26] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(2) :129–141, 2000. doi:10.1109/34.825752.

[27] F. Destrempes and M. Mignotte. Localization of shapes using statistical models and stochastic optimization. *IEEE Trans. Pattern Anal. Machine Intell.*, 29(9) :1603–1615, 2007. doi:10.1109/tpami.2007.1157.

[28] D. E. Ilea and P. F. Whelan. CTex- an adaptive unsupervised segmentation algorithm on color-texture coherence. *IEEE Trans. Image Processing*, 17(10) :1926–1939, 2008. doi:10.1109/tip.2008.2001047.

[29] A. Y. Yang, J. Wright, S. Sastry, and Y. Ma. Unsupervised segmentation of natural images via lossy data compression. *Computer Vision and Image Understanding*, 110(2) :212–225, May 2008. doi:10.1016/j.cviu.2007.07.005.

[30] M. Mignotte. A de-texturing and spatially constrained K-means approach for image segmentation. *Pattern Recognition letter*, 32(2) :359–367, January 2011. doi:10.1016/j.patrec.2010.09.016.

[31] M. S. Allili, N. Bouguila, and D. Ziou. Finite general gaussian mixture modeling and application to image and video foreground segmentation. *J. of Electronic Imaging*, 17(1) :1–13, 2008. doi:10.1117/1.2898125.

[32] Y. S. Chen, B. T. Chen, and W. H. Hsu. Efficient fuzzy c-means clustering for image data. *J. of Electronic Imaging*, 14(1) :013017, 2005. doi:10.1117/1.1879012.

[33] D. Mujica-Vargas, F. J. Gallegos-Funes, A. J. Rosales-Silva, and J. Rubio. Robust c-prototypes algorithms for color image segmentation. *EURASIP Journal on Image and Video Processing*, 63, 2013. doi:10.1186/1687-5281-2013-63.

[34] S. Xu, L. Hu, X. Yang, and X. Liu. A cluster number adaptive fuzzy c-means algorithm for image segmentation. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(5), 2013. doi:10.14257/ijsip.2013.6.5.17.

[35] M. M. Mushrif and A. K. Ray. A-IFS histon based multithresholding algorithm for color image segmentation. *IEEE Signal Processing Lett.*, 16(3) :168–171, 2009. doi:10.1109/lsp.2008.2010820.

[36] M. Mignotte. MDS-based multiresolution nonlinear dimensionality reduction model for color image segmentation. *IEEE Trans. Neural Networks*, 22(3) :447–460, March 2011. doi:10.1109/tnn.2010.2101614.

[37] M. Mignotte. MDS-based segmentation model for the fusion of contour and texture cues in natural images. *Computer Vision and Image Understanding*, 116 :981–990, September 2012. doi:10.1016/j.cviu.2012.05.002.

[38] D. Comaniciu and P. Meer. Mean shift : A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(5) :603–619, 2002. doi:10.1109/34.1000236.

[39] Q. Luo and T.M. Khoshgoftaar. Unsupervised multiscale color image segmentation based on MDL principle. *IEEE Trans. Image Processing*, 15(9) :2755–2761, September 2006. doi:10.1109/tip.2006.877342.

[40] M. A. Carreira-Perpinan. Fast nonparametric clustering with Gaussian blurring mean-shift. In *Proc. of the International Conference on Machine Learning (ICML'06)*, pages 153–160, 2006. doi:10.1145/1143844.1143864.

[41] S. Makrogiannis, I. Vanhamel, S. Fotopoulos, H. Sahli, and J. Cornelis. Watershed-based multiscale segmentation method for color images using automated scale selection. *J. Electronic Imaging*, 14(3) :033007, 2005. doi:10.1117/1.1995711.

[42] I. Mecimore and C. D. Creusere. Unsupervised bitstream based segmentation of images. In *Proc. of the Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop 2009*, pages 643–647, Jan. 2009. doi:10.1109/dsp.2009.4786002.

[43] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(8) :800–810, 2001. doi:10.1109/34.946985.

[44] S. R. Vantaram, E. Saber, S. A. Dianat, M. Shaw, and R. Bhaskar. Multiresolution adaptive and progressive gradient-based color-image segmentation. *J. Electronic Imaging*, 19(1), 2010. doi:10.1117/1.3277150.

[45] F. Deboeverie, P. Veelaert, and W. Philips. Image segmentation with adaptive region growing based on a polynomial surface model. *J. of Electronic Imaging*, 22(4) :043004–043004, 2013. doi:10.1117/1.jei.22.4.043004.

[46] Y. Ma, H. Derksen, W. Hong, and J. Wright. Segmentation of multivariate mixed data via lossy coding and compression. *IEEE Trans. Pattern Anal. Machine Intell.*, 29(9) :1546–1562, 2007. doi:10.1109/tpami.2007.1085.

[47] H. Wu, M. Li, M. Zhang, J. Zheng, and J. Shen. Texture segmentation via scattering transform. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(2), 2013. Available at : https://semanticscholar.org/Wu13.pdf.

[48] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6) :721–741, 1984. doi:10.1109/tpami.1984.4767596.

[49] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Three-class Markovian segmentation of high resolution sonar images. *Computer Vision and Image Understanding*, 76(3) :191–204, 1999. doi:10.1006/cviu.1999.0804.

[50] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Sonar image segmentation using a hierarchical MRF model. *IEEE Trans. Image Processing*, 9(7) :1216–1231, 2000. doi:10.1109/83.847834.

[51] P. Brault and A. Mohammad-Djafari. Unsupervised bayesian wavelet domain segmentation using potts-markov random field modeling. *J. Electronic Imaging*, 14(4) :043011, 2005. doi:10.1117/1.2139967.

[52] F. Destrempes, M. Mignotte, and J.-F. Angers. A stochastic method for Bayesian estimation of hidden Markov random field models with application to a color model. *IEEE Trans. Image Processing*, 14(8) :1096–1108, 2005. doi:10.1109/tip.2005.851710.

[53] F. Destrempes, J.-F. Angers, and M. Mignotte. Fusion of hidden Markov Random Field models and its Bayesian estimation. *IEEE Trans. Image Processing*, 15(10) :2920–2935, October 2006. doi:10.1109/tip.2006.877522.

[54] H. Rachid and M. Mignotte. A hierarchical graph-based Markovian clustering approach for the unsupervised segmentation of textured color images. In *Proc. of the IEEE International Conference on Image Processing (ICIP'09)*, pages 1365–1368, Cairo, Egypt, November 2009. doi:10.1109/icip.2009.5413555.

[55] S. Chatzis and G. Tsechpenakis. The infinite hidden Markov random field model. *IEEE Trans. Neural Networks*, 21(6) :1004–1014, 2010. doi:10.1109/tnn.2010.2046910.

[56] S. Chen, L. Cao, and Y. Wang. Image segmentation by ML-MAP estimations. *IEEE Trans. Image Processing*, 19(9) :2254 – 2264, 2010. doi:10.1109/tip.2010.2047164.

[57] X. He, Z. Song, and J. Fan. A novel level set image segmentation approach with autonomous initialization contour. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(4), 2013. Available at : https://semanticscholar.org/He13.pdf.

[58] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(8) :888–905, 2000. doi:10.1109/34.868688.

[59] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *International Journal on Computer Vision*, 59 :167–181, 2004. doi:10.1023/b:visi.0000022288.19776.77.

[60] J. Wang, Y. Jia, X-S Hua, C. Zhang, and L. Quan. Normalized tree partitionning for image segmentation. In *IEEE Computer Society Conference on computer vision and pattern recognition (CVPR'08)*, pages 1–8,, Anchorage, AK (USA), June 2008. doi:10.1109/cvpr.2008.4587454.

[61] L. Bertelli, B. Sumengen, B. Manjunath, and F. Gibou. A variational framework for multi-region pairwise similarity-based image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 30(8) :1400–1414, 2008. doi:10.1109/tpami.2007.70785.

[62] M. Donoser, M. Urschler, M. Hirzer, and H. Bishof. Saliency driven total variational segmentation. In *Proc. of the IEEE Int'l Conf. Computer Vision (ICCV'09)*, 2009. doi:10.1109/iccv.2009.5459296.

[63] M. Ben Salah, A. Mitiche, and I. Ben Ayed. Multiregion image segmentation by parametric kernel graph cuts. *IEEE Trans. Image Processing*, 20(2) :545–557, 2011. doi:10.1109/tip.2010.2066982.

[64] Q. Chen and C. He. Integrating clustering with level set method for piecewise constant mumford-shah model. *EURASIP Journal on Image and Video Processing*, 1, 2014. doi:10.1186/1687-5281-2014-1.

[65] M. Krninidis and I. Pitas. Color texture segmentation based on the modal energy of deformable surfaces. *IEEE Trans. Image Processing*, 7(18) :1613–1622, 2009. doi:10.1109/tip.2009.2018002.

[66] Y. Wang and C. He. Image segmentation algorithm by piecewise smooth approximation. *EURASIP Journal on Image and Video Processing*, 16, 2012. doi:10.1186/1687-5281-2012-16.

[67] S. Nath and K. Palaniappan. Fast graph partitioning active contours for image segmentation using histograms. *EURASIP Journal on Image and Video Processing*, 2009. doi:10.1155/2009/820986.

[68] Z. Li and J. Fan. Stochastic contour approach for automatic image segmentation. *J. of Electronic Imaging*, 18(4) :043004, 2009. doi:10.1117/1.3257933.

[69] Y. Chen and O.-C Chen. Image segmentation method using thresholds automatically determined from picture contents. *EURASIP Journal on Image and Video Processing*, 2009. doi:10.1155/2009/140492.

[70] F. Nie, J. Li, T. Tu, and P. Zhang. Image segmentation using two-dimensional extension of minimum within-class variance criterion. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(5), 2013. doi:10.14257/ijsip.2013.6.5.02.

[71] S. Chabrier, C. Rosenberger, B. Emile, and H. Laurent. Optimization-based image segmentation by genetic algorithms. *EURASIP Journal on Image and Video Processing*, 842029, 2008. doi:10.1155/2008/842029.

[72] H. Y. Huang, Y. S. Chen, and W. H. Hsu. Color image segmentation using a self-organizing map algorithm. *J. of Electronic Imaging*, 11(2) :136 148, 2002. doi:10.1117/1.1455007.

[73] W. Wang and R. Chung. Image segmentation by optimizing a homogeneity measure in a variational framework. *J. of Electronic Imaging*, 20(1) :013009, 2011. doi:10.1117/1.3543836.

[74] G. Bertrand, J.C. Everat, and M. Couprie. Image segmentation through operators based on topology. *J. Electronic Imaging*, 6(4) :395–405, 1997. doi:10.1117/12.276856.

[75] G. Uma Maheswari, K. Ramar, D. Manimegalai, V. Gomathi, and G. Gowrision. An adaptive color texture segmentation using similarity measure of symbolic object approach. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 4(4), 2011. Available at : https://semanticscholar.org/Maheswari11.pdf.

[76] T. Cour, F. Benezit, and J. Shi. Spectral segmentation with multiscale graph decomposition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005. doi:10.1109/cvpr.2005.332.

[77] A. Strehl and J. Ghosh. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal on Machine Learning Research, JMLR*, 3 :583–617, 2003. doi:10.1162/153244303321897735.

[78] A. Fred and A.K. Jain. Data clustering using evidence accumulation. In *In Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, pages 276–280, August 2002. doi:10.1109/icpr.2002.1047450.

[79] S. Vega-Pons and J. Ruiz-Shulcloper. A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence, IJPRAI*, 25(3) :337–372, May 2011. doi:10.1142/s0218001411008683.

[80] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336) :846–850, 1971. doi:10.1080/01621459.1971.10482356.

[81] Y. Jiang and Z-H. Zhou. SOM ensemble-based image segmentation. *Neural Processing Letters*, 20(3) :171–178, November 2004. doi:10.1007/s11063-004-2022-8.

[82] J. Keuchel and D. Kuttel. Efficient combination of probabilistic sampling approximations for robust image segmentation. In *28th Annual Symposium of the German Association for Pattern Recognition, DAGM-Symposium, Lecture Notes in Computer Science*, pages 41–50, Berlin, Germany, September 2006. doi:10.1007/11861898_5.

[83] P. Wattuya, K. Rothaus, J.-S. Prassni, and X. Jiang. A random walker based approach to combining multiple segmentations. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR'08)*, pages 1–4, Tampa, Florida, USA, December 2008. doi:10.1109/icpr.2008.4761577.

[84] R. Harrabi and E. B. Braiek. Color image segmentation using multi-level thresholding approach and data fusion techniques : application in the breast cancer cells images. *EURASIP Journal on Image and Video Processing*, 1 :1–11, 2012. doi:10.1186/1687-5281-2012-11.

[85] S. Ghosh, J. Pfeiffer, and J. Mulligan. A general framework for reconciling multiple weak segmentations of an image. In *Proceedings of the Workshop on Applications of Computer Vision, (WACV'09)*, pages 1–8, Snowbird, Utah, USA, December 2009. doi:10.1109/wacv.2009.5403029.

[86] A. Alush and J. Goldberger. Ensemble segmentation using efficient integer linear programming. *IEEE Trans. Pattern Anal. Machine Intell.*, 34(10) :1966–1977, 2012. doi:10.1109/tpami.2011.280.

[87] X. Ceamanos, B. Waske, J. Atli Benediktsson, J. Chanussot, M. Fauvel, and J. R. Sveinsson. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4) :293–307, 2010. doi:10.1080/19479832.2010.485935.

[88] B. Song and P. Li. A novel decision fusion method based on weights of evidence model. *International Journal of Image and Data Fusion*, 5(2) :123–137, 2014. doi:10.1080/19479832.2014.894143.

[89] L. Khelifi and M. Mignotte. EFA-BMFM : A multi-criteria framework for the fusion of colour image segmentation. *Information Fusion*, 38 :104–121, 2017. doi:10.1016/j.inffus.2017.03.001.

[90] L. Khelifi and M. Mignotte. A multi-objective decision making approach for solving the image segmentation fusion problem. *IEEE Transactions on Image Processing*, 26(8) :3831–3845, 2017. doi:10.1109/tip.2017.2699481.

[91] I. Daoudi. *Recherche par similarité dans les bases de données multimédia : application à la recherche par le contenu d'images*. PhD thesis, Université Mohammed V-Agdal, Faculté des Sciences, Rabat, 2009. Available at : http://theses.insa-lyon.fr/daoudi2009.pdf.

[92] N. Idrissi. *La navigation dans les bases d'images : prise en compte des attributs de texture*. PhD thesis, Université de Nantes ; Faculté des sciences de Rabat, 2008. Available at : https://hal.archives-ouvertes.fr/idrissi2008.pdf.

[93] S. Wang. *A robust CBIR approach using local color histograms*, 2001. doi:10.7939/R3NV99B8V.

[94] J. Huang, S.R. Kumar, M. Mitra, Wei-Jing Zhu, and R. Zabih. Image indexing using color correlograms. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 762–768. IEEE, 1997. doi:10.1109/cvpr.1997.609412.

[95] Nhu Van Nguyen. *Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images*. Theses, Université de La Rochelle, September 2011. Available at : https://tel.archives-ouvertes.fr/nguyen2011.pdf.

[96] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker. Query by image and video content : The QBIC system. *computer*, 28(9) :23–32, 1995. doi:10.1109/2.410146.

[97] John R. Smith and Shih-Fu Chang. Tools and techniques for color image retrieval. In *Storage and Retrieval for Image and Video Databases (SPIE)*, volume 2670, pages 2–7, 1996. doi:10.1117/12.234781.

[98] M. Tuceryan, Anil K Jain, et al. Texture analysis. *Handbook of pattern recognition and computer vision*, 2 :235–276, 1993. doi:10.1142/9789814343138_0010.

[99] L. Paulhac. *Outils et méthodes d'analyse d'images 3D texturées : application à la segmentation des images échographiques*. PhD thesis, Université François Rabelais-Tours, 2009. Available at : https://tel.archives-ouvertes.fr/paulhac2009.pdf.

[100] Robert M. Haralick, K.Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, SMC-3(6) :610–621, 1973. doi:10.1109/tsmc.1973.4309314.

[101] Paul J Curran. The semivariogram in remote sensing : an introduction. *Remote sensing of Environment*, 24(3) :493–507, 1988. doi:10.1016/0034-4257(88)90021-1.

[102] A.K. Jain, S.K. Bhattacharjee, Y. Chen. On texture in document images. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on*, pages 677–680. IEEE, 1992. doi:10.1109/cvpr.1992.223203.

[103] Chaman L. Sabharwal, S. R. Subramanya. Indexing image databases using wavelet and discrete fourier transform. In *Proceedings of the 2001 ACM symposium*

on Applied computing, pages 434–439. ACM, 2001. doi:10.1145/372202.372395.

[104] S.G. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7) :674–693, 1989. doi:10.1109/34.192463.

[105] Regina K. Ferrell, Shaun S. Gleason, and Kenneth W. Tobin. Application of fractal encoding techniques for image segmentation. In *Proceedings of SPIE*, volume 5132, pages 69–77, 2003. doi:10.1117/12.514943.

[106] Güray Erus. *Reconnaissance d'objets cartographiques dans les images satellitaires à haute résolution*. PhD thesis, Paris 5, 2008. Available at : https://www.grss-ieee.org/Erus_These_2008.pdf.

[107] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2) :179–187, 1962. doi:10.1109/tit.1962.1057692.

[108] Sahibsingh A. Dudani, Kenneth J. Breeding, and Robert B. McGhee. Aircraft identification by moment invariants. *IEEE transactions on computers*, 100(1) :39–46, 1977. doi:10.1109/tc.1977.5009272.

[109] H. Qjidaa, L. Radouane. Robust line fitting in a noisy image by the method of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11) :1216–1223, 1999. doi:10.1109/34.809115.

[110] J. Zunic, N. Sladoje. Efficiency of characterizing ellipses and ellipsoids by discrete moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4) :407–414, 2000. doi:10.1109/34.845384.

[111] Michael Reed Teague. Image analysis via the general theory of moments. *Journal of the Optical Society of America (920-930)*, 70, 1980. doi:10.1364/josa.70.000920.

[112] A. Khotanzad, Y.H. Hong. Invariant image recognition by zernike moments. *IEEE Transactions on pattern analysis and machine intelligence*, 12(5) :489–497, 1990. doi:10.1109/34.55109.

[113] R.Dahyot. *Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier*. PhD thesis, Université Louis Pasteur - Strasbourg I, 2003. Available at : https://www.scss.tcd.ie/dahyot2003.pdf.

[114] A. Auclair. *Méthodes rapides pour la recherche des plus proches voisins SIFT : application à la recherche d'images et contributions à la reconstruction 3D multi-vues*. PhD thesis, Paris 5, 2009. Available at : http://www.math-info.univ-paris5.fr/Auclair.pdf.

[115] Hans P. Moravec. Rover visual obstacle avoidance. In *IJCAI Proceedings of the 7th international joint conference on Artificial intelligence*, pages 785–790, 1981. Available at : https://dl.acm.org/Moravec.

[116] C. Harris, M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Manchester, UK, 1988. doi:10.5244/c.2.23.

[117] G. GALES. Détection et mise en correspondance de points d'intérêt pour la propagation de germes. *MS thèses, Institut de Recherche en Informatique de Toulouse*, 2007. Available at : https://www.irit.fr/GalesRapportMaster2007.pdf.

[118] D.G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999. doi:10.1109/iccv.1999.790410.

[119] G.Vincent. *Suivi d'objets d'intérêt dans une séquence d'images : des points saillants aux mesures statistiques*. Theses, Université Nice Sophia Antipo-

lis, December 2008. Available at : https://hal.inria.fr/Vincent_Garcia_thesis.pdf.

[120] H. Bay, T. Tuytelaars, L.V. Gool. SURF : Speeded up robust features. *Computer vision–ECCV 2006*, pages 404–417, 2006. doi:10.1007/11744023_32.

[121] Y. Ke, R. Sukthankar. PCA-SIFT : A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, 2004. doi:10.1109/cvpr.2004.1315206.

[122] G. Young, A.S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3(1) :19–22, 1938. doi:10.1007/bf02287916.

[123] Warren S. Torgerson. Multidimensional scaling : I. theory and method. *Psychometrika*, 17(4) :401–419, 1952. doi:10.1007/bf02288916.

[124] C. Stauffer, W.E.L. Grimson. Adaptive background mixture models for real-time tracking, CVPR'99. *Conference on Computer Vision and Pattern Recognition*, 2, June 1999. doi:10.1109/cvpr.1999.784637.

[125] S. P. Lloyd. Least squares quantization in PCM. *IEEE Trans. Inform. Theory*, 28(2) :129–136, February 1982. doi:10.1109/tit.1982.1056489.

[126] G. Schaefer. Interactive browsing of image repositories. *Computer Vision and Graphics*, pages 236–244, 2012. doi:10.1007/978-3-642-33564-8_29.

[127] G. Schaefer. A next generation browsing environment for large image repositories. *Multimedia Tools and Applications*, 47(1) :105–120, 2010. doi:10.1007/s11042-009-0409-2.

[128] I. Bartolini, P. Ciaccia, and M. Patella. Adaptively browsing image databases with PIBE. *Multimedia Tools and Applications*, 31(3) :269–286, 2006. doi:10.1007/s11042-006-0044-0.

[129] D. Heesch. A survey of browsing models for content based image retrieval. *Multimedia Tools and Applications*, 40(2) :261–284, 2008. doi:10.1007/s11042-008-0207-2.

[130] T. Leelanupab, Y. Feng, V. Stathopoulos, and Joemon M. Jose. A simulated user study of image browsing using high-level classification. In *SAMT*, pages 3–15. Springer, 2009. doi:10.1007/978-3-642-10543-2_3.

[131] J. Urban, Joemon M. Jose, and Cornelis J. van Rijsbergen. An adaptive technique for content-based image retrieval. *Multimedia Tools and Applications*, 31(1) :1–28, 2006. doi:10.1007/s11042-006-0035-1.

[132] M. Meila. Comparing clusterings - an axiomatic view. In *Proceedings of the 2005 22nd International Conference on Machine Learning (ICML'05)*, pages 577–584, Bonn, Germany, August 2005. doi:10.1145/1102351.1102424.

[133] M. Meila. Comparing clustering an information based distance. *Journal of multivariate analysis*, 98(5) :873–895, 2007. doi:10.1016/j.jmva.2006.11.013.

[134] M. Mignotte. A non-stationary MRF model for image segmentation from a soft boundary map. *Pattern Analysis and Applications*, 17(1) :129–139, January 2014. doi:10.1007/s10044-012-0272-z.

[135] M. Mignotte. Nonparametric multiscale energy-based model and its application in some imagery problems. *IEEE Trans. Pattern Anal. Machine Intell.*, 26(2) :184–197, February 2004. doi:10.1109/tpami.2004.1262180.

[136] M. Mignotte. A multiresolution markovian fusion model for the color visualization of hyperspectral images. *IEEE Trans. on Geoscience and Remote Sensing*, 48(12) :4236–4247, December 2010. doi:10.1109/tgrs.2010.2051553.

[137] G. Young and A.S. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3, January 1938. doi:10.1007/bf02287916.

[138] W.S. Torgerson. Multidimensional scaling : I. theory and method. *Psychometrika*, 17 :401–419, 1952. doi:10.1007/bf02288916.

[139] L. Cayton and S. Dasgupta. Robust euclidean embedding. In *In Proceedings of the 23rd International Conference on Machine learning*, pages 169–176. ACM Press, 2006. doi:10.1145/1143844.1143866.

[140] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th International Conference on Computer Vision (ICCV'01)*, volume 2, pages 416–423, Vancouver, British Columbia, Canada, July 2001. doi:10.1109/iccv.2001.937655.

[141] I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling : Theory and Applications*. Springer, 2005. doi:10.1007/0-387-28981-X.

[142] T. Kim, H. Adeli, C. Ramos, B. Kang *Signal processing, image processing and pattern recognition*. Springer, 2011. doi:10.1007/978-3-642-27183-0.

[143] C. Alvarez, A. Id-Oumohmed, M. Mignotte, and J.-Y. Nie. Toward cross-language and cross-media image retrieval. In *5th Workshop on Cross Langage Evaluation Forum, CLEF 2004 Lecture notes in Computer science, Multilingual information access for text, speech and images*, volume 3491, pages 676–688, Bath, United Kingdom, September 2004. doi:10.1007/11519645_66.

[144] A. Id-Oumohmed, M. Mignotte, and J.-Y. Nie. Semantic-based cross-media image retrieval. In *3rd International Conference on Advances in Pattern Recognition, ICAPR'05, Lecture Notes in Computer Science, volume LNCS 3686, Pattern Recognition and Data Mining, Proceedings Part 2*, pages 414–423, Bath, United Kingdom (UK), August 2005. doi:10.1007/11552499_47.

[145] H. Xiao, Y. Liu, and M. Zhang. Fast l 1-minimization algorithm for robust background subtraction. *EURASIP Journal on Image and Video Processing*, 2016(1) :45, 2016. doi:10.1186/s13640-016-0150-5.

[146] M. Cristani, M. Farenzena, D. Bloisi, and V. Murino. Background subtraction for automated multisensor surveillance : a comprehensive review. *EURASIP Journal on Advances in signal Processing*, 2010(1) :343057, 2010. doi:10.1155/2010/343057.

[147] S. Lee and C. Lee. Low-complexity background subtraction based on spatial similarity. *EURASIP Journal on Image and Video Processing*, 2014(1) :30, 2014. doi:10.1186/1687-5281-2014-30.

[148] T.H. Tsai, C.Y. Lin, and S.Y. Li. Algorithm and architecture design of human-machine interaction in foreground object detection with dynamic scene. *IEEE Trans. on Circuits and Systems for Video Technology*, 23(1) :15–29, 2013. doi:10.1109/tcsvt.2012.2202193.

[149] A. Bugeau and P. Pérez. Track and cut : simultaneous tracking and segmentation of multiple objects with graph cuts. *EURASIP Journal on Image and Video Processing*, 2008(1) :317278, 2008. doi:10.1155/2008/317278.

[150] S. Velipasalar, J. Schlessman, C.Y. Chen, W.H. Wolf, and J.P. Singh. A scalable clustered camera system for multiple object tracking. *EURASIP Journal on Image and Video Processing*, 2008(1) :542808, 2008. doi:10.1155/2008/542808.

[151] L.C. Belhadj and M. Mignotte. Spatio-temporal fastmap-based mapping for human action recognition. In *23rd IEEE International Conference on Image Processing, ICIP'16*, Phoenix, Arizona, US, September 2016. doi:10.1109/icip.2016.7532919.

[152] C. Zhao, X. Wang, and W.K. Cham. Background subtraction via robust dictionary learning. *EURASIP Journal on Image and Video Processing*, 2011. doi:10.1155/2011/972961.

[153] A. Shi, G. Gao, and S. Shen. Change detection of bitemporal multispectral images based on FCM and DS theory. *EURASIP Journal on Advances in Signal Processing*, 2016(1) :96, 2016. doi:10.1186/s13634-016-0397-0.

[154] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar. Cdnet 2014 : An expanded change detection benchmark dataset. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW'14*, pages 393–400, June 2014. doi:10.1109/cvprw.2014.126.

[155] B. Lee and M. Hedley. Background estimation for video surveillance. In *Image and Vision Computing New Zealand, IVCNZ'02*, pages 315–320, November 2002. Available at : https://publications.csiro.au/Lee2002.

[156] K. Karman and A. Brandt. Moving object recognition using an adaptive background memory. *Time-Varying Image Processing and Moving Object Recognition*, 2 :297–307, 1990. Available at : http://www.worldcat.org/time-varying-image-processing2.

[157] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower : principles and practice of background maintenance. In *Proceedings of the Seventh IEEE International Conference on Computer Vision, ICCV'98*, volume 1, pages 255–261, September 1999. doi:10.1109/iccv.1999.791228.

[158] J. Zheng, Y. Wang, N. Nihan, and M. Hallenbeck. Extracting roadway background image : Mode-based approach. *J. Transp. Res. Rep.*, 1944(1) :82–88, 2006. doi:10.3141/1944-11.

[159] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder : real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :780–785, 1997. doi:10.1109/34.598236.

[160] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE International Conference on com-*

puter vision and pattern recognition, CVPR'99, volume 2, pages 246–252, 1999. doi:10.1109/cvpr.1999.784637.

[161] Y. Benezeth, P.M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger. Comparative study of background subtraction algorithms. *Journal of Electronic Imaging*, 19(3) :1–12, 2010. doi:10.1117/1.3456695.

[162] D. Li, L. Xu, and E. D. Goodman. Illumination-robust foreground detection in a video surveillance system. *IEEE Trans. on Circuits and Systems for Video Technology*, 23(10) :1637–1650, October 2013. doi:10.1109/tcsvt.2013.2243649.

[163] L. Maddalena and A. Petrosino. A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection. *Neural Computing and Applications*, 19(2) :179–186, 2010. doi:10.1007/s00521-009-0285-8.

[164] G. Ramirez-Alonso and M. I. Chacon-Murguia. Auto-adaptive parallel SOM architecture with a modular analysis for dynamic object segmentation in videos. *Neurocomputing*, 175 :990–1000, 2016. doi:10.1016/j.neucom.2015.04.118.

[165] A. Miron and A. Badii. Change detection based on graph cuts. In *Proceedings of International Conference on Systems, Signals and Image Processing, IWSSIP'15*, pages 273–276, September 2015. doi:10.1109/iwssip.2015.7314229.

[166] G. Allebosch, F. Deboeverie, P. Veelaert, and W. Philips. EFIC : Edge based foreground background segmentation and interior classification for dynamic camera viewpoints. In *Advanced Concepts for Intelligent Vision Systems, 16th International Conference*, volume 9386 of *Lecture Notes in Computer Science (LNCS)*, pages 130–141, Catania, Italy, October 2015. Springer. doi:10.1007/978-3-319-25903-1_12.

[167] G. Allebosch, D. Van Hamme, F. Deboeverie, P. Veelaert, and W. Philips. C-EFIC : Color and edge based foreground background segmentation with interior classification. In José Braz, Julien Pettré, Paul Richard, Andreas Kerren, Lars Linsen, Sebastiano Battiato, and Francisco Imai, editors, *Computer Vision, Imaging and Computer Graphics Theory and Applications : 10th International Joint Conference, VISIGRAPP 2015, Berlin, Germany, March 11-14, 2015, Revised Selected Papers*, volume 598 of *Communications in Computer and Information Science*, pages 433–454. Springer International Publishing, 2016. doi:10.1007/978-3-319-29971-6_23.

[168] P.M. Jodoin, M. Mignotte, and J. Konrad. Light and fast statistical motion detection method based on ergodic model. In *13th IEEE International Conference on Image Processing, ICIP'06*, pages 1053–1056, Atlanta, GA, USA, October 2006. doi:10.1109/icip.2006.312736.

[169] Y. Chen and J. Wang. Learning sharable models for robust background subtraction. In *IEEE International Conference in Multimedia and Expo, ICME'15*, pages 1–6, June 2015. doi:10.1109/icme.2015.7177419.

[170] P.L. St-Charles, G.A. Bilodeau, and R. Bergevin. Flexible background subtraction with self-balanced local sensitivity. In *Proceedings of IEEE Workshop on Change Detection, CVPRW'14*, June 2014. doi:10.1109/cvprw.2014.67.

[171] L. Yang, H. Cheng, J. Su, and X. Li. Pixel-to-model distance for robust background reconstruction. *IEEE Trans. Circuits Syst. Video Techn.*, 26(5) :903–916, 2016. doi:10.1109/tcsvt.2015.2424052.

[172] H. Sajid and S.C. S. Cheung. Background subtraction for static & moving camera. In *IEEE International Conference on Image Processing, ICIP'15*, pages 4530–4534, September 2015. doi:10.1109/icip.2015.7351664.

[173] P.M. Jodoin, M. Mignotte, and J. Konrad. Statistical background subtraction methods using spatial cues. *IEEE Trans. on Circuits and Systems for Video Tech-*

nology, 17(12) :1758–1764, December 2007. doi:10.1109/tcsvt.2007.906935.

[174] B. N. Subudhi, P. K. Nanda, and A. Ghosh. A change information based fast algorithm for video object detection and tracking. *IEEE Trans. on Circuits and Systems for Video Technology*, 21(7) :993–1004, 2011. doi:10.1109/tcsvt.2011.2133870.

[175] C.C. Chiu, M.Y. Ku, and L.W. Liang. A robust object segmentation system using a probability-based background extraction algorithm. *IEEE Trans. on Circuits and Systems for Video Technology*, 20(4) :518–528, 2010. doi:10.1109/tcsvt.2009.2035843.

[176] J.M. Guo, C.H. Hsia, Y.F. Liu, M.H. Shih, C.H. Chang, and J.Y. Wu. Fast background subtraction based on a multilayer codebook model for moving object detection. *IEEE Trans. on Circuits and Systems for Video Technology*, 23(10) :1809–1821, October 2013. doi:10.1109/tcsvt.2013.2269011.

[177] V. Reddy, C. Sanderson, and B. Lovell. Improved foreground detection via block-based classifier cascade with probabilistic decision integration. *IEEE Trans. on Circuits and Systems for Video Technology*, 23(1) :83–93, 2013. doi:10.1109/tcsvt.2012.2203199.

[178] B. Klare and S. Sarkar. Background subtraction in varying illuminations using an ensemble based on an enlarged feature set. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 66–73. IEEE, 2009. doi:10.1109/cvpr.2009.5204078.

[179] M. Camplani and L. Salgado. Background foreground segmentation with rgb-d kinect data : An efficient combination of classifiers. *J. Visual Communication and Image Representation*, 25(1) :122–136, 2014. doi:10.1016/j.jvcir.2013.03.009.

[180] S. Bianco, G. Ciocca, and R. Schettini. Combination of Video Change Detection Algorithms by Genetic Programming. *IEEE Transactions on Evolutionary Computation*, 914–928, 2017. doi:10.1109/TEVC.2017.2694160.

[181] C. Mythili and V. Kavitha. Color image segmentation using ERKFCM. *International of computer Applications*, 41(20), 2012. Available at : https://pdfs.semanticscholar.org/ERKFCM.pdf.

[182] K. Chen, Y. Ma, and J. Liu. Segmentation by fusion of self-adaptive SFCM cluster in multi-color space components. *International Journal of Image Processing (IJIP)*, 6(2) :157–166, 2012. Available at : https://pdfs.semanticscholar.org/SFCM.pdf.

[183] C. Charrier, G. Lebrun, and O. Leezoray. Fusion de segmentations d'images microscopiques par SVM dans differents espaces couleur selon la theorie de l'evidence. In *20 ieme Colloque sur le traitement du signal et des images, GRETSI, Groupe d'etudes du Traitement du Signal et des Images*, France, 2005. Available at : http://documents.irevues.inist.fr/charrier.pdf.

[184] S.B. Chaabane, M. Sayadi, F. Fnaiech, and E. Brassart. Colour image segmentation using homogeneity method and data fusion techniques. *EURASIP journal on advances in signal processing*, 2010 :1, 2010. doi:10.1155/2010/367297.

[185] Z. Xue and R. S. Blum. Concealed weapon detection using color image fusion. In *Proceedings of the 6th International Conference on Information Fusion*, volume 1, pages 622–627, July 2003. doi:10.1109/icif.2003.177504.

[186] Z. Zhou, B. Wang, S. Li, and M. Dong. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters. *Information Fusion*, 30 :15 – 26, 2016. doi:10.1016/j.inffus.2015.11.003.

[187] Y. Du, H. Ai, and S. Lao. Evaluation of color spaces for person re-identification. In *Proceedings of the IEEE 21st International Conference on Pattern Recog-*

nition, ICPR'12, pages 371–1374, November 2012. Available at : http://ieeexplore.ieee.org/ICPR12.pdf.

[188] S. Chindaro, K. Sirlantzis, and F. Deravi. Texture classification system using colour space fusion. *Electronics Letters*, 41(10), 2005. doi:10.1049/el:20050594.

[189] L. Nanni and A. Lumini. Fusion of color spaces for ear authentication. *Pattern Recognition*, pages 1906–1913, 2009. doi:10.1016/j.patcog.2008.10.016.

[190] P.S. Hiremath and M. Hiremath. Fuzzy geometric face model for face detection based on skin color fusion model. In *Proceedings of International Conference on Advances in Intelligent Systems and Computing, ICAISC'13*, volume 174, pages 977–982, 2012. doi:10.1007/978-81-322-0740-5_118.

[191] P. Peer, J. Kovac, and F. Solina. Human skin colour clustering for face detection. In *Proceedings of EUROCON*, pages 144–148, September 2003. doi:10.1109/eurcon.2003.1248169.

[192] F.H. Xiang and S.A. Suandi. Fusion of multi-color space for human skin region segmentation. *Int. J. Inform. Electron. Eng.*, 3 :172–174, 2013. doi:10.7763/ijiee.2013.v3.292.

[193] Z.H. Al-Tairi, M.I. Saripan R. W. Rahmat, and P.S. Sulaiman. Skin segmentation using YUV and RGB color spaces. *J. Inf. Process. Syst.*, 10(2) :283–299, 2014. doi:10.3745/jips.02.0002.

[194] M.A. Sheha, M. S. Mabrouk, and A. Sharawy. Pigmented skin lesion diagnosis by automated imaging system. *Journal of Bioengineering & Biomedical Science*, 5(170), 2015. doi:10.4172/2155-9538.1000170.

[195] Z. Zhou, D. Li, D. Lin, L. Liang, and L. Liu. Application of multi-color space feature fusion in color difference processing. In *Proceedings of Fourth In-*

ternational Conference on Instrumentation and Measurement, Computer, Communication and Control, IMCCC'14, pages 751–755, September 2014. doi:10.1109/imccc.2014.159.

[196] T. Ding and D. Jia. Detection of similar targets using multiple color space feature fusion. *Journal of Tsinghua University (Science and Technology)*, 46(2) :176–179, 2006. Available at : http://en.cnki.com.cn/Article.

[197] T.H. Chen, P.H. Wu, and Y.C. Chiou. An early fire-detection method based on image processing. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 3, pages 1707–1710. IEEE, 2004. doi:10.1109/icip.2004.1421401.

[198] J. Chen, Y. He, and J. Wang. Multi-feature fusion based fast video flame detection. *Building and Environment*, 45(5) :1113–1122, 2010. doi:10.1016/j.buildenv.2009.10.017.

[199] L. Rossi, M. Akhloufi, and Y. Tison. On the use of stereovision to develop a novel instrumentation system to extract geometric fire fronts characteristics. *Fire Safety Journal*, 46(1) :9–20, 2011. doi:10.1016/j.firesaf.2010.03.001.

[200] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9) :1582–1596, 2010. doi:10.1109/tpami.2009.154.

[201] A. Setkov, M. Gouiffs, and C. Jacquemin. Color invariant feature matching for image geometric correction. In *Proceedings of Sixth International Conference on computer vision/computer graphics collaboration techniques and applications*, volume 7, pages 751–755, 2013. doi:10.1145/2466715.2466731.

[202] Color space. Wikipedia. accessed August 2, 2017. at https://en.wikipedia.org/wiki/Color_space.

[203] M. Wozniak, M. Grana, and E. Corchado. A survey of multiple classifier systems as hybrid systems. *Information Fusion*, 16 :3 – 17, 2014. Special Issue on Information Fusion in Hybrid Intelligent Fusion Systems. doi:10.1016/j.inffus.2013.04.006.

[204] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi. Multisensor data fusion : A review of the state-of-the-art. *Information Fusion*, 14(1) :28 – 44, 2013. doi:10.1016/j.inffus.2011.08.001.

[205] A. Buades, J.L. Lisani, and J.M. Morel. Dimensionality of color space in natural images. *Journal of the Optical Society of America A*, 28(2) :203–209, 2011. doi:10.1364/josaa.28.000203.

[206] J.P. Guilford. *Psychometric Methods*. New York : McGraw Hill, 83(343) :228 1937. doi:10.1192/bjp.83.343.228.

[207] P.L. St-Charles, G.A. Bilodeau, and R. Bergevin. Improving background subtraction using local binary similarity patterns. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision, WACV'15*, pages 990–997, Mars 2014. doi:10.1109/wacv.2014.6836059.

[208] Y. Wang, Z.M. Luo, and P.M. Jodoin. Interactive deep learning method for segmenting moving objects. *Pattern recognition letters*, 96 :66–75, 2017. doi:10.1016/j.patrec.2016.09.014.

[209] J. Ashburner and K.J. Friston. Voxel-based morphometry the methods. *Neuroimage*, 11(6) :805–821, 2000. doi:10.1006/nimg.2000.0582.

[210] R. Wolz, V. Julkunen, J. Koikkalainen, E. Niskanen, D.P. Zhang and D. Rueckert. Multi-method analysis of MRI images in early diagnostics of alzheimer's disease. *PloS one*, 6(10) :e25446, 2011. doi:10.1371/journal.pone.0025446.

[211] J.F. Mangin, D. Rivière, A. Cachia, D. Papadopoulos-Orfanos, D.L. Collins, A.C. Evans, and J. Régis. Object-based strategy for morphometry of the ce-

rebral cortex. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 160–171. IPMI, July 2003. doi:10.1007/978-3-540-45087-0_14.

[212] M. Toews, W. Wells, D.L. Collins, and T. Arbel. Feature-based morphometry : Discovering group-related anatomical patterns. *NeuroImage*, 49(3) :2318–2327, 2010. doi:10.1016/j.neuroimage.2009.10.032.

[213] O.B. Ahmed, J. Benois-Pineau, M. Allard, C. B. Amar, and G. Catheline. Classification of alzheimer's disease subjects from MRI using hippocampal visual features. *Multimedia Tools and Applications*, 74(4) :1249–1266, 2015. doi:10.1007/s11042-014-2123-y.

[214] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, and T.-F. Yuan. Detection of subjects and brain regions related to alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. *Frontiers in Computational Neuroscience*, 9 :66, 2015. doi:10.3389/fncom.2015.00066.

[215] L.M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C.M. Clark, P.S. Aisen, R.C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, et al. Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. *Annals of neurology*, 65(4) :403–413, 2009. doi:10.1002/ana.21610.

[216] R. Brian, R.A. Cohen, A. Gongvatana, O.C. Okonkwo, C.E. Johanson, E.G. Stopa, J.E. Donahue, G.D Silverberg, Alzheimer's Disease Neuroimaging Initiative, et al. Brain ventricular volume and cerebrospinal fluid biomarkers of alzheimer's disease. *Journal of Alzheimer's Disease*, 20(2) :647–657, 2010. doi:10.3233/jad-2010-1406.

[217] B. M. French, M. R. Dawson, and A. R. Dobbs. Classification and staging of dementia of the alzheimer type : a comparison between neural networks and linear discriminant analysis. *Archives of neurology*, 54(8) :1001–1009, 1997. doi:10.1001/archneur.1997.00550200057011.

[218] A. Savio, M. García-Sebastián, C. Hernández, M. Graña, and J. Villanúa. Classification results of artificial neural networks for alzheimer's disease detection. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 641–648, September 2009. doi:10.1007/978-3-642-04394-9_78.

[219] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali. Support vector machine-based classification of alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology*, 51(2) :73–83, 2009. doi:10.1007/s00234-008-0463-x.

[220] L. Mesrob, B. Magnin, O. Colliot, M. Sarazin, V. Hahn-Barma, B. Dubois, and H. Benali. Identification of atrophy patterns in alzheimer's disease based on SVM feature selection and anatomical parcellation. In *International Workshop on Medical Imaging and Virtual Reality*, pages 124–132, August 2008. doi:10.1007/978-3-540-79982-5_14.

[221] P.P. de Magalhães Oliveira Jr, R. Nitrini, G. Busatto, C. Buchpiguel, J. Sato, R. João, and E. Amaro Jr. Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect alzheimer's disease. *Journal of Alzheimer's Disease*, 19(4) :1263–1272, 2010. doi:10.3233/jad-2010-1322.

[222] A. Savio, M. Grańa, and J. Villanúa. Deformation based features for alzheimer's disease detection with linear SVM. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 336–343, May 2011. doi:10.1007/978-3-642-21222-2_41.

[223] E.S.A. El-Dahshan, T. Hosny, and A.B.M. Salem. Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Processing*, 20(2) :433–441, 2010. doi:10.1016/j.dsp.2009.07.002.

[224] Z. Lao, D. Shen, Z. Xue, B. Karacali, S.M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *Neuroimage*, 21(1) :46–57, 2004. doi:10.1016/j.neuroimage.2003.09.027.

[225] A. Savio, M. T. García-Sebastián, D. Chyzyk, C. Hernández, M. Graña, A. Sistiaga, A.L. De Munain, and J. Villanúa. Neurocognitive disorder detection based on feature vectors extracted from vbm analysis of structural MRI. *Computers in biology and medicine*, 41(8) :600–610, 2011. doi:10.1016/j.compbiomed.2011.05.010.

[226] A. Savio, M. García-Sebastián, M. Graña, and J. Villanúa. Results of an adaboost approach on alzheimer's disease detection on MRI. In *Bioinspired Applications in Artificial and Natural Computation*, pages 114–123, 2009. doi:10.1007/978-3-642-02267-8_13.

[227] H. Muller and T. M. Deserno. Content-based medical image retrieval. *Biomedical Image Processing*, pages 471–494, 2010. doi:10.1007/978-3-642-15816-2_19.

[228] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng. Content-based medical image retrieval : a survey of applications to multidimensional and multimodality data. *Journal of digital imaging*, 26(6) :1025–1039, 2013. doi:10.1007/s10278-013-9619-2.

[229] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner. Open access series of imaging studies (oasis) : cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9) :1498–1507, 2007. doi:10.1162/jocn.2007.19.9.1498.

[230] J J. Talairach and P Tournoux. Co-planar stereotaxic atlas of the human brain.3-dimensional proportional system : An approach to cerebral imaging.

39 :125, 01 1988. Available at : http://www.thieme.com/books/atlas-of-the-human-brain.

[231] S.M. Smith. Fast robust automated brain extraction. *Human brain mapping*, 17(3) :143–155, 2002. doi:10.1002/hbm.10062.

[232] T. Bailey and A.K. Jain. A note on distance-weighted *k*-nearest neighbor rules. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(4) :311–313, 1978. doi:10.1109/tsmc.1978.4309958.

[233] D. Chyzhyk, M. Graña, A. Savio, and J. Maiora. Hybrid dendritic computing with kernel-LICA applied to alzheimer's disease detection in MRI. *Neurocomputing*, 75(1) :72–77, 2012. doi:10.1016/j.neucom.2011.02.024.

[234] N. Villain, B. Desgranges, F. Viader, V. De La Sayette, F. Mézenge, B. Landeau, J.-C. Baron, F. Eustache, and G. Chételat. Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in alzheimer's disease. *Journal of Neuroscience*, 28(24) :6174–6181, 2008. doi:10.1523/jneurosci.1392-08.2008.

[235] M. Allard O. B. Ahmed, J. Benois-Pineau, G. Catheline, and C. B. Amar. Recognition of alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. *Neurocomputing*, 220 :98–110, 2017. doi:10.1016/j.neucom.2016.08.041.

[236] S. Lu, Y. Xia, W. Cai, M. Fulham, D.D. Feng, Alzheimer's Disease Neuroimaging Initiative, et al. Early identification of mild cognitive impairment using incomplete random forest-robust support vector machine and FDG-PET imaging. *Computerized Medical Imaging and Graphics*, 2017. doi:10.1016/j.compmedimag.2017.01.001.

[237] O.B. Ahmed. *Features-based MRI brain classification with domain knowledge : Application to Alzheimer's disease diagnosis*. PhD thesis, University of bordeaux ;

University of Sfax, 2015. Available at : http://www.olfa-benahmed.com/thesis2015.pdf.

[238] O.B. Ahmed, M. Mizotin, J.B. Pineau, M. Allard, G. Catheline, C.B. Amar, Alzheimer's Disease Neuroimaging Initiative, et al. Alzheimer's disease diagnosis on structural MR images using circular harmonic functions descriptors on hippocampus and posterior cingulate cortex. *Computerized Medical Imaging and Graphics*, 44 :13–25, 2015. doi:10.1016/j.compmedimag.2015.04.007.

[239] N. Aggarwal, B. Rana, and R. Agrawal. Detection of alzheimer's disease via statistical features from brain slices. In *The Twenty-Sixth International FLAIRS Conference*, 2013. Available at : https://www.aaai.org/FLAIRS2013.

[240] D. Chyzhyk and A. Savio. Feature extraction from structural MRI images based on VBM : data from OASIS database. *University of the Basque Country, Internal research publication*, 2010. Available at : http://www.ehu.eus/Chyzhyk.pdf.

[241] D. Chyzyk, M. Termenon, and A. Savio. A comparison of VBM results by SPM, ICA and LICA. *Hybrid Artificial Intelligence Systems*, pages 429–435, 2010. doi:10.1007/978-3-642-13803-4_53.

[242] S. Alam, M. Kang, S. Shin, J.Y. Pyun, and G.R. Kwon. Classification of alzheimer's disease using combination of each recognition method in MR scans. In *Biomedical Engineering and Science BIOENG*, 2015. Available at : http://worldcomp-proceedings.com/BIOENG15.pdf.

[243] S. Alam, M. Kang, J.Y. Pyun, and G.R. Kwon. Performance of classification based on PCA, linear SVM, and multi-kernel SVM. In *Ubiquitous and Future Networks (ICUFN), 2016 Eighth International Conference on*, pages 987–989. IEEE, 2016. doi:10.1109/icufn.2016.7536945.

[244] SK. Aruna and S. Chitra. Machine learning approach for identifying dementia from MRI images. *World Academy of Science, Engineering and Technology,*

International Journal of Computer, Electrical, Automation, Control and Information Engineering, 9(3) :881–888, 2015. Available at : https://www.waset.org/Machine_learning_approach.

[245] P. Zhang, G. Wu, Y. Gao, P.T. Yap, and D. Shen. A dynamic tree-based registration could handle possible large deformations among mr brain images. *Computerized Medical Imaging and Graphics*, 52 :1–7, 2016. doi:10.1016/j.compmedimag.2016.04.005.

[246] D.C.T. Nguyen, S. Benameur, M. Mignotte, and F. Lavoie. Superpixel and entropy-based multi-atlas fusion framework for the segmentation of X-ray images. In *18th International Conference on Image Analysis and Processing, ICIAP'15,*, pages 151–161, Genova, Italy, September 2015. doi:10.1007/978-3-319-23234-8_15.

[247] D.R. Martin, C.C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE transactions on pattern analysis and machine intelligence*, 26(5) :530–549, 2004. doi:10.1109/tpami.2004.1273918.

[248] J.M. Górriz M. López, J. Ramírez, D. Salas-González, I.A. Illan, F. Segovia, and C.G. Puntonet. Principal component analysis-based techniques and supervised classification schemes for the early detection of alzheimer's disease. *Neurocomputing*, 74(8) :1260–1271, 2011. doi:10.1016/j.neucom.2010.06.025.

[249] J.M. Górriz L. Khedher, J. Ramírez, A. Brahim, and F. Segovia. Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing*, 151 :139–150, 2015. doi:10.1016/j.neucom.2014.09.072.

[250] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *IEEE Conference on Computer Vision and Pattern Re-*

cognition, CVPR'10, pages 1434–1441, San Francisco, CA, USA, June 2010. doi:10.1109/cvpr.2010.5539802.

[251] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *in European Conference on Computer Vision, ECCV'02*, pages 82–96, Copenhague, Danemark, June 2002. doi:10.1007/3-540-47977-5_6.

[252] G. Vogiatzis, C. Hernández, P. Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12) :2241–2246, December 2007. doi:10.1109/tpami.2007.70712.

[253] C. A. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Processing*, 2(3) :296–310, 1993. doi:10.1109/83.236536.

[254] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60 :259–268, 1992. doi:10.1016/0167-2789(92)90242-f.

[255] R. Touati and M. Mignotte. MDS-based multi-axial dimensionality reduction model for human action recognition. In *Eleventh conference on Computer and Robot Vision, CRV'2014*, Montré al, Quebec, Canada, May 2014. doi:10.1109/crv.2014.42.

[256] S. Alpert, M. Galun, R. Basri, and A. Brandt". "image segmentation by probabilistic bottom-up aggregation and cue integration.". In *"Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition"*, "June" "2007". doi:10.1109/cvpr.2007.383017.

[257] S. Gould, R. Fulton, and D. Koller. Decomposing a scene into geometric and semantically consistent regions. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1–8. IEEE, 2009. doi:10.1109/iccv.2009.5459211.

[258] Jamie Shotton, John Winn, Carsten Rother, and Antonio Criminisi. TextonBoost : Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European conference on computer vision*, pages 1–15. Springer, 2006. doi:10.1007/11744023_1.

Annexe I

Échéancier de la thèse

TABLE I.1 : Échéancier de la thèse

Session	Description
A12	• Réussite des 2 cours obligatoire du prédoc : IFT 2015 (B+) & IFT 2125 (B+)
H13	• Réussite dans le première cours gradué : IFT 6141 (A) • Réussite dans le deuxième cours gradué : 6271 (B+)
É13	• Préparation de l'examen pré-doctoral écrit
A13	• Réussite dans l'examen pré-doctoral écrit (A+)
H14	• Activités de recherche & expérimentation
É14	• Rédaction du rapport pour l'examen oral (présentation du projet de recherche)
A14-H15	• Étude expérimentale de la partie 1. [Consensus de segmentation dans le cadre de la visualisation, navigation et la recherche dans les grandes bases de données d'images] • Rédaction des articles associés à cette partie
É15-H16	• Étude expérimentale de la partie 2. [Consensus de segmentation pour améliorer la précision des algorithmes de détection de mouvement] • Rédaction des articles associés à cette partie
É16-É17	• Étude expérimentale de la partie 3. [Consensus de segmentation pour la classification d'imagerie médicale : application au diagnostic de la maladie d'Alzheimer] • Rédaction des articles associés à cette partie
A17	• Rédaction de la thèse
H18	• Présentation de la thèse

Annexe II

Suppléments de résultats expérimentaux à la partie 1

Nous présentons ici les résultats de notre algorithme de visualisation appliqué sur les trois bases de données suivantes : Weizmann200 [256], Stanford715 [257] et Msrc590 [258].

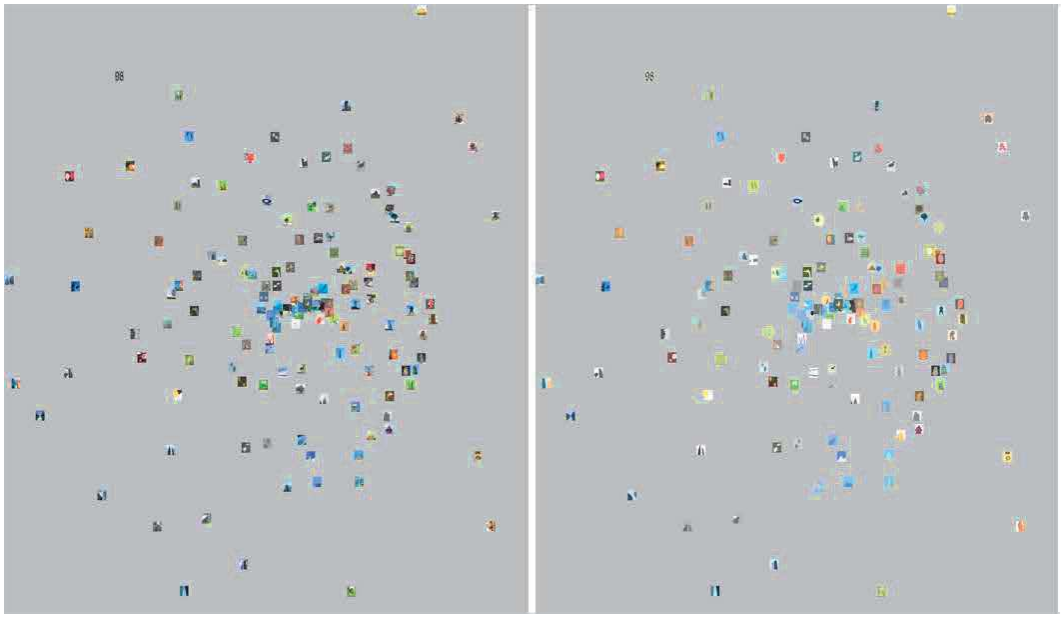


FIGURE II.1 : A MDS visualization map of the Weizmann200 based on the VoI distance between segmentations. Left : view map of the (200) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense.

FIGURE II.2 : Magnified details of Fig. II.1 showing the grouping of images of the Weizmann200 based on the similar layout and arrangement of the objects detected and segmented in the scene.

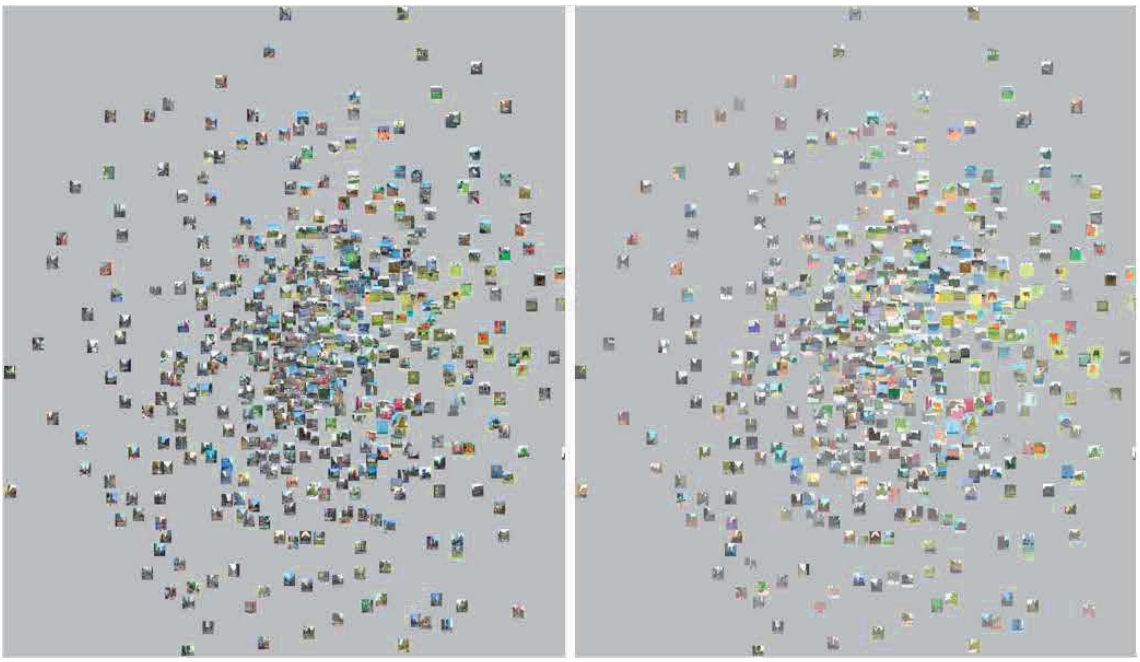


FIGURE II.3 : A MDS visualization map of the Stanford715 based on the VoI distance between segmentations. Left : view map of the (715) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense.

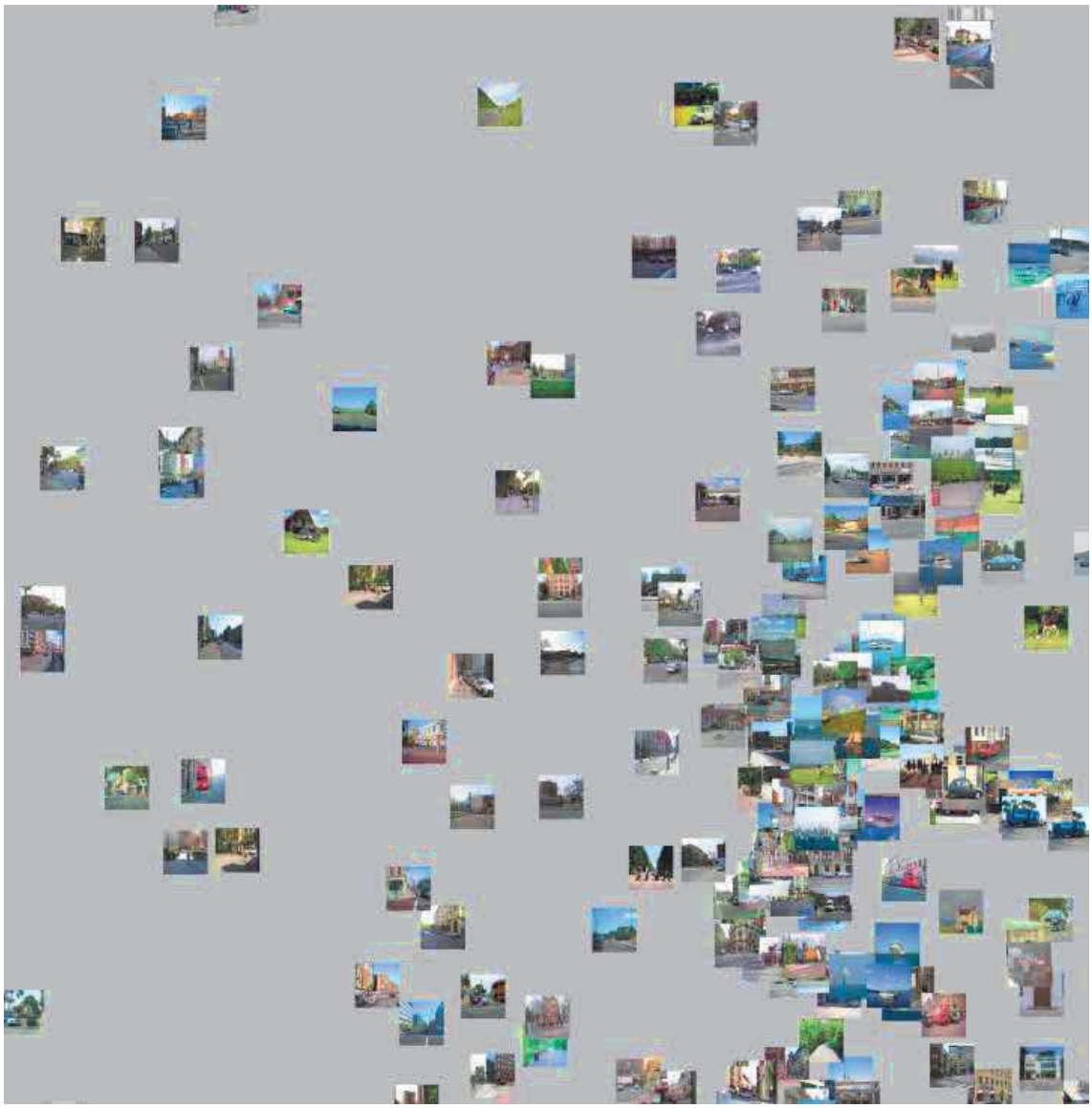


FIGURE II.4 : Magnified details of Fig. II.3 showing the grouping of images of the Stanford715 based on the similar layout and arrangement of the objects detected and segmented in the scene.

FIGURE II.5 : A MDS visualization map of the Msrc590 based on the VoI distance between segmentations. Left : view map of the (590) color images mapped according to their similarity in term of region-based segmentation result. Right : view map of the segmentation results according to their similarity in the VoI distance sense.

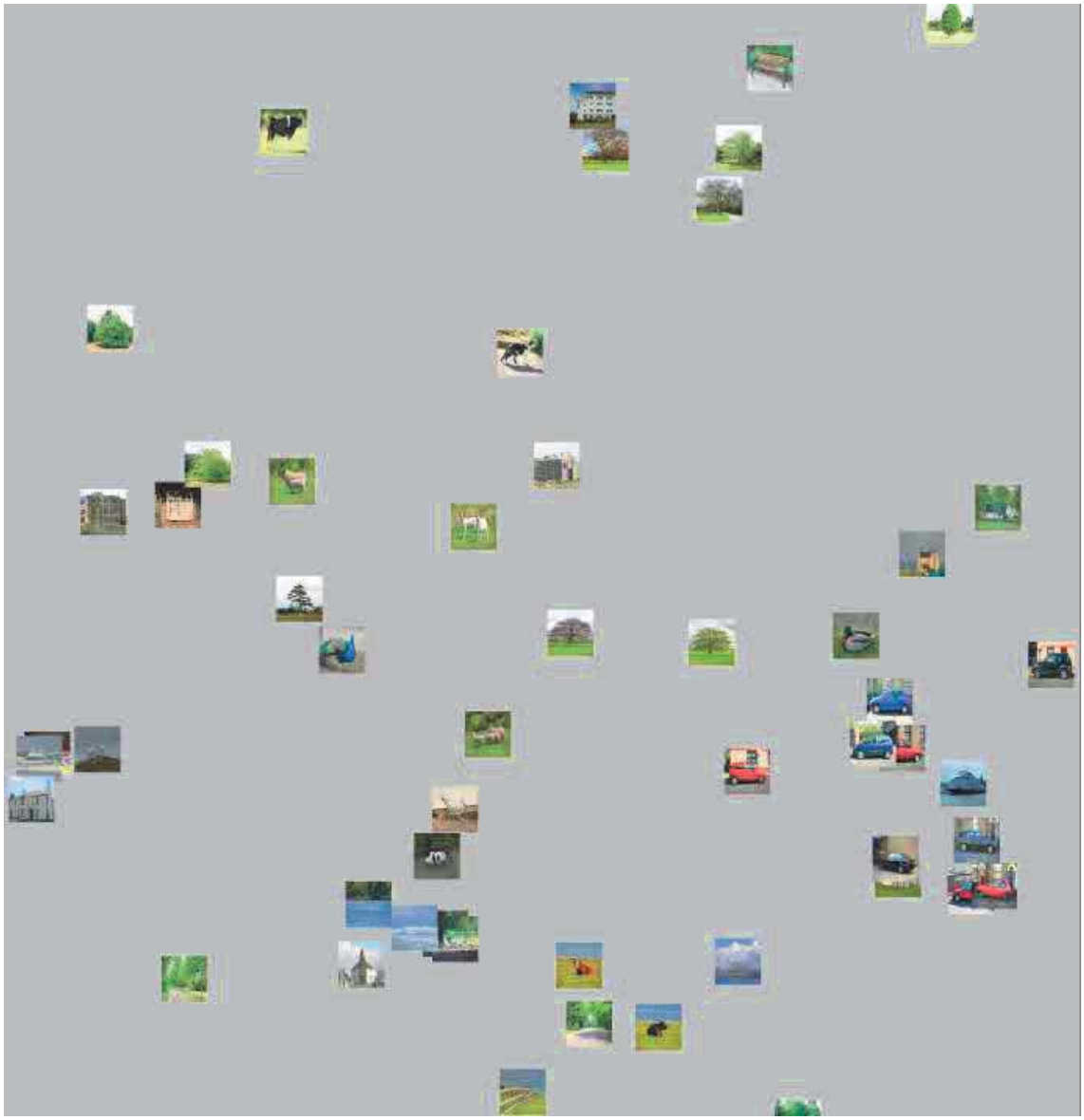


FIGURE II.6 : Magnified details of Fig. II.5 showing the grouping of images of the Msrc590 based on the similar layout and arrangement of the objects detected and segmented in the scene.