Université de Montréal

Contributions à la fusion de segmentations et à l'interprétation sémantique d'images

par

Lazhar Khelifi

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

août, 2017

RÉSUMÉ

Cette thèse est consacrée à l'étude de deux problèmes complémentaires, soit la fusion de segmentation d'images et l'interprétation sémantique d'images. En effet, dans un premier temps, nous proposons un ensemble d'outils algorithmiques permettant d'améliorer le résultat final de l'opération de la fusion. La segmentation d'images est une étape de prétraitement fréquente visant à simplifier la représentation d'une image par un ensemble de régions significatives et spatialement cohérentes (également connu sous le nom de « segments » ou « superpixels ») possédant des attributs similaires (tels que des parties cohérentes des objets ou de l'arrière-plan). À cette fin, nous proposons une nouvelle méthode de fusion de segmentation au sens du critère de l'Erreur de la Cohérence Globale (GCE), une métrique de perception intéressante qui considère la nature multi-échelle de toute segmentation de l'image en évaluant dans quelle mesure une carte de segmentation peut constituer un raffinement d'une autre segmentation. Dans un deuxième temps, nous présentons deux nouvelles approches pour la fusion des segmentations au sens de plusieurs critères en nous basant sur un concept très important de l'optimisation combinatoire, soit l'optimisation multi-objectif. En effet, cette méthode de résolution qui cherche à optimiser plusieurs objectifs concurremment a rencontré un vif succès dans divers domaines. Dans un troisième temps, afin de mieux comprendre automatiquement les différentes classes d'une image segmentée, nous proposons une approche nouvelle et robuste basée sur un modèle à base d'énergie qui permet d'inférer les classes les plus probables en utilisant un ensemble de segmentations proches (au sens d'un certain critère) issues d'une base d'apprentissage (avec des classes pré-interprétées) et une série de termes (d'énergie) de vraisemblance sémantique.

Mots clefs : Ensemble de segmentation, fusion, erreur de la cohérence globale (GCE), modèle de vraisemblance pénalisée, optimisation multi-objectif, prise de décision, segmentation sémantique d'image.

ABSTRACT

This thesis is dedicated to study two complementary problems, namely the fusion of image segmentation and the semantic interpretation of images. Indeed, at first we propose a set of algorithmic tools to improve the final result of the operation of the fusion. Image segmentation is a common preprocessing step which aims to simplify the image representation into significant and spatially coherent regions (also known as *segments* or *super-pixels*) with similar attributes (such as coherent parts of objects or the background). To this end, we propose a new fusion method of segmentation in the sense of the Global consistency error (GCE) criterion. GCE is an interesting metric of perception that takes into account the multiscale nature of any segmentations of the image while measuring the extent to which one segmentation map can be viewed as a refinement of another segmentation. Secondly, we present two new approaches for merging multiple segmentations within the framework of multiple criteria based on a very important concept of combinatorial optimization ; the multi-objective optimization. Indeed, this method of resolution which aims to optimize several objectives concurrently has met with great success in many other fields. Thirdly, to better and automatically understand the various classes of a segmented image we propose an original and reliable approach based on an energy-based model which allows us to deduce the most likely classes by using a set of identically partitioned segmentations (in the sense of a certain criterion) extracted from a learning database (with pre-interpreted classes) and a set of semantic likelihood (energy) terms.

Key words : Segmentation ensemble, fusion, global consistency error (GCE), penalized likelihood model, multi-objective optimization, decision making, semantic image segmentation.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES APPENDICES

LISTE DES SIGLES

ACA	Average per-Class Accuracy
ASD	Aerial image Segmentation Dataset
BCE	Bidirectional Consistency Error
BDE	Boundary Displacement Error
BSD	Berkeley Segmentation Dataset
CPU	Central Processing Unit
CRF	Conditional Random Field
CNNs	Convolutional Neural Networks
DMD	Multiplicative and Difference Model
EFA-BMFM	Multi-criteria Fusion Model Based on the Entropy-Weighted Formula Approach
EM	Expectation Maximization
ESE	Exploration/Selection/Estimation
FGM	Fast Global Minimization
FS	Feasible Solutions
GCE	Global Consistency Error
GCEBFM	Global Consistency Error Based Fusion Model
GPA	Global per-Pixel Accuracy
GPU	Graphic Processor Unit
HOG	Histogram of Oriented Gradients
ICM	Iterative Conditional Modes
KNN	K-Nearest Neighbor
LAP	Laplacian Operator

LBP	Local Binary Pattern
LMO	LabelMe Outdoor
LSD	Line Segment Detector
LRE	Local Refinement Error
MCDM	Multi-Criteria Decision Making
MC-SSM	Multi-Criteria Semantic Segmentation Model
ML	Maximum Likelihood
MO	Multi-objective Optimization
MOBFM	Multi-objective Optimization Based Fusion Model
MRF	Markov Random Fields
MRI	Magnetic Resonance Imaging
MSRC	Microsoft Research Cambridge Dataset
OCLBP	Opponent Color Local Binary Pattern
ODS	Optimal Data set Scale
OIS	Optimal Image Scale
PRI	Probabilistic Rand Index
PTA	Pareto Approach
RI	Rand Index
RLSF	Region-based model via Local Similarity Factor
SA	Simulated Annealing
SBD	Stanford Background Dataset
SIFT	Scale-Invariant Feature Transform
TOPSIS	Technique for Order Performance by Similarity to Ideal Solution
VoI	Variation of Information
WFA	Weighted Formula Approach

Je dédie cette thèse à :

Ma mère.
Pour son appui indéfectible durant mes études.
Je n'oublie pas ses énormes sacrifices.

L'âme de mon cher père.

A mes frères.

A mes soeurs.

A 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 tiens à exprimer en tout premier lieu mon immense gratitude à mon directeur de thèse le professeur Max Mignotte, d'avoir accepté de diriger mes travaux de recherche, de faire confiance à mes compétences et de m'offrir une grande autonomie. Je le remercie aussi pour son encadrement et pour son expertise dans le domaine segmentation d'images, ainsi que pour la grande disponibilité dont il a fait preuve 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.

Finalement, je remercie tous les professeurs qui ont contribué à ma formation universitaire.

CHAPITRE 1

INTRODUCTION

1.1 Contexte de recherche

La vision par ordinateur est une branche de l'intelligence artificielle qui permet à une machine de comprendre ce qu'elle « voit » lorsqu'on la connecte à une ou plusieurs caméras. En d'autres termes, c'est un traitement automatisé des informations visuelles par ordinateur. Cette discipline scientifique étant très vaste, elle englobe d'autres sous-domaines tels que le traitement d'images qui est une discipline riche et qui donne lieu à une profusion de travaux académiques et industriels chaque année. En effet, les connaissances en la matière s'appliquent de nos jours dans plusieurs contextes comme la retouche d'images, la reconnaissance faciale, l'analyse de scènes routières, l'imagerie multi-spectrale, la reconnaissance de l'écriture, l'imagerie médicale, etc. Cette richesse s'explique par l'importance de l'analyse, l'extraction de l'information et la compréhension de l'image. À cet égard, plusieurs techniques et méthodes ont été proposées afin de trouver les solutions adéquates pour résoudre les problèmes qui se présentent pendant les différentes phases de traitement de l'image :

- La phase de prétraitement (traitements photométriques et colorimétriques, réduction de bruit, restauration d'images, etc.), qui permet une meilleure visualisation de l'image, facilitant ainsi les traitements ultérieurs ;
- La phase de segmentation, qui consiste à partitionner l'image en un ensemble de régions connexes et cohérentes ;
- La phase de quantification (description de forme, caractéristiques géométriques d'un objet, etc.), qui a pour but de fournir des indices quantitatifs ou géométriques.

Dans cette thèse, nous nous intéressons, dans un premier temps, à la phase de segmentation. En effet, la segmentation d'image est une étape primordiale qui consiste à

regrouper les pixels de l'image en différentes régions selon des critères de ressemblance prédéfinis (il peut s'agir, par exemple, de séparer les objets du fond). Cette opération dite de bas niveau permet d'obtenir une représentation simplifiée de l'image. Elle n'est pas considérée comme un but, mais comme un moyen efficace qui permet ensuite d'effectuer des tâches de plus haut niveau visant à analyser le contenu de l'image.

La résolution de problèmes de segmentation d'images nécessite l'implémentation d'un algorithme qui permet de diviser l'image en zones de régions homogènes. Cependant, les expériences en segmentation nous ont montré qu'il est difficile d'obtenir un tel résultat en utilisant un algorithme classique de segmentation. À cette fin, au lieu de concevoir un algorithme de segmentation très compliqué, nous proposons dans ce travail une autre méthodologie qui consiste à segmenter l'image avec des algorithmes très simples, mais très différents, puis à fusionner les résultats (ou cartes de segmentation) à l'aide d'une procédure de fusion calculant une sorte de moyennage de segmentation pour générer une segmentation finale plus robuste. Suivant cette stratégie, nous proposons deux modèles de fusion de segmentation d'image, soit le modèle mono-objectif, basé sur un seul critère, et le modèle multi-objectif, basé sur différents critères et sur le concept de l'optimisation multi-objectif.

Notre démarche s'inspirant de la logique et de la perception humaine, nous nous penchons dans un deuxième temps sur un autre problème, soit l'interprétation sémantique d'images. À cet égard, nous présentons un nouveau système permettant d'identifier automatiquement les différentes régions d'une image segmentée.

1.2 La segmentation d'images

1.2.1 Définition

La segmentation d'images est une étape de prétraitement fréquente visant à simplifier la représentation d'image par un ensemble de régions significatives et spatialement cohérentes (aussi appelées « superpixels ») possédant des attributs similaires (tels que des parties cohérentes d'un même objet ou de l'arrière-plan). Cette tâche de vision de

bas niveau, qui modifie la représentation d'une image en quelque chose de plus facile à analyser, est souvent l'étape préliminaire et également critique dans le développement de nombreux algorithmes de compréhension de l'image et des systèmes de vision par ordinateur tels que les problèmes de reconstruction [21] ou la localisation/reconnaissance d'objet 3D [22, 23].

La segmentation consiste à partitionner une image I en n régions différentes $R_1, ..., R_n$. Les régions obtenues doivent respecter les propriétés d'homogénéité. Mathématiquement, soit $P(R_i)$ le prédicat logique qui définit l'homogénéité d'une région R_i. Ce prédicat est défini formellement par l'équation suivante :

$$P(R_i) = \begin{cases} vrai \quad si \quad R_i \quad est \quad homogéne \\ faux \quad sinon \end{cases} \tag{1.1}$$

Pour valider un résultat de segmentation, les régions générées par un algorithme doivent respecter les conditions suivantes [1] :

- Recouvrement : chaque pixel de l'image doit appartenir à une région R_i et l'union de toutes les régions correspond à l'image entière

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

- Connexité : les pixels qui appartiennent a une région doivent être connectés, plus précisément pour toute paire de pixels p et q d'une région R_i , il est possible de tracer un chemin de p vers q en ne passant que par des pixels de la région R_i [24]

$$R_i \text{ forme un ensemble connexe } \forall i = 1, 2, ...n.$$

- Disjonction : aucun pixel ne fait partie de deux régions différentes à la fois

$$R_i \bigcap R_j = \varnothing \ \forall i, j \,|\, i \neq j.$$

- Satisfiabilité : chaque région doit satisfaire un prédicat d'homogénéité P

$$P(R_i) = VRAI \ \ \forall i = 1, 2, ...n.$$

- Segmentabilité : un même prédicat ne se réalise pas pour l'union de deux régions adjacentes

$$P(R_i \bigcup R_j) = FAUX \ \forall i, j \,|\, i \neq j \ \text{ et } R_i, R_j \text{ étant adjacents dans } I.$$

D'un point de vue algorithmique, une région est un groupe de pixels connectés entre eux avec des propriétés similaires, par contre, une classe est un ensemble de pixels qui possèdent des caractéristiques texturales similaires, la figure 1.1 montre la différence entre ces deux notions.

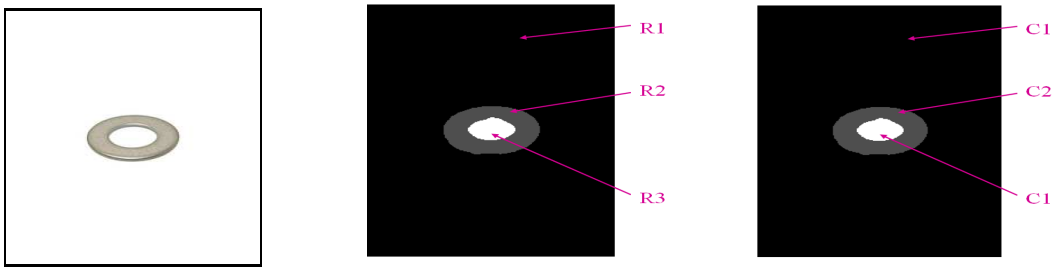


FIGURE 1.1 : De gauche à droite ; une image couleur, sa segmentation en régions (R_1, R_2 et R_3) et sa représentation en classes (c_1 : arrière-plan et c_2 : rondelle).

1.2.2 Stratégies de segmentation d'images

Une pléthore de méthodes de segmentation basées sur les régions a été proposée afin de résoudre le problème difficile de la segmentation non supervisée d'images naturelles texturées. La plupart de ces méthodes exploitent une première étape d'extraction de paramètres, pour caractériser chaque région texturée significative à segmenter, suivie d'une technique de classification, qui permet de regrouper selon des critères ou des stratégies différentes des régions spatialement cohérentes partageant des attributs similaires. Pendant des années, les recherches en segmentation se sont concentrées sur des caractéristiques plus sophistiquées d'extraction de caractéristiques et des techniques de classification plus élaborées. Ces travaux ont amélioré de façon significative les résultats finaux de segmentation, mais ont généralement augmenté la complexité du modèle et/ou de calcul. Ces méthodes comprennent des modèles de segmentation qui exploitent

directement des systèmes de regroupement (« clustering ») [2, 3, 19, 25] en utilisant la modélisation par mélange de gaussiennes [26], l'approche de classification floue [27,28], les ensembles flous [29] ou, après une approche de dé-texturation [3, 4, 6]), le « mean-shift » ou plus généralement des procédures basées sur la recherche des modes d'une distribution [30], les méthodes de ligne de partage d'eaux [31] ou les stratégies de croissance de la région [32], les modèles de codage et de compression avec perte [31, 33], la transformée en ondelettes [34], les champs aléatoires de Markov (MRF) [35–37], l'approche Bayésienne [38], l'approche basée sur le texton [39] ou les modèles basés sur le graphe [12,40,41], les méthodes variationnelles ou de l'ensemble du niveau [39,42–45], les modèles de surfaces déformables [46], de contour actif [47] (avec approche basée sur le partitionnement de graphe [48]) ou les techniques basées sur les courbes [49], la technique de seuillage non supervisée itérative [50, 51], l'algorithme génétique [52], les cartes auto-organisatrices [53], la technique de l'apprentissage de variétés [54], l'approche basée sur la topologie [55], les objets symboliques [56] et la classification spectrale [57], etc. pour en citer que quelques-uns.

1.3 La fusion de segmentation d'images

Une variante récente et efficace de segmentation consiste à combiner ou fusionner plusieurs cartes de segmentation grossièrement et rapidement estimées de la même scène et associée à un modèle de segmentation [1] simple, pour obtenir une segmentation finale améliorée. Au lieu de chercher le meilleur algorithme de segmentation avec ses paramètres internes optimaux, ce qui est difficile si l'on tient compte des différents types d'images existantes, cette stratégie privilégie la recherche d'un modèle de fusion de segmentations, ou plus précisément, la recherche du critère le plus efficace pour fusionner de multiples segmentations.

[1] Ces cartes de segmentations destinées à être fusionnées peuvent être générées par différents algorithmes (idéalement complémentaires) ou par le même algorithme ayant différentes valeurs des paramètres internes ou graines (pour les méthodes stochastiques), ou en utilisant des caractéristiques texturales différentes et appliquées à une image d'entrée éventuellement exprimée dans différents espaces de couleurs ou transformations géométriques (par exemple, facteur d'échelle, inclinaison, etc.) ou par d'autres moyens.

La combinaison de plusieurs segmentations peut constituer un cas particulier du *problème d'ensemble de classifieurs*, c'est-à-dire le concept qui combine plusieurs méthodes de classification pour améliorer le résultat final de classification (et qui fut d'abord exploré dans le domaine de l'apprentissage machine [58–60]). En effet, l'ordonnancement spatial est un aspect distinctif des données d'une image et la segmentation d'images est donc un processus de regroupement des données spatialement indexées. Par conséquent, le groupement des pixels doit non seulement tenir compte de la similitude de leur caractéristique (couleur, texture, etc.), mais aussi de leur cohérence spatiale. Il est intéressant de noter que ce problème de fusion de segmentation ou segmentation d'ensemble peut également être considéré comme étant un cas particulier d'un problème de débruitage dans lequel chaque segmentation à fusionner est en fait une solution bruitée ou une observation. L'objectif final est donc de trouver une solution de segmentation débruitée, qui serait en fait un consensus ou un compromis (en termes de clusters, de niveau de détails, de précision de contour, etc.) de toutes les segmentations. En un sens, la segmentation finale fusionnée représente la moyenne de toutes les segmentations individuelles à combiner selon un critère bien défini. Quand cette stratégie a d'abord été introduite en [61] [62], toutes les segmentations à fusionner devaient contenir le même nombre de régions. Un peu plus tard, cette stratégie fut utilisée sans cette restriction, avec un nombre arbitraire de régions [2, 63]. Depuis ces travaux novateurs, cette fusion de multiples segmentations de la même scène, pour obtenir un résultat de segmentation plus fiable et précis, est maintenant effectuée selon plusieurs stratégies et/ou des critères bien définis (Figure 1.2).

1.4 L'optimisation multi-objectif

Le problème de segmentation d'image est souvent formalisé sous la forme d'un problème d'optimisation. Un problème d'optimisation est défini, généralement, par un espace de recherche S et une fonction objectif f. Le but est de trouver la solution de meilleure qualité. Suivant le problème posé, nous cherchons soit le minimum soit le maximum de la fonction f [64]. Formellement, un problème d'optimisation peut être

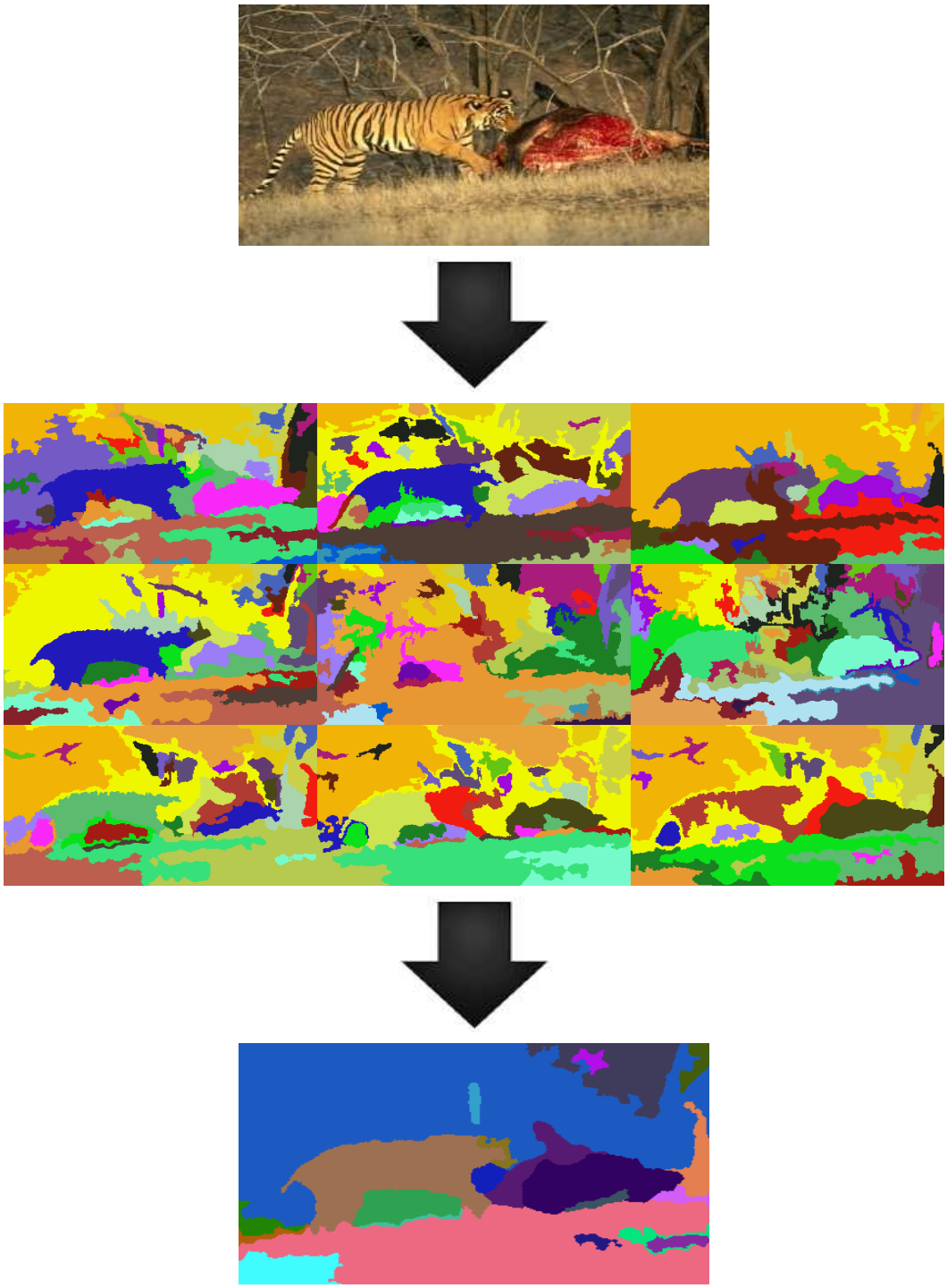


FIGURE 1.2 : Fusion de segmentations.

représenté de la manière suivante :

$$\left.\begin{array}{lll} \min & f(\overrightarrow{x}) & \text{(function à optimiser)} \\ \text{avec} & \overrightarrow{g}(\overrightarrow{x}) \leqslant 0 & \text{m contraintes d'inégalités} \\ \text{et} & \overrightarrow{h}(\overrightarrow{x}) = 0 & \text{p contraintes d'égalités} \end{array}\right\} \tag{1.2}$$

où $\overrightarrow{x} \in \Re^n, \overrightarrow{g}(\overrightarrow{x}) \in \Re^m, \overrightarrow{h}(\overrightarrow{x}) \in \Re^p$. Les vecteurs $\overrightarrow{g}(\overrightarrow{x})$ et $\overrightarrow{h}(\overrightarrow{x})$ représentent respectivement m contraintes d'inégalité et p contraintes d'égalité. Cet ensemble de contraintes permet de délimiter un espace restreint de recherche de la solution optimale pour un certain problème. L'optimisation mono-objectif consiste à maximiser (ou minimiser) une seule fonction objective par rapport à un ensemble de paramètres. Cependant, dans le cas multi-objectif, on cherche à satisfaire plusieurs objectifs souvent contradictoires devant être simultanément maximisés ou minimisés . Par conséquent, l'augmentation d'un objectif entraîne une diminution de l'autre objectif. Mathématiquement, dans le cas de la minimisation le problème s'écrit de la manière suivante :

$$\left.\begin{array}{lll} \min & \overrightarrow{f}(\overrightarrow{x}) & \text{(k function à optimiser)} \\ \text{avec} & \overrightarrow{g}(\overrightarrow{x}) \leqslant 0 & \text{m contraintes d'inégalités} \\ \text{et} & \overrightarrow{h}(\overrightarrow{x}) = 0 & \text{p contraintes d'égalités} \end{array}\right\} \tag{1.3}$$

où $\overrightarrow{x} \in \Re^n, \overrightarrow{f}(\overrightarrow{x}) \in \Re^k, \overrightarrow{g}(\overrightarrow{x}) \in \Re^m, \overrightarrow{h}(\overrightarrow{x}) \in \Re^p$ et f représente un vecteur qui regroupe k fonctions objectif.

1.5 L'interprétation sémantique d'images segmentées

L'interprétation sémantique d'images segmentées, également appelée la classification d'objets visuels, vise à diviser et étiqueter l'image en régions sémantiques ou objets, par exemple ; *montagne*, *ciel*, *bâtiment*, *arbre*, *etc*. Bien que cette tache soit triviale pour un être humain, elle est considérée comme l'un des problèmes les plus difficiles dans le domaine de la vision par ordinateur. Une des raisons de cette difficulté vient du fait que certains défis importants doivent être pris en compte afin d'avoir un bon résultat d'étiquetage, tels que ; la variation de point de vue, la variation d'échelle, la déforma-

tion, l'occultation, les conditions d'éclairage, la confusion d'arrière-plan et la variation intra-classe [2] (voir Figure 1.3).

1.6 Contributions

1.7 Contributions

Le but de cette thèse est l'étude de deux problèmes complémentaires, soit la segmentation (en régions) et l'interprétation sémantique d'images. La nature mal posée de ces deux problèmes et la proposition de nouveaux modèles non-paramétriques de minimisation d'énergie à base de fusion rendent ce travail distinct de la majorité des méthodes qui ont utilisé des approches purement paramétriques ou basé sur l'apprentissage machine. Le travail réalisé dans cette thèse se divise essentiellement en trois parties :

Fusion de segmentations mono-objectif :

L'approche de fusion de différentes segmentations d'une même scène afin d'obtenir un résultat de segmentation plus précis a été proposée récemment selon plusieurs stratégies ou critères. Nous pouvons mentionner le modèle de fusion introduit dans [2] qui fusionne un ensemble de segmentations en minimisant la dispersion (ou l'inertie) des étiquettes obtenues localement autour de chaque pixel de l'image en exécutant simplement une procédure de fusion à base de l'algorithme des *k-moyennes*. De la même manière, on peut également citer le modèle proposé dans [72] qui suit la même idée, mais au sens de l'inertie pondérée en exploitant cette fois l'algorithme des *k-moyennes* flou. Cette fusion de segmentations a également été réalisée en utilisant la version probabiliste du critère *Rand* (PRI) [70] grâce à une procédire de fusion basé sur un modèle Markovien permettant d'estimer la segmentation maximisant la compatibilité, des étiquettes, au sens de chaque paire de pixels, avec l'ensemble de segmentations à fusionner. De même, la combinaison de cartes de segmentation a été effectuée selon le critère de variation d'information (VoI) dans [76] en exploitant un modèle à base d'énergie et en appliquant une méthode de descente du gradient combinée avec des contraintes de

[2] http ://cs231n.github.io/classification/ (Vu le 15/05/2017).

cohérence spatiale. La fusion des segmentations a aussi été réalisée au sens de *l'accumulation de l'évidence* [59] *via* une stratégie de partitionnement hiérarchique, ou au sens de la précision et du rappel (F-mesure) [77] avec un modèle de minimisation d'énergie. Finalement, nous pouvons citer le modèle de fusion de segmentation d'image qui se base sur des méthodes de regroupement d'ensembles proposées dans [80], et l'approche présentée dans [81] basée sur un algorithme de consensus de regroupement, minimisant une fonction de distance avec une descente de gradient stochastique.

Dans ce travail nous présentons un nouveau modèle mono-objectif de fusion de segmentation basé sur le critère de l'erreur de la cohérence globale (GCE). Le GCE est une métrique de perception intéressante qui considère la nature intrinsèque multi-échelle de toute segmentation d'image en évaluant dans quelle mesure une carte de segmentation peut constituer un raffinement d'une autre segmentation. De plus, nous avons ajouté à ce modèle un terme de régularisation *a priori* permettant d'intégrer des connaissances sur la solution de segmentation (et définis *a priori* comme étant des solutions acceptables). Cette stratégie nous permet habilement d'adapter notre modèle avec la nature mal posée du problème de la segmentation.

Fusion de segmentations multi-objectif :

Comme mentionné ci-dessus, la résolution du problème de la fusion de segmentations est généralement basée sur l'optimisation d'un seul critère. Suivant cette stratégie, un seul critère ne peut pas modéliser toutes les propriétés géométriques ou statistiques d'une segmentation. Avec un seul critère, la procédure de fusion est intrinsèquement biaisée vers la recherche d'un ensemble particulier de solutions possibles (considérées comme acceptables) et ce choix mono-critère restreint l'exploration de certaines régions spécifiques de l'espace de recherche contenant les solutions à certaines zones où sont censées exister les solutions définies comme étant acceptables par ce seul critère. Cette stratégie peut limiter et biaiser la performance des modèles de fusion de segmentations. Pour éviter cet inconvénient, c'est-à-dire le biais inhérent causé par l'utilisation d'un seul critère, nous proposons une nouvelle approche pour la fusion des segmentations au sens de plusieurs critères basés sur un concept très important de l'optimisation combinatoire, soit l'optimisation multi-objectif. En effet, cette méthode de résolution, qui cherche à

optimiser plusieurs objectifs concurremment, a rencontré un vif succès dans divers domaines. De même, notre objectif est de concevoir de nouveaux modèles de fusion de segmentations qui profitent de la complémentarité de différents objectifs (critères), et qui permettent finalement d'obtenir un meilleur résultat de segmentation par consensus. Dans le cadre de cette nouvelle stratégie, nous introduisons, dans un premier temps, un nouveau modèle de fusion multicritères pondéré par une mesure basée sur l'entropie (EFA-BMFM). L'objectif principal de ce modèle est de combiner et d'optimiser simultanément deux critères de fusion de segmentation différents et complémentaires, à savoir le critère VoI (basé sur la région) et le critère F-measure (basé sur le contour) dérivé du rappel-précision. Dans un deuxième temps, afin de combiner et d'optimiser efficacement deux critères de segmentation complémentaires (l'erreur de la cohérence globale (GCE) et le critère du F-measure) nous intégrons le concept de dominance dans notre cadre de fusion. À cette fin, nous présentons une méthode hiérarchique et efficace pour optimiser la fonction d'énergie multi-objectif liée à ce modèle de fusion qui exploite une stratégie d'optimisation itérative, simple et déterministe combinant les différents segments d'image. Cette étape est suivie d'une tâche de prise de décision basée sur la technique de la performance de l'ordre par similarité à la solution idéale (TOPSIS).

Interprétation sémantique d'images :

Les méthodes d'interprétation sémantique d'images qui ont été proposées dans la littérature se divisent en trois catégories. La première est l'approche paramétrique qui utilise les techniques d'apprentissage automatique pour apprendre des modèles paramétriques en utilisant les catégories d'intérêt dans l'image. Selon cette stratégie il faut apprendre des classifieurs paramétriques pour reconnaître des objets (par exemple, bâtiment, vache ou ciel) [150]. Dans ce contexte, nous pouvons citer les techniques d'apprentissage profond [151] qui sont basées sur les réseaux de neurones convolutifs (CNN) [149] telles que ; FCN [152], R-CNN [153], SDS [155], DeepLab [156], multiscale net [157], les techniques par les machines à vecteurs de support [158] [159], et les forêts d'arbres décisionnels (ou forêts aléatoires) ; tels que OCS-RF [160] et Geof [161]. La deuxième est l'approche non paramétrique qui vise à étiqueter l'image d'entrée en faisant correspondre des parties d'images à des parties similaires dans une base d'images

étiquetée. Ici, l'apprentissage des classifieurs de catégories est remplacé en général par un champ aléatoire de *Markov* dans lequel les potentiels unaires sont calculés par la méthode de plus proche voisin [150]. Dans la troisième catégorie, le modèle non paramétrique est intégré avec le modèle paramétrique [167], dans ce contexte, pour tirer parti des avantages des deux méthodologies une méthode quasi paramétrique (hybride) qui intègre une méthode basée sur l'algorithme k plus proche voisin (KNN) et une méthode basée sur le CNN, a été proposée dans [168].

Bien que, récemment, l'approche paramétrique par apprentissage machine a connu un grand succès, toutes ces méthodes ont certaines limites en termes de temps d'apprentissage. Une autre source de problèmes vient du nombre d'objets à étiqueter. Ce nombre d'objets est réellement illimité dans le monde réel, ainsi une tâche de mise à jour est nécessaire pour adapter le modèle à un nouveau jeu de données d'apprentissage. Dans ce travail, nous suivrons une approche non paramétrique mais sans avoir recours à l'apprentissage machine et donc sans étape préalable d'apprentissage. Nous proposons un modèle de segmentation sémantique multicritères basé sur une minimisation d'une fonction d'énergie (MC-SSM). L'objectif principal de ce nouveau modèle est de prendre en avantages la complémentarité de différents critères ou caractéristiques. Ainsi, le modèle proposé combine efficacement différents termes de la vraisemblance globale, et exploite une base d'apprentissage d'image segmentée et pré-interprétée. Afin d'optimiser notre modèle énergétique, nous utilisons une simple procédure d'optimisation locale.

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 :

Le *Chapitre 2* présente notre première contribution avec un article portant sur la fusion de segmentation mono-objectif. Ce chapitre propose une nouvelle méthode de fusion de segmentation au sens du critère GCE (Erreur de Cohérence Globale). Cette

métrique de perception considère la nature multi-échelle de toute segmentation d'image en évaluant à quelle distance une carte de segmentation peut être considérée comme le raffinement d'une autre segmentation. De plus, afin de gérer la nature mal posée du problème de segmentation, nous ajoutons à ce modèle de fusion, un terme de régularisation permettant d'intégrer des connaissances sur le type de fusion de segmentation, défini *a priori* comme solutions acceptables.

Partie 2 :

Le modèle mono-objectif présenté au *Chapitre 2* soulève la nécessité de mettre en oeuvre des stratégies permettant d'effectuer le processus de fusion de segmentation au sens de différents critères en nous basant sur un concept très important issu du domaine de la recherche opérationnelle ; l'optimisation multi-critère ou multi-objectif. À cet égard, dans un premier temps, le *Chapitre 3* présente un modèle de fusion basé sur deux critères contradictoires et complémentaires (à base de région et contour) de segmentation, et une approche de résolution basée sur la méthode de pondération des fonctions objectives. Dans l'étape suivante, le *Chapitre 4* présente un deuxième modèle de fusion de segmentations multi-objectif basé sur approche Pareto. Une méthode efficace de prise de décision est utilisée pour choisir la solution finale qui résulte de notre modèle de fusion.

Partie 3 :

Le *Chapitre 5* présente notre quatrième contribution avec un article portant sur la segmentation sémantique d'image. À cette fin, nous proposons un nouveau système automatique d'étiquetage sémantique exploitant une base d'apprentissage d'image segmentée et pré-interprétée, et nous proposons un modèle à base d'énergie permettant d'inférer les classes les plus probables en nous basant sur les k segmentations les plus proches au sens du critère de l'Erreur de Cohérence Globale et minimisant la somme de différents termes de vraisemblances sémantiques utilisant différents critères.

1.8.2 Publications

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

- Travaux sur la fusion de segmentation mono-objectif

  - **L. Khelifi**, M. Mignotte. A novel fusion approach based on the global consistency criterion to fusing multiple segmentations. *IEEE Transactions on Systems, Man, and Cybernetics : Systems (TSMC)*, 47 (9) : 2489-2502, Septembre 2017.

    $\Rightarrow$ Article présenté dans le *Chapitre 2*.

  - **L. Khelifi**, M. Mignotte. GCE-based model for the fusion of multiples color image segmentations. *23rd IEEE International Conference on Image Processing (ICIP)*, pages 2574-2578, Phoenix, Arizona, USA, Septembre 2016.

- Travaux sur la fusion de segmentations multi-objectif

  - **L. Khelifi**, M. Mignotte. EFA-BMFM : A multi-criteria framework for the fusion of colour image segmentation. Information Fusion (IF), Elsevier, 38 : 104-121, Novembre 2017.

    $\Rightarrow$ Article présenté dans le *Chapitre 3*.

  - **L. Khelifi**, M. Mignotte. A new multi-criteria fusion model for color textured image segmentation. *23rd IEEE International Conference on Image Processing (ICIP)*, pages 2579-2583, Phoenix, Arizona, USA, Septembre 2016.

  - **L. Khelifi**, M. Mignotte. A Multi-objective decision making approach for solving the image segmentation fusion problem. *IEEE Transactions on Image Processing (TIP)*, 26 (8) : 3831-3845, Août 2017.

    $\Rightarrow$ Article présenté dans le *Chapitre 4*.

- **L. Khelifi**, M. Mignotte. A multi-objective approach based on TOPSIS to solve the image segmentation combination problem. *23rd IEEE International Conference on Pattern Recognition (ICPR)*, pages 4220-4225, Cancun, Mexico, Décembre 2016.

• Travaux sur l'interprétation sémantique des images

- **L. Khelifi**, M. Mignotte. MC-SSM : Nonparametric Semantic Image Segmentation with the ICM algorithm. *Pattern Recognition)*, Soumis Janvier 2018.

  $\Rightarrow$ Article présenté dans le *Chapitre 5*.

- **L. Khelifi**, M. Mignotte. Semantic image segmentation using the ICM algorithm. *24th IEEE International Conference on Image Processing (ICIP)*, Beiging, China, Septembre 2017.

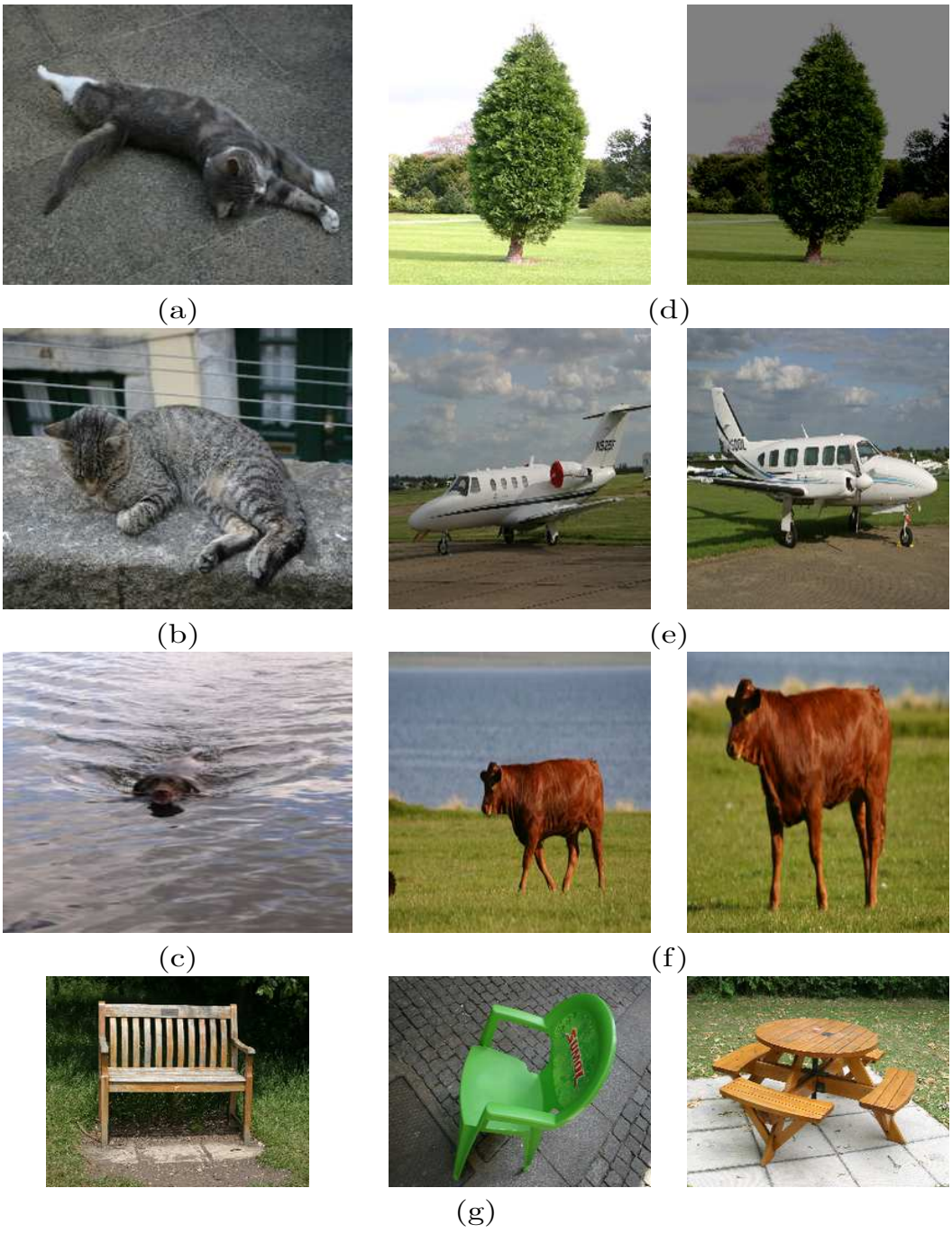


FIGURE 1.3 : Quelques défis liés à l'interprétation sémantique d'images : la déformation (a), la confusion d'arrière-plan (b), l'occultation (c), les conditions d'éclairage (d), la variation de point de vue (e), la variation d'échelle (f), et la variation intra-classe (g).

Première partie

Fusion de segmentations basée sur un modèle mono-objectif

CHAPITRE 2

A NOVEL FUSION APPROACH BASED ON THE GLOBAL CONSISTENCY CRITERION TO FUSING MULTIPLE SEGMENTATIONS

Cet article a été publié dans le journal *IEEE Transactions on Systems, Man, and Cybernetics : Systems* comme l'indique la référence bibliographique.

L. Khelifi, M. Mignotte. A Novel Fusion Approach Based on the Global Consistency Criterion to Fusing Multiple Segmentations

IEEE Transactions on Systems, Man, and Cybernetics : Systems (TSMC), 47 (9) :2489-2502, Septembre 2017.

Cet article est presenté içi dans une version légèrement modifiée.

Abstract

In this work, we introduce a new fusion model whose objective is to fuse multiple region-based segmentation maps to get a final better segmentation result. The suggested new fusion model is based on an energy function originated from the global consistency error (GCE), a perceptual measure which takes into account the inherent multiscale nature of an image segmentation by measuring the level of refinement existing between two spatial partitions. Combined with a region merging/splitting prior, this new energy-based fusion model of label fields allows to define an interesting penalized likelihood estimation procedure based on the global consistency error criterion with which the fusion of basic, rapidly-computed segmentation results appears as a relevant alternative compared with other (possibly complex) segmentation techniques proposed in the image segmentation field. The performance of our fusion model was evaluated on the Berkeley dataset including various segmentations given by humans (manual ground truth segmentations). The obtained results clearly demonstrate the efficiency of this fusion model.

2.1 Introduction

Combining multiple, quickly estimated (and eventually poor or weak) segmentation maps of the same image to obtain a final refined segmentation has become a promising approach, over the last few years, to efficiently solve the difficult problem of unsupervised segmentation [65] of textured natural images.

This strategy is considered as a particular case of the *cluster ensemble problem*. Originally investigated in machine learning[1], this approach is also known as the concept of fusing multiple data clusterings for the amelioration of the final clustering result [58–60, 66]. Indeed, an inherent feature of images is the spatial ordering of the data and thus, image segmentation is a clustering procedure for grid-indexed data. In this context, the partitioning into regions must consider both the closeness in the feature vector space and the spatial coherence property of the image pixels. This approach can also be considered as a special case of restoration/denoising procedure in which each rough segmentation (to be combined) is, in fact, assumed to be a noisy observation or solution and the final goal of a fusion model is to obtain a denoised segmentation solution which could be a compromise or a consensus (in terms of contour accuracy, clusters, number of regions, etc.) provided by each input segmentations. Somehow, the final combined segmentation is the average of all the putative segmentations to be fused with respect to a specific criterion. This approach has firstly been proposed in [61] [62] with a constraint specifying that all input segmentations (to be fused) must be composed of the same region number. Shortly after, other fusion approaches have been proposed with an arbitrary number of regions in [2, 63]. Since these pioneering works, this fusion of multiple segmentations[2] of the same scene in order to get a more accurate and reliable result of segmentation (which would be, in some criterion sense, the average of all the individual segmentation) is now implemented according to several strategies or well-defined criteria.

[1]The cluster ensemble problem, itself, is derived from the theory of merging classifiers to improve the performance of individual classifier and also known under the name of *classifier ensemble problem* or *ensemble of predictors*, *committee machine* or *mixture of expert classifier* [67–69].

[2]This strategy can also be efficiently exploited, more generally, for various other problems involving

Following this strategy, we can mention the combination model introduced in [2] which fuses the individual putative segmentations according to the within-point scatter of the cluster instances (described in terms of the set of local re-quantized label histogram produced by each input segmentations), by simply running a K-means based fusion procedure. By doing so, the author implicitly assumes, in fact, a finite distribution mixture based fusion model in [70] which the labels assigned to the different regions (given by each input segmentations to be fused), are modeled as random variables distributed according K spherical clusters with an equal volume (or gaussian distribution [71] with identical covariance matrix) which can be efficiently clustered with a K-means algorithm. In a similar way, we can also mention the combination model performed in [72] which follows the same idea but for the set of local *soft* labels (estimated with a multiscale thresholding technique) and for which the fusion operation is thus performed in the sense of the weighted within class/cluster inertia. This fusion of segmentations can also be carried out according to the Probabilistic version of the well-known Rand index [70] (PRI) criterion with an energy-based fusion model in order to estimate the segmentation solution with the maximum number of pairs of pixels having a compatible label relationship with the ensemble of segmentations to be fused. This PRI criterion can be minimized either with a stochastic random walking technique [63] (along with an estimator based on mutual information to estimate the optimal region number), or with an algebraic optimization method [73], or with an expectation maximization (EM) procedure [74] (combined with integer linear programming and performed on superpixels, initially estimated by a simple over-segmentation) or also in the penalized PRI sense in conjunction with a global constraint on the combination process [75] (constraining the size and the number of segments) with a Bayesian approach relying on a Markovian energy function to be minimized. Combination of segmentation maps can also be performed according to the variation of information (VoI) criterion [76] (by exploiting an energy-based model minimized by applying a pixel-wise gradient descent method strategy under a spatial coherence constraint). Fusion of segmentations can also be achieved

label maps other than spatial segmentations (e.g., depth field estimation, motion detection or estimation, 3D reconstruction/segmentation, etc.).

in the evidence accumulation sense [59] (and via a hierarchical agglomerative partitioning strategy), or in the F-measure (or precision-recall criterion) sense [77] (and via a hierarchical relaxation scheme fusing the different segments generated in the segmentation ensemble in the final combined segmentation). Finally, we can also mention the fusion scheme proposed in [78] in the optimal or maximum-margin hyperplane (between classes) sense and in which the hyperspectral image is segmented based on the decision fusion of multiple and individual support vector machine classifiers that are trained in different feature subspaces emerging from a single hyperspectral data set or the recent Bayesian [70] fusion procedure for satellite image segmentation proposed in [79]. In addition we can cite the image segmentation fusion model using general ensemble clustering methods proposed in [80] or the approach presented in [81] based on a consensus clustering algorithm, called filtered stochastic best one element move (filtered stochastic BOEM) minimizing a distance function (called symmetric distance function) with a stochastic gradient descent.

The fusion model, introduced in this work, is based on the global consistency error (GCE) measure. This graph theory based measure has been designed to directly take into account the following interesting observation : segmentations produced by experts are generally used as a reference or ground truths for benchmarking segmentations performed by various algorithms (especially for natural images). Even though different people propose different segmentations for the same image, the proposed segmentations differ, essentially, only in the local refinement of regions. In spite of these variabilities, these different segmentations should be interpreted as being consistent, considering that they can express the same image segmented at different levels of detail and, to a certain extend, the GCE measure [70] is designed to take into account this inherent multiscale property of any segmentations made by humans. In our fusion model, this GCE measure, which has thus a perceptual and physical meaning, is herein adopted and tested as a new consensus-based likelihood energy function of a fusion model of multiple weak segmentations.

In the remainder of this paper, we first describe the proposed fusion model and the optimization strategy used to minimize the consensus energy function related to this new

fusion model in Section 2.2. In Section 2.3 we present the generation of the segmentation ensemble to be combined with our model. Finally, an ensemble of experimental tests and comparisons with existing segmentation approaches is described in Section 2.4. In this section, our model of segmentation is tested and benchmarked in the Berkeley color image dataset.

2.2 Proposed Fusion Model

The fusion framework, proposed in this work is a hierarchical energy-based model with an objective consensus energy function derived from the global consistency error (GCE) [18], an interesting perceptual measure which takes into account the inherent multi-scale nature of an image segmentation by measuring the level of refinement existing between two spatial partitions. In addition, to include an explicit regularization hyper parameter overcoming the inherent ill-posed nature of the segmentation problem, we add to this fusion model a merging regularization term, allowing to integrate knowledge about the types of resulting fused segmentation, a priori considered as acceptable solutions. In this new model, the proposed resulting consensus energy-based fusion model of segmentation is efficiently optimized by simply applying a deterministic relaxation scheme on each region given by each individual segmentations to be combined.

2.2.1 The GCE Measure

There are a lot of (similarity) metrics in the statistic and vision literature for measuring the agreement between two clusterings or segmentation maps. Among others, we can cite [82] [83] ; the Jacquard coefficient [84], a variant of the counting pairs also called the Rand index [70] (whose the probabilistic version is the PRI), the Mirkin distance [85], the set matching measures (including the Dongen [86], the F-measure [77] and the purity and inverse purity [87]), and the information theory based metrics ; namely the VoI [76], V-measure [88] or kernel-based metrics (graph kernel or subset significance [89] based measures [90]) or finally the popular Cohen's kappa [91] [92] measure.

In our fusion model we use the global consistency error (GCE) [18] criterion which (is the only one, to our knowledge that) measures the extent to which one segmentation map can be viewed as a refinement of another segmentation. In this metric sense, a perfect correspondence is obtained if each region in one of the segmentation is a subset (i.e., a refinement) or geometrically similar to a region in the other segmentation. Segmentations with similar GCE can be interpreted as being consistent, inasmuch as they could express the same natural image segmented at different degree of detail, as it is the case of the segmented images generated by different human observers for which a finer level of detail will be (possibly) merged by another observer in order to give the larger regions of a segmentation thus estimated at a coarser level.

This GCE distance can be exploited as a segmentation measure to evaluate the correspondence of a segmentation machine with a ground truth segmentation. To this end, it was recently proposed in image segmentation [19, 33] as a quantitative and perceptually interesting metric to compare machine segmentations of an image dataset to their respective manually segmented images given by human experts (i.e., a ground truth segmentations) and/or to objectively measure and rank (based on this GCE criterion) the efficiency of different automatic segmentation algorithms[3].

Let $S^{\mathrm{t}} = \{C_1^{\mathrm{t}}, C_2^{\mathrm{t}}, \ldots, C_{R^{\mathrm{t}}}^{\mathrm{t}}\}$, $S^{\mathrm{g}} = \{C_1^{\mathrm{g}}, C_2^{\mathrm{g}}, \ldots, C_{R^{\mathrm{g}}}^{\mathrm{g}}\}$, R^{t}, and R^{g} be respectively the segmentation result, the manually segmented image, the number of regions[4] in S^{t} and in S^{g}. We consider, for a particular pixel p_i, the segments in S^{t} and S^{g} including this pixel. We denote these segments by $C^{\mathrm{t}}_{<p_i>}$ and $C^{\mathrm{g}}_{<p_i>}$ respectively. If one segment is a subset of the other, so the pixel is practically included in the refinement area, and the local error should be equal to zero. If there is no subset relationship, then the two regions overlap in an inconsistent way and the local error ought be different from zero [18]. The local

[3] In addition, as the semantic gap is generally considered as a difference between low-level segmentation (i.e., labeling decision based on a machine by using pixel information) and high-level segmentation (i.e., based on the human expert's labeling decision, the use of the GCE-based perceptually metric also leads to objectively measure and rank the semantic gap width as well.

[4] A region is a set of connected pixels grouped into the same class and a class, a set of pixels possessing similar textural characteristics.

refinement error (LRE) is therefore denoted at pixel p_i as :

$$\mathrm{LRE}(S^{\mathrm{t}}, S^{\mathrm{g}}, p_i) = \frac{|C^{\mathrm{t}}_{<p_i>} \setminus C^{\mathrm{g}}_{<p_i>}|}{|C^{\mathrm{t}}_{<p_i>}|} \tag{2.1}$$

where $\setminus$ represents the set differencing operator and $|C|$ the cardinality of the set of pixels C. As noticed in [18], this clustering (or segmentation) error measure is not symmetric and encodes a measure of refinement in only one sense. $\mathrm{LRE}(S^{\mathrm{t}}, S^{\mathrm{g}}, p_i)$ is equal to 0 specifically if S^{t} is a refinement of S^{g} at pixel p_i, but not *vice-versa*. A possible and natural way to combine the LRE at each pixel into a measure for the whole image is the so-called global consistency error (GCE) which constraints all local refinement to be in the same sense in the following way :

$$\mathrm{GCE}(S^{\mathrm{t}}, S^{\mathrm{g}}) = \frac{1}{n} \min \left\{ \sum_{i=1}^{n} \mathrm{LRE}(S^{\mathrm{t}}, S^{\mathrm{g}}, p_i), \sum_{i=1}^{n} \mathrm{LRE}(S^{\mathrm{g}}, S^{\mathrm{t}}, p_i) \right\} \tag{2.2}$$

where n is the pixels number p_i within the image. This segmentation error, based on the GCE, is a metric whose values belong to the interval $[0,1]$. A measure of 0 expressed that there is a perfect match between the two segmentations (identical segmentations) and an error of 1 represents a maximum difference between the two segmentations to be compared.

Although a fundamental problem with the GCE measure is that there are two bad, unrealistic segmentation types (i.e., degenerate segmentations) giving an unusually high score value (i.e., a zero error for GCE) [18]. These two degenerative segmentations are the two following trivial cases ; one pixel per region (or segment) and one region per the whole image. The former is, in fact, a detailed improvement (i.e., refinement) of any segmentation, and any segmentation is a refined improvement of the latter. This illustrates why, the GCE measure is useful only when comparing two segmentation maps with an equal number of regions.

In our application, in order to be able to define an energy-based fusion model, avoiding the two above-mentioned degenerate segmentation cases, and for which a reliable

consensus or compromise resulting segmentation map would be solution, *via* an optimization scheme (see Section 2.2.2), we have replaced the minimum operator in the GCE by the average operator :

$$\mathrm{GCE}^{\star}(S^{\mathrm{t}}, S^{\mathrm{g}}) = \frac{1}{2n}\left\{\sum_{i=1}^{n}\mathrm{LRE}(S^{\mathrm{t}}, S^{\mathrm{g}}, p_i) + \sum_{i=1}^{n}\mathrm{LRE}(S^{\mathrm{g}}, S^{\mathrm{t}}, p_i)\right\} \tag{2.3}$$

This new measure is slightly different, while being a tougher measure than the usual and classical GCE measure since $\mathrm{GCE}^{\star}$ is always greater than GCE for any automatic segmentation relatively to a given ground truth S^{g}[5].

The performance score, based on the GCE measure, was also lately used in the segmentation of natural image [94] as a score to compare an unsupervised image segmentation given by an algorithm to an ensemble of ground truth segmentations provided by human experts. This ensemble of slightly different ground truth partitions, given by experts, represents, in essence, the multiple acceptable ground truth segmentations related to each natural image and reflecting the inherent variation of possible (detailed) interpretations (of an image) between each human segmenter. Recently, this variation among human observers, modeled by the Berkeley segmentation database [18], comes from the fact that each human generates a segmentation (of a given image) at different levels of detail. These variations highlight also the fact that the image segmentation is inherently an ill-posed problem in which there are different values of the number of classes for the set of more or less detailed segmentations of a given image. Let us finally mention

[5] An alternative to avoid the above-mentioned degenerate segmentation cases was also proposed in [93] with the so-called bidirectional consistency error (BCE) :

$$\mathrm{BCE}(S^{\mathrm{t}}, S^{\mathrm{g}}) = \frac{1}{n}\sum_{i}\max\Big\{\mathrm{LRE}(S^{\mathrm{t}}, S^{\mathrm{g}}, p_i), \mathrm{LRE}(S^{\mathrm{g}}, S^{\mathrm{t}}, p_i)\Big\}$$

in which the problem of degenerate segmentations 'cheating' a benchmark also disappears. Nevertheless, this measure does not tolerate refinement at all (more precisely, BCE is a measure that penalizes dissimilarity between segmentations proportional to the degree of region overlap) contrary to our $\mathrm{GCE}^{\star}$ measure which tolerates, to a certain extent, a refinement between two segmentations (i.e., which considers, as consistent, two segmentations with a certain different degree of detail).

that, as already said, the GCE metric is a measure tolerant to this intrinsic variability between possible interpretations of an image by different human observers. Indeed, this variability is often due to the refinement between human segmentations represented at different levels of image detail, abstraction or resolution. Thus, in the presence of a set of various human segmentations (showing, in fact, a small fraction of all possible perceptually consistent spatial partitions of an image content [95]), this measure of segmentation quality, based on GCE criterion, has to quantify the degree of similarity between an automatic image segmentation (i.e., performed by an algorithm) and this set of possible ground truths. As proposed in [19], this variability can simply be taken into account by estimating the mean GCE value. More precisely, let us assume a set of L manually segmented images $\{S_k^g\}_{k\leq L} = \{S_1^g, S_2^g, \ldots, S_L^g\}$ related to a same scene. Let S^t be the segmentation to be compared to the manually labeled set, the mean GCE measure is thus given by :

$$\overline{\mathrm{GCE}}\big(S^t, \{S_k^g\}_{k\leq L}\big) = \frac{1}{L}\sum_{k=1}^{L} \mathrm{GCE}\,(S^t, S_k^g) \tag{2.4}$$

and equivalently, we can define :

$$\overline{\mathrm{GCE}^\star}\big(S^t, \{S_k^g\}_{k\leq L}\big) = \frac{1}{L}\sum_{k=1}^{L} \mathrm{GCE}^\star\,(S^t, S_k^g) \tag{2.5}$$

For example, this $\overline{\mathrm{GCE}}$ measure will return a high score (i.e., a low value) for an automatic segmentation S^t which is homogeneous, in the sense of this criterion, with most of the ground truth segmentations provided by human segmenters.

2.2.2 Penalized Likelihood Based Fusion Model

Let us assume now that we have an ensemble of L (different) segmentations $\{S_k\}_{k\leq L} = \{S_1, S_2, \ldots, S_L\}$ (of the same scene) to be combined in the goal of providing a final improved segmentation result $\hat{S}$ (*i.e.,* more accurate than the individual member of $\{S_k\}_{k\leq L}$). To this end, a classic strategy for finding a segmentation result $\hat{S}$, which would be a consensus or compromise of $\{S_k\}_{k\leq L}$, or equivalently, a strategy for combining/fusing

FIGURE 2.1 : Examples of initial segmentation ensemble and fusion results (Algo. GCE-Based Fusion Model). Three first rows ; Results of K-means clustering for the segmentation model presented in Section 2.3. The forth row ; Input image chosen from the Berkeley image dataset and final segmentation given by our fusion framework.

these L individual segmentations, consists in designing an energy-based model generating a segmentation solution which is as close as possible (with the GCE* considered distance) to all the other segmentations or, equivalently, a likelihood estimation model of $\hat{S}$, in the minimum $\overline{\text{GCE}}^\star$ distance sense (or according to the maximum likelihood (ML) principle for this $\overline{\text{GCE}}^\star$ criterion), since this measure, contrary to the $\overline{\text{GCE}}$ measure is not degenerate. This optimization-based approach is sometimes referred to as the *median partition* [60] with respect to both the segmentation ensemble $\{S_k\}_{k\leq L}$ and the $\overline{\text{GCE}}^\star$ criterion. In this framework, if $\mathcal{S}_n$ designates the set of all possible segmentations using n pixels, the consensus segmentation (to be estimated in the $\overline{\text{GCE}}^\star$ criterion sense) is then straightforwardly defined as the minimizer of the $\overline{\text{GCE}}^\star$ function :

$$\hat{S}_{\overline{\text{GCE}}^\star} = \arg\min_{S\in\mathcal{S}_n} \overline{\text{GCE}}^\star\left(S, \{S_k\}_{k\leq L}\right) \tag{2.6}$$

However, the problem of image segmentation remains an ill-posed problem providing different solutions for multiple possible values of regions number (of the final fused segmentation and/or of each segmentation to be fused) and which is *a priori* unknown.

To make this problem a well-posed problem characterized by a unique solution, it is essential to add some constraints on the segmentation process, favoring merging regions or conversely, an over-segmentation. From the probabilistic standpoint, these regularization constraints could be defined via a prior distribution on the segmentation solution $\hat{S}_{\overline{\mathrm{GCE}}^{\star}}$. Analytically, this requires to recast our likelihood estimation problem of the consensus segmentation in the penalized likelihood framework by adding, to the simple ML fusion model [see (2.6)], a regularization term, allowing to integrate knowledge about the types of resulting fused segmentation, *a priori* considered as acceptable solutions. In our case, we search to estimate a resulting segmentation map providing a reasonable number of segments or regions. In our framework, this property, regarding the types of segmentation maps that we would like to favor, can be efficiently modeled and controlled *via* a region merging or splitting regularization term related to the different (connected) region area of the resulting consensus segmentation map. In this optic, an interesting global prior, derived from the information theory, is the following region-based regularization term :

$$E_{\mathrm{Reg}}\Big(S=\{C_k\}_{k\leq R}\Big)=\left|-\sum_{k=1}^{R}\left[\frac{|C_k|}{n}\log\frac{|C_k|}{n}\right]-\overline{\mathcal{R}}\right| \tag{2.7}$$

where we remind that R denotes the region number (or segments) in the segmentation map S, n and $|C_k|$ are respectively the pixel number within the image and the pixel number in the k-th region C_k of the segmentation map S (i.e., the area, in terms of pixel number, of the region C_k). $\overline{\mathcal{R}}$ is an internal parameter of our regularization term that defines the mean entropy of the a priori defined acceptable segmentation solutions. This penalty term favors merging (i.e., leads to a decrease of the penalty energy term) if the current segmentation solution has an entropy greater than $\overline{\mathcal{R}}$ (i.e., in the case of an oversegmentation) and favors splitting in the contrary case. Contrary to the regularization term defined in [75], this one takes into account both the region number of the resulting segmentation solution, but also the proportion of these regions. In image segmentation, this information theoretic regularization term (without the absolute value and with $\overline{\mathcal{R}}=0$) has been used first to restrict the number of clusters of the classical objective function of the fuzzy K-means clustering procedure [96] (i.e., the class number of the segmentation problem)

in [97] and later, to efficiently restrict the number of regions of an objective function in a level set segmentation framework [98]. Finally, with this regularization term, a penalized likelihood solution of our fusion model is thus given by :

$$\begin{aligned}\hat{S}_{\overline{\mathrm{GCE}}^{\star}_{\beta}} &= \arg\min_{S\in\mathcal{S}_n}\left\{\overline{\mathrm{GCE}}^{\star}\left(S,\{S_k\}_{k\leq L}\right)+\beta\, E_{\mathrm{Reg}}(S)\right\}\\ &= \arg\min_{S\in\mathcal{S}_n}\overline{\mathrm{GCE}}^{\star}_{\beta}\left(S,\{S_k\}_{k\leq L}\right)\end{aligned} \tag{2.8}$$

with β allowing to weight the related contribution of the region splitting/merging argument in our energy-based fusion model.

It is also noteworthy to mention that the region splitting/merging regularization term remains essential in some relatively rare cases in which the segmentation solution may lead to a $\overline{\mathrm{GCE}}^{\star}$ measure which is minimal in the trivial one region segmentation case. The penalized likelihood approach allows to avoid these (relatively rare) situations. In addition and consequently, this penalized likelihood approach allows also to exploit the original $\overline{\mathrm{GCE}}$ measure with the minimum operator [see (2.2)]. A comparison of efficiency between these two error metrics, in our fusion based segmentation application, will be discussed later, in the experimental results section.

2.2.3 Optimization of the Fusion Model

Our fusion model of multiple label fields, based on the penalized $\overline{\mathrm{GCE}}^{\star}$ criterion, is therefore formulated as a global optimization problem involving a nonlinear objective function characterized by a huge number of local optima across the lattice of possible clusterings $\mathcal{S}_n$. In our case, this optimization problem is difficult to solve, mainly because (among other things) we are not able to express (for this $\overline{\mathrm{GCE}}^{\star}$ criterion) the local decrease in the energy function for a new label assignment at pixel p_i, and consequently, we cannot adopt the pixel-wise optimization strategy described in [76] in which a simple Gauss-Seidel type algorithm is exploited. This aforementioned Gauss-Seidel type algorithm is, in fact, a deterministic relaxation scheme or an approximate gradient descent where any pixel of the consensus segmentation to be classified are updated one at a time (by searching the minimum local energy label assignment also called the mode). Never-

FIGURE 2.2 : Example of fusion convergence result on three various initializations for the Berkeley image ($n^0187039$). Left : initialization and Right : segmentation result after 8 iterations of our GCEBFM fusion model. From top to bottom, the original image, the two input segmentations (from the segmentation set) which have the best and the worst $\overline{\text{GCE}}^{\star}_{\beta}$ value and one non informative (or blind) initialization.

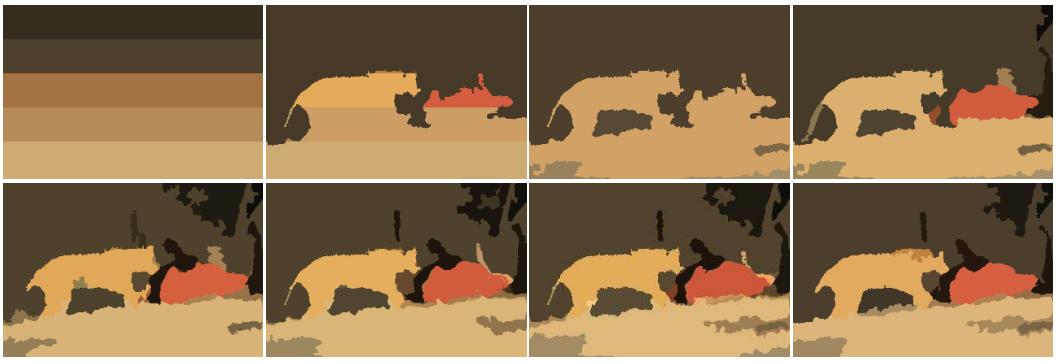


FIGURE 2.3 : Progression of the segmentation result (from lexicographic order) during the iterations of the relaxation process beginning with a non informative (blind) initialization.

theless, in our case, we can adopt the general optimization strategy proposed in [77], in which the strategy of optimization is based on the ensemble of superpixels belonging in $\{S_k\}_{k\leq L}$, i.e., the segments ensemble or regions provided by each individual segmentations to be fused. This approach has other crucial advantages. First, by considering this set of superpixels as the atomic elements to be segmented in the consensus segmentation (instead of the set of pixels), we considerably decrease the computational complexity of the consensus segmentation process. Second, it is also quite reasonable to think that, if individually, each segmentation (to be fused) might give some poor results of segmentation for some sub-parts of the image (i.e., bad regions or superpixels) and also conversely good segmented regions (or superpixels) for other sub-parts of the image, the superpixel ensemble created from $\{S_k\}_{k\leq L}$ is likely to contain the different individual pieces of regions or right segments belonging to the optimal consensus segmentation solution. In this semi-local optimization strategy, the relaxation scheme is based on a variant of the iterative conditional modes (ICM) [99] i.e., a Gauss-Seidel type process (see Algo. 1 for more details) which iteratively optimizes only one superpixel (in our strategy) at a time without considering the effect on other superpixels (until convergence is achieved). On the one hand, this iterative search algorithm is simple and deterministic, however on the other hand, the main drawback of this technique is to strongly depend on the initialization step, which should be not too far from the ideal solution (in order to prevent the ICM from getting stuck in a local minima far from the global one). To this end, we can take, as initialization, the segmentation map $\hat{S}^{\star[0]}_{\overline{\text{GCE}}}$ defined as follow :

$$\hat{S}^{[0]}_{\overline{\text{GCE}}^{\star}_{\beta}} = \arg \min_{S\in\{S_k\}_{k\leq L}} \overline{\text{GCE}}^{\star}_{\beta}\left(S, \{S_k\}_{k\leq L}\right) \tag{2.9}$$

i.e., from the L segmentation to be combined, we can select the one ensuring the minimal consensus energy (in the $\overline{\text{GCE}}^{\star}_{\beta}$ sense) of our fusion model. This segmentation will be considered as the first iteration of our penalized likelihood model (2.8)[6]. This

[6]Another efficient approach consists in running the ICM procedure, independently, with the first N_I optimal input segmentations extracted from the segmentation ensemble (in the $\overline{\text{GCE}}^{\star}_{\beta}$ sense) as initialization, and to select, once convergence is achieved, the result of segmentation associated with the lowest $\overline{\text{GCE}}^{\star}_{\beta}$ energy. This strategy will improve slightly the performance of our combination model, but will increase

iterative algorithm attempts to obtain, for each superpixel to be classified, the minimum energy label assignment. More precisely, it begins with an initialization $\overline{\mathrm{GCE}}^{\star}_{\beta}$ not far to the optimal segmentation [see (2.9)], and for each iteration and each atomic region (superpixel), iterative conditional modes assigns the label giving the largest decrease of the energy function (to be minimized). We summarize in Algo. 1, the overall penalized GCE-based fusion model (GCEBFM) algorithm based on the ICM procedure and superpixel set.

2.3 Generation of the Segmentation Ensemble

The initial ensemble of segmentations, which will be combined via our fusion model, is rapidly generated, in our case, through the standard K-means method [100] associated with 12 different color spaces in order to ensure variability in the segmentation ensemble, those are, YCbCr, TSL, YIQ, XYZ, h123, P1P2, HSL, LAB, RGB, HSV, i123, LUV (in paper [75] more explanation are given on the choice of these color spaces). Also, for the class number K of the K-means, we resort to a metric measuring the complexity relative to each input image, in terms of number of the different texture type present in the natural color image. This metric, presented in [101], is in fact the measure of the absolute deviation (L_1 norm) of the ensemble of normalized histograms obtained for each overlapping squared fixed-size (N_w) neighborhood included within the image. This measure ranges in $[0,1]$ and an image with different textured regions will provide a complexity value close to 1 (and conversely, a value close to 0 when the image is characterized by few texture types). In our framework,

$$K \;=\; \mathrm{floor}\Big(\frac{1}{2} + \big[K^{\max} \times \text{complexity value}\big]\Big) \tag{2.10}$$

where $\mathrm{floor}(x)$ is a function that gives the largest integer less than or equal to x and $K^{\max}$ is an upper-bound of the number of classes for a very complex natural image. It is noteworthy to mention that, in our application, we use three different values of $K^{\max}$ ($K_1^{\max} = 11$,

the computational cost.

$K_2^{\max} = 9$ and $K_3^{\max} = 3$) once again, in order to ensure variability in the segmentation ensemble.

Algorithm 1 Penalized GCE-Based Fusion Algorithm

Mathematical notation:

$\overline{\text{GCE}}_\beta^\star$	Penalized mean GCE (See (2.8))
$\{S_k\}_{k\leq L}$	Set of L segmentations to be fused
$\{b_k\}$	Set of superpixels $\in \{S_k\}_{k\leq L}$
$\{\mathcal{E}_k\}$	Set of region labels in $\{S_k\}_{k\leq L}$
$T_{\max}$	Maximal number of iterations (=8)
β	Regularization parameter

A. Initialization:

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

B. Steepest Local Energy Descent:

while $p < T_{\max}$ **do**
  for each b_k superpixel $\in \{S_k\}_{k\leq L}$ **do**
    Draw a new label x according to the uniform distribution in the set $\{\mathcal{E}_k\}$
    Let $\hat{S}^{[p],\text{new}}_{\overline{\text{GCE}}^\star_\beta}$ the new segmentation map including b_k with the region label x
    Compute $\overline{\text{GCE}}^\star_\beta\left(S, \{S_k\}_{k\leq L}\right)$ on $\hat{S}^{[p],\text{new}}_{\overline{\text{GCE}}^\star_\beta}$
    if $\overline{\text{GCE}}^\star_\beta(\hat{S}^{[p],\text{new}}_{\overline{\text{GCE}}^\star_\beta}) < \overline{\text{GCE}}^\star_\beta(\hat{S}^{[p]}_{\overline{\text{GCE}}^\star_\beta})$ **then**
      $\overline{\text{GCE}}^\star_\beta = \overline{\text{GCE}}^{\star,\text{new}}_\beta$
      $\hat{S}^{[p]}_{\overline{\text{GCE}}^\star_\beta} = \hat{S}^{[p],\text{new}}_{\overline{\text{GCE}}^\star_\beta}$
    end if
  end for
  $p \leftarrow p+1$
end while

In addition, as input multidimensional descriptor of feature, we exploited the ensemble of values (estimated around the pixel to be labeled) of the re-quantized histogram (with equal bins in each color channel). In our framework, this local histogram is re-quantized, for each color channels, in a $N_b = q_b^3$ bin descriptor, estimated on an overlapping, fixed-size squared ($N_w = 7$) neighborhood centered around the pixel to be classified with three different seeds for the K-means algorithm and with two different values of q_b, namely $q_b = 5$ and $q_b = 4$. In all, the number of input segmentations, to be combined, is $60 = 12 \times (3+2)$ [7].

[7] This process aims to ensure the diversity needed to achieve a reliable (i.e., good) set of putative segmentation maps on which the final result will depend. This diversity is crucial to guarantee the availability of more (reliable) information for the consensus function (on which the model of fusion is defined) [60,75]. The use of different segmentations associated with the same scene, expressed in diverse spaces of color, is

FIGURE 2.4 : An example of segmentation solutions generated for different values of $\overline{\mathcal{R}}$ ($\beta = 0.01$), from top to bottom and left to right, $\overline{\mathcal{R}} = \{1.2, 2.2, 3.2, 4.2\}$, respectively segmentation map results with $4, 12, 20, 22$ regions.

2.4 Experimental Results

2.4.1 Initial Tests Setup

In all the tests, the evaluation of our fusion scheme [see (2.8)] is presented for an ensemble of $L = 60$ segmentations $\{S_k\}_{k \leq L}$ with spatial partitions generated with the simple K-means based segmentation technique introduced in Section 2.3 (see Fig. 2.1).

(somewhat) equivalent to observing the scene with several sensors or cameras with different characteristics [79, 102] and also a necessary condition for which the fusion model can be efficiently carried out. On the other hand, it is easy to understand that the fusion of similar solutions of segmentation cannot provide a better reliable segmentation than an individual segmentation. The time of execution, related to each segmentation achieved by this simple K-means technique is rapid (less than 1 second) for a non-optimized sequential program in C++.

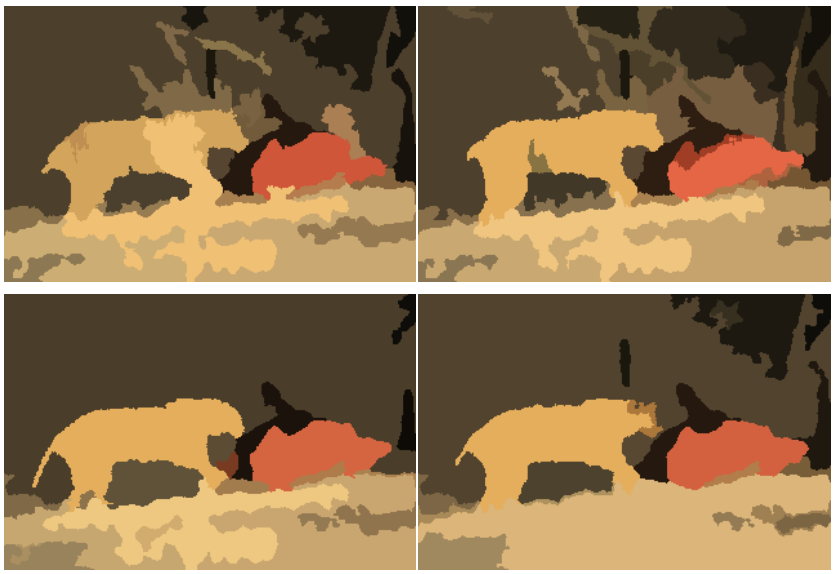


FIGURE 2.5 : Example of fusion result using respectively $L = 5, 10, 30, 60$ input segmentations (i.e., $1, 2, 6, 12$ color spaces). We can also compare the segmentation results with the segmentation maps given by a simple K-means algorithm (see examples of segmentation maps in the segmentation ensemble at Fig. 2.1).

Moreover, for these initial experiments, we have fixed, $\overline{\mathcal{R}} = 4,2$ and $\beta = 0,01$ [see (2.7) and (2.8)]. The justification of these internal parameter values (for the fusion algorithm) will be detailed in Section 2.4.2.

First of all, we have tested the convergence properties of our iterative optimization procedure based on superpixel by choosing, as initialization of our iterative local gradient descent algorithm, various initializations (extracted from our segmentation ensemble $\{S_k\}_{k \leq L}$) and one non informative (or blind) initialization by creating an image exhibiting K horizontal and identical rectangular regions, thus with K various region labels (see Figs. 2.2 and 2.3). Before all, we can notice that our proposed optimization procedure shows good convergence properties in its ability to achieve the optimization of our consensus function of energy. Indeed, the consensus energy function is perhaps not purely convex (three somewhat different solutions are obtained), nevertheless, the obtained final solutions (after 8 iterations) remain very similar. In addition, the final $\overline{\mathrm{GCE}}^{\star}_{\beta}$ score along with the resulting final segmentation map, is on average, all the better than the initial segmentation solution is associated to a good initial $\overline{\mathrm{GCE}}^{\star}_{\beta}$ score (while remaining

robust when the initialization is not reliable). Consequently, the combination of the use the superpixels of $\{S_k\}_{k\leq L}$ along with a good initialization strategy [see (2.9)] definitely gives good convergence properties to our fusion model. Secondly, we have tested the influence of parameter $\overline{\mathcal{R}}$ [see (2.7)] on the generated solutions of segmentation. Fig. 2.4 indicates unambiguously that $\overline{\mathcal{R}}$ can be clearly interpreted as a regularization parameter of the final number of regions of our combination scheme ; favoring under-segmentation, for low values of $\overline{\mathcal{R}}$ (and consequently penalizing small regions) or splitting, for great values of $\overline{\mathcal{R}}$. To further test the regularization role of $\overline{\mathcal{R}}$ in our fusion model, we have also plotted in Fig. 2.6, the average regions number for each image of the BSD300 as a function of the value of $\overline{\mathcal{R}}$. In our case, the value for $\overline{\mathcal{R}} = 4,2$ (see Section 2.4.2) allows to obtain 23 regions, on average, on the BSD300. It is worth recalling that the average regions number belonging to the set of human segmentation ensemble of the BSD300 is around this value (see [19]).

2.4.2 Performances and Comparison

In this section, we have benchmarked our model of fusion as algorithm of segmentation on the Berkeley segmentation dataset (BSD300) [18] (with images normalized to have the longest side equal to 320 pixels). The segmentation results are then super-sampled in order to obtain segmentation images with the original resolution (481×321) before the estimation of the performance metrics.

To this end, several performance measures computed on the full image dataset) will be indicated for a fair comparison with the other state-of-the-art segmenters proposed in the literature. These measures of performance include first and foremost the PRI [103] score, which seems to be among the most correlated (in term of visual perception) with manual segmentations [19] and which is generally exploited for segmentations based on region. This PRI score computes the percentage of pairs of pixel labels perfectly labeled in the result of segmentation and a value equal to PRI=0.75 means that, on average, 75% of pairs of pixel labels are correctly labeled (on average) in the results of segmentation on the BSD300.

To guarantee the integrity of the benchmark results, the two control parameters of

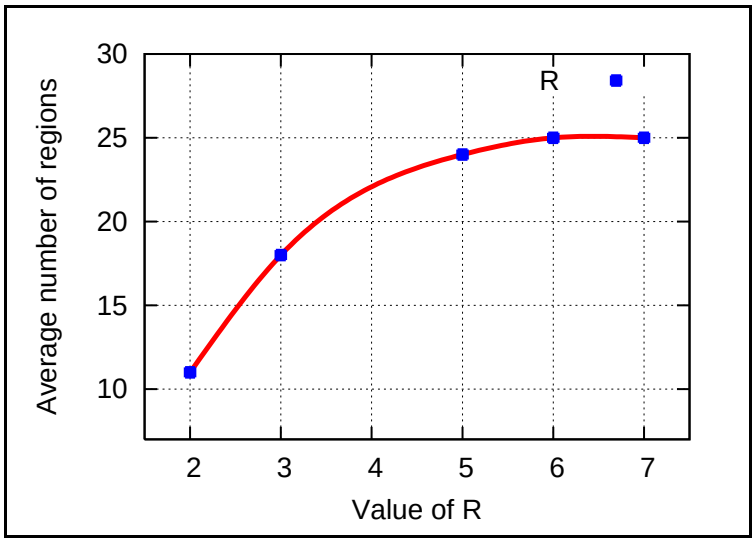


FIGURE 2.6 : Plot of the average number of different regions obtained for each segmentation (of the BSD300) as a function of the value of $\overline{\mathcal{R}}$.

our algorithm of segmentation [i.e., $\overline{\mathcal{R}}$ and β, see (2.7) and (2.8)] are optimized on the ensemble of training images by using a local search procedure (with a fixed step-size) on a discrete grid, on the (hyper)parameter space and in the feasible ranges of parameter values ($\beta \in [10^{-3} : 10^{-1}]$ [step-size = 10^{-3}] and $\overline{\mathcal{R}} \in [3 : 6]$ [step-size = 0.2]. We have found that $\overline{\mathcal{R}} = 4,2$ and $\beta = 10^{-2}$ are reliable hyper-parameters for the model yielding interesting $0,80$ PRI value (see Table 2.1).

For a fair comparison, we now present the results of our fusion model by displaying the same segmented images (see Figs. 2.7 and 2.8) as those presented in the model of fusion introduced in [75, 76]. The results concerning the whole dataset are accessible on-line *via* this link : "http ://www-etud.iro.umontreal.ca/∼khelifil/ResearchMaterial/gcebfm.html".

In order to ensure an effective comparison with other segmentation methods we have also used the variation of information (VoI) measure [106], the GCE [18] and the boundary displacement error (BDE) [107] (this metric measures the average displacement error of boundary pixels between two segmented images, especially, it defines the error of one boundary pixel as the euclidean distance between the pixel and the closest pixel in the other boundary image) (see Table 2.2, the lower distance is better). The results show

TABLE 2.1 : Average performance, related to the PRI metric, of several region-based segmentation algorithms (with or without a fusion model strategy) on the BSD300, ranked in the descending order of their PRI score (the higher value is the better) and considering only the (published) segmentation methods with a PRI score above 0.75.

	ALGORITHMS	**PRI** [103]
	-HUMANS- (in [19])	0,87
With Fusion Model	**-GCEBFM-**	**0,80**
	(2014) -VOIBFM- [76]	0,81
	(2014) -FMBFM- [77]	0,80
	(2010) -PRIF- [75]	0,80
	(2012) -SFSBM- [101]	0,79
	(2008) -FCR- [2]	0,79
	(2009) -Consensus- [73]	0,78
	(2007) -CTM- [19, 33]	0,76
Without Fusion Model	(2012) -MDSCCT- [6]	0,81
	(2011) -gPb-owt-ucm- [11]	0,81
	(2012) -AMUS [74]	0,80
	(2009) -MIS- [46]	0,80
	(2011) -SCKM- [3]	0,80
	(2008) -CTex- [25]	0,80
	(2004) -FH- [12] (in [19])	0,78
	(2011) -MD2S- [4]	0,78
	(2009) -HMC- [36]	0,78
	(2009) -Total Var- [43]	0,78
	(2009) -A-IFS HRI- [29]	0,77
	(2001) -JSEG- [15] (in [25])	0,77
	(2011) -KM- [44]	0,76
	(2006) -Av. Diss- [42] (in [11])	0,76
	(2011) -SCL- [104]	0,76
	(2005) -Mscuts- [57] (in [43])	0,76
	(2003) -Mean-Shift- [14] (in [19])	0,75
	(2008) -NTP- [41]	0,75
	(2010) -iHMRF- [37]	0,75
	(2005) -NCuts- [57] (in [11])	0,75
	(2006) -SWA- [105] (in [11])	0,75

FIGURE 2.7 : Example of segmentations obtained by our algorithm GCEBFM on several images of the Berkeley image dataset (see also Tables 2.1 and 2.2 for quantitative performance measures and "http ://www-etud.iro.umontreal.ca/∼khelifil/ResearchMaterial/gcebfm.html" for the segmentation results on the entire dataset).

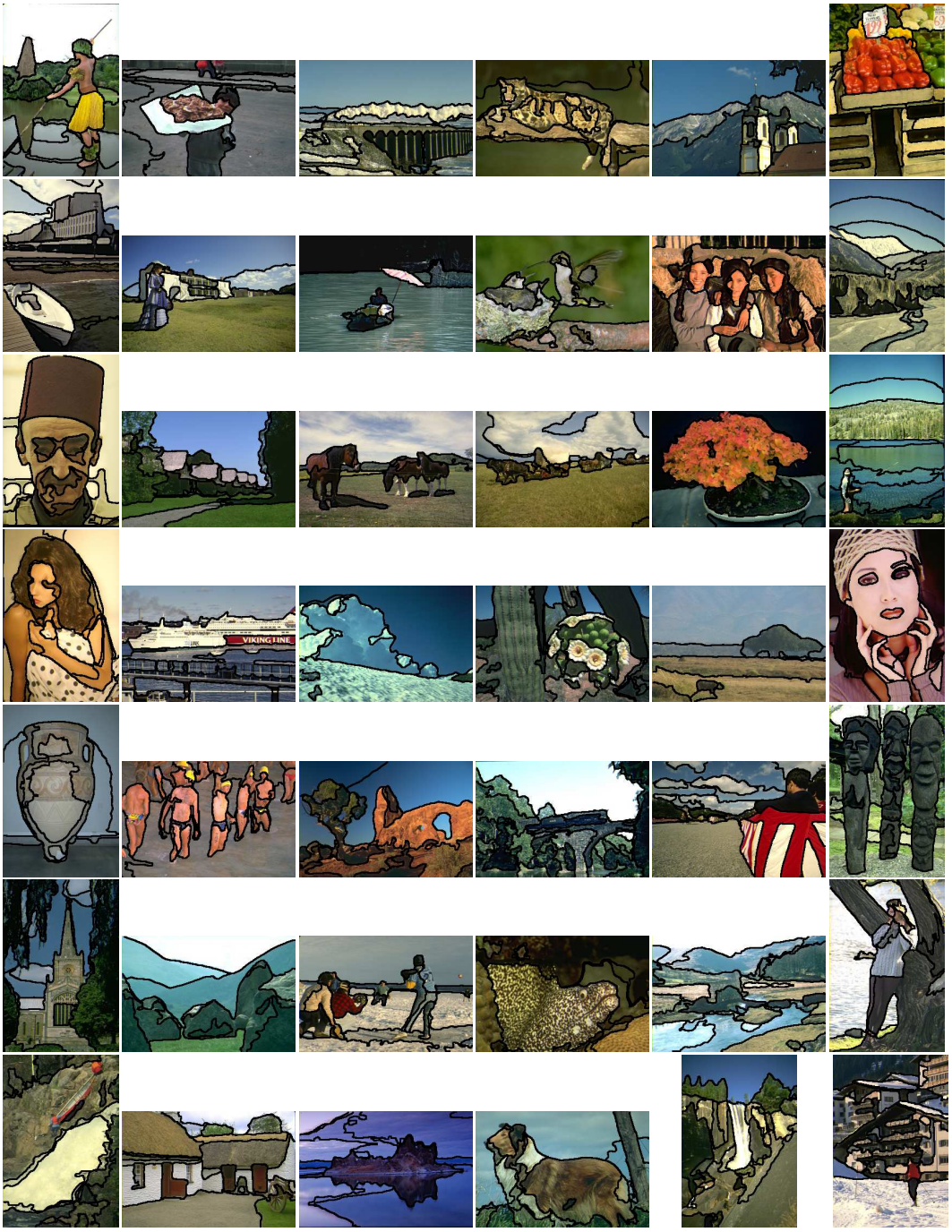


FIGURE 2.8 : Example of segmentations obtained by our algorithm GCEBFM on several images of the Berkeley image dataset (see also Tables 2.1 and 2.2 for quantitative performance measures and "http ://www-etud.iro.umontreal.ca/∼khelifil/ResearchMaterial/gcebfm.html" for the segmentation results on the entire dataset).

TABLE 2.2 : Average performance of diverse region-based segmentation algorithms (with or without a fusion model strategy) for three different performances (distance) measures (the lower value is the better) on the BSD300.

	ALGORITHMS	**VoI**	**GCE**	**BDE**
	-HUMANS-	1,10	0,08	4,99
With Fusion Model	**-GCEBFM-**	**2,10**	**0,19**	**8,73**
	-VOIBFM- [76]	1,88	0,20	9,30
	-FCR- [2]	2,30	0,21	8,99
	-CTM- [19, 33]	2,02	0,19	9,90
	-PRIF- [75]	1,97	0,21	8,45
Without Fusion Model	-MDSCCT- [6]	2,00	0,20	7,95
	-SCKM- [3]	2,11	0,23	10,09
	-MD2S- [4]	2.36	0.23	10,37
	-Mean-Shift- [14] (in [19])	2,48	0,26	9,70
	-NCuts- [40] (in [19])	2,93	0,22	9,60
	-FH- [12] (in [19])	2,66	0,19	9,95
	-AMUS- [74]	1,68	0,17	-

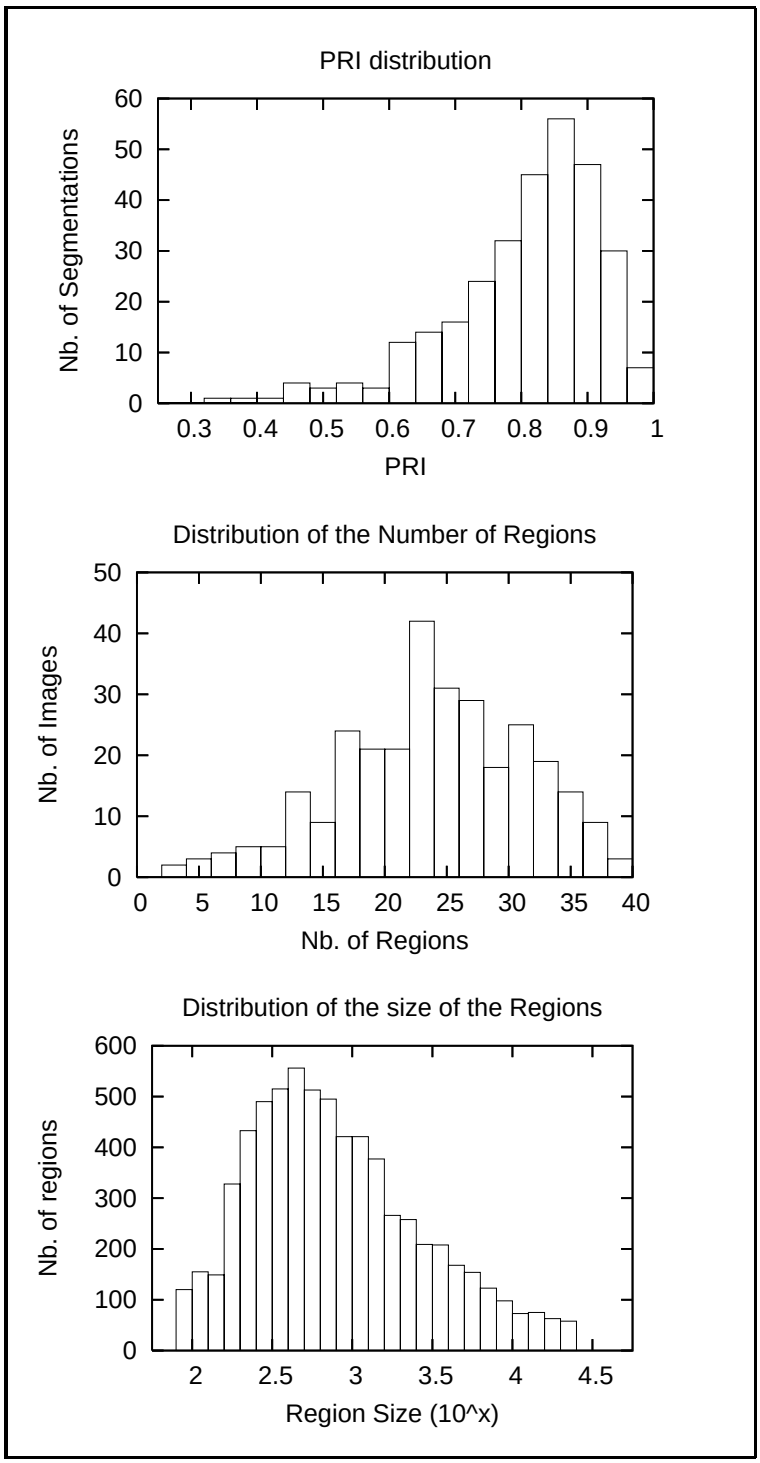


FIGURE 2.9 : Distribution of the PRI metric, the number and the size of regions over the 300 segmented images of the Berkeley image dataset.

that our method provides a competitive result for some other metrics based on different criteria and comparatively to state-of-the arts.

From Table 2.1 we can see that our model yields an interesting PRI value equals to $0,80$ (see Table 2.1). This result means that 80% of pairs of pixel labels are correctly labeled (on average) in the results of segmentation on the BSD300 dataset. Our proposed algorithm is better than different fusion based segmentation model such as CTM [19,33], Consensus [73], FCR [2] and SFSBM [101]. Also, our method is better than different region based segmentation algorithms (without fusion model) methods, for example ; the NCuts [40], Mean-Shift [14], JSEG [15], and MD2S [4]. The experimental results given on Table 2.2 show that our fusion model outperforms all other fusion approaches in term of GCE measure. This can be explained by the fact that our model is based on an energy function originating from the global consistency error (GCE). For example, our GCEBFM model gives better result than the VOIBFM fusion model [76] which is based on the variation of information criteria. Contrary, in terms of VoI measure, we find that the result achieved by our GCEBFM model is worse than the results obtained by the VOIBFM (it is quite logical since the criteria used in the VOIBFM is optimal in the VOI criterion sense). The second remark from Table 2.2 is the proven efficiency of our model in terms of BDE distance. The BDE measures the average displacement error of boundary pixels between two segmented images. The GCEBFM algorithm outperforms the FH [12], NCuts [40], Mean-Shift [14]. It is also worthly to mention that our model outperforms different algorithms based on the same strategy of fusion such as the VOIBFM [76], FCR [2] and CTM [19, 33].

In addition, and as it has been proposed in Section 2.2.2, we have used our penalized likelihood approach with the original $\overline{\text{GCE}}$ consensus energy function, with the minimum operator, [i.e., by using (2.2) instead of (2.3)] and tuned the internal parameters of our segmentation model, noted β and $\overline{\mathcal{R}}$ on the ensemble of training images *via* a local search approach on a discrete grid. We have found that $\overline{\mathcal{R}} = 4.2$ and $\beta = 0.0375$ are optimal hyper-parameters giving the following performance measures ; PRI=0.78, VoI=2.22, GCE=0.20 and BDE=10.43, significantly less better than our GCE* based fusion model.

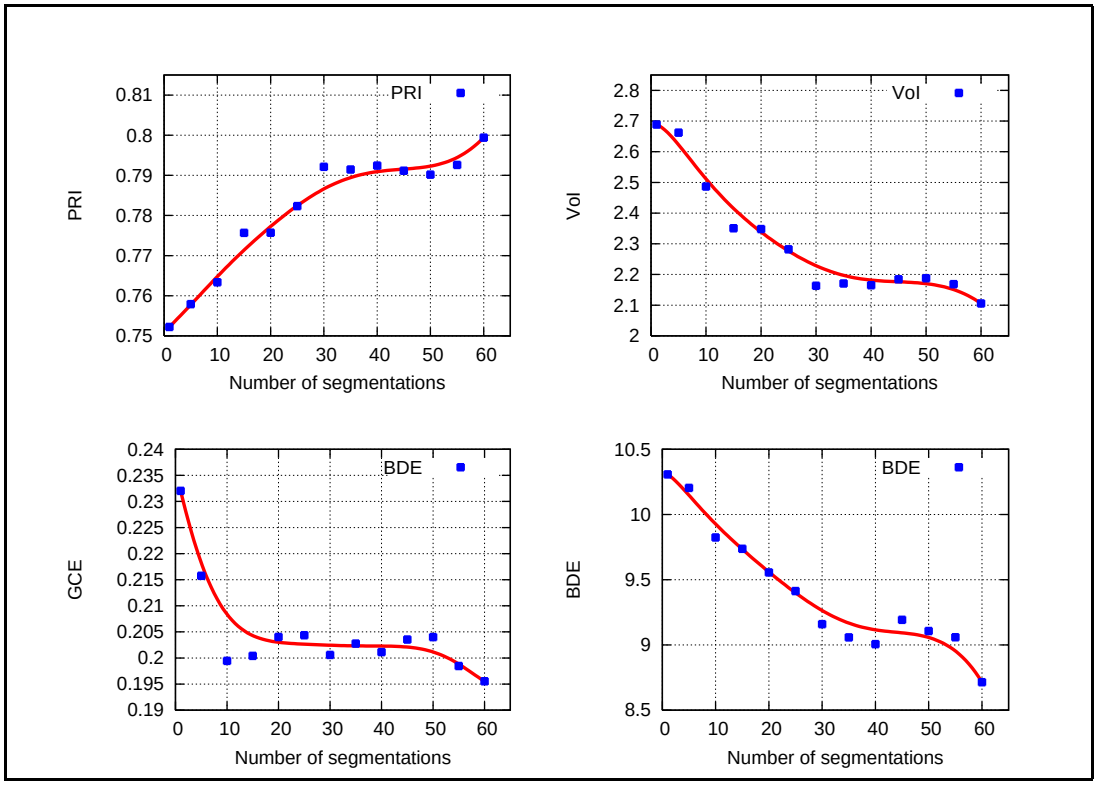


FIGURE 2.10 : From lexicographic order, progression of the PRI (the higher value is better) and VoI, GCE, BDE metrics (the lower value is better) according to the segmentations number (L) to be fused for our GCEBFM algorithm. Precisely, for $L = 1, 5, ..., 60$ segmentations (by considering first, one K-means segmentation (according to the RGB color space) and then by considering five segmentation for each color space and $1, 2, \ldots, 12$ color spaces).

2.4.3 Discussion

As we can notice, our fusion model of simple, rapidly estimated segmentation results is very competitive for different kinds of performance measures and can be regarded as a robust alternative to complex, computationally demanding segmentation models existing in the literature.

We have compared our segmentation algorithm (called GCEBFM) against several unsupervised algorithms. From Table 2.2 we can conclude that our method performs overall better than the others for different and complementary performance measures and especially for the PRI measure (which is important because this measure is highly correlated with human hand segmentations) and with the GCE measure which is closely related to the classification error *via* the computation of the overlap degree between two segmentations (and this good performance is also due to our fusion model which is based on this specific criterion). Statistics on the segmentation results of our method (e.g., the distribution of the PRI, the distribution of the number of regions and size of the regions of the segmented Berkeley database images), for our algorithm are given in Fig. 2.9. These statistics show us that the average number of regions, estimated by our algorithm, is close to the average value given by humans (24 regions) and the PRI distribution shows us that few segmentation exhibits a bad PRI score even for the most difficult segmentation cases.

Moreover, we can observe (see Figs. 2.10 and 2.5) that the PRI, VoI, BDE, GCE performance scores are better when L (the segmentation number to be merged) is high. This test shows the validity and the potentiality of our fusion procedure and demonstrates also that our performance scores are perfectible if the segmentation ensemble is completed by other (and complementary or different) segmentation maps (of the same image).

The experimental results (see Table 2.2) show that our fusion model outperforms all other fusion approaches in term of GCE measure. In the one hand, this result is driven by the fact that our model is based on an energy function originating from the global consistency error (GCE). In the other hand, with the addition of an efficient entropy-based regularization term (see Section 2.2.2), our model can accurately (and adaptively)

estimate a resulting segmentation map with the optimal number of regions, thus yielding a good similarity score between our segmentation results and the ground truth segmentations (given by humane expert). Also, in terms of BDE score, our model outperforms all other fusion approaches excepted the PRIF model. Let us finally add that our model gives a good compromise (comparing to other methods) between all complementary performance measures mentioned in Table 2.1 and Table 2.2 (this last point is important, since this good compromise between relevant complementary performance indicators, is also a clear indication of the quality of segmentations produced by our algorithm).

The PRI, VoI, BDE, GCE measures are quite different for a given image compared to the measures obtained by other approaches like the MDSCCT algorithm (see Table 2.3 and Fig. 2.11). It means that these two methods perform differently and well for different images. This is not surprising since these two methods are, by nature, very different from each other (the MDSCCT is a purely algorithmic approach, on the contrary, our GCEBFM algorithm is a fusion model whose objective is to combine different region-based segmentation maps). This fact may suggest that these two methods extract complementary region information and, consequently, could be paired up or combined together to achieve better results. Also, it is important to note that the GCEBFM method's performance strongly depends on the level of diversity existing in the initial ensemble of segmentations. This means that a better strategy for the generation of the segmentation ensemble could ensure better performance results for our fusion model.

TABLE 2.3 : Comparison of scores between the GCEBFM and the MDSCCT algorithms on the 300 images of the BSDS300. Each value points out the number of images of the BSDS300 that obtain the best score.

ALGORITHMS / MEASURES	GCEBFM	MDSCCT
-GCE-	121	179
-VOI-	166	134
-BDE-	120	180
-PRI-	132	168

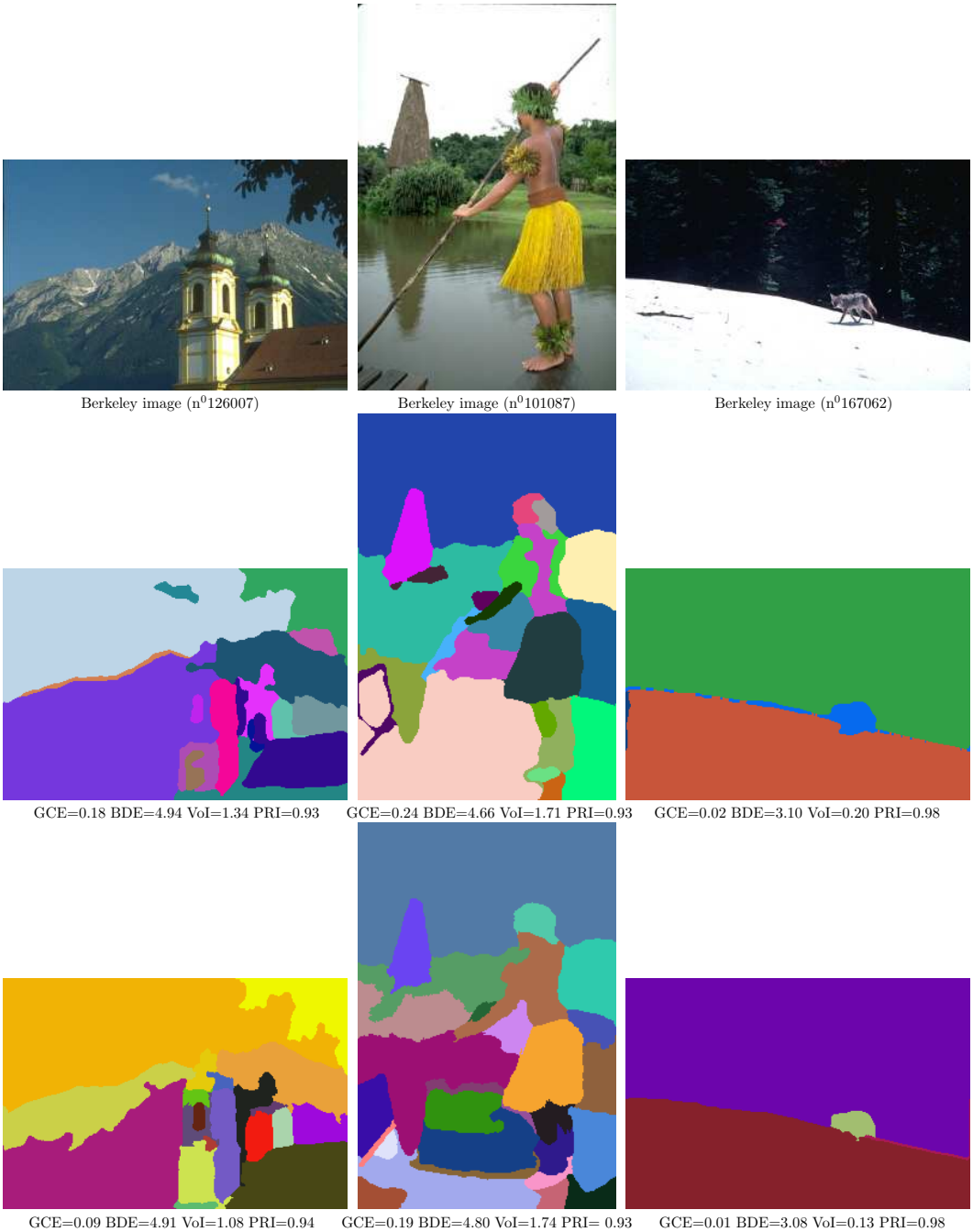


FIGURE 2.11 : First row ; three natural images from the BSD300. Second row ; the result of segmentation provided by the MDSCCT algorithm. Third row ; the result of segmentation obtained by our algorithm GCEBFM.

2.4.4 Computational Complexity

Due to our optimization strategy based on the ensemble of superpixels (see Algo.1), the time complexity of our fusion algorithm is $O\left(nLN_sN_o\right)$ where n, L, N_s, N_o are respectively the pixel number within the image, the number of segmentations to be fused, the number of superpixels existing in the set of segmentations (to be fused) and $N_o < T_{max}$, the number of iterations of the steepest local energy descent (since our iterative optimizer can stop before the maximum number of iterations T_{max}, when convergence is reached).

The segmentation operation takes, on average, about 2 and 3 minutes for an Athlon-AMD 64-Proc-3500+, 2.2 GHz, 4422.40 bogomips and non-optimized code running on Linux ; namely, the two steps (i.e., the estimations of the $L = 60$ weak segmentations to be combined and the minimization step of our fusion algorithm) takes respectively, on average, one minute to generate the segmentation ensemble and approximately two or three minutes for the fusion step and for a 320×214 image (Table 2.4 compares the average computational time for an image segmentation and for different segmentation algorithms whose PRI is greater than 0.76). Also, it is important to mention that the initial segmentations to be combined and the proposed energy-based fusion algorithm could easily be processed in parallel or could efficiently use multi-core processors. It is straightforward for the generation of the set of segmentations but also truth for our fusion model by an application of a Jacobi-type version of the Gauss-Seidel based ICM procedure [108]. The final energy-based minimization can be efficiently performed via the use of the parallel abilities of a Graphic Processor Unit (GPU) (integrated on most computers) which could significantly speed up the algorithm.

Finally, the source code (in C++ language) of our model and the ensemble of segmented images are publicly accessible via this link : "http ://www-etud.iro.umontreal.ca/∼khelifil/ ResearchMaterial/gcebfm.html" in the goal to make possible eventual comparisons with different performance measures and future segmentation methods.

TABLE 2.4 : Average CPU time for different segmentation algorithms.

	ALGORITHMS	**PRI**	**CPU time (s)**	**On [image size]**
With Fusion Model	**-GCEBFM-**	0,80	$\simeq 180$	[320 $\times$ 214]
	-VOIBFM- [76]	0,81	$\simeq 60$	[320 $\times$ 214]
	-FMBFM- [77]	0,80	$\simeq 90$	[320 $\times$ 214]
	-CTM- [19, 33]	0,76	$\simeq 180$	[320 $\times$ 200]
	-PRIF- [75]	0,80	$\simeq 20$	[320 $\times$ 214]
	-FCR- [2]	0,79	$\simeq 60$	[320 $\times$ 200]
Without Fusion Model	-MDSCCT- [6]	0,81	$\simeq 60$	[320 $\times$ 214]
	-CTex- [25]	0,80	$\simeq 85$	[184 $\times$ 184]
	-FH- [12]	0,78	$\simeq 1$	[320 $\times$ 200]
	-HMC- [36]	0,78	$\simeq 80$	[320 $\times$ 200]
	-JSEG- [15]	0,77	$\simeq 6$	[184 $\times$ 184]

2.5 Conclusion

In this work, we have introduced a novel and efficient fusion model whose objective is to fuse multiple segmentation maps to provide a final improved segmentation result, in the global consistency error sense. This new fusion criterion has the appealing property to be perceptual and specifically well suited to the inherent multiscale nature of any image segmentations (which could be possibly viewed as a refinement of another segmentation). More generally, this new fusion scheme can be exploited for any clustering problems using spatially indexed data (e.g., motion detection or estimation, 3D reconstruction, depth field estimation, 3D segmentation, etc.). In order to include an explicit regularization hyper parameter overcoming the inherent ill-posed nature of the segmen-

tation problem, we have re-casted our likelihood estimation problem of the consensus segmentation (or the so-called median partition) in the penalized likelihood framework by adding, to the simple ML fusion model a merging regularization term allowing to integrate knowledge about the types of resulting fused segmentation, a priori considered as acceptable solutions. This penalized likelihood estimation procedure remains simple to implement, perfectible, by incrementing the number of segmentation to be fused, adapted to lower outliers, general enough to be applied to different other problems dealing with label fields and is suitable to be implemented in parallel or to fully take advantage of multi-core (or multi CPU) systems.

Deuxième partie

Fusion de segmentations basée sur un modèle multi-objectif

CHAPITRE 3

EFA-BMFM : A MULTI-CRITERIA FRAMEWORK FOR THE FUSION OF COLOUR IMAGE SEGMENTATION

Cet article a été publié dans le journal *Information fusion* comme l'indique la référence bibliographique

L. Khelifi, M. Mignotte. EFA-BMFM : A Multi-Criteria Framework for the Fusion of Colour Image Segmentation

Information fusion (IF), 38 : 104-121, Novembre 2017.

Cet article est presenté içi dans sa version originale.

Abstract

Considering the recent progress in the development of practical applications in the field of image processing, it is increasingly important to develop new, efficient and more reliable algorithms to solve an image segmentation problem. To this end, various fusion-based segmentation approaches which use consensus clustering, and which are based on the optimization of a single criterion, have been proposed. One of the greatest challenges with these approaches is to select the best fusion criterion, which gives the best performance for the image segmentation model. In this paper, we propose a new fusion model of image segmentation based on multi-objective optimization, which aims to overcome the limitation and bias caused by a single criterion, and to provide a final improved segmentation. To address the ill-posedness for the search of the best criterion, the proposed fusion model combines two conflicting and complementary criteria for segmentation fusion, namely, the region-based variation of information (VoI) criterion and the contour-based F-measure (precision-recall) criterion using an entropy-based confidence weighting factor. To optimize our energy-based model, we propose an extended local optimization procedure based on superpixels and derived from the iterative conditional mode (ICM) algorithm. This new multi-objective median partition-based approach,

which relies on the fusion of inaccurate, quick and spatial clustering results, has emerged as an appealing alternative to the use of traditional segmentation fusion models which exist in the literature. We perform experiments using the Berkeley database with manual ground truth segmentations, and the results clearly show the feasibility and efficiency of the proposed methodology.

3.1 Introduction

The focus of image segmentation is to divide an image into separate regions which have uniform and homogeneous attributes [34]. This step is crucial and important in higher-level tasks such as feature extraction, pattern recognition, and target detection [45]. Several promising methods for segmentation of textured natural images have been recently proposed and reported in the literature. Of those, the ones which are based on the combination of multiple and weak segmentations of the same image to improve the quality of segmentation results are appealing from a theoretical perspective and offer an effective compromise between the complexity of the segmentation model and its efficiency.

Most of these approaches, which are used to compute the segmentation fusion result from a set of initial and weak putative segmentation maps, are theoretically based on the notion of median partition. According to a given specific criterion (which can also be expressed as a distance or a similarity index/measure between two segmentation maps), the median partition approach aims to minimize the average of the distances (or to maximize the average of similarity measures), separating the (consensus) solution from the other segmentations to be fused. To date, a large and growing number of fusion-segmentation approaches based on the result of the median partition problem, along with different criteria or different optimization strategies, have been proposed in the literature.
For example, a fusion model of weak segmentations was initially introduced in the evidence accumulation sense in [59] with a co-association matrix, and in [2], it is then based on a minimization of the inertia (or intra-cluster scatter) criterion across cluster instances (represented by the set of local re-quantized label histogram given by each input segmentation to be fused). The fusion of multiple segmentation maps has also been proposed with respect to the Rand Index (RI) criterion (or its probabilistic version), with either a stochastic constrained random walking technique [63] (within a mutual information-based estimator to assess the optimal number of regions), an algebraic optimization method [73], a Bayesian Markov random field model [75], a superpixel-based approach optimized by the expectation maximization procedure [74] or finally, according

to a similarity distance function built from the adjusted RI [81] and optimized with a stochastic gradient descent. It should also be noted that the solution of the median partition problem can be determined according to an entropy criterion, either in the variation of information (VoI) sense [76], using a linear complexity and energy-based model optimized by an iterative steepest-local energy descent strategy combined with a spatial connectivity constraint, or in the mutual information sense [109] using expectation maximization (EM) optimization. The fusion of clustering results can also be carried out according to the global consistency criterion (GCE) [20] (a perceptual measure which takes into account the inherent multiscale nature of an image segmentation by measuring the level of refinement existing between two spatial partitions) or based on the precision-recall criterion [77] using a hierarchical relaxation scheme. In this context, [80] proposed a methodology allowing the use of virtually any ensemble clustering method to address the problem of image-segmentation combination. The strategy is mainly based on a preprocessing step which estimates a superpixel map from the segmentation ensemble in order to reduce the dimensionality of the combinatorial problem. Finally, in remote sensing, there have been reports of the combining model based on the maximum-margin sense (of the hyperplanes between spatial clusters) [78] or the recent Bayesian fusion procedure proposed in [79], in which the class labels obtained from different segmentation maps (obtained from different sensors) are fused by the weights of the evidence model.

In fact, the performance of these energy-based fusion models is related both to the optimization procedure, with its potential ability to find an optimal solution (as quickly as possible), and it also largely depends on the chosen fusion criterion, which defines all the intrinsic properties of the consensus segmentation map to be estimated. However, by assuming that an efficient optimization procedure is designed and implemented (in terms of its ability to quickly find a global optimal and stable solution), it remains unclear whether it can find the most appropriate single criterion allowing both to extract all the useful information contained in the segmentation ensemble and also to model all the complex geometric properties of the final consensus segmentation map. Another

way to look at this problem is to understand that if the optimization problem is based on the optimization of a single criterion, the fusion procedure is inherently biased towards searching one particular family of possible solutions ; otherwise, some specific regions of the search space contain solutions, which are *a priori* defined (by the criterion), as acceptable solutions. This may bias and limit the performance of an image segmentation model. To overcome this main disadvantage (the bias caused by a single criterion), we propose an interesting solution to use approaches based on multi-objective optimization in order to design a new fusion-segmentation model which takes advantage of the (potential) complementarity of different objectives (criteria), and enables us to finally obtain a better consensus segmentation result. Following this new strategy, in this work, we introduce a new multi-criteria fusion model weighted by an entropy-based confidence measure (EFA-BMFM). The main goal of this model is to simultaneously combine and optimize two different and complementary segmentation-fusion criteria, namely, the (region-based) VoI criterion and the (contour-based) F-measure (derived from the precision-recall) criterion.

The remainder of the paper is organized as follows. In Section 3.2, we present basic concepts of multi-objective optimization. In Section 3.3, we describe the generation of the segmentation ensemble to be fused by our model, while in Section 3.4, we describe the proposed fusion model, *i.e.*, the used segmentation criteria, the multi-objective function and the optimization strategy of the proposed algorithm for the fusion of image segmentation. We explain the experiments and discussions in Section 3.5, and in Section 3.6, we conclude the paper.

3.2 Multi-objective Optimization

The motivation of using multi-objective (MO) optimization comes from all the drawbacks and limitations of using a mono-objective one, as mentioned in our preliminary work [110]. As previously mentioned, the final segmentation solution is inherently biased by the chosen single criterion as well as by the parameters of the model and the possible outliers of the segmentation ensemble. A MO optimization-based segmentation

fusion framework enables us to more efficiently extract the useful information contained in the segmentation ensemble according to different criteria or different viewpoints, as well as to model easily all the complex geometric properties of the final consensus segmentation map *a priori* defined as the acceptable solution. To this end, the challenge is to find two different and complementary criteria.

Contrary to the mono-objective optimization case, in the MO optimization case, there are often several conflicting objectives to be simultaneously optimized [111]. Existing approaches which are utilized to solve a MO problem can be distinguished into two classes [112]. The first class is called the Pareto approach (PTA), and aims to provide a set of solutions which are non-dominated with respect to each objective. The second class (adopted in our work) is called the weighted formula approach (WFA), which transforms a MO problem into a problem with a single objective function. This is typically achieved by first assigning a numerical (estimated data-driven) weight to each objective (evaluation criterion), and then combining the values of the weighted criteria into a single value either by adding all the weighted criteria. The formula to determine the quality (or cost) Z which is related to a given candidate model is written as :

$$Z = \mathrm{w}_1\,\mathrm{c}_1 + \mathrm{w}_2\,\mathrm{c}_2 + \ldots + \mathrm{w}_n\,\mathrm{c}_n \tag{3.1}$$

with n representing the number of evaluation criteria, and w_i are real-valued weights (assigned to criteria c_i) which satisfy the following relations : $0 \leqslant \mathrm{w}_i \leqslant 1$ and $\sum_{i=1}^{n} \mathrm{w}_i = 1$.

A geometric representation of the WFA approach is given in Fig 3.1. In fact, the minimization of Z can be analysed by searching the value of the tangency point A for which the line T with slope $-\mathrm{w}_1/\mathrm{w}_2$ (associated with $\mathrm{c}_2 = -\frac{\mathrm{w}_1}{\mathrm{w}_2}\,\mathrm{c}_1 + \frac{Z}{\mathrm{w}_2}$ in the case of two objectives) just touches the boundary of the set of feasible solutions (FS) (related to the couple [c_1, c_2]). Note that the estimation of the weights (also known as the *importance factors*) is an essential step, and should be based on the degree of information or the confidence levels regarding the ensemble of segmentations (to be fused) provided by each criterion, along with the difference of the scaling between these two criteria. This re-scaling is essential to prevent either of the two criteria from being assigned too much significance ;

otherwise, it would make the fusion of the two criteria ineffective. To address this issue, we propose an entropy-based confidence measure (see Section 3.4.3).

3.3 Generation of the Initial Segmentations

In our application, it is simple to acquire the initial segmentations (see Fig. 3.2) used by our fusion Framework. To do this, we employ a K-means [100] clustering technique, with an image expressed in 12 different colour spaces [1], namely : RGB, HSV, YIQ, XYZ, LAB, LUV, i123, h123, YCbCr, TSL, HSL and P1P2. For each input image of the BSDS300, we predict the cluster number of the K-means algorithm (K) using a metric which measures the complexity in terms of the number of distinct texture classes within the image. This metric, which is defined in [113], has a range of $[0,1]$, where a value close to 0 means that the image has few texture patterns, and a value close to 1 means that the image has several different types of texture. Mathematically, the value of K is written as :

$$K \quad = \quad \text{floor}\big(\frac{1}{2} + \big[K^{\max} \times \text{complexity value}\big]\big) \tag{3.2}$$

where $\text{floor}(x)$ is a function which gives the largest integer less than or equal to x, and $K^{\max}$ is an upper-bound of the number of classes for a very complex natural image. In our application, we used three different values of $K^{\max}$, namely$K_1^{\max} = 11$, $K_2^{\max} = 9$ and $K_3^{\max} = 3$. Additional details about the complexity value of an image are given in [76]. Note that in our case, the complexity is a measure of the absolute deviation (L_1 norm) of the set of normalized histograms or feature vectors for each overlapping squared fixed-size (N_w) neighbourhood contained within the input image.

Moreover, we used a set of values of the re-quantized colour histogram (as a feature vector for the K-means) with equidistant binning, which is estimated around the pixel to be classified. In our framework, this local histogram is equally re-quantized for each of the three-colour channels in a $N_b = q_b^3$ bin descriptor. This descriptor is computed on an overlapping squared fixed-size ($N_w = 7$) neighbourhood, which is centered around the pixel to be segmented using three different values of $K^{\max}$ for the K-means algorithm, and

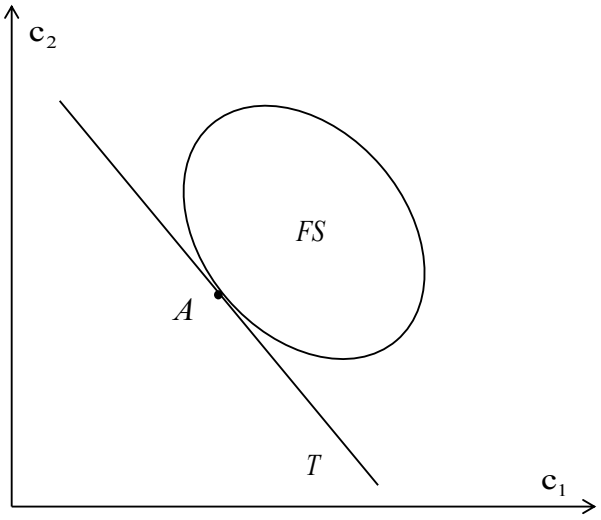


FIGURE 3.1 : The weighted formula approach (WFA).

using two different values of q_b, namely $q_b = 5$ and $q_b = 4$, for a total of $12 \times (3+2) = 60$ input segmentations to be fused.

It should be noted that different weak segmentations (resulting from a simple K-means expressed in different colour spaces) used in our fusion model can be easily viewed as different and complementary image channels, as provided by various sensors. In this context, our fusion model has the same goal of a multi-sensor data fusion scheme [114–116], which aims to take advantage of the complementarity in the data in order to improve the final result. In addition, different values of $K^{\max}$ (which is related to the cluster number) and q_b (related to the level of resolution of the texture model used in the K-means) enable us to generate a consistent variability in the segmentation ensemble, and considers the inherently ill-posed nature of the segmentation problem, which is due to the large number of possible partitions for a single image, and which can also be segmented at different levels of resolution or detail by different human observers.

[1] It should be noted that each colour space has an interesting specific property which is efficiently taken into account in our application in order to better diversify the segmentation ensemble (to be fused), and thus making a more reliable final fusion procedure. For example, RGB is an additive colour system based on trichromatic theory, and is nonlinear with visual perception. This space colour appears to be optimal for tracking applications [117]. The LAB colour system approximates human vision, and its component clo-

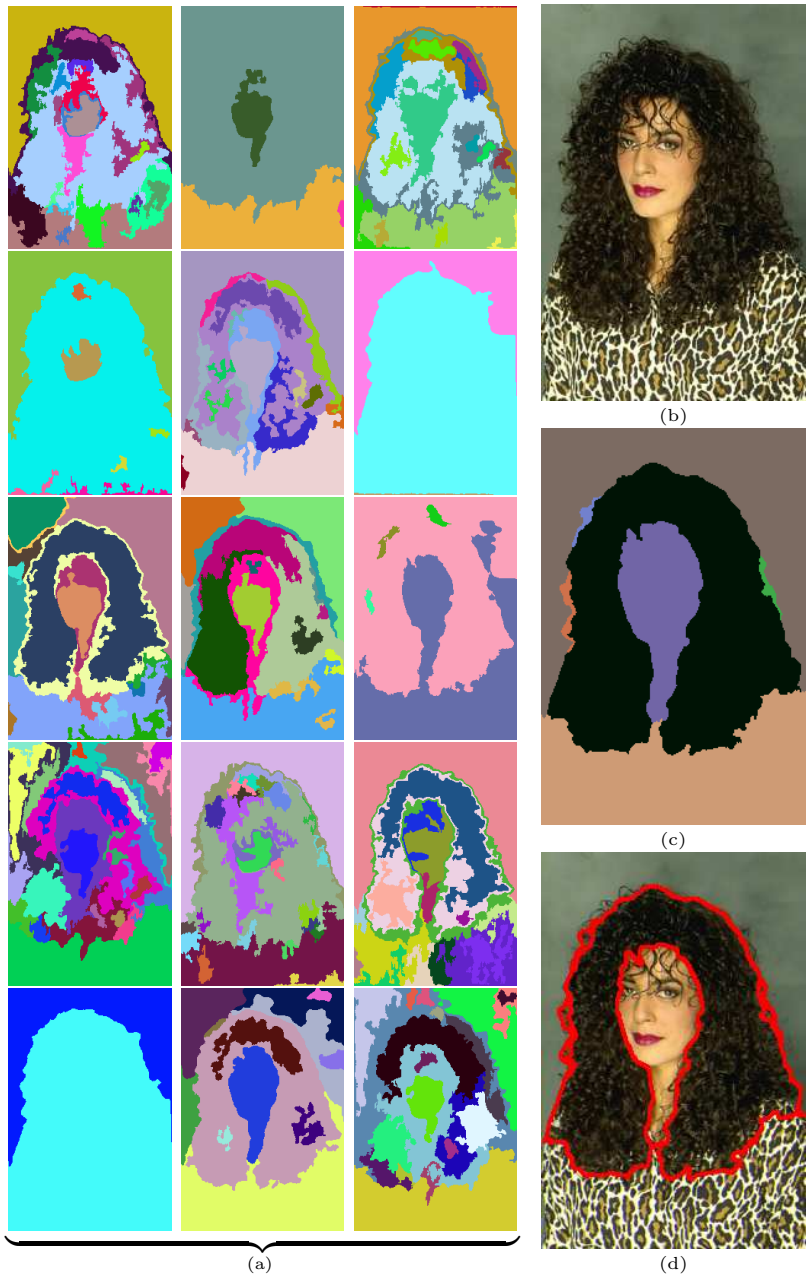


FIGURE 3.2 : Examples of initial segmentation set and combination result (output of Algorithm 1). (a) Results of K-means clustering. (b) Input image ID 198054 selected from the Berkeley image dataset. (c) Final segmentation given by our fusion framework. (d) Contour superimposed on the colour image.

3.4 Proposed Fusion Method

3.4.1 Region-based VoI criterion

The VoI [106] is an information theoretic criterion used for comparing two segmentations (partitions) or clusterings. By measuring the amount of information which is lost or gained while switching from one clustering to another, this metric aims to quantify the information shared between two partitions. In particular, the VoI takes a value of 0 when two clusterings are identical, but ≤ 1 otherwise. Similarly, it also expresses roughly the amount of randomness in one segmentation which cannot be explained by the other [121].

Let us assume that there is a machine segmentation to be computed (or compared) $S^{\mathrm{a}} = \{C_1^{\mathrm{a}}, C_2^{\mathrm{a}}, \ldots, C_{R^{\mathrm{a}}}^{\mathrm{a}}\}$ relative to a (ideal) manually segmented image $S^{\mathrm{b}} = \{C_1^{\mathrm{b}}, C_2^{\mathrm{b}}, \ldots, C_{R^{\mathrm{b}}}^{\mathrm{b}}\}$, where R^{a} represents the number of segments or regions (C) in S^{a} and R^{b} denotes the number of regions in S^{b}. The VoI distance between S^{a} and S^{b} can be written as follows :

$$\mathrm{VoI}(S^{\mathrm{a}}, S^{\mathrm{b}}) = \mathrm{H}(S^{\mathrm{a}}) + \mathrm{H}(S^{\mathrm{b}}) - 2\,\mathrm{I}(S^{\mathrm{a}}, S^{\mathrm{b}}) \tag{3.3}$$

where $\mathrm{H}(S^{\mathrm{a}})$ and $\mathrm{H}(S^{\mathrm{b}})$ denote the entropy associated with the segmentation S^{a}, and S^{b} and $\mathrm{I}(S^{\mathrm{a}}, S^{\mathrm{b}})$ represent the mutual information between these two spatial partitions. Let n be the number of pixels within the image, let n_i^a be the number of pixels in the i^{th} cluster i of the segmentation S^{a}, n_j^b the number of pixels in the j^{th} cluster j of the segmentation S^{b} and finally, n_j^i 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}. Note that the entropy is always positive or zero in the case where there is no uncertainty (when there is only one

sely matches the human perception of lightness [118]. The LUV components provide an Euclidean colour space yielding a perceptually uniform spacing of colour approximating a Riemannian space [119]. The HSV is interesting in order to decouple chromatic information from the shading effect [120]. The YIQ colour channels have the property of being able to code the luminance and chrominance information, which are useful in compression applications. Besides, this system is intended to take advantage of human colour characteristics. XYZ has the advantage of being more psycho-visually linear, although they are nonlinear in terms of linear-component colour mixing. Each of these properties will be efficiently combined by our fusion technique.

cluster), and is given by :

$$\mathrm{H}(S^{\mathrm{a}}) = -\sum_{i=1}^{R^{\mathrm{a}}} \mathrm{P(i)} \log \mathrm{P(i)} = -\sum_{i=1}^{R^{\mathrm{a}}} (\frac{n_i^a}{n}) \log (\frac{n_i^a}{n}) \tag{3.4}$$

$$\mathrm{H}(S^{\mathrm{b}}) = -\sum_{j=1}^{R^{\mathrm{b}}} \mathrm{P(j)} \log \mathrm{P(j)} = -\sum_{j=1}^{R^{\mathrm{b}}} (\frac{n_j^b}{n}) \log (\frac{n_j^b}{n}) \tag{3.5}$$

where $\mathrm{P(i)} = n_i^a/n$ represents the probability that a pixel belongs to cluster S^{a} (respectively $\mathrm{P(j)} = n_j^b/n$ being the probability that a pixel belongs to cluster S^{b}) in the case where i and j represent two discrete random variables with values of R^{a} and R^{b}, respectively, and uniquely related to the partition S^{a} and S^{b}. Now, let us assume that $\mathrm{P(i,j)} = n_{ij}/n$ represents the probability when a pixel belongs to C_i^{a} and to C_j^{b}, which is the mutual information between the partitions S^{a}, and S^{b} is equal to the mutual information between the random variables i and j, and is expressed as :

$$\mathrm{I}(S^{\mathrm{a}}, S^{\mathrm{b}}) = \sum_{i=1}^{R^{\mathrm{a}}} \sum_{j=1}^{R^{\mathrm{b}}} \mathrm{P(i,j)} \log \left(\frac{\mathrm{P(i,j)}}{\mathrm{P(i)P(j)}} \right). \tag{3.6}$$

3.4.2 Contour-based F-measure criterion

In the field of statistical analysis, the F-measure score (also called the F-score or F1 score) is defined as a measure of a test's accuracy. We obtained the results of the F-measure from a combination of two complementary measures, i.e. precision (Pr) and recall (Re). In the (contour-based) image segmentation domain, these two scores respectively represent the fraction of detections which are true boundaries and the fraction of the true boundaries detected [77]. In particular, a low precision value is typically the result of significant over-segmentation, and highlights the fact that a large number of boundary pixels have poor localization. On the contrary, the recall is low when there is significant under-segmentation or when there is a failure to capture the salient image structure (in terms of contours). In other words, precision and recall can be understood in terms of the rate of *false positives* and *missed detection*.

Mathematically, let us assume that a segmentation result $S^{\mathrm{a}} = \{C_1^{\mathrm{a}}, C_2^{\mathrm{a}}, \ldots, C_{R^{\mathrm{a}}}^{\mathrm{a}}\}$ has

to be compared with a manually segmented image $S^b = \{C_1^b, C_2^b, \ldots, C_{R^b}^b\}$ (considered as ground truth), where R^a represents the number of regions (C) in S^a and R^b denotes the number of regions in S^b. Now, let B_{C^a} be the set of pixels which belong to the boundary of the segment C^a in the segmentation S^a (B_{C^b} is the set of pixels belonging to the boundary of the segment C^b in the ground truth segmentation S^b). The precision (Pr) and recall (Re) are then respectively defined as :

$$Pr = \frac{|B_{C^a} \cap B_{C^b}|}{|B_{C^a}|} \quad , \quad Re = \frac{|B_{C^a} \cap B_{C^b}|}{|B_{C^b}|} \tag{3.7}$$

where $\cap$ denotes the intersection operator and $|X|$ represents the cardinality of the set of pixel X.

Generally, the performance of a boundary detector providing a binary output is represented by a point in the precision-recall plane. If the output is a soft or a probabilistic boundary representation, a precision-recall curve displays the trade-off between the absence of noise and the fidelity to the ground truth, considering that the threshold parameter of the boundary detector varies. A specific application[2] can characterise the relative cost α between these two amounts, which highlights a particular point on the precision-recall curve [94]. In this case, the new expression of the F-measure is given as follows :

$$F_\alpha = \frac{Pr \times Re}{\alpha \times Re + (1-\alpha) \times Pr} \tag{3.8}$$

which is within the range $[0,1]$ where a score equal to 1 indicates that two segmentations are identical (*i.e.* they have identical contours).

3.4.3 Multi-objective function

The VoI and F-measures, which are described in Section 3.4.1 and Section 3.4.2, are in fact frequently used to validate a new segmentation method [19, 75, 94] as two

[2]In the case of an algorithm performing a search task, it is usually preferable to have a lower rate of false positives (higher precision) than a low rate of missed detections (high recall).

complementary comparison measures which enable the assessment of an automatic segmentation (*i.e.* given by an algorithm) relative to a set of ground truth segmentations (provided by a set of human experts). This summarizes the possible (and consistent) interpretation of an input image segmented at different levels of detail or resolution levels (see Fig. 3.3). Let $\{S_k^b\}_{k\leq L} = \{S_1^b, S_2^b, \ldots, S_L^b\}$ be a finite ensemble of L manually obtained ground truth segmented images of the same scene (segmented by L different human experts at different levels of detail), and S^{a} be the spatial clustering result to be estimated by making a comparison with the manually labeled set $\{S_k^b\}_{k\leq L}$. The mean F-measure and the mean VoI metrics are simply the two metrics which consider this set of possible ground truth segmentations, *i.e.* :

$$\overline{\mathrm{C}}\big(S^{\mathrm{a}}, \{S_k^b\}_{k\leq L}\big) \;=\; \frac{1}{L}\sum_{k=1}^{L} \mathrm{C}\,(S^{\mathrm{a}}, S_k^b) \tag{3.9}$$

with $\mathrm{C}\in\{\mathrm{VoI}, \mathrm{F}_\alpha\}$. In particular, the $\overline{\mathrm{VoI}}$ distance function will give a low value (on the contrary, the $\overline{\mathrm{F}}_\alpha$ measure function will give a high value) to a segmentation result S^{a} which is in good agreement with the set of segmentation maps obtained from human experts.

In our case, we aim to obtain a final improved segmentation result $\hat{S}$ by the fusion of a family of L segmentations $\{S_k\}_{k\leq L} = \{S_1, S_2, \ldots, S_L\}$ (associated with the same scene or image), with the hope that the result is more accurate than that of each individual member of $\{S_k\}_{k\leq L}$. To this end, these two complementary criteria, namely the contour-based F-measure and the region-based VoI measure, can be used directly as an MO cost function in an energy-based model. From this point of view, the consensus segmentation $\hat{S}_{\overline{\mathrm{MO}}}$ is simply obtained as the result of the following bi-criteria optimization problem :

$$\hat{S}_{\overline{\mathrm{MO}}} \;=\; \arg\min_{S\in\mathcal{S}_n} \overline{\mathrm{MO}}\,(S, \{S_k\}_{k\leq L}) \qquad \text{with :} \tag{3.10}$$

$$\overline{\mathrm{MO}}\big(S, \{S_k\}_{k\leq L}\big) \;=\; \mathrm{w}_{\mathrm{VoI}}\,\overline{\mathrm{VoI}}\big(S, \{S_k\}_{k\leq L}\big) + \frac{\mathrm{w}_{\mathrm{F}\alpha}}{\overline{\mathrm{F}}_\alpha\big(S, \{S_k\}_{k\leq L}\big)} \tag{3.11}$$

where S is a segmentation map belonging to the set of possible segmentations ($S \in \mathcal{S}_n$). The importance (or weighting) factors $\mathrm{w}_{\mathrm{VoI}}$ and $\mathrm{w}_{\mathrm{F}_\alpha}$ must be data-driven and estimated

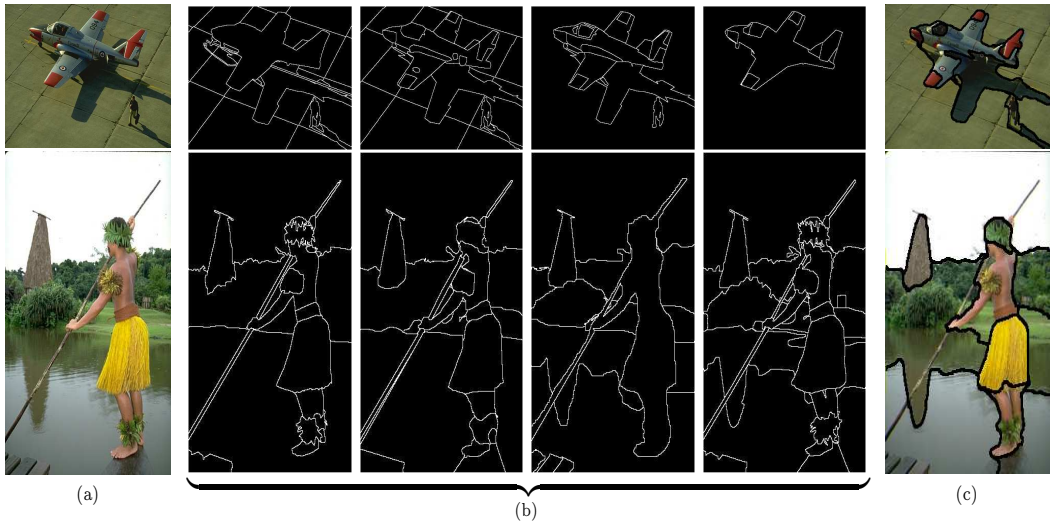


FIGURE 3.3 : Two images from the BSDS300 (a) and its ground truth boundaries (b). Segmentation results obtained by our EFA-BMFM are shown in (c).

based on the concept of the informational importance of the segmentation ensemble given a criterion, or according to the traditional multiple-criteria analysis decision making (MCDM) problem under uncertainty [122] based on the intrinsic information generated by the segmentation ensemble through each criterion.

In our model, we can use the entropy value to measure the amount of decision information contained in the segmentation ensemble and related to each criterion as follows (for the VoI criterion and similarly for the F-measure) :

$$\mathrm{e}_{\mathrm{VoI}} = -D\sum_{i=1}^{L}\left\{\frac{\overline{\mathrm{VoI}}\left(S_i,\{S_k\}_{k\leq L}\right)}{S_{\overline{\mathrm{VoI}}}}\log\frac{\overline{\mathrm{VoI}}\left(S_i,\{S_k\}_{k\leq L}\right)}{S_{\overline{\mathrm{VoI}}}}\right\} \tag{3.12}$$

$$\text{where :}\quad S_{\overline{\mathrm{VoI}}} = \sum_{j=1}^{L}\overline{\mathrm{VoI}}\left(S_j,\{S_k\}_{k\leq L}\right) \tag{3.13}$$

where $\mathrm{D} = 1/\log(\mathrm{L})$ is a constant which guarantees $0 \leqslant \mathrm{e}_{\mathrm{VoI}} \leqslant 1$. In this context, the degree of divergence of the intrinsic information (or the contrast intensity) of the VoI

and the F_α criterion can be measured as follows :

$$\mathrm{d}_{\mathrm{VoI}} = 1 - \mathrm{e}_{\mathrm{VoI}} \tag{3.14}$$

$$\mathrm{d}_{F_\alpha} = 1 - \mathrm{e}_{F_\alpha} \tag{3.15}$$

and finally, the objective weight for each criterion (VoI and F_α) is thus defined by :

$$\mathrm{W}_{\mathrm{VoI}} = \frac{\mathrm{d}_{\mathrm{VoI}}}{\mathrm{d}_{\mathrm{VoI}} + \mathrm{d}_{\mathrm{F}_\alpha}} \quad \in [0,1] \tag{3.16}$$

$$\mathrm{W}_{\mathrm{F}_\alpha} = \frac{\mathrm{d}_{\mathrm{F}_\alpha}}{\mathrm{d}_{\mathrm{VoI}} + \mathrm{d}_{\mathrm{F}_\alpha}} \quad \in [0,1]. \tag{3.17}$$

In this manner, the entropy generated by the set of mean pairwise VoI distances of each weak segmentation (*i.e.* the set of rough segmentations to be fused) is first computed to obtain e_{VOI} (in addition, the entropy generated by the set of mean pairwise F_α distances of each weak segmentation allows us to obtain e_{F_α}). Then, e_{VOI} and e_{F_α} enable us to estimate the degrees of divergence of the intrinsic information related to each criterion, *i.e.* $\mathrm{d}_{\mathrm{VoI}}$ or $\mathrm{d}_{\mathrm{F}_\alpha}$ (also referred to as the inherent contrast intensity [122]), and are finally both used to compute the weight W associated with each criterion.

Conceptually, the entropy e_{VOI} or e_{F_α} defines the uncertainty of distribution of mean pairwise distances (related to each criterion). For example, if the set of weak segmentation maps to be fused have similar pairwise mean distances relative to the VoI criterion, this VoI criterion transmits too little information (relative to the other F_α criterion) to the fusion (decision maker) model [123]. As a result, the weight $\mathrm{W}_{\mathrm{VoI}}$ of this VoI criterion is less because this criterion becomes less important for our fusion model.

3.4.4 Optimization of the fusion model

To enable us to solve this consensus function, in the bi-criteria sense, we resort to a deterministic search technique, which is called the iterative conditional mode (ICM), proposed by Besag [99] (*i.e.* a Gauss-Seidel relaxation), where pixels are updated one at a time. In this work, we used a much more effective enhancement of the ICM algorithm,

which involves utilizing a superpixel (*i.e.* the regions or segments given by each individual segmentation S_k generated by the K-means algorithm) concept instead of pixels. This superpixel-based strategy makes our consensus energy function nearly convex by adding several region-based constraints (among other advantages over the pixel-based fusion method [124]). However, with the lack of proper initialization, this algorithm will converge towards a bad local minima(*i.e.* a local minima which is far away from the global minimum, and which gives a poor segmentation result).

Again, to solve this problem, we resort to the entropy values of each criteria (see (3.12)). Thus, we select the criteria which gives the minimal entropy (*i.e.* the most informative criterion ; see Section 3.4.3), and for the first iteration of the ICM, of the L segmentations to be fused, we then choose the one which ensures the minimal consensus energy (in this selected criterion sense) of our fusion model. Because this iterative algorithm amounts to achieving simultaneously, for each superpixel to be labeled, the minimum value of (3.11), we call this segmentation algorithm a multi-criteria fusion model based on the entropy-weighted formula approach (EFA-BMFM). The pseudo-code of EFA-BMFM is shown in Algorithm 1.

3.5 Experimental Tests and Results

3.5.1 Data set and benchmarks

In order to measure the performance of the proposed fusion model, we validate our approach using the famous Berkeley segmentation database (BSDS300) [18]. Recently, this dataset has been enriched to BSDS500[3] [11] with 200 additional test colour images of size 481×321. In order to quantify the efficacy of the proposed segmentation algorithm, for each colour image, the BSDS300 and the BSDS500 offer a set of benchmark segmentation results (*i.e.* ground truth), given by human observers (between 4 and 7). In addition, we used the Matlab source code proposed in [19] with the aim of estimating the different quantitative performance measures (*i.e.* the four image segmentation indices presented in Section 3.5.3). This code is available online at : http ://www.eecs.berkeley.edu/∼yang/software/lossysegmentation. In addition, to test the effectiveness for other types

Algorithm 1 EFA-Based Fusion Model algorithm

Mathematical notation:

$\overline{\text{VoI}}$	Mean VoI
$\overline{\text{F}}_\alpha$	Mean F-measure
$\overline{\text{MO}}$	Multi-objective function
$\{S_j\}_{j\leq J}$	Set of J segmentations to be fused
$\{z_j\}_{j\leq J}$	Set of weights
$\{b_j\}$	Set of superpixels $\in \{S_j\}_{j\leq J}$
$\mathcal{E}$	Set of region labels in $\{S_j\}_{j\leq J}$
T_{max}	Maximal number of iterations (=10)
$S_{I_{best}}$	Fusion segmentation result
α	F-measure compromise parameter
e_{VoI}	Entropy of the VoI criterion
$\text{e}_{\text{F}_\alpha}$	Entropy of the F-measure criterion

Input: $\{S_j\}_{j\leq J}$

Output: $S_{I_{best}}$

A. Initialization:

Compute e_{VoI} *(see (3.12))*

Compute $\text{e}_{\text{F}_\alpha}$

if $\text{e}_{\text{VoI}} < \text{e}_{\text{F}_\alpha}$ **then**

  $S_I^{[0]} = \arg\min_{S\in\{S_j\}_{j\leq J}} \overline{\text{VoI}}(S, \{S_j\}_{j\leq J})$

else

  $S_I^{[0]} = \arg\min_{S\in\{S_j\}_{j\leq J}} \overline{\text{F}_\alpha}(S, \{S_j\}_{j\leq J})$

end if

B. Steepest Local Energy Descent:

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

  for each b_j superpixel $\in \{S_j\}_{j\leq J}$ **do**

    Draw a new label x according to the uniform distribution in the set $\mathcal{E}$

    Let $S_I^{[p],\text{new}}$ the new segmentation map including b_j with the region label x

    Compute $\overline{\text{MO}}\,(S_I, \{S_j\}_{j\leq J})$ on $S_I^{[p],\text{new}}$ *(see (3.10))*

    if $\overline{\text{MO}}(S_{\text{MO}}^{[p],\text{new}}) < \overline{\text{MO}}(S_{\text{MO}}^{[p]})$ **then**

      $\overline{\text{MO}} = \overline{\text{MO}}^{\text{new}}$

      $S_I^{[p]} = S_I^{[p],\text{new}}$

      $S_{I_{best}} = S_I^{[p]}$

    end if

  end for

  $p \leftarrow p + 1$

end while

of images, we tested our proposed method on the aerial image segmentation dataset (ASD)[4] [10], and we performed a quantitative evaluation using two medical images (a brain magnetic resonance imaging (MRI) and a cornea image) recently used in [7] and [9].

3.5.2 Initial tests

Our initial tests can be divided into two main stages. First, we tested the convergence properties of our ICM procedure based on superpixels by choosing as the initialization of our iterative local gradient-descent algorithm various initializations extracted from our segmentation ensemble $\{S_k\}_{k\leq L}$ (these convergence properties have been discussed in Section 3.5.7). From our results, the final energy value, along with the resulting final segmentation map, is on average better when the initial segmentation solution is associated with an initialization chosen by our proposed entropy-based method, while it remains robust to other initializations (see Section 3.4.4 and Fig. 3.4). We also found that the average error for the PRI performance measure (on the BSDS300) is lower when the initial segmentation solution is associated with an initialization chosen by our entropy-based method (Init$-$best in Fig. 3.5).

Secondly, we tested the effect of the number of initial segmentations on the accuracy of the final segmentation result. Qualitatively, Fig. 3.6 shows that the final consensus result is even better than the size of the segmentation ensemble L is high. Quantitatively, we observed that the different performance measures (see Section 3.5.3) are improved when we increase the number of initial segmentations. This test demonstrates the validity of the proposed fusion procedure, and shows that the segmentation results can be enhanced if the segmentation ensemble is completed by other segmentation maps of the same scene.

[3] The BSDS300 [18] and the BSDS500 [11] are available online at :
https ://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html

[4] The ASD [10] is available online at :
http ://web.ornl.gov/∼jiy/ASD/Aerial Image Segmentation Dataset.html

FIGURE 3.4 : Example of fusion convergence result for three various initializations. (a) Berkeley image ID 229036 and its ground-truth segmentations. (b) A non informative (or blind) initialization. (c) The worst input segmentation. (d) The best input segmentation (from the segmentation set) selected by the entropy method (see Section 3.4.4). (e), (f) and (g) segmentation results after 10 iterations of our EFA-BMFM fusion model (resulting from (b), (c) and (d), respectively).

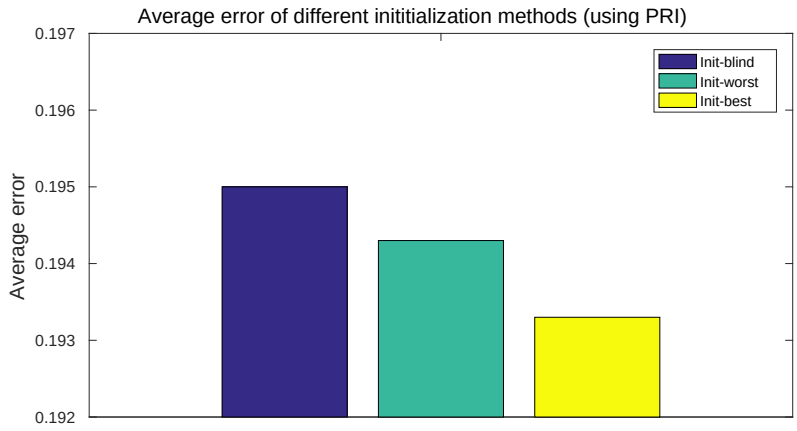


FIGURE 3.5 : Average error of different initialization methods (for the probabilistic Rand index (PRI) performance measure) on the BSDS300.

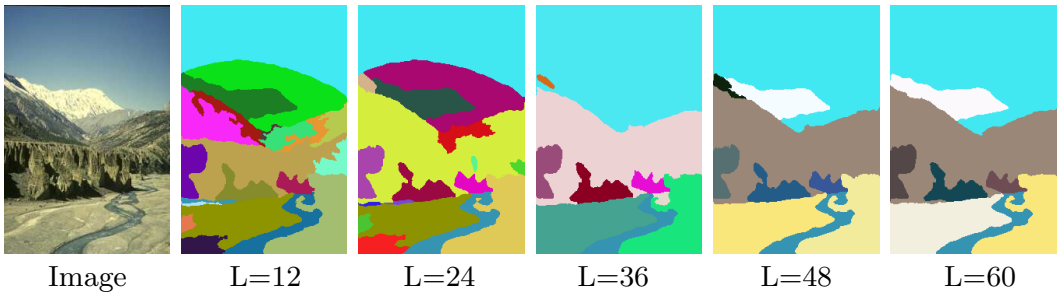


FIGURE 3.6 : Progression of the segmentation result as a function of the number of segmentations (L) to be fused for the EFA-BMFM algorithm. More precisely, for L= 12, 24, 36, 48 and 60 segmentations.

3.5.3 Performance measures and results

In an attempt to test and evaluate our fusion segmentation model, we employed four performance metrics which are most popular in the literature. These well-known performance measures [5] are :

1. The Probabilistic Rand index (PRI) [103] counts the fraction of pairs of pixels whose labels are consistent between the computed segmentation and the human segmentation, averaging through all of the ground-truth segmentation of a given image.

2. The boundary displacement error (BDE) [107] measures the average displacement error of boundary pixels between two segmented images. In particular, it defines the error of one boundary pixel as the distance between the pixel and the closest pixel in the other boundary image.

3. The variation of information (VoI) [106] defines the distance between two segmentations as the average conditional entropy of one segmentation given the other ; it measures the amount of information which is lost or gained while switching from one region to another.

4. The global consistency error (GCE) [18] determines the extent to which one segmentation map can be viewed as a refinement of another segmentation map. In this way, for a perfect match, every region in one of the segmentations must be a refinement (*i.e.*, a subset) of a region in the other segmentation.

As can be seen from the results given in Table 3.1 and Table 3.2, for the BSDS300, our method generally outperforms the state-of-the art algorithms in terms of the different distance measures with : BDE= 8.284, VoI= 1.870, GCE= 0.198 (a lower value

[5] The GCE metric is in the range [0 ;1], where a score of 0 indicates that there is a perfect match between two segmentations and an error of 1 represents a maximum difference between the two segmentations to be compared [5]. Also, the PRI metric is in the range [0 ;1], where higher values indicating greater similarity between two segmentations [75]. For the BDE measure, a value near-zero indicates high quality of the image segmentation, and its maximum value can be the length of the image segmentation [107]. The VOI metric taking a value of 0 when two segmentations are identical and positive otherwise. This metric is in the range [0 ;log(n)], where *n* denotes the number of pixels within the image [76].

is better) and PRI= 0.806 (a higher value is better). From the tables, we also see that if we compare our results to a mono-objective approach (FMBFM and VOIBFM) based on the same single criterion, we obtain significantly better results. This shows clearly that our strategy of combining two complementary (contour and region-based) criteria of segmentation (the VoI and the F-measure) is effective. In addition, from the data in Table 3.3 and Table 3.4, we observe that for the BSDS500, our method gives comparable performance results compared to different algorithms with or without the fusion model when : BDE= 7.90, VoI= 1.97, GCE= 0.21 (a lower value is better) and PRI= 0.81. Moreover, Fig. 3.7, we observe that the PRI and VoI performance scores are better when L (the segmentation number to be fused) is high. This test shows that our performance scores can be further improved if we increase the number of segmentations to be fused. In addition, for better comparison, in Fig. 3.8, we present a sample of results obtained by applying our algorithm to some images from the Berkeley dataset compared to other state-of-the-art algorithms. In addition, Fig. 3.9 displays a small number of segmented images which are similar to those shown in the mono-criterion fusion model (FMBFM and VOIBFM) proposed in [76] and [77], respectively. Fig. 3.10 shows the best and worst segmentation results (in the PRI sense) from the BSDS300. The results for the entire database will be available on the website of the author. Fig. 3.11 shows the distribution of the PRI, BDE, VoI and GCE measures. From this figure, we can conclude that few segmentations exhibit poor PRI and BDE scores even for the most difficult segmentation cases. Moreover, Fig. 3.12 shows the distribution of the number and size of regions obtained by our EFA-BMFM algorithm over the BSDS300.

3.5.4 Comparison of medical image segmentation

Medical image segmentation is an important part of medical analysis, and is also a process which is clearly different from the segmentation of natural (textured colour) images because input medical images are generally in grey levels, have low contrast and are noisy. We performed two experiments on medical images to demonstrate the effectiveness and flexibility of our segmentation approach. In the first experiment, we used a brain magnetic resonance imaging (MRI), as shown in Fig. 3.13. The results, which were

TABLE 3.1 : Performance of several segmentation algorithms (with or without a fusion model strategy) for three different performance measures : VoI, GCE and BDE (lower is better), on the BSDS300.

ALGORITHMS	**VoI**	**GCE**	**BDE**
HUMANS	1.10	0.08	4.99
Algorithms With Fusion Model			
EFA-BMFM	1.870	0.198	8.284
-2016- GCEBFM [5]	2.10	0.19	8.73
-2014- FMBFM [77]	2.01	0.20	8.49
-2014- VOIBFM [76]	1.88	0.20	9.30
-2014- SFSBM [113]	2,21	0,21	8,87
-2010- PRIF [75]	1.97	0.21	8.45
-2008- FCR [2]	2.30	0.21	8.99
-2007- $\text{CTM}_{\gamma=20}$ [19]	2.02	0.19	9.90
Algorithms Without Fusion Model			
-2016- DGA-AMS [125]	2,03	-	-
-2014- CRKM [126]	2.35	-	-
-2012- MDSCCT [6]	2,00	0,20	7,95
-2012- AMUS [74]	1,68	0,17	-
-2011- KM [44]	2.41	-	-
-2011- MD2S [4]	2.36	0.23	10,37
-2010- SCKM [3]	2,11	0,23	10,09
-2009- MIS [46]	1,93	0,19	7,83
-2009- HMC [36]	3,87	0,30	8,93
-2008- NTP [41]	2,49	0,24	16,30
-2008- Av. Diss [42]	2,62	-	-
-2005- $\text{NCuts}_{K=20}$ [57] (in [19])	2.93	0.22	9.60
-2004- $\text{FH}_{\Sigma=0.5,k=500}$ [12] (in [19])	2,66	0,19	9,95
-2002- Mean-Shift [14]	2.48	0.26	9.70

TABLE 3.2 : Performance of several segmentation algorithms (with or without a fusion model strategy) for the PRI performance measure (higher is better) on the BSDS300.

ALGORITHMS	**PRI**
HUMANS	0.87
Algorithms With Fusion Model	
EFA-BMFM	0.806
-2016- GCEBFM [5]	0.80
-2014- FMBFM [77]	0.80
-2014- VOIBFM [76]	0.81
-2014- SFSBM [113]	0,79
-2010- PRIF [75]	0.80
-2009- Consensus [73]	0,78
-2008- FCR [2]	0,79
-2007- $\text{CTM}_{\gamma=20}$ [19]	0.76
Algorithms Without Fusion Model	
-2016- LSI [127]	0.80
-2014- CRKM [126]	0.75
-2011- SCKM [3]	0,80
-2011- MD2S [4]	0,78
-2011- KM [44]	0,76
-2009- MIS [46]	0,80
-2009- HMC [36]	0,78
-2009- Total Var [43]	0,78
-2009- A-IFS HRI [29]	0,77
-2008- CTex [25]	0,80
-2004- $\text{FH}_{\Sigma=0.5,k=500}$ [12] (in [19])	0,78
-2005- $\text{NCuts}_{K=20}$ [57] (in [19])	0.72
-2002- Mean-Shift [14]	0.75
-2001- $\text{JSEG}_{c=255,s=1.0,m=0.4}$ [15] (in [25])	0,77

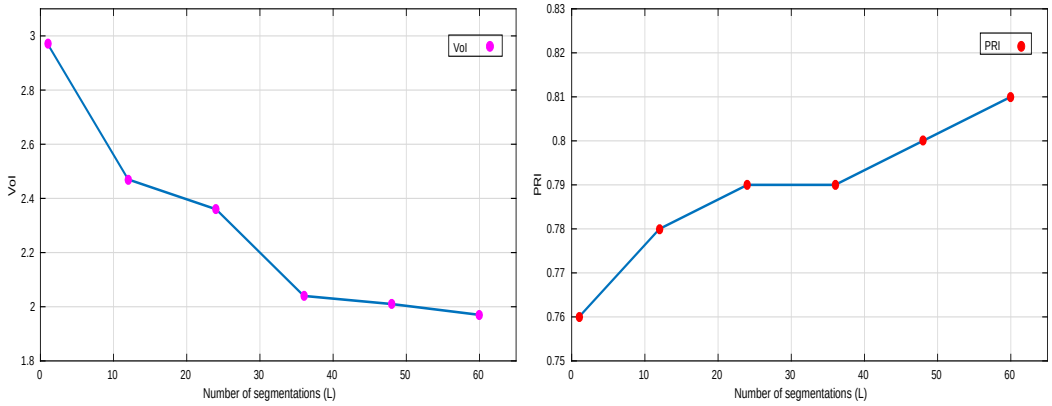


FIGURE 3.7 : Progression of the VoI, (lower is better) and the PRI (higher is better) according to the segmentation number (L) to be fused for our proposed EFA-BMFM algorithm (on the BSDS500). Precisely, for L = 1, 12, 24, 36, 48 and 60 segmentations.

TABLE 3.3 : Performance of several segmentation algorithms (with or without a fusion model strategy) for three different performance measures : VoI, GCE and BDE (lower is better), on the BSDS500.

ALGORITHMS	**VoI**	**GCE**	**BDE**
HUMANS	1.10	0.08	4.99
Algorithms With Fusion Model			
EFA-BMFM	1.97	0.21	7.90
-2016- GCEBFM [5]	2.18	0.20	8.61
-2014- FMBFM [77]	2.00	0.21	8.19
-2014- VOIBFM [76]	1.95	0.21	9.00
-2010- PRIF [75]	2.10	0.21	8.88
-2008- FCR [2]	2.40	0.22	8.77
-2007- CTM [19] (in [128])	1.97	-	-
Algorithms Without Fusion Model			
-2011- $WMS_{d_\Lambda=20}$ [129] (in [128])	2.10	-	-
-2004- $FH_{\Sigma=0.8}$ [12] (in [128])	2.18	-	-
-2002- Mean-Shift [14] (in [128])	2.00	-	-

TABLE 3.4 : Performance of several segmentation algorithms (with or without a fusion model strategy) for the PRI performance measure (higher is better) on the BSDS500.

ALGORITHMS	**PRI**
HUMANS	0.87
Algorithms With Fusion Model	
EFA-BMFM	0.81
-2016- GCEBFM [5]	0.80
-2014- FMBFM [77]	0.80
-2014- VOIBFM [76]	0.80
-2010- PRIF [75]	0.79
-2008- FCR [2]	0,79
-2007- CTM [19] (in [128])	0.73
Algorithms Without Fusion Model	
-2004- $FH_{\Sigma=0.8}$ [12] (in [128])	0,77
-2011- $WMS_{d_A=20}$ [129] (in [128])	0.75
-2002- Mean-Shift [14] (in [128])	0.77

FIGURE 3.8 : A sample of results obtained by applying our proposed algorithm to images from the Berkeley dataset compared to other algorithms. From left to right : original images, FCR [2], SCKM [3], MD2S [4], GCEBFM [5], MDSCCT [6] and our method (EFA-BMFM).

FIGURE 3.9 : Additional segmentation results obtained from the BSDS300.

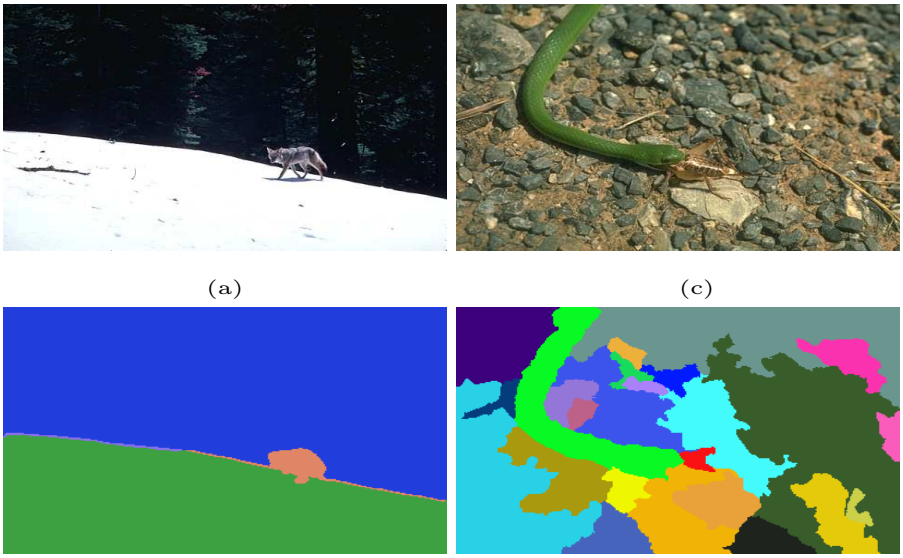


FIGURE 3.10 : Best and worst segmentation results (in the PRI sense) obtained from the BSDS300. First column : (a) image ID 167062 and (b) its segmentation result (PRI=0.99). Second column : (c) image ID 175043 and (d) its segmentation result (PRI = 0.37).

obtained by using the region-based model via local similarity factor (RLSF), the global active contour model (these two models which are based on active contour were recently proposed in [7]) and our EFA-BMFM model, are shown in Fig. 3.13 (b)-(d), respectively. As can be seen, our method outperforms the global active contour model and gives an interesting result compared to the segmentation achieved by the RLSF model. In the second experiment, we tested our model on a real cornea image, and we compared the segmentation result provided by our EFA-BMFM model with the results given by the fast global minimization (FGM) [8] and the double fitting terms of multiplicative and difference (DMD) [9] models (see Fig. 3.14). We observe that the quality of the segmentation obtained by the FGM model for this cornea image is not as good as those of the DMD and EFA-BMFM. The reason for this is (as mentioned in [9]) that the image with intensity inhomogeneity is too challenging for the FGM.

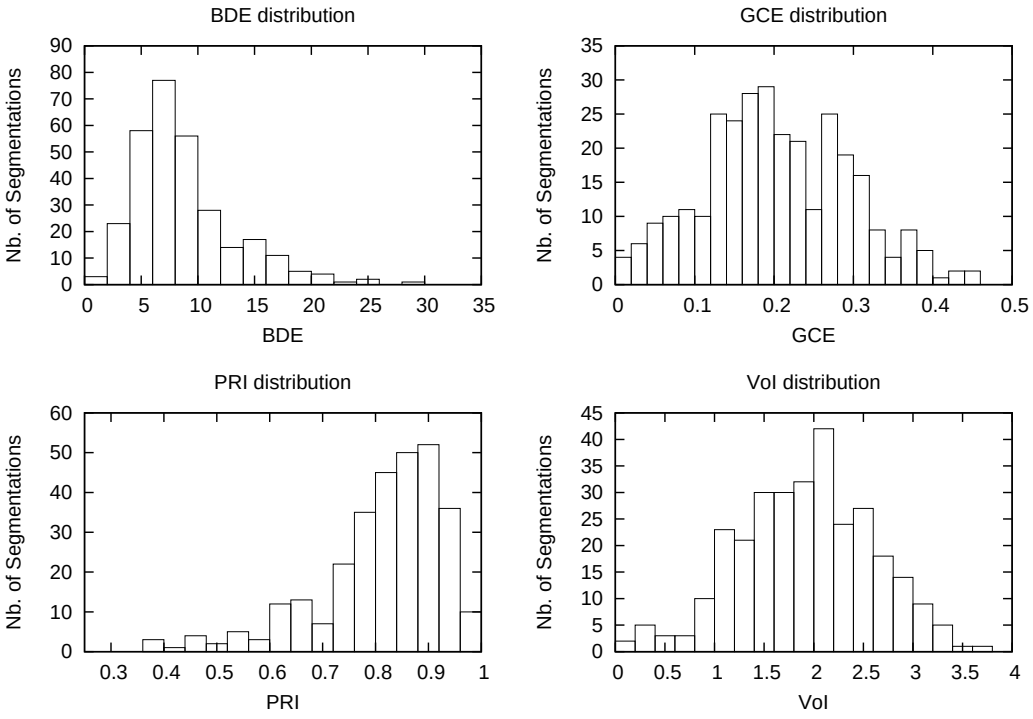


FIGURE 3.11 : Distribution of the BDE, GCE, PRI and VoI measures over the 300 segmented images of the BSDS300.

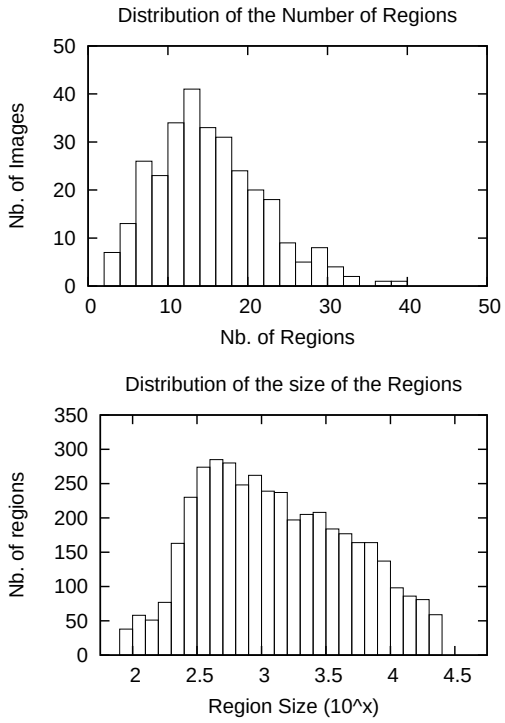


FIGURE 3.12 : Distribution of the number and size of regions over the 300 segmented images of the BSDS300.

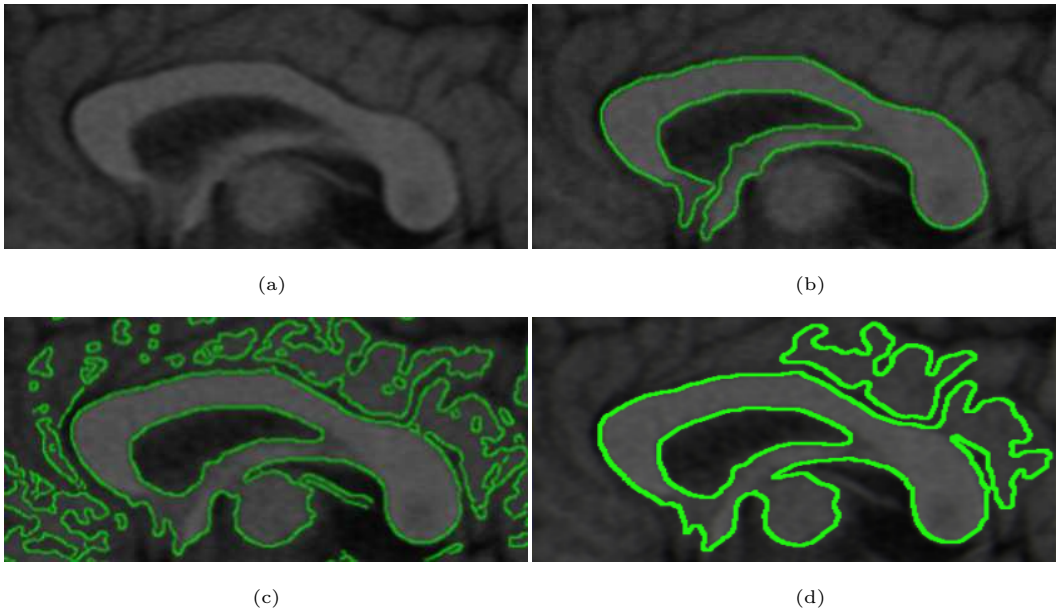
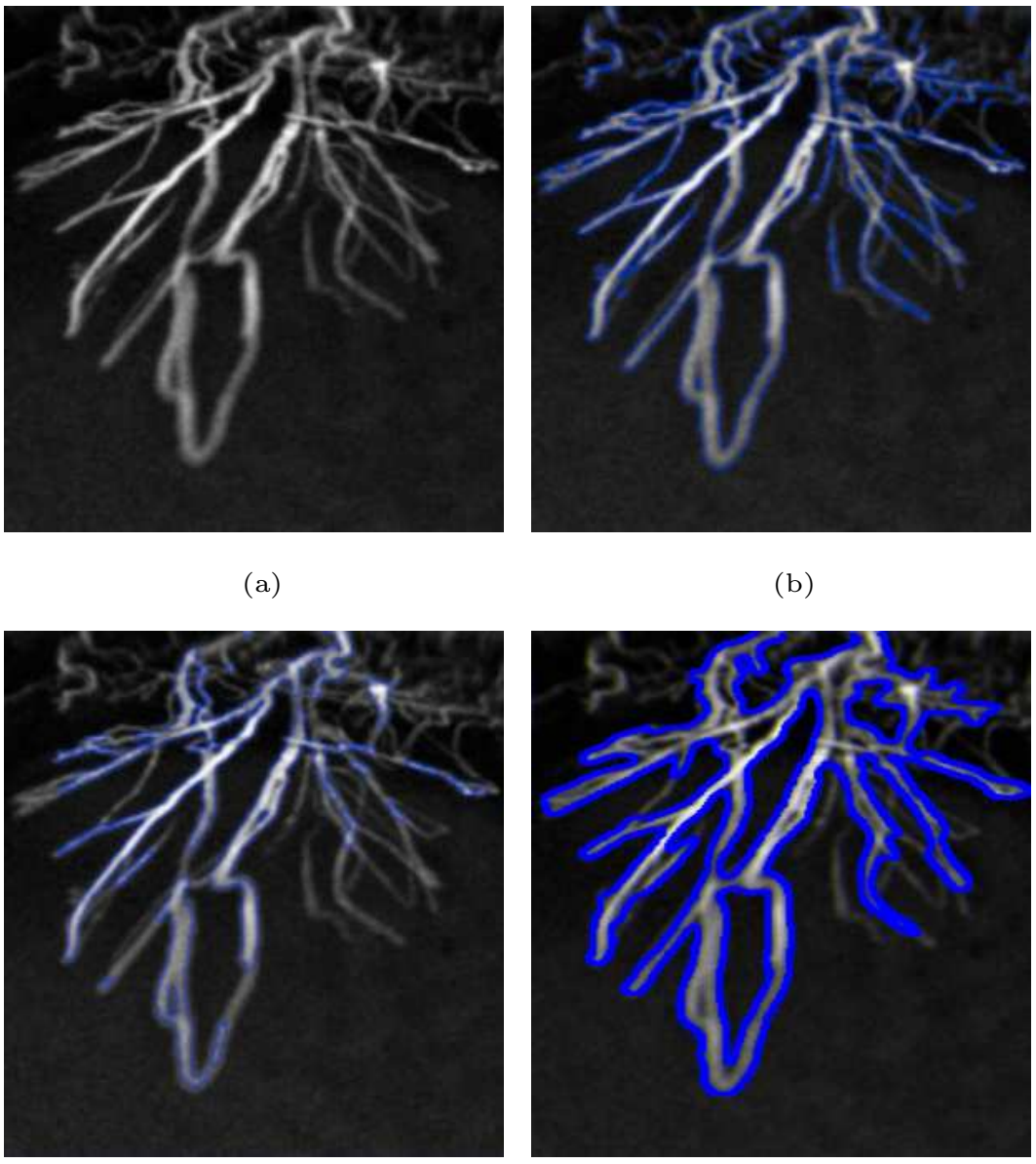


FIGURE 3.13 : Comparison of two region-based active contour models on a brain MRI. (a) original image. (b) segmentation of the RLSF model [7]. (c) segmentation of the global active contour model [7]. (d) segmentation achieved by our EFA-BMFM model.

3.5.5 Comparison of Segmentation Methods for Aerial Image Segmentation

We also benchmarked our fusion model as a segmentation method using the aerial image segmentation dataset (ASD) [10]. This new image dataset contains 80 high-resolution aerial images, with spatial resolutions ranging from 0.3 to 1.0 m, including different scenes as schools, residential areas, cities, warehouses and power plants. The images were normalized to realize a resolution of 312×312 pixels, and the segmentation results were then super-sampled in order to obtain segmentation images with the original resolution (512×512 pixels).

Table 3.5 shows the overall F-measure of different segmentation algorithms under two different scale settings. The first is the score under the optimal data set scale (ODS), and the average F-measure of 80 images at each scale is calculated and the best measure across scales is reported. The second is the score under the optimal image scale (OIS), which uses the best F-measure across scales for each image, and the average measure over images is reported[6]. As can be seen from the data on Table 3.5, our method out-



(a) (b)

(c) (d)

FIGURE 3.14 : Comparison of two segmentation methods on segmenting a real cornea image. (a) original image of size 256×256. (b) detection using the FGM method [8] (5000 iterations). (c) detection using the DMD method [9] (5 iterations). (d) detection resulting from our EFA-BMFM model (10 iterations).

performs the VOIBFM fusion model in terms of both the ODS and OIS, and it remains generally competitive compared to segmentation algorithms without a fusion strategy. In addition, and for better comparison, samples of the results obtained by applying our algorithm to some images from the ASD dataset compared to other state-of-the art algorithms are given in Fig. 3.15.

3.5.6 Algorithm complexity

With respect to the time complexity, the first step of our algorithm (the generation of the initial ensemble of segmentations) has a complexity equal to $O(N \cdot K \cdot I \cdot d)$, where N, K, I and d are the number of points of each cluster, the number of clusters, the number of iterations and the dimension of each point to be clustered, respectively. Moreover, the second step (fusion algorithm) is characterized by a complexity time of $O(Nbsup \cdot n)$, where n is the pixel number within the image and $Nbsup$ represents the number of superpixels existing in the set of segmentations to be fused (see Table 3.6 for a comparison with other methods).

As another important aspect, in terms of the execution time, the segmentation operation takes on average about 240 s for an Intel 64 Processor core i7-4800M Q, 2.7 GHz with 8 GB of RAM memory and non-optimized code running on Linux ; on average, it takes 60 s to generate the segmentation ensemble and approximately 180 s for the fusion step and for a 320 $\times$ 214 image (Table 3.7 compares the average computational time for an image segmentation and for different segmentation algorithms whose PRI is greater than 0.76). Further, it is important to note that the algorithm can easily be parallelized (using the parallel capabilities of a graphic processor unit) because its two steps (described above) are purely independent. Finally, to enable comparisons with future segmentation methods, the source code (in C++ language) of our model and the ensemble of segmented images are publicly accessible here : http ://www-etud.iro.umontreal.ca/∼khelifil/ResearchMaterial/efa-bmfm.html.

[6] The soft contour map is provided by averaging, 6 times, the set of hard (*i.e.* binary) boundary representations of our segmentation method with different values of K_{max} (the number of classes of the segmentation).

TABLE 3.5 : Boundary benchmarks on the aerial image segmentation dataset (ASD). Results obtained for different segmentation methods (with or without the fusion model strategy). The figure shows the F-measures (higher is better) when choosing an optimal scale for the entire dataset (ODS) or per image (OIS).

ALGORITHMS	**ODS**	**OIS**
HUMANS	0.68	0.69
Algorithms Without Fusion Model		
FH [12]	0.59	0.62
SRM [13]	0.58	0.60
Mean shift [14]	0.56	0.58
JSEG [15]	0.54	0.56
FSEG [16]	0.58	0.61
MSEG [17]	0.53	0.57
Algorithms With Fusion Model		
EFA-BMFM	0.50	0.50
VOIBFM [76]	0.36	0.36
FMBFM [77]	0.53	0.53

TABLE 3.6 : Fusion segmentation models and complexity.

	EFA-BMFM	GCEBFM [5]	VOIBFM [76]
K-means step (generation of initial segmentations)	$O(N \times K \times I \times d)$	$O(N \times K \times I \times d)$	$O(N \times K \times I \times d)$
Fusion step	$O(Nbsup \times n)$	$O(Nbsup \times n)$	$O(n)$

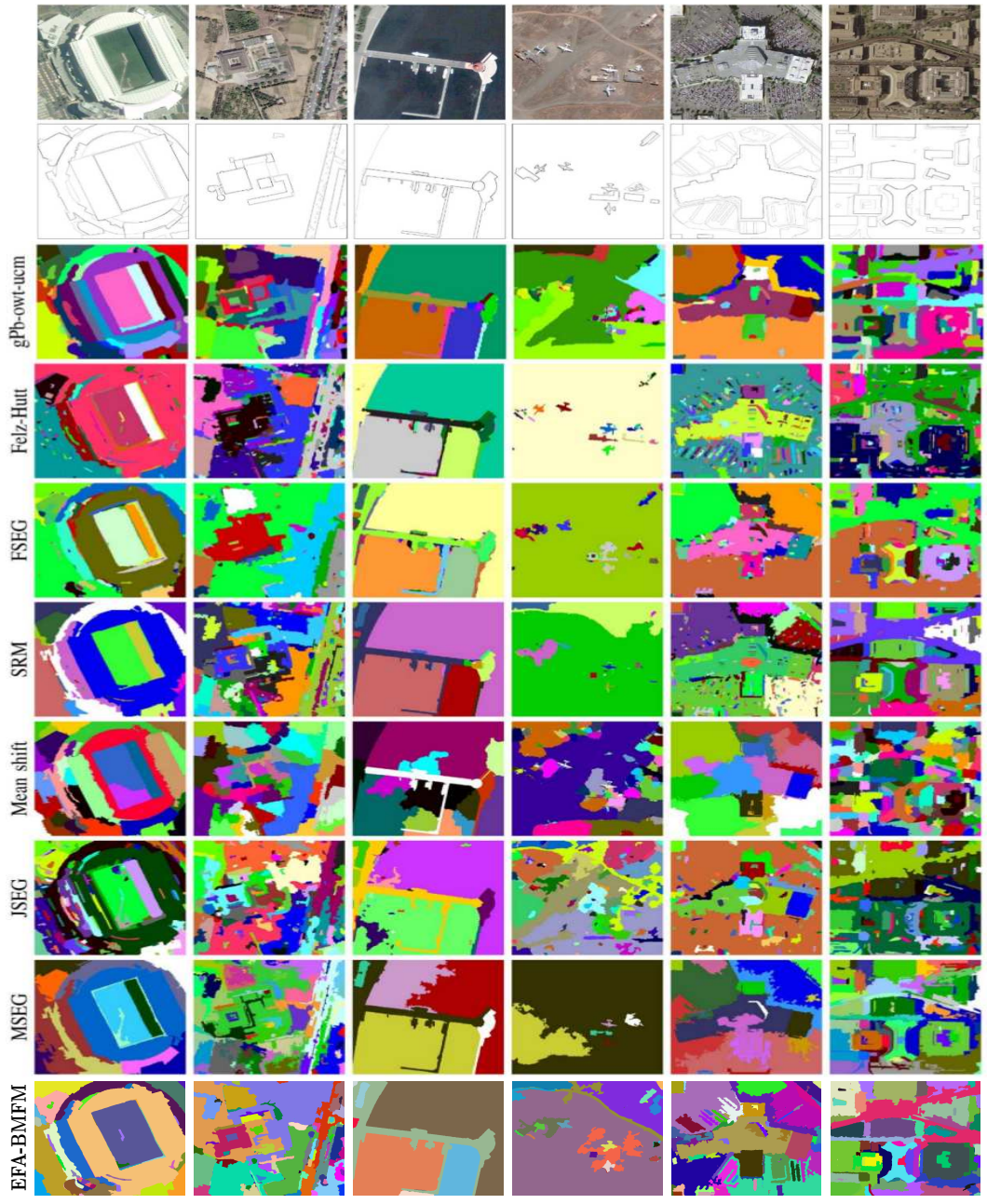


FIGURE 3.15 : A sample of results obtained by applying our algorithm to images from the aerial image dataset [10] compared to other popular segmentation algorithms (gPb-owt-ucm [11], Felz-Hutt (FH) [12], SRM [13], Mean shift [14], JSEG [15], FSEG [16] and MSEG [17]). The first row shows six example images. The second row overlays segment boundaries generated by four subjects, where the darker pixels correspond to the boundaries marked by more subjects. The last row shows the results obtained by our method (EFA-BMFM).

TABLE 3.7 : Average CPU time for different segmentation algorithms for the BSDS300.

ALGORITHMS	**PRI**	**CPU time (s)**	**[image size]**
-EFA-BMFM-	0,80	$\simeq 240$	[320 $\times$ 214]
-GCEBFM- [5]	0,80	$\simeq 180$	[320 $\times$ 214]
-VOIBFM- [76]	0,81	$\simeq 60$	[320 $\times$ 214]
-FMBFM- [77]	0,80	$\simeq 90$	[320 $\times$ 214]
-CTM- [19]	0,76	$\simeq 180$	[320 $\times$ 200]
-PRIF- [75]	0,80	$\simeq 20$	[320 $\times$ 214]
-FCR- [2]	0,79	$\simeq 60$	[320 $\times$ 200]
-MDSCCT- [6]	0,81	$\simeq 60$	[320 $\times$ 214]
-CTex- [25]	0,80	$\simeq 85$	[184 $\times$ 184]
-HMC- [36]	0,78	$\simeq 80$	[320 $\times$ 200]
-LSI- [127]	0,80	$\simeq 60$	[481 $\times$ 321]

3.5.7 Discussion

The most obvious finding to emerge from the above analysis is that the use of the MO optimization concept enables us to design a new fusion model that takes advantages of the complementarity of different segmentation criteria.

This interesting model appears to be very competitive for different kinds of performance measures, and it therefore appears as an alternative to complex and computationally demanding segmentation models which exist in the literature. Moreover, another possible alternative analysis is given in Table 3.8. In fact, from this table, we can confirm that the performance measures are quite different for a given image compared to the values obtained by other approaches. Thus, our model outperforms the VOIBFM [76] fusion model and the MDSCCT [6] algorithm (a purely algorithmic approach), in terms of the number of images of the BSDS300 which obtain the best GCE, BDE and PRI scores. These results provide further support for the hypothesis that our model appears to be very competitive against other methods with or without a fusion strategy. Compared to the mono-objective approach, the combination of two objectives makes our

fusion algorithm slower, confirming the hypothesis in [130], and indicating that a high number of objectives cause additional challenges. However, it appears that the choice of using super-pixels with the ICM (as an optimization algorithm) limits this problem as the execution time remains close to those of other algorithms. In this context, we present a convergence analysis of a Berkeley colour image, shown in Fig. 3.16. Fig. 3.16 shows (a) the original Berkeley image ID 187039 selected from the BSDS300, (b) the evolution of the segmentation map of our EFA-BMFM fusion model starting from a blind (or noninformative) initialization and (c) the evolution of the consensus energy function along the number of iterations of the EFA-BMFM. In Fig. 3.16 (c), we observe that our EFA-BMFM model converged to a minimum energy value after 5 iterations. It should be noted that this faster convergence speed of our model resulted from the use of superpixels.

As mentioned in Section 3.1, to date, there have been no reports of the application of current knowledge of MO optimization to the field of the fusion of colour image segmentation. These interesting results provided by our model are related both to the generality and the relative applicability of this MO concept with different segmentation criteria.

TABLE 3.8 : Comparison of scores between the EFA-BMFM and other segmentation algorithms for the 300 images of the BSDS300. Each value indicates the number of images of the BSDS300 which obtain the best score.

MEASURES	EFA-BMFM Vs GCEBFM [5]		EFA-BMFM Vs MDSCCT [6]		EFA-BMFM Vs VOIBFM [76]	
	EFA-BMFM	GCEBFM	EFA-BMFM	MDSCCT	EFA-BMFM	VOIBFM
GCE	**216**	84	**261**	39	**167**	133
VOI	143	**157**	122	**178**	134	**166**
BDE	**151**	149	**175**	125	**201**	99
PRI	147	**153**	**167**	133	**160**	140

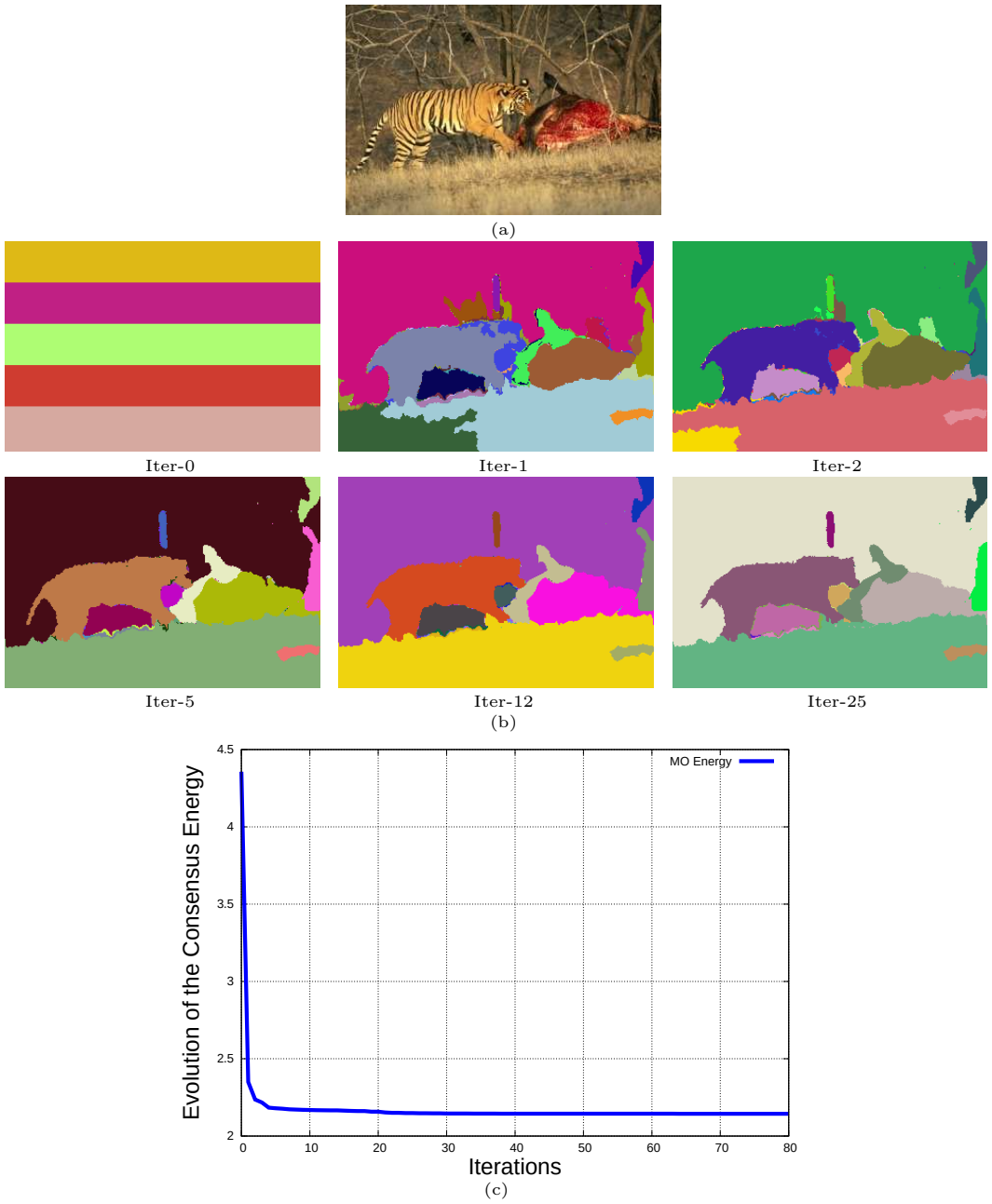


FIGURE 3.16 : Convergence analysis. (a) input image ID 187039 selected from the BSDS300. (b) change of the segmentation map of our EFA-BMFM fusion model starting from a blind (or non informative) initialization. (c) evolution of the consensus energy function along the number of iterations of the EFA-BMFM.

3.6 Conclusion

In this paper, we present a new and efficient multi-criteria fusion model based on the entropy-weighted formula approach (EFA-BMFM). The proposed model combines multiple segmentation maps to achieve a final improved segmentation result. This model is based on two complementary (contour and region-based) criteria of segmentation (the VoI and the F-measure criteria). We applied the proposed segmentation model to BSDS300, BSDS500, ASD and medical images, and the proposed model appears to be comparable to or even outperform other segmentation models, which proves the effectiveness and robustness of our multi-criteria fusion approach. In our model, the fusion process is performed at three different conceptual and hierarchical levels ; first, at the criterion level, because the proposed fusion model combines two conflicting and complementary criteria ; second, at the (segmentation) decision level by exploiting the combination of different and weak segmentations of the same image (expressed in different colour spaces) ; third, at the (pixel-)data level, and this is done by considering the set of superpixels as the atomic elements to be segmented in the consensus segmentation (instead of the set of pixels). Although our current multi-criteria fusion model is reasonably efficient and the superpixel strategy makes our energy function nearly convex, it would be interesting to optimize the consensus function with other optimization algorithms such as the exploration/selection/estimation (ESE) [131] or genetic algorithms. Thus, these algorithms are guaranteed to find the optimal solution ; however, they have the drawback of a huge computational time. To overcome this problem, we can use the parallel computing capabilities of a graphic processor unit (GPU) (based on its massively parallel architecture consisting of thousands of smaller, which are designed to handle multiple tasks simultaneously). For all these reasons, the proposed fusion method may therefore be seen as an attractive strategy for solving the difficult image segmentation problem.

CHAPITRE 4

A MULTI-OBJECTIVE DECISION MAKING APPROACH FOR SOLVING THE IMAGE SEGMENTATION FUSION PROBLEM

Cet article a été publié dans le journal *IEEE Transactions on Image Processing* comme l'indique la référence bibliographique

L. Khelifi, M. Mignotte. A Multi-objective Decision Making Approach for Solving the Image Segmentation Fusion Problem

IEEE Transactions on Image Processing (TIP), 26(8) :3831-3845, Août 2017.

Cet article est presenté içi dans sa version originale.

Abstract

Image segmentation fusion is defined as the set of methods which aim at merging several image segmentations, in a manner that takes full advantage of the complementarity of each one. Previous relevant researches in this field have been impeded by the difficulty in identifying an appropriate single segmentation fusion criterion, providing the best possible, i.e., the more informative, result of fusion. In this paper, we propose a new model of image segmentation fusion based on multi-objective optimization which can mitigate this problem, to obtain a final improved result of segmentation. Our fusion framework incorporates the dominance concept in order to efficiently combine and optimize two complementary segmentation criteria, namely, the global consistency error (GCE) and the F-measure (precision-recall) criterion. To this end, we present a hierarchical and efficient way to optimize the multi-objective consensus energy function related to this fusion model which exploits a simple and deterministic iterative relaxation strategy combining the different image segments. This step is followed by a decision making task based on the so-called "technique for order performance by similarity to ideal solution" (TOPSIS). Results obtained on two publicly available databases with manual ground truth segmentations clearly show that our multi-objective energy-based model gives better results than the classical mono-objective one.

4.1 Introduction

Image segmentation is one of the most crucial components of image processing and pattern recognition system whose aim is to represent the image content into different regions of coherent properties with homogeneous characteristics such as texture, color, movement and boundary continuity [132]. This pre-treatment is crucial because the resulting segments form the basis for the subsequent classification, which may be based on spectral, structural, topological, and/or semantic features [133, 134].

In order to solve the difficult unsupervised segmentation problem, different strategies have been proposed in the past [135, 136]. Among them, one can mention the region based segmentation which in fact assumes that neighboring pixels within the same region should have similar values [137] and more precisely segmentation models exploiting directly clustering schemes [3, 25] using Gaussian mixture modeling, fuzzy clustering approaches [27, 28] or fuzzy sets [29], region growing strategies [15], compression models [33], wavelet transform [34] or watershed transformation [31], Bayesian [38], or texton-based approaches [39], graph-based [12, 40, 41], deformable surfaces [46], or active contour model [47] or genetic algorithm [52] and spectral clustering [57], just to mention a few.

Another line of work has recently become the focus of considerable interest, which suggests that an improved segmentation result can be achieved through the combining of multiple, quickly estimated and weak segmentation maps of the same scene. To the best of our knowledge, Jiang et al. [61] was the first to investigate this merging strategy based on a defined criterion, but this approach has suffered from a constraint related to the initial segmentations which should include the same regions number. Afterward, this approach has also been implemented without this restriction, with an arbitrary number of regions [2, 63].

Fusion of segmentation has been extensively studied, in particular with respect to a single criterion. However, an inherent weakness of the mono-criterion based fusion model comes from the facts that, the segmentation is inherently an ill-posed problem related to the large number of possible partitioning solutions for any image, and also, by

the fact that a single criterion cannot model all the geometric properties of a segmentation solution or otherwise said, the single criterion optimization process is only dedicated to exploring a subset or a specific region of the search space.

Thus, a key problem with much of the literature on the fusion of segmentation consists in choosing the most appropriate criterion able to generate the best segmentation result. Motivated by the above observations, in this work, we focus on proving that a fusion model of segmentation, expressed as a multi-objective optimization problem, with respect to a combination of different and complementary criteria, is an interesting approach that can overcome the limitations of a single criterion and give a competitive final segmentation result for different images with several distinct texture types. In addition, the proposed strategy can be also viewed as a general framework for combining several *a priori* energy terms in any energy-based models or several prior distributions in a possible Bayesian multi-objective framework.

The remainder of this paper is organized as follows. In Section 4.2 we discuss the literature review concerning the fusion models of segmentations. In Section 4.3 we describe our proposed fusion model ; we start by introducing basic concepts about multi-objective optimization in the first part of the section, in the second part we define the two criteria used in our model, in the third part we present the multi-objective function relating to this novel fusion framework, in the fourth part we describe the optimization strategy used to minimize our multi-objective function and in the fifth part we outline the decision making method adopted for the selection of the best solution from an ensemble of non-dominated solutions. In Section 4.4 we describe the generation of the segmentation set to be combined by our model. In Section 4.5 we illustrate a set of experimental results and comparisons with existing segmentation algorithms. In this section, our strategy of segmentation is validated on two publicly available databases. Finally, we conclude the paper in Section 4.6.

4.2 Literature Review

In the literature, there are several examples of new fusion algorithms, which all solve the segmentation problem based on a single criterion. Here we only give a brief review of some popular criteria.

One of the first implementation of the fusion of region-based segmentations of the same scene was carried out by Mignotte [2], who proposed the merging of the initial input segmentations in the within-cluster variance sense, since the obtained segmentation result was achieved by exploiting a fusion scheme based on K-means algorithm. This fusion framework remains simple and fast, however, the final segmentation result closely depends on the distance choice and the value of K used in the final K-means based fusion procedure. Following this strategy, we can also mention the fusion model suggested by Harrabi et al. [72], which adopted the same approach, but for the set of local *soft* labels estimated with a multilevel thresholding scheme and for which the fusion procedure is thus provided in the sense of the weighted within-cluster inertia, with the same disadvantages of the previous method while requiring more computational time for estimating the mass functions of the information's to be combined.

Another widely used criterion is the Rand index [70] (RI) which was first used in [59], with the idea of evidence accumulation in a hierarchical agglomerative clustering model, for combining the results of multiple conventional clusterings. This RI measure of agreement can be also used in the case of two segmentations, by encoding the set of constraints, in terms of pairs of pixel labels (identical or not), achieved by each of the segmentations to be fused. This idea has been first proposed by [63] with a random walking stochastic approach and associated with an estimator based on mutual information to estimate the optimal regions number, and later by Ghosh et al. [73] with an algebraic optimization based fusion model using non-negative matrix factorization. The penalized version of the RI criterion has also been used in [75], by adding a global constraint on the fusion process, which restricts the size and the number of the regions, within a Markovian framework and an analytical optimization method and by Alush et al. [74] exploiting a constrained version of this RI criterion by an expectation maximization (EM)

algorithm applied on super-pixels preliminary provided by an over-segmentation process. The main drawback of the Rand Index criterion is due to its quadratic complexity in terms of data set size since it uses all pairs of pixel, and in terms of algorithm complexity of the fusion model.

Fusion of segmentation maps can also be accomplished with the entropy, or more precisely in the variation of information (VoI) sense [76] with an energy-based model optimized by exploiting an iterative steepest local energy descent strategy combined with a connectivity constraint. This criterion is interesting but some studies have shown than it is less correlated with human segmentation in term of visual perception compartively to the RI or the least square or within-cluster inertia criterion. It is also important to mention the fusion scheme proposed by Ceamanos et al. [78], which is based on the maximum-margin hyperplane sense and in which the hyperspectral image is segmented according to the decision fusion of multiple and individual support vector machine classifiers that are trained in different feature subspaces emerging from a single hyperspectral data set. Similarly, Song et al. [79] presented a recent Bayesian fusion procedure for satellite image segmentation, in which class labels obtained from different segmentation maps are fused by the weights of an evidence model which estimates each final class label with the maximum logit posterior odd. Recently, Khelifi et al. [5] proposed the fusion of multiple segmentation maps according to the global consistency criterion (GCE). In this metric sense, which measures the extent to which one segmentation map can be viewed as a refinement of another segmentation, a perfect correspondence is obtained if each region in one of the segmentation is a subset or geometrically similar to a region in the other segmentation.

It is important to mention, that all these above-described studies treat the image segmentation fusion problem with a single criterion. However, the major problem of the mono-criterion based fusion model comes from the fact that, the segmentation is inherently an ill-posed problem related to the large number of possible partitioning solutions for any image, and also, that a single criterion cannot model all the geometric properties of a segmentation solution or otherwise said, the single criterion optimization process is only dedicated to exploring a subset or a specific region of the search space.

The fusion model outlined in this work is called multi-objective optimization based-fusion model (MOBFM). The motivation of using multi-objective optimization is to design a new segmentation fusion model that takes advantage of the complementarity of different objectives to achieve a final better segmentation. Besides, in order to better constrain and to improve the optimization process, we resort to the iterative conditional modes (ICM) algorithm applied on pre-estimated super-pixel to be labeled. To this end, we have incorporated, in the ICM-based optimization strategy, the dominance concept in order to combine and optimize different segmentation criteria ; namely the (region-based) global consistency error (GCE) criterion and the (contour-based) F-measure (precision-recall) criterion. This strategy allows us to find a consensus segmentation resulting from the fusion of different and complementary criteria to enhance the quality of the final segmentation result.

4.3 Proposed Fusion Model

4.3.1 Multi-objective Optimization

In this work, we take advantage of the multi-objective optimization concept, also called vector optimization or multi-criteria optimization [138, 139], by regarding the segmentation problem from different points of view, in terms of different, complementary or contradictory criteria to be simultaneously satisfied with aim of achieving a better segmentation result.

As shown in the preliminary work [140], a mono-objective approach aims to optimize a single objective function with respect to a set of parameters. Otherwise, in the multi-objective case, there are several, often conflicting objectives to be simultaneously maximized or minimized [111]. Mathematically, in the case of minimization, the problem is generally formulated as follows :

$$\left.\begin{array}{l} \min \overrightarrow{f}(\overrightarrow{x}) \text{ (k functions to be optimized)} \\ \text{s.t } \overrightarrow{g}(\overrightarrow{x}) \leqslant 0 \\ \overrightarrow{h}(\overrightarrow{x}) = 0 \end{array}\right\} \tag{4.1}$$

where $\overrightarrow{x} \in \Re^n$, $\overrightarrow{f}(\overrightarrow{x}) \in \Re^k$, $\overrightarrow{g}(\overrightarrow{x}) \in \Re^m$, $\overrightarrow{h}(\overrightarrow{x}) \in \Re^p$. Note that the vectors $\overrightarrow{g}(\overrightarrow{x})$ and $\overrightarrow{h}(\overrightarrow{x})$ describe, respectively, m inequality constraints and p equality constraints. This set of constraints delimits a restricted subspace to be searched for the optimal solution [64]. In our case the number of functions k to be optimized is equal to 2 and without any inequality or equality constraints (i.e., $m = 0$ and $p = 0$).

The resolution of this problem consists of minimizing or maximizing these k objective functions without degradation of the optimal values obtained comparing with those obtained from a mono-objective optimization achieved objective by objective. Generally, approaches solving this problem are divided into three popular classes or types [64]. The first is the scalarization approach, also known as the weighted-sum ; according to this approach, a multi-objective problem is solved by assigning a numerical weight to each objective and combining its multiple objectives by adding all weighted criteria into a single composite function [141]. In addition to the scalarization technique, another alternative approach is the progressive preference technique. Here, the user refines his choice of the compromise during the progress of the optimization. A further important approach, which is increasingly used, includes a posteriori preference method. Thus, instead of transforming a multi-objective problem into a mono-objective problem, we can define a dominance relationship, where the overarching goal is to find the best compromise between objectives. Hence, several dominance relationships have already been presented, but the most famous and the most commonly used is the Pareto dominance, called also the Pareto Approach (PTA). This domination concept that will be used in our study is defined by :

Definition 1. *The solution $x^{(i)} \in S$ dominates another solution $x^{(j)} \in S$, denoted $x^{(i)} \prec x^{(j)}$ (in case of minimization), if and only if : $f_l(x^{(i)}) \leq f_l(x^{(j)})$ for all $l \in \{1,2,..,k\}$ and, $f_l(x^{(i)}) < f_l(x^{(j)})$ for some $l \in \{1,2,..,k\}$.*

where S denotes the search space and $f_l(.)$ represents the l-th objective function. In Fig. 4.1, we present the Pareto frontier (i.e., the set of solutions that dominate all other solutions) of a multi-objective problem in case of minimization.

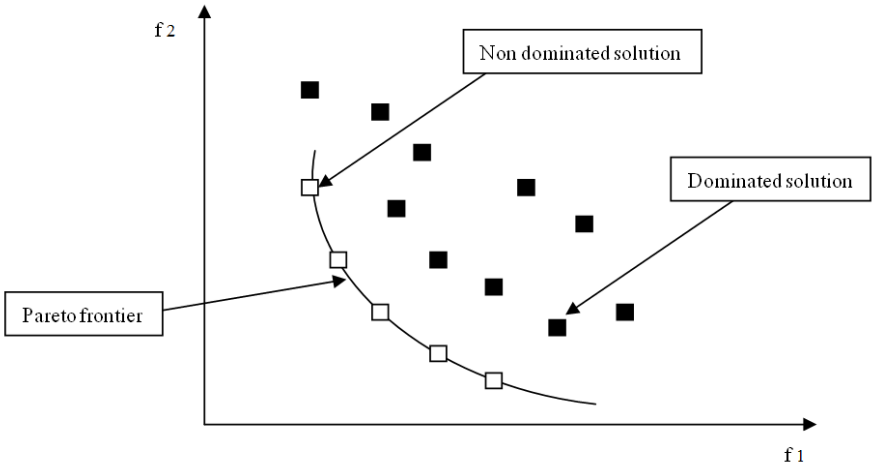


FIGURE 4.1 : Pareto frontier of a multi-objective problem in case of a minimization.

4.3.2 Segmentation Criteria

4.3.2.1 The F-measure Criterion

The F-measure is, a combination of two complementary measures ; precision and recall, which are commonly used by information retrieval theorists and practitioners [142]. In the contour-based image segmentation case, these two scores represent, respectively, the fraction of detections of the true boundaries and the fraction of true boundaries detected [77]. On the one hand, a low precision value is typically the result of over-segmentation[1] and indicates that a large number of boundary pixels have poor localization. On the other hand, the recall measure is low when there is significant under-segmentation[1], or when there is a failure to capture the salient image structure.

Mathematically, let $S_{\text{T}} = \{R^1_{\text{T}}, R^2_{\text{T}}, \ldots, R^{Nb_T}_{\text{T}}\}$ & $S_{\text{M}} = \{R^1_{\text{M}}, R^2_{\text{M}}, \ldots, R^{Nb_M}_{\text{M}}\}$ represent, respectively, the segmentation test result to be measured and the manually segmented image with Nb_T being the number of segments or regions (R) in S_{T} and Nb_M the number of regions in S_{M}. Let us now suppose that $B(R_{\text{T}})$ denotes the set of pixels that belongs to the boundary of the segment R_{T} in the segmentation S_{T} and let us also consider that

[1] In the over-segmentation : An object is partitioned into multiple regions after the segmentation and in the under-segmentation case : multiple objects are presented by a single region after the segmentation process [143].

$B(R_{\text{M}})$ is the ensemble of pixels belonging to the boundary of the segments R_{M} in the ground truth segmentation S_{M}. The precision (Pr) and recall (Re) are then respectively defined as follows :

$$Pr = \frac{|B(R_{\text{T}}) \cap B(R_{\text{M}})|}{|B(R_{\text{T}})|} \quad , \quad Re = \frac{|B(R_{\text{T}}) \cap B(R_{\text{M}})|}{|B(R_{\text{M}})|} \tag{4.2}$$

Here, $\cap$ represents the intersection operator and $|X|$ denotes the cardinality of the set of pixel X. While the precision assesses the amount of noise in the output of a detector, the recall evaluates the amount of ground-truth detected. An interesting measure that considers both the precision and the recall is called the F-measure. This combined measure aims to estimate a compromise between these two quantities and a specific application can determine a trade-off α between these two measures, describing the harmony between Pr and Re [94]. Then, the F-measure between the segmentations S_{T} and S_{M} can be evaluated as follows :

$$F_\alpha(S_{\text{T}}, S_{\text{M}}) = \frac{Pr \times Re}{\alpha \times Re + (1-\alpha) \times Pr} \quad \text{with } \alpha \in [0,1] \tag{4.3}$$

Where the F_α is in the interval of $[0,1]$, and the value of 1 proves that similar edges exists between the two segmentations, on the contrary, a value of 0 indicate the opposite situation.

4.3.2.2 The GCE Criterion

The global consistency error (GCE) [18] computed the extent to which one region-based segmentation map can be viewed as a refinement of another segmentation. This segmentation error measure is particularly useful in evaluating the agreement of a segmentation machine with a given ground truth segmentation (see Fig. 4.2) since different experts can segment an image at different levels of details.

Formally, let n be the number of pixels in the image and let $S_{\text{T}} = \{R^1_{\text{T}}, R^2_{\text{T}} \ldots, R^{Nb_T}_{\text{T}}\}$ & $S_{\text{M}} = \{R^1_{\text{M}}, R^2_{\text{M}}, \ldots, R^{Nb_M}\}$ be, respectively, the segmentation test result to be measured and the manually segmented image and Nb_T being the number of segments or regions

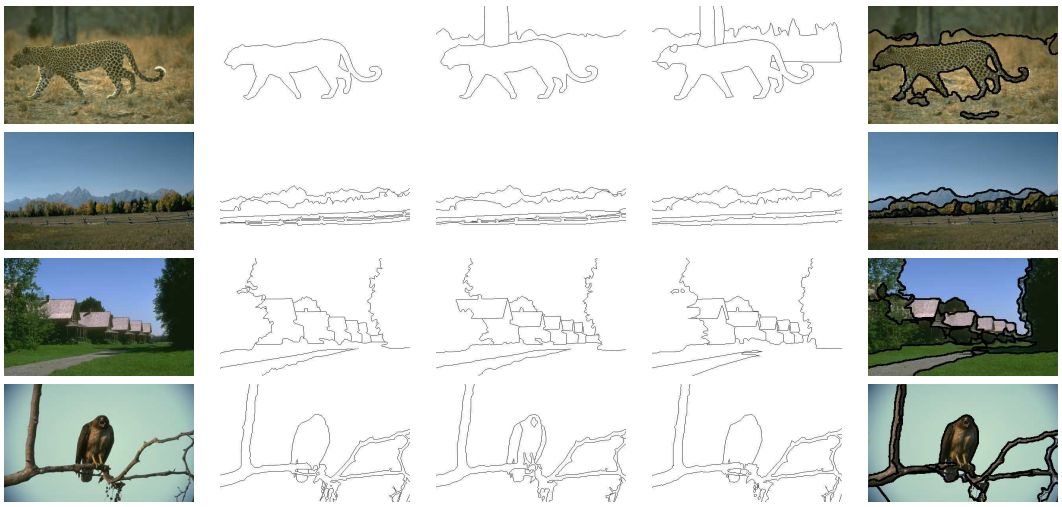


FIGURE 4.2 : Four images from the BSDS300 and their ground truth boundaries. The images shown in the last column are obtained by our MOBFM fusion model.

(R) in S_{T} and Nb_M the number of regions in S_{M}. Let now p_i be a particular pixel and the couple ($R_{\mathrm{T}}^{<p_i>}$,$R_{\mathrm{M}}^{<p_i>}$) be the two segments including this pixel (respectively in S_{T} and S_{M}). The local refinement error (LRE) can be computed at pixel p_i as :

$$\mathrm{LRE}(S_{\mathrm{T}}, S_{\mathrm{M}}, p_i) = \frac{|R_{\mathrm{T}}^{<p_i>} \setminus R_{\mathrm{M}}^{<p_i>}|}{|R_{\mathrm{T}}^{<p_i>}|} \tag{4.4}$$

where $\setminus$ represents the operator of difference and $|R|$ denotes the cardinality of the set of pixels R. Thus, a measure of 0 expresses that the pixel is practically included in the refinement area, and an error of 1 means that the two regions overlap in an inconsistent manner [18].

As it has been reported in [18], the major drawback of this segmentation measure, is that it encodes a measure of refinement in only one direction, i.e, not symmetric. To solve this issue, an interesting and straightforward way is to combine the LRE at each pixel into a measure for the whole image and for each sense. The combining result is the so-called global consistency error (GCE), which forces all local refinement to be in the same direction ; in this manner, every pixel p_i must be computed twice, once in each

sense, in the following manner :

$$\begin{aligned} &\mathrm{GCE}(S_{\mathrm{T}}, S_{\mathrm{M}}) = \\ &\frac{1}{n}\left\{\sum_{i=1}^{n}\mathrm{LRE}(S_{\mathrm{T}}, S_{\mathrm{M}}, p_i) + \sum_{i=1}^{n}\mathrm{LRE}(S_{\mathrm{M}}, S_{\mathrm{T}}, p_i)\right\} \end{aligned} \tag{4.5}$$

with this above representation, there is still considerable ambiguity, since we can find two degenerate segmentation cases ; one pixel per region and one region per image giving a GCE value equal to 0. To avoid these two problems, we can propose the new measure GCE* as follows [140] :

$$\begin{aligned} &\mathrm{GCE}^{\star}(S_{\mathrm{T}}, S_{\mathrm{M}}) = \\ &\frac{1}{2n}\left\{\sum_{i=1}^{n}\mathrm{LRE}(S_{\mathrm{T}}, S_{\mathrm{M}}, p_i) + \sum_{i=1}^{n}\mathrm{LRE}(S_{\mathrm{M}}, S_{\mathrm{T}}, p_i)\right\} \end{aligned} \tag{4.6}$$

Since the GCE* ranges in the interval of $[0,1]$, the GCE* reaches its best value at 0, this value expresses a perfect match between the two segmentations to be compared. However, it reaches the worst value at 1, this value represents a maximum difference between the two segmentations.

4.3.3 Multi-Objective Function Based-Fusion Model

Suppose now that we have a family of J segmentations $\{S_j\}_{j\leq J} = \{S_1, S_2, \ldots, S_J\}$ associated with a same scene to be combined for providing a final improved segmentation result, and let also S_I be a selected segmentation map belonging to the set $\{S_j\}_{j\leq J}$. The two complementary criteria ; namely the contour-based F-measure and the region-based GCE measure (see section 4.3.2), can be used directly, as cost functions, in an energy-based model. In this context, the consensus segmentation is simply obtained from the

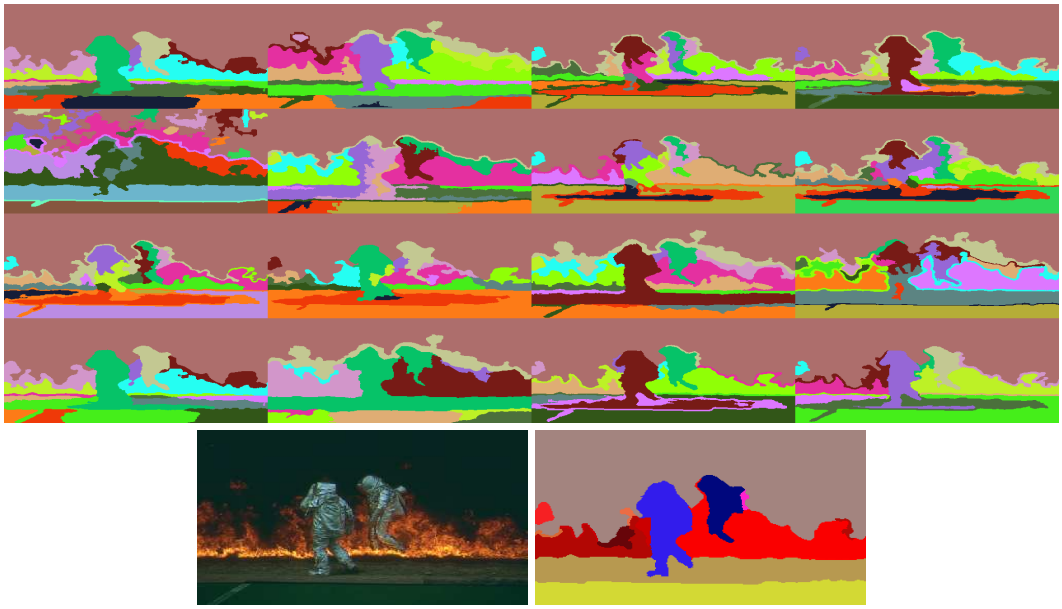


FIGURE 4.3 : A set of initial segmentations and the final fusion result achieved by MOBFM algorithm. From top to bottom ; Four first rows ; K-means clustering results for the segmentation model detailed in Section 4.4. Fifth row : Natural image from the BSDS500 and final segmentation map resulting of our fusion algorithm.

solution of the following multi-optimization problem :

$$\text{MOBJ}\left(S_I, \{S_j\}_{j\leq J}\right) = \begin{cases} \arg\max \overline{\text{F}}_\alpha\left(S_I, \{S_j\}_{j\leq J}\right) \\ \bigcap \\ \arg\min \overline{\text{GCE}}^\star\left(S_I, \{S_j\}_{j\leq J}\right) \end{cases} \tag{4.7}$$

with $\overline{\text{X}}\left(S_I, \{S_j\}_{j\leq J}\right) = \frac{1}{J}\sum_{j=1}^{J} \text{X}\left(S_I, S_j\right)$. To improve the accuracy of our segmentation result, we have made a modification in the multi-objective function (as proposed in [77]), by weighting the importance of each segmentation of $\{S_j\}_{j\leq J}$. This strategy allows us to penalize outliers and consequently aims to increase the robustness of our fusion model. So, we have weighted the first member (F-measure criterion), by a coefficient z_j proportional to its mean F-measure $\overline{\text{F}}_\alpha\left(S_I, \{S_j\}_{j\leq J}\right)$. This coefficient is defined as :

$$z_j = \frac{1}{H}\exp\left(\frac{\overline{\text{F}}_\alpha\left(S_I, \{S_j\}_{j\leq J}\right)}{d}\right) \tag{4.8}$$

where d is a parameter controlling the decay of the weights, and H is a normalizing constant ensuring $\sum_j z_j = J$. This modification allows us to ensure the robustness of our model when facing a possible bad segmentation map belonging to $\{S_j\}_{j\leq J}$ far away from the fused segmentation result. In addition, for the second member (GCE criterion), we have added a regularization term, allowing the incorporation of knowledge concerning the types of resulting fused segmentation, *a priori* defined as acceptable solutions. This term is defined as :

$$T_{\text{Reg}}\big(S_j\big) = \left| -\sum_{k=1}^{Nb_j} \left[\frac{|R_j^k|}{n} \log \frac{|R_j^k|}{n} \right] - \overline{\mathcal{Q}} \right| \tag{4.9}$$

with $S_j = \{R_j^k\}_{k\leq Nb_j}$ and Nb_j is the number of regions in the segmentation map S_j and where $\overline{\mathcal{Q}}$ is an internal parameter of our regularization term that represents the mean entropy of the *a priori* defined acceptable segmentation solutions. Thus, if the current segmentation solution has an entropy lower than $\overline{\mathcal{Q}}$, this T_{Reg} term favors splitting. On the contrary, if the current segmentation solution has an entropy greater than $\overline{\mathcal{Q}}$, T_{Reg} favors merging. Also, we have added a parameter γ to allow for weighting the relative contribution of the region splitting/merging term. Finally, with these two modifications in the multi-objective function, a penalized likelihood solution of our fusion model is thus given by the resolution of this following function :

$$\begin{aligned} &\text{MOBJ}\big(S_I, \{S_j\}_{j\leq J}\big) = \\ &\quad \begin{cases} \arg\max\Big\{ \overline{\text{F}}_\alpha \left(S_I, \{z_j\}, \{S_j\}_{j\leq J}\right) \Big\} \\ \quad \bigcap \\ \arg\min\Big\{ \overline{\text{GCE}}^\star \big(S_I, \{S_j\}_{j\leq J}\big) + \gamma T_{\text{Reg}}(S_I) \Big\} \end{cases} \end{aligned} \tag{4.10}$$

4.3.4 Optimization Algorithm of the Fusion Model

In our work, the fusion model of multiple segmentations in the bi-criteria sense (F-measure and GCE) is presented as a multi-objective optimization problem with a complex energy function. To solve this consensus function, several optimization algorithms

can be efficiently used, such as the stochastic simulated annealing or the genetic algorithms, which are both insensitive to initialization and are guaranteed to find the optimal solution but with the drawback of a huge computational load. Another alternative is to perform the optimization step by an iterative conditional modes (ICM) proposed by Besag [99], i.e. ; a Gauss-Seidel relaxation where pixels (superpixels[2] in our hierarchical approach) are updated one at a time. This iterative search technique is simple and deterministic, however, it can converge towards a bad local minima in case of an initialization by the segmentation map far from the optimal one. To solve this problem, we can choose for the first iteration of the optimization procedure, among the J segmentation to be combined, the one ensuring the minimal consensus energy of our fusion model, in the $\overline{\mathrm{GCE}}^{\star}_{\gamma}$ sense. This segmentation map $\hat{S}^{\star[0]}_{\overline{\mathrm{GCE}}}$ can be defined as :

$$\hat{S}^{[0]}_{\overline{\mathrm{GCE}}^{\star}_{\gamma}} = \arg \min_{S \in \{S_j\}_{j \leq J}} \overline{\mathrm{GCE}}^{\star}_{\gamma}\left(S_I, \{S_j\}_{j \leq J}\right) \tag{4.11}$$

In the mono-objective case, the ICM aims to accept a new solution for each pixel if this one is better than the current solution or decreases the energy function. On the contrary, in our multi-objective case, this iterative algorithm amounts to simultaneously obtain, for each (super)-pixel to be labeled, the minimum value of $\overline{\mathrm{GCE}}^{\star}_{\gamma}$ and the maximum value of $\overline{\mathrm{F}}_{\alpha}$. For this purpose, we have incorporated into the ICM a domination function (defined in section 4.3.1) ; Concretely, in each iteration, the modified ICM practically accepts a new solution to enter on the list of non-dominated solutions (L_{NDS}) only if this one is not dominated by any other solution contained in this L_{NDS} list and then updates the L_{NDS} by deleting solutions dominated by the new solution. Afterward, when the maximum number of iterations (T_{max}) is attained (and/or a sufficient number of solutions have been explored) and that no more non-dominated solution can not be found, the algorithm stops in a Pareto local optimum, and this set of non-dominated solutions is then given as input to TOPSIS technique (see Section 4.3.5). Finally, our MOBFM algo-

[2]Superpixels are given in our application by the set of regions given by each individual segmentations to be combined.

rithm with the iterative steepest local energy descent strategy and the Pareto domination is presented in pseudo-code in Algorithm 3.

4.3.5 Decision Making With TOPSIS

As soon as the generation of the Pareto frontier has been carried out [i.e., the output of Algorithm 1 (see Fig. 4.4)], one solution must be chosen, and consequently, we are faced to a multi-criteria decision making (MCDM) problem. To solve this issue we resort to a useful and efficient technique called TOPSIS (technique for order performance by similarity to ideal solution [144]). The TOPSIS technique is based on the selection of the alternative (solution) that is the closest to the ideal solution and the farthest from the negative ideal solution (see Figs. 4.5 and 4.6). The ideal solution is the one that maximizes the benefit criterion, i.e., criterion with larger value is better, and minimizes the cost criterion, i.e., criterion with smaller value is better, on the contrary, the negative ideal solution minimizes the benefit criterion and maximizes the cost criterion [145]. Let us note that these two ideal and negative-ideal solutions are, in fact, two virtual solutions or two virtual 2D points in the cost-benefit criterion space of the set of the non-dominated solutions since they are not associated with a non-dominated segmentation. Nevertheless, these two virtual solutions will be exploited by the TOPSIS technique in order to find the optimal solution according to this multi-criteria decision strategy.
As others have highlighted [146] [147], one of the advantages of this technique is its simple competition process, which allows for solving many real-problems in the research operation field (see paper [147] for more examples). Finally, the TOPSIS method is described in pseudo-code in Algorithm 1 and its graphical representation is presented in Fig. 4.5.

4.4 Segmentation Ensemble Generation

The initial segmentations used by our fusion framework are simply acquired, in our application, by a K-means [100] clustering algorithm, with 12 different color spaces, namely ; P1P2, YIQ, HSV, LUV, i123, YCbCr, LAB, TSL, RGB, HSL, h123, XYZ.

Algorithm 1 MO-Based Fusion Model algorithm

Mathematical notation:

$\overline{\text{GCE}}^{\star}_{\gamma}$	Penalized mean GCE
$\overline{\text{F}}_{\alpha}$	Mean F-Measure
$\{S_j\}_{j\leq J}$	Set of J segmentations to be fused
$\{z_j\}_{j\leq J}$	Set of weights
$\{b_j\}$	Set of superpixels $\in \{S_j\}_{j\leq J}$
$\mathcal{E}$	Set of region labels in $\{S_j\}_{j\leq J}$
L_{NDS}	List of non-dominated segmentations (Pareto set of solutions)
S_L	Solution $\in L_{NDS}$
T_{max}	Maximal number of iterations (=11)
γ	Regularization parameter
α	F-Measure compromise parameter

Input: $\{S_j\}_{j\leq J}$

Output: L_{NDS}

A. Initialization:

$$S_I^{[0]} \longleftarrow \arg\min_{S\in\{S_j\}_{j\leq J}} \overline{\text{GCE}}^{\star}_{\gamma}\left(S, \{S_j\}_{j\leq J}\right)$$

B. Steepest Local Energy Descent:

while $p < T_{\text{max}}$ **do**
  for each b_j superpixel $\in \{S_j\}_{j\leq J}$ **do**
    Draw a new label x according to the uniform distribution in the set $\mathcal{E}$
    Let $S_I^{[p],\textbf{new}}$ the new segmentation map including b_j with the region label x
    Compute $\overline{\text{GCE}}^{\star}_{\gamma}\left(S_I^{[p],\textbf{new}}, \{S_j\}_{j\leq J}\right)$
    Compute $\overline{\text{F}}_{\alpha}\left(S_I^{[p],\textbf{new}}, \{z_j\}, \{S_j\}_{j\leq J}\right)$
    if $S_I^{[p],\textbf{new}}$ **dominates** $S_I^{[p]}$ *(see Defintion 1)* **then**
      if $\nexists\ S_L \in L_{NDS}$ in which S_L **dominates** $S_I^{[p],\textbf{new}}$ **then**
        $\overline{\text{GCE}}^{\star}_{\gamma} \longleftarrow \overline{\text{GCE}}^{\star,\textbf{new}}_{\gamma}$
        $\overline{\text{F}}_{\alpha} \longleftarrow \overline{\text{F}}^{\textbf{new}}_{\alpha}$
        $S_I^{[p]} \longleftarrow S_I^{[p],\textbf{new}}$
        Update L_{NDS} *(see Algorithm 2)*
      end if
    else if $S_I^{[p],\textbf{new}}$ **not dominates** $S_I^{[p]}$ and $S_I^{[p]}$**not dominates** $S_I^{[p],\textbf{new}}$ **then**
      if $\nexists\ S_L \in L_{NDS}$ in which S_L **dominates** $S_I^{[p],\textbf{new}}$ **then**
        Update L_{NDS} *(see Algorithm 2)*
      end if
    end if
  end for
  $p \longleftarrow p+1$
end while

Algorithm 2 L_{NDS}-Updating algorithm

Mathematical notation:

$S_I^{[p],\text{new}}$	A new solution generated at iteration number p *(see Algorithm 1)*
L_{NDS}	List of non-dominated segmentations (Pareto set of solutions)
S_L	Solution $\in L_{NDS}$
$\setminus$	Private operator
$\cup$	Union operator

Input: L_{NDS}, $S_I^{[p],\text{new}}$

Output: L_{NDS}

1: Add the solution $S_I^{[p],\text{new}}$ to the list L_{NDS}
$L_{NDS} \longleftarrow L_{NDS} \cup S_I^{[p],\text{new}}$
2: **for** each solution $S_L \in L_{NDS}$ **do**
3: **if** $S_I^{[p],\text{new}}$ **dominates** S_L *(see Definition 1)* **then**
4: Delete the solution S_L from the list L_{NDS}
$L_{NDS} \longleftarrow L_{NDS} \setminus S_L$
5: **end if**
6: **end for**

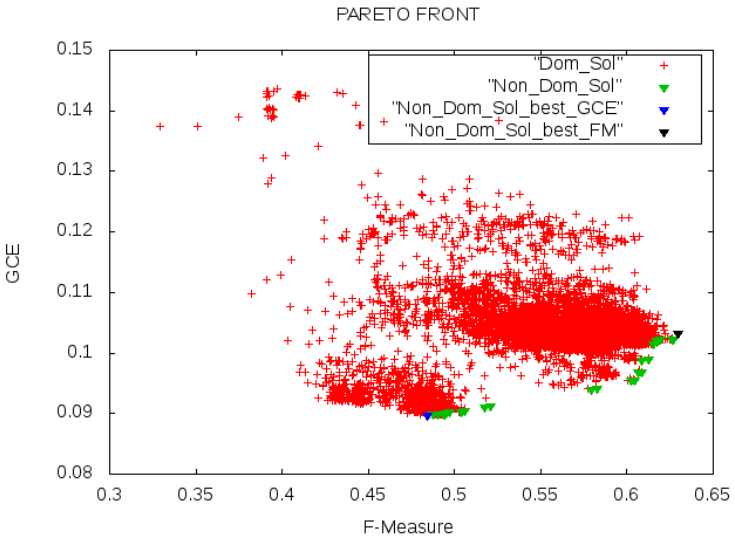


FIGURE 4.4 : First row ; a natural image ($n^0176035$) from the BSDS500. Second row ; the Pareto frontier generated by the MOBFM algorithm (cf. Algorithm 1).

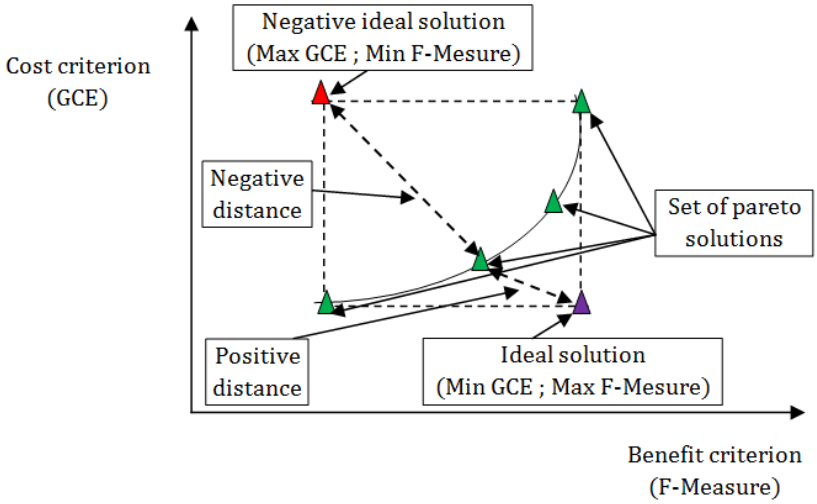


FIGURE 4.5 : Graphical representation of TOPSIS (technique for order performance by similarity to ideal solution).

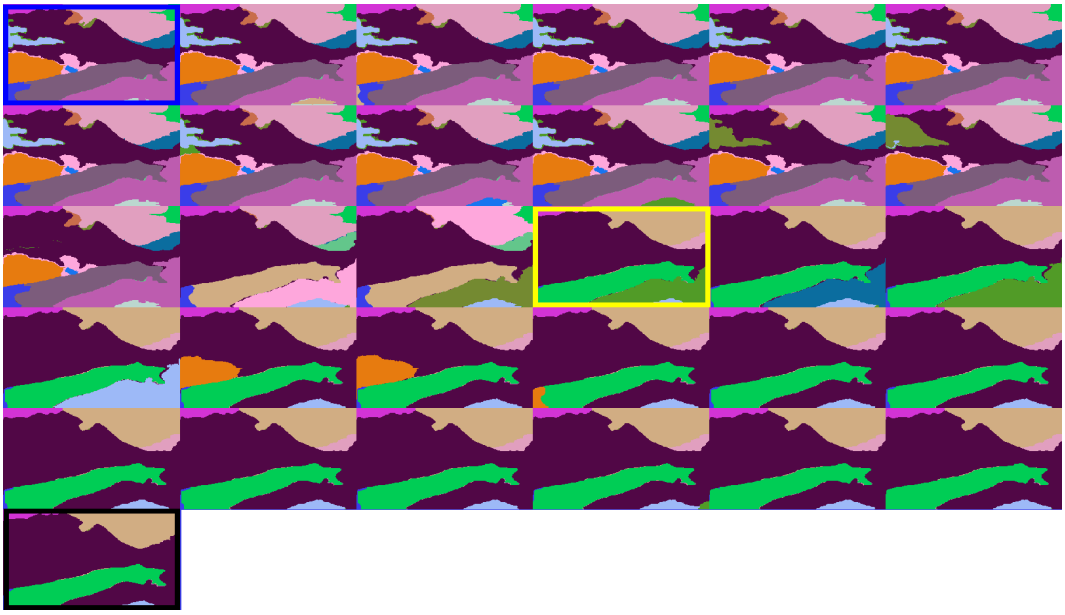


FIGURE 4.6 : The ordered set of solutions, i.e, segmentations, belonging to the Pareto-front ; The boxes marked in blue, black and yellow indicate, respectively, the solution which has the minimum $\overline{\text{GCE}}^{\star}_{\gamma}$ score, the solution which has the maximum $\overline{\text{F}}_{\alpha}$ score and the best solution chosen automatically by TOPSIS among these different solutions belonging to the Pareto frontier (cf, Fig. 4.4).

Algorithm 3 TOPSIS method

Mathematical notation:

n	Number of criteria
m	Number of alternatives (solutions)
J	Set of benefit criteria (larger is better)
$J^{'}$	Set of cost criteria (smaller is better)
W_j	The relative weight of the j-th criterion, $\sum_{j=1}^{n} W_j = 1$
L_{NDS}	List of non-dominated segmentations (Pareto set of solutions)
S^{bst}	Best solution (segmentation)

Input: L_{NDS} (output of Algorithm 1)

Output: S^{bst}

1: Construct the decision matrix X_{ij}; $i = 1, 2, .., m$ $j = 1, 2, .., n$

2: Calculate the normalized decision matrix (using vector normalization)

$$N_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^{m} X_{ij}^2}}; \; i = 1, 2, .., m \;\; j = 1, 2, .., n$$

3: Calculate the weighted normalized decision matrix
(in our case, $W_1 = 1/3$ and $W_2 = 2/3$)

$$V_{ij} = N_{ij} * W_j \; ; \; i = 1, 2, .., m \;\; j = 1, 2, .., n$$

4: Determine the ideal solution A^+ and the negative ideal solution A^-

$$A^+ = \{V_1^+, V_2^+, .., V_n^+\} = \{(max_i V_{ij} \mid j \in J), (min_i V_{ij} \mid j \in J^{'})\}$$
$$A^- = \{V_1^-, V_2^-, .., V_n^-\} = \{(min_i V_{ij} \mid j \in J), (max_i V_{ij} \mid j \in J^{'})\}$$

5: Calculate the separation measure from the ideal solution(E_i^+) and the negative ideal solution(E_i^-)(using Euclidean distance)

$$E_i^+ = \sqrt{\sum_{j=1}^{n} (V_{ij} - V_j+)^2} \; ; \; i = 1, 2, .., m$$
$$E_i^- = \sqrt{\sum_{j=1}^{n} (V_{ij} - V_j-)^2} \; ; \; i = 1, 2, .., m$$

6: Calculate the relative closeness $\overline{C_i^*}$ of each alternative to the ideal solution

$$\overline{C_i^*} = \frac{E_i^-}{E_i^+ + E_i^-} \; ; \; 0 \leq \overline{C_i^*} \leq 1$$

7: Choose an alternative with maximum of $\overline{C_i^*}$ (S^{bst})

The class number of the K-mean algorithm (K) is computed for each input image of the BSDS300 by using a metric measuring the complexity, in terms of its number of distinct texture classes within the image. This metric, defined in [101] ranges in $[0,1]$, where a value close to 0 means that we have an image with a low number of texture patterns, and a value close to 1 if we have an image with several different texture types (see Fig. 4.7). Mathematically, the value of K is written as :

$$K \quad = \quad \text{floor}\big(\frac{1}{2} + \big[K^{\max} \times \text{complexity value}\big]\big) \tag{4.12}$$

where floor(x) is a function that gives the largest integer less than or equal to x and $K^{\max}$ is an upper-bound of the number of classes for a very complex natural image. In our framework, we use three different values of $K^{\max}$, namely $K_1^{\max} = 11$ and $K_2^{\max} = K_1^{\max}$ - 2 and $K_3^{\max} = K_1^{\max}$ - 8. More details about the complexity value of an image are given in [76], but we can mention that the complexity in our case is simply the absolute deviation measure (L_1 norm) of the normalized histograms set or feature vectors for each overlapping, fixed-size squared (N_w) neighborhood included within the input image. Besides the points listed above, as input multidimensional descriptor of feature, we exploited the ensemble of values (estimated around the pixel to be labeled) of the requantized histogram (with equal bins in each color channel). In our framework, this local histogram is re-quantized, for each color channels, in a $N_b = q_b^3$ bin descriptor, estimated on an overlapping, squared fixed-size ($N_w = 7$) neighborhood, centered around the pixel to be classified with three different seeds for the K-means algorithm and with two different values of q_b, namely $q_b = 5$ and $q_b = 4$ for a total of $(3+2) \times 12 = 60$ input segmentations to be combined.

4.5 Experimental Results and Discussion

4.5.1 Initial Tests

It is important to recall that the proposed fusion model [see (4.10)] has been experimented from a segmentation ensemble $\{S_j\}_{j \leq J}$ with $J = 60$ initial segmentations

FIGURE 4.7 : Complexity values obtained on five images of the BSDS300 [18]. From left to right, value of complexity = 0.450, 0.581, 0.642, 0.695, 0.796 corresponding to the number of classes (k) (with the three different value of $K^{\max}$: $K_1^{\max}$, $K_2^{\max}$ and $K_3^{\max}$) of the k-means clustering algorithm respectively to $(5,4,2)$, $(6,5,2)$, $(7,6,2)$, $(8,6,2)$, $(9,7,3)$ in the k-means segmentation model.

acquired with the simple K-means based procedure, as indicated in Section 4.4 (see Fig. 4.3). In this case, the convergence properties of our iterative optimization procedure has been tested by considering as initialization of the ICM based iterative steepest local energy descent algorithm, respectively, two blind initializations (image spatially divided by $k = 5$ rectangles with k different labels), the input segmentation which has the $J/6 = 10$ th minimal (i.e. best) $\overline{\mathrm{GCE}}^{\star}_{\gamma}$ score, the $J/3 = 30$ th best score, the worst score, i.e., maximal, and the best score (see Fig. 4.8). It is clearly that the multi-objective cost function is certainly non-convex and complex with many local minima (see Fig. 4.9 and Fig. 4.10). Also, it is worth mentioning that the strategy, consisting of initializing the ICM procedure by the segmentation close to the optimal solution in terms of $\overline{\mathrm{GCE}}^{\star}_{\gamma}$ score, appears as a good initialization strategy that improves the final segmentation result. As a consequence, the combination of using the superpixels of $\{S_j\}_{j\leq J}$ with a good initialization strategy [see (4.11)] allows us to ensure the good convergence properties of our fusion model.

4.5.2 Evaluation of the Performance

For an objective comparison with other segmenters, we compare the use of different segmentation algorithms, with or without a fusion model strategy, evaluated on two segmentation datasets ; the BSDS300 [18] and the BSDS500 [11]. In addition, to provide

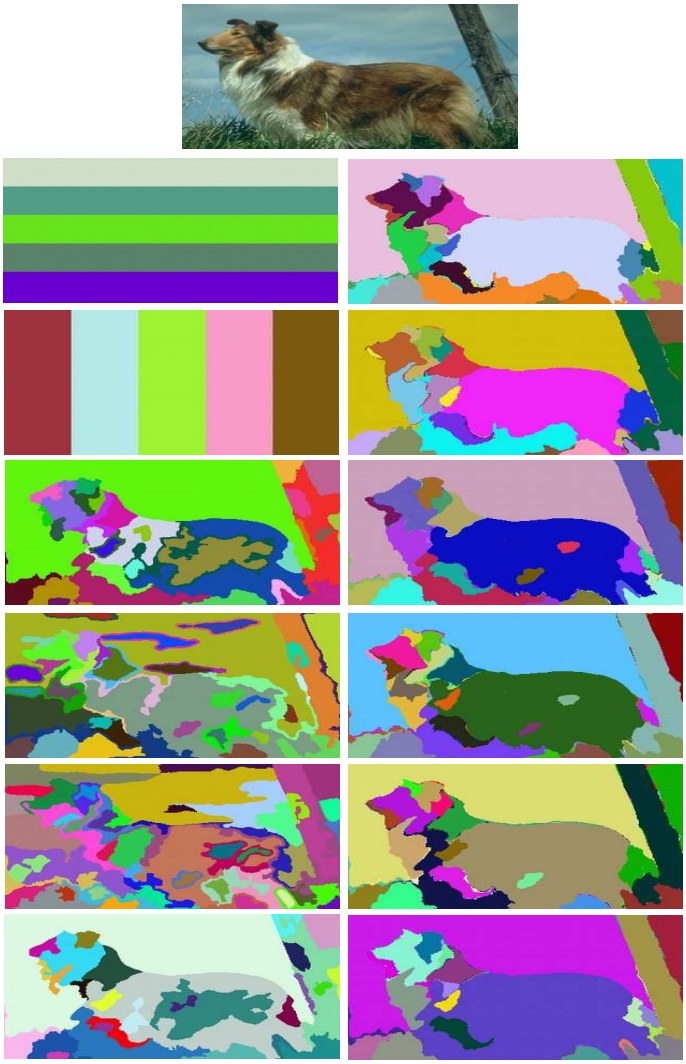


FIGURE 4.8 : Fusion convergence result on six different initializations for the Berkeley image n^{0}247085. Left : initialization and Right : result after 11 iterations of our MOBFM fusion model. From top to bottom, the original image, two blind initialization, the input segmentation which have the $J/6 = 10-th$ best $\overline{\text{GCE}}^{\star}_{\gamma}$ score, the input segmentation which have the $J/2 = 30-th$ best $\overline{\text{GCE}}^{\star}_{\gamma}$ score and the two segmentations which have the worst and the best score $\overline{\text{GCE}}^{\star}_{\gamma}$.

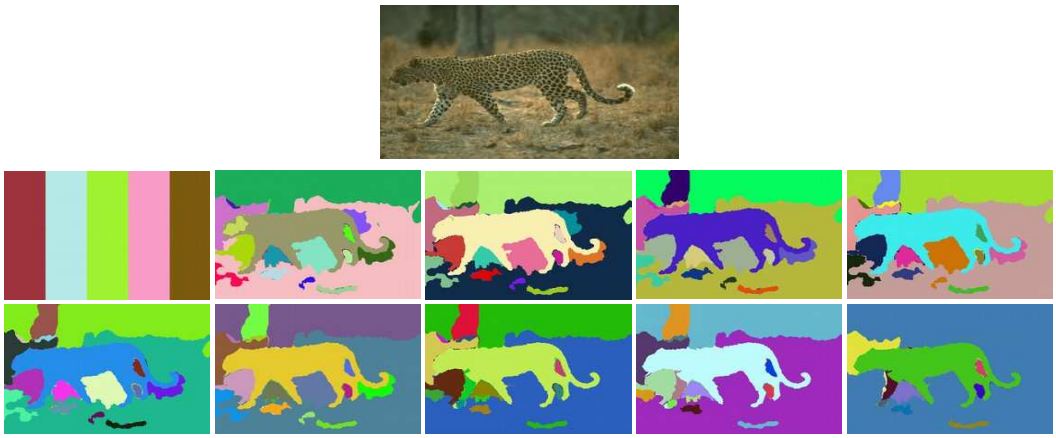


FIGURE 4.9 : First row ; a natural image ($n^0134052$) from the BSDS300. Second and third row ; evolution of the resulting segmentation map (0-th, 1-st, 2-nd, 4-th, 6-th, 8-th, 11-th, 20-th, 40-th, 80-th) (from lexicographic order) along the iterations of the relaxation process starting from a blind initialization.

a basis of comparison for the MOBFM model, we quantitatively evaluate the performance of the segmentation from two levels, namely, region level with the PRI [103], the GCE [18] and the VoI [106] and boundary level with the BDE [107]. It is important to mention that, in our application, all color images are normalized to have the longest side equal to 320 pixels. The segmentation results are then super-sampled in order to obtain segmentation images with the original resolution (481×321) before the estimation of the performance metrics.

4.5.2.1 BSDS300 Tests

The BSDS300 is a dataset of natural images that have been segmented by human observers. It contains 300 natural images divided into a training set of 200 images, and a test set of 100 images. This dataset serves as a benchmark for comparing different segmentation and boundary finding algorithms. First, in terms of region performance measures, the obtained final scores are : GCE=0.20, VoI=1.98 (for which a lower value is better) and PRI=0.80 ; this value indicates that, on average, 80 % of pairs of pixel labels are correctly labeled in the results of segmentation. It is worth noticing that our segmentation procedure gives a very competitive PRI score compared to the state-of-the-

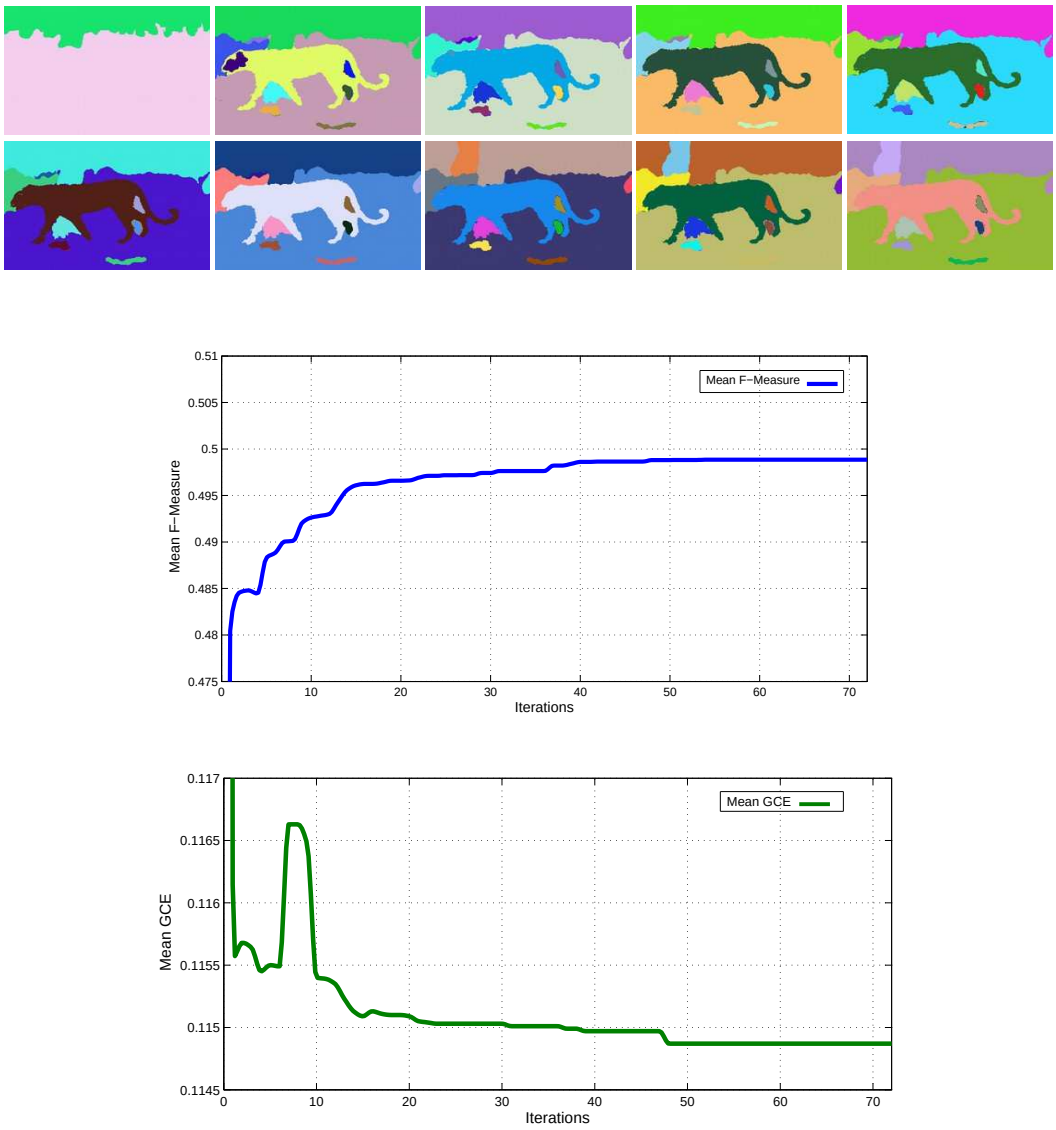


FIGURE 4.10 : First and second row ; evolution of the resulting segmentation map (0-th, 1-st, 2-nd, 4-th, 6-th, 8-th, 11-th, 20-th, 40-th, 80-th), from lexicographic order along the iterations of the relaxation process starting from the initial segmentation which have the best $\overline{\text{GCE}}^{\star}_{\gamma}$ score. Third row ; evolution of the Mean GCE value and the F-Measure value along iterations.

art segmentation methods recently proposed in the literature (see Table 4.1). Fig. 4.11 outlines, respectively, the distribution of the PRI measure and the number and size of segments provided by our MOBFM algorithm over the BSDS300. These results show us that the average number of regions estimated by our algorithm is close to the average value given by humans (24 regions) and that the PRI distribution shows us that few segmentations exhibit a bad PRI score even for the most difficult segmentation cases. Second, for the boundary performance measures, our MOBFM model performs well, with a BDE score at 8.25 (see Table 4.1). We can also observe (see Figs. 4.12 and 4.13) that the PRI, VoI, BDE and GCE performance measures are better when the number of segmentations to be fused J is high. It can be mentioned from this result that our performance scores are perfectible if the segmentation set is completed by other segmentation maps of the same image.

4.5.2.2 BSDS500 Tests

This new dataset is an extension of the BSDS300. It consists of 500 natural images divided into a training set of 300 images and a test set of 200 images, and each image was segmented by five different subjects on average. On the BSDS500, in terms of region-based metrics we obtained these following scores ; GCE=0.20, VoI=2.05 and PRI=0.80. Also, for the boundary performance measure the obtained final score is BDE=8.05 (see Table 4.2). These results prove the effectiveness and the scalability of our segmentation algorithm against different natural images and segmentation datasets.

4.5.3 Sensitivity to parameters

To ensure the integrity of the evaluation, the internal parameters of our segmentation algorithm, namely K_1^{max} required for the segmentation ensemble generation (see Section 4.4), and those required for the fusion step ; $\overline{\mathcal{Q}}$ [see (4.9)], γ [see (4.10)] and α [see (4.3)] was chosen after trial and error with a grid-type search approach applied on the train image set of the BSDS300 database.

The parameter K_1^{max} allows to refine the final segmentation map and allows, to a

TABLE 4.1 : Benchmarks on the BSDS300. Results for diverse segmentation algorithms (with or without a fusion model strategy) in terms of : the VoI, the GCE (the lower value is the better) and the PRI (the higher value is the better) and a boundary measure : the BDE (the lower value is the better)

	BSDS300			
	VoI ↓	GCE ↓	PRI ↑	BDE ↓
HUMANS	1.10	0.08	0.87	4.99
With Multi-Criteria Fusion Model				
MOBFM	**1.98**	**0.20**	**0.80**	**8.25**
With Mono-Criterion Fusion Model				
GCEBFM [5]	2.10	0.19	0.80	8.73
FMBFM [77]	2.01	0.20	0.80	8.49
PRIF [75]	1.97	0.21	0.80	8.45
FCR [2]	2.30	0.21	0.79	8.99
SFSBM [113]	2.21	0.21	0.79	8.87
Without Fusion Model				
CTM [19]	2.02	0.19	0.76	9.90
Mean-Shift [14] (in [19])	2.48	0.26	0.75	9.70
FH [12] (in [19])	2.66	0.19	0.78	9.95
DGA-AMS [148]	2.03	-	0.79	-
LSI [127]	-	-	0.80	-
CRKM [126]	2.35	-	0.75	-

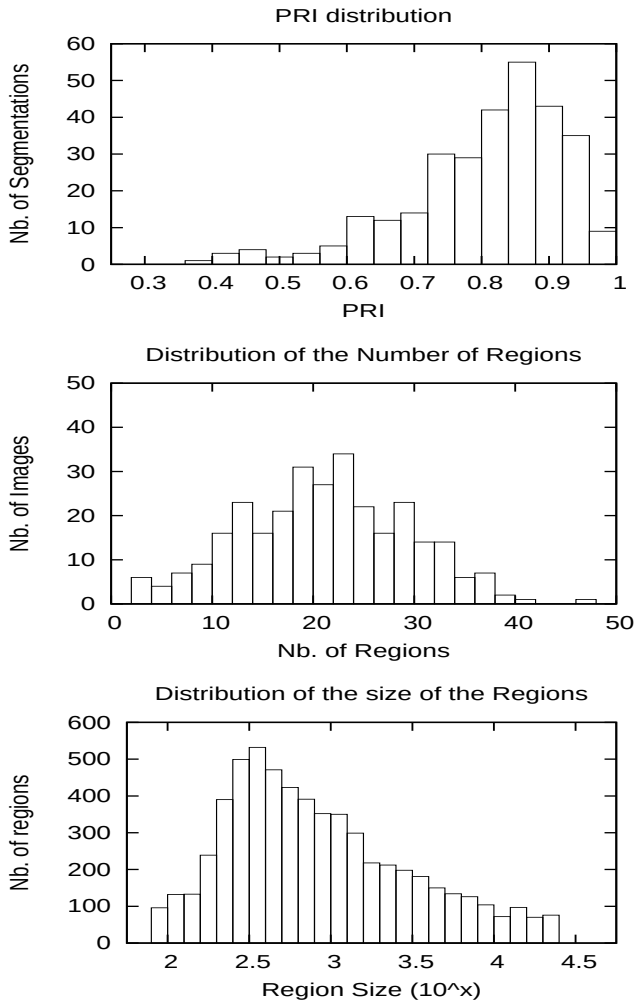


FIGURE 4.11 : From top to bottom, distribution of the PRI measure, the number and the size of regions over the 300 segmented images of the BSDS300 database.

FIGURE 4.12 : Example of fusion results using respectively $J = 5$, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 input segmentations (i.e., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 color spaces).

TABLE 4.2 : Benchmarks on the BSDS500. Results for diverse segmentation algorithms (with or without a fusion model strategy) in terms of : the VoI, the GCE (the lower value is the better) and the PRI (the higher value is the better) and a boundary measure : the BDE (the lower Value is the better).

	BSDS500			
	VoI ↓	GCE ↓	PRI ↑	BDE ↓
HUMANS	1.10	0.08	0.87	4.99
With Multi-Criteria Fusion Model				
MOBFM	**2.05**	**0.20**	**0.80**	**8.05**
With Mono-Criterion Fusion Model				
GCEBFM [5]	2.18	0.20	0.80	8.61
FMBFM [77]	2.00	0.21	0.80	8.19
PRIF [75]	2.10	0.21	0.79	8.88
VOIBFM [76]	1.95	0.21	0.80	9.00
FCR [2]	2.40	0.22	0.79	8.77
Without Fusion Model				
CTM [19] (in [128])	1.97	-	0.73	-
Mean-Shift [14] (in [128])	2.00	-	0.77	-
FH [12] (in [128])	2.18	-	0.77	-
WMS [129] (in [128])	2.10	-	0.75	-

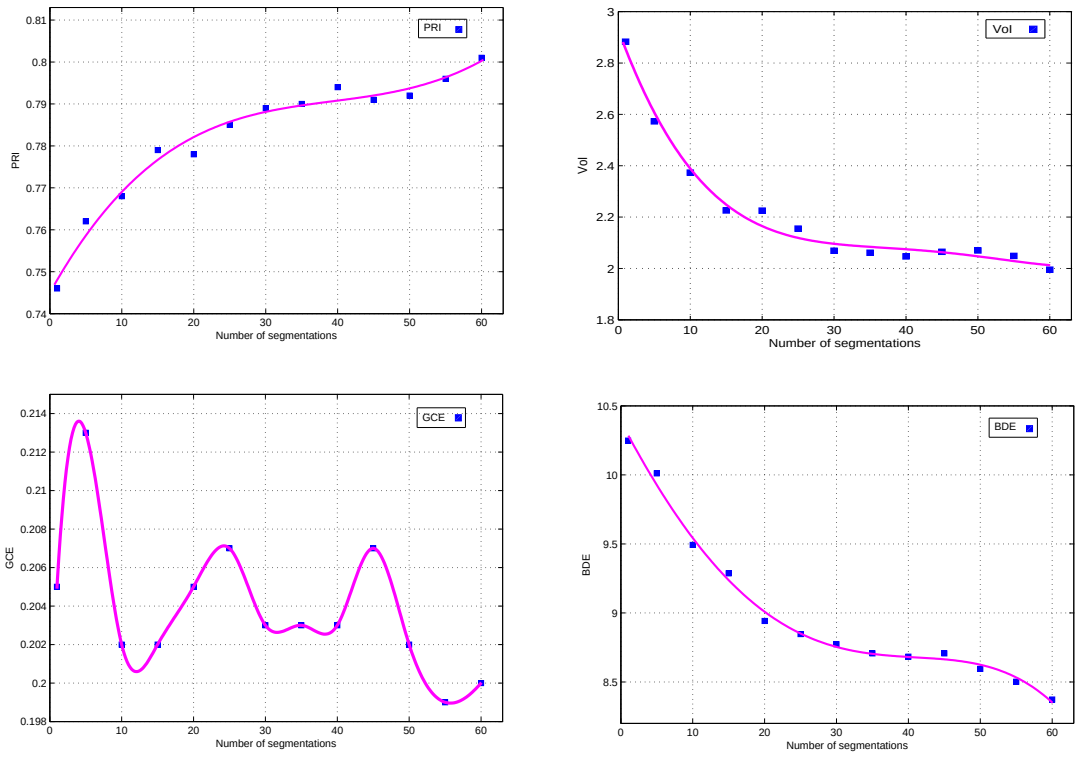


FIGURE 4.13 : From lexicographic order, evolution of the PRI (higher is better) and VoI, GCE, BDE measures (lower is better) as a function of the number of segmentations (J) to be combined for our MOBFM algorithm. More precisely for $J = 1, 5, 10, 15, 20, \ldots, 60$ segmentations, by considering first, one K-mean segmentation and then by considering five segmentations for each color space and $1, 2, 3, \ldots, 12$ color spaces.

certain extent, to avoid some over-segmented (especially when K_1^{max} is high) and under-segmented (when K_1^{max} is low) partition maps results. In order to quantify the influence of parameter K_1^{max}, we have compared the performance measures obtained with our method using three different values of K_1^{max} (see Table 4.3). Also, we have tested the role of the parameters α and $\overline{\mathcal{Q}}$ on the obtained segmentation solutions. Figs. 4.14 and 4.15 show clearly that α and $\overline{\mathcal{Q}}$ efficiently act as two regularization parameters of our fusion model. The parameter α favors over segmentation for value close to 0 and merging for value close to 1. Contrary, $\overline{\mathcal{Q}}$ favors under-segmentation, for low value and consequently splitting, for a higher value. In addition, tests show that the fusion method is sensitive to the number of segmentations to be fused (J), in the sense that the performance measures are all the more better than J is high (see Fig. 4.13).

Finally, we can notice that $K_1^{max} = 10$ or 11, $\overline{\mathcal{Q}} = 4.2$, $\gamma = 0.01$ and $\alpha = 0.86$ is a good set of internal parameters leading to a very good PRI score of 0.80 and a good consensus score for the other metrics (see Table 4.1). Further, it is important to note that we have used the same values of parameters both with the BSDS300 and BSDS500 and we have found similar values of performance measures. These results show that the parameters required for the fusion step of our algorithm do not depend on the used database and consequently that the proposed fusion model does not overfit and generalizes well. However, as the MOBFM fusion method's performance strongly depends on the level of diversity and complementarity existing in the initial ensemble of segmentations to be fused, this makes necessarily the four internal parameters of the MOBFM method highly sensitive to the pre-segmentation method (used to generate the segmentation ensemble).

4.5.4 Other Results and Discussion

Since the ICM algorithm depends on the choice of the initialization, a good initialization strategy should be used. In this context, we have used an initial segmentation based on $\overline{\mathrm{GCE}}_\gamma^\star$ score [see (4.11)] and we have found that this choice leads to the scores mentioned above. In addition, we have tested our approach with an initialization based on the F-Measure ($\overline{\mathrm{F}}_\alpha$) with the same internal parameters of our algorithm, and we have found

TABLE 4.3 : Influence of the value of parameter K_1^{max} (average performance on the BSDS300).

	BSDS300			
MOBFM (K_1^{max})	VoI ↓	GCE ↓	PRI ↑	BDE ↓
10	1.95	0.20	0.80	8.21
11	1.98	0.20	0.80	8.25
12	2.03	0.20	0.80	8.19
16	2.28	0.18	0.79	8.42
22	2.42	0.16	0.79	8.77

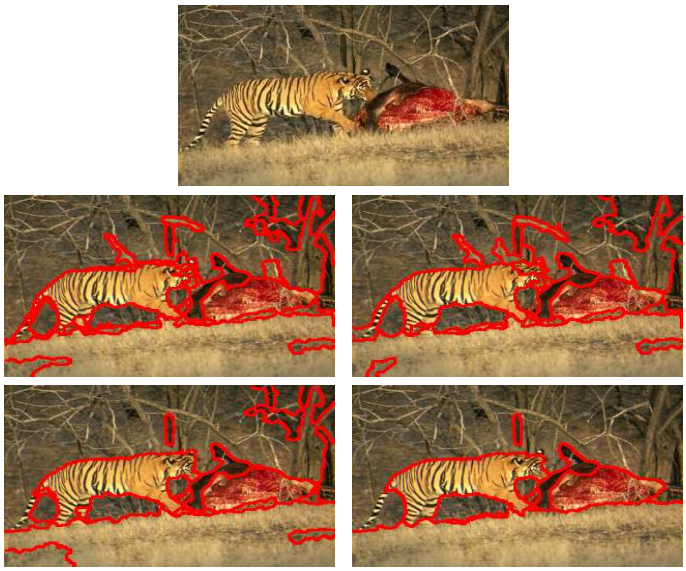


FIGURE 4.14 : Example of segmentation solutions obtained for different values of α, from top to bottom and left to right, α=$\{0.55, 0.70, 0.86, 0.99\}$.

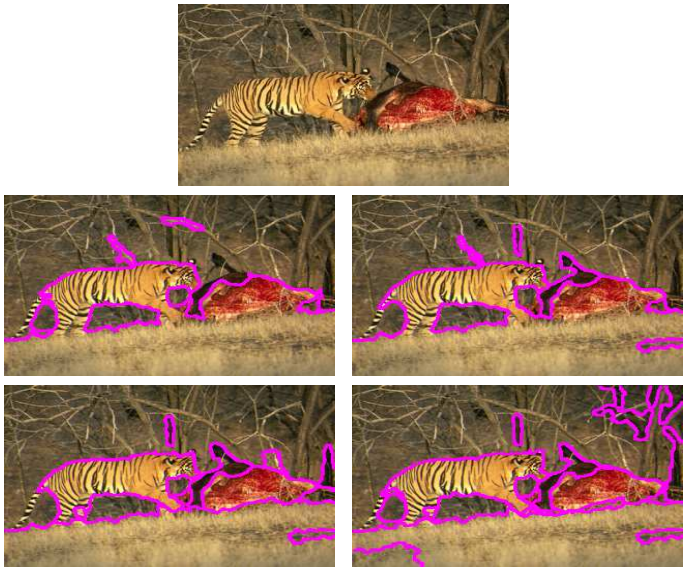


FIGURE 4.15 : Example of segmentation solutions obtained for different values of $\overline{\mathcal{Q}}$, from top to bottom and left to right, $\overline{\mathcal{Q}}$=$\{0.2, 1, 2, 4.2\}$.

that this strategy leads to the following performance measures : PRI=0.79, VoI=1.88, GCE=0.20 and BDE=8.62 on the BSDS300 ; which are slightly less better in terms of PRI and BDE than an initialization based on $\overline{\text{GCE}}^{\star}_{\gamma}$.

We can also see, from Table 4.4, that if we compare the average performances to those provided by using a single criterion, F-measure or GCE, we obtain significantly better performance rate. This shows clearly that our strategy of combining two complementary contour and region-based criteria of segmentation is effective. In order to test the robustness of our fusion approach with a third criterion, we have added to the cost function [see 4.10] the VoI (variation of information) objective, also used in [76] as the main and unique criterion of fusion of segmentations. This metric estimates the information shared between two partitions by measuring the amount of information that is gained or lost in changing from one clustering to an other [76]. The obtained final scores are ; PRI=0.80, VoI=1.97, GCE=0.19 and BDE=8.35 on the BSDS300. These results show some improvements, which can be explained by the addition of this new VoI-based criterion. But, the combination of three objectives makes our algorithm slower, with 6 minutes per image on average, and complexifies the optimization process, indicating that a high number of objectives cause additional challenges [130].

Also, as another strategy whose aim is to reduce the execution time of the algorithm, we have used the dominance function to converge directly to a solution close to the Pareto frontier, by comparing the current solution with new solutions without seeking the Pareto front ; this strategy gives us the following results : PRI=0.80, VoI=1.99, GCE=0.20, BDE=8.37 on the BSDS300 and an execution time equal to 4 minutes on average. For qualitative comparison, we now illustrate an example of segmentation results (see Fig. 4.16) obtained by our algorithm MOBFM on four images from the BSDS300 compared to other algorithms with or without a fusion model strategy (FCR [2], GCEBFM [5] and CTM [19]). From these qualitative results, we can notice that the strength of our fusion model relies in its ability to provide an appropriate set of segments for any kind of natural images.

Based on the PRI score which seems to be among the most correlated with human segmentation in term of visual perception. The results show that application of the MOBFM on the BDSD300 gives a PRI mean equal to 0.802 and a standard deviation equal to 0.1194, i.e., a significantly better mean performance along with a lower dispersion of score values than the CTM which provides a PRI mean equal to 0.761 and a standard deviation equal to 0.1427. In our case, this leads to a Z score[3] equal to 3.82, meaning that the two sample results are highly significantly different according to the Z-test. This significance of improvement is also visually and qualitatively confirmed in Fig. 16 where different segmentation results achieved by the CTM algorithm are illustrated and compared with the proposed segmentation method.

To sum up, our fusion method of simple segmentation results based on multi-objective optimization appears to be very competitive for different kinds of performance metrics and thus appears as an interesting alternative to mono-objective segmentation fusion models existing in the literature.

4.5.5 Discussion and Future Work

Let us recall that our fusion algorithm is composed of two stages, where in the first one, our algorithm estimates the set of the non-dominated solutions, constituting the so-

TABLE 4.4 : The Value of VoI, GCE, PRI and BDE as a function of the used criterion ; single-criterion (either F-Measure and GCE) and the tow combined criteria (GCE+F-measure)

	BSDS300			
Our Fusion Model	VoI ↓	GCE ↓	PRI ↑	BDE ↓
GCE	2.11	0.20	0.79	8.86
F-measure	2.04	0.20	0.78	8.52
GCE+F-measure	1.98	0.20	0.80	8.25

FIGURE 4.16 : Example of segmentation results obtained by our algorithm MOBFM on four images from the BSDS300 compared to other algorithms with or without a fusion model strategy (FCR [2], GCEBFM [5] and CTM [19].

called Pareto-front or Pareto-optimal set (see Algorithm 1 and Figs. 4.4 and 4.5).
Concretely, this set of non-dominated solutions necessarily includes the solution or the segmentation map that only optimizes (at least locally, since the ICM-based algorithm 1 is deterministic) the first criterion and also the solution that uniquely satisfies the second criterion (these two solutions are represented by the blue and the black triangle symbols, respectively, at the top right and bottom left in Fig. 4.4). The other non-dominated solutions ($\in L_{NDS}$), belonging to the Pareto-front, are, in fact, some "interesting" trade-offs or compromised solutions between the two considered criteria. Therefore, conceptually, the Pareto-front thus captures the whole set of "interesting" compromise solutions between the two considered criteria. By the word "interesting", we mean, more precisely, in fact, the set of non-dominated solutions according to the classical definition used in MCDM "*a non-dominated solution is a feasible solution where there does not exist another feasible solution better than the current one in some objective function without worsening other objective function*".
It is interesting also to note that this list or set of non-dominated solutions, belonging to the Pareto-front, can be easily ordered into a connected path of solutions, from the solution that minimizes the first criterion to the solution that optimizes the second criterion (see Fig. 4.6). This "linked chain" of segmentation maps, represented by the ordered triangles from right top to bottom left in Fig. 4.4, could help us to visually understand how the first criterion influences and characterizes a segmentation solution, in terms of the boundaries and region properties of the segments or, more generally, in terms of geometrical, aggregative, morphometric properties, compared to the second considered criterion, and this could be useful for finding a specific criterion or a pair of criteria for a specific vision application.
In addition, it is interesting to note that the length of the Pareto curve, in average for a diversified image database, is in fact a good indicator that could help us to know how a criterion is different, complementary, conflicting or contradictory from a second given criterion. Indeed, when the Pareto-front comes down to a single point or solution, it sim-

[3] $Z = (0.802 - 0.761)/\sqrt{(0.11942^2/300) + (0.1427^2/300)}$ is the distance from the sample mean to the population mean in units of the standard error.

ply means that the obtained solution is the one that simultaneously minimizes the first but also the second criterion. In this case, a mono-objective segmentation fusion model, using either the first or the second criterion, would have given the same segmentation result.

Besides, the set of plausible solutions, or candidate segmentation maps given by the Pareto-front, obtained for different given pair of criteria, could also be interestingly compared, in term of agreement, to the set of available manual segmentations estimated for each natural image, by several human observers, in the Berkeley segmentation dataset. We recall that this variability expressed by the multiple acceptable ground truth solutions associated with an image, represents, in fact, the different levels of detail and/or the possible interpretations of an image between human observers. This comparison could help us to find the pair of criteria which will give us the set of plausible solutions which would be consistent with the existing inherent variability existing between human segmenters.

Also, the Pareto-optimal set of plausible solutions could be exploited to adaptively estimate the optimal or the best compromise number of segments or regions of the segmented image.

Finally, it would be interesting to compare the length of the Pareto front, obtained for different given pair of criteria, for different segmentation ensembles (see Section 4.4) generated by different strategies. This measure could be a good indicator of the consistent diversity, as opposed to a noisy diversity, of the segmentation ensemble which is indispensable for a good fusion result.

4.5.6 Algorithm

The execution time takes, on average, between 4 and 5 minutes for an Intel® 64 Processor core i7-4800MQ, 2.7 GHz, 8 GB of RAM memory and non-optimized code running on Linux . More accurately, the first step in our segmentation procedure, i.e., estimations of the $J = 60$ weak segmentations to be fused, takes on average, 1 minute. The second step, i.e., minimization of our fusion procedure, takes approximately 3 or 4 minutes for the fusion step and for a 320×214 image. Our segmentation method

TABLE 4.5 : Average CPU time for different segmentation algorithms on the BSDS300.

ALGORITHMS	**CPU time (s)**	**On [image size]**
With Multi-Criteria Fusion Model		
MOBFM	$\simeq 240$	$[320 \times 214]$
With Mono-Criterion Fusion Model		
GCEBFM [5]	$\simeq 180$	$[320 \times 214]$
FMBFM [77]	$\simeq 90$	$[320 \times 214]$
SFSBM [113]	$\simeq 60$	$[320 \times 214]$
FCR [2]	$\simeq 60$	$[320 \times 200]$
PRIF [75]	$\simeq 80$	$[320 \times 214]$
VOIBFM [76]	$\simeq 60$	$[320 \times 214]$
Without Fusion Model		
CTM [19]	$\simeq 180$	$[320 \times 200]$
FH [12]	$\simeq 1$	$[320 \times 200]$
Mean-Shift [14] (in [128])	$\simeq 80$	$[320 \times 200]$
WMS [129] (in [128])	$\simeq 2$	$[320 \times 480]$

has acceptable computation time in comparison with some results given in the literature (see Table 4.5. However, improvements can be made, since these two steps can be easily computed in parallel by using the parallel abilities of any graphic processor unit (GPU). Moreover, the whole implementation was developed using the C++ language and the source code, data and all that is necessary for the reproduction of results and the ensemble of segmented images are available at this address ; http ://www-etud.iro.umontreal.ca/∼khelifil/ResearchMaterial/mobfm.html in order to make possible comparisons with future segmentation algorithms.

4.6 Conclusion

In this paper, we have proposed a novel and efficient fusion model based on multi-objective optimization (MOBFM), whose goal is to combine multiple segmentation maps with multiple different criteria to achieve a final improved segmentation result. This model is based on two complementary (contour and region-based) criteria of segmentation. To optimize our fusion model, we used a modified ICM algorithm, including a dominance function that allowed us to find a compromise between these different segmentation criteria. Besides that, we have used an efficient technique of decision making called TOPSIS, allowing us to find the most preferred solution from a given set of non-dominated solutions. Applied on the $BSDS300-500$, the proposed segmentation model gives competitive results compared to other segmentation models, which proves the effectiveness and the robustness of our bi-criteria fusion approach.

To sum up, we have shown that the strategy of fusion of different segmentations remains simple to implement and perfectible by incrementing the number and the complementarity of the segmentations to be fused. We have also shown that a fusion model of segmentations, expressed as a multi-objective optimization problem, with respect to a combination of different and complementary criteria, is an interesting approach that can overcome the limitations of a single criterion based fusion procedure. It gives a competitive final segmentation result for different images with several distinct texture types. Besides, the Pareto-optimal set of plausible segmentations given by this MO fu-

sion strategy can help to understand ambiguous natural scene, by providing different and plausible segmentations of an image in a similar way than the neural mechanisms of visual perception, which also provides many competing organizations making possible several conflicting interpretations of the same image. In our case, this set of multiple distinct segmentations, which corresponds to interesting compromise solutions between the two considered criteria, can be advantageously used in a last stage of computation for a specific higher level vision task.

In addition, this new multi-objective optimization strategy based on multiple different and complementary criteria remains enough general to be applied to other energy-based models, until now based on a single criterion, and extensively used in image processing, image understanding and computer vision applications. This idea is currently under investigation, especially for energy-based restoration models, denoising and deconvolution, where a fusion of different and complementary regularization terms could be appealing in order to better constrain the optimization process or to better incorporate (complementary or contradictory) knowledge or beliefs concerning the types of restorations *a priori* defined as being acceptable solutions in the associated inverse (ill-posed) optimization problem. Similarly, classification procedures, such as energy-based semantic interpretation model (scene parsing), consisting in semantically labeling every pixel in the segmented image, is also under investigation since it can also be efficiently done in a fusion framework with several complementary criteria, and on the basis of a training or learning set of segmentation with pre-interpreted classes.

Troisième partie

Interprétation sémantique d'images

CHAPITRE 5

MC-SSM : NONPARAMETRIC SCENE PARSING VIA AN ENERGY BASED MODEL

Cet article a été soumis au journal *attern Recognition* comme l'indique la référence bibliographique

L. Khelifi, M. Mignotte. MC-SSM : Nonparametric Scene Parsing Via an Energy Based Model,

Pattern Recognition, Janvier 2018.

Cet article est presenté içi dans sa version originale.

Abstract

In the last few years there has been considerable interest in scene parsing. This task consists of assigning a predefined class label to each pixel (or pre-segmented region) in an image. To best address the complexity challenge of this task, first, we propose a new geometric retrieval strategy to select nearest neighbors from a database containing fully segmented and annotated images. Then, we introduce a novel and simple energy-minimization model. The proposed cost function of this model combines efficiently different global nonparametric semantic likelihood energy terms. These terms are computed from the (pre-)segmented regions of the (query) image and their structural properties (location, texture, color, context and shape). Different from the traditional approaches, we use a simple and local optimization procedure derived from the iterative conditional modes (ICM) algorithm to optimize our energy-based model. Experimental results on two challenging datasets ; Microsoft research Cambridge dataset (MSRC-21) and Stanford background dataset (SBD) demonstrate the feasibility and the success of the proposed approach. Compared to existing annotation methods that require training classifiers for each object and learning many parameters, our method is easy to implement, has few parameters, and combines different criteria.

5.1 Introduction

Scene parsing, also called semantic image segmentation, has been attracting considerable interest in the last few years. This task aims to divide an image into semantic regions or objects, such as *mountain, sky, building, tree, etc.* The main challenge of scene parsing is that it combines three traditional problems ; detection [162], segmentation [163] [164] and multi-label recognition [165] in a single process [149]. This task aims to assign an object class label from a predetermined label set to each pixel (or super-pixel[1]) in an input image [166].

As an active research area, various methods for scene parsing have been proposed in the literature. The existing methods fall into three categories. The first one is the parametric approach that uses machine learning techniques to learn compact parametric models for categories of interest in the image. Following this strategy, we can learn parametric classifiers to recognize objects (for example, building or sky) [150]. In this field several deep learning techniques [151] have been applied to semantic segmentation, for example a parametric scene parsing algorithm based on the convolutional neural networks (CNNs) has been presented recently in [149]. In this algorithm, CNNs aim to learn strong features and classifiers to discriminate the local visual subtleties. The second is the nonparametric approach which aims to label the input image by matching parts of images to similar parts in a large dataset of labeled images. Here, the category classifier learning is replaced in general by a Markov random field in which unary potentials are computed by nearest-neighbor retrieval [150]. In the third category, a nonparametric model is integrated with a parametric model [167]. In this context, a quasi-parametric (hybrid) method, which integrates K-nearest neighbor (KNN)-based nonparametric method and CNN-based parametric method, has been proposed in [168]. Inspired by this method, a new automatic nonparametric image parsing framework towards leveraging the advantages of both parametric and nonparametric methodologies, has been also developed in [169].

Although the parametric approach has achieved great success on the scene parsing,

[1] In general, super-pixel is defined as a set of connected pixels having similar appearance [180] [182].

all current parametric methods have certain limitations in terms of training time. Another source of the problem is the retraining of models as new training dataset is added. This updating task is necessary and even important for such task, by the fact that the number of object labels in such parsing models is limited. However, the number of objects is actually unlimited in the real world. In contrast, for nonparametric approaches, no special accommodation is required when the vocabulary of semantic category labels is expanded, because there is no need to retrain category models when we add a new data [150].

To cope with these aforementioned problems related to parametric methods, in this paper, following the nonparametric approach, we propose a simple energy-minimization model called the multi-criteria semantic segmentation model (MC-SSM). The potential aim of this new model is to take advantages of the complementarity of different criteria or features. Thus, the proposed model combines efficiently different global likelihood terms either based on the spatial organization and distribution of the region semantic labels within the image or on region-based properties (location, texture, color, context and shape), and their training adequacy, in a multi-criteria cost function. In order to optimize our energy-model, we use a simple local optimization procedure derived from the iterative conditional modes (ICM) algorithm.

In the following, the paper is structured as follows : A literature review concerning the nonparametric approach for scene parsing is presented in Section 5.2. Then our semantic segmentation model is discussed in detail in Section 5.3. Experimental results and comparisons with existing scene parsing methods are illustrated in Section 5.4. In this section our method is validated on two publicly available databases. A summary of our method and discussion of the conclusions are presented in Section 5.5.

5.2 Related Work

In nonparametric scene parsing approach, methods can be generally classified into three groups based on the relationships (dependencies) which are encoded between different pixels in the image. The first type contains methods which solve the pixel-labeling

problem by classifying each pixel independently [170] [171]. Following this strategy, we can mention the system proposed by Liu *et al.* [172], which selects a subset of the nearest neighbors for an input image, using a large dataset that contains fully annotated images. In this system, a dense correspondence is established between the query image and each of the nearest neighbors using the SIFT flow algorithm [173]. Then, the annotations are transferred from the retrieved subset to the input image using a Markov random field (MRF) defined over pixels. However, the high computational cost of these types of methods and their inefficiency makes them unattractive to applications. The second type of methods is based on the pairwise MRF or conditional random field (CRF) models [174], where nodes in the graph represent the semantic label associated with a pixel, and potentials are created to define the energy of the system. Thus, a relationship between pairs of neighboring pixels is incorporated in the graph, which encourages adjacent pixels that are similar in appearance to take the same semantic label. However, in this type of framework, the learning and inference of complex pairwise terms are often expensive. In addition, this approach is still too local and not descriptive enough to capture long-range relationships observed between adjacent regions. In the third group, pixels are grouped into segments (or super-pixels) and a single label is assigned to each group [175]. Following this approach, an efficient nonparametric image parsing method called Superparsing [176] has been proposed by Tighe *et al.*, in this method, an MRF is applied over super-pixels instead of pixels, then labels are transferred from a set of neighbor images to the input image based on super-pixels similarity. Also, Zand *et al.* [177] have proposed recently an ontology-based semantic image segmentation using mixture models and multiple CRFs. By doing so, the problem of image segmentation is then reduced to that of a classification task where CRFs individually classify image regions into appropriate labels for each visual feature. Moreover, Xie *et al.* [166] have proposed a new semantic image segmentation method addressing multiscale features and contextual information. In their work, an over-segmentation is applied on a given image to generate various small-scale segments, and a segment-based classifier with a CRF model are used to generate large-scale regions, then the features of regions are exploited to train a region-based classifier. It is important to note that, there are two main questions that

need to be asked when we follow the nonparametric image parsing approach, which are : a) How to retrieve some similar images from a training dataset for a query test image ; b) How to parse the test image with the retrieved images by transferring the annotation associated with the retrieved images to the query image [178]. In this work, to solve the first problem, we propose a new selection process based on a new criterion called global consistency error. For the second issue, as shown in the preliminary work [179], we propose a novel energy-minimizing framework, which aims to assign to each region a single class label based on a global fitness function.

5.3 Model Description

As mentioned in Section 5.1, our main aim is to decompose an image I into an unknown number (K) of geometric regions, and then to identify their categories (*i.e.*, tree, building, mountain, etc.) by iteratively optimizing a multi-criteria energy function that evaluates the quality of the solution at hand. Fig. 5.1 illustrates the proposed system overview, which consists of following four steps : (i) Region generation creates a set of regions (*i.e.*, objects) for a given input image. (ii) Geometric retrieval set selects a subset of images from the entire dataset, by a new matching scheme based on the global consistency error (GCE) measure. (iii) Region features extract different types of features for each region, including color, texture, shape, image location and semantic contextual information (both for the input image and the retrieval set). (iv) Image labeling assigns each region with an object class label by using an energy minimization scheme. In the following subsections, each step of our model is discussed in detail.

5.3.1 Regions Generation

In this first step, a set of segments (regions) is generated by a new pre-segmentation algorithm called GCEBFM [5, 20]. This novel algorithm aims to obtain a final refined segmentation by combining multiple and eventually weak segmentation maps generated by the standard K-means algorithm. This algorithm is applied on 12 different color spaces in order to ensure variability in the segmentation ensemble, those are, YCbCr,

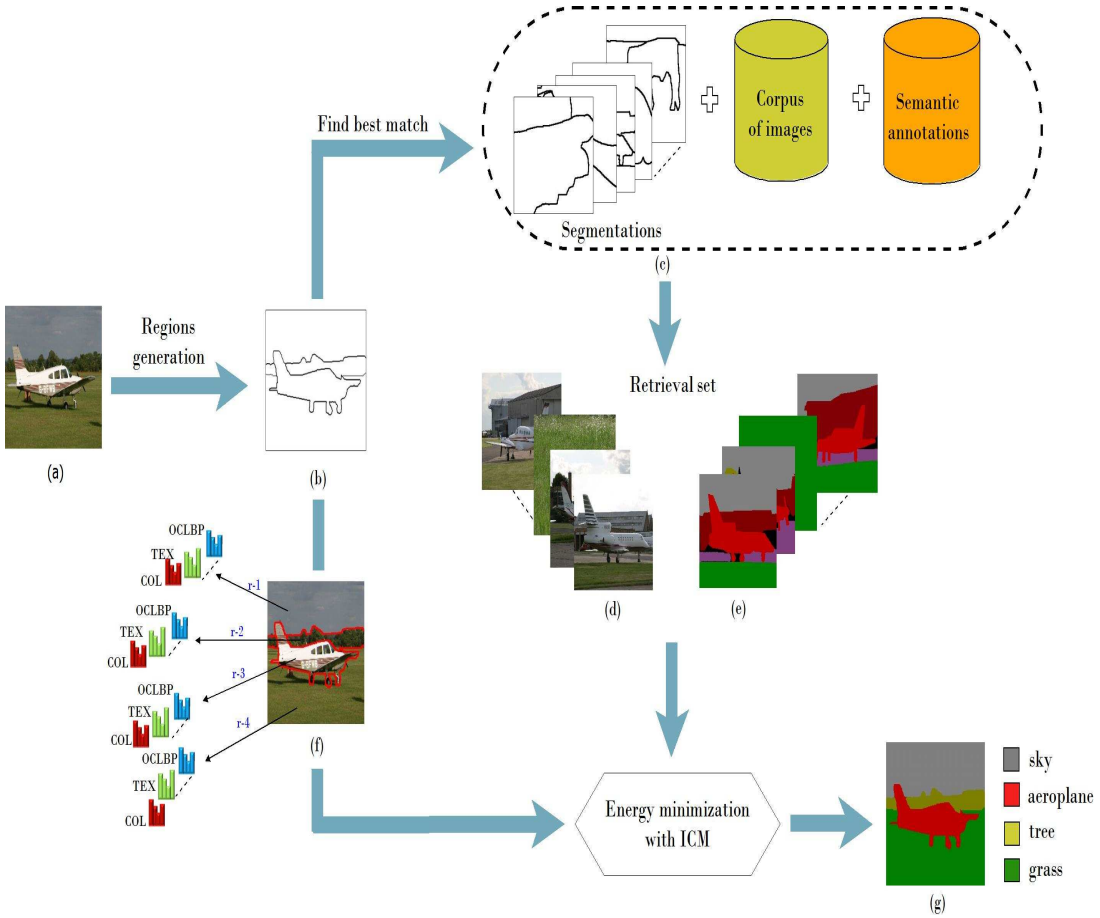


FIGURE 5.1 : System overview. Given an input image (a), we generate its set of regions with the GCEBFM algorithm (b), we retrieve similar images from the full dataset (c) using the GCE criterion, we extract different features both for the input image (f) and the retrieved images (d). Based on the labeled segmentation corpus (e), a single class label is assigned to each region (g) using energy minimization based on the ICM.

TSL, YIQ, XYZ, h123, P1P2, HSL, LAB, RGB, HSV, i123, and LUV. This new algorithm has been adopted in our work mainly for two reasons ; Firstly, as it has been mentioned in [5], this fusion algorithm remains simple to implement, perfectible, by incrementing the number of segmentations to be fused, and general enough to be applied to different types of images. Secondly, previously published studies [180] that use predefined super-pixels[1], generated by an over-segmentation, provide boundaries which are often inconsistent with the true region boundaries, and in most cases, objects are segmented into many regions, making an accurate decomposition of the image impossible. On the contrary, this algorithm aims to generate large regions which allow us to derive global properties for each region (see Section 5.3.3), and on the other hand, to reduce the complexity and the memory requirement of the full model. Also, it is important to note that the performance of this new fusion model was evaluated on the Berkeley dataset [18] including various segmentations given by humans (in [5] more explanations are given about this new algorithm). Fig. 5.2 shows examples of initial segmentation ensemble and fusion results of an input image chosen from the MSRC-21 dataset [181].

It is worth mentioning that is very difficult to act directly on the segments produced by GCEBFM, in order to deduce their appropriate semantic interpretation for example by following a full parametric approach. This difficulty is due to the higher number of class labels on the most available data sets and the type of the used criteria in our model. For example, the statistical distribution of color related to each object category is diverse in the MSRC-21 dataset [181]. Rather than building a complex scene parsing system (that uses, for example, a conditional random field (CRF) model [170] to learn the conditional distribution over the class labeling given an image), our goal is to propose a new simple model that based on the transfer of semantic labels from a retrieval set annotated images to the query segmentation (generated by the GCEBFM).

5.3.2 Geometric Retrieval Set

In our method, we follow the hypothesis, indicating that using a subset of images which are similar to the query image, instead of using the entire dataset, is more useful for the labeling task. Note that it could be meaningful to labeling an object as a tree if we

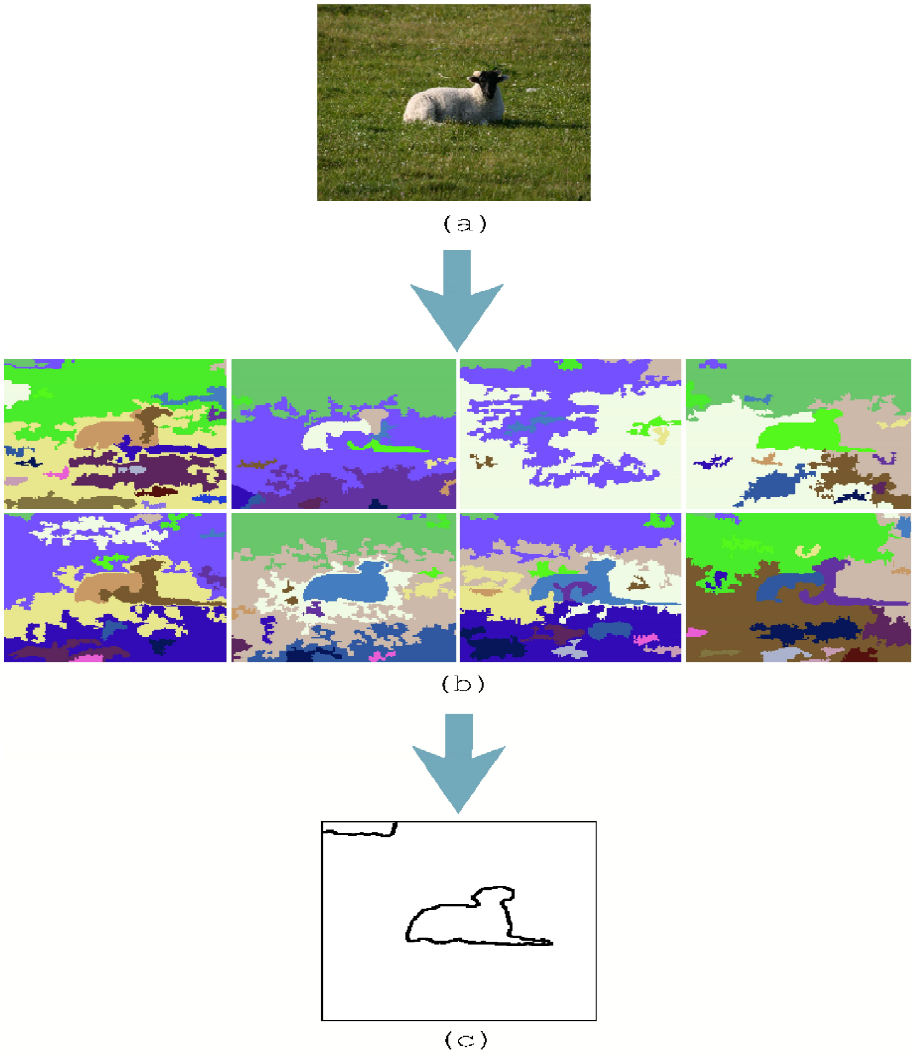


FIGURE 5.2 : Regions generation by the GCEBFM algorithm [20]. (a) input image. (b) examples of initial segmentation ensemble. (c) segmentation result.

search for the nearest neighbors in images of gardens and eliminate views from indoor scenes. With the aim of finding a relatively smaller and interesting set of images instead of using the entire training set, we use a new criterion called global consistency error (GCE) to find matches between the region map or the segmentation of the input image (see Section 5.3.1) and the region map of each image in the dataset. This new similarity criterion was recently proposed in the segmentation fusion framework [5] based on the *median partition* solution (which conceptually defines the consensus segmentation as being the partition that minimizes the average pairwise distance between itself and all other segmentations) and before that, as a quantitative metric to compare and evaluate a machine segmentation with multiple (possible) ground truths (*i.e.*, manually segmented images provided by experts) [19]. Based on this metric, a perfect correspondence is yielded if each region in one of the segmentation is a subset or geometrically similar to a region in the other segmentation (this appealing property inherent to GCE makes this criterion relatively invariant to a possible over-segmentation). The GCE measure is originated from the so-called local refinement error (LRE) [19] which is expressed at each pixel. Mathematically, let n be the number of pixels within the image I and let $R_I = \{r_I^1, r_I^2, \ldots, r_I^{nb_I}\}$ & $R_{\text{M}} = \{r_M^1, r_M^2, \ldots, r_M^{nb_M}\}$ be, respectively, the segmentation result of the input image to be measured and the segmentation of an image that belongs to the dataset, nb_I being the number of segments or regions in R_{I} and nb_M the number of regions in R_{M}. Let now p_i be a particular pixel and the couple $(r_{\text{I}}^{<p_i>}, r_{\text{M}}^{<p_i>})$ be the two segments including this pixel (respectively, in R_{I} and R_{M}). The local refinement error (LRE) can be computed at a pixel p_i as follows :

$$\text{LRE}(r_{\text{I}}, r_{\text{M}}, p_i) = \frac{|r_{\text{I}}^{<p_i>} \setminus r_{\text{M}}^{<p_i>}|}{|r_{\text{I}}^{<p_i>}|} \tag{5.1}$$

Where $|r|$ denotes the cardinality of the set of pixels r and $\setminus$ represents the algebraic operator of difference. Particularly, a value of 1 means that the two regions overlap, in an inconsistent manner, on the contrary, an error of 0 expresses that the pixel is practically included in the refinement area [18]. A great way of forcing all local refinement to be in the same direction is to combine the LRE. On this basis, every pixel p_i must be computed

twice, once in each sense, and in fact, gives as result the so-called global consistency error (GCE) :

$$\begin{aligned} &\mathrm{GCE}^{\star}(R_{\mathrm{I}}, R_{\mathrm{M}}) = \\ &\frac{1}{2n}\left\{\sum_{i=1}^{n}\mathrm{LRE}(r_{\mathrm{I}}, r_{\mathrm{M}}, p_i) + \sum_{i=1}^{n}\mathrm{LRE}(r_{\mathrm{M}}, r_{\mathrm{I}}, p_i)\right\} \end{aligned} \tag{5.2}$$

The GCE* value belongs in the interval of $[0,1]$, on the one hand, a value of 0 expresses a maximum similarity between the two segmentations R_{I} and R_{M}, on the other hand, a value of 1 represents a bad match or correspondence between the two segmentations to be compared.

Finally, based on this GCE distance and in ascending order from the query image, we rank all the images OF the entire dataset T. Then, we eliminate unhelpful images that have a higher GCE value, and we select a subset of images M from the entire dataset T as the retrieval set.

5.3.3 Region Features

A key idea with the proposed approach is that it simply uses large regions as the basic semantic unit. To perform the labeling process, we define the characteristics of those regions by extracting different features for each one. These used features are divided into five types ;

- Color : This feature gives a relevant information about the statistical distribution of color related to each region. For each pixel, we estimate the re-quantized color histogram, with equidistant binning ($P_{BIN} = 5$) for each color channel (RGB), by considering the set of color values existing in an overlapping squared neighborhood ($SN = 7$) centered around this pixel. A normalized re-quantized color histogram is then estimated for each region by simply averaging the local histograms of each pixel belonging to the same region.

- Texture : To quantify the perceived texture of different regions in an image we use three features :

– Histogram of oriented gradients (HOG) : We compute the 40-bin normalized HOG with 4 different directions (respectively, vertical, horizontal, right diagonal, and left diagonal) and 10 amplitude values. By doing so, each histogram is computed on the luminance component of each pixel contained in an overlapping squared neighborhood ($SN = 7$) centered around each pixel in the image. Then, we average all histograms of pixels which belong to the same region. Note that this region-based strategy of normalization aims to make this feature more invariant to changes in shading and illumination comparatively to a pixel-based approach.

– Opponent color local binary pattern (OCLBP) : The original LBP operator proposed by Ojala *et al.* [183] was aimed to represent statistics of micro patterns contained in an image by encoding the difference between the pixel value of the center point and those of its neighbors. Formally, let I be a color image and let q_c be the value of the center pixel c of a local neighborhood and let q_p $(p = 0, ..., P-1)$ be the values of P equally spaced pixels on a circle of radius R that form a circularly symmetric set of neighbors. If the coordinates of q_c are $(0,0)$, then the coordinates of q_p are given by $(R\sin(\frac{2\pi p}{P}), R\cos(\frac{2\pi p}{P}))$. Particularly, a bilinear interpolation is used to estimate the values of neighbors which do not fall exactly in the center of a pixel. The LBP operator on this pixel (c) is defined as follows :

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(q_s - q_c)2^p,\ s(x) = \begin{cases} 1 & ,x \geq 0 \\ 0 & ,x < 0 \end{cases} \tag{5.3}$$

In our method we apply the opponent color version of LBP (OCLBP) presented in [184] and used recently in [185]. The idea within this extended version is to take a center pixel from one color channel and neighborhood from other color channel. For example, the OCLBP operator for a pixel c and between

color channel pair (C_a, C_b) can be defined as :

$$OCLBP_{P,R}(C_a, C_b) = \sum_{p=0}^{P-1} s(q_s^{C_a} - q_c^{C_b}) 2^p \tag{5.4}$$

After computing the OCLBP for three pairs of color channels (red-green, red-blue and green-blue), as input multidimensional descriptor of feature, we compute the set of values of the re-quantized OCLBP histogram (in each OCLBP result of color channel pair), with equidistant binning, $P_{BIN} = 5$. Thus, each histogram of 125 bins (as feature descriptor) is estimated at an overlapping, fixed size squared ($N_w = 7$) neighborhood centered around the pixel. Finally, we average all histograms of pixels which belong to the same region (see Fig. 5.3).

– Laplacian operator (LAP) : In order to more efficiently capture local textural properties of each region, we also propose a new criterion derived from the Laplacian operator expressed in the logarithmic space [137] which efficiently complements the HOG features. The two steps of the estimation of this criterion are summarized in Algorithm 1.

- Context : As the context plays an important role in natural human recognition of objects and scene understanding [186], we decide to exploit the semantic contextual information around each region. More precisely, we compute the z-bin (z is the number of classes in the dataset) normalized histogram over the labels of the neighbors of each region excluding its own semantic label.

- Shape : Motivated by the efficacy of this classic feature, and in order to provide a simple geometric property, in our approach, we calculate the normalized area (i.e, the number of pixels in a region divided by the number of pixels within the image) of each region in the image.

- Location : This feature aims to capture the global position of each region with respect to the topmost pixel in the image (by computing the maximum y-coordinate).

For example, sky region tends to have the minimum distance to the horizon.

5.3.4 Image labeling

5.3.4.1 Principle

After extracting the feature descriptors used to describe regions and given an available labeled segmentation corpus, a single class label is assigned to each region by optimizing a global fitness function that measures the *quality* of the generated solution.

More formally, let us assume that we have an input image I and its region segmentation $R_I = \{r_I^1, r_I^2, \ldots, r_I^m\}$ to be semantically labeled, where m represents the number of regions (r) in R_I. Let also $\mathcal{C} = \{\mathcal{I}_k, \mathcal{S}_k\}_{k \leq K}$ represents respectively a set (or a training corpus) of K images $\mathcal{I}_k$ and their corresponding semantic segmentations $\mathcal{S}_k$. In our framework, if $\mathcal{S}_\Omega$ represents the set of all possible semantically labeled segmentation maps of I (based on its partition into regions R_I) then, our semantic labeling problem $\hat{S}_{\text{MC}} = \{s_I^1, s_I^2, \ldots, s_I^m\}$ is formulated as the result of the following multi-criteria optimization problem :

$$\hat{S}_{\text{MC}} = \arg\min_{S \in \mathcal{S}_\Omega} \overline{\text{MC}}\left(I, R_I, S, \{\mathcal{I}_k, \mathcal{S}_k\}_{k \leq K}\right) \tag{5.5}$$

$$\begin{aligned}
\text{with :}\overline{\text{MC}}\left(I, R_I, S, \{\mathcal{I}_k, \mathcal{S}_k\}_{k \leq K}\right) = \quad & \alpha_1 \sum_{i=1}^{m} \text{COL}\left(I, r_I^i, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \\
+ \alpha_2 \sum_{i=1}^{m} \text{TEX}\left(I, r_I^i, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \quad & + \alpha_3 \sum_{i=1}^{m} \text{OCLBP}\left(I, r_I^i, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \\
+ \alpha_4 \sum_{i=1}^{m} \text{LAP}\left(I, r_I^i, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \quad & + \alpha_5 \sum_{i=1}^{m} \text{SHA}\left(r_I^i, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \\
+ \alpha_6 \sum_{i=1}^{m} \text{LOC}\left(r_I^i, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \quad & + \alpha_7 \sum_{i=1}^{m} \frac{1}{h} \left\{ \sum \text{CTX}\left(r_I^m, s_I^i, \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}\right) \right\}
\end{aligned}$$

Where the parameters α_1, α_2, α_3, α_4, α_5, α_6 and α_7 are used to weight the different terms of this energy function. COL, TEX, OCLBP, LAP, SHA, LOC and CTX designate respectively the different energy terms, or nonparametric distance measures, of this cost function, reflecting the adequacy of a specific semantic label (existing in the training

corpus $\{\mathcal{I}_k, \mathcal{S}_k\}_{k \leq K}$) for each region of the image, in terms of its color, texture, shape, image location and semantic contextual information.

More precisely, let $\{\mathcal{C}\}^{s_I^i} = \{\mathcal{I}_k, \mathcal{S}_k\}^{s_I^i}$ denotes the set of images $\mathcal{I}_k$ and their associated semantic segmentation solutions $\mathcal{S}_k$ (belonging to the training corpus) that contain a region semantically labeled s_I^i and let also h be the total number of those semantic segmentations in the corpus $\{\mathcal{C}\}^{s_I^i}$ (see Table 5.1). Then, COL(.), TEX(.), OCLBP(.), LAP(.) and CTX(.) are, respectively, the minimum Ruzicka distance[2] between the p-bin normalized color histogram, the q-bin normalized histogram of oriented gradients (HOG), the p-bin normalized OCLBP histogram, the p-bin normalized LAP histogram,the z-bin normalized histogram of semantic labels of r_I^i and those of each region corresponding to the semantic label assigned to r_I^i (*i.e.*, s_I^i) and existing in $\{\mathcal{C}\}^{s_I^i}$. Also, LOC(.) and SHA(.) are, respectively, the minimum absolute distance between the normalized area, the height of the topmost pixel existing in the region r_I^i, and normalized area and the topmost pixel of each region corresponding to the semantic label assigned to r_I^i (*i.e.*, s_I^i) and existing in $\{\mathcal{C}\}^{s_I^i}$.

Algorithm 1 Estimation of the Laplacian operator

Mathematical notation:

r Radius (r=1)

for each pixel $x(i,j)$ with color value R_x, G_x,B_x **do**
  $x(i,j) = 1/3 \times (R_{x(i,j)} + G_{x(i,j)} + B_{x(i,j)})$
end for
for each pixel $x(i,j)$ **do**
  $X_0(i,j) = \log(1 + x(i,j+r) - 2 \times x(i,j) + x(i,j-r))$
  $X_1(i,j) = \log(1 + x(i+r,j) - 2 \times x(i,j) + x(i-r,j))$
  $X_2(i,j) = \log(1 + x(i,j+r) - 2 \times x(i,j) + x(i-r,j-r))$
end for

5.3.4.2 Optimization of the Energy Function

The proposed semantic segmentation model of multiple label fields is formulated as a global optimization problem incorporating a nonlinear multi-objective function. In order

[2] $\text{distance}_{\text{Ruzicka}} = 1 - \sum_i [\min(P_i, Q_i) / \max(P_i, Q_i)]$

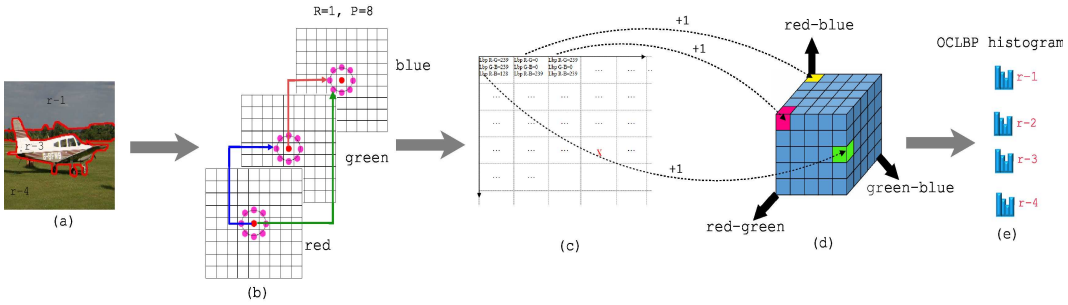


FIGURE 5.3 : Generation of the OCLBP histogram for each region. (a) The regions map of the input image. (b) Estimation of LBP value of a center pixel from one color channel based on neighborhoods from another channel [see (5.4)]. (c)-(d) Estimation, for each pixel X, of the N_b bin descriptor $q = 5$ in the cube of pair channels. Each $LbpR - G_X$,$LbpR - B_X$, $LbpG - B_X$ value associated with each pixel contained in a squared neighborhood region of size 7×7 centered at a pixel X, increments $(+1)$ a particular bin. (e) OCLBP histogram of each region.

TABLE 5.1 : Summary of the combined criteria used in our Model.

TYPE	CRITERION	DIMENSION
Color	Color histogram	125
Texture	Oriented gradient histogram	40
	Opponent color local binary pattern histogram	125
	Laplacian operator histogram	125
Shape	Pixel area	1
Location	Top height	1
Context	Context histogram	21

to achieve the minimum of this energy function [see (5.5)], approximation approaches based on different optimization algorithms such as the exploration/selection/estimation (ESE) [131], the genetic algorithm or the simulated annealing can be exploited. These algorithms are guaranteed to find the optimal solution, but with the drawback of a huge computational time. To avoid this problem, in this work we adopt the iterated conditional modes (ICM) method proposed by Besag [99] (i.e. ; a Gauss-Seidel relaxation), where pixels (semantic label of each region in our case) are updated one at a time. In our case, this algorithm turned out to be both easy to implement, fast and efficient in terms of convergence properties (the algorithm is fast converging after 100 iterations according to our experiments). The entire pseudo-code of our MC-SSM based on ICM is presented in Algorithm 2.

5.4 Experiments

5.4.1 Datasets

To evaluate the performance of our model, we compared it with different nonparametric methods, tested on two challenging semantic segmentation datasets ; Microsoft Research Cambridge dataset [181] and the Stanford background dataset [187].

5.4.1.1 Microsoft Research Cambridge Dataset (MSRC-21)

The MSRC-21 (v2) dataset[3] is an extension of the MSRC-9 (v1) dataset. It contains 591 color images with corresponding ground truth labelling for 23 object classes (building, grass, tree, cow, etc.). Among the 23 object classes, only 21 classes are commonly used. The unused labels are (void=0, horse=5, mountain=8) due to background or too few training samples.

5.4.1.2 Stanford Background Dataset (SBD)

The SBD dataset [4] contains a set of outdoor scene images imported from existing public datasets : LabelMe [188], MSRC [181], PASCAL VOC [189] and Geometric

Context [190]. Each image in this dataset contains at least one foreground object. The dataset is pixel-wise annotated (horizon location, pixel semantic class, pixel geometric class and image region) for evaluating methods for semantic scene understanding.

5.4.2 Evaluation Metrics

To provide a basis of comparison for the MC-SSM model, we quantitatively evaluate the annotation performance from two levels, which are widely used for evaluating the performances of related tasks. The first is the global (overall) per-pixel accuracy (GPA) which represents the total proportion of pixels correctly labeled. Mathematically, the global accuracy is computed as :

$$\text{GPA} = \frac{\sum_{i=1}^{n} \text{v}(x)}{n}, \;\; \text{v}(x) \begin{cases} 1 & y_i = l_i \\ 0 & otherwise \end{cases} \tag{5.6}$$

Where $v(.)$ denotes the indicator function, n is the number of pixels within the input image, y_i represents the label for pixel i predicted by the algorithm and l_i denotes the ground truth label for pixel i. The second level is the average per-class accuracy (ACA) which represents the average proportion of pixels correctly labeled in each category. Formally, the class-averaged accuracy is computed as follows :

$$\text{ACA} = \frac{1}{|C|} \sum_{c \in C} \frac{\sum_{i=1}^{n \times nb} \text{v}(y_i = l_i) \wedge \text{v}(l_i = c)}{\sum_{i=1}^{n \times nb} \text{v}(l_i = c)} \tag{5.7}$$

Where |C| denotes the number of classes within the input image, nb is the number of images in the dataset and $\wedge$ represents the logic operator *And*.

[3] The MSRC-21 dataset can be downloaded here :
http ://www.cs.cmu.edu/ tmalisie/projects/bmvc07/

[4] The SBD dataset is publicly accessible via this link :
http ://dags.stanford.edu/data/iccv09Data.tar.gz

5.4.3 Results and Discussion

To validate our model on the MSRC-21 dataset, we adopt the leave-one-out evaluation strategy. Thus, for each image, we use it as a query image and we classify its region based on the rest of the images in the dataset.

To guarantee the integrity of the benchmark results, the seven weight parameters of our algorithm [i.e., α_1, α_2, α_3, α_4, α_5, α_6 and α_7, see (5.5)] are optimized on the ensemble of 276 training images by using a local linear search procedure in the feasible ranges of parameter values ($[1:2]$) with a fixed step-size $= 10^{-2}$. We have found that $\alpha_1 = 1.83$, $\alpha_2 = 1.53$, $\alpha_3 = 1.55$, $\alpha_4 = 1.44$, $\alpha_5 = 1.35$, $\alpha_6 = 1.70$ and $\alpha_7 = 1$, are reliable hyper-parameters for the model yielding the best performance.

As we show in Table 5.2, MC-SSM outperforms the nonparametric SuperParsing method [196] with a GPA and ACA scores equal to, respectively, 0.75 and 0.63 (we perform tests on the 315 test images). Also, compared with state-of-the-art parametric methods, our method gives good results while not requiring expensive model training and being much simpler. It is worth mentioning that parametric scene parsing methods have a small advantage in accuracy over nonparametric methods. However they require large amounts of model training, making them less practical for open datasets [191]. The confusion matrix experimented from the MSRC-21 dataset is shown in Table 5.3. From this table we can see that better result in terms of class-accuracy is yielded for the following classes ; *sky*, *grass*, *aeroplane*, *sheep* and *book*, with values are higher than 80%. However, lower accuracy is achieved for the *chair* class with a value equal to 17.6%, this class is often confused with the *bird* class due to the similarity in color and texture between these two classes. Additionally, we present a qualitative comparison with other methods ; Unary [192], Auto context [193] and Geodesic [194] (see Fig. 5.4). Also, Fig. 5.5 and Fig. 5.6 show example results of success and failures on the MSRC-21 generated by our algorithm, respectively.

Also, we validated our model on the SBD dataset and we adopt the same evaluation strategy, the leave-one-out, but for the entire dataset as we used the same value of the parameters fixed on the training set of the MSRC-21 dataset. Table 5.4 shows that our

model is still competitive with different methods with a GPA value equal to 0.68 and ACA value equal to 0.53. These values are less better compared to those achieved on the MSRC-21 dataset. This result is not surprising, because the SBD dataset contains a foreground class that refers to different types of objects which increases significantly the intra-class variability.

Table 5.5 shows the confusion matrix for our model in the SBD dataset. From this table, we can note that better result in terms of class-accuracy is yielded for the following classes ; *sky* and *grass* classes, with values are higher than 80%. In contrast, lower accuracy is achieved for the *mountain* class with a value equal to 15.5%.

Algorithm 2 MC-Semantic Segmentation Model algorithm

Mathematical notation:

$\overline{\text{MC}}$	Multi-criteria function
$\{\mathcal{I}_k\}_{k\leq K}$	Set of K images
$\{S_k\}_{k\leq K}$	Set of K semantic segmentations (related to $\{\mathcal{I}_k\}_{k\leq K}$)
$\mathcal{E}$	Set of class labels in $\{S_k\}_{k\leq K}$
T_{max}	Maximal number of iterations (=100)
$\hat{S}_{MC}$	Semantic segmentation result
I	Image to be labeled
R_I	Region segmentation of image I

Input: I, $\{\mathcal{I}_k\}_{k\leq K}$, $\{S_k\}_{k\leq K}$
Output: $\hat{S}_{MC}$

A. Initialization:

Segment image I into different coherent regions R_I (with the GCEBFM algorithm)
Assign class label for each r_i region $\in R_I$ using random element from $\mathcal{E}$

B. Steepest Local Energy Descent:

while $p < T_{\text{max}}$ **do**
  for each r_i region $\in R_I$ **do**
    Draw a new class label y according to the uniform distribution in the set $\mathcal{E}$
    Let $R_I^{[p],\text{new}}$ the new semantic segmentation map including r_i with the class label y
    Compute $\overline{\text{MC}}\left(I, R_I^{[p],\text{new}}, S, \{\mathcal{I}_k, \mathcal{S}_k\}_{k\leq K}\right)$ *[see (5.5)]*
    if $\overline{\text{MC}}\left(I, R_I^{[p],\text{new}}, S, \{\mathcal{I}_k, \mathcal{S}_k\}_{k\leq K}\right) < \overline{\text{MC}}\left(I, R_I^{[p]}, S, \{\mathcal{I}_k, \mathcal{S}_k\}_{k\leq K}\right)$ **then**
      $\overline{\text{MC}} = \overline{\text{MC}}^{\text{new}}$
      $R_I^{[p]} = R_I^{[p],\text{new}}$
      $\hat{S}_{MC} = R_I^{[p]}$
    end if
  end for
  $p \leftarrow p + 1$
end while

We have also tested the effects of varying the retrieval set size *K* in Fig. 5.7. This test shows that $K = 197$ (the 1/3 of the dataset) is a reliable value that yielding the best

TABLE 5.2 : Performance of our model on the MSRC-21 segmentation dataset in terms of global per-pixel accuracy and average per-class accuracy (higher is better).

ALGORITHMS	PERFORMANCE MEASURES	
	Global (GPA)	Average (ACA)
Nonparametric (non-learning-based) methods		
MC-SSM	0.75	0.63
SuperParsing [196] in [197]	0.62	-
Parametric (learning-based) methods		
SVM on segment [166]	0.51	-
CRF on segment [166]	0.64	-
CRF+N=2 [198] in [166]	0.68	-
CRF+N=3 [198] in [166]	0.68	-
SVM on region [166]	0.69	-
Tree model [166]	0.70	-
TextonBoost [181]	0.72	0.58
Graphical model [199]	0.75	0.65
Auto-context [193]	0.75	-
GP [200]	0.72	-

TABLE 5.3 : Accuracy of segmentation for the MSRC 21-class dataset. Confusion matrix with percentages row-normalized. The overall per-pixel accuracy is 75%.

TRUE CLASS \ INFERRED CLASS	building	grass	tree	cow	sheep	sky	aeroplane	water	face	car	bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat
building	**53.6**	3.6	3.6		3.6	1.2	2.4	3.6		2.4			1.2	9.5			3.6	7.1	2.4	1.2	1.2
grass		**89.9**	2.9		0.7			0.7						0.7			3.6	0.7	0.7		
tree	5.6	12.5	**55.6**	2.8	2.8		1.4							12.5			1.4	4.2			1.4
cow				**72.7**	9.1									9.1					9.1		
sheep				5.0	**80.0**									5.0				5.0	5.0		
sky						**95.1**		2.4						1.2					1.2		
aeroplane	6.2						**87.5**			6.2											
water		2.5	2.5			10.0		**57.5**						2.5			25.0				
face									**69.7**	3.0				9.1			3.0		6.1	6.1	3.0
car	8.3						8.3			**58.3**											25.0
bicycle	11.8										**64.7**			11.8				11.8			
flower		5.6		5.6						5.6		**61.1**	5.6		5.6		5.6	5.6			
sign								5.6	5.6				**72.2**	5.6		5.6	5.6				
bird			5.0		10.0								5.0	**50.0**		5.0		15.0	5.0	5.0	
book			5.6		11.1										**83.3**						
chair	11.8			11.8	11.8							5.9		17.6		**17.6**		17.6	5.9		
road	5.7	2.3						8.0				1.1		1.1		1.1	**72.4**	2.3	3.4	1.1	1.1
cat				7.7	7.7										7.7			**76.9**			
dog	6.2			12.5	18.8									18.8	6.2			12.5	**18.8**	6.2	
body	2.7			2.7	2.7		2.7		2.7				2.7	24.3			2.7	5.4		**48.6**	2.7
boat	23.5						5.9			5.9			5.9					17.6	5.9		**35.3**

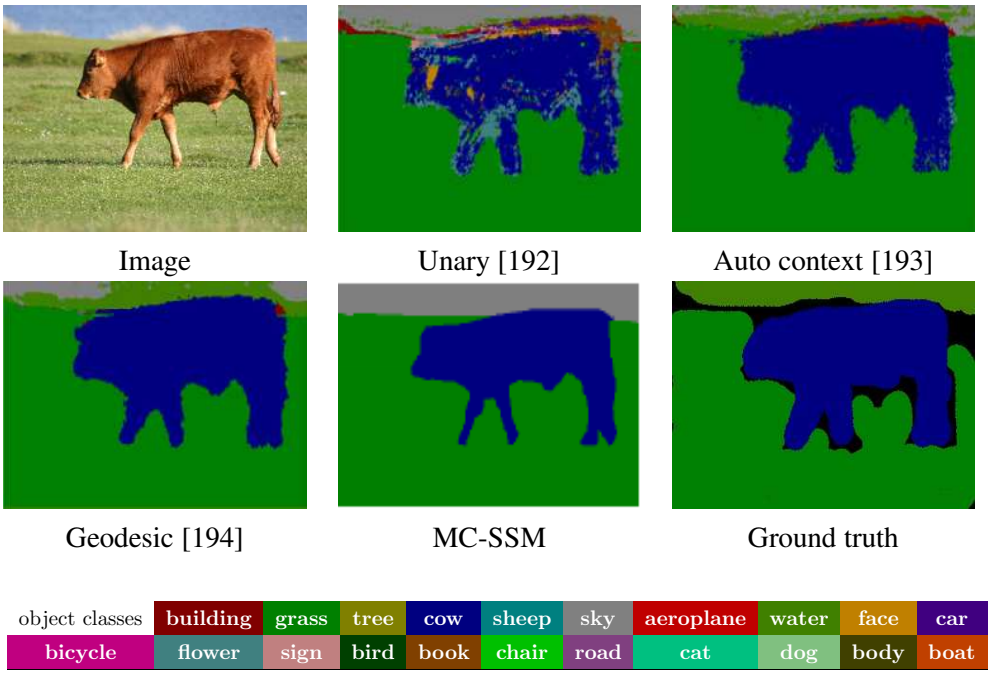


FIGURE 5.4 : Example of segmentation result obtained by our algorithm MC-SSM on an input image from the MSRC-21 compared to other algorithms.

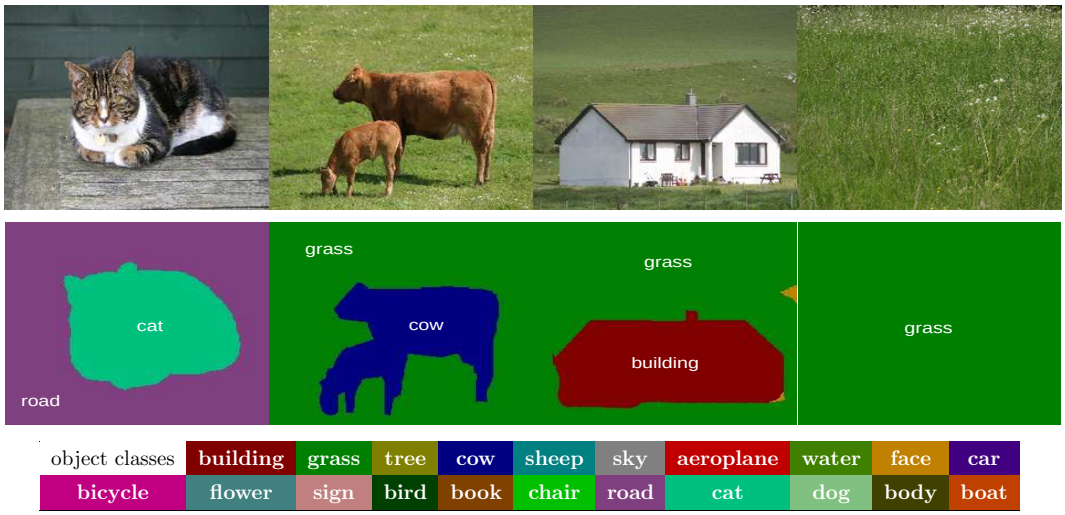


FIGURE 5.5 : Example results obtained by our MC-SSM model on the MSRC-21 dataset (for more clarity, we have superimposed textual labels on the resulting segmentations).

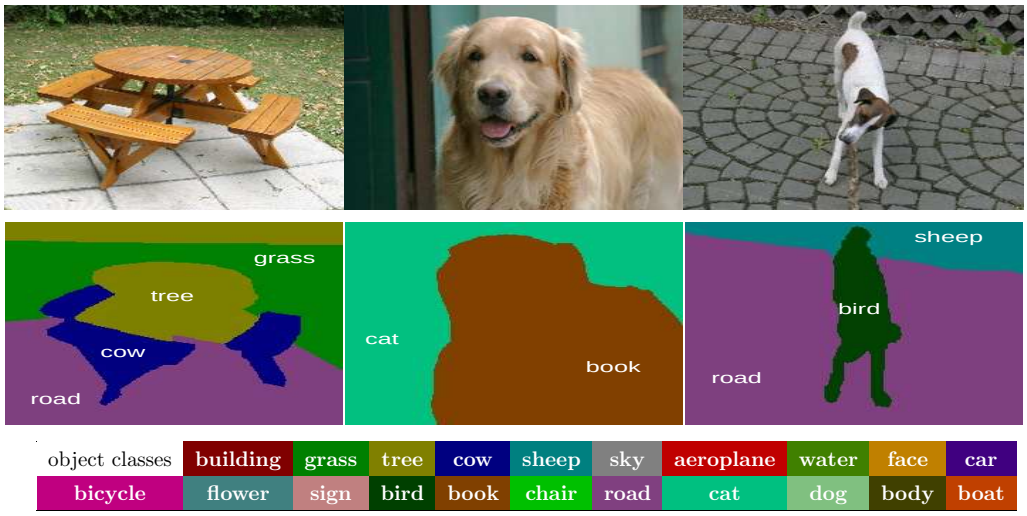


FIGURE 5.6 : Example results of failures on the MSRC-21 dataset. Top : query image, Bottom : predicted labeling.

TABLE 5.4 : Performance of our model on the Stanford background dataset (SBD) in terms of global per-pixel accuracy and average per-class accuracy (higher is better).

ALGORITHMS	PERFORMANCE MEASURES	
	Global (GPA)	Average (ACA)
Nonparametric (non learning-based) methods		
MC-SSM	0.68	0.62
SuperParsing [196]	0.76	-
Parametric (learning-based) methods		
SVM on segment [166]	0.51	-
CRF on segment [166]	0.62	-
CRF+N=2 [198] in [166]	0.67	-
CRF+N=3 [198] in [166]	0.66	-
SVM on region [166]	0.69	-
Tree model [166]	0.69	-
Leaf Level [195]	0.73	0.58

TABLE 5.5 : Accuracy of segmentation for the SBD dataset. Confusion matrix with percentages row-normalized. The overall per-pixel accuracy is 68%.

		INFERRED CLASS							
		sky	tree	road	grass	water	building	mountain	foreground
TRUE CLASS	sky	**92.3**	2.5	2.7		1.5	0.2	0.2	0.7
	tree	0.4	**32.1**	2.7	1.6	0.4	5.7	1.8	55.4
	road	1.3	2.0	**80.2**	1.4	8.1	1.8	1.8	3.4
	grass		9.8	9.3	**52.1**	2.1	18.0	4.6	4.1
	water	5.2	2.1	35.1	1.0	**43.3**	8.2	3.1	2.1
	building	0.4	12.5	3.7	0.7	0.4	**74.1**	1.1	7.1
	mountain	4.2	35.2	16.9	2.8	4.2	12.7	**15.5**	8.5
	foreground	0.6	33.6	3.9	1.1	2.0	15.5	5.0	**38.4**

accuracy for our model. As another evaluation test, in Table 5.6 we report the results of our model using single criterion and multiple criteria. We can see that color histogram, OCLBP and Laplacian operator histogram are the best criteria. Also, if we compare our results (in bold) to mono-criterion approach we obtain better results. This shows clearly that our strategy of combining different criteria is effective. In order to test the convergence properties of our iterative optimization procedure, we have tested our algorithm with different random initializations (step 2 in Algorithm 2) and we have found similar results, this result shows clearly that the consensus cost function [see Eq. (5.5)] is nearly convex. This also means that the proposed semantic labelling model is numerically rendered well-posed (and the optimization problem tractable) thanks to appropriate convex constraints or appropriate feature descriptors for this kind of problem. Also, we have evaluated the proposed model with different iteration numbers of the optimization algorithm and we have found that $T_{max} = 100$ is the best value which gives the asymptotic result in terms of GPA and ACA on the MSRC-21 dataset (see Fig. 5.8).

As we can notice, our multi-criteria semantic segmentation model (MC-SSM) is both simple and efficient and can be regarded as a robust alternative to complex, computationally demanding semantic segmentation models existing in the literature. Finally, it is worth mentioning that improvements can be made efficiently in our algorithm by adding other interesting invariant features (to the multi-criteria function) such as the SIFT (scale-invariant feature transform) or the LSD (line segment detector) descriptors or other similarity measures between segmentations.

5.4.4 Computation Time

The computational complexity of the proposed model depends on two factors ; the number of the images in the dataset and the number of the used criteria (combined as a global energy function). On the MSRC-21 dataset, the execution time takes, on average, between 5 and 6 minutes for an Intel 64 Processor core i7-4800MQ, 2.7 GHz, 8 GB of RAM memory and non-optimized code running on Linux for a 240×240 image. More accurately, the labeling process takes 0.14 second and the geometric retrieval step takes 0.32 second. However, the computation time of the proposed model

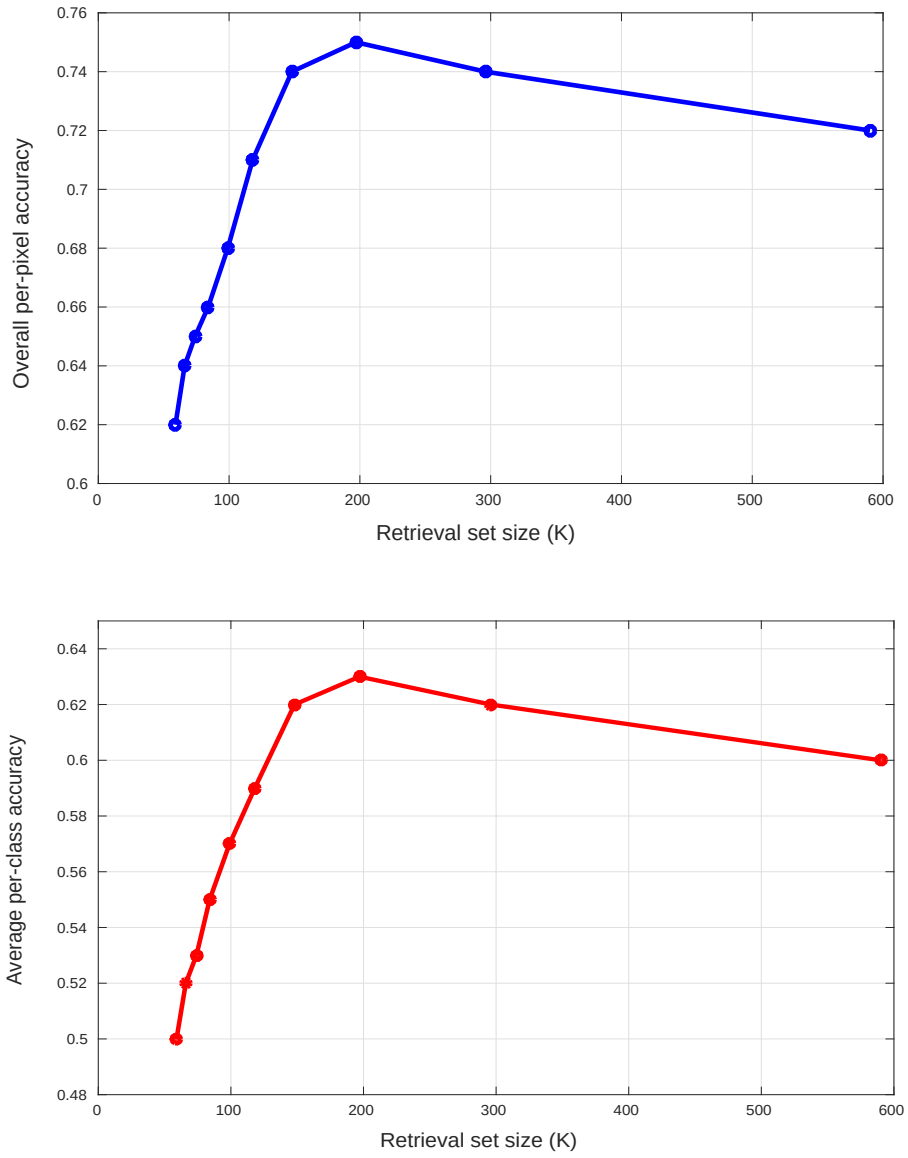


FIGURE 5.7 : Effects of varying the retrieval set size K for the MSRC-21 dataset ; shown are the overall per-pixel accuracy and the average per-class accuracy.

TABLE 5.6 : Performance of our model using single and multiple criteria (on the MSRC-21 dataset).

	CRITERION	PERFORMANCE MEASURES	
		Global (GPA)	Average (ACA)
SINGLE CRITERION	CTX	0.13	0.07
	LOC	0.15	0.13
	SHA	0.19	0.13
	TEX	0.26	0.18
	OCLBP	0.54	0.43
	LAP	0.59	0.49
	COL	0.65	0.55
MULTIPLE CRITERIA	TEX+CTX	0.27	0.19
	TEX+CTX+LOC+SHA	0.38	0.23
	TEX+CTX+LOC+SHA+COL	0.71	0.58
	TEX+CTX+LOC+SHA+COL+OCLBP+LAP	**0.75**	**0.63**

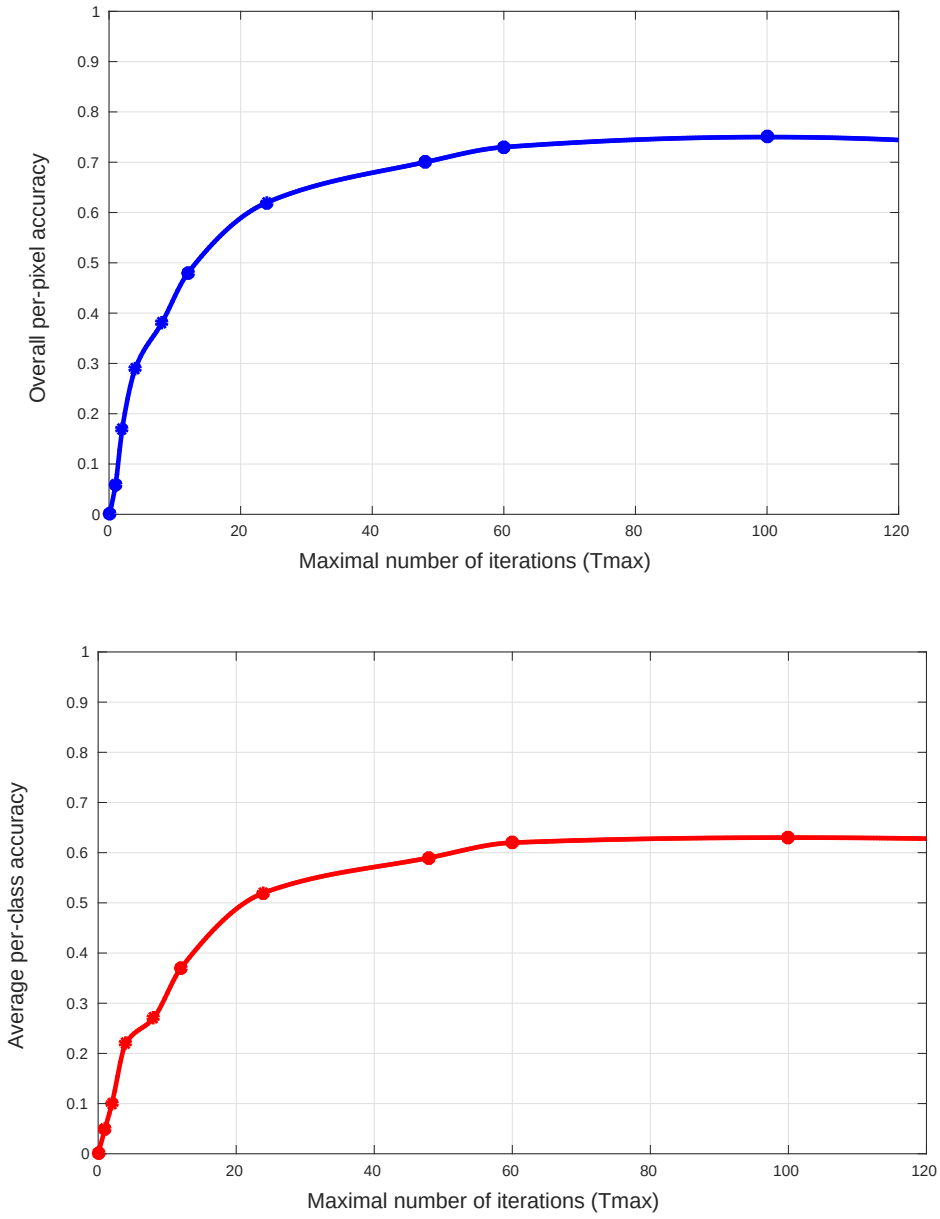


FIGURE 5.8 : Evolution of the overall per-pixel accuracy and the average global per–class accuracy along the number of iterations of the proposed MC-SSM starting from a random initialization on the MSRC-21 dataset.

(for each image) is mainly occupied by the region generation code with 205 seconds and the features extraction (from the full dataset) with 171 seconds. The former can be reduced by a parallelized implementation while the latter can be easily sped up by performing the extraction only once and then storing the extracted features on a data structure. The whole unoptimized and unparallelized implementation was developed using the C++ language and the source code, data and all that is necessary for the reproduction of results and the ensemble of labeled images are available at this http address ; http ://www-etud.iro.umontreal.ca/∼khelifil/ResearchMaterial/mc-ssm.html, in order to make possible comparisons with future scene parsing algorithms.

5.5 Conclusion

The aim of this present research was to address the problem of semantic segmentation (called also scene parsing). Towards this goal, we proposed a novel and simple energy-minimization based approach called the multi-criteria semantic segmentation model (MC-SSM). The proposed cost function of this model combines efficiently different global nonparametric semantic likelihood energy terms computed from the (pre-)segmented regions of the (query) image and defined according to their structural properties (location, texture, color, context and shape). To optimize our energy-based model we resort to a simple and local optimization procedure derived from the iterative conditional modes (ICM) algorithm. Our approach achieved state-of-the-art performance in two popular datasets (MSRC-21 and SBD). One area of future work will be, to improve further the classification accuracy by incorporating others criteria (possibly at different geometric and semantic abstraction levels).

CHAPITRE 6

CONCLUSION GÉNÉRALE ET PERSPECTIVES

L'objectif principal de notre thèse est d'apporter des solutions à deux problèmes importants de la vision par ordinateur, soit la segmentation et l'interprétation sémantique d'images. Dans un premier temps, nous synthétiserons nos contributions et dans un deuxième temps, nous discuterons les limitations ainsi que les orientations concernant les perspectives de ce travail.

6.1 Sommaire des contributions

La première partie de cette thèse a été consacrée à l'étude du problème de la fusion de segmentation mono-objectif. Nous avons présenté un nouveau modèle mono-objectif de fusion de segmentation basé sur le critère de l'erreur de la cohérence globale (GCE). De plus, nous avons ajouté à ce modèle un terme de régularisation permettant d'intégrer les connaissances concernant les types de segmentations résultantes fusionnées (définis à priori comme des solutions acceptables). Cette stratégie nous permet d'adapter le modèle avec la nature mal posée du problème de la segmentation. Les expérimentations faites sur la base de Berkeley ont montré l'efficacité de notre approche.

La deuxième partie de ce travail a porté sur la fusion de segmentation multi-objectif. Dans un premier temps, nous avons présenté un modèle de fusion basé sur deux critères complémentaires et contradictoires (la variation de l'information (VoI) et la F-mesure (précision-rappel)), l'optimisation de ce modèle est basée sur la méthode de pondération des fonctions objectives. Dans un deuxième temps, nous avons présenté un autre modèle multi-objectif qui s'appuie sur deux critères complémentaires (l'erreur de la cohérence globale (GCE) et la F-mesure (précision-rappel)). Pour optimiser notre modèle, nous avons utilisé une variante de l'ICM incluant une fonction de domination permettant de trouver un compromis (ensemble de solutions non dominées) entre ses différents critères

de segmentation. Puis, nous avons utilisé une technique efficace de prise de décision appelée TOPSIS, qui nous a permis de trouver la meilleure solution à partir de cet ensemble de solutions. Les tests que nous avons réalisés montrent des performances remarquables.

La troisième partie de ce travail a touché le sujet de l'interprétation sémantique d'images. En effet, nous avons proposé un nouveau système (non paramétrique) automatique d'étiquetage sémantique exploitant une base d'apprentissage d'image segmentée et pré-interprétée, et nous proposons un nouveau modèle à base d'énergie non paramétrique permettant d'inférer les classes les plus probables en nous basant sur différents critères dont celui de l'erreur de la cohérence globale (GCE) déjà utilisée pour le problème de la fusion de segmentation et combiné avec différents termes de vraisemblance sémantique non paramétrique. Le modèle ainsi proposé se réduit à un problème d'optimisation bien posé dont les différents termes d'énergie, permettent d'inférer la classe sémantique la plus adaptée, conduisent à une fonction d'énergie quasi convexe.

6.2 Limites et orientations futures de la recherche

D'autres pistes de recherche, liées à ce travail, méritent sans doute d'être approfondies et/ou explorées, offrant ainsi de nouvelles perspectives de recherche :

Fusion de segmentations mono-objectif

Compte tenu des limites de notre modèle de fusion de segmentations mono-objectif, nous n'avons pu analyser l'ensemble de ce sujet très vaste. Par exemple, nous avons remarqué que la tache de la fusion dépend de la qualité des cartes de segmentation initiales (à fusionner). Pour mieux diversifier cet ensemble de segmentations, il nous semblerait intéressant, à l'avenir, d'utiliser un ensemble de valeurs de l'histogramme de motifs binaires locaux (LBP) quantifiés ou un ensemble de valeurs de l'histogramme de quantification de phase locale (LPQ). Ces deux descripteurs pourraient être utilisés individuellement ou combinés avec le descripteur de l'histogramme couleur en tant que vecteur de fonctionnalité pour l'algorithme k-moyennes. Aussi, cette diversité peut être créée en utilisant plusieurs valeurs du paramètre de la taille du voisinage utilisé pour

définir la texture conduisant ainsi à une représentation multi-échelle des éléments de texture d'une image.

Fusion de segmentations multi-objectif

Pour le modèle de fusion de segmentations multi-objectif, certaines limitations peuvent être abordées et différentes orientations de travaux futures peuvent être explorées. Tout d'abord, le résultat final de fusion dépend de la combinaison de différents critères de fusion. Pour résoudre ce problème, nous travaillons à étendre notre approche par l'utilisation d'autres critères de fusion plus complémentaires. Aussi, pour surmonter le problème du temps de calcul, nous pouvons utiliser les capacités de calcul parallèle de processeur graphique (GPU), basé sur son architecture massivement parallèle, conçue pour gérer plusieurs tâches simultanément.

Interprétation sémantique d'images

Nous pensons qu'il est important d'améliorer ce système d'interprétation sémantique d'images en nous fondant sur d'autres critères, dans ce contexte, nous pourrions proposer dans de futurs travaux le même système, mais avec d'autres critères (descripteurs). Également, nous croyons que la sélection de k segmentations les plus proches au sens d'autres critères tels que le VoI, la F-mesure ou le PRI pourrait permettre l'amélioration du résultat final d'étiquetage. Aussi, une piste de recherche future consiste à combiner ce système non paramétrique avec un autre système paramétrique.

Imagerie fonctionnelle cérébrale

La méthode de fusion développée dans ce travail pourrait être appliquée dans le domaine de l'imagerie fonctionnelle cérébrale qui cherche à caractériser le cerveau :

- En effet, nous pourrons faire une segmentation moyenne consensuelle d'un ensemble de cerveaux au sens d'un certain critère qui sera intéressant pour une maladie spécifique telles que l'*Alzheimer* ou le *Parkinson*.

- À l'inverse, nous pourrons aussi imaginer une segmentation moyenne dissensus, plus précisément, nous pourrons chercher une segmentation qui donne la différence la plus grande à un ensemble de segmentations de cerveaux au sens d'un

critère pour quantifier ce qui serait la structure de l'Alzheimer dans ses différents modes de pathologie propre.

- Dans le domaine de l'imagerie fonctionnelle, différent machines qui captent un cerveau donnent différents résultats, suivant cette hypothèse un consensus ou un dissensus au sens d'un certain critère pourrait être intéressant pour étudier la similarité ou la différence en termes de mode d'acquisition et la fiabilité d'un appareillage et/ou son caractère reproductible.

- Également, nous pourrions générer une carte de segmentation hybride à travers de différentes cartes de segmentations fonctionnelles et des cartes de segmentations structurelles, dont le but d'avoir une structure de segmentation plus intéressante en matière de régions, et ainsi définir la position de la structure anatomique dans la cartographie fonctionnelle du cerveau.

Segmentation de textures dynamiques

Nous pourrions proposer une nouvelle approche basée sur la fusion des différents résultats de segmentation pour segmenter une séquence vidéo contenant des textures dynamiques naturelles. C'est une piste pour l'avenir, mais il importe de réfléchir à un descripteur capable de distinguer des textures similaires dans une même scène. Dans le même contexte, nous pourrons considérer aussi un cerveau humain, caractérisé par des données d'IRM fonctionnelle, comme une structure dynamique, composée de plusieurs textures dynamiques (en matière de signal fonctionnel) en action.

Classification de cerveaux segmentés structurellement

La notion de segmentation de consensus ou segmentation moyenne permettrait de générer en analyse fonctionnelle ou structurelle des prototypes de cerveaux ou des modes de prototypes de cerveaux (en termes de pathologies et en utilisant un algorithme de *clustering* tel que l'algorithme des K-moyennes exploitant une distance entre deux segmentations) permettant la classification de certaines pathologies structurelles ou fonctionnelles du cerveau et peut-être leurs liens.

Géo-imagerie

Le concept de carte de dissensus pourrait être aussi appliqué au domaine de géo-imagerie dans lequel on a une image avant et une image après, captée par différentes modalités (ex : SAR et optique). Le but est de chercher le changement de détection en matière de segmentation, donc il serait intéressant de réfléchir à un modèle spécifique pour ce genre du problème basé sur un ensemble de segmentations de l'image avant et un ensemble de segmentations de l'image après.

BIBLIOGRAPHIE

[1] S. W. Zucker. Survey : Region growing : Childhood and adolescence. *Computer Vision, Graphics, and Image Processing*,5(3) :382–399, 1976. doi:10.1016/S0146-664X(76)80014-7.

[2] M. Mignotte. Segmentation by fusion of histogram-based K-means clusters in different color spaces. *IEEE Transactions on Image Processing*, 17 :780–787, 2008. doi:10.1109/TIP.2008.920761.

[3] 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.

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

[5] 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*, 47 (9) :2489-2502, 2017. doi:10.1109/TSMC.2016.2531645.

[6] 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.

[7] S. Niu, Q. Chen, L. de Sisternes, Z. Ji, Z. Zhou, and D. L. Rubin. Robust noise region-based active contour model via local similarity factor for image segmentation. *Pattern Recognition*, 61 :104 – 119, 2017. doi:10.1016/j.patcog.2016.07.022.

[8] X. Bresson, S. Esedoğlu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *J. Math. Imaging Vis.*, 28(2) :151–167, June 2007. doi:10.1007/s10851-007-0002-0.

[9] H. Ali, N. Badshah, K. Chen, and G. A. Khan. A variational model with hybrid images data fitting energies for segmentation of images with intensity inhomogeneity. *Pattern Recognition*, 51 :27 – 42, 2016. doi:10.1016/j.patcog.2015.08.022.

[10] J. Yuan, S. S. Gleason, and A. M. Cheriyadat. Systematic benchmarking of aerial image segmentation. *IEEE Geoscience and Remote Sensing Letters*, 10(6) :1527–1531, Nov 2013. doi:10.1007/10.1109/LGRS.2013.2261453.

[11] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5) :898–916, May 2011. doi:10.1109/TPAMI.2010.161.

[12] 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.

[13] R. Nock and F. Nielsen. Statistical region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11) :1452–1458, Nov 2004. doi:10.1109/tpami.2004.110.

[14] D. Comaniciu and P. Meer. Mean shift : A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5) :603–619, 2002. doi:10.1109/34.1000236.

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

[16] X. Liu and D. L. Wang. A spectral histogram model for texton modeling and texture discrimination. *Vision Research*, 42(23) :2617 – 2634, 2002. doi:10.1109/10.1016/S0042-6989(02)00297-3.

[17] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *Journal of Photogrammetry and Remote Sensing*, 58(3-4) :239 – 258, 2004. doi:10.1016/j.isprsjprs.2003.10.002.

[18] 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 *Proc. 8th Int'l Conf. Computer Vision (ICCV'01)*, volume 2, pages 416–423, July 2001. doi:10.1109/ICCV.2001.937655.

[19] 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.

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

[21] 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 Transactions on Biomedical Engineering*, 52(10) :2041–2057, 2005. doi:10.1109/TBME.2005.855717.

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

[23] F. Destrempes and M. Mignotte. Localization of shapes using statistical models and stochastic optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9) :1603–1615, 2007. doi:10.1109/TPAMI.2007.1157.

[24] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006. DIP:1076432.

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

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

[27] 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*, 2013(1) :63, 2013. doi:10.1186/1687-5281-2013-63.

[28] 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) :191–204, 2013. doi:10.14257/ijsip.2013.6.5.17.

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

[30] 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.

[31] 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.

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

[33] Y. Ma, H. Derksen, W. Hong, and J. Wright. Segmentation of multivariate mixed data via lossy coding and compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9) :1546–1562, 2007. doi:10.1109/TPAMI.2007.1085.

[34] Y. Wu, P. Zhang, M. Li, Q. Zhang, F. Wang, and L. Jia. SAR image multiclass segmentation using a multiscale and multidirection triplet Markov fields model in nonsubsampled contourlet transform domain. *Information Fusion*, 14(4) :441 – 449, 2013. doi:10.1016/j.inffus.2012.12.001.

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

[36] R. Hedjam 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.

[37] S. Chatzis and G. Tsechpenakis. The infinite hidden Markov random field model. *IEEE Transactions on Neural Networks*, 21(6) :1004–1014, 2010.

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

[39] 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) :219–232, 2013. doi:10.1.1.399.5301.

[40] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :888–905, 2000. doi:10.1109/34.868688.

[41] 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.

[42] L. Bertelli, B. Sumengen, B. Manjunath, and F. Gibou. A variational framework for multi-region pairwise similarity-based image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8) :1400–1414, 2008. doi:10.1109/TPAMI.2007.70785.

[43] 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.

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

[45] Y. Chen, A. B. Cremers,and Z. Caoo. Interactive color image segmentation via iterative evidential labeling. *Information Fusion*, 20 :292 – 304, 2014. doi:10.1016/j.inffus.2014.03.007.

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

[47] Y. Wang and C. He. Image segmentation algorithm by piecewise smooth approximation. *EURASIP Journal on Image and Video Processing*, 2012(1) :16, 2012. doi:10.1186/1687-5281-2012-16.

[48] 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.

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

[50] 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.

[51] 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) :13–24, 2013. 10.14257/ijsip.2013.6.5.02.

[52] S. Chabrier, C. Rosenberger, B. Emile, and H. Laurent. Optimization-based image segmentation by genetic algorithms. *EURASIP Journal on Image and Video Processing*, 2008(1), 2008. doi:10.1155/2008/842029.

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

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

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

[56] G. U. 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) :63–76, 2011. sn:167075.

[57] 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.

[58] A. Strehl and J. Ghosh. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal on Machine Learning Research, JMLR*, 3 :583–617, 2001. doi:10.1162/153244303321897735.

[59] 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.

[60] 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, 2011. doi:10.1142/S0218001411008683.

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

[62] J. Keuchel and D. KÃ$\frac{1}{4}$ttel. Efficient combination of probabilistic sampling approximations for robust image segmentation. In *DAGM-Symposium, Lecture Notes in Computer Science*, pages 41–50, 2006. doi:10.1007/11861898_5.

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

[64] Y. Collette and P. Siarry. *Multiobjective optimization : principles and case studies.* Springer-Veralg BerlinHiedelberg, 2004. doi:10.1007/978-3-662-08883-8.

[65] W. Tao, H. Jin, and Y. Zhang. Color image segmentation based on mean shift and normalized cuts. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 37(5) :1382–1389, Oct 2007. doi:10.1109/TSMCB.2007.902249.

[66] D. Parikh and R. Polikar. An ensemble-based incremental learning approach to data fusion. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 37(2) :437–450, April 2007. doi:10.1109/TSMCB.2006.883873.

[67] L. I. Kuncheva. Switching between selection and fusion in combining classifiers : an experiment. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 32(2) :146–156, Apr 2002. doi:10.1109/3477.990871.

[68] A. J. Sharkey. *Combining artificial neural nets ensemble and modular multi-net systems*. Springer-Verlag, New York, Inc., ISBN :185233004X, 1999. doi:10.1007/978-1-4471-0793-4.

[69] T. Dietterich. Ensemble methods in machine learning. In Lecture Notes In Computer Science, editor, *Proceedings of the First International Workshop on Multiple Classifier Systems, LNCS, Multiple Classifier Systems*, volume 1857, pages 1–15. Springer, 2000. doi:10.1007/3-540-45014-9_1.

[70] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336) :846–850, 1971. doi:10.2307/2284239.

[71] T. M. Nguyen and Q. Wu. Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 42(1) :193–202, Feb 2012. doi:10.1109/TSMCB.2011.2161284.

[72] 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*, 2012. doi:10.1186/1687-5281-2012-11.

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

[74] A. Alush and J. Goldberger. Ensemble segmentation using efficient integer linear programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10) :1966–1977, 2012. doi:10.1109/TPAMI.2011.280.

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

[76] M. Mignotte. A label field fusion model with a variation of information estimator for image segmentation. *Information Fusion*, 20(0) :7–20, 2014. doi:10.1016/j.inffus.2013.10.012.

[77] C. Hélou and M. Mignotte. A precision-recall criterion based consensus model for fusing multiple segmentations. *International Journal of Signal Processing*, 7(3) :61–82, 2014. doi:10.14257/ijsip.2014.7.3.07.

[78] 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.

[79] 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.

[80] L. Franek, D. Abdala, S. Vega-Pons, and X. Jiang. *Image Segmentation Fusion Using General Ensemble Clustering Methods*, pages 373–384. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-19282-1_30.

[81] M. Ozay, F. T. Y. Vural, S. R. Kulkarni, and H. V. Poor. Fusion of image segmentation algorithms using consensus clustering. In *2013 IEEE International Conference on Image Processing*, pages 4049–4053, Sept 2013. doi:10.1109/ICIP.2013.6738834.

[82] S. Vega-Pons and J. Ruiz-Shulcloper. A survey of clustering ensemble algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(03) :337–372, 2011. doi:10.1142/S0218001411008683.

[83] R. Unnikrishnan and M. Hebert. Measures of similarity. In *Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05) - Volume 1 -*, pages 394–394, Washington, DC, USA, 2005. doi:10.1109/ACVMOT.2005.71.

[84] A. Ben-hur, A. Elisseeff and I. Guyon. A Stability Based Method for Discovering Structure in Clustered Data. in *Pacific Symposium on Biocomputing*, pages 6–17, 2002. doi:10.1142/9789812799623_0002.

[85] B. Mirkin. *Mathematical Classification and Clustering*. Non-convex Optimization and Its Applications. Springer US, 1996. doi:10.1007/978-1-4613-0457-9.

[86] S. V. Dongen. Performance criteria for graph clustering and markov cluster experiments. Technical report, Amsterdam, The Netherlands, 2000. doi:10.1.1.26.9783.

[87] Y. Zhao and G. Karypis. Criterion functions for document clustering : Experiments and analysis. Technical report, 2002. doi:10.1.1.16.6872.

[88] A. Rosenberg and J. Hirschberg V-measure : A conditional entropy-based external cluster evaluation measure. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning(EMNLP-CoNLL)*, pages 410–420, 2007. link.

[89] S. Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper. Weighted partition consensus via kernels. *Pattern Recognition*, 43(8) :2712 – 2724, 2010. doi:10.1016/j.patcog.2010.03.001.

[90] S. Vega-Pons, J. Correa-Morris, and J. Ruiz-Shulcloper. *Weighted Cluster Ensemble Using a Kernel Consensus Function*, pages 195–202. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-85920-8_24.

[91] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1) :37–46, 1960. doi:10.1177/001316446002000104.

[92] M. Banerjee, M. Capozzoli, L. McSweeney, and D. Sinha. Beyond kappa : A review of interrater agreement measures. *Canadian Journal of Statistics*, 27(1) :3–23, 1999. doi:10.2307/3315487.

[93] D. R. Martin. *An Empirical Approach to Grouping and Segmentation*. PhD thesis. University of California, 2002. link.

[94] D. Martin, 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, May 2004. doi:10.1109/TPAMI.2004.1273918.

[95] R. Unnikrishnan, C. Pantofaru, and M. Hebert. Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29 :929–944, 2007. doi:10.1109/TPAMI.2007.1046.

[96] L. Chen, C. Chen, and M. Lu. A multiple-kernel fuzzy c-means algorithm for image segmentation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics* , 41(5) :1263–1274, Oct 2011. doi:10.1109/TSMCB.2011.2124455.

[97] A. Lorette, X. Descombes, and J. Zerubia. Fully unsupervised fuzzy clustering with entropy criterion. In *Proc. International Conference on Pattern Recognition (ICPR'00)*, Barcelone, Espagne, September 2000. doi:10.1109/ICPR.2000.903710.

[98] I. Ben Ayed and A. Mitiche. A region merging prior for variational level set image segmentation. *IEEE Transactions on Image Processing*, 17(12) :2301–2311, 2008. doi:10.1109/TIP.2008.2006425.

[99] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, B-48 :259–302, 1986. doi:10.1080/02664769300000059.

[100] S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2) :129–136, 1982. doi:10.1109/TIT.1982.1056489.

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

[102] S. Chitroub. Classifier combination and score level fusion : concepts and practical aspects. *International Journal of Image and Data Fusion*, 1(2) :113–135, 2010. doi:10.1080/19479830903561944.

[103] R. Unnikrishnan, C. Pantofaru, and M. Hebert. A measure for objective evaluation of image segmentation algorithms. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Workshop on Empirical Evaluation Methods in Computer Vision*, volume 3, pages 34–41, June 2005. doi:10.1109/CVPR.2005.390.

[104] R. Huanga, N. Sangb, D.Luoc, and Q. Tangd. Image segmentation via coherent clustering in l*a*b* color space. *Pattern Recognition Letters*, 32(7) :891–902, 2011. doi:10.1016/j.patrec.2011.01.013.

[105] E. Sharon, M. Galun, D. Sharon, R. Basri, and A. Brandt. Hierarchy and adaptivity in segmenting visual scenes. *Nature*, 442 :810–813, 2006. doi:10.1038/nature04977.

[106] M. Meila. Comparing clusterings–an information based distance. *Journal of Multivariate Analysis*, 98(5) :873–895, 2007. doi:10.1016/j.jmva.2006.11.013.

[107] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi. Yet another survey on image segmentation : Region and boundary information integration. In *Proc. 7th European Conference on Computer Vision (ECCV02)*, pages 408–422, 2002. doi:10.1007/3-540-47977-5_27.

[108] P.-M. Jodoin and M. Mignotte. Markovian segmentation and parameter estimation on graphics hardware. *Journal of Electronic Imaging*, 15(3) :033015–1–15, July-September 2006. doi:10.1117/1.2238881.

[109] H. Wang, Y. Zhang, R. Nie, Y. Yang, B. Peng, and T. Li. Bayesian image segmentation fusion. *Knowledge-Based Systems*, 71(1) :162–168, 2014. doi:10.1016/j.knosys.2014.07.021.

[110] L. Khelifi and M. Mignotte. A new multi-criteria fusion model for color textured image segmentation. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2579–2583, Sept 2016. doi:10.1109/ICIP.2016.7532825.

[111] L. Khelifi, I. Zidi, K. Zidi, and K. Ghedira. A hybrid approach based on multi-objective simulated annealing and tabu search to solve the dynamic dial a ride problem. In *International Conference on Advanced Logistics and Transport (ICALT), 2013*, pages 227–232, May 2013. doi:10.1109/ICAdLT.2013.6568464.

[112] B. C. Wei and R. Mandava. Multi-objective nature-inspired clustering techniques for image segmentation. In *2010 IEEE Conference on Cybernetics and Intelligent Systems*, pages 150–155, June 2010. doi:10.1109/ICCIS.2010.5518564..

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

[114] 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.

[115] A. A. Goshtasby and S. Nikolov. Image fusion : Advances in the state of the art. *Information Fusion*, 8(2) :114 – 118, 2007. doi:10.1016/j.inffus.2006.04.001.

[116] Y. Liu, S. Liu, and Z. Wang. Multi-focus image fusion with dense SIFT. *Information Fusion*, 23 :139 – 155, 2015. doi:10.1016/j.inffus.2014.05.004.

[117] E. Maggio and A. Cavallaro. Multi-part target representation for color tracking. In *IEEE International Conference on Image Processing 2005*, volume 1, pages I–729–32, Sept 2005. doi:10.1109/ICIP.2005.1529854.

[118] M. Millnert. Signal processing, image processing and pattern recognition, s. banks, prentice-hall, englewood cliffs, nj, 1990, isbn 0-13-812579-1, xiv + 410 pp., Â£22.95. *International Journal of Adaptive Control and Signal Processing*, 6(5) :519–520, 1992. doi:10.1002/acs.4480060511.

[119] Z. Kato and T-C. Pong. A Markov random field image segmentation model for color textured images. *Image and Vision Computing*, 24(10) :1103–1114, 2006. doi:10.1016/j.imavis.2006.03.005.

[120] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. *Color-Based Probabilistic Tracking*, pages 661–675. Springer Berlin Heidelberg, 2002. doi:10.1007/3-540-47969-4_44.

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

[122] H. Deng, C.-H. Yeh, and R. J. Willis. Inter-company comparison using modified TOPSIS with objective weights. *Computers and Operations Research*, 27(10) :963 – 973, 2000. doi:10.1016/S0305-0548(99)00069-6.

[123] T.-C. Wang and H.-D. Lee. Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. *Expert Systems with Applications*, 36(5) :8980 – 8985, 2009. doi:10.1016/j.eswa.2008.11.035.

[124] J. J. Lewis, R. J. O'Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah. Pixel- and region-based image fusion with complex wavelets. *Information Fusion*, 8(2) :119 – 130, 2007. doi:10.1016/j.inffus.2005.09.006.

[125] M. A. Jaffar. A dynamic fuzzy genetic algorithm for natural image segmentation using adaptive mean shift. *Journal of Experimental and Theoretical Artificial Intelligence*, 29(1) :149–156, 2017. doi:10.1080/0952813X.2015.1132263.

[126] M. B. Salah, I. B. Ayed, J. Yuan and H. Zhang. Convex-relaxed kernel mapping for image segmentation. *IEEE Transactions on Image Processing*, 23(3) :1143–1153, March 2014. doi:10.1109/TIP.2013.2297019.

[127] L. Dong, N. Feng,and Q. Zhang. Lsi : Latent semantic inference for natural image segmentation. *Pattern Recognition*, 59 :282 – 291, 2016. doi:10.1016/j.patcog.2016.03.005.

[128] S. Li and D. O. Wu. Modularity-based image segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(4) :570–581, April 2015. doi:10.1109/TCSVT.2014.2360028.

[129] A. Browet, P.-A. Absil, and P. V. Dooren. *Community Detection for Hierarchical Image Segmentation*, pages 358–371. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. doi:10.1007/978-3-642-21073-0_32.

[130] D. Brockhoff and E. Zitzler. *Are All Objectives Necessary ? On Dimensionality Reduction in Evolutionary Multiobjective Optimization*, pages 533–542. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. doi:10.1007/11844297_54.

[131] 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 Transactions on Image Processing*, 14(8) :1096–1108, Aug 2005. doi:10.1109/TIP.2005.851710.

[132] X. Wang, Y. Tang, S. Masnou, and L. Chen. A global/local affinity graph for image segmentation. *IEEE Transactions on Image Processing*, 24(4) :1399–1411, April 2015. doi:10.1109/TIP.2015.2397313.

[133] T. Blaschke. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1) :2– 16, 2010. doi:10.1016/j.isprsjprs.2009.06.004.

[134] C. Witharana, D. L. Civco, and T. H. Meyer. Evaluation of data fusion and image segmentation in earth observation based rapid mapping workflows.

ISPRS Journal of Photogrammetry and Remote Sensing, 87 :1 – 18, 2014. doi:10.1016/j.isprsjprs.2013.10.005.

[135] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9) :1277 – 1294, 1993. doi:10.1016/0031-3203(93)90135-J.

[136] R. Dass and S. Devi. Image segmentation techniques 1. *InternatIonal Journal of Electronics and Communication Technology*, 3(1) :66–70, 2012. ISSN:2230-7109.

[137] F. Y. Shih. Image segmentation. *Wiley-IEEE Press. Image Processing and Pattern Recognition : Fundamentals and Techniques*, vol. 110, no. 2, pp. 119–178, Apr. 2010. doi:10.1002/9780470590416.

[138] C. A. Coello and A. D. Christiansen. Multiobjective optimization of trusses using genetic algorithms. *Computers and Structures*, vol. 75, no. 6, pp. 647–660, 2000. doi:10.1016/S0045-7949(99)00110-8.

[139] A. Osyczka. Design Optimization. Multicriteria optimization for engineering design. *Design Optimization*,J. S. Gero, Ed. Academic Press., pp. 193–227, 1985. doi:10.1016/B978-0-12-280910-1.50012-X.

[140] L. Khelifi and M. Mignotte. A multi-objective approach based on topsis to solve the image segmentation combination problem. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, December 2016. doi:10.1109/ICPR.2016.7900296.

[141] B. Chin-Wei and M. Rajeswari. Multiobjective optimization approaches in image segmentation - the directions and challenges. *International on Advances in Soft Computing and its Applications*, 2(1) :40 – 65, 2010. issn:2074-8523.

[142] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986. isbn:0070544840.

[143] X. Jiang. An adaptive contour closure algorithm and its experimental evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11) :1252–1265, 2000. doi:10.1109/34.888710.

[144] C. L. Hwang and K. Yoon. Multiple attribute decision making. In *Lecture Notes in Economics and Mathematical Systems*, volume 186. Springer-Verlag Berlin, 1981. doi:10.1007/978-3-642-48318-9.

[145] E. Ataei. Application of topsis and fuzzy topsis methods for plant layout design. *World Applied Science Journal*, 23(12) :48–53, 2013. doi:10.5829/idosi.wasj.2013.23.12.975.

[146] G. Kim, C. S. Park, and K. Yoon. Identifying investment opportunities for advanced manufacturing systems with comparative-integrated performance measurement. *International Journal of Production Economics*, 50(1) :23 – 33, 1997. doi:10.1016/S0925-5273(97)00014-5.

[147] H.-S. Shih, H.-J. Shyur, and E. S. Lee. An extension of TOPSIS for group decision making. *Mathematical and Computer Modelling*, 45(7–8) :801 – 813, 2007. doi:10.1016/j.mcm.2006.03.023.

[148] M. A. Jaffar. A dynamic fuzzy genetic algorithm for natural image segmentation using adaptive mean shift. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(1) :149–156, 2017. doi:10.1080/0952813X.2015.1132263.

[149] C. Farabet, C. Couprie, L. Najman, and Y. LeCun. Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8) :1915–1929, Aug 2013. doi:10.1109/TPAMI.2012.231.

[150] B. Tung and J. J. Little. Scene parsing by nonparametric label transfer of content-adaptive windows. *Computer Vision and Image Understanding*, 143 :191 – 200, 2016. doi:10.1016/j.cviu.2015.08.009.

[151] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution,

and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99) :1–1, 2017. doi:10.1109/TPAMI.2017.2699184.

[152] E. Shelhamer, J. Long, C. Fowlkes, and D. Darrell. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4) :640–651, April 2017. doi:10.1109/TPAMI.2016.2572683.

[153] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1) :142–158, January 2016. doi:10.1109/TPAMI.2015.2437384.

[154] Y. Li, H. Qi, J. Dai, X. Ji and Y. Wei. Fully Convolutional Instance-Aware Semantic Segmentation. In *Proc. of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4438–4446, 2017. arxiv.org/abs/1611.07709.

[155] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik. Simultaneous Detection and Segmentation. In *Proc. 13th European Conference on Computer Vision (ECCV2014)*, pages 297—312, 2014. doi:10.1007/978-3-319-10584-0_20.

[156] L-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. In *3rd International Conference on Learning Representations (ICLR)*, pages 1–14, 2015. arxiv:1412.7062.

[157] C. Farabet, C. Couprie, L. Najman, and Y. Lecun. Scene parsing with Multiscale Feature Learning, Purity Trees, and Optimal Covers. In *29th International Conference on Machine Learning (ICML12*, pages 575–582, 2012. arxiv:abs/1202.2160.

[158] J. Tighe, M. Niethammer, and S. Lazebnik. Scene parsing with object instances and occlusion ordering. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3748–3755, 2014. doi:10.1109/CVPR.2014.479.

[159] J. Tighe and S. Lazebnik. Finding Things : Image Parsing with Regions and Per-Exemplar Detectors. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3001–3008, 2013. doi:10.1109/CVPR.2013.386.

[160] F. Schroff, A. Criminisi, and A. Zisserman. Object Class Segmentation using Random Forests. In *British Machine Vision Conference (BMVC)*, pages 1–10, 2008. doi:10.5244/C.22.54.

[161] P. Kontschieder, P. Kohli, J. Shotton, and A. Criminisi. Geof : Geodesic forests for learning coupled predictors. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 65–72, 2013. doi:10.1109/CVPR.2013.16.

[162] S. P. Narote, P. N. Bhujbal, A. S. Narote, and D. M. Dhane. A review of recent advances in lane detection and departure warning system. *Pattern Recognition*, 76(C) :216–234, 2018. doi:10.1016/j.patcog.2017.08.014.

[163] K. Li, W. Tao, X. Liu, and L. Liu. Iterative image segmentation with feature driven heuristic four-color labeling. *Pattern Recognition*, 76(C) : 69–79, 2018. doi:10.1016/j.patcog.2017.10.023.

[164] S. K. Choy, S. Y. Lam, K. W. Yu, W. Y. Lee, and K. T. Leung. Fuzzy model-based clustering and its application in image segmentation. *Pattern Recognition*, 68(C) : 141–157, 2017. doi:10.1016/j.patcog.2017.03.009.

[165] O. Gupta, D. Raviv, and R. Raskar. Illumination invariants in deep video expression recognition. *Pattern Recognition*, 76 (C) : 25–35, 2018. doi:10.1016/j.patcog.2017.10.017.

[166] J. Xie, L. Yu , L. Zhu and X. Chen. Semantic Image Segmentation Method with Multiple Adjacency Trees and Multiscale Features. *PCognitive Computation*, 9(2) : 168–179, 2017. doi:0.1007/s12559-016-9441-5.

[167] B. Shuai, Z. Zuo, G. Wang, and B. Wang. Scene parsing with integration of parametric and non-parametric models. *IEEE Transactions on Image Processing*, 25(5) :2379–2391, 2016. doi:10.1109/TIP.2016.2533862.

[168] S. Liu, X. Liang, L. Liu, X. Shen, J. Yang, C. Xu, L. Lin, X. Cao, and S. Yan. Matching-CNN meets KNN : Quasi-parametric human parsing. In *CVPR*, pages 1419–1427. IEEE Computer Society, 2015. doi:10.1109/CVPR.2015.7298748.

[169] X. An, S. Li, H. Qin, and A. Hao. Automatic non-parametric image parsing via hierarchical semantic voting based on sparse-dense reconstruction and spatial-contextual cues. *Neurocomputing*, 201 :92 – 103, 2016. doi:10.1016/j.neucom.2016.03.034.

[170] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding : Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81(1) :2–23, January 2009. doi:10.1007/s11263-007-0109-1.

[171] J. Shotton and P. Kohli. *Semantic Image Segmentation*, pages 713–716. Springer US, Boston, MA, 2014. doi:10.1007/978-0-387-31439-6_25.

[172] C. Liu, J. Yuen, and A. Torralba. *Nonparametric Scene Parsing via Label Transfer*, pages 207–236. Springer International Publishing, Cham, 2016. doi:10.1007/978-3-319-23048-1_10.

[173] C. Liu, J. Yuen, and A. Torralba. Sift flow : Dense correspondence across scenes and its applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5) :978–994, May 2011. doi:10.1109/TPAMI.2010.147.

[174] Z. Liu, X. Li, P. Luo, C. C. Loy, and X. Tang. Semantic image segmentation via deep parsing network. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1377–1385, Dec 2015. doi:10.1109/ICCV.2015.162.

[175] N. W. Campbell, W. Mackeown, B. T. Thomas and T. Troscianko. Interpreting image databases by region classification. *Pattern Recognition*, 30(4) :555 – 563, 1997. doi:10.1016/S0031-3203(96)00112-4.

[176] J. Tighe and S. Lazebnik. Superparsing : Scalable nonparametric image parsing with superpixels. *International Journal of Computer Vision*, 101(2) :329–349, 2013. doi:10.1007/s11263-012-0574-z.

[177] M. Zand, S. Doraisamy, A. Abdul Halin, and M. R. Mustaffa. Ontology-based semantic image segmentation using mixture models and multiple crfs. *IEEE Transactions on Image Processing*, 25(7) :3233–3248, July 2016. doi:10.1109/TIP.2016.2552401.

[178] H. Zhang, T. Fang, X. Chen, Q. Zhao, and L. Quan. Partial similarity based nonparametric scene parsing in certain environment. In *CVPR 2011*, pages 2241–2248, June 2011. doi:10.1109/CVPR.2011.5995348.

[179] L. Khelifi and M. Mignotte. Semantic image segmentation using the ICM algorithm. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3080–3084, Sept 2017.

[180] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller. Multi-class segmentation with relative location prior. *International Journal of Computer Vision*, 80(3) :300–316, Dec 2008. doi:10.1007/s11263-008-0140-x.

[181] J. Shotton, J. Winn, C. Rother and A. Criminisi. *TextonBoost : Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation*, In Proceedings of 9th European Conference on Computer Vision, pages 1–15, 2006. doi:10.1007/11744023_1.

[182] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11) :2274–2282, Nov 2012. doi:10.1109/TPAMI.2012.120.

[183] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7) :971–987, July 2002. doi:10.1109/TPAMI.2002.1017623.

[184] T. Maenpaa, M. Pietikainen, and J. Viertola. Separating color and pattern information for color texture discrimination. In *Object recognition supported by user interaction for service robots*, volume 1, pages 668–671, 2002. doi:10.1109/ICPR.2002.1044840.

[185] A. Joshi and A. K. Gangwar. Color local phase quantization (CLPQ)-a new face representation approach using color texture cues. In *2015 International Conference on Biometrics (ICB)*, pages 177–184, May 2015. doi:10.1109/ICB.2015.7139049.

[186] S. Gould and X. He. Scene understanding by labeling pixels. *Communications of the ACM*, 57(11) :68–77, October 2014. doi:10.1145/2629637.

[187] S. Gould, R. Fulton, and D. Koller. Decomposing a scene into geometric and semantically consistent regions. In *IEEE 12th International Conference on Computer Vision*, pages 1–8, Sept 2009. doi:10.1109/ICCV.2009.5459211.

[188] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. Labelme : A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77(1) :157–173, 2008. doi:10.1007/s11263-007-0090-8.

[189] M. Everingham, L. Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2) :303–338, June 2010. doi:10.1007/s11263-009-0275-4.

[190] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *International Journal of Computer Vision*, 75(1) :151–172, 2007. doi:10.1007/s11263-006-0031-y.

[191] F. Tung, J .J. Little, D. Fleet, T. Pajdla, and B. Schiele. T. Tuytelaars, CollageParsing : Nonparametric Scene Parsing by Adaptive Overlapping Windows. In *Proceedings of 13th European Conference on Computer Vision*, pp. 511–525, 2014. doi:10.1007/978-3-319-10599-4_33.

[192] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Associative hierarchical random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6) :1056–1077, June 2014. doi:10.1109/TPAMI.2013.165.

[193] Z. Tu and X. Bai. Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10) :1744–1757, Oct 2010. doi:10.1109/TPAMI.2009.186.

[194] V. Haltakov, C. Unger, and S. Ilic. Geodesic pixel neighborhoods for 2d and 3d scene understanding. *Computer Vision and Image Understanding*, 148 :164 – 180, 2016. Special issue on Assistive Computer Vision and Robotics -. doi:10.1016/j.cviu.2015.11.008.

[195] D. Munoz, J. A. Bagnell, and M. Hebert. Stacked hierarchical labeling. In *Proceedings of the 11th European Conference on Computer Vision : Part VI*, ECCV'10, pages 57–70, Berlin, Heidelberg, 2010. Springer-Verlag. doi:10.1007/978-3-642-15567-3_5.

[196] J. Tighe and S. Lazebnik. *SuperParsing : Scalable Nonparametric Image Parsing with Superpixels*, pages 352–365. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. doi:10.1007/978-3-642-15555-0_26.

[197] A. Bassiouny and M. El-Saban. Semantic segmentation as image representation for scene recognition. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 981–985, Oct 2014. doi:10.1109/ICIP.2014.7025197.

[198] B. Fulkerson, A. Vedaldi, and S. Soatto. Class segmentation and object localization with superpixel neighborhoods, In *Proceedings of 12th IEEE Int. Conf. Comput. Vis.* , pp. 670–677, 2009. doi:10.1109/ICCV.2009.5459175.

[199] L. Zhang and Q. Ji. Image segmentation with a unified graphical model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8) :1406–1425, Aug 2010. doi:10.1109/TPAMI.2009.145.

[200] Q. Li, X. Chen, Y. Song, Y. Zhang, X. Jin, and Q. Zhao. Geodesic propagation for semantic labeling. *IEEE Transactions on Image Processing*, 23(11) :4812–4825, Nov 2014. doi:10.1109/TIP.2014.2358193.

Annexe I

Opérateurs de quantification de textures

Nous présentons ici les résultats des différents opérateurs utilisés pour quantifier la texture des différentes régions dans une image. Deux parmi eux ont été utilisés dans notre modèle d'étiquetage sémantique (voir la section 5.3.3 du chapitre 5).

FIGURE I.1 : Color input image from the MSRC-21 Dataset.

Local binary pattern (LBP)

FIGURE I.2 : Result of local binary pattern (LBP), with $r = 2$ and $P = 9$.

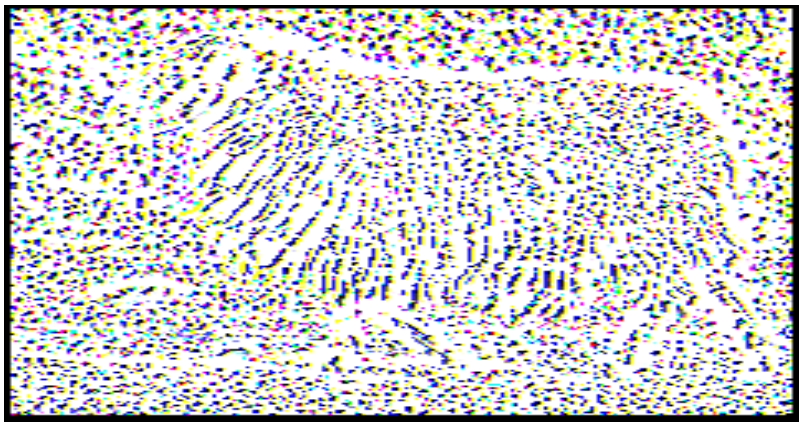


FIGURE I.3 : Result of local binary pattern (LBP), with $r = 2$ and $P = 16$.

Opponent color local binary pattern (OCLBP)

FIGURE I.4 : Result of opponent color local binary pattern (OCLBP), with $r = 1$ and $P = 9$ (red-green, red-blue and green-blue).

FIGURE I.5 : Result of opponent color local binary pattern (OCLBP), with $r = 2$ and $P = 16$ (red-green, red-blue and green-blue).

FIGURE I.6 : Result of opponent color local binary pattern (OCLBP), with $r = 1$ and $P = 9$ (green-red, blue-red and blue-green).

FIGURE I.7 : Result opponent color local binary pattern (OCLBP), with $r = 2$ and $P = 16$ (green-red, blue-red and blue-green).

Laplacian operator (LAP)

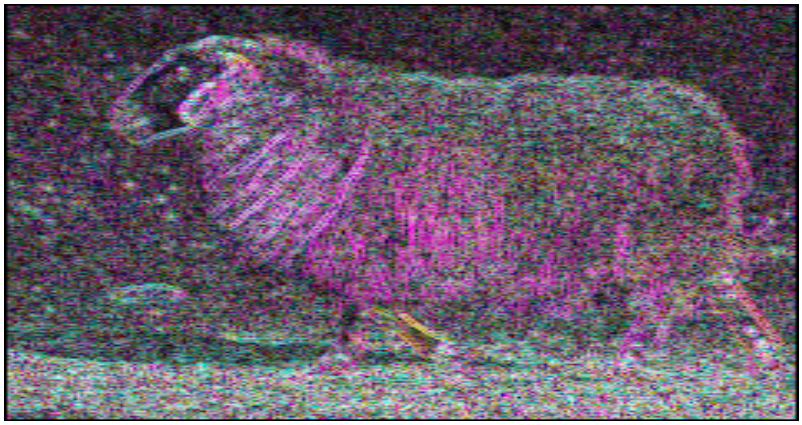


FIGURE I.8 : Result of Laplacian operator (LAP), with $r = 1$ and $P = 9$.

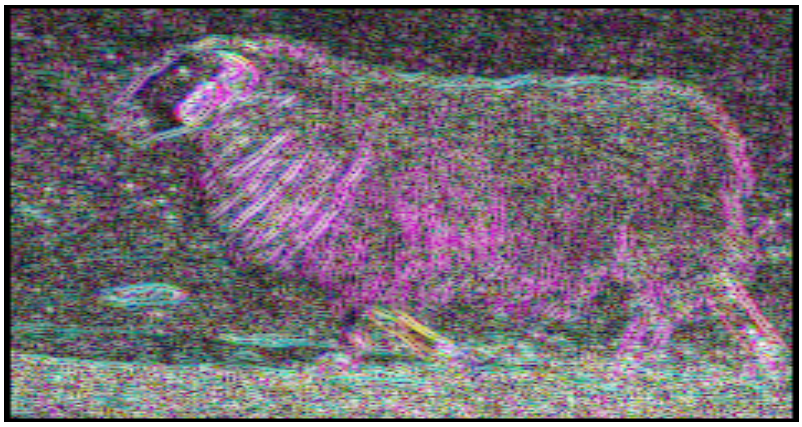


FIGURE I.9 : Result of Laplacian operator (LAP), with $r = 2$ and $P = 16$.

Annexe II

Échéancier de la thèse

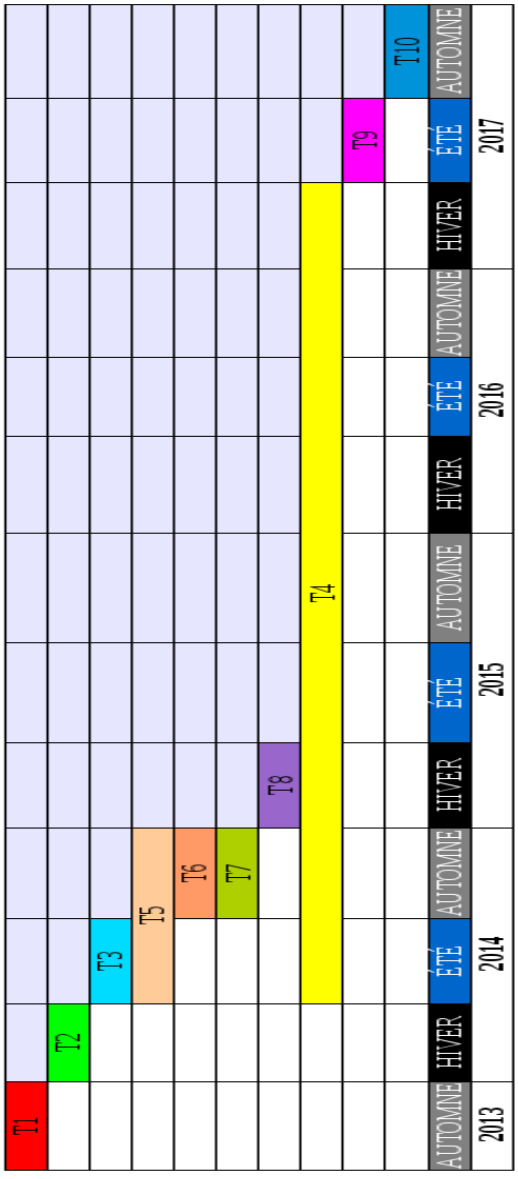


FIGURE II.1 : Échéancier de la thèse

- T1 : Première partie de l'examen générale de synthèse (cours IFT2015 et IFT2125).
- T2 : Deux cours gradués (obligatoire).
- T3 : Définition de la problématique et l'objectif de notre travail.
- T5 : Revue de la littérature.
- T4 : Réalisation du projet de recherche.
- T6 : Deux cours gradué (optionnel).
- T7 : Deuxième partie de l'examen générale de synthèse (examen de spécialité).
- T8 : Troisième partie de l'examen générale de synthèse (présentation du projet de recherche).
- T9 : Rédaction de la thèse.
- T10 : Présentation de la thèse.