HAL
open science

Algorithmes de correspondance et superpixels pour l'analyse et le traitement d'images

Rémi Giraud

▶ **To cite this version:**

Rémi Giraud. Algorithmes de correspondance et superpixels pour l'analyse et le traitement d'images. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Bordeaux, 2017. Français. NNT : . tel-01660492v1

HAL Id: tel-01660492

https://theses.hal.science/tel-01660492v1

Submitted on 10 Dec 2017 (v1), last revised 11 Jan 2018 (v2)

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

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

THÈSE

présentée à

L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

par

RÉMI GIRAUD

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

Algorithmes de correspondance et superpixels pour l'analyse et le traitement d'images

---

Date de soutenance : 29 novembre 2017

Devant la commission d'examen composée de :

Florence	Tupin	Pr.	Télécom ParisTech	Présidente
Pierre	Hellier		Principal Scientist - Technicolor Rennes	Examinateur
Vincent	Lepetit	Pr.	Université de Bordeaux	Examinateur
Patrick	Bouthemy	DR	Inria Rennes - Bretagne Atlantique	Rapporteur
Olivier	Lézoray	Pr.	Université de Caen- Normandie	Rapporteur
Nicolas	Papadakis	CR	CNRS, Université de Bordeaux	Directeur de thèse
Vinh-Thong	Ta	MCF	Bordeaux INP	Directeur de thèse

– 2017 –

Titre

Algorithmes de correspondance et superpixels pour l'analyse et le traitement d'images

Résumé

Cette thèse s'intéresse à diverses composantes du traitement et de l'analyse d'images par méthodes non locales. Ces méthodes sont basées sur la redondance d'information présente dans d'autres images, et utilisent des algorithmes de recherche de correspondance, généralement basés sur l'utilisation patchs, pour extraire et transférer de l'information depuis ces images d'exemples. Ces approches, largement utilisées par la communauté de vision par ordinateur, sont souvent limitées par le temps de calcul de l'algorithme de recherche, appliqué à chaque pixel, et par la nécessité d'effectuer un prétraitement ou un apprentissage pour utiliser de grandes bases de données.

Pour pallier ces limites, nous proposons plusieurs méthodes générales, sans apprentissage, rapides, et qui peuvent être facilement adaptées à diverses applications de traitement et d'analyse d'images naturelles ou médicales. Nous introduisons un algorithme de recherche de correspondances permettant d'extraire rapidement des patchs d'une grande bibliothèque d'images 3D, que nous appliquons à la segmentation d'images médicales. Pour utiliser de façon similaire aux patchs, des présegmentations en superpixels réduisant le nombre d'éléments de l'image, nous présentons une nouvelle structure de voisinage de superpixels. Ce nouveau descripteur permet d'utiliser efficacement les superpixels dans des approches non locales. Nous proposons également une méthode de décomposition régulière et précise en superpixels. Nous montrons comment évaluer cette régularité de façon robuste, et que celle-ci est nécessaire pour obtenir de bonnes performances de recherche de correspondances basées sur les superpixels.

Mots-clés

Algorithmes de correspondance, méthodes non locales basées sur les patchs, superpixels, métriques d'évaluation, régularité, voisinage de superpixels, segmentation, classification, étiquetage, images médicales, transfert de couleurs

Title

Matching algorithms and superpixels for image analysis and processing

Abstract

This thesis focuses on several aspects of image analysis and processing with non local methods. These methods are based on the redundancy of information that occurs in other images, and use matching algorithms, that are usually patch-based, to extract and transfer information from the example data. These approaches are widely used by the computer vision community, and are generally limited by the computational time of the matching algorithm, applied at the pixel scale, and by the necessity to perform preprocessing or learning steps to use large databases.

To address these issues, we propose several general methods, without learning, fast, and that can be easily applied to different image analysis and processing applications on natural and medical images. We introduce a matching algorithm that enables to quickly extract patches from a large library of 3D images, that we apply to medical image segmentation. To use a presegmentation into superpixels that reduces the number of image elements, in a way that is similar to patches, we present a new superpixel neighborhood structure. This novel descriptor enables to efficiently use superpixels in non local approaches. We also introduce an accurate and regular superpixel decomposition method. We show how to evaluate this regularity in a robust manner, and that this property is necessary to obtain good superpixel-based matching performances.

Keywords

Matching algorithms, patch-based non-local methods, superpixels, evaluation metrics, regularity, superpixel neighborhood, segmentation, classification, labeling, medical imaging, color transfer

Remerciements

Tout d'abord, je tiens à remercier Olivier Lézoray et Patrick Bouthemy pour avoir accepté de rapporter ma thèse et pour leurs retours précieux sur ce manuscrit. Merci également à Florence Tupin, Pierre Hellier et Vincent Lepetit pour avoir accepté de faire partie de mon jury et pour les échanges que nous avons eus sur mes travaux.

Durant ma thèse, j'ai été encadré de la meilleure des manières par Nicolas Papadakis et Vinh-Thong Ta. Leur investissement et leur disponibilité ont été exemplaires tout au long de ces trois années, et à leur contact j'ai pu progressivement découvrir les ficelles du métier. Un immense merci pour leur exigence, à tous les niveaux pour n'importe quelle tâche, qui m'a permis de progresser considérablement en me remettant moi, et mes travaux, toujours en question. Enfin, merci pour leur passion pour la recherche, qu'ils ont su me transmettre. J'espère pouvoir continuer à échanger avec eux dans le futur.

Durant cette thèse, j'ai également été amené à travailler avec Aurélie Bugeau et Pierrick Coupé. Merci à tous les deux pour leurs conseils toujours pertinents et leur bienveillance. Merci également à Jose Manjón pour son accueil très chaleureux à Valence. Quelle belle tradition que l'*esmorzaret* de 11h!

Merci à tous les copains du labo, Florent, Joachim, Kilian, Luce, Samah, Adrien, Mary, Mathias, Simon, David, Nicolas, Pierre, Fabien, et tous les autres, y compris à l'IMB et à l'IMS. Café des Moines, Barbecue, Raclette, et Café Michel 4 ever.

Merci à l'association AFoDIB, et aux gens qui l'ont entretenue depuis sa création. Les activités (laser quest, journée sport, pauses-café, karaoké, journée d'intég... d'accueil!) ont sans nul doute égayé la vie du laboratoire et de ses doctorants.

Deux éléments ont aussi participé à cette bonne ambiance et méritent que je leur rende hommage : le baby-foot et la table de ping-pong. Une pratique intensive pendant 3 ans nous a permis d'atteindre un niveau considérable et nous aurait même permis d'envisager des compétitions à l'échelle internationale, c'est certain. Une pensée toute émue pour ce tournoi de ping-pong resté inachevé. Pourvu qu'il demeure ainsi pour l'éternité. Une pensée également pour nos nombreux adversaires malheureux et notamment INRIA qui nous a permis à plusieurs reprises de représenter fièrement le LaBRI lors de sorties dans les bars munis d'un baby-foot.

Pour m'avoir soutenu durant ces trois ans, parfois difficiles du fait de mon investissement, il est important pour moi de remercier Justine, Jeanne et Adèle. Les vieux copains du lycée, Chevet, remi, Nathan, Keyron, Fransoa, ont également contribué indirectement, en me donnant le moral. Une pensée pour tous les copains, plus ou moins perdus de vue après la prépa, puis l'école.

Enfin, merci du fond du cœur à ma famille qui a toujours veillé à ce que je ne manque de rien, et pour leur soutien indéfectible depuis le début de mes études.

“Ces jeux de correspondances laissent libre cours à l’émotion. La beauté des images, qui se répondent, est source de surprise, de rêve et d’émerveillement. De découvertes. De vertige. Alors, la science et l’art se confondent.”

Quand l’art rencontre la science
Jean-Claude Ameisen

Sommaire

Remerciements **v**

Introduction générale **2**

I Algorithme de correspondance basé sur les patchs pour la segmentation d'images médicales **15**

1 Segmentation d'images médicales et algorithmes de correspondance **19**
1.1 Contexte et pathologies neurodégénératives 20
1.2 Segmentation d'images IRM 3D 21
1.3 Algorithmes de correspondance à partir de patchs 26
1.4 Conclusion 32

2 La méthode OPAL **35**
2.1 Introduction 36
2.2 Algorithme PatchMatch optimisé 36
2.3 Fusion d'étiquettes basées sur les patchs et segmentation 42
2.4 Agrégation d'estimateurs multi-échelles et multi-descripteurs 44
2.5 Conclusion 46

3 Résultats de segmentation d'images médicales **49**
3.1 Introduction 50
3.2 Validation d'OPAL sur la segmentation de l'hippocampe 50
3.3 Segmentation du cervelet 62
3.4 Extension à la classification de la maladie d'Alzheimer 65
3.5 Conclusion 70

II Décomposition en superpixels réguliers **73**

4 Algorithmes de superpixels et étude des métriques standards **77**
4.1 Utilisation des superpixels 78
4.2 Méthodes de décomposition en superpixels 81
4.3 Étude des métriques de superpixels 86
4.4 Conclusion 94

5 La méthode SCALP **97**
5.1 Contributions 98
5.2 Décomposition en superpixels avec SCALP 98
5.3 Contrainte depuis une segmentation initiale 106
5.4 Validation de SCALP sur les métriques superpixels standards 109
5.5 Conclusion 119

6 Évaluation de la régularité **123**
6.1 Introduction 124
6.2 Mesure de régularité globale 126
6.3 Comparaison de SCALP à l'état-de-l'art sur GR 136
6.4 Impact de la régularité sur des applications basées sur les superpixels 138
6.5 Conclusion 147

III Algorithme de correspondance basé sur les superpixels et applications **149**

7 La méthode SuperPatchMatch **153**
7.1 Introduction 154
7.2 Définition du superpatch 156
7.3 Comparaison de superpatchs 156
7.4 SuperPatchMatch 161
7.5 Conclusion 164

8 Application à la segmentation et l'étiquetage basés sur les superpixels **167**
8.1 Introduction 168
8.2 Schéma de fusion d'étiquettes 168
8.3 Validation de SuperPatchMatch sur l'étiquetage de visages 169
8.4 Conclusion 177

9 Application au transfert de couleurs **179**
9.1 Introduction 180
9.2 SuperPatchMatch contraint 182
9.3 Fusion des couleurs 184
9.4 Résultats de transfert de couleurs 186
9.5 Conclusion 188

Conclusion générale et perspectives **194**
1 Conclusion générale 194
2 Résumé des contributions 196
3 Perspectives 197

Liste des publications **206**

Bibliographie **208**

Liste des acronymes **230**

Table des matières **233**

Introduction générale

Les avancées technologiques et la popularisation des dispositifs d'acquisition d'images ont mené ces dernières années à une explosion de la quantité de données d'imagerie numérique. À l'heure de ces données massives, un des enjeux du domaine de la vision par ordinateur consiste à utiliser les nouvelles puissances de calcul offertes par la technologie pour extraire de l'information des très nombreuses images à disposition. Ces images peuvent être de natures très diverses, provenant d'appareils photos, de caméras, de télescopes, de radars ou encore de scanners médicaux. De nombreux domaines sont donc concernés par l'utilisation et le traitement de ces images, dont l'informatique graphique, la robotique, l'astronomie, la défense, la bioinformatique ou encore la recherche médicale. Dans cette thèse, nous proposons plusieurs méthodes générales pour exploiter efficacement et rapidement la redondance d'information contenue dans ces grandes bases de données d'imagerie. Les outils proposés peuvent être utilisés pour diverses applications et sur différents types d'images, naturelles et médicales. Dans la suite, nous introduisons le contexte général de cette thèse, que nous illustrons notamment avec des exemples liés au domaine de l'analyse d'images médicales, étudié en détails dans une première partie, puis nous présenterons nos motivations et nos contributions.

Dans tous les domaines d'imagerie, deux grandes familles d'application se distinguent : le traitement et l'analyse d'images. Le traitement d'images consiste à directement modifier le contenu des données, à traiter l'image pour corriger, améliorer ou transformer l'intensité de ses pixels. Par exemple, dans le domaine médical, des algorithmes de débruitage sont souvent utilisés pour améliorer la qualité de l'image en sortie de l'appareil d'acquisition. Pour d'autres applications qualitatives, des méthodes de réhaussement de contraste, permettent de mieux distinguer les éléments de l'image. Pour des applications artistiques on peut vouloir coloriser une image initialement en niveaux de gris, ou transférer un style de peinture ou les couleurs d'une image à une autre. L'uniformisation des couleurs entre plusieurs images peut par exemple servir à la reconstruction 3D d'objets ou de monuments depuis plusieurs prises de vues par des capteurs différents. Enfin, beaucoup d'images naturelles souffrent également de perturbation à l'acquisition, comme la présence d'objets occultant une partie de l'image. Les techniques d'*inpainting* permettent alors de corriger en partie ce problème. Des logiciels d'édition populaires incluent plusieurs de ces techniques, pour permettre à n'importe quel utilisateur d'utiliser de manière simplifiée, des algorithmes complexes de traitement d'images.

L'analyse d'images consiste à décomposer et reconnaître les éléments présents dans une image. Cette analyse peut donner une simple caractéristique à toute l'image, par exemple dans le cas de la classification de données où l'on cherche à identifier le type d'objet présent. Celle-ci peut également aller jusqu'à l'identification des différents objets présents dans une image, voire la reconnaissance du contexte et de l'action en cours. Au quotidien, l'analyse d'images permet la reconnaissance de caractères, de codes barres ou encore de signalisation routière

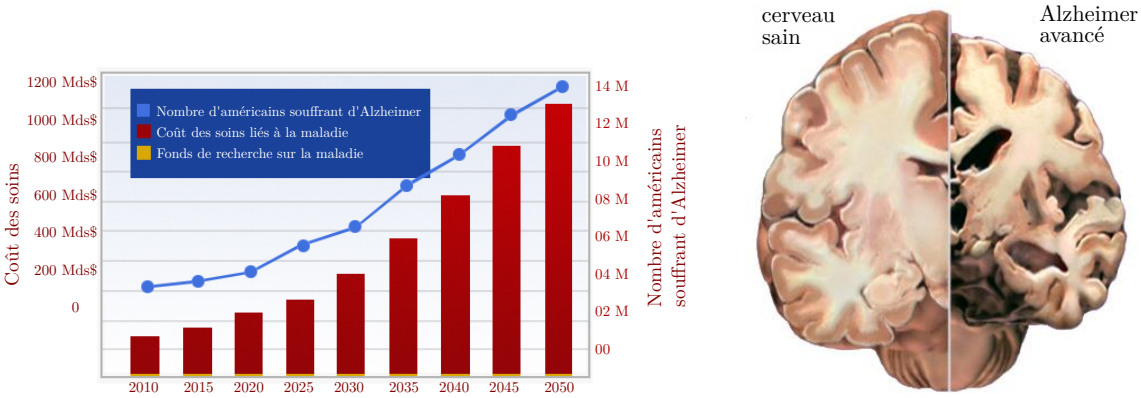
depuis son smartphone. Cette reconnaissance est généralement associée à une étape préliminaire de segmentation de ces objets. L'image à analyser est décomposée en régions, auxquelles sont affectées des caractéristiques indiquant le type d'objets contenu. L'analyse d'images fournit donc de l'information sur la donnée traitée, et celle-ci peut être localisée sur certaines parties de l'image.

Comprendre automatiquement le contenu d'une image est un des enjeux majeurs actuels. Sur des images naturelles, le cerveau humain est très souvent capable d'identifier les différents objets présents, ainsi que leurs contours. Sur des images plus complexes, médicales par exemple, la définition des structures est accessible uniquement aux experts, et leurs contours sont même parfois décorrelés des intensités de l'image. Sur ce type de données, des images d'exemples sont alors nécessaires pour déterminer comment analyser une autre image. De manière générale, pour permettre à un ordinateur de reconnaître un objet ou une action, celui-ci a besoin d'un a priori, d'une connaissance de l'analyse ou du traitement à effectuer. L'utilisation de grandes bases de données d'imagerie déjà traitées, offre alors cette connaissance.

De nombreux domaines bénéficient de l'utilisation de données d'exemple, comme celui du document, où la restauration et la numérisation de vieux manuscrits utilisent la reconnaissance préalable de documents d'exemple. L'utilisation de nombreuses données d'exemple peut également permettre d'extraire des caractéristiques invisibles à l'oeil, par exemple des marqueurs spécifiques de pathologies médicales, ou de reconnaissance faciale et digitale. En effet, dans le domaine de la sécurité, on cherche à reconnaître automatiquement des individus, et cette reconnaissance s'effectue sur des marqueurs du visage caractéristiques, déterminés a priori par l'algorithme.

Le domaine de la santé est particulièrement concerné par l'utilisation de techniques d'imagerie, avec par exemple l'imagerie par résonance magnétique (IRM). Les traitements et analyses informatiques de ces images offrent de nombreuses perspectives. Par exemple, les maladies neurodégénératives, que l'on peine encore à diagnostiquer ou à prévenir, apparaissent avec l'âge, et sont de plus en plus étudiées via l'imagerie cérébrale. Ces pathologies entraînent la destruction des structures du cerveau (Figure 1(b)), conduisant à la perte de capacités essentielles au fonctionnement et à l'autonomie du corps humain, comme la motricité, le langage ou la mémoire. La Figure 1(a), présente l'évolution de la maladie d'Alzheimer en termes de patients atteints et de répercussions de coût pour le système de santé américain. La prédiction et le traitement de ces maladies vont donc représenter un défi sociétal de taille dans les années à venir. Dans ce contexte, les images IRM permettent d'étudier l'évolution des structures cérébrales marqueurs de la maladie, et une analyse précise de l'état d'un patient permet alors d'adapter la médication en conséquence et faire gagner en qualité et espérance de vie.

L'utilisation de grandes bases d'images provenant de sujets différents, et déjà analysées par des experts, est alors cruciale pour extraire automatiquement de l'information. Celle-ci permet de mettre en évidence de nouveaux marqueurs, de déduire l'état d'un patient ou de segmenter ces structures d'intérêt (Coupé et al., 2012b). La segmentation est en effet la première étape permettant d'analyser la forme et la taille des structures impactées, et de faciliter leur suivi au fil des années. Cependant, une segmentation manuelle par un expert est une tâche très fastidieuse, coûteuse en temps, et les résultats obtenus par différents experts sont sujets à une forte variabilité. De ce fait, il est très important de proposer à la communauté médicale, mais également à d'autres domaines de recherche, des méthodes rapides d'analyse qui soient précises et fiables, afin de faciliter l'étude de ces images.



(a) (b)

FIGURE 1 – Évolution de la maladie d'Alzheimer en termes de populations impactées et de coût pour le système de santé américain (a), et de dégénérescence du cerveau (b). *Source : A national Alzheimer's strategic plan: The report of the Alzheimer's study group (2009).*

Pour utiliser l'information contenue dans des images d'exemples, fournies avec leur résultat, souvent déterminé manuellement, certaines méthodes utilisent un apprentissage a priori de ces données pour apprendre à l'algorithme le comportement à avoir pour une nouvelle image d'entrée. Différentes approches d'apprentissage existent et sont applicables à l'analyse ou au traitement d'images, comme les arbres de décision, l'utilisation de *support vector machine* (SVM), ou l'apprentissage de dictionnaires. Par exemple, dans le cas de l'analyse d'images médicales, l'apprentissage de dictionnaires consiste à déterminer les éléments discriminants dans une grande base d'images d'exemples, pour en extraire un sous-ensemble pertinent. Les éléments d'une image en entrée sont alors comparés à ceux du dictionnaire, pour en déduire automatiquement une analyse.

Les récentes avancées en termes de puissance de calcul ont offert de nouvelles perspectives dans le domaine de l'apprentissage profond, avec notamment les réseaux de neurones. Le principe initial du perceptron est alors étendu à plusieurs couches qui

contiennent une grande quantité de neurones, connectés les uns aux autres. Dans le cas de l'analyse d'images, chaque étape du réseau consiste à filtrer l'image en entrée pour en extraire des caractéristiques, réduisant progressivement sa dimension à chaque couche pour aboutir à une classification finale (Krizhevsky et al., 2012). L'apprentissage de ces modèles consiste à nourrir le réseau avec une très grande quantité de données d'exemple, afin de paramétrer les différentes connexions entre les couches.

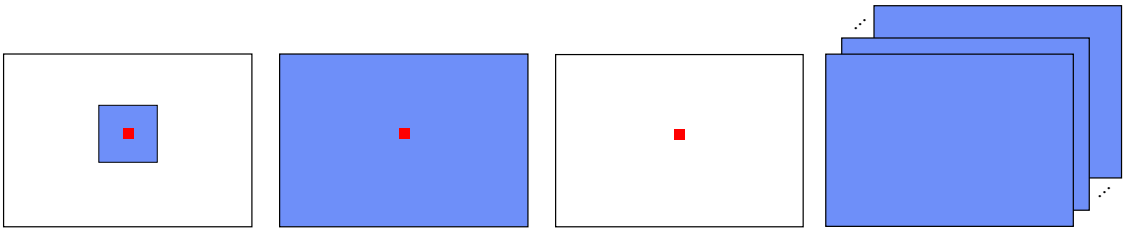
Bien que ces méthodes permettent d'obtenir de très bonnes performances, notamment sur des applications de classification d'images, ces approches par apprentissage souffrent de plusieurs inconvénients. Tout d'abord l'entraînement peut être très fastidieux, et nécessiter plusieurs jours sur des architectures matérielles très performantes (GPU). Les applications sont dépendantes des données à disposition, qui doivent être très nombreuses pour paramétrer efficacement le réseau. À titre d'exemple, la base de données CIFAR-10 (Krizhevsky et Hinton, 2009) contient 60000 imagettes de taille 32×32 pixels à fournir au réseau pour l'apprentissage.

Outre le temps d'entraînement conséquent et des difficultés d'optimisation qui peuvent en résulter, le fait d'apprendre une analyse particulière sur un ensemble spécifique de données peut limiter l'adaptabilité directe des réseaux à d'autres applications. Enfin, les connexions entre neurones sont complexes et ne permettent pas nécessairement de visualiser l'information extraite, de sorte que certaines architectures offrent peu de maîtrise sur le résultat en sortie. Des travaux tels que (Szegedy et al., 2013) ont d'ailleurs montré que l'ajout d'un bruit quasiment invisible modifiait significativement la décision de classification de l'algorithme, menant à des résultats aberrants.

Méthodes non locales à partir d'exemples

Dans cette thèse, nous nous plaçons dans le contexte des méthodes dites non locales à partir d'exemples, qui extraient directement de l'information en cherchant des éléments similaires dans une ou plusieurs images d'exemples. Le terme "non local" désigne l'espace de recherche de contenu similaire, quand le terme "à partir d'exemples" indique l'utilisation d'images d'exemples. Ces méthodes ne nécessitent aucune étape d'apprentissage, ce qui facilite leur adaptation à de nombreuses applications, et l'optimisation des différents paramètres. Nous montrons notamment dans cette thèse que des résultats équivalents ou supérieures à ceux de plusieurs approches basées sur des apprentissages peuvent être obtenus.

Des applications de traitement d'images comme le débruitage ou l'inpainting peuvent se limiter à considérer le contenu de l'image traitée (Buades et al., 2005). On peut alors chercher à annuler la contribution du bruit, par exemple en considérant l'information locale des pixels voisins (Figure 2(a)), ou utiliser des approches dites non locales pour extraire et transférer de l'information depuis des régions plus éloignées (Figure 2(b)). Comme mentionné précédemment, pour de nombreuses applications de traitement et d'analyse d'images, comme la reconnaissance et la classi-



(a) Local (b) Non local (c) Non local basé exemple(s)

FIGURE 2 – Illustration des différentes domaines de recherche de contenu similaire pour un pixel donné (rouge). La recherche peut être limitée à un voisinage (a), étendu à toute l'image (b), ou être effectuée dans des images d'exemples (c).

fication, l'utilisation d'images d'exemples est nécessaire pour introduire un a priori (Figure 2(c)). Les méthodes non locales basées exemple considèrent donc directement l'ensemble des images d'exemples à disposition pour extraire et transférer de l'information en fonction de l'image à traiter.

Même sur des images standardisées ou recalées, par exemple de type médical, la variabilité, même légère entre les images, peut rendre sous-optimales des approches de transfert global d'a priori, par exemple de segmentation. Une approche plus efficace consiste à extraire et à transférer des informations différentes selon les zones de l'image à traiter, considérées indépendemment. Les méthodes non locales standards définissent chaque pixel de l'image par un descripteur, généralement un patch carré, capturant les pixels dans son voisinage, et transfèrent l'information de descripteurs similaires extraits dans les images d'exemples.

L'utilisation des patchs a notamment été popularisée avec la synthèse de texture (Efros et Leung, 1999), puis avec le débruitage (Buades et al., 2005). Ceux-ci permettent de caractériser de façon robuste le contexte local d'un pixel. Deux patchs possèdent le même nombre d'éléments, et leur structure régulière permet d'optimiser numériquement le calcul de distance. Pour exploiter efficacement cette redondance de motifs similaires, des patchs voisins doivent être trouvés dans l'espace de recherche. Lorsque cette recherche de correspondances s'effectue dans une ou plusieurs images, la recherche exhaustive des plus proches patchs s'avère extrêmement coûteuse en temps de calcul. D'importants temps de calcul sont problématiques en pratique pour le traitement de nombreuses données, et peuvent limiter l'optimisation empirique des paramètres et l'investigation de nouvelles approches. Ainsi, à moins qu'une forte hypothèse de recalage ne permette de limiter la zone de recherche, la plupart des méthodes non locales utilisent des algorithmes de recherche approchés, par exemple Barnes et al. (2009); Korman et Avidan (2016). Ces méthodes ne fournissent que des plus proches voisins approchés ANN (pour Approximate Nearest Neighbor), mais qui suffisent à obtenir des performances proches de celles obtenues avec une recherche exhaustive.

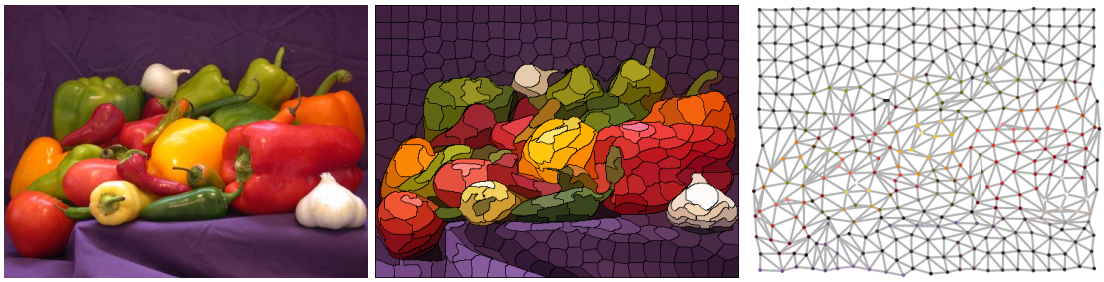
Ces algorithmes ont encore peu été appliqués à des données 3D, et sur de grandes bibliothèque d'images. Les approches ANN dans des données de grande dimension sont généralement hautement paramétrés et basées sur des apprentissages coûteux, par exemple Barnes et al. (2015), pour notamment limiter le temps de traitement. En effet, le traitement de tous les patchs de l'image et la recherche dans de grandes bibliothèques peuvent être coûteux, un patch étant considéré pour chaque pixel. Des pixels proches en termes spatial et colorimétrique dans l'image sont pourtant susceptibles de contenir la même information. De plus en plus d'approches utilisent alors des méthodes de présegmentation telles que les superpixels pour grouper a priori ces pixels homogènes, et limiter le nombre d'éléments considérés.

Superpixels

Sur certains types d'image, le résultat de traitement peut être amélioré par une étape de présegmentation, groupant les pixels adjacents susceptibles de recevoir la même information. Cette étape peut en effet réduire le bruit résultant d'un traitement indépendant de chaque pixel. De plus, considérer un nombre réduit d'éléments peut diminuer fortement le temps de calcul, qui peut être conséquent avec des descripteurs à partir de patchs, définis pour chaque pixel. Depuis leur introduction avec notamment Ren et Malik (2003), l'utilisation des superpixels est devenue très populaire, avec des méthodes de décomposition comme celle d'Achanta et al. (2012).

Les superpixels groupent les pixels de l'image en zones de couleurs homogènes, tout en essayant de respecter les contours des objets de l'image, réduisant ainsi considérablement le nombre d'éléments à traiter. On donne en Figure 3 un exemple de décomposition d'une image en superpixels. Cette étape de groupement a priori des pixels en zones plus larges, préserve l'information géométrique, perdue avec les approches multirésolution régulières. Ainsi, un résultat obtenu à l'échelle des superpixels peut être très proche de la vérité terrain à l'échelle pixellique, puisque les contours des superpixels tentent de suivre les contours des objets.

Les superpixels sont donc de plus en plus utilisés dans les procédés de traitement et d'analyse d'images classiques, par exemple pour la localisation d'objets (Fulkerson et al., 2009), le traitement de la stéréovision et des occultations (Zhang et al., 2011b; Pérez-Rúa et al., 2016), la segmentation d'objets multi-classes (Gould et al., 2014), ou sont adaptés aux architectures de réseaux de neurones (Gadde et al., 2016). Néanmoins, dans le contexte des méthodes ANN, et de la recherche de correspondances dans de grandes bibliothèques d'images, les superpixels ont été peu étudiés. L'irrégularité de forme et de taille des décompositions empêche l'utilisation directe des outils basés sur l'utilisation de patchs intégrant l'information de voisinage, nécessaire à la recherche de bonnes correspondances.



(a) (b) (c)

FIGURE 3 – Exemple de décomposition d'une image (a) en superpixels contenant les couleurs moyennes (b) avec la méthode SCALP présentée dans le chapitre 5. Le graphe d'adjacence, connectant les barycentres des superpixels est représenté en (c).

Motivations

Dans cette thèse, nous nous plaçons dans le contexte des approches non locales, en proposant des contributions visant à exploiter rapidement et efficacement la redondance d'informations contenues dans une ou plusieurs images d'exemples, sans étape d'apprentissage. Nous nous intéressons aux différents maillons de la chaîne de traitement, aussi bien au niveau des descripteurs, que des méthodes de recherche de correspondances. Nous montrons que les approches proposées sont compétitives en termes de performances avec des approches utilisant un apprentissage, et présentent de nombreux avantages en termes de temps de calcul et d'applicabilité.

Un des avantages des méthodes non locales est la capacité à extraire directement de l'information depuis une zone similaire dans la même ou dans une autre image. De façon similaire aux méthodes d'apprentissage, le nombre d'images d'exemples est généralement lié aux performances. La diversité des images d'exemples permet potentiellement d'accéder à de meilleures zones de correspondance et donc de transférer une information plus pertinente. Néanmoins, considérer un grand nombre d'images d'exemples peut augmenter le temps de calcul du procédé de recherche. De nombreuses méthodes utilisent alors une étape de présélection des images d'exemples pouvant diminuer les performances, ou présentent des temps de calcul prohibitifs, ne permettant pas d'optimiser le modèle.

Les méthodes de recherche de correspondances non locales que nous proposons dans cette thèse répondent aux propriétés suivantes :

— pas d'apprentissage nécessaire,
— capacité de gérer de grandes bibliothèques d'images d'exemples,
— faible temps de calcul.

Les méthodes proposées sont adaptées à plusieurs applications, et obtiennent des performances au niveau de l'état-de-l'art. Celles-ci se veulent générales et facilement adaptables à plusieurs contextes, afin de faciliter leur utilisation par la communauté.

La contrainte de complexité et de temps de calcul peut être clairement exprimée notamment lorsque l'on effectue la recherche d'ANN dans de grandes images, des vidéos ou des bibliothèques d'images 3D, pour par exemple la segmentation d'images médicales. Pour cette application, les approches empiriques utilisent des modèles coûteux de déformation pour obtenir une segmentation automatique depuis les images d'exemples. Des méthodes non locales ont alors été introduites, et présélectionnent quant à elles les images d'exemples pour considérer, en un temps de calcul modéré, de façon exhaustive l'information de tous les patchs de la bibliothèque. De manière générale, les algorithmes de recherche d'ANN ont peu été étudiés en 3D et sur des bibliothèques d'images.

Un premier travail dans cette thèse a donc consisté à adapter au domaine médical une méthode rapide de recherche d'ANN (Barnes et al., 2009), initialement proposée pour l'édition et notamment l'inpainting d'images naturelles 2D. La méthode proposée permet d'extraire efficacement de l'information d'une grande bibliothèque d'images 3D, sans présélection et en des temps de calcul largement inférieurs à ceux des méthodes de référence. Les informations extraites permettent alors la segmentation, la classification et potentiellement le recalage d'images médicales.

Pour ce premier travail, des patchs réguliers de pixels sont utilisés. Les avantages de ces descripteurs pour la recherche de motifs similaires ont été mentionnés précédemment. Les patchs sont adaptés à la caractérisation de petits voisinages sur des images médicales, dont les segmentations de vérité terrain ne suivent pas nécessairement les intensités de l'image. Néanmoins, pour d'autres types de données et images naturelles, dont les contours des objets sont plus corrélés au contenu de l'image, une présegmentation, notamment en superpixels, peut permettre d'améliorer la précision et le temps de calcul. Cependant, les superpixels segmentent l'image en régions irrégulières, et deux images n'ont potentiellement pas la même décomposition. Par conséquent, tous les outils basés sur les patchs réguliers ne s'appliquent pas directement à ces structures, et l'utilisation des superpixels dans les méthodes standards de traitement et d'analyse d'image est aujourd'hui limitée.

Un des objectifs de cette thèse est de proposer des outils permettant d'exploiter les superpixels de façon similaire à celle des patchs réguliers. Les méthodes proposées visent alors à combiner les avantages des méthodes basées sur les patchs incluant l'information de voisinage, nécessaire à la recherche de bonnes correspondances, et les superpixels permettent d'améliorer la précision et de diminuer la complexité de calcul. Celles-ci se veulent plus efficaces que des approches multirésolution qui réduisent la dimension de l'image en considérant des blocs réguliers. Si les outils basés sur les patchs peuvent être appliqués directement à ces approches, les blocs générés contiennent des pixels de couleurs hétérogènes qui lissent l'information contenue dans les descripteurs, par exemple si l'on considère la couleur moyenne. Les superpixels en revanche ont des formes irrégulières capturant les objets, et contiennent ainsi des pixels homogènes qui fournissent des descripteurs pertinents, pouvant donner accès à de meilleures correspondances. L'irrégularité (ou la régula-

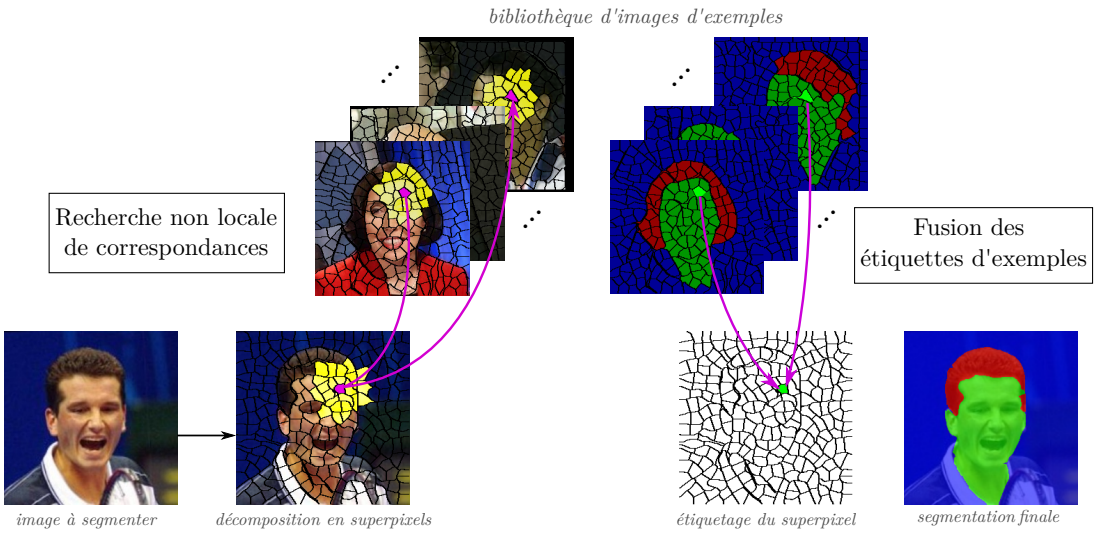


FIGURE 4 – Schéma de la méthode proposée au chapitre 7, illustrant plusieurs travaux menés dans cette thèse : décomposition en superpixels, méthode de recherche d'ANN dans une bibliothèque d'images, et fusion d'informations d'exemples pour la segmentation.

rité) a un impact direct sur les performances de recherche de correspondances, mais celle-ci est mal définie et mesurée dans la littérature. De plus, cette irrégularité de forme des superpixels est la principale limite d'extension à ces descripteurs, des outils basés sur l'utilisation du voisinage.

Pour se rapprocher de l'utilisation des patchs réguliers de pixels, nous proposons une nouvelle structure du voisinage contenant l'information de position des superpixels. Une distance adaptée pour permettre la comparaison de deux de ces structures doit également être définie. Ces outils permettent d'adapter l'algorithme de Barnes et al. (2009) pour proposer une méthode robuste et rapide de recherche d'ANN basés sur les superpixels, ayant les propriétés énoncées précédemment.

Nous illustrons en Figure 4 notre méthode de recherche de correspondance basée sur les superpixels dans une bibliothèque d'images d'exemples, que nous appliquons à la segmentation et l'étiquetage automatique à partir d'exemple. Une image est décomposée en superpixels et plusieurs ANN sont sélectionnés parmi de nombreuses images, en utilisant la structure de voisinage des superpixels. Les informations de segmentation correspondantes sont alors fusionnées et transférées à l'image à traiter pour produire une segmentation automatique de l'image en plusieurs classes. L'utilisation des superpixels permet de réduire le nombre d'éléments, et donc le temps de calcul, tout en obtenant une segmentation proche du contenu de l'image. Au travers de différentes applications, nous démontrons l'intérêt de la structure de voisinage de superpixels par rapport aux patchs, et aux simples superpixels pour calculer de bonnes correspondances.

Contributions

Dans cette thèse, nous considérons des domaines d'applications divers, et les performances et la facilité d'adaptation des outils proposés sont démontrées. Les mêmes principes de recherche d'ANN et de fusion d'informations d'exemples permettent à la fois l'analyse et le traitement d'images avec :

— la segmentation de structures anatomiques 3D (Ta et al., 2014; Giraud et al., 2015, 2016a; Romero et al., 2017),
— la détection de la maladie d'Alzheimer (Hett et al., 2016),
— la segmentation et l'étiquetage basés superpixels (Giraud et al., 2017a),
— le transfert de couleurs (Giraud et al., 2017f,g).

Les principales contributions de notre travail sont les suivantes :

— Nous étendons l'algorithme de correspondance de Barnes et al. (2009) à la recherche basée patchs de multiples ANN dans une bibliothèque d'images, et proposons un schéma de fusion d'étiquettes produisant une segmentation automatique depuis des images d'exemples. La méthode est appliquée avec succès à la segmentation de structures anatomiques sur des images 3D de type IRM (Ta et al., 2014; Giraud et al., 2016a; Romero et al., 2017; Giraud et al., 2015), avec un gain substantiel en termes de temps de calcul sur les méthodes de l'état-de-l'art. Nous montrons ensuite que l'information pathologique des modèles d'exemple peut également être transférée pour effectuer une classification de la maladie d'Alzheimer (Hett et al., 2016).

— Pour permettre l'utilisation du voisinage des superpixels dans un contexte non local, nous proposons une nouvelle méthode de décomposition en superpixels assurant une forte régularité, mais également une très bonne précision de segmentation (Giraud et al., 2016b, 2017b,e). La méthode proposée utilise l'information de couleur et de contour sur le chemin linéaire entre chaque pixel et le barycentre du superpixel concerné, pour assurer le groupement de pixels homogènes, en respectant les contours des objets de l'image. Notre méthode obtient des performances supérieures à celles des méthodes de l'état-de-l'art et s'étend naturellement à la décomposition d'objets 3D.

— Dans le contexte des superpixels, beaucoup de métriques ont été introduites pour évaluer le respect des objets de l'image, l'homogénéité de couleur ou la régularité. Nous étudions le processus global de validation des performances des méthodes de décomposition en analysant les limites des métriques de référence. Une nouvelle mesure de régularité est notamment introduite afin de pallier la faible robustesse de celles de l'état-de-l'art. Nous démontrons que la mesure proposée est hautement corrélée aux performances de recherche d'ANN, et à celles d'autres applications comme le suivi d'objets ou la détection de contours (Giraud et al., 2017d,c).

— Afin d'utiliser les superpixels de façon similaire à celle des patchs réguliers, nous proposons une structure de voisinage qui intègre l'information spatiale des superpixels adjacents. Un schéma de comparaison est défini afin de pouvoir mesurer une distance entre deux de ces structures, possédant potentiellement un nombre et des formes d'éléments différents (Giraud et al., 2017a).

— Nous proposons une adaptation de l'algorithme de Barnes et al. (2009), cette fois à la recherche d'ANN basés superpixels, et décrits par la structure de voisinage introduite (Giraud et al., 2017a). La méthode proposée fournit rapidement des correspondances robustes de superpixels dans une bibliothèque d'images, et nous obtenons des résultats état-de-l'art pour l'étiquetage de superpixels. Le gain en performances par rapport à de simples correspondances basées superpixels ou patchs est alors clairement démontré, et les résultats obtenus sont mêmes supérieurs à ceux de méthodes basées sur un apprentissage ou des réseaux de neurones.

— Enfin, nous montrons que la méthode de recherche d'ANN proposée est particulièrement intéressante du fait de sa modularité, puisqu'elle permet facilement d'intégrer des contraintes au processus de recherche. Nous proposons alors de l'adapter pour introduire une nouvelle méthode rapide de transfert de couleurs basée superpixels, en forçant la sélection uniforme d'ANN dans le domaine de l'image d'exemple (Giraud et al., 2017f,g).

Contenu du document

Cette thèse est décomposée en trois parties, chacune composée de trois chapitres qui sont résumés dans les paragraphes suivants.

Partie I : Algorithme de correspondance basé sur les patchs pour la segmentation d'images médicales

Dans cette partie, nous nous intéressons à la conception d'un algorithme rapide de recherche de correspondances basé sur les patchs au sein d'une grande bibliothèque d'images, pour la segmentation d'images médicales. Les différentes approches et limites des méthodes de segmentation existantes sont présentées et justifient l'utilisation d'un algorithme rapide de recherche de plus proches voisins approchés (ANN) pour obtenir de meilleures performances.

Compte tenu des contraintes d'utilisation d'une bibliothèque d'images, nous présentons une méthode de segmentation basée sur l'algorithme PatchMatch (Barnes et al., 2009), présentant une complexité calculatoire intéressante. Grâce au faible temps de calcul de la méthode proposée, nous étudions de nouvelles stratégies de segmentation multi-échelles et multi-descripteurs.

La méthode est ensuite appliquée à la segmentation de plusieurs structures cérébrales, sur plusieurs bases d'images médicales 3D de type IRM. Les performances sont comparées aux méthodes de l'état-de-l'art et démontrent un gain substantiel en termes de précision et de temps de calcul. Enfin, une extension de la méthode à la classification de la maladie d'Alzheimer est présentée.

Partie II : Décomposition en superpixels réguliers

Dans cette partie, nous nous intéressons au processus de décomposition d'une image en superpixels. Les différentes méthodes de décomposition sont présentées, avec une étude des nombreuses métriques de validation de référence.

Afin d'utiliser les superpixels efficacement dans un contexte non local, ceux-ci doivent être assez réguliers pour limiter la variabilité des décompositions d'une image à l'autre, et permettre la recherche de correspondances. Un second travail consistera donc à proposer une nouvelle méthode de décomposition, qui contrairement aux approches existantes, améliore simultanément la précision de segmentation et la régularité de forme. La méthode est validée et comparée aux méthodes de décomposition en superpixels de l'état-de-l'art sur les métriques pertinentes.

Nous proposons également de nouveaux outils d'évaluation, notamment pour mesurer de façon pertinente la régularité d'une décomposition. Nous démontrons que cet aspect impacte les performances de nombreuses applications, notamment la recherche de correspondance basée sur les superpixels.

Partie III : Algorithme de correspondance basé sur les superpixels et applications

Dans cette dernière partie, nous proposons des outils permettant d'utiliser les superpixels efficacement dans un contexte non local, de façon similaire aux patchs. Nous définissons le superpatch, une structure de voisinage de superpixels, ainsi qu'un schéma de comparaison permettant de calculer une distance entre deux de ces structures. L'algorithme SuperPatchMatch est alors proposé pour la recherche d'ANN de superpixels, décrits par ces superpatchs.

La recherche rapide et précise de multiples correspondances de superpixels dans une bibliothèque d'images peut être utilisée pour de nombreuses applications, et nous proposons un schéma de fusion d'étiquettes basée sur les superpixels. Notre méthode est alors validée sur une application de segmentation et d'étiquetage de superpixels, et comparée à des approches basées sur des apprentissages.

Nous montrons à nouveau l'intérêt de notre méthode, par sa rapidité et sa capacité à intégrer des contraintes, telles que la répartition des correspondances sélectionnées dans les images d'exemples. Nous montrons que forcer la diversité des correspondances de superpixels s'avère intéressant pour effectuer d'efficaces transferts de couleurs entre images.

Enfin, une conclusion générale présente une synthèse de ces travaux et donne plusieurs perspectives et pistes de recherche basées sur les outils proposés.

Première partie

Algorithme de correspondance basé sur les patchs pour la segmentation d'images médicales

Résumé

L'étude des structures anatomiques du cerveau sur des images IRM est cruciale dans le cadre de l'étude des maladies neurodégénératives, telles qu'Alzheimer. Dans cette partie, nous nous intéressons à la segmentation de ces structures sur des images médicales IRM 3D. Le processus de segmentation manuel étant très fastidieux, et coûteux financièrement, il est nécessaire de proposer des méthodes rapides et précises de segmentation automatique. De nombreuses approches utilisent des recalages non linéaires de modèles pré-segmentés vers le sujet à traiter. Le calcul de ces transformations est coûteux (jusqu'à plusieurs heures) et ces méthodes souffrent de la variabilité anatomique entre le sujet et les modèles utilisés. Des approches basées sur les patchs ont récemment permis de se limiter à des recalages affines peu coûteux, en considérant l'information de segmentation de modèles d'exemple dans un schéma de fusion d'étiquettes. Cependant, la considération exhaustive de tous les patchs de la bibliothèque de modèles limite la précision et l'amélioration du temps de calcul (plusieurs minutes par sujet).

Pour pallier les limites des approches de l'état-de-l'art et notamment celles par fusion d'étiquettes basée sur les patchs, nous proposons une nouvelle méthode de segmentation basée sur les patchs nommée OPAL (pour Optimized PatchMatch for Accurate Label fusion). Un algorithme rapide de recherche de plus proches voisins approchés (ANN) est adapté pour trouver très rapidement de bonnes correspondances de patchs dans une grande bibliothèque de modèles. Le faible temps de calcul d'OPAL (quelques secondes) permet d'utiliser de nouvelles stratégies de recherche multi-échelles et multi-descripteurs qui améliorent les performances de segmentation.

La méthode est appliquée à la segmentation de l'hippocampe, la principale structure d'intérêt dans le cadre de la maladie d'Alzheimer, des sous-structures du cervelet, et enfin, la méthode est étendue à la classification de la maladie d'Alzheimer. Sur ces expériences, OPAL surpasse les méthodes de l'état-de-l'art jusqu'à atteindre la fiabilité inter-experts de segmentation sur les bases de données considérées, avec des résultats obtenus en seulement quelques secondes de traitement par image.

Mots-clés

Algorithme de correspondances basées sur les patchs, Plus proches voisins approchés, PatchMatch, Fusion d'étiquettes, Segmentation, Classification, Images médicales 3D, Maladie d'Alzheimer, Hippocampe, Cervelet

Contenu

— Le chapitre 1 présente le contexte de la segmentation d'images médicales (section 1.1) et les méthodes populaires de segmentation existantes (sections 1.2.1 et 1.2.2). Une réponse aux limites des approches de l'état-de-l'art apparaît être l'utilisation d'un algorithme de recherche de plus proches voisins approchés basés sur les patchs. Les méthodes existantes applicables à notre

contexte de recherche de correspondances dans une bibliothèque d'images 3D sont alors brièvement présentées (section 1.3).

— Le chapitre 2 présente la méthode de segmentation proposée OPAL comprenant un nouvel algorithme de recherche de correspondances basées sur les patchs (section 2.2), une étape de fusion d'étiquettes (section 2.3), et une extension à une architecture multi-échelles et multi-descripteurs (section 2.4).

— Le chapitre 3 applique OPAL à la segmentation de l'hippocampe sur plusieurs bases de données (section 3.2) ainsi qu'à la segmentation des structures du cervelet (section 3.3). Enfin, une extension de la méthode à la classification de la maladie d'Alzheimer est proposée (section 3.4).

Publications associées

Revues internationales avec comité de lecture (2)

[1] *CERES: A New Cerebellum Lobule Segmentation Method*
Jose E. Romero, Pierrick Coupé, Rémi Giraud, Vinh-Thong Ta, Vladimir Fonov, et al.
NeuroImage (NIMG), 2016

[2] *An Optimized PatchMatch for Multi-scale and Multi-feature Label Fusion*
Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis, Jose V. Manjón, Louis Collins, et al.
NeuroImage (NIMG), 2015

Conférences internationales avec comité de lecture et actes (2)

[1] *Patch-based DTI grading: Application to Alzheimer's disease classification*
Kilian Hett, Vinh-Thong Ta, Rémi Giraud, Mary Mondino, Jose V. Manjón, et al.
International Workshop on Patch-based Techniques in Medical Imaging (Patch-MI, MICCAI), 2016

[2] *Optimized PatchMatch for near real time and accurate label fusion*
Vinh-Thong Ta, Rémi Giraud, Louis Collins et Pierrick Coupé
International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2014

Conférences nationales avec comité de lecture et actes (1)

[1] *Optimisation de l'algorithme PatchMatch pour la segmentation de structures anatomiques*
Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis, Louis Collins et Pierrick Coupé
Groupe d'Études du Traitement du Signal et des Images (GRETSI), 2015

Dépôt logiciel (1)

[1] *CERES: A New Cerebellum Lobule Segmentation Method*
Jose V. Manjón, Pierrick Coupé, Jose E. Romero, Vinh-Thong Ta et Rémi Giraud
IDDN.FR.001.470008.000.S.P.2015.000.21000, 2016

Chapitre 1

Segmentation d'images médicales et algorithmes de correspondance

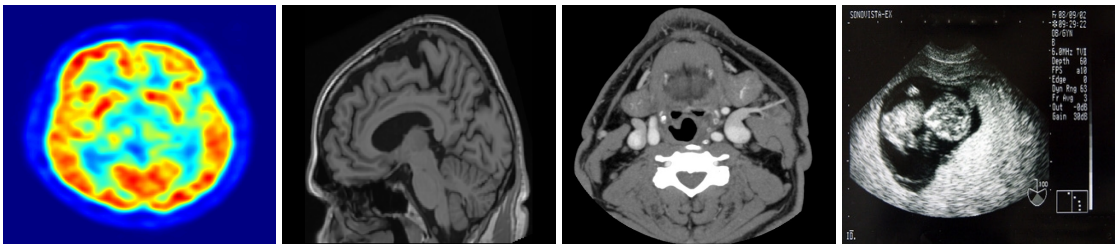
Sommaire

1.1 Contexte et pathologies neurodégénératives 20
1.2 Segmentation d'images IRM 3D 21
1.2.1 Approches par recalage non linéaire 21
1.2.2 Approche par fusion d'étiquettes basée sur les patchs 23
1.3 Algorithmes de correspondance à partir de patchs 26
1.3.1 Introduction des méthodes basées sur les patchs 26
1.3.2 Algorithmes de recherche de correspondances 30
1.4 Conclusion 32

1.1 Contexte et pathologies neurodégénératives

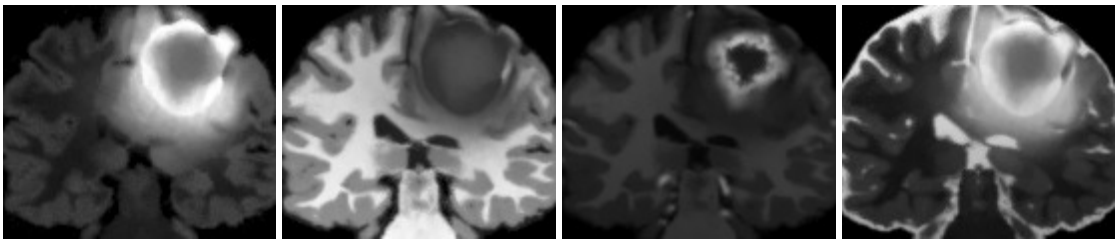
Le traitement et l'analyse d'images médicales sont des domaines largement étudiés, notamment pour la prévention, le suivi, et le diagnostic de pathologies. Différentes méthodes ont été proposées pour permettre d'améliorer la qualité visuelle, par exemple en corrigeant les intensités de l'image en sortie de l'appareil d'acquisition, pour faciliter l'étude d'un sujet en segmentant les structures anatomiques présentes à l'image, ou encore pour effectuer des études à grande échelle sur de grandes bases de sujets.

De nombreux moyens d'acquisition existent, fournissant des images aux propriétés différentes. Certains sont basés sur la radioactivité avec par exemple l'imagerie à tomographie par émission de positons (TEP), sur les champs magnétiques avec l'imagerie par résonance magnétique (IRM), sur les rayons X avec l'imagerie tomodensitométrique (TDM) ou encore sur les ultrasons avec l'échographie. Ces différentes méthodes d'acquisition sont illustrées dans la Figure 1.1. Pour un même type d'acquisition, par exemple l'IRM, plusieurs modes sont utilisables, avec des paramètres de scanner différents. Ces réglages permettent de générer plusieurs modalités, capturant des intensités différentes, qui sont préférées selon les applications. Des exemples de modalités FLAIR, T1, T1C (contraste augmenté) et T2 sont également donnés dans la Figure 1.2.



(a) (b) (c) (d)

FIGURE 1.1 – Exemples d'images médicales de type TEP (a), IRM (b), TDM (c) et d'échographie (d).



(a) (b) (c) (d)

FIGURE 1.2 – Modalités d'IRM d'un même sujet : FLAIR (a), T1 (b), T1C (c) et T2 (d).

Les organes vitaux tels que les poumons, le coeur et principalement le cerveau sont intensivement étudiés par imagerie médicale. Comme mentionné dans l'introduction générale, l'imagerie cérébrale offre en particulier de nombreuses perspectives dans le cadre des maladies neurodégénératives, qui sont à l'heure actuelle des pathologies que l'on peine encore à diagnostiquer ou à prévenir. L'étude du cerveau peut se faire grâce aux IRM et de nombreux travaux dédiés spécifiquement à ces images ont été proposés, par exemple pour le débruitage (Manjón et al., 2008b; Coupé et al., 2008), le recalage d'images (Klein et al., 2009; Avants et al., 2011), la classification de pathologies (Klöppel et al., 2008; Zhang et al., 2011a) ou la segmentation de structures anatomiques (Collins et al., 1995; Heckemann et al., 2006; Lötjönen et al., 2010; Tong et al., 2013; Wang et al., 2014; Pereira et al., 2016). Cette dernière application est une étape cruciale pour l'étude des structures d'intérêt et permet d'analyser leur forme, leur taille et de faciliter leur suivi.

1.2 Segmentation d'images IRM 3D

La segmentation des structures du cerveau sur les images 3D issues des scanners est une étape nécessaire pour beaucoup d'applications cliniques. La segmentation manuelle des structures par des experts est toujours considérée comme la référence. Cependant, la segmentation manuelle est une tâche très fastidieuse et coûteuse en temps. De plus, les segmentations générées manuellement sont sujettes à une variabilité inter et intra-expert. De ce fait, il est très important de proposer une méthode de segmentation rapide, précise et fiable facilitant l'étude des images IRM.

Pour de nombreuses structures anatomiques, y compris l'hippocampe, la définition anatomique donnée par les experts ne suit pas nécessairement le contenu de l'image IRM. Les frontières de la structure ne sont pas directement visibles sur l'image et ne suivent pas précisément ses gradients. Dans la Figure 1.3, les coupes d'une image IRM sont représentées, avec une segmentation manuelle de l'hippocampe (Mazziotta et al., 1995). La segmentation manuelle effectuée par l'expert ne suit pas les contours visibles sur l'IRM. La structure anatomique ne peut donc être segmentée précisément sans intégrer une information a priori dans la méthode de segmentation. Ainsi, de nombreuses méthodes de segmentation de structures anatomiques utilisent des modèles de référence, déjà segmentés par un expert.

1.2.1 Approches par recalage non linéaire

Approche du meilleur modèle

Depuis une vingtaine d'années, plusieurs paradigmes ont été présentés pour effectuer automatiquement la segmentation des structures du cerveau. Tout d'abord, des méthodes basées sur l'utilisation d'un modèle de référence, ou atlas, ont proposé d'effectuer un recalage non linéaire du modèle vers le sujet (Collins et al., 1995; Babalola et al., 2009). Une fois que la transformation non linéaire vers le sujet est

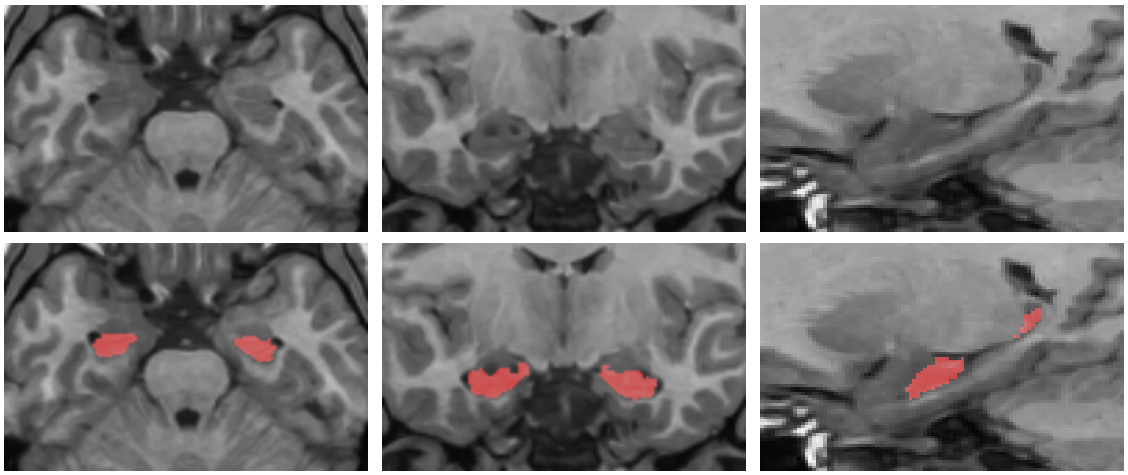


FIGURE 1.3 – Différence entre le contenu d'une image IRM (Mazziotta et al., 1995) et la définition anatomique de l'hippocampe donnée par une segmentation d'expert (ligne du bas).

calculée, la segmentation est obtenue en appliquant la même transformation à la segmentation du modèle vers le sujet. Cette approche de déformation d'un modèle de référence est illustrée dans la Figure 1.4. Ces méthodes ont été largement utilisées en raison de leur robustesse et de la facilité d'intégration des connaissances des experts via la déformation de segmentations manuelles. Cependant, ces méthodes peuvent s'avérer peu précises, et ne pas réussir à suffisamment capturer la variabilité entre sujets, en raison de l'hypothèse de correspondance globale entre un unique modèle et un sujet. De plus, elles sont sujettes à des erreurs de recalage.

Afin de minimiser les erreurs de recalage, l'approche du meilleur modèle a été proposée dans Barnes et al. (2008). L'idée est de réduire la distance anatomique entre le modèle utilisé et le sujet à segmenter, afin d'améliorer la précision du recalage. Le modèle le plus similaire est donc d'abord sélectionné dans une base de modèles et la même approche de segmentation par déformation est appliquée. Le modèle sélectionné est alors non linéairement recalé vers le sujet et la transformation calculée est appliquée à la segmentation manuelle du modèle pour obtenir la segmentation finale dans l'espace du sujet. Cette méthode permet d'améliorer les résultats de segmentation, mais est toujours soumise aux erreurs de recalage du fait de l'hypothèse de correspondance globale entre le sujet et un unique modèle.

Approches multi-modèles

Une amélioration significative des performances de segmentation a été obtenue avec l'utilisation simultanée de plusieurs modèles. Ces méthodes appliquent l'approche précédente de recalage à plusieurs modèles et fusionnent les informations de segmentation pour obtenir une meilleure précision (voir Figure 1.5).

Dans ces méthodes, les erreurs de recalage résultant de la variabilité inter-sujets

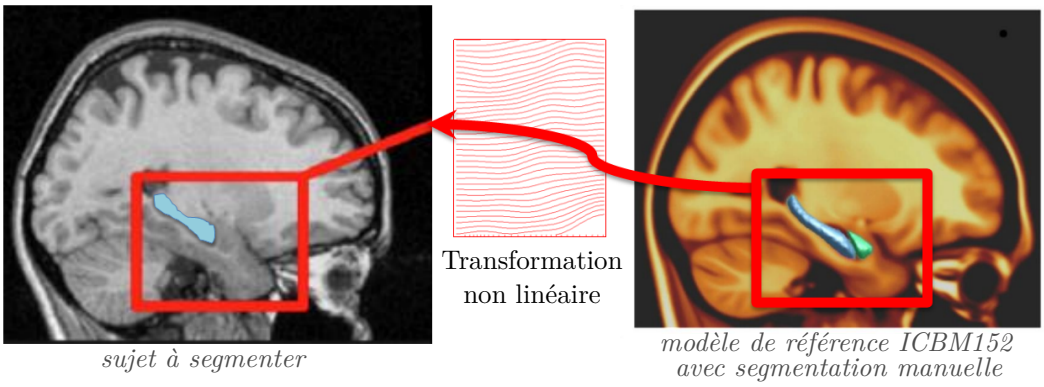


FIGURE 1.4 – Schéma de segmentation par recalage non linéaire (Collins et al., 1995). Une transformation non linéaire entre le modèle de référence ICBM152 et le sujet est calculée, puis appliquée à la structure de référence pour segmenter automatiquement le sujet. *Source : Collins et Pruessner (2010).*

sont considérées comme une variable aléatoire, et sont réduites avec l'utilisation de plusieurs sujets (Rohlfing et al., 2004; Heckemann et al., 2006). De nombreuses approches ont été proposées pour effectuer la phase de fusion des étiquettes de segmentation, comme la présélection des modèles les plus similaires, suivie par un vote à la majorité (Aljabar et al., 2009; Collins et Pruessner, 2010; Cardoso et al., 2013), les modèles d'intensité (Wolz et al., 2009; Lötjönen et al., 2010), les techniques de fusion d'étiquettes pondérées localement (Artaechevarria et al., 2009; Khan et al., 2011; Sabuncu et al., 2010) ou la correction systématique du biais de l'expert en utilisant une méthode d'apprentissage (Wang et al., 2011). Ces approches multi-modèles, bien que performantes en termes de précision de segmentation, sont obtenues en un temps important (jusqu'à plusieurs heures), dû au calcul coûteux de plusieurs transformations non linéaires.

1.2.2 Approche par fusion d'étiquettes basée sur les patchs

Plus récemment, l'algorithme de moyennes non locales (Buades et al., 2005) a été étendu à la fusion d'étiquettes basée sur les patchs sur une bibliothèque d'images 3D recalées affinement dans la méthode PBL (pour Patch-Based Label fusion, Coupé et al. (2011)). Cette méthode a été proposée pour réduire le temps de calcul des approches multi-modèles basées sur des recalages non linéaires. Le recalage affine utilise 12 paramètres pour déformer l'image avec une translation, une rotation et un éventuel zoom, et peut être calculé en seulement quelques secondes.

Plutôt que de calculer des transformations non linéaires, la méthode PBL nécessite uniquement un recalage affine, et s'appuie sur la comparaison de patchs de voxels, ou voisinages carrés, dans une zone de recherche restreinte entre le sujet à segmenter, et plusieurs images regroupées en une bibliothèque. Le recalage affine assure que les structures d'intérêt entre les images se trouvent dans des zones similaires.

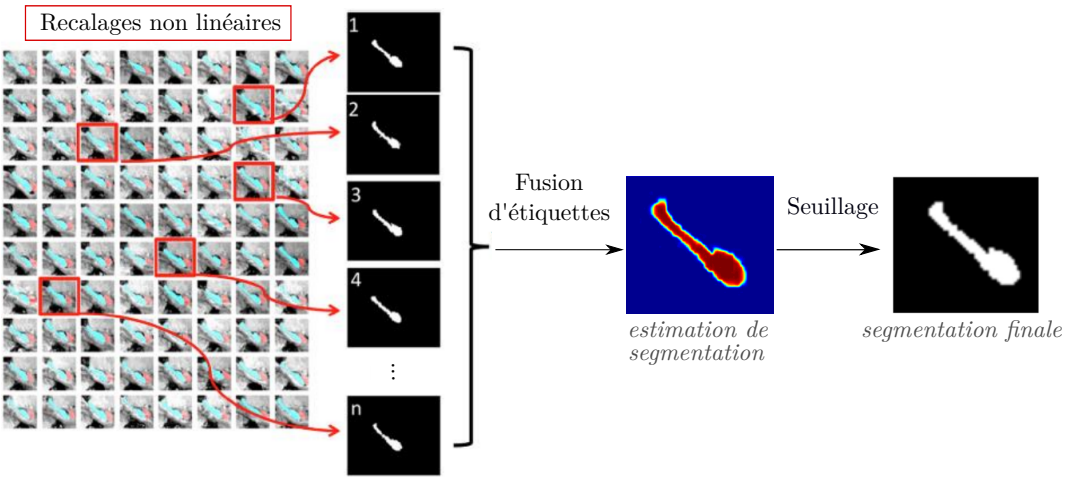


FIGURE 1.5 – Schéma de segmentation multi-modèles (Heckemann et al., 2006). Les plus proches modèles sont sélectionnés puis recalés vers le sujet à traiter. Les transformations sont alors appliquées aux segmentations manuelles qui sont enfin fusionnées pour fournir la segmentation finale du sujet. *Source : Collins et Pruessner (2010).*

Les patchs sont particulièrement efficaces pour capturer l'information contenue dans le voisinage d'un voxel. Chaque voxel est donc décrit localement par un patch, et l'étiquette de ce voxel est déterminée depuis l'information d'expert contenue dans les patchs des modèles de la bibliothèque.

La méthode PBL est illustrée dans la Figure 1.6. Les plus proches modèles de la bibliothèque sont présélectionnés afin de réduire le temps de calcul. Des distances sont alors calculées entre le patch à traiter dans le sujet (patch bleu) et les patchs de la bibliothèque (patchs noirs) dans une zone de recherche restreinte. Les contributions des patchs de la bibliothèque à l'étiquetage sont agrégées par une étape de fusion d'étiquettes basée sur les patchs, inspirée par la méthode des moyennes non locales (Buades et al., 2005). Les étiquettes des patchs les plus proches en termes d'intensité contribuent davantage à l'étiquetage final du voxel du sujet.

Cette stratégie non locale permet à PBL d'être robuste aux imprécisions de recalage et de capturer efficacement la variabilité inter sujets. PBL améliore ainsi les méthodes multi-modèles basées sur des associations globales de modèles d'exemple au sujet, grâce à l'utilisation de patchs qui décrivent localement le contenu des pixels dans le sujet et les modèles de la bibliothèque. Cette méthode produit des résultats de segmentation état-de-l'art en un temps de calcul réduit (quelques minutes).

Depuis son introduction, l'approche PBL a été beaucoup étudiée et étendue. Tout d'abord, la méthode peut être combinée avec d'autres approches multi-modèles (Rousseau et al., 2011; Hu et al., 2014; Wang et al., 2014). D'autres améliorations ont été proposées en utilisant un schéma multirésolution (Eskildsen et al., 2012), un apprentissage discriminant de dictionnaires et une approche de codage parcimonieux (Tong et al., 2013), ou des modèles de probabilité générative (Wu et al., 2014).

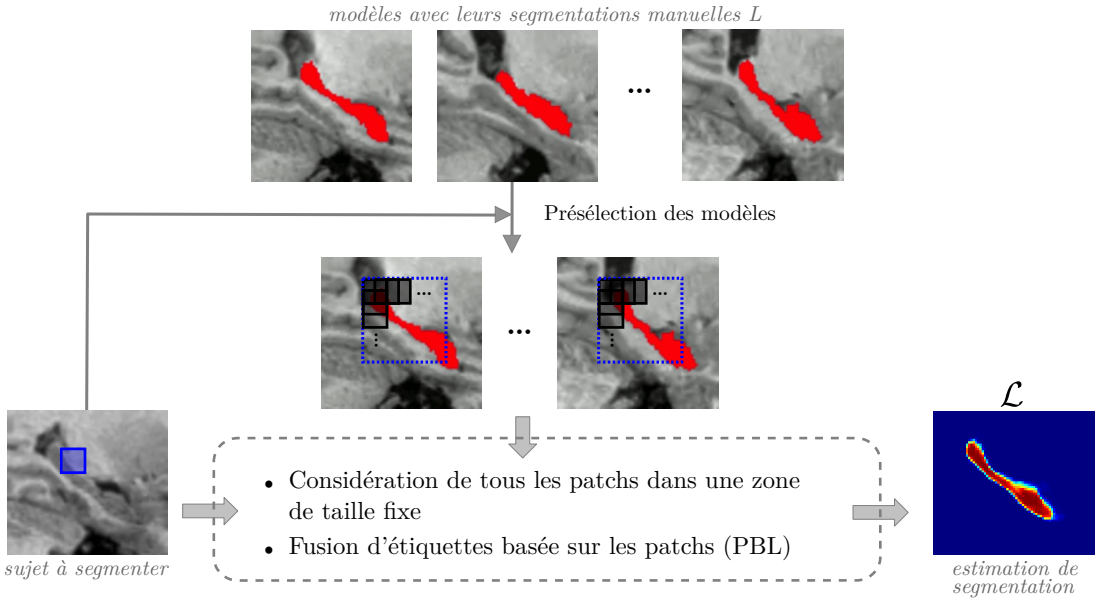


FIGURE 1.6 – Schéma de la méthode non locale de fusion d'étiquettes basée sur les patchs PBL (Coupé et al., 2011). Tous les patchs de la bibliothèque dans une zone de recherche restreinte sont considérés dans le schéma de fusion pour estimer la probabilité de segmentation d'un patch du sujet à traiter.

Cependant, PBL souffre encore de plusieurs limitations.

Limites

Tout d'abord, la recherche de patchs similaires est toujours coûteuse. Bien que la présélection de modèles (Rohlfing et al., 2004; Coupé et al., 2011) ou des stratégies multi-échelles (Eskildsen et al., 2012) aient été proposées, une grande quantité de calculs reste dédiée à la prise en compte exhaustive des patchs de la bibliothèque.

Ensuite, l'étape de présélection de modèles peut réduire la possibilité de trouver pour chaque patch, les voisins les plus similaires existant dans la bibliothèque. En effet, en sélectionnant des modèles selon une mesure globale de similarité avec le sujet, l'étape de présélection est susceptible d'éliminer des parties pertinentes de la bibliothèque, ce qui peut conduire à des résultats de segmentation sous-optimaux.

Enfin, dans les approches exploitant PBL, les comparaisons de patchs sont effectuées entre le patch traité et tous les patchs de la bibliothèque. La pertinence de chaque correspondance est ensuite pondérée en fonction de la similarité avec le patch du sujet à segmenter. Ainsi, des poids sont attribués à un grand nombre de patchs de la bibliothèque, y compris de nombreux patchs dissimilaires. Au-delà des calculs inutiles dédiés à l'estimation de poids négligeables, ces patchs dissimilaires peuvent diminuer la précision de segmentation (Tong et al., 2013). Les approches parcimonieuses ont tendance à limiter ce problème, mais souffrent d'un coût de calcul encore plus important (Tong et al., 2013; Wu et al., 2014).

Vers l'utilisation d'algorithmes de correspondance de patchs

Afin de pallier les limites de la méthode PBL, il est nécessaire de diminuer le temps de calcul tout en considérant tous les modèles de la bibliothèque sans étape de présélection, ainsi qu'un nombre limité de patchs de modèles dans l'étape de fusion. La recherche exhaustive des meilleures correspondances pour chaque patch étant trop coûteuse, nous proposons par la suite de nous limiter à la recherche de plus proches voisins approchés ANN (pour Approximate Nearest Neighbor).

En pratique, les méthodes ANN basées sur les patchs ont prouvé leur efficacité, en offrant des performances proches des approches exhaustives, pour des temps de calcul extrêmement réduits. En considérant uniquement de bonnes correspondances, il est possible de diminuer le nombre de patchs dans le modèle de fusion, et donc le temps de calcul associé à cette étape. Néanmoins, afin de s'affranchir de l'étape de présélection de modèles, l'algorithme de recherche d'ANN doit également pouvoir considérer une grande bibliothèque d'images 3D.

1.3 Algorithmes de correspondance à partir de patchs

1.3.1 Introduction des méthodes basées sur les patchs

Définition d'un patch

Un *patch* est défini comme un voisinage carré autour d'un pixel d'une image. Celui-ci permet de caractériser localement une région, et d'en extraire un motif. La redondance des mêmes motifs dans une ou plusieurs images permet alors d'extraire ou de transférer de l'information. Ces patchs peuvent contenir différentes informations, selon le descripteur utilisé, par exemple, les intensités, les couleurs des pixels, leur moyenne, leur gradient, etc.

Dans la suite, pour des raisons de simplicité, nous représentons les patchs en 2D, mais l'extension 3D est naturelle. Dans la Figure 1.7, un patch de taille 5×5 pixels est extrait d'une image 2D, et résume l'information locale autour du pixel central.

Un patch de l'image est uniquement défini par la position de son pixel central et une taille donnée. Il permet de représenter l'information de géométrie, c'est-à-dire, les positions relatives de ses pixels. Deux patchs $\mathrm{P}(p)$ et $\mathrm{P}(q)$ centrés respectivement sur des pixels p et q aux positions (x, y) et (x', y') d'une image A, peuvent alors être comparés par une distance pixel à pixel, par exemple, une somme des différences au carré sur les intensités, appelée SSD (pour Sum of Squared Differences) :

$$\mathrm{SSD}(\mathrm{P}(p), \mathrm{P}(q)) = \sum_{i=-s}^{s} \sum_{j=-s}^{s} \left(A(x+i, y+j) - A(x'+i, y'+j)\right)^2, \tag{1.1}$$

pour des patchs 2D $\mathrm{P}(p)$ et $\mathrm{P}(q)$ de taille $(2s+1)\times(2s+1)$ pixels.

FIGURE 1.7 – Illustration d'un patch régulier. Un patch de taille 5×5 pixels est extrait d'une image (a). Les valeurs de couleur des pixels (b), peuvent être résumées en un vecteur qui décrit le contenu du patch, dont un exemple est donné en (c).

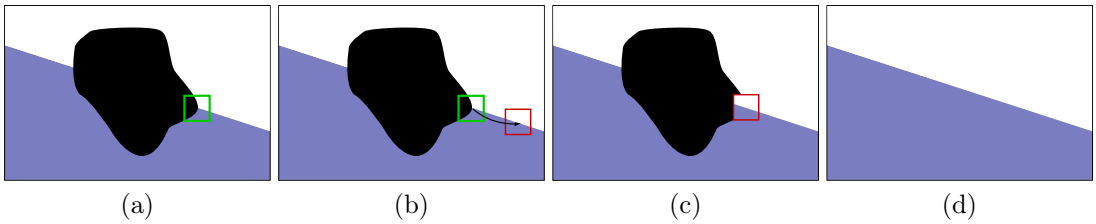


FIGURE 1.8 – Illustration de la méthode d'inpainting basée sur les patchs de Criminisi et al. (2004). Un patch (vert) est considéré sur le contour de la région à remplir (noire), en priorité sur les plus forts gradients (a). Un patch similaire (rouge) est trouvé dans le reste de l'image (b). Le patch sélectionné (rouge) vient alors remplir le patch vert (c). Ces étapes sont répétées jusqu'au remplissage complet de la région noire (d).

Méthodes basées sur les patchs

Les méthodes basées sur les patchs sont aujourd'hui utilisées pour de nombreuses applications telles que le calcul de flot optique (Chen et al., 2013; Bao et al., 2014), la détection de saillance (Borji et Itti, 2012), la segmentation (Coupé et al., 2011), ou encore la colorisation d'images (Bugeau et al., 2014). Elles ont été introduites très tôt en traitement d'images et sont devenues particulièrement populaires avec des applications de synthèse de texture (Efros et Leung, 1999) ou d'inpainting pour la reconstruction progressive d'une partie d'une image (Criminisi et al., 2004; Xu et Sun, 2010; Daisy et al., 2013).

La Figure 1.8 illustre un procédé d'inpainting basé sur les patchs où une région manquante est progressivement complétée par des patchs de l'image, et des exemples de résultats de remplacement d'objets de la scène par le fond sont donnés dans la Figure 1.9.

(a) (b)

FIGURE 1.9 – Exemples de résultats d'inpainting basé sur les patchs avec Criminisi et al. (2004) (a) et Daisy et al. (2013) (b).

Dans Buades et al. (2005), la redondance d'information dans une même image est utilisée pour le débruitage à partir de patchs, dans une méthode appelée moyennes non locales (*NL-means*). La redondance de motifs similaires permet d'annuler statistiquement la contribution du bruit altérant une image I. Pour un pixel $p \in I$, décrit par un patch $\mathrm{P}(p)$, les autres patchs $\mathrm{P}(q)$ de l'image sont considérés et contribuent à la valeur finale du pixel selon leur similarité avec le patch traité. L'intensité du pixel p dans l'image débruitée I_d est alors définie comme suit :

$$I_d(p) = \frac{\sum_{q \in I} \omega(p,q) I(q)}{\sum_{q \in I} \omega(p,q)}, \tag{1.2}$$

avec le poids ω, calculé selon une distance de type SSD (1.1) pour pondérer les contributions de chaque patch :

$$\omega(p,q) = \exp\left(-\frac{\mathrm{SSD}(\mathrm{P}(q), \mathrm{P}(q))}{h^2}\right), \tag{1.3}$$

avec h un paramètre qui dépend de l'intensité du bruit.

La Figure 1.10 illustre cette méthode de débruitage à l'échelle d'un patch. Les patchs de l'image ayant une forte similarité obtiennent des poids plus élevés et permettent de débruiter efficacement le patch traité. Un exemple de résultat est également donné en Figure 1.11, où la méthode de Buades et al. (2005) est comparée à des approches locales, de filtrage Gaussien et médian. L'utilisation de la redondance de motifs dans l'image permet un débruitage nettement plus efficace, conservant les textures, quand un traitement local des pixels a tendance à lisser l'information.

Ce schéma d'agrégation des patchs selon leur similarité s'avère particulièrement efficace, et peut permettre la fusion de différentes informations. Pour le débruitage, l'intensité des patchs d'une même image est considérée, mais des informations issues de correspondances dans d'autres images peuvent également être fusionnées. C'est le cas dans la méthode de segmentation d'images médicales PBL (Coupé et al., 2011) et dans les méthodes de fusion d'étiquettes proposées aux chapitres 2 et 8, et de fusion de couleurs au chapitre 9.

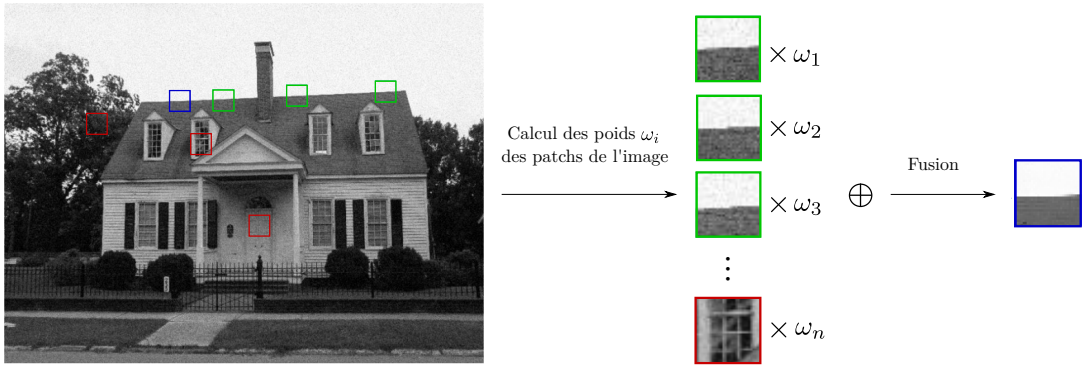


FIGURE 1.10 – Illustration de la méthode de débruitage par moyennes non locales (Buades et al., 2005). Le patch bleu est considéré. Des exemples de patchs ayant respectivement une forte (en vert) et faible (en rouge) similarité avec le patch bleu sont représentés. La fusion considère davantage les contributions des patchs verts qui permettent de supprimer le bruit en conservant le contour contenu dans le patch bleu.

FIGURE 1.11 – Comparaison des résultats de débruitage d'une image bruitée (b), depuis une image initiale (a), par filtage Gaussien (c), filtrage médian (d), et par moyennes non locales (Buades et al., 2005) (e). La méthode non locale permet un débruitage plus efficace, qui conserve notamment les textures.

Des correspondances entre les patchs d'une image et ceux d'une ou plusieurs autres peuvent être calculées pour extraire ou transférer de l'information. Dans le cas où la recherche de correspondances s'étend à une ou plusieurs images, la recherche exhaustive des plus proches patchs s'avèrent beaucoup trop coûteuse en temps de calcul. Des algorithmes de recherche approximatifs, ne fournissant qu'un ou plusieurs plus proches voisins approchés ANN (pour Approximate Nearest Neighbor) sont donc généralement utilisés.

1.3.2 Algorithmes de recherche de correspondances

Dans la Figure 1.12, nous considérons deux images proches A et B, et nous comparons la reconstruction de l'image A depuis les correspondances des plus proches patchs dans l'image B, calculées exhaustivement et approximativement avec une méthode ANN (Barnes et al., 2009). Les résultats sont visuellement très proches, quand la différence de temps de calcul est considérable. Les méthodes ANN peuvent donc fournir des correspondances très satisfaisantes permettant de se rapprocher des performances des approches exhaustives calculant les plus proches voisins.

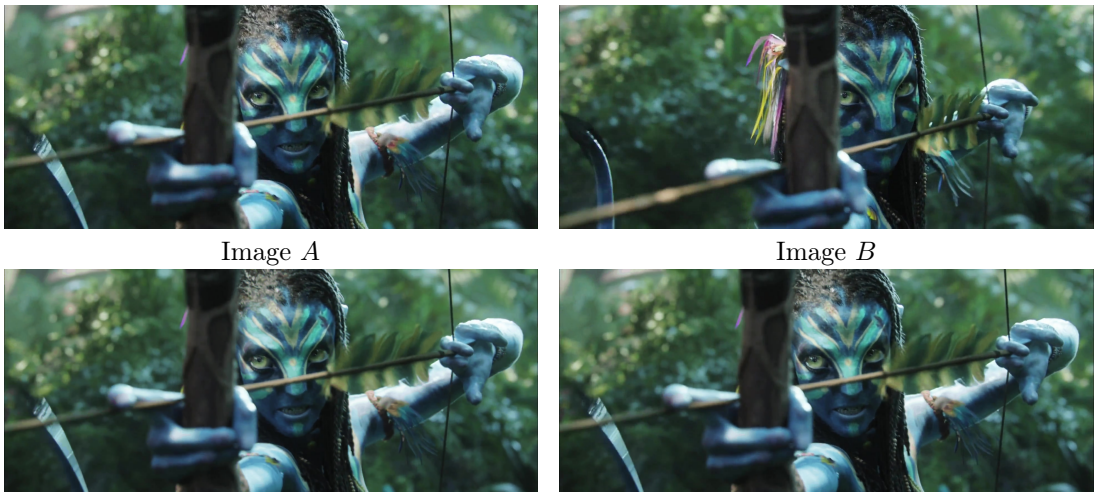


Image A — Image B

Image $\tilde{A}$ (Barnes et al., 2009) (t=14s) — Image $\tilde{A}$ (recherche exhaustive) (t=10h)

FIGURE 1.12 – Comparaison d'un algorithme de recherche exhaustive de plus proche voisin et d'ANN (Barnes et al., 2009). Deux images proches A et B de taille 1920×800 pixels issues de Korman et Avidan (2011) sont considérées, et l'image A est reconstruite depuis les patchs (de taille 5×5 pixels) sélectionnés dans B. L'algorithme d'ANN offre en quelques secondes un résultat quasi identique à celui de l'approche exhaustive, nécessitant plus de 10h de calcul avec du code C.

Pour calculer rapidement des ANN de patchs, des approches basées sur des arbres kd ont été proposées, par exemple, Arya et al. (1994); Wei et Levoy (2000); Muja et Lowe (2009); Olonetsky et Avidan (2012); He et Sun (2012), afin de décomposer les images en structures arborescentes facilitant la recherche. Des travaux tels que

Ashikhmin (2001); Tong et al. (2002) appliqués à la synthèse de texture, ont introduit la notion de cohérence spatiale (deux patchs similaires ont probablement des voisins similaires), pour favoriser la propagation des ANN. Ces approches basées sur des arbres kd souffrent néanmoins d'une complexité élevée, qui peut mener à d'importants temps de calcul lorsque de grandes images et/ou plusieurs images sont considérées (Indyk et Motwani, 1998).

Dans ce contexte, l'algorithme PatchMatch a été introduit pour l'édition rapide d'images (Barnes et al., 2009). L'approche est basée sur une recherche d'ANN itérative, en partie aléatoire. L'idée principale de l'algorithme est illustrée dans la Figure 1.13. La méthode consiste à associer aléatoirement tous les patchs de l'image A à un patch dans l'image B. Puis un processus de raffinement itératif est effectué, basé sur la propagation de bonnes de correspondances entre patchs adjacents, et sur la sélection aléatoire de candidats. La combinaison de ces deux étapes offre en pratique une très rapide convergence, et puisqu'aucun prétraitement n'est effectué, la complexité de la méthode dépend seulement du nombre de pixels à traiter dans l'image A.

L'algorithme a été étendu, par exemple dans Ramakanth et Babu (2012); Barnes et al. (2015), et a été appliqué notamment au calcul de flot optique (Bao et al., 2014; Hu et al., 2016) à la colorisation (Wang et al., 2012; Pierre et al., 2015), ou encore à l'inpainting (Newson et al., 2013). Pour cette dernière application, PatchMatch a notamment été intégré à des logiciels très populaires d'édition d'images.

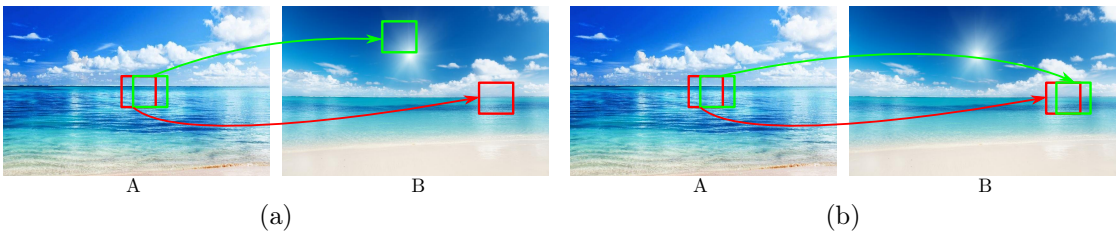


FIGURE 1.13 – Illustration du processus de propagation d'ANN de la méthode PatchMatch (Barnes et al., 2009). Deux patchs adjacents dans l'image A sont initialisés aléatoirement (a). Le patch vert est dans une zone de bonne correspondance. Le patch rouge teste alors la zone accédée par son voisin adjacent, le menant à une meilleure correspondance (b).

D'autres méthodes de recherche d'ANN ont été proposées, par exemple basées sur des architectures similaires aux réseaux de neurones (Revaud et al., 2016). Enfin, dans Korman et Avidan (2016), l'algorithme de hachage d'Indyk et Motwani (1998) et PatchMatch (Barnes et al., 2009) sont combinés. Dans le cas de la recherche d'ANN entre deux images, un hachage a priori des patchs permet de propager de plusieurs façons les bonnes correspondances. Les patchs similaires entre les deux images sont regroupés a priori dans une table de hachage, qui permet à un patch d'accéder directement à des candidats proches. Cependant, le hachage est effectué simultanément sur les images A et B considérées, avec des filtres 2D spécifiques. Ce prétraitement ne s'étend donc pas directement en 3D, sa complexité dépend de la taille de la bibliothèque et il doit être effectué pour chaque image à traiter.

1.4 Conclusion

L'analyse des structures cérébrales impactées par des maladies neurodégénératives, telles qu'Alzheimer, est une étape cruciale qui passe généralement par la segmentation. De nombreuses structures ont une définition anatomique qui ne correspond pas strictement aux contours des images IRM utilisées, et l'utilisation d'un a priori de segmentation d'expert est alors nécessaire. Les premières méthodes basées sur des déformations non linéaires de modèles pré-segmentés s'avèrent limitées par la variabilité anatomique entre les sujets, et sont coûteuses en termes de temps de calcul.

La méthode de fusion d'étiquettes basée sur les patchs PBL (Coupé et al., 2011) a été proposée pour réduire ce coût en recalant uniquement de façon affine les images, et en considérant les patchs des modèles dans une zone restreinte. Cependant, la méthode utilise une présélection des modèles, et la fusion de tous les patchs de la bibliothèque produit des résultats de segmentation sous-optimaux.

Nous proposons donc de considérer un algorithme rapide de recherche d'ANN pour ne considérer que quelques bonnes correspondances de patchs dans toute la bibliothèque de modèles. Bien que des méthodes récentes et efficaces de recherche d'ANN aient été proposées pour des images naturelles 2D, telles que (Muja et Lowe, 2009; Korman et Avidan, 2016), l'intérêt de ces méthodes est limité pour la segmentation par fusion d'étiquettes d'images médicales, puisqu'elles ne sont pas directement adaptables à une base d'images 3D, en un temps de calcul limité. En effet les approches de type kd-tree (Muja et Lowe, 2009) ou basées sur un hachage Korman et Avidan (2016) ont une complexité directement liée à la taille des images considérées, incluant celles de la bibliothèque.

Du fait de sa simplicité de mise en oeuvre (nombre réduit d'étapes), sa modularité (unique mesure de distance), et surtout de sa complexité proportionnelle à la taille de l'image du sujet, l'algorithme PatchMatch apparaît être un bon candidat dans le contexte de recherche d'ANN dans une grande bibliothèque d'images 3D. Néanmoins, comme présenté dans Barnes et al. (2009), PM reste limité à la recherche d'ANN entre deux images 2D naturelles, n'utilise aucun a priori de position pour les ANN, et ne retourne qu'une seule correspondance. Nous présentons donc dans le chapitre 2, les modifications proposées pour appliquer cet algorithme à notre contexte de segmentation d'images médicales 3D.

Chapitre 2

La méthode OPAL

Sommaire

2.1 Introduction 36
2.2 Algorithme PatchMatch optimisé 36
2.2.1 L'Algorithme PatchMatch 36
2.2.2 Initialisation contrainte 38
2.2.3 Propagation optimisée 40
2.2.4 Recherche aléatoire contrainte 40
2.2.5 PM multiples et calcul parallèle 41
2.3 Fusion d'étiquettes basées sur les patchs et segmentation 42
2.3.1 Fusion d'étiquettes basée sur les patchs 42
2.3.2 Filtrage spatial 44
2.4 Agrégation d'estimateurs multi-échelles et multi-descripteurs 44
2.4.1 Estimateurs multi-échelles 45
2.4.2 Estimateurs multi-descripteurs 46
2.4.3 Agrégation a posteriori d'estimateurs 46
2.5 Conclusion 46

2.1 Introduction

Dans ce chapitre, nous étendons l'algorithme PM au contexte de la segmentation d'images médicales, pour proposer un algorithme de recherche de multiples ANN de patchs 3D dans une bibliothèque d'images. Nous proposons plusieurs améliorations de l'algorithme initial permettant de réduire le temps de calcul, en tirant notamment profit de l'hypothèse de recalage linéaire des images considérées. La méthode proposée OPAL (pour Optimized PatchMatch for Accurate Label fusion) s'avère particulièrement rapide et efficace pour la recherche de correspondances dans une grande bibliothèque d'images, sans qu'une étape de présélection de modèles soit nécessaire. OPAL inclut l'algorithme de recherche de correspondances optimisé basé sur PM et une étape de fusion d'étiquettes de segmentation manuelle des modèles de la bibliothèque pour produire une segmentation automatique de l'image traitée.

Dans le chapitre 3, la méthode est validée et comparée à l'état-de-l'art sur la segmentation de plusieurs structures anatomiques. Bien que l'approche présentée ici soit adaptée au contexte de la segmentation d'images médicales 3D, les méthodes d'optimisation, pour obtenir plusieurs correspondances ainsi que de fusion d'étiquettes restent générales et sont notamment reprises dans les chapitres 7 et 8, où PM est cette fois adapté aux superpixels.

2.2 Algorithme PatchMatch optimisé

Dans cette section, nous présentons l'algorithme de recherche de plus proches voisins approchés (ANN), basé sur PM. Nous étendons l'algorithme, initialement introduit pour la correspondance de patchs entre deux images 2D, à la recherche de multiples ANN, décrits par des patchs 3D, dans une bibliothèque d'images. Notons que l'algorithme PatchMatch a été appliqué par exemple à la super-résolution d'images cardiaques IRM (Shi et al., 2013), mais la plupart des applications de l'algorithme concernent des problèmes d'édition d'images 2D.

2.2.1 L'Algorithme PatchMatch

L'algorithme PM original (Barnes et al., 2009) présenté dans la section 1.1 est une approche rapide qui calcule des correspondances de patchs entre deux images 2D. Le point clé de la méthode est que les bonnes correspondances peuvent être propagées aux patchs adjacents spatialement dans l'image traitée. Le coeur de la méthode est basé sur trois étapes : l'initialisation, la propagation, et la recherche aléatoire. L'initialisation consiste à associer aléatoirement chaque patch d'une image A à un patch dans une autre image B, afin d'obtenir une carte initiale de correspondance. Les deux étapes suivantes sont effectuées itérativement afin d'améliorer ces correspondances initiales, stockées dans une carte de correspondance notée $\mathcal{M}$.

L'étape de propagation se base sur l'hypothèse que lorsqu'un patch p centré sur $\mathbf{x_i} = (x, y) \in A$ correspond bien à un patch q centré sur $\mathbf{x_j} \in B$, alors les patchs adjacents de $p \in A$ devraient également bien correspondre aux patchs adjacents de $q \in B$. Le processus itératif suit un ordre de parcours de l'image (de gauche à droite, de haut en bas) lors des itérations paires, et inversement lors des itérations impaires. Durant l'étape de propagation, seuls les pixels adjacents traités pendant l'itération courante sont sélectionnés puisque les autres ont déjà été testés à l'itération précédente et leur correspondance n'a pas évolué depuis. Par exemple, lors d'une itération paire, pour un patch situé en $\mathbf{x_i} = (x, y) \in A$, seuls les patchs adjacents centrés sur $(x - 1, y)$ et $(x, y - 1)$ sont considérés durant la propagation. Soit $\mathbf{x'_j} \in B$ la correspondance du patch centré en $(x - 1, y) \in A$. Le candidat pour améliorer la correspondance de p est le patch centré en $\mathbf{x'_j} + (1, 0) \in B$.

Ensuite, l'étape de recherche aléatoire consiste en un échantillonnage aléatoire autour de l'ANN courant afin d'échapper à de potentiels minima locaux. Les candidats sont sélectionnés aléatoirement au sein de fenêtres de recherche de taille progressivement décroissante, centrées sur $\mathbf{x_j}$. La propagation de bonnes correspondances au sein du processus itératif de recherche, combiné à un échantillonnage aléatoire fournit en pratique une convergence très rapide de l'algorithme. La méthode PatchMatch est résumée dans l'Algorithme 1. Des versions plus détaillées de l'algorithme sont disponibles dans Ramakanth et Babu (2012); Besse et al. (2014).

Algorithme 1 PatchMatch(A, B)

1: **Pour** *chaque pixel* $\mathbf{x_i} \in A$ **faire**
2: Initialisation aléatoire de la carte de correspondance $\mathcal{M}(\mathbf{x_i}) = \mathbf{x_j} \in B$
3: **Pour** *chaque itération* **faire**
4: **Pour** *chaque pixel* $\mathbf{x_i} \in A$ **faire**
5: Propagation depuis les voisins $\mathbf{x_i} \pm [1, 0]$ et $\mathbf{x_i} \pm [0, 1]$
6: Recherche aléatoire autour de la meilleure correspondance $\mathcal{M}(\mathbf{x_i})$
7: **Retourner** $\mathcal{M}$

PatchMatch optimisé

Par rapport à Barnes et al. (2009) où des images 2D sont considérées, OPAL calcule des correspondances de patchs 3D entre une image S et une bibliothèque de n modèles $T = \{T_1, \ldots, T_n\}$. Un avantage majeur de PM est que sa complexité ne dépend que de la taille de l'image A, ou S, à segmenter et pas de l'image d'exemple B, ou T, dans le cas d'OPAL. Ce fait important motive l'utilisation de PM pour gérer de grandes bases d'images, et permet à OPAL de considérer une bibliothèque d'images T sans aucune étape de présélection à complexité constante. De plus, pour chaque patch dans S, OPAL calcule plusieurs ANN dans T et pas seulement une correspondance comme dans Barnes et al. (2009).

Dans la suite, nous détaillons la méthode OPAL. La Figure 2.1 résume les différentes étapes du processus de recherche, appelé OPM (pour Optimized PatchMatch).

Pour faciliter la compréhension des étapes de l'algorithme, l'initialisation et deux itérations de l'algorithme sont illustrées pour une bibliothèque T comprenant seulement 3 modèles, représentés avec leur segmentation manuelle sur une coupe 2D.

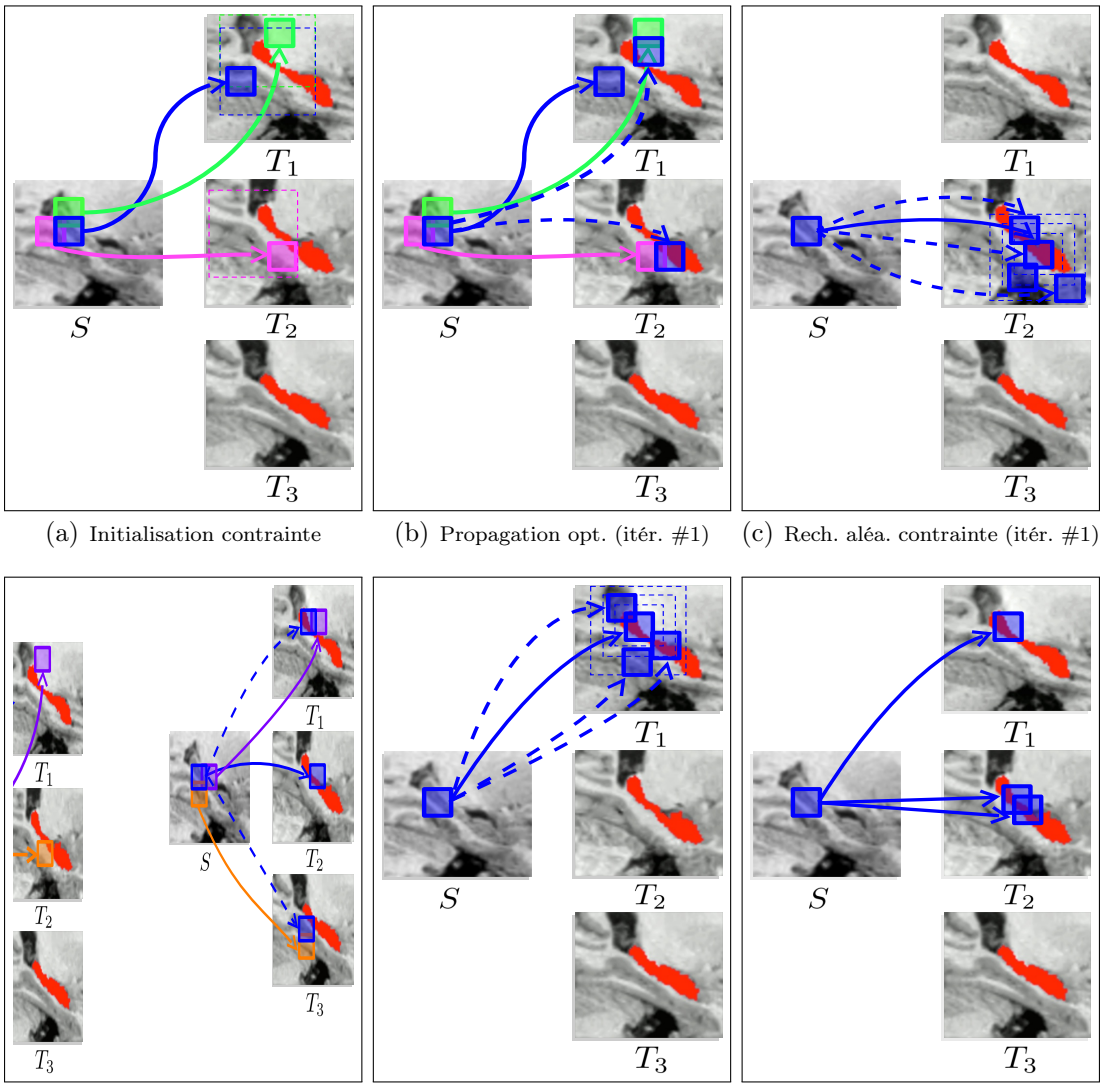
2.2.2 Initialisation contrainte

Dans le papier original de PM (Barnes et al., 2009), l'initialisation consiste à assigner, pour chaque patch à la position $(x, y) \in A$, une correspondance aléatoire qui peut être située à n'importe quelle position $(x', y') \in B$. Dans le cas de la recherche de correspondances dans plusieurs images 3D, l'extension naturelle de cette initialisation est d'assigner, pour chaque patch de l'image 3D à traiter S à la position $\mathbf{x_i} = (x, y, z) \in S$, une correspondance de patch aléatoire en $\mathbf{x_j} = \{(x', y', z'), t\}$, où $t \in \{1, \ldots, n\}$ représente le numéro du modèle T_t dans la bibliothèque T. Cependant, une telle initialisation entièrement aléatoire est faite pour traiter des images sans aucun a priori de zones de correspondance. Par conséquent, celle-ci peut s'avérer sous-optimale lorsque les images sont très similaires, ou que les zones d'intérêt ont peu varié spatialement.

Ici, nous considérons des images IRM 3D recalées linéairement dans un espace de référence (Avants et al., 2011). Par conséquent, les structures d'intérêt se trouvent dans une zone réduite de l'image. Nous proposons donc de contraindre spatialement l'initialisation pour que la correspondance (x', y', z') sélectionnée aléatoirement se trouve dans une fenêtre de recherche de taille fixe autour de la position (x, y, z) du voxel dans l'image S. Pour chaque voxel dans S, un indice $t \in \{1, \ldots, n\}$ est également sélectionné aléatoirement pour déterminer la position complète (x', y', z', t) de la correspondance. Chaque patch dans S est donc associé à une unique correspondance dans toute la bibliothèque T.

Du fait du nombre très important de patchs dans S, notons que tous les modèles de T sont largement susceptibles d'être tous sélectionnés plusieurs fois. Par exemple, dans les applications considérées dans le chapitre 3, les structures d'intérêt se situent dans des régions de taille $90 \times 60 \times 60$ voxels, et la bibliothèque contient environ une centaine de modèles. De plus, bien que le modèle correspondant soit sélectionné aléatoirement durant cette étape, les correspondances pourront passer d'un modèle à l'autre durant les étapes suivantes. La Figure 2.1(a) illustre l'étape d'initialisation. Pour chaque patch dans S (trois sont représentés), la fenêtre de recherche pour la sélection aléatoire à l'initialisation est représentée par les lignes pointillées.

Cette contrainte présente deux avantages. D'abord, elle permet d'accélérer la convergence en tirant profit du recalage entre l'image du sujet à traiter et les modèles de la bibliothèque. Ensuite, limiter l'initialisation à une zone de recherche fixée empêche l'algorithme de trouver des patchs proches en termes d'intensité (faible SSD), qui seraient spatialement trop éloignés pour être pertinents. Après cette étape d'initialisation contrainte, les étapes de propagation et de recherche aléatoire sont effectuées itérativement pour améliorer les correspondances initiales.



(a) Initialisation contrainte (b) Propagation opt. (itér. #1) (c) Rech. aléa. contrainte (itér. #1)

(d) Propagation opt. (itér. #2) (e) Rech. aléa. contrainte (itér. #2) (f) Multiples OPM

FIGURE 2.1 – Étapes de l'algorithme PatchMatch optimisé OPM illustrées pour 3 modèles en 2D. Les étapes sont illustrées pour le patch bleu dans S. Les patchs vert, rose, violet et orange représentent les voisins adjacents du patch bleu. Durant l'initialisation contrainte (a), les patchs du sujet S sont associés (lignes pleines) à un patch aléatoire de la bibliothèque dans une fenêtre de recherche dépendant de leur position dans S. L'étape de propagation est représentée pour les itérations #1 et #2 dans (b) et (d), respectivement. Les correspondances décalées des patchs adjacents déjà traités pendant l'itération sont testées pour améliorer la correspondance du patch bleu (lignes pointillées). La recherche aléatoire contrainte est représentée pour les itérations #1 et #2 dans (c) et (e), respectivement. Des tests de correspondance aléatoires sont effectués dans une zone de recherche autour du meilleur ANN actuel, dans le même modèle. Dans (f), plusieurs correspondances sont représentées, provenant d'OPM indépendants.

2.2.3 Propagation optimisée

La propagation est l'étape clé de l'algorithme PM et elle est ici étendue en 3D. Pour chaque patch en $(x, y, z) \in S$, on essaie d'améliorer la correspondance, c'est-à-dire minimiser la distance avec un patch dans T, en considérant l'information des voisins adjacents dans l'image. Comme dans PM, on utilise ici une distance de type SSD (1.1) entre les patchs 3D. Chaque patch étant initialisé aléatoirement, certains peuvent avoir trouvé une bonne zone de correspondance. Si une bonne zone a été trouvée pour un patch de S, celui-ci doit essayer de la propager à ses voisins spatiaux.

Les ANN des six voisins adjacents $(x \pm 1, y, z)$, $(x, y \pm 1, z)$, $(x, y, z \pm 1)$ seront testés au cours du processus. Selon le sens de parcours, les trois voisins déjà traités durant l'itération sont considérés. Les correspondances de ces voisins adjacents sont alors décalées dans T pour respecter la position relative des voisins dans S, et ces nouveaux patchs sont candidats pour améliorer la correspondance du patch traité.

Les Figures 2.1(b) et (d) illustrent cette étape, où les lignes bleues pointillées correspondent au test des correspondances décalées des voisins spatiaux. Dans cet exemple, la meilleure correspondance pour le patch bleu se déplace du modèle T_1 au modèle T_2 durant l'itération #1 et de T_2 à T_1 durant l'itération #2. L'étape de propagation est l'étape clé du processus de recherche OPM puisqu'elle permet de propager rapidement les bonnes zones de correspondances et de se déplacer entre les modèles de la bibliothèque.

Calcul rapide de distance

Puisque les patchs adjacents dans S et T se superposent, le calcul de distance peut être réduit en considérant celle du voisin avec sa correspondance, qui est nécessairement connue. Le calcul de distance peut alors être réduit aux extrémités des deux patchs qui ne se superposent pas. Pour des patchs de taille $n \times n \times n$ voxels, seules 2 aires de $n \times n$ voxels de part et d'autre des patchs ont besoin d'être considérées. Ainsi, le coût de la distance durant l'étape de propagation est réduit d'un facteur $n/2$. Des approches similaires de réduction de coût de calcul pour des distances entre patchs ont été proposées dans Xiao et al. (2011); Tapia (2011).

La superposition des patchs est illustrée en 2D dans les Figures 2.1(b) et (d), et également en 2D pour deux images A et B dans la Figure 2.2. Dans cette dernière, le patch bleu considère son voisin adjacent rouge pour le mener vers une meilleure correspondance. La distance entre les deux patchs rouges dans A et B étant connue, il suffit de lui retrancher la distance calculée sur l'aire verte et d'ajouter celle calculée sur l'aire jaune pour obtenir la distance exacte entre les deux patchs bleus.

2.2.4 Recherche aléatoire contrainte

Dans l'algorithme PM original (Barnes et al., 2009), la recherche aléatoire est effectuée sur toute l'image B. Des positions indiquant les patchs à tester sont tirées aléatoirement dans des zones de taille décroissante autour de la meilleure correspon-

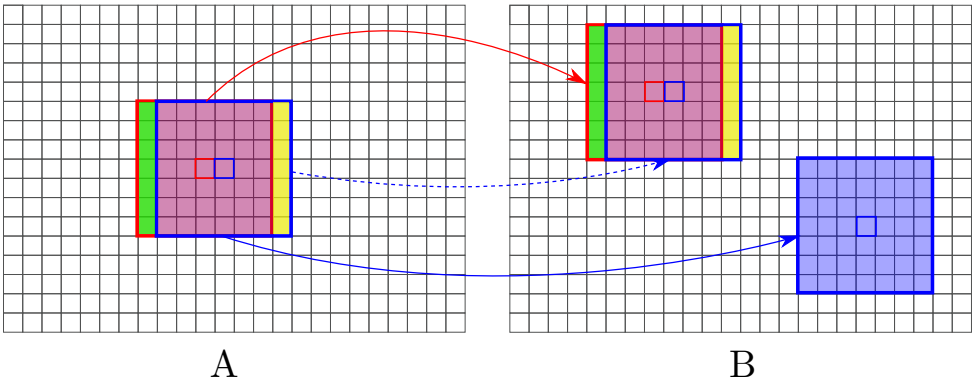


FIGURE 2.2 – Illustration 2D du calcul de distance optimisé lors de l'étape de propagation d'OPM. Le patch bleu teste la correspondance décalée de son voisin rouge dont la distance avec sa correspondance est connue. Par conséquent, pour avoir la distance entre les deux patchs bleus (ligne pointillée), il suffit de retrancher la distance entre les deux bandes vertes et d'ajouter celle entre les deux bandes jaunes. Dans le cas de 3D, seules les distances sur les deux carrés de part et d'autre des patchs doivent être calculées.

dance actuelle. Par rapport à la méthode originale, OPM considère une bibliothèque d'images recalées linéairement (voir section 3.2.1). Cette étape de recherche aléatoire est modifiée pour prendre en compte cet aspect. Afin d'assurer la cohérence spatiale, et de façon similaire à l'initialisation, nous effectuons la sélection aléatoire dans une zone restreinte autour de la meilleure correspondance dans le modèle courant. Autrement dit, t est fixé, et on sélectionne aléatoirement des positions $(x'_t, y'_t, z'_t) \in T_t$ dans des zones de recherche carrées dont la taille décroît progressivement d'un facteur 2 à chaque fois qu'un nouveau candidat est sélectionné. Dans la suite, la taille de la première fenêtre de recherche aléatoire est fixée à la taille de la fenêtre d'initialisation. Les Figures 2.1(c) et (e) illustrent cette étape.

À noter que durant les étapes de propagation et de recherche aléatoire, les correspondances peuvent sortir de la fenêtre d'initialisation. En pratique, peu d'ANN se situent en dehors de cette zone, et l'étape de fusion présentée dans la section 2.3, intègre une distance spatiale entre les patchs du sujet S et leur ANN dans T.

2.2.5 PM multiples et calcul parallèle

Contrairement à Barnes et al. (2009) qui ne fournit qu'un ANN avec PM, OPAL calcule k-ANN dans T. Une extension de l'algorithme PM original à la recherche de k-ANN a été proposée dans Barnes et al. (2010). La stratégie est alors de construire une pile des meilleurs patchs visités. À chaque nouveau patch, la distance est comparée à la moins bonne de la pile. Cependant, une telle approche est difficile à paralléliser puisque l'image S doit être découpée en plusieurs morceaux, et l'étape de propagation qui utilise les patchs adjacents, n'est alors pas optimale aux frontières.

Par conséquent, nous décidons d'implémenter la recherche de k-ANN en effectuant k recherches indépendantes. Un processus est alors dédié à chaque OPM sans dépendance aux autres. Avec cette approche, les mêmes ANN peuvent être sélectionnés par plusieurs processus. Les correspondances sélectionnées plusieurs fois ont alors plus de poids dans l'étape de fusion. Dans nos expériences de validation au chapitre 3, la sélection des mêmes ANN s'est avérée très rare. La Figure 2.1(f) illustre le résultat obtenu par plusieurs OPM indépendants ($k = 3$).

L'utilisation de l'algorithme PatchMatch et les différentes améliorations apportées permettent à la méthode OPAL de fournir une segmentation de structures cérébrales comme l'hippocampe en moins de 2s contre près de 10mn pour la méthode PBL (Coupé et al., 2011).

2.3 Fusion d'étiquettes basées sur les patchs et segmentation

Après convergence des multiples OPM, la position et la distance des k-ANN sont connues. Une étape de fusion d'étiquettes basée sur les patchs peut alors être effectuée pour transférer l'information des segmentations manuelles de la bibliothèque et générer une segmentation du sujet traité. Les étiquettes sont fusionnées selon la pertinence des correspondances trouvées dans la bibliothèque pour générer une carte d'estimation de segmentation du sujet. Par rapport à la méthode de fusion d'étiquettes basée sur les patchs originale PBL (Coupé et al., 2011), qui ne transfère que l'information du voxel central, OPAL effectue une fusion basée sur les patchs en transférant toute l'information du patch sélectionné, comme dans Rousseau et al. (2011); Wu et al. (2014); Manjón et al. (2014). De plus, comme proposé dans Manjón et al. (2014), OPAL utilise un filtrage spatial pour également tenir compte du recalage linéaire et ainsi renforcer la cohérence spatiale. La Figure 2.3 illustre ce processus de fusion basée sur les patchs des étiquettes de la bibliothèque, et le calcul des cartes d'estimateur de segmentation.

2.3.1 Fusion d'étiquettes basée sur les patchs

À la fin du processus de recherche de correspondance, k-ANN sont estimés pour chaque patch de S, et la position et la distance avec leurs k-ANN dans T sont connues. Pour obtenir une segmentation finale depuis ces correspondances, il faut transférer l'information de segmentation manuelle. La méthode de fusion basée sur les patchs PBL (Coupé et al., 2011) suit l'approche des moyennes non locales (Buades et al., 2005). Celle-ci considère tous les patchs d'une présélection de modèles dans l'étape de fusion et pondère leur information de segmentation en fonction de leur similarité avec les patchs du sujet. En revanche, comme mentionné dans la section 1.2.2, l'utilisation de seulement k correspondances par OPAL peut limiter l'erreur

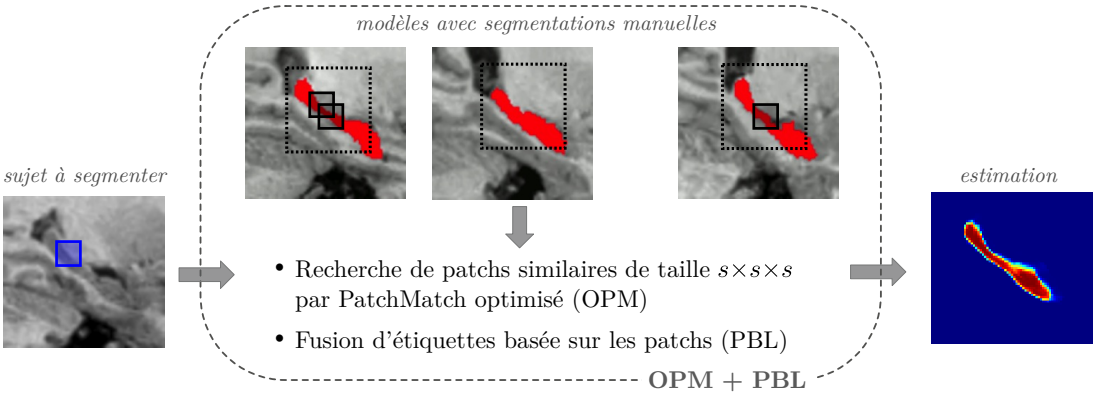


FIGURE 2.3 – Illustration de la méthode OPAL : recherche d'ANN par PatchMatch optimisé (OPM) et fusion basée sur les patchs des étiquettes (PBL). Pour chaque voxel du sujet à segmenter, une recherche de patchs similaires de taille $s \times s \times s$ est effectuée par OPM. L'information de segmentation des patchs sélectionnés est alors transférée par l'étape de fusion (PBL) pour calculer la carte d'estimateur de segmentation.

de segmentation, et effectuer cette recherche sur toute la bibliothèque peut augmenter la précision de segmentation. Considérons un patch 3D $\mathrm{P}(\mathbf{x_i})$ à la position $\mathbf{x_i} = (x, y, z) \in S$, et $\mathcal{K}_i = \{\mathbf{x_{j,t}}\}$ l'ensemble des positions de ses k-ANN, sa fusion d'étiquettes $\mathcal{L}(\mathbf{x_i})$ est définie par :

$$\mathcal{L}(\mathbf{x_i}) = \frac{\sum_{\mathbf{x_{j,t}} \in \mathcal{K}_i} \omega(\mathbf{x_i}, \mathbf{x_{j,t}}) l(\mathbf{x_{j,t}})}{\sum_{\mathbf{x_{j,t}} \in \mathcal{K}_i} \omega(\mathbf{x_i}, \mathbf{x_{j,t}})}, \tag{2.1}$$

où $\omega(\mathbf{x_i}, \mathbf{x_{j,t}})$ est le poids assigné à $l(\mathbf{x_{j,t}})$, l'étiquette, ici binaire (0 pour le fond, 1 pour la structure), donnée par la vérité terrain d'expert au voxel $\mathbf{x_{j,t}} = \{\mathbf{x_j}, t\} \in T$. Cette fusion peut être faite pour plusieurs étiquettes (voir chapitres 3 et 8).

Le poids $\omega(\mathbf{x_i}, \mathbf{x_{j,t}})$ dépend de la similarité entre le patch $\mathrm{P}(\mathbf{x_i}) \in S$, contribuant à l'étiquetage de $\mathbf{x_i}$, et le patch ANN $\mathrm{P}(\mathbf{x_{j,t}}) \in T$. Ce poids est défini comme :

$$\omega(\mathbf{x_i}, \mathbf{x_{j,t}}) = \exp\left(1 - \frac{\|\mathrm{P}(\mathbf{x_i}) - \mathrm{P}(\mathbf{x_{j,t}})\|_2^2}{h(\mathbf{x_i})^2}\right), \tag{2.2}$$

où $h(\mathbf{x_i})^2 = \alpha^2 \min_{\mathbf{x_{j,t}} \in \mathcal{K}_i} (\|\mathrm{P}(\mathbf{x_i}) - \mathrm{P}(\mathbf{x_{j,t}})\|_2^2 + \epsilon)$, avec ϵ une faible constante pour assurer la stabilité numérique, et α une constante de normalisation. Avec le paramètre $h(\mathbf{x_i})$, la distance de la contribution courante est divisée par la distance minimale parmi toutes les contributions des k-ANN.

La plupart des méthodes de fusion non locales effectuent uniquement le transfert de l'information de l'élément central (pixel ou voxel), ce qui peut mener à un manque de régularité de la segmentation finale. Ainsi, pour améliorer la qualité de la segmentation, la fusion des étiquettes est effectuée sur tout le patch dans OPAL, comme

dans Rousseau et al. (2011); Wu et al. (2014); Manjón et al. (2014). L'étiquetage basé sur les patchs est donc calculé comme suit :

$$\mathcal{L}(\mathrm{P}(\mathbf{x_i})) = \frac{\sum_{\mathbf{x_{j,t}} \in \mathcal{K}_i} \omega(\mathbf{x_i}, \mathbf{x_{j,t}}) l(\mathrm{P}(\mathbf{x_{j,t}}))}{\sum_{\mathbf{x_{j,t}} \in \mathcal{K}_i} \omega(\mathbf{x_i}, \mathbf{x_{j,t}})}. \tag{2.3}$$

et toute l'information de segmentation des patchs des k-ANN $\mathrm{P}(\mathbf{x_{j,t}})$ est transférée au patch $\mathrm{P}(\mathbf{x_i}) \in S$. Ainsi, la carte de probabilité d'étiquetage $\mathcal{L}$ est obtenue en moyennant les contributions de tous les patchs contenant $\mathbf{x_i}$.

2.3.2 Filtrage spatial

En plus de la stratégie de fusion basée sur les patchs, nous proposons d'utiliser, comme dans Manjón et al. (2014), une distance spatiale durant la fusion afin de renforcer la cohérence spatiale des k-ANN sélectionnés. Cette distance spatiale tire profit du recalage linéaire qui assure que les structures d'intérêt sont spatialement proches. Ainsi, de bonnes correspondances doivent être similaires en termes d'intensité mais également peu éloignées spatialement. Chaque contribution à l'étiquetage est donc également pondérée par la distance spatiale entre les centres des deux patchs $\mathbf{x_i} \in S$ et $\mathbf{x_{j,t}} = \{\mathbf{x_j}, t\} \in T$:

$$\omega(\mathbf{x_i}, \mathbf{x_{j,t}}) = \exp\left(1 - \left(\frac{\|\mathrm{P}(\mathbf{x_i}) - \mathrm{P}(\mathbf{x_{j,t}})\|_2^2}{h(\mathbf{x_i})^2} + \frac{\|\mathbf{x_i} - \mathbf{x_j}\|_2}{\sigma^2}\right)\right), \tag{2.4}$$

où σ^2 est un paramètre d'échelle.

2.4 Agrégation d'estimateurs multi-échelles et multi-descripteurs

Du fait du coût de calcul important, la plupart des méthodes de segmentation basées sur les patchs considèrent une seule métrique, c'est-à-dire, un seul type de descripteur. De nettes améliorations peuvent être apportées par des approches multi-échelles (Eskildsen et al., 2012; Wachinger et al., 2014; Wu et al., 2015) et multi-descripteurs (Kim et al., 2013b; Bai et al., 2015). Cependant, puisque chaque méthode requiert un temps de calcul non négligeable, elles ne proposent pas de considérer simultanément plusieurs descripteurs à plusieurs échelles.

Ces méthodes agrègent donc de façon précoce les différents descripteurs : toutes les échelles ou descripteurs sont fusionnés dans un seul vecteur. Ainsi, la recherche de patchs correspondants se fait sur tous les descripteurs à la fois. Cette stratégie de fusion précoce n'est pas nécessairement la meilleure (Snoek et al., 2005), et est généralement seulement utilisée par souci d'économie de calcul.

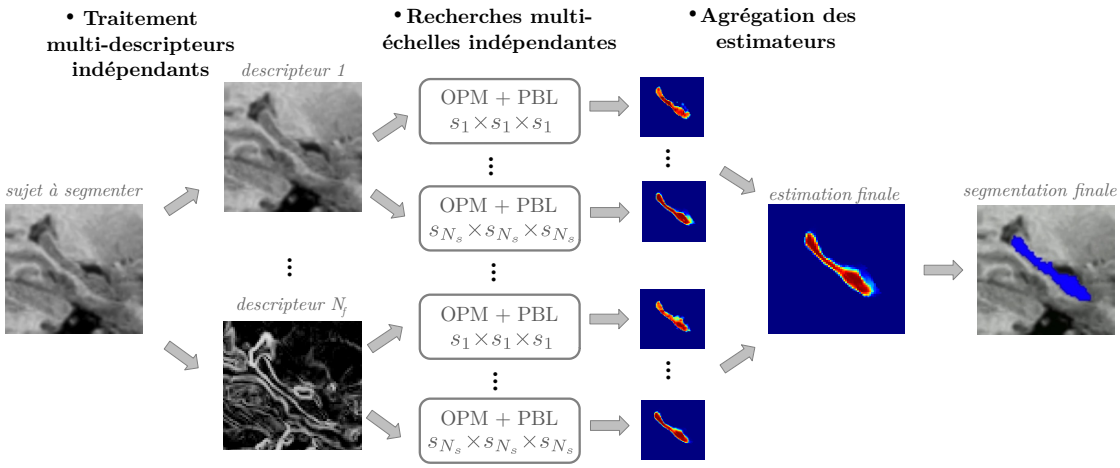


FIGURE 2.4 – Illustration de la méthode OPAL. Agrégation des estimateurs de segmentation multi-descripteurs et multi-échelles. L'algorithme est effectué sur N_s différentes tailles de patch, sur N_f différents descripteurs, donc $N = N_s \times N_f$ estimateurs sont calculés et agrégés pour générer la segmentation finale.

La rapidité de la méthode OPAL nous permet d'envisager une nouvelle stratégie de recherche multi-échelles et multi-descripteurs et de fusionner a posteriori les différents estimateurs. La Figure 2.4 illustre la méthode OPAL complète et la fusion a posteriori des estimateurs multi-descripteurs et multi-échelles.

2.4.1 Estimateurs multi-échelles

Dans les méthodes basées sur les patchs, la description efficace d'une structure ou d'un motif dépend grandement de la taille du patch. Bien que la taille de patch optimale sur une base de données puisse être déterminée empiriquement, les approches multi-échelles permettent d'augmenter significativement la précision de segmentation, comme l'ont montré de récents travaux (Wachinger et al., 2014; Pant et al., 2015; Wu et al., 2015). Dans ces travaux, la recherche d'ANN consiste à trouver des correspondances qui minimisent la distance sur toutes les échelles en même temps. Cette stratégie sélectionne des correspondances qui fournissent la meilleure similarité moyenne sur plusieurs échelles, mais cette approche peut s'avérer sous-optimale.

Par conséquent, avec OPAL, nous proposons d'effectuer la recherche d'ANN de façon indépendante sur chaque échelle. Notons qu'ici le terme multi-échelles désigne l'utilisation de patchs de N_s différentes tailles $s_{i,i\in\{1,\ldots,N_s\}}$ sur les images à pleine résolution. Dans la Figure 2.4, la recherche d'ANN et la fusion sont effectuées pour chaque descripteur sur les N_s tailles de patch.

2.4.2 Estimateurs multi-descripteurs

De façon similaire, la recherche d'ANN par OPM peut également être menée indépendemment sur différents descripteurs. En effet, les intensités des images initiales ne contiennent pas nécessairement la meilleure information possible pour effectuer la recherche des correspondances. Selon l'application, il peut être intéressant de privilégier l'information de texture ou de contour, et de nombreux travaux cherchent à proposer des descripteurs robustes intégrant ces informations (Matsakis et Wendling, 1999; Lowe, 2004; Dalal et Triggs, 2005; Bay et al., 2006; Sharma et al., 2013; Trulls et al., 2014). La Figure 2.4 montre comment OPAL est appliquée sur N_f descripteurs extraits du sujet S à segmenter. Dans le chapitre 3, nos expériences sur la segmentation de l'hippocampe montrent que l'utilisation de la norme du gradient des intensités, en plus de l'intensité des images initiales ($N_f = 2$) offre de très bonnes performances.

2.4.3 Agrégation a posteriori d'estimateurs

Les estimations sont calculées indépendemment par l'étape de fusion des étiquettes (2.3) pour chaque configuration multi-échelles, multi-descripteurs. La dernière étape consiste à agréger ces estimateurs pour générer la segmentation finale. Ici, OPAL est appliquée sur N_f descripteurs, avec N_s tailles de patchs différentes, donc $N = N_s \times N_f$ estimations $\mathcal{L}_i$ avec $i \in \{1, \ldots, N\}$ sont calculées et considérées pour générer la segmentation finale. L'estimateur final $\mathcal{F}$ est alors calculé simplement par la moyenne des estimations selon une approche de fusion tardive (Snoek et al., 2005) :

$$\mathcal{F} = \frac{\sum_{i=1}^{N} \mathcal{L}_i}{N}. \tag{2.5}$$

Enfin, pour générer la segmentation automatique finale $\mathcal{S}(\mathbf{x_i})$ en un voxel $\mathbf{x_i}$, la décision d'étiquetage finale est prise par seuillage comme suit :

$$\mathcal{S}(\mathbf{x_i}) = \begin{cases} 1, & \text{si } \mathcal{F}(\mathbf{x_i}) \geq 0.5, \\ 0, & \text{sinon.} \end{cases} \tag{2.6}$$

2.5 Conclusion

Dans ce chapitre, nous proposons une nouvelle méthode, OPAL, de recherche rapide de correspondances basée sur les patchs, avec une fusion d'étiquettes destinée à la segmentation de structures anatomiques. Les propriétés inhérentes à l'algorithme PatchMatch, comme sa complexité liée uniquement à la taille de l'image à traiter, le rendent intéressant dans notre contexte de grandes bases de données d'images d'exemple 3D. Nous étendons donc l'algorithme à la recherche de multiples ANN dans une bibliothèque d'images.

Nous proposons également d'optimiser la méthode à certaines étapes clés, l'initialisation et la recherche aléatoire. Une stratégie efficace pour calculer de multiples correspondances est également proposée, permettant la parallélisation des recherches. Le faible temps de calcul offert par OPAL permet de proposer une nouvelle approche de recherches multi-échelles et multi-descripteurs indépendantes, dont les estimateurs sont fusionnés a posteriori.

Dans le chapitre suivant, pour illustrer les performances d'OPAL en termes de précision et de temps de calcul, la méthode est appliquée à la segmentation de l'hippocampe et du cervelet sur plusieurs bases de données. La méthode proposée reste cependant générale et les stratégies pour obtenir de multiples correspondances ainsi que la fusion d'étiquettes basée sur les patchs sont également utilisées dans les chapitres 7 et 8 où PatchMatch est adapté aux superpixels.

Chapitre 3

Résultats de segmentation d'images médicales

Sommaire

3.1 Introduction 50
3.2 Validation d'OPAL sur la segmentation de l'hippocampe 50
3.2.1 Schéma de validation 50
3.2.2 Influence des paramètres 53
3.2.3 Comparaison aux méthodes de l'état-de-l'art 56
3.2.4 Ajout de segmentations automatiques à la bibliothèque 59
3.2.5 Transfert 61
3.3 Segmentation du cervelet 62
3.3.1 Contexte 62
3.3.2 Amélioration de la fusion des estimateurs 63
3.3.3 Comparaison à l'état-de-l'art 64
3.4 Extension à la classification de la maladie d'Alzheimer 65
3.4.1 Classification depuis le volume des segmentations automatiques 66
3.4.2 Classification depuis la fusion d'étiquettes de groupe 68
3.5 Conclusion 70

3.1 Introduction

Dans ce chapitre, nous appliquons la méthode OPAL présentée au chapitre 2 à la segmentation de plusieurs structures du cerveau. Nous considérons dans un premier temps deux bases de données d'images IRM fournies avec des segmentations d'expert de l'hippocampe. Une de ces bases contient des sujets atteints de pathologies neurodégénératives et qui ont donc entre eux une variabilité anatomique plus élevée.

Nous étudions alors l'influence des paramètres d'OPAL tels que la taille de patch ou le nombre d'ANN sur les performances de segmentation. Ces performances sont comparées à celles de méthodes de l'état-de-l'art appliquées sur les deux bases considérées, et l'évaluation des performances s'effectue avec l'outil standard, le coefficient Dice. La précision de segmentation dépend fortement de la variabilité des segmentations manuelles, et les méthodes automatiques tendent vers les performances de fiabilité inter-experts (coefficients Dice entre plusieurs segmentations manuelles du même sujet). Ainsi, même de faibles gains en termes de coefficients Dice s'avèrent très significatifs, et nous validons la supériorité des résultats obtenus avec OPAL en utilisant également des tests statistiques.

Des adaptations d'OPAL sont également présentées pour effectuer une segmentation état-de-l'art des sous-structures du cervelet sur une autre base de données (section 3.3). Enfin, une extension de la méthode à la classification de la maladie d'Alzheimer est également proposée (section 3.4).

3.2 Validation d'OPAL sur la segmentation de l'hippocampe

3.2.1 Schéma de validation

Bases de données

Deux bases de sujets ont été considérées. La première contient des images de jeunes adultes, obtenus depuis la base ICBM (pour International Consortium for Brain Mapping, Mazziotta et al. (1995)). La seconde est composée d'images d'adultes plus âgés, provenant du projet international ADNI (pour Alzheimer's Disease Neuroimaging Initiative, Jack et al. (2008)). Les images de la base ADNI sont obtenues sur des sujets pouvant être atteints de pathologies neurodégénératives ayant impacté les structures de leur cerveau.

Notre objectif est de démontrer la robustesse de la méthode proposée OPAL sur des données provenant de sources multiples avec des pipelines de prétraitement différents. Les détails d'acquisition des images pour les deux bases considérées sont donnés dans Giraud et al. (2016a).

ICBM. Nous considérons une partie de la base ICBM (Mazziotta et al., 1995) qui est composée de 80 IRM de jeunes adultes sains. Les coefficients Dice (3.1) correspondant aux fiabilités inter et intra-expert sur les segmentations manuelles de la base ont respectivement été estimés à 90% et 92%. Ces coefficients indiquent respectivement la similarité entre les segmentations effectuées par plusieurs experts pour un même sujet, et par le même expert effectuant plusieurs segmentations du même sujet à quelques jours d'intervalle.

Toutes les images sont prétraitées avec le pipeline suivant : estimation de l'écart type du bruit (Coupé et al., 2010) ; débruitage utilisant un filtrage basé sur la méthode des moyennes non locales (Coupé et al., 2008) ; correction des inhomogénéitées (Sled et al., 1998) ; recalage par une transformation affine vers le modèle de référence ICBM152 (Collins et al., 1994) ; normalisation linéaire des intensités de chaque sujet ; normalisation croisée des intensités des images IRM entre les sujets avec Manjón et al. (2008a). Ce pipeline complet de prétraitement est effectué en moins de 5min par sujet.

EADC-ADNI. Le projet ADNI vise à étudier la maladie d'Alzheimer à travers divers initiatives. ADNI propose notamment le suivi de patients au fil des années, en fournissant régulièrement les images IRM des cerveaux avec les segmentations des structures d'intérêt telles que l'hippocampe. Dans ce contexte, le protocole HarP (pour Harmonized Protocol [1]) a été défini afin de réduire les incohérences entre les différents protocoles de segmentation existants (Boccardi et al., 2014), et les segmentations basées sur ce protocole permettent de valider les algorithmes de segmentation automatique. La robustesse du protocole est illustrée sur des exemples de segmentation effectuées par plusieurs experts, par exemple dans Frisoni et al. (2015).

La base EADC-ADNI est composée de 100 IRM issues de la base ADNI (Jack et al., 2008), qui ont été manuellement segmentées par des experts en suivant le protocole HarP. La fiabilité inter-experts, ou valeur moyenne des coefficients Dice (3.1) entre les différentes segmentations d'un même sujet par des experts différents a été estimée à 89% selon Tangaro et al. (2014).

Toutes les images sont prétraitées en utilisant le pipeline de volBrain [2], dont les principales étapes sont les suivantes. La première étape du pré-traitement consiste à appliquer un filtre adaptatif basé sur la méthode des moyennes non locales (Manjón et al., 2010). L'inhomogénéité des images IRM débruitées est alors lissée avec Tustison et al. (2010). Ensuite, un recalage affine est effectué en utilisant Avants et al. (2011). Une étape de correction d'inhomogénéité est effectuée en utilisant Weiskopf et al. (2011). Enfin, une normalisation des intensités est effectuée sur les images (Manjón et al., 2008a). Comme pour la base ICBM, ce pipeline complet de traitement est effectué en moins de 5min par sujet.

1. www.hippocampal-protocol.net
2. http://volbrain.upv.es

Ces bases couvrent différents protocoles de segmentation manuelle et les images ont subi des prétraitements et des recalages affines différents. De cette façon, nous montrons dans la suite la robustesse d'OPAL aux différentes définitions de l'hippocampe et des procédés de pré-traitement. Sur ces deux bases, les structures des hippocampes sont contenues dans une zone d'environ 90×60×60 voxels.

Métriques d'évaluation

Puisque notre méthode ne nécessite aucune étape d'apprentissage, nous effectuons une validation croisée de type *leave-one-out*, c'est-à-dire que chaque sujet de la base est considéré à tour de rôle comme un sujet à traiter, et est segmenté par les sujets restants qui définissent la bibliothèque. La précision de segmentation est estimée avec le coefficient Dice standard, également appelé index kappa, introduit dans Zijdenbos et al. (1994), qui permet de comparer les segmentations manuelles des experts aux segmentations automatiques. Pour deux segmentations binaires $\mathcal{S}_1$ et $\mathcal{S}_2$, le coefficient Dice est calculé comme suit :

$$\mathrm{Dice}(\mathcal{S}_1, \mathcal{S}_2) = \frac{2 \mid \mathcal{S}_1 \cap \mathcal{S}_2 \mid}{\mid \mathcal{S}_1 \mid + \mid \mathcal{S}_2 \mid}. \tag{3.1}$$

Pour chaque sujet, les coefficients Dice des hippocampes droit et gauche sont moyennés. Les valeurs dans les Tableaux 3.1 et 3.2 correspondent au Dice médian sur toutes les moyennes des coefficients Dice des sujets de la base.

Ainsi, pour valider la supériorité d'une méthode par rapport à une autre, nous proposons également d'utiliser le test de Student standard sur les ensembles des coefficients Dice de chaque sujet. Ce test est notamment utilisé pour comparer deux échantillons, et la valeur-p résultant du test permet d'estimer si l'un présente des valeurs statistiquement supérieures à celles de l'autre. Des valeurs-p inférieures à 0.05 permettent généralement de valider l'hypothèse de supériorité d'un échantillon et donc de comparer les performances de différentes méthodes de segmentation.

Enfin, les temps de calcul donnés par OPAL incluent la recherche de correspondances sur toutes les échelles et descripteurs considérés, la fusion des étiquettes pour chaque estimateur, l'agrégation des estimateurs et la segmentation finale des deux hippocampes.

Durant notre processus de validation, nous étudions l'impact de paramètres tels que la fenêtre d'initialisation des correspondances, la taille de patch, le nombre de voisins (k), et l'impact des différentes configurations multi-échelles et multi-descripteurs sur la précision de segmentation et le temps de calcul. Les résultats obtenus par OPAL sont ensuite comparés à ceux de méthodes de l'état-de-l'art appliquées aux mêmes bases.

Réglage des paramètres

OPAL a été implémentée sous MATLAB en utilisant du code C-MEX multi-processus. Nos expériences sont menées sur un serveur de 16 coeurs à $2,6$ GHz avec 100 GO de mémoire. Les paramètres par défaut sont fixés pour traiter à la fois les bases ICBM et EADC-ADNI. Ces paramètres offrent un bon compromis entre précision de segmentation et temps de calcul. Dans les résultats suivants, le processus de recherche d'ANN (OPM) est effectué en 3 itérations. Dans (2.4), les paramètres α et σ sont empiriquement fixés à 2. Pour la configuration multi-descripteurs, les estimations sont calculées depuis les intensités de l'image IRM initiale et de la norme du gradient des intensités. Pour la configuration multi-échelles, OPAL est effectuée avec des patchs de taille $3\times3\times3$ et $5\times5\times5$ voxels. Enfin, le nombre de correspondances sélectionnées par voxel pour chaque estimateur est par défaut fixé à $k = 10$ ANN, et la taille de la fenêtre d'initialisation à $13\times13\times13$ voxels.

3.2.2 Influence des paramètres

Tout d'abord, comme mentionné dans la section 2.2.2, la fenêtre d'initialisation permet d'assurer une cohérence spatiale entre les voxels dans le sujet S et leurs correspondances dans la bibliothèque de modèles T. En réglant la taille de la fenêtre de façon optimale, l'algorithme de recherche converge plus rapidement puisque les premières correspondances se trouvent dans les zones d'intérêt. Pour un nombre fixé d'itérations, ce paramètre permet donc d'améliorer les performances de segmentation. Son réglage optimal dépend de la nature des images et de la base considérée.

La Figure 3.1 montre les coefficients Dice pour plusieurs tailles de fenêtre sur les deux bases étudiées. Pour ICBM, un plateau est atteint pour une taille de fenêtre de $7\times7\times7$ voxels, quand une zone de $13\times13\times13$ voxels offrent les meilleures performances pour la base EADC-ADNI. Cette deuxième base requiert en effet une zone de recherche plus large à l'initialisation puisqu'elle contient une variabilité anatomique plus élevée du fait de la présence de sujets adultes atteints de pathologies, avec des structures d'hippocampe plus ou moins atrophiées. Nous fixons la taille de la zone de recherche à l'initialisation par défaut à $13\times13\times13$ voxels.

La Figure 3.2 montre l'influence du nombre d'ANN (k) et de la taille de patch sur les performances de segmentation et le temps de calcul. Sans l'approche multi-échelles, les patchs de taille $5\times5\times5$ voxels fournissent les meilleurs résultats pour les deux bases, avec des coefficients Dice de 89.8% et 89.7% respectivement sur ICBM et EADC-ADNI pour $k = 10$. Cette taille de patch est en effet adaptée pour décrire les structures d'images IRM standards, comme observé dans Coupé et al. (2011); Tong et al. (2013). Avec l'approche proposée multi-échelles, nous pouvons cependant tirer profit de l'information apportée par d'autres tailles de patch qui peuvent localement

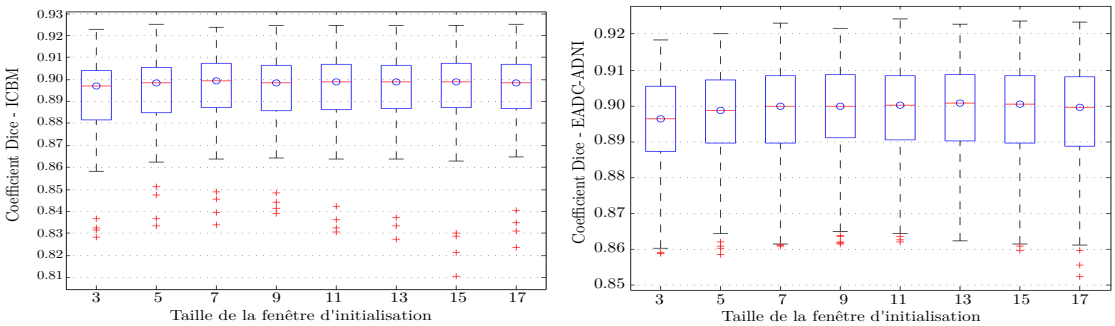


FIGURE 3.1 – Influence de la taille de la fenêtre de recherche à l'initialisation sur le coefficient Dice pour les bases ICBM (à gauche) et EADC-ADNI (à droite).

être mieux adaptées au contenu de l'image. En fusionnant les estimations générées depuis des patchs de taille 3×3×3 et 5×5×5 voxels, nous obtenons un coefficient Dice de 89.9% pour la base ICBM, avec les paramètres par défaut. En ajoutant les estimations obtenues avec des patchs de taille 7×7×7 et en augmentant le nombre k d'ANN, nous obtenons un coefficient Dice de 90.1%. Pour la base EADC-ADNI, nous atteignons un coefficient Dice de 90.1% pour $k = 20$ (90.05% avec les paramètres par défaut). Ces résultats soulignent l'importance de prendre en compte la diversité d'information obtenue depuis des tailles de patch différentes.

Nous notons que le coefficient Dice médian atteint un plateau autour de 10 ANN. Il est intéressant de remarquer que ce nombre est cohérent avec le nombre d'images d'exemples recommandé dans les méthodes de correspondance multi-modèles (Collins et Pruessner, 2010). Comme attendu, de plus grands patchs et un nombre plus élevé d'ANN requièrent des temps de calcul plus élevés. D'après nos expériences, l'utilisation de $k = 10$ ANN sur chaque descripteur semble offrir un bon compromis entre précision de segmentation et temps de calcul. Enfin notons que pour les deux bases, l'étape de segmentation est effectuée en moins de 2s par sujet.

Les résultats dans le Tableau 3.1 présentent l'impact de chaque contribution sur les coefficients Dice et le temps de calcul, avec les valeurs-p correspondant à la comparaison avec la configuration multi-échelles et multi-descripteurs. Pour les deux bases, l'utilisation de plusieurs échelles et de plusieurs descripteurs améliore significativement la précision de segmentation, comme le montrent les valeurs-p inférieures à 10^{-14}. De plus, dans les cas étudiés, l'approche complète multi-échelles et multi-descripteurs améliore les résultats par rapport à la configuration mono-échelle, multi-descripteurs, ce qui démontre la complémentarité de notre stratégie de fusion a posteriori des estimateurs.

Des estimations pour les différents descripteurs et échelles (tailles de patch) sont illustrées dans la Figure 3.3. On s'aperçoit d'abord que des tailles de patch plus grandes produisent des cartes plus lisses. Des tailles de patch plus réduites sont en

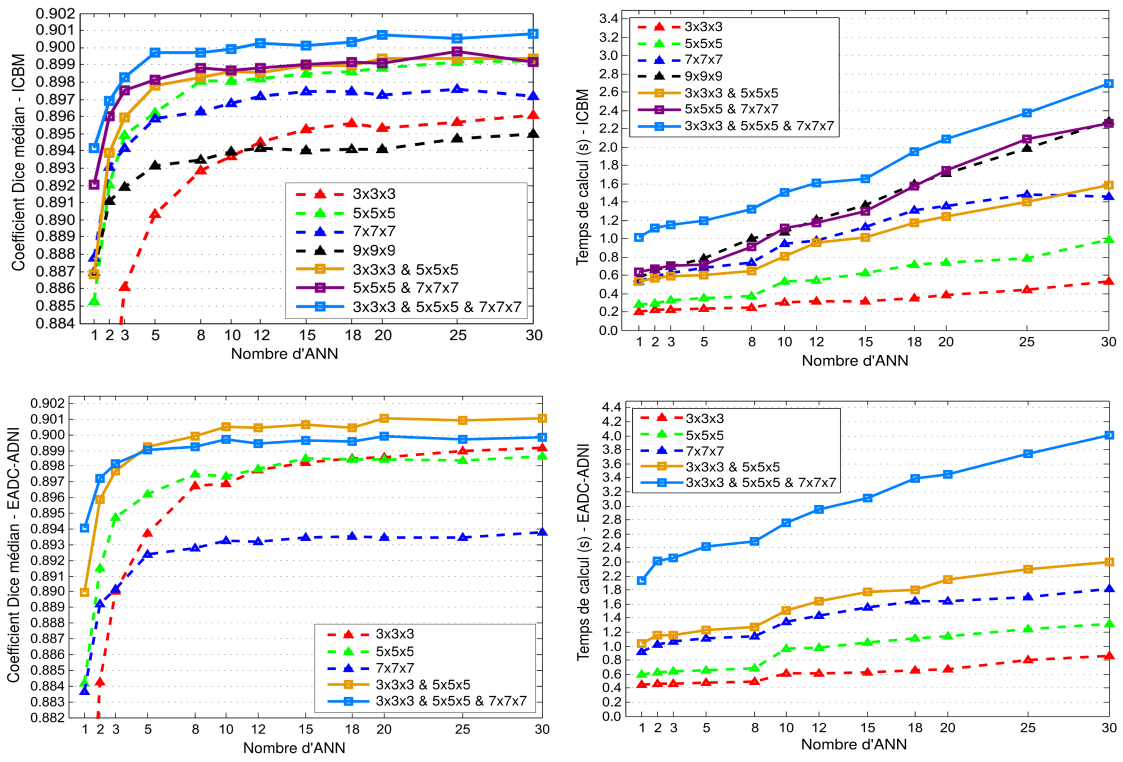


FIGURE 3.2 – Coefficient Dice médian pour différentes configurations mono et multi-échelles selon le nombre k de voisins (à gauche), avec le temps de calcul correspondant (à droite) sur la base de données ICBM (en haut) et EADC-ADNI (en bas). Ces résultats sont obtenus avec la configuration multi-descripteurs, c'est-à-dire, en utilisant la norme du gradient des intensités en plus des intensités de l'image IRM initiale.

revanche capable de mieux capturer les détails fins, au détriment d'une carte d'estimation de segmentation plus bruitée. Ensuite, les estimateurs basés sur la norme du gradient des intensités permettent de mieux capturer la surface de la structure, mais sont également moins réguliers à l'intérieur du volume. Enfin, l'agrégation de tous les estimateurs est capable de tirer profit des différentes configurations pour segmenter la structure de façon précise.

La Figure 3.4 présente les résultats de segmentation du meilleur, médian et pire sujets obtenus sur la base EADC-ADNI. D'abord, nous pouvons constater que les segmentations automatiques sont plus lisses que les segmentations manuelles. La fusion basée sur les patchs patchs des étiquettes permet d'obtenir une segmentation cohérente à la surface du volume de l'hippocampe, mais tend à remplir les trous présents dans les segmentations manuelles. Certains de ces trous semblent correspondre au fluide cérébrospinal quand d'autres sont dus à des erreurs de segmentation des experts.

Tableau 3.1 – Influence des configurations multi-échelles et multi-descripteurs en termes de précision de segmentation et de temps de calcul sur les bases ICBM et EADC-ADNI. Les résultats mono-échelle et mono-descripteur correspondent à une recherche d'ANN décrits par des patchs de taille 5×5×5 voxels sur l'image IRM initiale. Le résultat multi-descripteurs correspond à l'ajout de la norme du gradient des intensités de l'image. Le résultat multi-échelles correspond à l'ajout des estimations pour les deux modalités avec des patchs de taille 3×3×3 voxels. Les valeurs-p correspondent à la comparaison avec les résultats multi-descripteurs et multi-échelles. Enfin, les temps de calcul correspondent au temps de traitement moyen d'un sujet.

Base	Multi-descripteurs	Multi-échelles	Dice médian	Dice moyen	Valeur-p	Temps de calcul
ICBM	✗	✗	89.4%	$89.1 \pm 1.85\%$	$< 10^{-14}$	$0.27s$
	✓	✗	89.8%	$89.6 \pm 1.68\%$	0.0131	$0.53s$
	✓	✓	89.9%	$89.7 \pm 1.70\%$	×	$0.92s$
EADC-ADNI	✗	✗	89.4%	$89.2 \pm 1.55\%$	$< 10^{-25}$	$0.49s$
	✓	✗	89.7%	$89.6 \pm 1.45\%$	$< 10^{-8}$	$0.95s$
	✓	✓	90.1%	$89.8 \pm 1.46\%$	×	$1.51s$

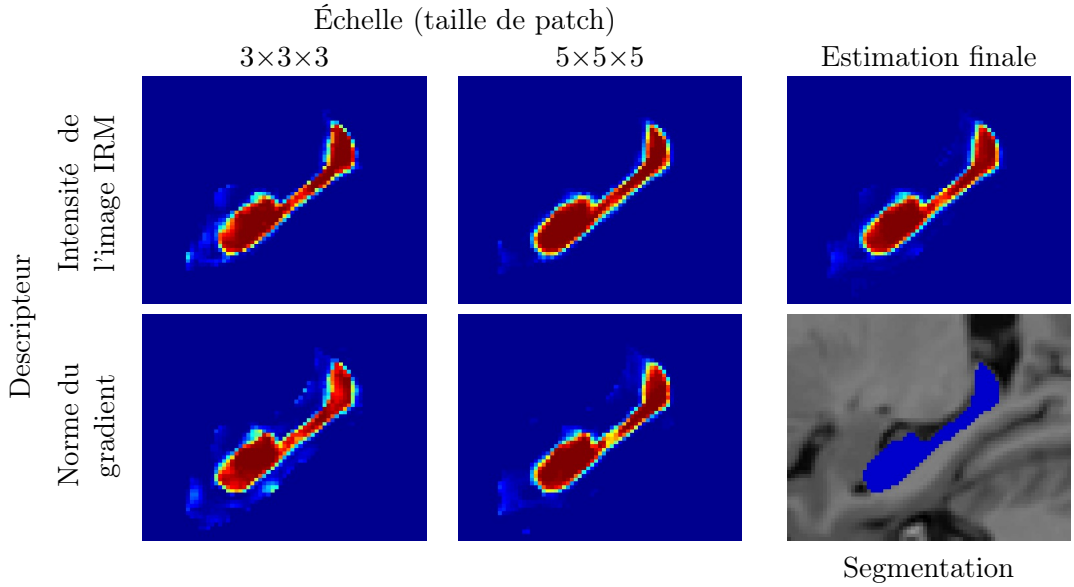


FIGURE 3.3 – Visualisation 2D des estimations de segmentation calculées depuis (2.3) pour les différents descripteurs et tailles de patchs sur une image de la base EADC-ADNI. Avec des patchs de taille 5×5×5 voxels, l'estimation semble plus stable dans le volume de l'hippocampe. Avec des patchs de taille 3×3×3 voxels, certaines régions sont segmentées plus précisément, par exemple la pointe haute de l'hippocampe.

3.2.3 Comparaison aux méthodes de l'état-de-l'art

Nous comparons à présent la méthode proposée OPAL aux méthodes de l'état-de-l'art appliquées aux mêmes bases. Pour la base ICBM, nous considérons la méthode initiale de fusion d'étiquettes basée sur les patchs PBL (Coupé et al., 2011), la méthode multi-modèles MTM (pour Multi-Template Method, Collins et Pruessner

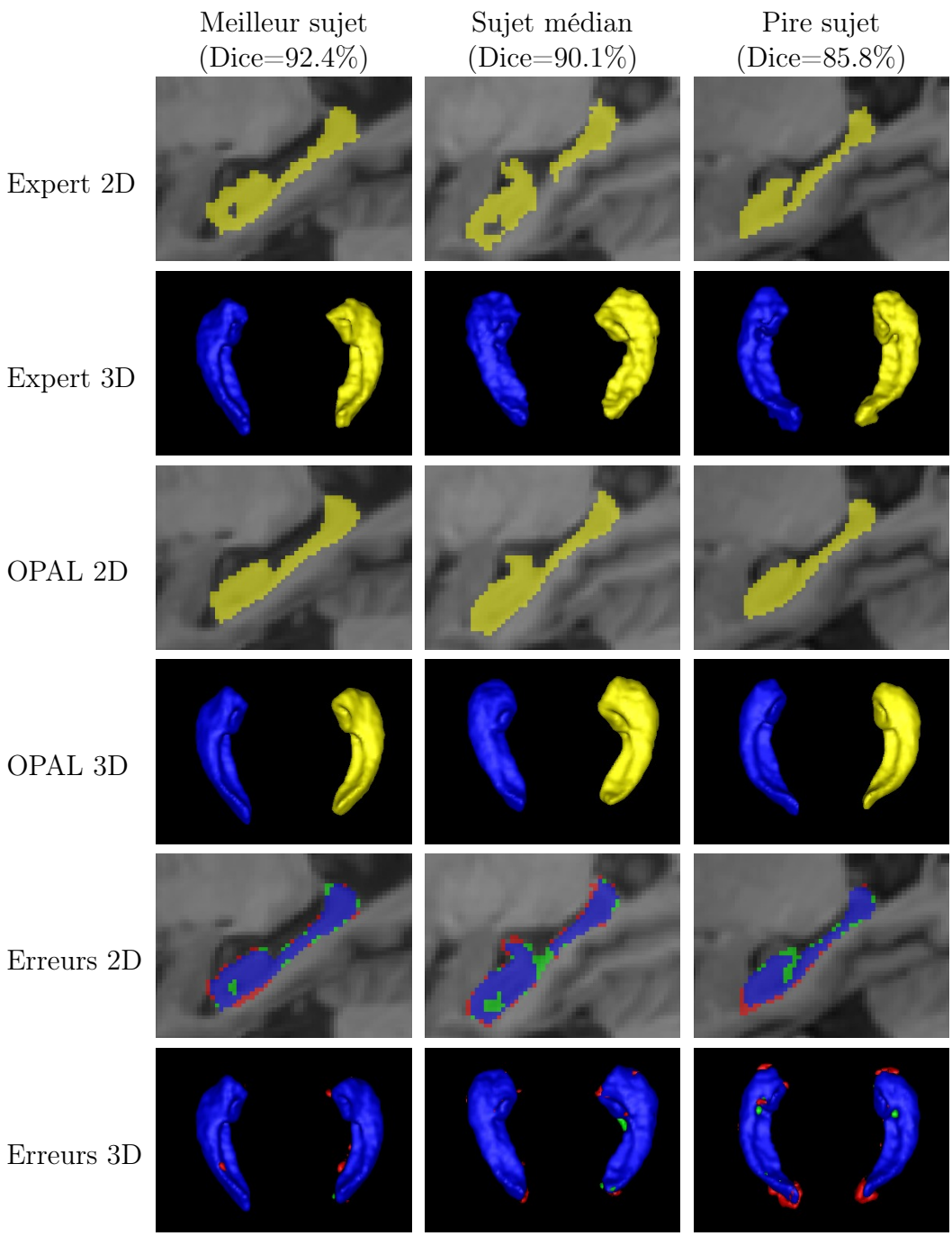


FIGURE 3.4 – Visualisations 2D et 3D des meilleur, médian et pire sujets EADC-ADNI segmentés avec OPAL. Aux cinquième et sixième lignes, les voxels bleus correspondent à la superposition avec la segmentation de l'expert, les verts aux faux positifs (segmentés par OPAL mais pas par l'expert), et les rouges aux faux négatifs (segmentés par l'expert mais pas par OPAL).

(2010)), une méthode de classification parcimonieuse SC (pour Sparse Coding, Tong et al. (2013)), et une méthode d'apprentissage de dictionnaire DDLS (pour Discriminative Dictionary Learning for Segmentation, Tong et al. (2013)). Les coefficients Dice moyens obtenus sur EADC-ADNI sont comparés aux résultats obtenus par une approche basée sur une forêt d'arbres décisionnels RF (pour Random Forest, Tangaro et al. (2014)), par deux approches multi-modèles, BMAS (pour BioClinica Multi-Atlas Segmentation, Roche et al. (2014)) et LEAP (pour Learning Embeddings for Atlas Propagation Gray et al. (2014)), par une méthode multi-échelles basée sur les patchs, notée PB-MS (pour Patch-Based Multi-Scale, Pant et al. (2015)), et par une méthode qui propose l'apprentissage d'une métrique optimisée pour la fusion d'étiquettes basée sur les patchs NLW-ML (pour Non Local patch-based Weighted voting with Metric Learning Zhu et al. (2017)).

Les performances obtenues par OPAL sur les bases ICBM et EADC-ADNI sont respectivement comparées à celles d'autres méthodes dans les Tableaux 3.2 et 3.3. Les valeurs présentées sont les résultats publiés par les auteurs, par conséquent des méthodes différentes sont comparées sur les deux bases. Les temps de calcul fournis correspondent seulement à l'étape de segmentation, et n'incluent pas les éventuelles étapes de présélection des modèles, dont OPAL n'a pas besoin. Par conséquent, les temps de calcul sont potentiellement sous-estimés à l'exception de celui d'OPAL.

Sur la base ICBM, par rapport à la méthode originale basée sur les patchs PBL (Coupé et al., 2011), OPAL améliore la précision de segmentation de $1,7\%$ tout en étant 700 fois plus rapide. Par rapport à la méthode de l'état-de-l'art la plus précise, basée sur l'apprentissage du dictionnaire DDLS (Tong et al., 2013), OPAL obtient un coefficient Dice médian plus élevé, pour un temps de calcul 1000 fois plus rapide, et avec une valeur p inférieure à 10^{-12} obtenue à partir d'un test de Student sur les ensembles des coefficients Dice de chaque sujet segmenté avec OPAL et DDLS. De plus, pour un coefficient Dice donné de 89.0%, équivalent à la précision de la méthode DDLS, OPAL nécessite moins de 0.22s de traitement sur la base ICBM.

Sur la base EADC-ADNI, les résultats d'OPAL sont comparés à ceux des autres méthodes uniquement en termes de précision de segmentation, puisque les temps de calcul ne sont pas fournis par les auteurs dans leurs publications hormis pour la méthode de (Pant et al., 2015), qui indique un temps de calcul de près de 5 minutes par sujet. Les résultats présentés avec OPAL sur la base EADC-ADNI dans le Tableau 3.3 sont obtenus en 1.51s de traitement par sujet. Cependant, on peut présumer que les méthodes BMAS (Roche et al., 2014) et LEAP (Gray et al., 2014) sont susceptibles d'avoir des temps de calcul comparables à ceux de la méthode multi-modèles MTM (Collins et Pruessner, 2010), puisqu'elles sont également basées sur des approches de déformation multi-modèles. On peut noter que ce type de méthode obtient de moins bonnes performances sur la base EADC-ADNI que sur ICBM. Cela peut être expliqué par la plus grande variabilité anatomique de la base, due à la présence de sujets atteints par des pathologies comme la maladie d'Alzheimer.

Tableau 3.2 – Comparaison des méthodes en termes de coefficient Dice sur la base ICBM. Les temps de calcul sur des machines similaires sont reportés, et nous donnons le facteur d'accélération par rapport au temps d'OPAL.

Méthode sur ICBM	Dice médian	Intervalle à 95%	Temps de calcul
Basé patchs (PBL)	$88.2 \pm 2.19\%$	$[87.7; 88.7]\%$	$662s$ ($\times 700$)
Multi-sujets (MTM)	$88.6 \pm 2.05\%$	$[88.2; 89.0]\%$	$3974s$ ($\times 4300$)
Codage éparse (SC)	$88.7 \pm 1.94\%$	$[88.3; 89.2]\%$	$5587s$ ($\times 6000$)
Dictionnaire (DDLS)	$89.0 \pm 1.90\%$	$[88.5; 89.4]\%$	$943s$ ($\times 1000$)
OPAL	$\mathbf{89.9 \pm 1.70}\%$	$[\mathbf{89.6}; \mathbf{90.3}]\%$	**0.92s**

Tableau 3.3 – Comparaison des méthodes en termes de coefficient Dice sur la base EADC-ADNI. Puisqu'aucune des publications de l'état-de-l'art ne mentionne les temps de calcul, la comparaison porte uniquement sur la précision de segmentation. Le résultat sélectionné pour OPAL a été obtenu en 1.51s de traitement par sujet.

Méthode sur EADC-ADNI	Dice moyen	Intervalle à 95%
Forêt d'arbres décisionnels (RF)	$76.0 \pm 7.00\%$	$[74.6; 77.4]\%$
Multi-sujets (BMAS)	$86.6 \pm 1.70\%$	$[86.3; 86.9]\%$
Multi-sujets (LEAP)	$87.6 \pm 2.07\%$	$[87.1; 88.0]\%$
Basé patchs (NLW-ML)	$88.3 \pm 2.50\%$	$[87.8; 88.8]\%$
Multi-échelles basé patchs (PB-MS)	$89.2 \pm 2.22\%$	$[88.8; 89.6]\%$
OPAL	$\mathbf{89.8 \pm 1.46}\%$	$[\mathbf{89.5}; \mathbf{90.1}]\%$

Dans toutes les configurations, OPAL produit les meilleurs résultats de segmentation avec un coefficient Dice moyen de 89.8% et un Dice médian de 90.1%. OPAL surpasse donc des méthodes récentes de l'état-de-l'art. Par exemple, comparée à une approche basée sur les forêts d'arbres décisionnels (Tangaro et al., 2014), OPAL améliore les résultats de 13.8%, et de 1.5% par rapport à une méthode récente basée sur l'apprentissage de métriques (Zhu et al., 2017). Les valeurs-p obtenues depuis un test de Student sont inférieures à 10^{-25} et valident les meilleures performances d'OPAL par rapport aux méthodes de l'état-de-l'art sur la base EADC-ADNI.

3.2.4 Ajout de segmentations automatiques à la bibliothèque

En raison du coût des segmentations manuelles, peu de sujets avec leur segmentation sont disponibles comme modèles. De plus, ces sujets segmentés manuellement sont sujets à de fortes variabilités inter et intra-expert (Frisoni et al., 2015). De nombreux travaux ont proposé d'augmenter la taille de la bibliothèque en utilisant des segmentations automatiques comme modèles pour segmenter de nouveaux su-

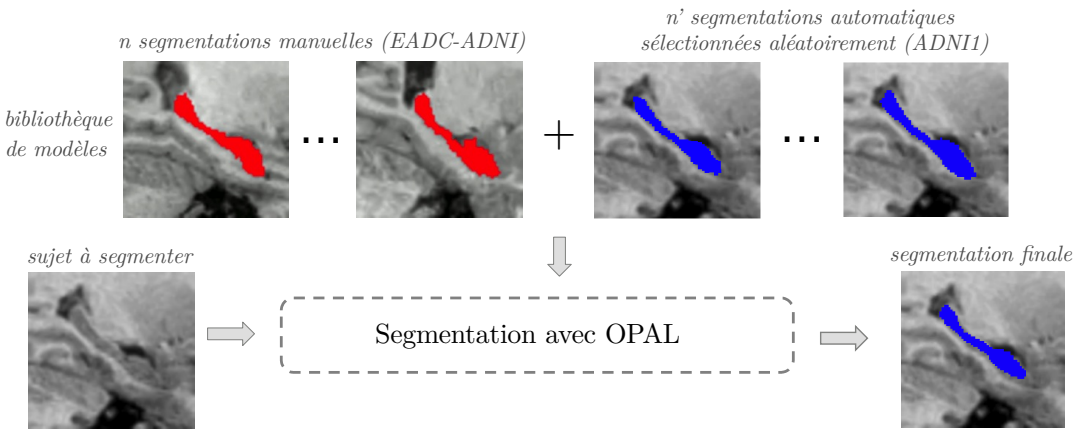


FIGURE 3.5 – Ajout de nouveaux sujets segmentés automatiquement à la bibliothèque d'exemples. Les segmentations automatiques des sujets fournis sans segmentation manuelle peuvent être ajoutées à la bibliothèque pour augmenter sa diversité. Les prochaines segmentations peuvent alors bénéficier de l'information contenue dans ces sujets supplémentaires.

jets. Les segmentations automatiques sont alors considérées avec les segmentations manuelles au sein de la bibliothèque (Eskildsen et al., 2012).

Dans Pipitone et al. (2014), une approche a été proposée pour propager des segmentations dans une bibliothèque de sujets en utilisant des recalages non linéaires. De façon similaire, la méthode LEAP (Gray et al., 2014) propose de propager les étiquettes à des sujets sans segmentation en segmentant itérativement les sujets les plus proches en termes d'entropie conjointe. Ces approches améliorent la précision de segmentation, en augmentant la diversité de la bibliothèque dans laquelle sont recherchées les correspondances. Notons toutefois que les segmentations automatiques doivent être de précision suffisante pour renforcer la pertinence de la bibliothèque, et que si l'on ajoute beaucoup plus de sujets, l'algorithme de recherche de correspondances peut nécessiter plus d'itérations pour converger.

Comme mentionné dans la section 2.2.1, le temps de calcul et la complexité de la méthode OPAL ne dépendent que de la taille du sujet à segmenter. Cette propriété permet d'étendre la taille de la bibliothèque sans impacter la complexité de l'algorithme. Par conséquent, comme illustré dans la Figure 3.5, nous proposons une expérience où des sujets issus de la base ADNI1 (Wyman et al., 2013) sont aléatoirement sélectionnés et segmentés par OPAL avant d'être ajoutés à la bibliothèque avec leur segmentation automatique. Les performances sont exprimées avec le coefficient Dice et toujours calculées avec une procédure de validation croisée sur les sujets EADC-ADNI fournis avec leur segmentation d'expert.

La Figure 3.6 montre l'impact de l'augmentation de la taille de la bibliothèque sur la précision et le temps de calcul. Ajouter de nouveaux sujets à la bibliothèque avec des segmentations automatiques utilisées comme modèles permet d'améliorer la précision de la segmentation. En effet, puisque la bibliothèque est étendue avec de nouveaux sujets, sa diversité est augmentée, et des correspondances plus pertinentes

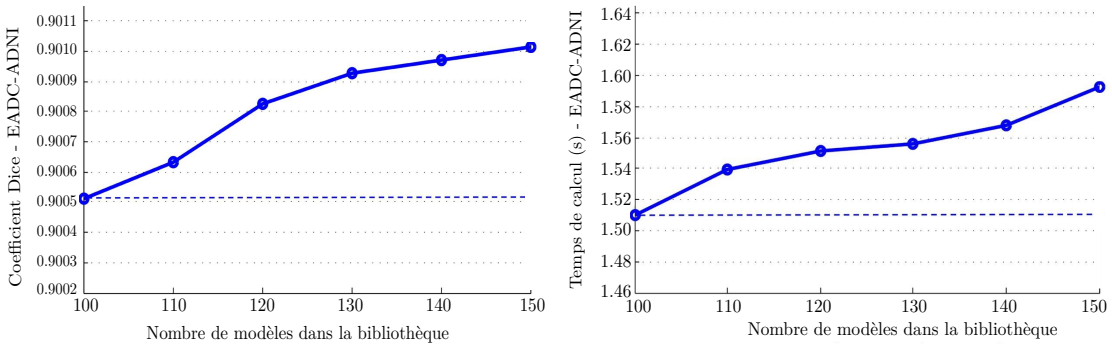


FIGURE 3.6 – Influence de l'ajout de sujets ADNI segmentés automatiquement à la bibliothèque de sujets d'entraînement EADC-ADNI. La précision de segmentation (coefficient Dice) est donnée à gauche, et le temps de calcul correspondant à droite. Les résultats obtenus avec 100 sujets (ligne pointillée) correspondent aux résultats du Tableau 3.1.

peuvent être trouvées. Augmenter le nombre de modèles améliore la précision au détriment d'une très faible augmentation du temps de calcul, due au stockage en mémoire des données. Avec 50% de sujets en plus, le temps de calcul d'OPAL est seulement augmenté de 6%.

3.2.5 Transfert

Dans cette section, nous avons appliqué avec succès notre méthode à la segmentation de l'hippocampe, la structure la plus étudiée dans le cadre de la maladie d'Alzheimer. Notre méthode obtient des résultats qui surpassent ceux des méthodes de l'état-de-l'art en seulement quelques secondes de traitement. Avec de telles performances en termes de temps de calcul et de précision de segmentation, OPAL ouvre la voie à de nouvelles applications du processus de segmentation automatique depuis la fusion d'étiquettes de modèles. L'intégration dans un logiciel de visualisation par exemple, faciliterait grandement l'analyse d'images IRM.

La méthode OPAL a été notamment incluse dans volBrain [3] un système en ligne d'étude volumétrique d'images IRM cérébraux décrit dans Manjón et Coupé (2016). Cette plateforme permet de télécharger une image d'IRM cérébral de modalité T1, et fournit en quelques minutes une analyse complète du sujet avec notamment la segmentation de structures anatomiques telles que l'hippocampe, les amygdales, ou le cervelet. Depuis mars 2015, volBrain a servi à plus de 1400 utilisateurs à travers le monde pour traiter automatiquement plus de 45000 images IRM.

La méthode OPAL peut être appliquée à la segmentation de n'importe quelle structure anatomique. Dans la section suivante, la méthode est appliquée efficacement à la segmentation des sous-structures du cervelet.

3. http://volbrain.upv.es

3.3 Segmentation du cervelet

3.3.1 Contexte

D'autres structures d'intérêt, comme le cervelet, interviennent dans le contexte des maladies neurodégénératives. Des patients atteints de schizophrénie ou de la maladie d'Alzheimer présentent des atrophies de certaines sous-structures du cervelet (Okugawa et al., 2002; Thomann et al., 2008). Cette structure joue un rôle important dans la coordination des mouvements et est liée aux capacités d'attention et de gestion des émotions. Les problématiques d'étude et de segmentation de cette structure sont donc les mêmes que pour l'hippocampe.

L'anatomie du cervelet est divisée en deux hémisphères remplis de matière blanche et entourés d'une couche de matière grise créant des plis, appelés lobules. Un exemple de segmentation manuelle du cervelet et de ses lobules est donné dans la Figure 3.7. La taille, la position et le nombre de lobules visibles sont très variables selon les sujets, ce qui rend la segmentation du cervelet particulièrement difficile. Dans cette section, OPAL est appliquée efficacement à la segmentation du cervelet, composé de plusieurs sous-structures mitoyennes, contrairement à l'hippocampe. La complexité de l'anatomie de cette structure peut mener à des segmentations irrégulières. Deux modifications de l'étape de fusion sont alors apportées à OPAL, afin d'obtenir de meilleures performances de segmentation, et cette nouvelle méthode de segmentation du cervelet est appelée CERES (pour CEREbellum Segmentation), et a également été intégrée à la plateforme volBrain (Manjón et Coupé, 2016).

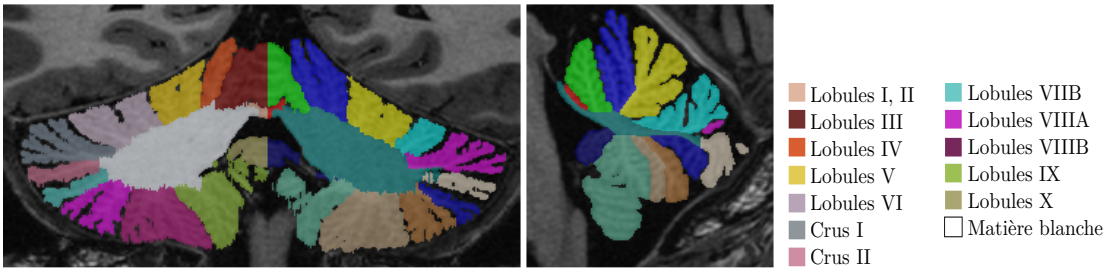


FIGURE 3.7 – Exemple de segmentation manuelle des sous-structures du cervelet. Les couleurs des étiquettes correspondent aux segmentations de l'hémisphère gauche.

Base de données

Le cervelet étant moins étudié à l'heure actuelle que des structures telles que l'hippocampe, et sa taille beaucoup plus grande rendant le processus de segmentation manuel plus fastidieux, il existe moins de bases de données du cervelet et de ses sous-structures avec leurs segmentations manuelles. Dans cette application, 5 images avec segmentations manuelles ont été utilisées pour construire la bibliothèque (Park et al., 2014), et les résultats présentés sont obtenus sur 20 images de tests fournies

par CoBrALab[4]. Plus de détails sur la construction de la base de données et du prétraitement sont disponibles dans Romero et al. (2017).

3.3.2 Amélioration de la fusion des estimateurs

Pondération

Une première modification de l'étape de fusion est la pondération des estimateurs de segmentation. Dans la méthode OPAL présentée dans le chapitre 2, toutes les cartes issues des différentes configurations multi-descripteurs et multi-échelles sont considérées équitablement. Dans le cas du cervelet, de nombreuses structures sont présentes avec une forte variabilité de formes et de tailles. Certains estimateurs s'avèrent nettement plus précis que d'autres pour certaines sous-structures et il est donc possible d'améliorer les performances en pondérant les cartes lors de l'étape d'agrégation finale. Le cervelet étant composé de plusieurs sous-structures adjacentes, l'étape de fusion calcule M cartes $\mathcal{F}^m$ avec $m \in \{1, \ldots, M\}$, correspondant aux M structures différentes. Ces cartes sont calculées depuis les contributions des N configurations différentes :

$$\mathcal{F}^m = \sum_{i=1}^{N} \mathcal{L}_i^m \alpha_i^m, \tag{3.2}$$

avec $\alpha_i^m \in [0, 1]$, les poids associés à chaque carte $\mathcal{L}_i^m$ issue du processus de fusion d'étiquettes (2.3), et tels que $\sum_{i=1}^{N} \alpha_i^m = 1$.

L'étiquette finale $\mathcal{S}(\mathbf{x_i})$ d'un voxel $\mathbf{x_i}$ est alors obtenue par :

$$\mathcal{S}(\mathbf{x_i}) = \underset{m}{\operatorname{argmax}}\ \mathcal{F}^m(\mathbf{x_i}). \tag{3.3}$$

Dans cette application, seuls deux estimateurs sont utilisés en considérant deux échelles de patchs de 3×3×3 et 9×9×9 voxels sur les intensités des images IRM. Les petits et grands patchs permettent de s'adapter à la grande variabilité des structures. Les paramètres α_i^m optimaux sont obtenus par une descente de gradient projetée, en considérant les résultats de segmentation des sujets de la base de test. En pratique, les paramètres α_i^m calculés atteignent des valeurs jusqu'à 0.82, ce qui revient à pondérer jusqu'à 4 fois plus, l'influence d'une des échelles.

Régularisation

Compte tenu de la grande variabilité des structures, même une étape de fusion basée sur les patchs peut entraîner des irrégularités, voire des trous dans la segmentation finale. Afin d'assurer une certaine régularité des structures segmentées automatiquement, une étape de régularisation des cartes peut être effectuée pour lisser les résultats de segmentation et obtenir de meilleures performances. Les cartes

4. http://cobralab.ca/atlases/Cerebellum.html

Tableau 3.4 – Impact de la pondération des estimateurs (adaptation des α_i^m) et de la régularisation sur les performances de CERES en termes de coefficient Dice moyen.

Structure	$\alpha_i^m = 1/N$ Sans régularisation	α_i^m adaptatif Sans régularisation	α_i^m adaptatif Avec régularisation
Lobule I-II	0.4226	0.4364	**0.5056**
Lobule III	0.7068	**0.7237**	0.7157
Lobule IV	0.7213	0.7281	**0.7320**
Lobule V	0.7514	0.7517	**0.7535**
Lobule VI	0.8667	0.8673	**0.8678**
Lobule Crus I	0.8970	0.9000	**0.9005**
Lobule Crus II	0.8024	0.8034	**0.8043**
Lobule VIIB	0.6742	0.6756	**0.6778**
Lobule VIIIA	0.7830	0.7848	**0.7862**
Lobule VIIIB	0.8499	0.8506	**0.8531**
Lobule IX	0.8849	0.8855	**0.8858**
Lobule X	0.7540	**0.7604**	0.7603
Cervelet	0.9358	0.9370	**0.9372**
Moyenne	0.7595	0.7640	**0.7702**

pour chaque étiquette de structure m sont alors régularisées par une convolution avec un noyau Gaussien H de taille 5×5×5 voxels, de la façon suivante :

$$\mathcal{F}_r^m = \mathcal{F}^m + \beta^m H * \mathcal{F}^m, \tag{3.4}$$

avec β^m un paramètre fixé empiriquement pour mélanger les cartes originale et convoluée afin d'éviter la disparition des très petites structures. Les cartes régularisées $\mathcal{F}_r^m$ sont alors agrégées par (3.2) pour générer l'estimation finale. À noter que dans le cas d'OPAL appliquée à la segmentation de l'hippocampe, la régularisation ne permet pas d'améliorer les performances, la structure étant déjà régulière, notamment car elle est segmentée depuis de nombreux modèles d'exemple.

L'impact de la pondération des estimateurs et de la régularisation sur les performances de segmentation de la méthode est montré dans le Tableau 3.4. L'optimisation de la fusion des estimateurs pour chaque étiquette permet d'améliorer les résultats de segmentation pour toutes les structures. La régularisation permet également d'améliorer globalement les performances, même si de très légères dégradations sont notables pour les lobules III et X.

3.3.3 Comparaison à l'état-de-l'art

La méthode CERES est comparée aux méthodes de l'état-de-l'art MAGeT (Chakravarty et al., 2013; Park et al., 2014) et RASCAL (Weier et al., 2014) sur la base utilisée dans Park et al. (2014). Les résultats de segmentation sur chaque structure et le cervelet complet pour ces deux méthodes et CERES sont donnés dans le Tableau 3.5. La fiabilité intra-expert est également donnée. Compte tenu de la plus grande variabilité des structures, l'écart avec les méthodes automatiques est plus important que pour la segmentation de l'hippocampe. CERES surpasse les méthodes

Tableau 3.5 – Coefficient Dice moyen et écart-type pour les méthodes MAGeT (Park et al., 2014), RASCAL (Weier et al., 2014) et CERES, avec la fiabilité intra-expert sur la base considérée.

Structure	MAGeT	RASCAL	CERES	Intra-expert
Lobule I-II	0.3960 ± 0.1424	0.3260 ± 0.2178	**0.5201** ± 0.1555	0.639
Lobule III	0.6800 ± 0.1741	0.6379 ± 0.2165	**0.7213** ± 0.1572	0.751
Lobule IV	0.6980 ± 0.1440	0.6627 ± 0.1611	**0.7271** ± 0.1346	0.818
Lobule V	0.7320 ± 0.1398	0.6666 ± 0.1560	**0.7561** ± 0.1332	0.881
Lobule VI	**0.8710** ± 0.0359	0.7969 ± 0.0523	0.8695 ± 0.0316	0.912
Lobule Crus I	0.8870 ± 0.0257	0.8383 ± 0.0351	**0.9007** ± 0.0152	0.904
Lobule Crus II	0.7780 ± 0.0679	0.7340 ± 0.0667	**0.8096** ± 0.0569	0.900
Lobule VIIB	0.5990 ± 0.1487	0.5820 ± 0.1137	**0.6850** ± 0.1205	0.863
Lobule VIIIA	0.7300 ± 0.0934	0.6757 ± 0.1426	**0.7926** ± 0.0759	0.860
Lobule VIIIB	0.7970 ± 0.0607	0.7783 ± 0.0931	**0.8533** ± 0.0390	0.833
Lobule IX	0.8560 ± 0.0384	0.8460 ± 0.0545	**0.8849** ± 0.0327	0.874
Lobule X	0.7540 ± 0.0490	0.7237 ± 0.0680	**0.7548** ± 0.0469	0.760
Cervelet	0.9250 ± 0.0094	0.9349 ± 0.0089	**0.9377** ± 0.0090	0.941
Moyenne	0.7320 ± 0.0568	0.6890 ± 0.0524	**0.7729** ± 0.0427	0.833

de l'état-de-l'art sur la majorité des structures, et segmente le cervelet entier avec une précision de coefficient Dice de 0.9377 pour une fiabilité intra-expert de 0.941. Dans la Figure 3.8, sont comparés les résultats de segmentation de CERES et des méthodes de l'état-de-l'art sur le pire, médian et meilleur sujet de la base.

En plus des performances de segmentation supérieures à l'état-de-l'art, CERES offre un temps de calcul très limité. Pour la segmentation complète du cervelet, CERES ne requiert qu'environ 1 minute de traitement, quand, avec l'utilisation de recalages non linéaires, les méthodes MAGeT (Park et al., 2014) et RASCAL (Weier et al., 2014) prennent respectivement environ 2h et 4h de traitement par sujet.

3.4 Extension à la classification de la maladie d'Alzheimer

Dans le contexte des pathologies neurodégénératives, la prévention précoce est le principal enjeu, les dommages entraînés sur le cerveau étant très souvent irréparables. Cette détection de l'apparition de la maladie peut se faire via l'étude des structures du cerveau sur des images IRM. Le projet ADNI vise d'ailleurs cet objectif en effectuant le suivi de patients sur plusieurs années, et en proposant des scanners de la même personne à plusieurs mois d'intervalle, en indiquant pour chaque nouvel IRM l'état clinique du patient. L'hippocampe est l'une des structures atteintes le plus précocement dans l'évolution de la maladie d'Alzheimer, et plusieurs méthodes ont été proposées pour déterminer automatiquement, depuis l'étude de cette structure, l'état clinique d'un sujet. La segmentation de l'hippocampe est alors une première étape clé pour l'étude clinique.

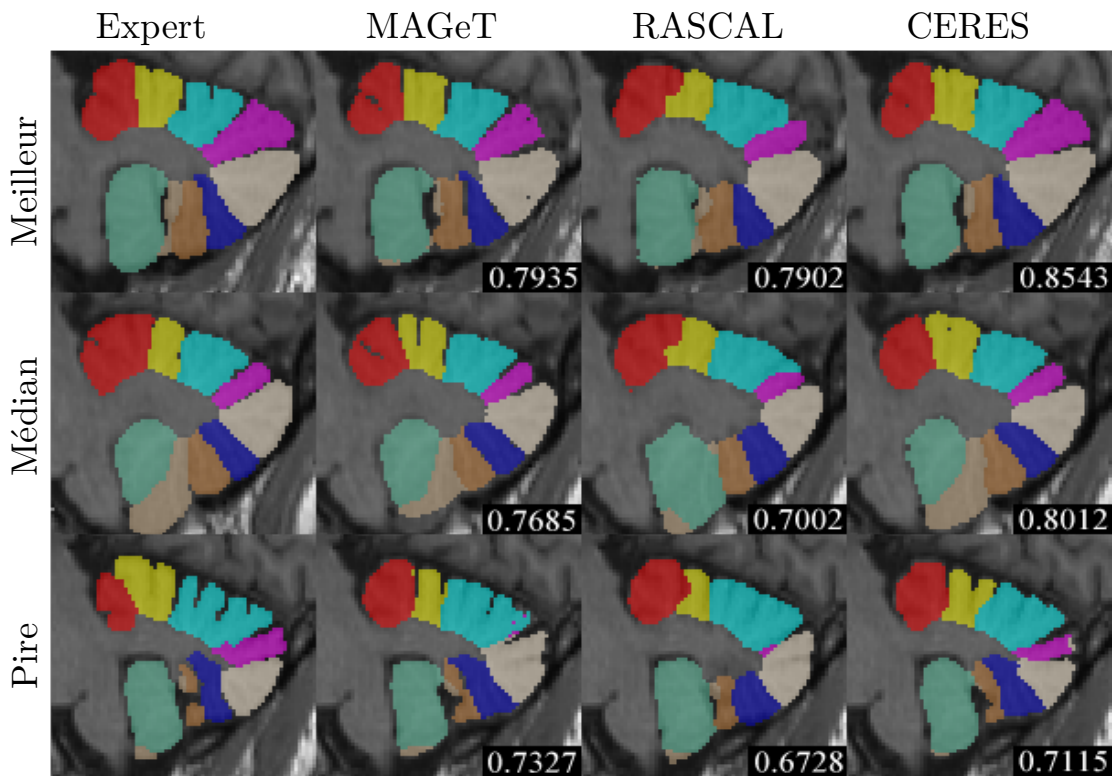
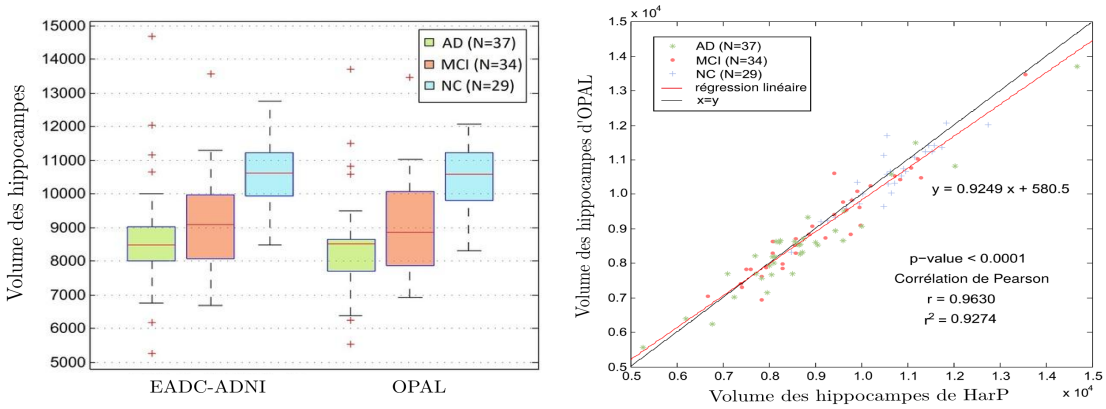


FIGURE 3.8 – Comparaison des méthodes de segmentation du cervelet sur le pire, médian et meilleur sujet de la base, avec les coefficients Dice correspondants.

Les bases telles que ADNI (Jack et al., 2008) sont divisées en trois catégories de populations : les sujets sains NC (pour Normal Control), les sujets présentant de légères déficiences cognitives MCI (pour Mild Cognitive Impairment), et les sujets malades AD (pour Alzheimer's Disease). L'enjeu est alors de pouvoir classifier automatiquement un sujet dans un de ces états, afin d'estimer une éventuelle apparition ou évolution de la maladie.

3.4.1 Classification depuis le volume des segmentations automatiques

Les états cliniques des modèles de la bibliothèque étant connus, une classification peut être effectuée directement grâce à l'étape de segmentation. On montre dans un premier temps que les volumes des segmentations générées par OPAL permettent de mieux séparer les trois groupes de population. Le volume des hippocampes est en effet un descripteur pertinent puisque la maladie d'Alzheimer entraîne la dégradation et l'atrophie de la structure. Ainsi, des sujets sains sont susceptibles d'avoir des hippocampes plus gros que ceux de patients malades.



(a) (b)

FIGURE 3.9 – Volumes des hippocampes et résultats de segmentation depuis le protocole de référence HarP d'EADC-ADNI et la méthode OPAL (a). Corrélation entre les volumes des hippocampes issus des segmentations de HarP et d'OPAL (b).

Nous proposons donc de comparer la précision de séparation des populations des trois groupes de la base EADC-ADNI, AD (N=37), MCI (N=34) et NC (N=29), en utilisant les segmentations manuelles issues du protocole HarP (Boccardi et al., 2014), et les segmentations générées par OPAL.

La Figure 3.9 montre la distribution pour chaque groupe des volumes des hippocampes et la corrélation entre les segmentation issues de HarP et d'OPAL. Le coefficient de corrélation de Pearson est de 0.9630, ce qui montre une forte corrélation entre les tailles des résultats de segmentation d'OPAL et ceux de HarP. Néanmoins, le Tableau 3.6 montre que les segmentations d'OPAL permettent de mieux séparer les groupes. En effet, les courbes ROC (pour Receiver Operating Characteristic), sont générées pour les différentes comparaisons de groupes de populations, et l'aire sous la courbe AUC (pour Area Under Curve), indique la capacité des estimateurs, ici la taille du volume des hippocampes, à séparer correctement les populations. OPAL obtient une plus haute AUC sur toutes les comparaisons de groupes, montrant la pertinence des volumes des hippocampes segmentés automatiquement.

Tableau 3.6 – Aire sous la courbe ROC (AUC) des volumes des hippocampes des segmentations depuis le protocole de référence HarP d'EADC-ADNI, et d'OPAL.

	EADC-ADNI HarP	OPAL
Volume moyen	9397 ± 1588	9272 ± 1525
AUC NC vs. AD	0.884	0.898
AUC NC vs. MCI	0.805	0.821
AUC MCI vs. AD	0.612	0.634

3.4.2 Classification depuis la fusion d'étiquettes de groupe

Le volume des hippocampes segmentés automatiquement peut être cependant limité. En effet, une forte atrophie de l'hippocampe témoigne de l'ancrage de la maladie, et son étude limite donc la possibilité de prévention. Des descripteurs robustes calculés sur différentes modalités d'IRM, ont été étudiés, notamment pour une analyse plus fine de l'évolution de la maladie lors du stade MCI.

On peut en effet distinguer également deux catégories parmi ces sujets : les sujets eMCI (pour early MCI), qui présentent de très légers troubles cognitifs, et les sujets lMCI (pour late MCI) dont les symptômes sont plus marqués, et qui risquent d'évoluer vers un stade AD.

Avec l'étude de différents descripteurs et modalités, des méthodes spécifiques de classification ont été proposées. Par exemple, Coupé et al. (2012a) proposent d'effectuer le même processus que celui de la segmentation par PBL (Coupé et al., 2011), mais en transférant cette fois une information pathologique. Les modèles NC et AD composent la bibliothèque, et la fusion d'étiquettes pour un voxel $\mathbf{x_i}$ est effectuée comme suit :

$$\mathcal{L}(\mathbf{x_i}) = \frac{\sum_{\mathbf{x_{j,t}} \in \mathcal{K}_i} \omega(\mathbf{x_i}, \mathbf{x_{j,t}}) l(\mathbf{x_{j,t}})}{\sum_{\mathbf{x_{j,t}} \in \mathcal{K}_i} \omega(\mathbf{x_i}, \mathbf{x_{j,t}})}, \tag{3.5}$$

avec ici le label transféré $l(\mathbf{x_{j,t}}) = \pm 1$, selon que le sujet de la base soit classé comme étant sain (+1 pour NC) ou malade (-1 pour AD).

OPAL peut alors naturellement être appliquée de la même façon en présentant le même intérêt de gain en temps de calcul par rapport à PBL que pour la segmentation. Une fois la recherche de correspondances et l'étape de fusion des étiquettes effectuée, la moyenne de l'estimateur est calculée sur toute la structure de l'hippocampe, segmentée à la volée. Le sujet est alors classé dans l'un des groupes selon cette valeur. L'architecture de la méthode de classification automatique utilisant OPAL pour la recherche de correspondances, et des exemples de fusion correspondant à chaque groupe de population sont respectivement illustrés dans les Figures 3.10 et 3.11.

Nous comparons alors les performances de classification d'un grand nombre de descripteurs standards, et plus complexes, issus des modalités T1 et d'IRM de diffusion DTI (pour Diffusion Tensor Imaging). De cette modalité DTI, peuvent être extraits les descripteurs suivants : l'anisotropie fractionnelle FA (pour Fractional Anisotropy), la diffusivité moyenne MD (pour Mean Diffusivity), la diffusivité axiale AxD (pour Axial Diffusivity), et la diffusivité radiale RD (pour Radial Diffusivity). Plus de détails sur ces différentes modalités et descripteurs sont disponibles dans Hett et al. (2016).

Dans le Tableau 3.7, sont données les performances de classification de ces différents descripteurs. Les résultats obtenus avec la méthode OPAL selon le pipeline de classification de la Figure 3.10, sont comparés avec ceux obtenus depuis la simple moyenne des descripteurs sur le volume de l'hippocampe préalablement segmenté.

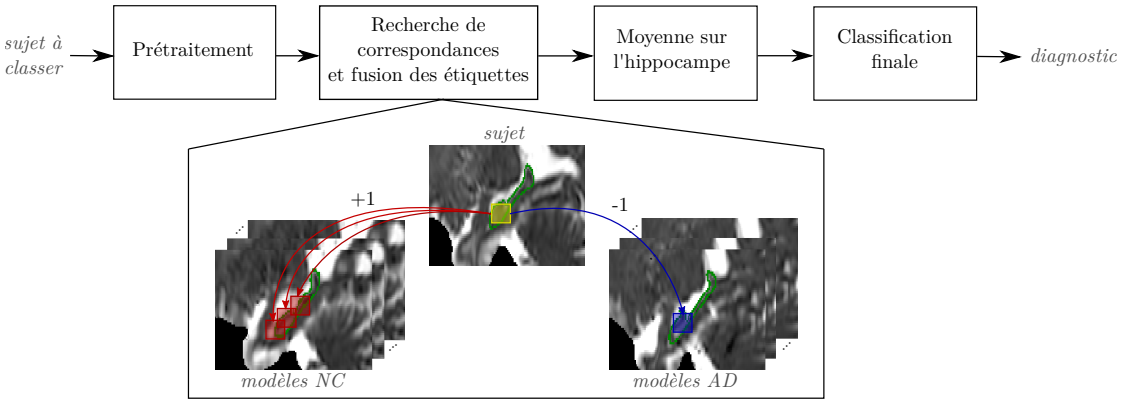


FIGURE 3.10 – Architecture de la méthode de classification automatique utilisant OPAL pour la recherche de correspondances. La fusion est effectuée sur les étiquettes de pathologie des modèles de la bibliothèque.

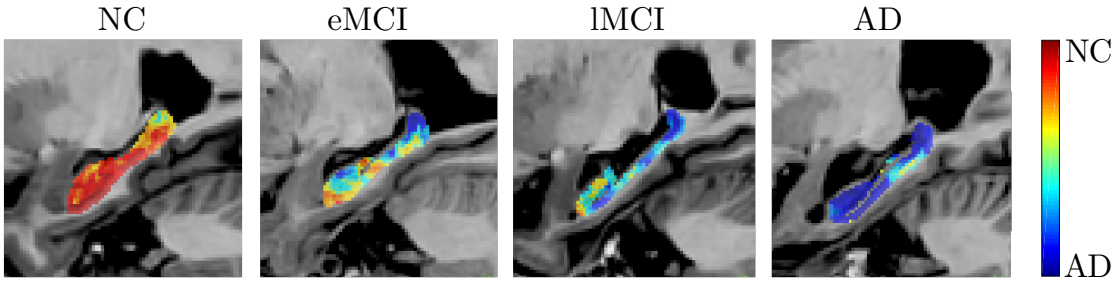


FIGURE 3.11 – Exemples de fusion des étiquettes de pathologie pour chaque groupe de population. La couleur bleu est associée aux modèles malades AD et indique la présence de pathologie, quand la couleur rouge est associée aux sujets sains.

Dans cette dernière configuration, la classification est donc effectuée depuis les valeurs moyennes des descripteurs de chaque sujet, par exemple en considérant uniquement le volume des segmentations. Une analyse discriminante linéaire LDA (pour Linear Discriminative Analysis) est utilisée pour classer chaque sujet et fournir les performances en termes d'AUC et de précision ACC (pour ACCuracy).

Le volume ne permet pas d'obtenir les meilleures performances de classification, et considérer les descripteurs moyens sur l'image s'avère également moins efficace que d'effectuer une fusion d'étiquettes de pathologies avec OPAL sur ces descripteurs. La méthode obtient des performances comparables à l'état-de-l'art. La méthode de Prasad et al. (2015) obtient pour la différenciation des groupes NC et AD, une précision de 78.5% contre des résultats allant de 80.1% jusqu'à 87.8% en utilisant la modalité T1 pour OPAL, et sur les groupes eMCI et lMCI, celle-ci obtient une précision de 63.4%, contre 67.6% pour OPAL en utilisant les descripteurs AxD issus du DTI. Selon les groupes de population, les meilleures performances de classification sont obtenues pour des modalités différentes. Étudier leur complémentarité serait donc une perspective intéressante (Hett et al., 2016).

Tableau 3.7 – Performances de classification (AUC/ACC) obtenues avec différents descripteurs, soit moyennés a priori sur l'hippocampe, soit utilisés avec OPAL selon le procédé illustré dans la Figure 3.10 (voir le texte pour plus de détails).

	Descripteurs	NC vs AD	NC vs MCI	AD vs MCI	eMCI vs lMCI
Moyenne	Volume	88.4/83.1	69.5/63.9	71.1/67.2	67.2/63.7
	FA	64.2/59.2	57.7/56.1	54.0/52.7	38.2/43.1
	MD	85.7/80.3	66.0/62.6	75.0/72.5	67.6/62.8
	AxD	83.5/81.4	63.5/58.0	74.3/70.2	68.9/66.8
	RD	86.2/79.2	66.5/62.3	74.8/70.5	66.0/61.5
OPAL	T1	**93.4/87.8**	**71.3/64.1**	82.0/73.4	68.7/66.2
	FA	85.0/80.1	63.5/60.1	74.9/70.3	63.0/60.7
	MD	90.6/86.5	68.8/60.7	80.4/76.3	70.4/65.8
	AxD	91.1/85.8	68.7/59.6	80.2/73.1	**71.8/67.6**
	RD	90.3/85.1	68.9/61.0	**80.0/76.5**	69.3/65.4

3.5 Conclusion

Dans ce chapitre, la méthode basée sur les patchs OPAL, qui permet de trouver rapidement de multiples correspondances dans une large base de sujets, sans étape de présélection ou d'apprentissage, a été appliquée à la segmentation de structures anatomiques sur des images d'IRM cérébral 3D. Grâce à la complexité réduite de notre algorithme et aux contraintes spatiales apportées, OPAL produit une segmentation en moins de 2 secondes, quand les méthodes de l'état-de-l'art basées sur des recalages non linéaires ou l'apprentissage de dictionnaires sont beaucoup plus coûteuses (jusqu'à plusieurs heures). Le faible temps de calcul d'OPAL a permis de définir une approche multi-échelles et multi-descripteurs augmentant significativement les performances de segmentation.

Les résultats d'OPAL dépassent ceux des méthodes de l'état-de-l'art sur la segmentation de l'hippocampe, jusqu'à atteindre la fiabilité inter-experts (coefficients Dice de 90% et de 89.0% respectivement pour les bases ICBM et EADC-ADNI). Ces coefficients représentent une barrière symbolique, et il serait difficile d'augmenter les performances de segmentation sans améliorer la qualité des bases de données en réduisant la variabilité des segmentations manuelles. La segmentation d'un seul hippocampe peut demander 30min à un expert (1h pour les deux) et OPAL produit donc une segmentation de qualité comparable en seulement 2s.

La méthode obtient également des résultats du niveau de l'état-de-l'art pour la segmentation des sous-structures du cervelet. Enfin, OPAL peut également être utilisée pour transférer non plus des étiquettes de structure, mais une information de pathologie, pour diagnostiquer la maladie d'Alzheimer en fonction des statistiques de correspondances trouvées dans la base de sujets.

Le faible temps de calcul d'OPAL offre un réel intérêt à la communauté d'analyse d'images médicales et facilite l'étude de nouvelles approches, comme celle proposée de fusion a posteriori d'estimateurs multi-échelles et multi-descripteurs.

La conception de la méthode OPAL et de l'algorithme de recherche PatchMatch optimisé se place dans le contexte de la segmentation d'images médicales recalées. Par conséquent, certaines étapes, comme l'initialisation et la recherche aléatoire contraintes sont proposées spécifiquement pour cette application. Néanmoins, d'autres applications peuvent bénéficier de ces adaptations, par exemple pour des images naturelles recalées, ou pour la recherche locale de correspondance dans le cadre du calcul de flot optique sur un flux vidéo. De plus, certaines améliorations telles que l'optimisation de l'étape de propagation, ou la recherche de multiples ANN sont indépendantes du contexte médical et peuvent améliorer les performances d'autres applications basées sur la recherche d'ANN.

Deuxième partie

Décomposition en superpixels réguliers

Résumé

Les superpixels ont progressivement été introduits depuis une vingtaine d'années afin de diminuer le coût de calcul et d'augmenter la précision d'applications d'analyse d'images, notamment de segmentation. De nombreuses méthodes et métriques ont été proposées pour décomposer une image en superpixels et évaluer la conformité de la décomposition à différentes propriétés comme l'homogénéité de couleur des pixels regroupés, le respect des contours des objets de l'image par les frontières des superpixels ou la régularité de forme et de taille des régions générées. Dans cette partie, nous présentons d'abord les méthodes de l'état-de-l'art existantes et étudions en détails les métriques de superpixels existantes. Nous démontrons la redondance ou les limites de certaines métriques de référence, afin de ne retenir que les critères d'évaluation pertinents.

Nous présentons ensuite une nouvelle méthode de décomposition nommée SCALP (pour Superpixels with Contour Adherence using Linear Path) qui tend à répondre aux limitations des approches existantes. L'utilisation de l'information de couleur et de contour des pixels le long du chemin linéaire entre le pixel et le barycentre spatial du superpixel considérés lors du processus de groupement permet de renforcer le respect des contours et la convexité des superpixels générés, tout en respectant l'homogénéité de couleur. La méthode proposée améliore donc conjointement les propriétés mentionnées précédemment tout en étant robuste au bruit, contrairement à la majorité des méthodes de l'état-de-l'art. SCALP surpasse également les méthodes récentes en offrant les meilleures performances de respect des objets de l'image et de détection de contours. Un gain en performances sur ces aspects est généralement obtenu en diminuant la régularité, mais SCALP réussit à surpasser les méthodes de référence tout en produisant des superpixels parmi les plus réguliers.

Concernant ce critère de régularité, nous démontrons qu'aucune métrique existante d'évaluation n'est satisfaisante. Nous proposons alors une nouvelle métrique de régularité globale basée sur l'évaluation conjointe de la régularité locale et la consistance de forme des superpixels. L'évaluation de cet aspect est d'ailleurs fondamental puisque nous démontrons sur plusieurs applications, que la régularité des superpixels peut impacter fortement les performances liées à leur utilisation. Puisque la plupart des méthodes de décomposition en superpixels tendent à permettre le réglage de la régularité. nous recommandons de comparer les méthodes sur les métriques de superpixels pour différents niveaux de régularité, évalués avec la métrique proposée. Ainsi, la comparaison permet de mesurer le spectre des possibles résultats d'une méthode, et permet d'avoir une comparaison entre méthodes qui ne soit pas biaisée par les paramètres de régularité choisis.

Dans cette partie, nous présentons donc : une étude des métriques existantes, une nouvelle méthode de décomposition (SCALP), et une nouvelle métrique d'évaluation de la régularité avec un nouveau schéma robuste de comparaison des méthodes de superpixels.

Mots-clés

Superpixels, Segmentation, Détection de contours, Métriques d'évaluation, Régularité

Contenu

— Le chapitre 4 présente les intérêts de l'utilisation des superpixels dans diverses applications de traitement d'images (section 4.1) et l'état-de-l'art des méthodes de superpixels (section 4.2). Une étude des métriques de superpixels standards est alors effectuée pour montrer les limites ou défauts des mesures de référence, afin d'évaluer les méthodes de superpixels sur des critères pertinents (section 4.3)

— Le chapitre 5 présente ensuite la méthode SCALP (section 5.2) avec une extension permettant d'intégrer une segmentation initiale au processus de décomposition (section 5.3). La méthode SCALP est validée sur les métriques de superpixels démontrées pertinentes dans le chapitre 4, et obtient les meilleures performances par rapport aux méthodes de l'état-de-l'art (section 5.4).

— Le chapitre 6 propose une nouvelle mesure d'évaluation de la régularité des décompositions en superpixels (section 6.2). Cette mesure évalue de façon globale la régularité en considérant une métrique de régularité de forme et de consistance de la décomposition. SCALP est alors comparée aux méthodes de l'état-de-l'art sur ce critère (section 6.3). L'évaluation et la possibilité de paramétrer la régularité s'avèrent cruciales, puisque nous montrons que cette dernière peut fortement influer sur les performances de nombreuses applications basées sur les superpixels (section 6.4).

Publications associées

Revues internationales avec comité de lecture (1)

[1] *Evaluation Framework of Superpixel Methods with a Global Regularity Measure*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Journal of Electronic Imaging (JEI), 2017
Special issue on Superpixels for image processing and computer vision

Conférences internationales avec comité de lecture et actes (2)

[1] *Robust Shape Regularity Criteria for Superpixel Evaluation*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
IEEE International Conference on Image Processing (ICIP), 2017

[2] *SCALP: Superpixels with Contour Adherence using Linear Path*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
International Conference on Pattern Recognition (ICPR), 2016

Conférences nationales avec comité de lecture et actes (1)

[1] *Décomposition en superpixels via l'utilisation de chemin linéaire*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Groupe d'Études du Traitement du Signal et des Images (GRETSI), 2017

Articles en révision (1)

[1] *Robust Superpixels using Color and Contour Features along Linear Path*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Computer Vision and Image Understanding (CVIU)

Chapitre 4

Algorithmes de superpixels et étude des métriques standards

Sommaire

4.1 Utilisation des superpixels 78
4.2 Méthodes de décomposition en superpixels 81
4.2.1 Méthodes de graphes 82
4.2.2 Méthodes hiérarchiques 83
4.2.3 Méthodes de montée de gradient 84
4.3 Étude des métriques de superpixels 86
4.3.1 Homogénéité du groupement sur la couleur 87
4.3.2 Respect des objets de l'image 89
4.4 Conclusion 94

4.1 Utilisation des superpixels

De nombreuses applications de traitement et d'analyse d'images peuvent bénéficier d'une étape de présegmentation groupant a priori les pixels de l'image. Par exemple, l'analyse d'images consiste généralement à reconnaître et localiser les différents objets présents dans une image. Cette reconnaissance peut se faire par une segmentation préalable des objets, qui sont ensuite classés. La segmentation est donc une étape essentielle qui peut avoir un impact fort sur la précision ainsi que le temps de calcul de la méthode de classification. Afin d'augmenter la précision de reconnaissance des objets, des méthodes de décomposition en superpixels ont été initialement introduites. Ces méthodes groupent les pixels en zones de couleurs homogènes, tout en essayant de respecter les contours des objets de l'image, réduisant ainsi le nombre d'éléments à traiter, et par conséquent la charge de calcul.

Pour des superpixels homogènes, de formes irrégulières, les intensités des pixels contenus dans chaque région ont une très faible variance, contrairement aux blocs d'une décomposition régulière qui ne s'adapte pas au contenu de l'image. Cette comparaison entre décomposition régulière et irrégulière avec les superpixels est illustrée dans la Figure 4.1.

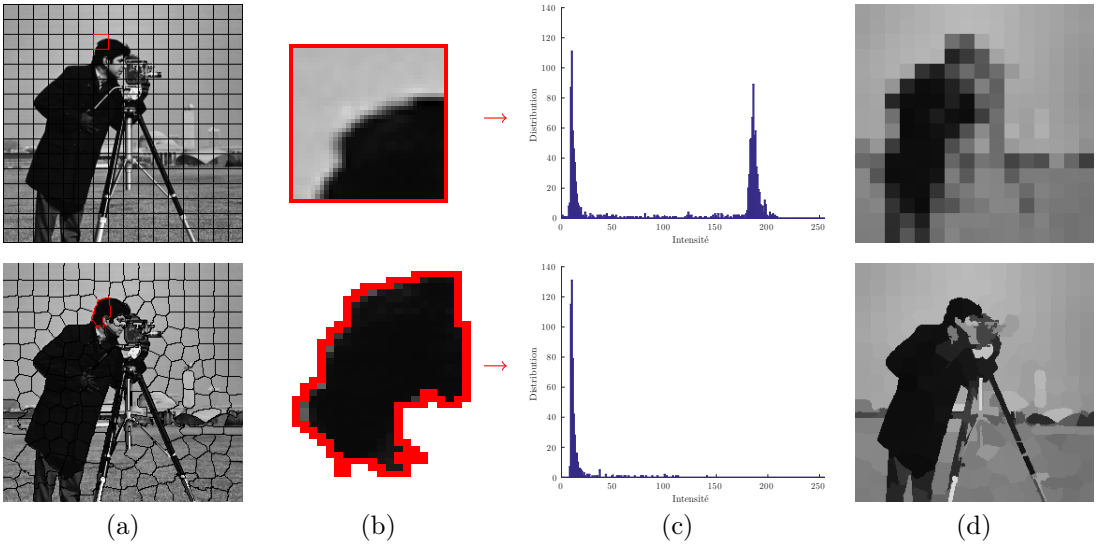


FIGURE 4.1 – Illustration de la faible variance d'intensité dans les superpixels. L'image *Cameraman* est décomposée en blocs réguliers (en haut) et en superpixels avec la méthode SCALP (en bas). Une région est sélectionnée pour chaque décomposition (b) et l'histogramme des intensités des pixels contenus dans cette région est donné en (c). L'intensité moyenne dans chaque région est alors représentée en (d). Les superpixels permettent de résumer efficacement le contenu de l'image.

La Figure 4.2 montre quant à elle les résultats d'une segmentation en trois classes effectuée sur plusieurs descripteurs : sur des patchs réguliers, sur des blocs d'une décomposition en grille régulière, sur des blocs d'une décomposition de type arbre

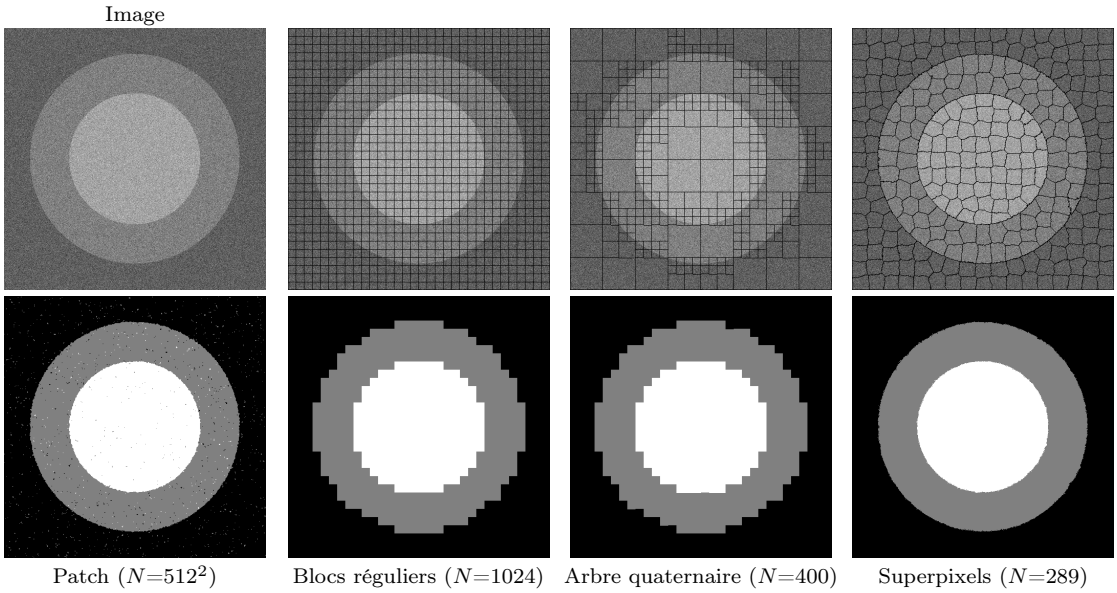


FIGURE 4.2 – Précision et robustesse de segmentation des superpixels par rapport à des descripteurs réguliers. La ligne du bas correspond au résultat de segmentation en trois classes d'une image bruitée avec l'algorithme d'Otsu depuis (de gauche à droite) des patchs réguliers de taille 5×5 en chaque pixel, des blocs réguliers, des blocs de différentes tailles issus d'une décomposition de type arbre quaternaire, et depuis des superpixels calculées avec la méthode SCALP. Le nombre d'éléments considérés N est donné pour chaque type de descripteurs. Les superpixels segmentent précisément l'image, en considérant un nombre réduit d'éléments, quand les patchs ne sont pas robustes au bruit et les autres descripteurs ne suivent pas les contours de l'image, et mènent à une segmentation imprécise.

quaternaire (Tanimoto et Pavlidis, 1975), et sur des superpixels. La segmentation est effectuée par l'algorithme d'Otsu sur les intensités moyennes de chaque descripteur. Les patchs réguliers sont peu robustes au bruit, quand les superpixels fournissent un résultat précis, proche du résultat optimal à l'échelle pixellique, contrairement aux autres descripteurs réguliers. De plus, les superpixels permettent de réduire considérablement le nombre d'éléments considérés et donc potentiellement le temps de calcul des méthodes basées sur les superpixels.

Les superpixels permettent donc de capturer efficacement les objets de l'image, en respectant leur contours, tout en réduisant significativement le nombre d'éléments à traiter. Ils sont donc de plus en plus utilisés dans d'autres processus de traitement et d'analyse d'images classiques, par exemple pour la localisation d'objets (Fulkerson et al., 2009), le traitement stéréo et les problèmes d'occultation (Zhang et al., 2011b), la détection de contours (Arbelaez et al., 2011), le suivi vidéo (Wang et al., 2011; Chang et al., 2013; Reso et al., 2013), la détection de saillance (Yang et al., 2013; Liu et al., 2014), la segmentation d'objets multi-classes (Yang et al., 2010; Tighe et Lazebnik, 2010; Kae et al., 2013; Gould et al., 2014), ou sont adaptés aux architectures de réseaux de neurones (Liu et al., 2015b; Gadde et al., 2016), et peuvent être décrits par des descripteurs standards (Tighe et Lazebnik, 2010).

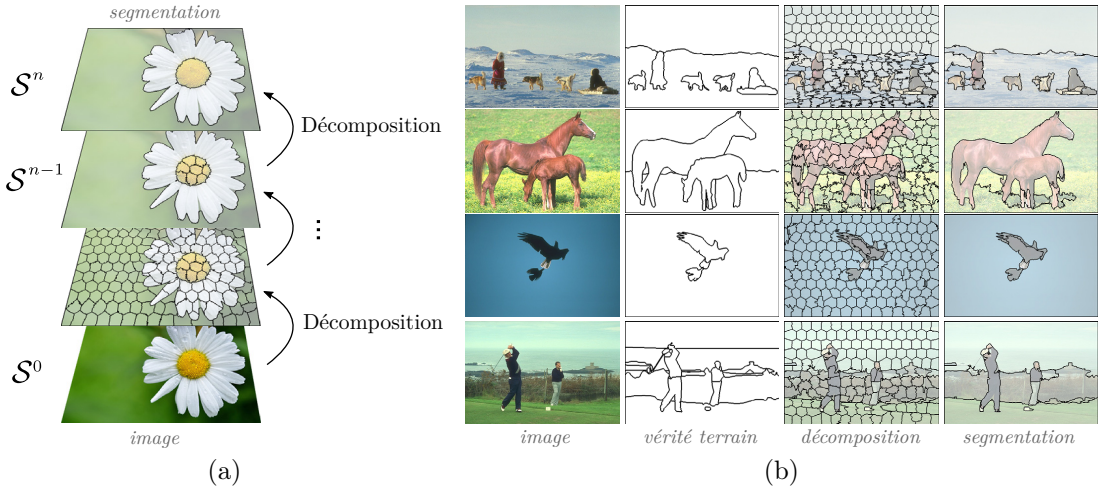


FIGURE 4.3 – Segmentation hiérarchique depuis une décomposition en superpixels (a), avec plusieurs exemples de résultats (b). L'image est décomposée en superpixels $\mathcal{S}^i$ à une échelle i, et ces superpixels sont progressivement regroupés pour segmenter les objets de l'image en n itérations. *Source : Nakamura et Hong (2017).*

Les superpixels permettent de segmenter l'image en un nombre généralement fixé de régions. Une image est alors souvent segmentée en plus de régions qu'elle ne contient d'objets, et les approches classiques ne permettent pas d'obtenir avec précision les objets de l'image en fixant un nombre très faible de superpixels. Pour obtenir une segmentation depuis une décomposition en superpixels, certaines approches considèrent un modèle hiérarchique, en fusionnant a posteriori les superpixels (Levinshtein et al., 2010; Li et al., 2012; Nakamura et Hong, 2017). Cette approche est illustrée dans la Figure 4.3, où une image est décomposée en un grand nombre de superpixels réguliers, progressivement fusionnés pour obtenir une segmentation pertinente des objets de l'image en quelques éléments. Ces modèles hiérarchiques sont également utilisés pour la détection de la saillance visuelle (Tong et al., 2014).

D'autres approches de segmentation utilisent des images d'exemples avec leur vérité terrain (Kae et al., 2013; Gould et al., 2014). Dans la partie I, les patchs réguliers ont permis de s'adapter à de fines structures, et de capturer des motifs similaires dans la bibliothèque d'exemples, permettant de segmenter des structures dont les contours ne suivaient pas le contenu de l'image. Néanmoins, sur d'autres types de données médicales, les superpixels, ou supervoxels en 3D, permettent de capturer efficacement certaines structures comme par exemple les tumeurs (Wang et Yushkevich, 2013; Irving et al., 2014).

La décomposition en superpixels s'étend également à la vidéo. La détection de saillance est par exemple étudiée en 3D sur la vidéo, où les supervoxels sont particulièrement intéressants pour assurer la cohérence de la détection au cours de la séquence (Liu et al., 2014). Des modèles de décompositions en supervoxels permettent en effet de décomposer l'image en *tubes*, où le même supervoxel peut être

étiré sur toute la séquence vidéo. Ces supervoxels permettent par exemple le suivi (*tracking*) des objets dans l'image (Wang et al., 2011; Chang et al., 2013; Reso et al., 2013), que nous aborderons aussi dans la section 6.4. La segmentation d'objets sur un flux vidéo peut également donner une information de mouvement, qui peut permettre la reconnaissance d'actions spécifiques (Jain et al., 2014).

Les superpixels sont donc très intéressants par rapport aux approches régulières de réduction de dimension, et sont de plus en plus utilisés dans de nombreux domaines de vision par ordinateur. Pour la majorité des applications visées, la précision de segmentation des superpixels est fondamentale, et pour certaines comme le suivi vidéo, la régularité de la décomposition est également essentielle (section 6.4). La méthode SCALP proposée dans le chapitre 5 vise à maximiser ces aspects, et elle s'étend naturellement en 3D, ce qui n'est pas le cas de toutes les méthodes de superpixels de l'état-de-l'art.

4.2 Méthodes de décomposition en superpixels

De très nombreuses méthodes de décomposition en superpixels ont été proposées au fil des années, et plusieurs travaux se sont attachés à dresser un état-de-l'art et comparer les performances des méthodes les plus populaires, par exemple, Neubert et Protzel (2012); Stutz et al. (2016). Nous présentons dans cette section les méthodes les plus populaires, afin de donner un aperçu des différentes approches utilisées pour décomposer une image en superpixels. Les propriétés recherchées pour les superpixels peuvent varier d'une méthode à l'autre. Néanmoins, la majorité des méthodes tendent à produire des superpixels qui respectent les propriétés suivantes : régularité en termes de forme et de taille, respect des contours des objets de l'image et homogénéité de couleur.

Dans la suite, nous grouperons les méthodes en trois types selon l'approche de décomposition utilisée : les méthodes de graphe, hiérarchiques ou de montée de gradient. Dans le Tableau 4.1, nous résumons les propriétés des méthodes de l'état-de-l'art FH (Felzenszwalb et Huttenlocher, 2004), Quick shift QS (Vedaldi et Soatto, 2008), Turbopixels TP (Levinshtein et al., 2009), ERS (Liu et al., 2011), SEEDS (Van den Bergh et al., 2012), SLIC (Achanta et al., 2012), ERGC (Buyssens et al., 2014), Waterpixels WP (Machairas et al., 2015), LSC (Chen et al., 2017), ETPS (Yao et al., 2015), et la méthode SCALP présentée dans le chapitre 5.

Nous considérons la complexité des algorithmes, la capacité de chaque méthode à contrôler le nombre de superpixels et leur régularité. Un exemple de décomposition pour chaque méthode est également illustré dans la Figure 4.4. Notons que le type de méthode ne reflète pas nécessairement les propriétés des superpixels générés. Par exemple, deux méthodes de graphe telles que Ren et Malik (2003) et Felzenszwalb et Huttenlocher (2004) produisent respectivement des superpixels très réguliers et très irréguliers.

Tableau 4.1 – Comparaison des méthodes de superpixels de l'état-de-l'art basées sur un graphe (bg), hiérarchiques (h) ou de montée de gradient (mg). Nous indiquons la capacité des méthodes à régler le nombre de superpixels en sortie et leur régularité, ainsi que la complexité des algorithmes pour une image de N pixels.

	FH	QS	TP	ERS	SEEDS	SLIC	ERGC	WP	LSC	ETPS	SCALP
Type	bg	mg	mg	bg	h	mg	bg	mg	mg	h	mg
Nombre	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
Régularité	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
Complexité ($\mathcal{O}$)	$N \log N$	N^2	N	$N^2 \log N$	N	N	$N \log N$	N	N	N	N

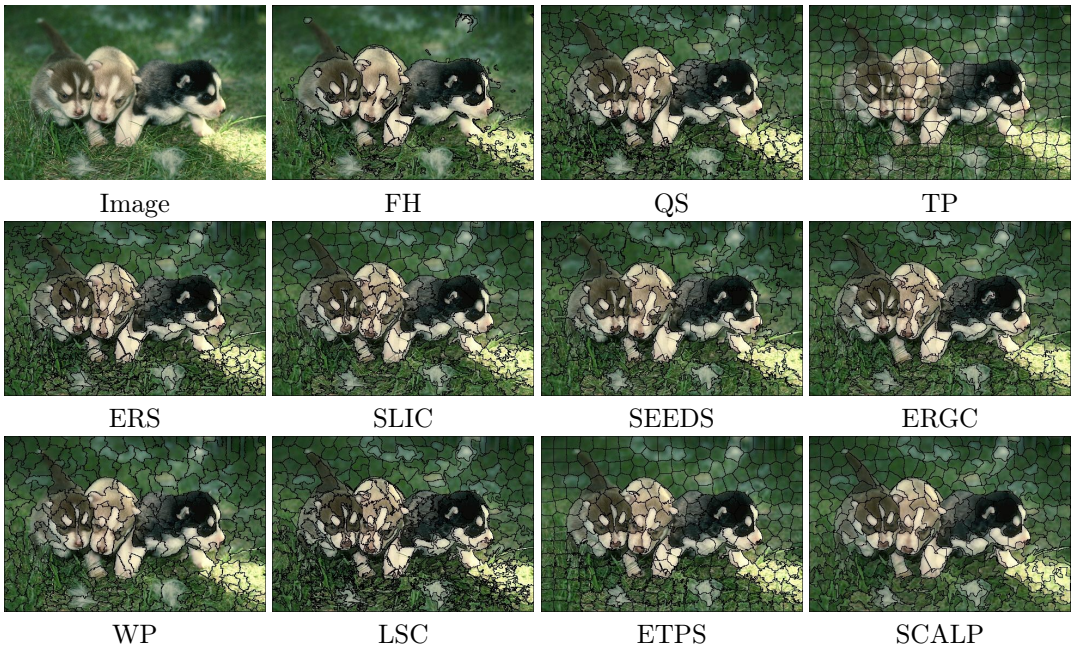


FIGURE 4.4 – Exemple de décomposition de plusieurs méthodes de l'état-de-l'art en approximativement 250 superpixels. Notons que les méthodes FH et QS ne permettent pas de fixer directement le nombre de superpixels.

4.2.1 Méthodes de graphes

La plupart de ces méthodes considèrent l'image comme un graphe et cherchent un minimum local d'une fonctionnelle donnée. Ces approches permettent aussi bien de produire des superpixels réguliers (Ren et Malik, 2003; Buyssens et al., 2014) ou irréguliers (Felzenszwalb et Huttenlocher, 2004). Le terme superpixel a été introduit dans Ren et Malik (2003) qui utilise un algorithme de type *normalized cuts* (Shi et Malik, 2000) pour décomposer une image en zones régulières. La charge de calcul de cet algorithme, qui segmente l'image selon des descripteurs de contour et de texture, est importante et a limité l'utilisation de cette méthode.

La méthode FH proposée dans Felzenszwalb et Huttenlocher (2004), considère chaque pixel comme un noeud du graphe et effectue un groupement rapide des pixels au détriment de contraintes de connectivité et de régularité. Les superpixels produits

sont donc très irréguliers en termes de formes et de tailles. De telles irrégularités peuvent poser différents problèmes. Par exemple, de trop petites régions peuvent empêcher de calculer un descripteur pertinent, quand de grands superpixels sont susceptibles de se superposer avec plusieurs objets de l'image.

Dans Moore et al. (2008), les superpixels sont contraints de respecter un maillage, calculé depuis une carte de contours, et sont raffinés par un algorithme de type *optimal cuts*. Cependant, malgré la contrainte du maillage, les superpixels produits ont des formes assez irrégulières, et la méthode apparaît très dépendante de la carte de contours utilisée. La méthode de Veksler et al. (2010) utilise une approche de type *graph cuts*, où la régularité est augmentée en associant les pixels à un ensemble de patchs réguliers se superposant à plusieurs échelles. Enfin, l'approche ERS, de type *graph cuts* considérée dans Liu et al. (2011) vise à maximiser le taux d'entropie de chaque superpixel au sein d'un processus de groupement agglomératif. La décomposition générée peut être assez irrégulière mais la méthode obtient des résultats proches de l'état-de-l'art en termes de respect des objets.

Enfin, la méthode ERGC utilise une formulation du problème eikonal sous forme de graphe pour itérativement raffiner les bordures des superpixels (Buyssens et al., 2014). Afin de compenser les limites du placement initial des superpixels, une étape de raffinement est proposée en considérant une distance géodésique sur la couleur afin d'ajouter de nouveaux superpixels dans les zones de forte variance.

4.2.2 Méthodes hiérarchiques

Les approches de type hiérarchique considèrent l'image à différentes échelles et convergent progressivement vers une décomposition à l'échelle pixellique. La méthode SEEDS (Van den Bergh et al., 2012), de façon similaire à une approche de type arbre quaternaire, segmente itérativement l'image en considérant des blocs de taille de plus en plus réduite jusqu'à ne contenir qu'un pixel. SEEDS est initialisé depuis une décomposition régulière en blocs larges et raffine les bordures des superpixels en effectuant des transferts de blocs. Cette méthode considère une énergie calculée sur les histogrammes de couleur des superpixels, qui s'avère relativement robuste au bruit (Stutz et al., 2016), et fournit des résultats en un temps de calcul réduit. Cependant, SEEDS fournit des superpixels irréguliers et ne permet pas de directement contrôler le nombre d'éléments en sortie.

Plus récemment, la méthode ETPS a été proposée dans Yao et al. (2015). De façon similaire à SEEDS (Van den Bergh et al., 2012), à chaque échelle, la méthode considère des blocs en bordure des superpixels, et le transfert n'est permis que s'il respecte la connectivité, c'est-à-dire si l'association ne sépare pas un superpixel en parties disjointes. Cette contrainte permet de maintenir la connectivité tout au long de la décomposition quand d'autres méthodes, telles que SLIC (Achanta et al., 2012) nécessitent une étape de post-traitement pour assurer cette propriété. Nous illustrons ce processus de transfert itératif dans la Figure 4.5. Notons que contrairement à SEEDS (Van den Bergh et al., 2012), ETPS permet de régler la régularité de la

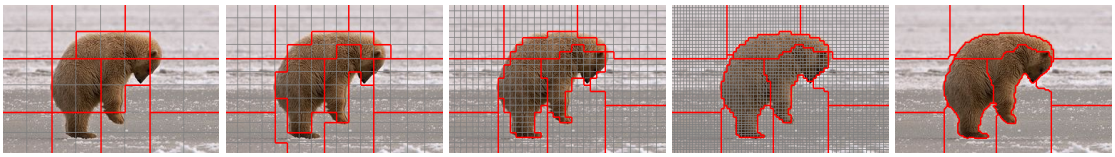


FIGURE 4.5 – Illustration du processus de décomposition de la méthode ETPS de type hiérarchique. *Source : Yao et al. (2015).*

décomposition. En général, les approches par blocs de type hiérarchiques peuvent facilement être optimisées numériquement et permettent d'atteindre de faibles temps de calcul en concentrant le raffinement de la décomposition sur les blocs en bordure des superpixels.

4.2.3 Méthodes de montée de gradient

Des méthodes telles que MS (pour Mean Shift, Comaniciu et Meer (2002)) et QS (pour Quick Shift, Vedaldi et Soatto (2008)) effectuent une recherche de mode pour grouper les pixels en régions. La méthode de Comaniciu et Meer (2002) génère des superpixels aux formes très irrégulières, en un temps de calcul important. La méthode de Vedaldi et Soatto (2008) quant à elle, considère une segmentation initiale pour améliorer la précision de la décomposition, mais elle requiert un temps de calcul supérieur à celui de la méthode de Comaniciu et Meer (2002). De plus, ces méthodes présentent l'inconvénient majeur de ne pas permettre le contrôle du nombre de superpixels générés.

Dans Levinshtein et al. (2009), des superpixels sont générés depuis une grille de départ régulière. Cependant, Levinshtein et al. (2009) ne permet pas de moduler la régularité des superpixels. Celle-ci est par défaut très élevée, c'est-à-dire que tous les superpixels ont une taille similaire et des formes quasi convexes, qui les empêchent d'adhérer correctement aux contours des objets de l'image.

Bien que les régions segmentées ne soient pas mentionnées comme étant des superpixels, Vincent et Soille (1991) proposent de décomposer une image en utilisant une approche basée sur les lignes de partage des eaux. Sans aucune contrainte, ce type de méthode, qui effectue une montée de gradient vers un minimum local, ne contraint pas la forme ou la taille des régions. Ainsi, la méthode de Vincent et Soille (1991) produit des régions aux formes très irrégulières. Pour contraindre les formes de ces régions, un terme spatial est ajouté au modèle de lignes de partage des eaux dans Neubert et Protzel (2014); Machairas et al. (2015). Dans l'ensemble, les méthodes de montée de gradient permettent à la fois de produire des superpixels très réguliers (Levinshtein et al., 2009) et très irréguliers (Vedaldi et Soatto, 2008), mais ces décompositions n'atteignent en général pas les performances de l'état-de-art en termes de précision de segmentation.

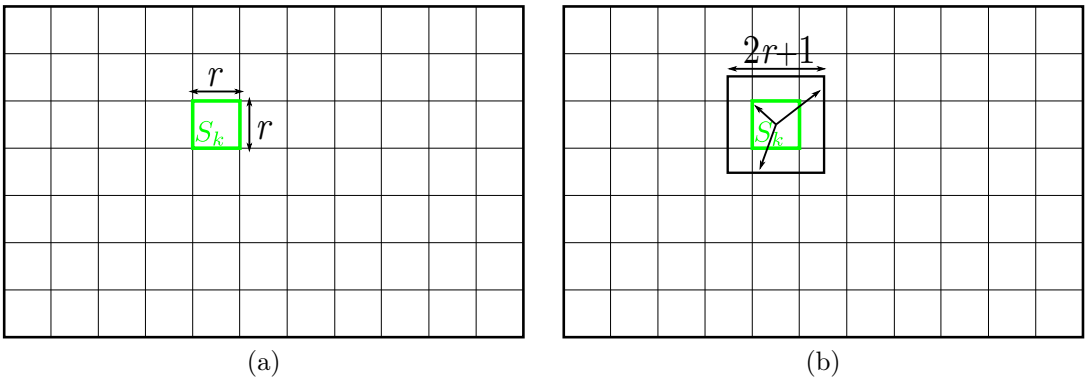


FIGURE 4.6 – Illustration de la méthode de groupement itératif SLIC (Achanta et al., 2012). (a) Initialisation des superpixels S_k avec une grille régulière en blocs de taille $r \times r$. (b) Zone de recherche d'association des pixels de taille $(2r+1) \times (2r+1)$ centrée sur le barycentre spatial du superpixel.

Depuis son introduction, la méthode SLIC (pour Simple Linear Iterative Clustering, Achanta et al. (2012)) a été intensivement étudiée. Comme son nom l'indique, la méthode est simple et effectue un groupement itératif des pixels, basé sur une approche de type K-moyennes localement contrainte, et initialisée depuis une grille régulière. Nous illustrons le processus de groupement itératif d'Achanta et al. (2012) dans la Figure 4.6. Le groupement est calculé selon une distance spatiale et couleur, et peut produire des décompositions régulières ou irrégulières, selon le réglage du compromis entre ces deux termes. Les auteurs d'Achanta et al. (2012) utilisent l'espace CIELab, mais d'autres espaces (Qazi et al., 2011) peuvent être utilisés pour calculer la distance en couleur, liée au contenu de l'image. Enfin, les décompositions sont calculées avec SLIC en un temps largement inférieur à ceux d'autres approches du même type comme Veksler et al. (2010) ou Liu et al. (2011), et celles-ci obtiennent également de bonnes performances sur les métriques de superpixels standards.

Pour ces raisons, SLIC a été utilisée dans un nombre important de pipelines de vision par ordinateur, par exemple pour la segmentation d'images médicales (Lucchi et al., 2012), le calcul de flot optique (Lu et al., 2013), ou la détection de saillance (Zhu et al., 2014). SLIC a également été étendu dans de nombreux travaux, par exemple, afin d'améliorer l'homogénéité de couleur en calculant des superpixels sur des variétés projetées dans l'espace spatial et colorimétrique (Liu et al., 2016b), ou en effectuant le groupement des pixels dans un espace de descripteurs de haute dimension (Chen et al., 2017). Certaines approches se sont intéressées à la diminution du temps de calcul (Neubert et Protzel, 2014; Huang et al., 2016; Achanta et Süsstrunk, 2017). Enfin, les performances d'adhérence aux contours ont été améliorées en modifiant la distance de groupement de l'algorithme initial dans des travaux tels que Kim et al. (2013a); Jia et al. (2015); Zhang et al. (2016); Rubio et al. (2016); Wang et al. (2017a).

4.3 Étude des métriques de superpixels

Comme indiqué en introduction, les principales propriétés des superpixels sont les suivantes : (i) l'homogénéité de la couleur au sein de chaque superpixel, (ii) le respect des objets de l'image par les frontières des superpixels, (iii) la régularité de forme et de taille des superpixels. Puisque ces propriétés ne peuvent être optimisées en même temps (Neubert et Protzel, 2014), la plupart des méthodes cherchent un compromis entre ces aspects dans leurs modèles.

Pour illustrer la nécessité de tels compromis, la Figure 4.7 considère des exemples synthétiques (a) et (d), décomposés différemment en trois superpixels. Dans les Figures 4.7(b) et (e), l'homogénéité de couleur (i) est optimale, quand le respect des contours de l'image (ii) et la régularité (iii) sont respectivement favorisés dans (c) et (f). De tels exemples permettent d'illustrer comment une méthode de décomposition en superpixels doit se comporter pour optimiser chaque critère, et que l'homogénéité de couleur n'est pas nécessairement corrélée avec les contours perceptuels d'un objet.

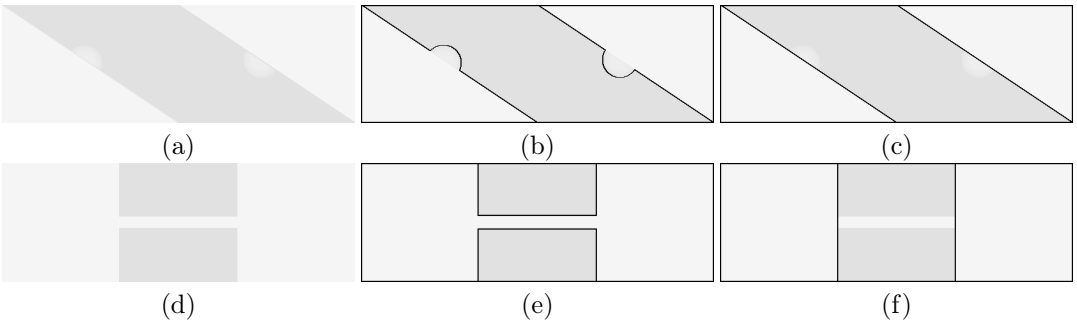


FIGURE 4.7 – Compromis entre les différentes propriétés pour la décomposition des images (a) et (d) en trois superpixels. Les décompositions en (b) et (e) sont optimales en termes d'homogénéité des couleurs, quand celles en (c) et (f) sont respectivement optimales en termes de respect des contours de l'image et de régularité.

Avec les nombreuses méthodes de superpixels introduites, beaucoup de métriques dédiées aux superpixels ont été proposées pour évaluer la conformité des décompositions avec les propriétés mentionnées précédemment (Levinshtein et al., 2009; Liu et al., 2011; Achanta et al., 2012; Schick et al., 2012; Neubert et Protzel, 2012). Les métriques les plus utilisées sont les suivantes :

(i) Critère d'homogénéité de couleur :
- ICV (pour Intra-Cluster Variation, Benesova et Kottman (2014)),
- EV (pour Explained Variation, Moore et al. (2008)).

(ii) Critère de respect des objets de l'image :
- UE (pour Undersegmentation Error, Levinshtein et al. (2009); Achanta et al. (2012); Neubert et Protzel (2012)),

— ASA (pour Achievable Segmentation Accuracy, Liu et al. (2011)),
— BR (pour Boundary Recall, Martin et al. (2004)).

(iii) Critère de régularité :
— C (pour Circularité, Schick et al. (2012)),
— MF (pour Mismatch Factor, Machairas et al. (2015)).

Ces métriques offrent différentes interprétations et évaluations des propriétés des superpixels. Par exemple, la métrique de circularité (Schick et al., 2012) évalue les formes circulaires comme les plus régulières, ce qui peut être discutable.

Dans cette section, nous présentons les métriques de l'état-de-l'art qui ont été progressivement introduites pour les critères (i) et (ii), et nous les analysons afin de concentrer l'évaluation de SCALP dans le chapitre 5, uniquement sur les mesures standards pertinentes. Dans le chapitre 6, nous proposons une nouvelle métrique de régularité, puisqu'aucune métrique de l'état-de-l'art n'apparaît satisfaisante, pour comparer SCALP aux méthodes de l'état-de-l'art sur ce critère. Dans la suite, nous considérons une image I, dont la décomposition en superpixels est notée $\mathcal{S} = \{S_k\}_{k\in\{1,\dots,|\mathcal{S}|\}}$, composée de $|\mathcal{S}|$ superpixels S_k, où $|.|$ représente la cardinalité de l'élément considéré, et telle que pour deux superpixels $S_k, S_j \in \mathcal{S}$, $S_k \cap S_j = \emptyset$.

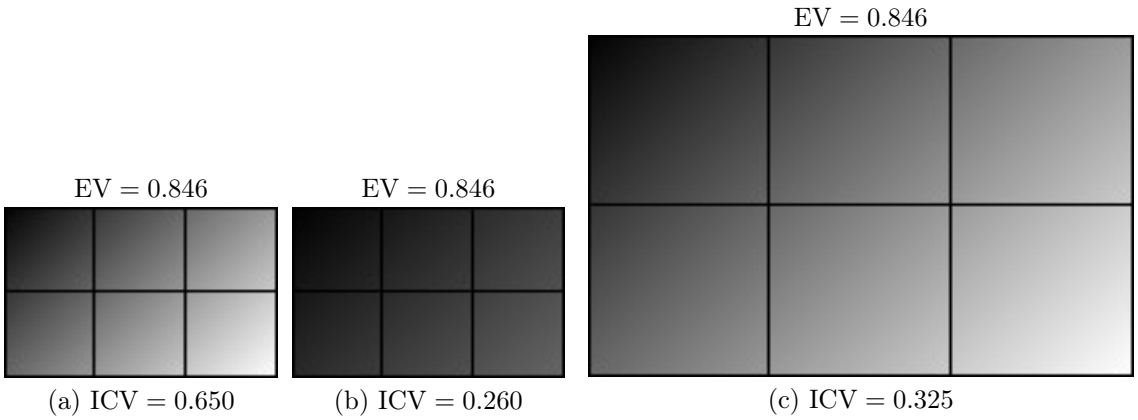
4.3.1 Homogénéité du groupement sur la couleur

L'homogénéité du groupement sur la couleur est une propriété importante de décomposition puisque les superpixels visent à réduire le nombre d'éléments de l'image en synthétisant notamment son information de couleur. De plus, bien que le groupement des pixels de couleurs proches (i) puisse permettre assez efficacement de capturer les frontières des objets (ii), les critères (i) et (ii) ne sont pas strictement corrélés (voir Figure 4.7), La plupart des méthodes, dont SLIC (Achanta et al., 2012), calculent dans leur modèle de décomposition, un compromis entre distances spatiale et de couleur pour calculer les superpixels. La capacité à grouper des pixels homogènes doit donc être considérée dans le processus de comparaison des méthodes, mais cet aspect est rarement évalué dans les travaux sur les superpixels.

L'homogénéité de couleur peut être évaluée en comparant les couleurs moyennes des superpixels aux couleurs des pixels de l'image initiale. La variation intra-cluster ICV (Benesova et Kottman, 2014) a été proposée pour mesurer une telle variation de couleur pour une décomposition $\mathcal{S}$:

$$\mathrm{ICV}(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{S_k} \frac{\sqrt{\sum_{p\in S_k} (I(p) - \mu(S_k))^2}}{|S_k|}, \tag{4.1}$$

avec $\mu(S_k)$, la couleur moyenne du superpixel S_k.



(a) ICV = 0.650 (b) ICV = 0.260 (c) ICV = 0.325

FIGURE 4.8 – Comparaison de la robustesse de l'ICV et de l'EV sur l'évaluation de l'homogénéité de couleur des superpixels. La dynamique et la dimension de l'image (a) décomposées en six superpixels sont respectivement modifiées en (b) et (c). EV est robuste à ces transformations contrairement à ICV.

Le critère EV (Moore et al., 2008) a également été proposé pour mesurer l'homogénéité de couleur et est défini comme :

$$\mathrm{EV}(\mathcal{S}) = \frac{\sum_{S_k} |S_k| \left(\mu(S_k) - \mu(I)\right)^2}{\sum_{p \in I} \left(I(p) - \mu(I)\right)^2}. \tag{4.2}$$

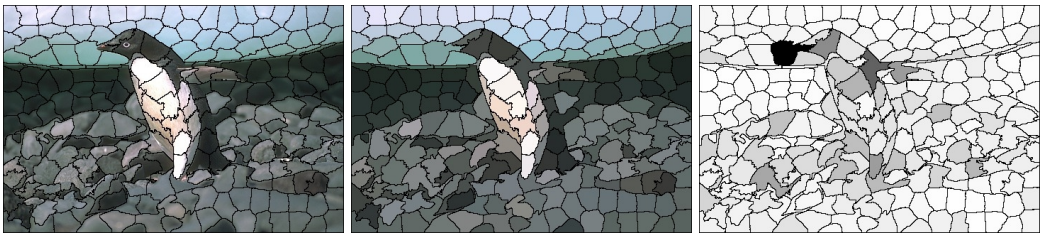
Analyse

La métrique ICV présente plusieurs inconvénients. Elle n'est pas normalisée par la variance de l'image pour permettre une comparaison pertinente entre différentes images (Stutz et al., 2016), et la taille des superpixels n'est pas considérée, donc la mesure n'est pas robuste à l'échelle. Dans la Figure 4.8, nous illustrons ces aspects sur des exemples synthétiques. La dynamique et la taille de la même image avec la même décomposition en superpixels sont modifiées et ces transformations ont une incidence sur la mesure ICV.

La métrique EV quant à elle est robuste à ces transformations. Elle exprime la variance couleur contenue dans chaque superpixel, et de simples opérations nous permettent de transformer l'Éq. (4.2), pour obtenir une formulation plus explicite de cette mesure :

$$\mathrm{EV}(\mathcal{S}) = 1 - \sum_{S_k} \frac{|S_k|}{|I|} . \frac{\sigma(S_k)^2}{\sigma(I)^2}, \tag{4.3}$$

où la variance $\sigma(S_k)^2$ est considérée selon la taille $|S_k|$ de chaque superpixel S_k. Plus les superpixels sont homogènes et plus le critère est optimal (proche de 1).



(a) Superpixels (b) Couleurs moyennes (c) Variance couleur

FIGURE 4.9 – Exemple de décomposition en superpixels (a) avec couleurs moyennes (b) et variance couleur $\sigma(S_k)^2$ dans chaque superpixel S_k (c). La couleur blanche indique une faible variance couleur.

Dans la Figure 4.9, nous illustrons la variance couleur $\sigma(S_k)^2$ de chaque superpixel S_k. Une variance élevée dans un superpixel réduira la mesure EV. À l'inverse, une valeur élevée d'EV exprime un groupement homogène des couleurs. Malgré la robustesse de ce critère, celui-ci n'a pas été considéré dans les principaux travaux de superpixels. Nous recommandons néanmoins l'utilisation de la métrique EV plutôt que celle du critère ICV pour évaluer le critère d'homogénéité de couleur.

4.3.2 Respect des objets de l'image

Dans cette section, les métriques évaluent le respect des objets de l'image, et comparent donc une décomposition à une vérité terrain de segmentation qui sera notée $\mathcal{G} = \{G_j\}_{j\in\{1,\dots,|\mathcal{G}|\}}$, avec G_j une région segmentée.

En ce qui concerne l'UE, plusieurs définitions ont été successivement proposées (Levinshtein et al., 2009; Achanta et al., 2012; Neubert et Protzel, 2012). L'UE évalue le nombre de pixels qui traversent les limites des régions de la vérité terrain. La formulation initiale de l'UE (Levinshtein et al., 2009), notée $\mathrm{UE_L}$, est définie comme :

$$\mathrm{UE_L}(\mathcal{S},\mathcal{G}) = \frac{1}{|\mathcal{G}|}\sum_{G_j}\frac{\left(\sum_{S_k,S_k\cap G_j\neq\emptyset}|S_k|\right)-|G_j|}{|G_j|}. \tag{4.4}$$

La mesure $\mathrm{UE_L}$ a été discutée dans plusieurs travaux comme Achanta et al. (2012); Van den Bergh et al. (2012); Neubert et Protzel (2012), puisque tout superpixel se superposant avec une région vérité terrain pénalise la métrique. Ainsi, $\mathrm{UE_L}$ est très sensible à de légères superpositions et ne reflète pas précisément le respect des objets de l'image. Une façon de réduire l'impact de ces superpositions mineures dans (4.4) a été proposée dans Achanta et al. (2012), mais la mesure introduite nécessite un paramètre de seuillage. Les travaux récents tendent à utiliser la formulation d'UE sans paramètre introduite par Neubert et Protzel (2012) :

$$\mathrm{UE}(\mathcal{S},\mathcal{G}) = \frac{1}{|I|}\sum_{S_k}\sum_{G_j}\min\{|S_k\cap G_j|,|S_k\backslash G_j|\}. \tag{4.5}$$

Cette formulation considère soit l'intersection, soit l'aire du superpixel ne se superposant pas avec la région de la vérité terrain, et corrige ainsi l'absence de robustesse des précédentes définitions d'UE.

La mesure ASA (Liu et al., 2011) vise aussi à évaluer la superposition des superpixels avec la vérité terrain. Elle est utilisée dans la plupart des travaux de superpixels comme suit :

$$\mathrm{ASA}(\mathcal{S}, \mathcal{G}) = \frac{1}{|I|} \sum_{S_k} \max_{G_j} |S_k \cap G_j|. \tag{4.6}$$

Pour chaque superpixel S_k, la plus large superposition possible avec une région de la vérité terrain G_j est considérée, et des valeurs élevées d'ASA indiquent de bonnes performances.

La mesure BR est également largement recensée pour évaluer l'adhérence aux contours des objets. Cette métrique évalue la détection des contours de vérité terrain $\mathcal{B}(\mathcal{G})$ par les bordures des superpixels $\mathcal{B}(\mathcal{S})$ et est définie comme suit :

$$\mathrm{BR}(\mathcal{S}, \mathcal{G}) = \frac{1}{|\mathcal{B}(\mathcal{G})|} \sum_{p \in \mathcal{B}(\mathcal{G})} \delta[\min_{q \in \mathcal{B}(\mathcal{S})} \|p - q\| < \epsilon], \tag{4.7}$$

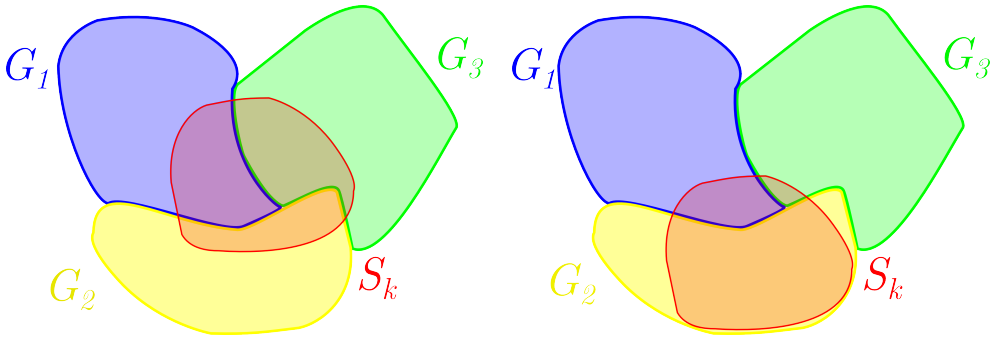
avec $\delta[a] = 1$ quand a est vrai et 0 sinon, et ϵ est un seuil de distance qui peut être réglé, par exemple à 2 pixels (Liu et al., 2011; Van den Bergh et al., 2012). Chaque pixel de contour de la vérité terrain est considéré comme détecté, si une frontière de superpixel est à une distance inférieure à ϵ.

Analyse

L'étude de Stutz et al. (2016) affirme que les mesures UE (4.5) et ASA (4.6) sont corrélées, de sorte que ASA $= 1 -$ UE, rendant alors redondante l'utilisation des deux métriques pour l'évaluation des méthodes de superpixels. Sous une condition pertinente, nous montrons que la vraie relation entre les deux mesures est :

$$\mathrm{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \mathrm{UE}(\mathcal{S}, \mathcal{G})/2. \tag{4.8}$$

La relation (4.8) est exacte quand tous les superpixels ont une superposition dite majeure, c'est-à-dire, qu'une région de la vérité terrain $\mathcal{G}$ se superpose avec plus de la moitié de l'aire du superpixel (Proposition 1). Ce cas est illustré dans la Figure 4.10(b), où un superpixel S_k a une superposition majeure avec une région G_2. Notons que cette condition est nécessairement vraie quand chaque superpixel ne se superpose qu'avec deux régions au plus (Proposition 2) ou quand la vérité terrain est binaire (Proposition 3).



(a) Pas de superposition majeure (b) Superposition majeure (4.9)

FIGURE 4.10 – Exemples de décomposition où un superpixel S_k se superpose avec plusieurs régions G_j. Dans (b), contrairement à (a), une région se superpose avec plus de la moitié de l'aire de S_k, ce qui correspond à la condition de superposition majeure dans (4.9).

Proposition 1. Si tous les superpixels S_k ont une superposition majeure, c'est-à-dire, si $\forall S_k, \exists G_i, |S_k \cap G_i| \geq |S_k|/2$, alors, $\mathrm{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \mathrm{UE}(\mathcal{S}, \mathcal{G})/2$.

Démonstration. L'hypothèse de superposition majeure, illustrée dans la Figure 4.10(b), signifie que chaque superpixel S_k a une superposition avec une région G_i couvrant plus de la moitié de son aire, c'est-à-dire :

$$\forall S_k, \exists G_i, |S_k \cap G_i| \geq |S_k|/2. \tag{4.9}$$

Dans ce cas, $\min\{|S_k \cap G_i|, |S_k \backslash G_i|\} = |S_k \backslash G_i|$, et $\forall G_{j,j\neq i}$, $\min\{|S_k \cap G_j|, |S_k \backslash G_j|\} = |S_k \cap G_j|$.
Par conséquent, nous obtenons :

$$\begin{aligned}
\sum_{G_j} \min\{|S_k \cap G_j|, |S_k \backslash G_j|\} &= |S_k \backslash G_i| + \sum_{G_j, j\neq i} |S_k \cap G_j|, \\
&= (|S_k| - |S_k \cap G_i|) + (|S_k| - |S_k \cap G_i|), \\
&= 2(|S_k| - |S_k \cap G_i|), \\
&= 2(|S_k| - \max_{G_j} |S_k \cap G_j|).
\end{aligned}$$

Puis :

$$\begin{aligned}
\mathrm{UE}(\mathcal{S}, \mathcal{G}) &= \frac{1}{|I|} \sum_{S_k} \sum_{G_j} \min\{|S_k \cap G_j|, |S_k \backslash G_j|\}, \\
&= \frac{2}{|I|} \sum_{S_k} \left(|S_k| - \max_{G_j} |S_k \cap G_j| \right), \\
&= 2 \left(1 - \frac{1}{|I|} \sum_{S_k} \max_{G_j} |S_k \cap G_j| \right), \\
&= 2(1 - \mathrm{ASA}(\mathcal{S}, \mathcal{G})).
\end{aligned}$$

D'après la définition de l'ASA (4.6), la relation (4.8) est donc vérifiée. □

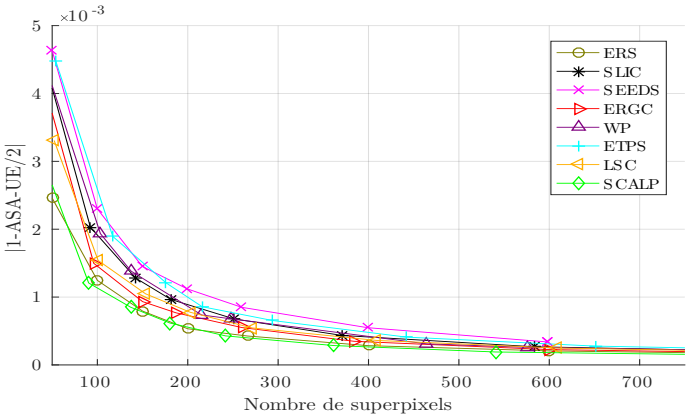


FIGURE 4.11 – Erreur entre l'ASA et l'UE mesurées, et l'approximation basée sur la condition (4.9). Les faibles valeurs obtenues par rapport aux métriques ASA et UE démontrent la pertinence de notre condition.

Proposition 2. Si tous les superpixels S_k se superposent avec au plus 2 régions, c'est-à-dire, si $\forall S_k, |I_k| \leq 2$, avec $\sum_{i \in I_k} |S_k \cap G_i| = |S_k|$, alors, $\mathrm{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \mathrm{UE}(\mathcal{S}, \mathcal{G})/2$.

Démonstration. Dans ce cas, un superpixel S_k peut se superposer avec une ou deux régions. Si S_k se superpose avec uniquement une région, par exemple, G_i, alors (4.9) est vérifiée. Sinon, si S_k se superpose avec deux régions, (4.9) est nécessairement vraie, puisqu'une de ces deux régions se superpose avec au moins la moitié de S_k. Ainsi, dans les deux cas, (4.8) est vérifiée. □

Proposition 3. Si $\mathcal{G}$ est une vérité terrain binaire, c'est-à-dire, $|\mathcal{G}| \leq 2$, alors, $\mathrm{ASA}(\mathcal{S}, \mathcal{G}) = 1 - \mathrm{UE}(\mathcal{S}, \mathcal{G})/2$.

Démonstration. Si $\mathcal{G}$ est une vérité terrain binaire, elle contient uniquement 2 régions telles que $\mathcal{G} = G_1 \cup G_2$, et tous les superpixels se superposent nécessairement avec au plus 2 régions. La condition de la Proposition 2 est ainsi vraie et (4.8) est vérifiée. □

La Figure 4.11 démontre la pertinence de cette relation en pratique. De très faibles différences sont calculées entre ce modèle et l'ASA et l'UE mesurées pour les décompositions des méthodes suivantes : ERS (Liu et al., 2011), SLIC (Achanta et al., 2012), SEEDS (Van den Bergh et al., 2012), ERGC (Buyssens et al., 2014), WP (Machairas et al., 2015), LSC (Chen et al., 2017), ETPS (Yao et al., 2015) et SCALP, appliquées aux images de test de la base BSD (pour Berkeley Segmentation Dataset, Martin et al. (2001)). Notons que l'amplitude de l'erreur diminue logiquement avec le nombre de superpixels, c'est-à-dire avec la précision de la décomposition.

Par conséquent, étant donné que la mesure de $\mathrm{UE_L}$ s'est révélée non robuste, et que nous avons démontré que l'étude de l'UE et de l'ASA était équivalente, nous recommandons d'évaluer le respect des objets de l'image avec la métrique ASA (4.6).

(a) Image avec les contours de la vérité terrain (b) Superpixels irréguliers (BR=1.00 | BR×(1-CD)=0.563) (c) Superpixels réguliers (BR=1.00 | BR×(1-CD)=0.858)

FIGURE 4.12 – Exemples de décompositions irrégulière (Yao et al., 2015) (b) et régulière (SCALP) (c) d'une image (a), avec une mesure BR maximale. BR évalue la détection des contours de vérité terrain, de sorte que les deux décompositions peuvent avoir une mesure BR maximale, bien que la l'irrégulière produise plus de pixels de bordure. Cet aspect est pris en compte par la mesure CD qui pénalise le grand nombre de pixels de frontières.

Bien que la mesure BR (4.7) ait été largement utilisée, et est recommandée dans Stutz et al. (2016), elle n'exprime pas directement le respect des objets de l'image, mais plutôt la capture des contours de ces régions. Cependant, la mesure à elle seule n'est pas significative, puisqu'elle ne considère que les vrais positifs, parmi les pixels de bordures de la vérité terrain détectés par les superpixels. Par conséquent, le nombre de pixels de contour des superpixels générés n'est pas considéré, et des méthodes très irrégulières peuvent obtenir des résultats de BR plus élevés. Pour ces raisons, de récents travaux tels que Machairas et al. (2015); Zhang et al. (2016) mesurent les résultats de BR selon la densité de contour CD (pour Contour Density), qui évalue le nombre de pixels de bordure des superpixels $\mathcal{B}(\mathcal{S})$ par rapport au nombre de pixel de l'image $|I|$, tel que :

$$\mathrm{CD}(\mathcal{S}) = \frac{|\mathcal{B}(\mathcal{S})|}{|I|}. \tag{4.10}$$

La Figure 4.12 compare deux décompositions avec le même nombre de superpixels et ayant une mesure de BR maximale, c'est-à-dire BR = 1. Une des décompositions est très irrégulière et produit un nombre élevé de pixels de bordure, et cet aspect n'est pas considéré par la métrique BR.

La mesure BR n'est donc pas significative utilisée seule, mais elle apporte toutefois une information différente de celle de l'ASA, puisqu'elle évalue uniquement la détection des contours des objets de l'image. Lorsque l'on considère des décompositions avec la même CD, c'est-à-dire le même nombre de pixels de bordure, les résultats BR peuvent être comparés de manière pertinente. Une mesure de BR supérieure avec la même CD indique que les superpixels produits détectent mieux les contours de l'image.

La mesure BR peut également être comparée à la précision P de détection de contours. Les superpixels produisant généralement plus de régions qu'il n'existe d'objets dans l'image, la précision diminue naturellement pour un plus grand nombre de superpixels. Néanmoins, cette mesure peut être utilisée pour étudier les perfor-

mances de détection de contours des méthodes de superpixels avec la représentation des courbes précision-rappel notées PR (Martin et al., 2004). Dans le contexte de la détection de contours, une carte de contours entre 0 et 1 est considérée, et les mesures de BR et P sont calculées après binarisation de la carte pour différents seuils. Pour obtenir cette carte de contours depuis des superpixels, on peut par exemple moyenner les bordures de décompositions obtenues à plusieurs échelles (Van den Bergh et al., 2012).

Les courbes PR considèrent donc la mesure BR (4.7), c'est-à-dire les vrais positifs, ou le pourcentage de contours de vérité terrain détectés, et la précision $\mathrm{P} = |\mathcal{B}(\mathcal{S}) \cap \mathcal{B}(\mathcal{G})|/|\mathcal{B}(\mathcal{S})|$, c'est-à-dire le pourcentage de vraies détections parmi les pixels de bordure des superpixels. L'information des courbes PR est généralement synthétisée par la mesure F, définie comme suit :

$$\mathrm{F} = \frac{2.\mathrm{P}.\mathrm{BR}}{\mathrm{P} + \mathrm{BR}}. \tag{4.11}$$

4.4 Conclusion

En résumé, la métrique EV est démontrée robuste contrairement à l'ICV, et est donc préférée pour évaluer l'homogénéité de couleur. L'ASA est suffisante pour mesurer la superposition des superpixels avec les objets de l'image, et la mesure BR doit être associée à la mesure CD pour être significative. La mesure CD donne alors une information sur la régularité des contours des superpixels. Enfin, les performances de détection de contours des superpixels peuvent être étudiées avec le schéma PR, et résumées par la mesure F.

Dans le chapitre 5, nous introduisons la méthode proposée SCALP, que nous validons et comparons aux méthodes de l'état-de-l'art sur ces métriques, et dans le chapitre 6, nous proposons une nouvelle métrique de régularité, puisqu'aucune métrique de référence ne se révèle satisfaisante pour valider SCALP sur ce critère.

Chapitre 5

La méthode SCALP

Sommaire

5.1 Contributions 98
5.2 Décomposition en superpixels avec SCALP 98
5.2.1 Schéma de groupement itératif 99
5.2.2 Information de couleur du voisinage 101
5.2.3 Information de couleur et contour sur le chemin linéaire 102
5.3 Contrainte depuis une segmentation initiale 106
5.3.1 Segmentation hiérarchique depuis une détection de contours . 107
5.3.2 Décomposition régulière en superpixels depuis une segmentation hiérarchique 107
5.4 Validation de SCALP sur les métriques superpixels standards 109
5.4.1 Cadre de validation 109
5.4.2 Impact des contributions 110
5.4.3 Comparaison avec les méthodes de l'état-de-l'art 114
5.4.4 Extension aux supervoxels 116
5.5 Conclusion 119

5.1 Contributions

Dans ce chapitre, nous proposons une nouvelle méthode de décomposition en superpixels appelée SCALP (pour Superpixels with Contour Adherence using Linear Path). Notre méthode est basée sur l'algorithme de groupement itératif de SLIC (Achanta et al., 2012) et améliore conjointement les trois propriétés principales : homogénéité de couleur, respect des objets de l'image et régularité des superpixels.

— La majorité des méthodes obtiennent des performances très dégradées lorsqu'elles sont appliquées à des images même légèrement bruitées (voir Figure 5.2). Dans Brekhna et al. (2017), la robustesse des méthodes de l'état-de-l'art est évaluée, et aucune n'apparaît satisfaisante sur tous les critères d'évaluation. Les auteurs soulignent la nécessité de proposer des algorithmes de décomposition robustes pour traiter des images naturelles pouvant être bruitées. Nous proposons de considérer le voisinage de chaque pixel dans le processus de décomposition. Nous montrons que cette information peut être intégrée en conservant la même complexité de calcul, tout en permettant d'améliorer la robustesse au bruit et la précision de segmentation des superpixels.

— Afin d'assurer l'homogénéité de couleur au sein d'une forme régulière, nous définissons le chemin linéaire entre un pixel et le barycentre d'un superpixel, et nous considérons l'information de couleur sur ce chemin dans le processus de décomposition. Contrairement aux distances géodésiques qui permettent des chemins irréguliers, conduisant à des formes irrégulières, notre chemin linéaire favorise naturellement la convexité des superpixels de la décomposition. Une carte de contours peut également être utilisée pour empêcher la traversée des contours des objets par les frontières des superpixels.

— Nous proposons également une méthode pour générer une décomposition en superpixels à partir d'une segmentation initiale, calculée depuis une carte de contours. Les superpixels produits sont réguliers, tout en respectant la segmentation initiale, utilisée pour garantir une forte adhérence aux contours.

— Nous effectuons une évaluation approfondie de SCALP sur l'ensemble de la base de données de segmentation BSD (pour Berkeley Segmentation Dataset, Martin et al. (2001)). Nos résultats sont supérieurs à ceux des méthodes de l'état-de-l'art sur les images initiales et bruitées, en termes de détection de contours et sur les métriques de superpixels standards.

5.2 Décomposition en superpixels avec SCALP

Dans cette section, nous présentons la méthode SCALP, basée sur l'algorithme de groupement itératif de SLIC, qui présente plusieurs avantages. D'abord, celui-ci permet de moduler le nombre de superpixels et leur régularité. Ensuite, la décomposition s'effectue en un temps de calcul réduit. Enfin, le modèle calcule de simples distances de couleur et spatiale, ce qui offre plusieurs possibilités d'amélioration.

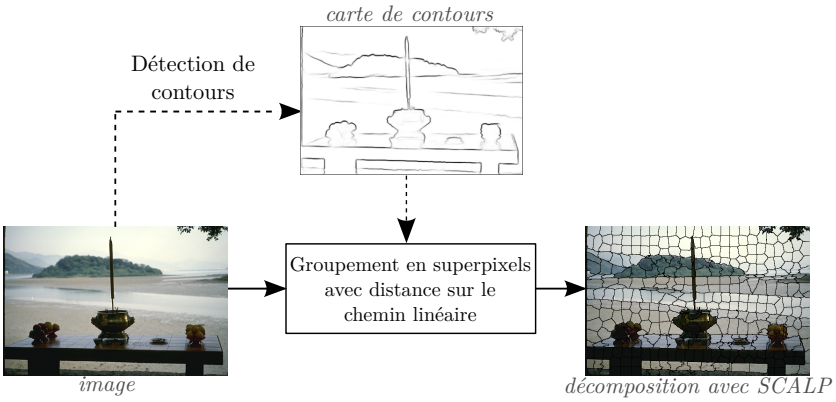


FIGURE 5.1 – La méthode SCALP. Une carte de contours peut être utilisée (flèches pointillées) pour assurer le respect des contours de l'image. Lorsque l'on cherche à associer un pixel à un superpixel, SCALP considère la couleur des pixels voisins et l'information de couleur et de contour sur le chemin linéaire vers le barycentre du superpixel pour produire une décomposition précise et régulière.

La méthode SCALP est résumée dans la Figure 5.1, et nous comparons la méthode proposée aux méthodes de l'état-de-l'art dans la Figure 5.2. SCALP fournit un résultat plus satisfaisant qui respecte les contours de l'image. De plus, contrairement aux autres méthodes de l'état-de-l'art, SCALP est robuste au bruit, puisque la méthode génère des superpixels réguliers qui suivent les contours des objets, même sur la partie de l'image où un bruit blanc Gaussien a été ajouté.

5.2.1 Schéma de groupement itératif

L'algorithme SLIC (Achanta et al., 2012) consiste en un groupement itératif basé sur l'approche des K-moyennes, qui associe les pixels proches dans l'espace couleur CIELab. Le processus de décomposition démarre d'une grille régulière, composée de blocs de taille $r \times r$ pixels, avec $r = \sqrt{N/K}$, c'est-à-dire le rapport entre le nombre de pixels N et le nombre de superpixels désiré K. Chaque superpixel S_k est représenté par un descripteur C_k qui contient des informations moyennes sur les pixels $p \in S_k$: la couleur moyenne $F_k = [l_k, a_k, b_k]$, et le barycentre spatial de S_k, $X_k = [x_k, y_k]$, tel que $C_k = [F_k, X_k]$. Pour chaque descripteur C_k, tous les pixels à l'intérieur d'une fenêtre de taille $(2r+1)\times(2r+1)$ pixels centrée sur le barycentre X_k sont testés à chaque itération, pour être associés au superpixel S_k. La forme du superpixel est ainsi contrainte dans cette zone carrée, centrée sur son barycentre. L'initialisation en grille régulière et la fenêtre d'association des pixels ont été illustrées précédemment dans la Figure 4.6 et un pseudo code de notre méthode basée sur SLIC est donné dans l'Algorithme 2. Un pixel $p = [F_p, X_p]$ est comparé au descripteur C_k par une distance spatiale d_s, et une distance de couleur d_c telles que :

$$d_s(p, C_k) = (x_p - x_k)^2 + (y_p - y_k)^2, \tag{5.1}$$

$$d_c(p, C_k) = (l_p - l_k)^2 + (a_p - a_k)^2 + (b_p - b_k)^2. \tag{5.2}$$

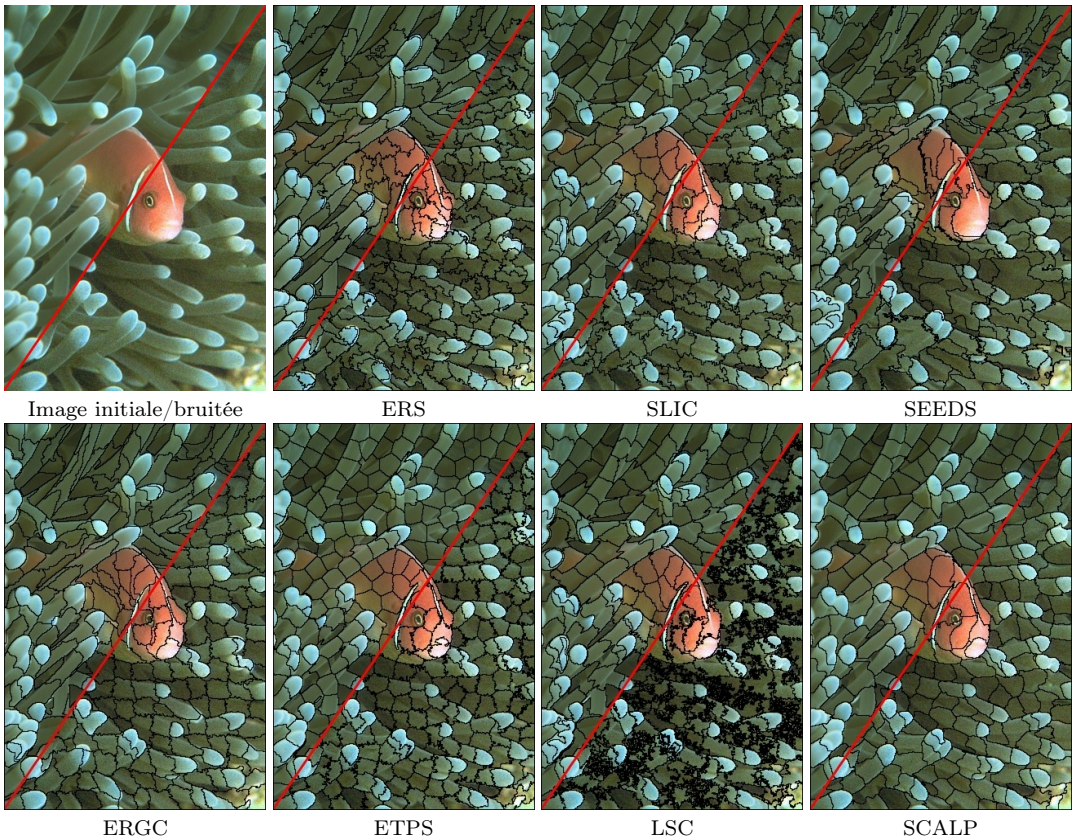


FIGURE 5.2 – Comparaison de la méthode proposée SCALP aux méthodes de l'état-de-l'art suivantes : ERS (Liu et al., 2011), SLIC (Achanta et al., 2012), SEEDS (Van den Bergh et al., 2012), ERGC (Buyssens et al., 2014), ETPS (Yao et al., 2015) et LSC (Chen et al., 2017). Un bruit blanc Gaussien a été ajouté à la partie bas-droite de l'image. La décomposition de SCALP suit les contours de l'image, et produit des superpixels réguliers, même sur la partie bruitée de l'image.

Un pixel p est associé à chaque itération au superpixel S_k minimisant la distance :

$$D(p, C_k) = d_c(p, C_k) + d_s(p, C_k)\frac{m^2}{r^2}, \tag{5.3}$$

avec m le paramètre de régularité qui fixe le compromis entre les distances de couleur et spatiale. De fortes valeurs de m produisent des superpixels plus réguliers, c'est-à-dire plus proches de la forme de la grille initiale, quand de faibles valeurs de m permettent aux superpixels de mieux coller aux contours de l'image, au détriment d'une plus grande variabilité dans la forme et la taille des superpixels.

Néanmoins, puisque ce paramètre est fixé globalement pour régler la régularité et que seule l'information de couleur du pixel est considérée dans (5.3), SLIC peut échouer à produire des superpixels respectant les contours de l'image, et obtient des performances limitées lorsque l'on souhaite optimiser toutes les propriétés des superpixels. Dans les sections suivantes, nous montrons comment améliorer le modèle pour obtenir des performances supérieures à celles des méthodes de l'état-de-l'art.

5.2.2 Information de couleur du voisinage

Dans la Figure 5.2, nous avons montré que la plupart des méthodes de superpixels de l'état-de-l'art n'étaient pas robustes au bruit, et pouvaient générer des décompositions très imprécises sur des images légèrement bruitées. Avec de telles méthodes, une étape de débruitage est nécessaire avant de pouvoir calculer une décomposition pertinente. L'approche de lignes de partage des eaux de Machairas et al. (2015) utilise par exemple une étape de pré-filtrage pour lisser les gradients locaux en fonction de la taille des superpixels. Néanmoins, cette étape n'est pas conçue pour gérer des images bruitées et son impact sur les performances de la méthode n'est pas reporté dans Machairas et al. (2015).

Nous proposons d'utiliser la couleur du voisinage des pixels dans le processus de décomposition afin d'améliorer la précision et la robustesse de SCALP. Considérer les couleurs des pixels voisins dans la distance (5.3) accroît le nombre de comparaison. Néanmoins, nous proposons une technique d'implémentation permettant d'intégrer cette information de voisinage dans le processus de décomposition en conservant la même complexité de calcul.

Distance sur le voisinage

De façon similaire aux approches basées sur les patchs, nous proposons de considérer l'information de couleur des pixels dans un voisinage carré, $P(p)$ centré sur le pixel p, dans la distance D_c :

$$D_c(p, C_k) = \sum_{q \in P(p)} (F_q - F_{C_k})^2 w_{p,q}. \tag{5.4}$$

Afin d'être robuste aux forts gradients locaux, tout en préservant les contours de l'image, nous définissons $w_{p,q}$ tel que :

$$w_{p,q} = \frac{1}{Z} \exp\left(-\frac{(F_p - F_q)^2}{2\sigma^2}\right), \tag{5.5}$$

avec Z le facteur de normalisation $Z = \sum_{q \in P(p)} \exp\left(-(F_p - F_q)^2/(2\sigma^2)\right)$, tel que $\sum_{q \in P(p)} w_{p,q} = 1$, et σ un paramètre fixé empiriquement.

Calcul rapide de la distance

La complexité de la distance proposée (5.4) est $\mathcal{O}(N)$, avec $N = (2n+1)^2 = |P(p)|$, le nombre de pixels dans le voisinage. Nous proposons alors une technique d'implémentation pour réduire la charge de calcul associée à (5.4). Puisque l'Éq. (5.4) calcule une différence entre les descripteurs d'un ensemble de pixels et celui d'un superpixel, elle peut être décomposée et partiellement pré-calculée.

Proposition 4. L'Éq. (5.4) peut être calculée avec une complexité $\mathcal{O}(1)$.

Démonstration. La distance entre les descripteurs F dans (5.4) s'écrit :

$$\begin{aligned}\sum_{q\in P(p)}(F_q - F_{C_k})^2 w_{p,q} &= \sum_{q\in P(p)}\left(F_q^2 + F_{C_k}^2 - 2F_qF_{C_k}\right)w_{p,q},\\ &= \sum_{q\in P(p)} F_q^2 w_{p,q} + \sum_{q\in P(p)} F_{C_k}^2 w_{p,q} - 2\sum_{q\in P(p)} F_qF_{C_k}w_{p,q},\\ &= \mathcal{F}_p{}^{(2)} + F_{C_k}^2\sum_{q\in P(p)} w_{p,q} - 2F_{C_k}\sum_{q\in P(p)} F_q w_{p,q},\\ &= \mathcal{F}_p{}^{(2)} + F_{C_k}^2 - 2F_{C_k}\mathcal{F}_p{}^{(1)}. \end{aligned} \tag{5.6}$$

Dans l'Éq. (5.6), les termes $\mathcal{F}_p{}^{(2)} = \sum_{q\in P(p)} F_q^2 w_{p,q}$, et $\mathcal{F}_p{}^{(1)} = \sum_{q\in P(p)} F_q w_{p,q}$, qui dépendent uniquement de l'image initiale, peuvent être pré-calculés. La complexité de la distance proposée D_c est ainsi réduite à $\mathcal{O}(1)$ au lieu de $\mathcal{O}(N)$. □

5.2.3 Information de couleur et contour sur le chemin linéaire

La qualité d'une décomposition est évaluée selon l'homogénéité de couleur de chaque superpixels, le respect des contours de l'image par les frontières, et par la régularité des formes produites. Pour améliorer ces deux derniers aspects, tout en respectant l'homogénéité de couleur, nous proposons de considérer l'information de couleur et de contour sur le chemin linéaire entre le pixel et le barycentre superpixel.

Définition du chemin linéaire

On définit le chemin $\mathbf{P}_p^k$, contenant les pixels entre la position $X_p = (x_p, y_p)$ d'un pixel p et le barycentre $X_k = (x_k, y_k)$ d'un superpixel S_k. Afin de limiter le nombre de pixels dans $\mathbf{P}_p^k$, l'algorithme de Bresenham (1965) est utilisé pour ne considérer que les pixels sur le chemin linéaire entre X_p et X_k, comme illustré dans la Figure 5.3. Les pixels considérés q (rouge) sont ceux qui intersectent le segment (flèche) entre X_p (bleu), et X_k (vert).

D'autres travaux considèrent une distance géodésique (Peyré et al., 2010), notamment pour améliorer l'homogénéité de couleur (Wang et al., 2013; Rubio et al., 2016). Les couleurs le long de la géodésique depuis un pixel doivent être proches de la couleur moyenne du superpixel, générant ainsi des formes irrégulières.

Nous illustrons cet aspect dans la Figure 5.4. Nous comparons une distance géodésique et la distance de couleur moyenne sur le chemin linéaire. Alors que la géodésique peut trouver un chemin sinueux en liant des pixels distants spatialement, notre chemin linéaire pénalise la traversée de pixels avec des couleurs différentes. Un exemple de décomposition de SCALP et d'une méthode basée sur une distance de couleur géodésique (Rubio et al., 2016) est donné dans la Figure 5.5. En considérant ce chemin linéaire, nous limitons le coût de calcul, qui peut être important pour calculer les distances géodésiques, et nous renforçons la convexité des superpixels.

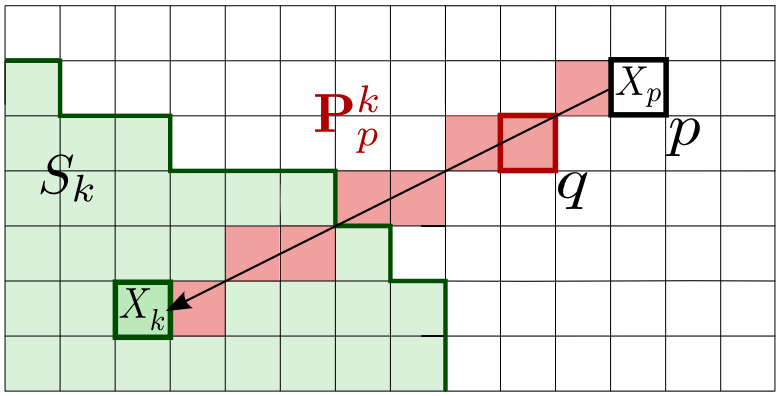
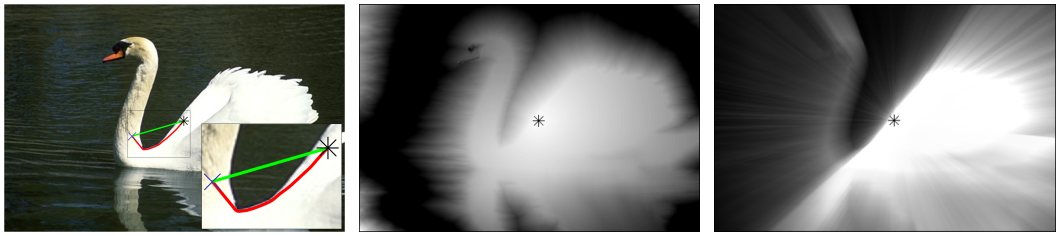


FIGURE 5.3 – Illustration du chemin linéaire $\mathbf{P}_p^k$ entre un pixel p et un superpixel S_k de barycentre X_k calculé avec Bresenham (1965). Le chemin $\mathbf{P}_p^k$ est uniquement composé des pixels q (rouges) qui intersectent le segment (flèche) entre X_p et X_k.



(a) Image (b) Distance géodésique (c) Distance chemin linéaire

FIGURE 5.4 – Comparaison du chemin géodésique (rouge) et du chemin linéaire (vert) entre la position d'un pixel initial (étoile) et final (croix) en (a). La distance géodésique et la distance de couleur moyenne sur le chemin linéaire du pixel initial à tous les pixels de l'image sont respectivement représentées en (b) et en (c). Notre chemin linéaire pénalise la traversée de régions avec des couleurs différentes.

(a) Image (b) Rubio et al. (2016) (géodésique) (c) SCALP (chemin linéaire)

FIGURE 5.5 – Exemple de résultat de décomposition d'une image (a) avec une méthode considérant une distance de couleur géodésique (Rubio et al., 2016) (b), et la méthode proposée SCALP, calculant une distance de couleur sur le chemin linéaire au barycentre du superpixel (c). SCALP décompose l'image en formes régulières quand la méthode basée sur une distance de couleur géodésique peut générer des superpixels très irréguliers.

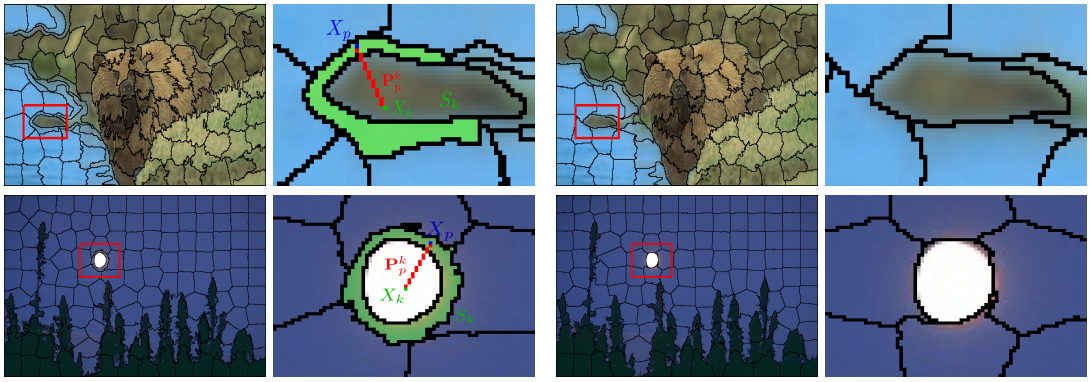
Distance de couleur

Dans Achanta et al. (2012), la distance à minimiser durant la décomposition est composée de deux termes couleur et spatial. Nous proposons de renforcer la précision de la distance couleur en la calculant également sur $\mathbf{P}_p^k$, c'est-à-dire, entre le descripteur du superpixel S_k et ceux des pixels sur le chemin linéaire entre p et le barycentre de S_k. La nouvelle distance de couleur est définie comme suit :

$$d_c(p, C_k, \mathbf{P}_p^k) = \lambda D_c(p, C_k) + (1-\lambda)\frac{1}{|\mathbf{P}_p^k|}\sum_{q \in \mathbf{P}_p^k} D_c(q, C_k), \tag{5.7}$$

où $\lambda \in [0, 1]$ pondère l'influence de la distance de couleur le long du chemin. Avec la distance proposée (5.7), la couleur sur le chemin linéaire au barycentre doit également être proche de la couleur moyenne du superpixel.

La distance (5.7) empêche l'apparition de formes irrégulières en favorisant la convexité des formes générées. Dans la Figure 5.6, nous montrons deux exemples de formes irrégulières, qui peuvent être calculées avec SLIC, par exemple dans des zones de dégradé de couleurs. Sans le chemin linéaire, les barycentres X_k ne sont pas contenus à l'intérieur de ces superpixels irréguliers S_k. Lorsque les pixels $q \in \mathbf{P}_p^k$ ont des couleurs très éloignées de la couleur moyenne du superpixel S_k, l'éq. (5.7) pénalise l'association du pixel p à ce superpixel durant l'itération suivante.



(a) (b)

FIGURE 5.6 – Illustration de l'impact de la distance de couleur proposée sur la régularité de la décomposition. Deux exemples de formes irrégulières obtenues avec SLIC (a), et rendues régulières avec la distance de couleur sur le chemin linéaire de SCALP (5.7) (b) sont représentées. Dans le cas des formes irrégulières, les barycentres ne sont pas contenus à l'intérieur des superpixels, et le chemin linéaire capture alors des couleurs éloignées de la couleur moyenne du superpixel.

Adhérence à la carte de contours

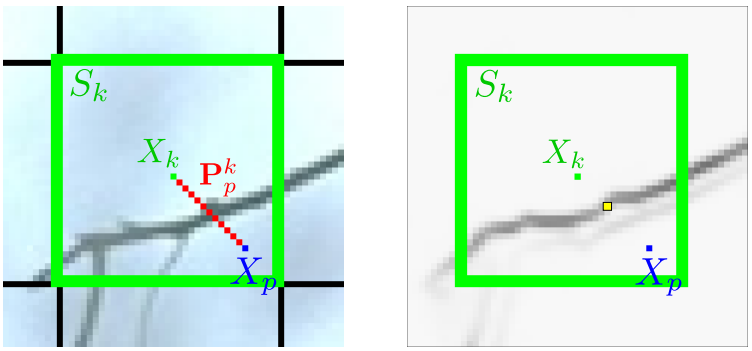
Puisque l'homogénéité de couleur optimale peut ne pas correspondre au respect des objets de l'image (voir Figure 4.7), une carte de contour $\mathcal{C}$ peut également être exploitée sur le chemin linéaire. Cette carte est telle que $\mathcal{C}(p) \approx 1$ si un contour est détecté au pixel p, et ≈ 0 sinon. Nous proposons un moyen rapide et efficace d'intégrer l'information de contour en pondérant la distance entre un pixel et un superpixel par un nouveau terme $d_{\mathcal{C}}(\mathbf{P}_p^k)$, qui considère le maximum d'amplitude de contour sur $\mathbf{P}_p^k$, pondéré par un paramètre $\gamma \geq 0$:

$$d_{\mathcal{C}}(\mathbf{P}_p^k) = 1 + \gamma \max_{q \in \mathbf{P}_p^k} \mathcal{C}(q). \tag{5.8}$$

La Figure 5.7 illustre la sélection du maximum d'amplitude de contour sur le chemin linéaire. Lorsqu'une forte amplitude de contour se trouve sur le chemin entre p et le barycentre de S_k, l'Éq. (5.8) empêche ce pixel d'être associé au superpixel.

Quand d'autres méthodes considèrent une carte binaire de contours, par exemple Wang et al. (2017a), SCALP permet de considérer n'importe quelle carte de contours, binaire, ou ayant des intensités entre 0 et 1. La méthode proposée permet également de s'adapter à des contours épais puisque seul le maximum d'intensité est considéré sur le chemin.

(a) Grille régulière initiale (b) Carte de contours



(c) Chemin linéaire $\mathbf{P}_p^k$ (d) Contour maximal sur $\mathbf{P}_p^k$

FIGURE 5.7 – Illustration de la première itération à partir d'une grille initiale (a) et utilisant une carte de contours (b). Le chemin linéaire $\mathbf{P}_p^k$ est défini pour un pixel p et un superpixel S_k (c), et l'intensité maximale de contour (pixel jaune) est considérée pour favoriser le respect des contours des objets de l'image par les frontières des superpixels (d).

Enfin, nous multiplions ce terme aux distances de couleur et spatiale afin d'assurer le respect des contours, et la distance proposée D à minimiser pour chaque pixel durant le processus de décomposition est définie comme suit :

$$D(p, C_k) = \left(d_c(p, C_k, \mathbf{P}_p^k) + d_s(p, C_k)\frac{m^2}{r^2} \right) d_{\mathcal{C}}(\mathbf{P}_p^k), \tag{5.9}$$

avec la distance spatiale d_s calculée par (5.1).

La méthode SCALP est résumée dans l'Algorithme 2 pour une image I décomposée en K superpixels. Notons que contrairement à d'autres méthodes nécessitant des adaptations conséquentes pour pouvoir être appliquées à des images 3D, SCALP s'étend naturellement au calcul de décompositions en supervoxels. La méthode considère une grille régulière 3D, et ajoute une dimension aux équations précédentes.

Algorithme 2 SCALP$(I, K, \mathcal{C})$

1: Initialisation des descripteurs $C_k \leftarrow [F_k, X_k]$ depuis une grille régulière
2: Initialisation des étiquettes des superpixels $\mathcal{S} \leftarrow 0$
3: Pré-calcul des descripteurs ${\mathcal{F}_p}^{(2)}$ et ${\mathcal{F}_p}^{(1)}$ (5.6)
4: **Pour** *chaque itération* **faire**
5: Distance $d \leftarrow \infty$
6: **Pour** *chaque* C_k **faire**
7: **Pour** *chaque pixel* p *dans une fenêtre de* $(2r+1)\times(2r+1)$ *centrée sur* X_k **faire**
8: Calcul du chemin linéaire $\mathbf{P}_p^k$ (Bresenham, 1965)
9: Calcul de $D(p, C_k)$ en utilisant $\mathcal{C}$ et $\mathbf{P}_p^k$ avec (5.9)
10: **Si** $D(p, C_k) < d(p)$ **alors**
11: $d(p) \leftarrow D(p, C_k)$
12: $\mathcal{S}(p) \leftarrow k$
13: **Pour** *chaque* C_k **faire**
14: Actualiser $[F_k, X_k]$
15: **Retourner** $\mathcal{S}$

5.3 Contrainte depuis une segmentation initiale

Dans cette section, nous proposons une méthode pour utiliser une carte de contours comme segmentation initiale pour la décomposition en superpixels. Générer une segmentation de l'image en régions depuis une carte de contours nécessite certaines étapes supplémentaires, mais peut permettre d'améliorer les performances de décomposition, notamment de respect des objets de l'image. L'idée est de tirer un profit maximal de la détection de contours en conservant la segmentation des objets, tout en générant une décomposition en superpixels réguliers qui s'adaptent localement au contenu de l'image. Notons que des méthodes telles que Arbelaez et Cohen (2008); Arbelaez et al. (2009) permettent de segmenter une image en régions depuis une carte de contours, mais ne contrôlent ni la taille, ni la forme, ni le nombre des régions, et ne peuvent donc pas être considérées comme des méthodes de superpixels.

Nous proposons ici une méthode utilisant une segmentation initiale, pour produire une décomposition régulière en superpixels au sein des régions pré-segmentées, avec contrôle sur le nombre d'éléments. De cette façon, nous profitons de la précision de la segmentation initiale tout en générant une décomposition de l'image en superpixels réguliers. En initialisant la décomposition par rapport aux régions pré-calculées, les superpixels initiaux s'adaptent mieux au contenu de l'image. Par exemple, de petites régions peuvent être initialement segmentées en un ou plusieurs superpixels. Sans initialisation adaptée, ces régions peuvent se trouver entre deux barycentres de superpixels initiaux, et ne pas être précisément capturées par le processus de décomposition.

5.3.1 Segmentation hiérarchique depuis une détection de contours

Afin d'adapter une segmentation initiale pour produire des superpixels réguliers, nous proposons d'utiliser une segmentation hiérarchique, qui peut être calculée à partir d'une carte de contours avec des méthodes telles que Arbelaez et Cohen (2008); Arbelaez et al. (2009).

Notée $\mathcal{U}$, une segmentation hiérarchique est définie comme une carte de probabilité de contours. Pour tout seuil, elle produit un ensemble de courbes fermées et donc de régions. Les régions segmentées avec une faible probabilité, c'est-à-dire dont le contour dans $\mathcal{U}$ a une faible intensité, peuvent donc être supprimées avec une étape de seuillage. La carte de contours fermés seuillée est notée $\mathcal{U}_\tau$, pour un seuil τ, et sa décomposition correspondante en régions est notée $\mathcal{R} = \{R_i\}$. La Figure 5.8, illustre le résultat de segmentation obtenu depuis une segmentation hiérarchique pour différents seuils.

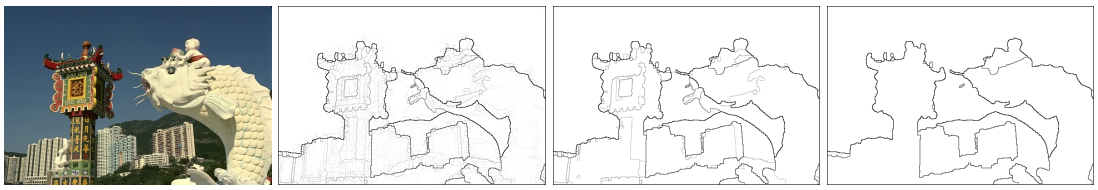
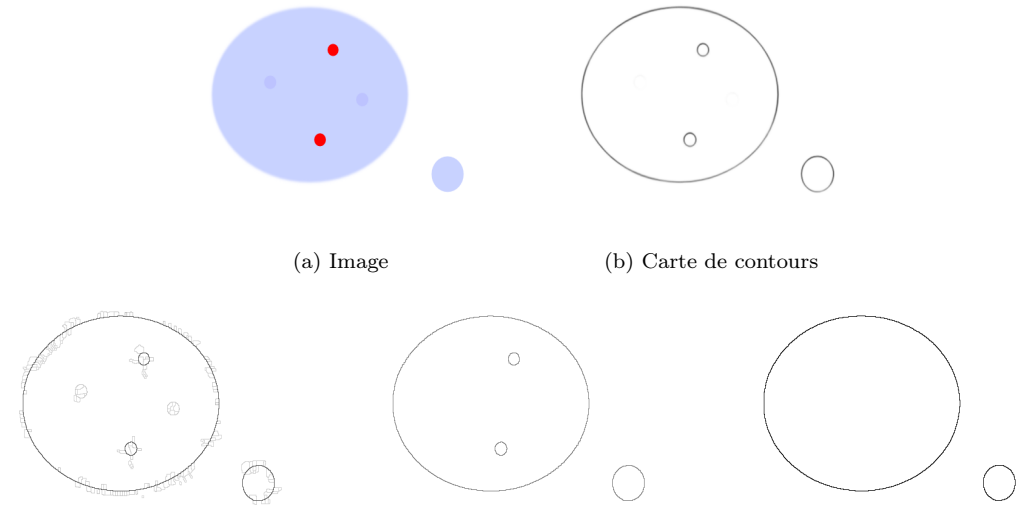


Image $\tau = 0$ $\tau = 0.2$ $\tau = 0.6$

FIGURE 5.8 – Exemple de segmentation hiérarchique calculée avec Arbelaez et al. (2009) depuis une carte de contours obtenue avec Dollár et Zitnick (2013). La segmentation hiérarchique est illustrée pour différentes valeurs du paramètre de seuillage τ.

5.3.2 Décomposition régulière en superpixels depuis une segmentation hiérarchique

La segmentation hiérarchique obtenue, une étape de fusion peut être effectuée pour supprimer les plus petites zones. Ces petites régions doivent être fusionnées



(a) Image (b) Carte de contours

(c) Segmentation hiérarchique (d) Seuillage (e) Fusion

FIGURE 5.9 – Illustration des étapes de seuillage et de fusion de la segmentation hiérarchique (c) calculée à partir de la carte de contours (b) d'une image (a). L'étape de seuillage (d) permet d'éliminer les zones segmentées avec une faible probabilité, c'est-à-dire, les petits cercles bleus et les artefacts de segmentation. Ensuite, selon la condition dans (5.10), les plus petites régions sont supprimées (e), c'est-à-dire, les cercles rouges, bien qu'ils aient une probabilité de segmentation plus élevée que les bleus.

à une région adjacente pour respecter la régularité de la décomposition en termes de taille. Avec K le nombre de superpixels et $|I|$ le nombre de pixels de l'image I, la taille moyenne des superpixels est $s = |I|/K$. Un seuil $t \in [0, 1]$ est défini pour fusionner les régions contenant moins de $s \times t$ pixels. La probabilité de segmentation d'une région R_i est la plus faible intensité de détection $\mathcal{U}_\tau(p)$ parmi ses pixels en bordure $p \in \mathcal{B}(R_i)$. La région R_i est donc fusionnée à la région adjacente R_j qui partage la bordure avec la plus faible probabilité de segmentation :

$$\text{si} \quad |R_i| < s \times t, \qquad \mathcal{R}(R_i) = \underset{j, p \in \mathcal{B}(R_i) \cap \mathcal{B}(R_j)}{\operatorname{argmin}} \mathcal{U}_\tau(p). \tag{5.10}$$

La Figure 5.9 illustre ces étapes. Le seuillage supprime les zones segmentées avec une faible probabilité et la fusion empêche la segmentation de petites régions.

Une étape (partition) ajoute ensuite de nouveaux superpixels dans les régions restantes. Si le nombre de régions est inférieur au nombre de superpixels K, des superpixels sont ajoutés en fonction de la taille de la région $|R_i|$. Dans une région R_i, $\lfloor |R_i|/s \rfloor$ sous-régions sont initialisées par une approche basée sur un algorithme de K-moyennes spatial (Lloyd, 1982), indépendamment de l'information de couleur. La méthode proposée utilisant la carte de contours comme contrainte dure de segmentation est illustrée dans la Figure 5.10, et sera désignée SCALP+HC dans la suite.

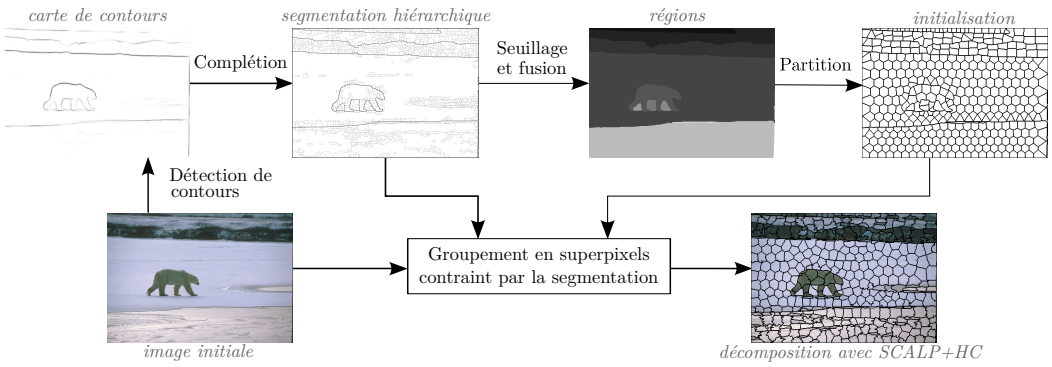


FIGURE 5.10 – Méthode SCALP+HC utilisant la carte de contours comme contrainte dure pour fournir une segmentation initiale. Une étape de complétion est effectuée pour produire une segmentation hiérarchique depuis la carte de contours. Les régions segmentées avec une faible probabilité sont éliminées par une étape de seuillage, et les régions trop petites par rapport à la taille donnée des superpixels sont fusionnées aux régions adjacentes. Ces régions peuvent ensuite être partitionnées pour fournir une meilleure initialisation en superpixels. SCALP est effectué indépendamment dans chaque région, en profitant de la précision de détection de contours, tout en produisant une décomposition régulière qui s'adapte bien au contenu local de l'image.

Cette approche s'adapte donc à la taille des superpixels et est modérément sensible aux paramètres de seuillage. Notons que bien que nous considérons ici la segmentation comme une contrainte dure pour assurer le respect des objets de l'image, la segmentation initiale peut être utilisée uniquement pour initialiser la répartition en superpixels, au lieu d'utiliser une grille régulière.

5.4 Validation de SCALP sur les métriques superpixels standards

5.4.1 Cadre de validation

Base de données

Nous considérons la base d'images standard BSD (Martin et al., 2001) pour évaluer la méthode proposée SCALP et la comparer aux méthodes de l'état-de-l'art. Cette base contient 200 images de test de taille 321×481 pixels. Au moins 5 décompositions de vérité terrain sont fournies par image pour calculer des performances de respect des objets de l'image, et d'adhérence aux contours.

Réglage des paramètres

SCALP a été implémenté avec MATLAB via du code C-MEX. Nous considérons dans d_c (5.4) et d_s (5.1) les descripteurs de Chen et al. (2017), plus performants que ceux de SLIC, et définis dans un espace de plus haute dimension (6 pour la

couleur, et 4 pour la position spatiale). Dans (5.4), le paramètre σ est fixé à 40 et $P(p)$ est défini comme un voisinage de taille 7×7 pixels ($n = 3$) autour d'un pixel p. Ces paramètres sont empiriquement réglés pour améliorer la précision de la décomposition à la fois sur les images initiales et celles modérément bruitées. Dans la distance de couleur proposée (5.7), λ est réglé à 0.5, et γ à 50 dans (5.8). Le paramètre de compacité m^2 est fixé à $0.075r^2$ dans la distance finale (5.9), comme dans Chen et al. (2017). Ce paramètre offre un bon compromis entre adhérence aux contours et régularité des superpixels.

Le nombre d'itérations du processus de décomposition est fixé à 5, contrairement à Chen et al. (2017) qui utilise 20 itérations, puisque notre méthode converge plus rapidement. Sauf mention contraire, lorsqu'elle est utilisée, la carte de contours considérée par SCALP est calculée avec la méthode de Dollár et Zitnick (2013). Lorsque la carte de contours est utilisée comme contrainte dure (SCALP+HC), nous fixons respectivement les paramètres t et τ dans (5.10) à 0.15 et 0.4, et la segmentation hiérarchique est calculée avec Arbelaez et al. (2009). Comme dans Van den Bergh et al. (2012), les résultats des courbes PR sont générés depuis la moyenne des bordures des superpixels obtenus à plusieurs échelles, allant de 25 à 1000 superpixels. Dans la suite, lorsque des résultats sont donnés pour des images bruitées, celles-ci ont été altérées par un bruit blanc Gaussien de variance 20.

5.4.2 Impact des contributions

Paramètres de distance

Nous mesurons d'abord l'influence des paramètres dans la distance dans (5.9) sur les performances de SCALP. Dans la Figure 5.11, nous donnons les résultats sur les courbes PR avec la mesure F maximale (4.11), ASA (4.6), CD (4.10) par rapport à BR (4.7) et EV (4.3) pour différents paramètres de distance, pour les images initiales et bruitées de la BSD. Tout d'abord, notons que l'utilisation du voisinage ($n = 3$ dans (5.4)) augmente la précision de la décomposition, notamment l'adhérence aux contours (courbe BR par rapport à CD). L'utilisation de l'information de couleur ($\lambda = 0.5$ dans (5.7)) améliore également les résultats en termes de respect des objets de l'image, d'adhérence et de détection de contours. Enfin, l'utilisation de la carte de contours ($\gamma = 50$ dans (5.8)) sur le chemin linéaire permet d'atteindre de hautes performances de détection de contours (courbe PR) et augmente les performances sur les métriques de superpixels ASA et BR par rapport à CD sur les images initiales.

Sur les trois premières métriques, les contributions améliorent chacune les performances de décomposition. Pour les images bruitées, la carte de contours est moins précise et n'apporte qu'un gain en termes de respect des objets de l'image (ASA). Logiquement, les performances d'homogénéité de couleur, évaluées avec EV décroissent légèrement à mesure que de nouveaux termes, autres que la simple distance de couleur du pixel au superpixel, sont ajoutés au modèle. Néanmoins, dans le cas bruité, la prise en compte du voisinage suffit à compenser la chute des performances.

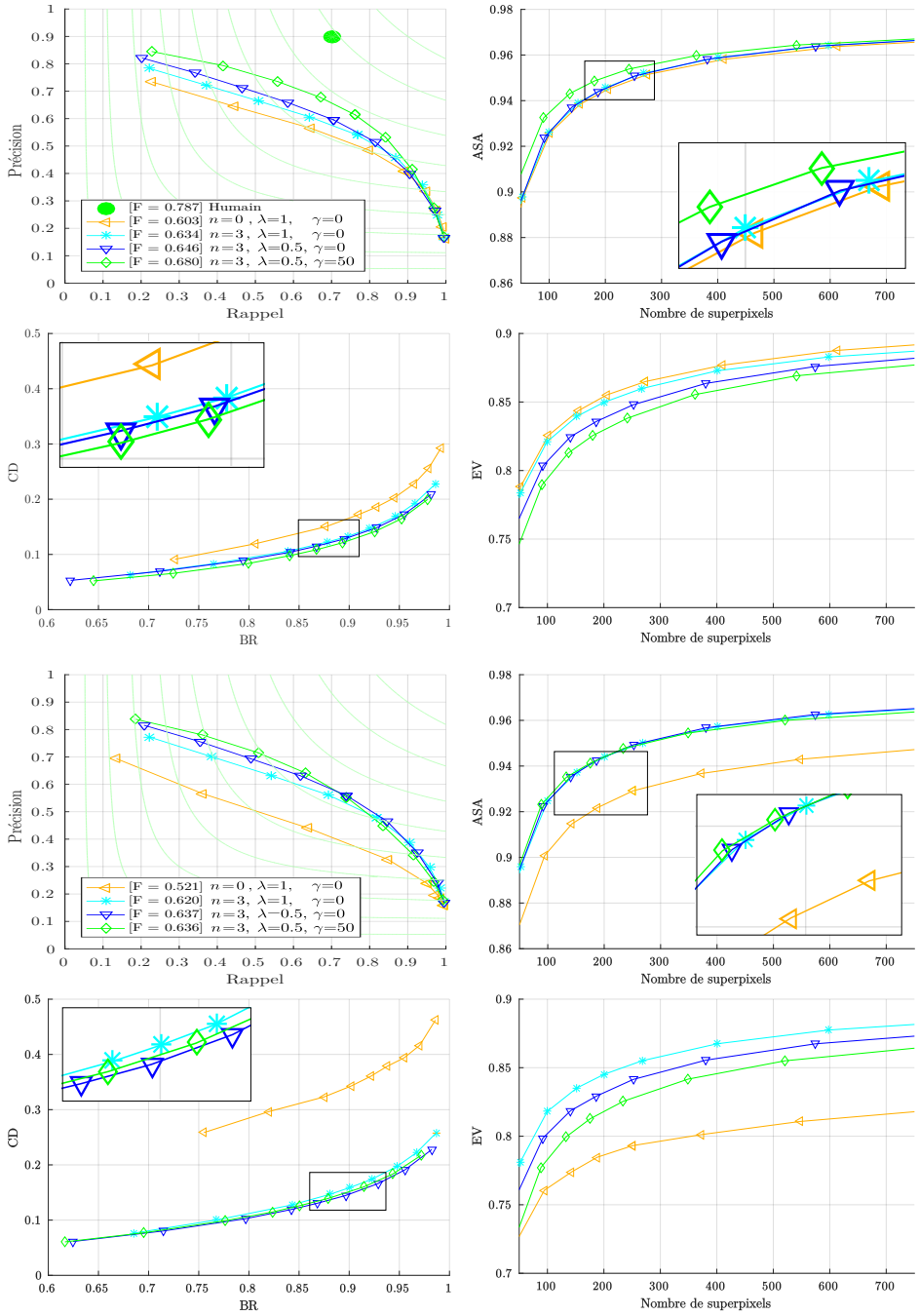


FIGURE 5.11 – Évaluation des paramètres de distance de SCALP sur les métriques PR, ASA, CD par rapport à BR et EV pour les images initiales (en haut) et bruitées (en bas). Chaque contribution augmente la précision de la décomposition à la fois pour les images initiales et bruitées. Le paramètre n définit l'utilisation du voisinage dans la distance, et λ et γ règlent respectivement l'influence de la couleur et de la carte de contours le long du chemin linéaire. Voir le texte pour plus de détails.

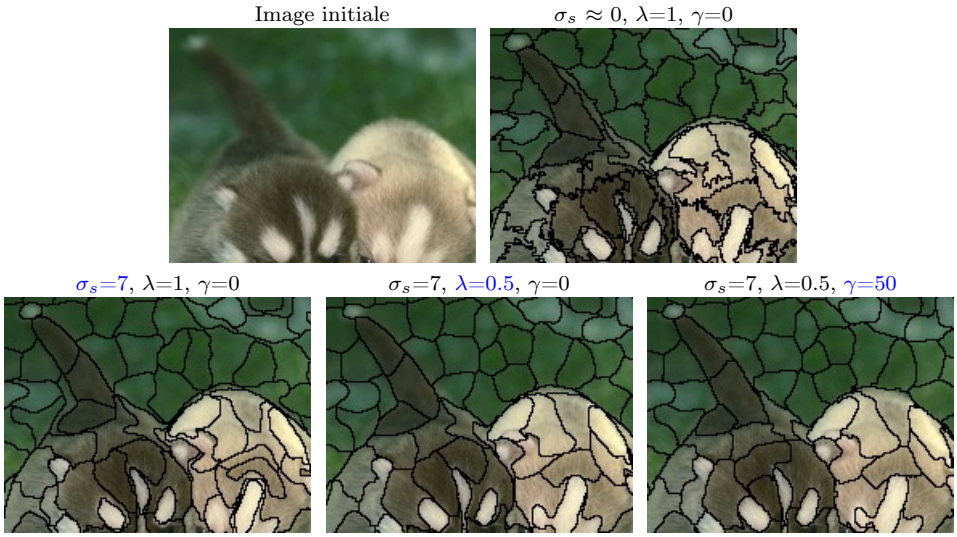


FIGURE 5.12 – Impact visuel des paramètres de distance sur une image de la BSD. Chaque contribution augmente progressivement la précision de la décomposition en ajoutant des descripteurs pertinents et complémentaires.

La Figure 5.12 montre les résultats de décomposition pour ces différents paramètres de distance sur un exemple. Avec seulement les descripteurs utilisés dans Chen et al. (2017), c'est-à-dire, avec $n = 0$, $\lambda = 1$, $\gamma = 0$, les contours des superpixels sont peu précis et très irréguliers. L'information de voisinage ($n = 3$) réduit considérablement le bruit en bordure des superpixels. La distance de couleur sur le chemin linéaire ($\lambda = 0.5$) améliore la régularité des superpixels. Enfin, l'information de contour ($\gamma = 50$) permet de capturer plus efficacement les structures des objets et de respecter leurs contours.

Carte de contours

Nous étudions également l'influence de la carte de contours. De nombreuses méthodes efficaces de détection de contours ont été proposées dans la littérature (voir les références dans Arbelaez et al. (2011)). Les performances de notre méthode avec l'utilisation d'une carte de contours sont logiquement corrélées à la précision de la carte utilisée. Néanmoins, nous démontrons que des performances supérieures sont obtenues, même avec une détection de contours basique.

Un moyen rapide d'obtenir une détection de contours, simple mais précise, consiste à moyenner les bordures des décompositions en superpixels obtenues à plusieurs échelles. Nous proposons de considérer le même ensemble d'échelles $\{\mathcal{K}\}$ utilisé pour calculer les courbes PR. Les frontières des superpixels résultant $\mathcal{B}(\mathcal{S}^K)$ d'une décomposition $\mathcal{S}^K$, calculée à l'échelle $K \in \{\mathcal{K}\}$ sont moyennées dans $\bar{\mathcal{B}}$:

$$\bar{\mathcal{B}} = \frac{1}{|\{\mathcal{K}\}|} \sum_{K \in \{\mathcal{K}\}} \mathcal{B}(\mathcal{S}^K). \tag{5.11}$$

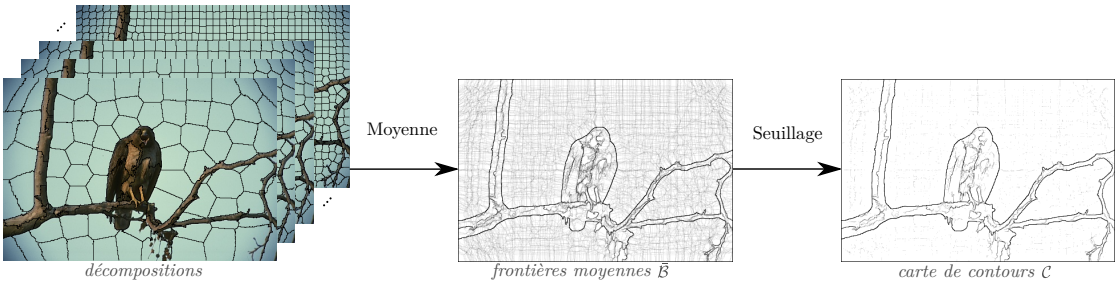


FIGURE 5.13 – Illustration de la détection de contours depuis les frontières des superpixels obtenus à plusieurs échelles. Les frontières sont moyennées et seuillées pour fournir, d'une façon simple et rapide, une carte de contours précise. Le seuil de la carte de contours (5.11) est fixé à 0.5.

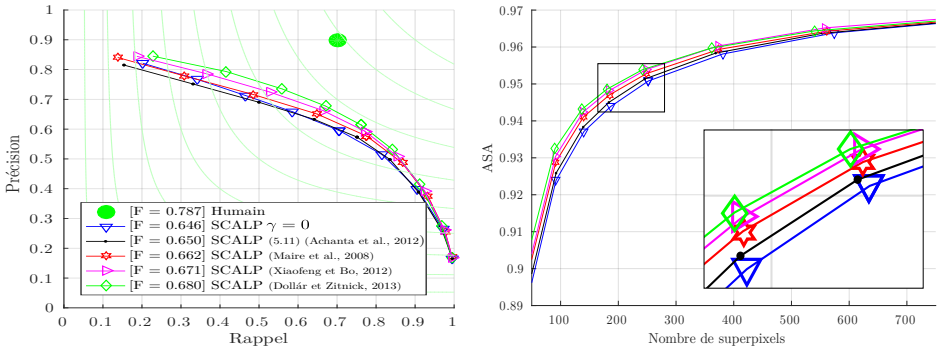


FIGURE 5.14 – Évaluation des performances de SCALP pour différentes méthodes de détection de contours. Toutes les méthodes utilisées permettent d'améliorer la précision, et même une détection de contours simple obtenue à partir des frontières des superpixels calculés à plusieurs échelles (5.11) améliore l'adhérence aux contours de l'image.

La moyenne $\bar{\mathcal{B}}$ à valeurs dans $\{0, \ldots, 1\}$ peut alors être seuillée pour supprimer les contours détectés avec une faible probabilité, c'est-à-dire, détectés à peu d'échelles, pour fournir une carte de contours précise $\mathcal{C}$. La Figure 5.13 illustre le calcul de la carte de contours $\mathcal{C}$ depuis les frontières des superpixels calculés avec Achanta et al. (2012). Notons que les décompositions à plusieurs échelles sont indépendantes et peuvent être calculées rapidement en parallèle.

Dans la Figure 5.14, nous donnons les résultats obtenus avec différentes méthodes de détection : la détection de contours obtenue depuis les décompositions de superpixels à plusieurs échelles avec un seuil de la carte de contours fixé à 0.5 (5.11), l'algorithme de probabilité globale de contours (Maire et al., 2008), une méthode d'apprentissage parcimonieuse de patchs de gradients (Xiaofeng et Bo, 2012), et une approche basée sur les forêts structurées (Dollár et Zitnick, 2013). Les résultats sont améliorés avec la précision de la détection de contours. Néanmoins, nous notons que même une simple détection de contours avec (5.11) permet d'améliorer les performances. Dans la suite, les résultats donnés pour SCALP sont calculés en utilisant Dollár et Zitnick (2013).

5.4.3 Comparaison avec les méthodes de l'état-de-l'art

Nous comparons la méthode SCALP aux approches populaires de l'état-de-l'art : ERS (Liu et al., 2011), SLIC (Achanta et al., 2012), SEEDS (Van den Bergh et al., 2012), ERGC (Buyssens et al., 2014), Waterpixels (WP) (Machairas et al., 2015), ETPS (Yao et al., 2015) and LSC (Chen et al., 2017). Les résultats donnés pour ces méthodes sont calculés avec les codes fournis par les auteurs, et les méthodes sont utilisées avec leurs paramètres par défaut.

Dans la Figure 5.15, nous donnons les performances pour les images initiales (haut) et bruitées (bas). SCALP surpasse clairement les méthodes comparées en termes de respect des objets de l'image et de leurs contours, donnant par exemple la plus haute mesure F (0.680), tout en produisant des superpixels aux contours lisses mais précis (faible CD pour fort BR). L'ASA évalue la conformité de la décomposition en superpixels par rapport aux objets de l'image. Ainsi, les meilleures performances d'ASA obtenues avec SCALP indiquent que les superpixels sont mieux contenus dans les objets de l'image.

En utilisant la carte de contours comme contrainte (SCALP+HC), notre méthode atteint même de plus hautes performances, avec F=0.709. Nous pouvons noter que les résultats de SCALP obtenus sans utiliser de carte de contours, c'est-à-dire, $\gamma = 0$ dans (5.8), surpassent toujours ceux des méthodes de l'état-de-l'art les plus précises ERS (Liu et al., 2011) et LSC (Chen et al., 2017). Enfin, malgré les contraintes de respect des contours et de régularité introduites dans SCALP, la méthode reste parmi les toutes meilleures en termes d'homogénéité de couleur des superpixels.

Le gain par rapport aux méthodes de l'état-de-l'art est encore plus large sur les images bruitées. Des méthodes telles que ERS (Liu et al., 2011), ERGC (Buyssens et al., 2014) ou LSC (Chen et al., 2017) obtiennent des performances très dégradées lorsqu'elles sont appliquées à des images légèrement bruitées, tandis que Van den Bergh et al. (2012); Yao et al. (2015) sont les seules autres méthodes robustes au bruit sur certains aspects évalués. Les méthodes de l'état-de-l'art peuvent en effet avoir des comportements très différents lorsqu'elles sont appliquées à des images bruitées. Elles produisent généralement des superpixels très bruitées aux contours (voir Figure 5.2). Cet aspect est notamment illustré par les performances dégradées de CD par rapport à BR dans la partie basse de la Figure 5.15. Enfin, sur la métrique ASA, SCALP fournit des résultats légèrement supérieurs à ceux de SCALP+HC pour ces images. La présence de bruit peut en effet diminuer la précision de la détection de contours qui ne doit pas être considérée comme une contrainte dure pour assurer le respect des objets. Ces résultats sont résumés dans le Tableau 5.1, où nous indiquons les performances de toutes les méthodes comparées sur les images initiales et bruitées pour $K = 250$ superpixels.

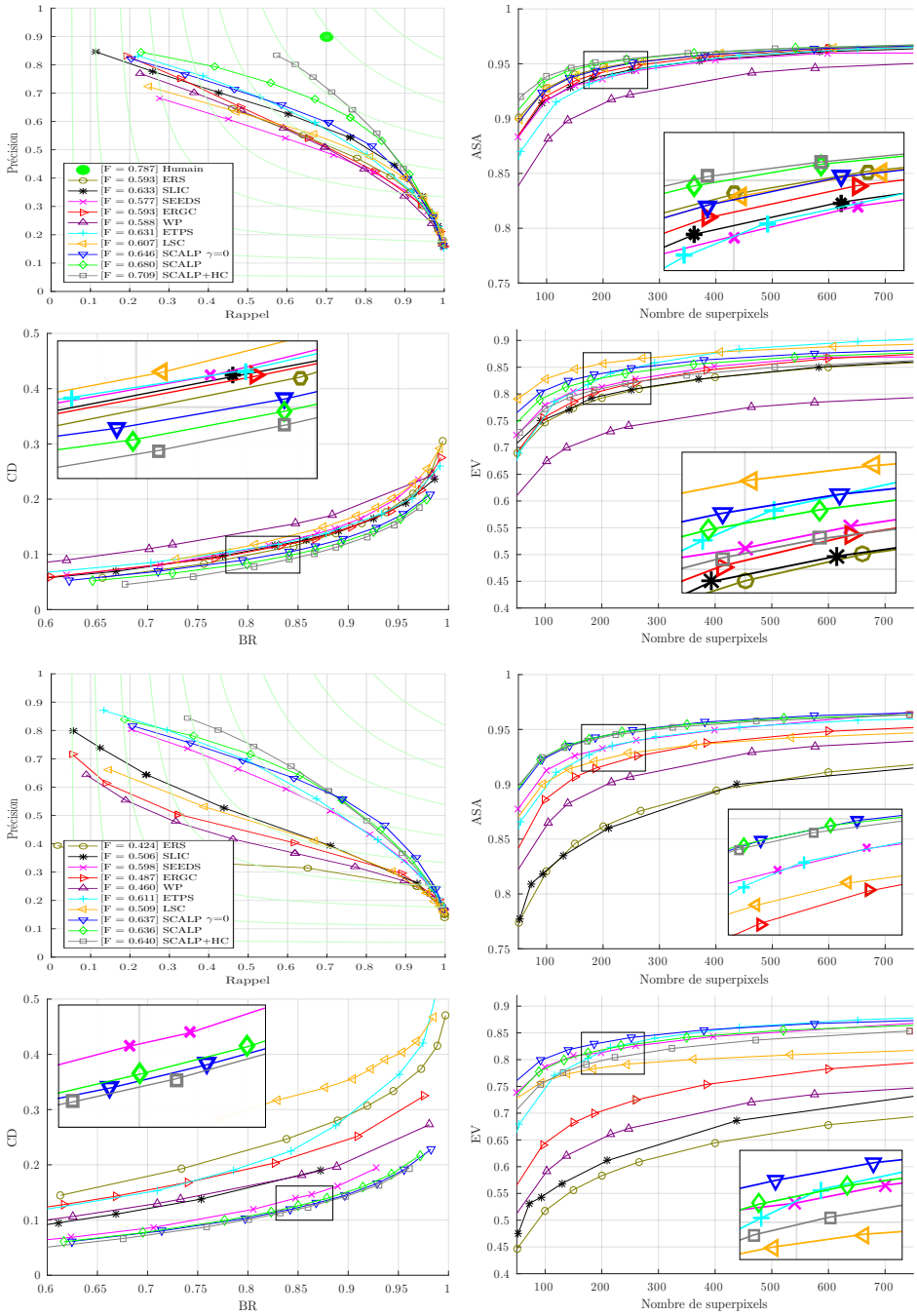


FIGURE 5.15 – Comparaison entre SCALP et les méthodes de l'état-de-l'art sur les performances de détection de contours (PR) et les métriques de superpixels standards (ASA, CD par rapport à BR et EV) sur les images de la BSD. SCALP surpasse les autres méthodes sur les images initiales (en haut) et bruitées (en bas). De plus, les résultats de SCALP sans l'utilisation d'une carte de contours (SCALP $\gamma = 0$), surpassent toujours les méthodes comparées les plus précises ERS (Liu et al., 2011) et LSC (Chen et al., 2017).

Tableau 5.1 – Comparaison de SCALP aux méthodes de l'état-de-l'art sur les images initiales|bruitées. La mesure F maximale (4.11) est calculée comme présentée dans le chapitre 4. Les résultats d'ASA (4.6) et d'EV (4.3) sont donnés pour $K = 250$ superpixels, et les résultats de CD (4.10) sont donnés pour BR = 0.8 (4.7). Les couleurs bleue (en gras) et rouge (souligné) indiquent respectivement les premier et second meilleurs résultats.

Méthode	F	ASA	CD/BR	EV
ERS	0.593|0.424	0.951|0.872	0.099|0.227	0.806|0.603
SLIC	0.633|0.506	0.944|0.867	0.106|0.156	0.808|0.624
SEEDS	0.577|0.598	0.943|0.939	0.109|0.118	0.825|0.824
ERGC	0.593|0.487	0.948|0.924	0.104|0.192	0.819|0.721
WP	0.588|0.460	0.932|0.907	0.124|0.162	0.740|0.672
ETPS	0.631|0.509	0.943|0.939	0.110|0.199	0.847|**0.830**
LSC	0.607|0.611	0.950|0.929	0.115|0.300	**0.864**|0.792
SCALP	0.680|0.636	0.954|**0.949**	0.084|0.107	0.839|0.828
SCALP+HC	**0.709**|**0.640**	**0.955**|0.947	**0.076**|**0.101**	0.821|0.808

Les Figures 5.16 et 5.17 illustrent respectivement les décompositions en superpixels obtenues avec SCALP et les meilleures méthodes comparées sur des images initiales et bruitées de la BSD. SCALP fournit des superpixels réguliers qui suivent précisément les contours de l'image. SCALP+HC permet de guider de façon plus précise la décomposition en contraignant les superpixels dans les régions pré-segmentées.

Enfin, le temps de calcul de SCALP reste raisonnable, c'est-à-dire moins de 0.6s sur les images de la BSD, sans architecture d'implémentation multi-processus, contrairement à d'autres implémentations telles que Yao et al. (2015). Ce temps de calcul correspond à ceux des méthodes de superpixels standards, et est même inférieur à ceux de méthodes telles que Levinshtein et al. (2009); Liu et al. (2011) qui peuvent aller jusqu'à 5s. Dans ce travail, nous nous concentrons uniquement sur les performances des décompositions et nous ne comparons pas de façon approfondie les temps de calcul, puisque leur mesure dépend grandement de l'implémentation de chaque méthode (Stutz et al., 2016). Notre méthode est basée sur l'approche d'Achanta et al. (2012), et de récents travaux ont montré que cet algorithme pouvait être implémenté en temps-réel (Neubert et Protzel, 2014; Ban et al., 2016; Choi et Oh, 2016). Ainsi, SCALP ayant la même complexité calculatoire, notre méthode peut potentiellement atteindre de telles performances avec une implémentation optimisée.

5.4.4 Extension aux supervoxels

Enfin, nous étendons de façon naturelle la méthode SCALP au calcul des supervoxels sur des volumes 3D, et appliquons notre méthode aux images médicales 3D. De nombreuses méthodes supervoxels sont consacrées à la segmentation de vidéos (voir par exemple les références de Xu et Corso (2012)). Ces méthodes segmentent

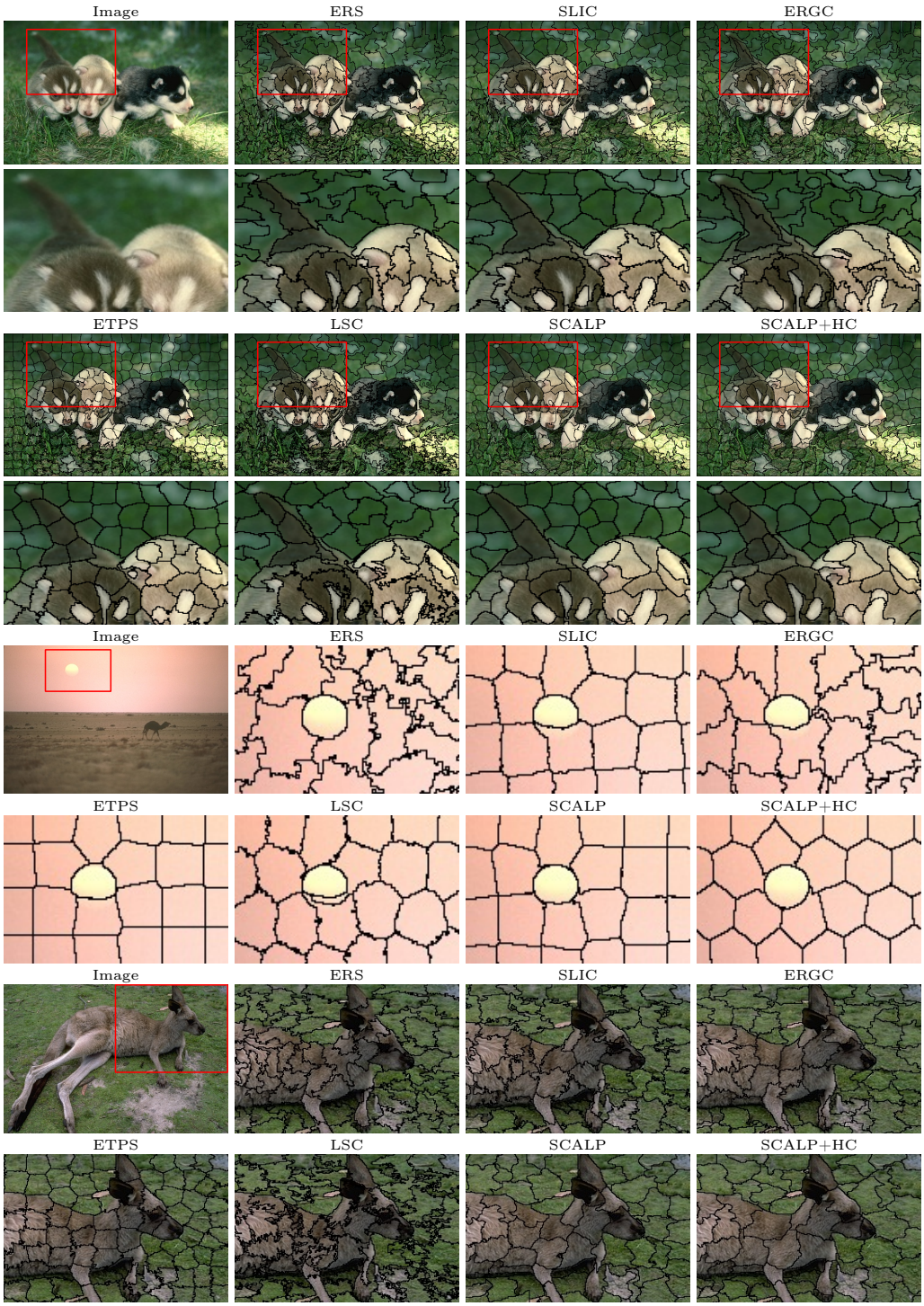


FIGURE 5.16 – Comparaison qualitative des décompositions de SCALP et des méthodes de l'état-de-l'art sur des images de la BSD. SCALP produit les résultats les plus satisfaisants, avec des superpixels réguliers qui adhèrent aux contours de l'image. SCALP+HC permet d'encore améliorer le respect des objets de l'image.

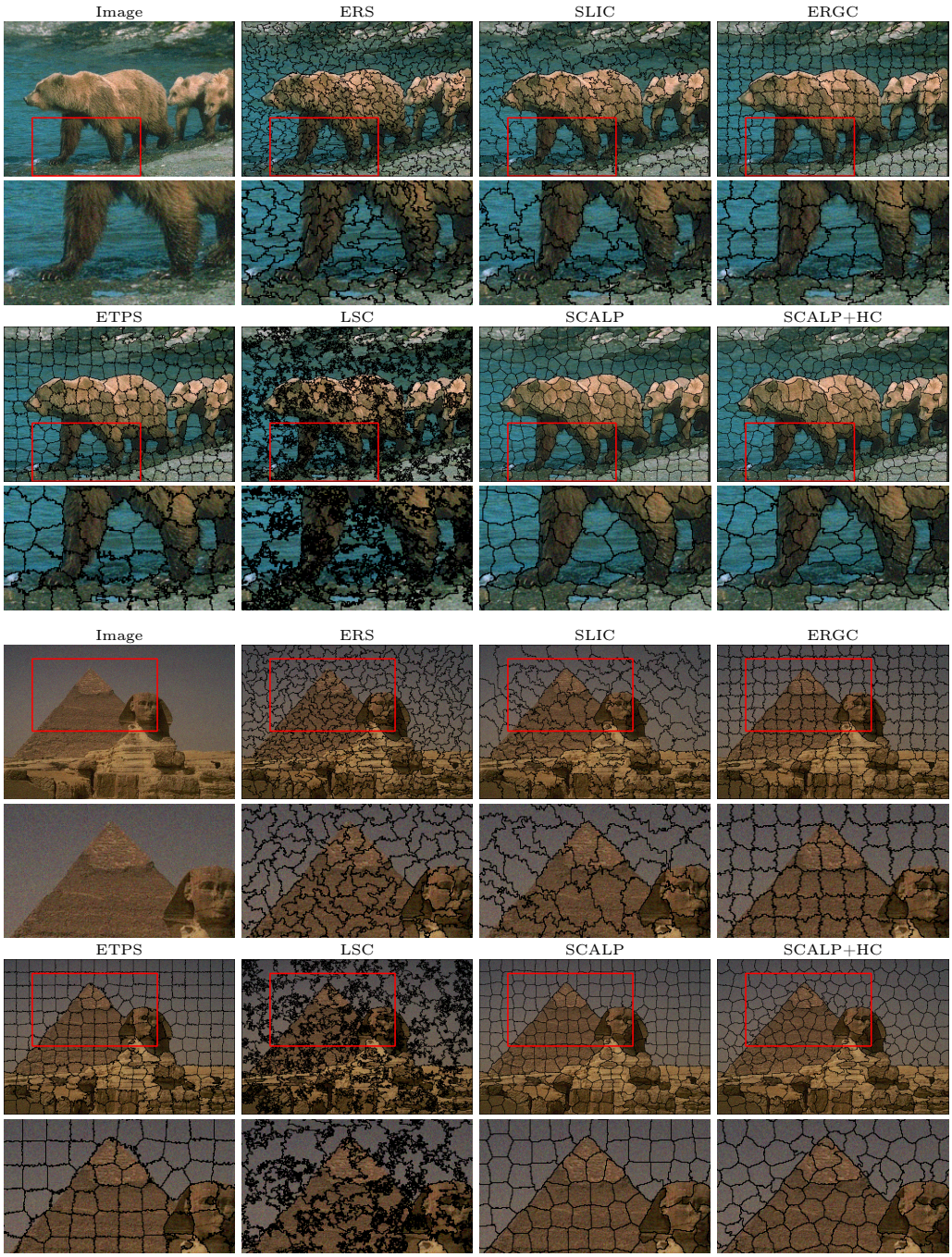


FIGURE 5.17 – Comparaison qualitative des décompositions de SCALP et des méthodes de l'état-de-l'art sur des images de la BSD bruitées. Les méthodes comparées génèrent des superpixels imprécis avec des frontières bruitées, quand SCALP est robuste au bruit et produit des superpixels réguliers avec des frontières lisses qui adhèrent correctement aux contours de l'image.

le volume (2D + temps) en superpixels temporels tubulaires et sont donc uniquement adaptées au contexte du traitement vidéo, par exemple pour effectuer le suivi d'actions (Jain et al., 2014). D'autres méthodes proposent d'effectuer un suivi des superpixels générés sur les trames successives, par exemple Wang et al. (2011); Chang et al. (2013); Reso et al. (2013), ce qui peut entraîner une segmentation similaire aux supervoxels tubulaires, et exiger un calcul de flot optique pour être efficace (Chang et al., 2013).

Contrairement à d'autres méthodes basées sur des géodésiques (Rubio et al., 2016), ou des approches de lignes de partage des eaux (Machairas et al., 2015), qui nécessiteraient des adaptations conséquentes pour être appliquées aux données 3D, nous étendons naturellement SCALP au calcul de décompositions en supervoxels sur des volumes 3D. Nous partons d'une grille 3D régulière et effectuons la décomposition en ajoutant une dimension aux équations. Pour valider cette généralisation, nous considérons les images 3D de type IRM de la base BRATS (Menze et al., 2015). Cet base contient 80 IRM de cerveau de patients souffrant de tumeurs. Les images sont segmentées en trois étiquettes : fond, tumeur et œdème, entourant la tumeur.

Nous illustrons des exemples de décomposition en supervoxels de SCALP avec la segmentation de vérité terrain dans la Figure 5.18, où la tumeur et l'œdème sont respectivement segmentés en rouge et en vert. Cette base de données est particulièrement difficile à traiter car la résolution des images est très faible.

Nous évaluons les performances sur la métrique ASA 3D par rapport aux étiquettes des structures, pour SCALP et les méthodes Achanta et al. (2012) et Buyssens et al. (2014), qui permettent de calculer des décompositions volumiques en supervoxels et dont les implémentations sont disponibles. SCALP obtient une ASA 3D de 0.9848, tandis que SLIC (Achanta et al., 2012) et ERGC (Buyssens et al., 2014) obtiennent respectivement une mesure moyenne de 0.9840 et 0.9652. Ces résultats montrent également l'intérêt de l'utilisation de SCALP pour des applications basées sur les supervoxels, puisque notre méthode obtient une précision similaire à celles des méthodes de l'état-de-l'art.

5.5 Conclusion

Dans ce travail, nous généralisons la méthode proposée dans Achanta et al. (2012), en considérant l'information de couleur et d'intensité de contour sur le chemin linéaire du pixel au barycentre du superpixel. Notre méthode est robuste au bruit et l'utilisation de descripteurs le long du chemin linéaire améliore le respect des objets de l'image et la régularité de la forme des superpixels. SCALP obtient également des résultats, surpassant les méthodes les plus récentes sur les métriques de détection de contours, de superpixels et supervoxels standards. De nombreuses applications de traitement et d'analyse d'images peuvent bénéficier d'une telle décomposition rapide, précise et régulière.

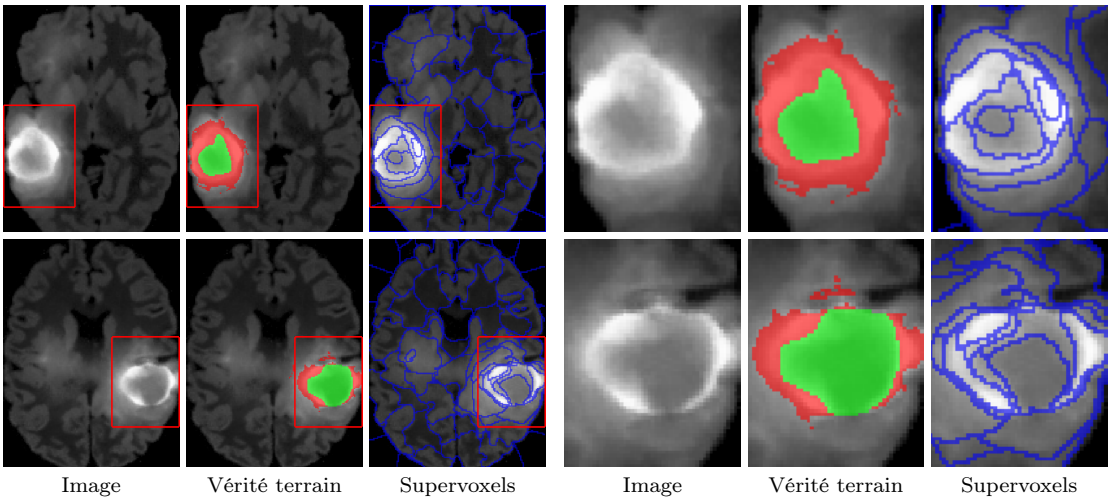


FIGURE 5.18 – Exemples de décomposition en supervoxels avec SCALP sur des images médicales (Menze et al., 2015) (à gauche), avec un zoom (à droite) sur la tumeur et l'œdème, respectivement segmentés en vert et en rouge. Les structures sont détectées par SCALP malgré le bruit et la faible résolution de l'image.

Dans ce chapitre, la régularité des superpixels générés par SCALP est attestée visuellement. Le chemin linéaire augmente naturellement la convexité des superpixels quand d'autres distances géodésiques tendent à créer des superpixels irréguliers. Dans le chapitre suivant, nous montrons qu'il n'existe pas de métriques satisfaisantes permettant d'évaluer cet aspect. Nous proposons alors une nouvelle métrique, qui évalue de façon robuste la régularité de forme et de taille des superpixels, et qui permet de renforcer la pertinence de l'évaluation des méthodes de décomposition.

Chapitre 6

Évaluation de la régularité

Sommaire

6.1 Introduction 124
6.2 Mesure de régularité globale 126
6.2.1 Régularité de forme 126
6.2.2 Consistance de forme 133
6.2.3 Mesure de régularité globale 135
6.3 Comparaison de SCALP à l'état-de-l'art sur GR 136
6.4 Impact de la régularité sur des applications basées sur les superpixels 138
6.4.1 Correspondance de plus proches voisins approchés 138
6.4.2 Suivi de superpixels 140
6.4.3 Compression d'image 142
6.4.4 Détection de contours et segmentation 144
6.4.5 Corrélation entre GR et performances 145
6.5 Conclusion 147

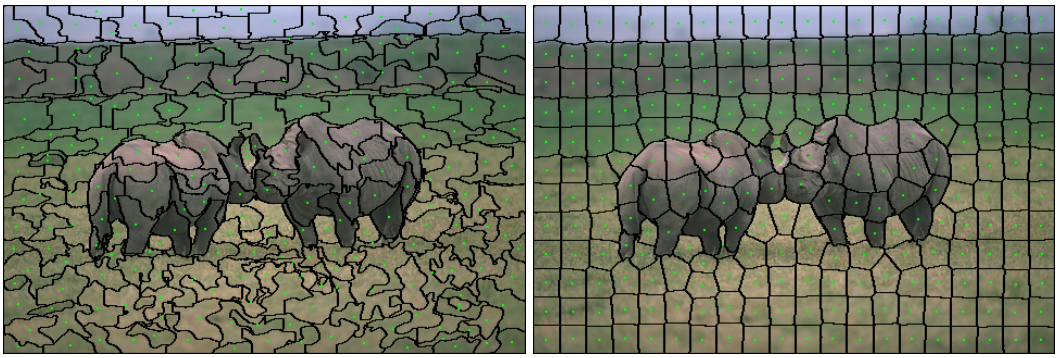
6.1 Introduction

Comme mentionné dans le chapitre 4, les propriétés d'une décomposition en superpixels sont les suivantes : homogénéité de couleur, respect des objets de l'image, et régularité des superpixels. Ce dernier aspect est souvent négligé dans le processus d'évaluation des méthodes. Cependant, des travaux tels que Strassburg et al. (2015) montrent que la régularité est en lien direct avec les performances d'applications de segmentation et d'étiquetage à partir d'exemples. Les superpixels réguliers permettent également d'atteindre de plus hautes performances pour des applications telles que le suivi vidéo (Chang et al., 2013; Reso et al., 2013). La régularité est donc une propriété fondamentale qui doit être évaluée de façon robuste pour comparer pertinemment les méthodes de superpixels.

Alors que des métriques permettent d'évaluer correctement les deux premières propriétés, il n'existe pas de mesure satisfaisante d'évaluation de la régularité d'une décomposition en superpixels, ni même de définition standard de ce que devrait être la régularité. Par exemple, la métrique de circularité introduite dans Schick et al. (2012) est très sensible à la régularité des contours et privilégie fortement les formes circulaires, quand la métrique locale de régularité introduite dans Wang et al. (2017b) considère uniquement les formes carrés comme étant les plus régulières. Ces deux métriques ont une définition différente de la régularité optimale, et sont toutes deux discutables. Ces mesures sont d'ailleurs uniquement locales et ne mesurent pas la consistance des formes en termes de taille dans la décomposition. Les travaux ayant essayé de tenir compte de cette consistance, par exemple Machairas et al. (2015), ne tiennent alors plus compte de la régularité de forme en elle-même et s'avèrent ne pas être pertinents sur certaines décompositions irrégulières.

Il est pourtant nécessaire d'évaluer efficacement la régularité puisque cet aspect peut définir le comportement même de la méthode de superpixels. En effet, puisque toutes les propriétés ne peuvent être optimisées en même temps (Neubert et Protzel, 2014), la plupart des méthodes de décomposition, dont SCALP, calculent un compromis entre ces aspects dans leur modèle, et utilisent un paramètre, qui peut être réglé par l'utilisateur, pour déterminer la contrainte de régularité sur la forme des superpixels (Achanta et al., 2012; Buyssens et al., 2014; Machairas et al., 2015; Chen et al., 2017; Yao et al., 2015). Nous montrons dans la Figure 6.1 qu'une même méthode, par exemple SCALP, peut produire des résultats de décomposition très différents, selon le réglage de ce paramètre. Quand la décomposition régulière fournit des superpixels visuellement cohérents, avec approximativement la même taille et la même forme, l'irrégulière respecte plus précisément les variations de couleur, au détriment de formes de superpixels très variables.

Ainsi, bien que beaucoup de travaux fassent une distinction claire entre les méthodes régulières et irrégulières, en considérant généralement les résultats produits avec les paramètres par défaut, ce qualificatif peut ne pas refléter le comportement potentiel d'une méthode. La distinction entre méthode régulière



(a) Décomposition irrégulière (b) Décomposition régulière

FIGURE 6.1 – Exemple de décompositions obtenues avec SCALP pour différentes valeurs du paramètre de régularité. La différence entre les décompositions peut être exprimée par la variance des distances entre les barycentres des superpixels (normalisée par la distance moyenne), qui est de 0.632 pour la décomposition irrégulière (a) et de 0.406 pour la régulière (b). SCALP peut donc produire des décompositions aux régularités très différentes.

et irrégulière devrait donc être basée sur la capacité de la méthode à produire des superpixels de tailles très différentes, et non pas de formes irrégulières. Par exemple, des méthodes telles que Felzenszwalb et Huttenlocher (2004); Van den Bergh et al. (2012); Rubio et al. (2016) génèrent des régions de tailles très variables, quand SCALP peut produire des superpixels aux formes irrégulières en relâchant la contrainte de régularité (voir Figure 6.1 (a)), mais aux tailles consistantes, puisque le groupement des pixels est contraint dans une zone de recherche.

La régularité peut donc être paramétrée dans SCALP et dans la plupart des méthodes récentes, et peut avoir un impact important sur les performances. L'impact de ce paramètre doit donc être étudié durant l'évaluation des méthodes de l'état-de-l'art. De plus, l'objectif est généralement d'utiliser les superpixels comme pré-traitement pour améliorer le temps de calcul d'une application. Le nombre de superpixels correspond donc au nombre d'éléments à traiter, et est en lien avec la complexité du processus s'appuyant sur les superpixels. La régularité devient alors le principal paramètre à régler. Enfin, lorsque l'on compare les méthodes de l'état-de-l'art, même si les résultats sont calculés avec les paramètres par défaut indiqués dans les papiers correspondants, il peut y avoir un biais important dans la comparaison, puisque la plupart des métriques d'évaluation ou des applications basées sur les superpixels obtiennent des performances très différentes selon la régularité de la décomposition. Dans l'étude récente de Stutz et al. (2016), qui compare les performances des méthodes de superpixels populaires, il n'est pas fait mention des paramètres utilisés, bien que le classement proposé entre les méthodes puisse être altéré avec différents paramètres de régularité.

Contributions

Dans la section 5.4, nous avons validé SCALP de façon standard sur les métriques existantes pertinentes, évaluant le respect des objets de l'image et l'homogénéité de couleur, pour plusieurs échelles de superpixels. Dans ce chapitre, nous proposons une nouvelle façon d'évaluer les méthodes de superpixels en considérant une mesure robuste de régularité. Ce travail vise à unifier le processus de comparaison des méthodes et à permettre une estimation claire et précise de leurs performances :

— Nous considérons d'abord les métriques existantes cherchant à mesurer la régularité d'une décomposition en superpixels et montrons leur limites.
— Nous répondons à l'absence de robustesse de ces métriques avec la mesure de régularité globale GR (pour Global Regularity) qui évalue efficacement la régularité de forme et la consistance des superpixels d'une décomposition.
— Nous évaluons alors les méthodes de superpixels de l'état-de-l'art sur la mesure GR. Contrairement à l'approche classique, pour réduire le biais dans le processus d'évaluation, nous recommandons d'effectuer l'évaluation des métriques retenues dans le chapitre 4 pour plusieurs niveaux de régularité, évalués avec GR, permettant ainsi d'illustrer le spectre des résultats potentiels de chaque méthode de superpixels.
— Enfin, nous démontrons que la contrainte de régularité a un impact considérable sur les performances de plusieurs applications, et que notre métrique GR est fortement corrélée aux résultats obtenus.

6.2 Mesure de régularité globale

Les principales mesures de régularité sont la circularité (Schick et al., 2012) qui évalue la compacité locale des superpixels, et la mesure globale de Machairas et al. (2015), qui évalue la consistance des formes au sein de la décomposition. Ces métriques présentent plusieurs défauts, auxquels nous répondons dans ce chapitre en introduisant deux nouvelles métriques (6.5) et (6.8), que l'on combine pour proposer une nouvelle mesure de régularité globale (GR). GR évalue à la fois l'aspect de régularité de forme et la consistance des superpixels dans la décomposition, et nous démontrons dans les sections 6.3 et 6.4 que cette mesure permet d'évaluer efficacement les performances des méthodes de superpixels.

6.2.1 Régularité de forme

La métrique de circularité C a été introduite dans Schick et al. (2012), et évalue localement la compacité d'une forme de superpixel S comme suit :

$$\mathrm{C}(S) = \frac{4\pi|S|}{|P(S)|^2}, \tag{6.1}$$

où $P(S)$ est le périmètre du superpixel. La régularité est ainsi considérée comme la capacité à produire une forme circulaire.

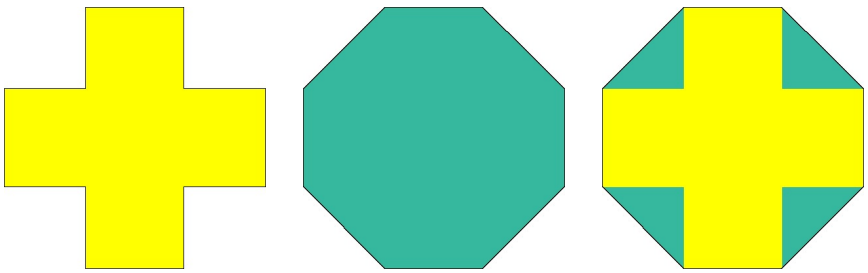
Bien que la circularité soit la seule mesure de régularité locale considérée dans les travaux de l'état-de-l'art (Reso et al., 2013; Buyssens et al., 2014; Schick et al., 2014; Duan et Lafarge, 2015; Zhang et al., 2016; Stutz et al., 2016; Brekhna et al., 2017), elle présente plusieurs défauts. La pertinence et la définition même de cette métrique ont d'ailleurs été discutées dans Machairas et al. (2015), puisque la circularité considère un cercle ou un hexagone comme étant plus régulier qu'un carré, et qu'elle s'avère très sensible aux contours irréguliers. Ainsi, plusieurs méthodes, dont Zhang et al. (2016), partent d'une décomposition hexagonale et définissent leur contrainte spatiale de sorte à favoriser les formes hexagonales, pour obtenir de plus hautes valeurs de circularité. En revanche, la plupart des méthodes génèrent une décomposition régulière carrée lorsque leur paramètre de régularité est réglé au maximum. Puisqu'il n'y a pas de raison a priori pour encourager des décompositions aux formes circulaires ou hexagonales, une mesure de régularité pertinente devrait donc également accorder une haute valeur aux formes régulières comme les carrés.

La régularité est généralement définie comme la capacité à fournir des formes convexes et équilibrées, comme les carrés, tout en pénalisant les contours irréguliers. Afin de définir une telle mesure, nous proposons de combiner ces trois aspects dans un nouveau critère de régularité de forme.

Solidité

Pour évaluer la convexité globale d'une forme S, nous proposons de considérer sa *solidité* (SO), c'est-à-dire sa superposition avec son enveloppe convexe H_S. Une telle enveloppe convexe, contenant toute la forme S, est illustrée dans la Figure 6.2 et peut être calculée, par exemple en utilisant une triangulation de Delaunay (Lee et Schachter, 1980). Des formes parfaitement convexes telles que les carrés, les hexagones ou les cercles obtiennent la plus haute valeur de SO :

$$\mathrm{SO}(S) = \frac{|S|}{|H_S|} \leq 1. \tag{6.2}$$



(a) (b) (c)

FIGURE 6.2 – Exemple d'enveloppe convexe d'une forme synthétique. La superposition entre la forme (a) et l'enveloppe convexe (b) est montrée en (c). La forme est contenue dans l'enveloppe et l'aire de la superposition est telle que SO = 0.781.

Répartition équilibrée

La superposition avec l'enveloppe convexe n'est cependant pas suffisante pour exprimer tous les aspects de la régularité. Des formes convexes telles que les ellipses ou les lignes obtiennent également de hautes valeurs de SO, bien qu'elles ne devraient être considérées comme régulières du point de vue de la répartition spatiale des pixels qui les composent. Afin de mesurer cet aspect, nous définissons un terme de variance $\mathrm{V_{xy}}$ tel que :

$$\mathrm{V_{xy}}(S) = \frac{\min(\sigma_x, \sigma_y)}{\max(\sigma_x, \sigma_y)} \leq 1, \tag{6.3}$$

avec σ_x et σ_y, les écarts-types des positions des pixels x et y dans S. $\mathrm{V_{xy}} = 1$ si, et seulement si, $\sigma_x = \sigma_y$. Dans ce cas, la répartition spatiale des pixels autour du barycentre est considérée comme équilibrée.

Régularité des contours

Enfin, la régularité des contours des superpixels doit également être considérée. La mesure de *convexité* (CO) compare le nombre de pixels de contours de la forme et celui de son enveloppe convexe :

$$\mathrm{CO}(S) = \frac{|P(H_S)|}{|P(S)|} \leq 1. \tag{6.4}$$

Bien que cette mesure soit généralement en lien avec SO, elle dépend principalement de la régularité des contours et pénalise les superpixels bruités.

Le critère de régularité de forme SRC (pour Shape Regularity Criterion) proposé, combine l'évaluation de ces trois aspects et est défini comme suit :

$$\mathrm{SRC}(\mathcal{S}) = \sum_k \frac{|S_k|}{|I|} \mathrm{SO}(S_k) \mathrm{V_{xy}}(S_k) \mathrm{CO}(S_k), \tag{6.5}$$

où $\mathcal{S} = \{S_k\}_{k \in \{1,\ldots,|\mathcal{S}|\}}$ est composé de $|\mathcal{S}|$ superpixels S_k, dont les tailles sont considérées afin de mieux refléter la régularité à l'échelle de la décomposition.

Notons que la combinaison de SO et CO est liée au ratio entre la constante de Cheeger de la forme S et de son enveloppe convexe H_S. Comme montré dans Caselles et al. (2009), la mesure de Cheeger est un outil pertinent pour évaluer à la fois la convexité et la régularité du contour. En pratique, nous utilisons la racine de $\mathrm{V_{xy}}$, afin que chaque critère ait la même gamme de variation. SRC évalue de façon robuste et conjointe la convexité, la régularité des contours et l'équilibre de la répartition des pixels.

	Formes régulières			Formes standards			Formes irrégulières		
	Carré	*Cercle*	*Hexagone*	*Ellipse*	*Croix*	*Haricot*	*W*	*Séparé*	*U*
C	**0.830**	**1.000**	**0.940**	**0.870**	**0.530**	**0.580**	**0.150**	**0.280**	**0.150**
SRC	**1.000**	**0.989**	**0.987**	**0.712**	**0.650**	**0.564**	**0.387**	**0.369**	**0.233**
SO	1.000	0.989	0.989	0.988	0.781	0.800	0.841	0.530	0.357
V_{xy}	1.000	1.000	0.997	0.718	1.000	0.811	0.990	0.888	0.942
CO	1.000	1.000	1.000	0.997	0.833	0.868	0.465	0.783	0.694
C	**0.480**	**0.430**	**0.420**	**0.410**	**0.340**	**0.440**	**0.100**	**0.210**	**0.070**
SRC	**0.716**	**0.633**	**0.625**	**0.474**	**0.515**	**0.500**	**0.296**	**0.321**	**0.136**
SO	0.925	0.923	0.917	0.927	0.757	0.796	0.763	0.547	0.227
V_{xy}	0.999	1.000	0.997	0.715	1.000	0.818	0.988	0.885	0.950
CO	0.774	0.685	0.683	0.997	0.680	0.767	0.392	0.664	0.630

FIGURE 6.3 – Comparaison des métriques de régularité sur des formes synthétiques aux contours lisses (en haut) et bruités (en bas). La circularité C favorise uniquement les formes circulaires et ne permet pas de différencier les formes régulières des formes standards dans le cas bruité. La métrique proposée SRC sépare quant à elle clairement les trois groupes dans les deux cas, et donne la plus forte mesure (≈ 1) aux formes régulières lisses.

C versus SRC

Afin de démontrer la robustesse et la pertinence de SRC (6.5) par rapport à la circularité C (6.1), nous considérons dans la Figure 6.3 des formes synthétiques séparées en trois groupes (régulières, standards et irrégulières), et générées avec des contours lisses (en haut) et bruités (en bas). Tout d'abord, nous présentons les défauts de la circularité, qui est plus faible pour le *Carré* que pour l'*Hexagone* ou même l'*Ellipse*. Puisque les méthodes telles que Machairas et al. (2015); Zhang et al. (2016) produisent des superpixels depuis une grille hexagonale, l'évaluation de la régularité est susceptible d'être très supérieure pour ces méthodes par rapport à d'autres initialisées depuis une grille régulière carrée.

Cet aspect est illustré dans la Figure 6.4 où les résultats de régularité évalués avec C et SRC sont donnés pour deux décompositions basées sur une grille régulière carrée et hexagonale. La circularité est également très sensible à la régularité des contours. Ainsi, les formes régulières et standards dans le cas bruité ne peuvent plus être différenciées (voir la partie basse de la Figure 6.3). Enfin, les formes standards non bruitées ont une circularité bien plus forte que les formes régulières bruitées.

Comme on peut le voir dans la Figure 6.3, SO, V_{xy} et CO considérées indépendemment ne sont pas suffisants pour exprimer la régularité d'une forme. Par exemple, SO est pertinent pour toutes les formes, excepté pour l'*Ellipse* et *W*, puisque ces deux formes ont une forte superposition avec leur enveloppe convexe. V_{xy} pénalise

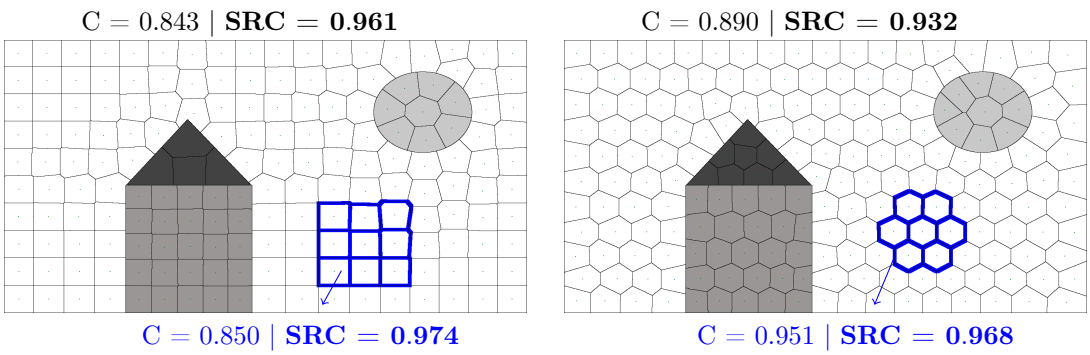


FIGURE 6.4 – Illustration de la forte régularité associée aux formes hexagonales avec la circularité C. Deux décompositions d'une image synthétique avec SCALP en considérant des grilles d'initialisation carrée (à gauche) et hexagonale (à droite) sont représentées. La circularité est bien plus élevée pour les hexagones et au total pour l'image de droite, bien que les formes des hexagones soient étirées en bordure de l'image. Le critère proposé SRC évalue équitablement carrés et hexagones comme étant des formes régulières.

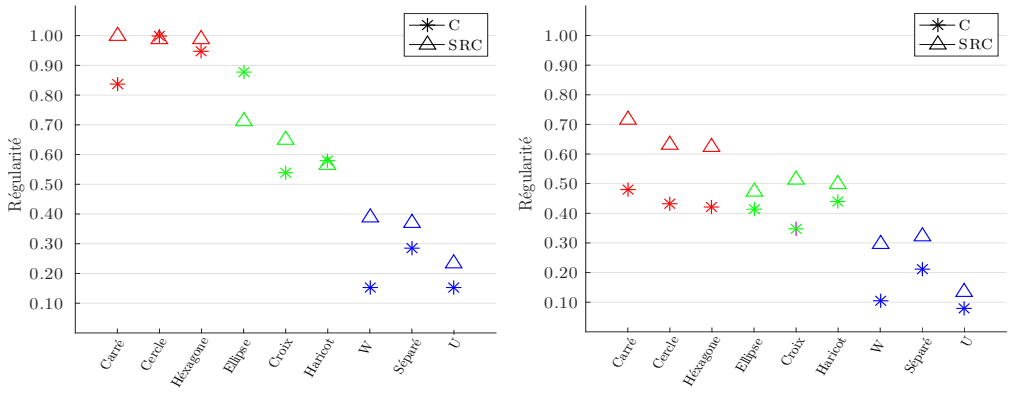


FIGURE 6.5 – Comparaison de la circularité C et du critère de régularité de forme SRC proposé sur la séparation des trois groupes dans le cas lisse (à gauche) et bruité (à droite).

l'*Ellipse* qui n'a pas une répartition équilibrée des pixels, et CO considère la grande quantité de pixels de contour de la forme *W*. La métrique proposée SRC combine ces trois mesures pour considérer tous les aspects de la régularité.

Avec SRC, nous notons d'abord que les trois formes régulières ont la plus haute régularité (≈ 1), et que les formes régulières dans le cas bruité ont des mesures de SRC similaires à celles des formes standards dans le cas non bruité. En général, puisque SRC est moins dépendant de la régularité des contours, dans chaque cas, les trois groupes sont clairement séparés, contrairement à C. Cette comparaison est également représentée dans la Figure 6.5, où SRC permet clairement de différencier les trois groupes.

Robustesse à l'échelle

La circularité apparaît également dans la Figure 6.6 très dépendante de la taille de la forme. Comme observé dans Roussillon et al. (2010), du fait d'un calcul discret, la circularité peut être supérieure à 1 (nous seuillons sa valeur dans la Figure 6.6), et celle-ci diminue quand la taille des formes augmente. Ainsi, les comparaisons des méthodes sur cette métrique ne sont pertinentes que si les décompositions ont le même nombre de superpixels, c'est-à-dire si les superpixels ont approximativement la même taille. Contrairement à la circularité, SRC fournit une mesure de régularité beaucoup plus consistante, et robuste à l'échelle des superpixels, par exemple le *Carré* a toujours une mesure de SRC égale à 1.

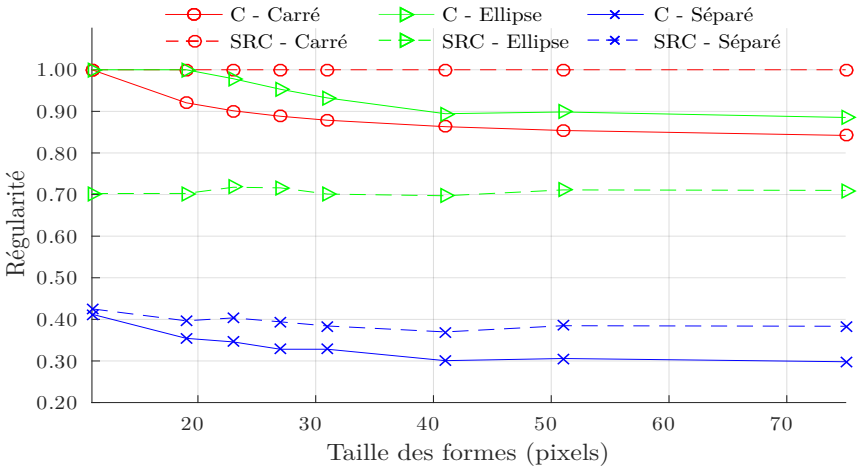


FIGURE 6.6 – Robustesse à l'échelle du critère SRC par rapport à C, sur des formes de chaque groupe. La valeur de C est instable et diminue avec la taille des formes, quand SRC reste constant. Les abscisses correspondent à la hauteur ou largeur maximale d'une forme.

Robustesse au bruit

Enfin, nous démontrons quantitativement la robustesse de SRC au bruit, c'est-à-dire aux frontières de superpixels très irrégulières. Nous proposons de perturber aléatoirement le compromis entre la distance de couleur et spatiale (5.3) dans la méthode de Achanta et al. (2012) afin de générer des décompositions bruitées, dont des exemples sont donnés dans la Figure 6.7. Nous calculons des décompositions sur toutes les images de la BSD pour différents niveaux de régularité (paramètre m) et évaluons la régularité moyenne avec C et SRC dans la Figure 6.8. La circularité s'avère fortement impactée par les frontières bruitées, jusqu'à ne plus être corrélée à l'évolution du paramètre de régularité m, puisqu'elle chute à partir de $m = 75$. À l'inverse, la métrique SRC est strictement positivement corrélée à l'évolution de m, démontrant qu'elle est moins sensible à la régularité des contours, et plus à celle de la forme, ce qui permet d'évaluer de manière plus robuste la régularité des superpixels.

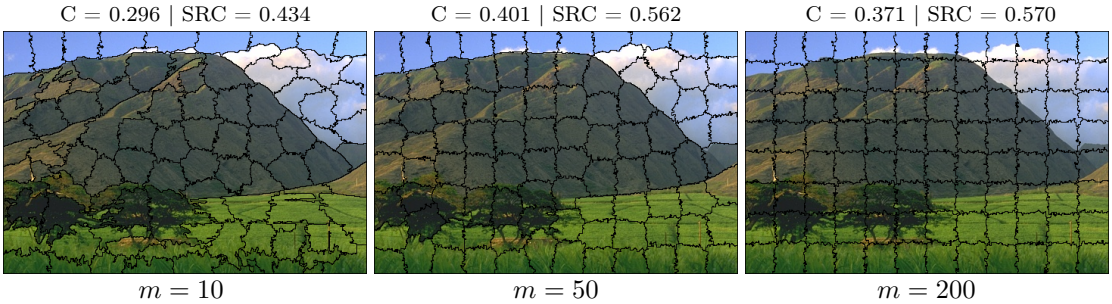


FIGURE 6.7 – Exemples de superpixels bruités calculés depuis Achanta et al. (2012) avec K=100 superpixels, pour différents paramètres de régularité m.

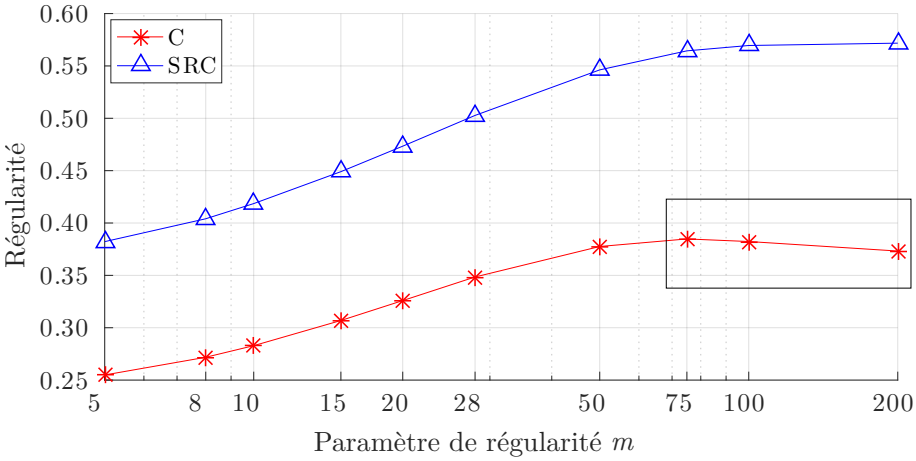


FIGURE 6.8 – Évolution de la régularité des superpixels bruités, calculés avec Achanta et al. (2012) pour différentes valeurs du paramètre de régularité m, sur les images de la BSD. La mesure de circularité chute à partir de $m = 75$.

Néanmoins, les métriques SRC ou C évaluent indépendamment chaque forme de superpixel sans considérer l'homogénéité globale de la décomposition. Des méthodes telles que Felzenszwalb et Huttenlocher (2004); Rubio et al. (2016), ou d'autres approches de segmentation telles que l'algorithme de partitionnement en arbre quaternaire (Tanimoto et Pavlidis, 1975) peuvent produire des régions de tailles très variables. Dans la Figure 6.9, nous montrons un exemple de décomposition avec des superpixels calculés avec SCALP ayant à peu près la même taille, et une décomposition de type arbre quaternaire en blocs réguliers (Tanimoto et Pavlidis, 1975), qui produit de plus larges régions dans les zones de variance couleur réduite. Puisque la circularité et SRC considèrent indépendamment chaque superpixel et fournissent une mesure moyenne d'évaluation locale de la régularité, la méthode de décomposition en blocs réguliers obtient la plus haute valeur bien que les régions segmentées soient de tailles très différentes. De telles métriques locales ne sont donc pas suffisantes pour exprimer la régularité globale. La consistance des formes dans la décomposition doit aussi être prise en compte.

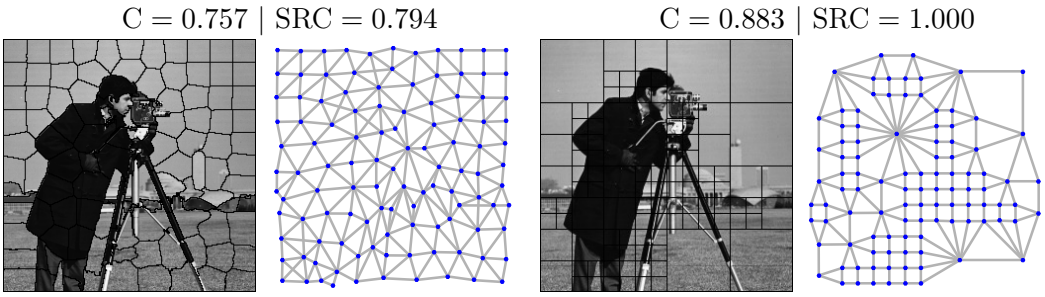


FIGURE 6.9 – Limitation des métriques locales de régularité. La consistance des formes n'est pas considérée. Ainsi, la décomposition par arbre quaternaire (Tanimoto et Pavlidis, 1975) donne de hautes mesures de régularité. Les décompositions sont représentées avec leur graphe de Delaunay, qui connecte les barycentres des superpixels adjacents.

6.2.2 Consistance de forme

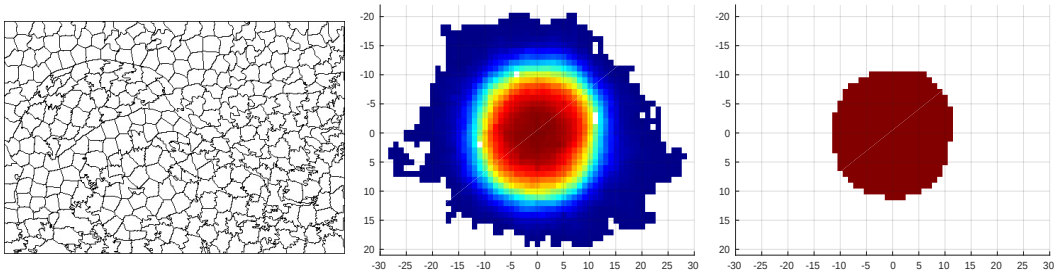
Dans Machairas et al. (2015), une mesure est introduite pour évaluer la consistance des formes des superpixels dans la décomposition. Une version adaptée du facteur de disparité (Strachan et al., 1990) est proposée pour mesurer cette cohérence. Le facteur de disparité est calculée par $1 - J(S_1, S_2)$ avec J l'index Jaccard standard (Jaccard, 1901), qui calcule la superposition entre deux régions S_1 et S_2 tel que : $J(S_1, S_2) = |S_1 \cap S_2|/|S_1 \cup S_2|$. La mesure globale J pour une décomposition en superpixels $\mathcal{S}$ est calculée comme suit (Machairas et al., 2015) :

$$\mathrm{J}(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{S_k} J(S_k^*, \hat{S}^*), \tag{6.6}$$

avec S_k^* la forme du superpixel S_k, recalé de sorte que son barycentre se trouve à l'origine du repère de coordonnées, et $\hat{S}^*$, la forme moyenne binaire de $\mathcal{S}$. Dans la suite, nous considérons la métrique J lorsque l'on se réfère au travail de Machairas et al. (2015), puisque, contrairement au facteur de disparité, l'interprétation de J est consistante avec celles des autres métriques utilisées (de fortes valeurs indiquent de meilleurs résultats). L'Éq. (6.6) compare chaque superpixel à la forme moyenne binaire $\hat{S}^*$ de la décomposition en utilisant l'opérateur J. Pour calculer $\hat{S}^*$, les superpixels recalés S_k^* sont d'abord moyennés dans S^* tels que :

$$S^* = \frac{1}{|\mathcal{S}|} \sum_{S_k} S_k^*. \tag{6.7}$$

Une forme seuillée S_t^* est alors définie par binarisation par rapport à un seuil t. La forme binaire moyenne est finalement définie comme $\hat{S}^* = S^*_{\underset{t}{\mathrm{argmax}}(|S_t^*| \geq \mu)}$, avec $\mu = |I|/|\mathcal{S}|$, la taille moyenne des superpixels. Nous illustrons ces définitions dans la Figure 6.10, en représentant une décomposition en superpixels $\mathcal{S}$, la moyenne des formes des superpixels S^*, et la forme moyenne binaire correspondante $\hat{S}^*$.



(a) Décomposition $\mathcal{S}$ (b) Forme moyenne S^* (c) Forme moyenne binaire $\hat{S}^*$

FIGURE 6.10 – Illustration de la définition de la forme moyenne d'une décomposition en superpixels (Machairas et al., 2015). Une décomposition est considérée dans (a). Les formes des superpixels sont recalées et moyennées dans S^* (b), puis seuillées pour fournir une forme moyenne binaire $\hat{S}^*$ (c).

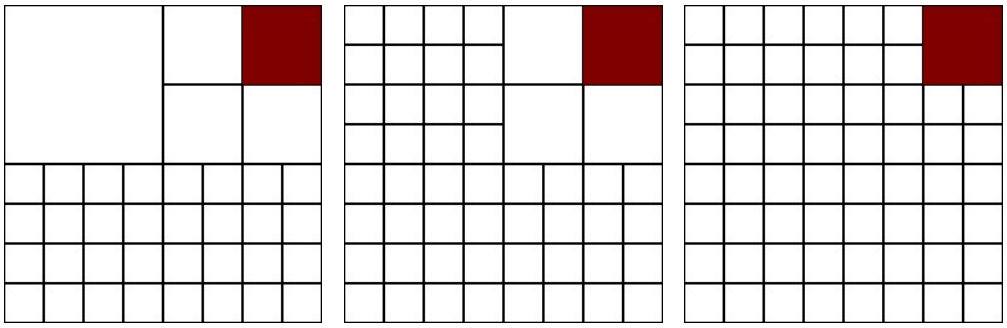
Le facteur de disparité, ou l'index Jaccard, (6.6) présente toutefois plusieurs défauts. D'abord, cette mesure ne considère pas la taille du superpixel, et chaque élément contribue de façon égale à la mesure de régularité. De plus, en considérant une forme moyenne seuillée, la métrique n'est pas robuste aux grands superpixels et peut mener à des mesures de régularité incohérentes.

Pour compenser ce problème, nous proposons le critère SMF (pour Smooth Matching Factor) qui compare directement les superpixels à la forme moyenne S^* :

$$\mathrm{SMF}(\mathcal{S}) = 1 - \sum_{S_k} \frac{|S_k|}{|I|} . \left\| \frac{S^*}{|S^*|} - \frac{S_k^*}{|S_k^*|} \right\|_1 /2. \tag{6.8}$$

SMF compare la répartition spatiale de la forme moyenne des superpixels S^* à chaque superpixel recalé S_k^*. Cette métrique est alors proche de 1 si les distributions des pixels parmi les formes sont similaires, et proche de 0 sinon. Notons que comme pour SRC (6.5), les tailles $|S_k|$ des superpixels sont considérées pour refléter la régularité à l'échelle de la décomposition. En considérant directement la forme moyenne, sans seuillage, l'évaluation est plus robuste aux grands superpixels. Nous illustrons ces aspects dans la Figure 6.11, où SMF est cohérent avec la consistance visuelle des décompositions, contrairement à la mesure J. Dans ces exemples, la forme moyenne binaire $\hat{S}^*$ est la même, et est représentée en rouge. La mesure J est faible et diminue alors que la consistance augmente visuellement. Enfin, des exemples de décomposition en superpixels sur des images naturelles sont donnés dans la Figure 6.12 et illustrent que SMF n'est pas corrélé à la mesure J.

Néanmoins, avec de telles métriques qui ne considèrent que la consistance de formes, décomposer l'image en des lignes ou en des rectangles étirés donneraient également la plus haute mesure. Il est donc nécessaire de considérer également la régularité de forme SRC (6.5) dans l'évaluation globale de la régularité.



J = 0.331 | SMF = 0.520 J = 0.308 | SMF = 0.719 J = 0.262 | SMF = 0.912

FIGURE 6.11 – Non robustesse de la métrique J sur plusieurs exemples. Des décompositions synthétiques sont illustrées avec les valeurs de J et de SMF correspondantes. La forme moyenne binaire $\hat{S}^*$ est représentée en rouge sur chaque exemple. J décroît alors que la régularité augmente visuellement.

J = 0.364 | SMF = 0.517 J = 0.291 | SMF = 0.536 J = 0.544 | SMF = 0.592

FIGURE 6.12 – Exemples de décompositions d'images naturelles avec les valeurs de J et de SMF correspondantes. Les deux mesures ne sont pas corrélées puisque J diminue entre la première et la deuxième image quand SMF augmente.

6.2.3 Mesure de régularité globale

Comme nous l'avons mentionné précédemment, une décomposition en superpixels parfaitement régulière doit être composée de superpixels aux formes convexes, équilibrées et aux contours réguliers, et qui soient consistants en termes de taille et de forme au sein de la décomposition. La métrique SRC permet de mesurer de façon robuste la régularité de forme locale quand la métrique SMF évalue la consistance des formes. Pour évaluer ces deux aspects en une seule mesure, nous proposons de combiner ces évaluations dans la mesure de régularité globale GR définie comme suit :

$$\mathrm{GR}(\mathcal{S}) = \mathrm{SRC}(\mathcal{S})\mathrm{SMF}(\mathcal{S}). \tag{6.9}$$

Dans la section 6.3 suivante, nous proposons une nouvelle façon d'évaluer les performances en comparant SCALP et les méthodes de l'état-de-l'art sur selon le critère GR, et en représentant les résultats des critères d'homogénéité de couleur (EV) et de respect des objets de l'image (ASA) par rapport à cette métrique de régularité.

6.3 Comparaison de SCALP à l'état-de-l'art sur GR

Dans cette section, nous comparons SCALP aux méthodes de l'état-de-l'art considérées dans le chapitre 5 sur les mêmes images de la BSD (Martin et al., 2001). Nous représentons tout d'abord dans la Figure 6.13, les performances de régularité des méthodes selon GR en fonction du nombre de superpixels. SCALP est parmi les méthodes de superpixels les plus régulières, tout en obtenant les meilleures performances de respect des objets et de bonnes performances d'homogénéité de couleur (voir le chapitre 5).

Comme on peut le voir dans la Figure 5.15, l'évaluation des performances selon le nombre de superpixels apparaît limitée puisque la hiérarchie des méthodes tend à être constante, c'est-à-dire que les courbes des performances ne se croisent plus au delà d'un certain nombre de superpixels. Comme mentionné en introduction, le réglage du paramètre de régularité peut avoir un impact important sur les performances. Ainsi, l'évaluation devrait également être effectuée sur ce critère pour mesurer la capacité de chaque méthode à produire le meilleur compromis entre les différentes propriétés. Cela permet également de s'affranchir du respect du paramètre par défaut de régularité, qui peut être réglé différemment selon les études. Dans la Figure 6.14, nous fixons le nombre de superpixels à $K = 250$, et nous calculons les performances sur les critères EV et ASA pour plusieurs niveaux de régularité. Le nombre de superpixels doit être réglé avec attention, puisque certaines méthodes peuvent produire un nombre d'éléments légèrement variable par rapport à celui demandé. Les marqueurs plus larges correspondent aux résultats obtenus avec les paramètres par défaut. Ainsi, pour TP (Levinshtein et al., 2009), ERS (Liu et al., 2011) et SEEDS (Van den Bergh et al., 2012), une seule valeur est donnée, puisque ces méthodes ne permettent pas de régler la régularité de la décomposition.

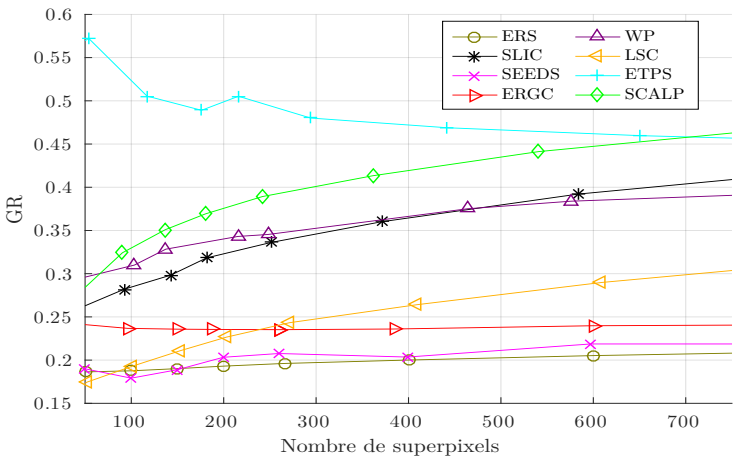


FIGURE 6.13 – Régularité de SCALP et des méthodes comparées évaluée avec GR (6.9).

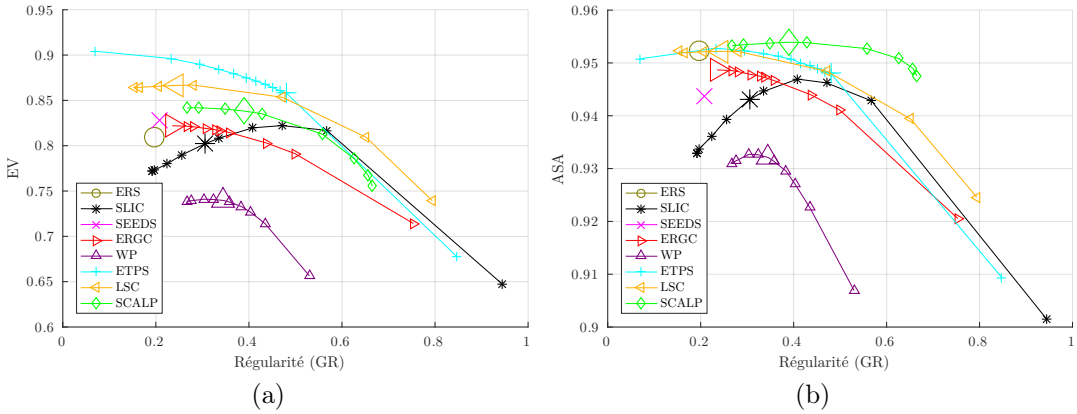
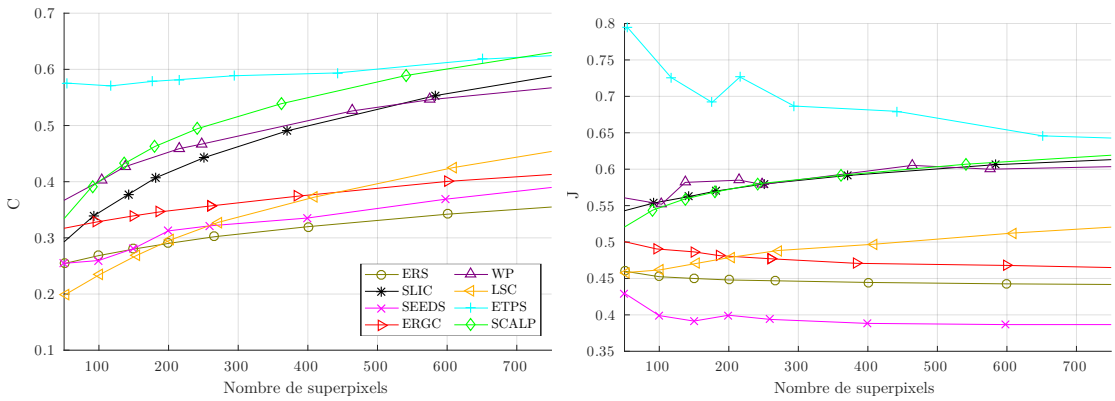


FIGURE 6.14 – Évaluation de SCALP et des méthodes de l'état-de-l'art sur EV (a) et ASA (b) pour différents niveaux de régularité évalués avec GR.

Tout d'abord, on peut observer que les résultats sont affectés par le réglage de la régularité, de sorte que pour chaque méthode, des résultats très différents peuvent être obtenus. SCALP apparaît néanmoins comme la méthode la moins affectée par la régularité et fournit de bons résultats pour chaque niveau de régularité. Ensuite, contrairement à l'évaluation standard selon le nombre de superpixels, les courbes des performances se croisent, et la hiérarchie entre les méthodes peut varier selon le niveau de régularité. Enfin, le paramètre par défaut des méthodes correspond généralement à un compromis entre les différentes propriétés mais n'est pas nécessairement optimal. Par exemple, le paramètre par défaut de la méthode ETPS (Yao et al., 2015) génère des superpixels très réguliers, et n'est pas optimal pour EV et ASA. Les performances de SLIC (Achanta et al., 2012) peuvent également être améliorés pour un plus fort paramètre de régularité. Notons que contrairement aux autres méthodes, les résultats d'EV et ASA diminuent avec une plus faible régularité pour SLIC. Cela peut être expliqué par l'étape de post-traitement de SLIC qui assure la connectivité des superpixels. Avec une contrainte de régularité relâchée, plus de pixels appartenant aux mêmes superpixels sont déconnectés, et regroupés avec moins de précision durant l'étape de post-traitement.

Cette nouvelle représentation des performances démontre que le réglage de la régularité est un paramètre crucial dans le processus d'évaluation. Un biais peut apparaître si ce paramètre n'est pas optimal, ou du moins réglé selon les recommandations des auteurs. Cet aspect est particulièrement important compte tenu du nombre important de travaux sur les méthodes de superpixels, dont les améliorations de performances sont parfois faibles et difficiles à mesurer. En comparant les performances, sur les métriques recommandées dans ce travail et en utilisant GR pour évaluer les différents niveaux de régularité, nous assurons une évaluation robuste des méthodes de superpixels. L'impact du paramètre de régularité est décrit en détails dans la section 6.4, où nous mesurons les performances de plusieurs applications pour différents niveaux de régularité évalués avec GR.



(a) (b)

FIGURE 6.15 – Régularité évaluée avec C (6.1) (a) et MF (6.6) (b) de SCALP et des méthodes de l'état-de-l'art.

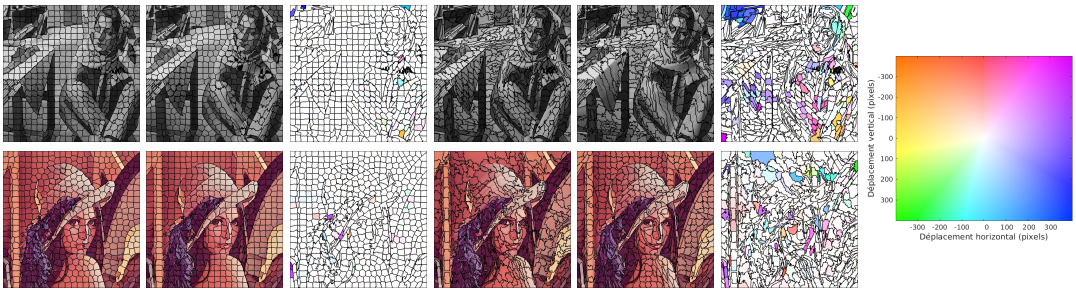
Enfin, les résultats obtenus avec les métriques C (6.1) et J (6.6) sont également représentés dans la Figure 6.15. Ces deux mesures montrent une évaluation différente de la régularité de celle obtenue avec GR. SCALP figure également parmi les meilleures méthodes sur ces critères.

6.4 Impact de la régularité sur des applications basées sur les superpixels

Dans cette section, nous considérons plusieurs applications, et démontrons comment la régularité des superpixels affecte les performances. Pour la plupart des applications basées sur les superpixels et des méthodes de décomposition elles-mêmes, les performances apparaissent très corrélées à la métrique GR proposée. Pour la recherche de correspondances ou le suivi de superpixels, des décompositions régulières permettent de mieux capturer des éléments similaires, quand pour la compression d'image ou des applications de détection de contours, une contrainte de régularité plus faible peut permettre d'atteindre de meilleures performances.

6.4.1 Correspondance de plus proches voisins approchés

Dans cette section, nous étudions l'impact de la régularité sur la précision de la recherche de plus proches voisins entre images. La corrélation entre la forme du superpixel et les performances d'étiquetage à partir d'exemples, a été démontrée dans Strassburg et al. (2015), où les méthodes basées sur une grille régulière, comme SCALP, donnent de meilleurs résultats à l'échelle pixellique. Nous proposons de calculer des correspondances de superpixels entre des décompositions régulières et irrégulières en utilisant la structure de voisinage introduite dans le chapitre 7.



(a) (b) (c) (d) (e) (f) (g)

FIGURE 6.16 – Exemples de correspondances entre les décompositions régulières (a), (b) et irrégulières (d), (e). Les déplacements entre les barycentres des superpixels associés sont respectivement illustrés pour les décompositions régulières et irrégulières en (c) et (f). Les déplacements sont illustrés en utilisant la représentation du flot optique standard (g), et sont affichés respectivement aux superpixels des images (a) et (d). Sur ces exemples, la mesure moyenne de GR est respectivement de 0.478 et de 0.205 pour les décompositions régulières (a) et (b) et irrégulières (d) et (e).

Dans la Figure 6.16, des décompositions sont calculées sur la même image, deux régulières (Figures 6.16(a) et (b)) et deux irrégulières (Figures 6.16(d) et (e)). Les plus proches voisins sont ensuite calculés entre les deux décompositions régulières et les deux irrégulières, et nous comparons l'amplitude du déplacement entre les positions respectives des correspondances dans les Figures 6.16 (c) et (f). Les correspondances sont calculées sur les descripteurs de couleur (histogrammes de couleur normalisés dans l'espace RGB), et l'amplitude du déplacement est illustré avec la représentation standard du flot optique (Figure 6.16 (g)).

Alors que les correspondances entre les décompositions régulières sont précises (faible déplacement entre les superpixels associés), les décompositions irrégulières faussent la recherche du plus proche voisin. Avec des méthodes irrégulières, les barycentres peuvent tomber à l'extérieur des superpixels, rendant difficile l'utilisation de l'information géométrique du voisinage dans le processus de recherche de plus proche voisin. Les méthodes régulières produisent des décompositions où tous les éléments ont à peu près la même taille et où les barycentres sont contenus à l'intérieur des superpixels. Par conséquent, la relation spatiale entre le barycentre d'un superpixel et ceux de ses voisins est plus pertinente, ce qui permet une recherche de correspondance plus précise.

La structure de voisinage de superpixel considérant les positions relatives des barycentres des superpixels est présentée dans le chapitre 7, et l'impact de la régularité sur les performances de recherche d'ANN est démontrée quantitativement sur une expérience d'étiquetage de visage dans le chapitre 8.

6.4.2 Suivi de superpixels

Le gain de temps de calcul obtenu avec les superpixels est particulièrement intéressant pour des applications visant le temps réel, comme le suivi d'objets sur des vidéos (*tracking*) (Wang et al., 2011; Chang et al., 2013; Reso et al., 2013). Deux principales approches existent pour la segmentation et le suivi vidéo de superpixels. La première calcule des superpixels 3D, c'est-à-dire, des supervoxels, ayant les mêmes propriétés que celles des méthodes 2D. Par conséquent, cette approche nécessite de disposer de toute la séquence pour produire une décomposition et ne peut pas être calculée au fur et à mesure de l'acquisition de nouvelles images.

La deuxième approche calcule les superpixels sur chaque image et la décomposition des images suivantes est contrainte par rapport aux segmentations précédentes (Chang et al., 2013). Dans ces méthodes, le calcul d'un flot optique est généralement nécessaire pour transférer le barycentre des superpixels d'une image à l'autre (Chang et al., 2013). En conséquence, un superpixel ne peut avoir qu'une unique correspondance dans l'image suivante, permettant alors le suivi des objets sur la séquence vidéo. L'inconvénient d'une forte irrégularité de taille entre les décompositions est illustré dans la Figure 6.17. Si une méthode décompose le même objet en un nombre variable d'éléments, il est difficile d'associer un à un des superpixels. Les décompositions irrégulières telles que SEEDS, générant des superpixels de tailles très variables mènent à des pertes dans le suivi (Figure 6.17(a)), quand les décompositions régulières permettent d'obtenir cette cohérence en favorisant les associations un à un de superpixels (Figure 6.17(b)),

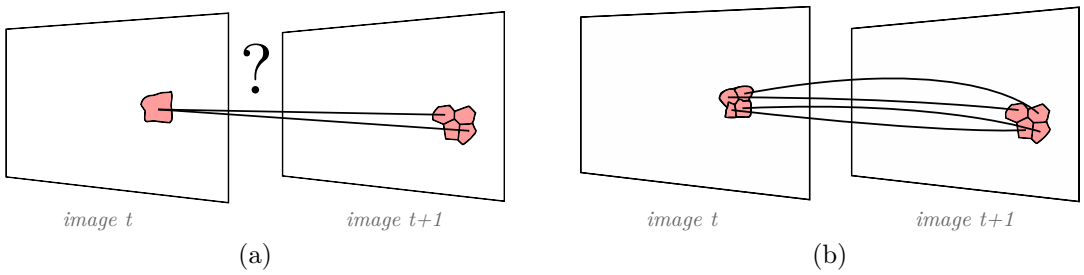
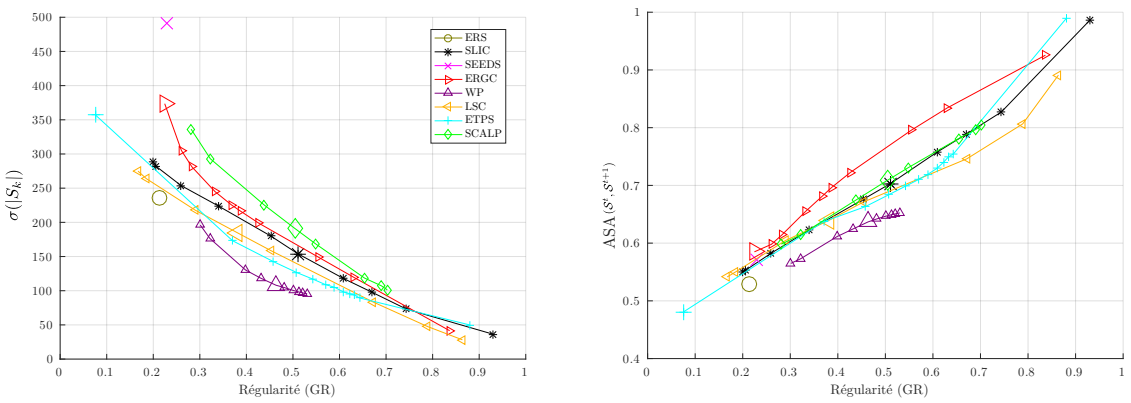


FIGURE 6.17 – Exemple de suivi de superpixels entre deux images successives décomposées avec une méthode irrégulière (a) et régulière (b). La consistance du suivi est favorisée avec une décomposition régulière, générant des superpixels aux tailles similaires.

Dans les Figures 6.18(a) et 6.18(b), nous représentons respectivement la variance de taille des superpixels et l'ASA entre les décompositions des images successives $\mathcal{S}^t$ et $\mathcal{S}^{t+1}$, sur une courte séquence vidéo issue de Tsai et al. (2012) pour SCALP et les méthodes de l'état-de-l'art. Ces valeurs sont représentées pour différents niveaux de régularité, évalués avec GR. La variance de taille est très élevée pour des méthodes irrégulières comme SEEDS, et celle-ci est inversement proportionnelle à la régularité. Augmenter le paramètre de régularité permet donc d'avoir une meilleure consistance de taille entre les décompositions, ce qui peut faciliter le suivi (voir



(a) (b)

FIGURE 6.18 – Cohérence des décompositions en superpixels dans une séquence vidéo. La variance de taille des superpixels sur toute la séquence est illustrée dans (a) et diminue avec la régularité. L'ASA entre les décompositions des images successives $\mathcal{S}^t$ et $\mathcal{S}^{t+1}$ est illustrée dans (b) et augmente avec la régularité. La séquence considérée (*girl*) est issue de Tsai et al. (2012), et des images de la séquence sont fournies dans la Figure 6.19.

Figure 6.17). L'ASA moyen en Figure 6.18(b) illustre le degré de superposition des superpixels d'une image à l'autre sur la séquence. Augmenter la régularité favorise la superposition des éléments et facilite donc le suivi. Bien sûr, celle-ci ne doit pas être trop élevée pour ne pas empêcher la capture des contours des objets. Néanmoins, les objets de l'image ne subissent généralement que de légères déformations entre deux images successives d'un flux vidéo. Sur la séquence considérée, l'ASA entre les vérités terrain de segmentation est en effet élevée (0.9580), ce qui montre le faible déplacement des objets entre deux images.

Nous étudions également l'impact de la régularité sur les performances de la méthode de suivi de superpixels TSP (Chang et al., 2013). Nous considérons les images de Tsai et al. (2012), qui contient de courtes séquences vidéo fournies avec un étiquetage binaire de la vérité terrain, et nous comparons les performances de suivi obtenues avec TSP (Chang et al., 2013) en générant des décompositions régulières et irrégulières. Dans le Tableau 6.1, nous indiquons la précision d'étiquetage (pourcentage de pixels étiquetés correctement) des pixels sur toutes les séquences, ce qui exprime la capacité de la méthode à suivre les superpixels dans la séquence. Avec des décompositions plus régulières, les formes et les tailles des superpixels sont cohérentes au fil du temps, ce qui améliore la précision du suivi. Nous montrons également le pourcentage moyen de nouveaux superpixels créés entre deux décompositions, ce qui correspond aux superpixels perdus dont le suivi a échoué. Des valeurs plus élevées pour les décompositions irrégulières indiquent l'échec du suivi, et la perte de correspondance entre les superpixels pendant le processus. Des exemples de suivi pour de telles décompositions régulières et irrégulières sont illustrés dans la Figure 6.19 pour toutes les séquences considérées. La cohérence des décompositions régulières permet de suivre plus efficacement les superpixels à travers les séquences.

Tableau 6.1 – Précision de suivi avec TSP (Chang et al., 2013) sur les séquences de Tsai et al. (2012). La précision d'étiquetage et le pourcentage de superpixels perdus sur les séquences sont donnés pour les décompositions régulières et irrégulières.

	Précision		Perte	
Séquence	Régulier	Irrégulier	Régulier	Irrégulier
birdfall2	98.3%	97.8%	1.0%	1.4%
girl	51.1%	50.4%	13.9%	24.8%
parachute	75.3%	73.9%	4.5%	5.1%
penguin	94.3%	85.0%	2.6%	8.8%
Moyenne	79.8%	76.7%	5.5%	10.0%

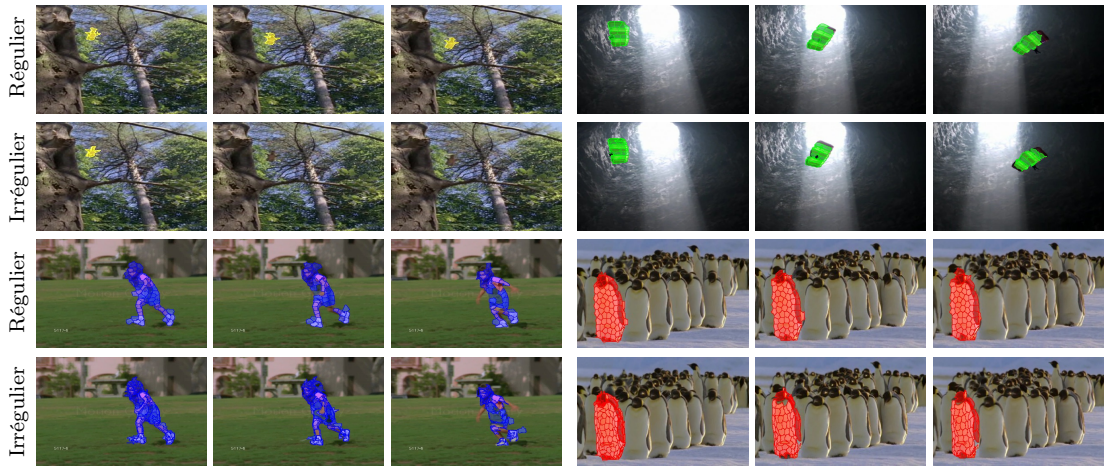


FIGURE 6.19 – Exemples de suivi de superpixels avec Chang et al. (2013) sur des séquences de Tsai et al. (2012). Sur ces séquences, les mesures moyennes de GR sont respectivement de 0.308 et 0.554 pour les décompositions irrégulières et régulières.

6.4.3 Compression d'image

Puisque les superpixels définissent une représentation de bas niveau de l'image, ils peuvent être utilisés pour compresser son contenu et notamment son information de couleur (Levinshtein et al., 2009; Wang et al., 2013). Nous proposons alors d'étudier l'impact de la régularité sur l'efficacité de la compression couleur de l'image. De la même manière que pour Wang et al. (2013), chaque canal de couleur des pixels contenus dans un superpixel est approché par un polynôme du troisième ordre. L'information de couleur est alors contenue dans les coefficients polynomiaux qui sont calculés par une approche des moindres carrés.

Un exemple de compression est illustré dans la Figure 6.20 pour une décomposition régulière et irrégulière. Nous montrons l'image reconstruite I_r à partir des coefficients polynomiaux stockés pour chaque superpixel. La décomposition irrégulière génère une reconstruction plus proche de l'image initiale I, tandis que les bords des superpixels sont nettement plus visibles avec la décomposition régulière.

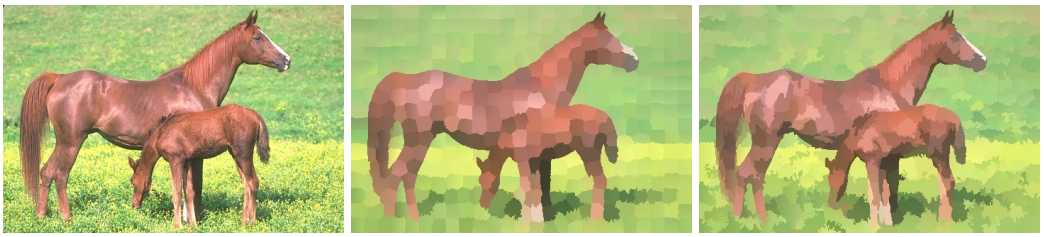


Image initiale I | Compression régulière I_r | Compression irrégulière I_r

FIGURE 6.20 – Exemple de compression d'image avec des décompositions régulières et irrégulières pour $K = 250$ superpixels calculées avec SCALP. Les couleurs reconstituées sont affichées dans chaque superpixel.

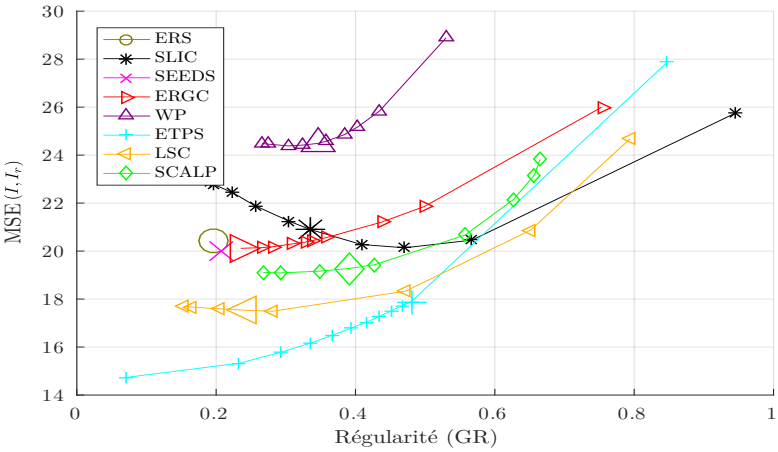
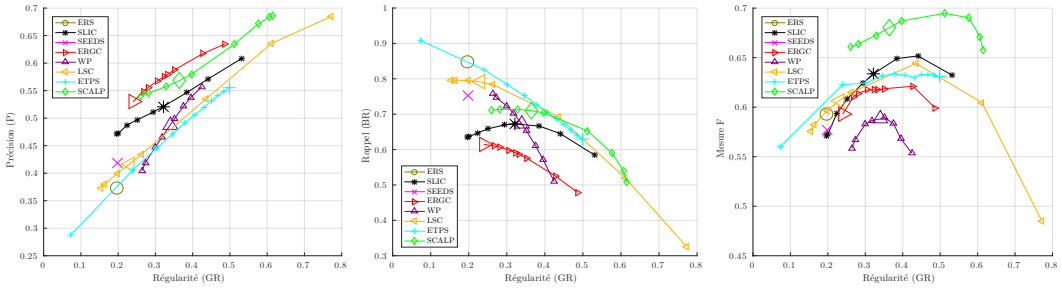


FIGURE 6.21 – Résultats des méthodes de superpixels de l'état-de-l'art sur la compression d'images de la BSD (erreur quadratique moyenne entre l'image initiale et celle reconstruite (6.10)).

Pour évaluer quantitativement l'efficacité de la compression, nous donnons dans la Figure 6.21 l'erreur quadratique moyenne (MSE) entre l'image initiale I et l'image reconstruite I_r :

$$\mathrm{MSE}(I, I_r) = \frac{1}{|I|} \sum_{p \in I} \left(I(p) - I_r(p) \right)^2 . \tag{6.10}$$

Les décompositions des méthodes de superpixels de l'état-de-l'art sont calculées pour plusieurs niveaux de régularité, pour $K = 250$ superpixels, et les résultats de MSE (6.10) sont moyennés sur l'ensemble des images de la BSD. Pour la plupart des méthodes, les paramètres de régularité par défaut ne sont pas optimaux pour la compression. Ces résultats sont logiquement corrélés à l'évaluation du critère EV (voir Figure 6.14(a)) puisque des superpixels homogènes en termes de couleur permettent de calculer des coefficients polynomiaux plus précis. Notons que de telles décompositions sont généralement obtenues pour des régularités plus faibles.



(a) (b) (c)

FIGURE 6.22 – Performances de détection de contours avec la précision P (a), la mesure BR (b) et la mesure F (c) selon plusieurs niveaux de régularité évaluée avec GR. SCALP offre les meilleures performances de détection de contours.

6.4.4 Détection de contours et segmentation

Enfin, les performances de détection de contours et de segmentation sont étudiées pour plusieurs niveaux de régularité. La capacité des méthodes de superpixels à fournir une détection précise des contours de l'image a été étudiée, par exemple dans Arbelaez et al. (2011); Van den Bergh et al. (2012). Une décomposition en superpixels peut naturellement produire plus de régions que le nombre d'objets dans l'image, selon le nombre de superpixels générés. Néanmoins, une carte de contours précise peut être obtenue facilement en moyennant les frontières des décompositions en superpixels obtenues à plusieurs échelles, comme dans la section 5.4.2. La carte moyenne obtenue présente donc des valeurs élevées aux pixels de contours détectés à plusieurs échelles.

Comme indiqué dans la section 4.3.2, la mesure BR n'est pas suffisante pour exprimer les performances de détection de contours, et doit être comparée par exemple à la mesure de précision P (4.11), qui évalue le pourcentage de détections correctes parmi les frontières de superpixels. Les performances de détection de contours peuvent alors être synthétisées par la mesure F (4.11). Dans la suite, les frontières des superpixels sont calculées à plusieurs échelles $K = \{25, \ldots, 1000\}$, et moyennées pour fournir la carte de contours. Pour souligner l'impact du paramètre de régularité sur les performances de détection de contours, les mesures de précision P et de rappel BR, ainsi que les valeurs de mesure F correspondantes sont calculées pour plusieurs niveaux de régularité respectivement dans les Figures 6.22 (a), (b) et (c). La régularité joue clairement sur les performances, puisque la précision et le rappel augmentent et diminuent respectivement avec la mesure GR dans les Figures 6.22 (a) et (b). Par conséquent, un bon compromis pour la mesure F maximale est obtenu pour les valeurs de GR entre 0.35 et 0.45 (Figure 6.22 (c)). De plus, les résultats semblent être significativement affectés par la régularité, de sorte que la plupart des méthodes pourraient obtenir de meilleures performances que celles obtenues avec leur paramètre par défaut. C'est le cas de SCALP, dont paramètre de régularité peut être réglé pour pour obtenir une mesure F allant jusqu'à 0.695.

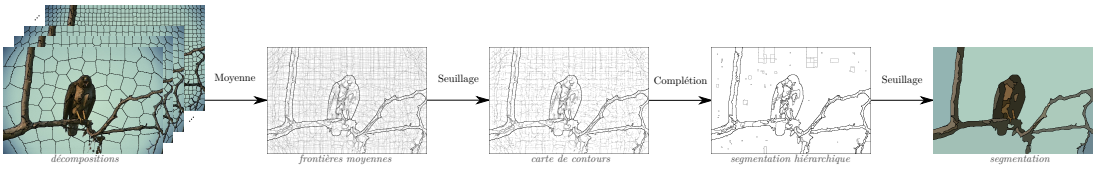


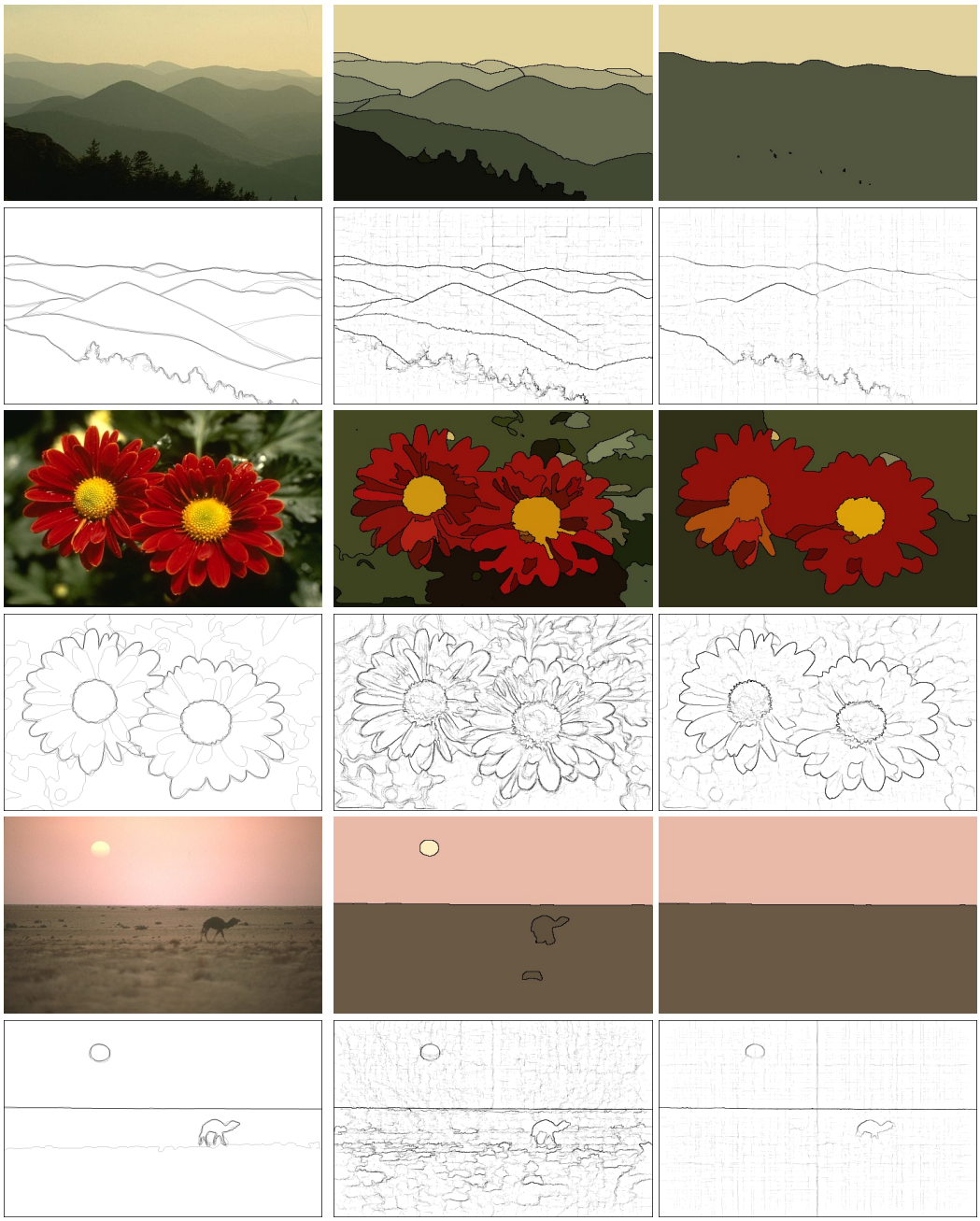
FIGURE 6.23 – Pipeline de segmentation par détection de contours depuis des superpixels. Les décompositions en superpixels sont calculées à plusieurs échelles et leurs frontières sont moyennées pour produire une carte de contours. Les pixels de frontière détectés avec une faible probabilité sont supprimés par une étape de seuillage. À partir de cette carte de contours, une segmentation hiérarchique peut être calculée et seuillée pour fournir une segmentation d'image, que nous illustrons avec les couleurs moyennes des régions.

Enfin, une segmentation peut être générée à l'aide d'une étape de complétion de contours qui calcule des régions fermées depuis la carte de contours (Arbelaez et al., 2009) comme montré dans la section 5.3.1. Le pipeline de détection de contours et de segmentation est illustré dans la Figure 6.23, et nous montrons dans la Figure 6.24 des exemples de détection de contours et de segmentation calculés à partir de décompositions régulières et irrégulières avec SCALP. La moyenne des frontières de superpixels réguliers ne permet pas de capturer efficacement les contours de la vérité terrain (voir Figure 6.24 (c)). La détection des contours des objets n'est pas assez précise pour conduire à une région segmentée avec Arbelaez et al. (2009). Les décompositions irrégulières, contraignent moins les formes des superpixels, qui peuvent mieux adhérer aux contours de l'image. Les contours issus de la moyenne des frontières des superpixels sont alors assez denses pour mener à une segmentation précise (voir Figure 6.24 (b)).

6.4.5 Corrélation entre GR et performances

Nous montrons de façon quantitative la corrélation entre la mesure globale de régularité proposée GR et les performances des métriques de superpixels et des applications présentées, en reportant dans le Tableau 6.2 le coefficient de corrélation de Pearson entre toutes les mesures de régularité : C (6.1), SRC (6.5), J (6.6), SMF (6.8) et GR (6.9), et les performances d'ASA (4.6), UE (4.5), BR (4.7), P (4.11), EV (4.3) et MSE (6.10). Les corrélations sont moyennées sur les décompositions, pour $K = 250$ superpixels, des méthodes qui permettent de paramétrer la régularité : SLIC (Achanta et al., 2012), ERGC (Buyssens et al., 2014), WP (Machairas et al., 2015), LSC (Chen et al., 2017), ETPS (Yao et al., 2015) et SCALP.

Les mesures d'ASA, de BR, de P, de d'EV sont négativement corrélées, quand les mesures d'UE et de MSE sont positivement corrélées avec la régularité. Les métriques SRC et SMF proposées améliorent respectivement la corrélation par rapport à la circularité C, et à la mesure J. Enfin, la mesure GR proposée obtient la plus forte corrélation avec les performances, ce qui démontre sa pertinence par rapport aux métriques de l'état-de-l'art. Enfin, notons que la corrélation entre ASA et UE est en



(a) Images / contours vérité terrain (b) Faible régularité (c) Forte régularité

FIGURE 6.24 – Exemples de détection et de segmentation avec une faible (b) et forte (c) régularité. Les images initiales avec les contours de vérité terrain moyens sont montrées dans (a). Les régions avec les couleurs moyennes sont obtenues après seuillage de la segmentation hiérarchique calculée depuis la carte de contours avec Arbelaez et al. (2009). La mesure GR moyenne est respectivement de 0.256 et 0.481 pour les décompositions irrégulières et régulières.

moyenne de −0.9969, ce qui confirme l'hypothèse de la section 4.3.2, que les deux mesures sont en pratique redondantes.

Tableau 6.2 – Comparaison des coefficients de corrélation de Pearson entre les performances des métriques et des applications considérées dans ce chapitre, et les différentes mesures de régularité. Les résultats moyens sont calculés par rapport à la valeur absolue des corrélations.

	GR (6.9)	SMF (6.8)	J (6.6)	SRC (6.5)	C (6.1)
ASA (4.6)	**−0.5473**	−0.5266	−0.5250	−0.5350	−0.5318
UE (4.5)	**0.5506**	0.5299	0.5284	0.5384	0.5353
BR (4.7)	**−0.9136**	−0.8972	−0.8974	−0.9049	−0.9034
P (4.11)	−0.9627	−0.9656	−0.9645	−0.9688	**−0.9712**
EV (4.3)	**−0.6641**	−0.6428	−0.6426	−0.6528	−0.6503
MSE (6.10)	**0.6760**	0.6554	0.6552	0.6655	0.6636
Moyenne	**0.7190**	0.7029	0.7022	0.7109	0.7093

6.5 Conclusion

Puisqu'aucune métrique de l'état-de-l'art ne permettait de mesurer de façon robuste la propriété de régularité, nous avons proposé dans ce chapitre une nouvelle mesure GR qui évalue la régularité de forme et la consistance des superpixels d'une décomposition. SCALP obtient de hautes performances de régularité, tout en étant la meilleure méthode en termes de respect des objets de l'image et de détection de contours, et en figurant parmi les meilleures en termes d'homogénéité de couleur.

Puisque les méthodes récentes, y compris SCALP, permettent de régler un paramètre de régularité, qui impacte grandement les performances de la décomposition, nous proposons de comparer les méthodes de superpixels à différents niveaux de régularité, évalués avec cette métrique GR. Cela permet notamment de ne pas biaiser le processus de comparaison en n'utilisant ni des paramètres optimaux pour un critère particulier, ni ceux recommandés par les auteurs.

Nous avons également montré que selon l'application, le paramètre de régularité devait être réglé en conséquence, pour permettre à une même méthode d'obtenir de bons résultats. La métrique GR apparaît hautement corrélée aux performances d'applications telles que la recherche de plus proches voisins, le suivi vidéo, la compression d'image ou la détection de contours. À travers ces expériences, nous démontrons également que SCALP convient à un large panel d'applications. La méthode obtient les meilleures performances de segmentation, en fournissant des superpixels réguliers qui permettent la recherche efficace de correspondances de superpixels. La contrainte de régularité peut également être relâchée si l'on souhaite favoriser l'homogénéité de couleur au sein des superpixels, pour par exemple effectuer une compression de l'image.

Troisième partie

Algorithme de correspondance basé sur les superpixels et applications

Résumé

Dans le chapitre 5, nous avons proposé une nouvelle méthode de décomposition en superpixels très précise, mais également très régulière afin de se rapprocher d'une décomposition en blocs réguliers utilisée dans les approches multirésolution. Les superpixels générés réduisent alors le nombre d'éléments de l'image tout en suivant les contours des objets. Néanmoins, la géométrie du voisinage varie d'un superpixel à l'autre, et entre plusieurs décompositions. Dans le contexte de la recherche d'ANN, les méthodes basées sur les patchs ont depuis longtemps démontré la nécessité de considérer l'information de voisinage, afin de capturer le contexte local. Pour étendre cette approche aux superpixels, nous proposons une nouvelle structure de patchs de superpixels, le *superpatch*. Une nouvelle distance de comparaison entre deux superpatchs est également introduite pour permettre la comparaison robuste de voisinage de superpixels.

Dans le chapitre 2, l'algorithme PatchMatch (PM) a été adapté pour la recherche d'ANN basés sur les patchs dans une bibliothèque d'images 3D recalées. Ici, nous introduisons *SuperPatchMatch*, la généralisation de l'algorithme PatchMatch à la recherche d'ANN basés sur les superpixels, décrits par la structure de superpatch. La méthode est également adaptée à la recherche de multiples correspondances dans une bibliothèque d'images. La méthode SuperPatchMatch fournit rapidement de robustes correspondances de superpixels dans une base d'images et peut être utilisée pour de nombreuses applications.

Pour démontrer la robustesse de la méthode proposée, nous appliquons SuperPatchMatch à la segmentation et l'étiquetage de visages depuis une grande bibliothèque d'images décomposées en superpixels. En structurant l'information du voisinage, les superpatchs fournissent de meilleures correspondances que les simples superpixels, et permettent de surpasser des approches de l'état-de-l'art hautement paramétrées et utilisant un apprentissage. Enfin, la méthode SuperPatchMatch de recherche d'ANN basés sur les superpixels est particulièrement intéressante en raison de sa modularité. En effet, nous montrons que l'algorithme peut facilement être contraint pour fournir des correspondances réparties de façon homogène dans l'espace de recherche. Cette contrainte permet alors d'appliquer la méthode au transfert de couleurs entre deux images.

Les contributions présentées dans cette partie visent à se rapprocher du formalisme des méthodes basées sur les patchs pour exploiter au mieux les superpixels et atteindre de meilleures performances. Les superpixels permettent en effet de réduire le nombre d'éléments considérés, tout en respectant les contours de l'image et peuvent mener à de meilleures performances qu'avec un traitement à l'échelle pixellique ou multirésolution considérant des blocs réguliers. Malgré le manque de structure de leur voisinage, les superpixels sont déjà utilisés dans plusieurs applications de vision par ordinateur. Ce travail offre alors de nouvelles perspectives d'amélioration des performances à ces méthodes basées sur les superpixels.

Mots-clés

Algorithmes de correspondance, Patchs de superpixels, PatchMatch, Fusion d'étiquettes, Segmentation, Transfert de couleurs

Contenu

— Le chapitre 7 introduit une nouvelle structure de voisinage de superpixels, le superpatch (section 7.2). Une distance entre ces descripteurs est ensuite proposée pour permettre de comparer efficacement les voisinages de superpixels (section 7.3). La méthode SuperPatchMatch est alors introduite pour trouver rapidement de bonnes correspondances de superpixels décrits par ces superpatchs (section 7.4).

— Le chapitre 8 applique SuperPatchMatch à la segmentation et l'étiquetage à partir d'exemples. Un schéma de fusion est proposé pour permettre l'étiquetage de superpixels depuis des correspondances avec vérité terrain (section 8.2). SuperPatchMatch obtient alors des performances du niveau de l'état-de-l'art sur une expérience de segmentation de visage (section 8.3).

— Le chapitre 9 applique SuperPatchMatch au transfert de couleurs entre deux images. L'algorithme est modifié pour sélectionner de façon globale dans l'image d'exemple, des couleurs à transférer (section 9.2), qui sont fusionnées selon une approche par moyennes non locales (section 9.3). La méthode proposée fournit des résultats compétitifs avec les méthodes de l'état-de-l'art en un temps de calcul très réduit (section 9.4).

Publications associées

Revues internationales avec comité de lecture (1)

[1] *SuperPatchMatch: An Algorithm for Robust Correspondences using Superpixel Patches*
Rémi Giraud, Vinh-Thong Ta, Aurélie Bugeau, Pierrick Coupé et Nicolas Papadakis
IEEE Transactions on Image Processing (TIP), 2017
Sélectionné pour présentation à ICIP 2017

Conférences internationales avec comité de lecture et actes (1)

[1] *Superpixel-based Color Transfer*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
IEEE International Conference on Image Processing (ICIP), 2017

Conférences nationales avec comité de lecture et actes (1)

[1] *Transfert de couleurs basé superpixels*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Groupe d'Études du Traitement du Signal et des Images (GRETSI), 2017

Chapitre 7

La méthode SuperPatchMatch

Sommaire

7.1 Introduction 154
7.2 Définition du superpatch 156
7.3 Comparaison de superpatchs 156
7.3.1 Distance entre deux superpatchs 156
7.3.2 Distance rapide entre deux superpixels 157
7.3.3 Généralisation des patchs réguliers de pixels 158
7.3.4 Robustesse du superpatch 159
7.4 SuperPatchMatch 161
7.4.1 Étape d'initialisation 161
7.4.2 Étape de propagation 162
7.4.3 Étape de recherche aléatoire 163
7.4.4 Bibliothèque d'images d'exemples 164
7.5 Conclusion 164

7.1 Introduction

Depuis leur popularisation pour la synthèse de texture avec Efros et Leung (1999) ou le débruitage avec Buades et al. (2005), les méthodes basées sur les patchs ont obtenu des résultats état-de-l'art sur de nombreuses applications de traitement et d'analyse d'images. Dans la plupart des méthodes basées sur les patchs, une méthode de recherche de plus proches voisins approchés (ANN) est utilisée pour trouver des correspondances. Plusieurs méthodes ont été proposées pour trouver des ANN au sein de la même image, entre deux images ou dans une base d'images (voir section 1.3). Lorsque l'on considère de grandes bases de données, des méthodes de réduction de dimension sont généralement utilisées pour calculer rapidement ces ANN, mais la complexité de ces approches dépend de la taille des données.

Dans le chapitre 2, nous avons étendu l'algorithme PM (Barnes et al., 2009), initialement introduit pour la recherche d'ANN de patchs réguliers 2D entre deux images, à la recherche de plusieurs ANN de patchs 3D dans une bibliothèque. De très bonnes performances de temps de calcul sont obtenues pour une application de la méthode à la segmentation de fines structures anatomiques sur des images médicales recalées. Néanmoins, lorsque la recherche est effectuée pour de larges images, ou dans une bibliothèque d'images, sans a priori de position, la recherche d'ANN peut devenir coûteuse, et nécessiter de nombreuses itérations du processus afin de converger. Un PM multirésolution (Barnes et al., 2010) peut être considéré pour initialiser la carte de correspondance, en fournissant une initialisation grossière. Cependant, comme on l'observe généralement avec de telles architectures hiérarchiques ne respectant pas les contours des objets, les détails sont perdus et des ANN peu précis sont estimés pour les petites structures.

Dans ce contexte, l'utilisation des superpixels peut être intéressante pour diminuer la charge de calcul. Les superpixels groupent en effet les pixels de l'image en zones homogènes tout en respectant les contours des objets (voir la partie II). Par conséquent, une décomposition en superpixels permet de réduire considérablement le nombre d'éléments considérés tout en conservant l'information géométrique, perdue avec les approches multirésolution. Des modèles locaux de correspondance de superpixels ont été proposés, par exemple pour le suivi vidéo (Wang et al., 2011; Reso et al., 2013). Cependant, des algorithmes de recherche d'ANN basés sur les superpixels ont peu été étudiés, ce qui limite l'utilisation de ces décompositions. De récents travaux tels que Rabin et al. (2014); Liu et al. (2016a), qui calculent des correspondances de superpixels entre les décompositions de deux images, utilisent des modèles complexes avec des temps de calcul conséquents.

L'irrégularité des décompositions en superpixels empêche l'utilisation directe des outils basés sur les patchs considérant le voisinage des pixels. La prise en compte de l'information contenue dans le voisinage autour d'un pixel est pourtant cruciale dans le processus de recherche de correspondances, et les descripteurs standards de vision par ordinateur intègrent cette information (Lowe, 2004; Bay et al., 2006; Calonder et al., 2010; Dellinger et al., 2015). Par exemple dans Garnier et al. (2012); Clément et

al. (2015), les histogrammes de forces (Matsakis et Wendling, 1999) sont utilisés pour stocker les relations spatiales entre objets. Les approches de *deformable part models* (Felzenszwalb et al., 2010; Trulls et al., 2014) ou de boîtes englobantes de *poselets* (Sharma et al., 2013) ont également été appliquées avec succès à la segmentation ou la reconnaissance d'objets. La nécessité d'inclure l'information spatiale dans les descripteurs a également été étudiée dans Bloch (2005) qui s'intéresse aux approches floues pour caractériser les relations spatiales entre des éléments proches.

Par conséquent, afin d'à la fois réduire le nombre d'éléments à traiter, de conserver l'information géométrique, et de trouver des correspondances précises, il apparaît nécessaire de considérer des descripteurs spatiaux intégrant le voisinage, afin de pouvoir définir des méthodes précises de correspondances de superpixels. Des tentatives d'utilisation de l'information de voisinage des superpixels ont été proposées (Pei et al., 2014; Sawhney et al., 2014). Cependant, ces méthodes ne sont pas adaptées au contexte de la recherche d'ANN, puisqu'elles effectuent une régularisation sur un graphe construit depuis les superpixels voisins, mais n'incluent pas directement l'information spatiale relative entre superpixels dans une structure dédiée.

À notre connaissance, la combinaison d'un algorithme de recherche d'ANN comme PM avec les superpixels a seulement été proposée dans Gould et al. (2014); Zheng et Li (2015), qui calculent des correspondances entre simples superpixels en utilisant des mouvements similaires à ceux de PM. Par exemple, dans Gould et al. (2014), la distance entre les descripteurs de superpixels est pré-calculée par apprentissage, mais les résultats de classification à partir d'exemples n'atteignent pas ceux de méthodes de l'état-de-l'art. Dans Lu et al. (2013), PM est considéré de façon plus réduite pour l'estimation de flot optique. L'algorithme est utilisé pour raffiner les résultats à l'intérieur de la boîte rectangulaire englobant le superpixel traité. Le but de notre travail est donc complètement différent puisque nous souhaitons comparer des voisinages de superpixels définis sur des domaines irréguliers de l'image.

Contributions

Dans ce chapitre, nous proposons une nouvelle structure de voisinage de superpixels appelée *superpatch*. Puisque les voisinages des superpixels ne sont pas nécessairement les mêmes, en termes de formes ou de nombre d'éléments, une mesure pour comparer des superpatchs est également introduite. Une nouvelle méthode, appelée *SuperPatchMatch*, qui étend l'algorithme PM, est proposée pour effectuer des recherches rapides et précises d'ANN de superpatchs dans une bibliothèque d'images.

Pour démontrer quantitativement l'intérêt de notre méthode, nous proposons d'appliquer SuperPatchMatch à la segmentation et l'étiquetage depuis une bibliothèque d'images d'exemples. SuperPatchMatch est particulièrement adapté pour gérer de grandes bases d'images en constante évolution puisqu'aucune phase d'entraînement n'est nécessaire, contrairement aux approches basées sur l'apprentissage automatique supervisé (He et al., 2004; Kae et al., 2013), ou les méthodes récentes utilisant des réseaux de neurones (Liu et al., 2015a; Long et al., 2015).

7.2 Définition du superpatch

De façon similaire au patch de pixels, un *superpatch* est un patch (un ensemble) de superpixels voisins spatialement. Considérons une image A, décomposée par une quelconque méthode de superpixels, en $|A|$ éléments. Un superpatch $\mathbf{A_i}$ est centré sur un superpixel A_i et est composé de ses superpixels voisins $A_{i'}$ tel que : $\mathbf{A_i} = \{A_{i'}, \text{ tel que } ||c_i - c_{i'}||_2 \leq R\}$, avec c_i le barycentre spatial des pixels contenus dans A_i. En d'autres termes, le superpatch centré sur un superpixel est défini en considérant tous les superpixels dans un rayon de recherche fixé $R \geq 0$. Notons que chaque superpatch $\mathbf{A_i}$ contient au moins le superpixel A_i. La Figure 7.1 illustre la définition du superpatch.

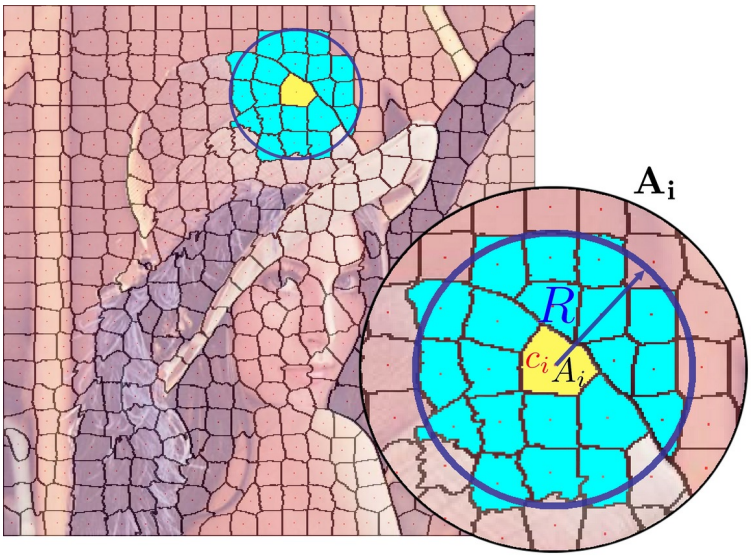


FIGURE 7.1 – Illustration de la définition du superpatch. En bleu : cercle de recherche de rayon R centré sur c_i, le barycentre de A_i. Le superpatch $\mathbf{A_i}$ est composé par tous les superpixels ayant leur barycentre à l'intérieur du cercle.

Cette définition basée sur le barycentre permet de considérer les superpixels réellement proches spatialement. Dans le cas de décomposition irrégulières, avec des superpixels très étirés, définir le superpatch avec par exemple tous les superpixels qui s'intersectent avec le disque de rayon R pourrait entraîner l'inclusion dans le superpatch de superpixels très éloignés.

7.3 Comparaison de superpatchs

7.3.1 Distance entre deux superpatchs

La comparaison entre deux patchs réguliers est généralement effectuée en calculant une somme des différences au carré ou SSD (1.1). Dans le cadre usuel, les deux

patchs réguliers ayant la même taille, cette distance est calculée entre les intensités des pixels correspondants spatialement dans chaque patch. Lorsque l'on considère deux superpatchs, leur nombre d'éléments et leur géométrie sont généralement différents, ce qui rend difficile leur comparaison. Dans la suite, pour illustrer la méthode, nous considérons deux superpatchs $\mathbf{A_i}$ et $\mathbf{B_j}$, dans deux images potentiellement différentes A et B, décomposées en superpixels de tailles similaires. Notons toutefois que les superpatchs peuvent aussi se trouver dans la même image. Pour comparer deux superpatchs, nous proposons tout d'abord de recaler les positions relatives de tous les superpixels appartenant aux superpatchs. Pour superposer deux superpatchs, toutes les positions $c_{j'}$ des superpixels $B_{j'} \in \mathbf{B_j}$ sont recalées par le vecteur $v_{ij} = c_i - c_j$, où c_i et c_j sont les barycentres spatiaux respectifs de A_i et B_j. Contrairement au schéma classique des patchs réguliers, le nombre et la géométrie des éléments contenus dans deux superpatchs sont susceptibles d'être différents. Ainsi, un superpixel recalé $B_{j'}$ peut se superposer avec plusieurs superpixels $A_{i'}$, et cette information doit être considérée. Le recalage d'un superpatch $\mathbf{B_j}$ sur un superpatch $\mathbf{A_i}$ est illustré dans la Figure 7.2.

Pour calculer une distance entre des structures irrégulières, telles que les superpixels, les auteurs de Sawhney et al. (2014) proposent d'utiliser la distance d'édition. Cependant, une telle distance calcule des correspondances uniques entre deux éléments de la structure et ne peut pas traiter efficacement la superposition de superpixels qui nécessite le calcul de correspondances entre un et plusieurs éléments. Une autre limite est le fait que cette distance mélange deux types d'information : la similarité de deux éléments et le coût d'enlever ou d'ajouter des superpixels. Ainsi, la distance d'édition doit être réglée avec précision selon l'application.

Pour définir une métrique pertinente entre des superpatchs, il est donc nécessaire de considérer leur géométrie. Nous proposons de définir la distance symétrique D entre deux superpatchs $\mathbf{A_i}$ et $\mathbf{B_j}$ comme :

$$D(\mathbf{A_i}, \mathbf{B_j}) = \frac{\sum_{i' \in \mathcal{I}_i^A} \sum_{j' \in \mathcal{I}_j^B} w(A_{i'}, B_{j'}) d(F_{i'}^A, F_{j'}^B)}{\sum_{i' \in \mathcal{I}_i^A} \sum_{j' \in \mathcal{I}_j^B} w(A_{i'}, B_{j'})}, \tag{7.1}$$

où d est la distance Euclidienne entre les descripteurs des superpixels $F_{i'}^A$ et $F_{j'}^B$, par exemple, les coordonnées de c_i, la couleur moyenne, ou tout descripteur de superpixel, w est un poids qui dépend de la position relative des superpixels $A_{i'}$ et $B_{j'}$, et $\mathcal{I}_i^A = \{i', \text{ tel que } A_{i'} \in \mathbf{A_i}\}$, est l'ensemble des indices des superpixels $A_{i'} \in \mathbf{A_i}$.

7.3.2 Distance rapide entre deux superpixels

Pour calculer le poids entre deux superpixels $A_{i'}$ et $B_{j'}$, nous voudrions idéalement mesurer l'aire de leur superposition relative, c'est-à-dire, calculer $w(A_{i'}, B_{j'}) = |A_{i'} \cap B_{j'}| / |A_{i'} \cup B_{j'}|$. Néanmoins, ce poids nécessite le calcul coûteux de tous les pixels qui se superposent, ce qui annulerait l'avantage de la représentation en superpixels.

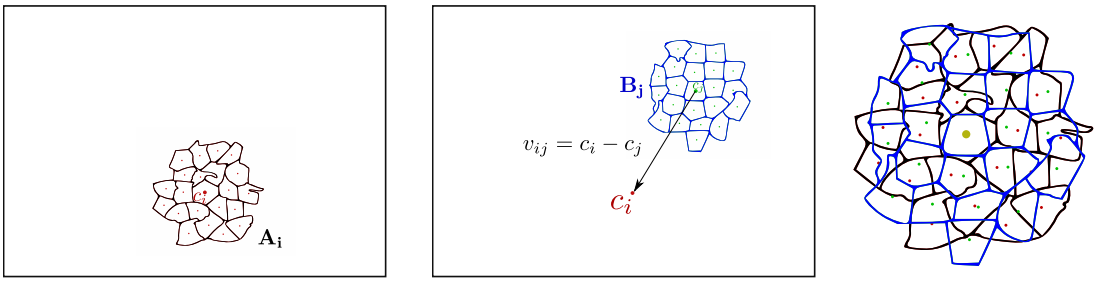


FIGURE 7.2 – Recalage d'un superpatch $\mathbf{B_j}$ sur un superpatch $\mathbf{A_i}$. Tous les superpixels de $\mathbf{B_j}$ sont recalés par le vecteur v_{ij}, afin que les barycentres des superpixels centraux A_i et B_j soient superposés.

Une méthode rapide serait de comparer un superpixel $A_{i'}$ uniquement au superpixel le plus proche $B_{j'}$, mais nous proposons une méthode plus robuste qui considère une distance spatiale entre les barycentres des superpixels. Nous définissons le poids symétrique spatial entre deux superpixels $A_{i'}$ et $B_{j'}$ tel que :

$$w(A_{i'}, B_{j'}) = \exp\left(-\frac{\|x_{i'j'}\|_2^2}{\sigma_1^2}\right) w_s(A_{i'}) w_s(B_{j'}), \tag{7.2}$$

où $x_{i'j'} = c_{j'} - c_{i'} + v_{ij}$ est la distance relative entre $c_{i'}$ et $c_{j'}$, $w_s(A_{i'})$ pondère l'influence de $A_{i'}$ par rapport à sa distance à A_i tel que $w_s(A_{i'}) = \exp(-\|c_i - c_{i'}\|_2^2 / \sigma_2^2)$, avec σ_1 et σ_2 des paramètres d'échelle. Le réglage de σ_1 dépend de l'échelle de la décomposition en superpixels. Puisque les superpatchs ont été recalés, et que le but est de comparer un superpixel $A_i \in \mathbf{A_i}$ aux plus proches $B_j \in \mathbf{B_j}$, σ_1 peut être fixé à la moitié de la taille moyenne d'un superpixel, c'est-à-dire, la moitié de la distance moyenne entre les barycentres de deux superpixels, tel que $\sigma_1 = \frac{1}{2}\sqrt{|I|/K}$, pour une image $|I|$ décomposée en K superpixels. Enfin, σ_2 dépend de la taille du superpatch et peut être réglé à $\sqrt{2}R$ pour favoriser la contribution des plus proches superpixels.

7.3.3 Généralisation des patchs réguliers de pixels

Dans le cas limite où chaque superpixel ne contient qu'un pixel, c'est-à-dire $A_{i'}{=}c_{i'}$ et $B_{j'}{=}c_{j'}$, alors $\mathbf{A_i}$ et $\mathbf{B_j}$ ont la même structure régulière et le même nombre d'éléments. Avec $\sigma_1 \to 0$ et $\sigma_2 \to \infty$ (7.2), $w(A_{i'}, B_{j'}) = w(c_{i'}, c_{j'}) = 1$ si $c_{i'} - c_i = c_{j'} - c_j$ et 0 sinon, et la distance proposée D (7.1) est une généralisation de la distance entre patchs réguliers, puisqu'elle est réduite a une SSD normalisée standard entre deux patchs de pixels :

$$D(\mathbf{A_i}, \mathbf{B_j}) = \frac{1}{|\mathbf{A_i}|} \sum_{i' \in \mathcal{I}_i^A} \sum_{j' \in \mathcal{I}_j^B} d(F_{i'}, F_{j'}) \delta_{c_{i'} - c_i = c_{j'} - c_j}, = \mathrm{SSD}(\mathbf{A_i}, \mathbf{B_j}), \tag{7.3}$$

où $\delta_a = 1$ quand a est vrai et 0 sinon.

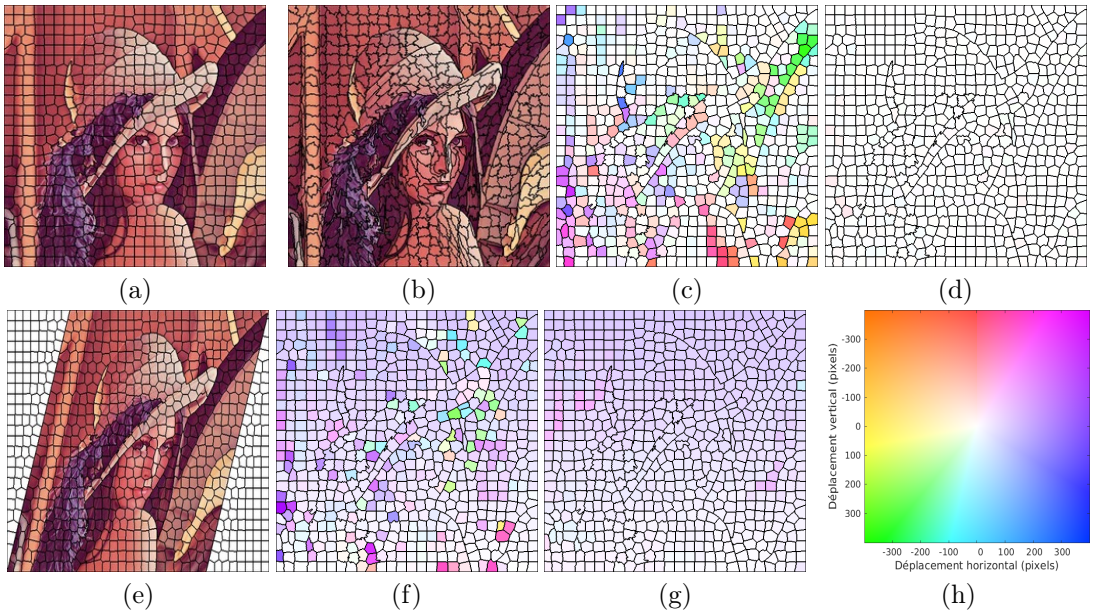


FIGURE 7.3 – Superpixels vs superpatchs pour la correspondance de superpixels. (a) et (b): deux décompositions par SCALP et Buyssens et al. (2014). (c) et (d): résultat de correspondances de superpixels (Gould et al., 2014) et notre résultat de correspondances de superpatch. La même expérience est effectuée entre (a) et l'image cisaillée (e), pour laquelle nous présentons les résultats de correspondances de superpixels (Gould et al., 2014) (f) et de superpatchs (g). Le déplacement est illustré avec la représentation standard du flot optique (h). Plus il est proche du blanc, plus le déplacement est faible.

7.3.4 Robustesse du superpatch

Les Figures 7.3 et 7.4 illustrent plusieurs expériences démontrant que les superpatchs permettent de trouver des correspondances de superpixels plus pertinentes que celles obtenues en utilisant uniquement les superpixels (Gould et al., 2014). Dans la Figure 7.3, deux décompositions sont calculées sur la même image en utilisant SCALP et Buyssens et al. (2014) (Figures 7.3(a) et (b)). Le but est de trouver la meilleure correspondance de superpixels entre (a) et (b) en termes de descripteurs de superpixels (ici norme ℓ_2 sur les histogrammes couleurs normalisés dans l'espace RGB). Nous représentons l'amplitude du déplacement entre les correspondances avec la représentation de flot optique standard (Figure 7.3(h)). Lorsque l'on utilise uniquement les superpixels, comme dans Gould et al. (2014), plusieurs erreurs de correspondances sont obtenues (Figure 7.3(c)), tandis que la correspondance de superpatchs fournit des correspondances très précises (Figure 7.3(d)).

La même expérience sur une image cisaillée décomposée avec Achanta et al. (2012) (Figure 7.3(e)), fournit un déplacement uniforme (Figure 7.3(g)) qui indique des correspondances pertinentes entre les superpatchs et démontre la robustesse de la structure proposée aux déformations géométriques par rapport aux méthodes utilisant uniquement des comparaisons de superpixels.

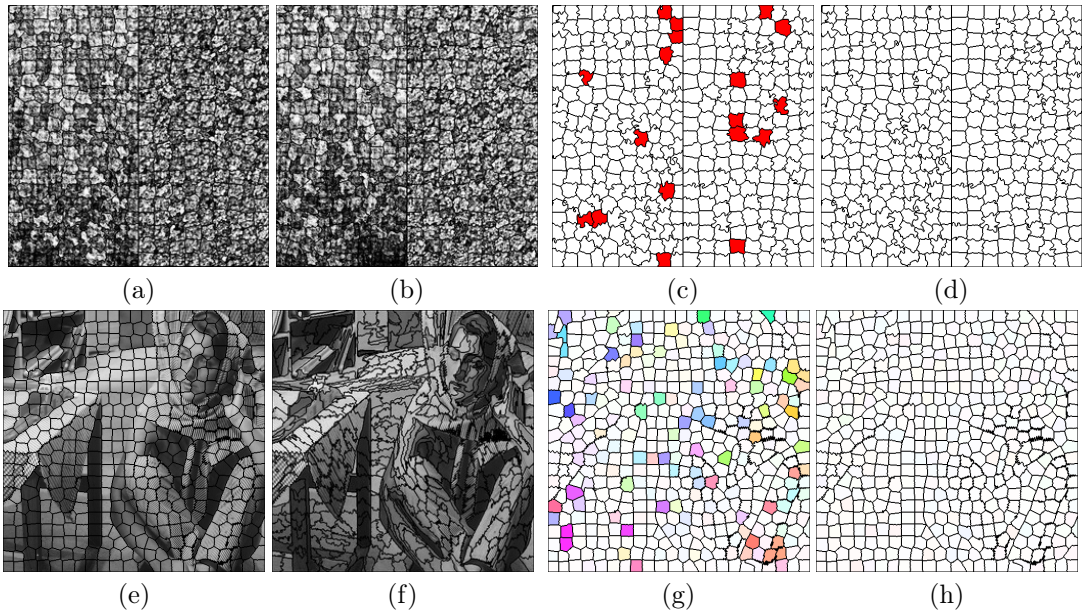


FIGURE 7.4 – Correspondances de superpixels sur des images texturées. Deux parties différentes de deux textures proches sont combinées dans (a) et (b) et décomposées avec SCALP. (c) et (d): résultat de correspondances de superpixels (Gould et al., 2014) et de superpatchs où les superpixels rouges indiquent une correspondance vers la mauvaise texture. (e) et (f): deux décompositions d'une image naturelle texturée en utilisant SCALP et Buyssens et al. (2014). Comparaison des correspondances de superpatchs obtenues avec un descripteur de couleur (g) et avec la combinaison de descripteurs de couleur et de texture (h).

Dans la Figure 7.4, deux images de textures proches sont chacune divisées verticalement en deux parties, qui sont combinées pour créer deux nouvelles images contenant les deux textures (Figures 7.4(a) et (b)). La même expérience de recherche de correspondance est effectuée entre ces deux images et nous représentons les mauvaises textures capturées par des superpixels rouges dans les Figures 7.4(c) et (d).

Enfin, nous montrons que pour une image naturelle contenant des textures (Figures 7.4(e) et (f)), la combinaison de descripteurs de couleur et de texture (histogrammes de gradients orientés (HoG) (Dalal et Triggs, 2005)) peut fournir des correspondances de superpatchs plus précises (Figure 7.4(h)).

Ces exemples démontrent clairement les avantages du superpatch qui contient l'information spatiale entre superpixels voisins. Ils démontrent également que notre méthode permet de considérer n'importe quelles structures irrégulières définies sur le domaine de l'image et qu'elle peut être appliquée à des images texturées.

7.4 SuperPatchMatch

Nous proposons à présent la méthode SuperPatchMatch (SuperPatchMatch (SPM)), une extension de l'algorithme PatchMatch (PM) (Barnes et al., 2009), décrit en détails dans les sections 1.3 et 2.2, dédiée à notre structure de superpatch pour la recherche rapide d'ANN basés sur les superpixels. Le manque de géométrie régulière entre superpatchs est le principal problème pour adapter l'algorithme PM. La notion de voisins adjacents doit notamment être définie pour une décomposition irrégulière de l'image.

Dans cette section, seules les relations d'adjacence directes ont besoin d'être considérées pour définir notre algorithme. Néanmoins, comme pour les pixels décrits par des patchs réguliers, nous montrons dans les Figures 7.3 et 7.4, ainsi que dans le chapitre 8 que l'information de voisinage avec les superpatchs est nécessaire pour que la méthode de recherche soit efficace pour calculer des correspondances précises de superpixels.

Dans la suite, comme dans la section 7.3, nous illustrons la méthode proposée en considérant deux superpatchs $\mathbf{A_i}$ et $\mathbf{B_j}$, dans deux images différentes A et B, mais notre approche peut être appliquée à une bibliothèque d'images, comme nous le montrons dans le chapitre 8. L'ANN d'un superpixel $A_i \in A$, est noté $B_{(i)} = B_{\mathcal{M}(i)}$, avec $\mathcal{M}$ la carte qui stocke pour les superpixels de A, l'indice dans B de leur ANN.

7.4.1 Étape d'initialisation

Comme dans l'algorithme initial, pour chaque superpixel $A_i \in A$, nous assignons aléatoirement un superpixel $B_{(i)} \in B$. La Figure 7.5 montre quelques exemples d'initialisation. Après cette étape, une propagation et une recherche aléatoire sont effectuées itérativement pour chaque superpixel afin d'améliorer les correspondances initiales.

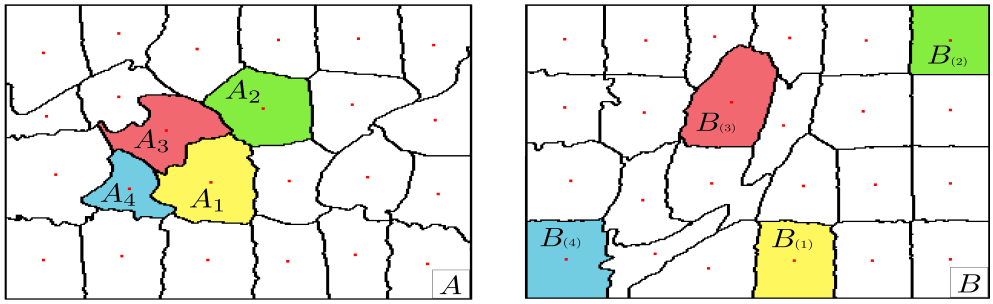


FIGURE 7.5 – Étape d'initialisation de SPM. Chaque superpixel $A_i \in A$ est associé aléatoirement à un superpixel $B_{(i)} \in B$.

7.4.2 Étape de propagation

Dans Barnes et al. (2009), l'étape de propagation tente d'améliorer les ANN actuels en utilisant ceux des voisins adjacents dans A. Les pixels sont traités selon un parcours pré-défini, par exemple depuis le coin haut-gauche au coin bas-droit. La propagation considère uniquement les éléments adjacents traités durant l'étape courante, par exemple les pixels haut-gauche. Leur ANN est décalé afin de respecter les positions relatives des pixels dans A. Le pixel à cette position décalée dans B est alors considéré comme candidat pour améliorer la correspondance du pixel dans A.

Pour les superpixels, nous définissons d'abord l'ordre de traitement des superpixels selon le même sens de parcours à l'échelle pixellique, et proposons de considérer les ANN de tous les superpixels adjacents déjà traités durant l'étape courante. La sélection des candidats depuis les superpixels adjacents est illustrée dans la Figure 7.6. La sélection des voisins adjacents n'est cependant pas directe puisqu'il n'existe pas de géométrie régulière entre les éléments.

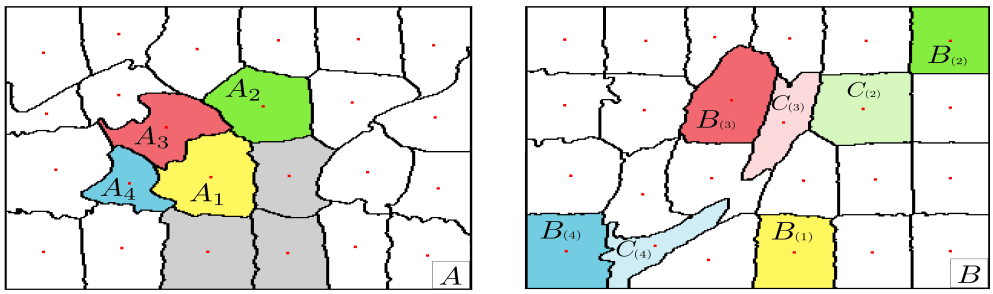


FIGURE 7.6 – Étape de propagation de SPM. Seuls les superpixels en position haut-gauche par rapport à A_1 sont considérés durant les itérations paires (A_2, A_3 et A_4). Les voisins restant (en gris) sont testés durant les itérations impaires. Les correspondances actuelles sont notées $B_{(i)}$, et les $C_{(i)}$ représentent les nouveaux candidats à tester comme ANN.

Quand pour un superpixel A_i, son superpixel adjacent $A_{i'}$ et son ANN $B_{(i')}$ sont considérés, un voisin de $B_{(i')}$ est sélectionné comme candidat pour tenter d'améliorer la correspondance de A_i. Cependant, l'ANN $B_{(i')}$ ne peut pas être décalé comme pour les patchs réguliers, puisque la géométrie autour des deux superpixels considérés est différente. Pour tenter d'améliorer la correspondance de A_i, un superpixel adjacent à $B_{(i')}$, noté $C_{(i')}$ doit donc être sélectionné comme candidat pour améliorer l'ANN de A_i. Il est alors important de conserver les positions relatives entre les voisins spatiaux dans les images A et B durant cette étape pour avoir une propagation plus rapide de grandes zones de correspondance (Barnes et al., 2009). Nous proposons de sélectionner le superpixel candidat $C_{(i')}$ en fonction de sa position relative à $B_{(i')}$, de sorte qu'elle soit la plus similaire à $\theta_{ii'}$, l'angle entre A_i et $A_{i'}$, calculé depuis les barycentres des superpixels :

$$C_{(i')} = \underset{k \in \mathcal{N}^B_{\mathcal{M}(i')}}{\operatorname{argmin}} \left\| (\theta_{ii'} + \pi) - \theta_{i'k} \right\|_1 , \tag{7.4}$$

avec $\theta_{i'k}$ l'angle entre $B_{(i')}$ et l'ensemble $\mathcal{N}^B_{\mathcal{M}(i')}$ de ses superpixels adjacents $B_k \in B$.

Dans la Figure 7.7, nous illustrons la sélection du candidat $C_{(3)}$ pour le voisin A_3, qui est en haut à gauche de A_1. Afin de conserver la même position relative, le nouveau superpixel à tester comme ANN pour A_1, est le voisin de $B_{(3)}$ qui est le plus proche de sa position bas-droite, par rapport à l'angle $\theta_{i'k}$.

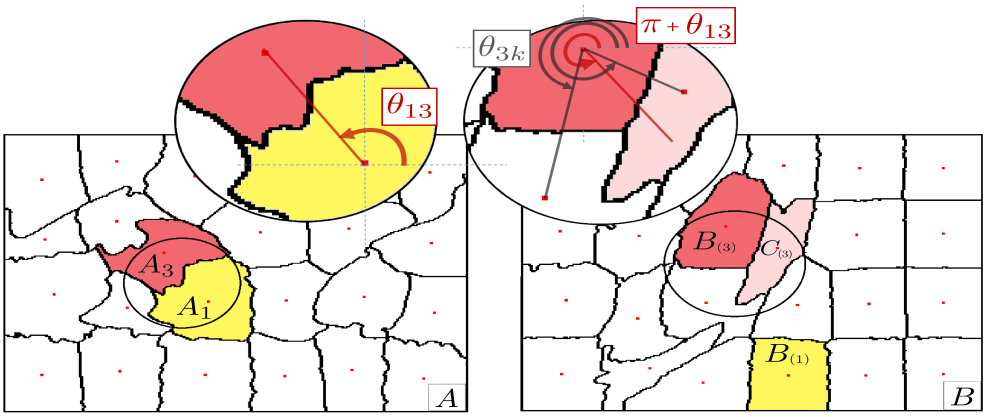


FIGURE 7.7 – Afin d'améliorer l'ANN du superpixel A_1, le voisin A_3 est considéré. Son ANN actuel est $B_{(3)}$. Le superpixel sélectionné $C_{(3)}$ à tester comme candidat est le superpixel adjacent à $B_{(3)}$ avec l'orientation relative la plus similaire, c'est-à-dire le superpixel B_k dont l'angle avec $B_{(3)}$, θ_{3k}, est le plus proche de $\pi + \theta_{13}$.

7.4.3 Étape de recherche aléatoire

L'étape de recherche aléatoire consiste à échantillonner dans B afin d'échapper à un potentiel minimum local (Barnes et al., 2009). Les candidats sont sélectionnés à l'intérieur de zones de recherche centrées autour du barycentre de l'ANN actuel, et dont la taille diminue d'un facteur 2 jusqu'à ne contenir qu'un seul pixel. Dans le cas des superpixels, une position dans B est sélectionnée aléatoirement et le superpixel la contenant est le candidat à tester. La Figure 7.8 illustre la recherche aléatoire, où les zones de recherche sont représentées par des lignes pointillées.

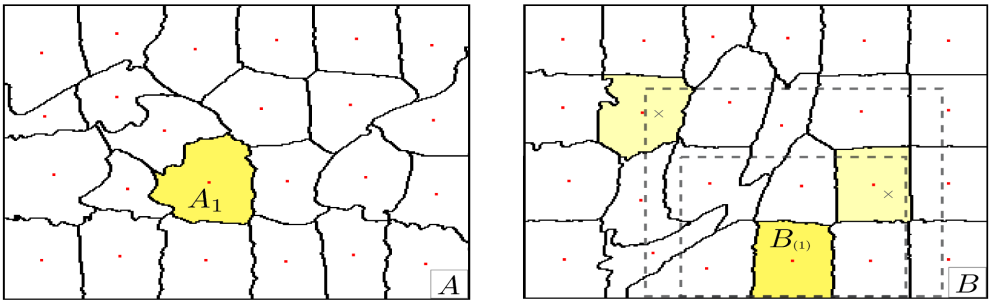


FIGURE 7.8 – Étape de recherche aléatoire de SPM. L'échantillonnage est effectué à une distance décroissante du barycentre de l'ANN actuel $B_{(1)}$ du superpixel A_1. Les superpixels contenant les positions sélectionnées (croix) sont testés pour améliorer l'ANN de A_1.

7.4.4 Bibliothèque d'images d'exemples

Comme pour PM, un avantage majeur de SPM est sa complexité, qui ne dépend que de la taille de l'image à traiter et pas de la taille de la bibliothèque d'images T considérée. De façon similaire à la méthode OPAL présentée dans le chapitre 2, de nombreuses images peuvent donc être considérées dans T sans étape de présélection. Les adaptations de SPM au cas d'une bibliothèque d'exemples sont directes. L'initialisation est étendue afin de sélectionner aléatoirement des correspondances dans T. L'étape de propagation teste toujours l'ANN décalée des voisins spatiaux dans l'image test. Ces voisins n'ont alors pas nécessairement des ANN se trouvant dans la même image de la bibliothèque, et permettent donc à un superpixel d'accéder potentiellement à de nouvelles images de T.

Enfin, concernant l'étape de recherche aléatoire, contrairement au cas d'OPAL dans le chapitre 2 et de son application à la segmentation d'images médicales dans le chapitre 3, avec les superpixels, le nombre d'éléments de l'image peut être largement inférieur à celui du nombre d'images d'exemples. Une image dans T qui ne serait pas accédée par une correspondance aléatoire durant l'initialisation ne pourrait donc pas être accédée durant le processus. Nous effectuons donc la recherche aléatoire dans la meilleure image courante et également dans une image tirée aléatoirement dans T, afin d'augmenter la probabilité d'accéder à toutes les images de la bibliothèque, comme dans Shi et al. (2013).

Enfin, contrairement à PM, qui ne calcule qu'un ANN, nous proposons de calculer k-ANN dans la bibliothèque T avec SPM. De façon similaire à la méthode OPAL, on choisit d'effectuer la recherche des k-ANN par k SPM indépendants afin de faciliter l'implémentation parallèle.

7.5 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle structure de patch de superpixels, qui peut considérer des décompositions irrégulières de l'image. Comme pour les patchs réguliers de pixels, ces superpatchs incluent la géométrie du voisinage des superpixels. De meilleures correspondances sont donc obtenues en comparant des superpatchs plutôt que de simples superpixels. La structure proposée répond alors aux limites actuelles d'utilisation du voisinage des superpixels.

Nous introduisons également SuperPatchMatch, l'extension de l'algorithme de correspondance PatchMatch à ces patchs de superpixels. L'irrégularité des décompositions nécessite d'adapter l'algorithme initial, que nous étendons également, à la recherche de multiples correspondances dans une bibliothèque d'images. Pour démontrer la robustesse de la méthode, nous appliquons respectivement SuperPatchMatch dans les chapitres 8 et 9, à la segmentation et au transfert de couleurs à partir d'exemple. Nous montrons notamment dans le chapitre 8, que les superpatchs permettent d'atteindre une précision similaire à celles obtenues avec des approches hautement paramétrées et utilisant un apprentissage.

Chapitre 8

Application à la segmentation et l'étiquetage basés sur les superpixels

Sommaire

8.1 Introduction 168
8.2 Schéma de fusion d'étiquettes 168
8.3 Validation de SuperPatchMatch sur l'étiquetage de visages 169
8.3.1 Réglage des paramètres 170
8.3.2 Influence des paramètres 170
8.3.3 Comparaison avec les méthodes de l'état-de-l'art 175
8.4 Conclusion 177

8.1 Introduction

Dans ce chapitre, nous appliquons la méthode SuperPatchMatch (SPM) à la segmentation et l'étiquetage d'images à partir d'exemples. L'algorithme SPM peut fournir pour chaque superpixel d'une image, une ou plusieurs correspondances dans plusieurs autres images. Dans la section 2.3, nous avons présenté un schéma de fusion d'étiquettes à l'échelle pixellique, appliqué à la segmentation d'images médicales dans le chapitre 3. De façon similaire, les correspondances de superpixels obtenues avec SPM peuvent permettre d'extraire de l'information sémantique, représentée par une étiquette. L'utilisation de SPM est particulièrement intéressante dans ce contexte, puisque les superpixels décomposent l'image en zones homogènes, en respectant les contours des objets de l'image, permettant ainsi un étiquetage plus précis et robuste qu'avec une approche par patch (voir Figure 4.2). Nous proposons un schéma de fusion d'étiquettes, pour permettre l'étiquetage d'une image présegmentée en superpixels, depuis des correspondances calculées par SPM dans des images disposant d'une vérité terrain d'étiquetage. Notre méthode obtient des performances état-de-l'art d'étiquetage de visages sur la base de référence LFW (pour Labeled Faces in the Wild, Huang et al. (2007b)).

8.2 Schéma de fusion d'étiquettes

Nous proposons d'adapter le schéma de fusion d'étiquettes décrit dans la section 2.3 pour permettre l'étiquetage multi-étiquettes de superpixels depuis une bibliothèque d'images d'exemples. On considère une bibliothèque d'exemples T, et on dénote $l(T_j) \in \{1, \ldots, M\}$ l'étiquette du superpixel $T_j \in T$. En sortie de l'algorithme SPM, k-ANN sont estimés pour chaque superpixel de l'image A. Afin d'obtenir l'étiquetage final d'un superpixel A_i, dont l'ensemble des k ANN T_j, ayant pour étiquettes $l(T_j) = m$, est noté $\mathcal{K}_i^m = \{T_j\}$, la carte de fusion d'étiquettes $L_m(A_i)$ est définie par :

$$L_m(A_i) = \frac{\sum_{T_j \in \mathcal{K}_i^m} \omega(A_i, T_j)}{\sum_{m=1}^{M} \sum_{T_j \in \mathcal{K}_i^m} \omega(A_i, T_j)}, \tag{8.1}$$

où $\omega(A_i, T_j)$ est le poids assigné à l'étiquette m, et dépend de la similarité entre le superpatch $\mathbf{A_i} \in A$, et le superpatch ANN $\mathbf{T_j} \in T$. Cette carte d'étiquettes L_m donne la probabilité d'assigner l'étiquette m au superpixel A_i.

Certains types de données, comme les images médicales ou certaines bases spécifiques, considèrent des images recalées où les structures d'intérêt entre les images sont proches spatialement. Sur ces images, de bonnes correspondances de superpatchs ne devraient pas être spatialement trop éloignées entre les images comparées, et cette proximité peut également être utilisée afin de renforcer la cohérence spatiale des k-ANN sélectionnés. Chaque contribution d'ANN à l'étiquetage peut donc être pondérée par la distance spatiale entre les barycentres $c_i \in A$ et $c_j \in T$, et le poids

utilisé dans l'Éq. (8.1) peut être défini comme :

$$\omega(A_i, T_j) = \exp\left(1 - \left(\frac{D(\mathbf{A_i}, \mathbf{T_j})}{h(A_i)^2} + \frac{\|c_i - c_j\|_2}{\beta^2}\right)\right), \tag{8.2}$$

où $h(A_i)^2 = \alpha^2 \min_{T_j \in \bigcup_m \mathcal{K}_i^m} (D(\mathbf{A_i}, \mathbf{T_j}) + \epsilon)$, où $\epsilon \to 0$, et α et β sont des paramètres d'échelle. Notons que si $\beta = \infty$, cela revient à ne pas considérer d'a priori de position. Avec la fonction $h(A_i)$, la distance des contributions à l'étiquetage de A_i est divisée par la distance minimale parmi toutes les contributions des k-ANN.

L'étiquetage final $\mathcal{L}(A_i)$ est enfin obtenu avec l'étiquette ayant la plus forte probabilité :

$$\mathcal{L}(A_i) = \underset{m \in \{1,\ldots,M\}}{\operatorname{argmax}}\ L_m(A_i). \tag{8.3}$$

La relation (8.3) donne une décision à l'échelle superpixel qui peut s'avérer irrégulière. Comme dans Pei et al. (2014), les cartes d'étiquetage L (8.1) peuvent être utilisées comme un terme de données multi-étiquettes et considérées dans le problème de régularisation suivant, où une énergie E est définie sur le graphe construit depuis les superpixels adjacents :

$$E(\mathcal{L}) = \sum_{i=1}^{|A|} \sum_{i' \in \mathcal{I}_i^A} \exp\left(-\frac{d(F_i, F_{i'})}{\gamma}\right) \delta_{\mathcal{L}(A_i) \neq \mathcal{L}(A_{i'})} + \sum_{i=1}^{|A|} (1 - L_{\mathcal{L}(A_i)}(A_i)), \tag{8.4}$$

où γ est un paramètre de régularisation, le terme de données $1 - L_{\mathcal{L}(A_i)}(A_i)$ est proche de 0 (respectivement 1) quand la probabilité de l'étiquette $\mathcal{L}(A_i)$ est forte (respectivement faible), $\delta_a = 1$ quand a est vrai et 0 sinon, et $\mathcal{I}_i^A = \{i', \text{ tel que } A_{i'} \in \mathbf{A_i}\}$ est l'ensemble des indices des superpixels $A_{i'} \in \mathbf{A_i}$. La minimisation de cette énergie par des méthodes telles que Boykov et al. (2001) incite à donner la même étiquette à deux superpixels voisins spatialement et ayant une faible distance d entre leurs descripteurs F.

8.3 Validation de SuperPatchMatch sur l'étiquetage de visages

La segmentation et l'étiquetage de visages sont des applications difficiles, en raison de plusieurs problèmes tels que la diversité de styles de coiffure, de couleurs de peau, les occultations partielles ou encore la présence de plusieurs visages sur une même image. Nous proposons d'évaluer notre approche SPM pour l'étiquetage de visages sur la base de données LFW (pour Labeled Faces in the Wild, Huang et al. (2007b)). Cette base contient 2927 images de taille 250×250 pixels, recalées linéairement (Huang et al., 2007a), et chaque image est fournie avec une décomposition en environ 225 à 250 superpixels.

LFW est une base de référence, utilisée pour valider les nouvelles méthodes basées sur les superpixels puisqu'elle contient des vérités terrains d'étiquetage (cheveux, visage et fond) associées à des décompositions en superpixels (Kae et al., 2013) fournies pour chaque image. Ainsi, contrairement aux autres bases standards, les comparaisons avec les méthodes de l'état-de-l'art ne sont pas biaisées par les décompositions en superpixels, qui impactent fortement la précision de segmentation.

8.3.1 Réglage des paramètres

SPM a été implémenté avec MATLAB via du code C-MEX multi-processus. Pour comparer la méthode à celles ayant des résultats de référence sur la base LFW (Kae et al., 2013; Liu et al., 2015a), nous utilisons les mêmes 1 500 images d'exemples dans notre bibliothèque T, et les mêmes 927 images de test comme images à étiqueter. Néanmoins, nous pouvons utiliser toutes les images dans une procédure de validation croisée de type *leave-one-out*, puisque SPM ne nécessite aucune étape d'apprentissage. En ajoutant de nouveaux sujets à la base, nous montrons dans la suite que nous améliorons la précision de l'étiquetage, sans augmenter le temps de calcul.

Le nombre d'itérations de SPM est fixé à 5, comme dans Barnes et al. (2009). Nous utilisons une norme ℓ_2 entre les histogrammes de gradients d'intensité orientés (HoG) (Dalal et Triggs, 2005) comme distance d dans (7.1). Dans l'Éq. (7.2), puisque les images sont de taille $h \times w = 250 \times 250$ pixels, et décomposées en approximativement $K = 250$ superpixels, σ_1 est fixé à 8. Dans (8.2), les paramètres α et β sont fixés empiriquement à 2 et 4. Dans le modèle de régularisation (8.4), nous fixons γ à 0.5 et utilisons l'algorithme α-*expansion* (Boykov et al., 2001) pour minimiser la fonctionnelle. Enfin, les temps de calcul dans la Figure 8.1 et le Tableau 8.1 incluent la recherche des k-ANN, la fusion d'étiquettes, la régularisation et l'étiquetage final.

8.3.2 Influence des paramètres

Influence de la taille du superpatch

Nous analysons tout d'abord l'influence de la taille du superpatch et du nombre d'ANN, c'est-à-dire de k, le nombre de SPM effectués en parallèle. La Figure 8.1 représente la précision de l'étiquetage à l'échelle superpixellique et le temps de calcul. La précision d'étiquetage est améliorée avec notre structure de superpatch et les meilleurs résultats sont obtenus avec R=50 pixels (95.08% avec k=50 ANN). Cette taille correspond en moyenne à la capture de trois anneaux de superpixels voisins, puisque les superpixels sont en moyenne de taille 16×16 pixels.

La Figure 8.1 représente également les courbes ROC correspondantes, obtenues avec k=50 ANN pour les trois étiquettes (visage, fond, cheveux). Sans la structure de superpatch, c'est-à-dire en calculant seulement la distance de correspondance sur le superpixel central (R=0 pixels), comme dans Gould et al. (2014), de moins bons

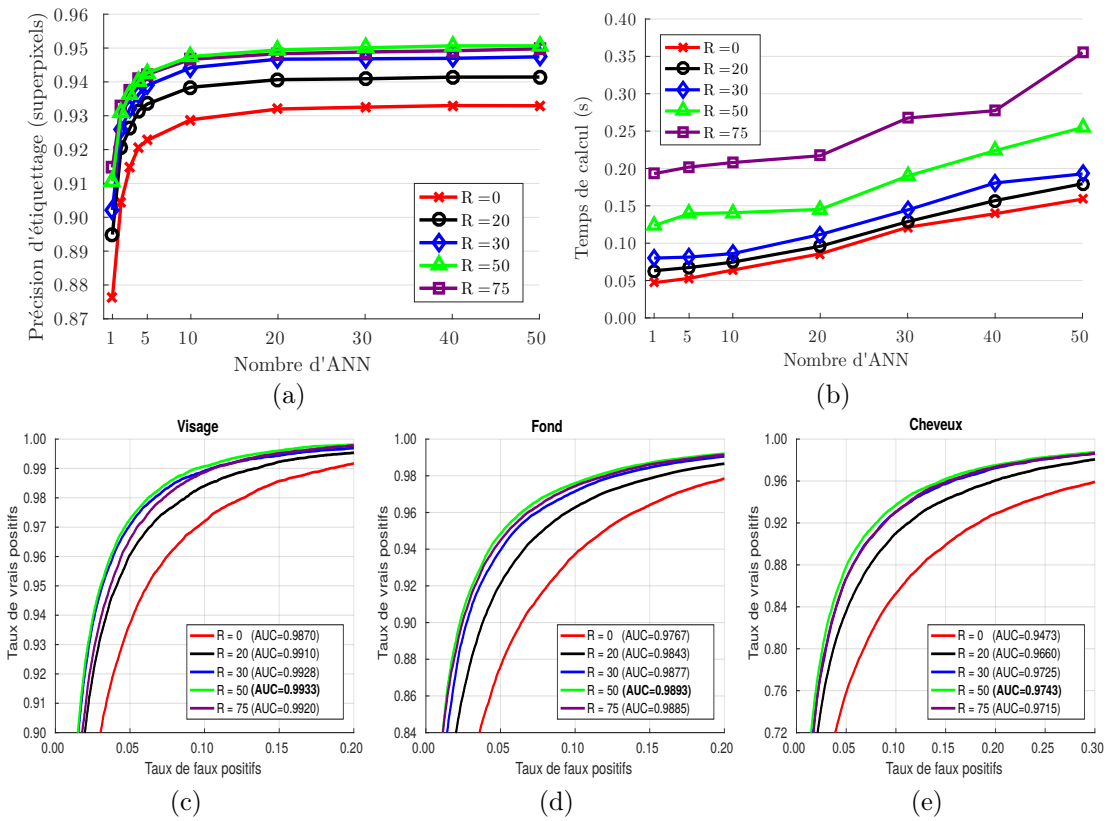


FIGURE 8.1 – Influence de la taille du superpatch sur la précision d'étiquetage (a) et temps de calcul correspondant (b). Les courbes ROC avec aires sous la courbe (AUC) pour différentes tailles de superpatch et k=50 sont données pour les trois classes visage (c), fond (d) et cheveux (e). $R = 50$ pixels donne les meilleurs résultats pour toutes les étiquettes.

ANN sont trouvés, ce qui décroît la précision d'étiquetage (93.29% avec k=50 ANN). La taille du superpatch doit être assez large pour capturer l'information contenue dans le voisinage des superpixels. Cependant, avec des superpatchs trop grands, c'est-à-dire $R > 50$, trop de superpixels voisins contribuent, menant à des ANN moins pertinents, à un étiquetage moins précis et à des temps de calcul plus élevés. Notons que l'on propose dans (8.2) une amélioration de l'étape de fusion d'étiquettes pour prendre en considération le recalage de la base, mais nous obtenons des résultats comparables (95.00% au lieu de 95.08%) sans a priori de position, c'est-à-dire β=∞ dans (8.2).

La Figure 8.2 illustre le processus de régularisation. Les probabilités d'étiquetage (8.1) obtenues par SPM sont représentées pour chaque étiquette (Figure 8.2(c)). La régularisation spatiale (8.4) donne des résultats plus cohérents (Figure 8.2(e)) qu'en assignant directement l'étiquette avec la plus forte probabilité (Figure 8.2(d)).

La Figure 8.3 montre enfin l'influence du superpatch sur différents exemples. Les erreurs d'étiquetage sont principalement dues à la similarité entre la couleur de cheveux et le fond, et à l'imprécision des segmentations en superpixels.

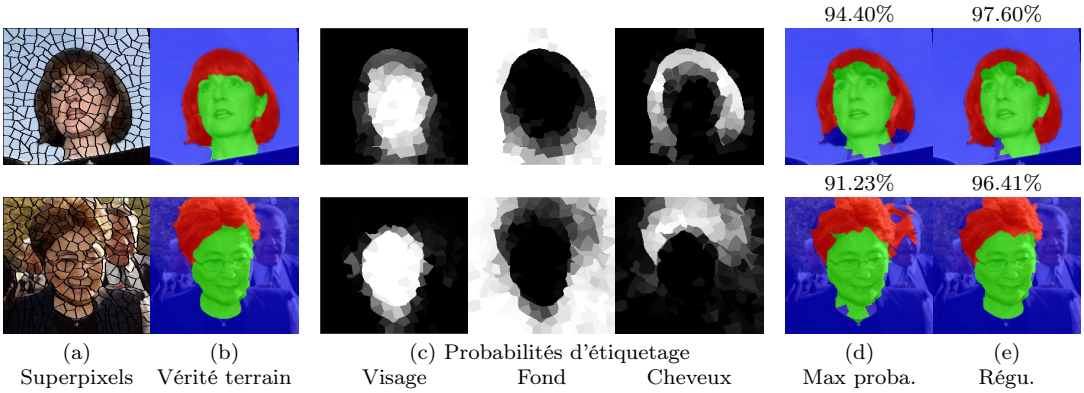


FIGURE 8.2 – Résultats d'étiquetage. (a) Décomposition en superpixels. (b) Vérité terrain associée. (c) Carte de fusion d'étiquettes pour les 3 classes (visage, fond, cheveux). (d) Étiquetage avec la plus haute probabilité (8.3). (e) Étiquetage après régularisation des cartes.

Influence de la propagation

Dans cette section, nous mesurons l'impact de l'étape de propagation proposée dans la section 7.4.2. Dans le cas des superpixels, l'irrégularité entre les décompositions des deux images considérées empêche de respecter strictement la même symétrie lors du décalage de la correspondance. Nous avons proposé dans (7.4), de prendre comme candidat, le superpixel dans l'image T_j ayant l'orientation la plus similaire à celle des superpixels dans l'image A, en considérant les angles entre les barycentres des superpixels. Avec cette méthode nous obtenons une précision d'étiquetage de 95.08%. En revanche, en considérant dans (7.4), un voisin aléatoire plutôt que celui avec l'orientation la plus similaire, nous obtenons une précision d'étiquetage de seulement 94.41%.

Dans la Figure 8.4, nous comparons ces deux configurations sur un exemple, en représentant les indices j des images T_j auxquelles appartiennent les correspondances. Plusieurs itérations sont considérées : l'étape d'initialisation (#0), la première (#1) et la dernière (#5). La méthode proposée de sélection du candidat durant l'étape de propagation (7.4), permet à l'algorithme de converger, comme en témoignent les larges zones d'indices similaires obtenues en sortie de l'algorithme. Autrement dit, SPM a pu efficacement propager les correspondances d'une même image d'exemple de la bibliothèque, en respectant les positions relatives entre les superpixels adjacents dans A et les superpixels adjacents dans T.

Influence de la régularité et validation de SCALP

À travers les chapitres 6 et 7, nous avons montré l'importance de la régularité pour limiter la variabilité entre les décompositions en superpixels, et ainsi utiliser plus efficacement l'information de leur voisinage. Cet aspect est notamment illustré dans la Figure 6.16, où le déplacement des correspondances calculées entre des

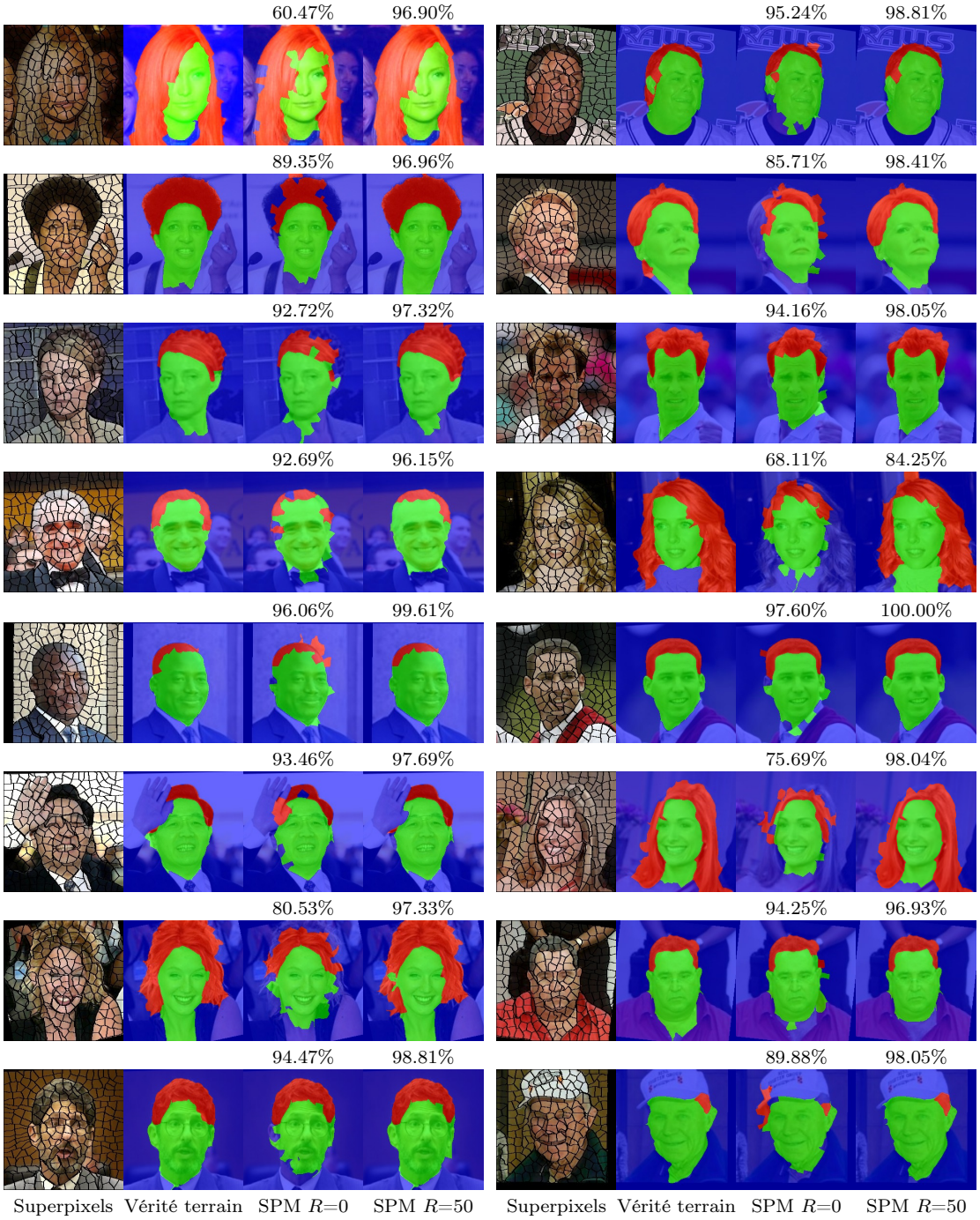


FIGURE 8.3 – Exemples d'étiquetages avec précision à l'échelle des superpixels obtenus avec SPM pour $R = 0$ et $R = 50$ pixels, c'est-à-dire, en utilisant respectivement les superpixels et les superpatchs.

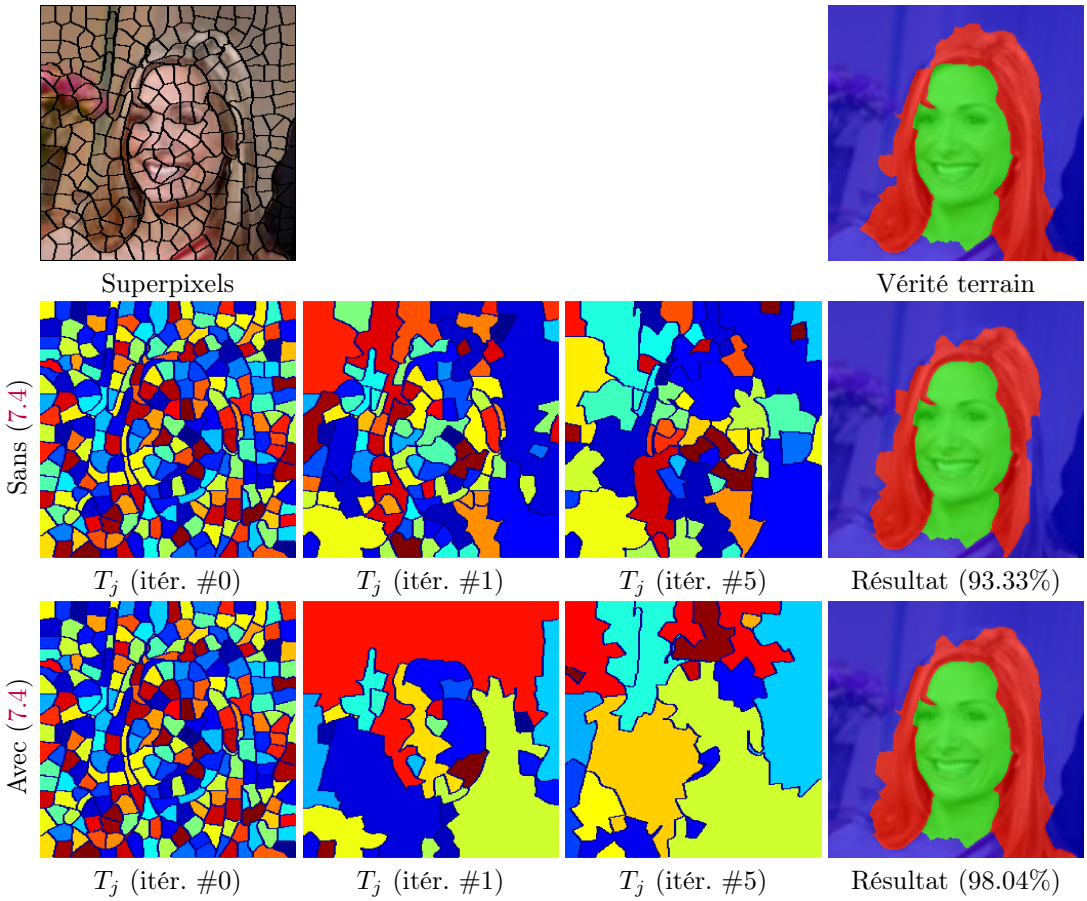


FIGURE 8.4 – Impact de la méthode de propagation proposée (7.4) sur la convergence et les performances de SPM. Les indices T_j correspondant au numéro de l'image dans la bibliothèque contenant l'ANN du superpixel sont représentés après l'initialisation (itér. #0), la première (itér. #1) et la dernière itération (itér. #5). Sans (7.4), en sélectionnant un voisin adjacent aléatoire durant la propagation, les candidats testés par SPM sont moins pertinents, les bonnes correspondances sont moins propagées et l'étiquetage moins précis. Avec la méthode proposée, les cartes sont plus lisses. SPM converge donc plus rapidement en propageant efficacement les correspondances dans une même image de la bibliothèque.

décompositions de la même image avec plusieurs méthodes est plus faible pour les décompositions régulières que celles irrégulières.

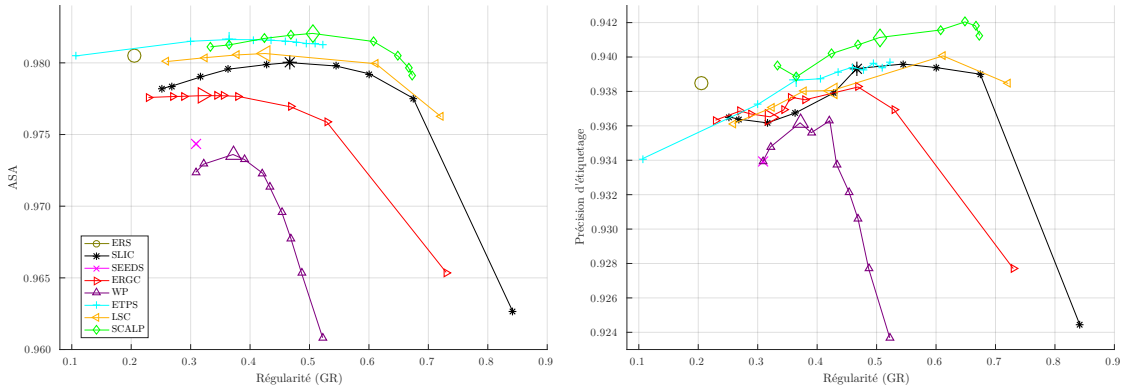
Sur l'application de recherche d'ANN basés sur les superpatchs présentée dans cette section, les décompositions des images sont déjà fournies par la base LFW. Pour mesurer l'impact de la régularité sur la précision de l'algorithme de recherche, nous proposons de décomposer pour plusieurs niveaux de régularité les images de la base de test par SCALP et par les méthodes superpixels de l'état-de-l'art considérées dans la partie II. Nous reportons alors dans la Figure 8.5 deux évaluations : celle de l'ASA (4.6), mesurant la superposition des superpixels générés avec la segmentation de la vérité terrain (Figure 8.5(a)), et celle de la précision de segmentation en trois

classes (visage, cheveux, fond) à l'échelle pixellique (Figure 8.5(b)). La régularité est évaluée avec GR (6.9).

La Figure 8.5 traduit deux résultats. Tout d'abord, les meilleures performances de segmentation sont obtenues avec SCALP. Cela démontre l'intérêt de la méthode de décomposition en superpixels proposée, qui permet par sa précision d'atteindre de meilleures performances de segmentation à partir d'exemples. Ensuite, pour toutes les méthodes comparées, les meilleures performances de segmentation sont obtenues pour des niveaux de régularité plus élevés que ceux fournissant la meilleure superposition avec la vérité terrain, évaluée avec l'ASA (Figure 8.5(a)). Autrement dit, il est bien préférable d'avoir une décomposition légèrement plus régulière, au détriment de l'adhérence aux contours, pour permettre une recherche de correspondances plus pertinente. Les larges marqueurs correspondent à l'ASA optimale pour chaque méthode, et sont reportés dans la Figure 8.5(b), pour souligner ce léger décalage, qui montre l'intérêt de la régularité. Des performances très proches sont néanmoins obtenues pour des niveaux de régularité plus faibles, montrant que les superpatchs et la méthode SPM s'adaptent également aux décompositions irrégulières. Notons également que les performances de segmentation sont légèrement inférieures à celles reportées dans la Figure 8.1 pour les paramètres par défaut, puisque nous n'utilisons plus les décompositions en superpixels correspondant à la vérité terrain. Néanmoins, les superpixels de SCALP permettent d'atteindre une précision de segmentation de 94.20%, surpassant celles de méthodes de l'état-de-l'art telles que CRBM (Kae et al., 2013) qui obtient 94.10%.

8.3.3 Comparaison avec les méthodes de l'état-de-l'art

SPM est comparé aux méthodes de référence appliquées à la base LFW dans le Tableau 8.1. Dans Kae et al. (2013), la méthode GLOC (pour GLObal and LOCal) utilise une machine de Boltzmann contrainte RBM (pour Restricted Boltzmann Machine) comme complément d'un champ aléatoire conditionnel, noté CRF (pour Conditional Random Field, Lafferty et al. (2001)). La combinaison de ces approches permet de réduire les erreurs d'étiquetage obtenues avec des modèles simples qui ne considèrent pas d'a priori de forme globale, aux dépens d'un temps de calcul plus élevé. Dans Liu et al. (2015a), une architecture basée sur un réseau de neurones convolutionnel profond DCNN (pour Deep Convolutional Neural Network) est proposée. Pour toutes les méthodes comparées, une étape d'apprentissage, qui peut durer plusieurs heures, est nécessaire afin d'entraîner les modèles. De plus, elles considèrent des a priori d'information sémantique dans leur processus d'étiquetage, par exemple, l'étiquette cheveux doit se trouver principalement au dessus de l'étiquette visage. Enfin, nous fournissons également les résultats obtenus avec l'algorithme PatchMatch à l'échelle pixellique, utilisé avec les mêmes paramètres et la même méthode de fusion d'étiquettes que celle présentée dans la section 8.2. Les correspondances sont alors calculées par une somme des différences au carré sur les intensités couleurs dans l'espace RGB de patchs de taille 9×9 pixels.



(a) (b)

FIGURE 8.5 – Comparaison des performances de SCALP et des méthodes superpixels de l'état-de-l'art obtenues sur les images de la base LFW. Les images à segmenter sont décomposées pour plusieurs niveaux de régularité, évalués avec GR (6.9), et les décompositions sont utilisées par SPM pour la recherche d'ANN et la fusion d'étiquettes. Les mesures d'ASA (4.6) et de précision de segmentation à l'échelle pixellique par SPM sont respectivement données en (a) et (b). SCALP offre les meilleures performances, et le résultat de segmentation optimal est obtenu pour une régularité plus élevée que celle générant la meilleure adhérence aux contours, démontrant l'importance de la régularité pour la recherche de correspondances dans notre méthode.

Afin de comparer notre méthode aux autres méthodes de l'état-de-l'art, nous fournissons dans le Tableau 8.1 les résultats d'étiquetage à l'échelle des superpixels et des pixels. La précision pixellique peut en effet être différente puisque la précision superpixellique ne considère pas la taille des superpixels. Les résultats sont ceux publiés par les auteurs, par conséquent, les valeurs de certaines métriques ne peuvent être reportées. La précision d'étiquetage superpixellique de SPM surpasse celles des méthodes comparées (95.08%). La précision pixellique (95.50%) surpasse également le résultat donné pour l'architecture de réseau de neurones DCNN (Liu et al., 2015a), qui a été optimisée pour être appliquée à la base LFW. L'augmentation de la précision pour l'échelle pixellique par rapport à l'échelle superpixellique montre que notre méthode échoue principalement à étiqueter des superpixels petits et étirés.

Le temps de calcul est un point de comparaison important. Les résultats de SPM donnés dans le Tableau 8.1 sont obtenus en $0.255s$ de traitement par image, et des temps similaires sont reportés pour la méthode GLOC (Kae et al., 2013). De plus, SPM surpasse les méthodes de état-de-l'art comparées en termes de précision d'étiquetage sans aucune étape d'entraînement. Contrairement aux autres méthodes nécessitant un ré-apprentissage, SPM permet donc d'intégrer directement de nouvelles images dans la bibliothèque d'exemples.

Pour illustrer ce point, nous étiquetons les images de la base avec une validation croisée de type *leave-one-out*. Autrement dit, pour chaque image à étiqueter,

nous ajoutons les 926 images restantes à la bibliothèque. Ainsi, SPM atteint une précision superpixellique de 95.26%. Ce résultat souligne l'impact de la diversité des images d'exemples qui peut améliorer la qualité des correspondances. De plus, ce résultat est obtenu sans augmentation du temps de calcul, puisque la complexité de notre algorithme ne dépend que de la taille de l'image traitée. Ainsi, SPM intègre facilement de nouvelles images dans la base, et est capable de fournir des résultats compétitifs en un temps de calcul limité sans aucun a priori de forme globale ou de modèle particulier.

Tableau 8.1 – Précision d'étiquetage sur la base de test LFW. Les résultats pour SPM sont obtenus avec k=50 ANN, et R=50 pixels. Les valeurs présentées sont les résultats publiés par les auteurs, ainsi, toutes les métriques d'évaluation ne peuvent être reportées.

Méthode	Précision à l'échelle des superpixels	Précision à l'échelle des pixels
PatchMatch	87.73%	87.02%
CRF spatial (Kae et al., 2013)	93.95%	*non reporté*
CRBM (Kae et al., 2013)	94.10%	*non reporté*
GLOC (Kae et al., 2013)	94.95%	*non reporté*
DCNN (Liu et al., 2015a)	*non reporté*	95.24%
SuperPatchMatch	**95.08**%	**95.43**%

8.4 Conclusion

Dans ce chapitre, nous avons démontré l'intérêt et le potentiel de la méthode SuperPatchMatch sur une application de segmentation et d'étiquetage. Les décompositions en superpixels permettent de grouper a priori les pixels, limitant ainsi le nombre d'éléments à traiter et le bruit contenu dans de plus petits descripteurs. Cependant, les superpixels seuls ne permettent pas de capturer le contexte local et donc de trouver de bonnes correspondances dans les bases d'images.

Utilisé avec la nouvelle structure de superpatch, SuperPatchMatch obtient alors des performances du niveau de l'état-de-l'art, et surpasse largement les résultats obtenus en utilisant uniquement des descripteurs basés sur les superpixels ou des patchs réguliers. Les résultats sont obtenus en un faible temps de calcul, et pour l'étiquetage de visages, les performances surpassent ceux de méthodes de l'état-de-l'art qui nécessitent de longues phases d'apprentissage pouvant durer jusqu'à plusieurs heures. Dans le chapitre suivant, SuperPatchMatch est appliqué au transfert de couleurs entre deux images.

Chapitre 9

Application au transfert de couleurs

Sommaire

9.1 Introduction 180
9.2 SuperPatchMatch contraint 182
9.2.1 Contrainte sur la diversité des correspondances 182
9.2.2 Problème d'assignement optimal 184
9.3 Fusion des couleurs 184
9.4 Résultats de transfert de couleurs 186
9.4.1 Paramètres 186
9.4.2 Influence de la diversité des ANN 187
9.4.3 Comparaison avec l'état-de-l'art 188
9.5 Conclusion 188

9.1 Introduction

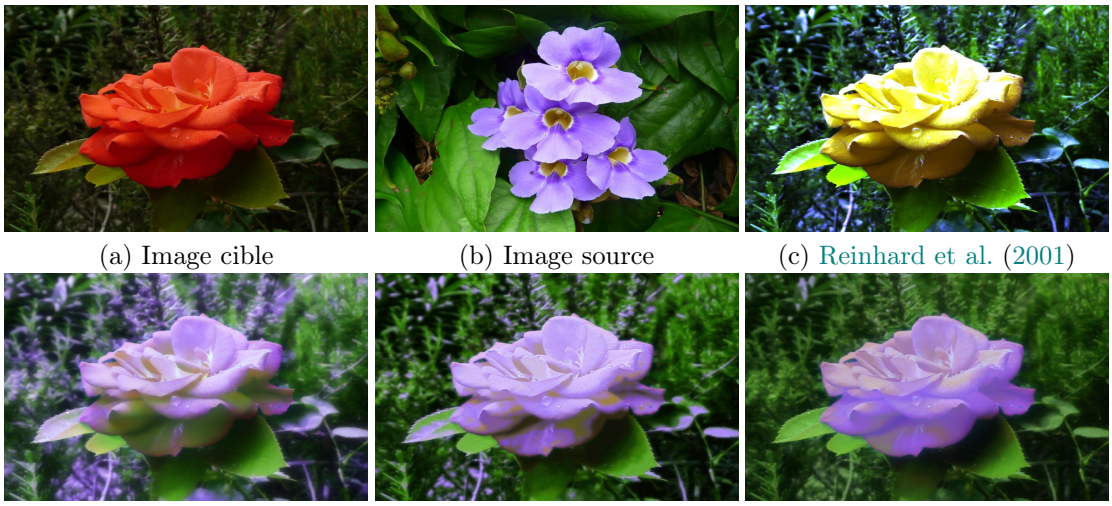
Les superpixels permettant de diminuer la résolution et donc de synthétiser l'information, notamment couleur d'une image, la méthode SuperPatchMatch peut être intéressante pour le transfert de couleurs rapide entre images. Cette application consiste à modifier la distribution des couleurs d'une image cible en utilisant une ou plusieurs images sources. Le résultat doit être cohérent avec la structure de l'image cible, et le temps de calcul limité pour permettre le traitement de grandes images ou de séquences vidéo. Depuis son introduction avec notamment Reinhard et al. (2001), beaucoup d'approches ont été proposées pour transférer les couleurs d'une image dans différents espaces (Pitié et al., 2005a; Xiao et Ma, 2006; Papadakis et al., 2011; Nguyen et al., 2014).

Les méthodes paramétriques transfèrent directement les statistiques couleurs entre les images. Par exemple, la méthode de Reinhard et al. (2001), transforme l'image cible de sorte qu'elle ait la même moyenne et variance couleur que l'image source. La méthode de Tai et al. (2005) utilise un algorithme de type espérance-maximisation (EM) pour estimer un modèle de mélanges de gaussiennes dans l'espace couleur et spatial, l'information de position aidant à préserver la structure de l'image. Ces méthodes paramétriques peuvent s'avérer rapides, mais le transfert direct des statistiques couleurs ne garantit pas la préservation des couleurs, comme le montre l'exemple de transfert dans la Figure 9.1(c).

Les outils de transport optimal sont également beaucoup utilisés pour transférer entièrement les distributions de couleurs (Pitié et al., 2007; Rabin et Peyré, 2011; Frigo et al., 2014). En plus des temps de calcul souvent conséquents avec ces approches, le transfert exact d'une palette de couleurs peut entraîner des aberrations visuelles lorsque les distributions et le contenu des images s'avèrent trop différents. De plus, en ne considérant que l'espace colorimétrique, des couleurs incohérentes peuvent être attribuées aux pixels voisins. Dans la Figure 9.1(d), la méthode de (Pitié et al., 2007) a transféré trop rigoureusement les couleurs de l'image cible, et le résultat n'est alors pas satisfaisant.

Des modèles de transport optimal relaxé, n'associant pas rigoureusement les distributions ont alors été proposés (Ferradans et al., 2014), et appliqués au transfert de couleurs à l'échelle des superpixels pour réduire le nombre d'éléments considérés, et donc le temps de calcul (Rabin et al., 2014). Dans la Figure 9.1(e), on constate que cette dernière approche permet un transfert de couleurs plus pertinent que l'approche de Pitié et al. (2007), mais toutefois perfectible. De plus, cette méthode présente toujours un coût de calcul important, même appliquée à l'échelle de superpixels.

Pour le transfert de couleurs, les superpixels sont particulièrement intéressants, puisqu'ils permettent de capturer des régions de couleur constante, résumant ainsi l'information couleur. Les approches basées sur les superpixels permettent ainsi une bonne adaptation au contenu de l'image mais nécessitent toujours un coût de calcul important dans le cas du transfert de couleurs par correspondance de graphes (Rabin et al., 2014; Liu et al., 2016a).



(a) Image cible (b) Image source (c) Reinhard et al. (2001)

(d) Pitié et al. (2007) (e) Rabin et al. (2014) (f) SCT

FIGURE 9.1 – Exemple de transfert de couleurs entre une image cible (a) et source (b) par une approche paramétrique (c), de transport optimal (d) et de transport optimal relaxé (e), sur un exemple illustrant les limites de chaque méthode. Ces résultats sont comparés à celui de SCT qui fournit un transfert de couleurs pertinent (f).

La méthode SuperPatchMatch proposée dans le chapitre 7 permet d'obtenir des correspondances de superpixels entre les images considérées en un temps de calcul très réduit, mais nécessite quelques adaptations pour être appliquée efficacement au transfert de couleurs. En effet dans ce contexte, bien que l'étape de propagation favorise la correspondance de larges zones, l'algorithme est libre de sélectionner la même correspondance pour plusieurs éléments de l'image cible. Pour le transfert de couleurs, ceci est un problème puisque l'on souhaite transférer globalement les couleurs de l'image source et pas uniquement quelques couleurs, au risque que celles-ci soient trop proches de celles de l'image cible. Ce problème de sélection répétée des mêmes éléments a également été montré dans Gutierrez et al. (2017) dans le cadre de la synthèse de texture à base de patchs.

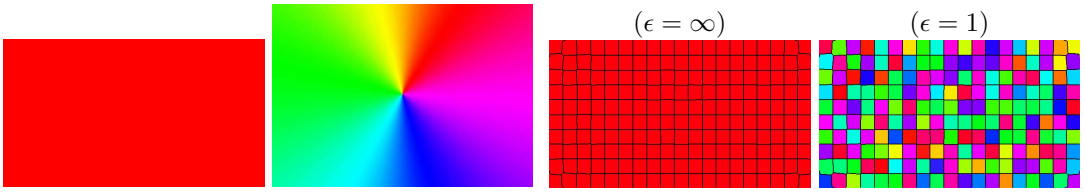
Contributions

Dans ce chapitre, nous proposons d'adapter SuperPatchMatch, pour définir une nouvelle méthode rapide de transfert de couleurs basée sur les superpixels nommée SCT (pour Superpixel-based Color Transfer). Afin de capturer globalement la palette de couleurs de l'image source, nous proposons une technique pour contraindre le processus de recherche d'ANN au sein de SuperPatchMatch et ainsi limiter la sélection des mêmes superpixels pour augmenter la diversité des correspondances. Les couleurs sélectionnées sont transférées par une approche inspirée par la méthode de moyennes non locales (Buades et al., 2005). Enfin, nous montrons que SCT produit des résultats satisfaisants en un temps de calcul limité grâce à SuperPatchMatch. Une exemple de transfert avec SCT est donné dans la Figure 9.1(f).

9.2 SuperPatchMatch contraint

Dans le chapitre 7, nous avons présenté la méthode SuperPatchMatch qui étend l'algorithme PM (Barnes et al., 2009) à la recherche de multiples ANN de superpixels. Dans notre contexte, une seule correspondance dans l'image source pour chaque superpixel de l'image cible est suffisante pour le transfert de couleurs. L'algorithme SuperPatchMatch pourrait donc être utilisé avec $k = 1$, le nombre d'ANN sélectionné pour chaque superpixel. Cependant, avec la méthode SuperPatchMatch classique, aucune garantie n'est donnée sur la répartition des correspondances, et pour le transfert de couleurs, il est important de capturer globalement la palette de couleurs de l'image source. En effet, en pratique, la recherche d'ANN peut converger pour trouver les meilleures correspondances et faire pointer un nombre important de superpixels de l'image cible vers seulement quelques éléments de l'image source. Si celle-ci contient un ou plusieurs éléments qui correspondent aux couleurs de l'image cible A, la recherche d'ANN peut mener aux mêmes correspondances, et le résultat du transfert est alors très proche de A.

La Figure 9.2(c) illustre ce problème. L'image source (Figure 9.2(b)) contenant également des éléments de couleur rouge, tous les superpixels de l'image cible (Figure 9.2(a)) trouvent une correspondance de couleur rouge. Le transfert de couleurs est alors quasi nul. Nous introduisons donc dans la section suivante, une méthode de correspondance qui impose une contrainte sur la sélection des éléments dans l'image source pour ne plus sélectionner le ou les quelques superpixels les plus proches.



(a) Image cible A (b) Image source B (c) Correspondances ($\epsilon = \infty$) (d) Correspondances ($\epsilon = 1$)

FIGURE 9.2 – Illustration de la capture globale de la palette source avec la contrainte proposée. Sans contrainte (c), c'est-à-dire $\epsilon = \infty$, tous les superpixels cibles de (a) sont associés à un élément source rouge de (b). Avec contrainte (d), c'est-à-dire $\epsilon = 1$, on force la sélection d'ANN à être répartie globalement sur toute l'image source, en n'autorisant qu'une seule fois la sélection de chaque superpixel de l'image source.

9.2.1 Contrainte sur la diversité des correspondances

Afin d'augmenter la diversité des correspondances, et de capturer globalement la palette des couleurs sources, nous proposons une méthode pour contraindre directement le processus de recherche d'ANN. Un paramètre ϵ est fixé pour définir le nombre maximal de sélection du même superpixel source.

Pour pouvoir appliquer cette méthode, il est d'abord nécessaire de s'assurer que la décomposition est telle que $|A| \leq \epsilon|B|$, et que l'initialisation respecte la contrainte fixée par ϵ lorsque l'on assigne aléatoirement les correspondances. Dans la Figure 9.2(d), avec $\epsilon = 1$, les superpixels de l'image cible capturent à présent de façon globale la palette de couleurs de l'image source. Ensuite, durant le processus itératif, on souhaite que l'algorithme diminue progressivement la distance globale $\sum_i D(A_i, B_{(i)})$ de correspondance entre tous les superpixels A_i et leur ANN $B_{(i)}$, avec D la distance sur les descripteurs des superpixels. La contrainte dure fixée par ϵ ne doit donc pas empêcher le changement d'association d'un superpixel de A vers un superpixel de B, déjà sélectionné ϵ fois, si celle-ci peut permettre de diminuer la distance globale.

En pratique, lorsque l'on souhaite procéder à un changement de correspondance, par exemple lorsqu'un superpixel $A_i \in A$ trouve un meilleur ANN $B_k \in B$, il existe deux cas de figure.

— Si moins de ϵ superpixels dans A pointent déjà vers $B_k \in B$, alors A_i peut changer de correspondance pour être associé à B_k, et la contrainte est respectée.

— Sinon, A_i ne peut être associé à B_k sans enfreindre la contrainte. Il est donc nécessaire de supprimer l'association d'un autre superpixel $A_j \in A$ pointant déjà vers B_k pour permettre à A_i d'y être associé. Pour faire cela, le superpixel A_j pointant initialement vers B_k doit maintenant être associé à une nouvelle correspondance.

Pour résoudre ce cas de figure, tout en assurant la diminution de la distance globale, et le respect de la contrainte, nous proposons d'associer un des A_j, correspondant initialement à B_k, vers $B_{(i)}$, la correspondance actuelle de A_i. On réalise ainsi un échange des correspondances assurant le respect de la contrainte fixée par ϵ. Pour tous les superpixels A_j tels que $B_{(j)} = B_k$, le coût de l'échange est calculé comme suit :

$$C(A_i, A_j) = (D(A_i, B_{(j)}) - D(A_i, B_{(i)})) + (D(A_j, B_{(i)}) - D(A_j, B_{(j)})). \tag{9.1}$$

Si pour au moins un superpixel A_j, on a $C(A_i, A_j) < 0$, on procède aux assignations suivantes qui permettent de réduire la distance globale :

$$\underset{A_j}{\operatorname{argmin}}(C(A_i, A_j)) \rightarrow B_{(i)}, \tag{9.2}$$

$$A_i \rightarrow B_{(j)} = B_k. \tag{9.3}$$

La Figure 9.3 illustre cet échange de correspondances dans le cas $\epsilon = 2$. A_i souhaite être associé à B_k, mais deux superpixels A_j lui sont déjà associés. Le superpixel A_j minimisant (9.1) et tel que $C(A_i, A_j) < 0$, est alors associé à $B_{(i)}$ et permet à A_i d'être associé à B_k. Cette adaptation de SuperPatchMatch définit un algorithme rapide de recherche d'ANN de superpixels dont la diversité est assurée pour fournir une sélection globale des couleurs à transférer.

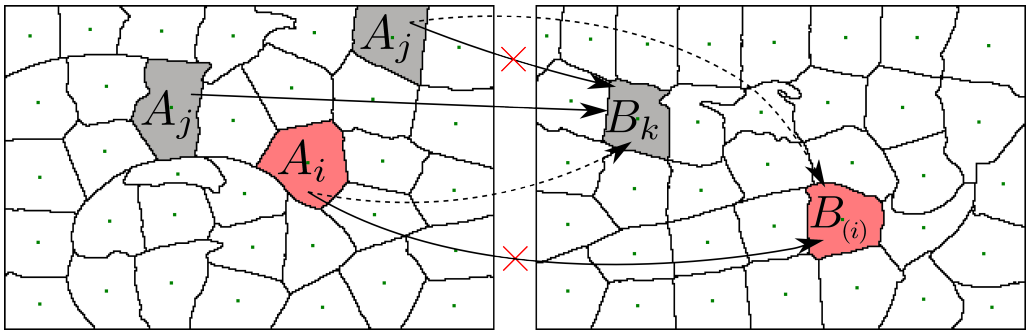


FIGURE 9.3 – Illustration de la contrainte de diversité appliquée à SuperPatchMatch. Dans cet exemple, un superpixel A_i trouve une meilleure correspondance B_k, à laquelle sont déjà associés deux superpixels A_j. La contrainte est fixée à $\epsilon = 2$. Par conséquent, pour associer A_i à B_k, un des superpixels A_j peut être envoyé vers $B_{(i)}$ pour assurer le respect de la contrainte, si cet échange permet de diminuer la distance globale de correspondance.

9.2.2 Problème d'assignement optimal

On peut remarquer que l'algorithme proposé de recherche d'ANN contraint par $\epsilon = 1$ approxime la résolution du problème d'assignement optimal. Ce problème est résolu de façon optimale par des approches telles que l'algorithme hongrois ou de Kuhn-Munkres (Munkres, 1957). Étant donné deux ensembles d'éléments $\{A_i\}_{i\in\{1,\ldots,|A|\}}$ et $\{B_j\}_{j\in\{1,\ldots,|B|\}}$ avec $|A| \leq |B|$, on cherche à caractériser une injection de A vers B qui minimise la distance globale. Autrement dit, le but est de trouver pour chaque élément A_i, une correspondance $B_{(i)}$, qui ne peut être sélectionnée qu'une seule fois, et qui minimise la distance globale entre les correspondances.

Nous considérons deux images proches issues de (Korman et Avidan, 2011). Dans la Figure 9.4, nous présentons la reconstruction de l'image cible $\widetilde{A}$ depuis les couleurs moyennes des superpixels sélectionnés dans l'image source, pour notre méthode et l'algorithme optimal (Munkres, 1957). Nous comparons également la distance totale entre les correspondances et le temps de calcul pour notre approche, l'algorithme de résolution optimale (Munkres, 1957), et des associations aléatoires, c'est-à-dire uniquement l'étape d'initialisation de notre méthode.

Notre méthode rapide de recherche d'ANN fournit des résultats proches de ceux de la résolution optimale, en étant significativement plus rapide. Par exemple pour des images décomposées en environ 1000 superpixels, les correspondances sont calculées en 1s par notre méthode quand l'algorithme optimal s'exécute en plus de 100s, alors que les reconstructions sont visuellement très proches.

9.3 Fusion des couleurs

Notre méthode fournit un ANN dans B pour chaque superpixel de A. La dernière étape consiste à transférer la couleur des correspondances pour calculer

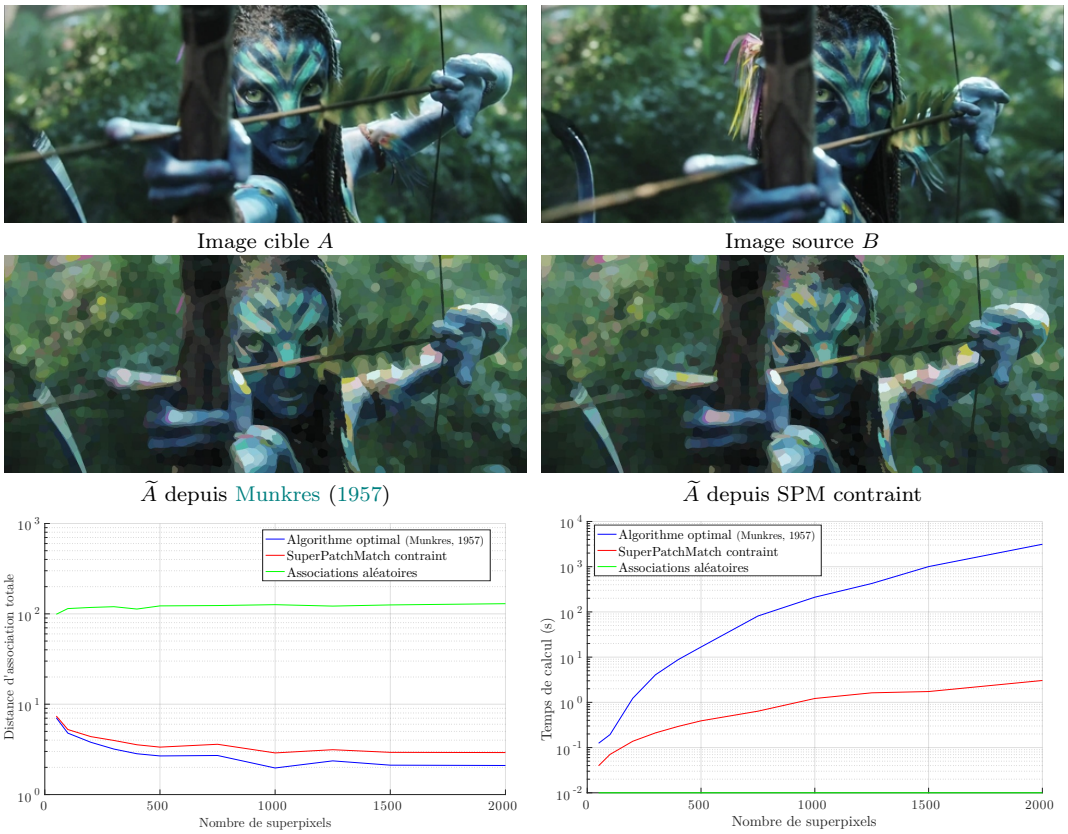


FIGURE 9.4 – Comparaison de l'algorithme d'assignement optimal (Munkres, 1957) et de notre méthode de correspondance, contrainte avec $\epsilon = 1$ sur des images de taille 1920×800 pixels issues de Korman et Avidan (2011). L'image cible $\widetilde{A}$ est reconstruite depuis les couleurs moyennes des superpixels sélectionnés dans la source ($K \approx 2000$ superpixels). Les performances en termes de distance totale et de temps de calcul sont également comparées pour plusieurs échelles.

l'image résultat du transfert de couleurs A_t, tout en préservant la structure de l'image cible A. Un superpixel dans A ayant une correspondance dans B, susceptible d'avoir une forme différente, il n'y a pas d'association unique pixel à pixel possible entre les deux éléments, et les couleurs ne peuvent pas être transférées directement à l'échelle pixellique. Les couleurs moyennes des ANN peuvent être transférées à leur correspondance dans A, mais cela produit un résultat constant par morceaux, comme dans la Figure 9.4.

Comme pour le transfert d'étiquettes dans les chapitres 2 et 8, nous proposons de considérer les couleurs moyennes des ANN de superpixels dans un schéma de fusion par moyennes non locales (Buades et al., 2005). Ainsi, les couleurs de l'image source sélectionnées contribuent au calcul de la couleur de chaque pixel dans A_t. On accroît ainsi le nombre de candidats couleur pour chaque pixel, ce qui permet de s'adapter à la structure de l'image cible. Un superpixel A_i est décrit par l'ensemble

des positions X_i et des couleurs C_i des pixels contenus $p \in A_i$, tel que $A_i = [X_i, C_i] = [(x_i/h, y_i/w), (r_i, g_i, b_i)/255]$, pour une image A de taille $h \times w$ pixels. Pour calculer la couleur $A_t(p)$ du pixel p, la fusion des couleurs des ANN est effectuée par une moyenne pondérée des correspondances sélectionnées :

$$A_t(p) = \frac{\sum_{A_j \in A} \omega(p, A_j) \bar{C}_{B_{(j)}}}{\sum_{A_j \in A} \omega(p, A_j)}, \tag{9.4}$$

avec $\bar{C}_{B_{(j)}}$, la couleur moyenne du superpixel $B_{(j)}$, la correspondance de A_j, et $\omega(p, A_j)$ un poids qui dépend de la distance entre le pixel considéré $p \in A_i$ et un superpixel $A_j \in A$. Afin de favoriser les contributions des superpixels similaires dans A, ce poids est calculé par une distance de Mahalanobis :

$$\omega(p, A_j) = \exp\left(-(p - \bar{A}_j)^T Q_i^{-1} (p - \bar{A}_j) + \sigma(p)\right), \tag{9.5}$$

où $\sigma(p)$ définit la dynamique de l'exponentielle et est fixé tel que $\sigma(p) = \min_j \left(-(p - \bar{A}_j)^T Q_i^{-1} (p - \bar{A}_j)\right)$, et Q_i inclut les covariances spatiales et colorimétriques des pixels dans A_i :

$$Q_i = Q(A_i) = \begin{pmatrix} \delta_s^2 Cov(X_i) & 0 \\ 0 & \delta_c^2 Cov(C_i) \end{pmatrix}. \tag{9.6}$$

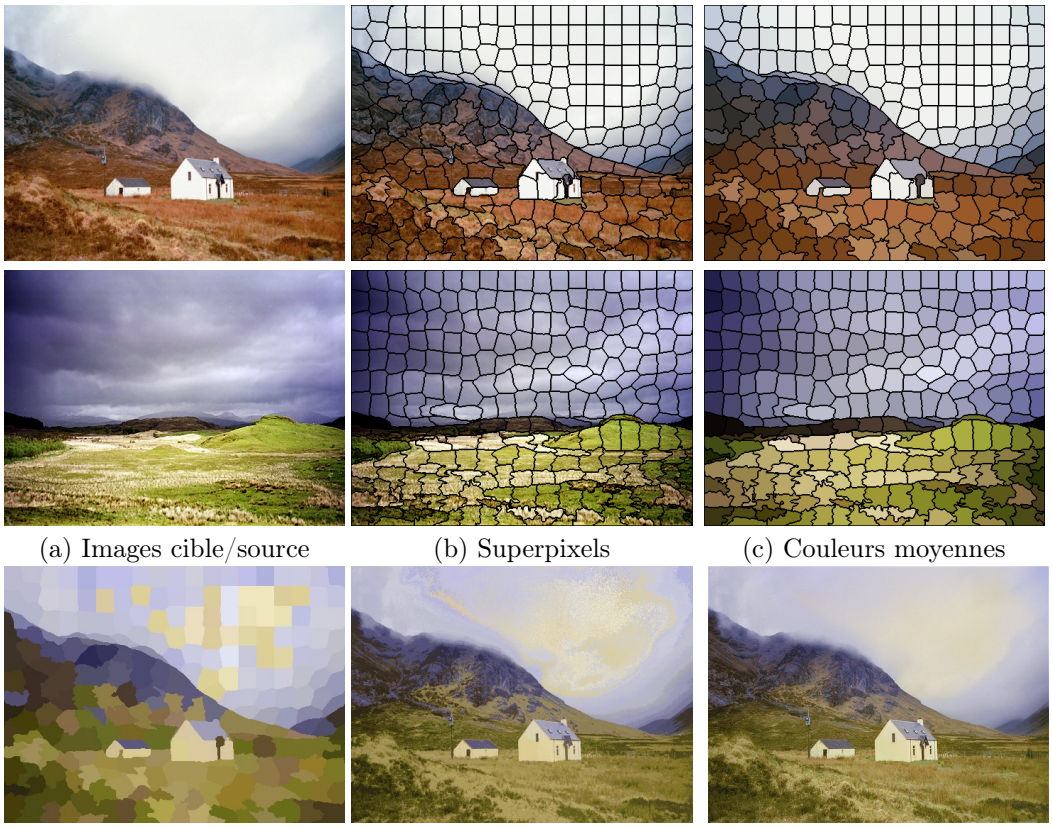
Ce poids permet de conserver la structure de l'image cible et le modèle de fusion (9.4) ne transfère que des couleurs présentes dans l'image source, limitant ainsi considérablement la création de nouvelles couleurs durant le transfert.

9.4 Résultats de transfert de couleurs

9.4.1 Paramètres

SCT est implémentée avec MATLAB via du code C-MEX. Les décompositions en superpixels sont calculées avec la méthode SCALP, de telle sorte que chaque superpixel contienne environ 500 pixels. L'utilisation de l'approche régulière SCALP permet d'obtenir des superpixels ayant approximativement la même taille, et qui peuvent donc être considérés équitablement dans le processus de fusion. Une décomposition qui générerait par exemple de très petits superpixels pourrait conduire au transfert d'une couleur qui ne soit pas significative visuellement dans la source.

La recherche d'ANN est effectuée sur les histogrammes cumulés de couleur des superpixels dans l'espace RGB. Dans l'Éq. (9.6), $\delta_s = 100\delta_c$ et $\delta_c = 0.1$ afin de favoriser la cohérence spatiale des couleurs fusionnées. Enfin, ϵ est fixé par défaut à 3, et les résultats sont post-traités par une étape de regrainage (Pitié et al., 2005b). Les différentes étapes de SCT sont illustrées dans la Figure 9.5. On représente la décomposition des images en superpixels, avec leur couleur moyenne, les couleurs des correspondances et les résultats de fusion des couleurs avec post-traitement.



(a) Images cible/source (b) Superpixels (c) Couleurs moyennes

(d) Correspondances (e) Fusion des couleurs (f) Fusion des couleurs +Pitié et al. (2005b)

FIGURE 9.5 – Étapes de la méthode proposée SCT. Représentation des décompositions des images (a) en superpixels (b) avec leur couleur moyenne (c), des correspondances affichées sur l'image cible (d), et des résultats de fusion des couleurs (e) avec post-traitement par une méthode de regrainage (Pitié et al., 2005b) (f).

Notre méthode produit un transfert de couleurs en un temps réduit grâce à l'utilisation des superpixels, c'est-à-dire moins de 1s pour des images de 480×360 pixels. Les décompositions sont calculées avec SCALP en moins de 0.4s, la recherche d'ANN est effectuée en approximativement 0.1s et la fusion de couleur nécessite 0.25s pour produire le transfert de couleurs.

9.4.2 Influence de la diversité des ANN

En plus de la Figure 9.2, on illustre sur des images naturelles dans la Figure 9.6 l'influence du paramètre ϵ qui limite la sélection des mêmes superpixels. Les cartes d'occurrence de sélection des superpixels dans l'image source sont représentées, et montrent que pour $\epsilon = \infty$, même une recherche rapide d'ANN avec SuperPatch-Match peut converger vers des correspondances exactes. En effet, dans ce cas de figure, seuls quelques superpixels de l'image source sont sélectionnés.

Par exemple, la plupart des superpixels de la fleur blanche dans l'image cible sont associés au plus blanc des superpixels de l'image source, ce qui produit un transfert de couleurs peu perceptible. Avec la méthode proposée, et la contrainte fixée par $\epsilon \neq \infty$, on capture efficacement et de manière globale la palette de couleurs de l'image source.

9.4.3 Comparaison avec l'état-de-l'art

Dans cette section, la méthode proposée SCT est comparée dans la Figure 9.7 à différentes méthodes de l'état-de-l'art basées sur le transport optimal (Pitié et al., 2007), le transfert d'histogrammes par méthode variationnelle (Papadakis et al., 2011), et une approche de transfert d'histogrammes par correspondance des gamut couleurs (Nguyen et al., 2014).

SCT produit des résultats plus satisfaisants ou équivalents aux autres méthodes. Notre approche réalise une sélection des couleurs sources, qui sont efficacement transférées à la cible avec respect de l'exposition et du grain initiaux, contrairement aux autres méthodes, qui par exemple utilisent les couleurs de l'image cible (Pitié et al., 2007; Papadakis et al., 2011) et peuvent entraîner la création de nouvelles couleurs (Nguyen et al., 2014).

Par exemple, sur le premier exemple, seule la méthode SCT réussit à transférer le bleu du ciel dans la mer. Sur le second exemple, les méthodes de Pitié et al. (2007); Papadakis et al. (2011) entraînent la disparition de structures, qui sont toujours visibles avec SCT. Sur le résultat de Papadakis et al. (2011) on constate par exemple la présence de couleur verte, qui n'apparaît pas dans l'image source. Sur le dernier exemple, seule la méthode SCT réussit à transférer la couleur orange des pierres, à l'herbe de la montagne, quand les autres méthodes modifient fortement l'exposition de la cible. Enfin, notre méthode SCT produit un transfert en moins de 1s. D'autres modèles tels que Papadakis et al. (2011) peuvent nécessiter des temps de calcul conséquents allant jusqu'à 120s.

9.5 Conclusion

Dans ce travail, nous avons proposé une nouvelle méthode de transfert de couleurs basé sur les superpixels. Nous adaptons SuperPatchMatch à la recherche globale de correspondances entre deux images. Nous avons introduit une méthode générale pour augmenter la diversité des ANN sélectionnés afin de capturer plus globalement la palette de couleurs de l'image source. Les couleurs sont efficacement transférées avec respect du grain et de l'exposition de l'image cible en un temps limité grâce aux superpixels. Les faibles temps de calcul permettent d'envisager l'application de la méthode au transfert de couleurs sur des vidéos, avec SuperPatchMatch appliqué à des supervoxels.

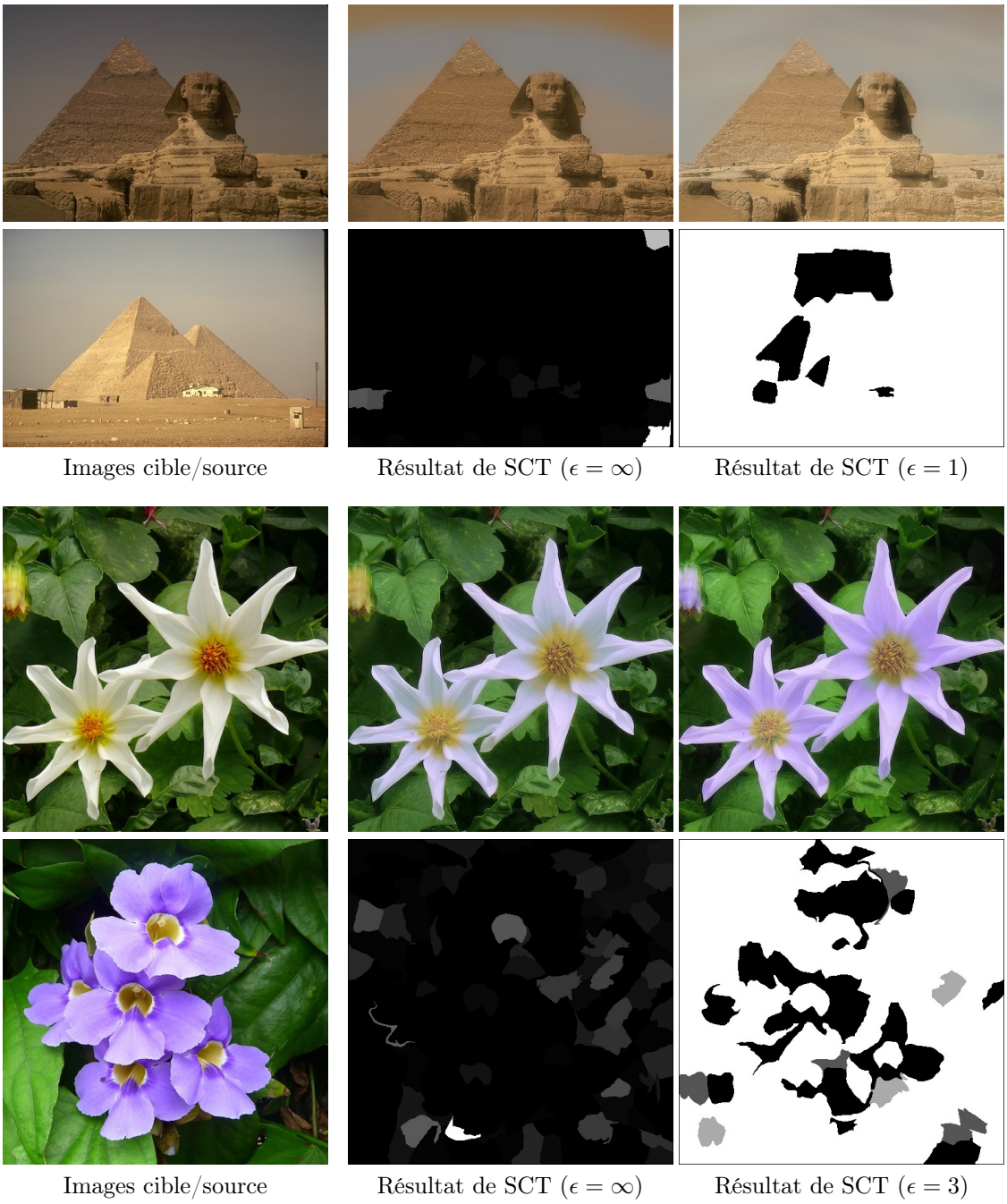


FIGURE 9.6 – Influence de la contrainte de diversité sur le transfert de couleurs avec SCT. Des exemples de résultats sont représentés pour différentes valeurs de ϵ, et comparés à ceux obtenus sans contrainte ($\epsilon = \infty$). Les cartes (ligne du bas) indiquent le nombre de sélections des superpixels dans l'image source (le noir représente zéro, et le blanc, le plus grand nombre de sélections). La contrainte permet de capturer globalement la palette de couleurs de la source.

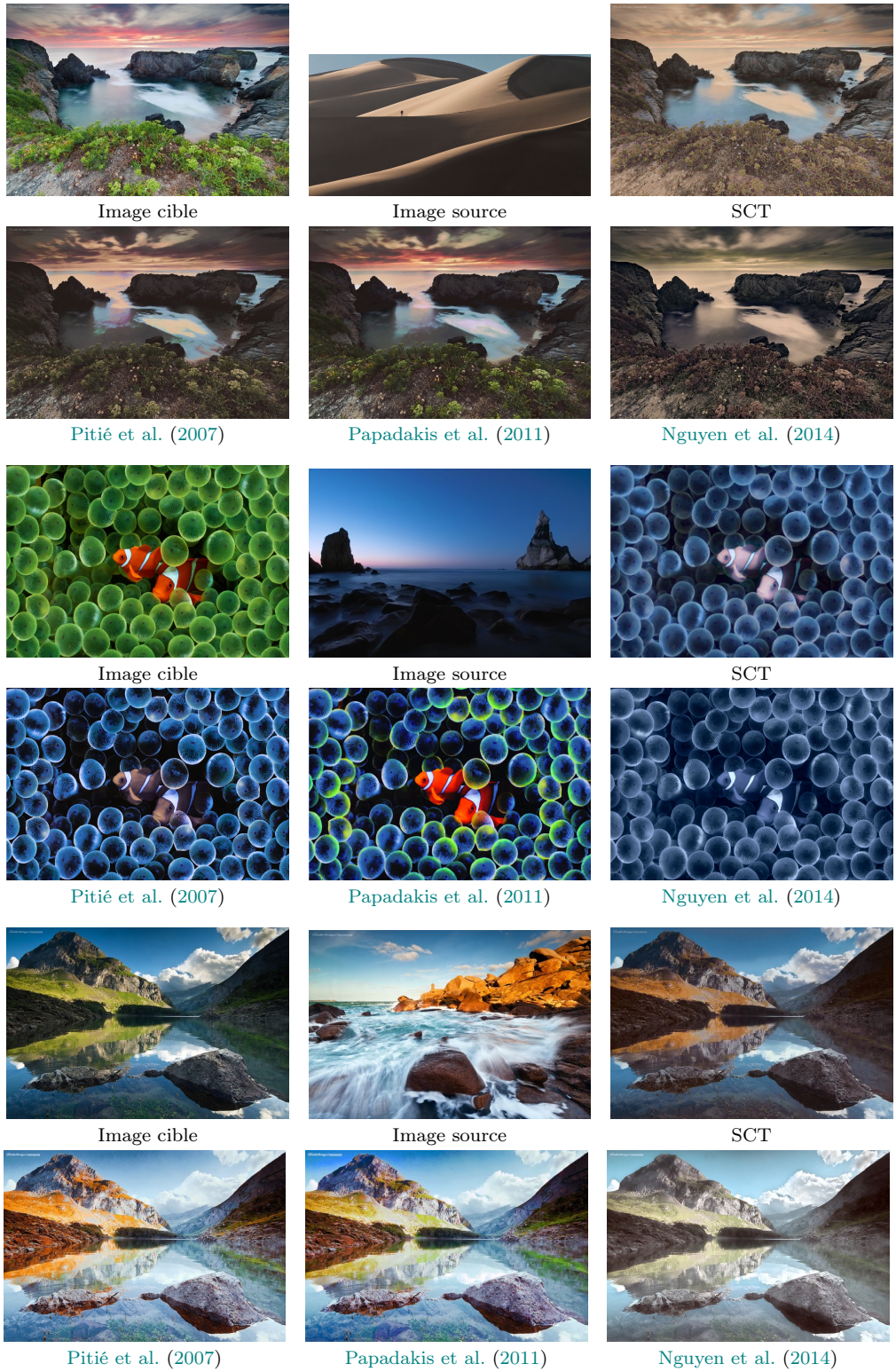


FIGURE 9.7 – Comparaison de SCT avec l'état-de-l'art. SCT produit des résultats plus satisfaisants ou équivalents à ceux des méthodes comparées.

Conclusion générale et perspectives

1 Conclusion générale

Dans cette thèse nous avons proposé plusieurs contributions dans le contexte du traitement et de l'analyse d'images par des approches non locales. Nous nous sommes intéressés aux différents niveaux de la chaîne de traitement, en proposant un travail sur les descripteurs de l'image avec une nouvelle méthode de décomposition en superpixels réguliers, et une nouvelle structure de voisinage de superpixels. Un travail a également été mené sur les méthodes rapides de recherche de correspondances de descripteurs similaires, notamment adaptées à de grandes bibliothèques d'images. Ces outils ont été étendus et comparés aux méthodes existantes sur des applications diverses telles que la segmentation d'images médicales 3D, l'étiquetage de superpixels ou le transfert de couleurs, et obtiennent des performances du niveau de l'état-de-l'art, avec de faibles temps de calcul.

Dans la partie I, nous nous sommes intéressés à la conception d'un algorithme rapide de recherche de correspondances de patchs dans une grande bibliothèque d'images. Les approches de l'état-de-l'art effectuant une recherche d'ANN dans de nombreuses images simultanément sont généralement hautement paramétrées et basées sur des prétraitements (Korman et Avidan, 2016) ou apprentissages (Barnes et al., 2015), limitant leur applicabilité. Dans le contexte de la segmentation d'images médicales 3D basée sur de grandes bibliothèques de modèles, ces contraintes peuvent affecter fortement le temps de calcul et réduire la précision de segmentation.

Nous avons proposé d'étendre l'algorithme de recherche d'ANN PatchMatch (Barnes et al., 2009) au sein d'une nouvelle méthode rapide de segmentation, nommée OPAL, présentant les propriétés suivantes :

— complexité liée uniquement à la taille de l'image à traiter,
— pas d'apprentissage nécessaire,
— réglage du nombre d'itérations possible,
— unique distance de comparaison entre les descripteurs à fixer.

Nous avons ensuite validé la méthode pour la segmentation d'images médicales de type IRM 3D, où nous obtenons des résultats en seulement quelques secondes de traitement, contrairement à plusieurs minutes voire heures, pour les méthodes de référence. En plus du gain substantiel en temps de calcul, la méthode obtient des performances supérieures à celles des méthodes de l'état-de-l'art, notamment celles basées sur la fusion exhaustive des patchs d'exemples des modèles de la bibliothèque. Pour les fines structures anatomiques étudiées que sont l'hippocampe et le cervelet, les contours des structures à segmenter ne suivent pas nécessairement les contours de l'image. Les patchs alors permettent de capturer efficacement le voisinage d'un voxel, et de transférer l'information de segmentation d'un patch similaire de la bibliothèque. Bien que la méthode ait ici été adaptée aux patchs pour la segmentation de ces structures, certains outils ont été repris pour adapter l'algorithme PatchMatch aux superpixels, et la méthode proposée dans la partie III présente les mêmes propriétés que celles de la méthode OPAL.

Dans la partie II, nous nous sommes intéressés à la décomposition d'une image en superpixels. Un des objectifs de cette thèse a été de proposer des outils permettant d'utiliser les superpixels efficacement dans un contexte non local. Ceux-ci doivent être réguliers pour limiter la variabilité des décompositions entre images, et permettre la recherche précise de correspondances. Dans l'état-de-l'art, cette régularité est souvent obtenue au détriment de la précision de segmentation.

Nous avons donc proposé une nouvelle méthode de segmentation en superpixels, nommée SCALP, qui utilise l'information de couleur et de contour sur le chemin linéaire au barycentre du superpixel durant le processus de décomposition, afin d'assurer le respect des contours des objets de l'image et la convexité des formes. La méthode obtient des résultats supérieurs à ceux des méthodes de l'état-de-l'art en termes de précision de segmentation, tout en générant des superpixels très réguliers.

Beaucoup de méthodes et de métriques de validation ont été introduites et varient selon les travaux. Afin d'analyser de façon pertinente la méthode SCALP, nous avons proposé une analyse du processus de validation, en étudiant la robustesse des métriques de l'état-de-l'art. Nous avons également proposé de nouveaux outils d'évaluation, notamment une métrique de régularité répondant aux limites de celles existantes. Nous avons démontré que la régularité est directement liée aux performances de plusieurs applications, et que celle de SCALP peut être réglée de sorte à générer d'excellents résultats pour un large panel de problèmes.

Dans la partie III, nous nous sommes intéressés à l'utilisation non locale des superpixels. Nous avons combiné les travaux des deux parties précédentes afin de proposer une méthode rapide et robuste de recherche de correspondances basées sur les superpixels dans une bibliothèque d'images. La méthode proposée étend l'algorithme PatchMatch aux superpixels, en associant des correspondances à un nombre réduit d'éléments, dont les frontières suivent les contours des objets.

Pour trouver des correspondances pertinentes, le contexte est nécessaire, et pour le cas pixellique, il est généralement obtenu avec les patchs, en considérant les pixels du voisinage. Néanmoins, le manque de régularité entre différentes décompositions en superpixels rend difficile l'utilisation de leur voisinage pour calculer ces correspondances. Nous avons donc proposé le superpatch, une nouvelle structure de voisinage de superpixels. Celle-ci intègre la position des superpixels voisins et une distance entre deux superpatchs peut être calculée en comparant leurs superpixels spatialement proches, généralisant les distances standards pour des patchs réguliers.

SuperPatchMatch a alors été validé sur la base de référence fournie avec des décompositions de visages en superpixels à étiqueter, et permettant une validation des performances indépendante de la méthode de segmentation en superpixels. SuperPatchMatch obtient des performances d'étiquetage supérieures à celles de méthodes basées sur un apprentissage et notamment des réseaux de neurones. Les performances obtenues en utilisant d'autres superpixels ont également été comparées, et SCALP fournit dans ce cas les décompositions menant aux meilleurs résultats d'étiquetage sur la base de visages. L'impact de la régularité, qui permet d'obtenir de meilleures correspondances, a alors à nouveau été mis en évidence.

Enfin, nous avons démontré l'intérêt de la méthode SuperPatchMatch et sa modularité en l'adaptant au transfert de couleurs entre deux images. Dans l'application précédente, SuperPatchMatch converge vers n'importe quel élément de la bibliothèque d'images. Dans cette application, il est nécessaire de transférer globalement les couleurs entre les deux images. Nous avons donc proposé de modifier la méthode pour influer directement sur la répartition des correspondances, et de forcer une distribution homogène des superpixels sélectionnés dans l'image d'exemple. Une nouvelle méthode de transfert de couleurs basée sur les superpixels, nommée SCT, a alors été proposée. Celle-ci obtient des résultats comparables à l'état-de-l'art en transférant efficacement les couleurs des superpixels d'une image à l'autre, en un temps de calcul très réduit.

2 Résumé des contributions

En conclusion, cette thèse s'inscrit dans le contexte du traitement et de l'analyse d'images par méthodes non locales en proposant à la communauté plusieurs outils :

- une méthode rapide de recherche de correspondances de patchs pour la segmentation de structures anatomiques sur des images médicales (OPAL), obtenant des performances du niveau de l'état-de-l'art, intégrée à la plateforme volBrain [1], et ayant fait l'objet d'un dépôt logiciel,
- une méthode de décomposition en superpixels (SCALP),
- une méthode de validation des méthodes de superpixels basée sur une nouvelle métrique d'évaluation de la régularité des superpixels (GR),
- une structure de voisinage de superpixels (superpatch),
- une méthode de recherche d'ANN de superpatchs (SuperPatchMatch), obtenant des performances d'étiquetage de superpixels du niveau de l'état-de-l'art,
- une extension de cette dernière méthode pour contraindre la répartition des ANN, appliquée au transfert de couleurs basé sur les superpixels (SCT).

Ces contributions visent à exploiter efficacement et simplement la redondance d'informations contenues dans de grandes bibliothèques d'images. Les méthodes proposées ne souffrent pas des contraintes liées à l'apprentissage et offrent de faibles temps de calcul favorisant leur adaptation à de nombreux domaines. Selon les applications, les superpixels peuvent être préférés aux pixels, et la structure de superpatch permet alors de les utiliser efficacement, de façon similaire aux patchs réguliers. L'algorithme de recherche rapide d'ANN (partie I), associé à la réduction du nombre d'éléments avec les superpixels (partie II), offrent alors des outils génériques et très rapides pour extraire de l'information de grandes bases d'images (partie III).

Les différents descripteurs, méthodes et applications proposés dans cette thèse obtiennent des performances similaires ou supérieures à celles de l'état-de-l'art, et offrent encore de nombreuses perspectives que nous décrivons dans la suite.

1. http://volbrain.upv.es/

3 Perspectives

Des perspectives de recherche directes pour étendre et envisager de nouvelles applications pour les méthodes OPAL, SCALP et SuperPatchMatch, sont tout d'abord présentées.

3.1 Segmentation d'images médicales (OPAL)

Compte tenu du faible temps de calcul offert par la méthode OPAL, celle-ci peut être appliquée à la segmentation de plus larges structures, ou même à la segmentation des structures du cerveau entier, comme dans (Fischl et al., 2002; Wu et al., 2014). Cette perspective est notamment donnée dans Manjón et Coupé (2016), qui prévoit d'étendre le pipeline volBrain à la segmentation complète du cerveau.

Dans le cas du cervelet, pour pallier la complexité des sous-structures et le faible nombre de modèles, les contributions à la segmentation des différentes tailles de patchs sont pondérées pour s'adapter à la structure considérée (voir section 3.3), et cette étape pourrait s'avérer également nécessaire dans le cas de la segmentation du cerveau. Les paramètres de pondération sont déterminés par une descente de gradient, quand la méthode originale OPAL est introduite sans aucune étape d'apprentissage, et les paramètres définis empiriquement, sont suffisants pour une bonne segmentation de l'hippocampe. Les différents estimateurs sont donc pondérées globalement dans (3.2) par un paramètre $\alpha_i^m \in [0,1]$ fixé selon le descripteur i pour la structure m.

Cet apprentissage de la pertinence de chaque descripteur pour chaque structure, pourrait être amélioré en pondérant localement les estimateurs. On désire donc en chaque voxel x, les paramètres $\alpha_i^m(x) \in [0,1]$ correspondant à la fiabilité de l'estimateur $i \in \{1,\dots,N\}$ pour la structure d'étiquette m et tels que $\sum_{i=1}^N \alpha_i^m(x) = 1$. Pour les obtenir, on peut chercher à résoudre le problème suivant :

$$\min_{\alpha_i^m(x)} \sum_{j=1}^{M} \left\| \sum_{i=1}^{N} \alpha_i^m(x) \mathcal{L}_{i,j}^m(x) - G_j^m(x) \right\|_2^2, \tag{9.7}$$

en considérant les cartes $\mathcal{L}_{i,j}^m$ et les segmentations binaires G_j^m de chaque image j.

Ce modèle peut être résolu par exemple par une descente de gradient projetée, mais nécessite cependant de distinguer la bibliothèque d'apprentissage des sujets à segmenter, pour ne pas introduire un biais avec une procédure de segmentation de type *leave-one-out* qui générerait les estimateurs $\mathcal{L}_{i,j}^m$ depuis les $M-1$ images restantes. Cette pondération locale des estimateurs multi-échelles et multi-descripteurs durant l'agrégation pourrait permettre d'améliorer la précision de segmentation.

3.2 Décomposition en superpixels (SCALP)

Comme mentionné dans le chapitre 5, la méthode proposée SCALP obtient des temps de calcul très raisonnables (inférieurs à 1s pour une image de taille 481×321 pixels), et comparables à ceux des méthodes de superpixels standards. Une implémentation temps-réel de la méthode, éventuellement basée sur une architecture multi-processus, ou GPU, pourrait néanmoins faciliter l'utilisation de SCALP par la communauté, en accélérant la décomposition de très grandes ou de nombreuses images. Ces performances de temps de calcul sont envisageables, puisque de récents travaux ont montré que l'algorithme de groupement itératif utilisé pouvait être implémenté pour fournir des résultats en temps-réel (Neubert et Protzel, 2014; Ban et al., 2016; Choi et Oh, 2016).

De nombreuses applications dont les performances dépendent directement de la décomposition en superpixels, pourraient également bénéficier de la précision de segmentation de la méthode SCALP, comme par exemple la localisation d'objets (Fulkerson et al., 2009), la détection de saillance (Liu et al., 2014), ou le suivi vidéo (Reso et al., 2013). Pour cette dernière application, l'utilisation de SCALP est prometteuse compte tenu de sa précision et de sa haute régularité. Nous avons en effet montré l'intérêt de la régularité pour le suivi des superpixels dans la section 6.4.

3.3 Méthodes d'ANN basés sur les superpixels (SuperPatchMatch)

La méthode SuperPatchMatch a été validée sur la base de référence LFW (Huang et al., 2007b), contenant des images de visages à étiqueter, déjà décomposées en superpixels. L'application à d'autres bases populaires telles que SIFT Flow (Liu et al., 2009), Stanford (Gould et al., 2009) ou MSRC (Shotton et al., 2009) peut être envisagée, en conservant le même pipeline. Cependant, sur ces bases, les performances de segmentation et d'étiquetage seraient dépendantes de la décomposition en superpixels, et ne permettraient pas une comparaison aux méthodes de l'état-de-l'art aussi pertinente.

Les outils proposés permettent l'utilisation robuste des superpixels dans un contexte non local de recherche de correspondances et ouvrent la voie pour de futures adaptations à des méthodes basées sur les superpixels, comme par exemple, la segmentation (Fulkerson et al., 2009; Trulls et al., 2014), l'étiquetage (Gould et al., 2014), la détection de saillance (Pei et al., 2014), ou le transfert de couleurs et de style (Liu et al., 2016a).

SuperPatchMatch peut par exemple être considéré pour initialiser une recherche d'ANN à l'échelle pixellique, lorsque la taille de la bibliothèque d'images d'exemples est trop grande, ou être utilisé au lieu des schémas multirésolution pour initialiser les méthodes de flot optique, afin de mieux capturer de grands déplacements de petits objets.

3.4 Estimation de transformation non linéaire

Dans le chapitre 1, nous avons présenté différentes méthodes de segmentation, basées sur des transformations non linéaires de modèles déjà segmentés. Bien que des méthodes comme OPAL, basées sur des recalages linéaires, soient plus efficaces pour la segmentation de structures anatomiques, le calcul d'une transformation non linéaire entre deux images IRM cérébrales est intéressant pour de nombreuses applications.

Ces transformations permettent de caractériser les déformations locales des structures vers un espace de référence et donc de comparer globalement les caractéristiques anatomiques de groupes de population ou d'effectuer le diagnostic de certaines pathologies (Frisoni et al., 2002; Chen et al., 2009). Ces transformations non linéaires sont définies par une carte, indiquant le déplacement de chaque pixel de l'image vers le sujet à transformer. En pratique, une interpolation est effectuée lors de l'application de la déformation au sujet. Le calcul de ces transformations souffre néanmoins toujours de temps de calcul importants (Avants et al., 2011), limitant l'applicabilité de ces méthodes.

Récemment, les auteurs de Kim et al. (2015) proposent d'estimer la transformation d'un sujet vers un modèle de référence, en utilisant des modèles d'exemples avec leurs cartes de déplacements vers celui-ci, précalculées de façon précise. Une approche par dictionnaire est appliquée à certains patchs *clés* du sujet. La déformation estimée est alors interpolée pour fournir une initialisation dense à une méthode de calcul coûteuse, par exemple (Avants et al., 2011), et celle-ci peut ainsi réduire le temps de calcul, voire améliorer la précision de la transformation finale. Néanmoins, l'approche de Kim et al. (2015) reste calculatoirement coûteuse (plusieurs minutes), et nécessite une interpolation des vecteurs de déplacements, puisqu'elle ne les estime pas directement pour tous les patchs du sujet.

L'algorithme PatchMatch optimisé proposé pour la recherche d'ANN pourrait s'appliquer dans ce contexte pour estimer la carte de déformation du sujet, depuis les cartes des modèles en un temps de calcul réduit. Le principe de la méthode est illustré dans la Figure 9.8. Les patchs les plus proches sont sélectionnés et contrairement à la segmentation, on transfère à présent dans le modèle de fusion l'information de déplacement du patch vers le modèle de référence.

La fusion de vecteurs de déplacement ne peut cependant pas s'effectuer comme la fusion des étiquettes de segmentation, qui permet d'obtenir un estimateur robuste en moyennant directement les contributions par une approche de moyennes non locales. En effet, la moyenne des vecteurs de déplacement s'avère plus sensible aux anomalies et lisse l'information, quand une approche éparse de type LASSO (Tibshirani, 1996) peut permettre de limiter les contributions de vecteurs trop différents (Kim et al., 2015). Des travaux tels que Fortun et al. (2016) se sont récemment intéressés à l'agrégation robuste de vecteurs de déplacements dans le contexte du calcul de flot optique. La Figure 9.9 illustre une première approche simple de fusion, qui consisterait à considérer le vecteur médian parmi un ensemble de vecteurs (Liu,

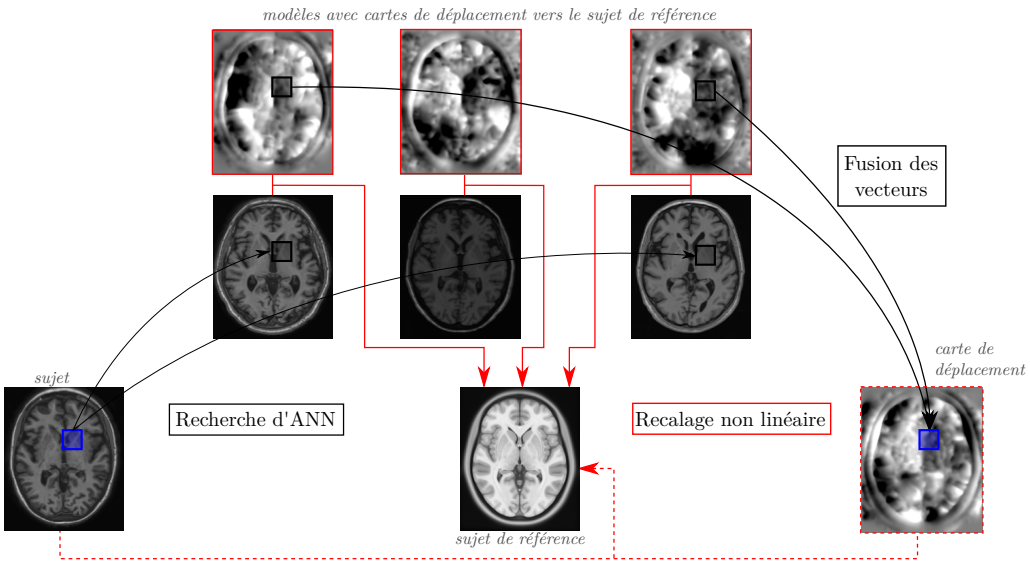
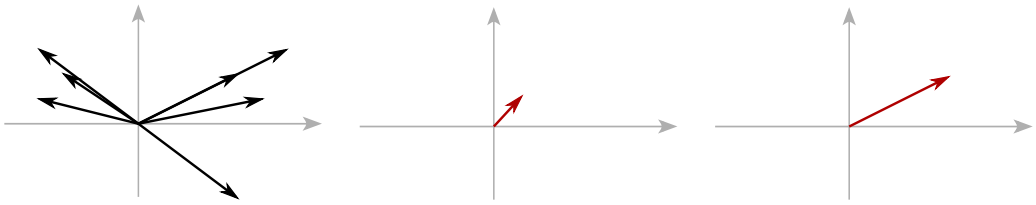


FIGURE 9.8 – Schéma d'application de l'algorithme de recherche d'ANN proposé à l'estimation de transformation non linéaire. Les modèles sont considérés avec leurs cartes de déplacement vers le sujet de référence, et une étape de fusion agrège les vecteurs correspondant aux patchs sélectionnés, pour estimer la déformation du sujet vers la référence.

2013). Cette approche fournit donc un vecteur existant parmi les vecteurs sélectionnés, quand la moyenne lisse l'information de déplacement. Ce vecteur médian, noté v^*, est déterminé comme suit parmi l'ensemble V des k vecteurs sélectionnés :

$$v^* = \underset{v_j \in V}{\operatorname{argmin}} \sum_{i=1}^{k} \|v_j - v_i\|_2 \,. \tag{9.8}$$



(a) Ensemble de vecteurs (b) Vecteur moyen (c) Vecteur médian

FIGURE 9.9 – Exemple de fusion de vecteurs 2D (a), par moyenne (b) et par l'approche du vecteur médian (9.8) (c).

L'algorithme rapide de recherche d'ANN, initialement proposé pour la segmentation, donnerait ici accès à l'information de déplacement des modèles vers le sujet de référence. Ces vecteurs de déplacement doivent alors être fusionnés de façon cohérente pour obtenir une carte de déplacement précise. Selon la précision de la transformation obtenue, celle-ci peut être utilisée directement ou comme initialisation et pourrait ainsi réduire considérablement le temps de calcul des approches usuelles (Kim et al., 2015).

3.5 Segmentation de tumeurs basée sur les supervoxels

Dans la partie I, nous avons proposé une méthode basée sur les patchs pour la segmentation d'images médicales 3D, que nous avons appliquée à la segmentation de structures telles que l'hippocampe ou le cervelet. D'autres structures, telles que les tumeurs affectent directement l'anatomie du cerveau, et leurs contours sont donc hautement corrélés au contenu de l'image. De plus, les tumeurs pouvant apparaître dans différentes régions du cerveau, il n'y a aucun a priori de position permettant d'exploiter un recalage linéaire entre un sujet et des modèles d'exemples. Les supervoxels offrent alors un intérêt pour la segmentation rapide et précise de ces structures par des approches non locales (Wang et Yushkevich, 2013).

Les supervoxels permettent en effet de capturer les contours de la tumeur, tout en réduisant le nombre d'éléments de l'image. Cette réduction du nombre d'éléments s'avère nécessaire pour permettre la recherche de correspondances dans la bibliothèque d'images en un temps de calcul limité. Cependant, dans des méthodes telles que Wang et Yushkevich (2013), la recherche des correspondances est exhaustive, calculée uniquement sur des supervoxels, et en considérant un large histogramme multi-modal des intensités, ce qui conduit à des temps de calcul prohibitifs. La méthode SuperPatchMatch permettrait donc potentiellement d'améliorer les performances de segmentation en utilisant les superpatchs, et de proposer des temps de calcul réduits en considérant des ANN.

Une application de la méthode, avec le même pipeline de fusion d'étiquettes présenté dans le chapitre 8, a été effectuée pour la segmentation d'images 2D issues de la base BRATS (Menze et al., 2015). Nous avons considéré les coupes centrales des images de la base, avec un processus de validation croisée de type *leave-one-out*. Cette base contient des données de faible résolution, d'images IRM de cerveau avec des tumeurs réelles et synthétiques.

Nous comparons dans la Figure 9.10, les résultats obtenus avec SuperPatchMatch (SPM) en utilisant des superpatchs, à ceux obtenus en considérant uniquement des superpixels, de façon similaire à Wang et Yushkevich (2013), et à ceux obtenus avec la méthode OPAL. L'approche basée sur les superpixels (Wang et Yushkevich, 2013) apparaît limitée puisqu'elle échoue à capturer le contexte de la tumeur et à trouver de bonnes correspondances dans d'autres images. Les patchs réguliers sont également limités dans ce contexte, du fait de la grande variation de forme des structures. Les superpatchs fournissent quant à eux un descripteur robuste, puisque la décomposition en superpixels permet de s'adapter aux intensités locales et le superpatch capture un large voisinage décrivant le contexte de chaque superpixel.

Cette simple expérience démontre le potentiel des superpatchs et de SuperPatchMatch, pour la segmentation rapide et précise de tumeurs, sur des images non recalées, en niveaux de gris et à faible résolution. Néanmoins, du fait de la variation importante de forme et de position des tumeurs, la méthode nécessite quelques adaptations pour obtenir des performances compétitives sur les volumes 3D des tumeurs.

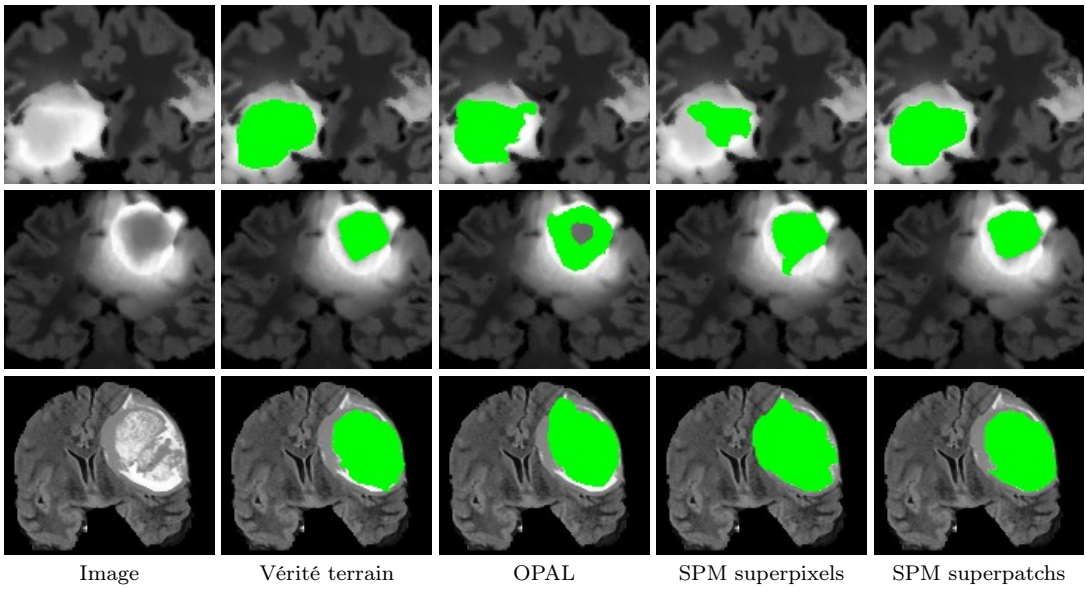


FIGURE 9.10 – Potentiel de SuperPatchMatch (SPM) pour la segmentation de tumeurs. Les exemples de segmentation 2D sont obtenus avec la méthode OPAL, et SPM en utilisant les superpixels ($R = 0$) (Wang et Yushkevich, 2013), et les superpatchs.

3.6 Transfert de style

Enfin, nous pouvons envisager d'utiliser les outils proposés pour le transfert de style entre images. Cette application consiste à transférer le style d'une image source (structure, couleur, texture) vers une image cible, afin de créer une nouvelle image résultant de la fusion des deux autres. Contrairement au transfert de couleurs qui conserve les structures de la cible, le transfert de style modifie donc directement sa structure.

Dans le chapitre 9, nous avons montré que l'algorithme SPM pouvait être contraint pour forcer la sélection globale des superpixels de l'image source. Cependant cette contrainte n'est pas suffisante pour éviter la recopie de larges zones contiguës, qui devient visible pour le transfert de style. La nécessité d'une telle contrainte, empêchant la recopie de larges zones est notamment illustrée dans Frigo et al. (2016). Cette approche utilise une partition régulière de l'image de type arbre quaternaire, qui s'adapte en fonction de la similarité du patch trouvé dans l'image source. Cependant, la considération exhaustive des patchs de l'image source dans le processus de fusion générant le transfert de style peut mener à des temps de calcul très importants.

De plus, la Figure 9.11 montre que contrairement aux approches de type réseau de neurones qui peuvent modifier sensiblement le contenu de l'image cible en s'inspirant du style source (Gatys et al., 2015), le résultat de Frigo et al. (2016) est obtenu depuis le transfert combiné de couleur et de texture de l'image source. Transférer uniquement les couleurs et les textures de la source avec Frigo et al. (2016) limite la modification des éléments de l'image cible et notamment des contours des structures.

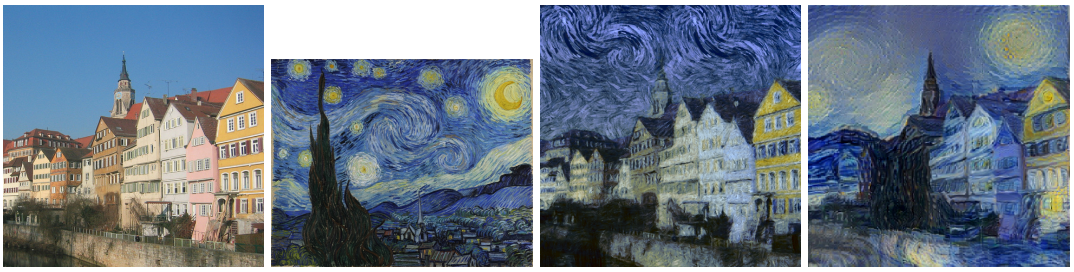


Image cible | Image source | Frigo et al. (2016) | Gatys et al. (2015)

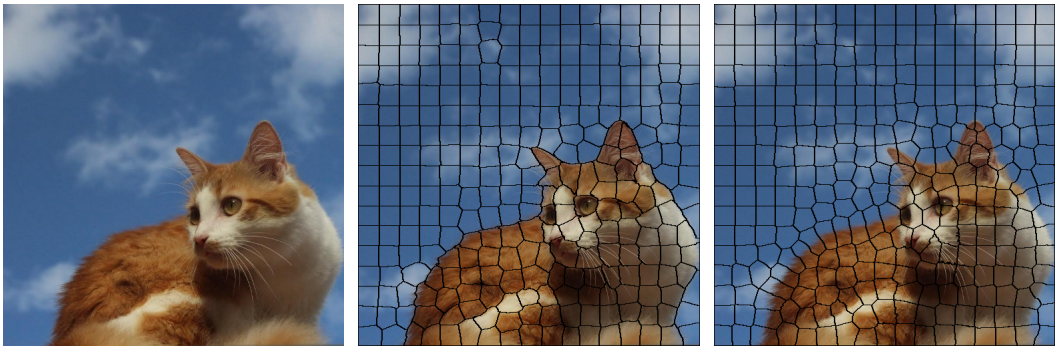
FIGURE 9.11 – Comparaison des résultats de transfert de style entre une approche basée sur le transfert de couleur et de texture depuis une décomposition en blocs réguliers (Frigo et al., 2016), et une approche basée sur un réseau de neurones (Gatys et al., 2015).

L'utilisation des superpixels et d'une méthode de recherche d'ANN peut donc être intéressante pour sélectionner les éléments de l'image cible à transférer, en réduisant le temps de calcul, et nous avons montré dans le chapitre 9 la capacité de notre algorithme basé sur PatchMatch à intégrer des contraintes spatiales.

De plus, afin de pouvoir adapter la structure des contours de l'image source au style souhaité, nous pouvons considérer des superpixels générés de façon à englober les contours, afin qu'ils soient affectés par le transfert. La distance considérée dans la méthode SCALP peut être modifiée afin de générer des superpixels réguliers mais dont la variance de la couleur est élevée et dont les contours des objets sont contenus dans les superpixels. Les contributions des termes de couleur et de contour dans la distance D (5.9) à minimiser pour chaque pixel p sont modifiées comme suit :

$$D(p, C_k) = \left(d_s(p, C_k)\frac{m^2}{r^2} - d_c(p, C_k, \mathbf{P}_p^k) \right) \frac{1}{d_{\mathcal{C}}(\mathbf{P}_p^k)}, \tag{9.9}$$

avec d_s la distance spatiale, et d_c et $d_{\mathcal{C}}$ les distances de couleur et de contour sur le chemin $\mathbf{P}_p^k$ au barycentre du superpixel S_k. Le résultat de ce SCALP *inversé* est illustré dans la Figure 9.12.



(a) Image | (b) SCALP | (c) SCALP *inversé* (9.9)

FIGURE 9.12 – Décomposition d'une image (a) avec la méthode proposée SCALP introduite dans le chapitre 5 (b), et la version modifié (9.9), englobant les contours (c).

L'utilisation d'un algorithme de recherche d'ANN basés sur les superpixels permettrait donc tout d'abord de réduire significativement le temps de calcul de méthodes basées sur les patchs telles que Frigo et al. (2016), et d'envisager le transfert sur des vidéos. De façon similaire à la méthode présentée dans le chapitre 9, la méthode SuperPatchMatch pourrait être contrainte pour éviter la recopie de larges zones contiguës afin de transférer de façon rapide et satisfaisante le style de l'image source. Enfin, afin d'éviter le strict respect des contours des objets de l'image cible, et donc de fournir un transfert de style plus marqué, et similaire à celui des méthodes basées sur les réseaux de neurones, l'algorithme de décomposition en superpixels peut être modifié pour contenir les contours des objets, et les forcer à adopter le style des contours de l'image source. Il existe donc un fort potentiel d'application des outils développés au domaine de l'infographie, et notamment du transfert de style.

Liste des publications

Revues internationales avec comité de lecture (4)

[1] *Evaluation Framework of Superpixel Methods with a Global Regularity Measure*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Journal of Electronic Imaging (JEI), 2017
Special issue on Superpixels for image processing and computer vision

[2] *SuperPatchMatch: An Algorithm for Robust Correspondences using Superpixel Patches*
Rémi Giraud, Vinh-Thong Ta, Aurélie Bugeau, Pierrick Coupé et Nicolas Papadakis
IEEE Transactions on Image Processing (TIP), 2017
Sélectionné pour présentation à ICIP 2017

[3] *CERES: A New Cerebellum Lobule Segmentation Method*
Jose E. Romero, Pierrick Coupé, Rémi Giraud, Vinh-Thong Ta, Vladimir Fonov, et al.
NeuroImage (NIMG), 2016

[4] *An Optimized PatchMatch for Multi-scale and Multi-feature Label Fusion*
Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis, Jose V. Manjón, Louis Collins, et al.
NeuroImage (NIMG), 2015

Conférences internationales avec comité de lecture et actes (5)

[1] *Superpixel-based Color Transfer*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
IEEE International Conference on Image Processing (ICIP), 2017

[2] *Robust Shape Regularity Criteria for Superpixel Evaluation*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
IEEE International Conference on Image Processing (ICIP), 2017

[3] *SCALP: Superpixels with Contour Adherence using Linear Path*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
International Conference on Pattern Recognition (ICPR), 2016

[4] *Patch-based DTI grading: Application to Alzheimer's disease classification*
Kilian Hett, Vinh-Thong Ta, Rémi Giraud, Mary Mondino, Jose V. Manjón, et al.
International Workshop on Patch-based Techniques in Medical Imaging (Patch-MI, MICCAI), 2016

[5] *Optimized PatchMatch for near real time and accurate label fusion*
Vinh-Thong Ta, Rémi Giraud, Louis Collins et Pierrick Coupé
International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2014

Conférences nationales avec comité de lecture et actes (3)

[1] *Décomposition en superpixels via l'utilisation de chemin linéaire*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Groupe d'Études du Traitement du Signal et des Images (GRETSI), 2017

[2] *Transfert de couleurs basé superpixels*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Groupe d'Études du Traitement du Signal et des Images (GRETSI), 2017

[3] *Optimisation de l'algorithme PatchMatch pour la segmentation de structures anatomiques*
Rémi Giraud, Vinh-Thong Ta, Nicolas Papadakis, Louis Collins et Pierrick Coupé
Groupe d'Études du Traitement du Signal et des Images (GRETSI), 2015

Dépôt logiciel (1)

[1] *CERES: A New Cerebellum Lobule Segmentation Method*
Jose V. Manjón, Pierrick Coupé, Jose E. Romero, Vinh-Thong Ta et Rémi Giraud
IDDN.FR.001.470008.000.S.P.2015.000.21000, 2016

Articles en révision (1)

[1] *Robust Superpixels using Color and Contour Features along Linear Path*
Rémi Giraud, Vinh-Thong Ta et Nicolas Papadakis
Computer Vision and Image Understanding (CVIU)

Bibliographie

[Achanta et al. 2012] Achanta, Radhakrishna ; Shaji, Appu ; Smith, Kevin ; Lucchi, Aurélien ; Fua, Pascal ; Süsstrunk, Sabine : SLIC superpixels compared to state-of-the-art superpixel methods. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 34 (2012), n. 11, pp. 2274–2282 7, 81, 83, 85, 86, 87, 89, 92, 98, 99, 100, 104, 113, 114, 116, 119, 124, 131, 132, 137, 145, 159

[Achanta et Süsstrunk 2017] Achanta, Radhakrishna ; Süsstrunk, Sabine : Superpixels and polygons using simple non-iterative clustering. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017 85

[Aljabar et al. 2009] Aljabar, Paul ; Heckemann, Rolf A. ; Hammers, Alexander ; Hajnal, Joseph V. ; Rueckert, Daniel : Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Dans : *Neuroimage* 46 (2009), n. 3, pp. 726–738 23

[Arbelaez et Cohen 2008] Arbelaez, Pablo ; Cohen, Laurent : Constrained image segmentation from hierarchical boundaries. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8 106, 107

[Arbelaez et al. 2009] Arbelaez, Pablo ; Maire, Michael ; Fowlkes, Charless ; Malik, Jitendra : From contours to regions: An empirical evaluation. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2294–2301 106, 107, 110, 145, 146

[Arbelaez et al. 2011] Arbelaez, Pablo ; Maire, Michael ; Fowlkes, Charless ; Malik, Jitendra : Contour detection and hierarchical image segmentation. 33 (2011), n. 5, pp. 898–916 79, 112, 144

[Artaechevarria et al. 2009] Artaechevarria, Xabier ; Muñoz-Barrutia, Arrate ; Ortiz-de-Solórzano, Carlos : Combination strategies in multi-atlas image segmentation: Application to brain MR data. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 28 (2009), n. 8, pp. 1266–1277 23

[Arya et al. 1994] Arya, S. ; Mount, D. ; Netanyahu, N. ; Silverman, R. ; Wu, A. : An optimal algorithm for approximate nearest neighbor searching. (1994), pp. 573–582 30

[Ashikhmin 2001] Ashikhmin, Michael : Synthesizing natural textures. Dans : *Proc. of Symposium on Interactive 3D Graphics (i3D)*, 2001, pp. 217–226 31

[Avants et al. 2011] Avants, Brian B. ; Tustison, Nicholas J. ; Song, Gang ; Cook, Philip A. ; Klein, Arno ; Gee, James C. : A reproducible evaluation of ANTs similarity

metric performance in brain image registration. Dans : *NeuroImage (NIMG)* 54 (2011), n. 3, pp. 2033–2044 21, 38, 51, 199

[Babalola et al. 2009] Babalola, Kolawole O. ; Patenaude, Brian ; Aljabar, Paul ; Schnabel, Julia ; Kennedy, David ; Crum, William ; Smith, Stephen ; Cootes, Tim ; Jenkinson, Mark ; Rueckert, Daniel : An evaluation of four automatic methods of segmenting the subcortical structures in the brain. Dans : *NeuroImage* 47 (2009), n. 1, pp. 1435–1447 21

[Bai et al. 2015] Bai, Wenjia ; Shi, Wenzhe ; Ledig, Christian ; Rueckert, Daniel : Multi-atlas segmentation with augmented features for cardiac MR images. Dans : *Medical Image Analysis (MIA)* 19 (2015), pp. 98–109 44

[Ban et al. 2016] Ban, Zhihua ; Liu, Jianguo ; Fouriaux, Jeremy : GLSC: LSC superpixels at over 130 FPS. Dans : *Journal of Real-Time Image Processing (JRTIP)* (2016), pp. 1–12 116, 198

[Bao et al. 2014] Bao, Linchao ; Yang, Qingxiong ; Jin, Hailin : Fast edge-preserving PatchMatch for large displacement optical flow. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3534–3541 27, 31

[Barnes et al. 2008] Barnes, Josephine ; Foster, Jo ; Boyes, Richard G. ; Pepple, Tracey et al. : A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Dans : *NeuroImage (NIMG)* 40 (2008), n. 4, pp. 1655–1671 22

[Barnes et al. 2009] Barnes, Connelly ; Shechtman, Eli ; Finkelstein, Adam ; Goldman, Dan B. : PatchMatch: A randomized correspondence algorithm for structural image editing. Dans : *ACM Trans. on Graphics (ToG)* 28 (2009), n. 3 6, 9, 10, 11, 12, 30, 31, 32, 36, 37, 38, 40, 41, 154, 161, 162, 163, 170, 182, 194

[Barnes et al. 2010] Barnes, Connelly ; Shechtman, Eli ; Goldman, Dan B. ; Finkelstein, Adam : The generalized PatchMatch correspondence algorithm. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2010, pp. 29–43 41, 154

[Barnes et al. 2015] Barnes, Connelly ; Zhang, Fang-Lue ; Lou, Liming ; Wu, Xian ; Hu, Shi-Min : PatchTable: Efficient patch queries for large datasets and applications. Dans : *ACM Trans. on Graphics (ToG)* 34 (2015), n. 4, pp. 97 7, 31, 194

[Bay et al. 2006] Bay, Herbert ; Tuytelaars, Tinne ; Van Gool, Luc : SURF: Speeded up robust features. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2006, pp. 404–417 46, 154

[Benesova et Kottman 2014] Benesova, Wanda ; Kottman, Michal : Fast superpixel segmentation using morphological processing. Dans : *Proc. of the Int. Conf. on Machine Vision and Machine Learning (MVML)*, 2014 86, 87

[Besse et al. 2014] Besse, Frederic ; Rother, Carsten ; Fitzgibbon, Andrew ; Kautz, Jan : PMBP: PatchMatch Belief Propagation for Correspondence Field Estimation. Dans : *International Journal of Computer Vision (IJCV)* (2014) 37

[Bloch 2005] Bloch, Isabelle : Fuzzy spatial relationships for image processing and interpretation: A review. Dans : *Image and Vision Computing.* 23 (2005), n. 2, pp. 89–110 155

[Boccardi et al. 2014] Boccardi, Marina ; Bocchetta, Martina ; Apostolova, Liana G. ; Barnes, Josephine ; Bartzokis, George ; Corbetta, Gabriele ; DeCarli, Charles ; Firbank, Michael ; Ganzola, Rossana ; Gerritsen, Lotte et al. : Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. Dans : *Alzheimer's & Dementia* 11 (2014), n. 2, pp. 126–138 51, 67

[Borji et Itti 2012] Borji, Ali ; Itti, Laurent : Exploiting local and global patch rarities for saliency detection. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 478–485 27

[Boykov et al. 2001] Boykov, Yuri ; Veksler, Olga ; Zabih, Ramin : Fast approximate energy minimization via graph cuts. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 23 (2001), n. 11, pp. 1222–1239 169, 170

[Brekhna et al. 2017] Brekhna, Brekhna ; Mahmood, Arif ; Zhou, Yuanfeng ; Zhang, Caiming : Robustness analysis of superpixel algorithms to image blur, additive Gaussian noise, and impulse noise. Dans : *Journal of Electronic Imaging (JEI)* 26 (2017) 98, 127

[Bresenham 1965] Bresenham, Jack E. : Algorithm for computer control of a digital plotter. Dans : *IBM Systems Journal* 4 (1965), n. 1, pp. 25–30 102, 103, 106

[Buades et al. 2005] Buades, Antoni ; Coll, Bartomeu ; Morel, Jean-Michel : A non-local algorithm for image denoising. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* 2, 2005, pp. 60–65 5, 6, 23, 24, 28, 29, 42, 154, 181, 185

[Bugeau et al. 2014] Bugeau, Aurélie ; Ta, Vinh-Thong ; Papadakis, Nicolas : Variational exemplar-based image colorization. Dans : *IEEE Trans. on Image Processing (TIP)* 23 (2014), n. 1, pp. 298–307 27

[Buyssens et al. 2014] Buyssens, Pierre ; Gardin, Isabelle ; Ruan, Su ; Elmoataz, Abderrahim : Eikonal-based region growing for efficient clustering. Dans : *Image and Vision Computing* 32 (2014), n. 12, pp. 1045–1054 81, 82, 83, 92, 100, 114, 119, 124, 127, 145, 159, 160

[Calonder et al. 2010] Calonder, Michael ; Lepetit, Vincent ; Strecha, Christoph ; Fua, Pascal : Brief: Binary robust independent elementary features. Dans : *Proc. of European Conference on Computer Vision (ECCV)* (2010), pp. 778–792 154

[Cardoso et al. 2013] Cardoso, Manuel Jorge ; Leung, Kelvin ; Modat, Marc ; Keihaninejad, Shiva ; Cash, David ; Barnes, Josephine ; Fox, Nick C. ; Ourselin, Sebastien et al. : STEPS: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation. Dans : *Medical Image Analysis (MIA)* 17 (2013), n. 6, pp. 671–684 23

[Caselles et al. 2009] CASELLES, Vicent ; FACCIOLO, Gabriele ; MEINHARDT, Enric : Anisotropic Cheeger sets and applications. Dans : *SIAM Journal on Imaging Sciences (SIIMS)* 2 (2009), n. 4, pp. 1211–1254 128

[Chakravarty et al. 2013] CHAKRAVARTY, Mallar ; STEADMAN, Patrick ; EEDE, Matthijs C. ; CALCOTT, Rebecca D. ; GU, Victoria ; SHAW, Philip ; RAZNAHAN, Armin ; COLLINS, D. Louis ; LERCH, Jason P. : Performing label-fusion-based segmentation using multiple automatically generated templates. Dans : *Human Brain Mapping (HBM)* 34 (2013), n. 10, pp. 2635–2654 64

[Chang et al. 2013] CHANG, Jason ; WEI, Donglai ; FISHER, John W. : A video representation using temporal superpixels. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2051–2058 79, 81, 119, 124, 140, 141, 142

[Chen et al. 2017] CHEN, Jiansheng ; LI, Zhengqin ; HUANG, Bo : Linear spectral clustering superpixel. Dans : *IEEE Trans. on Image Processing (TIP)* (2017) 81, 85, 92, 100, 109, 110, 112, 114, 115, 124, 145

[Chen et al. 2009] CHEN, Yasheng ; AN, Hongyu ; ZHU, Hongtu ; STONE, Taylor ; SMITH, J K. ; HALL, Colin ; BULLITT, Elizabeth ; SHEN, Dinggang ; LIN, Weili : White matter abnormalities revealed by diffusion tensor imaging in non-demented and demented HIV+ patients. Dans : *NeuroImage (NIMG)* 47 (2009), n. 4, pp. 1154–1162 199

[Chen et al. 2013] CHEN, Zhuoyuan ; JIN, Hailin ; LIN, Zhe ; COHEN, Scott ; WU, Ying : Large displacement optical flow from nearest neighbor fields. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2443–2450 27

[Choi et Oh 2016] CHOI, Kang-Sun ; OH, Ki-Won : Subsampling-based acceleration of simple linear iterative clustering for superpixel segmentation. Dans : *Computer Vision and Image Understanding (CVIU)* 146 (2016), pp. 1–8 116, 198

[Clément et al. 2015] CLÉMENT, Michaël ; GARNIER, Mickaël ; KURTZ, Camille ; WENDLING, Laurent : Color object recognition cased on spatial relations between image layers. Dans : *Proc. of Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2015, pp. 427–434 154

[Collins et al. 1994] COLLINS, D. Louis ; NEELIN, Peter ; PETERS, Terrence ; EVANS, Alan : Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. Dans : *Journal of Computer Assisted Tomography (JCAT)* 18 (1994), n. 2, pp. 192–205 51

[Collins et al. 1995] COLLINS, D. Louis ; HOLMES, Colin ; PETERS, Terrence ; EVANS, Alan : Automatic 3-D model-based neuroanatomical segmentation. Dans : *Human Brain Mapping (HBM)* 3 (1995), n. 3, pp. 190–208 21, 23

[Collins et Pruessner 2010] COLLINS, D. Louis ; PRUESSNER, Jens : Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Dans : *NeuroImage (NIMG)* 52 (2010), n. 4, pp. 1355–1366 23, 24, 54, 56, 58

[Comaniciu et Meer 2002] Comaniciu, Dorin ; Meer, Peter : Mean shift: A robust approach toward feature space analysis. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 24 (2002), n. 5, pp. 603–619 84

[Coupé et al. 2008] Coupé, Pierrick ; Yger, Pierre ; Prima, Sylvain ; Hellier, Pierre ; Kervrann, Charles ; Barillot, Christian : An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 27 (2008), n. 4, pp. 425–441 21, 51

[Coupé et al. 2010] Coupé, Pierrick ; Manjón, José V. ; Gedamu, Elias ; Arnold, Douglas ; Robles, Montserrat ; Collins, D. Louis : Robust Rician noise estimation for MR images. Dans : *Medical Image Analysis (MIA)* 14 (2010), n. 4, pp. 483–493 51

[Coupé et al. 2011] Coupé, Pierrick ; Manjón, José V. ; Fonov, Vladimir ; Pruessner, Jens ; Robles, Montserrat ; Collins, D. Louis : Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. Dans : *NeuroImage (NIMG)* 54 (2011), n. 2, pp. 940–954 23, 25, 27, 28, 32, 42, 53, 56, 58, 68

[Coupé et al. 2012a] Coupé, Pierrick ; Eskildsen, Simon F. ; Manjón, José V. ; Fonov, Vladimir S. ; Pruessner, Jens C. ; Allard, Michèle ; Collins, D. Louis et al. : Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. Dans : *NeuroImage: clinical* 1 (2012), n. 1, pp. 141–152 68

[Coupé et al. 2012b] Coupé, Pierrick ; Eskildsen, Simon F. ; Manjón, José V. ; Fonov, Vladimir S. ; Collins, D. Louis et al. : Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. Dans : *NeuroImage (NIMG)* 59 (2012), n. 4, pp. 3736–3747 4

[Criminisi et al. 2004] Criminisi, Antonio ; Pérez, Patrick ; Toyama, Kentaro : Region filling and object removal by exemplar-based image inpainting. Dans : *IEEE Trans. on Image Processing (TIP)* 13 (2004), n. 9, pp. 1200–1212 27, 28

[Daisy et al. 2013] Daisy, Maxime ; Tschumperlé, David ; Lézoray, Olivier : A fast spatial patch blending algorithm for artefact reduction in pattern-based image inpainting. Dans : *SIGGRAPH Asia Technical Briefs*, 2013, pp. 8:1–8:4 27, 28

[Dalal et Triggs 2005] Dalal, Navneet ; Triggs, Bill : Histograms of oriented gradients for human detection. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886–893 46, 160, 170

[Dellinger et al. 2015] Dellinger, Flora ; Delon, Julie ; Gousseau, Yann ; Michel, Julien ; Tupin, Florence : SAR-SIFT: a SIFT-like algorithm for SAR images. Dans : *IEEE Trans. on Geoscience and Remote Sensing (TGRS)* 53 (2015), n. 1, pp. 453–466 154

[Dollár et Zitnick 2013] Dollár, Piotr ; Zitnick, Lawrence : Structured forests for fast edge detection. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 1841–1848 107, 110, 113

[Duan et Lafarge 2015] Duan, Liuyun ; Lafarge, Florent : Image partitioning into convex polygons. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3119–3127 127

[Efros et Leung 1999] Efros, Alexei A. ; Leung, Thomas K. : Texture synthesis by non-parametric sampling. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)* 2, 1999, pp. 1033–1038 6, 27, 154

[Eskildsen et al. 2012] Eskildsen, Simon F. ; Coupé, Pierrick ; Fonov, Vladimir ; Manjón, José V. ; Leung, Kelvin K. ; Guizard, Nicolas ; Wassef, Shafik N. ; Østergaard, Lasse R. ; Collins, D. Louis et al. : BEaST: Brain extraction based on nonlocal segmentation technique. Dans : *NeuroImage (NIMG)* 59 (2012), n. 3, pp. 2362–2373 24, 25, 44, 60

[Felzenszwalb et al. 2010] Felzenszwalb, Pedro F. ; Girshick, Ross B. ; McAllester, David ; Ramanan, Deva : Object detection with discriminatively trained part based models. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 32 (2010), n. 9, pp. 1627–1645 155

[Felzenszwalb et Huttenlocher 2004] Felzenszwalb, Pedro F. ; Huttenlocher, Daniel P. : Efficient graph-based image segmentation. Dans : *International Journal of Computer Vision (IJCV)* 59 (2004), n. 2, pp. 167–181 81, 82, 125, 132

[Ferradans et al. 2014] Ferradans, Sira ; Papadakis, Nicolas ; Peyré, Gabriel ; Aujol, Jean-François : Regularized discrete optimal transport. Dans : *SIAM Journal on Imaging Sciences (SIIMS)* 7 (2014), n. 3, pp. 1853–1882 180

[Fischl et al. 2002] Fischl, Bruce ; Salat, David H. ; Busa, Evelina ; Albert, Marilyn ; Dieterich, Megan ; Haselgrove, Christian ; Van Der Kouwe, Andre ; Killiany, Ron ; Kennedy, David ; Klaveness, Shuna et al. : Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Dans : *Neuron* 33 (2002), n. 3, pp. 341–355 197

[Fortun et al. 2016] Fortun, Denis ; Bouthemy, Patrick ; Kervrann, Charles : A variational aggregation framework for patch-based optical flow estimation. Dans : *Journal of Mathematical Imaging and Vision (JMIV)* 56 (2016), n. 2, pp. 280–299 199

[Frigo et al. 2014] Frigo, Oriel ; Sabater, Neus ; Demoulin, Vincent ; Hellier, Pierre : Optimal transportation for example-guided color transfer. Dans : *Proc. of Asian Conference on Computer Vision (ACCV)* 9005, 2014, pp. 655–670 180

[Frigo et al. 2016] Frigo, Oriel ; Sabater, Neus ; Delon, Julie ; Hellier, Pierre : Split and match: Example-based adaptive patch sampling for unsupervised style transfer. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 553–561 202, 203, 204

[Frisoni et al. 2002] Frisoni, Giovanni B. ; Testa, Cristina ; Zorzan, Alessandra ; Sabattoli, Francesca ; Beltramello, Alberto ; Soininen, Hilkka ; Laakso, Mikko P. : Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry.

Dans : *Journal of Neurology, Neurosurgery & Psychiatry* 73 (2002), n. 6, pp. 657–664 199

[Frisoni et al. 2015] Frisoni, Giovanni B ; Jack, Clifford R. ; Bocchetta, Martina ; Bauer, Corinna ; Frederiksen, Kristian S. ; Liu, Yawu ; Preboske, Gregory ; Swihart, Tim ; Blair, Melanie ; Cavedo, Enrica et al. : The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: Evidence of validity. Dans : *Alzheimer's & Dementia* 11 (2015), n. 2, pp. 111–125 51, 59

[Fulkerson et al. 2009] Fulkerson, Brian ; Vedaldi, Andrea ; Soatto, Stefano : Class segmentation and object localization with superpixel neighborhoods. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 670–677 7, 79, 198

[Gadde et al. 2016] Gadde, Raghudeep ; Jampani, Varun ; Kiefel, Martin ; Kappler, Daniel ; Gehler, Peter V. : Superpixel convolutional networks using bilateral inceptions. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2016, pp. 597–613 7, 79

[Garnier et al. 2012] Garnier, Mickaël ; Hurtut, Thomas ; Wendling, Laurent : Object description based on spatial relations between level-sets. Dans : *Proc. of Int. Conf. on Digital Image Computing Techniques and Applications (DICTA)*, 2012, pp. 1–7 154

[Gatys et al. 2015] Gatys, Leon A. ; Ecker, Alexander S. ; Bethge, Matthias : A neural algorithm of artistic style. Dans : *arXiv preprint arXiv:1508.06576* (2015) 202, 203

[Giraud et al. 2015] Giraud, Rémi ; Ta, Vinh-Thong ; Papadakis, Nicolas ; Collins, D. Louis ; Coupé, Pierrick : Optimisation de l'algorithme PatchMatch pour la segmentation de structures anatomiques. Dans : *Actes du Groupe d'Études du Traitement du Signal et des Images (GRETSI)*, 2015 11

[Giraud et al. 2016a] Giraud, Rémi ; Ta, Vinh-Thong ; Papadakis, Nicolas ; Manjón, José V. ; Collins, D. Louis ; Coupé, Pierrick et al. : An optimized PatchMatch for multi-scale and multi-feature label fusion. Dans : *NeuroImage (NIMG)* 124 (2016), pp. 770–782 11, 50

[Giraud et al. 2016b] Giraud, Rémi ; Ta, Vinh-Thong ; Papadakis, Nicolas : SCALP: Superpixels with contour adherence using linear path. Dans : *Proc. of International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2374–2379 11

[Giraud et al. 2017a] Giraud, Rémi ; Ta, Vinh-Thong ; Bugeau, Aurélie ; Coupé, Pierrick ; Papadakis, Nicolas : SuperPatchMatch: An algorithm for robust correspondences using superpixel patches. Dans : *IEEE Trans. on Image Processing (TIP)* (2017) 11, 12

[Giraud et al. 2017b] Giraud, Rémi ; Ta, Vinh-Thong ; Papadakis, Nicolas : Décomposition en superpixels via l'utilisation de chemin linéaire. Dans : *Actes du Groupe d'Études du Traitement du Signal et des Images (GRETSI)*, 2017 11

[Giraud et al. 2017c] GIRAUD, Rémi ; TA, Vinh-Thong ; PAPADAKIS, Nicolas : Evaluation framework of superpixel methods with a global regularity measure. Dans : *Journal of Electronic Imaging (JEI)* (2017) 11

[Giraud et al. 2017d] GIRAUD, Rémi ; TA, Vinh-Thong ; PAPADAKIS, Nicolas : Robust shape regularity criteria for superpixel evaluation. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2017 11

[Giraud et al. 2017e] GIRAUD, Rémi ; TA, Vinh-Thong ; PAPADAKIS, Nicolas : Robust superpixels using color and contour features along linear path. Dans : *HAL preprint https://hal.archives-ouvertes.fr/hal-01510063* (2017) 11

[Giraud et al. 2017f] GIRAUD, Rémi ; TA, Vinh-Thong ; PAPADAKIS, Nicolas : Superpixel-based color transfer. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2017 11, 12

[Giraud et al. 2017g] GIRAUD, Rémi ; TA, Vinh-Thong ; PAPADAKIS, Nicolas : Transfert de couleurs basé superpixels. Dans : *Actes du Groupe d'Études du Traitement du Signal et des Images (GRETSI)*, 2017 11, 12

[Gould et al. 2009] GOULD, Stephen ; FULTON, Richard ; KOLLER, Daphne : Decomposing a scene into geometric and semantically consistent regions. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 1–8 198

[Gould et al. 2014] GOULD, Stephen ; ZHAO, Jiecheng ; HE, Xuming ; ZHANG, Yuhang : Superpixel graph label transfer with learned distance metric. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2014, pp. 632–647 7, 79, 80, 155, 159, 160, 170, 198

[Gray et al. 2014] GRAY, Katherine R. ; AUSTIN, Mark ; WOLZ, Robin ; MCLEISH, Kate ; BOCCARDI, Marina ; FRISONI, Giovanni ; HILL, Derek : Integration of EADC-ADNI Harmonised hippocampus labels into the LEAP automated segmentation technique. Dans : *Alzheimer's & Dementia* 10 (2014), pp. 555 58, 60

[Gutierrez et al. 2017] GUTIERREZ, Jorge ; RABIN, Julien ; GALERNE, Bruno ; HURTUT, Thomas : Optimal patch assignment for statistically constrained texture synthesis. Dans : *Proc. of Int. Conf. on Scale Space and Variational Methods in Computer Vision (SSVM)*, 2017, pp. 172–183 181

[He et Sun 2012] HE, Kaiming ; SUN, Jian : Computing nearest-neighbor fields via propagation-assisted kd-trees. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 111–118 30

[He et al. 2004] HE, Xuming ; ZEMEL, Richard S. ; CARREIRA-PERPIÑÁN, Miguel Á. : Multiscale conditional random fields for image labeling. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* 2, 2004 155

[Heckemann et al. 2006] Heckemann, Rolf A. ; Hajnal, Joseph V. ; Aljabar, Paul ; Rueckert, Daniel ; Hammers, Alexander : Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Dans : *NeuroImage (NIMG)* 33 (2006), n. 1, pp. 115–126 21, 23, 24

[Hett et al. 2016] Hett, Kilian ; Ta, Vinh-Thong ; Giraud, Rémi ; Mondino, Mary ; Manjón, José V. ; Coupé, Pierrick et al. : Patch-based DTI grading: Application to Alzheimer's disease classification. Dans : *Proc. of Int. Work. on Patch-based Techniques in Medical Imaging (Patch-MI, MICCAI)*, 2016, pp. 76–83 11, 68, 69

[Hu et al. 2014] Hu, Shiyan ; Coupé, Pierrick ; Pruessner, Jens C. ; Collins, D. Louis : Nonlocal regularization for active appearance model: Application to medial temporal lobe segmentation. Dans : *Human Brain Mapping (HBM)* 35 (2014), n. 2, pp. 377–395 24

[Hu et al. 2016] Hu, Yinlin ; Song, Rui ; Li, Yunsong : Efficient coarse-to-fine patchmatch for large displacement optical flow. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5704–5712 31

[Huang et al. 2007a] Huang, Gary B. ; Jain, Vidit ; Learned-Miller, Erik : Unsupervised joint alignment of complex images. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2007 169

[Huang et al. 2007b] Huang, Gary B. ; Ramesh, Manu ; Berg, Tamara ; Learned-Miller, Erik : Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Dans : *Technical Report 07-49, University of Massachusetts, Amherst* (2007) 168, 169, 198

[Huang et al. 2016] Huang, Chun-Rong ; Wang, Wei-An ; Lin, Szu-Yu ; Lin, Yen-Yu : USEQ: Ultra-fast superpixel extraction via quantization. Dans : *Proc. of International Conference on Pattern Recognition (ICPR)*, 2016 85

[Indyk et Motwani 1998] Indyk, Piotr ; Motwani, Rajeev : Approximate nearest neighbors: Towards removing the curse of dimensionality. Dans : *Proc. of the ACM Symposium on Theory of Computing (STOC)*, 1998, pp. 604–613 31

[Irving et al. 2014] Irving, Benjamin ; Cifor, Amalia ; Papież, Bartłomiej W. ; Franklin, Jamie ; Anderson, Ewan M. ; Brady, Michael ; Schnabel, Julia A. : Automated colorectal tumour segmentation in DCE-MRI using supervoxel neighbourhood contrast characteristics. Dans : *Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2014, pp. 609–616 80

[Jaccard 1901] Jaccard, Paul : Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Dans : *Bulletin de la Société Vaudoise des Sciences Naturelles* 37 (1901), pp. 547–579 133

[Jack et al. 2008] Jack, Clifford R. ; Bernstein, Matt A. ; Fox, Nick C. ; Thompson, Paul ; Alexander, Gene ; Harvey, Danielle ; Borowski, Bret ; Britson, Paula J. ; L Whitwell, Jennifer ; Ward, Chadwick et al. : The Alzheimer's disease neuroimaging

initiative (ADNI): MRI methods. Dans : *Journal of Magnetic Resonance Imaging* 27 (2008), n. 4, pp. 685–691 50, 51, 66

[Jain et al. 2014] Jain, Mihir ; Van Gemert, Jan ; Jégou, Hervé ; Bouthemy, Patrick ; Snoek, Cees G. M. : Action localization with tubelets from motion. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 740–747 81, 119

[Jia et al. 2015] Jia, Shaoyong ; Geng, Shijie ; Gu, Yun ; Yang, Jie ; Shi, Pengfei ; Qiao, Yu : NSLIC: SLIC superpixels based on nonstationarity measure. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4738–4742 85

[Kae et al. 2013] Kae, Andrew ; Sohn, Kihyuk ; Lee, Honglak ; Learned-Miller, Erik : Augmenting CRFs with Boltzmann machine shape priors for image labeling. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2019–2026 79, 80, 155, 170, 175, 176, 177

[Khan et al. 2011] Khan, Ali R. ; Cherbuin, Nicolas ; Wen, Wei ; Anstey, Kaarin J. ; Sachdev, Perminder ; Beg, Mirza F. : Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): Validation on hippocampus segmentation. Dans : *NeuroImage (NIMG)* 56 (2011), n. 1, pp. 126–139 23

[Kim et al. 2013a] Kim, Kwang-Shik ; Zhang, Dongni ; Kang, Mun-Cheon ; Ko, Sung-Jea : Improved simple linear iterative clustering superpixels. Dans : *Proc. of International Symposium on Consumer Electronics (ISCE)*, 2013, pp. 259–260 85

[Kim et al. 2013b] Kim, Minjeong ; Wu, Guorong ; Li, Wei ; Wang, Li ; Son, Young-Don ; Cho, Zang-Hee ; Shen, Dinggang : Automatic hippocampus segmentation of 7.0 Tesla MR images by combining multiple atlases and auto-context models. Dans : *NeuroImage (NIMG)* 83 (2013), pp. 335–345 44

[Kim et al. 2015] Kim, Minjeong ; Wu, Guorong ; Wang, Qian ; Lee, Seong-Whan ; Shen, Dinggang : Improved image registration by sparse patch-based deformation estimation. Dans : *NeuroImage (NIMG)* 105 (2015), pp. 257–268 199, 200

[Klein et al. 2009] Klein, Arno ; Andersson, Jesper ; Ardekani, Babak A. ; Ashburner, John ; Avants, Brian ; Chiang, Ming-Chang ; Christensen, Gary E. ; Collins, D. Louis ; Gee, James ; Hellier, Pierre et al. : Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Dans : *NeuroImage (NIMG)* 46 (2009), n. 3, pp. 786–802 21

[Klöppel et al. 2008] Klöppel, Stefan ; Stonnington, Cynthia M. ; Chu, Carlton ; Draganski, Bogdan ; Scahill, Rachael I. ; Rohrer, Jonathan D. ; Fox, Nick C. ; Jack Jr, Clifford R. ; Ashburner, John ; Frackowiak, Richard S. : Automatic classification of MR scans in Alzheimer's disease. Dans : *Brain* 131 (2008), n. 3, pp. 681–689 21

[Korman et Avidan 2011] Korman, Simon ; Avidan, Shai : Coherency sensitive hashing. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1607–1614 30, 184, 185

[Korman et Avidan 2016] Korman, Simon ; Avidan, Shai : Coherency sensitive hashing. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 38 (2016), n. 6, pp. 1099–1112 6, 31, 32, 194

[Krizhevsky et Hinton 2009] Krizhevsky, Alex ; Hinton, Geoffrey : Learning multiple layers of features from tiny images. (2009) 5

[Krizhevsky et al. 2012] Krizhevsky, Alex ; Sutskever, Ilya ; Hinton, Geoffrey E. : ImageNet Classification with Deep Convolutional Neural Networks. Dans : *Proc. of Int. Conf. on Neural Information Processing Systems (NIPS)*. 2012, pp. 1097–1105 5

[Lafferty et al. 2001] Lafferty, John ; McCallum, Andrew ; Pereira, Fernando C. : Conditional random fields: Probabilistic models for segmenting and labeling sequence data. Dans : *International Conference on Machine Learning (ICML)*, 2001, pp. 282–289 175

[Lee et Schachter 1980] Lee, Der-Tsai ; Schachter, Bruce J. : Two algorithms for constructing a Delaunay triangulation. Dans : *International Journal of Computer & Information Sciences* 9 (1980), n. 3, pp. 219–242 127

[Levinshtein et al. 2009] Levinshtein, Alex ; Stere, Adrian ; Kutulakos, Kiriakos N. ; Fleet, David J. ; Dickinson, Sven J. ; Siddiqi, Kaleem : TurboPixels: Fast superpixels using geometric flows. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 31 (2009), n. 12, pp. 2290–2297 81, 84, 86, 89, 116, 136, 142

[Levinshtein et al. 2010] Levinshtein, Alex ; Sminchisescu, Cristian ; Dickinson, Sven : Optimal contour closure by superpixel grouping. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2010, pp. 480–493 80

[Li et al. 2012] Li, Zhenguo ; Wu, Xiao-Ming ; Chang, Shih-Fu : Segmentation using superpixels: A bipartite graph partitioning approach. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 789–796 80

[Liu 2013] Liu, Yike : Noise reduction by vector median filtering. Dans : *Geophysics* 78 (2013), n. 3, pp. 79–87 199

[Liu et al. 2009] Liu, Ce ; Yuen, Jenny ; Torralba, Antonio : Nonparametric scene parsing: Label transfer via dense scene alignment. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1972–1979 198

[Liu et al. 2011] Liu, Ming-Yu ; Tuzel, Oncel ; Ramalingam, Srikumar ; Chellappa, Rama : Entropy rate superpixel segmentation. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2097–2104 81, 83, 85, 86, 87, 90, 92, 100, 114, 115, 116, 136

[Liu et al. 2014] LIU, Zhi ; ZHANG, Xiang ; LUO, Shuhua ; LE MEUR, Olivier : Superpixel-based spatiotemporal saliency detection. Dans : *IEEE Trans. on Circuits and Systems for Video Technology (TCSVT)* 24 (2014), n. 9, pp. 1522–1540 79, 80, 198

[Liu et al. 2015a] LIU, Sifei ; YANG, Jimei ; HUANG, Chang ; YANG, Ming-Hsuan : Multi-objective convolutional learning for face labeling. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3451–3459 155, 170, 175, 176, 177

[Liu et al. 2015b] LIU, Fayao ; SHEN, Chunhua ; LIN, Guosheng : Deep convolutional neural fields for depth estimation from a single image. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5162–5170 79

[Liu et al. 2016a] LIU, Jiaying ; YANG, Wenhan ; SUN, Xiaoyan ; ZENG, Wenjun : Photo stylistic brush: Robust style transfer via superpixel-based bipartite graph. Dans : *arXiv preprint arXiv:1606.03871* (2016) 154, 180, 198

[Liu et al. 2016b] LIU, Yong-Jin ; YU, Cheng-Chi ; YU, Min-Jing ; HE, Ying : Manifold SLIC: A fast method to compute content-sensitive superpixels. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 651–659 85

[Lloyd 1982] LLOYD, Stuart : Least squares quantization in PCM. Dans : *IEEE Trans. on Information Theory* 28 (1982), n. 2, pp. 129–137 108

[Long et al. 2015] LONG, Jonathan ; SHELHAMER, Evan ; DARRELL, Trevor : Fully convolutional networks for semantic segmentation. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440 155

[Lötjönen et al. 2010] LÖTJÖNEN, Jyrki M. ; WOLZ, Robin ; KOIKKALAINEN, Juha R. ; THURFJELL, Lennart ; WALDEMAR, Gunhild ; SOININEN, Hilkka ; RUECKERT, Daniel et al. : Fast and robust multi-atlas segmentation of brain magnetic resonance images. Dans : *NeuroImage (NIMG)* 49 (2010), n. 3, pp. 2352–2365 21, 23

[Lowe 2004] LOWE, David G. : Distinctive image features form scale-invariant keypoints. Dans : *International Journal of Computer Vision (IJCV)* 60 (2004), n. 2, pp. 91–110 46, 154

[Lu et al. 2013] LU, Jiangbo ; YANG, Hongsheng ; MIN, Dongbo ; DO, Minh N. : PatchMatch filter: Efficient edge-aware filtering meets randomized search for fast correspondence field estimation. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1854–1861 85, 155

[Lucchi et al. 2012] LUCCHI, Aurélien ; SMITH, Kevin ; ACHANTA, Radhakrishna ; KNOTT, Graham ; FUA, Pascal : Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 31 (2012), n. 2, pp. 474–486 85

[Machairas et al. 2015] Machairas, Vaïa ; Faessel, Matthieu ; Cárdenas-Peña, David ; Chabardes, Théodore ; Walter, Thomas ; Decencière, Etienne : Waterpixels. Dans : *IEEE Trans. on Image Processing (TIP)* 24 (2015), n. 11, pp. 3707–3716 81, 84, 87, 92, 93, 101, 114, 119, 124, 126, 127, 129, 133, 134, 145

[Maire et al. 2008] Maire, Michael ; Arbelaez, Pablo ; Fowlkes, Charless ; Malik, Jitendra : Using contours to detect and localize junctions in natural images. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8 113

[Manjón et al. 2008a] Manjón, José V. ; Tohka, Jussi ; García-Martí, Gracian ; Carbonell-Caballero, José ; Lull, Juan J. ; Martí-Bonmatí, Luís ; Robles, Montserrat : Robust MRI brain tissue parameter estimation by multistage outlier rejection. Dans : *Magnetic Resonance in Medicine* 59 (2008), n. 4, pp. 866–873 51

[Manjón et al. 2008b] Manjón, José V. ; Carbonell-Caballero, José ; Lull, Juan J. ; García-Martí, Gracián ; Martí-Bonmatí, Luís ; Robles, Montserrat : MRI denoising using non-local means. Dans : *Medical Image Analysis (MIA)* 12 (2008), n. 4, pp. 514–523 21

[Manjón et al. 2010] Manjón, José V. ; Coupé, Pierrick ; Martí-Bonmatí, Luis ; Collins, D. Louis ; Robles, Montserrat : Adaptive non-local means denoising of MR images with spatially varying noise levels. Dans : *Journal of Magnetic Resonance Imaging* 31 (2010), n. 1, pp. 192–203 51

[Manjón et al. 2014] Manjón, José V. ; Eskildsen, Simon F. ; Coupé, Pierrick ; Romero, José E. ; Collins, D. Louis ; Robles, Montserrat : Nonlocal intracranial cavity extraction. Dans : *Journal of Biomedical Imaging* 2014 (2014), pp. 10 42, 44

[Manjón et Coupé 2016] Manjón, José V. ; Coupé, Pierrick : volBrain: An online MRI brain volumetry system. Dans : *Frontiers in neuroinformatics* 10 (2016) 61, 62, 197

[Martin et al. 2001] Martin, David ; Fowlkes, Charless ; Tal, Doron ; Malik, Jitendra : A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)* 2, 2001, pp. 416–423 92, 98, 109, 136

[Martin et al. 2004] Martin, David R. ; Fowlkes, Charless C. ; Malik, Jitendra : Learning to detect natural image boundaries using local brightness, color, and texture cues. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 26 (2004), n. 5, pp. 530–549 87, 94

[Matsakis et Wendling 1999] Matsakis, Pascal ; Wendling, Laurent : A new way to represent the relative position between areal objects. Dans : *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)* 21 (1999), n. 7, pp. 634–643 46, 155

[Mazziotta et al. 1995] Mazziotta, John C. ; Toga, Arthur W. ; Evans, Alan ; Fox, Peter ; Lancaster, Jack : A probabilistic atlas of the human brain: Theory and rationale

for its development. Dans : *NeuroImage (NIMG)* 2 (1995), n. 2, pp. 89–101 21, 22, 50, 51

[Menze et al. 2015] Menze, Bjoern H. ; Jakab, Andras ; Bauer, Stefan ; Kalpathy-Cramer, Jayashree ; Farahani, Keyvan ; Kirby, Justin ; Burren, Yuliya ; Porz, Nicole ; Slotboom, Johannes ; Wiest, Roland et al. : The multimodal brain tumor image segmentation benchmark (BRATS). Dans : *IEEE Trans. on Medical Imaging (T-MI)* 34 (2015), n. 10, pp. 1993–2024 119, 120, 201

[Moore et al. 2008] Moore, Alastair P. ; Prince, Simon J. ; Warrell, Jonathan ; Mohammed, Umar ; Jones, Graham : Superpixel lattices. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8 83, 86, 88

[Muja et Lowe 2009] Muja, Marius ; Lowe, David G. : Fast approximate nearest neighbors with automatic algorithm configuration. Dans : *Proc. of Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)* 2 (2009), n. 2, pp. 331–340 30, 32

[Munkres 1957] Munkres, James : Algorithms for the assignment and transportation problems. Dans : *SIAM Journal on Imaging Sciences (SIIMS)* 5 (1957), n. 1, pp. 32–38 184, 185

[Nakamura et Hong 2017] Nakamura, Kensuke ; Hong, Byung-Woo : Hierarchical image segmentation via recursive superpixel with adaptive regularity. Dans : *Journal of Electronic Imaging (JEI)* 26 (2017), n. 6 80

[Neubert et Protzel 2012] Neubert, Peer ; Protzel, Peter : Superpixel benchmark and comparison. Dans : *Forum Bildverarbeitung*, 2012, pp. 1–12 81, 86, 89

[Neubert et Protzel 2014] Neubert, Peer ; Protzel, Peter : Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms. Dans : *Proc. of International Conference on Pattern Recognition (ICPR)*, 2014, pp. 996–1001 84, 85, 86, 116, 124, 198

[Newson et al. 2013] Newson, Alasdair ; Almansa, Andrés ; Fradet, Matthieu ; Gousseau, Yann ; Pérez, Patrick : Towards fast, generic video inpainting. Dans : *European Conference on Visual Media Production (CVMP)*, 2013, pp. 7 31

[Nguyen et al. 2014] Nguyen, Rang M. ; Kim, Seon J. ; Brown, Michael S. : Illuminant aware gamut-based color transfer. Dans : *Comput. Graph. Forum* 33, 2014, pp. 319–328 180, 188, 190

[Okugawa et al. 2002] Okugawa, Gaku ; Sedvall, Göran ; Nordström, Mikael ; Andreasen, Nancy ; Pierson, Ronald ; Magnotta, Vincent ; Agartz, Ingrid : Selective reduction of the posterior superior vermis in men with chronic schizophrenia. Dans : *Schizophrenia Research* 55 (2002), n. 1, pp. 61–67 62

[Olonetsky et Avidan 2012] Olonetsky, Igor ; Avidan, Shai : TreeCANN-kd tree coherence approximate nearest neighbor algorithm. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2012, pp. 602–615 30

[Pant et al. 2015] Pant, Abinash ; Rivest-Hénault, David ; Bourgeat, Pierrick : Efficient multi-scale patch-based segmentation. Dans : *Proc. of Int. Work. on Patch-based Techniques in Medical Imaging (Patch-MI, MICCAI)*, 2015, pp. 205–213 45, 58

[Papadakis et al. 2011] Papadakis, Nicolas ; Provenzi, Edoardo ; Caselles, Vicent : A variational model for histogram transfer of color images. Dans : *IEEE Trans. on Image Processing (TIP)* 20 (2011), n. 6, pp. 1682–1695 180, 188, 190

[Park et al. 2014] Park, Min Tae M. ; Pipitone, Jon ; Baer, Lawrence H. ; Winterburn, Julie L. ; Shah, Yashvi ; Chavez, Sofia ; Schira, Mark M. ; Lobaugh, Nancy J. ; Lerch, Jason P. ; Voineskos, Aristotle N. et al. : Derivation of high-resolution MRI atlases of the human cerebellum at 3T and segmentation using multiple automatically generated templates. Dans : *NeuroImage (NIMG)* 95 (2014), pp. 217–231 62, 64, 65

[Pei et al. 2014] Pei, Soo-Chang ; Chang, Wen-Wen ; Shen, Chih-Tsung : Saliency detection using superpixel belief propagation. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 1135–1139 155, 169, 198

[Pereira et al. 2016] Pereira, Sérgio ; Pinto, Adriano ; Alves, Victor ; Silva, Carlos A. : Brain tumor segmentation using convolutional neural networks in MRI images. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 35 (2016), n. 5, pp. 1240–1251 21

[Pérez-Rúa et al. 2016] Pérez-Rúa, Juan-Manuel ; Crivelli, Tomas ; Bouthemy, Patrick ; Pérez, Patrick : Determining occlusions from space and time image reconstructions. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1382–1391 7

[Peyré et al. 2010] Peyré, Gabriel ; Péchaud, Mickael ; Keriven, Renaud ; Cohen, Laurent D. : Geodesic methods in computer vision and graphics. Dans : *Foundations and Trends in Computer Graphics and Vision* 5 (2010), n. 3–4, pp. 197–397 102

[Pierre et al. 2015] Pierre, Fabien ; Aujol, Jean-François ; Bugeau, Aurélie ; Ta, Vinh-Thong : Luminance-hue specification in the RGB space. Dans : *Proc. of Int. Conf. on Scale Space and Variational Methods in Computer Vision (SSVM)*, 2015, pp. 413–424 31

[Pipitone et al. 2014] Pipitone, Jon ; Park, Min Tae M. ; Winterburn, Julie ; Lett, Tristram A. ; Lerch, Jason P. ; Pruessner, Jens C. ; Lepage, Martin ; Voineskos, Aristotle N. ; Chakravarty, Mallar et al. : Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. Dans : *NeuroImage (NIMG)* 101 (2014), pp. 494–512 60

[Pitié et al. 2005a] Pitié, François ; Kokaram, Anil C. ; Dahyot, Rozenn : N-dimensional probability density function transfer and its application to color transfer. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)* 2, 2005, pp. 1434–1439 180

[Pitié et al. 2005b] PITIÉ, François ; KOKARAM, Anil C. ; DAHYOT, Rozenn : Towards automated colour grading. Dans : *European Conference on Visual Media Production (CVMP)*, 2005 186, 187

[Pitié et al. 2007] PITIÉ, François ; KOKARAM, Anil C. ; DAHYOT, Rozenn : Automated colour grading using colour distribution transfer. Dans : *Computer Vision and Image Understanding (CVIU)* 107 (2007), n. 1–2, pp. 123–137 180, 181, 188, 190

[Prasad et al. 2015] PRASAD, Gautam ; JOSHI, Shantanu H. ; NIR, Talia M. ; TOGA, Arthur W. ; THOMPSON, Paul M. et al. : Brain connectivity and novel network measures for Alzheimer's disease classification. Dans : *Neurobiology of Aging* 36 (2015), pp. S121–S131 69

[Qazi et al. 2011] QAZI, Imtnan-Ul-Haque ; ALATA, Olivier ; BURIE, Jean-Christophe ; MOUSSA, Ahmed ; FERNANDEZ-MALOIGNE, Christine : Choice of a pertinent color space for color texture characterization using parametric spectral analysis. Dans : *Pattern Recognition (PR)* 44 (2011), n. 1, pp. 16–31 85

[Rabin et Peyré 2011] RABIN, Julien ; PEYRÉ, Gabriel : Wasserstein regularization of imaging problem. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2011, pp. 1541–1544 180

[Rabin et al. 2014] RABIN, Julien ; FERRADANS, Sira ; PAPADAKIS, Nicolas : Adaptive color transfer with relaxed optimal transport. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2014 154, 180, 181

[Ramakanth et Babu 2012] RAMAKANTH, S. Avinash ; BABU, R. Venkatesh : Feature match: An efficient low dimensional PatchMatch technique. Dans : *Proc. of Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2012, pp. 45 31, 37

[Reinhard et al. 2001] REINHARD, Erik ; ADHIKHMIN, Michael ; GOOCH, Bruce ; SHIRLEY, Peter : Color transfer between images. Dans : *IEEE Computer Graphics and Applications (CGA)* 21 (2001), n. 5, pp. 34–41 180, 181

[Ren et Malik 2003] REN, Xiaofeng ; MALIK, Jitendra : Learning a classification model for segmentation. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2003, pp. 10–17 7, 81, 82

[Reso et al. 2013] RESO, Matthias ; JACHALSKY, Jörn ; ROSENHAHN, Bodo ; OSTERMANN, Jörn : Temporally consistent superpixels. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 385–392 79, 81, 119, 124, 127, 140, 154, 198

[Revaud et al. 2016] REVAUD, Jerome ; WEINZAEPFEL, Philippe ; HARCHAOUI, Zaid ; SCHMID, Cordelia : Deepmatching: Hierarchical deformable dense matching. Dans : *International Journal of Computer Vision (IJCV)* 120 (2016), n. 3, pp. 300–323 31

[Roche et al. 2014] Roche, Florent ; Schaerer, Joël ; Gouttard, Sylvain ; Istace, Audrey ; Belaroussi, Boubakeur ; Yu, Hui J. ; Bracoud, Luc ; Pachai, Chahin ; DeCarli, Charles : Accuracy of BMAS hippocampus segmentation using the harmonized hippocampal protocol. Dans : *Alzheimer's & Dementia* 10 (2014), n. 4, pp. 56 58

[Rohlfing et al. 2004] Rohlfing, Torsten ; Brandt, Robert ; Menzel, Randolf ; Maurer, Calvin R. : Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Dans : *NeuroImage (NIMG)* 21 (2004), n. 4, pp. 1428–1442 23, 25

[Romero et al. 2017] Romero, Jose E. ; Coupé, Pierrick ; Giraud, Rémi ; Ta, Vinh-Thong ; Fonov, Vladimir ; Park, Min Tae M. ; Chakravarty, M. Mallar ; Voineskos, Aristotle N. ; Manjón, José V. : CERES: A new cerebellum lobule segmentation method. Dans : *NeuroImage (NIMG)* 147 (2017), pp. 916–924 11, 63

[Rousseau et al. 2011] Rousseau, Françcois ; Habas, Piotr A. ; Studholme, Colin : A supervised patch-based approach for human brain labeling. Dans : *Medical Imaging* 30 (2011), n. 10, pp. 1852–1862 24, 42, 44

[Roussillon et al. 2010] Roussillon, Tristan ; Sivignon, Isabelle ; Tougne, Laure : Measure of circularity for parts of digital boundaries and its fast computation. Dans : *Pattern Recognition (PR)* 43 (2010), n. 1, pp. 37–46 131

[Rubio et al. 2016] Rubio, Antonio ; Yu, LongLong ; Simo-Serra, Edgar ; Moreno-Noguer, Francesc : BASS: Boundary-aware superpixel segmentation. Dans : *Proc. of International Conference on Pattern Recognition (ICPR)*, 2016 85, 102, 103, 119, 125, 132

[Sabuncu et al. 2010] Sabuncu, Mert R. ; Yeo, B. T. Thomas ; Van Leemput, Koen ; Fischl, Bruce ; Golland, Polina : A generative model for image segmentation based on label fusion. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 29 (2010), n. 10, pp. 1714–1729 23

[Sawhney et al. 2014] Sawhney, Rahul ; Li, Fuxin ; Christensen, Henrik I. : GASP: Geometric association with surface patches. Dans : *Proc. of Int. Conf. on 3D Vision (3DV)*, 2014 155, 157

[Schick et al. 2012] Schick, Alexander ; Fischer, Mika ; Stiefelhagen, Rainer : Measuring and evaluating the compactness of superpixels. Dans : *Proc. of International Conference on Pattern Recognition (ICPR)*, 2012, pp. 930–934 86, 87, 124, 126

[Schick et al. 2014] Schick, Alexander ; Fischer, Mika ; Stiefelhagen, Rainer : An evaluation of the compactness of superpixels. Dans : *Pattern Recognition Letters* 43 (2014), pp. 71–80 127

[Sharma et al. 2013] Sharma, Gaurav ; Jurie, Frédéric ; Schmid, Cordelia : Expanded parts model for human attribute and action recognition in still images. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 652–659 46, 155

[Shi et Malik 2000] Shi, Jianbo ; Malik, Jitendra : Normalized cuts and image segmentation. 22 (2000), n. 8, pp. 888–905 82

[Shi et al. 2013] Shi, Wenzhe ; Caballero, Jose ; Ledig, Christian ; Zhuang, Xiahai ; Bai, Wenjia ; Bhatia, Kanwal ; Marvao, Antonio de ; Dawes, Tim ; O'Regan, Declan ; Rueckert, Daniel : Cardiac image super-resolution with global correspondence using multi-atlas PatchMatch. Dans : *Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2013, pp. 9–16 36, 164

[Shotton et al. 2009] Shotton, Jamie ; Winn, John ; Rother, Carsten ; Criminisi, Antonio : Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. Dans : *International Journal of Computer Vision (IJCV)* 81 (2009), n. 1, pp. 2–23 198

[Sled et al. 1998] Sled, John G. ; Zijdenbos, Alex P. ; Evans, Alan C. : A nonparametric method for automatic correction of intensity nonuniformity in MRI data. Dans : *Medical Imaging* 17 (1998), n. 1, pp. 87–97 51

[Snoek et al. 2005] Snoek, Cees G. M. ; Worring, Marcel ; Smeulders, Arnold W. M. : Early versus late fusion in semantic video analysis. Dans : *Proc. of ACM Int. Conf. on Multimedia*, 2005, pp. 399–402 44, 46

[Strachan et al. 1990] Strachan, Norval J. C. ; Nesvadba, Paul ; Allen, Alastair R. : Fish species recognition by shape analysis of images. Dans : *Pattern Recognition (PR)* 23 (1990), n. 5, pp. 539–544 133

[Strassburg et al. 2015] Strassburg, Johann ; Grzeszick, Rene ; Rothacker, Leonard ; Fink, Gernot A. : On the influence of superpixel methods for image parsing. Dans : *Proc. of Int. Conf. on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2015, pp. 518–527 124, 138

[Stutz et al. 2016] Stutz, David ; Hermans, Alexander ; Leibe, Bastian : Superpixels: An evaluation of the state-of-the-art. Dans : *Computer Vision and Image Understanding (CVIU)* (2016) 81, 83, 88, 90, 93, 116, 125, 127

[Szegedy et al. 2013] Szegedy, Christian ; Zaremba, Wojciech ; Sutskever, Ilya ; Bruna, Joan ; Erhan, Dumitru ; Goodfellow, Ian ; Fergus, Rob : Intriguing properties of neural networks. Dans : *arXiv preprint arXiv:1312.6199* (2013) 5

[Ta et al. 2014] Ta, Vinh-Thong ; Giraud, Rémi ; Collins, D. Louis ; Coupé, Pierrick : Optimized PatchMatch for near real time and accurate label fusion. Dans : *Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2014, pp. 105–112 11

[Tai et al. 2005] Tai, Yu-Wing ; Jia, Jiaya ; Tang, Chi-Keung : Local color transfer via probabilistic segmentation by expectation-maximization. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 747–754 180

[Tangaro et al. 2014] Tangaro, Sabina ; Amoroso, Nicola ; Boccardi, Marina ; Bruno, Stefania ; Chincarini, Andrea ; Ferraro, Giovanni ; Frisoni, Giovanni B. ; Maglietta, Rosalia ; Redolfi, Alberto et al. : Automated voxel-by-voxel tissue classification for hippocampal segmentation: Methods and validation. Dans : *Physica Medica* 30 (2014), n. 8, pp. 878–887 51, 58, 59

[Tanimoto et Pavlidis 1975] Tanimoto, Steven ; Pavlidis, Theo : A hierarchical data structure for picture processing. Dans : *Computer Graphics and Image Processing (CGIP)* 4 (1975), n. 2, pp. 104–119 79, 132, 133

[Tapia 2011] Tapia, Ernesto : A note on the computation of high-dimensional integral images. Dans : *Pattern Recognition Letters* 32 (2011), n. 2, pp. 197–201 40

[Thomann et al. 2008] Thomann, Philipp A. ; Schläfer, Christine ; Seidl, Ulrich ; Dos Santos, Vasco ; Essig, Marco ; Schröder, Johannes : The cerebellum in mild cognitive impairment and Alzheimer's disease–a structural MRI study. Dans : *Journal of psychiatric research* 42 (2008), n. 14, pp. 1198–1202 62

[Tibshirani 1996] Tibshirani, Robert : Regression shrinkage and selection via the lasso. Dans : *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), pp. 267–288 199

[Tighe et Lazebnik 2010] Tighe, Joseph ; Lazebnik, Svetlana : SuperParsing: Scalable nonparametric image parsing with superpixels. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2010, pp. 352–365 79

[Tong et al. 2002] Tong, Xin ; Zhang, Jingdan ; Liu, Ligang ; Wang, Xi ; Guo, Baining ; Shum, Heung-Yeung : Synthesis of bidirectional texture functions on arbitrary surfaces. Dans : *ACM Trans. on Graphics (ToG)* 21, 2002, pp. 665–672 31

[Tong et al. 2013] Tong, Tong ; Wolz, Robin ; Coupé, Pierrick ; Hajnal, Joseph V. ; Rueckert, Daniel et al. : Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. Dans : *NeuroImage (NIMG)* 76 (2013), pp. 11–23 21, 24, 25, 53, 58

[Tong et al. 2014] Tong, Na ; Lu, Huchuan ; Zhang, Lihe ; Ruan, Xiang : Saliency detection with multi-scale superpixels. Dans : *IEEE Signal Processing Letters* 21 (2014), n. 9, pp. 1035–1039 80

[Trulls et al. 2014] Trulls, Eduard ; Tsogkas, Stavros ; Kokkinos, Iasonas ; Sanfeliu, Alberto ; Moreno-Noguer, Francesc : Segmentation-aware deformable part models. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 168–175 46, 155, 198

[Tsai et al. 2012] Tsai, David ; Flagg, Matthew ; Nakazawa, Atsushi ; Rehg, James M. : Motion coherent tracking using multi-label MRF optimization. Dans : *International Journal of Computer Vision (IJCV)* 100 (2012), n. 2, pp. 190–202 140, 141, 142

[Tustison et al. 2010] Tustison, Nicholas J. ; Avants, Brian B. ; Cook, Philip A. ; Zheng, Yuanjie ; Egan, Alexander ; Yushkevich, Paul A. ; Gee, James C. : N4ITK: improved N3 bias correction. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 29 (2010), n. 6, pp. 1310–1320 51

[Van den Bergh et al. 2012] Van den Bergh, Michael ; Boix, Xavier ; Roig, Gemma ; de Capitani, Benjamin ; Van Gool, Luc : SEEDS: Superpixels extracted via energy-driven sampling. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2012, pp. 13–26 81, 83, 89, 90, 92, 94, 100, 110, 114, 125, 136, 144

[Vedaldi et Soatto 2008] Vedaldi, Andrea ; Soatto, Stefano : Quick shift and kernel methods for mode seeking. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2008, pp. 705–718 81, 84

[Veksler et al. 2010] Veksler, Olga ; Boykov, Yuri ; Mehrani, Paria : Superpixels and supervoxels in an energy optimization framework. Dans : *Proc. of European Conference on Computer Vision (ECCV)*, 2010, pp. 211–224 83, 85

[Vincent et Soille 1991] Vincent, Luc ; Soille, Pierre : Watersheds in digital spaces: An efficient algorithm based on immersion simulations. 13 (1991), n. 6, pp. 583–598 84

[Wachinger et al. 2014] Wachinger, Christian ; Brennan, Matthew ; Sharp, Greg C. ; Golland, Polina : On the importance of location and features for the patch-based segmentation of parotid glands. Dans : *Image-guided adaptive radiation therapy (IGART)* (2014) 44, 45

[Wang et al. 2011] Wang, Shu ; Lu, Huchuan ; Yang, Fan ; Yang, Ming-Hsuan : Superpixel tracking. Dans : *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1323–1330 23, 79, 81, 119, 140, 154

[Wang et al. 2012] Wang, Xiao-Hui ; Jia, Jia ; Liao, Han-Yu ; Cai, Lian-Hong : Affective image colorization. Dans : *Journal of Computer Science and Technology* 27 (2012), n. 6, pp. 1119 31

[Wang et al. 2013] Wang, Peng ; Zeng, Gang ; Gan, Rui ; Wang, Jingdong ; Zha, Hongbin : Structure-sensitive superpixels via geodesic distance. Dans : *International Journal of Computer Vision (IJCV)* 103 (2013), n. 1, pp. 1–21 102, 142

[Wang et al. 2014] Wang, Li ; Shi, Feng ; Li, Gang ; Gao, Yaozong ; Lin, Weili ; Gilmore, John H. ; Shen, Dinggang : Segmentation of neonatal brain MR images using patch-driven level sets. Dans : *NeuroImage (NIMG)* 84 (2014), pp. 141–158 21, 24

[Wang et al. 2017a] Wang, Hai ; Peng, Xiongyou ; Xiao, Xue ; Liu, Yan : BSLIC: SLIC superpixels based on boundary term. Dans : *Symmetry* 9 (2017), n. 3, pp. 31 85, 105

[Wang et Yushkevich 2013] Wang, Hongzhi ; Yushkevich, Paul A. : Multi-atlas segmentation without registration: A supervoxel-based approach. Dans : *Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2013), pp. 535–542 80, 201, 202

[Wang et al. 2017b] Wang, Murong ; Liu, Xiabi ; Gao, Yixuan ; Ma, Xiao ; Soomro, Nouman Q. : Superpixel segmentation: A benchmark. Dans : *Signal Processing: Image Communication* 56 (2017), pp. 28–39 124

[Wei et Levoy 2000] Wei, Li-Yi ; Levoy, Marc : Fast texture synthesis using tree-structured vector quantization. Dans : *Proc. of Int. Conf. on Computer graphics and interactive techniques (SIGGRAPH)*, 2000, pp. 479–488 30

[Weier et al. 2014] Weier, Katrin ; Fonov, Vladimir ; Lavoie, Karyne ; Doyon, Julien ; Collins, D. Louis : Rapid automatic segmentation of the human cerebellum and its lobules (RASCAL)—Implementation and application of the patch-based label-fusion technique with a template library to segment the human cerebellum. Dans : *Human Brain Mapping (HBM)* 35 (2014), n. 10, pp. 5026–5039 64, 65

[Weiskopf et al. 2011] Weiskopf, Nikolaus ; Lutti, Antoine ; Helms, Gunther ; Novak, Marianne ; Ashburner, John ; Hutton, Chloe : Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). Dans : *NeuroImage (NIMG)* 54 (2011), n. 3, pp. 2116–2124 51

[Wolz et al. 2009] Wolz, Robin ; Aljabar, Paul ; Rueckert, Daniel ; Heckemann, Rolf A. ; Hammers, Alexander : Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information. Dans : *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)* (2009), pp. 470–473 23

[Wu et al. 2014] Wu, Guorong ; Wang, Qian ; Zhang, Daoqiang ; Nie, Feiping ; Huang, Heng ; Shen, Dinggang : A generative probability model of joint label fusion for multi-atlas based brain segmentation. Dans : *Medical Image Analysis (MIA)* 18 (2014), n. 6, pp. 881–890 24, 25, 42, 44, 197

[Wu et al. 2015] Wu, Guorong ; Kim, Minjeong ; Sanroma, Gerard ; Wang, Qian ; Munsell, Brent C. ; Shen, Dinggang et al. : Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition. Dans : *NeuroImage (NIMG)* 106 (2015), pp. 34–46 44, 45

[Wyman et al. 2013] Wyman, Bradley T. ; Harvey, Danielle J. ; Crawford, Karen ; Bernstein, Matt A. ; Carmichael, Owen ; Cole, Patricia E. ; Crane, Paul K. ; DeCarli, Charles ; Fox, Nick C. ; Gunter, Jeffrey L. et al. : Standardization of analysis sets for reporting results from ADNI MRI data. Dans : *Alzheimer's & Dementia* 9 (2013), n. 3, pp. 332–337 60

[Xiao et al. 2011] Xiao, Chunxia ; Liu, Meng ; Yongwei, Nie ; Dong, Zhao : Fast exact nearest patch matching for patch-based image editing and processing. Dans : *IEEE Trans. on Visualization and Computer Graphics (TVCG)* 17 (2011), n. 8, pp. 1122–1134 40

[Xiao et Ma 2006] Xiao, Xuezhong ; Ma, Lizhuang : Color transfer in correlated color space. Dans : *ACM Int. Conf. on Virtual-Reality Continuum and its Applications in Industry (VRCAI)*, 2006, pp. 305–309 180

[Xiaofeng et Bo 2012] Xiaofeng, Ren ; Bo, Liefeng : Discriminatively trained sparse code gradients for contour detection. Dans : *Proc. of Int. Conf. on Neural Information Processing Systems (NIPS)*. 2012, pp. 584–592 113

[Xu et Corso 2012] Xu, Chenliang ; Corso, Jason J. : Evaluation of super-voxel methods for early video processing. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1202–1209 116

[Xu et Sun 2010] Xu, Zongben ; Sun, Jian : Image inpainting by patch propagation using patch sparsity. Dans : *IEEE Trans. on Image Processing (TIP)* 19 (2010), n. 5, pp. 1153–1165 27

[Yang et al. 2010] Yang, Yi ; Hallman, Sam ; Ramanan, Deva ; Fowlkes, Charless : Layered object detection for multi-class segmentation. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 3113–3120 79

[Yang et al. 2013] Yang, Chuan ; Zhang, Lihe ; Lu, Huchuan ; Ruan, Xiang ; Yang, Ming-Hsuan : Saliency detection via graph-based manifold ranking. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3166–3173 79

[Yao et al. 2015] Yao, Jian ; Boben, Marko ; Fidler, Sanja ; Urtasun, Raquel : Real-time coarse-to-fine topologically preserving segmentation. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2947–2955 81, 83, 84, 92, 93, 100, 114, 116, 124, 137, 145

[Zhang et al. 2011a] Zhang, Daoqiang ; Wang, Yaping ; Zhou, Luping ; Yuan, Hong ; Shen, Dinggang et al. : Multimodal classification of Alzheimer's disease and mild cognitive impairment. Dans : *NeuroImage (NIMG)* 55 (2011), n. 3, pp. 856–867 21

[Zhang et al. 2011b] Zhang, Yuhang ; Hartley, Richard ; Mashford, John ; Burn, Stewart : Superpixels, occlusion and stereo. Dans : *Proc. of Int. Conf. on Digital Image Computing Techniques and Applications (DICTA)*, 2011, pp. 84–91 7, 79

[Zhang et al. 2016] Zhang, Yongxia ; Li, Xuemei ; Gao, Xifeng ; Zhang, Caiming : A simple algorithm of superpixel segmentation with boundary constraint. Dans : *IEEE Trans. on Circuits and Systems for Video Technology (TCSVT)* (2016), n. 99 85, 93, 127, 129

[Zheng et Li 2015] Zheng, Jinghong ; Li, Zhengguo : Superpixel based patch match for differently exposed images with moving objects and camera movements. Dans : *Proc. of IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 4516–4520 155

[Zhu et al. 2014] Zhu, Wangjiang ; Liang, Shuang ; Wei, Yichen ; Sun, Jian : Saliency optimization from robust background detection. Dans : *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2814–2821 85

[Zhu et al. 2017] Zhu, Hancan ; Cheng, Hewei ; Yang, Xuesong ; Fan, Yong et al. : Metric learning for multi-atlas based segmentation of hippocampus. Dans : *International Journal of Computer Vision (IJCV)* 15 (2017), n. 1, pp. 41–50 58, 59

[Zijdenbos et al. 1994] Zijdenbos, Alex P. ; Dawant, Benoit M. ; Margolin, Richard A. ; Palmer, Andrew C. : Morphometric analysis of white matter lesions in MR images: method and validation. Dans : *IEEE Trans. on Medical Imaging (T-MI)* 13 (1994), n. 4, pp. 716–724 52

Liste des acronymes

ACC ACCuracy 69, 70

AD Alzheimer's Disease 66–70

ADNI Alzheimer's Disease Neuroimaging Initiative 50, 51, 61, 65, 66

ANN Approximate Nearest Neighbor 6, 7, 9–13, 16, 26, 30–32, 36–39, 41–46, 50, 53, 54, 56, 71, 148, 152, 153, 157, 159–162, 166–169, 172, 174, 175, 179–186, 192, 194, 196–199, 201, 202

ASA Achievable Segmentation Accuracy 87, 90–94, 110, 111, 113–115, 118, 133–135, 138, 139, 143, 145, 172–174

AUC Area Under Curve 67, 69, 70

BMAS BioClinica Multi-Atlas Segmentation 58, 59

BR Boundary Recall 87, 93, 94, 110, 111, 113–115, 142, 143, 145

BSD Berkeley Segmentation Dataset 92, 98, 109–111, 114–117

CD Contour Density 93, 94, 110, 111, 113–115

CERES CEREbellum Segmentation 62, 64, 65

DDLS Discriminative Dictionary Learning for Segmentation 58, 59

DTI Diffusion Tensor Imaging 68, 69

EADC-ADNI European Alzheimer's Disease Consortium and Alzheimer's Disease Neuroimaging Initiative 51, 53–61, 67, 70

EV Explained Variation 86, 88, 89, 94, 110, 111, 114, 115, 133–135, 141, 143, 145

GR Global Regularity 124, 133–138, 140, 142–145, 173, 174, 194

HarP Harmonized Protocol 51, 67

ICBM International Consortium for Brain Mapping 50, 51, 53–56, 58, 59, 70

ICV Intra-Cluster Variation 86, 88, 89, 94

IRM Imagerie par Résonance Magnétique 3, 11, 13, 16, 20–22, 32, 36, 39, 50, 51, 53, 55, 56, 61, 63, 65, 68, 70, 118, 192, 197, 199

LEAP Learning Embeddings for Atlas Propagation 58–60

LFW Labeled Faces in the Wild 166–168, 172–175, 196

MCI Mild Cognitive Impairment 66–68, 70

MF Mismatch Factor 87, 136

MSE Mean Square Error 141, 143, 145

MTM Multi-Template Method 56, 58, 59

NC Normal Control 66–70

NLW-ML Non Local patch-based Weighted voting with Metric Learning 58, 59

OPAL Optimized PatchMatch for Accurate Label fusion 16, 17, 36, 37, 41–46, 50, 52, 53, 56–64, 66–71, 192, 194, 195, 197, 199, 200

OPM Optimized PatchMatch 37, 38, 40–42, 45, 53

PB-MS Patch-Based Multi-Scale 58, 59

PBL Patch-Based Label fusion 23–26, 28, 32, 56, 58

PM PatchMatch 148, 152, 153, 159, 162, 180

PR Précision Rappel 94, 110–112, 114

RF Random Forest 58, 59

SC Sparse Coding 56, 59

SCALP Superpixels with Contour Adherence using Linear Path 8, 74, 75, 78, 79, 81, 82, 87, 92–94, 98–106, 108–119, 122–124, 128, 130, 131, 133–136, 138, 141–143, 145, 157, 158, 172–174, 184, 185, 193–196, 201

SCT Superpixel-based Color Transfer 179, 184–188, 194

SLIC Simple Linear Iterative Clustering 81, 82, 84, 85, 87, 92, 98–100, 104, 110, 113, 115–118, 135, 143

SMF Smooth Matching Factor 132, 133, 143, 145

SPM SuperPatchMatch 159–162, 166–168, 173, 174, 183, 200

SRC Shape Regularity Criteria 126–130, 132, 133, 143, 145

SSD Sum of Squared Differences 26, 28, 39, 154, 156

UE Undersegmentation Error 86, 89, 90, 92, 143, 145

Table des matières

Remerciements **v**

Introduction générale **2**

I Algorithme de correspondance basé sur les patchs pour la segmentation d'images médicales **15**

1 Segmentation d'images médicales et algorithmes de correspondance **19**
1.1 Contexte et pathologies neurodégénératives 20
1.2 Segmentation d'images IRM 3D 21
1.2.1 Approches par recalage non linéaire 21
1.2.2 Approche par fusion d'étiquettes basée sur les patchs 23
1.3 Algorithmes de correspondance à partir de patchs 26
1.3.1 Introduction des méthodes basées sur les patchs 26
1.3.2 Algorithmes de recherche de correspondances 30
1.4 Conclusion 32

2 La méthode OPAL **35**
2.1 Introduction 36
2.2 Algorithme PatchMatch optimisé 36
2.2.1 L'Algorithme PatchMatch 36
2.2.2 Initialisation contrainte 38
2.2.3 Propagation optimisée 40
2.2.4 Recherche aléatoire contrainte 40
2.2.5 PM multiples et calcul parallèle 41
2.3 Fusion d'étiquettes basées sur les patchs et segmentation 42
2.3.1 Fusion d'étiquettes basée sur les patchs 42
2.3.2 Filtrage spatial 44
2.4 Agrégation d'estimateurs multi-échelles et multi-descripteurs 44
2.4.1 Estimateurs multi-échelles 45
2.4.2 Estimateurs multi-descripteurs 46
2.4.3 Agrégation a posteriori d'estimateurs 46
2.5 Conclusion 46

3 Résultats de segmentation d'images médicales **49**
3.1 Introduction 50
3.2 Validation d'OPAL sur la segmentation de l'hippocampe 50
3.2.1 Schéma de validation 50
3.2.2 Influence des paramètres 53
3.2.3 Comparaison aux méthodes de l'état-de-l'art 56
3.2.4 Ajout de segmentations automatiques à la bibliothèque 59
3.2.5 Transfert 61
3.3 Segmentation du cervelet 62
3.3.1 Contexte 62
3.3.2 Amélioration de la fusion des estimateurs 63
3.3.3 Comparaison à l'état-de-l'art 64
3.4 Extension à la classification de la maladie d'Alzheimer 65
3.4.1 Classification depuis le volume des segmentations automatiques 66
3.4.2 Classification depuis la fusion d'étiquettes de groupe 68
3.5 Conclusion 70

II Décomposition en superpixels réguliers **73**

4 Algorithmes de superpixels et étude des métriques standards **77**
4.1 Utilisation des superpixels 78
4.2 Méthodes de décomposition en superpixels 81
4.2.1 Méthodes de graphes 82
4.2.2 Méthodes hiérarchiques 83
4.2.3 Méthodes de montée de gradient 84
4.3 Étude des métriques de superpixels 86
4.3.1 Homogénéité du groupement sur la couleur 87
4.3.2 Respect des objets de l'image 89
4.4 Conclusion 94

5 La méthode SCALP **97**
5.1 Contributions 98
5.2 Décomposition en superpixels avec SCALP 98
5.2.1 Schéma de groupement itératif 99
5.2.2 Information de couleur du voisinage 101
5.2.3 Information de couleur et contour sur le chemin linéaire 102
5.3 Contrainte depuis une segmentation initiale 106
5.3.1 Segmentation hiérarchique depuis une détection de contours . 107
5.3.2 Décomposition régulière en superpixels depuis une segmentation hiérarchique 107
5.4 Validation de SCALP sur les métriques superpixels standards 109
5.4.1 Cadre de validation 109

5.4.2 Impact des contributions 110
5.4.3 Comparaison avec les méthodes de l'état-de-l'art 114
5.4.4 Extension aux supervoxels 116
5.5 Conclusion 119

6 Évaluation de la régularité 123
6.1 Introduction 124
6.2 Mesure de régularité globale 126
6.2.1 Régularité de forme 126
6.2.2 Consistance de forme 133
6.2.3 Mesure de régularité globale 135
6.3 Comparaison de SCALP à l'état-de-l'art sur GR 136
6.4 Impact de la régularité sur des applications basées sur les superpixels 138
6.4.1 Correspondance de plus proches voisins approchés 138
6.4.2 Suivi de superpixels 140
6.4.3 Compression d'image 142
6.4.4 Détection de contours et segmentation 144
6.4.5 Corrélation entre GR et performances 145
6.5 Conclusion 147

III Algorithme de correspondance basé sur les superpixels et applications 149

7 La méthode SuperPatchMatch 153
7.1 Introduction 154
7.2 Définition du superpatch 156
7.3 Comparaison de superpatchs 156
7.3.1 Distance entre deux superpatchs 156
7.3.2 Distance rapide entre deux superpixels 157
7.3.3 Généralisation des patchs réguliers de pixels 158
7.3.4 Robustesse du superpatch 159
7.4 SuperPatchMatch 161
7.4.1 Étape d'initialisation 161
7.4.2 Étape de propagation 162
7.4.3 Étape de recherche aléatoire 163
7.4.4 Bibliothèque d'images d'exemples 164
7.5 Conclusion 164

8 Application à la segmentation et l'étiquetage basés sur les superpixels 167
8.1 Introduction 168
8.2 Schéma de fusion d'étiquettes 168
8.3 Validation de SuperPatchMatch sur l'étiquetage de visages 169

8.3.1 Réglage des paramètres 170
8.3.2 Influence des paramètres 170
8.3.3 Comparaison avec les méthodes de l'état-de-l'art 175
8.4 Conclusion 177

9 Application au transfert de couleurs 179
9.1 Introduction 180
9.2 SuperPatchMatch contraint 182
9.2.1 Contrainte sur la diversité des correspondances 182
9.2.2 Problème d'assignement optimal 184
9.3 Fusion des couleurs 184
9.4 Résultats de transfert de couleurs 186
9.4.1 Paramètres 186
9.4.2 Influence de la diversité des ANN 187
9.4.3 Comparaison avec l'état-de-l'art 188
9.5 Conclusion 188

Conclusion générale et perspectives 194
1 Conclusion générale 194
2 Résumé des contributions 196
3 Perspectives 197
3.1 Segmentation d'images médicales (OPAL) 197
3.2 Décomposition en superpixels (SCALP) 198
3.3 Méthodes d'ANN basés sur les superpixels (SuperPatchMatch) 198
3.4 Estimation de transformation non linéaire 199
3.5 Segmentation de tumeurs basée sur les supervoxels 201
3.6 Transfert de style 202

Liste des publications 206

Bibliographie 208

Liste des acronymes 230

Table des matières 233