

Segmentation hiérarchique d’images multimodales

Guillaume TOCHON¹, Mauro DALLA MURA¹, Jocelyn CHANUSSOT^{1,2}

¹GIPSA-lab

11 rue des Mathématiques, 38400 Saint Martin d’Hères, France

²Department of Electrical and Computer Engineering

University of Iceland, Sæmundargötu 2, 101 Reykjavik, Islande

prénom.nom@gipsa-lab.grenoble-inp.fr

Résumé – Les hiérarchies de partitions sont couramment utilisées pour segmentation d’images. Dans le cas d’images multimodales toutefois, la fusion de plusieurs hiérarchies reste un problème. Récemment, les tresses de partitions ont été proposées comme une possible solution à ce problème, mais n’ont jamais été implémentées dans un cas pratique. Nous proposons ainsi une nouvelle méthodologie, basée sur cette notion de tresse de partitions, pour effectuer la segmentation d’images multimodales. Cette méthode est appliquée dans un cas concret, à savoir la segmentation conjointe de données hyperspectrales et LiDAR. Les résultats obtenus confirment le potentiel de la méthode proposée.

Abstract – Hierarchies of partitions are widely used in the context of image segmentation. However, in the case of multimodal images, the fusion of multiple hierarchies remains a challenge. Recently, braids of partitions have been proposed as a possible solution to this issue, but have never been implemented in a practical case. In this paper, we propose a new methodology to achieve multimodal segmentation, based on this notion of braids of partitions. This new method is applied in a practical case, namely the joint segmentation of hyperspectral and LiDAR data. Obtained results confirm the potential of the proposed method.

1 Introduction

Les images multimodales (i.e., plusieurs images d’une même scène, mais acquises avec des capteurs différents) sont aujourd’hui monnaie courante dans le domaine du traitement de l’image. Cependant, un certain nombre de défis subsistent encore pour pouvoir exploiter pleinement les avantages apportés de telles images (tels que les corrélations et complémentarités entre chaque mode) [1]. La segmentation d’image est l’une des opérations qui bénéficierait sûrement du développement d’outils de traitement multimodaux, puisque son but est de diviser une image en régions qui ont du “sens” vis-à-vis d’un objectif donné. La segmentation d’une image multimodale devrait bénéficier de la complémentarité de ses différents modes pour garantir une meilleure délimitation de ses régions, en particulier quand celles-ci partagent certains attributs dans un mode donné, mais pas dans un autre.

Une image donnée peut bien souvent être segmentée à différents niveaux de détails, niveau qui dépend de l’objectif sous-jacent. Une solution potentielle à ce problème “pluri-échelle” est l’utilisation d’une hiérarchie de partitions, qui organise d’une manière emboîtée au sein de sa structure toutes les potentielles échelles d’intérêt. La hiérarchie peut dès lors être construite une seule fois, indépendamment de l’application, et son niveau d’exploration peut ensuite être ajusté pour produire la segmentation désirée. Dans [2] par exemple, ce réglage est basé sur un processus de minimisation d’énergie sur l’espace de toutes les partitions contenues dans la hiérarchie. L’échelle *optimale* dépend ainsi

de la définition de la fonction d’énergie. Cependant, la question demeure ouverte quant au cas d’images multimodales (et donc de plusieurs hiérarchies). Récemment, les tresses de partitions ont été introduites [4] comme une solution potentielle à ce problème. Nous proposons ainsi dans ce papier une méthode se basant sur les tresses de partitions pour segmenter conjointement plusieurs hiérarchies, permettant donc la segmentation d’images multimodales.

Dans la section 2, nous rappelons brièvement les travaux de [2,3] et [4] à propos de la minimisation d’énergie, respectivement sur des hiérarchies et tresses de partitions. Dans la section 3, nous décrivons la méthodologie pour accomplir la segmentation d’images multimodales. Dans section 4, la méthode proposée est appliquée dans un cas concret, à savoir la segmentation jointe de données hyperspectrales et LiDAR, et les résultats correspondants sont présentés. La conclusion et les perspectives sont données dans la section 5.

2 Minimisation d’énergie

2.1 Définitions et notations

Les mots segmentation et partition sont utilisés de manière interchangeable par la suite. Soit $\mathcal{I} : E \rightarrow V$, $E \subseteq \mathbb{Z}^2$ une image quelconque, dont les éléments (pixels) sont notés $\mathbf{x}_i \in E$. Une partition de E , notée π , est une collection de régions $\{\mathcal{R}_i \subseteq E\}$ de E telle que $\mathcal{R}_i \cap \mathcal{R}_{j \neq i} = \emptyset$ et $\bigcup_i \mathcal{R}_i = E$. L’ensemble

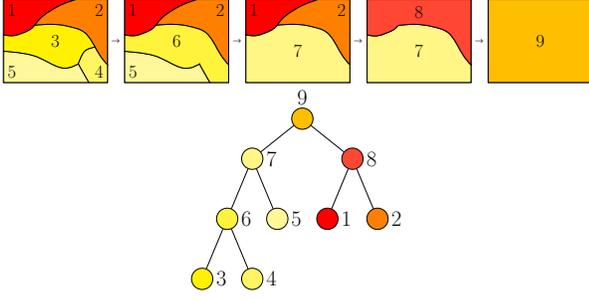


FIGURE 1 – Exemple de hiérarchie de partitions.

des partitions de E est noté Π_E . Pour deux partitions $\pi_i, \pi_j \in \Pi_E$, $\pi_i \leq \pi_j$ lorsque chaque région $\mathcal{R}_i \in \pi_i$ est incluse dans (ou égale à) une région $\mathcal{R}_j \in \pi_j$. Il est dit que π_i raffine π_j dans un tel cas. La minimisation d'une fonction d'énergie sur Π_E requiert la définition d'une énergie de région \mathcal{E} (qui assigne à chaque région $\mathcal{R} \subseteq E$ un nombre réel $\mathcal{E}(\mathcal{R}) \in \mathbb{R}^+$) et la définition d'une loi de composition \mathfrak{D} pour exprimer l'énergie de la partition comme étant une composition de l'énergie des régions la constituant.

$$\mathcal{E}(\pi = \{\mathcal{R}_i\}) = \mathfrak{D}_{\mathcal{R}_i \in \pi} \mathcal{E}(\mathcal{R}_i). \quad (1)$$

Cependant, trouver la partition optimale qui minimise une énergie donnée est une tâche difficile, principalement du fait de la grande cardinalité de Π_E . Les hiérarchies, en restreignant l'espace des partitions possibles, sont un outil attrayant pour réaliser la minimisation.

2.2 Minimisation sur une hiérarchie

Une hiérarchie de partitions de E est une collection $H = \{\mathcal{R} \subseteq E\}$ telle que $\emptyset \notin H$, $E \in H$ et $\forall \mathcal{R}_i, \mathcal{R}_j \in H$, $\mathcal{R}_i \cap \mathcal{R}_j \in \{\emptyset, \mathcal{R}_i, \mathcal{R}_j\}$ (n'importe quel couple de régions de la hiérarchie est soit disjoint, soit emboîté). Il est commun d'obtenir une décomposition hiérarchique d'une image en partant d'une certaine partition initiale π_0 et en fusionner les régions d'une manière itérative jusqu'à ce que le support entier de l'image soit atteint, produisant ainsi une séquence de partitions $\pi_0 \leq \pi_1 \leq \dots \leq \pi_n = \{E\}$ (voir figure 1). $\pi_n = \{E\}$ est appelée *racine* de la hiérarchie, et les régions de π_0 en sont les *feuilles*. Chaque autre région \mathcal{R} contient un ensemble de $S(\mathcal{R})$ régions enfants. Une *coupe* de H est une partition π de E dont les régions appartiennent à H . L'ensemble des coupes d'une hiérarchie H est noté $\Pi_E(H)$. $H(\mathcal{R})$ désigne la sous-hiérarchie de H de racine \mathcal{R} . Les coupes de la sous-hiérarchie $H(\mathcal{R})$ sont appelées *partitions partielles* de \mathcal{R} [4]. La coupe minimale de H vis-à-vis de l'énergie \mathcal{E} est définie comme :

$$\pi^* = \operatorname{argmin}_{\pi \in \Pi_E(H)} \mathcal{E}(\pi) \quad (2)$$

Les propriétés de \mathcal{E} pour lesquelles il est facile de retrouver la coupe minimale π^* ont été étudiées par [2] pour les énergies *séparables* (i.e., $\mathcal{E}(\pi) = \sum_{\mathcal{R} \in \pi} \mathcal{E}(\mathcal{R})$), par [7] pour les compositions par supremum (i.e., $\mathcal{E}(\pi) = \bigvee_{\mathcal{R} \in \pi} \mathcal{E}(\mathcal{R})$) puis

généralisées par [3] à des lois de compositions \mathfrak{D} plus générales, à savoir les *énergies h-croissantes* : \mathcal{E} est h-croissante lorsque pour deux partitions partielles quelconques $\pi_1(\mathcal{R})$ et $\pi_2(\mathcal{R})$ de $\mathcal{R} \in H$ telles que $\mathcal{E}(\pi_1(\mathcal{R})) \leq \mathcal{E}(\pi_2(\mathcal{R}))$, $\mathcal{E}(\pi_1(\mathcal{R}) \cup \mathcal{R}_0) \leq \mathcal{E}(\pi_2(\mathcal{R}) \cup \mathcal{R}_0)$ pour $\mathcal{R}_0 \in H$ disjoint de \mathcal{R} . Dans ce cas, la coupe minimale de H se trouve en résolvant pour chaque région \mathcal{R} le programme dynamique suivant :

$$\mathcal{E}^*(\mathcal{R}) = \min \left\{ \mathcal{E}(\mathcal{R}), \mathfrak{D}_{r \in S(\mathcal{R})} \mathcal{E}(\pi^*(r)) \right\} \quad (3)$$

$$\pi^*(\mathcal{R}) = \begin{cases} \{\mathcal{R}\} & \text{si } \mathcal{E}(\mathcal{R}) \leq \mathfrak{D}_{r \in S(\mathcal{R})} \mathcal{E}(\pi^*(r)) \\ \bigcup_{r \in S(\mathcal{R})} \pi^*(r) & \text{sinon} \end{cases} \quad (4)$$

La coupe optimale de \mathcal{R} est celle qui a la plus faible énergie entre $\{\mathcal{R}\}$ lui-même, ou l'union des coupes optimales de ses enfants. La coupe optimale de toute la hiérarchie est donnée par celle de sa racine, atteinte en scannant chaque région de la hiérarchie en une passe ascendante [2].

Les énergies dépendent souvent en pratique d'un certain paramètre réel λ qui agit comme un compromis entre simplicité (sous-segmentation) et fidélité aux données (sur-segmentation) de la segmentation. Ces énergies \mathcal{E}_λ génèrent des séquences de coupes optimales $\{\pi_\lambda^*\}$ quand λ varie. [3] a introduit la propriété de *croissance d'échelle* pour caractériser les variations de π_λ^* en fonction de λ : \mathcal{E}_λ est à croissance d'échelle si pour chaque $\mathcal{R} \in H$ et chacune de ses partitions partielles $\pi(\mathcal{R})$, l'inégalité $\mathcal{E}_{\lambda_1}(\mathcal{R}) \leq \mathcal{E}_{\lambda_1}(\pi(\mathcal{R}))$ ne change pas de sens pour $\lambda_2 \geq \lambda_1$. Dans le cas où l'énergie est h-croissante pour chaque λ et à croissance d'échelle par rapport à λ , la famille $\{\pi_\lambda^*\}$ de coupes optimales varie de manière hiérarchique, c'est-à-dire

$$\lambda_1 \leq \lambda_2 \Rightarrow \pi_{\lambda_1}^* \leq \pi_{\lambda_2}^*. \quad (5)$$

Dans un tel cas, il est possible d'extraire de H une sous-hiérarchie optimale H^* , générées à partir de coupes optimales de H .

2.3 Minimisation sur une tresse

Les tresses de partitions ont été introduites dans [4] comme suit : une famille de partitions $B = \{\pi_i \in \Pi_E\}$ est appelée *tresse* s'il existe une hiérarchie H_m , appelée *hiérarchie monotone*, telle que :

$$\forall \pi_i, \pi_j \in B, \pi_i \vee \pi_j \in \Pi_E(H_m) \setminus \{E\} \quad (6)$$

où $\pi_i \vee \pi_j$ est le suprémum par raffinement de π_i et π_j (la plus petite partition qui est raffinée par π_i et π_j). Les régions contenues dans une tresse ne sont pas forcément organisées de manière hiérarchique, mais celles de leurs suprémum par raffinement le sont. Pour cette raison, les tresses de partitions sont plus générales que les hiérarchies de partitions. Il est également important de noter que dans (6), le suprémum par raffinement de chaque couple de partitions doit différer de $\{E\}$. Autrement, n'importe quelle famille de partitions formerait une tresse avec $\{E\}$ comme suprémum, perdant ainsi toute structure intéressante. La coupe optimale d'une tresse de partitions est atteinte

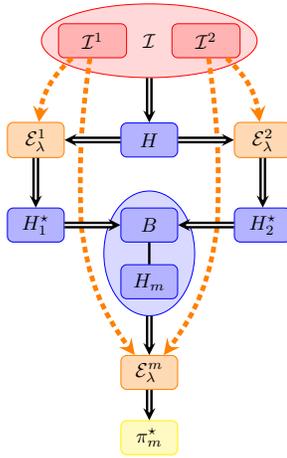


FIGURE 2 – Diagramme de la méthodologie de segmentation multimodale proposée.

en résolvant le programme dynamique (3) pour chaque région de la hiérarchie moniteur H_m , avec cependant une modification :

$$\mathcal{E}^*(\mathcal{R}) = \min \left\{ \mathcal{E}(\mathcal{R}), \bigvee_{r \in \mathcal{S}(\mathcal{R})} \mathcal{E}(\pi^*(r)), \bigwedge_{\pi_i(\mathcal{R}) \in B} \mathcal{E}(\pi_i(\mathcal{R})) \right\}. \quad (7)$$

En plus de comparer l'énergie de la région avec les énergies optimales de ses enfants, toutes les autres partitions partielles de \mathcal{R} qui peuvent être contenues dans la tresse doivent être considérées (puisque \mathcal{R} représente le suprémum par raffinement de régions de la tresse et non ces régions elles mêmes). La coupe optimale de \mathcal{R} est ainsi donnée par $\{\mathcal{R}\}$, l'union des coupes optimales de ses enfants ou une autre partition partielle contenue dans la tresse, selon qui a la plus faible énergie.

3 Méthodologie proposée

Nous décrivons maintenant la méthodologie proposée (illustrée par le diagramme de la figure 2) pour réaliser la segmentation d'images multimodales, utilisant les tresses de partitions pour fusionner les résultats de multiples hiérarchies. Soit $\mathcal{I} = \{\mathcal{I}^1, \mathcal{I}^2\}$ une image multimodale (supposée par la suite être composée de deux modes de même support spatial E , l'extension à un plus grand nombre de modes étant aisée). Premièrement, une hiérarchie H est construite sur l'un des modes de \mathcal{I} , par exemple \mathcal{I}^1 . Deux énergies \mathcal{E}_λ^1 et \mathcal{E}_λ^2 sont définies suivant le modèle d'énergie de Mumford-Shah constante par morceaux [5] dont le terme d'attache aux données (AOD) est défini par rapport à chaque mode \mathcal{I}^1 et \mathcal{I}^2 , et dont le terme de régularisation est égal à la moitié du périmètre de la région :

$$\mathcal{E}_\lambda^i(\pi) = \sum_{R \in \pi} \left(\Xi_i(\mathcal{R}) + \frac{\lambda}{2} |\partial \mathcal{R}| \right) \quad (8)$$

avec $\Xi_i(\mathcal{R}) = \sum_{\mathbf{x} \in \mathcal{R}} \|\mathbf{x} - \boldsymbol{\mu}_i(\mathcal{R})\|_2^2$ étant l'AOD s'appliquant au mode \mathcal{I}^i et $\boldsymbol{\mu}_i(\mathcal{R})$ est la valeur/vecteur moyen pour le mode \mathcal{I}^i des pixels appartenant à la région \mathcal{R} . Le modèle d'énergie

de Mumford-Shah constante par morceaux, par sa capacité à produire des segmentation cohérentes, est un choix populaire pour les méthode de minimisation d'énergie. D'autres définitions pourraient également être implémentées en fonction de l'application, la seule contrainte étant que les énergies \mathcal{E}_λ^1 et \mathcal{E}_λ^2 soient h-croissantes et à croissance d'échelle. C'est le cas pour le modèle de Mumford-Shah [3]. Ensuite, deux ensembles de coupes optimales H_1^* et H_2^* sont générées à partir de H en faisant varier λ pour chaque \mathcal{E}_λ^1 et \mathcal{E}_λ^2 . D'après la propriété de croissance d'échelle de \mathcal{E}_λ^i , H_1^* et H_2^* forment également deux hiérarchies, qui définissent la tresse $B = \{H_1^*, H_2^*\}$ une fois combinées. H_1^* et H_2^* étant des coupes extraites de la même hiérarchie H , il est garanti de fait qu'elles forment une tresse avec une certaine hiérarchie moniteur H_m , puisque le suprémum de deux coupes d'une hiérarchie est également une coupe de cette hiérarchie. Cela implique également que les régions qui composent la hiérarchie moniteur H_m sont un sous-ensemble des régions qui composent la hiérarchie initiale H . À noter cependant que cette garantie est perdue dès lors que H_1^* et H_2^* proviennent de deux hiérarchies indépendantes. Un nouveau terme d'énergie \mathcal{E}_λ^m dépendant des deux modes de l'image multimodale \mathcal{I} est finalement défini :

$$\mathcal{E}_\lambda^m(\pi) = \sum_{\mathcal{R} \in \pi} \left(\max \left(\frac{\Xi_1(\mathcal{R})}{\Xi_1(\mathcal{I}^1)}, \frac{\Xi_2(\mathcal{R})}{\Xi_2(\mathcal{I}^2)} \right) + \frac{\lambda}{2} |\partial \mathcal{R}| \right). \quad (9)$$

L'AOD de chaque région \mathcal{R} est maintenant définie comme le maximum par rapport aux deux modes des AODs normalisées. La normalisation permet de ramener les deux AODs dans la même plage de valeurs. \mathcal{E}_λ^m est également une énergie h-croissante et à croissance d'échelle. Sa minimisation sur H_m et B d'après le programme dynamique (7) résulte finalement d'une segmentation optimale π_m^* de \mathcal{I} , qui devrait être composée de régions saillantes communes aux deux modes, ainsi que des régions représentées expressément par \mathcal{I}^1 et \mathcal{I}^2 .

4 Résultats

Nous appliquons la méthodologie proposée sur une image multimodale composée d'une image hyperspectrale \mathcal{I}^1 comportant 144 bandes spectrales régulièrement distribuées entre 380 nm et 1050 nm, et un modèle de surface digitale (MSD) dérivée du LiDAR \mathcal{I}^2 , avec la même résolution d'échantillonnage au sol de 2.5 m. Les données présentent diverse structures urbaines typiques telles que des bâtiments, des routes, des pelouses, des arbres ainsi qu'une piste d'athlétisme.

La hiérarchie H est obtenue par construction d'un arbre de partition binaire sur l'image hyperspectrale, avec des paramètres standards [6]. Quatre coupes optimales sont extraites de H pour chaque mode, le paramètre λ étant réglé pour produire des coupes optimales comportant environ 200, 150, 100 et 50 régions. Les 8 partitions résultantes, composant B , produisent $\binom{8}{2}$ coupes possible de la hiérarchie monitrice H_m (certaines de ces coupes peuvent être identiques). La minimisation de \mathcal{E}_λ^m sur H_m est conduite avec λ étant réglé empiriquement à 5.10^{-5} , produisant une segmentation optimale π_m^* composée de 264

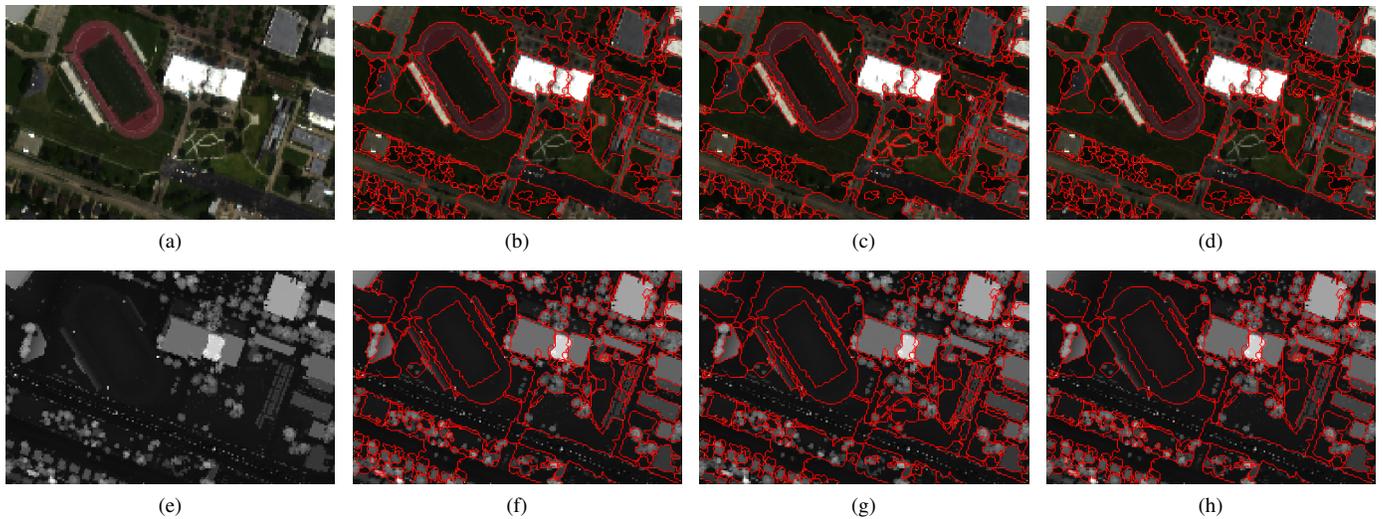


FIGURE 3 – ligne du haut : a) composition RVB de l’image hyperspectrale, de b) à d) segmentation optimales π_m^* , π_1^* et π_2^* superposées sur a). ligne du bas : e) MSD correspondant, de f) à h) segmentation optimales π_m^* , π_1^* et π_2^* superposées sur e).

régions. Une possible manière d’évaluer les améliorations apportées par la méthodologie proposée est de comparer π_m^* avec les deux segmentation optimales π_1^* et π_2^* , extraites de H par rapport à \mathcal{E}_λ^1 et \mathcal{E}_λ^2 , et ayant un nombre similaire de régions que π_m^* . En particulier, les deux segmentations π_1^* et π_2^* extraites de H contiennent respectivement 264 et 265 régions.

La figure 3 expose les résultats obtenus : la ligne du haut présente une composition RVB de l’image hyperspectrale (figure 3a) et les segmentations optimales π_m^* , π_1^* et π_2^* qui lui sont superposées (de la figure 3b à 3d). La ligne du bas présente l’image MSD (figure 3e) et les mêmes segmentations optimales π_m^* , π_1^* et π_2^* superposées dessus (de la figure 3f à 3h). La comparaison de π_m^* et π_1^* soulève plusieurs commentaires. Premièrement, toutes les régions saillantes notables du mode hyperspectral restent correctement segmentées en utilisant la structure de tresse, telles que la piste d’athlétisme, les différents bâtiments où bien les routes. Cependant, π_1^* est également composé de régions qui ne s’accordent pas avec le MSD. C’est le cas des différents groupes d’arbres présents dans l’image, qui apparaissent presque tous sous-segmentés (dans le coin inférieur gauche ou dans la partie supérieure de l’image). Cependant, ces régions sont correctement segmentées dans π_m^* et π_2^* . À l’inverse, certaines régions adjacentes ayant la même hauteur mais n’étant pas composées des mêmes matériaux sont sous-segmentées dans π_2^* mais pas dans π_m^* et π_1^* . C’est en particulier le cas de la piste d’athlétisme et de ses tribunes. La construction de la tresse et de la hiérarchie moniteur associée ainsi que la minimisation d’énergie consécutive fusionne effectivement l’information contenue dans chaque mode, de manière à produire une segmentation de l’image multimodale encore plus précise.

5 Conclusion

Nous présentons dans cet article une nouvelle méthode permettant d’effectuer la segmentation de données multimodales,

basée sur la minimisation hiérarchique d’une certaine fonction d’énergie. En particulier, nous utilisons le récent concept de tresses de partitions et hiérarchies moniteurs associées, pour lesquelles le programme dynamique utilisé pour effectuer la minimisation d’énergie sur des hiérarchies a été adaptée. La méthode proposée est appliquée sur une image multimodale composée d’un mode hyperspectral et d’un mode LiDAR. Les résultats montrent que l’approche proposée est capable produire une segmentation qui ne prend pas seulement en compte les régions saillantes communes aux deux modes, mais également les régions exprimées par seulement un mode de l’image multimodale.

Les travaux futurs incluent l’extension de la construction de la tresse à partir de plusieurs hiérarchies indépendamment construites sur leur mode respectif, ainsi que la construction de vérités terrain pour évaluer quantitativement les améliorations apportées par la méthode.

Références

- [1] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, and J. Chanussot. Challenges and opportunities of multimodality and data fusion in remote sensing. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European*, pages 106–110. IEEE, 2014.
- [2] L. Guigues, J. Cocquerez, and H. Le Men. Scale-sets image analysis. *International Journal of Computer Vision*, 68(3) :289–317, 2006.
- [3] B. Kiran and J. Serra. Global–local optimizations by hierarchical cuts and climbing energies. *Pattern Recognition*, 47(1) :12–24, 2014.
- [4] B. R. Kiran and J. Serra. Braids of partitions. In *Mathematical Morphology and Its Applications to Signal and Image Processing*, pages 217–228. Springer, 2015.
- [5] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5) :577–685, 1989.
- [6] S. Valero, P. Salembier, and J. Chanussot. Hyperspectral image representation and processing with binary partition trees. *Image Processing, IEEE Transactions on*, 22(4) :1430–1443, 2013.
- [7] M. Veganzones, G. Tochon, M. Dalla Mura, A. Plaza, and J. Chanussot. Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. *Image Processing, IEEE Transactions on*, 23(8) :3574–3589, Aug 2014.