

Application de la théorie de l'évidence à la combinaison de segmentations en régions.

Stéphane MATHEVET¹, Laurent TRASSOUDAIN¹, Paul CHECCHIN², Joseph ALIZON¹

¹LASMEA, UMR 6602 du CNRS, Université Blaise Pascal
24, Avenue des Landais, 63177 Aubière cedex, France

²CEMIF-SC Université d'Evry Val d'Essonne
40, rue du Pelvoux CE 1455, 91020, Evry cedex, France

Stephane.Mathevet@lasmea.univ-bpclermont.fr, Laurent.Trassoudaine@lasmea.univ-bpclermont.fr
Paul.Checchin@iup.univ-evry.fr, Joseph.Alizon@lasmea.univ-bpclermont.fr

Résumé – Cet article présente un algorithme de combinaison de segmentations décrites sous la forme de régions. La théorie de l'évidence est le cadre théorique de cette étude. La probabilité d'appartenance de chaque pixel à une région voisine permet de déduire le jeu de masse associé à une segmentation. Les jeux de masse ainsi obtenus sont combinés par la règle de Dempster. Le critère du maximum de plausibilité permet de choisir l'étiquette définitive du pixel. Une évaluation quantitative et comparative de l'apport de la combinaison a été menée sur un jeu de 40 images réelles.

Abstract – This paper deals with the combination of segmentations described with regions. For this study we use the evidential theory. The membership probability of each pixel of the neighbouring regions allows to calculate the mass function from the associate segmentation. The resulting mass functions are combined by Dempster rule of combination. The maximum plausibilities criterion is used to choose the final label of the pixel. A quantitative and comparative evaluation of the combination is done with a set of 40 real images.

1 Introduction

La fusion de données est souvent un passage nécessaire pour l'analyse de systèmes ou de situations complexes. Il en est de même en analyse d'images où elle permet de tirer profit de la redondance et de la complémentarité des données. Dans ce domaine, la segmentation des images est une étape préalable à l'interprétation des scènes. Si l'on fait abstraction de la fusion de bas niveau, on peut imaginer deux approches pouvant conduire à l'interprétation de la scène par combinaison d'informations issues de différents capteurs :

- les segmentations obtenues par différents capteurs sont combinées afin de conduire à une segmentation unique qui sera ensuite interprétée,
- chaque segmentation est interprétée et le fruit de ces interprétations est ensuite combiné.

La deuxième approche est souvent utilisée [1]. L'objectif de cette étude est de réaliser la combinaison de segmentations issues de sources et d'algorithmes différents pour en obtenir une nouvelle, plus révélatrice des attributs des objets présents dans la scène. Une telle stratégie est moins souvent appliquée mais on trouve cependant des exemples dans la littérature où l'approche contour est plus souvent utilisée [2][3][4], mais où l'approche région est également mise en oeuvre [5].

Pour cette étude, nous utilisons des données de profondeur et de réflectance fournies par un même capteur

laser. Il fournit pour chaque pixel les coordonnées X, Y, Z du point 3D correspondant. L'information réflectance est donnée, pour le même point, par l'intensité réfléchie de l'onde laser. De part la nature même de ce capteur, l'alignement spatio-temporel est inhérent aux données. La mise en correspondance des étiquettes d'une même région dans deux segmentations différentes est réalisée par la méthode sous optimale de Vinet. La théorie de l'évidence est le cadre théorique de cette étude.

2 La théorie de l'évidence

Dans cette partie nous présentons les éléments de la théorie de l'évidence qui sont utilisés dans le développement de l'algorithme.

Initialement introduite par Dempster [6], la théorie de l'évidence fut reprise, avec un formalisme mathématique plus abouti par Shafer. Elle est décrite en détails dans [7]. Cette théorie peut être considérée comme une généralisation de la théorie des probabilités. Elle nécessite la définition d'un ensemble θ composé de N hypothèses H_i exclusives et exhaustives, cet ensemble est appelé « cadre de discernement ». 2^θ représente l'ensemble des $2^N - 1$ parties A_k de θ , il sert de référentiel aux fonctions utilisées par la théorie de l'évidence.

Dans cette application, les hypothèses sont les étiquettes des régions d'une image ; le cadre de discernement est l'en-

semble de ces étiquettes. Les principales fonctions utilisées sont :

- la fonction de masse élémentaire $m_\theta(\cdot)$ définie par :

$$m_\theta(\phi) = 0 \text{ et } \sum_k^{2^N-1} m_\theta(A_k) = 1, \quad (1)$$

- la crédibilité $Cr_\theta(\cdot)$ définie par :

$$Cr_\theta(A) = \sum_{B \subseteq A} m_\theta(B), \quad (2)$$

- la plausibilité $Pl_\theta(\cdot)$ définie par :

$$Pl_\theta(A) = \sum_{A \cap B \neq \phi} m_\theta(B) \quad (3)$$

L'intérêt de la théorie de l'évidence en fusion de données est de pouvoir construire un jeu de masse unique par simple sommation orthogonale de M jeux de masse $m_{\theta_{S_k}}(\cdot)$ issus de M sources S_k distinctes, définies sur le référentiel commun 2^θ (règle de Dempster).

$$m_\theta(A) = \frac{\sum_{A_1 \cap \dots \cap A_M = A \neq \phi} \left\{ \prod_{k=1}^M m_{\theta_{S_k}}(A_k) \right\}}{1 - \sum_{A_1 \cap \dots \cap A_M = \phi} \left\{ \prod_{k=1}^M m_{\theta_{S_k}}(A_k) \right\}} \quad (4)$$

En pratique, il existe des cas où le cadre de discernement associé à une source n'est pas exhaustif. Dans ces conditions une étape de réajustement des cadres de discernement est nécessaire. Cette étape est réalisée grâce à la méthode de correction des plausibilités [9] lorsque les cadres sont différents mais avec une partie commune, et grâce à la méthode du minimum d'affectation [8] dans les cas de cadres disjoints.

3 Application à la combinaison de segmentations

L'objectif de l'algorithme (cf. figure 1) est de déterminer l'étiquette de chacun des pixels de l'image, connaissant leur étiquette dans chacune des segmentations préalables.

Chaque région de chaque image est modélisée par une distribution gaussienne de l'attribut: le niveau de gris pour l'image de réflectance, la distance du point au plan passant au mieux par l'ensemble des points de la région pour l'image de profondeur.

Pour chaque pixel de chaque segmentation, le cadre de discernement est composé de l'étiquette du pixel considéré et de l'ensemble des étiquettes des régions connexes à la région contenant le pixel. Le jeu de masse pour un pixel est déduit, pour chaque hypothèse (étiquette) du cadre de discernement, de la probabilité d'appartenance du pixel à la région considérée en utilisant l'un des deux modèles proposés par Appriou [10].

Une première combinaison des jeux de masse ainsi obtenus a lieu au sein de chaque segmentation afin d'obtenir, pour chaque pixel, les masses de chaque partie du cadre de discernement. Pour cette combinaison la règle de Dempster (cf. Eq. 4) est utilisée.

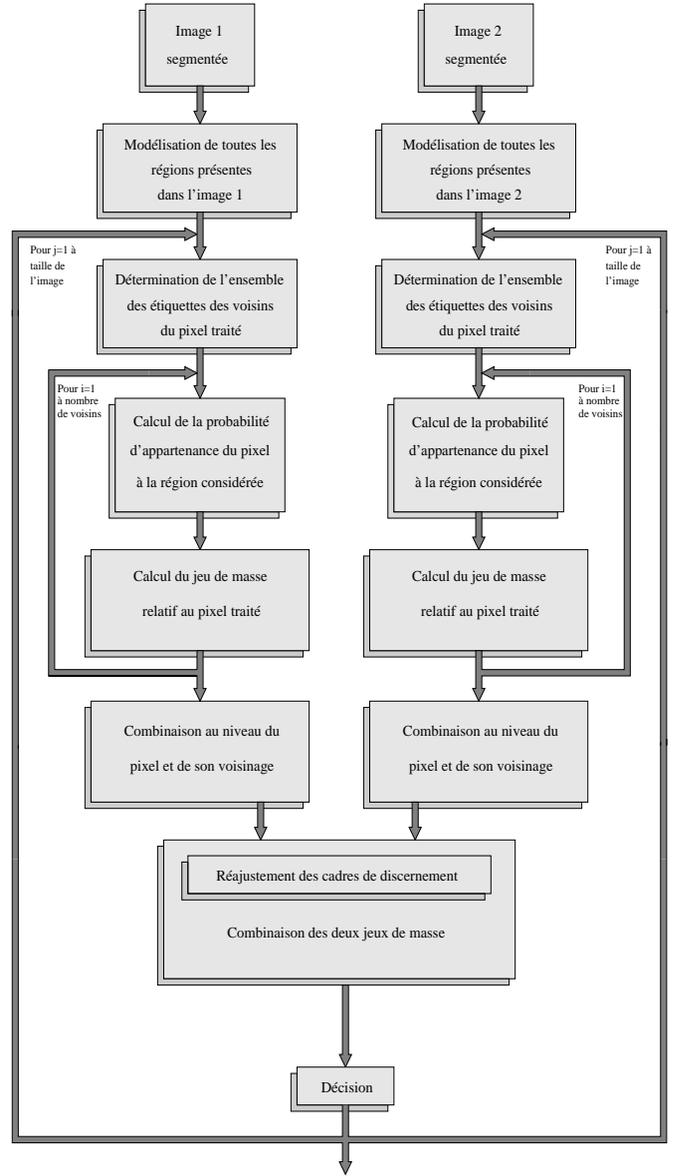


FIG. 1: Principe de la combinaison de deux segmentations.

Intervient ensuite la combinaison des informations proprement dites dans laquelle les jeux de masse d'un même pixel dans les deux segmentations sont combinés. Se pose ici le problème de l'exhaustivité des cadres de discernement :

- s'ils sont identiques, la règle de combinaison de Dempster est utilisée directement et les plausibilités sont déduites du jeu de masse combiné,
- s'ils sont disjoints, la combinaison est précédée d'une étape de réajustement des cadres de discernement par la méthode du minimum d'affectation,
- s'ils sont différents, mais avec une partie commune, on utilise alors la méthode de correction des plausibilités qui conduit directement aux plausibilités combinées.

La décision concernant l'étiquette du pixel considéré est prise selon la règle du maximum de plausibilité. La méthode ne garantissant pas la connexité des régions, une phase de re-étiquetage est nécessaire à la suite de laquelle les petites régions sont éliminées.

4 Résultats sur des images réelles

Dans cette section nous présentons les résultats obtenus avec la méthode proposée. Les images traitées ont été acquises avec un capteur télémétrique laser nommé PERCEPTRON. La figure 2 montre 4 des 40 images réelles utilisées pour l'évaluation de l'algorithme.

A chaque image réelle est associée une vérité de terrain (cf. figure 3) définie manuellement par un opérateur [11]. Cette banque d'images permet une évaluation qualitative mais surtout quantitative et comparative des résultats. Ces derniers, présentés ici, ont été obtenus avec la première modélisation des jeux de masse et avec filtrage des petites régions.

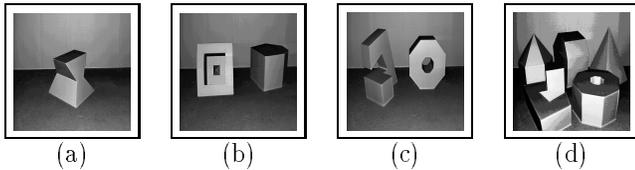


FIG. 2: Images de réflectance.

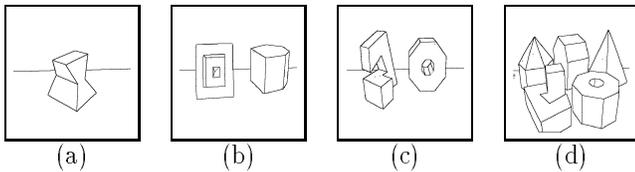


FIG. 3: Vérités de terrain.

Les segmentations 2D, présentées figure 4, sont obtenues à l'aide d'un algorithme de morphologie mathématique, fondé sur l'extraction de la Ligne de Partage des Eaux (LPE) effectuée sur l'image gradient. Ces segmentations sont de qualité moyenne étant donné le faible contraste existant sur les images originales. Les segmentations 3D de la figure 5 sont obtenues à l'aide d'un algorithme de segmentation d'images de profondeur, développé par Chечhin et al [12].

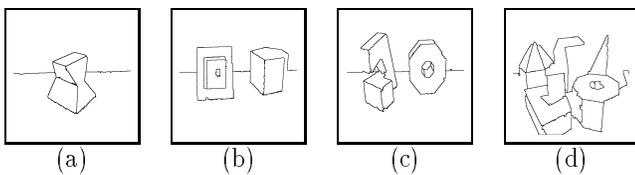


FIG. 4: Segmentations 2D.

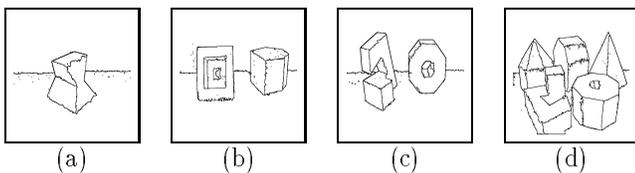


FIG. 5: Segmentations 3D.

Ces quelques segmentations, représentées par les contours des régions, font apparaître les principaux défauts rencon-

trés : régions sous-segmentées en 2D, nettement visible sur la troisième image (cf. figure 4c) ; imprécision des contours, surtout visible en 3D.

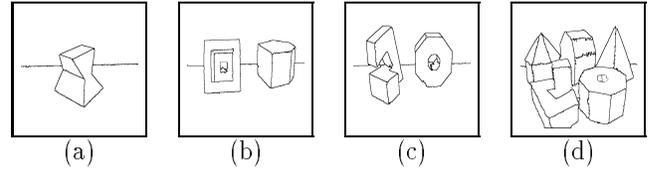


FIG. 6: Segmentations résultats.

Le premier résultat obtenu par la méthode proposée (cf. figure 6a) semble satisfaisant. L'amélioration apportée sur les contours est nettement visible en comparaison à la segmentation 3D. La segmentation 2D est plutôt bonne, ceci assure un bon résultat de combinaison.

Le deuxième résultat (cf. figure 6b) montre que la méthode permet de retrouver des régions qui avaient été sous-segmentées par l'une ou l'autre des segmentations de départ (ici la segmentation 2D). La partie supérieure ainsi qu'une face latérale de l'objet de droite n'ont pas été détectés (cf. figure 4b). Là encore, la combinaison permet d'améliorer les contours vis à vis des segmentations 3D.

Le troisième résultat présenté (cf. figure 6c) montre les limites actuelles de la méthode. Il est nettement visible que des régions n'ont pas été reconstruites correctement. Il s'agit des régions situées au centre de la pièce de droite. Sur la figure 3c, on observe cinq régions. Les segmentations 2D et 3D contiennent quatre régions, non identiques. Le résultat fait apparaître quatre régions mais avec une imprécision sur les bords. En revanche, les sous-segmentations sont gérées de façon satisfaisante sur la pièce de gauche. On peut dire que la gestion des sous-segmentations est faite convenablement seulement lorsque la disparité entre les attributs d'une région par rapport à ceux des régions voisines est grande. La même remarque peut être faite concernant les sur-segmentations.

Enfin, le dernier résultat (cf. figure 6d) synthétise l'ensemble des remarques faites précédemment. Il s'agit de l'image la plus complexe de la base utilisée. La vérité de terrain de cette image comporte 27 régions. La segmentation 2D en fait apparaître seulement 17, contre 26 pour la segmentation 3D ainsi que pour le résultat. Ce résultat imparfait est tout de même plus proche de la référence que les segmentations 2D (en terme de régions) et 3D (en terme de précision des contours).

Afin de caractériser les performances de l'algorithme présenté, une évaluation quantitative et comparative des résultats a été menée. La méthode d'évaluation utilise la mesure de Hoover et al [11] qui classe les régions en régions « correctement segmentées » ou non en fonction d'un taux de recouvrement T entre la région de la vérité de terrain et celle de la segmentation.

La figure 7 montre les résultats de combinaison obtenus grâce au premier modèle proposé par Appriou et amélioré par le filtrage des petites régions, ceux obtenus sur les segmentations 2D et 3D et enfin les résultats obtenus par l'Université d'Edinburgh avec une segmentation 3D seule.

En effet, au cours de récents travaux, quatre groupes de recherche, appartenant à l'Université de Floride du Sud, à l'Université de Bern, à l'Université de l'État de Washington et enfin à l'Université d'Edinburgh, ont proposé des algorithmes de segmentation d'images de profondeur en régions planes. Seuls les résultats de l'Université d'Edinburgh, présentant les scores les plus élevés, sont reportés sur la courbe. Nous avons choisi ces résultats dans le but de situer les performances de notre méthode.

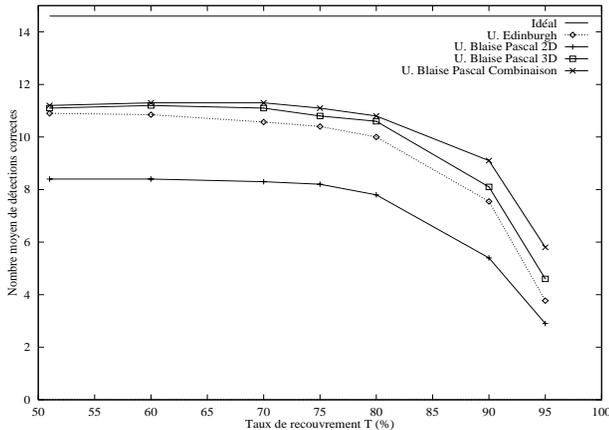


FIG. 7: Évolution des détections correctes pour l'ensemble des images issues du capteur PERCEPTON.

On note que la courbe illustrant les résultats obtenus à l'aide de la segmentation 2D est à un niveau très inférieur aux autres. En effet, de nombreuses régions sont sous-segmentées en 2D.

Dans tous les cas les scores de combinaison sont supérieurs aux segmentations 3D seules malgré la qualité moyenne de la segmentation 2D. Cette remarque confirme l'intérêt de la fusion. Si on se place à un taux de recouvrement de 95%, la méthode affiche encore un taux de bon classement de 40%, soit 8% de plus que la segmentation 3D seule (augmentation relative de 25%), 14% de plus que l'Université d'Edinburgh et 20% de plus que la segmentation 2D. L'apport de la combinaison est d'autant plus important que la contrainte est forte. En dessous de 80%, on améliore sensiblement les résultats; au-delà de 80% l'amélioration est très nette.

5 Conclusion

L'objectif de cette étude est de proposer un algorithme de combinaison de segmentations issues de capteurs différents, ou de données différentes fournies par un capteur unique, en utilisant le formalisme de la théorie de l'évidence. Nous avons décrit les aspects théoriques permettant de réaliser cette étude, puis l'algorithme mis en place pour y parvenir.

Les résultats obtenus avec cette méthode de combinaison sont satisfaisants. La détection des bords des régions des objets présents dans les images est très souvent améliorée. Malgré une segmentation 2D de relativement faible qualité et des segmentations 3D ne présentant pas toujours des bords d'objets très nets, l'algorithme fournit des

résultats convenables. Par ailleurs, une évaluation quantitative et comparative a permis de montrer que les scores de la segmentation combinée sont meilleurs que ceux obtenus par les autres méthodes prises séparément.

Cette étude renforce l'intérêt de la fusion pour obtenir des segmentations plus représentatives des scènes observées et cela même dans les cas où les segmentations initiales ne sont pas toutes de bonne qualité.

Références

- [1] I. Bloch. « Some Aspects of Dempster-Shafer Evidence Theory for Classification of Multi-Modality Medical Images Taking Partial Volume Effect into Account ». *Pattern Recognition Letters*, 17(8):905–919, July 1996.
- [2] Sateesha G. Nadabar and Anil K. Jain. « Edge Detection and Labeling by Fusion of Intensity and Range Images ». *Applications of Artificial Intelligence: Machine Vision and Robotics*, 1708:108–119, 1992.
- [3] M. Abdulghafour and M. A. Abidi. « Data Fusion Through Non-Deterministic Approaches - A Comparison ». *SPIE Sensor Fusion VI*, 2059:37–53, 1993.
- [4] B. Gil, A. Mitiche, and J.K. Aggarwall. « Experiments in Combining Intensity and Range Edge Maps ». *Computer Vision, Graphics, and Image Processing*, 21:395–411, 1983.
- [5] R. Kara-Falah. « Segmentation d'images : coopération, fusion, évaluation ». Thèse de Doctorat, Université de Savoie, Juin 1995.
- [6] A.P. Dempster. « A Generalization of Bayesian Inference ». *Journal of the royal statistical society*, 30(B), 1968.
- [7] G. Shafer. « *A Mathematical Theory of Evidence* ». Princeton University Press, 1976.
- [8] P. Smets. « The Combination of Evidence in the Transferable Belief Model ». *IEEE transactions on pattern analysis and machine intelligence*, 12(5):447–458, 1990.
- [9] F. Janez. « Fusion de sources d'information définies sur des référentiels non exhaustifs différents ». Thèse de Doctorat, Université d'Angers, Novembre 1996.
- [10] A. Appriou. « Probabilités et incertitude en fusion de données multi-senseurs ». *Revue scientifique et technique de la défense*, pages 27–40, Novembre 1991.
- [11] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. bunke, D. Goldgof, K.W. Bowyer, D. Eggert, A. Fitzgibbon, and R. Fisher. « An Experimental Comparison of Range Image Segmentation Algorithms ». *IEEE Transactions on PAMI*, 18(7):673–689, 1996.
- [12] P. Checchin, L. Trassoudaine, and J. Alizon. « Segmentation of Range Images into Planar Regions ». In *IEEE International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, pages 156–163, 12-15 May 1997. Ottawa Ontario Canada.