

Combinaison de segmentations en régions

Combination of image segmentation into regions

par Stéphane MATHEVET*, Laurent TRASSOUDAINE*, Paul CHECCHIN**, Joseph ALIZON*

* I.A.S.M.E.A. U.M.R. 6602 du C.N.R.S., Université Blaise Pascal de Clermont-Ferrand, 24, Avenue des Landais, 63177 Aubière cedex
email : Stephane.Mathevet, Laurent.Trassoudaine, Joseph.Alizon@lasmca.univ-bpclermont.fr

** CEMIF-SC, Université d'Evry Val d'Essonne, 40, rue du Pelvoux CE 1455 Courcouronnes, 91020 Evry cedex
email : Paul.Checchin@iup.univ-evry.fr

résumé et mots clés

La segmentation est une étape préalable à l'interprétation des images. Dans cet article, nous nous intéressons au problème de la fusion de données appliquée à la segmentation d'images. Le processus de segmentation doit extraire les informations « utiles » d'une image numérique afin de reconnaître les entités présentes dans cette image. La fusion de données permet de pallier les lacunes de certaines modalités en mettant à profit la redondance et la complémentarité de l'ensemble des informations. Nous proposons une méthode de combinaison de segmentations décrites sous la forme de régions dans le cadre de la théorie de l'évidence. Les masses sont déduites des probabilités d'appartenance de chaque pixel de l'image aux régions de celles-ci. Ces masses sont combinées par la règle de combinaison de Dempster en gérant préalablement la non-exhaustivité du cadre de discernement due aux sous-segmentations d'une image par rapport à l'autre. La prise de décision est réalisée par le maximum de plausibilité. Des résultats quantitatifs et comparatifs viennent illustrer la méthode.

Segmentation, fusion de données, théorie de l'évidence.

abstract and key words

Segmentation is a preliminary stage before image interpretation. In this article we are interested in the problem of data fusion applied to image segmentation. The aim of the study is to combine different image segmentations to obtain a more reliable one. Data fusion is actually an interesting tool to obtain more reliable data from uncertain and inaccurate ones. We propose a segmentation combination method within the context of Dempster-Shafer theory. The basic probability assignment (bpa) are computed from the probability that a pixel belong to a region of its neighbourhood. The bpa are then combined with the Dempster rule of combination. The non-exhaustivity of the frame of discernement due to the under-segmentation is taken into account. The maximum of plausibility is used for decision-making. We present quantitative and comparative results concerning intensity images and depth images.

Segmentation, data fusion, evidence theory.

1. introduction

La fusion de données est très 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 très souvent 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 (cf. figure 1),
- chaque segmentation est interprétée et le fruit de ces interprétations est ensuite combiné (cf. figure 2).

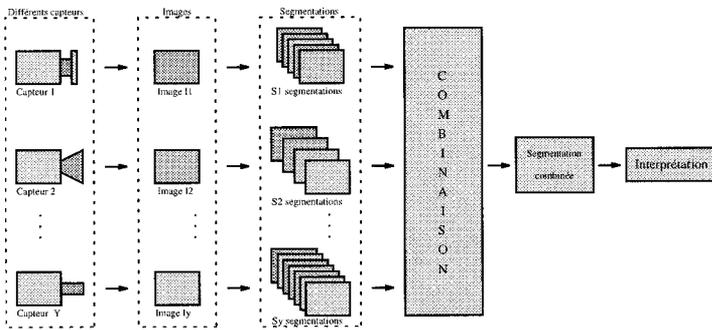


Figure 1. – Interprétation de la segmentation combinée.

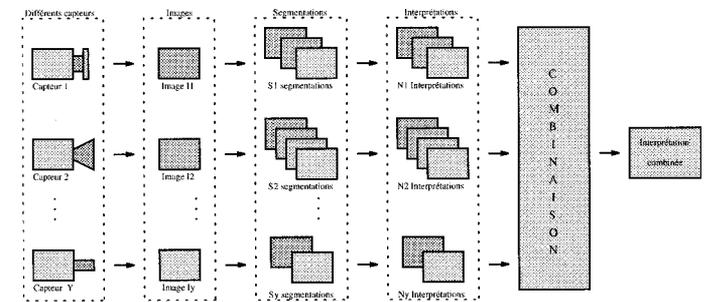


Figure 2. – Interprétation de chaque segmentation avant combinaison des interprétations.

Le deuxième schéma est souvent utilisé [1] [2]. 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.

Nadabar & Jain [3] proposent une méthode de détection de contours et d'étiquetage par fusion d'images de luminance et de profondeur. Leur méthode utilise la règle de Bayes pour effectuer la combinaison des différentes sources. La connaissance *a priori* dont ils disposent sur les étiquettes est modélisée par des champs de Markov.

Abdulghafour & Abidi [4] proposent une méthode où la fusion s'effectue selon différentes approches, telles que la logique floue, la règle de Bayes et la théorie de Dempster-Shafer. Il s'agit ici de combiner une image de contours issue de l'image d'intensité, une image de contours issue de l'image de profondeur et une image des normales. Les algorithmes sont testés sur des images réelles d'intensité et de profondeur acquises avec un capteur laser « Odetics Laser Range Scanner ». Le but est d'obtenir une meilleure description de la scène observée à travers la segmentation des deux images.

Citons également les travaux de Gil, Mitiche & Aggarwal [5]. Leur but est de réaliser la segmentation d'une scène en utilisant les informations de luminance et de profondeur. L'approche consiste en l'extraction de contours dans les deux images. La combinaison

est réalisée selon deux approches différentes : locale puis globale. L'approche locale consiste à travailler sur un pixel et son voisinage alors que l'approche globale consiste à travailler sur une liste de points de contour dans toute l'image. Une comparaison est effectuée entre l'étiquette du pixel traité et les étiquettes des points du voisinage ou de la liste.

Kara-Falah [6] propose une méthode reposant sur une représentation en régions pour chaque segmentation. Une mise en correspondance est réalisée entre les régions des différentes segmentations. A partir de n régions nettes, qui sont des représentations d'une unique région R de l'image, une région floue est déduite. Pour chaque pixel de l'image, l'information est représentée par un degré d'appartenance à une région par rapport à une autre. Plus ce degré est proche de un, plus l'information est sûre. La segmentation nette est obtenue en affectant chaque pixel à la région pour laquelle il a le degré d'appartenance maximum.

Les trois premières méthodes présentées combinent des segmentations en contour. Seule la dernière adopte une approche région comme nous nous proposons d'étudier en utilisant le formalisme de la théorie de l'évidence. Elle est appliquée aux segmentations d'images de luminance et de profondeur.

La combinaison de segmentations nécessite deux phases préalables que sont le calibrage des capteurs et la mise en correspondance des étiquettes utilisées dans chacune des segmentations. Pour cette étude, nous utilisons des données de profondeur et de réflectance fournies par un même capteur laser. L'information profondeur est issue du laser qui pour chaque pixel donne 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 par la nature même de ce capteur, l'alignement spatio-temporel est inhérent aux données. En ce qui concerne la mise en correspondance des étiquettes, l'objectif est d'attribuer la même étiquette à une région quelle que soit la segmentation dont elle est issue. La méthode optimale à utiliser est le calcul d'un couplage maximum dans un graphe biparti. Nous nous sommes contentés de la méthode sous optimale de Vinet [7].

Dans cet article, nous rappelons les bases de la théorie de l'évidence nécessaires à la combinaison de segmentations que nous exposons dans une seconde partie. Des résultats sur des images réelles viennent illustrer la méthode. Une évaluation quantitative et comparative est également menée.

2. la théorie de l'évidence

La théorie de l'évidence, initialement introduite par Dempster [8], fut reprise avec un formalisme mathématique plus abouti par Shafer. Elle est décrite en détails dans [9]. 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 ($i \in \{1, \dots, N\}$) exclusives et exhaustives,

dont l'une d'entre elles est la solution du problème. Cet ensemble est appelé « cadre de discernement ». 2^θ représente l'ensemble des $2^N - 1$ parties A_k ($k \in \{1, \dots, 2^N - 1\}$) de θ . Il sert de référentiel pour l'ensemble des fonctions utilisées par la théorie de l'évidence. Le paragraphe qui suit présente les principales fonctions utilisées dans cet article.

2.1. les principales fonctions

La fonction de masse élémentaire $m_\theta(\cdot)$ est définie de l'ensemble 2^θ sur l'intervalle $[0, 1]$ par les relations :

$$m_\theta(\phi) = 0 \quad (1)$$

$$\sum_k^{2^N-1} m_\theta(A_k) = 1 \quad (2)$$

Les éléments focaux sont alors les éléments A_k de 2^θ dont la masse $m_\theta(A_k)$ est non nulle. La théorie de l'évidence permet non seulement l'évaluation des hypothèses singletons H_i , mais également l'évaluation des hypothèses composées telles que $(H_1 \cup H_2)$ ou $(H_3 \cup H_4 \cup H_5)$, etc. Une telle souplesse permet de modéliser l'ignorance totale par le jeu de masse :

$$m_\theta(\theta) = 1 \text{ et } m_\theta(A) = 0 \forall A \neq \theta \quad (3)$$

La masse $m_\theta(A)$ représente le degré de croyance placé exactement sur la proposition A . Pour prendre en considération toute cette croyance, on définit une grandeur appelée **crédibilité**.

En effet, les ensembles B inclus dans la proposition A , contribuent également à sa croyance. Par conséquent, pour prendre en compte toute la croyance sur A , on calcule sa crédibilité définie de l'ensemble 2^θ sur l'intervalle $[0, 1]$ par :

$$Cr_\theta(A) = \sum_{B \subseteq A} m_\theta(B) \quad (4)$$

Une autre grandeur, duale de la crédibilité est la **plausibilité**. Elle est définie de l'ensemble 2^θ sur l'intervalle $[0, 1]$ par :

$$Pl_\theta(A) = 1 - Cr_\theta(\bar{A}) \quad (5)$$

d'où

$$Pl_\theta(A) = \sum_{A \cap B \neq \phi} m_\theta(B) \quad (6)$$

La plausibilité mesure l'intensité avec laquelle on ne doute pas de la proposition A , c'est-à-dire l'intensité avec laquelle on trouve A plausible.

Intuitivement, la crédibilité et la plausibilité peuvent être interprétées respectivement comme une mesure de vraisemblance minimale et maximale, encadrant la probabilité $p_\theta(A)$:

$$Cr_\theta(A) \leq p_\theta(A) \leq Pl_\theta(A) \quad (7)$$

La différence entre ces deux grandeurs, pour une même proposition, donne l'ignorance relative à cette dernière.

Une autre fonction manipulée par la théorie de l'évidence, et dont l'intérêt sera souligné par la suite, est la **fonction de communalité** également définie de l'ensemble 2^θ sur l'intervalle $[0, 1]$ par :

$$Q_\theta(A) = \sum_{B \supseteq A} m_\theta(B) \quad (8)$$

2.2. combinaison de sources

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, avec $k \in \{1, \dots, M\}$, définies sur le référentiel commun 2^θ . La loi de combinaison de Dempster est un outil qui permet de renforcer la croyance sur les événements pour lesquels les différentes sources sont en accord, et au contraire de l'atténuer lorsque les sources sont en conflit. Le jeu de masse résultant est alors défini par :

$$m_\theta(A) = m_{\theta_{S_1}}(A) \oplus m_{\theta_{S_2}}(A) \oplus \dots \oplus m_{\theta_{S_M}}(A) \quad (9)$$

$$m_\theta(A) = K_\theta \cdot \sum_{A_1 \cap A_2 \cap \dots \cap A_M = A \neq \phi} \left\{ \prod_{k=1}^M m_{\theta_{S_k}}(A_k) \right\} \quad (10)$$

A_k représente une partie quelconque de l'ensemble 2^θ . Le terme K_θ , coefficient de normalisation, est donné par la relation :

$$K_\theta = \frac{1}{1 - k_\theta} \quad (11)$$

avec

$$k_\theta = \sum_{A_1 \cap A_2 \cap \dots \cap A_M = \phi} \left\{ \prod_{k=1}^M m_{\theta_{S_k}}(A_k) \right\} \quad (12)$$

Le coefficient k_θ est l'inconsistant de la fusion, il représente le degré de conflit entre les sources. Dans le cas de deux sources, les formules se réduisent à :

$$m_\theta(A) = \frac{1}{1 - k_\theta} \sum_{A_1 \cap A_2 = A \neq \phi} m_{\theta_{S_1}}(A_1) \cdot m_{\theta_{S_2}}(A_2) \quad (13)$$

avec

$$k_\theta = \sum_{A_1 \cap A_2 = \phi} m_{\theta_{S_1}}(A_1) \cdot m_{\theta_{S_2}}(A_2) \quad (14)$$

Une telle loi de combinaison possède un certain nombre de propriétés intéressantes en fusion de données multi-capteurs, en particulier la commutativité et l'associativité.

2.3. hypothèse « monde fermé »

En pratique, il existe des cas où le cadre de discernement associé à une source n'est pas exhaustif. Il correspond alors à l'hypothèse « monde ouvert » proposée par Smets [10] par opposition à l'hypothèse « monde fermé » préconisée par la théorie de l'évidence. L'hypothèse « monde ouvert » propose de ne pas renormaliser la fusion ce qui permet d'avoir une masse non nulle, significative du conflit, sur l'ensemble vide. La masse conflictuelle k_θ est alors affectée à l'ensemble vide ϕ : $m(\phi) = k_\theta \neq 0$. La loi de combinaison de Dempster impose de combiner des sources distinctes dont les jeux de masse sont définis sur un même référentiel. Il faut donc se ramener à un cadre de discernement exhaustif ainsi qu'à l'hypothèse « monde fermé ».

Afin de pouvoir utiliser cette règle de combinaison, plusieurs méthodes existent dont le déconditionnement. Il s'agit de redéfinir la connaissance sur un ensemble plus large que celui sur laquelle elle est définie initialement. Nous avons choisi la méthode du minimum d'affectation. On peut s'affranchir du déconditionnement en utilisant une méthode proposée par Janez [11] : la méthode de correction des plausibilités. Cette dernière permet de combiner directement des sources définies sur des référentiels non exhaustifs différents sans réaliser préalablement de déconditionnement. Cette méthode est cependant limitée puisqu'elle ne peut déduire qu'une connaissance partielle : la plausibilité. Cette connaissance est suffisante lorsqu'on utilise le maximum de plausibilité comme critère de décision.

Principe de la méthode du minimum d'affectation

La méthode du minimum d'affectation, proposée par Smets, permet de réaliser le meilleur déconditionnement pour une source S_k , lorsqu'aucune connaissance des liens de compatibilité existant entre les différentes hypothèses n'est disponible. La seule connaissance dont on dispose est celle définie sur le référentiel θ_{S_k} . La méthode va alors permettre de redéfinir la connaissance sur un référentiel plus large et commun à l'ensemble des sources. Ce référentiel sera noté θ .

Le jeu de masse correspondant sur le cadre complet θ s'obtient par :

$$\forall A \subseteq \theta \begin{cases} m_\theta(A \cup (\theta - \theta_{S_k})) = m_{\theta_{S_k}}(A) & \text{si } A \in N_{\theta_{S_k}} \\ m_\theta(A) = 0 & \text{si } A \notin N_{\theta_{S_k}} \end{cases} \quad (15)$$

avec le noyau $N_{\theta_{S_k}}$ défini par :

$$N_{\theta_{S_k}} = \{A \subseteq \theta_{S_k} / m_{\theta_{S_k}}(A) > 0\} \quad (16)$$

De façon simple, le déconditionnement consiste à ajouter (au sens de l'union) aux éléments focaux de la source S_k les hypothèses qui lui manquent.

La méthode de correction des plausibilités

L'utilisation de cette méthode impose trois contraintes : d'une part les sources doivent disposer d'une partie commune, d'autre part elle impose l'utilisation de la plausibilité car la méthode repose sur des propriétés que seule celle-ci vérifie. Cette deuxième contrainte n'est pas préjudiciable, elle impose simplement d'utiliser le maximum de plausibilité comme critère de décision. Il est d'ailleurs très communément utilisé. La non associativité de la méthode de correction des plausibilités constitue la troisième contrainte. En effet, Janez a montré que les opérations doivent s'effectuer sur les plus grandes parties communes des référentiels et donc selon un certain ordre. Cette contrainte peut être éventuellement gênante lorsque le nombre de sources est supérieur à deux.

Dans le cas de deux sources S_1 et S_2 définies sur les référentiels θ_{S_1} et θ_{S_2} tels que $\theta_{S_1} = \{H_1, H_2, H_3\}$, $\theta_{S_2} = \{H_2, H_3, H_4\}$, la méthode s'applique comme suit. Dans un premier temps, on calcule les plausibilités pour toutes les hypothèses communes aux deux sources, soit $\theta_c = \{H_2, H_3\}$. Dans un second temps, on calcule les plausibilités des hypothèses non communes H_1 et H_4 . Pour les hypothèses considérées par une seule source, un réajustement de la connaissance est nécessaire, c'est-à-dire que l'on doit calculer la plausibilité de l'ensemble des hypothèses communes sur la source qui ne considère pas cette hypothèse. Un tel développement aboutit aux formules de plausibilités combinées suivantes :

– pour les hypothèses communes :

$$\begin{aligned} Pl_\theta(H_2) &= Pl_{\theta_{S_1}}(H_2) \cdot Pl_{\theta_{S_2}}(H_2) \\ Pl_\theta(H_3) &= Pl_{\theta_{S_1}}(H_3) \cdot Pl_{\theta_{S_2}}(H_3) \end{aligned} \quad (17)$$

– pour les hypothèses non communes :

$$\begin{aligned} Pl_\theta(H_1) &= Pl_{\theta_{S_1}}(H_1) \cdot Pl_{\theta_{S_2}}(\theta_c) \\ Pl_\theta(H_4) &= Pl_{\theta_{S_1}}(\theta_c) \cdot Pl_{\theta_{S_2}}(H_4) \end{aligned} \quad (18)$$

2.4. modélisation des jeux de masse

Traiter un problème par la théorie de l'évidence nécessite la détermination d'un jeu de masse élémentaire $m_{\theta_{S_k}}$ pour chaque source S_k ($k \in \{1, \dots, M\}$). Pour cela Appriou [12] propose de rechercher pour chaque source une modélisation des N probabilités *a priori* $p_{\theta_{S_k}}(H_i)$ ($i \in \{1, \dots, N\}$), pour chaque hypothèse H_i , ainsi que les N coefficients de confiance q_{ik} associés. Le coefficient q_{ik} représente le degré de confiance placé sur la probabilité $p_{\theta_{S_k}}(H_i)$; le cadre de discernement étant l'ensemble des N hypothèses H_i répertoriées *a priori*. Afin de passer des probabilités aux masses, Appriou a proposé deux modèles :

$$\begin{aligned} \text{Modèle 1 : } \quad m_{\theta_{S_k}}(H_i) &= q_{ik} \cdot p_{\theta_{S_k}}(H_i) \\ m_{\theta_{S_k}}(\bar{H}_i) &= q_{ik} \cdot [1 - p_{\theta_{S_k}}(H_i)] \\ m_{\theta_{S_k}}(\theta_{S_k}) &= 1 - q_{ik} \end{aligned}$$

Modèle 2 :

$$\begin{aligned}
 m_{\theta_{S_k}}(H_i) &= 0 \\
 m_{\theta_{S_k}}(\bar{H}_i) &= q_{ik} \cdot [1 - p_{\theta_{S_k}}(H_i)] \\
 m_{\theta_{S_k}}(\theta_{S_k}) &= 1 - q_{ik} \cdot [1 - p_{\theta_{S_k}}(H_i)]
 \end{aligned}$$

Ces deux modèles répondent aux exigences théoriques et utilisent les éléments focaux $H_i, \bar{H}_i, \theta_{S_k}$. De plus, Appriou a montré que le choix du facteur de confiance q_{ik} ne posait pas de problème particulier, il peut être choisi de façon relativement empirique sans que cela ne perturbe le système. Pour des densités peu représentatives de la réalité, il sera fixé à 0.9. Pour des densités correspondant bien à la réalité, il sera fixé à 1.

2.5. prise de décision

La prise de décision consiste à retenir l'hypothèse la plus vraisemblable H_i^* telle que :

$$Pl_{\theta}(H_i^*) = \text{Max}_{i \in \{1, \dots, N\}} [Pl_{\theta}(H_i)] \quad (19)$$

On parle, dans ce cas, de règle du maximum de plausibilité. Nous optons pour ce critère car la méthode de correction des plausibilités ne peut déduire que les plausibilités des hypothèses singletons. Elle est par ailleurs très communément utilisée.

2.6. utilisation de la fonction de communalité

La fonction de communalité décrite au paragraphe 2.1 est utilisée afin de simplifier les calculs de la combinaison. Pour cela, nous utilisons la transformation de Möbius, décrite dans [13], qui est fondée sur une transformée de Fourier généralisée.

Grâce à la fonction de communalité et à la transformation de Möbius on peut obtenir directement la fonction de communalité du jeu de masse combiné par :

$$Q_{\theta}(A) = \prod_{k=1}^M Q_{\theta_{S_k}}(A) \quad (20)$$

On déduit le jeu de masse combiné par la transformation :

$$\forall A \in \theta \quad m_{\theta}(A) = \sum_{A \subseteq X} (-1)^{|X-A|} Q_{\theta}(X) \quad (21)$$

Cette fonction est utilisée à la suite d'un déconditionnement par la méthode du minimum d'affectation.

3. application : combinaison de segmentations

3.1. introduction

Notre objectif est de réaliser la combinaison de deux segmentations, l'une issue d'une image de réflectance (2D, source S_1), l'autre issue d'une image de profondeur (3D, source S_2). La figure 3 présente le principe général de l'algorithme. Chaque image est tout d'abord segmentée. Puis, l'ensemble des régions est

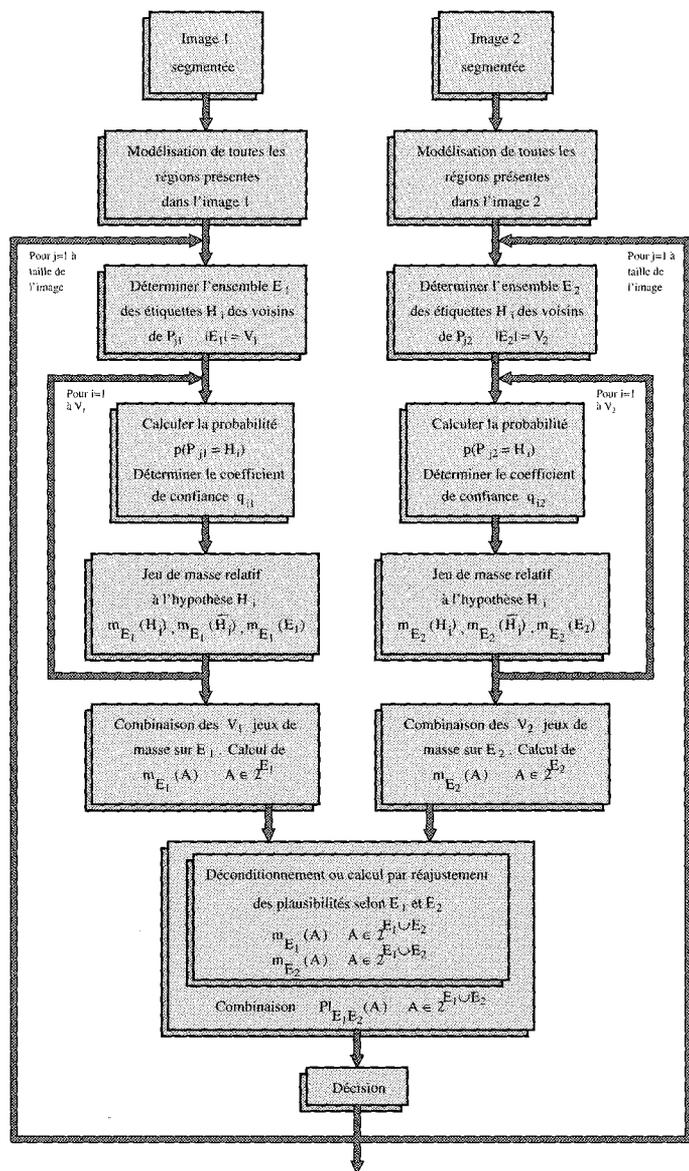


Figure 3. – Principe de la combinaison de deux segmentations.

caractérisé à partir d'une distribution gaussienne des attributs des pixels appartenant à chaque région. Les attributs sont : le niveau de gris pour l'image 2D et la distance du point 3D au plan auquel il est affecté pour l'image 3D. Un parcours de l'image permet de calculer, pour chaque pixel, la probabilité d'appartenance de celui-ci aux régions connexes à la sienne, on parlera de « régions voisines ». Cette probabilité dépend de l'attribut du pixel et des caractéristiques des régions. Dans le même temps un coefficient de confiance, relatif à chaque probabilité, est déterminé. On en déduit alors un jeu de masse. Ce calcul est répété pour toutes les régions voisines, l'ensemble des jeux de masse est ensuite combiné. Le jeu de masse final est obtenu en combinant les jeux de masse obtenus pour chaque segmentation. Un module de décision permet finalement d'étiqueter le pixel avec le label de la région qui convient le mieux. Avant de revenir en détails sur les principales phases de cet algorithme, nous précisons l'analogie entre les termes de la théorie de l'évidence et les termes employés en traitement d'images : les hypothèses sont les étiquettes des régions, le cadre de discernement est l'ensemble des étiquettes des régions voisines. Précisons enfin que P_{jk} est l'étiquette du pixel numéro j de l'image k , c'est aussi l'étiquette de la région à laquelle il appartient. L'étiquette d'une région i sera notée H_i .

3.2. détermination de l'ensemble E_k

Pour chaque pixel j appartenant à une région i de l'image k , l'ensemble E_k est déterminé. En toute rigueur on devrait utiliser la notation E_{ik} ; on ne l'utilisera pas afin de simplifier les notations. Cet ensemble est composé de l'étiquette de la région i et des étiquettes des régions voisines à la région i . Le nombre de régions est noté V_k , il représente le cardinal de E_k ($V_k = |E_k|$).

3.3. calcul des probabilités d'appartenance et des coefficients de confiance

Chaque région i d'une image k est modélisée par une distribution normale d'un attribut X (cf. Eq. 22). L'attribut retenu pour l'image de réflectance est le niveau de gris. Pour l'image de profondeur, il s'agit de la distance du point 3D au plan auquel il est affecté. Notons que cette modélisation pourrait s'appliquer à un vecteur de paramètres.

$$f_i(X) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(X-\mu_i)^2}{2\sigma_i^2}} \quad (22)$$

où μ_i est la moyenne et σ_i l'écart-type de la distribution gaussienne qui modélise la distribution de l'attribut sur une région i .

Puisque l'intégrale de $f_i(X)$ vaut 1, la probabilité pour que l'étiquette qu'un pixel j soit égale à l'étiquette d'une région i ,

connaissant la valeur de l'attribut val_j , peut être calculée.

$$I = \int_{val_j}^{\mu_i} f_i(X) dX \quad (23)$$

$$p(P_{jk} = H_i) = 1 - 2|I| \quad (24)$$

Cette formule garantit bien $0 \leq p(P_{jk} = H_i) \leq 1$. En normalisant par la somme des p on garantit également que

$$\sum_i p(P_{jk} = H_i) = 1.$$

Le coefficient q_{ik} , relatif à la probabilité $p(P_{jk} = H_i)$, est ensuite déterminé en fonction de la valeur de l'écart-type de la distribution gaussienne relevée pour la région i . Pour fixer ce coefficient, on choisit une fonction telle que :

- $q_{ik} = 1$ pour un écart-type minimum (σ_{min});
- $q_{ik} = q_{min} = 0.9$ pour un écart-type maximum (σ_{max}).

σ_{min} (σ_{max}) est le minimum (maximum) sur l'ensemble des régions de l'image.

L'évolution de q_{ik} est linéaire (cf. figure 4b).

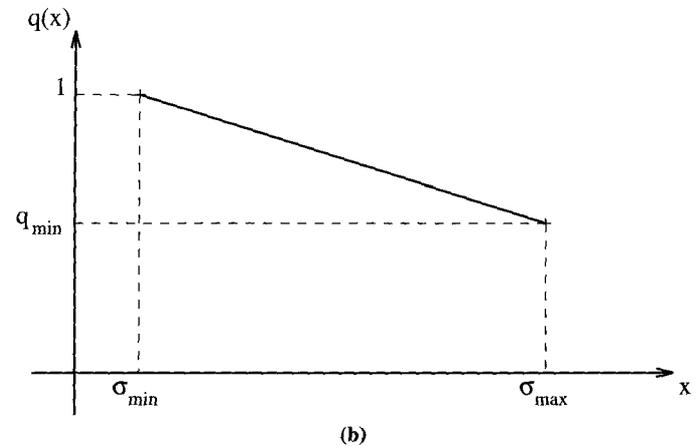
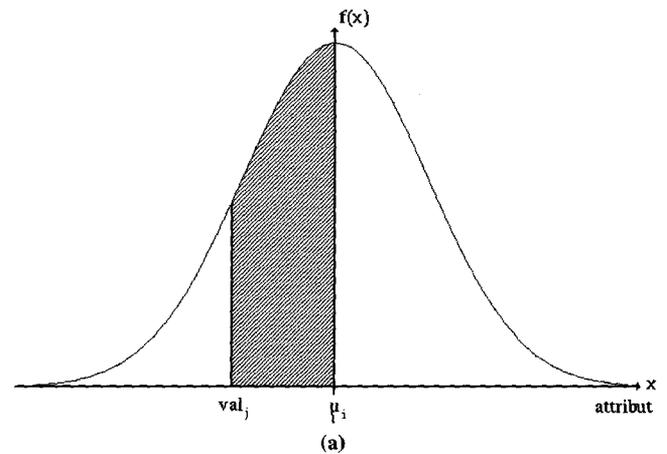


Figure 4. - a) Distribution normale $f(x)$, b) : fonction $q(x)$ permettant de fixer les coefficients de confiance.

Avec l'ensemble des probabilités et des coefficients de confiance, le jeu de masse relatif à chacune des régions i peut alors être déterminé.

3.4. calcul des jeux de masse

Le but est ici d'obtenir un jeu de masse représentatif de l'information sur le pixel traité. Un jeu de masse est calculé en utilisant l'un des modèles proposé par Appriou (cf. § 2.4). Pour chaque région i ($i \in \{1, \dots, V_k\}$) on détermine le jeu de masse sur les trois ensembles H_i , \bar{H}_i et E_k . E_k est l'ensemble composé de toutes les étiquettes des régions, H_i représente l'ensemble composé de l'étiquette de la région considérée et \bar{H}_i est le complémentaire de H_i dans E_k . Le jeu de masse est alors le suivant :

$$m_{E_k}(H_i), m_{E_k}(\bar{H}_i), m_{E_k}(E_k) \quad (25)$$

On dispose alors de V_k jeux de masse qu'il faut combiner.

3.5. combinaison des jeux de masse

La combinaison des jeux de masse est effectuée en deux temps. Tout d'abord, une combinaison est réalisée au niveau de chaque pixel relativement à l'ensemble des régions voisines E_k , cela sur les deux segmentations. Puis, une combinaison globale est faite entre les deux jeux de masse résultant de la première étape.

Première étape : combinaison au niveau pixel sur E_k

D'une façon générale, cette étape consiste en la combinaison des V_k jeux de masse sur E_k afin d'obtenir le jeu $m_{E_k}(A)$, $A \in 2^{E_k}$. Cette phase de combinaison est réalisée en utilisant l'apport de la fonction de communalité associée à la transformation de Möbius. Une telle transformation permet de réduire considérablement les calculs de combinaison, surtout lorsque le nombre de jeux de masse à combiner est supérieur à deux. Dans ce cas, au lieu de combiner les jeux de masse, il suffit de multiplier les fonctions de communalité (cf. Eq. 8) pour obtenir la fonction de communalité représentative de la combinaison. A l'issue de cette première étape, on dispose donc de deux jeux de masse, $m_{E_1}(A)$ et $m_{E_2}(A)$ avec $A \in 2^{E_1}$ et $A \in 2^{E_2}$.

Seconde étape : combinaison globale sur $\cup E_k$

La seconde étape consiste à combiner les jeux obtenus pour chaque segmentation. La difficulté de cette phase tient au fait que les deux ensembles E_1 et E_2 ne sont pas obligatoirement exhaustifs, c'est-à-dire que pour un même pixel j , les régions voisines dans les deux segmentations ne sont pas forcément les mêmes. Or, comme nous l'avons brièvement rappelé au paragraphe 2, la règle de combinaison de Dempster requiert des cadres de discernement exhaustifs donc des ensembles E_1 et E_2

identiques. Si cette condition n'est pas respectée ($E_1 \neq E_2$), une étape de réajustement des cadres de discernement est impérative.

Les méthodes utilisées sont :

- le **minimum d'affectation**, lorsque les deux ensembles E_1 et E_2 n'ont pas de régions communes ($E_1 \cap E_2 = \phi$);
- la **correction des plausibilités**, lorsque les ensembles E_1 et E_2 ont des régions en commun ($E_1 \cap E_2 \neq \phi$, $E_1 \neq E_2$). Cette dernière condition est en effet l'une de celles nécessaires à l'application de cette méthode.

Si l'on a utilisé le minimum d'affectation ou si les ensembles d'étiquettes sont identiques, les jeux de masse sont combinés et les plausibilités de chaque étiquette déduites. Dans les autres cas, la correction des plausibilités est utilisée et permet d'obtenir directement les plausibilités. Ainsi, à l'issue de cette deuxième étape de fusion, on obtient les plausibilités combinées $Pl_{E_1 E_2}(A)$, soit directement, soit par l'intermédiaire du jeu de masse combiné $m_{E_1 E_2}(A)$ avec $A \in 2^{E_1 \cup E_2}$.

3.6. module de décision

L'étiquette d'un pixel issue du module de décision est celle ayant la plausibilité maximale, elle est déterminée parmi l'ensemble des régions i avec $i \in \{1, \dots, |E_1 \cup E_2|\}$. Cette méthode ne garantit pas la connexité des régions. Une phase de réétiquetage est nécessaire à la suite de laquelle les petites régions (nombre de pixels inférieur à un seuil α_p) sont éliminées. Elles sont intégrées avec la région voisine qui présente la plus grande frontière commune.

4. résultats

4.1. introduction

Dans cette partie les résultats qualitatifs et quantitatifs obtenus avec notre méthode sont présentés. La présentation qualitative se borne aux résultats obtenus avec le modèle 1 de modélisation des jeux de masse et avec filtrage des petites régions. En revanche, l'évaluation quantitative et comparative fera apparaître les résultats obtenus avec les deux modèles, avec et sans filtrage des petites régions.

4.2. présentation des images et des segmentations initiales

A l'aide de la méthodologie proposée dans [14], nous avons mené une évaluation quantitative et comparative des résultats sur un

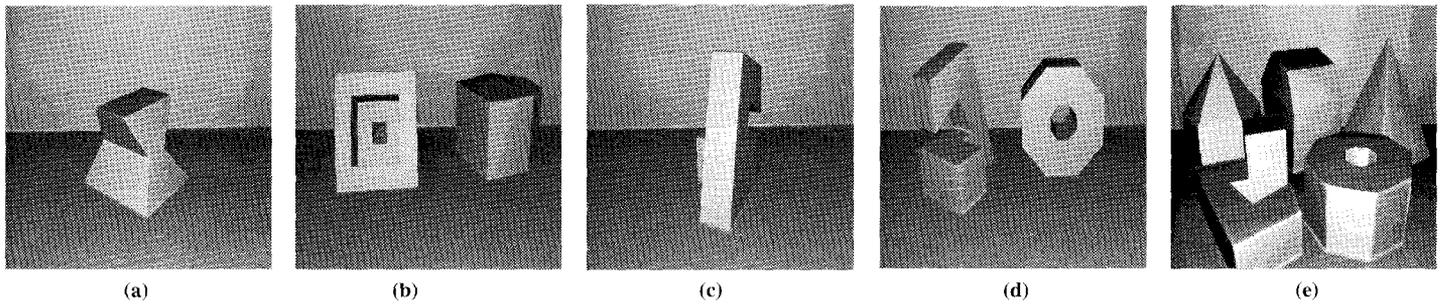


Figure 5. – Les images de réflectance.

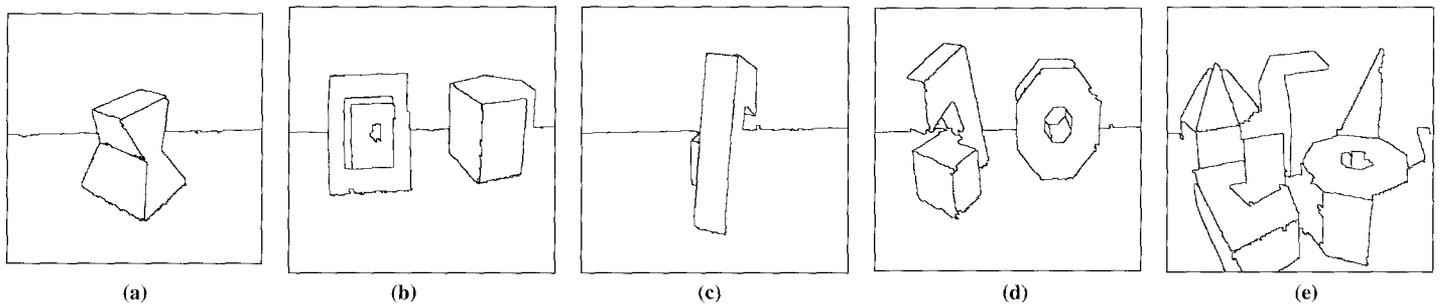


Figure 6. – Les segmentation 2D.

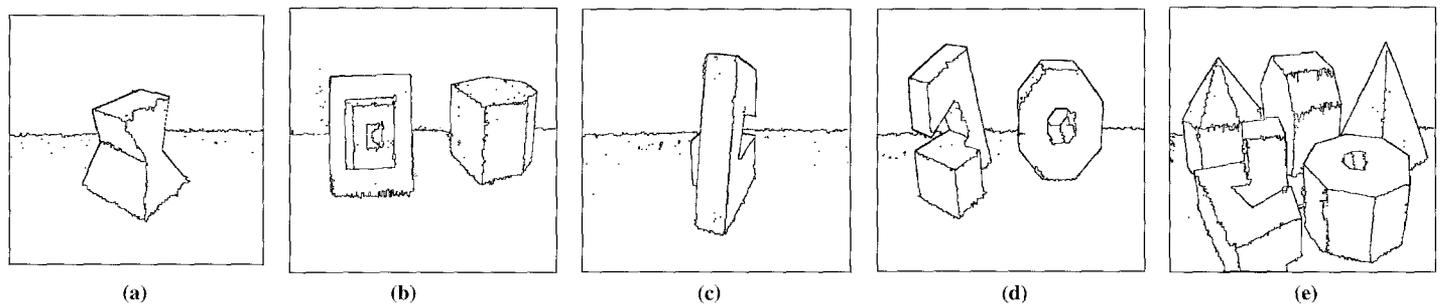


Figure 7. – Les segmentation 3D.

ensemble de 40 images réelles issues d'un télémètre laser nommé Perceptron [15].

Ces images représentent des objets composés de surfaces planes, de différentes tailles, de l'ordre du mètre, placés à moins de 20 mètres du capteur. Elles contiennent jusqu'à cinq objets polyédriques placés dans des attitudes variées. Ainsi, le nombre de régions croît avec le nombre d'objets dans la scène, et par conséquent, la taille des régions diminue. Pour chacune des scènes imagées nous disposons d'une image de réflectance (image 2D), d'une image de profondeur (image 3D) mais également d'une segmentation de référence appelée « vérité de terrain ». C'est grâce à cette segmentation de référence que nous pouvons évaluer nos résultats. Pour ce lot de 40 images, 10 images de mise au point ont été extraites par tirage au sort afin de régler les paramètres nécessaires à l'algorithme. Nous n'avons pas utilisé cette facilité car les deux seuls seuils utilisés (q_{min} et α_p) sont très peu sensibles. Les 30 autres images sont destinées à l'évaluation (images *test*). Pour ces 30 images *test*, les segmentations de référence totalisent 438 régions.

Les images initiales, réflectance et profondeur, provenant du même capteur sont intrinsèquement recalées. Par contre, les étiquettes des régions des deux segmentations sont mises en correspondance en utilisant la mesure de Vinet.

La figure 5 montre quelques-unes des images de réflectance réelles utilisées pour l'évaluation de notre algorithme. Les segmentations 2D, présentées figure 6, 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 7 sont obtenues avec un algorithme de segmentation d'images de profondeur en surfaces planes, développé par Checchin et al [16].

A chaque image réelle (cf. figure 5) est associée une vérité de terrain (cf. figure 8) définie manuellement par un opérateur.

L'opérateur humain délimite la frontière de chaque surface de la scène dans l'image. Pour l'aider dans ce travail, il dispose

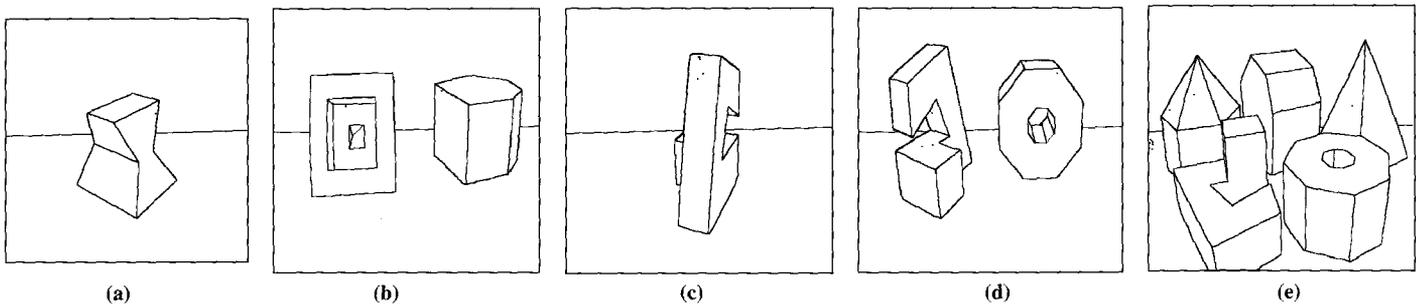


Figure 8. – Les vérités de terrain.

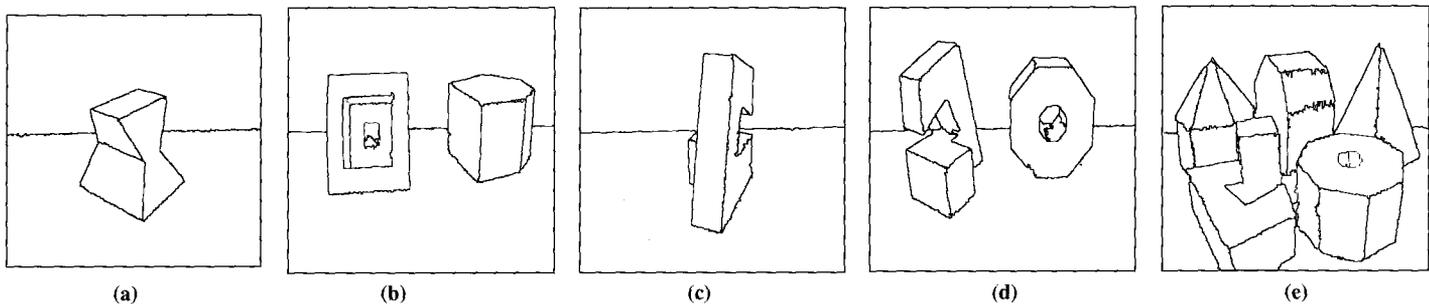


Figure 9. – Les segmentations résultat.

d'outils informatiques permettant de classer correctement chaque pixel. L'opérateur a également la possibilité d'étiqueter des pixels comme étant du bruit s'il considère que par manque d'information il n'est pas concevable qu'un algorithme puisse lui attribuer l'étiquette correcte.

Ces quelques segmentations, représentées par les contours des régions, font apparaître les principaux défauts rencontrés : le problème des régions sous-segmentées en 2D, nettement visible sur la troisième image (cf. figure 6c); l'imprécision des contours, surtout visible en 3D. Nous verrons par la suite que la méthode proposée permet d'améliorer ces aspects dans la plupart des cas traités.

4.3. résultats qualitatifs de la combinaison

Nous présentons, dans cette partie, les résultats (cf. figure 9) obtenus par la méthode proposée, correspondant aux images réelles de la figure 5.

Le premier résultat (cf. figure 9a) est satisfaisant. L'amélioration apportée est nettement visible surtout si on le compare à la segmentation 3D. On peut remarquer que les bords de celle-ci ne sont pas très précis. La segmentation 2D est plutôt bonne, ceci assure un bon résultat de combinaison.

Le deuxième résultat (cf. figure 9b) est intéressant dans le sens où il montre que notre méthode permet de retrouver des régions qui avaient été sous-segmentées par l'une ou l'autre des segmentations de départ, en l'occurrence ici la segmentation 2D. En effet,

la segmentation 2D n'a pas réussi à détecter la partie supérieure ainsi qu'une face latérale de l'objet de droite. La combinaison permet également d'améliorer les bords des objets par rapport à la segmentation 3D.

Un commentaire similaire peut être fait sur le troisième résultat (cf. figure 9c). Ici des régions sous-segmentées sont présentes dans les deux segmentations de départ.

Le quatrième résultat présenté (cf. figure 9d) permet de voir les limites actuelles de notre méthode. En effet, 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. Dans l'image de référence (cf. figure 8d), on observe cinq régions. Les segmentations 2D et 3D contiennent quatre régions, non identiques. Le résultat fait également 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 de façon convenable 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 9e) synthétise l'ensemble des remarques faites précédemment. Il s'agit de l'image la plus complexe de la base. On peut ajouter pour cette image que la vérité de terrain comporte 27 régions. La segmentation 2D en fait apparaître seulement 17, contre 26 pour la segmentation 3D ainsi que pour notre résultat. Ce résultat n'est certes pas parfait mais il est plus proche de la vérité que la segmentation 2D (en terme

de régions) et que la segmentation 3D (en terme de précision des contours).

Ces résultats montrent l'apport de notre algorithme au niveau des contours des objets lorsque ceux-ci ne sont pas bien définis par l'une des segmentations initiales. En revanche, des problèmes peuvent apparaître en cas de sous-segmentation comme de sur-segmentation. Actuellement notre algorithme ne règle pas systématiquement ce type de problème.

4.4. évaluation quantitative et comparative

Afin de caractériser les performances de notre algorithme, nous avons mené une évaluation quantitative et comparative des résultats de la combinaison de segmentations. Quatre évaluations ont été menées. Deux avec le modèle 1 de modélisation des jeux de masse, avec et sans filtrage des petites régions, et deux avec le modèle 2, avec et sans filtrage des petites régions. La méthode d'évaluation utilise la mesure de Hoover et al [14]. Cette mesure compare la vérité de terrain (VT) avec une image de régions (IR). Son principe est le suivant.

Soient :

- M le nombre de régions dans IR;
- N le nombre de régions dans VT;
- P_m le nombre de pixels de la région R_m (avec $m \in \{1, \dots, M\}$) dans IR;
- P_n le nombre de pixels de la région R_n (avec $n \in \{1, \dots, N\}$) dans VT.

On pose $O_{mn} = |R_m \cap R_n|$, le nombre de pixels appartenant à la fois à la région R_m et à la région R_n . S'il n'y a pas recouvrement entre les deux régions $O_{mn} = 0$, alors que si le recouvrement est total $O_{mn} = P_m = P_n$.

Une table de superposition des étiquettes TSE est déterminée comme suit :

$$TSE(m, n) = O_{mn}, m \in \{1, \dots, M\}, n \in \{1, \dots, N\}$$

On connaît donc le taux de recouvrement de chaque région. Par exemple, O_{mn}/P_m représente la proportion de pixels communs à R_m et à R_n recouvrant la région R_m . De la même façon O_{mn}/P_n représente la proportion de pixels communs à R_m et à R_n recouvrant la région R_n . Ces pourcentages sont utilisés afin de juger de la qualité d'une segmentation.

Une région sera considérée comme correctement segmentée (CS) si :

$$O_{mn} \geq T \times P_m \text{ et } O_{mn} \geq T \times P_n$$

où T est le taux de recouvrement mutuel minimum à atteindre.

Si une région n'est pas correctement segmentée, elle peut être classée « sur-segmentée » (Sur-Seg), « sous-segmentée » (Sous-Seg), « manquée » (M) ou « parasite » (P). Les mesures qui

définissent ces classifications sont détaillées dans [14]. La mesure de Hoover et al est appliquée sur l'ensemble des 30 images *test*.

Les figures 10 et 11 montrent l'évolution du nombre de régions correctement classées en fonction de T ($T \in [0.51, 1.00]$) respectivement pour le modèle 1, avec et sans filtrage des petites régions et pour le modèle 2, également avec et sans filtrage des petites régions. Nous positionnons ces résultats par rapport aux meilleurs scores à notre disposition soient ceux obtenus par l'algorithme 3D seul. A partir de l'analyse de ces résultats nous avons conservé le modèle 1 qui présente des scores meilleurs que le modèle 2 quel que soit le taux de comparaison. Le filtrage des petites régions est évidemment conservé, il permet d'obtenir des segmentations de meilleure qualité même si l'apport de la méthode est également présent sans filtrage.

La figure 12 montre les résultats obtenus sur les segmentations 2D et 3D, ainsi que ceux obtenus par l'Université d'Edinburgh

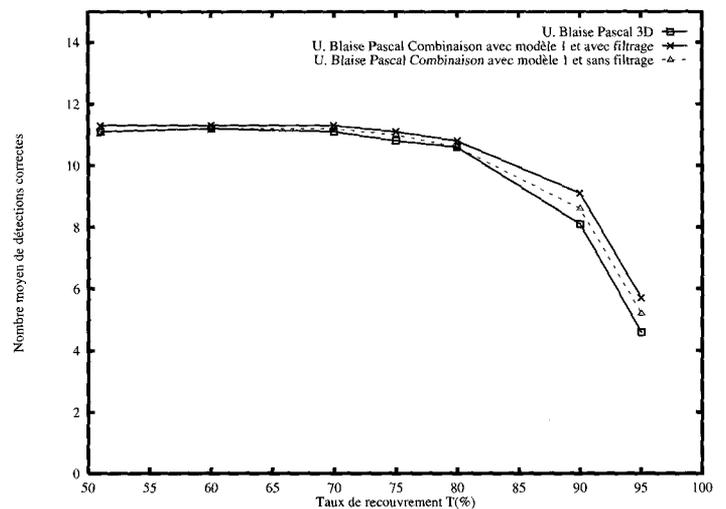


Figure 10. - Evolution des détections correctes avec le modèle 1.

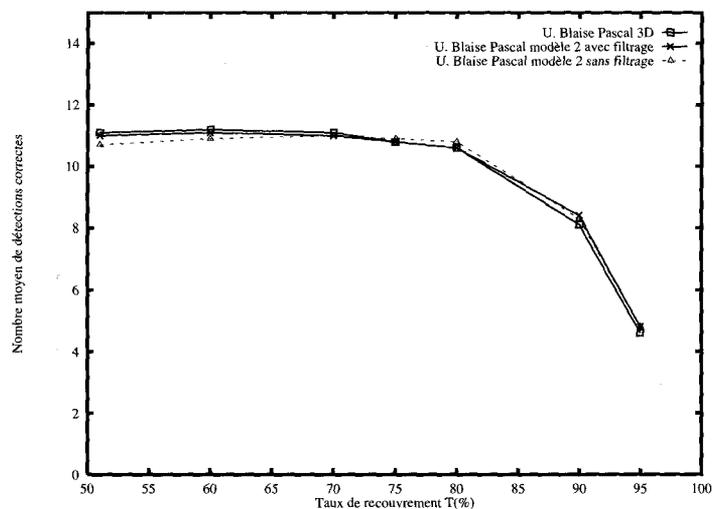


Figure 11. - Evolution des détections correctes avec le modèle 2.

[17] avec une segmentation 3D. 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 [14]. Seuls les résultats de l'Université d'Edinburgh, présentant les scores les plus élevés, sont reportés sur la courbe (cf. figure 12). Nous avons choisi ces résultats dans le but de situer les performances de notre méthode. Sur la figure 12, apparaît également le nombre de régions dans la vérité de terrain (Idéal).

Aucun des différents algorithmes de segmentation évalués ne réalise, sur aucune des images, un score parfait. Les progrès et les améliorations qu'il reste à réaliser pour approcher de la solution optimale sont encore nombreux.

Les courbes montrent que pour un taux de 51% ($T = 0.51$), le nombre de régions correctement segmentées n'atteint au mieux que 77% de la valeur idéale. Pour un taux de recouvrement plus élevé ($T = 0.75$), ce nombre passe à 76%. Au-delà de 80% ($T > 0.80$), il décroît fortement. Dans tous les cas, les scores de combinaison sont les plus élevés. Même si au-delà d'un taux de 80% les scores diminuent, il est intéressant de noter que les résultats de combinaison sont encore satisfaisants et ce malgré la qualité moyenne des segmentations 2D. Cette dernière remarque confirme l'intérêt de la fusion. En effet, même si pour un taux de 90% le score de la combinaison n'est plus que de 62% de régions correctement classées, il reste cependant supérieur de 7% au score obtenu par la méthode 3D (soit une augmentation relative de 13%). Si on se place à un taux de recouvrement de 95%, notre 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.

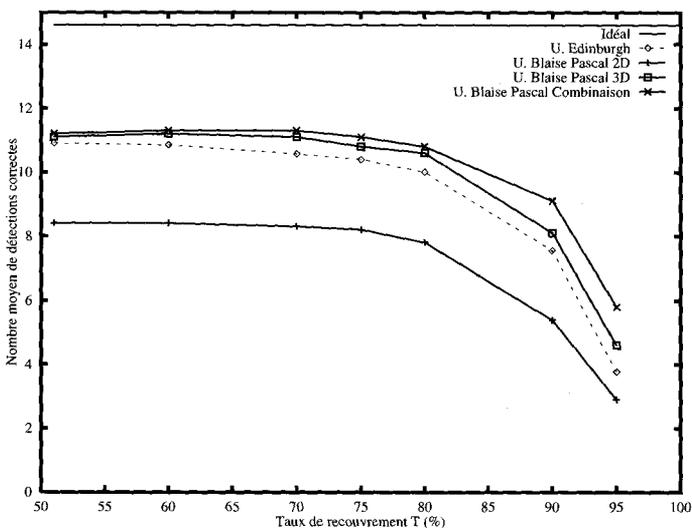


Figure 12. – Evolutions des détections correctes (CS) en fonction du taux de recouvrement T pour les images « test » issues du capteur PERCEPTRON.

Tableau 1. – Pourcentage de régions correctement segmentées par rapport à la valeur idéale en fonction de plusieurs valeurs du taux de recouvrement.

Taux de recouvrement (%)	51	75	90	95
Segmentation 2D	58	56	37	20
Segmentation 3D	76	74	55	32
Université d'Edinburgh	75	72	52	26
Combinaison (modèle 1 petites régions filtrées)	77	76	62	40
Combinaison (modèle 1 petites régions non filtrées)	76	75	59	36
Combinaison (modèle 2 petites régions filtrées)	75	74	58	33
Combinaison (modèle 2 petites régions non filtrées)	73	75	57	33

Nous voyons que l'apport de la combinaison se fait surtout sentir lorsque le taux de recouvrement est important. En dessous de 80%, on améliore sensiblement les résultats; au-delà de 80% l'amélioration est très nette.

Les courbes illustrant les résultats obtenus sur les segmentations 2D sont à un niveau très inférieur aux autres. Ceci s'explique par le fait que de nombreuses régions sont sous-segmentées en 2D. L'algorithme de segmentation 2D n'a pas fait l'objet d'une attention particulière. Nous avons utilisé un algorithme classique (cf. § 4.2), retenu sur la qualité visuelle des résultats.

L'ensemble des commentaires précédents sont relatifs aux résultats obtenus avec le modèle 1 de modélisation du jeu de masse et avec filtrage des petites régions. Cependant le tableau 1 résume l'ensemble des valeurs calculées pour les deux modèles.

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 sont très souvent améliorés. 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 testées. Même si nous avons opté pour un modèle, parmi les deux disponibles, il est intéressant de noter que ce choix ne remet pas en cause la méthode. En effet les deux modèles donnent des résultats corrects lorsque le taux de recouvrement devient important.

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.

BIBLIOGRAPHIE

- [1] Bloch I., « 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] Huet F. and Philipp S., « Fusion of Images After Segmentation by Various Operators and Interpretation by a Multi-Scale Fuzzy Classification ». In *14th International Conference on Pattern Recognition*, Brisbane, Queensland, Australia, 17-20 August 1998, ICPR98.
- [3] Nadabar S. G. and Jain A. K., « Edge Detection and Labeling by Fusion of Intensity and Range Images », *Applications of Artificial Intelligence : Machine Vision and Robotics*, 1708 :108-119, 1992.
- [4] Abdulghafour M. and Abidi M. A., « Data Fusion Through Non-Deterministic Approaches - A Comparison », *SPIE Sensor Fusion VI*, 2059 :37-53, 1993.
- [5] Gil B., Mitiche A. and Aggarwall J.K., « Experiments in Combining Intensity and Range Edge Maps », *Computer Vision, Graphics, and Image Processing*, 21 :395-411, 1983.
- [6] Kara-Falah R., « Segmentation d'images : coopération, fusion, évaluation », Thèse de Doctorat, Université de Savoie, Juin 1995.
- [7] Vinet L., « Segmentation et mise en correspondance des régions de paires d'images stéréoscopiques », Thèse de Doctorat, Université de Dauphine, Paris IX, 1991.
- [8] Dempster A.P., « A Generalization of Bayesian Inference », *Journal of the royal statistical society*, 30(B), 1968.
- [9] Shafer G., « *A Mathematical Theory of Evidence* », Princeton University Press, 1976.
- [10] Smets P., « The Combination of Evidence in the Transferable Belief Model », *IEEE transactions on pattern analysis and machine intelligence*, 12(5) :447-458, 1990.
- [11] Janez F., « Fusion de sources d'information définies sur des référentiels non exhaustifs différents », Thèse de Doctorat, Université d'Angers, Novembre 1996.
- [12] Appriou A., « Probabilités et incertitude en fusion de données multi-senseurs », *Revue scientifique et technique de la défense*, p. 27-40, Novembre 1991.
- [13] Kennes R., « Computational Aspects of the Möbius Transformation of Graphs », *IEEE transactions on systems man and cybernetics*, 22(2) :201-223, March April 1992.
- [14] Hoover A., Jean-Baptiste G., Jiang X., Flynn P.J., Bunke H., Goldgof D., Bowyer K.W., Eggert D., Fitzgibbon A. and Fisher R., « An Experimental Comparison of Range Image Segmentation Algorithms », *IEEE Transactions on PAMI*, 18(7) :673-689, 1996.
- [15] Perceptron Inc, 23855 Research Drive, Farmington Hills, Michigan 48335, « *LASAR Hardware Manual* », 1993.
- [16] Checchin P., Trassoudaine L. and Alizon J., « Segmentation of Range Images into Planar Regions ». In *IEEE International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, p. 156-163, 12-15 May 1997, Ottawa Ontario Canada.
- [17] Trucco E. and Fisher R.B., « Experiments in Curvature-Based Segmentation of Range Data », *IEEE Transactions on PAMI*, 17(2) :177-182, 1995.

Manuscrit reçu le 20 juillet 1998.

LES AUTEURS

Stéphane MATHEVET



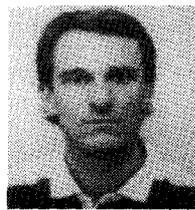
Titulaire d'un DEA Electronique et Systèmes, option Vision pour la Robotique, obtenu en 1995 à l'Université Blaise Pascal de Clermont-Ferrand. Il prépare actuellement une thèse de Doctorat sur la fusion de données appliquée à la combinaison de segmentations au sein du Groupe Automatique : Vision et Robotique (GRAVIR) du Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA), UMR 6602 du CNRS.

Paul CHECCHIN



Paul Checchin a effectué sa thèse de doctorat d'université, soutenue en 1996, au LASMEA, UMR 6602 du CNRS. Depuis 1997, il est maître de conférences à l'IUP d'Evry, Université d'Evry Val d'Essonne, où il intervient dans les enseignements d'informatique industrielle et du traitement d'images. Il exerce ses activités de recherche au sein du groupe Systèmes Complexes du Centre d'Etudes de Mécanique d'Ile-de-France (CEMIF). Ses centres d'intérêts concernent les systèmes de perception 3D de l'environnement et l'analyse des données issues de ces capteurs.

Laurent TRASSOUDAINE



Ingénieur du CUST (1990) et docteur de l'université Blaise Pascal (1993). Il est maître de conférences à l'I.U.P. Génie des Systèmes Industriels de cette université depuis 1993. Son activité de recherche est consacrée à l'analyse d'images de profondeur et à la fusion de données multisensorielles.

Joseph ALIZON



Chargé de Recherche au CNRS au Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA). Il est co-responsable du projet Aide à la Conduite au sein du Groupe Automatique : Vision pour la Robotique (GRAVIR). Ses centres d'intérêts portent sur les Capteurs de Vision Intelligents et la Fusion de Données.