

# Segmentation d'images de profondeur : une approche pyramidale

Paul Checchin<sup>(1)</sup>, Laurent Trassoudaine<sup>(2)</sup> et Joseph Alizon<sup>(2)</sup>

<sup>(1)</sup>HEUDIASYC, UMR CNRS 6599,  
Université de Technologie de Compiègne,  
BP 20529 – 60205 Compiègne cedex, FRANCE

<sup>(2)</sup> LASMEA UMR CNRS 6602,  
Université Blaise Pascal de Clermont-Ferrand,  
F-63177 Aubière cedex, FRANCE

## RÉSUMÉ

Cet article présente un algorithme de segmentation d'images de profondeur en surfaces planes. Les originalités de la méthode reposent sur le procédé d'estimation des attributs différentiels et sur le choix des germes de la croissance de régions. L'algorithme procède en deux temps. Dans un premier temps, les pixels sont regroupés en régions homogènes au sens d'attributs locaux déduits des données initiales. Les régions extraites sont représentées par un graphe d'adjacence. Dans un second temps, une stratégie pyramidale de fusion est mise en oeuvre pour aboutir au résultat de la segmentation. A l'aide de la méthodologie proposée dans [4], une évaluation quantitative des résultats a été menée sur un jeu de 80 images réelles issues de deux capteurs différents. Les résultats de cette étude sont présentés et comparés à ceux obtenus par d'autres techniques.

## ABSTRACT

This paper deals with the segmentation of range images into planar regions. The original aspects of our method are based on the computation of pixel's normal and the seed region selection for the region growing process. The segmentation algorithm consists of two parts. In the first one, the pixels are aggregated according to local properties derived from the input data. The extracted regions are represented by a region adjacency graph. In the second part, the segmentation result is achieved via a merging strategy based on the construction of an irregular pyramid. This algorithm has been evaluated on 80 real images acquired by two different range sensors using the methodology proposed in [4]. Experimental results are presented and compared to others obtained by other techniques.

## 1 Introduction

Nous avons conçu un algorithme adapté à la segmentation des données fournies par des capteurs de profondeur actifs, appelées *images de profondeur*. Ces informations tridimensionnelles sont une représentation point par point des surfaces ( $z = f(x_i, y_j)$ ) observées par ces capteurs. Dans cet article, nous nous limitons à l'analyse de scènes constituées d'objets uniquement polyédriques.

La double originalité de cet algorithme repose d'une part sur l'estimation des normales en chacun des pixels de l'image et d'autre part sur le choix des germes de la croissance de régions. Les régions extraites sont représentées par un graphe d'adjacence et forment la base d'une pyramide de graphes. La première étape consiste à regrouper les points constituant les lignes et les colonnes de l'image en segments par une technique d'approximation polygonale. Une estimation des attributs différentiels du premier ordre et de la normale à la surface en chaque pixel sont alors déduites. Afin d'agrèger les pixels, nous proposons de regrouper en régions les pixels connexes homogènes au sens de l'attribut normale et vérifiant un critère sur la longueur des segments horizontaux et verticaux auxquels ils appartiennent. A partir du graphe d'adjacence des régions ainsi obtenues, une stratégie pyramidale de fusion est mise en oeuvre pour aboutir au résultat de la segmentation.

Dans un premier temps, nous décrivons la méthode (cf. d' 2 et d' 3), puis nous présentons des résultats sur des images réelles issues de capteurs différents (cf. d' 4). Au paragraphe 5, ces résultats sont comparés avec ceux obtenus par d'autres chercheurs du domaine. Les résultats présentés enrichissent par conséquent une étude comparative de plusieurs méthodes dédiées au problème.

## 2 Notre approche

### 2.1 Estimation des normales

Un critère simple pour distinguer deux surfaces adjacentes d'un même objet polyédrique est de comparer leur normale respective. Si nous sommes capables d'estimer la normale à la surface en chaque pixel, nous pouvons regrouper tous les pixels connexes d'une même région en comparant leur attribut « normale ».

Le vecteur normal  $N$  en  $P$ , point de coordonnées  $(x_i,$

$y_j, z = f(x_i, y_j)$ ), est donné par :

$$\begin{aligned}
 N &= \frac{\partial OP}{\partial x} \wedge \frac{\partial OP}{\partial y} \\
 &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \\ z_x = f'_x(x_i, y_j) & z_y = f'_y(x_i, y_j) \end{vmatrix} \quad (1)
 \end{aligned}$$

où  $z_x = f'_x(x_i, y_j)$  et  $z_y = f'_y(x_i, y_j)$  sont les dérivées partielles en  $P$  de la fonction  $z = f(x_i, y_j)$  selon  $x$  et  $y$ , respectivement.

A la différence d'approches conventionnelles, la méthode proposée ne procède pas directement par une analyse bidimensionnelle de l'image. Au lieu de cela, elle débute par l'analyse de profils 1D définis par les directions des deux axes  $x$  et  $y$ , respectivement. Ces deux directions ne sont autres que les lignes et les colonnes de l'image. En effet, pour calculer le vecteur normal en chaque point, les attributs différentiels du premier ordre, selon ces deux directions de paramétrisation, suffisent (cf. Eq. 1). Avant d'estimer ces derniers, les observations suivantes peuvent être faites :

- les points d'intersection d'un objet à surface plane d'équation  $z = Ax + By + C$  avec le plan  $y = k = y_j$  (où  $k$  n'est autre que le numéro ligne de l'image) forment un segment de droite d'équation :  $z = Ax + By_j + C = Ax + a$  ;
- une conséquence de cette remarque est que tous les pixels appartenant à ce segment ont pour dérivée première par rapport à  $x$  la pente  $A$  ;
- le même raisonnement peut être tenu pour les points formant les colonnes de l'image.

Une façon simple pour détecter ces segments est d'utiliser un algorithme d'approximation polygonale sur les profils 1D de l'image constitués des lignes mais aussi des colonnes de celle-ci. Après segmentation de ces courbes, les différentes dérivées partielles sont alors estimées et il est donc possible de déterminer la normale aux surfaces en chaque pixel de l'image.

Il reste à agréger les pixels en régions. Nous avons choisi de regrouper ceux-ci en ensembles homogènes au sens du paramètre d'orientation, donné par le vecteur normal.

## 2.2 Sélection des germes et extraction des régions

L'agrégation des pixels en régions, conduite selon un double critère d'homogénéité et d'adjacence, est initialisée par un semencement de pixels de départ, nommés germes.

### 2.2.1 Choix des germes

Un segment obtenu à l'issue de l'approximation polygonale (cf. d' 2.1) est caractérisé non seulement par son équation mais aussi par sa longueur. Or, plus la longueur du segment est importante, plus la valeur de la dérivée est sûre et donc plus l'estimation du vecteur normal est précise. Afin de choisir les germes des régions, il paraît donc naturel de sélectionner les pixels qui appartiennent à des segments horizontaux et

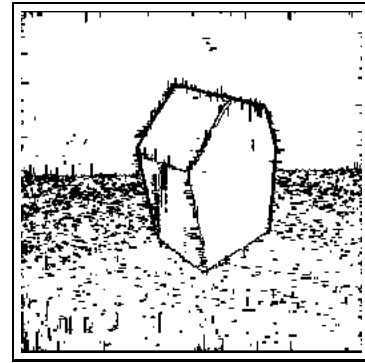


FIG. 1 — Résultat de la segmentation à l'issue de la croissance de régions fondée sur un critère d'homogénéité locale sur la colinéarité des vecteurs normaux. Le graphe d'adjacence correspondant constitue la base d'une pyramide de graphes.

verticaux d'une taille supérieure à un seuil préfixé. Ces segments de longueur importante correspondent à des zones faiblement bruitées de l'image. C'est pourquoi, pour avoir accès à ces pixels contenus dans ces zones, un tri de tous les pixels dans l'ordre des valeurs décroissantes d'un attribut  $\Sigma$  est effectué. Cet attribut  $\Sigma$  est calculé en sommant la longueur des deux segments orthogonaux (horizontal et vertical) auxquels appartient un pixel.

### 2.2.2 Extraction des régions

Le tri réalisé, on sélectionne alors le germe potentiel  $g$  d'une nouvelle région comme étant le point non affecté à une région déjà créée et dont la valeur  $\Sigma$  est maximale. La composante 4-connexe  $C_g$  autour de  $g$ , définie par les points satisfaisant un critère d'homogénéité locale sur la colinéarité des vecteurs normaux, est extraite.

La figure 1 présente les régions obtenues dans l'image à l'issue des traitements décrits ci-dessus. L'image est encore largement sur-segmentée, aussi nous avons recours à une étape de fusion. Les régions extraites sont représentées par un graphe d'adjacence et initialisent une pyramide de graphes.

## 3 La pyramide de graphes

La pyramide de graphes, également appelée pyramide irrégulière, est une approche multirésolution qui permet de segmenter une image en s'adaptant à son contenu [2].

### 3.1 Structure de la pyramide irrégulière

La pyramide irrégulière possède les caractéristiques suivantes :

- c'est un empilement de graphes d'adjacence, chacun d'eux correspondant à un niveau de la pyramide ;
- le niveau  $k + 1$  de la pyramide est construit à partir du niveau  $k$  ;
- chaque région de l'image, accompagnée de ses attributs, est modélisée par un sommet du graphe d'adjacence.

La pyramide au niveau  $k$  est un graphe non orienté, noté  $G[k]$ . Afin que cette construction tienne compte du contenu de l'image, un graphe orienté  $G'[k]$  est déduit de  $G[k]$ .  $G'[k]$  est appelé graphe de similarité. Il est obtenu en conservant dans  $G[k]$  les sommets « similaires » qui vérifient des contraintes locales.

La réduction du nombre de régions du niveau  $k$  au niveau  $k+1$  s'exprime dans notre approche orientée « graphes » par la réduction du nombre de sommets. Sur le graphe de similarité, les deux règles suivantes doivent être respectées :

**Règle 1** — Deux sommets adjacents au niveau  $k$  ne peuvent survivre tous les deux au niveau  $k+1$ .

**Règle 2** — Au niveau  $k$ , tout sommet non-survivant doit posséder au moins un sommet de son voisinage qui survit au niveau  $k+1$ .

### 3.2 Application à la segmentation d'images de profondeur

L'image (cf. figure 1) est encore largement sur-segmentée. Les régions qui composent celle-ci sont de deux types :

- des régions de taille importante possédant un nombre élevé de voisins ;
- des petites régions possédant peu de voisins, souvent encerclées par une seule grosse région.

La stratégie de fusion adoptée est de prendre en compte la taille des régions afin que les plus petites fusionnent avec les grandes. Le critère de similarité évalué localement que nous avons choisi est le suivant :

soit  $V(R_i)$  l'ensemble des régions voisines de  $R_i$ , avec  $R_i$  une région de taille  $N_i$  ; une région  $R_j \in V(R_i)$  et  $R_i$  peuvent fusionner si  $R_i$  est similaire à  $R_j$ , c'est-à-dire si  $C_{sim}(R_i, R_j) \leq C_{sim}(R_i, R_l), \forall R_l \in V(R_i) \mid N_l \geq N_i$  et si  $C_{sim}(R_i, R_j) < \varepsilon$  ; avec :

$$C_{sim}(R_i, R_j) = \frac{1}{N_i} \sum_{n=1}^{N_i} \frac{|z_n - A_j x_n - B_j y_n - C_j|}{\sqrt{A_j^2 + B_j^2 + 1}},$$

où  $(A_j, B_j, C_j)$  sont les paramètres du plan estimés à partir des  $N_j$  points de coordonnées  $(x_j, y_j, z_j)$  appartenant à la région  $R_j$ .

Le critère employé,  $C_{sim}(R_i, R_j)$ , traduit la distance moyenne des  $N_i$  points de la région  $R_i$ , de coordonnées  $(x_i, y_i, z_i)$ , au plan  $z = A_j x + B_j y + C_j$ . Ainsi la similarité des régions  $R_i$  et  $R_j$  est calculée en extrapolant la surface de la région  $R_j$  à la région  $R_i$ . Le seuil global  $\varepsilon$  permet d'arrêter la construction de la pyramide.

A l'aide de cet opérateur, le graphe de similarité orienté  $G'[k]$  est établi. En appliquant la règle 1 (relaxée dans le cas de circuits d'ordre impair) et la règle 2, les sommets non-survivants sont déduits et les régions correspondantes fusionnent avec les voisines, présentes au niveau supérieur de la pyramide. De plus, nous avons enrichi l'approche région pyramidale en prenant en compte les discontinuités de l'image pour interdire la fusion de certaines régions voisines. Le

résultat de la segmentation correspond aux régions au dernier niveau de la pyramide privé de celles d'une taille inférieure à la plus petite taille admise.

## 4 Résultats sur des images réelles

Dans cette section, deux résultats de segmentation obtenus par la méthode proposée sont présentés (cf. figure 2a et 3a). Les images de profondeur initiales ont été acquises par un capteur télémétrique laser nommé PERCEPTRON (cf. figure 2) et par un capteur à triangulation active nommé ABW (cf. figure 3). Pour ces images, nous disposons d'une segmentation de référence ou vérité de terrain réalisée par un opérateur humain (cf. figure 2b et 3b).

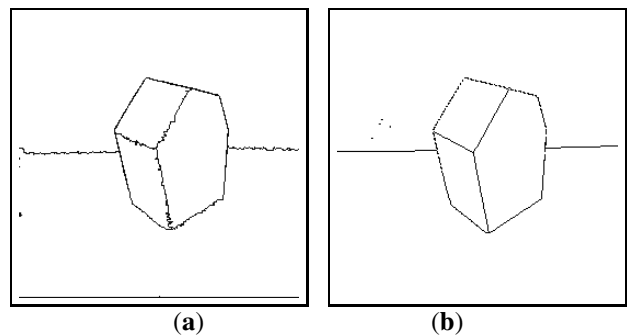


FIG. 2 — Image du capteur télémétrique PERCEPTRON (en noir, les contours des régions) : (a) Résultat de la segmentation. (b) Segmentation de référence.

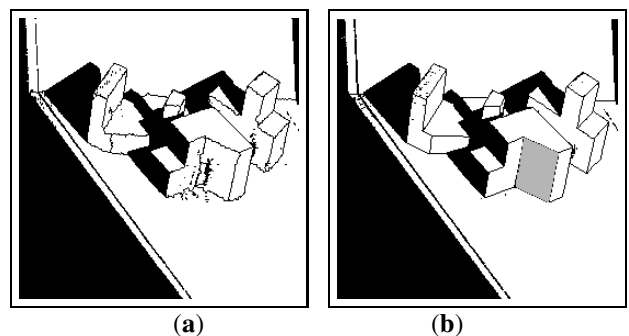


FIG. 3 — Image du capteur à triangulation active ABW (en noir, les contours des régions ou les zones occultées) : (a) Résultat de la segmentation. (b) Segmentation de référence.

Le résultat de la segmentation, figure 2a, sur l'exemple simple qui nous a permis d'illustrer notre méthode paraît proche de la segmentation idéale. De même, celui présenté figure 3a semble là encore visuellement proche du résultat idéal. On peut cependant remarquer que la région grisée dans la vérité de terrain a été sur-segmentée (cf. figure 3b).

Mais, nous avons choisi de ne pas nous contenter d'appréciations uniquement visuelles pour juger la qualité des segmentations, aussi nous avons effectué une évaluation quantitative et comparative de celles-ci.

## 5 Evaluation quantitative et comparative des résultats de la segmentation

Au cours de travaux récents, quatre groupes de recherche appartenant à l'Université de Floride du Sud (USF), à l'Université de Bern (UB), à l'Université de l'Etat de Washington (WSU) et enfin à l'Université d'Edinburgh (UE), [3] [4], ont proposé une méthodologie afin d'évaluer des algorithmes de segmentation d'images de profondeur en régions planes.

Cette évaluation est menée à l'aide d'une banque d'images réelles importante, constituée de 40 images acquises par un télémètre laser, nommé PERCEPTRON, et 40 images délivrées par un capteur à triangulation (ABW). Chaque image est accompagnée de la segmentation idéale ou vérité de terrain. Ces données sont disponibles via [http : / / marathon.csee.usf.edu/seg-comp/SegComp.html](http://marathon.csee.usf.edu/seg-comp/SegComp.html). De chacun des deux lots de 40 images, 10 images de mise au point ont été extraites. La méthode de comparaison employée classe les régions en régions « correctement segmentées » ou non en fonction d'un seuil de tolérance  $T$  [4].

Pour prouver l'efficacité de notre algorithme (UBP : Université Blaise Pascal), nous présentons figures 4 et 5 l'évolution du nombre de régions correctement segmentées par les cinq algorithmes sur les 60 images test, en fonction du seuil de tolérance  $T$  défini par Hoover et *al.*. Sur ces figures, est également reporté le nombre de régions dans la vérité de terrain (Idéal).

Les mesures effectuées (cf. figures 4 et 5) montrent que l'approche développée fournit des résultats de qualité similaire à ceux obtenus par ces autres techniques avec un avantage cependant dans le cas d'images bruitées (cf. figure 4). De plus, ces résultats sont obtenus en des temps de calcul acceptables (en moyenne 51 s et 22 s pour traiter, sur une HP9000/735, une image  $512 \times 512$  issue respectivement du PERCEPTRON et du ABW) [1].

## 6 Conclusion

Nous avons présenté dans cet article un algorithme de segmentation d'images de profondeur ainsi que les résultats d'une étude expérimentale visant à évaluer les performances d'algorithmes de segmentation en surfaces planes. De plus, une comparaison avec quatre autres méthodes de segmentation a été réalisée. Ce travail a été mené à l'aide d'un jeu de 80 images réelles. Cette étude quantitative et comparative montre que l'algorithme proposé est performant.

## Références

- [1] P. Checchin. *Segmentation d'images de profondeur*. Thèse de doctorat d'université, Université Blaise Pascal, Clermont-Ferrand, 1996.
- [2] J.P. Coqueret, S. Philipp, P. Bolon, J.M. Chassery, D. Demigny, C. Graffigne, A. Montanvert, R. Zéboudji, and J. Zérubia. *Analyses d'images : filtrage et segmentation*. Masson, Paris, 1995.

- [3] A. Hoover, G. Jean-Baptiste, D. Goldgof, and K.W. Bowyer. A Methodology for Evaluating Range Image Segmentation Techniques. In *2nd IEEE Workshop on Applications of Computer Vision*, pages 264–271, Coral Gables, Florida, December 1994.
- [4] 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.

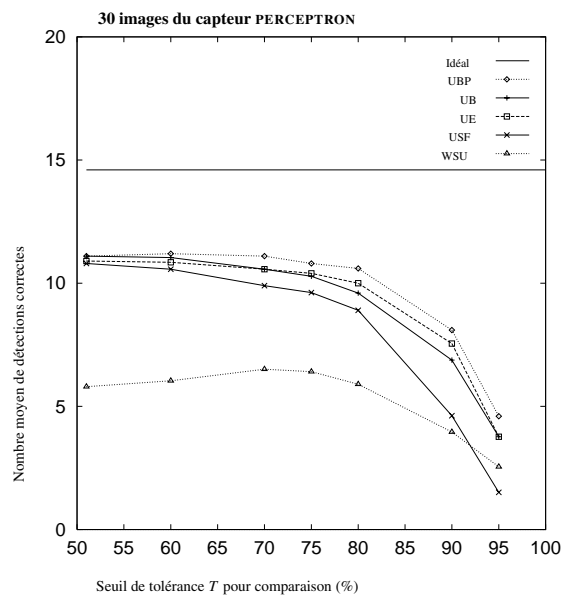


FIG. 4 — Evolution des détections correctes (C.S.) en fonction du taux de recouvrement  $T$  pour les 30 images test issues du capteur PERCEPTRON.

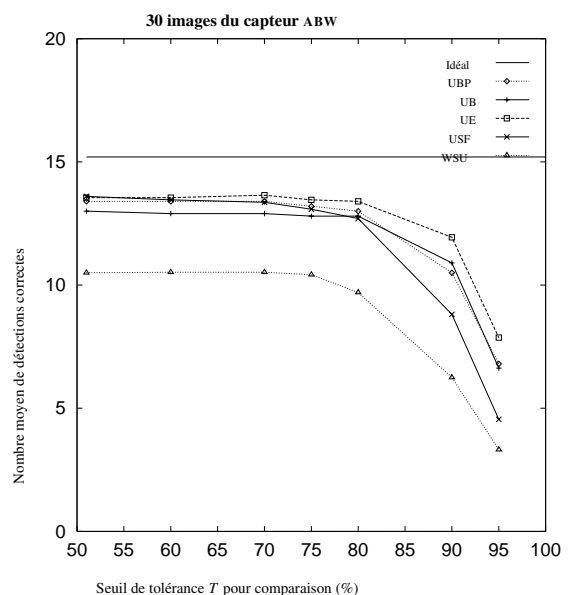


FIG. 5 — Evolution des détections correctes (C.S.) en fonction du taux de recouvrement  $T$  pour les 30 images test issues du capteur ABW.