

# Segmentation de scènes urbaines par combinaison d'information

M.-A. Bauda<sup>1,2</sup>

S. Chambon<sup>1</sup>

M. Spangenberg<sup>2</sup>

V. Charvillat<sup>1</sup>

<sup>1</sup> Institut de Recherches Informatique de Toulouse (IRIT) - VORTEX - Université de Toulouse

<sup>2</sup> imajing sas - 10, avenue de l'Europe - 31520 Ramonville St Agne

{mbauda, mspangenberg}@imajing.fr  
{sylvie.chambon, vincent.charvillat}@enseeih.fr

## Résumé

*Nous présentons dans cet article une méthode de segmentation en superpixels appliquée à des images de scènes urbaines. Ces images proviennent de séquences acquises au moyen d'un système de cartographie mobile qui assure des prises de vue géo-référencées et redondantes spatio-temporellement. Notre objectif est d'obtenir une segmentation sémantique d'un triplet d'images successives en utilisant la richesse des informations à notre disposition : positions des capteurs, reconstructions 3D éparses et redondances visuelles. Pour cela, nous mettons en œuvre une approche s'appuyant sur une sursegmentation en superpixels cohérents sur un triplet d'images. Notre algorithme fournit des résultats à partir d'images réelles, extraites d'un corpus de validation représentatif du problème posé.*

## Mots Clef

Segmentation, scène urbaine, superpixel, planéité.

## Abstract

*We present in this paper an image segmentation method for urban scene images. These images come from sequences captured with a mobile mapping system which provides geo-referenced and spatio-temporally redundant views. Our main goal is a semantic segmentation on the triplet of consecutive images using available information : position sensors, 3D scattered reconstruction and visual redundancy. To achieve this goal, this paper introduces an approach based on an over-segmentation with superpixels consistent on a triplet of images. Our algorithm provides results on real images, extracted from a corpus significant for the problem.*

## Keywords

Segmentation, urban scene, superpixel, planarity.

## 1 Introduction

De nombreux systèmes composés de capteurs permettent aujourd’hui, de collecter et d’exploiter une grande quantité de données de natures différentes. Dans le domaine de la cartographie mobile, ces données (images, points 3D, données GPS) acquises en milieu urbain, fournissent des informations utiles à la réalisation d’inventaires et d’outils de gestion d’équipements (signalisation horizontale et verticale, candélabres, obstacles latéraux).



FIGURE 1 – L’imajbox : système d’acquisition mobile pour le relevé de réseaux de transport.

Nous travaillons sur des données uniques produites par un système propriétaire de relevé mobile de réseaux de transport, proposé par l’entreprise imajing sas<sup>1</sup>. Le dispositif de capture, l’imajbox, figure 1, est installé sur un véhicule et fournit des images géoréférencées successives d’une scène urbaine comme illustré par le triplet de la figure 2.

Les données visuelles acquises, typiquement, tous les 5 mètres sont denses et riches : les positions du capteur sont estimées, les images successives se re-

1. <http://www.imajing.fr>

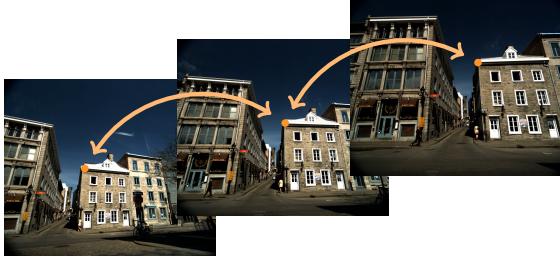


FIGURE 2 – Triplet d’images successives et géo-référencées d’une même séquence, à sur-segmenter à l’aide de la redondance d’information afin d’obtenir des zones cohérentes sur les trois images. Par exemple, le point orange doit se trouver dans un superpixel qui peut être mis en relation avec les superpixels des images adjacentes.

couvrent, une reconstruction 3D épars est aussi calculée sur chaque paire d’images, comme illustré par la figure 3. Afin de développer des applications et services d’inventaire ou de détection de changement, la société imajing, comme beaucoup d’autres acteurs du domaine, souhaite interpréter, segmenter les images acquises, reconnaître les objets d’intérêt (panneaux, façades des immeubles, etc.). Notre objectif est de segmenter les images de scènes urbaines en combinant les informations 2D/3D disponibles en vue d’une description sémantique conforme à une vérité terrain que nous avons établie sur un corpus de données de références.

Compte-tenu de la redondance disponible dans les images, il s’agit aussi d’un problème de cosegmentation. On voudra par exemple détecter un segment associé aux façades à partir des trois images de la figure 2. Pour détecter des façades, nous souhaitons regrouper des régions planes et verticales visibles dans les différentes images. Nous défendons ici une approche pour laquelle des superpixels (vus comme le résultat d’une étape préalable de sur-segmentation) vont permettre de bâtir des regroupements pertinents du point de vue de l’apparence, de la géométrie et de la sémantique.

Dans cet article, notre contribution est une méthode de sur-segmentation qui rend cohérents spatio-temporellement des superpixels calculés sur les 3 images d’un triplet. Des sur-segmentations cohérentes au sein d’un triplet d’images nous paraissent intéressantes pour exploiter les redondances disponibles dans les étapes ultérieures de notre interprétation d’images.

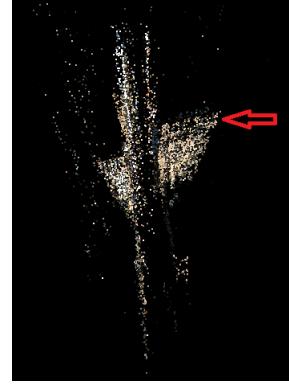


FIGURE 3 – Nuage de points 3D, obtenu par triangulation sur les points d’intérêt 2D extraits et appariés 2 à 2 sur 15 images successives de la séquence de travail. Nous apercevons la façade représentée sur le triplet figure 2.

**Travaux existants** Nous pouvons présenter trois domaines de recherche connexes à notre application de segmentation : la sur-segmentation, le suivi d’objets et la recherche de plans dans l’espace.

La sur-segmentation d’images est souvent une étape préliminaire à de nombreuses applications en vision par ordinateur telle que la segmentation [7] ou la détection d’objet d’intérêt [9]. Historiquement, les méthodes de segmentation s’appuyaient soit sur les **contours**, soit sur les **régions**. Les contours sont utilisés par les méthodes à base de lignes de partages des eaux [11] ou les approches de type contours actifs par minimisation d’énergie [8], alors que les régions interviennent, par exemple, dans l’approche *mean-shift* [3].

Plus récemment, est apparue la notion de **superpixel** qui est une structure locale et cohérente de niveau intermédiaire. Un superpixel possède un support spatial perceptiblement plus cohérent mais moins régulier en forme que celui d’un *patch* (zone de forme régulière). Certaines méthodes de construction de superpixels favorisent des propriétés de régularité topologique [10], de compacité [1, 10] ou encore de connexité [4]. Il semble encore difficile de trouver les liens entre les propriétés attendues et les applications qui en résultent. En plus de réduire la complexité algorithmique et le temps de calcul, nous pensons que les superpixels doivent adhérer au contour de l’image tout en respectant la géométrie de la scène.

Travaillant sur des séquences d’images, notre travail se situe à la frontière des méthodes de seg-

mentation sur une image unique et des méthodes de segmentation vidéo. D'une part, les travaux présentés dans [7] sont prometteurs car l'auteur propose une méthode de classification géométrique en trois classes mères (ciel, sol, objets verticaux) à partir d'une unique image. Pour cela il combine plusieurs sur-segmentations de [4] avec différentes valeurs de paramètres et un modèle d'apprentissage. D'autre part, les superpixels sont appliqués à la vidéo, avec des acquisitions faites par caméra fixe ou mobile pour suivre des objets d'intérêt [13, 12] qui peuvent être en mouvement [2].

La recherche de zones planes est souvent résolue à l'aide de méthodes statistiques standard d'estimation d'un modèle dans des données bruitées comme : la méthode de RANSAC (*Random Sample Consensus*) [5] ou la méthode de *JLinqage* [6].

**Méthodologie adoptée** Nous proposons une méthode utilisant des triplets d'images successives, cf. figure 2. Chaque triplet fournit de la double redondance d'information. D'une part, il y a la redondance spatiale, car les objets présents dans la scène apparaissent sur plusieurs images et, d'autre part, de la redondance temporelle car les acquisitions sont ordonnées dans le temps. Un triplet est constitué de deux paires d'images successives  $[0 - 1]$  et  $[1 - 2]$ . Chaque paire d'images contient des points d'intérêt, qui appariés et triangulés, fournissent un nuage épars de points 3D dans le repère du système d'acquisition de chaque image. Pour chaque triplet d'images, nous avons deux nuages de points 3D pouvant être mis en correspondance. La figure 3 présente le nuage de points 3D obtenu sur 15 images successives.

Dans la partie 2, nous présentons la méthode de sur-segmentation proposée. Nous exposons dans la partie 3 les résultats expérimentaux, évalués à partir de la vérité terrain. Cette vérité terrain est établie sur notre corpus de référence dont nous précisons les sources de variabilités.

## 2 Segmentation géométrique par combinaison d'information 2D/3D

Dans cette partie, nous décrivons la méthode de segmentation de scène urbaine proposée, que nous appellerons SEG par la suite. Dans la littérature, les superpixels sont généralement estimés et utilisés sur une seule image. Nous pouvons cependant imaginer rendre le calcul ainsi que l'exploitation des superpixels plus précis, en combinant l'information exis-

tante provenant de plusieurs images. Ainsi, l'approche proposée se décompose en quatre étapes principales :

1. construction des superpixels sur l'image centrale du triplet (2D), afin de regrouper localement l'information 3D par zone cohérente,
2. classification de ces superpixels suivant un critère de planéité (3D) pour permettre la description sémantique qui nous intéresse dans ce travail,
3. estimation des homographies existantes (2D) sur les zones planes pour extraire l'information nous permettant d'enrichir notre modèle,
4. construction de nouveaux superpixels cohérents sur le triplet (2D).

Comme attendu, l'ensemble de ces étapes combine à la fois de l'information 2D et 3D. Nous allons à présent détailler chacune de ces étapes.

### 2.1 Construction des superpixels sur l'image centrale

La construction des superpixels sur l'image centrale, nous permet de travailler localement sur l'information 3D extraite. Nous avons choisi d'initialiser notre algorithme avec une méthode de type *clustering* : nommée *Simple Linear Iterative Clustering* (SLIC) [1].



FIGURE 4 – (a) Image originale ( $2448 \times 2050$ ). (b) Image segmentée en utilisant les superpixels [1] avec les paramètres  $K = 20$ ,  $c = 20$ , cf. § 2.1 pour plus de détails.

Cette méthode récente, permet de construire des superpixels réguliers en surface. Tout d'abord les centres des superpixels sont initialisés sur une grille régulière, espacés de  $S$  pixels avec  $S = \sqrt{\frac{N}{K}}$  où  $N$  est le nombre total de pixels dans l'image et  $K$  le nombre de superpixels souhaités. Ils peuvent être éventuellement déplacés afin qu'ils ne se trouvent pas sur un contour de l'image. Cette méthode est itérative et comprend deux étapes :

1. l'assignation des pixels à un centre  $C_k$  suivant un critère d'appartenance,
2. la mise à jour des centres.

Cette approche tente de minimiser dans l'étape 1, le critère d'appartenance correspondant à une distance entre  $C_k$  et le pixel courant  $p$  définie par :

$$D_{SLIC}^2(C_k, p) = d_{lab}^2(C_k, p) + \frac{d_{xy}^2(C_k, p)}{S^2} c^2 \quad (1)$$

où  $d_{lab}$  est la distance colorimétrique et  $d_{xy}$  est la différence entre les positions dans l'image courante, telles que :

$$d_{lab}^2(p_j, p_i) = (l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2$$

$$d_{xy}^2(p_j, p_i) = (x_j - x_i)^2 + (y_j - y_i)^2.$$

Ainsi, les paramètres sont le nombre approximatif de superpixels  $K$  et leur compacité  $c$ . D'après [1],  $c \in [0 - 40]$  lorsque nous travaillons sur l'espace couleur CIELAB. Le terme  $c$  joue un rôle de pondération entre la couleur et la position. Lorsque  $c = 0$  les superpixels peuvent être très souples et adhèrent aux contours de l'image et lorsque  $c = 40$  ils se rapprochent d'une forme régulière. Les auteurs proposent de fixer  $c$  à 10 car avec cette valeur ils obtiennent des performances supérieures à celles obtenues avec [4].

Pour chaque centre est défini une zone de recherche d'appartenance de taille  $2S \times 2S$  et centrée sur  $C_k$ . Seuls les pixels  $p$  appartenant à cette zone sont parcourus comme le montre la figure 5. Cette restriction permet d'obtenir un algorithme de complexité  $\mathcal{O}(N)$  car chaque pixel n'appartient au maximum à 9 zones d'appartenances distincts.

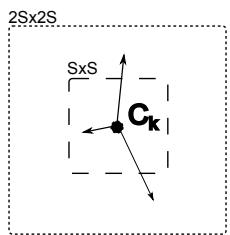


FIGURE 5 – Zone de recherche de pixels similaires au centre  $C_k$  de référence – Cette région est de taille  $2S \times 2S$ , centrée sur  $C_k$ , avec la surface  $S \times S$ .

Par la suite, nous évaluerons quantitativement ces superpixels, ce qui nous permettra de comparer nos résultats à ceux existants.

## 2.2 Classification géométrique des superpixels

Afin d'obtenir une description sémantique de la scène adaptée à notre problématique, cf. § 3.1, nous cherchons dans un premier temps à décrire la scène de manière géométrique, comparable à celle présentée par [7]. Pour cela nous proposons d'exploiter la redondance d'information issue du triplet étudié.

Ainsi, nous cherchons à classer les superpixels issus de la sur-segmentation de l'image centrale, comme présenté dans la figure 6(a), en trois classes géométriques : *zone plane*, *zone non plane* et *zone n'ayant pas suffisamment d'information pour déterminer la planéité*.

Pour chaque superpixel de l'image centrale, en orange sur la figure 4(b), nous pouvons extraire les points d'intérêts 2D détectés dans cette zone. Cette information 2D nous permet de travailler sur les coordonnées des points 3D associés pour estimer la planéité des zones.

Un plan est défini par 3 points 3D. Il nous faut donc au minimum  $P$  points non colinéaires ( $P \geq 4$ ) pour estimer un plan. S'il y a moins de  $P$  points d'intérêt dans le superpixel alors il est classé comme n'ayant pas suffisamment d'information pour déterminer la planéité. Sinon, s'il y a plus de  $P$  points d'intérêt, nous cherchons le plan dominant ayant l'erreur la plus faible avec le maximum de points validant le modèle. Ceci est fait dans le nuage de points 3D à l'aide de la méthode RANSAC. Si un plan dominant est extrait le superpixel est alors classé comme plan, sinon il est classé non plan.

Pour chaque plan estimé, nous construisons une boîte englobante sur les points d'intérêt 2D utilisés. Celle-ci nous permet de créer un masque binaire représentant l'ensemble des zones planes détectées, comme nous le montre la figure 6(b).

Cette classification en trois classes, nous permet de visualiser l'ensemble des zones, en vert sur 6(a) ou en blanc sur 6(b), sur lesquelles nous cherchons à améliorer la sur-segmentation.

## 2.3 Estimation des homographies et recalage d'images

Cette étape ne s'applique qu'aux superpixels classés « plan ». Une homographie est la transformation qui permet de passer de la représentation image d'un plan à son plan homologue dans une autre image. Pour chaque zone plane de l'image centrale nous pouvons estimer : soit une homographie avec l'image précé-

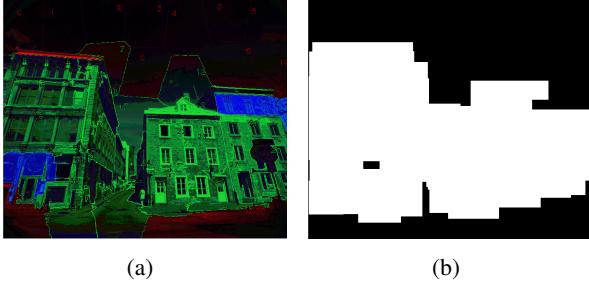


FIGURE 6 – (a) Résultat de la classification géométrique : zone plane en vert, zone non plane en bleu et zone sans information suffisante en rouge. (b) Masque binaire de l’ensemble des boîtes englobantes calculées sur les zones planes.

dente, soit une avec l’image suivante, soit les deux suivant le nombre de points d’intérêt détectés et mis en correspondance sur chaque paire d’images  $[0 - 1]$  et  $[1 - 2]$ .

Nous calculons les homographies existantes à partir de 4 paires de points 2D appariés parmi ceux utilisés par l’algorithme RANSAC. Comme nous le montre la figure 7, nous pouvons transformer et recaler la boîte englobante de la zone plane adjacente sur la zone d’intérêt de l’image centrale comme sur la figure 7(c).

Nous pouvons alors calculer  $\text{diff}_0$ , respectivement  $\text{diff}_1$ , la différence entre la zone originale de l’image centrale notée  $im_1$  et la zone recalée de l’image précédente notée  $H_0(im_0)$ , respectivement de l’image suivante,  $H_2(im_2)$ .

$$\text{diff}_0(p) = d_{\text{lab}}(p_0, p) \quad \text{et} \quad \text{diff}_1(p) = d_{\text{lab}}(p_2, p)$$

où  $p_0$ , respectivement  $p_2$ , représente le projeté orthogonal du point  $p$  dans  $H_0(im_0)$ , respectivement  $H_2(im_2)$ . Une homographie correcte fournit des zones planes bien recalées. La différence entre deux zones est alors faible, cf. les zones noires de la figure 7(d). Les zones rectifiées contenant des occultations ou des objets appartenant à un plan différent, comme le panneau de signalisation représenté à la fois en bleu et en jaune dans la figure 7(c), sont alors mis en évidence.

## 2.4 Construction de superpixels cohérents sur le triplet

Cette étape nous permet d’intégrer l’information de nature géométrique estimée à l’étape précédente pour affiner la construction des superpixels et la rendre cohérente sur le triplet. En effet, le recalage de zones

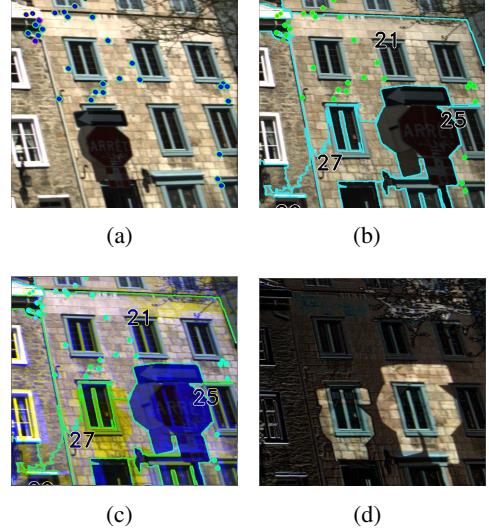


FIGURE 7 – Zoom sur une zone plane. (a)  $H_0(im_0)$  zone plane recalée de l’image précédente et (b)  $im_1$  zone plane d’intérêt de l’image centrale. (c) Superposition de ces deux zones ( $im_1$  est visualisée sur les canaux vert et rouge et  $H_0(im_0)$  est représentée sur le canal bleu). (d) Différence sur les trois canaux de couleur entre (a) et (b). Les zones mal recalées sont mises en évidence, comme, par exemple, le panneau.

d’images décrit précédemment, fournit une information complémentaire pour calculer une nouvelle segmentation en superpixels. De manière semblable à SLIC [1], notre algorithme SEG s’appuie sur deux étapes : le calcul du positionnement des centres puis l’assignation des pixels à un centre  $C_k$  suivant un critère d’appartenance.

Pour chaque zone plane de l’image centrale ( $im_1$ ), nous utilisons l’information provenant des homographies existantes ( $H_0$  et/ou  $H_2$ ). Le cas le plus favorable se produit lorsqu’il existe à la fois  $H_0$  et  $H_2$ , c’est-à-dire qu’un plan est estimé dans les 2 nuages de points 3D successifs.

Nous pourrions effectuer un « double SLIC » proche d’un « SLIC 3D », c’est-à-dire prendre en compte l’information de couleur et de position à la fois sur l’image centrale  $im_1$  avec  $D_{\text{SLIC}}(C_k, p)$  et sur les zones recalées, par exemple  $H_0(im_0)$  avec  $D_{\text{SLIC}}(C_{k_0}, p_0)$  où  $C_{k_0}$  et  $C_{k_1}$  ont les mêmes coordonnées que  $C_k$  dans la boîte englobante. Or  $d_{xy}(C_k, p) = d_{xy}(C_{k_0}, p_0)$ .

Les termes  $\text{diff}_0(C_k)$  et  $\text{diff}_1(C_k)$  sont constants sur un superpixel. Le « double SLIC » revient donc à étudier l’influence des variations de  $\text{diff}_0$ . C’est pourquoi

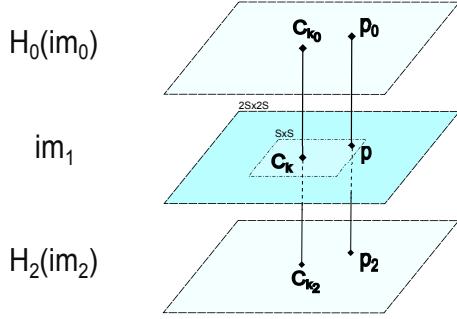


FIGURE 8 – Recalage des deux zones planes des images adjacentes afin de prendre en compte l’information couleur de ces zones.

lorsque le centre  $C_k$  se trouve dans une zone plane, il paraît judicieux de prendre en compte dans notre critère, la différence colorimétrique :  $\text{diff}_i$ ,  $i = \{0, 1\}$ . Nous calculons donc nos superpixels à partir d’un critère de similarité prenant en compte plusieurs couches comme illustré dans la figure 9.

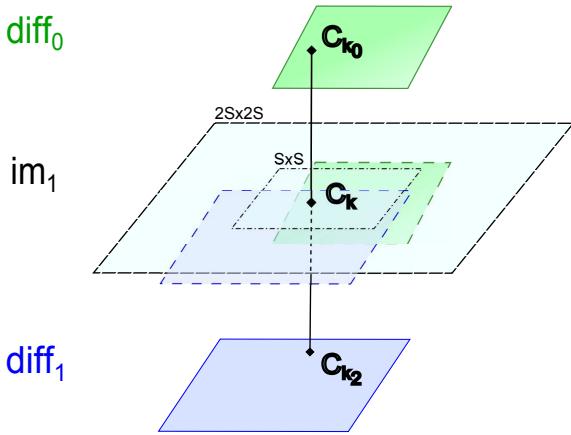


FIGURE 9 – Cas où  $C_k$  appartient à deux zones planes, il existe deux homographies qui permettent de rectifier les zones des images adjacentes et de calculer les différences  $\text{diff}_0$  et  $\text{diff}_1$  qui sont intégrées à la distance  $D$ , cf § 2.4 pour la définition de ces termes.

Le critère de similarité utilisé est défini comme suit :

$$D^2(C_k, p) = D_{SLIC}^2(C_k, p) + d_{plan}^2(p) \quad (2)$$

où

$$d_{plan}(p) = \begin{cases} \text{diff}_0(p) \mathbb{1}_{H_0} + \text{diff}_1(p) \mathbb{1}_{H_2} & \text{si } Z(p) \\ N_p & \text{sinon} \end{cases}$$

avec

$$N_p = \max_{p \in im_1} d_{plan}(p).$$

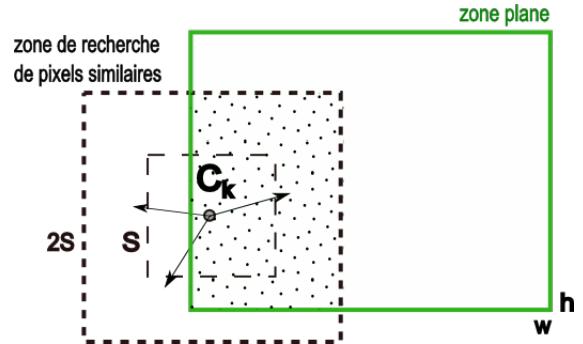


FIGURE 10 – Calcul de la pénalité – Pour chaque pixel de la zone de recherche centrée sur  $C_k$  qui appartient à une zone plane, la pénalité  $d_{plan}$  est prise en compte.

En d’autres termes, la pénalité  $d_{plan}$  n’est calculée que dans le cas où  $C_k$  se trouve dans une zone plane, i.e.  $Z(p)$  est vrai, sinon elle est nulle. La figure 10 schématise l’évaluation de celle-ci. Pour tous les pixels appartenant à la fois à la zone plane et à la zone de recherche de similarité (zone marquée en pointillée),  $d_{plan}$  vaut la somme des différences  $\text{diff}_i$  avec  $i = \{0, 1\}$  lorsque les homographies correspondantes existent. Pour les autres pixels, appartenant seulement à la zone de recherche de similarité,  $d_{plan}$  vaut la valeur maximale rencontrée  $N_p$ .

### 3 Résultats expérimentaux

Dans cette partie, nous commençons par présenter le corpus, représentatif de l’ensemble des données de référence, qui nous permet d’évaluer et de valider notre algorithme SEG. Nous exposons ensuite la description sémantique choisie puis les résultats obtenus.

#### 3.1 Corpus

**Variabilité de la base de données.** La base de données est d’une grande diversité. De nombreux paramètres ont une influence directe sur la variabilité des images comme la position de l’imajbox (paramètres extrinsèques), le type de support (voiture, train) et de réseau relevé ainsi que l’orientation de la trajectoire (rectiligne, virages à droite, à gauche). De plus, la distance et la hauteur des objets présents dans la scène modifient la surface de recouvrement entre deux images successives. Les images varient également en fonction des conditions d’acquisitions (luminosité, saison, météo).

**Vérité terrain.** Une trentaine d’images de référence ont été segmentées manuellement, cf. fi-

gure 11, à l'aide d'un programme s'appuyant sur *Object Labelling Tool (OLT)* de [7]. Nous l'avons adapté, d'une part en mettant en place un pré-traitement en superpixels [4] afin d'obtenir des contours plus précis et ainsi une meilleure résolution de la vérité terrain et d'autre part, en permettant de nommer les objets segmentés à partir d'une description sémantique appropriée à notre contexte.

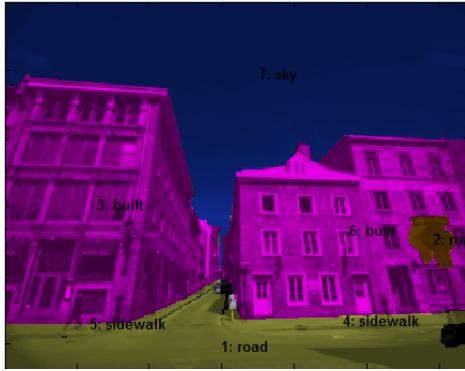


FIGURE 11 – Vérité terrain – Segmentation manuelle, permettant d'évaluer automatiquement les résultats obtenus par notre méthode.

**Description sémantique.** Nous proposons d'utiliser une description sémantique hiérarchique en six classes mères : ciel, sol, bâtiments, végétation, mobilier urbain et une classe divers dans laquelle seront classés tous les éléments rendant la segmentation difficile comme les piétons et les voitures. Ces classes pourront être divisées en sous-classes. Par exemple, la classe « sol » peut se décliner en : chaussée, trottoir, marquage au sol.

### 3.2 Évaluation

**Critère d'évaluation.** Nous cherchons à répondre à la question : « À quel point avons nous amélioré la sursegmentation initiale ? ». Pour cela, nous comparons nos résultats à la vérité terrain suivant un critère couramment utilisé dans la littérature et qui mesure l'adhérence des contours : l'erreur de sous-segmentation comme dans [1], définie par :

$$ErrSS = \frac{1}{N} \left[ \sum_{i=1}^M \left( \sum_{s_j | s_j \cap g_i > B} |s_j| \right) - N \right]$$

où  $N$  est le nombre total de pixels dans l'image,  $M$  est le nombre de superpixels,  $g_i$  est une région de la vérité terrain et  $s_j$  un superpixel. Ce critère s'appuie sur le calcul du nombre de pixels appartenant à la fois

à  $s_j$  et à chacun des  $g_i$ , si et seulement si  $s_j \cap g_i > B$ . Dans nos tests,  $B$  vaut zéro.

**Résultats.** Nous présentons les résultats obtenus sur le triplet d'images présenté figure 2. Le paramètre de compacité sera fixé à  $c = 10$  pour la suite, comme conseillé dans [1].

Le graphique 12 représente l'évolution de l'erreur de sous-segmentation en fonction du nombre de superpixels et permet de comparer notre méthode à l'approche SLIC. Nous constatons une amélioration pour  $K = 20$  l'erreur de sous-segmentation diminue de 1.71 pour SLIC à 1.42, c'est-à-dire que l'on a amélioré la sur-segmentation d'environ 16%.

Bien que l'amélioration ne soit pas constante, les superpixels semblent mieux délimiter les zones planes sur la figure 13(b) que sur la figure 13(c).

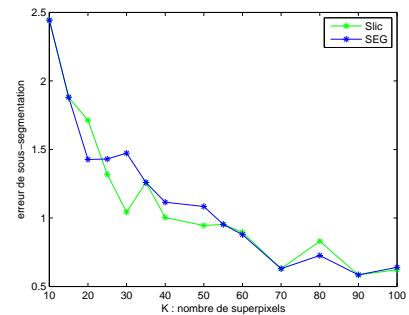


FIGURE 12 – Erreur de sous-segmentation pour  $c = 10$ .

**Analyses.** Nous travaillons sur des images représentant des scènes pouvant avoir une forte perspective. La distance entre les objets présents dans la scène et le système d'acquisition est variable et influence la taille de leur représentation dans l'image. C'est pourquoi, prendre des superpixels uniformes en surface ne semble pas être une contrainte adaptée à notre problème.

### Conclusion

Nous avons présenté une approche de sursegmentation combinant l'information géométrique et colorimétrique d'un triplet d'images. Cette méthode prend en compte à la fois la couleur des pixels et la planéité des régions.

Il nous semble intéressant d'utiliser l'information géométrique car elle est riche et donne une description réaliste et détaillée de la scène. Les superpixels que nous utilisons vérifient une propriété contraignante :

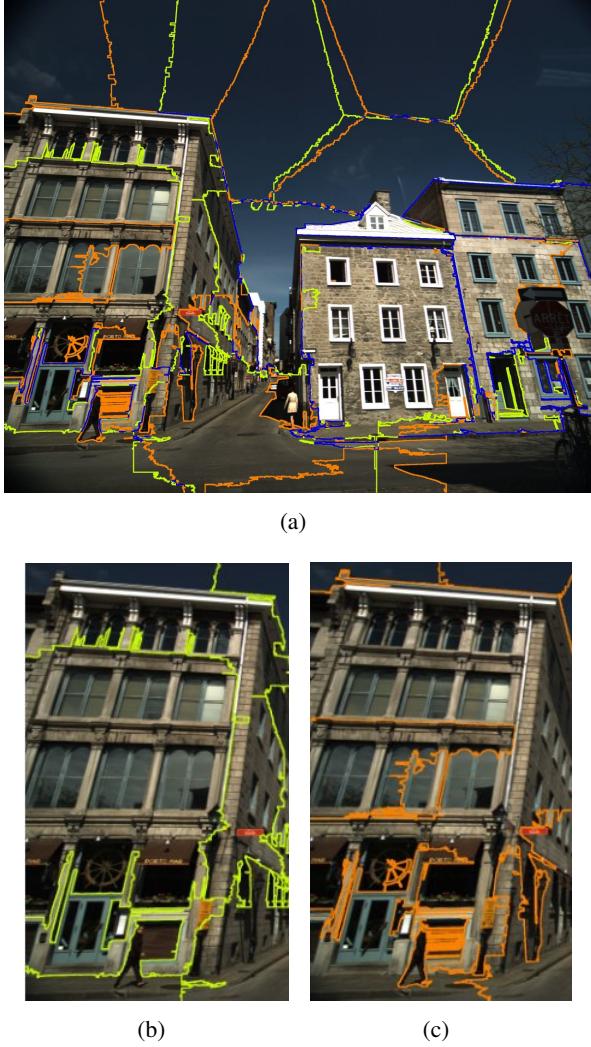


FIGURE 13 – Résultats – (a) Comparaison de nos résultats en jaune avec ceux obtenus avec l’algorithme SLIC en orange pour les paramètres  $c = 10$  et  $K = 20$ . Les contours communs sont en bleu. (b) et (c) zoom sur les résultats obtenus sur une façade.

l’uniformité en surface. Elle ne permet pas de visualiser clairement l’impact de notre critère sur la construction de nos nouveaux superpixels.

Nous pensons que l’information géométrique doit être extraite et estimée sur l’ensemble de la scène représentée car l’étude locale introduit des erreurs d’orientation de plan. Nous pourrons par exemple, modifier l’étape qui est au cœur de l’originalité de notre méthode : la détection de zones planes, en l’appliquant à l’ensemble de l’image à l’aide d’un seuil adaptatif.

Ceci nous amène à penser, que la prise en compte d’autres sources d’informations géométriques, comme les points de fuites et la ligne d’hor-

izon, pourraient nous permettre d’améliorer les performances de cette sur-segmentation.

Enfin, la prochaine étape de ce travail portera sur la classification de l’ensemble de ces superpixels suivant la description sémantique proposée.

## Références

- [1] R. Achanta and A. Shaji and K. Smith and A. Lucchi and P. Fua and S. Susstrunk, SLIC Superpixels Compared to State-of-the-art Superpixel Methods, *IEEE Transactions on PAMI*, 2011.
- [2] A. Ayvaci and S. Soatto, Motion segmentation with occlusions on the superpixel graph, *IEEE 12th ICCV Workshops*, 2009.
- [3] D. Comaniciu and P. Meer, Mean shift : a robust approach toward feature space analysis, *IEEE Transactions on PAMI*, pp. 603-619, 2002.
- [4] P. Felzenszwalb and D. Huttenlocher, Efficient Graph-Based Image Segmentation, *IJCV*, pp. 167-181, 2004.
- [5] M. Fischer and R. Bolles, Random Sample Consensus : A paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communication of the ACM*, pp. 381-397, 1981.
- [6] D. Fouhey and D. Scharstein and A. Briggs, Multiple Plane Detection in Image Pairs using J-Linkage, *IEEE ICPR*, 2010.
- [7] D. Hoiem, Geometric Context from a Single Image, *IEEE ICCV*, pp. 654-661, 2005.
- [8] M. Kass and A. Witkin and D. Terzopoulos, Snakes : Active contour models, *IJCV*, pp. 321-331, 1988.
- [9] G. Mori, Guiding model search using segmentation, *IEEE ICCV*, pp. 1417-1423, 2005.
- [10] A. Moore and S. Prince, Scene shape priors for superpixel segmentation, *IEEE ICCV*, pp. 771-778, 2009.
- [11] L. Vincent and P. Soille, Watersheds in Digital Spaces : An Efficient Algorithm Based on Immersion Simulations, *IEEE Transactions on PAMI*, 1991.
- [12] X. Ren and J. Malik, Tracking as Repeated Figure/Ground Segmentation, *IEEE CVPR*, 2007.
- [13] S. Wang, Superpixel tracking, *IEEE ICCV*, pp. 1323-1330, 2011.