
Exploration autonome et cartographie topologique en environnement inconnu référencées vision omnidirectionnelle

Romain Marie, Ouidad Labbani-Igbida, El Mustapha Mouaddib

*Laboratoire Modélisation, Information & Systèmes
Université de Picardie Jules Verne, Amiens, France
prenom.nom@u-picardie.fr*

RÉSUMÉ. Dans ce papier, nous présentons une nouvelle méthode pour l'exploration autonome d'environnements inconnus et la construction de carte topologique par un robot mobile terrestre à partir d'un capteur de vision catadioptrique. Nous proposons une approche incrémentale qui permet au robot d'extraire et de combiner plusieurs représentations spatiales construites à partir des images acquises : l'espace libre navigable obtenu par la propagation d'un contour actif, les topologies locale et globale déterminées à partir du squelette de l'espace libre et une signature de lieu fondée sur l'information photométrique pertinente. Des résultats expérimentaux en environnement extérieur à large échelle démontrent l'efficacité de l'approche et ouvrent de nouvelles perspectives pour les méthodes de navigation autonome basées vision, qui sont encore aujourd'hui un problème ouvert en robotique.

ABSTRACT. In this paper, we address the problem of exploration and topological map building in totally unknown environments for a mobile robot equipped with a sole catadioptric sensor. Multiple representations for spatial knowledge are built upon visual information only : free space detection, local space topology extraction, place signatures and topological links inferred from robot actions. We show how these spatial representations are nested together to incrementally build a spatial map of the explored robot space. The efficiency of the approach is evaluated in real world experiments, giving new solutions to real-time simultaneous localization, navigation and topological map building.

MOTS-CLÉS : exploration autonome, carte topologique, reconnaissance de lieu.

KEYWORDS: autonomous exploration, topological mapping, place recognition.

DOI:10.3166/TS.31.221-243 © 2014 Lavoisier

Extended abstract

This paper addresses the problem of autonomous exploration, navigation and topological map building in unknown environments. Considering a mobile robot equipped with an omnidirectional catadioptric sensor, we develop an incremental approach to merge successive local perceptions into a spatial model of the explored space. It is based on a set of local representations built from visual information only :

- The omnidirectional free space, extracted in the image using an active contour (Merveilleux *et al.*, 2011) initialized on the robot projection. It segments the image in two regions with complementary properties : the navigable space which carries the topological structure of the ground surroundings, and the obstacles which contain most of the relevant photometric information.

- The local Generalized Voronoi Diagram (GVD) computed directly in the image using an omnidirectional version of the *Delta Medial Axis* (DMA) on the extracted navigable space. It is based on an almost linear time skeletonization algorithm (Marie *et al.*, 2013), able to deal with boundary noise through an original pruning process using a single parameter. To consider distortions involved by the catadioptric sensor, and thus compute the skeleton of the free space instead of the skeleton of its projection, we replace the euclidian metric by an adapted version that considers both the projection model and the position of the sensor with respect to the ground.

- A visual signature built from the outer part of the image (the obstacles), which contains most of the relevant photometric information of the scene. We use the Haar integral invariants formalism to capture those information, and propose a combination of several kernel functions to produce a distribution-like signature with good discriminative properties.

Using a simple control law following the extracted skeleton, the robot navigates autonomously on the GVD, builds new signatures of discovered places, and identifies loop closures when needed. Each place is associated to a node in the topological map, and defined by a visual signature. A two threshold approach is used for localization purposes, and allows both new places identification and loop-closure detection. The entire process, including image acquisition, free space segmentation, skeleton extraction, signature computation and map update is performed online by the robot, in less than 200 ms. The efficiency of the approach is evaluated in real world experiments, giving new solutions to the real-time simultaneous localization, navigation and topological map building.

1. Introduction

La localisation, navigation et cartographie simultanées basées vision ou apparence est un domaine très actif de la robotique mobile, du fait notamment de l'information très riche fournie par le capteur visuel et des étendues des applications dans des environnements complexes et incertains. Une littérature abondante existe sur le sujet et nous permet de distinguer deux grandes familles : les approches métriques nommées souvent *visual slam*, qui aboutissent à des cartes basées sur des caractéristiques géométriques et l'estimation de pose du robot (Se *et al.*, 2002 ; Sim *et al.*, 2007), et les méthodes basées apparence (Valgren *et al.*, 2006 ; Murillo, Kosecka, 2009), dites topologiques, qui construisent une carte à un niveau plus structurel, en définissant des nœuds représentant des lieux, et des liens les reliant. Certains travaux (Lim *et al.*, 2012) utilisent la complémentarité de ces deux approches, en combinant l'information géométrique locale avec une information structurelle plus globale.

Les approches topologiques considèrent souvent une connaissance nécessaire a priori de l'environnement, et l'utilisent comme référence pour la localisation et la navigation. Cette connaissance est acquise par un processus préalable d'apprentissage où le robot est guidé dans l'environnement. Le travail présenté dans cet article s'affranchit de cette contrainte, en n'imposant aucune connaissance préalable du terrain. Il permet, simultanément, la navigation autonome et la construction de carte topologique par structuration de représentations spatiales acquises visuellement pendant l'exploration. L'approche proposée est fondée sur :

- une information topologique définie à partir du squelette de l'espace libre local (lui-même extrait depuis l'image) ;
- une information photométrique définissant une signature visuelle du lieu exploré et construite parallèlement à partir de l'information visuelle pertinente de l'image.

Ces connaissances, construites en ligne par le robot, constituent les représentations spatiales locales, qui sont ensuite fusionnées sous la forme d'une carte topologique manipulable par ce dernier. Le squelette de l'espace libre permet de définir des trajectoires sûres qui guident le robot dans l'environnement, générant ainsi des liens reliant les différents lieux visités. Ces derniers sont indexés en utilisant les signatures construites, caractérisant des nœuds dans le processus de construction de carte. De par leur invariance, elles permettent de localiser le robot et de gérer les problèmes de fermeture de boucle.

L'article est organisé comme suit. Après une brève présentation de l'état de l'art dans la section 2, nous introduisons le principe utilisé pour la détection de l'espace libre (section 3). Dans la section 4, un algorithme d'extraction de la topologie locale est présenté, suivi du processus de construction des signatures en section 5. La section 6 décrit comment ces deux représentations peuvent être imbriquées afin de construire un modèle spatial pertinent de l'environnement. Pour valider l'approche, des résultats expérimentaux sont présentés en section 7 avant de conclure section 8.

2. État de l'art

Les méthodes basées SLAM (*Simultaneous Localization And Mapping*) visuel (voir (Durrant-Whyte, Bailey, 2006 ; Muhammad *et al.*, 2009) pour une revue) proposent des outils de sélection et de mise en correspondance de caractéristiques visuelles (Eade, Drummond, 2006 ; Davison *et al.*, 2007 ; Konolige, Agrawal, 2008), menant à la construction de cartes géométriques précises de l'environnement.

De leur côté, les méthodes purement topologiques (Valgren *et al.*, 2006 ; Cummins, Newman, 2011) connaissent un essor important du fait des avancées récentes en reconnaissance d'images basée apparence. On peut ainsi caractériser et indexer un lieu visité par le robot en utilisant des signatures visuelles construites à partir de l'image. L'espace navigable du robot est représenté sous la forme d'un graphe où les nœuds correspondent à des lieux visités et les liens caractérisent les déplacements du robot entre les nœuds. Ces approches permettent une représentation compacte d'environnements très larges, sans utiliser une référence globale absolue, ni l'information géométrique. En revanche, elles sont généralement peu adaptées pour une navigation précise (et sécurisée). Dans (Milford, Wyeth, 2010), les auteurs proposent une carte persistante utilisant RatSLAM, un système SLAM bioinspiré, où le robot met en permanence à jour sa représentation du monde. Certaines approches hybrides (Lim *et al.*, 2012) étendent la notion de cartographie topologique en intégrant l'information métrique issue de l'odométrie du robot (Angeli *et al.*, 2009), ou bien une représentation géométrique permettant une mise à jour des nœuds (Dayoub *et al.*, 2011).

La principale problématique dans les approches basées topologie consiste à identifier de manière fiable les différents lieux, tout en étant robuste au problème d'ambiguïté perceptive où des lieux distants peuvent présenter une apparence fortement similaire. Plusieurs solutions ont été proposées dans la littérature. Elles sont fondées sur des descripteurs locaux de l'image (Se *et al.*, 2002 ; Cummins, Newman, 2011 ; Korrapati *et al.*, 2012) ou globaux (Murillo, Kosecka, 2009 ; Singh, Kosecka, 2010 ; Chapoulie *et al.*, 2013). Puisqu'elles considèrent l'ensemble de l'information visuelle, les approches globales sont, relativement aux approches locales, moins robustes aux occultations et changements d'illumination, et nécessitent donc une attention particulière lors de leur élaboration. Elles sont en revanche plus rapides et plus compactes pour des applications temps réel. Pour une étude comparative de ces méthodes (incluant notamment les invariants intégraux utilisés dans ce travail), nous renvoyons le lecteur à (Labbani-I. *et al.*, 2011).

Ce papier partage aussi quelques idées avec le modèle spatial sémantique hiérarchique de Kuipers (Kuipers *et al.*, 2004) et le contrôle bas niveau utilisant le Voronoï généralisé de Choset (Choset, Nagatani, 2001). Ces travaux posent en effet quelques prérequis pour des algorithmes incrémentaux de localisation et de cartographie dans une approche topologique sans connaissance a priori de l'environnement.

3. Segmentation de l'espace libre

Le processus d'extraction de l'espace libre visible dans l'image est réalisé à l'aide d'un algorithme de segmentation temps réel présenté dans (Merveilleux *et al.*, 2011). Il est basé sur la propagation d'un contour actif qui délimite le sol et intègre notamment la gestion des faux obstacles isolés. Son initialisation, tâche critique en général, est ici largement simplifiée par la nature des images omnidirectionnelles (voir figure 1) associée à la position de la caméra relativement au robot. Le contour actif est ainsi initialisé de telle sorte à englober la projection du robot dans l'image.

Bien entendu, l'environnement étant inconnu, l'algorithme n'a aucune information concernant les dimensions ou la forme de l'espace libre étudié.

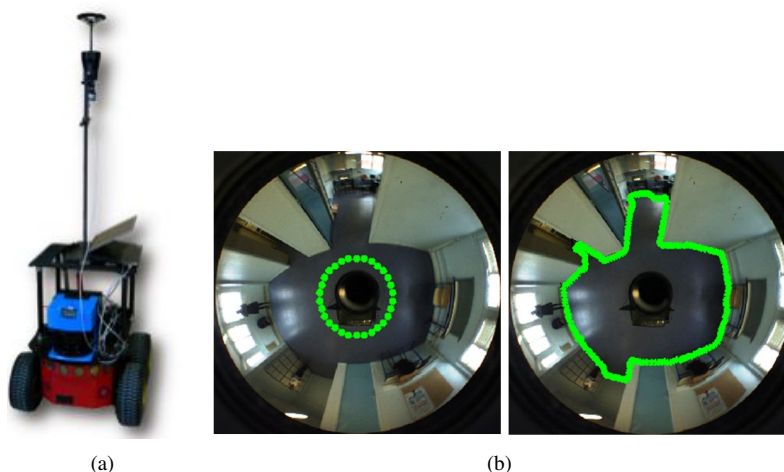


Figure 1. (a) : La plateforme expérimentale (robot Pioneer 3AT + capteur catadioptrique). (b) : Segmentation de l'espace libre en utilisant (Merveilleux *et al.*, 2011). Le contour actif est initialisé autour de la projection du robot dans l'image (centre), et se propage jusqu'à atteindre les frontières de l'espace libre

Au terme de la phase d'expansion, les frontières de l'espace libre produisent une partition de l'image en deux domaines bien distincts (figure 2) :

- la région intérieure FS , correspondant à l'espace libre, est utilisée pour l'extraction de la topologie locale (et globale par fusion), et sert de base à la commande du robot ;
- la partie extérieure (complémentaire) \overline{FS} , porteuse de la majorité de l'information photométrique pertinente, sert à construire une signature invariante utilisée comme représentation du lieu visible dans l'image.

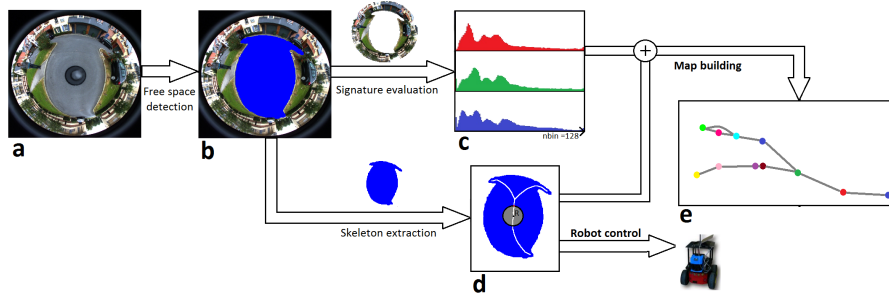


Figure 2. Etapes de construction de la carte topologique. (a) Image omnidirectionnelle, (b) Segmentation de l'espace libre, (c) Signature invariante construite depuis le complément de l'espace libre, (d) Squelette (topologie) de l'espace libre, et (e) Fusion squelette / signature de lieu, pour la construction incrémentale et automatique de la carte

4. Extraction de la topologie locale

Dans (Marie *et al.*, 2013), nous proposons le δ -medial axis, un algorithme de complexité quasi-linéaire, capable d'extraire en temps réel le squelette filtré d'une forme bruitée à partir d'un seul paramètre δ . Nous utilisons dans ce travail une version adaptée de cet outil capable de tenir compte des particularités liées au capteur catadioptrique afin d'extraire dans l'image la projection du diagramme de Voronoï de la scène observée. Ce dernier se trouve être une bonne représentation de la topologie locale.

Initialement défini par Blum (Blum, 1967), l'axe médian réel $S_{\mathbb{R}}$ d'une forme $\mathcal{X}_{\mathbb{R}} \subset \mathbb{R}^2$ peut être vu comme l'ensemble des points $x \in \mathcal{X}_{\mathbb{R}}$ qui sont le lieu d'une autocollision d'un front d'onde se propageant à vitesse constante et initialisé en tout point de $\partial\mathcal{X}_{\mathbb{R}}$, frontières de $\mathcal{X}_{\mathbb{R}}$. Autrement dit, il s'agit des points de $\mathcal{X}_{\mathbb{R}}$ à équidistance des deux points les plus proches sur la frontière $\partial\mathcal{X}_{\mathbb{R}}$:

DÉFINITION 1. — L'axe médian réel (Real Medial Axis) (RMA) $S_{\mathbb{R}}$ de $\mathcal{X}_{\mathbb{R}}$ est défini par l'ensemble des points $x \in \mathcal{X}_{\mathbb{R}}$ qui vérifient :

$$\exists y_1, y_2 \in \partial\mathcal{X} \text{ t.q. } \begin{cases} \|y_1 - x\| = \|y_2 - x\| \\ \forall z \in \partial\mathcal{X}, \|y_1 - x\| \leq \|z - x\| \end{cases} \quad (1)$$

L'utilisation de cet outil soulève généralement deux problématiques majeures :

- dans la plupart des applications, la forme étudiée est définie dans un espace discret (typiquement une image) qui rend la définition de Blum inadaptée. Une partie de la littérature s'attache donc à proposer une formulation alternative offrant une approximation discrète de l'axe médian réel ;
- le calcul d'axe médian est extrêmement sensible aux déformations aussi mineures soient-elles de la forme, générant pour chaque défaut une branche qui, la plu-

part du temps, n'est pas désirée (voir figure 3). De nombreux travaux (Chaussard *et al.*, 2011 ; Hesselink, 2007 ; Couprie *et al.*, 2007) proposent donc des processus de filtrage pour contrôler leur apparition.



Figure 3. Mise en évidence de la problématique liée aux déformations des frontières de la forme. En (a), axe médian obtenu à partir du Integer Medial Axis (IMA) (Hesselink, 2007), une méthode n'intégrant pas de processus d'élagage (suppression des branches parasites). En (b), la même image soumise au DMA pour une valeur du paramètre δ optimisée

Le principe du δ -medial axis a été spécialement défini afin de proposer une solution à ces deux problématiques tout en présentant une complexité quasi-linéaire permettant son utilisation dans un contexte temps réel.

Soit $\mathcal{X}_{\mathbb{Z}}$ une forme définie comme un sous-ensemble de \mathbb{Z}^2 . Pour tout \mathbf{x} de $\mathcal{X}_{\mathbb{Z}}$, on introduit la projection $\mathcal{P}(\mathbf{x})$ comme étant le point de la frontière $\partial\mathcal{X}_{\mathbb{Z}}$ à distance minimale de \mathbf{x} :

$$\mathcal{P}(\mathbf{x}, \partial\mathcal{X}_{\mathbb{Z}}) = \mathcal{P}(\mathbf{x}) = \{\mathbf{y} \in \partial\mathcal{X}_{\mathbb{Z}} \mid \forall \mathbf{z} \in \partial\mathcal{X}_{\mathbb{Z}}, d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z})\} \quad (2)$$

La distance correspondante, nommée distance au bord et notée indifféremment $\mathcal{D}(\mathbf{x}, \partial\mathcal{X}_{\mathbb{Z}})$ ou $\mathcal{D}(\mathbf{x})$ s'écrit donc :

$$\mathcal{D}(\mathbf{x}) = \min_{\mathbf{y} \in \partial\mathcal{X}_{\mathbb{Z}}} d(\mathbf{x}, \mathbf{y}) \quad (3)$$

A noter que $d(\mathbf{x}, \mathbf{y})$ désigne, sauf indication contraire, la distance euclidienne : $d(\mathbf{x}, \mathbf{y}) = d_E(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

Le δ -medial axis peut alors être défini comme suit :

DÉFINITION 2. — Soit $\delta \in \mathbb{R}_+^*$, le δ -medial axis de $\mathcal{X}_{\mathbb{Z}}$ est l'ensemble des points $\mathbf{x} \in \mathcal{X}_{\mathbb{Z}}$ qui vérifient (Marie *et al.*, 2013) :

$$\exists \mathbf{y} \in \mathcal{X}_{\mathbb{Z}}, \|\mathbf{x} - \mathbf{y}\| = 1 \text{ t.q. } \begin{cases} \exists \mathbf{z} \in [\mathcal{P}(\mathbf{x}), \mathcal{P}(\mathbf{y})] \mid \mathcal{D}(\mathbf{z}) > \delta \\ \& \\ \mathcal{D}(\mathbf{x}) \geq \mathcal{D}(\mathbf{y}). \end{cases} \quad (4)$$

Cette formulation évalue chaque couple de points voisins et cherche à déterminer, d'une part, s'il existe un point du RMA sur le segment continu les reliant ; et d'autre

part, de vérifier si celui-ci appartient à une branche parasite ou non. La première condition, $(\exists \mathbf{z} \in [\mathcal{P}(\mathbf{x}), \mathcal{P}(\mathbf{y})] \mid \mathcal{D}(\mathbf{z}) > \delta)$ est la clé de voûte de la méthode. Elle considère les projections respectives des deux points voisins, et vérifie s'il est possible d'insérer un disque de rayon δ centré sur un point \mathbf{z} du segment les reliant. Si tel est le cas, cela implique d'une part qu'il existe bien un point du RMA entre \mathbf{x} et \mathbf{y} (puisque dans ce cas on a $d(\mathcal{P}(\mathbf{x}), \mathcal{P}(\mathbf{y})) \geq 2\delta$), et d'autre part, que ce point n'appartient pas à une branche parasite au sens de δ (qui peut être vu comme la déformation minimale sur $\partial\mathcal{X}$ nécessaire pour générer une branche de l'axe médian). La seconde condition $(\mathcal{D}(\mathbf{x}) \geq \mathcal{D}(\mathbf{y}))$, garantit la finesse de l'axe médian produit, puisque seul le point qui constitue la meilleure approximation du RMA est ajouté (celui qui se trouve le plus loin de $\partial\mathcal{X}$).

Le DMA ainsi calculé est donc fondé sur une métrique euclidienne appliquée à l'image. Dans le but d'extraire la topologie locale de la scène perçue, nous proposons d'utiliser une nouvelle métrique adaptée à la géométrie du capteur catadioptrique omnidirectionnel et qui permet d'extraire directement dans l'image la projection de l'espace libre 3d réel.

Soient FS_W l'espace libre 3d (c'est-à-dire le sol, supposé plan, sur lequel le robot peut se déplacer sans risque), \mathcal{V} le diagramme de Voronoï qui en découle, et f_p la fonction qui projette un point 3d dans le plan image. Alors, nous cherchons à construire une version adaptée du DMA, notée ODMA (pour *Omnidirectional δ -Medial Axis*) telle que :

$$\text{ODMA}(FS) = f_p(\mathcal{V}(FS_W)) \quad (5)$$

Ce qui revient à définir une métrique adaptée \tilde{d} dans l'image qui soit proportionnelle à la distance euclidienne dans la scène :

$$\forall p_1, p_2 \in FS, \tilde{d}(p_1, p_2) = \alpha d_E(f_p^{-1}(p_1), f_p^{-1}(p_2)) \quad (6)$$

avec α un facteur d'échelle.

L'évaluation de cette métrique passe par la mise en équation puis l'inversion de la fonction de projection f_p . Dans (Barreto, Araújo, 2001 ; Geyer, Daniilidis, 2001), un modèle de projection unifié est proposé pour modéliser la projection d'un point 3d dans le plan image d'un capteur vérifiant la contrainte du point de vue unique. La fonction f_p peut alors être écrite comme la composition de deux fonctions de projection intermédiaires f_s et f_i , d'abord sur la sphère unitaire, puis sur le plan image, qui s'écrivent :

$$f_p = f_i \circ f_s \quad \text{avec}$$

$$f_s : P \in \mathbb{R}^3 \mapsto P_s \in \mathcal{S}^2 \quad t.q. \begin{cases} X_s = X/\rho \\ Y_s = Y/\rho \\ Z_s = Z/\rho \end{cases}, \quad \rho = \sqrt{X^2 + Y^2 + Z^2} \quad (7)$$

$$f_i : P_s \in \mathcal{S}^2 \mapsto p \in \mathbb{R}^2 \quad t.q. \begin{cases} x = X_s/(Z_s + \xi) \\ y = Y_s/(Z_s + \xi) \end{cases} \quad (8)$$

où ξ est un paramètre lié au type de miroir. Par les lois de composition, on a $f_p^{-1} = f_s^{-1} \circ f_i^{-1}$. L'opérateur f_i s'inverse sans difficulté, et on trouve :

$$P_s = f_i^{-1}(p) : \begin{cases} X_s = A(p)x \\ Y_s = A(p)y \\ Z_s = A(p) - \xi \end{cases} \quad \text{avec } A(p) = \frac{\xi + \sqrt{1 + (1 - \xi^2)(x^2 + y^2)}}{x^2 + y^2 + 1} \quad (9)$$

Pour l'opérateur f_s , on émet l'hypothèse que la caméra est positionnée de telle sorte que le plan du sol soit parallèle au plan image. On part également du principe que les points projetés appartiennent tous au plan du sol, et partagent donc tous la même altitude ($Z = H$). Alors, f_s peut être inversé, et on trouve :

$$P = f_s^{-1}(P_s) : \begin{cases} X = H X_s / Z_s \\ Y = H Y_s / Z_s \\ Z = H \end{cases} \quad (10)$$

En combinant (9) et (10), l'opérateur de projection inverse s'écrit :

$$P = f_p^{-1}(p) : \begin{cases} X = H A(p)x / (A(p) - \xi) \\ Y = H A(p)y / (A(p) - \xi) \\ Z = H \end{cases} \quad (11)$$

En repartant de (6), on peut écrire pour $P_1 = (X_1, Y_1, H)^t$ et $P_2 = (X_2, Y_2, H)^t$:

$$\tilde{d}(p_1, p_2)^2 = \alpha^2 ((X_1 - X_2)^2 + (Y_1 - Y_2)^2) \quad (12)$$

Qui, en posant $\alpha = 1/H^2$ et en utilisant (11) mène à l'expression finale de notre métrique adaptée :

$$\tilde{d}(p_1, p_2)^2 = \frac{A^2(p_1)}{(A(p_1) - \xi)^2} (x_1^2 + y_1^2) + \frac{A^2(p_2)}{(A(p_2) - \xi)^2} (x_2^2 + y_2^2) - 2 \frac{A(p_1)A(p_2)}{(A(p_1) - \xi)(A(p_2) - \xi)} (x_1 x_2 + y_1 y_2) \quad (13)$$

avec $A(p_i) = \frac{\xi + \sqrt{1 + (1 - \xi^2)(x_i^2 + y_i^2)}}{x_i^2 + y_i^2 + 1}$, $i = 1, 2$

Cette métrique adaptée permet maintenant de définir dans l'image une projection et une distance au bord adaptées qui s'expriment par :

$$\forall \mathbf{x} \in \mathbf{FS}, \begin{cases} \tilde{\mathcal{D}}(\mathbf{x}) = \min_{\mathbf{y} \in \partial \mathbf{FS}} \tilde{d}(\mathbf{x}, \mathbf{y}) \\ \tilde{\mathcal{P}}(\mathbf{x}) = \left\{ \mathbf{y} \in \partial \mathbf{FS} \mid \tilde{d}(\mathbf{x}, \mathbf{y}) = \tilde{\mathcal{D}}(\mathbf{x}) \right\} \end{cases} \quad (14)$$

On peut finalement proposer une définition pour le ODMA :

DÉFINITION 3. — Soit $\delta \in \mathbb{R}_+^*$. Le omnidirectional δ -medial axis (ODMA) de \mathbf{FS} , est l'ensemble des points $\mathbf{x} \in \mathbf{FS}$ qui vérifient les propriétés suivantes :

$$\exists \mathbf{y} \in \mathbf{FS}, \|\mathbf{x} - \mathbf{y}\| = 1 \text{ t.q. } \begin{cases} \exists \mathbf{z} \in [\tilde{\mathcal{P}}(\mathbf{x}), \tilde{\mathcal{P}}(\mathbf{y})] \mid \tilde{\mathcal{D}}(\mathbf{z}) \geq \delta \\ \tilde{\mathcal{D}}(\mathbf{x}) \geq \tilde{\mathcal{D}}(\mathbf{y}) \end{cases} \quad (15)$$

A noter que cette fois les points sont exprimés dans le plan image normalisé, et non plus dans le repère pixelique. La fonction qui permet de passer de l'un à l'autre est fondée sur la matrice des paramètres intrinsèques et nécessite donc que le capteur soit calibré (c'est de toute façon nécessaire dans le processus de construction de signature). De plus, le paramètre δ doit être choisi en corrélation avec la métrique adaptée, c'est-à-dire en rapport avec les proportions du plan sol 3d. Afin de rendre le processus totalement automatique, sa valeur est choisie directement à partir du contenu image. En l'occurrence, soit $\Gamma(0)$ le contour actif initial utilisé dans le processus de segmentation de l'espace libre et exprimé dans le plan image normalisé. En notant $c = (0, 0)$ le centre de projection de la caméra exprimé dans ce même repère, on pose $\delta = \tilde{d}(\Gamma(0), c)$. Autrement dit, les branches du diagramme de Voronoï qui ne seront pas supprimées dans le processus d'élagage du ODMA seront celles qui conduisent à une portion de l'espace libre suffisamment large pour que le robot s'y déplace, ce qui coïncide parfaitement avec l'un des objectifs majeurs du travail, à savoir garantir une navigation à la fois autonome et sûre.

Afin de justifier l'apport du ODMA relativement à la version standard de l'algorithme, la figure 4 propose les topologies extraites pour différentes situations ainsi qu'une vue 2d du sol reprojété. Clairement, la version standard, bien que fournissant une topologie cohérente avec la scène observée, ne propose en aucun cas des trajectoires sûres.

D'un point de vue algorithmique, les performances de la méthode sont maintenues temps réel (de l'ordre de 10 Hz en incluant le processus de segmentation d'espace libre, 25 Hz sans) par l'utilisation d'algorithmes de complexité linéaire optimisés, limitant ainsi la quantité de traitements requise. Pour le détail sur les différentes étapes algorithmiques, nous renvoyons le lecteur à (Marie *et al.*, 2013) où chacune d'entre elles est abordée. Nous nous contentons ici de considérer le cas particulier de l'évaluation de la métrique adaptée. L'équation (13) laisse en effet entrevoir un temps de calcul important, du fait que le nombre de distances à évaluer pour l'extraction de l'axe médian est considérable. En posant $B(p) = A(p)/(A(p) - \xi)$ et $C(p) = B(p)^2(x^2 + y^2)^2$, l'équation (13) s'écrit alors :

$$\tilde{d}(p_1, p_2) = C(p_1) + C(p_2) - 2B(p_1)B(p_2)(x_1x_2 + y_1y_2) \quad (16)$$

L'idée est de pré-évaluer les valeurs B et C par l'intermédiaire de deux tables de correspondance (look-up tables), réduisant ainsi considérablement le temps requis pour calculer une distance adaptée entre deux points image. L'algorithme 1 propose une vue d'ensemble du processus d'extraction du ODMA.

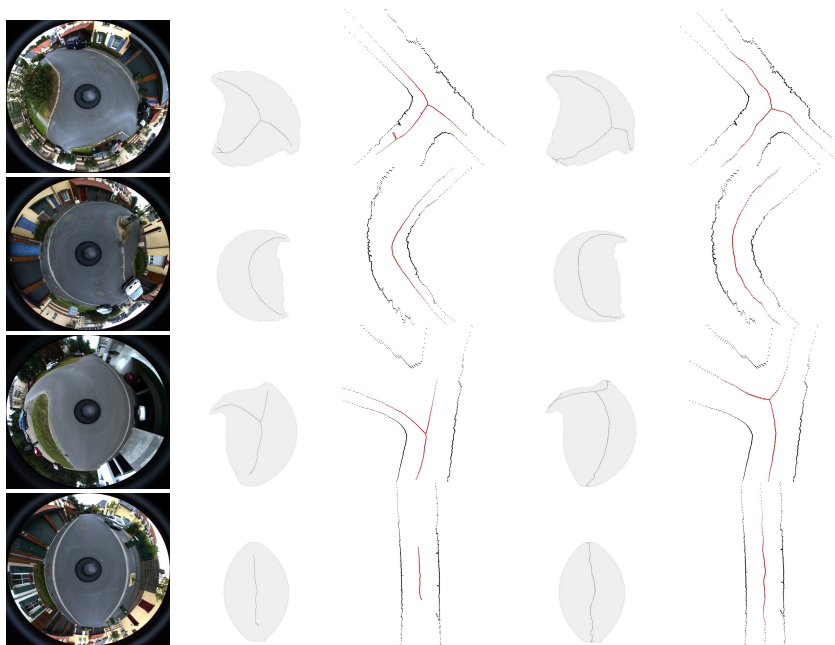


Figure 4. Résultats comparatifs entre le DMA et le ODMA. Première colonne : images en entrée choisies pour leur topologie intéressante. Deuxième et troisième colonnes : axe médian obtenu avec la métrique euclidienne et sa projection dans le plan sol. Quatrième et cinquième colonnes : idem dans le cas où la métrique adaptée est utilisée

Algorithme 1 Algorithme général du Omnidirectional δ -medial axis

Entrée : frontières de l'espace libre $\partial\mathbf{FS}$

Sortie : axe médian adapté et filtré ODMA

Pour chaque $\mathbf{x} \in \mathbf{FS}$
 Calcul de $\tilde{\mathcal{P}}(\mathbf{x})$ et $\tilde{\mathcal{D}}(\mathbf{x})$ (Danielsson, 1980)
Fin pour
Pour chaque $\mathbf{x}(i, j) \in \mathbf{FS}$, $\mathbf{y} \in \{(i-1, j), (i, j-1)\}$
 Si $d(\tilde{\mathcal{P}}(\mathbf{x}), \tilde{\mathcal{P}}(\mathbf{y})) > 2\delta$
 Si $\text{segmentSearch}(\tilde{\mathcal{P}}(\mathbf{x}), \tilde{\mathcal{P}}(\mathbf{y}))$
 Si $\tilde{\mathcal{D}}(\mathbf{x}) \geq \tilde{\mathcal{D}}(\mathbf{y})$, ajouter \mathbf{x} à ODMA **fin si**
 Si $\tilde{\mathcal{D}}(\mathbf{x}) \leq \tilde{\mathcal{D}}(\mathbf{y})$, ajouter \mathbf{y} à ODMA **fin si**
 Fin si
 Fin si
Fin pour

Procédure $\text{segmentSearch}(\tilde{\mathcal{P}}(\mathbf{x}), \tilde{\mathcal{P}}(\mathbf{y}))$

Entrée : deux extrémités du segment étudié

Sortie : vrai si $\exists \mathbf{z} \in [\tilde{\mathcal{P}}(\mathbf{x}), \tilde{\mathcal{P}}(\mathbf{y})]$ t.q. $\tilde{\mathcal{D}}(\mathbf{z}) > \delta$
 faux sinon

Pour chaque $\mathbf{z} \in [\tilde{\mathcal{P}}(\mathbf{x}), \tilde{\mathcal{P}}(\mathbf{y})]$
 Si $\tilde{\mathcal{D}}(\mathbf{z}) > \delta$
 Renvoyer vrai ; // Un point suffit, fin de la procédure
 Fin si
Fin pour
Renvoyer faux // Aucun point ne vérifie la condition

5. Construction de la signature de lieu

Nous utilisons la théorie de l'invariance de Haar pour définir une signature visuelle des lieux visités par le robot. Ce formalisme permet de combiner information visuelle et transformations induites par les mouvements du robot dans le processus d'indexation de lieu. Bien que les approches globales sont moins robustes aux occultations (Labrani-I. *et al.*, 2011), les travaux de (Charron *et al.*, 2006) ont montré de très bonnes performances d'indexation et de localisation basée apparence du robot dans des environnements intérieurs. En particulier, une telle signature est à la fois très rapide à calculer, et peu coûteuse en mémoire. Elle permet donc de s'ajouter à la boucle de traitement sans rompre la contrainte d'exécution temps réel, et permet la construction d'une carte topologique compacte et rapide d'utilisation.

Tout en conservant l'idée originale, nous introduisons une nouvelle formulation :

- nous proposons l'utilisation de descripteurs "région" à la place des descripteurs "contour" utilisés dans (Charron *et al.*, 2006), qui, bien qu'ils soient performants pour la représentation de lieu en environnement intérieur, manquent de stabilité et de robustesse dans les environnements extérieurs non structurés ;

- le processus d'intégration n'est réalisé que sur le complément de l'espace libre dans l'image, ce qui permet d'éliminer la zone la moins porteuse d'information. Le sol est en effet souvent homogène et commun aux différents lieux (goudron dans un scénario urbain par exemple), et, de surcroît, occupe souvent une zone importante de l'image ;

- l'étendue de l'espace libre est néanmoins implicitement prise en compte dans le processus de construction de la signature à travers une pondération utilisant les points frontières de l'espace libre.

Le formalisme des invariants intégraux de Haar est défini par l'équation :

$$H(\mathbf{x}) = \frac{1}{|G|} \int_G f(g\mathbf{x})dg; \quad |G| = \int_G dg; g \in G \quad (17)$$

L'invariant H d'une configuration \mathbf{x} (ici l'image acquise à un instant donné), est défini par l'application d'une fonction noyau f (à support compact) intégrée sur l'ensemble des transformations g d'un groupe G sur l'espace des configurations.

La configuration \mathbf{x} est obtenue en dépliant l'image omnidirectionnelle sur le plan $\Pi(\theta, \varphi)$ calculé à partir de la projection inverse déduite du modèle de la sphère équivalente (Geyer, Daniilidis, 2001). La région correspondante à l'espace libre est ensuite soustraite du domaine de \mathbf{x} afin de calculer l'invariant uniquement sur son complément $\Omega_{\overline{FS}} = \Pi_{\theta, \varphi} \setminus \Omega_{FS}$ (figure 5).

Le groupe de transformation G définit l'invariance de H . Il constitue dans ce travail l'ensemble des déplacements possibles du robot (translations et rotations), dans un monde supposé plan. En utilisant la représentation $\Pi(\theta, \varphi)$, le groupe G est paramétré en (θ, φ) , avec φ variant dans $[0; 2\pi[$ (champ de vue de l'image

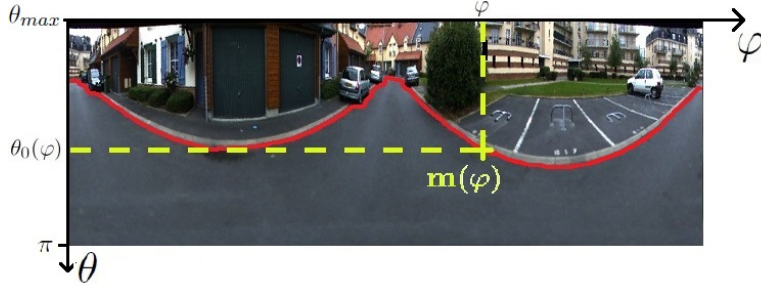


Figure 5. Représentation du plan $\Pi(\theta, \varphi) = \Omega_{\overline{FS}} \cup \Omega_{FS}$. En rouge, les frontières de l'espace libre séparent l'espace navigable Ω_{FS} (en dessous) et son complément $\Omega_{\overline{FS}}$ (au dessus). Elles constituent le lieu de projection des points $\mathbf{m}(\varphi) = \mathbf{x}(\theta_0(\varphi), \varphi)$, utilisés dans le calcul de $f_r(\mathbf{x}(\theta, \varphi))$

omnidirectionnelle), et θ dans $[\pi/2; \theta_0(\varphi)]$ où $\theta_0(\varphi)$ est la limite supérieure de θ qui dépend de l'étendue de l'espace libre dans la direction correspondante à φ (la valeur $\theta_0(\varphi) \rightarrow \pi$ si $\Omega_{\overline{FS}}$ couvre l'image entière) (figure 5).

La fonction noyau f est théoriquement une fonction intégrable à support compact, mais son choix peut s'avérer critique dans la pratique. Dans (Labbani-I. *et al.*, 2011), un noyau basé sur une différence de Gaussiennes est proposé et donne des résultats convaincants en environnement intérieur. Néanmoins, en extérieur, les scènes manquent souvent d'éléments présentant des contours structurés, ce qui réduit le taux de reconnaissance de la méthode.

Pour palier ce problème, nous proposons un noyau $f_{\theta, \varphi}$ qui combine une composante monomiale photométrique et une composante relationnelle basée sur les distances :

– Le noyau monomial f_m code la structure photométrique locale des pixels et est défini par :

$$f_m(\mathbf{x}(\theta, \varphi)) = \frac{1}{\sum_j} \sqrt{\prod \mathbf{x}^j(\theta', \varphi')} \quad (18)$$

où $\theta' = \theta + kd\theta$ et $\varphi' = \varphi + ld\varphi$ avec k, l, j entiers, et $\sum j$ l'ordre du noyau (on utilise ici un noyau d'ordre 2 généré par $(k=0, l=1)$ et $(k=2, l=0)$, ce qui donne : $f_m(\mathbf{x}(\theta, \varphi)) = \sqrt{\mathbf{x}(\theta, \varphi + d\varphi) \cdot \mathbf{x}(\theta + 2d\theta, \varphi)}$).

– Le noyau relationnel f_r ajoute quant à lui une information de distance entre les objets et le robot, basée sur les frontières de l'espace libre :

$$f_r(\mathbf{x}(\theta, \varphi)) = \frac{d_2(\mathbf{m}(\varphi), \mathbf{m}(\varphi - d\varphi)) + d_2(\mathbf{m}(\varphi), \mathbf{m}(\varphi + d\varphi))}{2} \quad (19)$$

où d_2 correspond à la distance euclidienne et $\mathbf{m}(\varphi)$ sont les coordonnées 2d métriques des points frontières de l'espace libre $\mathbf{x}(\theta_0, \varphi)$. Ces points (d'altitude zéro) sont reconstruits à partir des équations de reprojection du capteur catadioptrique (calibré).

Le noyau final $f_{\theta,\varphi} = f_r(\mathbf{x}(\theta, \varphi)) \cdot f_m(\mathbf{x}(\theta, \varphi))$ est appliqué séparément sur chacun des 3 canaux RGB de l'image couleur, afin d'augmenter la discriminance de la signature obtenue.

Le processus d'intégration permet finalement de calculer l'invariant de Haar H pour le groupe de transformations G . Pour chaque canal $c = r, g, b$, l'invariant H_c s'écrit dans le domaine discret :

$$\begin{aligned} H_c(\mathbf{x}) &= \frac{1}{|G_{\theta,\varphi}|} \sum_{\mathbf{x}(\theta,\varphi) \in \Omega_{\overline{FS}}} f_{\theta,\varphi}(\mathbf{x}(\theta, \varphi)) \sin \theta \, d\theta \, d\varphi \\ &= \frac{1}{|G_\varphi|} \sum_{\varphi=0}^{2\pi} \left[\frac{1}{|G_\theta|} \sum_{\theta=\pi/2}^{\theta_0(\varphi)} f_{\theta,\varphi}(\mathbf{x}(\theta, \varphi)) \sin \theta \, d\theta \right] d\varphi \end{aligned} \quad (20)$$

$$\text{avec } |G_\varphi| = \sum_0^{2\pi} d\varphi \text{ et } |G_\theta| = \sum_{\pi/2}^{\theta_0(\varphi)} \sin \theta \, d\theta.$$

Pour améliorer la discriminance de la signature, on remplace l'intégration sur φ , c-à-d $\sum_\varphi[\dots]d\varphi/|G_\varphi|$, par un histogramme normalisé (128 bins).

Les différentes étapes nécessaires à l'obtention de la signature de lieu sont résumées dans l'Algorithme (2).

Algorithme 2 Construction de la signature.

Entrées : Image omni, Frontières de l'espace libre $\partial\mathbf{FS}$ (Fig. 2-(b))

Sortie : Signature de Haar $H(\mathbf{x})$

```

Calcul de  $\Pi_{\theta,\varphi}$  et  $\mathbf{x}$  avec la résolution désirée;
Projection des limites de l'espace libre  $\partial\mathbf{FS}$  dans  $\Pi_{\theta,\varphi}$ ;
Pour chaque  $c$  dans  $r, g, b$ 
  Pour chaque  $\varphi$ 
     $sum_c = 0$  (initialisation)
    Pour chaque  $\theta$ 
      Si  $\mathbf{x}(\theta, \varphi) \in \Omega_{\overline{FS}}$  (complément de l'espace libre)
         $sum_c = sum_c + f_{\theta,\varphi}(\mathbf{x}(\theta, \varphi)) \cdot \sin \theta \, d\theta$ ; (Eq. 20)
      finSi
    finPour
    Ajout de  $sum_c$  dans l'histogramme  $Hist_c$ ;
  finPour
Normalisation de  $Hist_c$ ;
finPour
return  $H(\mathbf{x}) = \{Hist_c\}_{c=r,g,b}$  (Fig.2-(c)).

```

Afin d'évaluer la similarité entre deux signatures, on utilise la mesure basée sur l'**intersection d'histogrammes** (Eq. (21)). D'après (Barla *et al.*, 2003), cette mesure possède de bonnes propriétés pour la classification d'images couleur.

$$d_\cap(H, K) = \frac{1}{3} \sum_{c=r,g,b} \frac{\sum_i \min(h_{c,i}; k_{c,i})}{\sum_i k_{c,i}} \quad (21)$$

avec $H = \{h_i\}_{i=1,\dots,n}$ et $K = \{k_i\}_{i=1,\dots,n}$ les histogrammes des deux signatures comparées, n le nombre de bins de chaque histogramme. Etant donné que les signatures sont normalisées, on a $\sum_i k_{c,i} = 1$. A noter que deux signatures totalement identiques vont présenter une similarité de 1, et que plus la différence augmentera, plus le score obtenu sera faible.

6. Exploration autonome et construction de carte

A l'état initial, le robot commence son processus d'exploration. Il se déplace de manière autonome en suivant les chemins disponibles extraits à partir du squelette de l'espace libre dans l'image. La stratégie d'exploration est basée sur la sélection d'une des branches du squelette (figure 2d). En priorité, le robot suivra la branche la plus proche de sa trajectoire actuelle. Lorsqu'il atteint une intersection, la branche la plus à gauche dans l'image est sélectionnée par défaut (l'image étant inversée, le robot se déplacera donc vers la droite). Les branches non visitées sont enregistrées en tant que points d'intérêt topologiques, mais, à l'heure actuelle, aucune stratégie n'est définie pour revenir les explorer plus tard.

Dans le même temps, une signature $H_{t=0}$ correspondant à la position initiale est construite, et un premier nœud topologique n_0 est ajouté à la carte. Au cours de l'exploration, de nouvelles signatures sont créées et permettent au robot de se localiser et de définir de nouveaux lieux (voir algorithme 3). Un lieu est défini par un nœud de la carte. Pour changer de lieu, la différence de la signature courante avec la signature du nœud doit être inférieure à un seuil τ_1 ($\tau_1 = 0,9$ dans ce cas d'étude). Dans le cas où la mesure de similarité est inférieure à ce seuil, le robot élargit la recherche à l'ensemble des nœuds de la carte et trouve l'appariement le plus proche. Si la similarité est suffisamment marquée (plus grande qu'un second seuil $\tau_2 = 0,8$), le robot considère qu'il a atteint un lieu déjà visité (fermeture de boucle). Dans le cas contraire, il crée un nouveau nœud qui intègre la nouvelle signature et qui sera associé à un lieu pour l'instant inconnu. Cette stratégie à deux seuils offre une certaine robustesse au processus de construction de la carte : avec le premier seuil, on détecte un potentiel déplacement dans la carte topologique, tandis qu'avec le second, on décide si le nouveau nœud atteint existe déjà dans la carte (fermeture de boucle) ou bien s'il s'agit d'un nouveau lieu.

Dans tous les cas, à chaque fois que le robot change de lieu (de nœud), un lien reliant le dernier nœud visité et celui qui vient d'être atteint est créé. Ces liens étant générés à partir du squelette de l'espace libre, ils permettent d'ancrer les représentations photométriques (signatures de lieu) dans l'environnement parcouru par le robot.

L'algorithme global d'exploration, de construction et de mise à jour de la carte topologique est résumé dans l'algorithme 4. L'ensemble du processus, de l'acquisition de l'image jusqu'à la commande est réalisé de manière totalement autonome par le robot. Ce dernier s'arrête lorsqu'un temps maximal d'exploration (prédéfini) est atteint, ou si la carte n'évolue plus pendant une période de temps fixée (exploration supposée complète).

Algorithme 3 Construction et mise à jour de la carte

Entrées : Signature H_t du nœud courant n_i , Signature courante H_{t+1} , Map \mathcal{M}_t
Sortie : Nouvelle carte \mathcal{M}_{t+1} (mise à jour) avec nœud courant

```

CurrentNode ←  $n_i$ ;
Si  $d_{\cap}(H_t, H_{t+1}) < \tau_1$ 
  Trouver  $n^* = \arg \max_{n_k \in \mathcal{M}(t)} d_{\cap}(H_k, H_{t+1})$ ; ( $H^*$  signature associée à  $n^*$ )
  Si  $d_{\cap}(H^*, H_{t+1}) \geq \tau_2$ 
    Ajouter connexion  $n_i \mapsto n^*$ ;
    CurrentNode ←  $n^*$ ;
  Sinon
    Créer nœud  $n$  défini par la signature  $H_{t+1}$ ;
    Ajouter connexion  $n_i \mapsto n$ ;
    CurrentNode ←  $n$ ;
  finSi
   $\mathcal{M}_{t+1} = \mathcal{M}_t \cup \{CurrentNode\}$ ;
Sinon
   $\mathcal{M}_{t+1} = \mathcal{M}_t$ ;
finSi
return  $\mathcal{M}_{t+1}, CurrentNode$ 

```

A noter que les choix algorithmiques réalisés permettent un comportement temps réel du robot ainsi qu'une utilisation modérée de la mémoire. En particulier,

- chaque signature se présente sous la forme de trois histogrammes chacun formé de 128 bins (codés sur 4 octets). Quelle que soit la complexité de l'image menant à la génération d'un nouveau nœud, la taille de ce dernier est donc fixe, et égale à 384 octets ;

- le processus de localisation implique, lorsque nécessaire, de comparer la signature courante avec chaque nœud de la carte topologique construite. La mesure de distance utilisée, l'intersection d'histogramme, présente une complexité linéaire en nombre de bins. Ce dernier étant constant pour l'ensemble des nœuds, le temps de recherche de correspondance augmente donc linéairement avec le nombre de nœuds. A moins que le nombre de nœuds de la carte créée par le robot, n'explose de manière exponentielle, le temps nécessaire au processus de localisation est donc tout à fait négligeable relativement aux autres traitements.

Algorithme 4 Vue sommaire de l'algorithme.**Faire**

- Acquisition de l'image omnidirectionnelle ;
1. Segmentation de l'espace libre dans l'image (Algorithme (Merveilleux *et al.*, 2011));
2. Extraction de la topologie locale (Algorithme 1);
3. Calcul de la signature de lieu (Algorithme 2);
4. Création / Mise à jour de la carte (Algorithme 3);
5. Déplacement du robot le long du squelette de l'espace libre

Tant que *NotEnd*

7. Validation et expérimentations en conditions réelles

Toutes les expériences sont menées sur une plateforme Pioneer 3-AT (voir figure 1) équipée d'un capteur de vision omnidirectionnelle paracatadioptrique couleur. Les algorithmes sont exécutés en ligne sur un processeur Intel I7-2620m capable de réaliser, pour chaque image (de résolution 800×600), l'ensemble des traitements (segmentation de l'espace libre, extraction du squelette, construction de la signature, mise à jour de la carte et commande du robot), en moins de 200 ms (sans parallélisation des calculs).

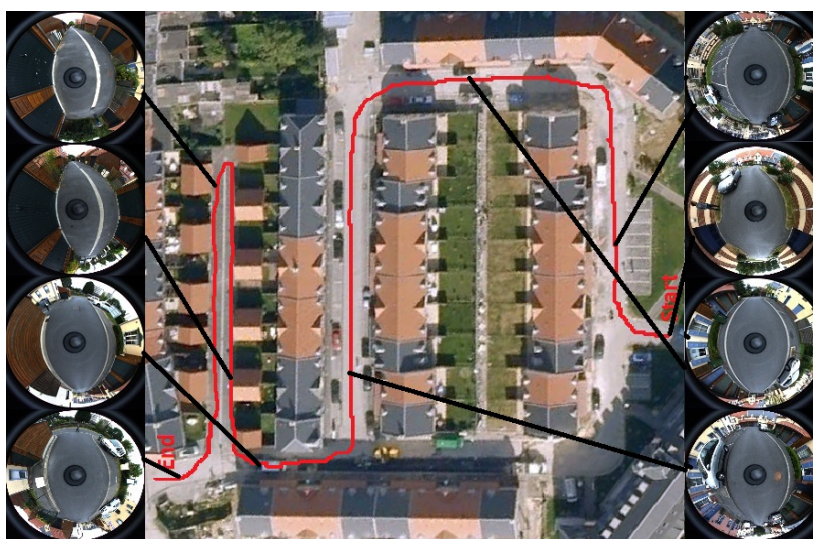


Figure 6. Vue Google Earth de l'environnement de test, avec un ensemble de vues présentées à gauche et à droite. En rouge, la trajectoire que le robot a suivie sans aucune intervention humaine

Plusieurs expériences ont été menées (en intérieur comme en extérieur) pour tester la capacité des algorithmes proposés, à explorer et à cartographier de manière autonome différents types d'environnements inconnus. Les scénarios proposés vont d'un simple couloir en intérieur jusqu'à plusieurs centaines de mètres dans différentes situations en extérieur. Dans ce papier, nous présentons l'étude d'un scénario en extérieur sur un peu plus de 350 m présentant des lieux différents avec ambiguïté perceptive importante (voir figure 6).

La figure 7 présente la carte construite par le robot lors de l'exploration autonome. Malgré les fluctuations qui surviennent dans la segmentation des espaces libres instantanés, l'extraction du squelette (construit pour être robuste au bruit), permet de guider le robot de manière sûre tout au long de l'exploration. De plus, le squelette porte les propriétés géométriques et topologiques de l'environnement. Le robot empruntant deux fois le même chemin suivra donc une trajectoire très similaire (sous réserve que

la topologie de l'environnement ne change pas trop). Cette propriété ajoute une stabilité et une robustesse supplémentaire à la construction de signatures et donc à la localisation du robot dans la carte.

Sur un parcours de plus de 350 m, un total de 50 nœuds avec des liens cohérents composent la carte (voir figure 7). La fermeture de boucle (passage par des nœuds déjà visités) à la fin de la séquence est correctement détectée, et la topologie globale de la carte est conservée. Ces résultats sont obtenus avec les valeurs de seuil $\tau_1 = 0,9$ et $\tau_2 = 0,8$.

Comme attendu, on peut observer que la répartition des nœuds sur la carte n'est pas homogène, ce qui est spécifique des représentations topologiques : le robot peut se déplacer de plus de 20 m avant de considérer que les informations visuelles ont suffisamment changé (il crée alors un nouveau nœud). À l'inverse, dans certaines situations (carrefour par exemple), les signatures risquent de changer très rapidement (par exemple apparition d'un objet caché par un obstacle). Il peut ainsi y avoir plusieurs nœuds construits sur une distance réduite (ce phénomène est bien visible figure 7 sur la partie en bas à gauche, où la disposition des bâtiments fait que la vision du robot change très rapidement).

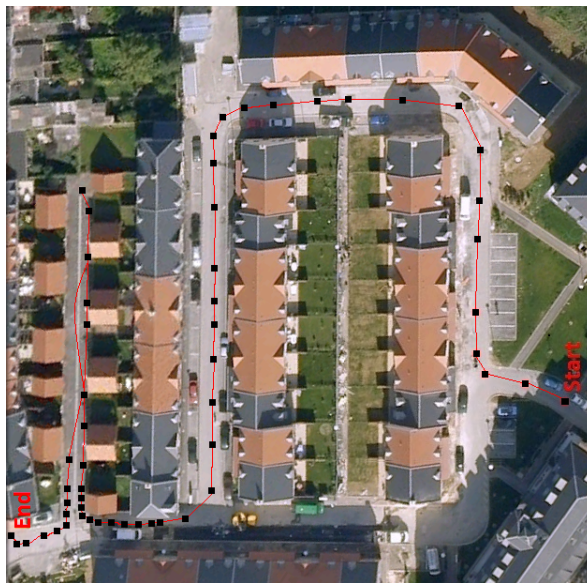


Figure 7. Carte topologique construite par le robot, superposée à la vue Google Earth. En noir, la position géographique des signatures ayant amenées à la création d'un nouveau nœud. En rouge, les liens existant entre les nœuds

Afin de valider la formulation des signatures comme représentation pertinente et empreinte de lieu, nous la comparons à deux variantes du processus de construction de signature. On s'intéresse aux effets sur la carte construite lorsque *i*) le noyau

relationnel est ignoré (seule l'information photométrique est prise en compte dans Eq. (18)-(20)); et *ii*) lorsque toute l'image est prise en compte au lieu de restreindre le calcul d'invariant au complément de l'espace libre.

Les graphiques figure 8 présentent pour chaque signature le nœud auquel elle a été associée (c-à-d, la position du robot dans la carte topologique à chaque image acquise). Commençons par comparer la signature proposée dans ce papier (figure 8a) avec celle utilisant uniquement le noyau monomial sur Ω_{FS} (figure 8b). L'aspect global des résultats est similaire, mais deux différences majeures sont observables.

Pour le même environnement, 82 nœuds sont nécessaires pour construire la carte topologique en utilisant le seul noyau monomial, contre seulement 50 avec la nouvelle formulation combinant noyau monomial et noyau relationnel. Cela s'explique par le fait que, pour les objets proches du robot (voitures, murs, ...), un faible déplacement de ce dernier engendre des variations perceptives importantes de ces objets. C'est exactement pour compenser ce phénomène que le noyau relationnel a été introduit.

La deuxième différence est l'incapacité du noyau monomial à localiser la fermeture de boucle dans certaines situations (voir fin de parcours figure 7). Dans cet exemple, la ligne continue de séparation des voies navigables délimite naturellement les frontières de l'espace libre dans l'image, le robot emprunte alors des chemins différents à l'aller et au retour sur ce parcours. Les variations de perception visuelle sur chacun des chemins conduisent à des signatures (basées noyau monomial) différentes. Ainsi, des nouveaux nœuds sont créés dans la carte et ne permettent pas l'identification des lieux visités. Le noyau relationnel en revanche permet de compenser cette différence, et notre méthode identifie correctement que le robot est déjà passé par ces lieux.

Les résultats pour les signatures calculées par intégration sur l'ensemble de l'image sont donnés figure 8c. Les performances chutent de manière significative. Les liens entre les nœuds ne sont pas fidèles à la réalité. Non seulement la fermeture de boucle n'est pas détectée, mais en plus de fausses fermetures de boucle le sont. Même en faisant varier les valeurs des seuils, il n'a pas été possible d'atteindre un résultat de cartographie cohérent avec la réalité terrain sur l'ensemble du parcours. Avec un capteur positionné à 1,35 m au dessus du sol, l'espace libre Ω_{FS} occupe environ 40 % de l'image. De ce fait, de petites variations dans la texture ou la taille de l'espace libre domine les changements qui peuvent survenir dans la scène (apparition d'un nouvel obstacle par exemple), et donc induit de larges variations et instabilités dans la signature.

Une vidéo de cette expérimentation, présentant l'ensemble des représentations construites (espace libre, axe médian, signature et carte topologique), est proposée à l'adresse ci-dessous ¹ afin d'étayer les résultats de ce travail.

1. Nous encourageons vivement le lecteur à visionner la vidéo disponible à cette adresse : <http://home.mis.u-picardie.fr/marie/TS14/TS14.wmv>

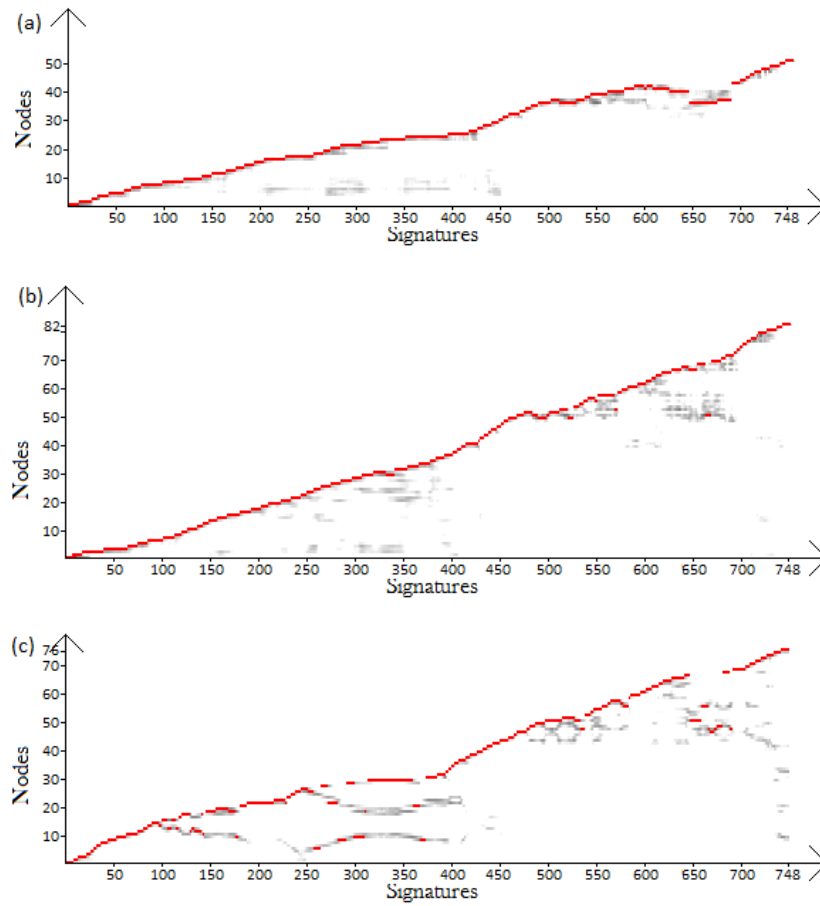


Figure 8. Graphe de correspondances entre nœuds et signatures. En rouge la position déduite de chaque signature pour : (a) l'approche proposée combinant noyaux monomial et relationnel, et intégrant sur $\Omega_{\overline{FS}}$; (b) noyau monomial seul sur $\Omega_{\overline{FS}}$; et (c) noyau monomial sur l'ensemble de l'image ($\Omega_{FS} \cup \Omega_{\overline{FS}}$). En gris, on présente également les nœuds dont la ressemblance est inférieure au nœud choisi, mais tout de même suffisamment élevée pour être notifiée (plus le pixel est sombre, plus la similarité est proche de celle du nœud choisi)

8. Conclusion

Dans ce papier, nous avons présenté une nouvelle approche pour l'exploration autonome et la construction de carte topologique en environnement inconnu. La carte construite par le robot est basée sur des caractéristiques perceptives extraites localement de l'image et structurées temporellement en représentations spatiales. Elle est enrichie incrémentalement pendant l'exploration.

La carte globale est organisée en temps réel autour d'une représentation topologique, définie par le squelette de l'espace libre de l'environnement, et d'une représentation photométrique, définie par les signatures des lieux visités pendant l'exploration. Les déplacements autonomes du robot sont générés sur le squelette de l'espace libre, qui permet une navigation au plus loin des obstacles, et définit un référencement naturel pour les signatures des lieux. La fermeture de boucle est gérée par l'identification de signatures visuelles construites par invariance à des transformations locales de l'image (invariants intégraux de Haar).

A travers l'étude d'un scénario en conditions réelles, nous avons validé à la fois l'exploration autonome et la construction de carte topologique dans un large environnement extérieur. Nous travaillons maintenant sur une stratégie d'exploration à plus haut niveau qui utilise les branches non explorées du squelette pour définir des objectifs précis et influencer directement sur la prise de décision lors de la détection de jonctions topologiques complexes.

Remerciements

Ce travail est financé par l'ANR (projet R-Discover), la Direction Générale de l'Armement (DGA) et le Fonds Européen de Développement Régional en Picardie (FEDER).

Bibliographie

- Angeli A., Doncieux S., Meyer J.-A., Filliat D. (2009). Visual topological slam and global localization. In *International conference on robotics and automation*, p. 2029–2034.
- Barla A., Odone F., Verri A. (2003). Histogram intersection kernel for image classification. In *International conference on image processing*, p. 513–516.
- Barreto J. P., Araújo H. (2001). Issues on the geometry of central catadioptric image formation. In *Computer vision and pattern recognition*, p. 422–427.
- Blum H. (1967). A transformation for extracting new descriptors of shape. In WWD (Ed.), *Models for the perception of speech and visual form*, p. 362–380. MIT Press.
- Chapoulie A., Rives P., Filliat D. (2013). Appearance-based segmentation of indoors/outdoors sequences of spherical views. In *Iros*, p. 1946–1951.
- Charron C., Labbani-Igbida O., Mouaddib E. M. (2006). On building omnidirectional image signatures using haar invariant features: Application to the localization of robots. In *Acivs*, p. 1099–1110.

- Chaussard J., Couprie M., Talbot H. (2011). Robust skeletonization using the discrete λ -medial axis. *Pattern Recognition Letters*, vol. 32, n° 9.
- Choset H., Nagatani K. (2001). Topological simultaneous localization and mapping (slam): Toward exact localization without explicit localization. *Transactions on Robotics and Automation*, vol. 17, n° 2, p. 125–137.
- Couprie M., Coeurjolly D., Zrou R. (2007). Discrete bisector function and euclidean skeleton in 2d and 3d. *Image and Vision Comp.*, vol. 25, n° 10, p. 1519–1698.
- Cummins M., Newman P. (2011). Appearance-only slam at large scale with fab-map 2.0. *International Journal of Robotics Research*, vol. 30, n° 9, p. 1100–1123.
- Danielsson P. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing*, vol. 14, n° 3, p. 227–248.
- Davison A. J., Reid I. D., Molton N. D., Stasse O. (2007). Monoslam: Real-time single camera slam. *PAMI Transactions*, vol. 29, n° 6, p. 1052–1067.
- Dayoub F., Cielniak G., Duckett T. (2011). Long-term experiments with an adaptive spherical view representation for navigation in changing environments. *Robotics and Autonomous Systems*, vol. 59, n° 5, p. 285–295.
- Durrant-Whyte H., Bailey T. (2006). Simultaneous localization and mapping (slam): Part i. *Robotics Automation Magazine*, vol. 13, n° 2, p. 99–110.
- Eade E., Drummond T. (2006). Scalable monocular slam. In *Computer vision and pattern recognition*, p. 469–476.
- Geyer C., Daniilidis K. (2001). Catadioptric projective geometry. *International Journal on Computer Vision*, vol. 45, n° 3, p. 223–243.
- Hesselink W. (2007). A linear-time algorithm for euclidean feature transform sets. *Information Processing Letters*, vol. 102, n° 5, p. 181–186.
- Konolige K., Agrawal M. (2008). Frameslam: From bundle adjustment to real-time visual mapping. *Transactions on Robotics*, vol. 24, n° 5, p. 1066–1077.
- Korrapati H., Courbon J., Mezouar Y., Martinet P. (2012). Image sequence partitioning for outdoor mapping. In *Icra*, p. 1650-1655.
- Kuipers B., Modayil J., Beeson P., MacMahon M., Savelli F. (2004). Local metrical and global topological maps in the hybrid spatial semantic hierarchy. In *International conference on robotics and automation*, p. 4845–4851.
- Labbani-I. O., Charron C., Mouaddib M. (2011). Haar invariant signatures and spatial recognition using omnidirectional visual information only. *Autonomous Robots Journal*, vol. 30, n° 3, p. 333–349.
- Lim J., Frahm J.-M., Pollefeys M. (2012). Online environment mapping using metric-topological maps. *I. J. Robotic Res.*, vol. 31, n° 12, p. 1394-1408.
- Marie R., Labbani-Igbida O., Mouaddib M. (2013). The delta-medial axis: a robust and linear time algorithm for euclidian skeleton computation. In *International conference on image processing*.
- Merveilleux P., Labbani-Igbida O., Mouaddib M. (2011). Robust free space segmentation using active contours and monocular omnidirectional vision. In *International conference on image processing*, p. 2877–2880.

- Milford M., Wyeth G. (2010). Persistent navigation and mapping using a biologically inspired slam system. *International Journal of Robotics Research*, vol. 29, n° 9, p. 1131-1153.
- Muhammad N., Fofi D., Ainouz S. (2009). Current state of the art of vision based slam. In *Spie image processing: Machine vision applications ii*, p. 72510F-72510F-12.
- Murillo A. C., Kosecka J. (2009). Experiments in place recognition using gist panoramas. In *Ieee workshop on omnidirectional vision, camera networks and non-classical cameras, held with iccv*, p. 2196-2203.
- Se S., Lowe D., Little J. (2002). Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, vol. 21, p. 735-758.
- Sim R., Elinas P., Little J. J. (2007). A study of the rao-blackwellised particle filter for efficient and accurate vision-based slam. *International Journal on Computer Vision*, vol. 74, n° 3, p. 303-318.
- Singh G., Kosecka J. (2010). Visual loop closing using gist descriptors in manhattan world. *International Conference on Robotics and Automation*.
- Valgren C., Lilienthal A. J., Duckett T. (2006). Incremental topological mapping using omnidirectional vision. In *International conference on intelligent robots and systems*, p. 3441-3447.

Reçu le 27/09/2013
Accepté le 12/05/2014

