

# Vers une cartographie sémantique d'environnements intérieurs

La recherche en cartographie, pour la robotique, porte de plus en plus sur des modèles d'environnements sémantiques contenant des concepts humains (tels que les pièces d'un bâtiment ou les objets). Nous avons développé un robot capable de construire de telles cartes dans le cadre d'une compétition d'exploration de bâtiments. Reposant sur les capteurs d'un type nouveau que sont les caméras de profondeur, ce système montre que la construction de tels modèles est possible, mais qu'elle repose sur des systèmes d'une grande complexité.

Par David FILLIAT\*

---

## INTRODUCTION

La robotique de service pour les particuliers en est encore à ses balbutiements. Elle fait l'objet de nombreux programmes de recherche, mais, hormis des succès dans des domaines très spécialisés (comme les robots aspirateurs), il n'existe pas aujourd'hui, commercialement, de robots pouvant assister efficacement des personnes dépendantes ou ranger un appartement. Les facteurs limitants sont multiples, que ce soit au niveau mécanique ou au niveau logiciel, mais l'une des principales limitations reste la capacité, pour le robot, de percevoir l'environnement et de l'interpréter. Ainsi, par exemple, apporter un objet à une personne demande des capacités (de détecter cet objet dans des situations très variées et de se déplacer dans un environnement changeant) qui ne sont pas encore suffisamment fiables pour être utilisées sans préparation, chez un particulier.

---

## LA CARTOGRAPHIE EN ROBOTIQUE

Pour des robots de service, la construction d'une carte de l'environnement est une fonction de base qui sert de support à leur navigation. Les cartes permettent, d'une part, d'obtenir une estimation fiable, à long terme, de la position du robot dans les environnements intérieurs où des méthodes telles que le GPS sont inutilisables et, d'autre part, de planifier des déplacements pour rejoindre des lieux distants. Un certain nombre d'applications sont bien sûr possibles sans cartographie : par exemple, certains robots aspirateurs se contentent de déplacements aléatoires qui finissent par couvrir tout l'environnement à traiter, sans planification préalable de trajectoire. Cependant, des applications de robotique de service plus complexes nécessiteront des capacités de localisation et

---

\* Enseignant-Chercheur à l'ENSTA ParisTech.

de planification évoluées (par exemple, pour aller chercher des objets et les apporter à leur utilisateur).

---

## UN DOMAINE MATURE

La qualité des capteurs utilisés pour la navigation est particulièrement importante pour l'obtention de bonnes performances. Ainsi, l'apparition des télémètres laser à balayage, dans les années 1990, a permis de quasiment résoudre le problème de la cartographie 2D, ce qui a permis l'apparition de solutions commerciales fiables. Le coût de ces capteurs reste assez élevé, mais il existe, par exemple, des modèles commercialisés de robots aspirateurs qui réalisent une cartographie leur permettant de nettoyer plus rapidement l'environnement. Pour des robots de plus grande taille, cependant, la cartographie 2D ne suffit plus et la majeure partie des recherches actuelles porte sur la cartographie 3D. De très nombreux travaux portent sur l'utilisation de la vision seule, qui peut fournir de bonnes performances, mais la démocratisation récente de caméras mesurant la profondeur, avec la commercialisation de la caméra Kinect (de Microsoft), a permis de faire des progrès rapides en cartographie 3D dense. Avec ce type de capteurs, il est non seulement envisageable de construire des cartes 3D, mais aussi d'aller plus loin, en intégrant des informations *sémantiques* au sein de ces cartes.

---

## UNE CARTOGRAPHIE SÉMANTIQUE

L'objectif de la cartographie sémantique est de construire des cartes contenant des informations de plus haut niveau que les cartes brutes (construites directement à partir d'un télémètre laser, par exemple). En effet, pour pouvoir agir efficacement, un robot a besoin de plus d'informations que sa position et celle des obstacles qui l'entourent. Pour chercher un objet, par exemple, il peut être utile de structurer l'environnement en pièces et de savoir si le robot se trouve dans une cuisine ou dans une chambre. De même, dans chaque pièce, il est utile de savoir que certains des obstacles sont des tables sur lesquelles on pourra trouver des objets, ou des placards qu'il faudra ouvrir pour y prendre des objets. Pour la navigation elle-même, une information de plus haut niveau aura son utilité, par exemple, pour déterminer qu'un obstacle est une chaise ou un rideau, c'est-à-dire des objets qui pourront être poussés, si besoin est, pour libérer le passage.

---

## COMPÉTITION CAROTTE

La cartographie sémantique est aussi utile pour un opérateur dans un contexte militaire. Lors de l'utilisation

d'un robot d'exploration, notamment, il serait plus aisé pour un opérateur dans un contexte de stress, d'analyser une carte qui fournirait directement des informations de haut niveau en termes de présence de couloirs, de pièces, de meubles et d'objets, que d'interpréter une simple carte des obstacles. Dans ce contexte, l'Agence Nationale pour la Recherche (ANR) et la Direction Générale de l'Armement (DGA) ont publié un appel d'offres, nommé CARTographie par ROboT d'un TErritoire (« Carotte »), dont l'objectif, pour les projets retenus, est de participer à un défi les opposant entre eux en vue de réaliser la cartographie sémantique d'un environnement intérieur. Le système présenté dans cet article a été développé dans le cadre de l'un des projets retenus, le projet Pacom (*Panoramic and Active Camera for Object Mapping*).

Ce défi se déroule dans des arènes modulaires composées de différentes pièces dans lesquelles sont disposés des objets de différentes tailles. L'objectif des compétiteurs est d'explorer ces arènes de manière complètement autonome, sans intervention de l'opérateur, et de fournir une carte de ces arènes en identifiant les pièces, les types de sols et de murs et les objets présents.

---

## LE PROJET PACOM

Le projet Pacom est commun à trois entités : l'École Nationale Supérieure de Techniques Avancées ParisTech (ENSTA ParisTech) (en charge de la navigation, de la cartographie, de l'exploration et d'une partie de la détection d'objets), l'Institut des Systèmes Intelligents et de Robotique (ISIR) de l'université Pierre et Marie Curie (en charge de la détection d'objets et des interfaces de contrôle) et la société Gostai (en charge de l'architecture logicielle).

Pour réaliser la cartographie sémantique proposée dans le défi Carotte, nous avons conçu un robot disposant de nombreux capteurs, auquel nous avons intégré plusieurs algorithmes de cartographie et de détection d'objets complémentaires afin d'obtenir un système qui présente une capacité d'adaptation suffisante et soit robuste.

---

## ARCHITECTURE MATÉRIELLE

Nous avons choisi de nous concentrer sur la problématique logicielle, sans chercher à développer une plateforme spécifiquement adaptée à un contexte militaire. La base mobile que nous avons choisie est le robot Pioneer 3 dx de la société Mobile Robots. Dans la version que nous avons utilisée, la plateforme est dotée de deux roues commandées indépendamment, de seize capteurs sonars disposés sur tout le pourtour du robot, d'un télémètre laser SICK LMS 200 et d'une caméra orientable Canon vcc50i. Le robot est équipé d'un microcontrô-

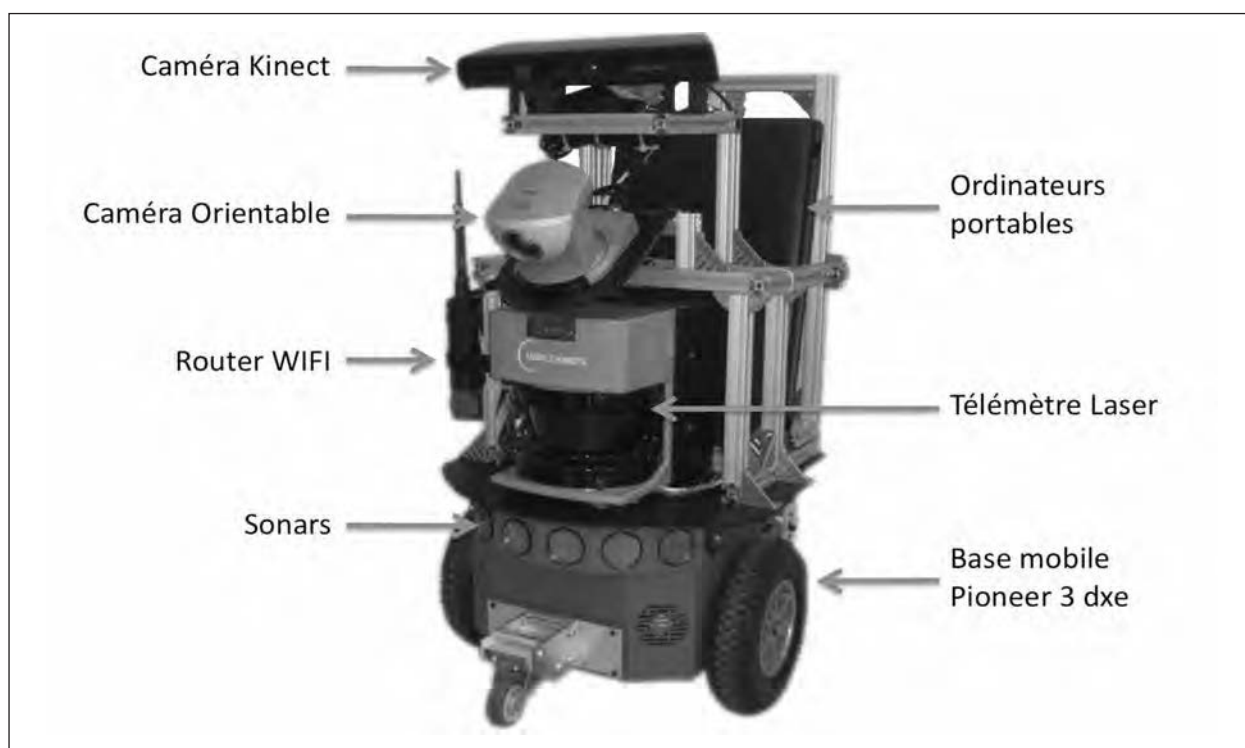


Figure 1 : Le robot développé dans le cadre du projet Pacom présenté au titre de la compétition Carotte.

leur assurant le contrôle des capteurs et l'asservissement des moteurs, et d'un ordinateur embarqué fonctionnant sous Linux (voir la figure 1).

Sur cette base, nous avons conçu une structure mécanique permettant d'ajouter une caméra de profondeur (Kinect de Microsoft) et deux ordinateurs portables, pour exécuter les différents algorithmes impliqués dans la cartographie sémantique. Pour diminuer les coûts et simplifier le développement, nous avons choisi d'utiliser des ordinateurs portables standards, qui ont, de plus, l'avantage de posséder leur propre alimentation. Les différents ordinateurs sont reliés par un réseau Ethernet au travers d'un routeur. Ce routeur est également un point d'accès Wifi qui permet de connecter une station de contrôle externe au robot, pour le contrôle d'exécution.

## ARCHITECTURE LOGICIELLE

L'architecture logicielle de l'ensemble du projet utilise le *framework* Urbi développé par la société Gostai. Ce *framework* est basé sur un *middleware* qui permet de distribuer sur plusieurs processeurs des composants écrits en langage C++ (*UObjects*). La mise en œuvre de ces composants se fait très simplement au travers d'un langage de script dédié à la robotique, urbiScript, qui permet de gérer de manière simple l'exécution parallèle des composants et le traitement des événements. L'utilisation de ce langage de script permet une grande souplesse pour la réalisation de tests interactifs et la modification du comportement global du robot.

Urbi permet donc de répartir simplement les calculs sur les différents ordinateurs embarqués sur le robot. De plus, le fait que les *UObjects* en mode distant soient des processus indépendants permet également de répartir naturellement les calculs sur les différents cœurs d'un processeur multi-cœurs. En effet, il suffit de lancer un *UObject* en mode distant sur le même ordinateur pour bénéficier d'un parallélisme certes assez simple, mais sans avoir besoin de programmer spécifiquement les composants dans ce but.

Pour le projet, nous avons donc développé un ensemble d'*UObjects* en langage C++ réalisant les différentes fonctionnalités nécessaires : interface avec le matériel, cartographies 2D et 3D, planification, suivi de trajectoire, exploration, détection d'objets et cartographie sémantique. L'ensemble de ces *UObjects* est réparti sur les différents processeurs du robot. Toute la mission du robot est codée en urbiScript, qui utilise les fonctionnalités de ces *UObjects* et coordonne leur activation.

## LA DÉTECTION D'OBJETS

La détection d'objets dans des environnements réalistes reste un problème difficile. Dans notre projet, elle est réalisée à partir des informations fournies par la caméra de profondeur en utilisant à la fois la structure 3D et l'apparence extraite de l'image couleur associée. Dans le cadre de la compétition, la majorité des objets sont connus à l'avance et il est donc possible d'entraîner les algorithmes de reconnaissance avant l'épreuve.

La première étape du processus de reconnaissance est la détection d'objets potentiels avant leur identification. A

partir du nuage de points 3D fourni par la caméra, nous commençons par détecter le plan du sol, dont la position est grossièrement connue, puis nous supprimons les points du sol et nous créons des groupes correspondant à tous les ensembles de points contigus restants. Les groupes sont ensuite filtrés et ceux présentant trop peu de points ou une taille trop faible sont supprimés. Les groupes restants correspondent alors aux objets et aux murs de l'environnement.

Les objets détectés doivent ensuite être identifiés. Pour cela, nous avons développé une nouvelle méthode multimodale qui utilise trois types d'informations pour permettre de traiter l'ensemble des objets : des informations de couleur (utiles pour des objets non texturés, comme les ballons), des informations de texture et des informations de forme 3D. Ces trois types d'informations sont traités selon la même approche, la méthode des sacs de mots visuels, qui représente les objets au moyen d'une collection d'informations locales. Cette approche est particulièrement intéressante pour gérer les occultations susceptibles de cacher une partie des objets.

Les informations nécessaires sont extraites de l'image couleur, dans laquelle, en projetant le nuage de points 3D, on a masqué tout ce qui n'est pas l'objet. Nous utilisons sur cette image des histogrammes de couleur locaux et des points d'intérêt SURF. L'information 3D utilise des *Surflets* calculées à partir de l'orientation relative des normales à l'objet, caractérisant ainsi sa forme globale. Pour chacun de ces types d'informations, nous

avons construit, d'avance, des modèles des objets. La reconnaissance se fait en comparant les modèles de l'objet inconnu à tous ceux des objets connus et en fusionnant les similarités obtenues pour les 3 types d'informations à l'aide d'une méthode d'apprentissage utilisant un réseau de neurones (voir la figure 2).

Notre méthode offre de très bonnes performances de reconnaissance lorsque les objets ont été correctement segmentés, mais une segmentation correcte est difficile à obtenir lorsque le robot explore son environnement de manière autonome. La performance globale de la détection d'objets dépend donc beaucoup de la disposition desdits objets : les objets isolés seront très bien reconnus, mais les objets placés dans des positions rendant difficile leur perception ne seront généralement pas détectés.

### CLASSIFICATION DES SOLS ET DES MURS

Pour la classification des sols et des murs, nous utilisons la caméra couleur orientable, que nous pointons vers le bas, ou vers l'avant. Ce processus est particulièrement important pour la détection de zones de gravier, que notre robot ne peut pas franchir. Nous avons utilisé, à cette fin, une méthode de classification de textures utilisant une forêt d'arbres aléatoires appliquée à des sous-images aléatoires. Cet algorithme d'apprentissage se base

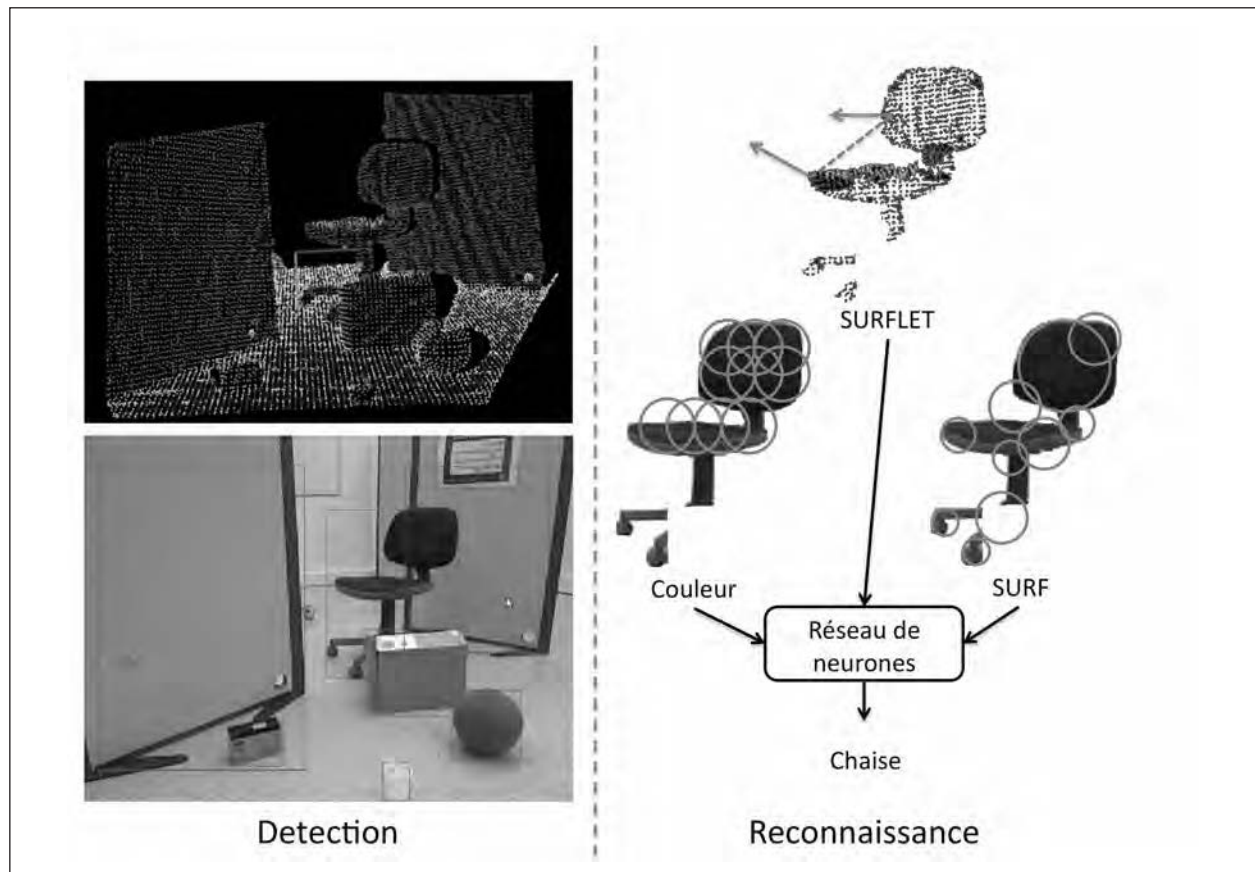


Figure 2 : Illustration de la méthode de détection et de reconnaissance d'objets.

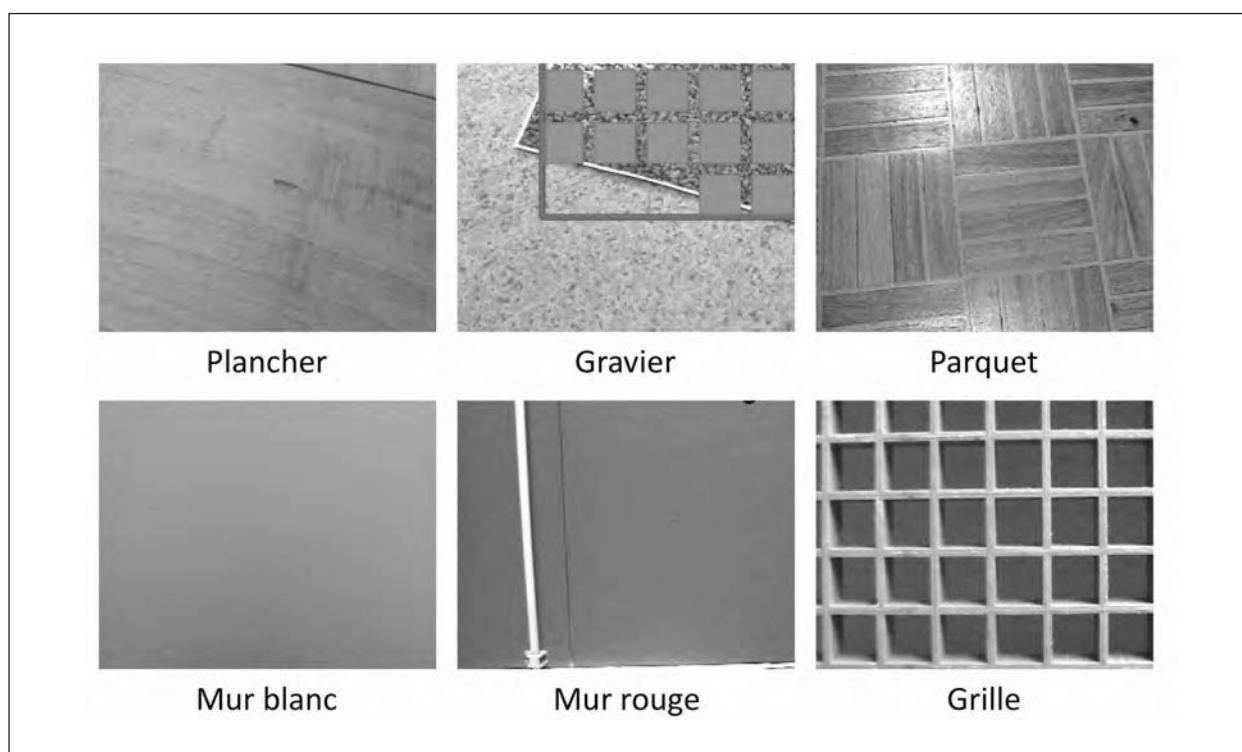


Figure 3 : Exemple de reconnaissances de types de sols et de murs.

simplement sur un ensemble d'images de chaque classe (voir la figure 3).

Cette méthode permet de prédire la catégorie des sols et des murs d'une manière très robuste et rapide. Elle a aussi l'avantage d'utiliser à la fois les informations portant sur la texture et la couleur, et donc de fournir une reconnaissance de couleur que l'on peut rendre très robuste aux variations d'éclairage, simplement en incluant ces variations dans la base d'apprentissage.

## LA GESTION DES OBSTACLES

La gestion robuste des obstacles, pour un robot de grande taille, requiert une gestion en 3D de l'environnement, une robustesse qui doit se maintenir même face à des vitres ou à des miroirs. A cette fin, nous intégrons donc les informations de tous nos capteurs. Le télémètre laser détecte de manière très précise les obstacles situés à environ 30 centimètres du sol, et ce sur toute la zone à l'avant du robot. Les capteurs sonars peuvent détecter des obstacles au sol ayant moins de 30 centimètres de hauteur, sur tout le pourtour du robot. Les sonars permettent également de détecter les vitres et les miroirs qui ne sont pas (ou mal) vus par le laser. Pour les obstacles plus grands, mais invisibles pour le laser et les sonars (tels que les tables ou les bancs), nous utilisons la caméra de profondeur. Nous projetons au sol les nuages de points des objets détectés et nous ajoutons le polygone formant l'enveloppe de ces points comme obstacle dans la carte 2D. Enfin, pour les zones de gravier (que notre robot ne peut pas franchir), nous utilisons la

méthode de classification des sols par vision et nous ajoutons à la carte le polygone englobant le gravier détecté. La carte résultante permet ainsi au robot de se déplacer de manière sûre.

## CARTOGRAPHIE

Notre objectif principal est de fournir des cartes sémantiques de l'environnement exploré. Dans notre approche, ces cartes sont basées sur des cartes 2D et 3D, sur lesquelles viennent s'ajouter des informations telles que les pièces, les types de sols et de murs, ainsi que les objets détectés.

La cartographie 2D est assurée par la bibliothèque commerciale Karto (de la société SRI). Ce module prend, en entrée, les données d'odométrie et les mesures du télémètre laser, et il fournit, en sortie, la carte de l'environnement sous la forme d'une grille d'occupation, ainsi que la position corrigée (voir la figure 4 de la page suivante).

Sur cette base, nous créons une cartographie 3D sous forme de nuages de points colorés obtenus à partir de la caméra de profondeur. Cette représentation est simplement créée en accumulant des nuages de points extraits par la caméra et associés aux différentes positions du robot. Cette carte permet d'avoir une représentation visuellement attrayante de l'environnement, mais qui est cependant difficile à analyser avec précision (voir la figure 5 de la page 55).

Enfin, la cartographie sémantique permet de détecter les pièces présentes dans l'environnement et d'associer à chacune d'elles la liste des objets et des types de sols et



Figure 4 : Exemple de carte 3D.

de murs détectés. Les pièces sont identifiées en cherchant tout d'abord les « portes » qui lient entre elles deux pièces, puis en analysant les espaces entourés de murs et de portes, qui correspondent alors aux pièces. La position des objets est estimée en fusionnant les différentes perceptions du même objet à l'aide d'un filtre de Kalman afin de confirmer les détections et de réduire les incertitudes de localisation.

---

## STRATÉGIE D'EXPLORATION

Notre robot devant être autonome, l'exploration de l'environnement doit répondre à deux objectifs : découvrir l'ensemble des pièces et détecter l'ensemble des objets. Le calcul d'un ensemble optimal de positions permettant de garantir une exploration exhaustive est difficile à réaliser, mais nous pouvons proposer des solutions de bonne qualité par des méthodes heuristiques. Nous avons donc utilisé une méthode d'échantillonnage stochastique pour rechercher une bonne séquence de positions. A partir de l'état courant de l'exploration, cette méthode va tirer aléatoirement de futures positions possibles et sélectionner celle qui permettra d'obtenir le plus d'informations.

Pour chaque position échantillonnée, un score intégrant plusieurs critères est donc calculé. Le premier critère mesure la surface de l'environnement potentiellement découvert par la caméra du robot depuis la position étu-

diée, avec pour objectif de chercher des objets dans toute la zone identifiée. Le second critère prend en compte les frontières entre les zones cartographiées par le télémètre laser et les zones encore inconnues afin de favoriser la découverte de nouvelles pièces. Enfin, le troisième critère prend en compte la distance à parcourir pour atteindre cette position depuis la position courante du robot afin d'en limiter les déplacements. Cette stratégie permet de réaliser une exploration rapide et exhaustive de l'environnement, en limitant notamment les allers-retours inutiles.

---

## CONCLUSION

Le système développé pour le projet Pacom permet donc d'obtenir des cartes sémantiques plus facilement interprétables par l'opérateur que des cartes construites simplement au moyen d'un télémètre laser. Ces résultats montrent que les techniques de cartographies 2D et 3D actuelles associées à des caméras de profondeur pour la détection d'objets offrent des solutions efficaces dans des environnements encore assez contrôlés.

Différents problèmes restent ouverts, tels que celui de la détection des objets dans des situations plus réalistes (par exemple, regroupés en tas ou rangés sur des étagères), situations dans lesquelles l'approche qui est actuellement la nôtre ne pourra pas les détecter. Il est également envisageable de ne plus utiliser de télémètre

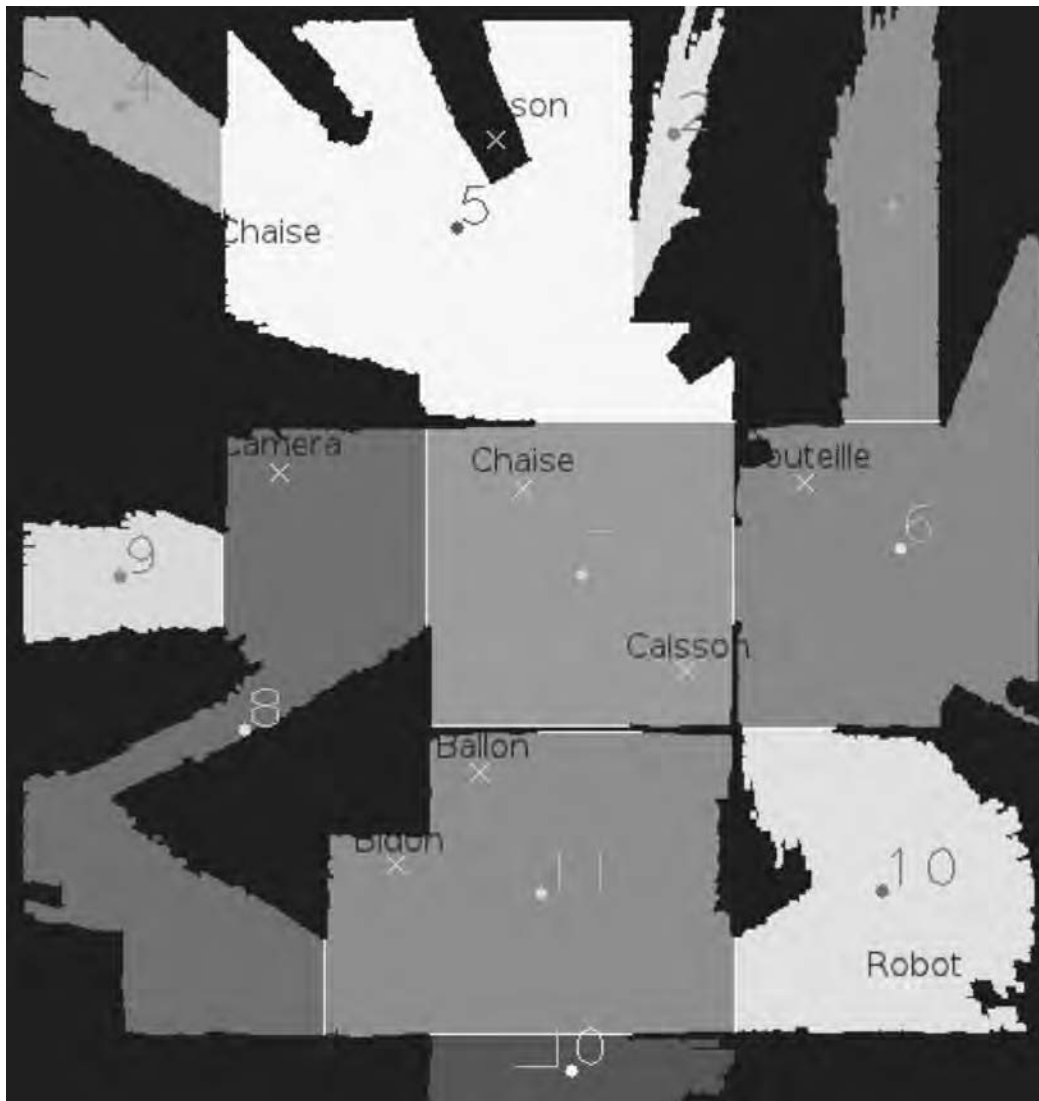


Figure 5 : Exemple de carte sémantique produite durant la compétition Carotte.

laser pour ne dépendre que d'une caméra de profondeur et de capteurs sonar complémentaires, ce qui permettrait de réduire fortement le coût de la plateforme. L'une des conclusions importantes du projet PACOM est que la réalisation d'une navigation efficace et d'une cartographie sémantique ne peut s'obtenir simplement avec un capteur et un algorithme. Concevoir un système robuste face à de nombreuses configurations courantes (comme la présence de vitres) ou capable de détecter une large gamme d'objets demande d'intégrer différents capteurs et plusieurs algorithmes complémentaires. Cela n'est pas, en soi, une surprise, la robotique étant un bon exemple de système complexe, mais cela fait reposer les performances du système beaucoup plus sur la qualité de l'intégration que sur la qualité des algorithmes utilisés pour les différentes tâches. Ainsi, même si les différents éléments du système ont été largement validés, notamment en obtenant les meilleurs résultats en matière de qualité des annotations sémantiques, la troisième place obtenue dans le cadre de la compétition

Carotte n'est pas à la hauteur des performances individuelles de chaque module. L'amélioration de ces performances est cependant moins un thème de recherche qu'une question, peut-être plus industrielle, d'optimisation du système.

Enfin, au-delà de la construction de cartes sémantiques pour la visualisation abordée dans ce projet, leur utilisation pour la réalisation de tâches complexes (telles que l'assistance aux personnes) est un domaine de recherche prometteur. Ce type de cartes permet en effet d'envisager des comportements beaucoup plus complexes pour le robot, qui sont indispensables pour qu'il puisse agir dans le contexte quotidien des humains.

---

## RÉFÉRENCE

[PACOM]  
<http://cogrob.ensta.fr/pacom/>