

Localisation d'un robot mobile par indexation d'images panoramiques

Mobile robot self-localisation by indexing panoramic images

Jose-Joel Gonzalez-Barbosa*

Simon Lacroix

LAAS-CNRS.

7, Av. du Colonel Roche

31077 Toulouse Cedex 4 France

{Jose-Joel.Gonzalez-Barbosa,Simon.Lacroix}@laas.fr

Résumé

Cet article présente une approche concernant la détermination de la position d'un robot mobile en milieu naturel non structuré, en utilisant des techniques d'indexation appliquées à des images omnidirectionnelles. Le principe est de construire dynamiquement une base d'images lors des déplacements du robot, en sélectionnant les images pertinentes dans le flot des images continûment acquises. Une fois cette base construite, le robot est capable de se localiser par rapport à elle, en mettant en correspondance une image perçue avec les images stockées. La représentation utilisée pour indexer et comparer les images est basée sur la densité de probabilité de caractéristiques locales. L'article décrit la technique d'indexation développée, et présente quelques résultats.

Mots Clefs : Localisation d'un robot mobile, caractéristiques locales, indexation par histogrammes, images omnidirectionnelles.

Abstract

This paper presents an approach to determine the position of a mobile robot in outdoor unstructured environments, using image indexing techniques on panoramic images. The principle of the approach is to dynamically build an image database as the rover moves, by selecting relevant images in the stream of acquired images. Once this database is built, the rover is able to localize itself with respect to it, by matching newly perceived images with the stored ones. Images are indexed by probability density functions of local characteristics. The article describes the indexing technique developed, and presents some results.

Keywords : Mobile robot self-localization, local characteristics, histogram indexing, omni-directional images.

1 Introduction - Contexte

La navigation autonome à long terme d'un robot dans un environnement naturel non structuré et initialement inconnu

*Le séjour de Jose-Joel Gonzalez-Barbosa au LAAS est financé par CONACyT.

nécessite l'intégration de nombreuses fonctionnalités, qui vont de l'asservissement de l'exécution des déplacements à la supervision de mission, en passant par la modélisation de l'environnement perçu et la planification de trajectoires et de stratégies de déplacements. Parmi ces différentes fonctionnalités, la localisation du robot, c'est à dire la capacité d'estimer à tout moment sa position et l'incertitude associée, est très importante. La connaissance de la position du robot est en effet indispensable à la réalisation de nombreuses fonctions : elle est d'une part nécessaire à l'exécution de missions globales, qui sont pour la plupart définies en termes de positions à rallier ou de zones à explorer ; elle est d'autre part nécessaire à la cohérence spatiale des différentes représentations de l'environnement construites ; et enfin elle permet de garantir et de corriger l'exécution de déplacements planifiés.

L'importance de la résolution du problème de la localisation en robotique est illustrée par le nombre et la variété des contributions qui l'on peut trouver dans la littérature ([2] est un ouvrage qui présente un bon aperçu des différentes approches, essentiellement en environnements intérieurs structurés). D'une manière générale, on peut classer les contributions en trois grands types de méthodes [12]:

- *Les techniques d'estimation de mouvement*, telles l'odométrie et la navigation inertielle, qui intègrent à haute fréquence des données proprioceptives, sans les structurer ni les mémoriser. Par essence, ces techniques produisent une erreur non bornée sur l'estimation de la position du robot.
- *Les techniques de recalage*, qui construisent un modèle de l'environnement à partir de données extéroceptives acquises par le robot, et qui utilisent ce modèle pour corriger la position estimée par les techniques d'estimation du mouvement.
- Et enfin les *techniques de localisation absolue*, qui mettent en correspondance des données perçues avec une modèle initial (obtenu par imagerie aérienne par

exemple), et produisent des estimées de position dont l'erreur est bornée.

C'est dans la seconde catégorie que l'on trouve le plus de contributions en robotique d'une manière générale, et aussi dans le contexte d'environnements extérieurs. Parmi ces contributions, de nombreuses portent sur la détection d'amers dans l'environnement : ces amers peuvent être des objets détectés dans l'environnement [1], ou définis de manière plus implicite, telles que des zones de modèles numériques de terrain présentant des caractéristiques (signatures) particulières [9, 17].

Lorsque l'imprécision sur l'estimée initiale de la position du robot est très grande (après exécution d'une longue trajectoire en boucle par exemple), la mise en correspondance entre les amers modélisés et les amers nouvellement perçus ne peut pas se faire par un simple raisonnement géométrique sur les positions des amers : il est alors nécessaire de les reconnaître. Ce problème est difficile en environnements naturels, car l'absence de structure des éléments que l'on y rencontre rend impossible la représentation de la géométrie des objets par des formes analytiques aisément extraites (droites et plans), et les erreurs sur les données tridimensionnelles acquises rendent inopérants les modèles basés sur des maillages déformables. Cet article est consacré à ce problème : il propose une solution, basée sur l'indexation d'images panoramiques, qui permet à un robot de déterminer qualitativement sa position lorsqu'il se retrouve une zone préalablement explorée.

Le principe de notre approche est le suivant (figure 1) :

- Dans une première étape, le robot construit dynamiquement la base d'images à indexer pendant ses déplacements dans des zones inconnues. Cette étape correspond à une phase de *modélisation* (ou d'apprentissage) : pour chaque image, à laquelle est associée la position courante déterminée par une ou plusieurs techniques d'estimation du mouvement (odométrie et odométrie optique [14] dans notre cas), des histogrammes représentant les statistiques sur des caractéristiques locales sont calculés. Seuls les histogrammes des images les plus représentatives sont mémorisés.
- Dans une seconde étape, lorsque le robot traverse après un long déplacement une zone préalablement traversée (ce qui est détecté grâce à l'estimée de sa position), l'image de la base de donnée la plus semblable à l'image courante est déterminée. Il s'agit d'une phase de *reconnaissance*, qui permet d'affiner qualitativement l'estimée de la position du robot, en lui associant l'estimée de l'image la plus semblable acquise durant la phase d'apprentissage.

Le paragraphe suivant présente un bref état de l'art des techniques d'indexation d'images, ainsi que de l'exploitation d'images panoramiques pour la navigation et la localisation d'un robot mobile. Le paragraphe 3 présente notre

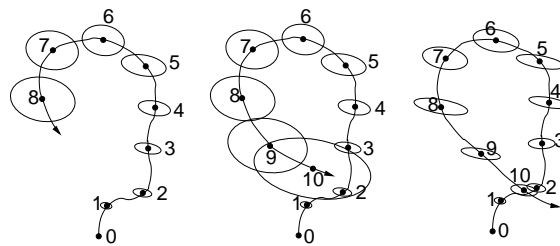


FIG. 1 – Principe de l'utilisation de la reconnaissance de lieux pour la localisation. À gauche, apprentissage : le robot effectue une trajectoire, le long de laquelle il acquiert et mémorise des images panoramiques, et leur associe la position estimée courante (dont l'imprécision est représentée par des ellipses). Au centre, reconnaissance : lorsque le robot croise à nouveau une zone préalablement traversée, l'image courante est mise en correspondance avec les images mémorisées situées dans la zone où le robot se trouve (ici, l'image 10 est mise en correspondance avec les images 1, 2 et 3). À droite, à l'issue de la phase de reconnaissance qui a ici associé l'image 10 à l'image 2, une meilleure estimée de la position courante est déterminée, et les estimées sur les positions précédentes sont affinées (cette dernière phase est hors du cadre de cet article).

technique d'indexation, basée sur l'utilisation d'histogrammes pour la représentation de statistiques sur des caractéristiques locales détectées dans les images. Des images tests d'objets non structurés sont exploitées pour déterminer la meilleure mesure de ressemblance entre des images. Dans le paragraphe 4, nous présentons comment nous avons appliqué cette technique d'indexation à des images panoramiques : une structuration des images est proposée. La manière dont est construite la base d'apprentissage à partir de l'ensemble des images acquises est présentée, et quelques résultats de reconnaissance sont décrits. Enfin, une discussion conclut l'article.

2 État de l'art

2.1 Indexation d'images

De nombreux articles récents dans la communauté de vision indiquent que l'indexation d'image est une approche prometteuse pour la reconnaissance d'objets, car elle permet d'établir des appariements robustes, sans nécessiter l'extraction de primitives géométriques dans les images. On peut distinguer trois grandes approches : les méthodes basées sur des attributs calculés globalement sur les images, les méthodes qui se basent sur une transformation de l'espace des images, et celles qui utilisent des attributs locaux.

Attributs globaux. Swain et Ballard [26] ont développé une technique d'identification d'objets fondée sur les histogrammes de couleurs. Le principe de base est de comparer un histogramme de couleur d'une région de l'image avec un histogramme d'apparences d'un objet. Leur technique est remarquablement fiable par rapport aux variations de

points de vue, même les changements de la forme de l'objet ne dégradent pas toujours la performance de leur méthode. La simplicité, la vitesse et la fiabilité de la comparaison des histogrammes de couleurs constituent les intérêts majeurs de cet algorithme de reconnaissance d'objets. Néanmoins, l'inconvénient principal de cette méthode est sa sensibilité aux variations d'éclairage. Dans [22], Schiele et Crowley ont proposé une technique de reconnaissance d'objets utilisant des histogrammes multidimensionnels de champs réceptifs. Ils proposent une formulation générale pour la représentation d'objets par une densité de probabilité d'un ensemble de descripteurs locaux d'apparences d'objets, fondés sur les dérivées gaussiennes et les filtres de Gabor. La technique permet l'identification de plusieurs objets dans des positions et orientations arbitraires, avec éventuellement des occultations partielles, et peut identifier plus de 100 objets avec des occultations importantes, avec une complexité algorithmique qui la rend utilisable en temps réel.

Changement d'espace. L'utilisation de l'extension de Karhunen-Loève (KL) pour la représentation d'images a suscité un grand intérêt: Kirby dans [11] et Turk et Pentland dans [28] représentent, détectent et identifient des visages en utilisant cette représentation. La méthode est connue sous le nom d'ACP (Analyse en Composantes Principales): avec cette méthode, chaque visage dans la base d'images peut être représenté comme combinaison linéaire d'un petit nombre de caractéristiques. On obtient l'information principale d'un ensemble de visages dans un espace où la redondance de la représentation de ces visages est minimale.

Attributs locaux. [24] est une contribution de référence pour l'utilisation d'attributs locaux pour la reconnaissance d'objets. La méthode est basée sur les invariants locaux de niveaux de gris qui sont calculés automatiquement, et permet une recherche efficace dans une base d'images de plus de 1000 images. Un algorithme de vote et une indexation multidimensionnelle rendent possible la recherche d'image. Cependant, le vote sur différents invariants est insuffisant pour garantir l'exactitude de la réponse dans l'indexation de la base d'images: une concordance semi-locale entre ces identifications est proposée, ce qui augmente l'efficacité de la reconnaissance. Les résultats expérimentaux montrent une reconnaissance correcte des objets qui peuvent apparaître dans des scènes complexes, et cela même s'ils sont partiellement visibles et s'ils sont observés de différents points de vues.

2.2 Localisation par images panoramiques

Une des contraintes de l'utilisation de caméras classiques pour la localisation des robots est leur champ de vue partiel. Les images panoramiques pallient cette contrainte, en fournissant une vue complète de la scène.

On trouve d'abord des contributions "classiques" au problème de la localisation, qui extraient des éléments remarquables des images pour la localisation d'un robot [27, 15]. Mais il est aussi possible de reconnaître des *lieux* sur la

base d'une analyse globale de telles images, sans nécessairement recourir à une phase de détection et de représentation d'amers. Ainsi, dans [10] et [30], des techniques de localisation basées uniquement sur l'apparence d'images panoramiques sont présentées. Dans ces contributions, les images sont représentées par une combinaison linéaire des vecteurs propres de la base d'apprentissage, produits par une technique d'ACP, et la position du robot est déterminée en trouvant l'image de la base la plus proche de l'image courante. Dans [18], chaque image est partitionnée en un ensemble de secteurs angulaires. La distribution des secteurs dans l'espace des vecteurs propres est représentée par un mélange de gaussiennes, et les auteurs proposent un formalisme bayésien pour identifier une image acquise, en introduisant des contraintes structurelles entre les différents secteurs. D'autres contributions mettent en correspondance des images panoramiques par simple corrélation d'anneaux de différents rayons [16].

3 Indexation par histogrammes de caractéristiques locales

Dans notre cas, la phase d'apprentissage est réalisée durant l'exécution d'une mission par le robot: elle n'est pas préalable à l'exécution d'une mission, et la base d'apprentissage est donc construite de manière dynamique. Ceci empêche toute méthode de reconnaissance par une technique d'ACP, dans laquelle les vecteurs propres sont déterminés à partir de la connaissance d'un ensemble fixe d'images. Par ailleurs, la répétabilité d'attributs locaux tels que les points d'intérêts est très affectée par la distorsion induite par notre miroir parabolique (au moins pour les points correspondant à des zones proches du robot). Nous avons donc opté pour une méthode qui représente les images par leur réponse globale à différents opérateurs locaux, comme proposé par Schiele et Crowley, afin de représenter la majeure partie de l'information des images dans une structure plus petite. Nous représentons les images par une fonction de densité de probabilité d'un ensemble de descripteurs locaux extraits des images: il s'agit donc d'une technique d'indexation basée sur des attributs globaux.

3.1 Caractéristiques locales

Les caractéristiques locales d'une image peuvent être obtenues par le filtrage de l'image avec une fonction dérivative [7, 25, 23]. La fonction dérivative que nous utilisons pour calculer les caractéristiques locales est la dérivée gaussienne. Ces dérivées gaussiennes sont souvent employées en vision par ordinateur, car elles présentent des caractéristiques intéressantes:

- Généricité: dans [20], il est montré que les images propres d'un grand nombre de régions d'images ressemblent aux dérivées gaussiennes.
- Il est possible de réaliser une implémentation récursive([29]).

- Robustesse aux changements d'échelle : [4] prouve que les dérivées gaussiennes sont robustes par rapport aux changements d'échelle d'approximativement $\pm 20\%$.
- Equivariance par rapport aux changements d'échelle. Lindeberg dans [13] a proposé une normalisation des dérivées de gaussiennes permettant d'obtenir un espace de caractéristiques invariant par changement d'échelle.

Les dérivées gaussiennes fournissent une base pour la décomposition du signal consistant en dérivations consécutives du signal d'entrée [3]. Ce sont les caractéristiques locales que nous utilisons.

Soit une image I , les caractéristiques locales sont définies par :

$$J[I(x, y)] = \{L_{i_1 \dots i_n}(x, y) \in I \times \mathbb{R}^+\}$$

où $L_{i_1 \dots i_n}$ est la convolution de l'image I avec les dérivées gaussiennes $G_{i_1 \dots i_n}$.

Les caractéristiques locales peuvent être décrites et caractérisées par la fonction logarithmique d'apparence locale :

$$Z_{n,\phi}^\sigma(x, y) = \log(1 + \|L_{n,\phi}^\sigma(x, y)\|)$$

où n est l'ordre de la dérivée considérée, σ l'écart type de la fonction gaussienne, et ϕ l'orientation de la dérivée. L'intérêt de cette mesure est qu'elle exprime la magnitude de la différence entre pixels voisins, et non le signe de cette différence. L'ajout du 1 dans cette définition sert à éliminer les variations abruptes de la fonction lorsque si $0 < \|L_{n,\phi}^\sigma\| < 1$, et à permettre de considérer les cas où $\|L_{n,\phi}^\sigma\| = 0$.

Les cinq équations suivantes présentent l'information de l'image extraite que nous utilisons à partir de caractéristiques locales du premier et deuxième ordre.

$$Z_x^\sigma(x, y) = \log(1 + \|L_x^\sigma(x, y)\|)$$

$$Z_y^\sigma(x, y) = \log(1 + \|L_y^\sigma(x, y)\|)$$

$$Z_{x,y}^\sigma(x, y) = \log(1 + \sqrt{L_x^\sigma(x, y)^2 + L_y^\sigma(x, y)^2})$$

$$Z_{xy}^\sigma(x, y) = \log(1 + \|L_{xy}^\sigma(x, y)\|)$$

$$Z_{xx,yy}^\sigma(x, y) = \log(1 + \sqrt{L_{xx}^\sigma(x, y)^2 + L_{yy}^\sigma(x, y)^2})$$

(1)

Ces 5 caractéristiques, qui ont été sélectionnées empiriquement, sont illustrées figure 2.

Nous représentons les images par une distribution statistique des caractéristiques locales $P(Z|\vec{p})$, où \vec{p} représente la position du robot. Ces distributions étant difficilement modélisées par des fonctions paramétriques, nous les représentons par des histogrammes :

$$H_{\vec{p}} = H(Z|\vec{p})$$

Au total, les images sont donc représentées par 5 histogrammes.

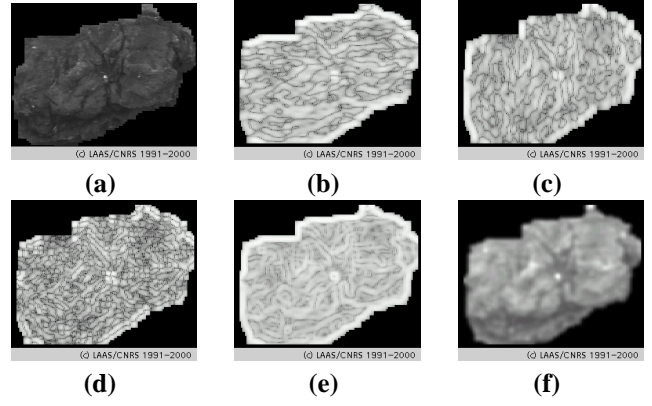


FIG. 2 – Illustration des caractéristiques locales retenues : (a) image originale, (d) Z_x , (c) Z_y , (d) $Z_{x,y}$, (e) Z_{xy} , (f) $Z_{xx,yy}$.

3.2 Mesure de similarité des histogrammes

Nous définissons la distance entre deux images I_1 et I_2 comme la moyenne de la distance entre les 5 histogrammes qui les représentent :

$$d(I_1, I_2) = \|I_1 - I_2\| \equiv \frac{1}{5} \sum_{k=1}^5 d(H_1^k, H_2^k)$$

De nombreuses définitions de la distance entre deux histogrammes ont été proposées et évaluées dans le contexte de l'indexation d'images par histogrammes :

Distance euclidienne : elle est définie par

$$d(H_1, H_2) = \sqrt{\sum_i (\underline{H}_{1,i} - \underline{H}_{2,i})^2}$$

où i représente l'indice d'une cellule d'un histogramme. En représentant un histogramme H par un vecteur monodimensionnel \underline{H} , cette distance peut s'écrire :

$$d(H_1, H_2)^2 = (\underline{H}_1 - \underline{H}_2)^t \cdot (\underline{H}_1 - \underline{H}_2)$$

La complexité du calcul de cette distance est donc en $O(n)$, n étant le nombre de cellules des histogrammes.

Intersection d'histogrammes : elle est définie par

$$d(H_1, H_2) = \sum_i \min(\underline{H}_{1,i}, \underline{H}_{2,i})$$

Cette définition de la distance a été utilisée pour comparer des images dans [26], sa complexité est en $O(n)$.

Distance quadratique : L'inconvénient des deux distances précédentes est qu'elles comparent les cellules une à une, sans tenir compte des cellules voisines. La distance quadratique, utilisée dans [6] et définie comme suit

$$d(H_1, H_2)^2 = (\underline{H}_1 - \underline{H}_2)^t A (\underline{H}_1 - \underline{H}_2) \quad (2)$$

pallie ce problème. Dans cette définition, la matrice A permet de pondérer le poids des cellules voisines en fonction de leur distance à la cellule considérée. La complexité du calcul de cette distance est en $O(n^2)$.

Distance de Mahalanobis : sa définition est la même que l'équation 2, dans laquelle la matrice A correspond à la matrice de covariance de l'ensemble d'histogrammes d'apprentissage. Le calcul de cette distance nécessite donc la connaissance de l'ensemble des images d'apprentissage, ce qui n'est pas compatible avec notre application.

Distance de Haussler : définie dans [8], elle s'écrit

$$d(H_1, H_2) = \sum_i \frac{|H_{1,i} - H_{2,i}|}{1 + H_{1,i} + H_{2,i}}$$

où les histogrammes H_1 et H_2 ne sont pas normalisés.

Statistique du χ^2 : Les test du χ^2 permet de déterminer la différence entre deux distributions statistiques :

$$d(H_1, H_2)^2 = \sum_i \frac{(H_{1,i} - H_{2,i})^2}{H_{1,i}}$$

Pour rendre symétrique cette mesure, la définition suivante a été introduite dans [22] :

$$d(H_1, H_2)^2 = \sum_i \frac{(H_{1,i} - H_{2,i})^2}{H_{1,i} + H_{2,i}}$$

"Earth Mover Distance" : cette distance est basée sur la minimisation du coût nécessaire pour transformer une distribution en une autre [21]. Le coût est défini comme un flux : il intègre la valeur des quantités f_{ij} déplacées d'une cellule i d'un histogramme à la cellule j de l'autre, multipliée par la distance d_{ij} entre les deux cellules. Une fois le flux global $\sum_i \sum_j d_{ij} f_{ij}$ minimisé, la distance est définie comme suit :

$$d(H_1, H_2) = \frac{\sum_i \sum_j d_{ij} f_{ij}}{\sum_i \sum_j f_{ij}}$$

La minimisation du flux global est un algorithme itératif, dont le temps de convergence croît exponentiellement avec le nombre de cellules des histogrammes en général. En pratique, ce temps peut être très réduit, suivant la "différence" entre les histogrammes considérés.

Comparaisons Afin d'évaluer ces différentes définitions de la distance entre histogrammes, nous avons construit une base d'apprentissage de près de 600 images sur 8 objets différents (figure 3). Pour ce faire, nous avons fait tourner le robot à une distance à peu près constante (de l'ordre de 1.5m) autour des objets posés sur le sol, en acquérant de manière continue un ensemble de couples d'images stéréoscopiques des objets. Les régions des images correspondant effectivement aux objets ont été déterminées par un simple seuillage sur l'élévation des points 3D reconstruits par stéréovision. À chaque image d'un objet est associée une orientation sous lequel le robot l'a perçu.

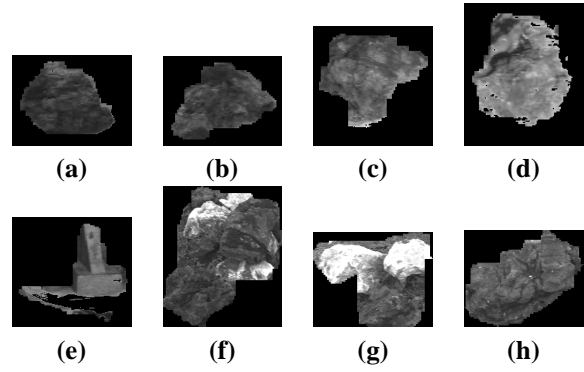


FIG. 3 – Images des 8 objets utilisés pour évaluer les distances entre histogrammes.

Un ensemble de 39 images tests a aussi été acquis, et chacune d'entre elle a été comparée avec toutes les images de la base d'apprentissage. Le tableau 1 présente le pourcentage d'identification obtenu pour les différentes fonctions de distance entre histogrammes. Au vu de ces comparaisons, nous avons opté pour la distance définie par le test du χ^2 . Avec des histogrammes de 100 cellules, son calcul est exécuté en 0.2 ms sur une station de travail Sun UltraSparc 5.

Distance	Complexité	Taux de succès (%)
Intersection	$O(n)$	46.1
Euclidienne	$O(n)$	35.9
Quadratique	$O(n^2)$	41.0
Mahalanobis	$O(n^2)$	53.8
Haussler	$O(n)$	43.6
statistiques χ^2	$O(n)$	69.2
EMD	-	66.6

TAB. 1 – Comparaison du taux de réussite de reconnaissance d'un objet avec les différentes distances entre histogrammes.

4 Indexation d'images panoramiques

Pour évaluer l'efficacité la reconnaissance de lieux avec des images panoramiques, nous avons effectué des déplacements avec le robot Lama (figure 4) sur le terrain d'expérimentation du laboratoire. Les images panoramiques ont été continuellement acquises et sauvegardées, leur position étant mesurée grâce à un récepteur GPS à différence de phase (précision centimétrique). La figure 5 présente les positions successives du robot lors d'une trajectoire consistant en 3 boucles. La résolution des images est de 768×576 pixels, et la distance moyenne entre deux acquisitions successives est d'environ 0.4m (les déplacements entre les images 1 à 35 et 320 à 394 sont cependant plus long).



FIG. 4 – Le robot Lama. La caméra panoramique n’est pas visible sur cette photographie - elle était montée en haut du mat central, près de l’antenne GPS blanche.

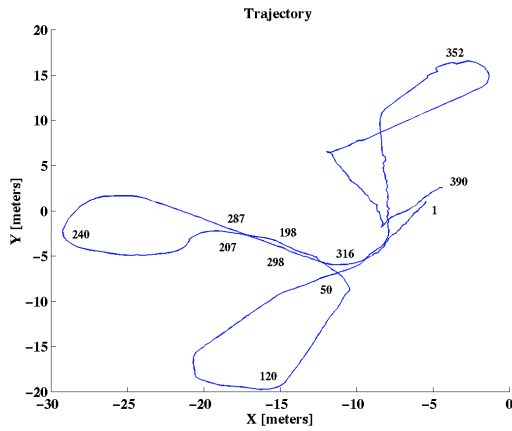


FIG. 5 – La trajectoire durant laquelle près de 400 images panoramiques ont été acquises.

4.1 Structuration des images

Afin de déterminer la position et l’orientation d’un robot à partir d’images panoramiques, plusieurs auteurs ont proposé de discrétiser les images en régions. Ainsi, dans [5], les images sont partitionnées en régions carrées. D’autres contributions redressent les images par une transformation cylindrique et les partitionnent en colonnes [19]. Dans ces deux cas, les partitions choisies rendent les algorithmes sensibles à l’orientation du robot.

Nous proposons plutôt de discrétiser les images en m anneaux (figure 6), de manière à ce que les histogrammes définis par chaque anneau ne dépendent pas de l’orientation du robot. L’avantage de cette partition est de prendre en compte le fait que pour les petits anneaux, qui correspondent à des zones proches du robot, l’apparence change beaucoup avec de petits déplacements, tandis qu’elle change bien moins pour les plus grands anneaux. La figure 7 montre la comparaison entre la distance calculée entre une image de la base et toutes les autres, en considérant l’image en-

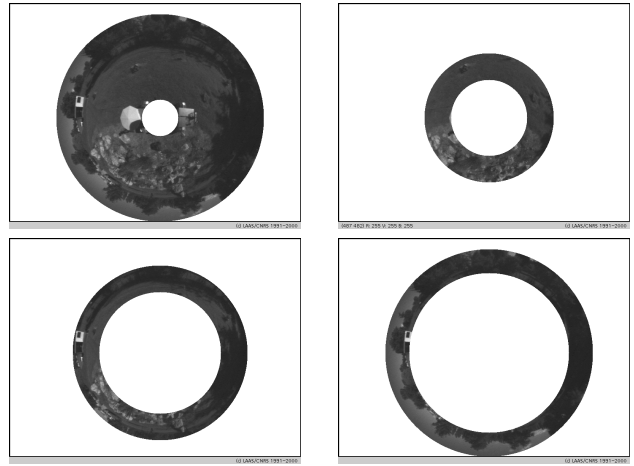


FIG. 6 – Une image panoramique, et les trois anneaux correspondants. Le quatrième anneau (le plus petit anneau) n’est pas représenté ici : il n’est en effet pas considéré, car il contient toujours les roues du robot et l’antenne GPS.

tière d’une part, et la moyenne des distances calculées pour chacun des anneaux d’autre part. La mesure est plus discriminante en partitionnant les images en anneaux.

4.2 Construction de la base de données

Durant la phase de découverte de l’environnement, le robot acquiert en continu des images, et calcule les histogrammes de caractéristiques locales correspondants. Afin de réduire la place mémoire nécessaire à leur stockage et les ultérieures reconnaissances, il est intéressant d’éliminer de la base les images semblables.

La figure 8 montre deux exemples de l’évolution de la distance entre les histogrammes d’une image et ceux des images suivantes durant les déplacements. Notons que de manière attendue, cette distance croît plus vite pour les petits anneaux que pour les grands. Sur la base de ce comportement, la sélection des histogrammes à conserver (les “histogrammes clefs”) est basée sur l’introduction d’une valeur seuil ξ sur la distance entre les anneaux. La procédure de construction de la base de donnée est la suivante : les histogrammes de tous les anneaux de l’image de départ sont sélectionnés comme histogrammes clefs. Durant le déplacement du robot, les histogrammes des différents anneaux des images continuellement acquises sont déterminés, et leur distance avec le dernier histogramme clef de l’anneau correspondant est calculée : quand cette distance excède le seuil ξ , l’histogramme courant est sélectionné comme un nouvel histogramme clef. Avec cette procédure, les 394 images acquises lors de la trajectoire de la figure 5 génèrent 99 histogrammes clefs pour le premier anneau, 84 pour le second et 74 pour le troisième. La base de donnée est donc réduite à près d’un quart, les anneaux plus petits générant naturellement plus d’histogrammes clefs que les plus grands : leur aspect change plus rapidement avec le déplacement du

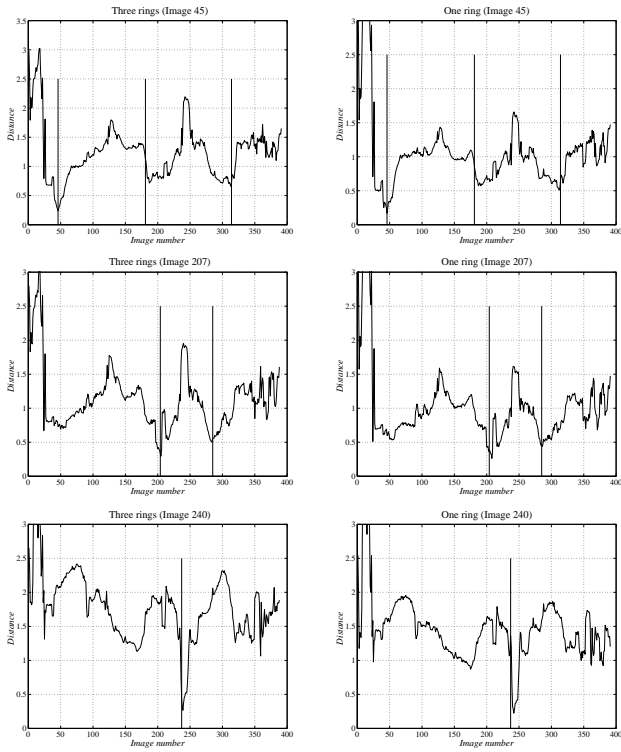


FIG. 7 – Comparaison, pour les images 45, 207 et 240, entre les distances calculées entre une image et les autres images de la base d'apprentissage en partitionnant les images en trois anneaux (à gauche), et en ne les partitionnant pas (à droite). Les lignes verticales indiquent les minima locaux bien marqués : on retrouve bien par exemple que le robot est repassé par deux fois près de l'image 45 - voir figure 5.

robot.

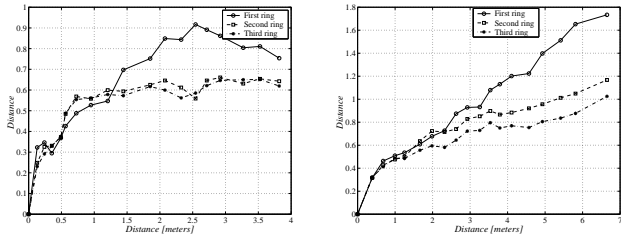


FIG. 8 – Deux exemples de l'évolution de la distance entre les histogrammes représentant une image et ceux des images suivantes, en fonction de la distance euclidienne séparant les lieux d'acquisition.

4.3 Reconnaissance

La phase de reconnaissance consiste simplement à déterminer parmi les images clefs de la base, celle qui est la plus proche de l'image courante (les images candidates sont déterminées grâce à l'estimée courante de la position du ro-

bot). La figure 9 montre la distance calculée entre l'image 287 et toutes les images retenues comme clefs de la base d'apprentissage (dans un cas réaliste, cette image n'aurait été comparée qu'avec les images autour de l'indice 200, voisinage dans lequel le robot est supposé se trouver).

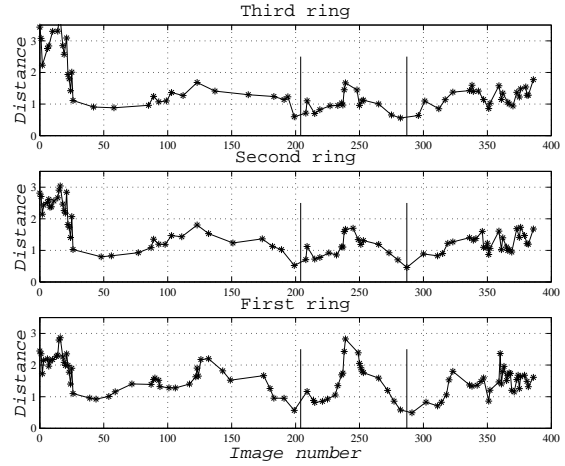


FIG. 9 – Distance entre les trois anneaux de l'image 287 et l'ensemble des images clefs retenues pour chaque anneau. Le temps de calcul pour établir ces courbes est de l'ordre d'une demi seconde.

A cet instant, les anneaux clefs de la base d'apprentissage les plus proches des anneaux de l'image courante sont connus. Mais la réduction de la base des images induit une discrétisation des positions plus grossière que la totalité des images acquises, et par ailleurs les anneaux clefs les plus proches ne correspondent pas forcément tous à la même position. Il est néanmoins possible de déterminer plus précisément à quelle image acquise lors de l'apprentissage correspond l'image courante, même si cette les anneaux de cette image n'ont pas été sélectionnés comme images clefs. Soient \vec{p} la position d'un anneau qui n'est pas sélectionné, k_+ and k_- les positions des histogrammes clefs qui l'entourent, et $d(\vec{p}, k_-)$ et $d(\vec{p}, k_+)$ les distances entre histogrammes correspondant. Alors, à la position courante n , on a (voir figure 10) :

$$d(\vec{n}, \vec{p}) \geq \max(d(\vec{n}, k_-) - d(\vec{p}, k_-), d(\vec{n}, k_+) - d(\vec{p}, k_+))$$

et

$$d(\vec{n}, \vec{p}) \leq \min(d(\vec{n}, k_+) + d(\vec{p}, k_-), d(\vec{n}, k_-) + d(\vec{p}, k_+))$$

La distance $d(\vec{n}, \vec{p})$ est alors simplement déterminée comme la moyenne des deux valeurs qui l'encadrent. Cette approximation permet d'associer à l'image courante une image qui n'a pas été sélectionnée comme image clef, elle ne nécessite de mémoriser lors de l'apprentissage que les distances aux anneaux clefs des anneaux non sélectionnés. La figure 11 montre le résultat des distances calculées entre une image et toutes les images de la base d'apprentissage,

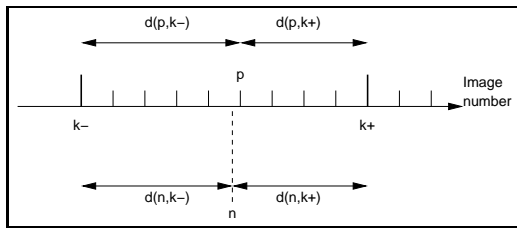


FIG. 10 – Principe de la détermination de l'image la plus proche à partir des images clefs.

en comparaison avec les distances calculées pour la même image avec les images clefs et l'approximation pour les images non sélectionnées.

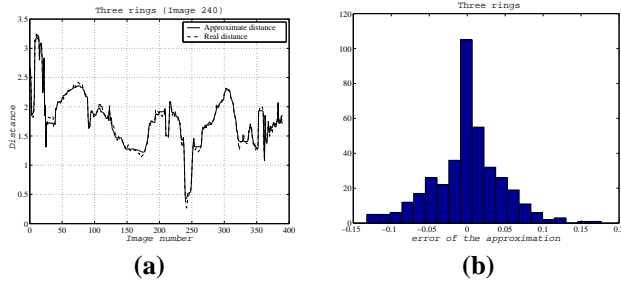


FIG. 11 – (a): Distances réelles et approximées à partir des images clefs, entre l'image 240 et la base d'apprentissage : les deux courbes sont difficilement distinguables. (b): Histogramme des erreurs de l'approximation.

4.4 Vers une estimée de la position courante

Les courbes de la figure 9 laissent supposer qu'il existe une relation qui lie la distance entre les histogrammes représentant une image et la distance euclidienne qui sépare les acquisitions. Pour s'en assurer, nous les avons calculées pour l'ensemble des 400 images de la base d'apprentissage. Les résultats sont présentés figure 12, ils montrent qu'en moyenne, il existe une corrélation entre ces deux distances.

Par ailleurs, la détermination de l'orientation du robot, une fois l'image de la base la plus proche déterminée, ne semble pas poser de problème: des essais de corrélation directe entre les pixels des anneaux, de manière analogue à [16] sont prometteurs. Bien entendu, il n'est pas raisonnable de stocker la totalité des images pour effectuer cette corrélation: nous étudions la possibilité de ne conserver que des anneaux très fins, de représenter les images par une image de résolution très grossière, voire d'exploiter le profil de la ligne d'horizon.

5 Conclusions

Dans cet article, nous avons présenté l'état de nos travaux sur le problème de la localisation d'un robot mobile en milieu naturel non structuré, en utilisant une technique d'indexation d'images panoramiques. Nous avons présenté une

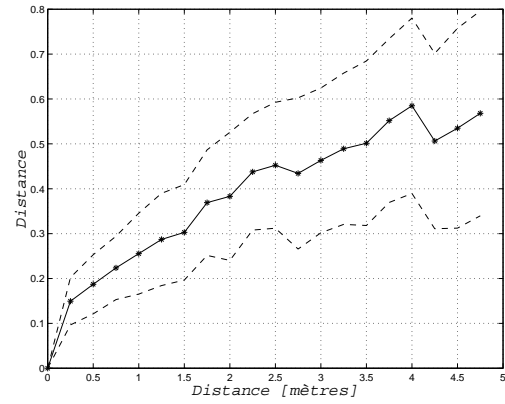


FIG. 12 – Évolution moyenne de la distance entre les histogrammes des images en fonction de la distance euclidienne séparant l'acquisition. Les courbes pointillées montrent les écarts types.

stratégie pour construire dynamiquement une base d'images lors des déplacements du robot, en structurant les images en anneaux et en sélectionnant les images pertinentes dans le flot des images continûment acquises. Une fois cette base construite, le robot est capable de se localiser par rapport à elle, en mettant en correspondance les anneaux d'une image perçue avec les images stockées grâce à une technique de comparaisons d'histogrammes de dérivées gaussiennes. Des premiers résultats encourageants ont été présentés.

Nos travaux futurs vont porter sur l'amélioration de la méthode. En particulier, une analyse discriminante va permettre de sélectionner plus rigoureusement les caractéristiques locales qui permettent une bonne identification. De manière analogue, la part relative des différents anneaux dans le calcul de la distance entre deux images mérite d'être étudiée. Enfin, les travaux présentés dans le paragraphe 4.4, qui visent à obtenir une détermination quantitative de la position du robot, sont encore préliminaires. Nos efforts vont se porter dans cette direction.

Références

- [1] S. Betge-Brezetz, R. Chatila, and M.Devy. Object-based modelling and localization in natural environments. In *IEEE International Conference on Robotics and Automation, Nagoya (Japan)*, pages 2920–2927, May 1995.
- [2] J. Borenstein, H.R. Everett, and L. Feng. *Navigating Mobile Robots: Sensors and Techniques*. A. K. Peters, Ltd., 1996.
- [3] Olivier Chomat, Vincent Colin de Verdière, Daniela Hall, and James L. Crowley. Local Scale Selection for Gaussian Based Description Techniques. *European Conference on Computer Vision*, pages 118–133, 2000.
- [4] Y. Dufournaud, C. Schmid, and R. Horaud. Appariement d'images à des échelles différentes. In *12ème Congrès AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Paris (France)*, Feb. 2000.
- [5] J. Gaspar, E. Grossmann, and J. Santos-Victor. Information sampling for optimal image data selection. *9th Interna-*

- tional Symposium on Intelligent Robotic Systems, Toulouse (France)*, 2001.
- [6] J. Hafner, H. Sawhney, W. Equitz and M. Flickner, and W. Niblack. Efficient Color Histogram Indexing for Quadratic Distance Functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(7):729–736, 1995.
- [7] C. Harris and M.J. Stephens. A Combined Corner and Edge Detector. In *4th Alvey Vision Conference. Manchester University*, pages 147–152, 1988.
- [8] David Haussler. Overview of the Probably Approximately Correct (PAC) Learning Framework. *Information and Computation*, 100(1):78–150, 1992.
- [9] D. Huber and M. Hebert. A new approach to 3-d terrain mapping. In *IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS '99)*, pages 1121–1127, 1999.
- [10] M. Jogan and A. Leonardis. Robust localization using eigenspace of spinning-images. *5th Int. Conf. on Pattern Recognition*, 2000.
- [11] M. Kirby and L. Sirovich. Application of Karhunen-Loève Procedure for the Characterization of Human Faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.
- [12] S. Lacroix and A. Mallet. Integration of concurrent localization algorithms for a planetary rover. In *6th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, June 2001.
- [13] Tony Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publisher, 1994.
- [14] A. Mallet, S. Lacroix, and L. Gallo. Position estimation in outdoor environments using pixel tracking and stereovision. In *IEEE International Conference on Robotics and Automation, San Francisco, Ca (USA)*, pages 3519–3524, April 2000.
- [15] B. Marhic, M. Mouaddib, D. Fofi, and E. Brassart. Localisation absolue par le capteur omnidirectionnel cyclop. *Traitement du Signal*, 17(3):195–206, 2000.
- [16] Toshiro Matsui, Hideki Asoh, and Simon Thompson. Mobile robot localization using circular correlations of panoramic images. *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2000.
- [17] C. Olson. Landmark selection for terrain matching. In *IEEE International Conference on Robotics and Automation, San Francisco, Ca (USA)*. JPL, April 2000.
- [18] Lucas Paletta, Simone Frintrop, and Joachim Hertzberg. Robust localization using context in omnidirectional imaging. *International Conference on Robotics and Automation*, 2001.
- [19] Lucas Paletta, Simone Frintrop, and Joachim Hertzberg. Robust Localization Using Context in Omnidirectional Imaging. *International Conference on Robotics and Automation*, pages 2072–2077, 2001.
- [20] Rajesh P.N. Rao and Dana Ballard. Object Indexing using an Iconic Sparse Distributed Memory. In *International Conference on Computer Vision*, Jan 1995.
- [21] Y. Rubner, C. Tomasi, and L. J. Guibas. A Metric for Distributions with Applications to Image Databases. *Proceedings of the 1998 IEEE International Conference on Computer Vision*, pages 56–66, 1998.
- [22] Bernt Schiele and James L. Crowley. Recognition without Correspondance using Multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, 2000.
- [23] Cordelia Schmid. A Structured Probabilistic Model for Recognition. In *Computer Vision and Pattern Recognition*, 1999.
- [24] Cordelia Schmid and Roger Mohr. Local Greyvalue Invariants for Image Retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.
- [25] Cordelia Schmid, Roger Mohr, and Christian Bauckhage. Comparing and Evaluating Interest Points. In *International Conference on Computer Vision*, Jan 1998.
- [26] Michel J. Swain and Dana H. Ballard. Color Indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
- [27] Simon Thompson, Toshihiro Matsi, and Alexander Zelinsky. Localisation using automatically selected landmarks from panoramic images. *Proceedings of Australian Conference on Robotics and Automation, Melbourne Australia*, 2000.
- [28] M. A. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.
- [29] Lucas J. van Vliet, Ian T. Young, and Piet W. Verbeek. Recursively Gaussian Derivative Filters. *Signal Processing*, 44(2):139–151, 1995.
- [30] Niall Winters, José Gaspar, Gerard Lacey, and José Santos Victo. Onmi-directional vision for robot navigation. *Proceedings of the 1st International IEEE Workshop on Omnidirectional Vision*, 2000.