

Navigation visuelle dans un environnement ouvert: reconnaissance de vues panoramiques

P. Gaussier, C. Joulain, S. Zrehen(*), J.P. Banquet (**), A. Revel

ENSEA ETIS, 6 Av du Ponceau, 95014 Cergy Pontoise Cedex, France

(*) USC USA, (**) CREARE Université Paris VI

e-mail: gaussier or joulain@ensea.fr

RÉSUMÉ

Nous présentons un système de navigation pour robot autonome dans un environnement ouvert. Le robot rejoint un objectif en associant des mouvements aux informations visuelles provenant de l'environnement. Il utilise un apprentissage simple et en ligne. Il ne crée aucune carte complexe de son environnement. Le mécanisme s'avère efficace et robuste, de plus il semble en accord avec les observations animales. Enfin, notre implémentation dans un environnement réel supporte des perturbations importantes.

ABSTRACT

In this paper we describe how a mobile robot controlled only by visual information can retrieve a particular goal location in an open environment. Our model does not need a precise map nor to learn all the possible positions in the environment. A simple and fast on line learning of a few places located near the goal allows the robot to reach the goal from anywhere in its neighborhood. The system uses only egocentric representation of the robot environment and presents very high generalization capabilities. We describe an efficient implementation tested on our robot in two real indoor environments.

1 Introduction

La plupart des robots mobiles actuels utilisent principalement pour naviguer des informations provenant de leur odométrie ou de systèmes de mesures actifs (télémètre ultra-son ou laser). Ces techniques donnent de bons résultats pour des tâches d'évitement d'obstacles mais semblent relativement inadaptées pour reconnaître précisément la position du robot dans son environnement (les systèmes ultra-son étant trop imprécis et les télémètres laser trop chers et trop lents pour fabriquer des cartes de profondeur de bonne qualité). Les caméras CCD, grâce aux renseignements qu'elles nous fournissent sur le contenu d'un champ de vision relativement étendu, semblent donc être des candidats idéaux pour ce type de problèmes. Jusqu'à présent, le rôle de la vision s'est surtout limité au recalibrage de la position des robots dans des environnements intérieurs bien structurés [1]. Dans le cas d'un environnement ouvert, lorsque, par exemple, le robot est au milieu d'une pièce ce type de technique ne peut plus s'appliquer. En effet, une position particulière ne peut que très rarement se définir grâce à la reconnaissance d'une seule image : l'imprécision sur le positionnement étant alors trop grande. Les techniques actuelles font donc intervenir la vision stéréoscopique pour récupérer une information de profondeur pertinente [4] (lorsque la distance n'est pas trop grande) et/ou fusionnent les informations visuelles avec des informations provenant de télémètres lasers (voir [3] pour le cas d'un environnement extérieur). Malheureusement, la complexité de ces méthodes, leur coût et la puissance de calcul nécessaire les rendent, pour l'instant, incompatibles avec des applications de grande série.

C'est dans ce contexte que nous nous sommes intéressés à

comprendre les mécanismes de navigation utilisés par les animaux [5, 7, 8]¹. En effet, de la guêpe[2] au primate en passant par le rat, toutes ces espèces sont capables de retourner à un endroit appris en utilisant seulement des informations visuelles monoculaires (bien que d'autres systèmes leur fournissent des informations complémentaires très précieuses (accélération, stéréovision...). Il a notamment été montré que le rat possède, dans une structure de son cerveau appelée l'hippocampe, des neurones s'activant uniquement lorsque l'animal se trouve dans un endroit particulier de son environnement et ce, quelque soit son orientation. Ces neurones sont appelés "cellules de lieux". Dans cet article, après un bref résumé des caractéristiques principales de notre modèle [10], nous montrons qu'il permet à un robot d'apprendre en ligne et en un seul essai à retourner à un endroit donné sans utiliser la moindre carte de l'environnement. La précision est de l'ordre du centimètre même en partant d'une position qui n'a jamais été explorée. Nous montrons enfin que notre algorithme supporte la présence d'objets imprévus tels qu'une personne traversant le champs de vision du robot.

2 Utiliser des amers pour naviguer

Afin de pouvoir reconnaître un lieu, un système utilisant uniquement la vision doit être capable d'isoler des amers (repères visuels), de mesurer leur angle par rapport à un référentiel relatif ou absolu (i.e. le Nord), pour enfin calculer la ressemblance entre la situation testée et la situation apprise.

¹Ce travail est soutenu par un projet du GIS sciences cognitives CNA43 et un projet DRET

Dans notre modèle, la partie correspondant à l'hippocampe fusionne les informations représentant le "quoi" (groupe Amer Fig. 2) avec les informations relatives à l'angle sous lequel est vu l'amer : le "où" (groupe Azimut). Cette recombinaison (groupe Amers×Azimuts) peut alors être apprise par des neurones appartenant à un groupe du type Winner Take All. Ces neurones simulés réagissent alors exactement comme les "cellules de lieux" enregistrées dans l'hippocampe des rats.

Durant le fonctionnement du robot, lorsque celui-ci trouve son but (ou qu'on le met dessus), il déclenche une procédure qui consiste à apprendre, dans les cellules de lieux, un petit nombre de positions dans le voisinage immédiat du but. Ces cellules sont associées avec la direction du mouvement à effectuer pour rejoindre le but (apprentissage hebbien - voir fig. 1). Le but est supposé toujours visible lors de cet apprentissage ou, tout du moins, le robot est supposé capable de maintenir en interne une information relative à la direction du but. La procédure détaillée permettant au robot d'apprendre de façon non supervisée est décrite dans [10].

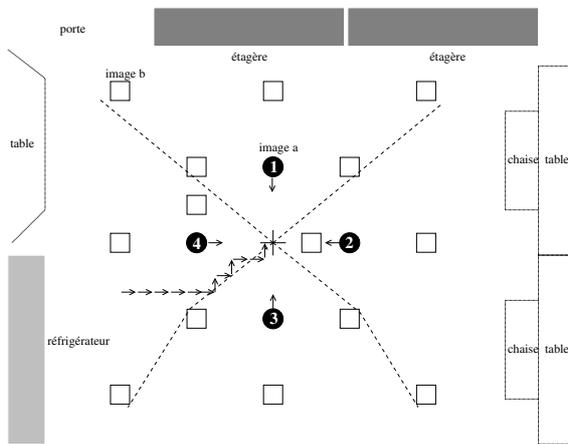


FIG. 1 — Quatres vues (cercles) sont apprises (i.e. associées à un mouvement élémentaire), les autres sont mises en correspondance avec une des vues apprises en fonction de leur similarité (rectangles). Si le robot apprend à rejoindre le but depuis chaque vue apprise, il est en mesure de le rejoindre depuis n'importe où. La suite de flèches représente un chemin possible. La zone représente une surface de $1,2m \times 1,2m$, les vues apprises sont à 30 cm du centre. L'échelle n'est pas respectée pour les positions des différents meubles (la distance au centre est de 1,5 m).

Le principe de l'architecture est décrit figure 2 et dans [6]. Lorsque le robot veut rejoindre son but, il tient compte des informations provenant des "cellules de lieux" et effectue le mouvement proposé par la cellule la plus active. Petit à petit, la distance à l'objectif est réduite (exemple Fig. 1,6). L'algorithme se contente de créer autour du but un bassin d'attraction qui entraîne systématiquement le robot en son centre. Le robot n'a pas besoin de connaître réellement sa position ni de comprendre ce que représentent les amers (pas besoin de comprendre les objets pour les utiliser). D'ailleurs, il ne lui est pas non plus nécessaire d'être sûr que les amers utilisés sont des objets. Si moins de la moitié des amers sont absents ou déplacés, le système est toujours capable de généraliser, il est capable de proposer un mouvement qui le rapproche du but. Néanmoins, pour que l'algorithme puisse fonctionner correctement, le robot doit apprendre un minimum de 3 lieux et il faut s'assurer que la mesure de similarité

entre les lieux est bien croissante et monotone lorsque le robot se rapproche du but ! C'est effectivement le cas lorsqu'une direction absolue est disponible pour mesurer la somme des erreurs sur les angles. Dans le cas où l'un des amers est utilisé comme référence, nous avons prouvé que l'algorithme peut toujours s'appliquer à condition de rester dans la zone circonscrite par les amers (la même chose semble être valable pour les rats). Une fois que l'objectif est en vue et est reconnaissable un autre mécanisme, un mécanisme permettant d'atteindre un objet dans une image, peut se mettre en place [11].

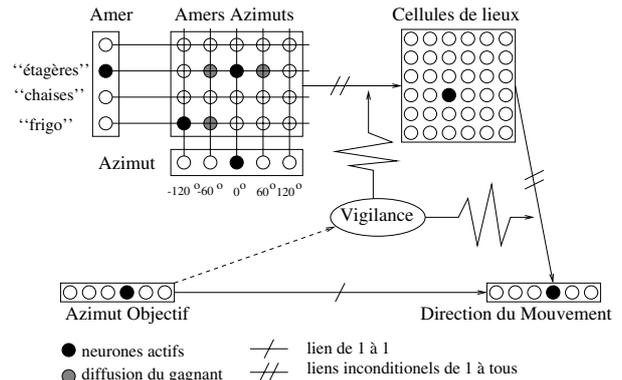


FIG. 2 — Réseau de neurone pour la navigation. Le groupe Amers-Azimuts fusionne les informations "où" et "quoi".

3 Résultats expérimentaux

Pour avoir un champ de vision assez large le robot utilise des images panoramiques, celles-ci ont été fabriquées en réunissant les parties centrales de 24 images obtenues après rotation de la caméra CCD du robot (résolution caméra de 384×288 en 256 niveaux de gris). Chaque image panoramique a une taille de (1066×288) et représente un angle de vue de 270° - Fig. 3,7). Pour trouver des amers potentiels dans l'image, le système fait une moyenne pondérée colonne par colonne de l'image panoramique pour obtenir une ligne de 1066 points (le poids des points situés sur les lignes centrales étant plus important). Ce signal est ensuite dérivé pour permettre de centrer la prise de vues locales (zone de 148×288 points assimilée à un amer) sur les zones ayant un fort contraste horizontal (Fig. 3). C'est un mécanisme de focalisation de l'attention simple et rapide, nous travaillons aussi sur des mécanismes plus complexes[12]. Chaque vue locale est réduite au format 32×32 en effectuant une réduction linéaire sur l'axe y et logarithmique sur l'axe x (centrée sur le point de focalisation, ici point de fort contraste horizontal).

L'algorithme a été testé dans plusieurs pièces différentes (voir plans Fig. 1, 5 et 8). La figure 4 montre pour chaque endroit de la Fig. 1 la "cellule de lieu" la plus active et le mouvement associé. La partition du plan en fonction des cellules gagnantes est comme on pouvait l'espérer centrée sur le but. Le robot est donc capable de rejoindre l'objectif quel que soit son point de départ dans la zone (voir Fig. 1).

Les résultats montrent que la reconnaissance décroît progressivement en fonction de la distance à l'objectif (tests effec-

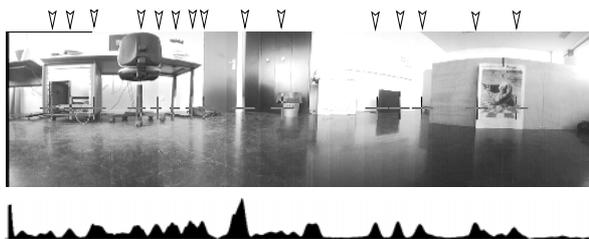


FIG. 3 — 15 vues locales apprises pour une image (environnement Fig. 5, cercle noir S), et la valeur absolue de la dérivée du signal utilisé pour la focalisation.

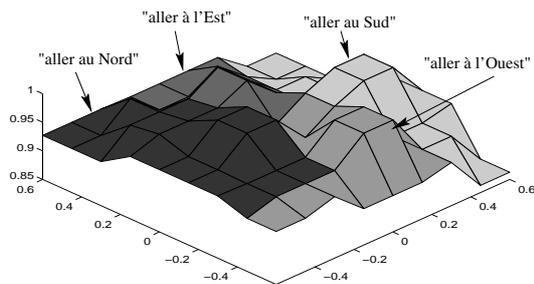


FIG. 4 — Niveau de réponse des "cellules de lieux" dans la zone expérimentale Fig. 1. Chaque niveau de gris est associé à une cellule. Les axes x et y sont en mètres

tués sur la ligne D Fig. 5). La précision du robot pour atteindre l'objectif est très bonne. Pour un déplacement élémentaire de 4,5 cm dirigé vers un des points cardinaux, le robot passe à moins de 2 cm de l'objectif (Fig. 6). Ce résultat est à comparer avec la taille du robot (20 × 20 cm) et la précision maximale théorique du positionnement : 7 mm (calculée à partir de la résolution de l'image panoramique et de la distance moyenne des amers).

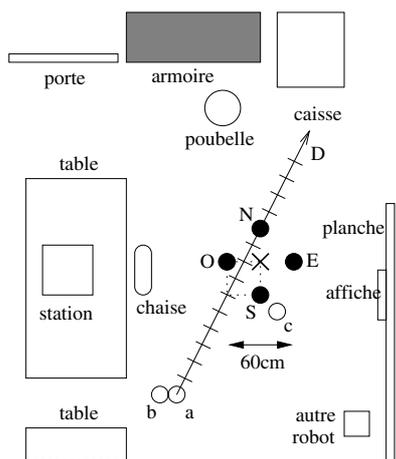


FIG. 5 — Une pièce pour les tests, les cercles blancs et la ligne D représentent plusieurs positions testées. Quatre vues sont apprises (cercles noirs). Le rectangle en pointillé est décrit en détails en fig.6, la croix représente l'objectif.

Nous avons aussi pu montrer que le système est robuste à des changements d'éclairage, des déplacements d'objets... Par exemple, lorsqu'on introduit des éléments distrayeurs, le robot utilise d'autres amers pour se repérer (Fig. 7). Il continue ainsi à effectuer un mouvement le rapprochant de l'objectif.

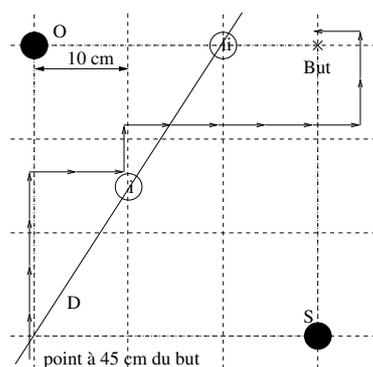


FIG. 6 — Exemple du chemin réellement pris par le robot pour atteindre l'objectif. Zoom du rectangle en pointillé de la fig.5.

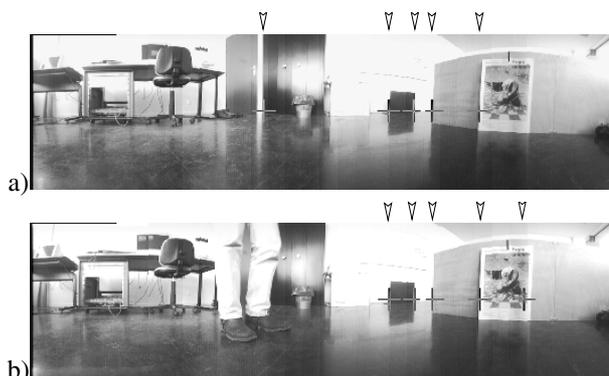


FIG. 7 — Changement du choix des amers lorsqu'un élément distracteur est introduit (environnement Fig. 5, position 'c').

D'autres expériences ont été menées dans le cas d'un environnement plus complexe pour lequel le robot doit apprendre à contourner des obstacles (fig. 8 et 9). En plus de lui apprendre des positions autour de l'objectif lui permettant de se diriger vers ce dernier nous lui avons appris des positions et leurs mouvements associés pour contourner des obstacles (flèches 2 et 6 fig.8 et 9). Dans cette expérience, le mécanisme pour sélectionner le mouvement tiens compte de l'activité de chaque neurone. Ainsi nous obtenons des trajectoires plus lissées. Nous pouvons voir sur la figure 8 qu'en ayant appris à associer 10 vues de l'environnement à 10 mouvements, le robot est en mesure de rejoindre l'objectif depuis n'importe quel point dans une surface d'à peu près 20m². La figure 8 montre différents chemins empruntés par le robot en fonction de sa position et de son orientation de départ.

4 Conclusion

Notre algorithme est efficace même dans des situations délicates, la disparition d'amers ou l'introduction d'éléments perturbateurs[6]. Il est capable d'apprendre à reconnaître des scènes relativement complexes et de naviguer correctement avec une architecture neuronale et un mécanisme d'apprentissage non supervisé très simple. Contrairement à la majorité des systèmes de navigation visuelle, ce système ne nécessite pas la présence d'un ingénieur pour identifier les vues à apprendre. De plus, lorsque le robot sort d'une zone ouverte pour entrer dans une autre, il peut se tromper de mouvement, mais alors

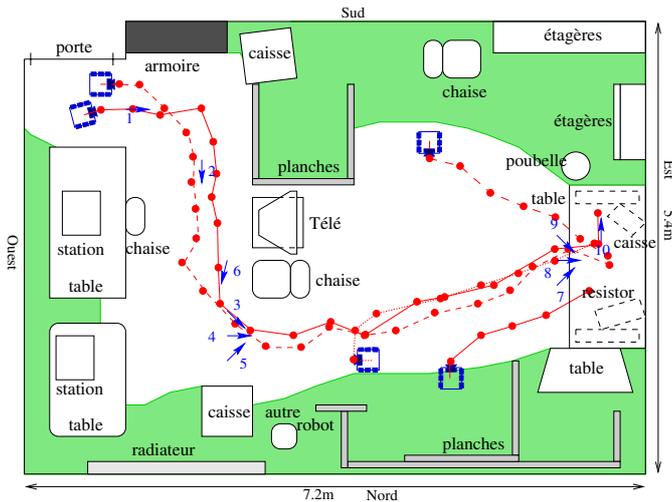


FIG. 8 — Les flèches numérotées représentent les positions apprises. Le but est sous la table (près de 10). Le robot ne peut rejoindre le but depuis les zones grisées. Le début d'une trajectoire et son orientation initiale sont représentés par les robots, les points reliés composent le reste de la trajectoire.

la somme des activités de ses "cellules de lieux" diminue très fortement. Nous prévoyons d'utiliser cette information pour permettre au robot d'apprendre par lui-même un nouveau lieu (sous-but) lorsque sa perception de l'environnement change trop (détection de nouveauté). Le robot pourrait ensuite utiliser ces sous-buts pour planifier ses déplacements [9].

Références

- [1] R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, R.A. 2(1) :14–23, March 1986.
- [2] B.A. Cartwright and T.S. Collett. Landmark learning in bees. *Journal Comp. Physiology*, 151 :521–543, 1983.
- [3] R. Chatila. Deliberation and reactivity in autonomous mobile robots. *Robotics and Autonomous System*, 16(2-4) :197–211, Dec 1995.
- [4] J.L. Crowley, P. Bobet, S. Mely, M. Kurek, and K. Sara-chik. Mobile robot perception using vertical line stereo. *Vision Based Vehicle Guidance*, June 1992.
- [5] P. Gaussier, C. Joulain, A. Revel, and J.P. Banquet. Are shaping techniques the correct answer for the control of visually guided autonomous robots? In *UKACC International Conference on Control'96*, pages 1248–1253, University of Exeter, September 1996. IEE.
- [6] P. Gaussier, C. Joulain, S. Zrehen, J.P. Banquet, and A. Revel. Visual navigation in an open environment without map. In *To appear in International Conference on Intelligent Robots and Systems - IROS'97*, Grenoble, France, September 1997. IEEE/RSJ.
- [7] P. Gaussier, A. Revel, C. Joulain, and S. Zrehen. Living in a partially structured environment : How to bypass the limitation of classical reinforcement techniques. *to appear in Robotics and Autonomous Systems*, 1997.

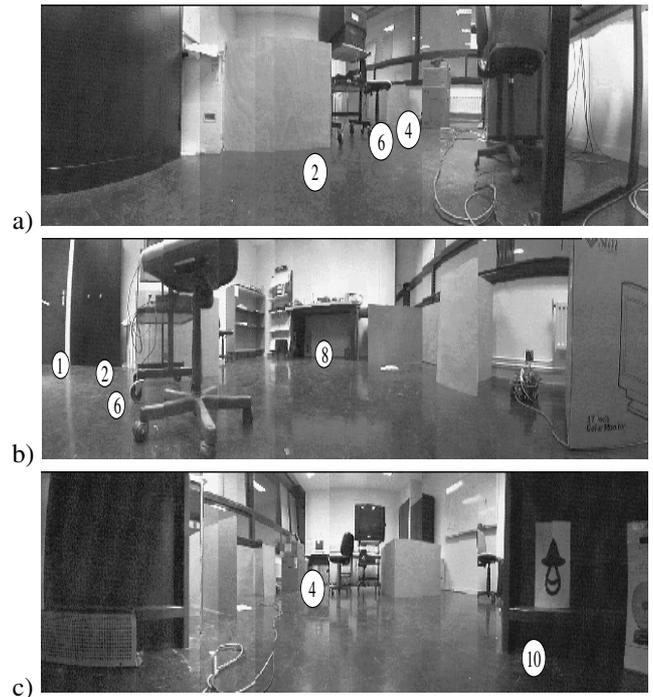


FIG. 9 — Les éléments sont ceux du plan fig. 8. a) Vue panoramique depuis le point de départ. b) Vue prise à la moitié de la trajectoire du robot. c) Vue panoramique près de la position du but (le champignon).

- [8] P. Gaussier and S. Zrehen. Avoiding the world model trap : An acting robot does not need to be so smart ! *Journal of Robotics and Computer-Integrated Manufacturing*, 11(4) :279–286, 1994.
- [9] P. Gaussier and S. Zrehen. Navigating with an animal brain : a neural network for landmark identification and navigation. In *Proceedings of Intelligent Vehicles*, pages 399–404, Paris, 1994.
- [10] P. Gaussier and S. Zrehen. Perac : A neural architecture to control artificial animals. *Robotics and Autonomous Systems*, 16(2-4) :291–320, 1995.
- [11] C. Joulain, P. Gaussier, and A. Revel. Learning to build categories from perception-action associations. In *To appear in International Conference on Intelligent Robots and Systems - IROS'97*, Grenoble, France, September 1997. IEEE/RSJ.
- [12] M. Quoy, O. Gallet, and P. Gaussier. Implémentation parallèle d'un système de focalisation de l'attention. In *A paraître dans les actes du GRETSI'97*, Grenoble, France, Septembre 1997.