

Invariants intégraux pour l’indexation d’images omnidirectionnelles

Cyril CHARRON¹, Ouiddad LABBANI-IGBIDA¹, El Mustapha MOUADDIB¹

¹Centre de Robotique, d’Électrotechnique et d’Automatique
7 rue du Moulin neuf, 80000 Amiens, France

Cyril.Charron@u-picardie.fr, Ouiddad.Labbani@u-picardie.fr, Mouaddib@u-picardie.fr

Résumé – Dans cette étude, nous nous intéressons à la localisation d’un robot mobile par indexation d’images omnidirectionnelles. La localisation est qualitative et le robot doit déterminer la pièce dans laquelle il se trouve. L’indexation est réalisée à partir d’invariants intégraux calculés sur les images omnidirectionnelles. Notre méthode utilise l’intégrale de Haar dans laquelle nous intégrons un modèle du capteur omnidirectionnel. L’approche intégrale nous permet de prendre en compte simplement la géométrie du capteur ainsi que les déplacements du robot pour le calcul des invariants. La robustesse de cette nouvelle méthode à des déplacements dans une pièce est supérieure à une méthode basée sur l’histogramme.

Abstract – The present study proposes an innovative approach to mobile robot’s localization using omnidirectional image indexation. The localization is qualitative and the robot has to determine the room it is situating in. Indexation is completed thanks to integral invariants computed on omnidirectional images. Our method starts from the Haar integral to which we include a model of the omnidirectional sensor. The integral approach allows us to cope simply with the sensor geometry and with the robot’s movements to compute the invariant. The robustness of this invariant to the movements of the robot in a given room is improved when compared to an histogram based method.

1 Introduction

La localisation dans des environnements intérieurs ou extérieurs est une condition importante pour l’autonomie des robots mobiles. De nombreuses méthodes pour la localisation d’un robot mobile ont été développées. Elles peuvent être classées en trois catégories : les méthodes odométriques [1], les méthodes basées sur des amers [2] et les méthodes basées sur l’indexation d’images [3]. Notre travail s’intègre dans cette dernière catégorie où nous proposons une méthode de construction de caractéristiques invariantes sur des images omnidirectionnelles en vue de localiser qualitativement un robot mobile. La localisation qualitative permet une estimation du lieu où se trouve le robot sans référence à une pose précise. L’intérêt de notre approche est qu’elle ne nécessite pas d’extraction de caractéristiques géométriques des images, ni de modélisation *a priori* de l’environnement. Ceci la rend attractive pour une application temps réel sur le robot.

Une revue intéressante des différentes méthodes d’indexation peut être trouvée dans [4] et [9]. Les histogrammes sont très souvent utilisés dans les systèmes de recherche d’images. Ils ont été appliqués à la localisation de robots mobiles à partir d’images omnidirectionnelles [3] pour leur représentation compacte des images, leur invariance aux rotations (ce qui est très intéressant pour les images omnidirectionnelles) et leur faible sensibilité à de petites translations. Toutefois, lors de grandes translations, de nombreux histogrammes de référence doivent être appris pour que la localisation soit fiable. En effet, les histogrammes ne sont pas robustes à de grands déplacements. L’idée de base de notre approche est d’appliquer un for-

malisme mathématique de calcul d’invariant (l’intégrale de Haar) pour construire des distributions de caractéristiques invariantes adaptées au capteur omnidirectionnel. Cette construction est adaptée de façon à tenir compte du modèle géométrique du capteur omnidirectionnel et des mouvements potentiels du robot. Les distributions résultantes sont robustes à des positions différentes du robot dans la scène.

Après avoir introduit les bases théoriques de la construction d’invariants intégraux de caractéristiques locales, nous présentons notre approche intégrant simultanément un modèle du capteur omnidirectionnel et les effets des mouvements du robot sur les images. Les résultats expérimentaux présentés ici montrent une amélioration de l’invariance de ces distributions par rapport à l’histogramme classique, et donc de la localisation qualitative.

2 Construction de l’invariant

La signature invariante d’une image est calculée à partir de l’intégrale de Haar et prend la forme d’une distribution. Dans ce qui suit, nous décrirons la méthode de construction de cet invariant et son adaptation à notre problème, une caméra omnidirectionnelle embarquée sur un robot mobile.

Le point de départ pour la construction de l’invariant est l’intégrale de Haar. Elle consiste en un parcours de l’espace des paramètres du groupe de transformation. Elle est généralement exprimée sous la forme :

$$I_{Haar} = \frac{1}{|G|} \int_G f \circ g(\mathbf{x}) dg \quad \text{avec} \quad |G| = \int_G dg$$

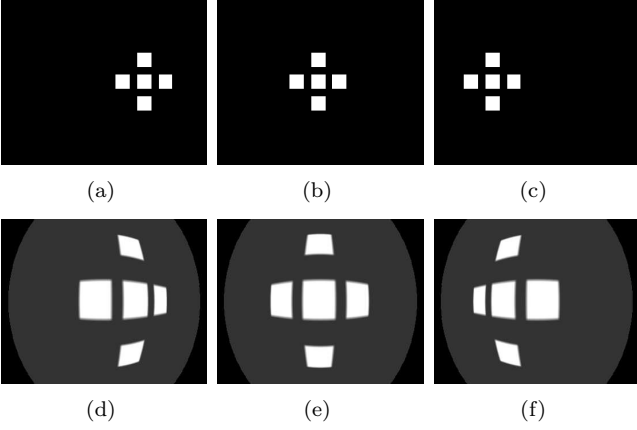


FIG. 1 – Translation dans le plan perpendiculaire à l'axe d'une caméra affine (a, b et c). Translation dans le plan perpendiculaire à l'axe d'une caméra munie d'un capteur omnidirectionnel (d, e et f).

où G est le groupe de transformation et $g(\mathbf{x})$ l'action d'un élément g de ce groupe sur le vecteur \mathbf{x} . Cet invariant a été utilisé en recherche d'image par Siggelkow [8] et pour la localisation de robot mobile par Wolf [5], bien que l'intégrale de Haar n'était pas explicitement utilisée.

2.1 Le groupe de transformation

Notre intérêt porte sur les transformations de l'image obtenue par une caméra. La nature des transformations est due aux mouvements du robot et au processus de projection (Fig. 1). Dans ce travail, l'étude des mouvements du robot a été limitée aux translations sur le sol. Les translations transforment le point 3-D (exprimé dans le référentiel du robot) en le point $\mathbf{x} + \mathbf{t}$ avec \mathbf{t} une translation dans le plan (Ox, Oy) .

La caméra est munie d'un capteur omnidirectionnel, produisant des transformations qui peuvent être séparées en une projection centrale sur le miroir parabolique suivie d'une projection orthographique sur le plan image. La projection du point \mathbf{x} sur le miroir est modélisée par l'équation (1) où \mathbf{x}_m définit le point correspondant à \mathbf{x} sur la surface du miroir (2). Le point $\mathbf{u} = (u, v)$ de l'image est obtenu par une projection orthographique de \mathbf{x}_m le long de l'axe optique de la caméra (3) (alignée suivant Oz). $\boldsymbol{\alpha} = (\alpha_u, \alpha_v)$ et $\mathbf{u}_0 = (u_0, v_0)$ sont les paramètres intrinsèques de la caméra, et h est le *latus rectum* du miroir parabolique.

$$\begin{pmatrix} x_m \\ y_m \\ z_m \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (1)$$

$$z_m = \frac{h^2 - (x_m^2 + y_m^2)}{2h} \quad (2)$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & 0 & u_0 \\ 0 & \alpha_v & 0 & v_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_m \\ y_m \\ z_m \\ 1 \end{pmatrix} \quad (3)$$

$$\lambda = \frac{h}{\sqrt{x^2 + y^2 + z^2 - z}} \quad (4)$$

$$g(\mathbf{M})[\mathbf{u}] = \mathbf{M}[\mathbf{u}'] \quad \text{and} \quad \mathbf{M}'[\mathbf{u}] = \mathbf{M}[\mathbf{u}'] \quad (5)$$

Après un mouvement, l'image $\mathbf{M}[\mathbf{u}]$ est transformée en l'image $\mathbf{M}'[\mathbf{u}] = \mathbf{M}[\mathbf{u}']$. Le point \mathbf{u}' défini par l'équation (5) est l'image du point $\mathbf{x}' = \mathbf{x} + \mathbf{t}$ résultant de la translation du robot dans l'environnement.

L'intégrale de Haar peut être formulée pour l'image \mathbf{M} en considérant les translations \mathbf{t} du robot. L'intégrale de Haar s'écrit alors :

$$I_{Haar} = \frac{1}{|G|} \int_G f \circ g(\mathbf{M}[\mathbf{u}]) d\mathbf{t}$$

Notre objectif est d'évaluer cette intégrale à partir d'une seule image, en simulant un parcours virtuel de la pièce dans laquelle se situe le robot. La fonction f est une fonction noyau à support compact choisie arbitrairement¹. Supposons, sans perte de généralité que $f(\mathbf{M}) = \mathbf{M}[\mathbf{0}]$. L'intégrale de Haar peut être écrite :

$$\begin{aligned} I_{Haar} &= \frac{1}{|G|} \int_G f(\mathbf{M}'[\mathbf{u}]) d\mathbf{t} \\ &= \frac{1}{|G|} \int_G \mathbf{M}'[\mathbf{0}] d\mathbf{t} \\ &= \frac{1}{|G|} \int_G \mathbf{M}[\mathbf{0}'] d\mathbf{t} \\ &= \frac{1}{|G|} \int_G \mathbf{M}[\mathbf{l}] d\mathbf{t} \end{aligned} \quad (6)$$

où $\mathbf{l} = (l, m)$ est la translation depuis l'origine exprimée dans les coordonnées images telle que

$$\mathbf{l} = \boldsymbol{\alpha} \frac{h\mathbf{t}}{\sqrt{\mathbf{t}^2 + z^2 - z}}. \quad (7)$$

L'intégrateur $d\mathbf{t}$ est formulé dans le référentiel lié au monde. Nous utilisons la formule du changement de variable pour des intégrales multidimensionnelles (8) afin d'exprimer l'intégrale de Haar dans les coordonnées images, en utilisant la relation $\mathbf{t} = r(\mathbf{l})$.

$$\int_G \mathbf{M}[\mathbf{l}] d\mathbf{t} = \int_{r(G)} \mathbf{M}[\mathbf{l}] D_r(\mathbf{l}) d\mathbf{l} \quad (8)$$

$D_r(\mathbf{l})$ est le module de la matrice jacobienne. Après calcul et en substituant $D_r(\mathbf{l})$ dans l'intégrale, on obtient :

$$I_{Haar} = \frac{1}{|G|} \int_{r(G)} \mathbf{M}[\mathbf{l}] \left| \frac{4 \cdot h^2 \cdot z^2}{(h^2 - \mathbf{l}^2)^3} (h^2 + \mathbf{l}^2) \right| d\mathbf{l}.$$

Tous les termes de cette expression, sauf z , sont endogènes à l'image. z est choisi arbitrairement sur une sphère de rayon R centrée sur le capteur, et peut dès lors s'écrire en fonction de \mathbf{l} (9) en utilisant (1), (2) et (4). Ce choix revient à faire l'hypothèse d'un monde hémisphérique entourant le robot. Cette contrainte peut être levée en intégrant les mouvements réels (et non virtuels) du robot (ce travail est en cours de validation).

$$z = R \frac{\mathbf{l}^2 - h^2}{\mathbf{l}^2 + h^2} \quad (9)$$

2.2 La fonction noyau

Dans le paragraphe précédent, le noyau $f(\mathbf{M}) = \mathbf{M}[\mathbf{0}]$ a été retenu. D'un point de vue photométrique, ce noyau prend uniquement la valeur du niveau de gris du pixel.

¹Dans la pratique, le choix du noyau f peut conditionner la stabilité et la robustesse de l'intégrale

L'évaluation de l'intégrale de Haar peut ainsi être interprétée comme une moyenne de groupe de l'information lumineuse de l'image. Le pouvoir descriptif et discriminant de la valeur moyenne est cependant pauvre comparé à celui d'un histogramme, par exemple. Dans le même sens, on peut définir un nouveau noyau [8] :

$$f_i(\mathbf{M}) = \begin{cases} \text{créneau}(\mathbf{M}[\mathbf{0}] - ndg_i) \\ = \begin{cases} \frac{-|\mathbf{M}[\mathbf{0}] - ndg_i|}{\Delta ndg_i} + 1 & \text{si } |\mathbf{M}[\mathbf{0}] - ndg_i| < \Delta ndg_i \\ 0 & \text{sinon} \end{cases} \end{cases}$$

où ndg_i est un niveau de gris de référence et Δndg_i est la distance séparant ndg_i et ndg_{i+1} . Une telle fonction noyau est préférée à un delta de Kronecker pour sa continuité et pour le recouvrement partiel des supports entre plusieurs de ces fonctions. Ceci permet d'éviter des effets de bord aux extrema du support (si un niveau de gris d'un pixel se trouve sur le bord du support d'une fonction f_i , une petite variation dans l'environnement ou dans le processus de formation de l'image peut le modifier et le faire tomber dans le support d'une des fonctions adjacentes à f_i).

En évaluant l'intégrale avec ce noyau, on peut calculer la probabilité pour chaque point de l'image d'appartenir à une certaine gamme de niveaux de gris (définie par la largeur du support de la fonction noyau), la probabilité de chaque point étant pondérée par $D_r(l)$. Considérons $F = \{f_i\}_{i=1,\dots,n}$, une famille de fonctions qui partitionne l'espace des niveaux de gris. On obtient alors une distribution d'invariants I_i .

Cette distribution ressemble à un histogramme classique. On montre que ce dernier est identique aux distributions d'invariants dans le cas d'une caméra affine au lieu du capteur omnidirectionnel utilisé ici (si on considère des transformations euclidiennes dans un plan parallèle au plan image d'une caméra affine, alors $D_r(l)=1$, et l'histogramme classique est obtenu à partir de l'intégrale de Haar).

Chaque f_i définit un invariant I_i et l'ensemble des I_i est une distribution discrète de caractéristiques locales invariante à des rotations, des changements d'échelle et de luminosité. La reconnaissance d'une image est alors réalisée en fonction d'une mesure de distance entre sa distribution associée et les distributions enregistrées.

3 Mesure de distance entre distributions

Le choix de la mesure de distance entre deux distributions discrètes est un problème qui a été abondamment discuté. Il est généralement admis que l'on peut le séparer en deux parties distinctes [7] : une mesure de distance dans l'espace des caractéristiques dont la distribution est étudiée, que l'on nomme la distance au sol (la distance entre deux couleurs dans l'espace colorimétrique par exemple), et une mesure de distance entre chaque distribution vue comme un tout.

Un premier type de mesure de distance entre une distribution $H = \{h_i\}_{i=1,\dots,n}$ et une distribution $K = \{k_i\}_{i=1,\dots,n}$ consiste en une comparaison des termes de H et K de même indice. Ce type de mesure est sensible à la largeur

de l'intervalle de partitionnement : si elle est trop étroite, des caractéristiques similaires ne seront pas appariées ; si elle est trop large, des caractéristiques différentes ne seront pas distinguées. L'intersection et la divergence de Jeffrey appartiennent à ce type de mesure. Ici, la distance entre les caractéristiques n'est pas prise en compte. Pour la prendre en compte plus finement et ainsi pouvoir tolérer de petites variations de la distribution, on peut autoriser la comparaison entre différents termes de H et de K , gardant à l'esprit la distance propre à l'espace des caractéristiques sous-jacent. La distance quadratique, la mesure de Kolmogorov-Smirnov et la distance des transporteurs de terre (10) (DTT, de l'anglais *EMD*)[7] appartiennent à cette catégorie. Dans (10), f_{ij} est le flux de matière entre le point i et le point j qui minimise un travail et les d_{ij} sont les coefficients d'une matrice représentant les distances entre ces points.

$$\text{EMD}(H, K) = \frac{\sum_i \sum_j d_{ij} f_{ij}}{\sum_i \sum_j f_{ij}} \quad (10)$$

Une étude comparative menée par [6] montre que la divergence de Jeffrey et la DTT donnent les meilleurs résultats. Dans nos expériences, la mesure DTT s'est avérée meilleure que la mesure de Jeffrey et nous l'avons retenue.

4 Application

Afin de valider notre méthode de localisation qualitative, nous la comparons à un histogramme classique. Nos expériences visent à étudier l'évolution de la mesure de distance entre deux distributions reliées par le mouvement du robot séparant deux prises d'images. De plus, notre but est aussi d'étudier la sensibilité de notre méthode aux rotations, qui n'ont pas été incluses au groupe de transformation.

L'implémentation a été réalisée sous Matlab et une *dll* a été compilée pour accélérer les calculs. Pour une image (576×768 pixels) et sur un P-IV à 2,4GHz, le calcul prend environ 0,5s.

Des prises de vue ont été effectuées dans quatre pièces de notre laboratoire dont les dimensions sont approximativement de 5m par 5m (excepté l'atelier environ deux fois plus grand). Le robot s'est déplacé sur le sol en suivant une grille de 3m par 3m centrée dans chaque pièce. 25 positions sont définies sur cette grille.

4.1 Invariance aux translations

De façon à clarifier la dépendance de la mesure de distance à l'amplitude du mouvement, la distance DTT est calculée pour chacun des couples de positions possibles de la grille (325 couples) et mise en relation avec la distance physique séparant les prises de vue. Ceci est effectué pour notre méthode et pour celle basée sur les histogrammes. Les résultats obtenus sont alors moyennés pour chacune des distances physiques (euclidiennes) possibles et ce calcul est répété pour l'ensemble des quatre pièces étudiées. Les résultats sont résumés dans le tableau 1. Notre méthode produit en moyenne des distances plus faibles de

16% par rapport à la méthode basée sur l’histogramme. En moyenne, notre méthode se comporte mieux, *i.e.* produit des variations plus faibles de distance au sein d’une pièce donnée.

Il faut maintenant vérifier le comportement de la mesure de distance lorsque des distributions de deux pièces différentes sont comparées. Pour ce faire, nous avons déterminé pour chaque pièce la distribution (resp. histogramme) dont les distances aux autres distributions (resp. histogrammes) de cette pièce sont minimales. Ceci nous permet de déterminer un bon prototype pour chaque pièce. Ce prototype est ensuite comparé à l’ensemble des distributions (resp. histogrammes) des autres pièces. La distance moyenne est présentée dans le tableau 1. Les deux méthodes offrent des résultats similaires. La distance moyenne est environ 4 fois supérieure à la plus grande distance que l’on a obtenue au sein d’une même pièce.

TAB. 1 – Variation de la mesure DTT (en unité arbitraire) pour des translations et des rotations. Notre méthode (*distri.*) et la méthode basée sur l’histogramme (*histog.*) sont comparées grâce à une différence relative (*différence rel.*).

	Translation (cm)					
	0	60	85	120	134	170
histog. ($\times 10^{-2}$ DTT)	0,0	15,8	25,0	17,6	20,8	22,2
distri. ($\times 10^{-2}$ DTT)	0	12,6	19,3	14,8	16,8	18,7
différence rel. (%)	0	-14	-16	-9	-14	-12
	Rotation (degré)					
	30	90	150	210	270	330
histog. ($\times 10^{-2}$ DTT)	1,5	2,4	2,4	2,2	2,4	2,4
distri. ($\times 10^{-2}$ DTT)	2,2	4,0	3,2	3,5	3,7	2,9
différence rel. (%)	33	69	34	63	54	24
	Changement de pièce moyenne sur le labo.					
	histog. ($\times 10^{-2}$ DTT)	90				
distri. ($\times 10^{-2}$ DTT)	88					
différence rel. (%)	2					

4.2 Invariance aux rotations

Pour tester les rotations, 35 prises de vues séparées par une rotation de 10° autour de l’axe du capteur (Oz) ont été stockées. La première distribution (correspondant à l’image prise à 0°) est choisie comme référence. Pour chaque distribution obtenue lors de la rotation du robot, la distance DTT à cette référence est calculée. Les deux méthodes produisent des distances restant dans une faible gamme de valeurs lorsqu’on les compare aux mesures obtenues lors des translations (tableau 1). Ainsi, les deux méthodes peuvent être considérées comme produisant des invariants aux rotations.

5 Conclusion

Nous avons présenté une méthode de construction d’invariants basée sur l’intégrale de Haar et nous l’avons adaptée au cas de la caméra omnidirectionnelle. L’avantage principal de cette méthode est qu’elle permet de tenir

compte des mouvements réels du robot pour construire l’invariant. Nous sommes partis des mouvements du robot dans le monde (translation dans le plan du sol) et nous avons considéré leur effet sur des images omnidirectionnelles. Nous avons montré que cette méthode améliore sensiblement les résultats en terme de réduction de distance entre des distributions d’une même pièce par rapport à ce qui est obtenu avec des histogrammes classiques. De plus, la distance entre des distributions de pièces différentes est aussi élevée que lorsque des histogrammes sont utilisés, ce qui est une bonne propriété en vue de l’intégration dans un système de classification. Enfin, l’invariant obtenu est aussi robuste aux rotations bien qu’elles n’aient pas été explicitement prises en compte dans le groupe de transformation retenu. En conclusion, cette méthode de construction et le choix d’un noyau géométrique adapté à la topologie spécifique du capteur omnidirectionnel pourrait d’avantage améliorer la reconnaissance de lieux (moyennant l’utilisation d’un système de classification pour traiter les distances fournies par notre méthode).

Remerciements

Ce travail a été financé par la région Picardie (Pôle DIVA).

Références

- [1] J. Borenstein, B. Everett, and L. Feng. *Navigating Mobile Robots : Systems and Techniques*. A. K. Peters, Ltd., Wellesley, MA, February 1996.
- [2] J. Castellanos and J. Tardos. *Mobile Robot Localization and Map Building - A Multisensor Fusion Approach*. Kluwer Academic Publishers, 1999.
- [3] J. Gozalez and S. Lacroix. Rover localization in natural environments by indexing panoramic images. In *IEEE Int. Conf. on Robotics and Automation*, pages 1365–1370, 2002.
- [4] F. Idris and S. Panchanathan. Review of image and video indexing techniques. *Journal of Visual Communication and Image Representation*, 8(2) :146–166, June 1997.
- [5] J. Wolf W. Burgard and H. Burkhardt. Using an image retrieval system for vision-based mobile robot localization. In N. Sebe M. S. Lew and J. P. Eakins, editors, *Proc. of the Int. Conf. on Image and Video Retrieval (CIVR)*, pages 108–119, 2002.
- [6] Jan Puzicha, Yossi Rubner, Carlo Tomasi, and Joachim M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. In *Int. Conf on Computer Vision (ICCV)(2)*, pages 1165–1172, 1999.
- [7] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. The earth mover’s distance as a metric for image retrieval. *Int. Journal of Computer Vision*, 40(2) :99–121, November 2000.
- [8] S. Sigglekow. *Feature histograms for content-based image retrieval*. PhD thesis, Universitat Freiburg im Breusgau, 2002.
- [9] M.J. Swain and D.H. Ballard. Color indexing. *Int. Journal of Computer Vision*, 7(1) :11–32, 1991.