

Invariant couleur basé sur la « transformation du boulanger » – Application à la localisation globale d'un robot d'intérieur

C. MONTAGNE*, S. LELANDAIS*, A. CHAARI*# et M. BEN AHMED#

**Laboratoire Systèmes Complexes - CNRS FRE 2494 - Université d'Evry
40, rue du Pelvoux - 91020 Evry Cedex - FRANCE*

cmontagne@iup.univ-evry.fr
s.lelandais@iut.univ-evry.fr

*#Laboratoire RIADI - Laboratoire de Recherche en Génies Documentiel & Logiciel
Ecole Nationale des Sciences de l'Informatique
Campus Universitaire de la Manouba - 2010 La Manouba - TUNISIE*

Résumé: Dans cet article nous proposons un nouvel invariant couleur défini en utilisant la transformation du boulanger. Cet invariant est utilisé comme descripteur d'une image. Il nous permet de rechercher, à travers une base de données constituée d'images d'intérieur, celle qui est la plus proche d'une image dite « requête ». L'objectif de ce travail est de pouvoir permettre à un robot, perdu dans son environnement défini par un appartement, de se localiser de manière assez grossière en identifiant la pièce où il se trouve et en déterminant son orientation dans cette pièce. Chaque image de la base de données possède son descripteur couleur, en l'occurrence l'invariant ici proposé. La projection de l'image « requête » à travers tous ces invariants nous permet d'obtenir une mesure de distance entre cette image et toutes celles de la base. Les premiers résultats obtenus au moyen de cette approche nous permettent de valider l'intérêt de l'utilisation de cet invariant.

Mots clé: Imagerie couleur, Indexation d'images, Localisation par vision, Transformation quasi-mixing.

1 Introduction

Le Laboratoire Systèmes Complexes (LSC) est impliqué dans le développement d'un robot d'aide aux personnes handicapées moteur, à travers le projet ARPH (Hoppenot & al., 2001). Ce projet bénéficie du support de l'AFM, Association Française contre les Myopathies.

Le système proposé comprend un poste de commande et une base mobile équipée d'un bras manipulateur. Ce système doit pouvoir fonctionner sous la direction d'une personne ou en mode totalement autonome (Hoppenot & al., 2001). Dans ce dernier cas, la base mobile doit donc être capable de se localiser dans l'appartement de la personne. Mais, pour des principes évidents de confort de l'occupant, il n'est pas prévu d'équiper l'appartement de dispositifs spécialisés permettant à la base de se repérer. Il nous faut donc mettre au point un système de localisation à deux niveaux :

- Une localisation fine, qui permettra à la base de connaître sa position et son orientation dans une pièce et de calculer sa distance aux objets présents dans la scène.

- Une localisation frustre qui permettra à la base, en cas de perte complète de repère, de savoir grossièrement dans quelle pièce elle se trouve et d'obtenir une indication sur son orientation.

Différents travaux menés au laboratoire portent sur la localisation fine, soit à partir d'une vision monoculaire et d'un modèle de la pièce (Ait Aider 2002), soit en vision stéréoscopique noir & blanc (Loaiza & al., 2001) ou couleur (Montagne & al., 2002) (Montagne & al., 2004). Par contre nous débutons les travaux sur la localisation grossière. Le principe que nous avons retenu est le suivant. Une prise exhaustive de photographies est réalisée dans l'appartement dans lequel se déplace le robot (i.e. la base mobile). Ces photographies constituent une base de données d'images qui sont accompagnées de renseignements concernant la pièce dans laquelle elles ont été prises : nom de la pièce, orientation, pièces

liées, vues connexes... Lorsque la base est perdue, elle fournira au système une prise de vue de son environnement qui représentera une interrogation pour la base de données d'images. C'est ce que nous appellerons l'image requête dans la suite de cet article. Le but sera de retrouver rapidement l'image la plus ressemblante à l'image requête de manière à permettre au système de pouvoir aborder l'étape de localisation fine. Cette procédure relève de la problématique générale de l'indexation d'images (Mulhem 2002). Nous proposons ici la définition d'un nouvel invariant couleur qui sera utilisé comme descripteur de chacune des images de la base de données. L'utilisation de la couleur est motivée par le fait qu'en robotique d'intérieur l'environnement est riche d'informations de cette nature qui varieront d'une pièce à l'autre, selon les fonctions de la pièce et les goûts de la personne qui y habite

Le descripteur que nous allons définir est dérivé d'un travail précédent, basé sur la technique nommée « quantification de la couleur » (Orchard & al., 1991) (Sharma & al., 1997), qui vise à limiter de manière adaptative le nombre de couleurs dans une image (Montagne & al., 2004 bis). Cette réduction adaptative des couleurs est fondée sur l'utilisation de la « transformation du boulanger » (Smolarz & al., 2002) (Cornu & al., 2002).

Dans la première partie de cet article nous présenterons le principe de cette transformation qui est issue de l'analyse des processus ergodiques. Dans une deuxième partie nous expliquerons comment il est possible d'utiliser cette transformation pour obtenir une palette caractéristique des couleurs de l'image d'origine, qui constituera notre invariant couleur. Une troisième partie présentera le principe de recherche de l'image la plus proche de la requête à l'intérieur de la base. La quatrième partie consistera en la présentation des résultats obtenus sur les images de la base de données. Différentes situations ont été testées qui seront discutées. Enfin nous concluons en évoquant les suites possibles de ce travail.

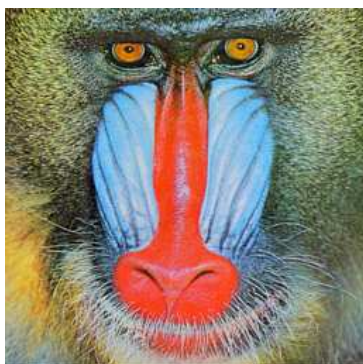


Figure 1. Image originale (256 x 256 pixels).

2 La transformation du boulanger

La transformation du boulanger – notée TB par la suite – est basée sur la définition du mélange des systèmes dynamiques (Arnold & al., 1967) (Billingsley 1965). L'intérêt principal de cette transformation est qu'elle mélange d'une manière très homogène tous les éléments de l'espace de travail.

Arnold & Avez (Arnold & al., 1967) ont donné un grand nombre d'exemples de ces transformations qui sont définies sur un carré de taille unité $[0,1] \times [0,1]$. Nous utilisons l'une d'entre elles : la transformation du boulanger. Il faut toutefois mentionner ici que tous les exemples donnés par Arnold & Avez sont définis sur des ensembles continus. Or une image est constituée d'un ensemble fini de points : les pixels. Malheureusement, il a été montré qu'une transformation définie sur un ensemble fini de points ne peut être une transformation « mélange ». Toutefois, pour des transformations « mélange » de type particulier, comme la TB, la restriction à des ensembles finis conserve certaines propriétés malgré tout. En particulier on peut démontrer que les pixels sont statistiquement convenablement mélangés quand on itère la TB un nombre convenable de fois.

Une itération de la TB est basée sur deux étapes. Tout d'abord on utilise une transformation affine qui conduit à l'obtention d'une image deux fois plus large et moitié moins haute que l'image d'origine (voir les figures 1 et 2), le nombre total de pixels restant inchangé. Puis on coupe verticalement par le milieu l'image résultante de l'étape 1 et on pose la moitié droite sur la moitié gauche (voir figure 3), l'image finale ayant exactement la même taille que l'image d'origine. La TB impose l'utilisation d'images carrées de taille $2^N \times 2^N$ pixels. On peut montrer que la TB est périodique. Ainsi au bout de $4N$ itérations, l'image obtenue est identique à l'image initiale. L'image est convenablement mélangée au bout de N itérations comme le présente la figure 4. Si nous prélevons dans l'image mélangée une fenêtre de taille $2^p \times 2^p$ pixels, nous obtenons une bonne représentation de l'image originale à une taille réduite après applications de $3p$ itérations de la TB à cette fenêtre (voir figure 5).



Figure 2. Première étape de l'itération initiale de la TB.



Figure 3. Deuxième de l'itération initiale de la TB.

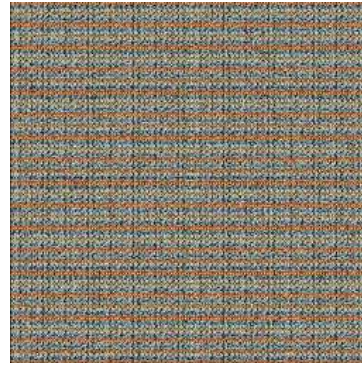


Figure 4. Image convenablement mélangée (8 itérations).

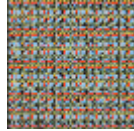


Figure 5. Fenêtre de taille $2^6 \times 2^6$ extraite de la figure 4 et résultat obtenu après 18 itérations.

D'un point de vue pratique, la TB est une transformation spatiale : pour une dimension d'image donnée, la position d'arrivée des pixels dans l'image mélangée est toujours la même. En conséquence, une « LUT », indiquant pour chaque pixel d'une image ses coordonnées dans l'image mélangée, permettra de s'affranchir de la transformation proprement dite et d'obtenir les « imagettes » plus rapidement. La TB ou l'emploi d'une « LUT » revient à extraire de façon homogène les pixels dans l'image d'origine et à les placer dans l'image mélangée de manière à former des « imagettes » successives. Ce prélèvement homogène peut être appliqué à des images rectangulaires et permet alors de s'affranchir de la condition sur la dimension des images (dimension en puissance 2).

3 Obtention de l'invariant

Comme nous venons de le montrer dans le paragraphe précédent, une « imagette » de taille $2^p \times 2^p$ donne une bonne représentation de l'image d'origine tant au niveau des formes que des textures et des couleurs. Aussi l'idée que nous avons eue consiste à utiliser une de ces fenêtres extraites de l'image mélangée en tant que palette pour réduire tous les niveaux de couleurs présents dans l'image d'origine. Avec une image de taille $2^N \times 2^N$ pixels, il est possible d'obtenir K palettes contenant $2^p \times 2^p$ couleurs ($p < N$). Le nombre de palettes disponibles différentes obtenues à partir d'une image est $K = 2^{2(N-p)}$. Dans ce papier nous présentons des exemples avec des palettes contenant 256, 64, 16 ou 4 couleurs, palettes correspondantes à des fenêtres extraites de taille $2^4 \times 2^4$, $2^3 \times 2^3$, $2^2 \times 2^2$, $2^1 \times 2^1$ (ie $p=4, 3, 2, 1$) à partir de l'image mélangée. Une fois la palette choisie, pour chaque pixel de l'image d'origine, nous calculons la distance euclidienne entre sa couleur initiale et les $2^p \times 2^p$ couleurs de la palette. La couleur retenue pour être

affectée comme couleur réduite au pixel est celle conduisant à la distance minimum. Un problème se pose cependant : « Comment choisir la fenêtre la plus représentative de l'image d'origine permettant d'obtenir la meilleure palette? » Pour se faire, nous proposons quatre solutions (Montagne & al., 2004 bis) qui s'intègrent au processus de la réduction comme l'illustre le schéma général présenté sur la figure 6. A partir de l'image initiale, nous itérons la transformation du boulanger jusqu'à obtenir l'image convenablement mélangée. Connaissant le nombre de couleurs souhaitées dans l'image finale, nous déterminons la taille de la fenêtre à extraire pour construire la palette. C'est à ce moment que les méthodes diffèrent entre elles. La première sélectionne aléatoirement une fenêtre. La deuxième crée une fenêtre contenant pour chaque position une couleur médiane. La troisième sélectionne une fenêtre médiane. La quatrième crée une fenêtre moyenne. Une fois déterminée la palette, il reste à affecter une couleur de la palette à chaque pixel de l'image originale comme indiqué précédemment.

L'évaluation qualitative des résultats se fait par un calcul de distance entre l'image d'origine et l'image réduite en couleurs. Cette distance, baptisée « delta », est calculée sur chacun des trois canaux de l'image (Rouge, Vert et Bleu) pour tous les pixels de l'image, en utilisant l'équation [1], où I_1 et I_2 représentent les niveaux de couleurs d'un pixel respectivement dans l'image d'origine et dans l'image réduite.

$$\text{delta} = \frac{\sum_{i=1}^{2^N} \sum_{j=1}^{2^N} |I_1(i,j) - I_2(i,j)|}{2^N \times 2^N} \quad (1)$$

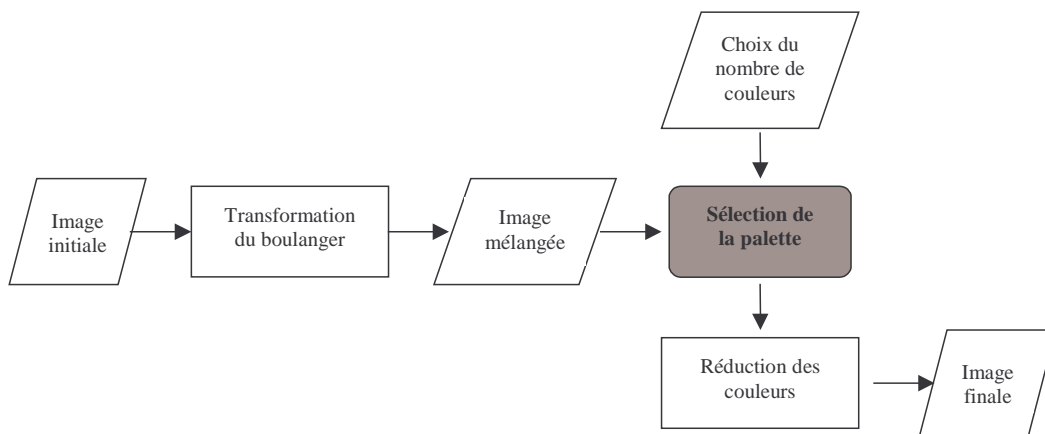


Figure 6. Principe du choix aléatoire de la palette.

Des quatre méthodes, la plus satisfaisante est celle qui aboutit à la palette dite « médiane ». Elle consiste à calculer, pour chaque « imagerie », un critère dont l'étude permet de choisir une palette parmi les K disponibles. Ce critère est une valeur baptisée « Sigma » et nous retenons la palette qui conduit à la valeur médiane de « Sigma », après classement par valeur croissante. La formule utilisée pour calculer « Sigma » est donnée dans l'équation [2]. Dans cette équation, *i* et *j* représentent les positions d'une couleur dans la palette (conservée sous forme matricielle) et *k*

est l'index de la palette. *R*, *G* et *B* sont respectivement les octets des canaux Rouge, Vert et Bleu d'une couleur.

$$Sigma_k = \sum_{i=1}^{2^{n/2}} \sum_{j=1}^{2^{n/2}} R_k(i,j) + G_k(i,j) + B_k(i,j) \quad (2)$$

La figure 7 présente les résultats obtenus avec cette méthode et le tableau 1 donne les valeurs de « delta » obtenues sur les trois canaux entre l'image d'origine et les images réduites.

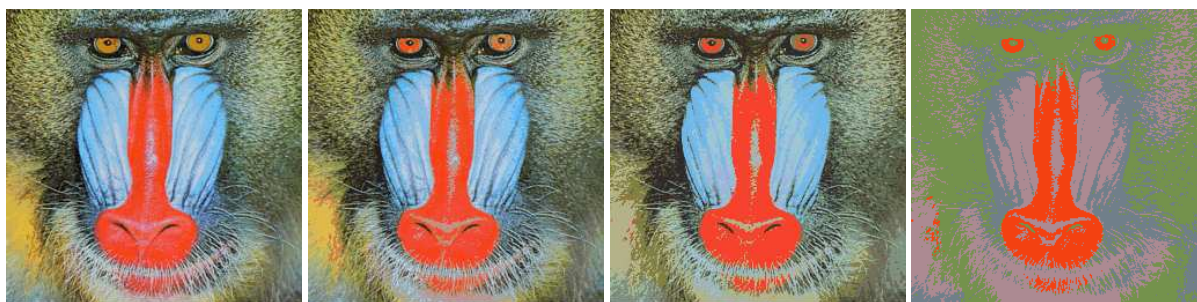


Figure 7. Image réduite à a) 256 couleurs ; b) 64 couleurs ; c) 16 couleurs ; d) 4 couleurs.

Nombre de couleurs	Canal Rouge	Canal Vert	Canal Bleu
256	4,9993	4,9006	4,8845
64	8,7838	8,0207	8,1871
16	13,2763	11,8202	15,7809
4	25,3266	38,3961	31,1236

Tableau 1. Valeur de la distance « delta » calculée à partir des quatre images précédentes.

4 Recherche de l'image proche de l'image requête

S'il est possible d'extraire d'une image un échantillon de pixels dont les couleurs sont représentatives de l'ensemble des couleurs de l'image, alors il est possible de qualifier un tel échantillon d'« invariant » couleur. Cet invariant sera utilisé comme descripteur indirect permettant d'obtenir des projections de l'image requête, à savoir la réduction du nombre de couleurs de cette image au moyen des différentes palettes disponibles. La recherche de la proximité entre images de la base et image requête est basée sur l'analyse de ces projections.

Soit la base d'images d'un lieu d'habitation : cette base comprend *x* images. De chacune de ces images, nous extrayons notre invariant. Nous avons vu que cet invariant était assimilable à une palette de couleurs : nous avons donc une seconde base de *x* palettes. Soit une image requête. Pour déterminer la pièce et l'orientation de la prise de vue, nous adoptons la démarche suivante :

- Nous procédons à une projection de l'image requête dans l'espace des couleurs défini par chacune des *x* palettes de la base. Ceci nous conduit à l'obtention de *x* images réduites.
- Nous calculons la différence entre l'image requête et chacune de ses réductions par une distance euclidienne.
- Nous classons les distances par ordre croissant.

- Nous déduisons de la distance minimale la palette conduisant à la réduction la plus proche de l'image requête. Ceci nous permet de remonter à l'image de la base correspondante et de conclure en proposant au système une localisation et une orientation grossière.

Le schéma de la figure 8 illustre cette manière de procéder.

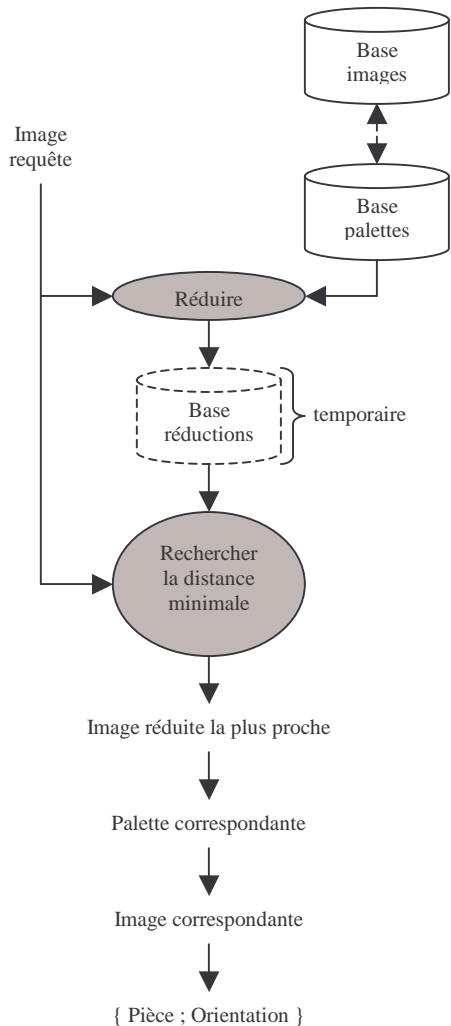


Figure 8. Principe de la localisation grossière.

5 Résultats

Nous avons expérimenté notre méthode sur une base d'images d'intérieur comprenant une série de 69 images. Ceci nous a permis d'obtenir des résultats préliminaires suffisamment intéressants pour valider notre approche. Nos premiers résultats montrent que la méthode décrite dans les paragraphes précédents fait bien la correspondance entre image requête et image(s) de la base représentant la même scène. Les figures suivantes montrent un exemple de résultat obtenu. La figure 9 est une image requête fournie par le capteur de vision du système robotisé. Les figures 10 correspondent à deux images résultant de la réduction du nombre de couleurs de l'image requête par rapport à deux des palettes des images de la base (palettes obtenues préalablement via la TB ou

assimilé). Ces palettes contiennent 256 couleurs. Nous constatons que l'image de la figure 10a conserve la même dynamique de couleurs que l'image requête. Par contre, l'image de la figure 10b est fortement dégradée. Le calcul de la distance « delta » entre image requête et images réduites mène à la même conclusion (cf. tableau 2).



Figure 9. Image requête.

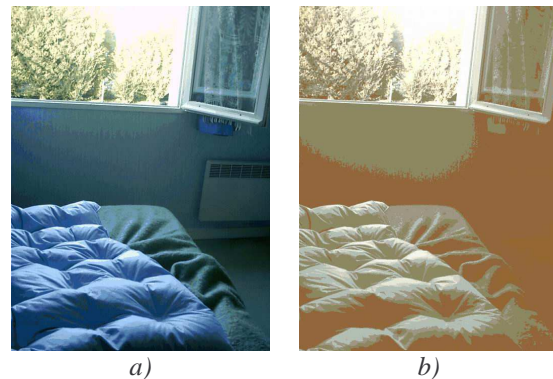


Figure 10. a) Image réduite proche ; b) Image réduite éloignée.

Figure	delta R	delta V	delta B	<delta>
10a	4.01	4.12	5.19	4.44
10b	73.19	30.49	23.86	42.52

Tableau 2. Différences entre image requête et réductions.

La palette ayant abouti à l'image de la figure 10a contient donc un ensemble de couleurs proches de celles de l'image requête, contrairement à l'image de la figure 10b. Une fois les palettes identifiées, nous remontons aux images de la base qui leur correspondent : ce sont les figures 11. Nous constatons que l'image de la figure 11a est similaire à l'image requête initiale. Notre méthode a donc correctement indiqué l'image de la base à laquelle pouvait se rattacher l'image requête. L'image de la figure 11b n'a par contre aucun rapport avec l'image requête comme nous pouvions nous y attendre à la lecture de « delta ». Nous pouvons parler de corrélation entre la distance « delta » et la qualité de la dynamique des couleurs d'une palette : plus la distance est faible, plus les couleurs d'une palette ont une dynamique proche de l'image requête, et *vice versa*.

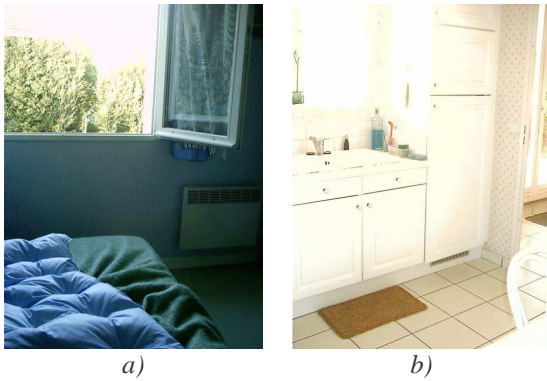


Figure 11. Images de la base ayant abouti aux figures 10a et 10b.

Conclusions et perspectives

Dans le cadre d'une localisation grossière d'un système robotisé égaré dans un lieu de vie, nous avons développé une méthode permettant une « indexation » d'images avec un invariant colorimétrique. Celui-ci se présente, pour chaque image d'une base, sous la forme d'un échantillon limité de couleurs, autrement dit une palette. Cette palette n'est pas utilisée à proprement parler comme descripteur mais comme outil de calcul : nous opérons de manière indirecte. Nos résultats préliminaires sont prometteurs mais il convient de les valider sur des bases d'images plus conséquentes, ce que nous avons commencé à faire sur deux bases de 300 images. Nous avons pris le parti d'utiliser des palettes de 256 couleurs mais nous comptons expérimenter la méthode en utilisant des palettes de plus ou moins grande taille. Notre objectif avec cette variation est de déterminer si une faible quantité de couleurs ne serait pas plus significative, un surplus d'informations pouvant être néfaste. D'un point de vue pratique, nous envisageons également la possibilité de répéter la méthode avec une seconde image requête prise selon une orientation différente de la première afin de confirmer ou d'infirmer le résultat initial.

Malgré son intérêt qui valide le concept d'invariant couleur, notre méthode est handicapée par un temps de calcul très important. La projection de l'image requête selon toutes les palettes de la base prend d'autant plus de temps que la base est volumineuse. Nous pourrions cependant considérer la palette comme un vrai descripteur et ainsi chercher à définir une distance qui permette de comparer les palettes entre elles : nous travaillerions avec une palette requête. Cette distance devra s'appuyer non seulement sur la comparaison des couleurs, avec une pondération possible du poids de chaque canal, mais aussi sur la prise en compte de l'organisation spatiale de ces couleurs pour présenter une certaine invariance aux translations et aux

changements de résolution de l'image requête par rapport aux images de la base.

Références

- (Ait Aider 2002) O. Ait Ader, « Localisation d'un robot mobile dans un environnement intérieur basée sur la vision monoculaire », Thèse de doctorat, Université d'Evry Val d'Essonne, décembre 2002.
- (Arnold & al., 1967) V. I. Arnold et A. Avez, « *Problème ergodiques de la mécanique classique* », Monographies Internationales de Mathématiques Modernes n°9, Gauthier-Villars, Paris, 1967.
- (Billingsley 1965) P. Billingsley, « *Ergodic Theory and Information* », John Wiley & Sons Inc., New-York, 1965.
- (Cornu & al., 2002) P. Cornu et A. Smolarz, « Caractérisation d'images par textures associées », *Traitement du Signal 2002*, 2002, vol.19, n°1, p.29-35.
- (Hoppenot & al., 2001) P. Hoppenot, E. Colle, O. Ait Aider and Y. Rybarczyk, « ARPH - Assistant Robot for Handicapped People - A pluridisciplinary project », *IEEE Roman'2001*, Bordeaux and Paris, France, 2001, pp.624-629.
- (Loaiza & al., 2001) H. Loaiza, J. Triboulet, S. Lelandais and C. Barat, « Matching segments in stereoscopic vision », *IEEE Instrumentation & Measurement Magazine*, March 2001, vol.4, n°1, pp.37-42.
- (Montagne & al., 2002) C. Montagne, S. Lelandais, J. Triboulet and C. N'Zi, « How to choose the best color space for the guidance of an indoor robot? », *CGIV 2002 - The First European Conference on Colour in Graphics, Image and Vision*, Poitiers, France, April 2002, pp.589-593.
- (Montagne & al., 2004 bis) C. Montagne, A. Smolarz, S. Lelandais and P. Cornu, « Adaptive color quantization using the baker's transform », *CGIV 2004 - The Second European Conference on Colour in Graphics, Image and Vision*, Aachen, Germany, April 2004, pp.353-358.
- (Montagne & al., 2004) C. Montagne, S. Lelandais, J. Triboulet, A. Smolarz and E. Colle, « New features color for indoor processing », *SETIT 2004*, Sousse, Tunisia, March 2004.
- (Mulhem 2002) P. Mulhem, « Indexation par apprentissage local et global pour la recherche symbolique d'images fixes », *RSTI - ISI, Recherche et filtrage d'informations*, juillet 2002, p.183-205.
- (Orchard & al., 1991) M.T. Orchard and C.A. Bouman, « Color quantization of images », *IEEE Transactions on Signal Processing*, 1991, vol.39, n°12, pp.2677-2690.
- (Sharma & al., 1997) G. Sharma and H.J. Trussel, « Digital Color Imaging », *IEEE Transactions on Image Processing*, July 1997, vol.6, n°7.
- (Smolarz & al., 2002) A. Smolarz and P. Cornu, « A Color Image Classification by Means of Image Transformations », *CGIV 2002 - The First European Conference on Colour in Graphics, Image and Vision*, Poitiers, France, 2002, pp.310-315.