

Le Modèle Nominatif de Régions : segmentation couleur et identification de régions par analyse de couleur et de texture

Patricia Lasserre, Rafael Murrieta Cid et Maurice Briot

L.A.A.S. -C.N.R.S.,
7 Av. du Colonel Roche,
31077 Toulouse Cedex 4, France

RÉSUMÉ

Nous nous intéressons ici au développement des capacités de perception d'un robot mobile destiné à évoluer en environnement naturel (extérieur). Dans ce cadre, la connaissance du terrain ou des objets situés dans l'environnement peut être améliorée en ajoutant des informations autres que géométriques, telles que la couleur et la texture. Nous avons proposé comme représentation complémentaire, le modèle nominatif de régions qui indique la nature de chaque région d'une image. Un algorithme de segmentation couleur donne une description très synthétique de la scène. Les régions issues de cette segmentation sont ensuite caractérisées par leur couleur et leur texture puis identifiées en utilisant une technique probabiliste afin d'obtenir leur nature (roche, herbe, terre . . .).

ABSTRACT

This paper deals with the development of the use of a video camera in order to add useful information other than geometrical, such as color and texture, for both localization and navigation of a mobile robot roving in natural environments. We have proposed and implemented as a complementary representation, the nominally model of regions which indicates each region nature in an image. A color segmentation algorithm provides a synthetic description of the scene. Regions issued from the segmentation stage are then characterized by their color and texture, and afterwards identified by a probabilistic method in order to obtain their nature (grass, rocks, ground . . .).

1 Introduction

Notre étude se situe dans le cadre de la perception pour un robot mobile évoluant dans un environnement naturel. Dans ce cadre, la caméra est un capteur privilégié par la richesse des informations qu'elle peut fournir. Généralement, seule l'information tridimensionnelle est exploitée pour modéliser ce type d'environnement. Cependant, l'information de profondeur est insuffisante pour décrire correctement l'environnement. D'autres informations telles que la couleur et la texture doivent être prises en compte.

La connaissance de la nature des éléments constituant la scène est utile pour les tâches de navigation et localisation du robot. En effet, un objet ayant des caractéristiques colorimétriques et de texture peut être un bon amer naturel pour la tâche de localisation. D'un autre côté, connaître la nature des éléments de la scène est fondamental pour déterminer les zones navigables (en fonction des capacités du robot).

Le modèle nominatif de régions que nous proposons fournit cette connaissance. Chaque région de l'image est caractérisée par sa couleur et sa texture puis identifiée par association à l'une des classes prédéfinies dans une base de connaissances. Dans notre cas, nous avons défini cinq classes correspondant aux principaux éléments contenus dans la scène : herbe, terre, arbres, roches et ciel.

2 Construction du modèle

Cette approche se décompose en quatre phases successives (figure 1) : l'extraction des régions par segmentation couleur, la caractérisation de ces régions par leur couleur et leur texture, leur identification et enfin la vérification de la cohérence du modèle.

Ces quatre étapes se font avec des images de résolution différentes. L'étape de segmentation qui doit fournir rapidement une description synthétique de la scène, est réalisée avec une résolution faible (images de tailles 140x120). En revanche, la caractérisation des régions obtenues, en particulier pour les attributs de texture choisis, demandent de revenir à la pleine résolution (images 560x480). Finalement, le modèle fournit sera, quant à lui, à la même résolution que la segmentation.

L'identification nécessite une phase préalable d'apprentissage (supervisé) pour construire la base de connaissances. Elle est réalisée ici directement à partir de régions issues de la segmentation. Il sera donc possible, par la suite, d'envisager d'augmenter la base de connaissances dans la mesure où de nouvelles classes apparaîtraient.

Le modèle est construit en affectant chaque région à l'une des classes prédéfinies. Les régions identifiées de même nature et connexes sont fusionnées. À partir de simples règles sur la position des régions dans l'image et le contexte (environnement du robot) la cohérence du modèle est vérifiée.

Nous présentons donc à présent chacune de ces quatre phases de manière plus détaillée.

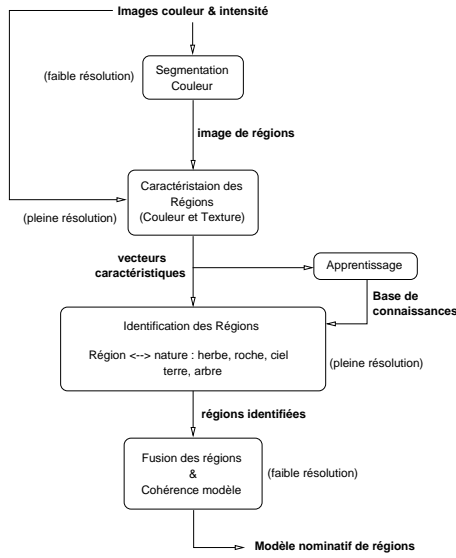


FIG. 1 — Construction du modèle nominatif de régions

2.1 Extraction des régions

Nous avons développé une méthode hybride de segmentation combinant les techniques de grossissement de régions et de *clustering* et qui profite des avantages liés à ces deux méthodes en limitant leurs inconvénients : elle est indépendante du point de départ de la segmentation ou de l'ordre d'adjacence des régions à fusionner et utilise un critère global de fusion (une préclassification des régions élémentaires est réalisée en étudiant l'allure générale des histogrammes).

Finalement, cette méthode, détaillée dans [4], permet d'obtenir une segmentation en grandes régions et de dégager, en grande partie, les principales composantes de la scène.

Différents espaces colorimétriques ont été testés : RVB, ITS, rvb, $I_1 I_2' I_3'$ [5] et ES [7]. L'étude montre que l'intensité ne peut pas être négligée durant la phase de segmentation (essentiellement pour les objets situés au loin où il y a un manque de saturation) et qu'il est plus intéressant d'utiliser des attributs non corrélés. Nous avons donc choisi l'espace $I_1 I_2' I_3'$ (dérivé de la transformation de Karhunen-Loeve) dont les équations sont :

$$I_1 = \frac{R+G+B}{3}, \quad I_2' = (R - B), \quad I_3' = \frac{2G-R-B}{2}$$

et qui s'est avéré le plus approprié pour le type de scènes traitées.

2.2 Caractérisation des régions

Chaque région de l'image est caractérisée par sa couleur et sa texture. Les opérateurs de texture sont basés sur l'analyse des lignes de contraste à l'intérieur de la région considérée et sont définis par les six paramètres suivants (calculés sur des sous-fenêtres 16x16 appelées *texel*) [2] :

1. **Densité** : elle précise le nombre de lignes de contraste contenues à l'intérieur du *texel*,

2. **Granularité** : ce paramètre quantifie la présence d'un grand nombre de points de gradient non concaténés dans le *texel*,
3. **Discontinuité** : elle indique le nombre de lignes discontinues dans le *texel*,
4. **Courbure** : elle mesure la densité des lignes de contraste non rectilignes dans le *texel*,
5. **Rectitude** : elle fournit le nombre de lignes droites dans le *texel*,
6. **Netteté** : une valeur élevée de ce paramètre indique que le *texel* présente des variations brusques d'intensité dans toutes les directions. Au contraire, une valeur faible indique que le *texel* présente des variations lentes d'intensité.

Finalement, toute région de l'image pourra être caractérisée à partir d'un vecteur contenant à la fois des informations sur sa couleur et sa texture. Ce vecteur contient les moyennes statistiques des six paramètres de texture définis précédemment, et les moyennes des attributs colorimétriques.

Une première étude a consisté à utiliser les attributs RVB pour caractériser la couleur d'une région [3]. Cependant, contrairement à la phase de segmentation, nous pensons qu'il est préférable de s'abstraire des problèmes de luminance. En conséquence, nous avons repris l'étude en conservant uniquement I_2' et I_3' . Nous montrerons un peu plus loin les premiers résultats obtenus.

2.3 Identification des régions

La nature des régions de la scène est obtenue en comparant les caractéristiques de la région à reconnaître à celle des différentes classes définies dans la base de données. La comparaison est effectuée en utilisant une technique probabiliste (classification bayésienne) [1, 6] avec pour estimation de la probabilité :

$$P(C_i | X) = \frac{P(X | C_i)P(C_i)}{\sum_{i=1}^n P(X | C_i)P(C_i)}$$

- $P(C_i)$ est la probabilité *a priori* sur la répartition des classes,
- $P(X | C_i)$ désigne la densité de probabilité sous-jacente à la distribution des observations provenant de la classe C_i .
- $P(C_i | X)$ représente la probabilité conditionnelle *a posteriori* pour que l'élément associé à X appartienne à la classe C_i .

Nous avons considéré que la probabilité *a priori* $P(C_i)$ était identique pour chacune des classes. Dans ce cas, le calcul de la probabilité *a posteriori* $P(C_i | X)$ est simplifié et se ramène uniquement au calcul de la probabilité conditionnelle $P(X | C_i)$. Cette dernière est estimée par la méthode du $k^{\text{ième}}$ plus proche voisin. Le calcul de $P(X | C_i)$ se ramène au calcul du volume englobant le $k^{\text{ième}}$ plus proche voisin. La seule variable intervenant dans le calcul de ce volume est la

distance du point X au $k^{\text{ième}}$ voisin choisi (X correspondant à la région à identifier et dont les coordonnées sont fournies par le vecteur d'attributs). Il suffit de calculer cette distance pour évaluer la probabilité $P(X | C_i)$ et donc $P(C_i | X)$. Finalement, l'observation X (soit la région) sera affectée à la classe C_i dont la distance au $k^{\text{ième}}$ voisin sera la plus petite.

Cette méthode est adaptée lorsque les classes ne sont pas bien séparées dans l'espace de classification. D'autre part, elle ne nécessite pas obligatoirement de connaître la loi de la densité de probabilité suivie par les classes.

2.4 Fusion des régions et construction du modèle

Chaque région de l'image est à ce stade assortie de sa nature. Bien que la segmentation fournisse de grandes régions, ces dernières ne correspondent pas obligatoirement aux grandes composantes de la scène. Une étape de fusion est donc nécessaire pour regrouper toutes les régions de même nature et connexes. À l'issue de cette étape de fusion, la cohérence du modèle est vérifiée. Celle-ci peut être étudiée en exploitant les caractéristiques de l'environnement dans lequel évolue le robot. Dans notre cas, nous pouvons remarquer que : le ciel et le sol doivent obligatoirement apparaître sur les images ; le sol ne peut être que plat ou en pente douce ; les arbres sont les objets les plus grands apparaissant dans nos scènes. Nous pouvons donc en déduire les règles suivantes :

1. une région de *terre*, d'*herbe* ou de *roche* ne peut pas être située entre du ciel et de l'arbre ;
2. une région *arbre* ne peut pas toucher le bord inférieur de l'image ni être entourée d'herbe ;
3. une région *herbe* ne peut pas toucher le haut de l'image ni être entourée d'arbre ;

La détection de ces erreurs nécessite de connaître uniquement des relations du type *au-dessus de*, *au-dessous de* et *autour de*. Ces relations sont obtenues en observant la position horizontale minimum et maximum des régions dans l'image. Le schéma 2 montre le raisonnement employé pour les déterminer. Soient y_{bottomR} et y_{topR} les positions minimales et maximales de la région R .

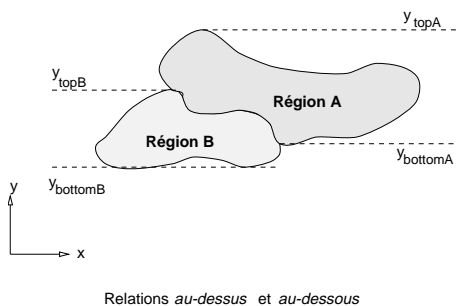


FIG. 2 — Principe du relevé de la position d'une région

- La région A sera dite *au-dessus de* la région B si $y_{\text{bottomA}} > y_{\text{bottomB}}$ et $y_{\text{topA}} > y_{\text{topB}}$.

- Réciproquement, la région B sera dite *au-dessous de* la région A si $y_{\text{bottomB}} < y_{\text{bottomA}}$ et $y_{\text{topB}} < y_{\text{topA}}$.
- Enfin la région A sera dite *autour de* la région B si la région B n'est en contact qu'avec la région A.

3 Résultats expérimentaux

Nous présentons dans le tableau 1 les résultats dans le cas où les attributs colorimétriques sont RVB. La base de connaissances contenait 135 régions prélevées sur 36 images. Nous avons effectué la reconnaissance sur 10 images différentes et avons obtenu un pourcentage de régions correctement identifiées de 78%.

	Nombre de régions N_{reg}	Nombre de régions reconnues $N_{Reconnues}$	pourcentage de réussite \mathcal{R}
Arbre	30	20	67%
Terre	4	3	75%
Ciel	14	13	93%
Herbe	15	12	86%
Roche	4	4	100%
Total	67	52	78%

TAB. 1 — Résultats de l'identification (première étude)

Les principales erreurs rencontrées au cours de l'identification sont du type *arbre* pris pour *roche* ou *terre* et *herbe* pris pour *arbre*. Le premier type d'erreurs est dû à des branchages clairsemés (beaucoup de ciel apparaît entre les branches) : en fonction de la quantité de branches dans la région (faible ou importante), l'identification va être *terre* ou *roche*. D'un autre côté, l'erreur *herbe* pris comme *arbre* provient du fait que l'herbe est très dense, très haute.

Ces erreurs peuvent cependant être détectées lors de la vérification de la cohérence du modèle. Il suffit alors de modifier les natures des régions mises en cause lors de la phase de vérification. Toute modification dans la nature d'une région entraînera la relance du processus de vérification. Ce dernier s'arrête une fois qu'il n'y a plus d'erreurs détectées.

La figure 6 montre le modèle obtenu pour l'image 3 : initialement la segmentation (figure 4) fournit 9 régions à identifier (3 régions dans le ciel, 2 régions d'arbre, 1 région de roche, 2 régions d'herbe, et une région non identifiable correspondant à un panneau solaire - indiquée par *rejet* sur les figures-). Lors de la phase d'identification (figure 4), toutes les régions identifiables sont correctement reconnues sauf une : la région de ciel à la frontière des arbres est identifiée *terre*. Après fusion et vérification du modèle (figure 5), cette erreur est détectée et remplacée par du ciel.

Cependant, l'environnement du robot était trop varié : arbres allant de la haie d'arbustes au platane ou au pin (cela se traduisait dans la base de connaissances par un petit nombre de différents types d'éléments pour une même classe). Nous



FIG. 3 — Image de la scène

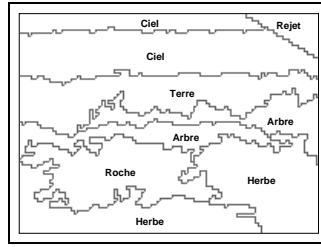


FIG. 4 — Segmentation et Identification

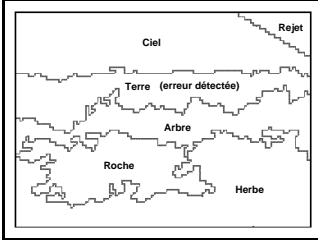


FIG. 5 — Fusion et Cohérence

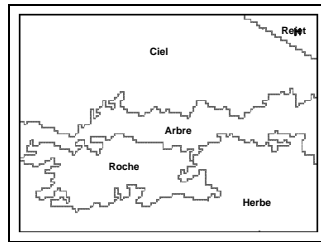


FIG. 6 — Modèle obtenu

avons donc repris l'étude sur un nombre réduit d'images plus homogènes. De nouveaux attributs colorimétriques, indépendants de la luminance ($I_2 I_3$), ont été introduits tandis que les mêmes attributs de texture ont été conservés. Seulement 8 images ont été utilisées pour la phase d'apprentissage. Bien entendu, cette deuxième étude doit être complétée en augmentant de manière significative la base de connaissances et le nombre d'images à identifier. Les premiers résultats montrent cependant que l'on peut espérer de bons résultats (figures 7 et 8).



FIG. 7 — Image de la scène

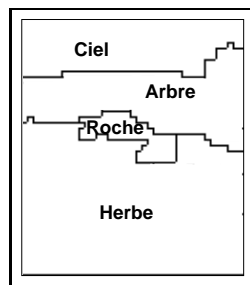


FIG. 8 — Modèle obtenu

4 Conclusion

L'information sémantique est une information riche aussi bien pour la navigation que pour la localisation du robot. Le modèle nominatif de régions que nous proposons s'appuie sur cette idée pour fournir une information sur la nature des éléments contenus dans une scène. Cette information est obtenue en utilisant des informations de couleur et de texture à différents niveaux de la construction du modèle (segmentation et caractérisation). Les attributs de texture ont été initialement choisis pour la possibilité qu'ils offriraient d'être réalisés sur carte câblée. Cependant, ils ont l'inconvénient d'être basés sur

l'étude des lignes de contraste et de dépendre donc fortement de la qualité de celle-ci (les contours doivent être différents sur chacune des classes). Il serait donc intéressant de tester la validité sur un plus grand nombre d'images d'une part, et d'étudier d'autres attributs d'autres part, afin de comparer les résultats et de dégager les plus pertinents pour notre étude. D'autres perspectives seraient d'associer un coefficient de confiance dédié à la segmentation et apportant la possibilité d'affiner celle-ci en fonction du résultat de l'identification... Enfin, de manière plus générale, notre objectif serait de faire coopérer ce modèle avec les représentations géométriques afin de dégager un modèle sémantique de la scène qui serait complété au fur et à mesure de nouvelles perceptions.

Références

- [1] H.C. Andrews. *Mathematical Techniques in Pattern Recognition*. Wiley-Interscience, 1972.
- [2] A. Casals, J. Amat, and A. Grau. Texture Parametrization Method for Image Segmentation. In *Second European Conference on Computer Vision, Santa Margherita Ligure (Italie)*, 1992.
- [3] P. Lasserre. Vision pour la robotique en environnement naturel. Thèse de l'Université Paul Sabatier, Toulouse (France), Laboratoire d'Analyse et d'Architecture des Systèmes, septembre 1996.
- [4] R. Murrieta Cid, P. Lasserre, and M. Briot. Color segmentation in principal regions for natural outdoor scenes. *Third Workshop on Electronic Control and Measuring Systems (Toulouse)*, juin 1997.
- [5] Y. Ohta. *Knowledge-Based Interpretation of Outdoor Natural Color Scenes*. Morgan Kaufman Publishers, Inc., Palo Alto, CA, 1985.
- [6] J. G. Postaire. *De l'image à la décision - analyse des images numériques et théorie de la décision*. DUNOD informatique, 1987.
- [7] E. Saber, A.M. Tekalp, R. Eschbach, and K. Knox. Automatic Image Annotation Using Adaptive Color Classification. *Graphical Models and Image Processing*, 58(2) :115–126, march 1996.