

Structuration d'image par utilisation de groupements perceptuels pertinents

Thierry Baron, Maurice Briot

Laboratoire d'Automatique et d'Analyse des Systèmes du CNRS
7, avenue du Colonel Roche
31077 Toulouse Cedex

Résumé

En robotique la problématique de la structuration d'image s'inscrit dans le contexte de l'interprétation de scène à partir de primitives images plus ou moins complexes. Les primitives sur lesquelles de nombreux travaux ont été réalisés sont principalement le point et le segment. Il faut toutefois constater que la description d'une image à partir de ces primitives n'est pas complète. C'est pourquoi il semble naturel d'évoluer vers une description de l'image sur la base de groupements perceptuels obtenus à partir de deux représentations duales de l'image : contours et régions. Cette démarche peut permettre d'envisager des applications de reconnaissance et d'identification beaucoup moins lourdes d'un point de vue algorithmique, puisque les primitives sont plus riches.

Introduction

La perception de l'environnement est l'un des problèmes majeur de la robotique autonome, dont le concept de base s'articule autour du paradigme "perception-décision-action". Dans ce contexte, le rôle de la perception est de fournir à un système le maximum d'information pour lui permettre de réagir en conséquence.

Ce domaine de recherche a permis de mettre en évidence nombre de techniques de traitement d'image pour extraire l'information des données issues de la perception sous forme de primitives bas-niveau telles que les segments, les courbes et les régions. Toutefois l'aspect élémentaire de ces primitives a orienté les travaux vers des descriptions mettant en jeu des relations entre primitives : graphes relationnels, groupements perceptuels, ...

Cette démarche reprend les capacités du système visuel humain à pouvoir déterminer dans une scène des structures qui aident à sa compréhension. Le nombre des percepts est relativement élevé pour la vision humaine et utilise la plupart des domaines de la vision par ordinateur tels que : les aspects purement géométriques (colinéarité, parallélisme, point de fuite, symétries,...), les aspects d'homogénéité des surfaces (couleur, texture), et ceux liés à l'éclairage d'une scène (réflectance).

Nos travaux sont liés plus particulièrement à l'utilisation de caméras achrome ou trichrome en robotique, et à l'extraction de primitives perceptuelles obtenues à partir de deux représentations duales de l'image : contours et régions.

Abstract

Image Structuration from features is essential for efficient scene interpretation tasks in robotics. Points and lines are the most commonly used features, but they don't provide a complete image description. Perceptual organisation, based on two dual image representations (regions and segments), is a natural evolution towards the improvement of recognition and identification tasks since the features are less ambiguous.

1 Groupements perceptuels

Le concept de groupement perceptuel présenté par Marr [Marr 82], a donné lieu à de nombreux travaux en perception. Son intérêt principal réside dans le regroupement de primitives de bas niveau au moyen de contraintes topologiques.

Les primitives les plus couramment employées sont obtenues à partir des contours extraits d'une image. Ces contours sont alors modélisés sous la forme de courbes et de droites. Avec ces deux types de primitives, on cherche à trouver des indices visuels caractéristiques qui seront identifiables facilement. Ces percepts sont généralement représentatifs des environnements créés par l'homme.

Les relations de base pour la formation d'indices sont principalement les intersections et les symétries locales image obtenues à partir des symétries locales 3D [Malik 85] [Lowe 85] [Mohan 89] [Heraud 90].

- les discontinuités (Cf. figure 1)
 - courbes et droites → discontinuité, tangence.
 - courbes → co-curvilinéarité.
 - droites → colinéarité.
- les symétries.

L'importance de ce type de description est de pouvoir mettre en avant les propriétés structurelle d'une scène et d'en faciliter ainsi sa compréhension. Toutefois les indices visuels que l'on peut utiliser sont fortement dépendant de l'environnement applicatif (Cf. paragraphe 2) et leur cohérence peut être envisagée sous différents angles (Cf. paragraphes 3 et 4).

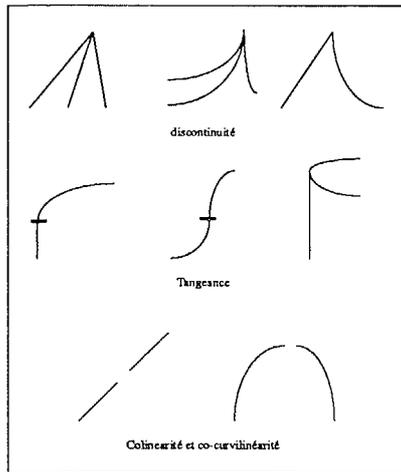


Figure 1 : Discontinuités d'après [Heraud 90]

2 Prise en compte de l'environnement

Dans le cas de scènes d'intérieur on peut noter que la plupart des objets sont constitués de segments de droites ce qui a incité de nombreux auteurs à utiliser la primitive segment pour obtenir une information synthétique des contours en milieu structuré.

Sur la base de cette description de l'image, de nombreux algorithmes ont été développés : la stéréovision passive binoculaire ou trinoculaire, l'estimation du mouvement, la reconnaissance d'objets à partir d'un modèle. Leur principe repose sur la prédiction-vérification d'hypothèses et nécessite l'utilisation de nombreuses contraintes pour limiter le nombre d'appariements potentiels et éviter les faux appariements.

Il est à noter qu'une partie de ces contraintes est imposée par la pauvreté de la primitive segment lorsqu'elle est prise hors contexte. C'est pourquoi le regroupement de segments en primitives plus élaborées, grâce à l'utilisation d'indices image, conduit à une description structurée de l'image beaucoup plus synthétique.

Il est important de pouvoir définir dans un environnement donné quels sont les percepts les plus représentatifs du contexte. Le choix retenu est composé de trois primitives contextuelles :

- *Méta-segment ou ensemble de segments colinéaires.* Leur recherche est importante car en général l'approximation polygonale fournit un sur-découpage des chaînes de contour.
- *Groupes de segments ou méta-segments parallèles.* On appellera rubans les groupes de segments localement proches.
- *Lignes brisées ou polygones.*
Ces groupes sont relatifs aux objets composants une scène.

Les primitives une fois définies, il faut pouvoir leur donner une consistance qui est liée à l'image. C'est pourquoi nous définissons le concept de cohérence image.

3 Cohérence image

Les groupements n'ont de pertinence que s'ils appartiennent aux frontières d'un même objet. Cette relation d'appartenance peut être obtenue par l'exploitation de l'information contenue dans l'image achrome ou trichrome qui a permis d'obtenir les segments.

Pour grouper ces derniers, nous avons choisi d'utiliser une segmentation en régions avec prise en compte des lignes de contours (Cf. figures 2 et 4). Les segments sont alors regroupés par appartenance à une même région (Cf. figures 3 et 5).

L'obtention de cette information permet de vérifier la consistance de chaque segment par :

soit S_i un segment, alors S_i est consistant

ssi S_i est bordé par uniquement deux régions (R_a, R_b)

Cette relation revient à dire que tout segment S_i ne possède pas de point de jonction ailleurs qu'aux extrémités.

Pour obtenir la cohérence image, on va rendre consistant, dans une première étape, tous les segments de l'image, et ensuite les regrouper avec le critère :

S_i appartient à $R_a \iff S_i$ borde R_a

On définit ainsi une structuration de l'image cohérente en terme de groupes de segments liés à des régions.

L'association de l'ensemble des concepts de cohérence contextuelle et image permet d'établir une description synthétique de l'image sous forme de primitives de haut niveau. Alors la formalisation d'algorithmes de mise en correspondance de segments peut être envisagée avec l'utilisation de ces nouvelles primitives d'image.

Le positionnement de chacune de ces structures les unes par rapport aux autres, peut être envisagé afin de décrire l'image sous forme d'un graphe relationnel utilisant les relations suivantes pour les régions :

- adjacence
- inclusion
- contiguïté

L'intérêt majeur de cette démarche réside dans l'utilisation de primitives déjà structurées ce qui réduit de manière importante la taille du graphe relationnel associé à l'image. La potentialité des graphes pour la mise en correspondance est bien connue depuis les travaux de Heraud et Skordas [Skordas 88a] [Skordas 88b].

Les primitives formées possèdent cependant des ambiguïtés, quant à l'appartenance d'un segment à une surface 3D correspondant à une région, qui ne peuvent être levées que par la stéréovision comme nous allons le voir dans le paragraphe suivant.

4 Cohérence Stéréoscopique

Les segments d'une image peuvent être répartis suivant trois catégories :

- Arête géométrique
 - + discontinuité de normale
→ le segment est une arête $S_i \in (R_a, R_b)$.
 - + discontinuité de profondeur
→ le segment est une frontière $S_i \in R_a$.

- segment d'occlusion \in un autre objet.
- segment photométrique (ombre portée, segment pictural).

La distinction entre ces trois groupes ne peut être faite a priori et nécessite l'utilisation de l'information additionnelle apportée par la stéréovision.

Dans l'hypothèse d'un environnement polyédrique la contrainte naturelle sera la coplanarité des segments appartenant à une même structure. Cette contrainte forte permet de pouvoir détecter les segments d'occlusion qui ont été attribués à une entité à laquelle ils n'appartiennent pas, et de différencier les discontinuités de normale de celles de profondeur.

5 Contraintes pour la mise en correspondance

Nous présentons ici, un ensemble de contraintes qui peuvent être utilisés pour réaliser de la mise en correspondance de régions, ou de segments dans le cas général.

En stéréovision la contrainte épipolaire est la plus importante. Elle permet de pouvoir sélectionner efficacement parmi les segments de l'image à apparier un nombre réduit de candidats potentiels. Cependant dans le cas d'une paire stéréoscopique, seule une technique de prédiction-vérification d'hypothèses appliquée globalement sur l'image permet d'apparier les segments de manière cohérente.

Avec la description région-segment, on ajoute une contrainte forte, la région. La cohérence de la mise en correspondance est alors liée aux caractéristiques de la région.

Sa forme détermine la topologie des segments qui constituent sa frontière. On peut alors se servir de cette topologie pour déterminer les appariements qui vérifieront au mieux cette contrainte, au moyen d'un algorithme de cliques maximales [Horaud 89] parmi l'ensemble des appariements potentiels obtenus avec la contrainte épipolaire.

Toutefois, l'information amenée par la région est plus riche, et permet de pouvoir limiter avant l'utilisation des cliques maximales le nombre de régions potentiellement compatibles. Il est cependant nécessaire de connaître les critères utilisables et leur domaine de validité.

Les caractéristiques retenues généralement sont :

- *Surface* :
L'utilisation de cette information est délicate, car elle est sensible aux occlusions et à la variation de point de vue entre deux caméras. Elle permet cependant de faire un tri en supprimant les régions trop différentes.
- *Centre de gravité* :
Les centres de gravités permettent dans le cas des régions totalement observées de vérifier la contrainte épipolaire (Cf figures 6 et 7) et de donner ainsi une indication fiable.
- *Couleur* :
La couleur est un attribut très discriminant lorsque l'on utilise l'information de chrominance contenue dans les systèmes colorimétriques I_{rv}^1 ou I_{ts}^2 [Coutance 91].

- *Intensité* : L'intensité est porteuse d'une information riche mais complexe [Baron 90]: forme 3D, position de la source lumineuse ("shape from shading"). On peut toutefois l'utiliser de la même manière que le critère de surface.

Cette ensemble de contraintes permet de pouvoir limiter le nombre de régions candidates à un appariement, et ainsi de fiabiliser les mises en correspondances de segments permettant d'obtenir une représentation 3D d'une scène.

Conclusion

Cette communication présente la possibilité d'utiliser les descriptions duales d'une image (régions et segments) pour réaliser une structuration monoculaire et stéréoscopique. L'intérêt de grouper ces primitives d'image vient de la quantité d'information qui en résulte, et la possibilité de pouvoir donner des primitives moins ambiguës à des tâches de haut-niveau qui peuvent s'inscrire dans les domaines de la reconnaissance ou de perception de l'environnement d'un robot mobile.

Bibliographie

- [Baron 90] T. Baron et P. Grandjean. *A criterion for region matching based on specular reflectance*. Rapport Interne, Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Toulouse (France), 1990.
- [Coutance 91] V. Coutance. *La couleur en vision par ordinateur, application à la robotique*. Thèse de docteur ingénieur, Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Toulouse (France), janvier 1991.
- [Horaud 89] R. Horaud et T. Skordas. *Stereo Correspondance Through Feature Grouping and Maximal Cliques*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(11):1168-1180, Novembre 1989.
- [Horaud 90] R. Horaud et F. Veillon. *Finding Geometric and Relational Structures in an Image*. In 1st European Conference on Computer Vision, Antibes (France), Avril 1990.
- [Lowe 85] D.G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.
- [Malik 85] Jitendra Malik. *Interpreting Line Drawings of Curved Objects*. Rapport technique, Stanford University, Department of Computer Science, DEC 1985.
- [Marr 82] D. Marr. *Vision*. W.H. Freeman and Co., 1982.
- [Mohan 89] R. Mohan et R. Nevatia. *Segmentation and Description Based on Perceptual Organisation*. In IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, California (USA), pages 333-341, Juin 1989.
- [Skordas 88a] T. Skordas. *Mise en correspondance et reconstruction stéréo utilisant une description structurelle des images*. Thèse de docteur ingénieur, Institut National Polytechnique de Grenoble, Mars 1988.
- [Skordas 88b] T. Skordas et R. Horaud. *Mise en Correspondance de Structures 2-D dans une Paire d'Images Stéréoscopiques*. In 2ème Atelier Scientifique : "Traitement d'Images : du Pixel à l'Interprétation" (TIPI), Aussois, pages III-1 - III-23. C.N.R.S., Avril 1988.

¹ I_{rv} : Intensité, rouge et vert normalisés

² I_{ts} : Intensité, teinte et saturation

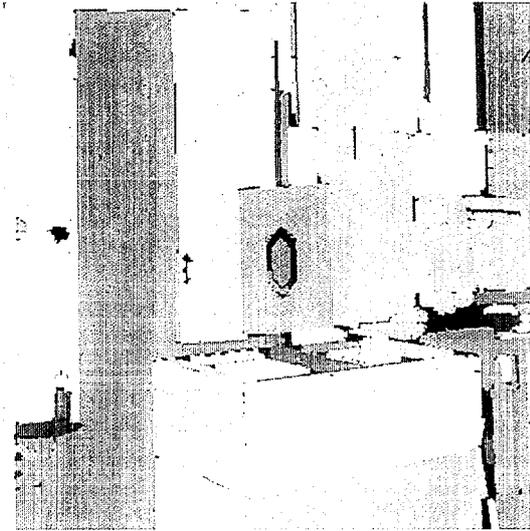


Figure 2 : Segmentation en régions (image droite)

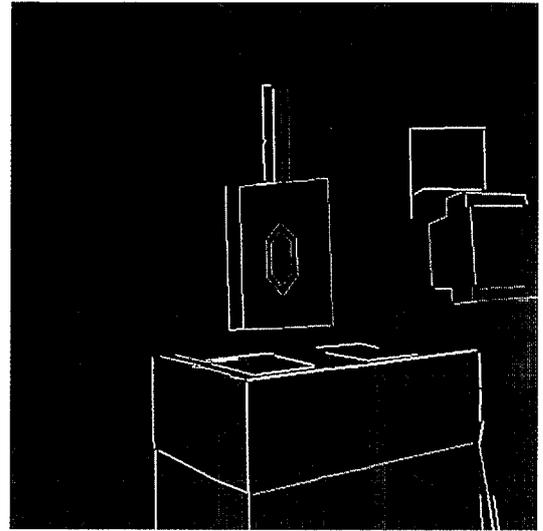


Figure 3 : Groupements des segments sur la base des régions de la figure 2

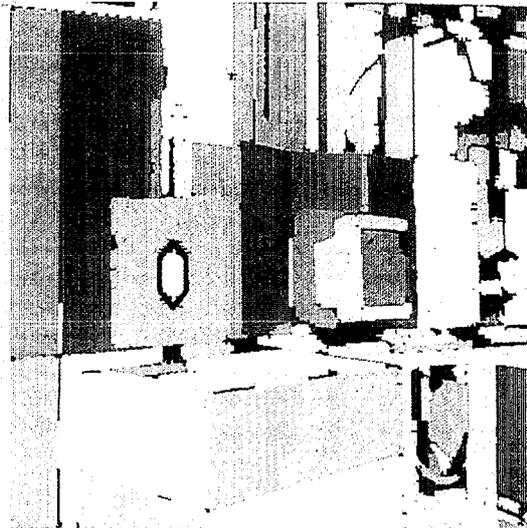


Figure 4 : Segmentation en régions (image gauche)

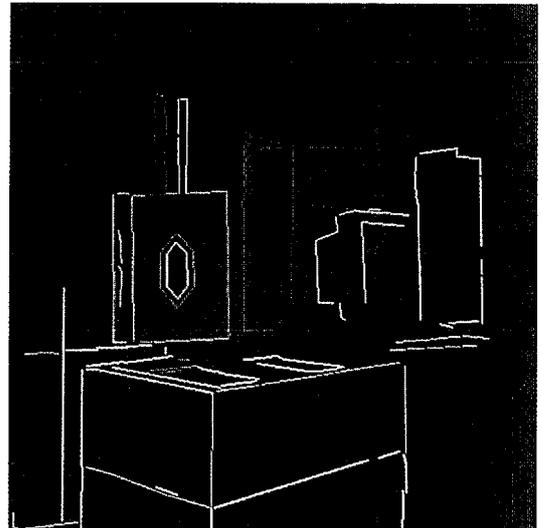


Figure 5 : Groupements des segments sur la base des régions de la figure 4

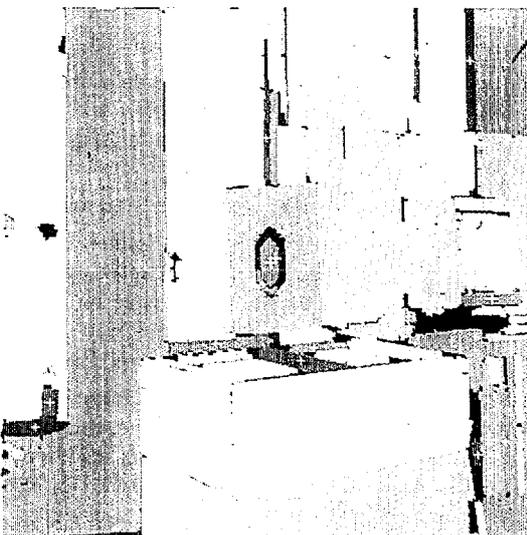


Figure 6 : Régions et centres de gravités (en jaune les centres donnant lieu aux lignes épipolaires de la figure 7)

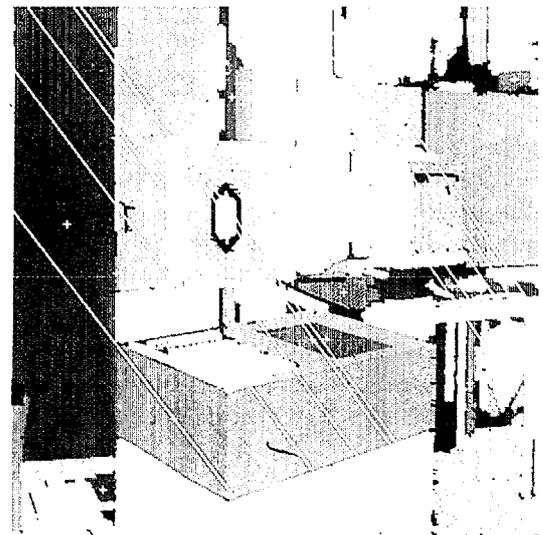


Figure 7 : Région, centres de gravité et lignes épipolaires