

Vers une Méthode d'Evaluation Comparée de Mécanismes de Regroupements de Pixels : Application à la Décomposition en Composantes Connexes et à la Croissance de Régions pour l'utilisation dans un Système de Vision Robotique Temps Réel

Patrick BONNIN^{1,2}, Olivier STASSE^{2,1}, Vincent HUGEL¹, Pierre BLAZEVIC¹ Maryline SILLY³

¹Laboratoire de Robotique de Versailles LRV
10-12 Av de l'Europe, 78 140 Vélizy, France

²Laboratoire Traitement et de Transport de l'Information L2TI
Av JB Clément, 93 430 Villetaneuse, France

³Institut de Recherche en Informatique de Nantes IRIN
Rue Christian Pauc, La Chantrerie, BP 50609,44 306 Nantes Cedex 03, France
bonnin@robot.uvsq.fr, stasse@l2ti.univ-paris13.fr, hugel@robot.uvsq.fr
pierre@robot.uvsq.fr, maryline.silly@iut-nantes.univ-nantes.fr

Résumé – Nous proposons une méthode de comparaison et d'évaluation d'algorithmes de regroupements rapides de pixels pour l'extraction de composantes connexes et la croissance de région couleur. L'évaluation est basée sur deux critères : la rapidité de traitement et la qualité des résultats pour des applications robotiques.

Abstract – We propose a method to compare and to evaluate the algorithms of fast pixel gathering mechanisms for the connected component extraction and the color region segmentation. The evaluation is based on two criteria: the speed of processing and the quality of the results for robotic applications.

1 Introduction

La rapidité est une propriété essentielle pour un algorithme de segmentation d'image que l'on cherche à utiliser dans un système de vision temps réel monté sur un robot mobile. En effet, obtenir avec un tel algorithme une cadence de traitement le plus proche possible de la cadence vidéo (25 Hz) est impératif dès que le robot mobile doit suivre un indice visuel de son environnement au cours de son déplacement, d'autant plus si le robot est un robot à pattes, de type quadupède ou bipède. La robustesse nécessaire de l'algorithme va dépendre de l'application et notamment des conditions d'éclairage, si celles-ci sont ou non fluctuantes au cours du temps.

La contrainte d'embarcabilité du système informatique est également à prendre en compte pour un robot mobile autonome. Elle implique généralement une puissance de calcul embarquée restreinte, partagée entre le système de vision mais également le système de contrôle commande du robot (du bas niveau : contrôle des moteurs, au haut niveau : stratégie d'accomplissement de la tâche).

Compte tenu qu'un système de vision doit équiper différents robots en projet dans le laboratoire pour assurer diverses tâches, il n'est pas envisageable de rechercher un système de vision unique, mais une approche méthodologi-

que s'adaptant aux différents cas de figure. Ainsi, nous recherchons différents opérateurs, avec des caractéristiques différentes en terme de puissance de calcul nécessaire, et de qualité des résultats. L'évaluation comparée de ces divers opérateurs en fonction des deux critères précédents est un point clef de notre recherche. Elle permettra de réaliser l'Adéquation entre l'opérateur de traitement d'image, le système informatique du robot et l'application envisagée.

Le système de vision pourra être classique, c'est-à-dire figé : une chaîne d'opérateurs de traitement unique est choisie à la conception, ou dynamique. Dans ce dernier cas nous nous intéressons à la tolérance aux fautes, c'est-à-dire à la possibilité de mettre en oeuvre une procédure de secours au cas où les informations de la procédure principale ne seraient pas délivrées à temps [1] et deviendraient par conséquent obsolètes. La rapidité est la caractéristique essentielle de la procédure de secours.

2 Les Mécanismes de Regroupement de Pixels

Les deux tâches essentielles, confiées au Système de Vision pour leur partie extraction des informations sont - d'une part la navigation, - d'autre part l'évitement d'obstacles.

C'est le cas par exemple pour l'application RoboCup [2], compétition internationale www.robocup.org, où les robots quadrupède AIBO jouent au football en équipe, ainsi que pour le projet RNTL CLEOPATRE, labellisé en 2001 www.cleopatre-project.org, pour l'équipement d'un robot à roues de type convoyeur industriel assurant le transport de charges entre ateliers. Dans les deux applications, l'environnement est rendu coopératif par l'adjonction de balises visuelles artificielles colorées. Compte tenu des impératifs temps réel, nous nous limitons à l'hypothèse suivante : le sol est de couleur(s) connue(s), et toute couleur autre au niveau du sol (la calibration 3D de la caméra est connue) est considérée comme un obstacle.

Ainsi, l'information visuelle à extraire de l'image est de type *Région Colorée* affectée d'attributs sémantique : couleur, et topologiques : surface, centre de gravité boîte englobante. La primitive image *Région Colorée* peut être considérée alors une primitive *Région MultiSpectrale* [2], car l'image couleur est composée de trois composantes spectrales Rouge, Verte et Bleue. Comme en Traitement d'Image classique, on dénombre dans la littérature en fusion de données de type image, trois niveaux de traitements clairement établis : bas, intermédiaire et haut, correspondants parfaitement [3]. Le bas niveau, appelé également *Fusion Pixel* est clairement le niveau de fusion le plus rapide [4]. On retrouve à ce niveau de traitement les algorithmes de classification, c'est-à-dire qu'en fonction de ses valeurs dans les différentes bandes spectrales, un pixel se voit affecter aucune, une ou plusieurs (possible mais dangereux) classes. Il suffit ensuite de regrouper entre eux les pixels adjacents affectés d'une même classe par une procédure de Décomposition en Composantes Connexes (DCC) sur une seule image mono spectrale résultant de la classification. Ainsi, il est possible d'obtenir rapidement des primitives *Régions Colorées* à partir d'une classification dans les espaces colorimétriques RVB, YUV ou autre, suivie d'une décomposition en Composantes Connexes.

En résumé, la rapidité de ce traitement : classification suivi d'une DCC provient :

- d'une part de la fusion de données à bas niveau par l'opérateur de classification. En effet, la DCC porte sur une seule image et non trois,
- du fait qu'une faible proportion des pixels de l'image seront classifiés. En effet, si l'on décide de ne pas extraire le fond (pixels non classifiés), la DCC ne sera que plus rapide.

Conformément à notre méthodologie d'implantation [5] permettant une optimisation temporelle, ces deux procédures sont regroupées au sein d'un même parcours de l'image.

L'inconvénient majeur de ce traitement provient de la procédure de classification. Celle-ci est réalisée à l'aide d'un algorithme de *multi seuillage* qui requiert un certain nombre de paramètres fixés a priori. Un tel algorithme n'admet quasiment pas de variations des conditions d'éclairage. Par conséquent il est quasiment inutilisable en environnement extérieur, où les conditions fluctuent au cours de la journée, sauf si les couleurs sont très discernables dans l'espace couleur RVB, YUV ou autre.

Seule, une vraie segmentation en régions multi spectrales [2] [6] est à même de s'adapter aux variations des conditions d'éclairage. En effet la segmentation obtenue, c'est-à-dire les régions colorées en terme de *regroupement de pixels* est quasi indépendante, dans une plage admissible, des conditions d'éclairage. Seuls les attributs photométriques des régions vont varier. Il suffit donc d'étudier expérimentalement leur évolution en fonction de l'éclairage pour une sémantique en terme de couleur donnée.

La décomposition en composantes connexes, comme la segmentation par croissance de régions sont basées sur des mécanismes de regroupement de pixels [7]. La différence essentielle est que pour une DCC tous les pixels regroupés possèdent une même sémantique ou en d'autres termes sont codés à même niveau de gris dans une image monochrome, alors qu'une région monospectrale ou multispectrale regroupe un ensemble de pixels d'intensité en monospectrale ou d'intensités en multispectral légèrement différentes. Dans le cas d'une segmentation en régions, les paramètres de contrôle portent sur les différences d'intensités locales admissibles. Ainsi, une recherche portant sur *Les Mécanismes Rapides de Regroupement de Pixels* est pour nous fondamentale.

Paradoxalement, cette problématique est peu étudiée par les chercheurs en Traitement d'Image de nos jours, car les ordinateurs fixes sont puissants et des temps de traitements de l'ordre de grandeur de quelques minutes sont admissibles. On retrouve en revanche cette problématique dans les travaux de années 60 et 70 : [8], [9], [10] etc ... très certainement car la puissance disponible dans ces années là pour des temps de traitements admissibles de l'ordre de la minute voire de l'heure sur des gros ordinateurs est voisine de celle dont on dispose aujourd'hui dans un système de vision temps réel embarqué devant fonctionner à la cadence vidéo de 25 Hz, sur un ordinateur classique.

3 Vers une Méthodologie d'Evaluation

3.1 Evaluation en Traitement d'Images

L'évaluation de la qualité des résultats des opérateurs de traitement d'image a fait l'objet de nombreux travaux dans la communauté internationale et nationale. L'ouvrage collectif *Analyse d'Images : Filtrage et Segmentation* [11] lui consacre deux chapitres.

La difficulté essentielle est de trouver une segmentation de référence, a priori correcte. La méthode généralement utilisée consiste à partir d'une image parfaite de synthèse, dont la segmentation (image des étiquettes par exemple) ne pose aucun problème et constituera la référence. L'image est ensuite bruitée. La segmentation obtenue par l'opérateur à tester sur cette image est comparée à la segmentation de référence [12]. Le score attribué est global sur l'image, résultant de comparaisons locales pixel à pixel entre les segmentations. L'évaluation porte uniquement sur la qualité de la segmentation résultante en terme d'image : le pixel a-t-il été correctement regroupé, par rapport à la segmentation de référence ?

3.2 Evaluation en Vision Robotique

En vision robotique, la problématique est différente. En effet, si le temps de traitement est trop long, l'information délivrée, quelle que soit sa qualité est obsolète et le robot ne peut pas accomplir sa tâche. Par exemple, la primitive visuelle à suivre va sortir du champ de vue, ce qui rend tout asservissement visuel impossible. D'autre part, l'information utile pour le robot est plus globale, elle se situe non pas au niveau du regroupement correct des pixels, mais au niveau des attributs exploitables des primitives issues de l'image. Par exemple, le centre de gravité d'une région peut être extrait avec suffisamment de précision c'est-à-dire à un ou deux pixels près avec une segmentation dont les frontières de régions ne sont pas précises au niveau de la qualité évoquée ci-dessus.

De plus, l'évaluation servira :

- d'une part, comme en traitement d'image pour comparer les algorithmes,
- mais, d'autre part pour réaliser l'Adéquation entre un algorithme et la puissance disponible du système embarqué, soit statiquement dans la plupart des cas, soit dynamiquement en fonction de **l'état instantané du système informatique**, pour un système de vision dynamique.

La difficulté essentielle est la détermination d'une procédure de secours, pour laquelle la robustesse va être sacrifiée aux dépens de la rapidité.

3.3 Mise en Oeuvre du Critère de Rapidité

Les tests temporels sont effectués sous Linux, sur une machine de puissance connue. Le système d'exploitation garantissant la précision à 1/100ème de seconde [13], nous effectuons de 100 à 1000 itérations des opérateurs, dont le temps de traitement sur des images de faible résolution (ex : 176 colonnes x 144 lignes) est de l'ordre de la milliseconde. Nous effectuons également des moyennes de tests sur la même image, pour tester l'influence de l'état du système informatique, ainsi que sur des images différentes. Nous effectuons aussi des tests à diverses résolutions. Cette batterie de tests nous permet ensuite d'extrapoler à d'autres machines.

Le choix des images de test est très important. En effet, si les variations de temps de traitement ne sont pas trop importantes d'une image à l'autre pour une segmentation en régions, car tous les pixels sont traités, il n'en est pas de même pour une classification suivie d'une décomposition en composantes connexes. En effet, la décomposition en composantes connexes n'affectera que les pixels classifiés.

Ainsi, l'image test doit nous permettre d'évaluer le **Pire des Cas**, surtout si l'on cherche à évaluer une procédure de secours. Mais ce pire des cas ne doit pas être trop éloigné de la réalité, sinon il conduirait à un sur dimensionnement du système informatique du robot, donc à un coût plus élevé ou une dimension du robot plus importante. Les indices visuels étant de taille modeste dans l'environnement du robot, nous choisissons des images des

images dans lesquelles l'indice recherché couvre environ 50% de l'image, c'est-à-dire que nous plaçons la caméra proche de l'indice visuel pour la prise de vue.

3.4 Mise en Oeuvre du Critère de Qualité

Pour la décomposition en composantes connexes, il existe une seule solution exacte, donc l'évaluation en terme de regroupements corrects est aisée. En revanche, le problème est plus complexe pour l'évaluation d'une segmentation en régions, où la référence est difficile à obtenir (cf § 3.1).

Le problème sous jacent est d'une manière générale l'influence de l'ordre d'examen des pixels puis des régions à fusionner sur le résultat de la segmentation. La segmentation de référence est obtenue par un algorithme quasi indépendant de cette influence, ce que nous avons démontré dans [7]. Les caractéristiques essentielles de cet algorithme est d'être guidé par les données, c'est-à-dire sans parcours rigide a priori. La fusion est locale et isotrope couronne par couronne autour de régions embryons, et sans remise à jour des paramètres globaux d'homogénéité en cours de croissance. Malheureusement, cet algorithme est *trop gourmand* en terme en ressources calculatoires.

Une seconde évaluation de la qualité, plus en adéquation avec les contraintes de la robotique est d'évaluer les différences entre les attributs des primitives des segmentations de référence et à tester : centre de gravité, surface, boîte englobante, qui seront utilisées pour le calcul des paramètres 3D nécessaires au robot.

3.5 Vers une Hybridation des Algorithmes des Opérateurs

L'opérateur évalué n'est pas considéré comme une "boîte noire". Au contraire, nous détaillons les caractéristiques de l'algorithme de manière qualitative et/ou même quantitative en introduisant des variables statistiques dans le code source, en terme de : nombre d'accès pixel pour le traitement du pixel courant, nombre d'accès à la structure de données, nombre de réétiquetages nécessaires etc ... Le but est de tenter d'établir des corrélations entre ces propriétés et les résultats de l'évaluation sur les plans de la rapidité et de la qualité.

Ces corrélations nous permettrons de mettre en avant les caractéristiques essentielles, et si possible de proposer de nouveaux opérateurs hybrides.

4 Evaluations Réalisées

4.1 Trois Mécanismes Évalués

Ce travail est en cours. Pour l'instant nous avons évalué trois mécanismes. Les deux premiers sont bien connus, il s'agit:

- de celui proposé par A.Rosenfield et JL Pfalz [8] selon le balayage vidéo et le voisinage passé 4 connexe,
- de celui par plages utilisé dans la bibliothèque CM

Nous avons également testé un mécanisme plus complexe comportant un regroupement pyramidal à deux niveaux, puis un regroupement parallèle et isotrope, à partir de composantes prises successivement comme embryon, similaire à la croissance de régions proposée dans [7]. La partie pyramidale consiste à diviser l'image en voisinages de 3x3 pixels, puis à tenter de regrouper les points autour du point central. Le principe du second niveau est identique, mais sur les voisinages précédents. Il consiste à regrouper les 9 voisinages 3x3 du voisinage 9x9 autour du voisinage 3x3 central.

4.2 Résultats d'Evaluation

Concernant la décomposition en composantes connexes, pour des résultats rigoureusement identiques, le regroupement de pixels par plages s'avère de 20 à 30% plus rapide que le regroupement 4 connexe. Le troisième regroupement, moins adaptatif, donne des résultats en image (cf 3.1) légèrement moins bons. De plus, nécessitant des réétiquetages intermédiaires lors du regroupement parallèle, il s'avère être de deux à trois fois plus long. En revanche, le regroupement pyramidal à deux niveaux est rapide.

Pour la Croissance de Régions, les conclusions des résultats temporels sont analogues. En revanche les résultats de segmentation sont meilleurs pour le troisième regroupement, les deux premiers générant des anisotropies de segmentation liés au balayage vidéo.

4.3 Suite des Travaux

Nous travaillons actuellement sur l'hybridation des mécanismes précédents. Le regroupement initial est le regroupement pyramidal à deux niveaux, qui à le double avantage d'être - rapide, et - isotrope. Ensuite, nous adaptons le balayage vidéo pixel des deux autres méthodes au parcours des regroupements 3x3 et/ou 9x9.

5 Conclusion

Nous avons proposé une méthode d'évaluation comparée de mécanismes de regroupements de pixels appliquée à la Décomposition en composantes connexes et à la croissance de régions. Cette méthode a été appliquée sur plusieurs regroupements.

Nous sommes entrain d'analyser les corrélations entre les caractéristiques des algorithmes et les résultats de l'évaluation, ainsi que de proposer un nouvel opérateur hybride.

References

[1] M. Silly, T. Garcia, F. Grelier *Open source components for embedded real time applications*. IEEE Int Conf on Control and Automation, Xiamen, Chine, Juin 2002.

[2] P. Bonnin *Vers une Méthode de Conception et de Réalisation de Système de Vision Temps Réel pour la Robotique Mobile*, Habilitation à Diriger les Recherches de l'Université de Versailles Saint Quentin soutenue le 25 janvier 2000

[3] M. Mangolini, *Apport de la fusion d'images satellitaires multicapteur au niveau pixel en télédétection et photointerprétation*. PhD thesis, Nice Sophia Antipolis University, 1995

[4] L. Sliwa, *Optimisation d'une Configuration Multicapteur donnée Fusion Pixel*, Thèse de doctorat soutenue le 14 février 96 à l'Université Paul Sabatier

[5] P. Bonnin, L. Cabaret, V. Hugel, P. Blazevic, N. M'Sirdi, P. Coiffet, *Méthodes d'Optimisation Temporelle d'Algorithmes de Vision*, 18 ème Colloque GRETSI, 10-13 Septembre 2001 Toulouse, pp 597-600

[6] P. Bonnin, L. Cabaret, V. Hugel, P. Blazevic, N. M'Sirdi, P. Coiffet, *Exemple de Conception et de Réalisation d'un Système de Vision pour la Robotique Mobile et Autonome*, Int. Conf on Image and Signal Processing ICISP 2001, Agadir Maroc 3-5 Mai 2001.

[7] P. Bonnin, *Méthode Systématique de Conception et de Réalisation d'applications en vision par ordinateur*, Thèse de Doctorat de l'Université Paris VII soutenue le 12 décembre 91

[8] A. Rosenfield, J.L. Pfalz, *Sequential operations in digital picture processing*, Journal of ACM, vol 13 n°4 1966.

[9] S.L. Horowitz, T. Pavlidis, *Picture segmentation by a directed split-and-merge procedure*, Actes de la 2ème Int. Joint. Conf. Pattern Recognition 1974 pp 424-433

[10] S.W. Zucker, *Survey: Region Growing: Childhood and Adolescence*, Computer Graphics and Image Processing vol 5 1976 pp 382-399

[11] J.P. Cocquerez et S. Philipp, *Analyse d'Images : Filtrage et Segmentation*, Ouvrage Collectif coordonné par Travaux du Groupe 134 de Recherche (CNRS), Editeur Masson 1995

[12] R. Kara-Falah, *Segmentation d'Images : Coopération, Fusion, Evaluation*, Thèse de Doctorat de l'Université d'Annecy soutenue le 1er Juin 95, sous la Direction de P.Bolon

[13] C. Blaess, *Programmation Système en C sous Linux*, Editions Eyrolle

[14] J. Bruce, T. Balch, M. Veloso, *Fast and Inexpensive Color Image Segmentation for Interactive Robots*, IROS 2000