MISE EN CORRESPONDANCE DANS UNE SÉQUENCE BINOCULAIRE: UNE APPROCHE COALESCENCE NON SUPERVISÉE

K.KAOULA* **M.BENJELLOUN*** **B.DUBUISSON****

* Groupe Image-DSR
Institut National des Télécommunications
91011 Evry

** Heudiasyc U.R.A. 817 C.N.R.S
Université de Technologie de Compiègne
60200 Compiègne

RÉSUMÉ

Nous présentons dans cet article une nouvelle approche à la vision dynamique binoculaire. Les étapes de segmentation de mouvement et de mise en correspondance stéréoscopique dans une séquence binoculaire sont perçus comme des problèmes de classification automatique non supervisée. Ces deux processus d'habitude séparés, peuvent fusionner en une unique opération. La stratégie d'appariement et de segmentation utilisée ici est basée sur le principe de la *coalescence* automatique. Néanmoins, les problèmes spécifiques de la vision dynamique binoculaire nécessitent une parfaite adaptation de cette stratégie.

ABSTRACT

We present in this paper a new approach to binocular dynamic vision. The motion segmentation and the stereoscopic matching are seen as classification problems. They are achieved with a statistical clustering strategy. These two usually separated tasks can be processed simultaneously. However, the particular problems of the binocular dynamic vision need a perfect adaptation of our clustering strategy.

1. INTRODUCTION

L'approche décrite dans ce papier consiste à traiter les processus de segmentation spatio-temporelle et de mise en correspondance stéréoscopique dans une séquence binoculaire d'images avec une stratégie de classification automatique non-supervisée. Pourquoi une telle approche? Les méthodes usuellement utilisées en vision dynamique arrivent difficilement à maîtriser la variabilité qui caractérise foncièrement les primitives spatio-temporelles extraites de la séquence d'images. Ceci se traduit d'une part, par des problèmes d'interprétation de la scène observée ou d'estimation de mouvement et d'autre part, par la nécessité d'hypothèses fortement restrictives sur la nature du mouvement.

Les méthodes de reconnaissance des formes en général et les techniques de classification en particulier ont montré dans de très nombreux domaines leur capacité naturelle à s'accommoder de la variabilité des primitives. Celle-ci n'apparaît plus comme un inconvénient mais est intégrée au coeur même de leur mode de fonctionnement. Il serait donc intéréssant de voir ce que donnerait l'application d'une technique de classification automatique. La vision dynamique nécessite l'étude de quelques points importants:

- Le niveau sémantique retenu doit permettre une description robuste et suffisamment discriminante des classes.
- L'espace de représentation dans lequel fonctionnent les techniques de reconnaissance des formes statistique doit être parfaitement robuste et adapté aux problèmes spécifiques de la vision dynamique binoculaire.
- La configuration du dispositif d'acquisition doit être déterminée de manière à optimiser les performances de la stratégie étudiée ici.

La première partie de notre article détaille l'application des techniques de classification automatique et plus exactement la coalescence statistique à la vision binoculaire ainsi que l'adaptation de ces techniques aux problèmes spécifiques de la vision dynamique binoculaire. La deuxième partie comportees résultats de segmentation de mouvement et de mise en correspondance stéréoscopique de séquences réelles.

2. VISION BINOCULAIRE ET COALESCENCE

2.1. Les tâches de la vision binoculaire: La vision binoculaire intéresse de nombreux domaines. On peut citer par exemple, la télévision en relief, la navigation de robots mobiles, l'imagerie médicale, etc.... Les tâches à réaliser sont principalement de deux sortes: la reconstruction 3D et l'estimation de mouvement. La qualité de ces deux processus dépend de deux points [1][2]:

- la segmentation de mouvement.
- la mise en correspondance.

2.2. Techniques de coalescence: Les techniques de coalescence sont une classe d'algorithmes appartenant à la famille des techniques de classification non-supervisée [6]. Elles interviennent souvent lors de la phase de prétraitement statistique en permettant l'estimation du nombre de regroupements ou de classes présents dans l'espace de représentation. L'intérêt de ces techniques vient de leur capacité à fournir des labels aux individus - de les classer- sans utiliser aucune règle de décision. Notre étude s'est porté principalement sur trois techniques de coalescence. Les deux premières sont basée sur le principe des nuées dynamiques [3]. Néanmoins, ces deux algorithmes pêchent par le manque de robustesse de la partition fournie et par la nécessité d'initialiser le nombre de classes souhaitées.. La troisième approche étudiée est basée sur les principes de la coalescence robuste [7][9]. Le principal avantage de celle-ci est la présence de la notion de rejet des individus peu fiables.

2.3. Adaptation de la coalescence à la vision dynamique: Les processus de segmentation de mouvement et de mise en correspondance stéréoscopique peuvent être considérés sous certaines conditions comme des problèmes de classification automatique.

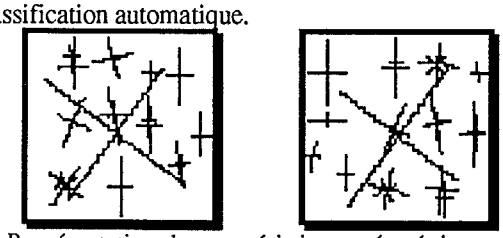

Fig.1. Représentation de caractéristiques géométriques dans un couple d'images stéréoscopiques

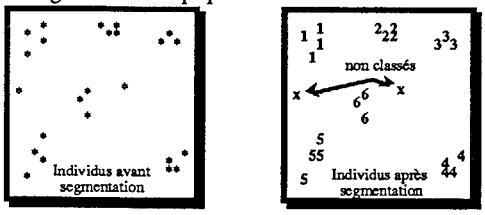

Fig.2. segmentation de 4(2x2) images stéréoscopiques

Plusieurs stratégies sont possibles, en fonction de l'application recherchée. La segmentation de mouvement et la mise en correspondance stéréoscopique ne sont donc plus forcément distinctes mais peuvent fusionner en un processus unique. Les deux séquences stéréoscopiques ne sont plus perçues comme des séquences parallèles mais comme formant une entité unique. La signification donnée à une classe statistique dépendra de la stratégie retenue. Si les deux processus sont distincts, une classe correspond, dans le cas de la segmentation de mouvement, à l'ensemble des primitives spatio-temporelles décrivant le même objet au cours du temps. Dans le cas de la mise en correspondance stéréoscopique, une classe sera formée par les éléments spatiaux -projections ou contours d'un objet 3D- perçus par le dispositif binoculaire.

Une retombée intéressante de notre approche est l'affranchissement de la recherche de correspondant stéréoscopique de la contrainte épipolaire habituellement utilisée pour réaliser les appariements.

a- Niveau sémantique: Un premier choix important à faire est lié à la détermination du niveau sémantique des primitives. Il est évident que pour permettre une discrimination entre les objets de la scène et pour assurer la rapidité du processus, il faudrait une description globale, concise et discriminante des objets de la scène. Si on considère la segmentation spatiale par contour ou segments, on remarque une forte sensibilité au bruit. De plus, les caractéristiques usuellement utilisées pour décrire globalement une portion de contour approximée par un segment -longueur, coordonnées du centre du segment,...- sont excessivement instables. Seule l'orientation de segments possédant un gradient important s'avère assez robuste. L'utilisation des régions paraît donc tout indiquée pour notre stratégie. En effet, un grand nombre de caractéristiques globales peuvent être utilisées.

b- Espace de représentation: Comme pour tout problème de reconnaissance des formes statistique, un espace de représentation -dans lequel doit s'effectuer la segmentation de mouvement *et/ou* la mise en correspondance stéréoscopique- doit être choisi de manière adéquate. Du point de vue de la classification automatique, les individus appartenant à une même classe doivent former un amas plus ou moins compact dans l'espace de représentation (fig.2). Il est donc nécessaire que les mesures extraites des deux vues gauche et droite soient invariantes par projection et indépendantes de l'orientation 3D supposée inconnue des régions observées.

indépendantes de l'orientation 3D supposée inconnue des régions observées.

Certaines propriétés géométriques globales telles que les moments peuvent être retenues pour la construction de cet espace. Les moments d'ordre supérieur à 2 sont cependant trop facilement perturbés par le bruit et sont inutilisables dans le cas de scènes réelles. Il faut cependant s'assurer que la géométrie des projections gauche et droite soit identique pour pouvoir décrire de manière unique la région 3D (fig.1) . Ceci est obtenu en considérant une configuration particulière des caméras -voir plus bas.

Par ailleurs, On constate que les propriétés photométriques habituellement utilisées pour caractériser globalement une région donnée de l'image ne sont pas suffisamment robustes. Des caractéristiques comme que : la moyenne des niveaux de gris, le minimum ou le maximum sont trop dépendantes du point de vue, de l'éclairage ou du niveau signal/bruit. La couleur par contre peut constituer un très bon critère de discrimination. En effet, cette propriété est généralement, invariante pour la plus grande partie de natures de surfaces pouvant être observées. De plus, les applications actuelles nécessitent de plus en plus la couleur, même si la segmentation en couleur est plus coûteuse en temps de calcul.

c- Configuration du système d'acquisition: Si on suppose que les objets de la scène sont modélisables par des surfaces planes, alors la transformation T liant la géométrie de la projection d'un objet donné observé par les deux caméras d'un système d'acquisition binoculaire est reflétée par un système dépendant de l'orientation 3D de la région, de la profondeur de la région de l'image observée et enfin de la calibration du système d'acquisition binoculaire [10].

L'examen de cette transformation montre qu'il suffit de positionner les plans caméras parallèles à la droite reliant les focales des deux caméras pour que d'une part, cette relation devienne indépendante de l'orientation de la normale de la région 3D et des paramètres de la calibration et d'autre part, qu'elle permette une distance minimale entre les caractéristiques géométriques des projections gauche et droite de la région 3D. En fait, les conditions d'acquisition de la TV3D -angles de rotation faibles et absence de décalages suivant l'axe vertical et l'axe de visée-, s'avèrent largement satisfaisantes.

3. RÉSULTATS

3.1.Segmentation de mouvement: La réalisation de la segmentation de mouvement à l'aide d'une stratégie de coalescence donne d'excellents résultats. La séquence couleur ci-dessous (Fig.3) comprend des objets se déplaçant à des vitesses fort différentes.

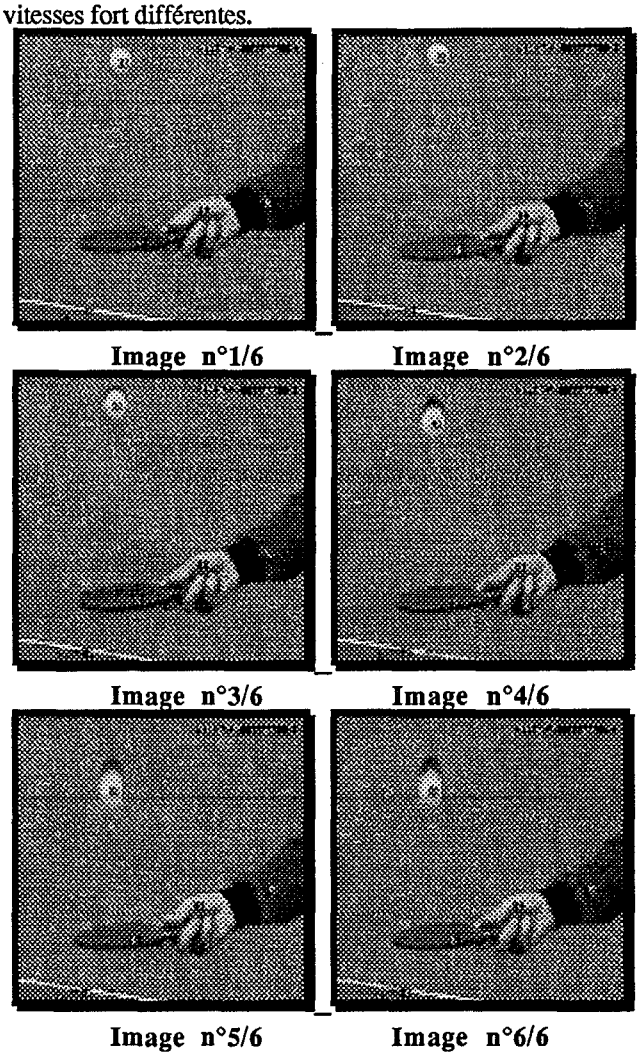

Fig.3. Segmentation de la séquence couleur "Ping-Pong".

3.2. Mise en correpondance: Les résultats de la segmentation de mouvement et de la mise en correspondance simultanés sont très intéressants. L'exemple montre une séquence réelle difficile (fig.4). L'espace de représentation utilisé est formé à partir de primitives géométriques et photométriques.

Aucune connaissance préalable des éléments de calibration n'a été utilisée. Néanmoins, on s'est assuré des conditions d'acquisition décrites plus haut.

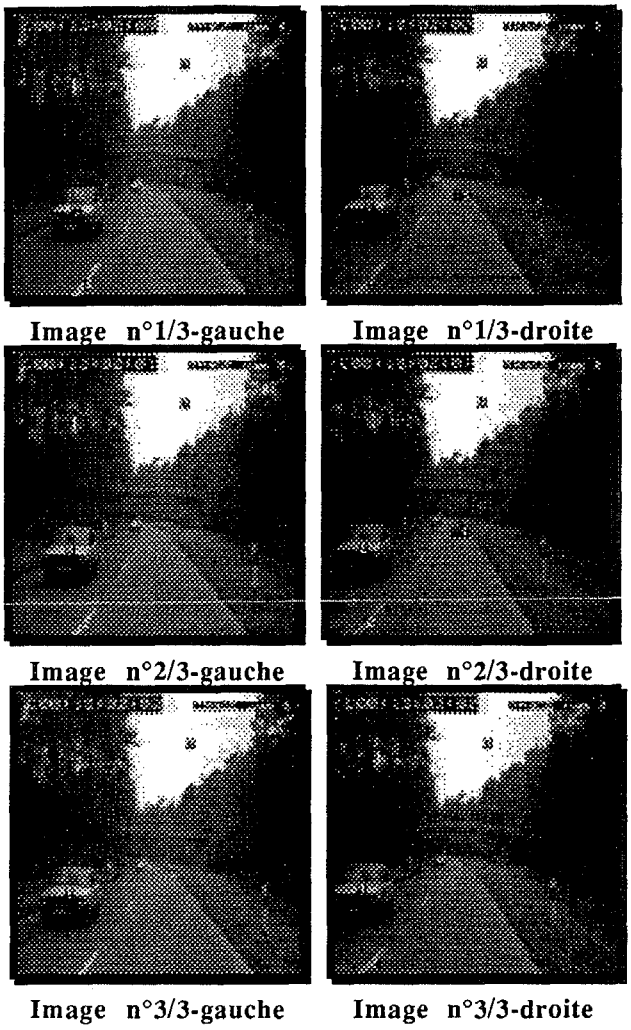

Fig.4. Mise en correspondance stéréoscopique.

4. CONCLUSION

Les avantages de la stratégie de segmentation dans une séquence binoculaire basée sur une technique de coalescence, décrite dans cet article, peuvent être très intéressants:

- les appariements stéréoscopiques ne dépendent plus de la contrainte épipolaire. Ce qui nous affranchit de l'étape coûteuse de la calibration.
- le traitement *simultané de* 2xn images d'une séquence binoculaire devient possible pour certaines applications de codage d'images.
- la stratégie de recherche de correspondants permet de segmenter des scènes comportant des objets animés de mouvements différents et quelconques. Aucune hypothèse fortement restrictive sur le mouvement n'est à faire.
- l'intégration de la variabilité dans le moteur de la stratégie rend possible le traitement de scènes comportant des occlusions.
- l'apparition ou la disparition de régions sont facilement appréhendées. Elles sont vues comme la création ou la destruction d'une classe statistique.
- la notion de rejet présente dans le classificateur robuste permet de mettre à l'écart les régions mal segmentées ou les pseudo-régions. Seules les régions suffisamment fiables seront traitées par le module de l'estimation de mouvement situé en aval.

Cependant, l'inconvénient majeur de cette approche provient de la validité statistique de certains tests de qualité situés au coeur des classifieurs. Ce problème se pose essentiellement quand on désire travailler par exemple sur un couple unique stéréoscopique ou quand on désire segmenter deux images. Une solution qui a fait ses preuves consiste à recourir aux techniques statistiques de bootstrapping [9]. Les effectifs sont alors artificiellement gonflés dans le but d'augmenter le nombre des individus réellement présents dans l'espace de représentation.

5. BIBLIOGRAPHIE

[1] N. Ayache,*Vision Stéréoscopique et Perception Multisensorielle: Applications à la Robotique Mobile*, Paris,InterEditions, 1989.

[2] P.Bouthémy, "Modèles et méthodes pour l'analyse du mouvement dans une séquence d'images", *Techniques et Science Informatique*, 1987.

[3] E.Diday and F.Lemarie, *Eléments d'analyse de données.* Paris: Dunod, 1970.

[4] R.Duda & P.Hart, Pattern Classification and scene analysis. New York:John Wiley & Sons, Inc, 1973.

[5] K.Fukunaga, Introduction to Statistical Pattern Recognition, 1972, Academic press, Inc.Orlando.

[6] A.K. Jain and C.Dubes, *Algorithms for Clustering Data*, Engelwoods CliffsPrentice Hall,

[7] J.M Jolion, Peter Meer, and Samira Bataouche, "Robust Clustering with Applications in Computer Vision, IEEE transactions on Pattern Analysis and Machine Intelligence", Vol 13, No. 8, August 1991.

[8] K.Kaoula & M.Benjell-oun, "Achieving The Spatio-temporal Segmentation in a Feature Space: A pattern recognition approach", IFAC/IEEE 1992, Perugia, Italy.

[9] P. J. Rousleeuw and A. M. Leroy, *Robust Regression & Outlier Detection.* New York: Wiley, 1987.

[10] J.M Vezien & A.Gagalowicz,*A region based 3D reconstruction of a stereo pair of images* , 1991.