

PREMIER COLLOQUE IMAGE
Traitement, Synthèse, Technologie et Applications

BIARRITZ — Mai 1984 —

VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART

COMPUTER VISION IN ROBOTICS : STATE OF THE ART

O.D. FAUGERAS

INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY CEDEX - FRANCE

RESUME

La Vision par Ordinateur est le moyen privilégié pour les Robots d'appréhender leur environnement.

Dans cet article nous passons en revue les différentes méthodologies qui ont été développées dans ce domaine en nous centrant sur celles susceptibles d'intervenir en Reconnaissance, Positionnement et Inspection d'objets.

La détection de contours et la caractérisation de textures sont analysées dans ce contexte et un besoin important d'intégration des algorithmes est identifié. Nous nous concentrons ensuite sur des systèmes de Reconnaissance et Positionnement bidimensionnels qui résolvent le problème du vrac planaire. Puis nous abordons les tâches tridimensionnelles avec d'une part le problème de l'acquisition de données distance par des techniques passives ou actives et le problème de la Reconnaissance et du Positionnement d'autre part.

Nous concluons en identifiant les possibilités de transfert technologique et les domaines où une recherche est encore nécessaire.

SUMMARY

Computer Vision is a privileged way for Robots to acquire knowledge about the outside world.

In this paper we review the various methodologies which have been developed in this field from the standpoint of object Recognition, Positioning and Inspection.

Edge detection and texture analysis are examined within this context and we identify a deep need for algorithm integration. We then concentrate on Object Recognition and Positioning Systems that solve the planar bin picking problem. Next we tackle three-dimensional tasks with on one hand the problem of acquiring range data by active and passive techniques and on the other hand with the Recognition and Positioning problem.

We conclude by pointing out areas where technological transfers from Computer Vision to Robotics are possible and areas where Research is still needed.



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
COMPUTER VISION IN ROBOTICS : STATE OF THE ART

O.D. FAUGERAS.

Introduction :

Le développement des Robots industriels de 3ème génération c'est-à-dire capables de s'adapter à des variations de leur environnement ou des Robots autonomes mobiles nécessite l'utilisation de capteurs pour appréhender cet environnement. La vision est un mode de perception privilégié du point de vue biologique et, sans faire d'antropomorphisme outrancier, on peut dire que pris dans un sens large c'est et ce sera la pierre angulaire des systèmes de perception des Robots. Les questions qu'il faut poser sont, quelles sont les performances qu'on peut attendre aujourd'hui des systèmes de Vision industrialisés et de ceux qui existent sous forme de prototypes dans les laboratoires de recherche, quelles sont les performances qu'on peut attendre demain de ces systèmes et dans quelles directions doivent porter les efforts de recherche.

Cet article s'efforce de répondre à ces questions en tentant de dresser un panorama des techniques, des algorithmes et des systèmes utilisés en Vision par Ordinateur et susceptibles d'avoir rapidement un impact en robotique. Nous l'avons structuré de la manière suivante : nous commençons par discuter les problèmes d'acquisition donc de capteurs, puis nous présentons les problèmes bidimensionnels qui constituent à l'heure actuelle la plus grosse demande en robotique et nous terminons par les problèmes tridimensionnels.

Systèmes d'acquisition

On peut distinguer utilement deux grandes classes de systèmes, les systèmes actifs qui concentrent une énergie lumineuse en une petite zone de la scène afin de la suréclairer, et les systèmes passifs. Les systèmes actifs s'attachent essentiellement à mesurer la distance au capteur de points situés sur les objets de la scène. Ce sont donc plutôt des capteurs de distance que des capteurs d'intensité. Passifs ou actifs les systèmes d'acquisition utilisent la plupart du temps les mêmes capteurs de type télévision c'est-à-dire fournissant soit un signal Vidéo standard qui est ensuite numérisé soit dans certains cas directement une information digitale.

Les caméras de type Vidicon très répandues parce que relativement bon marché du fait de l'existence du marché de la Vidéo conduisent à des résolutions spatiales de l'ordre de quelques pourcents du champ et des résolutions de l'ordre de .4 % du signal d'intensité. Elles présentent en général des déformations géométriques qui sont coûteuses à corriger.

Les caméras Etat-Solide utilisent des rétines photosensibles discrètes obtenues à partir d'une technologie silicium ce qui présente l'avantage d'éliminer les problèmes de géométrie puisqu'on peut contrôler la position des éléments photosensibles à 0.01 %. Ces caméras sont de deux types. Elles utilisent soit un arrangement linéaire de photodiodes pouvant en comporter jusqu'à 4096 soit encore un arrangement plan de cellules photosensibles pouvant aller jusqu'à des résolutions de 480 x 380 comme dans les caméras CID ou CCD.

Dans la plupart des applications il est nécessaire d'acquérir des images bidimensionnelles ce qui pour les caméras à barette nécessite l'introduction de la deuxième dimension à l'aide d'un miroir monté sur galvanomètre par exemple ou en utilisant le mouvement des objets étudiés qui sont très souvent transportés sur un tapis roulant. Cette dernière situation nécessite un contrôle précis de la

vitesse de déplacement (0.05 % pour une barette de 2048 éléments).

Les systèmes actifs sont basés sur l'idée de suréclairer un ensemble de points de la scène afin de déterminer par triangulation ou par mesure du temps de vol la distance des points correspondants. Le principe de la technique par triangulation est expliqué sur la figure 1.

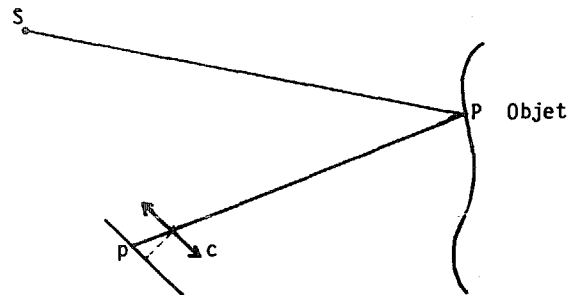


Figure 1

Connaissant la position du rayon incident SP et celle de l'image p du point P sur la rétine du capteur on peut calculer les coordonnées de P.

Si le capteur est une barette linéaire il faut au moins 1 milliseconde pour obtenir p ce qui limite fortement la cadence d'acquisition. Celle-ci est encore plus basse si l'on utilise une rétine bidimensionnelle qui nécessite au moins 17 millisecondes par image. Une façon simple d'augmenter la cadence d'acquisition est de mesurer les distances de plusieurs points en un temps de lecture de la rétine photosensible. Ceci s'obtient en projetant sur la scène une configuration de plusieurs points comme celles indiquées sur la figure 2.

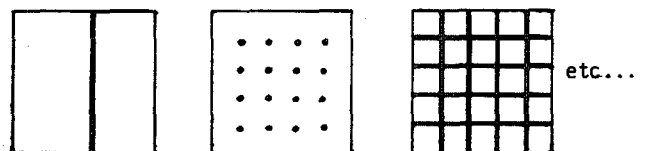


Figure 2

Un problème important avec ce genre de technique par triangulation est celui de l'occultation illustré sur la figure 3.

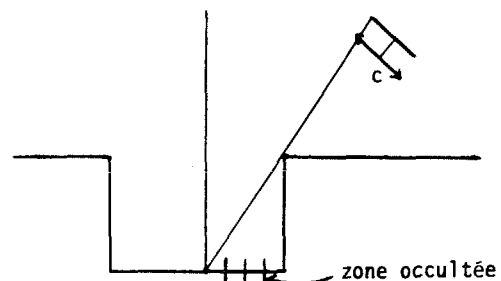


Figure 3

On peut remédier à ce problème en utilisant plusieurs capteurs ou en utilisant la méthode dite de mesure du temps de vol. Les systèmes commerciaux utilisant cette technique commencent à apparaître sur le marché avec des précisions géométriques de l'ordre de 0.5 % et des cadences d'acquisition de l'ordre de 500 à 10 000 points par seconde. Les adresses de quelques fabricants sont indiqués dans la table ci-dessous.



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

TABLE I

Technical Arts Corp.
 180 Nickerson
 Suite 102
 Seattle, WASHINGTON 98109, U.S.A.
 (206) 282.17.03

Digital Design
 Z.A. Orsay-Courtaboeuf
 Avenue de l'Océanie
 Immeuble C2
 B.P. 90, 91943 LES ULIS CEDEX
 (6) 928.01.31

Robotic Vision Systems Inc.
 536 Brood Hollow Road
 MELVILLE NY 11747, U.S.A.
 (516) 694.89.10

ITMI
 Chemin des Clos, Zirst
 38240 MEYLAN
 35 (76) 90.33.81

La référence (1) est un bon résumé des différentes technologies utilisées pour les capteurs.

Vision bidimensionnelle

Une très grande classe de problèmes posés par la Robotique Industrielle est celle de la Reconnaissance, du Positionnement et de l'Inspection de pièces et de matériaux (RPI). Pour une tâche donnée une seule ou plusieurs de ces opérations peut être nécessaire. Les données à traiter sont soit bi- soit tri-dimensionnelles. Autant il est possible d'envisager un système général de Reconnaissance et de Positionnement c'est-à-dire adapté à une très vaste classe d'objets, autant il est difficile d'envisager un tel système pour les tâches d'inspection. Une méthodologie de l'inspection automatique reste à définir. Dans la plupart des applications d'inspection les cadences de traitement sont d'au moins 1 million de points d'image par seconde afin de satisfaire les contraintes économiques et justifier l'investissement correspondant. Ceci implique l'utilisation de matériel spécialisé avec un haut degré de parallélisme et incorporé dans une structure pipeline. Ce type d'architecture n'est en général pas directement transposable d'une application à une autre. Une conséquence de ce coût élevé est que les applications ont jusqu'ici été pratiquement limitées à des tâches où le volume de produits est important et où une très haute qualité doit être maintenue : masques de circuits intégrés, circuits imprimés, pièces simples automobiles et produits pharmaceutiques.

Nous pensons néanmoins que les outils très généraux d'analyse d'image qui existent dans les laboratoires devraient permettre, une fois intégrés, de réaliser un système modulaire général capable de réaliser les tâches de RPI à bas coût pour une très grande gamme d'applications et réduire considérablement la diversité des systèmes actuels qui ne sont souvent que la mise en place de techniques heuristiques ad-hoc complètement dépendantes d'une application donnée.

Plutôt que de passer en revue une longue liste de systèmes d'Inspection à partir de laquelle le lecteur aurait du mal à se faire une idée synthétique nous préférons le renvoyer à la littérature. (2) comporte 288 références sur des sujets tels que l'inspection de circuits imprimés, de masques de circuits intégrés, de pièces automobiles,

de tissus, etc... Dans (3) et (4) on trouve à la fois des articles sur des méthodes générales que sur des systèmes commerciaux et/ou implantés sur des sites industriels.

Nous nous concentrerons ici sur les outils généraux susceptibles d'être utilisés efficacement dans la plupart des tâches de RPI et plus spécifiquement sur les outils d'extraction de contours et de textures qui sont fondamentaux.

Les outils généraux : caractérisation des contours et des textures

Extraction de contours :

Le problème de la détection de contours est l'un des problèmes les plus anciens en Vision par Ordinateur et la littérature dans ce domaine est immense. Le lecteur trouvera une bonne présentation d'un grand nombre de techniques proposées dans les références (5-7).

Il est important de distinguer la détection locale des contours (microcontours) d'une détection globale (macrocontours). La première s'effectue à l'aide de techniques à base de convolutions et fournit à partir d'une image une autre image. La seconde s'effectue à partir de techniques de suivi s'apparentant à des traversées de graphes et fournit des listes de points. L'extraction de macrocontours est d'autant plus simplifiée que celle des microcontours est de qualité et satisfait à des contraintes d'insensibilité au bruit et de précision spatiale.

Deux approches sont utilisées pour les microcontours, l'utilisation d'opérateurs de dérivation du premier et du second ordre, directionnels ou pas (8-17) et l'approximation de l'image par un ensemble de fonctions de base (18-20). Dans le cas d'opérateurs directionnels, ceux-ci doivent être appliqués à l'image dans plusieurs directions pour détecter les contours d'orientation quelconque. Un changement rapide d'intensité produit un pic à la sortie d'un opérateur de dérivation du premier ordre, un passage par zéro à la sortie d'un opérateur de dérivation du second ordre.

La détection des macrocontours s'effectue donc, pour un opérateur de dérivation du premier ordre, en chainant les maxima locaux supérieurs à un certain seuil de contraste avec tous les problèmes de discontinuité que cela implique. Un avantage considérable des techniques basées sur le passage par 0 est que le problème du suivi c'est-à-dire de la construction des macrocontours est plus simple puisque les lignes de passage par 0 sont par définition fermées ou sortent de l'image. Un désavantage est qu'elles donnent une réponse à tous les contours quel que soit leur contraste. L'élimination des contours "inintéressants" s'effectue alors en utilisant un critère de longueur (les macrocontours doivent être d'une longueur supérieure à un seuil) et/ou un critère sur le contraste (le contraste du contour doit être supérieur à un seuil). Ce filtrage est effectué dans (15) en considérant la pente du passage par zéro qui est liée au contraste du contour et dans (17) en combinant l'image des passages par zéro et l'image du gradient et en rejetant les macrocontours dont la valeur moyenne du gradient ne dépasse pas un certain seuil. Nous montrons sur la figure 4 le résultat d'une telle technique sur une scène comportant des pièces industrielles. En combinant un tel filtrage des macrocontours avec une approximation géométrique par une chaîne polygonale et une approximation des variations d'intensité au voisinage de ceux-ci par



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
COMPUTER VISION IN ROBOTICS : STATE OF THE ART
O.D. FAUGERAS

trois plans (figure 5) on obtient une riche représentation symbolique des frontières de l'image.

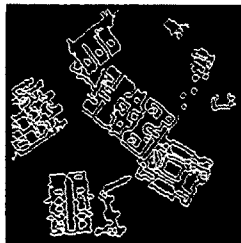
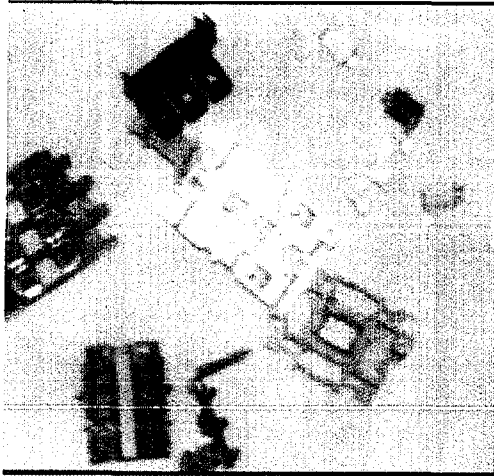


Figure 4

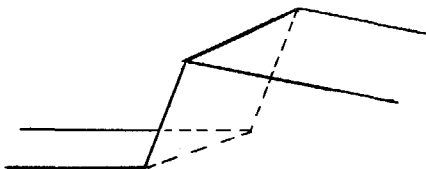


Figure 5

L'un des défis actuels en Vision par Ordinateur est l'intégration dans des processeurs spécialisés de ces algorithmes permettant de calculer rapidement une telle représentation à partir de l'image. La réalisation de ces processeurs à des coûts réduits est un point de passage obligé si la Vision par Ordinateur veut pouvoir déboucher au plan industriel.

Analyse de Textures :

Un autre serpent de mer en Vision par Ordinateur est celui de la caractérisation quantitative des textures. Le problème est compliqué parce qu'il est difficile de donner une définition précise de ce qu'est une texture. Il s'agit de la manifestation dans une image d'un phénomène d'organisation des intensités lumineuses qui, quand il est décomposable, comporte deux niveaux. Un premier niveau microscopique local permet de dégager la notion de primitives de texture. Un second niveau plus macroscopique permet de dégager la notion d'agencement, d'organisation de ces primitives les unes par rapport aux autres.

Les deux grandes classes de techniques d'analyse de textures qui sont les techniques statistiques et les techniques structurelles correspondent en gros aux deux niveaux précédents. Le lecteur intéressé trouvera dans (21) une bonne revue des différentes approches. On range sous le vocable

approches statistiques celles qui utilisent des transformations orthogonales locales sur l'image (Fourier, Haar, Hadamard, etc ...). Il est à noter que la transformation de Fourier peut être implémentée soit numériquement soit par des techniques de filtrage optique. Plus naturellement on rencontre dans cette rubrique les techniques utilisant les statistiques locales de l'intensité lumineuse principalement les lois du premier ordre (histogrammes) ou les moments du second ordre c'est-à-dire la fonction d'autocorrélation. Celle-ci est modélisée soit directement (22), soit à partir du spectre de puissance c'est-à-dire sa transformée de Fourier (23) soit encore à l'aide d'un processus autorégressif (24-27). L'utilisation de moments d'ordre supérieur a également été étudiée dans (27). Les lois jointes du second ordre dont l'utilité a été justifiée par des études psychovisuelles (26-27) ont prouvé leur pouvoir de séparation pour la reconnaissance de textures (28-29).

Dans un ordre d'idées très différent, les outils de la Morphologie Mathématique (MM) développés par Matheron et Serra (30-31) se sont avérés très utiles dans la caractérisation des textures. L'algèbre des opérateurs de la MM est susceptible d'être implémentée de manière très efficace dans des processeurs spécialisés.

Pour en terminer avec les techniques statistiques, j'aimerais citer le travail de Laws (32) qui définit à partir d'opérateurs de dérivation et de moyennage très simples (correspondant à des noyaux de convolution de taille 2x1 ou 1x2) toute une gamme d'opérateurs de convolution dont la combinaison fournit à ma connaissance l'une des meilleures, et potentiellement les plus rapides, techniques de caractérisation de textures naturelles.

Les techniques structurelles peuvent être présentées comme une généralisation des techniques statistiques de façon à incorporer le second niveau qui concerne l'organisation des primitives de texture. Dans un premier temps on effectue une croissance de région sur un ou plusieurs attributs (intensité, gradient, etc ...) de façon à identifier les primitives qu'on caractérise par leur forme et la valeur moyenne de l'attribut utilisé. Dans un second temps on calcule les histogrammes du premier ordre (33-34) ou du second ordre (35) de ces paramètres afin de caractériser leur répartition dans l'image.

Beaucoup de travail est encore nécessaire en vue d'une meilleure compréhension des textures visuelles tout particulièrement du côté de l'utilisation de modèles physiques ou statistiques 3-D des surfaces des objets observés pour caractériser la structure des textures mesurées sur la rétine du capteur utilisé. Quelques éléments semblent être disponibles dans la communauté Synthèse d'Images (36) inspirés en particulier des travaux de Mandelbrot (37).

L'utilisation des techniques d'analyse de texture n'est sans doute pas nécessaire pour la robotique industrielle d'assemblage par contre elles me paraissent être un élément important pour de nombreux systèmes d'inspection et pour des robots mobiles autonomes.

Reconnaissance et Positionnement bidimensionnels

L'un des facteurs inhibant l'automatisation à grande échelle est la difficulté à saisir des pièces de leur lieu de stockage et à les présenter à un poste de travail dans une position et une orientation donnée. Suivant la complexité du problème on peut soit utiliser une information 2-D soit une



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

information 3-D. Un exemple de problème très simple est la localisation d'un objet connu plan dans de bonnes conditions d'éclairage (problème A), un exemple de problème compliqué est la reconnaissance et le positionnement d'objets tridimensionnels en vrac. Nous réservons la discussion des problèmes 3-D à un paragraphe ultérieur et nous concentrons maintenant sur les problèmes bidimensionnels.

La plupart des systèmes de Vision industriels existant à l'heure actuelle ne savent résoudre que le problème A précédent. Le lecteur trouvera une bonne analyse comparative de ces systèmes dans (38). En simplifiant à peine on peut dire que ces systèmes sont basés sur celui du SRI (39). Les deux idées de base sont d'une part que puisque le contraste entre le fond et l'objet est important on peut se ramener par simple seuillage à une image binaire donc facilement exploitable et que d'autre part puisque les objets sont isolés on peut caractériser leur forme, leur position et leur orientation à l'aide de quelques attributs globaux comme le périmètre, la surface totale, la surface des trous, etc ... La reconnaissance s'effectue soit à partir d'arbres de décisions binaires dans les cas les plus simples soit à partir de tests bayésiens.

L'évolution de ces systèmes vers une plus grande généralité s'effectue dans deux directions principales qui correspondent à la relaxation des deux hypothèses de base du système du SRI. Ou bien on ne suppose plus que les conditions de contraste sont telles qu'un simple seuillage suffit à séparer les objets du fond ou/et on ne suppose plus que les objets sont isolés. Le premier cas implique la mise en oeuvre de techniques de segmentation plus élaborées tout particulièrement au niveau de la détection de contours. La tendance actuelle est à la réalisation d'opérateurs spécialisés rapides réalisant certains algorithmes mentionnés précédemment. Un exemple d'un tel système est le système ANIMA commercialisé par la société GIXI (40-41).

Le premier comme le second cas impliquent que seule une portion des objets à reconnaître et à positionner est visible et que l'utilisation d'attributs globaux n'est plus possible, ceux-ci n'étant pas stables par occultation. Avant de passer à la description plus détaillée d'un certain nombre de systèmes ne présupposant pas des objets isolés il est bon d'indiquer quels genres de scènes ils permettent d'analyser ; des scènes comprenant :

- un morceau d'objet
- plusieurs objets se touchant
- plusieurs objets pouvant se chevaucher partiellement
- plusieurs objets pouvant avoir des défauts (on rejoint alors le problème de l'inspection).

Ces systèmes se caractérisent donc par une capacité accrue de segmentation d'image ; ils sont capables d'isoler des objets ou des portions d'objets dans des conditions de contraste et d'éclairage assez mauvaises en utilisant des opérateurs d'extraction de contours performants. Ils se caractérisent aussi par le fait que les programmes qui les constituent manipulent des descriptions symboliques des modèles des objets et des images analysées et qu'en conséquence les procédures de reconnaissance et de positionnement tirent leurs paradigmes plutôt du côté de l'Intelligence Artificielle que de la Reconnaissance des Formes statistiques ou de l'Analyse de Données.

Probablement l'un des premiers systèmes de

ce type est celui de Perkins (42). La partie extraction de contours utilise un opérateur d'O à Hueckel (19) et une procédure de suivi et de chaînage complétée par un algorithme de segmentation des macrocontours obtenus en segments de droite et arcs de cercles. Ceci constitue la partie acquisition de modèles et description symbolique de l'image. Modèles et image sont donc décrits en termes d'ensembles de courbes généralisées (cg), chaque courbe étant elle-même un ensemble de segments de droites et d'arcs de cercles. La partie reconnaissance et positionnement consiste alors à mettre en correspondance les modèles et la description de l'image. Les couples de cg modèles et image sont ordonnés à l'aide d'attributs globaux. Le meilleur appariement est alors sélectionné et utilisé pour estimer la transformation géométrique (rotation, translation) appliquant le modèle sur l'objet. Cette estimation s'effectue en corrélant les représentations $t(s)$ des macrocontours associés où s est l'abscisse curviligne le long du contour mesuré à partir d'un point de référence arbitraire et t est l'orientation de la tangente. Dans le cas où les cg ont une symétrie de révolution (cercles complets) Perkins en utilise deux pour estimer la transformation totale.

A ce niveau, les transformations inacceptables peuvent être rejetées immédiatement si la corrélation dans l'espace $t(s)$ est mauvaise. Un test supplémentaire est en général nécessaire pour confirmer ou infirmer la transformation précédente. Ce test est basé sur une comparaison exhaustive locale des cg impliquées. Ce système a été testé avec succès sur des scènes comportant de l'ordre de vingt à trente cg, les temps calcul correspondants (programme PL/I exécutant sur IBM 370/168) sont de l'ordre de 20 secondes. Il est à noter que la robustesse du système à l'occultation n'a pas été testée de manière approfondie, que l'estimation de la transformation a toutes les chances d'être assez imprécise puisque basée uniquement sur l'appariement d'un ou deux couples de cg, enfin que le système n'a pas prévu explicitement de variations de facteur d'échelle entre le modèle et l'image. Un système un peu similaire, en particulier au niveau de l'utilisation de la corrélation entre des représentations intrinsèques des macrocontours, est celui de Dessimoz (43-44).

Une autre approche assez différente est celle de Bolles et Cain (45). Là encore l'idée est de représenter l'image et les modèles sous forme de listes d'attributs locaux tels que des petits trous circulaires, des coins rectangulaires ou des segments de droite. Au niveau du modèle, une procédure d'apprentissage partiellement automatique permet de regrouper les attributs en amas locaux organisés autour d'un attribut dit focalisateur. Chaque attribut focalisateur est donc caractérisé de manière non ambiguë par une liste d'attributs voisins dont l'identification permet de reconnaître l'attribut focalisateur auquel ils sont rattachés et d'établir la position et l'orientation de l'objet. Enfin la liste des attributs focalisateurs est partiellement ordonnée par ordre de mérite. Bolles et Cain n'analysent pas de manière détaillée les problèmes liés à l'extraction robuste de ces attributs dans des images bruitées. La procédure de reconnaissance et de positionnement utilise comme précédemment un paradigme du type production et vérification d'hypothèses. Plus spécifiquement, le système commence par rechercher dans l'image les attributs considérés par le modèle comme les plus prometteurs. En ayant trouvé un il détermine si les attributs voisins sont présents ou non. Ceci permet de générer une liste de correspondances possibles entre attributs du modèle et attributs de l'image. Cette



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

liste est convertie en un graphe dont les sommets sont les correspondances possibles et dont les arcs indiquent la compatibilité des correspondances qu'ils joignent. Un algorithme de recherche de cliques maximales c'est-à-dire de plus grands sous-graphes complètement connectés est alors appliqué. Chacun des sous-graphes identifiés permet de générer une hypothèse c'est-à-dire la rotation et la translation nécessaires pour passer du modèle à la scène. Chaque hypothèse est ensuite confirmée ou infirmée à l'aide de deux tests. Le premier test recherche dans l'image d'autres attributs compatibles avec l'hypothèse et les utilise pour raffiner son estimation de la transformation (d'une façon qui n'est pas précisée). A partir de cette meilleure estimation, le programme va tester dans l'image que les variations d'intensité lumineuse au voisinage de la frontière hypothétique de l'objet sont compatibles avec l'information stockée dans le modèle. Ce système a été testé sur des scènes comportant de l'ordre de 20 à 30 attributs avec de bonnes conditions de contraste et les temps de calcul sont de l'ordre de 10 à 25 secondes.

L'utilisation de techniques de recherche de cliques maximales dans des graphes d'appariement avait déjà été proposée par Barrow et al (46) en 1973. Leur principal désavantage est que le problème est NP - complet c'est-à-dire qu'il n'existe pas d'algorithme de complexité polynomiale par rapport au nombre de sommets du graphe permettant de le résoudre. Bolles et Cain contournent l'obstacle en limitant la taille du graphe par l'utilisation d'attributs voisins. Ceci limite la généralité de leur système. Il est à noter également que la robustesse du système à la dégradation de la qualité des images n'a pas été testée et que le système n'a pas prévu explicitement de variation de facteur d'échelle entre le modèle et l'image.

Une autre approche (47-49) utilise également une représentation de l'image et du modèle en termes de segments de droites, d'arcs de cercles et de coins. Contrairement à Bolles et Cain les attributs ne sont pas structurés en amas locaux. Ces attributs sont extraits de l'image qui peut être de qualité médiocre à l'aide d'une technique de détection de contours de type gradient. La procédure de reconnaissance et de positionnement ou d'appariement de modèle utilise là encore un paradigme de type prédiction et vérification d'hypothèses dont le moteur est ici un algorithme

d'exploration d'arbre de décision de type A* (50). Dans cette méthode, une hypothèse est l'appariement d'un attribut du modèle et de la scène. Si l'on exclut le prétraitement de l'image qui dans (48-49) est de 50 secondes pour une image 128 x 128 x 6 les temps de reconnaissance pour un modèle sont inférieurs à la seconde. La précision de la localisation ne semble pas très bonne et le système n'a pas été prévu pour prendre en compte de variations de facteur d'échelle entre le modèle et l'image.

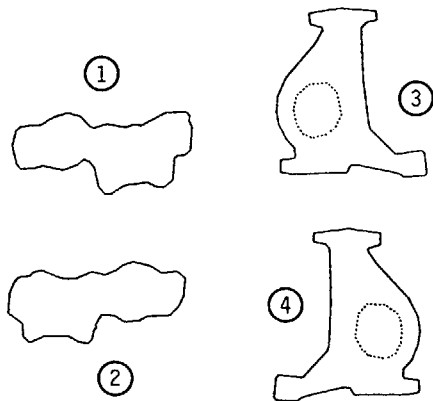
Une autre approche est celle du système PVV (51-52) commercialisé par la société ITMI mentionnée plus haut. PVV comporte deux modules : un module de description qui extrait des attributs de l'image (dans la version implémentée des segments de droite) et qui peut fonctionner soit en mode ascendant classique soit en mode descendant pour vérifier la présence dans l'image d'indices visuels donnés et un module de prédiction et vérification d'hypothèses qui interprète les attributs extraits par la segmentation en termes de modèles d'objets. L'originalité du système provient du fait que les deux modules sont implantés comme coroutines de façon à ce que l'interpréteur contrôle entièrement la description

c'est-à-dire qu'on ne commence pas comme dans tous les autres systèmes par extraire tous les attributs possibles dans l'image. PVV est implanté sur micro-ordinateur LSI-11/23 et a été testé sur des scènes relativement simples. L'information relative à la précision du positionnement et aux temps d'exécution n'est pas disponible mais par contre le système est indépendant de la position de la caméra.

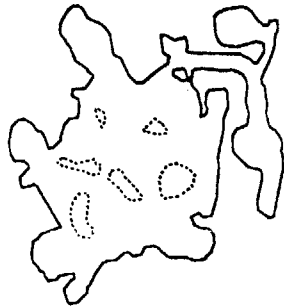
Le dernier système que nous décrivons est le système HYPER (Hypothèses, Prédiction et Evaluation Réursive) développé à l'INRIA (53-54) et commercialisé par la société Digital Design mentionnée plus haut. La partie extraction de contours est indépendante de la partie reconnaissance. Elle utilise la combinaison d'opérateurs de dérivation du premier et du second ordre (17) suivie d'une approximation polygonale. Les attributs image et modèle sont donc des segments de droite et le modèle privilégie un certain nombre de segments focalisateurs afin de limiter l'explosion combinatoire. La reconnaissance et le positionnement s'effectuent par prédiction et évaluation réursive d'hypothèses à l'aide d'une technique de Branch-and-Bound. Une hypothèse est l'appariement d'un segment privilégié du modèle et d'un segment de la scène. L'étape de vérification apparie le maximum d'autres segments en mettant à jour récursivement un score de qualité et à l'aide d'un filtre de Kalman l'estimation de la similitude transformant le modèle dans la scène. Le système qui prend en compte explicitement les variations du facteur d'échelle est implémenté sur un calculateur Perkin Elmer 32-40 en Fortran. Il a été testé sur un très grand nombre d'images dont les descriptions comportent plusieurs centaines de segments et des modèles en comportant également plusieurs centaines dans des conditions d'occultation importantes (60 %). Pour de telles scènes les temps de reconnaissance et de positionnement (n'incluant pas la partie segmentation) varient entre une et dix secondes. La précision géométrique est de l'ordre du degré en rotation et de l'ordre de 5 pixels en position sur des images 256 x 256. Des résultats sont présentés en Figure 6.



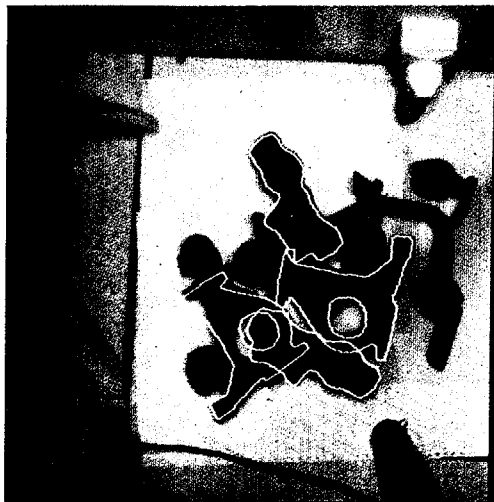
Image originale



Modèles des objets



Détection des contours dans l'image



Résultat de la Reconnaissance et du Positionnement

Figure 6

Le système a été couplé à un bras manipulateur pour réaliser des opérations de saisie automatique de pièces industrielles en vrac planaire, comme il est décrit dans (54).

Vision tridimensionnelle :

A l'inverse des systèmes 2-D dont il existe des versions industrielles, les choses sont beaucoup moins avancées au niveau des systèmes 3-D. Ceci est dû essentiellement à la difficulté qu'il y a à extraire l'information 3-D de façon robuste et précise par des techniques passives et la (presque) non disponibilité commerciale de systèmes actifs. Il est à noter cependant que la plupart des systèmes de

RPI tridimensionnels existants sont basés sur des systèmes d'acquisition actifs.

Nous commençons par passer en revue les techniques d'acquisition passives puis nous décrivons quelques systèmes de RPI. Les techniques d'acquisition passives se décomposent en deux grandes classes. La première classe utilise une ou plusieurs images de la scène supposée fixe en faisant éventuellement varier la position de la source d'éclairage supposée unique. Ces techniques de reconstruction de la forme à partir de l'intensité lumineuse ("shape from shading") (56-57), de reconstruction de la forme à partir de stéréoscopie photométrique ("shape from photometric stereo") (58-59), de reconstruction de la forme à partir de la texture (60) ne fournissent pas directement une information de distance absolue mais plutôt l'orientation de la normale aux objets. La deuxième classe utilise plusieurs images de la même scène prises de points de vue différents si les objets sont fixes ou à des instants différents si certains objets sont mobiles (11,61-66).

Acquisition passive de données distance :

Nous ne parlerons pas, faute de résultats probants, des techniques de reconstruction de la forme à partir de la texture. Les deux autres techniques mentionnées précédemment utilisent le fait que l'intensité apparente d'un morceau de surface d'un objet dépend de quatre paramètres qui sont le type de matériau dont est fait l'objet, la position de la source lumineuse, la position de l'observateur et la normale à la surface. Supposant connus les trois premiers paramètres et l'intensité de la lumière incidente, la mesure de l'intensité apparente en un point d'un objet fournit une relation (non linéaire) entre les composantes de la normale à la surface considérée.

Si l'on considère ces composantes comme les dérivées partielles du premier ordre de l'équation de la surface (56-57) on en déduit une équation différentielle qu'on peut intégrer à partir de diverses conditions aux limites comme par exemple la connaissance de la normale à la frontière de l'objet ("occluding boundary"). Les calculs sont assez compliqués et les résultats assez sensibles au bruit qu'il soit sur l'intensité ou sur les conditions initiales.

Si l'on s'autorise à déplacer de manière connue la source de lumière on peut effectuer une mesure d'intensité apparente par position et en déduire autant de relations liant les composantes de la normale en chaque point (à condition qu'il soit effectivement éclairé). En général trois ou quatre mesures suffisent à réduire complètement l'ambiguïté (58-59). Le gros problème avec cette approche est qu'elle suppose une connaissance exacte et très précise (problème de stabilité des résultats) de la fonction de réflectance. Une analyse assez fine de la robustesse de la technique aux erreurs de mesure est publiée dans (67). Un système de prise en vrac d'objets en forme de tore utilisant cette méthodologie a récemment été développé au MIT (69). Il est à noter cependant que pour éliminer le problème mentionné précédemment de la fonction de réflectance les objets ont été peints en blanc mat ce qui assure une réflectance lambertienne.

La deuxième classe de techniques d'acquisition passive concerne les problèmes de stéréo en général c'est-à-dire l'utilisation de plusieurs images d'une même scène pour inférer une information tridimensionnelle. Notre discussion sera basée sur l'utilisation de deux images ce qui ne



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

restreint en rien la généralité de notre propos.

Le problème de base de la vision stéréo est celui de l'appariement des points des deux images qui correspondent à un même point de la scène. Ce problème est fondamentalement ambigu c'est-à-dire que pour un point d'une image il existe en général plusieurs points candidats à l'appariement dans l'autre image.

La solution de ce problème passe par la réponse à deux questions essentielles : quels sont les attributs des images utilisés pour l'appariement et quelles sont les règles guidant cet appariement. On peut accélérer la mise en correspondance en jouant sur deux paramètres : soit au niveau du prétraitement on calcule à partir de chaque image une représentation suffisamment riche pour réduire considérablement la probabilité de fausses correspondances soit on utilise au niveau de l'appariement des règles de sélection qui réduisent l'explosion combinatoire. En pratique les deux approches sont utilisées.

Les attributs utilisés sont en général de deux types ; ce sont soit des mesures effectuées en chaque point comme l'intensité lumineuse, des dérivées de l'intensité lumineuse ou des fonctions de ces dérivées (70-71) soit des attributs d'image moins denses, principalement des contours (11,63,64,66).

Les règles de sélection implémentent trois contraintes qui représentent des hypothèses "raisonnables" sur le monde réel et la position de l'observateur par rapport à celui-ci. Ces hypothèses sont que les objets présents ont des surfaces qui sont continues presque partout, en général non transparentes (à chaque point d'image correspond un point physique et un seul) et que l'observateur occupe par rapport aux objets un point de vue général. Ces règles sont appliquées en utilisant des statistiques d'attributs calculées dans des fenêtres ou le long de contours, en imposant que l'ordre de primitives appariées soit conservé dans les deux images le long de lignes épipolaires ou en utilisant les résultats d'appariements à des résolutions plus grossières pour limiter les dynamiques de disparité à des résolutions plus fines.

Les techniques basées sur la corrélation d'intensité lumineuse ont été utilisées essentiellement pour des applications partiellement interactives en stéréophotogrammétrie. Plus récemment les systèmes (61-62) de Moravec et Gennery développés pour la navigation de véhicules autonomes sont également de ce type. Le système de Baker et Binford (11-63) utilise à la fois des attributs de type intensité et des attributs de type contours. Le premier système de Marr et Poggio (65) conçu pour les stéréogrammes utilise des attributs d'intensité, celui de Medioni (64) des attributs de contours ; les deux utilisent des techniques de relaxation (72-74) pour lever les ambiguïtés. Enfin, le second système de Marr et Poggio (66) utilise les passages par zéro de la dérivée seconde des images à plusieurs résolutions comme primitives.

Tous ces systèmes nécessitent actuellement plusieurs heures de temps de calcul sur des machines relativement importantes pour traiter des images de résolution 500 x 500. Il est donc clair qu'ils ne sont pas à l'heure actuelle envisageables pour des applications de type robotique. Cependant si l'on considère (63) trois besoins spécifiques de la robotique qui sont 1) savoir si de la matière est présente dans un certain volume de l'espace, 2) connaître l'orientation, la distance et la surface approximative d'un objet dont on connaît à peu près

la position et 3) connaître la position dans l'espace d'un contour physique étant donné une indication approximative de sa position, je pense que au niveau des tâches 1) et 2) le transfert technologique peut commencer.

Systèmes d'acquisition actifs :

Comme nous avons vu précédemment, l'idée de la stéréoscopie active est d'éliminer le problème de l'ambiguïté de l'appariement des primitives en suréclairant un ensemble de points de la scène. Le lecteur trouvera un bon résumé des différentes techniques dans (75). Nous ne rentrerons pas plus dans les détails de ces capteurs et passons maintenant à la description d'un certain nombre de systèmes de RPI manipulant explicitement des modèles tri-dimensionnels.

Systèmes de RPI tri-dimensionnels :

A une exception près, le système ACRONYM de Stanford (76-77), tous les systèmes existants utilisent des capteurs actifs. Nous ne décrivons pas ce système ici à cause de son inadéquation actuelle aux problèmes spécifiques de la robotique.

Les autres systèmes évitent le difficile problème de la stéréo et utilisent comme données de base des images d'un type un peu spécial où chaque pixel contient les coordonnées dans l'espace du point physique correspondant.

Se posent alors des problèmes nouveaux en Vision par Ordinateur qui sont

- 1) celui du traitement de ces données géométriques
- 2) celui des attributs qu'on veut extraire de ces données
- 3) celui des représentations (au niveau données et au niveau modèles) qu'on veut utiliser pour résoudre les problèmes de RPI.

Examinons tout d'abord le problème de la représentation des objets. Celle-ci peut être soit hiérarchique soit homogène. Même si les représentations hiérarchiques à plusieurs niveaux de résolution sont potentiellement très séduisantes (78-79) elles ne sont pas encore couramment utilisées dans des systèmes opérationnels. Les représentations actuelles sont donc homogènes et constituées de lignes tracées sur les objets correspondant à des discontinuités de la normale (80) (l'équivalent des coins en deux dimensions) ou de morceaux de surfaces qui sont soit des cylindres généralisés (78,81,82) soit des plans (83-85) ou des quadriques (84-88).

Ces primitives sont extraites (c'est le problème 1) des données à partir d'opérateurs de détecteurs de contours classiques (80,89) pour les lignes ou à partir de techniques de transformée de Hough (85,87) ou de croissance de régions (31-35,88,90) pour les surfaces. Notons qu'un effort de recherche important est encore nécessaire au niveau de ces problèmes de segmentation.

Nous décrivons maintenant trois systèmes qui permettent de reconnaître et de positionner des objets tridimensionnels à partir de données 3-D obtenues par stéréoscopie active. Le plus ancien est celui de Ohshima et Shirai (90). Scènes et modèles sont représentés par des graphes dont les sommets sont les régions (des plans ou des quadriques (84)) et les arcs des relations entre régions (type d'intersection c'est-à-dire convexe, concave ou inexistante, angle entre régions et distance des centres de gravité). Le problème de la reconnaissance

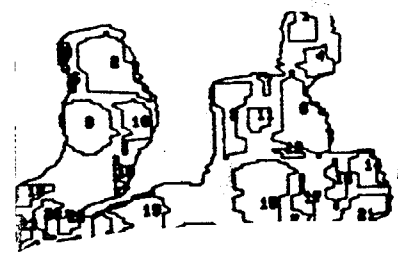
VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

d'un modèle dans la scène est donc un problème d'isomorphisme de sous-graphe : reconnaître un modèle c'est trouver un sous-graphe du graphe représentant la scène. Oshima et Shirai résolvent ce problème par une technique classique de recherche arborescente en profondeur.

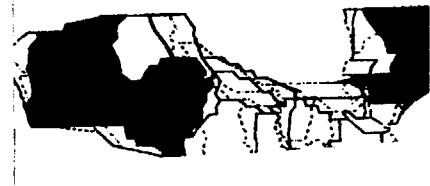
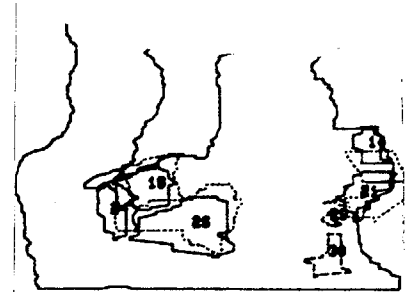
Le système a été testé sur quelques scènes contenant des objets réels et contenant 10 ou 20 primitives surfaciques. Les temps d'exécution sur une machine classique sont de l'ordre de trois minutes pour construire la description de la scène et de moins d'une seconde pour reconnaître un modèle. La précision du positionnement ne semble pas avoir été étudiée et les modèles utilisés sont très simples c'est-à-dire contenant un nombre très limité de primitives (moins de 10).

Le système de Bolles et Horaud (80,89) utilise des idées assez similaires à celles du système bidimensionnel de Bolles et Cain (45) décrit plus haut. Les primitives sont des segments de droite et des cercles et les surfaces planes ou cylindriques qui leur sont adjacentes. La méthode de reconnaissance et de positionnement utilise comme dans (45) une technique de génération et de test d'hypothèses. Rien n'est dit sur la précision du positionnement. Les programmes ont été développés en C sur un VAX-11/750, le temps de segmentation est très long (!) (89), le temps de reconnaissance de l'ordre de 10 secondes pour des scènes comportant trois objets.

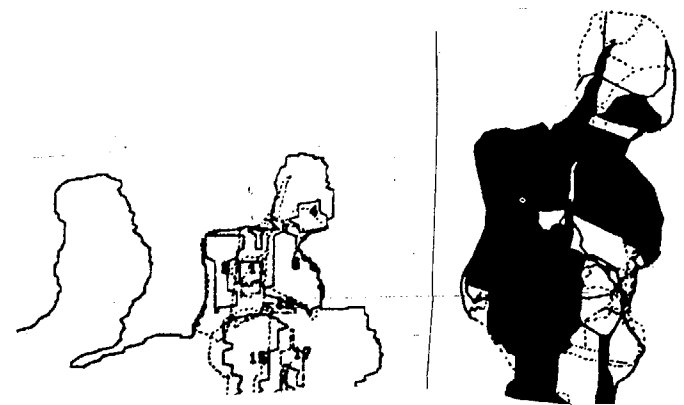
Le système de Faugeras et Hébert (91-93) utilise dans la version actuelle des primitives planes uniquement. La technique de reconnaissance et de positionnement est également une technique de génération et de test d'hypothèses implémentée par une recherche arborescente de type Branch-and-Bound. Une partie essentielle de l'algorithme de reconnaissance est l'estimation itérative de la transformation (Translation-Rotation) amenant le modèle reconnu sur la scène. Le système a été testé sur des scènes compliquées, comportant une trentaine de primitives, pour des objets occultés à 70 % et des modèles comportant une soixantaine de primitives. Les temps de reconnaissance et de positionnement sur un ordinateur Perkin-Elmer 32-40 et un programme Fortran sont de l'ordre de 10 secondes par modèle avec une précision absolue en translation de l'ordre du millimètre pour des objets d'une taille de l'ordre de dix centimètres à une distance de quatre mètres du capteur et une précision de l'ordre de 1 degré en rotation. La segmentation de la scène prend entre 10 secondes et une minute pour un nombre de primitives planes variant entre 10 et 50. Nous présentons en Figure 7 des résultats pour une scène relativement compliquée.



Résultat de la segmentation



Identification de la première pièce



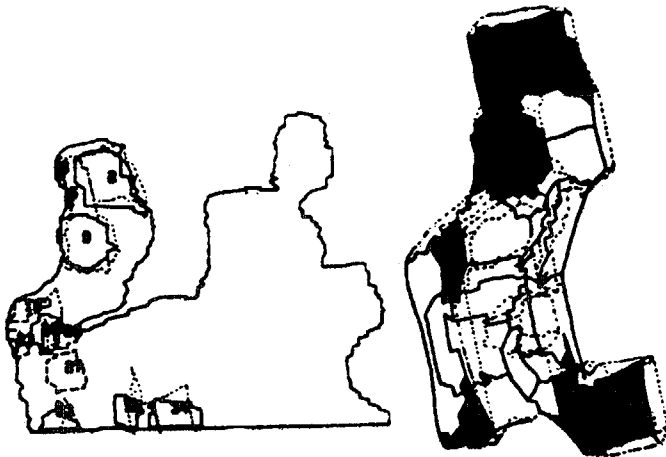
Identification de la deuxième pièce



Scène originale



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
COMPUTER VISION IN ROBOTICS : STATE OF THE ART
O.D. FAUGERAS



Identification de la troisième pièce

Figure 7

Conclusions :

De ce panorama rapide de la Vision par Ordinateur vu du point de la Robotique on peut tirer un certain nombre de conclusions. Tout d'abord il faut souligner le fait que la Vision par Ordinateur est d'ores et déjà présente en Robotique et que le besoin ne fait que s'accroître. Afin d'accélérer encore le transfert de technologie des laboratoires de recherche vers l'industrie il me paraît important d'identifier un certain nombre de besoins.

Au niveau des capteurs tout d'abord, on a besoin d'une meilleure précision géométrique et d'une meilleure définition spatiale ; il est nécessaire également de développer rapidement des systèmes d'acquisition 3-D actifs, rapides, précis et peu coûteux.

Au niveau de l'inspection qui constitue certainement à l'heure actuelle le plus gros marché pour la Vision en Robotique à la fois en 2-D et en 3-D il apparaît un besoin pressant de définition d'une méthodologie d'une part, et d'intégration dans des processus peu coûteux d'algorithmes existants d'autre part.

La commande en boucle fermée de manipulateurs par la Vision pour la Robotique industrielle nécessite également un gros effort d'intégration d'algorithmes dans des processeurs spécialisés rapides ; un autre besoin est celui de l'articulation du système de Vision sur une base de données CAO et son incorporation dans le langage de commande des manipulateurs ; enfin de la recherche est encore nécessaire pour développer des systèmes efficaces de stéréo passive ou plus généralement d'intégration de données provenant de plusieurs capteurs.

Au niveau des robots mobiles autonomes le besoin d'intégration des algorithmes est encore plus pressant vu la complexité accrue de l'environnement et une recherche de base sur l'apprentissage et les problèmes de génération de stratégies est encore nécessaire.

REFERENCES

- (1) J.L. Mundy, "Image Input Systems", dans Industrial Applications of Image Analysis, édité par N.J. Zimmerman et A. Oosterlinck, D.E.B. Publishers, The Netherlands, 1983.
- (2) R.T. Chin et C.A. Harlow, "Automated Visual Inspection : a Survey", IEEE Trans. on PAMI, Vol. PAMI-4, N° 6, p. 557-573, Novembre 1982.
- (3) "Industrial Applications of Image Analysis", Edité par N.J. Zimmerman et A. Oosterlinck, D.E.B. Publishers, The Netherlands, 1983.
- (4) Le numéro de Novembre 1983 des IEEE Transactions on PAMI est consacré aux "Applications industrielles de la Vision par Ordinateur", Vol. PAMI-5, N° 6, Novembre 1983.
- (5) A. Rosenfeld et A. Kak, "Digital Picture Processing", 2^{ème} Edition, Academic Press, 1982.
- (6) W. Pratt, "Digital Image Processing", Wiley, New York, 1978.
- (7) L. Davis, "A survey of edge detection techniques", Computer Graphics and Image Processing, 4, 1975, p. 248-270.
- (8) L.G. Roberts, "Machine perception of 3-Dimensional Solids", Optical and Electro-Optical Information Processing, J. Tippet, D. Berkowitz, L. Clapp, C. Koester, A. Vanderbergh Eds, M.I.T. Press, Cambridge, p. 159-197, 1965.
- (9) I.D.G. Macleod, "On finding structures in Pictures", Picture Language Machines, S. Kaneff Ed. Academic Press, New-York, p. 231, 1976.
- (10) E. Persoon, "A new edge detection algorithm and its applications", Computer Graphics and Image Processing 5, p. 425-446, 1976.
- (11) T.O. Binford, "Inferring surfaces from images", Artificial Intelligence, 17, p. 205-244, 1981.
- (12) R. Nevatia and R. Babu, "Linear feature extraction and description", Computer Graphics and Image Processing 13, p. 257-269, 1980.
- (13) J.W. Modestino and R.W. Fries, "Edge detection in noisy images using recursive digital filtering", Computer Graphics and Image Processing 6, p. 409-433, 1977.
- (14) D.C. Marr and E. Hildreth, "Theory of Edge Detection", Proc. R. Soc. Lond. B 207, p. 187-217, 1980.
- (15) E.C. Hildreth, "The detection of intensity changes by computer and biological vision systems", Computer Vision, Graphics and Image Processing, 22, p. 1-27, 1983.
- (16) J.F. Canny, "Finding Edges and Lines in Images", M.I.T. Artificial Intelligence Laboratory Technical Report N° 720, 1983.
- (17) N. Keskes, "Application des Techniques d'Analyse d'Images aux Signaux Sismiques", Thèse de Doctorat ès Sciences, Université Paris VI, Mars 1984.
- (18) J.M.S. Prewitt, "Object Enhancement and Extraction", Picture Processing and Psychopictorics, B. Lipkin and A. Rosenfeld Eds, Academic



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

- (19) Press, New-York, p. 75-149, 1970.
- (19) M.H. Hueckel,
 "An Operator which locates edges in digitized pictures",
 JACM 18, N° 1, p. 113-125, 1971.
- (20) R.M. Haralick,
 "Digital step edges from zero crossing of second directional derivatives",
 IEEE Trans. on PAMI, Vol. PAMI-6, N° 1, p. 58-68, 1984.
- (21) R.M. Haralick,
 "Statistical and structural approaches to texture",
 Proceedings of the IEEE, Vol. 67, N° 5, p. 786-804, Mai 1979.
- (22) O.D. Faugeras et W.K. Pratt,
 "Decorrelation methods of texture feature extraction",
 IEEE Transactions on PAMI, Vol. PAMI-2, N° 4, p. 323-332, 1980.
- (23) R. Bajcsy et L. Liebermann,
 "Texture gradient as a depth cue",
 Computer Graphics and Image Processing, Vol. 5, N° 1, p. 52-67, 1976.
- (24) B.H. Mc Cormick et S.N. Jayaramurthy,
 "Time series model for texture synthesis",
 Int. J. Comput. Inform. Sci., Vol. 3, N° 4, p. 329-343, Décembre 1974.
- (25) J.T. Tou, D.B. Kao et Y.S. Chang,
 "Pictorial texture analysis and synthesis",
 Third Int. Joint Conf. On Pattern Recognition, Coronado, Août 1976.
- (26) B. Julesz,
 "Visual pattern discrimination",
 IRE Trans. Inform. Theory, Vol. IT-8, p. 84-92, Février 1962.
- (27) A. Gagalowicz,
 "Vers un modèle de textures",
 Thèse de Doctorat ès Science, Université de Paris VI, 1983.
- (28) R.M. Haralick, K. Shanmugam et I. Dinstein,
 "Textural features for image classification",
 IEEE Trans. Syst., Man, Cybern, Vol. SMC-3, p. 610-621, Novembre 1973.
- (29) R.W. Connors, C.W. Mc Millin, Kingyao Lin et R.E. Vasquez-Espinosa,
 "Identifying and locating surface defects in wood : part of an automated lumber processing system",
 IEEE Transactions on PAMI, Vol. PAMI-5, N° 6, p. 573-583, Novembre 1983.
- (30) G. Matheron,
 "Eléments pour une théorie des milieux poreux",
 Paris, Masson, 1967.
- (31) J. Serra,
 "Mathematical morphology and image analysis",
 Academic Press, 1982.
- (32) K.I. Laws,
 "Textured image segmentation",
 University of Southern California, USCPI Report 940, 1980.
- (33) S. Tsuji et F. Tomita,
 "A structural analyzer for a class of textures",
 Proc. Third Int. Joint Conf. Artificial Intelligence, p. 564-571, 1973.
- (34) F. Tomita, M. Yachida et S. Tsuji,
 "Detection of homogeneous regions by structural analysis",
 Proc. Third Int. Joint Conf. Artificial Intelligence, p. 564-571, 1973.
- (35) L. Davis, S. Johns et J.K. Aggarwal,
 "Texture analysis using generalized co-occurrence matrices",
 Pattern Recognition and Image Processing Conf., Chicago, 1978.
- (36) A. Fournier, D. Fussell et L. Carpenter,
 "Computer rendering of stochastic models",
 CACM, Vol. 25, N° 6, p. 371-384, Juin 1982.
- (37) B.B. Mandelbrot,
 "Fractals : Form, Chance and Dimension",
 Freeman, San Fransisco, 1977.
- (38) N.J. Zimmerman, G.J.R. Van Boven et A. Oosterlinck,
 "Overview of industrial vision systems",
 publié dans la référence (3).
- (39) G.J. Agin et R.O. Duda,
 "SRI vision research for advanced industrial automation",
 Proc. 2nd USA-Japan Computer Conference, Tokyo, p. 113-117, 1975.
- (40) D. Juvin et B. Dupeyrat,
 "ANIMA, a quasi real time system",
 Proc. of PRIP 82, Las Vegas, p. 358-361, Juin 1982.
- (41) D. Juvin,
 "Contribution à la reconnaissance automatique des images appliquées à la robotique",
 Rapport CEA-R-5218, 1983.
- (42) W.A. Perkins,
 "A model-based vision system for industrial parts",
 IEEE Trans. On Computers, Vol. C-27, N° 2, p. 126-143, Février 1978.
- (43) J.D. Dessimoz, M. Kunt, G.H. Granlund et J.M. Zurcher,
 "Recognition and handling of overlapping industrial parts",
 Proc. 9th Int. Symp. On Indust. Robots, Washington D.C., p. 357-366, 1979.
- (44) J.D. Dessimoz,
 "Traitement des contours en reconnaissance de formes visuelles, applications en robotique",
 Thèse N° 387, Ecole Polytechnique Fédérale de Lausanne, 1980.
- (45) R.C. Bolles et R.A. Cain,
 "Recognising and locating partially visible objects : the local-feature focus method",
 dans Robot Vision édité par A. Pugh, IP'S Publications Ltd, U.K., Springer Verlag, p. 43-82, 1983.
- (46) H.G. Barrow, A.P. Ambler et R.M. Burstall,
 "Some techniques for recognising structures in pictures",
 dans Frontiers of Pattern Recognition, édité par S. Watanabe, p. 1-29, 1973.
- (47) W. Hättich, W. Schwerdtmann et H. Tropp,
 "Experience with two hybrid systems for the recognition of overlapping workpieces",
 Proc. 6th Int. Conf. On Pattern Recognition, Munich, p. 670-673, 1982.
- (48) P. Rummel et W. Beutel,
 "A model-based image analysis system for workpiece recognition",
 ibid p. 1014-1017.
- (49) P. Rummel et W. Beutel,
 "Workpiece recognition and inspection by a model-based scene analysis systems",
 Pattern Recognition, Vol. 17, N° 1, p. 141-148, 1984.
- (50) N.J. Nilsson,
 "Problem solving methods in artificial intelligence",
 Mc Graw-Hill, New York, 1971.
- (51) V. Souvignier,
 "PVV - un système d'interprétation d'images par prédiction et vérification",
 Thèse de Docteur de Troisième Cycle, Institut National Polytechnique de Grenoble, Juin 1983.



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

- (52) A. Lux et V. Souvignier,
 "PVV - Un système de vision appliquant une stratégie de Prédiction-Vérification",
 4^e Congrès Recon. Des Formes et Intel. Artificielle, Paris, p. 223-234, 1984.
- (53) N. Ayache et C. Darmon,
 "Reconnaissance récursive et localisation de formes planes partiellement visibles dans une image plane",
 9^{ème} colloque GRETSI, Nice, p. 611-617, 1983.
- (54) N. Ayache,
 "Un système de vision bidimensionnelle en robotique industrielle",
 Thèse de Docteur Ingénieur, Université Paris Sud, 1983 (Rapport INRIA IS9N 2-7261-0345-6).
- (55) N. Ayache, B. Faverjon, J.D. Boissonnat et B. Bollack,
 "Manipulation automatique de pièces industrielles en vrac planaire",
 Actes de ce colloque, 1984.
- (56) B.K.P. Horn,
 "Obtaining shape from shading information", dans The Psychology of Computer Vision, édité par P.H. Winston, New York, Mc Graw-Hill, p. 115-155, 1975.
- (57) K. Ikeuchi et B.K.P. Horn,
 "Numerical shape from shading and occluding boundaries", dans Computer Vision, édité par M. Brady, Amsterdam, North-Holland, p. 141-184, 1981.
- (58) R.J. Woodham,
 "Reflectance map techniques for analysing surface defects in metal casting",
 Rapport du Laboratoire d'Intelligence Artificielle du MIT AI-TR-457, Cambridge, MIT, USA, 1978.
- (59) K. Ikeuchi,
 "Determining surface orientations of specular surfaces by using the photometric stereo method",
 IEEE Trans. on PAMI, Vol. PAMI-2, N° 6, p. 661-669, 1981.
- (60) A.P. Witkin,
 "Recovering surface shape and orientation from texture", dans Computer Vision, édité par M. Brady, Amsterdam, North-Holland, p. 17-47, 1981.
- (61) H.P. Moravec,
 "Robot Rover Visual Navigation",
 UMI Research Press, 1980.
- (62) D.B. Gennery,
 "Modeling the environment of an exploring vehicle by means of stereo vision",
 Stanford Artificial Intelligence Laboratory Memo 339, 1980.
- (63) H.H. Baker et T.O. Binford,
 "Depth from edge -and intensity- based stereo",
 Proc. 7th Intern. Joint Conf. On A.I., Vancouver, British Columbia, p. 631-636, 1981.
- (64) G. Medioni,
 "Matching images using linear features",
 Rapport Intelligent Systems Group de l'Université de Californie du Sud, ISG Report 103, 1983.
- (65) D. Marr et T. Poggio,
 "Cooperative computation of stereo disparity",
 Science, 194, p. 283-287, 1976.
- (66) D. Marr et T. Poggio,
 "A computational theory of human stereo vision",
 Proc. R. Soc. Lond. B, 204, p. 301-328, 1979.
- (67) R. Ray, J. Birk et R.B. Kelley,
 "Error analysis of surface normals determined by radiometry",
 IEEE Trans. On PAMI, Vol. PAMI-5, N° 6, p. 631-644, 1983.
- (68) H.K. Nishihara et T. Poggio,
 "Stereo Vision for Robotics",
 Proc. First International Symposium of Robotics Research, Brettenwoods, USA, 1983.
- (69) K. Ikeuchi, B.K.P. Horn et S. Nagata,
 "Picking up an object from a pile of objects",
 ibid.
- (70) M. Kass,
 "A computational framework for the visual correspondence problem",
 Proc. ARPA Image Understanding Workshop, Washington D.C., 1983.
- (71) S. Castan et J. Shen,
 "A stereo vision algorithm taking into account the perspective distortions",
 soumis à 7th Int. J. Conf. On Pattern Recognition, Montréal, 1984.
- (72) A. Rosenfeld, R. Hummel et S. Zucker,
 "Scene labeling by relaxation operations",
 IEEE Trans. On SMC, 6, p. 420-433, 1976.
- (73) O.D. Faugeras et M. Berthod,
 "Improving consistency and reducing ambiguity in stochastic labeling: an optimization approach",
 IEEE Trans. On PAMI, Vol PAMI-3, p. 245, 1981.
- (74) R.A. Hummel et S.W. Zucker,
 "On the foundations of relaxation labeling processes",
 IEEE Trans. On PAMI, Vol. PAMI-5, p. 267-297, 1983.
- (75) R.A. Jarvis,
 "Perspective on range finding techniques for computer vision",
 IEEE Trans. On PAMI, Vol. PAMI-5, N° 2, p. 122-139, 1983.
- (76) R.A. Brooks,
 "Symbolic reasoning among 3-D models and 2-D images",
 Artificial Intelligence, Vol. 17, p. 285-348, 1981.
- (77) R.A. Brooks,
 "Model-based three-dimensional interpretations of two-dimensional images",
 IEEE Trans. on PAMI, Vol. PAMI-5, p. 140-150, 1983.
- (78) D. Marr et H.K. Nishihara,
 "Representation and recognition of the spatial organization of three-dimensional shapes",
 Proc. R. Soc. B, 200, p. 269-294, 1978.
- (79) O.D. Faugeras et J. Ponce,
 "Prism trees: a hierarchical representation for 3-D objects",
 Proc. 8th Int. J. Conf. On Artificial Intelligence, Karlsruhe, Allemagne, p. 982-988, 1983.
- (80) R.C. Bolles, P. Horaud et M.J. Hannah,
 "3DPO: a three dimensional part orientation system",
 ibid, p. 1116-1120.
- (81) G.J. Agin et T.O. Binford,
 "Computer description of curved objects",
 IEEE Trans. Comput., Vol. C-25, p. 439-449, 1976.
- (82) R. Nevatia et T.O. Binford,
 "Description and recognition of curved objects",
 Artificial Intelligence, Vol. 8, p. 77-98, 1977.
- (83) R.O. Duda, D. Nitzan et P. Barret,
 "Use of range and reflectance data to find



VISION PAR ORDINATEUR EN ROBOTIQUE : ETAT DE L'ART
 COMPUTER VISION IN ROBOTICS : STATE OF THE ART
 O.D. FAUGERAS

- planar surface regions",
 IEEE Trans. On PAMI, Vol. PAMI-1, N° 3, p.
 259-271, 1979.
- (84) M. Oshima et Y. Shirai,
 "A scene description method using
 three-dimensional information",
 Pattern Recognition, Vol. 11, p. 9-17,
 1979.
- (85) M. Hébert et J. Ponce,
 "A new method for segmenting 3-D scenes
 into primitives",
 Proc. 6th Int. J. Conf. On Pattern
 Recognition, p. 836-838, Munich, Allemagne,
 1982.
- (86) O.D. Faugeras, M. Hébert et E. Pauchon,
 "Segmentation of range data into planar and
 quadratic patches",
 Proc. CVPR 83, Washington D.C., USA, p.
 8-13, 1983.
- (87) Y. Muller et A. Mirgoux,
 "Détection de quadriques par transformée de
 Hough",
 Actes du 4ème Congrès de Reconnaissance des
 formes et d'intelligence artificielle, Vol.
 1, Paris, p. 315-326, 1984.
- (88) C.A. Dane III,
 "An object centered three-dimensional model
 builder",
 Thèse de Ph. D, Opt. De Computer Science,
 Univ. De Pennsylvanie, 1982.
- (89) P. Horaud, R.C. Bolles et C. Barrowil,
 "Reconnaissance d'objets 3-D par
 interprétation d'informations partielles à
 l'aide d'un modèle",
 Actes du 4ème Congrès de Reconnaissance des
 formes et d'intelligence artificielle, Vol.
 1, Paris, p. 327-336, 1984.
- (90) M. Oshima et Y. Shirai,
 "Object recognition using three-dimensional
 information",
 IEEE Trans. on PAMI, Vol. PAMI-5, N° 4, p.
 353-361, 1983.
- (91) O.D. Faugeras et M. Hébert,
 "A 3-D recognition and positioning
 algorithm using geometrical matching
 between primitive surfaces",
 ibid., p. 996-1002.
- (92) O.D. Faugeras, M. Hébert, J. Ponce et E.
 Pauchon,
 "Object representation, identification and
 positioning from range data",
 Proc. 1st Int. Symp. Of Robotics Research,
 Brettenwoods, USA, 1983.
- (93) M. Hébert,
 "Reconnaissance de formes
 tridimensionnelles",
 Thèse de Doctorat de 3ème cycle, Université
 de Paris Sud, 1983.