

PREMIER COLLOQUE IMAGE  
Traitement, Synthèse, Technologie et Applications

BIARRITZ - Mai 1984 -

---

ANALYSE DE SCENES PAR UNE METHODE DU TYPE "K2D"  
POUR LA MANIPULATION DE PIECES EN ROBOTIQUE AVANCEE

SCENE ANALYSIS WITH A "K2D" TYPE METHOD  
FOR PART MANIPULATION IN ADVANCED ROBOTICS.

K. BEN RHOUMA - L. PERALTA - L. HENNINGER - A. OSORIO

LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur) B.P.30 - F 91406 ORSAY Cedex

---

**RESUME**

Nous décrivons le système de vision stéréoscopique appliqué à la Robotique, utilisé au LIMSI. Ce système constitue l'essentiel du module de perception utilisé pour la manipulation d'objets. Les opérations d'assemblage automatique seront assurées par des capteurs d'efforts, l'information géométrique étant fournie par le module de vision.

Les caractéristiques de ce système sont les suivantes :

- utilisation de la méthode "K2D",
- positions des caméras déterminées à partir des images de quatre points d'étalonnage,
- angle de vision de la scène quelconque (30°-60°),
- images de 256 x 256 x 4 bits,
- les objets sont approximés par des polyèdres.

Les images sont d'abord traitées afin d'extraire les contours. Puis les contours sont approchés par des segments de droites. L'algorithme hongrois est utilisé pour trouver les segments homologues. La précision des résultats est limitée principalement par les dimensions des pixels et par la précision de l'extracteur de contours.

**SUMMARY**

A three dimensional vision system used for depth perception in robotics is described. This system is based on stereo-vision. It is used as a sensor for manipulator-type robots. The automated assembly operations will be carried out using other types of sensors as force sensors, but the geometrical informations are provided by the vision process.

The characteristics of our system are :

- the use of a "K2D" method,
- the positions of the cameras found using four calibration points,
- the vision angle is not fixed (between 30° and 60°),
- images of 256 x 256 x 16 grey levels,
- the objects are approximated by polyhedra.

First, edges are extracted from the images ; then they are approximated by segments. The hungarian algorithm is used to find homologous segments. The precision of the results are limited mainly by the resolution of the images and on the precision of the edge operator.



## INTRODUCTION

L'étape la plus délicate dans l'analyse de scènes est probablement le filtrage des informations dont on dispose en entrée, sans pour cela perdre les caractéristiques générales de l'image.

Nous avons abordé la conception du système de vision dans le cadre du projet robotique du LIMSI. Le taux considérable de problèmes industriels purement bidimensionnels en perception optique (probablement 70%) nous a amenés à approfondir les techniques 2D et, par ce biais, à mettre en oeuvre une méthodologie de ce type conduisant à identifier la notion de relief dans les scènes analysées. Pour ce faire trois étapes étaient nécessaires.

Nous avons d'abord cherché à filtrer l'image. Cela se fait par un extracteur de contour élaboré au LIMSI suivi d'une segmentation et d'un raccordement de tous les segments obtenus.

D'autre part, nous avons élaboré un module d'étalonnage du système de vision nous permettant de maîtriser les relations taille physique/taille image et les positions des caméras. Ce module est simple d'utilisation puisqu'il suffit que quatre points connus appartenant au plan de travail soient sur le moniteur.

Enfin, afin d'accéder au relief de la scène, nous avons utilisé une méthode de stéréovision du type K2D. Nous avons étudié un algorithme de recherche de segments homologues basé sur l'algorithme hongrois. La troisième dimension est alors obtenue par triangulation.

## ETALONNAGE DU SYSTEME DE VISION

Si on accepte une approximation "pin-hole" de la caméra (14), les coordonnées (u,v) d'un point appartenant à la surface de travail et les coordonnées (x,y) du point correspondant dans le plan de l'image sont reliées par la transformation bilinéaire :

$$u = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1}$$

$$v = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1}$$

Les coefficients  $a_1, \dots, a_8$  dépendent de la position de la caméra et son orientation, la position de l'axe optique dans l'image et le facteur d'échelle sur l'image.

Afin de déterminer les 8 coefficients de la transformation bilinéaire il suffit de connaître la position de 4 points sans triade colinéaire sur le plan de travail et la position de leurs points image. Néanmoins, ces coefficients ne permettent pas de déduire la position de la caméra.

Pour avoir accès à la position de la caméra il faut une donnée ou une hypothèse supplémentaire. En réalité, il est possible de déterminer la position de la caméra en utilisant moins de 4 points et plus d'hypothèses supplémentaires. Par exemple, on peut supposer connue la position de l'axe optique dans l'image. Si en plus de cette hypothèse on suppose connue la position du centre optique par rapport au plan image et la correspondance longueur en mm-pixel sur le plan image, le nombre de points nécessaire à l'étalonnage serait encore plus faible.

La position de la caméra peut être déterminée avec seulement quatre points d'étalonnage si on fait une hypothèse supplémentaire sur le facteur d'échelle,

la position de l'axe optique est toujours inconnue. En effet, nous avons supposé que sur le plan image, les échelles longueurs-pixel sont les mêmes dans les sens horizontal et vertical.

## DETERMINATION DE LA POSITION DE LA CAMERA

Dans la Figure 5, nous avons représenté le plan de travail et le plan image. En général, le plan de la figure n'est pas perpendiculaire au plan de travail, ni à la ligne horizontale et ne contient pas l'axe optique. Par contre il contient le segment s et le segment  $s_1$ . On constate que le rapport des longueurs de ces deux segments est égal à  $K = \frac{l_2}{l_1}$ . Ce rapport est

inconnu mais si on fait l'hypothèse supplémentaire que les échelles horizontale et verticale sont les mêmes on peut dire qu'il est le même pour toutes les configurations qui ont le point B en commun. Si on prend trois configurations différentes mais avec le point B commun, Figure 5, les distances  $d_1, d_2$  et  $d_3$  de la caméra à ligne d'horizon réciproque pour chacun des trois cas, satisfont les relations suivantes :

$$d_1 = Kb_1$$

$$d_2 = Kb_2$$

$$d_3 = Kb_3$$

Le lieu des points qui satisfont à ces relations est un cercle perpendiculaire d'une part à la ligne d'horizon réciproque et d'autre part au plan de travail. Son rayon est facilement obtenu à partir des trois rapports  $\frac{l_i}{l_j}$   $i \neq j$   $i = 1,2,3$   $j = 1,2,3$ . L'axe optique de la caméra est contenu dans le plan du cercle et est tangent à celui-ci.

Il est à remarquer que pour trouver ce cercle, il n'est pas nécessaire de faire d'hypothèse ni sur la position de centre de l'image, ni sur la focale de la caméra. Il s'agit de trouver à quel endroit du cercle se trouve la caméra.

Pour illustrer comment ceci peut être fait imaginons dans l'image une croix symétrique dont l'axe vertical est contenu dans la droite homologue de celle qui contient le diamètre du demi-cercle sur lequel se trouve la caméra. Les extrémités de cette croix seront vues par la caméra comme indiqué dans la Figure. L'hypothèse d'égalité des échelles verticale et horizontale implique que  $tg\delta_2 - tg\delta_1 = tg\delta_3 + tg\delta_4$  (Fig.8). Cela détermine la position du centre optique de la caméra c'est-à-dire l'intersection de l'axe optique avec le plan focal frontal de l'objectif de la caméra.

## PRETRAITEMENT DE L'IMAGE

### 1) EXTRACTION DE CONTOUR (5) :

Nous utilisons un opérateur d'extraction de contour conçu dans notre laboratoire qui est peu sensible au bruit. Si on désigne par  $f(x,y)$  la fonction image obtenue par un dispositif de digitalisation, deux étapes sont nécessaires :

1ère étape : Dans cette première étape nous calculons un gradient isotropique dans une fenêtre  $3 \times 3$  :

$$g(i,j) = X^2 + Y^2 \quad \text{avec} \quad X = X_2 - X_1 \quad \text{et} \quad Y = Y_2 - Y_1$$

$$X_1 = F(i-1,j-1) + \sqrt{2} F(i,j-1) + F(i+1,j-1)$$

$$X_2 = F(i-1,j+1) + \sqrt{2} F(i,j+1) + F(i+1,j+1)$$

$$Y_1 = F(i-1,j-1) + \sqrt{2} F(i-1,j) + F(i-1,j+1)$$

$$Y_2 = F(i+1,j-1) + \sqrt{2} F(i+1,j) + F(i+1,j+1)$$

Nous fixons un seuil  $t_0$  nous permettant de détecter les points de contour avec du bruit.

2ème étape : Pour chaque point détecté précédemment, tel que  $g(i,j) > t_0$ , nous calculons la direction du gradient  $\alpha = tg(Y/X)$  et nous le classons selon un système à 6 directions angulaires. Pour chacune des directions correspond un masque 5 x 5 contenant une direction théorique. Compte tenu du masque théorique, nous calculons un second gradient égal à la différence des sommes des valeurs des intensités des pixels de part et d'autre du contour :

$$g'(i,j) = \left\{ \sum_g F(m,n) - \sum_d F(m,n) \right\}^2$$

Le filtrage consiste à retenir le point (i,j), détecté et classifié suivant la n<sup>ème</sup> direction dans la première étape, seulement si le second gradient satisfait la relation :

$$a g(i,j) < g'(i,j) < b g(i,j) ;$$

La première partie de l'inégalité sert à éliminer les points bruités et la deuxième partie sert à réduire la bande de points.

2) OBTENTION DES SEGMENTS (3) :

Tout d'abord, nous procédons à une recherche de points voisins formant une chaîne  $C_i$ . Tous les points d'une chaîne ont des gradients de direction proche.

Ensuite, nous approximons cette chaîne de points par le meilleur segment de moindres carrés. Si l'erreur statistique est trop grande, nous divisons la chaîne en deux en utilisant comme nouveau sommet le point le plus éloigné du segment. Cette opération est répétée jusqu'à obtention de segments satisfaisants ; si le nombre de points de la chaîne devient trop petit, la chaîne est considérée comme un bruit.

3) RACCORDEMENT DES SEGMENTS :

Les quatre configurations principales retrouvées dans une image de contour avant raccordement des segments sont représentées dans la Figure 1.

Pour les cas de Figure a et b, les deux segments sont remplacés par un seul segment équivalent si le rapport des moments d'inertie longitudinal et transverse est inférieur à un seuil donné. Pour la Figure 1-c et 1-b, le nouveau sommet des segments serait le centre de gravité de l'ensemble des sommets voisins. Les petits segments sont éliminés.

RECHERCHE DES SEGMENTS HOMOLOGUES

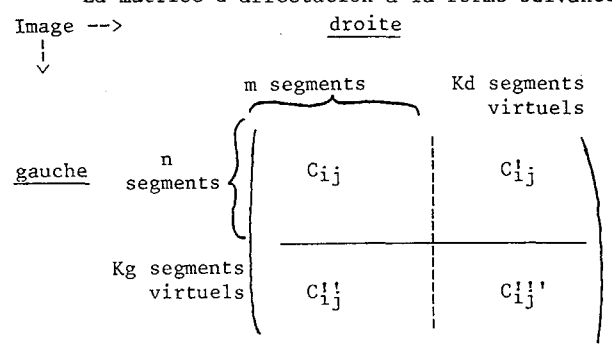
Les traitements étalonnage et prétraitement cités dans les paragraphes précédents sont effectués sur l'image de gauche ainsi que celle de droite.

Les segments de l'image de gauche ainsi que ceux de l'image de droite sont projetés sur le plan de travail à l'aide de la transformation bilinéaire (Figure 3). Ceci permet de chercher les segments homologues même si le cadrage respectif des deux images est fortement différent.

Pour la recherche des segments homologues, nous avons utilisé l'algorithme hongrois légèrement modifié. L'algorithme hongrois est un algorithme applicable à un cas particulier de problèmes de flots, plus précisément au problème de transport sans capacité. C'est un algorithme qui permet d'affecter n éléments constituant l'ensemble de départ à n éléments constituant l'ensemble d'arrivée en minimisant la fonction de coût. Celle-ci étant la somme des coûts des affectations individuelles.

Dans notre cas, il fallait d'abord définir une fonction de coût convenable, et adapter l'algorithme. En effet, le nombre de segments n'est pas nécessairement le même dans les deux images, et certains segments ne possèdent pas d'homologues dans l'autre image. Pour résoudre ce problème, nous avons ajouté un certain nombre de segments fictifs qui permettent d'une part de rendre carrée la matrice de coût et d'autre part de permettre la non affectation de certains segments.

La matrice d'affectation a la forme suivante :



Les éléments de la matrice  $C_{ij}''$  sont tous nuls car cela reflète bien l'affectation d'un segment virtuel à un segment virtuel.

Les éléments  $C_{ij}'$  et  $C_{ij}''$  représentent le coût pour qu'un segment quelconque n'ai pas d'homologue. Le regret pour ne pas trouver de segment homologue a été pris proportionnel à la longueur du segment.

Les éléments  $C_{ij}$  correspondent à l'affectation de deux segments des deux images. Le coût d'une telle affectation est une combinaison linéaire de la valeur absolue de la différence des coordonnées des sommets correspondants. Il est évident que le coefficient de coût pour des disparités dans le sens vertical est beaucoup plus élevé que les autres.

L'avantage fondamental de l'algorithme hongrois est sa vitesse de convergence : dans la majorité des cas, il arrive à la solution optimale à la première itération.

Les résultats des correspondances sont montrés dans la Figure 3. Deux segments homologues sont appelés par la même lettre.

LOCALISATION TRIDIMENTIONNELLE DES SEGMENTS

Avant de calculer les coûts utilisés dans l'algorithme hongrois, tous les segments étaient préalablement orientés dans les sens haut-bas et gauche-droite. Ce qui permet, une fois l'affectation obtenue, de calculer la position tridimensionnelle des sommets par simple triangulation puisque les positions des deux caméras ont été préalablement déterminées.

Dans la Figure 4, nous avons représenté les trois projections sur les trois plans du repère de référence. Ce sont les projections de l'image de trois cubes (Figure 2).

LE DISPOSITIF EXPERIMENTAL

Notre dispositif expérimental comporte deux robots à 6 degrés de liberté : un robot ACMA TH8-Renault et un robot transfert dont la structure a été construite par Micro-System. Les moyens informatiques sont constitués par un mini-ordinateur IBM7 relié canal à canal à un ordinateur Hote:le NAS du CIRCE.

Le système de vision est formé par deux camé-



ANALYSE DE SCENES PAR UNE METHODE DU TYPE "K2D"  
POUR LA MANIPULATION DE PIECES EN ROBOTIQUE AVANCEE

BEN RHOUMA  
L. PERALTA  
L. HENNINGER  
A. OSORIO

SCENE ANALYSIS WITH A "K2D" TYPE METHOD  
FOR PART MANIPULATION IN ADVANCED ROBOTICS.

ras de surveillance (Thomson-TAV 1030) reliées à un digitalisateur d'image NIM 628. L'image est digitalisée en 16 niveaux de gris et 256 x 256 pixels. Le NIM 628 optimise la digitalisation en faisant correspondre les 16 niveaux de gris aux variations du signal vidéo sur toute l'image. Le NIM 628 est relié à l'IBM 7.

Le système de vision a été testé dans plusieurs expériences canoniques d'assemblage. Les caméras sont disposées à l'extérieur du volume du travail des robots et regardent la scène sous un faible angle (environ 30°)

PRECISION DU SYSTEME DE VISION (13) :

Dans un système de vision idéal, la précision dépend seulement de la taille des pixels. En utilisant une caméra fixe telle que son axe optique est perpendiculaire au plan de travail, la précision obtenue est uniforme sur toute l'image. Si la caméra est inclinée, la précision décroît du côté éloigné de l'image.

Dans un système de vision "K2D", la précision obtenue pour la troisième dimension est fondamentalement plus mauvaise que celle obtenue sur le plan de travail. La raison est que l'angle formé par les axes optiques des deux caméras est faible (souvent de quelques dixièmes de radians) et les erreurs dans la détermination du point projeté sont multipliées par un facteur  $\sim \alpha^{-1}$  ( $\alpha$  est l'angle formé par les deux axes optiques). Si on augmente l'angle entre les deux axes optiques, on augmente la précision ; mais, la recherche de segments homologues devient de plus en plus difficile. Un compromis entre ces deux contraintes doit être fait. Nous avons pris un angle d'environ 0,3 radians.

En pratique, deux facteurs peuvent dégrader la précision : les non-linéarités électriques et optiques des caméras et les erreurs dues à l'étalonnage (détermination des points d'étalonnage, mauvais réglage des rapports des amplitudes des balayage horizontal et vertical).

Dans notre système, la précision obtenue est limitée par les erreurs dues à l'obtention et le raccordement des segments dans les deux images et surtout par la taille du pixel. La raison est que 256 x 256 est une faible résolution. Les faibles non-linéarités des caméras sont compensées par la méthode d'étalonnage.

Les erreurs observées dans l'obtention des segments sont en partie dues à la faible résolution en niveau de gris.

CONCLUSION

Nous avons décrit la méthode utilisée par l'analyse de scènes tridimensionnelle et appliquée dans des expériences d'assemblage. Cette méthode est relativement indépendante des conditions d'éclairage et peut être implémentée dans un petit ordinateur. Elle est utilisée comme un premier système de perception.

Les problèmes de dérive des caméras vidéo, ainsi que ceux provenant de leur manque de linéarité, sont partiellement résolus par le lancement d'une séquence d'étalonnage juste avant d'aborder l'analyse de la scène. L'accumulation d'erreurs conduit à la détermination de centres de gravité avec une précision qui est de l'ordre de 5 mm. La démarche choisie au Laboratoire consiste à rattraper cette erreur par l'intermédiaire de l'organe de préhension. En effet, l'objet étant indentifié, sa prise est réglée de manière à pouvoir assurer un recentrage selon deux axes orthogonaux, ce qui permet, après manipulation, de connaître la po-

sition des objets avec la précision de positionnement de l'outil de préhension.

Pour des applications industrielles, cette méthode a l'avantage de fournir un nombre important d'informations à partir de données produites par des capteurs peu coûteux.

Les informations obtenues par le système de vision sont utilisées pour résoudre le problème de contournement d'obstacles. Le processus proposé sera implémenté dans une unité spécialisée constituant un module autonome de perception stéréoscopique.

REFERENCES

1. R.DUDA, P.HART : "Pattern classification and scene analysis", Wiley 1973.
2. D.MARR : "Representing visual information, a computational approach", in HANSON A., RESEMAN E. "Computer vision systems", Ac. Press 1978.
3. A.IKONOMOPOULOS, A.OSORIO : "A general operator for the analysis of line drawings and its use for edge detection", EUSIPCO-80, 1<sup>st</sup> European Signal Processing Conference, Lausanne, Sept. 16-19, 1980.
4. G.BARROW, J.TENENBAUM : "Recovering intrinsic scene characteristics from images", in HANSON A., RESEMAN E., Computer vision systems, Ac. Press 1978.
5. A.IKONOMOPOULOS, A.OSORIO : "An approach to edge detection and representation by two-knowledge-free techniques", 5<sup>th</sup> Int.Conf.on Pattern Recognition, Miami, Dec.1-4, 1980.
6. A.OSORIO, K.BEN RHOUMA, D.TEIL, J.RIVAILLIER : "Un système de perception 3D pour un robot utilisant une caméra vidéo", in "Proceeding of the IASTED Intern. Symposium ROBOTICS", Davos, March 2-5, 1982.
7. MARTIN D.LEVINE : "Computer determination of depth maps", Computer Graphics and Image Processing 2, 1973, 131-150.
8. BARNARD, THOMPSON : "Disparity analysis of images", on Pattern anal. & mach. intell., PAMI-2, n°4, July 80.
9. Y.SHIRAI : "A context sensitive line finder for recognition of polyedre", A.I.4(2).
10. A.ROSENFELD, A.KAK : "Digital picture processing, Ac.Press 1978.
11. D.WALTS : "Understanding line drawings of scene with shadows", in P.WINSTON, "The psychology of computer vision", Mc Graw, 1975.
12. R.NEVATIA : "Depth Measurement by motion stereo", 6 Ca IP 5, 203-214, 1976.
13. K.BEN RHOUMA, L.PERALTA, A.OSORIO : "A "K2D" perception approach for assembly robots", EUSIPCO-EURASIP, Erlangen, Sept. 1983.
14. J.HWANG : "Computer stereo-vision for three dimensional object location", PH.D. 1980, The University of Tennessee.

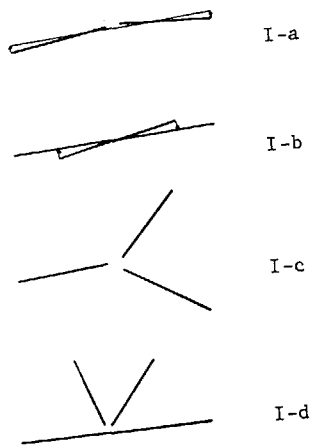
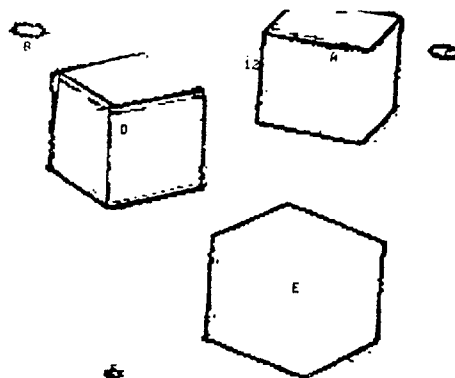


Figure 1



LA CAMERA GAUCHE EST EN (X,Y,Z) = -57 32 -1854 53 956.79

Fig.2 Contour et segments de l'image de gauche

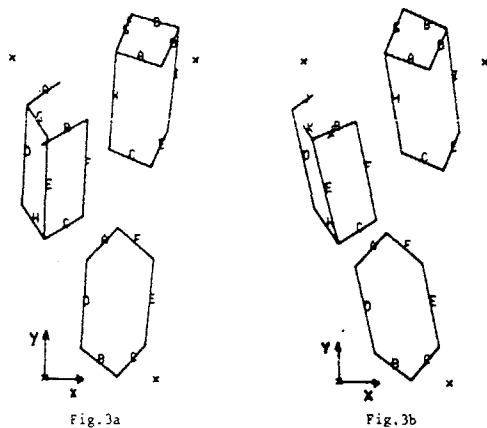


Fig.3a

Fig.3b

Fig.3 Projection des contours des objets sur le plan de travail  
Fig.3a : à partir de la caméra de gauche      Fig.3b : à partir de la caméra de droite

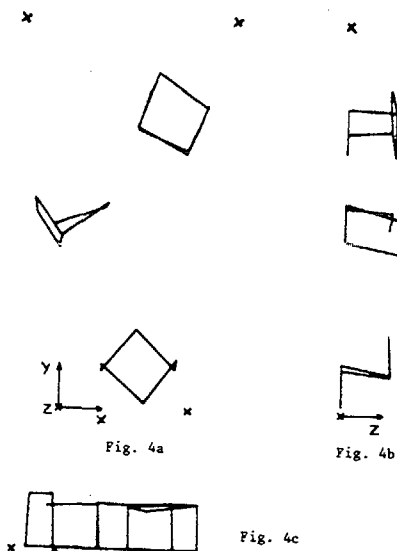


Fig. 4a

Fig. 4b

Fig. 4c

Fig. 4 : Projection des points tridimensionnels calculés des objets sur les trois plans :  
Fig. 4a : plan (x,y) - Fig. 4b : plan (z,y) - Fig. 4c : plan (x,z)

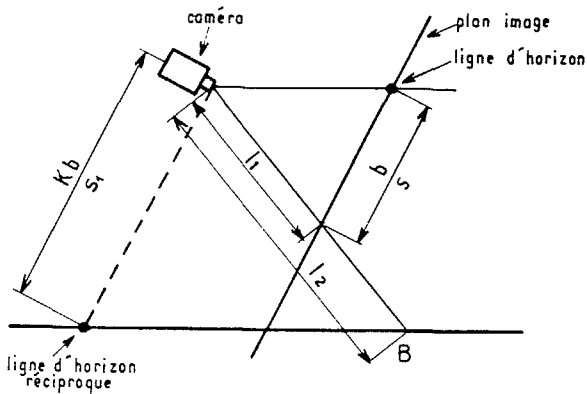


Figure 5

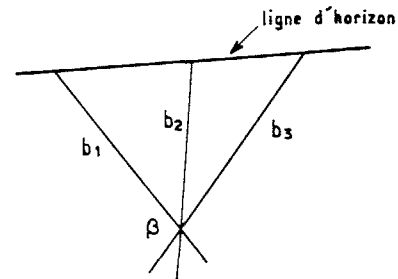


Figure 6



ANALYSE DE SCENES PAR UNE METHODE DU TYPE "K2D"  
 POUR LA MANIPULATION DE PIECES EN ROBOTIQUE AVANCEE

K. BEN RHOUMA  
 L. PERALTA  
 L. HENNINGER  
 A. OSORIO

SCENE ANALYSIS WITH A "K2D" TYPE METHOD  
 FOR PART MANIPULATION IN ADVANCED ROBOTICS.

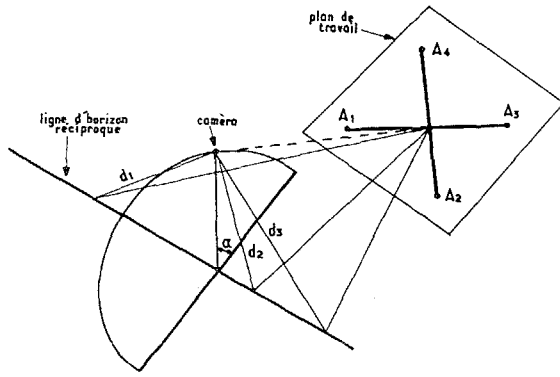


Figure 7

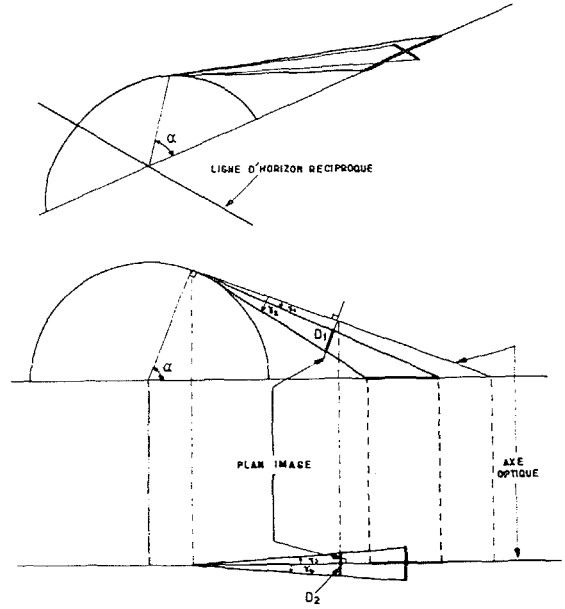


Figure 8