



Traitement, Synthèse, Technologie et Applications

BIARRITZ - Mai 1984 -

PROCEDURE D'APPRENTISSAGE POUR L'IDENTIFICATION D'OBJETS PLANS
PARTIELLEMENT OBSERVES
LEARNING PROCEDURE FOR THE IDENTIFICATION OF PARTIALLY OBSERVED
PLANAR OBJECTS

Abderrahman RIAD - Michel DEVY

LABORATOIRE d'AUTOMATIQUE et d'ANALYSE des SYSTEMES du C.N.R.S.
7, avenue du Colonel Roche
31400 TOULOUSE

RESUME

La communication présentée concerne la reconnaissance d'objets plans partiellement observés, déformés ou à structure déformable ; l'organe de perception étant une caméra.

L'image numérisée est réduite aux contours des objets qu'elle contient. Un traitement permet d'extraire un descripteur de type fonction de courbure, de chaque contour considéré.

Soient N classes d'objets à discriminer connues par les descripteurs associés. Durant la phase d'apprentissage, une procédure automatique met en relief l'information utile des descripteurs (courbures extrémales) et génère l'ensemble des primitives discriminantes. Une représentation syntaxique en terme de ces primitives pertinentes et de relations entre elles est déterminée pour chaque classe.

Ces modèles ainsi construits sont structurés en un ensemble de règles de décision, lesquelles sont compilées en un arbre de décision binaire. La phase d'identification d'un objet consiste alors à appliquer cet arbre de recherche à chacun des contours extraits et traités. Une primitive de départ est recherchée ; si elle est trouvée, l'arbre propose une autre primitive à rechercher ainsi que sa position éventuelle ou aboutit à une décision finale. Dans les deux phases, la recherche d'une primitive repose sur des techniques de corrélation.

SUMMARY

This paper is concerned with the identification of partially observed or distorted planar objects, using a solid-state video camera.

From outlines of a digitized picture, a preprocessing stage extracts the relevant local information : a curvature function-like description.

During the learning phase, an automatic procedure looks for the pertinent information (singular points of curvature) of the considered object classes descriptors and so generates a set of discriminant primitives. A syntactic representation in terms of these primitives and relations among them is sought.

Multilevels models of object classes are built and structured into a set of decision rules. Such rules determine during the identification phase which primitive to look for and where. Primitives are recognized by correlation techniques.



I. PRESENTATION DU PROBLEME

L'étude présentée ici s'inscrit dans le domaine de la vision par ordinateur destinée à équiper des cellules flexibles d'assemblage de pièces manufacturées. Elle s'intéresse essentiellement aux problèmes d'identification et de localisation d'objets plans pouvant être partiellement visibles, l'organe de perception étant une caméra vidéo. En effet, le problème d'identification est différemment envisagé selon la présence dans le champ de la caméra :

- d'une ou plusieurs pièces complètement perçues et isolées,
- d'une pièce partiellement perçue, de pièces se chevauchant ou présentant des déformations ou des parties additionnelles.

Les systèmes de vision, actuellement disponibles, /1,17/ peuvent reconnaître et localiser des objets isolés, la seule condition étant de disposer d'un fond contrasté. Ils reconnaissent des formes binaires en mesurant leurs caractéristiques globales et en comparant ces valeurs à des modèles déjà établis.

De tels systèmes sont suffisants pour plusieurs applications. Il en existe néanmoins de nombreuses autres pour lesquelles, il serait difficile (utilisation de moyens mécaniques coûteux pour isoler les pièces, ...), voire impossible, de faire en sorte que les pièces soient isolées et complètement visibles.

Plusieurs techniques et approches ont été explorées pour la résolution des problèmes de reconnaissance de formes planes. Elles diffèrent tant par le type de représentation choisi que par les procédures d'analyse mises en oeuvre.

La forme ou silhouette d'un objet, représentée par son contour extérieur et éventuellement ses contours internes, a toujours été considérée comme un de ses facteurs les plus discriminants.

Dans le cas des objets partiellement observés, les propriétés globales des formes ne sont plus valables. La compréhension de telles scènes ne peut alors se réaliser que par une étude des propriétés locales de l'image, notamment de la forme de ses contours.

Cette information locale peut aussi être nécessaire lorsqu'une approche globale n'a pu amener à l'identification d'un objet isolé, ou si cet objet ayant été identifié, il subsistait un problème de symétrie ou de précision dans la localisation.

Plusieurs chercheurs ont opté pour des solutions ad-hoc /7,11,13,16,19/. L'information issue des contours est directement utilisée ; la reconnaissance se résume alors à une simple comparaison de type corrélation du descripteur local utilisé (généralement la fonction de courbure), avec les descripteurs de référence obtenus lors d'une phase d'apprentissage.

Ces méthodes attractives par leur simplicité d'implémentation deviennent vite coûteuses lorsque le nombre d'objets considérés croît (corrélations systématiques) et même impraticables lorsque les détails intérieurs

sont importants ou que des accidents relativement graves surviennent (destruction, occultation ou déformation d'une sous-forme, insertion de formes inutiles) /9,18/.

L'objet est donc considéré comme une organisation structurée "d'éléments" reliés entre eux par des relations /2,4,9,12/. Si un graphe apparaît comme l'outil mathématique le plus simple pour représenter de telles structures relationnelles, son utilisation durant la phase de reconnaissance pose des problèmes de coût et de temps de calcul, même dans les solutions proposées par /2,5/.

Pour parer à ces problèmes de combinatoire, nous avons choisi de représenter les objets par des modèles hiérarchisés et multiniveaux /13/. Chaque niveau n'est utilisé, durant le processus de reconnaissance, que si ses précédents ont échoué ou fourni des résultats insuffisants.

Les primitives de base de ces modèles sont définies comme des sous-formes caractéristiques (discriminantes) des objets considérés. Leur extraction durant la phase d'apprentissage permet une réduction des données et ainsi un allègement de l'analyse. A cela s'ajoute l'utilisation d'une analyse ascendante-descendante, où des primitives sont d'abord recherchées et où l'information géométrique, ainsi disponible, est utilisée à chaque pas de la procédure d'identification pour la guider et donc la simplifier.

II. TRAITEMENT DE L'IMAGE

Nous exposons ici les principales méthodes utilisées pour extraire et traiter l'information nécessaire à la résolution d'un problème d'identification à partir de l'analyse d'une image.

Après pré-traitement (notamment, après filtrage), la segmentation permet d'extraire de l'image, les informations pertinentes pour la compréhension de la scène perçue ; selon la technique utilisée, cette information est sous la forme :

- de contours fermés disjoints, avec un graphe d'inclusion (segmentation utilisant les propriétés topologiques des images binaires) /6/,
- des régions et leurs contours, avec un graphe d'adjacence (même technique, mais sur images multiniveaux),
- des lignes de contraste, non connexes (segmentation utilisant des opérateurs locaux de différenciation spatiale).

Les contraintes de temps imposées par la robotique industrielle ne permettent pas l'emploi de techniques très complexes ; aussi nous avons, dans un premier temps, développé notre méthode à partir d'une image binaire, en ne retenant donc que les contours extérieurs (silhouettes) et intérieurs (en général, trous) des objets qu'elle contient (Figure 1).

Si cette technique convient dans de nombreux cas, elle hérite des graves inconvénients de la binarisation d'une image : grande réduction de l'information disponible (notamment, pour des objets très contrastés) et grande sensibilité aux variations d'éclairage.



Pour ces raisons, si, en phase d'apprentissage, nous considérons que les objets à identifier sont connus par leurs contours fermés, nous admettons qu'en phase d'identification, une quelconque des 3 méthodes de segmentation précitées peut être employée. Après segmentation, la scène est décrite par un ensemble de lignes (contours fermés, frontières des régions ou lignes de contraste) et de relations entre-elles ; ces diverses entités doivent être modélisées selon une description concise afin de rendre leur traitement plus aisé en phase d'apprentissage et d'identification.

Pour cette modélisation, nous avons choisi de travailler dans le plan courbure-abscisse curviligne ; l'extraction des primitives et leurs comparaisons aux entités de la scène perçue sont effectuées sur les fonctions de courbure.

La courbure est directement extraite des contours de la scène (pas d'approximation polygonale) ; elle est donc très bruitée et nécessite un filtrage avant toute analyse ; parmi les techniques existantes pour approcher cette fonction, nous pouvons employer soit le code de Freeman étendu, compensé et lissé /11/ (Figure 2.a.), soit l'intégrale de courbure, obtenue après lissage des contours considérés (Figure 2.b.c.) ; les algorithmes de lissage ou de filtrage curviligne sont dérivés de ceux présentés dans /7/. Ces deux méthodes conduisent à des résultats très proches ; soulignons qu'elles fournissent des descripteurs basés sur la courbure, indépendants de la rotation du modèle. Si la première citée conduit à des temps de calcul très rapides, la seconde peut s'appliquer quelque soit la segmentation initiale.

Pour résoudre notre problème, il faut savoir comparer entre-eux de tels descripteurs, notamment :

- durant l'apprentissage, afin de déterminer un ensemble de sous-formes caractéristiques (ou primitives), $P = \{p_1, \dots, p_N\}$
- durant l'identification, afin de rechercher une primitive de P dans un des descripteurs de la scène perçue.

Deux méthodes sont possibles pour effectuer ces comparaisons suivant que l'on utilise ou pas, un CODAGE des descripteurs en termes d'un alphabet Σ . Tout descripteur peut, en effet, être discrétisé en éléments de base (segment de longueur donnée, arcs, coins ou points convexes ou concaves, ...) auxquels des symboles de sont attachés.

La méthode, utilisant un codage préalable, s'appuie sur des techniques de "string matching" ; elle emploie un algorithme rapide de recherche de facteurs dans des mots /10/ et permet de trouver très rapidement l'ensemble P des primitives ; si l'on veut éviter d'avoir à coder la scène perçue durant la phase d'identification, on peut, par codage inverse, retrouver le sous-descripteur correspondant, que l'on recherchera par corrélation sur les descripteurs.

La difficulté essentielle de cette méthode, réside dans le codage ; l'expérience nous a montré que les représentations des objets étaient instables ; pour en tenir compte, l'algorithme de recherche d'un facteur dans

un mot devient beaucoup plus complexe. Ce problème n'étant pas encore résolu de façon satisfaisante, nous utilisons ci-dessous les techniques de corrélation pour déterminer l'ensemble P directement à partir des descripteurs associés aux objets à apprendre.

III. APPRENTISSAGE

III.1. Définition du modèle

Comme discuté précédemment, le modèle adopté est hiérarchisé et comporte plusieurs niveaux. Nous présentons ici ceux qui concernent directement le cas d'objets partiellement observés :

- Une représentation syntaxique utilisant :
 - .un ensemble P de primitives locales de type 1-D : p_1, p_2, \dots, p_n ;
 - .une relation " α " entre deux primitives consécutives permettant la prise en compte des régions non pertinentes ou n'ayant pu de manière stable être traduites sous forme de primitives (Figure 3). La relation " α " est définie comme une distance sur le descripteur entre 2 primitives. Une représentation structurelle utilisant :
 - .l'ensemble P
 - .l'ensemble $L = \{L_1, \dots, L_Q\}$ des caractéristiques de type 2-D telles que des trous de faible dimension. À ces éléments, sont associées des informations : de position (barycentre), et de type, sous la forme d'une caractéristique globale (rapport des rayons minimum et maximum),
 - une relation vectorielle " β " permettant de traduire l'information géométrique existant entre deux caractéristiques 1-D et/ou 2-D. La relation β est définie par deux angles et une longueur : orientation relative entre 2 vecteurs (Figure 4.a.b.).

III.2. Extraction des primitives

Soit O_j l'objet à apprendre et C_j le contour externe correspondant (ou éventuellement la j ème ligne de niveau interne), on recherche d'abord le descripteur local correspondant. L'acquisition de l'ensemble P des primitives est effectuée en deux étapes :

1. Recherche d'un ensemble de primitives potentielles : PT ;
2. Association d'une fonction de pertinence (évaluée itérativement) à chaque élément de l'ensemble PT . Un seuillage permet alors de faire un choix définitif.

III.2.1. Recherche des primitives potentielles

L'utilisation des facilités offertes par le système de traitement utilisé /8/, (mini-calculateur 32 bits SEL 32/77-80 couplé à un processeur de traitement d'images GRINNELL) permet à l'opérateur de choisir directement les éléments de l'ensemble PT . Un curseur commandé manuellement sert à définir, sur l'objet même (l'ensemble de ses contours) ou sur les descripteurs associés, les points de départ et d'arrivée des parties jugées "intéressantes". Les sous-tableaux ainsi obtenus constituent



l'ensemble des primitives potentielles. Une deuxième manière d'obtenir automatiquement ce "même" ensemble est possible en faisant l'hypothèse que les parties intéressantes d'un objet sont très liées aux variations particulières de sa fonction de courbure. La procédure débute en cherchant les points singuliers de l'objet (points d'inflexion, courbures extrémales, longs segments,...). Une liste d'indicateurs associés aux positions de ces points est établie. La longueur des primitives étant supposée fixée a priori, on positionne par rapport à ces indicateurs les primitives potentielles recherchées. Un test est toutefois nécessaire pour éviter des intersections ou des redondances entre les éléments obtenus. Si ce deuxième mode apparaît comme le plus intéressant, nous avons opté pour un compromis qui fournit à l'opérateur la possibilité d'intervenir pour compléter l'ensemble PT.

III.2.2. Fonction de pertinence

A chaque primitive p_j de P , vont être associés deux ensembles S_j et E_j qui représentent respectivement l'ensemble des classes d'objets qu'elle sélectionne et celui qu'elle exclut. La fonction de pertinence d'une primitive définit son degré de discrimination vis-à-vis de ces ensembles ainsi que les risques d'erreur associés à la possibilité d'observer cette primitive ailleurs que sur les objets de l'ensemble S . Soit w_i une classe d'objets, on note $d(p_j/w_i)$ la distance maximale lorsqu'on compare p_j aux descriptions des objets de la classe w_i :

avec : $d(p_j/w_i) = \text{Max} (d(p_j/O_i), O_i \in w_i)$
Si $R = \{r_1, \dots, r_n\}$ est le descripteur associé à p_j
et $S = \{s_1, \dots, s_m\}$ est le descripteur associé à O_i
alors : $d(p_j/O_i) = \text{Min} (\sum_{i=1}^n |r_i - s_{i+\kappa}|)$, $\kappa = 0, m-1$

Soient S_j et E_j les ensembles des classes sélectionnées et exclues par la primitive p_j . On définit alors la fonction de pertinence par :

$$\rho(p_j) = \frac{\text{Max}(d(p_j/w_i), w_i \in S_j)}{\text{Min}(d(p_j/w_k), w_k \in E_j)}$$

III.2.3. Recherche des primitives

L'ensemble P des primitives définitivement choisies est un sous-ensemble de l'ensemble PT dont les éléments ont été validés par un programme basé sur l'évaluation de leurs fonctions de pertinence /15/.

La validation d'une primitive potentielle p_j signifie qu'elle assure une forte transition entre les ensembles associés S_j et E_j ($\rho(p_j)$ proche de zéro) et que le nombre de classes de S_j ne dépasse pas un seuil fixé. Deux cas se présentent pour les primitives non validées :

Si $\rho(p_j)$ voisin de 1 alors

$$\exists (w_i, w_k) \in S_j \times E_j \quad \text{tq} \quad d(p_j/w_i) \approx d(p_j/w_k)$$

On met w_k dans S_j et on recommence le processus.

Si la transition entre S_j et E_j existe mais n'est pas assez importante, on abandonne alors p_j .

III.3. Génération des règles de décision

Cette dernière tâche de l'apprentissage consiste à générer un ensemble de règles permettant d'organiser la recherche des caractéristiques 1D ou 2D sur les descripteurs de la scène. Deux types de règles sont considérés : celles utilisant uniquement la représentation syntaxique et celles utilisant en plus l'information structurelle. D'une manière générale, ces règles se présentent de la manière suivante: "Si e , partie d'un contour, comporte la séquence discriminante :

$$"p_{i1}[(\beta_1 L_1), (\beta_2 L_2), \dots] \alpha_{i1} p_{i2} [\dots] \dots",$$

alors e appartient à un objet de la classe w_i ".

Le premier type de règles est obtenu en se restreignant d'abord à la recherche de séquences discriminantes sans relations bidimensionnelles.

Partant des représentations " $p_{i1}, \alpha_{i1}, \dots, p_{ij}, \alpha_{ij}, \dots$ ", pour $1 \leq j \leq m$ (m = nombre de classes apprises) et connaissant l'ensemble des classes sélectionnées par chaque primitive p_j , les séquences discriminantes sont recherchées d'une façon similaire à celle développée dans /10/ pour la recherche de facteurs discriminants. Les relations bidimensionnelles ne sont alors ajoutées qu'à titre de vérification sur les séquences obtenues. Ces séquences peuvent être utilisées directement ou alors compilées en un arbre de décision, ce qui se traduit par une plus grande rapidité au niveau de la phase d'identification. A chaque noeud de cet arbre, on a les informations suivantes :

- la primitive à rechercher en ce noeud : $\alpha_j p_j$ ($\alpha_j < 0$ ou $\alpha_j > 0$ selon le sens de parcours),
- l'ensemble des classes sélectionnées associé à la séquence de primitives trouvée en ce noeud.

Un exemple d'arbre de ce type est présenté dans la Figure 5.

IV. IDENTIFICATION

Lors de la phase de reconnaissance, un ensemble d'objets étant présent dans le champ de la caméra, il s'agit d'assigner à chacun le nom de la classe à laquelle il appartient. Un prétraitement de la scène conduit à un ensemble de "contours" qui sont traités et décrits par des descripteurs. Identifier la scène revient alors à appliquer l'arbre de recherche généré lors de la phase d'apprentissage sur l'ensemble de ces descripteurs. Il est à noter, cependant, que suivant le prétraitement appliqué, plusieurs cas se présentent :

- Recherche de lignes de contraste

La notion de connexité des lignes n'est plus valable. Chaque ligne est traitée



PROCEDURE D'APPRENTISSAGE POUR L'IDENTIFICATION D'OBJETS PLANS PARTIELLEMENT OBSERVES
LEARNING PROCEDURE FOR THE IDENTIFICATION OF PARTIALLY OBSERVED PLANAR OBJECTS

A. RIAD - M. DEVY

séparément si sa longueur permet la recherche des primitives 1-D. L'utilisation des relations " α " est très réduite. A l'inverse, l'information 2-D est plus sollicitée, sous la forme des relations " β ".

- Extracteur de contours multiniveaux

La connexité de la segmentation est assurée par ce prétraitement. Dans le cas de deux objets qui se chevauchent, on a, par exemple, le type de Figure 6.a. (cas idéal).

On peut effectuer une recherche de l'objet "complet" puis traiter ensuite la ligne restante, ou encore traiter chaque ligne séparément. Une hypothèse étant faite sur la nature d'un objet (pour (1) par exemple), on fait une vérification sur l'ensemble des lignes (1) et (2) (ou (1) et (3)).

- Extracteur bi-niveaux

Notons d'abord que c'est le prétraitement utilisé dans notre étude. Si la connexité de la segmentation est encore assurée, les lignes de transitions entre deux objets qui se chevauchent ne sont plus visibles (pour l'exemple précédent, Figure 6.b.).

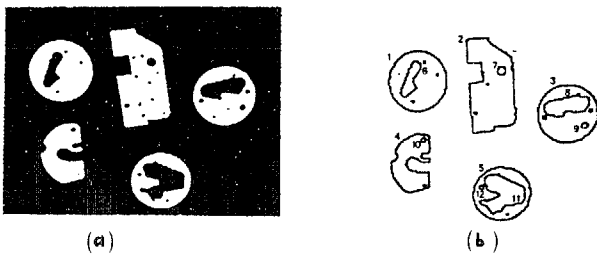
L'identification se fait alors de la manière suivante : on traite tout le contour jusqu'à avoir une hypothèse sur un objet "possible". Le descripteur modèle de cet objet est positionné par rapport au descripteur de la scène en utilisant l'information géométrique associée à l'hypothèse. La comparaison de ces deux descripteurs permet de dégager les zones similaires entre l'objet et la scène. Celles-ci sont retranchées de la scène et les lignes restantes sont alors retraitées.

BIBLIOGRAPHIE

- 1 G.J. AGIN, "An experimental vision system for industrial application", 5th Int. Symp. on Ind. Robots, Chicago, USA, Sept. 1975.
- 2 A.P. Ambler, H.G. Barrow and al, "A versatile computer-controlled assembly system", 3th IJCAI, Stanford, CA, USA, Aug. 1973.
- 3 N. Ayache, "Un système de vision bidimensionnelle en robotique industrielle", Thèse de Docteur-Ingénieur, Université Paris-Sud, Juin 1983.
- 4 H.G. Barrow, A.P. Ambler, R.M. Burstall, "Some techniques for recognizing structures in pictures", Frontiers of Pattern Recognition, S. Watanabe, Ed. 1972, pp.1-29.
- 5 R.C. Bolles, "Locating partially visible objects : the local-feature-focus method", Proc. 1st Annual National Conf. on Artificial Intelligence, pp.41-43, Stanford, Aug. 1980.
- 6 C. Bourdeau, M. Briot, J.M. Pons, J.C. Talou, "Etude et réalisation d'un extracteur rapide de contours fermés d'une image binarisée", Congrès AFCE-
- 7 J.D. Dessimoz, "Traitement des contours en reconnaissance de formes visuelles : applications à la robotique", Thèse Ph.D., n°387, E.P.F., Lausanne, 1980.
- 8 M. Devy, "Manuel d'utilisation de la mémoire d'images Grinnell", Note Technique LAAS-MITRI, n°82.043, Aug. 1982.
- 9 K.S. Fu, "Syntactic methods in pattern recognition", New-York Academic, 1974.
- 10 M. Ghallab, "Optimisation de processus décisionnels pour la robotique", Thèse d'Etat, U.P.S., Toulouse, 1982.
- 11 J.W. McKee, J.K. Aggarwal, "Computer recognition of partial views of curved objects", IEEE Trans. Comp., vol.C-22, pp.790-800, Sept. 1977.
- 12 T. Pavlidis, "Structural pattern recognition", Springer-Verlag, 1977.
- 13 W.A. Perkins, "A model-based vision system for industrial parts", IEEE Trans. Compt., vol.C-27, pp.126-143, Feb. 1978.
- 14 A. Riad, M. Briot, "A hierarchical vision system for object identification and localization", 13th Int. Symp. on Ind. Robots, Chicago, USA, April 1983.
- 15 A. Riad, "Identification d'objets partiellement observés en vue de leur manipulation en robotique", Thèse de 3ème Cycle, n°2890, U.P.S., Toulouse, Oct. 1983.
- 16 Y. Shirai, "Edge finding, segmentation of edges and recognition of complex objects", in Proc. 4th Int. Conf. on Artificial Intelligence, pp.674-681, Tbilissi, Georgia, URSS, 1975.
- 17 F. Stuck, "Réalisation d'un système autoadaptatif de traitement d'images pour l'identification et la localisation de pièces en robotique", Thèse de 3ème Cycle, U.P.S., Toulouse, 1980.
- 18 W.H. Tsai, K.S. Fu, "A syntactic-statistical approach to recognition of industrial objects", Proc. 5th Int. Conf. on Pattern Recognition, pp.251-259, Miami Beach, Florida, Dec. 1980.
- 19 M. Yashida, S. Tsuji, "A versatile machine vision system for complex industrial parts", IEEE Trans. Compt., vol. C-26, pp.882,894, Sept. 1977.

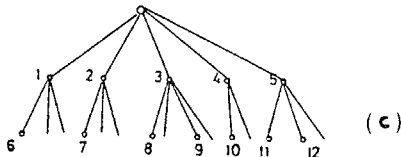


PROCÉDURE D'APPRENTISSAGE POUR L'IDENTIFICATION D'OBJETS PLANS PARTIELLEMENT OBSERVÉS
 LEARNING PROCEDURE FOR THE IDENTIFICATION OF PARTIALLY OBSERVED PLANAR OBJECTS
 A. RIAD - M. DEVY



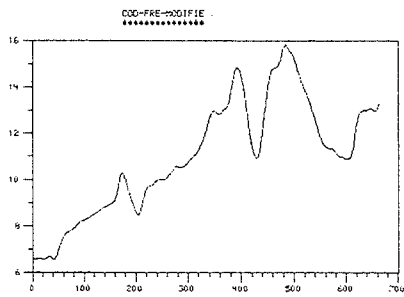
(a)

(b)

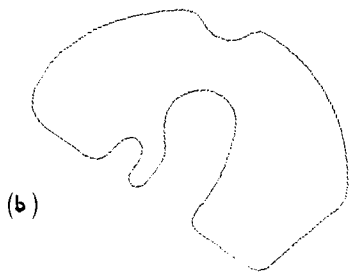


(c)

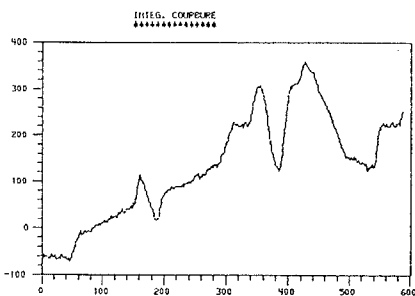
FIGURE 1. a) Image binaire
 b) Contours correspondants
 c) Arbre d'inclusion



(a)



(b)



(c)

FIGURE 2. a) SCEFC du contour 4 (Figure 1)
 b) Contour 4 lissé
 c) Intégrale de courbure du contour 4 lissé

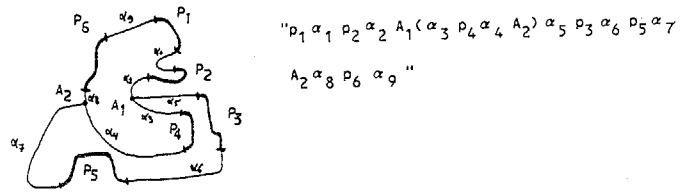
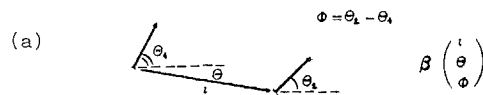
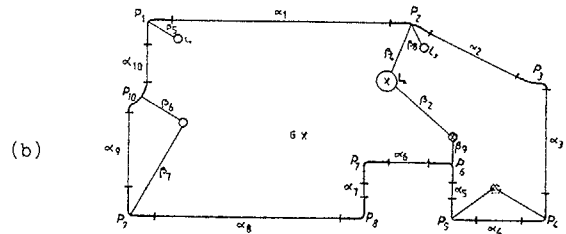


FIGURE 3. Exemple de représentation syntaxique



(a)



(b)

FIGURE 4. a) Définition de β
 b) Exemple d'une représentation complète

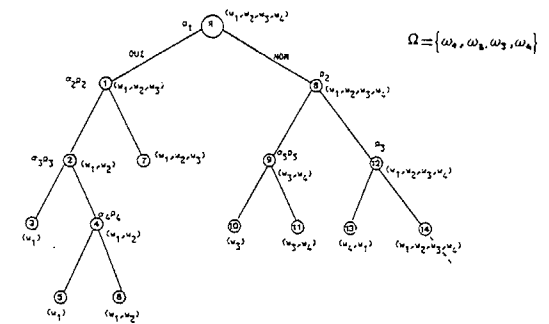
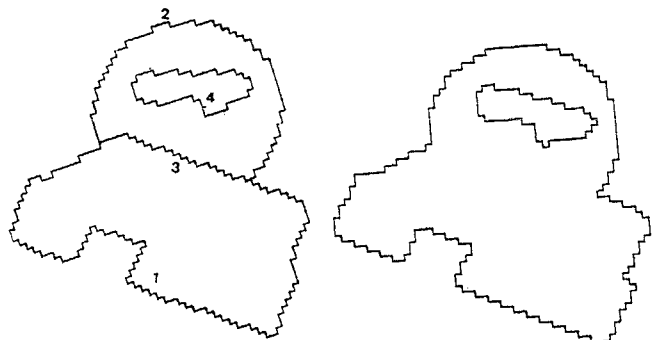


FIGURE 5. Exemple d'arbre de décision



(a)

(b)

FIGURE 6. Objets en recouvrement
 a) Cas idéal
 b) Extracteur biniveaux