

Deux méthodes de comparaison d'images pour l'identification d'objets à partir de données prospectives

S. COSTA *, P.Y. ARQUES*#, H. SEDANO *, A. HILLION&

* ECN TOULON - B.P. 28 - CTSN/LSA/RAD/TTS - 83800 TOULON NAVAL - FRANCE

Tél. : 04 94 16 21 14 - Fax : 04 94 16 22 81 - E-mail : costa@isitv.univ-tln.fr

I.S.I.T.V. - Université de Toulon et du Var - B.P. 32 - 83957 LA GARDE CEDEX - FRANCE

& TELECOM BRETAGNE - B.P. 832 - 29285 BREST CEDEX - FRANCE

RESUME

Cette étude aborde le problème de l'identification d'objets mobiles à partir de données délivrées par un senseur prospectif dont la conception est actuellement en cours. Le but est d'estimer la faisabilité d'une telle identification à l'aide d'outils disponibles à ce jour en reconnaissance des formes. On présente dans ce papier la réalisation complète d'une chaîne de simulation, comprenant à la fois la génération des données (non disponibles) et la mise en place de processus capables de les exploiter dans un but d'identification. Des paramètres variables contrôlent la nature des images (richesse, niveau de bruit) tout au long de la simulation, ceci afin de pouvoir prendre en compte des données de qualité variable.

ABSTRACT

This survey is about the problem of mobile objects identification using data given by a prospective sensor. Its aim is to evaluate the feasibility of such an identification using nowadays pattern recognition tools. This paper shows the global achievement of an identification device, from the radar data simulation (non available yet) to the implementation of processes able to use these data in order to identify objects. Some simulation parameters controlling images characteristics can vary so that we are able to take into account data with several quality levels.

1 Introduction

Cet article traite du problème de la classification d'objets mobiles à l'aide de données provenant d'un senseur, par exemple un radar. Le senseur mis en jeu étant prospectif, nous ne disposons pas des données réelles correspondantes. On décrit donc ici un processus complet de simulation, allant de la création des données à la mise en place d'un dispositif capable de les exploiter.

Aujourd'hui, la plupart des radars bidimensionnels produisent des images dans le plan "distance \times azimut". Ce type de construction d'image utilise par exemple les techniques S.A.R. (ouverture synthétique) ou I.S.A.R. (ouverture synthétique inverse). Cependant, dans le cas d'une cible en mouvement, en particulier non coopérative et fortement manoeuvrante, ces techniques présentent des limites provenant du découplage arbitraire des problèmes de détection et d'estimation: en pratique, elles se révèlent incapables d'estimer le mouvement de la cible, qui doit pourtant être compensé avant la formation de l'image.

Pour pallier ces imperfections, le Laboratoire d'Analyse et d'Architecture des Systèmes (L.A.A.S.) de Toulouse propose une modélisation complète du problème de détection-estimation appliquée au traitement radar, utilisant une technique basée sur le filtrage particulière [11]. Celle-ci confère au senseur une capacité intrinsèque d'imagerie [2] qui bénéficie d'une intégration plus longue du signal, en accord avec la trajectoire de la cible. En conséquence, les

performances actuelles des radars en traitement et en poursuite devraient s'en trouver améliorées. En prenant en compte ces nouvelles caractéristiques, des hypothèses ont été émises quant à la nature des données délivrées par des senseurs prospectifs susceptibles d'utiliser de telles techniques.

Ainsi, notre processus d'identification est basé sur l'exploitation de deux types de données [3]:

- des informations cinématiques concernant la cible, contenues dans son vecteur vitesse,
- une image bidimensionnelle de la cible.

La simulation de ces données est schématisée par la figure 1. En ce qui concerne les données cinématiques, elle consiste en un tirage aléatoire des valeurs dans un ensemble réaliste, et à l'ajout d'un bruit gaussien représentant l'erreur de mesure du senseur.

La simulation des données de type image a été réalisée par l'Equipe de Synthèse d'Image de l'Université de Marne-La-Vallée. Elle est basée sur une technique de lancer de rayons adaptée au domaine radar. En outre, cette simulation permet de "régler" la richesse en informations des images [2]. Divers niveaux de "qualité" d'image ont donc pu être pris en compte lors de l'identification [13].

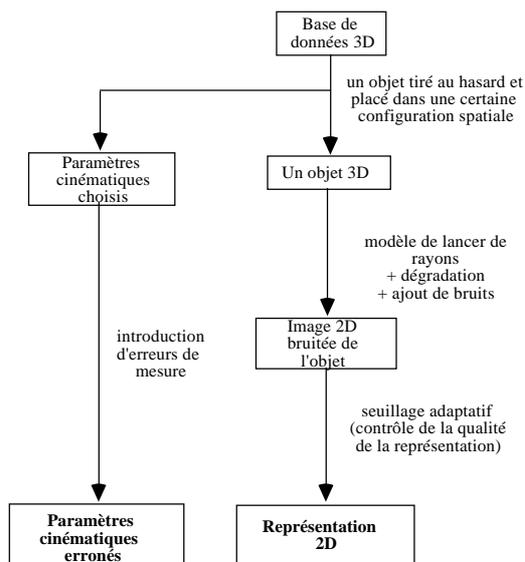


Fig. 1: Simulation des données synthétiques

2 Processus d'identification

Le principe de l'identification est d'effectuer des comparaisons entre l'image issue du senseur et les éléments tridimensionnels contenus dans une base de données regroupant l'ensemble des objets connus susceptibles d'être reconnus.

On aborde pour cela deux démarches différentes correspondant à deux types différents d'images:

- la première exploite des images pauvres en informations, obtenues en ajustant le paramètre qualitatif cité ci-dessus à un niveau bas. Elle effectue une comparaison d'images 2D à images 2D issues des modèles 3D, et est basée sur l'extraction d'attributs de type squelette de l'image et sur une prise de décision organisée de manière syntaxique,

- la seconde exploite des images riches en informations, qu'elle compare aux modèles 3D de cibles en utilisant la notion d'invariance par projection.

Ces deux approches sont décrites dans les paragraphes suivants.

3 Méthode syntaxique

Le premier dispositif mis en place utilise des images pauvres en informations, obtenues par un seuillage grossier des images en niveaux de gris issues du lancer de rayons. On obtient ainsi des images en noir ou blanc représentant les parties de l'objet les plus réfléchissantes pour la longueur d'onde utilisée. Un exemple de telles images est donné par la figure 2.

Le processus d'identification se déroule en deux étapes [4, 7]:

- une extraction de primitives de type squelette,
- une prise de décision syntaxique.

Squelettiser un objet revient à le représenter par son axe de symétrie principal [1]. Cette notion permet de résumer la structure générale d'un objet sans tenir compte de son épaisseur. La squelettisation de nos images s'effectue grâce à une méthode discrète à partir de cartes de distances. Construire la carte de distances d'une image binaire revient à

étiqueter chaque point par sa distance minimale (euclidienne) au fond. Les points de l'axe médian de l'objet sont alors les maxima locaux de cette carte des distances. Le squelette s'obtient ensuite après une phase dite de reconnexion de l'axe médian, qui consiste à "propager" des chemins le long des lignes de crête de la carte de distances. Le résultat est un squelette pondéré, dont chaque point est étiqueté par sa plus proche distance au bord de l'objet.



Fig.2: Donnée synthétique 2D en l'absence de bruit

La comparaison dans l'espace des squelettes fait appel à une méthode syntaxique utilisant la notion de grammaire [5, 10, 12]. Ce choix a été motivé par le fait que les images à partir desquelles se fait l'identification sont constituées principalement d'éléments géométriques simples (points, segments, arcs de cercle, ...) pouvant évoquer des mots à partir desquels se construisent des phrases, en l'occurrence les images.

Une méthode syntaxique est généralement structurée en trois étapes:

- la mise en place d'un vocabulaire dit "terminal": l'ensemble des formes élémentaires non décomposables,
- la recherche de règles d'assemblage,
- la définition d'un vocabulaire dit "auxiliaire".

Une grammaire permet de représenter un objet ou une classe d'objets (selon la finesse de la classification), et est définie par le quadruplet:

$$G = (V_T, V_N, S, P)$$

avec:

V_T : vocabulaire terminal,

V_N : vocabulaire auxiliaire (non terminal),

S : axiome de la grammaire, $S \in V_N$,

P : ensemble des règles de production.

Une règle de production grammaticale s'écrit:

$$\alpha \rightarrow \beta \text{ ("}\alpha \text{ se réécrit en } \beta \text{")}$$

avec α et $\beta \in (V_T \cup V_N)^*$, où $(V_T \cup V_N)^*$ représente l'ensemble de toutes les chaînes générées en concaténant un nombre quelconque d'éléments de V_N ou de V_T .

Si une chaîne d'éléments de V_T peut être générée à partir de l'axiome S et en utilisant une série de réécritures respectant les règles de la grammaire, on dit que cette chaîne appartient au langage généré par cette grammaire.

Les grammaires sont affectées à différentes catégories suivant la forme de leurs règles de production. On a choisi d'utiliser une grammaire dite de "type 2" (ou "context-free"), dont les règles de production sont de la forme:

$$N \rightarrow \beta \text{ avec } N \in V_N \text{ et } \beta \in (V_T \cup V_N)^*$$

L'axiome de base est commun à l'ensemble des grammaires qui décrivent nos classes d'objets. Il s'agit d'une image représentant trois "éléments de base": un point, un segment et un arc de cercle.

L'ensemble de nos règles de production est le suivant:

- déplacement spatial d'un ou plusieurs éléments de base,
- modification structurelle (allongement, raccourcissement, ...) d'un ou plusieurs éléments de base,
- duplication d'un ou plusieurs éléments de base.

Ces règles permettent, à partir de l'image-axiome, de représenter l'ensemble des images de référence correspondant aux classes d'objets présentes dans la base de données 3D. Reste à déterminer ensuite si l'image issue du radar appartient ou non aux langages grammaticaux ainsi construits.

Les tests de simulation effectués sur cette méthode ont donné les résultats présentés en figure 3. A titre de comparaison, ceux fournis par une identification par corrélation sont également indiqués.

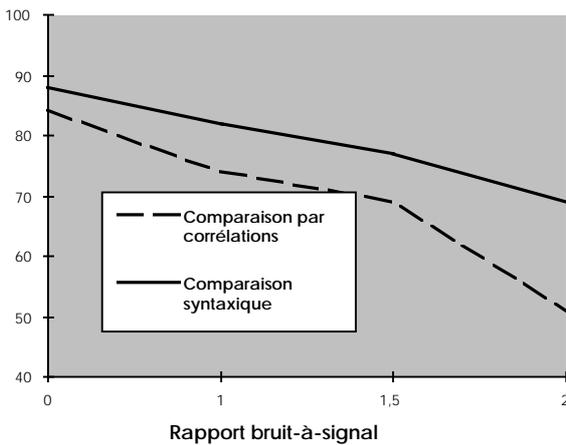


Fig. 3: Résultats de classification obtenus par reconnaissance syntaxique ou par corrélation

Les différentes valeurs de performances ont été obtenues en faisant varier le niveau de bruit introduit dans la simulation des images radar.

4 Méthode par invariance

Les images sont maintenant plus "riches" en informations que précédemment, semblables à celle de la figure 4.

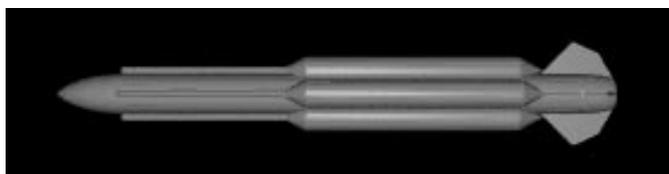


Fig. 4: Donnée synthétique 2D en l'absence de bruit.

Les paramètres invariants [6, 15] sont les propriétés des configurations géométriques qui restent inchangées par une certaine classe de transformations. Ici, on est plus particulièrement intéressé par la recherche d'invariants par une projection en perspective sur une image: étant donné le contexte de simulation des images, la projection à considérer comprend une transformation euclidienne tridimensionnelle de l'objet, suivie par une projection en perspective sur une image.

On cherche donc les éléments 3D restant invariants par des transformations projectives de la forme:

$$k \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ t_{41} & t_{42} & t_{43} & t_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

où k est un scalaire de valeur telle que la quatrième coordonnée soit égale à 1.

Pour que ces invariants existent, on doit faire l'hypothèse que les objets 3D sont rigides [13, 14], ce qui entre parfaitement dans le cadre de notre application.

Le paramètre invariant le plus utilisé peut être considéré comme le "cross-ratio" de volumes tétraédriques calculé en prenant les déterminants des coordonnées des points quatre par quatre [8, 9].

Un invariant pour quatre points 3D est donné par:

$$C_R(M_1, M_2, M_3, M_4) = \frac{|\overline{M}_1 \overline{M}_2 \overline{M}_3 \overline{M}_4| |\overline{M}_2 \overline{M}_1 \overline{M}_4 \overline{M}_3|}{|\overline{M}_3 \overline{M}_4 \overline{M}_1 \overline{M}_2| |\overline{M}_4 \overline{M}_3 \overline{M}_1 \overline{M}_2|}$$

où $\overline{M}_i = (X_i, Y_i, Z_i, 1)^T$

On note que la valeur du cross-ratio dépend de l'ordre dans lequel les M_i sont pris en compte.

On a choisi d'utiliser, toujours dans le même ordre, quatre points caractéristiques situés sur quatre éléments toujours présents sur les objets (voir figure 5). Le cross-ratio des quatre points est préalablement calculé pour chaque objet 3D de la base de données, puis pour l'objet 2D présent sur l'image issue du senseur.

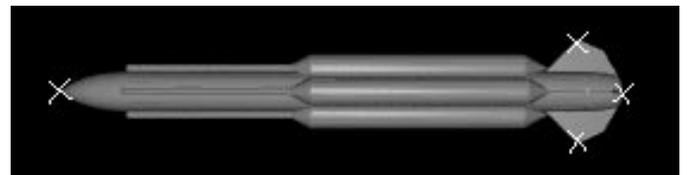


Fig. 5: Emplacement des points choisis pour le calcul du cross-ratio.

Le système de reconnaissance basé sur le cross-ratio est décrit par la figure 6. Il utilise le cross-ratio comme un index pointant sur la base de modèles.

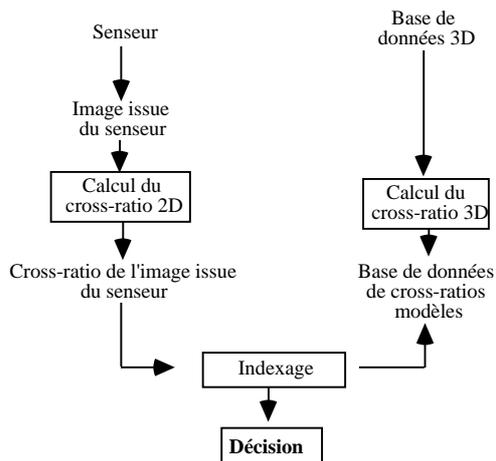


Fig. 6: Processus d'identification basé sur le cross-ratio.

Les résultats fournis par sa mise en pratique sont présentés en figure 7. Ils sont, là encore, confrontés aux résultats obtenus par des comparaisons d'images par corrélation.

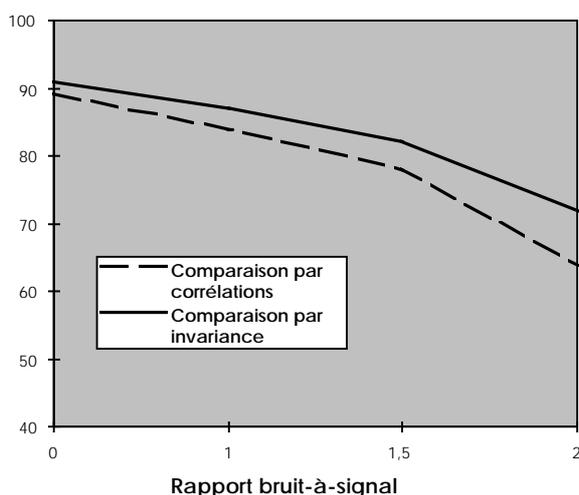


Fig. 7: Résultats obtenus par notion d'invariance ou par corrélation.

5 Conclusion

On remarque, à la vue des résultats des figures 5 et 7, que les deux approches (adaptées à deux types différents d'images) se comportent nettement mieux qu'une comparaison par corrélation. En particulier, l'approche syntaxique présente une robustesse très intéressante au bruit introduit dans l'image issue du senseur. L'atout principal de l'approche par invariance est sa rapidité, en effet, l'utilisation d'une fonction d'indexage permet un accès direct aux modèles de la base de données, sans utiliser d'information spécifique sur le modèle lui-même ou sur sa position dans l'espace.

En conclusion, chacune des deux méthodes permet d'une manière efficace de traiter un certain type de données. Les tests menés actuellement montrent en outre que chacune d'entre elles convient particulièrement à certains angles d'approche de l'objet à identifier. Il serait donc intéressant, dans le même contexte d'identification, d'étudier une fusion

des informations provenant d'un système de capteurs fournissant divers types d'images, et permettant, selon la position relative de l'objet mobile à identifier, de prendre une décision grâce à l'une ou à l'autre des deux méthodes décrites ici.

6 Références

- [1] ATTALI D. "Squelettes et graphes de Voronoï 2D et 3D", Thèse de l'Université Joseph Fourier-Grenoble I, 1995.
- [2] CHAMON M.A. "Filtrage particulière et ouverture synthétique inverse sur cibles RADAR non-coopératives", Thèse du LAAS-CNRS Toulouse, Université Paul SABATIER, 1996.
- [3] COCQUEREZ J.P., PHILIPP S. "Analyse d'images: filtrage et segmentation", MASSON Paris, 1995.
- [4] COSTA S., ARQUES P.Y., SEDANO H., HILLION A. "Two-dimensional energetic representations simulation and object identification by extracting edge and skeleton", IEEE CESA'96 IMACS Multiconference, Lille, FRANCE, Symposium on Robotics and Cybernetics, pp. 433-436, 9-12 juillet 1996.
- [5] COSTA S., ARQUES P.Y., SEDANO H., HILLION A. "Syntactic identification using skeleton extraction from energetic images", ICSPAT' 96. Boston, USA, pp. 1142-1145, 7-10 octobre 1996.
- [6] FORSYTH D., MUNDY J. L., ZISSERMAN A., COELHO C., HELLER A., ROTHWELL C. "Invariant descriptors for 3-D object recognition and pose" IEEE Transactions on Pattern Analysis and Machine Intelligence (USA), Vol. 13, n°10, pp.971-991, octobre 1991.
- [7] GAILLAT G. "Méthodes statistiques de reconnaissance des formes", Ecole Nationale Supérieure de Techniques Avancées, 1983.
- [8] GROS P. "How to use the cross-ratio to compute projective invariants from two images", Second Joint European - US Workshop, Ponta Delgada, Azores, Portugal, 9-14 octobre 1993.
- [9] MAYBANK S. J. "Classification based on the cross ratio".
- [10] MICLET L. "Méthodes structurelles pour la reconnaissance des formes", EYROLLES, Collection Technique et Scientifique des Télécommunications, 1984.
- [11] NOYER J.C. "Traitement non-linéaire du signal RADAR par foltrage particulière", Thèse du LAAS-CNRS Toulouse, Université Paul SABATIER, 1996.
- [12] PAVLIDIS T. "Structural pattern recognition", Springer-Verlag, Springer Series in Electrophysics, 1977.
- [13] POSTAIRE J. G. "De l'image à la décision. Analyse des images numériques et théorie de la décision", DUNOD Paris, 1987.
- [14] SAMU T., SADJADI F., HALL E. "Three dimensional moment invariants for automated target recognition", Automatic Object Recognition VI, Proceedings of the SPIE, Vol. 2756, pp. 230, Orlando, FL, USA, 9-10 avril 1996.
- [15] ZISSERMAN A., FORSYTH D., MUNDY J. L., ROTHWELL C., LIU J., PILLOW N. "3D object recognition using invariance", Journal Artificial Intelligence (Netherlands), ELSEVIER, Vol. 78, pp. 239-88, octobre 1995.