

NEUVIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

TRAITEMENT DE SEQUENCE D'IMAGES
ORIENTE VERS L'ANALYSE DE SCENE

Y. LE GUILLOUX

ETAB. TECH. CENT. ARME - CTME/OP/TI - 94114 ARCUEIL CEDEX (France)

RESUME

Après avoir situé la traitement de séquences d'images et l'analyse de scène dynamique, leurs buts et leurs problèmes, on décrit une méthode de correspondance entre deux images consécutives d'une séquence. Cette technique, basée sur la programmation dynamique, n'utilise aucun modèle abstrait ; elle réalise simultanément la séparation des objets en mouvement et l'estimation de leurs vitesses. Le principe est de regrouper les structures invariantes entre deux images. On présente des résultats obtenus à partir d'une séquence d'images infra-rouge 64 x 64.

SUMMARY

After situating image sequence analysis and dynamic scene analysis, their goals and problems, we describe a method for matching two successive images from a sequence. This technique, based on dynamic programming, does not use any abstract model, it realizes simultaneously the segmentation of moving objects and the estimation of their velocities. The principle is to gather structures that stay invariable through both images. We present the results obtained from a 64 x 64 infrared image sequence.

INTRODUCTION

Les progrès techniques d'intégration rendent possibles des traitements opérant sur plusieurs images. Ceux-ci sont effectués avec des moyens informatiques courants en des temps raisonnables.

On a donc cherché à employer le complément de données fourni pour rendre plus performantes les techniques de détection, d'amélioration et de codage d'images prises dans une séquence.

Depuis 1969, on a ainsi assisté à la naissance d'un domaine de recherches spécifique. Signalons notamment les noms de J.K. AGGARWAL (1), T.S HUANG(1), H.H. NAGEL (1) et l'ouvrage récapitulatif "IMAGE SEQUENCE ANALYSIS" (1981) (1) qui fournit une bibliographie impressionnante.

Comme le traitement d'image a engendré l'analyse de scène, le traitement de séquence comprend l'analyse de scène dynamique. On vise alors une description des objets de la scène et des évolutions en cours.



INTERET DU TRAITEMENT DE SEQUENCES D'IMAGES

Les apports sont variés et intéressent de nombreux domaines.

La redondance entre deux images consécutives d'une séquence permet d'envisager une réduction du débit nécessaire pour le stockage et la transmission. La téléconférence et le visiophone sont notamment concernés.

L'accumulation de clichés d'une même scène, indépendamment de l'aspect temporel, permet d'obtenir une vue synthétique meilleure que tous les clichés analysés. On obtient ainsi par exemple des clichés satellites sans nuages (2).

La détection et la compréhension d'une évolution temporelle forment le champ privilégié du traitement de séquences d'images. Les domaines d'application les plus motivés sont :

- la vision du robot qui doit modéliser l'environnement et les actions en cours,
- l'observation de la terre en vue d'analyser automatiquement les modifications (couverture, neigeuse, érosion des côtes, évolution des forêts ...),
- la surveillance de la circulation routière,
- les applications militaires que sont la poursuite robuste de plusieurs cibles et la veille automatique,
- les applications médicales opérant sur les images de cellules ou de traces de radiopharmaceutiques dans le corps par exemple,
- la météorologie, pionnier dans ce type de traitement, qui s'intéresse aux déplacements des nuages et à la formation de phénomènes atmosphériques locaux.

Un autre apport de l'utilisation de plusieurs vues d'une même scène est la possibilité de construire un modèle tridimensionnel des objets présentés. L'observation de la terre et la vision du robot encore une fois, l'architecture et bien d'autres secteurs d'activités sans doute s'intéressent ainsi à l'automatisation de tels procédés.

Les séquences spatiales, suites de coupes, permettent de reconstruire une distribution tridimensionnelle de densité par tomographie. L'intérêt d'un tel outil pour la médecine est bien connu.

ANALYSE DE SCENE - ANALYSE DE SCENE DYNAMIQUE

L'analyse de scène donne une description d'image. Elle regroupe des parties de l'image en objets référencés comme un tout parfois morcelable (description par un arbre). Il s'agit donc d'une réduction d'information redondante. L'objet situé dans l'image étant d'autre part connu sous forme de modèle, on ne garde que l'information nouvelle : la présence de l'objet dans la scène, sa position et quelques autres caractéristiques macroscopiques.

Remarquons que si le modèle est fourni par une autre image, il s'agit de réduire la redondance d'information entre deux images, ce qui relève déjà de l'analyse de séquence d'images.

L'analyse de scène dynamique doit assurer une tâche similaire, le problème étant plus complexe. Les redondances spatiales et temporelles devront être réduites entre les images de la séquence et éventuellement entre la séquence et une base de modèles. La description qui en résulte ne sera donc pas une suite de descriptions statiques. Elle aura typiquement la forme d'un scénario. Si ce scénario ne signale rien, mais que la séquence n'évolue pas. C'est l'équivalent du codage différentiel d'un signal suréchantillonné. L'hypothèse fondamentale est la lenteur de l'évolution de la scène relativement à la cadence de prise de vue, condition vérifiée dans les applications.

Toutefois, contrairement au cas du signal, ce n'est pas une valeur numérique qui varie faiblement, mais la description de chaque image. Si une caméra effectue un panoramique, la différence point à point de deux images consécutives n'a pas de raisons d'être faible.

Il s'agit donc a priori d'interpréter chaque image de la séquence, puis de confronter les descriptions pour en déduire une évolution. Deux remarques s'imposent :

- La comparaison de deux descriptions devra de façon certaine identifier un même objet dans deux images malgré son déplacement, son masquage (ou apparition) partiel, le bruit d'observation ... On aura donc besoin d'une interprétation robuste et précise de chaque image, ce qui renvoie au problème du traitement d'image idéal.
- Le traitement envisagé ci-dessus suppose une quantité de calcul considérable pour chaque image. On peut noter que l'interprétation effectuera souvent les mêmes traitements dans des images successives. C'est cette répétition que l'on pourrait peut-être éviter.

L'analyse de séquences d'images a créé ses outils propres, et mis en lumière un problème fondamental : celui de la correspondance. En effet, dans une séquence typique, la séparation en informations redondante et nouvelle suppose l'identification des composantes de la scène, vues éventuellement sous des angles différents. Si l'on ne signale que les changements survenus, c'est après vérifié toutes les identifications. C'est sans doute là le travail le plus long.

On a vu que la correspondance pouvait être envisagée entre des descriptions abstraites des images. L'analyse de scène dynamique propose aussi des stratégies différentes. Certaines techniques travaillent directement sur les images, d'autres à des niveaux variés intermédiaires entre l'image et le modèle abstrait. Les problèmes essentiels du traitement de séquences d'images semblent être l'établissement de correspondances et le choix du niveau auquel les réaliser.

En pratique, l'application à des images réelles reste souvent liée à des techniques de faible niveau d'abstraction. Seule l'utilisation de scènes très spécifiques permet le recours à des modèles "intelligents". En effet, le passage à un niveau abstrait est généralement sensible aux imperfections de l'image.

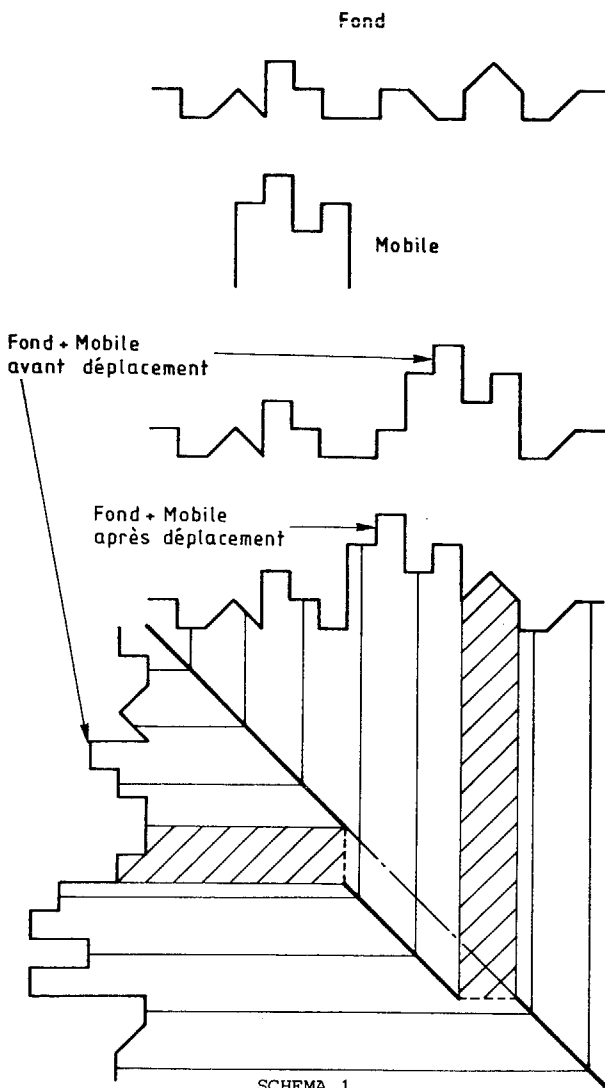
UNE TECHNIQUE DE CORRESPONDANCE AU NIVEAU DES IMAGES

Comme on l'a vu, on sera amené à réduire l'information par identification des parties de la scène entre plusieurs images, les évolutions étant nécessairement limitées d'une image à l'autre. Ce travail est le rôle spécifique du traitement de séquence d'images.

Si l'on veut insister sur l'apport des techniques de séquences, il est intéressant de faire abstraction des traitements d'images préliminaires. Ainsi on met en correspondance directe deux images numériques. La faible dimension des données utilisées (64 x 64) rendait très envisageable une telle stratégie.

Expérimentalement, on a utilisé une hypothèse simplificatrice importante qui ramène le problème à plusieurs problèmes monodimensionnels. En effet, supposons que les mouvements observés, y-compris celui de la caméra, sont horizontaux. Alors à chaque ligne d'image correspond exactement la ligne de même rang de l'image suivante.

Visualisons la correspondance entre deux lignes d'images dans le cas du déplacement d'un mobile devant un fond (sch.1).



SCHEMA 1

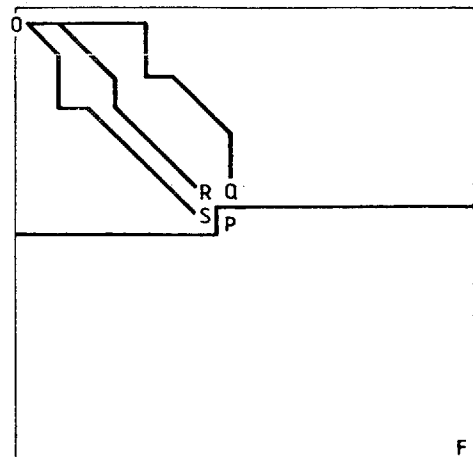
Les points du fond ici immobile se correspondent sur la diagonale du carré, les points du mobile sur une parallèle à la diagonale, dont l'écart horizontal à celle-ci donne le déplacement du mobile. Les points du fond disparus ou apparus n'ont pas de correspondant.

Si l'on décide de joindre les segments décrits, cela revient à associer tous les points cachés au premier point visible non apparu.

Bien qu'artificielle, cette mesure permet d'associer à une correspondance un chemin monotone sur les deux indices joignant les angles supérieur gauche et inférieur droit du carré.

Si l'on admet que deux points mis en correspondance doivent avoir des niveaux semblables, le chemin ne doit passer que par des cases associant des valeurs proches. Le carré contenant les écarts absolus entre les niveaux, le chemin doit emprunter les cases les moins coûteuses.

Un algorithme bien connu de programmation dynamique résoud le problème du chemin à coût minimal monotone sur les deux indices. Le principe en est le suivant : tout sous-chemin d'un chemin optimal est lui-même optimal.



SCHEMA 2

- Le chemin optimal menant en P est donc nécessairement;
- soit le chemin optimal menant à Q prolongé jusqu'à P
 - " " " R " "
 - " " " S " "

Si l'on connaît les coûts des chemins jusqu'à Q, R, S, on en déduit la direction à prendre en P vers l'origine et le coût minimal jusqu'en P.

Par un balayage de tout le carré (n x n), on trouve le chemin optimal menant en F et son coût.

Il est donc nécessaire d'effectuer et de mémoriser tout les choix (n^2). Pratiquement, si l'on majore les déplacements (d) on ne scrutera plus qu'une bande diagonale (2d x n). Ces valeurs sont relatives



à une paire de lignes. Toutes les lignes peuvent être considérées comme indépendantes et donc traitées à la suite ou en parallèle.

Cette méthode, que l'on retrouve en reconnaissance de la parole (3) et en reconnaissance syntaxique des formes (4) nécessite quelques adaptations mais semble présenter plusieurs avantages.

Elle réalise des choix locaux d'après des critères globaux ce qui lui permet "d'avalier" certaines différences en restant précise.

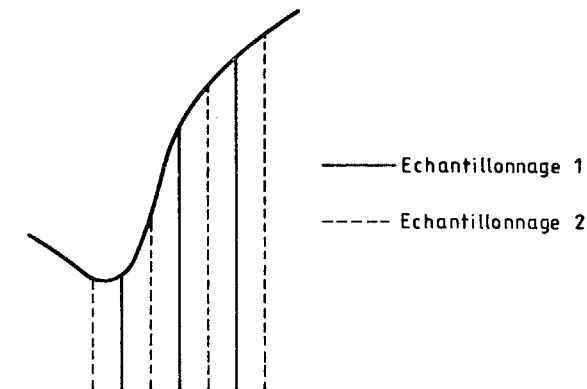
L'intérêt majeur semble être de ne pas réaliser a priori la partition de l'image en objets pour calculer ensuite leurs mouvements. Cela supposerait du reste un modèle des objets ici absent. Les choix sont effectués sans hypothèse particulière, le résultat donnant à la fois la position dans l'image et les mouvements des différentes composantes.

Les limites dues au principe de la méthode sont claires : on ne distingue pas deux objets en contact animés du même mouvement. Ils forment en effet une seule structure.

Notons les difficultés pratiques rencontrées :

- Les correspondances postulées par continuité sont virtuelles. Elles seront donc taxées de façon arbitraire. Trop coûteuse, elles pénalisent les objets en mouvement. Trop faibles, elles conduisent à ne plus rien identifier.

- L'échantillonnage étant assez lâche, un décalage d'un demi-point modifie les valeurs. (sch 3).

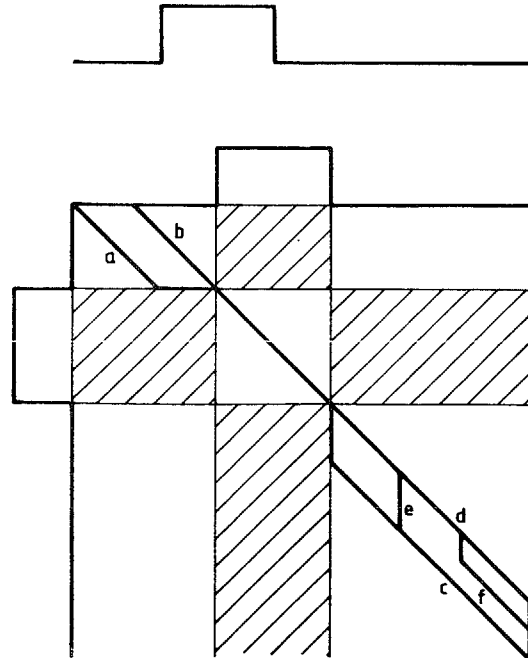


SCHEMA 3

- On rencontre des ambiguïtés dans les choix (sch.4)

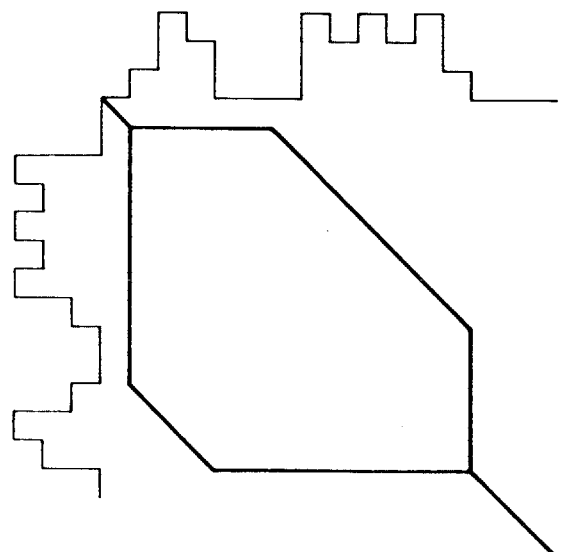
Les options a et b d'une part, c, d, e, et f d'autre part sont équivalentes. A l'interprétation, cela revient à attacher tout ou partie de la plage uniforme au mobile ou au fond, ce qui est logique, la

partition a priori des lignes étant inconnue. Seul un programme de niveau supérieur saura trancher.



SCHEMA 4

- Enfin dans le cas de plusieurs véhicules, notons que notre algorithme fonctionne convenablement tant qu'il n'y a pas de croisements, auquel cas on n'identifie qu'un véhicule parmi les deux (sch.5).



SCHEMA 5

RESULTATS EXPERIMENTAUX

Les résultats illustrent la séparation obtenue par le mouvement entre des mobiles de forme mal définie, parfois dans des situations de masquage. On présente des images dans lesquelles la couleur traduit le déplacement de chaque point homogène.

L'aspect infra-rouge ne devrait pas être important dans les expériences.

Extensions possibles :

On étudie en ce moment la généralisation aux translations non horizontales. D'autre part, par des techniques d'interpolation, on devrait accéder à une évaluation de déplacement plus précise. Une telle perspective pourrait permettre notamment de suivre un objet en rotation, les correspondances apparaissant sur un segment non parallèle à la diagonale. Enfin peut-être est-il envisageable à ce niveau de traitement d'intégrer une information relative à plus de deux images ou issue d'un modèle.

CONCLUSIONS

L'intérêt de la méthode semble être de travailler sur des petites images, et de donner des résultats de déplacements et de segmentations sans connaissance a priori sur ces structures.

BIBLIOGRAPHIE

Toutes les idées générales sur le traitement de séquences d'images sont illustrées dans (1), ou dans des références contenues dans (1).

- 1. (T.S HUANG) Sp. Verlag, Image Sequence Analysis
- 2. (D. CANO) Etude de l'ennuagement par analyse de séquences d'images satellite. Thèse de D.I. (ENST-E-82009).
- 3. (H. SAKOE) Two-Level DP-Matching. A dynamic programming-based pattern matching algorithm for connected word recognition. IEEE transactions on acoustics, speech and signal processing, Vol ASSP-27, n° 6, december 1979.
- 4. (S.Y.LU) A string - to - string correlation algorithm for image skeletonization - proceedings of the 6 th ICPR (MUNICH) October 1982.
- 5. Special issue on motion and time-varying imagery IEEE tr on PAMI. vol-PAMI-2, n°6, nov. 1980.

