TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

Segmentation de Régions Spatio-temporelles dans une Séquence d'Images Acquise par une Caméra en Mouvement *

Jean-Pierre Gambotto

MATRA/LTIS
38, Bd. Paul Cezanne - F78052 Saint-Quentin-en-Yvelines Cedex (France)

Résumé. Cet article présente un algorithme pour la segmentation d'une séquence d'images en régions spatio-temporelles. L'algorithme proposé associe une méthode de croissance de régions et une méthode de poursuite de régions. Cette approche utilise la cohérence temporelle afin d' éviter la phase de mise en correpondance inter-images. L'algorithme est appliqué à plusieurs séquences d'images acquises par une caméra en mouvement.

Abstract. This paper presents an algorithm for the segmentation of an image sequence into spatio-temporal regions. This algorithm combines region growing and region tracking. The proposed approach relies on the temporal coherence of the image sequence and no correspondence analysis is needed. The algorithm is applyed to image sequences which have been acquired with a moving camera.

1 Introduction

Il existe deux grandes catégories d'approches pour calculer le mouvement. Dans la première, les primitives sont extraites de chaque image, puis une mise en correspondance permet d'estimer le mouvement. Dans la seconde, la correspondance est obtenue directement entre les niveaux de gris et la segmentation du flot optique fournit des zones de mouvement homogènes. Les techniques qui calculent le mouvement à partir d'un grand nombre d'images sont cependant plus robustes. Bolles [3] utilise une séquence d'images acquise à cadence élevée. Cette séquence est considérée comme un volume spatio-temporel et les trajectoires des primitives sont analysées dans des coupes du volume suivant l'axe temporel. Dans cette approche, on suppose que le mouvement de la caméra est connu et que la scène est fixe. Une généralisation est proposée par Peng [11].

Dans un article récent, Baker [1] décrit un algorithme permettant de construire des surfaces 3D dans le volume spatio-temporel. La convolution de l'image par le laplacien d'une gaussienne fournit une segmentation où chaque région a pour unique attribut le signe du laplacien. La simplicité de cet algorithme provient de deux choix essentiels. Tout d'abord, les dimensions spatiales et temporelle ne sont pas traitées de la même façon. En effet, la construction récursive des volumes a lieu suivant une seule direction (l'axe temporel). D'autre part, la dichotomie entre régions de signes différents facilite la création de volumes connexes et de surfaces fermées. Ce second choix est en fait trop restrictif dans la mesure où l'on souhaite caractériser une région par différents attributs et donc utiliser d'autres algorithmes de segmentation .

Nous proposons un nouvel algorithme pour la segmentation d'une séquence d'images en régions 3D. L'approche adoptée associe un algorithme de croissance de régions et un algorithme de poursuite de régions afin d'éviter la phase de correspondance inter-images. Contrairement à [1], une segmentation complète est nécessaire seulement pour la première image. Dans les images suivantes, la cohérence temporelle est utilisée pour trouver le noyau de chaque région. Nous utilisons la définition suivante: il y a cohérence temporelle lorsque les régions correspondantes dans deux images consécutives se superposent. Cet algorithme est utilisé pour la poursuite d'objets acquis par une caméra en mouvement. Il contient deux phases principales: (1) segmentation de la première image à l'aide d'un algorithme de croissance de régions [8]; à partir de cette segmentation, un ensemble de régions est selectionné sur la base de critères de forme, de contraste ou de convexité; (2) segmentation spatio-temporelle des régions ainsi déterminées. Dans cet article nous présentons cette seconde phase.

*Cette étude a été en partie financée par le contrat DRET No. 88.329

2 Segmentation Spatio-temporelle

L'algorithme utilise la région segmentée dans l'image $I(t-1)$ pour trouver la région correspondante dans l'image $I(t)$. Il comprend les étapes suivantes:

- mise à jour du modèle de région,
- estimation et interprétation élémentaire du mouvement,
- détermination du noyau dans l'image $I(t)$,
- croissance du noyau et détection du contour de la région.

2.1 Le modèle de region

Chaque région de l'image $I(t)$ est représentée par un modele $Reg(t)$ qui comprend trois composantes.

Modèle Radiométrique: Le modèle radiométrque comprend la moyenne et la variance des niveaux de gris sur la région. Deux autres paramètres niv_{min} et niv_{max} sont obtenus à partir de l'histogramme. Leur valeur est définie afin d'avoir une proportion importante des niveaux de gris de la région dans l'intervalle $[niv_{min}, niv_{max}]$.

Modèle géométrique: Il comprend :

- la surface S de la région ,
- la fenêtre sur laquelle est définie la région ,
- une approximation de l'enveloppe convexe ,
- une approximation polygonale du contour ,
- une structure décrivant le masquage de la région .

Modèle dynamique: les paramètres utilisés sont:

- le centre de gravité ,
- la translation inter-images optimale (T_x^*, T_y^*),

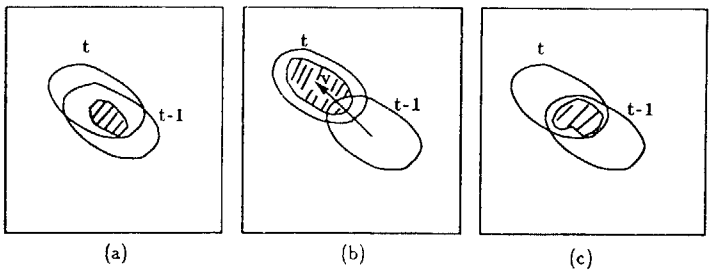

Figure 1: Détermination du noyau (zone hachurée)

• un vecteur de changement de taille $(\kappa x, \kappa y)$.

2.2 Estimation du mouvement

Cette étape consiste à prédire les changements géométriques dans le plan image. Dans cette étude, nous considérons le mouvement de translation, les changements de forme des régions ainsi que les masquages sur les bords de l'image. Le but de cette analyse est de fournir des informations fiables pour la détermination des noyaux. L'estimation optimale du mouvement sera effectuée ultérieurement, lorsque les deux régions auront été segmentées. Soient $Mask(t)$ le masque de la région $R(t)$ et $\tilde{M}ask(t)$ le masque du noyau au temps t, l'idée qui est mise en oeuvre consiste à rechercher pour chaque région, une translation et un changement de taille entre $Mask(t-2)$ et $Mask(t-1)$ afin de prédire la position et la taille de $\tilde{M}ask(t)$.

2.2.1 Translation optimale

Les techniques suivantes ont été explorées:

• recalage des centres de gravité des deux masques,
• recalage des sommets de l'approximation polygonale,
• correlation de masques binaires.

La première méthode fournit une très bonne estimation du mouvement en l'absence de masquage ou d'erreurs de segmentation trop importantes. La seconde méthode est bien adaptée aux régions présentant des arêtes rectilignes, mais est assez difficile à utiliser dans le cas général. La dernière méthode est intéressante car elle permet de prendre en compte les déformations dues aux masquages entre régions [10] . Elle consiste à rechercher la translation (T_x^*, T_y^*) qui donne une intersection maximale entre les deux régions.

2.2.2 Paramètres de changement de taille

Le changement de taille entre deux régions $R1$ et $R2$ est obtenu à partir du rectangle de surface minimale [7] associé à la région $R1$. Ce rectangle est décrit par son orientation $\theta 1$ et ses dimensions $\delta x1$ et $\delta y1$ dans le système de coordonnées $x_{\theta 1}$ et $y_{\theta 1}$. Les sommets de l'enveloppe convexe de la région $R2$ sont ensuite projetés sur les axes $x_{\theta 1}$ et $y_{\theta 1}$. Les extrémums de ces projections donnent un nouveau rectangle de dimensions $\delta x2_{\theta 1}$ et $\delta y2_{\theta 1}$. On obtient les paramètres: $\kappa x = \delta x2_{\theta 1}/\delta x1$ et $\kappa y = \delta y2_{\theta 1}/\delta y1$.

2.3 Détermination d'un noyau

Trois méthodes sont définies pour obtenir un noyau dans l'image $I(t)$ à partir d'une région segmentée dans l'image $I(t-1)$. Ces méthodes sont schématisées sur la figure 1.

2.3.1 Erosions multiples

La première méthode consiste à éroder plusieurs fois la région $R(t-1)$. Le nombre d'érosions doit être au moins égal au déplacement inter-images maximum Δmax. Cette méthode est très générale car seule la connaissance de Δmax est nécessaire, et la direction du mouvement peut être inconnue. Elle présente cependant certains inconvénients. Tout d'abord, l'érosion étant effectuée sur tout le contour de la région, la surface érodée peut être relativement grande par rapport à la surface totale de la région. D'autre part, le critère de cohérence temporelle n'est pas une condition suffisante pour l'obtention d'un noyau. En effet, comme le suggère la figure 1a, le noyau obtenu par érosions successives disparait alors que l'intersection entre les masques est encore relativement importante.

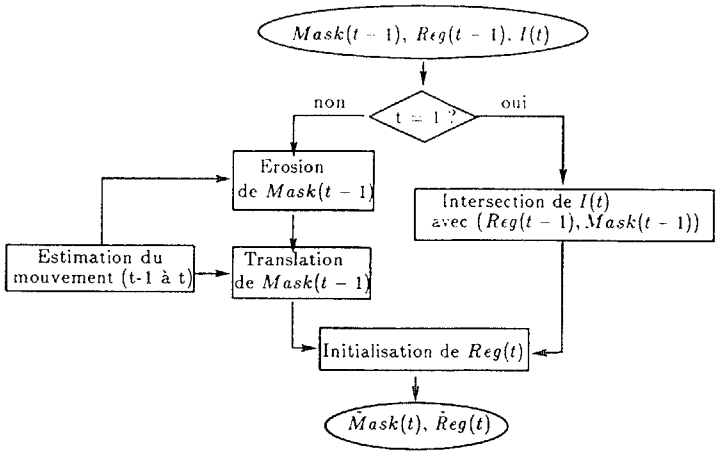

Figure 2: Création du masque des noyaux $\tilde{M}asq(t)$ et initialisation des régions correspondantes $\tilde{R}eg(t)$.

2.3.2 Erosion et prédiction du mouvement

La connaissance du mouvement permet de déterminer le noyau de façon plus précise. Cette méthode comprend deux étapes : (1) érosion du masque $R(t-1)$, (2) translation du masque érodé. L'érosion permet de prendre en compte les changements de forme et de taille de la région. Deux paramètres modélisent ces erreurs. Le premier d_{err} rend compte des petites erreurs de segmentation. Le second $d_{incr} = \max[(\kappa x - 1)\delta x, (\kappa y - 1)\delta y)]$ modélise les déformations. Ces déformations ont pour causes principales le changement de perspective et les masquages. Le nombre d'érosions est égal à la valeur entière de $(d_{err} + d_{incr} + 1)$. La seconde étape utilise l'estimation du mouvement pour recentrer le noyau sur la position prédite de la région. Cette méthode résout un certain nombre de problèmes. Elle peut en particulier être utilisée avec des régions très petites. D'autre part, l'érosion étant très faible, la phase de croissance est très rapide. Notons aussi que la méthode est utilisable dans le cas d'un déplacement supérieur à Δ_{max}.

2.3.3 Intersection entre le modèle radiométrique de $R(t-1)$ et l'image $I(t)$

La méthode utilise la cohérence entre les radiométries des deux images pour calculer le noyau. Elle consiste à rechercher les pixels de $I(t)$ dont les niveaux de gris appartiennent à l'intervalle $[niv_{min}, niv_{max}]$ et qui intersectent la région $R(t-1)$. En l'absence d'information sur le mouvement, on obtient un noyau appartenant à l'intersection des deux régions (cf. figure 1c).

2.3.4 Stratégie adoptée

Une étude expérimentale a permis de définir une stratégie pour créer le masque des noyaux $\tilde{M}ask(t)$ et pour initialiser toutes les régions de l'image $I(t)$. La première méthode n'a pas été retenue, car dans la plupart des séquences étudiées, la valeur du paramètre Δmax est relativement élevée par rapport à la taille des régions. Cependant, cette méthode est efficace pour définir les noyaux des grandes régions. La stratégie adoptée est représentée sur la figure 2. Le masque des noyaux $\tilde{M}ask(t)$ est obtenu à partir du masque des régions $Mask(t-1)$ et du modèle de région $Reg(t-1)$. La poursuite est initialisée en utilisant la troisième méthode.

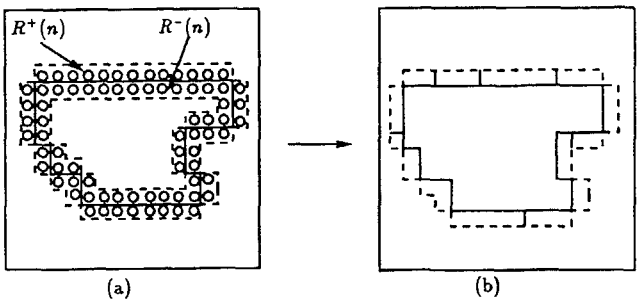

Figure 3: Région $R(n)$ (trait plein) et ensembles des pixels frontière $R^+(n)$ et $R^-(n)$ de part et d'autre du contour de la région (a), et segments de frontière (b).

Par la suite, le module d'estimation du mouvement prédit les changements inter-trames (translations et changements de taille) et la seconde méthode est utilisée.

2.4 Segmentation de chaque région

L'approche est fondée sur la coopération entre un algorithme de croissance de région et un algorithme de détection de contours. Le signal image est considéré comme une surface continue par morceaux. Cependant, la méthode est plus simple que celle de Besl et Jain [2] car le modèle de surface est d'ordre faible et l'approximation de surface n'est pas utilisée pour arrêter la croissance. Celle-ci est réalisée par une détection de contour. Le signal sur la région R est représenté par une surface continue $SM(R)$ perturbée par un bruit blanc $n_{i,j}$ de moyenne nulle et d'écart-type σ:

$$Y_{i,j} = Ai + Bj + C + n_{i,j}$$

Deux modèles de surface ont été étudiés: le modèle de facette plane décrit par les paramètres A, B et C, et le modèle correspondant à une surface constante ($B = C = 0$). Soit $d[Y_{i,j}, SM(R)]$ la distance euclidienne entre le pixel $Y_{i,j}$ et le modèle $SM(R)$, nous calculons l'histogramme suivant:

$$|d[Y_{i,j}, SM(R)] - d[Y_{i-k,j-l}, SM(R)]| \text{ où } k, l = 0, 1$$

La médiane thg de cet histogramme permet de décrire les variations du signal par rapport au modèle de surface. Nous donnons ici les grandes lignes de l'algorithme [9].

2.4.1 Croissance itérative

A l'itération n, la croissance est réalisée en analysant les pixels situés de part et d'autre de la frontière de la région (cf. figure 3a). Une itération comprend les étapes suivantes:

1. Segmentation de la frontière extérieure $R^+(n)$:
Cette étape a pour but d'obtenir des informations fiables sur la frontière de la région. L'algorithme de regroupement hiérarchique [6] est utilisé pour partitionner $R^+(n)$ en segments de longueurs variables.

2. Regroupement des segments de frontière avec $R(n)$:
Un segment S_k de la frontière $R^+(n)$ est regroupé avec $R(n)$ si sa distance au modèle de région est inférieur au seuil thg. Si aucun segment ne vérifie ce critère, le segment le plus proche de $R(n)$ est regroupé. Il y a donc au moins un segment qui est regroupé à chaque itération.

3. Mise à jour du modèle de région:
Le modèle est mis à jour à chaque itération. Ce mode de fonctionnement permet de segmenter des régions qui n'ont pas toujours des propriétés statistiques très homogènes.

2.4.2 Détection des contours et fin de la croissance

L'arrêt de la croissance est obtenu en calculant, à chaque itération, le gradient moyen sur la frontière de la région:

$$F(n) = \sum G(k,l)/P(n)$$

où $P(n)$ est le périmètre de la région et le gradient $G(k,l)$ au point (k,l) est une des fonctions suivantes:

1. $G(k,l) = |Y_{i,j} - Y_{k,l}|$ où $Y_{i,j}$ et $Y_{k,l}$ sont deux pixels adjacents tels que $Y_{i,j} \in R^+(n)$ et $Y_{k,l} \in R^-(n)$.
2. Le gradient de Canny-Deriche [5] ou de Castan-Shen [4].

Le critère d'arrêt consiste à rechercher un maximum au cours des itérations ainsi qu'un maximum localement dans le plan image.

2.4.3 Décroissance

Cette dernière étape permet de retrouver le contour optimal. Elle consiste à modifier les étiquettes d'une liste de pixels créée à l'étape précedente.

L'algorithme traite les régions en séquence ou en parallèle. Dans ce dernier cas, un critère d'inhibition évite le recouvrement des régions. D'autre part, le gradient pouvant diminuer au cours du temps, une contrainte temporelle limite les changements de taille et assure la stabilité du processus de propagation.

3 Résultats

Nous présentons les résultats obtenus sur deux séquences. La première contient 45 images et a été acquise par une caméra installée dans une voiture. La segmentation spatio-temporelle (figure 4a) est correcte malgré un mouvement rapide des régions sur le bord de l'image. On remarque aussi que les positions des noyaux sont assez fluctuantes par rapport aux positions réelles des régions. Ceci est dû principalement à l'absence de stabilisation de la caméra. La seconde séquence a été obtenue par synthèse d'images à l'aide d'un modèle géométrique et d'une image réelle de la scène. L'image correspondant à une position donnée est obtenue par reprojection des radiométries de l'image réelle sur le modèle 3D. La séquence de 36 images ainsi obtenue (cf. figure 5) modélise la vision d'une caméra se déplaçant sur une trajectoire rectiligne. La figure 5 montre les régions segmentées ainsi que les volumes spatio-temporels obtenus.

4 Conclusions

Nous avons présenté un nouvel algorithme pour segmenter une séquence d'images en régions spatio-temporelles. Les résultats obtenus démontrent la faisabilité de l'approche spatio-temporelle qui permet en particulier de: (1) segmenter des régions dans une image, puis propager cette segmentation dans les images suivantes, (2) maintenir un modèle de région au cours du temps afin d'assurer la stabilité de la méthode de propagation. La méthode est utilisée pour poursuivre des régions dans une séquence d'images acquise par une caméra en mouvement. Elle est efficace même dans le cas de mouvements importants.

Références

[1] H. H. Baker. Building surfaces of evolution: the weaving wall. In *DARPA Workshop on Image Understanding*, pages 1031–1040, 1988.

[2] P. Besl and R. Jain. Segmentation through variable-order surface fitting. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 10(2):167–192, March 1988.

[3] R.C. Bolles et al. Epipolar-plane image analysis: an approach to determinig structure from motion. *International Journal of Computer Vision*, 7–55, 1987.

[4] S. Castan and J. Shen. Optimal filter for edge detection. methods and results. In *Proceedings of the European Conference on Computer Vision*, pages 13–17, Antibes, France, April 1990.

[5] R. Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):, 1987.

[6] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New-York, 1973.

[7] H. Freeman. Computer processing of line-drawing images. *ACM Computing Surveys*, 6(1):57–97, 1974.

[8] J.P. Gambotto. Estimation récursive de la moyenne de signaux bidimensionnels: vers une approche parallèle de la segmentation d'images. In *Colloque GRETSI*, page , Nice, May 1985.

[9] J.P. Gambotto. A new approach to combining region growing and edge detection. *submitted*, 1991.

[10] J.P. Gambotto. Segmentation and interpretation of infrared image sequences. In Huang T.S., editor, *Time-Varying Imagery*, pages 1–38, JAI Press, 1988.

[11] S.L. Peng and G. Medioni. Interpretation of image sequence by spatio-temporal analysis. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 344–351, Irvine, CA, March 1989.

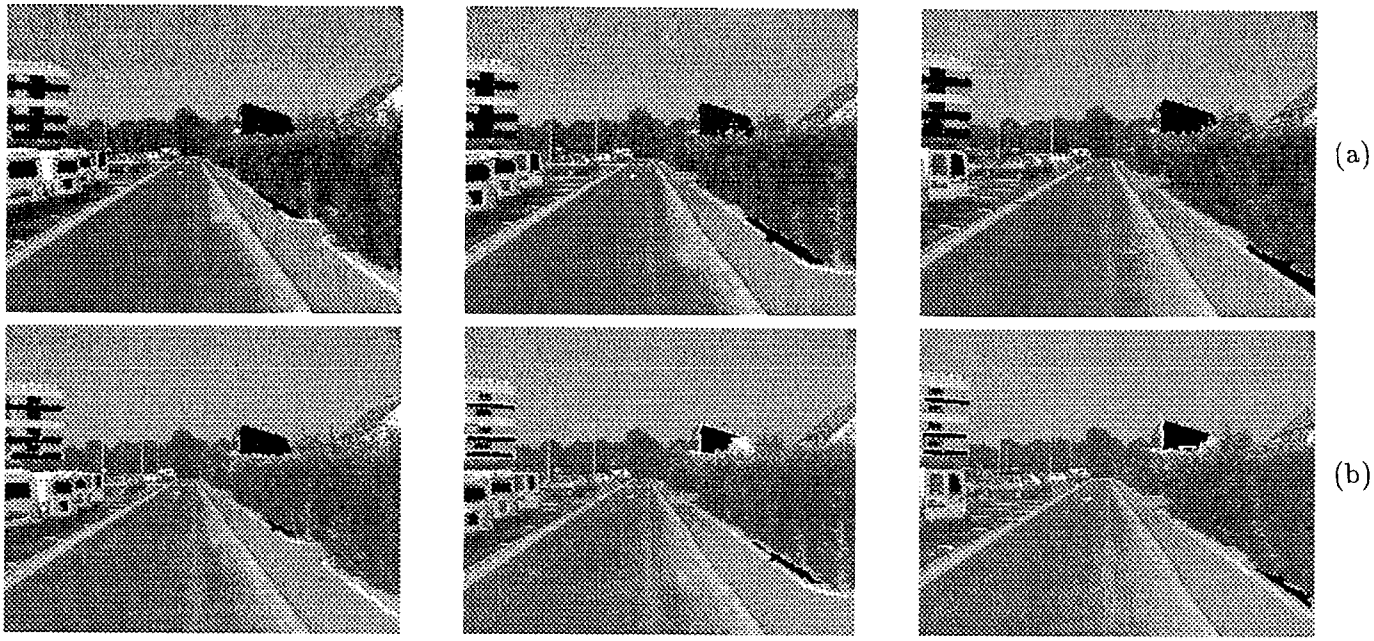

Figure 4: Séquence 1: segmentation spatio-temporelle (a) et noyaux (b).

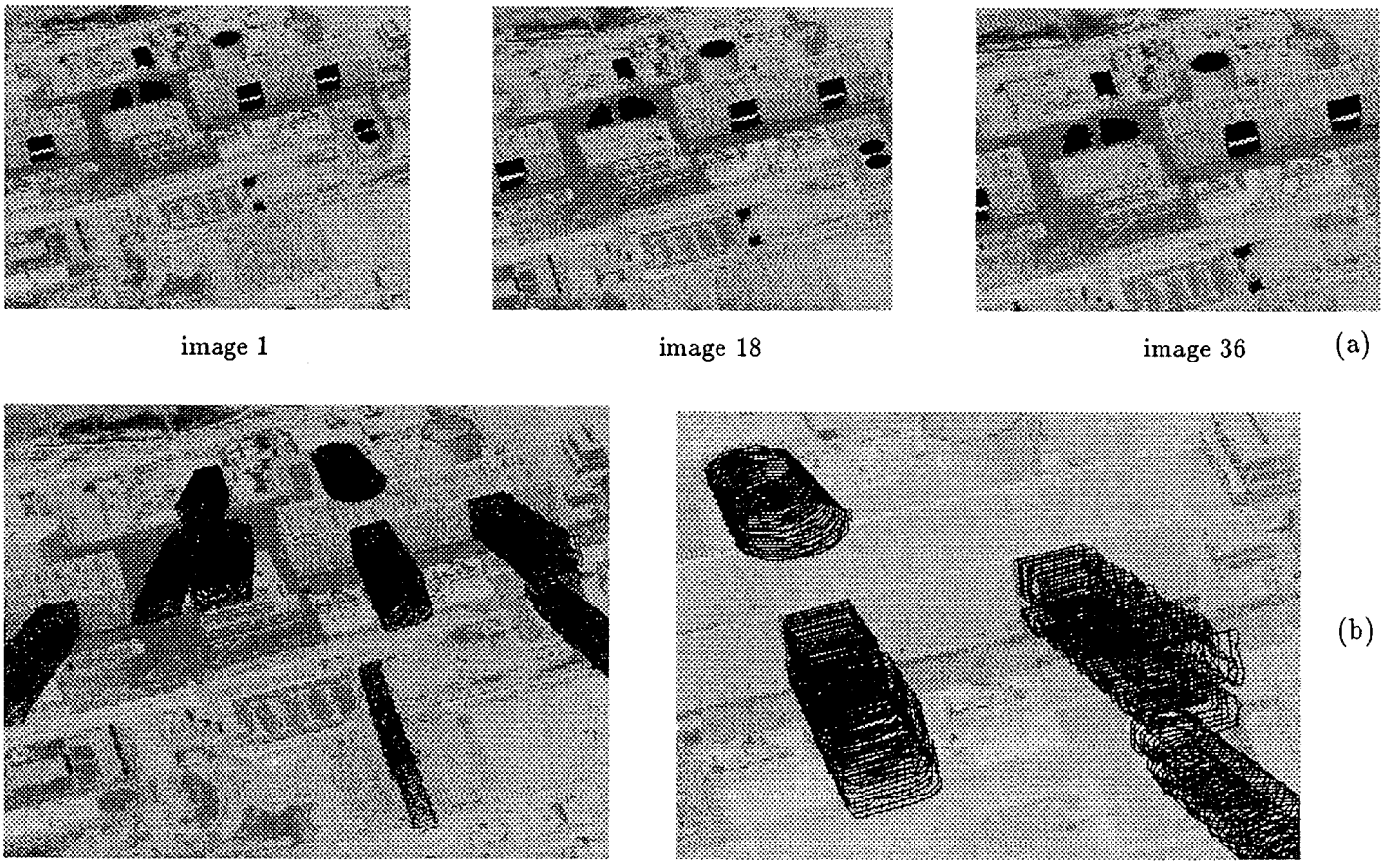

Figure 5: Séquence 2: segmentation spatio-temporelle (a) et volumes spatio-temporels (b).