

Détection et poursuite de cibles par minimisation de la complexité stochastique

Jean-François Boulanger¹, Frédéric Galland¹, Philippe Réfrégier¹ et Pascal Martin²

1 : Institut Fresnel, CNRS, Aix-Marseille Université, Ecole Centrale Marseille, Campus de Saint Jérôme, 13013 Marseille, France

2 : Kaolab, ZAC les Frênes, 209 rue du Bouleau, 13109 Simiane Collongue, France

Contact : jf.boulanger@fresnel.fr

Résumé

L'accroissement de la puissance des ordinateurs, la miniaturisation des capteurs optiques et la baisse de leurs coûts de production ont favorisé le développement d'applications de vidéosurveillance. Dans ce contexte, il est nécessaire de développer des algorithmes rapides et non supervisés de détection et de poursuite de cibles. Nous avons récemment proposé une méthode de poursuite de cibles fondée sur la minimisation de la complexité stochastique. Nous généralisons cette approche au cas de la détection de cibles et nous présentons ensuite le couplage de ces algorithmes afin de suivre l'évolution de plusieurs cibles sur des séquences d'images.

Abstract

Video surveillance applications have been largely developed due to the growth of computer power, the miniaturisation and the production cost decrease of optical sensors. In this context, it is necessary to develop fast and unsupervised target detection and tracking algorithms. We have recently proposed a tracking method based on the minimisation of the stochastic complexity. We generalize this approach to target detection. The coupling of these algorithms is then illustrated on video sequences.

Mots-clés : Détection et poursuite de cibles, Vidéosurveillance, Complexité stochastique.

Keywords: Target detection and tracking, Video surveillance, Stochastic complexity.

1. Introduction

De nombreuses applications de vidéosurveillance nécessitent la détection et la poursuite des différentes cibles en mouvement présentes sur l'image. Cette étape est nécessaire afin de permettre une analyse précise des événements qui se déroulent dans une scène (interaction entre deux personnes, détection de comportements dangereux, vitesse excessive, chute, etc). De plus, avec l'augmentation du nombre de caméras ainsi que de leur résolution, il est nécessaire de développer des solutions qui soient d'une part très rapides afin de pouvoir respecter la contrainte temps-réel requise pour de nombreuses applications (voire même de pouvoir traiter en parallèle plusieurs séquences vidéos) et d'autre part qui nécessitent peu d'intervention de la part d'un utilisateur.

Nous avons récemment proposé une technique de poursuite mono-cible [1] par minimisation de la complexité stochastique (CS) [4] adaptée aux séquences vidéos avec caméra fixe. Cette technique repose sur la minimisation d'un critère qui nécessite peu de temps de calculs et qui combine les avantages des techniques de soustraction de fond et ceux des techniques reposant sur des mesures de similarité avec un modèle de cible. Cependant, comme pour la plupart des algorithmes de poursuite de cibles, il est nécessaire d'initialiser l'algorithme en déterminant la position et la forme de la cible sur la trame où celle-ci apparaît, ce qui revient à coupler l'étape de poursuite à une étape préalable de détection de cibles.

Dans cette communication, nous généralisons l'approche proposée pour la poursuite de cible au cas de la détection de cibles afin de définir un cadre théorique permettant d'obtenir une méthode

rapide de détection et de poursuite de cibles par caméra fixe nécessitant le réglage d'un nombre réduit de paramètres de la part d'un utilisateur.

2. Poursuite de cibles par minimisation de la complexité stochastique

Dans cette partie, nous présentons succinctement l'approche introduite dans [1] qui repose sur la minimisation de la CS de chaque trame.

Dans le cas qui nous intéresse, la caméra est supposée fixe et l'arrière plan de l'image reste fixe tout au long de la séquence. Il peut donc être appris (soit au cours d'une période d'initialisation, soit directement au cours de la séquence), ce qui permet d'estimer la densité de probabilité (ddp) $P_{i,j}^F$ des fluctuations des niveaux de gris en chaque pixel (i, j) de l'arrière plan de l'image. Différentes méthodes de poursuite de cibles fondées sur la détection sur chaque trame des pixels dont les valeurs ont significativement évolué par rapport au modèle d'arrière plan ont été proposées [5,7]. D'autres approches reposent sur la recherche de la similarité entre les cibles poursuivies sur les trames précédentes et différentes régions de l'image analysée [2,3].

Le critère de complexité stochastique permet de prendre en compte simultanément l'écart par rapport à un modèle d'arrière plan et la similarité à un modèle de cible [1]. Soit s^t la trame à l'instant t . Soit Ω une région de l'image correspondant à une cible potentielle et $\bar{\Omega}$ son complémentaire dans l'image. La CS de l'image associée à la région cible Ω correspond à la longueur de code nécessaire pour décrire Ω (position, forme, etc) ainsi que les niveaux de gris des pixels dans chacune des deux régions Ω et $\bar{\Omega}$.

Dans [1], la CS est déterminée à partir des hypothèses suivantes :

1. A l'intérieur de la région Ω , les niveaux de gris des pixels doivent être en adéquation avec la ddp P^C estimée sur la cible sur les trames précédentes.
2. A l'extérieur de cette région (dans $\bar{\Omega}$), si on suppose que seule la cible est en mouvement, les niveaux de gris des pixels doivent être en adéquation avec la ddp $P_{i,j}^F$ de l'arrière plan de la scène.

Pour obtenir de faibles temps de calculs, les hypothèses simplificatrices suivantes sont également introduites :

- La région cible Ω est de forme rectangulaire suivant les axes verticaux et horizontaux : seules sa position, sa hauteur et sa largeur sont donc estimées, ce qui permet de mettre en œuvre une implémentation fondée sur l'utilisation d'images intégrales [6].
- Les ddp $P_{i,j}^F$ et P^C sont supposées gaussiennes.

La moyenne m_C et l'écart type σ_C de la ddp P^C sont estimés, sur la région Ω au sens du maximum de vraisemblance sur la trame où la cible apparaît.

L'expression de la CS obtenue dans [1] est alors :

$$\Delta_P(s^t, \Omega) = 2 \log N + \left[N_\Omega \log \sigma_C + \sum_{(i,j) \in \Omega} \frac{(s^t(i,j) - m_C)^2}{2\sigma_C^2} \right] + \sum_{(i,j) \in \bar{\Omega}} \left[\log \sigma_F(i,j) + \frac{(s^t(i,j) - m_F(i,j))^2}{2\sigma_F^2(i,j)} \right] \quad (1)$$

où N et N_Ω correspondent aux nombres de pixels dans l'image et dans la région Ω . La moyenne et l'écart type de la ddp P^C de la cible et de la ddp $P_{i,j}^F$ de l'arrière plan au pixel (i, j) sont notés respectivement m_C , σ_C et $m_F(i, j)$, $\sigma_F(i, j)$.

L'objectif est alors de déterminer la région Ω qui minimise la CS dans une fenêtre de recherche autour de la localisation de la région cible sur la trame précédente. Pour les résultats qui seront présentés, cette fenêtre est obtenue en dilatant la région cible estimée sur la trame précédente de 20 pixels. Cette optimisation se fait en deux étapes : à partir de la position de la cible sur la trame précédente, nous estimons d'abord sa translation puis sa déformation dans le cas où une seule cible est présente dans l'image. Cette approche permet d'estimer la position et la forme de la cible en quelques milli-secondes par trame sur des images 320×240 pixels (voir [1] pour plus de détails).

3. Détection de cibles

Il est possible de généraliser l'approche décrite précédemment pour détecter l'apparition de nouvelles cibles dans l'image. L'hypothèse 2 de la section précédente reste valide : la région $\overline{\Omega}$ est toujours supposée constituée uniquement de pixels appartenant à l'arrière plan. En revanche, l'hypothèse 1 doit être modifiée dans la mesure où Ω sur la trame courante ne correspond plus à une cible déjà détectée, mais à une nouvelle cible. Afin d'aboutir à des algorithmes rapides, les niveaux de gris des pixels de la région Ω sont toujours supposés distribués selon une ddp gaussienne, mais dont les paramètres sont cette fois-ci inconnus. Les estimés au sens du maximum de vraisemblance \hat{m} et $\hat{\sigma}$ de la moyenne et de la l'écart type de la ddp dans Ω sur la trame courante sont alors substitués aux vraies valeurs qui sont inconnues afin de déterminer la CS :

$$\Delta_D(s^t, \Omega) = 2 \log N + \left[N_\Omega \log \hat{\sigma} + \frac{N_\Omega}{2} + \log N_\Omega \right] + \sum_{(i,j) \in \overline{\Omega}} \left[\log \sigma_F(i,j) + \frac{(s^t(i,j) - m_F(i,j))^2}{2\sigma_F^2(i,j)} \right] \quad (2)$$

Dans cette équation, seul le deuxième terme est modifié par rapport à l'équation (1). Ce terme représente la longueur de code nécessaire pour décrire à la fois les fluctuations des niveaux de gris à l'intérieur de Ω , connaissant les paramètres \hat{m} et $\hat{\sigma}$ de la ddp, auquel il est nécessaire d'ajouter la longueur de code $\log N_\Omega$ pour décrire ces paramètres. Afin de déterminer si Ω constitue effectivement une cible, on considère les deux modèles d'image qui correspondent respectivement à l'apparition d'une cible ou à l'absence de cible et on choisit celui qui minimise la CS. Cela revient à comparer la CS $\Delta_D(s^t, \Omega)$ à la CS $\Delta_D(s^t, \emptyset)$ obtenue sous l'hypothèse qu'aucune cible n'est présente :

$$\Delta_D(s^t, \emptyset) = \sum_{(i,j) \in \text{Image}} \left[\log \sigma_F(i,j) + \frac{(s^t(i,j) - m_F(i,j))^2}{2\sigma_F^2(i,j)} \right] \quad (3)$$

Une cible est alors détectée si $\Delta_D(s^t, \Omega) < \Delta_D(s^t, \emptyset)$. Pour détecter l'ensemble des cibles présentes dans l'image, l'image est divisée en un maillage régulier délimitant des carrés de 8×8 pixels qui constituent des régions de cibles potentielles. L'algorithme de détection consiste alors à tester pour chacune de ces régions si elle correspond à une cible en comparant $\Delta_D(s^t, \Omega)$ et $\Delta_D(s^t, \emptyset)$. Cette méthode de détection de cible requiert une dizaine de milli-secondes par trame pour des images de dimensions 320×240 pixels. On peut montrer que cette approche peut être également interprétée comme correspondant à un test d'un rapport de vraisemblance généralisé dont le seuil est déterminé automatiquement à l'aide du principe de minimisation de la CS.

4. Exemples de résultats

Nous présentons sur la première ligne de la figure 1 un exemple de résultats sur une séquence vidéo lorsque l'on couple le résultat de détection de cibles avec la méthode de poursuite mono-cible que nous avons définie dans [1].

Les paramètres $m_F(i,j)$ et $\sigma_F(i,j)$ du modèle de fond en chaque pixel ont été calculées sur les 200 premières images de la séquence vidéo à l'aide d'estimateurs robustes (utilisation de la médiane pour l'estimation de $m_F(i,j)$ et du premier quartile pour l'estimation de $\sigma_F(i,j)$). Les valeurs de $m_F(i,j)$ sont affichées sur l'image (a) de la figure 1. Nous exposons sur l'image (b) les résultats de la détection de cibles sur la trame 201 de la séquence vidéo. Comme les cibles correspondent souvent à plusieurs régions issues du maillage, nous définissons alors des boîtes englobantes qui contiennent toutes les régions de cibles adjacentes issues de la détection. Ces boîtes englobantes (voir fig 1.c) servent alors à initialiser l'algorithme de poursuite mono-cible. La cible est alors poursuivie trame à trame, nous affichons sur l'image (d) un résultat de poursuite obtenu.

5. Perspectives

La première perspective est de généraliser l'algorithme de poursuite mono-cible proposé dans [1] au cas de la poursuite de plusieurs cibles. Une possibilité consiste à appliquer l'algorithme de poursuite d'une cible unique sur chacune des cibles séparément en ajoutant une contrainte interdisant le recouvrement de deux régions cibles et en réalisant une recherche de la localisation

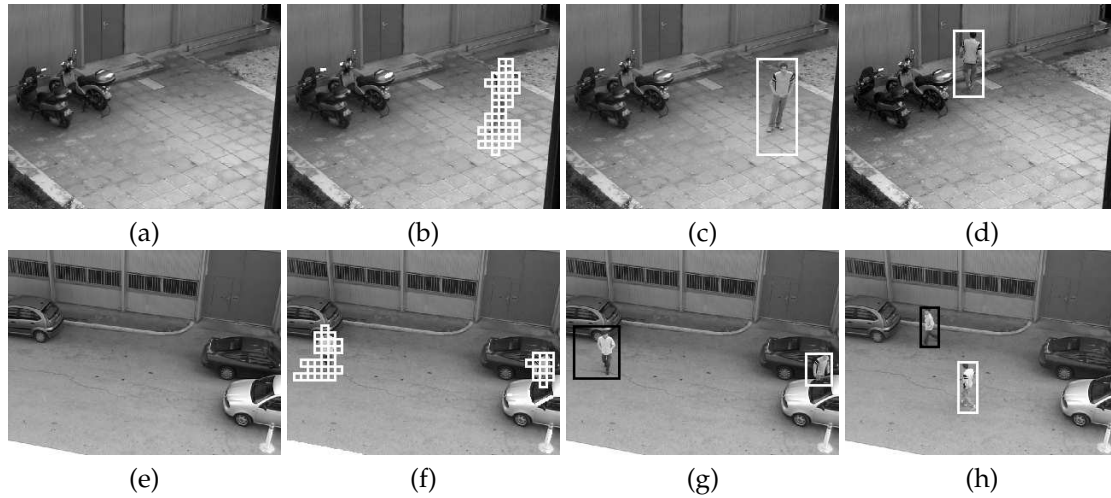


FIG. 1 – Exemples de résultats du couplage de la détection et de la poursuite dans le cas d’une unique cible qui pénètre et évolue dans la scène surveillée sur la ligne 1 ou de plusieurs cibles sur la ligne 2. (a,e) Images des moyennes $m_F(i, j)$ du fond estimées à l’aide d’un estimateur robuste (médiane) au cours d’une période précédant la poursuite (trames 1 à 200). (b,f) Résultat de la détection appliquée sur la trame 201. (c,g) Création des boîtes englobantes à partir de la détection obtenue en (b) ou (f) et qui sert d’initialisation à l’algorithme de poursuite. (d,h) Images extraites de la séquence de résultats de l’algorithme de poursuite (trames 900 pour la première ligne et trame 720 pour la deuxième ligne). Les images sont de dimension 320×240 pixels.

de la cible non plus sur une fenêtre de recherche mais sur toute l’image. Nous présentons sur la deuxième ligne de la figure 1 un premier résultat de détection et de poursuite obtenu sur une séquence où deux cibles évoluent.

De plus, il serait intéressant d’évaluer si l’utilisation de ddp plus complexes que des ddp gaussiennes (mélanges de gaussiennes, histogrammes, ddp estimées par des estimateurs à noyaux, ...) permet d’améliorer les performances de la méthode de manière significative sans un trop grand surcoût en temps de calculs.

Bibliographie

1. J.-F. Boulanger, F. Galland, P. Martin, et P. Réfrégier. Stochastic complexity integral image based technique for fast video tracking. *Optics Letters*, 33 :2521–2523, 2008.
2. D. Comaniciu, V. Ramesh, et P. Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(5) :564–575, 2003.
3. P. Pérez, C. Hue, J. Vermaak, et M. Gangnet. Color-based probabilistic tracking. In *European Conf. in Computer Vision (ECCV’2002)*, pages 661–675, Copenhague, Denmark, jun 2002.
4. J. Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 sur *Series in Computer Science*. World Scientific, Singapore, 1989.
5. C. Stauffer et W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8) :747–757, 2000.
6. Paul Viola et Michael Jones. Rapid object detection using a boosted cascade of simple features. *copr*, 01 :511, 2001.
7. C. R. Wren, A. Azarbayejani, T. Darrell, et A. Pentland. Pfunder : Real-time tracking of the human body. *IEEE Trans. on Patt. Anal. and Mach. Intel.*, 19(7) :780–785, 1997.

Nous remercions le Conseil Régional Provence-Alpes-Côte d’Azur pour son soutien financier dans le cadre de la thèse de Jean-François Boulanger.
