Détection et suivi simultanés dans une séquence vidéo en couleur par filtres particulaires

Jacek Czyz[1], Branko Ristic[2], Benoit Macq[1]

[1] UCL-TELE, Batiment Stévin, Place du Levant 2, 1348 Louvain-la-Neuve, Belgique

[2] DSTO, ISRD 200 labs, PO Box 1500, Edinburgh SA 5111, Australie

czyz@tele.ucl.ac.be

Résumé − Nous nous intéressons au suivi d'objets dans une séquence vidéo. Le problème est formulé sous forme d'estimation bayesienne récursive et implémenté par un filtre particulaire. Les histogrammes des couleurs des objets que l'on souhaite suivre sont utilisés pour les caractériser. Nous utilisons un filtre particulaire hybride qui permet la détection et le suivi simultanés des objets. Le filtre incorpore une variable discrète, modélisée par une chaîne de Markov, qui représente le nombre d'objets présents dans la scène. L'approche permet de détecter et de suivre un ou plusieurs objets de couleur similaire même dans le cas où l'arrière plan n'est pas fixe. L'approche sur base des histogrammes de couleur permet aux objets de subir des déformations et des occultations partielles.

Abstract − We study object tracking in a video sequence. The problem is formulated as bayesian recursive estimation and is implemented as a particle filter. Color histograms are used as observation features. Joint detection and tracking is performed thanks to a hybrid particle filter in which a discrete variable, representing the number of objects present in the scene, is estimated. The approach allows to detect and track several objects sharing the same color distribution and undergoing deformations and partial occlusions.

1 Introduction

Parmi les opérations de haut-niveau qui aide à l'interprétation d'une scène filmée par une caméra, le suivi automatique (parfois appelé traçage) des objets d'intérêt présents dans la scène est incontournable. Les applications du suivi sont par exemple la visio-surveillance automatique, les interfaces gestuelles homme-machine, l'analyse automatique du mouvement, etc. Le suivi d'un objet (ou d'une personne) dans un flux vidéo consiste à déterminer la position de l'objet d'intérêt dans l'image courante à partir de sa position dans l'image précédente.

Classiquement [2] le suivi s'effectue en une étape de détection et d'association de données. L'étape de détection consiste à trouver dans l'image des caractéristiques qui pourraient correspondre aux objets que l'on souhaite suivre. Pour cela des caractéristiques stables de l'objet doivent être connues, puisque l'apparence d'un objet peut varier fortement avec l'éclairage de la scène, la position des objets et d'autre facteurs. Deuxièmement ces caractéristiques doivent être associées aux objets que l'on est en train de suivre (association de données). Shi et Tomasi [9] ont introduit une technique dans laquelle des points d'intérêt locaux à l'image sont détectés et suivis au cours du temps. Les points sont définis de façon à être très stables malgré les variations d'apparence et les déformations que peuvent subir les objets.

Cox et Hingorani ont [2] introduit le suivi multi-hypothèses. A chaque image de la séquence vidéo, on constitue une liste d'hypothèses (ainsi que la probabilité que l'hypothèse soit vérifiée) de correspondances entre les caractéristiques détectées et les objets suivi.

Comaniciu *et al.* [1] utilisent comme caractéristique de l'image les histogrammes de couleurs. Ceux-ci ont la particularité de ne pas dépendre de la configuration spatiale des objets et conviennent donc très bien pour les objets déformables.

Isard et Blake [4] ont utilisé le filtrage particulaire [3] pour résoudre le problème du suivi par contours. La caractéristique de l'objet que l'on détecte est donc le contour de l'objet. Le problème du suivi est formulé dans le cadre de l'estimation bayesienne récursive et il est résolu en utilisant un filtre particulaire. Le suivi par filtre particulaire a été étendu au cas ou les caractéristiques des objets sont des histogrammes de couleur [6, 7]. La méthode qui en résulte, appelée filtre particulaire couleur (FPC), permet de suivre de façon robuste des objets déformables, à la cinématique complexe, dans des séquence vidéo où l'arrière-plan est quelconque.

Cependant dans la méthode du filtre particulaire couleur telle que proposée dans [6, 7], il faut passer par une étape de détection externe qui initialise le suivi. De plus, le filtre ne peut gérer le suivi plusieurs objets ayant des histogrammes de couleurs similaires. Dans cette communication, nous proposons un filtre particulaire couleur qui effectue *simultanément* la détection et le suivi d'un nombre quelconque d'objets ayant des histogrammes de couleur similaires. Notre approche se base sur l'estimation bayesienne dans un *espace d'état hybride*, dans lesquels le vecteur d'état peut contenir à la fois des variables continues

et discrètes. L'addition au vecteur d'état traditionnel de [6] d'une variable discrète, qui représente le nombre d'objets d'intérêt dans la scène, permet la détection et le suivi conjoint des objets. L'approche de Isard et MacCormick utilise elle aussi un vecteur d'état hybride [5]. Cependant, les particularités de la trame utilisées pour caractériser les objets sont les réponses d'un banc de filtres appliquées à l'image. Cette caractérisation nécessite intrinsèquement l'estimation de l'arrière-plan, contrairement à la méthode présentée ici. De plus les histogrammes de couleurs permettent de représenter les objets de façon très succincte. Ceci permet d'utiliser des vecteurs d'état de dimension faible réduisant ainsi la dimension de l'espace d'état et donc la complexité de l'exploration.

La structure de l'article est la suivante : nous commençons par décrire les modèles dynamique et d'observation utilisés dans notre approche. Le filtre particulaire hybride est décrit dans la section suivante. Nous terminons par quelques résultats qualitatifs de suivi d'objets obtenus sur une séquence vidéo réelle.

2 Filtrage particulaire couleur

Soit le vecteur d'état $\mathbf{x}_t$ au temps t associé à un objet d'intérêt, $\mathbf{x}_t$ contient la position de l'objet ainsi que les paramètres de la région dans laquelle on calcule l'histogramme qui représente l'objet. Soit Z_t la séquence des trames disponibles $\{\mathbf{z}_1, ..., \mathbf{z}_t\}$. L'estimation bayesienne récursive consiste à estimer $p(\mathbf{x}_t|Z_t)$ à partir de $p(\mathbf{x}_{t-1}|Z_{t-1})$. L'idée centrale du filtrage particulaire est de représenter la densité de probabilité conditionnelle de $\mathbf{x}_t$ par un ensemble d'échantillons pondérés $\{\mathbf{x}_t^n, w_t^n\}$. Dans sa forme la plus simple, l'évolution des échantillons est décrite par le modèle dynamique du système. Après évolution, les échantillons sont pondérés en utilisant le modèle d'observation et $\mathbf{z}_t$.

2.1 Modèle dynamique

Comme énoncé précédemment, le vecteur d'état $\mathbf{x}_t$ contient les paramètres de région où est calculé l'histogramme. En particulier, les régions que nous utilisons sont des rectangles, on a donc $\mathbf{x} = (x\ y\ H\ W)^T$ où (x, y) sont les coordonnées du centre du rectangle dans l'image, et H et W sont respectivement la hauteur et la largeur du rectangle. Pour modéliser la cinématique des objets d'intérêt, nous utilisons le modèle dynamique de type Gauss-Markov. L'équation d'état est linéaire et s'écrit

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_{t-1}$$

où le bruit $\mathbf{w}_t$ est supposé gaussien de moyenne nulle et de matrice de covariance Q. D'autre types de modèles dynamiques sont envisageables et peuvent être mieux adaptés selon l'application.

2.2 Modèle d'observation

De même que [7, 6], l'histogramme q_t de couleurs extrait de la trame courante est utilisé comme observation. L'histogramme est calculé dans la région rectangulaire définie par le vecteur d'état. Le modèle d'observation s'écrit

$$p(q_t|\mathbf{x}_t) \propto \mathcal{N}(D_t; 0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{D_t^2}{2\sigma^2}\} \qquad (1)$$

où D_t est la distance entre l'histogramme de référence q^* qui caractérise l'objet d'intérêt et l'histogramme q_t calculé à partir de la trame courante. La grandeur σ est un paramètre à fixer par le concepteur du système. On voit donc que $p(q_t|\mathbf{x}_t)$ aura une valeur d'autant plus élevée si la distance en l'histogramme de référence et l'histogramme q_t est petite, c'est-à-dire si les régions de l'image sont semblables. Soit U le nombre de catégories dans les histogramme q^* et q_t, la distance D_t^2 entre ces histogrammes est définie dans [1] comme étant

$$D_k^2 = 1 - \sum_{u=1}^{U} \sqrt{q^*(u)\ q_k(u)}.$$

Notons que les histogrammes peuvent être calculés dans un espace de couleur tel que RGB, HSV ou autre.

2.3 Modélisation du nombre d'objets dans la scène

La modélisation du nombre d'objets d'intérêt dans la scène se fait en introduisant une variable aléatoire discrète $E \in \mathbb{E} = \{0, 1, \ldots, M\}$. E représente le nombre d'objets présents dans la scène. Nous supposons que E est une chaîne de Markov à M états dont les probabilités de transition sont définies par une matrice de transition $\mathbf{\Pi} = [\pi_{ij}]$, où

$$\pi_{ij} = Pr\{E_t = j|E_{t-1} = i\}, \qquad (i, j \in \mathbb{E}) \qquad (2)$$

est la probabilité de transition d'un nombre d'objets présents égale à i au temps $t-1$ à un nombre d'objets égale à j au temps t. A titre d'exemple, imaginons que nous souhaitons détecter et suivre un seul objet (c'est-à-dire $M = 1$), la matrice de transition est 2×2 et est donnée par

$$\mathbf{\Pi} = \begin{bmatrix} (1 - P_n) & P_n \\ P_m & (1 - P_m) \end{bmatrix}$$

où P_n et P_m représentent la probabilité de voir l'objet respectivement entrer et quitter la scène.

3 Filtre particulaire hybride

Chaque particule $\mathbf{y}_t^n$ du filtre particulaire hybride contient la variable aléatoire E_t^n ainsi que le nombre correspondant de vecteurs d'état $\mathbf{x}_{i,t}^n$ $(i = 1, \ldots, E_t^n)$ de chaque objet présent dans la scène, c'est-à-dire

$$\mathbf{y}_t^n = \begin{cases} E_t^n & & & \text{if } E_t^n = 0 \\ [(\mathbf{x}_{1,t}^n)^T & E_t^n]^T & & \text{si } E_k = 1 \\ [(\mathbf{x}_{1,t}^n)^T & (\mathbf{x}_{2,t}^n)^T & E_t^n]^T & \text{si } E_t^n = 2 \\ \vdots & & & \vdots \end{cases} \qquad (3)$$

où n varie de 1 à N, le nombre de particules.

Les étapes de l'algorithme sont résumées dans la table 1. Elles sont décrites plus en détail ci-dessous.

TAB. 1 – Filtre particulaire hybride (pseudo-code)

$[\{\mathbf{y}_k^n\}_{n=1}^N] = \text{PF}[\{\mathbf{y}_{k-1}^n\}_{n=1}^N, \mathbf{z}_k]$

1. Transitions de la variable E_{t-1} (transition aléatoire du nombre d'objets présents) : $[\{E_t^n\}_{n=1}^N] = \text{ETrans}\ [\{E_{t-1}^n\}_{n=1}^N, \mathbf{\Pi}]$
2. FOR $n = 1 : N$
   a. Sur base des paires (E_{t-1}^n, E_t^n) pair, tirer aléatoirement $\mathbf{x}_{1,t}^n, \ldots, \mathbf{x}_{E_t^n,t}^n$;
   b. Evaluer les poids $\tilde{w}_k^n$ (à une constante près) par l'équation (4).
3. END FOR
4. Normaliser les poids
   a. Calculer poids total : $t = \text{SOMME}\ [\{\tilde{w}_k^n\}_{n=1}^N]$
   b. FOR $n = 1 : N$
      – Normaliser : $w_k^n = t^{-1}\tilde{w}_k^n$
      END FOR
5. Re-échantillonage : $[\{\mathbf{y}_k^n, -\}_{n=1}^N] = \text{RESAMPLE}\ [\{\mathbf{y}_k^n, w_k^n\}_{n=1}^N]$

En premier lieu, on simule les transitions aléatoires de E_{t-1}^n à E_t^n grâce aux probabilités contenues dans la matrice de transitions $\mathbf{\Pi}$. Donc, en reprenant l'exemple $M = 1$ de la section 2.3, si E_{t-1}^n valait 0 au cycle précédent, E_t^n sera tiré aléatoirement et vaudra 1 avec une probabilité P_n.

Suivant le résultat de la transition de E_{t-1}^n, on distingue les trois cas suivant.

1. Si $E_{t-1}^n = E_t^n$, les $\mathbf{x}_{i,t}^n$ $(i = 1, \ldots, E_t^n)$ subissent l'évolution définie par le modèle dynamique.
2. Si $E_{t-1}^n < E_t^n$, les $\mathbf{x}_{i,t}^n$ correspondant aux objets qui existaient déjà à la trame précédente subissent l'évolution définie par le modèle dynamique, tandis que pour les "nouveaux" objets, les $\mathbf{x}_{i,t}^n$ $(i = E_{t-1}^n, \ldots, E_t^n)$ sont tirées d'une densité $p_b(\mathbf{x}_k)$ qui représente la connaissance *a priori* sur l'apparition de nouveaux objets. Cette densité doit être spécifiée par le concepteur du système. Lorsqu'aucune connaissance *a priori* n'est disponible, elle peut être modélisée par une densité uniforme sur l'entièreté de l'image.
3. Si $E_{t-1}^n > E_t^n$, on tire uniformément E_t^n vecteurs $\mathbf{x}_{i,t}^n$ qui correspondent aux objets qui continuent à être présents. Les autres objets disparaissent du vecteur d'état composite $\mathbf{y}_t^n$. Les objets qui continuent à être présent subissent l'évolution définie par le modèle dynamique.

Les particules $\mathbf{y}_t^n$ sont pondérées grâce au modèle d'observation ce qui implique les poids suivants

$$\tilde{w}_t^n = \begin{cases} 1, & \text{si } E_t^n = 0 \\ C_B^{E_t^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{E_t^n}\left(D_{i,t}^n\right)^2\right\}, & \text{si } E_t^n > 0 \end{cases} \quad (4)$$

où C_B est une constante fixée par l'utilisateur, qui tient compte de la similarité entre l'histogramme de référence et l'arrière plan de l'image. Aussi le poids d'une particule est fixé à 0 lorsque deux objets, définis par cette particules, sont trop proches. Ceci pour éviter l'apparition de plusieurs objets factices sur une même région de l'image.

La dernière étape consiste à ré-échantillonner l'ensemble des particules en utilisant un algorithme standard tel que présenté dans [8].

Lorsque les étapes précédentes sont accomplies, on peut estimer le nombre d'objets présents ainsi que leur vecteur d'état respectif. En effet, l'ensemble des particules forme une approximation de la densité $p(\mathbf{y}_t|Z_t)$. Et lorsque celle-ci est connue, la probabilité $P_m = Pr\{E_t = m|Z_t\}$ qu'il y ait m objets présents dans la scène à l'instant est simplement la probabilité marginale de $p(\mathbf{y}_t|Z_t)$, c'est-à-dire

$$P_m = \int \ldots \int p(\mathbf{x}_{1,t}, \ldots, \mathbf{x}_{m,t}, E_k = m|Z_k) d\mathbf{x}_{1,t} \ldots d\mathbf{x}_{m,t} \quad (5)$$

pour $m = 1, \ldots, M$. Une estimation du maximum *a posteriori* du nombre d'objets est donc donnée par

$$\hat{m}_t = \arg \max_{m=0,1,\ldots,M} P_m. \quad (6)$$

En transposant ceci pour les particules nous avons

$$Pr\{E_t = m|Z_t\} = \frac{1}{N}\sum_{n=1}^{N}\delta(E_t^n, m) \quad (7)$$

et $\delta(i,j) = 1$, si $i = j$, et zéro dans le cas contraire (delta de Kronecker). Les estimations des états des objets $i = 1, \ldots, \hat{m}$ donc

$$\hat{\mathbf{x}}_{i,t|t} = \frac{\sum_{n=1}^{N}\mathbf{x}_{i,t}^n\,\delta(E_t^n, i)}{\sum_{n=1}^{N}\delta(E_t^n, i)}. \quad (8)$$

4 Résultats expérimentaux

Plusieurs séquences vidéo traitées par l'algorithme présenté dans cette communication peuvent être télé-chargées sur le site internet :
http ://euterpe.tele.ucl.ac.be/Tracking/pf.html. Nous décrivons dans cette section les paramètres adoptés pour le traitement. Dans la matrice de transition $\mathbf{\Pi}$, seules les transitions d'un nombre d'objets m_t à une nombre d'objets $m_t \pm 1$ au temps $t+1$ sont autorisées, avec un probabilité 0,05. La matrice $\mathbf{\Pi}$ est donc tri-diagonale. La probabilité pour que le nombre d'objets reste inchangé est donc 0,9. Cette simplification de la matrice de transition a comme conséquence que si deux objets apparaissent en même dans la scène, l'estimation du nombre d'objets $\hat{m}_t$ sera incrémentée en deux étapes.

La distribution $p_b(\mathbf{x}_{i,k})$, introduite à la section 3, est une distribution uniforme sur toutes les variables du vecteur d'état $\mathbf{x}_{i,k}$. Les histogrammes de couleur sont calculés dans l'espace RGB avec un nombre de catégories de 8x8x8 comme dans [6]. L'histogramme de référence q^* est créé à partir de quelques trames d'initialisation. La région contenant l'objet à suivre est sélectionnée manuellement et un

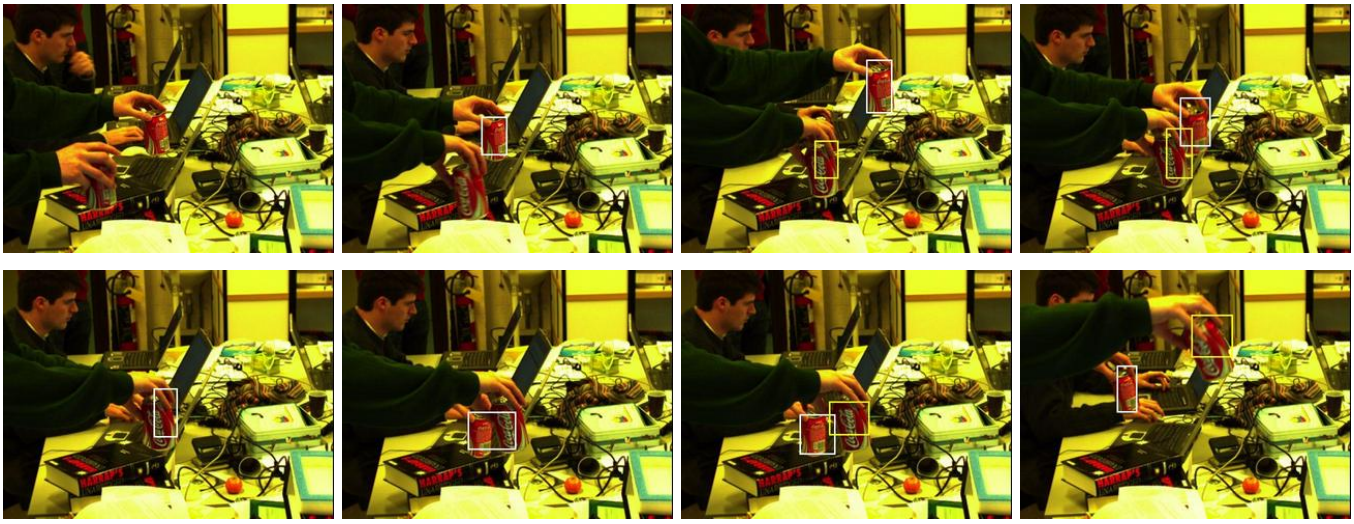

FIG. 1 – Le but est de détecter et de suivre les boîtes de soda. Les boîtes détectées sont encadrées par un rectangle. Deux boîtes sont présentes dès la première trame. La première boîte, puis la seconde, sont rapidement détectées et suivies. Lorsqu'une des deux boîtes occulte l'autre, l'algorithme ne détecte plus qu'une boîte. Dès que les deux boîtes sont à nouveau visibles, la deuxième boîte est détectée et suivie et ce, jusqu'à la fin de la séquence.

histogramme est extrait de cette région. L'histogramme de référence s'obtient en moyennant les histogrammes obtenus pour chaque trame.

Le nombre de particules nécessaires au bon fonctionnement de la détection et de l'estimation dépend de plusieurs facteurs. Principalement il s'agit du nombre maximal M d'objets et de la connaissance *a priori* sur l'apparition des objets (donc le choix de $p_b(\mathbf{x_{k,i}})$. Pour $M = 1$ un nombre de particules de 150 suffit pour obtenir un résultat satisfaisant tant en détection qu'en suivi. Pour $M = 6$ objets identiques pour utilisons 5000 particules.

Sur la figure 1, on peut voir quelques trames extraites d'une séquence test. Le but est de détecter et de suivre les boîtes de soda rouges. Les boîtes détectées sont encadrées par un rectangle. Le filtre utilise $N = 1000$ particules avec comme paramètre $\sigma = 0.6$ et $C_B = 70$. La taille de l'image est 640 x 480. Notons que l'arrière plan est pour le moins complexe. Deux boîtes sont présentes dès la première trame. La première boîte, puis la seconde, sont rapidement détectées et suivies. Lorsqu'une des deux boîtes occulte l'autre, l'algorithme ne détecte plus qu'une boîte. Dès que les deux boîtes sont à nouveau visibles, la deuxième boîte est détectée et suivie et ce, jusqu'à la fin de la séquence.

5 Conclusion

Nous avons présenté un algorithme de suivi d'objet grâce à leur distribution de couleurs. La particularité de notre approche est de considérer la détection est le suivi conjointement, en utilisant un filtre particulaire hybride. Le filtre incorpore une variable discrète, modélisée par une chaîne de Markov, qui représente le nombre d'objets présents dans la scène. L'approche permet de détecter et de suivre un ou plusieurs objets de couleur similaire même dans le cas où l'arrière plan n'est pas fixe.

Références

[1] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. IEEE Conf. Comp. Vision Pattern Recog.*, pages II :142–149. Hilton Head, SC, June 2000.

[2] I. J. Cox and S. L. Hingorani. An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. In *International Conf. on Pattern Recognition*, pages 437–443. 1994.

[3] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Springer, New York, 2001.

[4] M. Isard and A. Blake. Visual tracking by stochastic propagation of conditional density. In *Proc. European Conf. Computer Vision*, pages 343–356. 1996.

[5] M. Isard and J. MacCormick. BraMBLe : a bayesian multiple blob tracker. In *Proc. Int. Conf. Computer Vision*, pages 34–41. 2001.

[6] K. Nummiaro, E. Koller-Meier, and L. Van-Gool. An adaptive color-based particle filter. *Image and Vision Computing*, 21 :99–110, 2003.

[7] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In A. H. et al., editor, *Proc. European Conf. Computer Vision (ECCV)*, pages 661–675. Springer-Verlag, 2002. LNCS 2350.

[8] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter : Particle filters for tracking applications*. Artech House, 2004.

[9] J. Shi and C. Tomasi. Good features to track. In *IEEE Int. Conference on Computer Vision and Pattern Recognition*, pages 593–600. 1994.