Suivi d'individus dans des cartes de profondeurs par champ de Markov, dans un contexte de détection de chute

Geoffroy CORMIER[1, 2, 3, 4], Jean-Marc LAFERTÉ[1], Guy CARRAULT[2, 3], Jean-louis DILLENSEGER[2, 3], Vincent GAUTHIER[4]

[1]ECAM RENNES - Louis de Broglie, Campus de Ker Lann - Bruz, CS 29128 - 35091 Rennes Cedex 9, France

[2]INSERM, U1099, Rennes, F-35000, France

[3]Université de Rennes 1, LTSI, Campus de Beaulieu, Bât. 22, 35042 Cedex, Rennes, France

[4]Neotec-Vision, Bâtiment Club Services, 7 allée de la Planche Fagline, 35740 Pacé, France

geoffroy.cormier@ecam-rennes.com, guy.carrault@univ-rennes1.fr

Résumé – Cet article propose une application des champs de Markov spatio-temporels (ST-MRFs) à l'analyse de cartes de profondeurs pour séparer et suivre des objets détectés dans une scène acquise par un capteur de profondeurs. La méthode comprend trois étapes, qui traitent la carte de profondeurs en amont du ST-MRF : la segmentation, l'estimation de mouvement, et la création et mise à jour de cartes d'objets. La méthode a été testée sur des séquences synthétiques et réelles : les résultats montrent la capacité de la méthode à séparer les objets se croisant dans l'image.

Abstract – This paper proposes a method for the detection, separation, and tracking of objects in depth maps. The method is based on an application of spatio-temporal Markov Random Fields (ST-MRF) which uses the results of 3 preprocessing steps of the depth map: a segmentation step, a motion estimation step, and finally, an objects map updating step. The method has been tested on synthetic and real depth map sequences, and yielded results which show its ability to discriminate objects with crossing image trajectories.

1 Introduction

Alors que l'on estime qu'en France, 30% de la population aura plus de 65 ans d'ici 2050 [3], il est nécessaire de se doter d'outils qui permettent d'améliorer l'autonomie des personnes âgées. Les chutes représentent la première cause de mortalité de cette tranche d'âge, et génèrent en France des coûts de santé de l'ordre de deux milliards d'euros par an. Ces constats imposent de déployer des systèmes de suivi et de détection d'événements robustes tenant le temps réel. Pour respecter l'anonymat des personnes observées, et nous affranchir des conditions d'éclairage et de texture des scènes acquises, notre système de suivi se base sur l'analyse de cartes de profondeurs obtenues par des capteurs travaillant dans le proche infrarouge. Le suivi d'objets dans une séquence vidéo peut être basé sur le filtrage de Kalman [9] ; sur des algorithmes à base de filtres particulaires, tels que CONDENSATION [4] ; ou encore sur des algorithmes itératifs à base de champs aléatoires ou pseudo-aléatoires, tels que les champs de Markov spatio-temporels (ST-MRFs) [6]. Cependant, que le suivi d'individu repose sur une squelettisation d'une silhouette détectée [8] ou sur de l'analyse de formes, il nécessite au préalable une bonne segmentation. Par ailleurs, les méthodes de squelettisation présentent des défauts rédhibitoires dans notre contexte, parce qu'elles ont été prévues pour le jeu, où les individus sont clairement visibles, et debout face au capteur. Une segmentation en composantes connexes seule ne suffit pas pour suivre les objets détectés dans la carte de profondeurs : si deux objets détectés entrent en contact physique ou contact image (occultation de l'un par l'autre), la segmentation aura tendance à les fusionner. Cette communication vise à proposer une solution pour pallier à cette difficulté. Les ST-MRFs ont de nombreuses applications en traitement d'images et de vidéos, que ce soit pour le suivi d'objets ou de postures dans des séquences couleur [2], la restauration de couples d'images couleur et de profondeurs [5], ou encore pour répondre aux problèmes de segmentation d'images et de superrésolution de cartes de profondeurs [7]. Dans un contexte de suivi d'objets, il n'existe pas à notre connaissance d'application des ST-MRFs à la seule carte de profondeurs. Nous proposons une application des ST-MRFs à l'analyse de cartes de profondeurs, pour opérer la séparation et le suivi des objets détectés dans la scène. Dans un premier temps, nous présentons les différentes étapes de la méthode. Dans un second temps, nous présentons nos premiers résultats.

2 Méthode

L'algorithme de suivi adopté se déroule comme suit :

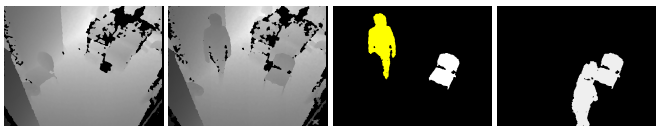
- La carte de profondeurs D_t acquise à l'instant t est comparée à une référence calculée au préalable, puis segmentée et étiquetée en composantes connexes pour donner

une carte segmentée S_t,

- La carte d'objets O_t calculée à l'instant $t-1$ est mise à jour à partir de S_t,
- Le déplacement image des objets entre les instants $t-1$ et t est estimé à partir de O_t, de D_{t-1} et de D_t, et enregistré dans une carte de vecteurs de mouvements M_t, de même taille que O_t,
- O_t est optimisée par le ST-MRF à partir de D_{t-1} et de D_t, de O_{t-1}, de S_t, et de M_t.

2.1 Segmentation de la carte de profondeurs

La carte de profondeurs D_t (fig. 1b) est comparée à une référence (fig. 1a) calculée au préalable pour une scène sans individu, afin de détecter les objets apportés ou déplacés dans la scène. Les régions détectées dans D_t par rapport à la référence sont filtrées afin de ne conserver que des régions d'intérêt et d'éliminer le bruit de détection dû au capteur. La carte de détection ainsi obtenue est segmentée en composantes connexes (8-connexité) pour obtenir une carte segmentée S_t (fig. 1c).



(a) (b) (c) (d)

Figure 1 – Segmentation dans la carte de profondeurs.
a : Référence. b : D_t. c : S_t. d : Occultation d'objets et fusion de régions dans S_t.

2.2 Carte d'objets

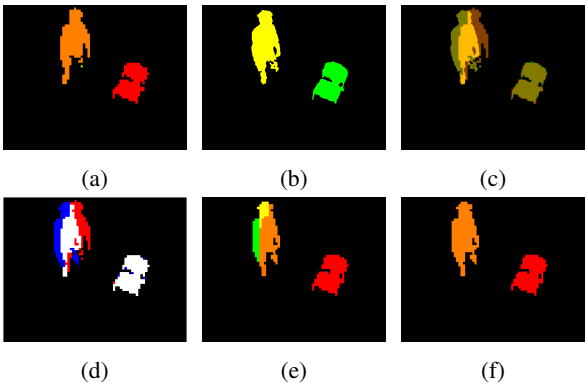
La carte d'objets O_t mémorise les étiquettes des objets détectés, même en cas de contact (au sens image) de deux régions segmentées. Chaque bloc de la carte d'objets correspond à un bloc de n par n pixels de la carte de segmentation S_t. O_t est créée à partir de S_t en deux étapes : une étape de mise à jour et une étape de correction. Le passage en cartes de blocs est motivé par le choix de la méthode d'estimation de mouvement mise en œuvre (block-matching). Ce choix permet aussi de réduire le coût calculatoire de l'algorithme.

2.2.1 Mise à jour de la carte d'objets

L'étape de mise à jour de la carte d'objets O_t consiste à y faire apparaître les blocs détectés dans la carte de segmentation S_t, et à en effacer les blocs non détectés, selon l'algorithme 1 .
$B(S)$ représente le bloc de pixels de S_t qui correspond au bloc B, $\mathcal{V}(B)$ désigne le voisinage en blocs du bloc B (voisinage 3×3), et $L(B)$ représente l'étiquette du bloc B. Le voisin B' choisi dans $\mathcal{V}(B)$ est le premier voisin non nul, dans l'ordre lexicographique. $\#\bullet$ est l'opérateur cardinal.

Algorithme 1 Mise à jour de la carte d'objets

Pour chaque bloc B de O_t, faire
  Si $\left(\#\left\{pixel \in B(S) ; pixel == 0\right\} > \frac{n^2}{2}\right)$ alors
    $B \leftarrow 0$
  Sinon
    Si $(L(B) == 0)$ alors
      Si $\left(\exists B' \in \mathcal{V}(B) ; L\left(B'\right) \neq 0\right)$ alors
        $L(B) \leftarrow L\left(B'\right)$
      Sinon
        $L(B) \leftarrow min\left(i \in \mathbb{N}^* ; i \notin O\right)$ // nouvelle étiquette
      Fin Si
    Fin Si
  Fin Si
Fin Pour



(a) (b) (c)

(d) (e) (f)

Figure 2 – Mise à jour de la carte d'objets O_t.
a : O_{t-1}. b : S_t. c : Superposition de la S_t et de O_{t-1}. d : Mise en évidence des blocs à retirer ou à ajouter dans O_t. e : O_t mise à jour. f : O_t corrigée.

La figure 2 illustre le processus d'effacement et d'ajout des blocs dans la carte d'objets : si comme le montre la figure 2c l'on superpose la S_t (fig. 2b) à O_{t-1} (fig. 2a), l'on met en évidence les zones devant être effacées de O_t, les zones à conserver, et les zones à ajouter à O_t (respectivement, en rouge, blanc et bleu dans la figure 2d). La carte d'objets présentée en figure 2e résulte de ce processus d'effacement/ajout de blocs dans la carte d'objets précédente. O_t étant parcourue dans l'ordre lexicographique, cette étape de mise à jour fait apparaître des zones connexes aux objets déjà présents dans la scène mais étiquetées différemment (cf. fig. 2e) : il est nécessaire d'opérer une correction de la carte obtenue, pour ne conserver que des régions qui correspondent à une réalité tant image que physique.

2.2.2 Correction de la carte d'objets

L'étape de correction de la carte d'objets O_t consiste à rattacher les blocs étiquetés avec des identifiants non présents dans O_{t-1} aux régions étiquetées avec d'anciens identifiants. Il s'agit d'un étiquetage en composantes connexes (8-connexité) qui interdit de modifier un bloc étiqueté avec un « ancien » identifiant. À la fin de la correction, la O_t comprend d'anciens objets mis à jours, et éventuellement de nouveaux objets, qui correspondent aux blocs qui n'ont pas été rattachés à un ancien objet, et qui gardent un nouvel identifiant. La figure 2f présente un résultat

du ré-étiquetage en composantes connexes de O_t (à comparer à la figure 2e).

Malgré l'étape de correction, le calcul de la carte d'objets seul ne suffit pas à assurer la séparation des régions détectées, lorsqu'elles entrent en contact dans l'image de segmentation (cf. fig. 1d). Il reste nécessaire d'optimiser la carte d'objets, ce que nous faisons au moyen d'un ST-MRF (cf. 2.4).

2.3 Estimation de mouvement

L'estimation du mouvement apparent des objets détectés est nécessaire pour le ST-MRF : elle porte l'information temporelle utilisée pour le suivi des objets.
Le déplacement en pixels des blocs non nuls de O_t est estimé au moyen d'un algorithme de block matching, le Diamond Search (DS) [10]. Pour un bloc non nul de O_t, le bloc de profondeurs associé dans D_t est déplacé dans D_{t-1}, dans la direction qui minimise une métrique de distorsion donnée (la métrique adoptée est la distance 3D moyenne entre les deux blocs, calculée sur les pixels non nuls des deux cartes de profondeurs). Le vecteur de mouvement retenu pour un bloc est la somme des vecteurs de mouvements déterminés à chaque itération du Diamond Search. La carte des mouvements M_t obtenue (fig. 3b) présente des vecteurs pointant vers le passé, dans notre implémentation. Pour des raisons de lisibilité, les vecteurs de mouvements de la figure 3b ont été amplifiés d'un facteur 10.

(a) Encodage chromatique des vecteurs de mouvement. (b) Vecteurs de mouvements estimés (amplifiés).

Figure 3 – Estimation de mouvement.

2.4 Champ de Markov

Nous utilisons un ST-MRF pour optimiser la distribution des blocs objets issus de l'étape décrite en 2.2. Pour cela, nous définissons les trois fonctions d'énergies suivantes, calculées pour chaque bloc, et pour chaque identifiant non nul présent dans O_t : une fonction d'énergie spatiale

$$E_S(k) = 1 - \frac{\#\{B' \in \mathcal{V}; B' = k\}}{8}, \tag{1}$$

une fonction d'énergie potentielle

$$E_P(k) = \left| -\frac{1}{2}\left(\frac{\overline{d}(B) - \mu_k(d)}{\sigma_k(d)}\right)^2 - \ln\left(\sigma_k(d)\sqrt{2\pi}\right)\right|, \tag{2}$$

et une fonction d'énergie cinétique

$$E_C(k) = \left| -\frac{1}{2}\left(\frac{\|\overrightarrow{m}\|(B) - \mu_k(\|\overrightarrow{m}\|)}{\sigma_k(\|\overrightarrow{m}\|)}\right)^2 - \ln\left(\sigma_k(\|\overrightarrow{m}\|)\sqrt{2\pi}\right)\right|. \tag{3}$$

L'équation (1) est liée à la probabilité qu'a le bloc B d'être étiqueté k dans O_t, sachant le nombre de blocs du voisinage $\mathcal{V}$ du bloc B appartenant à l'objet A_k, soit :

$$E_S(k) \leftrightarrow p(B = k \,|\, \mathcal{V}). \tag{4}$$

L'équation (2) est liée à la probabilité qu'a le bloc B d'être étiqueté k dans O_t, sachant sa profondeur moyenne $\overline{d}(B)$ et connaissant la distribution de la profondeur (moyenne, écart-type, modèle Gaussien) de l'objet A_k, calculée à partir de D_{t-1}, soit :

$$E_P(k) \leftrightarrow p\left(B = k \,|\, \overline{d}(B), A_k \sim \mathcal{N}(\mu_k(d), \sigma_k(d))\right). \tag{5}$$

L'équation (3) est liée à la probabilité qu'a le bloc B d'être étiqueté k dans O_t, sachant la norme de son vecteur vitesse, et connaissant la distribution des normes des vecteurs vitesses (moyenne, écart-type, modèle Gaussien) de l'objet A_k, calculée à partir de M_t, soit :

$$E_C(k) \leftrightarrow p\left(B = k \,|\, \|\overrightarrow{m}\|(B), A_k \sim \mathcal{N}(\mu_k(\|\overrightarrow{m}\|), \sigma_k(\|\overrightarrow{m}\|))\right). \tag{6}$$

La probabilité qu'a un bloc B de O_t de prendre l'étiquette k est donc

$$p(B = k) = \frac{1}{Z_1}e^{-E_S(k)} \cdot \frac{1}{Z_2}e^{-E_P(k)} \cdot \frac{1}{Z_3}e^{-E_C(k)}, \tag{7}$$

avec Z_1, Z_2 et Z_3 les constantes de normalisation. Le problème revient donc à trouver, pour chaque bloc B l'étiquette qui minimise une combinaison linéaire des trois énergies :

$$\begin{aligned} k &= \underset{k \in \{Etiquettes\}}{\operatorname{argmax}} \; p(B = k) \\ &= \underset{k \in \{Etiquettes\}}{\operatorname{argmin}} \; C_1 E_S(k) + C_2 E_P(k) + C_3 E_C(k), \end{aligned} \tag{8}$$

avec C_1, C_2 et C_3 les poids qui permettent de régler l'importance relative de chaque composante énergétique. L'optimisation est réitérée (selon le principe décrit dans [1]) jusqu'à ce que la carte se soit stabilisée.

3 Résultats

La méthode décrite dans cet article a été testée sur une séquence synthétique de 560 trames, où deux ellipsoïdes se croisent, ainsi que sur deux séquences réelles de 152 et 557 trames, où un individu passe devant une chaise détectée (cf. fig. 4). Les séquences réelles ont été acquises par un capteur Asus Xtion Pro, à 30 Hertz. Les cartes de profondeurs avaient une résolution de $X_R = 640$ pixels par $Y_R = 480$ pixels. Les blocs des cartes d'objets et de mouvements correspondaient à des blocs carrés de côté $n = 8$ pixels. Les poids C_1, C_2, et C_3 ont été fixés de manière *ad hoc* à $C_1 = 0.6$, $C_2 = 0.1$, et $C_3 = 0.3$.
Les performances de la méthode ont été évaluées en comparant à chaque trame le résultat de l'optimisation de la carte d'objets par le ST-MRF avec une vérité terrain établie par étiquetage

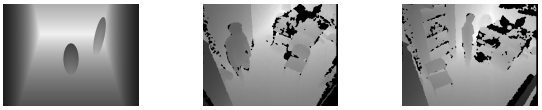


Figure 4 – Séquences testées : synthétique (gauche) et réelles (milieu : 152 trames, droite : 557 trames).

Table 1 – Erreurs moyennes pour la méthode.

	Trames avant croisement		Trames après croisement	
Séquence testée	Sans MRF	Avec MRF	Sans MRF	Avec MRF
Synthétique (560 trames)	0.00 %	1.28 %	43.56 %	1.03 %
Réelle (152 trames)	0.41 %	2.75 %	46.20 %	7.02 %
Réelle (557 trames)	0.00 %	6.24 %	51.06 %	0.34 %

manuel de la carte d'objets. L'équation (9) définit l'erreur ε commise par la méthode par rapport à la vérité terrain GT, pour la trame de l'instant t, calculée sur les blocs détectés.

$$\varepsilon_t = \frac{\sum\limits_{i,\,j} \left((1 - \delta_{0,\,m}) * \delta_{m,\,n} \right)}{\sum\limits_{i,\,j} (1 - \delta_{0,\,m})}. \tag{9}$$

Dans l'équation (9), $m = GT_t\,(i,\,j)$, $n = O_t\,(i,\,j)$ et δ est le symbole de Kronecker, dont l'équation (10) rappelle la définition.

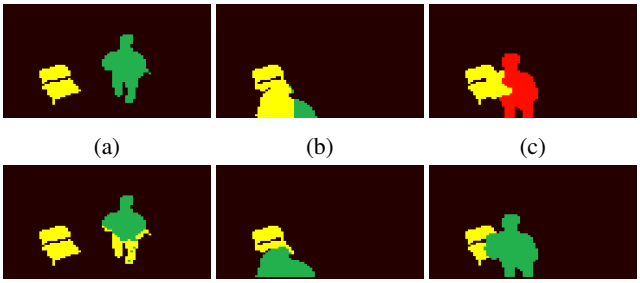
$$\forall\,(x,\,y) \in \mathbb{R}^2,\ \delta_{x,\,y} = \begin{cases} 1 & \text{si } x = y \\ 0 & \text{sinon} \end{cases}, \tag{10}$$

Autrement dit, ε_t représente le pourcentage de blocs non nuls mal étiquetés dans O_t par rapport à GT_t.
Pour chaque séquence testée, la table 1 donne l'erreur moyenne $\bar{\varepsilon}$ de la méthode pour les trames précédant le premier croisement, puis pour les trames suivant le premier croisement (croisement compris).

La table 1 montre le gain (colonnes de droite) apporté par la méthode dans la séparation des objets après croisement. En moyenne le taux d'erreur est diminué d'environ 44 % en comparaison avec le cas sans MRF. Cependant, on remarque que c'est au prix d'une erreur résiduelle avant croisement, mais ceci pourra être améliorer en optimisant les poids des fonctions d'énergie. Les figures 5e et 5f, en comparaison avec 5b et 5c illustrent la capacité du ST-MRF à conserver les étiquettes des objets qui entrent en contact dans l'image, là où l'absence de l'outil conduit à la fusion progressive des objets en contact, d'où les erreurs importantes après croisement dans la table 1.

4 Conclusion et perspectives

Dans cet article, une méthode à base de champ de Markov pour le suivi et la séparation d'objets détectés dans une carte de profondeurs a été proposée. Les premiers résultats montrent la robustesse de l'approche lorsque deux objets se croisent. Ce travail doit être considéré comme préliminaire, et les développements actuels portent sur l'utilisation de poids adaptatifs et l'ajout de connaissance *a priori* qui empêcherait une même région d'être indûment scindée dans la carte d'objets. Nous envisageons également de comparer notre méthode à différents



(a) (b) (c)

(d) (e) (f)

Figure 5 – Apports du champ de Markov.
(a, b, c) : Sans MRF. (d, e, f) : Avec MRF.

algorithmes de suivi de la littérature, tels que le filtrage de Kalman étendu ou CONDENSATION. Les perspectives à plus long termes étant le déploiement de cette méthode dans un système de suivi de personnes âgées à domicile ou en établissement, les contraintes liées au temps réel devront également être examinées.

Remerciements : Ces travaux ont été réalisés dans le cadre de la convention CIFRE 2012/0668.

Références

[1] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 259–302, 1986.

[2] N.-G. Cho, A. Yuille, and S.-W. Lee, "Nonflat observation model and adaptive depth order estimation for 3d human pose tracking," in *Pattern Recognition (ACPR), 2011 First Asian Conference on*. IEEE, Novembre 2011, pp. 382–386.

[3] INSEE, *Tableaux de l'économie française*. Institut national de la statistique et des études économiques , 18, boulevard Adolphe Pinard, 75675 PARIS CEDEX 14 ,www.insee.fr, 2015, ch. Population par âge, pp. 26–27.

[4] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *International journal of computer vision*, vol. 29, no. 1, pp. 5–28, 1998.

[5] S. Jonna, V. S. Voleti, R. R. Sahay, and M. S. Kankanhalli, "A multimodal approach for image de-fencing and depth inpainting," in *Advances in Pattern Recognition (ICAPR), 2015 Eighth International Conference on*, Janvier 2015, pp. 1–6.

[6] S. Kamijo, "Spatio-Temporal MRF model and its Application to Traffic Flow Analyses," in *Data Engineering Workshops, 2005. 21st International Conference on*, Avril 2005, pp. 1203–1203.

[7] M.-K. Kang, D.-Y. Kim, and K.-J. Yoon, "Adaptive support of spatial-temporal neighbors for depth map sequence up-sampling," *Signal Processing Letters, IEEE*, vol. 21, no. 2, pp. 150–154, Février 2014.

[8] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[9] K. Van Beeck, T. Goedemé, and T. Tuytelaars, "Towards an automatic blind spot camera: robust real-time pedestrian tracking from a moving camera," in *Proceedings of the twelfth IAPR Conference on Machine Vision Applications*, Juin 2011, pp. 528–531.

[10] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *Image Processing, IEEE Transactions on*, vol. 9, no. 2, pp. 287–290, Février 2000.