

Filtrage particulière avec une loi de proposition quasi-optimale utilisant la détection souple pour le suivi visuel

Mehdi OULAD AMEZIANE^{1,2}, Christelle GARNIER¹, Yves DELIGNON¹, Emmanuel DUFLOS², François SEPTIER¹

¹Institut Mines-Telecom / Telecom Lille, CRISAL, UMR CNRS 9189, Villeneuve d'Ascq, 59650, France

²Ecole Centrale de Lille, CRISAL, UMR CNRS 9189, Villeneuve d'Ascq, 59650, France

mehdi.ouladameziane@telecom-lille.fr, christelle.garnier@telecom-lille.fr
yves.delignon@telecom-lille.fr, Emmanuel.Duflos@ec-lille.fr,
francois.septier@telecom-lille.fr

Résumé – Les filtres particuliers sont aujourd’hui largement utilisés pour le suivi visuel. Pour améliorer l’exploration de l’espace d’état, nous proposons d’enrichir le modèle d’observation avec une information de détection souple, plus riche et fiable que le résultat de détection binaire usuel, et d’approcher la loi d’importance optimale à partir d’une approximation de la vraisemblance définie à partir de cette information. Le filtre résultant permet un suivi d’une grande précision et d’une grande robustesse face aux mouvements abrupts.

Abstract – Particle filters are currently widely used for visual tracking. In order to improve state space exploration, we propose to enrich the observation model with soft detection information, more reliable than the usual binary detection result, and to derive a near-optimal importance function from an approximation of the likelihood defined from this information. The resulting PF achieves high tracking accuracy and has the advantage of coping with abrupt movements.

1 Introduction

Le suivi visuel constitue une tâche essentielle en vision par ordinateur. Parmi les méthodes de suivi [1], les filtres particuliers (FPs) sont aujourd’hui largement utilisés. Leur performance dépend fortement de la loi d’importance qui sert à explorer l’espace d’état. Le choix le plus courant est la loi a priori liée au modèle dynamique. Mais il est difficile de prédire le déplacement d’un objet en cas de mouvements abrupts liés à un changement de caméra, de dynamique ou un flux vidéo bas débit. D’après [2], la loi d’importance optimale prend aussi en compte l’observation courante. L’inconvénient est qu’il est rarement possible de la calculer parce que l’expression analytique est indisponible ou les calculs sont trop coûteux. L’enjeu en suivi visuel est de sélectionner les observations les plus pertinentes et de les utiliser au mieux pour propager les particules dans l’espace d’état.

Les images brutes contiennent une très grande quantité d’informations. Un pré-traitement est nécessaire pour en extraire des informations plus succinctes, telles que des histogrammes couleur, des mesures de mouvement, des patches, des résultats de détection... Pour exploiter ces informations, deux types d’approches ont été développées. Les approches implicites utilisent la loi a priori et intègrent une étape pour guider les particules à partir des observations : étape de préselection des particules avec le FP auxiliaire [3] ou bien étape d’optimisation par "mean shift" [4] ou méthodes heuristiques [5]. Les approches explicites sont plus directes et cherchent à construire une loi d’im-

portance à partir des observations. Cette loi est généralement modélisée par un mélange de Gaussiennes (MdG) composé de la loi a priori et de lois centrées sur des points particuliers, comme les centres de silhouettes obtenues par détection [6]. Plusieurs détecteurs peuvent être combinés pour pallier leur manque de fiabilité [7].

Dans ce papier, nous proposons d’enrichir le modèle d’observation avec une information de détection souple et d’approcher la loi de proposition optimale [2] pour améliorer l’exploration de l’espace d’état en suivi visuel par FP. Cette information intermédiaire, obtenue dans des détecteurs avant décision dure (classification), est plus fiable que la sortie usuelle binaire, qui peut fournir des détections fausses ou manquées. Cette information a déjà été exploitée dans les FPs pour calculer les poids [8], mais à notre connaissance jamais pour propager les particules. Contrairement aux méthodes explicites qui utilisent une modélisation par MdG et aux méthodes implicites qui peuvent sortir du cadre théorique des FPs, notre méthode repose sur une approximation précise de la loi d’importance optimale. Le filtre résultant permet un suivi d’une grande précision et d’une grande robustesse face aux mouvements abrupts.

2 Suivi visuel par FP

Soit une séquence d’observations $y_{1:k} = (y_1, \dots, y_k)$. L’état x_k de l’objet est estimé à l’aide de la loi a posteriori $p(x_k|y_{1:k})$ approchée récursivement par un ensemble de N_p particules pon-

dérées $\{x_k^{(i)}, w_k^{(i)}\}_{i=1}^{N_p}$. Pour obtenir les particules à l'instant k , les échantillons $x_k^{(i)}$ sont tirés suivant une loi d'importance $q(x_k^{(i)} | x_{k-1}^{(i)}, y_k)$, puis les poids $w_k^{(i)}$ sont mis à jour selon l'expression réursive :

$$w_k^{(i)} \propto w_{k-1}^{(i)} \cdot \frac{p(y_k | x_k^{(i)}) \cdot p(x_k^{(i)} | x_{k-1}^{(i)})}{q(x_k^{(i)} | x_{k-1}^{(i)}, y_k)} \quad (1)$$

où la loi a priori $p(x_k | x_{k-1})$ représente l'évolution de x_k et la vraisemblance $p(y_k | x_k)$ mesure l'adéquation de y_k à x_k .

2.1 Modèle dynamique et loi a priori

L'objet est représenté par une fenêtre englobante. Le vecteur d'état s'écrit $x_k = \{c_k, s_k\}$ avec $c_k = \{c_k^x, c_k^y\}$ la position et $s_k = \{s_k^x, s_k^y\}$ la taille de la fenêtre. Pour pouvoir traiter tout type de mouvement, on choisit un modèle peu informatif. On suppose que les éléments de x_k évoluent indépendamment suivant une marche aléatoire Gaussienne : $x_k | x_{k-1} \sim \mathcal{N}(x_{k-1}, \Sigma)$ où $\Sigma = \text{diag}(\sigma_c^2, \sigma_c^2, \sigma_s^2, \sigma_s^2)$ est la matrice de covariance qui définit l'espace de recherche autour de l'état précédent. Dans les scénarios réels, l'objet peut effectuer des changements de position de grande amplitude, alors que sa taille évolue lentement. En conséquence la variance de la position σ_c^2 est beaucoup plus grande que la variance de la taille σ_s^2 .

2.2 Modèle d'observation et vraisemblance

Le modèle d'observation intègre l'information couleur usuelle et est enrichi avec l'information de détection souple extraite de chaque image I_k .

L'information couleur est exprimée sous forme d'un ensemble d'histogrammes RVB : $y_k^H = \text{hist}(I_k \cdot \mathbb{1}_{R(x_k)})$ avec $R(x_k)$ la région définie par x_k . Comme dans [9], la région $R(x_k)$ est divisée en bandes pour prendre en compte la distribution spatiale de la couleur.

L'information de détection souple est fournie par un détecteur de mouvement [10] qui a l'avantage de pouvoir détecter tout type d'objet. Il repose sur une modélisation des zones homogènes du fond et du premier plan par un MdG adaptatif dans un espace spatio-colorimétrique. La classification des pixels est effectuée par maximum de vraisemblance et fournit une carte de détection dure binaire. Ici, nous exploitons une information plus riche disponible avant la classification. Il s'agit d'une carte de probabilités : $y_k^D = [P_{i,j}]$ où $P_{i,j}$ est la probabilité que le pixel situé à la position (i, j) appartienne au premier plan. La figure 1 illustre les détections souple et dure.

Ces informations sont fusionnées classiquement en supposant leur indépendance conditionnellement à l'état [11], d'où la vraisemblance : $p(y_k | x_k) = p(y_k^H | x_k) \cdot p(y_k^D | x_k) = L_H \cdot L_D$.

L_H est la vraisemblance couleur définie à partir de la distance de Bhattacharyya D_B [11] entre H_k^{ref} les histogrammes de référence à N_b bins et y_k^H ceux de la région candidate $R(x_k)$

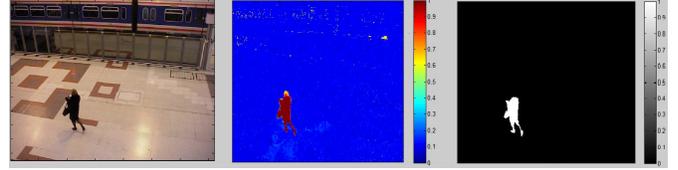


FIGURE 1 – Cartes de détection (gauche à droite : image originale, détection souple, détection dure)

pour les 3 canaux RVB et les S bandes :

$$L_H \propto \exp \left(-\lambda \sum_{p=1}^3 \sum_{r=1}^S D_B^2 \left(y_k^H(p, r), H_k^{ref}(p, r) \right) \right) \quad (2)$$

Et L_D est la vraisemblance de détection définie à partir de la carte de probabilités $y_k^D = [P_{i,j}]$ comme :

$$L_D \propto \exp \left(\lambda_1 \cdot \sum_{(i,j) \in R(x_k)} P_{i,j} - \lambda_2 \cdot N(s_k) \right) \quad (3)$$

avec $N(s_k)$ le nombre de pixels de la région $R(x_k)$ de taille s_k . Cette formulation évite de privilégier systématiquement les régions les plus grandes. λ , λ_1 et λ_2 sont des paramètres fixés de façon expérimentale.

2.3 Loi d'importance optimale

La loi d'importance optimale [2], au sens de la minimisation de la variance des poids, prend en considération les observations et s'écrit :

$$p(x_k | x_{k-1}, y_k) = \frac{p(y_k | x_k) \cdot p(x_k | x_{k-1})}{\int p(y_k | x_k) \cdot p(x_k | x_{k-1}) \cdot dx_k} \quad (4)$$

D'après les sections précédentes, la loi a priori est simple, mais la vraisemblance est complexe. Il est donc impossible d'obtenir l'expression analytique de la loi optimale. Par ailleurs, une évaluation point par point pour toutes les valeurs pouvant être prises par x_k est trop coûteuse en calculs.

3 FP avec loi d'importance quasi-optimale

Pour explorer efficacement l'espace d'état, l'objectif est de trouver un compromis entre la complexité de calculs et l'optimalité de la loi d'importance. Notre approche repose sur une approximation de la loi optimale à partir d'une approximation de la vraisemblance, qui est la loi la plus coûteuse en calculs dans l'expression (4).

3.1 Approximation de la loi optimale

D'après les hypothèses de la section 2.1, la position c_k et la taille s_k évoluent de façon indépendante et même en cas de variations abruptes de la position, la taille varie lentement. La vraisemblance varie donc beaucoup plus en fonction de c_k que de s_k . Pour réduire de façon significative le coût de calcul de

la vraisemblance, celle-ci est évaluée pour une unique valeur de s_k : $\tilde{s}_k = E[s_k | \hat{s}_{k-1}]$ avec \hat{s}_{k-1} la taille estimée à l'instant $k-1$.

Dans le même but, comme la vraisemblance de détection souple est plus simple à calculer que L_H , seule $L_D = p(y_k^D | c_k, s_k)$ est considérée dans l'expression de la loi d'importance optimale. En effet, d'après les expressions (2) et (3), L_D requiert uniquement $N(s_k)$ additions, tandis que L_H nécessite le calcul de $3S$ histogrammes et des distances de Bhattacharyya ce qui revient à $3(N(s_k) + 3N_b S)$ opérations. Néanmoins L_H intervient toujours dans la mise à jour des poids.

La loi d'importance optimale est donc approchée par :

$$\hat{p}(x_k | x_{k-1}, y_k) = \frac{p(y_k^D | c_k, \tilde{s}_k) \cdot p(c_k | c_{k-1}) \cdot p(s_k | s_{k-1})}{\hat{p}(y_k^D | c_{k-1}, s_{k-1})}$$

avec :

$$\hat{p}(y_k^D | c_{k-1}, s_{k-1}) = \int \int p(y_k^D | c_k, \tilde{s}_k) \cdot p(c_k | c_{k-1}) \cdot p(s_k | s_{k-1}) \cdot dc_k \cdot ds_k = \int p(y_k^D | c_k, \tilde{s}_k) \cdot p(c_k | c_{k-1}) \cdot dc_k$$

Finalement, la loi d'importance quasi-optimale s'écrit comme le produit de deux lois, une loi quasi-optimale pour c_k et la loi a priori pour s_k :

$$\hat{p}(x_k | x_{k-1}, y_k) = p(c_k | c_{k-1}, \tilde{s}_k, y_k^D) \cdot p(s_k | s_{k-1}) \quad (5)$$

où :

$$p(c_k | c_{k-1}, \tilde{s}_k, y_k^D) = \frac{p(y_k^D | c_k, \tilde{s}_k) \cdot p(c_k | c_{k-1})}{p(y_k^D | c_{k-1}, \tilde{s}_k)} \quad (6)$$

Pour pouvoir échantillonner c_k , $p(c_k | c_{k-1}, \tilde{s}_k, y_k^D)$ est considérée comme une distribution discrète avec un support fini, d'où $p(y_k^D | c_{k-1}, \tilde{s}_k) = \sum_{c_k} p(y_k^D | c_k, \tilde{s}_k) \cdot p(c_k | c_{k-1})$.

3.2 Algorithme de suivi

L'algorithme de suivi repose sur une implémentation du FP avec la loi de proposition quasi-optimale. D'après (5) l'échantillonnage se fait en deux temps : la position est tirée suivant la loi discrète (6) et la taille est tirée suivant la loi a priori. En introduisant les expressions de ces lois dans (1), on obtient l'expression suivante pour la mise à jour des poids :

$$w_k^{(i)} = w_{k-1}^{(i)} \cdot \frac{p(y_k | c_k^{(i)}, s_k^{(i)}) \cdot p(y_k^D | c_{k-1}^{(i)}, \tilde{s}_k)}{p(y_k^D | c_k^{(i)}, \tilde{s}_k)} \quad (7)$$

4 Résultats expérimentaux

Afin de montrer l'intérêt d'exploiter l'information de détection souple pour optimiser l'exploration de l'espace d'état, des simulations ont été réalisées sur deux vidéos intitulées "Marche" et "Course" issues de bases de données publiques. Pour simuler des flux vidéo à débit variable, les séquences sont sous-échantillonnées avec un taux noté DS. On compare plusieurs FPs utilisant des lois d'importance différentes :

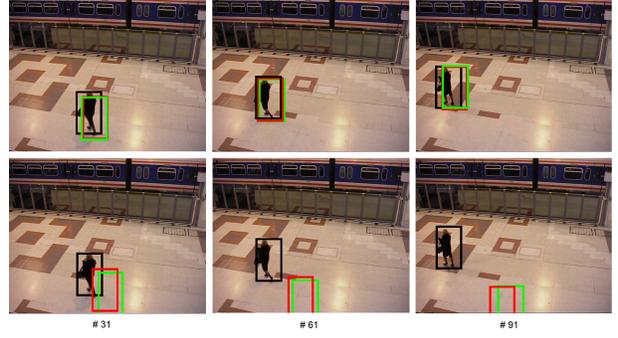


FIGURE 2 – Résultats qualitatifs de suivi sur la séquence "Marche" pour différents filtres : PF (vert), BPF avec $\alpha = 2/3$ (rouge) et NOPF (noire). En haut : DS=15, en bas : DS=30.

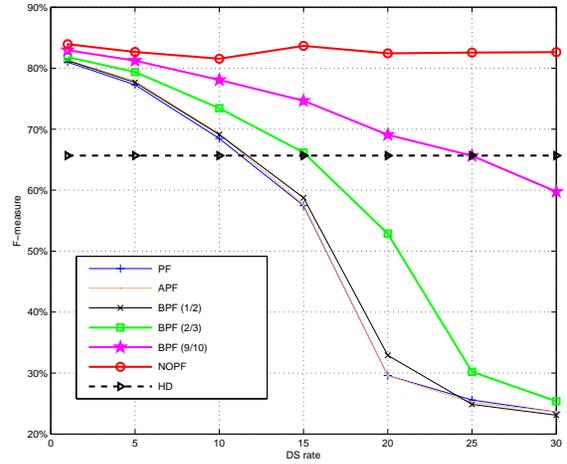


FIGURE 3 – Résultats quantitatifs de suivi : F-mesure moyen en fonction de DS sur la séquence "Marche".

- Une loi a priori Gaussienne pour le **FP conventionnel (PF)** [2] et le **FP auxiliaire (APF)** [3] qui intègre en plus une pré-sélection des particules.
- Une loi basée sur une information de détection dure pour le **FP "boosté" (BPF)** [6]. Il s'agit d'un MdG composé de la loi a priori et d'une loi centrée sur l'objet détecté de proportion α .
- La loi proposée basée sur l'information de détection souple pour le **FP quasi-optimal (NOPF)**.
- On considère en plus un **détecteur dur (HD)** appliqué indépendamment sur les images sans tenir compte de leur corrélation temporelle.

L'initialisation du suivi est manuelle et les paramètres sont les suivants : $N_p = 100$ particules, $S = 4$ bandes, $N_b = 10$ bins pour chaque canal RVB, $\lambda = 3$, $\lambda_1 = 4, 55 \cdot 10^{-4}$, $\lambda_2 = 5, 5 \cdot 10^{-5}$ et $\Sigma = \text{diag}(1600, 1600, 2, 2)$. Les résultats sont moyennés sur 100 expérimentations.

La figure 2 montre quelques résultats de suivi sur la séquence "Marche". La figure 3 montre le F-mesure moyen obtenu en fonction de DS. Cet indicateur traduit le taux de recouvrement entre la fenêtre estimée et la fenêtre correspondant à la vérité

TABLE 1 – Résultats quantitatifs de suivi : Taux de succès en fonction de DS sur la séquence "Course".

DS	PF/APF	BPF(1/2)	BPF(2/3)	BPF(9/10)	NOPF
2	34%	55%	66%	79%	100%
5	27%	37%	55%	78%	100%
10	24%	31%	36%	65%	94%

terrain. Malgré une variance élevée pour la position, les performances des filtres PF et APF décroissent rapidement quand DS augmente. La loi a priori ne permet pas une exploration correcte de l'espace d'état en cas de mouvements abrupts. La préselection des particules dans l'APF n'apporte pas d'amélioration car le bruit d'état est trop grand. Le BPF permet de mieux explorer l'espace d'état, surtout quand la proportion de la détection augmente dans le MdG ($\alpha = 2/3$ et $9/10$), ce qui montre l'apport de l'information de détection dans la loi de proposition. Le NOPF offre les meilleures performances de suivi face aux mouvements abrupts avec un F-mesure constant et supérieur à 80% jusqu'à DS= 30. La loi d'importance proposée, basée sur une information de détection souple, constitue une meilleure approximation de la loi optimale que la loi modélisée par un MdG à partir d'une information de détection dure. Enfin les résultats obtenus avec le détecteur seul montrent l'intérêt de recourir au filtrage pour $DS \leq 10$. En revanche, pour $DS \geq 25$, quand les déplacements ont une grande amplitude, ses résultats surpassent les FPs, sauf le NOPF qui exploite une information de détection souple plus riche et plus fiable.

La séquence "Course" est plus difficile : la cible a une vitesse plus élevée, une forme qui varie et des couleurs similaires à d'autres zones des images. La figure 4 montre quelques résultats de suivi et le tableau 1 résume le taux de succès obtenu sachant qu'un suivi est considéré comme réussi si F-mesure > 50%. Le détecteur HD n'apparaît pas car il donne de moins bons résultats sur cette séquence (F-mesure < 50%), les informations de détection dure ne sont pas précises. Par rapport à la vidéo précédente, les performances diminuent plus rapidement dès que DS= 2, du fait de la vitesse de l'objet. Le filtre NOPF, grâce aux informations de détection souple et à l'approximation de la loi optimale, surpasse les autres méthodes avec un taux de succès qui reste proche de 100%.

5 Conclusion

Dans cet article, nous avons proposé un filtre particulière avec une loi de proposition quasi-optimale utilisant une information de détection souple pour optimiser l'exploration de l'espace d'état dans le cas de déplacements peu prédictibles et de grande amplitude en suivi visuel. La loi d'importance proposée repose sur une approximation de la loi optimale et assure la propagation des particules dans les zones les plus probables de l'espace d'état. Les résultats montrent l'intérêt d'exploiter l'information de détection souple dans la loi de proposition,

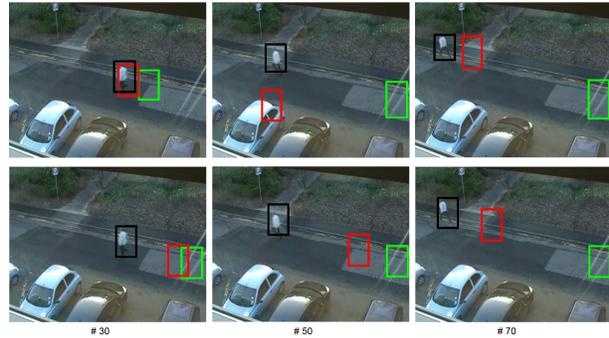


FIGURE 4 – Résultats qualitatifs de suivi sur la séquence "Course" pour différents filtres : PF (vert), BPF avec $\alpha = 2/3$ (rouge) et NOPF (noire). En haut : DS=2, en bas : DS=10.

confortent les hypothèses formulées et prouvent que notre algorithme offre un suivi précis et robuste face aux mouvements abrupts.

Références

- [1] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, "Recent advances and trends in visual tracking : A review," *Neurocomputing*, vol. 74, pp. 3823–3831, November 2011.
- [2] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, pp. 197–208, March 2000.
- [3] M. Pitt and N. Shephard, "Filtering via simulation : Auxiliary particle filters," *Journal of the American Statistical Association*, vol. 94, pp. 590–599, June 1999.
- [4] E. Maggio and A. Cavallaro, "Hybrid particle filter and mean shift tracker with adaptive transition model," in *ICASSP*, pp. 221–224, IEEE, March 2009.
- [5] J. Pantrigo, A. Sanchez, A. Montemayor, and A. Duarte, "Multi-dimensional visual tracking using scatter search particle filter," *Pattern Recognition*, vol. 29, pp. 1160–1174, June 2008.
- [6] K. Okuma, A. Taleghani, N. D. Freitas, J. Little, and D. Lowe, "A boosted particle filter : Multitarget detection and tracking," in *ECCV*, pp. 28–39, 2004.
- [7] I. Zuriarrain, A. Mekonnen, F. Lerasle, and A. Nestor, "Tracking-by-detection of multiple persons by a resample-move particle filter," *Machine Vision and Applications*, vol. 24, pp. 1751–1765, November 2013.
- [8] M. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *IEEE Transactions on PAMI*, vol. 33, pp. 1820–1833, September 2011.
- [9] D. T. Cong, F. Septier, C. Garnier, L. Khoudour, and Y. Delignon, "Robust visual tracking via MCMC-based particle filtering," in *ICASSP*, pp. 1493–1496, IEEE, 2012.
- [10] D. T. Cong, L. Khoudour, C. Achard, and A. Flancquart, "Adaptive model for object detection in noisy and fast-varying environment," in *ICIAP*, Lecture Notes in Computer Science, 2011.
- [11] P. Perez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particles," *Proceedings of the IEEE*, vol. 92, pp. 495–513, February 2004.