Suivi temps réel de personnes dans des séquences d'images couleur

G. MOSTAFAOUI[1], C. ACHARD[1], M. MILGRAM[1]

[1]L'aboratoire des Instruments et Systèmes d'Iles de France, 3 rue Galilée 94200 Ivry-sur-seine

ghiles.mostafaoui@lis.jussieu.fr, {achard, maum}@ccr.jussieu.fr

Résumé – Nous proposons dans cet article un algorithme temps réel de suivi de personnes, il est entièrement non supervisé, il ne nécessite aucune initialisation ni sur les modèles de pistes, ni sur leur nombre qui peut évoluer dans le temps. Il permet de gérer divers problèmes tels que les occlusions et les sous ou sur segmentations. La première étape du processus consiste à détecter les zones en mouvement. Les différentes régions ainsi obtenues seront affectées à des trajectoires en utilisant le concept de pistes élémentaires. Ces dernières nous permettent d'une part de faciliter le suivi et d'autre part de détecter les sorties d'occlusions en introduisant des ensembles cohérents de régions sur lesquels des modèles cinématiques, de forme ou de couleur pourront être définis. . Des résultats significatifs seront présentés sur des séquences réelles avec vérité de terrain.

Abstract *– We propose a real time algorithm to track moving persons without any a priori knowledge neither on the model of person, nor on their size or their number, which can evolve with time. It manages several problems such as occlusion and under or over-segmentations. The first step consisting in motion detection, leads to regions that have to be assigned to trajectories. This tracking step is achieved using the concept of elementary tracks. They allow on the one hand to manage the tracking and on the other hand, to detect the output of occlusion by introducing coherent sets of regions. Those sets enable to define temporal kinematical, shape and colour models. Significant results will be presented on several sequences with ground truth.*

1. Introduction

La détection et suivi d'objets temps réel est une problématique difficile qui se pose dans un grand nombre d'applications de traitement d'images comme l'interaction homme-machine, la surveillance civile et militaire, la réalité virtuelle, l'analyse du mouvement humain ou encore la compression d'images. Cette difficulté est accentuée dans les environnements sans contraintes où le système de suivi devra s'adapter à la variabilité importante des objets, aux variations de luminosité, aux occlusions (partielles ou totales) ainsi qu'aux problèmes de détection de mouvement.

Il existe dans la littérature un nombre important d'algorithmes de suivi parmi lesquels on peut citer les méthodes heuristiques simples à mettre en œuvre [6], le filtrage particulaire [8] et son extension à plusieurs pistes [7], les méthodes itératives utilisant la prédiction de la position des objets (JPDAF : Joint Probabilistic Data Association Filter) [2] , le MHT (Multiple Hypothésis Tracker) qui formule de manière récursive toutes les possibilités d'association des régions aux pistes [12], ou encore le Mean Shift [3]. Pour tous ces algorithmes, différentes sources d'information peuvent être utilisées pour la modélisation et le suivi des objets, les plus courantes étant la cinématique [1], la forme (silhouettes[4], modèles articulés 2D ou 3D [14] [9]) et les modèles d'apparence [13] [11]. Ces informations peuvent être utilisées simultanément pour accroître la robustesse par rapport aux différents bruits (segmentation, luminosité, occultation…). La méthode présentée ici s'inscrit dans cette lignée.

L'algorithme de suivi que nous proposons dans cet article est entièrement non supervisé, il ne nécessite aucune initialisation ni sur les modèles de pistes, ni sur leur nombre qui peut évoluer dans le temps. Il est réalisé grâce à une méthode heuristique utilisant à la fois la cinématique, la forme et un modèle d'apparence des objets basé sur la couleur. L'originalité de la méthode consiste à introduire la notion de pistes élémentaires. Celles-ci joueront un rôle important dans le suivi puisqu'elles permettront, d'une part de faciliter le suivi et d'autre part de détecter les sorties d'occlusion. Cet algorithme, d'exécution temps réel, utilise les résultats d'une segmentation en région au sens du mouvement. La méthode d'association des régions aux pistes, décrite dans la section 2, permet de gérer les problèmes de segmentation et les occlusions partielles ou totales, comme nous le montrerons dans les résultats de la section 3, sur des séquences réelles avec vérité de terrain (séquences proposées lors de la conférence PETS2004).

2. L'algorithme de suivi

L'algorithme, dont le synoptique est présenté ci-dessous, débute par une détection des zones en mouvement réalisée par différence de l'image courante et d'une image de fond, suivie par une relaxation markovienne [5].

Le suivi consiste alors à associer les régions obtenues aux différentes pistes, la brique de base pour l'association étant la piste élémentaire que nous définissons maintenant :

Soient R_i^t et R_j^{t+1} deux régions en mouvement appartenant respectivement aux images I_t et I_{t+1}. Ces deux régions sont voisines si leur taux de recouvrement (lorsqu'elles sont projetées dans la même image) n'est pas nul.

Pour appartenir à une même piste élémentaire, R_i^t doit avoir un seul voisin au temps t+1 (R_j^{t+1}) et R_j^{t+1} doit avoir un seul voisin au temps t (R_i^t). Les deux régions sont alors liées par une arête stricte.

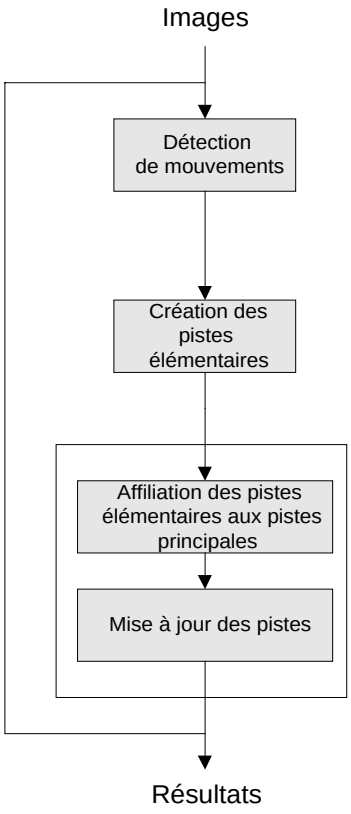

FIG. 1 : Synoptique

Sur le graphe ci-dessous où les nœuds représentent des régions et les arcs la relation de voisinage, la région R_1^4 est voisine avec deux régions au temps t-1, ces trois régions ne peuvent donc pas faire parti de la même piste élémentaire. Les régions présentes dans le graphe vont ainsi amener à la construction de 5 pistes élémentaires.

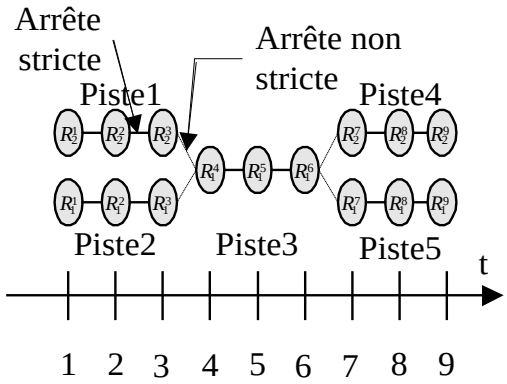

FIG. 2 : Graphe des pistes élémentaires

L'association des régions aux pistes élémentaires est qualifiée de sûre, elle ne sera pas remise en cause ultérieurement. Elle permet d'introduire des ensembles cohérents de régions qui pourront être modélisées comme des pistes avec:

- des paramètres cinématiques : centre de gravité, poids (probabilité de la piste, fonction de sa durée dans le temps) et vélocité.
- des paramètres de forme : ellipse d'axes $2\sigma_x$ et $2\sigma_y$ où σ_x et σ_y sont les écarts types en x et y.
- un modèle d'apparence : il est constitué par un modèle couleur et un vecteur de probabilités d'apparence. Le modèle couleur utilisé ici dérive de celui introduit dans [10]. Le principe est de projeter sur l'axe vertical la moyenne des couleurs des pixels appartenant à la piste et ce pour chaque plan R, V et B (voir figure 3). Nous utilisons une projection plutôt qu'un vrai modèle d'apparence 2D car celui-ci est plus robuste aux changements de forme des personnes en mouvement et, de plus, l'ajustement du modèle à la forme est simplifié (corrélation 1D). Le vecteur des probabilités d'apparence représente, pour chacune de ses lignes, le nombre d'occurrences pour lesquelles un mouvement a été détecté sur cette ligne. Il servira de pondération lors du calcul de la distance entre les modèles d'apparence.

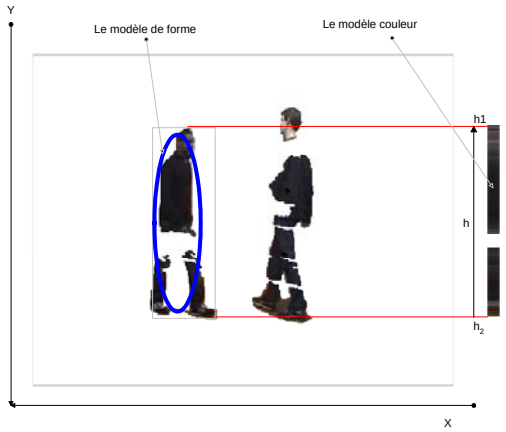

FIG. 3 : Modèle de forme et de couleur

L'algorithme de suivi va maintenant consister à regrouper les pistes élémentaires liées en pistes, chaque piste représentant idéalement la trajectoire d'une personne. Deux pistes élémentaires Pe_1 et Pe_2 sont dites liées s'il existe un chemin constitué d'arrêtes non-strictes (voir graphe ci-dessus) permettant d'aller de la piste Pe_1 à la piste Pe_2. Chaque piste sera alors composée d'un ensemble de pistes élémentaires liées. Notons que plusieurs régions de la même image peuvent appartenir à la même piste.

L'algorithme de suivi utilisé est le suivant :

Soient P_i^{t-1} i=(1..N) les pistes présentes au temps t-1 et Pe_k^t (k=1..M) les pistes élémentaires aux temps t. Il faut associer chaque piste élémentaire à une piste (ou à une nouvelle piste). Soit S_i^t l'ensemble des pistes élémentaires liées à la piste P_i^{t-1} (et donc susceptible d'appartenir à cette piste).

Plusieurs cas peuvent se produire :

a) Si S_i^t est vide alors la piste P_i^{t-1} s'arrête.

b) Si S_i^t contient une piste élémentaire qui n'est pas liée à une autre piste alors la piste P_i^{t-1} est mise à jour au temps t.

c) Si S_i^t contient une ou plusieurs pistes élémentaires liées à d'autres pistes (P_j^{t-1} par exemple), il faut opter pour la présence d'une occlusion ou d'une erreur de segmentation. Pour cela, nous étudions les vecteurs vitesses des pistes concernées (P_i^{t-1} et P_j^{t-1}).

- Si ceux-ci ont eu la même évolution au cours du temps, un problème de segmentation a été détecté et la piste ayant le poids le plus faible est détruite tandis que l'autre est mise à jour.
- Si ce n'est pas le cas, une occlusion est détectée et une nouvelle piste dite d' « occlusion » constituée par les pistes P_i^{t-1}et P_j^{t-1} est alors créée (les pistes P_i^{t-1}et P_j^{t-1} ne sont évidemment pas détruites).

d) Si S_i^t contient plusieurs pistes élémentaires qui ne sont pas liées à une autre piste, on peut être en présence d'un seul objet ou de deux objets qui se séparent. La décision va de nouveau être prise en étudiant la cinématique des pistes élémentaires concernées (une sur-segmentation est détectée en cas de cinématiques semblables, tandis que pour des cinématiques différentes, on optera pour la présence de deux objets). Distinguons les cas suivants :

- Dans le cas d'une sur-segmentation, la piste P_i^{t-1} est mise à jour.
- Dans le cas de la séparation de deux objets (détectés comme une seule piste), les pistes élémentaires liées à cette piste sont divisées en deux classes, en utilisant la cinématique des pistes élémentaires et l'algorithme des k-means. Parmi ces deux ensembles de pistes, celui qui est le plus proche de P_i^{t-1} au sens du modèle d'apparence est affecté à cette piste. Le deuxième ensemble sera affecté à une nouvelle piste qui correspondra à la nouvelle trajectoire issue de la fission.
- Si la séparation des objets est détectée et que la piste étudiée est une piste d'occlusion, les pistes élémentaires vont être divisées en deux ensembles, de la même manière que précédemment. Un modèle d'apparence va être créé pour chacun de ces deux ensembles et comparé aux modèles d'apparence des pistes ayant créé l'occlusion. Si les deux ensembles n'appartiennent pas au même modèle de piste avant occlusion et qu'il y a cohérence du point de vue de la forme des pistes, alors les deux pistes avant occlusion sont mises à jour et la piste d'occlusion est détruite (fin de l'occlusion) sinon, l'occlusion n'est pas rompue.

3. Résultats et évaluation

Notre méthode a été testée sur un grand nombre de séquences. Nous présentons ici les résultats obtenus sur une dizaine de séquences de test proposées lors de la conférence pets2004. Les scénarios sont variés allant de la simple personne qui marche seule avec divers comportements (Browse2,Browse4 et Bww2,Rsf,Rff) à plusieurs personnes qui interagissent (Mwt1,Mws : deux ou plusieurs personnes qui se croisent et marchent cote à cote, Mc : quatre personnes marchant groupées qui se séparent et se croisent entre elles, Fra1, Fomd : deux personnes qui se battent et se séparent), ou encore une personne qui vient déposer un objet à terre avant de repartir (Lbox).

L'avantage certain de ces séquences réside dans le fait d'avoir pour chacune d'entre elles, une « Vérité de terrain » (Nombre de personnes, localisation des personnes…etc). Afin de comparer nos résultats à la vérité de terrain, nous proposons l'évaluation de notre méthode comme suit : considérons une arrête (stricte ou non) qui lie deux régions. Soient E^1_{vt} et E^2_{vt} les étiquettes (numéro des pistes) de ces régions données par la vérité de terrain et E^1_r et E^2_r les étiquettes résultantes de notre méthode. L'arrête est **positive** si : ($E^1_{vt} = E^2_{vt}$ et $E^1_r = E^2_r$) ou ($E^1_{vt} \neq E^2_{vt}$ et $E^1_r \neq E^2_r$) sinon elle est **négative.** Le tableau ci-dessous résume les résultats obtenus sur toutes les séquences. En plus de l'évaluation des arrêtes (nombre d'arrêtes positives « NAP » et négatives « NAN »), nous comparons le nombre totale de personnes trouvées sur toute la séquence (NPr) à celui de la vérité de terrain (NPvt). Un certain nombre de statistiques sur ces séquences sont aussi données dans le tableau (NI : nombre d'images, NR : nombre de régions détectées en mouvement, NPvtI : nombre moyen de personnes par images, NRI : nombre moyen de régions par images). Ces statistiques permettent notamment de donner, pour chaque séquence, une idée globale des problèmes de sur ou sous-segmentations ainsi que d'occlusions. Ainsi, en comparant le nombre moyen de personnes par image de la vérité de terrain (NPvtI), au nombre moyen de région détecté par image (NRI), on remarque le nombre important de sur-segmentations. Celui-ci est dû aux fortes variations d'éclairage présentes tout au long de la séquence et gênant fortement l'algorithme de segmentation.

Pour toutes ces séquences, aucune initialisation n'a été réalisée manuellement et aucun seuil n'a été changé (l'algorithme possède un seul seuil utilisé pour l'étude de la cinématique des pistes). D'autre part, aucune information a priori n'a été introduite sur les objets à suivre. Les résultats ci-dessous montrent que notre algorithme est capable de suivre plusieurs personnes, et donc, de détecter et de gérer les interactions entre ses dernières (occlusions, sous-segmentation). Nous pouvons néanmoins constater un taux d'erreurs plus élevé pour certaines séquences (Mwt1 et Mc). Ces erreurs sont le plus souvent dues au fait que les personnes qui interagissent ont un comportement cinématique quasi identique ce qui rend difficile la détection des occlusions et des fissions. De plus pour la séquence Mws, les deux personnes qui se croisent ont des caractéristiques colorimétriques identiques, ce qui a provoqué une erreur d'étiquetage lors de la sortie de l'occlusion. Généralement, sur ces séquences difficiles, l'algorithme a un fonctionnement correct avec un nombre d'arrêtes négatives (NAN) très petit devant le nombre d'arrêtes positives(NAP)

TAB. 1 : Résultats expérimentaux

	Bro wse2	Bro wse4	Rsf	Lbox	Mwt 1	Mws	Mc	Fra1	Fom d
NI	875	1138	910	862	706	622	490	550	959
NR	3176	3654	3106	3515	1902	2005	995	3823	3056
NPvt	2	3	3	4	5	9	4	7	9
NPr	2	3	4	5	5	12	4	10	10
NPvt I	1.10	0.77	3.17	3.4	1.8	2.7	2.3	4.9	2.76
NRI	3.62	3.21	3.38	4.07	2.69	3.23	2.02	6.9	3.18
NAP	3169	3662	3108	3521	1885	1995	876	3815	3073
NAN	26	20	25	15	56	37	135	28	31

Conclusion

Nous avons présenté une méthode de suivi d'objets en mouvements temps réel et robuste utilisant simultanément la cinématique, la forme et un modèle d'apparence. Afin de gérer correctement les problèmes d'occlusions et de sur-segmentations, nous avons introduit un nouveau concept : les pistes élémentaires. Celles-ci regroupent de manière sure (arrête stricte) des régions de la séquence, ce qui permet d'accéder à des modèles temporels (modèle de forme, de vitesse, colorimétriques) qui ne pouvaient pas être définis sur des régions isolées. Des résultats expérimentaux réalisés sur diverses séquences avec vérité de terrain issues de la conférence pets2004 ont été présentés et ont permis d'évaluer les performances de notre système. Celui-ci produit de bons résultats sur ces séquences particulièrement difficiles. La suite de ce travail va consister à incorporer le système de suivi dans une application réelle de comptage de personnes en introduisant des connaissances a priori sur les objets à suivre, de manière à accroître la robustesse du système.

Références

[1] C.Achard, G.Mostafaoui, M.Milgram, *Object tracking based on kinematics with spatio-temporal blob*, MVA2005

[2] Y. Bar-Shalom, XR Li, *Multitarget-Mulisensor tracking*, Publisher: Yaakov Bar-Shalom, 1995.

[3] D.Comaniciu, P.Meer , *Mean Shift Analysis and Application* , ICCV 1999

[4] L.S.Davis,D.Harwood, I.Haritagolu, *Ghost: A Human Body Part Labeling System Using Silhouettes*, ICPR98

[5] J.Denoulet, G.Mostafaoui, L.Lacassagne, A.Merigot, *Robust Embbeded Hardware implementation of Motion Markov Detection and hysteresis thresholding in colors sequences*, CAMP2005

[6] I. Haritaoglu, D. Harwood, L.S. Davis, *W4S : a real time system for detecting and tracking people in 2,5D,* European Conference Computer Vision, 1998, Maryland.

[7] C. Hue, J.P. Le cadre, P. Perez, *Tracking multiple objects with particle filtering*, RR INRIA n° 4033, 2000

[8] M. Isard, A. Blake, *Condensation | conditional density propagation for visual tracking*, Int. J. Computer Vision, 29, 1, 5--28, 1998.

[9] H.Moon, R.Chellappa, A.Rosenfeld,*T racking of Human Activities Using Shape-encoded Particle Propagation*, ICIP2001

[10] A.Mittal, L.S.Davis, M_2 Tracker : *A Multi-View Approach to Segmenting and tracking people in a Cluttered Scene*, IJCV2003

[11] S.Park, J.K.Aggarwal, *Segmentation and tracking of interacting human body parts under occlusion and shadowing*, Motion2002

[12] D.B. Reid, *An algorithm for Tracking Multiple Targets*, IEEE Trans. on Automatic Control, Vol. AC-24, N° 6, pp 843-854, 1979.

[13] A.Senior, *Tracking People with Probabilistic Appearance Models*, Pets2002

[14]L.Wang, H.Ning, T.Tan, W.Hu, *Fusion of static and dynamic body biometrics for gait recognition*, ICCV2003