Suivi par ré-identification dans un réseau de caméras à champs disjoints

Boris Meden, Patrick Sayd, Frédéric Lerasle

Suivi par ré-identification dans un réseau de caméras à champs disjoints

Boris Meden[1] Patrick Sayd[1] Frédéric Lerasle[2 3]

[1] CEA, LIST, Laboratoire Vision et Ingénierie des Contenus, BP 94, F-91191 Gif-sur-Yvette, France,
[2] CNRS ; LAAS ; 7 avenue du Colonel Roche, F-31077 Toulouse Cedex 4, France,
[3] Université de Toulouse ; UPS, INSA, INP, ISAE ; UT1, UTM, LAAS ; F-31077 Toulouse Cedex 4, France

{boris.meden, patrick.sayd}@cea.fr, lerasle@laas.fr

Résumé

Cet article adresse le problème de suivi automatique de piétons au travers de réseaux de caméras à champs de vue disjoints. Le suivi dans l'image est traité de manière locale par un algorithme de Suivi-par-Détections et ré-identification. Avec du filtrage particulaire à état mixte, nous introduisons la notion d'identité globale dans un algorithme de suivi multi-pistes pour caractériser les personnes au niveau du réseau et pallier aux discontinuités d'observations. Nous venons renforcer la décision de ré-identification en proposant un schéma décisionnel haut niveau intégrant les hypothèses de chaque traqueur confrontées à la topologie du réseau.
La composante suivi multi-personnes et ré-identification est d'abord testée en contexte monocaméra. Nous évaluons ensuite notre approche complète sur un réseau de 3 caméras à champs de vue disjoints et un ensemble de 7 personnes. La seule connaissance a priori requise est la carte topologique du réseau.

Mots Clef

Ré-identification, suivi de personnes, réseau de cameras, champs de vue disjoints, filtrage particulaire.

Abstract

This article tackles the problem of automatic multi-pedestrian tracking in non-overlapping fields of view camera networks, using monocular, uncalibrated cameras. Tracking is locally addressed by a Tracking-by-Detection and re-identification algorithm. We propose here to introduce a notion of global identity into a multi-target tracking algorithm, qualifying people at the network level, to allow us to rebound observation discontinuities. We embed that identity into the tracking loop thanks to the mixed-state particle filter framework, thus including it in the search space. Doing so, each tracker maintains a mutli-modality on the identity in the network of its target. We increase the decision strength introducing a high level decision scheme which integrates all the trackers hypothesis over all the cameras of the network with previous re-identification results and the topology of the network.

FIGURE 1 – Réseau de caméras à champs disjoints.

The tracking and re-identification module is first tested with a single camera. We then evaluate the whole framework on a 3 non overlapping fields of views network with 7 identities. The only a priori knowledge assumed is a topological map of the network.

Keywords

Re-identification, tracking, camera network, non-overlapping fields of view, particle filtering.

1 Introduction

Les travaux présentés dans cet article ont pour objectif le suivi de personnes en environnement grande échelle. Les contraintes matérielles/économiques limitent en général le nombre de caméras et empêchent une couverture totale de l'espace, ce qui engendre des discontinuités dans le champ de vue du réseau. On parle de réseaux à champ de vue disjoint comme celui présenté en figure 1.
L'enjeu pour le processus de suivi est alors de gérer ces discontinuités dans le champ de vue du réseau pour assurer la cohérence spatio-temporelle. Outre le suivi des personnes dans chaque image, le système doit ré-identifier les personnes lors de leur apparition dans les différents champs

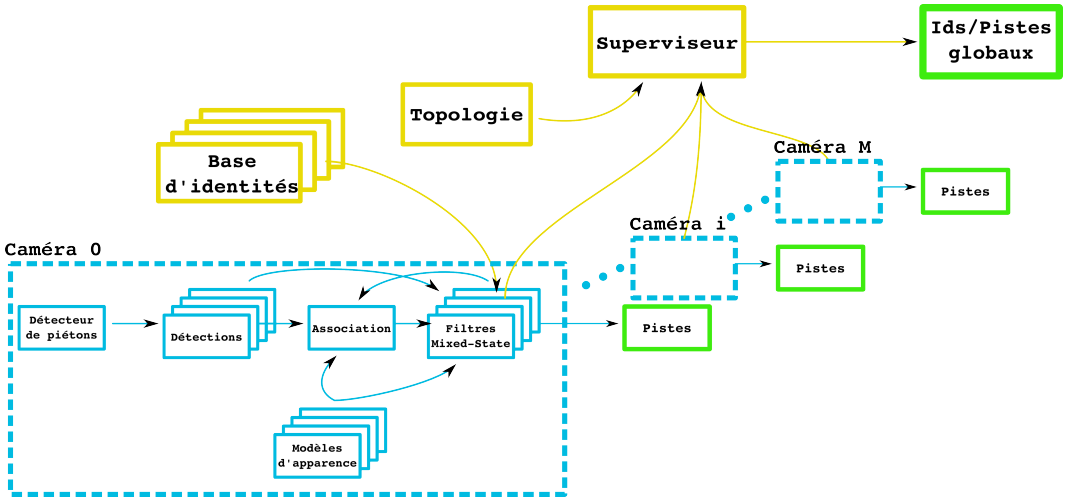


FIGURE 2 – Achitecture du système perceptuel. Les traitements «suivi et ré-identification dans l'image» sont localisés au niveau des caméras, alors que le superviseur a la connaissance de tout le réseau et confronte les distributions d'identités à sa topologie.

de vue.

Nous proposons ici d'intégrer le formalisme de filtrage particulaire à état mixte pour la ré-identification [13] dans un algorithme de suivi-par-détection multi-piste [2]. Ceci permet une stratégie de ré-identification en ligne, intégrée au suivi, se basant sur une description colorimétrique des identités. La seconde contribution de ce papier réside dans l'ajout d'un superviseur intégrant les résultats d'identification des traqueurs à la manière de tracklets, tout en les validant par rapport à la topologie du réseau.

La figure 2 schématise l'ensemble de l'architecture proposée, avec les traitements locaux aux caméras, et la centralisation des résultats de ré-identification.

Les travaux antérieurs sont passés en revue dans la section 2. Puis nous détaillons le traqueur par ré-identification inhérent à chaque caméra dans la section 3. La supervision de ces ré-identifications de pistes est ensuite détaillée en section 4. Finalement, la section 5 ouésente des évaluations quantitatives de la fonction de base de ré-identification, ainsi que de l'apport de contraintes topologiques lorsqu'appliquées à un réseau.

2 État de l'art

La ré-identification de personnes devient nécessaire dès que les trajectoires des cibles suivies incluent des discontinuités dues à des pertes d'observabilité temporaires. La notion sous-jacente est celle d'*identité globale* au sein du réseau par opposition à l'identité locale propre à un traqueur suivant une piste localement dans une caméra. La qualité d'un suivi multi-pistes est évaluée notamment par la capacité des traqueurs à demeurer sur la même personne dans l'intervalle de temps où la personne est observable (*i.e.* conserver la même *identité locale* [1]). Cette notion d'identité, se limite à la continuité spatio-temporelle du suivi (la sortie et la réapparition de personnes dans le champs de vue de la caméra ne génèreront pas la même *identité locale*). Le problème du suivi et de l'identification conjointe de personnes au sein de réseaux de caméras à champs joints *e.g.* [15] est très similaire. La combinaison de plusieurs flux vidéo issus de plusieurs capteurs passe généralement par une calibration du dispositif, permettant de travailler dans un repère commun, et l'identification de pistes s'appuie sur leur continuité spatio-temporelle.

En revanche, un réseau à champs disjoints (figure 1) présente des discontinuités d'observations correspondant aux temps de transit des cibles entre les divers noeuds (caméras) du réseau ou aux entrées/sorties au sein d'une même caméra. Ce problème est appelé la ré-identification de personnes, et nous introduisons ici la notion d'*identité globale*, qui sera l'identité de cette personne lors de chacune de ses apparitions dans les caméras du réseau.

Ce problème de ré-identification est classiquement traité par «requête dans une base d'image», inspiré des technologies web, et met l'accent sur la description de l'apparence de la personne à ré-identifier. Ainsi Gray *et al.* [6] proposent d'entrainer un classifieur sur les composantes invariantes par changement de caméra. Farenzena *et al.* [5] adoptent la même approche en s'affranchissant de la composante apprentissage. Ceci passe notamment par le calcul d'axes de symétrie des silhouettes et en imposant des caractéristiques dédiées à la représentation de piétons. Ces méthodes coûteuses en temps de calcul se prêtent bien à un traitement *a posteriori*.

Pour une application sur réseau de caméras, le module de ré-identification doit permettre un traitement vidéo temps réel. En effet nous visons ici un maintien en continu de l'identité des pistes au sens réseau. Une problématique similaire est abordée par [3, 12, 17], cependant [17] fait l'hypothèse de passages de personnes seules dans le réseau, [3] ne donne pas de détails sur sa composante de suivi et [12]

simule un réseau de caméra à champs disjoints, et donc n'est pas non plus confronté au problème de suivi dans chacune d'elles. Ces travaux ne considèrent pas de manière conjointe le problème de suivi et ré-identification et occultent les difficultés inhérentes au suivi multi-personnes.
Le suivi multi-personnes est largement abordé dans la communauté Vision ; plusieurs constats associés ont motivé notre approche. L'intérêt des algorithmes de filtrage particulaires pour le suivi est acquis depuis les travaux initiaux de Isard et Blake dans [8], notamment en contexte multi-cibles. Depuis [14], les approches *tracking-by-detection* émergent et en particulier l'intégration temporelle de *tracklets* (portion de trajectoire d'une piste), dont la robustesse a été prouvée par Kaucic *et al.* dans [9]. L'optimisation de *tracklets* a par ailleurs été étendue à deux caméras présentant un champ disjoint entre elles [11]. Toutefois cette méthode ne travaille pas en ligne, car réalise une optimisation sur une fenêtre temporelle.
Par opposition, notre approche se place en contexte markovien au niveau du module de suivi. Notre approche s'inspire de [2] et de [16]. A l'instar de [2], elle repose sur des filtres particulaires distribués mais elle rajoute la composante ré-identification via une variable discrète relative à l'identifiant de la cible. On parle alors de filtrage particulaire à état mixte. A l'instar de [16], une intégration temporelle (tracklet) sur les identités de pistes est mise en oeuvre mais non à l'échelle d'une ou plusieurs caméras à champs joints mais au niveau du réseau de part notre problématique.

3 Suivi par ré-identification au sein d'une caméra

Dans cet article, nous proposons une extension aux réseaux à champs disjoints de l'algorithme de suivi-par-détection proposé par Breitenstein *et al.* dans [2], en introduisant la notion d'*identité globale*. Nous présentons dans cette section notre implémentation de [2] et comment l'utilisation du filtrage particulaire à état mixte pour la ré-identification [13] vient étendre cette approche.

3.1 Description des cibles

FIGURE 3 – Image-clés relatives à chaque ID (caméra 1).

Apprentissage des identités du réseau. Tout algorithme de ré-identification nécessite d'avoir vu une personne au préalable pour être capable de la ré-identifier. Nous supposons ici la phase de consitution d'une telle base acquise. Pour cela, nous extrayons une collection d'image-clés d'une des caméras du réseau (*e.g.* positionnée dans le hall d'entrée du bâtiment à surveiller), et utilisons celles-ci comme descriptions de nos identités. Le choix des image-clés est réalisé par K-means sur des séquences de suivi dans la caméra choisie à la manière de [13]. Ainsi, ces image-clés encodent la variabilité d'apparence obtenue pour cette identité au cours de son suivi initial. La figure 3 présente un exemple de base d'identités utilisé pour traiter le réseau de la figure 1, apprise ici dans la caméra 1.

Modélisation de l'apparence d'une piste. Nous utilisons le même modèle d'apparence que [13] pour décrire les pistes de tracking ainsi que les identités de la base : des bandes horizontales de distributions couleur calculées dans l'espace RGB. La mesure de similarité entre deux descripteurs est la somme des distances de Bhattacharyya, évaluées sous un noyau gaussien. Ceci nous permet de calculer les similarités par rapport au modèle d'apparence d'un traqueur ainsi que par rapport à une identité de la base, notées respectivement $w_{App}(.)$ et $w_{Id}(.)$.

3.2 Gestion des détections

Association aux détections. Notre approche privilégie une stratégie «tracking-by-detection» via le détecteur classique proposé dans [4]. Ces détections sont intégrées dans le processus de suivi par une étape d'association préalable de type algorithme glouton. À la fin de cette étape, chaque traqueur est potentiellement associé à une détection qui va servir à la mise à jour de ses particules. Pour ce faire, nous construisons une matrice d'association entre les détections (lignes) et les traqueurs (colonnes). Le score de chaque paire détection d, traqueur tr donné par l'équation (1), fait intervenir :
– la distance des particules du traqueur à la détection évaluées sous une loi normale $p_{\mathcal{N}}(.) \sim \mathcal{N}(., \sigma^2)$,
– l'aire de la boîte du traqueur $\mathcal{A}(tr)$ relativement à celle de la détection aussi évaluée sous une loi normale,
– l'évaluation du modèle d'apparence du traqueur en la détection ($w_{App}(.)$).

$$S(d,tr) = \underbrace{\sum_{p\in tr}^{N} p_{\mathcal{N}}(d-p)}_{\text{distance euclidienne}} \times \underbrace{p_{\mathcal{N}}\left(\frac{|\mathcal{A}(tr)-\mathcal{A}(d)|}{\mathcal{A}(tr)}\right)}_{\text{taille relative}} \times \underbrace{w_{App}(d,tr)}_{\text{modèle d'apparence}} \quad (1)$$

Ainsi, le traqueur et la détection doivent présenter simultanément une cohérence en terme de position, de taille et de contenu colorimétrique. Une fois cette matrice de similarité construite, on extrait itérativement les maxima, avec suppression de leurs lignes et colonnes. On itère tant que les maxima sont supérieurs au seuil d'appariement. Une telle heuristique est préférée à la solution optimale fournie par l'algorithme Hongrois [10], écarté pour sa complexité.

Initialisations / terminaisons automatiques de traqueurs. Toute détection récurrente temporellement donne lieu à l'instanciation d'un nouveau traqueur. Par ailleurs, tout traqueur n'ayant pas de détection associée sur un intervalle de temps supérieur au seuil de suppression se voit arrêté.

3.3 Filtrage particulaire

Modèle de prédiction à état mixte. Chaque piste initialisée par une détection est suivie par un filtre à particules. Étant donnée la base d'identités, nous avons des descripteurs de référence supplémentaires auxquels se comparer. Pour cela, à l'instar de [13], nous utilisons des filtres de type Mixed-State CONDENSATION, introduit dans [7]. Nous cherchons à estimer un vecteur d'état mixte, ajoutant un terme discret aux paramètres continus, soit

$$\mathbf{X} = (\mathbf{x}, id)^\mathsf{T},\ \mathbf{x} \in \mathbb{R}^4,\ id \in \{1, \dots, N_{id}\}$$

La partie continue de l'état $\mathbf{x} = [x, y, v_x, v_y]^\mathsf{T}$ se compose de la position dans le plan image $(x, y)^\mathsf{T}$ et du vecteur vitesse $(v_x, v_y)^\mathsf{T}$. La partie entière id renvoie à l'une des N_{id} identités de la base. Le suivi se passe dans le plan image, et la dimension des boîtes de suivi est fixée et mise à jour sur les détections associées à ces traqueurs. Le modèle d'apparence est lui aussi mis-à-jour sur la détection associée. Étant donné ce vecteur d'état étendu, la densité du processus d'échantillonnage à l'image t peut être décomposée comme dans [7] :

$$p(\mathbf{X}_t|\mathbf{X}_{t-1}) = p(\mathbf{x}_t|id_t, \mathbf{X}_{t-1}) \cdot P(id_t|\mathbf{X}_{t-1})$$

$$\begin{array}{cc} P(id_t|\mathbf{X}_{t-1}) : & P(id_t = j|\mathbf{x}_{t-1}, id_{t-1} = i) = T_{ij}(\mathbf{x}_{t-1}) \\ p(\mathbf{x}_t|id_t, \mathbf{X}_{t-1}) : & p(\mathbf{x}_t|\mathbf{x}_{t-1}, id_{t-1} = i, id_t = j) = p_{ij}(\mathbf{x}_t|\mathbf{x}_{t-1}) \end{array}$$

où $T_{ij}(\mathbf{x}_{t-1})$ est la probabilité de transition de l'identité i vers j, appliquée au paramètre discret d'identité, et $p_{ij}(\mathbf{x}_t|\mathbf{x}_{t-1})$ est l'échantillonnage de la loi appliquée à la partie continue de l'état. La matrice de transition $T = [T_{ij}]$ est construite sur l'ensemble des images clés. L'élément T_{ij} est la similarité $w_{id}(.)$ entre les identités i et j de la base, calculée entre les images clés les plus dissemblables. Les particules sont propagées selon un modèle de mouvement d'ordre 1 :
$p_{ij}(\mathbf{x}_t|\mathbf{x}_{t-1})$:

$$\begin{cases} (x, y)_t = (x, y)_{t-1} + (v_x, v_y)_{t-1} \cdot \Delta t + \epsilon_{(x,y)} \\ (v_x, v_y)_t = (v_x, v_y)_{t-1} + \epsilon_{(v_x,v_y)} \end{cases}$$

où les bruits $\epsilon_{(x,y)}$ et $\epsilon_{(x,y)}$ suivent des lois normales et où Δt est l'intervalle de temps séparant deux images.

Modèle d'observation intégrant les détections. Le poids $w_{tr}^{(p)}$ attribué à la p^e particule du traqueur tr est calculé en intégrant la distance de la particule à la détection d^* qui lui a été associée, la similarité colorimétrique au modèle d'apparence du traqueur $w_{App}(.)$ et la similarité colorimétrique à la l'identité de la particule $w_{Id}(.)$. $Id(p)$ représente l'identité choisi par p. Il s'agit du terme mixte.

$$w_{tr}^{(p)} = \underbrace{\alpha \cdot \mathcal{I}(tr) \cdot p_{\mathcal{N}}(d^* - p)}_{\text{distance à la détection}} + \underbrace{\beta \cdot w_{App}(d, tr)}_{\text{modèle d'apparence}} + \underbrace{\gamma \cdot w_{Id}(d, id(p))}_{\text{identité}} \quad (2)$$

où α, β et γ sont des coefficients dont la somme est égale à 1, et $\mathcal{I}(tr)$ un booléen signifiant l'existence ou non d'une détection associée au traqueur. Comme dans [13], l'introduction d'une similarité relative à l'identité dans la pondération de la particule permet de diriger le nuage de particule vers les identités les plus probables aux vues des observations reçues. En ce sens, chaque traqueur maintient une multi-modalité sur les *identités globales* les plus vraisemblables pour la personne qu'il suit.
L'estimation de l'état est un processus en deux étapes. Nous commençons par calculer le Maximum A Posteriori sur le paramètre discret par rapport à l'observation courante $\mathbf{Z}_t$ avec l'équation (3), *i.e.* l'identité la plus probable à l'instant t.

$$\hat{id}_t = \arg\max_j P(id_t = j|\mathbf{Z}_t) = \arg\max_j \sum_{p \in \Upsilon_j} w_{tr}^{(p)}(t), \quad (3)$$
$$\text{où } \Upsilon_j = \left\{ p|\mathbf{X}_t^{(p)} = (\mathbf{x}_t^{(p)}, j) \right\}$$

Ensuite, les composantes continues sont estimées sur le sous-ensemble de particules $\hat{\Upsilon}$ qui possèdent l'identité la plus vraisemblable, selon l'équation (4).

$$\hat{\mathbf{x}}_t = \sum_{p \in \hat{\Upsilon}} w_{tr}^{(p)}(t) \cdot \mathbf{x}_t^{(p)} / \sum_{p \in \hat{\Upsilon}} w_{tr}^{(p)}(t), \quad (4)$$
$$\text{où } \hat{\Upsilon} = \{p|\mathbf{X}_t^{(p)} = (\mathbf{x}_t^{(p)}, \hat{id}_t)^\mathsf{T}\}$$

De cette manière, en plus de l'estimation de la position image de la cible, chaque filtre fournit une distribution d'identités pour cette cible.

4 Supervision topologico-temporelle des ré-identifications

La section 3 a présenté une stratégie de ré-identification intégrée au suivi. Cette stratégie a été établie comme supérieure à une comparaison exhaustive à la base par [13]. Sa limitation réside dans le caractère distribué des filtres à état mixte. En effet, les densités de probabilités sur l'identité de la personne suivie sont indépendantes d'un filtre à l'autre. Deux filtres peuvent produire simultanément la même identité. Nous souhaitons ici contraindre ceci de manière à produire un appariement filtre/identité exclusif via leur interaction au niveau du réseau.

4.1 Génération de tracklets sur les identités

Ajout de la topologie. Dans cette partie, nous supposons disposer de la topologie du réseau sur lequel nous travaillons. Cette topologie est représentée par un graphe non orienté $G = (V, E)$ où les noeuds V représentent les zones d'entrées/sorties des caméras, et les arêtes E donnent les transitions possibles entre ces zones, comme présenté en figure 4.
Cet *a priori* fixé ici pourrait être appris en ligne à l'aide de méthodes telles que [3].

Intégration temporelle. Chaque traqueur produit à chaque instant une distribution discrète de probabilités sur l'ensemble des identités, calculée comme le ratio de particules dédiées à une identité. Ces probabilités sont aggrégées

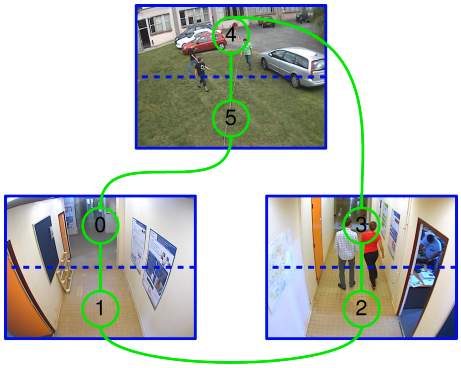


FIGURE 4 – Modélisation de la topologie du réseau de caméras. Un graphe non orienté relie les zones d'entrées/sortie des caméras.

sur une fenêtre temporelle dans un formalisme de type Programmation Dynamique. On parle alors de tracklet à l'instar de [16]. Ce faisant, nous construisons une matrice d'association entre les traqueurs et les identités de la base selon l'équation (5).

L'utilisation de la topologie du réseau intervient à ce niveau. Elle permet de supprimer les associations traqueur/identité impossibles. Nous partons d'une localisation initiale des identités dans le réseau. À chaque terminaison de traqueur, nous mettons à jour cette localisation avec sa ré-identification. Nous utilisons cette localisation pour mettre à zéro les associations incohérentes avec celle-ci. Une association est dite incohérente si la zone d'entrée/sortie dans laquelle se trouve le traqueur n'est pas connexe avec la dernière localisation enregistrée de l'identité testée.

$$S(tr_{t_0+T}, id_{t_0+T}) = p\left(id_{t_0+T}|zone(tr_{t_0})\right) \cdot \sqrt[T]{\prod_{t=t_0+1}^{t_0+T} \text{Card}\left(\Upsilon_{tr,id_t}\right)} \quad (5)$$

$$\text{où } \Upsilon_{tr,id_t} = \left\{p|\mathbf{X}_t^{(p)} = (\mathbf{x}_t^{(p)}, id_t)^\mathsf{T}\right\}$$

et où

$$p\left(id|zone(tr)\right) = \begin{cases} 1 & \text{si localization[id]=zone(tr)} ; \\ 0 & \text{sinon.} \end{cases}$$

Exclusivité de l'association. Une affectation exclusive de même type que celle décrite en section 3, travaillant à partir de la fonction de similarité (5) permet d'obtenir une association exclusive traqueurs/identités en fin de fenêtre temporelle. La topologie et les ré-identifications précédentes interviennent pour supprimer des possibilités. Finalement, l'association traqueurs/détections impose une exclusivité entre les paires résultantes.

La gestion des identités au sein du suivi permet d'éviter les problèmes de combinatoire inhérent à la gestion de pistes multiples et de maintenir constamment à jour la répartition dans le réseau des *identités globales* évoluant dedans.

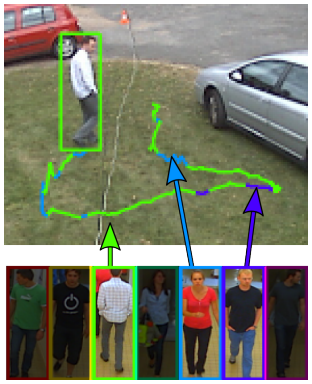


FIGURE 5 – Tracklets d'identités au cours d'une séquence de suivi. (meilleur rendu en version couleur)

4.2 Optimisation des tracklets sur une séquence de suivi

Ces affectations supervisées interviennent à la fin de chaque fenêtre temporelle. Chaque fenêtre temporelle fournit une identification pour toute sa durée. Nous obtenons ici des tracklets d'identités. La figure 5 présente les différentes tracklets d'identités inférés par le superviseur pour une séquence de suivi. Les différentes couleurs réfèrent à différentes identités de la base.

Pour ne pas biaiser le processus de ré-identification sur le début de la séquence de suivi, nous venons remettre la distribution des identités du mixed-state à l'équiprobabilité à chaque fin de fenêtre temporelle. Ainsi, le processus de recherche d'identité de la cible propose à nouveau toutes les identités de la base, et converge à nouveau vers une en particulier, selon les observations qu'il reçoit.

Pour chacun des traqueurs actifs nous mémorisons ces ré-identifications dans un accumulateur indexé sur les identités de la base. Suivant les principes de la programmation dynamique, à chaque instant l'identité affectée à ce traqueur est le mode prédominant dans cet accumulateur (vote majoritaire). De la même manière, lorsqu'un traqueur s'arrête, il se voit affecter l'identité ayant eu le plus de votes dans son accumulateur, et la localisation dans le graphe topologique de cette identité est mis à jour.

5 Implémentation et évaluations associées

5.1 Implémentation

Notre réseau IP fournit une fréquence moyenne de 16 images par secondes pour le flux vidéo à traiter. Nous fixons donc $\Delta t = 1/16s$ dans le modèle d'évolution des filtres particulaires. Dans le modèle d'observation des particules, équation (2), nous fixons de manière empirique :

$$\begin{cases} \alpha = 0.90,\ \beta = 0.05 \text{ et } \gamma = 0.05 & \text{si } \mathcal{I}(tr) = 1 \text{ [2]} \\ \alpha = 0.0,\ \beta = 0.8 \text{ et } \gamma = 0.2 & \text{sinon, [13].} \end{cases}$$

Dans le superviseur la durée des intégrations temporelles est fixée à 7 images, ce qui correspond au temps moyen de convergence des filtres mixed-state sur leur identité.

5.2 Évaluations

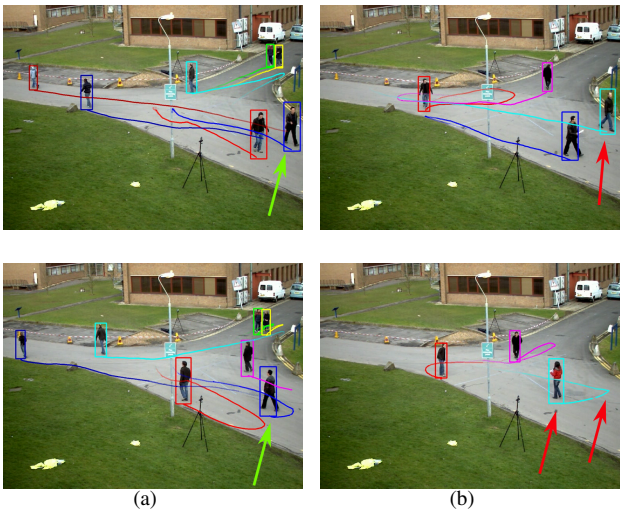
Jeux de données. Nous évaluons les différentes composantes de notre approche sur deux jeux de données. Tout d'abord nous testons le traqueur dédié à chaque noeud/caméra, sans, puis avec le module de ré-identification actif, sur la séquence PETS'09 S2L1 [1]. Cette séquence publique, longue de 795 images, présente un espace ouvert, dans lequel évoluent 10 individus, avec croisements et entrées/sorties. Ayant labellisé ce jeu de données, nous sommes en mesure de quantifier les résultats de notre algorithme de suivi.
Au niveau du réseau à champs disjoints et étant donné l'absence de jeux de données publics associés, nous évaluons la composante de supervision sur une séquence privée notée NOFOVNetwork (Non Overlapping Field Of View Network) dans la suite. La séquence présente un ensemble de 7 personnes transitant entre 3 caméras. Il n'y a pas de champs de vue commun entre les caméras, deux sont placées en intérieur de bâtiment, alors que la $3^{ème}$ surveille un espace ouvert extérieur avec une configuration similaire à la séquence PETS'09. La séquence représente 837 images. Notre objectif est de rendre ces données publiques.

Critères et modalités évalués. Nous utilisons les métriques CLEAR MOT [1] pour la quantification des résultats de suivi. Nous obtenons un score de précision : MOTP (Multi-Object Tracking Precision), calculé comme le rapport de l'intersection sur l'union des boîtes de suivi avec celles de la vérité terrain, et un score d'«accuracy » : MOTA (Multi-Object Tracking Accuracy) prenant en compte les faux positifs, les faux négatifs et les changements de cibles des traqueurs.
De plus nous évaluons les capacités de ré-identification par un taux de ré-identification correcte TRR (True Re-identification Rate), calculé comme le rapport du nombre de ré-identifications correctes sur le nombre de ré-identifications total. Étant donné que le superviseur opère sur une fenêtre temporelle, les TRR sont estimés en fin de fenêtre temporelle.

5.3 Performances au niveau caméra

FIGURE 6 – Base des 10 identités de la séquence PETS.

Notion d'*identité globale*. La figure 6 présente la base d'identités utilisée pour traiter la séquence PETS. Il s'agit ici d'un contexte monocaméra. Les images de la base sont donc issues de la caméra où est réalisé le suivi. Nous présentons ici quelques images de la séquence, pour chaque identité.



(a) (b)

FIGURE 7 – Résultats issus de [2] (a) Entre les images 204 et 241 la personne fléchée sort puis entre à nouveau. Un traqueur est maintenu et apparie la bonne personne, sur un critère spatial. (b) La même situation se produit entre les images 390 et 445. Mais cette fois, une nouvelle personne entre, un traqueur déjà existant lui est affecté, la trajectoire est reprise. Il s'agit d'une erreur de ré-identification.

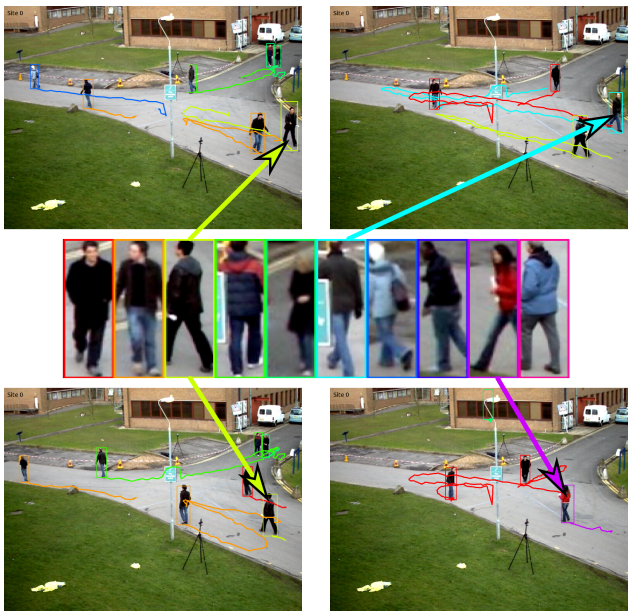


FIGURE 8 – Intérêt de la ré-identification intégrée au suivi multipiste : en (a) comme en (b), le système ré-identifie la piste suivie par rapport à la base d'identités et permet de détecter que la personne est différente (b).

Les figures 7 et 8, présentent la limitation d'une simple gestion d'*identités locales* et l'apport de notre modalité

1. http ://www.cvg.rdg.ac.uk/PETS2009/a.html

de ré-identification. Sur la figure 7, lorsqu'une personne sort et une personne différente entre, les trajectoires sont raboutées. Un simple critère spatial est utilisé dans [2]. La figure 7 (b) met en exergue cette limitation lorsque la personne sortie n'est pas la même que celle qui entre. Le traqueur serait également pris en défaut si la personne suivie réapparait dans une autre région de l'image, typiquement dans un réseau de couloir.
Dans notre cas (figure 8), à chaque instant, chaque traqueur propose une distribution de probabilité d'identité observée. Ceci permet d'accepter des périodes de non observabilité comme une sortie de caméra puis de ré-initialiser le traqueur avec le bon identifiant. Lorsqu'une personne entre, le traqueur qui la suit va converger vers des identités de la base.

Performances quantificatives. La table 1 présente nos résultats quantitatifs sur les deux séquences utilisées. PETS'09 nous a permis dans un premier temps de valider notre implémentation de [2], dont certains aspects n'ont pas été implémentés (utilisation de la confiance du détecteur dans le modèle d'observation, modèle d'apparence de type Boosting Online).
Cependant, notre approche dispose d'une modalité supplémentaire avec la notion d'*identité globale*. Nous montrons d'abord que l'introduction du filtrage particulaire à état mixte ne dégrade pas les performances de suivi, en comparant MOTP et MOTA pour notre implémentation sans et avec le module de ré-identification actif. Puis cette modalité supplémentaire permet d'exprimer le taux de ré-identification pour la séquence. Finalement, nous comparons ces résultats de ré-identification seule à l'approche filtres supervisés, dans laquelle l'exclusivité entre les ré-identifications est imposée (section 4). Cette contrainte d'exclusivité induit des scores de ré-identification meilleurs.
L'aspect stochastique du filtrage particulaire est pris en compte : la table 1 présente les résultats moyens de chaque score, sur un ensemble de 10 répétitions.

Séquence PETS'09	MOTP	MOTA	TRR
Suivi-par-détection [2]	**56.3%**	**79.7%**	-
Suivi-par-détection implémenté	42.7%	77.9%	-
Suivi-par-Réidentification	42.5%	77.7%	59.7%
Suivi-par-Réidentification supervisé	42.4%	75.9%	**64%**

TABLE 1 – Résultats de suivi selon les métriques CLEAR MOT [1] et taux de ré-identification sur la séquence mono-caméra PETS'09 S2L1. Nous donnons ici les Multi-Object Tracking Precision (MOTP), Multi-Object Tracking Accuracy (MOTA), et True Re-identification Rate (TRR) définis en section 5.2.

5.4 Performances du superviseur

La séquence NOFOVNetwork n'étant pas annotée pour le suivi, nous ne présentons que des taux de ré-identifications pour cette séquence. Nous comparons ici la méthode se basant uniquement sur les informations de couleur introduite dans le filtrage particulaire (approche inspirée de [13]), avec le système supervisé que nous proposons en section 4 et son ajout de contraintes topologiques.

La table 2 présente les taux de ré-identification par caméra, puis au niveau du réseau global. La base étant construite à partir de la caméra 1, ceci explique les taux de ré-identification supérieurs dans cette caméra. Ces résultats illustrent l'apport du superviseur : chaque identité correctement ré-identifiée contraint le système dans la suite de la topologie.

Séquence NOFOV	cam0	cam1	cam2	réseau
Suivi-par-Réidentification	43.7%	67.3%	55.5%	54.6%
Suivi-par-Réidentification supervisé	**67.7%**	**76.9%**	**63.8%**	**68.2%**

TABLE 2 – Taux de ré-identifications correctes TRR pour chacune des caméras du réseau NOFOVNetwork : comparaison des approches sans, et avec superviseur sur le réseau.

Finalement, la figure 9 représente la sortie de notre méthode, avec les suivis dans les images et la localisation des identités dans la topologie du réseau.

6 Conclusion et perspectives

Cet article traite de la surveillance de personnes évoluant dans des réseaux de caméras à champs disjoints, *i.e* localiser en ligne les pistes dans la topologie. Ceci passe par la notion d'*identité globale*. Nous avons présenté une méthode de suivi par ré-identification, travaillant à deux niveaux en se basant respectivement sur des signatures colorimétriques et sur des contraintes spatio-temporelles dans le réseau.

Le niveau caméra est traité par un tracking-par-détection markovien inspiré de [2], enrichi par la notion d'*identité globale* prise en compte au sein des filtres particulaires avec le formalisme d'état mixte. Ainsi, chaque traqueur maintient une multi-modalité sur l'identité qu'il est en train de suivre. Ce faisant, il intègre la capacité de ré-initialisation après disparition puis ré-appariation dans le champ de vue de la caméra.

Ces distributions d'identités, considérées comme des tracklets sur les identités, sont filtrées spatio-temporellement au niveau du réseau par un superviseur. Celui-ci impose l'exclusivité des ré-identifications et s'assure de leur cohérence dans la topologie du réseau.

Une première extension vise l'apprentissage en ligne des image-clés et leur mise à jour au fur et mesure du suivi. Un modèle d'apparence plus évolué entrainé en ligne sur la cible qu'il caractérise rendrait par ailleurs ceci plus simple. Finalement, une analyse dans le plan du sol et non image pour le traqueur serait à considérer.

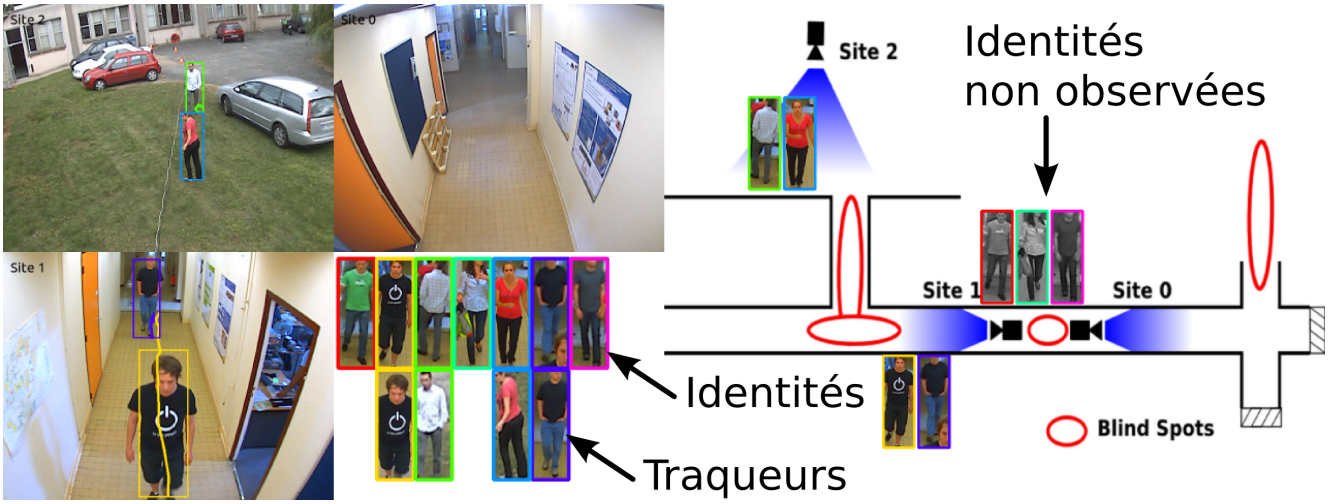


FIGURE 9 – Exemple de suivi dans le réseau avec maintien d'*identités globales* sur les pistes, permettant de les localiser dans la topologie du réseau.

Références

[1] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance : the clear mot metrics. *Journal on Image and Video Processing*, 2008.

[2] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multi-person tracking-by-detection from a single, uncalibrated camera. *Pattern Analysis and Machine Intelligence*, 2010.

[3] K.W. Chen, C.C. Lai, Y.P. Hung, and C.S. Chen. An adaptive learning method for target tracking across multiple cameras. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2008.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2005.

[5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2010.

[6] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Europ. Conf. on Computer Vision*, 2008.

[7] M. Isard and A. Blake. A mixed-state CONDENSATION tracker with automatic model-switching. In *Int. Conf. on Computer Vision*, 1998.

[8] M. Isard and A. Blake. BraMBLe : a Bayesian multiple blob tracker. In *Int. Conf. on Computer Vision*, 2001.

[9] R. Kaucic, A.G. Perera, G. Brooksby, J. Kaufhold, and A. Hoogs. A unified framework for tracking through occlusions and accross sensor gaps. In *Int. Conf. on Computer Vision and Pattern Recognition*, 2005.

[10] H.W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 1955.

[11] C.H. Kuo, C. Huang, and R. Nevatia. Inter-camera association of multi-target tracks by on-line learned appearance affinity models. In *Europ. Conf. on Computer Vision*, 2010.

[12] A. Lev-Tov and Y. Moses. Path recovery of a disappearing target in a large network of cameras. In *Int. Conf. on Distributed Smart Cameras*, 2010.

[13] B. Meden, P. Sayd, and F. Lerasle. Mixed-State Particle Filtering for Simultaneous Tracking and Re-Identification in Non-Overlapping Camera Networks. In *Scandinavian Conference on Image Analysis*, 2011.

[14] K. Okuma, A. Taleghani, N. De Freitas, J. Little, and D. Lowe. A boosted particle filter : multitarget detection and tracking. In *Europ. Conf. on Computer Vision*, 2004.

[15] W. Qu, D. Schonfeld, and M. Mohamed. Distributed bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. *Int. Journal EURASIP*, 2007.

[16] C. Wojek, S. Roth, K. Schindler, and B. Schiele. Monocular 3D scene modeling and inferences : understanding multi-object traffic scenes. In *Europ. Conf. on Computer Vision*, 2010.

[17] W. Zajdel and B. Kröse. A sequential bayesian algorithm for surveillance with nonoverlapping cameras. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 2005.