



**HAL**  
open science

## Tracking-by-reidentification in a non-overlapping fields of view cameras network

Boris Meden, Frédéric Lerasle, Patrick Sayd, Christophe Gabard

► **To cite this version:**

Boris Meden, Frédéric Lerasle, Patrick Sayd, Christophe Gabard. Tracking-by-reidentification in a non-overlapping fields of view cameras network. International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Feb 2012, Rome, Italy. 9p. hal-02024840

**HAL Id: hal-02024840**

**<https://hal.science/hal-02024840>**

Submitted on 19 Feb 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TRACKING-BY-REIDENTIFICATION IN A NON-OVERLAPPING FIELDS OF VIEW CAMERAS NETWORK

Boris Meden<sup>1</sup>, Frédéric Lerasle<sup>2,3</sup>, Patrick Sayd<sup>1</sup> and Christophe Gabard<sup>1</sup>

<sup>1</sup>CEA, LIST, Vision and Content Engineering Laboratory, Point Courier 94, F-91191 Gif-sur-Yvette, France

<sup>2</sup>CNRS ; LAAS ; 7 avenue du Colonel Roche, F-31077 Toulouse Cedex 4, France

<sup>3</sup>Université de Toulouse ; UPS, INSA, INP, ISAE; UT1, UTM, LAAS ; F-31077 Toulouse Cedex 4, France  
{boris.meden, patrick.sayd, christophe.gabard}@cea.fr, lerasle@laas.fr

Keywords: Reidentification, tracking, camera network, non-overlapping fields of view, particle filtering.

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 reidentification algorithm. We propose here to introduce the concept 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 multi-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 reidentification results and the topology of the network. The tracking and reidentification module is first tested with a single camera. We then evaluate the whole framework on a 3 non-overlapping fields of view network with 7 identities. The only a priori knowledge assumed is a topological map of the network.

## 1 INTRODUCTION

This article addresses the problem of pedestrian tracking in large scale environment. Material and economical reasons generally limit the number of cameras thus yielding discontinuities/blind spot in the network field of view. We use the term non overlapping fields of view networks (abbreviated NOFOV networks). Figure 2 provides an example of such a network.

The goal of the tracking module is then to cope with these discontinuities and to still guarantee spatio-temporal consistency. Beyond the image plane tracking, the system should be able to re-identify the targets when it appears in a new camera.

We propose here to integrate the reidentification mixed-state particle filter framework (Meden et al., 2011) in a multi-target tracking-by-detection algorithm (Breitenstein et al., 2010). This allows an online reidentification, embedded in the multi-target tracking process, based on a colorimetric signature of the identities. The second contribution of that paper resides in the addition of a supervision module, working at the network level, that integrates and compares the reidentification results and validates them relatively to

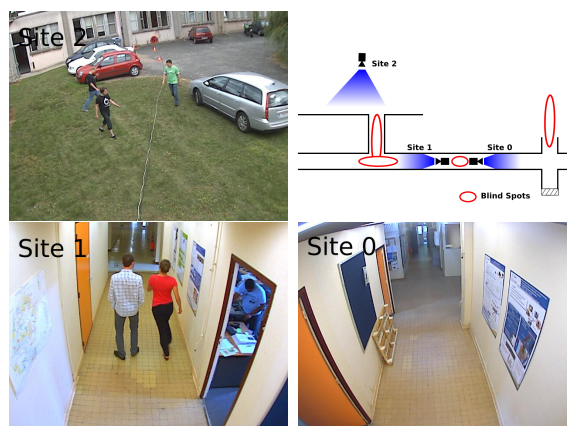


Figure 2: Camera network with non-overlapping fields of views.

the network topology.

Previous works are reviewed in the section 2. Then, we describe the tracking-by-reidentification module, that operates on each camera in the section 3. The supervision module is detailed in 4. Finally, section 5 presents both qualitative and quantitative analysis of the camera level module, and of the addition

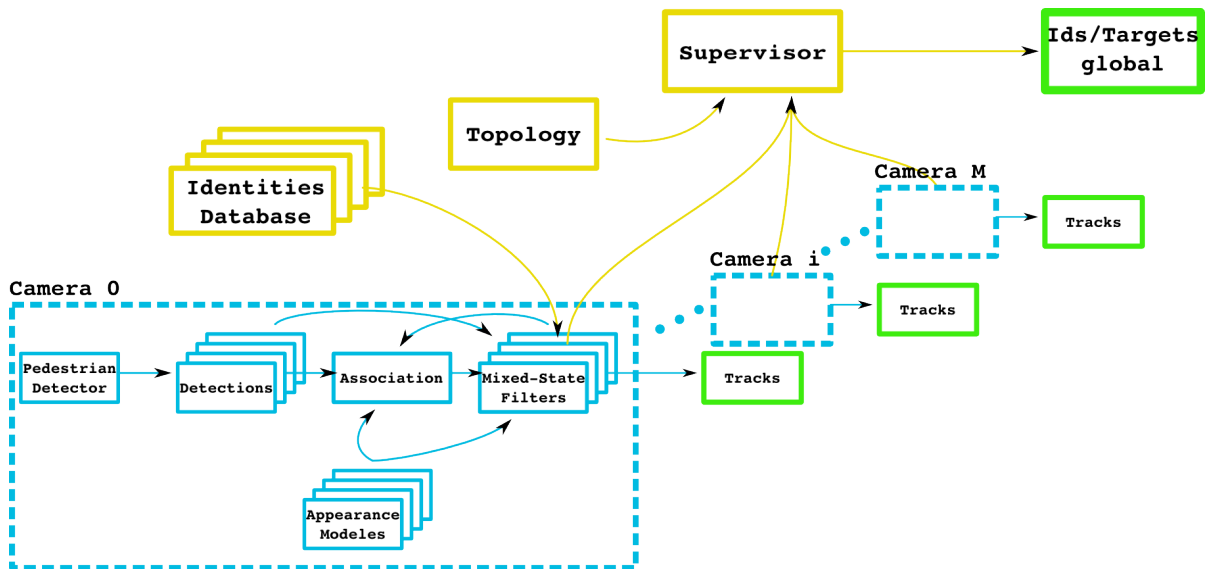


Figure 1: Architecture of our system. Tracking and reidentification in the image are localized at the cameras level, whereas the supervisor works at the network level, confronting identities distributions between them and to the topology.

of topological constraints when applied to a NOFOV network.

Figure 1 summarizes the proposed architecture, with local treatments of the cameras and the centralisation of the reidentification results.

## 2 STATE OF THE ART

Pedestrian reidentification becomes a necessity when targets' trajectories present discontinuities due to the lack of observability. The underlying notion is the one of *global identity* within the network, opposed to the *local identity* of each tracker tracking locally a target during its visibility time in the camera. The quality of a multi-target tracking framework is evaluated by its capacity to keep trackers on the targets they follow while these targets are visible, *i.e.* to keep the same *local identity* (Bernardin and Stiefelhagen, 2008). However this notion of identity is limited to the spatio-temporal continuity of the tracking (after exiting, a re-entering target would receive a new *local identity*). The problem of joint tracking and identification in overlapping fields of view network *e.g.* (Qu et al., 2007) is really similar. The combination of video streams from different sensors comes usually with the calibration of the system, allowing to work in common coordinates, and thus identification is based on the trackers spatio-temporal continuity.

However, a NOFOV network (figure 2) presents discontinuous observations, corresponding to the targets' transit times between the different cameras of

the network or to entry/exit within the same camera. This problem is called pedestrian reidentification, and we introduce here the notion of *global identity* to qualify a target in the network, which will be his/her identity at each of his/her periods of observability in the cameras.

This reidentification problem is classically treated as a request in a database, inspired from web technologies, and put the focus on the pedestrian appearance description to re-identify. Thus, (Gray and Tao, 2008) propose to train a classifier on the invariant parts during a camera change. (Farenzena et al., 2010) adopt the same approach without any learning, proposing a robust fixed signature based on symmetry and asymmetry of the appearance and well positioned colorimetric features. These methods are costly in terms of computation time and are well suited to *a posteriori* treatments.

For a camera network application, the reidentification module should allow real time computation of video streams. Here we target an online update of targets' *global identities*. A similar problem has been tackled by (Chen et al., 2008; Lev-Tov and Moses, 2010; Zajdel and Kröse, 2005). However (Zajdel and Kröse, 2005) suppose to have single pedestrians passing in the network, (Chen et al., 2008) do not report on their tracking process and (Lev-Tov and Moses, 2010) just simulate a NOFOV network and do not work on images. These works do not consider tracking and reidentification jointly, and thus occult the difficulties of multi-target tracking.

Mono-camera multi-target tracking is a largely

tackled problem in the Computer Vision community: our approach is based on different associated assessments. First, particle filtering algorithms’ interest for tracking (CONDENSATION) have been established since the initial work of Isard and Blake in (Isard and Blake, 2001), notably for multiple targets. Then, since (Okuma et al., 2004), *tracking-by-detection* has emerged and particularly the temporal integration of *tracklets*, which robustness has been proven by Kaucic *et al.* in (Kaucic et al., 2005). *Tracklets* optimisation has also been extended to two cameras presenting a disjoint field of view by (Kuo et al., 2010). This method yet does not work online, as the optimisation is conducted on a temporal window.

In opposition to them, our approach places itself in the markovian formalism for the tracking module. Our approach is inspired of (Breitenstein et al., 2010) and (Wojek et al., 2010). Like (Breitenstein et al., 2010), it is based on distributed particle filters enhanced by a reidentification component coming from a discrete identity variable also sampled. They are termed mixed-state particle filters. Then, in the vein of (Wojek et al., 2010), we perform a tracklet temporal integration, but on the identities here, and not for cameras but on the whole network.

### 3 TRACKING-BY-REIDENTIFICATION WITHIN A CAMERA

In this article, we propose an extension to NOFOV networks of the tracking-by-detection algorithm proposed by Breitenstein *et al.* in (Breitenstein et al., 2010), introducing the notion of *global identity* that we seek to retrieve for each target. We present in this section our implementation of (Breitenstein et al., 2010) and how the use of mixed-state particle filtering for reidentification (Meden et al., 2011) comes to extend that approach.

#### 3.1 Targets Description

##### 3.1.1 Global Identities Learning



Figure 3: Key-frames of each identity for the NOFOV network sequence (issued from camera 1).

Each reidentification algorithm needs a first view before allowing any reidentification. Here, we assume that such a database is acquired offline. To do so, we extract a collection of key-frames from one of the cameras (*e.g.* positioned in the entrance hall of the building to monitor), and we use these as description of our *global identities*. The choice of the key-frames is done with K-means on tracking sequences from the chosen camera as detailed in (Meden et al., 2011). Thus, these key-frames encode the variability of the identity during its first tracking. Figure 3 presents the identity database used for the network of figure 2, learned in camera 1.

##### 3.1.2 Target Appearance Modelling

We use the same appearance model as depicted in (Meden et al., 2011) to describe the targets and their identities in the database: horizontal stripes of color distributions, computed in the RGB space. The similarity between two descriptors is the Bhattacharyya distances between corresponding stripes, normalized by a gaussian kernel. This allows us to compute similarities to the appearance model of a tracker, and also to the key-frames of an identity in the database, respectively noted  $w_{App}(\cdot)$  and  $w_{Id}(\cdot)$ .

#### 3.2 Detections Integration

##### 3.2.1 Association to Detections

Our approach favor a tracking-by-detection strategy via the classical HOG detector proposed by Dalal and Triggs in (Dalal and Triggs, 2005). These detections are integrated in the tracking process by a greedy association stage. After that association, each tracker has potentially received a detection which will be used to update the particles. To do so, an association matrix is built between trackers and detections. The score of pair detection  $d$  vs. tracker  $tr$  given by equation (1), involves:

- the distance between the tracker’s particles and the detection, evaluated under a gaussian kernel  $p_{\mathcal{X}}(\cdot) \sim \mathcal{N}(\cdot, \sigma^2)$  ;
- the tracker’s box area  $\mathcal{A}(tr)$  relatively to the detection’s one also evaluated under a gaussian kernel ;
- the tracker’s appearance model evaluated on the detection ( $w_{App}(\cdot)$ ).

$$S(d, tr) = \underbrace{\sum_{p \in tr}^N p_{\mathcal{X}}(d-p)}_{\text{euclidean distance}} \times \underbrace{p_{\mathcal{X}}\left(\frac{|\mathcal{A}(tr) - \mathcal{A}(d)|}{\mathcal{A}(tr)}\right)}_{\text{relative size}} \times \underbrace{w_{App}(d, tr)}_{\text{appearance model}} \quad (1)$$



Thus, tracker and detection should present simultaneously a similar position, a similar size and a similar colorimetric response. Once the matrix is built, maxima are extracted on greed manner, suppressing lines and columns after affectation. The process is iterated till the pairing threshold is reached. Such a heuristic is preferred to the optimal method, the Hungarian Method (Kuhn, 1955), unsuited for its complexity.

### 3.2.2 Automatic Tracker Initialisations / Terminations

Every temporally recurrent detection, which is not associated to any tracker, yields the instantiation of a new tracker. On a similar manner, every tracker which has not been associated with a detection for a time period longer than the suppression threshold is stopped.

## 3.3 Particle Filtering

### 3.3.1 Mixed-State Prediction Model

Each target initialized on a detection is tracked by a particle filter. Given the identity database, we have extra reference descriptors to compare with. To do so, following (Meden et al., 2011), we use Mixed-State CONDENSATION filters, introduced in (Isard and Blake, 1998). We aim to estimate a mixte state vector, composed of several continuous terms and a discrete one.

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

The continuous part of the state  $\mathbf{x} = [x, y, v_x, v_y]^\top$  is composed of the position in the image plane  $(x, y)^\top$  and of the speed vector  $(v_x, v_y)^\top$ . The integer part  $id$  refers to one of the  $N_{id}$  identities in the database. The tracking is conducted in the image plane, and tracking box dimension is updated on the associated detections. The appearance model is also updated on the associated detection. Given this extended state vector, the density of sampling process at image  $t$  can be decomposed (Isard and Blake, 1998):

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

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

where  $T_{ij}(\mathbf{x}_{t-1})$  is the transition probability from identity  $i$  to  $j$ , applied to the discrete identity parameter, and  $p_{ij}(\mathbf{x}_t | \mathbf{x}_{t-1})$  is the sampling applied to the continuous part. The transition matrix  $T = [T_{ij}]$  is built over the set of key-frames. The element  $T_{ij}$  is the similarity  $w_{id}(\cdot)$  between identities  $i$  and  $j$  of

the database, computed between the most different key-frames. Particles are propagated according to a first order motion model:

$$p_{ij}(\mathbf{x}_t | \mathbf{x}_{t-1}) :$$

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

where the noises  $\boldsymbol{\varepsilon}_{(x,y)}$  and  $\boldsymbol{\varepsilon}_{(v_x, v_y)}$  are drawn from normal distribution and where  $\Delta t$  is the time interval between two images.

### 3.3.2 Observation Model Integrating Detections

The weight  $w_{tr}^{(p)}$  associated with the  $p$ -th particle of tracker  $tr$  is computed integrating the distance to the associated detection  $d^*$ , the colorimetric similarity to the appearance model  $w_{App}(\cdot)$  and the colorimetric similarity to the identity of the particle  $w_{Id}(\cdot)$ .  $Id(p)$  represents the identity taken by particle  $p$ . This is the discrete parameter of  $p$ .

$$w_{tr}^{(p)} = \underbrace{\alpha \cdot I(tr) \cdot p_{\mathcal{X}}(d^* - p)}_{\text{distance to the detection}} + \underbrace{\beta \cdot w_{App}(d, tr)}_{\text{appearance model}} + \underbrace{\gamma \cdot w_{Id}(d, id(p))}_{\text{identity}} \quad (2)$$

where  $\alpha$ ,  $\beta$  and  $\gamma$  are weighting coefficients, and  $I(tr)$  is a boolean signifying the existence or not of an associated detection to the tracker. As in (Meden et al., 2011), the introduction of similarity relative to the identity in the particle weighting drives the particle cloud towards the most likely identities given the received observations. In that way, each tracker maintains a discrete distribution over the *global identities*, the modes of that distribution being the most likely identities.

The state estimation is a two-stage process. First we compute the Maximum A Posteriori over the discrete parameter relatively to the current observation  $\mathbf{Z}_t$  with equation (3), *i.e.* the most likely identity at time step  $t$ .

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

$$\text{where } \Upsilon_j = \{p | \mathbf{X}_t^{(p)} = (\mathbf{x}_t^{(p)}, j)\}$$

Then, the continuous components are estimated over the subset of particles  $\hat{\Upsilon}$  which have that most likely identity, following equation (4).

$$\hat{\mathbf{x}}_t = \sum_{p \in \hat{\Upsilon}} w_{tr}^{(p)}(t) \cdot \mathbf{x}_t^{(p)} / \sum_{p \in \hat{\Upsilon}} w_{tr}^{(p)}(t), \quad (4)$$

$$\text{where } \hat{\Upsilon} = \{p | \mathbf{X}_t^{(p)} = (\mathbf{x}_t^{(p)}, \hat{id}_t)^\top\}$$

That way, on top of target image position estimation, each filter provides a discrete identity distribution for its target.

## 4 TOPOLOGICO-TEMPORAL SUPERVISION OF THE REIDENTIFICATIONS

Section 3 has presented a reidentification strategy integrated to the image plane tracking. That strategy has been established as superior to an exhaustive comparison to the database by (Meden et al., 2011). Its limitation resides in the distributed aspect of the mixed-state filters. Indeed, the probability densities over the target identity are independent from one filter to another. Thus, two filters may produce the same identity at the same time for two different targets. We wish here to constrain the process, so that it produces exclusive trackers/identities pairing. This is done at the network level.

### 4.1 Identities Tracklets Generation

#### 4.1.1 Using the Topology

In this part, we suppose to have access to the topology of the network we monitor. This topology is represented by a non-oriented graph  $G = (V, E)$ , which vertices  $V$  represent entry/exit areas in the cameras, and which edges  $E$  give existing transitions between these areas, as illustrated in figure 4.

This fixed *a priori* here, could be learned online with methods such as (Chen et al., 2008).

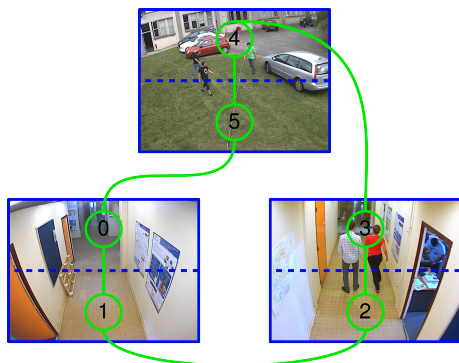


Figure 4: Network topology modelling. A non-oriented graph link entry/exit areas of the adjacent cameras.

#### 4.1.2 Temporal Integration

Each tracker produces at each time step a discrete probability distribution over the set of identities, computed as the ratio of particles dedicated to one identity. These probabilities are aggregated over a time window in a Dynamic Programming manner. Following (Wojek et al., 2010), we speak here of track-

lets over the identities. Doing so, we build an association matrix between trackers and identities using equation (5).

The use of the network topology comes in at that point. It is used to suppress the impossible tracker/identity associations. We start from an initial localization of the identities in the network. At every termination of a tracker, this localization is updated with its reidentification. We use that localization to set to null the associations that violate this localization. An association is violating it if the tracker's area is not connected to the last localization of the proposed identity.

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

$$\circ \Upsilon_{tr, id_t} = \{p | \mathbf{X}_t^{(p)} = (\mathbf{x}_t^{(p)}, id_t)^T\}$$

and where

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

#### 4.1.3 Association Exclusivity

A greedy exclusive association similar to the one described in section 3, is performed. It works using the similarity function (5) and produces an exclusive association tracker/identity at the end of the time window. The topology, and the preceding reidentification come to suppress impossible configuration. Finally, the association forces exclusivity in the pairing.

The management of the *global identities* at the core of the tracking process allows us to skip combinatorial problems inherent when handling multiple targets and to maintain up-to-date the positions of these identities in the network.

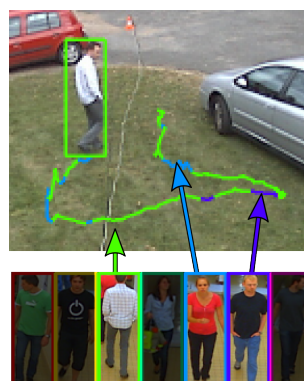


Figure 5: Different identity tracklets during a tracking sequence. (better viewed in colors)

## 4.2 Tracklets Optimisation over a Tracking Sequence

These supervised affectations come at the end of each time window, and give the re-identification during the next time window. We obtain here short period re-identifications, which we call identity tracklets. Figure 5 presents different tracklets of identities inferred by the supervisor for a tracking sequence. Each color refers to an identity in the database.

To avoid a reidentification process biased on the beginning of the tracking sequence, we settle the identity distribution in the mixed-state filters back to equiprobability at the end of each time window. That way, the mixed-state reidentification filter explores again each identity and converge towards the most likely, relatively to the observations it receives.

For each active tracker, these reidentifications are binned into an histogram indexed over the identities. Following Dynamic Programming principles, the current affectation trackers / identities displayed is the best solution found so far, *i.e.* the strongest mode of that histogram. In the same way, when a tracker is stopped, the re-identification affected is the strongest mode, and the localization in the topological graph of that identity is updated.

## 5 IMPLEMENTATION AND ASSOCIATED EVALUATIONS

### 5.1 Implementation

Our IP network has an average framerate of 16 images per second. We thus fix  $\Delta t = 1/16s$  in the evolution model of the particles filters. In the observation model, equation (2), we fix empirically:

$$\begin{cases} \alpha = 0.90, \beta = 0.05 \text{ and } \gamma = 0.05 & \text{if } I(tr) = 1 \\ \text{(Breitenstein et al., 2010)} \\ \alpha = 0.0, \beta = 0.8 \text{ and } \gamma = 0.2 & \text{else,} \\ \text{(Meden et al., 2011).} \end{cases}$$

In the supervisor, the time window length is set to 7 images, which correspond to the average convergence time of the mixed-state filters towards their identity.

### 5.2 Evaluations

#### 5.2.1 Datasets

We evaluate the different component of our approach on two different datasets. First, we test the tracking module without and with reidentification activated on

the sequence PETS'09 S2L1<sup>1</sup>. This public dataset, composed of 795 frames, presents an open outdoor area, where 10 pedestrians wander, with crossings and enterings/exitings. Having labeled these data, we are able to quantify the quality of our tracking algorithm.

Considering the lack of public datasets in terms of NOFOV network, we evaluate the supervision part on a private sequence which we call the NOFOVNetwork sequence in the sequel. It presents a total of 7 pedestrians wandering between 3 cameras. There is no overlapping between the cameras field of view. Two of them are placed in a building corridor, the third one monitoring an outside area with a configuration similar to PETS'09. The dataset has 837 frames. We plan to release publicly these data.

#### 5.2.2 Metrics

We use the CLEAR MOT (Bernardin and Stiefelagen, 2008) to quantify tracking results. We obtain a precision score MOTP (Multiple-Object Tracking Precision) computed as the intersection over the union between tracking boxes and ground truth, and an accuracy score MOTA (Multiple-Object Tracking Accuracy) taking into account false positives, false negatives and switching trackers between targets.

Moreover, we evaluate the reidentification performances by a True Reidentification Rate (TRR), computed as the ratio of correct reidentification over the number of trials. Given that the supervisor operates on a time window, these TRRs are updated only at the end of these time windows.

### 5.3 Camera Level Performances

#### 5.3.1 Global Identity Notion



Figure 6: Key-frames of the 10 identities in the PETS sequence.

Figure 6 gives an overview of the identity database we used on the sequence PETS'09. Here there is only one camera. Thus, the database images are taken from the test sequence. The goal here is to illustrate the tracking-by-reidentification, compare its performances without reidentification and highlights the new modality it offers.

Figures 7 and 8 illustrate the limitation of handling only *local identities* when targets exit and then

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

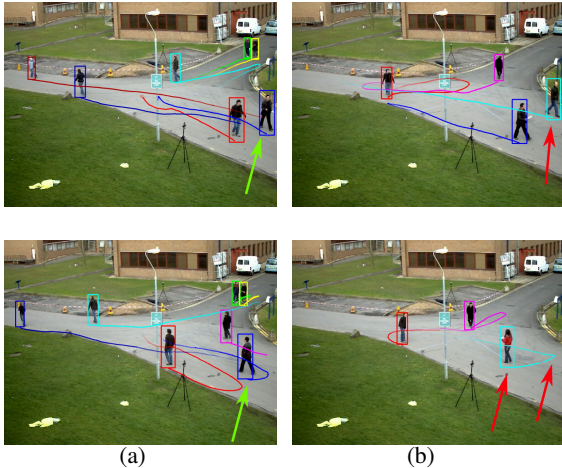


Figure 7: Results from (Breitenstein et al., 2010) (a) Between frames 204 and 241 the highlighted person exit and enter again the scene. The new tracker is matched with the old trajectory on a spatial criterion. (b) A similar situation happens between images 390 and 445. However that time, it is a different person who enters. The new tracker is matched again with the previous trajectory. Here this is a reidentification failure as the person is different.

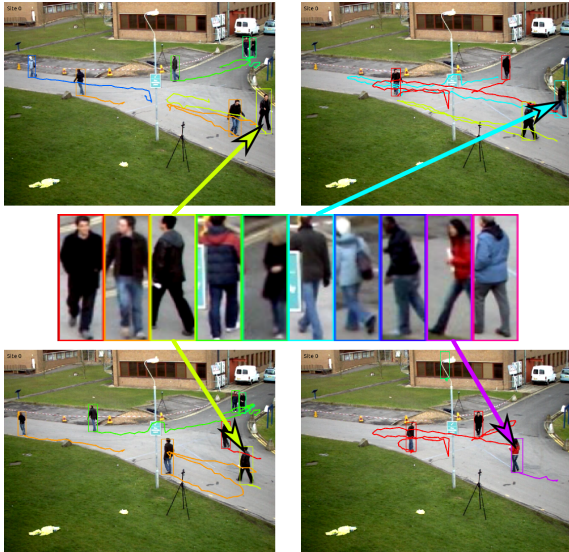


Figure 8: Interest of using reidentification embedded in the multi-target tracking: in (a) as like as in (b), the system reidentifies the target relatively to the identity database, and detects that in (b) the person entering is not the same as the exiting one.

re-enter the scene. On figure 7, when the person exits and a new one enters, the trajectory of the preceding is matched with the new one. (Breitenstein et al., 2010) just use a simple spatial criterion, no reidentification is involved in that matching. It works on figure 7(a) as the person is the same, but not on figure 7(b) as the person is different.

In our case (figure 8), at each time step, each tracker provides a probability distribution for the observed identity. This allows us to tolerate periods without observations like exits from the camera. When the target re-enter, the tracker searches again the correct identity.

### 5.3.2 Quantitative Analysis

Table 1 presents quantitative results on the PETS'09 sequence. First, we validate our partial implementation of (Breitenstein et al., 2010) (without HOG + ISM detector, detector confidence use in the observation model, and Boosting Online based appearance model).

However, our approach presents an extra modality with the notion of *global identity*. We show first that the introduction of mixed-state particle filtering does not decrease much tracking performances. To do so, we compare MOTP and MOTA for our implementation without and with the reidentification module activated. Then, this extra modality allow us to compute TRR for the sequence. Finally, we compare the reidentification results of the distributed mixed-state filters alone against the supervised ones. There, exclusivity constraints (section 4) yield improved results.

The stochastic aspect of particle filtering has been taken into account in our experiences: table 1 shows results averaged over ten repetitions of tracking.

Table 1: CLEAR MOT metrics tracking results (Bernardin and Stiefelhagen, 2008) and true reidentification rates on the monocamera sequence PETS'09 S2L1. We give here Multi-Object Tracking Precision (MOTP), Multi-Object Tracking Accuracy (MOTA), and True Reidentification Rate (TRR) defined in section 5.2.

Sequence PETS'09	MOTP	MOTA	TRR
Tracking-by-detection (Breitenstein et al., 2010)	<b>56.3%</b>	<b>79.7%</b>	-
Tracking-by-detection implemented	42.7%	77.9%	-
Tracking-by-Reidentification	42.5%	77.7%	59.7%
Tracking-by-Reidentification supervised	42.4%	75.9%	<b>64%</b>

## 5.4 Supervisor Performances

The NOFOVNetwork sequence being not annotated for the tracking, we just present true reidentification rates for that sequence. We compare here the method based only on colorimetric information and particle filtering inspired by (Meden et al., 2011), with the supervised system we propose in section 4 which includes exclusivity and topological constraints.



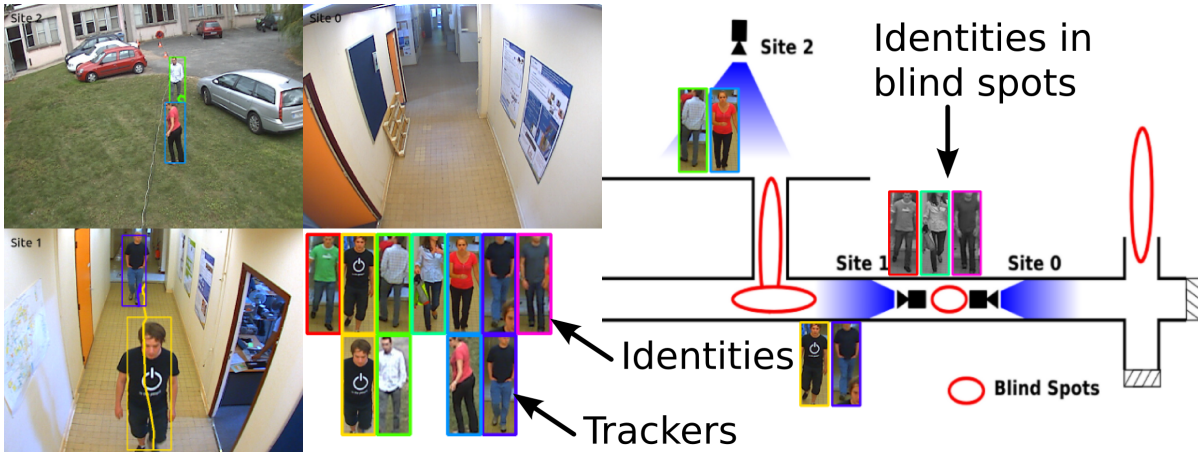


Figure 9: Network tracking example: output of our system with camera tracks matched to *global identities* (left), allowing to localize them in the topology (right).

Table 2 presents true reidentification rates per camera, and for the whole network. The database being built with descriptors taken from camera 1, this explains better TRR in that camera. These results illustrate the contribution of the supervisor. Each correctly re-identified target constrains the system in the sequel through the topology.

Table 2: True Reidentification Rates for each camera of the sequence NOFOVNetwork: comparison of the approaches without, and with supervisor on the network.

NOFOV Sequence	cam0	cam1	cam2	network
Tracking-by-Reidentification	43.7%	67.3%	55.5%	54.6%
Tracking-by-Reidentification supervised	<b>67.7%</b>	<b>76.9%</b>	<b>63.8%</b>	<b>68.2%</b>

Finally, figure 9 gives an overview of our system output. Left, the cameras of the network display current tracks, and right, the identities are localized in the topology.

## 6 CONCLUSION

This article deals with non-overlapping fields of view cameras networks monitoring, aiming at localizing the targets in the topology. This is achieved through the concept of *global identity*. We present here a two stage tracking-by-reidentification method, based respectively on colorimetric signatures and spatio-temporal constraints in the network.

The camera level is treated by a markovian tracking-by-detection inspired by (Breitenstein et al., 2010), enhanced by the concept of *global identity* taken into account in the mixed-state particle filter framework. Thus, each tracker builds a discrete iden-

tity distribution for its target. Doing so, it integrates a re-initialisation capacity after the target's exit.

These identity distributions, considered as tracklets over the identities are filtered spatio-temporally by a supervisor. This one forces exclusivity between reidentifications and insure consistency regarding the network topology.

A first extension resides in the database online learning and updating to achieve a fully automatic system. Further work will also investigate a more enhanced appearance model, *e.g.* trained online on its target. Finally, additional knowledge about the scene (*e.g.*, a ground plane to improve targets size estimation) would be beneficial.

## REFERENCES

- Bernardin, K. and Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: the clear mot metrics. *Journal on Image and Video Processing*.
- Breitenstein, M., Reichlin, F., Leibe, B., Koller-Meier, E., and Van Gool, L. (2010). Online multi-person tracking-by-detection from a single, uncalibrated camera. *Pattern Analysis and Machine Intelligence*.
- Chen, K., Lai, C., Hung, Y., and Chen, C. (2008). An adaptive learning method for target tracking across multiple cameras. In *Int. Conf. on Computer Vision and Pattern Recognition*.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Int. Conf. on Computer Vision and Pattern Recognition*.
- Farenzena, M., Bazzani, L., Perina, A., Murino, V., and Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. In *Int. Conf. on Computer Vision and Pattern Recognition*.
- Gray, D. and Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Europ. Conf. on Computer Vision*.
- Isard, M. and Blake, A. (1998). A mixed-state CONDENSATION tracker with automatic model-switching. In *Int. Conf. on Computer Vision*.
- Isard, M. and Blake, A. (2001). BraMBLe: a Bayesian multiple blob tracker. In *Int. Conf. on Computer Vision*.
- Kaucic, R., Perera, A., Brooksby, G., Kaufhold, J., and Hoogs, A. (2005). A unified framework for tracking through occlusions and accross sensor gaps. In *Int. Conf. on Computer Vision and Pattern Recognition*.
- Kuhn, H. (1955). The hungarian method for the assignment problem. *Naval research logistics quarterly*.
- Kuo, C., Huang, C., and Nevatia, R. (2010). Inter-camera association of multi-target tracks by on-line learned appearance affinity models. In *Europ. Conf. on Computer Vision*.
- Lev-Tov, A. and Moses, Y. (2010). Path recovery of a disappearing target in a large network of cameras. In *Int. Conf. on Distributed Smart Cameras*.
- Meden, B., Sayd, P., and Lerasle, F. (2011). Mixed-State Particle Filtering for Simultaneous Tracking and Re-Identification in Non-Overlapping Camera Networks. In *Scandinavian Conference on Image Analysis*.
- Okuma, K., Taleghani, A., De Freitas, N., Little, J., and Lowe, D. (2004). A boosted particle filter: multitarget detection and tracking. In *Europ. Conf. on Computer Vision*.
- Qu, W., Schonfeld, D., and Mohamed, M. (2007). Distributed bayesian multiple-target tracking in crowded environments using multiple collaborative cameras. *Int. Journal EURASIP*.
- Wojek, C., Roth, S., Schindler, K., and Schiele, B. (2010). Monocular 3D scene modeling and inferences: understanding multi-object traffic scenes. In *Europ. Conf. on Computer Vision*.
- Zajdel, W. and Kröse, B. (2005). A sequential bayesian algorithm for surveillance with nonoverlapping cameras. *Int. Journal of Pattern Recognition and Artificial Intelligence*.