MCMC Supervision for People Reidentification in Nonoverlapping Cameras

Boris Meden[1]
boris.meden@cea.fr
Frédéric Lerasle[2]
lerasle@laas.fr
Patrick Sayd[1]
patrick.sayd@cea.fr

[1] CEA, LIST,
Laboratoire Vision et Ingénierie des Contenus,
BP 94, F-91191 Gif-sur-Yvette, France

[2] CNRS ; LAAS ;
Université de Toulouse ; UPS, LAAS ;
F-31077 Toulouse Cedex 4, France

We present a pedestrian tracking system that uses re-identification to monitor non-overlapping cameras. As tracking, re-identification is an assignment problem, the difficulties being to generate an accurate representation and to prune unlikely pairings. The assignments are realised in two stages.

First, a Markovian multi-target tracking-by-detection framework which includes identification in the search space is run in the cameras, following the approach of [1]. Inspired by [2], the appearance model used is composed of horizontal stripes of HSV histograms weighted by their distances to the symmetry axis. The use of topology allows to instantiate new identities from the feeding areas in an identity database, which we compare with to perform re-identification. The mixed-state formalism [3] uses that database and samples in this identity space. That way the tracker produce a tracklet and re-identification probabilities in the database representing the belief of the tracker. The resulting tracklets are sent to the supervisor along with their probabilities of identity, their time of existence and their areas.

At the network level, the supervisor resorts to deferred logic to optimize the assignment between the received tracklets using re-identification distributions and network topology information. The combinatorial space is efficiently explored through MCMC sampling. Tracks output by the supervisor are optimized to represent the activity of the same person, as shown in figure (1).

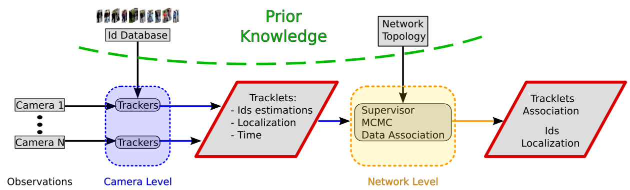

Figure 1: Synoptic diagram of the combination between markovian tracking-by-reidentification and MCMC data association at the network level.

Camera Level: Mixed-state Tracking by Reidentification 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 $w_{App}(.)$ and the colorimetric similarity to the identity of the particle $w_{Id}(.)$. $Id(p)$ represents the identity taken by particle p. This is the discrete parameter of p.

$$w_{tr}^{(p)} = \underbrace{\alpha \cdot \mathcal{I}(tr) \cdot p_{\mathcal{N}}(d^* - p)}_{\text{distance to the detection}} + \underbrace{\beta \cdot w_{App}(d, tr)}_{\text{appearance model}} + \underbrace{\gamma \cdot w_{Id}(d, id(p))}_{\text{identity}} \quad (1)$$

where α, β and γ are weighting coefficients empirically set, and $\mathcal{I}(tr)$ is a boolean signifying the existence or not of an associated detection to the tracker.

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 (2), *i.e.* the most likely identity at time step t.

$$\hat{id}_t = \arg\max_j P(id_t = j|\mathbf{Z}_t) = \arg\max_j \sum_{p \in \Upsilon_j} w_{tr}^{(p)}(t), \quad (2)$$

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

Then, the continuous components are estimated over the subset of particles $\hat{\Upsilon}$ which have that most likely identity, following equation (3).

$$\hat{\mathbf{x}}_t = \sum_{p \in \hat{\Upsilon}} w_{tr}^{(p)}(t) \cdot \mathbf{x}_t^{(p)} / \sum_{p \in \hat{\Upsilon}} w_{tr}^{(p)}(t), \quad (3)$$

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

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

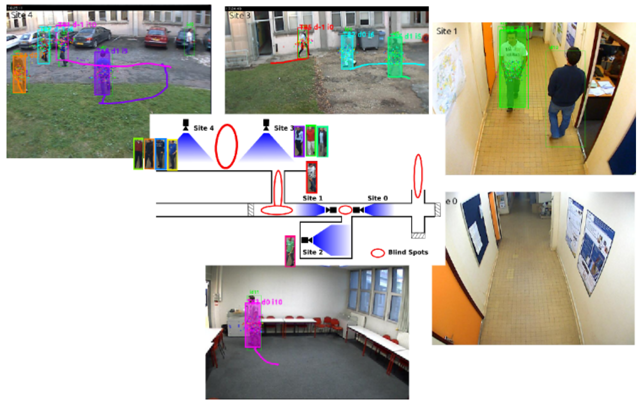

Figure 2: Overview of the monitored network.

Network Level: Topologic and Appearance Driven MCMC Optimisation The likelihood a collection of tracklets τ_i have to be associated to an identity id mixes topology and identity distributions: $p(Y|H) = \mathcal{P}_{Topo}(Y|H) \cdot \mathcal{P}_{MSR}(Y|H)$, with:

$$\mathcal{P}_{Topo}(Y|H) = \prod_{i=1}^{|\tau_n|-1} p_{\mathcal{N}}(d_{topo}(a_{i-1}^{out}, a_i^{in})), \quad (4)$$

where $d_{topo}(.)$ is the distance between two nodes of the topological graph, $a_i^{in/out}$ are the area of beginning (*resp.* ending) of the i-th tracklet, $p_{\mathcal{N}}(.)$ is a gaussian kernel to transform the distance into a similarity between 0 and 1 and $|\tau_n|$ is the cardinal of the tracklet set τ_n.

As comparing directly descriptors taken from different cameras yields an homogeneity problem, we use instead the mixed-state trackers belief on the tracklet identity, resulting from online comparison with the database:

$$\mathcal{P}_{MSR}(Y|H) = \prod_{i=1}^{|\tau_n|-1} ids_i(id), \quad (5)$$

where ids_i is the discrete probability distribution over the identity database for the i-th tracklet. That way $ids_i(id)$ represents the probability that tracklet i has the identity id.

We use MCMC to optimise the tracklet-to-identities assignment $H = \{\tau_i\}_{i=1...N}$ The tracking results obtained on a large ground-truthed dataset demonstrate the effectiveness of the approach. Figure 2 provides an example of the method running on 5 cameras.

[1] 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. *PAMI*, 2010.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[3] B. Meden, P. Sayd, and F. Lerasle. Mixed-State Particle Filtering for Simultaneous Tracking and Re-Identification in Non-Overlapping Camera Networks. In *SCIA*, 2011.