

Optimizing people tracking for a video-camera network

Optimisation du suivi de personnes dans un réseau de caméras

Julien Badie

November 17, 2015

Supervisor: François Brémond



Région
Provence
Alpes
Côte d'Azur

Contents

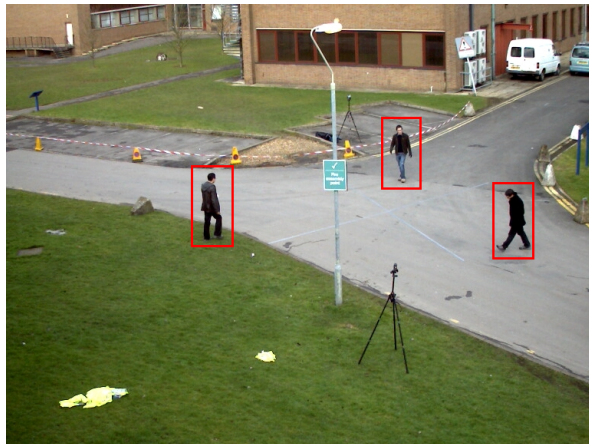
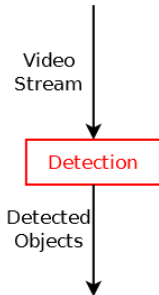
- 1 Introduction
- 2 Global Tracker
- 3 Online evaluation of tracking results
- 4 Tracklet matching over time
- 5 Conclusion and future works

Section 1

Introduction

Scene understanding: detection

Step 1 - Detection: where are the people?



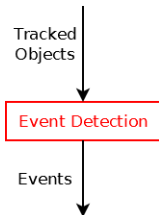
Scene understanding: tracking

Step 2 - Tracking: where are the people going?



Scene understanding: event recognition

Step 3 - Event recognition: what are the people doing?



Challenges



Commercial vision-based systems

- Average performance
- Relying on human operators
- Different kinds of contexts, scenarios and applications

Main issues with state of the art tracking algorithms

Challenges

- **Parameter tuning**
- **Adapt tracking to new situations:** detection and tracking of one specific type of object in one specific type of environment
- **Error management:** difficulties to detect and manage erroneous input data

Tracking algorithms: state-of-the-art

Trackers	Approaches	Limitations
Multi-target tracking by on-line learned discriminative appearance models [Kuo et al., CVPR 2010]	Online learning of people appearance	Very sensitive to detection errors
Multi-features tracker [Chau et al., ICDP 2011]	Computes context and tunes parameters according to its changes	Heavy offline learning - Sensitive to detection errors
Detection- and Trajectory-Level Exclusion [Milan et al., CVPR 2013]	Conditional random field	Does not handle long-term occlusions
Robust Online Multi-Object Tracking [Bae et al., CVPR 2014]	Tracklet confidence computation and online discriminative appearance learning	Requires confidence data from detection - Important number of ID switches compared to the state-of-the-art
Target Identity-aware Network Flow for Online Multiple Target Tracking [Dehgha et al., CVPR 2015]	Discriminative learning and global data association	Requires manual annotation for initialization

Global tracking algorithms: state-of-the-art

Trackers	Approaches	Limitations
Hybridboosted multi-target tracker [Li et al., CVPR 2009]	Trajectories association + AdaBoost variant	Offline process
GMCP-tracker [Zamir et al, ECCV 2012]	Computes cliques on a graph of detections	Very slow (4.4s/frame) with a good setup (4 cores) - Complexity grows exponentially with the number of detections
Multiple Object Tracking by efficient Graph Partitioning [Kumar et al., ACCV 2014]	Graph partitioning	Does not handle wrong tracklets (mixing two people)
CRF-based Multi-Person Tracking [Heili et al., Transactions on Image Processing 2014]	Conditional random field	Does not handle long-term occlusions
GMMCP Tracker [Dehghan et al., CVPR 2015]	Computes cliques on a graph of detections + models occlusions and missed detections	4 empirically defined parameters - Complexity grows exponentially with the number of detections

Section 2

Global Tracker

Definitions

Definitions

- **Context:** all elements that can influence the vision-based system such as illumination, indoor/outdoor scene, entrance/leaving areas, background obstacles, etc.
- **Error:** significant difference between a result given by the detection or tracking algorithm and the Ground-Truth. The border between an error and an acceptable result is defined by a metric.
- **Anomaly:** variation of feature larger than usual; considered as a potential error.
- **Tracklet:** segment of trajectory representing one tracked object. Result given by the tracking algorithm.

Assumptions

Assumptions

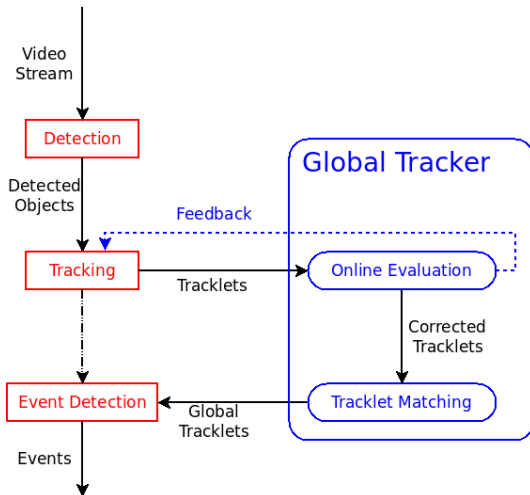
- **Fixed cameras:** the vision-based system is composed of one or multiple overlapping or non-overlapping fixed cameras.
- **Limited prior knowledge:** prior knowledge about scene/context is not required but can be used if available.
- **No Ground-Truth available:** due to online constraint.
- **Near real-time**

Approach

Contributions

- **Tracking quality estimation:**
 - Estimate the quality of the tracking algorithm by analyzing tracking results (tracklets) without using Ground-Truth.
 - Identify potential errors (anomalies) and classify them (real errors, natural phenomena)
- **Tracking results improvement:**
 - Correct the errors, either by repairing them directly or by sending feedback to detection or tracking modules.
 - Improve overall tracking results by merging segments of trajectory representing the same object.

Global Tracker framework overview

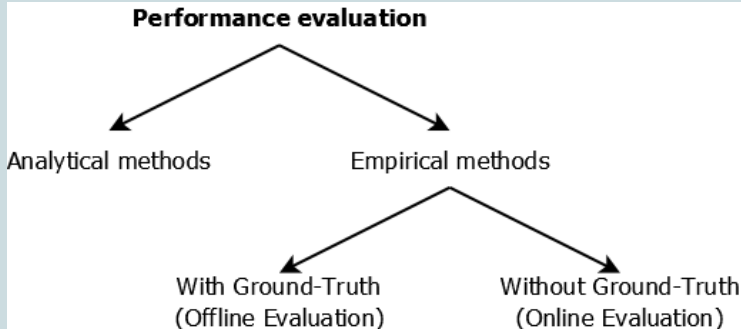


Section 3

Online evaluation of tracking results

Different methods of tracking performance evaluation

Types of performance evaluation¹



¹E. Maggio and A. Cavallaro. Video tracking: Theory and practice. Wiley, 2010.

Empirical methods with Ground-Truth

Required data

- Objects tracked by algorithm
- Ground-Truth defined by human
- Metrics (VACE¹, CLEAR², trajectory-based³, ETISEO⁴, ...)

¹R. Kasturi, D. Goldgof and P. Soundararajan. PAMI 2009.

²K. Bernardin and R. Stiefelhagen. EURASIP Journal on Image and Video Processing 2008.

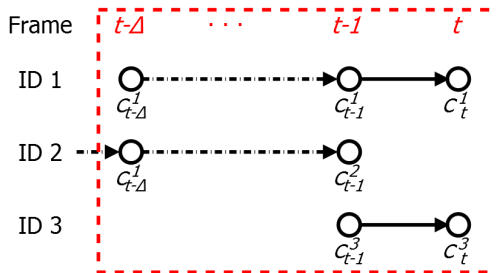
³B Wu and R Nevatia. CVPR 2006.

⁴A. T. Nghiem, F. Brémond, M. Thonnat and V. Valentin. AVSS 2007.

Metrics

Projects	Metrics
VACE	Sequence Frame Detection Accuracy (SFDA)
	Average Tracking Accuracy (ATA)
CLEAR	Multiple Object Tracking Accuracy (MOTA)
	Multiple Object Tracking Precision (MOTP)
Trajectory-based	Mostly Tracked (MT)
	Partially Tracked (PT)
	Mostly Lost (ML)
ETISEO	Tracking time
	ID persistence
	ID confusion




Tracking result: tracklets



- time window of size $[t - \Delta, t]$
- each tracklet is defined on an interval $[T_{start}^i, T_{end}^i]$
- one object detected on one frame corresponds to one node C_t^i
- each node contains a pool of features \mathcal{F}_t^i

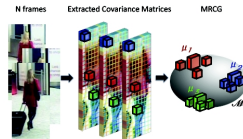
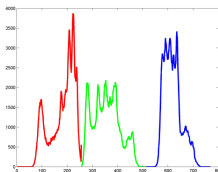
Feature pool

$$\mathcal{F}_t^i = \{\mathcal{F}_t^{O,i}, \mathcal{F}_t^{OO,i}, \mathcal{F}_t^{OE,i}\}$$

$\mathcal{F}_t^{O,i}$	Object alone	
$\mathcal{F}_t^{OO,i}$	Object vs Object	
$\mathcal{F}_t^{OE,i}$	Object vs Environment	

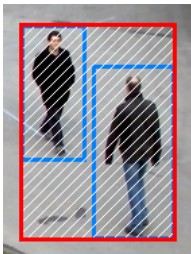
Feature pool - Object alone

Feature pool	Feature description
\mathcal{F}^O	bounding box dimension
	trajectory (direction + speed)
	color histogram
	covariance matrices



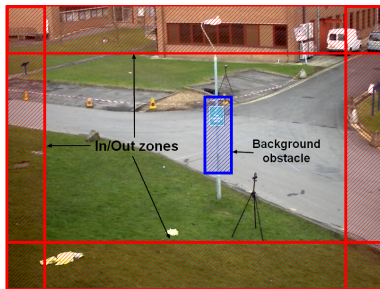
Feature pool - Object vs Object

Feature pool	Feature description
\mathcal{F}^{OO}	density with other objects
	spatial overlap level with other objects
	frame-to-frame overlap with other objects



Feature pool - Object vs Environment

Feature pool	Feature description
\mathcal{F}^{OE}	object appearing/disappearing in zone
	overlap level with background elements



Tracklet quality computation using \mathcal{F}^0

Weighted mean

$$\mu(f^i) = \frac{\sum_{t=T_{start}^i}^{T_{end}^i} w(t) * f_t^i}{\sum_{t=T_{start}^i}^{T_{end}^i} w(t)}$$

f_t^i : feature value at frame t
from $\mathcal{F}_t^{O,i}$

w : weight function (linear or exponential)

Weighted standard deviation

$$\sigma(f^i) = \sqrt{\frac{\sum_{t=T_{start}^i}^{T_{end}^i} w(t) * (f_t^i - \mu(f^i))^2}{\sum_{t=T_{start}^i}^{T_{end}^i} w(t)}}$$

Tracklet coherency computation

Coefficient of variation

$$c(f^i) = \frac{\sigma(f^i)}{\mu(f^i)}$$

Tracklet coherency

$$\delta_t^i = \left| 1 - \frac{c(f^i)_t}{c(f^i)_{t-1}} \right|$$

- $\delta_t^i \in [0, \epsilon] \implies$ no anomaly detected
- $\delta_t^i \in [1 - \epsilon, 1] \implies$ anomaly detected

Anomaly classification

Anomaly classification is performed using the remaining feature pool \mathcal{F}^{OO} and \mathcal{F}^{OE} .



Error

Natural phenomenon

Error recovering strategies

Basic approach

- Erroneous nodes of the tracklets are removed.
- Replaced with the interpolation of the nodes before and after.

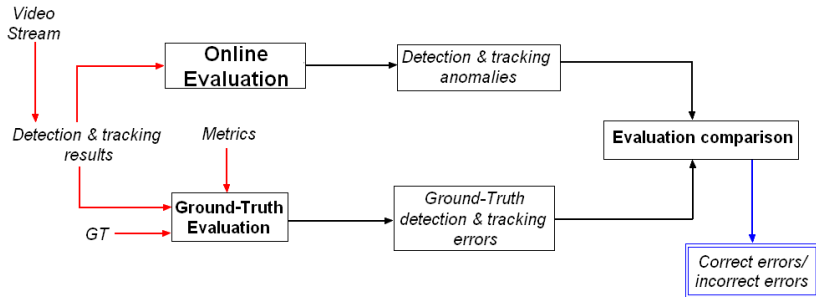
Feedback-based approach

- Feedback sent to the tracking algorithm to tune its parameters.

Evaluation protocol of the online evaluation

Tracking algorithms used

- Tracker 1 [Chau et al., ICDP 2011]: multi-feature tracker using 3D position, shape, dominant color and HOG descriptors. Provides short but reliable tracklets.
- Tracker 2 [Kumar et al., ACCV 2014]: based on graph partitioning.

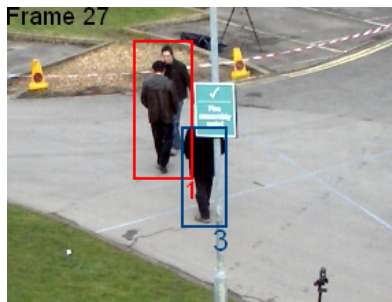
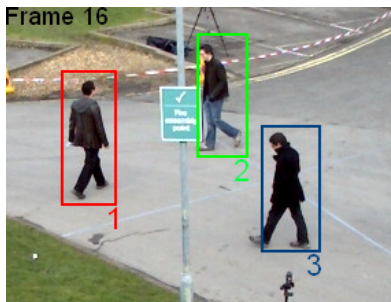


Online evaluation errors

Trackers	Errors from GT	TP (%)	FP (%)	FN (%)
Tracker 1	306	65.60%	6.86%	34.4%
Tracker 2	165	60.61%	9.09%	39.39%

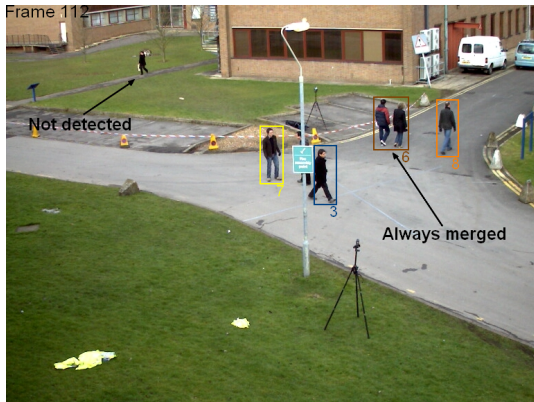
Percentage of errors found with the online evaluation compared to the errors given by the Ground-Truth evaluation for sequence S2.L1 of PETS2009

Detection of easy errors



ID2 is occluded by the pole and by ID1.

Limitations of the online evaluation



Non-detected objects and consistent errors over a long time interval cannot be detected.

CLEAR metrics on PETS2009 dataset

Methods	MOTA	MOTP
[Henriques et al., 2011]	0.85	0.69
[Zamir et al., 2012]	0.90	0.69
[Milan et al., 2013]	0.90	0.74
Tracker 1	0.62	0.63
Tracker 1 + online evaluation (basic recovery)	0.85	0.71
Tracker 1 + online evaluation (feedback recovery)	0.88	0.72
Tracker 2	0.85	0.74
Tracker 2 + online evaluation (basic recovery)	0.90	0.74

Results on CAVIAR dataset

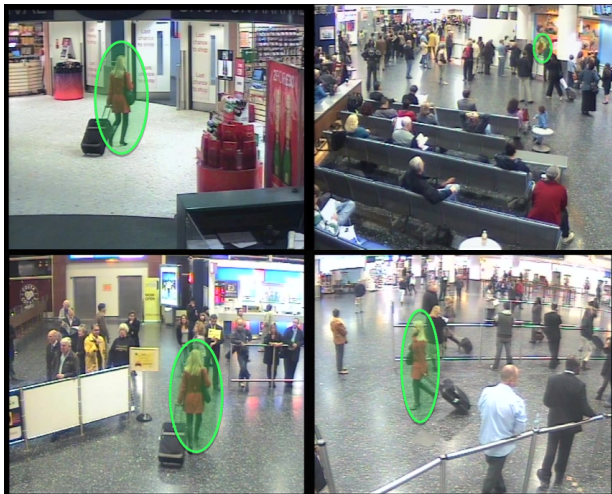
Methods	MT (%)	PT (%)	ML (%)
[Li et al., 2009]	84.6	14.0	1.4
[Kuo et al., 2010]	84.6	14.7	0.7
Tracker 1 alone	78.3	16.0	5.7
Tracker 1 + online evaluation (basic recovery)	82.6	11.7	5.7
Tracker 1 + online evaluation (feedback recovery)	83.8	10.3	5.9

- ML: Mostly tracked (more than 80% of the trajectory is tracked)
- PT: Partially tracked (between 20% and 80% of the trajectory is tracked)
- ML: Mostly lost (less than 20% of the trajectory is tracked)

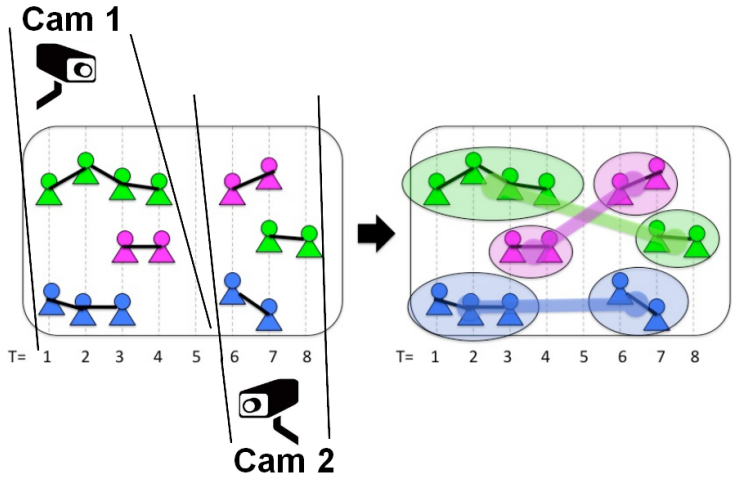
Section 4

Tracklet matching over time

Tracklet matching over time



Tracklet matching over time



Lost tracklets on mono camera

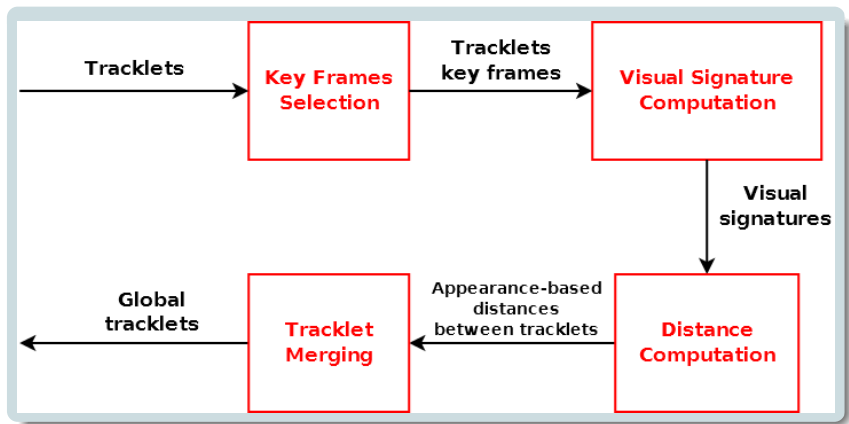
ID change due to tracking errors (occlusions, missed detections)



ID change due to the person reentering the scene



Method overview



Key frames selection

Mean

$$\mu(f^i) = \frac{\sum_{t \in [T_{start}^i, T_{end}^i]} f_t^i}{|[T_{start}^i, T_{end}^i]|}$$

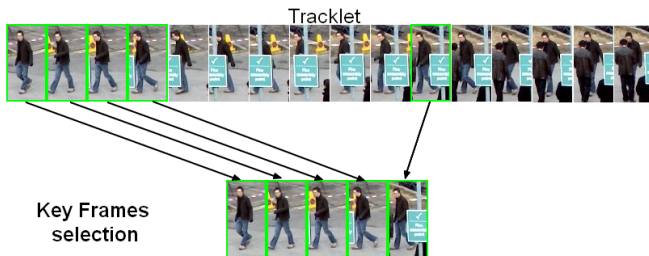
Standard Deviation

$$\forall t \in [T_{start}^i, T_{end}^i] : \sigma(f_t^i) = \frac{\sum_{f_t^i \in \mathcal{F}_t^i} |f_t^i - \mu(f^i)|}{|\mathcal{F}_t^i|}$$

Energy Function

$$E(C^i) = \sum_{g=1}^5 \sigma(f_{t_g}^i) \implies \min_{t_g \in \{1, \dots, 5\} \in [[T_{start}^i, T_{end}^i]]} (E(C^i)) \text{ gives 5 key frames}$$

Key frames selection

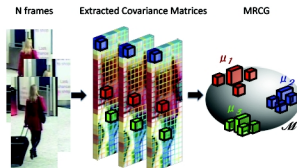


Selection of the 5 key frames that are the closest to the mean value of the features



Visual signature computation

Mean Riemannian Covariance Grid (MRCG) descriptor [S. Bak et al., AVSS 2011]

- Shows very good results on the classical re-identification challenge
- Multi-shot approach
- Holds information on feature distribution, their spatial correlations and their temporal changes throughout the tracklet



Two lists of distances between tracklet pairs

Rank	Possible matches	Distance	Rank	Impossible matches	Distance
1	  ID 7 ID 8	0.12	1	  ID 1 ID 2	0.25
2	  ID 2 ID 10	0.14	2	  ID 10 ID 15	0.28
...



Unsupervised learning

- Parameter p : number of impossible matches used to validate the possible matches

Merging tracklets corresponding to the same person

Constrained clustering algorithm

- Must Link Constraint: tracklets representing the same object must be in the same cluster
- Cannot Link Constraint: given by the impossible matches list

Mean-shift based constrained clustering algorithm

- Computes distance threshold Θ_p using the p tracklet pairs with the smallest distances of the impossible matches list
- Selects a tracklet pair with a distance below threshold Θ_p .
- Adds the pair to the cluster if the intra-distance of the cluster is below threshold Θ_p .

Results on PETS dataset

Methods	p	Correctly merged	Incorrectly merged	Not merged
Visual signature (without key frames selection)	1	21.7%	1.6%	76.7%
	5	53.5%	3.9%	42.6%
	10	59.7%	12.4%	27.9%
Visual signature (with key frames)	1	51.4%	0%	48.6%
	5	71.1%	5.2%	23.7%
	10	78.9%	6.5%	14.6%

A high p value increases the number of both correctly and incorrectly merged tracklets.

Results on CAVIAR dataset

Methods	MT (%)	PT (%)	ML (%)
[Li et al., 2009]	84.6	14.0	1.4
[Kuo et al., 2010]	84.6	14.7	0.7
Tracker 1 alone	78.3	16.0	5.7
Tracker 1 + tracklet matching	84.6	9.5	5.9

Final results on PETS2009 dataset

Methods	MOTA	MOTP
[Berclaz et al., 2011]	0.80	0.58
[Ben Shitrit et al., 2011]	0.81	0.58
[Henriques et al., 2011]	0.85	0.69
[Zamir et al., 2012]	0.90	0.69
[Milan et al. 2013]	0.90	0.74
Online evaluation	0.90	0.74
Tracklet matching	0.83	0.68
Global Tracker	0.92	0.76

Final results on CAVIAR dataset

Methods	MT (%)	PT (%)	ML (%)
[Xing et al. 2009]	84.3	12.1	3.6
[Huang et al. 2008]	78.3	14.7	7
[Li et al. 2009]	84.6	14.0	1.4
[Kuo et al. 2010]	84.6	14.7	0.7
Online Evaluation	82.6	11.7	5.7
Tracklet matching	84.6	9.5	5.9
Global Tracker	86.4	8.3	5.3

Section 5

Conclusion and future works

Conclusion

Global Tracker

- Online evaluation: detects errors by analyzing tracklet coherency and corrects them
- Tracklet matching: improves tracking results by merging tracklets representing the same object

Results

- Tested on several datasets (PETS2009, CAVIAR, TUD, I-LIDS, VANAHEIM), reaching or outperforming the state-of-the-art
- Used in different scenarios (tracking associated with a controller, 3D camera, camera network with overlapping or distant cameras)

Future works and possible improvements

Online evaluation

- Characterize the origin of the errors (detection or tracking)
- Investigate other tracking recovery techniques such as re-detection, re-tracking or backtracking
- Add additional features such as interest points or body parts

Tracklet matching

- Design a set of visual signatures per person if the appearance of the tracklets changes over time
- Overcome the sensitivity of tracklet matching w.r.t. parameter ρ and threshold Θ_ρ

Appendix 1: Related publications

- Julien Badie, François Brémond, *Global tracker: an online evaluation framework to improve tracking quality*, AVSS 2014
- Carolina Garate, Sofia Zaidenberg, Julien Badie, François Brémond, *Group Tracking and Behavior Recognition in Long Video Surveillance Sequences*, VISAPP 2014
- Baptiste Fosty, Carlos Fernando Crispim-Junior, Julien Badie, François Brémond, Monique Thonnat, *Event Recognition System for Older People Monitoring Using an RGB-D Camera*, ASROB 2013 - Workshop on Assistance and Service Robotics in a Human Environment
- Duc Phu Chau, Julien Badie, François Brémond, Monique Thonnat, *Online Tracking Parameter Adaptation based on Evaluation*, AVSS 2013
- Julien Badie, Slawomir Bak, Silviu-Tudor Serban, François Brémond, *Recovering people tracking errors using enhanced covariance-based signatures*, PETS 2012
- Slawomir Bak, Duc-Phu Chau, Julien Badie, Etienne Corvee, François Brémond, Monique Thonnat, *Multi-target Tracking by discriminative analysis on Riemannian Manifold*, ICDP 2012

Appendix 2 : Tracklet matching results on PETS2009 dataset

Methods	MOTA	MOTP
[Zamir et al, 2012]	0.90	0.69
[Milan et al. 2013]	0.90	0.74
Tracker alone	0.82	0.65
Tracklet matching	0.83	0.68

- State-of-the-art results do not take into account people leaving and re-entering the scene.
- State-of-the-art metrics on this dataset are not adapted to evaluate tracklet matching performance.