Occlusion Geodesics for Online Multi-Object Tracking

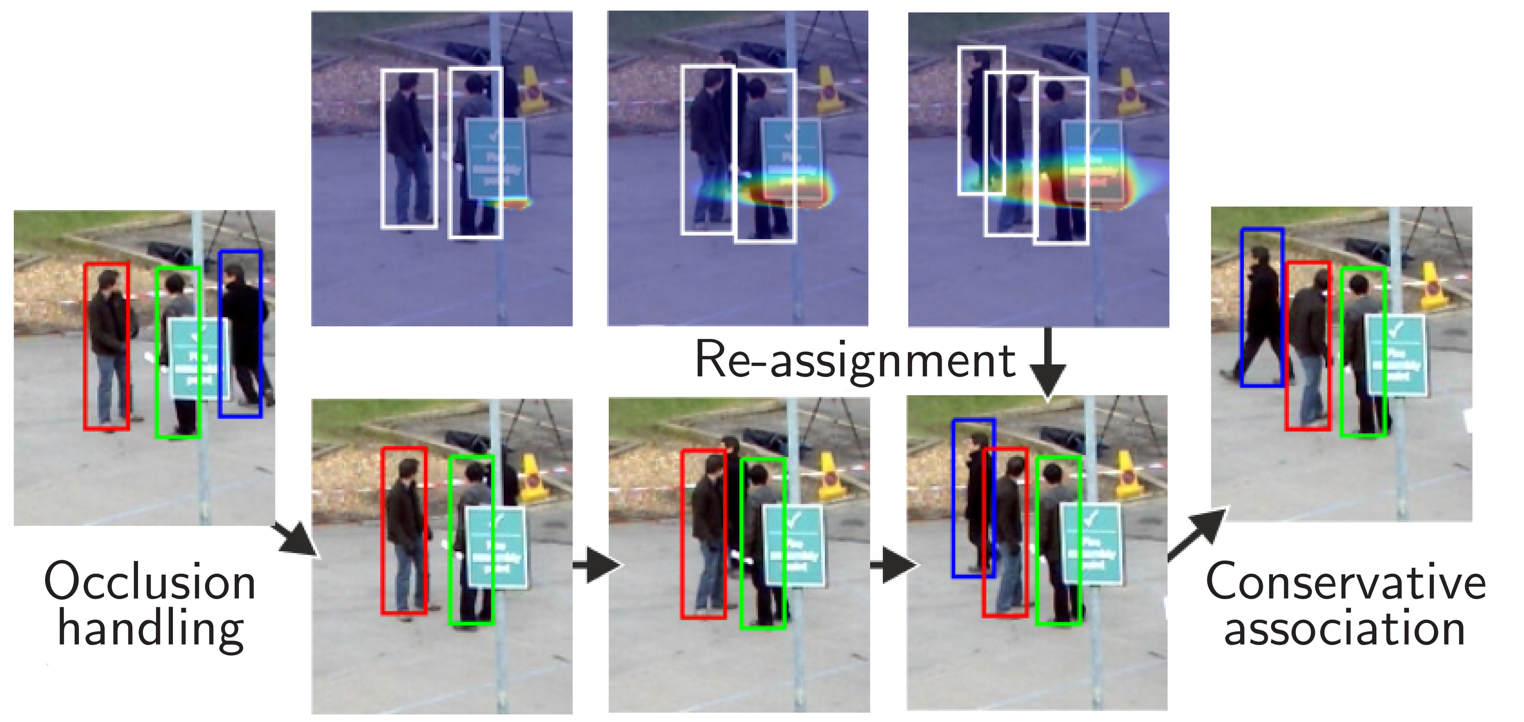
Horst Possegger, Thomas Mauthner, Peter M. Roth, and Horst Bischof

{possegger,mauthner,pmroth,bischof}@icg.tugraz.at
Institute for Computer Graphics and Vision, Austria

Motivation

- State-of-the-art multi-object tracking-by-detection
  - Robust offline linking schemes require detections over a large number of frames or the whole sequence in advance
  - Real-time capable online algorithms suffer from long term occlusions and detection failures
- We focus on re-assignment of missed/occluded targets once they re-appear rather than hallucinating trajectories
- Targets without corresponding detection are more likely moving within occluded regions than being missed by the detector
- Exploit occlusion information and motion prediction to find physically plausible paths
- Implementation publicly available (scan QR code)

Data Association



- Compute assignment matrix $\mathbf{A}^* = \left[a_{ij}^{(t)}\right], a_{ij}^{(t)} \in \{0,1\}$ between N_D detections at time t and N_O object trajectories using Hungarian algorithm:

$$\mathbf{A}^* = \arg\min_{\mathbf{A}} \sum_{i=1}^{N_O} \sum_{j=1}^{N_D} \psi_{ij}^{(t)} a_{ij}^{(t)},$$

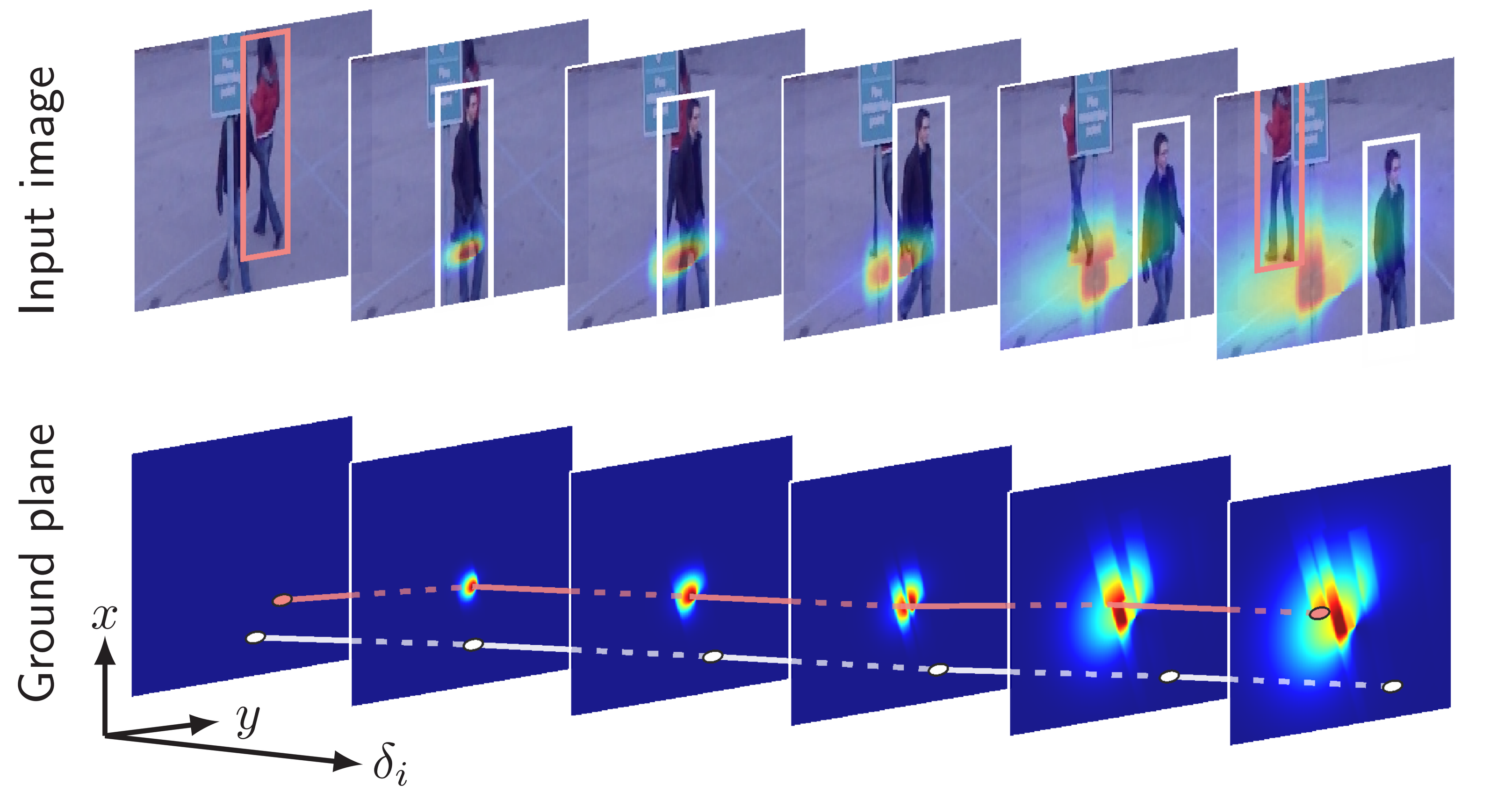
$$\text{s.t.} \sum_{i=1}^{N_O} a_{ij}^{(t)} = 1, \forall j \in \{1,\ldots,N_D\}, \quad \sum_{j=1}^{N_D} a_{ij}^{(t)} = 1, \forall i \in \{1,\ldots,N_O\}$$

- Spatial proximity to decide on safe (isolated and visible) assignments:

$$\psi_{ij}^{(t)} = \begin{cases} \|\mathbf{x}_j^{(t)} - \mathbf{x}_i^{(t-1)}\| & \text{if } \|\mathbf{x}_j^{(t)} - \mathbf{x}_i^{(t-1)}\| < \tau_c \\ \infty & \text{otherwise} \end{cases}$$

- Re-assign candidate detections to previously missed targets based on occlusion geodesics $\psi_{ij}^{(t)} = \Psi_i^{(\delta_i)}\left(\mathbf{x}_j^{(t)}\right)$

Occlusion Geodesics



- Compute instance-specific confidence maps φ_i for missed targets at each time step
- Check each re-assignment candidate at position $\mathbf{x} = (x,y)^\top$ for existence of a valid and plausible path to the last known object position $\hat{\mathbf{x}}_i$
- Efficiently compute cost $\Psi_i^{(\delta_i)}$ of a valid path to a candidate location $\mathbf{x}$ via recursive accumulation:

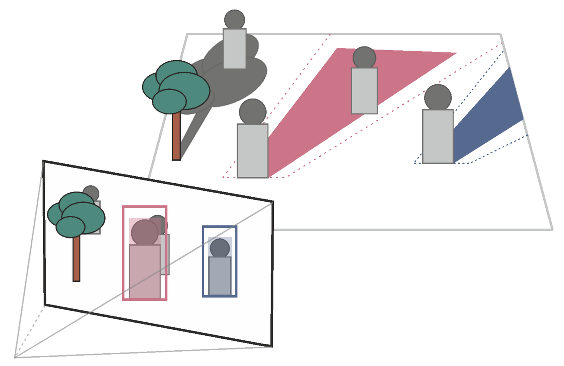
$$\Psi_i^{(\delta_i)}(\mathbf{x}) = 1 - \varphi_i^{(\delta_i)}(\mathbf{x}) + \inf_{\mathbf{z}} \Psi_i^{(\delta_i - 1)}(\mathbf{x} + \mathbf{z}), \quad \|\mathbf{z}\| \leq v_{\text{avg}}$$

Confidence Scores

- Confidence score $\varphi_i^{(\delta_i)}$ indicating the presence of object i at location $\mathbf{x}$ after being missed for δ_i time steps:

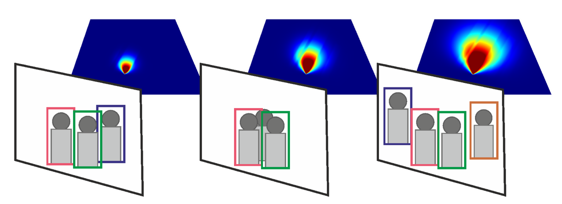
$$\varphi_i^{(\delta_i)}(\mathbf{x}) = c_{o,i}^{(\delta_i)}(\mathbf{x})\, c_{p,i}^{(\delta_i)}(\mathbf{x})\, c_{d,i}^{(\delta_i)}(\mathbf{x})$$

- Occlusion information and detector reliability factor $\beta \in [0,1]$:



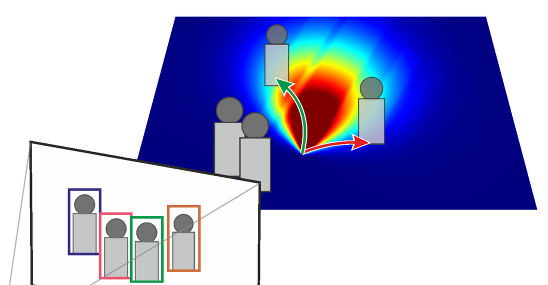
$$c_{o,i}^{(\delta_i)}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \mathcal{P}_s \cup \mathcal{P}_d^{(t)} \\ 1 - \beta^{\delta_i} & \text{otherwise} \end{cases}$$

- Physically plausible distance:



$$c_{p,i}^{(\delta_i)}(\mathbf{x}) = \exp\left(- \frac{\|\mathbf{x} - \hat{\mathbf{x}}_i\|^2}{2\sigma_p^2 \delta_i^2 \max\left(\|\hat{\mathbf{d}}_i\|, v_{\text{avg}}\right)^2} \right)$$

- Inertia model based on predicted motion direction $\hat{\mathbf{d}}_i$ (IQM):

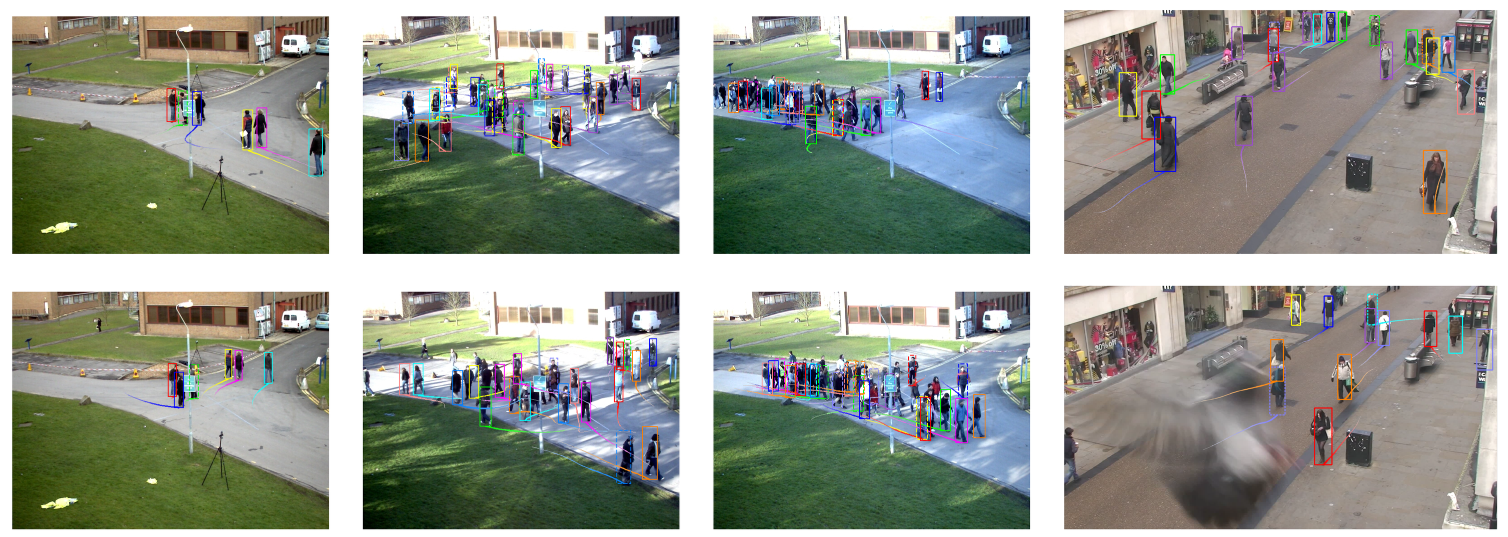


$$c_{d,i}^{(\delta_i)}(\mathbf{x}) = \exp\left(- \frac{\left(\left\langle \hat{\mathbf{d}}_i, \mathbf{d}_j \right\rangle - \|\hat{\mathbf{d}}_i\|\|\mathbf{d}_j\|\right)^2}{2\sigma_d^2 \|\hat{\mathbf{d}}_i\|^2 \|\mathbf{d}_j\|^2} \right)$$

Evaluation and Results

- State-of-the-art performance while being fully online and real-time capable (~ 11 fps, MATLAB)
- Single view experiments on standard benchmark datasets (See paper for detailed evaluation):

Dataset	Approach	Online	App.	MOTA	MOTP	FM	IDS
PETS '09 S2.L1	[2]	no	no	98.0	82.8	11	10
	[4]	no	yes	90.6	80.2	6	11
	[5]	yes	yes	93.3	68.2	-	19
	Proposed	yes	no	98.1	80.5	16	9
PETS '09 S2.L2	[2]	no	no	75.8	65.1	252	234
	[4]	no	yes	56.9	59.4	73	99
	[5]	yes	yes	66.7	58.2	-	215
	Proposed	yes	no	66.0	64.8	315	181
PETS '09 S2.L3	[2]	no	no	62.8	70.5	217	225
	[4]	no	yes	45.5	64.6	27	38
	[5]	yes	yes	40.4	56.4	-	80
	Proposed	yes	no	62.5	62.6	98	59
Town Centre	[1]	yes	no	64.3	80.2	343	222
	[3]	no	no	71.3	71.8	363	165
	[5]	yes	yes	73.6	68.8	-	421
	[6]	no	no	69.1	72.0	440	243
	Proposed	yes	no	70.7	68.6	321	157



PETS'09 S2.L1 PETS'09 S2.L2 PETS'09 S2.L3 Town Centre

References and Acknowledgments

[1] B. Benfold and I. Reid. Stable Multi-Target Tracking in Real-Time Surveillance Video. In *Proc. CVPR*, 2011.
[2] M. Hofmann, D. Wolf, and G. Rigoll. Hypergraphs for Joint Multi-View Reconstruction and Multi-Object Tracking. In *Proc. CVPR*, 2013.
[3] L. Leal-Taixé, G. Pons-Moll, and B. Rosenhahn. Everybody needs somebody: Modeling social and grouping behavior on a linear programming multiple people tracker. In *Proc. ICCV Workshops*, 2011.
[4] A. Milan, S. Roth, and K. Schindler. Continuous Energy Minimization for Multi-Target Tracking. *PAMI*, 2014.
[5] J. Zhang, L. Lo Presti, and S. Sclaroff. Online Multi-Person Tracking by Tracker Hierarchy. In *Proc. AVSS*, 2012.
[6] L. Zhang, Y. Li, and R. Nevatia. Global Data Association for Multi-Object Tracking Using Network Flows. In *Proc. CVPR*, 2008.

This work was supported by the Austrian Science Foundation (FWF) project Advanced Learning for Tracking and Detection in Medical Workflow Analysis (I535-N23).