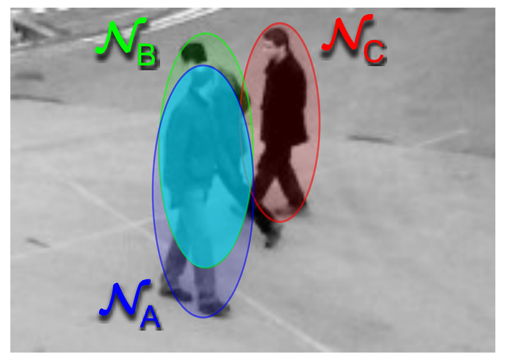
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

An Analytical Formulation of Global Occlusion Reasoning for Multi-Target Tracking

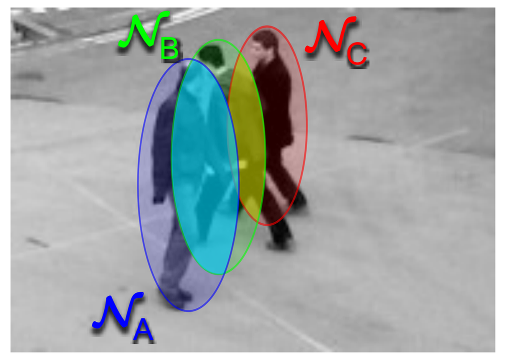
Anton Andriyenko[1] Stefan Roth[1] Konrad Schindler[2]
[1]Department of Computer Science, TU Darmstadt, Germany
[2]Photogrammetry and Remote Sensing Group, ETH Zürich, Switzerland

TECHNISCHE UNIVERSITÄT DARMSTADT

Motivation and Overview



- Frequent and long-term occlusions pose a challenging problem for multi-target tracking due to the complex interdependences between (potentially) all targets and should not be ignored.



- We propose a **global, analytical occlusion model** that allows computing the amount of occlusion and its derivative efficiently in **closed form** and in **continuous space**.

Multi-Target Tracking Framework [1]

- A continuous energy function (1) that accurately captures many important aspects of multi-target tracking.
- The continuous-valued state $\mathbf{X}$ is composed of *all* targets in *all* frames.
- A data term (2), three physically based constraints (3-5) and a regularizer (6):

$$E = E_{\text{obs}} + \alpha E_{\text{dyn}} + \beta E_{\text{exc}} + \gamma E_{\text{per}} + \delta E_{\text{reg}} \quad (1)$$

$$E_{\text{obs}}(\mathbf{X}) = \sum_t \sum_i \left[v_i^t(\mathbf{X}) \cdot \lambda - \sum_g \omega_g^t \frac{s_g^2}{\|\mathbf{X}_i^t - \mathbf{D}_g^t\|^2 + s_g^2} \right] \quad \text{observation} \quad (2)$$

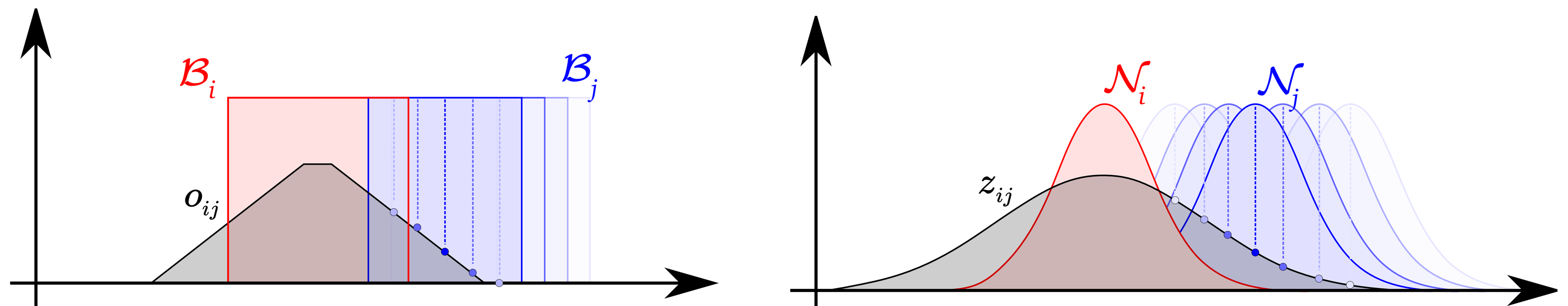
$$E_{\text{dyn}}(\mathbf{X}) = \sum_t \sum_i \left\| \mathbf{X}_i^t - 2\mathbf{X}_i^{t+1} + \mathbf{X}_i^{t+2} \right\|^2 \quad \text{dynamics} \quad (3)$$

$$E_{\text{per}}(\mathbf{X}) = \sum_t \sum_{i, j \neq i} \frac{s_g^2}{\|\mathbf{X}_i^t - \mathbf{X}_j^t\|^2} \quad \text{exclusion} \quad (4)$$

$$E_{\text{exc}}(\mathbf{X}) = \sum_{t \in \{1, F\}} \sum_i \frac{1}{1 + \exp\left(-q \cdot b(\mathbf{X}_i^t) + 1\right)} \quad \text{persistence} \quad (5)$$

$$E_{\text{reg}}(\mathbf{X}) = N + \sum_i \frac{1}{F(i)} \quad \text{regularizer} \quad (6)$$

Analytical Occlusion Reasoning



(a) (b)

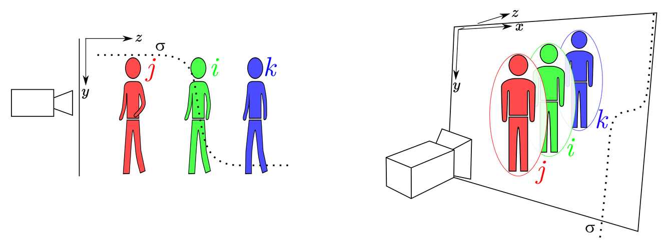
- Relative overlap between two bounding boxes (a) can be computed as a product of two simple indicator functions: $o_{ij} = \frac{1}{\int \mathcal{B}_i(\mathbf{x})\, d\mathbf{x}} \int \mathcal{B}_i(\mathbf{x}) \mathcal{B}_j(\mathbf{x})\, d\mathbf{x}$.
- This occlusion function is not differentiable.
- We propose to model the overlap by a product of two Gaussians (b).
- Their product is proportional to another Gaussian:

$$z_{ij} = \mathcal{N}(\mathbf{c}_i; \mathbf{c}_j, \mathbf{C}_{ij}) = \int \mathcal{N}_i(\mathbf{x}) \cdot \mathcal{N}_j(\mathbf{x})\, d\mathbf{x} \quad (7)$$

- The 'occlusion' is then the unnormalized z_{ij}:

$$V_{ij} = \exp\left(-\tfrac{1}{2} [\mathbf{c}_i - \mathbf{c}_j]^\top \mathbf{C}_{ij}^{-1} [\mathbf{c}_i - \mathbf{c}_j] \right) \quad (8)$$

- The **depth ordering** is modeled with a vertical sigmoid function.
- The occlusion function remains **differentiable**.



$$\sigma_{ij} = \frac{1}{\left(1 + \exp(y_i - y_j)\right)} \quad (9)$$

The total visibility of target i is defined as:

$$v_i(\mathbf{X}) = \exp\left(-\sum_j \sigma_{ij} V_{ij} \right). \quad (10)$$

Limitations

- Gaussians only provide an approximation to the actual shape.
- The level of occlusion may be overestimated.
- Targets are assumed to be roughly the same size on a common ground plane.

Ground Truth

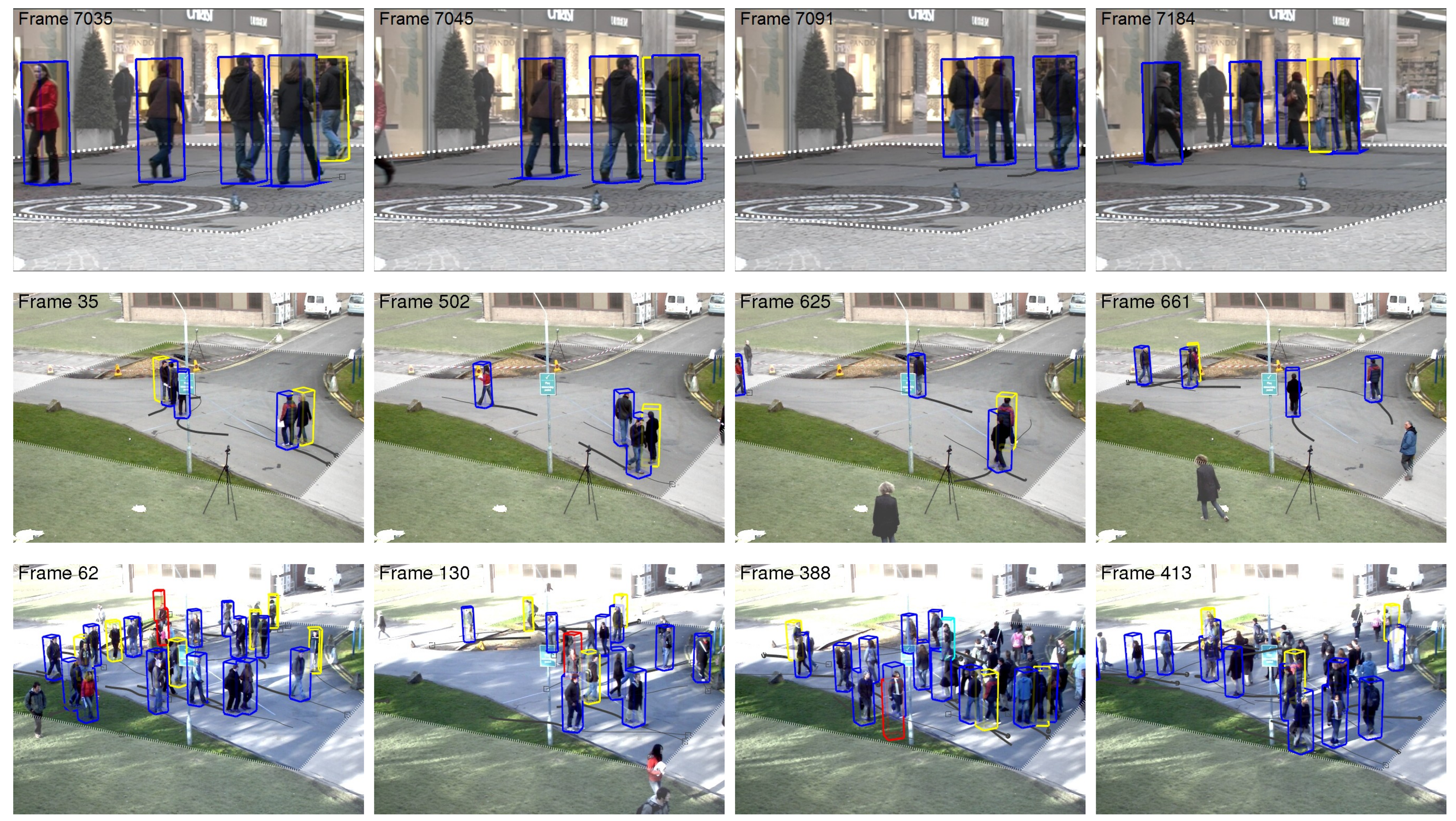
- Only very few public datasets with ground truth exist for crowded scenes.
- We make **all our annotations available** to the community:

goo.gl/3mBeS

Experiments

- Publicly available datasets: TUD-Stadtmitte [2] and PETS'09 [3].
- Varying maximal number of concurrent targets: Between 8 and 42.
- Challenging sequences, originally acquired for **crowd density estimation**.
- Standard CLEAR MOT metrics [4] for quantitative evaluation.

Qualitative Comparison



Tracked only with OM — Tracked only without OM — Tracked with both methods — False positives

Quantitative Evaluation

Crowd density	Method	GT	MT	ML	MOTA	MOTP
medium	OM	*13.0*	**10.7**	**0.3**	**84.4 %**	**74.6 %**
	[1]	*13.0*	10.3	**0.3**	82.5 %	73.9 %
	EKF	*13.0*	4.0	**0.3**	65.4 %	72.2 %
high	OM	*49.5*	**15.5**	**14.2**	**49.4 %**	**63.5 %**
	[1]	*49.5*	13.0	17.8	47.3 %	63.3 %
	EKF	*49.5*	1.5	30.2	22.5 %	61.3 %

Conclusion

- We presented a global, analytical occlusion model for multi-object tracking.
- The proposed occlusion function is continuous and differentiable.
- Its value and gradient are computed efficiently in closed form, thus making it perfectly suitable for gradient based optimization methods.
- Our experiments on crowded scenes reveal the importance of explicitly handling occlusions.

References

[1] A. Andriyenko and K. Schindler. Multi-target tracking by continuous energy minimization. In *CVPR*, 2011.

[2] M. Andriluka, S. Roth, and B. Schiele. Monocular 3d pose estimation and tracking by detection. In *CVPR*, 2010.

[3] J.M. Ferryman and A. Shahrokni. Pets2009: Dataset and challenge. pages 1–6, 2009.

[4] R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa, and P. Soundararajan. The CLEAR 2006 evaluation. In *CLEAR*, 2006.