Occlusion Geodesics for Online Multi-Object Tracking

Horst Possegger Thomas Mauthner Peter M. Roth Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology
{possegger,mauthner,pmroth,bischof}@icg.tugraz.at

Abstract

Robust multi-object tracking-by-detection requires the correct assignment of noisy detection results to object trajectories. We address this problem by proposing an online approach based on the observation that object detectors primarily fail if objects are significantly occluded. In contrast to most existing work, we only rely on geometric information to efficiently overcome detection failures.

In particular, we exploit the spatio-temporal evolution of occlusion regions, detector reliability, and target motion prediction to robustly handle missed detections. In combination with a conservative association scheme for visible objects, this allows for real-time tracking of multiple objects from a single static camera, even in complex scenarios. Our evaluations on publicly available multi-object tracking benchmark datasets demonstrate favorable performance compared to the state-of-the-art in online and offline multi-object tracking.

1. Introduction

One of the most important tasks in many video analysis applications (*e.g.*, visual surveillance or sports analysis) is to robustly estimate the location of objects in a scene. Due to the rapid progress in object detection (*e.g.*, Poselets [8], HOG [11], and DPM [12]), recent research in object tracking has focused on the *tracking-by-detection* principle. Thus, multiple object tracking becomes a *data association* problem where detection responses need to be reliably linked to form target trajectories. However, this is still a difficult and only partially solved problem. In fact, state-of-the-art object detectors often miss objects or are prone to false positive detections due to dynamic backgrounds or changing illumination conditions.

Several recent tracking algorithms address the association problem offline, *i.e.*, by optimizing detection assignments over large temporal windows, *e.g.*, K-shortest paths [6], Hungarian algorithm [16], and hypergraphs [18]. By exploiting information from future time steps, these approaches overcome detection failures, such as missed detections over long occlusion periods. However, processing video sequences in large frame batches (*e.g.*, dynamic programming [14]) or even optimizing over whole sequences (*e.g.*, continuous energy minimization [25]) leads to a significant temporal delay between object observation and estimating its location. Thus, such offline approaches cannot be applied for time-critical video analysis applications (*e.g.*, traffic safety tasks).

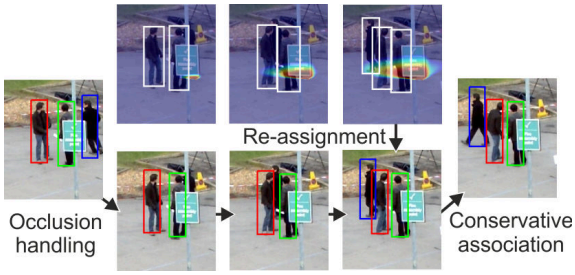


Figure 1: To solve the association problem, *i.e.*, assigning detections (white rectangles, top) to trajectories, we combine a conservative linking scheme for visible objects (red and green) and a novel confidence measure (hot color overlay, top) for occluded objects (blue). By finding physically plausible paths through occlusion regions w.r.t. these confidences, occluded objects can robustly be re-assigned.

Instead, such applications require online tracking methods which only consider observations up to the current frame and provide robust location estimates in real-time (*i.e.*, without temporal delay). To model the uncertainty which arises from occluded targets or missed detections, such trackers often rely on probabilistic frameworks (*e.g.*, Sequential Monte Carlo methods [9, 31]). However, online approaches tend to drift if objects are occluded for longer periods of time and may fail to reliably re-assign missed or occluded objects due to simplified motion models.

Hence, the goal of this work is to overcome these limitations for online multi-object tracking and to achieve high quality results similar to offline approaches. To overcome the drifting problem of existing online trackers, we introduce a novel confidence measure to predict the location of

missed objects, solely based on geometric cues such as occlusion information, detector reliability, and motion prediction. By introducing *occlusion geodesics*, *i.e*., shortest paths (from the location an object first was lost up to its re-detection) w.r.t. these instance-specific confidences, detections of re-appearing objects can reliably be assigned to the corresponding trajectories (*e.g*., the blue target in Figure 1). Additionally, inspired by the low-level tracklet generation of offline approaches such as [19, 21], we use a conservative association scheme to link detections to trajectories of isolated and visible objects (*e.g*., the red and green targets in Figure 1). Combining these association strategies allows for efficiently tracking multiple objects in complex real-world scenarios, as demonstrated by our experimental results.

The remainder of this paper is organized as follows. First, related state-of-the-art multi-object tracking approaches are discussed in Section 2. Next, multi-object tracking by occlusion geodesics is introduced in Section 3. Finally, a detailed evaluation on several challenging real-world datasets is presented in Section 4.

2. Related Work

The major issue of tracking-by-detection approaches is the data association problem, *i.e*., how to correctly assign (possibly noisy) detection results to target trajectories. Until recently, this problem has primarily been addressed by online methods incorporating Joint Probabilistic Data Association Filters [15], Multi-Hypothesis Tracking [32], Markov chain Monte Carlo methods (*e.g*., [5, 28]), and particle filter-based approaches (*e.g*., [29, 33]). Such methods maintain multiple hypotheses until enough observations are available to resolve ambiguities. However, due to the combinatorial hypotheses space such methods often suffer from the exponentially increasing complexity.

Alternatively, the Hungarian algorithm [27] or greedy association schemes (*e.g*., [9, 10, 34]) can be used to solve the association problem. For example, Breitenstein *et al*. [9] use a greedy association scheme in combination with particle filtering based on a constant velocity model. In particular, they use the continuous confidence density output of detectors and online learned instance-specific classifiers to resolve occlusion scenarios. In contrast to [9], we rely on the final detection output and focus on the robust re-assignment of detections to missed objects. Therefore, we allow missed targets to move along physically plausible paths which are defined by combining motion prediction, detector reliability, and geometric knowledge of occluded regions. This allows for handling missed and occluded objects by finding the shortest plausible paths.

Recently, several approaches focused on optimizing trajectories over whole sequences (*e.g*., [26, 39]) or large temporal windows (*e.g*., [14, 21]) to solve the association problem. Such offline approaches often discretize the space of target locations to simplify the underlying optimization problem (*e.g*., [6, 3, 18]). For example, Berclaz *et al*. [6] propose a flow model on a 2D discretization of the ground plane, where detection results are efficiently linked to trajectories using the K-shortest paths algorithm. However, as their method operates offline on a graph built over large frame batches, it cannot handle arbitrarily dense discretizations due to memory limitations. Therefore, other approaches estimate the final object locations by continuous fitting problems to obtain smoother trajectories (*e.g*., [1, 2, 25]) to improve the accuracy of the tracking results.

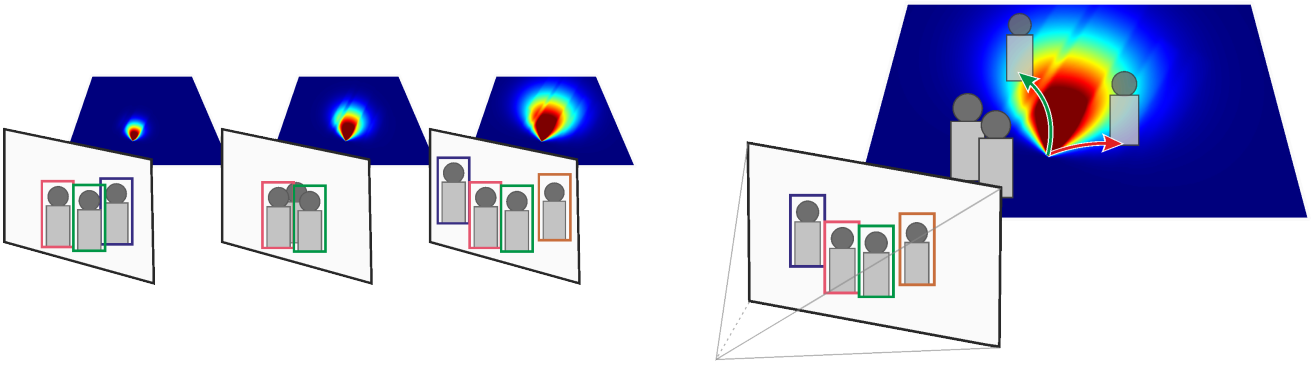
Another group of successful offline approaches follows a hierarchical tracking scheme where subsequent detections are linked together at a low-level pre-processing step, to form short but reliable trajectories, *i.e*., tracklets (*e.g*., [16, 19, 21]). Thus, the key issue becomes to correctly link tracklets to form the final object trajectories, *e.g*., by combining motion and appearance models [17], or by learning tracklet associations from training data [24].

However, since offline approaches require detection results of future frames to perform robust linking, these cannot be used within time-critical applications (*e.g*., surveillance). Here, the main focus lies on robustly linking detections to visible objects and correctly re-assigning detections to previously occluded (or missed) objects in real-time.

3. Tracking by Occlusion Geodesics

We propose to solve the data association problem for online multi-object tracking-by-detection by two complementary steps. First, we compute reliable associations using a conservative linking strategy, as discussed in Section 3.1. This allows for assigning detections to isolated, visible objects (*e.g*., the red and green targets in Figure 1). Second, we introduce instance-specific cost functions which model physically plausible paths through occluded regions to handle missed detections. Using occlusion geodesics (*i.e*., paths with minimal costs), future detections can be reliably re-assigned to missed objects (*e.g*., the blue target in Figure 1), as detailed in Section 3.2.

The proposed occlusion geodesics build on the observation that object detectors fail whenever objects are severely occluded, either dynamically by other objects or by static scene occluders (*e.g*., benches, statues, and trees). In order to re-assign a candidate detection to a previously lost target there must be a physically plausible path through occluded regions, as illustrated in Figure 2. Since missed detections are either caused by detection failures or the object being fully occluded, we propose a novel confidence measure to weight such paths. Therefore, we combine occlusion information, target motion prediction, and object detector reliability to define these confidences, as detailed in Section 3.3.



(a) Evolution of confidence scores (visualized as ground plane overlay; hot colors indicate high confidence) w.r.t. the blue object. Note that occluded regions yield higher confidences.

(b) The green arrow denotes a path with minimal costs based on the blue object's motion confidence and the spatio-temporal evolution of the occlusion regions.

Figure 2: Resolving a typical occlusion scenario (a). Using shortest paths w.r.t. the proposed confidence scores (b), the temporarily occluded blue target can be correctly re-assigned. Relying on Euclidean distances the brown detection would be chosen, as it is closer to the location where the blue target originally has been missed by the detector. Best viewed in color.

3.1. Conservative Data Association

Similar to recent state-of-the-art approaches such as [1, 18], we exploit the scene geometry and perform tracking in real-world coordinates. Thus, given a set of N_D object detections at time t, we project the bottom center point of each detection bounding box $j = \{1, \dots, N_D\}$ onto the ground plane to obtain its 2D real-world location $\mathbf{x}_j^{(t)}$. Then, detections can be assigned to isolated and visible objects based on spatial proximity. In particular, given N_O object trajectories at time $t-1$, we define the cost $\psi_{ij}^{(t)}$ of assigning detection j to object i using the Euclidean distance to the previously observed object location $\mathbf{x}_i^{(t-1)}$ as

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

where τ_c is a conservative distance threshold, and $\psi_{ij}^{(t)} = \infty$ denotes impossible assignments. To obtain the optimal assignment of reliable matches at time t, we use the Hungarian algorithm [27] for computing the assignment matrix $\mathbf{A}^* = \left[a_{ij}^{(t)}\right], a_{ij}^{(t)} \in \{0, 1\}$, which minimizes the total association cost[1]:

$$\mathbf{A}^* = \arg\min_{\mathbf{A}} \sum_{i=1}^{N_O} \sum_{j=1}^{N_D} \psi_{ij}^{(t)} a_{ij}^{(t)}, \quad (2)$$

$$\text{s.t.} \sum\nolimits_{i=1}^{N_O} a_{ij}^{(t)} = 1, \forall j \in \{1, \dots, N_D\},$$

$$\sum\nolimits_{j=1}^{N_D} a_{ij}^{(t)} = 1, \forall i \in \{1, \dots, N_O\}.$$

[1]Although the original formulation assumes $N_D = N_O$, the Hungarian algorithm can easily be extended for rectangular assignment matrices where $N_D \neq N_O$.

All objects which could not be assigned by this conservative association scheme are considered to be missed by the detector. Such false negative detections are either caused by static and dynamic occluders (see Figure 3) or detection failures. Thus, future detections must be re-assigned to the corresponding trajectories whenever missed objects are re-detected (*e.g.*, after exiting occluded regions). In the following, we introduce occlusion geodesics to solve this association problem efficiently.

3.2. Occlusion Geodesics for Data Association

To overcome missed detections, we introduce a novel confidence measure predicting the location of a missed object w.r.t. occlusion information, detector reliability, and motion prediction. This allows for weighting physically plausible paths from the location a target first was missed up to its re-detection. Using occlusion geodesics (*i.e.*, minimal paths w.r.t. these confidence weights), we can reliably re-assign detections to previously missed objects. In contrast to state-of-the-art approaches such as [9, 16, 38], which rely on appearance information to resolve occluded trajectories, we only exploit the available geometric information to highlight the favorable performance of the proposed occlusion geodesics.

In particular, let i denote a missed object and δ_i the occlusion length, *i.e.*, for how long object i has been missed. Moreover, let $c_{o,i}^{(\delta_i)}$ be the occlusion confidence to account for occluded objects and detection failures, $c_{p,i}^{(\delta_i)}$ the motion range confidence to limit physically plausible movement, and $c_{d,i}^{(\delta_i)}$ the object's inertia model. Then, for a location $\mathbf{x} \in \mathbb{R}^2$ on the ground plane, we define

$$\varphi_i^{(\delta_i)}(\mathbf{x}) = c_{o,i}^{(\delta_i)}(\mathbf{x})\, c_{p,i}^{(\delta_i)}(\mathbf{x})\, c_{d,i}^{(\delta_i)}(\mathbf{x}) \quad (3)$$

to indicate the confidence of object i being present at location $\mathbf{x}$ after being missed by the detector for δ_i frames. The corresponding confidence terms will be discussed in more detail in Section 3.3.

Since occluded regions change over time (*e.g.*, whenever occluding objects move), the spatio-temporal evolution of occlusions has to be considered to reliably re-assign detections to a missed object. Therefore, assuming an average object velocity v_{avg} between subsequent frames, we weight physically plausible paths by the recursive cost function

$$\Psi_i^{(\delta_i)}(\mathbf{x}) = 1 - \varphi_i^{(\delta_i)}(\mathbf{x}) + \inf_{\mathbf{z}} \Psi_i^{(\delta_i - 1)}(\mathbf{x} + \mathbf{z}) \,. \quad (4)$$

Accumulating the infima within the spatial neighborhood $\mathbf{x} + \mathbf{z}, \|\mathbf{z}\| \leq v_{\text{avg}}$ over time ensures that $\Psi_i^{(\delta_i)}(\mathbf{x})$ represents the minimum cost of all paths reachable by object i, which lead from its last known position up to location $\mathbf{x}$. The initial re-assignment cost for the recursive computation is set to $\Psi_i^{(0)} = 0$.

Similar to the conservative association scheme, we use the Hungarian algorithm (recall Eq. (2)) to obtain the optimal assignment between missed objects and candidate re-detections at time t. In particular, given the ground plane location $\mathbf{x}_j^{(t)}$ of detection j, we set the assignment costs to $\psi_{ij}^{(t)} = \Psi_i^{(\delta_i)}\left(\mathbf{x}_j^{(t)}\right)$.

3.3. Confidence Scores

Considering the occluded regions at a specific time t, we combine occlusion information and motion prediction to estimate the confidence measure $\varphi_i^{(\delta_i)}$ (recall Eq. (3)). In particular, we expect the object detector to miss an object whenever it is fully occluded or environmental conditions cause detection failures (*e.g.*, illumination changes). Therefore, we define the occlusion term $c_{o,i}^{(\delta_i)}$ as

$$c_{o,i}^{(\delta_i)}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \mathcal{P}_s \cup \mathcal{P}_d^{(t)} \\ 1 - \beta^{\delta_i} & \text{otherwise} \end{cases}, \quad (5)$$

where $\beta \in [0, 1]$ is a detector reliability factor to account for missed detections of visible objects; $\mathcal{P}_s$ and $\mathcal{P}_d^{(t)}$ denote the currently occluded regions at time t caused by static and dynamic occluders, respectively.

To obtain the dynamic occlusion regions $\mathcal{P}_d^{(t)}$, we exploit the geometric knowledge of the currently visible objects. Assuming that partially occluded objects can be handled by state-of-the-art detectors (*e.g.*, part-based models [12]), we can project the center region of each detected bounding box onto the ground plane, as illustrated in Figure 3. Occlusion regions $\mathcal{P}_s$ caused by static scene structures can easily be provided as a predefined mask.

In order to restrict the re-assignment candidates to detections which can be reached via physically plausible motion

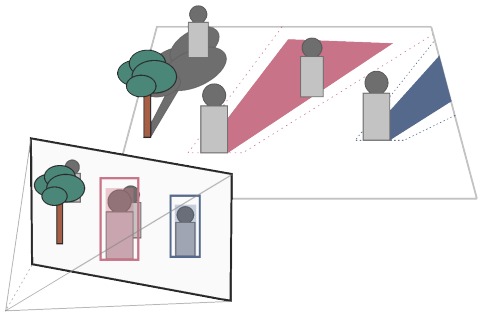


Figure 3: Projected occlusion regions for static occluders ($\mathcal{P}_s$, gray) and dynamic inter-object occlusions ($\mathcal{P}_d^{(t)}$, red and blue).

of the target, we define the plausibility term

$$c_{p,i}^{(\delta_i)}(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \hat{\mathbf{x}}_i\|^2}{2\sigma_p^2 \delta_i^2 \max\left(\|\hat{\mathbf{d}}_i\|, v_{\text{avg}}\right)^2}\right), \quad (6)$$

where σ_p denotes the motion variance, $\hat{\mathbf{x}}_i$ is the last known position of the occluded object i at $\delta_i = 0$, and $\hat{\mathbf{d}}_i$ is the predicted motion direction of object i. To estimate $\hat{\mathbf{d}}_i$, we consider the previously observed target motion between subsequent frames and compute the interquartile mean to robustly handle outliers (*e.g.*, due to inaccurate localization by the detector). To enforce the hard constraint that the distance between the last known target position $\hat{\mathbf{x}}_i$ and the ground plane location $\mathbf{x}$ must lie within physically plausible limits, we use the predefined cut-off threshold τ_p to set $c_{p,i}^{(\delta_i)} = -\infty$ if $c_{p,i}^{(\delta_i)} < \tau_p$.

Additionally, we exploit the available previous observations of the object trajectory to model its inertia confidence:

$$c_{d,i}^{(\delta_i)}(\mathbf{x}) = \exp\left(-\frac{\left(\left\langle \hat{\mathbf{d}}_i, \mathbf{d}_j \right\rangle - \|\hat{\mathbf{d}}_i\| \|\mathbf{d}_j\|\right)^2}{2\sigma_d^2 \|\hat{\mathbf{d}}_i\|^2 \|\mathbf{d}_j\|^2}\right), \quad (7)$$

where $\mathbf{d}_j = \mathbf{x} - \hat{\mathbf{x}}_i$. The directional variance σ_d can be used to penalize significant changes of the motion direction. In particular, choosing a small directional variance proves to be useful in scenarios where the object direction can be predicted, *e.g.*, when observing pedestrians on a sidewalk.

3.4. Automatic Initialization and Termination

In order to enable fully automatic tracking of an unknown number of targets, we exploit the object detector to initialize and cancel trajectories. In particular, we define entry and exit regions near the image borders, similar to recent approaches, such as [9, 14, 18]. Thus, a new trajectory is initialized for subsequent nearby detections within the entry regions. Similarly, trajectories are terminated if the corresponding objects exit the field-of-view.

4. Evaluation

In the following, we give a detailed analysis of our approach compared to the state-of-the-art in multi-object tracking.

4.1. Metrics

We use the widely accepted CLEAR performance metrics [7], *Multiple Object Tracking Accuracy* (MOTA [2]) and *Precision* (MOTP [2]). The precision metric MOTP evaluates the alignment of true positive trajectories w.r.t. the ground truth, whereas the accuracy metric MOTA combines 3 error ratios, namely false positives, false negatives (*i.e.*, missed objects), and identity switches. Additionally, we report the measures defined by Li *et al*. [24], which denote the percentage of *mostly tracked* (MT [2]) and *mostly lost* (ML [3]) ground truth trajectories, as well as the number of *fragments* (FM [3]) and *identity switches* (IDS [3]).

4.2. Datasets

To demonstrate the performance of our online multi-object tracker, we use two publicly available benchmark datasets which impose several challenges.

PETS'09. The *PETS'09* benchmark [13] shows an outdoor scene with numerous pedestrians recorded from multiple cameras at 7 fps, where we only use the first camera (*i.e.*, View 1). The major challenges of this dataset are frequent occlusions, either caused dynamically by people occluding each other, or static occlusions due to a traffic sign which covers large parts of the crossroads. Additionally to the widely used *S2L1* sequence, we also evaluate our approach on the more challenging *S2L2* and *S2L3* sequences, which capture much denser crowds.

For a fair comparison, we use the ground truth provided by Milan *et al*. [25], where all occurring persons have been annotated. Following their evaluation protocol, we compute the distances between tracker hypotheses and ground truth annotations on the ground plane, using a hit/miss threshold of 1 m. Similar to Hofmann *et al*. [18], we use DPM detections [12] as input for our tracking algorithm. As the *PETS'09* sequences are captured using up to 7 cameras, we also include state-of-the-art multi-camera approaches into our comparison.

Town Centre. The *Town Centre* dataset [5] shows a busy town centre street from a single elevated camera. On average, 16 people are visible at any time, resulting in frequent dynamic occlusions. Furthermore, many people are not detected due to partial occlusions caused by static scene structures such as benches.

The dataset provides manually annotated ground truth trajectories as well as pre-computed HOG detections [11]. Similar to Leal-Taixé *et al*. [22], we track at 2.5 fps (*i.e.*, use only every 10^{th} frame) to demonstrate the robustness of our approach even at low frame rates. Following their evaluation protocol, the assignment of tracker hypotheses to ground truth annotations is computed via bounding box overlap using the PASCAL overlap criterion as hit/miss threshold.

[2]Higher is better. [3]Lower is better.

4.3. Experimental Settings

To track all objects throughout the benchmark sequences, the proposed tracking algorithm relies on several intuitive parameters. In particular, we used the following default parameter settings for our experiments.

The average velocity is set to $v_{\text{avg}} = 4.5$ m/s to account for both pedestrians and cyclists. To obtain reliable matches by conservative linking, only detections within a vicinity of $\tau_c = 0.5$ m between subsequent frames at 7 fps are considered for valid assignments. The variances for the range and inertia models are set to $\sigma_p^2 = 1$ and $\sigma_d^2 = 0.5$, respectively. To discard candidate detections outside the physically plausible motion range, we choose a range cut-off threshold of $\tau_p = 0.01$. Since we use state-of-the-art object detectors which achieve high recall, we set the expected detector reliability factor to $\beta = 0.9$. For a fair comparison to other approaches, we linearly interpolate missing detections of the final object trajectories.

4.4. Results and Discussion

Quantitative results of our comparison on the two benchmark datasets are listed in Tables 1 and 2, while illustrative results are shown in Figure 4. As can be seen from the tracking results on the *PETS'09 S2L1* benchmark, we achieve excellent results in contrast to competing online multi-object trackers, such as [9, 35, 37], even outperforming most offline approaches, *e.g.*, [2, 6, 38]. Furthermore, we achieve competitive results compared to the offline hypergraph formulation of Hofmann *et al*. [18], which currently is the best-performing approach on the *PETS'09* sequences. Note that our approach also achieves similar performance results compared to offline multi-camera approaches.

Considering the more challenging *PETS'09 S2L2* and *S2L3* sequences, our proposed approach also performs favorably compared to most state-of-the-art trackers. In particular, the proposed re-assignment approach using occlusion geodesics leads to significantly less identity switches compared to the best-performing approach of Hofmann *et al*. [18]. Furthermore, the experimental results indicate that additional appearance information (*e.g.*, as used by [35]) may improve tracking performance for moderately crowded scenarios, such as *S2L2*.

The results on the much longer *Town Centre* dataset also confirm that the proposed occlusion geodesics are beneficial compared to standard occlusion handling techniques.

Although using only every 10^{th} frame, we outperform other online tracking approaches, such as [5, 30, 35, 36]. Similar to the *PETS'09* sequences, we again achieve competitive results to the best-performing offline approaches on the *Town Centre* dataset. Note however, that both Izadinia *et al*. [20] and Zamir *et al*. [38] use custom detections and additionally exploit appearance information. Thus, considering only trackers which rely on the provided HOG detections, our approach performs on par with the best-performing offline tracker by Leal-Taixé *et al*. [22].

Furthermore, to emphasize the real-time capability of our approach, we report the average frame rate on a standard PC with a 3.4 GHz Intel CPU. In particular, our unoptimized single-threaded MATLAB prototype runs at 11.2 fps for moderately crowded scenarios (*e.g*., *PETS'09 S2L1*) and still achieves a frame rate of 2.1 fps for much denser crowds where up to 38 objects are simultaneously visible (*e.g*., *PETS'09 S2L2 and S2L3*). In contrast, competing online approaches report significantly lower runtime performances of $0.4 - 2$ fps [9], or $1 - 2$ fps [35] for the same scenarios, also excluding the detection step. Hence, our experiments confirm that combining efficient object detectors (*e.g*., [4]) with our tracking approach allows for robust online localization of multiple objects in real-time applications.

5. Conclusion

In this paper, we proposed an online multi-object tracking-by-detection approach for real-time applications. To account for detection failures, we exploit geometric cues such as the spatio-temporal evolution of occlusion regions, motion prediction, and detector reliability. Using these cues to model physically plausible paths of missed objects, we can reliably re-assign detections to re-appearing objects. In combination with a conservative association strategy for visible objects, multiple objects can robustly be tracked, even in crowded scenarios.

Our evaluations on several challenging real-world datasets demonstrate significant improvements compared to the state-of-the-art in online and offline multi-object tracking. In particular, although using only observations up to the current frame, our results are on par with the best-performing offline approaches which require detections for each frame of a sequence in advance. Thus, the proposed tracking approach can be applied for time-critical applications where location estimates of multiple objects are required in real-time.

Acknowledgments We would like to thank Martin Hofmann, Laura Leal-Taixé, and João Henriques for providing their results and helpful discussions. This work was supported by the Austrian Science Foundation (FWF) under the project Advanced Learning for Tracking and Detection in Medical Workflow Analysis (I535-N23).

References

[1] A. Andriyenko and K. Schindler. Multi-target Tracking by Continuous Energy Minimization. In *Proc. CVPR*, 2011.

[2] A. Andriyenko, K. Schindler, and S. Roth. Discrete-Continuous Optimization for Multi-Target Tracking. In *Proc. CVPR*, 2012.

[3] H. Ben Shitrit, J. Berclaz, F. Fleuret, and P. Fua. Multi-Commodity Network Flow for Tracking Multiple People. *PAMI*, 2014. To appear.

[4] R. Benenson, M. Mathias, R. Timofte, and L. Van Gool. Pedestrian detection at 100 frames per second. In *Proc. CVPR*, 2012.

[5] B. Benfold and I. Reid. Stable Multi-Target Tracking in Real-Time Surveillance Video. In *Proc. CVPR*, 2011.

[6] J. Berclaz, F. Fleuret, E. Türetken, and P. Fua. Multiple Object Tracking using K-Shortest Paths Optimization. *PAMI*, 33(9):1806–1819, 2011.

[7] K. Bernardin and R. Stiefelhagen. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP JIVP*, 2008.

[8] L. Bourdev, S. Maji, T. Brox, and J. Malik. Detecting People Using Mutually Consistent Poselet Activations. In *Proc. ECCV*, 2010.

[9] 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*, 33(9):1820–1833, 2011.

[10] Y. Cai, N. de Freitas, and J. J. Little. Robust Visual Tracking for Multiple Targets. In *Proc. ECCV*, 2006.

[11] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *Proc. CVPR*, 2005.

[12] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. *PAMI*, 32(9):1627–1645, 2010.

[13] J. Ferryman and A. Shahrokni. PETS2009: Dataset and Challenge. In *Proc. Winter-PETS*, 2009.

[14] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multi-Camera People Tracking with a Probabilistic Occupancy Map. *PAMI*, 30(2):267–282, 2008.

[15] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffé. Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association. *J. Oceanic Eng.*, 8(3):173–184, 1983.

[16] J. F. Henriques, R. Caseiro, and J. Batista. Globally Optimal Solution to Multi-Object Tracking with Merged Measurements. In *Proc. ICCV*, 2011.

[17] M. Hofmann, M. Haag, and G. Rigoll. Unified Hierarchical Multi-Object Tracking using Global Data Association. In *Proc. PETS*, 2013.

[18] M. Hofmann, D. Wolf, and G. Rigoll. Hypergraphs for Joint Multi-View Reconstruction and Multi-Object Tracking. In *Proc. CVPR*, 2013.

[19] C. Huang, B. Wu, and R. Nevatia. Robust Object Tracking by Hierarchical Association of Detection Responses. In *Proc. ECCV*, 2008.

[20] H. Izadinia, I. Saleemi, W. Li, and M. Shah. $(MP)^2T$: Multiple People Multiple Parts Tracker. In *Proc. ECCV*, 2012.

	Method	Appearance	Cam IDs	MOTA [%]	MOTP [%]	MT [%]	ML [%]	FM	IDS
online	Breitenstein *et al*. [9]	yes	1	79.7	56.3	-	-	-	-
	Wu *et al*. [35]	yes	1	92.8	74.3	**100.0**	**0.0**	**11**	**8**
	Yang *et al*. [37]	yes	1	76.0	53.8	-	-	-	-
	Proposed	no	1	**98.1**	**80.5**	**100.0**	**0.0**	16	9
offline	Andriyenko & Schindler [1]	no	1	81.4	76.1	82.6	**0.0**	21	15
	Andriyenko *et al*. [2]	no	1	95.9	78.7	**100.0**	**0.0**	8	10
	Berclaz *et al*. [6]	no	1	80.3	72.0	73.9	8.7	22	13
	Henriques *et al*. [16]	yes	1	83.3	71.1	89.5	**0.0**	45	19
	Hofmann *et al*. [17]	yes	1	97.8	75.3	**100.0**	**0.0**	8	**8**
	Hofmann *et al*. [18]	no	1	**98.0**	**82.8**	**100.0**	**0.0**	11	10
	Izadinia *et al*. [20]	yes	1	90.7	76.0	-	-	-	-
	Milan *et al*. [25]	yes	1	90.6	80.2	91.3	4.3	**6**	11
	Milan *et al*. [26]	no	1	90.3	74.3	78.3	**0.0**	15	22
	Zamir *et al*. [38]	yes	1	90.3	69.0	89.5	**0.0**	54	10
offline	Hofmann *et al*. [18]	no	1+5	**99.4**	82.9	**100.0**	**0.0**	**1**	**1**
	Hofmann *et al*. [18]	no	1+5+7	**99.4**	**83.0**	**100.0**	**0.0**	**1**	2
	Leal-Taixé *et al*. [23]	no	1+5	76.0	60.0	-	-	-	-
	Leal-Taixé *et al*. [23]	no	1+5+6	71.4	53.4	-	-	-	-

(a) *PETS'09 S2L1*.

	Method	App.	Cam IDs	MOTA [%]	MOTP [%]	MT [%]	ML [%]	FM	IDS
onl.	Wu *et al*. [35]	yes	1	**73.3**	**73.2**	**68.9**	4.1	**113**	**122**
	Proposed	no	1	66.0	64.8	41.5	**0.0**	315	181
offline	Berclaz *et al*. [6]	no	1	24.2	60.9	9.5	54,0	38	**22**
	Hofmann *et al*. [17]	yes	1	57.1	56.4	39.5	18,4	**59**	67
	Hofmann *et al*. [18]	no	1	**75.8**	**72.1**	**65.1**	**0.0**	252	234
	Milan *et al*. [25]	yes	1	56.9	59.4	37.8	16.2	73	99
	Milan *et al*. [26]	no	1	46.0	59.8	33.8	10.8	105	126
offl.	Hofmann *et al*. [18]	no	1+2	**87.6**	73,5	**86.0**	**0.0**	**128**	**111**
	Hofmann *et al*. [18]	no	1+2+3	79.7	**74.2**	69.8	2.3	129	132

(b) *PETS'09 S2L2*.

	Method	App.	Cam IDs	MOTA [%]	MOTP [%]	MT [%]	ML [%]	FM	IDS
onl.	Wu *et al*. [35]	yes	1	58.3	**69.7**	**47.7**	18.2	**39**	**41**
	Proposed	no	1	**62.5**	62.6	31.8	**13.6**	98	59
offline	Berclaz *et al*. [6]	no	1	28.8	61.8	11.4	70.5	**12**	**7**
	Hofmann *et al*. [17]	yes	1	41.5	65.0	34.1	31.8	67	46
	Hofmann *et al*. [18]	no	1	**62.8**	**70.5**	**54.5**	**11.4**	217	225
	Milan *et al*. [25]	yes	1	45.5	64.6	20.5	40.9	27	38
	Milan *et al*. [26]	no	1	39.8	65.0	18.2	43.2	22	27
offl.	Hofmann *et al*. [18]	no	1+2	**68.5**	72.3	**54.5**	**20.5**	149	156
	Hofmann *et al*. [18]	no	1+2+4	65.4	**73.9**	40.9	25.0	**88**	**116**

(c) *PETS'09 S2L3*.

Table 1: Quantitative results on the *PETS'09* benchmark sequences. Result listings are grouped into online monocular, offline monocular, and offline multi-camera approaches. Bold scores highlight the best results of each group.

[21] C.-H. Kuo and R. Nevatia. How does person identity recognition help multi-person tracking? *Proc. CVPR*, 2011.

[22] 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.

[23] L. Leal-Taixé, G. Pons-Moll, and B. Rosenhahn. Branch-and-price global optimization for multi-view multi-target tracking. In *Proc. CVPR*, 2012.

[24] Y. Li, C. Huang, and R. Nevatia. Learning to Associate: HybridBoosted Multi-Target Tracker for Crowded Scene. In *Proc. CVPR*, 2009.

[25] A. Milan, S. Roth, and K. Schindler. Continuous Energy Minimization for Multi-Target Tracking. *PAMI*, 36(1):58–72, 2014.

[26] A. Milan, K. Schindler, and S. Roth. Detection- and Trajectory-Level Exclusion in Multiple Object Tracking. In *Proc. CVPR*, 2013.

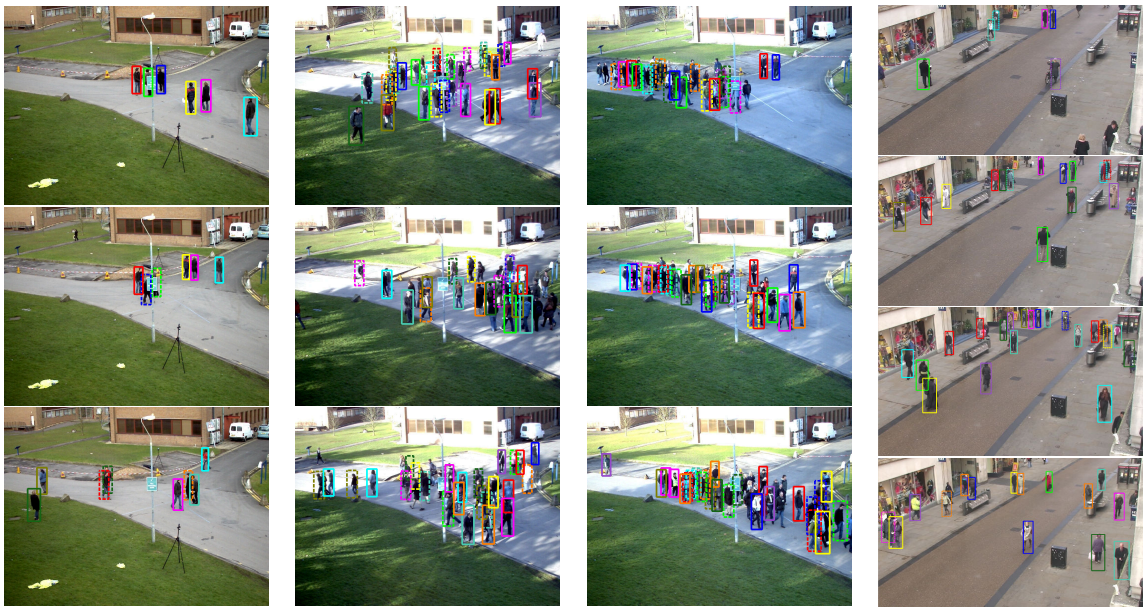
[27] J. Munkres. Algorithms for the Assignment and Transportation Problems. *J. Soc. Ind. Appl. Math.*, 5(1):32–38, 1957.

[28] S. Oh, S. Russel, and S. Sastry. Markov Chain Monte Carlo Data Association for Multiple-Target Tracking. *Trans. Automatic Control*, 54(3):481–497, 2009.

[29] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe. A Boosted Particle Filter: Multitarget Detection and

	Method	App.	MOTA [%]	MOTP [%]	MT [%]	ML [%]	FM	IDS
online	Benfold & Reid [5]	no	64.3	**80.2**	**67.4**	**6.5**	343	222
	Pellegrini *et al*. [30]	no	65.5	71.8	59.1	7.0	499	288
	Wu *et al*. [35]	yes	69.5	68.7	64.7	7.9	453	209
	Yamaguchi *et al*. [36]	no	66.6	71.7	58.1	**6.5**	492	302
	Proposed	no	**70.7**	68.6	56.3	7.4	**321**	**157**
offline	Izadinia *et al*. [20]	yes	**75.7**	71.6	-	-	-	-
	Leal-Taixé *et al*. [22]	no	71.3	71.8	**58.6**	**7.0**	**363**	**165**
	Zamir *et al*. [38]	yes	75.6	71.9	-	-	-	-
	Zhang *et al*. [39]	no	69.1	**72.0**	53.0	9.3	440	243

Table 2: Evaluation on the *Town Centre* dataset. Results of [30, 36, 39] have been provided by the authors of [22]. In contrast to [22], we use the stricter definition of identity switches by Li *et al*. [24].



(a) *PETS'09 S2L1*. (b) *PETS'09 S2L2*. (c) *PETS'09 S2L3*. (d) *Town Centre*.

Figure 4: Sample tracking results on the evaluated public datasets. Dashed rectangles indicate interpolated results for occluded objects. Additional results are included in the supplemental material. Best viewed in color.

Tracking. In *Proc. ECCV*, 2004.

[30] S. Pellegrini, A. Ess, K. Schindler, and L. Van Gool. You'll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking. In *Proc. ICCV*, 2009.

[31] H. Possegger, S. Sternig, T. Mauthner, P. M. Roth, and H. Bischof. Robust Real-Time Tracking of Multiple Objects by Volumetric Mass Densities. In *Proc. CVPR*, 2013.

[32] D. B. Reid. An Algorithm for Tracking Multiple Targets. *Trans. Automatic Control*, 24(6):843–854, 1979.

[33] J. Vermaak, A. Doucet, and P. Perez. Maintaining Multi-Modality through Mixture Tracking. In *Proc. ICCV*, 2003.

[34] B. Wu and R. Nevatia. Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors. *IJCV*, 75(2):247–266, 2007.

[35] Z. Wu, J. Zhang, and M. Betke. Online Motion Agreement Tracking. In *Proc. BMVC*, 2013.

[36] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and Where are you going? In *Proc. CVPR*, 2011.

[37] J. Yang, P. A. Vela, Z. Shi, and J. Teizer. Probabilistic Multiple People Tracking through Complex Situations. In *Proc. PETS*, 2009.

[38] A. R. Zamir, A. Dehghan, and M. Shah. GMCP-Tracker: Global Multi-object Tracking Using Generalized Minimum Clique Graphs. In *Proc. ECCV*, 2012.

[39] L. Zhang, Y. Li, and R. Nevatia. Global Data Association for Multi-Object Tracking Using Network Flows. In *Proc. CVPR*, 2008.