Bundle Adjustment for Multiple Object Tracking - a Stereo-based Approach

Master's Thesis Presentation

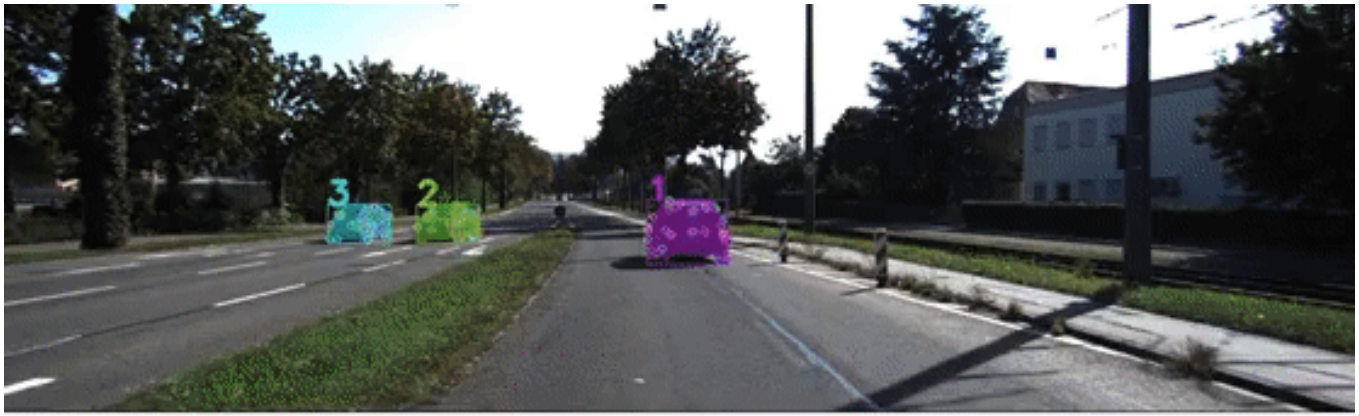
Anselm Coogan

March 31st, 2021

Technische Universität München

BAMOT: Bundle Adjustment for Multiple Object Tracking

Demo



Motivation

- vital for autonomous navigation $\rightarrow$ dynamic map of environment
- static SLAM benefits from tight integration with 3D MOT

Related Work

Dynamic SLAM systems

- dynamic environments are difficult for *classical* SLAM systems
- solutions to improve static SLAM performance:
  - detect and remove non-static elements, e.g. *RGBD-SLAM in Dynamic Environments*[1]
  - detect and track dynamic objects, e.g. *CubeSLAM*[2]

[1] Li et al., "RGB-D SLAM in Dynamic Environments Using Static Point Weighting", *IEEE Robot. Autom. Lett.* 2017.

[2] Yang et al., "CubeSLAM", *IEEE Trans. Robot.* 2019.

MOT Metrics

- What constitutes a *good* tracker?
  - detection ability
  - detection accuracy
  - tracking accuracy
- Which tracker performance metrics exist?
  - CLEAR[3]: MOTA, MOTP
  - IDF1[4] & Track-mAP[5]
  - HOTA[6]

---

[3] Bernardin et al., "Evaluating Multiple Object Tracking Performance", *EURASIP J. Image Video Process.* 2008.

[4] Ristani et al., "Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking", *ECCV*, 2016.

[5] Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge", *Int. J. Comput. Vis.* 2015.

[6] Luiten et al., "HOTA", *Int. J. Comput. Vis.* 2021.

3D MOT

- (dense) stereo-based systems
  - *CIWT*[7]
  - *MOTSFusion*[8]
  - use 3D info to improve 2D MOT
- *AB3DMOT*[9]
  - *custom* 3D MOT evaluation (+ LiDAR method)
- other LiDAR methods[10]

---

[7] Osep et al., "Combined Image- and World-Space Tracking in Traffic Scenes", *ICRA*, 2017.

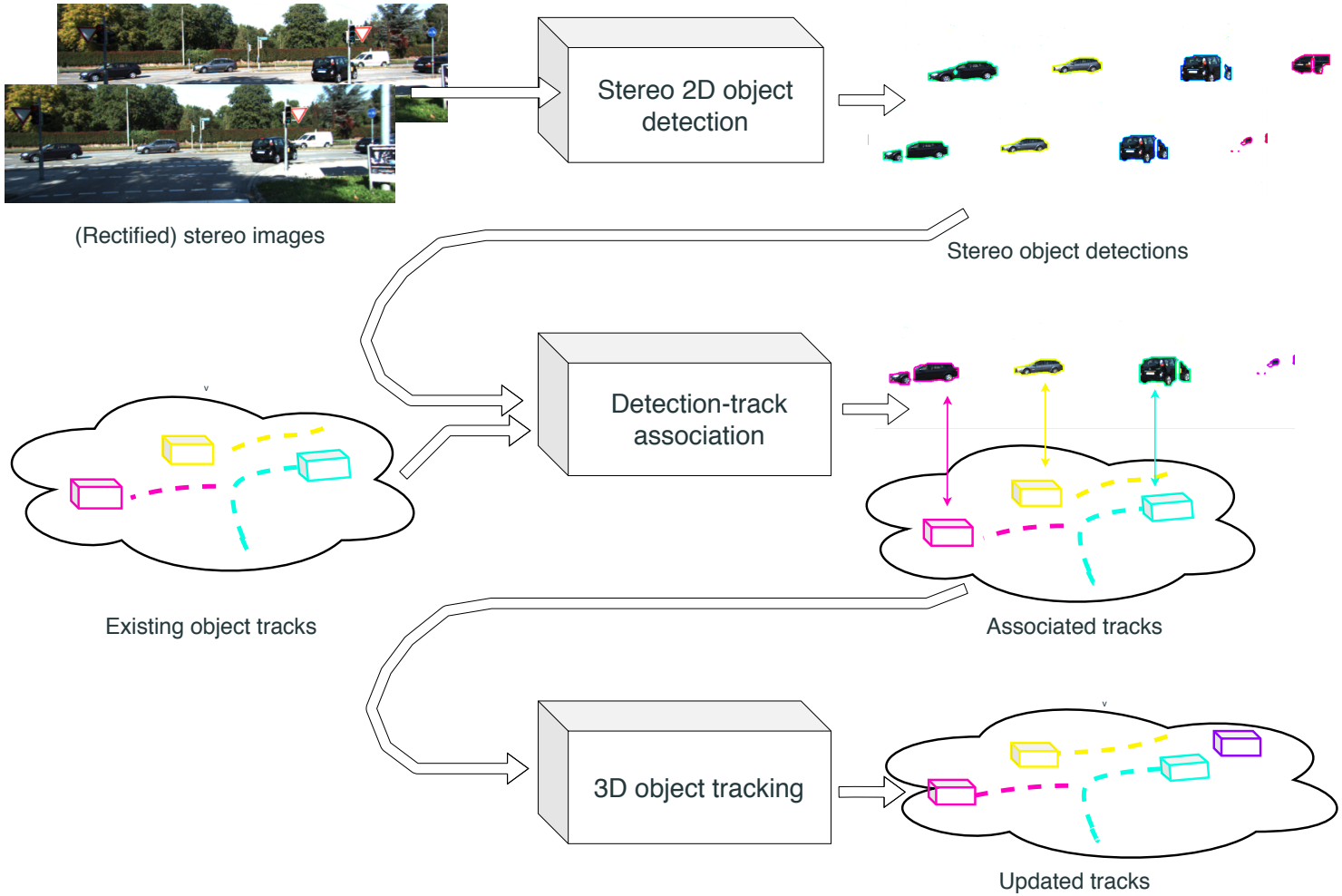
[8] Luiten et al., "Track to Reconstruct and Reconstruct to Track", *IEEE Robot. Autom. Lett.* 2020.

[9] Weng et al., "3D Multi-Object Tracking: A Baseline and New Evaluation Metrics", *IROS*, 2020.

[10] Yin et al., "Center-Based 3D Object Detection and Tracking", *arXiv:2006.11275 [cs]*, 2021; Chiu et al., "Probabilistic 3D Multi-Object Tracking for Autonomous Driving", *arXiv:2001.05673 [cs]*, 2020.

Method

Overview

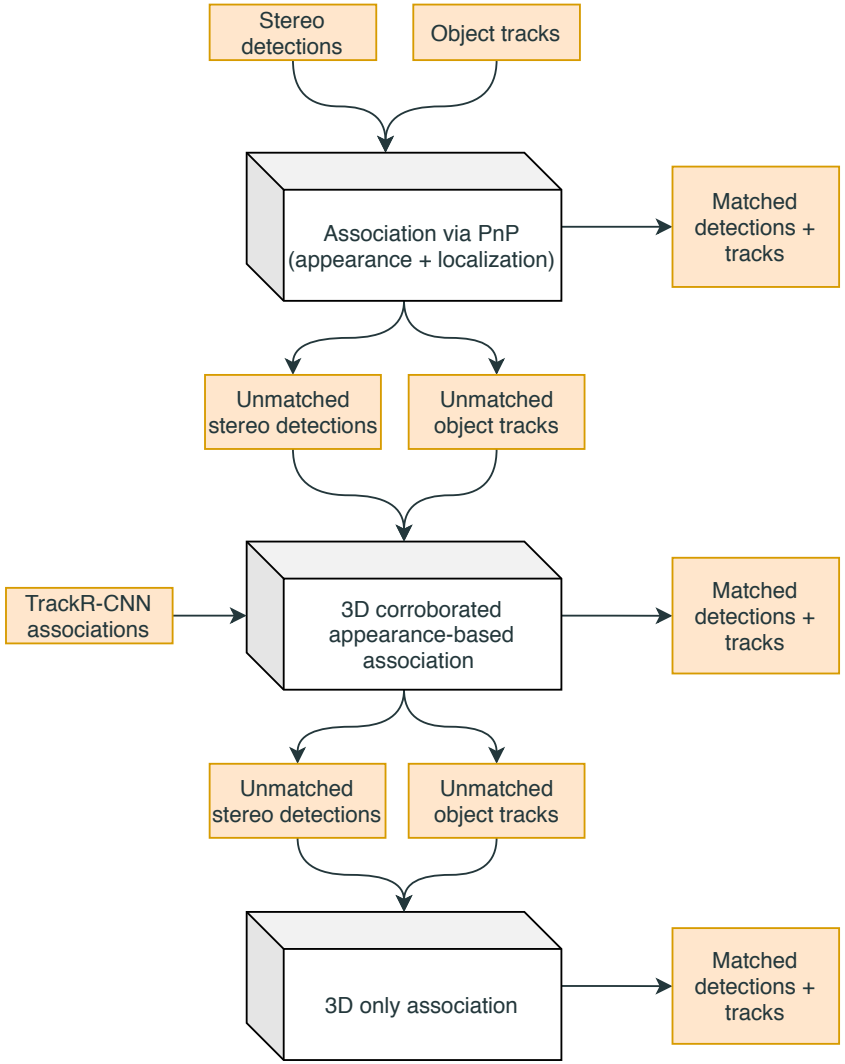

Stereo 2D object detection
(Rectified) stereo images
Stereo object detections
Detection-track association
Existing object tracks
Associated tracks
3D object tracking
Updated tracks

”Stereo” Object Detections

- pseudo-stereo detections via matched monocular 2D detections from TrackR-CNN[11]
- similarity
  - 2D location not applicable
  - appearance based $\rightarrow$ match features

[11] Voigtlaender et al., “MOTS”, *CVPR*, 2019.

Detection-Track-Association



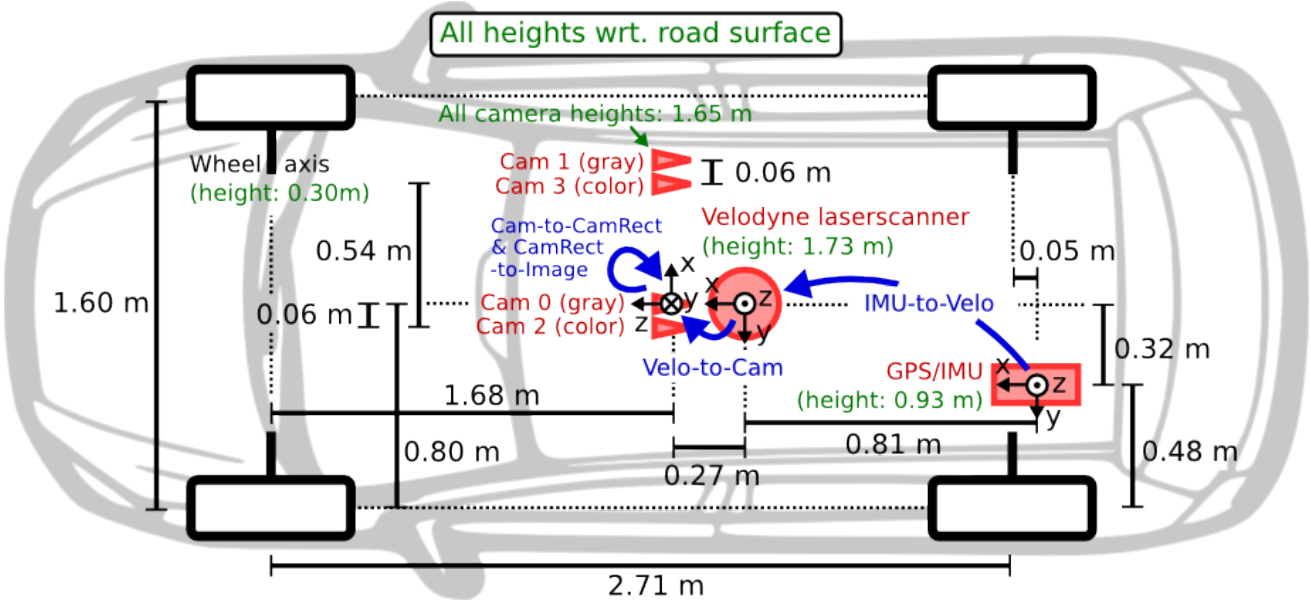
3D Object Tracking

- ORB features[12]
- PnP for object pose estimation
- object-level Bundle Adjustment w/ constant motion error term
- landmark culling based on known object shape

[12] Rublee et al., "ORB", *ICCV*, 2011.

Results

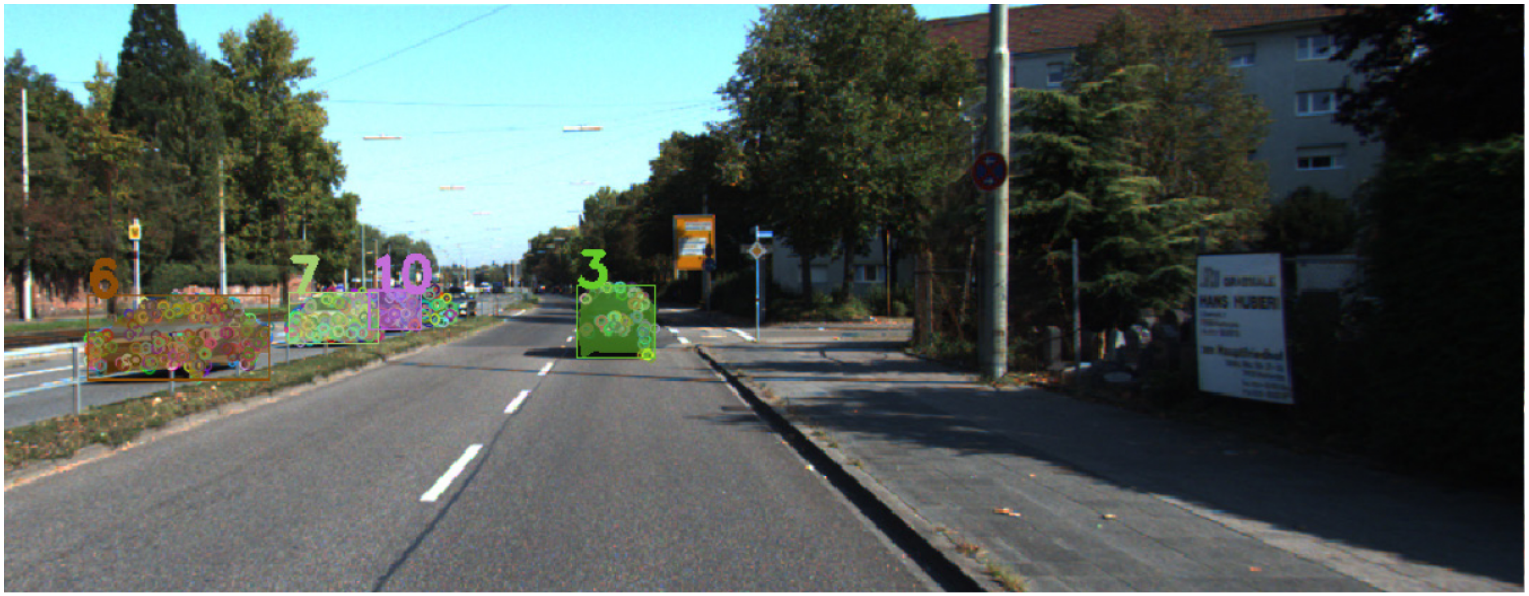
KITTI[13]

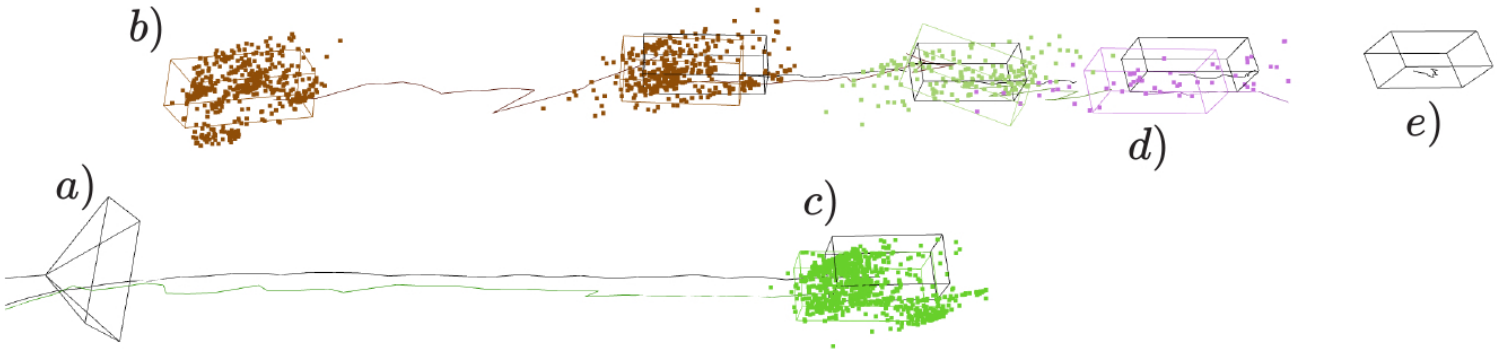


- 21 training scenes
- 29 testing scenes
- input: rectified stereo frames + ego motion

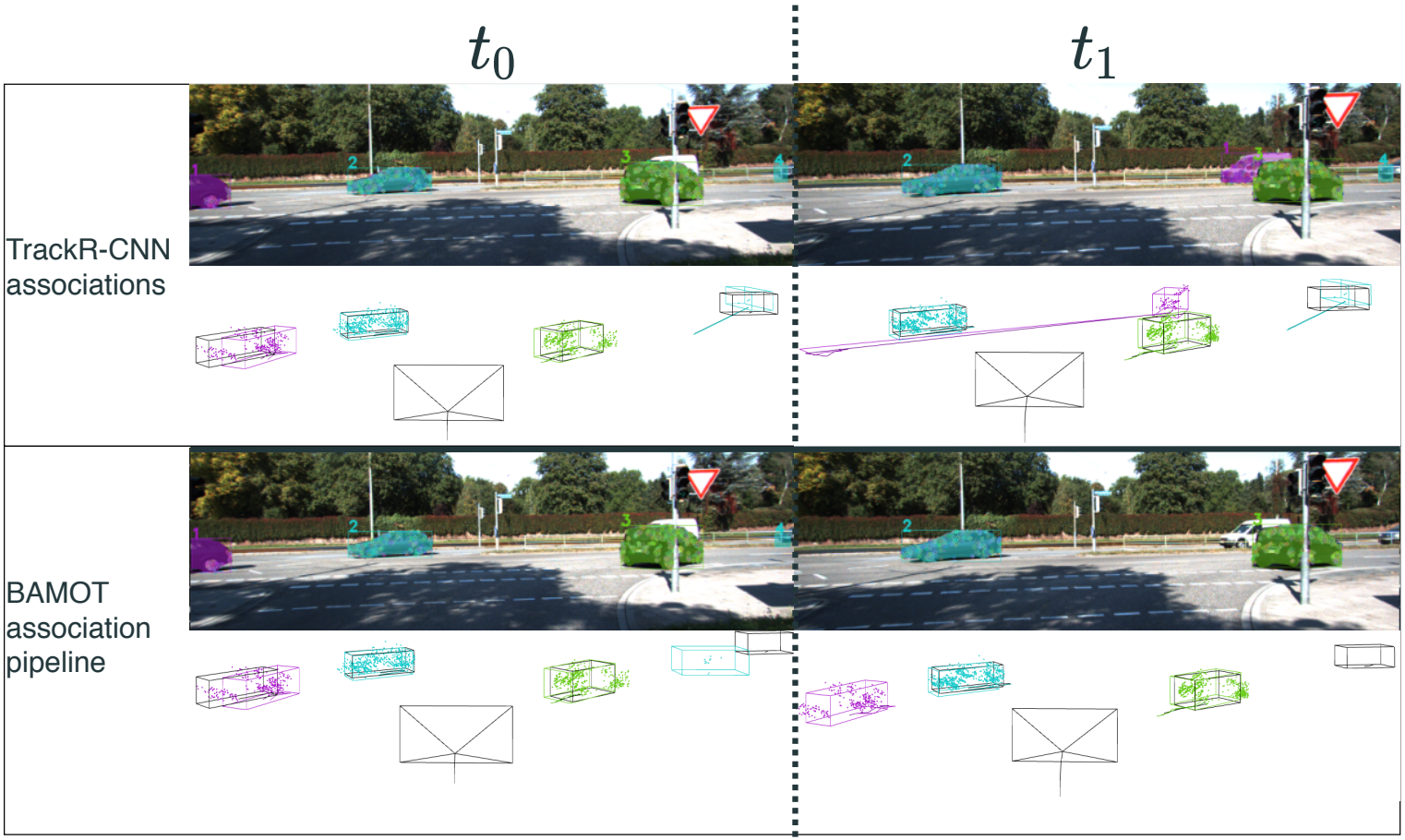
[13] Geiger et al., "Vision Meets Robotics", *Int. J. Robot. Res.* 2013.

Qualitative results: Example

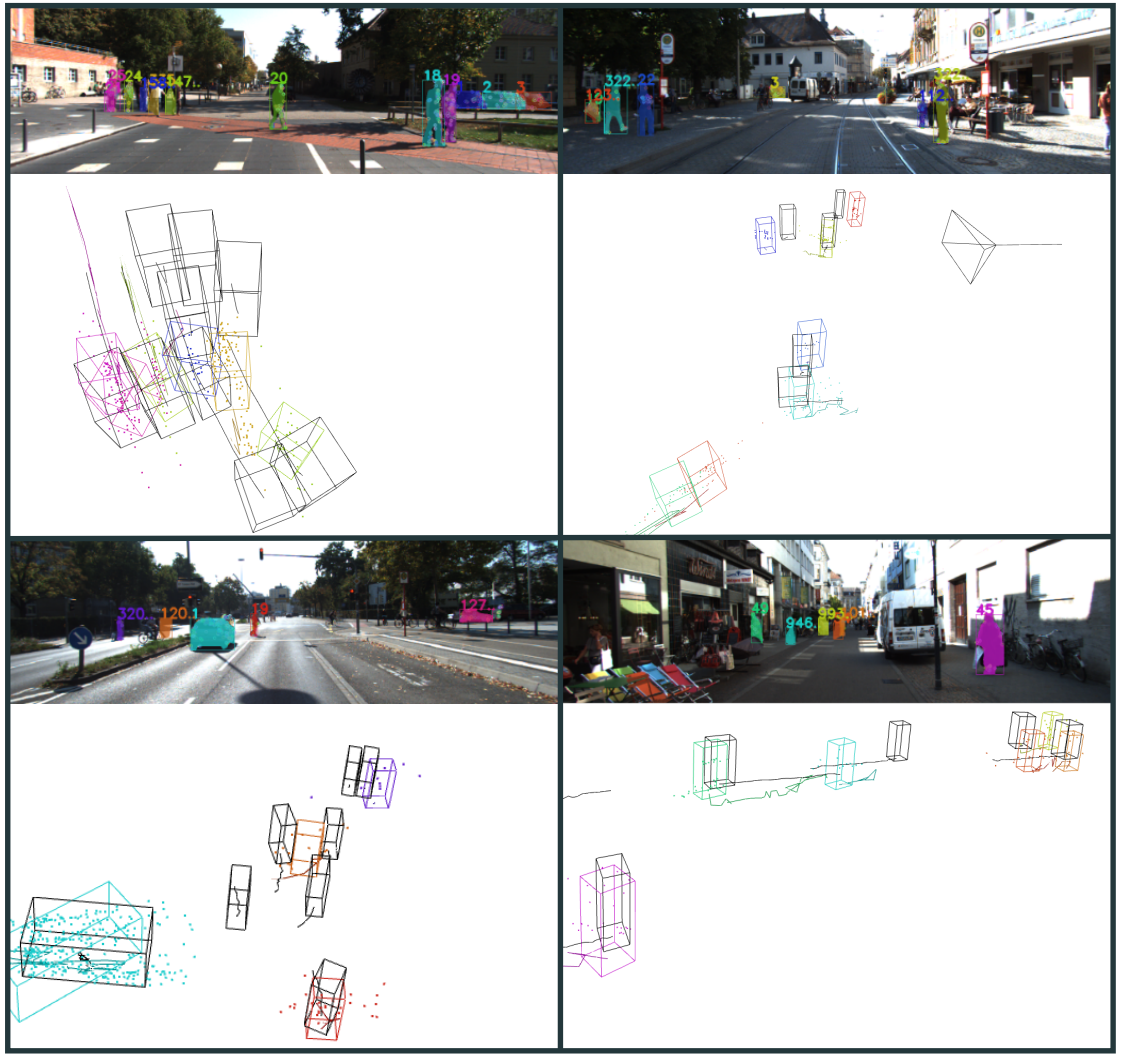




Qualitative results: Improved tracking

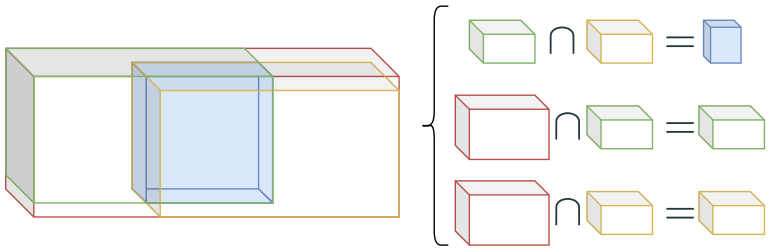


Qualitative results: Pedestrians



Quantitative Evaluation procedure

- parameters optimized for HOTA
- LiDAR-based AB3DMOT[14] as comparison
- similarity: Generalized Intersection-over-Union



$$\mathrm{IoU} = \frac{\square \cap \square}{\square \cup \square}$$

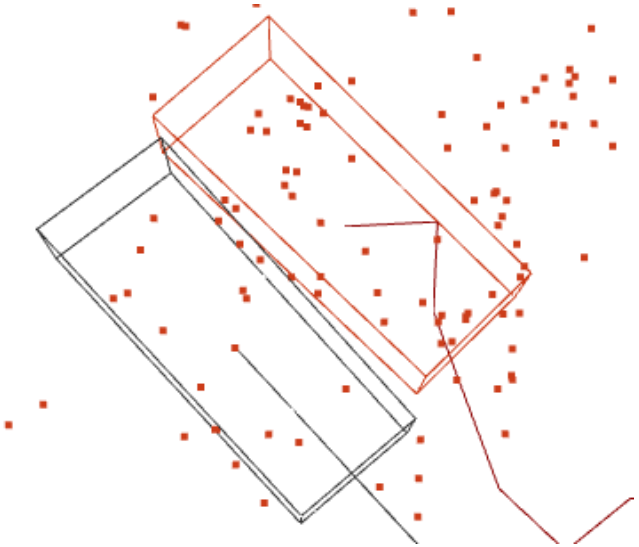
$$\mathrm{GIoU} = \mathrm{IoU} - \frac{\square \setminus (\square \cup \square)}{\square}$$

[14] Weng et al., "3D Multi-Object Tracking: A Baseline and New Evaluation Metrics", *IROS*, 2020.

Why GIoU?

$\mathrm{IoU} = 0$

$\text{normalized GIoU} \approx 0.5$



BAMOT vs AB3DMOT

- *BAMOT*
  - sparse feature point-based triangulation from 1.4 MP RGB cameras
  - triangulation accuracy decreases rapidly at approx. 20 times stereo baseline[15] ($\rightarrow$ 10m)
- *AB3DMOT*
  - dense LiDAR input (1.3 million points/sec)
  - 120m range
  - 2cm distance accuracy
  - 0.09 degree angular resolution

[15] Mur-Artal et al., "ORB-SLAM", *IEEE Trans. Robot.* 2015.

Quantitative results: BAMOT vs AB3DMOT

- Detection accuracy (DetA) → detection ability
- Localization accuracy (LocA) → detection precision
- Association accuracy (AssA) → tracking ability

	HOTA	DetA	AssA	LocA
BAMOT (car)	40.7	35.8	48.2	66.6
AB3DMOT (car)	71.3	65.8	77.8	88.5
BAMOT (ped.)	19.7	22.7	18.4	65.0
AB3DMOT (ped.)	55.5	55.3	55.9	82.5

BAMOT vs AB3DMOT using IoU

	HOTA	DetA	AssA	LocA
BAMOT (car, IoU)	22.9	18.3	30.4	61.5
AB3DMOT (car, IoU)	66.3	63.3	70.4	82.1
BAMOT (car, GIoU)	40.7	35.8	48.2	66.6
AB3DMOT (car, GIoU)	71.3	65.8	77.8	88.5

	HOTA	DetA	AssA	LocA
BAMOT (ped, IoU)	10.6	9.83	12.3	63.9
AB3DMOT (ped, IoU)	46.6	45.4	48.4	74.5
BAMOT (ped., GIoU)	19.7	22.7	18.4	65.0
AB3DMOT (ped., GIoU)	55.5	55.3	55.9	82.5

BAMOT using TrackR-CNN vs GT masks

	HOTA	DetA	AssA	LocA
BAMOT (car, TrackR-CNN)	40.7	35.8	48.2	66.6
BAMOT (car, GT masks)	43.7	37.3	53.0	67.8
BAMOT (ped., TrackR-CNN)	19.7	22.7	18.4	65.0
BAMOT (ped., GT masks)	27.7	24.6	31.9	65.1

Ablation study

	HOTA	DetA	AssA	LocA
BAMOT (car)	40.7	35.8	48.2	66.6
no impr. assoc. (car)	37.2 ↓	35.9 ↑	41.2 ↓	66.4 ↓
no const. motion (car)	40.4 ↓	36.1 ↑	47.1 ↓	66.9 ↑
no track retention (car)	39.1 ↓	35.8 -	44.8 ↓	66.9 ↑
no robust init. (car)	39.1 ↓	35.9 ↑	44.6 ↓	66.0 ↓
none of the above (car)	35.0 ↓	34.9 ↓	37.3 ↓	65.6 ↓
BAMOT (ped.)	19.7	22.7	18.4	65.0
no impr. assoc. (ped.)	20.7 ↑	24.3 ↑	18.1 ↓	64.2 ↓
no const. motion (ped.)	13.7 ↓	11.5 ↓	17.0 ↓	63.6 ↓
no track retention (ped.)	11.6 ↓	11.4 ↓	12.5 ↓	65.4 ↑
no robust init. (ped.)	14.6 ↓	12.0 ↓	18.6 ↑	63.5 ↓
none of the above (ped.)	15.5 ↓	12.4 ↓	19.8 ↑	64.0 ↓

Conclusion

Conclusion & Future work

- contribution
  - sparse, vision-based 3D MOT system
  - promising results w.r.t. static objects
  - 3D MOT evaluation (HOTA + 3D GIoU)
- future work
  - 3D MOT comparison against other vision-based methods necessary
  - integration with static SLAM system

Thank you!

Thesis + code:

github.com/AnselmC/bamot

References

- Bernardin, Keni et al. “Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics”. en. In: *EURASIP J. Image Video Process.* 2008 (2008), pp. 1–10. ISSN: 1687-5176, 1687-5281. DOI: 10.1155/2008/246309.
- Chiu, Hsu-kuang et al. “Probabilistic 3D Multi-Object Tracking for Autonomous Driving”. en. In: *arXiv:2001.05673 [cs]* (Jan. 2020). arXiv: 2001.05673 [cs].

- Geiger, A et al. “Vision Meets Robotics: The KITTI Dataset”. en. In: *Int. J. Robot. Res.* 32.11 (Sept. 2013), pp. 1231–1237. ISSN: 0278-3649, 1741-3176. DOI: 10.1177/0278364913491297.
- Li, Shile et al. “RGB-D SLAM in Dynamic Environments Using Static Point Weighting”. In: *IEEE Robot. Autom. Lett.* 2.4 (Oct. 2017), pp. 2263–2270. ISSN: 2377-3766, 2377-3774. DOI: 10.1109/LRA.2017.2724759.
- Luiten, Jonathon et al. “HOTA: A Higher Order Metric for Evaluating Multi-Object Tracking”. en. In: *Int. J. Comput. Vis.* 129.2 (Feb. 2021), pp. 548–578. ISSN: 0920-5691, 1573-1405. DOI: 10.1007/s11263-020-01375-2.

- Luiten, Jonathon et al. “Track to Reconstruct and Reconstruct to Track”. In: *IEEE Robot. Autom. Lett.* (Apr. 2020). DOI: 10.1109/LRA.2020.2969183. arXiv: 1910.00130.
- Mur-Artal, Raul et al. “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”. In: *IEEE Trans. Robot.* 31.5 (Oct. 2015), pp. 1147–1163. ISSN: 1552-3098, 1941-0468. DOI: 10.1109/TRO.2015.2463671. arXiv: 1502.00956.
- Osep, A. et al. “Combined Image- and World-Space Tracking in Traffic Scenes”. In: *ICRA*. May 2017, pp. 1988–1995. DOI: 10.1109/ICRA.2017.7989230.

- Ristani, Ergys et al. “Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking”. In: *ECCV* (Sept. 2016). arXiv: 1609.01775.
- Rublee, Ethan et al. “ORB: An Efficient Alternative to SIFT or SURF”. en. In: *ICCV*. Barcelona, Spain, IEEE, Nov. 2011, pp. 2564–2571. ISBN: 978-1-4577-1102-2 978-1-4577-1101-5 978-1-4577-1100-8. DOI: 10.1109/ICCV.2011.6126544.
- Russakovsky, Olga et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *Int. J. Comput. Vis.* 115 (Jan. 2015), pp. 211–252. arXiv: 1409.0575.

- Voigtlaender, Paul et al. “MOTS: Multi-Object Tracking and Segmentation”. en. In: *CVPR*. Long Beach, CA, USA, IEEE, June 2019, pp. 7934–7943. ISBN: 978-1-72813-293-8. DOI: 10.1109/CVPR.2019.00813.
- Weng, Xinshuo et al. “3D Multi-Object Tracking: A Baseline and New Evaluation Metrics”. In: *IROS* (2020).
- Yang, Shichao et al. “CubeSLAM: Monocular 3D Object SLAM”. en. In: *IEEE Trans. Robot.* 35.4 (Aug. 2019), pp. 925–938. ISSN: 1552-3098, 1941-0468. DOI: 10.1109/TRO.2019.2909168. arXiv: 1806.00557.
- Yin, Tianwei et al. “Center-Based 3D Object Detection and Tracking”. en. In: *arXiv:2006.11275 [cs]* (Jan. 2021). arXiv: 2006.11275 [cs].