JAN 4-8 WACV 2024 WAIKOLOA HAWAII

Paper ID #1810

So you think you can track?

Dept. of Computer Science and the Institute for Software Integrated Systems at Vanderbilt University

Derek Gloudemans, Gergely Zachár, Junyi Ji, Yanbing Wang, Matt Nice, William Barbour, Matt Bunting,
Benjamin Seibold, Benedetto Piccoli, Maria Laura delle Monache, Alexandre Bayen, Jonathan Sprinkle, Daniel B. Work

I-24 MOTION | ISIS Institute for Software Integrated Systems | VANDERBILT UNIVERSITY

Motivation

- Existing tracking algorithms and benchmarks are short in duration and are thus tailored towards high localization accuracy, precision, and recall, underemphasizing the challenges of long-term occlusions, appearance changes, and increasing chance of fragmentations and ID swaps with increasing track length.
- In the field of traffic science, high-quality *trajectory data* (tracking data for every vehicle in a traffic flow) is critical to understand microscopic traffic phenomena and the impact of mixed autonomy in traffic
- This data is extremely labor-intensive to produce at large scales (minutes and miles of data) by manual annotation, and GPS instrumentation does not capture every vehicle.
- Such data requires high tracking performance over long time durations; a new benchmark is needed to develop algorithms targeting long-term multi camera tracking performance.

The I-24 Video Dataset

- Using the I24-MOTION camera system, we produce largest multi-camera video dataset:
  - 234 cameras and 234 hours of video covering an 8-10 lane interstate roadway scene
  - high object density (>500 objects typically present) and long object durations (6.6 minutes, 11,880 frames average)
- We create sparse set of 270 GPS-produced annotations corresponding to 1782 minutes of labeled vehicle trajectory.
- We provide precise scene information utilizing novel methods for precisely re-aligning camera homographies to account for slight drift over time, outperforming existing image stabilization techniques in over 99% of cases.
- We define a unified curvilinear coordinate system aligned along the roadway direction of travel for the entire 4.2-mile roadway section, useful for filter-based tracking and downstream traffic science.

Example Video

Data Access

I-24 MOTION Camera System

- I-24 Mobility Technology Interstate Observation Network (MOTION) is a 4.2-mile section of Interstate-24 in Nashville, Tennessee, USA equipped with 294 4K cameras densely covering 4 miles of interstate roadway.
- This system offers an unprecedented opportunity to solve massively multi-camera tracking problems and to produce vehicle tracking datasets at new scales.
- This dataset uses a subset of 234 cameras from the I-24 MOTION system to produce a multi-camera video dataset.

I24-MOTION System

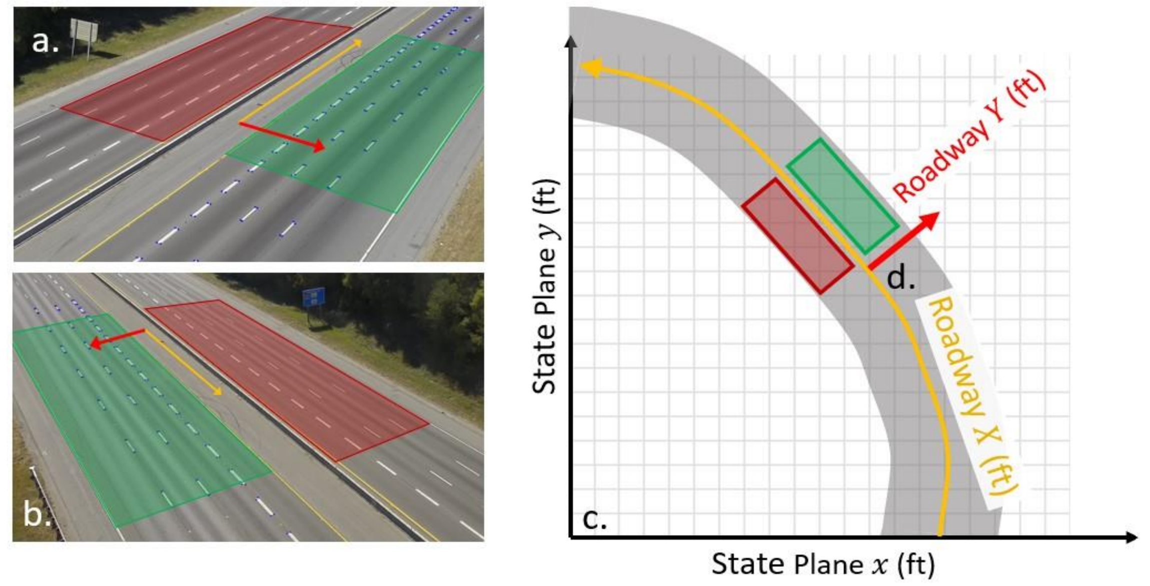
Close-up of 110-foot-tall camera poles along the interstate

Six-camera mount (at top of one pole)

Dataset Comparison

Dataset	Cameras	Video (hr)	Scene (min)
DukeMTMC	8	11.3	85
Wildtrack	7	1.0	8.6
CityFlow	25	3.3	6.5
Synthehicle	7	17	3
EPFL-Terrace	4	14	3.5
PETS	8	0.2	0.3
pNEUMA Vision	10	3.9	13
I24-3D	17	1.0	1.5
I24-Video (proposed)	234	234	60

Existing multi-camera datasets by number of *Cameras* covering a single scene, total hours of *Video*, and duration of a typical *Scene*.

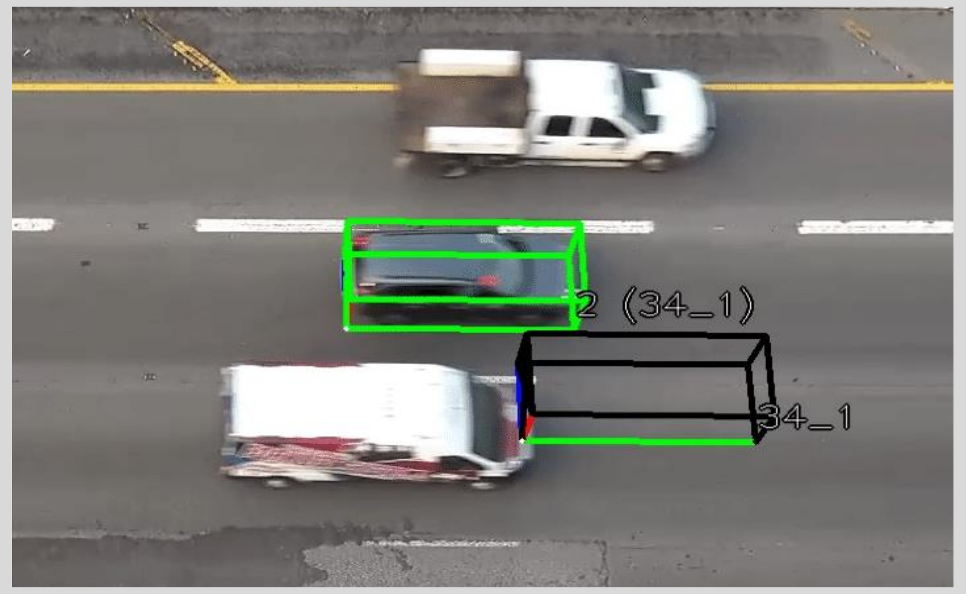


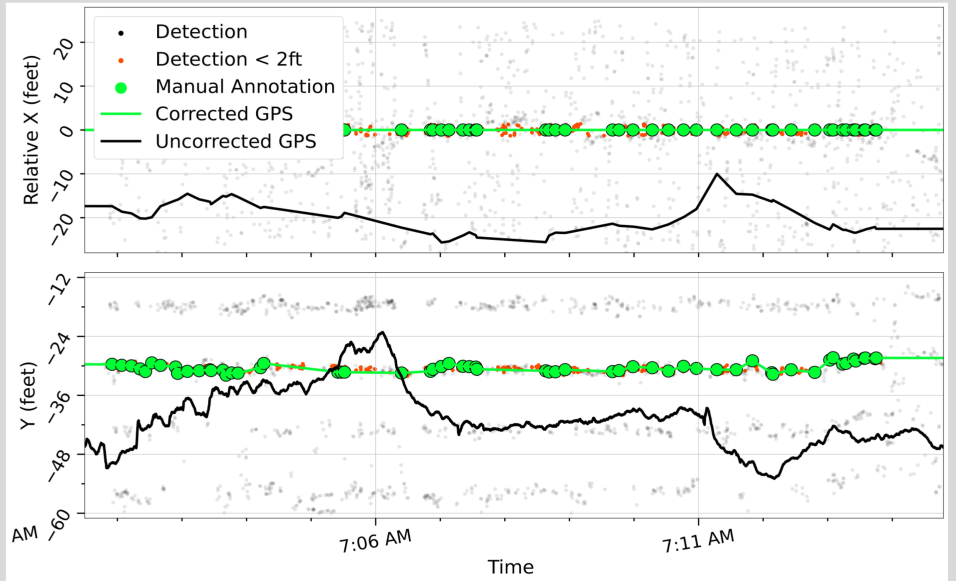
Two camera fields of view (a) and (b) are mapped to the same unified coordinate system. The X-axis is always aligned with the roadway direction of travel and the Y-axis is always perpendicular.

Example fields of view from each of the 234 cameras covering the scene in the I24-Video Dataset.

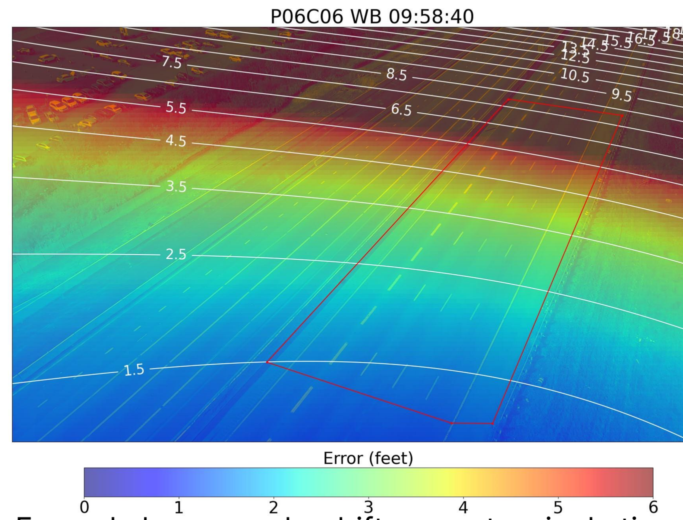
GPS Trajectories

- Concurrent with video recording, a fleet of 103 GPS instrumented vehicles was driven through the testbed and a total of 270 vehicle trajectories were recorded.
- GPS trajectories contained bias and variance in positional data (black box/curve), which was corrected by rectifying each trajectory against vehicle positions manually annotated in video data (light green box/curve shows result)



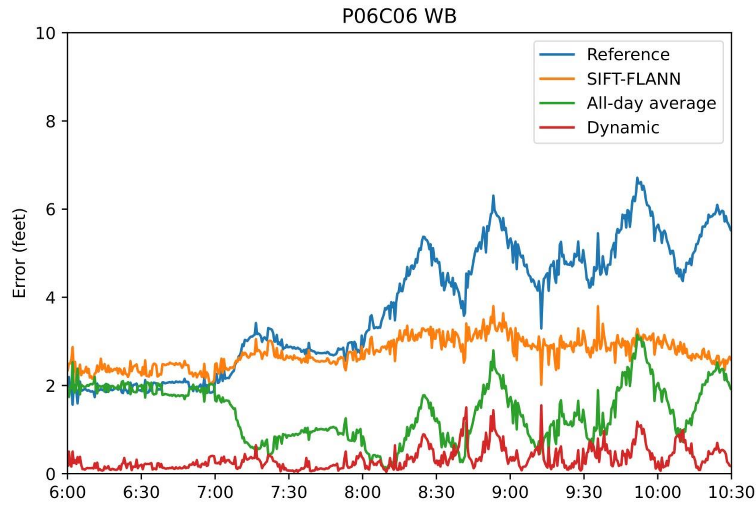


Homography Realignment

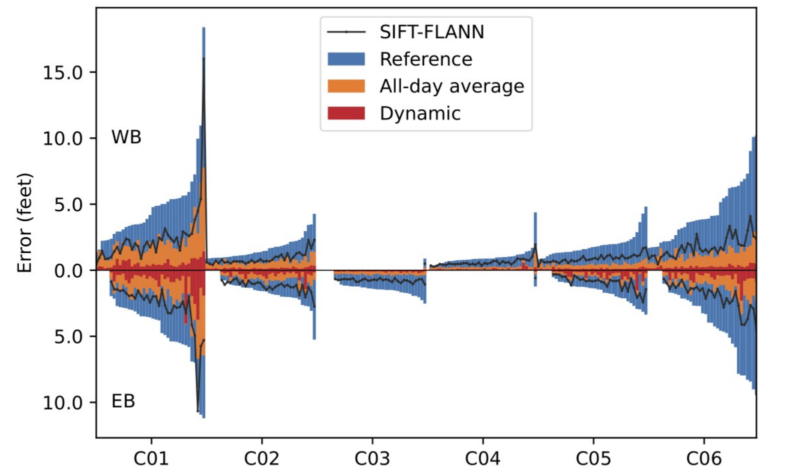


Example homography drift error at a single time

- Camera fields-of-view are constantly changing due to differential heating of the sun and shade-facing sides of the pole (the *Sunflower Effect*), producing errors often over 10 feet in scene homographies.
- We develop an automatic method for re-aligning each camera's homography using re-detection of semantically meaningful lane markings (rather than semantically meaningless SIFT-based features used in the existing SOTA method.)
- The proposed methods (an all-day static average and a dynamic homography based on re-detected lane, reduce the mean error of 2.78 feet to 1.03ft and 0.33ft average error, respectively, and outperform the existing method in 88.1% and 99.7% of cases



Homography error after correction, relative to uncorrected reference, over entire scene (one camera)



Mean all-day error for each camera before and after correction

Benchmarking Experiments

- A limited set of detection-fusion algorithms was benchmarked on the dataset
- An *Oracle* tracker was run by selecting the best object detections available corresponding to each trajectory (to disentangle tracking performance issues from detection failures)
- HOTA is low for all trackers; driven primarily by low AssA scores. This indicates that object tracklets are not strongly persistent (this is also supported by the relatively low LCSS and mean tracklet durations compared to the 6.6 minute mean trajectory length, and high average IDs per ground truth)
- By comparison, Oracle tracker HOTA is fairly high, indicating that the detection set itself is suitable to provide precise object positions given a well-designed tracker
- Thus, a performance gap exists that must be filled before such trajectory data is useful for downstream traffic science
- A more thorough benchmarking of other methods (especially offline methods) is required.

Tracker	HOTA	DetA	AssA	Recall	IDs/GT ↓	$LCSS_t$ (s)	$LCSS_d$ (ft)	$MOTP_i$	$MOTP_e$ (ft) ↓	TD (s)
SORT [4]	**9.5**	51.3	**1.8**	73.6	53.1	**51.9**	2609	68.0	**2.70**	12.3
IOU [5]	1.1	7.4	0.2	20.4	60.0	16.8	53.2	36.7	7.31	8.4
KIOU [5, 10]	8.5	51.2	1.4	73.9	**47.9**	40.6	2181	66.9	2.72	**15.1**
ByteTrack (L2) [55]	**9.5**	51.5	**1.8**	73.6	53.3	51.5	**2575**	**70.0**	2.71	12.4
ByteTrack (IOU) [55]	8.5	**53.1**	1.4	**75.9**	50.3	44.1	2390	67.1	2.72	14.9
Oracle	53.1	55.1	51.0	86.4	1.2	636	14699	75.3	2.53	690

Tracking results for implemented tracking algorithms.

Acknowledgements

This work is supported by NSF grant Nos. 2135579 and DGE-1937963, USDOT Grant Nos. 693JJ32245006 and 693JJ322NF5201, DOE award No. CID DE-EE0008872, and CMAQ award No. TN20210003. This work is conducted with support from the Tennessee Department of Transportation. The views expressed herein do not necessarily represent the views of the Tennessee Department of Transportation, U.S. Department of Energy, or the U.S. Government.