The I-24 Multi-Camera 3D Dataset

A new benchmark for 3D multi-camera vehicle tracking

Dept. of Computer Science at Vanderbilt University, Institute for Software Integrated Systems at Vanderbilt University
Derek Gloudemans, Gracie Gumm, Yanbing Wang, William Barbour and Daniel B. Work

BMVC 2023
Paper ID #15

Motivation

- High-quality *trajectory data* (positional 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.
- Automatic methods (object detection and tracking algorithms) offer a tremendous but underexploited means to produce vehicle trajectory data.
- Few benchmarking datasets exist on which to develop, train, and evaluate object trackers for precise 3D vehicle tracking across many cameras.
- Only the Synthehicle dataset, developed in parallel with this work, offers the necessary dataset attributes.

Dataset Comparison

Dataset	Resolution	Detection		MOT		MCT		Boxes	Frames	Cameras
		2D	3D	2D	3D	2D	3D			
WILDTRACK	1920×1080	✓		✓		✓	✓	38k	61k	7
KITTI	1382×512	✓	✓	✓	✓			200k	15k	1
NuScenes	1600×900	✓	✓	✓	✓	✓	✓	12M	40k	6
BoxCars116k	*varies*	✓	✓					116k	116k	1
UA-DETRAC	1920×1080	✓		✓				1.2M	140k	1
CityFlow	960×540	✓		✓		✓		229k	117k	25*
Synthehicle	1920×1080	✓	✓	✓	✓	✓	✓	4.62M	6.7k	7
I24-3D (Ours)	3840×2160	✓	✓	✓	✓	✓	✓	877k	105k	16-17

Comparison by suitable uses (✓) for Multiple Object Tracking (*MOT*) and Multi-Camera Tracking (*MCT*), number of annotated bounding *Boxes*, total number of *Frames*, and total number of *Cameras* covering a scene.

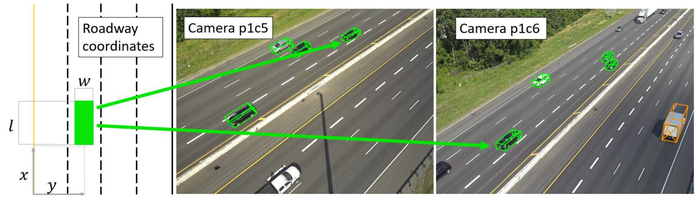
Dataset Attributes

- Video data for 3 *scenes,* each consisting of video from 16-17 partially overlapping cameras
- 3D tracking bounding box annotations for 6 classes of vehicles, over 877,000 in total
- Vehicles are manually annotated in a shared roadway coordinate space
- Scene homography allowing lossless conversion from image coordinates to 3D space

Scene	Time (s)	Cameras	Frames	Boxes	IDs	VMT	Description
1	90	17	45900	291k	324	118	Free-flow traffic
2	60	16	30600	146k	114	24.4	Slow traffic, snow conditions
3	60	16	28800	440k	282	67.0	Congested traffic
Total	210	-	105300	877k	720	209	-

Scene details , including duration, unique IDs, and total vehicle miles traveled.

Examples for each vehicle class, and total annotation counts for each class



Homography example – the same bounding box in 3D space (*left*, green solid rectangle), can be projected losslessly into multiple camera fields of view covering that portion of the roadway (*right*).

Benchmarking Experiments

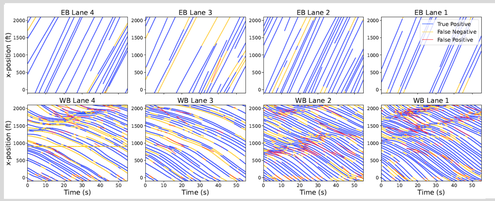
Detector	Tracker	DF	TF	HOTA	MOTA	Rec	Prec	GT%	Pred%	MT	ML	Sw/GT
Crop	Byte	✓	✓	23.6	21.3	53.4	64.0	90.5	72.9	25.6	25.0	1.1
Crop	KIOU	✓	✓	24.6	21.4	54.4	64.2	90.5	71.2	27.6	22.3	1.1
Dual3D	Byte	✓	✓	30.9	50.0	65.6	81.9	90.6	93.4	35.9	15.0	0.9
Dual3D	KIOU	✓	✓	39.7	71.6	76.5	93.7	91.5	95.3	52.5	10.4	0.7
Single3D	Byte	✓	✓	27.5	49.3	62.8	83.9	92.1	91.8	29.7	15.4	0.9
Single3D	KIOU	✓	✓	39.9	71.6	76.3	94.1	93.4	**95.6**	51.6	8.8	0.7
Crop	Byte		✓	15.2	-16.5	43.5	47.0	90.4	59.8	14.5	32.2	1.5
Crop	KIOU		✓	20.7	-2.1	51.6	52.9	90.2	53.7	25.5	26.4	1.5
Dual3D	Byte		✓	38.7	75.0	80.2	93.8	93.0	93.6	59.0	8.0	0.8
Dual3D	KIOU		✓	**44.8**	77.0	**83.0**	93.2	91.7	92.3	**63.8**	8.8	**0.5**
Single3D	Byte		✓	38.2	72.6	80.6	90.9	**94.9**	92.5	58.7	**4.7**	1.1
Single3D	KIOU		✓	**44.8**	**77.1**	**83.0**	93.4	93.3	91.3	62.2	7.8	**0.5**
Crop	Byte	✓		19.2	34.7	58.3	72.8	91.9	81.4	27.1	16.6	2.4
Crop	KIOU	✓		19.3	32.1	57.9	71.4	91.9	79.7	25.9	17.3	2.4
Dual3D	Byte	✓		20.9	60.2	64.2	94.8	92.7	94.7	29.5	8.7	2.8
Dual3D	KIOU	✓		21.1	60.4	64.0	95.2	92.6	94.7	29.7	8.9	2.7
Single3D	Byte	✓		21.3	60.3	63.9	95.3	93.8	94.2	27.1	7.5	2.6
Single3D	KIOU	✓		21.4	60.3	63.7	**95.5**	94.2	93.9	26.5	7.5	2.6
Crop	Byte			17.6	18.7	59.8	64.0	91.9	73.0	30.4	15.1	3.0
Crop	KIOU			16.9	10.8	57.5	60.0	91.9	66.0	28.2	18.5	3.2
Dual3D	Byte			15.0	55.1	72.8	81.7	93.2	87.5	42.5	6.8	7.3
Dual3D	KIOU			15.1	55.6	72.7	82.2	93.1	87.8	42.3	7.0	7.3
Single3D	Byte			15.1	54.0	72.3	80.8	94.3	85.9	40.5	5.8	7.2
Single3D	KIOU			15.2	54.4	72.2	81.3	94.5	86.1	39.4	5.6	7.1

Tracking results for each tracking pipeline over all scenes along a variety of metrics, including recall (*Rec),* precision (*Prec),* percentage of ground truths tracked by at least one object (*GT%),* and average identity switches per GT (*Sw/GT).* See paper for more details.

Scene	HOTA	MOTA	MOTP	Rec	Prec	GT%	Pred%	MT	ML	Sw/GT
1	**58.5**	**89.7**	69.2	**92.9**	**96.7**	**95.3**	**98.4**	**86.3**	**2.2**	**0.02**
2	46.9	77.7	**74.5**	86.2	91.1	90.4	82.4	64.0	9.6	0.49
3	29.1	63.5	64.8	69.9	91.7	89.3	96.1	40.9	14.6	1.05
avg	44.8	77.0	69.5	83.0	93.2	91.7	92.3	63.8	8.8	0.52

Results for best tracking pipeline, per individual scene.

- Existing object detectors, trackers, and multi-camera methods are combined into tracking pipelines to benchmark performance on the dataset.
- A best overall score of 44.8% HOTA is obtained, indicating substantial room for improvement



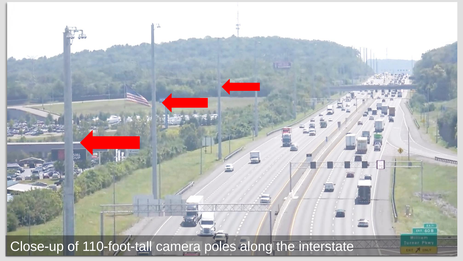
Scene 3 vehicle x-positions versus time for each lane. Blue indicates a correctly predicted vehicle position, yellow indicates a false negative, and red indicates a false positive. The farthest lane (WB Lane 4) has significantly more false negatives than more interior lanes, likely due to significant occlusion of these vehicles by vehicles in interior lanes.

Future Work

- Ground truth video dataset with many more (200+) cameras to allow massively multi-camera tracking algorithm development.
- Develop tracking algorithms explicitly well-suited for vehicle tracking informed by strong explicit motion models and scene understanding

I-24 MOTION Camera System

- I-24 Mobility Technology Interstate Observation Network (MOTION) is a four-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 18 cameras from the I-24 MOTION system to produce a multi-camera vehicle tracking dataset.


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


Six-camera mount (at top of one pole)

Example frames from each of the 18 cameras used in this work. For selected cameras (*red green, blue, orange)* the corresponding field of view is shown on an overhead view of the roadway coordinate space. The selected cameras densely cover a 2000-foot portion of the roadway with overlaps.

Learn More

Data Access | Example Video | I24-MOTION System

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1853913 and the USDOT Dwight D. Eisenhower Fellowship program under Grant No. 693JJ32345054. This work is conducted with support from the Tennessee Department of Transportation.