PathTrack: Fast Trajectory Annotation with Path Supervision

Santiago Manén / ETH Zürich Michael Gygli / ETH Zürich Dengxin Dai / ETH Zürich Luc Van Gool / ETHZ, KU Leuven

www.varcity.eu

1. Summary

Problem:

- Insufficient data to train fully data-driven trackers.

Goal:

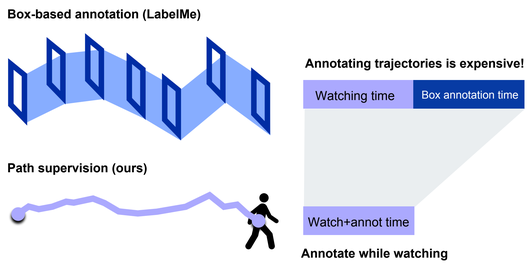
- *More data* to train Multi-Object Trackers (MOT).

Contributions:

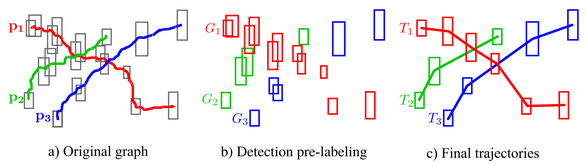
- Efficient way to annotate trajectories based on *path supervision*. It specially shines for quick quantity--over-quality data collection, ideal for training data.
- The new PathTrack MOT dataset provides abun--dant training data (from 720 videos) to learn fully data-driven trackers.
- Insights into MOT train data collection:
  - Saturation point not yet reached for MOT training data.
  - Quantity over quality to learning to link detections into trajectories.

2. Annotation with Path Supervision

Most trajectory annotation frameworks interpolate through a set of manually annotated boxes, which are time-consuming and expensive to obtain. Instead, we annotate rough *paths* by loosely following the objects in the scene with the cursor in an intuitive manner.



Paths represent a mode of weak-superivision on MOT trajectories. They provide no scale information and might be imprecise. So obtaining dense box-trajectories from them is not trivial.



Our energy minimization framework promotes the linkage of affine object detections while enforcing path constraints

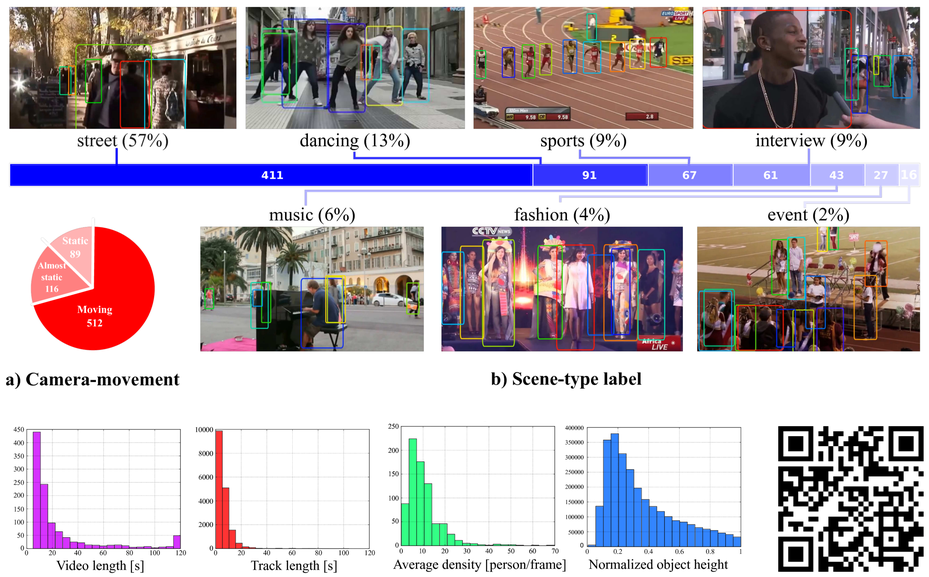
3. The PathTrack dataset

We use path supervision to annotate a MOT dataset of unprecedented scale (720 sequences). The path trajectories were crowdsourced, since they are intuitive and natural to annotate. We hope it encourages and supports richer and fully data-driven MOT systems.

Table 1: Comparison of PathTrack with other popular MOT datasets.

Dataset	Train			Test			Total			Classes (P = Person, C = Car)	Camera (S=Static M=Moving)
	# seqs	Duration (mins)	# tracks	# seqs	Duration (mins)	# tracks	# seqs	Duration (mins)	# tracks		
VirtualKITTI [1]	-	-	-	-	-	-	5	4	261	C	car-mounted
KITTI [2]	21	13		29	18	-	50	30	-	C + P	car-mounted
MOT15 [3]	11	6	500	11	10	721	22	16	1221	P	S+M
MOT16 [4]	7	4	512	7	4	830	14	8	1342	C+P	S+M
PathTrack (ours)	640	161	15,380	80	11	907	720	172	16,287	P	S+M

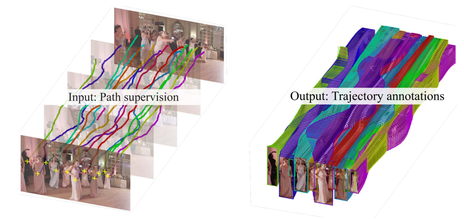
Apart from quantity, diversity is a central goal of the dataset. It includes 7 different kind of scenes taken from both stationary and moving cameras. Each sequence is labeled with its *scene-type* and *camera-movement*, allowing for fine-grained performance analysis.



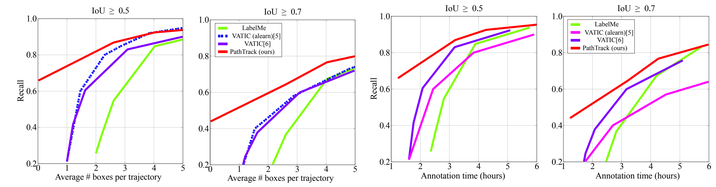
The full dataset can be downloaded from: http://people.ee.ethz.ch/~smanenfr/pathtrack/

5. Experiments

Annotation efficiency with path supervision

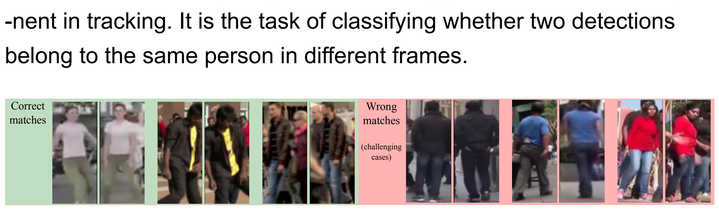


Annotating with path supervision is more efficient than competing methods for any annotation quality, particularly for quantity-over-quality annotations.

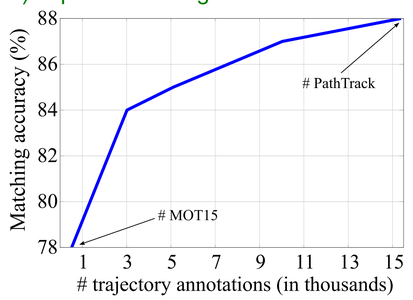


Insights into MOT data collection

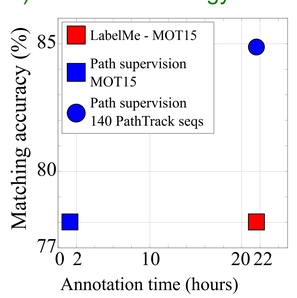
We use as case study the *person matching* problem, a critical compo--nent in tracking. It is the task of classifying whether two detections belong to the same person in different frames.



a) Impact of training data



b) Annotation strategy



The community can still benefit from *even more* training data for the matching problem. It can be efficiently collected with path supervision, e.g., the MOT15 data can be annotated 20 times faster while achieving the same matching accuracy.

Tracking results

First tracking results on the dataset demonstrate the usefulness of our training data.

LP Tracker trained on	MOTA ↑	MOTP ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Switch ↓
MOT15 [3]	24.5	81.4	44.2%	19.2%	42,502	37,720	1,827
PathTrack (ours)	27.6	81.5	47.3%	18.2%	40,614	36,508	1,576

7. References

[1] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig. Virtual worlds as proxy for multi-object tracking analysis. In CVPR, 2016.

[2] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In CVPR, 2012.

[3] L. Leal-Taixe, A. Milan, I. Reid, S. Roth, and K. Schindler. MOTChallenge2015: Towards a Benchmark for Multi-Target Tracking. arXiv:1504.01942[cs], 2015.

[4] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler. MOT16: A Benchmark for Multi-Object Tracking. arXiv:1603.00831[cs], 2016.

[5] C. Vondrick and D. Ramanan. Video Annotation and Tracking with Active Learning. In NIPS, 2011.

[6] C. Vondrick, D. Patterson, and D. Ramanan. Effciently Scaling Up Crowdsourced Video Annotation. IJCV, 2013.

The authors gratefully acknowledge support by Toyota and the European Research Council (ERC) under the project VarCity (#273940).