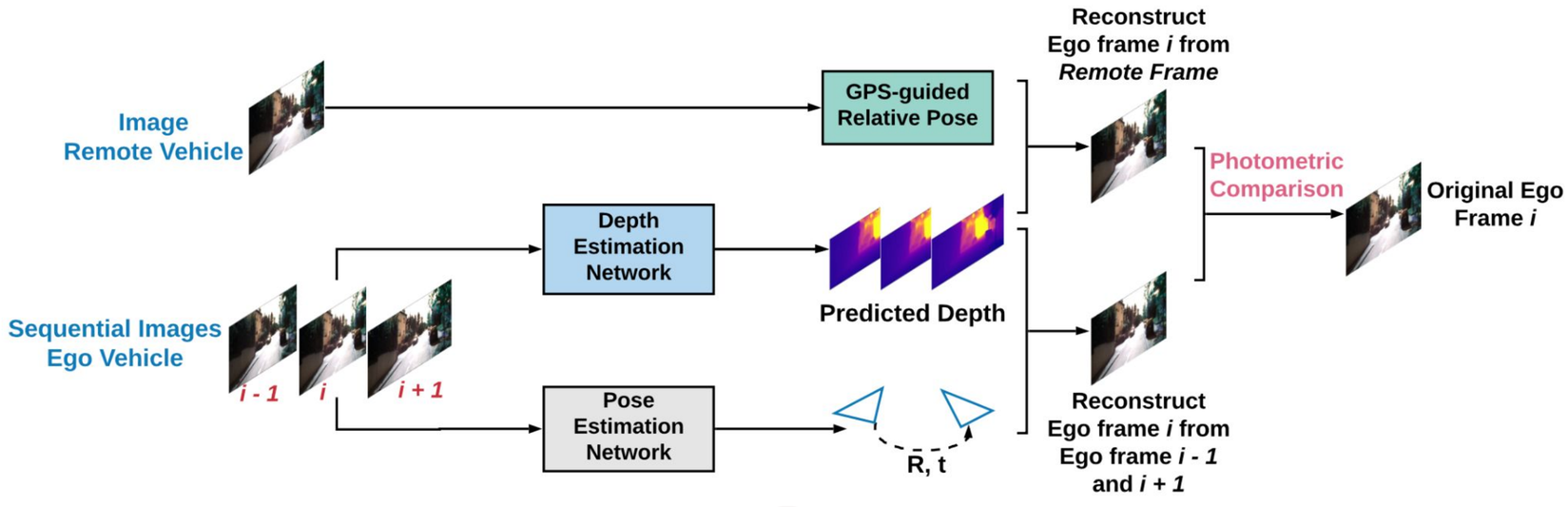
VISTA: *VI*rtual *ST*ereo based *A*ugmentation for Depth Estimation in Automated Driving

Bin Cheng, Kshitiz Bansal, Mehul Agarwal, Gaurav Bansal, Dinesh Bharadia

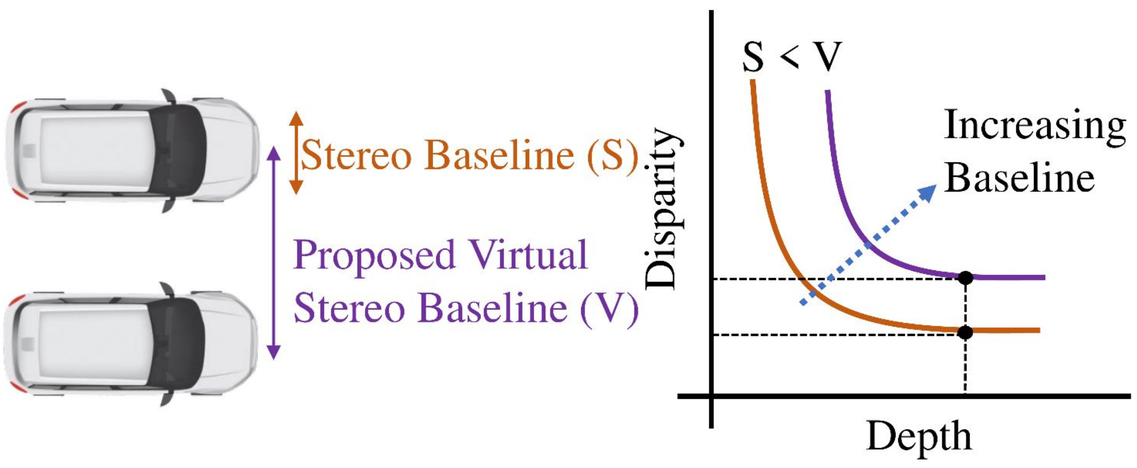


Motivation

- Classic LiDAR or stereo based depth estimation requires expensive setup but can produce accurate results
- Recent monocular depth estimation only needs simple configuration but the results are less accurate

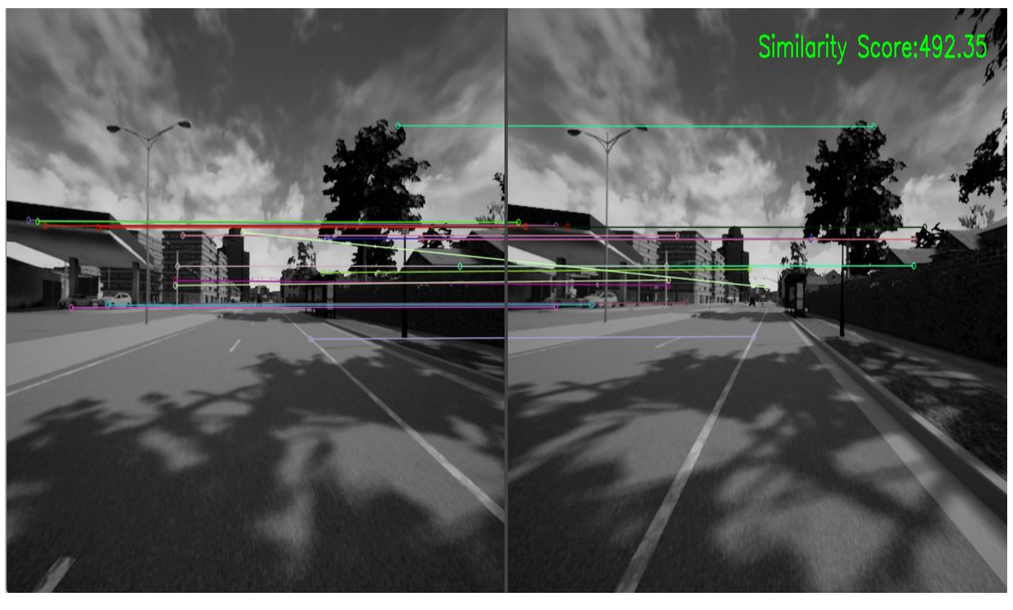
Virtual Stereo for Training

- Automotive companies launch data collection with a large number of fleets or crowdsourcing customer vehicles
- Nearby data collection vehicles can form ***virtual stereo*** pairs
- In training, virtual stereo pose additional learning loss to existing monocular depth estimation



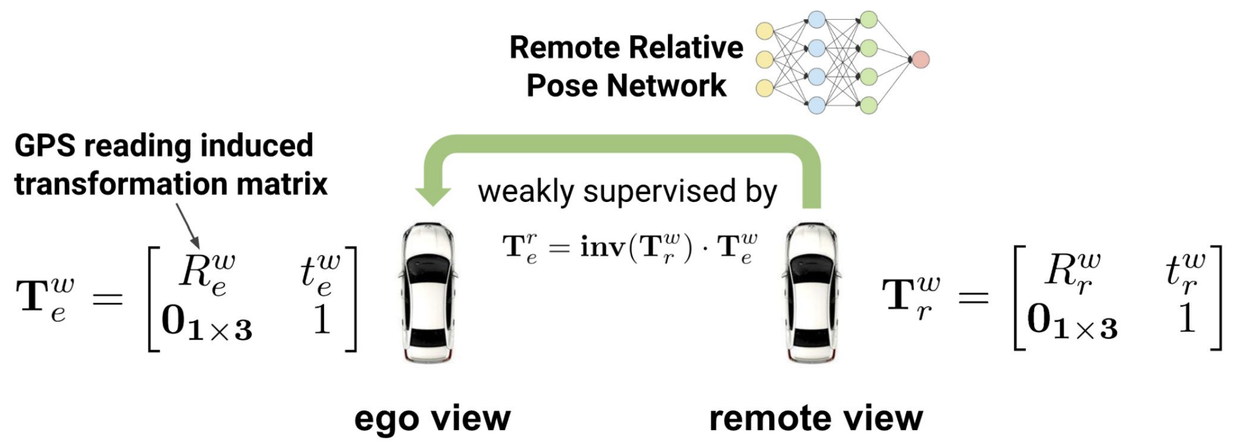
Methodology – Image Similarity based Search

- Select images from different perspectives but with similar scene into training



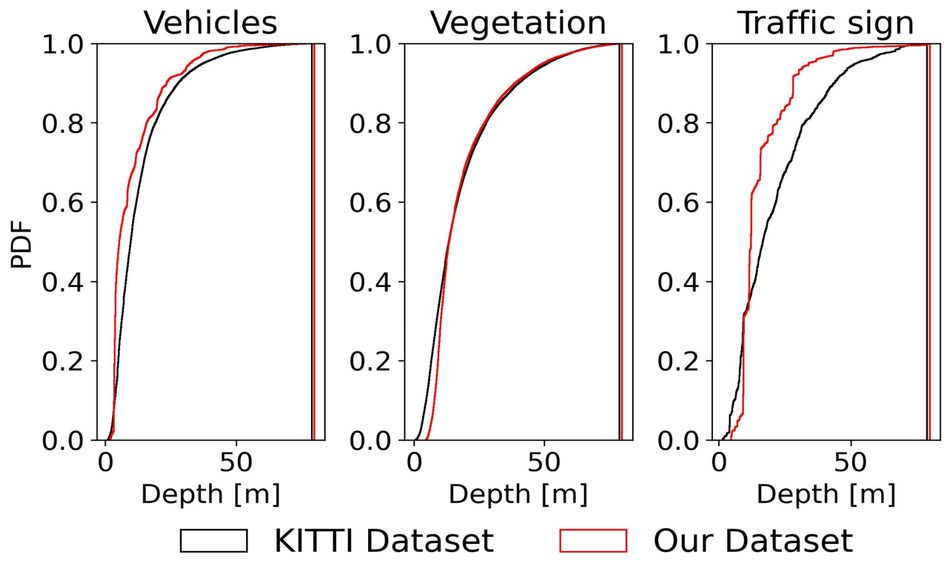
Methodology – GPS-guided Remote Relative Pose Estimation

- GPS readings are in lane-level accuracy and robustness
- Remote relative pose estimation uses GPS information as weak learning signal



Training Data Generation

- No publicly available datasets for training
- CARLA simulates realistic scenarios with similar camera configurations and pixel distribution as KITTI



Improved Depth Estimation Accuracy

	Absolute Relative Error (<=80 m)	RMSE (<= 80 m)
Virtual Stereo	0.1041	7.2735
Local Stereo	0.1059 (+1.7%)	7.4440 (+2%)
Monocular Only	0.1132 (+8%)	7.7336 (+5.9%)