Imitation Learning of Path-Planned Driving using Disparity-Depth Images

Sascha Hornauer, Karl Zipser, Stella X. Yu

International Computer Science Institute,
University of California, Berkeley, USA

Introduction

- Sensor data representation is a defining factor for the performance of autonomous driving systems
- Segmented camera images provide the most elementary driving cues but need human annotation
- We evaluate an End-to-End trained autonomous system, driving only based on disparity images
- Disparity images annotate depth sufficiently for *Free-Roaming* - collision avoiding space traversal
- In performed experiments, disparity-images are generated from stereo-RGB images and we compare driving based on each image type

Network Training

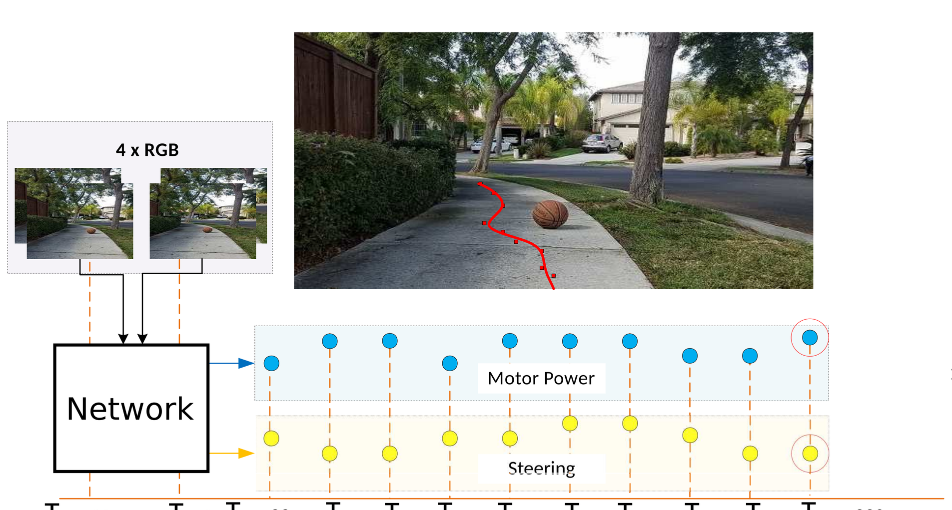


Figure: Network training, as further explained in [1]

- Images from past timesteps are collated as input to the network
- The output of n-steering and motor commands, from the present to n timesteps in the future, form a trajectory
- Only one steering-motor pair is used for actual control
- Generation of all steering-motor pairs is a side-task, improving trajectory output

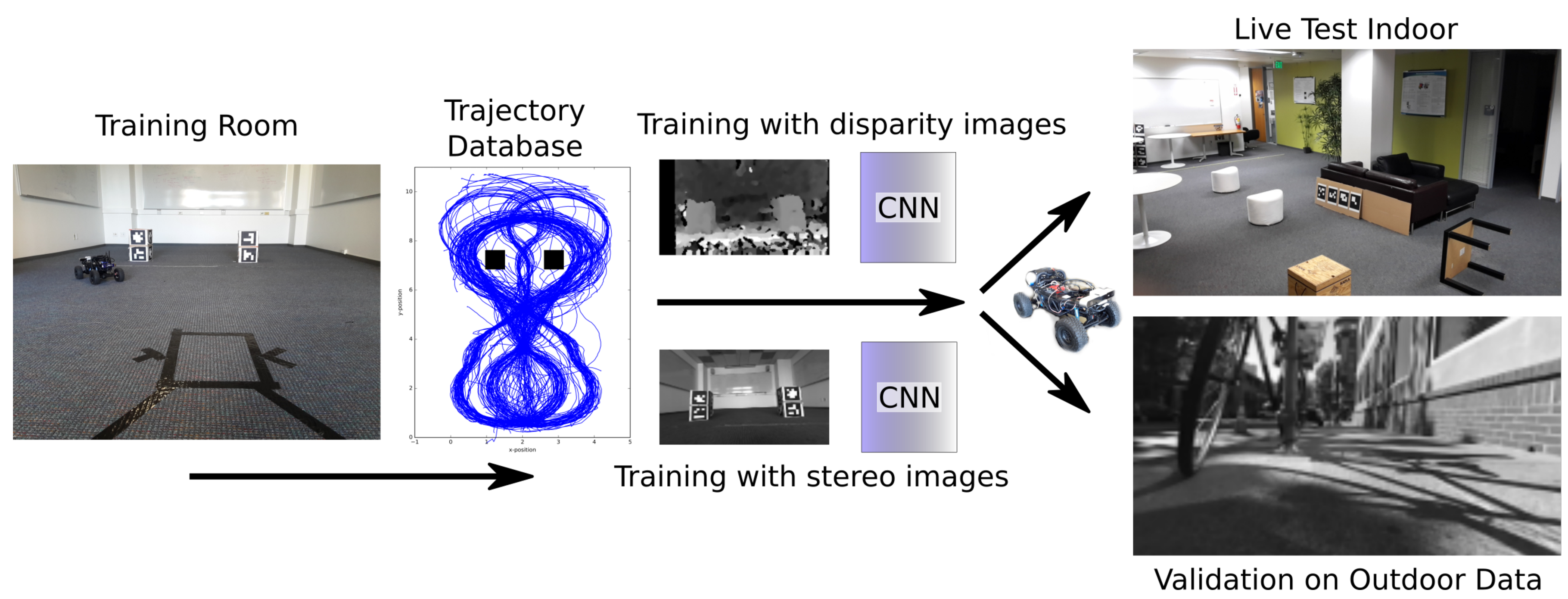
Acknowledgements

We would like to thank the Berkeley DeepDrive program sponsors for their support. Furthermore we thank Nvidia for supplying Jetson TX1 and TX2 platforms on which all cars are based.

Contact Information

- Email: saschaho@icsi.berkeley.edu

Overview



Disparity Images

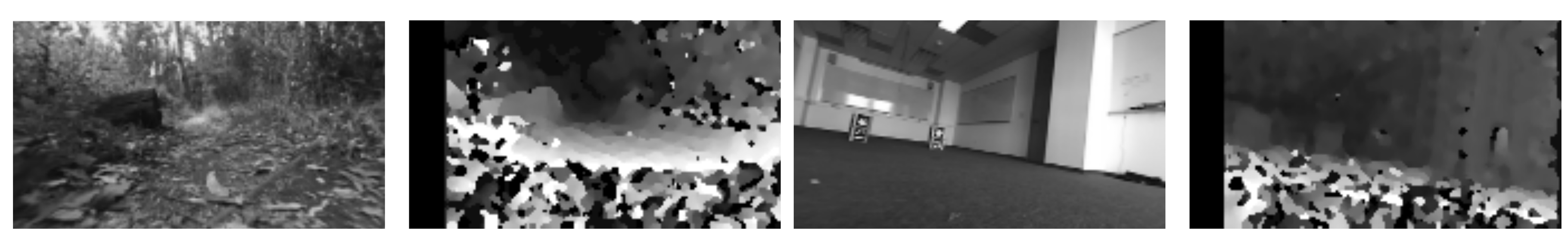


Figure: Left to right: Stereo image from driving in a park. Reconstructed disparity image from the same scene. Image from the data collection room for training. Reconstructed disparity image with noise on the ground.

Trajectories in Training and Test Environment

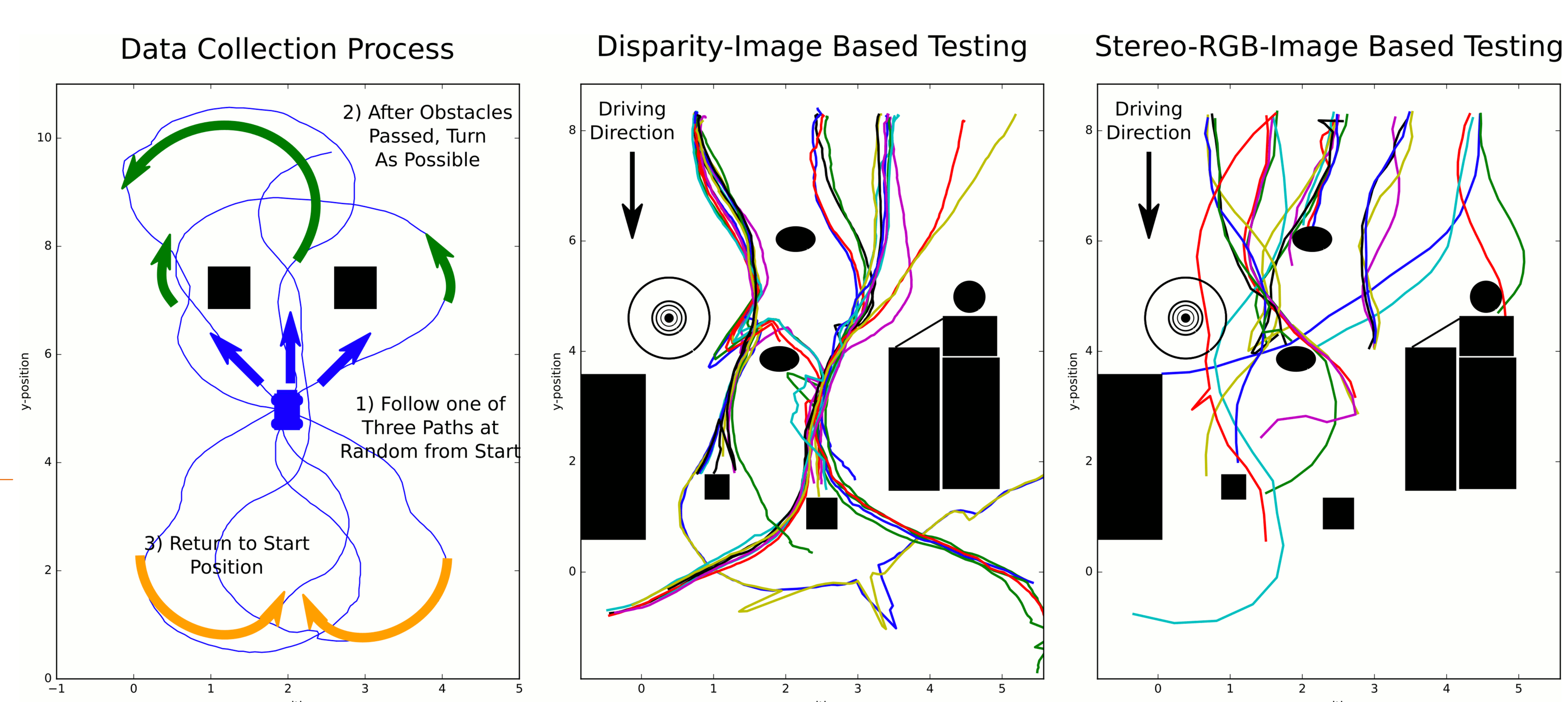


Figure: Data collection maneuvers (left) and comparison of trajectories in test-environment. Localization errors are visible as short spikes. Driving direction in the evaluation (center and right image) is from top to bottom.

Results

Images Type used / Test Environment	# Trajectories Driven	Avg. Length	σ Length	Longest Trajectory
Stereo / Cluttered Room	24	$5.32\ m$	$2.22\ m$	$11.23\ m$
Depth / Cluttered Room	28	**9.78** m	$3.09\ m$	**18.80** m
Depth / Stage Simple	20	$7.44\ m$	$4.03\ m$	$10.97\ m$
Depth / Stage Complex	20	$5.73\ m$	$4.32\ m$	$11.67\ m$
Depth / Stage Real	20	$3.63\ m$	$3.35\ m$	$11.03\ m$

Table: Results of disparity- against stereo-image based driving, compared by trajectory length in different environments. Stage refers to the *Stage* simulator with simple, complex and real being different maps.

Additional Simulation Experiments

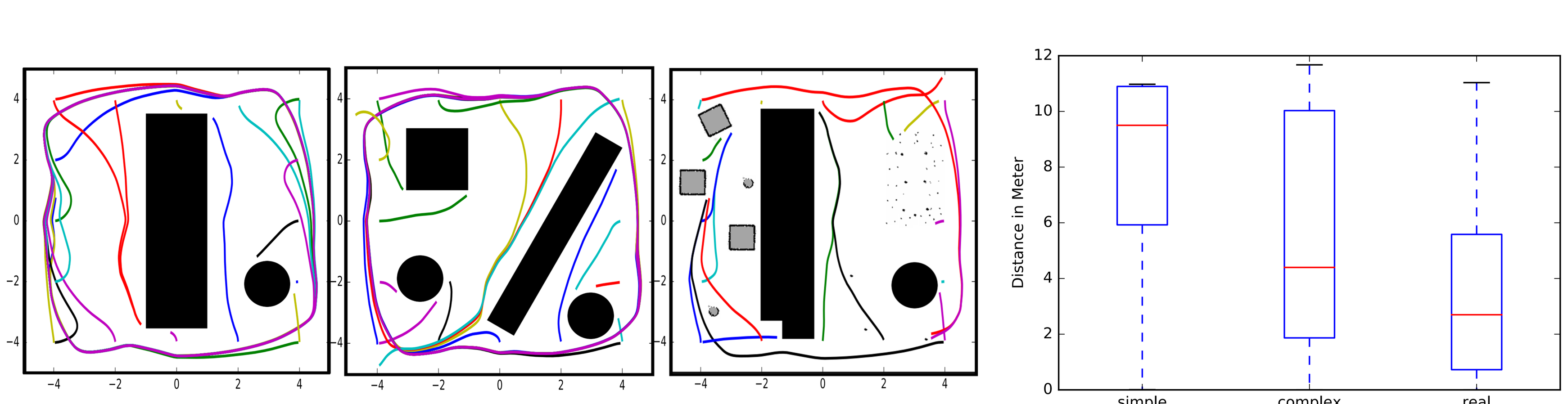
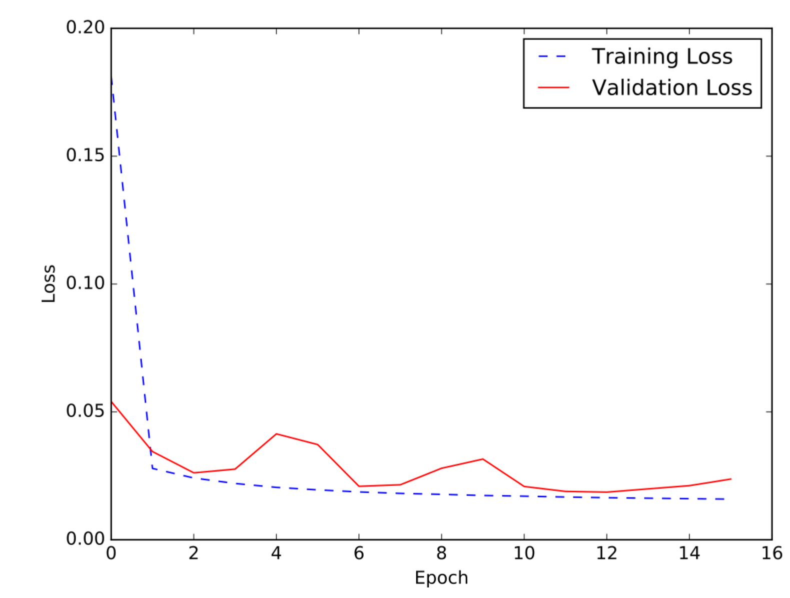
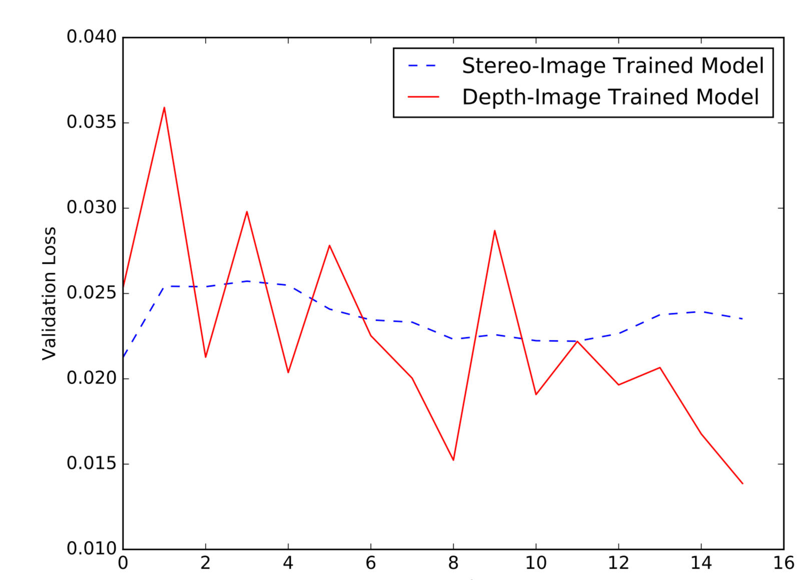


Figure: *Stage*-Simulator experiments. Start positions of trajectories are along the edges in maps called **Simple**, **Complex** and **Real** (FLTR) with driving-distance comparison. On each map the other side is reached though less often with increasing complexity.

Training and Validation Loss



(a) Training results when training with disparity images



(b) Outdoor video-validation, comparing steering prediction using each visual representation against human steering decisions.

Figure: Shown are MSE training- and validation-loss. Indoor training (fig. 5a) shows good convergence and robustness against over- and under-fitting. Evaluation of the additional test of the trained model on outdoor video-data is shown in (fig. 5b). In later epochs the depth-image based method outperforms stereo-based steering angle prediction.

Conclusion

- Generalization to new environment possible from 7 hours of driving examples.
- Obstacle avoidance based on disparity images is successful after one epoch training.
- This enables trajectory generation in simpler settings with a path planner for application in more complex scenarios

References

[1] Sauhaarda Chowdhuri, Tushar Pankaj, and Karl Zipser. Multi-Modal Multi-Task Deep Learning for Autonomous Driving. *arXiv preprint arXiv:1709.05581*, 2017.