Drive&Act: A Multi-modal Dataset for Fine-grained Driver Behavior Recognition in Autonomous Vehicles

Manuel Martin*, Alina Roitberg*, Monica Haurilet, Matthias Horne, Simon Reiß, Michael Voit, Rainer Stiefelhagen

Fine-Grained Driver Behaviour Recognition

- Activities during both, autonomous and manual driving (83 classes in total at all levels of abstraction)
- Multi-modality: color, depth, infrared- and body pose
- Multi-view: six synchronized and calibrated camera views
- Hierarchical activity labels on three levels of abstraction
- Fine-grained actions (e.g. opening bottle and closing bottle)
- Diversity of action duration/complexity (e.g. opening door from inside – seconds; reading a magazine – minutes).

Drive&Act – Key Properties

- Recorded in a static driving simulator
- 5 NIR cameras with active IR illumination and band pass filter
- 1 Kinect for Xbox One
- Tasks presented in random order on a touch screen
- 3D Body pose by triangulation of OpenPose results

Main Contribution

- First large-scale dataset for fine-grained driver behavior recognition in context of manual and autonomous driving
- Twelve hours (96.9 Mio. frames) annotated and publicly available

Data Collection: Environment

Hierarchical Annotation Scheme:

- Scenarios / Tasks: 12 high-level tasks completed each session (e.g. read a magazine to answer a question)
- Fine-grained Activities: alternate freely i.e. the driver is not told how to execute the task (3A in total)
- Atomic Action Units: Interactions with the environment: Action (5), Object (17), Location (14) triplets

Distribution of Scenarios/Tasks

- 45% Interior
- 32% Sentinel
- 23% Sentinel: Door

Data Collection: Annotation

Motivation

- Looking at humans inside the cabin crucial for human-vehicle communication, dynamic driving adaptation and safety
- Lack of large-scale public datasets for driver-activity recognition

Fine-Grained Activities on different Views

LSTM-based independent streams

- Depth first walk of kinematic model

Results

Fine-Grained Activities

- Type: Model Validation Test

Sample Frequency of Fine-grained Activities (Left) and Atomic Actions (Right) by Class

- 98% vehicle
- 94% road
- 59% driving

Pose-based Recognition

- Based on mid-level representation:
  - Body Pose
  - Static car interior
  - LSTM-based independent streams with weighted late fusion

Pose-able Recognition

- Body pose vector over time
- Pose: Body pose vector over time
- Structure: Body structure using depth first walk of kinematic model

Cross View Evaluation (Fine-Grained Act.)

- Front top 68.57 63.64
- Right view 65.16 60.80
- Left view 54.10 54.54
- GPS 60.73 67.83
- Kinetik Depth 69.30 60.95
- Kinetik Off 60.75 64.96
- Combined 72.80 68.76

All combined (vote averaging) 54.44 45.05

*equal contribution, alphabetical order