Initiative Seminar 2016

Sharing large naturalistic driving datasets while respecting privacy

Helena Gellerman, SAFER

Date: 2016-10-27
Content

- Introduction to Naturalistic Driving Data
- Challenges when sharing automation data
- Data Sharing Framework using remote desktops and Data Protection Concept
- Anonymisation of personal data
- Feature extraction from video - manual => automated
Digitalisation and Naturalistic Driving Data

- 15 years ago – memory capacity allowed continuous video collection in vehicles
  - Opportunity to understand driver behaviour
- Active Safety Systems has become standard
- Vehicles are connected
- Bike and motorcycle riding data are collected
- Autonomous vehicles need large amount of traffic environment data – tests in Göteborg 2017
- How will connected, automated vehicles change society?

Naturalistic Data is collected to understand how road users interact and the benefits of vehicle functions to society.
Naturalistic Driving Data

Driver and forward camera

Logger including vehicle data

Eye/head tracker

Rear camera

Feet camera
Data enrichment

- Reference photos of drivers
- Driver identification
- Questionnaires
- External web Questionnaire handler
- Weather data
- Map data
- Events

SAFER

27/10/2016
Sharing large naturalistic driving datasets while respecting privacy
Push towards larger use of naturalistic driving data

- Large datasets collected globally
- Usefulness of driving data
  - Research on many topics; safety, environment, mobility,
  - ADAS and automation-normal driving, driving styles
- General global trend towards open data
- Remote access techniques enables access to global datasets

- Personal data, i.e. video and GPS, and IPR data need protection
Automation related data

- Mixed traffic - other road user’s reactions
- Larger amount of sensor data
- Data to know baseline in societal scenarios that is affected by automation
  - Such as today’s vehicle sharing activities
  - Mobility patterns from transport planners
- Questionnaires to potential automated vehicle users
Challenges in sharing automation data

- Data from competitive systems (sensors, algorithms), earlier - mature systems
- Sharing for common good/policy for automation, common goals incentive for sharing
- Liability; driver => OEMs
- Data protection – data anonymization/annotations (i.e. video)
- Common data format – OEM collect with their tools and change format before sharing the data
Data Sharing Framework

Current frameworks meet automation challenges
UDRIVE Remote Desktop

- UDRIVE use remote desktop for connecting 11 pan-European Analysis Sites to one Central Data Center, hosted by SAFER.
- 6 data collection sites (140 cars, 40 trucks, 40 PTW:s)
- In UDRIVE, a Data Protection Concept (DPC) has been developed, stating the necessary requirements for analysing and managing sensitive data.
- Sensitive = video and GPS traces.
Data Protection Concept

Why data protection?

- Personal Integrity Data (PID) regulated by law
- Intellectual Property Right (IPR) regulated by agreements
- Data protection is the key to create the trust between a data provider and the research organisation
- Data protection must cover the full chain of events, during data collection, processing, hosting, and analysing.
Anonymisation – main issues
video and GPS

Anonymise while keeping the features essential to analysis

- **Internal video**
  - Identify the emotions of the driver
  - Follow the head and eye movements
  - Identify body movements and tasks

- **External video**
  - Identify scenario
  - Identify detailed information in interaction with other traffic participants

- **GPS**
  - Keep the start and end of trips as long as possible
Example of anonymization - avatar
Chalmers, Volvo, SmartEye, Räven
Anonymisation in the Past, Today and in the future

The past
- Images, contact info, GPS

Today
- Video, GPS for research

In the near future
- Low-cost, real-time feature extraction or anonymisation
- Position privacy in connected automation
Automated feature extraction

Challenges

- Automation challenge - Uncontrolled environment
  - Low resolution
  - People moving out of scope
  - Time of day, weather
  - Differences in camera mounting, glasses, driving position and drivers’ length

- Manual annotations - time-consuming
- Automated annotations
  - Machine learning/
  - deep learning algorithms
Conclusions

- Naturalistic driving data valuable for developing and evaluating automation
- Current data sharing frameworks applicable
- Specific challenges with sharing automation data remain to be solved
- Reliable automated feature extraction key to easier access to data
Thank you for your attention!

Contact info:
Helena Gellerman
SAFER
helena.gellerman@chalmers.se
+46 31 7721095