SAFER100Car: a toolkit to analyze data from the 100 Car Naturalistic Driving Study

Marco Dozza, PhD – Chalmers University of Technology – Göteborg, Sweden.

The 100-Car Naturalistic Driving Study (DOT HS 810 593) was the first large-scale naturalistic driving study in the world. In this study, 100 cars instrumented with various sensors collected observational data in real time in real traffic. Two million miles of data were collected during 43000 hours of driving. These data included: vehicle dynamics (such as acceleration and yaw rate), driver control inputs (such as gas and brake pedal position), radar data (such as range and range rate), and lane marking information. Further, video data was also collected and annotated for specific events such as crashes and near-crashes. Video annotation included environmental and driving situation as well as driver's glances before, during, and after the event. Analysis of these event annotations proved, for instance, that drowsiness increases relative risk for a crash or near-crash event by 4-6 times, complex secondary tasks increases it by 3 times and moderate secondary tasks do so by 2 times (Klauer et. al 2006; The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data).

Part of the 100-Car Naturalistic Driving Study data (768 near-crashes and 60 crashes events) has been made available to the public (http://www.access.vtti.vt.edu/). Time series data are provided in text files format (one file for each event). Tables presenting event narratives and sensor operation status are provided in pdf format. Glance data and video annotations are also available in two additional text files.

By parsing these different files, it was possible to obtain a file system where all data from multiple sources was merged and synchronized for each individual event. Thus by opening any of these files, all data from all data sources related to a specific event (time series, narratives, glances, etc) can be assessed. Further, a matrix organizing all data by event and containing all information but the time series was developed in Matlab to allow fast and easy querying of the data set. For instance, a few lines of code are enough to generate a query for "the average speed at event start in rear-end crash events in which, at the event start time, the driver was not looking forward". So far, such queries have been used in the ASSESS project (http://www.assess-project.eu/) to couple accident databases and field operational test data from the 100-Car data to specific scenarios in order to combine and cross-validate the data.

In order to further facilitate analysis, a graphical user interface (GUI) was also developed. This GUI was developed in Matlab and is intended to assist the user in browsing and visualizing the 100-Car data. As shown in Figure 1, this GUI enables a user to select events from the 100-Car dataset and plot time series and glances synchronized in time. Further, this GUI provides the user with narratives, all video annotations, and sensor status information related to the current loaded event.

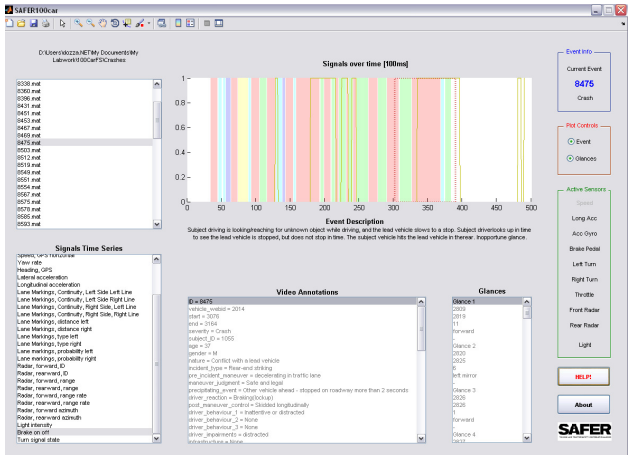

Figure 1 – Screenshot of the SAFER100Car GUI

In conclusion, the SAFER100Car toolkit consists of a file system, a Matlab event matrix, and a GUI to support further analysis of the public 100 car data set. This toolkit is freely available at (http://forums.vtti.vt.edu/index.php?/files/category/4-file-exchange/). Future versions of this GUI include new features such as calculation of safety-relevant parameters (e.g. glance frequency and time to collision).

Author Contacts:

Marco Dozza, PhD – CHALMERS - University of Technology, Dept. of Applied Mechanics

Tel: +46 31 772 3621, e-mail: marco.dozza@chalmers.se

SAFER - Box 8077 - S-402 78 - Göteborg - Sweden