# CHALMERS



# Recognizing safety-critical events from naturalistic driving data

*Master's Thesis in the Master's programme of Automotive Engineering*

## NIEVES PAÑEDA GONZÁLEZ

MASTER'S THESIS 2011:38

# Recognizing safety-critical events from naturalistic driving data

Master's Thesis in the Master's programme of Automotive Engineering

NIEVES PAÑEDA GONZÁLEZ

Department of Applied Mechanics
*Division of Vehicle Safety*
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2011

Recognizing safety-critical events from naturalistic driving data
Master's Thesis in the Master's programme of Automotive Engineering
NIEVES PAÑEDA GONZÁLEZ

Recognizing safety-critical events from naturalistic driving data
Master's Thesis in the Master's programme of Automotive Engineering
NIEVES PAÑEDA GONZÁLEZ
Department of Applied Mechanics
Division of Vehicle Safety
Chalmers University of Technology

ABSTRACT


New trends in research on traffic accidents involve conducting Naturalistic Driving Studies (NDS). NDS are based on large-scale data collection of driver, vehicle and environment information in real-traffic. NDS provide large data sets which have proven to be extremely valuable for the analysis of safety-critical events such as near crashes and incidents.

NDS data needs to be filtered to recognize safety-critical events. Filtering safety-critical events has been traditionally achieved by using kinematics triggers (e.g. searching for deceleration below a certain threshold signifying harsh braking). The low sensitivity and specificity of this filtering procedure, however, requires manual annotation of video data to decide whether the events individuated by the triggers are actually safety-critical. Such reviewing procedure is based on subjective decisions, time-consuming, and often tedious for the analysts.

This project looked into improving this reviewing procedure using video data collected from 100 Volvo cars during one year in Gothenburg within a NDS called euroFOT. More than 400 videos from the triggered events have been reviewed, concluding that driver's reaction may be the key to discriminate safety-critical events. In fact, whether an event if safety-critical or not depends on the driver. Several statistical procedures have been then applied to automatically recognize driver reaction from video data. In this project, we showed how combining automated video analysis with kinematics triggers increases sensitivity of near crash recognition from NDS data. These results open up to new ways to use video frames in NDS.


Key words: naturalistic driving, driver behavior, traffic safety, near crashes, safety-critical events, driver's reaction, euroFOT

# Contents

# Preface

This project ends the academic formation that I held in Spain over the last years. "Recognizing safety-critical events from naturalistic driving data" has given me the opportunity to learn about traffic safety in the multicultural environment of an open area at SAFER. Personally and professionally, I will never forget this experience in Sweden. There are many *thanks* that I would like to share:

Thanks to those who make the naturalistic driving studies possible. Specially, thanks to Volvo Cars for allowing me access to their database in this research. Thanks to all the participants who have been recorded while driving for their collaboration in gaining knowledge about driver behaviour. Without them this project wouldn't be possible.

Thanks to SAFER, where I was working during the last months. It was a pleasure to be part of this *family* and the incredible work of this group to save lives.

It's said that a good teacher teaches, and the best teacher inspires. To my supervisor Marco Dozza, thanks for *inspiring* me during this project. Thanks for this opportunity, for trusting me from the beginning and for your guidance during these months. I feel very lucky to have worked not only with a great professional, but a great person.

Thanks to the University of Oviedo for letting me participate in this international exchange. Specially, thanks to my supervisor in Spain Ramón Rubio.

To my friends and to everyone I've shared experience, thanks for making unforgettable this year in Göteborg.

To my family, *muchas gracias* for your unconditional support in my life plan.


Göteborg June 2011

Nieves Pañeda González

# 1 Introduction

*This chapter presents the reader with an overview about Naturalistic Driving Studies (NDS) and their implementation together with Field Operational Tests (FOTs). In particular, the euroFOT project is introduced as a base of this project. This chapter also covers the limitations found in previous studies and formulates the research question and objectives for the present project.*

## 1.1 Naturalistic Field Operational Tests: real-traffic data

Statistics said than more than 1.2 million people die on the roads in traffic accidents every year (WHO, 2009). Technological advances allow the development of new systems in cars to mitigate road accidents by automatically detecting risk situations. To make it possible, it is essential to know which the real causes of accidents are.

New trends in research on traffic accidents involve conducting Naturalistic Driving Studies. **Naturalistic Driving Study (NDS)** as concept refers to a "method of observation that captures driver behaviour in a way that does not interfere with the various influences that govern those behaviours" (Boyle *et* al., 2009). Statistics and crash investigations rarely provide information about behavioural issues before the incident. In simulations, test subjects are well aware of the experimental conditions. Thus, NDS aim collecting data on driver behaviour in a natural setting. In this naturalistic observations drivers use, preferably, their own car equipped with cameras during their daily driving. Experience in this field shows that drivers quickly forget the presence of cameras.

On the other hand, new technologies enable the collection of an extended amount of data, such as vehicle dynamics or the environment, in real traffic within large-scale testing programmes called **Field Operational Tests (FOTs)**. FOTs are studies undertaken to evaluate the efficiency of intelligent in-vehicle systems as well as the impact on safety and the driver acceptance, among others (ERTICO, 2009). The main purpose of these systems is to assist and inform drivers while driving. This concept applied to the field of safety embraces alerting the driver or automatically acting in the car in presence of what the system understands as a risky situation.

To sum up, FOTs are a complementary step to the development of intelligent in-vehicle systems. The procedure is mainly based on:

-Instrumenting cars with loggers to collect information from the CAN bus (signals from accelerometers, gyroscopes, turn indicators, etc.), GPS and/or extra sensors.

-Driving such equipped cars to collect data.

-Performing analysis from collected data.

Although FOTs and NDS pursue different objectives, this view is changing. Combination of both, called **Naturalistic Field Operational Test (N-FOT)**, allow the use of this *unobtrusive* observation of drivers to evaluate their relationship with the car and the environment under crash-risk and the effectiveness of intelligent in-vehicle systems.

### 1.1.1    State of the art of N-FOTs: EuroFOT

During the last years, FOTs and N-FOTs have been conducted in the United States, Asia and, relatively new, in Europe. Particularly, US has extensive experience in NDS with programs as 100-Car study, 250-Truck study, the Commercial Vehicle Operation study or the Strategic Highway Research Programme (SHRP2).

The 100-Car Naturalistic Driving Study (Dingus *et.* Al, 2006) was the first large-scale program where data from 100 drivers were collected during one year.  The main goal of this research project was the study of contributing and associative factors (such as driver behavior, kinematic characteristics and corrective actions) in critical situations. In the ongoing SHRP2 project (TRB, 2011), data from 3000 volunteer drivers in instrumented cars will be collected. Main goals are to redesign highways (congestion reduction, planning, environmental conditions) and to study human behavior for a safer highway.

Among the European experience in this field can be highlighted the contributions of SAFER, the Vehicle and Traffic Safety Centre at Chalmers University, in Sweden. Programs as SeMiFOT (Victor *et.* al., 2010) in collaboration with Michigan, carried out the development of a N-FOT methodology. Data were collected from 14 vehicles during six months, with the participation of 39 drivers that made 12.571 trips.The methodology is widely used in accident research and evaluation of safety and acceptance.

The ongoing second version SeMiFOT2 is using the data collected in the first version of the program. New statistical methods, such as extreme value theory, are being explored to identify and model outliers. This provides useful information for insurance companies, for instance, to establish a link between rare events and catastrophic consequences (García, 2004). In addition, the analyses of visual motion in drivers are one of the main lines of research.

Other ongoing European projects are TeleFOT, 2BeSafe NDS, INTERACTION, TSSFOT, simTD and euroFOT (ERTICO, 2010). Particularly, this research has accessed the data collected in euroFOT. Characteristics of this program are further explained below.

Co-founded by the European Commission, **euroFOT** began in May 2008 and will last until February 2012 supported by 28 partners (vehicle manufacturers, automotive suppliers, and research institutes among others). As stated in the previous section, intelligent in-vehicle systems are tested to explore potential ways to improve European road traffic.

 The tested applications in euroFOT may be classified as (ERTICO, 2010):

> •Assisting the driver in forward/rear directional safety:
> - Adaptive cruise control
> - Forward collision warning
> - Speed Control System
>
> • Assisting the driver to detect hazards at the sides of the car:
> - Blind Spot Information System
> - Lane departure warning / Lane Assist / Impairment Warning
>
> • Advanced applications:
> - Curve Speed Warning
> - Fuel Efficiency Adviser
> - Safe Human/Machine Interface

These functions have been tested in a fleet of 1000 instrumented cars from nine different brands across France, Germany, Italy and Sweden. This has led one of the largest and most completed FOT's databases in Europe for public research.

As can be seen in Figure 1.1. FOTs are operated on fleets managed by different OEMs around Europe.



*Figure 1.1* *Geographical coverage of euroFOT: OEMs and operation sites. (Mure S., 2010, EuroFOT [electronic print] Available at: <http://wiki.fot-net.eu/index.php?title=File:Eurofot.jpg> [Accessed May 2011]).*

Depending on the project and the OEM, various devices are part of the test equipment to collect data. These may be classified according to the source of the recorded signals:

-CAN bus.

-CAN bus and video cameras.

-CAN bus, video cameras and extra sensors (as eye tracker).

In addition to the test and evaluation of intelligent in-vehicle systems, some research focuses on naturalistic observation, hence the implementation of cameras in the cars. In any case, the resources for data collection and storage are common in both types of projects. Another type of drivers' data comes from interviews and questionnaires.

Both the kinematics of the car from loggers and camera images have proved very useful when studying the interaction between driver, vehicle and the environment during a crash risk situation. The knowledge on driver behaviour and dynamics of the car before an accident allow for hypothesising possible causes. This is a step towards the inclusion of new measures in accident prevention.

### 1.1.2 Available data from VCC (euroFOT)

In particular, this research has accessed the data collected from 100 **Volvo Cars driving for a year in Gothenburg** within euroFOT program. After a certain period of continuously data collection, information from loggers was downloaded and transferred to a network. Then, these signals have been post-processed and stored into MatLab variables.

The available data are mostly signals from the CAN bus sampled at 10 Hz, GPS information, video images and signals from the eye tracker. These provide information on, for example, kinematic values (such as speed, lateral and longitudinal acceleration, brake pressure, yaw rate, steering wheel jerk, among others) or signals from intelligent in-vehicle systems and turn indicators.

A total of four cameras are installed in each of the instrumented cars. Two are located in the front and back of the cars to mainly reconstruct rear-end crashes and evaluate the traffic flow. One is located under the steering wheel, to record the pedals and the feet movements. Finally, another camera is located in the rear-mirror, focusing the driver. The eye tracking is also available.

## 1.2 Data reduction approach: triggering data

To understand the causes of road accidents and be able to further develop countermeasures is essential to analyze safety critical situations. The identification of safety critical situations among hours of normal driving is a limitation when loggers and cameras are continuously recording. Therefore, once data are collected, a filtering process is carried out before performing analysis (see in Figure 1.2). This process is commonly called *triggering* the data. The main goal of this data reduction approach is the discrimination between normal driving situations (*negative* situations) and the critical *events* (*positive* situations) while driving.
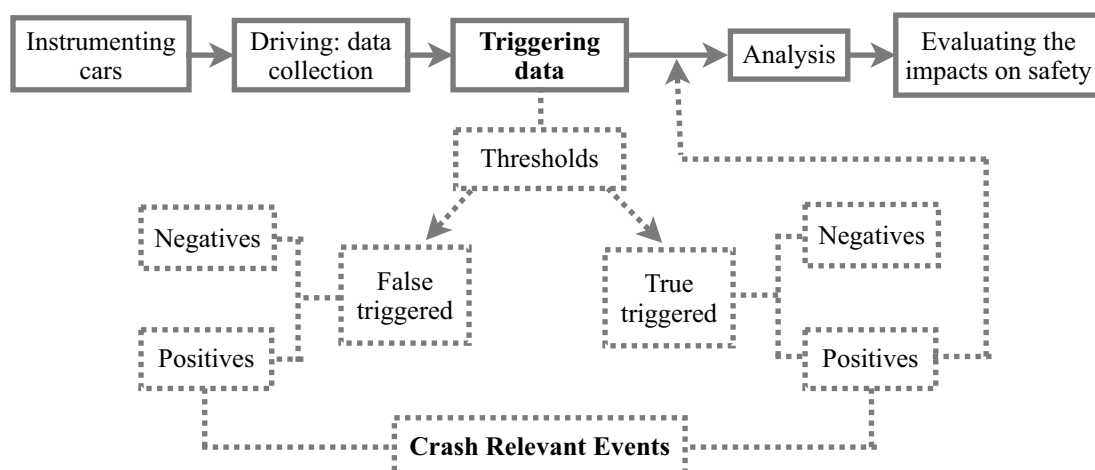


*Figure 1.2      General steps before the evaluation of safety in FOTs.*

A more precise definition of what those critical situations are, is given in the first large-scale FOT conducted in US, the 100-Car study. The distinction is done as follows (Dingus *et* al., 2006):

> **-Crash:** situations in which there is *physical contact between the subject vehicle and another vehicle, fixed object, pedestrian, cyclist or animal.*
> **-Near-Crash:** situations requiring a *rapid, severe, evasive maneuver to avoid a crash.*
> **-Incident:** situations requiring *an evasive maneuver occurring at less magnitude than a near crash.*

These safety critical situations are grouped under the name *Crash Relevant Events* (CREs). Once they are located in the database, the next steps are to conduct a detailed description (of the driver behavior, the environment, traffic conditions, etc), draw conclusions and evaluate possible solutions.

Conventionally, CREs from naturalistic driving data have been isolated from the large database using kinematic triggers. These are pieces of code that run throughout the database and record situations with certain kinematic values. Most of these triggers are associated with common evasive maneuvers and acceleration peaks. For example, one of the most typical responses in drivers is to slam on the brakes to avoid a rear-end collision, which leads to peaks in longitudinal acceleration. Therefore, situations in which deceleration is below a certain threshold[1] may indicate that there is a CRE. In that case, the recorded situations have been *true triggered* and constitute a list of candidates to CRE.

However, as evidenced by triggering with kinematic values, some CREs are missing (positives that haven't been triggered*, usually called *false positives*) and many normal driving situations are wrongly triggered (*false negatives*). This is mainly due to some cutoff kinematic values related to evasive maneuvers may be identical to those obtained while normal driving because of the diversity of drivers and ways of driving. For instance, the same acceleration value may or may not be indicative of risk depending on the aggressiveness of the driver and his/her driving experience. Taking as reference signals such as braking, incidents in which the driver is distracted would be lost. Hence the importance of a precise definition of what is a CRE and the development of intelligent triggers.

Among all the possible types of CRE, crashes may be more likely to be detected. This is due to the involvement of contact is likely to cause sudden changes in kinematic parameters. However, **near-crashes and incidents** are closer to normal actions while driving. Thus, trying to locate these situations, which are also relevant from a safety and statistical point of view, creates a high rate of *false negative events.*

The low sensitivity and specificity of triggering with kinematic values require the intervention of reviewers, who decide whether the situation is critical by watching the video segments from the candidates to CRE. Therefore, only the true triggered events that have been considered positive by the annotators pass into the analysis phase. Such reviewing procedure it's mostly based on subjective decisions, time-consuming and often tedious for the annotators.

---

[1] values taken as references for each trigger to save results (if keeping decelerations below -4 $m/s^2$, then acceleration is the trigger and -4 is the threshold).

## 1.3    What is safety-critical? Driver behaviour in NDS

The 100-Car study defines CRE as:

*"A subjective judgment of any circumstance that requires, but is not limited to, a crash avoidance response on the part of the subject-vehicle driver, any other vehicle, pedestrian, cyclist, or animal that is less severe than a rapid evasive maneuver (as defined in near-crash event), but greater in severity than a normal maneuver to avoid a crash(...)"* (Klauer *et* al., 2006)

When annotators review the list of candidates to CRE from the triggering process, their *subjective judgment* it's based primarily on their perception of how critical the situation seems. This concept is under the above definition, since annotators should evaluate whether *the circumstance requires a crash avoidance response on the driver* or other involved.

Given the limitation of answer this question by just checking the kinematic values of the car or its proximity to other vehicles (objective judgment), each annotator mostly bases his/her opinion on the own driving experience. This hypothesis casts a question: what I think it's critical, is it also critical for you?. It may be that the fairest answer to this issue requires some *empathy* with the subject-vehicle driver. This changes the question into: Does the driver think that the situation is safety-critical?.

The answers to this question in previous studies were based, for instance, on the force with which the driver depresses the brake pedal[2] or on changes in the speech under threatening conditions (Malta *et* al., 2009). This is also related with the fact that around the 60% of drivers brakes before a crash (Molinero *et* al., 2009). The main limitation arises in those critical situations closer to normal driving in kinematic terms, such as near-crashes and incidents. These provide a large source of information and a definite benefit in safety and statistical analysis concerning NDS (Guo *et* al., 2010).

There are many literature about how driving is affected by factors such as country, gender, age, or lifestyle among others (Evans, 2004). These factors imply a diversity of driving modes, hence the importance of using the driver as part of the analysis. This conclusion was also pointed out in 100-Car study (Klauer *et* al., 2006).

The analysis of driver behaviour in NDS has been used, for instance, in the development of a model based on multi-modal signals (Takeda, 2010), or in the study of situations when drivers approach to intersections. In this case, it has found a relationship between distance to other vehicles and the location of covering the brake pedal (Sato and Akamatsu, 2007). The movements of the head and eyes are also objects of study in the distractions at the wheel (Nagase *et* al. *2009).*

Regarding to the driver behaviour prior to a CRE, Molinero *et* al. (2009) define *key events* in situations with failure or not presence of manoeuvres. These include *excessive speed* and *inappropriate reaction*, which they relate to *driver panic*. This concept is present in so-called *oops reactions* in SeMiFOT, used in the study of driver inattention associated with poor driving performance (Victor *et.* al*, 2010). They also highlight the importance of optimizing the CRE triggers.

---

[2] brake pressure signal in combination with speed is a potential trigger detected while triggering an initial euroFOT dataset (see Appendix 1).

The main limitations in the identification of CRE in a large data set are the variety of drivers and the wide range of situations. This procedure based on what the driver is expected to do, such as evasive maneuvers, leads to loss CRE and results in a high rate of negative situations. Although the perception of what is risky and what can be done depends on the person, there may be a common feeling when someone realizes that something is wrong. This feeling may materialize in a particular body language, before whatever evasive action, if any.

## 1.4    Purpose

Conventional triggering does not seem very efficient to find critical situations among hours of normal driving in a large database. Although kinematic filters can run automatically into the database, the high rate of false events requires the manual intervention of reviewers. Such reviewing procedure is mostly based on the drivers' reactions in images from cameras inside the cars. In addition, this procedure is time-consuming and often tedious for analysts. Furthermore, comparison of results between different NDSs may also be inaccurate given that the validations are subjective decisions of reviewers opening for inter-subject and intra-subject reliability concerns.

A traditional triggering procedure applied to the initial euroFOT data set suggested the hypothesis that there is a relationship between driver motion and CRE. This idea came after watching more than 400 videos containing 40 positive situations[3].

The main objective of this thesis is to test such hypothesis by creating an *algorithm able to automatically identify CREs among the events triggered with kinematics values in euroFOT database. Such algorithm is based on the recognition of driver's reaction from video images.*

By defining a training sample from the initial triggered procedure, several methods were applied to recognise the driver's reaction using images from cameras inside the car. Once possible algorithms had been defined and tested in the training sample, the next step was to evaluate them in a larger data set. Conclusions of these procedures and suggestions for future research are also addressed in the last chapters of this thesis.

The scope of this thesis has excluded the use of images other than 1) the driver's body and 2) the search for kinematic values related to the driver's reactions. Further, this thesis moves a first step toward the integration of video information for triggering CRE focusing on the driver reaction and not on the current possibilities of image-processing algorithms.

---

[3] further information in Appendix 1

# 2 Methods

*The following chapter proposes the algorithms employed in this thesis to recognize drivers' reaction from cameras inside the cars. The different algorithms were tested on a training sample containing two normal driving situations and a CRE for eleven different drivers. The intermediate goal was to find a method that allowed for an automatic discrimination between true and false CRE.*
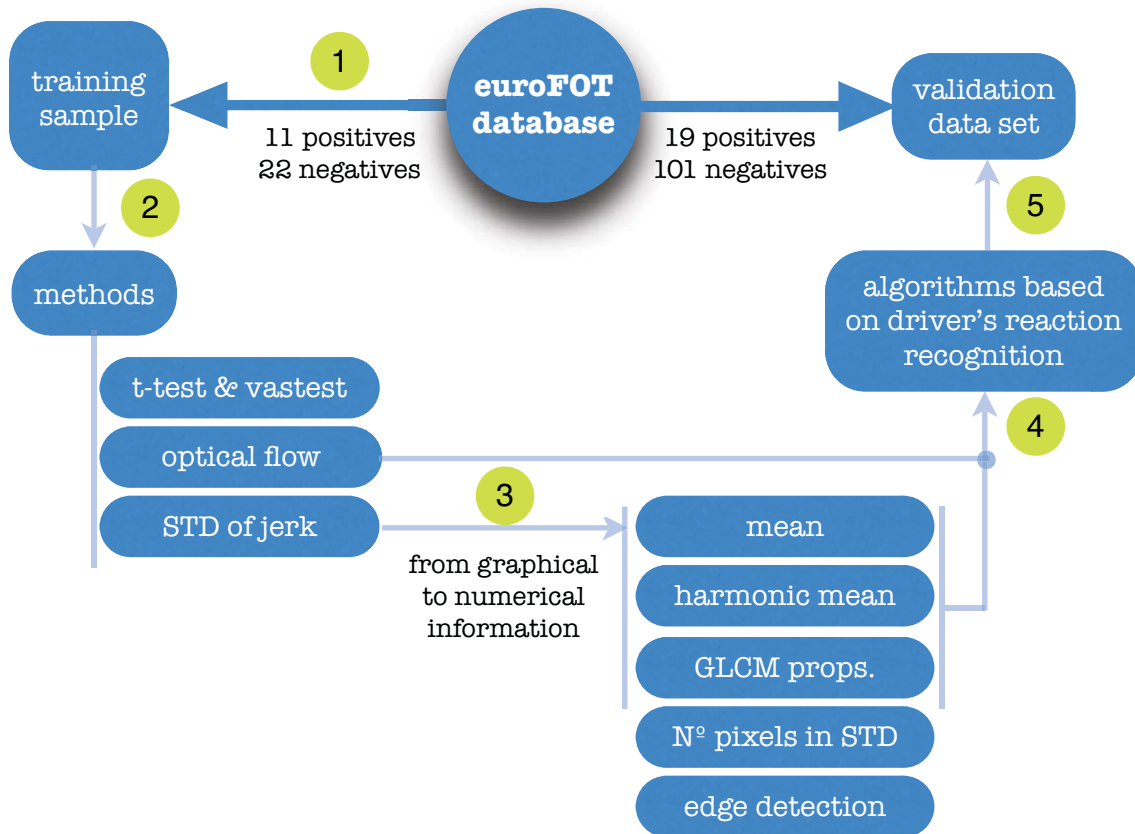


*Figure 2.1    Methodology.*

The Figure 2.1 contains a schema of the followed methodology, whose steps are addressed in more detail throughout the following sections. To have an overall idea, these can be summarized as follows:

**1**  33 sequences, containing positives and negatives situations, were extracted from the euroFOT database to define a **training sample**.

**2**  Then, three methods were applied in the training sample to discriminate between positives and negatives: the **t-test&vartest**, the **Optical Flow** calculation and the **STD of jerk**. The last two were identified as potential algorithms and entered the next phase.

**3**  STD of jerk required an intermediate step to convert its graphical information to numerical. Among several methods, the **mean**, **harmonic mean** and **GLCM properties** were used as three different convertors that allow an automatic detection.

**4** The three convertors of STD of jerk together with the Optical Flow criterion defined four potential **algorithms based on driver's reaction recognition in the identification of CREs**.

**5** Finally, the four algorithms were first **tested** on the **training sample** and then on the **validation data set**. This was formed by 120 situations (101 negatives and 19 positives) extracted from euroFOT database. The results from this phase are explained in the next chapter.

## 2.1 Driver's reaction recognition. General assumptions

As suggested by the viewing of videos of candidates to CRE triggered in an initial euroFOT data set, the key to discriminate between normal driving situations and CRE may be the driver's reaction. In fact, it's the driver who decides whether the situation is critical (*positive event*) or not (*negative event*).

For instance, harsh braking is one of the most typical responses when drivers presence a critical situation. A high decelerations is used as trigger to detect such CREs. However, there are more aggressive driving styles, so the same deceleration level may be achieved in drivers that are totally aware of the situation. Due to the diversity of drives and personalities, reviewers examine which is the driver attitude in the videos to guess whether the situation is critical for him/her.

In euroFOT, these images are taken from cameras located in the rear-view mirror inside the cars. These are oriented toward the driver, making it possible to observe his/her torso[4]. In the sequences of CREs is observed a rigid body motion common to all drivers. This reaction is characterized by sudden movements, such as suddenly grab the steering wheel with both hands and tilt the body forward. This theory also fits with the findings in a study of emotions and associated motions, which relates the surprise with an acceleration of the whole-body portions (Kobayashi, 2008).

Prior to the beginning the search for possible methods, assumptions and requirements should be defined. Based on the findings of the initial triggering procedure, **assumptions** are:

1) **Driver reaction is an indicator of CREs.**

2) **Motion in the driver's body from euroFOT cameras can be used to detect driver reaction.** Given the presence of kinematic changes while driving, driver reaction implies movement (it may not be just a change in face expression).

3) On the basis of the second assumption, individual movements may be not enough self-explanatory. Thus, a **sequence of movements** seems the best indicator of driver's reactions.

The main requirement is that the greatest number of positive events should be detected with the least possible number of negative events. This means to **increase the sensitivity of CRE recognition**. The main challenge in this point is to identify near

---

[4] front-seat passengers are not included in the camera's field-of-vision.

crashes and incidents, since they have kinematic values closer to normal driving situations.

Other issues to consider are the variety of drivers and the computational time. It's important to create an algorithm able to detect different drivers' reactions in the shortest possible time. This can be generalized considering the images as matrices containing numbers (pixel intensities). In addition, a **statistical approach** can contribute to measure changes in these matrices and to save computational time.

Due to privacy and ethical issues, throughout this project the faces of drivers are hidden to remain anonymous. The tools used in the analysis were placed in locked rooms following the requirements on personal data handling. Only authorized analysts were able to see the displayed data in such rooms. An information document was signed before getting access to ensure that individual drivers could not be identified by anyone except authorized persons. Finally, the extracted data have been revised to include them in this report.

## 2.2    Definition of training sample

The training sample is a part of the entire data set used for testing methods. This involves testing and searching alternatives to recognize driver's reaction within a limited collection of data. Results from this procedure allow the development of potential algorithms able to identify CREs among negative and positive situations. This will be further evaluated in a larger data set.

For the results to be enough consistent, the training sample should be representative of the population. In this case, it contains two-second sequences of eleven different drivers randomly selected among positive events. This positive events were obtained using the kinematic triggers defined in 100-Car (Dingus *et* al, 2006) in an initial euroFOT data set. By watching those videos is possible to identify the whole driver's reaction within two-seconds (starting half-a-second before the triggered time). Events have been further described in Appendix 2.

The training sample also contains two additional negative events for each driver (see Figure 2.2). These have been recorded in the sequences that take place four and two seconds before the positive event. Such sequences are related to normal driving, thus they are defined as negative events.
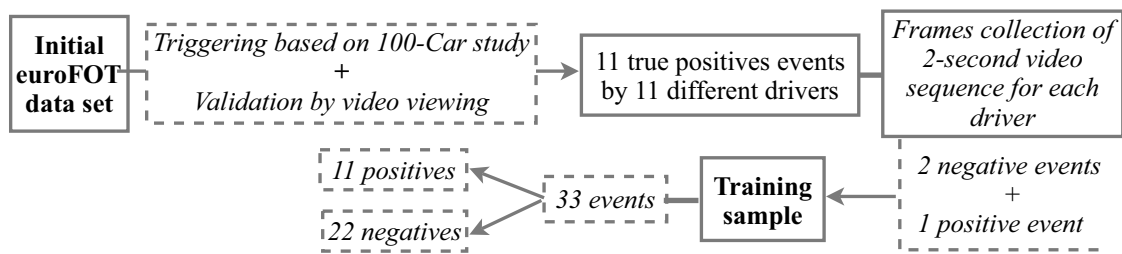


*Figure 2.2     Procedure to define the training sample.*

To sum up, the training sample contains 33 situations, of which eleven are positive events. The fact that these are experimented by different drivers is based on the requirements established for the algorithm, given that the database is formed by 100 drivers.

## 2.3    Recognition of driver's reaction. General structure

Several methods were applied to the training sample in order to:

> 1) Recognize the positive events among the rest of the sample. *Which are the differences in driver's reactions between positive and negative events?*
> 2) Once the differences were established, efforts were focus on finding a way *to automatically detect as many positive events as the lowest possible of negatives.* At this point, it was important to save computational time.

Possible solutions to address both research questions are presented below together with some initial steps to prepare the images. Note that this research aims to identify in a *rough* and fast way the reactions of the drivers. Therefore, more specific and accurate image processing methods, such as defining specific features in the images and analyzing the movement, were not considered. This is mainly limited by the size of the database and the variety of drivers.

### 2.3.1    Data description & Image pre-processing

As specified in previous sections, issues as the computational time and the diversity of drivers play an important role together to rightly identify driver's reaction. Therefore, images were treated as matrices containing pixel intensity values. Since the collected data in euroFOT is available in MatLab, the scripts to access and evaluate the data were also programmed in its language. This section covers technical issues about the structure of the data and initial steps to extract and prepare the images (see Figure 2.3).
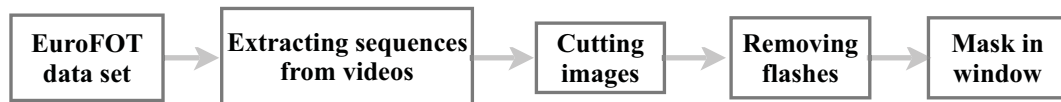


*Figure 2.3      Steps of data acquisition and image pre-processing.*

■ **Extracting sequences from videos_** Three video sequences of two-second duration were extracted from files in format *.avi* for each of the drivers of the training sample[5]. The original images are in grayscale with 288x352 pixels. The frame rate is 12,5 fps[6]. Each frame was saved to a level of array, which is defined by two other structures: *cdata* and *colormap* (see Figure 2.4).
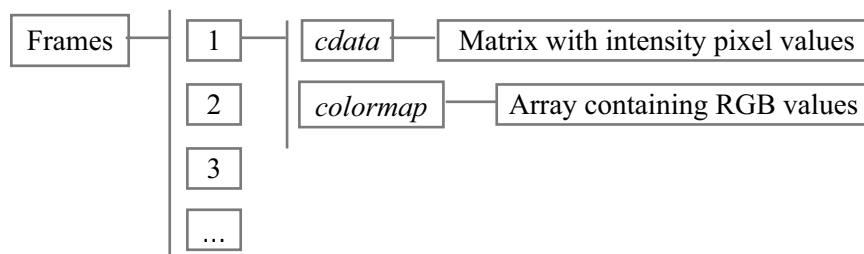


*Figure 2.4      Unfold of array structure and frames information.*

---

[5] some problems to access the frames in long trips with the MatLab function *aviread* were solved by using an application called *videoIO*, developed by Geral Dalley (2006).

[6] frames per second

**◼ Cutting images_** Since images are in grayscale, one plane was enough for defining the pixel values (colored images are defined by three planes). Only a certain area of the matrix stored in *cdata* was saved around the driver's torso to remove superfluous information (see Figure 2.5). Thus, the final sizes of the images were 283x231 pixels.
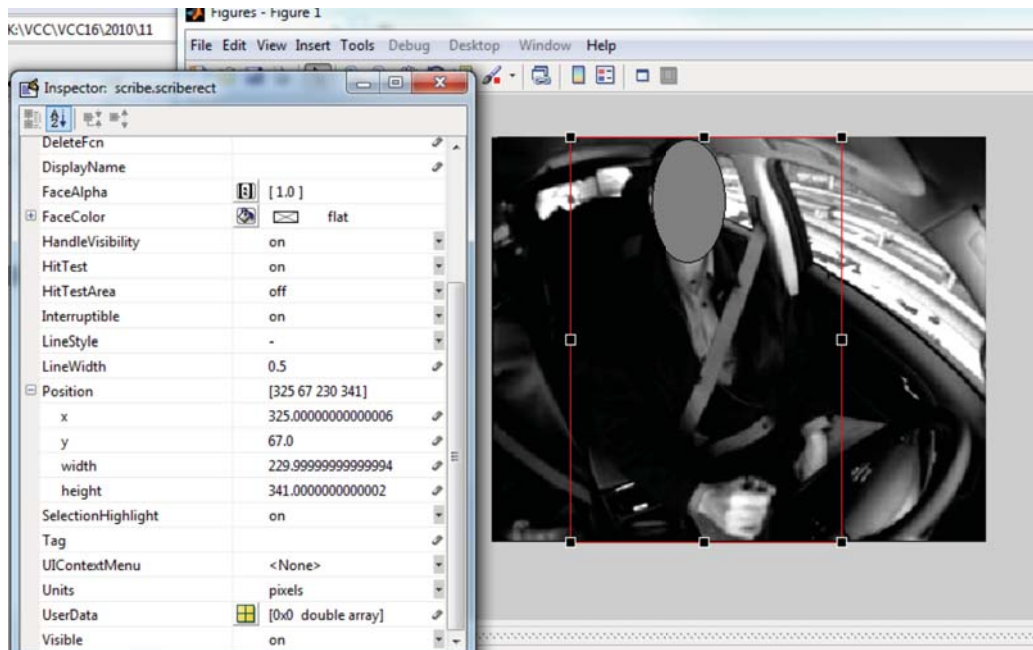


*Figure 2.5     Clipping the torso of the driver[7].*

**◼ Removing flashes_** Over-bright images were removed from the sequences to avoid false changes in pixel intensity. Observations from the training sample indicate a constant frequency of one flashed frame each five. This effect was observed in some of the drivers, but this pre-filtering script was applied to all the sequences without distinction. Although this implies to eliminate right information in some cases and makes the sequences faster than in reality, it is preferable to false intensity changes.



**◼ Mask in window_** Superfluous information, as outside movements in the window's area, may affect the results by generating changes in pixels intensities not related with driver's motion. Therefore, a binary mask changed the pixel's intensities to null values in the window's area. A situation in which the driver's body leans forward was taken as dimensional reference to not lose information of the driver's motion.

*Figure 2.6     Mask polygon in window's area.*

---

[7] the driver's face has been covered due to confidentiality issues.

### 2.3.2 Recognition of driver's reaction in sequences

Following this first phase, the images were already preprocessed and the training sample was defined by a collection of frames for each situation in 33 arrays, of which 11 were positive events. Below, there is an explanation of the three different methods applied in the identification of those events based on driver's reaction recognition

#### 2.3.2.1 T-test and Vartest. Comparison of false&positive events

Frames are defined as matrices containing values of pixel intensity. These values change depending on the motion in the scene. Since negative and positive situations are recorded for each driver, it is possible to compare both to see how different the distribution of pixel values in each case is.

A way to use this information is conducting a **t-test of the null hypotheses that data in a certain pixel position along both arrays of each sequence are from the same normal distribution**. This theory was applied using two different functions in MatLab:

- *Ttest2:* tests the null hypothesis that values for each pixel position come from populations with equal means, against the alternative that means are unequal (unequal variance is assumed).
- *Vartest2:* tests the null hypothesis that values for each pixel position come from populations with equal variance, against the alternative that variance is unequal.

Under 5% of significance level (by default), functions return *h=1* if the null hypothesis is rejected and *h=0* on the contrary, so results can be represented as binary images. In addition, it computes a *p-matrix* containing the probability of observing the values as extremes. Three different populations were considered in this calculation:

- *Intensities* in the same pixel position over time in both sequences.
- *First derivative* values for each pixel position over tie in both sequences: deriving also takes into account the time changes. Those most obvious (the largest change in intensity in less time) were expected to be blank areas in the *h-matrix*.
- *Square of the first derivative* values for each pixel position over time in both sequences: if the pixel intensity decreases during the sequence, the first derivative becomes negative. Then, the square values consider whether this effect can affect the results.

For each population, two binary images resulted from the calculation of two different *t-tests*: one with two negative events and the another with a positive and a negative event. The procedure was the same when performing a *vartest*.

It was expected that the binary image resulting from the comparison of a negative and a positive events contained more white areas than the resulting from the two negatives. This would mean that the intensities in those pixels have experienced more changes and, therefore, rejected the null hypothesis. According to this theory, the positive event could be recognized in the following steps (see Figure 2.7):
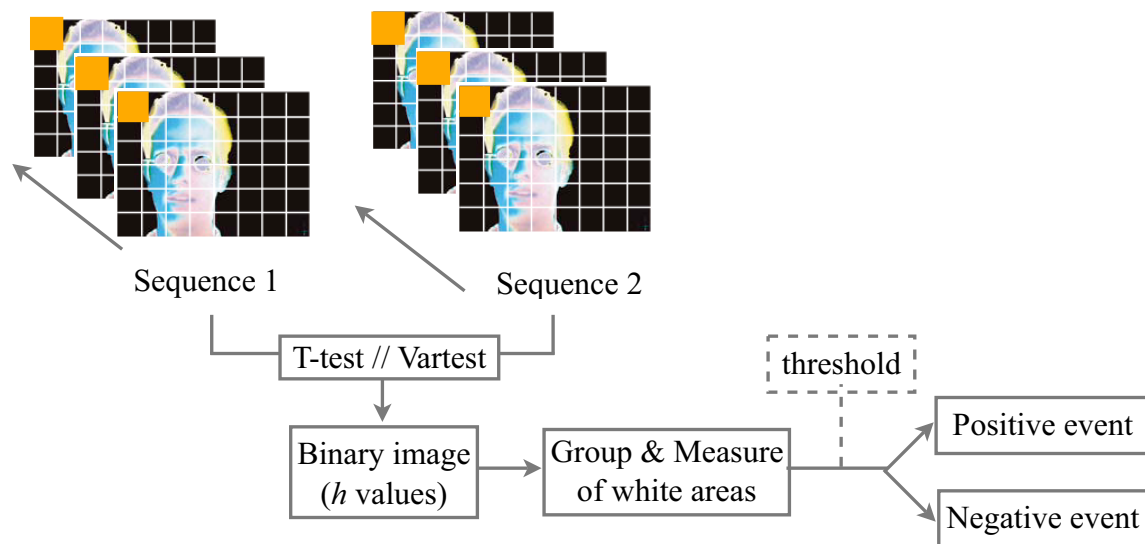
*Figure 2.7     Procedure of recognition using t-test&vartest as potential triggers.*

First, *t-tes*t and *vartest* were applied to compare a negative and a positive sequence for the three different populations. The aim was to detect which population and which test were most suitable for recognizing the driver's motion. Results of the first approach are shown below (see Figures 2.8 and 2.9):
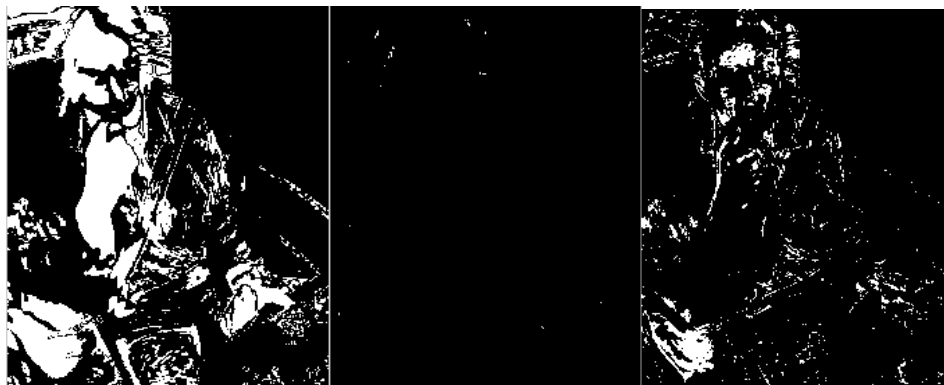


*Figure 2.8     Binary images from vartest of: intensities, first derivative of intensities and square of first derivative of intensities (driver A686).*



*Figure 2.9    Binary images from vartest of: intensities, first derivative of intensities and square of first derivative of intensities (driver A686).*

As can be seen in Figure 2.8, the binary image from the t-test of square-of-first-derivative of intensities seemed to be the most representative of the driver's motion. It highlights the areas that have undergone major changes: the hand and the driver's head.

Using the square-of-first-derivative of intensities as population, t-tests were performed in two negative sequences and in a positive and a negative sequences applying the procedure detailed in Figure 2.7. When comparing the positive and the negative events, more white areas were expected on the resulting binary image that collects the $h$ values. Finally, grouping and measuring these white areas might be used to determinate which of two comparisons belong to a positive event (driver's reaction).
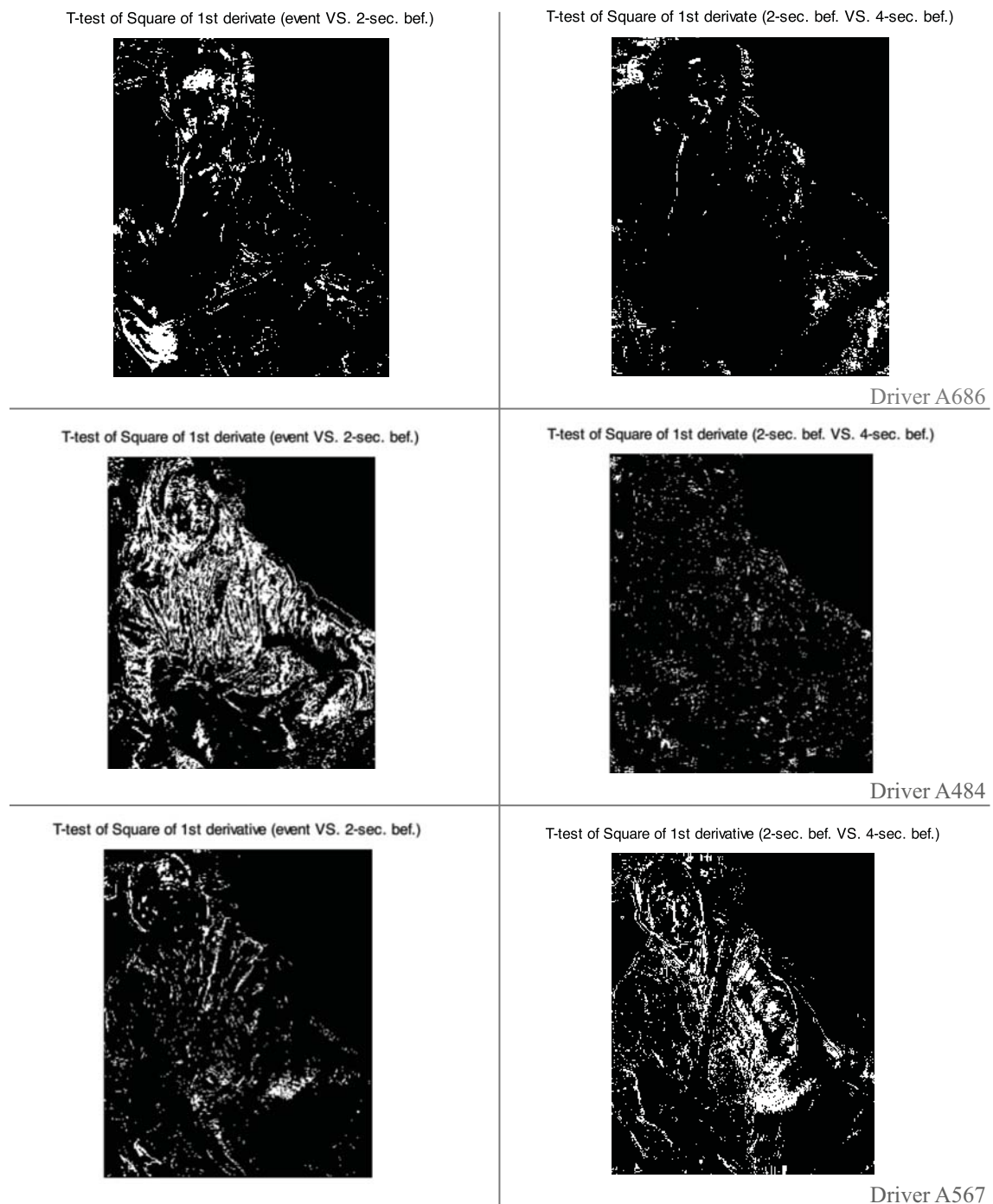


Figure 2.10    Binary images from t-test of square of first derivative of intensities.

In most of the drivers, the t-test comparing positive and negative sequences (*event VS. 2-second bef.*) resulted in more white areas than the obtained from two negative sequences. This effect is observed in drivers A686 and A484 in Figure 2.10, since the images on the left have more white areas than the images on the right resulting of the comparison of two negative sequences. Nevertheless, the opposite occurred in the driver A567. Similar results are also observed in some of the drivers of the training sample.

The binary images are quite noisy, which makes difficult to relate the driver's reactions with the white areas from the rejected null hypothesis. Then, it seems unclear to define a certain white-area threshold to highlight the reaction. The main problem is that **some maneuvers while driving may be related with a broad change in the rate of pixel intensities**. For instance, turning the steering wheel creates a larger white area when is compared via t-test with a sequence of driving on a straight road. Then, alternative methods were explored to distinguish the driver's reactions.

### 2.3.2.2  Standard Deviation of Jerk

Given that results from the t-test were unclear in the discrimination between positive and negative events, possible alternatives are discussed below.

As noted with the conventional triggering procedure, the reactions of drivers are mostly related to sudden quick motions of the driver's torso. Therefore, the time in which the action takes place appears to be an important factor. One way to take the motion's time into account is by deriving intensities in each pixel position over time.

Velocity (first derivative) and acceleration (second derivative) represent rate of change of position and velocity over time, respectively. If each derivation level is related to a rate of change of what is deriving, then the third derivative represents a rate of change of acceleration. Young relates *control* with *third derivative* in "The Reflexive Universe" to explain any fact of the daily life. He illustrates that controlling the car can be expressed with the third derivative since it is related to changes in acceleration (Young, 2004). The third derivative is also called *jerk*, and its application can be extrapolated to various fields of mathematics and engineering (Iradier, 2006).

**The jerk of intensity values at each pixel location over time can give an idea of how sudden these changes are.** Calculating the jerk for each driving sequence allows studying whether the positive sequences have values significantly different from the negatives. From this calculation, an array containing matrices with jerk values for each sequence was obtained as result. Two ways to look into these arrays were considered:

- **Computing the standard deviation (STD):** The wider variance in a normal distribution of jerk values, the more different that they have been over time. It was expected from this analysis that the highest changes in accelerations were represented as whiter areas in a grayscale image.

- **Representing the maximum square of jerk values:** Peaks in jerk distribution can also be represented as white areas in a grayscale image, without taking into account how different these values have been in the distribution. The squared values are used to avoid any influence from negative numbers when the pixel intensity changes to a lower value by deriving.

Prior to compute the jerk and its variance, over-bright images were removed from the sequence to avoid false changes in pixel intensity. Observations from the training sample indicated a constant frequency of one flashed frame each five in most of the drivers. This pre-filtering algorithm was applied to all the sequences without distinction. Thus, it was possible to lose information by removing right frames and it also made sequences faster than in reality.

The resulting images from both calculations in the three sequences for one of the drivers (two negatives and one positive) of the sample are presented below. The goal was to distinguish the positive event (called just *event* from now) from the other two situations:
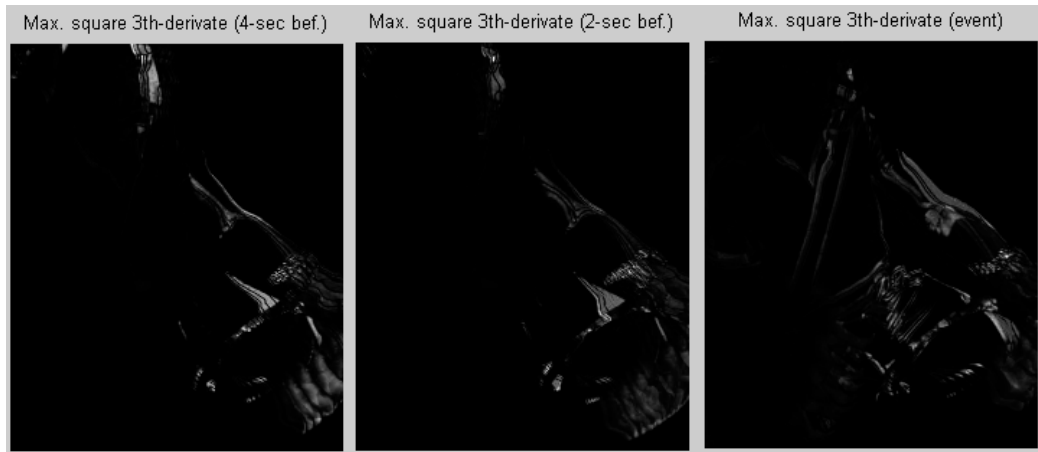


*Figure 2.11 Maximum of square of jerk values: negative, negative and positive events.*



*Figure 2.12    STD of jerk values: negative, negative and positive events (driver A241).*

The rates of acceleration changes during the event were not significantly different from those obtained in negative sequences. In fact, maximums were mainly achieved when drivers were maneuvering in sequences previous to the event. Thus, driver's reaction might not be related to peaks in jerk values. However, differences between negative and positive images from the *standard deviation of jerk* (STD of jerk) seemed noticeable, as shown in Figure 2.12.

As can be seen in the image from the positive event in Figure 2.12, a **white silhouette of the driver** appeared when calculating STD of jerk during the event. If driver had

remained in the same position, a dark image was obtained as a result. On the other hand, maneuvers seemed to generate a certain white area in the image. This effect can be observed in the image on the middle in Figure 2.12, when the driver turns the steering wheel. In this context, it's interesting the case of one of the drivers of the sample:



*Figure 2.13    STD of jerk values: negative, negative and positive events (driver A1064).*

As commented before, the sequences of the training sample were chosen randomly from the previous triggering procedure. In this specific case, it was unexpected clearly recognize the driver's reaction since the motion was almost imperceptible in the video sequence. However, by looking at the Figure 2.13, the driver silhouette obtained in the positive sequence enables to make a distinction respect to the other two previous states.

These results support the theory that the **driver's reaction during a CRE seems to involve the whole body (rigid body motion in the driver's torso), while maneuvers seem to involve just a certain part**. This fact makes important to consider the area in which the changes take place to make the distinction.

In most of the drivers of the sample, the resulting driver's silhouette from STD of jerk allowed an intuitive recognition of the driver's reaction and hence the positive events. Therefore, the possibilities were to post-process the STD-of-jerk images to identify the threshold that relates the graphic silhouette with the driver's motion (a conversion from graphical to numerical information), or to keep trying other methods.

### 2.3.2.3   Optical Flow

A numerical alternative to the graphical method discussed in the previous paragraph was the calculation of the *optical flow*. Its original formulation came from Horn and Schunck (1981), who defined optical flow as "*the distribution of apparent velocities of movement of brightness patterns in an image*". This distribution provides information about the object motion in terms of spatial allocation and rate of change. The optical flow constrains equation is defined as:

$$Ix \cdot u + Iy \cdot v + It = 0 \quad \begin{cases} Ix, Iy, It: \text{spatiotemporal image brightness derivatives.} \\ u = \text{horizontal optical flow} \\ v = \text{vertical optical flow} \end{cases} \quad (2.1)$$

This equation relates the intensity changes in a sequence of images with three-dimensional object motion. Nevertheless, this relationship could result unclear in some cases. For instance, the optical flow is zero in all the points of a rotating movement of a sphere. However, assuming the surfaces are flat, motion may be related to changes in brightness. Therefore, velocities of object motion can be estimated by solving u and v[8].

During the last years, several studies have been carried to improve the performance of the classical formulation. In this context, D. Sun, S. Roth and M. J. Black (2010a,b) have recently contributed developing a more accurate optical flow. They published a public Matlab code to compute this new optical flow formulation under educational proposes (Sun, 2010). Outputs are two matrices with the horizontal and vertical components of optical flow (OF from now) for each pair of processed images.

■ **Estimating flow in drivers_** The OF code developed by Sun *et* al. (2010) were implemented in the training sample with the parameters established by default[9]. The main objective was to assess **whether the body's motion of the driver can be estimated with changes in brightness in a two-dimensional image.**

Such script returned two matrices containing speed components for each pair of computed images. In this case, matrices were combined into a single keeping the magnitude of speed, since this value seemed more significant than the flow's direction. Each matrices collection was kept in an array for each of the sequences in the training sample.

As commented in the previous section, the highest changes in acceleration in STD of jerk images weren't reached during the driver's reaction. This led to consider other alternatives that peaks in speed along the arrays to make the distinction.

As happened in the calculation of jerk, pre-filtering alters the results since intermediate values are lost. However, this was preferable to false intensity changes due to flashes. Taking as reference one of the drivers of the sample, initial estimates consisted on using speed data from the OF calculation in each array (with and without filtering frames) to calculate:

- **Local maximum from Optical Flow** (OF): peak in speed of optical flow over the array of each sequence.
- **Maximum sum of the whole array:** maximum value in a matrix resulting from the sum of individual matrices with velocities of the entire OF array for each sequence.
- **Number of pixels above or equal to the 95% of the peak** in a matrix resulting from the sum of individual matrices of the entire OF array for each sequence.

---

[8] see further information about Optical Flow and its formulation in Appendix 3

[9] it's also possible to test different methodologies of flow calculation; by default the program use the "Classic+NL-Fast" (only pixels from certain regions are weighted to save computational time).

*Table 2.1      Comparative of OF values in negative and positive events (driver A686).*

| Sequences (driver A686) | Original sequence | | | Filtered sequence | | |
|---|---|---|---|---|---|---|
| | Local max. | Max. sum of the whole array | Nº of pixels with speed>=95% of the highest speed in total sum matrix | Local max. | Max. sum of the whole array | Nº of pixels with speed>=95% of the highest speed in total sum matrix |
| Negative (4 sec. before) | 39,79 | 74,5798 | 1 | 38,98 | 59,7046 | 155 |
| Negative (2 sec. before) | 30,49 | 93,219 | 122 | 28,65 | 84,4021 | 277 |
| Positive event | 18,88 | 112,603 | 206 | 11,01 | 96,6913 | 298 |

By looking at the Table 2.1, some differences might be identified between the calculation with the original and the filtered sequence. Anyway, maximum values were achieved in the same categories. Local peaks of speed were not recorded during the positive event. However, it registered the maximum when considering in the calculation a single matrix containing the sum of speeds over time. Besides, this matrix had more pixels with higher sum of velocities than the negative sequences.

Anyway, differences were not enough consistent to establish these values as criterion of discrimination. For instance, the number of pixels containing the highest sum of velocities was 277 in the negative filtered sequence (with 84.4021 of peak speed) and 298 in the positive (with 96.6913 of peak speed). This suggested **analyzing the speed of the optical flow between single frames instead of using a matrix containing the sum of values along the array.**

Results of performing the same calculation on single frames from each of the sequences for the same driver are presented below:

■ **Average speed on single frames**
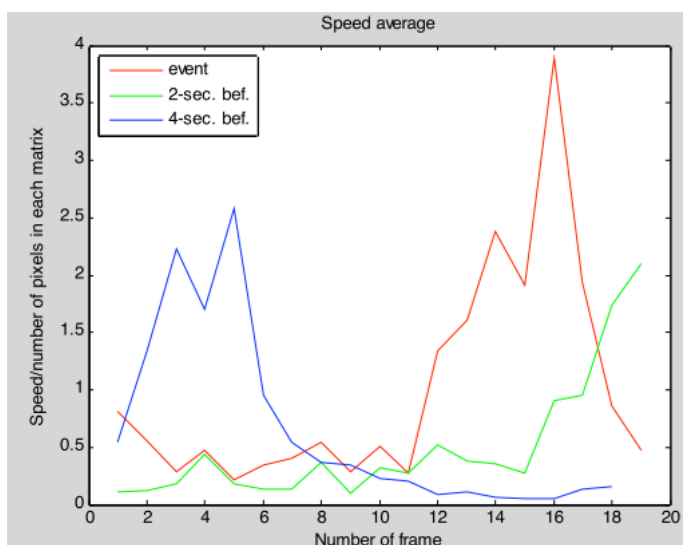


As can be seen in the Figure 2.14, the average speed was significantly greater during the event (red line in the graph) than in negative sequences (obtained two and four seconds before the event). The peak was achieved in the 16th matrix of the OF array resulted from processing the images of the positive sequence.

*Figure 2.14    Speed average over OF frames in positive and negative sequences.*

**■ Peak speed on single frames**
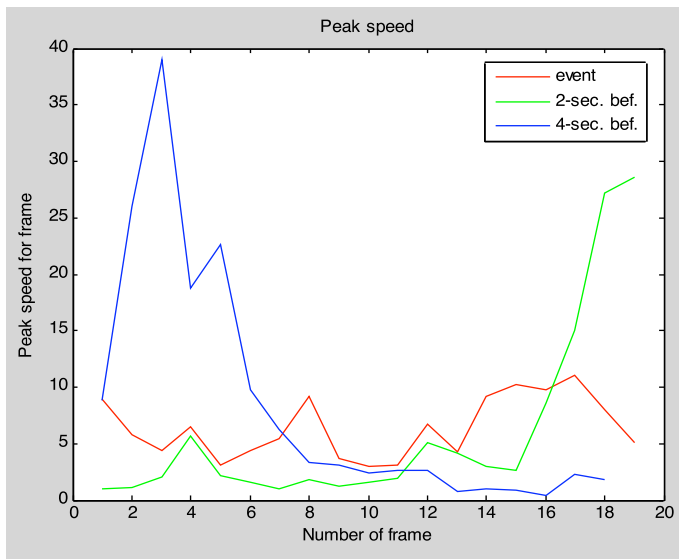


However, the peak speed wasn't obtained in the OF of the positive event. As can be seen in the Figure 2.15, the highest peak speed was recorded in the negative sequence, which takes place four seconds before the event.

*Figure 2.15    Peak speed over OF frames in positive and negative sequences.*

**■ Number of pixels sharing highest peak speeds on single frames**
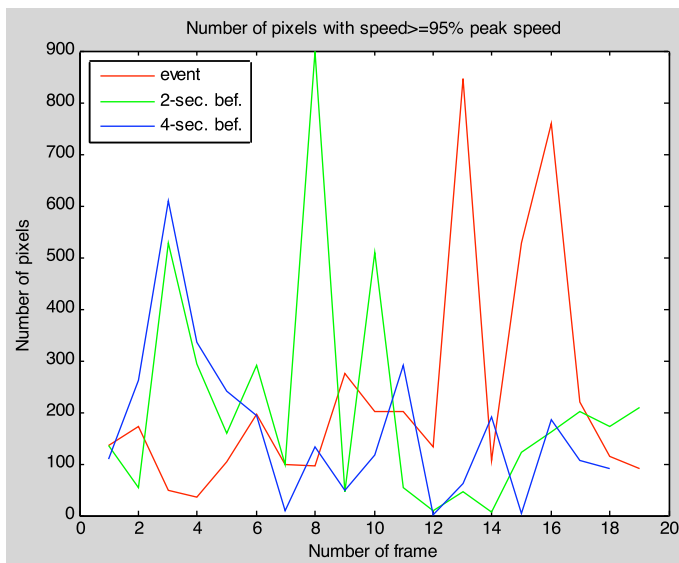


As shown in Figure 2.16, at the time of maximum speed average ($16^{th}$ matrix of the event sequence, see in Figure 2.14) the number of pixels with, at least, the 95% of the peak speed was 760. Another maximum was observed on the $13^{th}$ of the array with 847 pixels. However, in comparison with other sequences, the maximum was registered in one of the negatives two seconds before the event.

*Figure 2.16    Peak speed over OF frames in positive and negative sequences.*

The most significant speed rate during the event was recorded in the $16^{th}$ matrix of the OF array. It resulted from the estimation of speeds between $16^{th}$ and $17^{th}$ frames of the original filtered sequence. Note that OF interpolation warps the second image and its derivative toward the first[10]. Those original frames are shown below to know the significance of such peak speed:

---

[10] see optical flow formulation in Appendix 3.

*Figure 2.17    Frames of the event sequence corresponding to peak speed (driver A686).*

As can be seen in Figure 2.17, *Frame 16* corresponded to the biggest change in motion during the reaction, when the body leaned forward. In *Frame 17* driver was back to the original state. The change in motion was evident in this part of the sequence.

The greatest differences between the event and second-before sequences seemed to be related to the average speed on single frames. According to this, sudden changes might be recognized by deriving. The Figure 2.18 shows the second derivative values of the average speed vectors for each sequence.

## ◼ Jerk of average speed on single frames
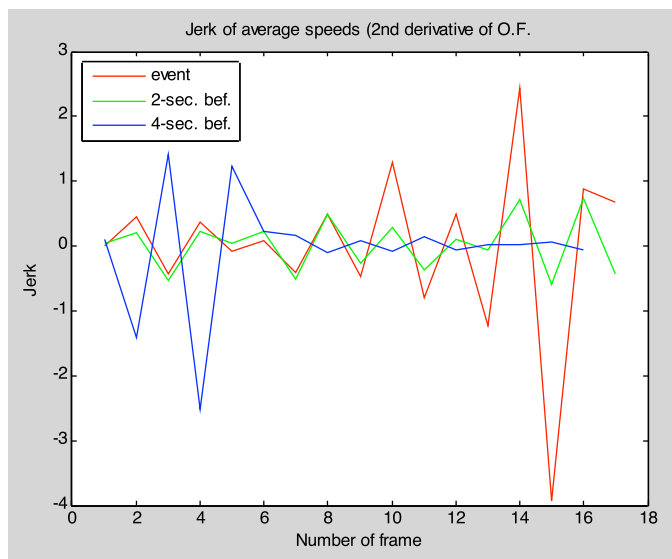


The biggest slope matched with the change in motion between the 14th and 15th values during the event sequence. At that moment, driver's body leaned forward due to the inertia of harsh braking.

Rates of acceleration changes recorded two seconds before the event kept values into a rate during the whole array. However, other significant change occurred between the 4th and the 5th derivative values four seconds before the event.

*Figure 2.18    Jerk of average speeds over OF frames in sequences.*

As can be seen in Figure 2.19, the second biggest slope in the sequence that occurs four seconds before the event was due to a change in the driver's position:

*Figure 2.19    Frames of the negative corresponding to peak speed (driver A686).*

Given these findings, it seemed that **peaks in the distribution of jerk from OF velocities were associated with the driver's motion**. Nevertheless, the main obstacle is the computational time required to estimate the OF

## 2.3.3    Silhouette detection in STD of Jerk images

In the last section, several methods were applied in the recognition of driver's reaction in presence of CREs. Among these methods, the OF and the STD of jerk were identified as potential algorithms. The main limitation of the OF was the computational time spending. Although it didn't concern the STD of jerk, its results were graphical. So, this graphical information should be converted into numerical to facilitate an automatic detection method. This automatic detection mostly involved the study of the properties which characterize the images of the STD of jerk in the event among the negative sequences.
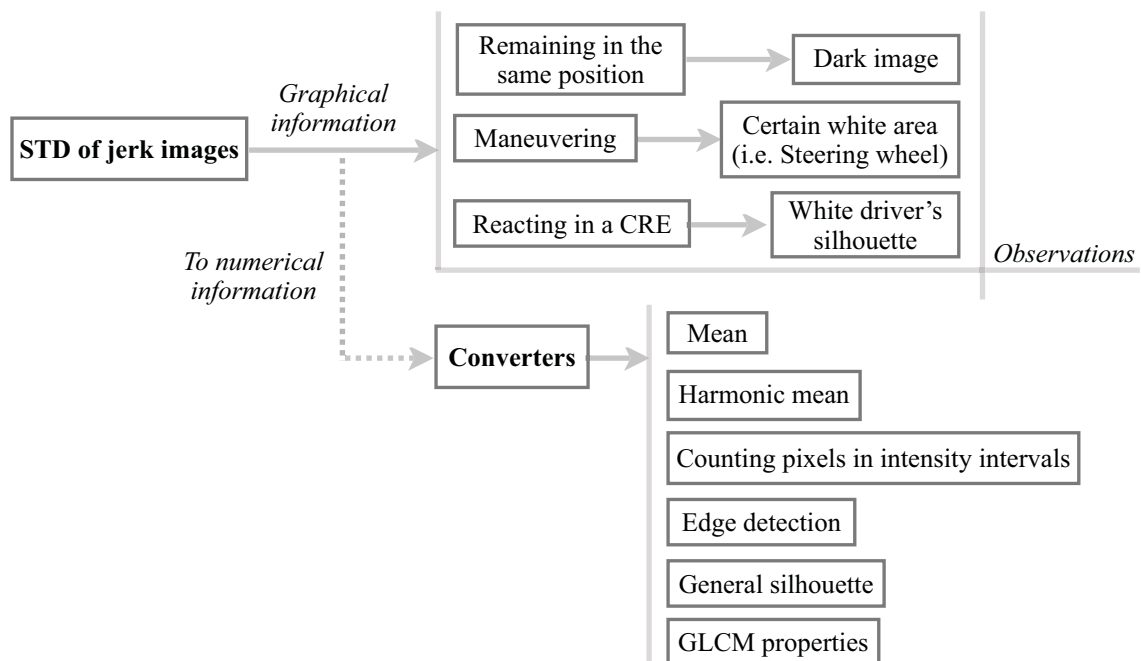


*Figure 2.20    Converters from graphical to numerical information in STD of jerk.*

### 2.3.3.1 Mean

A first alternative to the use of graphic information from the STD of jerk was conducted by plotting the sum of the values along rows. The reason was to try to find out a silhouette in the images corresponding to the driver's reaction. According to this hypothesis, it was thought that the mean[11] would be higher during the event than in previous sequences. This was based on the dispersion of pixels over the image to generate the driver's silhouette.

Taking as reference a driver from the training sample, the STD of jerk values were sum along rows for each of the sequences. The resulting vectors are plotted below together with the STD of jerk images of the driver (see Figure 2.20 and 2.21).
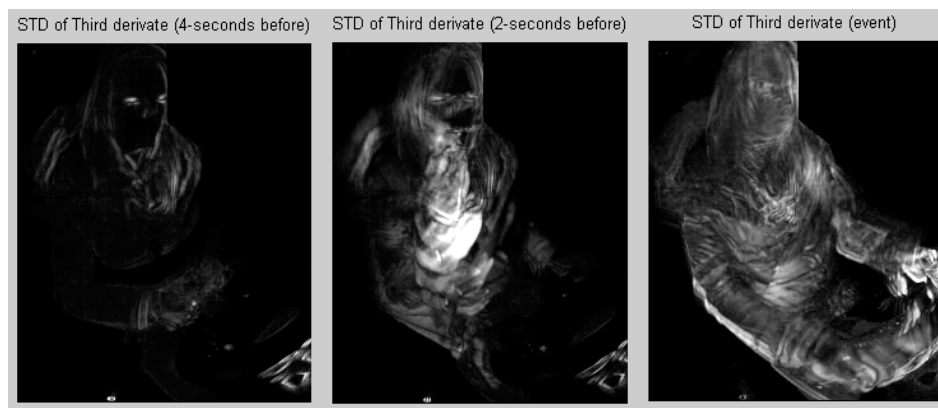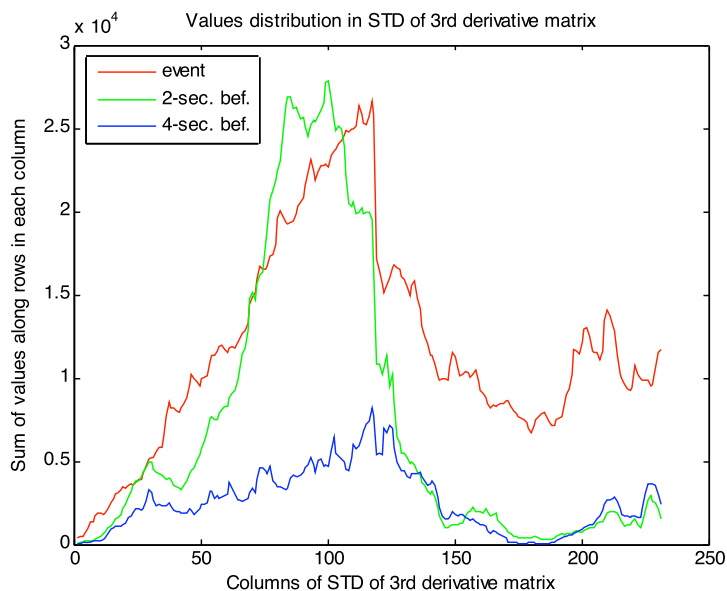


*Figure 2.21    STD of jerk values: negative, negative and positive events (driver A936).*



*Figure 2.22    Distribution of sum of STD of jerk values along rows (driver A936).*

As evidenced in Figure 2.20, the sudden motion from the driver's reaction generated a white silhouette in the event sequence. Its distribution of STD values along rows reached a mean of 11900, while means in previous sequences are 2598 y 7299, respectively. The peak in the distributions of STD corresponded to the two-second before image, where a bright white area is concentrated in the middle of the figure.

This result was unexpected since in the pre-filtering procedure over-bright images were removed. By reviewing the video it was checked that this area corresponds to a movement of the driver, who moves the arm from the steering wheel to the mouth.

---

[11] arithmetic average of the values over the distribution.

### 2.3.3.2 Harmonic mean

In educational fields, the harmonic mean is commonly used to calculate the final grades of students to ensure a reasonable level of work during the academic year (Wilson, 2006). This case was also somewhat related given that mean was affected by local peaks in distributions. Hasna and Alouini (2002) used this formulation to study the performance of wireless communication. They defined the harmonic mean as follows:

*"Given two numbers $X_1$ and $X_2$, the harmonic mean of $X_1$ and $X_2$, $\mu_H(X_1,X_2)$, is defined as the reciprocal of the arithmetic mean of the reciprocals of $X_1$ and $X_2$, that is:*

$$\mu_h(x_1,x_2) = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} \tag{2.2}$$

*It is clear that the harmonic mean of two numbers is equal to the square of their geometric mean divided by their arithmetic mean."* (Hasna and Alouini, 2002).

The harmonic mean is not affected by the outliers (related to maneuvers) and it's also useful when the data are resulted from indirect calculations. In this case, jerk belongs to derivatives of intensity changes in pixels over time. Therefore, this method was thought as an alternative of the mean calculation when distributions of STD of jerk were affected by normal driving maneuvers and position changes in the driver.

As can be seen in Table 2.2, the mean of the negative sequence that takes place 4-seconds before the event was 7299, which corresponds to a 61,33% of the mean achieved during the event (11900). This influence was lower in the case of the harmonic mean. Assuming as 100% the harmonic mean achieved during the event (6393), the value recorded in the 4-seconds-bef. sequence represents a 18,62% (1190,5).

*Table 2.2    Comparison between mean and harmonic mean in the distribution of sum of STD of jerk values along rows in sequences of driver A936:*

| Criterion | 2-sec. bef. | 4-sec. bef. | Event |
|---|---|---|---|
| Mean | 2598 | 7299 | 11900 |
| Harmonic mean | 369,8246 | 1190,5 | 6393 |

### 2.3.3.3 Counting pixels in intensity intervals

Since the maximum values of STD of jerk weren't achieved during the event, another option was to consider the number of pixels with a certain STD within an interval. This same concept is in the calculation of the image histogram. This method allows representing the intensity levels respect to the number of pixels that share such intensities. Histograms can be used to obtain the parameters of a texture (Alba *et al.*, 2006). In some way, the driver's silhouette is related to a texture given that it's defined by a relationship between pixels. Some of the properties of histograms can be summarized as follows (Olmos, 2008):

- Images can't be rebuilt from their histograms.

- Two images can be associated to the same histogram.

- Histograms not contain spatial information about the image.

In the next trial, the driver *A686* was took as reference to analyze the intensity levels generated by changes in driver's position. Images of STD of jerk from the event and a previous sequence are represented below in a three-dimensional graphic (see Figures 2.23 and 2.24). This graphical representation gave an idea about the rates of STD and their location over the image in a negative and in a positive situation.
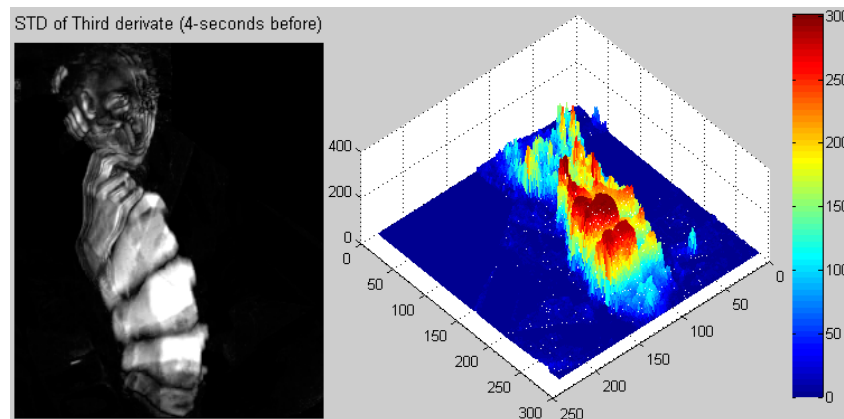


*Figure 2.23    Tridimensional distribution of STD of jerk over the image in one of the negative sequences of driver A686.*
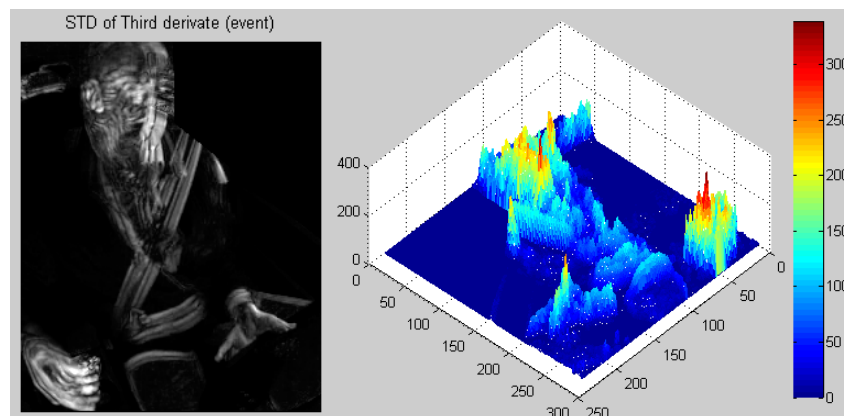


*Figure 2.24    Tridimensional distribution of STD of jerk over the image of the positive sequence (event) of driver A686.*

As shown in the graphs above, the negative sequence concentrated highest values of STD in the area generated by the hand movement. It seems that the widest variances might be not related to the driver's reaction. Results from counting the number of pixels within certain intervals of STD of jerk in each images are presented as follows:
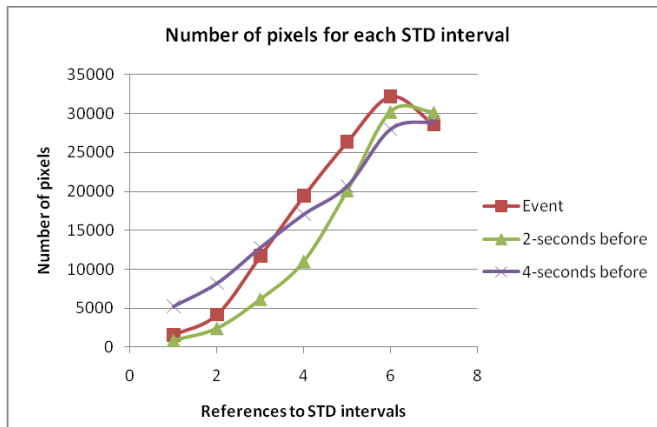
As can be seen in Figure 2.29, the differences were more significant in the fourth and fifth interval of STD. These intervals were groups of pixels with STD of jerk between 50 and 20. In this range of STD, a higher number of pixels were counted during the event than in the previous sequences.

*Figure 2.25    Distributions of number of pixels within intervals of STD in the event and in previous sequences for the driver A686.*

The numerical values associated to Figure 2.25 are represented in Table 2.3:

*Table 2.3      Number of pixels within intervals of STD of jerk during the event and previous sequences.*

| Driver A686 | Number of pixels within a certain interval of STD of jerk | | | | | | |
|---|---|---|---|---|---|---|---|
| Nº of interval (STD) | 1 (300-150) | 2 (150-100) | 3 (100-50) | 4 (50-20) | 5 (20-10) | 6 (10-5) | 7 (0) |
| Event | 1542 | 4178 | 11708 | 19422 | 26384 | 32199 | 28559 |
| 2-second before | 871 | 2430 | 6167 | 11030 | 20147 | 30255 | 30152 |
| 4-second before | 5229 | 8195 | 12788 | 17063 | 20705 | 28007 | 28962 |

By adding the values of the fourth and fifth intervals (columns "4 (50-20)" and "5 (20-10)" in Table 2.3), the numerical difference between the event and the previous sequences was not enough significant to discriminate between both situations. The number of pixels during the event was 45806, while 31177 and 37768 pixels were counted in previous sequences, respectively. This suggested taking into account the **spatial distribution of pixels** in next tests.

### 2.3.3.4   Edge detection: Hough transform

The Image Processing Toolbox in MatLab contains several procedures to detect edges in an image. In the following test, the Hough transform method was applied to a positive and a negative situation for the same driver. This method is based on the parametric representations of lines in a plane (MathWorks, 2011):

$$\rho = x \cdot \cos\Theta + y \cdot \sin\Theta \tag{2.3}$$

The general procedure consists in detecting edges using *Sobel* or *Canny* algorithms. The resulting images may have open forms and isolated points. Then, the correction is possible by taking an initial point and drawing straight lines in a polar coordinate system. ρ and θ values are accumulated in a matrix called *Standard Hough Transform* (SHT) to guess which pixel is more likely to belong to each edge. Peaks in SHT represent potential lines in the input image. Finally, *houghlines* command finds the extremes of the lines and fills the small gaps.

This method was applied to a pair of images from the same driver, one obtained from an event and another from 2-second before the event. The *Hough* transform was represented in a graph and its peaks (potential lines) appeared in squares. Then, the detected lines were colored on the input images.
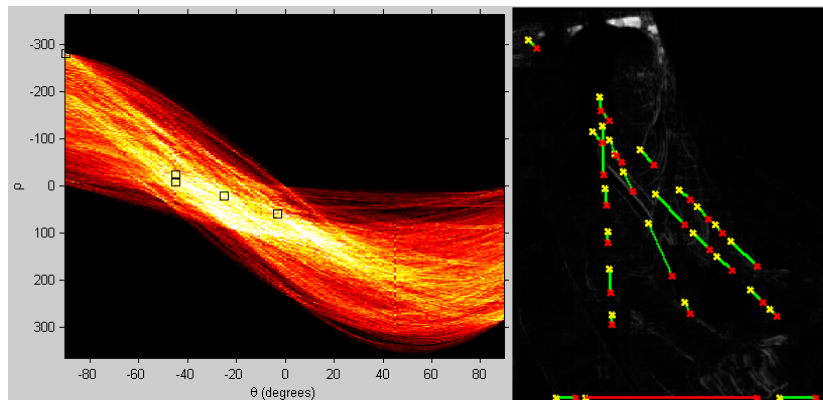


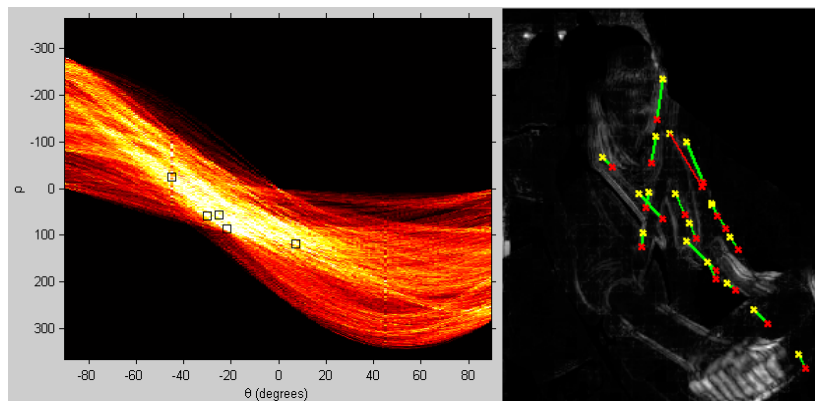*Figure 2.26    Hough transform and detected lines from sequence 2-sec. before the event*



*Figure 2.27    Hough transform and detected lines from the event's sequence.*

By looking at Figures 2.28 and 2.29, the detected lines were not clearly different to discriminate between both situations. This method is usually useful in detecting roads in aerial images. However the straight lines seem not fit with the driver's silhouette. This method can be also applied with curve lines by previously defining an original shape. This shape might be not clear definition in this case, due to the variety of drivers and the camera positions.

### 2.3.3.5 General silhouette

Since maneuvers and changes into positions seem to generate white areas in the STD of jerk images, another possibility in the identification of events was to define a certain area where the reactions were likely to take place to avoid false positives.

To define this area, several silhouettes obtained from different drivers of the sample were combined into a single one. Three different procedures were applied using the command *wfusimg* in MatLab. This program merges two images using fusion methods. Thus, the images containing the drivers' silhouettes from STD of jerk during the events were merged in pairs (see procedure in Figure 2.28):
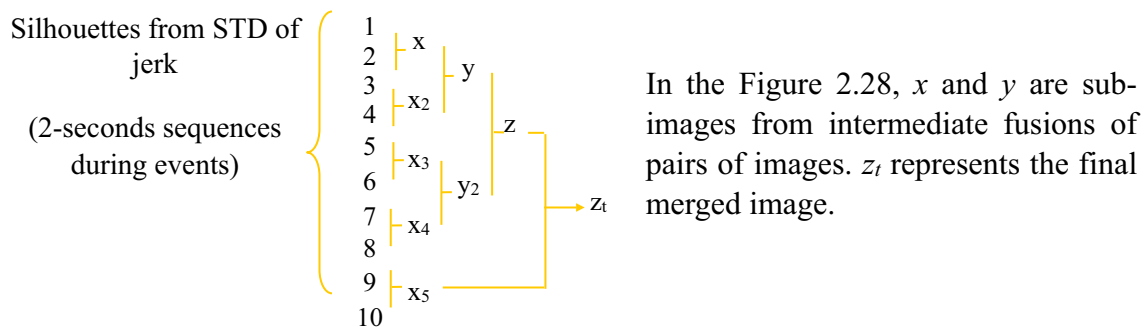


In the Figure 2.28, *x* and *y* are sub-images from intermediate fusions of pairs of images. $z_t$ represents the final merged image.

*Figure 2.28    Schema of combination of silhouettes.*

The command *wfusimg* allows to define levels of approximations and details. The following are *$z_t$ images* resulting from variations in these parameters:



-Maximum for approximations and minimum for details:

-Maximum absolute for approximations and details:
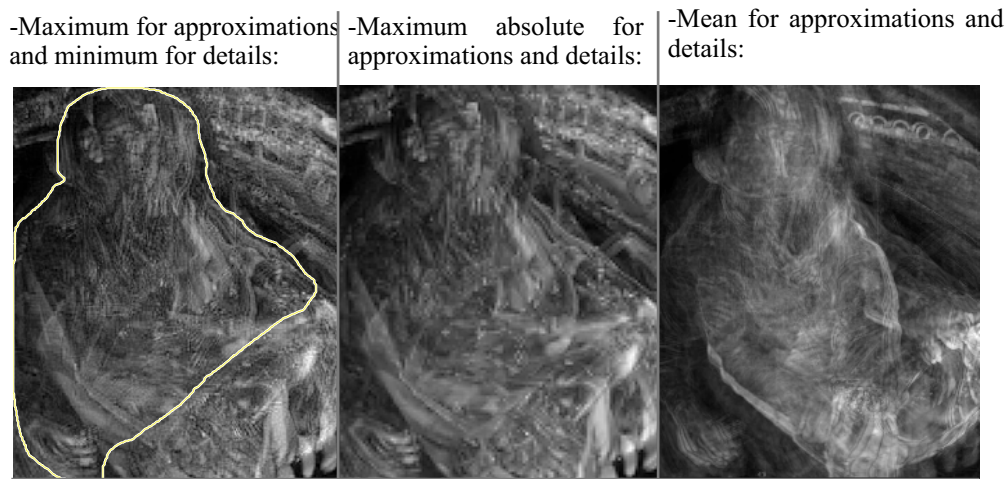
-Mean for approximations and details:

*Figure 2.29    Resulting images using different inputs in the fusion command.*

Matrices of STD of jerk contain different values depending on the movement of the driver during the sequence and the illumination conditions, for instance. Merging images based on mean values for approximations and details (see image on the right in Figure 2.29) tends to highlight the drivers with widest variances in such matrices.

However, the merged silhouette should provide an idea about the area in which commonly reactions take place, regardless of STD values.

Since the merged images with maximum and minimum levels were quite similar, one of them was selected and a freehand region was drawn around the driver's place (see image on the left in Figure 2.29). Areas of windows, rear seats and the steering wheel were not taken into account to avoid false positives. Although several reactions and evasive maneuvers are related to turn the steering wheel, this area seems to tend to confusion when discriminating between positive and negative events.

This procedure was just an approximation to facilitate the study of changes in pixel intensities in a given area. The position of this Region Of Interest (ROI) was saved into an N-by-2 array in MatLab. This ROI could be applied as binary mask to the image in combination with the rest of the methods, aiming to improve their performance.

### 2.3.3.6   Gray level co-occurrence matrix

Texture filters often use the image's histogram to statistically evaluate the texture. Although this provides information about its properties, shape or spatial distribution over the image are unknown (IZMIRAN, 2005).

Another statistical procedure of texture analysis that considers the spatial distribution is the Gray Level Co-occurrence Matrix (GLCM). GLCM contains **how often pairs of different combination of pixel intensities occur in an image** (see procedure in Figure 2.35). This texture analysis is originally from Heralick *et al.*(1973) and today is commonly used in medical image processing, modeling of forests attributes or studying the sea-ice, among others. In this case, this method was thought to identify the driver's silhouette in images of STD of jerk based on its distribution.
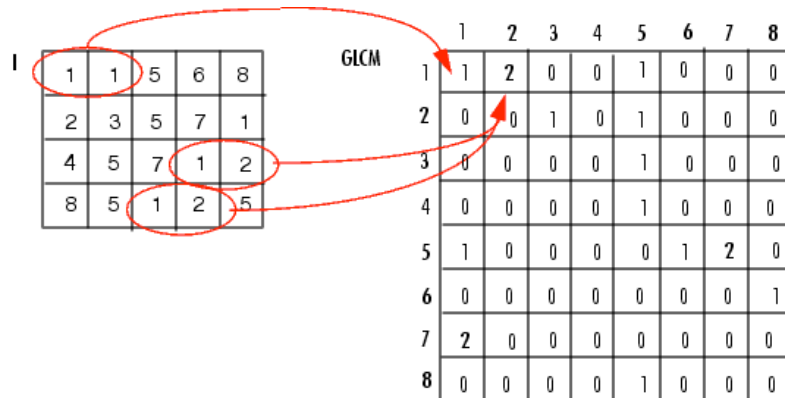


*Figure 2.30   Process Used to Create the GLCM, [electronic print] Available at <http://matlab.izmiran.ru/help/toolbox/images/enhanc15.html>[Accessed May 2011].*

This method can be applied in two main steps:

■ **Definition of GLCM:** Calculating the frequency of certain relationship between pixels requires the choice of:

-**Offset:** distance between the related pair of pixels.

-**Direction of offset:** direction in which the pair of pixels are going to be evaluated. This choice is based on a visual examination of what it's likely to be more characteristic of the texture.

-**Gray levels:** the input image is scaled in a certain number of intensity levels. The lower scales, the lower computational time. Besides, the statistical study is improved by reducing the number of levels.

■ **Calculation of statistics using GLCM:** Once the GLCM is defined, several statistical methods can be used to identify the texture's properties. Hall-Beyer (2007) has created an online tutorial about how to define a GLCM and its possibilities. She defines three main groups derived from GLCM calculations, which are summarized as follows together with the possibilities offered in MatLab:

Measures group 1: distance to the GLCM diagonal (**contrast**)

> **Contrast:** the diagonal of the GLCM contains pairs of pixels with the same gray level. If there is a high frequency of these combinations, then the image doesn't have much contrast. This measure is the sum of square of variances and increases away from the diagonal(=0 if constant image).

> **Homogeneity:** closeness in the distribution of combinations in the GLCM. It increases with less contrast (=1 in the diagonal).

Measures group 2: how **regular** the pixels are within the image
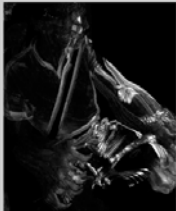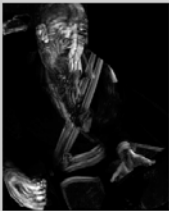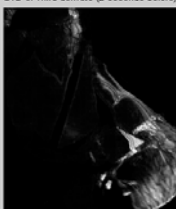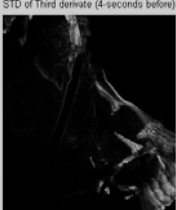
> **Energy:** uniformity in the image that is measured by adding the squared elements (moment of inertia) in the GLCM (=1 for uniform image).

Measures group 3: descriptive **statistics** of GLCM

> **GLCM correlation:** dependency of gray levels between neighboring pixels (=+1 or -1 for perfectly correlated image). This doesn't take into account the frequency of occurrence of a pixel, but its frequency together with a given pixel value.

The Table 2.4 contains some results of testing the GLCM together with a Fusion mask in two drivers of the sample. The drivers were turning the steering wheel and moving the hand in previous sequences. These situations were chosen to define problematic situations that could interfere in the recognition of the driver's silhouette.

*Table 2.4     Properties in four directions of the GLCM in drivers A241 and A686 (offset=200).*

| | Driver A241 | Driver A686 | Contrast | Correlation | Energy |
|---|---|---|---|---|---|
| Event |  |  | **A241:**<br>[0.1598 0.0007 0.2925 0.0031]<br><br>**A686:**<br>[0.6412 2.1862 1.9693 0] | **A241:**<br>[0.3671 -0.0004 0.2036 -0.0016]<br><br>**A686:**<br>[-0.0565 -0.2229 -0.0400 NaN] | **A241:**<br>[0.8380 0.9984 0.6473 0.9938]<br><br>**A686:**<br>[0.7625 0.3703 0.3482 1] |
| 2-sec. bef. |  |  | **A241:**<br>[0.0595 0 0.1335 0]<br><br>**A686:**<br>[0.0283 0.0696 0.3453 0] | **A241:**<br>[-0.0085 NaN -0.0289 NaN]<br><br>**A686:**<br>[-0.0143 -0.0360 -0.0402 NaN] | **A241:**<br>[0.9580 1 0.8710 1]<br><br>**A686:**<br>[0.9451 0.8705 0.6109 1] |
| 4-sec. bef. |  |  | **A241:**<br>[0.0304 0 0.1766 0]<br><br>**A686:**<br>[0 0 1.4572 0] | **A241:**<br>[-0.0085 NaN -0.0356 NaN]<br><br>**A686:**<br>[NaN NaN 0.0481 NaN] | **A241:**<br>[0.9625 1 0.8439 1]<br><br>**A686:**<br>[1 1 0.5422 1] |

The contrast was one of the properties resulted in more significant differences between the event and previous sequences. These differences were more evident when increasing the offset between the pair of pixels. This might be due to the sizes of the driver's torso in the silhouette. This dispersion might not be large enough in the area from maneuvering or moving the hand.

Respect to the correlation, in driver A241 the value recorded in the first direction (horizontal) was positive during the event and negative in previous sequences. However, this effect wasn't observed in driver A686, since values were quite similar in both negative and positive sequences[12]. In the case of the energy, some values were higher in previous sequences than during the event, depending on the direction of the GLCM. This fact might indicate a higher uniformity in images from negative situations.

---

[12] note that the NaN values obtained in some directions when calculating the correlation mean that the GLCM variance is null. So, the image is completely uniform according to the defined combination of pair of pixels.

## 2.4    Evaluation criteria. Data set definition

The goal of this project is to create an algorithm able to run throughout the triggered events of the database and to save only those in which the drivers react in presence of CRE. Potential methods of recognition of driver's reaction have been commented in previous sections using some positive and negative situations. The performance in the training sample will provide an idea about which combinations are more likely to identify the CREs.

Nevertheless, the evaluation of the proven methods requires the use of a larger data set containing different situations from those used previously. This data set, called *validation data set* from now, comes from a triggering process with kinematic triggers in the euroFOT database and a subsequent evaluation by the annotators. The validation data set contains 120 different situations chosen randomly among the events that have been considered positive or have been rejected by the annotators when watching the videos of candidates to CRE.
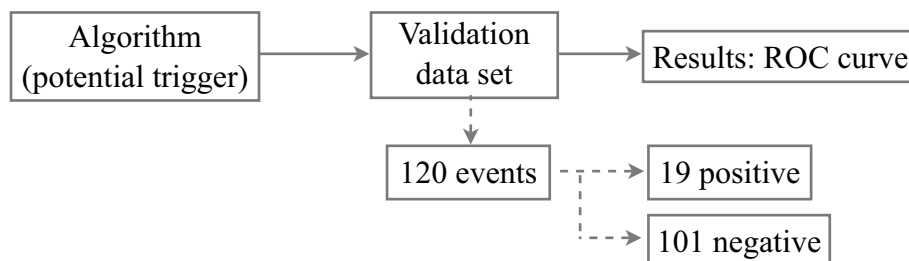


*Figure 2.31    Schema of procedure of algorithms' evaluation.*

Several thresholds have been considered when implementing the algorithms in the baseline. If the threshold is not strict, then it will result in a greater number of true positive events (CRE rightly triggered), but also false negatives (normal driving situations wrongly triggered as positives). The ideal situation would only capture the 19 positive events without any negative (19 true positives and 101 true negatives). Since this only could be possible in a further study with the adequate adjustments from this preliminary project, there should be a compromise between the true positives to be achieved at the expense of false negatives. Such compromise can be represented in terms of *specificity* and *sensitivity* using Receiver Operating Characteristic (ROC) curves.

The ROC curve is a graphical representation of a rate of true positives against the rate of false positives for different thresholds in a diagnostic test (Tape, n.d.). This method was originally developed in the World War II in radar-signal detection (Mason and Graham, 2002). Nowadays is common and widespread used in the medical field for diagnosis of diseases.

There is a trade-off between sensitivity (rate of positives well diagnosed by the test) and specificity (rate of negatives well diagnosed by the test). If the sensitivity increases, then the specificity decreases and vice-versa. In this case, true positives are CRE rightly triggered (Y-axis) and false positives are normal driving situations wrongly triggered (X-axis).

Comparing ROC Curves
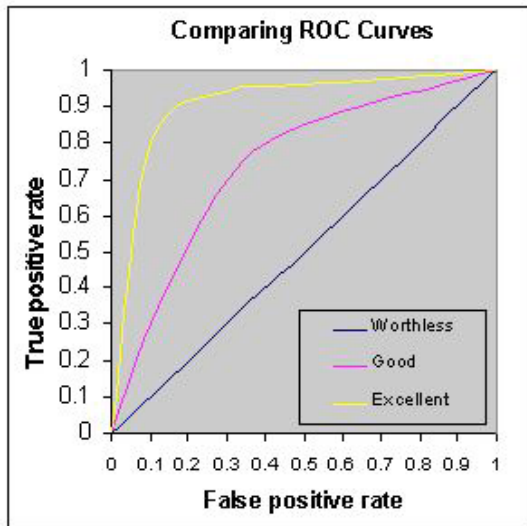
The closer the curve is to the left point, the more accurate the test (see representation in Figure 3.2). On the other hand, the less accurate test, the closer to the diagonal. The area under the ROC curve is commonly used as a measure of accuracy. The following values can be used as guide (Tape, n.d.):

Area Under Curve (AUC)

.90-1 = excellent (A)
.80-.90 = good (B)
.70-.80 = fair (C)
.60-.70 = poor (D)
.50-.60 = fail (F)

*Figure 2.32    Tape T., The Area Under an ROC Curve, [electronic print] Available at <http://gim.unmc.edu/dxtests/ROC3.htm>[Accessed on May 2011].*

In this case, the main limitation of using the AUC when comparing methods is that is more important to save true positives even if this means an increased number of false negatives. Thus, the evaluation of the methods for the whole set of false positives (the area under the entire ROC curve) seems not be the most appropriate in this case. One alternative is to analyze a portion of the ROC curve (Katzman, 1989; Cleveland, 2011).

An estimation of the **relevant portion of the curve** can be defined by a **range of false positives below 60% and a range of true positives above 80%**. The main reason is to keep almost all the true positives (sensitivity) even if it means increasing the false negatives (1-specificity). Therefore, the negative events in the database may be reduced in at least 40%, without losing more than 20% of positive events. The numerical meaning according to the dimensions of the validation data set is to reduce by 40 the total of 101 negative events and keep at least 16 of the 19 positives.

# 3 Results

*In the previous chapter, several algorithms were considered and tuned up in a training sample with eleven drivers and three different situations for each driver with the aim to distinguish CREs from a collection of negative and positive events. This chapter covers the performance of such algorithms in the training sample and their validation within a larger data set. The ideal algorithm should be able to identify as many positive events with the minimal negative situations. Results are presented below making use of ROC curves.*

## 3.1 Performance in the training sample

Along the last chapter, several methods have been applied in the recognition of driver's reaction to identify safety critical situations. Initial assumptions, as classification based on t-test results, seem to generate noisy images and an unclear definition of the state of the driver's motion. Nevertheless, analyzing changes in pixel intensities over time suggests that the **sought motion may be related to a sudden change in a group of pixels intensities.**

This same concept is behind the images of STD of jerk and the OF calculations. By looking at the grayscale images from STD of jerk is possible to identify which is the positive event without any additional information in most of the cases. The key of this identification is the silhouette of the driver, which means that there are a group of pixels that share a wide variance of jerk distribution over time. In the case of the OF, peaks in jerk distribution from the average of OF velocities in each frame contribute to discriminate between previous sequences and the event. The calculation is based on average speeds, so a group of pixels change quickly between frames.

To assess the validity of this theory, these calculations must be performed throughout the entire training sample. The following is the example of results obtained for one of the drivers. Jerk distributions are presented together with the images from STD of jerk, which are also evaluated as distributions of sum of values along rows. The Appendix 4 covers the same calculations for all the drivers of the training sample.

It is expected that a driver's silhouette appears during the event when plotting the STD of jerk. Distribution of sum of values in rows for each column pursues to distinct between normal driving maneuvers (just certain white areas in the images) and reactions in CREs. Thus, normal maneuvering may be related to local peaks in these curves, while a higher mean may be related with the positive events. This is because white areas are dispersed along the image to reproduce driver's silhouette.
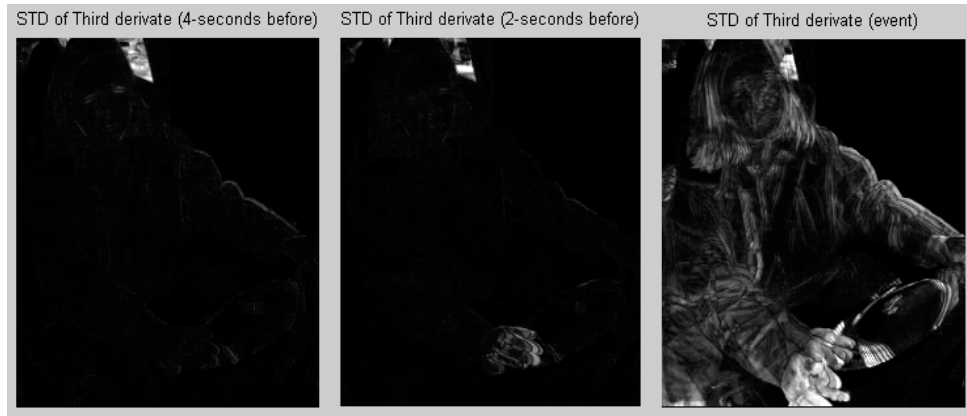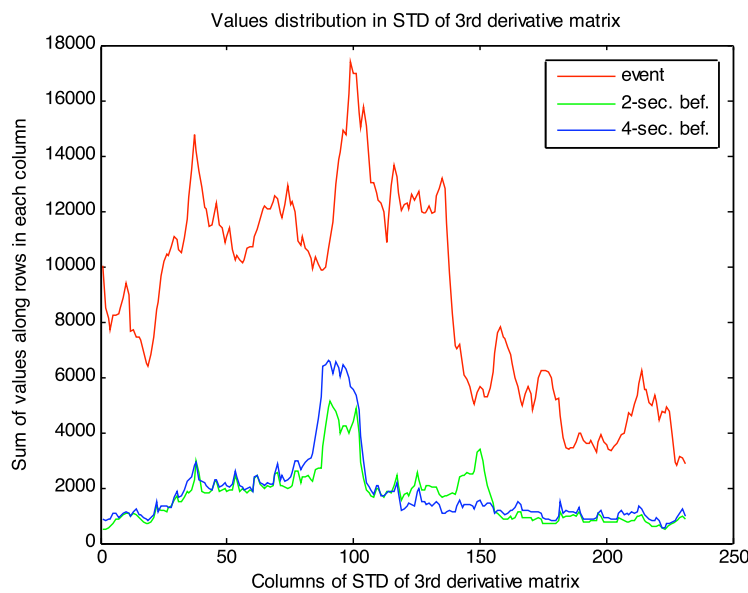
*Figure 3.1    STD of jerk: 4-sec. before the event, 2-sec. bef.  and event (driver A484).*



The driver remains in the same position in sequences before the event. Therefore, the images of STD of jerk seem to be a clear indicator of when the driver reacts. As can be seen in Figure 3.2, the sum of values along the rows for each column is also significantly higher during the event than in previous sequences.

*Figure 3.2    Distribution of STD of jerk values along rows and columns.*
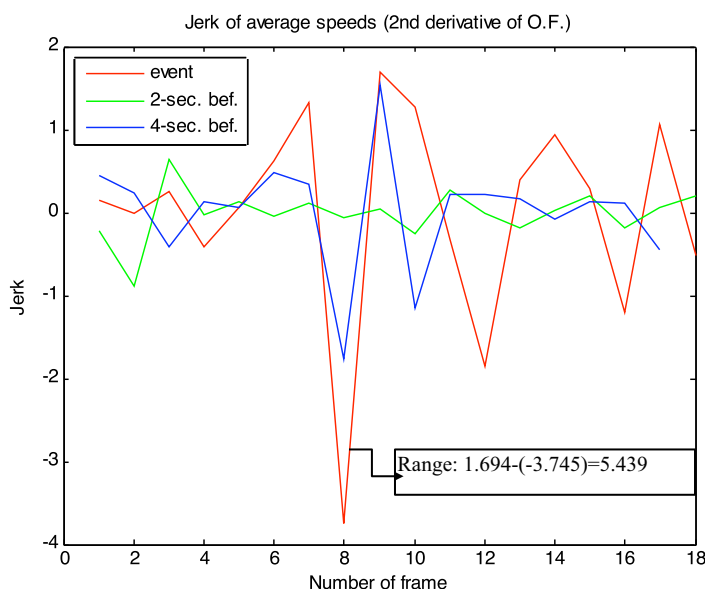


Since the driver remains in the same position over time, distributions of jerk should be relatively constant before the event. Some unexpected results were obtained four seconds before the event at $8^{th}$ iteration, as shown in Figure 3.3. Anyway, the maximum jerk is reached during the event.

*Figure 3.3    Distribution of jerk from OF velocities.*

### 3.1.1 Optical Flow

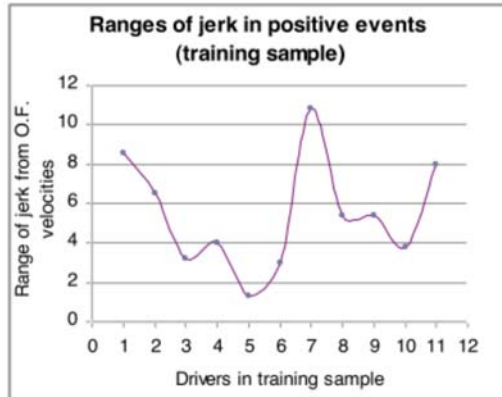Findings in the training sample support that **peaks in distribution of jerk from OF velocities are related to the drivers' reaction in presence of a CRE**. This hypothesis is accomplished in ten of the eleven drivers of the sample, in which the ranges of jerk are significantly higher during the event than those obtained in previous sequences.



Uncertainties are which **range of jerk** is related to the driver's reaction, since these values are **different for each driver** (see Figure 3.4) and the computational cost of running the optical flow code. Given the dimensions of the database, the computational time is an important limitation.

*Figure 3.4      Ranges of jerk for drivers of the training sample.*

To sum up, jerk peaks from OF velocities and images of STD of jerk were identify as potential indicators of positive events. Both methods base the discrimination in the presence of the drivers' reaction when CREs occur. The **main limitation** when calculating the OF is the **consumption of computational time**. On the other hand, it was observed a relationship between the driver's reaction and images of STD of jerk. In most of the cases was possible to identify what the positive event is by just looking at the driver's silhouette. As this is graphic information, several *converters* have been addressed in the last chapter to transform this information into numerical.

### 3.1.2 Mean criterion

The figure below includes the distribution of mean values for each sequence in all the drivers of the training sample:
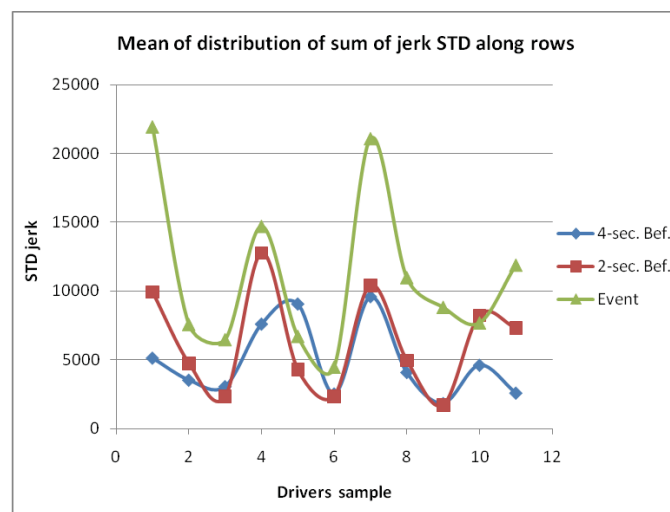


*Figure 3.5      Mean of distribution of sum of STD of jerk values along rows for all the drivers of the training sample*

As shown in Figure 3.5, the mean values are higher during the event than in previous sequences in nine of the eleven drivers of the sample. In both exceptions ( 5th and 10th position in the sample), maneuvering in sequences before the event generates peaks in the distribution of STD of jerk values and, consequently, the mean value increases. Images of STD of jerk of both cases are presented below to analyze why the mean value differs from those obtained in the rest of the sample.
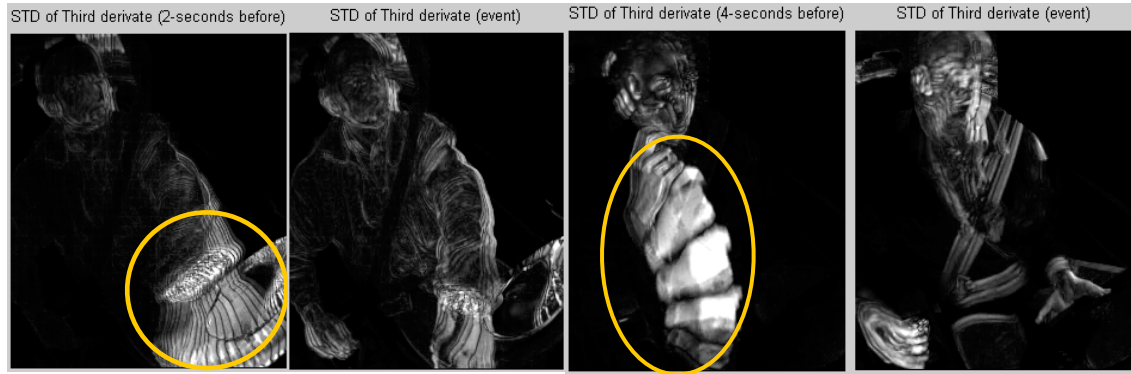


*Figure 3.6      STD of jerk in negative sequences of drivers A567 and A686.*

Marked areas of moving the hand and turning the steering wheel stay pixels with wide variance in jerk values over time. This causes the mean increases in such situations in comparison with the figure obtained from the positive event. Thus, the mean criterion not seems consistent enough in itself to discriminate between positive and negative situations.

Since the values were added only along rows, distributions can be also tested in another direction. This involves calculating the mean of the distribution of sum of STD of jerk values along columns instead of rows. Distributions in both directions using one of the drivers from the exceptions are plotted in the figures below together with the number of zeros (black color) in the images:



*Figure 3.7      Distribution of sum of STD of jerk values along rows (driver A567).*

*Figure 3.8     Distribution of sum of STD of jerk values along columns (driver A567).*

As can be seen in Figures 3.7 and 3.8, peaks of sum of STD values are higher in the previous sequence than during the event along rows and columns. Then, the white area when turning the steering wheel still generates a higher mean in this case. Besides, the number of zeros is very similar in all the sequences.
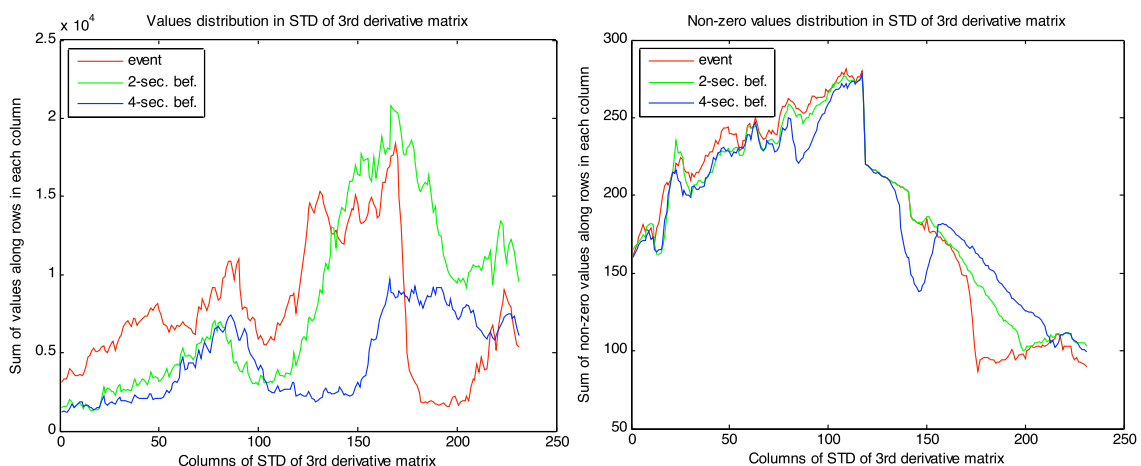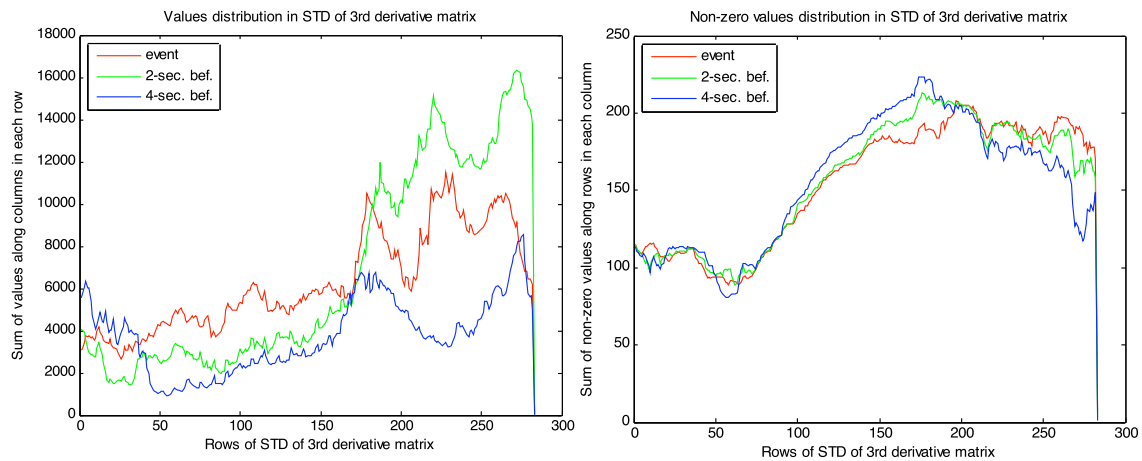
If comparing this result with those obtained for the rest of the sample, it is observed that in both exceptions (drivers A567 and A686) the mean is also higher in other sequences than during the event (see Table 3.1). Regarding the number of zeros, this value is not significant enough to distinct between positive and negative sequences.

*Table 3.1     Mean of distributions of sum of STD over columns and Number of non-zero values in the training sample.*

| | Mean of STD distribution over columns | | | Number of non-zero values | | |
|---|---|---|---|---|---|---|
| Driver | 4-sec. Bef. | 2-sec. Bef. | Event | 4-sec. Bef. | 2-sec. Bef. | Event |
| A34 | 4190 | 8092 | 17920 | 144,2 | 148,4 | 172,5 |
| A241 | 2885 | 3869 | 6198 | 126,8 | 130,5 | 152,7 |
| A481 | 2505 | 1892 | 5295 | 183,7 | 181,9 | 191 |
| A501 | 6204 | 10410 | 12010 | 162,9 | 178,7 | 181,9 |
| A686 | 7393 | 3485 | 5489 | 128,7 | 124,5 | 130,1 |
| A1064 | 2071 | 1916 | 3656 | 122,8 | 124,6 | 132,9 |
| A131 | 9519 | 10330 | 20950 | 168,4 | 163,6 | 200,2 |
| A352 | 3342 | 4045 | 8982 | 205,2 | 206,9 | 217,9 |
| A484 | 1505 | 1381 | 7214 | 158,6 | 156 | 172 |
| A567 | 3767 | 6708 | 6287 | 154,8 | 156,2 | 154,7 |
| A936 | 2121 | 5958 | 9711 | 98,91 | 108,2 | 142,9 |

In conclusion, the **mean values of distributions of sum of STD of jerk along rows and columns** have been calculated to recognize the driver's silhouette as a wider dispersion of intensities over the images. Since this value is **affected by concentrated areas in the image from maneuvering and changes in position**, other statistical measures are taken into account.

### 3.1.3 Harmonic mean

As was performed with the mean in the last section, the harmonic mean is estimated in the distributions of sum of STD of jerk values along rows and columns. Again, the reason is to try to locate the silhouette by the dispersion of pixels in the image with the difference that harmonic mean is not as affected by outliers. The sum of each par of values (harmonic means in rows and columns) is presented below as "Combination of harmonic means" for all the drivers of the training sample:
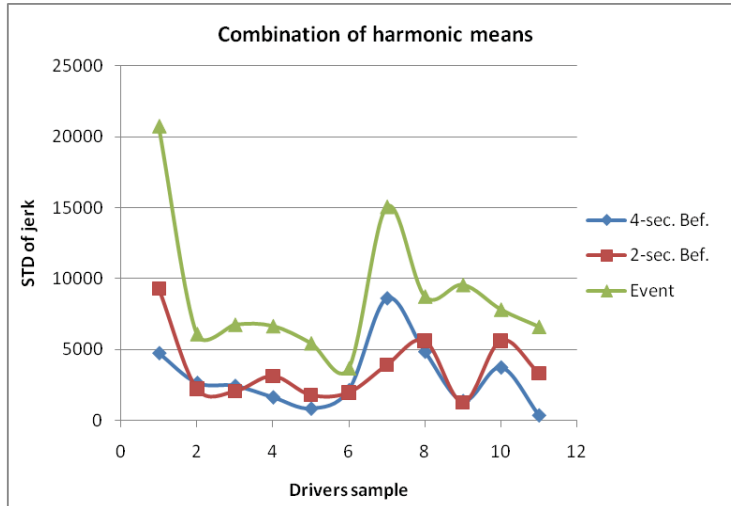


As shown in Figure 3.9, combinations of **harmonic means reach higher values during the event than in previous sequences for all the drivers of the training sample.** This result in also observed in drivers in 5th and 10th position in the sample, exceptions of the mean criterion,who register in this case a higher sum of harmonic means during the event.

*Figure 3.9    Distribution of combination of harmonic means in the training sample.*

Looking at the range of values of harmonic means in different drivers (see in Figure 3.9), the main issue is to establish a threshold able to identify as many events at the expense of negative situations.

### 3.1.4 Mean&General mask

The mean value was also calculated considering the Fusion mask, binary mask created from the combination of several driver's silhouettes.
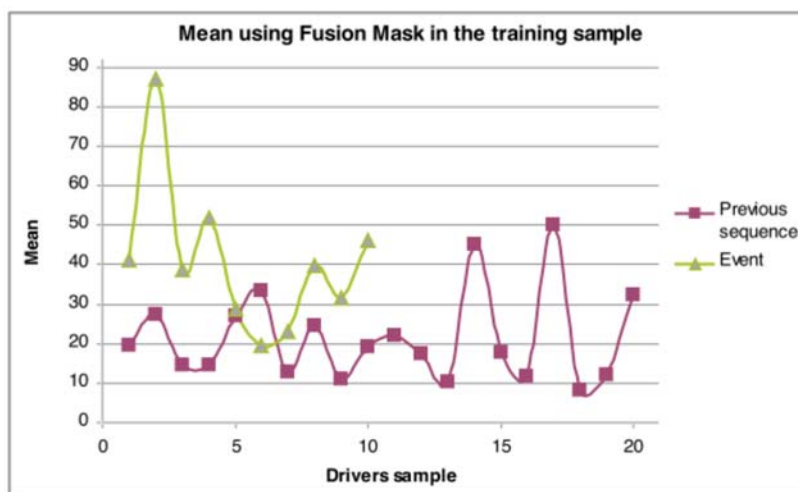


The Figure 3.10 shows the mean values using the binary Fusion mask in images of STD of jerk. Despite the fact that the size of the training sample is not large enough to have a statistical sense, these results suggest higher means in those images from events than from previous sequences.

*Figure 3.10    Mean values by applying Fusion mask in STD of jerk images.*

### 3.1.5   GLCM properties

The contrast and the energy of the GLCM have been tested as well in the entire training sample considering an offset of 200:
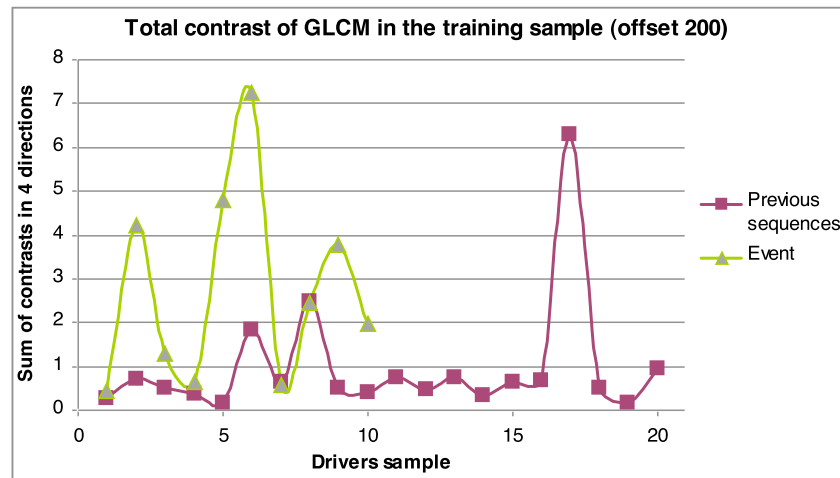


*Figure 3.11     Sum of contrast in four directions of the GLCM in the training sample.*
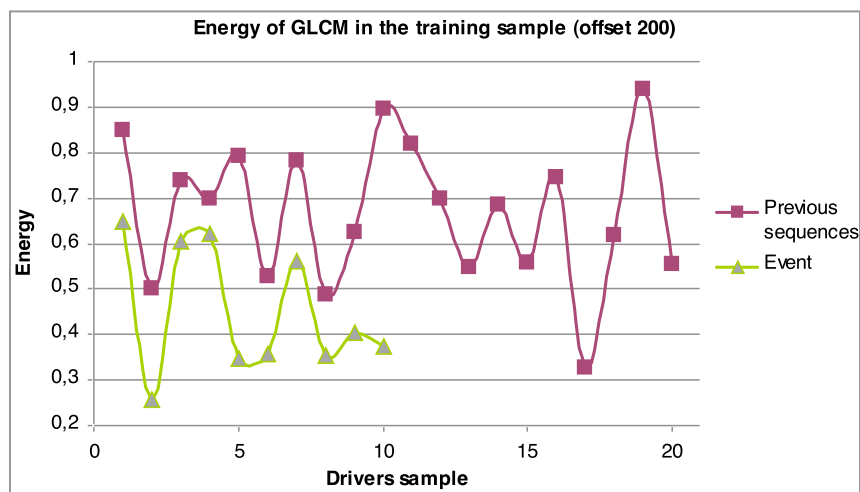


*Figure 3.12     Energy of the GLCM in the training sample*

The Figure 3.11 shows the sum of contrasts in four different directions of the GLCM with an offset of 200 pixels for the sequences of the training sample. This value appears to be greater in some of the positive events regarding the previous sequences, but is not a clear discriminator in some cases. The same occurs using energy as property of study in GLCM. It seems that values are generally lower during the events in comparison to previous sequences. **The main limitation would be to set a value that discriminates between both situations.**

## 3.2 Results in the validation data set

As mentioned previously, the training sample is used to test different methods and identify potential features to discriminate between positive a negative events. However, its dimensions are not large enough for a statistical sense. Therefore, the evaluation requires the use of the validation data set.

Below, ROC curves are plotted for each method with different combinations of masks in the images and thresholds. The range of variation of threshold values has been chosen according to the results of the training sample. These are represented as dots on the graph for the entire false positive rate. This gives an idea of the accuracy of the curve. However, only a certain area under a portion of the curve is relevant. It is bounded by two lines on the graphs. The largest area within these boundaries determines which method is the most accurate based on the requirements specified in the Evaluation criteria in Chapter 2.

Another consideration when comparing the methods arises in the computational time. This is estimated in terms of how long (in seconds) processing each second of trip takes. This is calculated by taking the time of computing all the iterations when changing the threshold values and considering the two-second duration of each file in the baseline.

### 3.2.1 Mean criterion

The mean criterion evaluates the presence of driver's silhouette in images of STD of jerk by adding STD values along rows and columns. Both vectors containing partial sums are combined into a single. The mean is calculated in its distribution.

60 iterations have been considered by changing the threshold values with a step of one unit. Three different input images have been considered:

> -without mask: original image crop around the torso.
>
> -BW mask: binary mask hiding the window.
>
> -Fusion mask: binary mask around the area in which driver's silhouettes commonly take place.

In comparison with the commented evaluation criteria, the curves in Figure 3.13 are closer to the shape of good accuracy. However, better results would obtain if the curve was closer to the upper left corner. The area under the bounded portion of the curve seems larger without using any mask. The second best option according to this area suggests the use of fusion mask.
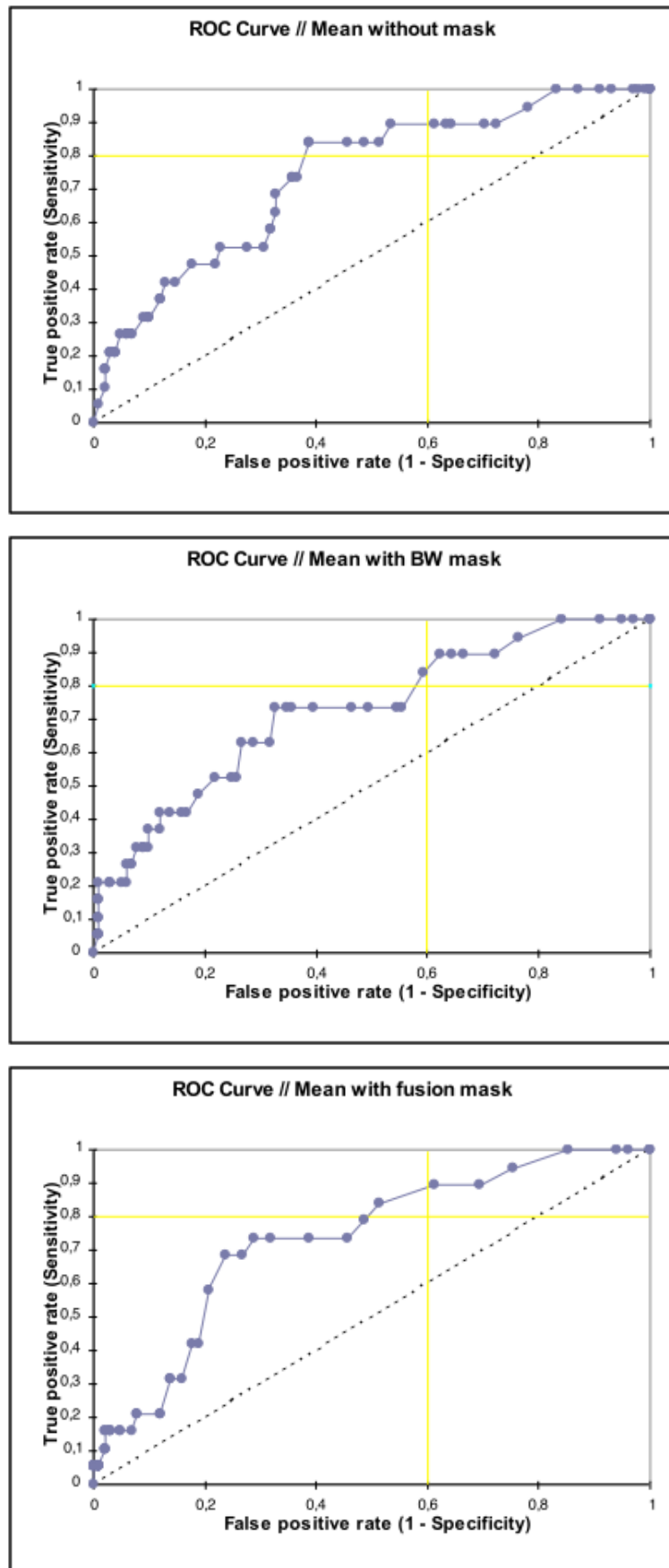
*Figure 3.13    ROC curves of thresholding with different combinations of mean criterion.*

### 3.2.2 Harmonic mean



The distribution of partial sums of STD of jerk along rows and columns is now evaluated using the harmonic mean. Unlike the previous test, this method discriminates the outliers of such distribution, mostly regarding with maneuvers.

The threshold values vary between harmonic means of 400 and 10000, resulting in a stepped ROC curve.
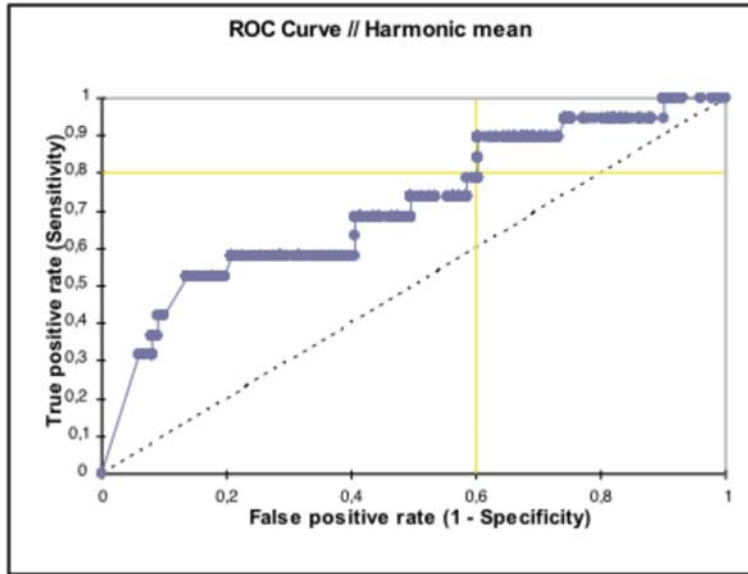
*Figure 3.14    ROC curves of thresholding with harmonic mean criteria.*

The harmonic mean criterion emerged as an alternative of using the mean. Among the 33 cases of the training sample, the harmonic mean was higher in the 11 that were positive. However, as can be seen in Figure 3.14, the bounded area is null. The slope at the beginning of the ROC curve is positive in terms of a further increase in sensitivity against (1-specificity). Nevertheless, variations from a certain threshold values don't seem to affect the rate of true and false triggered. The best result obtained with this method according to the initial criterion is achieved with harmonic means above 2635. In that case, the false positive rate is 60,4%, while the true positive is 89,4%.

### 3.2.3 Ranges of jerk from OF



The calculation of the optical flow (OF) is a numerical alternative to the use of STD images in the estimation of rates of change in pixel intensities.

Events in the baseline have been triggered above three different intermediate values of range of jerk (1, 2 and 6). This low sampling rate is mainly due to the computational cost of implementing the OF

*Figure 3.15    ROC curves of thresholding with OF Criteria.*

**ROC Curve // O.F. BW mask**

OF velocities have been calculated using the original images and them combinations with binary masks. For the three cases observed in Figure 3.15, ROC curves are closer to the diagonal. Although the results are slightly better with the use of fusion mask, the method seems inaccurate for identifying positive events.

**ROC Curve // O.F. fusion mask**

This contrasts with the results of evaluating the OF in the training sample. In ten of the eleven drivers, the peak in the jerk distribution in positive situations was clearly significant in comparison with those obtained in negative sequences. The main limitation arise in the threshold value, since it changes for each driver.

*Figure 3.15    ROC curves of thresholding with OF Criteria.*

In any case, the potential application of this method together with the use of fusion mask reduce by 40% the number of negatives while triggering 17 of the 19 positive events (see table 3.1).

*Table 3.2    Results of triggering with ranges of Jerk of OF speeds above one in combination with fusion mask over the original images.*

| Range>1 | | |
|---|---|---|
| | **TRUE** | **FALSE** |
| **Positive** | 17 | 2 |
| **Negative** | 38 | 63 |

## 3.2.4    GLCM properties



A statistical approach in the identification of the drivers' silhouette in images of STD can be done according to the spatial distribution of pair of pixels over the images. This involves estimating how the properties of GLCM change in positive and negative events.

As can be seen in Figure 3.6, the trend at the beginning of the ROC curve when thresholding with Energy values is better than the obtained using the Contrast. This is due to an increase in sensitivity at a low rate of (1-specificity). However, the bounded area seems to indicate a better diagnostic using Contrast as threshold in order to trigger almost all the positives of the baseline.
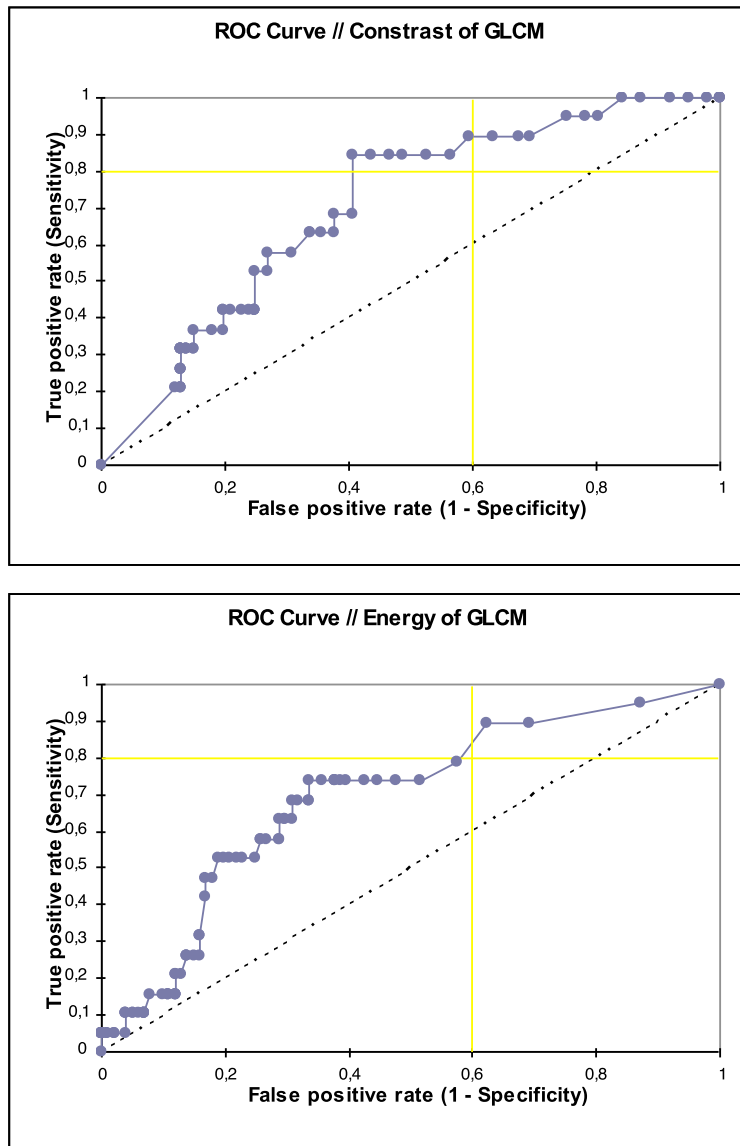
*Figure 3.16    ROC curves of thresholding with properties of GLCM.*

The iterations have been done with a GLCM Contrast between 0 and 4 (step of 0,05) and GLCM Energy between 0 and 1 (step of 0,02). The best result is obtained when thresholding with Contrast above one. In that case, the rate of negatives is reduced in a 60%, while keeping the 84% of the positives.

### 3.2.5 Analysis of false negatives and positives

The comparison between bounded areas suggests that triggering with GLCM Contrast and with Mean Criterion without mask are the most accurate methods when keeping above the 80% of the positive events and losing at least 40% of the negatives. Between both cases, the bounded area is slightly higher in the case of the Mean criterion. Means above a threshold of 24 reaches a sensitivity of 84,2% together with a specificity of 61,4%. This result extrapolated to the baseline's dimensions means that 55 events are triggered among 120, containing 16 of the 19 positives.

The analysis of the false positives and negatives from this procedure might provide a better understanding of the failures of the algorithm. This information is essential in order to improve the algorithm and use images as triggers in future research.

*Table 3.2      Descriptions of false positives from Mean Criterion.*

| False positives | Description |
|---|---|
|  | The driver was overtaking when another car appeared in the blind spot. He reacted by turning the steering wheel. Although the motion is not so evident, the driver's silhouette appears in the image from STD of jerk. Note that these trials were doing based on Mean criterion without mask, therefore the area in the window is not hidden by a binary mask. |
|  | The driver was looking to the right. When he looked back on the road, a car has appeared in front of him. The driver reacted by braking. There wasn't an evident motion in driver's torso. Anyway, a driver's silhouette appeared in the STD of jerk image (although the mean values is not enough higher to be triggered with the proposed threshold). |
|  | In this situation driver was overtaking when another car tried to change to its lane. The driver braked and pressed the horn. Its seems like the driver is aware during the whole sequence that the car can take its lane, so the situation could be unexpected for him. |

Although the mean values wasn't enough higher to be triggered, the silhouette of the driver appears in the false positives, as can be seen in Table 3.2. Among the false negatives, three main categories of failures have been observed:

**■ Rough pavement / bumps on the road:**
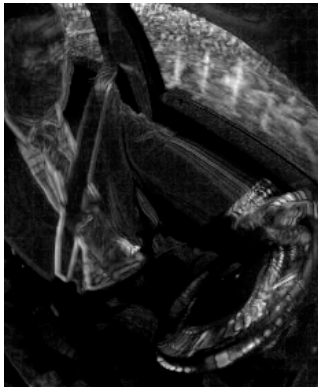


Due to the state of the pavement or the presence of bumps on the road, the motion in the driver is closer to jump in the seat. This situations are mainly related with peaks in longitudinal acceleration.

*Figure 3.17    Example of false negative due to a bump on the road.*

**■ Closed curves:**
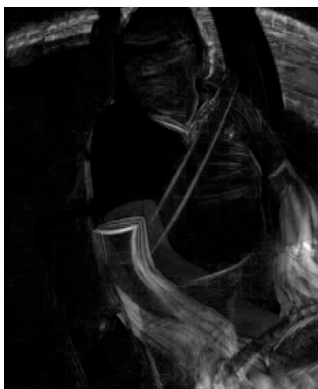


The driver is taking a closed curve. Due to kinematic forces, the motion generates a silhouette in the plot of STD of jerk. Looking at the image, the shoulders don't generate a white shadow. This detail could be study as possible discriminator factor in future research. A STD of jerk image for a CRE by the same driver is plotted below, together with two other normal driving situations that take place two and four seconds before the event. Comparing both triggered situations (the negative from the triggering procedure with Mean criterion and the one from the real CRE), it´s possible to appreciate a different motion in the driver's torso.

*Figure 3.18    Example of false negative due to kinematic forces in a closed curve.*
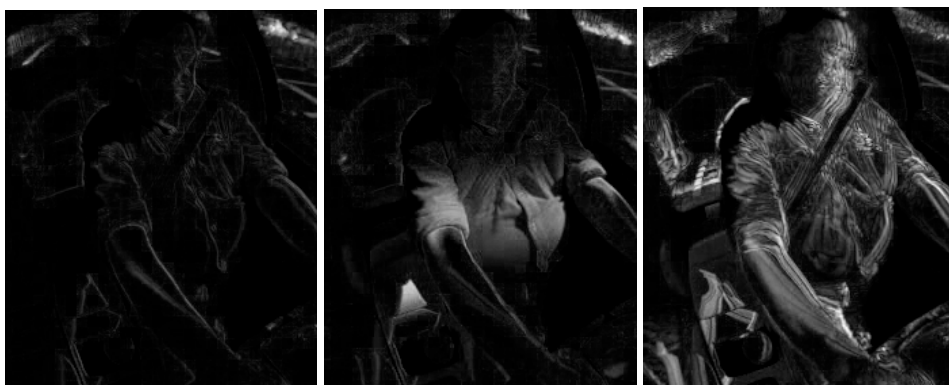


*Figure 3.19    STD of jerk: 4-sec. before the event, 2-sec. bef. and CRE.*

■ **Occlusion in the camera / Irregular illumination:**

The irregular illumination originates continuous shadows in the frames collection. Pre-filtering just removes those frames that are supposed to be flashed by the devices of the instrumented car.



*Figure 3.20    Example of false negatives due to camera occlusions and irregular illumination.*

Throughout this project, the driver's motion has been estimated from changes in pixel intensity. In the case of closed curves and bumps on the road, false negatives come from a motion in the driver other than the reaction when a CRE. This suggests a more clear definition of the specific motion in the driver's reaction. However, the driver's motion can't be estimated by variations in pixel intensity when the illumination is irregular. In such situations, these effects should be corrected before performing the analysis.

### 3.2.6    Mean criterion in motion's detection

Assuming that, once the illumination is corrected, false negatives will arise due to the detection of a motion different from the reaction, results can be recalculated without considering these situations in the validation data set. This would provide a true interpretation of the algorithm failure detection in motion.

The mean criterion has been applied in the validation data set without sequences with irregular illumination. This represents a total of 93 situations containing 19 positives. Results are presented below (see in Figure 3.21).
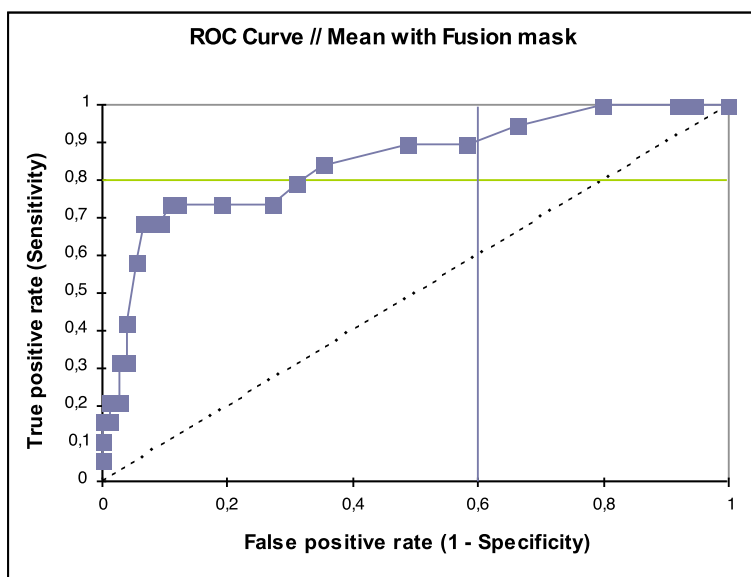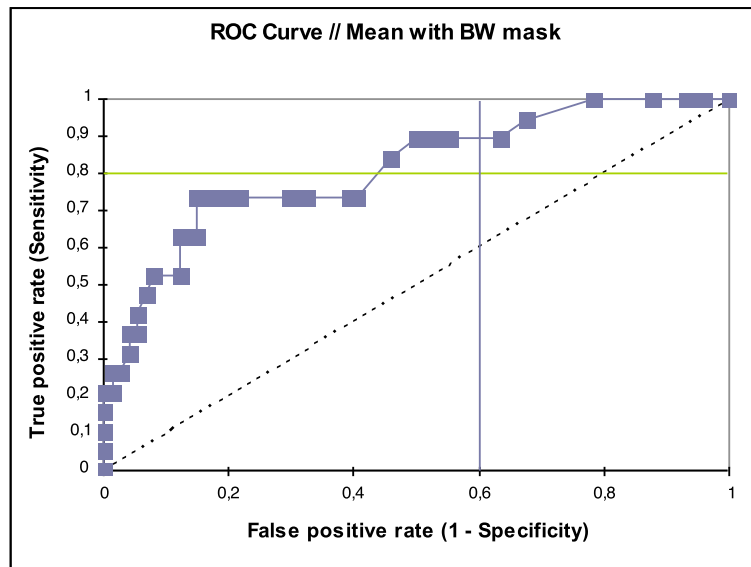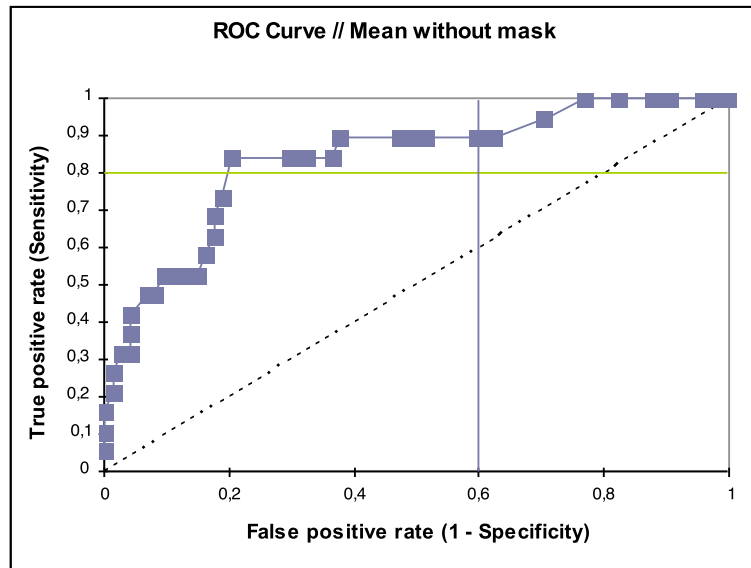
*Figure 3.21 ROC curves of thresholding with different combinations of mean criterion.*
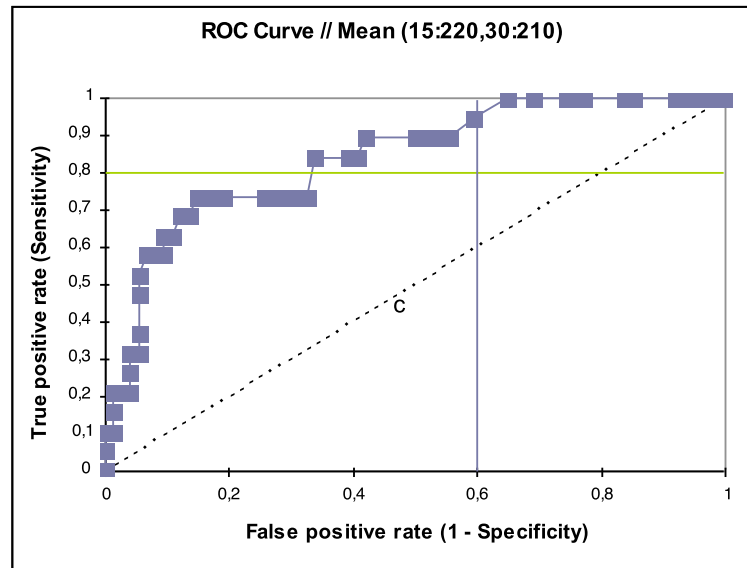
*Figure 3.22    ROC curve from applying the mean criterion without mask in a central part of the image (rows: 15 to 220; columns: 30 to 210).*

### 3.2.7    Comparison

The Table 3.3 contains a compilation of results for each algorithm in the original validation data set (19 positive and 101 negative events). The information about each algorithm is given by:

- ⁃ The range of thresholds used along the iterations.
- ⁃ The highest sensitivity and specificity achieved based on the minimum requirements (portion of the ROC curve)[13] and the corresponding number of true positives and false negatives.
- ⁃ The entire area under the ROC curve as general measure of accuracy (AUC).
- ⁃ The computational time requires for each algorithm in processing the whole baseline with a specific threshold.
- ⁃ The ratio between trip/computational time: taking into account the number of files in the validation data set and a duration of two seconds each one, it's possible to calculate how long processing each second of trip takes. Results of these estimations are listed in *Ratio between trip/computational time*. This value is dimensionless, since it's a ratio of two quantities in seconds. The lower the ratio is, the higher the computational time is saved.

As can be seen in Table 3.3, the best results were achieved using the mean and the GLCM contrast criteria. According to the accuracy levels based on the AUC[14], these are in the range of *fair accuracy* level of recognition (AUC between 0,7 and 0,8). In both cases is reached a sensitivity of 84,21% together with a specificity of 61,39% and 59,41%, respectively. The numerical meaning according to the data set dimensions is to keep 16 of the 19 positives with 39 and 41 of the negatives in each case. Although the ideal is to achieve a higher specificity, these results suggest a reduction of the 60% in

---

[13] defined by a range of false positives below 60% and a range of true positives above 80%.

[14] see Evaluation criteria in Chapter 2.

the quantity of negatives. Thus, the dimension of the list of candidates to CRE is reduced from 120 to 55 situations.

Given that the OF algorithm requires several iterations in its pyramidal process, this affects the computational time. The consumption is significantly higher in comparison with the rest of the methods. It requires processing during 200 seconds each second of recorded trip[15]. Nevertheless, the low accuracy achieved with the OF in unexpected. In ten of the eleven drivers of the training sample the peaks in the jerk distribution of average of OF speeds were significantly higher during the event than in previous sequences. However, the ranges were different in each case, hence the limitation of establishing a threshold.

*Table 3.3        Compilation of results in the original validation data set (120 situations).*

| Algorithm | 19 positives / 101 negatives | | | | | | | Computa-tional time (by one iteration) (sec) | Ratio trip/ compu-tational time |
|---|---|---|---|---|---|---|---|---|---|
| | Thresholds | TP | Sensitivity | FN | Specificity | Thres. | AUC | | |
| **Mean (without mask)** | 1 to 80 | 16 | 0,8421 | 39 | 0,6139 | 24 | 0,7387 | 19,16 | 0,08 |
| **Mean (BW mask)** | 1 to 60 | 16 | 0,8421 | 60 | 0,4059 | 13 | 0,7241 | 17,36 | 0,07 |
| **Mean (fusion mask)** | 1 to 60 | 16 | 0,8421 | 52 | 0,4851 | 9 | 0,7327 | 22,16 | 0,09 |
| **Harmonic mean** | 400 to 10000 | 16 | 0,8421 | 61 | 0,3960 | 2745 | 0,6899 | 27,02 | 0,11 |
| **Optical flow (fusion mask)** | 0 to 4 | 17 | 0,8947 | 63 | 0,3762 | 1 | 0,6087 | 48000 | 200 |
| **Optical flow (BW mask)** | 0 to 4 | 17 | 0,8947 | 73 | 0,2772 | 1 | 0,5862 | 48000 | 200 |
| **Optical flow (without mask)** | 0 to 4 | 19 | 1,0000 | 94 | 0,0693 | 1 | 0,5735 | 48000 | 200 |
| **GLCM energy (fusion mask)** | 0 to 1 | 17 | 0,8947 | 63 | 0,3762 | 0,02 | 0,6959 | 21,00 | 0,09 |
| **GLCM contrast (fusion mask)** | 0 to 4 | 16 | 0,8421 | 41 | 0,5941 | 1 | 0,7030 | 21,63 | 0,09 |

The study of the false negatives from this first trial showed that some sequences of the validation data set were affected by an irregular illumination. Thus, in those cases the driver's motion can't be rightly estimated without a previous preprocessing (changes in pixel intensity are due to variations in the illumination instead of motion in the images). Therefore, those sequences were removed to have a true interpretation of the algorithms' performance in driver's reaction detection among other kind of motions. The Table 3.4 contains the results of applying some of the algorithms to a validation data set with 19 positives and 74 negatives situations.

---

[15] note that a multi-core computer is used in processing all the algorithms.

*Table 3.4     Compilation of results removing irregular illumination and camera occlusions from the validation data set (93 situations).*

| Algorithm | 19 positives / 74 negatives | | | | | | |
|---|---|---|---|---|---|---|---|
| | Thresholds | TP | Sensitivity | FN | Specificity | Thres. | AUC |
| Mean (without mask) | 1 to 80 | 16 | **0,8421** | 15 | **0,7973** | 24 | 0,8364 |
| Mean (BW mask) | 1 to 60 | 16 | 0,8421 | 34 | 0,5405 | 13 | 0,8133 |
| Mean (fusion mask) | 1 to 60 | 16 | 0,8421 | 26 | 0,6486 | 9 | **0,8471** |
| Mean (80:170,80:140) | 400 to 10000 | 17 | 0,8947 | 35 | 0,5270 | 24 | 0,7809 |
| Mean (15:220,30:210) | 1 to 60 | 16 | 0,8421 | 32 | 0,5676 | 18 | 0,7760 |
| Harmonic mean | 1 to 60 | 16 | 0,8421 | 25 | 0,6622 | 26 | 0,8417 |
| GLCM contrast (fusion mask) | 0 to 4 | 16 | **0,8421** | 18 | **0,7568** | 1,1 | 0,8350 |

The best level of accuracy based on the area under the entire ROC curve is achieved using the mean criterion together with the fusion mask (binary mask with general driver's silhouette). In that case, the AUC is the 0,8471, which means a *good accuracy* level of recognition (areas between 0,8 and 0,9). Nevertheless, according to the minimum requirements, the use of mean without mask achieves a sensitivity of 84,21% together with a specificity of 79,73%. By removing the irregular illuminated sequences, the specificity has increased in a 18% from the result obtained in the first trial. The GLCM contrast also achieved the same sensitivity with a 75,68% of specificity.

# 4    Discussion & Conclusions

*Throughout the past chapters, several algorithms to recognize the drivers' reaction in a CRE have been described and tuned up using a training sample. Then, such algorithms have been evaluated in a larger sample containing 120 events and adapted to achieve a maximal increase in sensitivity and specificity. In this chapter, conclusions from this evaluation, reasons to rely on driver's reaction recognition, and possibilities for future research are addressed.*

## 4.1    Where did the idea of recognizing driver's reaction come from? Triggering in euroFOT based on the 100-Car study algorithms

The idea of developing a trigger based on driver's reaction recognition comes from a conventional process of data reduction carried out using the triggers from the 100-Car Naturalistic Driving study (Dingus *et* al., 2006) in an initial collection of euroFOT data[16]. Reasons were to address the problem of triggering the data and to try to locate CRE based on the experience of a public large-scale project as the 100-Car Naturalistic Driving study.

The following explains the technical bases of this initial data reduction and the conclusions when triggering with kinematic values of cars. Note that creating an algorithm able to identify the driver's reaction and use it as potential trigger wasn't the first idea at the beginning of this project. This idea came after watching more than 400 videos containing not even two dozen positive situations.

Rydström *et* al. (2009) compared the most common kinematic triggers and associated thresholds implemented in driving studies. They refers to longitudinal acceleration, lateral acceleration and TTC[17]. This comparative doesn't show significant differences between NDS with the same type of vehicle (if conducting a NDS with a fleet of trucks the thresholds would be lower than in cars).

Table 4.1 shows the performance of the triggers applied in 100-Car study. The most common detected cases were rear-end events, so that longitudinal acceleration and forward TTC reach more positive events. It's worth noting the high percentage of negative events obtained in all categories. Therefore, a multivariate statistical analysis[18] was carried out after triggering the data to improve the identification performance (the scope of this procedure was limited to only lead/following vehicle conflicts).

---

[16] at the time of this first triggered process (January 2011), only part of the data is already available (note that data collection lasts until December 2010).

[17] time remaining before a collision that can be used to indicate the severity of the incident (the lower TTC, the higher risk) (Sidaway *et* al., 1996).

[18] building a complex classification matrices by mathematically synthesizing the data (Dingus *et* al.2006).

*Table 4.1      Results of triggering in 100-Car Naturalistic Driving study (Dingus et al., 2006).*

| Trigger type | %Positives | %Negatives |
|---|---|---|
| Lateral acceleration | 3,5 | 91,3 |
| Longitudinal acceleration | 44,7 | 66,4 |
| Event button | 8,4 | 69,9 |
| Forward TTC | 56,5 | 86,4 |
| Rear TTC | 4,6 | 59,9 |
| Yaw rate | 21,7 | 91,1 |
| Lane tracker | 0,6 | 96,1 |
| Side conflicts | 3,1 | 96,5 |

Although results from other studies do not seem promising, an initial collection of data from euroFOT was triggered taking the triggers and associated thresholds used in 100-Car Naturalistic Driving study as reference. Thus, it allowed a better understanding of the problem and the situations that were triggered. Additionally, situations in which Lane Departure Warning system[19] alerts drivers were also analyzed. The following table gives a summary of trigger definitions and results obtained[20].

As can be seen in Table 4.2, percentages of valid events are quite low compared to the total number of cases triggered. Longitudinal acceleration and LDW registered best performances by detecting, approximately, a 30% of positive events. Other potential kinematic triggers observed are the *brake pressure* signal combined with the car speed and the longitudinal acceleration. In this context, several events were detected in highways with velocities above 100 km/h and around 30 bar of brake pressure. Such events weren't triggered since longitudinal decelerations were around -3 m/s² (below the 6 m/s² used as threshold[21]). Note that triggering were carried out when the transference of data was starting, so some videos were missing at this time. Also, the size of the data is not enough to generalize these results.

Nevertheless, more than 400 videos from the triggered events were visualized and those situations were described in a further report. Thus, results seemed enough consistent to have an idea about the complications when triggering with kinematics and the typical wrongly triggered situations.

---

[19] intelligent in-vehicle system that alerts the driver in case the car is unintentionally leaving the own lane.

[20] see report at Appendix 1.

[21] a threshold of 6 m/s² is quite strict if comparing with the used in other naturalistic driving studies. A typical value is around 4 m/s², but also it is likely to trigger more negative events.

*Table 4.2*    *Summary of results of triggering initial euroFOT data based on 100-Car.*

| Trigger definition | Total cases triggered | % Positives | %Video not available |
|---|---|---|---|
| Lat. Accel. $\geq$\|6,867m/s$^2$\| | 34 | 0 | 1 (2,9%) |
| Long. Accel. $\geq$\|5,886m/s$^2$\| | 119 | 40(33,61%) | 14 (11,77%) |
| Long. Accel. $\geq$\|4,905m/s$^2$\| TTC $\leq$ 4s | 12 | 3 (25%) | 3 (25%) |
| \|4,905m/s$^2$\| $\leq$ lat. accel. $\leq$\| 3,924m/s$^2$\| TTC $\leq$ 4s At min. TTC, FW$\leq$30,48m | 14 | 3 (21,43%) | 3 (21,43%) |
| Min. Variation: -4 to 4 degrees/s (or vice.) (3s window) Neutral position at the beginning and end of the range | 7 | 1(14,28%) | 4 (57,14%) |
| BLISS signal Turn indicator=On Speed $\geq$ 24,12 km/h (+/-1s window) | 153 (turn to left) only 49 checked<br><br>96 (turn to right) only 15 checked | 4 (8,16%)<br><br>2 (13,33%) | 0<br><br>0 |
| Min. Variation: -2 to 2 degrees/s (or vice.) (3s window) Neutral position at the beginning and end of the range -Speed $\geq$ 24,12km/h | 17 | 1(5,88%) | 1 (5,88%) |
| -LDW State=3 -LDW warning -Not activated turn indicator -Speed $\geq$ 100 km/h | 1266 (167 checked) | 58 (34,73%) | 0 |

Throughout this analytical work it is concluded that **driver's reaction** may be the key to discriminate between *normal driving* and *crash relevant events*. In fact, whether an event if safety-critical or not depends on the driver. Styles and experiences while driving are different, so the same acceleration value can be critical in some drivers while others are fully aware. Given this diversity of drivers and personalities, reviewers examine which is the driver attitude in the videos to guess whether the situation is critical for him/her. While conventionally this classification is a subjective decision of reviewers, an automatic process can be developed by recognising drivers' reactions.

## 4.2    Recognizing drivers' reaction as potential trigger

In Chapter 2, recognizing the drivers' reaction has been used in a second phase to reduce the number of false negatives after a triggering procedure with kinematics triggers. However, since a percentage of CRE is missing when triggering with kinematics (overall near-crashes and incidents), an alternative is using the driver's reaction recognition as potential trigger directly in the database. Although the developed algorithms have been previously used as complement of kinematic triggers, in the next trial the performance of using one of them as main trigger is analyzed. Findings when applying the Harmonic mean criterion in a 40-minute trip are presented below.

■ **Recognizing drivers' reaction (trials with Harmonic mean):** The Harmonic mean criterion has been used in the past chapters to recognize the driver's reaction in complement to the kinematic triggers. Results from this criterion in the training sample were promising in the discrimination between CRE and previous situations. To analyze the potential use of Harmonic mean as main trigger, the algorithm was applied to a long trip instead of using a list of candidates to CRE from kinematics triggers.

The selected video is a trip of approximately 40 minutes with a different driver from those used previously. An event at 1174 seconds was recorded in the conventional triggering with an initial euroFOT data set. Longitudinal acceleration was used as trigger. The registered values of decelerations at such time were -6,714 and -6,723 m/s$^2$. In the video, the driver is overtaking on the fast lane (at, approx., 80 km/h) on a two-lane road. The leading vehicle starts braking and the driver of the subject vehicle slams on the brakes (two hands on the wheel and 60 bar of brake pressure) and turns the steering wheel to avoid a possible contact. Due to the delay in the reaction, the driver's response is clear in the video sequence.

A first approximation in the training sample using 6000 as threshold of harmonic mean detects nine of the eleven events (81,81%) together with two negatives (6,06%). However, the sample is not enough large to take these results in consideration. They could be used as approximation to define thresholds.

The script runs into the trip calculating every two seconds the STD of jerk. Then, harmonic means are calculated along rows and columns, respectively, to try to identify driver's silhouette. If the sum of harmonic means is above 6000, then the number of the frame and the harmonic mean are saved.

The following figure (Figure 4.1) shows the triggered frames and the harmonic mean values associated. A total of 86 situations were triggered with sum of harmonic means above 6000. Some of these results are explaining below.
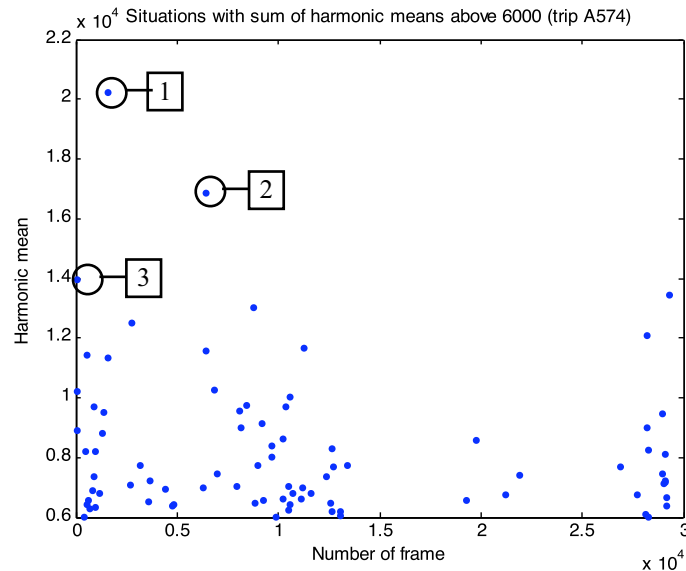
*Figure 4.1    Triggered frames and associated harmonic means in a 40-minute trip.*

**Point 1_**The maximum sum of harmonic means is 20238. This corresponds to a negative event described in the following frames from its sequence (also the STD of jerk image is represented at the right):



*Figure 4.2    Single frames and STD of jerk of the sequence of Point 1 in Figure 4.1.*

The changes in pixel intensities during the sequence may affect the results by increasing the variance. However, this action is supposed slower than a sudden reaction in the driver. Anyway, some frames have been removed when preprocessing the image, which makes the action faster than in reality.

**Point 2_** The sum of harmonic means at this point is 16875.



*Figure 4.3    Single frames and STD of jerk of the sequence of Point 2 in Figure 4.1.*

In this case, the STD of jerk image is affected by changes in pixel intensities due to constant shadows in the image. On the contrary to Point 1, changes are not belonged to movement in the image. The luminosity is not uniform and pixels are changing between bright and dark values during the sequence. This also might result in a false positive when calculating optical flow vectors, since such algorithm is based on a constant brightness.

**Point 3_** The sum of harmonic mean at this point is 13997. Another false positive is detected due to occlusion in the camera during the sequence:



*Figure 4.4        Single frames and STD of jerk of the sequence of Point 3 in Figure 4.1.*

The event has not been triggered since the sum of harmonic values are 2238,8 (below 6000). Nevertheless, the silhouette of the driver is represented in the image from the STD of jerk, similar to those obtained in the training sample (see in Figure 4.5). This difference is also more evident in comparison with the image of STD of jerk obtained in two-second before sequence.
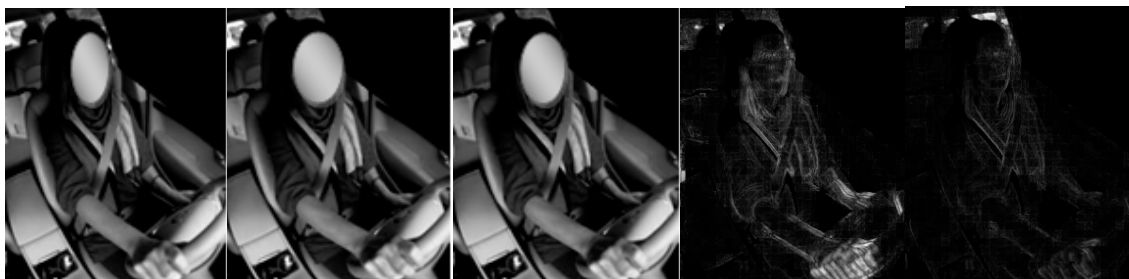


*Figure 4.5        Single frames and STD of jerk of the CRE sequence. Image on the right: STD of jerk of sequence 2-sec. before the event (false event).*

As exemplified, **occlusions in the camera and luminosity changes create changes in pixel intensity not related with driver's motion.** Therefore, is important to consider these false indicators when using driver's reaction recognition in images as potential trigger. However, the driver's silhouette appears during the event in contrast with the image from two-second before sequence. This makes possible to continue relying on a relationship between the driver's reaction and the silhouette when plotting the STD of jerk.

## 4.3    Final conclusions

■ *Recognition of driver's reaction to identify CREs:* Throughout this project has been shown that **driver's reaction is the key to validate and identify CREs**. This idea came after watching more than 400 videos from a conventional triggering procedure based on the 100-Car NDS algorithms.

NDS are an important source of information for a better understanding of the causes of road accidents. Commonly, just a few crashes (if any) are registered in the whole data set given their low frequency of occurrence. However, to have statistical sense is essential that a large enough population take part of the analysis. This can be just achieve with near-crashes and incidents. The sensitivity and specificity of kinematic triggers in the identification of such situations is very low. This is mainly due to kinematic values are quite similar to those obtained while normal driving. Therefore, one of the most important reasons for recognizing driver's reaction is mostly built in the basis of the identification of near-crashes and incidents.

■ *Quantifying the empathy:* Driver's reaction has been identified during this work as a **sudden motion of the driver's torso that can be quantified by given changes in pixel intensities**. This allows an **objective definition of what safety-critical is**.

The identification of safety-critical situations in a large data set requires empathy with the drivers. In previous NDS this was an exercise carried out by reviewers when watching videos of candidates to CRE. However, these are subjective decisions mostly based in their experience at the wheel, which makes inaccurate the comparison between different NDS.

■ *How identify driver's reaction?:* Among the several methods tested based on recognizing the driver's reaction, the **STD of jerk has emerged as a possibility in the detection of CREs by identifying a driver's silhouette.**

Given the variety of drivers and the computational time, a statistical approach has been considered in the recognition of driver's reaction. Sudden intensity changes in a group of pixels have been estimated by computing the optical flow and the standard deviation of jerk. This last option allows important savings in computational time while changes are represented as whiter pixels in a grayscale image. Throughout this work, three different situations have been detected:

> -dark image when driver remains in the same position.

> -certain white area in driving maneuvers.

> -driver's silhouette in critical situations.

■ *Automatic detection of driver's silhouette:* The **combination of STD of jerk together with the Mean criterion achieves a sensitivity of 84,21% and a specificity of 79,73%.** This procedure requires images with regular illumination, is **not computationally demanding** and allows reducing in a 66,67% the validation time. This implies a reduction of human and technical resources in the identification of CRE among the candidates while keeping above the 84% of the positives.

■ *Future research:* On basis to these findings, possibilities in future research could be:

- In order to develop an algorithm able to **complement or substitute** in the future the conventional **kinematic triggers is essential to uniform the illumination**. This can be done by implementing an adaptive threshold that differences between ranges of intensity changes due to luminance changes and due to real motion. Alternatives are the photometric normalization algorithms, mainly used to address the irregular lighting in faces verification systems or histogram manipulation functions (Struct, 2009; Heusch *et* al., 2005). It is also possible the use of techniques of detection of different intensities light events that affect locally or globally the frames collection (Ekiza and Marqués, 2010).

- A more specific definition of motion when drivers react in presence of CREs and when they drive in closed curves or over bumps on the road. The driver's shoulders seem to be an important area to make possible the discrimination.

- More specific methods of the field of **image processing** can be applied for an accurate definition of the driver's reaction and its associated motion, such as:

  - Scale-Invariant Feature Transform (SIFT) is used in the recognition of objects by taking features (specially high-contrast regions, as edges) and individually comparing each feature and finding candidates in an image collection (Lowe, 2004).

  - A multi-scale sequential image differencing can be carried out to estimate individual intensity differences in a group of pixels between pair of frames (Down, n.d.). Is possible relate this motion with specific kinematic values.

# 5    References

World Health Organization (2009): *Global status report on road safety: time for action.* Geneva, [Online] Available at: <http://www.who.int/violence_injury_prevention/road_safety_status/2009> [Accessed November 2010].

Boyle L., Lee J., Neyens D., McGehee D., Hallmark S., Ward N. (2009): *S02 Integration of Analysis Methods and Development of Analysis Plan.* SHRP 2 Phase 1 Report, University of Iowa. [Online] Available at: <http://onlinepubs.trb.org/onlinepubs/shrp2/draftS02_Phase1.pdf> [Accessed November 2010].

ERTICO-ITS Europe (2010): *euroFOT.* [Online] Available at: <http://www.eurofot-ip.eu> [Accessed November 2010].

Dingus, T. A., Klauer, S.G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., and Jermeland, J. (2006): *The 100-car naturalistic driving study, Phase II - results of the 100-car field experiment* (Contract No. DTNH22-00-C-07007). Washington, DC: National Highway Traffic Safety Administration.

Transportation Research Board (2011): *Strategic Highway Research Program 2 (SHRP2).* [Online] Available at: <http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/SHRP2ResearchReports.aspx> [Accessed January 2011].

Victor T., Bärgman J., Hjälmdahl M., Kircher K., Svanberg E., Hurtig S., Gellerman H., Moeschlin F. (2010): *Sweden-Michigan Naturalistic Field Operational Test (SeMiFOT) Phase 1: Final Report,* SAFER Vehicle and Traffic Safety Centre at Chalmers.

García Pérez (2004): *A Contribution to Solvency Analysis: The Extreme Value Theory*, [Online] Available at: <http://dspace.uah.es/jspui/bitstream/10017/637/4/Texto%20completo.pdf> [Accessed January 2011].

ERTICO-ITS Europe (2010): *Field Operational Tests: Evaluating ITS-applications in a real world environment*, FOT-NET, [Online] Available at: <http://www.fot-net.eu> [Accessed December 2010].

Klauer, S.G., Dingus, T. A., Neale, V. L., Sudweeks, J.D., and Ramsey, D.J. (2006): *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data,* Technical Report August 2004 to August 2005, NHTSA NVS-331, [Online] Available at <http://www.nhtsa.gov/DOT/NHTSA/NRD/Articles/HF/Reducing%20Unsafe%20behaviors/810594/pages/Glossary.htm> [Accessed May 2011].

Malta L., Miyajima C. and Takeda K. (2009): *A Study of Driver Behavior Under Potential Threats in Vehicle Traffic*, IEEE Transactions on Intelligent Transportation Systems, 10, No. 2, pp. 201-210.

Molinero A., Evdorides H., Naing C., Kirk A., Tecl J., Barrios J.M., Simon M.C., Phan V., Hermitte T. (2009): *Accident causation and pre-accidental driving situations. Part 2. In-depth accident causation analysis*, TRACE, Project No. 027763.

Guo F., Klauer S., Hankey J., Dingus T. (2010): *Near Crashes as Crash Surrogate for Naturalistic Driving Studies*, Transportation Research Record: Journal of the

Transportation Research Board, Transportation Research Board of the National Academies, ISSN 0361-1981, Volume 2147 / 2010, pp. 66-74.

Evans L. (2004): *Traffic Safety*. Bloomfield Hills, MI, Science Serving Society [Online] Available at <http://www.scienceservingsociety.com/ts/text/ch01.pdf> [Accessed May 2011].

Klauer S., Sudweeks J., Hickman J., Neale V. (2006): *An assessment of the relative risk of engaging in potentially unsafe driving behaviors*, Virginia Tech Transportation Institute, AAA Foundation for Traffic Safety. [Online] Available at <http://www.aaafoundation.org/pdf/RiskyDrivingReport.pdf> [Accessed May 2011].

Takeda K. (2010): *A Signal Processing Approach to the Analysis and Modeling of Driving Behavior,* Seminar at Graduate school of information science, Nagoya University, Graduate School CE.

Sato T. and Akamatsu M. (2007): *Analysis of Naturalistic Driving Behavior While Approaching an Intersection and Implications for Route Guidance Presentation,* Lecture Notes in Computer Science, Volume 4558/2007, pp. 618-627.

Nagase A., Kawanaka H., Shoaib Md., Oguri K. (2009): *Multi-class identification of driver's cognitive distraction with error-correcting output coding (ECOC) method*, Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA.

Kobayashi Y. (2008): *The Emotion Sign: Human Motion Analysis Classifying Specific Emotion*, Journal of computers, Vol. 3, N0. 9. pp. 20-28.

Young A. (2004): *The Reflexive Universe*, Anodos Fundation. [Online] Available at <http://www.arthuryoung.com/ruexc.html> [Accessed May 2011].

Iradier M (2006): *La tercera derivada del tiempo, las ondas y el análisis del pulso*, [e-book] Available through <http://www.scribd.com/doc/17296148/Iradier-Miguel-La-Tercera-Derivada-Del-Tiempo> [Accessed May 2011].

Horn B. and Schunck B. (1981): *Determining Optical Flow*, Artificial Intelligence 17, pp. 185-203.

Sun D., Roth S., Black M. J. (2010): *Secrets of Optical Flow Estimation and Their Principles*, IEEE Int. Conf. on Comp. Vision & Pattern Recognition.

Sun D., Roth S., Black M. J. (2010): *A Quantitative Analysis of Current Practices in Optical Flow Estimation and The Principles Behind Them,* Technical Report Brown-CS-10-03.

Sun D. (2010): *Matlab code for methods in CVPR2010 paper,* Brown University, Providence, RI. USA, [Online] Available at <http://www.cs.brown.edu/~dqsun/research/software.html> [Accessed March 2011].

Wilson B. (2006): *Understanding the Harmonic Mean*, [Online] Available at <http://www.cse.unsw.edu.au> [Accessed 6 April 2011].

Hasna M. and Alouini M. (2002): *Application of the harmonic mean statistics to the end-to-end performance of transmission systems with relays*, University of Minnesota, Minneapolis, USA IEEE.

Alba J.L., Cid J., Mora I. (2006): *Métodos de análisis de imágenes: Extracción de características*, Universidades Carlos III, Rey Juan Carlos and de Vigo. [Online] Available at <http://www.gts.tsc.uvigo.es/pi/Analisis%20de%20imagenes.pdf> [Accessed April 2011].

Olmos I. (2008): *Curso Procesamiento Digital de Imágenes*, Universidad Autónoma de Puebla, México. [Online] Available at <http://www.cs.buap.mx/~iolmos/pdi/Sesion3_Histograma.pdf> [Accessed 27 April 2011].

MathWorks (2011): *Image Processing Toolbox: Hough Transform.* [Online] Available at <http://www.mathworks.com/help/toolbox/images/ref/hough.html> [Accessed April 2011].

IZMIRAN (2005): *Using a Gray-Level Co-Occurrence Matrix (GLCM). [Online] Available at <http://matlab.izmiran.ru/help/toolbox/images/enhanc15.html>* [Accessed May 2011].

Hall-Beyer M. (2007): *The GLCM Tutorial. Version 2.10,* University of Calgary, [Online] Available at <http://www.fp.ucalgary.ca/mhallbey/tutorial.htm> [Accessed April 2011].

Tape T. (n.d.): Interpreting Diagnostic Tests, University of Nebraska Medical Center, [Online] Available at <http://gim.unmc.edu/dxtests/ROC1.htm> [Accessed May 2011].

Mason S.J. and Graham N. E. (2002): *Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation*, Q. J. R. Meteorol. Soc., 128, pp. 2145–2166.

Katzman D. (1989): *Analyzing a Portion of the ROC Curve,* Society for Medical Decision Making, SAGE Publications, 190.

Cleveland Clinic Lerner Research Institute (2011): *ROC Analysis*, Department of Quantitative Health Sciences, [Online] Available at <http://www.lerner.ccf.org/qhs/software/roc_analysis.php> [Accessed May 2011].

Rydström A., Moeschlin F., Sandberg D., Hjälmdahl M. (2009): *SeMiFOT Task Report: Task no.1 Crash-relevant event analysis,* V. 1.7.

Sidaway B., Fairweather M., Sekiya H., Macnitt-Ray J. (1996): *Time-to-Collision Estimation in a Simulated Driving Task*, Human Factors journal, Vol. 38.

Heusch G., Cardinaux F. and Marcel S. (2005): *Lighting Normalization Algorithms for Face Verification,* IDIAP Research Institute, Martigny, Switzerland.

Struct V. (2009): *INface: A toolbox for illumination invariant face recognition*, University of Ljubljana, [e-book] Available through <http://uni-lj.academia.edu/VitomirStruc/Papers/107120/INface_A_Toolbox_for_Illumination_Invariant_Face_Recognition> [Accessed June 2011].

Ekiza C. And Marqués F. (2010): *Use of morphological filters in detection of flashes and other light events in video sequences*, UPC (Cataluña, Spain) and Thomson Corporate Research Lab (Princeton, USA). [Online] Available at <http://upcommons.upc.edu/pfc/bitstream/2099.1/10419/1/PFC-Christian_Ekiza-

Morphological%20Filters%20in%20Detection%20of%20Flashes%20and%20Other %20Light%20Events%20in%20Video.pdf>[Accessed June 2011].

Lowe D. (2004): *Distinctive image features from Scale-Invariant keypoints*, University of British Columbia, Vancouver, Canada. International Journal of Computer Vision, 2004.

Dow M. (n.d.): *Video motion estimation and quantification*, University of Oregon Brain Development Laboratory, [Online] Available at <.http://lcni.uoregon.edu/~mark/ Geek_software/Video_motion/Video_motion_estimation.html> [Accessed June 2011]

# 6    Appendix 1



**January 2011** Triggering an initial euroFOT dataset

### Introduction

This appendix contains the results from triggering an initial **euroFOT data set using 100-Car triggers** in MatLab. These have been programmed by following the triggers description in 100-Car report (Dingus *et* al., 2006). Most of them are associated with common evasive maneuvers and acceleration peaks. Additionally, situations in which the **Lane Departure Warning** system alerts drivers have also been analyzed.

**Validation** of triggered events is based on **video viewing**. According to the driver's reaction, the triggered candidates to CRE are classified as positive or negative, depending on how critical the situation is by the driver. This report just contains general results, although each triggered situation has been described in another document.

### Triggers&Results

The following results are described in terms of :

-Trigger definition

-Number of valid events

-Most common actions

-Observation

■ **Lateral acceleration**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| Lat. Accel. $\geq|6,867 m/s^2|$ | 34 | 0 | 1 (2,9%) |

| Most common situations | Cases(%) |
|---|---|
| Roundabouts | 3 (8,8%) |
| Closed curves | 10 (29,4%) |
| Direction changes | 2 (5,8%) |
| Turning in intersections | 5 (14,7%) |
| Full-curves road | 2 (5,8%) |
| Parking maneuvers | 4 (11,7%) |
| Error values when downloading data | 4 (11,7%) |

As can be seen in the table, the most common situations are **overspeed taking closed curves and turning intersections.** In all the analyzed events, drivers are aware and don't loose the control of the car. In this context, a 11,7% of triggered events are parking maneuvers, which suggests taking speed into account.

Has also been observed some peaks around 12-15 $m/s^2$ in lateral acceleration when downloading data.

.

**Other observations:** In one of the cases, driver is playing with a kid in the rear seat just turning the steering wheel in both directions.

■ **Longitudinal acceleration**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| Long. Accel. $\geq|5,886 m/s^2|$ | 119 | 40 (33,61%) | 14 (11,77%) |

One of the most typical response in drivers is to hit the brakes trying to avoid a possible incident or collision on the road. These hard braking situations are usually related with peaks in longitudinal acceleration. Therefore, situations in which longitudinal acceleration is above a certain threshold can indicate that an incident happens.

One possible classification of triggered events can be done in terms of **why the driver brakes:**

| Most common situations | Cases | Valid |
|---|---|---|
| Leading vehicle is braking/stopped: | | |
| -Surprise&Wrong calculation of stop distance | 27 | 19 |
| -Eyes out-off the road | 5 | 5 |
| Bus incorporation | 3 | 1 |
| Intersections (low visibility) | 10 | 5 |
| Speed bump on the road | 3 | 0 |

| Most common situations | Cases | Valid |
|---|---|---|
| Roundabout | 8 | 3 |
| Vehicles changing lane | 6 | 6 |
| Traffic jam | 2 | 0 |
| Traffic lights | 8 | 0 |
| Dodging obstacles | 1 | 0 |
| Vehicle in opposite way | 2 | 1 |
| Overspeed in curves/intersections | 11 | 1 |
| Parking | 3 | 2 |

**Particular cases:**

- On a one-lane road, the driver is talking with a child that is in the rear seat. He takes his eyes off the road for approx. 4 seconds at 20 km/h. When he looks back at the road (two hands on the wheel), he realizes that is closer to crash with some shrubs. He doesn't brake, only steers the wheel (from 1023 to -632 deg/s$^2$ of steering angle jerk), but there is a brake pressure of 63 bar before turning the head to talk with the rear passenger. So, the real incident is not a result from triggering with longitudinal deceleration, it starts just before the triggered event.

- Driver is continuously turning the wheel in both directions to entertain a child in the rear seat. This event was also triggered using Lateral accel. Thresholds

- Driving on one-lane road, at approx. 50 km/h, driver slams on the brakes and stops the car without apparently reason (brake pressure: 163 bar), and, then, she continues driving. It seems that she is talking with somebody else (it's night, ice in the road, not images from more possible passengers in the car).

- Driving at 42 km/h (one hand on the wheel) on an urbanization road, driver is using a device that is between the two front seats (undetectable in the camera), when a ball appears on the road. In this moment, he is aware of the situation and reacts by braking (47 bar) but few seconds ago he was checking the device and not looking to the road. (73)

**Longitudinal acceleration & TTC**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| Long. Accel. $\geq$\|4,905m/s$^2$\| TTC $\leq$ 4s | 12 | 3 (25%) | 3 (25%) |

In one of the cases, there is not apparently reason for a TTC low value. There is not traffic in front of the car (another car is coming for the next lane in an opposite direction, but is quite far when the trigger happens). Three valid cases are found with this trigger, but two of them are, also, found by only triggering with longitudinal acceleration.

| Most common situations | Cases | Valid | Observations |
|---|---|---|---|
| Leading vehicle is braking/ stopped | 6 | 2 | -Both cases: Speed 60 km/h<br><br>-Leading vehicle starts braking and also the subject car (car is braking itself). When it's too closer, driver slams on the brake (60 bar) and turning the wheel (from 706 to -635 degr/s$^2$ of steering angle jerk variation).<br><br>-41 km/h, 31 bar |
| Parking | 1 | 0 | |
| Roundabout | 1 | 1 | -Driver is taking a roundabout. He checks the traffic situation into the roundabout. He realizes in the last moment (before a possible contact) that the leading car is stopped and waiting to get into. He stops from a speed of 35 km/h, (123 bar, one hand on the wheel, also when the incident happens) |

 **Longitudinal acceleration & TTC & FW:**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| $\|4,905\text{m/s}^2\| \leq$ lat. accel. $\leq \|3,924\text{m/s}^2\|$<br>TTC $\leq$ 4s<br>At min. TTC, FW$\leq$30,48m | 14 | 3 (21,43%) | 3 (21,43%) |

| Most common situations | Cases | Valid |
|---|---|---|
| Leading vehicle is braking/stopped | 7 | 32 |
| Traffic lights | 1 | 0 |
| Roundabout | 3 | 1 |

**■ Yaw rate:**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| Min. Variation: -4 to 4 degrees/s (or vice.)<br>(3s window)<br>Neutral position at the beginning and end of the range | 7 | 1(14,28%) | 4 (57,14%) |

Triggering with yaw rate aims to identify rapid changes in vehicle heading due to evasive maneuvers. Assuming that the yaw rate is in a neutral position at the beginning and at the end of the range, it's supposed that driver is on a straight road. Possible variations of this method could consider only changes in a yaw rate within a certain time window.

Only is possible to check the video in three of the seven triggered events.

| Most common situations | Cases | Valid | Observations |
|---|---|---|---|
| Vehicle in opposite direction | 2 | 1 | -One lane / two-way traffic<br>-Giving way to another car (narrow road) |
| Parking maneuver | 1 | 0 | |

**■ Side turn light:**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| BLISS signal<br>Turn indicator=On<br>Speed ≥ 24,12 km/h<br>(+/-1s window) | 153 (turn to left) only 49 checked | 4 (8,16%) | 0 |
| | 96 (turn to right) only 15 checked | 2 (13,33%) | 0 |

This trigger aims to identify events in which **any object is detected by BLISS system within +/-1 second of turn signal activation.** This means that the driver tries to change lanes or turn, and another car is on the blind spot. The original trigger in 100-Car is defined by using side radar signal. Since the system is not included in euroFOT fleet, BLISS is used instead (in the events the system is switched-off so drivers not receive warnings). Vehicle speed also has to be higher than 6,7 m/s (24,12 km/h).

As evidenced by filtering, dynamics of the car is not enough to **discriminate between valid and invalid** events. Throughout the analytical work in this report the key to validate is based on **driver reaction**. However, in lane changes driver is commonly aware, so that another important factor to take into account is the **behavior of others drivers** (Do they allow the lane change? Is the subject vehicle forcing other cars to stop?).

**Example of valid event:**

-On a three-lane road (traffic jam), driver is in the middle lane trying to change lanes. First attempt: car is stopped and there is a van in the left lane. Driver starts accelerating, tries to change lane but aborts maneuver. Second attempt: At approximately speed of 25 km/h, there is a car on the right side on the subject vehicle, driver is aware that such car is there, but continues with the maneuver. The driver brakes to avoid a collision between both cars.

**Example of invalid event:**

-On a three-lane road, driver is in the middle lane and changes to the left one. A car stops to allow the changing. Driver in the next lane is aware and allows the subject vehicle to change lane (also subject vehicle driver is aware).

Applying this criterion, these are the results from the checked videos (49 trying to turn left and 15 to right):

| Most common situations | Cases | Valid | Observations |
|---|---|---|---|
| Changing lanes | 27 | 6 | -Average speed: 30 km/h<br>-Several lanes road (2 or 3 lanes in most of the cases)<br>-Driver using a phone in one case |
| Joining from additional lane | 13 | 1 | -Average speed: 26 km/h |
| **False BLISS warnings** | | | |
| Vehicles in different lanes turning at the same time | 3 | 0 | -BLISS warns when vehicles in different lanes turn together in intersections |
| Intersections | 10 | 0 | -In two cases car is carrying a towing<br>-Average speed: 33 km/h |
| Roundabouts | 5 | 0 | -Average speed: 35 km/h |

As can be seen in the results, this trigger not performs well related to event validation, but can be useful to define a scenario with several lanes and traffic jam situation.

**■ Side yaw rate:**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| Min. Variation: -2 to 2 degrees/s (or vice.)<br>(3s window)<br>Neutral position at the beginning and end of the range<br>-Speed ≥ 24,12km/h | 17 | 1(5,88%) | 1 (5,88%) |

This trigger aims to identify events in which yaw rate go to the neutral, to +2 degr./s, oscillated back to -2 degr./s (or vice versa) and then return to neutral within a 3-second-window-time. It's less restrictive than the Yaw Rate filter, however speed is take into account:

| Most common situations | Cases | Valid | Observations |
|---|---|---|---|
| Turning steering wheel before taking an intersection | 4 | 0 | -Average speed: 32 km/h |
| Speed bump | 2 | 0 | |
| Dodging a pothole | 4 | 0 | |
| Road works | 1 | 0 | |
| Stopping on the road to talk with a pedestrian | 1 | 0 | |
| Controlling the steering wheel with one hand while talking on the phone | 1 | | -Speed: 30 km/h |

In this case, the only triggered event considered valid is:

-On an urban road, driving at 35 km/h, driver turns the steering wheel because a vehicle parked on the right suddenly opens its door. Surprise reaction in the driver.

■ **Lane Departure Warning:**

| Trigger definition | Total cases triggered | % Valid | %Video not available |
|---|---|---|---|
| -LDW State=3 <br> -LDW warning <br> -Not activated turn indicator <br> -Speed ≥ 100 km/h | 1266 <br><br> (167 checked) | 58 (34,73%) | 0 |

This trigger aims to identify events in which LDW system detects a car deviation above 100 km/h of speed. The defined state of LDW is 3, which means that the lane tracker is performing well and there is a good visibility. Velocities above 100 km/h can also be a referent of high speed roads where driver is more likely to be distracted (maybe drivers are more aware on urban roads). An event is considered valid if car deviation is unintended. It's not sure to assess that car corrections belong to warnings.

In most of the cases, vehicles are driving in a highway and drivers are involved in secondary tasks, as talking by phone, taking pictures or using the computer. Also in some cases, the system was activated due to symptoms of driver's sleepiness when driving at night.

# 7 Appendix 2

## Definition of training sample

The training sample contains a two-second sequence of video of eleven different events. These have been true triggered from a conventional triggering process. The triggers used in each case are explained below together with a description of the situation. The different states experimented by some drivers during the whole sequence are explained as example.

<div align="right">A686</div>

---

**Triggered by:** Longitudinal acceleration (-7,9270; -7,3620 m/s^2)

**Description:** Car is stopped waiting to turn left in an intersection. Leading car starts driving so the driver presses the accelerometer and at the same time checks the phone (no hand on the wheel). When the driver realizes that the leading car is braking he slams on the brakes to prevent the collision (82 bar of brake pressure) and grabs with one hand the steering wheel (but he doesn't move the steering wheel).

**States:**
-Initial state: driver is checking the mobile phone (without hands on the steering wheel).
-Second state: change in driver's face expression, looking ahead, one hand on the wheel (other hand on the phone along whole the sequence).
-Third state: the driver's body leans forward (it's detected a change in the visible area of the rear window regarding driver's head position).
-Forth state: driver returns to the starting position but with the head tilted slightly back.

<div align="right">A484</div>

---

**Triggered by:** Longitudinal acceleration (-8,8430; -9,8090; -9,0350; -8,0840; -8,0340; -7,7460)

**Description:** On a two-lane road, driver is manipulating some buttons in the wheel (two hands on the wheel) at speed of 50 km/h. The leading vehicle starts stopping, so she also brakes, but she is looking at something in the middle of the car (probably the cd player, during a 3-seconds glance). When she realizes the leading vehicle is too closer, so she slams on the brakes (97,5 bar) and turns the steering the wheel to the right (steering angle jerk from 1502 to -1042 degrees/s^2).

**States:**
-Initial state: driver manages with both hands on the steering wheel (not looking at the road).
-Second state: there is a change in driver's face expression, looking ahead, and same position with hands on the wheel.
-Third state: the driver's body leans forward (it's detected a change in the visible area of the rear window regarding driver's head position).

-Forth state: driver returns to the starting position and looks back to the same object from the beginning.

**Triggered by:** Longitudinal acceleration (-8,3560; -7,1660 m/s^2)

**Description:** On an urban road, driver manages with left hand on the steering wheel (in the other he carries an object, apparently a phone). In the moment of the incident, he is driving at low speed (around 30-20km/h) and he is looking at right to a truck that is maneuvering. He doesn't realize that the leading car is braking and he reacts by slamming on the brakes (brake pressure=80bar).

**States:**
-Initial state: river with head turned to the right, one hand on the steering wheel and the other holding a phone.
-Second state: change in driver's face expression, looking ahead, same hands position.
-Third state: the driver's body leans backward (due to slams-on-the-brakes action).
-Forth state: driver backs to the initial position.

**Triggered by:** Longitudinal acceleration (-7,399 m/s^2) &TTC (1,822s)

**Description:** The subject vehicle gets an exit at 60 km/h (one hand on the wheel). The leading vehicle starts braking and, also, the subject car brakes (not movement in the right foot, seems that it's the car which is breaking itself). When it's too closer to the leading car, driver slams on the brakes (60 bar) and turns the steering wheel (from 706 to -635 degr/s^2 of steering angle jerk variation).

**Triggered by:** Longitudinal acceleration (-6,8220 m/s^2) &TTC (0,9030 s)

**Description:** The subject vehicle intends to take a roundabout. He is checking the traffic situation into the roundabout. He realizes in the last moment (before a possible contact) that the leading car is stopped and waiting to also get into the roundabout. He stops from a speed of 35 km/h, (123 bar of brake pressure, one hand on the wheel remaining in the same position during the incident).

**Triggered by:** Longitudinal acceleration (6,6120 m/s^2)

**Description:** The subject vehicle is in a parking at speed of 30 km/h. Driver is talking with a kid that is in the front seat. Suddenly, a parked car reverses. Then the driver of the subject vehicle slams on the brakes (44 bar) and turns the steering wheel (from 5713 to -3981 degr./s^2).

**Triggered by:** Longitudinal acceleration (-6,2730 m/s^2)

**Description:** On an unpaved road in a country side, the subject vehicle is driving at low speed (10 km/h). Driver is looking outside (his head is turned) and he doesn't realize until the last moment that there is a tree in front of the car. He reacts by braking (23 bar of brake pressure) and by turning the steering wheel (7000 to -3287 degr./s^2).

**Triggered by:** Longitudinal acceleration (-6,3490 m/s^2)

**Description:** The driver of the subject vehicle slams on the brakes given that a car from he next lane intends to change to its lane. The motion is very evident in driver's body, who also releases both hands on the steering wheel.

**Triggered by:** Longitudinal acceleration (-6,0470; -6,1420 m/s^2)

**Description:** The subject vehicle is driving at 65 km/h. It's night. The driver releases the steering wheel and starts opening a candy. Six-seconds later she manages the steering wheel with both hands again, and she starts looking outside. Then she turns to look at the road and she realizes that the leading vehicle it's braking. She reacts by turning the steering wheel (from 2065 to -4372 degr. /s^2 of steering angle jerk) and by slamming on the brakes (109 bar of brake pressure).

**Triggered by:** Side yaw rate

**Description:** On an urban road, driving at 35 km/h, driver turns the steering wheel because a vehicle parked on the right suddenly opens its door. The driver of the subject vehicle reacts by turning the steering wheel.

**Triggered by:** Longitudinal acceleration (-9,5850 m/s^2)

**Description:** On a two-lane road, the subject vehicle is on the right lane at 82 km/h. It seems that the driver is talking, but its night and it's not possible to identify in the images if there are more passengers in the car. The road is wet. The leading car stops. The driver reacts later by slamming on the brakes (101 bar of brake pressure).

# 8    Appendix 3

## More specific information about Optical Flow

■ **Classical flow formulation: Horns and Schunck** The following method was developed by Horns and Schunck (1981) to compute the optical flow in sequences of images. Assumptions are that the illumination is uniform across the surface and the brightness varies smoothly, without discontinuities (so it's differentiable). However discontinuities appear when one object hide another, so results can be not clear in the steering wheel area, for instance. The relationship between brightness and speed motion is expressed as follows:

$$E_x u + E_y v + E_t = 0$$

Horn and Schunck state that additional constrains are needed to calculate the optical flow in independent pixels (without information from the neighborhood). This is because pixel speed requires two components to be defined and the change in brightness due to the motion is only one. Then, an additional constraint called smoothness constraint is used to minimize the square of optical flow velocity gradient. So, Horn-Schunck' method estimates the velocity vector $[u\ v]^T$ along time by minimizing this equation:

$$E = \iint (I_x u + I_y v + I_t)^2 \, dxdy + \alpha \iint \left\{ \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \right\} dxdy$$

Spatiotemporal image brightness derivatives     Global smoothness     Spacial derivatives of each optical velocity component

Where is a (x,y) point of the image at time t, brightness is denoted by E(x,y,t).

**Brightness derivatives approximation:** Since brightness' measurements are discrete values, brightness derivatives can be estimated using as reference a point in the center of a cube formed by eight measurements. Then, Ix, Iy and It are the average of first differences along each parallel edges of the cube.

$$
\begin{aligned}
E_x &\approx \tfrac{1}{4}\{ E_{i,j+1,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i+1,j,k} \\
&\quad + E_{i,j+1,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i+1,j,k+1} \}, \\
E_y &\approx \tfrac{1}{4}\{ E_{i+1,j,k} - E_{i,j,k} + E_{i+1,j+1,k} - E_{i,j+1,k} \\
&\quad + E_{i+1,j,k+1} - E_{i,j,k+1} + E_{i+1,j+1,k+1} - E_{i,j+1,k+1} \}, \\
E_t &\approx \tfrac{1}{4}\{ E_{i,j,k+1} - E_{i,j,k} + E_{i+1,j,k+1} - E_{i+1,j,k} \\
&\quad + E_{i,j+1,k+1} - E_{i,j+1,k} + E_{i+1,j+1,k+1} - E_{i+1,j+1,k} \}.
\end{aligned}
$$

*Figure 8.1    Figure with equations of derivatives reproduced from: Horn B. and Schunck B. (1981): Determining Optical Flow, Artificial Intelligence 17, pp. 185-203 .*

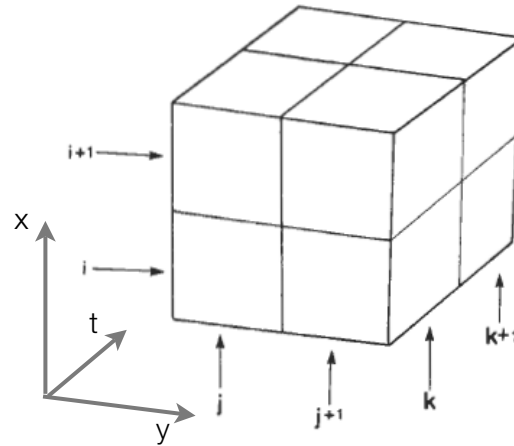*Figure 8.2    Representation of brightness derivatives approximations reproduced from: Horn B. and Schunck B. (1981): Determining Optical Flow, Artificial Intelligence 17, pp. 185-203.*

**Estimating Laplacian of smoothness constrain:** The square of the gradient of optical flow velocity is used as smoothness parameter. The Laplacian is approximated by calculating an average of weighted values in neighboring pixels:

$$\nabla^2 u \approx \kappa(\bar{u}_{i,j,k} - u_{i,j,k}) \quad \text{and} \quad \nabla^2 v \approx \kappa(\bar{v}_{i,j,k} - v_{i,j,k}),$$

*Figure 8.3    Figure with Laplacian equation reproduced from: Horn B. and Schunck B. (1981): Determining Optical Flow, Artificial Intelligence 17, pp. 185-203.*

The weights of neighboring pixels are:



*Figure 8.4    Weights of neighbour pixels, reproduced from: Horn B. and Schunck B. (1981): Determining Optical Flow, Artificial Intelligence 17, pp. 185-203.*

**Iterative solution:** By taking these approximations, Horn-Schunck' method minimizes the general equation to obtain the velocity vector [u v] for each pixel in the image, which is defined by:

$$u^{n+1} = \bar{u}^n - E_x[E_x\bar{u}^n + E_y\bar{v}^n + E_t]/(\alpha^2 + E_x^2 + E_y^2),$$
$$v^{n+1} = \bar{v}^n - E_y[E_x\bar{u}^n + E_y\bar{v}^n + E_t]/(\alpha^2 + E_x^2 + E_y^2).$$

*Figure 8.5    Figure with equations of OF velocities reproduced from Horn B. and Schunck B. (1981): Determining Optical Flow, Artificial Intelligence 17, pp. 185-20*

**■ Contributions to classical optical flow: Sun, Roth & Black_** The classical formulation of optical flow relies on constant brightness and smoothly variations. However, it seems that the quadratic formulation is not enough robust to outliers. Solutions are to replace the quadratic error function with a robust function, as using smoothness values or image segmentation. Therefore, variations using as base this classical formulation have been developed during the last years.

D. Sun, S. Roth and M. J. Black (2010) find that applying a median filter to intermediate flow values every warping step improves the accuracy. They distinguish two concepts depending on optical flow calculations: the *model* ("objective function that defines the problem"), and the *method* ("optimization algorithm and implementation used to minimize it"). A "non-local term" is added to the classical formulation "to integrate information over large spatial neighborhoods" (Sun, Roth & Black, 2010).

Before the flow estimation, they propose to convert the image to LAB space. By doing this, the brightness is ignored and the white, pink and blue colors are used to define images. Therefore, images can be represented using three layers:

L\*a\*b - - - L=luminosity layer
a=chromacity layer (red-green)
B=chromacity layer (blue-yellow)

Then, the texture is decomposed in a linear combination of structure and texture by using the *Rudin-Osher-Fatemi* (ROF) method. Optical flow is estimated in a pyramidal process that warps the second image and its derivative (using a 5-point derivative filter) toward the first one in a bicubic interpolation. Standard deviation of the *Gaussian anti-aliasing filter* is used as down-sampling factor.

They perform an analysis on the *Middlebury* training set and evaluate the results using a *Wilcoxon* rank, considering three different penalty functions:
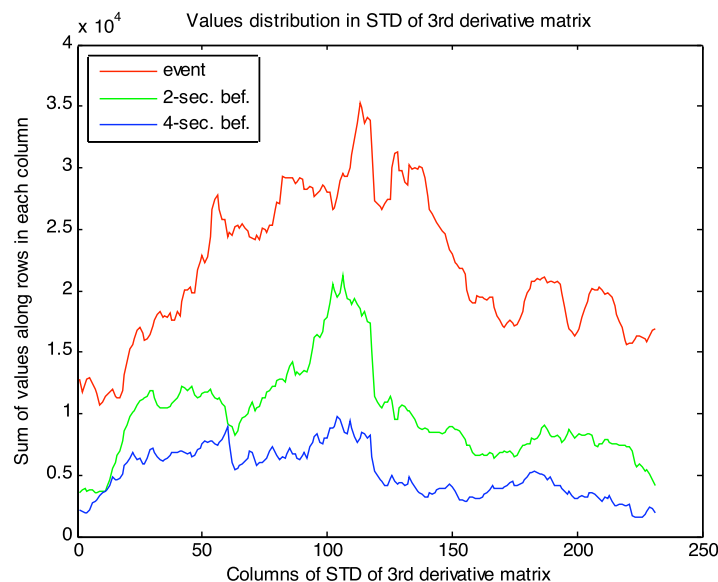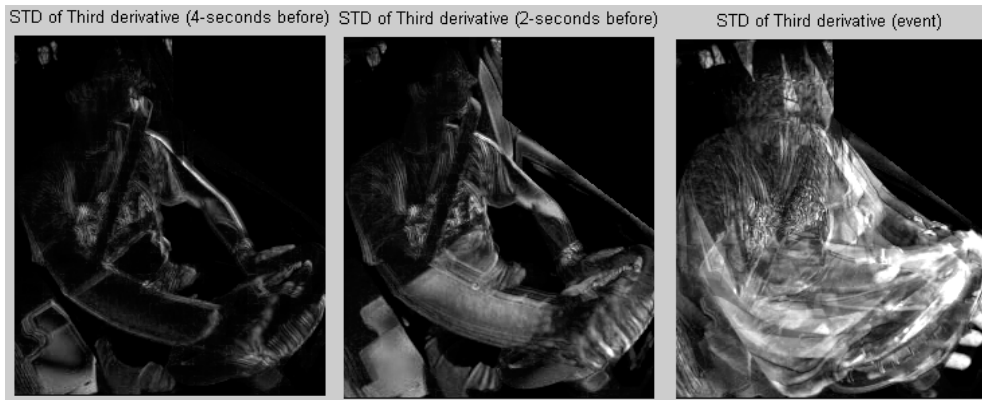
-Quadratic (HS)
-Charbonnier (Classic-C)
-Lorentzian (Classic-L)

They find that the best performance is achieved by using a *Charbonnier* penalty function, although is not as robust as a *Lorentzian* one[22]. Also, they conclude that median filtering allows to achieve better accuracy by making the flow less noisy. However, it seems that median filtering changes the objective function that is optimized, which makes higher energy in flow values (*oversmoothing*). To solve it, they propose to weight the non-local term taking into account which pixels are from the same surface, using spatial, color-value distance, or occlusion state.

To sum up, the ***main and recent contributions*** in their CVPR 2010 publication (Sun, Roth & Black, 2010) to optical flow can be addressed as follows:

---

[22] optimization is easier with convex functions.

- Image conversion to the *LAB* space and texture decomposition.

- Flow estimation in a pyramidal process, by warping the second image and its derivative (using 5-point derivative filter) toward the first in a bicubic interpolation.

- Use *Charbonnier* penalty function

- Apply median filtering to intermediate flow values every warping step. Although this improves the accuracy by making the flow less noisy, some values are over smoothed. To solve it, they propose to weight the non-local term taking into account which pixels are from the same surface. Boundaries are defined by *Sobel* edge detection method.

# 9 Appendix 4

STD of Third derivative (4-seconds before) — STD of Third derivative (2-seconds before) — STD of Third derivative (event)



Values distribution in STD of 3rd derivative matrix

Images from the standard deviation of the $3^{rd}$ derivative suggest that there is a relationship between the driver's silhouette and the response reaction at the event. One way to detect the difference it's by looking at values distribution of each image.

This graph shows the distribution of sum of values along rows in each column. Since the white area is more significant at the moment of the event, values are relatively higher respect seconds-before sequences.



Jerk of average speeds (2nd derivative of O.F.)

Range:3.456-(-5.123)=8.579
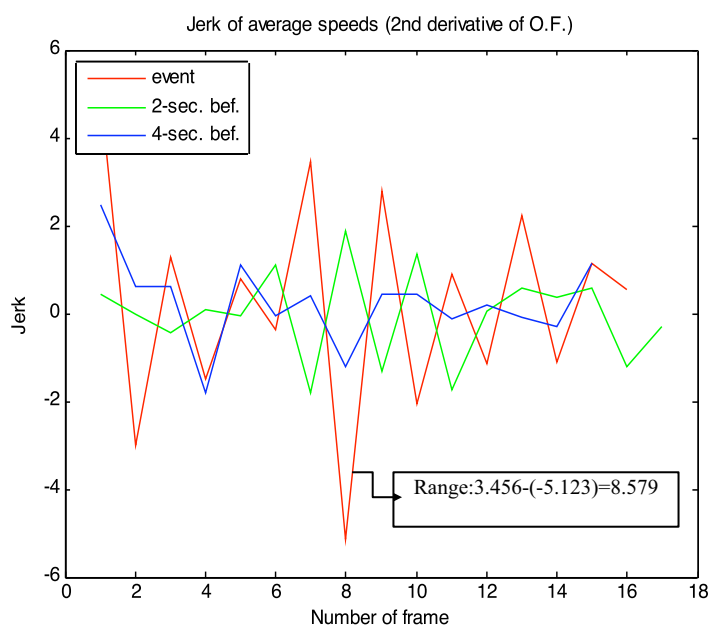
From optical flow values, the rate of change of acceleration is considerably higher in the event than seconds before, with a range of jerk of 8,579 (red line in the graph).

STD of Third derivate (4-seconds before)    STD of Third derivate (2-seconds before)    STD of Third derivate (event)



Values distribution in STD of 3rd derivative matrix
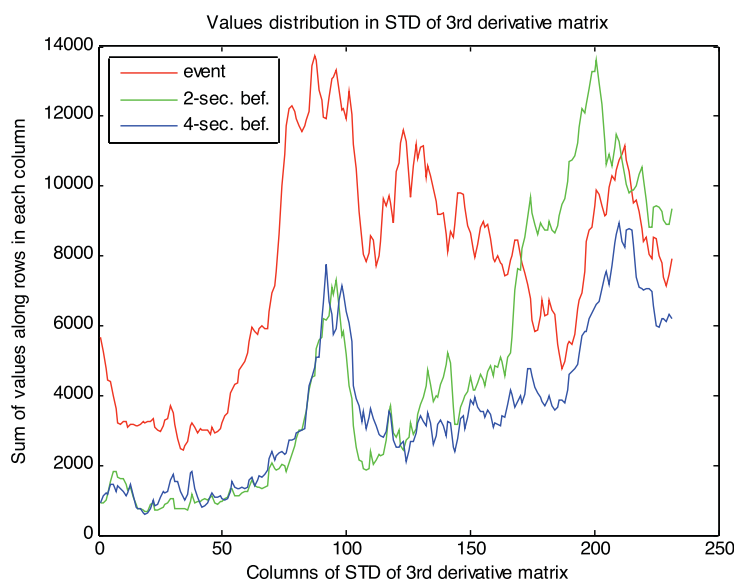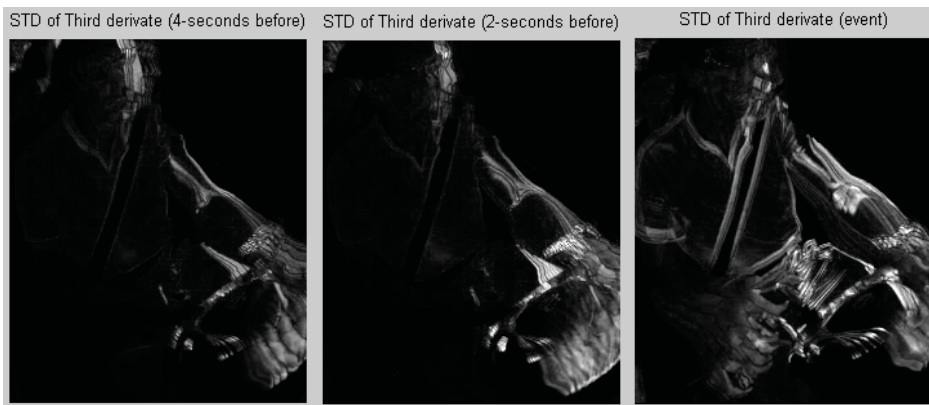
The movements when manoeuvring are located in certain areas of the standard deviation pictures. The main difference with a response reaction in the driver is that it involves the whole body. Then, it's expected a higher distribution of sum of non-zero values along the columns in such images.

As can be seen in the graph, that is the tendency in the event line until 170-column. Then sequences from seconds before the event achieve highest sum of values, due to white areas from turning the steering wheel.



Jerk of average speeds (2nd derivative of O.F.)

Range: 2.056-(-4.484)=6.540
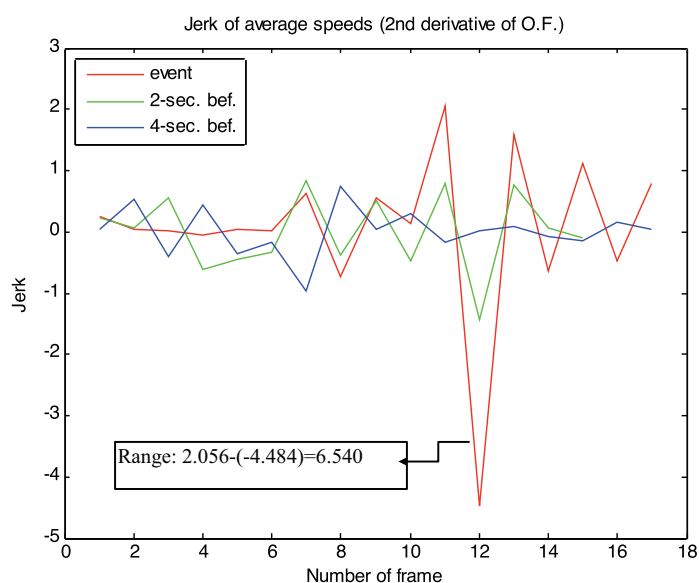
Trends of jerk distribution in sequences before the event are rather stable along the frames in the "Jerk of average speeds" graph. The steepest slope is given during the event over again, with a range of change in acceleration of 6,540.

STD of Third derivate (4-seconds before)    STD of Third derivate (2-seconds before)    STD of Third derivate (event)
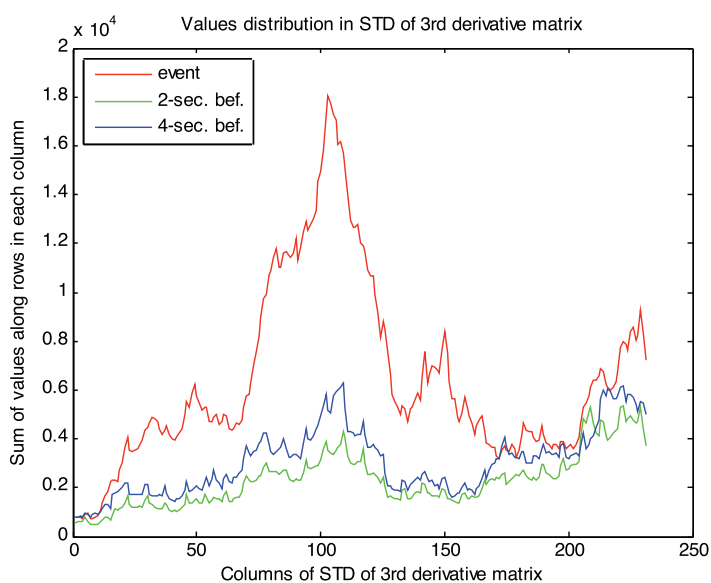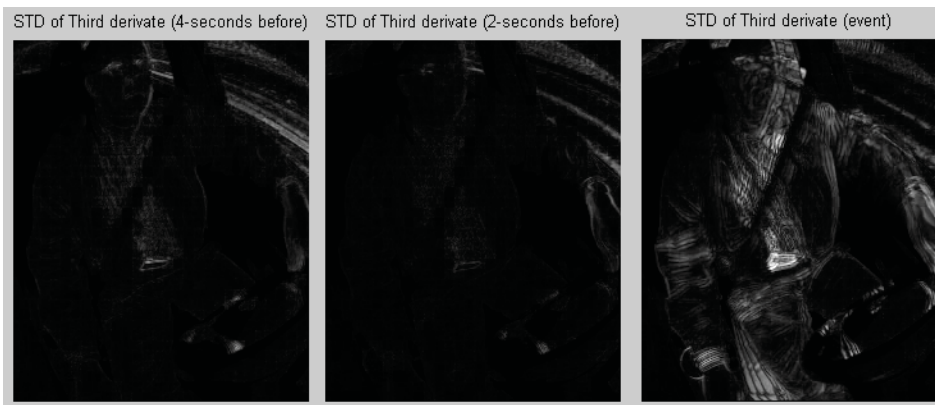


The graph shows a clear influence of white areas in the sequence of the event over previous times. Black areas are represented as zeros in the images, so the greater sum of non-zero values along the rows, the more white pixels in such column.

By looking at the images from standard deviation of 3rd derivative seems like the driver remained in the same position seconds before the event.
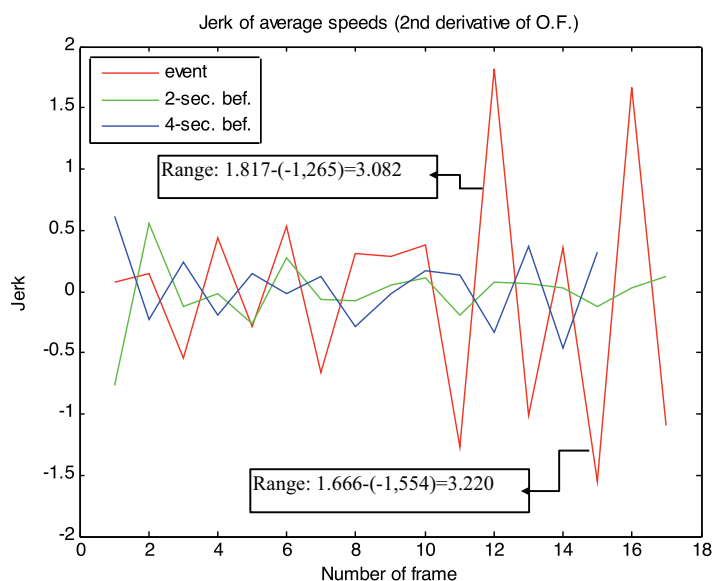


The jerk distribution from optical flow calculation reaches the minimum and maximum values during the event. Unlike in previous cases, two main ranges can be distinguished in the graph.

During the sequence there is a clear forward movement of the driver due to hard braking. Both ranges may correspond to this forward movement and its return to the original position, respectively.

STD of Third derivative (4-seconds before) — STD of Third derivative (2-seconds before) — STD of Third derivative (event)



Values distribution in STD of 3rd derivative matrix



Jerk of average speeds (2nd derivative of O.F.)
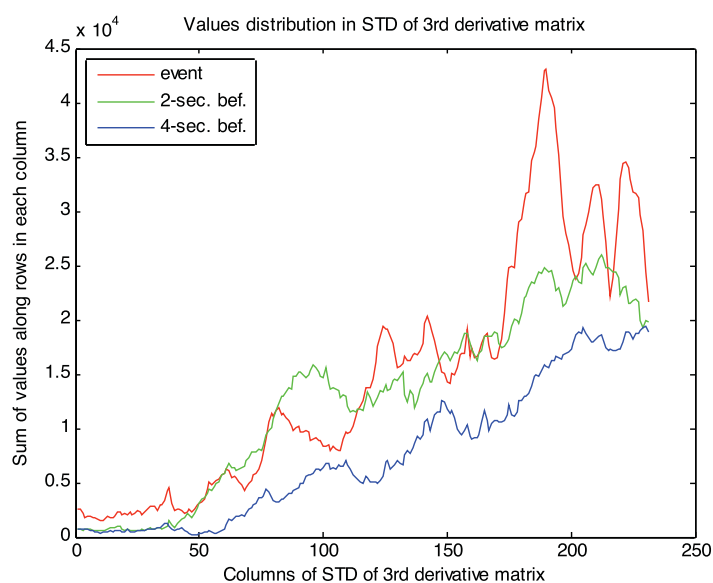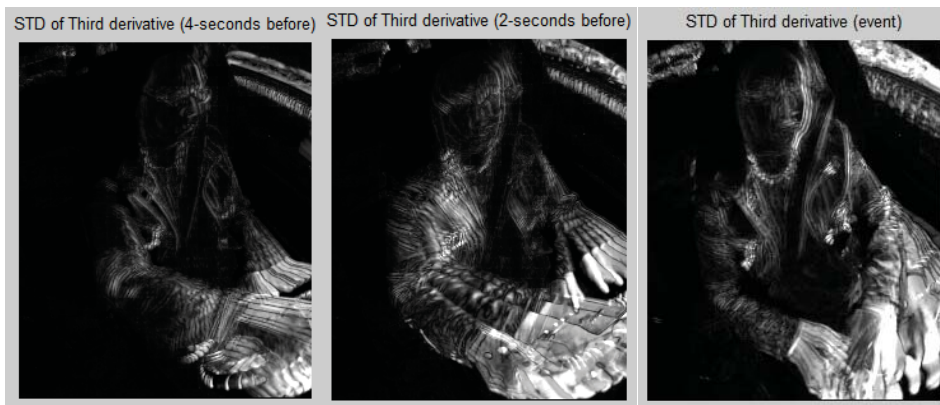
In this case the event is not as evident as in other drivers in the images from standard deviation calculus. Driver is turning the steering wheel seconds before the event and she also reacts in that way during the real event.

As can be seen in the graph, distributions are really close, although there is an appreciable difference in 190's column. This shape is more related with manoeuvres than with the driver response reaction, since the main white area is concentrated in one part of the image. Given that the response should affect the whole body, distribution should be more uniform to asses this theory.
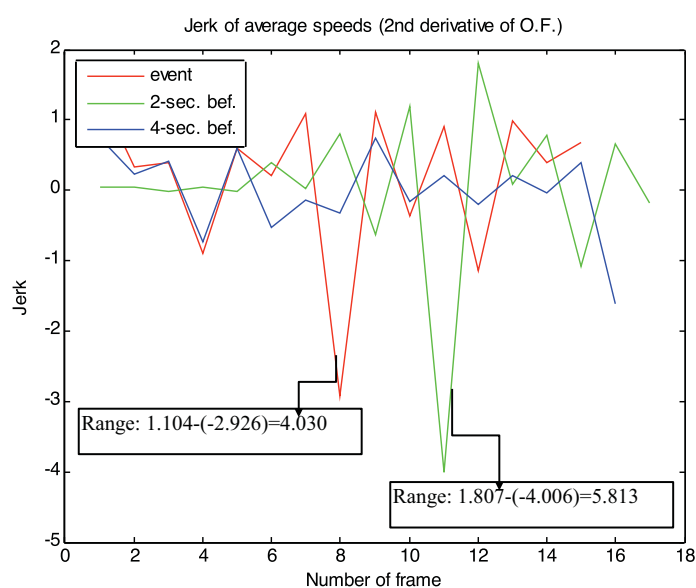
Results from the jerk calculation are also unexpected. The highest rate in acceleration changes occurs two seconds before the real event, in a completely normal driving situation (green line).

Given that some frames were removed from the sequence, accelerations may have been affected. During this sequence in particular, the flashes are not observed. Anyway, the filter is applied without distinction, removing those frames that have more light every certain frequency. To verify the repercussion of this fact, calculations were carried out again taking the entire sequence:
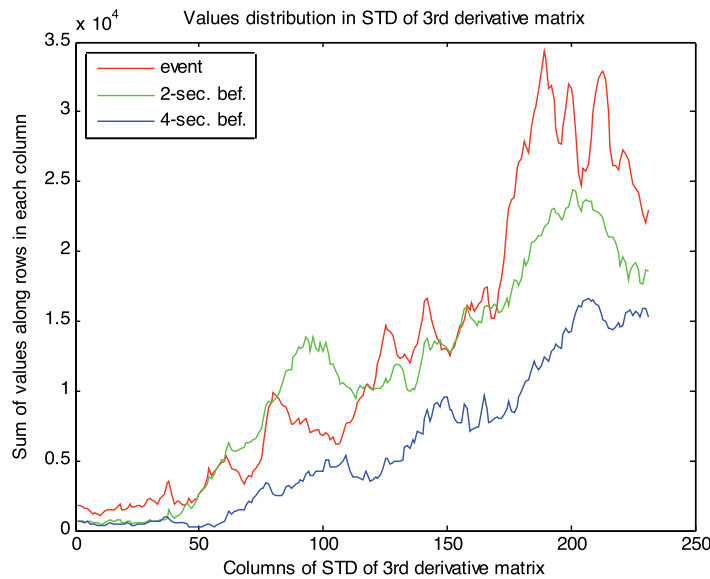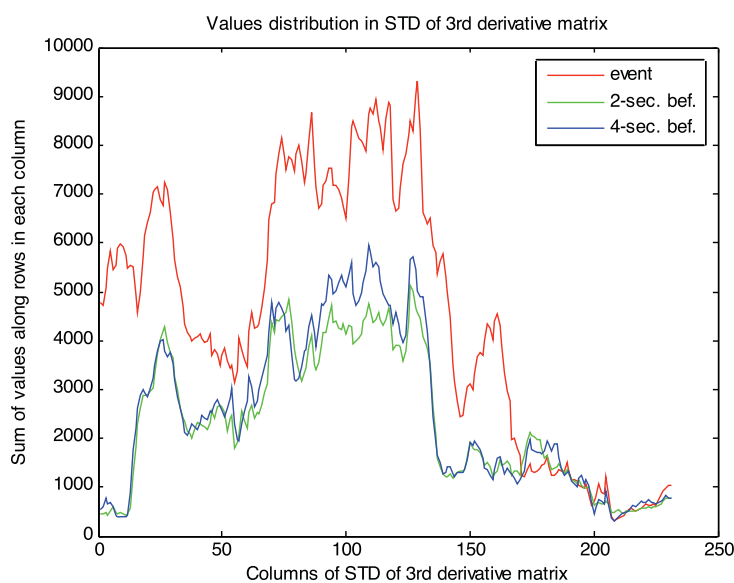


The distribution of sum of row values from STD of jerk is very similar to the previous one. Although there are few differences after the 200's columns, the shape remains.
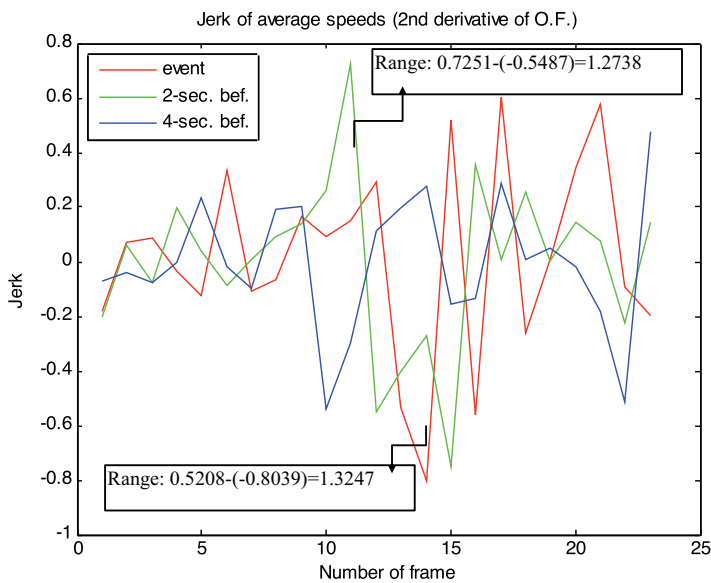


However, the curves of jerk distribution are quite different. It seems that removing some frames affects the results from the optical flow estimation of velocities.

In this figure, it can be seen the two peaks from the previous case. Anyway, the highest ranges are also reached during the event and in the two seconds-before sequence.
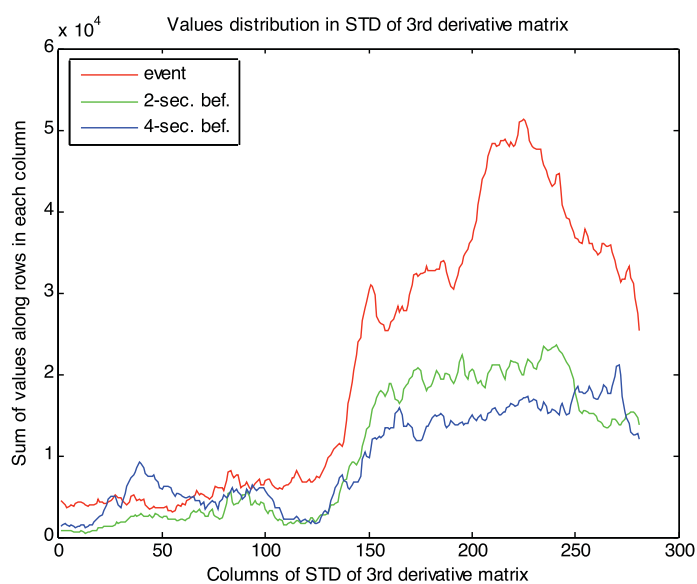
In comparison with previous drivers, this value of acceleration rate is not very significant.

STD of Third derivative (4-seconds before)    STD of Third derivative (2-seconds before)    STD of Third derivative (event)



Values distribution in STD of 3rd derivative matrix

This case can be complex since there is not an appreciable change in driver's motion during the event. Anyway, results from STD of jerk are surprisingly positive in terms of reaction recognition (the driver's silhouette appears during the event, as can be seen in the images above). In addition, the distribution of sum of values along rows is relatively higher during the event than seconds before, as shown in the graph.



Jerk of average speeds (2nd derivative of O.F.)

Range: 1.122-(-1.848)=2.970
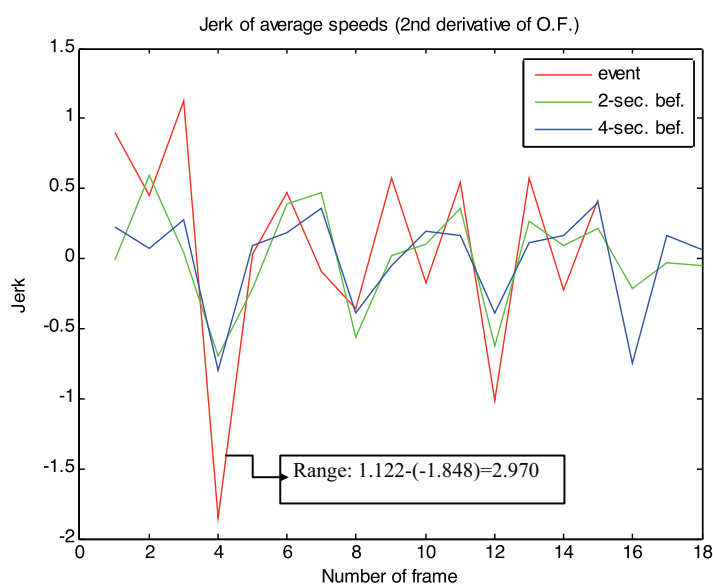
About the jerk distribution, a peak is registered during the event with a rate of change slightly lower than in previous drivers, but significantly higher in comparison with the other sequences.

STD of Third derivative (4-seconds before)    STD of Third derivative (2-seconds before)    STD of Third derivative (event)



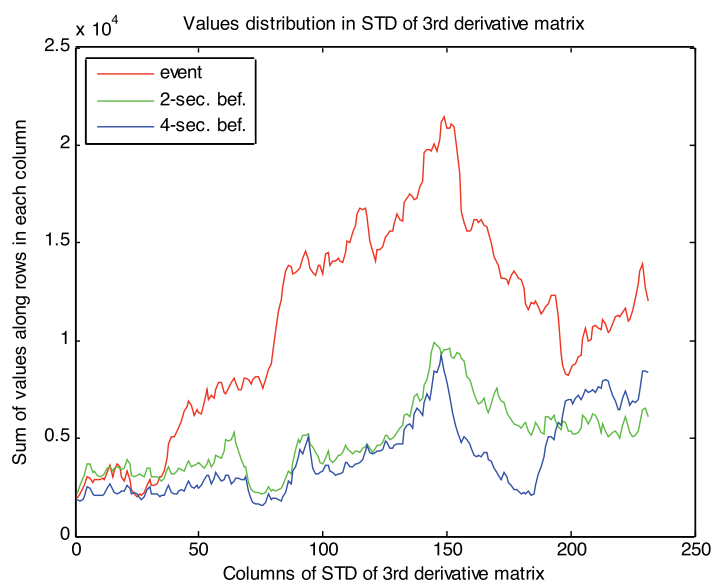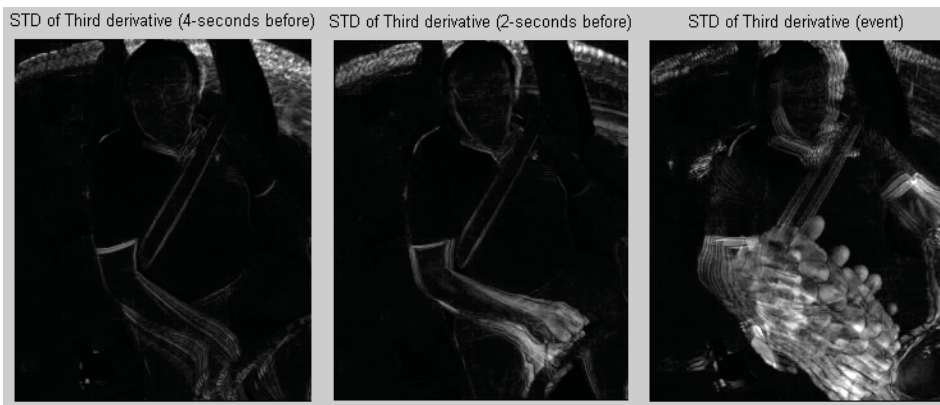Values distribution in STD of 3rd derivative matrix

In this case, the driver's silhouette is in all the images from STD of jerk, although the greatest white area is concentrated in the event due to an evasive manoeuvre. Therefore, the distribution is relatively similar in all sequences along the first columns of images but is significantly higher when driver turns the steering wheel during the reaction.



Jerk of average speeds (2nd derivative of O.F.)
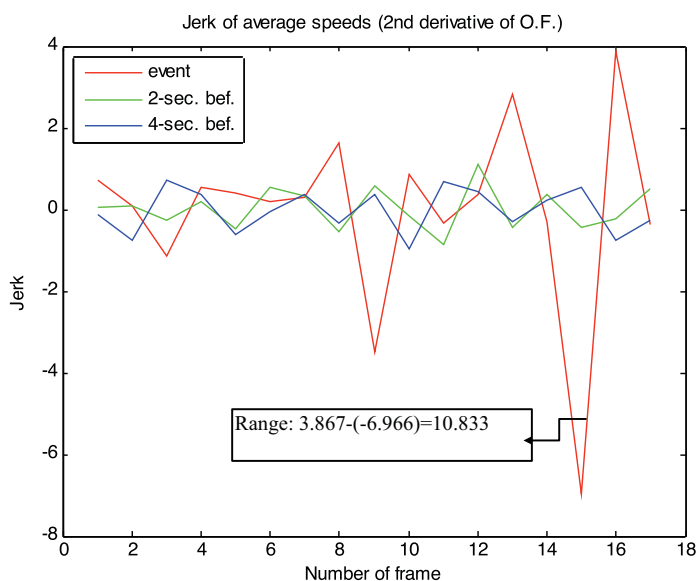
Respect to the jerk of average speeds, the highest range in the whole data sample is achieved in this case with 10.833 of rate in acceleration changes.

Trends of previous sequences are relatively stable along both arrays.

Range: 3.867-(-6.966)=10.833

STD of Third derivative (4-seconds before)   STD of Third derivative (2-seconds before)   STD of Third derivative (event)
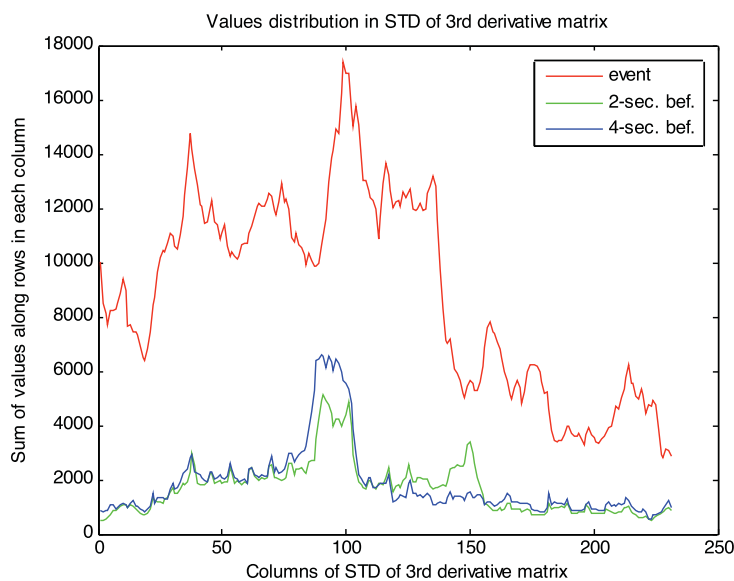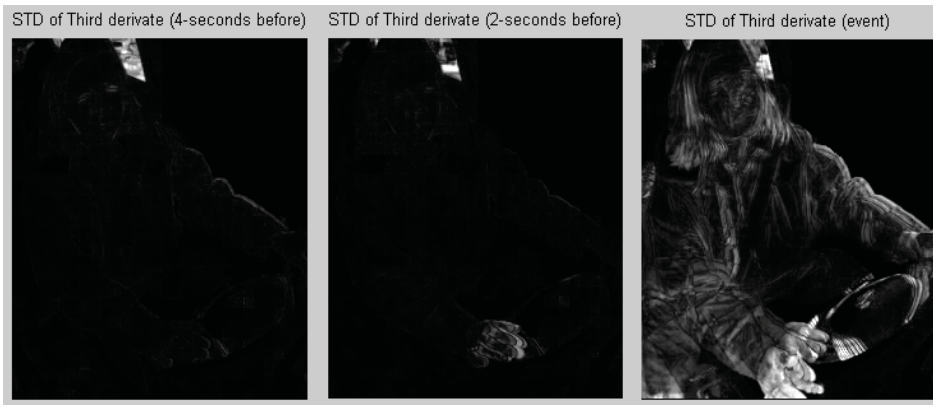


This distribution is a good example of what it's expected from the STD of jerk calculus, because in almost all the columns the sums of values along rows are higher during the event than in seconds before sequences. This means that there are more pixels containing white values along the columns in the image from the event, so they can represent a possible silhouette.

It is also possible due to the reaction of the driver is accompanied by movement in the body, especially in the area of the hand.
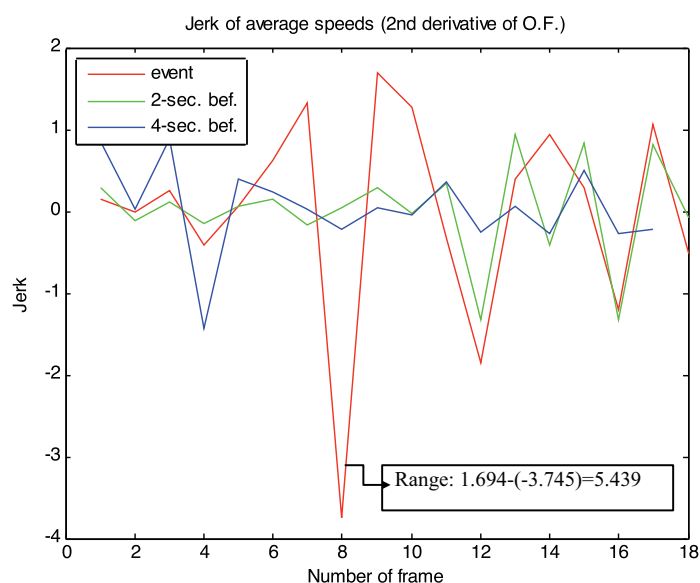


The peak in jerk distribution is also achieved during the event. Sequences from previous time remain stable along their arrays.

STD of Third derivate (4-seconds before) — STD of Third derivate (2-seconds before) — STD of Third derivate (event)



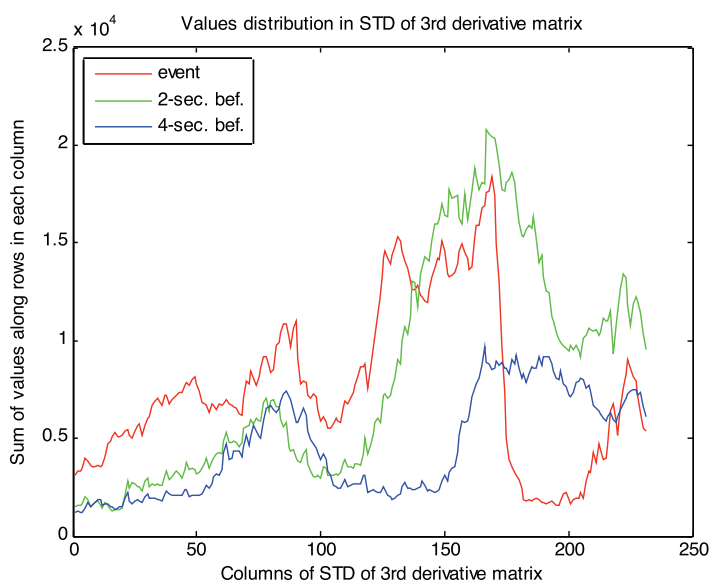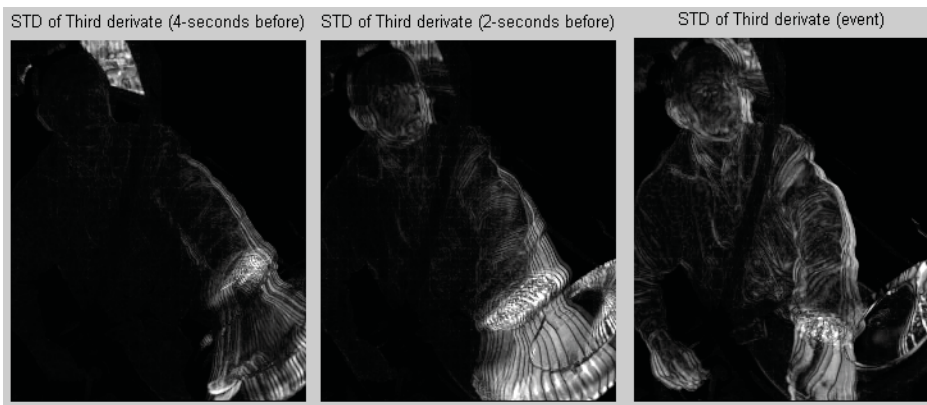Values distribution in STD of 3rd derivative matrix

The driver remains in the same position in both sequences before the event. Therefore, the image from the STD of jerk may be an indicative of the driver's reaction in this case. The sum of values along the rows for each column is also significantly higher than in previous time as can be seen in the graph.



Jerk of average speeds (2nd derivative of O.F.)

Range: 1.694-(-3.745)=5.439
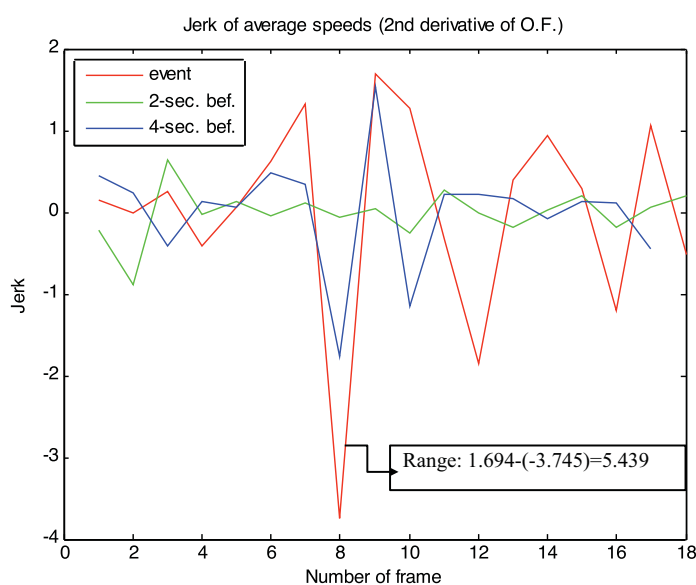
Since the driver remains in the same position over time, the distributions of rate in acceleration changes should be relatively constant before the event. Some unexpected results were obtained four seconds before at 8th iteration of jerk calculation from optical flow values. Anyway, the maximum jerk is reached during the event.

In this case, the sequence is affected by flashes every five images, but these have been removed in a previous step before the optical flow estimation.

STD of Third derivate (4-seconds before)    STD of Third derivate (2-seconds before)    STD of Third derivate (event)
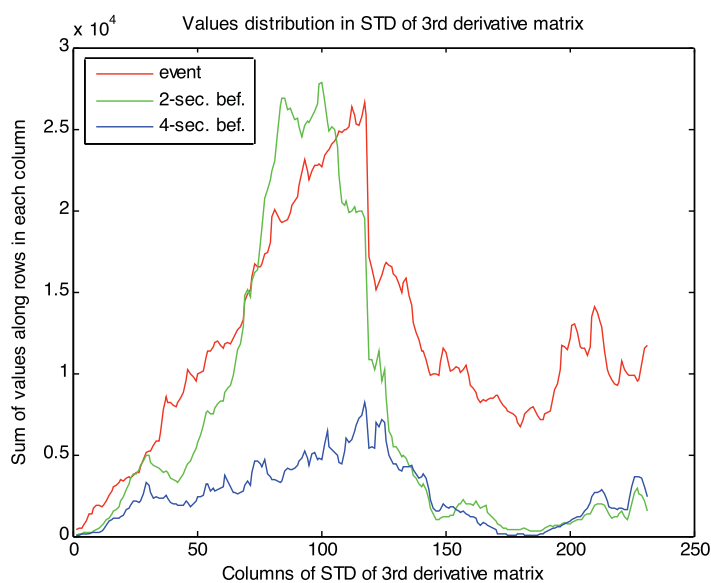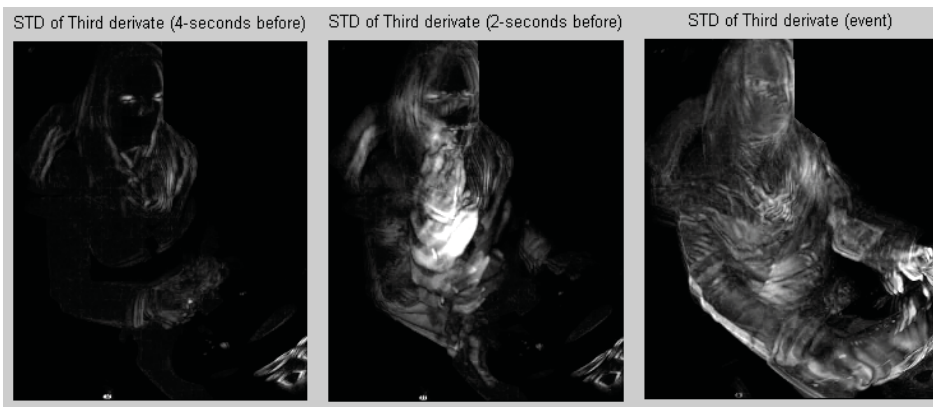


The driver's silhouette begins to be visible in the image from the event. Therefore, sums of the STD of jerk values are higher in the first columns than in previous sequences.

Nevertheless, driver was turning the steering wheel two seconds before the event. Then, as can be seen in the graph, the green line reaches the maximum at almost the middle of the image. This might be the main difference between the driver's reaction and a normal manoeuvre: the presence of concentrated white areas in a certain part of the image.
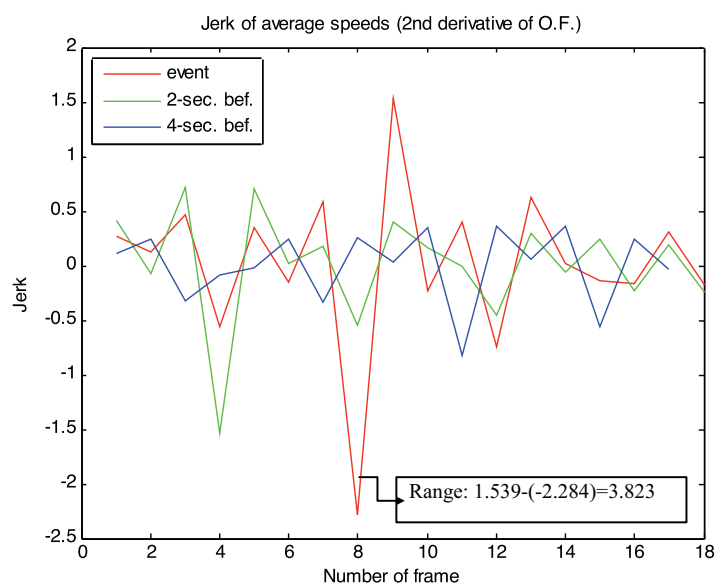


Regarding the jerk distribution, the minimum and maximum correspond to the event sequence. The peak in the green line from two-seconds-before values can be due to position changes in the driver, who moves the head and turns the steering wheel.

It is interesting to notice as the changes in acceleration ranges vary depending on the driver. In this case, the main range of jerk of average speeds is 3.823 (in terms of acceleration changes).

STD of Third derivate (4-seconds before) — STD of Third derivate (2-seconds before) — STD of Third derivate (event)



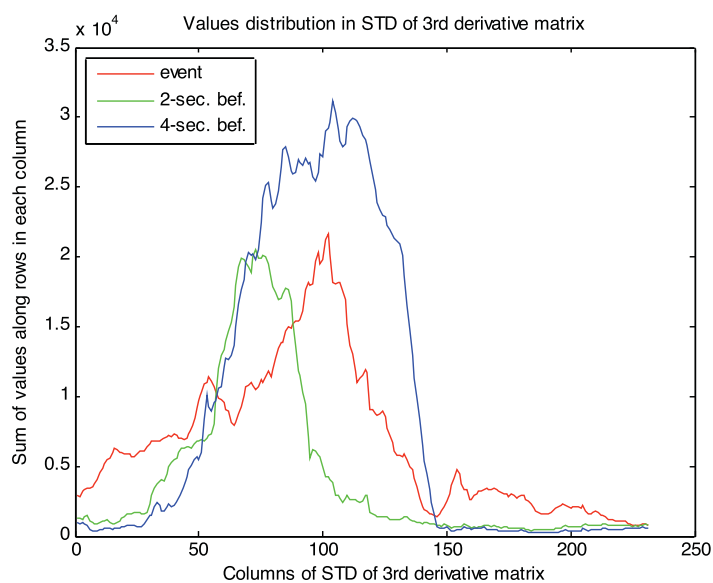Values distribution in STD of 3rd derivative matrix

In general, distribution of STD of jerk reaches higher values along the columns of the image from the event than in those from second-before sequences.

The peak corresponds to the second image (green line), where a bright white area is concentrated in the middle of the figure. This may seem strange given that in the pre-filtering process over-bright images have been removed. By reviewing the video it's checked that this area corresponds to a movement of the driver, who moves the arm from the steering wheel to the mouth.



Jerk of average speeds (2nd derivative of O.F.)
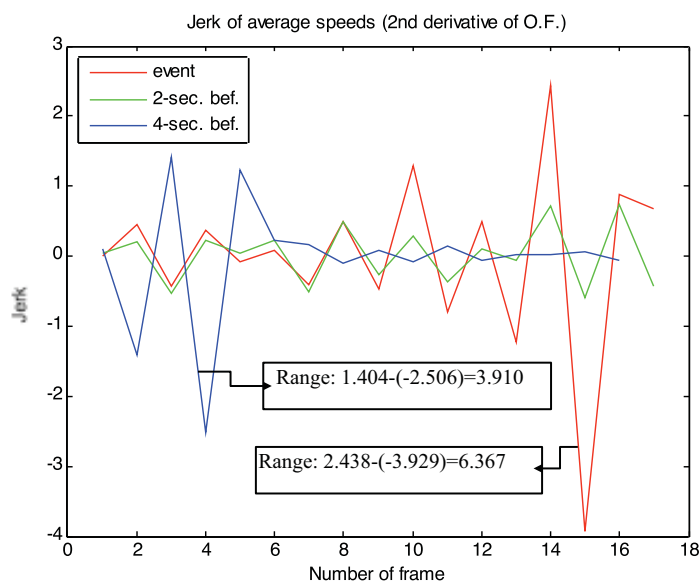
Range: 4.03-(-3.968)=7.998
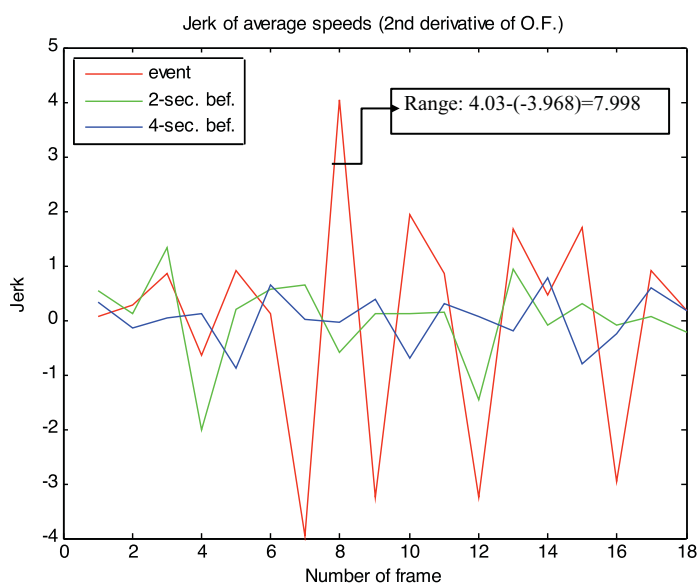
The jerk distribution of the event is characterized by several peaks, probably due to the excessive reaction of the driver. It also can be seen that the movement of the arm previously mentioned in the 2-second-before sequence is not significant by looking at jerk values in the graph.

STD of Third derivate (4-seconds before)    STD of Third derivate (2-seconds before)    STD of Third derivate (event)



Values distribution in STD of 3rd derivative matrix

As shown in the figure, distributions of sum of values from STD images are characterized with local peaks due to movements of the arm in second-before sequences.

Therefore, the mean value in 4-sec. bef. curve is higher than during the event, although it ranges greater sums before and after its peak. This happens because white areas are more dispersed in the image to draw driver's silhouette.



Jerk of average speeds (2nd derivative of O.F.)

Range: 1.404-(-2.506)=3.910

Range: 2.438-(-3.929)=6.367

The jerk distribution is characterized by two peaks, one reached during the event and another due to the change of position discussed before. The range of the last one is closer to other drivers' ranges in true events. This makes unclear at what value of range to establish the difference between a mere change of position and a true reaction in the driver.