3
Real-World Data Collection with "UYANIK"

Hüseyin Abut, Hakan Erdoğan, Aytül Erçil, Baran Çürüklü, Hakkı Can Koman, Fatih Taş, Ali Özgür Argunşah, Serhan Coşar, Batu Akan, Harun Karabalkan, Emrecan Çökelek, Rahmi Fıçıcı, Volkan Sezer, Serhan Danış, Mehmet Karaca, Mehmet Abbak, Mustafa Gökhan Uzunbaş, Kayhan Eritmen, Mümin Imamoğlu, and Çağatay Karabat

Abstract In this chapter, we present data collection activities and preliminary research findings from the real-world database collected with "UYANIK," a passenger car instrumented with several sensors, CAN-Bus data logger, cameras, microphones, data acquisitions systems, computers, and support systems. Within the shared frameworks of Drive-Safe Consortium (Turkey) and the NEDO (Japan) International Collaborative Research on Driving Behavior Signal Processing, close to 16 TB of driver behavior, vehicular, and road data have been collected from more than 100 drivers on a 25 km route consisting of both city roads and The Trans-European Motorway (TEM) in Istanbul, Turkey. Challenge of collecting data in a metropolis with around 12 million people and famous with extremely limited infrastructure yet driving behavior defying all rules and regulations bordering madness could not be "painless." Both the experience gained and the preliminary results from still on-going studies using the database are very encouraging and give comfort.

Keywords Drive-Safe · NEDO · UYANIK · Data collection · Istanbul route · Brake pedal pressure sensor · Gas pedal pressure sensor · Laser ranger finder · EEG · CAN-Bus · IVMCTool · Driver modeling · Distraction · Fatigue · Safety · Abnormal driving · Reference driving · On-line banking · Navigational dialog · Road-sign reading · Driver verification · Speech recognition · Multi-classifier · Facial feature tracking · 3D head tracking

3.1 Introduction

Throughout the world, driver error has been blamed as the primary cause for approximately 80% of traffic accidents. For instance, in 2005 there were more than 3,200 fatalities and 135,000 plus bodily injuries in over 570,000 traffic

H. Abut (✉)
San Diego State University (Emeritus), San Diego, CA, 92182, USA
e-mail: abut@anadolu.sdsu.edu

K. Takeda et al. (eds.), *In-Vehicle Corpus and Signal Processing for Driver Behavior*, DOI 10.1007/978-0-387-79582-9_3,

accidents in Turkey according to the Turkish Traffic Education and Research Directorate. Furthermore, it has been estimated that the long-term economic loss emanating from these accidents is over US $6 billion, which puts a rather significant burden on national budget. Albeit some variations, statistics from many other countries spanning the globe are very similar.

In 2004, Drive-Safe (DSC) Consortium, an academia and industry research partnership has been established in Turkey to create conditions for prudent driving on highways and roadways with the purposes of reducing accidents caused by driver behavior and to offer more amenities [1]. The objectives of this program consist of but are not restricted to

- DSP Technologies for driver identification based on multi-sensory behavioral and analog driving signals
- Driver modeling for personalization of driving environment
- Framework for automated detection of abnormal driving behavior from multiple sensors (audio, video, and driving signals)
- Development of technologies for "passive" assistance to mitigate abnormal driving and technologies for "active" assistance to reduce abnormal driving
- Ad hoc vehicle-to-vehicle communication for driving assistance

To achieve its objectives, collection of critical real-world data from multiple sensors (cameras, microphones, CAN-Bus, and other means) is required to build a coherent database on driver behavior, vehicle, and road conditions.

It is not difficult to guess characteristics in driving behavior differ from country to country according to their cultural and social backgrounds. Under the flagship of Nagoya University, an international alliance has been established in 2005 with several research groups in the United States, Japan, Turkey, Italy, and Singapore to share the worldwide driving, road, and vehicle-specific data obtained from 450 drivers (NEDO Alliance). Toward that end, three vehicles—"UYANIK" in Istanbul, "Nagoya Vehicle" in Japan (see Chapter 4), and "UT-Drive" in Dallas, Texas, USA (see Chapter 5)—have been equipped in cooperation with each other [2,3]. Data collection is underway in the three countries with potential of an ISO application in this area [4].

In this chapter, we will be focusing on the data collection effort with "UYANIK" and the preliminary findings by a number of members of the DSC in Turkey, who are also participating in the NEDO Alliance. The chapter is organized as follows. Section 3.2 discusses the instrumented vehicle, joint requirements of DSC/NEDO, sensors, and data acquisition systems. The selected data collection route, tasks, driver profile, and some challenges faced—both resolved and un-resolved "pains"—are addressed in Section 3.3. Signal samples are presented in Section 3.4. Next, research activities of Drive-Safe Consortium using collected data are highlighted in Section 3.5 as "Gains Are Coming." In continuing, preliminary findings from the student projects are mentioned in Section 3.6. Finally, Section 3.7 concludes the planned study with future work.

3.2 "Uyanik" and Sensors

Vehicle: As depicted in Fig. 3.1, "UYANIK,"—awake—is a Renault sedan donated to the consortium by OYAK-Renault of Turkey after many task-specific retrofitting done in the factory to serve the needs of this project. Modifications include special fortified front bumper, high-power battery, 1500 W DC–AC converter, CAN-Bus output socket, navigator seat and instrument bench, and re-wiring for power and signaling.

Sensors: The complete system layout for sensors, data acquisition systems, and wiring is shown in Fig. 3.2. There are two sets of cameras—one set for daylight and another for night vision—configured to give facial shots of the driver from both sides and a third one pointed to the road ahead as seen in Fig. 3.3. Four audio recordings are made by a lapel microphone or a headset, and two microphones on the rearview mirror and at the back of the head rest on the driver seat to capture the chamber noise and the conversation with the navigator. The human navigator is normally silent and handles the recording process except assistance when needed. Fourth microphone is the audio recording made from the microphone of the mobile phone and placed on the chest of the driver used for hands-free dialog.

An 180° laser range finder reads 181 *x*–*y* distances of the objects/vehicles in its range. Brake and gas pedal pressure readings from sensors connected to the pedals, several sensor outputs from the CAN-Bus socket of the vehicle, acceleration in *xyz* directions by an IMU device, location information from a GPS receiver are recorded at the sampling rate of the CAN-Bus. A 20-channel portable EEG subsystem is available for ground truth experiments.

Joint Requirements: In the data collection phase, the DSC/NEDO teams have come with the following desirable data set:

(1) **Three channels of uncompressed video:** They are captured (left and right view of the driver and the road ahead) at 30 frames per second. This corresponds to 140–200 GigaBytes (GB) of data per driver for frame-accurate audio–visual face tracking and recognition tasks. For Nedo Alliance applications, only two channels of MPEG-compressed video at 3.0 Mbits/s are recorded.

Fig. 3.1 "UYANIK" data collection vehicle

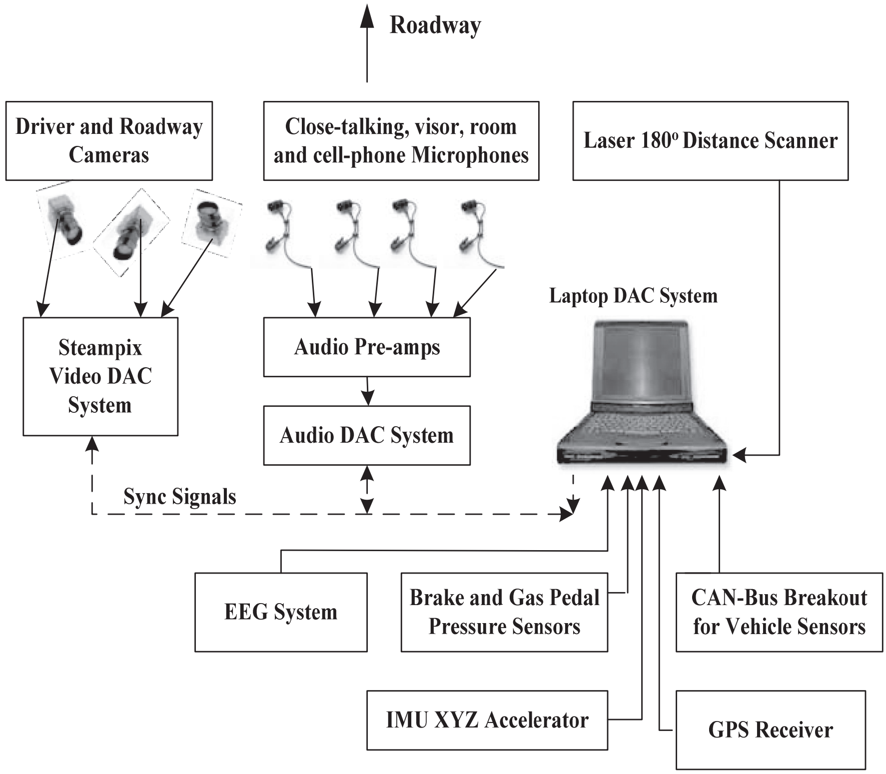

Fig. 3.2 System layout for sensors, data acquisition systems, and signal paths

Fig. 3.3 Vehicle sensors: cameras, navigator area and instruments bench, laser range finder, 3D accelerator, brake pressure sensor, microphone, and EEG cap (from *top left clockwise*.)

(2) **Three audio recordings**: Lapel/headset, rearview mirror, chamber noise, and the dialog over the mobile phone. They are digitized at 16,000 samples per second with 16-bit resolution either in raw format or in wav format.

Vehicle speed, engine RPM, steering wheel angle, head distance, location, EEG (only for grown-truth experiments), brake pedal, and gas pedal status readings are to be recorded not more than 1.0 kHz sampling rate.

(3) **CAN-Bus readings:** Even though most manufacturers claim that their unit complies with the standards the formatting (data packaging) is proprietary. Brake pedal status (pressed/idle) and gas pedal engagement percent are to be recorded together with the speed and RPM at the rate permitted by the manufacturer. Renault Megane is designed to sample at either 10 or 32 Hz.
(4) **Pedal sensors:** Brake pedal and the gas pedal pressure sensor readings are digitized at the CAN-Bus sampling rate by an independent two-channel A/D. They are bundled with CAN-Bus signals and recorded in a laptop computer.
(5) **Site-specific sensors:** Different from Nagoya and Dallas data collection sites, UYANIK was equipped with a laser distance measuring device in the front bumper and also with an IMU XYZ Accelerator measuring sensor set-up. In addition, a number of drivers have used the vehicle with an EEG device to form the ground truth for experiments.

Data Acquisition Systems: Data are collected by three acquisition systems synchronized with each other.

Video Acquisition: Uncompressed digital recordings of video from three cameras with 30 frames per second and a frame size of 640 × 480 were achieved by a semi-custom StreamPix digital video recorder from NORPIX. Each video channel is recorded into a separate 750 GB HD with a separate firewire connection and the total HD budget per driver is about 150 GB. At the end of each week, the data are archived into HQ archiving tapes.

Audio Acquisition: Alesis ADAT HD24 Data Acquisition System is used for audio recordings. Four microphone channels and a sync signal between the two acquisition systems were sampled at 48 kHZ and 24 bits per sample. Later these are converted to 16 kHZ and 16 bits off-line in wav format.

Acquisition of CAN-Bus signals, laser range finder, GPS receiver, brake pedal sensor, and IMU XYZ Accelerator was realized over USB and RS232 PCMCIA ports of a notebook computer using a custom software developed by two of the authors listed above from the Mekar Labs and Autocom Center at the Technical University (ITU) of Istanbul—a DSC partner.

Data Acquired on a Notebook Computer: Engine speed, RPM, steering wheel angle, brake pedal status, percent gas pedal position, brake pedal pressure, gas pedal pressure, clutch status, rear gear engagement status, and individual tire speeds are recorded at 32 samples per second. It is worth noting that no more than 10 Hz sampling rate was possible at the beginning of experiments and the data sets for some drivers need to be re-converted to 32 Hz for uniformity. In addition, the laser range finder readings (1 per second), IMU readings, and location information are recorded by this notebook. Undependable behavior of the GPS receiver was overcome by a second location tracking system donated by Satko, Inc. of Istanbul, a DSC sponsor. This alternate location information is tracked and recorded at the base control center located at the ITU campus.

3.3 Tasks and Pains

Route: Data collection starts and ends at the OTAM Research Center in the ITU Campus in Ayazağa, where UYANIK was housed. The navigator and the server for data storage are also located there. The data collection route is little over 25 km. It consists of a short ride inside the campus, followed by two 1.5 km very busy city thoroughfare sections, where a major construction is taking place to build a metro station. TEM Motorway toward the airport is the next segment. The route exits the first exit and makes a U-turn and travels toward the FSM Bridge. Highway driving ends at the Etiler exit and the rest of the route goes through city streets in Etiler, Akatlar, Levent, 4.Levent, Ayazağa, and back to OTAM at ITU campus via the ever-busy Büyükdere Caddesi (Fig. 3.4).

Data collection in this route has been a major challenge in a city of 12 million, famous with extremely limited infrastructure, drivers defying all rules and regulations, and the complete lack of courtesy to other vehicles and pedestrians around. Hence, the driving experience can be best portrayed as an art bordering madness.

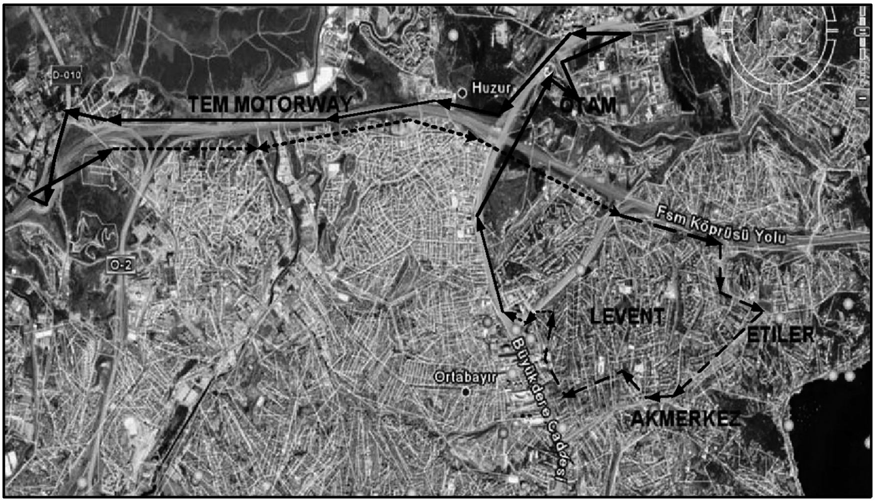

Fig. 3.4 DSC/NEDO route for Istanbul data collection. (Legend: *Solid lines*: free driving and radio tuning; *dotted lines*: on-line banking using ASR; *dashed lines*: Road/building signs, navigational dialog with human operator, and license-plate reading

Data Collection Tasks: As marked on the route map, there are four primary tasks each driver goes through:

(1) *Reference driving*: Here the driver gets used to the vehicle, the route, and the tasks. Most drivers turn on the radio and they tune to their favorite station. It was planned to have an ASR-based music query in this segment. A homegrown package was experimented but the results were not satisfactory.
(2) *Query dialog*: In this segment, each driver performs on-line banking using the cell phone mounted on the dashboard, which is programmed for speed dialing. He/she queries market prices of several national stocks using the on-line ASR-based service of a national bank. Here is a synopsis of the dialog:

ASR:	*Please tell the full name of the stock after beep.*
Driver:	*Arcelik*
ASR:	*6 YTL & 58*. (If successful, female voice)
ASR:	*I did not understand please repeat after beep*. (If unsuccessful, male voice)

(3) *Signboard reading and navigational dialog*: The driver announces the road and other signs posted on boards and buildings. At Etiler exit on TEM, base station is dialed and the next phase of the route is verified. Sign reading continues for about 2 km together with the license plates of the vehicles around. Both audio signals from the input and the speaker of the cell phone are recorded.
(4) *Pure navigational dialog*: After completing the very busy segment of the road in Etiler, the driver frequently contacts the navigator and conducts a human-to-human dialog.

Upon completion of the final segment of the route on Büyükdere Caddesi (again free driving), the experiment ends in front of the OTAM where the driver completes a couple of forms and a questionnaire.

Driver Profile: In Istanbul data collection effort, 108 drivers (100 were required) have driven the vehicle in the 25 km route, 19 of them were female and the remaining 89 male. The age range for female drivers was 21–48, and the corresponding male range was 22–61. Driver population was mostly pulled from the academic partners of the DSC together with their family and friends. However, due to equipment malfunction and a major disk crash affected data from several drivers. This brings the usable driver size to 101 in total.

Challenges Faced: As it was mentioned earlier, data collection on live traffic with human subjects is a major challenge by itself, which should not be seen as a natural extension of experiments conducted with simulators in a controlled lab environment. They are affected by equipment failures, technical difficulties, climate and weather, traffic density and type, driving local culture, effectiveness of law enforcement, and probably the most critical ones are the driver's physical and mental conditions and his/her driving behavior. In this undertaking, there were challenges in each area and many of them came all at once. Some examples with their current status/solution in parenthesis are

- Digital recording of uncompressed video at 30 frames/s with 640 $\times$ 480 resolution. (Solution: three-channel StreamPix system.)
- Synchronization of three different data logging systems. (Solution: common system clock and sharing a sync signal.)
- Data store archives approximately 16 TB of data from 108 drivers and their backup into tapes. (Solution: employment of a full-time engineer and part-time students to archive data into HQ digital back tapes on an HP backup subsystem. Each experiment of 40–50 minutes has required 4–5 hours to access, align, and write into a tape.)
- Off-line compression of data into MPEG4 format and alignment of various sensor readings. (Solution: custom-designed software package for aligning, compressing, and browsing/re-recording.)
- CAN-Bus data readings. (Solution: special software package.)
- CAN-Bus data cannot be read at a programmable rate, it fluctuates around, either 10 or 32 Hz. Earlier experiments were done at about 10 Hz and the final data sets were collected at 32 Hz. (Solution: rate for each run is recorded and re-sampling is done subsequently by users of the database.)
- Unacceptably noisy signal from brake/gas sensor pedal pressure. (Solution: semi-custom-designed A/D for these signals.)
- More reliable location information (unresolved).
- ASR for Music Search, i.e., query (under study).
- Complaints from drivers using pedal with sensors mounted. (Solution: this is mitigated by opening the gap between pedals at the Renault manufacturing plant and advising drivers to use light-weight shoes.)
- Complaints from drivers in “multi-tasking” and losing attentiveness. (Solution: do on-line banking on the curb side or earlier inside ITU campus if they feel losing control.)
- Wrong turns from unfamiliarity with the route, which results in significant time loss and the stored data size. (Solution: emergency assistance by the navigator and frequent assistance from the control center at OTAM via cell phone.)

3.4 Signal Samples

A screen shot from a male driver is shown in Fig. 3.5. Here, video feeds from two cameras mounted in the vehicle pointing at the driver and a third one toward the road showing a residential neighborhood along the route are displayed together with four speech waveforms down-sampled to 16 kbits/s and 16-bit resolution (recorded in wav format). Driving signals fed from the CAN-Bus are also displayed. Video feeds could be either uncompressed or compressed with MPEG4 at 3 MB/s avi format using DivX codec.

Fig. 3.5 Browser synopsis of the multi-channel multimedia display for Driver IF1014. Three video feeds and four audio waveforms together with steering wheel angle, vehicle speed, and engine RPM dials are included

As can be seen from Fig. 3.5, steering wheel angle in degrees, vehicle speed in km/h, and the engine RPM are illustrated. There are four visible dots on the needle of speed dial, which represent the actual tire speeds. Normally, they move together. However, they register different values during skids and slides in rainy weather or inclined surface, and sudden brakes.

Brake and gas pedal pressures are displayed in Fig. 3.6. In addition, the vehicle speed and the steering wheel angle (sampled at 34 Hz) are shown in Fig. 3.7 for an interval of 640 seconds from a test run. Vehicle-related signals recorded from CAN-Bus data are also synchronized with video but the re-sampled versions are not recorded, i.e., originals are kept. There are few other readings from the CAN-Bus reporting the status of clutch, rear gear, and brake pedal.

One hundred and eighty one distance measurements between UYANIK and other vehicles/objects around are measured by a laser ranger finder at the rate of 1 Hz and Fig. 3.8 shows both the plot and the actual photo at that instant, which is explicitly marked.

In addition, steering wheel angle velocity, yaw rate, clutch status, and rear gear status readings are recorded from the CAN-Bus.[1] IMU readings showing

[1] These readings are information-rich for projects carried out with active/passive vehicle control and avoidance systems.

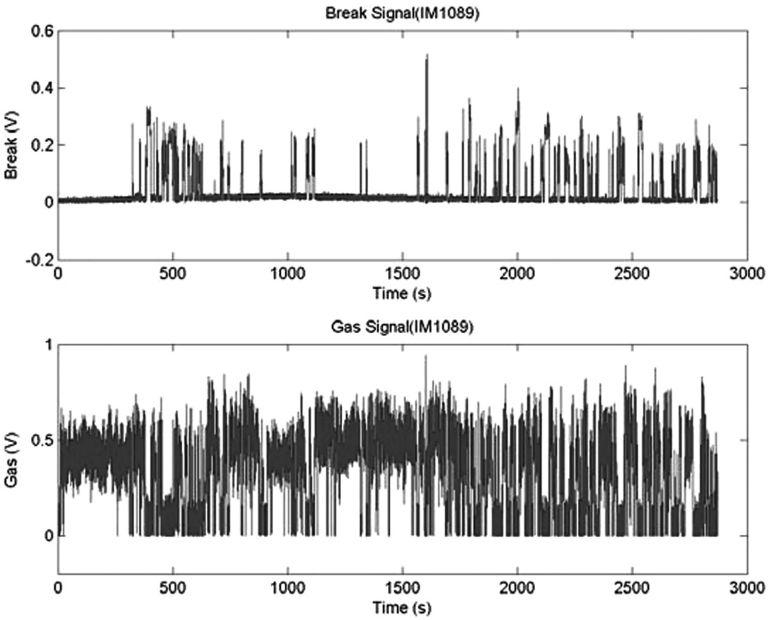

Fig. 3.6 Brake and gas pedal pressure readings

the *xyz* directional accelerations and the location information from GPS receiver are recorded into separate files. A number of application-specific plots are illustrated in several related works briefly discussed in Section 3.6.

Finally, we have recorded in a number of experiments 16-channel EEG signals to act as ground-truth for fatigue and attention/inattention determination. For the driver shown in Fig. 3.9, plots from eight channels—specifically, C3, C4, F3, F4, O1, O2, A1, and A2, which are important for fatigue prediction—are shown in Fig. 3.9. On the other hand, EOG channels are recorded for the estimation and the rejection of eye blinks. ECG and EMG are also used for both artifact rejection and fatigue estimation. Alpha signal power and signal power ratio between right and left hemispheres are used for fatigue prediction from EEG measurements after rejection of ECG, EMG, and EOG artifacts, the importance of these are currently under study. For fatigue and distraction understanding, long-run simulator experiments (3–4 hours long or more) with/without EEG are in progress.

3.5 Gains are Coming: Part 1

Using the database generated during the data collection process, the preliminary findings from three on-going projects carried by scholars and their students at Sabancı University in Istanbul, Turkey, will be briefly discussed in this and the following sections.

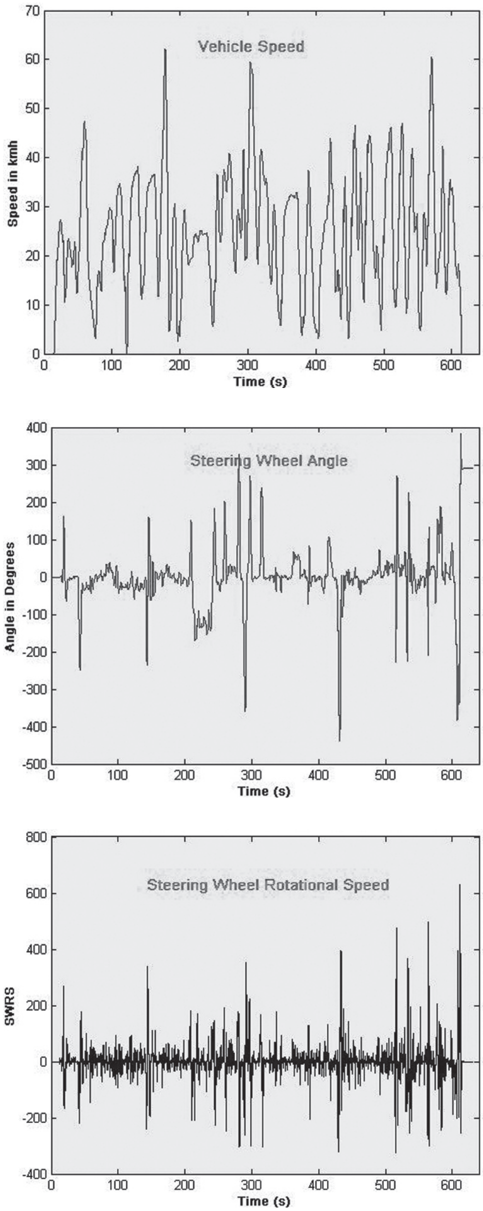

Fig. 3.7 Vehicle speed, engine, steering wheel, angle, and steering wheel rotational speed plots are for the first 380–560 seconds of one test

3.5.1 Audio–Visual Speech Recognition in Vehicular Noise Using a Multi-classifier Approach by H. Karabalkan and H. Erdoğan

Speech recognition accuracy can be increased and noise robustness can be improved by taking advantage of the visual speech information acquired from

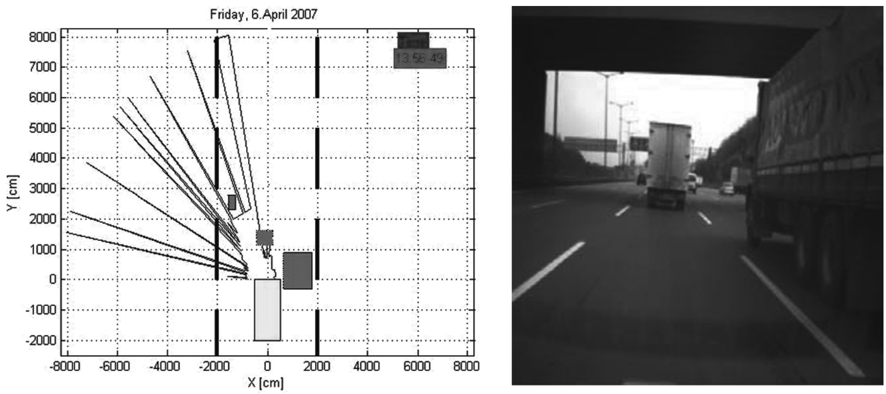

Fig. 3.8 Laser range finder readings and the view of Driver IF1015. Truck on the right is between −200 to +800 cm relative to UYANIK (overlapping by 2 m and another 8 m ahead and the white truck in front is 12 m ahead)

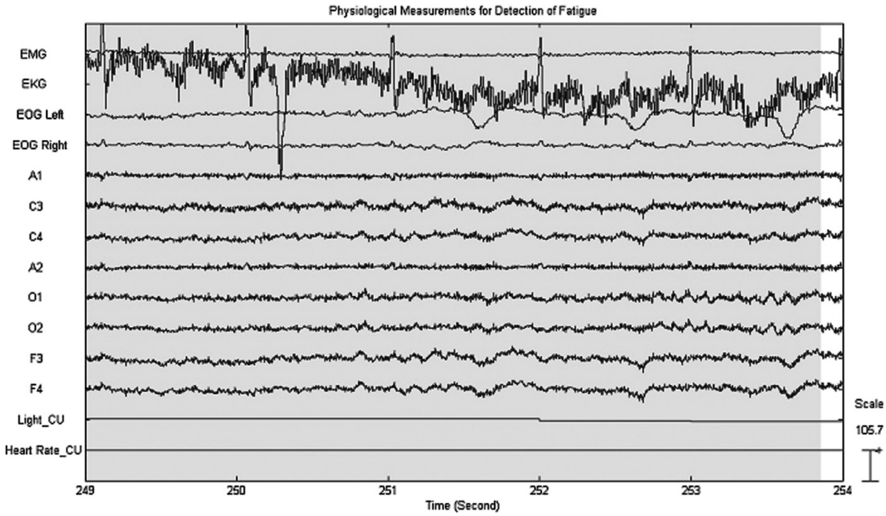

Fig. 3.9 Eight channels of the EEG recordings for the driver in Fig. 3.3

the lip region. To combine audio and visual information sources, efficient information fusion techniques are required. In this paper, we propose a novel SVM–HMM tandem hybrid feature extraction and combination method for an audio–visual speech recognition system. From each stream, multiple one-versus-rest support vector machine (SVM) binary classifiers are trained where each word is considered as a class in a limited-Vocabulary speech recognition scenario.

The outputs of the binary classifiers are treated as a vector of features to be combined with the vector from the other stream and new combining binary classifiers are built. The outputs of the classifiers are used as observed features in hidden Markov models (HMM) representing words.

The complete process can be considered as a nonlinear feature dimension reduction system which extracts highly discriminatory features from limited amounts of training data. To simulate the performance of the system in a real-world environment, we add vehicular noise at different SNRs to speech data and perform extensive experiments.

In the approach reported here, 12-dimensional cepstral coefficients and the energy in the window are extracted to obtain static MFCC coefficients. Then, the feature vector is extended to a 39-dimensional vector by taking the first and second time differences of static coefficients. The audio feature vector is passed through a multiple binary classifier structure, as shown in Fig. 3.10, to obtain an 11-dimensional audio feature vector which is more discriminative and noise tolerant. The dimension is 11 since there are 10 different words uttered in the database and an additional class is considered for silence.

To extract visual features, principal component analysis (PCA) is applied to the lip region which returns the eigenlips. Only the 30 most significant eigenlips are considered to come up with 30-dimensional visual feature vector. This reduces dimensionality as well as speaker independence. The 30-dimensional visual feature is also passed through a multiple binary classifier structure as in the case of audio features, and finally an 11-dimensional visual feature vector is obtained.

Multiple paralleled binary classifier structure is used for combining audio and visual information, too. The audio feature vector and the visual feature vector, which are both 11-dimensional, are concatenated to form a 22-dimensional audio–visual feature vector and this feature vector is passed through the multiple binary classifier structure. The resulting 11-dimensional audio–visual

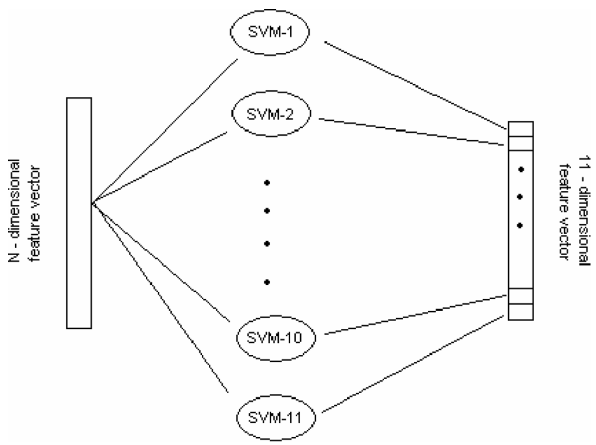

Fig. 3.10 Multiple parallel binary classifier structure

feature vector is the observation for HMM. The proposed fusion technique is planned to be compared with linear discriminant analysis (LDA). The experiments for SVM-HMM tandem hybrid approach are still on-going but the method promises good accuracy rates throughout different SNR values [5].

3.5.2 Graphical Model-Based Facial Feature Point Tracking in a Vehicle Environment by S. Coşar

Facial feature point tracking is an important step in problems such as video-based facial expression analysis, human–computer interaction (HCI), and fatigue detection. Generally, such analysis systems consist of three components: feature detection, feature tracking, and expression recognition. Feature detection involves detecting some distinguishable points that can define the movement of facial components. This may involve detection of eyes, eye brows, mouth, or feature points of these components. Next is the tracking part which consists of tracking the detected feature points. Finally, according to tracking results of these feature points, the recognition component outputs results such as happy, sad, or tired.

For feature point tracking, roughly there are two classes of methods in literature: general purpose approaches and face-specific approaches. Generally, feature point tracking is done by using a temporal model that is based on pixel values. Consequently, these methods are sensitive to illumination and pose changes, and ignore the spatial relationships between feature points. This affects the tracking performance adversely, causes drifts and physically unreasonable results when the data are noisy or uncertain due to occlusions.

In this work, feature point tracking is performed in a statistical framework that incorporates not only temporal information about feature point movements but also information about the spatial relationships between such points. This framework is based on graphical models that have recently been used in many computer vision problems. The model is based on a parametric model in which the probability densities involved are Gaussian. The parametric nature of the models makes the method computationally efficient. Spatial connections between points allow the tracking to continue reasonably well by exploiting the information from neighboring points, even if a point disappears from the scene or cannot be observed. Feature values from video sequences are based on Gabor filters. Filters are used in a way to detect the edge information in the image, to be sensitive to different poses, orientations, and feature sizes. Based on this model, an algorithm that achieves feature point tracking through a video observation sequence is implemented. The current method is applied on 2D gray scale real video sequences taken in the vehicle environment, UYANIK, and the superiority of this approach over existing techniques is demonstrated in Figs. 3.11 and 3.12.

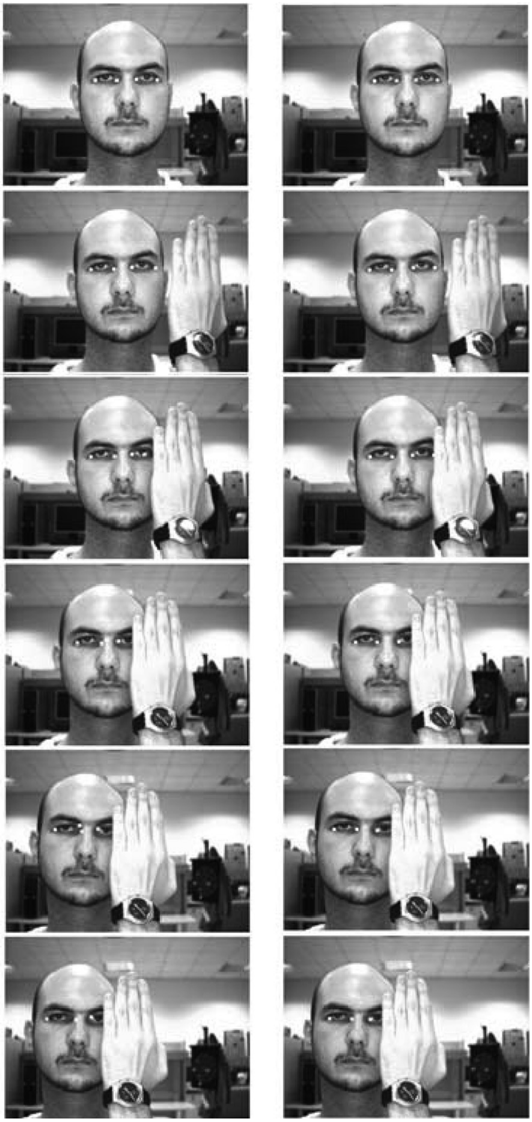

Fig. 3.11 Ideal environment tracking results from earlier methods (*left*) and the proposed method (*right*) sample shots from a laboratory setting

3.5.3 3D Head Tracking Using Normal Flow Constraints in a Vehicle Environment by B. Akan

Head tracking is a key component in applications such as human–computer–interaction, person monitoring, driver monitoring, video conferencing, and

Fig. 3.12 Driver head tracking based on an existing method (*left*) and the proposed method (*right*). Video captured from UYANIK

object-based compression [6]. The motion of the head of a driver head can tell a lot about his/her mental state, e.g., whether he/she is drowsy, alert, aggressive, comfortable, tense, or distracted. This chapter reviews an optical flow-based method to track the head pose, both orientation and position, of a person and presents results from real-world data recorded in a car environment.

Driver behavior modeling and fatigue detection is an important feature in developing new driver assistance systems and smart cars. These intelligent vehicles are intended to be able to warn or activate other safety measures when hazardous situations have been detected such as fatigued or drunk driver, so that a system can be developed to actively control the driver before he/she becomes too drowsy, tired, or distracted. The pose of the head can reveal numerous clues about alertness, drowsiness, or whether the driver is comfortable or not. Furthermore knowing the pose of the head will provide a basis for robust facial feature extraction and feature point tracking.

In this, study, we propose a method for tracking the driver's head using normal flow constraint (NFC) which is an extension of the original optical flow algorithm. Optical flow is the 2D vector field which is the projection of the 3D motion onto an image plane. It is often required to use complex 3D models or nonlinear estimation techniques to recover the 3D motion when depth information is not available. However, when such observations are available from devices such as laser range finders (laser scanners) or stereo cameras, 3D rigid body motion can be estimated using linear estimation techniques [7].

We have tested the algorithm in a real car environment. A bumblebee stereo camera system has been used for data acquisition. The camera hardware analyzes the stereo images and establishes correspondence between pixels in each image. Based on the camera's geometry and the correspondences between pixels in the images, it is possible to determine the distance to points in the scene. Without any special optimizations the tracker can update pose estimations based on 2,000–3,000 pixels per frame at a rate of 60 Hz on a Celeron 1.5 GHz laptop.

Performance of the tracker has been tested using the data collected from "UYANIK." Several sequences of length 500 frames or roughly 30s of video with both intensity and disparity images have been recorded. The sequences involve all natural head movements: throughout the video the driver rotates his head checking out left, right, and rear mirrors of the car and looks down at the gear. Some outputs from the tracking algorithm can be seen in Fig. 3.13.

Fig. 3.13 Results of the driver head tracker at frames 0, 60, 110, 150, 230, 400

3.6 Gains are Coming: Part 2

Along with the three applications in the previous section, two additional projects are under study by several students at Sabancı University, Istanbul, Turkey. Because, these studies are still in progress, we will briefly identify these problems and present preliminary findings.

3.6.1 Pedal Engagement Behavior of Drivers by M. Karaca, M. Abbak, and M.G. Uzunbaş

In Fig. 3.14, we illustrate a sample of brake pedal and the gas pedal engagement status for a driver and the correlation between these two waveforms, we obtain zero-crossings per minute (zero-crossing rate) indicating the transitions in the

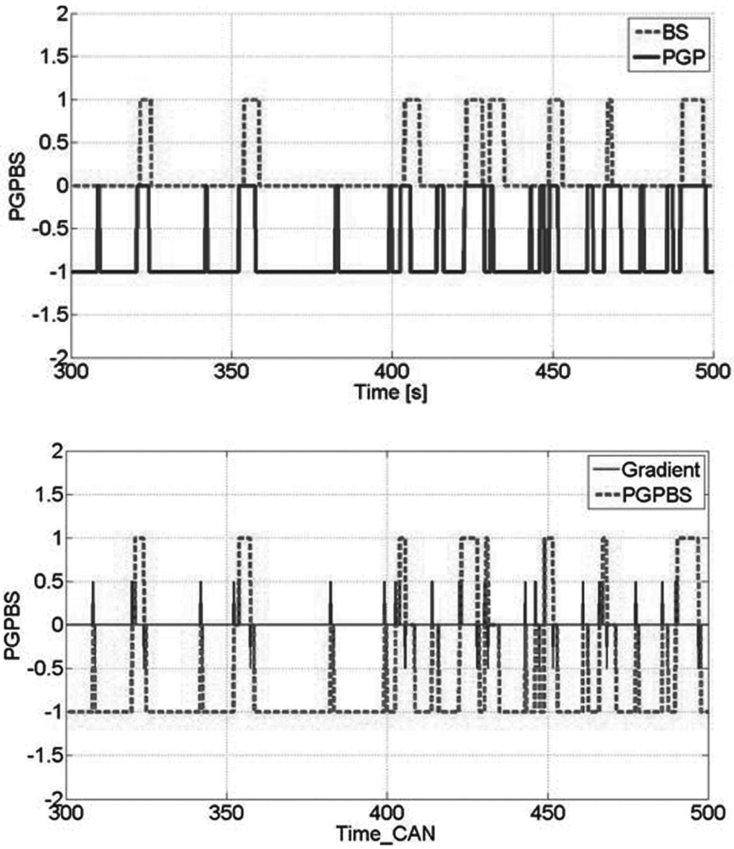

Fig. 3.14 Brake and gas pedal status plots in a 200-s interval

pedal engagement from brake-to-gas and vice versa (pedal shift). Also, the head distance between the test vehicle and other vehicles ahead is depicted in Fig. 3.8.[2]

This multi-sensory information is explored for understanding the gender-specific behavior of drivers to the traffic ahead by obtaining the statistics of intra-/inter-gender zero-crossings per minute rates for several male and female drivers in the database. Interesting patterns have been observed but not ready to interpret them until more extensive studies are done. However, one local folk theorem is proven.

The average inter-gender pedal shifts per minute at various speeds are shown in Fig. 3.15. It is clearly observable that the female drivers in Turkey are driving more smoothly across all speeds when compared to their male counterparts. Furthermore, male drivers are very impatient and make frequent brake/gas shifts at a wide range of speeds of 40–80 km/h. Should the male drivers in Turkey leave the driving to ladies?

3.6.2 Speaker Verification and Fingerprint Recognition by K. Eritmen, M. Imamoğlu, and Ç. Karabat

In this application, drivers use their fingerprint and speak their names to start the vehicle. The purpose of this application is twofold: (a) access/deny to the vehicle with one physical and one behavioral signature—fingerprint and speech—for improved performance and (b) still access/deny in the case if only one of the sensory modes is available.

In biometrics, signal processing, and forensic communities, performance of biometric verification systems is measured by using receiver operating characteristics (ROC) curve which is the plot of false accept rate (FAR) versus false reject rate (FRR) for changing threshold values [5].

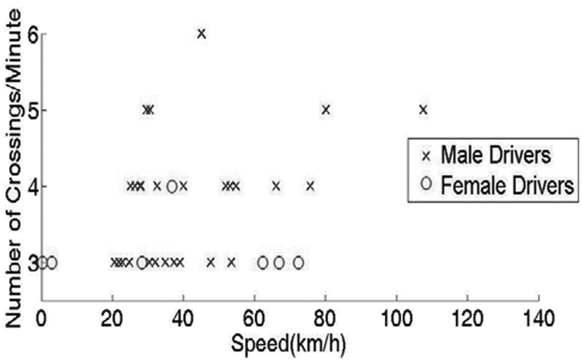

Fig. 3.15 Gas and brake pedal zero-crossing rates for female and male drivers

[2] Vehicle identification at a given time is done manually by studying the picture and the coordinates of the distances to the objects recorded by the laser scanner simultaneously.

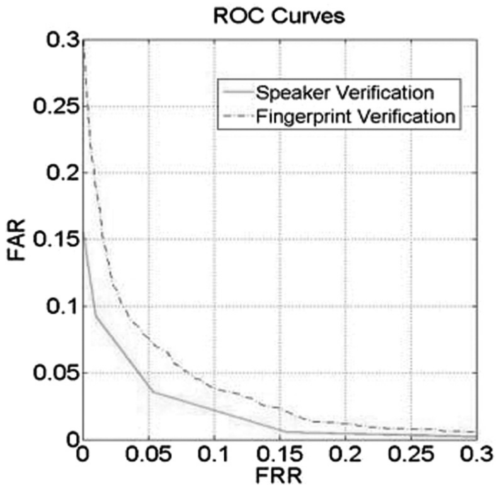

Fig. 3.16 Individual ROC curves for fingerprint and speaker verification before score-level fusion

In Fig. 3.16, we show the false acceptance (FAR) versus false reject (FRR) rates obtained for several drivers. In these experiments, we have used the familiar GMM-based speaker models obtained from MFCCs. Individual equal error rates are approximately 4–7% for these two signatures, which are very similar to results reported in the literature.

At the time of this writing, reliable and meaningful score-level classifier experiments were performed to fuse the results with the anticipation of improved performance as it was reported in many data fusion application classification and in person/object identification problems including the works by some of the authors [4,5].

3.7 Conclusions and Future Work

In this chapter, we report the progress on real-world data collection with "UYANIK" in Istanbul (Turkey) as part of an international collaboration research with Nagoya University (Japan) and the University of Texas in Dallas (USA) and in partnership with the Drive-Safe Consortium in Turkey. A total of 101 drivers participated in the experiment to date resulting at a data storage size of more than 16 TB with very rich information on driver behavior, the vehicle performance, and the road and traffic conditions on a 25.6 km route in Istanbul.

We will complement the road data with experiments carried on a recently installed full-fledged driving simulator with research focus on driver modeling, fatigue, and distraction detection. At the same time, we would transcribe and analyze cognitive task-loaded and free-driving segments to better understand the impact of certain tasks on drivers and to make technology recommendations for safer driving.

Acknowledgments This work is partially supported by the State Planning Organization of Turkey (DPT) under the umbrella initiative called "Drive-Safe Consortium," the NEDO collaborative grant titled "International Research Coordination of Driving Behavior Signal Processing Based on Large Scale Real World Database" from Japan, and the European Commission under grant FP6-2004-ACC-SSA-2 (SPICE).

The authors would like to acknowledge Mr. Hakan Tandoğdu, his team at Renault Manufacturing Plant in Bursa for retrofitting UYANIK, and his management for donating the vehicle. Turkcell, OPET, and Satko of Turkey have been sponsoring the data collection effort with goods and services.

We kindly appreciate the contributions and advice from Dr. Levent Guvenc and his students at ITU, Dr. Ali G. Göktan and Yunus Canlı at the OTAM Center of ITU in Turkey, and Dr. Engin Erzin of Koç University in Istanbul, and Dr. John H.L. Hansen and his team in Dallas. Without Dr. Kazuya Takeda of Nagoya University in Japan, the undertaking of this magnitude could not have happened. We are indebted to him and his colleagues in Nagoya.

References

1. A. Erçil, H. Abut, E. Erzin, E. Göçmençelebi, A. Göktan, L. Güvenç, E. Özatay, H. Tandoğdu, "The DriveSafe Project," Proceedings, Autocom Workshop on Preventive and Active Safety Systems for Road Vehicles, September, 19–20, 2005, Istanbul, Turkey.
2. C. Miyajima, T. Kusakawa, T. Nishino, N. Kitaoka, K. Itou, K. Takeda, "On-Going Data Collection of Driving Behavior Signals," Chapter 4 in this book.
3. P. Angkititrakul, J.H.L. Hansen, "UTDrive: The Smart Vehicle Project," Chapter 5 in this book.
4. H. Abut, J.H.L. Hansen, K. Takeda (Editors), *Advances for In-Vehicle and Mobile Systems: Challenges for International Standards*, Springer Science-Business, May 2007, ISBN: 0-387-33503-X.
5. E. Erzin, Y. Yemez, A. M. Tekalp, A. Erçil, H. Erdoğan, H. Abut, "Multimodal Person Recognition for Human–Vehicle Interaction," *IEEE Multi-Media Transactions*, April–-June 2006, pp. 18–31.
6. L.P. Morency, A. Rahimi, N. Checka, T. Darrell, "Fast Stereo-Based Head Tracking for Interactive Environments," Proceedings of Conference on Automatic Face and Gesture Recognition, 2002.
7. M. Harville, A. Rahimi, T. Darrell, G.G. Gordon, J. Woodfill, "3D Pose Tracking with Linear Depth and Brightness Constraints," International Conference on Computer Vision, pp. 206–213, 1999.

http://www.springer.com/978-0-387-79581-2

In-Vehicle Corpus and Signal Processing for Driver Behavior
Takeda, K.; Erdogan, H.; Hansen, J.; Abut, H. (Eds.)
2009, XIV, 286 p. 125 illus., Hardcover
ISBN: 978-0-387-79581-2