Baseline Face Detection, Head Pose Estimation, and Coarse Direction Detection for Facial Data in the SHRP2 Naturalistic Driving Study

J. Paone, D. Bolme, R. Ferrell, *Member, IEEE*, D. Aykac, and
T. Karnowski, *Member, IEEE*
Oak Ridge National Laboratory, Oak Ridge, TN

***Abstract*—Keeping a driver focused on the road is one of the most critical steps in insuring the safe operation of a vehicle. The Strategic Highway Research Program 2 (SHRP2) has over 3,100 recorded videos of volunteer drivers during a period of 2 years. This extensive naturalistic driving study (NDS) contains over one million hours of video and associated data that could aid safety researchers in understanding where the driver's attention is focused. Manual analysis of this data is infeasible; therefore efforts are underway to develop automated feature extraction algorithms to process and characterize the data. The real-world nature, volume, and acquisition conditions are unmatched in the transportation community, but there are also challenges because the data has relatively low resolution, high compression rates, and differing illumination conditions. A smaller dataset, the head pose validation study, is available which used the same recording equipment as SHRP2 but is more easily accessible with less privacy constraints. In this work we report initial head pose accuracy using commercial and open source face pose estimation algorithms on the head pose validation data set.**

I. Introduction

A LARGE component of the Strategic Highway Research Program (SHRP2) program is an extensive naturalistic driving study (NDS), which features video and other sensor recordings from over 3,100 volunteer drivers for up to 2 years each [1]. A custom data acquisition system (DAS) was developed by engineers at the Virginia Tech Transportation Institute (VTTI) [8] to capture and record the data, which has a resulting size of over 2 Petabytes. A large part of this data is video from the cameras in the DAS, which offers exciting possibilities for analysis and insight into the driver practices. Many areas of intelligent vehicle technology, such as driver alertness monitoring, predicting driver actions, improved warning systems, advanced driver assistance, and human machine interaction could all benefit from the NDS data. Indeed, the real-world nature, volume, and acquisition conditions are unmatched in the transportation community, but there are also challenges because the data has relatively low resolution, high compression rates, and differing illumination conditions. Furthermore, manual analysis of such a large collection can be slow, tedious, and error-prone. Consequently, the Federal Highway Administration (FHWA) is leading efforts to develop and deploy automated feature extraction algorithms to measure driver and surrounding vehicle behavior to increase the utility of the NDS [2]. The algorithms under development include tools to automatically extract data regarding the driver disposition, surrounding roadway information, passenger information, and general driving performance. A particularly useful indicator of driver state can be found by analysis of the face and head, which can indicate drowsiness, distraction, or other important characteristics. As a result, automating the analysis of the face and head are of primary importance. In this paper we discuss the performance of a variety of baseline methods on a SHRP2-like data set, including detection, pose estimation, and coarse direction of focus. This data set is publicly available from VTTI at no cost, although a simple data sharing agreement must be established.

II. Background

The processing of imagery to identify and characterize faces is an important application area in computer vision. The basic core methods of face detection including facial landmark detection serve as preliminary steps to more informative applications such as subject recognition, and subject state estimation such as emotional expression or drowsiness. These topics are highly researched with the state-of-the-art still improving at levels from basic detection and landmark identification [7, 23, 28] to higher level

Manuscript received January 23, 2015. This work was supported by the Federal Highway Agency under the Exploratory Advanced Research Program (Agreement Officer's Representative: Lincoln Cobb). Notice: This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

The authors are with the Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: karnowskitp@ornl.gov)

.

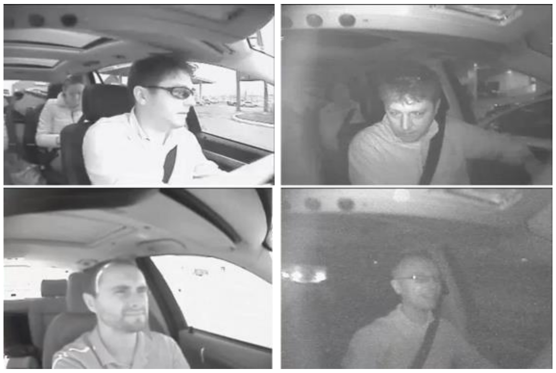

Figure 1. Example images from sample SHRP2 data during (left) daytime and (right) low-light conditions [5]

representations such as emotion [9]. There are also many excellent survey papers such as [26, 27] for face detection, [20] for head pose estimation, and [6, 15] for face recognition. Applications for these techniques range from social and commercial media knowledge discovery to security and surveillance. Transportation (specifically monitoring for distraction and interpreting driver behavior) is also an area of high interest [11, 16, 30]. Much research focuses on providing real-time feedback to the driver to alert them of dangerous situations [10, 22, 24], but there is also interest in video interpretation [19] which can help improve the analysis of face data for safety research [18].

III. METHODS

A. Data

As in virtually all computer vision applications, the state-of-the-art is usually advanced most significantly when data sets exist that can be used to develop and compare algorithms. For NDS work, there are a number of privacy constraints, which make free, full sharing of the data difficult. As an example the GPS location of trip start and ends cannot be shared as they could be used to identify the driver. Another example is the face video data, which also contains identifiable information. However, there is an alternate set of data, the "Head Pose Validation" (HPV) study, which is available to qualified researchers with less restrictive privacy agreements than the SHRP2 NDS [5]. This data was recorded by VTTI to serve as a means of measuring head pose in drivers and testing different methods for head pose estimation. Therefore, the HPV is an excellent source of data for computer vision researchers in naturalistic driving studies.

As a dedicated data collection system, the DAS interfaces to several sensors installed by VTTI and SHRP2 contractors as well as signals available on the controller area network (CAN) bus. There are five NTSC analog cameras in the system, which obtain images of the roadway to the front and rear of the vehicle, as well as the interior (hands/steering wheel and face). The DAS combines four of the camera images into a single compressed 720x480 pixel frame for storage. An additional cabin view camera is used, which blurs the image to obscure passengers in the vehicle. The cabin view snapshot is recorded at intervals of 10 minutes. The main subject for this paper is the face camera, which has a focal length of 3.3 mm, with a 51.7 by 40.0 field of view, and is scaled to 360x240 before recording to disk at 15 frames a second. In addition, the camera contains an IR pass filter to improve dynamic range in the different environmental conditions of real-world driving. To illuminate the driver at night, the DAS has a built-in IR light source. Some example images are shown in Figure 1.

A subset of the data, called the "clipped" set, is primarily available for automated algorithm development. The clipped set contains 41 participants, with 20 involved in "static" trials (where the vehicle was not driven) and 21 in "dynamic" trials (where the vehicle was driven in a fixed route for roughly 30 minutes.) A single vehicle was used for all trials. In the static tests, participants performed a variety of "tasks", such as putting on / taking off glasses, simulating a cell phone conversation, etc. In the dynamic tests some tasks were also performed but common driving activities (such as a merge action) were also identified for some frames. Data was recorded at night, during the day, and in transition periods from day to night. Two DAS systems were used to capture different, but related, video streams: one for lower-resolution imagery data compatible with the SHRP2 study, and a second which was modified to record only the face camera. This second "high resolution" video stream records the face camera at 720x480 pixels and does not contain the dash, front or rear camera video.

For some frames, a ground truth face detection and head pose are provided. These ground truth values were provided by facial landmarks manually annotated by a set of video reviewers on the high resolution DAS system, followed by post-processing. A small percentage of frames have the manually annotated landmarks but do not include rotation estimates. Of a total of 911,018 frames from 41 SHRP2 videos, roughly 7% have facial landmarks and rotation estimates. These frames occur often consecutively for several short periods during the video.

The rotation estimate data provided for the HPV study was based on the high resolution video taken of the participant. It was not performed on SHRP2 format images. In order to use this data with the SHRP2 images, the frame offsets between the high and low resolution video had to be found, using semi-automated image comparison followed by manual verification.

B. Baseline Algorithms

Geometric - The first method of interest was a geometric method used by VTTI during the initial collection and analysis of the head pose study data. The method uses seven manual face landmark annotations consisting of (1) outer corner of right eye, (2) inner corner of right eye, (3) inner corner of left eye, (4) outer corner of left eye, (5) tip of nose, (6) right corner of mouth, and (7) left corner of mouth. For

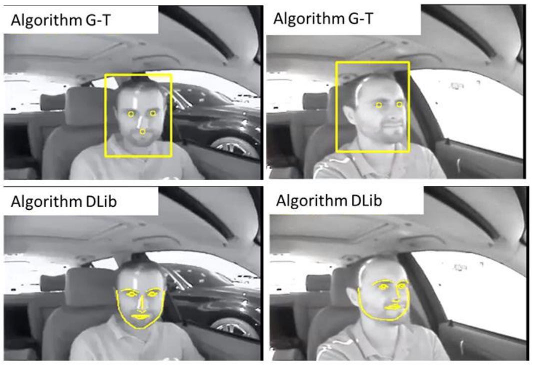

Figure 2. Examples of (top) GOTS landmarks with facial region-of-interest and (bottom) DLIB landmarks on SHRP2 sample data [5]

many frames at least two analysts made annotation marks. These analysts were trained according to a protocol to establish consistency. The landmarks were located in the high-resolution frame, but there is a constant scaling that can be used to locate them in the corresponding low-resolution (SHRP2) frame.

The resulting 7-point annotations were then converted to head pose measurements using the methods of [12] (pitch) and [14] (yaw and roll). In addition to these landmarks, the geometric method utilized facial feature measurements taken of each participant's face. Most frames have 2 sets of rotation estimates – a raw rotation estimate of the yaw, pitch and roll and an estimation of each of these values filtered with a 4th order Butterworth filter. On some frames the filtered values are not available. The geometric method is used in this work as the ground truth to compare to other methods.

GOTS and **GOTS Tracked (G-T)** – This method uses a particular "Government Off-The-Shelf" software which includes face detection, tracking, landmark detection, pose estimation, and face recognition that is well suited for non-frontal faces in uncontrolled lighting. (For this paper we will refer to it simply as GOTS for simplicity.) This particular tool has become a standard for evaluation due to its government use rights and performance. For this study we operated GOTS in two different modes: detection only and detection with tracking. In the detection only mode each frame is treated independently and GOTS operates like a standard face detector. In the detection with tracking mode, GOTS attempts to follow the face over time in order to improve detection rates using one of multiple tracking options (the "Serial Mode" was used in this study). GOTS typically provides 3 landmarks for faces that are approximately frontal including the two eyes and nose. For faces significantly non-frontal 5 points are provided which include the near eye, nose, and points near the cheek and ear. The algorithm also provides estimates of the head roll and yaw (but not pitch) and seems to be effective for a +/-90 degree yaw range centered on frontal.

DLib – This is an open source machine learning library [17] that provides face detection and landmark detection algorithms. Both of these algorithms appear to operate well on a variety of poses from +/-60 degrees yaw and pitch from frontal and under various lighting conditions. Performance of the detector appears to drop off significantly in the range from 60 to 90 degrees yaw. The software provides 72 landmark points.

DLib does not provide face pose or recognition capabilities so for pose estimation we implemented our own algorithm that is based on the face landmark locations. For training we used the Pointing04 dataset [13] which included 15 people with 2790 total images. The landmark location was clustered in a manner similar to *k*-means with the exception that the distances to the cluster centers are computed after aligning the landmarks to the cluster center using the "AffineFromPointsLS" function from the PyVision library [4]. Fifteen cluster centers were used. Once the cluster centers are determined the training data is aligned to the centers and the top 1/3 of the training data was used to train support vector regression using the automatically tuned implementation in OpenCV [3] to estimate the yaw and pitch. This provides significant overlap with the neighboring clusters and results in a smooth transition between clusters for the pose estimates. Roll is estimated by measuring the slope of the line passing through the eye centers. Landmarks from both the GOTS and DLib methods are shown in Figure 2 on sample SHRP2 data.

IV. RESULTS

We compare the results of baseline face detection and head pose estimation with the ground truth results from the Head Pose Validation set. Note that the head pose comparison can only be generated for frames for which a rotation estimate is available and that the particular technique being evaluated was able to generate a face detection and rotation estimate. All analysis was performed on the low-resolution data stream, which is representative of the true SHRP2 data.

A. HPV Face Detection Results

The first step in most automatic face analysis algorithms is face detection. The percentage of frames detected are shown in Table 1 and 2 for the non-driving (static) and driving (dynamic) trials. We separate the trials because the dynamic trials represent more normal driving conditions, but the static offer a variety of different actions during a smaller period of time. The "combined" method simply counts frames where at least one of the other three methods detects a face. There is very little difference in the face detection performance in GOTS and DLib on the Daytime data. For the Night video DLib shows a significant drop in performance while GOTS shows a small improvement. There is significantly less light in the Night video because the scene is illuminated by infrared LEDs on the DAS. The difference in performance may be due to the algorithms sensitivity to that particular lighting condition. Regardless, this shows that GOTS has a significant advantage on the Night videos. Additionally

TABLE 1: PERCENTAGES OF FRAMES WITH DETECTED FACES FOR DIFFERENT BASELINE METHODS: STATIC TRIALS

Algorithm	Night	Transition	Day
G-T	84.7%	92.9%	75.5%
GOTS	79.0%	89.1%	66.0%
DLib	63.0%	79.7%	76.7%
Combined	87.1%	94.5%	86.1%

TABLE 2: PERCENTAGES OF FRAMES WITH DETECTED FACES FOR DIFFERENT BASELINE METHODS: DYNAMIC TRIALS

Algorithm	Night	Transition	Day
G-T	87.9%	88.9%	83.8%
GOTS	79.1%	83.7%	75.5%
DLib	70.7%	69.0%	77.1%
Combined	90.8%	91.1%	89.4%

TABLE 3: LANDMARK MEAN-ABSOLUTE ERROR IN PIXEL FOR DIFFERENT BASELINE METHODS

Algorithm	Night	Transition	Day
G-T	2.50	2.32	2.40
GOTS	2.50	2.31	2.36
DLib	5.16	4.67	4.62

TABLE 4: MEAN ABSOLUTE ERROR FOR YAW IN DEGREES

Algorithm	Night	Transition	Day
G-T	5.89	4.48	5.89
GOTS	5.98	4.47	5.80
DLib	13.6	12.7	13.5

TABLE 5: MEAN ABSOLUTE ERROR FOR PITCH IN DEGREES

Algorithm	Night	Transition	Day
G-T	11.4	10.2	9.57
GOTS	11.3	10.2	9.43
DLib	10.7	10.6	10.9

TABLE 6: MEAN ABSOLUTE ERROR FOR ROLL IN DEGREES

Algorithm	Night	Transition	Day
G-T	1.72	1.58	2.04
GOTS	1.72	1.51	1.94
DLib	1.89	1.85	1.99

these results show that a 8% to 9% improvement can be gained by enabling the GOTS face tracking in both modes. Furthermore, incorporating all the methods improves performance significantly for all cases (between 1% to 10%). The dynamic and static results fit intuition for the day and night cases: dynamic trials have much face movement that is representative of "regular" driving conditions, such as long periods of little head motion with occasional glances, while static trials include actions such as putting on caps which occlude the face and cause problems with detection. Another interesting note is that the transition cases perform better for static trials than dynamic trials. A likely explanation for this is that transition trials have more sudden changes in temporal lighting condition because the vehicle is moving.

The landmark performance evaluation is shown in Table 3. The comparison was done by comparing the detected eye/nose landmarks from GOTS with the annotated ground-truth eye/nose landmarks, with the inner and outer eyes averaged for comparison with GOTS. For DLib, the landmarks corresponding to the inner and outer eye were used, along with the outer mouth and the nose tip. The mean absolute error (MAE) was found for each trial and then the mean of these values were used for the overall estimate. There is not much change with GOTS from case to case, although generally the error seems slightly less during the day. The DLib error is greater overall, but we note that DLib produces an order of several magnitude more points, which can have advantages in some cases as it may be able to leverage the relative position of many more values for more robustness. Because the DLib is open source, and the face detection and landmark detection are two distinct function calls, it may be possible to pair the landmark detection with another face detector to get the best of both worlds. Additionally, it should be possible to retrain the DLib detectors to obtain better results on this particular data set. In general retraining is something that is infeasible with other commercial face detection algorithms because training APIs are typically not available.

B. HPV Pose Estimation Results

The pose estimation error is shown in Tables 4, 5, and 6 for yaw pitch, and roll, separated by night, transition and day trials (there was not a significant difference for the dynamic vs static cases). Note that GOTS does not provide a pitch calculation, so we simply used a constant value of 0 for its estimate and give the error in that case. Both GOTS implementations outperform the DLib case by roughly 10 degrees for Yaw. The DLib pitch values are roughly 3 degrees better than a constant estimate with a roughly 12 degree mean absolute error. The roll performs well for all cases; this is a fairly straightforward measurement.

C. HPV Pose Consistency

One particularly useful attribute of the HPV is the definition of coarse pose, or rough driver attention direction, which can be obtained from the annotations in the data. As an example, in Figure 3 we show the ground truth head pose for ten of the 20 static trial participants in the "calibration" task, where they are looking at the camera mounted by the rear-view mirror. We see that for each participant, there is a distinct clustering of the angles that define this particular coarse head pose, as well as extent of the measurement. However, there is considerable variation between the participants themselves, with the largest difference approximately 25 degrees in pitch. This suggests that coarse head pose measurements may require some level of normalization or calibration for an individual participant,

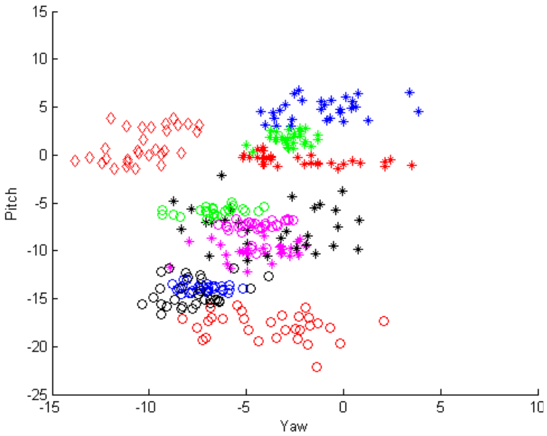

Figure 3. Scatter plot showing Yaw and Pitch of ten of the static trial participants performing the "calibration" task (looking at the camera). Each subject is represented by a unique shape and color.

although this would hopefully be consistent from trip to trip. There are several cases of task' like this in the HPV, which can comprise ground-truth events for coarse head, pose measurements, but we must consider this intra-driver variability when finding an overall method or standard for coarse head pose identification.

With this in mind, we performed an evaluation to understand how the results compare if a calibration step is applied to the data. We test for a good fit by allowing the pose estimates *P* to be adjusted such that *P'=sP+b*, where *s* and *b* are a scale and a bias that are selected on a per-video basis. This provides a mechanism to find an optimal linear adjustment to the estimates to align them with the ground truth. One problem with solving for an optimal *s* and *b* is that the vast majority of the driving data has the driver looking directly out the front window. When solving for a calibration using a least squares fit, we found that *b* would be selected to correspond to that frontal pose in the ground truth and *s* would be very small to compress all the data to cluster all the estimates close to that value.

Instead we implemented a method that evenly weighs the frontal estimate and non-frontal estimates to produce a more useful calibration. We assume that the median of the ground truth is consistent with when the driver is looking out the front windshield and therefore can be used as an estimate to define a region that the driver is considered to be looking forward, say +/- 5 degrees around the center. A tolerance is also defined such that errors within that +/-5 degree range are considered to be correct. (Note that the 5 degree tolerance on head pose was intentionally chosen as it is difficult to achieve this level of accuracy, yet offers room for improvement for future development.) The score is based on the number of poses correctly estimated such that they are within the range defined by the tolerance, however to maintain balance between when the drivers are looking forward and when they are looking to the sides, the score is split into those two categories based on the ground truth and then they are evenly weighted. This means that the frontal and non-frontal can each contribute at most a value of 0.5 for a total maximum score of 1.0. This score indicates what fraction of the estimates fall within the 5 degree range after calibration. To optimize *s* and *b* with respect to this score we used the GENITOR [25] algorithm.

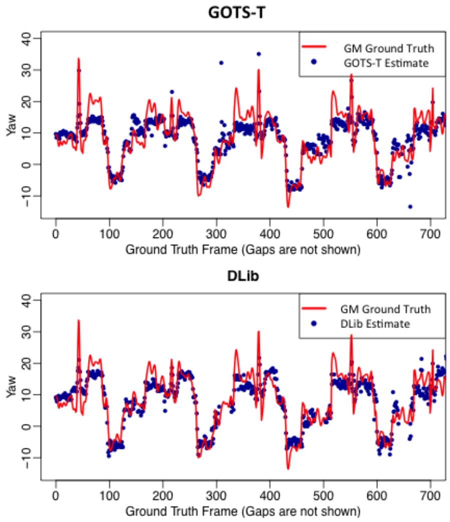

Figure 4. Plots of corrected automated yaw estimates for (top) G-T and (bottom) DLib after applying robust calibration on a dynamic trial. Note that the ground truth is not available for all frames so the x-axis is not a direct correlation with all video frames.

Figure 4 shows an example of how effective the robust calibration process aligned the pose estimates on a dynamic trial from the HPV. In this particular instance G-T detected faces in 99.1% of the ground truth frames and produced an overall score of 0.757 which indicates that approximately 75.7% of poses were within the 5 degree tolerance. DLib detected faces in 87.9% for the frames resulted in an overall score for that video of 0.652. This particular example worked very well, but there are other cases where the agreement is not as high, suggesting either an issue with the particular video or even the ground truth. For the entire HPV dataset, the MAE dropped to 5.34 degrees and 5.54 degrees for G-T and DLib respectively, which is comparable for G-T (from 5.51 degrees uncalibrated) and a dramatic improvement for DLib (from 13.3 degrees uncalibrated).

V.CONCLUSIONS

In this work we presented the application of baseline face detection and pose estimation to naturalistic driving study data, specifically the Head Pose Validation data set which is representative of the SHRP2 NDS, including issues such as relatively low resolution, high compression rates, and differing illumination conditions. We applied a commercial package with government-use, GOTS, which detected faces on 80-90% of the available frames of data in the set and achieved a yaw estimate that was within 4-6 degrees of the ground truth on average. We also used an open source face detector, DLib, which we trained to produce head pose. Given the reduced accuracy of this method, and variation within the data set as shown by the "calibration" task, we explored methods for achieving more consistent estimates with automated methods. Further work in this area will include more automation and blind testing as well. While our immediate goal is to use this data set, and others, to create standards and methodologies to facilitate efforts to automate feature extraction for the SHRP2 NDS, the HPV will be beneficial for the development of algorithms in many areas of intelligent vehicle technology, such as driver alertness monitoring, predicting driver actions, improved warning systems, advanced driver assistance and human machine interaction.

VI. ACKNOWLEDGEMENTS

We would like to acknowledge the support from Federal Highway Administration Exploratory Advanced Research Program, Lincoln Cobb and David Kuehn; the support and assistance of VTTI, especially Jon Hankey, Jon Antin, Suzanne Lee, and Miguel Perez. We would also like to acknowledge the assistance of the Transportation Research Board, especially Ken Campbell and Andy Horosko.

The findings and conclusions of this paper are those of the authors and do not necessarily represent the views of VTTI, SHRP2, the Transportation Research Board, or the National Academies.

REFERENCES

[1] Analyzing Driver Behavior Using Data from the SHRP2 Naturalistic Driving Study. `http://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2_PB_S08_2013-05.pdf`, May 2013. Accessed: 2015-01-23.

[2] Automated Video Feature Extraction Workshop Summary Report. `http://www.fhwa.dot.gov/advancedresearch/pubs/13037/001.cfm`, December 2012. Accessed: 2015-01-23.

[3] OpenCV. `http://www.opencv.org`. Accessed: 2015-01-23.

[4] PyVision. `https://github.com/bolme/pyvision`. Accessed: 2015-01-23.

[5] Transportation Research Board of the National Academies of Science. *The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset*. 2013. Available from the SHRP2 NDS InSight Data Dissemination web site: https://insight.shrp2nds.us. Accessed: 2015-01-23

[6] A. Abate, et al. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885-1906, 2007.

[7] P.N. Belhumeur, et al. Localizing parts of faces using a consensus of exemplars. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition 2011*, June 2011.

[8] K. Campbell. The SHRP2 Naturalistic Driving Study. *TR News*, vol. 282, pp. 30-35, 2012.

[9] A. Dhall, et al. Emotion recognition in the wild challenge 2014: Baseline, data and protocol. *Proc. of the 16th Int'l. Conf. on Multimodal Interaction*, pp. 461-466, 2014.

[10] Y. Dong, et al. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 596-614, December 2010.

[11] A. Doshi and M.M. Trivedi. Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions. *Proc. IEEE Intelligent Vehicles Symposium 2009*, pp. 887-892, June 2009.

[12] A. Gee and R. Cipolla. Determining the gaze of faces in images. *Image and Vision Computing*, vol. 12, no. 10, pp. 639-647, 1994.

[13] N. Gourier, D. Hall, and J.L. Crowley. Estimating Face Orientation from Robust Detection of Salient Facial Features. *Proc. Int'l. Conf. on Pattern Recognition Workshop Visual Observation of Deictic Gestures*, pp. 17-25, August 2004.

[14] T. Horprasert, Y. Yacoob, and L. Davis. Computing 3-d head orientation from a monocular image sequence. *Proc. Int'l. Conf. Automatic Face and Gesture Recognition*, pp. 242-247, 1996.

[15] R. Jafri and H.R. Arabnia. A Survey of Face Recognition Techniques. *Journal of Information Processing Systems*, vol. 5, no. 2, pp. 41-68, 2009.

[16] Q. Ji and X. Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, vol. 8, no. 5, pp. 357-377, October 2002.

[17] D.E. King. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research*, vol. 10, pp. 1755-1758, 2009.

[18] S. Martin, A. Tawari, and M.M. Trivedi. Toward privacy-protecting safety systems for naturalistic driving videos. *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1811–1822, Aug. 2014.

[19] S. Martin, et al. Understanding head and hand activities and coordination in naturalistic driving videos. *Proc. IEEE Intelligent Vehicles Symposium 2014* pp. 884-889, June 2014.

[20] E. Murphy-Chutorian and M.M. Trivedi. Head Pose Estimation in Computer Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 607-626, April 2009.

[21] G.A. Pelaez C, F. Garcia, A. de la Escalera, and J.M. Armingol. Driver Monitoring Based on Low-Cost 3-D Sensors. *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1855-1860, Aug. 2014.

[22] M. Rezaei and R. Klette. Look at the driver, look at the road: No distraction! no accident!. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 129-136, June 2014.

[23] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1891-1898, June 2014.

[24] C. Tran and M.M. Trivedi. Towards a vision-based system exploring 3d driver posture dynamics for driver assistance: Issues and possibilities. *Proc. IEEE Intelligent Vehicles Symposium 2010*, pp. 179-184, 2010.

[25] L.D. Whitley. The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best. *Proc. 3rd Int'l. Conf. on Genetic Algorithms*, vol. 89, pp. 116-123. 1989.

[26] M.-H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, January 2002.

[27] C. Zhang and Z. Zhang. A Survey of Recent Advances in Face Detection. *Microsoft Research Technical Report MSR-TR-2010-66*, June 2010.

[28] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2879-2886, 2012.

[29] Y. Zhu and K. Fujimura. Head pose estimation for driver monitoring. *Proc. IEEE Intelligent Vehicles Symposium 2004*, pp. 501-506, June 2004.

[30] Z. Zhu and Q. Ji. Real time and non-intrusive driver fatigue monitoring. *Proc. 7th Int'l. IEEE Conf. on Intelligent Transportation Systems 2004*, pp. 657-662, October 2004.