2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)

Challenges in Head Pose Estimation of Drivers in Naturalistic Recordings Using Existing Tools

Sumit Jha and Carlos Busso
Multimodal Signal Processing (MSP) Lab, Eric Jonsson School of Engineering and Computer Science
The University of Texas at Dallas, Richardson, Texas 75080, USA
Email: sumit.jha@utdallas.edu, busso@utdallas.edu

***Abstract*—Head pose is an important feature to understand the driver's behavior and their level of attention. While current *head pose estimation* (HPE) algorithms are suitable for many applications under controlled conditions, the performance drops in driving environments where images commonly have varying illumination, occlusions, and extreme head rotations. It is important to understand the limitations of current HPE algorithms to create computer vision solutions that target in-vehicle applications. This paper analyzes the HPE algorithms OpenFace, IntraFace and ZFace, with images recorded under natural driving environment where the goal is to find consistent conditions that affect the performance of current HPE systems. The key feature of the recordings is the use of a headband with AprilTags, which are used to estimate reference head pose angles. We study the effect of different factors including head pose angles, illumination in the frames and occlusions due to glasses. We identify the range of yaw and pitch rotations for which these HPE algorithms provide reliable estimations. While the HPE algorithms work reasonably well for normal/rimless glasses, occlusion due to thicker glasses is a major problem. We identified frames where all the HPE algorithms failed to provide an estimate. We are releasing the data to serve as a benchmark for future HPE algorithms in naturalistic driving conditions.**

I. Introduction

The driver's head pose is one of the most important cues in designing smart human vehicle interfaces. The information about the head pose of the driver can facilitate the coarse estimation of gaze [12], and the detection of driving actions such as checking the mirrors [15] or changing lanes [5]. The ability to monitor driver behavior can prevent hazardous situations on the road, where estimating visual attention of the driver is key to assess awareness and distraction. Head pose estimation can also help in designing more accessible in-vehicle infotainment systems that respond to the driver's gesture (e.g., pulling out information on landmarks in the driver's view, helping locating places) [16].

Head pose estimation has been a very important part of face analysis, where many studies have attempted to formulate accurate and robust algorithms for general applications [2], [11], [19], [24], as well as for specific driving environments [6], [7], [18], [25]. However, many of the proposed solutions for in-vehicle applications have limited scope, where the methods are specific to the experimental setup [9], [17], are difficult to reproduce [4], [22], [25], or rely on experimental evaluation conducted in simulation environment [8], [23], which do not replicate the challenges observed under naturalistic conditions. Varying illumination, wide head pose variations and occlusions affect the efficiency of state-of-the-art head pose estimation tools (see Fig. 1). It is important to understand the limitations of current stand-alone tools for head pose estimation when applied to naturalistic driving conditions. By analyzing the conditions where current solutions tend to fail, we can create algorithms that work in vehicle environments.

(a) High head rotation (b) Occlusion/Glasses (c) Illumination

Fig. 1. Challenges in head pose estimation in naturalistic driving conditions.

This paper analyzes three state-of-the-art head pose estimation tools in naturalistic recordings (IntraFace [24], OpenFace [2], and ZFace [11]). These tools are either completely or partially available to researchers for a wide array of applications. Hence, several research groups have used these algorithms across applications, including solutions for in-vehicle systems. The analysis relies on a database with real driving recordings. 16 drivers wore a headband with AprilTags [20], providing ground truth for the head poses of the drivers (5h and 51m of data). The data collection setting is ideal for evaluating head pose algorithms for in-vehicle applications. We study different scenarios in which the algorithms work well, and the conditions where the algorithms either do not provide estimations (e.g., face is not detected), or the estimations are not accurate. We study the performance of the tools with respect to various head rotation angles, illumination, and occlusions, discussing the range of conditions where current systems provide accurate information.

II. Related Works

Estimating head pose is an important problem in computer vision (see review by Murphy-Chutorian and Trivedi [19]). Algorithms that work on controlled environments tend to fail under the challenging conditions imposed by in-vehicle applications. As Figure 1 illustrates, common problems include changes in illumination, occlusions and extreme head orientations. Even detecting faces is a challenge in these cases. We review proposed approaches for in-vehicle applications.

Features are important in building robust *head pose estimation* (HPE) algorithms for applications in naturalistic driving conditions. Walger et al. [23] analyzed the reliability of four feature sets in estimating the head pose in a setting with varying illumination: *Histogram of Oriented Gradients* (HOG), Gabor filters, *Active Shape Models* (ASM) and a weak perspective geometry. They concluded that ASM consistently performed better than other feature sets. Most of the algorithms use appearance-based features. Murphy-Chutorian et al. [17] employed *Localized Gradient Orientation* (LGO) histograms with *support vector regression* (SVR) to design a light-invariant system that predicts the yaw and pitch angle of the driver. Fridman et al. [6] used HOG and Linear *support vector machine* (SVM) to detect the face, predict 56 landmarks, and estimate three head rotation angles. Lee et al. [13] proposed a system that is robust to lighting conditions and occlusions. They used the vertical edge at the boundaries and the center of the face to detect the yaw angle, and the normalized mean and standard deviation of the horizontal edge for the pitch angle.

Some studies have implemented hybrid methods so one algorithm overcomes the limitation of another. Murphy-Chutorian and Trivedi [18] presented a hybrid approach to detect head pose that combines the static pose estimator proposed in Murphy-Chutorian et al. [17] with a head tracker based on a *sampling importance resampling* (SIR) particle filter. They mapped the face into a 3D model. Zhu and Fujimura [25] proposed a hybrid method to estimate the driver head pose that combines an appearance-based method using *principal component analysis* (PCA) with a 3D motion estimation method using optical flow. More recently, Breidt et al. [3] used depth cameras to obtain a robust estimate of the head pose of the driver in all 6 degrees of freedom.

An alternative solution consists of using multiple cameras in driving environments. Some studies have focused on augmenting the driver's information with the road information [9], [21], while others have used multiple cameras to record the driver's face from different perspectives [8], [22]. Rezaei and Klette [21] used a Fermat point transform to model a 3D shape from the 2D face image and relate it with the information from the road. Tawari et al. [22] proposed a head pose estimation method with three cameras placed at different angles. The head pose is estimated using a *constrained local model* (CLM) and *mixture of pictorial structures* (MPS). The best view is chosen to obtain the final value for the yaw, pitch, and roll angles. Hoffken et al. [8] proposed a method to estimate the head pose from 3D information obtained from a stereo camera setup. They stated that this method is robust in the driving environment, since it does not depend on appearance-based features that tend to work poorly in varying illumination. Cheng et al. [4] used a multi-camera array with LWIR cameras and RGB cameras to track the driver's body.

Many of the previous algorithms require special equipments, the algorithms are not available, or they are difficult to reproduce. Therefore, this study relies on three HPE algorithms developed for general applications showing reliable performance: IntraFace [24], OpenFace [2], and ZFace [11]. We discuss the effect of various factors commonly observed in driving conditions in the performance of these algorithms.

(a) AprilTag (b) Headband with AprilTags

Fig. 2. Headband with AprilTags used for head pose estimation.

III. Database

A. Protocol

The study uses naturalistic recordings collected with the UTDrive platform [1]. The motivation in recording this corpus was to evaluate the relation between head pose and visual attention [12]. The driver's face and the road scene are recorded using a commercial two-channel dash camera (Blackvue DR650GW-2CH), which can be easily installed in any vehicle. Following a fixed protocol, 16 subjects participated in the recording where the main task was to look at predefined markers in the windshield, mirrors, side windows, speedometer panel, radio, and gear. All the sessions were collected between noon and 4pm. The database has three phases: (1) driver looks at the markers while the car is parked, (2) driver looks at the markers while driving, and (3) driver directs his/her head toward the markers while the car is parked. The study only uses recordings from phase 2, which includes not only cases when participants look at target markers while driving, but also natural head movements related to driving tasks.

A key feature of this corpus is the headband with AprilTags worn by the drivers (Fig. 2(b)). The AprilTags, which are discussed in Section III-C, give reliable information about the location and orientation of the driver's head, providing an ideal dataset for the study of automatic HPE algorithms. We have 632,347 frames with head pose estimations with AprilTag (5h and 51m). We collected seven additional recordings using the same protocol without the headband to evaluate the effect of the headband in the performance of HPE algorithms.

B. Distance Metric between Two Rotation Matrices

This paper uses the geodesic distance between two rotation matrices [10] to calculate the difference between the HPE and headband angles.

$$\Delta(R_1, R_2) = \|\log(R_1 R_2^T)\| \quad (1)$$

C. Head Pose Estimation with AprilTags

AprilTags are 2D fiducial markers with unique black and white patterns [20] (Fig. 2(a)). These patterns and their relative orientation are easy to detect in an image, since the shape and size of the markers are known. AprilTags find applications in areas such as robotics, augmented reality, and camera calibration. We use AprilTags to benchmark the orientation of the head pose of the drivers. We designed a rigid headband with 17 square faces, each containing a 2cm x 2cm AprilTag (Fig. 2(b)). Since the geometry of the headband is fixed, we can infer the position and orientation of the headband

(a) High Quality (b) Regular Quality (c) Regular Quality with Illumination

Fig. 3. 3D rendering of talking head with AprilTags based headband.

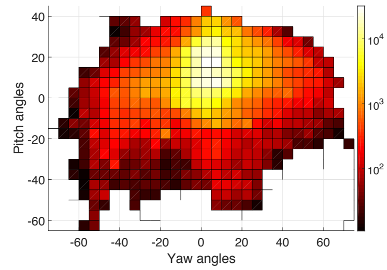


Fig. 4. Distribution of head orientation for yaw and pitch movements across frames.

even with one visible AprilTag. In most cases, more than one AprilTag are visible giving a robust estimate of the head pose (we use the median value of the estimates). From the position and orientation of the headband, we derive the head orientation of the drivers with homogeneous transformations. We individually estimate the head orientation for each frame without tracking, avoiding propagating errors across frames. We temporally smooth the estimation of the head orientation with a median filter of order five.

Evaluation of AprilTag-based headband: We create a synthetic dataset to assess the performance of the AprilTag-based headband. We render a head model wearing the headband with given head rotation (figure 3). The virtual headband has the same AprilTag as the one used in the evaluation. We created multiple frames by changing the angles as follows: *yaw* = (0°, ±30°, ±60°, ±90°), *pitch* =(0°, ±30°), *roll* =(0°, ±20°). We also evaluate frames with combinations of angles (e.g., pitch + yaw). We create three sets of data. The first set corresponds to high quality renderings using 3850×2160 images. The second set includes rendering with 960×540 images, approaching the resolution in our corpus (the size of the AprilTags in these renderings was $\approx 20 \times 20$, which is equivalent to their size in our recordings). The third set also has 960×540 images, but it includes high illumination and shadow, replicating challenges in real driving conditions (Fig. 3(c)). In spite of the extreme rotations, the median angle errors for each set are:

- High Quality (3840 x 2160p) 0.89°
- Med Quality (960 x 540p) 1.26°
- Data with added illumination 2.69°

The estimates are reliable, with reduced accuracy only in the presence of high illumination falling directly on the headband.

D. Distribution of Head Orientation

We estimate the head orientation of the drivers for each frame using the AprilTags. While roll angle provides discriminative information about cognitive distractions [14], we focus the analysis on pitch (vertical) and yaw (horizontal) angles, since it is difficult to visualize results with 3D plots, and the yaw and pitch angles are the more important angles to determine in in-vehicle applications. Figure 4 shows the distribution of head orientation for pitch and yaw angles across frames. As expected, most of the frames have reduced yaw and pitch values as the drivers tend to look at the center of the road most of the time. We observe that most of the data range between -20° and 20°, for yaw and -30° and 10°, for pitch. This is the critical range that HPE algorithms should robustly estimate. Some key driving tasks require head movements beyond this range, which will challenge HPE algorithms, as discussed in Section V.

IV. Head Pose Estimation Algorithms

A. Selected Head Pose Estimation Algorithms

This paper studies the performance of three state-of-the-art algorithms in detecting head orientation in our corpus.

IntraFace [24]: This algorithm uses the *supervised descent method* (SDM) to perform face alignment. Using SDM, a *non-linear least squares* (NLS) function of the image features is minimized to obtain the face landmarks.

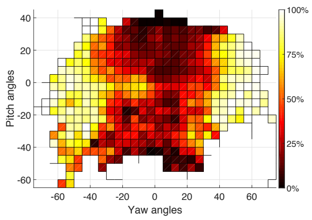
OpenFace [2]: This algorithm uses *constrained local neural field* (CLNF) to detect facial landmarks. It uses probabilistic patch expert to learn the spatial alignment of the landmarks. From the landmarks, head pose is obtained using a *perspective-n-point* (PnP) method.

ZFace [11]: This algorithm builds a dense 3D mesh with a manageable size by storing only the updated landmark displacements. The algorithm uses cascade regression to perform face alignment based on the mesh.

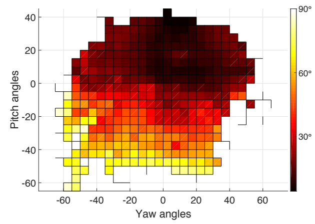
We evaluate the algorithms with the default settings, without attempting to optimize the implementation for the task at hand. The objective of this study is not to compare these algorithms, but to understand better the challenges that we need to solve before using HPE algorithms in vehicle applications. We evaluate each video sequence in the corpus with these HPE algorithms. We unify the coordinate reference system for these algorithms as follows. We normalize the data from each HPE algorithm by subtracting its mean and dividing by the standard deviation. Then, we translate the data, matching the first and second order statistics of the AprilTag angles. This normalization process unifies the angles' coordinate system across HPE algorithms and AprilTag headband. The yaw angle is positive in the right direction and negative in the left direction, and the pitch is positive looking up and negative looking down. The output of IntraFace has a few false positive faces where the size is too small or the location is not correct. We use contextual information to eliminate these detections.

B. Effect of the Headband on Head Pose Estimation

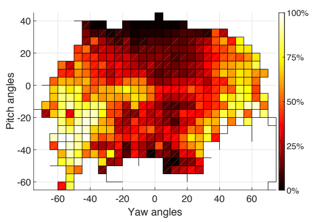
We can benchmark the performance of the HPE algorithms with precise metrics by adding the headband in the recordings. An implicit assumption in the analysis is that the headband



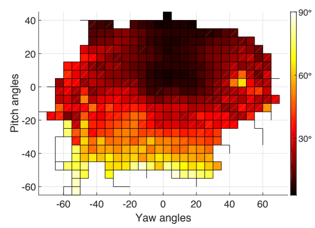
(a) IntraFace - detection failure



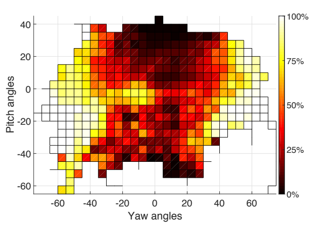
(b) IntraFace - angular error



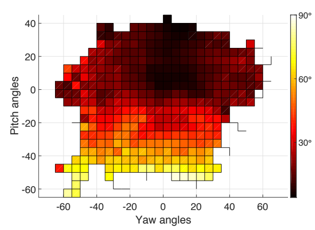
(c) OpenFace - detection failure



(d) OpenFace - angular error



(e) ZFace - detection failure



(f) Zface - angular error

Fig. 5. Detection failure rate and angular error of the HPE algorithms as a function of pitch and yaw angles.

does not dramatically affect the performance of the HPE algorithms. The headband occludes a portion of the face, so it is important to quantify its effect on HPE algorithms. For this purpose, we compare the recordings with the headband (16 participants) and without the headband (7 participants). We only estimate the percentage of frames that the HPE algorithms fail to detect, since we do not have ground truth angles for the head poses when the driver does not use the headband. Table I lists the percentage of frames that were not detected by each method. We cannot estimate the head orientation from the AprilTags in 5.3% of the frames due to saturated images, blurred images, or significant head rotations. Without the headband, ZFace provides head orientation estimates for 91.1% of the frames, which is remarkable. The table shows that the headband affects the HPE algorithms. The most significant case is for ZFace, where the percentage of frames without head orientation estimates increases from 8.9% to 21.9% when we include the headband. ZFace constructs a 3D mesh of the whole face, and the addition of the headband affects this process. For OpenFace (21.8% to 24.1%) and IntraFaces (19.0% to 27.3%), the problem is not as important. Notice that over 72% of the frames are available for the comparison (over 550K frames), so we can still study the performance of the HPE algorithms under the proposed experimental evaluation. Notice that once the face is detected, the headband should not affect the accuracy of the algorithms in estimating the head pose of the driver, since they depend on multiple facial landmarks. Only the facial markers near the eyebrows may be

TABLE I
PERCENTAGE OF FRAMES WHERE THE HPE METHODS FAILED TO PROVIDE AN ESTIMATE.

Estimation Method	AprilTag [%]	IntraFace [%]	OpenFace [%]	ZFace [%]
Without Headband	**-**	**19.0**	**21.8**	**8.9**
With Headband	**5.3**	**27.3**	**24.1**	**21.9**

slightly affected, therefore, the head pose estimations should be accurate for detected faces.

V. Analysis of Head Pose Estimation

The robustness of HPE algorithms is affected by various variables such as head rotation, occlusion, or illumination. This section studies the effect of these factors in the performance of the HPE algorithms.

A. *Performance as a Function of Yaw and Pitch Angles*

One of the most important variables that affects the robustness of HPE algorithms is the variation of head pose. While it is widely accepted that drivers tend to look at the center of the road, it is important to estimate cases where the head pose is not frontal, which convey important information. The plots in the left column in Figure 5 show the percentage of frames when the HPE algorithms were not able to provide an estimate. We do not include the bins which have fewer than 10 frames in our analysis. The robustness of the HPE algorithms decreases when the driver turns his head away from the road, especially for extreme values of yaw. Almost all the frames do not have an estimate for a yaw angle less than -40° and greater than 40°. OpenFace provides an estimate for 40-60% of the frames, even with yaw angles close to 60° (Fig. 5(c)).

We also compare the estimations of the HPE algorithms with the estimations derived from the headband. While the head pose angles derived from the AprilTag is not perfect, it gives reliable values across conditions, providing a reasonable benchmark to evaluate the HPE algorithms. The plots in the right column in Figure 5 show the difference between the angles estimated by each HPE algorithm and the angles estimated with the headband. For pitch angles between -40° and 20° and yaw angles between -60° and 60° the difference in estimation is very low. The figures show that the HPE algorithms are very precise for yaw angles, which is very important for many in-vehicle applications. The performance drops when the pitch angle is below -20°. While pitch angles below -40° are highly unlikely (Fig. 4), the error at these angles is more than 60°, so they are completely unreliable.

B. *Occlusion Due to Glasses*

The presence of glasses affects the performance of HPE algorithms. In our database, we had eight subjects out of 16 who wore eye glasses during the recordings. Three of them wore thick frame glasses, which caused major occlusion of the face (see Fig. 1(b)). Table II lists the performance for three groups of participants: drivers wearing thick-frame glasses (3), drivers wearing normal/rimless glasses (5), and drivers without glasses (8). We measure performance with the percentage of

TABLE II
PERCENTAGE OF MISSED FRAMES WITH AND WITHOUT GLASSES

Method	Intraface		Openface		Zface	
	Failed [%]	Δ [°]	**Failed** [%]	Δ [°]	**Failed** [%]	Δ [°]
TF Glasses	**67.7**	**10.0**	**67.5**	**12.9**	**64.1**	**12.3**
NF Glasses	**18.5**	**14.0**	**13.7**	**11.3**	**13.8**	**11.5**
No Glasses	**15.4**	**9.0**	**12.1**	**9.0**	**8.8**	**8.5**

[TF:Thick-frame Glasses; NF: Normal-frame/rimless Glasses; No: No Glasses]

frames where HPE algorithms failed to provide an estimate. We also estimate the average difference between the predicted angles and AprilTag-based angles (Δ).

The percentage of frames missed by the HPE algorithms significantly increases when drivers use thick-frame glasses (over 64%). we do not observe this problem for drivers using normal or rimless glasses, where the percentage of frames without detection is only slightly worse than the ones for drivers without glasses. For frames where the head pose is detected, the average deviation from the reference head pose (AprilTag) is between 6.8° and 14.1°, where drivers with thicker glasses have the larger angular error.

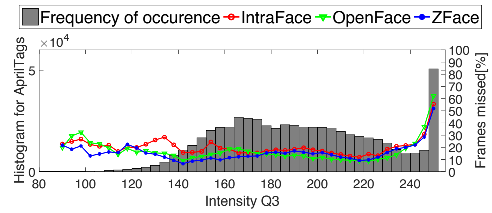
C. Effect of illumination

Varying illumination is one of the important challenges in computer vision algorithms for vehicle applications, as shadows and rapid changes in illumination are commonly observed (see Fig. 1(c)). When the vehicle turns, the lighting condition can dramatically change, which negatively impacts the performance of HPE algorithms to detect faces. We quantify illumination with the third quartile (Q3) of the pixels' intensity over the facial region (i.e., square region under the detected headband). Q3 provides useful information about illumination. Q3 increases for bright images, so we expect lower performance for high values of Q3. Also, if the image is too dark, the value for Q3 will decrease.

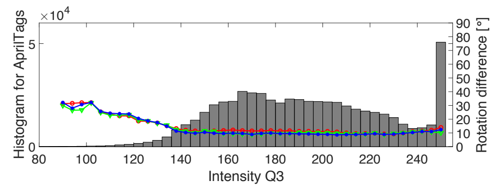
Figure 6(a) shows the percentage of frames where the head pose was not detected using each of the HPE algorithms as a function of Q3. As a reference, the bars represent the Q3 intensity distributions of frames detected by the AprilTag-based headband, showing the general values observed for Q3 in natural driving conditions. The HPE algorithms are not reliable when Q3 is above 240, since the image is too bright. This is important as we have many samples with high values of Q3. We also evaluate the difference between the head pose estimated by the headband and each of the HPE algorithms. Figure 6(b) shows the average angular error as a function of Q3, where the bars represent the distribution of frames in the corpus. When the HPE algorithms provide an estimate, the performance is consistent for Q3 between 140 and 255. The angular error increases for lower values of Q3 less than 150, where the image becomes too dark.

D. Ideal Scenarios (IS) versus Challenging Scenarios (CS)

The final analysis identifies two groups of frames. The first group is referred to as *Ideal Scenarios* (IS), and consists of frames where the three HPE algorithms provide the head pose

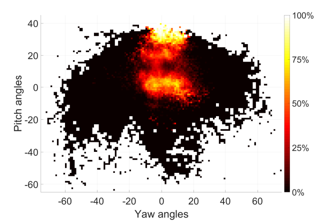


(a) Percentage of frames missed by the PHE algorithm

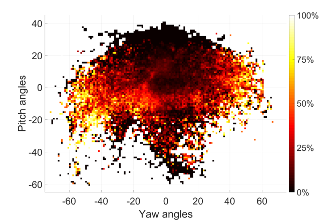


(b) Angle difference between HPE algorithms & AprilTag

Fig. 6. Performance of HPE algorithms as a function of the third quartile (Q3) of the pixel's intensity. The figures show the role of illumination.



(a) Ideal Scenarios (IS) group



(b) Challenge Scenarios (CS) group

Fig. 7. The percentage of frames per cell that belong to either the IS or CS groups. Bright colors indicate that most of the frames in the corpus for a given head pose belong to the corresponding group.

of the driver and the estimates do not deviate more than 10° from the reference angles derived from the headband. The IS group has 214,261 frames (33.9% of the corpus). The second group is referred to as *Challenging Scenarios* (CS), and consists of the frames where none of the HPE algorithms was able to estimate the head pose of the drivers. The CS group has 80,785 frames (12.8% of the corpus). Figure 1 shows three frames from this group.

We study the distribution of IS and CS frames as a function of yaw and pitch angles. Figure 7 shows the percentage of frames for each cell that belong to either the IS group (Fig. 7(a)) or the CS group (Fig. 7(b)). On the one hand, most of the frames with yaw angles between -20° to 20°, and pitch angles between 40° and -20° are included in the IS group. The HPE algorithms provide reliable information within this range. On the other hand, Figure 7(b) shows that most of the frames with yaw angles below -20° or above 20° are included in the CS group.

We also analyze whether the illumination conditions are different for the IS or CS groups. We estimate the third quartile of the intensity across frames around the face. For a given value of Q3, we estimate the percentage of frames belonging to either the IS or CS groups. Figure 8 shows that almost 50% of the frames belong to the CS group when Q3 is above 240 (saturated frames). This percentage also increases for frames with Q3 values lower than 140 (dark frames).

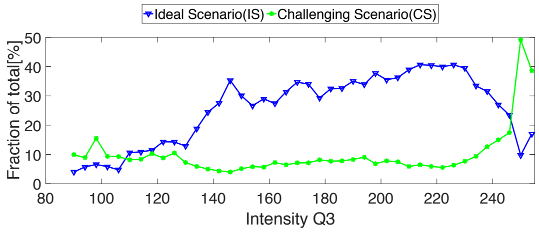


Fig. 8. Percentage of frames belonging to the IS and CS groups as a function of Q3. The HPE algorithms tend to fail when the Q3 values are outside the range [140,240].

VI. Conclusions

This paper analyzed the performance of three HPE algorithms in naturalistic driving conditions. We studied multiple factors that strongly affect the robustness of these tools such as extreme head orientations, occlusion and illumination. We used a novel corpus which provides head pose estimation for each frame using a headband with AprilTags. The analysis identified conditions under which current state-of-the-art frameworks are effective in recognizing the head pose of the driver. By studying the weaknesses of current HPE algorithms, more robust solutions can be created addressing the computer vision challenges imposed by these conditions. To facilitate this task, we will release the database with this paper, so other research groups can evaluate their solutions in these challenging conditions.

While the performance without the headband may be slightly better, we expect that the trends are representative of the capabilities provided by state-of-the-art HPE algorithms. We use the headband to estimate the reference head pose of the driver, which was key to visualize the range of yaw and pitch angles where the HPE algorithms provide reliable information. Likewise, there are other challenges that reduce the efficiency of HPE which were not considered in this study (presence of beard, caps, makeup, jewelry, and driving during night). We plan to extend our database adding more realistic scenarios. Our future work also includes the developments of robust HPE solutions for vehicle applications. Advances in this area can have clear benefits in in-vehicle safety systems that can detect when the attention of the drivers is not on the road, indicating lack of situational awareness.

References

[1] P. Angkititrakul, M. Petracca, A. Sathyanarayana, and J. Hansen. UT-Drive: Driver behavior and speech interactive systems for in-vehicle environments. In *IEEE Intelligent Vehicles Symposium*, pages 566–569, Istanbul, Turkey, June 2007.

[2] T. Baltru, P. Robinson, L.-P. Morency, et al. Openface: an open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10, Lake Placid, NY, USA, March 2016. IEEE.

[3] M. Breidt, H. H. Bülthoff, and C. Curio. Accurate 3d head pose estimation under real-world driving conditions: A pilot study. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pages 1261–1268, Rio de Janeiro, Brazil, November 2016. IEEE.

[4] S. Y. Cheng, S. Park, and M. M. Trivedi. Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis. *Computer Vision and Image Understanding*, 106(2):245–257, May-June 2007.

[5] A. Doshi and M. Trivedi. On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems*, 3(10):453–462, September 2009.

[6] L. Fridman, P. Langhans, J. Lee, and B. Reimer. Driver gaze estimation without using eye movement. *IEEE Intelligent Systems*, 31(3):49–56, May-June 2016.

[7] Z. Guo, H. Liu, Q. Wang, and J. Yang. A fast algorithm face detection and head pose estimation for driver assistant system. In *2006 8th international Conference on Signal Processing*, volume 3, Beijing, China, November 2006. IEEE.

[8] M. Höffken, E. Tarayan, U. Kreßel, and K. Dietmayer. Stereo vision-based driver head pose estimation. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pages 253–260, Dearborn, Michigan, USA, June 2014. IEEE.

[9] K. S. Huang, M. M. Trivedi, and T. Gandhi. Driver's view and vehicle surround estimation using omnidirectional video stream. In *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pages 444–449, Columbus, OH, USA, June 2003. IEEE.

[10] D. Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, June 2009.

[11] L. A. Jeni, J. F. Cohn, and T. Kanade. Dense 3d face alignment from 2d videos in real-time. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–8, Ljubljana, Slovenia, May 2015. IEEE.

[12] S. Jha and C. Busso. Analyzing the relationship between head pose and gaze to model driver visual attention. In *2016 IEEE Intelligent Transportation Systems Conference*, Rio de Janeiro, Brazil, November 2016. IEEE.

[13] S. J. Lee, J. Jo, H. G. Jung, K. R. Park, and J. Kim. Real-time gaze estimator based on driver's head orientation for forward collision warning system. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):254–267, March 2011.

[14] N. Li and C. Busso. Predicting perceived visual and cognitive distractions of drivers with multimodal features. *IEEE Transactions on Intelligent Transportation Systems*, 16(1):51–65, February 2015.

[15] N. Li and C. Busso. Detecting drivers' mirror-checking actions and its application to maneuver and secondary task recognition. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):980–992, April 2016.

[16] T. Misu. Visual saliency and crowdsourcing-based priors for an in-car situated dialog system. In *International conference on Multimodal interaction (ICMI 2015)*, pages 75–82, Seattle, WA, USA, November 2015.

[17] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi. Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In *2007 IEEE Intelligent Transportation Systems Conference*, pages 709–714, Bellevue, WA, USA, September-October 2007. IEEE.

[18] E. Murphy-Chutorian and M. Trivedi. HyHOPE: Hybrid head orientation and position estimation for vision-based driver head tracking. In *IEEE Intelligent Vehicles Symposium (IV 2008)*, pages 512–517, Eindhoven, The Netherlands, June 2008.

[19] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 31(4):607–626, April 2009.

[20] E. Olson. AprilTag: A robust and flexible visual fiducial system. In *IEEE International Conference on Robotics and Automation (ICRA 2011)*, pages 3400–3407, Shanghai, China, May 2011.

[21] M. Rezaei and R. Klette. Look at the driver, look at the road: No distraction! no accident! In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, pages 129–136, Columbus, OH, June 2014.

[22] A. Tawari, S. Martin, and M. M. Trivedi. Continuous head movement estimator for driver assistance: Issues, algorithms, and on-road evaluations. *IEEE Transactions on Intelligent Transportation Systems*, 15(2):818–830, April 2014.

[23] D. J. Walger, T. P. Breckon, A. Gaszczak, and T. Popham. A comparison of features for regression-based driver head pose estimation under varying illumination conditions. In *Computational Intelligence for Multimedia Understanding (IWCIM), 2014 International Workshop on*, pages 1–5, Paris, France, November 2014. IEEE.

[24] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 532–539, Portland, Oregon, June 2013. IEEE.

[25] Y. Zhu and K. Fujimura. Head pose estimation for driver monitoring. In *IEEE Intelligent Vehicles Symposium (IVS 2004)*, pages 501–506, Parma, Italia, June 2004.