Alioua *et al. EURASIP Journal on Image and Video Processing* (2016) 2016:2
DOI 10.1186/s13640-016-0103-z

EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

Driver head pose estimation using efficient descriptor fusion

Nawal Alioua[1,2*], Aouatif Amine[4], Alexandrina Rogozan[3], Abdelaziz Bensrhair[3] and Mohammed Rziza[2]

Abstract

A great interest is focused on driver assistance systems using the head pose as an indicator of the visual focus of attention and the mental state. In fact, the head pose estimation is a technique allowing to deduce head orientation relatively to a view of camera and could be performed by model-based or appearance-based approaches. Model-based approaches use a face geometrical model usually obtained from facial features, whereas appearance-based techniques use the whole face image characterized by a descriptor and generally consider the pose estimation as a classification problem. Appearance-based methods are faster and more adapted to discrete pose estimation. However, their performance depends strongly on the head descriptor, which should be well chosen in order to reduce the information about identity and lighting contained in the face appearance. In this paper, we propose an appearance-based discrete head pose estimation aiming to determine the driver attention level from monocular visible spectrum images, even if the facial features are not visible. Explicitly, we first propose a novel descriptor resulting from the fusion of four most relevant orientation-based head descriptors, namely the steerable filters, the histogram of oriented gradients (HOG), the Haar features, and an adapted version of speeded up robust feature (SURF) descriptor. Second, in order to derive a compact, relevant, and consistent subset of descriptor's features, a comparative study is conducted on some well-known feature selection algorithms. Finally, the obtained subset is subject to the classification process, performed by the support vector machine (SVM), to learn head pose variations. As we show in experiments with the public database (Pointing'04) as well as with our real-world sequence, our approach describes the head with a high accuracy and provides robust estimation of the head pose, compared to state-of-the-art methods.

Keywords: Driver monitoring, Head pose estimation, Support vector machine, Feature selection

1 Introduction

The increasing number of traffic accidents in the last years becomes a serious problem. The enhancement of traffic safety is a high-priority task for different government agencies over the world such as "National Transportation Safety Administration" (NTSA) in USA and "Observatoire National Interministériel de la Sécurité Routière" (ONISR) in France. In addition, automotive manufactures and researcher laboratories are also contributing to this important mission. Some preventive systems such as alcohol test and speed measurement radar are deployed to reduce the number of traffic accidents, but it is obvious that hypovigilance remains one of the most principal causes. In fact, hypovigilance is responsible for 20–30 % of road deaths and this statistic reaches 40–50 % in particular crash types, such as fatal single vehicle semi-trailer crashes [1]. Moreover, there are no standard rules to measure the driver vigilance level; the unique solution is to observe the signs. The first hypovigilance signs are itchy eyes, neck stiffness, back pain, yawning, difficulty to stabilize speed and to maintain trajectory, frequent position changes, and inattention to environment (road signs, pedestrian). Fatigue, sleep deprivation, soporific drugs, driving more than 2 h without break, and driving in a monotone road are the main causes of hypovigilance. The appropriate reactions when those signs appear are to stop driving immediately and take a break, but unfortunately, the drivers are not aware of their vigilance level and overestimate it. For this purpose, several studies have been conducted to develop intelligent systems for continuously

*Correspondence: nawal.alioua@yahoo.fr
[1] Ibn Zohr University, Morocco, BP 32/S, Agadir, Morocco
[2] LRIT-CNRST 29, Faculty of Sciences, Mohammed V University, Ibn Battouta Avenue, Rabat, Morocco
Full list of author information is available at the end of the article

estimating driver vigilance level and emitting visual and acoustic alarms to avert the driver against abnormal state. The warning signals could also activate the vibration of driver's seat or even a mechanism that stops the car at the roadside.

The literature regroups *three categories* of safety systems distinguished by the type of signals used to determine the driver vigilance level. (i) Studying *physiological signals* consists on measuring signal changes represented by brain waves or heart rate using special sensors such as electroencephalography (EEG), electrocardiography (ECG), and electromyography (EMG) [2]. Only few works are proposed in this category since the process is highly intrusive because of the necessity to connect sensing electrodes to the driver body. (ii) Monitoring *vehicle signals* can reveal abnormal driver actions indirectly, by studying several parameters such as vehicle velocity changes, steering wheel motion, lateral position, or lane changes. Some commercial systems already use these techniques since the signals are significant and their acquisition is quite easy compared to the previous category. Mercedes-Benz proposes in 2009 a commercial system named "Attention Assist" based on sensitive sensors allowing precise monitoring of the steering wheel movements and the steering speed. The system is active at 80–180 km/h and calculates an individual behavioral pattern during the first minutes of each trip. Audible and visual signals are emitted when typical indicators of hypovigilance are detected. The major disadvantages of such system are the limitations caused by the dependence to vehicle type, driver experience, and road conditions. (iii) Approaches based on *physical signals* utilize image processing techniques to measure the driver vigilance level reflected through the driver's face appearance and head/facial feature activity. These techniques are based principally on studying facial features, especially eye state [3–5], head pose [6, 7], or mouth state [8]. According to the study performed in [9], monitoring driver eye closure and head pose are the most relevant indicators of hypovigilance. Different kinds of cameras have been used for such systems: visible spectrum (VS) camera [10], infrared (IR) camera [11], stereo cameras [12], and also the Kinect sensor [13]. The Kinect sensor provides color images, IR images, and 3D information. However, this sensor is not very adapted to the real driving conditions since it is designed for indoor use and it is conceived to be placed in a minimal distance of 1.8 m from the target. The IR camera is adapted when driving at night, but it is not recommended when driving at daylight conditions, since the acquisition will suffer from color distortion. The VS camera is the cheapest one, and it provides robust acquisition even if the light is reduced. However, it is a big challenge to monitor the driver vigilance level using a single VS camera without depth information and IR information.

In our previous work [10], we have proposed a real-time system using a very cheap VS camera to determine driver fatigue and drowsiness by analyzing mouth and eyes, respectively. This system suffers from missed detection when the specific facial features are not visible because of non-frontal head position. The aim of this paper is to develop a head pose estimation approach that reveals rapidly driver inattention from monocular visible spectrum images, without prior facial feature extraction. To construct a robust head pose estimator, we follow an appearance-based head pose estimation architecture instead of a model-based one. These two architectures are detailed in Section 2. In fact, the model-based architecture is incompatible with our problem since it requires facial features to construct the face geometrical model, whereas the appearance-based one uses the whole head structure characterized by an image descriptor. Actually, the performance of the appearance-based estimator depends strongly on the image descriptor, which should be chosen carefully in order to reduce the information about identity and lighting contained in the face appearance. In this work, as detailed in Section 3, we first propose a novel descriptor resulting from the fusion of four most relevant orientation-based head descriptors, namely the steerable filters (SF), the histogram of oriented gradients (HoG), the Haar features, and an adapted version of SURF descriptor. Second, in order to construct a compact, robust, and pertinent subset of the descriptor's features, a comparative study is conducted on some well-known feature selection algorithms. Finally, the obtained subset is subject to the classification process, performed by the support vector machine (SVM), to learn head pose variations. In Section 4, an evaluation of the proposed head pose estimator on the public Pointing'04 database is performed to validate our approach and to compare it with the most representative and the best state-of-the-art methods. After that, we have acquired and annotated a driver video sequence simulating attention and inattention states in order to validate the proposed estimator in a real environment. Finally, we present a conclusion and discussion in Section 5.

2 Related works

2.1 Overview of head pose estimation techniques

In computer vision, head pose estimation can be defined as the ability to deduce head orientation relatively to a view of camera and it can refer to different interpretations [14]. At coarse level, a head is identified by a few discrete poses, but it might be estimated by a continuous angular measurement according to multiple degrees of freedom. The discrete representation is adapted to the applications requiring the knowledge of limited number of pose classes instead of the whole possible pose angles corresponding to the continuous representation. Even if muscular rotation

of head influences its orientation, it is often ignored and human head is considered as an incorporeal rigid object. This hypothesis allows to represent head pose using only three degrees of freedom which are *pitch*, *yaw*, and *roll*. Pitch corresponds to up and down motion around the X axis, yaw refers to left and right direction around the Y axis, and roll represents tilting the head towards left and right direction around the Z axis (see Fig. 1). Another hypothesis to be considered when building head pose estimator is the *pose similarity assumption*, which means that different people at the same pose look more similar than the same person at different poses. In literature [14–16], three requirements are established to define an efficient head pose estimator.

(R1) Perform head pose estimation from monocular cheap camera. Potentially, the accuracy can be improved using stereo techniques that need additional equipment cost, computation, and memory requirements.
(R2) Ensure autonomy by avoiding manual initialization or adjustment.
(R3) Guarantee invariance to identity and environment in order to make the system more efficient and robust.

In literature, many techniques have been proposed to estimate head pose for diverse applications including monitoring driver state systems. These techniques can be categorized into two main groups [17], namely model-based techniques and appearance-based techniques.

2.1.1 Model-based head pose estimation

Model-based techniques require specific facial features to estimate head pose. In this category, we can find geometric approaches that determine head pose from the relative locations of facial features such as eye corners, mouth corners, and nose tip. The most recent systems based on facial geometry are proposed in [12, 17, 18]. In [18], the authors propose a method for automatic head pose estimation using three features (the eyes and nose locations) ruled by the concept of golden ratio, whereas the majority of geometric approaches require at least five features. The golden ratio is the proportionality constant adopted by Leonardo Da Vinci in his master-work called The Vitruvian Man.

Flexible models based on fitting non-rigid models to the facial structure of each subject also belong to this category since comparisons at feature level are made rather than comparisons at global appearance level. Flexible models include methods such as active shape models (ASM), active appearance models (AAM), and elastic graph matching (EGM). In [19], authors present a probabilistic framework which do not need user initialization unlike most of flexible models which do not respect the requirement (R2).

Model-based techniques are dependent to the performance of the facial feature localization which is, in addition to high-resolution requirement, the major disadvantage.

2.1.2 Appearance-based head pose estimation

Appearance-based techniques work under the assumption that the 3D face pose and some properties of the 2D facial image are linked by a certain relationship [20]. Appearance template methods [21, 22] define this relationship by matching new head images into discrete head templates using image-based comparison metrics. These

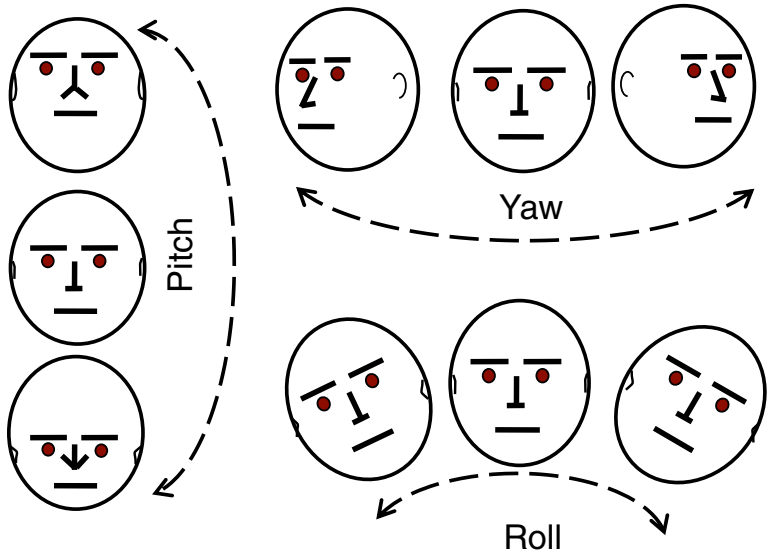

Fig. 1 Head pose representation using three degrees of freedom

methods are the most sensitive to lighting conditions. In our previous work [21], we use a robust descriptor based on steerable filters to construct a reference template for each discrete pose of the training set, and a likelihood parametrized function is learned to match the templates to new entries.

Another way to determine the relationship is the use of classification techniques on a large number of training data in order to learn an efficient separation between pose classes. SVM classifiers are quite used in literature to classify head poses [23–25] since they are adapted to real-time applications. In [23], the authors show that the multi-scale Gaussian derivatives, which are a particular case of steerable filters, combined to SVM give good results. In [26], normalized faces are used to train an auto-associative memory using the Widrow-Hoff correction rule in order to classify head poses. In one of the most recent works [27], the authors consider that object detection and continuous pose estimation are interdependent problems and they jointly formulate them as a structured prediction problem, by learning a single and continuously parameterized object appearance model over the entire pose space. After that, they design a cascaded discrete-continuous inference algorithm to effectively optimize a non-convex objective, by generating a diverse proposal to explore the complicated search space. Then, the model is learned using a structural SVM for joint object localization and continuous state estimation and a new training approach which reduces the processing time. Among the experiments, the authors perform the head pose estimation over the Pointing'04 database without considering the detection task, since they note that the images contain clean backgrounds. Before applying their method, the heads are cropped manually and the HOG descriptor is applied on three scales. Based on the relationship between the symmetry of the face image and the head pose, the authors in [28] propose a face representation method for head yaw estimation which is robust against rotations and illumination variations. First, they extract the multi-scale and multi-orientation Gabor representations of the face image, and then they use covariance descriptors to compute the symmetry between two regions in terms of Gabor representations under the same scale and orientation. Second, they apply a metric learning method named KISS MEtric learning (KISSME) to enhance the discriminative ability and reduce the dimension of the representation. Finally, the nearest centroid (NC) classifier is applied to obtain the final pose.

Regression techniques are also utilized to address head pose estimation problem when the pose angles are ordered, but they are more complex since they need powerful unit process to respect real-time constraints. In this case, the relationship is defined by learning continuous mapping functions between the face image and the pose space [29–31]. In [31], authors extract head feature vector using the robust 3-level HOG pyramid and then the partial least square (PLS) regression is used to determine the coefficients modeling the relationship between the head and its pose. In [24], authors use a dense scale invariant feature transform (SIFT) descriptor to construct feature vector and the random projection (RP) is applied to reduce the vector dimension. Similar to [32], the authors combine classification and regression to obtain an accurate estimation of head pose but this kind of approach is time consuming.

One can also include tracking approaches in the appearance-based techniques since they are based on head appearance to estimate poses in addition to temporal continuity and smooth motion of the heads in the video sequence. Particle filters (PF) [6, 33] are the most used technique to track head poses; in [6], authors propose a hybrid head orientation and position estimation system for driver head tracking based on PF. While tracking techniques can achieve high accuracy, they usually require an initial step such as frontal view or manual initialization which does not respect the requirement (R2).

The major part of the appearance-based techniques presented above is applied on features that verify the pose similarity assumption. In addition, the descriptor must be fast, must be robust to variations of lighting conditions, and should be representative of head orientations in order to respect the requirements (R2) and (R3). Gabor filter [34], steerable filters (SF) [21], SIFT [24], and HOG [33] are the most used descriptors verifying these requirements. Some dimensionality reduction methods can be used to seek a low-dimensional continuous manifold constrained by the pose variations. Principal component analysis (PCA) and linear discriminant analysis (LDA) are the most used dimensionality reduction techniques for head pose estimation [14]. In [35], authors propose to represent each head pose appearance neighborhood by a query point to reduce the size and then apply a piece-wise linear local subspace learning method to map out the global nonlinear structure for head pose estimation.

Each category suffers from some disadvantages. Even if model-based methods are fast and simple, they are sensitive to occlusion and require high-resolution images since the difficulties lie in detecting the specific facial features with high precision and accuracy. Appearance-based approaches are not affected by these limitations, but they are sensitive to information about identity and lighting contained in the face appearance. However, when using a robust and efficient head pose descriptor, the appearance-based techniques become invariant to identity and lighting.

In the following, we expose some head pose estimation techniques for monitoring driver vigilance state.

2.2 Driver head pose estimation

A great interest is focused on driver assistance systems that use driver head pose as a cue to visual focus of attention and mental state [6, 11, 36, 37]. A commercial product called Smart Eye AntiSleep [36, 38] is developed and corresponds to a compact system equipped with one VS camera and two IR flashes designed for automotive applications. AntiSleep measures 3D head position and orientation, gaze direction, and eyelid closures. Authors use a tracking approach and a geometric method as initialization step based on a 3D head model containing the relative distances between specific facial points localized using local Gaussian derivatives [39], SIFT, and Gabor jets [40]. The probability distribution of each point descriptor is learned from a large set of facial training images. Then, an initial head pose is estimated from the positions of the facial features and the generic head model. The detected facial features are then tracked using structure-from-motion algorithms. During tracking, the driver-specific appearance of each generic feature is learned for different views. The obtained information is used to stabilize and speed up tracking. This commercial product is limited to controlled environments and therefore is essentially used for simulation purposes.

The most popular research laboratory working on driver assistance systems is the CVRR Laboratory at the University of California, USA. This team proposes several approaches to monitor driver vigilance [6, 7, 16, 37, 41]. In [6], the problem of estimating driver head pose is addressed using a localized gradient orientation descriptor on 2D video frames acquired by a special camera (sensitive to IR and VS lights) as the input to two support vector regressions (SVRs), one for pitch and the other for yaw. This team has equipped a prototype car with many sensors allowing to look in and look out of a vehicle. Such equipment is too expensive to be widely used in car industry. Unfortunately, we cannot compare with these approaches since their database is not accessible and the systems are not detailed enough to allow reproduction.

The goal of our global work is to propose a system for monitoring driver vigilance level based on low cost equipment. In this paper, we focus our attention on estimating driver head pose respecting the requirements (R1), (R2), and (R3). In Table 1, we summarize the properties of some methods presented above and we precise with the signs "$*$" and "$+$" the approaches that will be used for comparison in Section 4. The sign "$*$" is associated to the most used references for benchmarks in literature, and the sign "$+$" corresponds to the recent works providing the best results. From literature, it is obvious that appearance-based techniques are more adapted to our purpose since they respect the requirement (R3) when the descriptor used to construct the feature vector is chosen carefully. Therefore, we propose an efficient and robust fusion of the most pertinent head pose descriptors and we decide to use the SVM classifier since it is adapted to the real-world applications and it proves its efficiency in literature.

3 Discrete head pose estimation for monitoring driver vigilance level

When analyzing the impact of head orientations on driver inattention, we can observe that the driver is attentive to

Table 1 Overview of the most relevant literature approaches

Reference	Year	Type	Methods	R1	R2	R3
Our$^{(+,0,4)}$	2015	CI	Descriptor fusion + SVM	√	√	√
[17]$^{(1)}$	2012	GM	Face symmetry	√	√	×
[12]$^{(0)}$	2012	GM	3D geometry	√	√	√
[18]$^{(+,4)}$	2013	GM	Golden ratio	√	√	×
[19]$^{(2,3)}$	2010	FM	Face model	√	√	√
[21]$^{(+,4)}$	2013	AT	SF + LPF	√	√	√
[31]$^{(+,4)}$	2012	Rg	HOG + PLS Rg	√	√	√
[24]$^{(+,4)}$	2012	Rg	SIFT + RP	√	√	√
[23]$^{(+,4)}$	2013	CI	Multi-scale SF + SVM	√	√	√
[27]$^{(+,4)}$	2014	CI	Joint detection and estimation + SVM	√	√	√
[28]$^{(+,4)}$	2014	CI	Gabor + covariance + learning	√	×	√
[26]$^{(*,4)}$	2007	CI	Associative memory	√	√	√
[32]$^{(*,4)}$	2008	CI+Rg	SVM + SVR	√	×	√
[35]$^{(*,4)}$	2007	DR+CI	LDA + linear learning	√	√	√
[6]$^{(0)}$	2008	Tr	Tracking using particle filters	√	√	√

Best result approaches (*plus sign*), most used references for benchmarks (*asterisk*). Databases: "0": Own; "1": FacePix [48]; "2": BU [49]; "3": MIT [50]; "4": Pointing'04 [34]
GM geometric model, *FM* flexible model, *AT* appearance template, *CI* classification, *Rg* regression, *Tr* tracking, *DR* dimensionality reduction

the road in frontal position. However, the driver needs to look at the dashboard, the rear-view mirror, and the side-view mirrors which correspond to moving the head to down, up, left, and right positions for a brief time. These positions must be maintained for few seconds; otherwise, they are representative of inattention. We can also conclude that the driver attention is not influenced by the orientation according to the roll angle, which allows us to reduce our degrees of freedom to pitch and yaw angles. According to [32], when one or some head pose labels are considered as a class, the head pose estimation is addressed as a classification problem and if the pose angles are ordered, the problem can be thought as a regression problem. After these observations, we can formulate our problem of estimating head pose to detect driver inattention as the problem of classifying head poses into 3 classes for pitch and 3 classes for yaw presented as follows:

- Pitch: frontal head, up position, down position
- Yaw: frontal head, left profile, right profile

In this work, we study different head pose descriptors able to detect variations in driver head pose and we propose an efficient fusion approach providing a good discrimination of pose variations. Since we address the problem of classifying human heads into discrete poses, we evaluate the ability of these descriptors to represent pose variations by testing their efficiency using the SVM classifier. In the following, we present a brief overview of our global system for monitoring driver vigilance level.

3.1 Global overview of our system for monitoring driver vigilance level

In this subsection, we present an overview of our global system for assessing driver vigilance level, while in the next sections, we focus our attention on studying driver inattention by estimating head pose. The principle of detecting inattention is based on the assumption that driver head is in abnormal position when it is maintained for a certain duration in a non-frontal pose for both pitch and yaw angles. Our system illustrated in Fig. 2 and can be synthesized as follows:

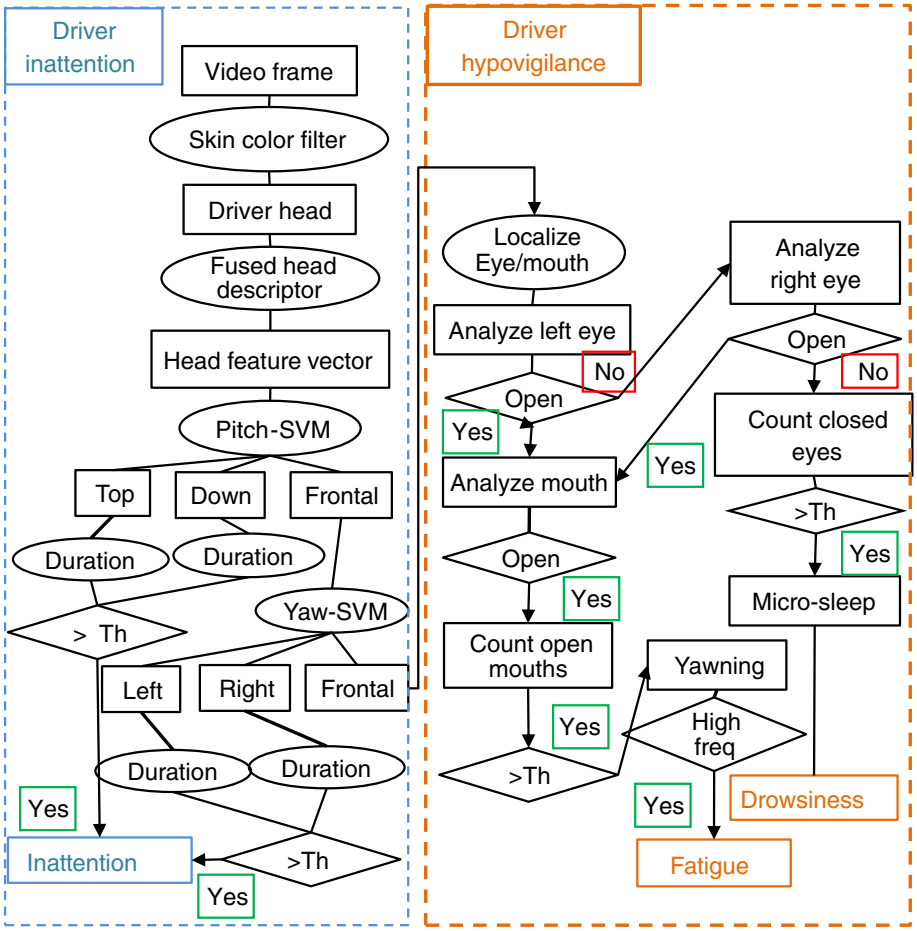

Fig. 2 Our system for monitoring driver vigilance level

Detecting driver inattention:

1. Extract the head from video frame using skin color segmentation.
2. Extract head pose descriptors to obtain a representative feature vector.
3. Apply Pitch-SVM at first since we assume that maintaining head in down position for a certain duration corresponds to the most critical pose and can reveal sleep.
4. If head is in down or up position, observe duration of fixed position and emit inattention warning when it is important.
5. If head is frontal according to pitch, apply Yaw-SVM.
6. If head is in left or right profile, observe duration of fixed position and emit inattention warning when it is important.
7. If head is frontal for both angles, proceed to *hypovigilance detection* (Fig. 2) detailed in [10].

In the following, we focus our attention on studying the most appropriate features to propose a robust fused descriptor representing driver head pose. Moreover, we evaluate the performance of the SVM classifier for estimating head poses.

3.2 Proposed approach for head pose estimation

As mentioned above, we present in this subsection several image descriptors frequently used in literature and judged to be the most representative of head pose variations. Next, we expose feature selection techniques allowing to select the most pertinent attributes among these descriptors. We use the SVM classifier to decide in which class each head image (characterized by its feature vector) is related.

3.2.1 Head pose descriptors

We chose to study four descriptors to characterize head pose variations which are SF, HOG, Haar features, and speeded up robust features (SURF). These descriptors are invariant to the common image transformations corresponding to image rotation, scale changes, and illumination variation. Hence, they respect the requirement (R3) allowing them to be used in order to build a head pose estimator.

- *Steerable filters*: The steerable filters [42] are used due to their ability to analyze oriented structures in images. We have proved their robustness in our previous work [21] for estimating head pose using likelihood parametrized function (LPF). Another motivation is given by their capacity to filter an image at any orientation using only a linear combination of its filtered versions obtained by a small set of basis filters. This concept reduces considerably the processing time. We chose a simple SF corresponding to the derivatives of the circularly symmetric Gaussian function $f(x,y) = \exp(-\frac{(x^2+y^2)}{2\sigma^2})$ to describe head poses. In this case, the basis filters are the first derivatives of f according to x and y and correspond to the filters at orientations $0°$ and $90°$, respectively. Hence, a filtered image by an orientation θ can be expressed by $R_1^\theta = \cos(\theta)R_1^{0°} + \sin(\theta)R_1^{90°}$, where $R_1^{0°}$ and $R_1^{90°}$ correspond to the image filtered by the two basis filters (see [21] for more details). The performance of the SF depends of the number of filters applied on the image and also the orientation of each filter. We have conducted several experiments, and we find that the following values provide the best result:
  - Number of filers $= 2$
  - Size of reduced patch image $= 15$
  - Angular displacement $= 50°$ (i.e., Filter 1 at $\theta = 0°$ and Filter 2 at $\theta = 50°$)
  - SF feature size: $450\ (15 \times 15 \times 2)$
- *Haar features*: The Haar features [43] represent a dense overcomplete representation using wavelets. The two-dimensional Haar decomposition of a square image with n^2 pixels consists of n^2 wavelet coefficients corresponding to a distinct Haar wavelet. The first wavelet is the mean pixel intensity value of the whole image; the rest of the wavelets are computed as the difference in mean intensity values of horizontally, vertically, or diagonally adjacent squares. The contrast variances between the pixel groups are used to determine relative light and dark areas. The Haar coefficient of a particular Haar wavelet is computed as the difference in average pixel value between the image pixels in the black and white regions. From the experiment, we find that the following number of wavelets provides the best estimation of head pose:
  - Number of wavelet $= 32$
  - Haar feature size : $1024\ (32 \times 32)$
- *Speeded up robust features*: SURF [44] is a fast and enhanced version of SIFT. It is an algorithm for local, similarity invariant representation and comparison. The algorithm is structured into three steps: detecting interest point, building the descriptor for each interest point, and performing descriptor matching. In our paper, we use an adapted version of SURF since we do not need to perform descriptor matching allowing image comparison. Hence, after obtaining the descriptors of interest points, we sort them according to their orientations. Then, we divide the sorted descriptors in groups before computing

the average of elements of each group. The descriptor size and the number of groups for decomposition are the parameters that influence the SURF performance. We find experimentally that these following values provide the best result of head pose estimation:

  - Descriptor dimension = 64
  - Number of descriptors = 4
  - SURF feature size: 256 (64×4)

- *Histogram of oriented gradients*: The basic idea of HOG [45] is that object appearance and shape can be represented by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. This concept can be implemented by splitting the image into small regions (cells) with a defined size adapted to the size and resolution of the object. For each cell, the occurrences of gradient orientation over all the pixels are accumulated in a local histogram. Each orientation histogram divides the gradient angle range into a fixed number of bins. The parameters influencing the HOG are the number of cells per rows and per column in addition to the number of bins. The best performance for head pose estimation using HOG is given by the following configuration:

  - Number of cells per image row = 3
  - Number of cells per image column = 3
  - Number of histogram bins = 10
  - HOG feature size: 90 $(3 \times 3 \times 10)$

3.2.2 Feature selection techniques

In our driver head pose estimator, different descriptors are used to extract image features. We choose to extract features as diverse and rich as possible in order to take advantage of their complementarity, but we did not ignore the possibility of redundancy. The aim of the feature selection step is to find a compact, relevant, and consistent set of features for classification task. Feature selection searches through all possible combinations of attributes in the data to find which subset works best for prediction by employing two tasks: search method and attribute evaluator. The search method generates subsets of features and attempts to find an optimal subset while the attribute evaluator determines how good a proposed feature subset is, returning some measures of goodness to the search method. We have evaluated three popular search methods (BestFirst, GreedyStepwise, and Ranker), and we find that the Ranker provides the best results. This can be explained by the individual evaluation of features by the Ranker instead of subset evaluation performed by the two other methods. Therefore, we study three attribute evaluators that can be associated with the Ranker method. The gain ratio (GR) evaluates the worth of an attribute by measuring the gain ratio with respect to the class. The OneR performs evaluation using a simple classification that generates one rule for each predictor in the data and selects the rule with the smallest total error as its "one rule." The evaluation performed by the ReliefF (RF) consists on repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. In Section 4, we will evaluate these feature selection techniques and the best one will be retained to construct the fused feature vector of head pose.

3.2.3 SVM classifier

The SVM is based on structural risk minimization theory [46]. Given a set of training vectors $(x_1, y_1), \ldots, (x_l, y_l)$ composed of observations $x_i \epsilon R^n$ and interpretations $y_i \epsilon \{-1, +1\}$, the binary SVM optimizes a hyperplane to separate positive and negative training samples using their feature vectors. Different kernels could be used to map the classification problem to a higher dimensional feature space. For multiclass problems, the original learning problem must be decomposed into a series of binary learning problems. A standard solution for this problem is the one-against-all approach, which constructs one binary classifier for each class. A faster and more accurate approach for small number of classes is the pairwise classification [47] which is based on transforming the c-class problem into $\frac{c(c-2)}{2}$ binary problems, one for each pair of classes. For our experiments, we used the pairwise classification multiclass SVM with RBF kernel, available in the free software WEKA.

4 Experimental results

Since there is no public database containing various driver head poses, we have acquired video sequences representing a driver in different head poses to perform our experiment. However, this is not enough to prove the robustness of our system which requires to be compared with the state-of-the-art approaches. To guarantee unbiased comparison, we perform experiments on the public Pointing'04 database [34], which is the most used database in literature for head pose estimation [14]. Moreover, this database could represent the driving environment since the distance between the subjects and the camera is comparable to the one between the driver and the dashboard, where the camera is mounted.

4.1 Experiments on public database

4.1.1 The Pointing'04 database

The Pointing'04 database contains head poses labeled according to pitch and yaw angles, and it is composed from 15 sets of near-field images. Each set contains two series of 93 images of the same person at 93 discrete head poses [34]. These ones span both pitch and yaw

included in the set $\{0; \pm 90; \pm 60; \pm 30; \pm 15\}$ and the interval $[-90°; +90°]$ with a displacement of 15°, respectively. The subjects range in age from 20 to 40 years old, five possessing facial hair and seven wearing glasses. Each subject was photographed against a uniform background, and head pose ground truth was obtained by directional suggestion. In Fig. 3, we show the frontal pose (pitch = yaw = 0°) of the thirty Pointing'04 folds.

We perform several series of experiments on the Pointing'04 database using 80 % as training set (2232 images), 10 % as validation set (279 images), and 10 % as test set (279 images). In the following, the results are given on the test set. We first present the results of the optimization step to fix the best system parameters and also the performance of separate and combined descriptors.

4.1.2 System optimization

In our paper, we deal with the problem of estimating driver head pose according to two degrees of freedom (pitch and yaw angles) in order to identify three classes (cl) for each angle. However, the Pointing'04 database is composed of 9 poses for pitch and 13 poses for yaw. We propose to cluster the poses into three classes for pitch and three classes for yaw to match our problem formulation. For SVMs, we find that the RBF kernel with $\gamma = 0.15$ provides the best classification results. The optimal values of descriptor parameters are presented in the Section 3.2.1 (SF = 450, HOG = 90, SURF = 256, and Haar = 1024).

In Table 2, we show the results of each descriptor evaluated separately in addition to all possible combinations of these descriptors (two, three, and four elements), in terms of accuracy and kappa statistic for both pitch and yaw angles. The accuracy (Acc) is the overall correctness of the model, and it is calculated as the sum of correct classifications divided by the total number of instances, while the kappa statistic (κ) is a chance-corrected measure of agreement between the classifications and the true classes. The highest values of Acc and κ correspond to the best system performance. We also show the processing time in seconds (time) needed to classify one image by pitch-SVM and yaw-SVM. It is obvious that increasing the number of descriptors conduce to increase the processing time of one frame. However, the Haar features are more expensive in terms of computational time because of the large size of their feature vector. From this table, we observe that SF features provide the best result when the descriptors are evaluated separately. The best result of combining two and three descriptors are given respectively by the feature vectors SF, HOG and SF, HOG, SURF. When we combine the four descriptors, the results are less advantageous than those of the best combination of two or three descriptors. This could be explained by an interaction between the attributes of the overall feature vector, which produces contradictions at the decision process performed by the multi-class SVM. This problem could be solved by introducing a feature selection step on the combined descriptor

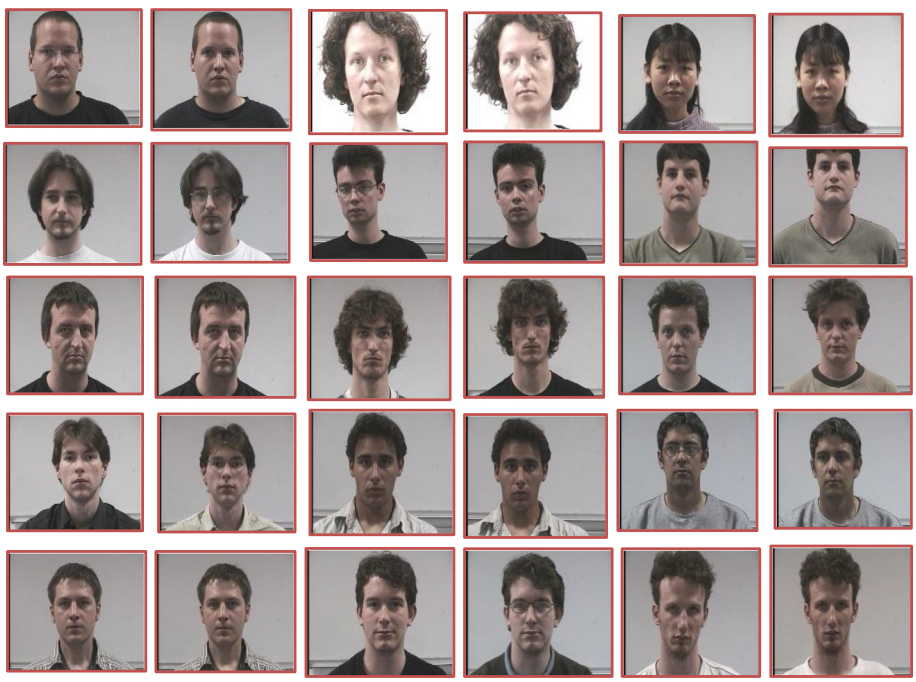

Fig. 3 The frontal pose of the thirty Pointing'04 folds

Table 2 Evaluation of separate descriptors and all possible combinations of them on the test set

	3 classes for pitch-SVM			3 classes for yaw-SVM		
Descriptor	Acc.	κ	Time	Acc.	κ	Time
SF	*87.2*	*0.80*	0.05	*94.3*	*0.91*	0.03
HOG	85.6	0.77	0.01	94	0.90	0.01
SURF	83.8	0.75	0.04	93.7	0.90	0.04
Haar	85.9	0.78	0.2	93	0.89	0.2
(SF, HOG)	*89.3*	*0.83*	0.07	*96.3*	*0.94*	0.06
(SF, SURF)	89.0	0.83	0.13	95.3	0.92	0.12
(SF, Haar)	86.7	0.79	0.5	94.7	0.91	0.47
(HOG, Haar)	88.8	0.82	0.24	94.5	0.91	0.21
(HOG, SURF)	87.2	0.80	0.06	94.9	0.92	0.06
(SURF, Haar)	87.3	0.80	0.35	94.6	0.91	0.32
(SF, HOG, SURF)	*89.1*	*0.83*	0.15	*95.4*	*0.93*	0.11
(SF, HOG, Haar)	85.6	0.77	0.28	95.1	0.92	0.29
(SF, SURF, Haar)	77.9	0.64	0.53	92.3	0.88	0.48
(HOG, SURF, Haar)	87.8	0.81	0.19	94.9	0.92	0.17
(SF, HOG, SURF, Haar)	*87.5*	*0.80*	*0.53*	*94.9*	*0.91*	*0.52*

Italic values in Table 2: Best results obtained by all possible combinations of one, two, three and four descriptors

that allows us to keep the most relevant attributes and reduce the processing time.

4.1.3 Evaluating feature selection techniques

In Table 3, we show the results of evaluating the feature selection techniques presented in Section 3.2.2 using the Ranker as search method, which is equivalent to the evaluation of the performance of three attribute evaluators (Attr. Eval.): GR, OneR, and RF. A first set of tests is conducted on the best combination of three descriptors (SF, HOG, SURF) using the 400 most relevant variables from a total of 796, which corresponds to a reduction of attributes by half. According to Table 3, the best result of these tests is given by the ReliefF algorithm. Hence, in a second set of tests, we apply the ReliefF attribute evaluator on the combination of the four descriptors and we evaluate the impact of varying the number of selected variables on the system performance. We chose to retain 400 relevant variables using ReliefF as the best configuration, since it provides a good compromise between processing time, accuracy, and kappa coefficient. In the last test of this subsection, we show the result of the best configuration when using the k-fold cross validation (CV) process with $k = 10$. The cross-validation reorders the database and divided it into 10 equal parts. Then, for each iteration, one part is used for the test and the other nine parts for learning the classifier. All results are collected and averaged at the end of the cross-validation. From the last line of Table 3, we note that the result obtained by cross-validation (CV) improves the conventional test, which proves that the

Table 3 Performance on the test set of the studied attribute evaluators on the best combination of three and four descriptors using the Ranker search method

		3 classes for pitch-SVM			3 classes for yaw-SVM		
Descriptor	Attr. Eval.	Acc.	κ	Time	Acc.	κ	Time
(SF, HOG, SURF)	(GR,400/796)	87.0	0.79	0.06	94.5	0.91	0.05
(SF, HOG, SURF)	(OneR,400/796)	86.4	0.79	0.05	94.6	0.91	0.05
(SF, HOG, SURF)	(RF,400/796)	90.1	0.84	0.05	95.4	0.93	0.05
(SF, HOG, SURF, Haar)	(RF,600/1820)	90.5	0.85	0.34	96.7	0.94	0.32
(SF, HOG, SURF, Haar)	(RF,400/1820)	90.5	0.85	0.09	96.6	0.94	0.08
(SF, HOG, SURF, Haar)	(RF,200/1820)	88.1	0.80	0.06	94.2	0.91	0.05
(SF, HOG, SURF, Haar)	(RF,400/1820)	91.9	0.87	CV	96.4	0.94	CV

proposed approach allows a good classification of poses even when varying samples.

To visualize which descriptors are more pertinent, we present in Table 4 the total number of each descriptor features (TN), the number of selected features from each descriptor (SN), the rate of the features extracted from each descriptor (FiD), and the participation rate of each descriptor in the fusion (DiF). If we analyze the column (FiD), we can observe that the SF and HOG are the most pertinent descriptors since more than 50 % of their features are selected while less than 10 % of Haar and SURF features are selected. Moreover, the analysis of the column (DiF) shows that the SF features are the most present ones in the final descriptor with more than 65 % of features.

4.1.4 Comparison with existing approaches

The major part of approaches using the Pointing'04 database for evaluation uses its standard representation of poses which corresponds to 9 angles for pitch and 13 angles for yaw. To provide a fair comparison, we increase the number of classes considered by our system in order to respect the standard representation. Therefore, in this experiment, the pitch-SVM and yaw-SVM must classify 9 and 13 head angles, respectively. Moreover, we present the results in terms of angular mean absolute errors (MAE) between the estimated and ground-truth angles for botch pitch and yaw, since all considered approaches for comparison use them. In Table 5, we present the result of our approach compared to the best approaches in literature and also to the most referenced ones (see Table 1). In [26], Gourier et al. measure the human performance for estimating head poses on the Pointing'04 database and find that the angular MAE correspond to 11° for pitch and 11.9° for yaw. From Table 5, we can conclude that our head pose estimator is more precise than the human performance. As can be seen, it provides the best results among all studied approaches.

In the next experiment, we can show the result obtained when using our head pose estimation technique on real video sequence representing driver with various head poses.

Table 4 Number and percentage of selected features from the fused descriptor

		3 classes for pitch-SVM			3 classes for yaw-SVM		
Descriptor	TN	SN	FiD	DiF	SN	FiD	DiF
SF	450	263	58 %	66 %	275	61 %	69 %
HOG	90	62	68 %	16 %	70	78 %	18 %
SURF	256	0	0 %	0 %	11	4 %	2 %
Haar	1024	75	7 %	18 %	44	4 %	11 %

TN the total number of descriptor features, *SN* the number of selected descriptor features after fusion, *FiD* the rate of SN in the descriptor, *DiF* the rate of SN in the fusion

Table 5 Comparison with existing techniques in terms of angular MAE using Pointing'04 database with 9 poses for pitch and 13 poses for yaw

Approach	Year	Pitch	Yaw
Our approach	2015	*4.6°*	*6.1°*
HOG + structural SVM [27]	2014	5.25°	5.91°
Dense SIFT + RP [24]	2012	5.84°	6.05°
Kernel PLS regression [31]	2012	6.61°	6.56°
Gabor + covariance + learning [28]	2014	7.14°	6.24°
Multi-scale SF + SVM [23]	2013	8°	6.9°
SF + LPF [21]	2012	8°	9.37°
Geometric approach (golden ratio) [18]	2013	13.6°	9.6°
Cropped head + SVM + SVR [32]	2008	7.69°	9.23°
Human performance [26]	2007	11°	11.9°
Associative memory [26]	2007	15.9°	10.3°
LDA + linear learning [35]	2007	30.7°	19.1°

Italic values in Table 5: Best angular MAE

4.2 Experiment on driver video sequence

We have acquired a video sequence, as shown in Fig. 4, with a cheap visible spectrum phone camera representing a driver in various head poses and composed from 2636 video frames. Each frame has a resolution of 1280 × 720 pixels.

Fig. 4 Driver acquisition system with cheap phone camera

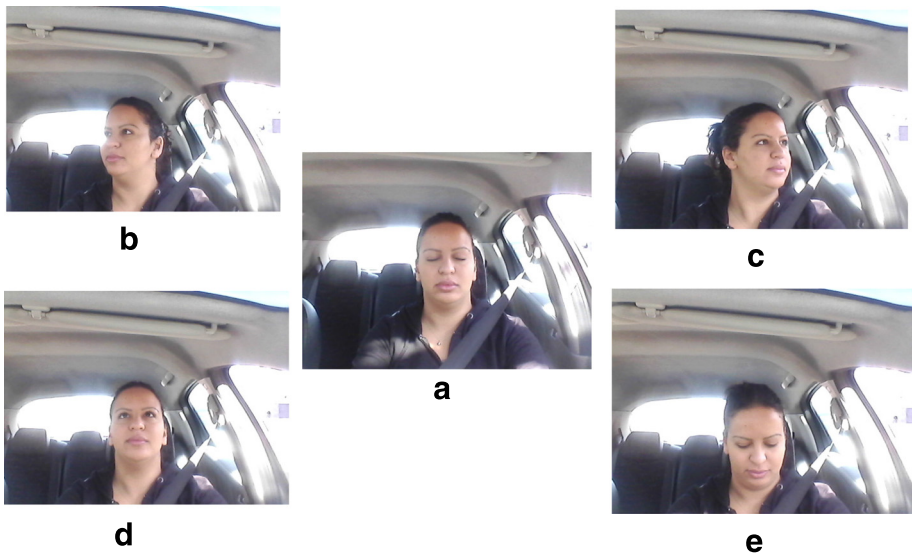

Fig. 5 Examples of driver video frames. **a** Frontal position (pitch and yaw). **b** Left profile (yaw). **c** Right profile (yaw). **d** Up position (pitch); **e** Down position (pitch)

Since we deal with estimating driver head pose, we have annotated our sequence using three classes for pitch and three classes for yaw. Figure 5 represents an example of frames corresponding to each class according to the pitch and yaw angles. The result obtained when applying our head pose estimator on driver video sequence using the best parameters determined in Section 4.1 is given by the second row in Table 6. The first result in this table reports the same experiment applied on Pointing'04 database, previously presented in line 5, Table 3.

Even if our sequence is acquired in real conditions, the results obtained in this experiment are better than the one obtained on Pointing'04 database. This fact might be explained by the inherent problem of annotation caused by the important number of poses in the Pointing'04 database while in our sequence, we annotate 3 poses for each angle.

5 Conclusions

In this paper, we have proposed a head pose estimation approach using a single camera in order to identify driver inattention. Our approach is based on a robust fusion of multiple significant descriptors (SF, HOG, Haar, and SURF) in order to construct an efficient feature vector representing head pose variations. Then, two SVMs are learned to classify the feature vectors according to pitch and yaw angles. Our head pose estimator is not restricted to monitoring driver inattention level and can also be used by diverse applications requiring knowledge of human activity such as human-machine interfaces and game industry. Before applying our estimator, it is important to identify the number of poses that must be estimated for each angle depending on the application requirements. In our paper, we use three classes for both pitch and yaw angles since we deal with the problem of estimating driver head pose to determine its inattention level. Since no public database is available for estimating driver head pose, we perform several experiments on the public database Pointing'04 to validate our approach and compare it with the recent and the most cited state-of-the-art techniques. We have also acquired a video sequence using a cheap visible spectrum camera representing a driver in various attention levels and we find that our head pose estimator can achieve an accuracy of 97.5 % for pitch and 98.2 % for yaw.

As future work, we can improve our global system for monitoring driver vigilance level by adding a gaze estimation approach in order to determine driver focus of attention. Since we use a visible spectrum camera, the acquisition can be perturbed at night and the usage of IR light could be considered to resolve this problem.

Table 6 Results of our head pose estimation on the driver video sequence using 3 classes for pitch and yaw

	3 classes for pitch-SVM			3 classes for yaw-SVM		
Database	Accuracy	Kappa	Time	Accuracy	Kappa	Time
Pointing'04	90.5	0.85	0.09	96.6	0.94	0.08
Our Sequence	97.5	0.96	0.03	98.2	0.98	0.02

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Ibn Zohr University, Morocco, BP 32/S, Agadir, Morocco. [2] LRIT-CNRST 29, Faculty of Sciences, Mohammed V University, Ibn Battouta Avenue, Rabat, Morocco. [3] LITIS, INSA-Rouen, Avenue de l'Université, Saint-Etienne-du-Rouvray, France. [4] LGS, ENSA-Kenitra, Ibn Tofail University, Avenue de l'Université, Kenitra, Morocco.

Received: 20 February 2015 Accepted: 4 January 2016
Published online: 09 January 2016

References

1. State of the Road, A fact sheet of CARRS-Q, Centre of Accident Research and Road Safety-Queensland, Queensland, Australia. Date accessed: March 2013 (2012). http://www.carrsq.qut.edu.au/publications/corporate/hooning_fs.pdf
2. C Berka, DJ Levendowski, MN Lumicao, A Yau, G Davis, VT Zivkovic, RE Olmstead, PD Tremoulet, PL Craven, EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviat. Space Environ. Med. **78**, 231–244 (2007)
3. TW Victor, JL Harbluk, JA Engstrom, Sensitivity of eye-movement measures to in-vehicle task difficulty. Transp. Res. Part F: Traffic Psychol. Behav. **8**, 167–190 (2005)
4. N Papanikolopoulos, M Eriksson, Driver fatigue: a vision-based approach to automatic diagnosis. Transpo. Res. Part C: Emerg. Technol. **9**, 399–413 (2001)
5. B Hrishikesh, S Mahajan, A Bhagwat, T Badiger, D Bhutkar, S Dhabe, L Manikrao, Design of DroDeASys (drowsy detection and alarming system). Adv. Comput. Algo. Data Anal. **14**, 75–79 (2009)
6. E Murphy-Chutorian, MM Trivedi, in *Proceeding of the Intelligent Vehicles Symposium (IV)*. Hyhope: Hybrid head orientation and position estimation for vision-based driver head tracking (IEEE, Eindhoven, The Netherlands, 2008), pp. 512–517
7. KS Huang, MM Trivedi, in *Proceedings of the 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems, in Conjunction with the 3rd International Conference on Computer Vision Systems*. Driver head pose and view estimation with single omnidirectional video stream (IAPR, Graz, Austria, 2003), pp. 44–51
8. T Azim, MA Jaffar, M Ramzan, AM Mirza, in *Signal Processing, Image Processing and Pattern Recognition. Communications in Computer and Information Science, vol. 61*. Automatic fatigue detection of drivers through yawning analysis (Springer, Jeju Island, Korea, 2009), pp. 125–132
9. JF May, LC Baldwin, Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transpo. Res. Part F: Traffic Psychol. Behav. **12**, 218–224 (2009)
10. N Alioua, A Amine, M Rziza, D Aboutajdine, in *Computer Analysis of Images and Patterns. Lecture Notes in Computer Science, vol. 6855*. Driver's fatigue and drowsiness detection to reduce traffic accidents on road (Springer, Seville, Spain, 2011), pp. 397–404
11. L Bergasa, J Nuevo, M Sotelo, R Barea, E Lopez, Real-time system for monitoring driver vigilance. IEEE Transac. Intell. Transpo. Syst. **7**, 63–77 (2006)
12. S Gurbuz, E Oztop, N Inoue, Model free head pose estimation using stereovision. Pattern Recognit. **45**, 33–42 (2012)
13. L Li, K Werber, CF Calvillo, KcD Dinh, A Guarde, A Konig, in *Online Conference on Soft Computing in Industrial Applications. Advances in Intelligent Systems and Computing, vol. 223*. Multi-sensor soft-computing system for driver drowsiness detection (Springer, 2012)
14. E Murphy-Chutorian, MM Trivedi, Head pose estimation in computer vision: a survey. IEEE Transac. Pattern Anal. Mach. Intell. (PAMI). **31**, 607–626 (2009)
15. X Liu, H Lu, H Luo, in *Proceeding of the 16th International Conference on Image Processing (ICIP)*. A new representation method of head images for head pose estimation (IEEE, Cairo, Egypt, 2009), pp. 3585–3588
16. E Murphy-Chutorian, MM Trivedi, Head pose estimation and augmented reality tracking: an integrated system and evaluation for monitoring driver awareness. IEEE Transac. Intell. Transpo. Syst. **11**, 300–311 (2010)
17. A Dahmane, S Larabi, C Djeraba, IM Bilasco, in *Proceeding of the 21st International Conference on Pattern Recognition (ICPR)*. Learning symmetrical model for head pose estimation (IEEE, Tsukuba, Japan, 2012), pp. 3614–3617
18. G Fadda, G Marcialis, F Roli, L Ghiani, in *International Conference on Image Analysis and Processing (ICIAP). Lecture Notes in Computer Science, vol. 8156*. Exploiting the golden ratio on human faces for head-pose estimation (Springer, Naples, Italy, 2013), pp. 280–289
19. LP Morency, J Whitehill, J Movellan, Monocular head pose estimation using generalized adaptive view-based appearance model. Image Vis. Comput. **28**, 754–761 (2010)
20. B Ma, X Chai, T Wang, A novel feature descriptor based on biologically inspired feature for head pose estimation. Neurocomputing. **115**, 1–10 (2013)
21. N Alioua, A Amine, A Bensrhair, M Rziza, D Aboutajdine, in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*. Head pose estimation based on steerable filters and likelihood parametrized function (EURASIP, Marrakech, Morocco, 2013)
22. M Demirkus, B Oreshkin, JJ Clark, T Arbel, in *Proceeding of the 18th International Conference on Image Processing (ICIP)*. Spatial and probabilistic codebook template based head pose estimation from unconstrained environments (IEEE, Brussels, Belgium, 2011), pp. 573–576
23. V Jain, JL Crowley, in *Proceeding of the 18th Scandinavian Conference on Image Analysis*. Head pose estimation using multi-scale gaussian derivatives (Springer, Espoo, Finlande, 2013)
24. HT Ho, R Chellappa, in *Proceeding of the 19th International Conference on Image Processing (ICIP)*. Automatic head pose estimation using randomly projected dense sift descriptors (IEEE, Orlando, Florida, USA, 2012), pp. 153–156
25. R Munoz-Salinas, E Yeguas-Bolivar, A Saffiotti, R Medina-Carnicer, Multi-camera head pose estimation. Mach. Vis. Appl. **23**, 479–490 (2012)
26. N Gourier, J Maisonnasse, D Hall, JL Crowley, in *Multimodal Technologies for Perception of Humans. Lecture Notes in Computer Science, vol. 4122*. Head pose estimation on low resolution images (Springer, Uthampton, UK, 2007), pp. 270–280
27. K He, L Sigal, S Sclaroff, in *European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, vol. 8692*. Parameterizing object detectors in the continuous pose space (Springer, Zurich, Switzerland, 2014), pp. 450–465
28. B Ma, A Li, X Chai, S Shan, CovGa: a novel descriptor based on symmetry of regions for head pose estimation. Neurocomputing. **143**, 97–108 (2014)
29. Y Ma, Y Konishi, K Kinoshita, S Lao, M Kawade, in *Proceeding of the 18th International Conference on Pattern Recognition (ICPR)*. Sparse Bayesian regression for head pose estimation (IEEE, Hong Kong, China, 2006), pp. 507–510
30. G Fanelli, M Dantone, J Gall, A Fossati, LV Goo, Random forests for real time 3D face analysis. Int. J. Comput. Vision (IJCV). **101**, 437–458 (2013)
31. M Al-Haj, J Gonzalez, LS Davis, in *Proceeding of Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. On partial least squares in head pose estimation: how to simultaneously deal with misalignment (IEEE, Providence, USA, 2012), pp. 2602–2609
32. G Guo, Y Fu, CR Dyer, TS Huang, in *Proceeding of the 19th International Conference on Pattern Recognition (ICPR)*. Head pose estimation: classification or regression? (IEEE, Tampa, Florida, USA, 2008), pp. 1–4
33. E Ricci, JM Odobez, in *Proceeding of the 18th International Conference on Image Processing (ICIP)*. Learning large margin likelihoods for realtime head pose tracking (IEEE, Cairo, Egypt, 2009), pp. 2593–2596
34. N Gourier, D Hall, J Crowley, in *Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*. Estimating face orientation from robust detection of salient facial structures (IEEE, Cambridge, UK, 2004)
35. Z Li, Y Fu, J Yuan, TS Huang, Y Wu, in *Proceeding of the International Conference on Multimedia and Expo (ICME)*. Query driven localized linear discriminant models for head pose estimation (IEEE, Beijing, China, 2007), pp. 1810–1813
36. L Bretzner, M Krantz, in *Proceeding of the International Conference on Vehicular Electronics and Safety*. Towards low-cost systems for measuring visual cues of driver fatigue and inattention in automotive applications (IEEE, 2005), pp. 161–164
37. MM Trivedi, T Gandhi, J McCall, Looking-in and looking-out of a vehicle: computer-vision-based enhanced vehicle safety. IEEE Trans. Intell. Transpo. Syst. **8**, 108–120 (2007)
38. Smart Eye AntiSleep 4. Date accessed: December 2013 (2013). http://smarteye.se/wpcontent/uploads/2014/12/Product-Sheet-AntiSleep.pdf
39. V Vogelhuber, C Schmid, in *Proceeding of the International Conference on Pattern Recognition (ICPR)*. Face detection based on generic local descriptors and spatial constraints (IEEE, Barcelona, Spain, 2000), pp. 1084–1087
40. L Wiskott, JM Fellous, N Kruger, CVD Malsburg, Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI). **19**, 775–779 (1997)

41. J Wu, MM Trivedi, A two-stage head pose estimation framework and evaluation. Pattern Recognit. **41**, 1138–1158 (2008)
42. WT Freeman, EH Adelson, The design and use of steerable filters. IEEE Trans. Pattern Anal. Mach. Intell. **13**, 891–906 (1991)
43. C Papageorgiou, T Poggio, A trainable system for object detection. Int. J. Comput. Vis. **38**, 15–33 (2000)
44. H Bay, A Ess, T Tuytelaars, LV Gool, SURF: speeded up robust features. Comput. Vision and Image Underst. **110**, 346–359 (2008)
45. N Dalal, B Triggs, in *Proceeding of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1*. Histograms of oriented gradients for human detection (IEEE, San Diego, CA, USA, 2005), pp. 886–893
46. C Burges, Tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**, 121–167 (1998)
47. SH Park, J Furnkranz, in *Machine Learn. Lecture Notes in Computer Science, vol. 4701*. Efficient pairwise classification (Springer, Warsaw, Poland, 2007), pp. 658–665
48. J Black, M Gargesha, K Kahol, P Kuchi, S Panchanathan, in *Internet Multimedia Systems II (ITCOM)*. A framework for performance evaluation of face recognition algorithms (SPIE, Boston, USA, 2002)
49. ML Cascia, S Sclaroff, V Athitsos, Fast, reliable head tracking under varying illumination: an approach based on registration of textured-mapped 3D models. IEEE Trans. Pattern Anal. Machine Intell. (PAMI). **22**, 322–336 (2000)
50. LP Morency, A Rahimi, T Darrell, in *Proceeding of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Adaptive view-based appearance model (IEEE, Madison, USA, 2003), pp. 803–810