Applied Soft Computing 45 (2016) 98–107

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A reduced feature set for driver head pose estimation

Katerine Diaz-Chito [a,b,*], Aura Hernández-Sabaté [a,b], Antonio M. López [a,b]

[a] Centre de Visió per Computador, Spain
[b] Universitat Autònoma de Barcelona, Spain

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form 23 March 2016
Accepted 19 April 2016
Available online 27 April 2016

Keywords:
Head pose estimation
Driving performance evaluation
Subspace based methods
Linear regression

ABSTRACT

Evaluation of driving performance is of utmost importance in order to reduce road accident rate. Since driving ability includes visual-spatial and operational attention, among others, head pose estimation of the driver is a crucial indicator of driving performance. This paper proposes a new automatic method for coarse and fine head's yaw angle estimation of the driver. We rely on a set of geometric features computed from just three representative facial keypoints, namely the center of the eyes and the nose tip. With these geometric features, our method combines two manifold embedding methods and a linear regression one. In addition, the method has a confidence mechanism to decide if the classification of a sample is not reliable. The approach has been tested using the CMU-PIE dataset and our own driver dataset. Despite the very few facial keypoints required, the results are comparable to the state-of-the-art techniques. The low computational cost of the method and its robustness makes feasible to integrate it in massive consume devices as a real time application.

1. Introduction

Driver fatigue/drowsiness and distraction are known to be behind a large amount of traffic accidents. Accordingly, different systems have been developed to detect such situations [1–4]. Distractions are specially challenging because many times are difficult to predict in advance since they may be due to sudden events in the environment or in the cabin. Indeed, a more general challenge including distractions is driving performance. Evaluation of driving performance is of utmost importance in order to reduce road accident rate. Behavior analysis while driving generally points out the abilities of the driver, which include cognitive (attention, executive functions, and memory) and (visual-spatial) perception skills, as well as, their fatigue levels or attention capability [5]. These abilities can be analyzed from several points of view, either measuring non-visual features like heart rate variability [6], or analyzing visual features such as eye blinking behavior [7], gaze direction estimation [8] or analysis of motion of the hands [9] or the feet [10]. In particular, head pose is a crucial indicator of driver's field of view and his/her attention focus [11] and, like most of the indicators named above, it deserves further consideration.

Head pose estimation is a challenging problem in itself due to the variability introduced by factors such as driver's identity and expression, cabin and outdoor illumination, etc. [12]. In fact, during the last decade there has been an increasing interest in developing methods to estimate head pose [13] for different applications such as security and surveillance systems [14], meeting rooms [15], intelligent wheelchair systems [16], and driving monitoring [1,17]. In the particular case of driving performance, when drivers are paying attention to the road ahead, their facial direction is within approximately $\pm 15^{\circ}$ from the straight normal [18]. Thus, the yaw angle of the driver could contribute to determine if he/she perceives road elements such as traffic lights, roundabouts, crosswalks and the attention he/she devotes. Accordingly, in this paper we focus on the computation of such angle from still images. Such a method can be very useful as part of a multi-cue system [19,20] for early detection of abnormal driving performance in common situations. For example, if the driver does not pay attention to the correct direction in a roundabout, or if he/she attends in a crosswalk but does not see a pedestrian crossing on, an advanced driver assistance system (ADAS) combining driving monitoring and exterior hazard assessment, can decide to elevate the warning level or even braking the car.

1.1. Related work and contribution

Murphy-Chutorian and Trivedi [13] divide head pose estimation methods in 8 categories. Three of them are regression methods, geometric methods and manifold embedding ones.

Regression methods apply a regression model on a training set in one or more directions (angles). The regression model usually

* Corresponding author at: Centre de Visió per Computador, Spain.
E-mail addresses: kdiaz@cvc.uab.es (K. Diaz-Chito), aura@cvc.uab.cat (A. Hernández-Sabaté), antonio@cvc.uab.cat (A.M. López).

http://dx.doi.org/10.1016/j.asoc.2016.04.027
1568-4946/

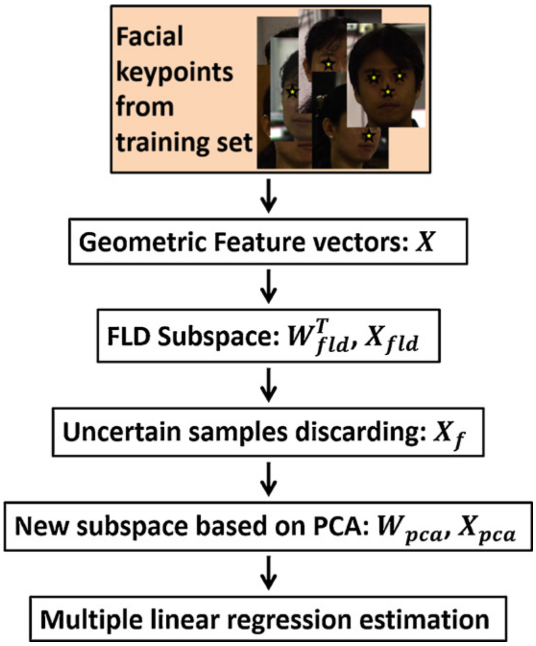

Fig. 1. Workflow of the system training.

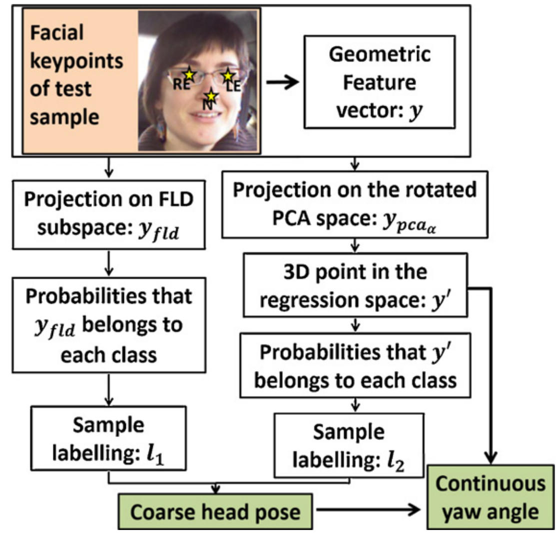

Fig. 2. Workflow of the test step.

is a non-linear one, such as, support vector regressors [21], sparse Bayesian regression [22] or neural networks [23]. One drawback of these methods is that they take into account the whole face image, so that its high dimensionality decreases the efficiency. In some cases, the dimensionality can be reduced, like when the face is well localized. Still, this high dimensionality makes not clear if an specific regression tool is capable of learning the proper curve modeling the directions arch.

Geometric methods are an alternative to directly explode most influencing properties on human head pose estimation, which are usually based on human perception. These features can be divided on two types, those based on face appearance, such as, orientation information of head images [24,25], skin color histogram [26] or facial contours [27], and those relying on a set of (usually 5–6) local facial keypoints [28]. In the first case, computational cost is still high, since they need to analyze the whole face image. In the second case, facial features detection needs to be highly precise and accurate. To overcome the limitations of a single category method, manifold embedding methods are usually combined with them, gaining in accuracy [12,29,30].

In the same fashion, this paper combines the three methods explained above to estimate the continuous angle of head pose of a driver. Roughly, given an image of the driver's head, we rely on a small set of geometric features computed from just three representative facial keypoints, namely the center of the eyes and the nose tip. Our method is based on a combination of subspace projections, as Fisher's linear discriminant (FLD) and principal component analysis (PCA), as well as multiple linear regression adjusted for each pose interval. Figs. 1 and 2 sketch the main steps of the method, split in training the system and testing new samples. For the training (Fig. 1), from a set of samples, we extract the facial keypoints to compute a geometric feature vector. A projection of the samples on a FLD allows the system to suppress some samples not useful to train. Then, the new set of samples is projected on another subspace based on PCA and a multiple linear regression is computed to estimate the regression parameters. When a new sample is given (Fig. 2), it is projected and then classified on the FLD subspace and on the one based on PCA. A combination of both classifications gives us the final coarse yaw angle estimation while the regression parameters serve to compute the continuous yaw angle. Besides, the method integrates a mechanism to self-evaluate the likelihood of the generated hypothesis and discard non likely poses by comparing the discrete angle obtained from FLD and the continuous angle from the regression.

The analysis of the results assessing the reliability of the method shows that, although the very few facial keypoints required, the approach has high accuracy and precision, which makes it comparable to the methods present in the literature. Besides, the computational cost is as low as it can run in real time, making easy to integrate it in massive consume devices such as tablets or mobiles and be part of a multi-cue system for driving performance evaluation. As well, the robustness of the method against noise in the facial keypoint detection is proven.

The remains of the paper are organized as follows. Section 2 describes the mathematical tools involved in the driver's yaw angle estimation. Section 3 describes the detection of the facial keypoints and the geometric features derived from them while Section 4 presents the workflow of the method. The experimental setting and the measures used to assess the reliability of the method are detailed in Section 5. Results and their analysis, are shown in Section 6, while the method is compared with the ones in the literature in Section 7. Last section, 8, is devoted to conclude the paper with some final remarks.

2. Mathematical tools

In this section, we explain the mathematical tools used along the method.

2.1. Subspace based methods

Statistical methods based on subspaces have been broadly used in computer vision and related fields for recognition and classification tasks due to their appealing capabilities and good practical behavior. Among the most popular methods we find FLD and PCA.

FLD computes a linear transformation that minimizes the scatter of the samples within each class, while maximizing the scatter between classes. The main goal of FLD is to find a projection matrix, W_{fld}, of the linear subspace W that maximizes the Fisher's criterion:

$$W_{fld} = arg \max_{W} \frac{|W^T S_B W|}{|W^T S_W W|} \quad (1)$$