Real-Time Gaze Estimator Based on Driver's Head Orientation for Forward Collision Warning System

Article *in* IEEE Transactions on Intelligent Transportation Systems · April 2011

DOI: 10.1109/TITS.2010.2091503 · Source: IEEE Xplore

CITATIONS
54

READS
383

5 authors, including:

Sung Joo Lee
Electronics and Telecommunications Research Institute, South Korea
41 PUBLICATIONS **629** CITATIONS
SEE PROFILE

Jaeik Jo
Yonsei University
10 PUBLICATIONS **192** CITATIONS
SEE PROFILE

Real-Time Gaze Estimator Based on Driver's Head Orientation for Forward Collision Warning System

Sung Joo Lee, Jaeik Jo, Ho Gi Jung, *Member, IEEE*, Kang Ryoung Park, and Jaihie Kim, *Member, IEEE*

***Abstract*—This paper presents a vision-based real-time gaze zone estimator based on a driver's head orientation composed of yaw and pitch. Generally, vision-based methods are vulnerable to the wearing of eyeglasses and image variations between day and night. The proposed method is novel in the following four ways: First, the proposed method can work under both day and night conditions and is robust to facial image variation caused by eyeglasses because it only requires simple facial features and not specific features such as eyes, lip corners, and facial contours. Second, an ellipsoidal face model is proposed instead of a cylindrical face model to exactly determine a driver's yaw. Third, we propose new features—the normalized mean and the standard deviation of the horizontal edge projection histogram—to reliably and rapidly estimate a driver's pitch. Fourth, the proposed method obtains an accurate gaze zone by using a support vector machine. Experimental results from 200 000 images showed that the root mean square errors of the estimated yaw and pitch angles are below 7 under both daylight and nighttime conditions. Equivalent results were obtained for drivers with glasses or sunglasses, and 18 gaze zones were accurately estimated using the proposed gaze estimation method.**

***Index Terms*—Driver monitoring system, forward collision warning (FCW) system, gaze estimation, head orientation estimation, precrash system.**

I. Introduction

ACCORDING to statistics from the United States, driver inattention is one of the major causes of traffic accidents. The U.S. Department of Transportation reported that 9.4% of fatal crashes in 2008 occurred because of driver inattention [1]. To reduce this problem, forward collision warning (FCW) systems have been developed to warn drivers when potential crash hazards exist [2]–[6]. Unfortunately, FCW systems based only on exterior observations (e.g., the distance between cars) distract and bother drivers because these systems generate false warnings, irrespective of the driver's status (e.g., eyes off the road ahead) [7]–[9]. Therefore, to reduce false warnings, FCW systems need not only exterior observations but interior observations of the driver's attention as well. In this paper, we focus on estimating a driver's gaze zone on the basis of the head orientation, which is essential in determining a driver's inattention level.

Many researchers have proposed methods to estimate head orientation and gaze from image streams. Recent surveys on these topics can be found in [10] and [11]. Among various methods, we focused on previous gaze estimation methods developed for vehicular environments. These methods can be categorized into methods considering both eye and head orientation and methods that consider only head orientation.

Methods considering both eye and head orientation can be categorized into hardware- and software-based methods. Hardware methods use two ring-type infrared (IR) light-emitting diodes: one located near the camera's optical axis and the other located far from it [12]–[14]. The light source near the camera's optical axis makes a bright pupil image caused by the red-eye effect, and the other light source makes a normal dark pupil image. The pupil was then easily localized by using the difference between bright and dark pupil images. Ji *et al.* used the size, shape, and intensity of pupils, as well as the distance between the left and right pupil, to estimate a driver's head orientation. Further, they used the pupil-glint displacement to estimate nine discrete gaze zones [12], [14]. Batista used dual Purkinje images to estimate a driver's discrete gaze direction [13]. The major advantage of these methods is the exact and rapid localization of the pupil. However, according to [11], [15], and [16], these methods could not work with drivers wearing glasses because the lenses create large specular reflections and scatter near-IR (NIR) illumination. In addition, during daytime, sunlight is usually far stronger than NIR light sources; thus, the red-eye effect may not occur. Software-based methods used intensity, color, and shape information to detect facial features. Smith *et al.* analyzed color and intensity statistics to find facial features, including both eyes, lip corners, and the bounding box of the face [17]. By using these features, they estimated continuous head orientation and gaze direction. However, this method cannot always find facial features when the driver wears eyeglasses or makes conversation [17]. Kaminski *et al.* analyzed the intensity, shape, and size properties to detect the pupils, nose bottom, and pupil glints. Based on these features and an anthropomorphic model, they estimated continuous head orientation and gaze direction [18].

By using the foregoing methods considering both eye and head orientation, detailed and local gaze direction can be estimated. However, the eye orientation cannot always be measured in vehicular environments because the eye region can be

Manuscript received June 23, 2009; revised January 22, 2010 and September 30, 2010; accepted October 23, 2010. Date of publication January 17, 2011; date of current version March 3, 2011. This work was supported in part by Mando Corporation Ltd. and in part by the National Research Foundation of Korea through the Biometrics Engineering Research Center, Yonsei University, under Grant R112002105070030(2010). The Associate Editor for this paper was A. Amditis.

S. J. Lee, J. Jo, and J. Kim are with the School of Electrical and Electronic Engineering, Biometrics Engineering Research Center, Yonsei University, Seoul 120-749, Korea.

H. G. Jung is with Mando Corporation, Yongin 446-901, Korea.

K. R. Park is with the Division of Electronics and Electrical Engineering and with the Biometrics Engineering Research Center, Dongguk University, Seoul 100-715, Korea.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2010.2091503

Fig. 1. Eye occlusions caused by (a) sunlight reflections on eyeglasses, (b) the eye blink of the driver, and (c) a large head rotation.

occluded by 1) sunlight reflections on eyeglasses; 2) the eye blink of the driver; and 3) a large head rotation, as shown in Fig. 1. In addition, the vehicular environment contains illumination variations between day and night. According to a recent survey on eye detection, all of these factors cause performance degradation of current eye detection algorithms [11]. Finally, FCW systems do not require such detailed gaze direction but need coarse gaze direction to reduce false warnings. For example, Ohue *et al.* determined whether the driver sees in the frontal direction to reduce false warnings; this system was commercialized by a car company [8]. Coarse gaze direction can be obtained by using head orientation since a person's effective visual field is limited, and usually, a person moves the head to a comfortable position before orienting the eye [7], [11]. Because of these reasons, many researchers measured coarse gaze direction by using only head orientation with an assumption that the coarse gaze direction can be approximated by the head orientation.

The methods only considering head orientation can be categorized into methods based on shape features with the eye position, methods based on shape features without the eye position, methods based on texture features, and methods based on hybrid (shape and texture) features.

Methods based on shape features with the eye position analyze the geometric configuration of facial features to estimate the head orientation [13], [16], [19]. A practical problem with these methods is that it is very difficult to reliably find facial features because the driver's facial image varies according to factors such as illumination and eyeglasses [10]. Nuevo *et al.* used an active appearance model (AAM) to find facial features in various situations. However, they used a person-specific AAM model so that the driver was required to cooperate with the system to train the AAM model [16]. Batista and Ji *et al.* [13], [19] found the location of two pupils by using the hardware-based method [12]–[14] and used the pupil location information to constrain the size, location, and orientation of the face when they fitted an ellipse to the face. From the fitted ellipse, continuous pitch and yaw values were estimated. However, ellipse fitting was sensitive to imaging conditions; therefore, these methods can be used in restricted background and illumination conditions [19].

Using a method based on shape features without the eye position, Ohue *et al.* found simple facial features—the left and right borders and the center of the face—instead of detailed facial features such as eyes, nose, and facial contour [8]. Based on these features, the researchers used a cylindrical face model to find the driver's yaw. This method has two advantages over methods based on shape features with the eye position. The first advantage is that the method does not need to train a person-specific model; hence, the driver's cooperation is not required. The second advantage is that the method is relatively robust to environmental variation because it does not require detailed facial features. For example, the method does not find the eye position; therefore, so it is relatively robust to the wearing of glasses. A disadvantage is that the cylindrical face model is inadequate to model the driver's head; thus, the estimated yaw is not exact when the driver's head rotates significantly.

The methods based on texture features find the driver's face in the entire image and analyze the intensity pattern of the driver's facial image to estimate the head orientation. Numerous manifold learning techniques such as PCA, KPCA, LDA, and kernel discriminant analysis have been used for extracting texture features, and these features are classified to obtain the discrete head orientation [20]–[23]. Murphy-Chutorian *et al.* used local gradient orientation (LGO) and support vector regression (SVR) to estimate the driver's continuous yaw and pitch [9]. Ma *et al.* analyzed the asymmetry of the facial image by using a Fourier transform to estimate the driver's continuous yaw [24]. The methods based on texture features are relatively reliable because specific facial features do not need to be localized. However, accuracy can degrade when the face detection module cannot give a consistent result [10].

Finally, methods based on hybrid features combine shape and texture features to estimate the head orientation. Wu *et al.* found the driver's discrete yaw and pitch by using a coarse-to-fine strategy [23]. Texture features—multiresolution Gabor wavelets responses—were used to obtain the coarse head orientation, and shape features obtained from bunch graph matching were used to obtain the fine head orientation. Murphy-Chutorian *et al.* found the initial head orientation by using an LGO-based head orientation method [9], and detailed head orientation was found by using 3-D face model fitting and tracking [25], [26]. This method showed excellent performance but required good initialization. A general drawback of the hybrid methods is the relatively high computational complexity caused by combining two feature extraction methods [23]. The pros and cons of previous methods and the proposed method are summarized in Table I.

In this paper, we propose a real-time continuous head orientation and discrete gaze zone estimation method to reduce the false warning of FCW systems. To develop a fast and robust algorithm, our method is based on shape features without the eye position and texture features. To estimate a driver's head yaw, shape features without the eye position (the left border, the right border, and the center of the driver's face) suggested in [8] were extracted. On the basis of these features, the yaw of the driver's head is found using the proposed ellipsoidal face model instead of the cylindrical face model in [8]. In addition, to estimate the driver's head pitch, we propose new texture features that are the normalized mean and standard deviation obtained from the histogram of the horizontal edge projection. On the basis of these features, the driver's head pitch was found by using SVR. By using the estimated head orientation, an approximate gaze direction was obtained based on the assumption that the approximate gaze direction and the head orientation are identical. From the estimated gaze direction, the face location, and face size, coarse gaze positions called

TABLE I
COMPARISON BETWEEN PREVIOUS METHODS AND THE PROPOSED METHOD

Category	Method	Strength	Weakness
Considering eye and head orientation	Hardware-based eye detector: -Active IR [12], [13], [14]	-Exact and rapid localization of the pupil -High resolution of gaze direction	-Red-eye effect decreases in the case of strong sunlight and glasses -Eye occlusion caused by head rotation, blink, and specular reflection of sunlight
	Software-based eye detector: -Glint-based method [18], -Color and intensity- based method [17]	-High resolution of gaze direction	-Color and intensity variations of eye region caused by illumination, pose, and glasses. -Eye occlusion caused by head rotation, blink, and specular reflection of sunlight
Considering only head orientation	Shape feature (with eye detection): -Person-specific AAM [16], -Ellipse fitting [13], [19]	-High resolution of head orientation	-Intensity variations of eye region caused by illumination, pose, and glasses. -Accuracy is affected by the performance of eye detection -Requires strict constraint for initialization
	Shape feature (without eye detection): -Cylindrical face model [8]	-Reliable performance irrespective of eye detection	-Low accuracy of head orientation estimation caused by simple head model
	Texture feature (without eye detection): -PCA, LDA [20], [21], [22] -KPCA [23], KLDA [23] -LGO [9] -Facial asymmetry [24]	-Reliable performance irrespective of eye detection	-Performance degradation caused by inconsistent results of face detection
	Shape+texture feature for estimating yaw and pitch (with eye detection): -KLDA + EGM [23] -3D model fitting [25],[26]	-High resolution of head orientation	-Relatively high computational cost -Requires the strict constraint for initialization
	Shape feature for estimating yaw +texture feature for estimating pitch (without eye detection) -Proposed method	-Reliable performance irrespective of eye detection -Performance is little affected by the accuracy of face detection -Fast processing speed	-Medium resolution of gaze zone

gaze zones were estimated by using a support vector machine (SVM). The proposed method can operate in both day and night conditions and is robust to facial image variation caused by eyeglasses since it does not need to find specific facial feature points such as eyes, lip corners, and facial contours. In addition, the complexity of the proposed method is very low; thus, it works in real time using a laptop computer.

The rest of this paper is organized as follows: In Section II, we describe the proposed method to estimate the driver's head orientation and the driver's gaze zones on the basis of the estimated head orientation. In Section III, we present our experimental environment and experimental results. Conclusions are given in Section IV.

II. PROPOSED GAZE ZONE ESTIMATION METHOD

A. Overview of Proposed Method

The proposed method consists of face detection, yaw and pitch estimation, and gaze zone estimation steps, as shown in Fig. 2. In the face detection step, the face region is found within the driver's entire facial image to remove unnecessary background and to set the regions of interest (ROIs) used in the yaw and pitch estimation steps. In the yaw estimation step, the left and right borders and the center of the driver's face are extracted; the driver's yaw is determined by the proposed ellipsoidal face model. In the pitch estimation step, the new features that are the normalized mean and standard deviation of the horizontal edge projection histogram are extracted; the driver's pitch is determined by SVR. Finally, in the gaze zone estimation step, the driver's gaze zones are determined on the basis of the driver's head orientation and position information by using SVM.

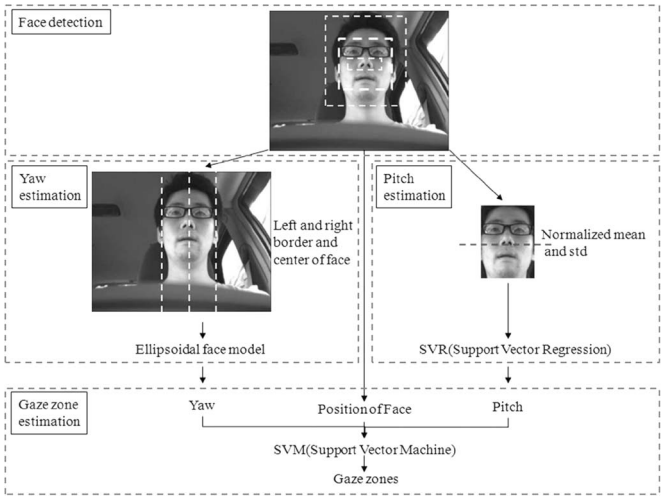


Fig. 2. Flow chart of proposed gaze zone estimation method.

B. Face Detection

In vehicle gaze estimation systems, it is important to rapidly find both frontal faces and rotated faces. To accomplish this, a single Adaboost face detector attuned to only frontal faces and adaptive template matching was combined by using the structure shown in Fig. 3 [27]. With adaptive template matching, rotated faces were found with little processing time.

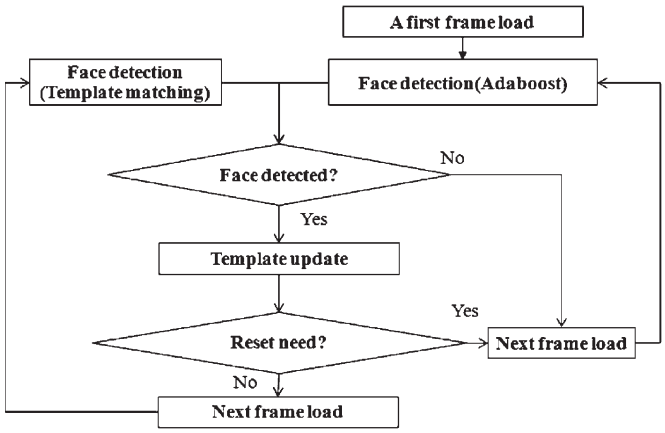


Fig. 3. Flow chart of face detection step.

Fig. 4. Corrected face region using left and right borders of face. (a) Biased face region. (b) Corrected face region.

A drawback of this method is that once the face detection fails, the facial template becomes erroneous. Consequently, a false face could continuously be detected with adaptive template matching. To solve this error propagation problem, a reset function was designed. The reset function is a switch function. The Adaboost face detector is used when the reset function is activated, whereas the adaptive template matching is used when the function is deactivated. The reset function is periodically activated at a certain frame interval (200 frames in our system). It is also activated when the previously estimated head orientation indicates that the driver is looking toward the frontal side after a large head rotation (15° in our system) because the adaptive template matching showed an error when the driver's head rotation was significant. In other cases, the reset function is deactivated. In addition, the detected face region can be biased toward the direction of head rotation, as shown in Fig. 4(a). To solve this problem, the face region is corrected by using the left and right borders of the face obtained in the yaw estimation step, as shown in Fig. 4(b).

In addition, three ROIs were set from the face detection result (ROI 2), as shown in Fig. 5(a). ROI 1, which was used for detecting the left and right borders of the face and for template matching, was selected by enlarging ROI 2. ROI 3, which was used for detecting the center of the face, was selected based on the ratio, as shown in Fig. 5(b). In addition, ROI 3 was biased when there was severe head rotation in the last frame, as shown in Fig. 5(c).

C. Yaw Estimation

1) Feature Extraction for Yaw Estimation: To estimate the driver's yaw, simple shape features proposed by Ohue *et al.* [8]

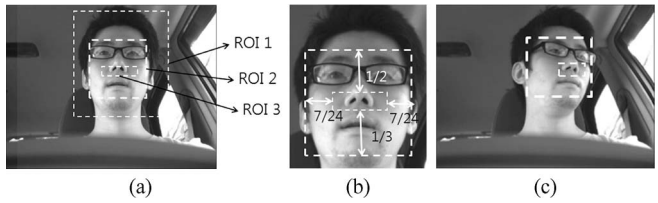


Fig. 5. ROI selection. (a) Three ROIs. (b) Selection of ROI 3. (c) Biased ROI 3.

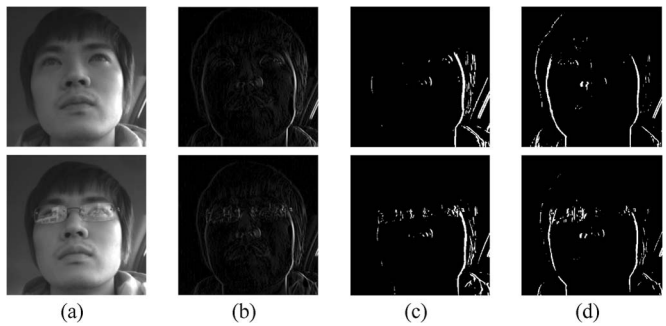


Fig. 6. Adaptive thresholding. (a) Original image. (b) Sobel masked image. (c) Binarized image obtained from an entire facial image. (d) Binarized image obtained from the left and right facial images.

were extracted, and these features include the left and right borders and the center of the driver's face, as shown in Fig. 2.

Let $\mathbf{I}_{\mathrm{ROI1}}$ be the image obtained from ROI 1. The binarized vertical edge image $\mathbf{B}_{\mathrm{ROI1\_}V}$ of $\mathbf{I}_{\mathrm{ROI1}}$ can be obtained from

$$\mathbf{B}_{\mathrm{ROI1\_}V} = \mathrm{T}(\mathbf{I}_{\mathrm{ROI1}} * \mathbf{M}_V) \tag{1}$$

where $*$ and $\mathbf{M}_V$ refer to the convolution operator and the Sobel vertical mask, respectively, and T refers to the p-tile thresholding function [28]. By using the p-tile thresholding function, we can adaptively obtain the threshold value, but when asymmetric illumination incident to the driver's face occurs, the edge in a relatively dark area may disappear, as shown in Fig. 6(c). To solve this problem, $\mathbf{I}_{\mathrm{ROI1}}$ is divided into left and right images, and each vertical edge image is binarized using the p-tile thresholding function. As a result, both sides of the edge are obtained as shown in Fig. 6(d). Then, $\mathbf{B}_{\mathrm{ROI1\_}V}$ is projected in a vertical direction to find the histogram $\mathbf{H}_V$, i.e.,

$$\mathbf{H}_V = \sum_i \mathbf{B}_{\mathrm{ROI1\_}V}[i,j]. \tag{2}$$

The left and right borders of the face can be found by using the left and right peaks of $\mathbf{H}_V$. The center of the face can be found as the center of the two nostrils. The image of ROI 3, i.e., $\mathbf{I}_{\mathrm{ROI3}}$, is binarized using the p-tile thresholding method, and the x positions of the dark points are averaged to find the center of the two nostrils.

2) Yaw Estimation Using Proposed Ellipsoidal Face Model: After the left and right borders and the center of the driver's face were extracted, the driver's yaw was estimated by using the proposed ellipsoidal face model instead of the cylindrical face model by Ohue *et al.* [8]. The cylindrical face model assumes

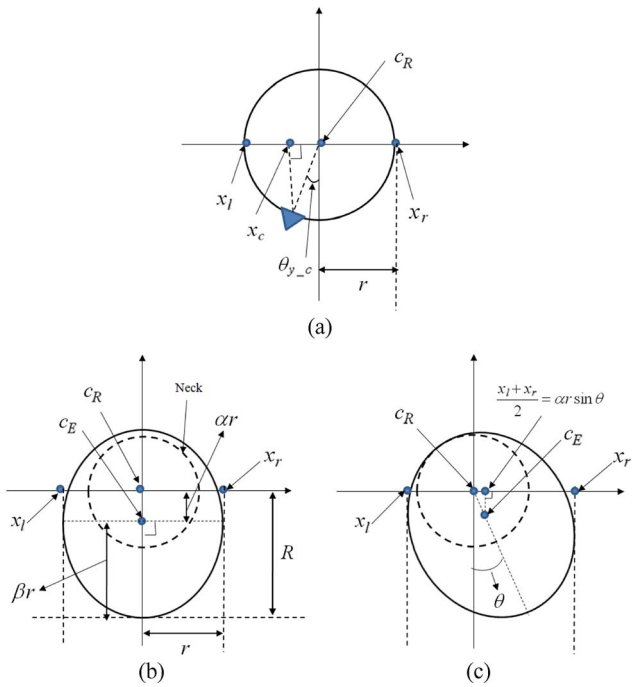


Fig. 7. Face models. (a) Cylindrical face model [8]. (b) Ellipsoidal face model. (c) Ellipse rotated counterclockwise by θ.

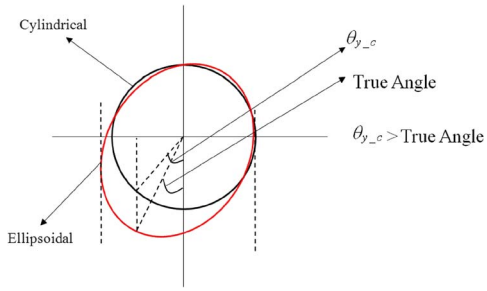


Fig. 8. Difference between cylindrical and ellipsoidal models.

that the driver's head is cylindrical, as shown in Fig. 7(a). The yaw angle can be found by

$$\theta_{y\_c} = \arcsin\left(\frac{c_R - x_c}{r}\right) \quad c_R = \frac{x_r + x_l}{2} \quad r = \frac{x_r - x_l}{2} \tag{3}$$

where θ_{y_c} is the yaw angle obtained from the cylindrical face model; x_r, x_l, and x_c refer to the right and left borders and the center of the face in the image, respectively; and c_R and r refer to the center and radius of rotation, respectively.

However, this assumption is generally not valid because the human head is ellipsoidal, and the center of rotation c_R is not the center of the ellipsoid c_E but rather the center of the neck, as shown in Fig. 7(b) [29]. Because of inaccurate modeling, θ_{y_c} tends to be larger than the true yaw angle. Fig. 8 shows that θ_{y_c} is larger than the true yaw angle, although we have the same x_c and c_R.

To find an accurate yaw angle on the basis of the ellipsoidal model, the ellipse shown in Fig. 7(b) was modeled using

$$\frac{x^2}{r^2} + \frac{(y+\alpha r)^2}{(\beta r)^2} = 1 \tag{4}$$

where α is the ratio between the minor radius of the ellipse r and the distance between c_R and c_E, and β is the ratio between the minor and major radius of the ellipse. The ellipse, rotated counterclockwise by θ, as shown in Fig. 7(c), can be written as

$$\frac{(\cos\theta x + \sin\theta y)^2}{r^2} + \frac{(-\sin\theta x + \cos\theta y + \alpha r)^2}{(\beta r)^2} = 1. \tag{5}$$

x_r and x_l can be found by the two intersection points between the x-axis and the two lines that are parallel to the y-axis and the tangent to the ellipse. The center between x_r and x_l is found as

$$\frac{x_r + x_l}{2} = \alpha r \sin\theta. \tag{6}$$

Therefore, the center of rotation c_R is given by

$$c_R = \frac{x_r + x_l}{2} - \alpha r \sin\theta. \tag{7}$$

The yaw angle based on the ellipsoidal model θ_{y_e} can be expressed as

$$\theta_{y\_e} = \arcsin\left(\frac{c_R - x_c}{R}\right), R = (\alpha + \beta)r \tag{8}$$

where R is the radius of rotation, and x_r, x_l, and x_c are obtained from the input image. Hence, to find θ_{y_e}, we must determine α, β, and θ. On the basis of anthropometry, α and β were set to 0.25 and 1.25, respectively [29]. In addition, θ was estimated by the yaw angle based on the cylindrical model θ_{y_c}.

D. Pitch Estimation

1) Feature Extraction for Pitch Estimation: To estimate the driver's pitch, Murphy-Chutorian *et al.* proposed a texture feature called LGO [9]. To find the LGO, the image of ROI 2 was divided into 16 subblocks, and eight quantized orientations were extracted for each subblock and smoothed [9]. These features can reliably be extracted because we do not need to localize specific facial feature points. If the face region (ROI 2) is detected, then these features could automatically be extracted. However, in general, face detectors cannot always give a consistent result. Consequently, local subblocks of ROI 2 could be changed, and LGO features could become inconsistent.

In this paper, we propose new texture features that are the normalized mean and the standard deviation of the horizontal edge projection histogram. The proposed features are extracted from the entire face image so that they are relatively robust to inconsistent face detection results. Let $\mathbf{B}_{\mathrm{ROI1\_}h}$ be a binarized horizontal edge image of $\mathbf{I}_{\mathrm{ROI1}}$. Then, $\mathbf{B}_{\mathrm{ROI1\_}h}$ can be obtained by

$$\mathbf{B}_{\mathrm{ROI1\_}h} = \mathrm{T}(\mathbf{I}_{\mathrm{zROI1}} * \mathbf{M}_h) \tag{9}$$

where $\mathbf{M}_h$ refers to the Sobel horizontal mask. The horizontal edge projection histogram $\mathbf{H}_h$ can be found from

$$\mathbf{H}_h = \sum_j \mathbf{B}_{\mathrm{ROI1\_}h}[i,j]. \tag{10}$$

(a) (b) (c)

Fig. 9. Binarized horizontal edge image. (a) When orientation is frontal. (b) Upside. (c) Downside.

Let the mean and the standard deviation of $\mathbf{H}_h$ be m_h and s_h, respectively. In our system, the camera is located at the front of a dashboard. As a result, if the driver's head rotates upward, then m_h and s_h decrease because facial features such as the eyes, nose, and lips converge to the upper part of the image, and the distances among facial features decrease, as shown in Fig. 9(b). Likewise, if the driver's head rotates downward, then m_h and s_h increase because the facial features move to the lower part of the image and the distances among facial features increase, as shown in Fig. 9(c).

However, the m_h and s_h values depend on the sitting height, appearance variation caused by glasses, and image scale. Therefore, they were normalized using

$$m_{h\_n} = \frac{(m_h - m_{\text{frontal}})}{f_s} \tag{11}$$

$$s_{h\_n} = \frac{(s_h - s_{\text{frontal}})}{f_s} \tag{12}$$

where m_{h_n} and s_{h_n} refer to the normalized mean and the standard deviation, respectively. m_{frontal} and s_{frontal} refer to the mean and the standard deviation values when the driver's head orientation is frontal, and f_s is the face size. In (11) and (12), the sitting height and the appearance variation caused by glasses can be normalized by subtracting m_{frontal} and s_{frontal}. Image scales can be normalized by dividing f_s. f_s can be obtained from the difference between the left and right borders of the face. m_{frontal} and s_{frontal} can be obtained using the moving average and because, in most cases, the driver's head orientation is frontal. Fig. 10(a) shows the ground truth pitch angles in one of our test sequences. From these data, we can know that the driver intentionally rotated his head frequently, unlike in the case of normal driving conditions. Although there was frequent head rotation, the moving average value of the ground truth pitch angles (red dotted line) converged to the true frontal angle (0°). The final moving average value is −0.1604, which is very close to the true frontal angle (0°). Likewise, the moving average values of m_h and s_h converged to the true m_{frontal} and s_{frontal} values. Fig. 10(b) shows the m_h values extracted in the same test sequence and their moving average value. The true m_{frontal} was 216.18, and the final moving average value was 214.85. Fig. 10(c) shows the s_h values and their moving average. The true s_{frontal} was 60.12, and the final moving average value was 60.70. From these results, we can find that the true m_{frontal} and s_{frontal} values can reliably be found using the moving average method.

2) Pitch Estimation Using Support Vector Regression: To estimate the driver's pitch from the normalized mean and standard deviation, a nonlinear function f_p was found as follows:

$$\theta_p = \mathrm{f}_p(\mathbf{x}_p), \quad \mathbf{x}_p = [m_{h\_n}, s_{h\_n}] \tag{13}$$

where $\mathbf{x}_p$ is a feature vector composed of the normalized mean and the standard deviation, and θ_p refers to the pitch angle. f_p is found using SVR [30]. SVR is a supervised learning method for the nonlinear regression of a scalar function. The basic idea is to project the input data into a high-dimensional space by using a nonlinear kernel function; then, linear regression is used for fitting a hyperplane to the high-dimensional space. The hyperplane f, which is the generalized version of f_p, is described as

$$\mathrm{f}(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) - b. \tag{14}$$

The fitting is accomplished by flattening the hyperplane, i.e., by minimizing $\|\mathbf{w}\|^2$ and simultaneously minimizing the sum of the error, which is larger than a margin ε. To solve the dual problem of (14) and make a nonlinear function f, kernel functions K are defined as

$$\mathrm{K}(\mathbf{a}, \mathbf{b}) = \Phi(\mathbf{a}) \cdot \Phi(\mathbf{b}). \tag{15}$$

In this paper, a radial basis function kernel was used because generally it has less numerical difficulties [31]. The radial basis function kernel is written as

$$\mathrm{K}_{\mathrm{RBF}}(\mathbf{a}, \mathbf{b}) = \exp\left(\frac{-\|\mathbf{a} - \mathbf{b}\|^2}{\sigma}\right). \tag{16}$$

For implementation, an optimized software package for SVR was used [32].

E. Normalization of Estimated Yaw and Pitch

The estimated yaw θ_{y_e} and pitch θ_p were unbiased by subtracting the yaw at the frontal view and the pitch angle at the frontal view, respectively, because θ_{y_e} and pitch θ_p may be not zero, although the driver's orientation is the frontal view. The unbiased yaw $\theta_{y\_e\_u}$ and pitch θ_{p_u} were found as follows:

$$\begin{aligned} \theta_{y\_e\_u} &= \theta_{y\_e} - \theta_{y\_e\_f} \\ \theta_{p\_u} &= \theta_p - \theta_{p\_f} \end{aligned} \tag{17}$$

where $\theta_{y\_e\_f}$ and θ_{p_f} refer to yaw at the frontal view and the pitch angle at the frontal view, respectively. These angles were obtained by using the moving average of θ_{y_e} and θ_p.

F. Gaze Zone Estimation

In this paper, we estimated not only the driver's head orientation but the gaze positions as well. Most previous methods estimate the gaze directions in vehicles [17], [18], but a method related to the estimation of the gaze positions has not been adequately researched. Even with equivalent gaze directions, gaze positions may be different because they vary according to the distance between the camera and the driver. We estimated both the gaze directions and the gaze positions.

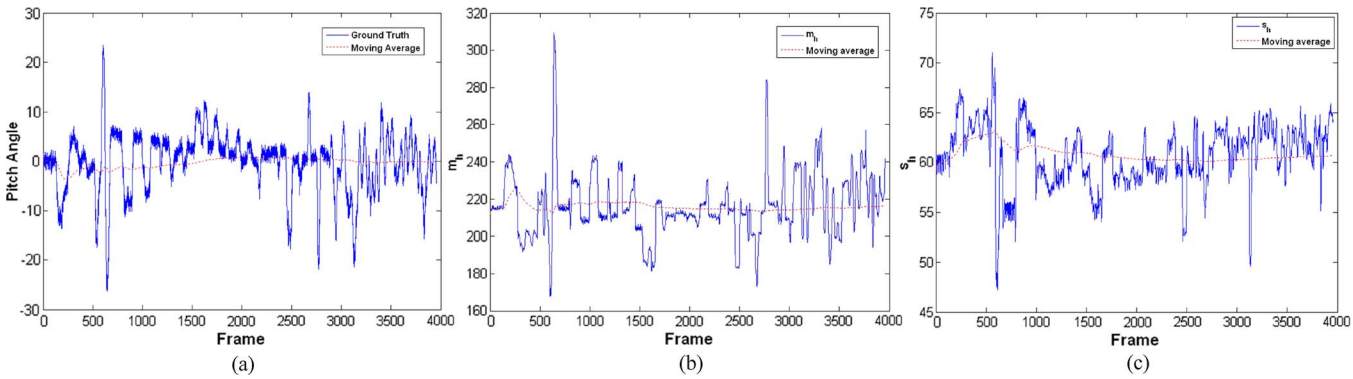


Fig. 10. Robustness of moving average. (a) Ground truth pitch angles and their moving average. (b) m_h values and their moving average. (c) s_h values and their moving average.

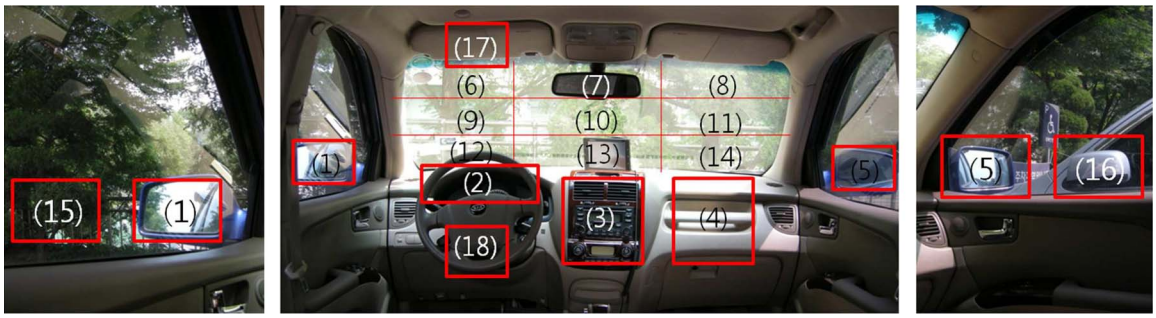


Fig. 11. Eighteen discrete gaze zones in a vehicle.

To find the exact continuous gaze position, we need the driver's 3-D position and gaze direction. Therefore, more than one camera with camera calibration, or a monocular camera with the driver's cooperation, is needed. This either increases the system cost or makes the drivers uncomfortable.

As an alternative, we propose a gaze position estimation method based on a statistical learning method to estimate 18 discrete gaze zones, as shown in Fig. 11. The proposed method does not require the driver's cooperation or stereo cameras. To achieve this goal, it is assumed that the driver's gaze direction and depth information can be estimated using the head orientation information and the face size measured from the image, respectively. On the basis of this assumption, a classifier f_G was found as follows:

$$G_p = \mathrm{f}_G(\mathbf{x}_G), \quad G_p \in \{1, 2, \ldots, 18\} \tag{18}$$

where G_p refers to 18 gaze zones, as shown in Fig. 11, and $\mathbf{x}_G$ refers to the feature vector. $\mathbf{x}_G$ has seven dimensions, as follows:

$$\mathbf{x}_G = [\theta_{y\_e\_u}, \theta_{p\_u}, x_f, y_f, x_d, y_d, f_s] \tag{19}$$

where $\theta_{y\_e\_u}$ and θ_{p_u} refer to the unbiased yaw and pitch angles, respectively; x_f and y_f refer to the x and y pixel positions at the frontal side (these values were obtained using a moving average of the center of ROI 2 in Fig. 5); x_d and y_d refer to the deviation of the current x and y pixel positions with respect to x_f and y_f; and f_s refers to the face size.

To find f_G, an SVM was used. An SVM performs pattern recognition for two-class problems by determining the optimal linear decision hyperplane on the basis of the concept of structural risk minimization with maximum distance to the closest points of the training set that are called support vectors [33]. In general, an SVM can be represented as [33]

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{k} \alpha_i y_i K(x, x_i) + b\right) \tag{20}$$

where k is the number of data points, and y_i -1, 1 is the class label of training point x_i. The coefficients α_i are found by solving a quadratic programming problem with linear constraints, and b is a bias. An SVM can be extended to nonlinear decision surfaces by using a kernel function $K(x, x_i)$. In this paper, a radial basis function kernel was used as shown in (16) because it generally has few numerical difficulties [31]. In addition, a multiclass SVM was used for obtaining 18 gaze zones. The multiclass SVM was obtained using many two-class SVMs because the single multiclass problem can be transformed into multiple binary problems. For the implementation, an optimized software package for the SVM was used [32].

III. Experiment

A. Image Acquisition Device for Estimating Gaze Zones

To build a gaze zone estimation system that can operate in various vehicle environments, the following three conditions were considered. First, the system should operate during daytime and nighttime. For night conditions, additional illumination to capture the driver's facial image was provided. In addition, the effect of a headlight from another car should be

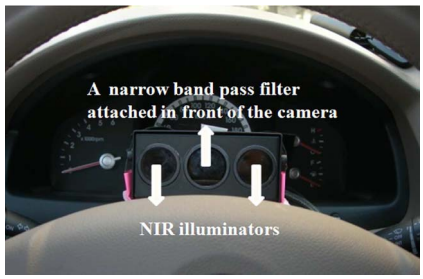


Fig. 12. Developed image acquisition device.

(a) (b)

Fig. 13. Sample images. (a) Image captured during daytime. (b) Image captured at nighttime.

minimized to capture a stable facial image. To accomplish this, a visible light blocking filter was attached in front of the camera lens. The shutter speed of the camera was also automatically adjusted to prevent image saturation. Second, the system should work for wearers of sunglasses. Most sunglasses can transmit NIR illumination. To find the optimal wavelength of illumination, we conducted an experiment with an electronically tunable liquid crystal filter with a bandwidth of 10 nm and a passband range from 650 to 1100 nm [34]. As a result, 850 nm was selected because it is less obtrusive than 750 nm, and the spectral response of the camera is still sufficient. Third, the system should operate irrespective of the driver's movement and sitting height. To satisfy this condition, a sufficient capture volume was set, and the viewing angle of illumination and the focal length of the lens were found based on the calculated capture volume.

The developed image acquisition device is shown in Fig. 12. Our system consists of two 850-nm NIR illuminators, a camera with a 6-mm lens, and a visible light blocking filter in front of the camera [35], [36]. Fig. 13 shows some examples of facial images captured by the proposed system.

B. Database

To evaluate the proposed methods, the developed device was attached in front of a dashboard, as shown in Fig. 14. We collected approximately 300 000 frames from 12 subjects.[1] The collected database contains images at daytime and nighttime and the images of wearers of glasses and sunglasses. In addition, an electromagnetic sensor called Patriot was used to obtain the ground truth data about the head orientation, as shown in Fig. 14 [37].

We built two types of databases. Database 1 contains approximately 85 000 images from 12 subjects. It was used to train the pitch estimation function f_p of (13) and the gaze estimation function f_G of (18) and to test the suggested gaze estimation method. In database 1, the subjects were requested to look frontally for 5 s and to look at one of the 18 gaze zones for another 5 s. This procedure was repeated 18 times for each subject to obtain 18 image sequences containing the driver's ground truth head orientation and ground truth gaze zones. Database 2 contains approximately 200 000 frames from 11 subjects. In database 2, considerably longer image sequences were captured for each subject to test the proposed head orientation method. The subjects were requested to see 1) frontally for 5 s; 2) the 18 gaze zones sequentially without a time interval; 3) one of the 18 gaze zones and frontally in sequence at intervals of 5 s; and 4) in random directions for 1 min. In addition, subjects who did not wear glasses were asked to wear sunglasses to be captured again. In a similar manner, the subjects who wore glasses were asked to take off their glasses or wear sunglasses to be captured again. When databases 1 and 2 were collected, subjects were not constrained to move only their head to orient target gaze zones because people's eyes and head cannot always be aligned in natural driving conditions. Therefore, the subjects were simply instructed to see target gaze zones freely. As a result, every subject rotated his/her head differently, even to gaze at the same gaze zone. Tables II and III show the contents of the two databases in detail.

[1]The database is available on request http://cherup.yonsei.ac.kr/research/research_BERC_Face_Database.html.

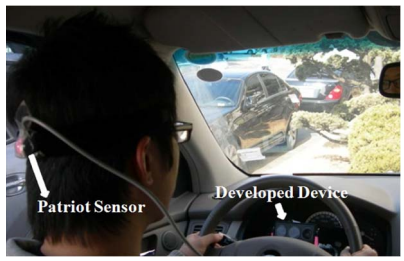


Fig. 14. Patriot sensor attached behind driver's head.

C. Experimental Results

Some qualitative results of the proposed yaw and pitch estimation method are shown in Figs. 15–17. The proposed method was robust to image variations caused by sunlight, glasses, sunglasses, blinking, mouth movement, and specular reflection on glasses.

1) Face Detection: We combined an Adaboost face detector attuned to only frontal faces with adaptive template matching to detect the driver's face. From database 2, the success face detection rate (the number of face-detected frames divided by the total number of frames) was measured. The success face detection rate was higher than 99% for both daytime and nighttime, as shown in Table IV.

2) Yaw Estimation: To estimate the driver's yaw angle, we propose an ellipsoidal face model and compared it with a cylindrical face model. From database 2, the unbiased yaw angle based on the cylindrical face model $\theta_{y\_c\_u}$ and the unbiased yaw angle based on the ellipsoidal face model $\theta_{y\_e\_u}$ were obtained, respectively. In a real application, $\theta_{y\_c\_f}$ and $\theta_{y\_e\_f}$ could be obtained using a moving average of θ_{y_c} and θ_{y_e}, respectively, because drivers see frontally in most cases. However, in database 2, subjects were asked to severely rotate

TABLE II
SPECIFICATION OF DATABASE 1

Daytime	Number of subjects	12 persons (Male: 9, Female: 3, without glasses: 2, with glasses: 10)
	Number of images	44,197 images from 216 sequences
	Number of images used for pitch experiment	35,039 from 197 sequences
	Number of images used for gaze experiment	9,850 images (197 image sequences*50 images)
Nighttime	Number of subjects	11 persons (Male: 8, Female: 3, without glasses: 3, with glasses: 8)
	Number of images	41,155 images from 198 sequences
	Number of images used for pitch experiment	32,416 from 178 image sequences
	Number of images used for gaze experiment	8,900 images (178 image sequences * 50 images)

TABLE III
SPECIFICATION OF DATABASE 2

Daytime	Number of subjects	11 persons (Male: 9, Female: 2, without glasses: 11, with glasses: 10, with sunglasses: 5)
	Number of images	102,022 images from 26 image sequences
	Number of images used for head orientation evaluation	100,478 images from 26 image sequences (1,544 images were excluded because of face detection failure(894) and frontal data gathering (650))
Nighttime	Number of subjects	11 persons (Male: 8, Female: 3, without glasses: 11, with glasses: 11)
	Number of images	97,349 images from 22 image sequences
	Number of images used for head orientation evaluation	96,486 images from 22 image sequences (863 images were excluded because of face detection failure(313) and frontal data gathering (550))

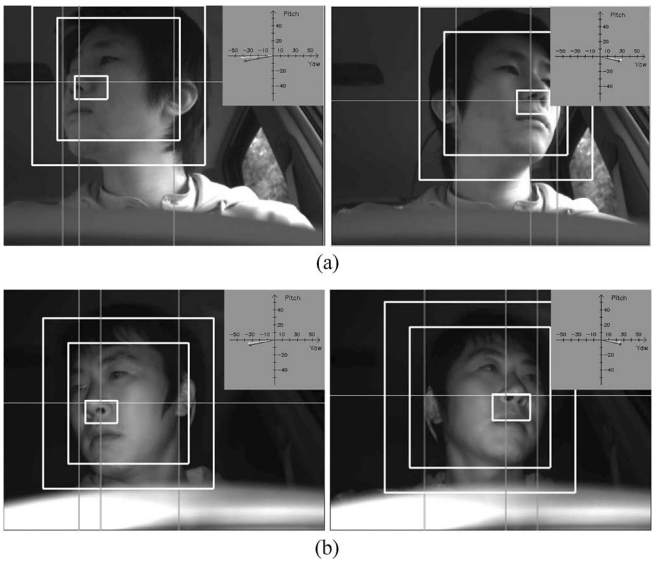


Fig. 15. Results of the proposed yaw and pitch estimation method from images captured during (a) daytime and (b) nighttime. Three vertical lines refer to the left, right, and center of the face. The white arrow in the upper right of image refers to the estimated yaw and pitch angles. Black arrow refers to the ground truth.

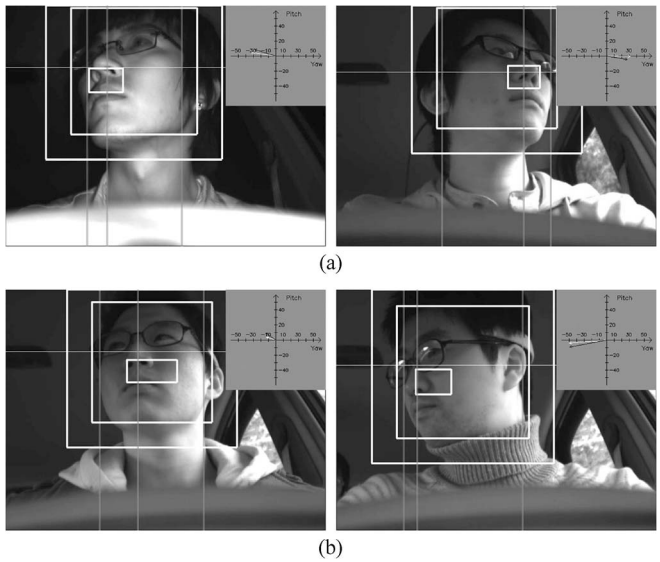


Fig. 16. Results of the proposed yaw and pitch estimation method for subjects who (a) wore glasses and (b) wore sunglasses.

their heads in various directions. Consequently, we cannot use the first several hundred image frames for the moving average. Hence, to test as many image frames as possible, we obtained $\theta_{y\_c\_f}$ and $\theta_{y\_e\_f}$ by averaging the first 25 frames (in these frames, subjects gazed frontally) instead of the moving average. In addition, to make the ground truth yaw angle zero when the driver saw frontally, an average ground truth yaw angle of the first 25 frames was subtracted from the ground truth yaw angles of the remnant frames.

As explained in Section II-C, $\theta_{y\_c\_u}$ was often larger than the ground true yaw angle. This problem was mitigated by the proposed ellipsoidal face model, as shown in Fig. 18.

To quantitatively evaluate the accuracy of the yaw estimation methods, the average RMS error between the ground truth and the estimated yaw angles was measured. If the RMS error was larger than 20°, then this case was regarded as a completely failed case and was removed when we measured the average RMS error. The completely failed case was caused by errors in the feature extraction step, and Fig. 19 shows some examples of errors in this step. Most errors in the feature extraction step were caused by inconsistent face detection results. The Adaboost algorithm did not always find the optimal location of the face; rather, it sometimes found the face region biased to the upper part of the face, as shown in Fig. 19(a). Consequently, ROI 3 was also biased to the upper part of the face, and the optimal face center was not found in such an erroneous ROI 3. In addition, when a driver's hair intruded ROI 3, the face center could not be found exactly, as shown in Fig. 19(b). However, these errors occurred in very special cases, and the errors

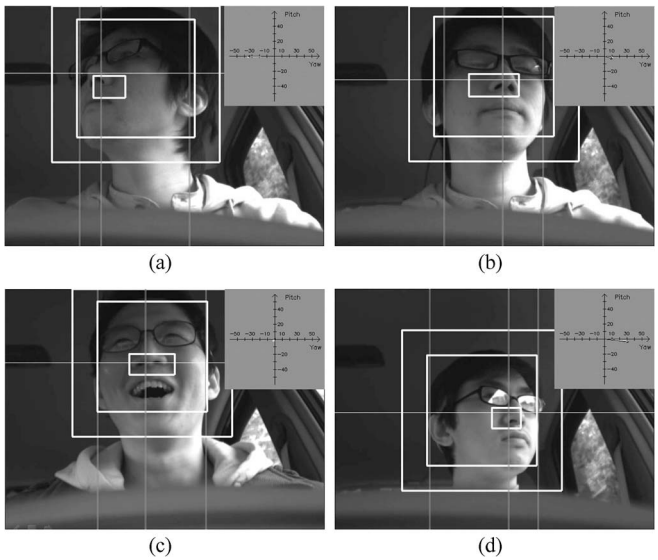


Fig. 17. Results of the proposed yaw and pitch estimation method from images with (a) eye occlusion by head rotation, (b) eye occlusion by blinking, (c) mouth movement, and (d) sunlight specular reflection on glasses.

TABLE IV
FACE DETECTION RESULTS

	Total frames	Face-detected frames	Success rate
Daytime	102,022	101,128	99.12%
Nighttime	97,349	97,036	99.68%

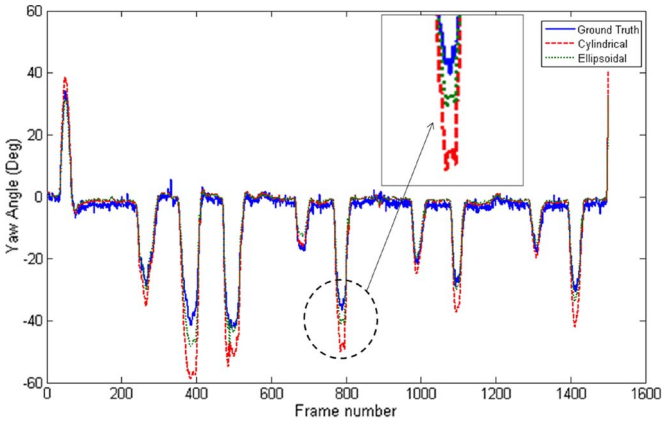


Fig. 18. Yaw estimation results with cylindrical and ellipsoidal face models.

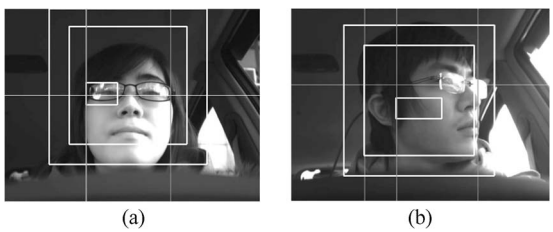


Fig. 19. Example of error cases. (a) Error caused by inconsistent face detection result. (b) Error caused by intrusion of hair.

were not propagated because of the reset function in the face detection step.

To measure the reliability of the yaw estimation methods, a success detection rate (SDR), which refers to the ratio of the number of frames whose RMS error is less than 20° divided by the number of total frames, was measured. Table V shows the average RMS error and SDR of all the images in database 2. From Table V, it is clear that the yaw estimation method based on the ellipsoidal face model was more accurate and reliable than that based on the cylindrical face model for all the conditions, such as daytime, nighttime, wearing glasses, and wearing sunglasses.

TABLE V
YAW ESTIMATION RESULTS ON IMAGES OF DATABASE 2 TAKEN UNDER EACH CONDITION

		Cylindrical[8]		Ellipsoidal	
Conditions	Frame#	RMS (°)	SDR (%)	RMS (°)	SDR (%)
Daytime	100,464	6.07	87.84	5.91	90.08
Nighttime	96,486	6.67	87.44	6.25	93.07
No glasses	89,331	6.27	90.15	5.97	93.56
Glasses	85,283	6.66	84.45	6.37	89.48
Sunglasses	22,336	5.67	89.78	5.34	91.35
Total	196,950	6.37	87.64	6.08	91.54

TABLE VI
PITCH ESTIMATION RESULTS ON THE TRAINING AND TESTING DATABASES

		LGO + SVR[9]		Proposed	
Database	Frame#	RMS (°)	SDR (%)	RMS (°)	SDR (%)
Training	35,039	4.32	99.75	4.68	98.8
Testing	196,950	6.54	97.72	5.57	97.77

3) Pitch Estimation: To estimate the driver's pitch angle, we propose a new feature vector $\mathbf{x}_p$ that consists of the normalized mean m_{h_n} and standard deviation s_{h_n}. $\mathbf{x}_p$ was compared with the previous LGO, which consisted of 128-D features [9]. After LGO and $\mathbf{x}_p$ were extracted, pitch angles were estimated from nonlinear functions f_{LGO} and f_p, which were found using SVR. To train f_{LGO} and f_p, the ground truth pitch angle, i.e., LGO, and $\mathbf{x}_p$ were extracted from the training data composed of approximately 70 000 images from database 1. To remove errors in the training data, face detection was manually performed, and some image sequences that contained feature extraction errors were manually excluded, as shown in Table II.

The RMS error and SDR were measured on both the testing database (database 2) and the training database (database 1) to determine which features were robust to face image variations caused by inconsistent face detection, sunlight, glasses, and sunglasses. Table VI shows the average RMS and SDR values of the training database and testing database, respectively. The SDRs of the two methods were similar and above 97% for both training and testing databases. However, the RMS errors of the two methods were different. The RMS error of the LGO-based method was smaller than that of the proposed features-based method in the case of the training database. However, the RMS error of the LGO-based method was higher than that of the proposed features-based method in the case of the testing database. The major difference between the tests on the training database and the testing database is the following: In the training database, the face region was detected manually, but in the testing database, the face region was detected automatically using the Adaboost face detector combined with adaptive template matching. In other words, in the testing database,

TABLE VII
PITCH ESTIMATION RESULTS ON IMAGES OF DATABASE 2 TAKEN UNDER EACH CONDITION

		LGO + SVR[9]		Proposed	
Conditions	Frame#	RMS (°)	SDR (%)	RMS (°)	SDR (%)
Daytime	100,464	7.27	96.61	5.65	97.53
Nighttime	96,486	5.69	98.87	5.49	98.03
No glasses	89,331	6.35	97.69	5	98.33
Glasses	85,283	6.62	97.56	5.91	97.15
Sunglasses	22,336	6.95	98.46	6.39	97.91
Total	196,950	6.54	97.72	5.57	97.77

TABLE VIII
PROCESS TIME OF THE PROPOSED AND LGO-BASED METHOD

	LGO+SVR[9]	Proposed
Feature extraction	1.4ms	1ms
Estimate pitch	3.3ms	0.7ms
Total	4.7ms	1.7ms

relatively inconsistent face regions were detected. Therefore, the proposed pitch estimation method is more robust to inconsistent face detection results than the LGO-based method. In addition, the proposed method showed better RMS results than the LGO-based method on daytime and nighttime images, as shown in Table VII, and showed better RMS results than the LGO-based method for drivers who wore glasses and sunglasses and who did not wear glasses. Therefore, the proposed method is more robust to facial image variations caused by sunlight, glasses, and sunglasses than the LGO-based method. In addition, the proposed pitch estimation method requires just 2-D features, whereas the LGO-based method requires 128-D features. Hence, the computational cost of the proposed method was far lower than that of the LGO-based method. Table VIII shows the processing time of the proposed pitch estimation method and the LGO-based method measured on Intel Pentium M processor at 1.60 GHz with 500-MB RAM.

4) Gaze Estimation: The driver's gaze was divided into 18 gaze zones, which are more specific than the seven gaze zones found in [38]. To estimate the 18 gaze zones, as shown in Fig. 11, three kinds of gaze classifiers were implemented and compared. The first is an SVM classifier, the second is a minimum-distance classifier, and the third is the nearest-neighbor classifier [39]. Frontal information was collected from the first 25 frames of each image sequence in database 1, and the gaze feature vector $\mathbf{x}_G$ and the ground truth gaze zone G_p were extracted from the last 50 frames of each image sequence. Hence, approximately 18 000 gaze feature vectors and ground truth gaze zones were obtained, as shown in Table II. We divided this total database into 12 subdatabases for each subject and performed leave-one-out cross validation. The performance of the gaze estimation method was measured by using the strictly correct estimation rate (SCER) and the loosely correct estimation rate (LCER). The SCER refers to the ratio of the number of strictly correct frames divided by the number of total frames. A strictly correct frame indicates a frame where the estimated gaze zone is equivalent to the ground truth gaze zone of a single target zone. The LCER refers to the ratio of the number of loosely correct frames divided by the number of total frames. A loosely correct frame indicates a frame where the estimated gaze zone is placed within multiple ground truth gaze zones obtained from a target zone and its neighbors. In this paper, at most eight neighbors were set for a target zone, as shown in Table X.

TABLE IX
SCER AND LCER OF THREE DIFFERENT CLASSIFIERS

Classifiers	SCER (%)	LCER (%)
k-NN (Nearest Neighbor), k=1	38.47	80.08
k-NN (Nearest Neighbor), k=3	38.92	80.07
k-NN (Nearest Neighbor), k=5	39.22	80.29
k-NN (Nearest Neighbor), k=10	39.05	80.44
Minimum distance	42.53	86.87
SVM	47.44	87.29

Table IX shows the SCER and LCER results for each classifier in database 1. From Table IX, we can find the following: 1) The SVM classifier showed the best performance among the three classifiers, and 2) the LCER was far higher than the SCER. In other words, it is difficult for the gaze estimation method based on head orientation information to estimate specific gaze zones; however, the method can estimate coarse gaze zones reliably. These coarse estimate results should be very useful in reducing the false alarms in FCW systems [7], [8].

To analyze the error of the gaze estimation method, the SCER, the LCER, and the mean and standard deviations of ground truth yaw angles and pitch angles for each gaze zone were measured, as shown in Table X. Here, the SCER and the LCER were measured using the SVM classifier. From Table X, we can find that the standard deviations of ground truth yaw angles and pitch angles increased as their mean values increased. This result shows that yaw and pitch angles overlapped more at gaze zones far from the frontal view than gaze zones close to the frontal view. Consequently, the SCER and the LCER generally decreased at these gaze zones (e.g., 5, 8, 11, 14, and 16). Fundamentally, these results occurred because every person rotated his/her head differently, even to see the same gaze zone. For example, one person slightly rotated his head and greatly rotated his eyes, but another person greatly rotated his head and slightly rotated his eyes, as shown in Fig. 20.

Gaze zones 5 and 14 showed a very low SCER for two reasons. The first reason is that the standard deviation values of the ground truth yaw at these gaze zones are relatively high. Consequently, most samples of these gaze zones were classified as their neighbor gaze zones. We can know this fact by observing that the LCER was far higher than the SCER at these gaze zones. The second reason is that the mean values of the ground truth yaw at these gaze zones are similar to those at other gaze zones that are not their neighbors. For example, the difference between the mean values of the ground truth yaw at gaze zone 14 and that at gaze zone 7 was only 5.5, and the difference between gaze zones 5 and 8 was only 4.16. As a result, many samples of gaze zones 14 and 5 were misclassified as 7 and 8, respectively. These results occurred because in these zones subjects rotated their heads less than in the other gaze zones. For instance, the mean value of the ground truth yaw at gaze zone 14 was the smallest among those of the gaze zones in the same column (gaze zones 4, 8, and 11).

TABLE X
SCER, LCER, MEAN, AND STANDARD DEVIATION OF GROUND TRUTH YAW AND PITCH ANGLES FOR EACH GAZE ZONE

Target zone	Neighbors	SCER(%)	LCER(%)	Ground truth yaws		Ground truth pitches	
				Mean (°)	STD (°)	Mean (°)	STD (°)
1	15, 2, 12, 18	85	100	22.05	10.18	-7.51	4.83
2	1, 3, 12, 13, 18	46	98	-1.42	1.93	-8.1	4.58
3	12, 13, 14, 2, 4, 18	41	95	-21.27	10.13	-6.16	4.3
4	13, 14, 3, 5	57	83	-30.52	14.33	-6.28	4.74
5	4, 14, 16	4	54	-38.1	17.77	-4.58	4.74
6	17, 7, 9, 10	64	92	1.52	2.11	6.78	4.58
7	6, 17, 8, 9, 10, 11	33	86	-22.85	11.01	5.35	3.71
8	7, 10, 11	35	76	-33.94	15.87	2.11	3.14
9	6, 7, 10, 12, 13	64	100	0.45	1.66	2.31	2.85
10	6, 7, 8, 9, 11, 12, 13, 14	48	93	-17.09	10.1	1.56	2.36
11	7, 8, 10, 13, 14	29	81	-31.22	15.32	0.05	2.91
12	9, 10, 13, 1, 2, 3	52	98	-0.34	1.98	-2.9	2.71
13	9, 10, 11, 12, 14, 2, 3, 4	21	93	-17.97	9.07	-1.36	2.81
14	13, 10, 11, 3, 4, 5	3	72	-28.35	14.12	-2.68	3.53
15	1	88	100	40.49	16.9	-13.6	7.56
16	5	46	61	-48.99	21.1	-5.98	5.97
17	6, 7	46	75	5.62	5.24	21.77	10.06
18	1, 2, 3	87	100	-3.26	3.92	-21.77	10.23

Fig. 20. Two different subjects gazed at the same gaze zone (zone 15).

TABLE XI
PROCESSING TIME OF PROPOSED GAZE ESTIMATION METHOD

Process	Process time
Face detection (Adaboost)	25.4 ms
Face detection (adaptive template matching)	5.8 ms
Feature extraction for yaw+ estimate yaw	4.4 ms
Feature extraction for pitch	1 ms
Estimate pitch by using SVR	0.7 ms
Estimate gaze by using SVM	2.2 ms
Total using Adaboost	33.7 ms
Total using adaptive template matching	14.1 ms

Both gaze zones showed a low SCER, but gaze zone 14 showed a higher LCER than gaze zone 5 because 1) the standard deviation values of the ground truth yaw at gaze zone 14 were smaller than that at gaze zone 5, and 2) gaze zone 14 had more neighbors than gaze zone 5.

5) Processing Time: The proposed gaze estimation method was very efficient and worked in real time. It took just 33.7 ms on an Intel Pentium M processor with 1.60 GHz and 500-MB RAM when the Adaboost face detector was used and 14.1 ms when adaptive template matching was used, as shown in Table XI.

IV. CONCLUSION

In this paper, we have proposed an efficient gaze area estimator based on a driver's head orientation composed of yaw and pitch. Compared with previous works, our method offers the following contributions.

1) We analyzed the histogram of horizontal and vertical edge projections and extracted simple shape features to estimate the driver's yaw and pitch. The proposed method can operate at both day and night and is robust to facial image variation caused by eyeglasses since it does not need to find specific facial feature points such as eyes, lip corners, and facial contours.
2) To determine the driver's exact yaw, we proposed an elliptical face model instead of a cylindrical face model.
3) To find the driver's pitch reliably and efficiently, we proposed the use of the normalized mean and the standard deviation obtained from the histogram of horizontal edge projections.
4) We proposed an efficient gaze estimation method based on the driver's head orientation using SVM. We estimated both the gaze directions and the gaze zones.

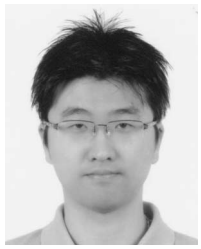
Experimental results showed that the RMS errors of the estimated yaw and pitch angles were below 7, irrespective of whether it was daytime or nighttime or whether the subject was wearing glasses or sunglasses. In addition, the LCER of the proposed gaze estimation method was approximately 90%, and the method worked in real time on a laptop computer with an Intel Pentium M processor at 1.60 GHz. In a future work, we will detect blinks when the driver gazes frontally to estimate the fatigue level of drivers. The fatigue information will be combined with inattention information to reduce traffic accidents caused by inattention and fatigue.

REFERENCES

[1] U.S. Dept. Transp., *2008 Traffic Safety Facts Annual FARS/GES Report (early ed.)*. [Online]. Available: http://www-nrd.nhtsa.dot.gov/Pubs/811170.PDF

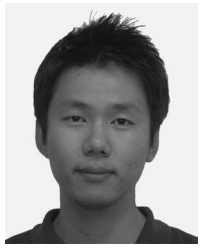
[2] J. Huang and H.-S. Tan, "Error analysis and performance evaluation of a future-trajectory-based cooperative collision warning system," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 175–180, Mar. 2009.

[3] T. Kim and H.-Y. Jeong, "Crash probability and error rates for head-on collisions based on stochastic analyses," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 896–904, Dec. 2010.
[4] C.-Y. Chang and Y.-R. Chou, "Development of fuzzy-based bus rear-end collision warning thresholds using a driving simulator," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 360–365, Jun. 2009.
[5] M. Brännström, E. Coelingh, and J. Sjöberg, "Model-based threat assessment for avoiding arbitrary vehicle collisions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 658–669, Sep. 2010.
[6] G. K. Mitropoulos, I. S. Karanasiou, A. Hinsberger, F. Aguado-Agelet, H. Wieker, H.-J. Hilt, S. Mammar, and G. Noecker, "Wireless local danger warning: Cooperative foresighted driving using intervehicle communication," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 539–553, Sep. 2010.
[7] A. Hattori, S. Tokoro, M. Miyashita, I. Tanaka, K. Ohue, and S. Uozumi, "Development of forward collision warning system using the driver behavioral information," presented at the Soc. Automotive Eng. World Congr., Detroit, MI, 2006, SAE Technical Paper Series, Paper 2006-01-1462.
[8] K. Ohue, Y. Yamada, S. Uozumi, S. Tokoro, A. Hattori, and T. Hayashi, "Development of a new pre-crash safety system," presented at the Soc. Automotive Eng. World Congr., Detroit, MI, 2006, SAE Technical Paper Series, Paper 2006-01-1461.
[9] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi, "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2007, pp. 709–714.
[10] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 4, pp. 607–626, Apr. 2009.
[11] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.
[12] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imaging*, vol. 8, no. 5, pp. 357–377, Oct. 2002.
[13] J. P. Batista, "A real-time driver visual attention monitoring system," in *Proc. Iberian Conf. Pattern Recog. Image Anal.*, vol. 3522, *Lecture Notes in Computer Science*, 2005, pp. 200–208.
[14] Q. Ji and X. Yang, "Real time visual cues extraction for monitoring driver vigilance," in *Proc. 2nd Int. Workshop Comput. Vis. Syst.*, vol. 2095, *Lecture Notes in Computer Science*, 2001, pp. 107–124.
[15] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," *Pattern Recognit.*, vol. 40, no. 8, pp. 2341–2355, Aug. 2007.
[16] J. Nuevo, L. M. Bergasa, M. A. Sotelo, and M. Ocana, "Real-time robust face tracking for driver monitoring," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2006, pp. 1346–1351.
[17] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 4, pp. 205–218, Dec. 2003.
[18] J. Y. Kaminski, D. Knaan, and A. Shavit, "Single image face orientation and gaze detection," *Mach. Vis. Appl.*, vol. 21, no. 1, pp. 85–98, Nov. 2009.
[19] Q. Ji and R. Hu, "3D face pose estimation and tracking from a monocular camera," *Image Vis. Comput.*, vol. 20, no. 7, pp. 499–511, May 2002.
[20] P. Watta, N. Gandhi, and S. Lakshmanan, "An eigenface approach for estimating driver pose," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2000, pp. 376–381.
[21] S. Lakshmanan, P. Watta, Y. L. Hou, and N. Gandhi, "Comparison between eigenfaces and fisherfaces for estimating driver pose," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2001, pp. 889–894.
[22] P. Watta, S. Lakshmanan, and Y. Hou, "Nonparametric approaches for estimating driver pose," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2028–2041, Jul. 2007.
[23] J. Wu and M. M. Trivedi, "A two-stage head pose estimation framework and evaluation," *Pattern Recognit.*, vol. 41, no. 3, pp. 1138–1158, Mar. 2008.
[24] B. Ma, S. Shan, X. Chen, and W. Gao, "Head yaw estimation from asymmetry of facial appearance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 6, pp. 1501–1512, Dec. 2008.
[25] E. Murphy-Chutorian and M. M. Trivedi, "HyHOPE: Hybrid head orientation and position estimation for vision-based driver head tracking," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 512–517.
[26] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 300–311, Jun. 2010.
[27] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 1, pp. I-511–I-518.
[28] R. Jain, R. Kasturi, and B. G. Schunck, "Automatic thresholding," in *Machine Vision*. New York: McGraw-Hill, 1995, pp. 76–86.
[29] L. G. Farkas, *Anthropometry of the Head and Face*. New York: Raven, 1994, pp. 244–270.
[30] N. Cristianini and J. Shawe-Taylor, "Support vector regression," in *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000, pp. 93–124.
[31] C.-W. Hsu, C.-C. Chang, C.-J. Lin. A practical guide to support vector classification. Dept. Comput. Sci. Inf. Eng., Nat. Taiwan Univ., Taipei, Taiwan, Tech. Rep. [Online]. Available: http://w.csie.org/~cjlin/papers/guide/guide.pdf
[32] C.-C. Chang and C.-J. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[33] V. Vapnik, "Support vector estimation of functions," in *Statistical Learning Theory*. New York: Wiley, 1998, pp. 375–570.
[34] *Specification of the Electronically Tunable Liquid Crystal Filter*. [Online]. Available: http://www.cri-inc.com/support/components.asp
[35] *Illumination Specification*. [Online]. Available: http://www.roithner-laser.com/datasheets/led_multi/to/LED850-66-60.pdf
[36] *Camera Specification*. [Online]. Available: http://www.graftek.com/pages/EC650.htm
[37] *Specification of Motion Tracking Device*. [Online]. Available: http://www.polhemus.com/polhemus_editor/assets/Tech%20compare%20with%20ResolutionvsRange%20graphs.pdf
[38] S. Y. Kim, H. C. Choi, W. J. Won, and S. Y. Oh, "Driving environment assessment using fusion of in- and out-of-vehicle vision systems," *Int. J. Automot. Technol.*, vol. 10, no. 1, pp. 103–113, Feb. 2009.
[39] R. O. Duda, P. E. Hart, and D. G. Stork, "Bayesian decision theory," in *Pattern Classification*. New York: Wiley, 2001, pp. 20–83.



Sung Joo Lee received the B.S. degree in electrical and electronic engineering and the M.S. degree in biometric engineering in 2004 and 2006, respectively, from Yonsei University, Seoul, Korea, where he is currently working toward the Ph.D. degree in electrical and electronic engineering.

His current research interests include driver monitoring system, 3-D face reconstruction, biometrics, pattern recognition, and computer vision.



Jaeik Jo received the B.S. degree in electrical and electronic engineering in 2008 from Yonsei University, Seoul, Korea, where he is currently working toward the M.S.-Ph.D. joint degree.

His current research interests include driver monitoring systems, 3-D face reconstruction, biometrics, pattern recognition, and computer vision.

Ho Gi Jung (M'05) received the B.E., M.E., and Ph.D. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1995, 1997, and 2008, respectively.

He was with Mando Corporation Global R&D H.Q., from 1997 to April 2009. He developed environment recognition algorithms for intelligent parking-assist systems, collision warning and avoidance, and the active pedestrian protection system. Since May 2009, he has been with Yonsei University. His interests are automotive vision, driver assistant systems, active safety vehicles, biometrics, and intelligent surveillance.

Dr. Jung is a member of the Society of Automotive Engineers, Society of Photo-Optical Instrumentation Engineers, Institute of Electronics Engineers of Korea, and Korean Society of Automotive Engineers.

Kang Ryoung Park received the B.S. and M.S. degrees in electronic engineering and the Ph.D. degree in computer vision from Yonsei University, Seoul, Korea, in 1994, 1996, and 2000, respectively.

He was an Assistant Professor with the Division of Digital Media Technology, Sangmyung University, Seoul, from March 2003 to February 2008. He has been an Associate Professor with the Division of Electronics and Electrical Engineering, Dongguk University, Seoul, since March 2008. He has also been a Research Member of the Biometrics Engineering Research Center. His research interests include computer vision, image processing, and biometrics.

Jaihie Kim (M'84) received the B.S. degree in electronic engineering from Yonsei University, Seoul, Korea, in 1979 and the M.S. degree in data structures and the Ph.D. degree in artificial intelligence from Case Western Reserve University, Cleveland, OH, in 1982 and 1984, respectively.

Since 1984, he has been a Professor with the School of Electrical and Electronic Engineering, Yonsei University. He is currently the Director of the Biometric Engineering Research Center, Korea. His research areas include biometrics, computer vision, and pattern recognition.

Prof. Kim is currently the Chairman of Korean Biometric Association and a member of the National Academy of Engineering of Korea.