Robotica (2012) volume 30, pp. 731–741. © Cambridge University Press 2011
doi:10.1017/S0263574711000749

# Face salient points and eyes tracking for robust drowsiness detection

J. Jimenez-Pinto and M. Torres-Torriti*

*Dept. of Electrical Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Casilla 306-22, Santiago, Chile. E-mail: jejimenp@puc.cl*

(Received in Final Form: June 21, 2011; accepted June 15, 2011. First published online: September 7, 2011)

**SUMMARY**

Measuring a driver's level of attention and drowsiness is fundamental to reducing the number of traffic accidents that often involve bus and truck drivers, who must work for long periods of time under monotonous road conditions. Determining a driver's state of alert in a noninvasive way can be achieved using computer vision techniques. However, two main difficulties must be solved in order to measure drowsiness in a robust way: first, detecting the driver's face location despite variations in pose or illumination; secondly, recognizing the driver's facial cues, such as blinks, yawns, and eyebrow rising. To overcome these challenges, our approach combines the well-known Viola–Jones face detector with the motion analysis of Shi–Tomasi salient features within the face. The location of the eyes and blinking is important to refine the tracking of the driver's head and compute the so-called PERCLOS, which is the percentage of time the eyes are closed over a given time interval. The latter cue is essential for noninvasive driver's alert state estimation as it has a high correlation with drowsiness. To further improve the location of the eyes under different conditions of illumination, the proposed method takes advantage of the high reflectivity of the retina to near infrared illumination employing a camera with an 850 nm wavelength filter. The paper shows that motion analysis of the salient points, in particular cluster mass centers and spatial distributions, yields better head tracking results compared to the state-of-the-art and provides measures of the driver's alert state.

KEYWORDS: Alert state assessment; Fatigue detection; Drowsiness detection; Driver assistance; IR eye tracking; Yawn detection; Eyebrow rising detection.

## 1. Introduction

### *1.1. Motivation*

The financial and social cost of traffic accidents to national economies has been estimated to be 1% of the gross national product in low-income countries, 1.5% in middle-income countries, and 2% in high-income countries, resulting in an aggregated global cost of US$518 billion per year.[24] This amount is expected to increase 67% by year 2020 if no actions are taken to improve education, law enforcement, infrastructure, and technology. Road accidents statistics point to inattention, fatigue, drowsiness, and drinking as the main causes of 10% to 30% of all traffic accidents (see ref. [10] and references therein). In particular for the trucking industry, drowsiness appears as the cause of about 60% of the fatal truck accidents.[2,17] In the United States, the National Highway Traffic Safety Association (NHTSA) reported as much as 56,000 accidents back in 1996,[21] a number that is now estimated to be around 70,000. Even if the latter figure is about 0.7% of the reported accidents, the number of persons that yearly get killed because of drivers falling asleep behind the wheel is about 1500. These figures are even larger if other accidents related to the driver's state-of-alert, such as distracted driving (3.5%) and cell-phone use accidents (0.1%), are included. Some other alarming accident statistics of accidents due to fatigue, stress, or distraction can also be found in refs. [10, 25]. The situation in Europe is similar, for example, statistics of the German insurance organization (GDV) reveal that one of every four fatal autobahn accidents are caused by tired or drowsy drivers. In this context, it is essential to develop effective systems for monitoring a driver's state of awareness, as well as other forms advanced driver assistance systems (e.g., pedestrian detection, lane departure warning, or headway keeping[1,4,22]), to ensure that the senses of a human driver are preserved and hopefully broaden with information that may fail to perceive.

### *1.2. State of the art*

During the last two decades, significant understanding of the causes and psychophysiological signals of driver fatigue, drowsiness, and distraction has been gained.[18,20,36] This has lead to the development of different monitoring and prevention countermeasures, which can be divided into three main classes: (i) methods based on bioelectric and nervous signals; (ii) methods based on driver steering and driving behavior monitoring; (iii) methods based on face and head monitoring. The first of these approaches relies on electrocardiograms, electroencephalograms, or electro-oculograms,[18] and its main drawback is that the measurement techniques involved are invasive. On the other hand, despite bioelectric signal patterns have high intraindividual correlations with drowsiness, they are difficult to employ as a general reference because they have significant interindividual variabilities that depend on multiple factors, such as age, gender, spatial ability, and intro-extroversion.[18] Methods based on driver steering monitoring measure

* Corresponding author. E-mail: mtorrest@ing.puc.cl

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

reaction rates and unexpected lane departures to warn the driver. However, despite claims that these approaches have low false alarms rates, there are no public research results to support these claims.[20] Some recent efforts to understand better changes in driver–vehicle interaction due to sleepiness are presented in ref. [36]; however, the interwoven complexity of driving tasks and cognitive functions makes it difficult to reliably isolate symptoms of fatigue. In the third category, it is possible to find head-nodding technology, which despite being cheap and simple to implement is unreliable due to its high false alarm rates and the fact that warns drivers too late as to their sleepiness.[20] Another alternative for face and head monitoring is to employ computer vision-based approaches to measure PERCLOS (the proportion of time during 1 min blocks that the eyelid covers 80% or more of the pupil) and other visual cues, such as yawns and blinks, head movements, and eye-gaze variations, which are the main changes occurring due to increased levels of fatigue, drowsiness, or distraction.

Video monitoring of the eyes, and in particular the measurement of PERCLOS, has been found to be the most reliable indicator of fatigue, and notably superior to eye-blink detection software and head-nodding technology.[20] PERCLOS values exhibit high correlations with lane departures, subjective sleepiness, and decrements in driver performance measured in terms of reaction times through psychomotor vigilance task experiments.[20] Thus, in recent years, a variety of noninvasive monitoring systems based on computer vision techniques have been proposed (see the comparative summary presented in Table I). The different approaches have been grouped according to the method employed to extract the area of the head. It is apparent from Table 1.2 that a large number of approaches[6,14,26,27,30,32] employ color-based segmentation, while another large group relies on the Viola–Jones detector.[10,13,19,29,35,37] The comparison of the approaches is not easy, because results are reported in different nonstandard ways. On the other hand, some approaches focus only on eye tracking, while other just consider particular facial cues, such as yawns.[8,27] However, it is possible to say that approaches based on color analysis are limited by illumination conditions and thus cannot be employed at night. This has motivated some researchers to use near infrared (IR) cameras, exploiting the retinas' high reflectivity to 850 nm wavelength illumination.[11,23] Some approaches employ neural networks to extract the head and main features,[7,31] while other rely on a variety of template matching schemes.[5,8,9,15,34] For a comprehensive survey on eye and gaze tracking models and approaches see ref. [12].

### *1.3. Research summary*

This paper presents an approach to robustly track a driver's eyes/head and detect the main visual cues that allow to monitor and predict the state of alertness: blinks, yawns, and eyebrow risings.[2,17] The novelty of the approach is in that it employs the Lucas–Kanade algorithm to track Shi–Tomasi salient points around the mouth and eyes, in addition to the Viola–Jones face detector. The salient points do not only provide information useful to track the motion of the head, but their analysis also makes possible to determine if the driver is yawning or rising eyebrows. The group of salient points also provides a reference relative to which the pupils can be located when the camera does not perceive the glare of the retina under IR illumination, and therefore, better tracking rates can be achieve despite temporary occlusions. The approach differs from refs. [11, 17] in that it employs continuous IR illumination without frame differencing to remove the background for eye detection purposes. The background subtraction may lead to the appearance of other blobs in addition to the eyes, and hence, a classification stage is required to select the right eye candidates. In ref. [17], a support vector machine classifier is proposed to verify the eye candidates. Another difference of our approach is that eye tracking is achieved in the image plane with the Lucas–Kanade tracker and a simple first-order infinite-impulse response discrete linear filter, while ref. [17] uses mean-shift tracker combined with a Kalman filter. The mean-shift procedure tracks the eyes based on appearance and is used when the orientation of the face is not frontal to the camera and the Kalman filter tends to fail or when the heads move to fast violating dynamic continuity assumptions of the Kalman filter. The approach in ref. [17] uses a number of fiducial points on the eyes, nose, and lips, and local graph representations to model and analyze facial expression. The fiducial points are described by their response to a Gabor wavelet filter, while our approach does not require the definition of specific point arrangements. Furthermore, our approach only requires simple gradient-based computations to obtain the Shi–Tomasi salient points. Eye detection and tracking in ref. [2] is solved in a similar way as proposed in ref. [17]. The two major differences between ref. [2] and ref. [17] are in that ref. [2] adaptively enlarges the pupils search when the Kalman filter fails instead of using the mean-shift procedure, and that unlike ref. [17], the approach in ref. [2] does not consider the analysis of yawns and eyebrow rising.

A detailed description of the proposed approach is presented in the next section. The results obtained show that our approach yields high detection and tracking rates with low false alarms (see Section 4). The main conclusions are presented in Section 5.

## 2. Proposed Approach

Let $I_k : \Omega \subset \mathbb{N}^2 \to \mathbb{Z}_+,\ p \mapsto I_k(p)$ for $p \in \Omega$, denote an image frame at sampling instant $k \in \mathbb{N}$, i.e., the set of intensity values $I_k(p)$ associated to each location $p$ of the optical plane $\Omega$. For clarity of exposition, it is also convenient to define image regions of interest within $\Omega$ for the location of the head, left eye, right eye, and mouth. These regions are specified in terms of bounding boxes $\mathcal{B}_k^s$ defined by three-component vectors $\mathcal{B}_k^s \stackrel{\text{def}}{=} (b_k^s, w_k^s, h_k^s)$, $s = h, l, r, m$, where $b_k^s \in \mathbb{N}^2$ is the upper-left coordinate of the region, $w_k^s, h_k^s$ are its width and height, respectively, and the labels $s = h, l, r, m$ indicate the head, left eye, right eye, and mouth. The sizes of the eyes and mouth bounding boxes are proportional to the size of the head's bounding box (see 1). The subimages corresponding to these regions of interest are denoted by $I_k^s$, $s = h, l, r, m$. It will also be useful to define a list of $n_k^s$ salient points within a given region $s$ at instant $k$ as an ordered

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

Table I. Summary of existing approaches for driver state of alert detection.

| Face detection technique | Approach characteristics | Remarks | Eyes detection rate (%) | Blinking detection/ false alarm rates (%) | Publication year |
|---|---|---|---|---|---|
| Color segmentation | | | | | |
| [14] | Eyes location from horizontal projection. Tracking by template matching. | 4 test subjects. | 99.1 | NA | 2006 |
| [6] | Eyes location from horizontal projection. Tracking by Dynamic Template Matching. | 2 test subjects. | 98.0 | NA | 2005 |
| [32] | Skin and eye color-based segmentation. | 37 test subjects. Requires illumination. | 96.9 | NA | 2008 |
| [30] | Eyes location from horizontal projection. | – | 95.0 | NA | 1999 |
| [27] | Lips and skin segmented with Fischer classifier and connected component analysis. Mouth is tracked in real-time using a Kalman Filter. | Results obtained from 150 images. Yawns detection rate of 95.3% and true negatives of 100%. | NA | NA | 2004 |
| [26] | Binarization and clustering. | – | 96.0 | NA | 2007 |
| Viola–Jones | | | | | |
| [13] | Eyes location from horizontal projection. Dynamic thresholding. | Poor accuracy for tilted faces (detection rate of 60.1%). | 88.2 | 89.3/14.1 | 2007 |
| [37] | Eyes location from projection. Tracking with Unscented Kalman Filter. | 3 test subjects. Few validations samples. | 99.5 | NA | 2006 |
| [29] | Eyes location and PERCLOS from horizontal projection. | 3 test subjects. Results obtained from 28,000 frames (20 min). | NA | 97.8/6.3 | 2009 |
| [35] | Binarization and histogram of the horizontal projection. | Results obtained from 5646 images. | NA | 97.1/0.3 | 2009 |
| [19] | Rectangle and texture features to extract eyes with Adaboost and SVM. | 4000 images from FERET database. | 96.8 | NA | 2009 |
| [10] | Face poses not detected by Viola–Jones are detected using a neural network. | – | 96.0 | 95.7/– | 2010 |
| Neural Networks | | | | | |
| [7] | Iris geometrical information and symmetry. | 3 test subjects. | NA | 90.0/– | 2004 |
| [31] | Horizontal projection and $2^{nd}$ derivative employed to detect eyelids using a SVM. | 200 images used for training and validation. | NA | 96.0/– | 2004 |
| IR-based Retina Detection | | | | | |
| [11] | Two different IR wavelengths to detect the retina. Gabor kernels employed to detect facial expressions. Tracking with Kalman Filter. | 1 test subject. | 99.1 | NA | 2002 |
| [23] | Illumination compensation and SVM to validate the detected eye candidates. | 10 test subjects. | 98.4 | NA | 2006 |
| Other | | | | | |
| [9] | Local Binary Pattern measure of texture around eyes and Gabor wavelets around mouth combined with Adaboost. | 30 test subjects. Yawns detection/misdetection = 90.8/6.9% | NA | 99.67/0.99 | 2008 |

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

Table I. Continued.

| Face detection technique | Approach characteristics | Remarks | Eyes detection rate (%) | Blinking detection/ false alarm rates (%) | Publication year |
|---|---|---|---|---|---|
| [5] | Curvature and upper eyelid curvature and aperture are employed using fuzzy fusion to detect driver fatigue. Tracking eyes withh Kalman Filter. | – | NA | 92.1/– | 2008 |
| [15] | A *separability filter*, which is a circular template matching approach is used together with the gradient of grayscale values to locate the eyes and detect blinking. | 10 test subjects. | NA | 95.45/– | 2002 |
| [34] | A binary eye pair template is used to roughly find eyes and SVM to validate candidates. | Only detects are of the eyes an not directly the pupils. | 97.0 | NA | 2006 |
| [8] | Employs a *Gravity-Center* template, together with grayscale projection and Gabor wavelets to detect yawns. | Yawns detection/misdetection = 95.0/6.0% | NA | NA | 2007 |

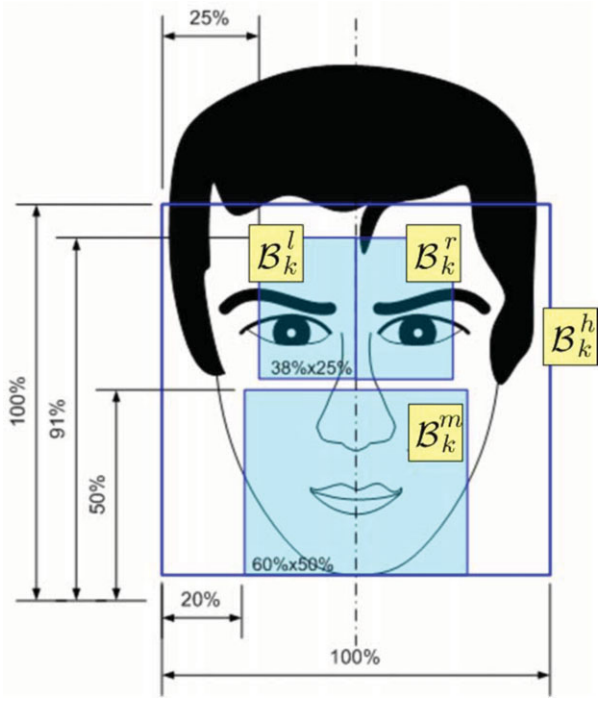



Fig. 1. Bounding boxes for the head, eyes, and mouth.

sequence of points $S_k^s \stackrel{\text{def}}{=} \{s_{k,i}^s \in \mathbb{N}^2, i = 1, 2, \ldots, n_k^s\}$. With this notation, the proposed approach for drowsiness and state of alert detection is summarized in the following steps:

(1) Initial head location and pose estimation using the Viola–Jones detector[33] to compute $\mathcal{B}_k^s$, and compute $\mathcal{B}_k^l$, $\mathcal{B}_k^r$, and $\mathcal{B}_k^m$ according to the layout rules shown in Fig. 1.
(2) Computation of set $S_k^s$ of Shi–Tomasi interest points[28] within the bounding boxes $\mathcal{B}_k^s$ and computation of each set's centroid $c_k^s$.
(3) Detection of the pupils' position $e_k^l$, $e_k^r$, computation of the *eyes' relative location vector* $d_k^s \stackrel{\text{def}}{=} e_k^s - c_k^s$, $s = l, r$, (corresponding to the pupils' position relative to the centroid), and update of the relative location vector estimate $\tilde{d}_k^s$.
(4) Detection of eyebrow rising and yawning by analyzing the temporal variation of the spatial distribution of salient points in sets $S_k^s$, $s = l, r, m$.
(5) Detection of eye blinking from intensity averages for image gradients of $I_k^s$, $s = l, r$.
(6) Rule-based classification of the state of alert.

The first two steps rely on the well-known computer vision techniques that will not be explained here for the sake of brevity. The remaining steps will be presented in next subsections.

### *2.1. Eyes location and head-pose estimation*

Because of the retina's high reflectivity to 850 nm wavelengths,[11,23] a simple adaptive threshold filter[16] allows to locate the pupils. The adaptive threshold filter takes the average pixel intensities in a $3 \times 3$ block and sets the output on the central pixel $p$ to 1, i.e., $I(p) = 1$ whenever

$$I(p) - \frac{\sum_{i \in \mathcal{N}_{3\times3}(p)} I(i)}{9} > 0.25 \frac{\sum_{i \in \mathcal{B}_k^s} I(i)}{w_k^s \cdot h_k^s}, \qquad (1)$$

where $\mathcal{N}_{3\times3}(p)$ is the $3 \times 3$ neighborhood of pixel $p$. The 0.25 coefficient was found experimentally to guarantee an adequate discrimination of the retina from other elements

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

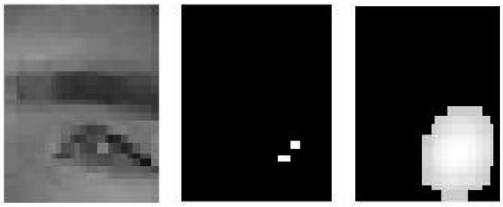


Fig. 2. Original image, adaptive threshold binarization image, and Gaussian filtered image.

in the image. As a result of the adaptive threshold filtering, binary images for the regions $\mathcal{B}_k^l$ and $\mathcal{B}_k^r$ are generated and convolved with an $11 \times 11$ Gaussian mask ($\sigma = 2.15$). This smoothing step is important because it allows to obtain a more accurate location of the pupils than the one that could be computed from the centroid of the nonzero values in the binary images due to the fact that filtering spreads out the nonzero values in the binary images, thus eliminating noise and generating peaks from which the position of the pupils can easily be obtained as

$$e_k^s = \arg\max_{p \in \mathcal{B}_k^s} \breve{I}_k^s(p), \ s = l, r, \tag{2}$$

where $\breve{I}_k^s$ represents the smoothed version of the binary image for region $s$. The results of the process of binarization by adaptive thresholding and posterior Gaussian smoothing of the area around an eye are illustrated in Fig. 2.

When daylight illumination is strong, pupils are smaller making the previous approach difficult to apply. To cope with this limitation, a second method based on the computation of the so-called horizontal and vertical projections in $\mathcal{B}_k^s$, $s = l, r$ is employed. The horizontal projection is a one-dimensional vector computed here by the following column-wise summation of elements in $I_k^s$:

$$h_k^s(j) \stackrel{\text{def}}{=} \sum_i I_k^s(i, j), \quad s = l, r. \tag{3}$$

The minimum of $h_k^s$ provides a good reference of the horizontal position of the pupil $e_k^s$ because the pupil appears less bright during daylight when it is more contracted and the eyelids and iris are darker than the rest of the average intensity of image $I_k^s$.

To find the vertical component of $e_k^s$, the vertical projection of $I_k^s$ cannot be used directly since often dark eyebrows may produce a wrong position measurement. To avoid this problem, a variant of the Laplacian kernel in the horizontal direction, defined as

$$\nabla_x^2 \stackrel{\text{def}}{=} \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & -2 & 1 \end{bmatrix}, \tag{4}$$

is applied to $I_k^s$, $s = l, r$. Observing that the vertical projection of $\nabla_x^2 * I_k^s$, given by the row-wise summation

$$v_k^s(i) \stackrel{def}{=} \sum_j \left[\nabla_x^2 * I_k^s\right](i, j), \tag{5}$$

presents strong components for the eye, while weaker ones for the eyebrow, a good reference of the pupils' vertical position can be obtained from the point at which $v_k^s$ is largest. More precisely, the horizontal and vertical positions of the pupils $e_{x,k}^s$ and $e_{y,k}^s$, $s = l, r$ are calculated as

$$e_{x,k}^s = \arg\min_{j \in \mathbb{N}} h_k^s(j), \tag{6}$$

$$e_{y,k}^s = \arg\max_{i \in \mathbb{N}} v_k^s(i). \tag{7}$$

In order to validate that pupils have been correctly located, the next geometrical constraints are applied

$$\left(e_{x,k}^l - b_{x,k}^l\right) - \left(b_{x,k}^r + w_k^r - e_{x,k}^r\right) \ < \ \delta_x, \tag{8}$$

$$\left|e_{y,k}^l - e_{y,k}^r\right| \ < \ \delta_y, \tag{9}$$

with thresholds $\delta_x, \delta_y$ defining the maximum horizontal and vertical distances.

The estimation of the pupils' location is made more robust to head movements and possible illumination variations by exploiting the good traceability of the Shi–Tomasi salient points[28] computed and stored in sets $S_k^s$ for the corresponding regions $\mathcal{B}_k^s$ $s = l, r$. The centroid of points $s_{k,i}^s \in S_k^s$, $i = 1, 2, \ldots, n_k^s$, given by

$$c_k^s = \frac{1}{n_k^s} \sum_{i=1}^{n_k^s} s_{i,k}^s, \tag{10}$$

provides a reference position from which it is possible to locate the center of the pupil, even when it may not be possible to detect the pupil directly in some frames due to the driver's motion or illumination changes. The pupil's estimated position relative to $c_k^s$ is given by

$$\hat{e}^s = c_k^s + \hat{d}_k^s, \tag{11}$$

where $\hat{d}_k^s$ is an estimate of the pupil's location vector relative to the centroid. A practical estimate of the relative position vector $d_k^s$ can be found by means of a simple first-order IIR running average filter of the form

$$\hat{d}_k^s = (1 - \alpha)\hat{d}_{k-1}^s + \alpha d_k^s, \tag{12}$$

with $d_k^s = e_k^s - c_k^s$, $\hat{d}_0^s = d_0^s$, $s = l, r$, where $\alpha$ is the update rate, normally set at 0.05.

Finding the head pose, in terms of standard pitch, roll, and yaw angles, can be solved by computing the deviation of the driver's nose from a reference point corresponding to the position of the nose when the person is looking forward. Finding the reference position is simple because it corresponds to the position for which the distance between the eyes, as observed in the image, is the largest. The roll angle $\phi$ is found by calculating the angle of the segment joining the eyes and a horizontal line, while the pitch $\psi$ and yaw $\theta$ angles are computed as $k_1 \Delta p_\psi$, $k_2 \Delta p_\theta$, where $k_1$ and $k_2$ are scaling factors and $\Delta p_\psi$ and $\Delta p_\theta$ are vertical and horizontal deviations (in pixels) with respect to the reference

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

point. This approximation is valid since $\sin(\alpha) \approx \alpha$ for small angles $\alpha$. On the other hand, true values of the pose angles are not necessary to determine if the person is distracted and looking away from the road ahead for too long periods of time.

### 2.2. Detection of eyebrow rising, yawns, and blinks

Analyzing the temporal variation of the spatial distribution of salient points $s^s_{k,i}$, $s = l, r, m$, with respect to the centroid $\hat{c}^s_k$ provides a good measure of eyebrow rising and yawns. To this end, the spatial distribution of salient points is first characterized in terms of their standard deviation at sampling instant $k$

$$\sigma^s_k = \sqrt{\frac{1}{n^s_k} \sum_{i=1}^{n^s_k} \left(s^s_{k,i} - c^s_k\right)^2}. \tag{13}$$

A reference value for the spatial distribution of salient points is defined simply as the running average of the salient points' standard deviation

$$\hat{\sigma}^s_k = (1 - \alpha)\hat{\sigma}^s_{k-1} + \alpha\sigma^s_k, \tag{14}$$

with an update rate $\alpha$ set to 0.05. Yawns or eyebrow risings are said to occur in frame $k$, if $\|\hat{\sigma}^s_{k,y} - \sigma^s_{k,y}\| > \lambda^s$,where $\hat{\sigma}^s_{k,y}$ and $\sigma^s_{k,y}$ are the vertical components of the distributions, and $\lambda^s$ is a threshold that must be found experimentally for the eyes ($s = l, r$) and mouth ($s = m$) regions.

Since the number of vertically oriented lines (i.e., lines with stronger horizontal gradients) increases when the eyes and mouth are open, the average intensities in the horizontal gradient image, given by

$$\bar{G}^s_{x,k} = \frac{1}{|\mathcal{B}^s_k|} \sum_{(i,j)\in\mathcal{B}^s_k} G^s_{x,k}(i, j),\ s = l, r, \tag{15}$$

provides a measure to determine if the person is blinking or yawning. An analogous filter to Eq. (14) provides a reference value $\hat{G}^s_{x,k}$ against which $\bar{G}^s_{x,k}$ must be compared to according to $\|\hat{G}^s_{x,k} - \bar{G}^s_{x,k}\| > \eta^s$, in order to establish if a blink or yawn has occurred. Here, $\eta^s$ is a threshold determined in such a way as to maximize the rate of detection, while minimizing the rate of false alarms. Figure 3 illustrates an open eye (upper right), a closed eye (lower left), and the corresponding responses obtained from the application of the Laplacian kernel $\nabla^2_x$ along the horizontal direction. When the eyes are open, the values of $\hat{G}^s_{x,k}$ and $\bar{G}^s_{x,k}$ are similar. However, when the eyes are closed, there is only a horizontal line, thus the average brightness is weaker and $\bar{G}^s_{x,k} < \hat{G}^s_{x,k}$ instead of $\bar{G}^s_{x,k} \approx \hat{G}^s_{x,k}$ as when eyes are open.

## 3. Experimental Setup

A standard off-the-shelf security camera with 22 integrated IR leds (see Fig. 4) was employed to film five different subjects during periods of more than 60 min. A sequence showing one of the test drivers waiting at a traffic light while being tracked by the system is presented in Fig. 5.

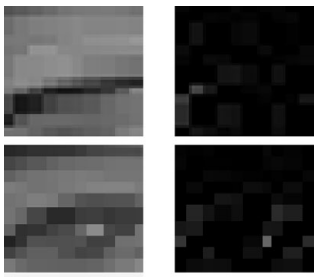


Fig. 3. Closed and open eye images, and their Laplacian-filtered versions.

The images in Fig. 5 show the bounding boxes for the head, eyes, and mouth detected by the system. The dots within each bounding box correspond to the salient points that are being tracked. Two larger dots show the location of the eyes as detected by the approach. When the eye markers disappear, as in the fourth frame of the sequence, this means that the system has detected a blink. The circle around the mouth that is shown in the third and fourth frames indicates that the spread of salient points around the mouth has increased above the threshold and hence that the driver may be yawning. It is to be noted that the possibility of confusing yawns with laugh or other gestures in which the driver opens the mouth above the threshold is high, and hence, the analysis of points around the mouth provides a less reliable cue for drowsiness estimation. This possibility of confusion is confirmed by the results presented in the next section.

The ideal camera location that minimizes misdetection and false alarms is at a position where the camera is centered with respect to the driver's head and normal to the face. However, for practical reasons and to avoid distracting the driver, the camera was mounted on the dashboard. Furthermore, initial tests revealed that there is no appreciable difference in detection and false alarm rates when the camera is mounted on the dashboard or the sun visor, since the reduction of the tracking accuracy was less than 1%. Both the dashboard and sun visor locations are better than lateral positions of the camera (e.g., car frame or rear mirror) because they lie on the plane of symmetry of the face; however, due to possible adjustments of the sun visor by the driver, this second alternative is less adequate for practical implementation.

An IR 850 nm wavelength filter was added to the camera in order to block other sources of illumination and glare of the sun. The camera's original resolution of 640×480 pixels was resampled to 384×288 pixel frames in order to achieve real-time processing at frame rates of 16.5 fps. A composite video to USB converter was employed to capture the frames directly on a laptop PC with a 2.2 GHz CPU and 3 GB of available RAM. The proposed approach was programmed using the OpenCV computer vision library.

The videos of the five drivers were first analyzed manually at slow playback speeds to detect and count blinks, eyebrow risings, and yawns correctly. A subset of video sequences containing the facial gestures of interest was selected for the evaluation of the proposed approach. The total length of the sequences amounts to 23 min of video and corresponds to 22,672 frames. This amount of frames guarantees the statistical significance of the results because all frames

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press




Fig. 4. IR camera 22 leds and the 850 nm filter.

containing facial gestures were included, and the ratio of frames containing blinks, eyebrow risings, and yawns to frames without facial gestures was reduced only by an order of 10. On the other hand, we noticed that false alarms occurred mainly due to difficulties in clearly observing some facial features or because some subjects would tend to move or talk more than others while driving, i.e., the false alarm rates are subject-dependent rather than time-dependent, as shown by the results presented in the next section.

In order to confirm the statistical validity of the results, a classic bootstrapping approach was employed.[3] Samples were drawn randomly with replacement from the input data sets to create 200 bootstrapped observation datasets. Confidence intervals for detection and false alarms rates were computed using the bootstrapped datasets.

## 4. Results

Detection and false alarm rates obtained for blinks, eyebrow risings, and yawns are summarized in Tables II–IV. These results were obtained selecting values for the thresholds $\lambda^l$, $\lambda^r$ (eyebrow risings), $\lambda^m$ (yawns), and $\eta^l$, $\eta^r$ (eye blinks) that maximized detection rates while minimizing the false positives. In other words, the threshold values correspond to the points in the ROC curves of Figs. 6 and 7 closest to the (0, 1) point.

Blinking was effectively detected 98.37% of the time on average with a false alarm rate of 0.98%, as shown in Table II. In three out of five subjects, blinking was detected every time it occurred. However, the blinking of the other two subjects was harder to detect on every occurrence because they tended to move more. The time duration of blinks was measured on each occasion they were detected, thus providing a good estimate for PERCLOS. The 95% confidence interval for the detection rate is ±1.49%, while the confidence interval for false alarm rate is ±0.38%, thus if the experiment would be

Table II. Blink detection results.

| Subject | Events | Detection rate (%) | False alarm rate (%) |
|---|---|---|---|
| 1 | 24 | 100.00 | 1.05 |
| 2 | 20 | 100.00 | 1.61 |
| 3 | 24 | 96.00 | 0.4 |
| 4 | 40 | 97.56 | 0.88 |
| 5 | 13 | 100.00 | 0.85 |
| Mean | | $98.37 \pm 1.49$ | $0.98 \pm 0.38$ |

Table III. Eyebrow rising detection results.

| Subject | Events | Detection rate(%) | False alarm rate (%) |
|---|---|---|---|
| 1 | 7 | 77.78 | 2.17 |
| 2 | 8 | 80.00 | 0.49 |
| 3 | 9 | 81.82 | 1.15 |
| 4 | 8 | 72.73 | 2.68 |
| 5 | 9 | 100.00 | 1.71 |
| Mean | | $82.00 \pm 9.91$ | $1.56 \pm 0.72$ |

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press



Fig. 5. Video sequence showing a driver's tracked eyes (frames 1, 2, 3, 5) and detected yawn (frames 3, 4).

Table IV. Yawn detection results.

| Subject | Events | Detection rate (%) | False alarm rate (%) |
|---|---|---|---|
| 1 | 5 | 71.43 | 4.34 |
| 2 | 10 | 100.00 | 2.94 |
| 3 | 8 | 90.00 | 13.80 |
| 4 | 8 | 66.67 | 3.76 |
| 5 | 10 | 100.00 | 1.71 |
| Mean | | $86.96 \pm 13.45$ | $5.29 \pm 4.90$ |

repeated again, in 95% of the repetitions the average detection rate would never be below 96.88%. On the other hand, the false alarm's confidence interval is also very narrow, and in 95% of the cases, the average false alarm would be less than 1.36%. These results are quite encouraging when compared to other results reported in the literature (see discussion in Section 1.2).

Eyebrow risings were detected on average 82.00% of the time with a very low false alarm rate of 1.56%, as shown in Table III. In this case, the confidence interval for the average detection rate is $\pm 9.91\%$, which is large compared to the confidence interval for blinkings. Similarly, Table IV presents yawn detection and false alarm rates, which are on average 86.96% and 5.29%, respectively. Once again, the 95% confidence intervals for yawn measurements are relatively large when compared to those for blinks. This is largely due to the fact that the computation of eyebrow risings and yawns are done in terms of the motion of salient points, unlike blink detection that is based on the analysis of the illuminated retina. Comparing Tables II–IV, it is possible to observe that blinking is a rather stable feature among the different subjects, while on the other hand, the motion of salient points varies greatly from one driver to another depending on their facial gestures and movements, and thus, eyebrow rising and yawns are less reliable for drowsiness detection. It is also to be noted that the false alarm rate

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

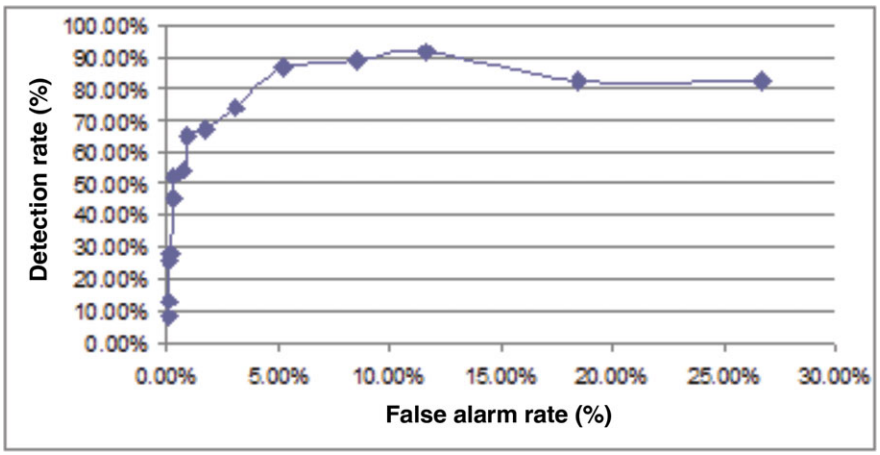



Fig. 6. ROC curve for yawns.

Table V. Eyes tracking results.

| Subject | Frames | Tracking rate (%) |
|---|---|---|
| 1 | 2768 | 93.39 |
| 2 | 6122 | 99.46 |
| 3 | 5219 | 98.58 |
| 4 | 3310 | 94.46 |
| 5 | 5253 | 98.50 |
| Mean | | $97.10 \pm 2.39$ |

for yawns is larger than the one for blinking and eyebrow rising. This is explained by the fact that it is easy to confuse yawns with other gestures typical of a person laughing or talking with exaggerated lip movements. Another difficulty in detecting eyebrow rising and yawning correctly arises from occlusions occurring when the drivers move their hands close to their faces, e.g., when scratching or covering their mouths. If the hands or some other occluding object moves in front of the face, some salient points are lost thus temporarily shifting the centroids. If the shifting lasts for a long 60–120 s interval, the filter will not be able to eliminate this perturbation and errors in eye tracking will be induced.

Drivers' eyes were tracked correctly 97.10% of the time on average as shown by the results reported in Table V. Considering the 95% confidence interval, the average tracking rate could lie in the interval between 94.71% and 99.49%, which is better than several of the existing methods summarized in Section 1.

## 5. Conclusion

A computationally effective method for real-time eyes tracking and driver's drowsiness estimation was presented. The proposed method is robust to occlusions because it relies on the motion analysis of a cloud of salient points, which are selected using the Shi–Tomasi approach and tracked with the Lucas–Kanade tracker. Unlike some of the existing methods, the proposed approach does not require an explicit geometric model for the eyes, nose, and mouth regions. Instead, our approach measures changes in the spatial distribution of the salient points around the eyes and mouth to determine whether the person may be yawning or blinking. Tracking of the centroids of the groups of selected points within regions of interest around the eyes and mouth allows to keep track of the driver's pupils at all times without computationally demanding face detection processes executed on every frame. The Viola–Jones approach for face detection is employed only to obtain an initial estimate of the eyes and mouth location, and every certain frames, when a large movement

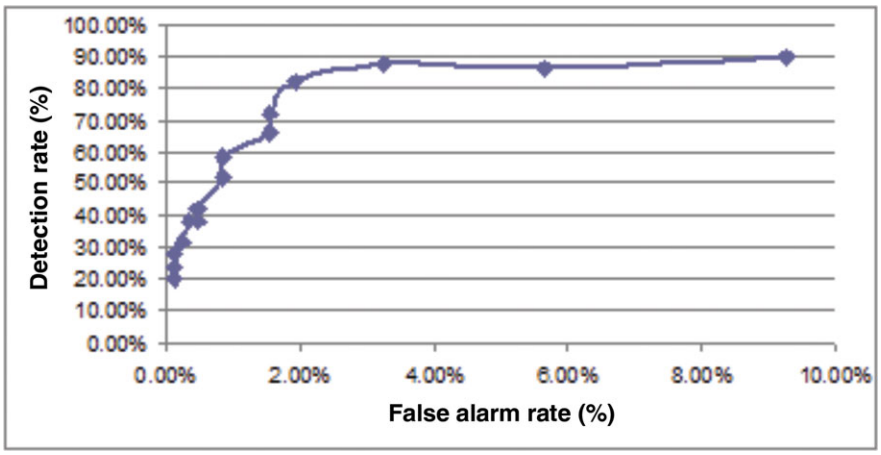



Fig. 7. ROC curve for eyebrow rising.

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

of the head is detected. The approach performs under day or night illumination conditions because it uses a circular array of infrared leds integrated into a standard off-the-shelf IR security camera and an 850 nm infrared passband filter added to block illumination variations produced during the day by external sources.

Results showed that blinking can be detected on average $98.37 \pm 1.49\%$ of the time with an average false alarm rate of $0.98 \pm 0.39\%$. Detection of yawns and eyebrow rising is more difficult because head movements and changes in illumination can influence the selection of some of the salient points. Nonetheless, reasonable detection rates between 82%–87% can also be achieved with low false alarm rates in the interval 1.5%–5.5%. Increasing detection rates is possible by reducing the thresholds, but false alarms rates will increase since some facial gestures or talking are easier to confuse with yawns. Finally, it is to be noted that the performance of the proposed approach compares favorably with respect to other methods reported in the literature. Furthermore, the approach can easily be implemented with currently available low-cost security cameras that include IR leds.

Ongoing research is concerned with the modeling of driver motion dynamics and the development of algorithms to detect occlusions. Addressing these issues is necessary to implement more sophisticated estimation and prediction strategies for an improved handling of those situations in which salient points are affected by driver actions, like moving hands close to the face, that difficult the detection of yawns and eyebrow risings.

## Acknowledgment

This project has been supported by the National Commission for Science and Technology Research of Chile (Conicyt) under Fondecyt Grant 11060251.

## References

1. S. Alvarez, M. A. Sotelo, M. Ocaña, D. F. Llorca, I. Parra and L. M. Bergasa, "Perception advances in outdoor vehicle detection for automatic cruise control," *Robotica* **28**(5), 765–779 (Sep. 2010).
2. L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea and M. E. López, "Real-time system for monitoring driver vigilance," *IEEE Trans. Intell. Transp. Syst.* **7**(1), 63–77 (Mar. 2006).
3. A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Application* (Cambridge University Press, Cambridge, UK, 1997).
4. A. de la Escalera and J. M. Armingol, "Vehicle detection and tracking for visual understanding of road environments," *Robotica* **28**(6), 847–860 (Dec. 2010).
5. W. Dong, P. Qu and J. Han, "Driver Fatigue Detection Based on Fuzzy Fusion," *Proceedings of the Chinese Control and Decision Conference (CCDC '08)*, Yantai, Shandong, China (Jul. 2008) pp. 2640–2643.
6. W. Dong and X. Wu, "Driver Fatigue Detection Based on the distance of Eyelid," *Proceedings of the IEEE International Workshop on VLSI Design and Video Technology*, Suzhou, China (May 2005) pp. 365–368.
7. T. D'Orazio, M. Leo, P. Spagnolo and C. Guaragnella, "A Neural System for Eye Detection in a Driver Vigilance Application," *Proceedings of the Seventh International IEEE Conference on Intelligent Transportation Systems*, Washington, D.C., USA (Oct. 2004) pp. 320–325.
8. X. Fan, B.-C. Yin and Y.-F. Sun, "Yawning Detection for Monitoring Driver Fatigue," *Proceedings of the International Conference on Machine Learning and Cybernetics*, Hong Kong (Aug. 2007) vol. 2, pp. 664–668.
9. X. Fan, B. Yin and Y. Sun, "Nonintrusive Driver Fatigue Detection," *Proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC '08)*, Hainan, China (Apr. 2008) pp. 905–910.
10. M. J. Flores, J. M. Armingol and A. Escalera, "Real-time warning system for driver drowsiness detection using visual information," *J. Intell. Robot. Syst.* **59**, 103–125 (Jun. 2010).
11. H. Gu, Q. Ji and Z. Zhu, "Active Facial Tracking for Fatigue Detection," *Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV '02)*, Orlando, Florida, USA (2002) pp. 137–142.
12. 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.* **32**(3), 478 –500 (Mar. 2010).
13. T. Hong, H. Qin and Q. Sun, "An Improved Real Time Eye State Identification System in Driver Drowsiness Detection," *Proceedings of the IEEE International Conference on Control and Automation (ICCA '07)*, Guangzhou, China (May 30–Jun. 1, 2007) pp. 1449–1453.
14. W.-B. Horng, C.-Y. Chen, Y. Chang and C.-H. Fan, "Driver Fatigue Detection Based on Eye Tracking and Dynamic Template Matching," *Proceedings of the IEEE International Conference on Networking, Sensing and Control*, Taipei, Taiwan (Mar. 2004) vol. 1, pp. 7–12.
15. T. Ito, S. Mita, K. Kozuka, T. Nakano and S. Yamamoto "Driver Blink Measurement by the Motion Picture Processing and Its Application to Drowsiness Detection," *Proceedings of the IEEE Fifth International Conference on Intelligent Transportation Systems*, Singapore (2002) pp. 168–173.
16. A. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall, Englewood Cliff, NJ, USA, 1986).
17. Q. Ji, Z. Zhu and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Trans. Veh. Technol.* **53**(4), 1052–1068 (Jul. 2004).
18. S. K. L. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.* **55**(3), 173–194 (2001).
19. H. Lu, W. Zhang and D. Yang, "Eye Detection Based on Rectangle Features and Pixel-Pattern-Based Texture Features," *Proceedings of the International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS '07)*, Xiamen, China (Nov. 28–Dec. 1, 2007) pp. 746–749.
20. J. F. May and C. L. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transp. Res.* **12**(3), 218–224 (2009).
21. NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness, "Drowsy driving and automobile crashes," Report No. DOT HS 808 707, National Center on Sleep Disorder Research, National Heart, Lung, and Blood Institute, and National Highway Traffic Safety Administration, Washington, D.C. (1998).
22. D. Olmeda, A. de la Escalera and J. M. Armingol, "Far infrared pedestrian detection and tracking for night driving," *Robotica* **29**, pp. 495–505 (2010).
23. I. Park, J.-H. Ahn and H. Byun, "Efficient Measurement of Eye Blinking under Various Illumination Conditions for Drowsiness Detection Systems," *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, Hong Kong (2006) vol. 1, pp. 383–386.
24. M. Pedan, R. Scurfield, D. Sleet, D. Mohan, A. A. Hyder, E. Jarawan and C. Mathers (eds.), *World Report on Road Traffic Injury Prevention* (WHO Press, Geneva, Switzerland, 2004).
25. C. A. Pickering, K. J. Burnham and M. J. Richardson, "A Review of Automotive Human Machine Interface Technologies and Techniques to Reduce Driver Distraction," *Proceedings of the Second Institution of Engineering and*

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press

*Technology International Conference on System Safety*, London, UK (Oct. 2007) pp. 223–228.
26. H. Qin, Y. Gao and H. Gan, "Precise Eye Location in Driver Fatigue State surveillance System," *Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICVES '07)*, Beijing, China (Dec. 2007) pp. 1–6.
27. W. Rongben, G. Lie, T. Bingliang and J. Lisheng, "Monitoring Mouth Movement for Driver Fatigue or Distraction with One Camera," *Proceedings of the Seventh International IEEE Conference on Intelligent Transportation Systems*, Washington, D.C., USA (Oct. 2004) pp. 314–319.
28. J. Shi and C. Tomasi, "Good Features to Track," *Proceedings of the 9th IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, Washington, USA (Jun. 1994) pp. 593–600.
29. M. H. Sigari, "Driver Hypo-Vigilance Detection Based on Eyelid Behavior," *Proceedings of the Seventh International Conference on Advances in Pattern Recognition (ICAPR '09)*, Kolkata, West Bengal, India (Feb. 2009), pp. 426–429.
30. S. Singh and N. P. Papanikolopoulos, "Monitoring Driver Fatigue Using Facial Analysis Techniques," *Proceedings of the IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems*, Tokyo, Japan (1999) pp. 314–318.
31. M. Suzuki, N. Yamamoto, O. Yamamoto, T. Nakano and S. Yamamoto, "Measurement of Driver's Consciousness by Image Processing—A Method for Presuming Driver's Drowsiness by Eye-Blinks Coping with Individual Differences," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '06)*, Taipei, Taiwan (Oct. 2006) vol. 4, 2891–2896.
32. P. R. Tabrizi and R. A. Zoroofi, "Open/Closed Eye Analysis for Drowsiness Detection," *Proceedings of the First Workshops on Image Processing Theory, Tools and Applications (IPTA '08)*, Sousse, Tunisia (Nov. 2008) pp. 1–7.
33. P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proceedings of the 16th IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii (Jun. 2001) vol. I, pp. 511–518.
34. Q. Wang, W. Yang, H. Wang, Z. Guo and J. Yang, "Eye Location in Face Images for Driver Fatigue Monitoring," *Proceedings of the Sixth International Conference on ITS Telecommunications*, Chegdu, China (Jun. 2006) pp. 322–325.
35. C. Xu, Y. Zheng and Z. Wang, "Efficient Eye States Detection in Real-Time for Drowsy Driving Monitoring System," *Proceedings of the International Conference on Information and Automation (ICIA '08)*, ZhangJiaJie, Hunan, China (Jun. 2008), pp. 170–174.
36. J. H. Yang, Z.-H. Mao, L. Tijerina, T. Pilutti, J. F. Coughlin and E. Feron, "Detection of driver fatigue caused by sleep deprivation," *IEEE Trans. Syst. Man Cybern.* **39**(4), 694–705 (Jul. 2009).
37. Z. Zhang and J. S. Zhang, "Driver Fatigue Detection Based Intelligent Vehicle Control," *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, Hong Kong (2006) vol. 2, 1262–1265.

https://doi.org/10.1017/S0263574711000749 Published online by Cambridge University Press