PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

# DRIVER ALERT STATE AND FATIGUE DETECTION USING OPTICAL FLOW ANALYSIS AND DRIVER'S KINEMATICS TO ACHIEVE ROBUST EYE TRACKING

**JAVIER E. JIMÉNEZ PINTO**

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
MIGUEL A. TORRES T.

Santiago de Chile, January 2010

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

# DRIVER ALERT STATE AND FATIGUE DETECTION USING OPTICAL FLOW ANALYSIS AND DRIVER'S KINEMATICS TO ACHIEVE ROBUST EYE TRACKING

**JAVIER E. JIMÉNEZ PINTO**

Members of the Committee:
MIGUEL A. TORRES T.
MARCELO GUARINI H.
RENATO SALINAS S.
JOAQUÍN DE CEA C.

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, January 2010

*Gratefully to my family, my mentor*

*and friends*

## ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I want to thank my supervisor Prof. Dr. M. Torres-Torriti from the Pontificia Universidad Católica de Chile whose assistance, encouragement and advice helped me in all the time of research for and writing of this thesis.

I gratefully thank my father Luis Jiménez, my mother Monica Pinto and my brother Camilo Jiménez for giving me the support and strength to complete this work.

Finally i gratefully acknowledge my friends that helped on the validation tests of my work, whom without their help this investigation would not have been possible.

## TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF TABLES

**ABSTRACT**

Assessing a driver's state of awareness and fatigue is especially important to reduce the number of traffic accidents often involving bus and truck drivers, who must work during several hours under monotonous driving conditions. The main challenge, in resolving the state of alert, is determine the driver's blinking (PERCLOS). To do so, an IR camera is used, processing the images in order to find the eyes, and determine if they are open or closed. Direct detection methods such as Viola-Jones, have the disadvantage of operating only with a particular pose of the object sought, in this case, the head and eyes. To overcome this problem, our approach combines the driver's kinematics and the motion analysis of Shi-Tomasi salient features within the face to determine head pose. With this information is possible to achieve a continuous and robust tracking of the eyes, resistant to occlusions, and computationally fast. The tracking has shown to be highly effective and allows a successful 99.41% rate detection of the eyes in normal conditions. To detect blinking, a set of filters that enhances horizontal intensity changes, and the advantage of the high reflectivity of the retina to near infrared illumination, employing a camera with an 850 nm wavelength filter, allows a 98.2% rate detection of blinking. Head pose, besides its use on the eye tracking, helps to determinates the driver's attention on the road and can also be used as a drowsiness cue, thought, blinking (PERCLOS) it will be used as the principal indicator.

**Keywords: Alert state assessment , fatigue detection , drowsiness detection , driver assistance , IR eye tracking, PERCLOS, image processing, driver's kinematics**.

## RESUMEN

Obtener el estado de alerta y la fatiga del conductor es particularmente importante para reducir el número de accidentes de tránsito que usualmente involucran a conductores de buses y camiones, los que están sometidos a largas horas de manejo bajo condiciones monótonas. El desafío principal en la resolución del estado de alerta, es determinar el parpadeo (PERCLOS). Para ello, una cámara infrarroja es utilizada, procesando las imágenes a fin de encontrar los ojos, y determinar si están abiertos o cerrados. Métodos de detección directa como Viola-Jones, tienen la desventaja de que sólo funcionan con una pose en particular del objeto buscado, en este caso, la cabeza y los ojos. Para superar este problema, nuestro enfoque combina la cinemática del conductor y el análisis del movimiento de los puntos salientes de Shi-Tomasi dentro de la cara para determinar la pose de la cabeza. Con esta información es posible lograr un tracking continuo y robusto de los ojos, resistente a las oclusiones y computacionalmente rápido. El tracking ha demostrado ser altamente eficiente, y permite una detección exitosa del 99.41% en condiciones normales. Para detectar el parpadeo, un grupo de filtros que resaltan los cambios horizontales de intensidad en el área del ojo, y la alta reflectividad de la retina a la iluminación en infra-rojo cercano usando una cámara con filtro de 850nm de longitud de onda permitió una deteccion del 98.2% de los parpadeos. La pose de la cabeza, ademas de su uso en el tracking de los ojos, ayuda a determinar la atención del conductor en el camino asi como también, como una característica del sueño, sin embargo, el parpadeo (PERCLOS) será usado como indicador principal.



**Palabras Claves: Evaluación de estado de alerta, detección de la fatiga, detección de somnolencia, asistencia al conductor, tracking de los ojos en IR, PERCLOS, procesamiento de imágenes, cinemática del conductor**.

# 1. INTRODUCTION

Road accidents take a heavy financial and social toll on national economies. The economic cost of traffic incidents is estimated to be 1% of gross national product in low-income countries, 1.5% in middle-income countries and 2% in high-income countries, totaling a global cost of US$518 billion per year (Pedan et al., 2004). Without appropriate actions to improve education, law enforcement, infrastructure and technology, a global increase of 67% is expected by year 2020. Although global statistics about accidents attributed to fatigue and drowsiness are not available because in many countries such details are not reported or classified, the number of incidents in high-income countries is not negligible. For example, the National Highway Traffic Safety Administration (NHTSA) reported as much as 56.000 accidents back in 1996 (Administration, 1998), which increased to 1.35 million in 2002 (Royal, 2002). The latter is about 0.7% of the reported accidents. 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 due to fatigue, stress or distraction can be found in (Flores, Armingol, & Escalera, 2008 ; Pickering, Burnham, & Richardson, 2007). In this context, developing systems to monitor a driver's state of awareness is fundamental. Though, a number of methods, like EEG (electroencephalogram) or EOG (electro-oculogram) are efficient to measure drowsiness (Chang, Lim, Kim, & Seo, 2007 ; Liang et al., 2005 ; Wright & McGown, 2001), they are invassive, producing disconfort and eventually the disconnection of the measuring system by the driver, so the development of a non invassive system is essential.

## 1.1. Objectives

The main objective is to determine, in real time, the driver's drowsiness state in a non-invasive way. The proposed system relies on CV(computer vision) techniques and the analysis of images obtained using a near IR camera with a 850nm filter. Drowsiness state must be classified in to three different levels: awake (when the driver is fully attentive), drowsy (when the driver is becoming sleepy) and asleep (when the driver has fallen asleep, requiring an alarm to be triggered). The proposed drowsiness detection system must also fulfill the following requirements:

- Must be robust against partial occlusion whenever the driver moves or blocks the camera momentarily.
- Must have detection rates close to 100% in order to be reliable.
- Must detect head pose, which is also an indicator of level of attention.
- Must have low false alarm rates.

## 1.2. Hypotheses

The main hypotheses are:

- Percentage of closure time of the eyes (PERCLOS) can be obtained using an IR camera with an 850nm filter.
- Rules can be defined using PERLCLOS to determine the driver's level of drowsiness.
- The face can be modeled as a coplanar set of points (SPG).
- Head pose can be estimated comparing the SPG projection into the 2D camera coordinates with the displacement of these points using Lucas-Kanade's optical flow algorithm.

## 1.3. Existing Approaches

Several studies exist about physiological cues that can be used to assess a driver's awareness. Some techniques can be invasive, but fortunately, there are many behavioral changes that provide visual cues, namely, eye-blinking frequency and closure percentage over some window of time (PERCLOS), yawn frequency, head movement, eye-gaze, among other facial expressions. Hence, a variety of systems based on computer vision techniques have been proposed. A summary is presented in Table 1.3, in which the approaches have been grouped according to the technique employed to extract the area of the head. As may be seen from Table 1.3, a large number of them (Dong & Wu, 2005 ; Horng, Chen, Chang, & Fan, 2004 ; Qin, Gao, & Gan, 2007 ; Rongben, Lie, Bingliang, & Lisheng, 2004 ; Singh & Papanikolopoulos, 1999 ; Tabrizi & Zoroofi, 2008) employ color-based segmentation approaches, while another important number of approaches relies on the Viola-Jones detector (Flores et al., 2008 ; Hong, Qin, & Sun, 2007 ; Lu, Zhang, & Yang, 2007 ; Sigari, 2009 ; Xu, Zheng, & Wang, 2008 ; Zhang & Zhang, 2006). The comparison of the approaches is not easy because results are reported in different non-standard ways. Moreover, some approaches only track the eyes, while other focus on particular facial cues, such as yawning (Fan, Yin, & Sun, 2007 ; Rongben et al., 2004). However, it is possible to say that approaches based on color analysis are limited by illumination conditions and often cannot be applied at night. This has motivated some researchers to use near infrared (IR) cameras, exploiting the retinas' high reflectivity to 850 nm wavelength illumination (Gu, Ji, & Zhu, 2002 ; Park, Ahn, & Byun, 2006). Some approaches employ neural-networks to extract the head and main features (D'Orazio, Leo, Spagnolo, & Guaragnella, 2004 ; Suzuki, Yamamoto, Yamamoto, Nakano, & Yamamoto, 2006), while other rely on a variety of template matching schemes (Dong, Qu, & Han, 2008 ; Fan et al., 2007 ; Fan, Yin, & Sun, 2008 ; Ito, Mita, Kozuka, Nakano, & Yamamoto, 2002 ; Wang, Yang, Wang, Guo, & Yang, 2006).

## 1.4. Summary of Contributions

The main contributions of this work are in that:

- The approach combines the kinematics driver's model, Shi-Tomasi salient points and Nelder-Mead simplex minimization to overcome the problem of partial detection. Increasing in the case of Viola-Jones from 38.02% to a 99.41% detection rate.
- By matching salient points in 3D space to points on the 2D image plane, head pose is estimated and loss of information can be minimized whenever occlusions occur. This improves the ability of the method to track the eyes and correctly

| Face Detection Technique | Approach Characteristics | Remarks | Eyes Detection Rate [%] | Blinking Detection / False Alarm Rates [%] | Publication Year |
|---|---|---|---|---|---|
| Color Segmentation | | | | | |
| (Horng et al., 2004) | Eyes location from horizontal projection. Tracking by template matching. | 4 test subjects. | 99.1 | N.A. | 2006 |
| (Dong & Wu, 2005) | Eyes location from horizontal projection. Tracking by Dynamic Template Matching. | 2 test subjects. | 98.0 | N.A. | 2005 |
| (Tabrizi & Zoroofi, 2008) | Skin and eye color-based segmentation. | 37 test subjects. Requires illumination. | 96.9 | N.A. | 2008 |
| (Singh & Papanikolopoulos, 1999) | Eyes location from horizontal projection. | – | 95.0 | N.A. | 1999 |
| (Rongben et al., 2004) | 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. Yawning detection rate of 95.3% and true negatives of 100%. | N.A. | N.A. | 2004 |
| (Qin et al., 2007) | Binarization and clustering. | – | 96.0 | N.A. | 2007 |
| Viola-Jones | | | | | |
| (Hong et al., 2007) | Eyes location from horizontal projection. Dynamic thresholding. | Poor accuracy for tilted faces (detection rate of 60.1%). | 88.2 | 89.3/14.1 | 2007 |
| (Zhang & Zhang, 2006) | Eyes location from projection. Tracking with Unscented Kalman Filter. | 3 test subjects. Few validations samples. | 99.5 | N.A. | 2006 |
| (Sigari, 2009) | Eyes location and PERCLOS from horizontal projection. | 3 test subjects. Few validations samples. | N.A. | 97.8/6.3 | 2009 |
| (Xu et al., 2008) | Binarization and histogram of the horizontal projection. | Results obtained from 5646 images. | N.A. | 97.1/0.3 | 2009 |
| (Lu et al., 2007) | Rectangle and texture features to extract eyes with Adaboost and SVM. | 4000 images from FERET database. | 96.8 | N.A. | 2009 |
| (Flores et al., 2008) | Face poses not detected by Viola-Jones are detected using a neural network. | – | 97.3 | 97.0/– | 2008 |

TABLE 1.1. *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 |
|---|---|---|---|---|---|
| Neural Networks | | | | | |
| (D'Orazio et al., 2004) | Iris geometrical information and symmetry. | 3 test subjects. | N.A. | 90.0/– | 2004 |
| (Suzuki et al., 2006) | Horizontal projection and $2^{nd}$ derivative employed to detect eyelids using a SVM. | 200 images used for training and validation. | N.A. | 96.0/– | 2004 |
| IR-based Retina Detection | | | | | |
| (Gu et al., 2002) | Two different IR wavelengths to detect the retina. Gabor kernels employed to detect facial expressions. Tracking with Kalman Filter. | 1 test subject. | 99.1 | N.A. | 2002 |
| (Park et al., 2006) | Illumination compensation and SVM to validate the detected eye candidates. | 10 test subjects. | 98.4 | N.A. | 2006 |
| Other | | | | | |
| (Fan et al., 2008) | Local Binary Pattern measure of texture around eyes and Gabor wavelets around mouth combined with Adaboost. | 30 test subjects. Yawning detection/misdetection = 90.8/6.9% | N.A. | 99.67/0.99 | 2008 |
| (Dong et al., 2008) | Curvature and upper eyelid curvature and aperture are employed using fuzzy fusion to detect driver fatigue. Tracking eyes with Kalman Filter. | – | N.A. | 92.1/– | 2008 |
| (Ito et al., 2002) | 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. | N.A. | 95.45/– | 2002 |
| (Wang et al., 2006) | A binary eye pair template is used to roughly find eyes and SVM to validate candidates. | – | 97.0 | N.A. | 2006 |
| (Fan et al., 2007) | Employs a *Gravity-Center* template, together with grayscale projection and Gabor wavelets to detect yawning. | Yawning detection/misdetection = 95.0/6.0% | N.A. | N.A. | 2007 |

TABLE 1.2. *Summary of existing approaches for driver state of alert detection. PART II*

estimate PERCLOS measure, necessary to determine the driver level of attentiveness.

## 2. PROPOSED APROACH

The proposed approach can be divided into two main groups of tasks (see fig. 2.1). The first one, initializes the algorithm when the driver is looking to the front, providing the driver's nominal pose. The second group is the recursive loop that tracks the head pose, taking as starting point the driver's nominal pose determined at the initialization stage.

The first step in the initialization process detects the driver's face using the classic Viola-Jones algorithm (Viola & Jones, 2001), which is briefly explained in section 2.1. The box bounding the area of the head found will be called $\mathcal{B}^h$. The next step finds the salient points inside $\mathcal{B}^h$, which will be used for tracking. The algorithm to find salient points will be described in section 2.4. Once the salient points are determined, a mesh or grid of salient points is created. The grid of salient points will be referred to as $SPG$ (Salient Points Grid). The $SPG$ is defined as a group of coplanar points forming a non deformable structure that will have 5 degrees of freedom (DOF), replicating the driver's kinematics. The $SPG$ will be described in greater detail in section 2.5. The initialization process ends with the detection of the eyes using the Viola-Jones algorithm trained for such purpose. The search is carried out in an area defined within $\mathcal{B}^h$ where the eyes are expected to be found according to the head's anatomy.

Once the initialization is completed , the head tracking task is carried out by comparing the motion of the $SPG$ in 3D space with the motion of salient points in the 2D image plane. To this end, the salient points in the image are tracked individually using the Lucas-Kanade (LK) approach to optical flow computation (details on the LK algorithm will be provided in section 2.2). The next step involves solving a matching problem that requires to find the motion of the $SPG$ using the kinematic model of the driver, such that the projection back to 2D coordinates in the image plane of salient points in the $SPG$, which are in 3D space, coincides as best as possible with the new location of the salient points tracked in 2D space using the LK approach. The driver's estimated pose is obtained as a solution of the matching problem that is further explained in section 2.8.

Finally, the driver's eyes are sought within expected regions according to the newly determined head pose. Solving the head tracking problem significantly improves tracking of the eyes, even if during several frames the eyes cannot be detected directly by analyzing the image. Continuous tracking of the eyes allows knowing the location of the pupils with great accuracy and if they are visible or not due to blinking by the driver. Analyzing the blinking yields measures such as PERCLOS (Wierwille, Ellsworth, Wreggit, Fairbanks, & Kirn, 1994), which provide an indication of the driver's state of drowsiness More details about PERCLOS will be provide in section 2.3 .

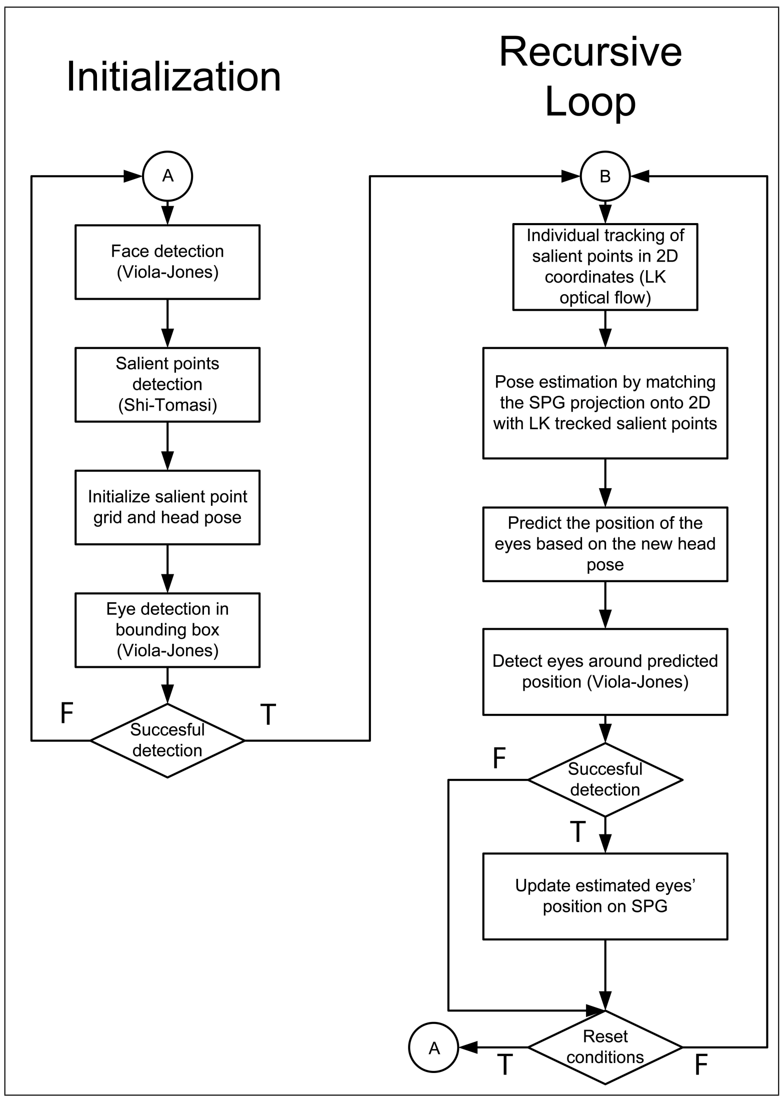


FIGURE 2.1. *Flow chart of the proposed method.*

### 2.1. Viola-Jones Detector

The Viola-Jones detector uses features obtained from the image response to Haar-like kernels, such as the shown in fig. 2.2.

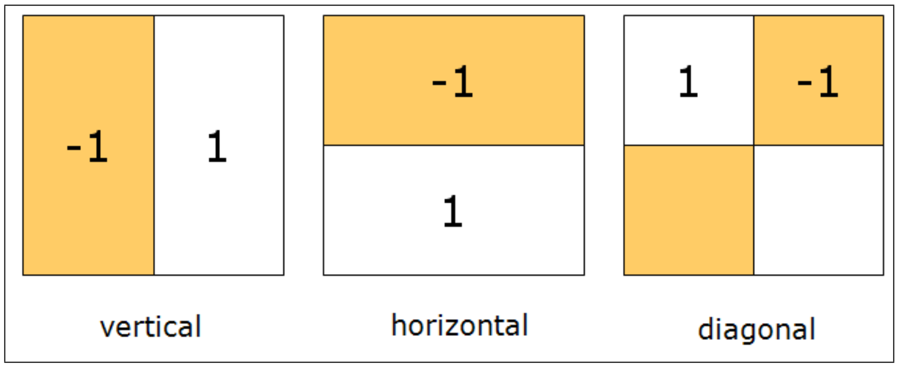


FIGURE 2.2. *Examples of Haar-like features.*

One of the contribution of Viola-Jones was to use integral image-based representations (Crow, 1984) for the faster computation of the Haar-like features. Viola and Jones propose the use of the adaptive boosting technique (known as Adaboost) to select simple linear classifiers with high detection rates and weak false positive rates. The use of these classifiers in cascade (fig. 2.3) allows excellent detection rates and reduced false alarms. From a practical stand point, the advantage of this approach is in that classification of features can be computed efficiently in short time by the cascade of simple classifiers, even if the training stage to solve the classifier parameters is computationally more demanding than that of other classification approaches, but that are slower on the classification stage.

### 2.2. Optical Flow

The optical flow approach proposed by Lucas and Kanade (Lucas & Kanade, 1981) rests on three assumptions: the first one is that a pixel of an object on the scene does not change its intensity when the object moves from frame to frame, the second one is that the objects only make small movements from frame to frame, and the final assumption is that neighboring points belong to the same surface, have similar motion, and are neighbors in the image plane if they are neighbors in the scene.

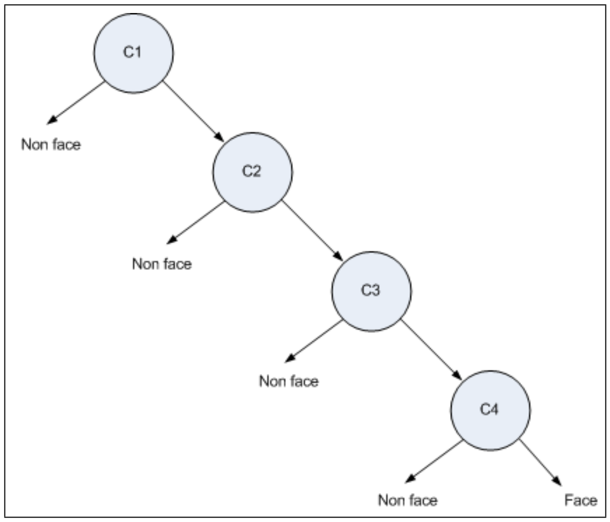


FIGURE 2.3. *Face detection using a cascade of linear clasifiers ($C_i$).*

If $I(x(t), t) : \mathbb{N}^2 \times \mathbb{R} \rightarrow \mathbb{R}$ denotes the brightness of a moving pixel $x(t) \in \mathbb{N}^2$ at time $t$, the basic idea of the Lucas-Kanade approach is to take the spatial derivative at pixel $x$, denoted by $I_x = \frac{\partial I}{\partial x}$, and the temporal derivative of the same pixel at time $t$, denoted by $I_t = \frac{\partial I}{\partial t}$, and then calculate how much the point moves from one frame to another using the following approximation:

$$\|\vec{v}\| \approx -\frac{I_t}{I_x}. \tag{2.1}$$

This approximation is valid for the one-dimensional case and is possible because:

$$\left.\frac{dI}{dt}\right|_{x(t),t} = \underbrace{\left.\frac{\partial I}{\partial x}\right|_t}_{I_x} \underbrace{\frac{dx}{dt}}_{\vec{v}} + \underbrace{\left.\frac{\partial I}{\partial t}\right|_{x(t)}}_{I_t} = 0 \;\Rightarrow\; v = -\frac{I_t}{I_x} \tag{2.2}$$

by the brightness constancy assumption. Considering fig. 2.4, which shows a smooth edge consisting of bright values on the left and dark values on the right and that is moving towards the right along the $x$-axis, the motion velocity $\vec{v}$ is the ratio between the rise in time and the displacement in space. Since brightness undergoes small variations in real images, this process must be iterated using the computed value of $\vec{v}$ as starting solution for the next iteration. Another important aspect is that in 2D images, there are two spatial variables, but just one equation, i.e. the system is under determined, and therefore a block of neighboring pixels must be employed to add constraints. The problem becomes an over determined system of equations, which can be solved using a pseudoinverse to find the two-dimensional velocity vector $\vec{v}$.

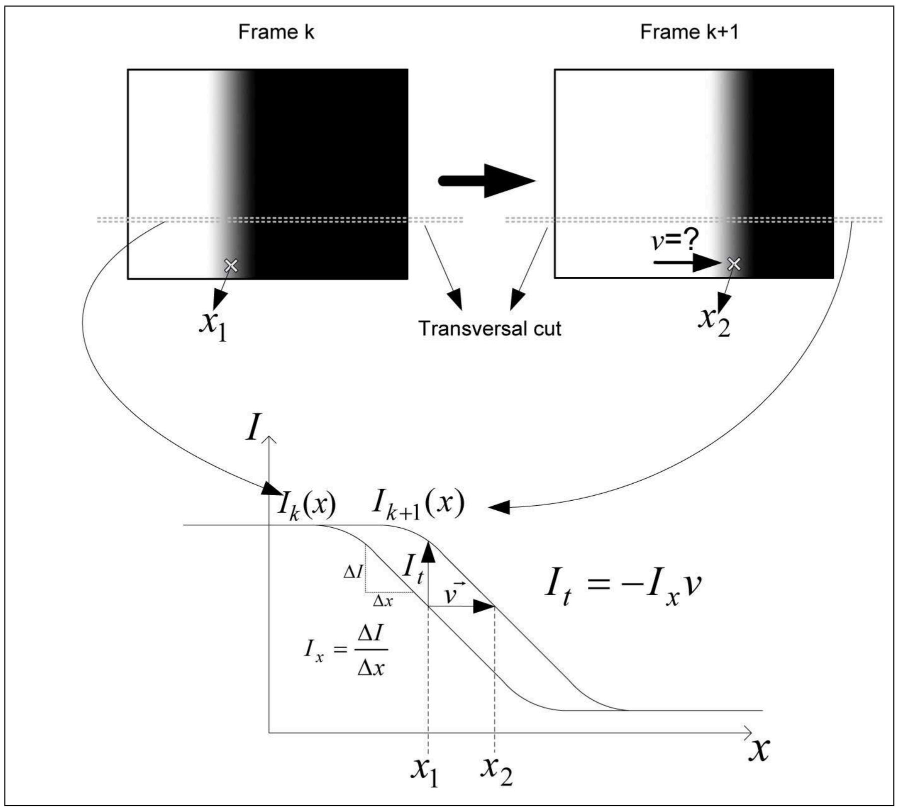


FIGURE 2.4. *Example of optical flow on 1D.*

To achieve more accuracy the operation is repeated until the difference of the intensities of the point in both frames is below certain threshold, or a maximum number of iterations is reached.

## 2.3. PERCLOS

PERCLOS is a measure that has been proven to have a high correlation with the level of drowsiness of persons (Sigari, 2009 ; Xu et al., 2008 ; Grace et al., 1998). The most important study was made by the Federal Highway Administration of the United States (Wierwille et al., 1994), and shows how the PERCLOS of a person increases directly with the level of fatigue. The test was made keeping ten subjects awake for 42 hours and taking tests of PERCLOS and reaction time every two hours. The test results show an average correlation between the reaction time and drowsiness of 0.878.

PERCLOS is calculated as a ratio between the amount of time the eyes are closed with respect to the total time:

$$PERCLOS = \frac{t_c}{t_c + t_o} \tag{2.3}$$

where $t_c$ is the time the eyes are closed and $t_o$ is the time the eyes are open. This measure is typically computed over moving windows lasting one minute. It will be shown in the results that our approach can measure PERCLOS with high accuracy in normal conditions (the direction of the head does not separate more than 15 degrees from the nominal position and the subjects did not use glasses or sunglasses). The results obtained from our tests with subjects in simulated driving conditions are consistent with the previous studies about PERCLOS reported in the literature.

## 2.4. Salient Points of the Face

Once face detection succeeds, the next step is to localize salient points that will be easy to track from frame to frame. To determine these points, the Shi-Tomasi corner detector (Shi & Tomasi, 1994) is implemented and applied to the image region $\mathcal{B}^h$. The objective is to use the points on the face that are good candidates for tracking regardless of their position. In this way, every characteristic, including the ones that change from face to face like scars, beard, freckles, etc., will be used to estimate head's pose and eyes' position.

The Shi-Tomasi corners are a variant of Harris corners (Shi & Tomasi, 1994), in which the autocorrelation matrix of first order horizontal ($I_x$) and vertical ($I_y$) image derivatives is used:

$$M(x,y) = \begin{bmatrix} \sum\limits_{-K\le i,j\le K} w_{i,j} I_x^2(x+i,y+j) & \sum\limits_{-K\le i,j\le K} w_{i,j} I_x(x+i,y+j) I_y(x+i,y+j) \\ \sum\limits_{-K\le i,j\le K} w_{i,j} I_x(x+i,y+j) I_y(x+i,y+j) & \sum\limits_{-K\le i,j\le K} w_{i,j} I_y^2(x+i,y+j) \end{bmatrix} \tag{2.4}$$

where 2K is the size of the surrounding region used to calculate the matrix.

The Shi-Tomasi approach select a salient point whenever $min(\lambda_1, \lambda_2)$ are higher, than a certain threshold, where $\lambda_1$ and $\lambda_2$ are the eigenvalues of $M(x, y)$.The number of salient points selected is restricted to a maximum number denoted by $N_{max}$ that in our implementation was set to 96.

### 2.5. Salient Points Grid

Visual tracking using a single camera of any object in 3D space is a challenging problem because, as it is well known, some information is lost due to the perspective projection that maps points in 3D space to the bidimensional space of the optical plane of the camera. However, if a priori knowledge of the object geometry is available, then it is possible to recover 3D motion and pose information. The approach in this work exploits the fact that all salient points belong to the driver's head (see fig 3.1 in the implementation section for an example of salient points detected for one of the test subjects), which is a rigid object of standard size and at a regular nominal distance from the camera. Since the salient points belong to a 3D object, the geometric constraints which they satisfy are only fulfilled in 3D coordinates, and therefore, tracking them correctly requires projecting salient points in 2D coordinates back to 3D coordinates in order to comply with the constraints of motion in 3D space.

The method for tracking the driver using a single camera uses three matrices,

$$P_{2D:3D} : x \in \mathbb{R}^2 \to X \in \mathbb{R}^3 \tag{2.5}$$

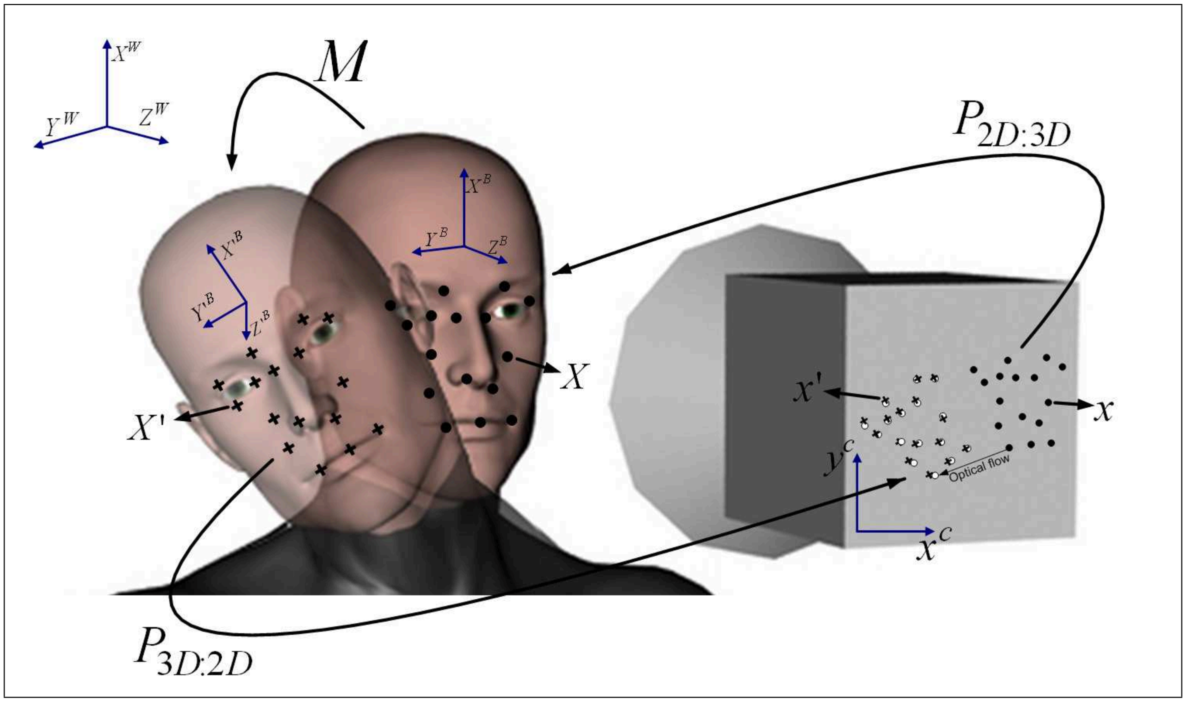


FIGURE 2.5. *Transformation of the salient points made by the proposed approach*

$$M : X \in \mathbb{R}^3 \rightarrow X' \in \mathbb{R}^3 \tag{2.6}$$

$$P_{3D:2D} : X' \in \mathbb{R}^3 \rightarrow x' \in \mathbb{R}^2 \tag{2.7}$$

and for simplicity of exposition it will be divided as follows:

- Applying the projection $P_{2D:3D}$ to salient points $x$ in the image back to points $X$ to the facial plane, as shown in fig. 2.5.
- Finding a motion matrix $M$ that maps points $X$ to points $X'$ according to the driver's body motion, such that the projection of $X'$ back to the image plane using the projection matrix $P_{3D:2D}$ results in points $x'$ that closely match the salient points found using the LK algorithm in the newly acquired frame.

The proposed approach just described relies on the following assumptions:

- The face can be modeled as a plane.
- The face is a non-deformable rigid object.

- The driver's motion can be represented by a two-link system with two articulated joints providing 5 DOF (shown in fig. 2.6).
- The distance between the joints is equal to the average length $H_b$ of an adult's torso.

By the above assumptions, the group of salient points form a non-deformable structure that will be called salient points grid or $SPG$. The salient points belonging to the $SPG$ of one of the test subjects is shown in fig 3.1 of the implementation section.

In the next sections, the kinematic model represented by the transformation matrix $M$ will be divided, as well as the projective transformation from camera to world coordinates ($P_{2D:3D}$) and vice versa ($P_{2D:3D}$).

## 2.6. Driver's kinematic model

The motion of the driver can be decomposed into head motion and torso motion as shown in fig. 2.6.

The motion model employs three coordinate systems: $\mathcal{S}^h$,$\mathcal{S}^b$,$\mathcal{S}^w$ for the head , body and world coordinates, respectively. It is assumed that initially $\mathcal{S}^h$,$\mathcal{S}^b$ and $\mathcal{S}^w$ are aligned and share the same origin coordinate. The first step in representing the position of the $SPG$ involves rotating $\mathcal{S}^h$ (and the $SPG$ fixed to $\mathcal{S}^h$) around $\mathcal{S}^b$ and translating, an $H_b$ distance, the rotated $\mathcal{S}^h$ along the x-coordinate of $\mathcal{S}^b$. This operation is generated by matrix:

$$M_h = \begin{bmatrix} c(\alpha_1)c(\beta_1) & c(\alpha_1)s(\beta_1)s(\gamma_1) - s(\alpha_1)c(\gamma_1) & c(\alpha_1)s(\beta_1)c(\gamma_1) + s(\alpha_1)s(\gamma_1) & H_b \\ s(\alpha_1)c(\beta_1) & s(\alpha_1)s(\beta_1)s(\gamma_1) + c(\alpha_1)c(\gamma_1) & s(\alpha_1)s(\beta_1)c(\gamma_1) - c(\alpha_1)s(\gamma_1) & 0 \\ -s(\beta_1) & c(\beta_1)s(\gamma_1) & c(\beta_1)c(\gamma_1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.8}$$

The next step rotates $\mathcal{S}^b$ around $\mathcal{S}^w$ using the rotation matrix:

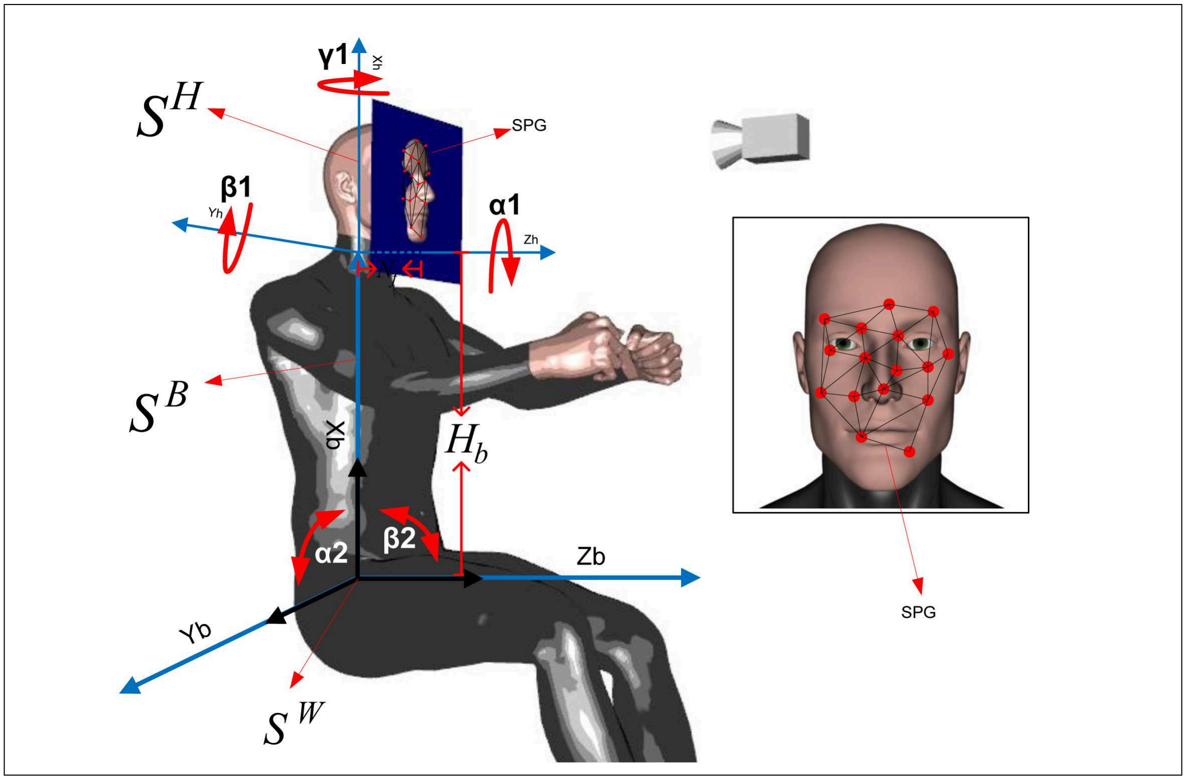


FIGURE 2.6. *Driver's kinematics.*

$$M_b = \begin{bmatrix} c(\alpha_2)c(\beta_2) & -s(\alpha_2) & c(\alpha_2)s(\beta_2) & 0 \\ s(\alpha_2)c(\beta_2) & c(\alpha_2) & s(\alpha_2)s(\beta_2) & 0 \\ -s(\beta_2) & 0 & c(\beta_2)c(\gamma_2) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.9}$$

Therefore, given angles $\alpha_1$,$\beta_1$,$\gamma_1$,$\alpha_2$,$\beta_2$ which define the driver's pose, the location of points $X^{spg} \in SPG$ in $\mathcal{S}^w$ coordinates will be given by:

$$X^W = M_B M_H X^{spg} \tag{2.10}$$

## 2.7. The $SPG$ and the Perspective Projection Model

In order to simplify the transformation of salient points in the image to points of the $SPG$ in 3D coordinates some useful coordinate transformations are introduced first.

The first transformation takes salient points in standard image coordinates (which assume the origin at the upper-left corner with the x-axis aligned with pixels rows towards the right and the y-axis aligned vertically with the columns), and expresses those coordinates with respect to a new coordinate frame whose origin is defined in terms of coordinates ($facepivot_x$,$facepivot_y$) located horizontally at the center of the head and vertically 1/3 below the center of the head. This coordinate also defines the position around which the facial plane in 3D coordinates can tilt when the head moves. To express salient points with respect to pivot coordinates the following transformation is employed:

$$x^p = CSx \tag{2.11}$$

where

$$S = \begin{bmatrix} Pix_h & 0 & 0 \\ 0 & Pix_w & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.12}$$

is a scaling matrix which transforms pixel units to metric units and

$$C = \begin{bmatrix} -1 & 0 & facepivot_x \\ 0 & 1 & facepivot_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.13}$$

transfers the coordinates of salient points in standard image coordinates to salient points in pivot coordinates, $Pix_h$ and $Pix_h$ are calculated with the camera intrinsic parameters (resolution and size of the CCD). It is to be noted that the latter are still 2D coordinates.

To take 2D points to the 3D space, more information besides the points' position on the camera's optical plane is needed. This information is the distance between the camera and the driver's head. Determining the approximate distance at which the head of the driver is located is possible using the distance between the eyes as yardstick since the interpupilar distance is relatively invariant for adult people. Here it will be assumed that the separation between the eyes is 63mm (Dodgson, 2004). If $\delta$ denotes the distance between the projection of the eyes onto the optical plane in millimeters, then the distance between the camera and the driver's face is given by:

$$D_{hc} = 63\frac{f}{\delta} \tag{2.14}$$

where $f$ is is the focal distance of the camera in millimeters. Using $D_{hc}$ (which usually takes values between 50cm and 80cm), the transformation matrix that projects the salient points back to the $SPG$ is

$$D = \begin{bmatrix} D_{hc}/f & 0 & 0 \\ 0 & D_{hc}/f & 0 \\ 0 & 0 & N_f \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.15}$$

where $N_f$ is the distance along z-coordinate between the pivot of the head coincident with the origin of $\mathcal{S}^{\mathcal{H}}$ and the $SPG$'s plane as shown in fig. 2.7. This distance must be taken into account because the face rotates about the neck, and the rotation point is not coplanar with the $SPG$'s plane.

With the previously defined matrices, the backprojection of 2D to 3D coordinates can be constructed as follows:

$$P_{2D:3D} = DCS. \tag{2.16}$$

The projection of SPG element coordinates back to the image plane requires two additional transformation matrices. One of this matrices is

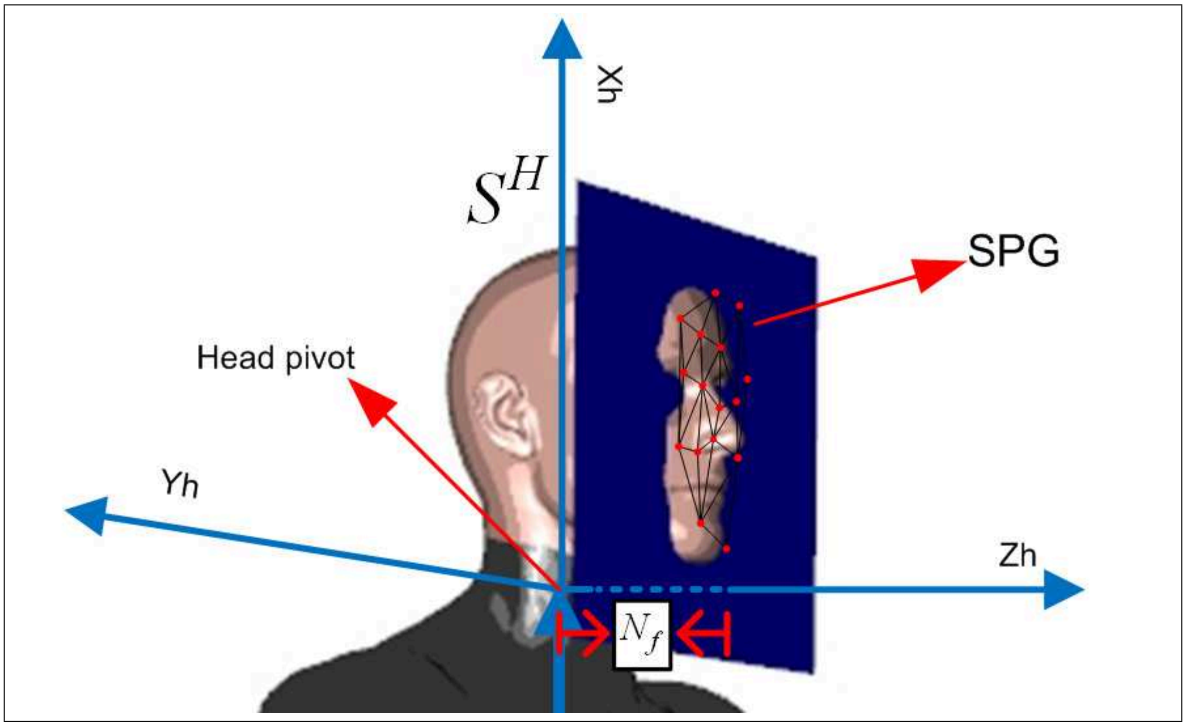


FIGURE 2.7. *Face pivot and $N_f$ distance.*

$$M_{pov} = \begin{bmatrix} -1 & 0 & 0 & H_b \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -(D_{hc} + N_f) \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.17}$$

whose point is to adjust the camera coordinates relative to the global world coordinate frame $\mathcal{S}^W$. The other matrix is the standard projection matrix:

$$H = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2.18}$$

Finally, the mapping of 3D point coordinates to 2D image coordinates can be defined by:

$$P_{3D:2D} = C^{-1} S^{-1} H M_{pov}. \tag{2.19}$$

It is to be noted that the combined transformation defined as:

$$T(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2,) = P_{3D:2D} M P_{2D:3D} \tag{2.20}$$

is a 3x3 matrix with full rank.

### 2.8. Head Pose Estimation

The proposed approach use the well known Lucas-Kanade's method to optical flow computation (Lucas & Kanade, 1981) to track point by point of the $SPG$ projected in the 2D image coordinates. As shown in fig. 2.8, some points may not be tracked correctly, especially if they are close to the boundaries of the head and they become occluded when the head turns. Bad tracking can also occur when the points are occluded by an external object like the driver's hand. In order to determine reliable points that were correctly tracked and discard wrongly tracked points, pixels neighboring the salient point in two consecutive frames are compared. This comparison is carried out in terms of the convolution of corresponding pixel blocks. The resulting value provides a measure of $w_i$ of the quality and reliability of the match which is used to assign weights to the different salient points.

If $x^i$ denotes the i-th salient point of the $SPG$ and $t^i$ the corresponding point tracked by LK algorithm, the head pose can be estimated by solving the following minimization problem:

$$\min_{\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2} \sum_i w_i \left\| T(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2) x^i - t^i \right\| \tag{2.21}$$

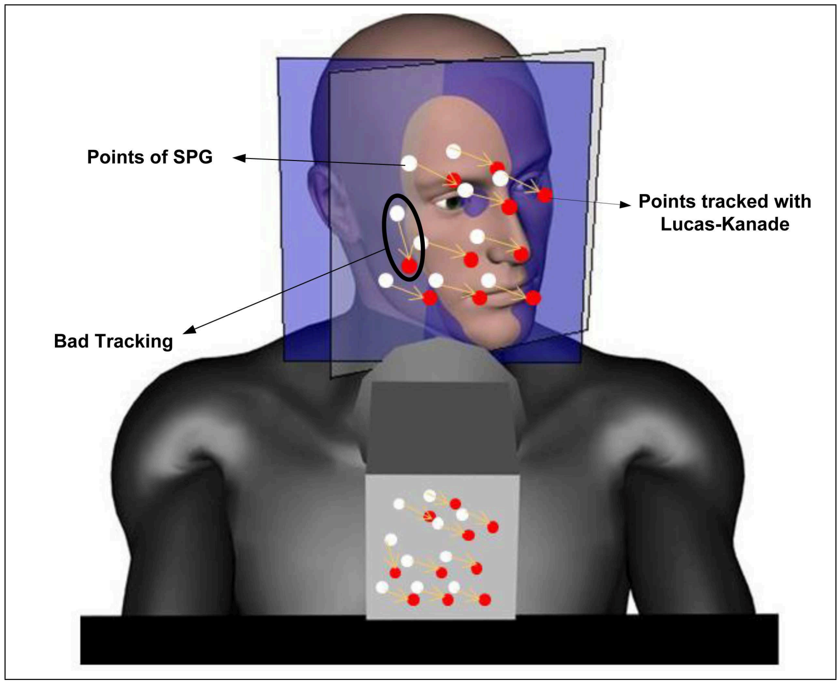


FIGURE 2.8. *Example of LK tracking between two frames with a point bad tracked.*

Once the head parameters are found it is possible to project the $SPG$, such that the $SPG$ points closely match the points tracked by the LK algorithm. If most of the salient points are well tracked by LK, the projected points of the $SPG$ will give a corrected position of the salient points, fixing the few bad points tracked as shown in fig. 2.9. The solution of the minimization problem employs as initial value the solution obtained in previous iterations. By doing so, falling into local minimum is avoided and the computation time is reduced, as compared to the one starting always from some nominal posture.

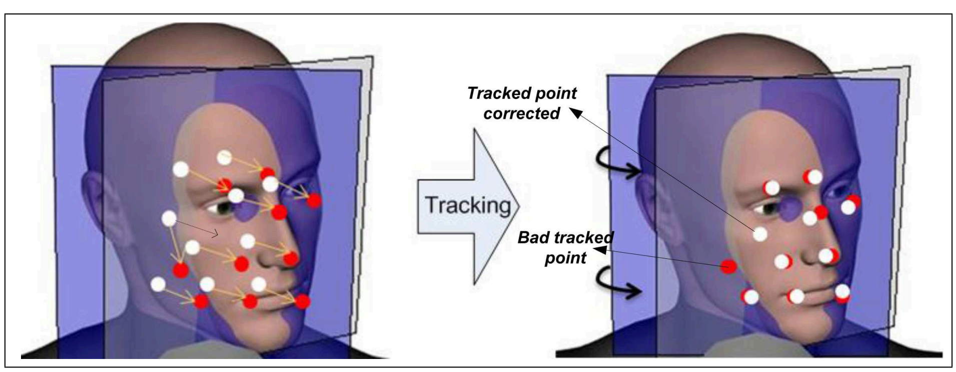


FIGURE 2.9. *Example of SPG tracking between two frames with a corrected point.*

## 2.9. Eyes Location and Tracking

The eyes' location is obtained using the Viola-Jones approach in a subwindow within $\mathcal{B}^h$. To achieve a continuous and reliable detection whenever the Viola-Jones algorithm fails to detect the eyes, especially due to changes in pose or due to occlusions, the proposed approach takes advantages of the fact that the $SPG$ is a non-deformable structure and that therefore prior identifications of the eyes (that are fixed points of the face) can be related to two additional points in the $SPG$. On every successful detection of the eyes using the Viola-Jones approach, an indication of the eyes' location is obtained. Each new hint of where the eyes are located can be employed to improve the location of the eyes within the $SPG$, which by the matrix transformations can be used to obtain the location of the eyes on the camera's optical plane in every frame, regardless of whether the Viola-Jones algorithm fails to find them. In this case, if $e_k^{c,l}$ and $e_k^{c,r}$ denote the respectively location of the left ($l$) and right ($r$) eye in image coordinates (indicated by $c$) for frame $k$, then

$$e_k^{spg,i} = T^{-1}(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2) e_k^{c,i},\ i = l, r \tag{2.22}$$

are the location of the eyes in $SPG$ coordinates. Hence the update of the estimated position of the eyes on the $SPG$ (denote by $\hat{e}_k^{spg,i}$) every time Viola-Jones successfully finds the eyes is computed as

$$\hat{e}_k^{spg,i} = (1-\alpha)\hat{e}_{k-1}^{spg,i} + \alpha e_k^{spg,i},\ i = l, r, \tag{2.23}$$

where $\alpha$ indicates the percentage of relevance of the raw measurement $e_k^{spg,i}$ that will be used to update $\hat{e}_{k-1}^{spg,i}$ i.e $\alpha$ is the so-called learning rate of the first order running average filter implemented by (2.23).

Having an estimated position of the eyes on the $SPG$ on every frame, the estimated position of the eyes in image coordinates is calculated by

$$\hat{e}_k^{c,i} = T^{-1}(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2)\hat{e}_k^{spg,i},\ i = l, r. \tag{2.24}$$

It will be shown in the results section that $\hat{e}_k^{c,i}$ has proven to be a very precise estimation of the eyes' position.

## 2.10. Reset Rules

As seen in the previous sections, the proposed approach estimates head pose with respect to an initial nominal pose, in other words, the head pose measure is differential. This is why, if small errors occur on some frames, these errors accumulate over time. Despite this, the eye detection system maintains its accuracy because every time a direct detection of the eyes succeeds, the estimation is updated correcting the cumulative error that may exist. To fix the accumulative error of the head tracking a set of rules to reset the head pose estimation is defined. If the head does not move outside $20°$ for any of the DOF ($\alpha_1$,$\beta_1$,$\gamma_1$,$\alpha_2$,$\beta_2$), or if the eyes are directly detected at least one time within a certain number of frames (100 frames were used in the tests), the tracking will be reset in a relatively long number of frames denoted as $n_l$, if any of the mentioned rules is broken, the tracking will be reset on a smaller number of frames that will be denoted $n_s$. On the tests, $n_l$ was set to 5000 and $n_s$ to 100, giving good tracking results by keeping the cumulative error to the minimum.

## 2.11. Detection of Blinking

A Laplacian filter is applied to an area of the image arround $\hat{e}_k^c$ denoted $\mathcal{E}_k^c$ with height equal to 33% of the height of the found face and width equal to 18% of the width of the found face. As result, an horizontal gradient image denote by $G_{x,k}^s(i, j)$ is obtained. Since the number of lines with vertical directions, i.e. lines with stronger horizontal gradients, increases when the eyes and mouth are open, the average intensities in the horizontal gradient

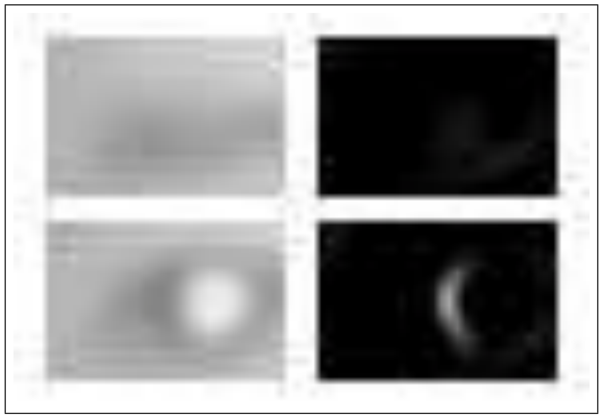

FIGURE 2.10. *Eye image and filter image for both open and close eye.*

image, given by

$$\bar{G}^s_{x,k} = \frac{1}{|\mathcal{B}^s_k|} \sum_{(i,j)\in\mathcal{B}^s_k} G^s_{x,k}(i,j),\ s = l, r, \tag{2.25}$$

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 blinking has occurred. Here $\eta^s$ is a percentage of $\hat{G}^s_{x,k}$ determined in such a way as to maximize the rate of detection, while minimizing the rate of false alarms. Figure 2.10 illustrates a closed eye (upper-left), an open eye (lower-left) and the corresponding responses obtained from the application $\nabla^2_x$ Laplacian filter in the horizontal direction. When the eyes are open, the value 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, resulting $\bar{G}^s_{x,k} < \hat{G}^s_{x,k}$ instead of $\bar{G}^s_{x,k} \approx \hat{G}^s_{x,k}$ as when eyes are open.

## 3. IMPLEMENTATION AND TESTING METHODOLOGY

To determine the efficacy of the proposed approach, a number of subjects participated in driving experiments, which lasted 45 minutes each. These experiments were carried in a simulator implemented to acquire driving behavior and measure each individual's reaction time. A snapshot of the software implemented to extract salient points and compute the PERCLOS measure is shown in fig 3.1.

To compare reaction times under drowsiness and full rest, the participants would drive in two sessions: one after sleep deprivation for one night, and the other a few days later after having had a full night of sleep. The next sections will describe how the driving tests and reaction time were obtained, as well as the details on the specifications and construction of the driving simulator.

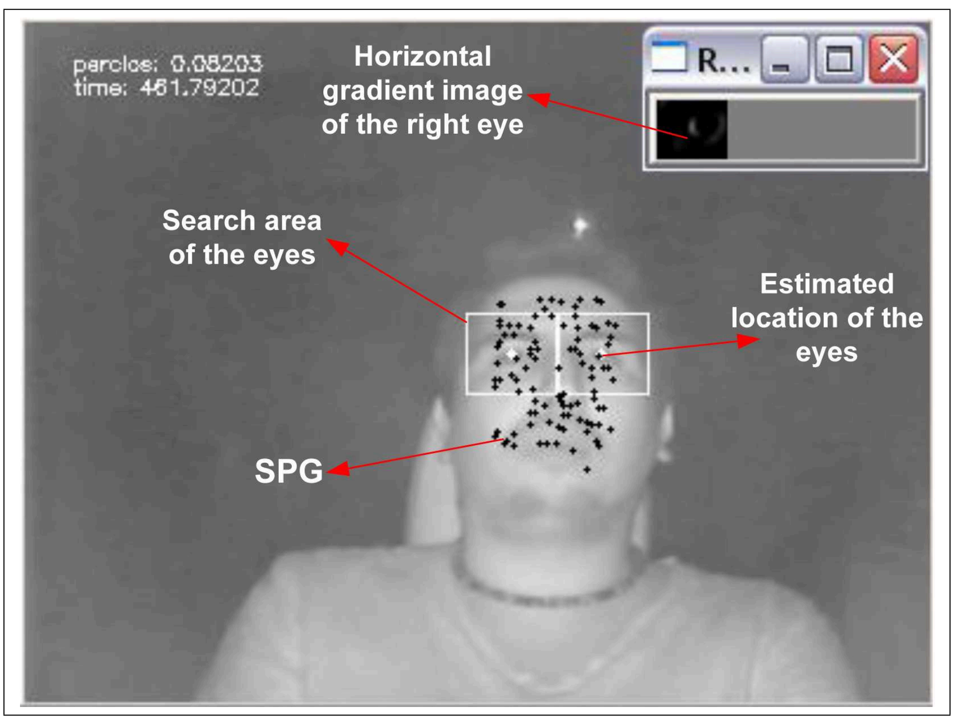


FIGURE 3.1. *Snapshot of the system running.*

### 3.1. Reaction Time and Driving Test

Measurement of reaction times was done by asking drivers to press a button as fast as possible whenever a green mark on the simulator's projection screen would turn red. The procedure was repeated 50 times before the driver would start driving. The elapsed time between each reaction test was random and could last between 2 to 10 seconds.

Once the tests to measure reaction time were completed, participants had to drive for 45 minutes along a rather monotonous track scenario simulating a dessert with hills and very few turns. The purpose of the chosen scenario was to induce drivers to fall asleep, while keeping visual distractors that could arouse driver's attention to a minimum.

### 3.2. Driving Simulator

The car simulator was built inside a closed lab with no external light sources using a Ford Escape 2009 seat and a Momo Racing Force Feedback Steering Wheel by Logitech, which included gas, brake and clutch pedals. A Viewsonic high resolution digital projector was used to project the scenes on a cylindrical projection plane, whose purpose was to immerse the driver into the virtual driving scenario and contribute to the realism perceived by the person by considering the effects of video motion on the peripheric vision. In other words, the curved backdrop surrounding the driver provides him or her an enhanced velocity sensation than what a planar surface would. The simulator also includes speaker to provide realistic sound. The software employed to create the driving environment is the open source driving simulator Racer (www.racer.nl). The simulator was configured to limit driving speed to $100\,km/h$

Fig. 3.2 shows the simulator layout, composed of the semicircular projection screen of $1.8\,m$ radius, the projector located $5.8\,m$ from the projection screen and $2.7\,m$ above the ground to avoid the car seat structure from casting shadows on the screen. As shown in fig. 3.2, the rear of the seat structure is $0.9\,m$ away from the center of the semicircular

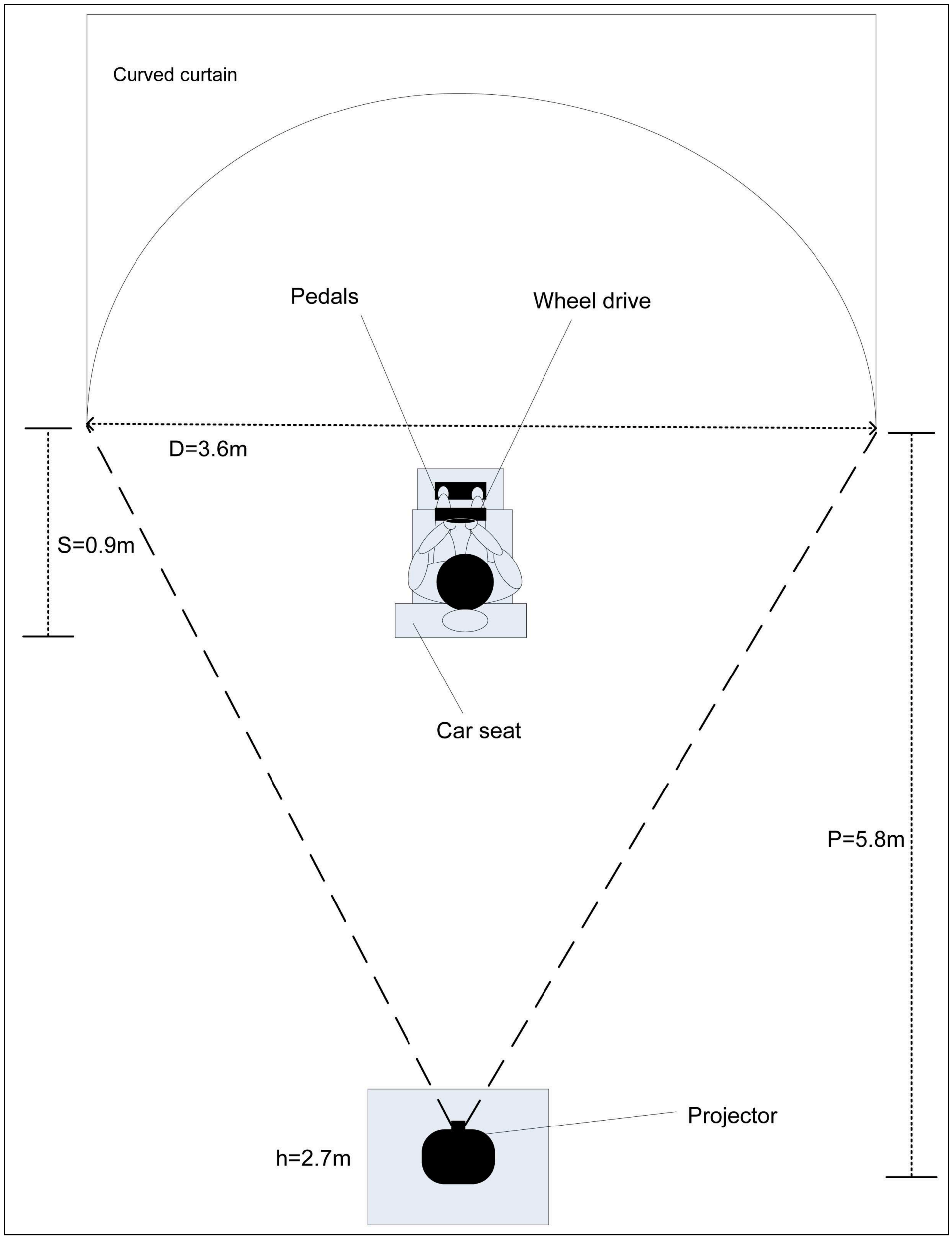


FIGURE 3.2. *Driving simulator layout.*

projection screen. This location ensures that the driver field of view subtends the whole projection screen and not just the central portion (see fig. 3.5). Fig. 3.3 shows the dimensions of the seat structure whose construction is shown in fig. 3.4.

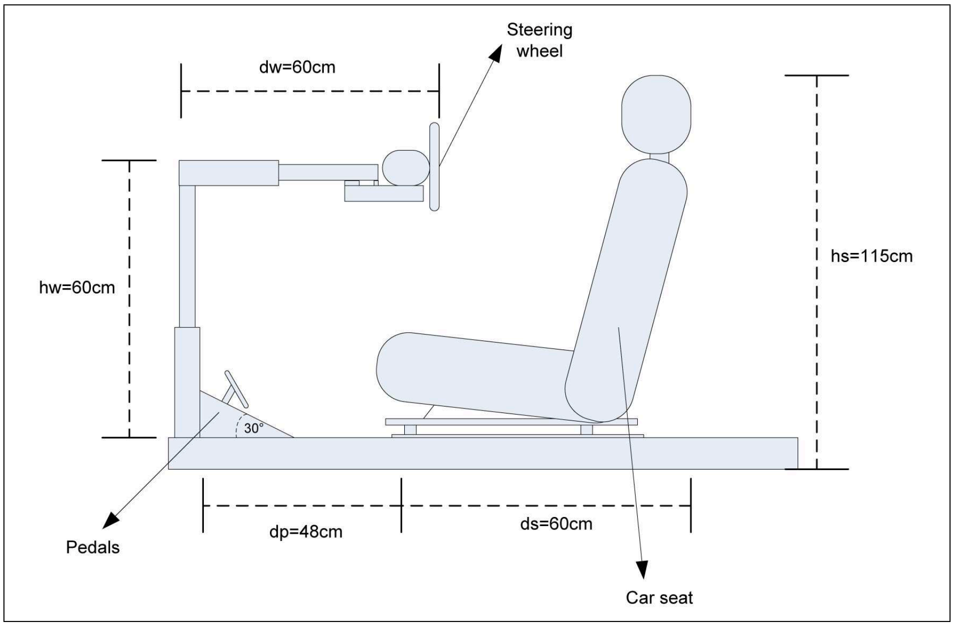


FIGURE 3.3. *Dimensions of the vehicle simulator steering-wheel and seat structure.*

FIGURE 3.4. *The constructed vehicle simulator structure.*

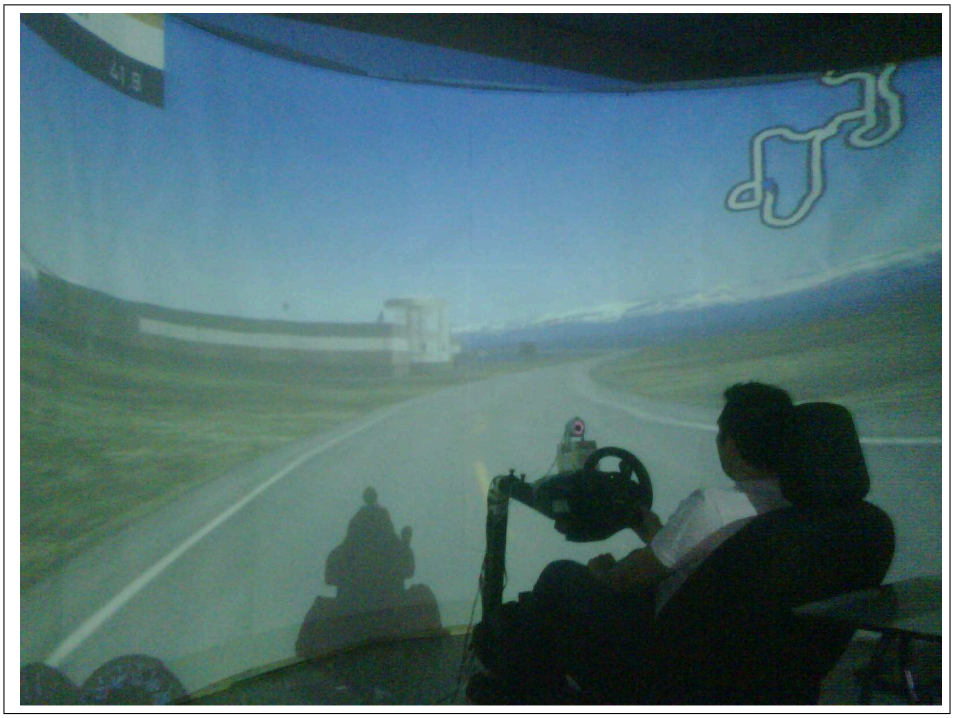

FIGURE 3.5. *Driving simulator during an test drive.*

### 3.3. Drowsiness Detection Sensor

The drowsiness detection sensor was implemented using a near infrared camera with a resolution of 640x480 pixels (see fig. 3.6). The camera is an of-the-shelf security camera available for less than USD 100.00. The camera includes an array of 21 infrared emitters. A $850\,nm$ filter was added to block interference from other sources of infrared radiation, especially the sun. A composite-video to USB converter was employed to capture the frames directly on a PC (approximate value USD 50.00). The processing algorithms were implemented in using OpenCV (the official website is http://sourceforge.net/projects/opencv/) and the Microsoft Visual C/C++ compiler executed on a laptop with a 2.2 GHz CPU and 3 GB of available RAM, delivering a frame rate of 16.5 fps for 384x288 pixel frames to achieve real-time processing capabilities.

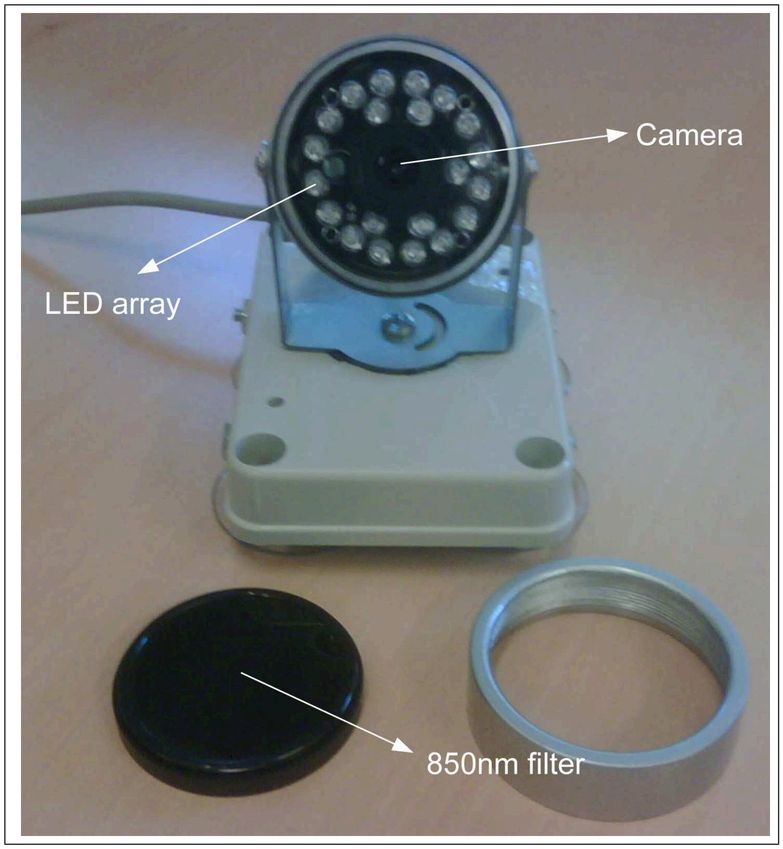


FIGURE 3.6. *IR camera with a 850nm filter and an array of 10 infrared emitters.*

## 4. EXPERIMENTAL RESULTS

In the next two sections this chapter presents and discusses the results concerning the detection of eyes and blinking, as well as the drowsiness detection results using the PERCLOS measure.

### 4.1. Eye and Blink Detection Results

Eye detection, eye tracking and blinking rates were calculated using a subset of five randomly chosen subjects out of the seventeen persons that volunteered for the experiments. The reason for using a subset and not the whole collected data is that the length of the videos ($45$ min per experiment) would have required an enormous amount of time of manual processing for ground truth extraction.

Eye blinking was effectively detected 98.37% of the time on average with a rate of false alarm of 0.98%, as shown in table 4.1. In three out of five subjects, blinking was detected every time it occurred. However, the third and fourth subjects were harder to detect on every occurrence because they tended to move more and change their eye gaze direction. The time duration of blinks was measured on each occasion they were detected, thus providing a good estimate for PERCLOS. Threshold values for eye blinking thresholds $\eta^s$, $s = l, r$ (defined in section 2.11), were chosen so as to yield the highest detection rates with lowest possible false alarm rate.

The proposed approach proved highly effective for eye tracking, yielding average tracking rates above 99%, as shown in Table 4.2. Except for two subjects that would move a lot while driving, the rest of the five drivers had perfect eye tracking rates of 100%, in spite of their motion and changes in the external illumination. In other words, the system is able to determine the location of the head and its pose 99.41% of the time on average.

If the Viola-Jones approach to eye detection is used alone, only 38.02% of the time the eyes can be detected and tracked, as shown in Table 4.3. This confirms the importance of

| 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 | 0.98 |

TABLE 4.1. *Blink detection results.*

| Subject | Frames | Tracking Rate [%] |
|---|---|---|
| 1 | 2768 | 100 |
| 2 | 6122 | 100 |
| 3 | 5219 | 100 |
| 4 | 3310 | 96 |
| 5 | 5253 | 99 |
| Mean | | 99.41 |

TABLE 4.2. *Eye tracking results with the proposed approach.*

| Subject | Frames | Tracking Rate [%] |
|---|---|---|
| 1 | 2768 | 22.24 |
| 2 | 6122 | 43.16 |
| 3 | 5219 | 21.18 |
| 4 | 3310 | 58.29 |
| 5 | 5253 | 44.33 |
| Mean | | 38.02 |

TABLE 4.3. *Direct eye detection results with the Viola-Jones aproach.*

including the kinematic model for the head motion as a way to improve the detection and tracking rates.

FIGURE 4.1. *Subjects used to calculate the tracking rate of the eyes and the detection rate of blinking.*

The subjects used to estimate the tracking rate of the eyes and the detection rate of blinking are shown in fig. 4.1.

## 4.2. Drowsiness Detection Using the PERCLOS Measure

To determine the state of drowsiness of the driver and establish its relation with the PERCLOS's value found by the proposed approach, the different states of drowsiness were separated in; awake, semi-drowsy and drowsy. The awake state is defined as the state of a person who had a full night sleep (between 6 and 8 hours) and that does not yawn or fall asleep during the test (falling asleep is defined as the total closure of the eyes for more than 2 seconds). The semi-drowsy state is defined according to any of the two criteria: (i) the subject had a full night of sleep, but did yawn during the test, or (ii) the subject did not sleep all night, but does not fall asleep during the test. Finally, the drowsy state is defined as the one for which subjects fall asleep at least one time during the test.

The PERCLOS measure computed using a moving window of 15 minutes for subjects that were awake is shown in table 4.4. Assuming the PERCLOS measurements have a normal distribution, the confidence interval (CI) with a confidence level of 95% is found to be $0.0319 \pm 0.0021$ with a standard deviation of 0.0066, i.e, the eyes were closed 3% of the time during the observation period. These results differ from those obtained for subjects in semi-drowsy state (see table 4.5) and for which the CI is within $0.0880 \pm 0.0084$ with a standard deviation of 0.0570. This means that the PERCLOS measure for semi-drowsy drivers is significantly larger than that for awake drivers. The semi-drowsy state can be interpreted as the transition from the awake state to the drowsy state considering that the average PERCLOS measure for drowsy drivers was 0.18, a remarkable increase with respect to not fully drowsy drivers. PERCLOS results for subjects in a drowsy state are shown in table 4.6. The CI for the PERCLOS measure for drowsy drivers is $0.1802 \pm 0.0146$ with a standard deviation of 0.0690.

| Subject | PERCLOS |
|---|---|
| 1 | 0.028057 |
| 2 | 0.032882 |
| 3 | 0.029127 |
| 4 | 0.044358 |
| 5 | 0.025534 |
| 6 | 0.031695 |
| Mean | 0.03194 |
| Std.Dev | 0.0066209 |

TABLE 4.4. *Average PERCLOS for the group of subjects in awake state.*

| Subject | PERCLOS |
|---|---|
| 7 | 0.10885 |
| 8 | 0.065954 |
| 9 | 0.090533 |
| 10 | 0.108824 |
| 11 | 0.066297 |
| Mean | 0.0880916 |
| Std.Dev | 0.0213996 |

TABLE 4.5. *Average PERCLOS for the group of subjects in semi-drowsy state.*

Fig. 4.1 shows the average PERCLOS values and the CI of the different states of alert confirming a clear difference between the states. Fig. 2 shows the normal distribution of the three states and providing the necessary information to establish the classification rules for the alert states of the driver. For example, the means and covariances can be assumed to specify the probability distributions for each class, and used to select the class for which the measurement has the highest probability to belong to. This sort of minimum distance estimator can be implemented in terms of the following rules:

- $0.000 <$PERCLOS$\leq 0.048$ indicates the driver is fully awake (awake state).
- $0.048 <$PERCLOS$\leq 0.125$ indicates the driver is in a semi-drowsy state.
- $0.125 <$PERCLOS$\leq 1.000$ indicates the driver is in a drowsy state.

| Subject | PERCLOS |
|---|---|
| 12 | 0.189135 |
| 13 | 0.176758 |
| 14 | 0.209923 |
| 15 | 0.097456 |
| 16 | 0.226241 |
| 17 | 0.181416 |
| Mean | 0.18015483 |
| Std.Dev | 0.0445936 |

TABLE 4.6. *Average PERCLOS for the group of subjects in drowsy state.*

In order to avoid incorrect triggering of alarms to wake up the driver when laughing, frowning, scratching or making other gestures not related to a drowsy state, the search for eye closures lasting more than two seconds to determine if the driver has fallen asleep are done only if the PERCLOS value is greater than $0.0881$, which is the mean PERCLOS value for drivers in semi-drowsy state. Because the semi-drowsy state is a transition between the awake and the drowsy states, is reasonable to assume that subjects in the semi-drowsy state closer to the drowsy state will have higher probability of falling asleep than the ones closer to the awake state.

Tests made to measure the reaction time of the subjects show that in most cases the reaction time increases with drowsiness, as shown in table 4.7. However, it is difficult

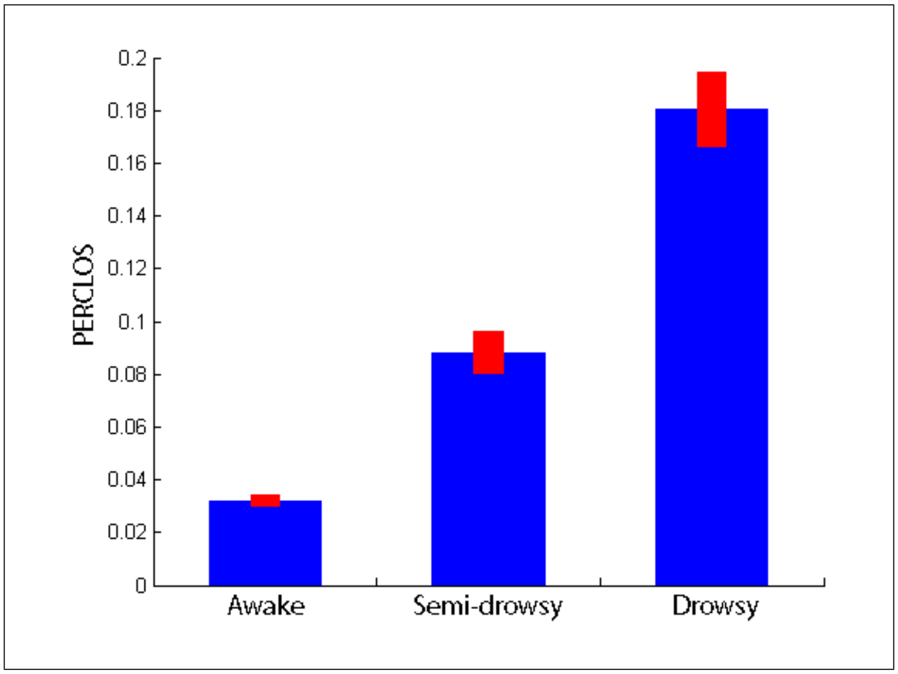


FIGURE 4.2. *Mean and confidence interval of the tested subjects for the three states of alert.*

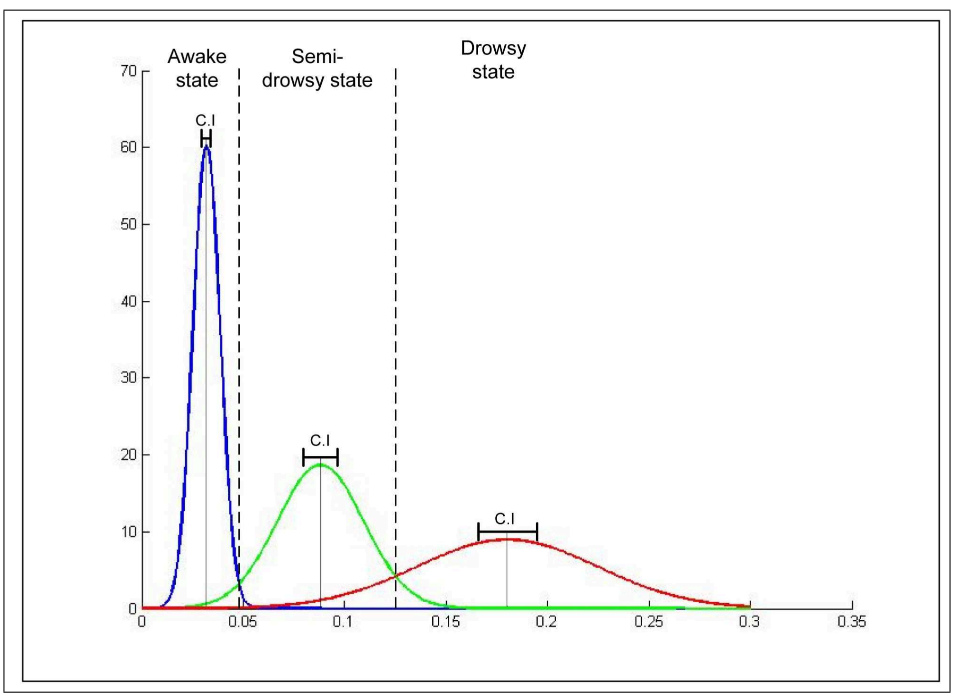


FIGURE 4.3. *Normal distributions of the different states of alert calculated with the tested subjects.*

to establish the state of the subject because the reaction time can vary significantly from subject to subject, even if when are equally awake or fatigued. For this reason the reaction time was not used as a criteria to determine the state of the driver, despite weak correlations between the mean reaction time of the different groups and their level of drowsiness.

| State | Reaction time[ms] |
|---|---|
| Awake | 199 |
| Semi-drowsy | 204.6 |
| Drowsy | 262.5 |

TABLE 4.7. *Mean reaction time of the group of subjects in the different states.*

# 5. CONCLUSION AND FUTURE RESEARCH

A robust and computationally effective approach for estimating the drivers state of drowsiness was presented. The approach relies on the motion analysis of salient points, which are selected using the Shi-Tomasi approach and tracked with the Lucas-Kanade tracker. Tracking the facial points and estimating the head pose allows keeping track of the driver's pupils at all times without requiring a computationally expensive face detection process on every frame. The Viola-Jones approach applied to face detection is used only to obtain an initial estimate of the face location, and every certain frames, when a large movement of the head is detected. The approach performs under day or night illumination conditions because it uses a standard camera together with a circular array of infrared leds integrated to the camera and an 850 nm infrared passband filter added to block illumination variations during the day due to external sources. The approach delivers the estimated position of the eyes $99.41\%$ of the time on average, regardless of whether a direct detection was possible or not, thus proving to be quite effective and robust to occlusions due to driver actions, such as yawning, scratching, laughing, or other facial gestures. It is also to be noted that blinking is detected a $98.37\%$ of the time on average with low false alarm rates (below $1\%$). Higher detection rates are possible, but with higher false alarm rates due to changes in the intensity of the pupil whenever the driver changes the direction view.

The approach performs in real-time comparing favorably with respect to other approaches reported in the literature. It can be easily implemented with currently available low-cost security cameras that include IR leds.

It was shown that PERCLOS measurements can be used to identify the driver's main states of alert: awakeness, partial drowsiness and full drowsiness. Basic rules to determine the driver's drowsiness take into account the mean PERCLOS value and the standard deviation to define the threshold for being in certain state. A remarkable finding was that the group of subjects in the awake state presents a mean PERCLOS value of $0.0319 \pm 0.0022$ (CI $95\%$) with a standard deviation of $0.0066$, while subjects in the drowsy state have a mean PERCLOS of $0.1801 \pm 0.0146$ (CI $95\%$) with a standard deviation of $0.0690$, thus exhibiting a significant difference between the two states that would allow a driver assistance system to warn the driver about having reached dangerous levels of fatigue, which could lead to an imminent accident, and that stopping the car to get some rest should be a priority.

The proposed approach may be improved by reducing cumulative error arising from head tracking or improving the direct detection of face and eyes, for example, through some improvement of the Viola-Jones algorithm. Therefore, future research is concerned with incorporating the RANSAC iterative method to discard outliers from the tracked points. Also, given that a 5 DOF kinematic model of the driver's motion in space is already being used, an extended Kalman filter could be added to minimize the variance in the head pose estimation from a frame to the next. The Viola-Jones algorithm used in our approach was trained using images taken with a regular camera. In order to improve the detection rates, IR images of faces will be used to re-train the Viola-Jones algorithm for eye and head detection.

Though some test with drivers wearing eyeglasses and sunglasses were done, no conclusive results were obtained. More tests must be made in the future to analyze how the system works under these conditions and how can it be improved to work robustly in both cases, with or without eye glasses.

# REFERENCES


Administration, N. H. T. S. (1998). *Drowsy driving and automobile crashes, report no. dot hs 808 707* (pp. 30 éd.; N. E. P. on Driver Fatigue & Sleepiness, Eds.). National Center on Sleep Disorders, Research, National Heart, Lung, and Blood Institute, and National Highway Traffic, Safety Administration, Washington, D.C.: NCSDR/NHTSA.

Chang, B., Lim, J., Kim, H., & Seo, B. (2007, Sept.). A study of classification of the level of sleepiness for the drowsy driving prevention. *The Society of Instruments and Control Engineers Conference*, 3084-3089.

Crow, F. (1984). Summed-area tables for texture mapping. *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, 207 - 212.

Dodgson, N. (2004, Jan.). Variation and extrema of human interpupillary distance. *Stereoscopic Displays and Virtual Reality Systems XI*, *2*, 36-46.

Dong, W., Qu, P., & Han, J. (2008, July). Driver fatigue detection based on fuzzy fusion. *Control and Decision Conference, 2008. CCDC 2008. Chinese*, 2640-2643.

Dong, W., & Wu, X. (2005, May). Driver fatigue detection based on the distance of eyelid. *VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on*, 365-368.

D'Orazio, T., Leo, M., Spagnolo, P., & Guaragnella, C. (2004, Oct.). A neural system for eye detection in a driver vigilance application. *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, 320-325.

Fan, X., Yin, B., & Sun, Y. (2008, April). Nonintrusive driver fatigue detection. *Networking, Sensing and Control, 2008. ICNSC 2008. IEEE International Conference on*, 905-910.

Fan, X., Yin, B.-C., & Sun, Y.-F. (2007, Aug.). Yawning detection for monitoring driver fatigue. *Machine Learning and Cybernetics, 2007 International Conference on*, *2*, 664-668.

Flores, M., Armingol, J., & Escalera, A. (2008, June). Real-time drowsiness detection system for an intelligent vehicle. *Intelligent Vehicles Symposium, 2008 IEEE*, 637-642.

Grace, R., Byrne, V., Bierman, D., Legrand, D., J.M. andGricourt, Davis, B., Staszewski, J., et al. (1998). A drowsy driver detection system for heavy vehicles. *Digital Avionics Systems Conference*, *2*, 136/1-136/8.

Gu, H., Ji, Q., & Zhu, Z. (2002). Active facial tracking for fatigue detection. *Applications of Computer Vision, 2002. (WACV 2002). Proceedings. Sixth IEEE Workshop on*, 137-142.

Hong, T., Qin, H., & Sun, Q. (2007, 30 2007-June 1). An improved real time eye state identification system in driver drowsiness detection. *Control and Automation, 2007. ICCA 2007. IEEE International Conference on*, 1449-1453.

Horng, W.-B., Chen, C.-Y., Chang, Y., & Fan, C.-H. (2004, March). Driver fatigue detection based on eye tracking and dynamk, template matching. In (Vol. 1, p. 7-12).

Ito, T., Mita, S., Kozuka, K., Nakano, T., & Yamamoto, S. (2002). Driver blink measurement by the motion picture processing and its application to drowsiness detection. *Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on*, 168-173.

Liang, S. F., Lin, C. T., Wu, R. C., Chen, Y. C., Huang, T. Y., & Jung, T. P. (2005, Sept.). Monitoring drivers alertness based on the driving performance estimation and the eeg power spectrum analysis. *Engineering in Medicine and Biology 27th Annual Conference*.

Lu, H., Zhang, W., & Yang, D. (2007, 28 2007-Dec. 1). Eye detection based on rectangle features and pixel-pattern-based texture features. *Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on*, 746-749.

Lucas, B., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of Imaging Understanding Workshop*, 121-130.

Park, I., Ahn, J.-H., & Byun, H. (2006, 0-0). Efficient measurement of eye blinking under various illumination conditions for drowsiness detection systems. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, *1*, 383-386.

Pedan, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., et al. (2004). *World report on road traffic injury prevention*. WHO Press.

Pickering, C., Burnham, K., & Richardson, M. (2007, Oct.). A review of automotive human machine interface technologies and techniques to reduce driver distraction. *System Safety, 2007 2nd Institution of Engineering and Technology International Conference on*, 223-228.

Qin, H., Gao, Y., & Gan, H. (2007, Dec.). Precise eye location in driver fatigue state surveillance system. *Vehicular Electronics and Safety, 2007. ICVES. IEEE International Conference on*, 1-6.

Rongben, W., Lie, G., Bingliang, T., & Lisheng, J. (2004, Oct.). Monitoring mouth movement for driver fatigue or distraction with one camera. *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, 314-319.

Royal, D. (2002). *National survey of distracted and drowsy driving attitudes and behavior: 2002, report no. dot hs 809 566*. NHTSA.

Shi, J., & Tomasi, C. (1994, June). Good features to track. In *Proc. of the* $9^{th}$ *ieee conf. on computer vision and pattern recognition* (pp. 593–600).

Sigari, M. H. (2009, Feb.). Driver hypo-vigilance detection based on eyelid behavior. *Advances in Pattern Recognition, 2009. ICAPR '09. Seventh International Conference on*, 426-429.

Singh, S., & Papanikolopoulos, N. (1999). Monitoring driver fatigue using facial analysis techniques. *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, 314-318.

Suzuki, M., Yamamoto, N., Yamamoto, O., Nakano, T., & Yamamoto, S. (2006, Oct.). Measurement of driver's consciousness by image processing -a method for presuming driver's drowsiness by eye-blinks coping with individual differences -. *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, *4*, 2891-2896.

Tabrizi, P. R., & Zoroofi, R. A. (2008, Nov.). Open/closed eye analysis for drowsiness detection. *Image Processing Theory, Tools and Applications, 2008. IPTA 2008. First Workshops on*, 1-7.

Viola, P., & Jones, M. (2001, June). Rapid object detection using a boosted cascade of simple features. In *Proc. of the 16$^{th}$ ieee conf. on computer vision and pattern recognition* (Vol. I, pp. 511–518).

Wang, Q., Yang, W., Wang, H., Guo, Z., & Yang, J. (2006, June). Eye location in face images for driver fatigue monitoring. *ITS Telecommunications Proceedings, 2006 6th International Conference on*, 322-325.

Wierwille, W., Ellsworth, L., Wreggit, S., Fairbanks, R., & Kirn, C. (1994). Research on vehicle based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness. *National Highway Traffic Safety Administration Final Report*.

Wright, N., & McGown, A. (2001). Vigilance on the civil flight deck: incidence of sleepiness and sleep during long-haul flights and associated changes in physiological parameters. *Ergonomics*, *44*(1), 82-106.

Xu, C., Zheng, Y., & Wang, Z. (2008, June). Efficient eye states detection in real-time for drowsy driving monitoring system. *Information and Automation, 2008. ICIA 2008. International Conference on*, 170-174.

Zhang, Z., & Zhang, J. shu. (2006, 0-0). Driver fatigue detection based intelligent vehicle control. *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, *2*, 1262-1265.