A Drowsiness and Point of Attention Monitoring System for Driver Vigilance

Jorge Batista
ISR-Institute of Systems and Robotics
Department of Electrical Engineering and Computers
FCTUC - University of Coimbra, Coimbra, Portugal

***Abstract*— This paper presents a framework that combines a robust facial features location with an elliptical face modelling to measure driver's vigilance level. The proposed solution deals with the computation of eyelid movement parameters and head (face) point of attention.**

The most important facial feature points are automatically detected using a statistically anthropometric face model. After observing the structural symmetry of the human face and performing some anthropometric measurements, the system is able to build a model that can be used in isolating the most important facial feature areas: mouth, eyes and eyebrows. Combination of different image processing techniques are applied within the selected regions for detecting the most important facial feature points.

A model based approach is used to estimate the 3D orientation of the human face. The shape of the face is modelled as an ellipse assuming that the human face aspect ratio (ratio of the major to minor axes of the 3D face ellipse) is known. The elliptical fitting of the face at the image level is constrained by the location of the eyes which considerable increase the performance of the system.

The system is fully automatic and classifies rotation in all-view direction, detects eye blinking and eye closure and recovers the principal facial features points over a wide range of human head rotations. Experimental results using real images sequences demonstrates the accuracy and robustness of the proposed solution.

I. Introduction

The ever-increasing number of traffic accidents in the EC due to the diminished driver's vigilance level has became a serious problem to society. Driver fatigue resulting from sleep deprivation or sleep disorders is an important factor in the increasing number of accidents on today's roads. Statistics shows that a leading cause for fatal or injury-causing traffic accidents is due to drivers with a diminished vigilance level. Automatically detecting the visual attention level of drivers early enough to warm them about their lack of adequate visual attention due to fatigue may save a significant amount of lives and personal suffering. Therefore, it is important to explore the use of innovative technologies for solving the driver visual attention monitoring problem.

Many efforts have been reported in the literature on developing non-intrusive real-time image-based fatigue monitoring systems [2], [7], [9], [10], [12], [13], [14]. Measuring fatigue in the workplace is a complex process. There are four kinds of measures that are typically used in measuring fatigue: physiological, behavioral, subjective self-report and performance measures [1]. An important physiological measure that has been studied to detect fatigue has been eye-movements. Several eye-movements were used to measure fatigue like blink rate, blink duration, long closure rate, blink amplitude, saccade rate and peak saccade velocity.

*This Work was funded by FCT Project POSI/EEA-SRI/61150/2004

The present solution focuses on rotation of the head and eye blinking, two important cues for determining driver visual attention, to gather statistics about the driver's visual attention level. Since the motion of a person's head pose and gaze direction are deeply related to his/her intention and attention, the ability to detect the presence of visual attention, and/or determine what a person is looking at by estimating the direction of eye gaze and face orientation is useful to measure driver's level of attention. One of our main purpose is to recover and track the three degrees of rotational freedom of a human head, without any prior knowledge of the exact shape of the head and face being observed. This automatic orientation estimation algorithm is combined with a robust facial features detection and tracking to provide additional constraints to estimate the sequence of head orientations observed.

II. Anthropometric Face Model

Face Shape is dynamic, due to the many degrees of articulative freedom of the human head, and the deformations of the face and its parts induced by muscular action. Face shape variability is also highly limited by both genetic and biological constraints, and is characterized by a high degree of (approximate) symmetry and (approximate) invariants of face length scales and ratios. Anthropometry is a biological science that deals with the measurements of the human body and its different parts [20]. It is concerned with tabulation and modelling of the distributions of these scales and ratios, and can serve as a useful source of shape and parts location constraints for analyzing image sequences of human faces.

After performing anthropometric measurement on several frontal face images taken from different human's subjects, an anthropometric model of the human face is build that can be used to locate the most important facial feature areas from face images [19]. The facemarks points that have been measured to build the face anthropometric model are represented

TABLE I

PROPORTION RATIOS OBTAINED FROM THE ANTHROPOMETRIC MEASUREMENTS D_i.

Ratios	*Description*	*Value*
$D_{4,5}/D_1$	Proportion of the eye-eyebrow distance to the mouth width	≈ 0.35
D_2/D_1	Proportion of the nose tip to mouth center distance to the mouth width	≈ 0.65
D_3/D_1	Proportion of the distance between the midpoint of the eyes center and the mouth center to the mouth width	≈ 1.40
D_6/D_1	Proportion of the inter-eyes distance to the mouth width	≈ 1.20

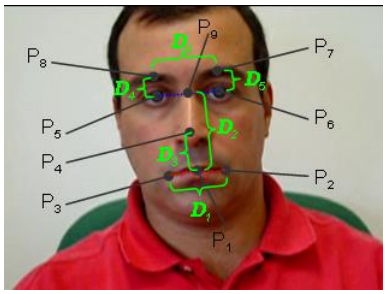

Fig. 1. Anthropometric face model used for facial feature area localization. *Facemarks* (P_i) and anthropometric measurements (D_i) of our anthropometric face model are displayed.

in fig. 1. Some statistics of proportion were obtained from these points, and the mouth statistics serves as the principal parameter for measuring the center location and size of the other facial feature regions. Table I show the proportion of distances ($\frac{D_i}{D_1}$) taking the mouth endpoints distance as the principal parameter of measurement. The mouth statistics is used, instead of inter-eyes statistics, mainly because eyes-centers (pupils) can not be detected with eyes shut.

III. IMAGE FACE DETECTION

Though people differ in color and length of hair it is reasonable to assume that the amount of skin that can be seen and the position of the skin pixels within the frame is a relatively invariant cue for a person's face detection in a static image. Many researchers have exploited the relative uniqueness of skin color to detect and track faces [4], [17]. To automatically select the face skin pixels on the image, a color skin histogram is defined taking samples from different scenarios [23]. A hand-select face skin region is performed on a sequence of learning frames to compute a (normalized) skin colour histogram model in RGB-space. We then compute the probability that every pixel in the face images is drawn from the predefined skin histogram model. Using the skin histogram, and taking the approach proposed on [23], each pixel in each face image is drawn from a specific RGB bin and so is assigned the relevant weight which can be interpreted as a probability that the pixel cames from the skin model. Figure 2 show the results of the skin color detection on a human face image. Using the Birchfield [4] solution, that combines intensity gradients and color histograms, the face projection on the image plane is modelled as an ellipse. Fig. 2 show the filtered skin

Fig. 2. The image face skin color detection with the face detected area and the fitted ellipse.

detected area and the ellipse fitted to that area. This ellipse is the starting point for our automatic facial features detection algorithm. The ellipse center and the ellipse's axis will be used as the new image face coordinate system.

IV. FACIAL FEATURE POINTS DETECTION

Since the proposed anthropometric face model was obtained for frontal face images, a starting frontal head pose is required to ensure a robust detection of the facial feature points. To robustly detect the facial features over a sequence of human face images taken from different head poses, a kalman filter feature tracking is combined with the proposed anthropometric face model features detection.

A. Identification of facial feature areas

The rotation angle (θ) between the frame coordinate system and the ellipse coordinate system (see fig. 2) encodes (at this stage) the roll rotation of the face, and to fit the anthropometric face model with the image face the whole image is rotated by this amount.

Since the mouth width D_1 serves as the principal parameter for measuring the center locations of the other facial feature regions, the implementation of our automatic facial feature point detection begins with the detection of the mouth area. Once the mouth area is correctly detected, points P_2 and P_3 can be accurately detected and distance D_1 used to locate points P_5, P_6, P_7 and P_8 by calculating the distances D_2, D_4, D_5 and D_6 using the proportionality ratios proposed by the anthropometric face model. Rectangular areas for confining the facial feature regions are then approximated using the mouth width as the measurement criteria.

B. Mouth features detection

Once the face is aligned to fit with the anthropometric model, the mouth detection starts by searching over the bottom area of the ellipse the region not segmented as skin that is closer to the major ellipse axis. The end points of this area and the maximum thickness value are used to accurately define the lips corner points and the inter-lips line.

Taking the coordinates of the end points (P_l, P_r) of the detected area as the initial lips corner points, the location of the correct lips corners is obtained by using the following approach:

1) Covert from color to gray scale and apply a contrast stretching to increase contrast;
2) Let P_l and P_r represent the coordinates of the initial lips corner points, and let T_{ck} represent the maximum thickness of the detected area;
3) For each column extending beyond both lips corners, consider a vertical line $\Im_c$ (of height T_{ck} pixels and centered on the previous detected lip

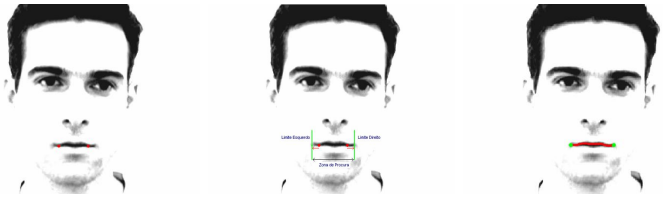

Fig. 3. Lips corner detection. Left to right: Grey level image after contrast stretching with the initial lip corners superimposed; The searching area; The correctly located lip corners and the extracted inter-lips line.

corner) and find the darkest pixel on this vertical line [7]. The darkest pixel will generally be a pixel in the gap between the lips.

4) To determine where the lip corners are the system obtains

$$f(x,y) = \frac{1}{D(x,y)} + \frac{1}{I(x,y)} \qquad \& \qquad D(x,y) < 3 \tag{1}$$

where $D(x,y)$ is the distance of a pixel (x,y) from the closest corner of the mouth, and $I(x,y)$ is the gray level intensity at (x,y). This will give a pixel that is close to the previous lip corner, and that is not too bright. The function maximum is the new lip corner.

5) This searching process is stopped when the gradient along the vertical line is below a certain threshold.

$$|\nabla \Im_c| < 0.1$$

Figure 3 shows the result of the proposed algorithm for lip corner location and inter-lips line detection.

C. Eyebrows features detection

The detection of the eyebrow feature points is accomplished prior to the eye pupil and eye corner detection for two basic reasons: i) to take into account the eyes shut situations and ii) because the correct detection of the eye's corners are difficult to achieve for large yaw head rotation angles due to occlusions.

Since the anthropometric proportion rations fail for non-frontal face images, to deal with large yaw and pitch head rotations additional constraints must be adopted. The following strategy is used to robustly detect both eyebrows:

1) Using the face model estimate the location and size of both eyebrow feature regions. The dimension of the feature region is related with the the anthropometric measure D_1, being defined as : $Width = 1.25 \times D_1$, $Height = 0.8 \times D_1$.
2) Perform the eyebrow detection within each of the feature regions using
   a) Covert from color to gray scale and apply a contrast stretching to increase contrast;
   b) Detect horizontal features using a vertical gradient mask (ex. Sobel) and threshold the outcome;
   c) Perform some noise filtering;
   d) Select the lowest extreme points of the segmented area as the eyebrow corner points;
3) To accomplish the correct detection of the end-points, in special for large yaw head rotations, adjust the the size of the feature region (10% steps) and repeat from step 2 [1];
4) Stop the process when the end-points location remain unchanged.

Figure 4 shows the defined eyebrow regions for two yaw head gaze rotations with the eyebrow segmentation result superimposed.

D. Eye features detection

The eye features detection is the most challenging task due to the variability of shapes. The eye region is composed of an upper eyelid with eyelash, lower eyelid, pupil, bright sclera and the skin region that surrounds the eye. Because both

[1]The black-box shown in fig. 4 represents the adjusted eyebrow region.

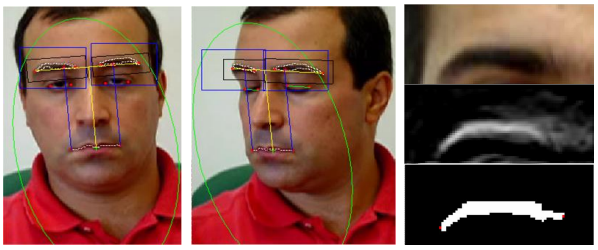

Fig. 4. Eyebrows detection. Defined eyebrow regions for different head rotations with the segmented eyebrow superimposed; right column: The eyebrow segmentation and eyebrow corner detection.

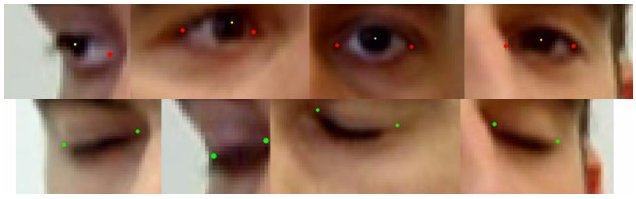

Fig. 5. The eyes features variability

pupil and sclera change their shape with various possible appearances of eyes, especially when the eye is closed or partially closed, robust detection of pupil center and eye corners is not an easy task. Most of the approaches found on the literature model the eyelid and detects the eye features (pupil center and eye corners) mainly for frontal face images [25], [26], [27]. Fig. 5 shows the variability of eye shapes under different head poses.

Our purpose is to be able to accurately detect the most important features of the eye (pupil center and eyes corners) for the large set of different appearances of an eye that can occur within our working scenario.

The corners of each eyebrow are used as the starting point for an accurate selection of the eye region. Taking the anthropometric face model, the location of the eye region is placed immediately above the eyebrow region. To deal with non-frontal face images, the size of each eye region is constrained by the size of the correspondent eyebrow. Representing $D_{eyebrow}$ the distance between both eyebrow corners, the width (M) and height (N) of each eye region is defined as

$$M = D_{eyebrow} + 0.4 \times \max(D_1, D_{eyebrow}) \tag{2}$$
$$N = 0.7 \times \max(D_1, D_{eyebrow}).$$

Eye features detection is accomplished in two steps. At the first step the presence of the iris is checked on each eye region and if the eye is labelled as an open eye, then a second processing step is applied to find the pupil center and to fit an ellipse to the iris shape.

1) Eyes Open vs. Eyes Shut: The detection of eye open *vs.* eye shut is accomplished by computing the variance projection function [24] on each eye feature region. The variance projection function (VPF) is defined as

$$\sigma_H^2(y) = \frac{1}{M} \sum_{i=1..M} [I(x_i, y) - H(y)]^2 \tag{3}$$

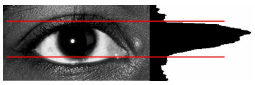

Fig. 6. The Variance Projection Function used to bound the iris area.

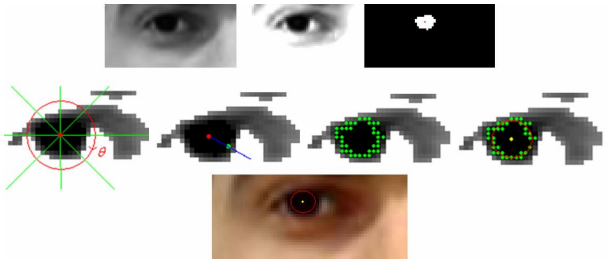

Fig. 7. Pupil center detection and iris elliptical shape model.

where $H(y)$ is the mean intensity value for the row y and $I(x,y)$ is the intensity of pixel (x,y). This variance projection function is applied in both directions and it is used to bound the area of the iris (fig. 6). To increase the robustness of the iris checking process, the detected iris is cross-checked based on size and proportion constraints.

2) Pupil center detection and iris shape modelling: Independently of the color of an iris, the pupil will always be its darkest region. The use of this cue to locate the pupil center is constrained by the level of detail of the eye region and by the color or gray level quantization of the image. For our purpose, the precise location of the pupil center is not a major issue, and the pupil center was considered to be coincident with the iris center.

The following approach was used to detect the pupil center and to model the iris shape.

1) Covert from color to gray scale and apply a contrast stretching to increase contrast;
2) Threshold the bounded iris area and consider the center of mass of the segmented region as the iris center;
3) Obtain the iris contour points by measuring the gradient maximum values along the radial lines starting at the iris center;
4) fit an ellipse to the detected contour points.

Fig. 7 shows the results obtained with the proposed approach.

3) Eyes corners Detection: The detection of corners of the eye is a challenging task in special for large yaw head rotations. In these cases, the correct location of the eye corners is extremely difficult mainly due to occlusions. The solution the we propose try to overcame this problem, detecting the eye corners either in the presence of eyes open or eyes shut.

a) Eyes-Shut:: To detect the corners of a closed eye we will take advantage of the dark region that is created with the union of both eyelashes. In these cases, the following algorithm is used:

1) Covert from color to gray scale and apply a contrast stretching to increase contrast;
2) Apply a vertical gradient mask to enhance the horizontal edges and threshold;
3) Obtain the skeleton of the segmented region followed by a pruning operation;
4) Select the end-points of the skeleton as the eye corners.

Figure 8 presents the evolution of the algorithm and some results.

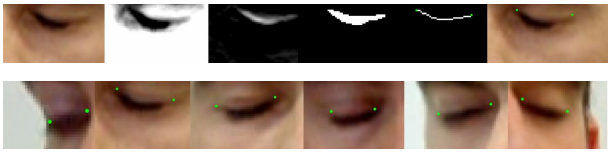

Fig. 8. Eye-Shut corners detection. Top row: Evolution of the proposed algorithm; Bottom row: Corners detected on several Eyes-Shut images.

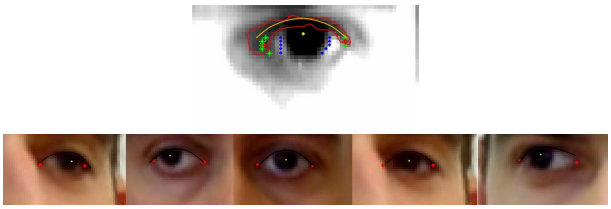

Fig. 9. Eye-Open corners detection. Top row: Evolution of the proposed algorithm; Bottom row: Corners detected on several Eyes-open images.

b) Eyes-Open:: For Open-eyes, we define the corner as the farthest transition point between the sclera (brightest pixels) and the skin (darkest pixels). This cue is easily visible for frontal face images, but lacks visibility when the eyes are gazed on the sizes and when the face image is not frontal.

To overcame this lack of visibility an additional cue has been added. The upper eyelid and eyelash are normally associated to the darkest region just above the iris, and the proposed solution will take advantage of this fact to extract the shape of the upper eyelid. The shape of the eyelid is used to constraint the location area for the corners. The following strategy is used to locate the eye corners:

1) Covert from color to gray scale and apply a contrast stretching to increase contrast;
2) Taking the previously segmented iris, obtain the VPF vales on the horizontal vicinity of the iris; These values will encode the gray level variability that exist between the dark regions of the eyelids and the bright area of the sclera.
3) Obtain an estimate for the location of the eye corners by thresholding the VPF values;
4) Threshold the image in order to enhance the iris and upper eyelid areas, and remove the pixels that belong to the previously segmented area of the iris.
5) Obtain the skeleton of the remaining region and fit a polynomial function to the skeleton;
6) The end-points of the skeleton polynomial function are used to locate the eye corners.
7) Combine the information supplied by both approaches.

Figure 9 shows the detection results of the proposed algorithm.

E. Experimental results

The performance of the proposed solutions for facial feature location was evaluated using five video sequences of a human face gazing at different 3D points and blinking his eyes. The algorithm was tested with different human subjects and under different image scales. Table II presents the detection accuracy of the automatic facial feature point detector described on the paper and figure 10 shows the facial feature detection on different human image faces extracted from the image sequence database.

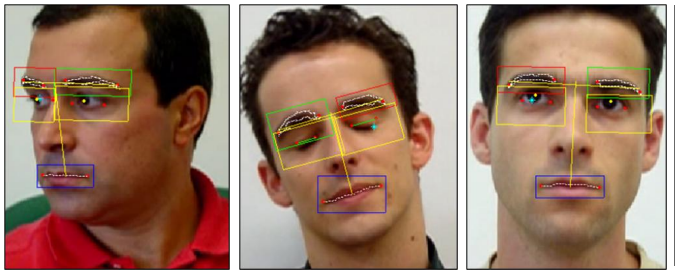

Fig. 10. Facial feature detection for different subject and with different face gaze orientation. The feature regions obtained using the anthropometric face model and the detected features are superimposed on the images.

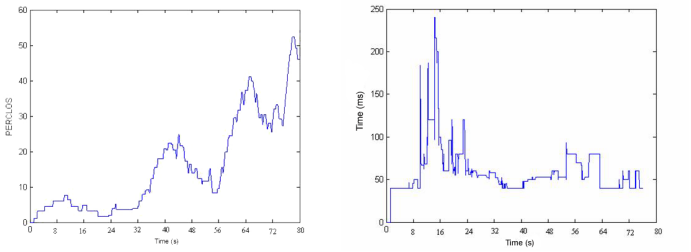

Fig. 11. PERCLOS (left) and AECS (right) measurements over a period of 80 seconds.

TABLE II

DETECTION ACCURACY (IN PERCENT) OF THE AUTOMATIC FACIAL FEATURE POINT DETECTOR

		Facial Feature Points Detection Rate (%)		
Seq. no.	# frames	% Mouth	% *eyebrows* left-right	% *eyes corners* left-right
1	360	99%	100%-99%	96.6%-98.3%
2	360	100%	98.2%-98%	97.7%-96.6%
3	360	100%	100%-96.6%	98.0%-96.6%
4	360	97%	99.4%-98.7%	95.5%-88.3%
5	346	99%	98.2%-100%	96.5%-98.3%

	Eyes-Shut Detection Rate		
Seq. no.	# Eyes-Shut	False Positives left-right	% left-right
1	95	3-1	96.8%-99.0%
2	10	1-0	99%-100%
3	90	4-4	95.5%-95.5%
4	47	13-3	72.3%-93.6%
5	105	3-1	97.1%-99.0%

	Eyes-Open Detection Rate		
Seq. no.	# Eyes-Open	False Negatives left-right	% left-right
1	265	1-1	99.6%-99.6%
2	350	1-3	99.7%-99.1%
3	270	2-1	99.2%-99.6%
4	313	1-7	99.6%-97.7%
5	241	2-5	99.1%-97.9%

V. DRIVER VISUAL ATTENTION STATISTICS

Of the drowsiness-detection measures, the measure referred to as PERCLOS [1] was found to be the most reliable and valid determination of a drivers alertness level. PERCLOS is the percentage of eyelid closure of the pupil over time and reflects slow eyelid closures (droops) rather than blinks. To measure eyelid closure of the pupil, the size of the pupil was taken as the average size of both pupils and the rate of closure is defined as $rate_{closure} = 1-(pupilsize)/max(pupilsize)$, defining a closed eye if $rate_{closure} \geq 0.8$.

AECS is the average eye closure speed [10], which means the amount of time needed to fully close the eyes and to fully open the eyes. An individual eye closure speed is defined as the time period during which the $0.2 \leq rate_{clossure} \leq 0.8$. Figure 11 shows the PERCLOS and AECS for a period of 80 seconds.

VI. HUMAN HEAD ORIENTATION

The presented approach models the shape of the human face with an ellipse, since human faces can be accurately modelled with an ellipse and is less sensitive to facial expression changes. To recover the 3D face pose from a single image, it is assumed that the ratio of the major and minor axes of the 3D face ellipse is know. This ratio is obtained through the anthropometric face statistics. Our purpose is to recover the three angles of rotation: yaw (around vertical axis), pitch (around horizontal axis) and roll (around the optical axis).

A. Image face ellipse detection and tracking

In order to correctly detect the face ellipse, some constraints must be considered, in special size, location and orientation. The distance between the detected pupils and their location are used to constrain the size and location of the image face ellipse. The orientation of the line that passes through both pupils is directly related to the 3D face roll rotation. For roll free face poses this line remains horizontal, which means that it is invariant to the yaw and pitch rotations. Under this constraints, the roll angle (γ) is defined by $\gamma = atan[(y_{p_l} - y_{p_r})/(x_{p_l} - x_{p_r})]$, where $P_l = (x_{p_l}, y_{p_l})$ and $P_r = (x_{p_r}, y_{p_r})$ are the image location of the left and right detected pupil, respectively.

For frontal orientation, a weak perspective projection can be assumed and the face symmetry for the location of the eyes within the 3D face ellipse hold for the image face ellipse. This means that the major axis of the face ellipse is normal to the line connecting the two eyes and pass through the center of the line. In fact, these constraints doesn't hold for non-frontal orientation and the orientation of the major line is not normal to the connecting line. Although, the solution adopted kept the constrain that the major axis of the ellipse pass through the center of the line, considering the existence of an angle α between the major axis and the normal to the line that connect the two eyes.

Assuming the existence of an ellipse coordinate frame located at the middle point of the eyes connecting line, with the X and Y axis aligned with the minor and major axes of the ellipse, respectively, the image face ellipse is characterized by a 4−tuple $e = (m_i, n_i, d, \alpha)$, where m_i and n_i are the lengths of the major and minor semi-axis of the ellipse, respectively, d is the distance to the image ellipse center and α is the rotation angle.

Taking the approach proposed by Birchfield [4], the image face ellipse can be detected as the one that minimizes the normalized sum of the gradient magnitude projected along the directions orthogonal to the ellipse around the perimeter of the ellipse and the color histogram intersection of the face's

interior. This can be formulated has $\varepsilon_g(e) = \frac{1}{N}\sum_{i=1}^{N}|n(i)\cdot g(i)|$ where $n(i)$ is the unit vector normal to the ellipse at pixel i, $g(i)$ is the pixel intensity gradient and $(\cdot)$ denotes dot product and $\varepsilon_c(e) = \frac{\sum_{i=1}^{N} min(I_e(i),M(i))}{\sum_{i=1}^{N} I_s(i)}$ where $I_e(i)$ and $M(i)$ are the numbers of pixels in the ith bin of the histograms, and N is the number of bins.

The best face ellipse is $\chi = arg\ max_{e\in E}(\bar{\varepsilon}_g(e) + \bar{\varepsilon}_c(e))$ where the search space E is the set of possible ellipses produced by varying the 4-tuple parameters of the ellipse and $\bar{\varepsilon}_g$ and $\bar{\varepsilon}_c$ are normalized values. The 4-tuple parameters of the ellipse are filtered via a kalman filter and the *a posterior* estimated tuple is used to define an initial estimate for best face ellipse search. Figure 12 show several results of the image face ellipse fitting.

B. Face orientation

Consider an object coordinate frame attached to the 3D face ellipse, with its origin located on the center of the ellipse and its X and Y axes aligned with the major and minor axes of the ellipse. The Z axis is located normal to the 3D ellipse plane. The camera coordinate frame is located at the camera optical center with the X_c and Y_c aligned with the image directions with the Z_c along the optical axis. Since the 3D face ellipse is located on the plane $Z = 0$, the projection equation that characterizes the relationship between an image face ellipse point $p_i = (x,y,1)^T$ and the corresponding 3D face ellipse point $P_i = (X,Y,1)^T$ is given by $p_i = \beta K[R|t]P_i$ where K represents the camera intrinsic parameters matrix, $M = [R|t] = [r_1\ r_2|t]$ is the extrinsic parameters matrix and $\beta = \lambda/f$ is an unknown scalar.

Representing

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} a & c/2 & d/2 \\ c/2 & b & e/2 \\ d/2 & e/2 & f \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \quad (4)$$

the matricial generic formula of an ellipse, the 3D face ellipse and the image face ellipse can be defined, respectively, as

$$\begin{bmatrix} X & Y & 1 \end{bmatrix} \mathbf{Q} \begin{bmatrix} X & Y & 1 \end{bmatrix}^T = 0 \quad (5)$$

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \mathbf{A} \begin{bmatrix} x & y & 1 \end{bmatrix}^T = 0. \quad (6)$$

Substituting $p_i = \beta KMP_i$ to Eq. 6 lead to

$$\begin{bmatrix} X & Y & 1 \end{bmatrix} \beta M^T K^T AKM \begin{bmatrix} X & Y & 1 \end{bmatrix}^T = 0. \quad (7)$$

Denoting $B = K^T AK$, the 3D ellipse matrix Q yields $Q = \beta M^T BM$.

Let the length of the major and minor axis of the 3D face ellipse be m and n, respectively, and since the object frame is located on the center of the ellipse, the ellipse matrix Q is parameterized as

$$Q = \begin{bmatrix} 1/m^2 & 0 & 0 \\ 0 & 1/n^2 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (8)$$

resulting the equation

$$\begin{bmatrix} 1/m^2 & 0 & 0 \\ 0 & 1/n^2 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \beta \begin{bmatrix} r_1^T Br_1 & r_1^T Br_2 & r_1^T Bt \\ r_2^T Br_1 & r_2^T Br_2 & r_2^T Bt \\ t^T Br_1 & t^T Br_2 & t^T Bt \end{bmatrix} \quad (9)$$

Due to the symmetry of the matrix, there are only six equations (constraints) for a total of nine unknowns.

Since the roll angle was already obtained, the face orientation can be defined just by the yaw and pitch rotation. Assuming a null translation vector, the rotation matrix obtained from the yaw and pitch rotation is

$$\begin{aligned} R &= R_\sigma R_\upsilon = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} = \quad (10) \\ &= \begin{bmatrix} cos(\sigma) & sin(\sigma)sin(\upsilon) & -sin(\sigma)cos(\upsilon) \\ 0 & cos(\upsilon) & sin(\upsilon) \\ sin(\sigma) & -cos(\sigma)sin(\upsilon) & cos(\sigma)cos(\upsilon) \end{bmatrix}. \end{aligned}$$

Assuming that the ratio between the major and minor axis of the 3D face ellipse is know by anthropometric face analysis [20], and letting $c = m^2/n^2$ represent this ratio, the 2×2 sub-matrix yields

$$\beta \begin{bmatrix} r_1^T Br_1 & r_1^T Br_2 \\ r_2^T Br_1 & r_2^T Br_2 \end{bmatrix} = \begin{bmatrix} 1/m^2 & 0 \\ 0 & 1/n^2 \end{bmatrix} \quad (11)$$

resulting the following constraint equations

$$r_1^T Br_2 = 0 \quad (12)$$

$$\frac{\beta r_1^T Br_1}{1/m^2} = \frac{\beta r_2^T Br_2}{1/2^2} \Leftrightarrow r_1^T Br_1 = \frac{n^2}{m^2} r_2^T Br_2 \Leftrightarrow r_2^T Br_2 - cr_1^T Br_1 = 0. \quad (13)$$

Using these two equations it is possible to solve for the pitch and yaw iteratively. The initial estimates of 0^o for both angles has been used for the first frame of the sequence.

Given r_1 and r_2 the translation T can be calculated up to a scale factor by using

$$\beta \begin{bmatrix} r_1^T BT \\ r_2^T BT \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (14)$$

Let $T = (t_x, t_y, t_z)$, t_x/t_z and t_x/t_z can be solved by using the above equation analytically. This approach was tested with several real images with good results. Although, the accuracy obtained with this approach is highly dependent on the image face ellipse obtained. A robust and accurate detection of the eyes pupil is fundamental for the overall performance of this approach. Figure 12 shows the results obtained with the face orientation estimation approach.

C. Monitoring driver point of attention

Besides the drowsiness, visual attention is another aspect of driver vigilance. A common way to characterize such level of inattention is duration and driver's point of attention. Driver point of attention is obtained using the face orientation solution described on the previous subsection, measuring the rate of visual inattention on a fixed time interval (60 s) and the largest period of visual inattention. The proposed solution is able to robustly measure yaw head rotation over $[-30°..+30°]$ interval and pitch head rotation over

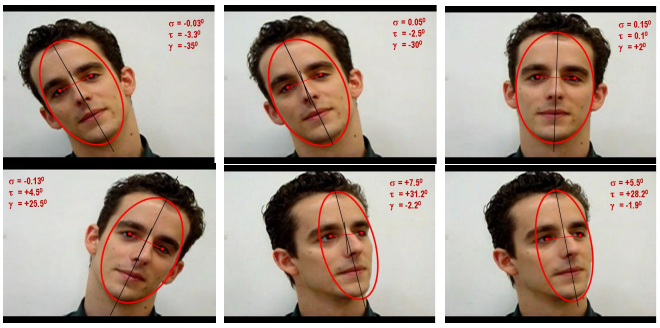

Fig. 12. The elliptical face model and 3D face gaze estimation. Results concerning the 3D face orientation estimation are superimposed on the images (τ: yaw; σ: pitch; γ: roll).

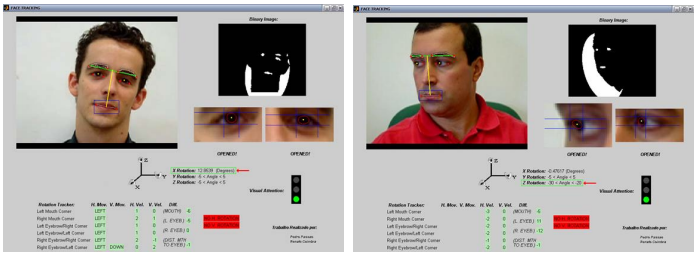

Fig. 13. The Graphic Interface Application used for driver visual attention Monitoring. Drowsiness level and point of attention monitoring is displayed by the semaphore shown on the GI.

$[-20°..+20°]$ interval. The head gaze interval is divided into several gaze regions (R_i) according to the following mapping

	Yaw Angle	Pitch Angle
R_{-3}	$-30°..-20°$	N.D.
R_{-2}	$-20°..-10°$	$-20°..-10°$
R_{-1}	$-10°..0°$	$-10°..0°$
R_0	$-5°..+5°$	$-5°..+5°$
R_1	$0°..+10°$	$0°..+10°$
R_2	$+10°..+20°$	$+10°..+20°$
R_3	$+20°..+30°$	N.D.

Figure 13 shows the point of attention region detection using the proposed head gaze orientation.

VII. Conclusions

A framework that combines a robust facial features location with an elliptical face modelling to determining the gaze of faces in image sequences was presented. The detection of facial feature point uses an anthropometric face model to locate the feature regions in the face. The statistical parameters of the mouth was used to model a set of face proportions that helped in the localization of facial feature regions, introducing a considerable reduction in computational time. A efficient and simple solution for pupil detection was presented that were used to take some drowsiness measures in real-time. The proposed solution is independent of the scale of the face image and performs well even with yaw and pitch rotation of the head. To determine the gaze of the face an elliptical face modelling was used taking the eye's pupil locations to constraint the shape, size and location of the ellipse. A robust and accurate detection of the eyes pupil is fundamental for the overall performance of this approach. This solution has proven to work relatively well, although the iterative approach that was used to obtain the 3D face gaze parameters.

References

[1] P. Sherry et al., Fatigue Countermeasures in the Railroad Industry: Past and Current Developments, *Published by Association of American Railroads*, 2000.

[2] A. Yilmaz, et al., Automatic feature detection and pose recovery for faces. *ACCV2002*, Melbourne, Australia, 2002.

[3] Y. Bar-Shalom et al., Tracking and Data Association. *Academic Press, Inc*, New-York 1988.

[4] S. Birchfield, Elliptical head tracking using intensity gradients and color histograms, *IEEE CVPR*, 1998.

[5] A.T. Horprasert et al., Computing 3D head orientation from a monocular image sequence *SPIE, 25th AIPR workshop: Emerging Applications of Computer Vision* 2962, 1996.

[6] A. Nikolaidis et al., Facial feature extraction and determination of pose, *http://www.citeseer.nj.nec.com/cs*.

[7] P. Smith et al., Determining Driver Visual Attention with one camera, *IEEE Trans. on Intelligent Transportation Systems, vol.4, n.4*, Dezember, 2003.

[8] A. Fitzgibbon et al., Direct least square fitting ellipse, *IEEE Trans. PAMI, vol. 21*, May 1999.

[9] Qiang Ji et al., Real-time eye, gaze and face pose tracking for monitoring driver vigilance, *Real-Time Imaging, 8*, 2002.

[10] Qiang Ji et al., 3D pose estimation and tracking from a monocular camera, *Image and Vision Computing*, 2002.

[11] J.G. Wang et al. Study on Eye Gaze Estimation, *IEEE Trans. on Systems, Man, and CyberneticsPART B: Cybernetics, vol. 32, no. 3*, JUNE 2002.

[12] M. Kaneda et al., Development of a drowsiness warning system, *11th Int. Conf. on Enhanced Safety of Vehicles*, Munich, 1994.

[13] R. Grace, A drowsy driver detection system for heavy vehicles, *Conf. on Ocular Measures of Driver Alertness*, 1999.

[14] D. Cleveland, Unobstusive eyelid closure and visual of regard measurement system, *Conf. on Ocular Measures of Driver Alertness*, 1999.

[15] M. Mallis et al., Ocular measurement as an index of fatigue and as the basis for alertness management: experiment on performance-based validation of technologies, 1999.

[16] S. Phimoltares, et al., Locating Essential Facial Features using Neural Visual Model, *Int. Conf. on Machine Learning and Cybernatics*, 2002.

[17] Z. Xin, X. Yanjun, D. Limin, Locating Facial Features with Color Information. *IEEE Int. Conf. on Signal Processing*, vol.2, 1998.

[18] A.H. Gee et al., Determining the gaze of faces in images, *Image and Vision Computing*, 12 (10), 1994.

[19] A. Sohail, P. Bhattacharya, Localization of Facial Feature Regions using Anthropmetric Face Model, *Int. Conf. on Multidisciplinary Information Sciences and Technologies*, 2006.

[20] L. Farkas, Anthropometry of the Head and Face, *Raven Press*, New-York, 1994.

[21] Y. Matsumoto, A. Zelinsky. An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement, *Int. Conf. on Automatic Face and Gesture Recognition*, 2000.

[22] S.Y. Ho et al., An analytic solution for the pose determination of human faces from a monocular image, *In Pattern Recognition Letters, 19*, 1998.

[23] M. Swain, D. Ballard, Color Indexing, *Int. Journal of Computer Vision*, 7(1), 1991.

[24] G. Feng, P.Yuen, Variance Projection Function and its application to eye detection for human face recognition, *Pattern Recognition Letters*, vol.19(9), 1998.

[25] V. Vezhnevets, A. Degtiareva, Robust and Accurate Eye Contour Extraction. *Graphicon-2003*, Moscow, Russia, 2003.

[26] P. Kuo, J. Hannah, An Improved Eye Feature Extraction Algorithm based on Deformable Templates, *ICIP*, 2005.

[27] Y. Wu, H. Liu, and H. Zha, A new method of human eyelids detection based on deformable templates, *IEEE Int Conf. on Systems, Man and Cybernetics (SMC'04)*, Hague, Netherlands, 2004.