Adaptive Haar-like Classifier for Eye Status Detection
Under Non-ideal Lighting Conditions

Mahdi Rezaei
The University of Auckland
Auckland, New Zealand
m.rezaei@auckland.ac.nz

Reinhard Klette
The University of Auckland
Computer Science Department
Auckland, New Zealand

ABSTRACT
The paper introduces a novel methodology to enhance the accuracy, performance and effectiveness of Haar-like classifiers, specially for very dark and complicated lighting conditions. Performing a dynamic statistical intensity analysis on input sequences, the technique provides a very fast and robust eye status detection via a low resolution VGA camera, without application of any infra-red illumination or image enhancement. A field test of driver monitoring in real-world conditions has been considered under challenging lighting conditions including ‘very bright’ lightings conditions at daytime as well as ‘very dark’ or ‘artificial’ lightings at nights. Adaptive Haar Classifier adjusts the detection parameters according to a dynamic level-based intensity measurements in given regions of interest. Experimental result and performance evaluation on various datasets show higher detection rate comparing to the standard Viola-Jones classifier.

Categories and Subject Descriptors
I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Object Recognition; I.5.4 [Pattern Recognition]: Applications; I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search—Heuristic methods

Keywords
Adaptive Haar-like classifier, Eye detection, Face and eye monitoring, Challenging lighting conditions

1. INTRODUCTION
During past decade many researchers such as Viola and Jones [15], Froba and Ernst [19], Jianxin et al. [20], and Hsu et al. [7] developed well recognized model-based and learning based face detection techniques. Apart from general improvements, there are still many rooms for face detection and more specifically eye detection enhancement in non-ideal environments.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IVCNZ’12 November 26 - 28 2012, Dunedin, New Zealand
Copyright 2012 ACM 978-1-4503-1473-2/12/11 ...$15.00.

Driver drowsiness monitoring is a good instance of eye status analysis in a challenging environment. Difficulties for a proper detection are mainly due to non uniformity of the light source, asymmetric shades on driver’s face and eye region, and rapid change in light intensity while real-world driving due to e.g. shades of building, bridges, trees, or driving into or out of a dark tunnel.

Reviewing recent research in face and eye detection shows many robust and precise algorithms such as works done by [16, 18, 21, 22]; However, majority of them have been done for fairly frontal face positions, ideal lighting condition, or in indoor environments. There are many complicated and sensitive applications such as driver eye status monitoring (open, blinking, or closed), which are far from being solved accurately.

Among related works in addressed area of application, Castrillon et al. [5] has applied performance analysis on publicly available Haar-classifiers presenting their pros and cons for face and eye detection using Haar-like techniques.

Liu and Liu [23] proposed an integrated eye tracker which can robustly detect and track eyes under variable rotations and face angle in real time. They applied multiple filtering techniques such as Zernike Moments and Support Vector Machine; performance test has been applied on a indoor database with the average accuracy of 94%. In another work, Majumdar [10] introduced a hybrid approach for eye detection based on Haar-like classification in HSV colour mode; however the validation test has been applied on a very limited number of frontal faces. Cehn et al. [24] present an eye detection algorithm using histogram back-projection, elliptical separability filter, and vector quantization. Experimental results only show normal lighting condition. Zhua and Ji [17] introduce another eye detection and tracking for variable lighting environments, however they need to use IR illumination to overcome lighting challenges.

In brief, all above discussed research suffer from a lack of verification and performance analysis on a wide-range video, image dataset, and lighting variations. In this paper, we pursue three goals. 1) to improve noisy measurements of a Haar-classifier into a more stable solution for detecting and localizing features in the image plane. 2) to reduce total computational cost by minimizing the search region for “eye status” detection 3) to overcome eye-detection failures due to complicated and difficult lighting condition by introducing a novel technique, an adaptive Haar classifier, for the first time.

The structure of this paper is as follows: In Section 2 we outline the main idea. Section 3 discusses in details of adaptive classifier. Section 4 discusses on validation test and experimental results to compare adaptive classifier with
Section 5 concludes as well as suggestions for future works.

2. BRIEF METHODOLOGY- MAIN IDEA

Figure 1 shows a brief hierarchical representation of our idea. The methodology mainly involves to resolve weakness of Haar-like feature in extreme low light or high light scenes, while keeping the robustness of such classifiers in normal situation.

Utilizing Haar-feature based classifiers, two search options has been considered. In Option 1 (left path), the initial region of interest (ROI) is quantified as 100% that means the entire area of the input image should be analyzed to find the given object (e.g. a face). For multiple feature detections like face, open eyes, or closed eyes detection three different classifiers need to search the whole input, each one individually with the computational cost of 3 × 100%.

This is a common way that standard Haar-classifiers apply to search and hit the objects; however, according to our assessment on more than 1000 recorded frames from different driving conditions and comparing with ground truth information, we determined the average face size of 185 × 185 pixel which covers only 11% of the VGA input images. In order to increase our search speed and also reducing the chance of false eye detections we define Option 2 (right path) with the idea of partial search instead of full search in image plane. Furthermore, based on eye localization information of FERET and YALE databases, we define a border for eye movements in upper part of the detected face as in range of 0.55-0.75 (region A) as shown in Fig. 2.

Assuming more complicated scenarios for head tilt and rotation, a movement range of 0.35-0.95 is also estimated as region B. Having an already detected face, we could run eye classifier on regions A and B only (ROI2) which includes 5.2% of the input image. In case of face detection failures, so we could not have any prior estimation for eye location, and consequently the search cost for eye detection will again increase to 100% (ROI3). Following to previous information obtained for open eye location, then closed-eye classifier will be applied in region C (ROI4) which only covers 3% of the input image frame.

In brief, having an already detected face in the first step will cause a maximum search cost of 108.2% for a complete eye status analysis (ROI1+ROI2+ROI4), while failure in face detection causes a searching cost of equal to 203% (ROI1+ROI2+ROI3+ROI4). Aiming a fast and successful driver monitoring (including drowsiness detection, head pose estimation, and distraction detection) with a minimized search (as per Path 2), requires proper face localization from the first stage, as well as accurate and robust eye state detectors in all lighting situations.

Non-adaptive eye detection solutions, such as [2, 13] and those work mentioned in section Introduction report up to 95% successful detection under “normal” driving conditions. However, none of them are successful for irregular and challenging conditions like sample situations shown in Fig. 3. That points to a need for eye detectors to be adaptive and robust in extremely challenging lighting conditions.

Feedback cycles in Fig. 1 demonstrate classifier adaptation and tuning phase, that tries to overcome the weakness of standard Haar-like based detectors introduced by Viola and Jones [14, 15]. Improvement is based on a hybrid lighting intensity measurement which is detailed in next section.

3. ADAPTIVE CLASSIFICATION

This section discusses on adaptation module detailed in three sub-sections. In order to create a V-J based detector [14], we combine three techniques:

1. Application of a wide and comprehensive range of Haar-like masks or features that are in analogy to base function of the Haar-wavelet (samples are given in Fig. 4).
imposed Haar-feature on eye region.

Figure 4: Sample Haar-like features, and two superimposed Haar-feature on eye region.

2. Usage a boosting algorithm to train and select appropriate et of features for each weak classifier, and

3. Creation of a strong cascaded classifier by merging the trained week classifiers.

Selection process of an object is based on value distributions in dark or light regions of Haar-features that models expected intensity distributions. For example, the superimposed features in Fig. 4, right, shows the idea that in any face there are always darker regions of eyes compared to a brighter bridge of nose.

3.1 Weakness of Viola-Jones Method in Challenging Lighting Conditions

We tried five well recognized and publicly available [12] Haar classifiers developed by Castrillon, Lenhart, Yu, and Hameed, in our nominated application: “driver monitoring”. Although they are quite robust for non-challenging and normal lighting scenes, we realized that due to frequent shadows and artificial lighting in day and night a Haar-like classifier could easily fail. The situation becomes even more complicated when a part of the driver’s face is lighter than the other part (due to lights that falling in through a side-window). These make eye status detection extremely difficult. We also compiled and trained our own classifier based on a large dataset of +12,000 positive images from YALE, FERET, BioID, PICS, and FTD.

Continuing our previous work [11], and applying an AdaBoost machine learning technique [6], we got the best results for our trained Haar-like classifier with the following given parameters:

- Positive images’ size: 21 × 21 pixel
- Number of weak classifier (stages): 15
- Minimum expected hit rate for each stage: 99.8%
- Maximum acceptable false alarm at each stage: 40%
- Trimming threshold: 0.95

The next step, after training and creation of the classifier, we needed to utilizing the classifier in the real-world with parameters which are normally similar to the trained parameters. The main parameters are:

- Initial search window size (SWS) that should normally be the same as the scale size of positive images (e.g. 21 × 21 as above)
- Scale factor (SF) to increase the SWS in each subsequent search iteration (e.g. 1.2 which means 20% increase in window size for each search iteration)
- Minimum expected number of detected neighbours or MNN which is needed to confirm an object, when there are multiple object candidates in a small region (e.g. 3)

In general, a smaller SF means a more detailed search in each iteration, but it also causes more computational cost. Regarding MNN, if we decrease it then the detection rate will increase; this will also increase the rate of false detections as a disadvantage. Larger values for MNN lead to more strictness to confirm a candidate for face or eye, thus a reduced detection rate. Figure 5 shows potential issues related to the MNN parameter in the Viola-Jones methodology for non-ideal lighting such as driving. The figure shows 10 initial eye candidates before trimming those by the parameter “minimum number of neighbours”. Detections are distributed in 5 regions, each region shows 1 to 4 overlapping candidates. In order to minimize this problem, it is common to assign a trade-off value for the MNN parameter to gain the best possible results. Figure 5, top right, shows one missed detection with an MNN of 3 or 4, and Fig. 5, bottom right, shows one false detection with an MNN of 2. MNN equal to 1 will cause 3 false detections, and any MNN greater than 4 will lead to no detection at all; So, as can be seen there is no successful MNN value for this example.

We concluded that although we generally can define a trade-off values for SWS, SF, and MNN to get the optimum average detection rate for a ideal video sequences, however, a diversity of changes in light intensity over the target object can still significantly affect the performance of the given classifier in terms of TP and FP rates.

3.2 Hybrid Intensity Averaging

Aiming to cope with above mentioned issues, we arrived at the idea that Haar-classifier parameters could be adaptive, vary with time, and depending on temporal information and lighting changes while driving. Figures 6 and 7 illustrates the reasons that we can not measure lighting changes by a simple intensity averaging over the input frame: In a driving application, there can be strong white back-lights from back windshield, strong white pixel values around the driver’s cheek or forehead (due to light reflections) or dark shadows on a driver’s face. All these conditions may negatively affect the overall mean intensity measurement. Analysing various recorded sequences, we realized that the pixel intensity around eyes can change more independently than surround-

Figure 5: Left: Initial results of eye detection. Right: Finally detected eyes after trimming by two different MNN factors.

Figure 6: Mean intensity measurement for eye detection, by excluding very bright and very dark regions. Green: thresholding range 210-255. Blue: thresholding range 0-35.
3.3 Parameter Adaptation

I

Figure 7: Selected regions to sum up the mean intensity around eyes. Images Source: Yale database.

mports) as well as of very bright lights (i.e. green segments) consequently will reduce the impact of eye colour (i.e. blue segments) in the region surrounding an eye can guarantee robust eye detection. Following this consideration, we defined white rectangles around eyes (Fig. 6, right) which can not only provide a good approximation of both vertical and horizontal light intensities around the eyes, but they are also very marginally influenced by green or blue (very bright or very dark) regions.

Considering an already detected face, and expected eye regions A and C based on Fig. 2, we can geometrically define $C_r$, $F_r$, $C_l$, and $F_l$ as being the optimum regions in order to gain an accurate estimation of vertical or horizontal lighting intensity around the eyes (see Fig. 7). We also consider independent classifier parameters for the left and right half of the face, as each half of the face may receive different and non-uniform light exposures. Performing a further analytical step, Fig. 6, right, shows that few small green or blue segments (extreme dark or light intensities) have entered the regions of white rectangles, what can affect the actual mean intensity calculations in the $C$ or $F$ regions. Thus, in order to reduce the effect of this kind of noise into our measurements, we apply hybrid averaging by combining mean and mode ($\text{Mo}$) of pixel intensities as follows:

$$I_r(\alpha) = \frac{1}{2} \left[ \alpha \cdot \text{Mo}(C_r) + \frac{(1 - \alpha)}{m} \sum_{i=1}^{m} C_i \right]$$
$$+ \left[ \alpha \cdot \text{Mo}(F_r) + \frac{(1 - \alpha)}{n} \sum_{j=1}^{n} F_j \right]$$

(1)

where $I_r(\alpha)$ is the hybrid intensity value of the right eye region of the face, $m$ and $n$ are the total numbers of pixels in $C_r$ and $F_r$ regions; $C_r$ and $F_r$ are in $[0, 255]$ point to cheek and forehead light intensity, respectively.

An $\alpha$ value of 0.66 assumes a double importance of mode intensity measurement compared to mean intensity; this consequently will reduce the impact of eye colour (i.e. blue segments) as well as of very bright lights (i.e. green segments) for our adaptive intensity measurement. Similarly, we can calculate $I_l(\alpha)$ as hybrid intensity value of the left eye region.

### 3.3 Parameter Adaptation

The final step of detection phase is classifier parameter adjustment based on the measured $I_r$ values in previous step, to make our classifier adaptive for every single input frame. At this stage we need to find optimum parameters (SWS, SF, MNN) for all the intensity ranges between 0 to 255, which is a highly time consuming practical tasks. Instead, we defined optimum parameters for 10 selected intensities, followed by a data interpolation method to extend those parameters to all intensity ranges. Table 1 shows the optimum parameter values as for 10 data points obtained based on 20 recorded videos in different weather and driving conditions. These factors are adjusted for leading to the highest rate of true detections for each of 10 given intensity.

Since the parameter values in Table 1 show a non-linear behaviour over intensity changes, therefore we applied non-linear cubic interpolation and Lagrange interpolation [3] to extend adapted values for intensity ranged from 0-255.

We selected Lagrange interpolation than cubic interpolation results, because of less fluctuation around existing data points so a reduced error probability for determined points.

Assuming the interpolation function as polynomial, i.e. $f(x)$ of the type $f(x) = a_0 + a_1x + a_2x^2 + ... + a_kx^k$ and considering a series of 10 data points $(x_1, y_1), (x_2, y_2), ..., (x_{10}, y_{10})$, the

$$f(x) = a_0 + a_1x + a_2x^2 + ... + a_{10}x^{10}$$

Table 1: Optimum Haar-classifier parameters for 10 selected grey levels in terms of Search Window Size, Scale Factor, and Minimum Number of Neighbours. FS: detected Face Size.

<table>
<thead>
<tr>
<th>Light Intensity</th>
<th>SWS</th>
<th>SF</th>
<th>MNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>FS/5.0</td>
<td>1.10</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>FS/4.5</td>
<td>1.12</td>
<td>2</td>
</tr>
<tr>
<td>50</td>
<td>FS/3.5</td>
<td>1.30</td>
<td>9</td>
</tr>
<tr>
<td>75</td>
<td>FS/4.0</td>
<td>1.15</td>
<td>7</td>
</tr>
<tr>
<td>90</td>
<td>FS/4.0</td>
<td>1.30</td>
<td>10</td>
</tr>
<tr>
<td>120</td>
<td>FS/4.2</td>
<td>1.25</td>
<td>16</td>
</tr>
<tr>
<td>155</td>
<td>FS/5.0</td>
<td>1.35</td>
<td>15</td>
</tr>
<tr>
<td>190</td>
<td>FS/4.5</td>
<td>1.30</td>
<td>14</td>
</tr>
<tr>
<td>220</td>
<td>FS/4.6</td>
<td>1.25</td>
<td>9</td>
</tr>
<tr>
<td>255</td>
<td>FS/4.0</td>
<td>1.35</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 8: MNN values defined by Lagrange and cubic interpolations for a grey-level range of 0 to 255.

Figure 9: Fitted curves for optimum SF and SWS values via cubic or Lagrange interpolation.
Table 2: Performance analysis for Standard and Adaptive Classifier.

<table>
<thead>
<tr>
<th></th>
<th>Standard V-J Classifier</th>
<th>Adaptive Classifier</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Face</td>
<td>Open</td>
</tr>
<tr>
<td>Video 1</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td></td>
<td>97.5</td>
<td>0.01</td>
</tr>
<tr>
<td>Video 2</td>
<td>81.1</td>
<td>1.02</td>
</tr>
<tr>
<td>Yale DB</td>
<td>86.3</td>
<td>0.05</td>
</tr>
<tr>
<td>Closed Eye DB</td>
<td>92.2</td>
<td>0.06</td>
</tr>
</tbody>
</table>

coefficients $a_0$ to $a_k$ identified by Lagrange interpolation

$$P_n(x) = \sum_{i=0}^{n} L_i(x) f(x_i)$$

where the $L_i$'s are polynomials

$$L_i(x) = \prod_{j=1, j \neq i}^{n} \frac{x - x_j}{x_i - x_j}$$

of degree $n$ that form a basis of $P_n$. We also applied cubic interpolation (with a 3rd-order spline) for comparison. Figure 8 and 9 show the results of interpolation for MNN, SF, and SWS for $x$-coordinate (intensity level) in range of 0 to 255.

4. EXPERIMENTAL AND VALIDATION RESULTS

A grey-scale VGA camera at the distance of about 60cm to driver seat and filed of view of 30 degrees is used to take continuous recording of the driver seat area. The classifier re-trained by integrating the parameter values obtained after interpolation. In order for validation tests, 20 recorded video sequences from extremely light varying conditions while driving as well as two indoor and outdoor images database has been examined. Table 2 provides details of TP, FP detection rates performed on selected videos and image datasets with high diversity of lighting and illumination changes; Adaptive classifier shows considerable improvements comparing to a standard Haar-classifier. Figures 10 to 13 show the results of face and eye status detection before and after implementing of adaptive framework.

![Figure 10: Face and eye detection under sharp sun-strikes; standard classifier (top) vs. adaptive classifier (bottom).](image1)

5. CONCLUSIONS

The paper introduced a real-time and effective detection framework that enables possibility of multiple facial feature tracking under difficult lightings; clearly better than Viola-Jones based classifiers. In proposed method, while we try to minimize the search region, the adaptive module focuses on the given reduced region to adapts the SWS, SF, MNN parameters depending on a pixel by pixel intensity changes in eye-pair surrounding area. In overall we gained about two times faster processing as well as more accurate results only with a low resolution VGA camera, without application of IR light source, or any preprocessing and illumination normalization techniques. The solution could be recognized as a common amendment for any kind of Haar-like classifiers to outperform in challenging environments. The paper compared the results with standard Viola-Jones classifiers for a wide range of video and image databases. Incorporating pre-processing techniques and comparison with illumination invariant methods such as SIFT or LBP, could be suggested for future works.

6. REFERENCES


![Figure 11: Face detection with sunglasses and under light-varying conditions; standard classifier (top) vs. adaptive classifier (bottom).](image2)
Figure 12: Face and Eye Detection under sophisticated lightning condition; standard classifier (top) vs. adaptive classifier (bottom). Image source: Yale database.

Figure 13: Closed eye detections; S: standard classifier (top) vs. A: adaptive classifier (bottom).

Proc. Audio Video-Based Biometric Person Authentication, LNCS 2091 (2001) 90–95


