Off-Line Driver’s Eye Detection: Multi-Block Local Binary Pattern Histogram Vs. Gabor Wavelets

Djamel Eddine Benrachou, Salah Bensaoula
University Badji Mokhtar Annaba
Laboratory of Automatic and Signals- Annaba
Department of electronics
BP 12, 23000, Annaba, Algeria
Email: djamelben.univ@gmail.com

Brahim Boulebtateche, Mohamed-Mourad Lafifi
University Badji Mokhtar Annaba
Laboratory of Automatic and Signals- Annaba
Department of electronics
BP 12, 23000, Annaba, Algeria
Email: brahim.boulebtateche@univ-annaba.dz

Abstract—Eye detection is a complex issue applied in several applications, such as human gaze estimation, Human-Robot interaction and driver’s aid framework for automatic drowsiness detection. This paper presents an algorithm that detects the human eye in still gray-scale images. The proposed scheme is based on learning statistical appearance model, which implies; the extraction of features and their classification. A comprehensive comparison of two photometric feature descriptors is performed between Gabor wavelets and Multi-block Local Binary Pattern histogram (BHLBP) features. Facial images are normalized and the eye features are extracted using the precedent descriptors, then Support Vector Machine classifier (SVM) is used to distinguish eye from non-eye classes. The presented schemes are built and tested with different frames, collected from a real driving video sequences of the RobeSafe Driver Monitoring Video dataset (RS-DMV), the driving sequences are recorded under realistic scenarios and with different subjects, which are exposed to a real driving environment. Discriminative performance of the descriptors are reported.

Keywords—Eye detection, eLBPh, Gabor wavelets, Driver’s drowsiness.

1. INTRODUCTION AND RELATED WORKS

The detection of the eyes is an active research subject and a challenging topic, which has shown expanding interest these last decades and demonstratizes an efficient applicability in different areas, it is still widely studied and used in numerous engineering applications including: driver drowsiness state monitoring framework, gaze direction estimation for Human-Robot interaction applications. Detecting the ocular region serves a crucial role in face normalization and thus facilitates further localization of other facial features. Some facial expressions are clearly distinguished through the eye state, since proven that the driver’s fatigue has a strong correlation with the PERCLOS measure[1]. These characteristics are increasingly finding used in safety applications to detect situations such as sleepiness and lack of attention while driving [2], [3], [4] or using hazardous machinery.

In automotive security framework the driver’s somnolence state is an important issue, as there are few direct measures used to detect this phenomenon and most of them are related to observed outcome symptoms gathered from the person’s behavior while driving. Recently, González-Ortega et al [4], proposed a real-time vision-based eye state detection framework to monitor the driver’s alertness, it consists to locate the eyes to recognize their states. This approach is implemented with a consumer-grade computer and an Universal Serial Bus camera with passive illumination. Lately, Benrachou et al [5] proposed an eye localization framework that extends the conventional Local Binary Pattern (LBP) to pyramid transform domain for encoding the eye in 3—levels of the image pyramid with different LBP variants. For the decision issue on eye detection, two well established classifiers were used, Support Vector Machine (SVM) and Multi-layer Perceptron (MLP) for binary classification between eye/not-eye classes. In the same fashion, Benrachou et al [6] exposed an algorithm for On-line eye detection in real world conditions, their framework is based on uniform LBP histogram to describe the eye patches and Long-Short Terms Memory Recurrent Neural Network (LSTM-RNN) that estimates the eye position on the running video frames. Nanni et al [7] presented an eye classification module based on multi-resolution LTP and LPQ descriptors. Ying and Wang [8] proposed a revised projection algorithm, to determine the eye position called locally selective projection (LSP). Sun and Zheng [9] improve the LSP scheme, by proposing an algorithm called LBP + SVM mode, this technique is able to localize the eyes and recognize their state as well. In their framework, the projection-based algorithm is avoided, for the reason that is sensitive to the image low quality and the projection calculation tends to fail in the lower gray-scale regions, with highly predictable regions that contain the true position of eyes.

In fact, there exist commercial eye localization systems, which perform well under relatively controlled environmental conditions, but tend to fail when the variation appears from different factors such as, facial expressions, the partial occlusion, the pose variation and lighting conditions. Several eye localization techniques have been emerged, and are broadly categorized into three main types; Methods based on the measurement of the eye characteristics [10], others Exploiting structural information [11], [12] and those using Learning statistical appearance model [8], [9].

In this paper, the proposed algorithm is based on conventional learning statistical appearance model, three stages are involved; i) preprocessing (noise removal, illumination correction etc.) ii) the extraction of useful visual features of eyes and iii) their classification. The advantage of this type of method
compared to the aforementioned ones is, that provide richer and more reliable information about the eye for the subsequent classification. However, the employed feature descriptors must have some essential properties such as, identifiability of the interesting object, rotation and scale invariance, robust against environmental illumination conditions and lighting changes.

In agreement with these conditions, Local Binary Patterns have proven their effectiveness for texture description, they are naturally tolerant to changes in monotonous gray scale, lighting variations, simple to compute and discriminative. To extract useful visual features from photometric appearance of eyes, we adopt the spatially enhanced Local Binary Pattern histogram (i.e BHLBP) to represent the micro- and macro-structure of the eye image patches. The analysis of the discriminant patterns is done by resorting to advanced machine learning approaches: Support Vector machine (SVM).

Thus, our eye appearance-based framework behaves robustly even under challenging imaging conditions (pose, illumination changes, facial expressions). In our experiments, discriminative performance of BHLBP are compared with those of Gabor wavelets descriptors.

The remainder of this paper is organized as follows, Section II devotes to the details of the proposed approach, in Section III the multi-block LBP histogram descriptor modeling is presented in detail and the conventional Gabor wavelets are briefly introduced. Section V presents the implementation of our experiments and finally Section VI contains conclusion, remarks and future work.

II. METHODOLOGY

In this section, the performance of an extended LBP based approach for eye detection are analyzed under gray-scale facial images. Experiments are conducted on the RobeSafe Driver Monitoring Video (RS-DMV) database [13], to show the effectiveness of these LBP extension (i.e BHLBP). Discriminative performance of BHLBP are compared against GW for eye description.

The global architecture of the proposed eye detection system is given in Fig. 1. For a given test image, the ocular region is described and segregated from other facial features (i.e nose, mouth, eyebrow etc). The search for the eye location is carried out by applying the associated detector over test image $I$, with size $(N_x \times N_y)$ pixels for the outdoor videos and $(300 \times 103)$ pixels for the indoor video frames. The image $I$ is normalized in the range of $(0, 1)$ and fully scanned with the precedent image descriptor over a shifting a sub-window with size $(27 \times 18)$ pixels through the entire image map. The algorithm needs a training image set, which is manually collected for this purpose. Either positive (eye) or negative (non-eye) used for training and testing are illustrated in Fig. 2 (a) and (b), respectively. The eye detection results are illustrated in Fig. 5.

III. FEATURE EXTRACTION

Features extraction step consists in transforming the input raw data into meaningful information. Thus, we obtain a reduction of the decision space which may accelerate the processing time.
called LBP code.

\[ LBPP_{R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \]  

(4)

where \( s(z) \) is the signum function

\[ s(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases} \]  

(5)

The equation 5 represents the signs of the computed differences in 4. \( g_c \) is the gray value of the center pixel, \( g_p \) is the value of its neighbors, \( P \) is the total number of sampling points, which controls the quantization of the angular space and \( R \) is the radius of the neighborhood (radii). The sampling is performed commonly in the clockwise direction, around the central pixel according to a particular radius value, which determines the spatial resolution of the distributed sampling points. In practical tasks, LBP histogram (LBPH) is usually used, the binary pattern of each pixel in the input image is identified. Thereafter, the LBPH is computed over the whole LBP image, i.e.,

\[ H(k) = \sum_{m=1}^{M} \sum_{n=1}^{N} f(LBPP_{R}(m,n),k), \quad k \in [0,K] \]  

(6)

\[ f(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases} \]  

(7)

Where \( K \) is the maximal LBP pattern value. The ocular

region is considered a dynamic and non-rigid object with large variations, this clearly indicates that textures of the ocular region is not close to be uniform due to intrinsic and environmental variations that can manifest in this facial region, and preprocessing such as subregion division can mitigate these large variations for a given eye image to certain extent. Hence, this process indicates a sort of rationality to use Multi-block histogram LBP (BHLBP) as an eye feature descriptor.

In this paper, BHLBP is used to describe the ocular region, it is formed on the basis of an eye image with size \((27 \times 18)\) pixels, which is divided into 6 non-overlapped sub-blocks of size \(9 \times 9\) pixels; and then the \(59\)-label \( LBPP_{8,1}^{2} \) operator is adopted to extract the LBP features. Hence, the local histogram is computed over the sub-regions and the global spatial histogram of \(354\) bins (i.e. \(59\) bins \(\times 6\)) is obtained, by concatenating all sub-regional histograms. These parameter settings for eye representation were suggested in [6]. The utilized \( LBPP_{8,1}^{2} \) is an extension of the original LBP operator, which indicates a coding and histogram mapping scheme in the \(u2\)-uniform LBP codes of eight neighbors at a distance of one pixel to the origin. The LBP operator is considered uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa, when the bit pattern is considered circular, \(LBPP_{u2}^{2}\) is statistically stable and less sensitive to noise [16].

IV. DATA AND SETTINGS

The RobeSafe Driver Monitoring Video (RS-DMV) database is a publicly available dataset, which is constituted from a set of video sequences of different subjects while driving. It contains 10 video sequences, 7 recorded outside (Type A) as shown in Fig. 3, the videos have been captured in car cockpit moving around the university campus. The (type B) as shown in Fig. 4 sequences include 3 video records, captured in a realistic truck simulator cockpit, under low-light conditions, that approached night time scenarios as shown in Fig. 4. In both cases, the cameras were installed over the dashboard in front of the driver. The persons were fully awake, talked frequently and were asked to look regularly to rear-view mirrors and operate the car sound system. Sequences includes some vision challenging scenarios; partial occlusions, illumination changes and other elements that are helpful for driver’s behavior monitoring systems using computer vision. Frames are recorded in gray-scale, at 30 frames per second, and stored as RAW video. The sizes of the outdoor and the indoor video frames are \((960 \times 480)\) pixels and \((1390 \times 480)\) pixels, respectively.

V. EXPERIMENTAL RESULTS

Eye detection application, requires a certain ability to distinguish the ocular region through the human face, despite the inherent variations in the intrinsic eye properties and its external appearance changes.
TABLE I. STATISTICAL RESULTS ON ROBE SAFE DRIVER MONITORING VIDEO (RS-DMV) DATABASE

<table>
<thead>
<tr>
<th>Approach</th>
<th>TP (%)</th>
<th>FP (%)</th>
<th>TN (%)</th>
<th>FN (%)</th>
<th>Prec</th>
<th>Rec</th>
<th>F1-Score</th>
<th>Acc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHLBP+SVM(Linear)</td>
<td>4.7</td>
<td>1.1</td>
<td>94.1</td>
<td>0.1</td>
<td>0.8085</td>
<td>0.9743</td>
<td>0.8836</td>
<td>98.57</td>
</tr>
<tr>
<td>Gabor+SVM(Linear)</td>
<td>69.0</td>
<td>0.6</td>
<td>29.1</td>
<td>1.3</td>
<td>0.9915</td>
<td>0.9810</td>
<td>0.9862</td>
<td>98.08</td>
</tr>
<tr>
<td>BHLBP+SVM(Quadratic)</td>
<td>3.2</td>
<td>2.6</td>
<td>94.0</td>
<td>0.2</td>
<td>0.5531</td>
<td>0.9285</td>
<td>0.6932</td>
<td>97.17</td>
</tr>
<tr>
<td>Gabor+SVM(Quadratic)</td>
<td>68.0</td>
<td>0.9</td>
<td>26.7</td>
<td>3.7</td>
<td>0.9872</td>
<td>0.62</td>
<td>0.7616</td>
<td>95.42</td>
</tr>
</tbody>
</table>

In this section, we investigate the classification performance of the spatially enhanced LBP and Gabor wavelet descriptors, both are applied for eye detection problem. For this purpose a large image set is collected from (RS-DMV) database, including face images of subjects taken under unconstrained conditions: hard pose variation, facial expression, different illumination conditions and occlusions.

Searching for eye location, is carried out by applying the associated detector that scans a full image of size $N_x \times N_y$ pixels over multiple shifting windows with size $27 \times 18$ pixels. The algorithm needs a training image set, which is manually collected for this purpose. 2476 images are used, which are confounded between eye and non-eye image patches, the collected samples are randomly divided, with 67% used for training, and the remaining 33% images are used for testing. The discriminative performance of BHLBP is compared against Gabor wavelets features. In this work, BHLBP and Gabor wavelets are trained and validated with SVM classifiers, by using the open source SVM implementation Libsvm version 3.12 [17].

The size of the test frames are reduced to $(300 \times 163)$ pixels for the outdoor video frames and $(300 \times 103)$ pixels for the indoor scenario frames, in order to speed up the detection process. BHLBP is applied with a subregion division and histogram concatenating, this image decomposition strategy yields a global histogram with a length of 354 bins, please refer to Sec. III-B, this representation represents the micro- and macro-structure of the eye patch.

Nevertheless, Gabor wavelets are applied with 40 filters (8 orientations and 5 scales) on $(27 \times 18)$ pixels eye patches, the resulting feature vector has 19440-dimensional vector.

In our experiments, Gabor filter bank appears to be quite perspective and has several advantages such as invariance to homogeneous illumination changes, robust against small changes in head poses and partial occlusions, but it is relatively long comparing the that of BHLBP descriptor and may implies a computational complexity in the feature extraction step.

In the classification phase, two SVM’s kernels (i.e Linear and Quadratic kernels) are used, to estimate the separability potential of the collected image, and preserving the trade off between the implementation simplicity and detection effectiveness, in terms of detection accuracy (Acc).

The use of such classification method tends to suffer from surrounding environmental variations implies a generalization problem, which is theoretically discussed from Vapnik-Chervonenkis (VC) in [18]. In the real-world applications, generalization problem constitutes a huge bottleneck for Automatic face and facial feature detection.

To improve the proposed approach, training data is collected from the driver’s environmental space, which makes the trained classifier significant, and less exposed to false positives.

In practice this process can be applied by using a web-camera embedded in the driver cockpit, that records surrounding driver’s environment and enriches the dataset used for training.

The performance validation of the detector and the final results are presented, including the calculation of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). The final average accuracy (Acc) for each descriptor is presented in Table I.

Moreover, precision and recall are calculated. Thanks to these two metrics, we compute the F-Score interpreted as a harmonic mean of the precision and recall for further comparison of the obtained results. When the trained classifier is predicting a positive class, we mostly have a high recall and a low precision, in contrast if the trained model predicts a negative class, a high precision is obtained against a very low recall. Thus, model achieves the highest precision and recall simultaneously is often desirable.

The proposed eye detection method, was tested on some RS-DMV image set. Therein images stress the real-world conditions, it includes variations in view (resolution changes), pose variations, illumination and expressions changes. Visual inspection is used as an evaluation measure for demonstrating the robustness of our method against these variations. To be more rigorous in evaluating the results, examples of successful localization of eyes from these datasets are shown in Figure. 5.

![Eye detection results on some images of the RS-DMV database.](image)

From Table I, it is clearly shown that Gabor wavelets transform, are well applied for eye detection providing a good results, which means that this descriptor is effective for facial features detection, besides being efficient for face description. Although, LBP based descriptor has some attractive proprieties while being very simple to compute and less complex than Gabor wavelets, it is invariant to ambient environment changes, illumination and alignment. Moreover, it should be pointed out that both approaches detect well the
human eyes in different states (open, partially closed and closed).

VI. CONCLUSION AND FUTURE WORK

In this paper, we have exposed an algorithm for eye detection in video recordings from automotive cockpit. Two feature extraction methods have been assessed for ocular region description; Gabor wavelets representation and BHLBP. In terms of eye presence classification and from conducted experiments, it concludes that the BHLBP with linear SVM approach gets a slight better performance than linear SVM trained with Gabor features. Thus, BHLBP and SVM(linear) realized a best recognition results of 98.57%.

This allows us to conclude that proposed Off-line eye detection framework, achieved a good detection results and conducted experimentations validate the performance of BHLBP/SVM scheme, which is favorably applied in an On-line eye detection framework on an embedded platform such as Raspberry Pi. Moreover, LBP based approach with sub-region division and histogram concatenating strategy, enhances the discriminative power of the descriptor. However, more complex image decomposition strategies exist (radial devisions, pyramidal image decomposition . . . etc) that can be joint to the LBP descriptor and tested in the future, such as for example.

In the future this work could be extended to track the position of the eyes while detecting their states, within a video stream to monitor the driver’s vigilance while he is conducting a car. The output of this algorithm, could constitute an input to the Long Short Term Memory- Recurrent Neural Networks (LSTM-RNNs), this deep-neural architecture can handle the temporal information dependencies through time without significant loss. Our system runs in a standard computer and camera, that makes it easy the implementation of the vision-based automotive security system.

REFERENCES


