Extended Eye Landmarks Detection for Emerging Applications

Laura Florea, Corneliu Florea, and Constantin Vertan

Abstract In this chapter we focus on the eye landmarking and eye components identification in the framework of emerging psychology-related eye tracking applications. Traditional eye landmarking separates the identification of eye centers and of eye corners and margins, while here we discuss their joint use for face expression analysis in unconstrained environments and precise estimation of non-visual gaze directions, as suggested by the Eye Accessing Cues (EAC) of the Neuro-Linguistic Programming (NLP). Such a system involves a combination of low-level feature extraction, heuristic pre-processing and trained classifiers. The approach is extensively tested across several classical image databases and compared with state of the art traditional methods.

1 Introduction

The exceptional developments in computer vision from the last decade make the automatic analysis of human behavior a goal that seems achievable. An important part of this process is the automatic face and face elements identification and interpretation and a lot of research has been particulary dedicated to eye detection. Eye data and the details of eye movements have numerous applications in face detection, biometric identification, and particularly in human-computer interaction tasks.

One also used to say that eyes are the gate to the soul and witness for various internal cognitive or emotional processes; this observation opened lately a plethora of less-traditional areas of research involving eye detection and tracking. Among these new investigated directions, we note the detection of deception as part of hostile intention perception [2], the estimation of pain intensity via facial expression analysis [3, 4], interpersonal coordination of mother–infant [5], assistance in marketing [6], etc. Recently, literature also reported attempts to interpret more complex

L. Florea (✉) • C. Florea • C. Vertan
Image Processing and Analysis Laboratory (LAPI), University "Politehnica" of Bucharest, Bucharest, Romania
e-mail: laura.florea@upb.ro; corneliu.florea@upb.ro; constantin.vertan@upb.ro

M. Kawulok et al. (eds.), *Advances in Face Detection and Facial Image Analysis*,
DOI 10.1007/978-3-319-25958-1_15

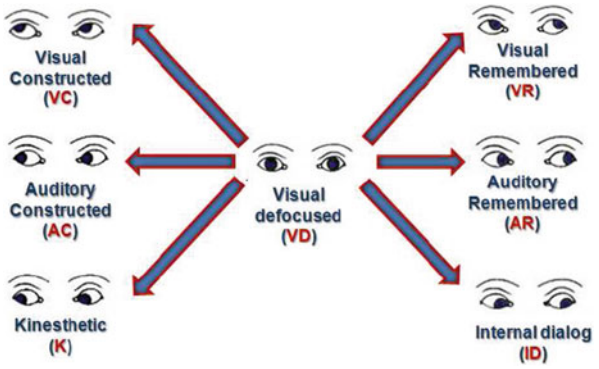

Fig. 1 The seven classes of Eye Accessing Cues [1]: When eyes are not used for visual tasks, their position can indicate how people are thinking (using visual, auditory or kinesthetic terms) and the mental activity they are doing (remembering, imagining, or having an internal dialogue). Image inspired from [1]

situations, such as dyadic social interactions for the diagnosis and treatment of developmental and behavioral disorders [7] and to experiment within new areas of psychology, as pointed in the recent review by Cohn and De La Torre [8]. The review of Friesen et al. [9] evaluates the social impact of gaze direction and concludes that many opportunities arise upon the understanding of the perceived direction of gaze.

Such an opportunity is offered by the Neuro-Linguistic Programming (NLP) theory, which presents unexplored opportunities for understanding the human patterns of thinking and behavior. One such model is the Eye-Accessing Cue (EAC) from the NLP theory that uses the positions of the iris inside the eye as an indicator of the internal thinking mechanisms of a person. The *direction of gaze*, under the NLP paradigm, can be used to determine the internal representational system employed by a person (see Fig. 1), who, when given a query, may think in visual, auditory or kinesthetic terms, and the mental activity of that person, of remembering, imagining, or having an internal dialogue.

In this chapter we direct the reader's attention to a system [10] that exploits an usual digital video camera (for instance a webcam—as its price makes it widely accessible) to infer eye features and landmarks positions in order to identify the direction of gaze for recognition of the Eye Accessing Cues, which are a potential mean to unravel one's background psychological process.

2 Eye Based Communications in Emergent Applications

The origins of the idea that involuntary eye movements point to inner mechanisms goes back until the nineteenth century [11]. In a review about the perception of interlocutor's gaze, Friesen et al. [9] concluded that "people's eyes convey a wealth of information about their direction of attention and their emotional and mental states". They further note that "eyes and their highly expressive surrounding region can communicate complex mental states such as emotions, beliefs, and desires" and "observing another person's behavior allows the observer to decode and make inferences about a whole range of mental states such as intentions, beliefs, and

emotions". Furthermore, the gaze based social mechanisms are specifically decoded by a part of the cortex, namely the *posterior superior temporal sulcus* region, that has been found [12] to responds to the inferred intentionality of social cues.

The underlying expression or the mental process of a person may be enquired by other means than face analysis, such as gaze direction. Liversedge and Findlay [13] showed that saccades parameters are correlated with the underlying cognitive process, namely the duration of fixations and the choice of saccade target emphasize continuities between biological and cognitive descriptions. The connection between underlying emotion, expression and gaze was discussed by Adams and Kleck [14], who proved that, indeed, when gaze direction matches the underlying behavioral intent (approach-avoidance) communicated by an emotional expression, the perception of that emotion is enhanced. Following these findings, a palette of applications based on recording the eye movements have been proposed. Typically developing such applications involves two steps: first, the hypotheses regarding the correlation between eye movements and some behavior pattern or social process is formulated and validated on a series of experiments; next based on the previous found conclusions, practical applications are proposed.

Such a distinct category is the analysis of the reading process with the aim of understanding the learning to read process or how attention correlates with understanding. For instance, Joseph et al. [15] investigated, by means of gaze tracking, insights of children process to *learn to read* and the words frequency impact in sentence reading. Godfroid et al. [16] used eye tracking measurements to test hypothesis concerning words complexity, attention persistence and short/long term memory. In the same line Rayner et al. [17] monitored subjects eye movements while read sentences containing high- or low-predictable target words; their findings showed word predictability (due to contextual constraint) and word length have strong and independent influences on word skipping and fixation durations. Possible application of the eye movement control in teaching are discussed, for instance, in [18].

Moving further, Chun [19] showed that while reading or simply scanning an image, the eye movement may give hints about the observers self build context. This idea was further developed by Bulling and Zander [20] who suggested an application that provides additional information relevant to the context; the various possibilities of the adaptive context are retrieved from analysis of the eye movements.

Another important category is related to the use of eye movements in gaming. For instance Meijering et al. [21] discussed the possibility and show evidence that eye movements are correlated with the plan type used in a strategy game; more precisely forward reasoning (where a player proceed from the initial point to finish) or backward reasoning (where the path from end to start is retrieved) are distinguishable by overlapping the eye movement with the game board. Furthermore, Krejtz et al. [22] investigated the degree of enjoyment when visual cues influences gaming experience and conclude that not only in such a scenario there is no additional cognitive effort but there are many arcade game optimization possible in such a context that would increase the pleasure of users.

2.1 Eye Accessing Cues in Neuro-Linguistic Programming

The Neuro-Linguistic Programming was introduced in the 1970s by Brandler and Grinder [1], as a different model for detecting, understanding and using the patterns that appear between brain, language and body. The NLP theory jumped over intensive and extensive academic investigation and made path very fast into the commercial market. Rigorous investigation was expected after the initial publication and consensus has not been reached yet in the academic world.

The Eye Accessing Cues from the NLP theory are not unanimously accepted, with some of the most recent research on the topic calling for further testing [23]. A recent experiment by Vrânceanu et al. [24], with the scope to gain better insight of the facts, showed that while not 100 % accurate (i.e. universal), the correct apparition rates were higher than random chance, especially between visual, auditory and kinesthetic ways of thinking (corresponding to a separation along the vertical axis of the gaze direction).

2.2 Recognizing Gaze Direction: Premises

The problem of identifying one's direction of gaze is intensively studied in computer vision. These systems may be classified by the position of the recording device as:

1. Head mounted devices (e.g. glasses or head mounted cameras);
2. Stationary and/or remote devices.

The head mounted devices are closer to the eye and because of that they have access to a higher resolution and better precision. But they are rather expensive (their price spans from several thousand dollars, for a professional commercial solution, down to a hundred dollars for the more affordable ones, compared to a few dollars for a normal webcam). Another shortcoming of the head mounted devices that restricts their area of usability is the fact that they are wearable. This may be a distinct indicator that the user is subject to investigation by non-traditional means and it has been showed [25] that voluntary control is exercisable over non-visual eye movements. This is why a stationary webcam is preferable for investigating the eye accessing cues.

Another way of classifying the gaze direction estimation systems may be performed according to the illumination source domain. Here, we may note:

1. Active, infra-red (IR) based illumination;
2. Passive, visible spectrum illumination.

The commercial eye-trackers, which have higher reported precisions, rely on the information from the IR domain. But again we note that this implies a distinct, specialized device (because the IR source is not typically incorporated in webcams). As in the case of wearable eye tracking, the use of specialized recording devices

needed by active illumination sources limits the applicability of methods to data that was recorded accordingly. Thus a system having the goal the recognizing the Eye Accessing Cues should be based on a normal digital video camera.

As for the usability of the NLP-EAC hypothesis one can imagine many areas. We will give just two simple examples.

The first example of how can one use the NLP-EAC hypothesis refers to online interviews. Small and medium enterprises may look for employees at a distance and the interview is usually online. In this case the candidate is recorded (sometimes by his/her own device) and, given a query, discrimination between remembering type of activities (looking left) and the constructing ones (looking right) can differentiate experience from creativity. But it is imperative that the interviewed person is not aware that his/hers non-verbal messages are recorded and analyzed. The method described in this chapter requires typical recoding devices for video transmission; thus no distinct means of recording (head mounted camera or/and active light sources) are involved.

The second use-case deals with interactive communication for marketing and training. If the interaction is face-to-face, the meeting may be recorded and analyzed either real-time (with conclusions being shown to the presenter), or before the next session (such that the trainer/seller will ensure that maximum of information reaches his interlocutors). If the communication is online, the restrictions are similar with online interviews: the subject must have access to a recording device which is usually a typical webcam.

3 Databases

3.1 Iris Center Annotated Databases

To set the introductory reports on iris center localization performance, the BioID database[1] is the most popular choice. This database contains 1521 gray-scale, frontal facial images of size 384×286, acquired with frontal illumination conditions in a complex background. The database contains 16 tilted and rotated faces, people that wear eye-glasses and, in very few cases, people that have their eyes closed (i.e. blink) or pose various expressions. The database was released with annotations for iris centers. Being one of the first databases that provided facial annotations, BioID became the most used database for face landmarks localization accuracy tests, even if it provides limited variability and reduced resemblance with real-life cases.

As many methods use the entire eye area as learned template, robustness to face expression should be envisaged, as it induces eye shape changes. In this sense the most appropriate choice would be the Cohn-Kanade database[2] [26]. This database

[1] http://www.bioid.com/downloads/software/bioid-face-database.html.

[2] http://www.pitt.edu/~emotion/ck-spread.htm.

was developed for the study of emotions, contains frontal illuminated portraits and it is challenging through the fact that eyes are in various poses (near-closed, half-open, wide-open).

Further, one should systematically evaluate the robustness of the iris localization methods with respect to lighting and pose changes. Appropriate tests may be conducted onto the Extended Yale Face Database B (B+)[3] [27]. The Extended Yale B database contains 16,128 gray-scale images of 28 subjects, each seen under 576 viewing conditions (9 poses × 64 illuminations). The size of each image is 640×480.

The BioID, Cohn-Kanade and Extended YaleB databases include specific variations as they are acquired under controlled lighting conditions with frontal faces only. In contrast, there are databases like the Labelled Face Parts in the Wild (LFPW) [28] and the Labelled Faces in the Wild[4] (LFW) [29], which are randomly gathered from the Internet and contain large variations in the imaging conditions. While LFPW is annotated with facial point locations, only a subset of about 1500 images is made available and contains high resolution and rather qualitative images. In opposition, the LFW database contains more than 12,000 facial images, having the resolution 250 × 250 pixels, with 5700 individuals that have been collected "in the wild" and vary in pose, lighting conditions, resolution, quality, expression, gender, race, occlusion and make-up. The face landmarks[5] and iris centers are publicly available.

3.2 Iris Center and Eye Landmarks Annotated Databases

To evaluate the performance of the eye landmarking algorithm, four annotated databases are at hand, with publicly available ground-truth: EyeChimera, HPEG, ULM and PUT.

To study the specifics of the EAC detection problem, the *Eye Chimera* Database[6] [24, 30] was developed so that it contains all the seven cues. In generating the database, 40 subjects were asked to move their eyes according to a predefined pattern and their movements were recorded. The movements between consecutive EACs were identified, the first and last frame of each move were selected and labelled with the corresponding EAC tag and eye points were manually marked. In total, the database comprises 1170 frontal face images, grouped according to the seven directions of gaze, with a set of five points marked for each eye: the iris center and four points delimiting the bounding box. Additionally, for more extensive

[3] vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html.

[4] The database is available at http://vis-www.cs.umass.edu/lfw/.

[5] At http://blog.gimiatlicho.webfactional.com/?page_id=38.

[6] http://imag.pub.ro/common/staff/cflorea/EyeChimeraReleaseAgreement.pdf.

testing, the still Eye Chimera database was extended with all the consecutive frames that are part of each basic eye movement; this part was named Eye Chimera Sequences.

The HPEG database[7] [31] is given as videos and we have extracted frames with relevant gaze variation, resulting in 233 images (640×480 resolution) of 10 persons, who's eye gaze varies from left to right (no vertical gaze direction is available). The head position includes yaw variations from −30° to +30°. The dataset contains two sessions, one in a close-up arrangement while the other with people placed more distantly from the camera. The database comes with annotation related to the head angle, but without landmark positions; these were added in [30] and are available online.

The ULM head and gaze database[8] [32] contains images (1600 × 1200 pixels) of 20 persons. Variations include gaze direction (left to right), and head pose on both yaw and pitch. The database contains annotation for six eye landmarks: inner eye limits, outer eye limits and pupil centers. Because not all the characters have the images marked and in tests the images with yaw or pitch angles higher in absolute value than 30°, are excluded; only 335 images have been kept and were used in the current study.

The PUT database[9] [33], is built in a similar manner with ULM. However the marking set is more complete as it contains all ten landmarks envisaged for the eye regions. Overall, it contains slightly more than 1000 annotated images.

4 System Overview

Approaches to the mentioned problem [10, 30, 34, 35] assume a scenario where the image acquisition is done with a single camera with fixed, near-frontal position, under free natural lighting.

The discussed algorithms rely solely on gray-scale images and a coarse-to-fine approach is used for localization, succeeded by gaze direction recognition. The possible schematic of such a system may be followed in Fig. 2.

First, the face bounding box is retrieved; the preferred method is derived from the Viola-Jones algorithm [36]. Given the face bounding box, the image is re-scaled at a fixed size. Next a rough estimation of the iris center should be retrieved. While in some works the iris center is treated as one of the landmarks, yet many of the current solutions make use of the physical particularities of the iris–pupil structure and are specifically optimized for iris center. The rough iris center, even it may be improved

[7] http://emotion-research.net/toolbox/toolboxdatabase.2010-02-03.4835728381.

[8] http://www.uni-ulm.de/in/neuroinformatik/mitarbeiter/g-layher/image-databases.html.

[9] https://biometrics.cie.put.poznan.pl.

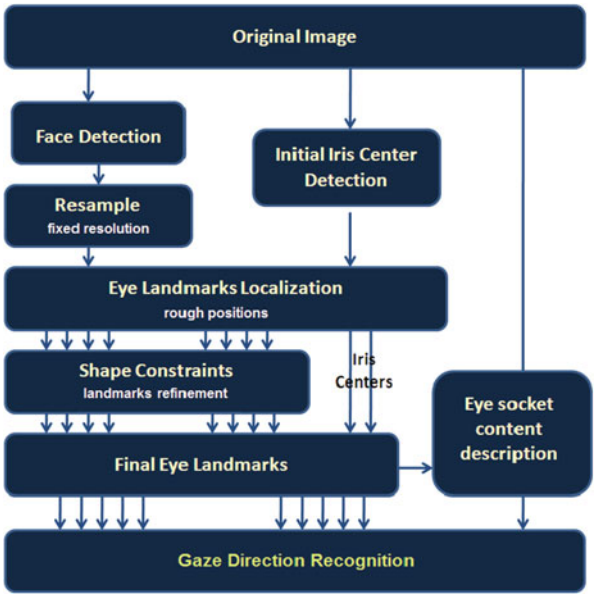

Fig. 2 The system for recognizing the gaze direction

and does vary with respect to gaze direction, is considerable more precise than the face square, thus it acts as a better initialization reference for the position of other landmarks.

The subsequent major step of the procedure refers to the localization of the eye socket landmarks. The prerequisite here are some potential areas deduced using anthropometric criteria and the initial iris position. The process of localization, due to take place in a rather uncommon condition set (gaze variation), usually contains a two step procedure; the initial rough positioning is followed by a refinement step in which case the eye socket landmarks make use of shape constraints.

Once the eye region is determined it is separated and further analyzed for recognizing the gaze direction. While in previous similar solutions this step is referred as gaze (eye) tracking, for the application set envisaged here the outcome is a categorical set of directions (e.g. three or seven directions), thus it seems more appropriate to name it gaze direction recognition.

4.1 Face Detection and Localization

The first step of the system is locating the face bounding box. While many method have been proposed, by far, the most popular is still the boosted cascade of Haar features introduced by Viola and Jones [36]. Currently there exist many public implementations, the OpenCV version being one of the most used.

A recent reevaluation by Mathias et al. [37] showed that, if trained properly, the Vanilla Deformable Part Models [38] reaches top performance. Additional choices

for face detection problem may be found on the Face Detection Data Set and Benchmark web page[10] and more recently on the Fine-grained Evaluation on Face detection in the Wild.[11]

5 Iris Center Localization

The problem of iris center localization was well investigated in literature, within a long history, as showed in the review by Song et al. [39]. Methods for eye center (or iris or pupil) localization in passive, remote imaging may approach the problem either as a particular case of pattern recognition application, [40, 41] or by using the physical particularities of the eye, like the high contrast with respect to the neighboring skin [42] or the circular shape of the iris [43]. More recent methods combine the two approaches.

As a general observation, we note that while older solutions [42, 44], tried to estimate also the face position, since the appearance of the Viola-Jones face detection solution [36], eye center search is limited to a subarea within the upper face square. Taking into account the recent advances on the face detection problem, one may truthfully assume that reconsideration of older eye methods may lead to better results than initially stated.

5.1 State of the Art Solutions

In this chapter we will point the attention of the reader to a very fast and robust iris center localization method based on zero-crossing encoded image projection. A list of other methods used for iris centers localization may be retrieved from the review by Song et al. [39] and from the summary presented in Table 1 and following paragraphs.

5.1.1 Projections Based Iris Localization Methods

The same image projections as in the work of Kanade [45] are used to extract information for eye localization in a plethora of methods [46–48]. Feng and Yuen [46] started with a snake based head localization followed by anthropometric reduction (relying on the measurements of Verjak and Stephancic [49]) to the

[10] http://vis-www.cs.umass.edu/fddb/results.html.

[11] http://www.cbsr.ia.ac.cn/faceevaluation.

Table 1 Iris center localization methods in remote imaging

Method	Face detection	Features	Machine learning	Public code
Jesorsky et al. [44]	No	Edges	MLP	No
Feng and Yuen [46]	No	VPF	No	No
Zhou and Geng [47]	No	GPF	No	No
Turkan et al. [48]	Yes	EPF	SVM	No
Cristinacce et al [50]	Yes	Image pixels	PRFR	Yes (AAM)
Campadelli et al. [51]	Yes	Haar wavelets	No	No
Hamouz et al. [40]	No	Gabor filters	SVM	No
Niu et al. [52]	Yes	Haar wavelets	AdaBoost	No
Kim et al. [53]	Yes	Image pixels	AdaBoost	No
Asteriadis et al. [41]	Yes	DVF	minDist	No
Valenti and Gevers [43]	Yes	Isophote	Yes	Yes
Asadifard [54]	Yes	Gradient	No	No
Ding and Martinez [55]	Yes	Pixels	AdaBoost/DA	Yes
Timm and Barth [56]	Yes	Image pixels	No	Yes
Kawulok and Szymanek [57]	Yes	Edges	SVM	Yes
Florea et al. [58]	Yes	Image pixels	MLP	Yes

so-called eye-images and introduce the variance projections for localization. The key points of the eye model are the projections particular values, while the conditions are manually crafted.

Zhou and Geng [47] described convex combinations between integral image projections and variance projections that are named generalized projection functions (GPF). These are filtered and analyzed for determining the center of the eye. The analysis is also manually crafted and requires identification of minima and maxima on the computed projection functions. Yet in specific conditions, such as intense expression or side illumination, the eye center does not correspond to a minima or a maxima in the projection functions.

Turkan et al. [48] introduced the edge projections and used them to roughly determine the eye position. Given the eye region, a feature is computed by concatenation of the horizontal and vertical edge image projections. Subsequently,

a SVM-based identification of the region with the highest probability is used for marking the eye. Florea et al. [58] also employed a projection based description of the eye region coupled with machine learning; yet this method aims for higher precision and robustness, so multiple projection types are used and simple but efficient dimensionality reduction techniques are employed for speeding up the process.

5.1.2 Pattern Recognition Based Methods

There are many other approaches to the problem of eye localization. Jesorsky et al. [44] proposed a face matching method based on the Hausdorff distance followed by a MLP eye finder. Wu and Zhou [42] even reversed the order of the typical procedure: they used eye contrast specific measures to validate possible face candidates.

Cristinacce et al [50] relied on the Pairwise Reinforcement of Feature Responses algorithm for feature localization. Campadelli et al. [51] used SVM on optimally selected Haar wavelet coefficients.

Hamouz et al. [40] refined with SVM the Gabor filtered faces, for locating ten points of interest; yet the overall approach is different from the face feature fiducial points approach. Niu et al. [52] used an iteratively bootstrapped boosted cascade of classifiers based on Haar wavelets. Kim et al. [53] use multi scale Gabor jets to construct an Eye Model Bunch. Asteriadis et al. [41] used the distance to the closest edge to describe the eye area. Valenti and Gevers [43, 59] used isophote's properties to gain invariance and follow with subsequent filtering with Mean Shift (MS) or nearest neighbor on SIFT feature representation for higher accuracy. Asadifard [54] relied on thresholding the cumulative histogram for segmenting the eyes. Ding and Martinez [55] trained a set of classifiers (with a SVM of with discriminant analysis—DA) to detect multiple face landmarks, including explicitly the pupil center, by using a sliding window approach and test in all possible locations and inter-connect them to estimate the overall shape. Timm and Barth [56] relied their eye localizer on gradient techniques and search for circular shapes. Kawulok and Szymanek [57] fit a multilevel ellipsoid regressed from Hough accumulator planes over the face and the eyes and optimize the localization using SVM.

5.2 Robust Eye Centers Localization with Zero-Crossing Encoded Image Projections

Recently, Florea et al. [58] proposed a framework for the eye centers localization by the joint use of encoding of normalized image projections and a Multi Layer Perceptron (MLP) classifier. This encoding consists in identifying the zero-crossings and extracting the relevant parameters from the resulting modes.

5.2.1 Integral and Edge Image Projections

The integral projections, also named integral projection functions (IPF) or amplitude projections, are tools that have been previously used in face analysis. They appeared as "amplitude projections" [60] or as "integral projections" [45] for face recognition. For a gray-level image sub-window $I(i,j)$ with $i = i_m \ldots i_M$ and $j = j_n \ldots j_N$, the projection on the horizontal axis is the average gray-level along the columns (1), while the vertical axis projection is the average gray-level along the rows (2):

$$P_H(j) = \frac{1}{i_M - i_m + 1} \sum_{i=i_m}^{i_M} I(i,j), \forall j = j_n, \ldots, j_N \tag{1}$$

$$P_V(i) = \frac{1}{j_N - j_n + 1} \sum_{j=j_n}^{j_N} I(i,j), \forall i = i_m, \ldots, i_M \tag{2}$$

To increase the robustness to side illumination, edge projection functions (EPF) could be used to complement the integral ones. To determine them, the classical horizontal and vertical Sobel contour operators are applied, resulting in S_H and S_V, which are combined in the $S(i,j)$ image used to extract edges:

$$S(i,j) = S_H^2(i,j) + S_V^2(i,j) \tag{3}$$

The edge projections are computed on the corresponding image rectangle $I(i,j)$ by replacing I with S in Eqs. (1) and (2).

5.2.2 Fast Computation of Projections

While sums over rectangular image sub-windows may be easily computed using the concept of summed area tables [61] or integral image [36], a fast computation of the integral image projections may be achieved using the *prefix sums* [62] on rows and respectively on columns. A prefix-sum is a cumulative array, where each element is the sum of all elements to the left of it, inclusive, in the original array. They are the 1D equivalent of the integral image, but they definitely precede it.

For the fast computation of image projections, two tables are required: one will hold prefix sums on rows (a table which, for keeping the analogy with the integral image, will be named horizontal 1D integral image) and respectively one vertical 1D integral image that will contain the prefix sums on columns. It should be noted

that computation on each row/column is performed separately. Thus, if the image has $M \times N$ pixels, the 1D horizontal integral image, on the column j, $\mathscr{I}_H^j$, is:

$$\mathscr{I}_H^j(i) = \sum_{k=1}^{i} I(k,j) \ , \forall i = 1, \ldots M \tag{4}$$

Thus, the horizontal integral projection corresponding to the rectangle $i = [i_m; i_M] \times [j_n; j_N]$ is:

$$P_H(j) = \frac{1}{i_M - i_m + 1} \left(\mathscr{I}_H^j(i_M) - \mathscr{I}_H^j(i_m - 1) \right) \tag{5}$$

Using the oriented integral images, the determination of the integral projections functions on all sub-windows of size $K \times L$ in an image of $M \times N$ pixels requires one pass through the image and $2 \times M \times N$ additions, $2 \times (M-K) \times (N-L)$ subtractions and two circular buffers of $(K+1) \times (N+1)$ locations, while the classical determination requires $2 \times K \times L \times (M-K) \times (N-L)$ additions. Hence, the time to extract the projections associated with a sub-window, where many sub-windows are considered in an image, is greatly reduced.

The edge projections require the computation of the oriented integral images over the Sobel edge image, $S(i,j)$ which needs to be found on the areas of interest.

5.2.3 Encoding and ZEP Feature

To reduce the complexity (and computation time), the projections are compressed using a zero-crossing based encoding technique. After ensuring that the projections values are in a symmetrical range with respect to zero, one will describe, independently, each interval between two consecutive zero-crossings. Such an interval is called an *epoch* and for its description three parameters are considered (as presented in Fig. 3):

- *Duration*—the number of samples in the epoch;
- *Amplitude*—the maximal signed deviation of the signal with respect to 0;
- *Shape*—the number of local extremes in the epoch.

The proposed encoding is similar with the TESPAR (Time-Encoded Signal Processing and Recognition) technique [63] that is used in the representation and recognition of 1D, band-limited, speech signals. Depending on the problem specifics, additional parameters of the epochs may be considered (e.g. the difference between the highest and the lowest mode from the given epoch). Further extensions are at hand if an epoch is considered to be the approximation of a probability density function and the extracted parameters are the statistical moments of the said distribution. In such a case the *shape* parameter corresponds to the number of modes of the distribution.

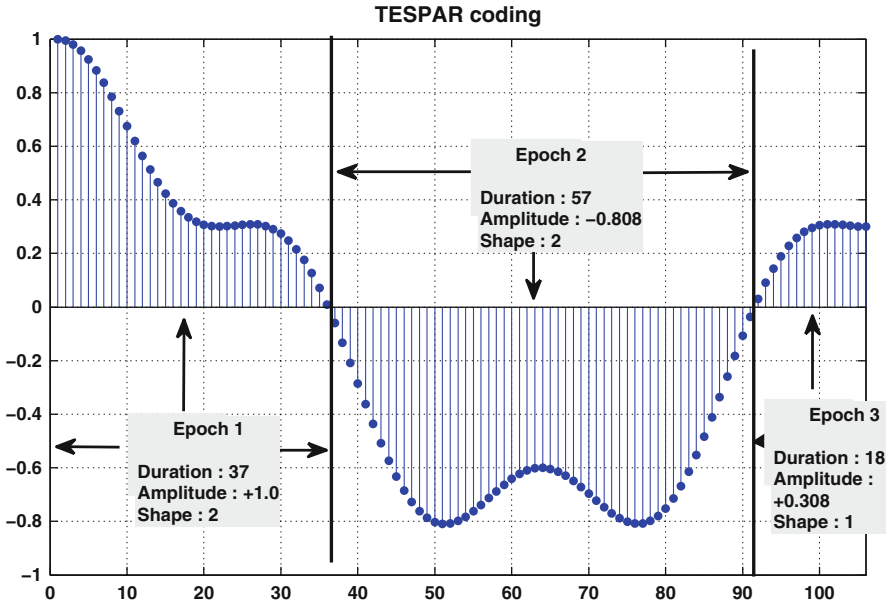

Fig. 3 Example of 1D signal (vertical projection of an eye crop) and the associated encoding. There are three epochs, each encoded with three parameters. The associated code is: $[37, +1, 2;\ 57, -0.808, 2;\ 18, 0.308, 1]$

The reason for choosing this specific encoding is twofold. First, the determination of the zero-crossings and the computation of the parameters can be performed in a single pass through the target 1D signal, and, secondly, the epochs have specific meaning when describing the eye region.

Given an image sub-window, the ZEP feature is determined by the concatenation of four encoded projections as described in the following:

1. Compute both the integral and the edge projection functions (P_H, P_V, E_H, E_V);
2. Independently *normalize* each projection within a symmetrical interval. For instance, each of the projections is normalized to the $[-1; 1]$ interval. This will normalize the amplitude of the projection;
3. Encode each projection as described; allocate for each projection a maximum number of epochs;
4. Normalize all other (i.e. duration and shape) encoding parameters;
5. Form the final Zero-crossing based Encoded image Projections (ZEP) feature by concatenation of the encoded projections. Given an image rectangle, the ZEP feature consists of the epochs from all the four projections: (P_H, P_V, E_H, E_V).

Image projections are simplified representations of the original image, each of them carrying specific information; the encoding simplifies even more the image representation. The normalization of the image projections, and thus of the epochs amplitudes, ensures independence of the ZEP feature with respect to uniform

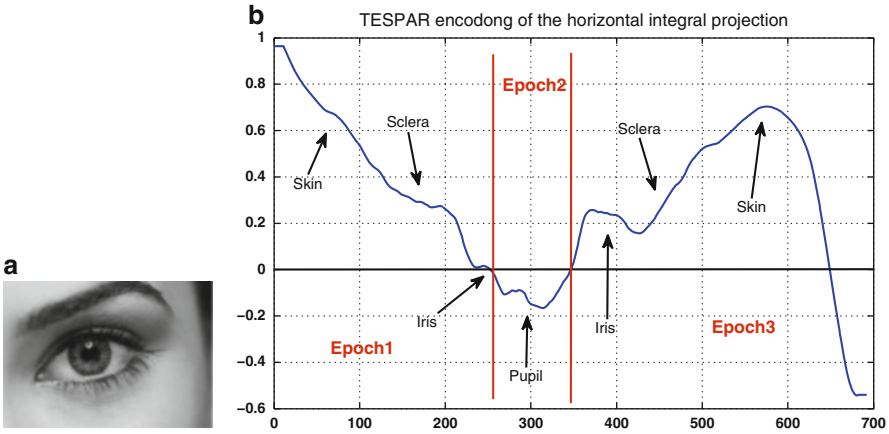

Fig. 4 Horizontal image projection function from a typical eye patch: (**a**) eye crop, (**b**) the integral projection on the eye crop and the physical features marked. The *vertical lines* mark the zero crossing that are typically found on all eye examples

variation of the illumination. The normalization with respect to the number of elements in the image sub-window leads to partial scale invariance: horizontal projections are invariant to stretching on the vertical direction and vice versa. The scale invariance property of the ZEP feature is achieved by completely normalizing the encoded durations to a specific range (e.g. the encoded horizontal projection becomes invariant to horizontal stretching after duration normalization).

In [47] it was noted that image projections in the eye region have a specific sequence of relative minima and maxima assigned with to skin (relative minimum), sclera (relative maximum), iris (relative minimum), etc.

Considering a rectangle from the eye region including the eyebrow (as showed in Fig. 4a), the associated integral projections have specific epochs, as showed in Fig. 4b. The particular succession of positive and negative modes is precisely encoded by the this technique. On the horizontal integral projection there will be a large (one-mode) epoch that is assigned to skin, followed by an epoch for sclera, a triple mode, negative, epoch corresponding to the eye center and another positive epoch for the sclera and skin. On the vertical integral projection, one expects a positive epoch above the eyebrow, followed by a negative epoch on the eyebrow, a positive epoch between the eyebrow and eye, a negative epoch (with three modes) on the eye and a positive epoch below the eye.

The ZEP feature, due to invariance properties already discussed, achieves consistent performance under various stresses and is able to discriminate among eyes (patches centered on pupil) and non-eyes (patches centered on locations at a distance from the pupil center).

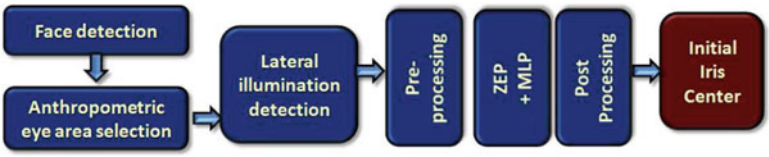

Fig. 5 The work flow of the rough iris center localization algorithm

5.2.4 Rough Iris Center Localization

The block schematic of the initial rough eye center localization algorithm is summarized in Fig. 5. The main problem when employing projections for iris localization is their susceptibility to lateral illumination. In such a case, to avoid loosing accuracy, Florea et al. [58] proposed a preemptive detection of the lateral illumination case by observing that for a near-frontal placed light source the distribution of intensity should be symmetrical with respect to the nose area. This is a simple example of the *face relighting* problem and further details may be retrieved from [64].

The conceptual steps of the actual eye localization procedure are: preprocessing, machine learning and postprocessing.

A simple preprocessing is applied for each eye candidate region to accelerate the localization process. Wu and Zhou [42] noted that the pupil center is significantly darker than the surrounding; thus the pixels that are too bright with respect to the eye region (and are not plausible to be pupil centers) are discarded. The "too bright" characteristic is encoded as gray-levels higher than a percentage from the maximum value of the eye region. In the lateral illumination case, this threshold is higher due to the deep shadows that can be found on the skin area surrounding the eye.

In the area of interest, using a sliding window, all plausible locations are investigated by ZEP+MLP. To further accelerate the algorithm only some of the values should be further considered [58]. The potential positions are recorded in a separate image which is further post-processed for eye center extraction. Further attention needs to be given to the confusion between eye and eyebrows; a possible procedure is a pre-segmentation and to locate the iris one will look only to the lower darker region [10].

For the frontal illumination case, in the case of training with L2 distance as objective, one expects a symmetrical shape around the true eye center. Thus the final eye location is taken as the weighted center of mass of the previously selected eye regions. For the lateral illumination, the binary trained MLP is supposed to localize the area surrounding the eye center and the final eye center is the geometrical center of the rectangle circumscribed to the selected region. In both cases, the specific way of selecting the final eye center is able to deal with holes (caused by reflections or glasses) in the eye region.

The training of the machine learning system should be performed with crops of eyes and non-eyes. The positive examples are to be taken near the eye ground truth

while the patches corresponding to the negative examples should also overlap with the true eye but to a lesser degree. This choice forces the machine learning to give importance in precise discrimination of patches centered on the iris and patches centered elsewhere.

5.2.5 Iris Center Refinement

If challenged by the gaze direction variation, the rough method does not perform very well and further refinement is necessary [30].

To improve the performance of some initial iris center localization method one may consider a small, centered region of interest (ROI) (e.g. of size $\frac{d_{eye}}{5} \times \frac{d_{eye}}{5}$, where d_{eye} is the inter-ocular distance) around the reported eye center. The improvement of the iris centers, as well as the detection of the eye socket limits require position and intensity priors and template matching. These are identical with eye landmarks localization procedure and will be discussed in the next paragraphs.

5.3 Evaluation of the Iris Center Localization Methods

5.3.1 Evaluation on the BioID, Cohn-Kanade, Yale B+ and LFW Databases

The iris centers localization performance is typically evaluated according to the stringent localization criterion [44]. The eyes are considered to be correctly determined if the specific localization error ϵ, defined in Eq. (6) is smaller than a predefined value.

$$\epsilon = \frac{\max\{\varepsilon_L, \varepsilon_R\}}{D_{eye}} \tag{6}$$

In the equation above, ε_L is the Euclidean distance between the ground truth left eye center and determined left eye center, ε_R is the corresponding value for the right eye, while D_{eye} is the distance between the ground truth eyes centers. Typical error thresholds are $\epsilon = 0.05$ corresponding to eyes centers found inside the true pupils, $\epsilon = 0.1$ corresponding to eyes centers found inside the true irises, and $\epsilon = 0.25$ corresponding to eyes centers found inside the true sclera. This criterion identifies the worst case scenario.

To give an initial overview of the problem in state of the art, we report in Table 2 the results of multiple methods performance on the BioID database.

Considering as most important criterion the accuracy at $\epsilon < 0.05$, it should be noted that Timm and Barth [56] and Valenti and Gevers [59] provide the highest accuracy. Yet, the best performance achieved by a variation of the method described in [59], namely Val.&Gev.+SIFT contains a tenfold testing scheme, thus using nine

Table 2 Comparison with state of the art (listed in chronological appearance) in terms of localization accuracy on the BioID database

Method	Accuracy		
	$\epsilon < 0.05$	$\epsilon < 0.1$	$\epsilon < 0.25$
Florea et al. [58]	70.46	91.94	98.89
Jesorsky et al. [44]	40.0	79.00	91.80
Wu and Zhou [42]	10.0*	54.00*	93.00*
Zhou and Geng [47]	47.7	74.5	97.9
Cristinacce et al. [50]	55.00*	96.00	98.00
Campadalli et al. [51]	62.00	85.20	96.10
Hamouz et al. [40]	59.00	77.00	93.00
Turkan et al. [48]	19.0*	73.68	*99.46*
Kroon et al. [65]	65.0	87.0	98.8
Asteriadis et al. [41]	62.0*	89.42	96.0
Asadifard et al. [54]	47.0	86.0	96.0
Timm and Barth [56]	82.50	*93.40*	98.00
Val.&Gev. [59]+MS	81.89	87.05	98.00
Val.&Gev. [59]+SIFT†	*86.09*	91.67	97.87

The correct localization presents results reported by authors; values marked with "*" were inferred from authors plot. While Zhou [47] reports only the value for $\varepsilon < 0.25$, the rest is reported by Ciesla and Koziol [66]. The method marked with † relied on a tenfold training/testing scheme, thus, at a step, using nine parts of the BioID database for training

parts of the BioID database for training. Furthermore, taking into account that BioID database was used for more that 10 years and provides limited variation, it has been concluded [28, 67] that other tests are also required to validate a method.

Valenti and Gevers [59] provide results on other datasets and made public the associated code for their baseline system (Val.&Gev.+MS), which is not database dependent. Timm and Barth [56] do not provide results on any other database except BioID or source code, yet there is publicly available[12] code developed with author involvement. Thus, in continuation, we will compare the method from [58] against these two on other datasets. Additionally, we include the comparison against the eye detector developed by Ding and Martinez [55] which has also been trained and tested on other databases, thus is not BioID dependent.

As mentioned in the introduction, the purpose of this chapter is to investigate emergent application with potential on behavior inference. Thus we will report eye localization performance with respect to facial expressions, as real-life cases with fully opened eyes looking straight are rare. We tested the performance of various methods on the Cohn-Kanade database [26]. This database was developed for the study of emotions, contains frontal illuminated portraits and it is challenging

[12] http://thume.ca/projects/2012/11/04/simple-accurate-eye-center-tracking-in-opencv/.

Table 3 Percentage of correct eye localization on the Cohn-Kanade database

Method	Type	Accuracy		
		$\epsilon < 0.05$	$\epsilon < 0.1$	$\epsilon < 0.25$
Florea et al. [58]	Neutral	76.0	99.0	100
	Apex	71.9	95.7	100
	Total	73.9	97.3	100
Valenti and Gevers [59]	Neutral	46.0	95.7	99.6
	Apex	35.1	92.4	98.8
	Total	40.6	94.0	99.2
Timm and Barth [56]	Neutral	66.0	95.4	99.0
	Apex	61.4	85.1	93.4
	Total	63.7	90.2	96.2
Ding and Martinez [55]	Neutral	14.3	75.9	100
	Apex	11.8	72.8	100
	Total	13.1	74.4	100

We report results of the methods form [55, 56, 58, 59] on the neutral poses, expression apex and overall. We marked with italics the best achieved performance for each accuracy criterion and respectively for each image type

Table 4 Comparative results on the Extended YaleB database. We marked with bold letters the best performance for each accuracy category

Method	$\epsilon < 0.05$	$\epsilon < 0.1$	$\epsilon < 0.25$
Florea et al. [58]	***39.9***	***67.3***	*97.3*
Valenti and Gevers [59]	37.8	66.6	**98.5**
Timm and Barth [56]	20.1	34.5	51.5
Ding and Martinez [55]	19.7	47.8	58.6

through the fact that eyes are in various poses (near-closed, half-open, wide-open). We tested only on the neutral pose and on the expression apex image from each example.

We note that solutions that try to fit a circular or a symmetrical shape over the iris, like the ones from [59] or [56], and thus, performs well on open eyes, do encounter significant problems when facing eyes in expressions (as it is shown in Table 3). Taking into account the achieved results, which are comparable on neutral pose and expression apex images, it is to be seen which method performs very well under such complex conditions. Results indicate that [58] achieved higher accuracy when compared with rest of the method tested.

A systematic evaluation of the robustness of iris center localization algorithms with respect to lighting and pose changes may be done using the Extended Yale Face Database B (B+) [27]. Here, by a small margin the best performance is obtained by the method from [58], followed closely by the one from [59]. The methods proposed by Timm and Barth [56] and respectively Ding and Martinez [55] have greater susceptibility to errors due to uneven illumination (Table 4).

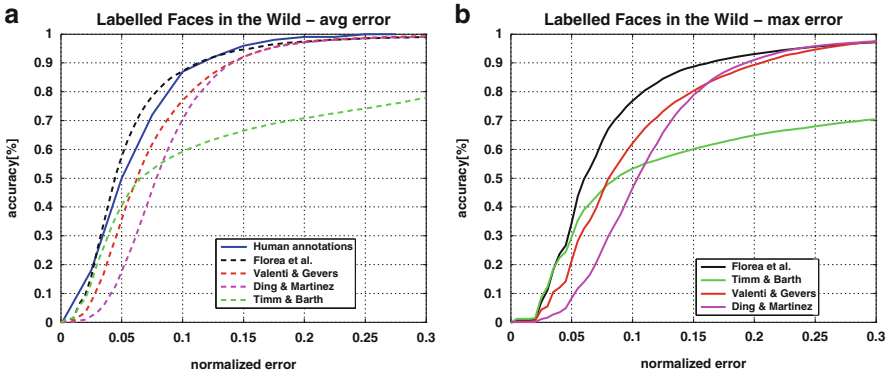

Fig. 6 Results achieved on the LFW database: (**a**) average error and (**b**) maximum error as imposed by the stringent criterion—Eq. (6). With *dashed blue line* is the average error for human evaluation

The difficulty of localizing features on the LFW database is certified by the performance of human evaluation error as reported in [67]. While the ground truth is taken as the average of human markings for each point normalized to inter-ocular distance, human evaluation error is considered as the averaged displacement of the one marker. Regarding the achieved results, the ZEP based method [58] provides the most accurate results, as one can see in Fig. 6. The improvement is by almost 50 % at $\epsilon < 0.05$ compared with [59] and from [56] and with more over [55].

5.3.2 Evaluation on Databases with Gaze Variability

The main purpose of this chapter is to discuss the potential application of computer vision methods for the communication by gaze induced methods; thus we will evaluate the performance of various methods on databases illustrating gaze variability. In Table 5, for comparison purposes, we report the performance the method from [58] and of its refinement from [30]. Additional we report the results achieved with the method of Sun et al. [68]. This localizes face landmarks; yet from their searched set, of interest for the current problem are only the iris centers; thus we report their performance specifically to the iris center section.

The highest accuracy is achieved by the deep convolution network based system [68], with very high results especially for medium accuracy ($\varepsilon = 0.1$). Yet, multi-stage method from [30] clearly outperforms it for high accuracy ($\varepsilon = 0.05$), which is critical in achieving high EAC recognition rate.

Table 5 Iris center localization accuracy measured as percentage of the inter-ocular distance (stringent criterion [44])

	Database					
	Eye Chimera		HPEG		ULM	
Method	$\varepsilon = 0.05$ (%)	$\varepsilon = 0.1$ (%)	$\varepsilon = 0.05$ (%)	$\varepsilon = 0.1$ (%)	$\varepsilon = 0.05$ (%)	$\varepsilon = 0.1$ (%)
Florea et al. [58]	42.7	68.9	53.5	78.8	32.5	74.1
Timm and Barth [56]	33.6	67.8	46.9	82.0	61.2	85.2
Valenti and Gevers [59]	16.1	50.5	24.6	55.7	50.1	77.0
Sun et al. [68]	59.9	91.2	54.2	90.3	60.8	93.2
Florea et al. [30]	65.3	78.7	71.1	83.2	76.6	92.7

Fig. 7 The five eye landmarks searched

6 Eye Landmarking

In order to locate properly the position of the eye for further processing while determining the direction of gaze, one needs to identify the eye landmarks. The typical set of eye landmarks is showed in Fig. 7. While there have been developed methods focusing only on the eye, most of the state of the art methods are general face landmarking methods. An overview of some of the most relevant such methods may be followed in Table 6.

Facial landmarking originates in the classical holistic approaches of Active Shape Models (ASMs) [69], Active Appearance Models (AAMs) [70] and Elastic Graph Matching [71]. Active Shape Models describe an eigenspace of the geometrical shape having as vertices the landmarks, while the AAM complement the information with pixel values from shape interior.

Building on the ASMs/AAMs versatility, a multitude of extensions appeared. For instance in [72], a 2D profile model and a denser point set are used. In [73], higher independence between the facial components is encouraged while the actual fitting step is further optimized by a Viterbi optimization process.

In the later period, the ASM underlying holistic fitting switched to independent models built on top of local part detectors, to form the so-called Constrained Local Models (CLMs) [74], or to a combination of local shape models and PCA-based

global shapes, as in [75]. The CLM model was extended with full voting from a random forest in [76] or by probabilistic interpretation for optimization of the shape parameters in [77].

Another direction assumes independent point localizer followed by aggregation of location. This is a rich class of methods, including some of the most recent and accurate solutions. Thus Valstar et al. [78] complemented the SVM regressed feature point location with conditional MRF to keep the estimates globally consistent. Belhumeur et al. [28] proposed a Bayesian model combining the outputs of the local detectors (formed by SVM classifier trained over SIFT features) with a consensus of non-parametric global models for part locations; Zhu and Ramanan [79] relied on a connected set of local templates described with HOG. Dantone et al. [67] constructed multiple random forrest that use image patches as input conditioned by the head pose to estimate in real time a set of face landmarks. Yu et al. [80] used 3D deformable shape models to iteratively fit over 2D data and identify without respect to pose the facial landmarks positions.

Martinez et al. [81] trained SVM regressors with selected Local Binary Patterns to formulate initial predictions that are further iteratively aggregated for improving accuracy. In [68], the relation between fiducial points is encoded directly in the localization system, which is based on deep convolutional networks. Florea et al. [30] used a multi stage method for precisely fitting the landmarks in the eye are. In the next subsection we will detail this method as it is focussed on the eyes.

6.1 Multi-level Eye Landmark Localization

6.1.1 Position and Intensity Priors

During training, a bounding box is computed on the range of each feature location within the region found by the face detector, as in the classical CLM [74]. For each candidate landmark, its *position prior* ic constructed as a probability map spanning its region of interest (ROI), such that each position is given the likelihood of being close to the ROI center. This is in fact the two-dimensional histogram of the positions.

The position prior is further denoted as $p_1(i,j)$ and an example for the left eye outer corner position prior map is shown in Fig. 8a.

While for eye center localization it is common to investigate only the darkest pixels [82], this idea may be extended to all landmarks with appropriate conditions (such as considering that the inner eye corner is darker than most of its neighboring pixels while the upper limit of the eye is brighter than most of its neighboring pixels). Thus, for each candidate landmark, its *intensity prior* is constructed as a probability map spanning its region of interest, such that each position is given the probability of occurrence of its corresponding graylevel within the ROI. This prior is denoted as $p_2(i,j)$ and is exemplified in Fig. 8b for the case of the outer corner of the left eye.

Table 6 Face landmarks localization methods in remote imaging

Method	Face detection	Features	Learning	Public code
Cootes et al. [69]—ASM	No	Edges	Yes	Yes
Cootes et al. [70]—AAM	No	Edges and pixels	Yes	Yes
Leung et al. [71]—EGM	No	VPF	No	No
Cristinacce et al. [74]—CLM	Yes	Edge	Yes	Yes
Tresadern et al. [75]	Yes	Edge+pixels	MRF	No (AAM)
Saragih et al. [77]—PAAM	Yes	Edge+pixels	Yes	Yes
Milborrow and Nichols [72]—STASM	Yes	Edges	Yes	Yes
Valstar et al. [78]—Borman	Yes	Image pixels	SVR+MRF	Yes
Zhu and Ramanan [79]	Yes	pHoG	Bagged trees	Yes
Belhumeur et al. [28]	Yes	SIFT	SVM	No
Dantone et al. [67]	Yes	Pixels	RF	No
Martinez et al. [81]—Lear	Yes	Gradient	SVR	Yes
Sun et al. [68]	Yes	Pixels	CNN	Yes
Yu et al. [80]	Yes	HoG	EM	Yes
Florea et al. [30]	Yes	IPF+EPF	MLP	Yes

The learning column signals that method parameters are regressed on a training database; we nominate when a specific machine learning system is employed

6.1.2 Template Matching

It is typical in the landmark localization [28, 30, 67] that the bulk of the search to be performed by a template matching procedure. In such a case, the problem is to determine for each pixel, within a reasonable neighborhood of an initial landmark, the probability of that pixel being the true landmark. A typical procedure implies considering consecutive sub-windows in the search area. Each sub-window is centered in the investigated location and it is represented in a descriptor space; a machine learning system is then trained to determine how likely is for the sub-window to be centered on the true position of the landmark.

Similar with the iris center localization procedure, given a rectangular sub-windows centered in the truth landmark position and a Multi-Layer Perceptron (MLP) is trained with the integral and edge projections within that image sub-window. The definition of the IPF and EPF were presented in Sect. 5.

To ensure a better robustness to illumination variation, each of the projections should be independently normalized to the $[-128, 127]$ range. Each sub-window, for

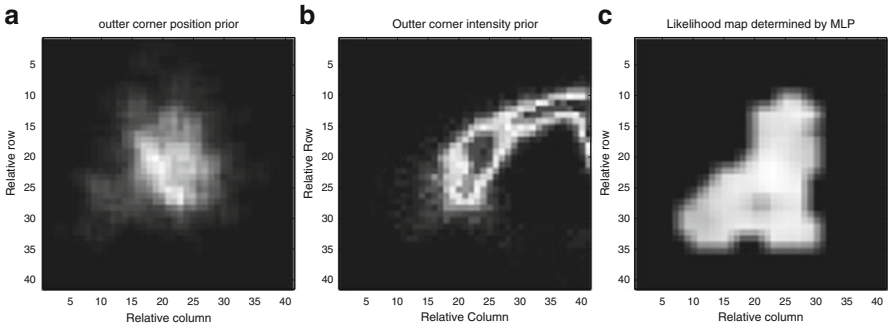

Fig. 8 (**a**) Position prior map p_1 for left eye outer corner, gathered from all training images; (**b**) Intensity prior map p_2 for left eye outer corner; (**c**) Likelihood Map determined by the MLP given the actual data, p_3 for a specific example of left eye outer corner. Image taken from [30]. Copyright of the authors

each landmark, is thus described by a feature vector formed by the concatenation of the four projections: horizontal/vertical, integral/edge. Thus, the length of the features is $2 \cdot W + 2 \cdot H = 120$.

A MLP is separately trained for each landmark (with feature vectors from the sub-windows as input) to output the Euclidian distance between the center of the input sub-window and the true landmark position (thus performing regression). Thus, given an original image and a landmark to localize, the complement of the output of the MLP, $c(i,j)$ is the landmark likelihood map $p_3(i,j)$, with $p_3(i,j) = 1 - c(i,j)$. A MLP with one hidden layer with 30 neurons is good choice [30]. A typical p_3 map is showed in Fig. 8c for the location of the outer left eye corner.

Alternatively, Vranceanu et al. [10] feed the concatenated image projection into a logistic regressor to determine only one dimension of the landmarks in order to obtain the bounding box of the eye.

6.1.3 Initial Landmarks Estimates

To fuse the information from intensity and position priors with template matching data, the final probability map is:

$$p_{123}(i,j) = \alpha \cdot p_1(i,j) + \beta \cdot p_2(i,j) + \gamma \cdot p_3(i,j) \tag{7}$$

where α, β and γ weight the confidence in each type of map; their values are to be deduced on the training database independently for each landmark, for each eye. Florea et al. [30] suggested the following empirical values $\alpha = \beta = 0.25$, $\gamma = 0.5$. The *weighted center of mass* of the p_{123} probability map for each landmark represents an accurate initial estimate of that landmark.

6.1.4 Shape Constraints

Even if gaze varies across the tested cases, the relative position of the eye remains stable enough, thus the eye socket shape may be further constrained. To achieve this, either the Constrained Local Model [74] or the Global Models Locally Constrained [30] are at hand. The second solution, instead of gathering local constrains and maximizing the global output, for each landmark (locally), iteratively uses the global shape to construct a local constraint. As eye landmarks are investigated in terms of gaze variation, the iris center position has very large variations, thus it is not included in the shape. For each face there will be two shapes, one for each eye, refined independently; a shape contains only four points, namely the limits of the eye socket.

Distinct shapes for the left and the right eye are envisaged. In ASM [69], given s sets of points $\{\mathbf{x}_i\}$, after alignment (centering all shapes in origin and aligning them by a specific rule), one may compute the mean shape $\overline{\mathbf{x}}$, and the projection matrix on a PCA-reduced space $\mathbf{P}$. Any of the input shapes is given as:

$$\mathbf{x} \approx \overline{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{8}$$

The $\mathbf{P}$ matrix is made by the first t eigenvectors of the covariance matrix of all the shapes $\{\mathbf{x}_i\}$ in the training database. Thus, the transformation vector is found as:

$$\mathbf{b} = \mathbf{P}^T (\mathbf{x} - \overline{\mathbf{x}}) \tag{9}$$

Given a large number of shapes $\{\mathbf{x}_i\}$, one may compute a histogram for the transformation vector, $\mathbf{b}$. More precisely, for each shape, $\{\mathbf{x}_i\}$ in the database, using Eq. (9), one will obtain a t-dimensional transformation vector $\mathbf{b}_i$.

Since the dimensionality reduction is performed by means of decorrelation (i.e. PCA), and the histogram is Gaussian-like, the histogram of $\mathbf{b}$ may be approximated with t-dimensional independent Gaussian distributions with $\mathbf{0}$-mean and $\Sigma = diag(\lambda_i)$ covariance matrix, where λ_i is the shapes eigenvalue on the dimension i, in the original shape space.

Given an initial shape $\mathbf{x}'$ and keeping all the landmarks fixed with the exception of the current one (which is to be improved), if one assumes that this landmark is at (i,j) location, a new value, $\mathbf{b}'$ will be obtained. Given the original histogram of $\mathbf{b}$, the newly obtained value, $\mathbf{b}'$, is back-propagated into a probability value, which will be named $p_s(i,j)|\mathbf{x}'$.

Thus, for each location in the searched area and for each of the landmarks, $p_s(i,j)|\mathbf{x}$ the probability to have the landmark at position (i,j) given the remainder of the shape $\mathbf{x}$ to its initial position is computed. Practical choices are $t = 2$ and the alignment of the shapes such that to have a horizontal outer-inner axis [30].

Taking into account that at any step, for each landmark, an estimate of the shape is available by computing the weighted center of mass from Eq. (7) and observing that some landmarks (e.g. eye outer corners) are more reliable than others (upper and lower boundaries), by keeping all points fixed with the exception of the

least reliable, the likelihood of various positions for the current landmark is built. The order of the landmarks is with respect to their reliability. Then, the procedure is repeated iteratively for each landmark N_{it} times (typically $N_{it} = 3$).

The *final landmark position* is taken as the weighted center of mass of the convex combination, $p_F(i,j)$, between the initial stages and the shape fitting likelihood:

$$p_F(i,j) = \delta \cdot p_{123}(i,j) + (1-\delta) \cdot p_S(i,j)|\mathbf{x}, \tag{10}$$

where δ= 0.75 was experimentally chosen.

6.2 Evaluation

6.2.1 Evaluation Procedure

For multiple landmarks the proximity measure [74] m_e is the common measure. It is computed as:

$$m_e = \frac{1}{t \cdot D_{eye}} \sum_{i=1}^{t} \varepsilon_i \tag{11}$$

In Eq. (11), ε_i are the point to point errors for each individual landmark location and t is the number of feature points searched (ten points in the current considered case, thus the measure being further referred as m_{e10}). Again, the interest is in obtaining a higher accuracy for low threshold values.

6.2.2 Eye Landmarks Localization

For this specific task, we compare the performance of the methods from [30, 78–81]. The accuracies are shown in Figs. 9 and 10. We note that the method from [81] does not locate the iris centers, thus in this case we take into account only eight landmarks and all methods are trained outside the tested databases.

The results show that mostly the best performance is achieved by Florea et al. [30]. The largest difference is on the Eye-Chimera database, where other methods, that are general face landmarking methods, significantly under-perform due to the significant variation of gaze on both horizontal and vertical directions.

7 Gaze Direction Recognition

In computer vision, extensive research was done in the field of detecting the direction of gaze [83, 84], by means of so-called eye trackers. Usually, *eye tracking* technology relies on measuring reflections of the infrared/near-infrared light on the

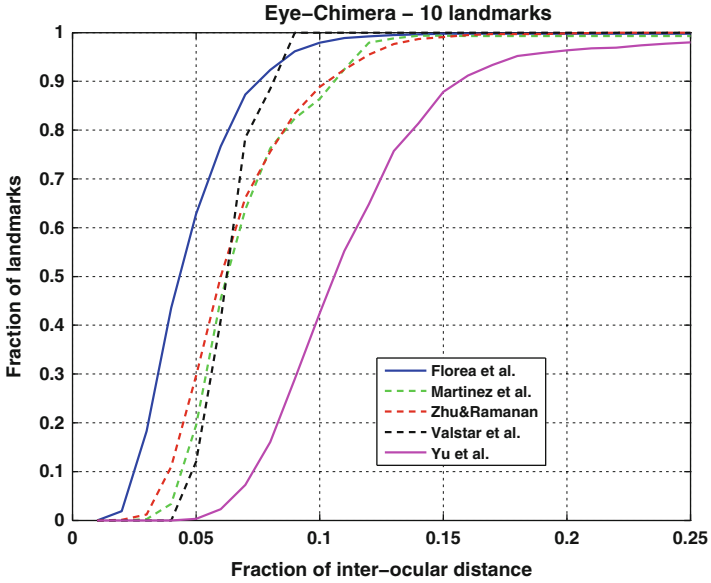

Fig. 9 Eye landmarks localization performance on the Eye-Chimera database of several methods: Valstar et al. [78], Zhu-Ramanan [79], Martinez et al. [81], Yu et al. [80] and Florea at al. [30] on the Eye-Chimera, HPEG and ULM databases

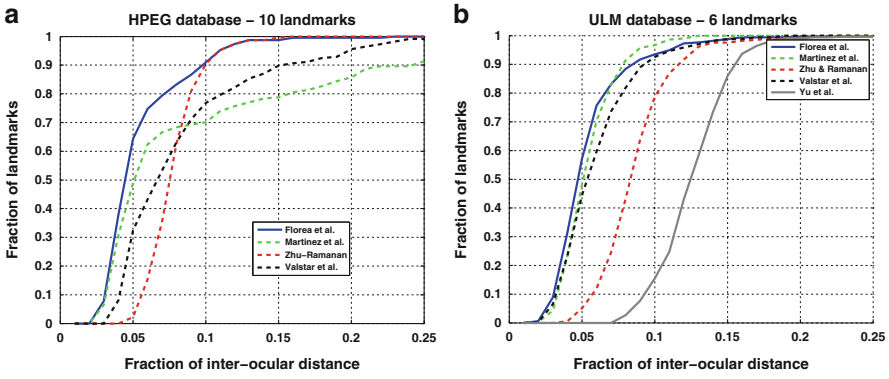

Fig. 10 Eye landmarks localization performance of several methods: Valstar et al. [78], Zhu-Ramanan [79], Martinez et al. [81], Yu et al. [80] and Florea at al. [30] on the HPEG (**a**) and ULM (**b**) databases

eye: the first Purkinje image (P1) is the reflection from the outer surface of the cornea, while the fourth (P4) is the reflection from the inner surface of the lens; these two images form a vector that is used to compute the angular orientation of the eye, in the so-called "dual Purkinje" method [84]. An example of such an eye

tracking systems is, for instance, found in the work of Yoo and Chung [85] that relies on two cameras and four infrared sources to achieve high accuracy.

A method relying on a head mounted device with visible spectrum illumination is found in the work of Pires et al. [86] who extracted the iris contour followed by a Hough transform to detect the iris center and, respectively, by the localization of the eye corners contours; the head-mounted device, permits high resolution for the eye image thus extending the range of a wearable eye-tracking for sport. Due to reasons detailed in the previous subsection, we will avoid both the IR-based and, respectively, the head mounted category of solutions.

The alternative is to develop non-intrusive, low-cost techniques that directly measure the gaze direction, such as the approaches used in [34, 82, 87–89]. Wang et al. [82] selected recursive nonparametric discriminant features from a topographic image feature pool to train an Adaboost that locates the eye direction. Hansen and Pece [87] modelled the eye contour as an ellipse and use Expectation-Maximization to locally fit the actual contour. Cadavid et al. [88] trained a Support Vector Machine with spectrally projected eye region images to identify the direction of gaze. Heyman et al. [89] used correlation-based methods (more precisely the so-called Canonical Correlation Analysis) to match the new eye data with marked data and to find the direction of gaze. Wolf et al. [34] applied the eye landmark localizer provided by Everingham and Zisserman [90] to initialize the fit of the eye double parabola model. We note that all these methods first localize eye landmarks and subsequently analyze the identified eye regions.

Florea et al. [30] used the eye landmarks and the interior of the eye shape to directly estimate the gaze into seven possible directions. The same set of directions, which matches the NLP identified direction, was also searched by the methods form [10, 35]. Radlak et al. [35] employed Hybrid Image Projections functions as defined by Zhou and Geng [47] followed by either a SVM or a random forest. Vranceanu et al. [10] complemented the projection based information with the geometrical location of four segments extracted from the eye region.

Also in the direction of gaze recognition in terms of EAC-NLP we note the work of Diamantopoulos [91] which used a head mounted device. Taking into account that Laeng and Teodorescu [25] showed that, even for non-visual tasks, voluntary control affects eye movement, we may conclude that they explore the theme only from a computer vision perspective, without direct practical applications. Furthermore, the head mounted device has the un-realistic advantage of being closer to the eye and, thus, of having access to higher resolution and more precisely located eye image patches. For images with high resolution, the method implied by Pires et al. [86] (iris contour detection followed by Hough transform for circles) works very well. However, for the lower resolution images, which are associated with remote

Fig. 11 Separation (segmentation) of the eye components using K-Means

acquisition devices, the contours in the eye region are no longer sharp and the accumulation in the Hough transform, very often, points to wrong locations (eye socket or brow).

Once the eye bounding box has been delimited by means of the eye landmarks, the specific EAC is retrievable by analyzing the positions of different eye components. The natural choice is to analyze the position of the iris inside the eye bounding box. Yet as eye localizers are imperfect [10], especially when challenged by gaze variation, to improve the accuracy, a separation of the components of the eye within the bounding box and use their relative position as indicators of the EAC, improves the landmark achievable performance.

7.1 Separating the Eye Components

7.1.1 Segmentation

The segmentation is a well known problem and many solutions have been proposed through the years. For the specific problem of the eye components separation for gaze direction estimation, it should be required that the segmentation of the eye components allows a good EAC recognition in a reasonable amount of time. According to the tests performed by Vranceanu et al. [10], the best compromise is achievable using a *K-Means* segmentation (Fig. 11).

7.1.2 Post-processing and Classification

The eye area, given by the detected landmarks is normalized to a standard size and position. The coordinates of each of the resulting eye components' centers of mass in the normalized bounding box and the average luminance are used as features describing the eye. To improve the region separation resulted from segmentation, the integral projections functions (IPF) complements the bounding box in a variation of the Appearance Models. Therefore, for a more general description inside the bounding box, the vertical and horizontal integral and edge projections are added as features for the classifier, next to the segmented regions center of mass and landmarks.

Table 7 Influence of the number of regions on the EAC recognition rate, RR (%), when simple K-Means segmentation is used (without additional projection information). We have marked with bold letters the best results for each EAC case

Regions no.	C = 2	C = 3	C = 4
RR (%) (7 EAC classes)	50.73	62.08	**64.56**
RR (%) (3 EAC classes)	62.77	77.28	**82.32**

In order to recognize the seven EAC classes, the feature vector is composed by:

- $3 \times C$ elements (which correspond to the centers of mass coordinates and the average luminance for each of the C regions);
- the concatenated horizontal and vertical integral and edge projections;
- landmarks

Various classification methods are considered and, as the number of features is small, the Logistic Classification [92] was found [10] to give good results.

7.2 Results

While computing the EAC recognition rate, two scenarios are evaluated: the seven-case and the three-case. The complete seven EACs set contains all the situations described by the NLP theory and presented in Fig. 1. Additionally, as the vertical direction of gaze is harder to identify [84], one may consider only three cases assigned to: looking forward (center), looking left and looking right; in terms of EACs, here, the focus is on the type of mental activity, while the representational systems are merged together. This particular test is relevant for the interview scenario, where, when given a query, if the subject remembers the solution, it indicates experience in the field, while if he/she constructs the answers, it points to creativity.

7.2.1 Segmentation Influence

Eye region segmentation is an important step for the accurate recognition of the EAC and a critical aspect is the number of classes, C, in which the input data should be divided. As can be seen in Table 7 a larger number of regions increases the EAC recognition rate; therefore, the eye space should be in fact divided in four regions corresponding to all the eye components present in the bounding box: the iris and the sclera, the eyelashes and the surrounding skin area.

Table 8 Individual recognition rate for each EAC case on the still Eye Chimera database

VD	VR	VC	AR	AC	ID	K
88.62	74.66	80.00	71.83	61.43	71.76	80.43

The acronyms for the EACs are presented in Fig. 1

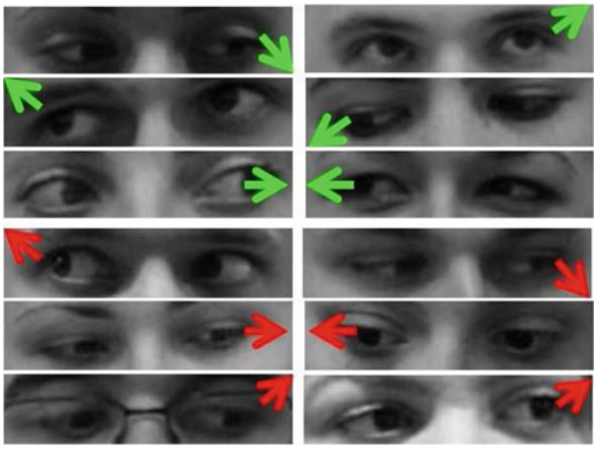

Fig. 12 Automatic recognition examples: correct (*green arrow*) and false (*red arrow*). Image taken from [10]. Copyright by Springer

7.2.2 EAC Recognition

The final solution that gives the best results for EAC recognition consists of using iris-oriented K-Means segmentation together with projection information and landmarks. Vrancenu et al. [10] showed that the extra use of the integral projections in the feature vector leads to an improvement of approx. +5 % in the recognition rate.

Furthermore, the recognition rates for each individual EAC are presented in Table 8. It can be seen that a higher confusion rate appears vertically, between eyes looking to the same side. In a NLP interpretation, this corresponds to a better separability between the internal activities and a poorer separability between representational systems. Visual examples of correct and false recognitions are shown in Fig. 12 and it should be noted that even for a human observer it is difficult, in some cases, to correctly classify the direction of gaze.

As said, one intuitive way to recognize the EAC is to use the coordinates of eye fiducial points. Thus, we consider as relevant several foremost such methods. First, the BoRMaN algorithm [78] can be employed for detecting the eye bounding box and a good iris center localization can be obtained using the maximum isophote algorithm presented in [43]. The eye landmarking method proposed in [30] also provides the required points for such an analysis. Finally, using the landmarking technique proposed in [79], out of a larger number of detected fiducial points, the points delimiting the eye and the iris center can be selected for the EAC analysis.

Comparative results are presented in Table 9. As one can see the best results are retrieved by the methods from [10]. There, the refined bounding box (or more precisely its height) is necessary to differentiate between looking down and looking

Table 9 Recognition rate (%) on the still Eye Chimera database for the three EAC cases scenario (when the focus is on the type of mental activity) and for the seven EAC cases scenario (the complete EAC set)

Bounding box method	Iris detection method	RR (%)	
		7 classes	3 classes
Manual	Manual	73.98	94.52
Manual	Valenti [43]	32.30	36.40
BoRMaN [78]	Valenti [43]	32.00	33.12
Zhu [79]	Zhu [79]	39.21	45.57
Zhu [79]	Sun [68]	47.25	68.15
Florea [30]	Florea [30]	48.64	78.57
Radlak [35]	Kawulok [57]	49.47	77.00
Vranceanu [10]	Vranceanu [10]	77.54	89.92

Table 10 EAC recognition rates (%) of various solutions, when information from both eyes is used

Method	RR (%)	[43] + [78]	[79]	[10] (1 eye)	[10] (2 eyes)
Still Eye	7 classes	39.83	43.29	77.54	83.08
Chimera	3 classes	55.73	63.01	89.92	95.21

elsewhere, while the iris position inside it, actually defines the direction of gaze. The pre-processing step removes the eye-lashes as it interferes with the iris separation from the rest of the eye components. The integral projections functions added in the post-processing step supplement the information used by the classifier for the EACs recognition. Due to these facts, when all seven EAC classes are considered, the algorithm from [10] surpasses the upper limit of a point-based analysis, which is obtained when only the five manual markings are used.

Both Eyes Information In order to further improve the detection rate, information from both eyes can be concatenated in the feature vectors. Vranceanu et al. [10] showed (also in Table 10) that this leads to an improvement of approx. $+6\,\%$ in the detection rate, in both the three cases scenario as well as for the complete EAC set.

Other Databases For a thorough evaluation, we report results on other databases, where the eye cues are partially represented. Since these databases are not designed for an EAC-NLP application, each poses different challenges and are somewhat incomplete from the EAC point of view. The HPEG database does contain all seven EACs, but in a small number, the UUlm database contains only three of the eye cues: Visual Defocus (VD—looking straight), Auditory Remember (AR—looking center-right) and Auditory Constructed (AC—looking center-left). The PUT database contains all seven cues but disproportionably represented. Furthermore, all three databases have a considerable head pose variation.

Comparative results can be seen in Table 11. Although the results vary considerably across databases, the eye components based method [10] offers the best results

Table 11 EAC recognition rate (%) computed on ULM, HPEG and PUT databases

Method	RR (%)	[43] + [78]	[79]	[10] (1 eye)	[10] (2 eyes)
HPEG	7 classes	18.52	31.82	43.71	50.00
	3 classes	29.34	49.15	68.54	75.17
ULM	7 classes	40.63	29.37	23.57	29.29
	3 classes	41.28	44.39	70.35	80.89
PUT	7 classes	11.01	31.11	55.68	62.18
	3 classes	13.18	44.11	63.76	71.43

in all scenarios. While testing on the ULM database, we looked also for seven cases, and any output different from the correct one is marked as an error; to make the test more relevant to the work, we ignored that, for this specific test, only three possible outputs could exist.

8 Discussion and Conclusions

The purpose of this research was to discuss some potential alternatives for an automatic solutions that recognize the direction of gaze in images that contain a frontal face. Such a solution would facilitate advances in areas such as non-conventional teaching, gaming industry and, at last but not at least, for the Eye Accessing Cues.

Eye Accessing Cues are a hypothesis from the Neuro-Lingvistic-theory and they have been only partially validated; more precisely, it has been found that correlation greater than random chance is possible. These results, which are in line with most prior art on the topic, in fact, motivates large scale intensive tests to find the truth behind. If validated, the direction of gaze may be used for better understanding of the mental patterns of a person. We nominated two applications: on-line interviews and interactive presentation.

From a computer vision point of view, while several efficient approaches were investigated, the best results for the recognition of the direction of gaze were reported by Vranceanu et al. [10]. There, consecutively, the face square, iris center, face landmarks and eye components are extracted.

The results on specifically built Eye Chimera database show that the method from Vranceanu et al. [10] surpasses in accuracy some of the most efficient state of the art methods for detecting landmarks and implicitly eye points. It was also shown that it surpasses the pure eye landmarking techniques proving that a region-based solution provides better accuracy than a point-only based approach. These findings were confirmed on other databases too.

Finally, using the video (sequence) part of the Eye-Chimera database, it was proven that, when dealing with sequences, the recognition rate can be increased

by considering the temporal redundancy and the correlation between consecutive frames. This observation, together with the low computational cost, offers potential for an implementation where eye cues are detected, tracked and interpreted for mass applications.

The current maximum reported performance reached 77.54 % accuracy on the Eye-Chimera database, which is assumed to be the most relevant for the EAC-NLP theme. This means that in average 1 out of 4 cases is mis-labelled. Thus, if the goal is to validate the hypothesis behind the EAC-NLP than the existing solution is reliable only for automatic initialization of the annotations followed by manual verification. Yet even for this case automatic solution brings speed-ups up to 10×. If the EAC-NLP hypothesis are validated, than the described application may work in real cases as an estimator, with the observation that it needs to be applied independently in consecutive cases, and the overall conclusion needs to be manually validated.

Some additional issues remain for further investigation and development. First, the hypothesis of the Eye Accessing Cues needs to be fully confirmed and most likely bounded. Secondly the EACs are related to non-visual tasks and, therefore, separation between visual and non-visual tasks is required. In normal conditions, the difference between voluntary eye movements (as for seeing something) and involuntary ones (as part of non-verbal communication) is retrievable by the analysis of duration and amplitude [83] as non-visual movements are shorter and with smaller amplitude. However, in both visual memory related task [25] as in the NLP theory, the actual difference between visual and non-visual tasks is achieved by integrating additional information about the person specific activities. More precisely, the Eye Accessing Cues are expected to appear following specific predicates (such as immediately after a question marked by “How?” or “Why?”). Thus, for a complete autonomous solution, the labels required for segmenting the video in visual and non-visual tasks should be inferred from an analysis of the audio channel, that should complement the visual data. To the moment, a completely functional system would be the one where the trainer/interviewer marks the beginning and the end of the non-visual period, as he is aware of the nature of communication.

Concluding, the gaze direction estimation from passive remotely acquired image is an interesting area with many un-explored development directions.

Acknowledgements This work was partially supported by the Romanian Sectoral Operational Programme Human Resources Development 2007–2013 through the European Social Fund Financial Agreements POSDRU/159/1.5/S/134398 (Knowledge).

References

1. R. Bandler, J. Grinder, *Frogs into Princes: Neuro Linguistic Programming* (Real People Press, Moab, 1979)
2. P. Tsiamyrtzis, J. Dowdall, D. Shastri, I.T. Pavlidis, M.G. Frank, P. Ekman, Imaging facial physiology for the detection of deceit. Int. J. Comput. Vis. **71**, 197–214 (2007)

3. A.B. Ashraf, S. Lucey, J.F. Cohn, T. Chen, Z. Ambadar, K.M. Prkachin, P. Solomon, The painful face – pain expression recognition using active appearance models. Image Vis. Comput. **27**, 1788–1796 (2009)
4. C. Florea, L. Florea, C. Vertan, Learning pain from emotion: transferred hot data representation for pain intensity estimation, in *Proceedings of European Conference on Computer Vision Workshop on ACVR* (2014)
5. D.S. Messinger, M.H. Mahoor, S.M. Chow, J. Cohn, Automated measurement of facial expression in infant-mother interaction: a pilot study. Infancy **14**(3), 285–305 (2009)
6. D. McDuff, R.E. Kaliouby, R. Picard, Predicting online media effectiveness based on smile responses gathered over the internet, in *IEEE Face and Gesture* (2013), pp. 1–8
7. J. Rehg, G. Abowd, A. Rozga et al., Decoding children's social behavior, in *Proceedings of Computer Vision and Pattern Recognition* (2013), pp. 3414–3421
8. J.F. Cohn, F. De la Torre, Automated face analysis for affective computing, in *The Oxford Handbook of Affective Computing* (Oxford University Press, Oxford, 2014)
9. A. Frischen, A.P. Bayliss, S.P. Tipper, Gaze cueing of attention. Psychol. Bull. **133**, 694–724 (2007)
10. R. Vranceanu, C. Florea, L. Florea, C. Vertan, Gaze direction estimation by component separation for recognition of eye accessing cues. Mach. Vis. Appl. **26**(2–3), 267–278 (2015)
11. W. James, *The Principles of Psychology* (Harvard University Press, Cambridge, 1890)
12. L. Nummenmaa, A. Calder, Neural mechanisms of social attention. Trends Cogn. Sci. **13**, 135–43 (2009)
13. S. Liversedge, J. Findlay, Saccadic eye movements and cognition. Trends Cogn. Sci. **4**(1), 6–14 (2000)
14. R. Adams, R.E. Kleck, Effects of direct and averted gaze on the perception of facially communicated emotion. Emotion **5**, 3–11 (2005)
15. H. Joseph, K. Nation, S.P. Liversedge, Using eye movements to investigate word frequency effects in children's sentence reading. Sch. Psychol. Rev. **42**, 207–222 (2013)
16. A. Godfroid, F. Boers, A. Housen, An eye for words: gauging the role of attention in incidental l2 vocabulary acquisition by means of eye-tracking. Stud. Second Lang. Acquis. **35**, 483–517 (2013)
17. K. Rayner, T.J. Slattery, D. Drieghe, S.P. Liversedge, Eye movements and word skipping during reading: effects of word length and predictability. J. Exp. Psychol. Hum. Percept. Perform. **37**, 514–528 (2011)
18. K. Rayner, B.R. Foorman, C.A. Perfetti, D. Pesetsky, M.S. Seidenberg, How psychological science informs the teaching of reading. Psychol. Sci. Public Interest **2**, 31–74 (2001)
19. M.M. Chun, Contextual cueing of visual attention. Trends Cogn. Sci. **4**, 170–178 (2000)
20. A. Bulling, T. Zander, Cognition-aware computing. IEEE Trans. Pervasive Comput. **13**, 80–83 (2014)
21. B. Meijering, H. van Rijn, N.A. Taatgen, R. Verbrugge, What eye movements can tell about theory of mind in a strategic game. PLoS One **7**(9) (2012) doi:10.1371/journal.pone.0045961
22. K. Krejtz, C. Biele, D. Chrzastowski, A. Kopacz, A. Niedzielska, P. Toczyski, A. Duchowski, Gaze-controlled gaming: immersive and difficult but not cognitively overloading, in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (2014), pp. 1123–1129
23. J. Sturt, S. Ali, W. Robertson, D. Metcalfe, A. Grove, C. Bourne, C. Bridle, Neurolinguistic programming: systematic review of the effects on health outcomes. Br. J. Gen. Pract. **62**, 757–764 (2012)
24. R. Vranceanu, C. Florea, L. Florea, C. Vertan, NLP EAC recognition by component separation in the eye region, in *Proceedings of Computer Analysis and Image Processing* (2013), pp. 225–232
25. B. Laeng, D.S. Teodorescu, Eye scanpaths during visual imagery reenact those of perception of the same visual scene. Cogn. Sci. **26**, 207–231 (2002)
26. T. Kanade, J.F. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in *IEEE Face and Gesture* (2000), pp. 46–53

27. K. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 684–698 (2005)
28. P. Belhumeur, D. Jacobs, D. Kriegman, N. Kumar, Localizing parts of faces using a consensus of exemplars, in *Proceedings of Computer Vision and Pattern Recognition* (2011), pp. 545–552
29. G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, 2007
30. L. Florea, C. Florea, R. Vranceanu, C. Vertan, Can your eyes tell me how you think? A gaze directed estimation of the mental activity, in *Proceedings of British Machine Vision Conference* (2013)
31. S. Asteriadis, D. Soufleros, K. Karpouzis, S. Kollias, A natural head pose and eye gaze dataset, in ACM Workshop on Affective Interaction in Natural Environments (2009), pp. 1–4
32. U. Weidenbacher, G. Layher, P. Strauss, H. Neumann, A comprehensive head pose and gaze database, in *IET International Conference on Intelligent Environments* (2007), pp. 455–458
33. A. Kasiński, A. Florek, A. Schmidt, The PUT face database. Image Process. Commun. **13**, 59–64 (2008)
34. L. Wolf, Z. Freund, S. Avidan, An eye for an eye: a single camera gaze-replacement method, in *Proceedings of Computer Vision and Pattern Recognition* (2010), pp. 817–824
35. K. Radlak, M. Kawulok, B. Smolka, N. Radlak, Gaze direction estimation from static images, in *Proceedings of IEEE Multimedia Signal Processing* (2014), pp. 1–4
36. P. Viola, M. Jones, Robust real-time face detection. Int. J. Comput. Vis. **57**, 137–154 (2004)
37. M. Mathias, R. Benenson, M. Pedersoli, L.V. Gool, Face detection without bells and whistles, in *Proceedings of the European Conference on Computer Vision*, vol. 8692 (2014), pp. 720–735
38. P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models. Pattern Recogn. Lett. **19**, 899–906 (2010)
39. F. Song, X. Tan, S. Chen, Z. Zhoub, A literature survey on robust and efficient eye localization in real-life scenarios. Br. J. Gen. Pract. **46**, 3157–3173 (2013)
40. M. Hamouz, J. Kittlerand, J.K. Kamarainen, P. Paalanen, H. Kalviainen, J. Matas, Feature-based affine-invariant localization of faces. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 643–660 (2005)
41. S. Asteriadis, N. Nikolaidis, I. Pitas, Facial feature detection using distance vector fields. Pattern Recogn. **42**, 1388–1398 (2009)
42. J. Wu, Z.H. Zhou, Efficient face candidates selector for face detection. Pattern Recogn. **36**, 1175–1186 (2003)
43. R. Valenti, T. Gevers, Accurate eye center location and tracking using isophote curvature, in *Proceedings of Computer Vision and Pattern Recognition* (2008), pp. 1–8
44. O. Jesorsky, K. Kirchberg, R. Frischholz, Robust face detection using the Hausdorff distance, in *Proceedings of International Conference on Audio- and Video-Based Biometric Person Authentication* (2001), pp. 90–95
45. T. Kanade, Picture processing by computer complex and recognition of human faces. Technical Report, Kyoto University, Department of Information Science, 1973
46. G.C. Feng, P.C. Yuen, Variance projection function and its application to eye detection for human face recognition. Pattern Recogn. Lett. **19**, 899–906 (1998)
47. Z. Zhou, Projection functions for eye detection. Pattern Recogn. **37**, 1049–1056 (2004)
48. M. Turkan, M. Pardas, A.E. Cetin, Edge projections for eye localization. Opt. Eng. **47**, 047–054 (2008)
49. M. Verjak, M. Stephancic, An anthropological model for automatic recognition of the male human face. Ann. Hum. Biol. **21**, 363–380 (1994)
50. D. Cristinacce, T. Cootes, I. Scott, A multi-stage approach to facial feature detection, in *Proceedings of British Machine Vision Conference* (2004), pp. 277–286
51. P. Campadelli, R. Lanzarotti, G. Lipori, Precise eye localization through a general-to-specific model definition, in *Proceedings of British Machine Vision Conference*, **I**, 187–196 (2006)

52. Z. Niu, S. Shan, S. Yan, X. Chen, W. Gao, 2D cascaded adaboost for eye localization, in *Proceedings of International Conference of Pattern Recognition* (2006), pp. 1216–1219
53. S. Kim, S.T. Chung, S. Jung, D. Oh, J. Kim, S. Cho, World Academy of Science, Engineering and Technology, in *WASET*, vol. 21 (World Academy of Science, Engineering and Technology, 2007), pp. 483–487
54. M. Asadifard, J. Shanbezadeh, Automatic adaptive center pupil detection using face detection and CDF analysis, in *Proceedings of International Multimedia Conference of Engineers and Computer Scientist* (2010), pp. 130–133
55. L. Ding, A.M. Martinez, Features versus context: an approach for precise and detailed detection and delineation of faces and facial features. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 2022–2038 (2010)
56. F. Timm, E. Barth, Accurate eye centre localisation by means of gradients, in *Proceedings of International Conference on Computer Theory and Applications* (2011), pp. 125–130
57. M. Kawulok, J. Szymanek, Precise multi-level face detector for advanced analysis of facial images. IET Image Process. **6**, 95–103 (2012)
58. C. Florea, L. Florea, C. Vertan, Robust eye centers localization with zero-crossing encoded image projections. Pattern Anal. Applic. 1–17 (2015), DOI:10.1007/s10044-015-0479-x, http://dx.doi.org/10.1007/s10044-015-0479-x
59. R. Valenti, T. Gevers, Accurate eye center location through invariant isocentric patterns. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 1785–1798 (2012)
60. H.C. Becker, W.J. Nettleton, P.H. Meyers, J.W. Sweeney, C.M. Nice, Digital computer determination of a medical diagnostic index directly from chest X-ray images. IEEE Trans. Biomed. Eng. **11**, 62–72 (1964)
61. F. Crow, Summed-area tables for texture mapping. Proc. SIGGRAPH **18**, 207–212 (1984)
62. G.E. Blelloch, Prefix sums and their applications. synthesis of parallel algorithms. Technical report, University of Massachusetts, 1990
63. R.A. King, T.C. Phipps, Shannon, TESPAR and approximation strategies. Comput. Secur. **18**, 445–453 (1999)
64. X. Chen, H. Wu, X. Jin, Q. Zhao, Face illumination manipulation using a single reference image by adaptive layer decomposition. IEEE Trans. Image Processing **22**(11), 4249–4259 (2013)
65. B. Kroon, A. Hanjalic, S.M. Maas, Eye localization for face matching: is it always useful and under what conditions, in *Proceedings of International Conference on Content-Based Image and Video Retrieval* (2008), pp. 379–387
66. M. Ciesla, P. Koziol, Eye pupil location using webcam. CoRR, (2012) http://arxiv.org/abs/1202.6517
67. M. Dantone, J. Gall, G. Fanelli, L.V. Gool, Real-time facial feature detection using conditional regression forests, in *Proceedings of Computer Vision and Pattern Recognition* (2012), pp. 2578–2585
68. Y. Sun, X. Wang, X. Tang, Deep convolutional network cascade for facial point detection, in *Proceedings of Computer Vision and Pattern Recognition* (2013), pp. 3476–3483
69. T. Cootes, C. Taylor, D. Cooper, J. Graham, Active shape models - their training and application. Comput. Vis. Image Underst. **61**, 38–59 (1995)
70. T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 681–685 (2001)
71. T. Leung, M. Burl, P. Perona, Finding faces in cluttered scenes using random labeled graph matching, in *Proceedings of International Conference on Computer Vision* (1995), pp. 637–644
72. S. Milborrow, F. Nicolls, Locating facial features with an extended active shape model, in *Proceedings of European Conference on Computer Vision* (2008), pp. 504–513
73. V. Le, J. Brandt, Z. Lin, L. Bourdev, T.S. Huang, Interactive facial feature localization, in *Proceedings of European Conference on Computer Vision* (2012), pp. 679–692
74. D. Cristinacce, T. Cootes, Feature detection and tracking with constrained local models, in *Proceedings of British Machine Vision Conference* (2006), pp. 929–938

75. P. Tresadern, H. Bhaskar, S. Adeshina, C. Taylor, T. Cootes, Combining local and global shape models for deformable object matching, in *Proceedings of British Machine Vision Conference* (2009)
76. T. Cootes, M.C. Ionita, C. Lindner, P. Sauer, Robust and accurate shape model fitting using random forest regression voting, in *Proceedings of European Conference on Computer Vision* (2012)
77. J. Saragih, S. Lucey, J. Cohn, Deformable model fitting by regularized landmark mean-shift. Int. J. Comput. Vis. **91**, 200–215 (2011)
78. M. Valstar, T. Martinez, X. Binefa, M. Pantic, Facial point detection using boosted regression and graph models, in *Proceedings of Computer Vision and Pattern Recognition* (2010), pp. 2729–2736
79. X. Zhu, D. Ramanan, Face detection, pose estimation, and landmark localization in the wild, in *Proceedings of Computer Vision and Pattern Recognition* (2012), pp. 2879–2886
80. X. Yu, J. Huang, S. Zhang, W. Yan, D.N. Metaxas, Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model, in *Proceedings of International Conference on Computer Vision* (2013), pp. 1944–1951
81. B. Martinez, M.F. Valstar, X. Binefa, M. Pantic, Local evidence aggregation for regression based facial point detection. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1149–1163 (2013)
82. P. Wang, M.B. Green, Q. Ji, J. Wayman, Automatic eye detection and its validation, in *IEEE Workshop on FRGC, Computer Vision and Pattern Recognition* (2005), p. 164
83. A. Duchowski, *Eye Tracking Methodology: Theory and Practice* (Springer, Berlin, 2007)
84. D. Hansen, J. Qiang, In the eye of the beholder: a survey of models for eyes and gaze. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 478–500 (2010)
85. D. Yoo, M. Chung, A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. Comput. Vis. Image Underst. **98**, 25–51 (2005)
86. B. Pires, M. Hwangbo, M. Devyver, T. Kanade, Visible-spectrum gaze tracking for sports, in *WACV* (2013)
87. D. Hansen, A. Pece, Eye tracking in the wild. Comput. Vis. Image Underst. **98**, 182–210 (2005)
88. S. Cadavid, M. Mahoor, D. Messinger, J. Cohn, Automated classification of gaze direction using spectral regression and support vector machine, in *Proceedings of Affective Computing and Intelligent Interaction* (2009), pp. 1–6
89. T. Heyman, V. Spruyt, A. Ledda, 3d face tracking and gaze estimation using a monocular camera, in *Proceedings of International Conference on Positioning and Context-Awareness* (2011), pp. 23–28
90. M. Everingham, A. Zisserman, Regression and classification approaches to eye localization in face images, in *IEEE Face and Gesture* (2006), pp. 441–446
91. G. Diamantopoulos, Novel eye feature extraction and tracking for non-visual eye-movement applications. Ph.D. thesis, University of Birmingham, 2010
92. S. le Cessie, J. van Houwelingen, Ridge estimators in logistic regression. Appl. Stat. **41**, 191–201 (1992)