Gaze Direction Estimation by Component Separation for Recognition of Eye Accessing Cues

Ruxandra Vrânceanu RVRANCEANU@IMAG.PUB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Corneliu Florea CORNELIU.FLOREA@UPB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Laura Florea LAURA.FLOREA@UPB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Constantin Vertan CONSTANTIN.VERTAN@UPB.RO
Image Processing and Analysis Laboratory
University "Politehnica" of Bucharest, Romania, Address Splaiul Independenţei 313

Abstract

This paper investigates the recognition of the Eye Accessing Cues (EACs) used in the Neuro-Linguistic Programming (NLP) as a method for inferring one's thinking mechanisms, since the meaning of non-visual gaze directions may be directly related to the internal mental processes. The direction of gaze is identified by separating the components of the eye (i.e. iris, sclera and surrounding skin) followed by retrieving the relative position of the iris within the eye bounding box, that was previously extracted from an eye landmarks localizer. The eye cues are retrieved via a logistic classifier from features that describe the typical regions within the eye bounding box. The simultaneous investigation of both eyes, as well as the eye tracking over consecutive frames are shown to increase the overall performance. The here proposed solution is tested on four databases proving to have superior performance when compared in terms of recognition rate with methods relying on state of the art algorithms.

1. Introduction

Along with entering into the digital era and fostered by the growth of computer usage in daily life, there are considerable efforts of creating systems to facilitate a better automatic understanding of human thinking and emotional mechanisms as part of establishing ways for non-verbal communication (Pentland, 2008). In the computer vision part of the mentioned area, most of the research is related to the understanding of the functioning of the human mind. More precisely, it is aimed at interpreting facial expressions (Fasel and Luettin, 1999), (Zeng et al., 2009) or establishing their underlying emotions which were shown to be universally correlated (Ekman, 1982). Recently, the literature reported attempts to interpret more complex situations, such as dyadic social interactions for the diagnosis and treatment of developmental and behavioral disorders (Rehg et al., 2013) and to experiment within new areas of psychology, as pointed in the recent review by Cohn and De La Torre (Cohn and De la Torre, 2014). Among newer directions investigated, we note the detection of deception as part of hostile intention perception (Tsiamyrtzis et al., 2007), the estimation of pain intensity via facial expression analysis (Ashraf et al., 2009), (Florea et al., 2014), the interpersonal coordination of mother–infant (Messinger et al., 2009), the assistance in marketing (McDuff et al., 2013), etc. Another direction of investigation is offered by the Neuro-Linguistic Programming (NLP) theory, which

presents unexplored opportunities for understanding the human patterns of thinking and behavior.

NLP was introduced in the 70s by Brandler and Grinder (Bandler and Grinder, 1979), as a different model for detecting, understanding and using the patterns that appear between brain, language and body. One such model is the Eye-Accessing Cue (EAC) 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* (Fig. 1), under the NLP paradigm, can be used to determine the internal representational system employed by a person, 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.

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 (Sturt et al., 2012). Thus, we performed our own experiment to gain better insight of the facts: we gave various persons queries and we checked if the reaction followed the NLP rules. The recorded results were reported in (Vranceanu, Florea and Florea, 2013), and while we did not find 100% accuracy (i.e. universality), the correct apparition rates were higher than random chance.

The problem of identifying one's direction of gaze is intensively studied in computer vision. One may classify these systems by:

1. Recording position of the device. Here, we may distinguish:
   (a) Head mounted devices (e.g. glasses or head moun- ted camera);
   (b) Stationary and/or remote devices.

   While being closer to the eye, the head mounted devices have access to higher resolution and, thus, better precision. Yet, they do come with two inherent shortcomings. First, their price (spanning from several thousand dollars, for a professional commercial solution, down to a hundred dollars for the more affordable ones compared to few dollars for a normal webcam) restricts the area of usability. The second aspect is related to the fact that they are wearable, which is a distinct indicator that the user is subject to exploration and investigation by non-traditional means. Consequently, we will rely on a stationary webcam.
2. Illumination source domain. Here we note:
   (a) Active, infra-red (IR) based illumination;
   (b) Visible spectrum illumination.

   While the high performance of the commercial eye-trackers relies on the information from the IR domain, its use implies a distinct device. This is due to the fact that the IR source is not typically incorporated in webcams, and it creates similar problematic as the head-mounted category.

Furthermore, the use of distinct means (such as wearable eye tracking and/or respectively active illumination sources) limits the applicability of methods to data which was recorded accordingly; post–processing for investigation analysis of other data is futile. We constructed our system having in mind a normal digital video camera and two main applications.

The first use-case relates to online interviews. Small and medium companies look for additional employees at a distance and, due to budgetary constraints, the interview is online, with the applicant being subject to recording (sometimes by his/her own means). In such cases, given a query, discrimination between the remembering type of activities (looking left) and the constructing one (looking right) would differentiate experience from creativity. For the technique to be effective in such a case, it is mandatory for the interviewed person to be unaware of the fact that his/hers nonverbal messages are also recorded and analyzed. We construct our method assuming recoding devices typical for webcam transmission; thus, no distinct mean, such as a head mounted camera or an active illumination source, can be involved.

The second use-case is about interactive communication for marketing and training. In such a case, if the communication is online, the restrictions are similar with the previous case (the subject must have access to a recording device). For the face–to–face interaction, the meeting is recorded and analysis may be performed 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.

1.1. Related work

In computer vision, extensive research was done in the field of detecting the direction of gaze (Duchowski, 2007), (Hansen and Qiang, 2010), by means of so-called eye trackers. Usually, Eye tracking technology, relies on measuring reflections of the infrared / near-infrared light on the 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

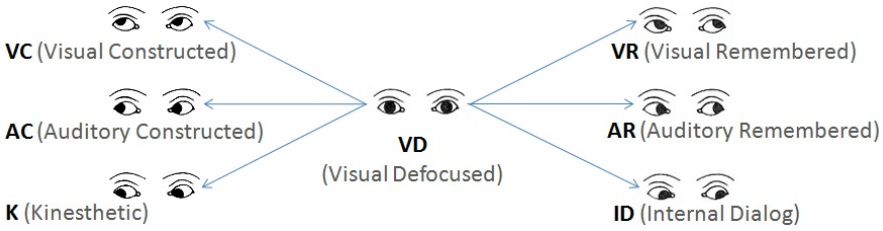

Figure 1. The 7 classes of EACs (Bandler and Grinder, 1979): When eyes are not used for visual tasks, their position can indicate how people are thinking (using visual, auditory or kinesthetic terms) and the mental activity they are doing (remembering, imagining, or having an internal dialogue).

that is used to compute the angular orientation of the eye, in the so–called "dual Purkinje" method (Hansen and Qiang, 2010). An example of such an eye tracking systems is, for instance, found in the work of Yoo and Chung (Yoo and Chung, 2005) who 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. (Pires et al., 2013) who extracted the iris contour followed by 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 (Wang et al., 2005), (Hansen and Pece, 2005), (Cadavid et al., 2009), (Heyman et al., 2011), (Wolf et al., 2010). Wang et al. (Wang et al., 2005) select recursive nonparametric discriminant features from a topographic image feature pool to train an Adaboost that locates the eye direction. Hansen and Pece (Hansen and Pece, 2005) model the eye contour as an ellipse and use Expectation-Maximization to locally fit the actual contour. Cadavid et al. (Cadavid et al., 2009) train a Support Vector Machine with spectrally projected eye region images to identify the direction of gaze. Heyman et al. (Heyman et al., 2011) use 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. (Wolf et al., 2010) used the eye landmark localizer provided by Everingham and Zisserman (Everingham and Zisserman, 2006) 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.

As we choose to locate landmarks in the eye region, our method is part of the category of face fiducial points locators. This is a rich class of methods, including some of the most recent and accurate solutions, as the ones proposed by Valstar et al. (Valstar et al., 2010) or Zhu and Ramanan (Zhu and Ramanan, 2012). The BoRMaN algorithm described in (Valstar et al., 2010) iteratively improves an initial facial landmark estimate by features processed with Markov Random Fields and Support Vector Regression. Zhu and Ramanan (Zhu and Ramanan, 2012) rely on a connected set of local templates described with Histogram of Oriented Gradient.

For the recognition of the direction of gaze in terms of NLP–EAC, we note the works from (Diamantopoulos, 2010), (Florea et al., 2013) and (Vranceanu, Florea, Florea and Vertan, 2013). In the work of Diamantopoulos (Diamantopoulos, 2010) a head mounted device is used. Taking into account that Laeng and Teodorescu (Laeng and Teodorescu, 2002) 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. (Pires et al., 2013) (iris contour detection followed by Hough transform for circles) works very well. However, for the lower resolution images, which are associated with remote 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. In (Florea et al., 2013), the focus is placed on correctly identifying the eye landmarks while the direction of gaze is only seen as a possible application limited by

the chosen approach.

The here proposed work is a direct extension of the method presented in (Vranceanu, Florea, Florea and Vertan, 2013). By comparison, here, we increase the accuracy of the method, we improve the results by fusing information from the analysis of both eyes and considering consecutive frames (in a tracking framework) and we extend the testing by considering three supplemental public databases.

1.2. Paper structure

The solution introduced in the current work assumes a scenario where the image acquisition is done with a single camera with fixed, near-frontal position, under free natural lighting. The algorithm relies solely on gray-scale images and a coarse–to–fine approach is used for localization, succeeded by gaze direction recognition. First, we precisely determine the eye bounding box, followed by a pre-processing that enhances the separation of the iris from the eyelashes. Once the eye region is segmented, and the eye components are retrieved, the relative position of the iris is extracted by template matching mechanisms and processed for the EAC recognition.

Our contribution is two-fold. On the technical side, while we rely on individual, known techniques, as building blocks, we refine them and combine them in a novel autonomous system that localizes the eye bounding box precisely, and recognizes 7 directions of gaze in real-time. On the application side, we propose the first easy–to–use system that specifically focuses on recognizing eye accessing cues in terms of NLP, that may be further incorporated in on-line communication for interviewing or training.

Thus, in section 2 we describe the detection of the eye bounding box; inside this bounding box various methods are employed for detecting the relative iris position and the corresponding EAC, as detailed in 2.2. Section 3 is some implementation details. Finally, the method is tested on the Eye-Chimera (both still and sequence parts), HPEG, UUlm and PUT databases and the achieved results are presented in section 4; the paper ends with discussions and conclusions.

2. Method

A visual overview of the proposed method is presented in Fig. 2. As the method is fully automatic, first the face crop is obtained using the classical Viola-Jones algorithm (Viola and Jones, 2004), trained with frontal faces ($\approx \pm 30^o$ yaw angle with respect to frontal positions). The yaw angle limitation originates in the constrain that both eyes should be completely visible. All the processing is then performed inside the face square, rescaled at a 100×100 pixels size. Our approach uses image projections functions to precisely determine the limits of the eye bounding box and then applies a segmentation that separates the eye components. Finally, by combining projection and segmentation description, a classification process is employed to recognize the EAC.

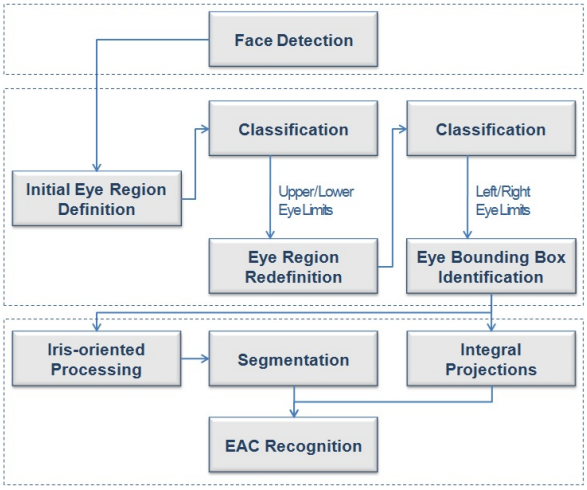

Figure 2. Workflow for the proposed EAC recognition method.

2.1. Detecting the Eye Bounding Box

While it is possible to estimate a rough bounding box of the eye directly based on the face square (as in the case presented in (Valenti and Gevers, 2008)), we aim at an improved precision. For this task we depend on the integral and edge image projections functions.

We recall that the integral image projections functions (IPF), in an image rectangle given by $[x_1, x_2] \times [y_1, y_2]$, are computed as:

$$P_V(i) = \sum_{j=x_1}^{x_2} I(i,j), \forall i = y_1, \ldots, y_2. \quad (1)$$

$$P_H(j) = \sum_{i=y_1}^{y_2} I(i,j), \forall j = x_1, \ldots, x_2.$$

where the $I(i,j)$ term stands for the luminance image at location (i,j). Similarly, the edge projections functions (EPF) are computed using a Sobel operator to obtain the magnitude edge image S from the luminance image I and then to apply Eq. (1), where I is

replaced by S:

$$E_V(i) = \sum_{j=x_1}^{x_2} S(i,j), \forall i = y_1, \ldots, y_2. \quad (2)$$

$$E_H(j) = \sum_{i=y_1}^{y_2} S(i,j), \forall j = x_1, \ldots, x_2.$$

The literature comprises many approaches related to eye localization, many of them being based on image integral projections functions (Feng and Yuen, 1998), (Zhou, 2003). Later, it was shown (Turkan et al., 2008), (Florea et al., 2012) that a combination of normalized integral image projections and machine learning systems has a high discriminative power in localizing the eye center. Here, we will take these works as a starting point and we will improve performance under the constraint to cover the specificity of the EAC testing (that is, under gaze variation).

The eye bounding box is first roughly initialized in the middle-upper face quarter band (that is the lines from $y_1 = 25$ to $y_2 = 50$) and symmetrical in the vertical image quarter bands (within $x_1 = 20$ and $x_2 = 45$ for the left eye and from $x_1 = 55$ to $x_2 = 80$ for the right eye), as shown in Fig. 3(a). Such a procedure is employed by Valenti and Gevers (Valenti and Gevers, 2008), in their iris center localization method: they have examined a set of four databases and located the relative position of the ground truth iris center with respect to Viola-Jones reported face rectangle; the result is the bounding box further used for search. Starting from those values, we enlarge the bounding box to include completely the eye lashes and to add more robustness for cases where the face detector reported rectangle is less precise.

Using the horizontal integral and edge projections of such crops (Fig. 3(c)/(d)) a logistic classifier (le Cessie and van Houwelingen, 1992) (10^{-7} ridge in log-likelihood, iterate until convergence) is first trained to find the upper and the lower limits of the eye bounding box. Next, re–cropping the eye area on the found limits and using the vertical integral and edge projections as input feature vector (Fig. 3(e)/(f)), the classifier is similarly trained to detect the left and the right limits.

The results obtained with this method are highly consistent with the ground truth, as can be seen in Fig. 4. As detailed in section 4, a precision of 95.51% is obtained for $\epsilon \leq 0.05$, while in 99.27% of the cases the error is $\epsilon \leq 0.1$.

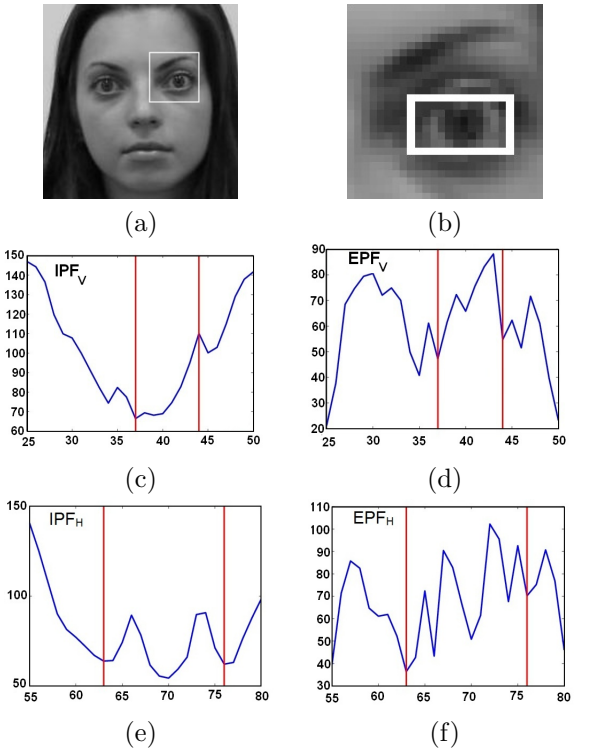

Figure 3. Eye limits in image projections functions: (a) Face crop with initial coarse eye selection; (b) Coarse eye selection with final bounding box; (c) Vertical IPF (red line marking eye upper/lower limits); (d) Vertical EPF (red line marking eye upper/lower limits); (e) Horizontal IPF (red line marking eye lateral limits); (f) Horizontal EPF (red line marking eye lateral limits).

2.2. EAC Recognition

Once the eye bounding box has been delimited, 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. The iris may be found either by the use of an eye center localizer (such as the one from (Valenti and Gevers, 2008)), or by separating the eye regions. As eye localizers are imperfect especially when challenged by gaze variation, for improving the accuracy, we focus on segmenting the components of the eye within the bounding box and use their relative position as indicators of the EAC.

Pre-processing The segmentation is performed inside the bounding box and the aim is to separate the iris, the sclera and the surrounding skin area in 3 distinct classes. Yet, as the iris and the eyelashes tend to be spatially connected and to have similar luminance values, before the actual segmentation, a preprocessing is required to separate the iris from the

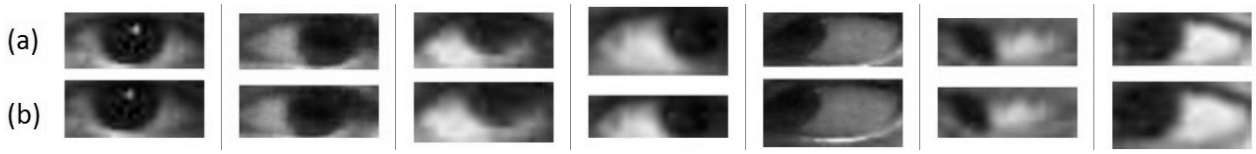

Figure 4. Eye refined bounding box: (a) Ground truth (manual markings); (b) Detected using integral and edge projections

eyelash.

Starting from the observation that the iris is a large, dark region of the eye (Wu and Zhou, 2003), we look for the darkest, smooth neighborhood within the bounding box. This is found by selecting the areas that are darker than the median luminance value within the bounding box, in both the original image and a Gaussian low-pass smoothed image. The luminance of each pixel in the remaining of the eye bounding box is then multiplied by a factor of 2, such that the segmentation will generally detect the iris as a stand alone region (Fig. 5(d)).

Segmentation Segmentation is a well known problem and many solutions have been proposed through the years. As we did not aim at good segmentation per se, we require a combination of good EAC recognition in a reasonable amount of time. According to the tests performed (visually shown in Fig. 5 and numerically quantified in section 4), the best compromise is achievable using a *K-Means* segmentation. It is possible to refine these results using *Graph Cuts* (Boykov and Kolmogorov, 2004) (which imposes a smoothness constraint to reduce the number of disconnected regions and provide more compact classes (Fig. 5(c)), yet the time overhead (75 msec in average for a portrait, compared to 10 msec for K-Means) is considerable when compared with the marginal accuracy improvement. Other tested segmentation methods are *Mean Shift* (Comaniciu and Meer, 2002) and *Watershed* (Meyer, 1994) (Fig. 5(e), (f)). These methods typically lead to over-segmentation. Yet, even though a new dynamic region merging technique is employed in order to consider light regions more similar and separate the darker areas (the iris and the eyelashes), the results under-perform the K-Means me- thod.

Post-processing and Classification The space given by the detected, refined, bounding box is normalized to a standard width (of 25 pixels), while preserving the aspect ratio. Also, since the height is variable, all eyes are aligned at the lower limit of the refined bounding box (i.e. always the bottom limit of the eye has the y coordinate equal with 0) to ensure a better separation between eyes looking down, which are contained in narrow boxes, respectively, looking up (and opened wider), within larger boxes.

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, we build upon the same integral projections functions (IPF), as we recalled their efficiency in describing the eye structure. Therefore, for a more general description inside the bounding box, the vertical and horizontal integral and edge projections are, once again, added as features for the classifier, next to the segmented regions center of mass.

In order to recognize the 7 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) and
- the concatenated horizontal and vertical integral and edge projections.

Various classification methods are considered and, as the number of features is small, the same Logistic Classification (le Cessie and van Houwelingen, 1992) gave good results.

3. Implementation

3.1. Databases

To study the specifics of the EAC detection problem, we have developed the *Eye Chimera* Database (Vranceanu, Florea, Florea and Vertan, 2013) containing all the 7 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 7 directions of gaze, with a set of 5 points marked for each eye: the iris center and 4 points delimiting the bounding box.

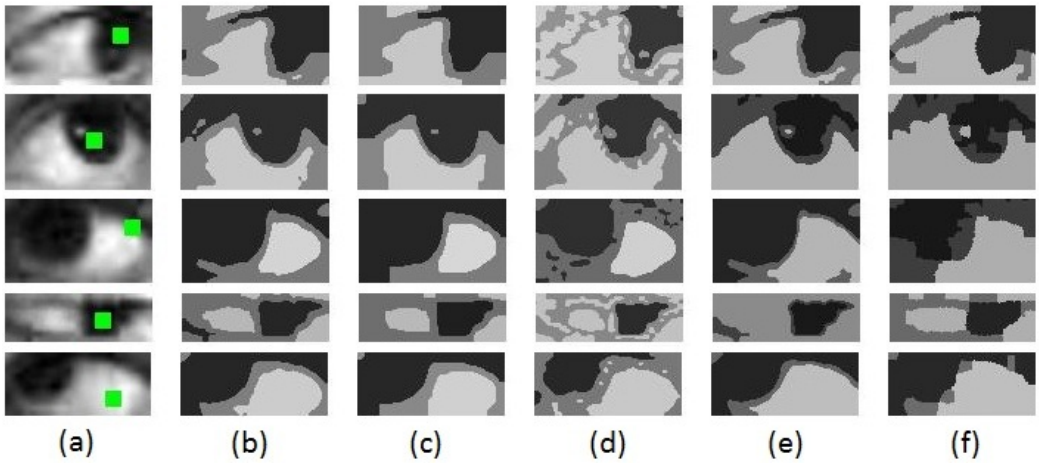

Figure 5. Eye Features: a) Iris center using (Valenti and Gevers, 2008); Segmentation in 3 classes using: b) K-Means; c) K-Means refined with Graph-Cuts; d) K-Means refined with pre-processing step; e) Mean Shift + region merging; f) Watershed + region merging.

Additionally, for a more extensive testing, we extended the basic Eye Chimera database with all the consecutive frames that are part of each basic eye movement. This part was named Eye Chimera Sequences.

Furthermore, in order to support the extensive research on eye gaze, there were introduced in literature a number of state of the art databases that contain these particular eye movements. While in some, the gaze movement appears only to be highly correlated with head pose as in the case of the Boston database (Cascia et al., 2000), we selected databases with gaze variability uncorrelated with the head pose. Three such databases are selected: the *Head Pose and Eye Gaze* (HPEG) by Asteriadis et al. (Asteriadis et al., 2009), the *PUT Face* by Kasinski et al. (Kasiński et al., 2008) and the *UUlm Head and Gaze* by Weidenbacher et al. (Weidenbacher et al., 2007) databases, which contains only sideways gazes with different head rotations. Still, it should be noted that the HPEG, PUT and UUlm databases, when compared to Chimera, introduce the variability of the head pose within the frontal detected posture.

3.2. Training and testing

The proposed method (for all experiments) is trained on randomly selected cases from the still Eye Chimera database. Half of the database is used for training, as well as for various parameter tuning.

While testing on the still Eye Chimera database, the training and testing parts are rotated such that the two-fold procedure is implied. In tests on HPEG, UUlm and PUT, the system values are the ones found while training on the Eye Chimera database.

While computing the EAC recognition rate, two scenarios are evaluated: the 7-case and the 3-case. The complete 7 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 (Hansen and Qiang, 2010), we consider only 3 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.

4. Results

In this section, we will use the still part of the Eye Chimera database to assess the influence of various parameters, as it contains the specifics of the EACs, which is the main concern for the current work. Furthermore, for the incipient tests, we will apply the method independently on each eye (therefore a face providing 2 cases). Later, we will show that simultaneously using information from both eyes increases the accuracy.

4.1. Bounding Box Detection

In order to evaluate the accuracy of the bounding box localization, we use the error for each point of interest normalized with the inter-ocular distance, as suggested in (Cristinacce and Cootes, 2006). This prox-

Table 1. Average localization rate [%] of the 4 bounding box limits on the still part from the Eye-Chimera database. We report the results when only the Integral Projections Functions (IPF) are used and when they are combined with Variance Projections Functions (VPF) and respectively with Edge Projections Functions (EPF). We also report the results achieved by our initial solution (Vranceanu, Florea, Florea and Vertan, 2013) and respectively by one of the foremost state of the art methods (Valstar et al., 2010).

Method	$\varepsilon < 0.05$	$\varepsilon < 0.1$
IPF	91.99	98.36
IPF+VPF	92.05	98.54
IPF+EPF	**94.51**	**99.27**
BoRMaN	54.85	81.40
Vranceanu et al.	89.75	98.50

imity measure m_e is computed as:

$$m_e = \frac{1}{t \cdot D_{eye}} \sum_{i=1}^{t} \varepsilon_i \qquad (3)$$

where the ε_i is the absolute distance between the ground truth for the coordinate i and the automatic finding of the same coordinate.

In evaluating the bounding box limits, only the x dimension is considered for the left and the right limits, and only the y dimension for the upper and lower limits. To provide accurate enough results, the normalized localization error should be below 0.05. A comparative evaluation of the proposed method to the coordinates given by the solution from (Valstar et al., 2010) and respectively from (Everingham and Zisserman, 2006), can be seen in Fig. 6 for various errors. While being more simple, the proposed solution generally gives more accurate results and provides an acceptable error for over 90% of the database.

The influence of the type of image projections functions used for computing the feature that describes the eye is investigated and the results are presented in Table 1 as the average error for the 4 eye limits. Best results are obtained when the integral and edge projections are combined, slightly outperforming (+3%) the version when only integral projections are used. Since a higher precision is critical for further analysis, this is the version used through the rest of the tests. However, adding the variance projections (Feng and Yuen, 1998) as well, does not bring much extra information and these types of projections are, therefore, discarded.

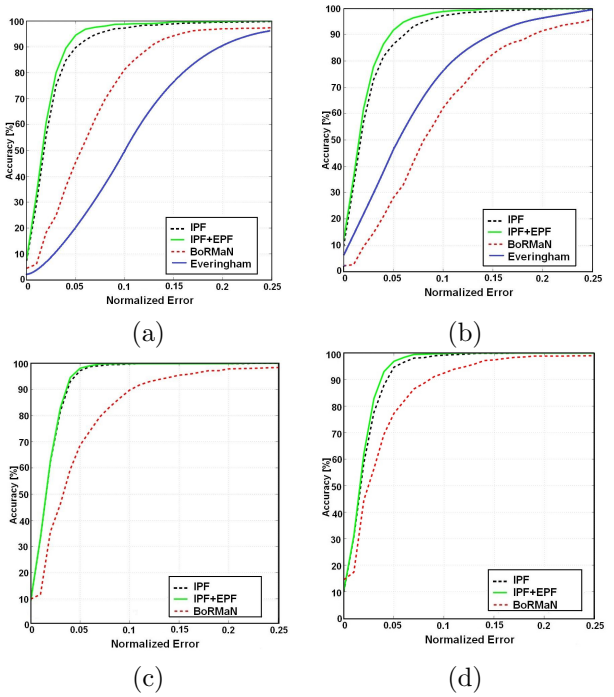

Figure 6. Eye limits detection accuracy computed on the still part of the Eye-Chimera database for different normalized errors (0:0.25), using BoRMaN solution (Valstar et al., 2010) (dashed red line), the method proposed by Everingham and Zisserman (Everingham and Zisserman, 2006) (continuous blue line) and the proposed solution based on image projection functions information: IPF (dashed black line) and IPF + EPF (solid green line), for the 4 bounding box limits: a) left; b) right; c) up; d) down. Everingham and Zisserman (Everingham and Zisserman, 2006) reported only the lateral eye limits.

4.2. Segmentation

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 2 a larger number of regions increases the EAC recognition rate; therefore, the eye space should be in fact divided in 4 regions corresponding to all the eye components present in the bounding box: the iris and the sclera, the eyelashes and the surrounding skin area.

The results obtained with various segmentation algorithms are presented in Table 3; we note that only the results for the best combination of parameters are presented. Although the Mean-Shift and Watershed segmentation has the highest accuracy when used as stand-alone methods, the K-Means segmentation gives better results, when the iris oriented pre-processing step is used (as described in section 2.2.a).

Table 2. Influence of the number of regions on the EAC recognition rate, RR[%], when simple K-Means segmentation is used (without additional projection information).

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**

Table 3. Execution time vs. EAC recognition rate for different segmentation methods using C = 4 classes of the eye region: K-Means (**KM**), K-Means refined with Graph Cuts (**KM+GC**), Mean Shift (**MS**) and Watershed (**WS**) post-processed with Region Merging (**RM**), Iris-oriented Preprocessed K-Means (**IP+KM**), which is then combined with projection information (**Proposed**).

Segmentation Method	**Execution Time [msec]**	**RR [%]**
KM	10	64.56
KM+GC	75	67.12
MS+RM	18	69.34
WS+RM	28	69.51
IP+KM	13	72.59
Proposed	14	**77.54**

4.3. EAC Recognition

The final proposed solution that gives the best results for EAC recognition consists of using iris-oriented K-Means segmentation together with projection information. As it can be seen in Table 3, 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 4 and the confusion matrix is shown in Fig. 7. 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. 8 and it should be noted that even for a human observer it is difficult, in some cases, to correctly classify the direction of gaze.

Comparison with related work. Given the state of the art, 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 (Valstar et al., 2010) 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

Table 4. Individual Recognition Rate for each EAC case on the still Eye Chimera database. The acronyms for the EACs are presented in Fig. 1.

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

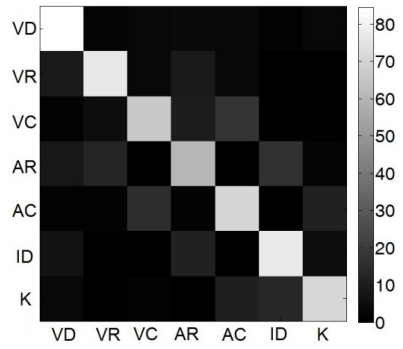

Figure 7. EAC Confusion Matrix as computed on the still Eye Chimera database.

(Valenti and Gevers, 2008). The eye landmarking method proposed in (Florea et al., 2013) also provides the required points for such an analysis. Finally, using the landmarking technique proposed in (Zhu and Ramanan, 2012), 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 5. As one can see, the proposed solution outperforms the state of the art. For the proposed method, the refined bounding box (or more precisely its height) is necessary to differentiate between looking down and looking elsewhere, while the iris position inside it, actually defines the direction of gaze. The pre-processing step removes

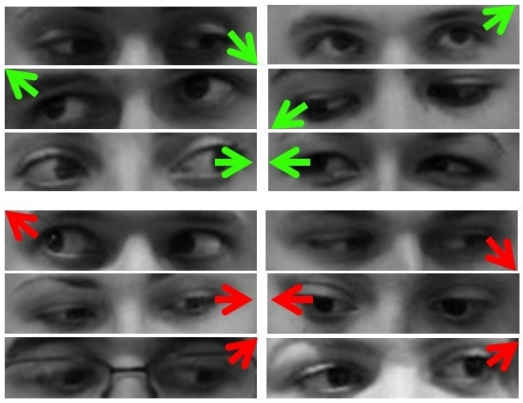

Figure 8. Automatic Recognition examples: correct (green arrow) and false (red arrow).

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 7 EAC classes are considered, the proposed algorithm surpasses the upper limit of a point-based analysis, which is obtained when only the 5 manual markings are used.

Both Eyes Information. In order to further improve the detection rate, information from both eyes is concatenated in the feature vectors. It can be seen in Table 6 that this leads to an improvement of approx. +6% in the detection rate, in both the 3 cases scenario as well as for the complete EAC set.

Temporal Redundancy. Taking into account that temporal redundancy appears between consecutive frames, we have also tried to filter potentially incorrect labels using the preceding and succeeding neighbor frames. This procedure is performed by filtering the individual labels. An alternative would be to consider a multi-frame feature, yet in such a case, the dimensionality would be much too high and the classifier losses performance.

The increase of accuracy is better for the case when a single eye is used for detection (approx. +3%), and results are only marginally improved when both eyes are used (approx. +1%). Table 7 shows these results, when our solution is applied to our larger database, comprised of consecutive frames for each motion sequence.

Other Databases. For a thorough evaluation, the proposed method (for single and both eyes information) is also compared to the state of the art 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 7 EACs, but in a small number, the UUlm database contains only 3 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 7 cues but disproportionably represented. Furthermore, all three databases have a considerable head pose variation.

Comparative results can be seen in Table 8. Although the results vary considerably across databases, the proposed method offers the best results in all scenarios. While testing on the ULM database, we looked also for 7 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.

Computational efficiency Even if computational efficiency has not been the main focus of this paper, the proposed method is fast enough, requiring an average of 125 msec per eye (95 msec for the eye bounding box localization, 10 msec for the segmentation and 20 msec for the classification), with single thread Matlab implementation and some binary routines. Even un-optimized, the performance (250 msec per face) is sufficient for real time EAC estimation. The testing was performed on an Intel I7 at 2.7 Ghz, running in single thread mode.

5. Discussion and Conclusions

If associated with non-visual movement, the direction of gaze is in fact an Eye Accessing Cue, under the NLP paradigm, and it may be used for better understanding the mental patterns of a person. Therefore, the purpose of this research was to develop an automatic solution for recognizing Eye Accessing Cues in images that contain a frontal face, and to implicitly determine the corresponding mental process behind them.

While a number of efficient approaches were investigated, the best results were obtained by precisely detecting the bounding box of the eye and performing a region-based analysis through an iris-oriented K-Means segmentation.

Multiple tests were performed on the Eye-Chimera database, a dataset that was specially designed for EAC analysis by having all the 7 eye cues well represented. The results show that the proposed method surpasses in accuracy two of the most efficient state of the art methods for detecting landmarks and implicitly eye points. The proposed method was also shown to surpass the eye landmarking technique proposed in (Florea et al., 2013) for EAC analysis, proving that a region-based solution provides better accuracy than a point-based approach.

Furthermore, in a more thorough testing under various acquisition conditions, the superior results of the proposed method were also confirmed on other external databases.

It was also shown that the accuracy of determining the eye cues can be further improved by using information from both eyes.

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

Table 5. Recognition rate [%] on the still Eye Chimera database for the 3 EAC cases scenario (when the focus is on the type of mental activity) and for the 7 EAC cases scenario (the complete EAC set) when using iris relative position inside the eye space.

Bounding Box Method	**Iris Detection Method**	**RR [%] 7 classes**	**RR [%] 3 classes**
Manual	Manual	73.98	94.52
Manual	Darkest Region	66.35	88.39
Manual	(Valenti and Gevers, 2008)	32.30	36.40
Proposed	Manual	67.21	91.03
Proposed	Darkest Region	65.24	86.34
BoRMaN (Valstar et al., 2010)	(Valenti and Gevers, 2008)	32.00	33.12
(Zhu and Ramanan, 2012)	(Zhu and Ramanan, 2012)	39.21	45.57
(Florea et al., 2013)	(Florea et al., 2013)	48.64	78.57
Proposed	Proposed	**77.54**	**89.92**

Table 6. EAC Recognition Rate [%] for the proposed solution versus state of the art, when information from both eyes is used.

Method	**RR [%]**	**(Valenti & Gevers,2008) +(Valstar et al., 2010)**	**(Zhu et al.,2012)**	**Proposed (1 eye)**	**Proposed (2 eyes)**
Still Eye Chimera	7 classes	39.83	43.29	77.54	**83.08**
	3 classes	55.73	63.01	89.92	**95.21**

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

Some additional issues remain for further investigation and development. First, Eye Accessing Cues 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 (Duchowski, 2007) as non-visual movements are shorter and with smaller amplitude. However, in both visual memory related task (Laeng and Teodorescu, 2002) 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.

Future research will focus on finding more features and increasing the precision of the bounding box detection, which are critical for good EAC detection. Supplementary inclusion of the audio retrieved data should enhance the trust of extracting eye cues and move on step closer to complete and easy to use solution.

Acknowledgment

This work has been co-funded by the Sectoral Operational Program Human Resources Development (SOP HRD) 2007-2013, financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/107/1.5/S/76903, POSDRU/89/ 1.5/S/62557 and POSDRU/159/1.5/S/134398.

References

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Ambadar, Z., Prkachin, K. and Solomon, P. (2009). The painful face – pain expression recognition using active appearance models, *Image Vis Comput.* **27**(12): 1788–1796.

Asteriadis, S., Soufleros, D., Karpouzis, K. and Kollias, S. (2009). A natural head pose and eye gaze

Table 7. EAC Recognition Rate [%] for the proposed solution over the NLP sequences (i.e. computed on the Eye Chimera Sequences) database.

Database	**NLP Sequences**		**NLP Sequences + Filtering**	
Recognition	1 Eye	2 Eyes	1 Eye	2 Eyes
7 classes	78.01	83.67	81.04	**84.37**
3 classes	89.28	96.07	90.99	**96.24**

Table 8. EAC Recognition Rate [%] for the proposed solution versus state of the art, computed on the external databases.

Method	**RR** [%]	**(Valenti & Gevers,2008) +(Valstar et al., 2010)**	**(Zhu et al.,2012)**	**Proposed (1 eye)**	**Proposed (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**

dataset, *ACM Workshop on Affective Interaction in Natural Environments*, pp. 1–4.

Bandler, R. and Grinder, J. (1979). *Frogs into Princes: Neuro Linguistic Programming*, Real People Press, Moab, UT.

Boykov, Y. and Kolmogorov, V. (2004). An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9): 1124–1137.

Cadavid, S., Mahoor, M., Messinger, D. and Cohn, J. (2009). Automated classification of gaze direction using spectral regression and support vector machine, *ACII*, pp. 1–6.

Cascia, M. L., Sclaroff, S. and Athitsos, V. (2000). Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models, *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(4): 322–336.

Cohn, J. F. and De la Torre, F. (2014). *The Oxford Handbook of Affective Computing*, Oxford University Press, chapter Automated Face Analysis for Affective Computing.

Comaniciu, D. and Meer, P. (2002). Mean Shift: A robust approach toward feature space analysis, *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5): 603–619.

Cristinacce, D. and Cootes, T. (2006). Feature detection and tracking with constrained local models, *BMVC*, pp. 929–938.

Diamantopoulos, G. (2010). *Novel eye feature extraction and tracking for non-visual eye-movement applications*, PhD thesis, Univ. of Birmingham.

Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag.

Ekman, P. (1982). *Emotion in the Human Face*, Cambridge Univ. Press.

Everingham, M. and Zisserman, A. (2006). Regression and classification approaches to eye localization in face images, *IEEE FG*, pp. 441–446.

Fasel, B. and Luettin, J. (1999). Automatic facial expression analysis: A survey, *Pattern Recognition* **36**(1): 256–275.

Feng, G. C. and Yuen, P. C. (1998). Variance projection function and its application to eye detection for human face recognition, *Pattern Recognition Letters* **19**(9): 899–906.

Florea, C., Florea, L. and Vertan, C. (2014). Learning pain from emotion: Transferred hot data representation for pain intensity estimation, *ECCV workshop on ACVR*.

Florea, L., Florea, C., Vertan, C. and Vranceânu, R. (2012). Zero-crossing based image projections encoding for eye localization, *EUSIPCO*, pp. 150 – 154.

Florea, L., Florea, C., Vranceanu, R. and Vertan, C. (2013). Can your eyes tell me how you think? A gaze directed estimation of the mental activity, *BMVC*.

Hansen, D. and Pece, A. (2005). Eye tracking in the wild, *Computer Vision and Image Understanding* **98**(1): 182–210.

Hansen, D. and Qiang, J. (2010). In the eye of the beholder: A survey of models for eyes and gaze, *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(3): 478–500.

Heyman, T., Spruyt, V. and Ledda, A. (2011). 3d face tracking and gaze estimation using a monocular camera, *Proc. of International Conference on Positioning and Context-Awareness*, pp. 23–28.

Kasiński, A., Florek, A. and Schmidt, A. (2008). The PUT face database, *Image Processing & Communications* **13**(3-4): 59–64.

Laeng, B. and Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene, *Cognitive Science* **26**: 207–231.

le Cessie, S. and van Houwelingen, J. (1992). Ridge estimators in logistic regression, *Applied Statistics* **41**(1): 191–201.

McDuff, D., Kaliouby, R. E. and Picard, R. (2013). Predicting online media effectiveness based on smile responses gathered over the internet, *IEEE FG*.

Messinger, D. S., Mahoor, M. H., Chow, S. M. and Cohn, J. (2009). Automated measurement of facial expression in infant-mother interaction: A pilot study, *Infancy* **14(3)**: 285–305.

Meyer, F. (1994). Topographic distance and watershed lines, *Signal Processing* **38**: 113–125.

Pentland, A. (2008). *Honest Signals: How they shape our world*, MIT Press, Cambridge, MA.

Pires, B., Hwangbo, M., Devyver, M. and Kanade, T. (2013). Visible-spectrum gaze tracking for sports, *WACV*.

Rehg, J., Abowd, G., Rozga, A. and et. al. (2013). Decoding childrens social behavior, *IEEE CVPR*, pp. 3414–3421.

Sturt, J., Ali, S., Robertson, W., Metcalfe, D., Grove, A., Bourne, C. and Bridle, C. (2012). Neurolinguistic programming: systematic review of the effects on health outcomes, *British Journal Of General Practice* **62**(604): 757–764.

Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I. T., Frank, M. G. and Ekman, P. (2007). Imaging facial physiology for the detection of deceit, *Int. Journal of Computer Vision* **71**: 197–214.

Turkan, M., Pardas, M. and Cetin, A. E. (2008). Edge projections for eye localization, *Optical Engineering* **47**: 047–054.

Valenti, R. and Gevers, T. (2008). Accurate eye center location and tracking using isophote curvature, *IEEE CVPR*, pp. 1–8.

Valstar, M., Martinez, T., Binefa, X. and Pantic, M. (2010). Facial point detection using boosted regression and graph models, *IEEE CVPR*, pp. 2729–2736.

Viola, P. and Jones, M. (2004). Robust real-time face detection, *Int. Journal of Computer Vision* **57**(2): 137–154.

Vranceanu, R., Florea, C., Florea, L. and Vertan, C. (2013). NLP EAC recognition by component separation in the eye region, *CAIP*, pp. 225–232.

Vranceanu, R., Florea, L. and Florea, C. (2013). A computer vision approach for the eye accesing cue model used in neuro-linguistic programming, *Scientific Bulletin of Univ. Politehnica of Bucharest, series C* **75**(4): 79–90.

Wang, P., Green, M. B., Ji, Q. and Wayman, J. (2005). Automatic eye detection and its validation, *IEEE Workshop on FRGC, CVPR*, p. 164.

Weidenbacher, U., Layher, G., Strauss, P. and Neumann, H. (2007). A comprehensive head pose and gaze database, *IET International Conference on Intelligent Environments.*, pp. 455–458.

Wolf, L., Freund, Z. and Avidan, S. (2010). An eye for an eye: A single camera gaze-replacement method, *IEEE CVPR*, pp. 817–824.

Wu, J. and Zhou, Z.-H. (2003). Efficient face candidates selector for face detection, *Pattern Recognition* **36**(5): 1175–1186.

Yoo, D. and Chung, M. (2005). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion, *Computer Vision and Image Understanding* **98**(1): 25–51.

Zeng, Z., Pantic, M., Roisman, G. and Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions, *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(1): 39–58.

Zhou, Z. (2003). Projection functions for eye detection, *Pattern Recognition* **37**(5): 1049–1056.

Zhu, X. and Ramanan, D. (2012). Face detection, pose estimation, and landmark localization in the wild, *IEEE CVPR*, pp. 2879–2886.