Can Your Eyes Tell Me How You Think? A Gaze Directed Estimation of the Mental Activity

Laura Florea
http://alpha.imag.pub.ro/common/staff/lflorea
Corneliu Florea
http://alpha.imag.pub.ro/common/staff/cflorea
Ruxandra Vrânceanu
rvranceanu@alpha.imag.pub.ro
Constantin Vertan
http://alpha.imag.pub.ro/common/staff/vertan

Image Processing and Analysis Laboratory, LAPI
University "Politehnica" of Bucharest
Bucharest, Romania

Abstract

Many applications pointed to the informative potential of the human eyes. In this paper we investigate the possibility of estimating the cognitive process used by a person when addressing a mental challenge, according to the Eye Accessing Cue (EAC) model from the Neuro-Linguistic Programming (NLP) theory [3]. This model states that there is a subtle, yet firm, connection between the non-visual gaze direction and the mental representation system used. From the point of view of computer vision, this work deals with gaze estimation under passive illumination. Using a multistage fusion approach, we show that it is possible to achieve highly accurate results in both terms of eye gaze localization or EAC case recognition.

1 Introduction

Psychologists noted the importance of efficient perception of people's gaze direction in non verbal communication. Human estimation of other's gaze direction is performed by specific neuronal circuitry involving areas from the superior temporal sulcus [23]. Eye gaze direction and eye movement may be perceived as tools for inferring one's cognitive processes [28], respectively emotional states, or for communicating in interpersonal relations [1].

In this paper we focus on identifying one's cognitive process by analyzing the direction of gaze. By analyzing non-visual movements of the eyes, Bandler and Grinder introduced the Eye Accessing Cues (EAC) model as part of the Neuro-Linguistic Programming (NLP) theory [3]. The NLP theory aims at describing the behavior patterns created by the interaction between the brain, the language and the body. The EAC model describes the eye-movements that are not used for visual tasks (non visual movements) and suggests that the direction of gaze, in such a case, can be an indicator for the internal representational system used by a person facing a given query: he/she may respond by thinking in visual terms (i.e. images),

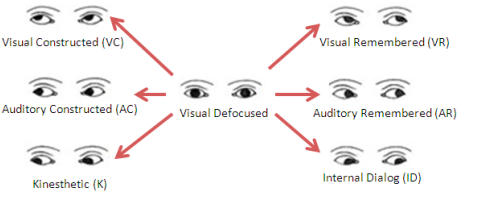


Figure 1: EACs as defined in the NLP theory. The model states that the specifics of the thinking mechanism is connected with a non-visual gaze orientation.

auditory (sounds) or kinesthetic terms (feelings, tastes, smells), as well as for the mental activity of that person, of remembering, imagining or having an internal dialog.

The here proposed approach first analyzes the eye region by localizing the eye landmarks and demonstrate that high accuracy is achievable using a multi stage fusion approach. Later, the landmarks positioning will be transformed into EACs. The proposed algorithm combines luminance and position prior information with standard machine learning process (feature extraction from a sliding sub-window, followed by classification) and iterative shape fit for landmarks localization. Given the positions, we identify the EACs, formulating, to our best knowledge, the first accurate solution to the problem of automatically recognizing the EACs.

1.1 Eye Accessing Cues

The origins of the idea that involuntary eye movements point to a person internal representational system goes back until the 19th century [18]. Dilts et al. [13] had proved the existence of such connection by using electrodes to track simultaneously eye movements and brain wave characteristics of people while asking them questions that were meant to stimulate their sense of sight, hearing or feeling. The results of these tests lead to the identification of the basic NLP EACs, which are shown in figure 1.

Since its introduction, several studies have attempted to validate this model. While some studies have shown a correlation between the representational systems and the direction of eye gaze, mostly for the visual system [4], [6], other studies were, at best, neutral, [15], [24], with the most recent calling for additional research [26]. The studies revealed an interdependence between the appearance of certain predicates or words associated with one of the representational systems (visual, auditive or kinesthetic), and the first EAC that follows them. Yet, the cognitive response differs from person to person and all the eye movements in response to a question should be considered.

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 [14]. Here, we concentrate our efforts only on correctly identifying the EAC from manually annotated still images, that are acquired under minimal invasive conditions (i.e. without a head mounted camera), as they may disturb the natural cognitive process. The actual EAC is thought to be identified by distinguishing between the relative position of the iris and the eye socket (lid edge). Our approach is to determine the four limits of the eye socket: the inner and outer corners, the upper and lower lids and the iris center and to subsequently analyze the identified region.

1.2 Related work

As far as we know, no previous work addressed the specific task of EAC recognition. Yet, the method specifics place it close to iris center localization, gaze detection or facial landmark estimation. Despite the rich computer vision literature in all these aspects, due to paper length limitations, we will consider only a partial review and we refer the reader to the recent surveys in [17] (for eye localization and gaze tracking) and in [8] (for facial landmark estimation).

There were many computer vision approaches to gaze direction identification [14], [17], most based on the use of an eye tracker. Eye tracking technology usually relies on intrusive techniques based on active illumination (such as measuring the reflection of light, or analyzing the corneal reflections – the Purkinje images), which are a potential perturbation to one's inner cognitive process and thus are excluded from the current work. The alternative is to develop non-intrusive techniques that directly measure the gaze direction, such as the approaches used in [33], [7], [35], that locate also facial landmarks. Also, we restricted for the moment the use of head movement, tracking and orientation, such as the many approaches surveyed in [22].

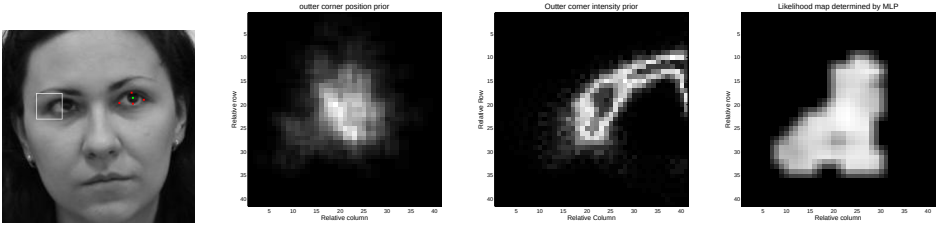
Facial landmarking originates in the classical holistic approaches of Active Appearance Models (AAMs) [9] and Elastic Graph Matching [21]. In the recent times, it switched to independent models built on top of local part detectors, sometimes known as Constrained Local Models (CLMs) [12]. The CLM model was extended with full voting from a random forest in [11] or in [25] where statistical shape model parameters are optimized via probabilistic interpretation. All other notable work assumes a connected spatial model, requiring the need for approximate matching algorithms. Thus Valstar *et al*. [31] complemented the SVM regressed feature point location with conditional MRF to keep the estimates globally consistent. Belhumeur *et al*. [5] proposed a Bayesian model combining the outputs of the local detectors with a consensus of non-parametric global models for part locations; Zhu *et al*. [36] relied on a connected set of local templates described with HOG.

As many facial landmarking methods do not localize the iris center, we consider here as foremost state of the art the maximum isophote method proposed by Valenti *et al*. [29]; especially since the necessary circular and symmetrical shape of the iris is not occluded in the envisaged situation.

2 Localization of eye landmarks

Similarly to [12], the here proposed algorithm for eye landmarks (inner corner, outer corner, upper lid, lower lid and iris center) localization has two pre-requisites (face square and eye centers) and uses a multi stage detection based on likelihood maps for position and intensity, followed by a more elaborate template matching and shape constraining techniques.

Pre-requisites: Face Detector and Eye Localizer The used face detector is the classical Viola-Jones [32] followed by an iris center localizer . Motivated by the fact that eye impairment may bring additional information about the gaze direction, we restricted our analysis to images which have *both* eyes completely visible, constraint which leads to a maximum head angle of 30^o. These cases are covered by a frontal face detector. All subsequent computation is performed on a normalized face square of 300×300 pixels, obtained by bilinear re-sampling of the detected face box. For the problem of eye (iris/pupil) center localization,



(a) (b) (c) (d)

Figure 2: Searched eye landmarks (showed in figure (a) on the right eye) and an outer corner maps example: (a) Face crop from an original image with highlighted outer eye corner search area;(b) Position prior map p_1 gathered from all training images; (c) Intensity prior map p_2; (d) Likelihood Map determined by the MLP given the actual data p_3.

we compared the method proposed by Valenti et al. [30] with the method from [16]. The comparative evaluation results are in table 1. Taking note that neither is sufficiently accurate when challenged by gaze variation, we significantly improve the performance of the eye localizer from [16] by considering a small, centered region of interest (ROI) 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 procedure of the eye center and the localization of the other eye landmarks, requires position and intensity priors and template matching.

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 [12]. For each candidate landmark, we construct its position prior as a probability map spanning its region of interest, such that each position is given the likelihood of being close to the ROI center. This prior is denoted as $p_1(i, j)$ and is exemplified in figure 2(b) for the case of the outer corner of the eye.

While for eye center localization it is common to investigate only the darkest pixels [33], we, here, extend this idea 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, we construct its intensity prior 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 figure 2(c) for the case of the outer corner of the eye.

Template matching The problem to be solved is to determine for each pixel, within a reasonable neighbourhood of an initial landmark, the probability of that pixel being the true landmark. As proposed in [27], we considered rectangular patches centered in the given point and we trained a Multi-Layer Perceptron (MLP) with the integral and edge projections within that image patch. While in [27] the edge projections were taken from a wavelet transform, we use the simple Sobel edge intensity image, that has the benefit of being invariant to additive illumination changes. We recall that for a gray-level image sub–window $I(i, j)$ with $i = i_m \dots i_M$ and $j = j_n \dots j_N$, the projection on the horizontal axis, P_H is the average gray–level along the columns, while the vertical axis projection, P_V is the average gray–level along

the rows (1):

$$P_H(j) = \frac{1}{H}\sum_{i=i_m}^{i_M} I(i,j), j = j_n,\ldots,j_N;\ \ P_V(i) = \frac{1}{V}\sum_{j=j_n}^{j_N} I(i,j), i = i_m,\ldots,i_M \tag{1}$$

where $H = i_M - i_m + 1$ and $V = j_N - j_n + 1$, and we empirically found that the best values are $H = W = 30$. The edge projections are found using eq. (1), where the original image values, $I(i,j)$ are replaced with Sobel edge intensity image values. To ensure a better robustness to illumination variation, we normalized the projections to the $[-128, 127]$ range. Each patch is thus described by a feature vector formed by the concatenation of the four projections: horizontal/vertical, integral/edge.

A MLP is separately trained for each landmark (with patches as input) to output the Euclidian distance between the center of the input patch 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 typical p_3 map is showed in figure 2(d) for the location of the outer eye landmark. For each landmark, a final probability map is computed as:

$$p_{123}(i,j) = \alpha \cdot p_1(i,j) + \beta \cdot p_2(i,j) + \gamma \cdot p_3(i,j) \tag{2}$$

where α, β and γ weight the confidence in each type of map; their values were deduced on the training database independently for each landmark for each eye. Good general choices are $\alpha = \beta = 0.25$, $\gamma = 0.5$. The weighted center of mass of the p_{123} probability map for each landmark represents an accurate estimate to that landmark.

Shape constraints A modified form of the Constrained Local Model [12] is further used: instead of gathering local constrains and maximizing the global output, for each landmark (locally), we iteratively use the global shape to construct a local constraint. As we investigate eye landmarks in terms of gaze variation, the iris center position has very large variations, thus it is not included in the shape; the shape contains only four points, namely the limits of the eye socket.

From the ASM model [10] we recall that given s sets of points $\{\mathbf{x}_i\}$, after alignment, one may compute the mean shape $\bar{\mathbf{x}}$, and the projection matrix on a PCA reduced space $\mathbf{P}$. Any of the input shapes is given as:

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \ \ \Rightarrow \mathbf{b} = \mathbf{P}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{3}$$

Given a large number of shapes, one may compute a histogram for the transformation vector, $\mathbf{b}$. This histogram may be approximated with a t-dimensional independent (due to PCA) Gaussians distributions with $\mathbf{0}$ mean and $\Sigma = \mathrm{diag}(\lambda_i)$ covariance matrix, where λ_i is the shape eigenvalue on the dimension i. Thus for each location, we compute $p_s(i,j)|\mathbf{x}$ as the probability to have the landmark at position (i,j) given a shape $\mathbf{x}$.

Taking into account that at the current step, for each landmark, an estimate of the shape is available 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, we build the likelihood of various positions for the current pixel. Then, the procedure is repeated iteratively for each landmark N_{it} times (typically $N_{it} = 2$).

The *final landmark position* is taken as the weighted center of mass of the convex combination (with typical weight δ= 0.7), $p_F(i,j)$, between the initial stages and the shape fitting likelihood:

$$p_F(i,j) = \delta \cdot p_{123}(i,j) + (1-\delta) \cdot p_S(i,j)|\mathbf{x} \tag{4}$$

3 Recognizing individual EACs

The recognition of the EAC case (gaze direction) is done by identifying the position of the iris center inside the eye socket. The here proposed methodology uses eye socket limits, as given by the eye landmarks, the eye center and the description of the eye socket content via integral projections. More precisely, the four landmarks on the eye socket form a quadrilateral. The H and V limits used in equation (1) vary on consecutive image rows/colums, as they are equal with the number of pixel on the specific row/column inside the given quadrilateral. Each of the resulting integral projections is normalized to 32 samples. The final feature describing the EAC of a given eye consists thus from the 10 coordinates of the normalized eye landmarks and the 64 elements form the horizontal/vertical projections.

For the actual recognition we have trained a classifier to take as input the EAC feature and to return the EAC case. After extensive testing we determined that simple classifiers (as MLP, SVM or boosting comparators) are unable to cope with the many possibilities encountered. Good performances are achieved only by meta-classifiers, and we thus considered that a very good compromise is achieved by the use of a regression random forest with 40 trees (mean square error as a splitting criterion, unlimited depth and 35 features used at each node).

4 Implementation and results

4.1 Databases

Existing gaze databases We have used three separate databases: HPEG by Asteriadis et al. [2], ULM head and gaze by Weidenbacher et al. [34] and PUT by Kasinski et al. [20]. HPEG database is given as videos. For this experiment we have extracted frames with relevant gaze variation. Out of these databases we have selected only images that include a head angle (on any direction) smaller than 30^0. While ULM and PUT databases include ground truth for landmarks (ULM only iris center and inner and outer corners, while PUT all eye landmarks), we have annotated the frames selected from HPEG database.

Eye-Chimera database Although there exist databases designed for the study of eye gaze, none deals with the specifics of the seven EAC situations (including the neutral – visually defocused pose). We have thus constructed the Eye-ChIMERA (Eye part from the **C**ognitive process **I**nference by the **M**utual use of the **E**ye and exp**R**ession **A**nalysis) database. It contains several parts, both videos and still and was recorded with Canon 600D (640×480 resolution) and Panasonic HDC-TM 60 (1920×1080) cameras on 40 subjects in the 20-30 age range. Illumination is typical indoor and coming from one side. For the unconstrained part, the subjects were given various queries, as recommended in [3]. For the simulated part the subjects were asked to follow specific movement patterns. While all data has been recoded as video, for now, the movements between consecutive EACs are identified and cropped out. The first and last frame of each move are selected and labelled with the corresponding EAC

Table 1: Iris center localization accuracy measured as percentage of the inter-ocular distance (stringent criterion [19]).

Method	Valenti *et al.* [30]		Florea *et al.* [16]		[16] + improvement	
Database	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.05$	$\varepsilon = 0.1$
Eye-Chimera	16.1 %	50.5%	42.7%	68.9%	65.3%	78.7%
HPEG	24.6 %	55.7%	53.5%	78.8%	71.1%	83.2%
ULM	50.1 %	77.0%	32.5%	74.1%	76.6%	92.7%

tag. The five eye landmarks are manually marked by three experienced annotators and their average was taken as ground truth. In all, the Eye-Chimera database comprises 1172 frontal face images, grouped according to the 7 gaze directions, with a set of 5 points marked for each eye: the iris center and 4 points delimiting the bounding box.

4.2 Eye landmark detection

Training We have trained the eye landmark localization system (with all four stages) on the PUT database (where 1000 images were randomly selected). Testing was done separately on the Eye-Chimera (1172 images), HPEG (233 images) and ULM (335 images).

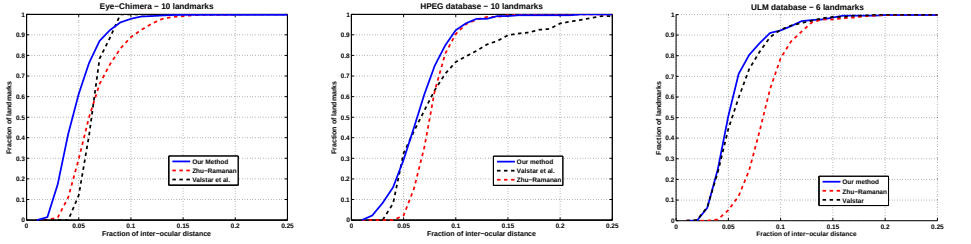
Evaluation procedure The standard measures for evaluation of the localization accuracy are the stringent criterion for the eye centers [19], ε, and the proximity measure [12] for multiple landmarks, m_e. Formally, this two are given by:

$$\varepsilon = \frac{\max\{\varepsilon_L, \varepsilon_R\}}{D_{eye}}; \quad m_e = \frac{1}{t \cdot D_{eye}} \sum_{i=1}^{t} \varepsilon_i \tag{5}$$

In the left hand part of the equations above, ε_L and ε_R are the Euclidean distances between the ground truth left/right eye center and the determined left/right eye center, while D_{eye} is the distance between the ground truth eyes centers. In the right hand part, ε_i are the point to point errors for each individual landmark location and t is the number of feature points searched (10 in our case, thus the measure will be referred as m_{e10}).

Iris center localization In table 1 we report the comparative performance of the state of the art method [30], the initial localization accuracy [16] and the overall improved results.

Eye landmark localization For this specific task we have compared our algorithm with the same landmarks returned by state of the art methods described in [31] and [36]. The comparison error may be seen in figure 3. Full results are in supplementary materials 2,3 and 4. As one can see, in all cases we provide at least the same accuracy as state of the art methods, but more often we outperform them. The most striking difference is on the Eye-Chimera database, where we noted that other methods significantly under-performs for gaze variation on both horizontal and vertical. Nevertheless, the proposed method produces improvement, even small, on all considered databases.



(a) (b) (c)

Figure 3: Eye landmarks localization and comparison with prior art on Eye-Chimera, HPEG and ULM databases.

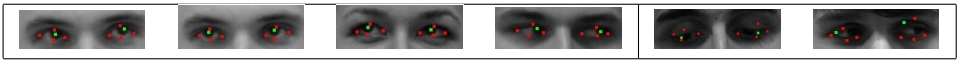


Figure 4: Examples of eye landmarks localization from Chimera database. The first four are correct localization ($d_{me10} < 0.05$), while the last two are examples with $d_{me10} > 0.1$.

4.3 EAC recognition

For the EAC recognition, as only the Eye-Chimera database is available, given the 2340 available eyes (2 eyes per image) we have selected 30 images (i.e. 60 examples) from each category and trained the random forest on the ground truth data.

We consider two type of recognition situations: three cases (looking left corresponding to constructed type of mental activity, looking straight - defocused and looking right - remembering) and 7 cases (described in figure 1). To have an idea of the chosen recognition approach upper limit, we tested it on the *manual landmarks*, achieving a recognition rate of *91.2%* for the three case and of *77.0%* for the 7 case situation. When applied on the actual data (from which the training cases were removed), the recognition rate was **78.57%** for three classes and **48.64%** for 7 classes. The confusion matrices are presented in table 2. Individually, only landmarks produced a 43.93% recognition rate for the 3–case and 71.27% for the 7–case, while only projections lead to 34.45% and respectively 55.94%.

4.4 Discussion

First, we have to point out that for the *eye landmarks localization*, when challenged by gaze variation and limited head angle, the hereby proposed method is an improvement compared to state of the art methods [31] and [36] both in accuracy and in duration. The multistage procedure also improves eye center localization, outperforming state of the art method [30].

In terms of the *EAC recognition rate*, noting that the training/testing cases ratio is 1 train example to 6 testing ones, we used a stringent test, pointing to most likely a lower limit of the method. Next, since the recognition rate is almost twice as large for the 3–case scenario than for the 7–case scenario, we may conclude that gaze direction separation on left - right direction is more at hand than separation on up-down direction; this conclusion is also confirmed by other gaze direction methods surveyed in [17].

Currently, the proposed algorithm is based on a Matlab implementation, with some dependencies in binary format, running on a single core i7 at 2.7 Ghz in 450 msec. This time includes face detection (7 msec), initial eye center localization (7 msec) and EAC classification (18 msec). Thus, the landmark localization method requires around 40 msec for each of

Table 2: Confusion matrix for the 3 cases scenario (when we focus on the mental activity type) and for the 7 cases when we look for both activity type and information type - complete EAC (using the acronyms from figure 1).

Gaze dir	*Activity*	*No. Cases*	Center	Right ⇒	Left ⇐
Center	Defocused	608	**349**	156	103
Right	Remembering	658	43	**537**	78
Left	Constructing	656	19	91	**546**

Gaze dir	*EAC*	*Cases*	Center	⇑⇒	⇐⇑	⇒	⇐	⇓⇒	⇐⇓
Center	VD	608	**349**	118	82	36	18	2	3
UpRight, ⇑⇒	VR	232	12	**169**	4	39	6	1	1
UpLeft ⇐⇑	VC	220	5	8	**137**	18	94	8	10
Right ⇒	AR	224	11	107	1	**85**	13	5	2
Left ⇐	AC	220	8	8	74	13	**113**	2	2
DownRight ⇓⇒	ID	202	20	26	3	82	37	**23**	11
DownLeft ⇐⇓	K	216	6	2	44	29	98	7	**30**

the ten landmarks, a time which is much lower than state of the art methods ([31] requiring $\approx$ 2500 msec/landmark or [36] with $\approx$ 200msec/landmark). These results suggest that a C implementation will take less than 50 msec/face, making it possible to discriminate between posed and unconstrained non-visual gaze directions.

5 Conclusion

In this paper, we presented in premier an eye accessing cue classifier that can discriminate with high accuracy in a 3-case challenge and with medium accuracy in a 7-case challenge on a newly introduced and annotated database. The recognition method is completely automatic and makes use of eye landmarks found with a novel method that may be beneficial for other eye-related tasks.

References

[1] R.B. Adams and R. E. Kleck. Effects of direct and averted gaze on the perception of facially communicated emotion. *Emotion*, 5:3 – 11, 2005.

[2] S. Asteriadis, D. Soufleros, K. Karpouzis, and S. Kollias. A natural head pose and eye gaze dataset. In *ACM Workshop on AFFINE*, pages 1 – 4, 2009.

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

[4] C.E. Beck and E.A. Beck. Test of the eye movement hypothesis of neurolinguistic programming: a rebuttal of conclusions. *Perceptual and Motor Skills*, 58(1):175 – 176, 1984.

[5] P. Belhumeur, D. Jacobs, D. Kriegman, and N. Kumar. Localizing parts of faces using a consensus of exemplars. In *IEEE CVPR*, pages 545 – 552, 2011.

[6] M. Buckner, N.M. Meara, E.J. Reese, and M. Reese. Eye movement as an indicator of sensory components in thought. *J. Counseling Psychology*, 34(3):283 – 287, 1987.

[7] S. Cadavid, M. Mahoor, D. Messinger, and J. Cohn. Automated classification of gaze direction using spectral regression and support vector machine. In *ACII*, pages 1 – 6, 2009.

[8] O. Çeliktutan, S. Ulukaya, and B. Sankur. A comparative study of face landmarking techniques. *EURASIP JIVP*, 13, 2013. doi:10.1186/1687-5281-2013-13.

[9] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE T. PAMI*, 23(6):681 – 685, 2001.

[10] T.F. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *CVIU*, 61:38 Ű 59, 1995.

[11] T.F. Cootes, M.C.Ionita, C.Lindner, and P.Sauer. Robust and accurate shape model fitting using random forest regression voting. In *ECCV*, 2012.

[12] D. Cristinacce and T. Cootes. Feature detection and tracking with constrained local models. In *BMVC*, pages 929 – 938, 2006.

[13] R.B. Dilts, J. Grinder, R. Bandler, and J. DeLozier. *Neuro-Linguistic programming: Volume 1: the study of the structure of subjective experience*. Meta Publications: Cupertino CA, 1980.

[14] A. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, 2007.

[15] M. Elich, R.W. Thompson, and L. Miller. Mental imagery as revealed by eye movements and spoken predicates: a test of neurolinguistic programming. *J. Counseling Psychology*, 32(4):622 – 625, 1985.

[16] L. Florea, C. Florea, C. Vertan, and R. Vranceânu. Zero-crossing based image projections encoding for eye localization. In *EUSIPCO*, pages 150 – 154, 2012.

[17] D. W. Hansen and Q. Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE T. PAMI*, 32(3):478 – 500, March 2010.

[18] W. James. *The Principles of Psychology*. Harvard University Press, 1890.

[19] O. Jesorskya, K. Kirchberg, and R. Frischolz. Robust face detection using the Hausdorff distance. In *AVBPA*, pages 90–95, 2000.

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

[21] T. Leung, M. Burl, and P. Perona. Finding faces in cluttered scenes using random labeled graph matching. In *ICCV*, pages 637 – 644, 1995.

[22] E. Murphy-Chutorian and M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE T. PAMI*, 31(4):607 – 626, 2009.

[23] D.I. Perrett, J.K. Hietanen, M.W. Oram, and P.J. Benson. Organization and functions of cells responsive to faces in the temporal cortex. *Phil. Trans. Roy. Soc. B*, 335:23 – 30, 1992.

[24] S.A. Poffel and H.J. Cross. Neurolinguistic programming: a test of the eye movement hypothesis. *Perceptual and Motor skills*, 61(3):23 – 30, 1985.

[25] J.M. Saragih, S. Lucey, and J.F. Cohn. Deformable model fitting by regularized landmark mean-shift. *IJCV*, 91:200 – 215, 2011.

[26] J. Sturt, S. Ali, W. Robertson, D. Metcalfe, A. Grove, C. Bourne, and C. Bridle. Neurolinguistic programming: systematic review of the effects on health outcomes. *British Journal Of General Practice*, 62(604):757 – 64, 2012.

[27] M. Turkan, M. Pardas, and A. Enis Cetin. Edge projections for eye localization. *Optical Engineering*, 47:047007, 2004.

[28] G. Underwood. *Cognitive Processes in Eye Guidance*. Oxford Univ. Press, 2005.

[29] R. Valenti and T. Gevers. Accurate eye center location and tracking using isophote curvature. In *CVPR*, pages 1–8, 2008.

[30] R. Valenti and T. Gevers. Accurate eye center location through invariant isocentric patterns. *IEEE T. PAMI*, 34(9):1785–1798, 2012.

[31] M. Valstar, B. Martinez, X. Binefa, and M. Pantic. Facial point detection using boosted regression and graph models. In *CVPR*, page 2729Ű2736, 2010.

[32] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57(2):137–154, 2004.

[33] P. Wang, M. B. Green, Q. Ji, and J. Wayman. Automatic eye detection and its validation. In *IEEE Workshop on FRGC, CVPR*, page 164, 2005.

[34] U. Weidenbacher, G. Layher, P. Strauss, and H. Neumann. A comprehensive head pose and gaze database. In *IET International Conference on Intelligent Environments.*, pages 455–458, 2007.

[35] L. Wolf, Z. Freund, and S. Avidan. An eye for an eye: A single camera gaze-replacement method. In *CVPR*, pages 817 – 824, 2010.

[36] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *CVPR*, pages 2879 – 2886, 2012.