Automatic Blinking Detection towards Stress Discovery

Alvaro Marcos-Ramiro
University of Alcala
Spain
amarcos@depeca.uah.es

Daniel Pizarro-Perez
University of Clermont-Ferrand
France
dani.pizarro@gmail.com

Marta Marron-Romera
University of Alcala
Spain
marta@depeca.uah.es

Daniel Gatica-Perez
Idiap Research Institute and
EPFL
Switzerland
gatica@idiap.ch

ABSTRACT

We present a robust method to automatically detect blinks in video sequences of conversations, aimed to discovering stress. Psychological studies have shown a relationship between blink frequency and dopamine levels, which in turn are affected by stress. Task performance correlates through an inverted U shape to both dopamine and stress levels. This shows the importance of automatic blink detection as a way of reducing human coding burden. We use an off-the-shelf face tracker in order to extract the eye region. Then, we perform per-pixel classification of the extracted eye images to later identify blinks through their dynamics. We evaluate the performance of our system with a job interview database with annotations of psychological variables, and show statistically significant correlation between perceived stress resistance and the automatically detected blink patterns.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and behavioral sciences; G.3 [**Mathematics and Computing**]: Miscellaneous; D.2.8 [**Image Processing and Computer Vision**]: Metrics—*Scene analysis*

General Terms

Blink detection; stress discovery; random forests

1. INTRODUCTION

Psychological studies have shown that nonverbal communication encodes information relative to the internal state of a person [7], and that humans unconsciously perceive nonverbal cues in order to form impressions of others. In situations like job interviews its importance is highlighted, as the interaction time is limited [4].

ICMI'14, November 12–16, 2014, Istanbul, Turkey.

ACM 978-1-4503-2885-2/14/11 ...$15.00.
http://dx.doi.org/10.1145/2663204.2663239.

In this work, we focus on the effect of involuntary blinks, and present a new method for their automatic detection in videos of conversations. Involuntary blinking has shown to be related with dopamine levels, which can be used as a proxy for stress discovery [1, 6]. Given that task performance has been shown to be correlated to both dopamine and stress levels through an inverted U shape [2], blinking detection can provide insights about stress in a non-intrusive fashion. Blink frequency and consistency along time are particularly revealing [6].

Traditionally, in psychological studies there has been the need for an annotator, which manually codes events in videos. Experimental video data to label can often be in tens (or even hundreds) of hours long, resulting in a very tedious task to endure. To address this problem, we build a real time, automatic blink detector from videos through computer vision and machine learning techniques, which is our first contribution. We first track and extract the eye region from frontal videos with an off-the-shelf face tracker [16]. Then, we normalize the color appearance of this region of interest, and we employ a non-linear Random Forest classifier in order to segment the different parts of the eye. Finally, we analyze the dynamics of the segmented eyes and evaluate the performance with a 11.5 hour real job interview corpus that has annotations, amongst other attributes, of stress resistance and personality [11, 12]. In addition, as our second contribution we show the viability of our approach by demonstrating significant correlation between the automatically discovered blink patterns and the perceived resistance to stress of job candidates.

The rest of the paper is structured as follows: in Section 2 the previous works are reviewed, in Section 3 we present our method, in Section 4 the used data is described, in Section 5 the results are presented and discussed, and finally we provide conclusions in Section 6.

2. PREVIOUS WORK

Given the psychology background presented in the previous section, and that blinking can be used to design human-machine interfaces in situations of reduced mobility, some works in the literature address the problem of blinking detection by means of computer vision.

Existing systems can be grouped into active sensing, in which special illumination or intrusive systems are used; and passive sensing, in which color images are used. Active sens-

ing like infrarred illumination [15] might cause damage to the retina or be intrusive. We focus on passive methods.

In passive methods, previous approaches include skin color models [8], template matching [5], generalized projection functions [17], or eye detection using Haar-like features [9].

To our knowledge, the most similar work to ours is [14], in which blink detection is also motivated towards inferring higher level psychological constructs. Hidden Markov Models and Support Vector Machines are used to estimate the likelihood of an open or close eye, given its dynamics. This method is more computationally complex than ours, as it involves the use of Histogram of Oriented Gradients, Gabor filters, and optical flow.

3. METHOD

Our automatic blink detection system consists of a three step process. First, the face is tracked and an eye region of interest is extracted. Then, using the eye region of interest as input to a classifier, the different parts of the eye are segmented. Finally, the variation of openness of the eye lid is measured to later detect blinks. The process is described with more detail as follows.

3.1 Face tracking and eye region extraction

We obtain and track a series of face points with an off-the-shelf method [16], from which we extract the eye region image $\mathbf{I}_e$. We first use the interocular line as reference (see Figure 1, in green), since it is the only face measure that is not affected by face deformations associated with changes of expression. Then, we obtain the head tilt angle and compensate for it, in order to have a stable region of interest (see Figure 1). We use the normalized V channel after converting the eye region into the HSV color space, and scale $\mathbf{I}_e$ to 147×51 pixels (as it is the average region of interest size) to provide consistency across different subjects and face poses.

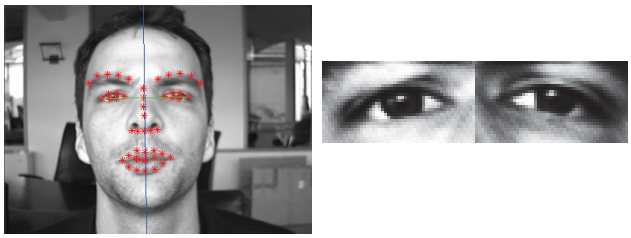

Figure 1: Eye region extraction. Left: face tracking in an image taken from the BioID Face Database. Right: extracted eye region of interest.

3.2 Eye segmentation

Similar to [13], we use an offset sampling idea and a Random Forest classifier in order to associate every pixel of $\mathbf{I}_e$ with an eye part label: iris and pupil, sclera (white part), eyelids, and background (i.e. non-eye regions). An example can be seen in Figure 2. Random Forests have shown real time performance and very good generalization capabilities.

A set $\mathcal{U}$ of pixel offset features is built: $\mathcal{U} = \{\vec{u_{\delta i}}\}_{i=1}^{n_\delta} = \{(u_{\delta i}, v_{\delta i})\}_{i=1}^{n_\delta}$. For a given pixel $\vec{u}$, the feature response is computed with feature parameters $\vec{u_{\delta i}}$ that describe a number of n_δ 2D pixel offsets $(u_{\delta i}, v_{\delta i})$. In [13], features are normalized with the distance to the camera in order to make them depth-invariant. We use contrast values instead, thus defining local contrast differences as per-pixel features.

Let $L(\vec{u}, \vec{u_{\delta i}}, \mathbf{I}^1)$ be a lookup function that returns the feature associated with pixel $\vec{u}$, given a single-channel image $\mathbf{I}^1$ and an offset $\vec{u_{\delta i}}$ (that is, $\mathbf{I}_e(\vec{u} + \vec{u_{\delta i}})$). A feature f_e for a given pixel $\vec{u}$ is then expressed as:

$$f_e(\vec{u}|\vec{u_{\delta i}}) = L(\vec{u}, \vec{u_{\delta i}}, \mathbf{I}_e)\frac{1}{\mathbf{I}_e(\vec{u})}. \quad (1)$$

This feature encodes local appearance as contrast differences within a spatial window. It provides a rich description of the several eye parts, given the high contrast between the iris/pupil and sclera.

3.2.1 *Training the classifier*

A subset of data from the real job interview dataset is annotated in order to serve as a training set. For classification, annotations consist in a number of manually-labeled pixels in the image, in which the nature of each label corresponds to a given eye part. We compute the previously introduced per-pixel offset features for each labeled pixel in the training subset. We then train a Random Forest with the extracted features and the associated eye part labels. Given an unseen image $\mathbf{I}_e$, the classifier outputs the per-pixel predicted eye part and an associated confidence score (see Figure 2).

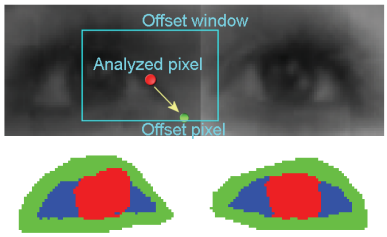

Figure 2: Classification features. Top: offset feature example. Bottom: manually annotated labels used to train the classifier: iris and pupil (red), sclera (blue), eyelids (green), and background (white). Best viewed in color.

3.3 Blinking dynamics

Blinking is defined as a quick closing and opening of the eyes. Therefore, we assume that the accumulated variation of openness of the eye along short periods of time peaks during blinking. We define an eye openness measure $e_{o,t}$ at time t as the number of pixels classified as pupil/iris and sclera, since they constitute the visible parts of the eye when it is open. The differential feature $e_{k,t}$ is therefore defined as the difference of openness between two time instants: $e_{k,t} = e_{o,t} - e_{o,t-1}$. Finally, we accumulate $e_{k,t}$ along a three-frame time window, in order to obtain the short-term accumulated feature $e_{a,t}$.

Since images $\mathbf{I}_e$ are normalized to the same size, a global threshold can be used to segment the $e_{a,t}$ signal, and works well in practice. Blinks are detected in instants in which the threshold is exceeded, resulting into the b_t signal. It is set to 1 during blinking, and 0 otherwise. See Figure 3 for an example, in which an empirically-set threshold of 0.1 is used.

4. DATA

We employed the data used in [11, 12]. An intelligent room and a set of experiments to record information of 60 participants in a real, for a sales-like job interviews were designed.

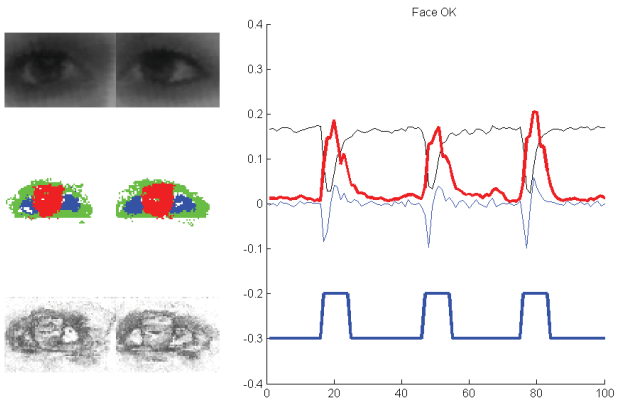

Figure 3: Blinking dynamics. Left: eye region of interest and classification results. Right: eye movement dynamics (thin black line: $e_{o,t}$; thin blue line: $e_{k,t}$; red line: accumulated variation of openness $e_{a,t}$; thick blue line: blink detection signal b_t). Best viewed in color.

An open position was advertised local universities using multiple communication channels. The average participant age was 24 years. Before starting the interview, applicants filled in a consent form. They then completed a questionnaire to assess their Big-Five personality scores [7], encoded with a 1-5 scale.

The interviews are dyadic, with the recruiter facing the job applicant. The protagonists were seated at both sides of a table. A total of 11.5 hours were recorded in 720p resolution. For the interview itself a structured design was used, where the sequence of instructions and questions remained constant across interviews in order to ensure that comparisons could be made between job candidates. The job applicants were asked to answer four behavioral questions related to past experiences in situations that require specific social skills, including resistance to stress, intelligence, and communication skills, in a five-point scale. Human resources experts annotated the perceived attributes of the participants while looking at audio only, at video only, and at both.

5. RESULTS

5.1 Experiments and measures

We define two experiments to assess the performance of our proposal: **experiment #1** is designed to evaluate the performance of the blink detection system. Blink events in multiple sequences containing different subjects from the real job interview corpus are manually labeled, and used as groundtruth to be compared with the automatic detections. Subjects are different to those used during training. In total, 330 seconds of video (five and a half minutes) and 10 subjects with different skin tones are employed for testing, containing 133 blinking events. The Random Forest classifier is trained with a disjoint subset of 32 eye images from 11 different subjects. As performance measure, we chose global accuracy and the F_1 score, which is defined as the harmonic mean between precision and recall:

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}. \quad (2)$$

F_1 is a strict measure that takes into account class balance. This is specially relevant in our case since blink events constitute very little proportion of the total video length. We define correct detections as blinking events that are detected within a four frame window of the blinking events contained in the groundtruth, as we accumulate the eye energy signal $e_{a,t}$ over a three-frame window, resulting in a slight delay in the detection. Performance results can be seen in Section 5.2.

In **experiment #2**, we assess the validity of our method to extract measures related to stress signals. We extract a series of features from the automatically extracted blink information, which we then correlate with the job candidates data obtained from the questionnaire data described in Section 4. The features that we extract are: (1) Shannon's entropy of the whole blinking event signal b_t; (2) entropy, mean, and standard deviation of the time lapses between blinking events; (3) entropy, mean, and standard deviation of the accumulated signal $e_{a,t}$.

Since the literature has shown that the regularity of blinking patterns is a proxy for stress discovery, we hypothesized that extra information can be extracted from the entropy of the signals.

5.2 Results and discussion

When evaluating our system with **experiment #1**, we obtain a F_1 value of 93.65% for event-based detection. We get 5 false positives (3.76%) and 11 undetected blinks (8.27%), out of the total 133 events. Upon visual inspection, the errors come from two sources: face tracking misplacements and eye part classification malfunctions due to a high rotation angle between the face and the camera. The latter could be corrected by using similar time moments during the training phase of the per-pixel classifier, while the former can only be addressed by discarding the mis-tracked periods through outlier detection. Nevertheless, the obtained F_1 value remains highly satisfactory, specially given the complexity of the task (high appearance variance across different subjects, face rotations, looking down to the table, or the distance from the face to the camera). In Figures 4 and 5, quantitative and qualitative results can be seen.

After conducting **experiment #2**, we present a series of statistically significant ($p < 0.05$) correlations, see Table 1. To obtain the correlation coefficients, we discarded lost tracking periods for the face, and used data for the 60 participants.

Table 1: Perceived attributes and modulus of the correlation coefficient obtained with our features (60 subjects).

Attribute	Corr	Feature
Stress resistance (audio)	0.27	Entropy of $e_{a,t}$
Openness to experience	0.31	Entropy of blink time lapses
Agreeableness	0.26	Standard deviation of $e_{a,t}$
Agreeableness	0.27	Mean of blink time lapses
Intelligence	0.28	Entropy of b_t
Communication skills	0.29	Entropy of $e_{a,t}$
Conscientiousness	0.33	Mean of $e_{a,t}$

The significant correlation between perceived stress resistance from annotating only with the audio channel validates our hypothesis, while also highlighting the multimodal nature of the stress perception (which is consistent with previous works [10]). Regarding correlation with other variables, interestingly, the highest correlation coefficient is obtained

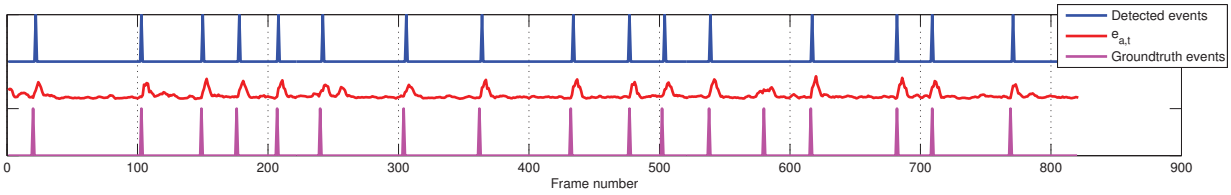

Figure 4: Blink extraction for one of the test sequences. The accumulated eye movement signal is used to estimate blink events.

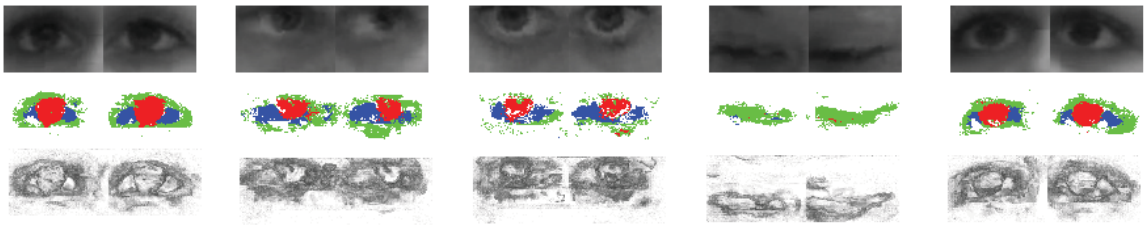

Figure 5: Qualitative results. Top row: eye region of interest. Central row: eye segmentation. Bottom row: classification confidence.

for conscientiousness. This goes in line with the literature on deception detection from blinking [14], given the known relationship between deception and conscientiousness levels [3].

The results also show that by processing directly the signal $e_{a,t}$, we are able to extract more information than just by using the binary blink events. This allows for a higher granularity in the intensity and duration of the eye blinks. As future work, we plan to approach the problem as a classification or regression task, in order to quantify the stress levels. In addition, blinking detection on subjects wearing glasses should be explored, given the problems that their use arise, such as reflexions and occlussions.

6. CONCLUSIONS

We have presented an automatic blink detection system towards stress discovery by using a computer vision approach. We first obtain track the face with an off-the-self method in order to extract the eye region of interest. Then, we employ a Random Forest classifier to segment the eyes into their different parts. Finally, we analyze the dynamics of the blinking thanks to their energy signal. We evaluated our computer vision approach with a challenging dataset, showing high performance in an unconstrained conversational setting. Then, we found moderate, yet significant correlation between blink detections and expert human annotations. This result provides a starting point to further investigate the connections of stress, amongst other attributes, with automatic blink discovery.

7. ACKNOWLEDGMENTS

This research was funded by SNSF UBImpressed project, the Spanish Ministry of Science and Innovation under project VISNU (ref. TIN2009-08984) and the University of Alcala FPI internship program.

8. REFERENCES

[1] C. Adler, I. Elman, N. Weisenfeld, L. Kestler, D. Pickar, and A. Breier. Effects of acute metabolic stress on striatal dopamine release in healthy volunteers. *Neuropsychopharmacology*, 22, 2000.

[2] P. L. Broadhurst. Emotionality and the yerkes-dodson law. *Journal of Experimental Psychology*, 54, 1957.

[3] R. B. Cattell. *The Scientific Analysis of Personality*. Penguin, London, 1965.

[4] J. Curhan and A. Pentland. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first five minutes. *Journal of Applied Psychology*, 2007.

[5] M. S. Devi and P. R. Bajaj. Driver fatigue detection based on eye tracking. In *International Conference on Emerging Trends in Engineering and Technology*, 2008.

[6] C. Karson. Spontaneous eye-blink rates and dopaminergic systems. *Brain*, 106, 1983.

[7] M. Knapp and J. Hall. *Nonverbal Communication in Human Interaction*. 2009.

[8] A. Krolak and P. Strumillo. Fatigue monitoring by means of eye blink analysis in image sequences. In *ICSES*, 2006.

[9] A. Krolak and P. Strumillo. Eye-blink detection system for human-computer interaction. *Universal Access in the Information Society*, 11, 2012.

[10] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *ACM Ubiquitous Computing*, 2012.

[11] L. Nguyen, A. Marcos, M. Marron, and D. Gatica. Multimodal analysis of body communication cues in employment interviews. In *ACM International Conference on Multimodal Interaction (ICMI)*, 2013.

[12] L. S. Nguyen, D. Frauendorfer, M. Schmid Mast, and D. Gatica-Perez. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Transactions on Multimedia*, 2014.

[13] J. Shotton, T. Sharp, A. Kipman, A. W. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. In *IEEE CVPR*, 2011.

[14] Y. Sun, S. Zafeiriou, and M. Pantic. A hybrid system for on-line blink detection. In *Hawaii International Conference on System Sciences*, 2013.

[15] P. Thoumie, J. R. Charlier, M. Alecki, D. D. Erceville, A. Heurtin, J. F. Mathe, G. Nadeau, and L. Wiart. Clinical and functional evaluation of a gaze controlled system for the severely handicapped. *Spinal Cord*, 36, 1998.

[16] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *IEEE CVPR*, 2013.

[17] Z.-H. Zhou and X. Geng. Projection functions for eye detection. *Pattern Recognition*, 37, 2004.