

Eye tracking and pupil size variation as response to affective stimuli: a preliminary study

Antonio Lanatà, Antonino Armato, Gaetano Valenza and Enzo Pasquale Scilingo
Interdepartmental Research Centre "E.Piaggio" Faculty of Engineering, University of Pisa, Pisa, Italy
Email: a.lanata@centropiaggio.unipi.it

Abstract—Eye Gaze Trackers (EGTs) and pupil size variation are generally developed for scientific investigation in controlled environments or laboratories and data are used in several fields of application such as ophthalmology, neurology, or psychology, with the aim of studying oculomotor characteristics and abnormalities. Very often, the focus of these studies is the identification of cognitive and mental states. This preliminary work aims at investigating if eye tracking and pupil size variation can provide useful cues to discriminate emotional states induced by viewing images at different arousal content. Here we propose a new wearable and wireless EGT, hereinafter called HATCAM, able to robustly enable eye tracking and pupil area detection. Although very preliminary, results are very promising for affective computing applications.

I. INTRODUCTION

Emotions are important psychological conditions that reflect several human states, such as pleasant or unpleasant feelings, human relations, process and results of action. They are present in all mental processes, and any human activity manifestation is accompanied by emotional experiences. Many research works have shown that emotional processing can have primacy over cognition [1]. The famous naturalist Darwin stated that emotions emerged in the course of evolution as the means by which living creatures determine the significance of certain conditions to meet their urgent needs [2]. They are the most important factors in the regulation of cognition. In addition, several works have shown how emotion regulation is an essential feature of mental health. In particular, they highlighted how emotion and its regulation have an important role in various aspects of normal functioning. For example, emotions become dysregulated in major depressive episodes, and some theoretical views of depression are based on emotion changes which have implications in assessment, treatment, and prevention of the pathology [3]. Moreover, it has been shown the strong relationship between emotion and anxiety [4] as well as brain damages of emotional processing areas and decision-making process [5]. Emotion recognition, by means of intelligent systems, is a crucial issue to be addressed for understanding human behavior, investigating mental health, interpreting social relations, etc. There are several physiological correlates of the autonomic nervous system (ANS), e.g., heart rate variability, electrodermal activity, pupil size and eye movement variation, with subject behavior or emotional status [6].

The associations of emotions and physiological reactions controlled by the ANS are complex, but anger, for example,

has been associated with higher heart rate than happiness, and on the other hand, anger has been associated with higher finger temperature than fear [7],[8]. In this work we aim at investigating the relationship between emotions and information coming from the eyes, i.e. pupil size variation and eye tracking. It is known that pupil dilations and constrictions are governed by the ANS [9]. Previous studies have suggested that pupil size variation is also related to both cognitive and affective information processing. Nowadays, eye-tracking technology development (e.g. ease of use, improved accuracy, and enhanced sampling rate), offers the possibility for unobtrusive monitoring of emotion-related reactions because no sensors need to be attached to the user. In order to be able to evaluate the possibilities of using pupil size measurement as well as pupil tracking for detecting emotional responses, we need to understand how emotions and eye features relate to each other.

Concerning eye tracking, people can consider and discard various aspects of a task rather quickly (in less than 200 ms), eye movements can provide detailed estimates of what information an individual is considering. Eye tracking is becoming an increasing popular measure of cognitive and affective information processing [10]. By gathering data on the location and duration of eye fixations, many inferences about the cognition structure could be done. The use of eye tracking in estimating cognitive or affective states can be focused on two assumption: the immediacy (people process information as it is seen) and the eye-mind (the eye remains fixated on an object while the object is being processed). A variety of eye-tracking methods exist. In terms of the data collected from the eye, two popular methods are: 1) shining a light on the eye and detecting corneal reflection and 2) simply taking visual images of the eye and then locating the dark pupil area.

Generally, the choice of the best method depends upon the external lighting conditions. To compute where a person is fixating, there are three popular methods. The first method simplifies the calculations by having fixed geometries by forcing the person to hold still by biting on a bar or putting the head in a restraint. In the second method the person wears a head sensor that tracks the head orientation and location in three dimensions and then combines this information with eye-direction. The third method, used in our experiments, places the eye-tracking apparatus on the person head along with a camera so that a visual image is captured showing what the person is currently looking at, with a point on the image indicating the object being fixated. A crucial issue is

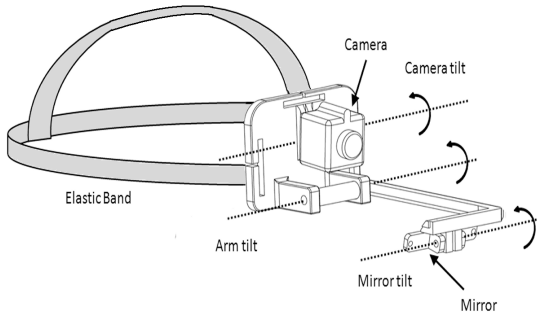


Fig. 1. HATCAM configuration.

the methodology for eliciting different emotional states. In this view, several works, [11], report a set of systematically studied affective image stimuli belonging to the International Affective Pictures System (IAPS). These stimuli have been studied using the Self-Assessment Manikin, which is a method to study differences in affective experiences using nine-point bipolar rating scales. In these works, subjects look at the stimuli and ranked them on three bipolar dimensions: emotional valence, arousal, and dominance. However, valence and arousal are the most frequently used dimensions to capture the nature of emotional information. The valence dimension varies from negative to positive emotional experience, and the arousal dimension varies from calm to highly excited. This kind of elicitation is used in our experiments. This work is a preliminary study which aims at investigating eye tracking and pupil area variation in response to stimulation using images from IAPS. We wanted to explore the association between eye information and emotional image categories. More specifically, by using an innovative head-mounted eye tracking system (named HATCAM) we were able to acquire pupil variation together with eye gaze trajectory and time of fixation as well, during exposition of subjects to affective images. The goal of the work was to identify characteristic features from pupil size variation and eye tracking and by means of classification methods to distinguish the neutral from arousal elicited states.

II. HATCAM - WEARABLE EGT SYSTEM

The system used in this experiment, HATCAM, is a new wearable and wireless eye tracking system which can be tailored to both adults and children. It is comprised of only one lightweight camera able to capture, by means of a mirror, the eyes of the subject and the scene in front of him, simultaneously. The system configuration is shown in fig. 1. In detail, the system is comprised of a wireless CMOS camera (CP294) having low weight (20g), low size (2x2x2cm), and an A/V transmitter up to 30m of distance. The camera has a resolution of 628x586 pixels with $F2.0$, $D45^\circ$ optic, and 25 frames per second (f.p.s.). The original lens of the camera was removed and substituted with a wide-angle-lens without IR filter. This operation allows enlarging the view angle and acquiring infrared components, which emphasize the

contrast between pupil and iris. This system is able to capture simultaneously, without latency, the visual scene in front of the subject and the position of his eyes. This is achieved using a mirror (4x0.6cm) placed on a shaft linked to the head (see fig. 1). Tilt and shaft of the mirror and the camera orientation can be tailored to the forehead profile of the user (see fig.1).

III. EXPERIMENTAL PROTOCOL

Ten subjects, nine males and one female, volunteered to participate in the experiment. All subjects do not suffer from evident mental pathologies. Six subjects had dark eyes and 4 had bright eyes. The average age was of 26.8 with a standard deviation of 1.5. The experiment was performed in a room with illumination condition achieved by white neon lighting equally distributed in the room with a power of 50 lumens. Subjects were asked to sit on a comfortable chair in front of a screen at a fixed distance of 70 cm. The HATCAM system was also equipped with a chin-support in order to avoid unwanted head movements. They were presented with a sequence of images, gathered from the IAPS database, while wearing headsets for acoustical insulation. The slideshow is comprised of 5 sessions of images N, A, N, A, N , where N is a session of 5 neutral images, and A are sets of 5 images having maximum level of arousal and the lowest valence, i.e. high negative affective impact. During the experimental test all the subjects were asked to look at the picture which appeared on the screen for 10 seconds. Each trial lasted about 25 minutes.

IV. PUPILLOMETRY AND GAZE POINT

This section deals with the processing techniques used to detect the center of the eye and how its movements are mapped into the image plane. This technique is often referred to as Video Oculography (VOG) and involves visible spectrum imaging. It is a passive approach that captures ambient light reflected from the eye. The mounted camera is modified to acquire also the IR components of natural light. Therefore, the system keeps the advantages of IR illumination in increasing the contrast between pupil and iris, and at same time preventing any possible injuries due to artificial IR illuminators, which are not required because of the presence of natural light. Figure 2 shows the block diagram of the algorithmic process used to classify visual stimuli having different affective arousal. The upper block implements the pupillometry and gaze point identification. The outputs are then processed to extract a specific set of features used for the classification. More in detail, the pupillometry and gaze point block is comprised of a sub-chain of blocks implementing eye extraction algorithm, photometric normalization algorithm of illumination, pupil contour and mapping of the eye center into the image scene. In the next sections, each block is described more in depth.

A. Eye Extraction

Figure 3 shows how the HATCAM is able to acquire simultaneously the eyes of the user and the scene in front of him using the mirror. Eye extraction procedure is constituted

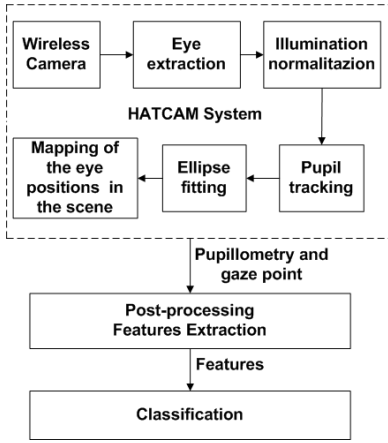


Fig. 2. Block diagram showing all the algorithmic stages of the processing of eyes and outside scene.



Fig. 3. Example of a single frame captured by the camera. The rectangular area marked by red represents the ROI.

of visual inspection of the first video frame, in which a rectangular area including the eye is freehand selected (see fig. 3 and 4). This region is called Region Of Interest (ROI). Being the system mounted on the head the ROI does not change throughout the experiment. In addition, only the red-image-component (see fig. 4a) is converted in gray scale (see fig. 4b) and used as input to the other processing blocks, as this component is specifically helpful in enhancing the contrast between pupil and background.

B. Photometric normalization technique

The purpose of the illumination normalization is to reduce or eliminate some variations in the captured eyes due to different conditions of illumination. It normalizes and enhances the eye image to improve the recognition performance

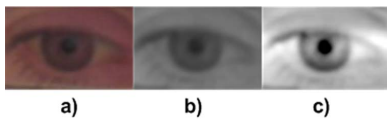


Fig. 4. Red component of the ROI.

of the system. In this work, we use the Discrete Cosine Transform (DCT) already proposed by Chen et al. [12]. This approach is based on the Retinex theory (from the words "retina" and "cortex", suggesting that both eye and brain are involved in the processing) developed by Land, [13]. This theory is based on color constancy assumption which ensures that the perceived color of objects remains relatively constant under varying illumination conditions. Land and his colleagues assume that the stimulus is not the result of the light source and surface reflectivity only, but that the visual system processes the stimulus, integrating the spectral radiance and generating a ratio of integrated radiance of any region of the scene with that of the brightest region. This stimulus is called *lightness*. This model eliminates the effect of a non uniform illumination and is completely independent of any a-priori knowledge of the surfaces reflectance and light source composition. According to this theory, the image intensity $I(x, y)$ can be simplified and formulated as follows:

$$I(x, y) = R(x, y)L(x, y) \quad (1)$$

where $R(x, y)$ is the reflectance and $L(x, y)$ is the illuminance at each point (x, y) . The luminance L is assumed to contain low frequency components of the image while the reflectance R mainly includes the high frequency components of the image. The DCT technique compensates for illumination variations by truncating the low frequency components of the discrete cosine transform (DCT) in the logarithm domain. In the logarithm domain the theory is formulated as follows:

$$\log I(x, y) = \log(R(x, y)) + \log(L(x, y)) \quad (2)$$

Fig. 4c shows the output of the DCT algorithm applied to gray scale image reported in fig. 4b.

C. Pupil tracking

Pupil tracking algorithm extracts the contour of the pupil exploiting the higher contrast of the pupil than the background due to the IR components of the natural light. Figure 5 shows the algorithm block diagram. More in detail, the first block binarizes the image by means of a threshold. Fig. 6 reports the histogram of the eye, i.e. the distribution of the image pixel vs the gray levels from 0 to 255. The threshold divides the histogram into two groups of pixels having only two levels of gray; the zero level should group all the pixel belonging to the pupil whereas the 255 level should identify the background. The criterion implies choosing the threshold as the absolute minimum value in the range comprised between the two highest peaks of the eye histogram as reported in fig. 6. An example of the binarization process is reported in fig.7. After binarization, two sheafs of lines starting from the middle points of the vertical sides of the image, with an angular aperture of 30° , are drawn. As result of the binarization process, the image borders are expected to belong to the background, therefore the starting point of each line has a value of 255 in terms of gray level. Analogously, the pupil is expected to be placed roughly in the middle of the image (this is assured by the an accurate freehand selection of the ROI). When each line

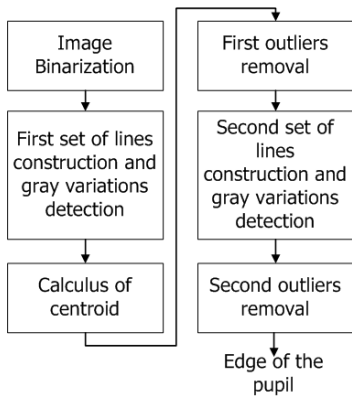


Fig. 5. Block diagram of the pupil tracking algorithm.

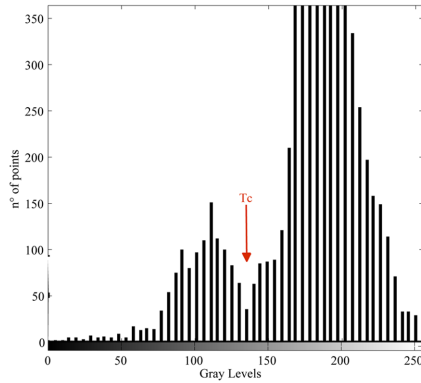


Fig. 6. Example of the histogram of the eye: T_c refers to the threshold identifying the eye and the sclera region.

encounters, along its path, a dark pixel, this latter can be thought to belong to the contour of the pupil. Afterwards, the centroid of these points is calculated. After removing all the outliers, being these points very far from the centroid with respect to the large point density (pupil edge), a large-grain approximation of the contour is obtained. Next, a sheaf of lines starting from the centroid with an angular aperture of 360° , and detects all discontinuities, but now from black to white. Finally, outliers are again removed. The result of this algorithm is a set of points constituting the pupil edge. This set will be the input of the fitting algorithm (see 8).



Fig. 7. Example the eye image after the binarization process.

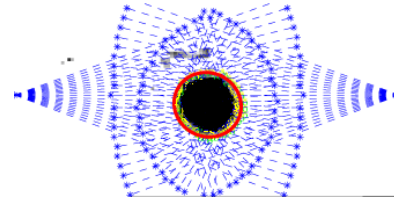


Fig. 8. Pupil tracking algorithm. Sheafs of lines are in blue; black points identify the eye including the outliers; yellow points highlight the pupil contour which is interpolated by the ellipse marked up in red.

D. Ellipse fitting

Ellipse fitting algorithm is implemented for constructing the pupil contour and detecting the center of the eye. Ellipse is considered as the best geometrical figure approximating the eye contour. According to the ellipse construction, it can be expressed by an implicit second order polynomial, being a central conic (with $b^2 - 4ac < 0$), such as:

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (3)$$

Ellipse fitting algorithms present in literature can be divided into two broad techniques: the clustering/vot-ing (CV) and the least square (LS) techniques. The first one uses two main approaches such as RANSAC and Hough Transform which are extremely robust but they are time-demanding or excessively resource consuming for real time machine vision [14],[15]. The LS method is based on finding a set of parameters that minimize the distance between the data points and the ellipse. According to literature this technique fulfills the real time requirement. One implementation of the LS technique has been introduced by Fitzgibbon et al., which is a direct computational method (i.e. B2AC) based on the algebraic distance with a quadratic constraint, [16]. In this work, we use a custom B2AC algorithm, where a gaussian noise is added for algorithm stabilization, [17], to calculate the center of the pupil (that coincides with the ellipse center), the axes dimensions as well as the eccentricity.

E. Mapping of the eye position

The mapping procedure associates the eye center position to the image plane of the scene, providing as result the gaze point. An experimenter guides this procedure. Firstly, the camera is positioned to capture both the scene (in our case the screen) and the mirror. In detail, tilt of the camera is adjusted as well as the tilt of mirror shaft and the tilt of the mirror to reflect the eyes. Each participant is asked to look at some specific points of the screen. These points are identified by coordinates $s_i = (x_{si}, y_{si})$ referred to the image plane (i.e. the image plane captured by the camera), (see fig.3). The participants were instructed to keep their head as still as possible and to carefully look at each target point without blinking until looking at the next one.

The mapping function gets as input the center of the eye coming from Ellipse fitting block, and the coordinates of the gaze point on the image plane.

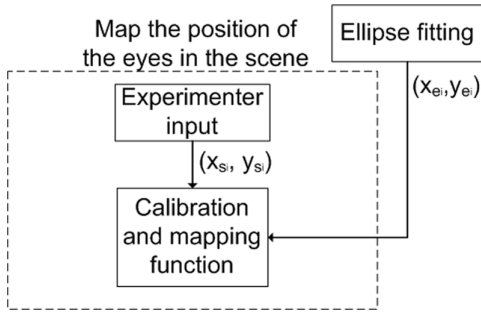


Fig. 9. Block diagram of the mapping function calculation process.

Mapping functions are quadratic polynomials defined as:

$$x_{si} = a_{11} + a_{12}x_{ei} + a_{13}y_{ei} + a_{14}x_{ei}y_{ei} + a_{15}x_{ei}^2 + a_{16}y_{ei}^2 \quad (4)$$

$$y_{si} = a_{21} + a_{22}x_{ei} + a_{23}y_{ei} + a_{24}x_{ei}y_{ei} + a_{25}x_{ei}^2 + a_{26}y_{ei}^2 \quad (5)$$

where x_{si} , y_{si} are the coordinates of the image plane (i.e. the coordinates of the point on the screen mapped into the image plane captured by the camera), and x_{ei} , y_{ei} are the coordinates of the center of the eye coming from the ellipse fitting block, referred to the image plane as well. The coefficients $a_{1,1-t0-6}$, and $a_{2,1-t0-6}$ are unknowns. Since each calibration point defines 2 equations, the system is over constrained with 12 unknowns and 18 equations, and can be solved using Least Square Method (LSM).

F. Post-processing and feature extraction

Post-processing phase is based on parameters extracted by Recurrence Quantification Analysis (RQA), along with fixation time and pupil area detection. As described in the experimental protocol, each image of the sequence is watched by the subjects for a time interval of 10 seconds. As the camera acquires with 25 fps, we obtain a grand total of 250 frames for each image. Each image can be represented as a matrix of 628×586 pixels. The gaze point is mapped into a pixel in each frame. We constructed a matrix of 628 rows and 586 column where any position corresponding to the pixel coordinates associated to the gaze point is set to 1. In order to minimize errors due to the eye blinking, only the pixels retained for at least five frames (0.2 sec) are set to 1.

1) *Recurrence Quantification Analysis*: RQA [18] is a nonlinear analysis method which quantifies the number and duration of recurrences of a dynamical system represented by state space trajectories.

Let RP be the Recurrence Plot matrix. Let $L_{ij}(x, y)$ be the set of eye-gaze points from the frame i of the image j , where x and y are spatial coordinates. We applied the RQA to the set $G_m(x, y)$ defined as:

$$G_m(x, y) = \bigcup_i L_{ij}(x, y)$$

Therefore, for each image the following features are calculated:

Recurrence Rate (RR) is the percentage of recurrence points in an RP and it corresponds to the correlation sum:

$$RR = \frac{1}{N^2} \sum_{x, y=1}^N G_m(x, y)$$

where N is the number of points on the phase space trajectory.

Determinism (DET) is the percentage of recurrence points which form diagonal lines:

$$DET = \frac{\sum_{l=lmin}^N lP(l)}{\sum_{i,j=1}^N R_{i,j}} \quad (6)$$

where $P(l)$ is the histogram of the lengths l of the diagonal lines.

Trapping Time (TT) is the average length of the vertical lines:

$$TT = \frac{\sum_{v=vmin}^N vP(v)}{\sum_{v=vmin}^N P(v)} \quad (7)$$

Averaged diagonal line length (L) is the average length of the diagonal lines:

$$L = \frac{\sum_{l=lmin}^N lP(l)}{\sum_{l=lmin}^N P(l)} \quad (8)$$

Entropy (ENTR) is the Shannon entropy of the probability distribution of the diagonal line lengths $p(l)$:

$$ENTR = - \sum_{l=lmin}^N p(l) \ln p(l) \quad (9)$$

Longest diagonal line (L_{max}) The length of the longest diagonal line:

$$L_{max} = \max(\{l_i; i = 1, \dots, N_l\}) \quad (10)$$

where N_l is the number of diagonal lines in the recurrence plot.

2) *Fixation time*: While watching each image, subject eye can be caught by specific details. We define as fixation time of each pixel, the absolute time during which the subject is dwelling on that pixel. We obtain a statistical distribution of fixation time over the the fixed pixels, whose mode, which we define here as $Tmax$, is used as additional feature. $Tmax$ is calculated for each image and each subject during both neutral and arousal elicitation as:

$$Tmax = \max_{i=0}^N (t(P_i)) \quad (11)$$

where N is the number of points of gaze in the image, P_i is the i_{th} point of gaze, $t(P_i)$ is the fixation time of the i_{th} point of gaze, respectively.

3) *Pupil Area Detection*: The pupil area was approximated as an ellipse whose area is calculated for the pupillometry. To increase the robustness of the algorithm an averaged areas of both eyes are considered:

$$A_p = \frac{\pi r_a^l r_b^l + \pi r_a^r r_b^r}{2} \quad (12)$$

where A_p is the pupil area, r_a^l and r_b^l are the ellipse semi-axes of the left eye, r_a^r and r_b^r are the ellipse semi-axes of the right eye.

G. Classification

This work aims at classifying two classes of images, labeled as neutral and arousal. Pattern recognition was performed by means of non parametric K-Nearest Neighbor (K-NN) algorithm [19]. It is based on the ‘‘proximity’’ concept, i.e. an object is supposed to belong to the closest class in the n -dimensional feature space. After the K-NN training process, the performance of the classification task is commonly evaluated using the confusion matrix [20]. The generic element r_{ij} of the confusion matrix indicates how many times in percentage a pattern belonging to the class i was classified as belonging to the class j . More the confusion matrix is diagonal and better is the classification. The matrix has to be read by columns. The training phase is carried out on 80% of the feature dataset while the testing phase to the remaining 20%. We performed 40-fold cross-validation steps in order to obtain unbiased classification results, i.e. it allowed us to consider gaussian the classification result distributions, which can be therefore described as mean and standard deviation among the obtained 40 confusion matrixes.

Our implementation was performed according to the following steps:

- 1) In the training phase, the K-NN algorithm just stores the training feature sets together with the labels.
- 2) In the test phase, the K-NN algorithm calculates the n Euclidean distances between the new feature set and the n features of the whole training feature set as follows:

$$D_i = \sqrt{(p_{i1} - q_1)^2 + \dots + (p_{in} - q_n)^2} = \sqrt{\sum_{j=1}^n (p_{ij} - q_j)^2} \quad (13)$$

where $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$ is i -th training feature set, $Q = (q_1, q_2, \dots, q_n)$ is the new feature test set, n is the number of features.

Afterwards, the K-NN algorithm finds the K training feature sets that have the minimum distance from the new feature set. Among these, m training feature sets belong to the neutral class and $K - m$ belong to the arousal class. The new feature

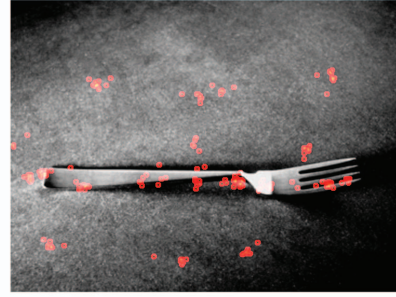


Fig. 10. Example of the points of gaze detected during a neutral elicitation. Gaze points are marked in red.

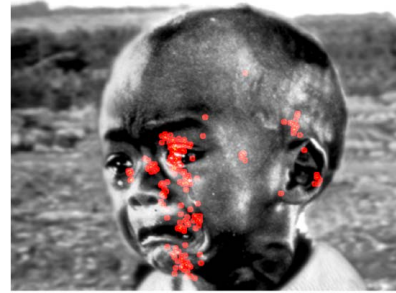


Fig. 11. Example of the points of gaze detected during a arousal elicitation. Gaze points are marked in red.

set is supposed to belong to the neutral class if $m > (K - m)$, i.e. $m > \frac{K}{2}$, to the arousal class otherwise.

V. EXPERIMENTAL RESULTS

Figures 10 and 11 show the map of points of gaze over two sample images, having neutral and the highest value of arousal, respectively. Gaze points are marked in red. Already at glance, the neutral image (fig. 10 shows a more sparse spatial distribution of the gaze points than the arousal image, in which gaze points are mainly concentrated in confined areas. RQA analysis was used to quantify the gaze point distribution for each frame of each image. All the extracted features from RQA and pupil area variation are not normally distributed, as confirmed by the Lilliefors test, [21], which returns a p -value ($p < 0.05$) rejecting the null hypothesis of normality. Accordingly, we used the Kruskal-Wallis test, [22], which is a non parametric one-way analysis of variance by ranks for testing equality of population medians. Kruskal-Wallis is performed on ranked data, so the measurement observations are converted to their ranks in the overall data set. This test does assume an identically-shaped and scaled distribution for each group, except for any difference in medians. The null hypothesis is stated as the probability that the samples come from identical populations, regardless their distributions. In place of the mean of distributions we considered the median as a measure of location [23]. Having only two sets of

features, one for arousal and one for neutral, Kruskal-Wallis test returned the probability that the two samples were not belonging to the same population, in other words, if there was a statistical difference between the two samples.

Median and Median absolute deviation of all RQA features are reported in Table I. Statistical differences between neutral and arousal elicitation was found (* $p < 0.01$ and ** $p < 0.001$).

TABLE I
FEATURES EXTRACTED FROM RQA

Features	Neutral	Arousal
RR*	0.0018±0.0002	0.0019±0.0002
DET**	0.7311±0.0782	0.6373±0.0798
TT**	2.5862±0.9024	2.0345±0.4368
L**	2.9184±0.4564	2.6513±0.3181
ENTR**	1.2592±0.2306	1.0501±0.2041
Lmax **	5.0000±1.2800	5.0000±1.0583
Tmax **	1.9600±0.3509	1.5600 ±0.3501
Pupil È	195.39±19.689	197.27±17.106

In the table II, the confusion matrix obtained from K-NN classifier after 40 fold-cross-validation steps is shown.

TABLE II
CONFUSION MATRIX OF K-NN CLASSIFIER BY USING RQA AND PUPIL SIZE FEATURES

	Neutral	Arousal
Neutral	90.2273±5.9622	20.1136±10.8881
Arousal	9.7727±5.9622	79.8864±10.8881

VI. CONCLUSION AND DISCUSSION

In this work we investigated eye tracking and pupil size variation in response to emotional elicitation induced by IAPS images. In particular, the goal was to identify a set of features from pupil size variation and eye tracking in order to distinguish between neutral to arousal states. In detail, we used an innovative wearable and wireless head-mounted eye tracking system (HATCAM) to acquire pupil variation together with eye-gaze trajectory as well as time of fixation. In addition, we adopted a novel methodology to characterize differences between neutral and arousal elicitation, in eye-gaze acquisitions, by means of features extracted from RQA. This choice is motivated by the analogy between the bi-dimensional image containing eye-gaze points and the matrix commonly used for Recurrence Plot [24]. Both, indeed, are matrixes of zero and ones. In addition to the features from RQA, we included in the feature space for arousal and neutral classification, also the elliptic area of pupil and the fixation time. Pattern recognition was performed by means of K-Nearest Neighbor (K-NN) classifier. After the K-NN training process, the performance of the classification task was evaluated by using the confusion matrix. It was randomized for 40-fold cross-validation steps to avoid bias. Results are reported in Table II. As it can be seen results are very satisfactory. The percentage of successful recognition is about 90% for neutral images and about 80% for images at high arousal. It means that eye gaze, both in terms of pupil tracking and size, can be a viable means to discriminate

different affective states. Future work will progress to exploit eye gaze together with peripheral physiological signals in the field of mental care.

ACKNOWLEDGMENT

This research is partially supported by the EU Commission under contract ICT-247777 Psyche, and partially supported by the EU Commission under contract ICT-258749 CEEDs.

REFERENCES

- [1] R. Zajonc, "On the primacy of affect," *American Psychologist*, vol. 39, no. 2, pp. 117–123, 1984.
- [2] C. Darwin, P. Ekman, and P. Prodger, *The expression of the emotions in man and animals*. Oxford University Press, USA, 2002.
- [3] J. Gross and R. Muñoz, "Emotion regulation and mental health," *Clinical Psychology: Science and Practice*, vol. 2, no. 2, pp. 151–164, 1995.
- [4] R. Lazarus and J. Averill, "Emotion and cognition: With special reference to anxiety," *Anxiety: Current trends in theory and research*, vol. 2, pp. 242–284, 1972.
- [5] A. Damasio, *Descartes' error: Emotion, reason, and the human brain*. Quill New York:, 2000.
- [6] R. Picard, *Affective computing*. The MIT Press, 2000.
- [7] P. Ekman, R. Levenson, and W. Friesen, "Autonomic nervous system activity distinguishes among emotions." *Science*, 1983.
- [8] R. Levenson, "Autonomic nervous system differences among emotions," *Psychological Science*, vol. 3, no. 1, p. 23, 1992.
- [9] J. Andreassi, *Psychophysiology Human behavior and physiological response*. Lawrence Erlbaum Assoc Inc, 2006.
- [10] G. Lohse and E. Johnson, "A comparison of two process tracing methods for choice tasks," in *System Sciences, 1996., Proceedings of the Twenty-Ninth Hawaii International Conference on*, vol. 4. IEEE, 2002, pp. 86–97.
- [11] P. Lang, M. Bradley, and B. Cuthbert, "International affective picture system (IAPS): Technical manual and affective ratings," *NIMH Center for the Study of Emotion and Attention*, 1997.
- [12] W. Chen, M. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, no. 2, pp. 458–466, 2006.
- [13] E. Land and J. McCann, "Lightness and retinex theory," *Journal of the Optical society of America*, vol. 61, no. 1, pp. 1–11, 1971.
- [14] D. Forsyth and J. Ponce, *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference, 2002.
- [15] N. Bennett, R. Burrige, and N. Saito, "A method to detect and characterize ellipses using the Hough transform," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 7, pp. 652–657, 2002.
- [16] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 5, pp. 476–480, 2002.
- [17] E. Maini, "Robust ellipse-specific fitting for real-time machine vision," *Brain, Vision, and Artificial Intelligence*, pp. 318–327, 2005.
- [18] J. Zbilut and C. Webber Jr, "Recurrence quantification analysis."
- [19] T. Cover and P. hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, January 1967.
- [20] R. Kohavi and F. Provost., "Glossary of terms," *Machine Learning*, vol. 30, p. 271 274, 1998.
- [21] H. Lilliefors, "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967.
- [22] W. Kruskal and W. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.
- [23] G. Stavig and J. Gibbons, "Comparing the mean and the median as measures of centrality," *International Statistical Review/Revue Internationale de Statistique*, vol. 45, no. 1, pp. 63–70, 1977.
- [24] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.