E.Y.E. C. U.: an Emotional eYe trackEr for Cultural heritage sUpport

Davide Maria Calandra, Dario Di Mauro, Daniela D'Auria and Francesco Cutugno

Abstract Enjoying a painting, a sculpture or, more in general, a piece of art and, at the same time, to receive all the information you need about it: in this paper, we present E.Y.E. C. U. (read "I see you"), a modular eye tracking system which supports art galleries fruition without diverting visitors attention. Every time a visitor lingers on a painting detail, a hidden camera detects her gaze and the framework beams, in real time, the related illustrative contents on the wall region around it, deeply implementing the augmented reality meaning. E.Y.E. C. U. enhances the gaze detection functionalities with an emotional analysis module: as pupil is well known to reflect the emotional arousal, we monitor its size, in order to detect radius variations. Once the visitor has completed her visit, the system summarizes the observed details and the emotional reactions in a report.

Keywords Emotion tracking · Affective computing · Pupil dilatation

1 Introduction

In recent years, with the widespread adoption of mobile devices, museum visitors tend to compensate the lack of information available in the caption aside the artworks, with web researches or dedicated applications usually based on image

D.M. Calandra (✉) · D. Di Mauro · D. D'Auria · F. Cutugno
Department of Electrical Engineering and Information Technology - DIETI,
University of Naples "Federico II", via Cinthia SNC, 80125 Naples, Italy
e-mail: davidemaria.calandra@unina.it

D. Di Mauro
e-mail: dario.dimauro@unina.it

D. D'Auria
e-mail: daniela.dauria4@unina.it

F. Cutugno
e-mail: cutugno@unina.it

T. Torre et al. (eds.), *Empowering Organizations*, Lecture Notes in Information
Systems and Organisation 11, DOI 10.1007/978-3-319-23784-8_13

recognition and retrieval [15]. These solutions satisfy the thirst for knowledge of the visitor but detach her interest from the piece of art which loses the centrality of the visit, as a mobile device interposes between her and the artwork. This scenario gave us the idea to design a software application which supports the artistic fruition without diverting user's attention from the subject of the visit. Thus, a possible solution is just intercepting the visitors' gaze, while they are enjoying the work of art and providing them the contents related to the point of gaze. These last could be soundscapes heared by means of smart headphones [4] or multimedia contents projected on the wall. Moreover, once the point of gaze is known, it can be used to analyse which are the most seen details of each painting, how many users looked at them and for how long time. Then, it could be interesting to know which are the emotional reactions of the visitors while they are enjoying their visit, in order to understand which details arouse pleasure.

In this paper, we present E.Y.E. C. U. (read "I see you"), an emotional eye tracking system which detects the details of the painting observed by the visitor and shows, in real-time, the related deepening contents on the wall around the piece of art; moreover, the *emotional* component computes visitors' reactions while they are enjoying the piece of art. Meanwhile, the logging module stores the information about the points of gaze, the duration of a fixation and which reactions it caused.

Pupil dilation (*mydriasis*) represents a reliable information source in the emotional arousal analysis; this was firstly proved by [8] and we discussed this matter in [2]. In this view, we also monitor the pupillary radius of the visitor to keep track of her emotions during the visit. As we wanted to provide an application working on any expression of art, we generically consider the visual two-dimensional plane beyond the observed object as divided in sections and we aim to detect which section the visitor is gazing; consequently, we can project the multimedia contents concerning the artwork, on the wall region around it, as shown in Fig. 1.

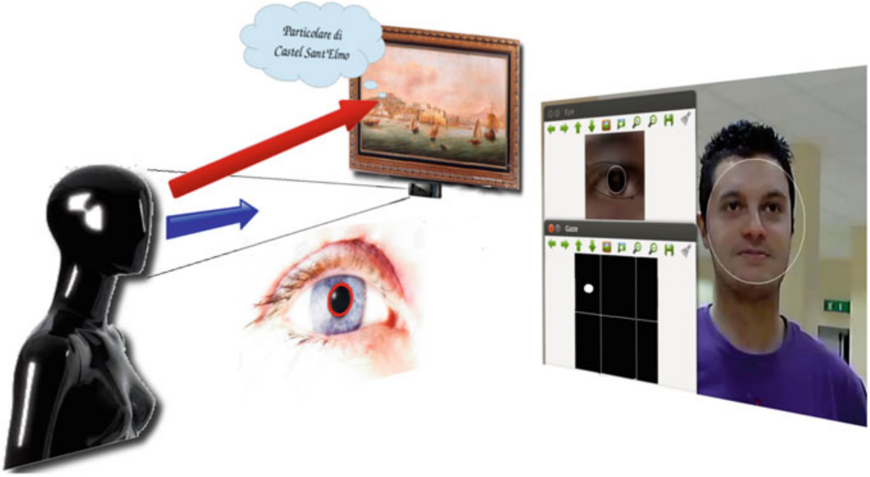

Fig. 1 E.Y.E. C. U

This work represents an evolution of our AutoMyDe [2], presented at itAIS 2013 (http://www.itais.org/itais2013/), in which we developed an automatic mydriasis detector to measure users reactions while they are using a web interface. The mydriatic events were interpreted in terms of cognitive workload.

The paper is organized as follows: Sect. 2 presents the different approaches used to estimate the gaze orientation; Sect. 3 describes the steps needed to detect gaze and to monitor the pupil status; Sect. 4 exposes a case study; Sect. 5 concludes the paper, discussing the obtained results.

2 Point of Gaze Estimation

Before designing a gaze detector, we have to decide if evaluating the points of gaze (PoG) from the eyes movements, from the head pose or both; moreover, we have to establish if users have to wear a device and the precision degree that we want to achieve. In some scenarios, such as the medical diagnosis, patients are usually not allowed to move their head [7] or they have to wear head mounted cameras pointed towards their eyes [11]. In these cases, to estimate the point of gaze means to compute the pupil center position respect to the ellipse formed by the eyelids, while the head position, when considered, is detected through IR sensors in-built on the head of subjects. These systems grant an error threshold lower than 5 pixels [11] due to strict constraints, such as the fixed distance between eye and camera but, on the other hand, they induce users to not natural behaviours. Most remote trackers, such as ones presented in [6], consider the gaze direction determined by the head orientation; these systems do not limit users' movements and do not require they wear a device; they are particularly indicated for scenarios not requiring a high accuracy in identification of details.

To provide a fine-grained remote eye tracker, both head poses and pupil positions have to be estimated, as shown in [16]. Head poses are usually computed by considering 3 degrees of freedom (DoF) [13]: the rotations along the 3 axis of symmetry in the space, *x, y,* and *z,* shown in Fig. 2. Once the head pose in the space is known, the pupil center position will refine its orientation.

3 E.Y.E. C. U

We aimed to develop a non intrusive and highly interactive framework which does not constrain users to maintain a fixed distance from the camera, neither to wear an external device. To achieve our task, we designed a remote gaze detector. However, while remote detectors grant non invasiveness, to track the head in its 3 DoF, increases the computational complexity, making interaction slower.

During our analysis, we asked users to perform maximum 15° of head rotations on *x* axis and maximum 8° on both *y* and *z* and to look at paintings of variable size

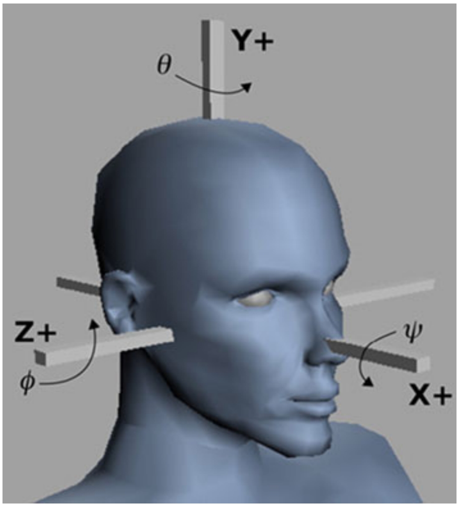

Fig. 2 Head movements

between 50 × 30 and 180 × 75 (sizes are expressed in centimeters) from at least 50 cm of distance. Interpreting the nose tip as center of head rotations and dividing the painting in a number of sections between 2 (1 × 2 matrix) and 6 (2 × 3 matrix), we observed that the head pose variations in the above cited range did not influence the PoG results obtained only by means of pupil center positions. These considerations allowed us to mediate between precision and real time interaction: instructing users to perform limited head rotations on *x* and *y* axis, we estimated wide angular variations along *z* axis only (1 DoF). Thus, geometrically projecting the nose tip coordinates on the observed surface, we evaluate if they fall in the higher or in the lower half of the painting; then, the pupil center position refines the results, indicating the specific observed section. In this way, in the case of a 2 × 3 matrix, the nose tip position indicates if users are looking at the first or the second row; the pupil position reveals if users are looking at the first, the second or the third column, depending on its position is respectively at the left, centered or at the right of the eye area.

Once the gaze direction is known, we have to understand if the user is only looking at the detected PoG or she is just observing it; this distinction allows to understand where the focus of the attention is oriented and it is well explained in Sect. 3.4. If the user is interested on a specific detail of the piece of art, we then provide her the related in depth content, by projecting it on the wall.

Pupils are usually easily detected with IR lights because they reflect the infrared, becoming well visible in the eyeball. However, we chose to detect the input stream by means of a webcam. In particular, we used a 720p webcam. We chose to use a common camera, in order to realize an efficient and cheap solution which does not affect users eyes with IR lights.

The open source OpenCV library (http://opencv.org/) has been used to perform image processing.

3.1 The Multithreading Architecture

The usage of classifiers, the image processing and providing multimedia information are not trivial tasks, computationally speaking. Thus, if by one side they are fundamental operations for our goal, on the other hand they could delay the interaction, causing a decrease of the usability. For this reason, we considered that some operations could be executed in parallel, in order to reduce the computation time. We solved the parallelism with multithreading.

Thus, E.Y.E. C. U. is designed as a multithreading software application: the visual processing thread works on the video stream, in order to detect the needed facial features, while the emotional thread monitors pupil size variations; the reporting thread fills, at regular intervals, a data structure with boolean values representing the mydriatic events, while the attentional thread populates a buffer with the detected points of gaze; when the concentration of the PoG belonging to the same section reaches a given threshold, we consider the visitor interested to it and the attentional thread starts the timed projection of the deepening content in the related wall region, as better shown in Sect. 4.

Once the visitor left the detection range of the camera, E.Y.E. C. U. produces a report indicating the observed sections and the related emotional reactions. E.Y.E. C. U. architecture is shown in Fig. 3. In the figure, threads are identified by the *T*, while the tuple (*t, m, g*) specifies the mydriatic event *m* and the gaze direction *g* measured at the time *t*.

3.2 Features Location

As first step we detect the face. To do this, we use the default pre-trained frontal face detector provided by the OpenCV library. When the application starts, we scan the entire image; then, we track the detected face in a smaller area. Face detection is

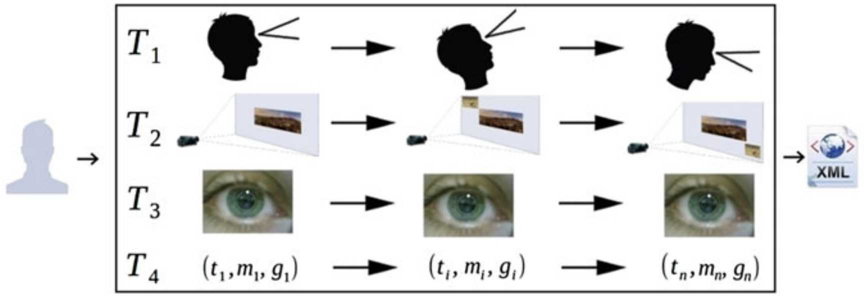

Fig. 3 E.Y.E. C. U. multithreading architecture

performed to constrain the face feature locations to regions inside the facial region, in order to increase both precision and computational efficiency. To obtain the gaze orientation, we firstly need to detect the nose and the eyes: the nose tip will be used to estimate the head position along the z axis; then we search the area of pupils (we actually consider only one eye in computing). Again, the facial features are found using the respective trained classifiers provided by OpenCV; they are Haar-based classifier which extend the set of Haar-like features proposed by Viola and Jones [17]. Once the features have been detected, we computed the detection time: performing the location on images of size 1280 × 960 pixels processed on an Intel Core i7 with 2.2 GHz, the detection time was about 100 ms. Then, we considered that this performances could be improved by taking advantage from the facial geometry: in particular, eyes are located in the upper half of the face and the nose can be easily found, starting from the facial axis on y axis and from the middle point of the face, for both x and z axis. We observed that the performed optimizations allowed to locate the face, the nose and the eyes in a total average time of 35 ms, reducing the computation time of the 65 %.

3.3 Pupil Detection

The detected ocular region contains eyelids, eyelashes, shadows and light reflexes. These represent noise for pupil detection and they could interfere with the correctness of the results. Thus, the eye image has to be refined, before searching the pupil. The following steps have been executed, in order to perform the refinement:

1. the gray scaled image (Fig. 4a) has been blurred by means of a median filter, in order to highlight well defined contours;
2. the well known Sobel partial derivate on the x axis revealed the significant changes in color, allowing to isolate the eyelids;
3. a threshold operation identifies the *sclera*.

As result, these steps produce a mask, which allows to isolate the eye ball from the source image. Pupil detection is now performed on the source image as follows:

1. we drop down to zero (black) all the pixels having cumulative distribution function (CDF) value greater than a certain threshold [1] (Fig. 4b);
2. we morphologically transform the resulting binary image by means of a dilation process, to remove the light reflexes on the pupil;
3. a contours detection operation identifies some neighborhoods (Fig. 4c).
4. selecting the region having maximum area, the pupillary area is found (Fig. 4d);
5. the center of the ellipse (Fig. 4e) best fitting the pupillary area, approximates the pupil center (Fig. 4f).

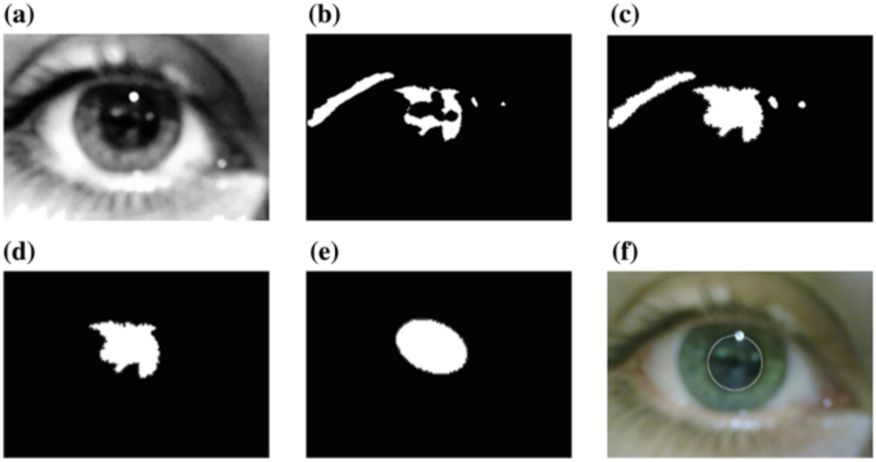

Fig. 4 Pupil detection: main steps. **a** a, **b** b, **c** c, **d** d, **e** e, **f** f

3.4 Focus of Attention

Eye movements, *saccades*, are the fastest produced by the human body: their speed can reach 900° per second and alternate to fixations, which can be interpreted quite differently depending on the context. According to [9], higher fixation frequency on a particular area can be indicative of greater interest in the target or it can be a sign that the target is complex in some way and more difficult to encode. Moreover, more fixations on a particular area indicate that it is more noticeable, or more important to the viewer than other areas [14]. Duchowski [5] reports a mean fixation duration of 1079 ms, during cognitive activities. In order to classify the observed sections of interest, we stored PoG in a buffer of 30 elements (one per frame), working like *history* and we observed that, in the time indicated by Duchowski, the buffer fills the 75 % of its size; for this reason, when the number of PoG belonging to the same section reaches the 75 % buffer size, we consider the user interested to the related section and the projection starts.

3.5 The Emotional Contribute

A wide range of medical studies [8, 10] proved that the brain reacts to emotional arousal with involuntary actions performed by sympathetic nervous system. These changes manifest themselves in a number of ways like increased heart-beat, body temperature, muscular tension and mydriasis. As shown in Sect. 3.1, a dedicated thread monitors the pupillary radius to detect significative variations.

Pupils are larger in children and smaller in adults and the normal size varies from 2 to 4 mm in diameter in bright light to 4–8 mm in the dark [3]. Moreover, pupils react to stimuli in 0.2 s, with the response peaking in 0.5–1.0 s [12]. Hess [8] presented 5 visual stimuli to male and female subjects and he observed that the increase in pupil size varied between 5 and 25 %.

Once we detected the pupil, to calculate the mydriasis we made a comparison between the first stored radius and those computed during the following iterations: according to Hess, when the comparison exceeded the 5 %, a mydriasis has been signaled.

3.6 *The Logging Module*

During the interaction, a parallel thread keeps track of the observed sections and the related emotional reactions. In particular, at fixed steps of 300 ms, the thread stores the current timestamp, the index of the observed section and a value representing the pupil status. If no section is observed, the section index is −1. If the pupil has normal size, the pupil status is 0, otherwise it is 1. At the end of the interaction, a XML document is built with the collected data. The structure of the XML document is shown in the Listing 1.1.

```
<?xml version=”1.0” encoding=”UTF-8”?>
<report>
  <track idTs=”1402674690300” section=”-1” mydriasis =”0” />
  <track idTs=”1402674690600” section=”1” mydriasis =”0” />
  <track idTs=”1402674690900” section=”1” mydriasis =”0” />
  <track idTs=”1402674691000” section=”1” mydriasis =”0” />
  <track idTs=”1402674691300” section=”1” mydriasis =”0” />
  <track idTs=”1402674691600” section=”1” mydriasis =”0” />
  <track idTs=”1402674691900” section=”1” mydriasis =”0” />
</report>
```

Listing 1.1: Report example

4 A Case Study

In this paper, we analyzed user reactions and the gaze orientation in front of the painting *Borgo di Chiaia* by Caspar Adriaans Van Wittel, located at Diego Aragona Pignatelli Cortes museum, Naples. It is an oil on canvas of size 75 × 174 cm. We imaged the painting as divided in a matrix of 2 rows and 3 columns which identify six sections of uniform size. For each one of the six sections, we prepared a list of

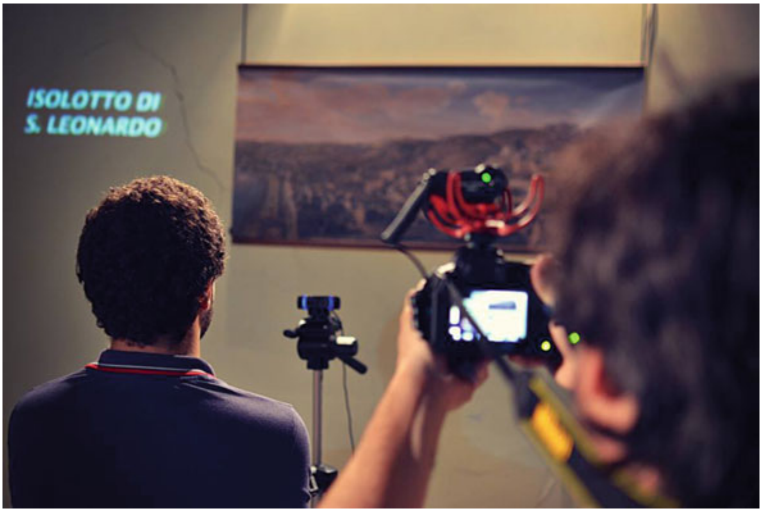

Fig. 5 Case study execution

contents located in corresponding directories. Each time the user lingered on a section for the discussed minimum time, the current content in the related list is projected on the wall region adjacent to the observed area, for a minimum of 3 s after that, if the visitor is still pointing her attention to the contents, these are updated and shown for other 3 s and so on. Meanwhile, the pupil of the visitor is monitored to detect her emotions variations. A real execution is shown in Fig. 5.

5 Results

We tested the system on a sample survey of 23 subjects: 18 adults of age between 23 and 50, and 5 children of age between 4 and 6 years. We asked them to look at the painting, according to the limits exposed in Sect. 3 and to declare which detail they were observing. We considered the test succeeded, when the shown content corresponded to the declared detail.

Experiments registered 75 % of success in adults and 78 % in children: in these cases the projected content referred to the detail that users were observing. System failures were due to a wrong pupil identification: glasses and heavy make-up caused the failure. In 63 % of adult testers, we observed mydriatic reactions; in children the percentage reached 75 %. Results are shown in Fig. 6.

As we stored the indexes of the observed sections, we could count how many times a section has been observed. We represent here the collected data by means of a heatmap, shown in Fig. 7. We built the heatmap using a points distribution function: the higher is the density of the points, the higher is the number of times

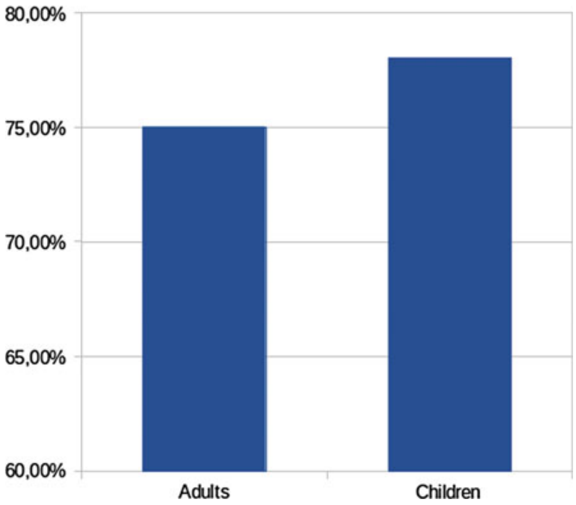

Fig. 6 Results

that the related section has been observed. In particular, Fig. 7 is divided in a 2 × 3 matrix representing the six sections in which the painting has been divided. The figure shows that the most observed section has been the second section on the first row: it has been observed by 80 % of users.

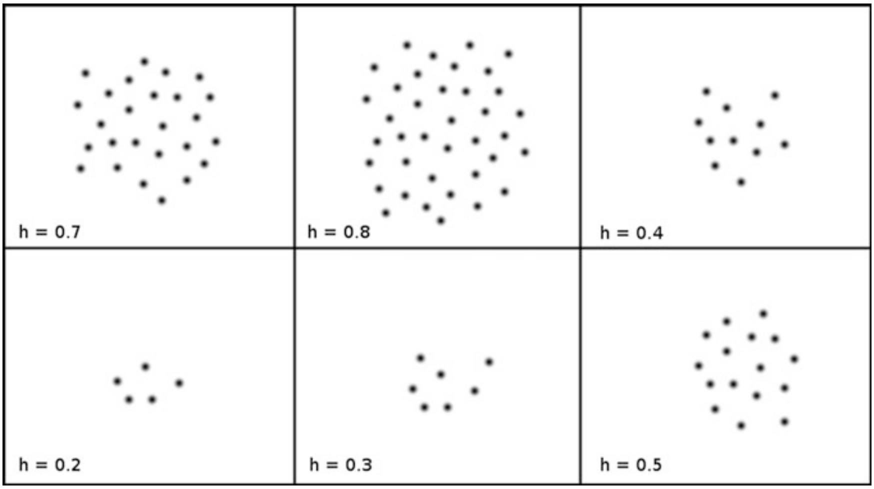

Fig. 7 Heatmap

6 Conclusions

Analysing the software solutions which support the artistic fruition, we observed that most of them divert users' attention from the visit: some require users point their smartphone on the piece of art [15]; others require users type the search key. In both cases, a mobile device interposes between visitor and art.

In order to provide the in depth contents about the object of the visit without deviate users' attention, we realized E.Y.E. C. U., a software application which intercepts users' gaze while they are looking at a painting. E.Y.E. C. U. detects the head movements and the pupil position, in order to understand the observed detail of the painting and then provides the related contents by projecting them on the wall. Just augmenting the reality. As it has been proved that pleasant images cause emotional reactions [8], we even thought to detect pupil dilation, as the pupil is a reliable source of information for the affective computing. In this way we could know which reactions arouse users while they are observing a specific detail.

We tested the system on a real painting, an oil on canvas, and the results showed that E.Y.E. C. U. can correctly detect user's gaze in most of cases. We also observed that people had emotional reactions while they were enjoying the visit, principally the children.

Future works will consist in improving the software application, by reducing users limitations and increasing the precision of the gaze detection. Moreover, we are going to provide audio contents too that could be listened by smart headphones [4] and we are going to enrich the emotional module, by detecting the facial expressions.

Acknowledgment Work supported by the European Community and the Italian Ministry of University and Research and EU under the PON Or.C.He.S.T.R.A. (ORganization of Cultural HEritage and Smart Tourism and Real-time Accessibility) project.

References

1. Asadifard, M., Shanbezadeh, J.: Automatic adaptive center of pupil detection using face detection and cdf analysis. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1, p. 3 (2010)
2. Calandra, D., Cutugno, F.: Automyde: a detector for pupil dilation in cognitive load measurement. In: Caporarello, L., Di Martino, B., Martinez, M. (eds.) Smart Organizations and Smart Artifacts. Lecture Notes in Information Systems and Organisation, vol. 7, pp. 135–147. Springer International Publishing, New York (2014)
3. Clark V.L., Kruse, J.A.: Clinical methods: the history, physical, and laboratory examinations. JAMA **264**(21), 2808–2809 (1990). http://dx.doi.org/10.1001/jama.1990.03450210108045
4. D'Auria, D., Di Mauro, D., Calandra, D.M., Cutugno, F.: Caruso: interactive headphones for a dynamic 3d audio application in the cultural heritage context. In: 2014 IEEE 15th International Conference on Information Reuse and Integration (IRI), pp. 525–528 (2014)
5. Duchowski, A.T.: Eye Tracking Methodology: Theory and Practice. Springer, New York (2007)

6. Fanelli, G., Gall, J., Van Gool, L.: Real time head pose estimation with random regression forests. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 617–624. IEEE (2011)
7. Gómez, E.S., Sánchez, A.S.S.: Biomedical instrumentation to analyze pupillary responses in white-chromatic stimulation and its influence on diagnosis and surgical evaluation (2012)
8. Hess, E.H., Polt, J.M.: Pupil size as related to interest value of visual stimuli. Science **132**, 349–350 (1960)
9. Jacob, R.J., Karn, K.S.: Eye tracking in human-computer interaction and usability research: ready to deliver the promises. Mind **2**(3), 4 (2003)
10. Kahneman, D., Beatty, J.: Pupil diameter and load on memory. Science **154**(3756), 1583–1585 (1966)
11. Kassner, M., Patera, W., Bulling, A.: Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction (April 2014). http://arxiv.org/abs/1405.0006
12. Lowenstein, O., Loewenfeld, I.E.: The pupil. The Eye **3**, 231–267 (1962)
13. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 607–626 (2009)
14. Poole, A., Ball, L.J., Phillips, P.: In search of salience: a response-time and eye-movement analysis of bookmark recognition. In: People and Computers XVIII—Design for Life, pp. 363–378. Springer (2005)
15. Ruf, B., Kokiopoulou, E., Detyniecki, M.: Mobile museum guide based on fast sift recognition. In: Proceedings of the 6th International Conference on Adaptive Multimedia Retrieval: Identifying, Summarizing, and Recommending Image and Music, pp. 170–183, AMR’08. Springer, Berlin (2010)
16. Valenti, R., Sebe, N., Gevers, T.: Combining head pose and eye location information for gaze estimation. IEEE Trans. Image Process. **21**(2), 802–815 (2012). http://www.science.uva.nl/research/publications/2012/ValentiTIP2012
17. Viola, P.A., Jones, M.J.: Rapid object detection using a boosted cascade of simple features. In: CVPR (1), pp. 511–518 (2001)