Affective computing with eye-tracking data in the study of the visual perception of architectural spaces

Magdalena Chmielewska[1,*], *Mariusz* Dzieńkowski[1], *Jacek* Bogucki[2], *Wojciech* Kocki[3], *Bartłomiej* Kwiatkowski[3], *Jarosław* Pełka[4]*, and *Wioletta* Tuszyńska-Bogucka[5]

[1]Lublin University of Technology, Institute of Computer Science, Nadbystrzycka 36B, 20-618 Lublin, Poland
[2]Medical University of Lublin, Clinical Genetics Department, Radziwiłłowska 11, 20-080 Lublin, Poland
[3]Lublin University of Technology, Department of Architecture and Urban Planning, Nadbystrzycka 40, 20-618 Lublin, Poland
[4]Addiction Treatment Center of Lublin, Karłowicza 1, 20-027 Lublin, Poland
[5]University of Economics and Innovations of Lublin, Institute of Psychology and Human Sciences, Projektowa 4, 20-209 Lublin, Poland

Abstract. In the presented study the usefulness of eye-tracking data for classification of architectural spaces as stressful or relaxing was examined. The eye movements and pupillary response data were collected using the eye-tracker from 202 adult volunteers in the laboratory experiment in a well-controlled environment. Twenty features were extracted from the eye-tracking data and after the selection process the features were used in automated binary classification with a variety of machine learning classifiers including neural networks. The results of the classification using eye-tracking data features yielded 68% accuracy score, which can be considered satisfactory. Moreover, statistical analysis showed statistically significant differences in eye activity patterns between visualisations labelled as stressful or relaxing.

1 Introduction

Application of neuroscience methods to analyse and understand human behaviour in controlled environments or laboratories have recently gained researchers attention. Eye-tracking applications cover several domains, such as psychology, engineering, and computer science. The study of automatic recognition of human emotional states and their utilisation in a computer system (affective computing) has had much interest lately due to its multidisciplinary applications.

Psychological research on pupil dilation has shown that not only light contributes to the pupil's response, but also memory load, cognitive difficulty, pain and emotional state [1]. The relationship between the size of the pupil and the emotional reactions is confirmed by the fact that pupil dilation is linked to the activation of the sympathetic or parasympathetic nervous systems [2–6]. One should, however, remember that this relationship is complex since the size of the pupil is also related to a load of cognitive processing [7] and the quantity of light or colour in the sight stimuli [3], so the results obtained cannot be generalized to an excessively high degree. The relationship between eye movements and emotional design or emotions is confirmed [8–10] as well, and the nature of this relationship is the subject of works undertaken also recently.

The approach based on hybrid data-mining was adopted in this study. It involves inferring of algorithms that explore the data, develop the model and discover previously unknown patterns, which are very useful to integrate information and theory from many fields of science.

2 Related works

In research, different pupillary responses to various affective pictures were found [5]. The looking patterns were related to emotional reactions [11–15]. Partala et al. [5] reported pupil dilatation even in the case of participants who were listening to affectively absorbing, as compared to neutral, sounds, suggesting that emotional affection contributed to pupil dilation even if the perceptive concept was not visual. Today, most research suggests that pupil diameter changes when people process stimuli which are emotionally stimulating, irrespective of their hedonic valence [2, 16, 17]. Although, historically speaking it has been indicated that pupil dilation is typical for a negative emotional evaluation of a presented picture [18].

Depending on the fact, that emotional stimuli influence pupil diameter as well as the looking patterns, several attempts of classification based on eye-tracking data can be found in the literature. Lanatà et al. [19] investigated the usefulness of eye tracking and pupil size data in discrimination of emotional states (arousal, neutral) induced by viewing images at different arousal content. They reported about 90% of successful classification for neutral images and about 80% of successful classification for images at high arousal using K-Nearest Neighbor (kNN) algorithm. In their research with emotional movie clips, Alghowinem et al. [20], achieved 66% of average recall results for positive vs.

---

* Corresponding author: m.borys@pollub.pl

negative emotions classification using eye movement, pupil dilation, and pupil invisibility as features and the Support Vector Machine (SVM) as a classifier. Jaques et al. [21] succeeded in predicting emotions relevant to learning, specifically boredom and curiosity using eye-tracking data for an intelligent tutoring system purpose. Four classifiers were tested: using Random Forests (RF), Naïve Bayes (NB), Logistic Regression (LR) and Support Vector Machines (SVM), and the best results obtained for boredom was 69% whereas for curiosity 73%. On the other hand, in the research of predicting like and dislike of movie trailers based on movie features, Hou et al. [22] used eye-tracking data to interpret the underlining reasons of viewers' "like/dislike" decision.

3 Materials and methods

3.1 Participants

The study included a group of 202 adult volunteers (103 women and 99 men) from 18 to 49 years of age (M=23.5, SD=6.11). In the beginning, the participants declared in writing that there were no contra-indications for eye-tracking measurements. The participants were not informed about the aim of the study and had normal or corrected-to-normal vision.

3.2 Apparatus and software

During the experiment, a remote eye tracker Tobii Pro TX300 recorded eye movements and pupil diameter parameters. Tobii Pro TX300, i.e. a binocular video-based eye tracking system, calculates the gaze position by means of near-infrared technology and the dark pupil and corneal reflection method. The eye tracking equipment has tolerance for head movements and therefore the participants could move freely and naturally face the stimuli. Throughout the experiment, the eye activity was recorded with the binocular sampling rate of 300 Hz and a vendor-reported spatial accuracy of 0.5°.

The presentation of the stimuli were conducted using a computer (Laptop Asus G750JX-T4191H with Intel Core i7-4700HQ and 8GB of RAM), running a custom application for stimuli presentation and Tobii Studio 3.3.2. employed for experiment control. The stimuli were displayed on a 23-in. TFT monitor running at a resolution of 1920x1080 and at 60 Hz, equipped with Tobii Pro TX300. The distance between the participants and the screen unit ranged from 50 to 75 cm.

The custom web application for displaying of the stimuli constituted the main part of the experiment launched in the Tobii Studio. Additionally, the app was used to collect metric data concerning the participants of the study, collect ratings of the displayed visualisations (stimuli) and prepare the obtained results for further processing. The developed software displaying the stimuli as separate web pages with a unique URL supported the automation of the eye tracking data extraction and their visualisations.

3.3 Procedure

The study was preceded by obtaining consents from the people interested in enrolling in the experiment who also

had to complete a general demographic questionnaire in which they had to give information about their age, cultural heritage, physical and mental health, etc. Afterward, the participants were briefed about the study and individually tested.

Each participant was asked to sit in an upright chair and was instructed to minimise body movements and keep their gaze directed toward the screen during experimental tasks. The study was conducted in a quiet testing room with an artificial lighting where the natural light was blocked in order to ensure stable conditions when the experiment was in progress. The measurements of light intensity in the room showed it reached approximately 350 lux. The participants' eye movements were calibrated by a nine-point calibration screen since a successful calibration was required in order to conduct the experiment. The next stage of the study was showing the participants a few screens presenting the instructions. Fifteen visualisations (stimuli), divided into five visualisation groups, were shown to participants. Within each of visualisation groups, the participants were shown the stimuli displayed in random order. Each participant decided about the time of exposure to a stimulus as it was possible to switch the subsequent screens with visualisations by means of a computer mouse. After showing each visualisation, every participant was asked to complete a post-questionnaire in order to describe their perceptions of emotions, i.e. presenting them as attractive/ unattractive, relaxing/ stressful, friendly/ unfriendly.

3.4 Data processing

The study enabled to collect data about the eye movements and the pupil diameters size. The pupil diameter was measured for the left and the right eye separately and the exact pupil size was expressed in mm. The algorithms for the pupil size estimation took into consideration the magnification effect produced by the spherical cornea and the distance to the eye [23]. In order to estimate the pupil size for samples in which the pupil was corrupted due to blinking or artifacts, a linear interpolation was used [24]. An artifact was defined as sudden pupil size increase or decrease of 0.1 mm, within a 3 ms time span [5]. The pupil size data were also smoothened using Savitzky-Golay filter [25], with the window length of 51 samples and the 2nd order polynomial. Interpolation and smoothing were separately applied for the data relating to the left and the right eye. The initial light reflex during the exposition of the stimulus was investigated on the basis of the pupil constriction after stimulus onset [12]. The average waveform was used during the exposition of the stimuli and therefore the initial light reflex was estimated at 1500 ms which was not included in the analysis of pupil size changes. In order to obtain more reliable results, the further statistical analysis excluded recordings where the presentation of the stimulus was shorter than 2 s. Afterward, the data from the left and the right eye were averaged using the arithmetic mean.

In order to enable extraction of fixations and saccades the recorded eye tracking data were processed by means of the Tobii Velocity-Threshold Identification

(I-VT) Fixation Filter built in the Tobii Studio 3.4.5 [23] from an average of the position data from the left and the right eye. The data gaps shorter than 75 ms were linearly interpolated and the noise was reduced using moving average filter with the window size of 5 samples. The window length for velocity calculation was set to 20 ms and the velocity threshold was set to 30 degrees per second, and the set distance threshold was 35 pixels. The adjacent fixations were merged and fixations shorter than 60 ms were discarded.

The tracking ratio, defined as the proportion of time that eye movements were measured during the exposition of the visualisation (stimulus), was calculated for each exposition of the stimulus so that the quality of data could undergo further examination. The stimulus expositions with at least 80% tracking ratio were selected for further analysis.

Data collected during individual visualisation's (stimulus) expositions for each participant were considered as a single observation in further statistical analysis and classification. However, due to the low quality of data, where tracking ratio was lower than 80%, and corrupted data for which I-VT algorithm was not able to identify eye movements, the data set has only 2424 observations. Each observation was assigned to one of two classes based on the participant's assessment (Class 0 – stressful, Class 1 – relaxing).

3.5 Feature extraction

For each observation statistical features were calculated from identified eye-movements and pupillary response. Among others, the following features were obtained:

- For fixations: mean, standard deviation, skewness and maximum of duration (in seconds); a number of fixations.
- For saccades: minimum, mean and standard deviation of duration (in seconds); mean, median, standard deviation and a maximum of amplitude (in degrees); a number of saccades; a number of vertical saccades; a number of horizontal saccades.
- For pupil diameter change: mean of normalized pupil diameter (in mm).
- Horizontal to vertical saccades ratio, fixations to saccades ratio, fixations number to exposition time ratio.
- Approximation of scan path length (as the sum of saccades amplitudes values, in degrees).

4 Results

The main objective was to correctly *classify the visualisation as relaxing or stressful based on the participant's eye activity*. The performance of a classification can be evaluated using indicators such as accuracy, recall, specificity, precision or F1 score. In this study, the accuracy and F-1 score were computed to compare classification algorithms' performance. The general organisation of data processing and analysis is presented in Fig. 1.

4.1 Statistical test

The distributions of all features in both groups (Class 0, Class 1) were investigated and differ significantly statistically from the normal distribution. (The

histograms were evaluated as well as the Shapiro-Wilk tests were performed for data). Therefore, to examine whether two groups (Class 0, Class 1) have the same distribution, the Mann-Whitney U test was employed.

The statistically significant differences were found for the features presented in Tab. 1. The U-values and p-values of test are presented in Tab. 2. The statistically significant differences between groups provide a basis for further automated classification of participant's affect.
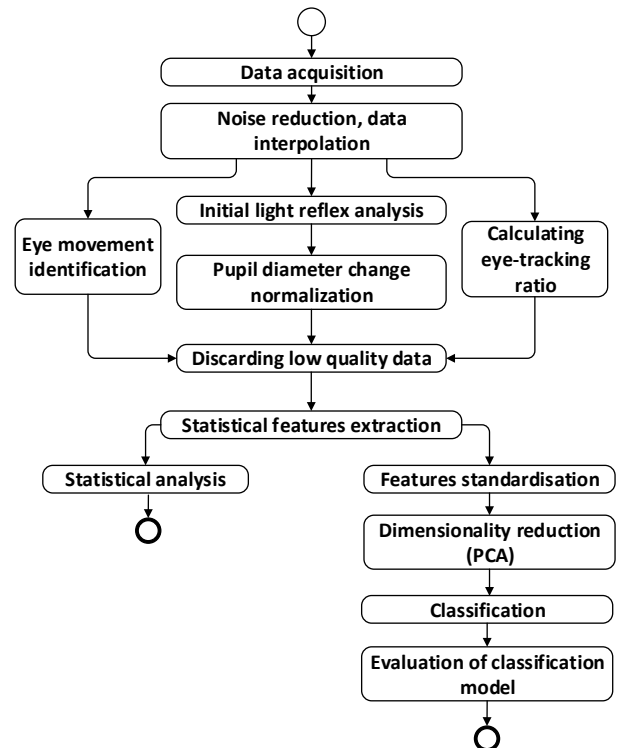

Fig. 1. The schema of data processing and analysis.

Table 1. The mean, median and standard deviation of selected features in the classes.

Feature	Class 0 *Stressful*	Class 1 *Relaxing*
N	1109	1315
minimum duration of saccades	M=9.7142 SD=8.0332 Median=7	M=8.7148 SD=7.1129 Median=7
mean duration of saccades	M=38.2765 SD=6.4687 Median=37.9825	M=37.2515 SD=6.2195 Median=36.7143
number of vertical saccades	M=6.6564 SD=3.9193 Median=6	M=7.8989 SD=4.0767 Median=7
number of horizontal saccades	M=13.9883 SD=8.3265 Median=12	M=12.3703 SD=7.5284 Median=10
horizontal to vertical saccades ratio	M=2.4731 SD=1.3596 Median=2.20	M=1.8028 SD=1.1350 Median=1.50
fixations number to exposition time ratio	M=3.3745 SD=0.5972 Median=3.3829	M=3.3065 SD=0.6049 Median=3.3145

Table 2. The U-values and p-values of the Mann-Whitney U test.

Feature	U-value	Z	p-value
minimum duration of saccades	689209.0	2.32761	0.019934
mean duration of saccades	659743.0	4.04404	0.000053
number of vertical saccades	577739.5	-8.82086	0.000000
number of horizontal saccades	631416.0	-5.6941	0.000000
horizontal to vertical saccades ratio	484132.5	-14.2736	0.000000
fixations number to exposition time ratio	682846.0	-2.6983	0.006971

4.2 Classification results

All transformations of data, as well as the classifications, were performed using Python 3.6.3 with compiler GCC 7.2.0, Numpy 1.14.0, Sklearn 0.19.1 running on Debian x86_64 platform.

Data collected during individual visualisation's expositions for each participant is considered as a single observation. All observations were randomly assigned to the training set and test set, however, the equal distribution of classes in the sets has been preserved. The cardinality and distribution of classes are presented in Tab. 3.

To reduce the dimensionality of data set the Principal Component Analysis (PCA) was conducted proceeded by scaling each feature to a (0,1) range. The cumulative explained variance in the principal component analysis is presented in Fig. 2. After five components more than 80% of the total variance is captured. Fifteen components allow to explained 100% of the total variance.

Table 3. Training and test sets size and distribution of classes.

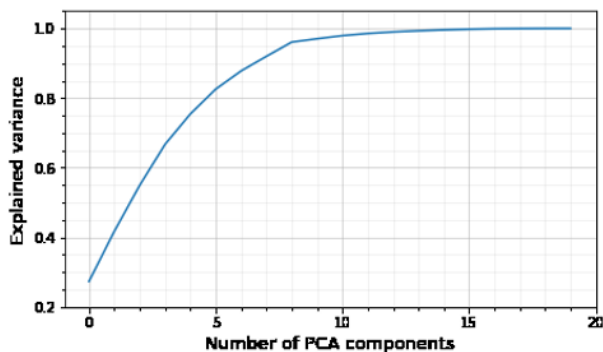|	Number of samples	Distribution of Class 1
Training set	2060 (85%)	45.73%
Test set	364 (15%)	45.88%

Fig. 2. The cumulative explained variance in the principal component analysis.

The several types of classifiers were used for binary classification of visualisations assessed as relaxing (Class 1) or stressful (Class 0) by participants. The following models were tested:
- from linear models: LR, Linear SVM with stochastic gradient descent learning (SGD), Randomized LR, Ridge Regression, Perceptron;
- Gaussian Naïve Bayes classifier;
- kNN classifier;
- from decisions tree models: Decision Tree (DT) and Extra Tree classifiers;
- from ensemble methods: Random Forest, Bagging, Gradient Boosting, and Ada Boost classifiers;
- from Support Vector Machine algorithms: Linear SVM, SVM with the Radial Basis Function kernel;
- from neural networks: Multi-layer Perceptron (MLP);
- as well as Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA).

The performance metrics results for selected classification algorithms are presented in Tab. 4. The best accuracy score was achieved by Linear SVM with stochastic gradient descent learning. However, DT classifier achieved the accuracy almost on the same level but has F1-score, which considers both precision and recall of binary classification, to be higher by about 3%. Moreover, DT model is simpler than SGD model in the implementation. What is noteworthy, simple models such as QDA and LDA also achieved high accuracy rate compared to other, more advanced models.

Table 4. The performance metrics results in (%) for stressful vs. relaxing emotions classification.

Classifier	Accuracy	Recall	Precision	F1-score
SGD	68.4	59.9	67.6	63.5
DT	68.1	69.5	64.1	66.7
MLP	67.6	57.5	67.1	61.9
Linear SVM	66.5	55.7	66.0	60.4
QDA	65.9	57.5	64.4	60.8
LDA	65.9	55.7	65.0	60

5 Discussion

The results from this study demonstrate that eye movement and pupil size data alone is sufficient to classify the visualisation of architectural space as relaxing or stressful. The accuracy results (68%) obtained are noteworthy in the field of affective computing.

The obtained results are close to the outcome of Jaques et al. [21] and Alghowinem et al. [20] works. Although, the visualisations in the study were not designed to arouse strong emotions and therefore the associated physiological (pupillary) response could be more subdued than in [20]. On the other hand, it is hard

to evaluate the results to Lanatà et al. [19] work since they classify neutral and arousal states.

In the future work, additional features associated with blinking to improve the classification will be added. Due to the fact that, voluntary blinks are linked with emotional expressions [26] and can be extracted from eye-tracking recordings. Moreover, the further enhancement of applied classifiers is planned using a grid search technique for tuning the values of specific parameters for an estimator.

References

1. A. T. Duchowski, *Behav. Res. Methods, Instruments, Comput.* **34**, 455–470 (2002)

2. M. M. Bradley, L. Miccoli, M. A. Escrig, P. J. Lang, *Psychophysiology*. **45**, 602–607 (2008)

3. E. Granholm, S. R. Steinhauer, *Int. J. Psychophysiol.* **52**, 1–6 (2004)

4. P. J. Lang, M. K. Greenwald, M. M. Bradley, A. O. Hamm, *Psychophysiology*. **30**, 261–273 (1993)

5. T. Partala, M. Jokiniemi, V. Surakka, in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (ACM, New York, NY, USA, 2000; http://doi.acm.org/10.1145/355017.355042), *ETRA '00*, pp. 123–129

6. S. R. Steinhauer, G. J. Siegle, R. Condray, M. Pless, *Int. J. Psychophysiol.* **52**, 77–86 (2004)

7. J. Beatty, B. Lucero-Wagoner, others, *Handb. Psychophysiol.* **2** (2000)

8. B. Park, L. Knörzer, J. L. Plass, R. Brünken, *Comput. Educ.* **86**, 30–42 (2015)

9. M. Husić-Mehmedović, I. Omeragić, Z. Batagelj, T. Kolar, *J. Bus. Res.* **80**, 145–154 (2017)

10. A. Wiebe, A. Kersting, T. Suslow, *J. Behav. Ther. Exp. Psychiatry*. **55**, 1–5 (2017)

11. J. R. Bardeen, T. A. Daniel, *Cognit. Ther. Res.* **41**, 853–866 (2017)

12. R. R. Henderson, M. M. Bradley, P. J. Lang, *Psychophysiology*. **55**, e13050 (2018)

13. V. Ferrari, A. De Cesarei, S. Mastria, L. Lugli, G. Baroni, R. Nicoletti, M. Codispoti, *Biol. Psychol.* **113**, 75–82 (2016)

14. R. J. Snowden, K. R. O'Farrell, D. Burley, J. T. Erichsen, N. V Newton, N. S. Gray, *Psychophysiology*. **53**, 1217–1223 (2016)

15. J. Coyne, C. Sibley, *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **60**, 37–41 (2016)

16. V. L. Kinner, L. Kuchinke, A. M. Dierolf, C. J. Merz, T. Otto, O. T. Wolf, *Psychophysiology*. **54**, 508–518 (2017)

17. M. M. Bradley, P. J. Lang, *Psychophysiology*. **52**, 1186–1193 (2015)

18. E. H. HESS, J. M. POLT, *Science*. **132**, 349–350 (1960)

19. A. Lanata, A. Armato, G. Valenza, E. P. Scilingo, *Proc. 5th Int. ICST Conf. Pervasive Comput. Technol. Healthc.*, 78–84 (2011)

20. S. Alghowinem, M. AlShehri, R. Goecke, M. Wagner, in *Intelligent Systems for Science and Information: Extended and Selected Results from the Science and Information Conference 2013*, L. Chen, S. Kapoor, R. Bhatia, Eds. (Springer International Publishing, Cham, 2014; https://doi.org/10.1007/978-3-319-04702-7_15), pp. 261–276

21. N. Jaques, C. Conati, J. M. Harley, R. Azevedo, in *Intelligent Tutoring Systems*, S. Trausan-Matu, K. E. Boyer, M. Crosby, K. Panourgia, Eds. (Springer International Publishing, Cham, 2014), pp. 29–38

22. Y. Hou, T. Xiao, S. Zhang, X. Jiang, X. Li, X. Hu, J. Han, L. Guo, L. S. Miller, R. Neupert, T. Liu, *IEEE Trans. Affect. Comput.* **7**, 29–44 (2016)

23. Tobii Technology AB, *User Manual - Tobii Studio* (2014; www.tobii.com)

24. S. L. Merritt, A. P. Keegan, P. W. Mercer, *Nurs. Res.* **43**, 56–59 (1994)

25. O. Bergamin, R. H. Kardon, *Invest. Ophthalmol. Vis. Sci.* **44**, 1546 (2003)

26. G. Cardona, N. Quevedo, *Curr. Eye Res.* **39**, 239–244 (2014)