

Detecting Mental Fatigue from Eye-Tracking Data Gathered While Watching Video

Yasunori Yamada^(✉) and Masatomo Kobayashi

IBM Research - Tokyo, Tokyo, Japan
{ysnr,mstm}@jp.ibm.com

Abstract. Monitoring mental fatigue is of increasing importance for improving cognitive performance and health outcomes. Previous models using eye-tracking data allow inference of fatigue in cognitive tasks, such as driving, but they require us to engage in a specific cognitive task. A model capable of estimating fatigue from eye-tracking data in natural-viewing situations when an individual is not performing cognitive tasks has many potential applications. Here, we collected eye-tracking data from 18 adults as they watched video clips (simulating the situation of watching TV programs) before and after performing cognitive tasks. Using this data, we built a fatigue-detection model including novel feature sets and an automated feature selection method. With eye-tracking data of individuals watching only 30-seconds worth of video, our model could determine whether that person was fatigued with 91.0% accuracy in 10-fold cross-validation (chance 50%). Through a comparison with a model incorporating the feature sets used in previous studies, we showed that our model improved the detection accuracy by up to 13.9% (from 77.1 to 91.0%).

Keywords: Mental fatigue · Cognitive fatigue · Feature selection · Natural viewing · Free viewing · Visual attention model

1 Introduction

Health monitoring in a smart environment such as adaptive workplaces and smart houses has been increasingly recognized for its importance in improving health outcomes [1]. Especially for supporting a rapidly aging population, research has been focused on developing technologies that help mitigate critical situations and enable individuals to manage their own health, with the dual aim of increasing quality of lives and reducing healthcare costs [2].

One area of an individual's daily health status that has yet to be utilized is mental fatigue, which refers to the feeling people might experience during or after cognitive activities [3]. Mental fatigue is a common problem in modern everyday life and comes at a huge public health cost [4]. It is a warning sign of harmful accumulation of stress that can have a detrimental effect on one's health [5] and an important symptom in general practice due to its association with a large number of chronic medical conditions such as cancer, Alzheimer's disease, and Parkinson's disease [6].

Previous studies for monitoring mental fatigue have primarily focused on detecting fatigue during cognitive tasks such as driving [6–8]. Unobtrusive methods, such as those that use remote- or webcam-based eye tracking, typically monitor changes in an individual’s pupil response, blinking behavior, and eye movement to determine levels of fatigue during cognitive tasks [9]. Although these methods have shown the usefulness for monitoring fatigue during a specific cognitive task that requires visual processing, no study has yet developed a model that enables us to infer mental fatigue from eye-tracking data in natural-viewing situations when the individual is not performing cognitive tasks. Such a system would enable people to monitor mental fatigue in a condition close to everyday life. Moreover, it would be used to infer fatigue induced by not only specific cognitive visual tasks, but also various factors such as cognitive auditory tasks, multiple cognitive tasks, or poor health [6].

In contrast, recent studies have attempted to discriminate clinical populations from eye-tracking data in natural viewing conditions [10,11]. For example, Crabb and colleagues demonstrated that patients with neuro-degenerative eye diseases can be separated from healthy controls by using eye-tracking data collected while the patients freely watched TV movies [11]. However, there has been no investigation on the associations of eye-tracking data in natural viewing conditions with mental fatigue.

In this paper, we present a novel model that detects mental fatigue in natural viewing situations in which people watched video clips such as a TV program. More specifically, we collected eye-tracking data from 18 adults as they watched video clips before and after they performed auditory cognitive tasks. From this data, we extracted 181 quantitative features, categorized into six feature sets related to oculomotor-based metrics, blinking behavior, pupil measurements, gaze allocation, eye-movement directions, and saliency-based metrics using a saliency model (a computational model of visual attention). Although the last three feature sets have already been used for characterizing eye movements, especially in natural viewing situations [10,11], they have not been used for inferring mental fatigue. Using these features and an automated feature selection method, we built a two-class classifier for detecting mental fatigue. With eye-tracking data of individuals watching only 30s worth of video, our model could determine whether that person was fatigued or not with 91.0% accuracy in 10-fold cross-validation (chance 50%). To make a comparison with a model based on the existing work, we also built a model using the three feature sets related to oculomotor-based metrics, blinking behavior, and pupil measurements used in a previous study, where the detection accuracy was 77.1%.

2 Data Collection

To build a model for inferring mental fatigue from eye-tracking data in a natural viewing situation, we collected data while participants watched video clips (simulating the situation of watching a TV program) before and after performing an auditory cognitive task.

2.1 Participants

We collected data from 20 participants (8 females, 12 males; 24–76 years; mean \pm SD age 47.5 ± 20.5 years). All participants were well-rested and in good health, as measured by self-reports, and they had normal or corrected-to-normal vision. They were unaware of the purpose of the experiment. Written informed consent was obtained prior to the study. Eye-tracking data from two participants (one female, one male) were excluded from our analysis because of problems calibrating the eye tracker. Thus, our sample size was $N = 18$.

2.2 Experimental Design and Procedure

The experimental procedure is summarized in Fig. 1A. Participants performed a 17-minute mental calculation task designed to induce mental fatigue two times (Fig. 1B). They were asked to take questionnaires and watch video clips prior to and following each mental calculation task. Prior to the experiment, all participants were given oral instructions about the experiments and allowed to practice the mental calculation task.

In regard to the questionnaires, we used numerical rating scales to measure the participant's current ("right now, at this moment") perceived intensity of feelings regarding mental and physical fatigue, sleepiness, and motivation. The intensity was scaled from 0 to 10, with zero indicating an absence of those feelings and 10 indicating the strongest feeling ever experienced.

To collect eye-tracking data, the participants were asked to watch video clips approximately five minutes in length during each phase. As in previous studies [10], they were instructed to simply "watch and enjoy the videos."

As an auditory cognitive task to induce mental fatigue, we used a modified version of the Paced Auditory Serial Attention Test (mPASAT) [12] (Fig. 1B). Participants listened to a series of numbers ranging from one to nine. They were asked to add the number they had just heard to the number they had heard before and then to press a button whenever the sum of the two consecutive numbers equaled ten. One phase consisted of five 3-minute on-periods and four 30-second off-periods for a total 17 min. Each number was presented every 1.5 s. Participants were also asked to visually focus on three numbers on the display, which randomly changed every 0.5 s. These visual numbers were intended to distract and interfere with the primary auditory task, thereby increasing the complexity and attentional demands of the task in order to induce further mental fatigue.

2.3 Stimuli and Eye-Tracking Data Acquisition

To simulate the situation of watching a TV program, we used video clips made in the same manner as previous studies that investigated how neurodevelopmental and neurodegenerative disorders affect eye movements in natural viewing situations [10, 13] (Fig. 1C).

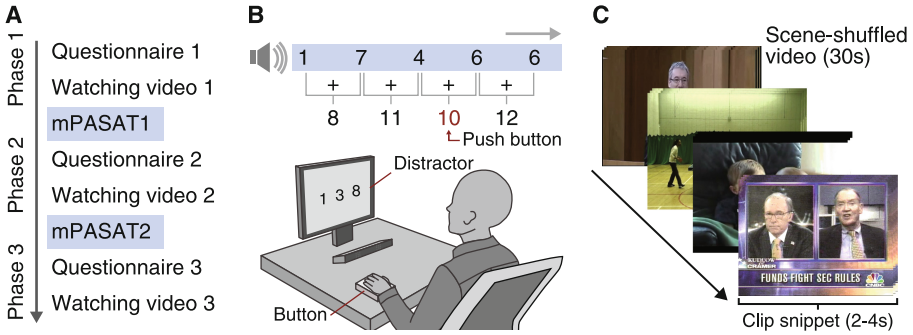


Fig. 1. Experimental setup: (A) overall procedure, (B) mental calculation task (mPASAT), (C) examples of scene-shuffled video clips.

Each 5-minute phase consisted of nine scene-shuffled videos (SVs), approximately 30 s each. Between the SVs, there were 5-second off-periods for rest. The SVs were made by assembling randomly extracted snippets from video clips. The lengths of the snippets were determined so that they were within the range of typical television programs [14, 15]. Specifically, the lengths of the snippets were uniformly distributed between two and four seconds, so each SV consisted of nine to eleven snippets with no temporal gaps in between. For the original video clips, we utilized two datasets: CRCNS-ORIG [16] and DIEM [17], consisting of heterogeneous sources with different styles of programs that are commonly watched on a daily basis.

The participants’ eye movements and pupil data were recorded using a noninvasive infrared EMR ACTUS eye-tracking device at a sample rate of 60 Hz (nac Image Technology Inc.; spatial resolution for eye movements and pupil diameter less than 0.5° and 0.1 mm, respectively). The eye tracker was calibrated using 9-point calibration at the beginning of each recording phase.

3 Mental Fatigue Detection Model

Our model uses 30 s worth of eye-tracking data in each SV to make a decision whether a participant is in a fatigued or non-fatigued state at that time. Thus, we obtained 9 samples for each 5-min phase of video watching.

First, we extracted 181 features categorized into six feature sets from the eye-tracking data that may change according to an individual’s state of mental fatigue. Next, we built a two-class classifier for inferring mental fatigue using a subset of the features selected by a feature selection method through recursive evaluation and selection to avoid over-fitting (Fig. 2).

3.1 Data Preprocessing and Features

The raw eye-position data were segmented into blink, saccade, and fixation (or smooth-pursuit) periods. First, we extracted blink periods by using eyelid

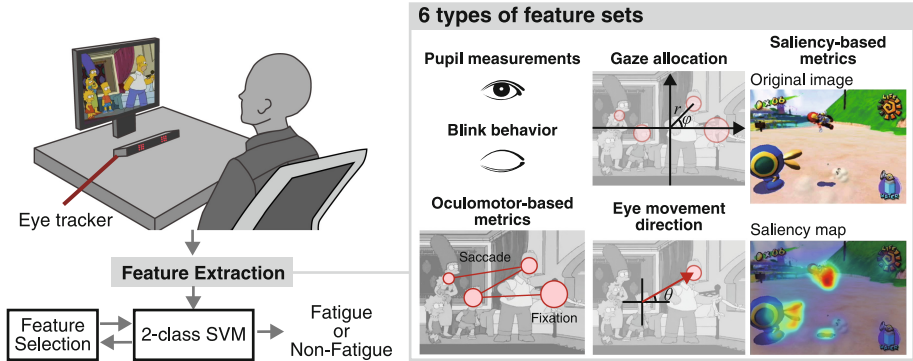


Fig. 2. Overview of our fatigue-detection model.

occlusion of both eyes. Apart from the blink periods, artifacts detected by the eye tracker were removed by using a linear interpolation algorithm. Finally, we used the mean-shift clustering method in the spatio-temporal domain to identify saccade and fixation periods [18].

We extracted six feature sets that we hypothesized would be differentially influenced by mental fatigue from each 30-second-long SV (Fig. 2). The first three feature sets related to oculomotor-based metrics, blinking behavior, and pupil measurements were used in previous studies on mental fatigue during cognitive tasks [6–8]. The other three feature sets were related to gaze allocation, eye movement directions, and saliency-based metrics. Although these feature sets have been used for characterizing eye movements in natural viewing conditions as well as inferring neurodevelopmental and neurodegenerative disorders [10, 11, 19], they have not been used for inferring mental fatigue.

The oculomotor-based features consisted of nine features: saccade amplitude, saccade duration, saccade rate, inter-saccade interval (mean, standard deviation, and coefficient of variance), saccadic mean velocity (mean and median), and fixation duration. We calculated seven features related to blinking behavior: blink duration, blink rate, blink duration per minute (the total time of all durations), and inter-blink interval (mean, standard deviation, and coefficient of variance). The pupil measurements were subdivided into six features related to pupil diameter, constriction velocity, and amplitude of each eye, and nine features related to the coordination of the pupil diameters of both eyes. Of these nine features, one was computed using Pearson’s correlation coefficient. The other eight features were extracted using the phase locking value [20], which can identify transient synchrony over shorter time scales than Pearson’s correlations. We used the mean and maximum values of the phase locking values with four different time windows (5, 10, 30, 60 frames).

The fourth feature set was calculated from a time-series of gaze allocation. We first converted gaze allocation values into radius and angle (r, ϕ) in a polar coordinate system situated at the center of the display. We then defined two time

series of gaze allocations during all periods and only during fixation periods as $(\mathbf{r}, \phi)_{\text{all}}$ and $(\mathbf{r}, \phi)_{\text{fx}}$, respectively. We discretized each time series with k bins of uniform width. We set $k = 8$ for \mathbf{r}_{all} and \mathbf{r}_{fx} , $k = 36$ for ϕ_{all} , and $k = 12$ for ϕ_{fx} . As features, we used the probability of each bin and entropy estimated using these histograms and also calculated the mean and median values of \mathbf{r}_{all} and \mathbf{r}_{fx} . In total, we obtained seventy-two features from the gaze-allocation data.

The fifth feature set related to eye-movement directions was calculated in a similar manner to the gaze allocation features. We discretized the time series of eye-movement directions θ during all periods and saccades periods into 12 and 36 bins of uniform width, respectively. We then computed the probability of each bin and entropy estimated using these histograms as features. In total, we obtained fifty features.

The sixth and final set consisted of features using a saliency model. The saliency model was proposed as a biologically-inspired computational model of human attention [21, 22]. The saliency model computes a topographic map of conspicuity for every location in each video frame, highlighting locations that may attract attention in a stimulus-driven manner. We used the graph-based visual saliency model, where conspicuity maps of six low-level features (intensity contrast, color contrast, intensity variance, oriented edges, temporal flicker, and motion contrast) are linearly combined and normalized to form a saliency map [23]. Using both the saliency map and the six conspicuity maps, we obtained $4 \times 7 = 28$ saliency-based features in total. For more details about how to calculate saliency-based features, please see the original papers [10, 23].

3.2 Classification and Feature Selection

For a two-class classification model for detecting mental fatigue, we used support vector machine (SVM) models [24, 25] with a radial basis function kernel as follows: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$. We set $\gamma = b_{\text{SVM}}/n_f$, where n_f is the number of features and b_{SVM} is a hyper-parameter. We used the algorithm for SVM implemented in MATLAB (MathWorks Inc., Natick, MA) and LIBSVM toolbox [25].

To identify useful features and avoid over-fitting of the model, we performed a feature selection through recursive evaluation and selection. One of the well-known methods is support vector machine recursive feature elimination (SVM-RFE) in a wrapper approach [26]. However, when the candidate feature set has highly correlated features, the ranking criterion of SVM-RFE tends to be biased, which would have a negative effect on the results. Our feature set contained highly correlated features such as features about saccade duration, amplitude, and velocity. We then used an improved SVM-RFE algorithm with a correlation bias reduction strategy in the feature elimination procedure [27].

4 Results

We first determined whether or not the tasks succeeded in inducing mental fatigue in the participants by using subjective ratings and objective

measurements of mental fatigue that have been used in previous studies. We then evaluated our mental fatigue detection model in terms of their average scores after 20 iterations of 10-fold cross-validation.

4.1 Mental Fatigue After the Cognitive Tasks

We first investigated the participants' reported mental fatigue before and after performing the mPASAT, with subjective ratings from 0 to 10 (Fig. 3A). Compared with the subjective ratings in phase 1, i.e., before engaging in the cognitive task, 12 and 14 out of 18 participants in phases 2 and 3, respectively, reported increased ratings of mental fatigue. We performed a repeated-measures Friedman non-parametric ANOVA followed by Dunn's multiple comparisons and found a significant increase in mental fatigue from phases 1 to 3 ($p < .05$), but no significant difference between phases 1 and 2 or between phases 2 and 3.

We also tried to determine whether objective measurements of mental fatigue changed after the participant performed cognitive tasks. We used pupil diameters and blink behaviors, which are widely used as fatigue-related biomarkers [6–8] as the objective measurements. One-way repeated measures ANOVA with post hoc Bonferroni multiple comparisons were used to calculate the statistical significance over the phases. In this analysis, we computed these measurements by taking averages during each 5-minute phase. As a result, we found a significant decrease of pupil diameters from phase 1 to 2 and 3 ($p < .05, p < .005$, respectively) for the left eye and from phase 1 to 3 ($p < .005$, Fig. 3B; from phase 1 to 2, $p = .14$) for the right eye. In regard to the blink behaviors, we also found significant changes indicative of increased mental fatigue in duration, blink rate, and blink duration per minute over the phases. Among them, the blink duration per minute showed the biggest difference ($\eta_p^2 = .348$; from phase 1 to 2, $p < .05$; from phase 1 to 3, $p < .001$; Fig. 3C).

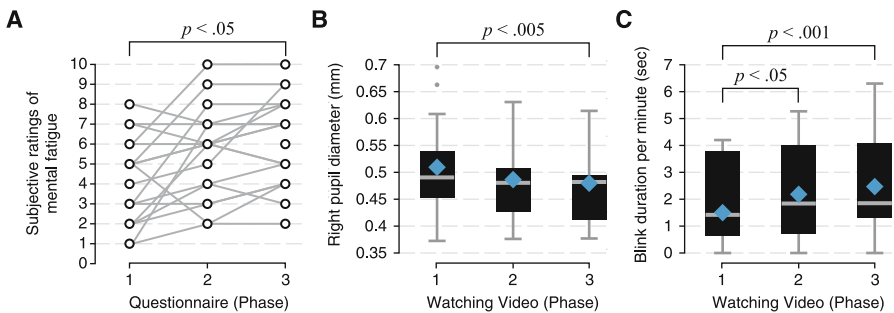


Fig. 3. Changes in subjective and objective measurements for mental fatigue after performing mPASAT. (A) Subjective ratings for mental fatigue on an 11-point numerical rating scale from 0 to 10. (B), (C) Right pupil diameter and blink duration per min. Boxes show the median, 25th, and 75th percentiles, filled symbols show outliers, and squares represent mean values.

Table 1. Fatigue-detection-model performance in 10-fold cross validation. F_{pre} : three feature sets related to oculomotor, blinks, and pupil measurements used in the previous studies, F_{sal} : saliency-based features, F_{emd} : features related to eye movement directions, and F_{ga} : features related to gaze allocation.

Model	Detection performance (%)			
	Accuracy	Precision	Recall	F-measure
F_{pre}	77.1	78.6	72.9	75.6
$F_{\text{pre}} + F_{\text{sal}}$	80.7	79.4	83.0	81.0
$F_{\text{pre}} + F_{\text{emd}}$	82.9	83.2	82.4	82.7
$F_{\text{pre}} + F_{\text{ga}}$	84.7	84.6	84.9	84.7
$F_{\text{pre}} + F_{\text{sal}} + F_{\text{emd}} + F_{\text{ga}}$	91.0	91.4	90.3	90.8

Through these analyses, the results regarding the subjective and objective measurements indicate that the participants experienced increased mental fatigue after engaging in the cognitive tasks two times, i.e., in phase 3. We thus regarded phase 1 as a non-fatigued circumstance and phase 3 as a fatigued circumstance and proceeded to build a model that classifies the eye-tracking data of phases 1 and 3.

4.2 Model Performance

We built a fatigue detection model to differentiate eye-tracking data before and after performing the cognitive tasks. We used eye-tracking data of 18 participants in phases 1 and 3. In our model, features were extracted from each 30-second SV trial, and each phase consisted of nine SVs. Thus, the number of samples was $18 \times 9 \times 2 = 324$.

As a result of 20 iterations of 10-fold cross-validation, our model detected mental fatigue with 91.0% accuracy (Table 1). The feature selection process selected 55 of the 181 features as the most discriminative for classifiers for detecting mental fatigue and selected the features of all six groups. We also evaluated our model by leave-one-subject-out cross-validation, where classifiers were trained using data collected from all participants except one and then were tested on data of the one participant left out of the training data set. We repeated this process for all participants, and obtained an accuracy of 88.5% accuracy.

We next investigated the contribution of the feature sets related to gaze allocation, eye movement directions, and saliency predictions proposed in this study. First, we built a model using only the three feature sets related to oculomotor-based metrics, blink behavior, and pupil measurements used in the previous studies. We did the feature selection and hyper-parameter optimization in the same way as our model. The model performance was 77.1% accuracy in 10-fold cross validation. Next, we separately added each feature set to this model. As a result, the model accuracies increased to 84.7%, 82.9%, and 80.7% as a result of adding gaze-location features, eye-movement direction features, and

saliency-based features, respectively (Table 1). Therefore, we found that the novel use of three feature sets each improved the model's performance, and when taken together improved the model's performance by up to 13.9% (from 77.1 to 91.0%).

5 Conclusion

In contrast to previous studies focusing on detecting mental fatigue during cognitive tasks, we aimed to develop a system enabling us to infer mental fatigue in natural-viewing situations when an individual is not performing cognitive tasks. To this end, we devised a fatigue-detection model including novel feature sets and an automated feature selection method. Through experimentation with 18 adults, we showed that our model could detect mental fatigue with an accuracy of 91.0% in 10-fold cross-validation. One of the limitations in this study is that the study took place in a lab setting. We need to investigate whether our model can infer mental fatigue induced by everyday tasks. In addition, there is a possibility that the controlled setting might influence the way people watch video clips. Thus, future work will include an in-situ study to test our model in more realistic situations.

Acknowledgments. This research was partially supported by the Japan Science and Technology Agency (JST) under the Strategic Promotion of Innovative Research and Development Program.

References

1. Alemdar, H., Ersoy, C.: Wireless sensor networks for healthcare: a survey. *Comput. Netw.* **54**(15), 2688–2710 (2010)
2. Favela, J., Castro, L.A.: Technology and aging. In: García-Peña, C., Gutiérrez-Robledo, L.M., Pérez-Zepeda, M.U. (eds.) *Aging Research-Methodological Issues*, pp. 121–135. Springer, Cham (2015)
3. Boksem, M.A., Tops, M.: Mental fatigue: costs and benefits. *Brain Res. Rev.* **59**(1), 125–139 (2008)
4. Avlund, K.: Fatigue in older adults: an early indicator of the aging process? *Aging Clin. Exp. Res.* **22**(2), 100–115 (2010)
5. Maghout-Juratli, S., Janisse, J., Schwartz, K., Arnetz, B.B.: The causal role of fatigue in the stress-perceived health relationship: a MetroNet study. *J. Am. Board Family Med.* **23**(2), 212–219 (2010)
6. Hopstaken, J.F., Linden, D., Bakker, A.B., Kompier, M.A.: A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology* **52**(3), 305–315 (2015)
7. Schleicher, R., Galley, N., Briest, S., Galley, L.: Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics* **51**(7), 982–1010 (2008)
8. Di Stasi, L.L., Renner, R., Catena, A., Cañas, J.J., Velichkovsky, B.M., Pannasch, S.: Towards a driver fatigue test based on the saccadic main sequence: a partial validation by subjective report data. *Transp. Res. Part C: Emerg. Technol.* **21**(1), 122–133 (2012)

9. Dawson, D., Searle, A.K., Paterson, J.L.: Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Med. Rev.* **18**(2), 141–152 (2014)
10. Tseng, P.H., Cameron, I.G., Pari, G., Reynolds, J.N., Munoz, D.P., Itti, L.: High-throughput classification of clinical populations from natural viewing eye movements. *J. Neurol.* **260**(1), 275–284 (2013)
11. Crabb, D.P., Smith, N.D., Zhu, H.: What’s on TV? Detecting age-related neurodegenerative eye disease using eye movement scanpaths. *Frontiers in Aging Neurosci.* **6**, 312 (2014)
12. Cook, D.B., O’Connor, P.J., Lange, G., Steffener, J.: Functional neuroimaging correlates of mental fatigue induced by cognition among chronic fatigue syndrome patients and controls. *Neuroimage* **36**(1), 108–122 (2007)
13. Carmi, R., Itti, L.: The role of memory in guiding attention during natural vision. *J. Vis.* **6**(9), 4 (2006)
14. Cutting, J.E., DeLong, J.E., Brunick, K.L.: Visual activity in Hollywood film: 1935 to 2005 and beyond. *Psychol. Aesthet. Creat. Arts* **5**(2), 115 (2011)
15. Bordwell, D.: Intensified continuity visual style in contemporary American film. *Film Q.* **55**(3), 16–28 (2002)
16. Itti, L., Carmi, R.: Eye-tracking data from human volunteers watching complex video stimuli (2009)
17. Mital, P.K., Smith, T.J., Hill, R.L., Henderson, J.M.: Clustering of gaze during dynamic scene viewing is predicted by motion. *Cogn. Comput.* **3**(1), 5–24 (2011)
18. Xu, P., Ehinger, K.A., Zhang, Y., Finkelstein, A., Kulkarni, S.R., Xiao, J.: TurkGaze: crowdsourcing saliency with webcam based eye tracking. arXiv preprint [arXiv:1504.06755](https://arxiv.org/abs/1504.06755) (2015)
19. Zhang, Y., Wilcockson, T., Kim, K.I., Crawford, T., Gellersen, H., Sawyer, P.: Monitoring dementia with automatic eye movements analysis. In: Czarnowski, I., Caballero, A.M., Howlett, R.J., Jain, L.C. (eds.) *Intelligent Decision Technologies 2016*. SIST, vol. 57, pp. 299–309. Springer, Cham (2016). doi:[10.1007/978-3-319-39627-9_26](https://doi.org/10.1007/978-3-319-39627-9_26)
20. Tass, P., Rosenblum, M., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., Freund, H.J.: Detection of n: m phase locking from noisy data: application to magnetoencephalography. *Phys. Rev. Lett.* **81**(15), 3291 (1998)
21. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. *Cogn. Psychol.* **12**(1), 97–136 (1980)
22. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1254–1259 (1998)
23. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*, pp. 545–552 (2006)
24. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152. ACM (1992)
25. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2**(3), 27 (2011)
26. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**(1–3), 389–422 (2002)
27. Yan, K., Zhang, D.: Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sens. Actuators B: Chem.* **212**, 353–363 (2015)