Psychological potential field and human eye fixation on binary line-drawing images: A comparative experimental study

Naoyuki Awano¹ (✉), Yuki Hayashi²

© The Author(s) 2020.

Abstract Quantitatively evaluating the psychological and perceptual effects of objects is an important issue, but is difficult. In cognitive studies, the psychological potential field (PPF), which represents psychological intensities in vision and can be calculated by applying computational algorithms to digital images, may help with this issue. Although studies have reported using the PPF to evaluate psychological effects, such as impressions, detailed investigations on how the PPF represents psychological perception and its limitations have not yet been performed. Another relevant tool is the fixation map, which visualizes human eye fixations; this map is generated from actual measurements acquired by eye-tracking and does not represent psychological effects directly. Although the PPF and the fixation map are based on visual imaging, they have never been compared. In this paper, we do so for the first time, using psychological and perceptual properties of line-drawing images. The results demonstrate the difference between these methods, including their representation of different properties with respect to visual perception. Moreover, the similarity between the two methods highlights the possibility of assessing perceptual phenomena such as categorization and cognition of objects based on human vision.

Keywords psychological potential field (PPF); eye fixation; eye-tracking; visual perception; line drawing

1 Introduction

Visual recognition from visual media, including object recognition and natural scene categorization, is a

fundamental issue in cognitive and computer sciences [1]. Determining the cognitive mechanism in human visual processing is important for resolving this issue. In general, because real objects have colors and textures, it is desirable to use color photographs for investigations. However, recent studies have reported that natural scene categorization of color photographs and line drawings generates similar neural activity [2, 3]. Line drawings have less information than color photographs, so if perceptual findings from viewing color photographs can be obtained by viewing line drawings, there is a possibility of simplifying investigations.

The psychological potential field (PPF) [4] (also called the field of visual perception and the induction field in vision) may be helpful in solving this issue. The PPF is derived by investigating the psychological effects projected from a shape to its surrounding area. The PPF represents psychological intensities produced by shape contours as potential values, as shown in Fig. 1, and can be calculated for digital images, although only binary images have been studied thus far. While studies using the PPF to evaluate psychological effects such as impressions have been reported [5], further clarification is needed to determine which type of psychological perception is represented by the PPF.

In contrast, a fixation map [6], a visualization of human eye fixations, can be derived from measurements acquired by tracking eye movements, as shown in Fig. 1(c). Several fixation map databases have been provided on the Internet; they are primarily generated by tracking human eye movements for a duration of a few seconds. The fixation map represents visual information, such as attention, but does not directly represent psychological effects.

Although the PPF and fixation map are both based

1 Osaka University of Economics, Osaka, Osaka 533-8533, Japan. E-mail: awanonaoyuki@gmail.com (✉).

2 Osaka Prefecture University, Sakai, Osaka 599-8531, Japan.

Manuscript received: 2019-12-04; accepted: 2020-03-20

Fig. 1 PPF and fixation map. Red to blue indicates high to low values.

on visual images, they have never been compared. In this study, we compare the PPF and fixation map for the first time to clarify their differences; then we determine the psychological and perceptual properties based on these differences. For instance, the images shown in Figs. 1(b) and 1(c) are different. However, the differences need clarification; specifically, the two images may be identical, depending on the stimuli, or the fixation map may gradually approach the PPF if the measurement duration is sufficiently long. Therefore, this paper introduces experimental methods to compare the PPF and the fixation map using simple binary line drawings to simplify the comparison.

2 Related work

2.1 Psychological potential field

Yokose [4] investigated the psychological effects of shape contours in vision using a light threshold method [8] and discovered the presence of a psychological field, similar to an electrostatic field, around shapes. This field has been demonstrated in physiological and psychological experiments [9, 10]; however, the theory has only been established for two-dimensional binary images, such as black letters on a white background. For example, from line segments, the PPF can be produced by calculating potential energies around the segments. The potential energy M_p of a point p can be calculated by Eq. (1).

$$M_p = \sum_{i=1}^N \int \frac{ds}{f_i(s)} \quad (1)$$

where N is the number of line segments, s is a point on a line segment, and $f(s)$ is the distance function from p to s . Note that s is a point of a line segment that is not occluded by another line segment, similar to the case in which p is a light source and s is the portion exposed to light.

Nagaishi [11] proposed a computational PPF model for digital images based on the above theory. The PPF is generally determined for binary digital images, with black pixels belonging to objects and white pixels to background. The potential energy M_p of a background pixel p is computed from the contour pixels of objects using Eq. (2).

$$M_p = \frac{1}{n} \sum_{i=1}^n \frac{1}{d_i} \quad (2)$$

where n is the number of black pixels not occluded by other contour pixels, and d is the distance from p to each of the black pixels not occluded. Figure 1(b) visualizes such a result, assigning red to high potential values and blue to low values.

Some studies have aimed to apply the PPF as an image feature for quantitative evaluation of impressions. Researchers have proposed different evaluation methods for their objectives, such as the spacing method for readable arrangements of characters within a specified area [12, 13], the design of arch bridges [14], and the impression of female faces with various hairstyles [5]. These analytic metrics are based on shape analyses for equipotential lines, defined by connecting equipotential values, and have demonstrated potential for evaluating impressions.

2.2 Fixation map

Numerous studies have adopted eye-tracking for various objectives because the eyes can provide important information about human visual perception. For instance, eye-tracking has been used for user interfaces [15, 16], for predicting visual saliency [17], and for analyzing borderline personality disorder patients' emotions [18]. Eye-tracking information includes eye saccades and eye fixations as primary information. An eye saccade is defined as rapid movement of the fovea from one point of interest to another, while eye fixations are defined

as periods of time during which the eye remains aligned with a target [19]. A fixation map presents a visual image based on eye fixations for a target. In general, a fixation map can be generated from eye-tracking data, although predictive maps have also been studied, such as saliency maps [17, 20, 21] and cursor-based attention-tracking [22, 23], as eye-trackers are expensive.

In general, four metrics are applied when creating a fixation map: fixation count, absolute fixation duration, relative fixation duration, and participant percentage (see Ref. [24] for a discussion of each metric). Below, we describe the absolute fixation duration metric, which is adopted in this study to confirm our comparative results based on the viewing duration.

A fixation map is created in three steps [6]. First, a blank image is prepared, and fixation points are rendered and accumulated in the image with a scaling function. A Gaussian function is most commonly used for scaling [25], with σ depending on the situation. Second, the fixation intensities are normalized. For example, short-duration fixations may be disregarded while significant fixations on objects of interest are highlighted. Third, colorization is applied for visualization. The most popular colorization tools are brightness and rainbow gradient colormaps. Figure 1(c) shows a fixation map with a rainbow gradient colormap, with red assigned to high-fixation areas and blue to low-fixation areas.

3 Comparison methodology

We compare the PPF and fixation map to determine whether the two results are identical in their stimulus response and to assess whether a fixation map will gradually approach the PPF if measured for a sufficient duration. For our objective, we used multiple stimulus images and conducted eye-tracking for short and long durations to generate two types of fixation maps. Below, we describe the experimental settings for eye-tracking, the stimulus images, the eye-tracking experiments, and the similarity metric applied for comparison.

3.1 Experimental setup

Our experimental setup is shown in Fig. 2. Each participant sat in front of a display, with head fixed using a static chin rest [26]. The Tobii Eye Tracker 4C was adopted in this work [27]. The distance from

Fig. 2 Experimental setup. The angle of foveal vision was 2° ; σ is a Gaussian parameter used in generating a fixation map.

the participant's eye to the center of the display was approximately 70 cm, and the angle of foveal vision was 2° , as determined for a general foveal area (0° – 2°) [6]. The σ value of the Gaussian function for generating the fixation map was derived from these parameters.

3.2 Stimulus images

As PPF theory has only been established for binary images, and aiming to reduce the burden on the participants, we used ten stimulus images depicting simple line-drawing objects, as shown in Fig. 3. The leftmost eight images are from the MIT saliency benchmark [7]; the original grayscale images were binarized to produce the PPF. The two rightmost images are meaningless shapes not associated with common objects, that multiple observers could not identify in a previous study [28]. The stimulus image resolution was 1080×1080 , and each image was shown at the center of the display. Pixels not corresponding to the stimulus image are rendered in gray.

3.3 Eye-tracking experiments

We conducted eye-tracking experiments for a short duration as a free-viewing task and for a long duration as an observation task. A total of 20 experimental participants were included in this study, with 10 (five males and five females) assigned to the free-viewing task and 10 (five males and five females) assigned to the observation task. All the participants were Japanese students of Osaka University of Economics. Their mean age was 20.0 years old, with a standard deviation (SD) of 1.3 years. Stimuli names (see Fig. 3) were not given for either task.

3.3.1 Free-viewing task

In a preliminary experiment with other, different, participants, we confirmed that they stopped viewing each stimulus after a few seconds when instructed to freely view the stimulus. Hence, as the short-duration

Fig. 3 Stimulus images. The left eight images were downloaded from the MIT saliency benchmark [7] and binarized. The two rightmost images are meaningless shapes from Ref. [28]. From upper left to lower right: Door, Cup, Mouse, Dolphin, Meaningless shape 1 (MS1), T-shirt, Eye, Umbrella, Dog, and Meaningless shape 2 (MS2).

free-viewing task, the participants were asked to freely view each stimulus. However, the participants could stop viewing a stimulus and move on to the next stimulus at an arbitrary timepoint because the viewing time for each stimulus was not defined. The order in which stimuli were displayed was randomly determined per participant.

3.3.2 Observation task

As noted in the previous section, the participants generally stopped viewing the stimuli after a few seconds. If stimuli are presented to the participants for a long duration, they may be tired of viewing them, resulting in complicated eye fixation results. Thus, it is important to take care when designing a task in which the participants must view the stimuli for a long duration.

Hence, for the long-duration observation task, the participants were asked to sketch each stimulus after the viewing step. Specifically, the participants viewed a stimulus, which was automatically hidden after 30 s, and then created a line drawing of the stimulus on paper. The sequential order for displaying and drawing each stimulus was random. Because the participants must memorize each stimulus while it is displayed, they must observe each stimulus carefully. We note that the pictures drawn by the participants were not used for gaze analyses.

3.4 Similarity metric

To evaluate the similarity of fixation maps and saliency maps, similarity metrics may be classified into location-based metrics and distribution-based metrics. Location-based metrics, which include

area under ROC curve (AUC) [29–31], normalized scanpath saliency (NSS) [32], and information gain (IG) [33], evaluate saliency values at discrete fixation locations. The distribution-based metrics, which include Pearson's correlation coefficient (CC) [34] and similarity (SIM) [35], evaluate both fixation maps and saliency maps as continuous distributions. Currently, NSS and CC have been recommended as the fairest comparison [36].

To calculate a similarity value in the PPF and the fixation map, all the above metrics can be similarly applied. However, here, it should be considered that the PPF does not produce potential values on black pixels. Usually, when viewing line drawings, many important fixation points would be placed on black pixels. This means that many important fixations cannot be evaluated when a location-based metric is adopted. However, there is little significant impact on the similarity of distribution-based metrics even when excluding black pixels because of continuous distributions. Therefore, we adopted CC excluding the values of foreground pixels as the similarity metric in this study. However, the general rule of thumb that a strong correlation is assumed for values exceeding 0.6 does not hold, because the PPF and the fixation map are based on obviously different functions.

4 Results and discussion

4.1 Gender difference in fixation maps

The cultural or educational backgrounds of the participants can be considered almost identical, as

noted in Section 3.3. However, gaze data include gender differences in some cases [37]. To confirm the necessity of comparing the PPF and the fixation map separately by gender, we determined the presence of a gender difference on the fixation maps for both tasks. Specifically, the fixation maps by gender in each task were generated and then compared using CC. In the free-viewing task, where the viewing durations differed for each participant, the fixation maps of each stimulus were normalized by the time for each participant. Specifically, fixation maps for each participant were generated separately and normalized by each participant's viewing duration, and were averaged to generate the final fixation map. When the results were grouped by gender, the mean CC of the free-viewing fixation maps for all stimuli was 0.94 (SD = 0.04) and 0.95 (SD = 0.03) for the observation task, indicating a lack of gender difference. Nevertheless, in the free-viewing task, it appears that females viewed the stimuli for longer, as shown in Fig. 4; however, it is difficult to calculate the statistical differences because the number of participants is low and the SDs are large. Although it cannot be determined whether this is a general trend, further investigation is needed to determine whether viewing durations differ according to gender in a statistically significant way, and fixation maps have no gender difference. In the following discussion, the fixation map results were generated from all ten participants, including both males and females.

4.2 Comparison and discussion

The PPF and fixation map results are shown in Fig. 5 for each task. The PPFs in Fig. 5 (and Fig. 1(b)) are highlighted to aid visualizing the results, as the values exponentially increase near the shapes; thus,

Fig. 4 Mean viewing duration and SD for the free-viewing task.

the PPFs display M'_p instead of M_p , where

$$M'_p = \min(1.0, \alpha M_p / M_{\max}) \quad (3)$$

and $M_{\max}$ is the maximum value of M_p . M_p has a range of $0 \leq M_p \leq 1$, and the bias value is $\alpha = 20$. Note that M'_p is only applied for visualization, while M_p is used in similarity calculations.

Clearly, the PPFs and fixation maps look different. Moreover, for most stimuli, the fixation maps show a broader view for the observation task than the free-viewing task. Considering the difference of approximately 3- to 8-fold in each task's viewing duration, as shown in Fig. 4, the fixation maps for the two tasks show little difference. This trend suggests that the PPF and the fixation map are different, regardless of the viewing duration.

The similarity metric values, as described in Section 3.4, are shown in Fig. 6. The similarity values are not constant, with different results for each stimulus. Additionally, Fig. 7 shows that the similarity changes with viewing duration for the observation task. For all stimuli, the similarities increase sharply for the first few seconds of viewing and then stabilize. Although the PPFs and fixation maps could be identical if the similarity values of all stimuli converged at a single time point, such a trend is not observed. Moreover, considering the maximum similarity value of 1.0, the results are far from completely identical. Additionally, although we cannot exclude the possibility that the two results would be identical at viewing durations exceeding 30 s, this possibility is highly unlikely because the similarities stabilize with increasing duration. However, there is a relationship between the two results, and all similarity values are positive, indicating a positive correlation. In conclusion, the PPF and fixation map represent different visual perception properties; moreover, the PPF does not represent noticeability and cannot replace human gaze measurements.

Therefore, it is important to determine what results can be derived from the similarity value. As shown for the two tasks in Fig. 6, for most stimuli, the similarity values in the observation task are higher than the free-viewing task values. However, they are not substantially higher; rather, drastic increases occur within the first few seconds, as shown in Fig. 7. Since the participants stopped viewing the stimulus after a few seconds, as shown in Fig. 4, this

Fig. 5 Visualization of PPFs and fixation maps for each stimulus shown in Fig. 3. Top: PPF results. Middle: fixation maps for the free-viewing task. Bottom: fixation maps for the observation task.

trend may be related to the viewing duration. This relationship should be clarified through additional experiments because this similarity may capture important human visual properties, such as cognition

and understanding.

Moreover, the similarity values differ for each stimulus, as shown in Figs. 6 and 7, with high similarity for the Dolphin, Eye, and Dog and

Fig. 6 Similarities between the PPFs and fixation maps.

Fig. 7 Similarity values for the observation task for each duration time. The similarity values at 30 s correspond to the data in Fig. 6.

low similarity for MS1 and MS2 (see Fig. 3 for stimulus names). This trend indicates that animate objects exhibit high similarity while inanimate objects exhibit low similarity. In general, for rapid categorization, human perceptual processing categorizes animals and non-animal objects on a priority basis [38, 39]. Supposing that this animal/non-animal categorization is based on animate/inanimate categorization, the finding that the similarity values for animate objects are high may represent rapid categorization. However, the results cannot distinguish whether the similarity represents either or both animate/inanimate and animal/non-animal categorization. Furthermore, as described in Section 3.2, the meaningless shapes are not associated with common objects. If the meaningless shapes can be interpreted as “unknown shapes that have never been viewed before”, with the other shapes

categorized as “known shapes”, the similarity results may indicate a degree of cognition or awareness.

In relation to these findings, in natural scenes in color photographs, categorical information is extracted in parallel from across the visual field independently of spatial attention [40]. In addition, it has been reported that natural scene categorization from color photographs and line drawings generates similar neural activity [2]. Because we used a single object line drawing for each stimulus, this study is related to single object categorization as a part of natural scene categorization. Therefore, our findings may be helpful in clarifying a cognitive mechanism in the early stages of scene categorization. In addition, a computational cognition model of perception, memory, and judgement (PMJ), which partitions the cognitive process into three stages, has been proposed [41]. Furthermore, based on PMJ, a computational cognition framework has been proposed to simulate natural scene categorization from line drawings [42]. As PPF and PMJ are different theories, our findings also indicate the possibility that categorical information can be extracted from different computational approaches.

5 Conclusions

To determine whether the PPF and fixation map are different, we qualitatively and quantitatively compared them, for the first time, in this study. With ten stimulus images depicting binary line drawings, we designed two eye-tracking experiments, including a free-viewing task and an observation task, to generate fixation maps with short and long durations. The free-viewing task participants viewed the stimuli for

an arbitrary duration, while the observation task participants viewed the stimuli for 30 s to prepare for subsequent sketching. In addition, a dedicated similarity metric based on correlations was adopted for the comparison.

The comparative results demonstrated that the PPF and fixation map for binary line drawings are different, regardless of the viewing duration, and represent different properties of visual perception. Moreover, the similarity results suggest the potential for indicating perception in animate/inanimate or animal/non-animal categorization as well as a degree of cognition or awareness in human vision. Additionally, although most conventional PPF studies have independently analyzed the PPF, our results indicate the importance of analyzing the PPF in conjunction with other methods, such as the fixation map. Future research can be performed to explore the above-mentioned possibilities through additional experiments, for example, by using other stimuli. In addition, further investigation is required to clarify gender differences in viewing duration and whether there is a better similarity metric than CC.

Acknowledgements

We would like to express our great appreciation for the experimental participants and the staff of Osaka University of Economics.

References

- [1] Liu, Y. J.; Yu, M. J.; Fu, Q. F.; Chen, W. F.; Liu, Y.; Xie, L. X. Cognitive mechanism related to line drawings and its applications in intelligent process of visual media: A survey. *Frontiers of Computer Science* Vol. 10, No. 2, 216–232, 2016.
- [2] Walther, D. B.; Chai, B.; Caddigan, E.; Beck, D. M.; Fei-Fei, L. Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences of the United States of America* Vol. 108, No. 23, 9661–9666, 2011.
- [3] Fu, Q. F.; Liu, Y. J.; Dienes, Z.; Wu, J. H.; Chen, W. F.; Fu, X. L. Neural correlates of subjective awareness for natural scene categorization of color photographs and line-drawings. *Frontiers in Psychology* Vol. 8, 210, 2017.
- [4] Yokose, Z. A study on character-patterns based upon the theory of psychological potential field. *Japanese Psychological Research* Vol. 12, No. 1, 18–25, 1970.
- [5] Awano, N.; Morohoshi, K. Objective evaluation of impression of faces with various female hairstyles using field of visual perception. *IEICE Transactions on Information and Systems* Vol. E101.D, No. 6, 1648–1656, 2018.
- [6] Courtemanche, F.; Léger, P. M.; Dufresne, A.; Fredette, M.; Labonté-Lemoyne, É.; Sénécal, S. Physiological heatmaps: A tool for visualizing users' emotional reactions. *Multimedia Tools and Applications* Vol. 77, No. 9, 11547–11574, 2018.
- [7] MIT saliency benchmark. Available at <http://saliency.mit.edu/>.
- [8] Ichikawa, N. The measurement of the figure-effect in the third dimension by the light threshold method. *The Japanese Journal of Psychology* Vol. 38, No. 5, 274–283, 1967.
- [9] Kaji, S.; Yamane, S.; Yoshimura, M.; Sugie, N. Contour enhancement of two-dimensional figures observed in the lateral geniculate cells of cats. *Vision Research* Vol. 14, No. 1, 113–117, 1974.
- [10] Fukouzu, Y.; Itoh, A.; Yoshida, T.; Shiraishi, T. An analysis of the figure by the visual space transfer model: A study on elucidation and estimation of a scene on figure recognition (6). *Bulletin of Japanese Society for Science of Design* Vol. 45, No. 4, 75–82, 1998.
- [11] Nagaishi, M. Identifying ability of a recognition method based on the field of induction. In: *Proceedings of the 2nd International Conference on Document Analysis and Recognition*, 926–929, 1993.
- [12] Miyoshi, M.; Shimoshio, Y.; Koga, H.; Shimoda, M.; Uchimura, K. A spacing method for readable arrangement of characters. In: *Proceedings of the International Conference on Human–Computer Interaction*, Vol. 1, 1348–1352, 2001.
- [13] Miyoshi, M.; Shimoshio, Y.; Koga, H. Automatic lettering design based on human kansei. In: *Proceedings of the International Conference on Human–Computer Interaction*, 2005.
- [14] Onaga, H.; Fukouzu, Y.; Yoshida, T.; Shiraishi, T. Relation between potential and curvature optical illusion: A study on elucidation and estimation of a scene on figure recognition (report 5). *Bulletin of Japanese Society for the Science of Design* Vol. 43, No. 2, 77–84, 1996.

- [15] Majaranta, P.; Bulling, A. Eye tracking and eye-based human–computer interaction. In: *Advances in Physiological Computing. Human–Computer Interaction Series*. Fairclough, S.; Gilleade, K. Eds. Springer London, 39–65, 2014.
- [16] Lutteroth, C.; Penkar, M.; Weber, G. Gaze vs. mouse: A fast and accurate gaze-only click alternative. In: *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 385–394, 2015.
- [17] Borji A.; Itti, L. CAT2000: A large scale fixation dataset for boosting saliency research. *arXiv preprint arXiv:1505.03581*, 2015.
- [18] Kaiser, D.; Jacob, G. A.; van Zutphen, L.; Siep, N.; Sprenger, A.; Tuschen-Caffier, B.; Senft, A.; Arntz, A.; Domes, G. Biased attention to facial expressions of ambiguous emotions in borderline personality disorder: An eye-tracking study. *Journal of Personality Disorders* Vol. 33, No. 5, 671S8, 2019.
- [19] Types of eye movements. Available at <https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/types-of-eye-movements/>.
- [20] Jiang, M.; Huang, S. S.; Duan, J. Y.; Zhao, Q. SALICON: Saliency in context. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1072–1080, 2015.
- [21] Pan, J. T.; Sayrol, E.; Giro-I-nieto, X.; McGuinness, K.; O'Connor, N. E. Shallow and deep convolutional networks for saliency prediction. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 598–606, 2016.
- [22] Tavakoli, H. R.; Ahmed, F.; Borji, A.; Laaksonen, J. Saliency revisited: Analysis of mouse movements versus fixations. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6354–6362, 2017.
- [23] Kim, N. W.; Bylinskii, Z.; Borkin, M. A.; Gajos, K. Z.; Oliva, A.; Durand, F.; Pfister, H. BubbleView: An interface for crowdsourcing image importance maps and tracking visual attention. *ACM Transactions on Computer–Human Interaction* Vol. 24, No. 5, Article No. 36, 2017.
- [24] Bojko, A. Informative or misleading? Heatmaps deconstructed. In: *Human–Computer Interaction. New Trends. Lecture Notes in Computer Science, Vol. 5610*. Jacko, J. A. Ed. Springer Berlin Heidelberg, 30–39, 2009.
- [25] Holmqvist, K.; Nystrom, M.; Andersson, R.; Dewhurst, R.; Jarodzka, H.; van de Weijer, J. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.
- [26] NAMOTO TKD-UK1. Available at <http://www.namoto.com/vision/chinorest.htm>.
- [27] Tobii Eye Tracker 4C. Available at <https://gaming.tobii.com/tobiieye-tracker-4c/>.
- [28] Aramaki, E.; Nakamura, T.; Usuda, Y.; Kubo, K.; Miyabe, M. Naming of meaningless sketch image. In: *Interactive Information Access and Visual Mining*, 47–51, 2013.
- [29] Judd, T.; Ehinger, K.; Durand, F.; Torralba, A. Learning to predict where humans look. In: *Proceedings of the IEEE 12th International Conference on Computer Vision*, 2106–2113, 2009.
- [30] Borji, A.; Sihite, D. N.; Itti, L. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing* Vol. 22, No. 1, 55–69, 2013.
- [31] Borji, A.; Tavakoli, H. R.; Sihite, D. N.; Itti, L. Analysis of scores, datasets, and models in visual saliency prediction. In: *Proceedings of the IEEE International Conference on Computer Vision*, 921–928, 2013.
- [32] Peters, R. J.; Iyer, A.; Itti, L.; Koch, C. Components of bottom-up gaze allocation in natural images. *Vision Research* Vol. 45, No. 18, 2397–2416, 2005.
- [33] Kümmerer, M.; Wallis, T. S. A.; Bethge, M. Information-theoretic model comparison unifies saliency metrics. *Proceedings of the National Academy of Sciences of the United States of America* Vol. 112, No. 52, 16054–16059, 2015.
- [34] Rodgers, J. L.; Nicewander, W. A. Thirteen ways to look at the correlation coefficient. *The American Statistician* Vol. 42, No. 1, 59–66, 1988.
- [35] Riche, N.; Duvinage, M.; Mancas, M.; Gosselin, B.; Dutoit, T. Saliency and human fixations: State-of-the-art and study of comparison metrics. In: *Proceedings of the IEEE International Conference on Computer Vision*, 1153–1160, 2013.
- [36] Bylinskii, Z.; Judd, T.; Oliva, A.; Torralba, A.; Durand, F. What do different evaluation metrics tell us about saliency models? *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 41, No. 3, 740–757, 2019.
- [37] Sammaknejad, N.; Pouretamad, H.; Eslahchi, C.; Salahirad, A.; Alinejad, A. Gender classification based on eye movements: A processing effect during passive face viewing. *Advances in Cognitive Psychology* Vol. 13, No. 3, 232–240, 2017.

- [38] Zhu, W. N.; Drewes, J.; Gegenfurtner, K. R. Animal detection in natural images: Effects of color and image database. *PLoS One* Vol. 8, No. 10, e75816, 2013.
- [39] Zhu, W. N.; Drewes, J.; Peatfield, N. A.; Melcher, D. Differential visual processing of animal images, with and without conscious awareness. *Frontiers in Human Neuroscience* Vol. 10, 513, 2016.
- [40] Peelen, M. V.; Li, F. F.; Kastner, S. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* Vol. 460, No. 7251, 94–97, 2009.
- [41] Fu, X. L.; Cai, L. H.; Liu, Y.; Jia, J.; Chen, W. F.; Yi, Z.; Zhao, G. Z.; Liu, Y. J.; Wu, C. X. A computational cognition model of perception, memory, and judgment. *Science China Information Sciences* Vol. 57, No. 3, 1–15, 2014.
- [42] Yu, M. J.; Liu, Y. J.; Wang, S. J.; Fu, Q. F.; Fu, X. L. A PMJ-inspired cognitive framework for natural scene categorization in line drawings. *Neurocomputing* Vol. 173, 2041–2048, 2016.

Naoyuki Awano received his Doctor of Information Science degree from Osaka Institute of Technology in 2012. He was a research associate at Osaka Institute of Technology from 2009 to 2013 and an assistant professor at Seikei University from 2013 to 2018. He has been a lecturer at Osaka University of Economics since

2018. His research interests include image processing, computer graphics, and perceptual information processing.

Yuki Hayashi received his Ph.D. degree in information science from Nagoya University in 2012. From 2009 to 2012, he was a recipient of a JSPS research fellowship for young scientists (DC1). From 2012 to 2014, he was an assistant professor at Seikei University. He is currently an associate professor in the

College of Sustainable System Sciences and the Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University. His research interests include computer-supported collaborative learning and human–computer interaction.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Other papers from this open access journal are available free of charge from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.