Frontiers of Architectural Research (2020) 9, 726–738

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/foar

Research Article

Wayfinding design in transportation architecture — are saliency models or designer visual attention a good predictor of passenger visual attention?

Ran Xu [a], Haishan Xia [a], Mei Tian [b,*]

[a] School of Architecture and Design, Beijing Jiaotong University, Beijing 100044, China
[b] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Received 29 January 2020; received in revised form 23 May 2020; accepted 24 May 2020

KEYWORDS
Transportation architecture design;
Passenger wayfinding;
Path choice;
Visual attention;
Eye fixation

Abstract In transportation architecture, wayfinding quality is a crucial factor for determining transfer efficiency and level of service. When developing architectural design concepts, designers often employ their visual attention to imagine where passengers will look. A saliency model is a software program that can predict human visual attention. This research examined whether a saliency model or designer visual attention is a good predictor of passenger visual attention during wayfinding inside transportation architecture. Using a remote eye-tracking system, the eye-movements of 29 participants watching 100 still images depicting different indoor scenes of transportation architecture were recorded and transformed into saliency maps to illustrate participants' visual attention. Participants were categorized as either "designers" or "laypeople" based on their architectural design expertise. Similarities were compared among the "designers'" visual attention, saliency model predictions, and "laypeople's" visual attention. The results showed that while the "designers'" visual attention was the best predictor of that of "laypeople", followed by saliency models, a single designer's visual attention was not a good predictor. The divergence in visual attention highlights the limitation of designers in predicting passenger wayfinding behavior and implies that integrating a saliency model in practice can be beneficial for wayfinding design.
© 2020 Higher Education Press Limited Company. Production and hosting by Elsevier B.V. on behalf of KeAi. All rights reserved. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

* Corresponding author.
E-mail address: 331650773@qq.com (M. Tian).
Peer review under responsibility of Southeast University.

https://doi.org/10.1016/j.foar.2020.05.005
2095-2635/© 2020 Higher Education Press Limited Company. Production and hosting by Elsevier B.V. on behalf of KeAi. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

www.manaraa.com

1. Introduction

In inside transportation architecture, such as airports, train stations, or multimodal transit hubs, wayfinding quality not only has an effect on transfer efficiency, travel quality (Schmitt et al., 2015), and business performance (Harding, 2019), it is also crucial for evacuation in dangerous situations (Galea et al., 2017). Providing succinct navigation clues to destinations is also effective in lowering passenger anxiety (Chang, 2012; Smahel, 2017). Architects and urban planners (hereinafter referred to as "designers") are responsible for developing wayfinding design concepts in transportation architecture. Walking passengers (hereinafter also referred to as "laypeople") who need to look actively for navigational clues inside transportation architecture are the main beneficiaries of high-quality wayfinding design.

Kevin Lynch's (Lynch, 1960) "The Image of the City" is the foundation of human wayfinding research. According to Lynch, the legibility of the city or "the ease with which [a city's] parts can be recognized and organized into a coherent pattern," is a significant aid for wayfinding tasks. Ensuring that human visual attention can *detect* city parts "with ease" is the premise of being able to *recognize* them "with ease." Studies have shown the close relationship between visual attention and wayfinding. Wiener et al., 2009, 2011, 2012 discovered that people exhibited a gaze bias in the direction of the eventually chosen path option. At present, wayfinding design in architecture is still conducted mainly by imagining the "immersed scenic view" of a user and anticipating their thoughts or information requirements in solving a wayfinding task (Brösamle and Hölscher, 2018). Simply put, designers have to use their visual attention to imagine where laypeople will look. This phenomenon is also ubiquitous in design practice in other sectors. As Rosenholtz et al. (2011) observed, designers often have to use their visual system to predict how their work will be perceived by laypeople. However, studies from other areas have already discovered that experts' eye movements are often differ from those of laypeople (Koide et al., 2015; Crowe et al., 2018). While the difference in eye movement is part of the designers' expertise and allows them to do a better job, it may be a hindrance for transportation architecture designers. If designers do not have similar eye movements to that of passengers, and still use their visual attention to deduce that of the passengers, it will affect their wayfinding design in a negative way. More importantly, an open question is whether designers can trust their visual attention during wayfinding design inside transportation architecture. Hence, in the present study, *Research Question 1* is as follows: in terms of wayfinding design inside transportation architecture, is the designer's visual attention a good predictor of laypeople's visual attention?

In recent years, an increasing number of methods have been developed with the aim of predicting visual saliency during wayfinding in the hope that they can offer reliable pre-evaluations for designs (Caduff and Timpf, 2008; Götze and Boye, 2016; Takemiya and Ishikawa, 2012; Xi et al., 2016). However, most of these methods were not based on human visual attention. Developments in computer vision have made it possible to predict human visual attention via saliency models with high accuracy (Bylinskii et al., 2019a). Compared to the use of eye-tracking to record human eye movements, using saliency models to understand visual attention in wayfinding has the following benefits: (1) workload can be reduced significantly, especially in the preparation and execution of an eye-tracking experiment that involves a large number of participants and covers a large investigation area; (2) designers can integrate saliency models in design processes to receive timely feedback on passengers' visual interactions with architectural factors and make improvements when necessary; and (3) using an appropriate saliency model, pre-evaluations of wayfinding quality can be conducted conveniently and repeatedly for multiple projects in the same type of architecture. Several studies that show how saliency models can provide effective assistance to design practices in other sectors (Rosenholtz et al., 2011; Wilson et al., 2015). However, saliency models are usually oriented towards common visual scenes and humans use different wayfinding clues in different environments (Brosset et al., 2008). Hence, *Research Question 2* is as follows: in terms of wayfinding design inside transportation architectures, how accurately can saliency models predict human visual attention?

This paper aims to answer the two *Research Questions* for the benefit of designers. The answer to the first question reveals how accurate designers are in anticipating passengers' wayfinding behavior when designing transportation architecture. The answer to the second question reveals whether integrating advanced tools from computer vision in design processes can be beneficial for designers. The remainder of the paper is organized as follows: Section 2 provides a literature review on human visual attention and saliency models. In Section 3, the research methodology is explained. The description of the experiments can be found in Section 4. Comparison results can be seen in Section 5. Section 6 offers a discussion of the comparison results and their implications. A conclusion of the paper and a discussion on future work are presented in Section 7.

2. Background

2.1. Human visual attention

Humans are capable of prioritizing the information received from the enormous amount of visual sensory inputs contained in complex visual scenes in real-time. Using visual attention, a subset of information can be selected rapidly and passed on for intermediate and higher visual processing. The eyes are believed to be able to scan a visual scene in a rapid, bottom-up, saliency-driven, and task-independent manner and in a slower, top-down, volition-controlled, and task-dependent manner (Itti et al., 1998; Itti and Koch, 2001). During the free viewing of a visual scene, the bottom-up mechanism is activated so that the most salient locations are fixated on preferentially (Stentiford, 2016). However, researchers are still completing the list of attributes that guide bottom-up visual attention. Currently, much evidence has accumulated in favor of color, orientation, size, and motion as guidance

www.manaraa.com

attributes (Nothegger et al., 2004; Wolfe and Horowitz, 2004). When looking for a particular target, the top-down mechanism helps the eyes fixate on locations that share features with the known target (Tummeltshammer and Amso, 2016).

Research has shown that eye movements can reveal visual attention and cognitive processes during perception (Bang and Wojdynski, 2016; Fogarty and Stern, 1989). For instance, eye fixation represents visual attention towards the object/region of interest, and it can be recorded using an eye-tracking system (see Fig. 1a). Eye fixations can also be transformed into a saliency map, where values are assigned to image regions from 0 to 1, with higher values representing a higher concentration of eye fixations (see Fig. 1b). Studies have also shown that eye-movements in laboratories are, in many respects, comparable to eye movements in the real world (Foulsham et al., 2011).

2.2. Saliency models

In this research, we examined how accurately saliency models can predict human visual attention during way-finding inside transportation architecture. The saliency model, which is a popular research area in the computer vision community, is a computational method based on human cognitive vision processing. It is a mathematical and logical expression of the human ability to analyze and prioritize visual information. The output of a saliency model is a saliency map, where saliency values are assigned to image regions on a scale of 0—1. Higher values represent regions with higher probability of receiving attention from the human visual system (Itti et al., 1998; Zhao and Koch, 2011; Veale et al., 2017). An example of a saliency map generated from a saliency model can be seen in Fig. 1c.

The popularity of saliency models has grown in the past few decades due mainly to two types of demand. The first is that many artificial intelligence applications require a selection mechanism that acts as a substitute for the human visual system to prioritize the vast amounts of information received from the sensory input quickly. Therefore, saliency models are applied in areas where machines need to "see" to react instantly to their environment. For instance, signage detection, monitoring vehicle blind spots (Han and Han, 2017), human body part detection (Jalal et al., 2019), environment exploration for aerial robotics (Dang et al., 2018), and identifying indoor navigation landmarks (Dubey et al., 2019).

The second type of demand, understanding which parts of a scene and which individual patterns in particular attract the viewer's eyes, has been the subject of psychological research for decades. Saliency models have been developed to predict human visual attention as accurately as possible. They also help in better understanding how human visual attention works. The Itti, Koch, and Niebur model (IKN) (Itti et al., 1998) and the Graph-Based Visual Saliency model (GBVS) (Harel et al., 2006) are classic saliency models that imitate early visual processing in the human visual system. Both models extract color, intensity, and orientation (of lines) from the input image as their first step because the cones in the retina are sensitive to colors, the ganglion cells are sensitive to intensity contrast, and the simple primary visual cortex cells are sensitive to lines/edges with different orientations (Hubel, 1988). Then, the IKN calculates the contrast between image regions within each feature channel, thereby producing a series of feature maps. For each channel, one final conspicuity map is produced using these feature maps. Then, the conspicuity maps across all feature channels are combined linearly into one final saliency map. Meanwhile, the GBVS applies weights between every two pixels based on the differences between their locations and feature values. For a particular pixel, the smaller the location distances and the larger the feature differences between it and other pixels, the larger the weight accumulated to it, thereby making it more salient. Central pixels in the image are given higher saliency values automatically due to the nature of the algorithm, which aligns with the fact that human visual attention is usually biased towards the center of a scene.

A large number of recently developed saliency models integrate deep neural networks in their model architectures and need to be trained with special image datasets. This integration enables them to learn independently which features are likely to be salient image regions. It should be noted that deep neural networks are a machine learning method: integrating them into the saliency model architecture does not necessarily mean the model itself resembles the biological working principles of the human visual system. The SALICON (Huang et al., 2015) is based on a deep neural network (i.e., AlexNet (Krizhevsky et al., 2012)) pre-trained for object recognition using millions of training images. It had the highest-ranked prediction accuracy in 2015 according to the MIT Saliency Benchmark (Bylinskii et al., 2019a). The Saliency Attentive Model (SAM) (Cornia et al., 2018b) was published in 2018 and has one of

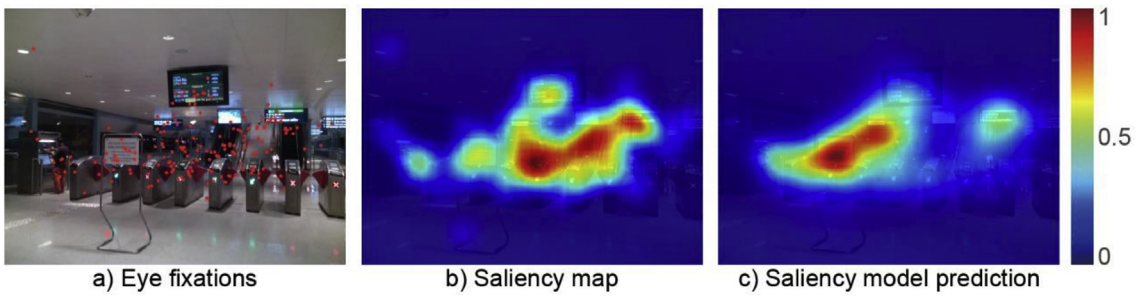


Fig. 1 An example of recorded eye fixations, the corresponding saliency map, and saliency model prediction. In image a), each red dot represents one eye fixation.

www.manaraa.com

the highest prediction accuracies to date, according to the MIT Saliency Benchmark. The model focuses on the most salient regions of the input image to refine the predicted saliency map iteratively due to the integrated Long-Short-Term-Memory in its architecture. It should be pre-trained by SALICON (Jiang et al., 2015), MIT1003 (Judd et al., 2009), MIT300 (Judd et al., 2012), and CAT2000 (Borji and Itti, 2015) image datasets in advance to achieve ideal prediction accuracy.

3. Methodology

Two comparisons need to be performed to address the research questions stated in Section 1.

3.1. Comparison 1

The first comparison aims to identify whether a group of designers, a single designer, or a saliency model can predict laypeople's visual attention in a wayfinding task accurately inside transportation architecture. The comparison is achieved by comparing laypeople's saliency map with that of a group of designers, a single designer, and saliency models. The concrete working steps are as follows:

(1) First, an eye-tracking experiment is performed to obtain human eye fixation data. Participants are given a wayfinding task while looking at static stimuli images of different indoor transportation architecture scenes. Their eye fixation data are recorded using a remote eye-tracking system.
(2) Participants are divided into two separate groups depending on their architectural design expertise: the "designers" group and the "laypeople" group. For each image, the eye fixations of all "laypeople" participants are averaged and transformed into the "laypeople" saliency map and treated as the ground truth. The eye fixations of all "designers" participants are averaged and transformed into a "designers" saliency map, and the eye fixations of every single participant in the "designers" group are transformed into a corresponding "single designer" saliency map. Section 3.4 offers a more detailed description of averaging and transforming procedures.
(3) A "prediction" saliency map is generated for each of the same stimuli images in step (1) using the saliency models.
(4) For each stimuli image, the following similarity evaluations should be made:
    - "Designers" saliency map vs. "laypeople" saliency map
    - "Single designer" saliency maps vs. "laypeople" saliency map, which involves comparing each "single designer" saliency map with the "laypeople" saliency map to obtain the corresponding Similarity Score (SIM) and Pearson's Correlation Coefficient (CC) score. The mean SIM and CC scores are calculated for each image subsequently.
    - "Prediction" saliency map vs. "laypeople" saliency map.

3.2. Comparison 2

The purpose of the second experiment is to determine whether saliency models can predict human visual attention accurately during a wayfinding task inside transportation architecture. The human observer saliency maps are compared with saliency model predictions to accomplish this purpose. The concrete steps are as follows:

(1) The eye-tracking experiment results from step (1) in Comparison 1 (see Section 3.1) are used again. For each stimuli image, eye fixation data of all valid observers are averaged and transformed into an "all observers" saliency map.
(2) The same "prediction" saliency maps from step (3) in Comparison 1 (see Section 3.1) are used.
(3) For each stimulus image, the similarity between the respective "all observers" saliency map and the corresponding "prediction" saliency map is calculated.

3.3. Image selection and categorization

With developments in technology, the airports, train, metro, and bus stations of today are evolving into multimodal transit hubs characterized by a very high volume of passengers, very high departure and arrival frequency, complex architectural spaces, and multiple designed walking routes that interfere with each other. These characteristics are the main causes of wayfinding difficulties in the new era and set new challenges for wayfinding design.

Therefore, the images used in Comparison 1 (see Section 3.1) and Comparison 2 (see Section 3.2) should represent a large variety of indoor scenes of airports, train, metro, and bus stations. They are divided into three groups based on the complexity level, namely the "easy," "medium," and "complex" groups. This arrangement also corresponds loosely to the wayfinding environment development pattern inside transportation architecture from design to actual use.

- In the "easy" group images, architectural spaces, landmarks, and interior decorations are the dominant factors. These factors are not only considered to be important visual guidance elements in wayfinding tasks (Hamburger and Röser, 2014; Hubregtse, 2016; An et al., 2019), they are also the main factors that designers manipulate when developing an architectural wayfinding concept. In these images, visual noise (such as advertisements or crowds) likely to have a negative or complicating effect on wayfinding performance is low.
- In the "medium" group images, in addition to architectural spaces, landmarks, and interior decorations, signage is added to the list of dominant factors. Even though signage is an important visual guidance element for wayfinding (Li and Xu, 2019), it is often designed by graphic designers only after the architects have completed the basic architectural design concept. Visual noise in these images is moderate. For instance, a few passengers can be seen.

www.manaraa.com

- In the "complex" group, the images display visual scenes that are more similar to the architecture after it is put into actual use. The level of visual noise is high. For instance, many advertisements and people can be seen. Therefore, observers may need to invest more attention in finding relevant wayfinding information. In addition to photographic images, architectural renderings (artificial photorealistic images used commonly in architectural design) are also included in this group because they are often produced in such a way that they can simulate the atmosphere after the architecture is put into use.

A further reason for categorizing the images by their complexity level and not by function or context is the growing global need to develop a consistent wayfinding environment across all transit modes to enable seamless multi-modal travel (Massachusetts Bay, 2015; Metrolinx, 2018; City and County of Honolulu, 2019). This study aims to provide general guidance for wayfinding design in transportation architecture, and we expect the acquired knowledge can be transferred to other types of public transportation architecture (including multi-model transit hubs) to help designers meet the wayfinding design challenges in the new era.

3.4. Transforming eye fixation data into saliency maps

The eye tracker's output (described in step 1, Section 3.1 Comparison 1) consists of discrete 2D spatial coordinates of the participant's gaze as it moves from point to point over time. These 2D spatial coordinates form a discrete saliency map with binary classifiers of which image pixels were fixated on and which were not. After obtaining the eye fixation data of each participant, the eye fixations of N participants within the "designers," "laypeople," and "all observers" groups need to be averaged. Based on the study of LeMeur and Baccino (2013), the averaged discrete saliency map f is given by

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} f^i(x). \tag{1}$$

For further analyses, discrete saliency maps should be transformed into continuous saliency maps. The common practice is to blur each fixation location using a Gaussian filter with sigma equal to one degree of visual angle (LeMeur and Baccino, 2013; Bylinskii et al., 2019b). In this research, it was 33 pixels.

3.5. Measurement of similarity between saliency maps

Different methods can be used to compare two saliency maps to evaluate their similarity. In general, evaluation metrics can be divided into location-based and distribution-based metrics (LeMeur and Baccino, 2013; Riche et al., 2013; Bylinskii et al., 2019b). Location-based metrics emphasize the comparison of discrete fixation locations on the maps. Distribution-based metrics compare the underlying distribution of the maps, making the evaluation more robust and tolerant. However, the discrete fixation map has to be transformed into a continuous saliency map, using, for instance, a Gaussian filter, to use the latter metrics. In architectural design, determining which broad regions attract a user's visual attention is more relevant than knowing very exact locations. Thus, the convenient and widely used distribution-based metrics Similarity Score and Pearson's Correlation Coefficient were preferred in this research.

- Similarity Score (SIM)

This score measures the similarity between two distributions. Each distribution is scaled so that they each sum to 1. The similarity is measured by the sum of the minimum values of two respective bins in the two distributions. Given two continuous saliency maps P and Q, they are transformed into two histograms with n bins. Their similarity can be calculated as follows:

$$S(P, D) = \sum_i \min(P_i, Q_i), \tag{2}$$

where $\sum_i P_i = \sum_i Q_i = 1$.

On a scale from 0 to 1, $S = 0$ indicates the absence of similarity between these two maps and 1 means that they are identical. SIM is very sensitive to false negatives (locations shown to be fixated in the ground truth but are not fixated in the prediction saliency map) and penalizes predictions that fail to account for all ground truth density (Bylinskii et al., 2019b).

- Pearson's Correlation Coefficient (CC)

This metric can be used to measure the linear relationship between two continuous saliency maps, H and P:

$$CC = \frac{cov(H, P)}{\sigma_H \sigma_P}, \tag{3}$$

where $cov(H, P)$ is the covariance between maps H and P. σ_H and σ_P are the standard deviations of maps H and P, respectively. The CC varies between -1 and 1, with -1 or 1 indicating a perfect correlation between two maps. Negative/positive values indicate the direction of the correlation. The closer the CC value is to 0, the lower the correlation between the two maps. The CC penalizes false positives (locations that are shown to be fixated in the prediction saliency map, but not fixated in the ground truth) and false negatives equally. High positive pixel-wise CC values occur when both the ground truth and prediction saliency maps have values of similar magnitudes at the same locations.

In this study, the codes for calculating the SIM and CC metrics were downloaded from the MIT Saliency Benchmark website (Bylinskii et al., 2019a). Only absolute values were generated for the CC evaluation.

www.manaraa.com

4. Experiments

Two experiments needed to be performed. The first experiment involves acquiring participants' eye fixation data, which could be transformed into "laypeople," "designers," "single designer," and "all observers" saliency maps. The second experiment involves acquiring "prediction" saliency maps.

4.1. Eye-tracking experiment

4.1.1. Participants

Judd et al. (2012) shows that eye fixation data from around 10—40 observers is sufficient to represent the general eye fixation distribution of an infinite number of observers. Therefore, 29 participants (15 men and 14 women) were recruited to participate in this eye-tracking experiment. All had a normal or corrected-to-normal vision. Their ages ranged from 21 to 58 years old (mean = 26.4, SD = 8.4). Thirteen of them either majored in architectural design or majored in urban planning and received basic architectural design education. The other 16 participants majored in other disciplines.

4.1.2. Stimuli images

In total, 100 still images representing a large variety of indoor scenes from different airports and trains, metro, and bus stations were used as stimuli. The images were categorized into three groups (as explained in Section 3.3). The "easy," "medium," and "complex" image groups had 35, 35, and 30 images, respectively. The "complex" group images included 10 architectural renderings. Care was taken to ensures that the image resolution and details were sufficient for the experiment. The images were either downloaded from the Internet or taken by the authors themselves. These parameters for stimuli images were also not fixed to obtain human eye fixation data comprehensively because renderings come in all kinds of focal lengths and proportion in architectural design practice. Fig. 2 shows some image examples for each group.

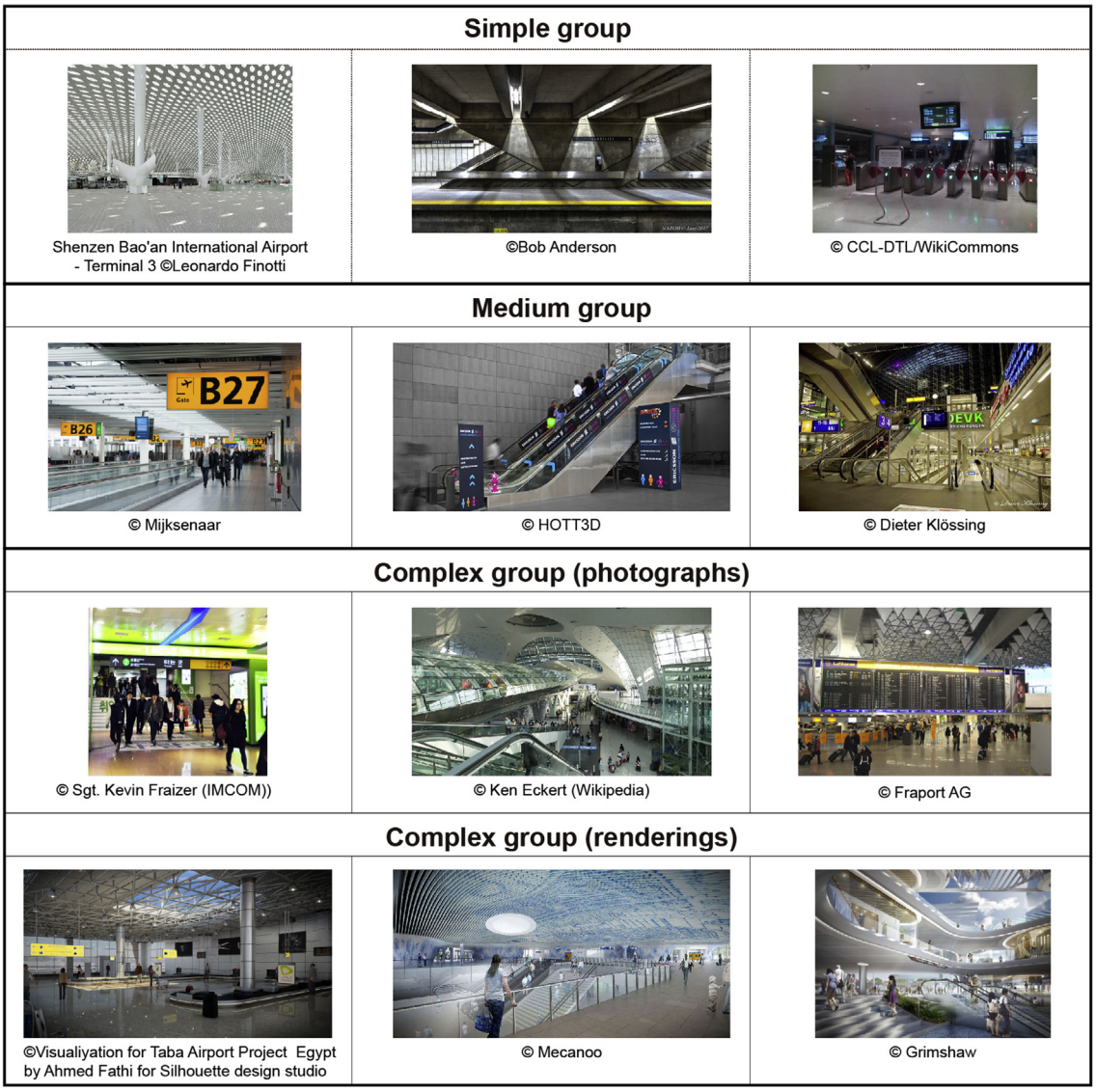


Fig. 2 Examples of the stimuli images in each image group.

www.manaraa.com

4.1.3. Eye-tracking equipment

Participants' eye movements were recorded using a remote RED eye-tracking system developed by Sensor Motoric Instruments, Germany. The apparatus belongs to and is located in the School of Computer and Information Technology, Beijing Jiaotong University, China. Eye-tracking was performed by sending low-power infrared light into the participants' eyes and collecting its reflection. The fixation locations, lengths, and sequences of both of the participant's eyes while watching images displayed on a 22-inch screen (resolution 1280×768) were registered. The sampling rate for eye tracking was set to 120 Hz. The settings for the eye tracker were determined by the associated software iView X. This software was also responsible for rescaling the stimuli images proportionally to fit on the 22-inch monitor. Dark grey frames were added to the blank spaces to minimize visual distractions. The transformation of the eye movement data into fixation data was performed using BeGaze software. The distance between each participant and the screen was set to 700 mm.

4.1.4. Experiment procedure

Before the eye-tracking took place, the aim and content of the experiment were explained to each participant. Several examples were shown to them to ensure that they understood the task fully. An Informed Consent form was also collected. Each participant was given the task of "find your way to board the airplane/train/bus and when you have found the target, press the spacebar on the keyboard immediately." The three groups of stimuli images were shown in order from "easy" to "medium" to "complex," which loosely matches the evolution of the wayfinding environment inside architecture from design to actual use. Before each group of measurements, the eye-tracking system was calibrated for each participant to ensure the tracking accuracy error was below 0.5°. The images were displayed in random order within each image group. For each image, the participant's eye fixations while searching for the target were recorded by the eye-tracking system. After each image, a grey image was displayed on the screen for 1s to allow the participant to rest their mind and eyes. The recordings of eye movements lasted about 10 min per participant.

4.1.5. Experiment results

In total, 2480 eye movement records were collected. For some participants, keeping the eye-tracking accuracy error below 0.5° was not possible and thus, we had to discard the invalid data of two, four, and seven participants in the easy, medium, and complex groups, respectively. Thus, the "easy" group had 27 valid participants (12 designers and 15 laypeople), 25 valid participants in the "medium" group (11 designers and 14 laypeople), and 22 valid participants in the "complex" group (12 designers and 10 laypeople).

After transforming eye fixation data into saliency maps using the method described in Section 3.4, the final experiment results were 12, 11, and 12 "single designer" saliency maps for each of the images in the "easy," "medium," and "complex" groups, respectively. Hence, the study had an overall total of 1165 "single designer" saliency maps. Additionally, for each stimuli image, one averaged saliency map was produced for the "designers", "laypeople", and "all observers" groups, respectively, giving each group a total of 100 averaged saliency maps.

4.2. Saliency model experiment

Four representative saliency models were chosen to predict human visual attention when watching stimuli images. For the classic saliency model, the IKN (improved) and GBVS saliency models were selected (the MATLAB code for both can be downloaded from (Harel, 2019)). SALICON and SAM were chosen as representatives of the advanced type of saliency model. The SALICON developers provide a demo version of the model online (Huang et al., 2019). The SAM developers have published the Python code (Cornia et al., 2018a) of the model pre-trained with the CAT2000 (Borji and Itti, 2015) image dataset. The same 100 stimuli images as in Section 4.1.2 were also loaded into saliency models to generate the corresponding "prediction" saliency maps.

The final experiment results were four different "prediction" saliency maps for each stimuli image, resulting in 400 "prediction" saliency maps in total.

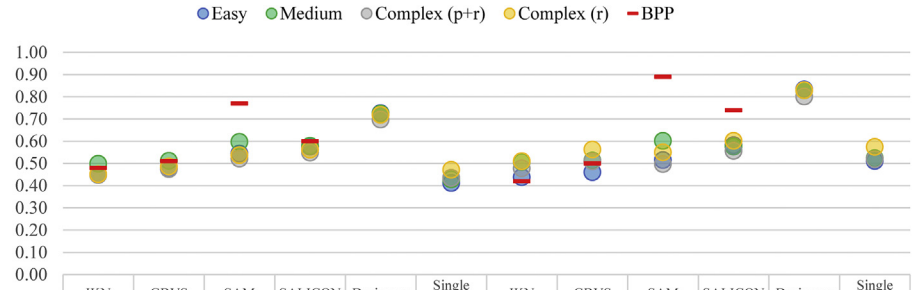
5. Result analysis

5.1. Comparison 1

In Comparison 1, "laypeople" saliency maps were compared with "designers," "single designer," and "prediction" saliency maps. Fig. 3 shows the CC and SIM scores calculated for each image and then averaged within each image group. Fig. 4 shows the standard deviations of the respective averaged SIM and CC scores. In Figs. 3 and 5, the Best Possible Performance (BPP) figure for each saliency model was accessed from the MIT saliency benchmark website (Bylinskii et al., 2019a). The developers of the saliency models calculate these figures as follows. First, standard test images are downloaded from the MIT saliency benchmark website and used to generate saliency maps for these images. Subsequently, the management team of the MIT saliency benchmark website evaluates the average accuracy of these saliency maps by comparing them with the corresponding human eye fixation data. The average accuracy is the BPP of the saliency model, which is published online. This procedure description can also be found on the MIT saliency benchmark website.

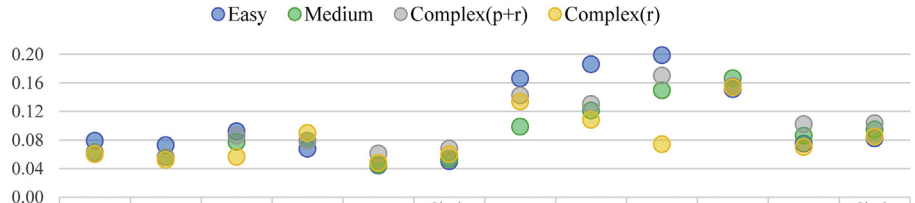
Table 1 shows the SIM and CC scores and the respective standard deviations averaged over all stimuli images. In both SIM and CC evaluations, the "designers" saliency map was the best predictor for that of "laypeople." However, "single designer" saliency maps were the least similar to that of "laypeople" in the SIM evaluation, and the fourth most similar to that of "laypeople" in the CC evaluation. The similarity of the "prediction" and "laypeople" saliency maps lay between the "designers" and "single designer" saliency maps in the SIM evaluation. In the CC evaluation, the IKN and GBVS saliency maps were less similar to that of the "laypeople" saliency map than the SALICON, SAM, "designers," and "single designer" saliency maps. Among the four saliency models, SALICON was the best predictor of "laypeople" visual attention, and its CC and SIM values

www.manaraa.com



	SIM						CC					
	IKN	GBVS	SAM	SALICON	Designers	Single Designer	IKN	GBVS	SAM	SALICON	Designers	Single Designer
Easy	0.45	0.48	0.54	0.58	0.73	0.41	0.44	0.46	0.51	0.58	0.83	0.51
Medium	0.50	0.51	0.60	0.58	0.72	0.43	0.51	0.51	0.60	0.58	0.83	0.52
Complex (p+r)	0.45	0.47	0.52	0.55	0.70	0.43	0.48	0.51	0.50	0.56	0.80	0.52
Complex (r)	0.45	0.49	0.54	0.56	0.72	0.47	0.51	0.56	0.55	0.60	0.83	0.57
BPP	0.48	0.51	0.77	0.60			0.42	0.50	0.89	0.74		

Fig. 3 The Similarity Score (SIM) and Pearson's Correlation Coefficient (CC) score in Comparison 1. The colored blobs represent the SIM or CC scores that have been averaged within the respective image group. The SIM or CC values can be read either from the vertical axis of the graph or from the table below the graph. The table labels easy, medium, complex (p+r), complex(r) represent the image groups "easy", "medium", "complex" (including photographic images and renderings), and "complex" (renderings only), respectively. BPP stands for the Best Possible Performance of the corresponding saliency model.



	SIM						CC					
	IKN	GBVS	SAM	SALICON	Designers	Single Designer	IKN	GBVS	SAM	SALICON	Designers	Single Designer
Easy	0.08	0.07	0.09	0.07	0.04	0.05	0.17	0.19	0.20	0.15	0.07	0.08
Medium	0.06	0.06	0.08	0.08	0.04	0.05	0.10	0.12	0.15	0.17	0.09	0.09
Complex(p+r)	0.06	0.06	0.09	0.08	0.06	0.07	0.14	0.13	0.17	0.16	0.10	0.10
Complex(r)	0.06	0.05	0.06	0.09	0.05	0.06	0.13	0.11	0.07	0.15	0.07	0.08

Fig. 4 Corresponding standard deviations of the Similarity Score (SIM) and Pearson's Correlation Coefficient (CC) in Comparison 1. The colored blobs represent the standard deviation of the respective SIM or CC scores. The standard deviation values can be read either from the graph's vertical axis or from the table below the graph.

Table 1 Similarity Score (SIM), Pearson's Correlation Coefficient (CC) score, and the corresponding standard deviations averaged over all images in Comparison 1.1. Colors have been used to better visualize the value distribution. Green represents higher values and yellow represents lower values.

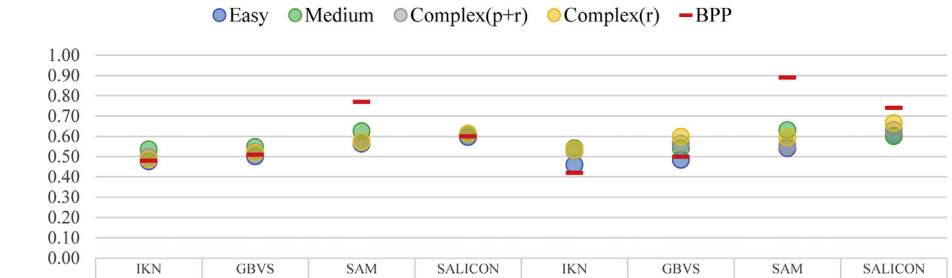
	SIM						CC					
	"Prediction" saliency map				"Designers" saliency map	"Single designer" saliency map	"Prediction" saliency map				"Designers" saliency map	"Single designer" saliency map
	IKN	GBVS	SAM	SALICON			IKN	GBVS	SAM	SALICON		
Average Similarity Value	0.46	0.49	0.55	0.57	0.72	0.42	0.47	0.49	0.54	0.57	0.82	0.52
Average Standard Deviation	0.07	0.06	0.09	0.08	0.05	0.06	0.14	0.15	0.17	0.16	0.09	0.09

www.manaraa.com

were the closest to that of "designers", followed by SAM, GBVS, and IKN.

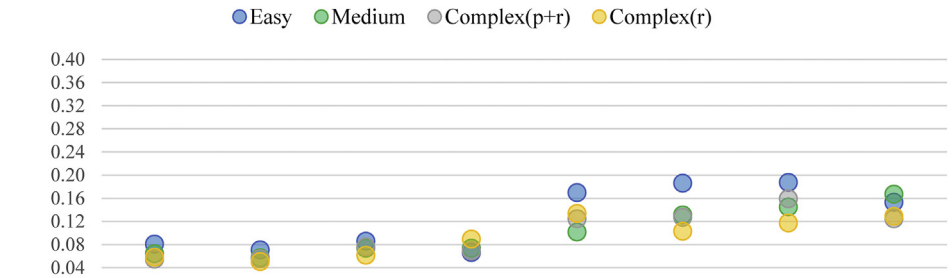
5.2. Comparison 2

In Comparison 2, each "all observers" saliency map was compared with the corresponding "prediction" map using the SIM and CC metrics. The results were then averaged within each image group. Fig. 5 shows the mean SIM and CC scores for the four saliency models while Fig. 6 shows the corresponding standard deviations. Table 2 shows the evaluation results averaged over all images. Overall, each saliency model performed better in Comparison 2 than in Comparison 1. With regards to BPP, in our experiment, the performance of IKN, GBVS, and SALICON was similar (see IKN, GBVS, and SALICON in SIM metric),



	SIM				CC			
	IKN	GBVS	SAM	SALICON	IKN	GBVS	SAM	SALICON
Easy	0.48	0.50	0.56	0.60	0.46	0.48	0.54	0.60
Medium	0.54	0.55	0.63	0.61	0.54	0.54	0.63	0.60
Complex(p+r)	0.50	0.52	0.57	0.61	0.53	0.56	0.55	0.63
Complex(r)	0.49	0.52	0.57	0.61	0.54	0.60	0.59	0.67
BPP	0.48	0.51	0.77	0.60	0.42	0.50	0.89	0.74

Fig. 5 Similarity Score (SIM) and Pearson's Correlation Coefficient (CC) score of saliency models when predicting the "all observers" saliency maps. The colored blobs stand for SIM or CC scores that have been averaged within the respective image group. The SIM or CC values can be read either from the graph's vertical axis or from the table below the graph. The table labels easy, medium, complex (p+r), complex(r) represent the image groups "easy", "medium", "complex" (including photographic images and renderings), and "complex" (renderings only), respectively. BPP stands for the Best Possible Performance of the corresponding saliency model.



	SIM				CC			
	IKN	GBVS	SAM	SALICON	IKN	GBVS	SAM	SALICON
Easy	0.08	0.07	0.09	0.07	0.17	0.19	0.19	0.15
Medium	0.06	0.06	0.07	0.07	0.10	0.13	0.14	0.17
Complex(p+r)	0.05	0.06	0.08	0.07	0.12	0.13	0.16	0.12
Complex(r)	0.06	0.05	0.06	0.09	0.13	0.10	0.12	0.13

Fig. 6 Corresponding standard deviations of the Similarity Score (SIM) and Pearson's Correlation Coefficient (CC) score of saliency models when predicting the "all observers" saliency maps. The colored blobs represent the standard deviation of the respective SIM or CC scores. The standard deviation values can be read either from the graph's vertical axis or from the table below the graph.

Table 2 Similarity Score (SIM), Pearson's Correlation Coefficient (CC) score, and standard deviations of saliency models when predicting the "all observers" saliency maps, averaged over all images. Colors have been used to better visualize the value distribution. Green represents higher values and yellow represents lower values.

	SIM				CC			
	IKN	GBVS	SAM	SALICON	IKN	GBVS	SAM	SALICON
Average Similarity Value	0.50	0.52	0.59	0.61	0.51	0.53	0.57	0.61
Average Standard Deviation	0.07	0.06	0.08	0.07	0.13	0.15	0.16	0.15

www.manaraa.com

or even better (see IKN and GBVS in CC metric), or only slightly worse (see SALICON in CC metric). The ranking of the saliency models in terms of prediction accuracy was unchanged. The SALICON was still the best, followed by SAM and GBVS, whereas IKN had the worst performance.

6. Discussion

6.1. Comparison 1

In Comparison 1, the "designers" saliency maps were the most similar to "laypeople" saliency maps (average SIM = 0.72, average CC = 0.82), while the "single designer" saliency maps were a lot less similar (average SIM = 0.42, average CC = 0.52). However, because the SIM and CC metrics compare the underlying distribution of the maps, it is natural to expect that the larger and the more similar the participant number contained in the respective maps, the higher the similarity. In this research, a high similarity between two saliency maps was more related to participant number and less to design expertise.

Except for the SAM model, Fig. 3 shows that the better the BPP of the saliency model, the closer its SIM and CC metrics were to that of "designers." The SALICON saliency model had the highest scores (average SIM = 0.57, average CC = 0.57), followed by SAM (average SIM = 0.55, average CC = 0.54), GBVS (average SIM = 0.49, average CC = 0.49), and IKN (average SIM = 0.46, average CC = 0.47). Thus, it can be inferred from the results that the SALICON prediction was the closest to that of human visual attention. The performance of the saliency models was generally worse than their published BPPs. The reason may be due to the smaller participant number in the "laypeople" group ($\leq$15 participants), making the "laypeople" saliency maps more individual and less general.

6.2. Comparison 2

Comparison 2 results supported the Comparison 1 results in two aspects. First, the performance ranking of the four saliency models stayed unchanged. Among all four saliency models, SALICON (average SIM = 0.61, average CC = 0.61) was still the most accurate in predicting human visual attention, followed by the SAM (average SIM = 0.59, average CC = 0.57), GBVS (average SIM = 0.52, average CC = 0.53), and IKN (average SIM = 0.50, average CC = 0.51). Second, saliency models were developed to predict general and not individual visual attention. In this comparison, the participant number increased, making the "all observers" saliency map more general and less individual, and thus more similar to the "prediction" saliency map. Therefore, the performance of all four saliency models improved in Comparison 2. Performance improvements for predicting wayfinding visual attention can be expected from the SALICON and SAM models if they are trained specifically and fine-tuned.

In both comparisons, SAM's evaluation results were the furthest from the published BPP, which is due to the SAM model being pre-trained only by the CAT2000 image dataset, while the developer's model used to submit saliency maps to the MIT Saliency Benchmark was pre-trained by several other image datasets.

6.3. Implications for designers

During wayfinding design, designers often use their visual attention to deduce that of laypeople's. The experimental results in this study showed that under certain circumstances, a designer's visual attention can be a good predictor of laypeople's visual attention. The results of Comparison 1 shows the two saliency maps are more likely to be similar if both maps have a large and comparable number of participants. For a small-sized design project, integrating as many designers as possible in the wayfinding design will enhance the overall understanding of passengers' visual interaction with architectural factors.

Thousands to hundreds of thousands of passengers walk through a medium or large-sized transportation architecture every day. It is very difficult for a design project to collect a sufficient number of designers to accurately estimate passenger visual attention for wayfinding design. Not to mention the fact that developing a design usually involves many revisions and adjustments to the concept. Standard practice is that a very limited number of designers are in charge of the design and use their visual attention to estimate that of passengers during wayfinding. The results of Comparison 1 exposed the limitations of a single designer in wayfinding design, which is that anyone person's visual attention is less likely to be similar to a crowd's overall visual attention. This will affect wayfinding design in a negative way. Even though involving as many designers as possible can make up for this limitation, being able to integrate more than 10 designers for the same design task is quite rare in practice. Under these circumstances, applying a saliency model can be very beneficial. As in Comparison 1, the SIM results showed that saliency models are more accurate in predicting laypeople's visual attention than one single designer.

6.4. Discussion of the choice and application of saliency models

If designers want to integrate saliency models to help their design processes, they should choose the model concerning the BPP as published on the MIT Saliency Benchmark website. High BPP certainly means that the model has high prediction accuracy and similarity to human visual attention, as can be seen with the SALICON model. However, designers need to have sufficient time and computer knowledge to pre-train and fine-tune the model. Without proper pretraining, the saliency model's prediction accuracy may be very different from the published BPP (as can be seen in Figs. 3 and 5 regarding SAM). The designer also needs to consider what specific characteristics are important for the design analysis (for instance text signage or architectural space configuration) and conduct fine-tuning when necessary. Once pretraining and fine-tuning have been completed, the model can be applied to design projects of the same architecture type repeatedly with high prediction accuracy. The results of the analysis from the

www.manaraa.com

saliency model will also be more meaningful because they can provide a better estimation of the general human visual attention than a single designer. The low BPP values of IKN and GBVS mean these models have low prediction accuracy and similarity to human visual attention. However, applying these models is relatively convenient and may be suitable if the designer does not have enough time or computer knowledge and only wants a quick check of the wayfinding design.

When integrating saliency models into design processes, designers can input rendered images into the model to generate "prediction" saliency maps. In this way, they can obtain a better understanding of passengers' visual interaction with architectural factors and make improvements when necessary. Additionally, designers can use "prediction" saliency maps as a better way of visualizing their wayfinding design concept. This communication method with other stakeholders has already proven to be very helpful in other design sectors (Rosenholtz et al., 2011).

6.5. Discussion of the difference between the "all observers" and the "prediction" saliency map

The SALICON model has a high mean SIM and CC scores. However, for some stimuli images (e.g., Fig. 7, left images), the difference between the respective "prediction" and "all observers" saliency maps was distinct. In Fig. 7, the respective SALICON saliency maps show the model places high visual saliency values on signage. The model also places more weight on the signage board if it is located on the left (see Fig. 7, SALICON saliency map in the second row). This reaction is probably because the model was instructed to emphasize the text area during the training. Additionally, because humans usually read from left to right, the model was also trained to apply more weight automatically to the left image region. However, both "all observers" saliency maps in Fig. 7 shows high saliency values in image regions that show the vanishing points of the architectural spaces. This phenomenon is coherent with the empirical evidence provided by O'Neill, 1991, who discovered that architectural space usually has a significant influence on humans during wayfinding, regardless of signage. Therefore, when applying the SALICON saliency model in the wayfinding design of transportation architecture, it is necessary to fine-tune it in these aspects.

6.6. Limitations of the research

This research examined whether designers and laypeople exhibit similar visual attention when observing 2D images displayed on a 22-inch screen. Some might consider using Virtual Reality (VR) or panoramic images to be more suitable for this study because these images normally offer a better imitation of real-life environments. However, to this day, work in architectural design practice and education is still based predominantly around 2D renderings (photo-realistic artificial images). According to Lin and Hsu (2017), "it is still not common for architectural design education to use VR platform as a presentation tool" and the majority of students had no previous contact with VR (Brandão et al., 2018).

Additionally, saliency models for panoramic images are still in the early stages of development. For instance, panoramic images are usually stored in equirectangular projections. Methods to cope with this kind of distortion are still under discussion (Zhang et al., 2018). Moreover, the training of new saliency models requires large amounts of data. At present, no commonly used training data for panoramic images exist, which is the same case for saliency models conducting 2D image prediction. It would be very difficult to train the saliency detection for a panoramic image without reliable data and not knowing the relationship between the content features and saliency in 360-degree images. Inappropriate feature extraction and model prediction would lead to unacceptable saliency detection performance. Research on how to combine 360-

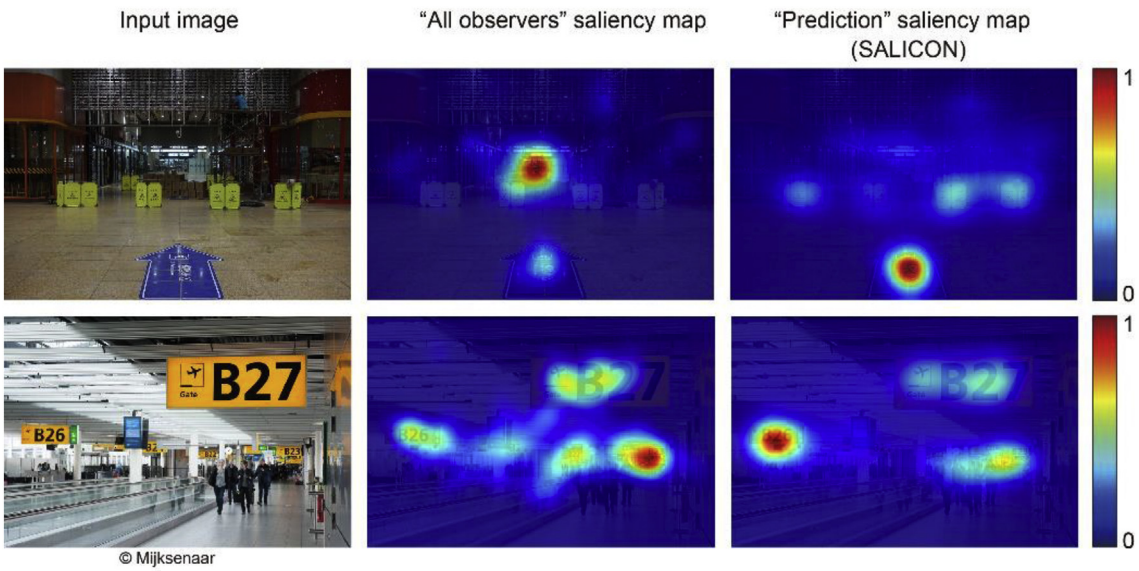


Fig. 7 The input image, the corresponding "all observers" saliency map, and SALICON "prediction" saliency map.

www.manaraa.com

degree image saliency detection with other types of data is still ongoing. For instance, head movement prediction is a crucial factor in predicting visual attention in panoramic images, but a limited number of head tracking datasets can be used for creating panoramic saliency, such as that of David et al. (2018). Moreover, Albayrak et al. explored a set of saliency models to assess how applicable they are for estimating visual saliency in 2D desktop viewing and 3D VR viewing using a head-mounted display (Albayrak et al., 2019). It was observed that integrating depth cues into saliency models did result in a dramatic change in performance. They concluded that integrating depth cues into outdoor scenes could have counterproductive results.

Lastly, using VR devices increases the likelihood of simulator sickness, which may result in early experiment termination and invalid eye fixation data. The factors mentioned above show that at this stage, it may be too early to evaluate visual attention in 3D panoramic images or VR.

7. Conclusion and future research

This study shows that when the numbers of designers and laypeople are relatively large and comparable, designer visual attention can be a good predictor of laypeople's visual attention during wayfinding inside transportation architecture. However, one single designer was not a good predictor in this case. Thus, integrating saliency maps into the design process will be beneficial for the final wayfinding design outcome. Saliency models can be applied to simulate general human visual attention. The designer can make a better prediction of passenger wayfinding behavior inside transportation architecture and they would know where to make the necessary adjustments to the architectural design factors to achieve a better guiding effect. In this experiment, SALICON was the most accurate saliency model in simulating human visual attention, followed by SAM, GBVS, and IKN. According to the SIM evaluation, all four saliency models performed better than one single designer in predicting laypeople's visual attention during wayfinding inside transportation architecture.

In our future research, we intend to verify the comparison results further and integrate more realistic factors, such as motion and depth perception. The eye fixation data collected in this study can be used to improve the prediction accuracy of saliency models. The models can then be applied in actual transportation architecture design to improve the quality of its indoor wayfinding environment.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgment

This material is based upon work supported by The Fundamental Research Funds for the Central Universities Grant No. 2019JBM317.

References

Albayrak, D., Askin, M.B., Capin, T.K., Celikcan, U., 2019. Visual Saliency Prediction in Dynamic Virtual Reality Environments Experienced with Head-Mounted Displays: an Exploratory Study. 2019 International Conference on Cyberworlds, Kyoto, Japan.

An, D.-D., Ye, J.-N., Ding, W., 2019. Spatial features and elements affecting indoor wayfinding—a case study in a Transit Hub. In: Krömker, H. (Ed.), HCI in Mobility, Transport, and Automotive Systems. HCII 2019. Lecture Notes in Computer Science. Springer, Cham.

Bang, H.-J., Wojdynski, B.W., 2016. Tracking users' visual attention and responses to personalized advertising based on task cognitive demand. Comput. Hum. Behav. Part B 55, 867—876.

Borji, A., Itti, L., 2015. CAT2000: A Large Scale Fixation Dataset for Boosting Saliency Research. IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, USA.

Brandão, G.V.L., do Amaral, W.D.H., de Almeida, C.A.R., 2018. Virtual Reality as a Tool for Teaching Architecture. International Conference of Design, User Experience, and Usability, Las Vegas, USA.

Brösamle, M., Hölscher, C., 2018. Approaching the architectural native: a graphical transcription method to capture sketching and gesture activity. Des. Stud. 56, 1—27.

Brosset, D., Claramunt, C., Saux, E., 2008. Wayfinding in natural and urban environments: a comparative study. Cartographica: Int. J. Geogr. Inf. Geovisualization 43, 121—130.

Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., Torralba, A., 2019a. MIT Saliency Benchmark. Retrieved from url. http://saliency.mit.edu/. (Accessed 23 March 2019).

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., Durand, F., 2019b. What do different evaluation metrics tell us about saliency models? IEEE Trans. Pattern Anal. Mach. Intell. 41 (3), 740—757.

Caduff, D., Timpf, S., 2008. On the assessment of landmark salience for human navigation. Cognit. Process. 9 (4), 249—267.

Chang, H.-H., 2012. Wayfinding strategies and tourist anxiety in unfamiliar destinations. Tourism Geogr.: Int. J. Tourism Space, Place Environ. 15 (3), 529—550.

City and County of Honolulu, Department of Planning and Permitting, 2019. Transit-oriented development- wayfinding master plan (phase 1 planning). Retrieved from url. http://www.honolulu.gov/.

Cornia, M., Baraldi, L., Serra, G., Cucchiara, R., 2018a. Python code for saliency attentive model (SAM). Retrieved from url. https://github.com/marcellacornia/sam. (Accessed 23 March 2019).

Cornia, M., Baraldi, L., Serra, G., Cucchiara, R., 2018b. Prediction human eye fixations via an LSTM-based saliency attentive model. IEEE Trans. Image Process. 27 (10), 5142—5154.

Crowe, E.M., Gilchrist, I.D., Kent, C., 2018. New approaches to the analysis of eye movement behaviour across expertise while viewing brain MRIs. Cognitive Res.: Princ. and Implic. 3 (12), 1—14.

Dang, T., Papachristos, C., Alexis, K., 2018. Visual Saliency-Aware Receding Horizon Autonomous Exploration with Application to Aerial Robotics. IEEE Internatonal Conference on Robotics and Automation, Brisbane, Australia.

David, E.J., Gutierrez, J., Coutrot, A., 2018. A Dataset of Head and Eye Movements for 360◦ Videos. 9th ACM Multimedia Systems Conference, Amsterdam, Netherlands.

Dubey, R.K., Sohn, S.S., Thrash, T., 2019. Identifying indoor navigation landmarks using a hierarchical multi-criteria decision framework. MIG' 19: Motion, Interact. Games 15, 1—11.

Fogarty, C., Stern, J.A., 1989. Eye movements and blinks: their relationship to higher cognitive processes. Int. J. Psychophysiol. 8 (1), 35—42.

www.manaraa.com

Foulsham, T., Walker, E., Kingstone, A., 2011. The where, what and when of gaze allocation in the lab and the natural environment. Vis. Res. 51, 1920—2931.
Galea, E., Xie, H., Deere, S., Cooney, D., Filippidis, L., 2017. Evaluating the effectiveness of an improved active dynamic signage system using full scale evacuation trials. Fire Saf. J. 91, 908—917.
Götze, J., Boye, J., 2016. Learning landmark salience models from user's route instructions. J. Locat. Based Serv. 10 (1), 47—63.
Hamburger, K., Röser, F., 2014. The role of landmark modality and familiarity in human wayfinding. Swiss J. Psychol. 73 (4), 205—213.
Han, W.-S., Han, I.-S., 2017. Bio-Inspired Neuromorphic Visual Processing with Neural Networks for Cyclist Detection in Vehicle's Blind Spot and Segmentation in Medical CT Images. Computing Conference, London UK.
Harding, J., 2019. A research-based approach for improving the airport wayfinding experience. J. Airpt. Manag. 13 (2), 133—143.
Harel, J., 2019. Saliency Map Algorithm: MATLAB Source Code. Retrieved from url. http://www.vision.caltech.edu/~harel/share/gbvs.php. (Accessed 23 March 2019).
Harel, J., Koch, C., Perona, P., 2006. Graph-based visual saliency. Adv. Neural Inf. Process. Syst. 19, 545—552.
Huang, X., Shen, C.-Y., Boix, X., Zhao, Q., 2015. SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks. The IEEE International Conference on Computer Vision Santiago, Chile.
Huang, X., Shen, C.-Y., Boix, X., Zhao, Q., 2019. SALICON Demo. Retrieved from url. http://salicon.net/demo/. (Accessed 12 March 2019).
Hubel, D., 1988. Eye, Brain, and Vision. W. H. Freeman & Co., New York.
Hubregtse, M., 2016. Passenger movement and air terminal design: artworks, wayfinding, commerce, and kinaesthesia. Inter. Deseign/Architecture/ Culture 7 (2—3), 155—179.
Itti, L., Koch, C., 2001. Computational modelling of visual attention. Nat. Rev. Neurosci. 2, 194—203.
Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20 (11), 1254—1259.
Jalal, A., Nadeem, A., Bobasu, S., 2019. Human Body Parts Estimation and Detection for Physical Sports Movements. International Conference on Communication, Computing and Digital systems, Islamabad, Pakistan.
Jiang, M., Huang, S.-S., Duan, J.-Y., Zhao, Q., 2015. SALICON: Saliency in Context. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA.
Judd, T., Ehinger, K., Durand, A.F., 2009. Learning to Predict where Humans Look. IEEE International Conference on Computer Vision, Kyoto, Japan.
Judd, T., Durand, F., Torralba, A., 2012. A Benchmark of Computational Models of Saliency to Predict Human Fixations. Massachusetts Institute of Technology DSpace@MIT. MIT-CSAIL-TR-2012-001.
Koide, N., Kubo, T., Nishida, S., Shibata, T., Ikeda, K., 2015. Art expertise reduces influence of visual salience on fixation in viewing abstract-paintings. PloS One 10 (2), 1—14.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems, Lake Tahoe, Nevada, USA.
LeMeur, O., Baccino, T., 2013. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. Behav. Res. Methods 45, 251—266.
Li, Z., Xu, W.-A., 2019. Path decision modelling for passengers in the urban rail transit hub under the guidance of traffic signs. J. Ambient Intell. Humaniz. Comput. 10 (1), 365—378.
Lin, C.-H., Hsu, P.-H., 2017. Integrating Procedural Modelling Process and Immersive VR Environment for Architectural Design Education. 2nd International Conference on Mechanical, Manufacturing, Modeling and Mechatronics (IC4M 2017) — 2017 2nd International Conference on Design, Engineering and Science (ICDES 2017), Kortrijk, Belgium.
Lynch, K., 1960. The Image of the City. MIT Press, Cambridge, Massachusetts.
Massachusetts Bay, Transportation Authority, 2015. Wayfinding by Design. Retrieved from url. https://www.mbta.com/.
Metrolinx, 2018. MAKING IT HAPPEN - Advancing the 2041 Regional Transportation Plan for the Greater Toronto and Hamilton Area. Retrieved from url. http://www.metrolinx.com.
Nothegger, C., Winter, S., Raubal, M., 2004. Selection of salient features for route directions. Spatial Cognit. Comput. 4 (2), 113—136.
O'Neill, M.J., 1991. Effects of signage and floor plan configuration on wayfinding accuracy. Environ. Behav. 23, 553—574.
Riche, N., Duvinage, M., Mancas, M., Gosselin, B., Dutoit, T., 2013. Saliency and Human Fixations, State-Of-the-Art and Study of Comparison Metrics. IEEE International Conference on Computer Vision, Sydney, Australia.
Rosenholtz, R., Dorai, A., Freeman, R., 2011. Do predictions of visual perception aid design? Trans. Appl. Percept. 8 (2), 12.
Schmitt, L., Currie, G., Delbose, A., 2015. Lost in transit? Unfamiliar public transport travel explored using a journey planner web survey. Transportation 42, 101—122.
Smahel, T., 2017. Airport features most likely to affect international traveler satisfaction. Transport. Res. Rec.: J. Transport. Res. Board 2626, 34—41.
Stentiford, F., 2016. Bottom-up visual attention for still images: a global view. In: Mancas, V.P.F.M., Riche, N., Taylor, J.G. (Eds.), From Human Attention to Computational Attention - A Multidisciplinary Approach. Springer, New York.
Takemiya, M., Ishikawa, T., 2012. Computationally determining the salience of decision points for real-time wayfinding support. J. Spatial Inf. Sci. 4, 57—83.
Tummeltshammer, K., Amso, D., 2016. Top-down contextual knowledge guides visual attention in infancy. Dev. Sci. 21 (4), 1—9.
Veale, R., Hafed, Z., Yoshida, M., 2017. How is visual salience computed in the brain? Insights from behavior, neurobiology and modelling. Philos. Trans. Royal Soc. B 372 (1714), 20160113.
Wiener, J., DeCondappa, O., Hölscher, C., 2011. Do you have to look where you go? Gaze behaviour during spatial decision making. Proc. Annu. Meet. Cogn. Sci. Soc. 33 (33), 0.
Wiener, J., Hölscher, C., Büchner, S., Koniecynz, L., 2012. Gaze behaviour during space perception and spatial decision making. Psychol. Res. 76 (6), 713—729.
Wiener, J., Hölscher, C., Büchner, S., Konieczny, L., 2009. How the geometry of space controls visual attention during spatial decision making. Cognitive Sci. Soc. 10, 137.
Wilson, R.T., Baack, D.W., Till, B.D., 2015. Creativity, attention and the memory for brands: an outdoor advertising field study. Int. J. Advert. 34 (2), 232—261.
Wolfe, J., Horowitz, T., 2004. What attributes guide the deployment of visual attention and how do they do it? Nat. Rev. Neurosci. 5 (6), 495—501.
Xi, D.-P., Fan, Q.-J., Yao, X.-B.A., Jiang, W.-P., Duan, D.-C., 2016. A visual salience model for wayfinding in 3D virtual urban environments. Appl. Geogr. 75, 176—187.
Zhang, Z.-H., Xu, Y.-Y., Xu, J.-Y., Gao, S.-H., 2018. Saliency Detection in 360◦ Videos. The European Conference on Computer Vision (ECCV), Munic, Germany.
Zhao, Q., Koch, C., 2011. Learning a saliency map using fixated locations in natural scenes. J. Vis. 11 (3), 74—76.

www.manaraa.com