Understanding and evaluating visual guidance quality inside passenger terminals - a cognitive and quantified approach

Ran Xu, Xing Su and Haishan Xia

*School of Architecture and Design, Beijing Jiaotong University, Beijing, China; †College of Computer Science, Beijing University of Technology, Beijing, China

ABSTRACT

Using the findings from visual cognitive psychology and computer vision, this study proposes a method that can predict how passenger visual attention to indoor visual guidance elements will affect the visual guidance quality inside passenger terminals. A saliency model is used to simulate the human visual attention, so as to understand how the visual guidance elements and visual noise are cognitively perceived by passengers. For every possible origin and destination node combination, the length and probability of the path that passengers are most likely to take (LP) are compared with the respective shortest path (SP). The overall evaluation of the terminal's visual guidance quality can be expressed by the Extra Walking Index. The validity of the developed evaluation method is verified and it is then applied in a case study. The method can be used as a supporting tool for architects to identify relevant architectural features in the design phase and optimize them accordingly; in addition, it can also provide existing passenger terminals with specific improvement suggestions.

1. Introduction

Psychological research shows that our visual and cognitive abilities are fundamental in the wayfinding process (Bosch and Gharaveis 2017). In a passenger terminal, a passenger uses visual guidance elements around him/her to navigate. Traditionally, studies on passenger wayfinding performance are conducted by on-site surveys and interviews (Foljanty 2017; Hölscher et al. 2009). This kind of method not only requires tedious work and tremendous time investment but is also impossible to perform before the building is completed. Therefore, recent researches propose predicting the passenger’s wayfinding performance by evaluating the quality of the terminal’s directional signage system, often using a 3D perception approach (Churchill et al. 2008; Lam et al. 2003; Becker-Asano et al. 2014). However, in addition to directional signage, what are the other important visual guidance elements in the wayfinding task? The visual information inside a passenger terminal can be divided into useful visual information and disturbing visual noise. How will the nearby visual noise (such as advertisement posters and shop boards) affect passenger visual cognition and wayfinding performance?

Recent findings in visual cognitive psychology disclosed the relationship between wayfinding and visual attention (Büchner, Wiener, and Hölscher 2012; Wiener, DeCondappa, and Hölscher 2011). As is shown on the MIT Saliency Benchmark website (Borji et al. 2018), the latest developments in computer vision have made it possible to simulate and predict human visual attention with high precision using saliency models. Therefore, with the saliency model as leverage, this research not only targets answering the questions raised above but also strives to provide an evaluation method to assess the overall visual guidance quality inside a passenger terminal using a cognitive, objective and quantified approach.

The necessity to make pedestrian guidance systems measurable was previously confirmed (Bauer et al. 2018). Compared to other public buildings, the design of the circulation route in a passenger terminal is much more complicated. The quality of transfer inside a passenger terminal imposes significant time and cost expenses on passengers (Guo and Wilson 2011). The efficiency of wayfinding is also crucial in evacuations during a fire or earthquake. Evacuees retrieve indications for their escape route from visual clues such as directional signage (Galea et al. 2017) and architectural spaces (Sun and de Vries 2006). To guide passengers to their destinations safely, efficiently and accurately in any situation, having an in-time, adequate and effective visual guidance system is of vital importance in a passenger terminal building.

Developments in technology and rapid urbanization cause growth in passenger numbers and building scales. Architects find it more and more difficult to use only their empirical knowledge to balance larger variations of design attributes. They often utilize computer
analyses and simulations to help them predict the quality of their design work. This paper bridges the findings in visual cognitive psychology and computer vision with architectural design. It proposes a method that helps architects obtain a clearer understanding of the passenger terminal’s visual guidance quality during the design stages and when a passenger flow simulation cannot yet be performed. Architects can use this information to improve their design when necessary. Terminal management teams can also use this method to identify the problems that currently exist in the best-case scenario (without passengers) and optimize the guidance system’s setup.

2. Literature review

Wayfinding in architecture is built on findings from fields of cognition and environmental psychology to the design of built spaces. It articulates spatial features and circulation systems as well as environmental communication and uses these as main elements to facilitate the movement of people through individual buildings. Our navigation is realized largely with the help of visual information originating from the real 3D world. Signage, visual access and the design of the architecture itself are the visual guidance elements that mainly influence our wayfinding performance within built environments (Weisman 1981). Section 2.1 offers a closer examination of these elements separately. However, these visual guidance elements do not only cooperate with each other to guide passengers to their destination, but also compete with the nearby visual noise for passenger visual attention. This cooperate-compete relationship for human visual attention can be visualized with the help of a saliency model, which will be explained in Section 2.2.

2.1. Visual guidance elements

The architectural space is the fundamental visual guidance element in a passenger terminal and the focus of architects. It was observed that symmetrical, regular and continuous (Canter 1974) or clearly organized spaces (Arthur and Passini 1992) have positive guidance effects on people in their wayfinding tasks. The direct visual contact to the destination point creates the most effective visual guidance (Carpman, Grant, and Simmons 1985). Raubal, Pfoser, and Tryfona (1997) proposed the “image schemata” to evaluate how architectural space facilitates people in wayfinding. The method reconstructed the built space by interviewing people regarding their spatial experiences as they were performing a wayfinding task. Surveys show that 78% of the interviewees consider their wayfinding performance to be related to architectural design (Foljanty 2015, 2017). In the space syntax approach, parameters such as visibility, viewshed and connectivity are often used to calculate the “reachability” between two locations and to predict passenger wayfinding performance (Ueno, Nakazawa, and Kishimoto 2009). This kind of method uses a purely geometric and quantitative approach and provides a very sound analysis for the relationships between different spaces. However, human visual and cognitive abilities in this situation are rarely considered. For instance, even if the visibility between two locations was theoretically unblocked, the visual noise (such advertisement posters or even unorthodox architectural spaces) will likely negatively affect the wayfinding performance.

Directional signage is an essential visual guidance element that assists passenger wayfinding within an architectural space. These two visual guidance elements cooperate closely with each other. It was determined that wayfinding, when assisted by proper signage, appeared more natural when the building function was also reasonably designed (Peponis, Zimring, and Choi 1990). Certain researchers detected that during wayfinding, a reasonable signage design could not compensate for the difficulties caused to humans by unreasonably organized architectural spaces (Arthur and Passini 1992). Reasonably designed directional signage was not only effective in reducing the number of wrong turns or requests for extra information, but its presence also significantly lowered the feeling of crowding, discomfort, anger, and confusion (Wener and Kaminoff 1983). There have been studies that examine directional signage system as an independent visual guidance element. A research was conducted to determine the relationship between the conspicuity and illuminance of the signage in Swiss railway stations and recommended 100 cd/m² to be the optimal value (Lasauskaite and Reisinger 2015). Interviews and surveys were also conducted to quantify how the signage in the Japan Railway East railway system influence cognitive cost in wayfinding (Chang et al. 2018). There have also been studies that investigated directional signage together with architectural spaces. For instance, the influence of the signage on passenger path choice behavior was studied (Li and Xu 2019); by considering only the impact of the directional signage, the researchers reconstructed the path that the passenger was likely to choose, and then the signage was optimized via the Ant Colony Algorithm. An agent-based simulation model was also used to assess the signage’s visibility (Nassar 2011), the humans were modeled as intelligent agents that move in spaces with predefined patterns, and their perceptual attention was modeled considering their field of view as well as the signage location and design.

Landmark as a visual guidance element can also add convenience to the wayfinding process (Epstein and Vass 2014; Hamburger and Röser 2014). Landmarks are used as milestones to form a cognitive map inside the human
brain. Researchers examined the verbal route instructions and sketch maps collected from participants (Schwering, Li, and Anacta 2013). They determined that the inclusion of global landmarks in both verbal instructions and sketch maps were supportive for wayfinding.

A fact that has been largely omitted by the past studies is that visual guidance elements are in constant conflict with the surrounding visual noise. Architectural spaces, directional signage and landmarks intend to note the shortest path to destinations, while distractions such as advertisement posters and shop boards are constantly attempting to lure passengers away from the shortest path. It is very difficult to quantify the impact of visual noise because visual noise cannot be defined, and its content changes constantly with each visual scene. The challenge is that the human visual attention must be simulated to understand how visual guidance elements and visual noise are perceived by passengers. This difficulty is tackled by integrating a saliency model into the evaluation method, as is proposed in this paper.

### 2.2. Saliency model

To understand how visual guidance elements and visual noise are cognitively perceived by passengers, a saliency model is used as leverage in this study. The research on the saliency model is a popular topic in computer vision and its objective is to simulate human visual attention as accurately as possible. Since the saliency model can effectively predict human visual attention on visual scenes (Veale, Hafed, and Yoshida 2017), it has been widely used in areas such as scene recognition (Zhang, Du, and Zhang 2014; Hu et al. 2016), identification and tracking of people (Aguilar et al. 2017b, 2017a), navigation (Wang and Tian 2011), detection and recognition of road signage (Zahabi et al. 2017; Won, Lee, and Son 2008) and many more. The model was inspired by the following visual cognition mechanism: the visual world contains enormous amounts of information, and human visual attention is a scarce resource. To use this resource efficiently, the human eyes developed a selective attention mechanism to prioritize certain parts of the incoming visual information for further processing. First, the highest resolution in human eyes is limited to the central fovea, instead of having an equal resolution throughout the entire field of view; the acuity declines sharply into a low-resolution periphery. Second, the visual system has integrated attention mechanisms that rapidly pre-select locations with certain qualities for further processing, such as areas with higher color contrast, and ignore the areas with less contrast (Wolfe and Horowitz 2004; Koch and Ullman 1985; Das, Bennett, and Dutton 2007; Boynton 2005). A saliency model is the mathematical and logical expression of this pre-selection of exterior information by human visual attention. The model calculates the conspicuity by comparing a particular image region to its surrounding, thus predicting the probability of this region to attract human visual attention. Saliency maps are generated using saliency models, and saliency values are assigned to image regions in the range of 0 to 1, with higher values representing higher probabilities for the eyes to notice this region (Veale, Hafed, and Yoshida 2017; Itti, Koch, and Niebur 1998; Zhao and Koch 2011) (see Figure 1).

Itti, Koch and Niebur proposed the first saliency model in 1998 (Itti, Koch, and Niebur 1998), which divides an input image into three channels: color, intensity and orientation (of lines). Thereafter, filters are used to calculate the contrast between image regions within each channel, thus producing a series of feature maps. For each channel, one final conspicuity map is produced using these feature maps, and then the conspicuity maps across all channels are combined linearly into one final saliency map. This model is based on comparing contrasts between different image regions; Figure 2 shows a simple workflow. The imitation of the early visual processing of the human visual system makes this model the basis of many new saliency models; it remains to be widely applied in many areas.

Traditional saliency model uses feed-forward neural network, meaning that its processing method is not altered with the input data. The recently

![](image1.png)

**Figure 1.** The input image (left) and its saliency map (right).
developed Saliency Attentive Model (SAM) (Cornia et al. 2018) is capable of processing serial data by the integration of Long-Short-Term-Memory (LSTM) (a recurrent neural network), and the accuracy of its prediction results is greatly improved. The LSTM is inspired by the neural cell, which can “memorize” the previously processed input data and frequently update its means of processing based on this memory. The manner in which the data are processed can be called a “cell state”, and three gates are used inside the model to abandon/add information from/to the cell state. These three gates are the forget gate, input gate and output gate. The forget gate decides which information the cell state should forget; the input gate decides which information to retain and generates data that can be appended to the new cell state; and the output gate and the new cell state decide the cell output.

The SAM saliency model focuses on the most salient regions of the input image to iteratively refine the predicted saliency map. Furthermore, one can use standard image datasets and the affiliated eye-tracker data (obtained by tracking the real human eye fixations when watching input data from the dataset) to train the model to achieve higher prediction accuracy. The SALICON dataset is the largest available, with 10,000 training images, 5,000 validation images and 5,000 testing images (Jiang et al. 2015). Experiments show that when observers examine the images, their gazes are biased towards the center of the images (Tatler 2007). This center-biased phenomenon is also included in the SAM model by learning certain priors. The accuracy of saliency models can be measured by comparing the prediction results with real human eye fixations using different metrics (Bylinskii et al. 2016). After training the SAM model with the image dataset, its accuracy was measured as one of the best in the MIT Saliency Benchmark (Borji et al. 2018) and LSUN2017 competition (Yu et al. 2017).

As a tool that can effectively predict human visual attention by processing input image data, saliency models can provide solutions to many visual cognition-related design problems and therefore have been largely applied in design area. For instance, models have been applied in 3D animation modelling to determine which part of the computer model was most likely to attract human visual attention and therefore needed further refinement (Lee, Varshney, and Jacobs 2005). In the advertising sector, designers used saliency models to determine if their design projects would lead to the results they were hoping for (Wilson, Baack, and Till 2015). Dupont et al. (2016) introduced saliency models into landscape design; they provided a series of landscape photos to observers and recorded their eye movements using an eye tracker, and when comparing the record with the calculated saliency maps, they obtained satisfying results. Although saliency models are widely applied in design sectors, their application in architectural design remains limited. Xu et al. used saliency maps to evaluate the visual impact caused by the installation of Building Integrated Photovoltaics on existing buildings with an objective, quantified and cognitive approach (Xu and Wittkopf 2014, 2017; Xu 2016).

As can be concluded from the literature review, signage and landmarks are important visual guidance elements in the wayfinding task, but their effectiveness is very dependent on the organization of architectural spaces. Also, passenger visual attention towards visual guidance elements may be disturbed by the nearby visual noise. Due to the large variations of design attributes, it is often difficult for architects to use their empirical knowledge to estimate the overall visual guidance effect in passenger terminal designs. In this case, saliency model can be integrated as a tool to simulate the human visual attention and to examine the visual guidance quality.

3. The proposed evaluation method

This paper proposes a method that can evaluate the visual guidance quality inside passenger terminals. The Markov chain model has often been used to model human movement (Xia, Zeephongsekul, and Packer 2010) or route choice (Manley, Cheng, and Haworth 2013). Inspired by this notion, a passenger terminal can be modeled as a walkable network $G(V,E)$ consisting of decision nodes $X$ in the set $V$ and links in the set $E$. Figure 3(a) is an example of

![Figure 2. A simplified workflow of the Itti-Koch-Niebur saliency model.](image-url)
a passenger terminal’s floor plan. \( d_{ij} \) is the distance between two connecting nodes \( X_i \) and \( X_j \). For the origin and destination node combination \( X_a \) and \( X_b \), connected by nodes \( X_1, X_2, \ldots, X_n \), the classic Dijkstra (Dijkstra 1959) method can be used to determine the shortest path (SP).

Figure 3(b) is the walkable network \( G(V, E) \) transformed from the floor plan. \( p_{ij} \) is the transition probability between two connecting nodes \( X_i \) to \( X_j \) only under the influence of nearby visual guidance elements. In other words, when a passenger is standing on node \( X_i \), \( p_{ij} \) is the probability of him/her “discovering” the relevant visual information that will guide him/her to the connecting node \( X_j \). For the origin and destination node combination \( X_a \) and \( X_b \), connected by intermediate nodes \( X_1, X_2, \ldots, X_m \), the probability that the passenger is going to choose this path can be calculated as:

\[
P(X_0|X_0) = p(X_0|X_n) \times p(X_n|X_{n-1}) \times \cdots \times p(X_2|X_1) \times p(X_1|X_0)
\]

(1)

Different choice of intermediate nodes will result in different \( P(X_0|X_0) \) value. The path with the highest \( P(X_0|X_0) \) is the path that passengers are most likely to choose (LP).

For the origin and destination node combination \( X_a \) and \( X_b \), the visual guidance quality can be expressed with the Extra Walking Index, which is defined as:

\[
    w_{ab} = (D_{LP(a,b)} - D_{SP(a,b)}) \times \frac{P_{LP(a,b)}}{P_{SP(a,b)}}
\]

(2)

where \( D_{LP(a,b)} \) stands for the length of LP and \( D_{SP(a,b)} \) stands for the length of SP. \( P_{LP(a,b)} \) is the probability of passengers choosing LP and \( P_{SP(a,b)} \) is the probability of choosing SP. The SP path is treated as standard because it is supposed to be the optimal route. The length difference illustrates the extra walking effort that poor visual guidance quality yields. However, using only the length difference is not enough to describe whether only a small percentage or the vast majority of passengers are likely to prefer the LP over the SP. To fill this gap, the probability comparison \( \frac{P_{LP(a,b)}}{P_{SP(a,b)}} \) is used to describe the severity of the issue. A perfect 1 means that the LP and SP are equally likely. A high \( \frac{P_{LP(a,b)}}{P_{SP(a,b)}} \) value indicates that at certain nodes, the visual guidance elements are providing incorrect wayfinding clues to passengers. Taking both the length difference and probability comparison into account, the closer the \( w_{ab} \) is to 0, the better the visual guidance quality between the origin and destination node combination \( X_a \) and \( X_b \).

The final evaluation for the passenger terminal’s visual guidance quality is expressed by the overall Extra Walking Index:

\[
    EWI = \sum_{n} w
\]

(3)
where $\sum w$ is the sum of $w$ of all origin and destination node combinations, and $N$ is the total number of origin and destination node combinations. The closer the $EWI$ is to 0, the more optimal the visual guidance quality inside the passenger terminal.

In the following, the selection of nodes $X$ as well as the determination of transition probability $p_{i,j}$ will be explained in detail.

1. The selection of the nodes $X$

The nodes are selected using the information provided by the architectural floor plans and the designed passenger circulation route. The main purpose of a passenger terminal is to let passenger access transportation vehicles. Therefore, entrances, exits from the metro station, railway platforms, etc. are treated as origin nodes; exits, entrances to the metro station, railway platforms, etc. are set as destination nodes. Important intersections, where passengers would confront direction choices, are treated as intermediate nodes. The intermediate nodes should be located on the main circulation route of the building, and the passenger should be able to reach no less than 2 other nodes from it. The decision regarding where to set a node is dependent on the designed architectural floor plan and not on how visual guidance elements are located. For instance, passengers would not be confronted with a direction choice in the middle of a corridor, therefore setting a node there would be unnecessary. Direct visibility between two connecting nodes is not necessary because wayfinding can be accomplished with the help of nearby visual guidance elements. An entrance node can also be an intermediate node or even an exit node simultaneously and vice versa.

2. The transition probability $p_{i,j}$ between two connecting nodes

When a passenger is standing on the node $X_i$, the probability of him/her “discovering” the relevant visual information that will guide him/her to the connecting node $X_j$ is $p_{i,j}$. To obtain $p_{i,j}$, photos are taken around the node $X_i$ first. Using these photos, respective saliency maps are generated and then manually processed to determine the probability of a passenger “discovering” the relevant visual guidance elements that will take him/her from node $X_i$ to node $X_j$. During manual processing of the saliency maps around node $X_i$, one should concentrate on these three visual guidance elements:

- The architectural spaces

The visual guidance effect of an architectural space can usually be expected from its outstanding brightness due to daylight or natural light and its geometry, such as the horizon line and the perspective vanishing point (see Figure 4); this will be reflected in the saliency values.

- Directional signage

This element mainly refers to signage that provides direction information to other nodes (see Figure 5). Other prominent but irrelevant signs, such as warning signs or advertisement posters, are not included.

- Landmarks

Landmarks are important for passengers in forming a cognitive memory map for navigation. In passenger terminals, landmarks may refer to help desks, meeting points or even an artistic sculpture (see Figure 6).

On each saliency map around node $X_i$, the top 10 local maximum values will be selected to check if they overlap with visual guidance elements that provide information to node $X_j$. A high saliency value either means that this image region is very salient to the passenger, or that he/she is very likely to spend a relative long time processing the information on this image region. Low saliency values either mean that a passenger is less likely to discover this region, or that a passenger will not spend much time processing the corresponding information. It needs to be emphasized that in saliency

![Figure 4. The perspective vanishing point of the architectural space and its saliency value. Photo taken in the Beijing South Railway Station.](image-url)
maps, when only parts of the visual guidance element are marked with high saliency values, it is assumed that the entire element is visually perceived by the passenger.

If none of the visual guidance elements overlap with the top 10 local maximum values on the saliency map, it is deemed that the passenger is unlikely to arrive at node $X_i$ from $X_j$. If positive, the highest saliency value across all the saliency maps around node $X_i$ will be chosen as the temporary visual guidance probability $o_{ij}$ for the connection node $X_i$ to $X_j$. To ensure that the sum of all the outgoing visual guidance probabilities on node $X_i$ equals 1, the $o_{ij}$ is normalized to $p_{ij}$ as follows:

$$p_{ij} = \frac{o_{ij}}{\sum o_i}$$  \hspace{1cm} (4)

The visual guidance probability $p_{ij}$ is the normalized visual guidance probability of the passenger to arrive at node $X_j$ from its connecting node $X_i$ only with the help of nearby visual guidance elements. The higher the number of nodes that $X_i$ is linked with, the richer the visual guidance information that is contained on $X_i$. Then, the normalization step can ensure that the corresponding $p_{ij}$ will be lower because more visual distractors will result in longer reaction time to discover the right target (Wolfe, Palmer, and Horowitz 2010).

4. Validation of the proposed evaluation method

In the following, two validations of the proposed evaluation method will be made based on empirical results from wayfinding studies

(1) Validation 1

In the wayfinding studies of Wiener et al. (2009, 2011, 2012), it was discovered that people displayed a gaze bias in direction of the eventually chosen path option and had a tendency to choose the path option that featured the longer line of sight. Vilar et al. (2013) confirmed that people prefer brighter pathways at intersections. Validation 1 examines the proposed evaluation method based on these notions.

A very simple computational 3D model was created and its floor plan can be seen in Figure 7. There are three different paths connecting the origin node $X_A$ with destination node $X_B$. All paths are of the same width. Path 1 is the longest (71 m), it is very bright and is featured with a long line of sight. Path 2 has a length of 36 m, but it is very dark and has a twist so that its line of sight is shorter. Path 3 is 56 m long, it has the shortest line of sight but is very well lit. Based on the empirical results from wayfinding studies mentioned above, the hypothesis of the path preference order from high to low should be path 1, path 3 and path 2.
To test this hypothesis, three renderings were generated around the origin node $X_A$ (see Figure 8). The viewpoints of the renderings are denoted with arrows and letters $a$, $b$, $c$ in Figure 7. The resulting saliency maps in Figure 8 show the probabilities of passengers to visually detect each path (marked out by red dotted boxes). Thanks to the longer line of sight and high brightness, the temporary visual guidance probability of path 1 is $o_{path1} = 1$. Due to the darkness and shorter line of sight, the temporary probability for passengers to see path 2 is $o_{path2} = 0.11$. Even though path 3 has the shortest line of sight, thanks to its high brightness, its temporary visual guidance probability $o_{path3} = 0.58$. After normalization, the path probabilities are $P_{path1} = 0.60$, $P_{path2} = 0.07$ and $P_{path3} = 0.35$. These results are coherent with the hypothesis mentioned above. The path 1 is the LP and the path 2 is the SP. The overall Extra Walking Index $EWi/(71 - 36) \times (0.60/0.07) = 300.00$.

(2) Validation 2

In an experiment, O’Neill (1991) discovered that even though the addition of signage resulted in an improvement in wayfinding, the architecture’s spatial configuration was found to exert a significant influence regardless of signage. Validation 2 examines the evaluation method based on this observation. This time, the rendering generated from viewpoint as in Figure 7 is equipped with a signboard or a landmark (see Figure 9). The respective saliency maps indicate that a signboard or a landmark will increase the probability of passengers to see path 2. The renderings from viewpoints $b$ and $c$ in Figure 7 remain unchanged. Now, the temporary probabilities have become $o_{path1} = 1$, $o_{path2} = 1$ and $o_{path3} = 0.58$. After normalization, $P_{path1} = 0.39$, $P_{path2} = 0.39$ and $P_{path3} = 0.22$. The overall Extra Walking Index $EWi/(71 - 36) \times (0.39/0.39) = 35$.

After adding the signage or landmark, the decreased $EWi$ value indicates an improvement in the visual guidance quality. Generally speaking, the impact of the architectural space is still very strong regardless of signage or landmark, which is coherent with the observation made by O’Neill (1991) as mentioned above. Even though in both Validation 1 and Validation 2, the length difference between the LP and SP is the same (35 m), but the change in visual guidance elements results in a very different wayfinding environment for passengers. This change is reflected in the visual guidance probability comparison between path 1 and path 2.

In this section, the proposed evaluation method is applied on a simple 3D computational model. The model features an origin node and a destination node. The two nodes are linked by three different

![Figure 7. Floor plan of a simple 3D computer model. The arrows and letters a, b and c around $X_A$ stand for the viewpoints used to generate the respective renderings in Figure 8.](image)
paths that vary in the setting of visual guidance elements. The method’s intermediate analyses as well as final results correspond with empirical results from wayfinding studies. Thus, the validity of the proposed method is verified.

5. Case study

The new Beijing South Railway Station was put into use in 2008 and covers an area of 420,000 m². This station is a transfer station for multiple transportation
types: high-speed railway, buses, cars and the metro. In 2014, more than 100,000 passengers entered the station every day, and the station’s capacity is to be expanded to 190 million passengers by 2020 (Chen et al. 2014). Public opinions (Huo 2018), private interviews with the railway station staff and private observations all indicate that the station’s current visual guidance quality is problematic. The proposed evaluation method was used to test if it can find out the causes behind the dissatisfaction. The station’s overall visual guidance quality was also evaluated.

The Beijing South Railway Station is a five-level modern railway terminal that includes two levels above ground and three underground. The station’s architectural space is provided in Figure 10. Figure 11 shows a simplified floor plan of the Beijing South Railway Station’s basement, which has a simple oval shape. As can also be seen in Figure 12 that the entrances, exits and lifts to the second floor waiting hall are located on the south and north ends of the floor plan. The metro station is situated at the heart of this floor and is surrounded by train arrival exits and commercial facilities on its four sides. Four access points for passengers arriving by private cars and taxis can be found on the west and east sides. Although the passenger circulation route is clearly designed, unobstructed visual contact cannot always be guaranteed due to numerous columns and indoor partitions. In addition, advertisement posters and shop boards also disturb the perception of the existing visual guidance elements. The proposed evaluation method was applied to see if it is possible to evaluate the current visual guidance quality and to identify problems.

There are 9 origin nodes, 13 destination nodes and 22 intermediate nodes in total. Origin nodes are exits from the 2nd floor, from the metro and from the parking garage or entrances from the exterior. Destination nodes are entrances to the 2nd floor, to the metro and to the parking garage or exits to the exterior. To reduce the number of unnecessary nodes, node points are not positioned directly at the entrances; instead, they are positioned where the passenger must pass after entering the terminal. Origin and destination node combinations between node points of the same category are omitted. For instance, after a passenger exits from the metro station, it is highly unlikely that his/her next move is to reenter it. Intermediate nodes are set at main junctions and where passengers would confront direction choices. In total, there are 92 variations of origin and destination node combinations (see Figure 12).

All input photos for the saliency analysis were taken around midnight, when the passenger terminal was nearly empty, with only a few passengers. Thus, the disturbing effect of the crowds was minimized. Photos were taken from each of the node’s four cardinal directions whenever possible. The SAM saliency model was used for this case study. To enhance the model’s prediction accuracy, the SALICON image dataset (Jiang et al. 2015) was used for training.

The distance difference analysis is shown in Table 1 and the probability comparison analysis can be found in Table 2. The overall Extra Walking Index EWI of the Beijing South railway station is 9.08.

Table 1 shows that the maximum distance difference (63 m) can be found between the origin and destination node combination of node 33 to 16.
Figure 12. The different types of nodes.

Table 1. Different origin and destination node combinations and the related distance differences between the LPs and SPs.

| Origin node | 1 | 3 | 4 | 8 | 10 | 13 | 16 | 17 | 21 | 24 | 26 | 29 | 32 |
|-------------|---|---|---|---|----|----|----|----|----|----|----|----|----|
| 8           | 0 | 0 | 0 | - | 0  | -  | 0  | -  | 0  | -  | 34 | 0  | -  |
| 13          | 0 | 0 | 0 | - | 0  | -  | 0  | -  | 0  | -  | 34 | 0  | -  |
| 15          | 9 | - | 9 | 9 | 9  | 0  | 0  | 0  | 0  | -  | 6  | -  | -  |
| 17          | 9 | 9 | 9 | 9 | 9  | 0  | -  | 0  | 0  | 0  | 0  | 6  | -  |
| 21          | 0 | 49| 23 | - | 0  | -  | 0  | 0  | -  | 0  | 0  | -  | -  |
| 22          | 0 | 38| 12 | - | 0  | -  | 0  | 0  | 0  | -  | 0  | 0  | -  |
| 26          | 0 | 38| 12 | - | 0  | -  | 0  | 0  | 0  | -  | 0  | 0  | -  |
| 28          | 0 | 0 | 0 | 0 | 8  | 8  | 21 | 0  | 0  | 0  | 0  | 0  | 4  |
| 33          | 0 | 0 | 0 | 0 | -  | 0  | 63 | 63 | 0  | -  | 0  | 0  | -  |

Table 2. Different origin and destination node combinations and the related probability comparisons between the LPs and the SPs. The abbreviation m means that the LP and SP are identical.

| Origin node | 1 | 3 | 4 | 8 | 10 | 13 | 16 | 17 | 21 | 24 | 26 | 29 | 32 |
|-------------|---|---|---|---|----|----|----|----|----|----|----|----|----|
| 8           | m | m | m | - | m  | -  | m  | -  | m  | -  | 1.4 | m  | -  |
| 13          | m | m | m | - | m  | -  | m  | -  | m  | -  | 1.4 | m  | -  |
| 15          | 1.33| - | 1.33| 1.33| 1.33| 1.33| m  | m  | 2.27| 2.27| 2.27| -  | 1.43|
| 17          | 1.33| 1.33| 1.33| 1.33| 1.33| 1.33| m  | m  | 2.27| 2.27| 2.27| 2.27| 1.43|
| 21          | 1.11| 2.1 | 2.36 | - | m  | -  | m  | -  | m  | -  | m  | m  | m  |
| 22          | 1.11| 1.55| 1.74 | - | m  | -  | m  | -  | m  | -  | m  | m  | m  |
| 26          | 1.11| 1.55| 1.74 | - | m  | -  | m  | -  | m  | -  | m  | m  | m  |
| 28          | m  | m  | m  | 1.37| 1.37| 1.97| m  | m  | m  | m  | m  | m  | 4.23|
| 33          | m  | m  | m  | -  | m  | 1.54| 1.54| m  | -  | m  | m  | -  | -  |
where the LP = [33,12,14,15,16] and SP = [33,34,15,16] (see Figure 13 left). The same distance difference can be found between the origin and destination node combination 33 to 17, where LP = [33,12,14,15,17] and SP = [33,34,15,17] (see Figure 13, right). This is caused by a construction site located between the nodes 33 and 34 (see Figure 14), resulting in a blocked visual contact between these two nodes. In addition, the available signage is not very salient visually, thus the passengers are more likely to make a turn to arrive at their destination, instead of following the originally intended straight line between the origin and final nodes.

The probability comparison analysis in Table 2, in 37 out of 92 origin and destination node combinations, the LPs are not coherent with the SPs. The maximum value (4.23) can be found between the origin and destination node combination 28 to 32, with the LP node sequence being [28,27,25,31,32] and the SP being [28,2,30,31,32] (see Figure 15). This occurs because the node 2 on the SP path is an important junction node that is connected to 6 other nodes (node 1, 3, 5, 28, 29, 30). The visual information around the node 2 is very rich because it has to provide visual guidance information to all 6 connecting nodes (see Figure 16). After normalization, the corresponding transition probability \( p \) towards each connecting node becomes quite low. This also corresponds to the phenomenon that passengers usually hesitate at an unfamiliar important junction node, because a lot of visual information needs to be scanned. In comparison, each node on the LP path is linked to fewer connecting nodes, the corresponding transition probability \( p \) is therefore higher, making this path more preferable.

Figure 17 presents a visual comparison between origin and destination node combinations where the LPs and SPs are not identical. The darker the color of the connecting lines, the more likely this path is preferred by passengers. In both LPs and SPs, the right aisle with the

![Figure 13](image13.png)  
**Figure 13.** LP and SP of origin and destination node combination 33 to 16 (left image) and 33 to 17 (right image). The blue line stands for the SP and the orange line stands for the LP.

![Figure 14](image14.png)  
**Figure 14.** Visual analysis of connection node 33 to 34.

![Figure 15](image15.png)  
**Figure 15.** LP and SP of origin and destination node combination 28 to 32. The blue line stands for the SP and the orange line for the LP.
connections [18, 19, 20, 22, 23, 25, 27, 28] (blue line in Figure 18) are less preferable in comparison to left aisle with the connections [14, 12, 11, 9, 7, 6, 5] (orange line in Figure 18). The reasons are the longer distances between nodes 19 to 20 and 23 to 25 as well as smaller corridor width at certain locations. These have exacerbated the condition of the existing visual pathway. Figure 19(a) shows an example for the connection node 23 to 25 in the right aisle, and Figure 19(b) presents an example for the connection node 9 to 7 in the left aisle.

To conclude the application of the proposed evaluation method on this case study, one can discover that:

- The length difference and probability comparison are as important as the final evaluation result $EWI$ because they can illustrate the difference between the LP and SP, hence identifying the node connections with significant visual guidance problems (for instance the connection node 33 to 34).
- Generally speaking, the evaluation method considers the station’s architectural space to be a prominent factor in visual guidance. This effect is especially apparent in architectural space that provides a long line of sight or a clear horizon line (for instance, architectural spaces displayed in Figure 16).
- Important junction nodes that are linked to a lot of other nodes are physically very convenient for accessing to different destinations. However, this type of node may also contain a large amount of visual information that requires longer reaction
time to find the information that the passenger needs. For the node 2 in this case study, the corresponding saliency maps were able to visualize the visual information overload, and the impact was reflected in the corresponding LP and SP.

- In the visual saliency analyses, the effect of signage and landmarks on wayfinding is equivocal. If located in a complex and cluttered image region, then the signage or landmark will not have high visual saliency values (Figure 19(a), upper part). When situated in a monotonous image region, then the signage or landmark will be considered as visually salient (Figure 19(b), lower part).

- Advertisement posters and shop boards are causing disturbance in detecting the necessary visual guidance elements (such as Figure 16(b and d); Figure 19(a)).

- An overall evaluation of the visual guidance quality based on visual guidance elements can be provided for the Beijing South Railway station ($EWI= 9.08$).

By taking only the LP and SP into account, the calculation of the evaluation result is rapid, convenient and efficient. This advantage will be more...
apparent when evaluating large-scale passenger terminals where the number of nodes and links are also larger.

6. Discussion and conclusion

This study bridges two very distinct disciplinary worlds, that of visual cognitive psychology and computer vision on the one hand and that of architectural design on the other hand. The former is concerned with understanding the human mind in a theoretical and quantified manner; the latter focuses on the design and analysis of physical places using an empirical approach.

This paper proposes a method that integrates saliency model to quantitively analyze the visual guidance quality for passengers inside a passenger terminal. The main novelty of the method is the use of a visual cognitive analysis approach to understand how architectural spaces, directional signage, and interior landmarks influence passenger path choice. For each origin and destination node combination, the path that passengers are most likely to choose (LP) is compared with the shortest path (SP). Using this information, an overall estimation of the terminal’s visual guidance quality can be expressed by the overall Extra Walking Index (EWI). The validity of the proposed method is verified in reference to empirical results from wayfinding studies. By using the Beijing South Railway Station as a case study, the evaluation method and its intermediate analyses can reveal weaknesses in the original architectural design, as well as identify problems in the terminal’s current setup. When compared to other widely used analysis methods, such as site surveys, tracking individuals with think aloud protocols, passenger questionnaires or space syntax, the proposed method is characterized by the following qualities:

- Cognitive and quantified approach: By integrating saliency model, it is possible to understand the passenger’s visual perception and attention. There is no need for passengers to tell what is on their minds or to passively observe their behavior. Dynamic weights can be applied to visual guidance elements depending on each specific visual scene, instead of assigning identical and fixed weights to these elements throughout all the scenes.

- Pre-assessment and high efficiency: Surveys and on-site investigations can be very expensive and time consuming. In addition, these investigations can only be conducted after the building is completed. By using architectural renderings as input images for the proposed evaluation method, the terminal’s visual guidance quality can be identified in early stages quickly and conveniently, and design revisions can be made when necessary.

- Identifying problems in a terminal’s current setup: The terminal’s management team can use this method to identify problems that exist even in the best-case scenario (without disturbances from passenger flow and crowds). For instance, weaken the visual saliency of existing shop boards or advertisement posters can reduce visual noise and thus improve passenger wayfinding performance.

The implication of the case study is that visual guidance quality is a combined effect of all visual

![Figure 19. Visual saliency analyses for connection node 23 to 25 in the right aisle (a), and for node 9 to 7 (b) in the left aisle. Note that in a, the highest saliency value is applied to a shop board instead of useful visual guidance elements.](image)
guidance elements plus visual noise. While one might expect that visual guidance elements only cooperate with each other in guiding a passenger to his/her destination, the evaluation analyses reveal that even visual guidance elements can compete against each other for passenger visual attention, for instance, on important junction nodes.

At this stage of the study, the passenger crowd’s impact was excluded in the evaluation because even though the crowd may guide passengers in wayfinding, but they may also randomly interrupt the perception of visual guidance elements. In the future study, it is necessary to develop a method that can appropriately model the complex “guidance” and “interruption” effect. In the future development, it is also considered to integrate vertical movement via stairs/escalators/lifts into the visual guidance quality evaluation. Stairs/escalators/lifts could be treated as transition nodes between the current network $G_i(V, E_i)$ and the network $G_j(V_j, E_j)$ of another floor.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by the China Postdoctoral Science Foundation [2018M640057].

**Notes on contributors**

Ran Xu is a lecturer in the School of Architecture and Design, Beijing Jiaotong University, China. She received her B.Sc. in Architecture design from Beijing Jiaotong University in 2008 and then graduated from Technical University of Munich, Germany in 2011 with a Dipl.-Ing. Degree in Architecture Design. In 2016, she received her D.Sc. degree from Ecole polytechnique fédérale de Lausanne (EPFL), Switzerland. From 2012-2017, she worked as a research assistant in the Competence Centre of Envelopes and Solar Energy, Lucerne University of Applied Sciences and Arts (HSLU). Her research interests include the application of saliency models on architecture design, visual impact assessment as well as the design of Building Integrated Photovoltaics.

Xing Su is a lecturer in the Faculty of Informatics, Beijing University of Technology, Beijing, China. He received his B.Sc in the school of software engineering from Beijing University of Technology in 2007. He received his M.Sc and PhD in computer science from University of Wollongong, Australia in 2012 and 2015. His research interests include distributed artificial intelligence, multi-agent systems, disaster management and service-oriented computing.

Haishan Xia is a professor in the School of Architecture and Design, Beijing Jiaotong University, China. From 2004-2009, he was the dean of Architecture faculty in China University of Mining and Technology. He was the dean of School of Architecture and Design, Beijing Jiaotong University from 2011-2019. His research interests include sustainable architecture design, railway-based urban planning and architecture design, integrated above-ground and underground design of urban spaces etc.

**References**

Aguilar, W., M. Luna, J. Moya, V. Abad, H. Ruiz, H. Parra, and C. Angulo. 2017a. “Pedestrian Detection for UAVs Using Cascade Classifiers and Saliency Maps.” Paper presented at the International Work-Conference on Artificial Neural Networks, Cadiz, Spain, June 14–16.

Aguilar, W., M. Luna, J. Moya, V. Abad, H. Ruiz, H. Parra, and W. Lopez. 2017b. “Cascade Classifiers and Saliency Map Based People Detection.” Paper presented at the International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Ugento, Italy, December 12–14.

Arthur, P., and R. Passini. 1992. Wayfinding: People, Signs and Architecture. New York: McGras-Hill.

Bauer, D., V. Seegast, H. Schrom-Feiertag, and A. Millionig. 2018. “Making the Usage of Guidance Systems in Pedestrian Infrastructures Measurable Using the Virtual Environment DAVE.” Transportation Research Part F 59 (2018): 298–317. doi:10.1016/j.trf.2018.09.012.

Becker-Asano, C., F. Ruzzoli, C. Hölscher, and B. Nebel. 2014. “A Multi-agent System Based on Unity 4 for Virtual Perception and Wayfinding.” Transport Research Procedia 2 (2014): 452–455. doi:10.1016/j.tra.2014.09.059.

Borji, A., Z. Bylinskll, F. Durand, L. Itti, T. Judd, A. Oliva, and A. Torralba. “MIT Saliency Benchmark.” Accessed 27 August 2018. saliency.mit.edu/results.html.

Bosch, S., and A. Gharaveis. 2017. “Flying Solo: A Review of the Literature on Wayfinding for Older Adults Experiencing Visual or Cognitive Decline.” Applied Ergonomics 58 (2017): 327–333. doi:10.1016/j.apergo.2016.07.010.

Boynton, G.-M. 2005. “Attention and Visual Perception.” Current Opinion in Neurobiology 15 (4): 465–469. doi:10.1016/j.conb.2005.06.009.

Büchner, S., J. Wiener, and C. Hölscher. 2012. “Methodological Triangulation to Assess Sign Placement.” Paper presented at the ACM ETRA 2012, Eye-tracking Research and Applications, Santa Barbara, USA, March 28–30. doi:10.1145/PDIS-11-11-0999-PDN.

Bylinskii, Z., T. Judd, A. Oliva, A. Torralba, and F. Durand. 2016. “What Do Different Evaluation Metrics Tell Us about Saliency Models?” arXiv:1604.03605.

Canter, D. 1974. Psychology for Architects. New York: Wiley.

Carpman, J., M. Grant, and D. Simmons. 2016. “Hospital Design and Wayfinding: A Video Simulation Study.” Environment and Behavior 17 (3): 296–314. doi:10.1177/0013916585173002.

Chang, X.Y., Y. Ikeda, S. Tsujimura, and K. Sakamoto. 2018. “Examining The Provision Of Railway Transit Information to Foreign Visitors in The Tokyo Metropolitan Area and Strategies for Improvement.” Transportation Research Record 2672 (8): 546–556. doi:10.1177/0361198118790839.

Chen, S.-K., Y. Leng, B.-H. Mao, and S. Liu. 2016. “Integrated Weight-based Multi-criteria Evaluation on Transfer in Examining The Provision Of Railway Transit Information to Foreign Visitors in The Tokyo Metropolitan Area and Strategies for Improvement.” Transportation Research Procedia 333. doi:10.1016/j.trpro.2014.09.059.

Chen, S.-K., Y. Leng, B.-H. Mao, and S. Liu. 2017a. “Towards a Comprehensive Methodology for the Design and Evaluation of Airport Wayfinding Systems.” Journal of Airport Transport Management 14 (3): 151–158. doi:10.1016/j.jairtraman.2008.03.005.
Weisman, J. 1981. “Evaluating Architectural Legibility: Way-finding in the Built Environment.” Environment and Behavior 13 (2): 189–204. doi:10.1177/0013916581132004.

Wener, R., and R. Kaminoff. 1983. “Improving Environmental Information: Effects of Signs on Perceived Crowding and Behavior.” Environment and Behavior 15 (1): 3–20. doi:10.1177/0013916583151001.

Wiener, J., C. Hölscher, S. Büchner, and L. Koniecynz. 2012. “Gaze Behaviour during Space Perception and Spatial Decision Making.” Psychological Research 76 (6): 713–729. doi:10.1007/s00426-011-0397-5.

Wiener, J., C. Hölscher, S. Büchner, and L. Koniecynz. 2009. “How the Geometry of Space Controls Visual Attention during Spatial Decision Making.” Paper presented at the 31st Annual Meeting of the Cognitive Science Society, Amsterdam, Netherlands, July 29-August 1.

Wiener, J., O. DeCondappa, and C. Hölscher. 2011. “Do You Have to Look Where You Go? Gaze Behaviour during Spatial Decision Making.” Paper presented at the 33rd Annual Meeting of the Cognitive Science Society, Boston, USA, July 20–23.

Wilson, R. T., D. W. Baack, and B. D. Till. 2015. “Creativity, Attention and the Memory for Brands: An Outdoor Advertising Field Study.” International Journal of Advertising 34 (2): 232–261. doi:10.1080/02650487.2014.996117.

Wolfe, J., E. M. Palmer, and T. S. Horowitz. 2010. “Reaction Time Distributions Constrain Models of Visual Search.” Vision Research 50 (14): 1304–1311. doi:10.1016/j.visres.2009.11.002.

Wolfe, J., and T. Horowitz. 2004. “What Attributes Guide the Deployment of Visual Attention and How Do They Do It?” Nature Reviews: Neuroscience 5 (6): 495–501. doi:10.1038/nrn1411.

Won, W., M. Lee, and J. Son. 2008. “Implementation of Road Traffic Signs Detection Based on Saliency Map Model.” Paper presented at the IEEE Intelligent Vehicles Symposium, Eindhoven, Netherlands, June 4–6.

Xia, J.-H., P. Zeephongsekul, and D. Packer. 2010. “Spatial and Temporal Modelling of Tourist Movements Using Semi-Markov Processes.” Tourism Management 32 (4): 844–851. doi:10.1016/j.tourman.2010.07.009.

Xu, R. 2016. “Visual Impact Assessment of BIPV in Building Retrofits Using Saliency Models.” PhD diss., Ecole Polytechnique Federale de Lausanne.

Xu, R., and S. Wittkopf. 2014. “Visual Assessment of BIPV Retrofit Design Proposals for Selected Historical Buildings Using the Saliency Map Method.” IOS Press: Journal of Facade Design and Engineering 2014 (2): 235–254.

Xu, R., and S. Wittkopf. 2017. “Quantitative Evaluation of BIPV Visual Impact in Building Retrofits Using the Saliency Method.” Energies 10 (5): 668. doi:10.3390/en10050668.

Yu, F., P. Kontschieder, S. Song, M. Jiang, Y. Zhang, C. Q. Zhao, T. Funkhouser, and J. Xiao. 2017. “LSUN 2017 - Large-scale Scene Understanding Challenge.” Accessed 27 August 2018. http://lsun.cs.princeton.edu/2017/

Zahabi, M., C. Pankok, D. Kaber, P. Machado, M. Lau, J. Hummer, and W. Rasdorf. 2017. “On-Road Visual Sign Salience, Driver Attention Allocation, and Target Detection Accuracy.” Transportation Research Record: Journal of the Transportation Research Board 2663 (1): 40–47. doi:10.3141/2663-06.

Zhang, F., B. Du, and L.-P. Zhang. 2014. “Saliency-Guided Unsupervised Feature Learning for Scene Classification.” IEEE Transactions on Geoscience and Remote Sensing 43 (4): 2175–2184.

Zhao, Q., and C. Koch. 2011. “Learning a Saliency Map Using Fixed Locations in Natural Scenes.” Journal of Vision 11 (3): 9. doi:10.1167/11.3.9.