International Journal of
Geo-Information

MDPI

Article

A Visual Attention Model Based on Eye Tracking in 3D Scene Maps

Bincheng Yang and Hongwei Li *

School of Geoscience and Technology, Zhengzhou University, Zhengzhou 450001, China; ybc_2545@163.com
* Correspondence: lhw29691518@zzu.edu.cn; Tel.: +86-136-7371-2015

Abstract: Visual attention plays a crucial role in the map-reading process and is closely related to the map cognitive process. Eye-tracking data contains a wealth of visual information that can be used to identify cognitive behavior during map reading. Nevertheless, few researchers have applied these data to quantifying visual attention. This study proposes a method for quantitatively calculating visual attention based on eye-tracking data for 3D scene maps. First, eye-tracking technology was used to obtain the differences in the participants' gaze behavior when browsing a street view map in the desktop environment, and to establish a quantitative relationship between eye movement indexes and visual saliency. Then, experiments were carried out to determine the quantitative relationship between visual saliency and visual factors, using vector 3D scene maps as stimulus material. Finally, a visual attention model was obtained by fitting the data. It was shown that a combination of three visual factors can represent the visual attention value of a 3D scene map: color, shape, and size, with a goodness of fit (R^2) greater than 0.699. The current research helps to determine and quantify the visual attention allocation during map reading, laying the foundation for automated machine mapping.

Keywords: visual attention; eye tracking; map cognition; visual cognition

Citation: Yang, B.; Li, H. A Visual Attention Model Based on Eye Tracking in 3D Scene Maps. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 664. https://doi.org/10.3390/ijgi10100664

Academic Editors: Stanislav Popelka, Zdeněk Stachoň, Peter Kiefer, Arzu Çöltekin and Wolfgang Kainz

Received: 7 July 2021
Accepted: 28 September 2021
Published: 1 October 2021

1. Introduction

This article is a contribution to the Special Issue "Eye Tracking in Cartography", which is dedicated to the research problem of understanding a person's cognitive state through eye movement analysis and interpreting their cognitive processes when performing visuospatial tasks (such as map reading, route learning, and navigation). To evaluate and optimize map design, geographic and other information is used, including visualization and various cartographic products such as map-like displays, including 3D representations. The article quantifies the visual attention allocation values during the reading of a 3D scene map by analyzing eye-tracking data. The research facilitates understanding the influence of map visual attributes on visual attention allocation when performing a 3D scene map-reading task. It provides a basis for studying the visual cognitive mechanisms of 3D scene maps.

The visual attention mechanism originates from the study of human vision. It is an essential psychological regulation mechanism of the human visual system in visual information processing activities [1]. Visual activities are closely related to almost everything humans do. Humans selectively focus on some pieces of information and ignore others to effectively use limited visual resources to process information. The study of map visual cognition involves the investigation of how humans read maps and obtain geospatial information, starting from human visual characteristics. It is usually based on subjective feelings, focusing on the reader's evaluation of the design of abstract map symbols and the rationality of improving the map's design [2]. As a result of the rapid development of disciplines such as cognitive science, psychology, and computer vision, and the advent of new technologies such as eye tracking, electrocardiograph (ECG), electroencephalograph (EEG), and nuclear magnetic resonance (NMR) [3], the quantitative and practical research of map visual cognition has received significant supported. Thus, visual cognition has attracted increasing attention of researchers and has become a popular research topic with in the field of map cognition [4–7].

The cognitive mechanism of 3D map visual variables is considered to be one of the challenges of map visual cognition [8]. At present, there are no universally agreed principles for visual cognition theory and the design of 3D maps, and conflicting conclusions exist regarding the research of 3D maps [8]. Because map visual variables are the basic elements that constitute a map, the cognitive study of 3D maps may yield a breakthrough related to the cognitive mechanism of 3D map visual variables. In this study, we placed 3D maps into a specific geographical scene. We defined 3D maps with specific scene content as 3D scene maps, such as Google Street View Map, 360 3D Street View Map, and E-city 3D City Map [9]. The object of this study was a 3D scene map.

Visual variables can affect the allocation of human visual attention. Visual variables affect the distribution of human visual attention, and generally include five aspects: shape, size, orientation, color, and reticle [10]. Existing studies have shown that size, color, orientation, and shape have apparent effects on the efficiency of visual attention [11–13]. Furthermore, texture is likely to affect the efficiency of visual attention, whereas gloss has no pronounced effect on the efficiency of visual attention. The human visual system is the main organ for obtaining the visual information of objects. This study focused on establishing the relationship between visual variables and human visual attention, which is essential for quantifying visual attention.

The research objectives of this study can be summarized as following.

1. To study the quantification of visual attention in the reading process of a 3D scene map. Analyzing the eye-tracking data revealed the key eye movement indicators in the map-reading process, combined with the visual variable factors of the ground objects that affect visual attention. Finally, we attempted to establish a visual attention model.
2. Ultimately, to lay the foundation for our subsequent study of the cognitive mechanism of the visual variables of the 3D scene map.

In Section 2, we present the related works. In Section 3, we focus on the modeling approach, including the design of the experiments, data collection, data processing methods, and model validation methods. Section 4 presents the results of data processing, model building, and model validation. In Section 5, we discuss the results. Finally, in Section 6, we draw conclusions from these results and provide an outlook for future research.

2. Related Works

This section describes the related works, including visual attention calculations in computing and eye-tracking research in 3D maps. The deficiencies of the current research can be summarized as follows: (a) numerous visual attention models consider only one aspect of the position or object, and lack eye movement data for verification; (b) for a 3D map of the scene, there are relatively few studies on eye tracking to quantify the visual attention distribution of different users.

2.1. Visual Attention Research in Computing

When a scene is observed, particular objects or positions more easily attract the viewer's attention due to gazing at objects in the area of interest for a longer period than for unimportant objects [14]. To understand the cognitive mechanism of visual attention, it is important to quantitatively calculate visual attention. Visual saliency is usually used to express the value of visual attention. Numerous researchers have undertaken efforts to calculate visual attention, which can be divided into two categories.

The first category is location-based attention assessment saliency. The core of this approach is to pay attention to the "spotlight" areas in the scene. Independent features, such as color, size, orientation, tilt rate, and curvature, are essential in visual activity. The combination of these features facilitates directing attention to the search task [15]. Itti et al. modelled visual attention in three feature dimensions: color, intensity, and orientation [16].

Research demonstrates that the three dimensions of color, intensity, and direction are related to eye movement behavior [17,18]. Subsequently, numerous researchers, improving on Itti's model, have proposed new models [19–21]. Eriksen et al. proposed an attentional focus model [22]. Koch et al. proposed a saliency graph model [23]. Several researchers have introduced attention mechanisms into deep learning neural networks to improve image processing in recent years [24–26]. A common feature of these models is that bottom-up attentional stimulus information is processed in parallel.

The second category is object-based attention assessment saliency. Position-based attention ignores the geometric properties of spatial objects, in particular their shape, dimensions, and orientation [27]. Prominent objects are described in terms of human vision, and the corresponding visual saliency is calculated [28]. Object-based attention assumes that the structural features of objects direct attention to them rather than via discontinuities in particular locations in the visual scene [29].

These visual attention models have been proposed to provide a basis for the quantitative computation of visual attention. However, these models tend to focus on image pixel-level computation, lack accurate human eye movement data as a basis, and do not explain the role of actual eye-tracking data in the visual attention process [30,31]. Therefore, ideas from computer vision related to computing visual attention can be used in cartography for computing the visual attention of map viewers.

2.2. Eye-Tracking Research in 3D Maps

Compared with traditional 2D maps, 3D maps can provide more one-dimensional spatial information and a three-dimensional perspective that is more appropriate to humans. The cognition and design of 3D maps is a large research area. This section focuses on the application of eye-tracking methods to the design and perception of 3D map symbols. The most numerous map eye-tracking studies are design usability evaluation studies of maps. However, because of maps' complexity, there is no unified theory to reveal the impact of map elements on spatial perception.

In recent years, cartographic eye movement research has begun to focus on the visual perception of 3D maps, and several researchers have investigated 3D map symbols and visual covariates. Liu et al. investigated the effect of field of view and viewing angle on the processing of 3D map information [32]. Liu et al. explored the guidance and constancy of visualization variables in 3D visualization using eye-tracking techniques [33,34]. Popelka et al. used eye tracking to study the level of abstraction of 3D symbols [34]. Lei et al. found through eye movement experiments that the average individual gaze duration was longer, the gaze points were denser, and the viewing angle was smaller when using 3D maps compared to 2D maps [35]. Popelka et al. used eye tracking to investigate differences in user perceptions of 2D and 3D topographic maps in terms of comprehensibility, applicability, and aesthetics [36]. Popelka et al. used eye tracking to evaluate the 3D visualization model they built [37]. Lee et al. used eye tracking to understand the effect of specific architectural elements on viewers' visual attention [38]. Balzarini et al. evaluated the effectiveness of panoramic map design by studying visual attention through eye-tracking techniques [39]. Banitalebi-Dehkordi et al. validated their proposed study for predicting 3D video saliency using an eye movement dataset [40]. Herman et al. described a new tool for the analysis of eye-tracking data and interactive 3D models [41]. Brazil et al. used eye-tracking technology and Google Street View to study the ability of cyclists to assess potential hazards in complex urban environments [42].

Visual variables are the carriers of information. Different map visual variables create complex and varied map representations. Studies have shown that, regardless of subjective factors, different visual variables guide visual attention differently, with the most attractive visual variables being color, orientation, animation, and size, and the potentially non-leading visual variables being semantic information, name, and 3D volume [43]. Color is critical because it is the first visual impression that attracts attention to a feature [44–46]. Similarly, a study of the effectiveness and efficiency of map visual variables in representing information found that size was the most accurate and fastest visual variable, whereas orientation was the least efficient visual variable [47]. Dong et al. demonstrated the effectiveness and efficiency of traffic flow data by comparing different scales and colors, and playback efficiency [48,49]. Because of the complexity of maps themselves, no unified theory reveals the impact of map elements on spatial perception.

The study of eye tracking in cartography in different directions provides ideas for research. To date, few studies have used eye tracking to quantify different visual attention distributions for 3D scene maps. Thus, the current research is highly relevant.

3. Materials and Methods

This section analyses and summarizes the influencing factors of the visual attention of objects. Specifically, it includes three points: (1) the formulation of the experimental plan, which involves the selection of participants, the production of experimental materials, and the arrangement of experimental tasks; (2) collection of the operation and eye movement data of the participants, and pre-processing of the data; (3) presentation of the methodology for statistical analysis and validation of the modeling.

3.1. Data Collection

3.1.1. Participants

A total of 30 (mean age = 23.76, SD = 0.76; 15 males and 15 females; surveying and mapping, computer science, and water resource-related backgrounds) student volunteers from the authors' university were recruited as participants in this experiment. Before the formal experiment, the basic information of the participants was collected via questionnaire. The participant' unadjusted eyesight or corrected visual acuity reached a normal level of 1.0 or higher. None of the participants suffered from color blindness or color weakness. They were all right-handed and proficient in operating computers. All the participants participated in the eye movement experiment for the first time, and did not receive similar training, or know the content of the experiment in advance.

3.1.2. Apparatus

This experiment used the X-series Tobii Pro X3-120 desktop eye tracker from Tobii Sweden for data acquisition. It used the accompanying ErgoLAB3.0 software to count eye movement data and IBM SPSS Statistics 26.0 for data analysis. The experiment was conducted using binocular tracking, with a sampling rate of 120 HZ, sampling accuracy of 0.24°, sampling accuracy of 0.4°, and a delay of fewer than 11 milliseconds. The screen size for displaying the stimulus material was 23.8 inches (16:9), and the screen resolution was 1920 × 1080. The whole experiment was conducted in a softly lit laboratory environment with no noise interference. Participants sat on a seat opposite to the screen, their eyeline was as high as the center of the screen, and their eyes were about 67 cm from the screen. They had a comfortable posture during the test.

3.1.3. Material

In this study, we first designed Experiment 1 and then conducted Experiment 2 based on the results of Experiment 1. In Experiment 2, three experiments were conducted, Experiment 2.1, Experiment 2.2, and Experiment 2.3, corresponding to the three modes of action of visual attention. The stimulus material of Experiment 1 was taken from the 3D scene maps of Beijing, Shanghai, Guangzhou, Zhengzhou, and ten other Chinese cities in the "Earth Online Street View Map". Available online: https://www.earthol.org/ (accessed on 4 April 2021). A total of 30 street pictures were selected as stimulus materials for Experiment 1, with a resolution of 1920 × 1080, and the content was familiar street scenes (Figure 1, left). The stimulus material for Experiment 2 was downloaded from the 3D scene map of "E City". Available online: www.edushi.cn (accessed on 22 April 2021). There were 25 pictures in total, with a scale of 1:2256 and a resolution of 0.597 m. The map comprehensively and realistically presents the city's geographic features, including complex visual features such as the color, size, and shape of the features (Figure 2, left).

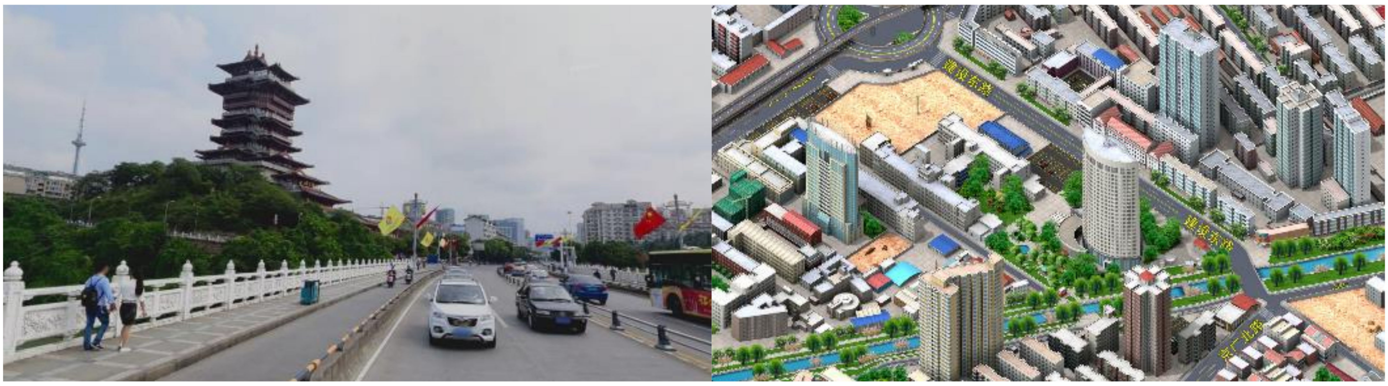


Figure 1. An example picture showing two different types of 3D scene maps (perspective projection in **left**, orthographic projection in **right**).

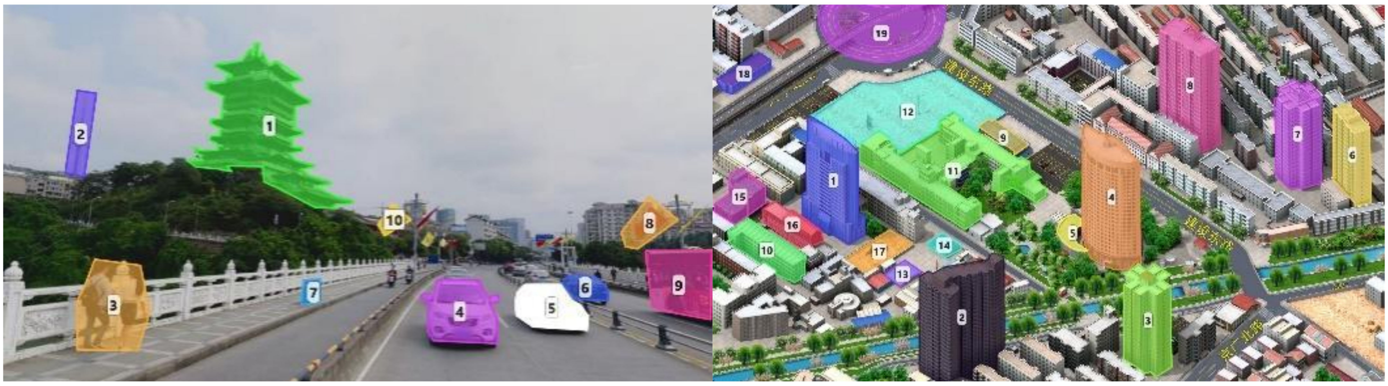


Figure 2. An example of AOI division.

3.1.4. Procedure

In this experiment, a single-factor intra-group experimental design was used. Using the pixel as the smallest unit, the area of pixels where the study target was located was divided into areas of interest (AOI), as shown in Figure 2. After the pre-experiments, the participants were asked to look at the presented 3D scene maps and identify the underground features of interest. Finally, the eye-tracking parameters in the AOI area were analyzed.

Experiment 1

The experimental variables were as follows: (a) the independent variable was the content of the street view picture. Experiment 1 had 30 levels, P1–P30; (b) the dependent variable was the eye movement data automatically recorded by the eye tracker.

The experiment steps were as follows.

1. Welcome the participants and briefly introduce the contents of the experiment, and then use the questionnaire function of Ergo-LAB software to record the participants' gender, age, education, and professional degree;
2. Use the five-point method to calibrate the eyes of the participants until the calibration reached the "Accept" level to ensure the accuracy of eye-tracking;
3. Conducting a pre-experiment. Participants were asked to perform sample pictures to practice and familiarize themselves with the operation process;
4. The formal experiment was started. The red "十" picture in the center was presented for 1 s to return the participant's eyes to the center of the screen, and then the street

view pictures appeared. When participants believed they had finished viewing a picture, they could move to the next. This was done until the subject finished viewing all of the stimulus material. The data was automatically recorded and saved by the eye-tracking device's ErgoLAB 3.0 software throughout the experiment.

5. At the end of the eye movement test, the playback of the experimental recording was observed with the participant. Participants were asked to think aloud, and if the instrument recording matched their actual observation.

Experiment 2

The experimental variables were as follows: (a) the independent variable was the content of the three-dimensional scene maps. Experiment 2 had 25 levels, M1–M25; (b) the dependent variable was the eye movement data automatically recorded by the eye tracker.

The experiment steps were as follows.

1. Steps (1)–(3) were identical to those of Experiment 1;
2. The formal experiment was started. The red "十" picture in the center was presented for 1 s to return the participant's eyes to the center of the screen, and then the three-dimensional scene maps appeared. Three sets of comparative experiments were conducted depending on how the maps were switched from one to the next. In Experiment 2.1, the timing of map switching was determined by the participants. When participants thought they had finished viewing the current map, they switched to the next by themselves until they had finished viewing all of the stimulus material. In Experiment 2.2, the map switching time was fixed at 6 s. When the map was presented on the screen for 6 s, it was automatically switched to the next map until participants had finished viewing all of the stimuli. In Experiment 2.3, the timing of the map switch was arbitrarily switched by the researcher. When the map was presented on the screen, the researcher could randomly switch to the next map as he/she wished, regardless of whether the participant had finished viewing the map or not, until the subject had finished viewing all of the stimulus material. The data was automatically recorded and saved by the eye-tracking device's ErgoLAB 3.0 software throughout the experiment.
3. At the end of the eye movement test, the playback of the experimental recording was observed with the participant, and the participant was asked to think aloud and if the instrument recording matched their actual observation.

Relationship between Experiment 1 and Experiment 2

The two experiments were separate but complementary experiments. The aim of Experiment 1 was to determine the relationship between the eye-tracking index and visual saliency. The aim of Experiment 2 was to determine the relationship between the eye-tracking index and the visual factor. The two experiments used the eye-tracking index as a bridge between the visual factor and the visual saliency. Finally, a visual attention model was developed.

The methods of visual attention perception can be divided into three types. First, the bottom-up approach is data-driven, and is the perception process of automatic salient area selection in natural scenes. Second, the top-down approach is a task-related perception process that is affected by the execution of the task (such as navigation, wayfinding, and sightseeing) and the target's characteristic distribution. The third method comprised the combination of the bottom-up and top-down methods. To better explore the impact of these three methods, three experiments were conducted in Experiment 2, which were recorded as Experiment 2.1, Experiment 2.2, and Experiment 2.3. The resulting models were called bottom-up models, top-down models, and mixed models, respectively.

3.2. Modeling Methods

3.2.1. Basic Ideas

The method of visual attention modeling in computer vision was used as the methodological basis of this study. We used visual saliency to show the value of visual attention, where the participants' visual attention to visual objects can be represented by a linear combination of visual saliency values of various visual factors, as shown in Equation (1):

$$S(y) = \sum_{i=1}^{n} k_i \times O_i(y) + C,\ i = 1, 2, \ldots, n. \tag{1}$$

where $S(y)$ is the visual saliency, $O_i(y)$ is $S(y)$ of the i-th visual factor, k_i represents the weighting values, C is a constant, i is the number of visual factors, and y is the number of objects.

The basic aim of solving Equation (1) can be expressed as follows:

1. To determine the visual factor index system and the calculation method of $O_i(y)$, and to construct the experimental environment based on the visual factors.
2. To collect the eye-tracking data using the eye-tracking technique, and to characterize the visual saliency ($S(x)$) by the combination of eye movement indexes.
3. To construct the experimental environment using the 3D scene map. The calculation of each visual influence factor is combined with the eye movement observation data to fit and solve the percentage k_i and constant C.

3.2.2. Modeling between Eye-Tracking Index and Visual Saliency

The acquisition of visual information is mainly through the eyes, and the process of eye movement reflects the process of visual thinking. Therefore, it is feasible to use eye tracking to obtain eye movement data to analyze the distribution rule of observers' visual attention. There is no uniform regulation relating to the interpretation of eye movement indexes in cartography [50]. The main eye-tracking indexes according to Dong et al. are: number of fixation points, fixation point duration, average fixation point gaze duration, time spent before first entry into AOI, and proportion of fixation points within AOI [8]. Li et al. considered common eye movement indexes as: fixation, sequence of fixation points, fixation point duration, fixation count, fixation frequency, fixation breadth, initial fixation time, and total fixation count [50]. Zheng argued that the main common eye movement indexes are: first entry time, first fixation duration, duration of fixation points, total fixation duration, number of fixation points, and number of fixation times [51]. We considered the eye movement indexes selected from previous studies and the purpose of the experiment in this study.

The eye movement indicators used in this study were Time to first fixation, First fixation duration, Total visit duration, Average visit duration, Visit count, Total fixation duration, Average fixation duration, and Fixation count. The specific meaning of each index is shown in Table 1.

Using a combination of eye movement indicators to show visual saliency,

$$\hat{S}(z) = \sum_{i=1}^{m} \lambda_i \times e_i(z),\ i = 1, 2, \ldots, m. \tag{2}$$

where $\hat{S}(z)$ is the estimated value of $S(x)$, e_i is the eye movement index, λ_i is the weight of influence of each eye movement index, and m is the number of eye movement indexes.

Table 1. Definition of eye movement indexes.

Name	Description
Time to first fixation(s)	The time from the start of the stimulus display until the subject fixates on the AOI for the first time.
First fixation duration(s)	Duration of the first fixation on the AOI.
Total visit duration(s)	Duration of all visits within the AOI.
Average visit duration(s)	Duration of each individual visit within the AOI.
visit count(N)	Number of subjects visit the AOI.
Total fixation duration(s)	Duration of all fixation within the AOI.
Average fixation duration(s)	Duration of each fixation in the AOI.
Fixation count(N)	Number of fixation points in the AOI.

We designed a visual saliency calculation experiment in a virtual experimental environment, in which the participants were allowed to browse freely through street scene photos as stimulus materials, and used an eye tracker to observe and record the participants' eye movement data when observing various objects. Finally, the analytic formula of Equation (2) was obtained by fitting the eye movement data of the participant.

The cognitive process of a visually ordinary observer for a feature is based on sensory input from the surroundings and the current task (sightseeing, walking to a destination). The observer rapidly allocates attention and distinguishes between different spatial objects. For local visual attention, allocation essentially relies on human selective attention mechanisms. This consists of two processes: first, bottom-up pattern creation, i.e., constructing patterns of visual features (such as color, shape, size). The contrast between the object and its environment at the feature level determines how much human visual attention is allocated to it. Second, the top-down attention process. According to the observer's objective, the relevant features are purposefully searched for. The visual attention process is related to the task. However, the pattern creation process reflects the general characteristics of human visual thinking and is universal. Eye movements reflect the visual thinking process. Therefore, it is feasible to use eye-tracking to obtain eye movement data to analyze observers' visual attention distribution. $\hat{S}(z)$ can be used instead of $S(y)$ for the later construction of the visual attention model.

3.2.3. Modeling between Eye-Tracking Index and Visual Factors

Wolf comprehensively summarized the effects of multiple visual attributes on attentional efficiency [11,12]. This study showed that there are significant effects of size, color, orientation, and shape on visual attentional efficiency; texture is likely to affect visual attentional efficiency. In terms of visual features, the features that arouse the observer's interest should have significant and unique visual features or spatial locations compared to the objects, which are mainly reflected in the color, size, shape, direction, texture, and other object features that can attract the observer's sensory attention (visual, olfactory, auditory, etc.). Due to the difficulty associated with visual feature attributes in existing research, the current study focused on the visual aspect. Three visual features, namely, color, size, and shape, were selected to construct the visual factor system. The three visual features are analyzed below.

Color is a visual nerve sensation [52]. Ground features with vivid colors and strong contrasts with their surroundings are more likely to attract attention. The choice of color space is significant for color factor analysis. Usually, RGB space is used to describe color, and can directly describe the physical quantity of the three primary colors. However, visual attention describes the human psychological quantity, and HSV color space is more appropriate. Compared with RGB color space, the three components of HSV color space are independent of each other and more appropriate for human visual characteristics.

Therefore, in this study, three components of HSV were used as the secondary index factors of feature color.

Size is quantified using the minimum bounding rectangle (MBR) of the visible surface of the feature, i.e., height × width.

A shape is a form of existence or expression of a specific underground feature or substance. Experiments have verified [13,53] that the six factors of aspect ratio, rectangularity, area convexity, perimeter convexity, sphericity, and form factor can describe and calculate the shape characteristics of ground objects, and are suitable for similarity comparisons.

In summary, the visual factors of visual saliency of the ground objects are summarized in Table 2.

Table 2. Visual saliency impact factor parameters and their quantification methods.

Name		Description of the Calculation Method	Quantification Method
Color	Hue	The hue/saturation/value is divided into 10 categories with equal spacing, and the quantization value is also divided into equal spacing	$(n-1)\times 36^\circ \leq \Delta H \leq n \times 36^\circ$ $0.1 \times n$
	Saturation		$(n-1)\times 10^\circ \leq \Delta S/\Delta V \leq n \times 10^\circ$ $0.1 \times n$
	Value		
Size	MBR	The sum of the bounding rectangles of the smallest area of all visible surfaces of the ground feature	$\mathrm{MBR} = \sum_{i=1}^{n} length\_MBR_i \times width\_MBR_i$
Shape	Aspect Ratio	The sum of the ratio of the length and width of the minimum bounding rectangle of the smallest area of all visible surfaces of the ground feature	$\mathrm{AspectRatio} = \sum_{i=1}^{n} \frac{length\_MBR_i}{width\_MBR_i}$
	Rectangularity	The sum of the ratio of the area of all visible surfaces of the ground feature to the area of the minimum area bounding rectangle	$\mathrm{Rectangularity} = \sum_{i=1}^{n} \frac{Area\_surface_i}{Area\_MBR_i}$
	Area Convexity	The sum of the ratio of the area of all visible surfaces of the ground feature to its convex hull area	$\mathrm{Area\ Convexity} = \sum_{i=1}^{n} \frac{Area\_surface_i}{Area\_convexity_i}$
	Perimeter Convexity	The sum of the ratio of the perimeter of all visible surfaces of the ground feature to the perimeter of its convex hull	$\mathrm{Perimater\ Convexity} = \sum_{i=1}^{n} \frac{Perimater\_surface_i}{perimeter\_MBR_i}$
	Sphericity	The calculated value of the area of all visible surfaces of the ground feature and its convex hull perimeter	$\mathrm{Sphericity} = \sum_{i=1}^{n} \frac{4\Pi \times Perimater\_surface_i}{(Perimeter\_convexity)^2}$
	Form Factor	Calculated value of all visible surface area and perimeter of features	$\mathrm{Form\ Factor} = \sum_{i=1}^{n} \frac{4\Pi \times Perimater\_surface_i}{(Perimeter\_surface)^2}$

3.2.4. Normalization and Different Degree Calculation

The calculated value of the visual saliency influence factor may have different units. For the convenience of subsequent calculations, the normalization method was used to eliminate the dimension, and the value was mapped to [0, 1]. The normalization formula is as follows:

$$\mathrm{N}_1 = \frac{n - n_{min}}{n_{max} - n_{min}} \tag{3}$$

where n_{min} is the minimum value, n_{max} is the maximum value, n is the attribute value of the visual factor, and N_1 is the normalized value of the attribute value of the visual factor.

From the perspective of visual perception, the object of interest may be quite different in the observer's field of vision from the surrounding features in terms of color, shape, and size. The visual saliency of an object depends not only on its attribute characteristics, but also on the degree of difference between the object and its surroundings. Therefore, when calculating the visual saliency of an object, the degree of difference between it and the surrounding objects should be calculated. Tobler's First Law of Geography states that everything is related to everything else, but near things are more related to each other [54]. The calculation formula for the discrepancy degree of the attribute value is as follows:

$$\mathrm{D(x)} = \left| \mathrm{f(d)} - \sum_{i=1}^{n} \frac{f(d_i)}{\Delta L_i} \right| \quad (4)$$

where $g(x)$ is the degree of difference; i is 1, 2, … , n, which is the neighboring object of the spatial object; $f(xi)$ is the attribute value of the neighboring object; ΔL_i is the distance between the spatial object and the neighboring object.

3.3. Model Coefficient Solving Methods

The model was solved using a multiple stepwise regression approach.

- First, the independent variables were included in the regression model sequentially. The first independent variable entering the regression model is the most closely related to the dependent variable, i.e., the independent variable that shows the greatest correlation with the dependent variable.
- The second independent variable to enter the regression is the one that is most highly correlated with the dependent variable and the first independent variable. In turn, all independent variables are included in the regression model by this rule.
- At each step of the execution, an F-test was used to test the independent variables entering the regression model.

3.4. Model Validation Methods

Model validation experiments were conducted to verify the validity and applicability of the visual attention model developed in this study. To ensure a comprehensive validation, four different types of 3D scenes were selected for the study: building-intensive low-rise area, building-intensive high-rise area, building-sparse low-rise area, and building-sparse high-rise area.

The principle of validation is to compare the value $S_i(x)$ calculated by our model with the true value $\hat{S}_i(z)$ of the eye-tracking data collected using the eye-tracking device. The degree of accuracy (DOF) of our model can be expressed as follows:

$$DOF = \frac{S_i(x)}{\hat{S}_i(z)} \times 100\% \quad (5)$$

where $S_i(x)$ is the model calculated value, $\hat{S}_i(z)$ is the true value, and i is the number of the area of interest.

4. Results

This section focuses on the analysis of the experimental results. Based on the preprocessing of the experimental data, a more detailed treatment was carried out. First, the eye-tracking index related to the visual saliency of the objects was selected by statistical methods, and the mathematical relationship model between visual saliency and the eye-tracking index was constructed experimentally. Secondly, the influence weights of visual factors were obtained using multiple regression methods. The visual attention model under the 3D scene map was constructed, and the model was analyzed and tested. Thirdly, the scientific validity and applicability of the model was assessed.

4.1. Solving between Eye Movement Index and $\hat{S}(z)$

This section corresponds to the results of Experiment 1. The purpose of the experiment was to obtain the quantitative relationship between eye movement indicators and visual saliency (i.e., to obtain the analytical formula of Equation (2)).

4.1.1. Data Analysis

Because the experimental apparatus is significantly affected by the participant's head offset and the intensity of the light, the collected data are missing or noisy. To ensure the reliability of the experimental data, five sets of data with a large number of saccades, continuous discontinuities in the fixation point, and sampling rate lower than 60%, were eliminated. A total of 25 sets of data were retained. Figure 3 shows the heat map and the track map of 25 participants in a street view image. There is no unified understanding of the interpretation of eye movement indicators in cartography [9,51,52]. After comprehensive consideration, eight eye movement indicators, namely, "Time first fixation", "First fixation duration", "Total visit duration", "Average visit duration", "Visits count", "Total fixation duration", "Average fixation duration", and "Fixation count" were selected for model construction.

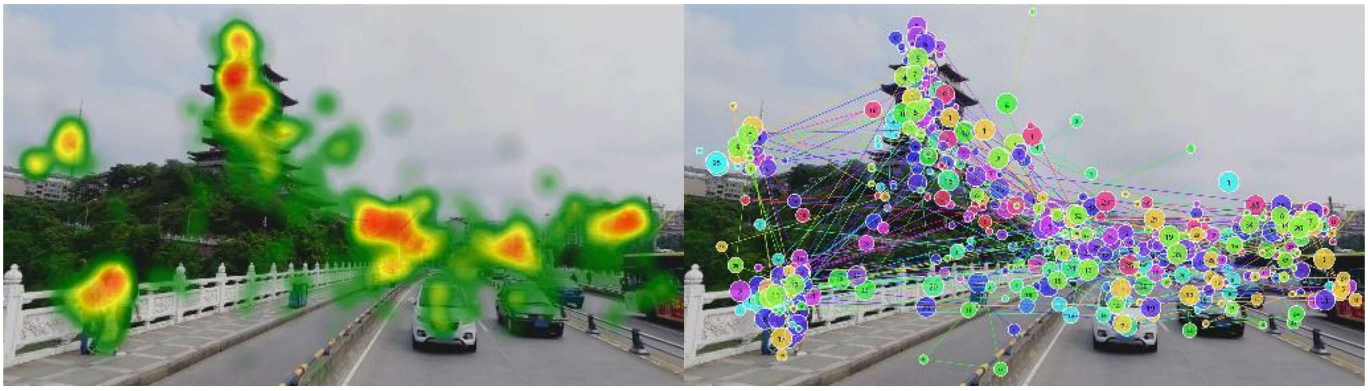


Figure 3. An example picture showing a heat map (**left**) and track map (**right**). The heat map shows the different distribution of the participants' fixation, and the darker the color, the longer the fixation time. The track map shows the movement of the participants' fixation track, and the circle represents the location of the fixation point, and the numbers inside represent the order of fixation.

Visual saliency characterizes the attractiveness of a feature to the observer. During an experiment, the greater the number of people entering an area of interest, the greater the attractiveness. Therefore, the visual saliency value can be replaced by the number of subjects entering an area of interest as a percentage of the total number of subjects. This calculated value was used as the dependent variable, and the eight eye movement indicators are used as independent variables for regression modeling.

4.1.2. Solving the Model

First, we conducted a correlation analysis between independent variables and dependent variables; the results of the analysis are shown in Table 3. Table 3 shows the correlation coefficient between First fixation duration and Visual saliency is 0.293, so the two are linearly weakly correlated. The Average fixation duration and Visual saliency correlation coefficient is 0.363, showing a low linear correlation. The absolute values of the correlation coefficients between the remaining six eye movement factors and visual saliency are between 0.609 and 0.775. Therefore, six eye movement factors, namely, Time to first fixation, Total visit time, Average visit duration, Visits count, Total fixation duration, and Fixations count, were selected for the next analysis.

Table 3. Coefficient of association between visual saliency and eye movement indexes.

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
y	−0.609	$\underline{0.239}$	0.738	0.695	0.775	0.745	$\underline{0.363}$	0.764

e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8: Time to first fixation, First fixation duration, Total visit duration, Average visit duration, Visit count, Total fixation duration, Average fixation duration, Fixation count. y: Visual saliency.

Second, the eye movement factors were used to performed linear regression [55]. To solve the problem of collinearity between independent variables, ridge regression analysis was adopted [56]. Finally, a mathematical model of the relationship between visual saliency and eye movement indexes was established. The model summary (Table 4) and expression (Equation (6)) are as follows:

$$\hat{S}(z) = -0.177e_1 + 0.307e_4 + 0.465e_5 + 0.083 \tag{6}$$

where $\hat{S}(z)$ is the estimated value of $S(x)$, e_1 represents Time to first fixation, e_4 is Average visit duration, and e_5 is the Visits count.

Table 4. Model summary.

R	**R^2**	**Adjusted R^2**	**Errors in Standard Estimation**	**Change Statistics**					
				R^2 Variation	**F Variation**	***DOF* 1**	***DOF* 2**	**Sig F Variation**	**D-W**
0.825	0.681	0.677	0.169	0.019	13.054	1	128	0.000 *	2.086

* $p < 0.05$; R: Coefficient of Determination; R^2: Goodness of Fit; DOF: Degree of freedom; D-W: Durbin–Watson test.

4.1.3. Model Analysis

From the regression Equation (6), it can be seen that: (a) the coefficient of e_1 is negative, indicating that "Time first fixation" is negatively related to "Visual saliency" of AOI; and (b) the coefficients of e_4 and e_5 are positive, indicating that these two variables are positively correlated with "Visual fixation" of AOI.

4.2. Solving between Eye Movement Index and Visual Factors

This section corresponds to Experiment 2, which is based on Equation (6). We simulated the process of human visual attention distribution in the real environment, in the virtual environment of the three-dimensional scene map where the visual factors can be calculated, and then fitted and solved the weighted value k_i for each visual factor and the constant C.

4.2.1. Data Analysis

Thirty volunteers participated in Experiment 2, and experimental data with an effective sampling rate of less than 60% were eliminated. The numbers of valid data bars for the three experiments (2.1, 2.2, and 2.3) were 27, 28, and 25, respectively.

Figure 4 shows the heat map and the track map. It can be seen from Figure 4 that most of the fixation points fall into the delineated AOI, indicating the accuracy of the delineated AOI. The three different browsing methods showed different fixation tracks and fixation points of the objects. This also shows that these three browsing methods lead to different visual attention distributions of the participants.

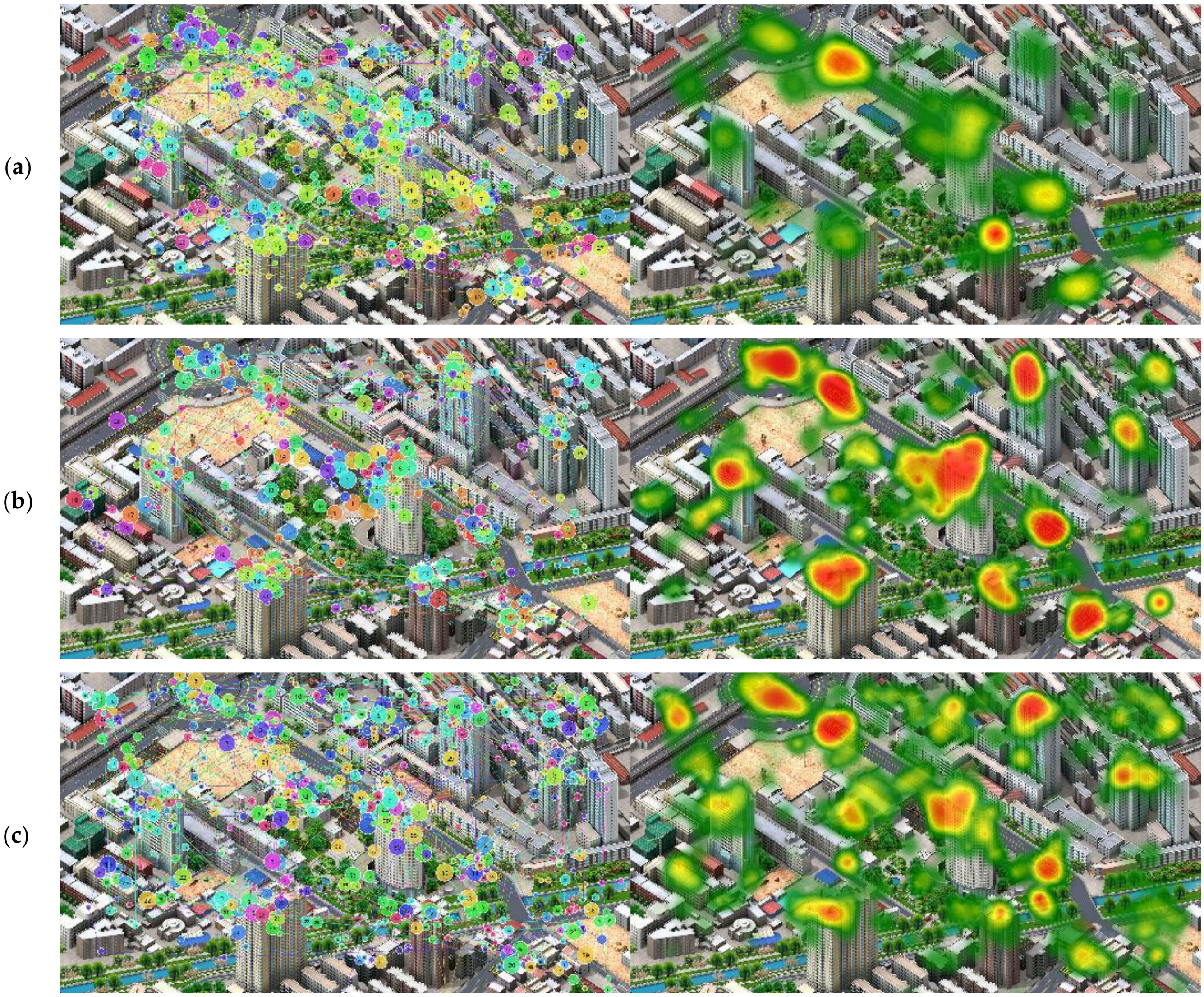
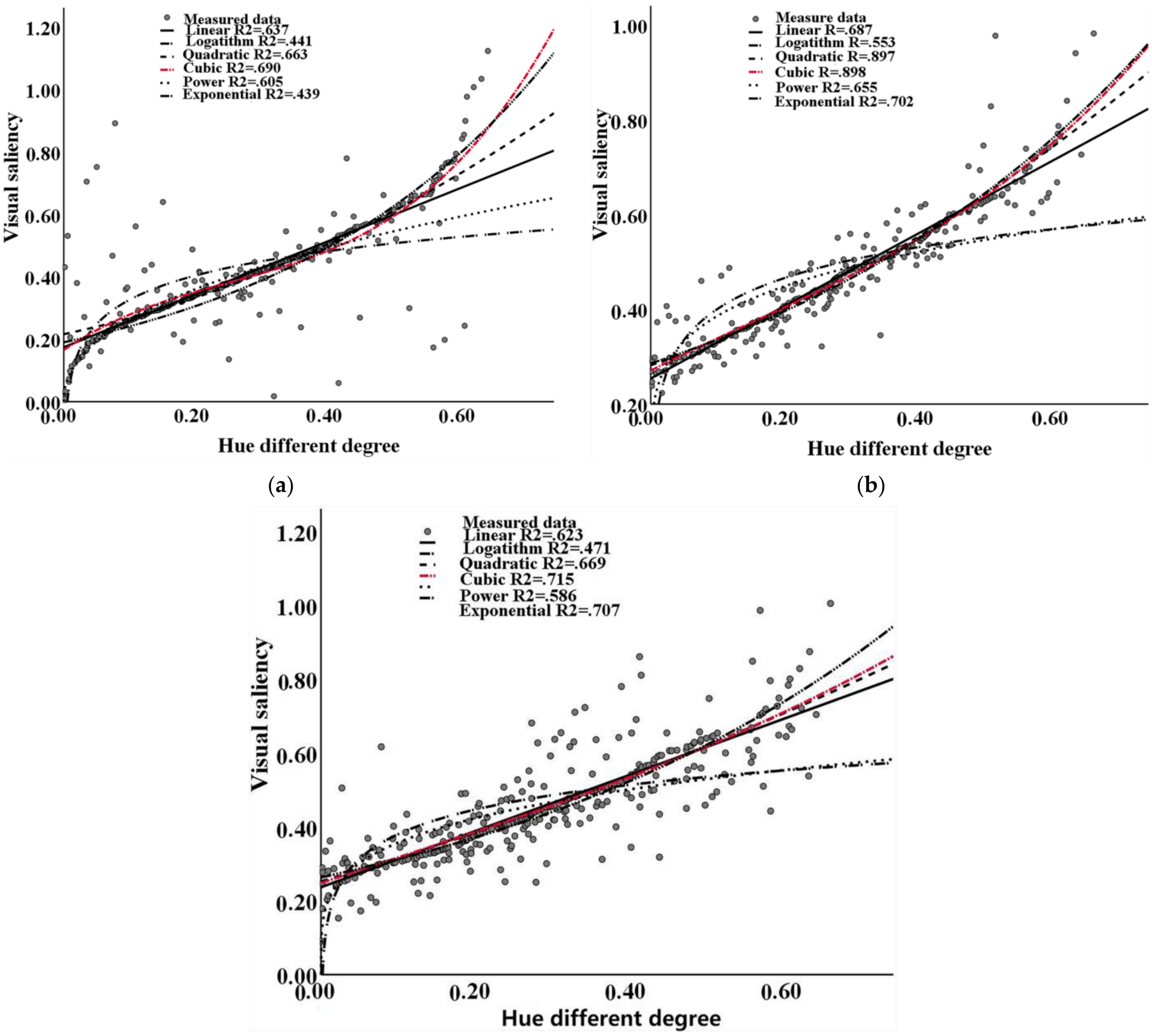


Figure 4. The figure shows the track maps and heat maps shown by "active browsing" (**a**), "timed browsing" (**b**), and "passive browsing" (**c**). The heat map shows the different distributions of the participants' fixation, and the darker the color, the longer the fixation time. The track map shows the movement of the participants' fixation track, the circle represents the location of the fixation point, and the numbers inside represent the order of fixation.

4.2.2. Solving Model

The derived eye movement data of "Time to first fixation", "Average visit duration", and "visit count" were used in Equation (6). The calculated value of "Visual saliency value" was used as the dependent variable, and the visual factor calculated value from the formula in Table 2 was used as the independent variable to solve the mathematical relationship model. The solving steps are as follows.

In the first step, considering that there may be a non-linear relationship between visual factors and visual saliency, each visual factor was spread onto a plane that was judged to roughly conform to its distribution law. Figure 5 shows the curve fitting diagram of the hue of the independent variable and the visual saliency of the dependent variable in Experiment 2.1, Experiment 2.2, and Experiment 2.3. The other independent variables use the same method. The curve law with the goodness of fit statistic R^2 was selected for the subsequent analysis according to the relevant results.



(**a**) (**b**)

(**c**)

Figure 5. As show in the figure, the hue different degree and visual saliency curve fitting graph, where R^2 represents the Pearson Goodness of Fit Statistic. ((**a**) is for Experiment 2.1, (**b**) is for Experiment 2.2, (**c**) is for Experiment 2.3.)

The first step indicated that some independent variables have a nonlinear relationship with the dependent variable. Thus, in the second step, it was considered that there may be an unknown correlation between the independent variables. The direct use of multiple linear regression analysis methods may result in a low goodness of fit statistic for the model. In this study, we used the idea of "nonlinear" to "linear" transformation [57,58]. We rationally introduced parameters and multiple regressions to achieve the conversion of nonlinearity to linearity, thereby eliminating covariance factors. Finally, a multivariate mathematical relationship model between multiple visual factors and visual saliency was constructed.

Introducing new parameters into the regression equation requires that the newly added variables only depend on the original variables and do not contain unknown parameters. This can be undertaken by changing the original independent variables as a function. The first step showed that in Experiments 2.1, 2.2, and 2.3, the cubic function relationship between the color factor and visual saliency is more significant, with correlation coefficients of 0.711, 0.856, and 0.752, respectively. The linear relationship between the size factor and visual saliency is more significant, with correlation coefficients of 0.699, 0.821, and 0.877, respectively. The linear relationship between the shape factor and visual significance was more significant, with correlation coefficients of 0.723, 0.789, and 0.688, respectively.

Therefore, we made combined changes to the independent variables, such as third power, reciprocal, and linear, and 27 new variables are added. The color factor is denoted x_1–x_3, the size factor is denoted x_4, the shape factor is denoted x_5–x_{10}, and the newly added factor is denoted x_{11}–x_{38}. Visual significance was used as the dependent variable, and the original visual factor and the added visual factor were used as independent variables. ANOVA was undertaken to remove the factors with correlation coefficients less than 0.6. The screened visual factors were used as independent variables to conduct multiple regressions, and the factors that led to the problem of multicollinearity between independent variables in the regression process were excluded. After several regressions, the problem of multicollinearity was eliminated and the model was tested. Finally, the visual saliency was obtained by factor reduction.

The summary (Table 5) and Equation (7) of model 2.1 are as follows:

$$\begin{aligned} S_1(x) = {} & -1.090x_1^3 + 0.069x_2^3 + 0.033x_3^3 + 0.493x_4 - 0.276x_5 + 0.492x_6 \\ & -0.256x_7 + 1.786x_9 - 0.301x_{10} - 1.694\frac{x_6}{x_4} + 1.338\frac{x_9}{x_4} \\ & +0.311\frac{x_5+x_6+x_7+x_8+x_9+x_{10}}{x_4} + 0.208 \end{aligned} \tag{7}$$

Table 5. Bottom-up model summary.

R	R^2	Adjusted R^2	Errors in Standard Estimation	Change Statistics					
				R^2 Variation	F Variation	*DOF* 1	*DOF* 2	Significant F Variation	D-W
0.841	0.707	0.699	0.089	0.707	87.352	12	435	0.000 *	2.037

Independent Variable: x_{38}, x_9, x_{21}, x_6, x_{13}, x_{11}, x_{12}, x_{18}, x_7, x_5, x_4, x_{10}. Dependent Variable: Visual saliency. * $p < 0.05$; R: Coefficient of Determination; R^2: Goodness of Fit; *DOF*: Degree of freedom; D-W: Durbin–Watson test.

The summary (Table 6) and Equation (8) of model 2.2 are as follows:

$$\begin{aligned} S_2(x) = {} & -1.305x_1^3 + 0.056x_2^3 + 0.294x_3^3 + 0.155x_4 + 1.048x_5 + 0.285x_6 - 0.122x_7 - 0.313x_8 + 1.294x_9 \\ & -0.214x_{10} - 0.445\frac{x_5+x_6}{x_4} - 0.638\frac{x_6+x_7}{x_4} + 0.826\frac{x_7+x_9}{x_4} + 0.271 \end{aligned} \tag{8}$$

Table 6. Top-down model summary.

R	R^2	Adjusted R^2	Errors in Standard Estimation	Change Statistics					
				R^2 Variation	F Variation	*DOF* 1	*DOF* 2	Significant F Variation	D-W
0.954	0.911	0.907	0.040	0.911	268.598	13	342	0.000 *	2.036

Independent Variable: x_{33}, x_9, x_6, x_{23}, x_{12}, x_{11}, x_{13}, x_4, x_{28}, x_7, x_{10}, x_5, x_8. Dependent Variable: Visual saliency. * $p < 0.05$; R: Coefficient of Determination; R^2: Goodness of Fit; *DOF*: Degree of freedom; D-W: Durbin–Watson test.

The summary (Table 7) and Equation (9) of model 2.3 are as follows:

$$\begin{aligned} S_2(x) = {} & -1.510x_1^3 - 0.224x_2^3 + 0.144x_3^3 - 0.538x_4 + 0.863x_5 - 0.163x_6 + 0.509x_7 + 0.350x_8 - 0.025x_9 \\ & +0.493x_{10} - 0.009\frac{x_6}{x_4} + 0.012\frac{x_7}{x_4} - 0.003\frac{x_9}{x_4} - 0.004\frac{x_6+x_8}{x_4} + 0.261 \end{aligned} \tag{9}$$

Table 7. Mixed model summary.

R	R^2	Adjusted R^2	Errors in Standard Estimation	Change Statistics					D-W
				R^2 Variation	F Variation	*DOF* 1	*DOF* 2	Significant F Variation	
0.855	0.731	0.720	0.077	0.731	64.351	14	331	0.000 *	2.024

Independent Variable: x_{29}, x_9, x_{21}, x_6, x_{11}, x_{12}, x_{18}, x_4, x_{13}, x_5, x_7, x_{19}, x_{10}, x_8. Dependent Variable: Visual saliency. * $p < 0.05$; R: Coefficient of Determination; R^2: Goodness of Fit; *DOF*: Degree of freedom; D-W: Durbin–Watson test.

In Equations (7)–(9), x_1 is the degree of “Hue” difference, x_2 is the degree of “Saturation” difference, x_3 is the degree of “Value” difference, x_4 is the degree of “Minimum area Bounding Rectangle” difference, x_5 is the degree of “Aspect Ratio” difference, x_6 is the degree of “Rectangularity” difference, x_7 is the degree of “Area Convexity” difference, x_8 is the degree of “Perimeter Convexity” difference, x_9 is the degree of “Sphericity” difference, and x_{10} is the degree of “Form Factor” difference.

4.2.3. Model Analysis

We analyzed the model from the form, and the analysis results showed the following:

(a) From the overall form of the regression equation, it can be seen that the visual saliency is affected by the color factor difference degree (x_1, x_2, x_3), size factor difference degree (x_4), and shape factor difference degree (x_5~x_{10}), and that the influence of each factor degree is different.

(b) From the analysis of the various factors of the equation, the influence of each factor on the visual saliency is different. There are linear and non-linear effects, single factor and compound factor effects, and positive and negative correlation effects. The regression coefficients of each influence factor were taken as absolute values, and the effect size of each factor was analyzed. In the bottom-up model, sphericity had a maximum coefficient of 1.786, and value had a minimum coefficient of 0.033 on visual saliency, and the other factors fell between these two values. In the top-down model, hue had a maximum coefficient of 1.305, and saturation had a minimum coefficient of 0.056 on visual saliency, whereas the other factors were between these two values. In the mixed model, the maximum coefficient of hue relative to visual saliency was 1.510, and the minimum sphericity coefficient was 0.025, with other factors in between.

In summary, the influence coefficient of each model parameter is shown in Figure 6.

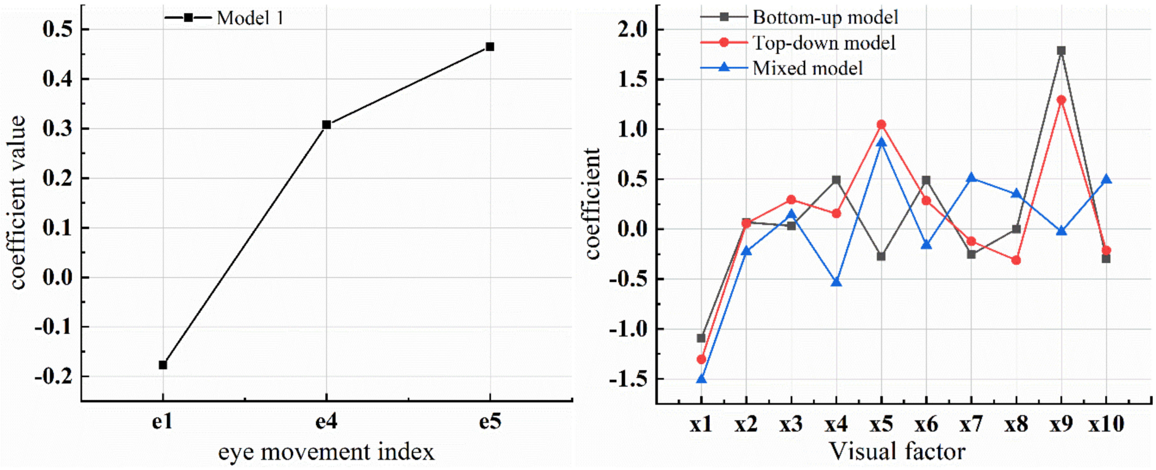


Figure 6. Statistical results of model independent variables’ influence coefficients. *e*1 is “Time to first fixation”, *e*4 is “Average visit duration”, *e*5 is “Visit count” (**Left**). *x*1–*x*10 are different degree of “Hue”, “Saturation”, “Value”, “MBR”, “Aspect Ratio”, “Rectangularity”, “Area Convexity”, “Perimeter Convexity”, “Sphericity”, and “Form Factor”, respectively (**Right**).

4.3. Model Validation Results

We collected eye-tracking data from 40 participants for four types of 3D scenes and counted and calculated them. The results are shown in Figure 7.

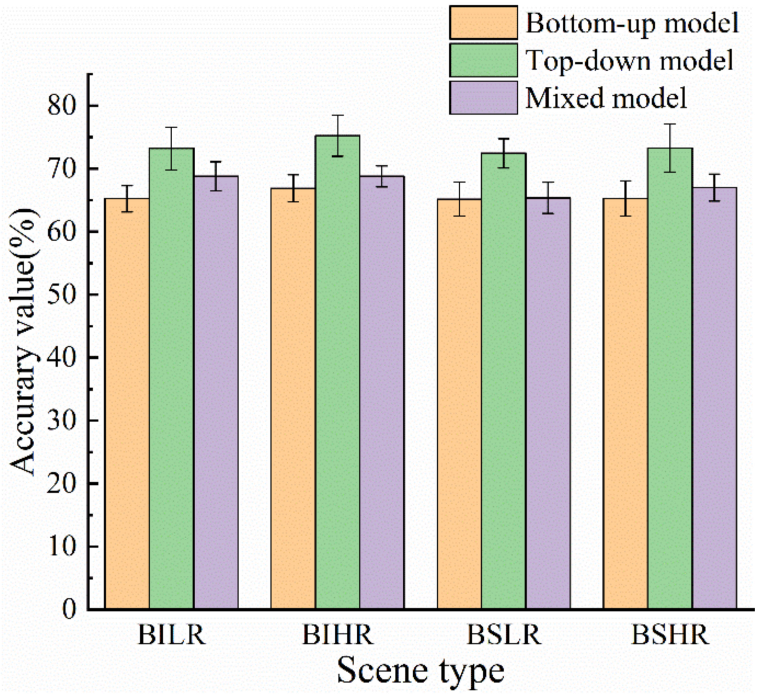


Figure 7. Statistics on the accuracy of model validation in four different scenes. Bars and error bars represent mean values and standard deviations, respectively. (BILR is building intensive low-rise area. BIHR is building intensive high-rise area. BSLR is building sparse low-rise area. BSHR is building sparse high-rise area).

The prediction accuracy of visual attention allocation for different 3D scenes was different for different models. Among the three models, the top-down model had the highest average accuracy (*DOF* = 74.05%, SD = 1.21). The bottom-up and mixed models' average accuracy was 65.65% (SD = 0.73) and 67.49% (SD = 1.43), respectively. Among the four different types of 3D scenes, the highest average accuracy was achieved by BIHR (*DOF* = 70.31%, SD = 3.56). The mean accuracies of BILR, BSLR, and BSHR were 69.08 (SD = 3.26), 67.66% (SD = 3.39), and 68.54% (SD = 3.44), respectively.

5. Discussion

In this section, we analyze the important factors that affect the visual attention model.

5.1. Important Factor for Visual Saliency Model

In model 1 (Equation (6)), the adjusted goodness of fit for the model was 0.677 and the model as a whole passed the F-test (significance level $p < 0.05$), suggesting that visual salience can be expressed to some extent as a linear combination of "Time to first fixation", "Average visit duration", and "Visits count". The three eye movement indexes together explained 67.7% of the visual saliency. The discussion in this section is based on the results of Equation (6).

The correlation coefficient for "Time to first fixation" was −0.177, which was negatively correlated with visual salience. It was shown that the later the participant entered the region for the first time from the appearance of the stimulus material, the weaker the region's attraction. "Time to first fixation" represents the time of initiation of cognitive processes after perception. It is both the time to form the gaze and process the first information, in addition to the initial perception of the target in the area as a whole. Nevertheless, the long "Time to first fixation" may be because of the attraction by the stimulus or a lack of understanding due to a certain difficulty in processing. The case must be analyzed in conjunction with other indicators (Average visit duration, Visits count). Thus, "Time to first fixation" primarily indicates the extent to which the object in the AOI attracts the gaze.

The correlation coefficient for "Average visit duration" was 0.307, and was positively correlated with visual salience. This indicates that the longer the average time that participants spend on each visit to an area of interest, the more attractive that area of interest. The average visit duration is influenced by the difficulty of information processing and the complexity of the material. The number of gaze points reflects the number of times information is processed, with a higher number of gaze points indicating more information processing, but not necessarily that the stimulus material is more difficult to understand. This is related to the individual participant's information processing strategy and should be considered in conjunction with the total gaze time. Therefore, "Average visit duration" is indicative of comprehension.

The correlation coefficient for "Visits count" was 0.465, and was positively correlated with visual salience. This indicates that the more visits a participant makes to an area of interest, the more attractive that area of interest. The second entry into the same area of interest was to review and confirm the target in that area of interest. Between the two gaze behaviours that occurred within the area of interest, participants achieved a transfer to another area on the map outside the area of interest for comparison and validation, and returned to gaze at the target after having a short-term memory of what was outside the area of interest. In this process, memory extraction, comparison, and judgement take place. First, the stimulus signal enters the visual system and is registered. Then, with attention, the relevant information is identified and transferred to short-term memory, which is matched with information extracted from long-term memory. Finally, comparisons and judgements are made on the basis of professional knowledge. "Visits count", therefore, indicates the degree of confirmation.

5.2. Important Factor for Visual Attention Model

The three experiments in Experiment 2 represent the three modes of visual attention: bottom-up, top-down, and bottom-up and top-down combined. Our results suggest that visual attention allocation can be characterized by the degree of variation in visual factors (color, size, and shape). The fit coefficients of the independent variables are different in each model, and they represent different cognitive meanings.

Bottom-up model (Equation (7)) represents the visual attention model with a bottom-up mode of action. The adjusted goodness of fit for the model was 0.699, and the model as a whole passed the F-test (significance level $p < 0.05$). In this model, the coefficient of influence was 0.397 for color variability, 0.493 for size variability, and 0.622 for shape variability. The adjusted goodness of fit for the model was 0.699, and the model as a whole passed the F-test (significance level $p < 0.05$). This suggests that when looking at a scene in a bottom-up mode of action, the greater the difference in the shape from the surrounding features, the less attractive the difference in size, and the least attractive the color difference.

The top-down model (Equation (8)) represents the visual attention model with a top-down mode of action. The adjusted goodness of fit for the model was 0.907, and the model as a whole passed the F-test (significance level $p < 0.05$). In this model, the coefficient of influence of color variability is 0.552, the coefficient of influence of size is 0.155, and the coefficient of influence of shape is 0.546. This suggests that when looking at a scene in a top-down mode of action, features with greater color differences compared to their surroundings attract more visual attention of the observer, with shape differences being the next most attractive, and size differences the least attractive.

The mixed model (Equation (9)) represents a visual attention model with a combination of bottom-up and top-down modes of action. The adjusted goodness of fit for the model was 0.720, and the model as a whole passed the F-test (significance level $p < 0.05$). In this model, the coefficient of influence for color variance was 0.626, that for size variance was 0.538, and that for shape variance was 0.401. This suggests that when looking at a scene in a top-down and bottom-up mode of action, participants first notice features with more significant color variance compared to their surroundings and then notice features with more significant size variance. Shape differences with surrounding features were least attractive.

In summary, the color variability of features tends to work alone, whereas shape variability and size variability work together. The degree of visual factor variability causes differences in the allocation of visual attention, which results in differences in the eye movement index [59]. Because the brain can only process information when the human eye is looking, the first fixation time indicates when cognition begins after the stimulus is connected to sight. The most closely related index is color. It is clear from cognitive psychology research that, of the most critical design elements—color, image, text, and composition—color is particularly important [45,46], because color is the first visual impression of a location that attracts attention [44]. Participants' gaze at features was also influenced by size and shape. Because the determination of information in maps is often similar to the determination of material differences, the eye can quickly focus on important information, because it spends greater time looking at objects in the area of interest than at unimportant objects [14]. The degree of contrast in color, size, and shape can directly impact the prominence of a feature. Physical features, such as the size, color, and spatial location of the stimulus, have an impact on attention. Strong contrasts in these features tend to make the stimulus stand out from the background and guide the individual's selective processing in a bottom-up manner [44].

During the data processing, we found that semantic information (such as signage on the road and business signs), and the textural information of the features themselves appealed, to the participants' visual attention. However, we did not quantitatively evaluate the semantic and textural information. There are two main reasons for this. First, we have not yet found a suitable way to quantify this information. Second, semantic information is challenging to model with regard to other visual factors (size, color, and shape).

6. Conclusions and Future Research

In this article, the correlation between the eye movement index and visual saliency was investigated using an eye movement cognitive experiment based on street images, and a mathematical relationship model based on the relationship between three eye movement indicators (Time to first fixation, Average visit duration, and Visits count). Visual saliency values were established. On this basis, eye movement experiments were conducted via an experimental desktop environment with a 3D scene map in which multiple visual factors could be calculated. The effect of each visual factor was analyzed using regression and statistical methods, and the weight of each factor indicator was obtained. Then, a multi-factor-based visual attention model was experimentally developed. This research contributes to the quantitative study of human visual attention when reading 3D scenes, and exploring the cognitive mechanisms of visual variables in 3D maps, laying the foundation for fully automated machine mapping.

We identified some limitations that can be improved in future studies. First, in this study we only statistically tested and analyzed the constructed model, and only some of the factors were studied. In the future research, more visual information, such as texture and semantic information can be added. Second, the model is currently only applicable to a single laboratory environment with a single set of conditions. To determine if it is universally applicable, experiments should be conducted on eye movement cognition based on map navigation in an outdoor environment with complex conditions. We will also compare and analyze the results with the model we developed to further demonstrate the validity of the results and improve them. Third, the current method of processing the data is relatively homogeneous. There is an urgent need to explore new methods for mining eye movement data to extract greater value, such as via the introduction of artificial neural networks and deep learning, to analyze individuals' cognitive patterns in greater depth. Fourth, experiments comprising a combination of oculomotor, electrocardiograph, electrodermal, and facial expression and behavioral analysis can be undertaken to comprehensively explore individuals' behavioral characteristics. Fifth, our ultimate goal is to write visual attention models as computer programs embedded in robots to automate machine mapping.

Author Contributions: Methodology, validation, formal analysis, investigation, data curation, writing—original draft preparation and editing, Bincheng Yang; writing—review, project administration, and funding acquisition, Hongwei Li. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Research Start-up Fund for Distinguished Professors of Zhengzhou University (No.135-32310276)—"Theory and Methodology of Map Spatial Cognition under Human-Machine-Environment Synergy".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated during the current study are not publicly available as they contain information that could compromise privacy and consent of research participants, but they are available from the corresponding author on reasonable request.

Acknowledgments: The authors would like to thank all the reviewers and editors for their helpful comments and suggestions. Many thanks to the volunteers for participating in the experiment. Many thanks to all the staff and students in the Map and Spatial Cognition Research Unit for their help and guidance.

Conflicts of Interest: The authors declare no conflict interest.

References

1. Bull, D.R. *Chapter 2—The Human Visual System*; Elsevier Ltd.: Amsterdam, The Netherlands, 2014.
2. Burian, J.; Popelka, S.; Beitlova, M. Evaluation of the Cartographical Quality of Urban Plans by Eye-Tracking. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 192. [CrossRef]
3. Keskin, M.; Ooms, K.; Dogru, A.O.; De Maeyer, P. Exploring the Cognitive Load of Expert and Novice Map Users Using EEG and Eye Tracking. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 429. [CrossRef]
4. Göbel, F.; Kiefer, P.; Raubal, M. FeaturEyeTrack: Automatic matching of eye tracking data with map features on interactive maps. *Geoinformatica* **2019**, *23*, 663–687. [CrossRef]
5. Popelka, S.; Vondrakova, A.; Hujnakova, P. Eye-tracking Evaluation of Weather Web Maps. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 256. [CrossRef]
6. Liu, B.; Ding, L.; Meng, L. What is the difference between augmented reality and 2D navigation electronic maps in pedestrian wayfinding. *Cartogr. Geogr. Inf. Sci.* **2021**, *48*, 305–319. [CrossRef]
7. Cybulski, P. Effectiveness of Memorizing an Animated Route—Comparing Satellite and Road Map Differences in the Eye-Tracking Study. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 159. [CrossRef]
8. Dong, W.; Liao, H.; Zhan, Z.; Liu, B.; Wang, S.; Yang, T. New research progress of eye tracking-based map cognition in cartography since 2008. *Acta Geogr. Et Cartogr. Sin.* **2019**, *74*, 599–614. [CrossRef]
9. Lv, G.; Yu, Z.; Yuan, L.; Ro, W.; Zhou, L.; Wu, M.; Sheng, Y. Is the Future of Cartography the Scenario Science? *J. Geo-Inf. Sci.* **2018**, *20*, 1–6. [CrossRef]
10. Mao, Z. *New Course in Cartography*, 2nd ed.; Higher Education Press: Beijing, China, 2008.
11. Wolfe, J.M. Guided Search 2.0 A revised model of visual search. *Psychon. Bull. Rev.* **1994**, *1*, 202–238. [CrossRef]
12. Wolfe, J.M. Guided Search 6.0 A revised model of visual search. *Psychon. Bull. Rev.* **2021**, *28*, 1060–1092. [CrossRef]
13. Jia, F.; Tian, J.; Zhi, M. A visual salience model of landmark based on virtual geographical experiments. *Acta Geogr. Et Cartogr. Sin.* **2018**, *47*, 1114–1122. [CrossRef]
14. Chao, X.; Yufen, C.; Yingjie, W.; Xilin, K. Electronic map design based on parametric template technology. *Geomat. Inf. Sci. Wuhan Univ.* **2009**, *34*, 956–960.
15. Treisman, A.M.; Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **1980**, *12*, 97–136. [CrossRef]
16. Itti, L. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans* **1998**, *20*. [CrossRef]
17. Parkhurst, D.; Law, K.; Niebur, E. Modeling the role of salience in the allocation of overt visual attention. *Vis. Res.* **2002**, *42*, 107–123. [CrossRef]
18. Itti, L. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Vis. Res.* **2005**, *12*, 1093–1123. [CrossRef]

19. Itti, L.; Koch, C. Computational modelling of visual attention. *Nat. Rev. Neurosci.* **2001**, *2*. [CrossRef]
20. Walther, D.; Koch, C. Modeling attention to salient proto-objects. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **2006**, *19*, 1395–1407. [CrossRef]
21. Mirian, M.S.; Ahmadabadi, M.N.; Araabi, B.N.; Siegwart, R.R. Learning active fusion of multiple experts' decisions: An attention-based approach. *Neural Comput.* **2011**, *23*, 558–591. [CrossRef]
22. Eriksen, C.W.; James, J.D.S. Visual attention within and around the field of focal attention: A zoom lens model. *Percept. Psychophys.* **1986**, *40*, 225–240. [CrossRef]
23. Koch, C.; Ullman, S. Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry. *Hum. Neurobiol.* **1987**, *4*, 219–227. [CrossRef]
24. Li, N.; Zhao, X.; Ma, B.; Zou, X. A Visual Attention Model Based on Human Visual Cognition. In Proceedings of the Advances in Brain Inspired Cognitive Systems, Xi'an, China, 7–8 July 2018; pp. 271–281.
25. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
26. Hu, X.; Yang, K.; Fei, L.; Wang, K. ACNet: Attention Based Network to Exploit Complementary Features for RGBD Semantic Segmentation. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 1440–1444.
27. Scholl, B. Objects and attention: The state of the art. *Cognition* **2001**, *80*, 1–46. [CrossRef]
28. Einhäuser, W.; Spain, M.; Perona, P. Objects predict fixations better than early saliency. *J. Vis.* **2008**, *8*, 1–26. [CrossRef]
29. John, D. Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* **1984**, *113*, 501–517. [CrossRef]
30. Borji, A.; Ahmadabadi, M.N.; Araabi, B.N.; Hamidi, M. Online learning of task-driven object-based visual attention control. *Image Vis. Comput.* **2010**, *28*, 1130–1145. [CrossRef]
31. Yaoru, S.; Rober, F. Object-based visual attention for computer vision. *Artif. Intell.* **2003**, *146*, 77–123. [CrossRef]
32. Liu, B.; Dong, W.; Wang, Y.; Zhang, N. The Influence of FOV and Viewing Angle on the Visual Information Processing of 3D Maps. *J. Geo-Inf. Sci.* **2015**, *17*, 1490–1496. [CrossRef]
33. Liu, B.; Dong, W.; Meng, L. Using Eye Tracking to Explore the Guidance and Constancy of Visual Variables in 3D Visualization. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 274. [CrossRef]
34. Popelka, S.; Doležalová, J. Non-photorealistic 3D Visualization in City Maps: An Eye-Tracking Study. In *Modern Trends in Cartography: Selected Papers of CARTOCON 2014*; Springer International Publishing: Cham, Switzland, 2015; pp. 357–367.
35. Lei, T.C.; Wu, S.C.; Chao, C.W.; Lee, S.H. Evaluating differences in spatial visual attention in wayfinding strategy when using 2D and 3D electronic maps. *Geojournal* **2016**, *81*, 153–167. [CrossRef]
36. Popelka, S.; Brychtova, A. Eye-tracking Study on Different Perception of 2D and 3D Terrain Visualisation. *Cartogr. J.* **2013**, *50*, 240–246. [CrossRef]
37. Popelka, S.; Dedkova, P. Extinct Village 3D visualization and its Evaluation with Eye-Movement Recording. In Proceedings of the ICCSA 2014, Guimaraes, Portugal, 30 June–3 July 2014; pp. 786–795.
38. Lee, S.; Cinn, E.; Yan, J.; Jung, J. Using an eye teacking to study three-dismensional environmental aesthetics: The impact of architectural elements and educational training on viewer's visual attention. *J. Archit. Plan. Res.* **2015**, *32*, 145–167.
39. Balzarini, R.; Murat, M. The effectiveness of panoramic maps design: A preliminary study based on mobile eye-tracking. *Remote Sens. Spat. Inf. Sci.* **2016**, 361–368. [CrossRef]
40. Banitalebi-Dehkordi, A.; Nasiopoulos, E.; Pourazad, M.; Nasiopoulos, P. Benchmark three-dimensional eye-tracking dataset for visual saliency prediction on stereoscopic three-dimensional video. *J. Electron. Imaging* **2016**, *25*, 013008. [CrossRef]
41. Herman, L.; Popelka, S.; Hejlova, V. Eye-tracking Analysis of Interactive 3D Geovisualization. *J. Eye Mov. Res.* **2017**, *10*. [CrossRef]
42. Brazil, W.; O'Dowd, A.; Caulfield, B. Using eye-tracking technology and Google street view to understand cyclists' perceptions. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems, Yokohama, Japan, 16–19 October 2017; pp. 1–6.
43. Wolfe, J.M.; Horowitz, T.S. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neuroence* **2004**, *5*, 495–501. [CrossRef]
44. Anderson, J.; Yulin, Q. *Cognitive Psychology and Its Enlightenment*; People's Posts and Telecommunications Press: Beijing, China, 2012.
45. Tian, Y.; Mao, B.; Wang, F. *Design Psychology*; Electronic Industry Press: Beijing, China, 2013.
46. Wu, X.; Zhou, F. *Design Cognition: Design Psychology and User Research*; Southeast University Press: Nanjing, China, 2013.
47. Garlandini, S.; Fabrikant, S.I. Evaluating the effectiveness and efficiency of visual variables for geographic information visualization. In Proceedings of the Spatial Information Theory: 9th International Conference, Aber Wrac'h, France, 21–25 September 2009; pp. 195–211.
48. Dong, W.; Liao, H.; Fang, X.; Liu, Z.; Zhang, S. Using eye tracking to evaluate the usability of animated maps. *Sci. China Earth Sci.* **2014**, *57*, 512–522. [CrossRef]
49. Dong, W.; Ran, J.; Wang, J. Effectiveness and Efficiency of Map Symbols for Dynamic Geographic Information Visualization. *Am. Cartogr.* **2012**, *39*, 98–106. [CrossRef]
50. Li, W.; Chen, Y. Cartography eye movements study and experimental parameter analysis. *Bull. Surv. Mapp.* **2012**, *10*, 16–20.

51. Zheng, S. *Research on Personalized Map Cognition Mechanism*; Chinese People's Liberation Army Information Engineering University: Zhengzhou, China, 2015.
52. Chang, Y.; Chen, X.; Xu, L.; Wang, H.; Wang, S. *Color Composition*; Chongqing University Press: Chongqing, China, 2015.
53. Wang, X.; Huang, D.; Du, J.; Zhang, G. Research on Leaf Image Feature Extraction and Recognition Technology. *Comput. Eng. Appl.* **2006**, *42*, 190–193.
54. Tobler, W.R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* **1970**, *46*, 234–240. [CrossRef]
55. Zhang, H.; Zhang, H. *SPSS Statistical Analysis Practical Collection*; Tsinghua University Press: Beijing, China, 2012.
56. Yang, N. The unique role of ridge regression analysis in solving multicollinearity problems. *Stat. Decis.* **2004**, *3*, 14–15. [CrossRef]
57. Bates, W. *Non-Linear Regression Analysis and Its Application*; China Statistical Press: Beijing, China, 1997.
58. Jiang, T.; Zhang, Q.; Zhou, L.; Jiao, M.; Wang, X. Research on nonlinear regression model based on wavelet method. *Acta Geogr. Et Cartogr. Sin.* **2006**, *35*, 337–341.
59. Negi, S.; Mitra, R. Fixation duration and the learning process: An eye tracking study with subtitled videos. *J. Eye Mov. Res.* **2020**, *13*. [CrossRef]