buildings

Article

Visual Perception of Environmental Elements Analysis in Historical District Based on Eye-Tracking and Semi-Structured Interview: A Case Study in Xining, Taishan

Xing Jiang, Xinxiang Wu, Fangting Chen, Zonghan Chen and Ziang Li *

College of Architecture and Urban Planning, Guangzhou University, Guangzhou 510006, China; jiangxing@gzhu.edu.cn (X.J.); 2112209005@e.gzhu.edu.cn (X.W.); 2112209002@e.gzhu.edu.cn (F.C.); chenzonghan@e.gzhu.edu.cn (Z.C.)
* Correspondence: lza@gzhu.edu.cn

Abstract: The style and overall urban texture of historic districts embody rich social and cultural values. Therefore, how to make relevant environmental elements effectively perceived visually has become the key to protecting and displaying historic streets. Based on this, the non-subjective eye movement data and subjective impression of the subject were collected through an eye-tracking experiment and semi-structured interview. ErgoLAB was used to generate eye-tracking metrics and heat maps based on eye movement data, and ROST-CM6 software was used to generate word frequency and emotional degree data for interview text. Through comparative analysis, it is found that the subjective and objective evaluation indexes of the subjects tend to be consistent in general, but the visual behavior characteristics of different environmental elements' types are different. The greater the variety of elements involved in visual perception, the longer the time required for participants to identify the relevant elements. The extent of element distribution also influenced differences in visual perception. Additionally, visual perceptions from partial elevation views and overall human perspective angles were largely similar, with distinctive elements attracting more interest. This study has an exploratory nature, and its findings contribute to the preservation and enhancement of the visual quality of historic districts.

Keywords: historical and cultural district; eye-tracking; semi-structured interview; environmental element

Academic Editors: Alexandros A. Lavdas and Nikos A. Salingaros

Received: 7 April 2025
Revised: 1 May 2025
Accepted: 3 May 2025
Published: 5 May 2025

Citation: Jiang, X.; Wu, X.; Chen, F.; Chen, Z.; Li, Z. Visual Perception of Environmental Elements Analysis in Historical District Based on Eye-Tracking and Semi-Structured Interview: A Case Study in Xining, Taishan. *Buildings* **2025**, *15*, 1554. https://doi.org/10.3390/buildings15091554

1. Introduction

With the rapid development of the economy and society, the preservation of historical and cultural districts has increasingly become a focus point of public concern. As an indispensable component of urban public life, historical and cultural districts not only provide shelter for residents through their integrated urban fabric but also profoundly define the cultural connotation and historical identity of local communities [1]. Consequently, the contribution of the overall urban texture and architectural heritage of districts is crucial to their social and cultural values, as well as the visual quality of the landscape [2–4]. Historical and cultural districts possess distinctive urban characteristics and historical value, attracting considerable tourist interest [5]. Serving as an "interface" between people and the environment, they establish the foundation for the perception, cognition, and evaluation of the surroundings by both residents and visitors [6,7]. Therefore, by enhancing the environmental elements' visual quality of historical and cultural districts, we can strengthen people's perception and recognition of these areas, effectively revitalizing their vitality and optimizing their conservation and utilization [8].

Vision is the primary pathway through which humans acquire information about the external environment. Eye movements serve as behavioral manifestations of visual information processing, and their metrics can reflect the holistic cognitive process of perceiving external information [9,10]. To understand people's visual perception of historical and cultural districts, eye-tracking technology must be employed. Eye-tracking technology is a technique that detects and records eye movements to track gaze positions. Utilizing specialized equipment such as eye trackers and image-processing algorithms, one can precisely measure gaze direction, fixation locations, and eye position relative to the head. Currently, the most commonly used eye movement metrics include fixation metrics, saccade metrics, and pupillary metrics. Fixation refers to maintaining visual focus on an object for a certain duration. Fixation metrics include first fixation duration, fixation count, and heat map. The longer the gaze time, the more it indicates that the object can attract the subject's attention [11,12]. Saccades represent the rapid movement of gaze from one fixation point to another. Saccade metrics include the saccade count and average saccade count. Pupillary metrics reflect changes in pupil diameter due to shifts in observer interest, including mean pupil diameter, maximum pupil diameter, and minimum pupil diameter [13,14].

In recent years, eye-tracking technology has been widely applied in fields such as medicine and consumer psychology [15,16], revealing individuals' attention focus and information processing by analyzing eye movement records and fixation points and promoting development and innovation in multiple areas. In architecture, the application of eye-tracking technology from a non-professional perspective offers a new approach to studying the built environment. Yin et al. combined traditional restorative scale questionnaires with eye-tracking to identify that green plants, people, and vehicles are important streetscape elements affecting the healing nature of the street and established an exact correspondence between street elements and healing nature [17]. Fu et al. combined questionnaire data with eye-tracking and concluded that the visual elements of the arcade buildings and the street style of a small park in Shantou were correlated with local identity, and the recognition degree of windows was the highest [18]. Lisińska et al. concluded that when people observe, they perceive the observed photo as a complete object, and the perception is based on its shape, light, color, and texture [19]. In their study of the visual behavioral characteristics of rural architectural heritage, Li et al. believed that characteristic elements attracted more visual attention. Visual behavior characteristics are related to area, relative area, distance from the center, and the perimeter [20]. Deng et al. used eye-tracking technology to reveal that for the architectural cultural heritage of southern Fujian, locals pay more attention to the patterns and shapes, while non-locals pay more attention to the characters [21].

This is mainly used in researching the restorative streetscape and tourists' sense of place in streets, as well as evaluating the visual quality of streets and traditional villages, showing a trend of increasingly diverse research directions. By studying the quantified eye movement indicators and combining them with subjective evaluation methods, such as the Semantic Differential (SD) Method, the Analytic Hierarchy Process (AHP), and Scenic Beauty Estimation (SBE), researchers can conduct analyses of interest distribution, attention preferences, and emotional arousal levels in urban streets and in traditional village architecture. Although the research on the protection and development of historical and cultural districts is extensive, studies focusing on historical districts dominated by commercial operations and delving into their unique environmental elements' visual perception characteristics are relatively scarce [22,23].

Research in this field urgently needs to be expanded to more comprehensively address how to shape and preserve the visual identity of historic districts within commercial areas, thereby enhancing their cultural ambiance and heritage continuity. This experiment primarily integrates objective quantitative metrics from eye-tracking technology with subjective semi-structured interviews, mitigating some limitations of traditional survey methods. This approach facilitates a deeper analysis of participants' psychological responses and influencing factors during observation [24]. We aim to ensure that while the commercial value of historic and cultural districts is utilized, their overall value can be better perceived visually and preserved [25].

Therefore, our study mainly investigates the following questions:

Q1: What are the visual cognitive patterns and distribution characteristics when people observe different types of environmental elements in historical and cultural districts?

Q2: Are there discrepancies or correlations between subjective interview data and objective eye-tracking metrics regarding visual cognition across different types of environmental elements in historic and cultural districts?

2. Materials and Methods

This study primarily consists of two parts: an eye-tracking experiment and semi-structured interviews. The eye-tracking experiment captures participants' visual perception data regarding the value characteristics of a historical and cultural district using the Tobii Glasses 3 wearable eye tracker. After completing the eye-tracking experiment, participants underwent semi-structured interviews to obtain their subjective perceptions and evaluations on whether they recognized the district's value characteristics. In this section, the experimental data generation process is detailed, including a description of the material acquisition and research methodology steps taken in order to lay the foundation for data analysis, as shown in the methodological flow chart (Figure 1).

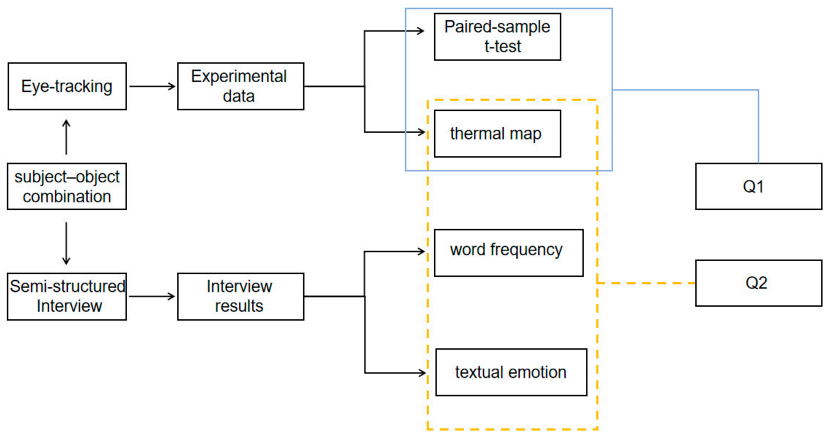


Figure 1. Methodological framework.

2.1. Study Area

This study was conducted in the Xining historical and cultural district in Taishan County, Jiangmen City, Guangdong Province, China (Figure 2). Taishan County is a renowned hometown of overseas Chinese in China. Its central towns and small urban areas have undergone modern municipal improvements in the Republic of China, showing typical regional features with arcade buildings as the main feature. In 1924, Taishan County was granted by Sun Yat-sen to implement local autonomy. The urban texture of the Xining historical and cultural district is highly distinctive. It was planned as a new urban district of the county town, adopting a grid layout to integrate the originally separate Xining Market and Ximen Xu (West Gate Market) outside the old county town. Roads and buildings were uniformly standardized in planning and construction, with all old structures along the streets demolished and replaced by commercial arcade buildings, merging the original "one town, two markets" pattern into a cohesive development. As the essential route connecting the old urban area and the Xining Railway Station, the Xining historical and cultural district experienced rapid urban expansion driven by robust overseas Chinese investment in commerce. Within the district, arcade buildings with diverse Western-style facades stand in great numbers, showing a unique blend of Chinese and Western architectural charm, and it is still the most prosperous part of Taishan County, gathering many important public buildings and residences of prominent people. The district also preserves many buildings with modern architectural heritage, such as the Tianqiao Building, the church, and the Taixi Business Council, while traditional buildings remain well preserved.

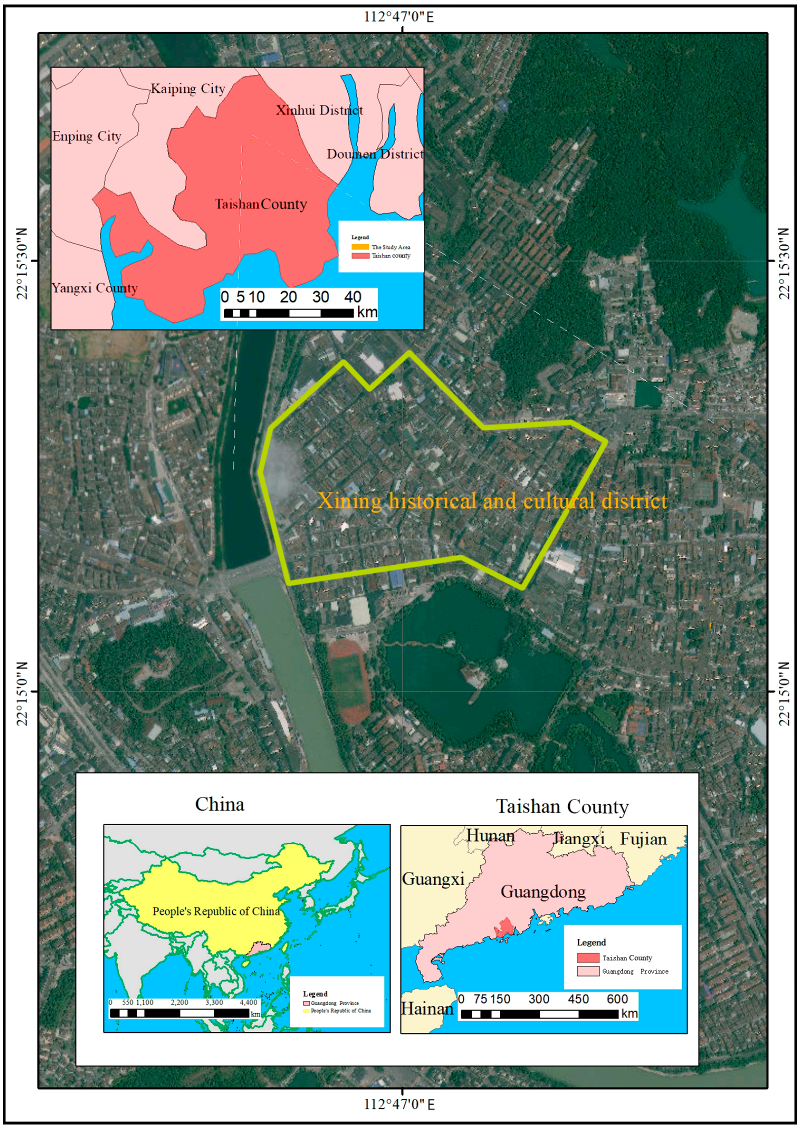


Figure 2. Location of the study area.

Therefore, the Xining historical and cultural district has both a regular grid layout and a road layout formed by the connection between the old and new cities at the urban texture level. Public buildings and arcade buildings demonstrated the prevalence of the combination of Western-style facades and traditional architecture during the Republic of China period, presenting extremely high typicality and representativeness. Their visual quality was unique and could be regarded as a model of the development level of county-level cities in China at that time. The Xining historical and cultural district has always existed as a new commercial development area, which can better realize our hope that the commercial value and overall value of the historical and cultural blocks can be better visually perceived and protected. In view of this, we have chosen it as the study area to explore how its environmental elements' visual perception characteristics influence and shape people's perceptual preferences, providing new perspectives and insights for the conservation and utilization of cultural heritage.

2.2. Experiment Preparation

2.2.1. Photograph Selection

Research indicates that with the advancement of photographic imaging technology, using photos as observational objects can effectively reflect real-world scenes while facilitating experiment execution and modifications [26]. In the eye-tracking experiment, static visual stimuli were selected from photographs of the Xining historical and cultural district captured via drone and smartphone that best represented the district's value characteristics and visual quality. To better observe and identify spatial visual perception, four groups of photos were chosen, depicting the following: urban texture, public buildings, arcade buildings, and partial elevation. Photos of urban texture and partial elevation were taken using a drone. Photos of public buildings and arcade buildings were captured at their front elevations using a smartphone at an approximate eye-level height of 1.60 m to simulate a natural human viewing perspective. From the collected images, 16 of the most representative photos were selected as observational stimuli, ensuring comprehensive coverage of the district's overall environmental elements' visual perception.

2.2.2. Participants

We selected 30 undergraduate and graduate students (15 males and 15 females) from a university in China with different majors, aged 21 to 25, as the participants. All subjects had never been to the Xining historical and cultural district in Taishan County, saw the samples for the first time, and had an uncorrected or corrected visual acuity of 1.0 or higher and normal color acuity. Before the experiment began, the participants were told the overall experiment process in detail, and the importance of collecting eye movement indicators was emphasized.

2.3. Procedure

The experiment was carried out in a bright room to eliminate the interference of outdoor light sources and was carried out in the form of a PowerPoint presentation. After the participant wore an eye tracker and calibrated it, the warm-up pictures were played to eliminate interference from the participants' eye movements. Then, four groups of experimental photos were played according to urban texture, public building, arcade building, and partial elevation. Each photo was played for 10 s with a 4 s interval, and with an 8 s gray image at the end of each set to alleviate visual fatigue; the playback time lasted for nearly four minutes (Figure 3).

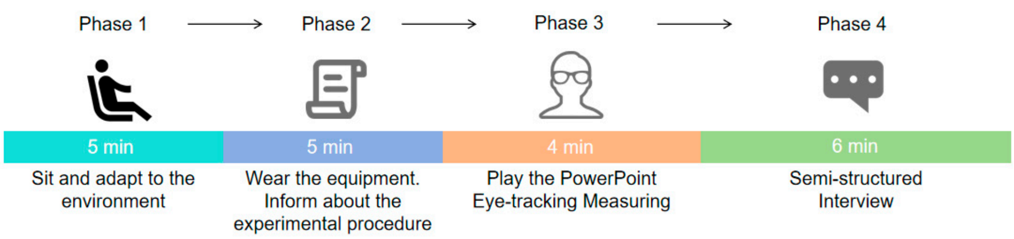


Figure 3. Experimental procedure.

2.4. Data Analysis

The eye movement data were processed by Ergolab software to extract the required indicators. The eye movement data results were imported into SPSS26.0 for statistical analysis. Semi-structured interview data were analyzed using the textual analysis software ROST-CM6.

2.4.1. Eye-Tracking Metrics Selection

We analyzed the characteristics of the visual behavior of those viewing the historical and cultural district from the perspectives of fixation and saccade, including Fixation Count (FC), Average Fixation Time (AFT), Saccade Count (SC), and Average Saccade Count (ASC). Table 1 shows their specific meanings. Heatmaps were also generated to visualize each participant's attention distribution.

Table 1. The meaning of eye movement metrics.

Metrics	Abbreviation	Basic Significance
Fixation Count (N)	FC	The fixation count refers to the number of all fixation points during fixation. A higher number means that the amount of information is too large, and it is difficult to identify all of them in a short time.
Average Fixation Time (s)	AFT	The average fixation time is the average duration of all fixation points during fixation. The longer the average fixation duration, the more difficult the information processing is.
Saccade Count (N)	SC	Saccade count refers to the number of saccades in the staring process. The more the number of saccades, the more information in the area, and the difficulty in identification.
Average Saccade Count (N/s)	ASC	The average saccade count is the average saccade count per second. The higher the number, the more the factors are affected when searching for the next fixation point, and the more information in the region is reflected.

2.4.2. Semi-Structured Interview Data Analysis

ROST-CM6 is a software developed by Wuhan University in China for Chinese text processing [27]. It integrates visualization technologies of text mining and content analysis and can perform a variety of text analysis functions, including text segmentation, word frequency statistics, social network construction, semantic network analysis, and emotional tendency analysis. At present, it has been widely used in China's domestic policy texts, the optimization of tourist attractions, spatial image perception, and other multidisciplinary fields [28–30], and has become an important tool for in-depth text analysis.

The semi-structured interview was processed using this software. The interview data were imported into ROST-CM6 software, word segmentation was carried out, and the feature words with high frequency (top 50%) were extracted to obtain the word frequency analysis results. The higher the frequency of a single word, the more effectively it reflects the consistency of the subjects' cognition of the photo content. Meanwhile, the ranking of word frequencies intuitively reveals which terms are more generalizable and better capture the core features of the photos.

Additionally, emotion analysis was conducted based on the text content to determine the observers' preference level for the photos. The emotion analysis function of ROST-CM6 was used to assign emotion scores to the text data of each subject. A score greater than 5 represented the positive evaluation and positive emotion of the reviewer on the visual perception type of a certain group, a score less than 5 represented the negative evaluation and negative emotion of the reviewer, and 5 was the median, that is, neutral emotion.

3. Results

3.1. Analysis of Eye-Tracking Data

Paired-sample T-tests were applied to examine and compare the mean values of the eye-tracking indicators (Figure 4). It indicates significant differences in the participants' observations when viewing the different scenes in the historical and cultural district, including FC, AFC, SC, and ASC (Tables 2–5).

Table 2. T-test results of fixation count (FC) for different types of carriers.

Metrics	**Correlation Coefficient t-Value/Significance Level *p*-Value**			
Group	Urban texture	Public buildings	Arcade buildings	Partial elevation
Urban texture	1	8.649/0.000 (**)	7.602/0.000 (**)	5.722/0.000 (**)
Public buildings		1	−0.658/0.516	0.671/0.508
Arcade buildings			1	0.940/0.355
Partial elevation				1

Note: ** Correlation is significant at the 0.01 level (two-tailed).

Table 3. T-test results of average fixation time (AFC) for different types of carriers.

Metrics	**Correlation Coefficient t-Value/Significance Level *p*-Value**			
Group	Urban texture	Public buildings	Arcade buildings	Partial elevation
Urban texture	1	−0.203/0.841	0.247/0.807	−0.384/0.704
Public buildings		1	0.423/0.675	0.492/0.626
Arcade buildings			1	0.046/0.964
Partial elevation				1

Table 4. T-test results of saccade count (SC) for different types of carriers.

Metrics	**Correlation Coefficient t-Value/Significance Level *p*-Value**			
Group	Urban texture	Public buildings	Arcade buildings	Partial elevation
Urban texture	1	5.336/0.000 (**)	3.079/0.005 (**)	3.741/0.001 (**)
Public buildings		1	−2.140/0.041 (*)	0.820/0.419
Arcade buildings			1	1.735/0.090
Partial elevation				1

Note: ** Correlation is significant at the 0.01 level (two-tailed); * correlation is significant at the 0.05 level.

Table 5. T-test results of average saccade count (ASC) for different types of carriers.

Metrics	**Correlation Coefficient t-Value/Significance Level *p*-Value**			
Group	Urban texture	Public buildings	Arcade buildings	Partial elevation
Urban texture	1	1.672/0.105	−0.334/0.741	1.231/0.228
Public buildings		1	−2.065/0.048 (*)	0.317/0.753
Arcade buildings			1	1.289/0.208
Partial elevation				1

Note: * correlation is significant at the 0.05 level.

3.2. Analysis of Eye-Tracking Heatmaps

In order to further explore the visual perception of participants' focusing range when observing photos, this study applied heatmaps to conduct a more intuitive analysis, which could more quickly identify which part of which elements could attract more attention (Table 6). Eye-tracking heatmaps can reflect the participants' browsing and fixation in the photo, as well as which areas of attention are more concentrated and the overall distribution of one's gaze.

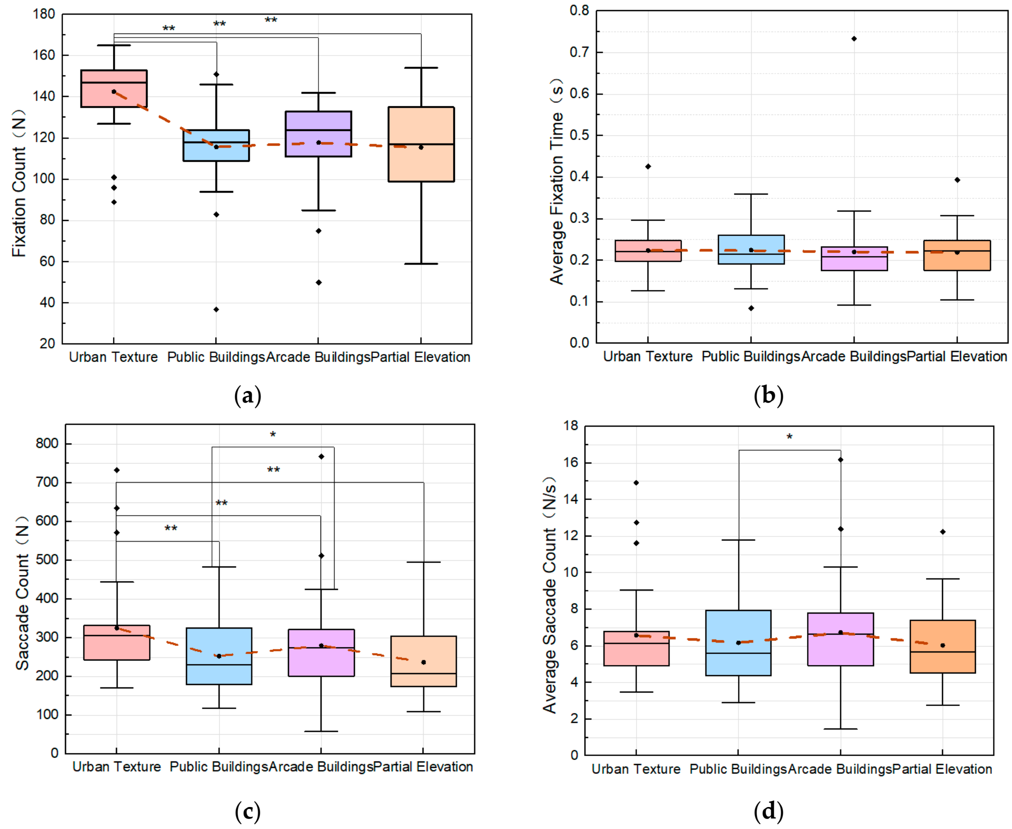
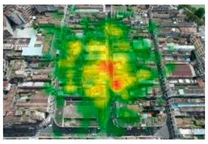
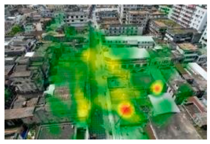
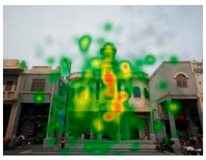
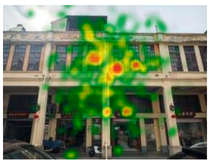


Figure 4. Results of eye-tracking data. (**a**) Inter-group comparative analysis of fixation count. (**b**) Inter-group comparative analysis of average fixation time. (**c**) Inter-group comparative analysis of saccade count. (**d**) Inter-group comparative analysis of average saccade count.

Table 6. Fixation characteristics of historical and cultural district in different groups.

Group	**Heatmaps**			
Urban texture				
Public buildings				
Arcade buildings				
Partial elevation				

3.3. Analysis of Word Frequency

By using the word frequency analysis function in ROST-CM6, semi-structured interview content was imported to obtain high-frequency interview data words about four groups of spatial perception types (Figure 5).

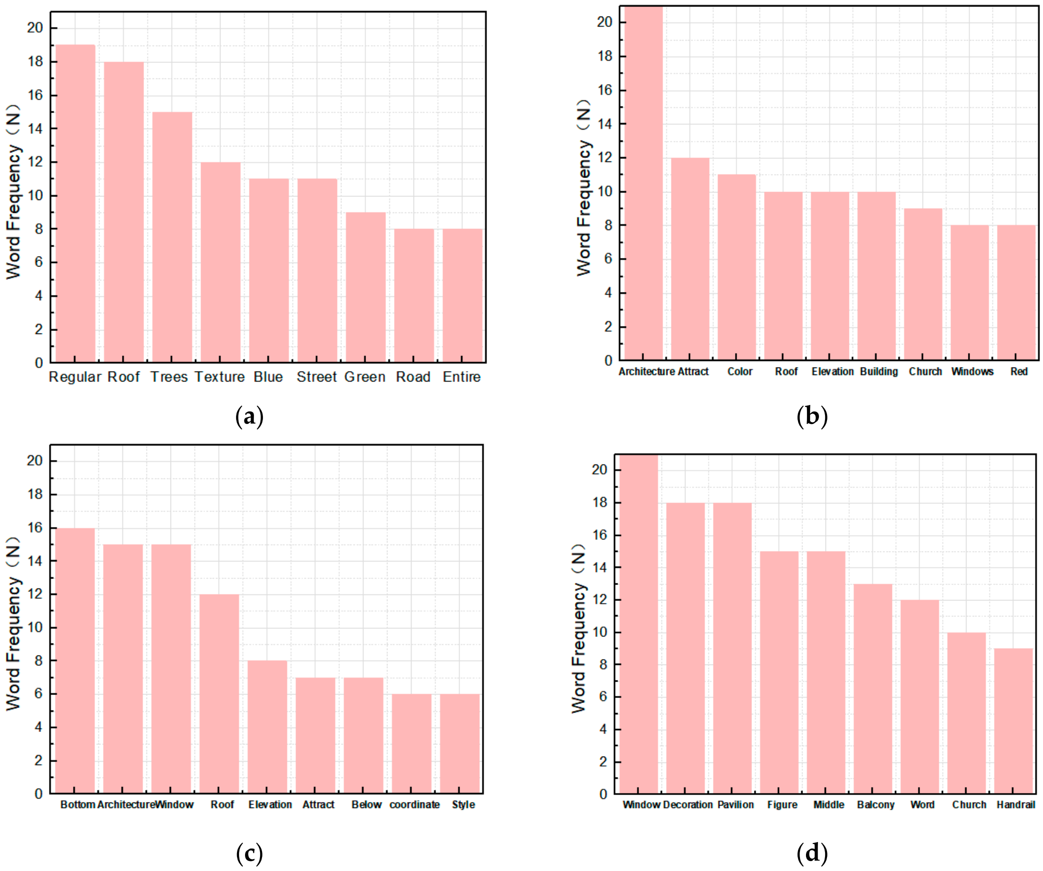


Figure 5. Results of word frequency analysis. (**a**) High-frequency words of urban texture. (**b**) High-frequency words of public buildings. (**c**) High-frequency words of arcade buildings. (**d**) High-frequency words of partial elevation.

3.4. Analysis of Emotion Degree

The statistical results (Tables 7–10) show that in the partial elevation type, the vast majority of participants held positive evaluations, with a ratio of positive to negative emotions at 29:1. This suggests that when only the decorative elements of a partial elevation are displayed, the whole elevation effect is further enhanced, creating a sense of visual comfort. Additionally, in the public building type, the ratio of positive to negative emotions was 4:1. Although general emotional intensity accounted for over half of the positive emotions, moderate and high-intensity negative emotions were absent. Highly positive emotions were primarily related to the windows and roof decorations on the upper part of the public building facades, which were perceived as aesthetically harmonious.

Table 7. Semi-structured interview (urban texture) of emotion analysis data statistics.

Emotional Type	Number of Evaluations	Proportion (%)	Emotional Intensity Level	Positive (N)	Negative (N)
Positive	17	56.67	Low (5, 15], [−15, 5)	11	12
Neutral	0	0	Moderate (15, 25], [−25, −15)	4	1
Negative	13	43.33	High (25, +∞), (−∞, −25)	2	0

Table 8. Semi-structured interview (public buildings) of emotion analysis data statistics.

Emotional Type	Number of Evaluations	Proportion (%)	Emotional Intensity Level	Positive (N)	Negative (N)
Positive	24	80.00	Low (5, 15], [−15, 5)	15	4
Neutral	0	0	Moderate (15, 25], [−25, −15)	7	0
Negative	6	20.00	High (25, +∞), (−∞, −25)	2	2

Table 9. Semi-structured interview (arcade buildings) of emotion analysis data statistics.

Emotional Type	Number of Evaluations	Proportion (%)	Emotional Intensity Level	Positive (N)	Negative (N)
Positive	17	56.67	Low (5, 15], [−15, 5)	6	11
Neutral	0	0	Moderate (15, 25], [−25, −15)	7	1
Negative	13	43.33	High (25, +∞), (-∞, −25)	4	1

Table 10. Semi-structured interview (partial elevation) of emotion analysis data statistics.

Emotional Type	Number of Evaluations	Proportion (%)	Emotional Intensity Level	Positive (N)	Negative (N)
Positive	29	96.77	Low (5, 15], [−15, 5)	7	1
Neutral	0	0	Moderate (15, 25], [−25, −15)	15	0
Negative	1	3.23	High (25, +∞), (−∞, −25)	7	0

4. Discussion

This study conducts an analysis of the visual behavior of observers in the Xining historical and cultural district, providing unique insights into four types of spatial visual perception based on the results of eye-tracking experiments and semi-structured interviews.

The observers exhibited significant differences in FC and SC for urban texture visual perception compared to the other three types. Specifically, FC showed statistically significant differences ($p < 0.01$), with the t-values of the correlation coefficients being 8.649, 7.602, and 5.722, respectively, when compared to the other three types. Similarly, SC also demonstrated significant differences ($p < 0.01$), with t-values of 5.336, 3.079, and 3.741, respectively. These results indicate that observers spent more cognitive time processing urban textures. However, no significant differences were found in AFC or AFT between urban texture and the other three types. Therefore, this suggests that the visual perception of urban textures involves more diverse environmental and spatial elements compared to the other types, offering a greater number of points of interest that capture attention and requiring longer processing time to identify relevant features [31].

In addition, there is a significant difference in SC and ASC between public building type and arcade building type. Among them, the two types of results showed $p < 0.05$ in SC, and the t-value of the correlation coefficient was −2.140. In ASC, the results of the two types showed $p < 0.05$, and the t-value of the correlation coefficient was −2.065. The most exciting elevation of public buildings is mainly concentrated in their visual center formed by symmetry. The arcade buildings have different architectural forms and their own characteristics are widely distributed from the second floor of the arcade to the top, which have gorgeous decoration. There is no obvious visual focus, which causes the increase of SC. Therefore, in terms of SC and ASC, there are significant differences between public building types and arcade building types [32–34].

From the eye-tracking heatmap results, it can be concluded that in terms of urban texture, participants tend to focus significantly on roads and streets, while roofs with more vivid and noticeable color blocks also attract attention. In contrast, homogeneous textures, lacking distinct focal points, draw the gaze toward the visual center. Moreover, the participants paid certain attention to the junctions formed by roads and the places with strong perspectives and vanishing points while ignoring the buildings on both sides of the roads. As for public buildings, the participants focused on the places with concentrated elements and decorations and also looked at the plaques on the buildings to identify the functions of the public building. For arcade buildings, the observations mainly focused on the spaces above the second floor and the overhanging plaques on the ground floor, and the observation of the plaques is the same as that for public buildings, which are used to identify the functional attributes of the ground floor shops of the arcade building. The observation of the second floor and above of the arcade building and the decorations of the balconies and windows above the second floor are the manifestations of architectural value characteristics, with the detailed and ornate carvings and concave and arc-shaped balconies and corridors being the focus of attention [35].

As for the partial elevation, the first image shows that the participants primarily focused on the central vertical elements, including the windows, decoration, and plaque bearing the name “Church.” The second image draws attention to the protruding pavilion at the center. The third image leads observers to examine the plaque, the columns on the balcony, and the railings, while the fourth image concentrates on the central column, the window on the left, and its carved decorative details. The eye-tracking heatmaps of these four images are generally consistent with the results from the previous two groups of human-perspective heatmaps. This further supports the conclusion that these visually perceived elements urgently require preservation, as they are crucial for maintaining the architectural value and characteristics [36].

In the word frequency statistics table of urban texture, the frequencies of “regular”, “roof”, and “trees” are 19, 18, and 15, respectively. It shows that the participants’ perception of urban texture is regular, and the grid layout of roads and streets is straight. However, at the same time, due to the presence of large areas of trees and blue factory roofs that attract people’s attention, there will be deviations in the observation of elements [37–39]. The frequency word statistics table of public buildings shows that the frequencies of “architecture”, “attract”, and “color” are 24, 12, and 11, respectively. The word frequency statistics of public buildings reflect that the participants have a relatively high degree of attraction and evaluation towards the color of the building itself and the roof part. “Red” and “church” have even shown that the color itself has the highest degree of attraction for people. The word frequency statistics table of arcade buildings shows that the frequencies of “bottom”, “architecture”, and “window” are 16, 15, and 15, respectively. The frequency word statistics of arcade buildings express that most subjects have a relatively high attention to the ground floor space, and the frequencies of other parts such as the roof and windows are unified and evaluated as the overall arcade building style being unified and coordinated. The word frequency statistics table of partial elevation reveals that the frequencies of “window”, “decoration”, and “pavilion” are 24, 18, and 18, respectively. This proves that when certain elements are magnified, people will indeed pay more attention to the decorative details.

In the emotional degree analysis, the intensity of emotions of different urban texture types is generally similar. In the arcade buildings group, the moderate and high degrees of positive and negative emotions are relatively high. Positive evaluations mainly involve the decorative style of the art and local details of the arcade buildings themselves, while negative comments mainly focus on the disconnection between the first floor space and the exquisite facade space above the second floor, as well as the unsympathetic handling relationship between the new and old parts, resulting in a distinct boundary demarcation.

5. Conclusions

This study provides valuable insights and references for preserving the characteristic values of historical and cultural districts and enhancing spatial visual perception. This study found a high level of consistency between the objective metrics of the eye-tracking data and the subjective semi-structured interview, indicating that visual cognition across different types of environmental elements in historic and cultural districts consistently elicits positive emotional responses.

While certain limitations exist, this experiment highlights the influence of different types of environmental elements on observers' visual attention. Relevant government departments should consider leveraging these spatial perception patterns to improve visual quality. Specifically, this research offers the following conclusions and optimization strategies:

Firstly, in terms of urban texture, buildings in historical and cultural districts should try to maintain the uniformity of the original style. The roofs should abandon iron roofs with overly striking colors, which are particularly prominent in the eye-tracking heatmap. It is necessary to avoid disrupting the overall harmonious beauty. Secondly, the T-test of eye-tracking data also clearly reveals the unique advantages of urban texture in attracting attention. Compared with other types, its complexity and regularity show more significant differences, which emphasizes the importance of deeply exploring and protecting the urban texture of the district and enhancing the subjective perception of visual quality [40,41]. Furthermore, our focus should go beyond merely enhancing the visual experience of walking and comprehensively cover all aspects of the district planning to ensure the harmonious integration of historical features and modern functions. As Chan said, it is necessary to focus on improving the objective characteristics of the community, as well as taking into account people's perceptions of the community environment and their attitudes towards walking [42].

At the road level, homogenized and regularized road surfaces offer a smooth passage experience but they also tend to draw the observer's attention to the road itself, causing them to overlook the buildings on both sides. Therefore, the optimization strategy of roads should focus on enhancing their overall aesthetic appeal, including the use of textured paving, elaborately designed local embellishments and greenery, creative structures, etc. As for irregular roads, they have the ability to enhance the display effect of building facades. Therefore, special attention should be paid to the meticulous creation at the intersections of streets and alleys, as well as the preservation and presentation of the individuality and artistry of the building facades around them.

Last but not least, at the architectural level, public buildings received the highest positive subjective evaluations [43]. Xing's research also supports this conclusion that a unique visual impression should be shaped for each public space. These emotional responses are an important part of tourists' travel experiences and directly affect their emotional attachment to heritage sites and subsequent behavioral intentions [34]. Meanwhile, as the most important component of the partial elevation, the position and design of signboards are equally crucial. Interestingly, the experimental data and semi-structured interviews show that people will identify the functions of the building to enhance the overall visual experience. For example, an arcade building complex has long facades along the street and each has its own characteristics. In the eye-tracking experiment data, their ASC and SC far exceed those of public buildings, highlighting their unique visual charm.

Consequently, conservation efforts should prioritize the protection of second floors and above for each arcade building. Visual perception analysis, based on eye-tracking heatmaps and semi-structured interviews, indicates that special attention must be given to strict preservation and restoration of windows, decorative elements, and architectural focal points. However, the first floor of arcade buildings, due to disharmony between newer constructions and upper levels, scored poorly in the emotional analysis. In order to improve this situation, we should restructure first floors that have been added or newly built and enhance their distinctive features. This will help tourists identify the information of the shops and ensure that the overall style is harmonious while maintaining modern commercial functions.

Furthermore, this study also has some limitations. The samples in this study were all from university students and postgraduate students on campus and were not grouped according to age or academic level. The research population was not expanded either, and so these two classifications would affect the experimental results [44,45].

Future research can further explore the use of eye-tracking devices in field experiments to eliminate the guiding problem caused by the possible shooting perspectives of photos. At the same time, it can enable participants to be present on the scene or adopt virtual reality, resulting in different experimental outcomes [46]. At the same time, multi-sensory methods can also be employed for experiments, using equipment such as EEG to measure and record the physiological activities of people, thereby enhancing the accuracy of objective data [47]. In this way, we can better reflect the real situation through the experimental results, more effectively improve the quality of spatial visual perception in historical and cultural districts, promote the protection of the districts by the government and the public's perception of historical relics, preserve local characteristics, and respect sustainable development.

Author Contributions: Conceptualization, X.J. and X.W.; methodology X.J. and X.W.; software, F.C. and Z.C.; data curation, F.C. and Z.C.; validation, X.W., F.C. and Z.C.; formal analysis, X.J. and F.C.; investigation, X.W. resources X.J. and Z.L.; writing—original draft preparation; X.W. and F.C.; writing—review and editing, X.J. and Z.L.; visualization X.W. and Z.C.; supervision, X.J.; funding acquisition, Z.L. and X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Philosophical and Science Foundation of Guangdong Province, grant number GD24YYS05; the joint construction project of Guangdong Philosophy Society in 2023, grant number GD23XLN18; the Special Funding Project of the Graduate School of Liwan Research Institute, Guangzhou University, grant number 0309-4100124; and the Special Topic on Basic and Applied Basic Research in Guangzhou, grant number 2024A04J4214.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Guangzhou University (Nos. [2025]093 and 2025-04-23) for studies involving humans.

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

Data Availability Statement: This study did not report any publicly archived datasets.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, J.; Yang, L.; Xiong, Y.; Yang, Y. Effects of Soundscape Perception on Visiting Experience in a Renovated Historical Block. *Build. Environ.* **2019**, *165*, 106375. [CrossRef]
2. Zamani-Farahani, H.; Musa, G. Residents' Attitudes and Perception towards Tourism Development: A Case Study of Masooleh, Iran. *Tour. Manag.* **2008**, *29*, 1233–1236. [CrossRef]
3. Wang, Z.; Shen, M.; Huang, Y. Exploring the Impact of Facade Color Elements on Visual Comfort in Old Residential Buildings in Shanghai: Insights from Eye-Tracking Technology. *Buildings* **2024**, *14*, 1758. [CrossRef]
4. Liu, M.; Nijhuis, S. Mapping Landscape Spaces: Methods for Understanding Spatial-Visual Characteristics in Landscape Design. *Environ. Impact Assess. Rev.* **2020**, *82*, 106376. [CrossRef]
5. Urry, J.; Larsen, J. *The Tourist Gaze 3.0*; SAGE: Newcastle upon Tyne, UK, 2011.
6. Zheng, C.; Jing, Q.; Hyeon, N.K. Research on Tourist Perception of Grand Canal Cultural Heritage Based on Network Text Analysis: The Pingjiang Historical and Cultural District of Suzhou City as an example. *J. Intell. Inf. Syst.* **2023**, *29*, 215–231.

7. Snowden, R.J.; Thompson, P.; Troscianko, T. *Basic Vision: An Introduction to Visual Perception*; Oxford University Press: Oxford, UK, 2012; pp. 206–207.
8. Katahenggam, N. Tourist Perceptions and Preferences of Authenticity in Heritage Tourism: Visual Comparative Study of George Town and Singapore. *J. Tour. Cult. Change* **2019**, *18*, 1–15. [CrossRef]
9. Golkar, K. Conceptual Evolution of Urban Visual Environment; from Cosmetic Approach through to Sustainable Approach. *Environ. Sci.* **2008**, *5*, 95–113.
10. Jin, T.; Lu, J.; Shao, Y. Exploring the Impact of Visual and Aural Elements in Urban Parks on Human Behavior and Emotional Responses. *Land* **2024**, *13*, 1468. [CrossRef]
11. Meso, A.I.; Fletcher, K.; Barton, J.J.S. Review: Eye Tracking: A Comprehensive Guide to Methods and Measures, Vision Rehabilitation: Multidisciplinary Care of the Patient Following Brain Injury. *Perception* **2012**, *41*, 1286–1288. [CrossRef]
12. Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*; Springer: London, UK, 2003.
13. Rayner, K. Eye Movements and Attention in Reading, Scene Perception, and Visual Search. *Q. J. Exp. Psychol.* **2009**, *62*, 1457–1506. [CrossRef]
14. Nikolov, T.Y.; Klein, C.; Ettinger, U. (Eds.) Eye Movement Research: An Introduction to Its Scientific Foundations and Applications. *Perception* **2020**, *49*, 994–995. [CrossRef]
15. Laxton, V.; Maratos, F.A.; Hewson, D.W.; Baird, A.; Stupple, E.J.N. Standardised Colour-Coded Compartmentalised Syringe Trays Improve Anaesthetic Medication Visual Search and Mitigate Cognitive Load. *Br. J. Anaesth.* **2023**, *130*, 343–350. [CrossRef] [PubMed]
16. Grewal, D.; Ahlbom, C.-P.; Beitelspacher, L.; Noble, S.M.; Nordfält, J. In-Store Mobile Phone Use and Customer Shopping Behavior: Evidence from the Field. *J. Mark.* **2018**, *82*, 102–126. [CrossRef]
17. Yin, Y.; Shao, Y.; Xue, Z.; Thwaites, K.; Zhang, K. An Explorative Study on the Identification and Evaluation of Restorative Streetscape Elements. *Landsc. Archit. Front.* **2020**, *8*, 76–90. [CrossRef]
18. Fu, H.; Wang, P.; Zhou, J.; Zhang, S.; Li, Y. Investigating Influence of Visual Elements of Arcade Buildings and Streetscapes on Place Identity Using Eye-Tracking and Semantic Differential Methods. *Buildings* **2023**, *13*, 1580. [CrossRef]
19. Lisińska-Kuśnierz, M.; Krupa, M. Suitability of Eye Tracking in Assessing the Visual Perception of Architecture—a Case Study Concerning Selected Projects Located in Cologne. *Buildings* **2020**, *10*, 20. [CrossRef]
20. Li, N.; Zhang, S.; Xia, L.; Wu, Y. Investigating the Visual Behavior Characteristics of Architectural Heritage Using Eye-Tracking. *Buildings* **2022**, *12*, 1058. [CrossRef]
21. Deng, J.; Chen, J.; Lei, Y. A Study on the Visual Perception of Cultural Value Characteristics of Traditional Southern Fujian Architecture Based on Eye Tracking. *Buildings* **2024**, *14*, 3529. [CrossRef]
22. Birer, E.; Çalışır Adem, P. Role of Public Space Design on the Perception of Historical Environment: A Pilot Study in Amasya. *Front. Archit. Res.* **2021**, *11*, 13–30. [CrossRef]
23. Zheng, S.; Zhang, J.; Zu, R.; Li, Y. Visual Perception Differences and Spatiotemporal Analysis in Commercialized Historic Streets Based on Mobile Eye Tracking: A Case Study in Nanchang Wanshou Palace, China. *Buildings* **2024**, *14*, 1899. [CrossRef]
24. Ma, L.; Xu, Y.; He, S.; Tang, Z.; Li, T.Z. Exploring Cognition Differences in the Traditional Village Landscape Based on Eye-tracking Experiment. *Mod. Urban Res.* **2023**, *2*, 86–94. (In Chinese)
25. Xu, K.; Kong, C.; Li, J.; Zhang, L.; Wu, C. Suitability Evaluation of Urban Construction Land Based on Geo-Environmental Factors of Hangzhou, China. *Comput. Geosci.* **2011**, *37*, 992–1002. [CrossRef]
26. Kalivoda, O.; Vojar, J.; Skřivanová, Z.; Zahradník, D. Consensus in Landscape Preference Judgments: The Effects of Landscape Visual Aesthetic Quality and Respondents' Characteristics. *J. Environ. Manag.* **2014**, *137*, 36–44. [CrossRef]
27. Nowacki, M. Heritage Interpretation and Sustainable Development: A Systematic Literature Review. *Sustainability* **2021**, *13*, 4383. [CrossRef]
28. Fang, L.; Yong, J. ROST data mining system analysis of the text of the accrediting management policy of inclusive private kindergartens in China. *Rev. Educ. Econ.* **2019**, *4*, 108–121. (In Chinese)
29. Zhang, X. Hohhot city image enhancement strategy: Network content analysis based on ROST CM6 software. *Inn. Mong. Sci. Technol. Econ.* **2022**, *16*, 23–25. (In Chinese)
30. Yuan, X.; Liang, H.; Xie, C. Research on public space image perception of Haizhu Square in Guangzhou based on online evaluation text. *Guangdong J. Landsc. Archit.* **2023**, *45*, 59–62. (In Chinese)
31. Dupont, L.; Ooms, K.; Duchowski, A.T.; Antrop, M.; Van Eetvelde, V. Investigating the Visual Exploration of the Rural-Urban Gradient Using Eye-Tracking. *Spat. Cogn. Comput.* **2016**, *17*, 65–88. [CrossRef]
32. Liu, Y.; Li, Z.; Tian, Y.; Gao, B.; Wang, S.; Qi, Y.; Zou, Z.; Li, X.; Wang, R. A Study on Identifying the Spatial Characteristic Factors of Traditional Streets Based on Visitor Perception: Yuanjia Village, Shaanxi Province. *Buildings* **2024**, *14*, 1815. [CrossRef]
33. Sadeghifar, M.; Pazhouhanfar, M.; Farrokhzad, M. An Exploration of the Relationships between Urban Building Façade Visual Elements and People's Preferences in the City of Gorgan, Iran. *Archit. Eng. Des. Manag.* **2018**, *15*, 445–458. [CrossRef]

34. Li, X.; Wang, P.; Li, L.; Liu, J. The Influence of Architectural Heritage and Tourists' Positive Emotions on Behavioral Intentions Using Eye-Tracking Study. *Sci. Rep.* **2025**, *15*, 1447. [CrossRef] [PubMed]
35. Liu, R.; Neisch, P. Measuring the Effectiveness of Street Renewal Design: Insights from Visual Preference Surveys, Deep-Learning Technology, and Eye-Tracking Simulation Software. *Landsc. Urban Plan.* **2025**, *256*, 105291. [CrossRef]
36. Doğan, H.A. Implementation of Eye Tracking Technology on Cultural Heritage Research and Practice. *Igra Ustvarjalnosti-Creativity Game* **2019**, *7*, 16–21. [CrossRef]
37. Amati, M.; Ghanbari Parmehr, E.; McCarthy, C.; Sita, J. How Eye-Catching Are Natural Features When Walking through a Park? Eye-Tracking Responses to Videos of Walks. *Urban For. Urban Green.* **2018**, *31*, 67–78. [CrossRef]
38. Wang, Z.; Shen, M.; Huang, Y. Combining Eye-Tracking Technology and Subjective Evaluation to Determine Building Facade Color Combinations and Visual Quality. *Appl. Sci.* **2024**, *14*, 8227. [CrossRef]
39. Askari, A.H.; Soltani, S. Contribution of Building Façades to Attractive Streetscapes: Study of Two Main Streets in Kuala Lumpur City. *J. Des. Built Environ.* **2018**, *18*, 29–40. [CrossRef]
40. Wang, M.; Zhao, M.; Lin, M.; Cao, W.; Zhu, H.; An, N. Seeking lost memories: Application of a new visual methodology for heritage protection. *Geogr. Rev.* **2020**, *110*, 556–574. [CrossRef]
41. Meng, Y.; Lyu, M.; Sun, D.; Shi, J.; Fukuda, H. Evaluating and Diagnosing Urban Function and Perceived Quality Based on Multi-Source Data and Deep Learning Using Dalian as an Example. *Buildings* **2025**, *15*, 998. [CrossRef]
42. Chan, E.T.H.; Schwanen, T.; Banister, D. The Role of Perceived Environment, Neighbourhood Characteristics, and Attitudes in Walking Behaviour: Evidence from a Rapidly Developing City in China. *Transportation* **2019**, *48*, 431–454. [CrossRef]
43. Xing, Y.; Leng, J. Evaluation of Public Space in Traditional Villages Based on Eye Tracking Technology. *J. Asian Archit. Build. Eng.* **2023**, *23*, 125–139. [CrossRef]
44. Rusnak, M.A. Eye tracker as a pro-social tool of managing urbanistic and architectural heritage. *Prot. Cult. Herit.* **2020**, *9*, 97–115. [CrossRef]
45. Rusnak, M.; Szmigiel, M.; Geniusz, M.; Koszewicz, Z.; Magdziak-Tokłowicz, M. Exploring the Impact of Cultural Context on Eye-Tracking Studies of Architectural Monuments in Selected European Cities: Sustainable Heritage Management. *J. Cult. Herit.* **2024**, *66*, 326–342. [CrossRef]
46. Jeon, J.Y.; Jo, H.I.; Lee, K. Psycho-Physiological Restoration with Audio-Visual Interactions through Virtual Reality Simulations of Soundscape and Landscape Experiences in Urban, Waterfront, and Green Environments. *Sustain. Cities Soc.* **2023**, *99*, 104929. [CrossRef]
47. Chynał, P.; Sobecki, J.; Rymarz, M.; Kilijańska, B. Shopping Behaviour Analysis Using Eyetracking and EEG. In Proceedings of the 2016 9th International Conference on Human System Interactions (HSI), Portsmouth, UK, 6–8 July 2016.