INTEGRATION OF VIRTUAL REALITY, 3-D EYE-TRACKING, AND PROTOCOL ANALYSIS FOR RE-DESIGNING STREET SPACE

LE-MIN ZHANG¹, TAY-SHENG JENG² and RUO-XI ZHANG³
¹Department of Architecture, National Cheng Kung University, Tainan; School of Architecture and Civil Engineering, Xiamen University, Xiamen
²zlmzj@126.com
³Department of Architecture, National Cheng Kung University, Tainan
⁴tsjeng@mail.ncku.edu.tw
⁵School of Architecture and Civil Engineering, Xiamen University
⁶zhangruoxi@xmu.edu.cn

Abstract. The objective of this paper is to develop an eye-tracking technology combined with a virtual reality system for an experimental study of an historical street design. Using protocol analysis, a set of design objects, parameters, and subjects are randomly selected for evaluation of the virtual street space of an ancient city. 3-D point-cloud data of spatial behaviors are tracked and analyzed. It is concluded that people with different cultural backgrounds each have a considerably different perception of the street space’s characteristics. The methodology described in this paper can be used for spatial design of urban space in the future.

Keywords. Virtual Reality; Eye-Tracking; Protocol Analysis; Street Space.

1. INTRODUCTION

Virtual Reality (VR) technology offers new opportunities for the development of interactive and simulation tools for urban space and building design. VR headsets offer a high quality immersive environment and interactive tools for historical site analysis and preservation design. Traditional street spaces can be viewed as “social media” where people walk, sightsee, perceive, and cognize. The important components of historic urban districts include sitting spaces, sunlight, trees, signs, and decorations (Guo and Jeng 2008). The relationship of these components to the street is integral and goes far beyond the traditional urban design. The development of an in-depth understanding of the connotations of street spaces is critical for renovating and restoring these districts.

The objective of this paper is to develop an eye-tracking technology, combined with a VR system, for an experimental study of an historical street design. The paper reports a visual thinking method that includes integrated eye-tracking technology in the VR headset and a protocol analysis. The aim is to query subjects so as to better understand attentiveness as well as head and body movements as
they relate to particular features of the virtual scene. This suite of technology and analytic protocols is applied to evaluate a proposed historic restoration of the Hong Kong Streetspace in the ancient city of Zhangzhou (Fujian, China). It also addresses a second objective, which is to assess the differential perceptions of professionals (architects) and naive user (non-architects) of the renovated street space. Based on the experimental studies of attentiveness in the VR environment, the traditional street space may be redesigned in such a way as to optimize the experience with the cultural nuances of the architectural elements of traditional shop-lined streets.

2. RELATED WORK

2.1. ADVANCES IN VR SYSTEMS

Much synergy is expected to be gained from the combination of VR with eye-tracking technologies, until recently, the limitations of VR impeded this progress (Kiefer, Giannopoulos et al. 2017). Currently, the research in this field is more about graphics acceleration and interactive control. Little work has been done to study eye-tracking with the subject wearing a mobile eye tracker in the VR environment. For example, study on feature of users’ eye movements during a distributed and synchronized VR meeting using cloud computing, this paper attempted to apply the eye-tracking technology for spatial design (Fukuda and Taguchi 2013). Evaluation of indoor guidance systems often use an immersive virtual environment in combination with a mobile eye tracking system. The paper presents a new approach to combining an immersive virtual environment with a mobile eye tracking system for 3D gaze analysis and visualization (Schrom-Feiertag, Settgast et al. 2017). Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research, the combination provides a unique opportunity for shopper research (Meißner, Pfeiffer et al. 2017). These studies have achieved good results, greatly expanded the study of spatial behavior. Although eye tracking in virtual reality has a wealth of data acquisition capabilities, many experiments in the VR environment still require that the subject wear the traditional mobile eye tracker. Hence, the ability of the VR helmet to assist with the determination of spatial location has not been fully exploited.

Eye tracking has an intrinsic relationship with VR, the progress in hardware and data collection has opened up more opportunities for eye-tracking research in the VR environment (Jacob and Karn 2003). It is already the case that the user wears a head-mounted display, adding a small head-mounted eye tracker and illuminator has negligible impact on the equipment size and weight. As for data processing, the eye tracking data can be merged with the user spatial behavior data quickly and accurately. Meanwhile, eye tracking accuracy and data analysis efficiency can be improved considerably by combining the components of the 3-D model in the VR environment with the database and object-oriented spatial positioning technologies.
2.2. AN EXAMPLE

The ancient city of Zhangzhou is about 86 hectares in area and is located at the center of Zhangzhou, Fujian. It has over 1300 years of history, ranging from the Tang, Song, Yuan, Ming, and Qing dynasties to the modern Republic of China. Hong Kong Street is at the core of ancient city preservation. It once received the Habitat Scroll of Honor Award, but the cityscape was heavily modified due to its long history of occupation and urbanization. How to understand the current composition of cityscape along Hong Kong Street and also highlight its original, traditional street space in southern Fujian province are both critical to architectural preservation (Zhang and Zhang 2015). The virtual image of the Hong Kong Street can be constructed more accurately through attention to spatial characteristics and also the special elevations of particular features. The key to researching the design of a street space in the ancient city of Zhangzhou is to design and analyze the human-space-interface elements as an organic whole. That is, the form or element cannot be treated separately because during the formation of a person’s perception of the ancient city, the street space is not only a carrier of the Street and the boundary, but also the interface between boundaries.

3. THE EXPERIMENT

This paper takes a protocol analysis approach to capturing the cognitive thinking and spatial cognition of the people who walk in the VR environment. The study starts with the street scene drawn from the Hong Kong Street historic street of ancient Zhangzhou, described above. A group of subjects with different backgrounds is selected for VR experiment and data on their perception of the scene is collected using experiment, behavioral record analysis and questionnaire survey. Analysis is then performed on the data collected from the three methods.

3.1. METHODS: 3-D EYE-TRACKING ANALYSIS

Motivated by the traditional eye-tracking analytic principles, one of the innovations in this paper is to combine the spatial depth perception with the people’s visual sensitivity in a 3-D VR environment. According to Besharse and Bok (2011), visual acuity declines by about 50% every 2.5° from the center axis to 30°, at which point visual acuity declines more steeply (Besharse and Bok 2011). In this way, 3-D eye-tracking analysis more approximates people’s actual visual behavior, facilitating research on spatial behavior. As we know, light travels along a line. Thinking analogously, we might regard the eye as a light source. That is, while observing an object, the subject emits a cone-type beam of light particles from the eye. According to established visual principles, the beam’s energy attenuates from its center to the periphery. And the energy attenuation coefficient is equal to the dynamic distribution of attenuation of visual sensitivity.

Then, the object’s visual thermal value is the accumulated value of the duration of attention on object multiplied by visual acuity, where visual acuity is defined as the a/2.5°th power of 1/2.
\[ C = \sum_{i=1}^{n} T \times V_{A}, \quad i = 1, 2, 3 \ldots n \]  

(1)

Where: visual thermal value is \( C \), attention object time is \( T \), tracking scan order is \( n \).

\[ V_{A} = \left( \frac{1}{2} \right)^{\frac{\pi \alpha}{180}}, \quad \in [0, 30^\circ] \]  

(2)

Where: visual acuity is \( V_{A} \), viewing angle is \( \alpha \).

The particle at the center of beam-like line of sight has more energy than those at the edge. Where the subjects can see is where the beam of light shines. The point attended to by the eye for a long time has more energy from the radiation of beam light, as shown in Table 1. Afterwards, the distribution of particle energy on the surface of 3-D object is determined thereby producing a 3-D eye-tracking thermal map.

3.2. PARTICIPANTS

Initially using random sampling, 100 experimental subjects were chosen. Due to visual limitations of 18 of the participants, we only considered the data of 82 subjects (36 males and 46 females) with an average age of 24.5 (SD=5.3) and ranging in age from 18 to 50. Subjects have lived in southern Fujian for an average of 5.1 (SD=6.9) years.

3.3. APPARATUS

The Spatial Behavior Simulation System is a multi-component interconnected system that enables the rapid exchange of data between the Simulation Monitoring terminal and Data Acquisition terminal through a shared server. In the Simulation Monitoring component, VR headsets (Hardware: HTC-Vive) and eye-tracking technology (Hardware: Optical-type eye tracker, Program: C++ and Unity3D) were used to construct the VR scenario and track user behavior in the VR environment in real-time, as shown in Figure 1(a). In the Data Acquisition terminal, the data from each experimental stage is recorded and aggregated. Visualization analysis is conducted on 3-D point-cloud data of spatial behaviors, as shown in Figure 1(b).

The core of this experimental system is the 3-D eye-tracking and spatial positioning data collection. Optical-type eye tracker and VR headsets are integrated in a single unit and the low-power infrared LED is used for multi-direction illumination. The eye-tracking data is then obtained from High-speed Eye Tracking Sensor. The optical-type eye tracker is located in the VR headset and thus is able to obtain accurate spatial location data from the VR headset using the signal of the base station. Finally, the spatial activity data is overlaid on the virtual scene using the software to provide the 3-D point-cloud data of spatial behavior.
INTEGRATION OF VIRTUAL REALITY, 3-D EYE-TRACKING, AND PROTOCOL ANALYSIS FOR RE-DESIGNING STREET SPACE

3.4. VR ENVIRONMENT

The virtual Hong Kong Street has a south-north length of 95.2 m, width of 3.2-5.8 m and height of 3.4-9.2 m. The moving range of subjects is the street between two memorial arches with a spacing of 26.8 m, as shown in Figures 2(a)(b). First, the 3-D scanner is used to survey and draw the existing Hong Kong Street, as shown in Figure 2(c). The obtained models are then simplified and corrected via 3Dmax. Afterwards, the buildings and environments of the street space are redesigned. Finally, the redesigned models are imported to Unity Engine 4, as shown in Figure 2(d). Meanwhile, there are 3348 objects in this VR scenario, all objects in the model are sequentially numbered for automatic data calculation.

3.5. EXPERIMENT SETUP FOR PROTOCOL ANALYSIS

The subjects enter the virtual street space of Hong Kong Street wearing the helmet with integrated eye-tracking and VR. Each subject experienced the virtual Hong Kong Street for two minutes. The 3-D point-cloud data of each subject is collected in this way. Finally, the entire experience is recorded by the camera for subsequent...
query and review of activities. The experiment consists of three stages, details are as follows.

First, the subject complete questions providing personal information and information on cultural background, and then get familiarized with the VR environment and how to operate in that environment. After wearing the equipment, they adjust their eyeballs to the monitoring zone before the experiment begins and match the 9-point eye position. The subjects are asked to answer the following three questions during their VR process.

- "Please identify the most representative architectural element of the traditional culture of southern Fujian province in the scene."
- "Please identify the most attractive architectural element of the traditional culture of southern Fujian province in the scene."
- "Please confirm whether this scale of street space is appropriate."

Next, during the experiment, three types of streetspaces with different dimensions were simulated, and subjects experienced these spaces for periods of two minutes. In the VR traditional street space, the subjects describe their experience and answer the three questions at their leisure. If they appear to be lost, they are reminded of the need to note their experiences in order to answer the three questions and describe their feelings with little time for deep reflection. Meanwhile, during the process of experiencing the streetspaces, spatial behaviors were recorded using a camera, and an inquiry and review of activity during the experience was conducted after completion. The quantitative analytical methods used in the questionnaire include descriptive statistical analysis and Pearson correlation analysis.

Finally, the 3-D point-cloud visual data of spatial behavior collected via the Data Acquisition Terminal is used to show subjects their individual eye-tracking and the features they attended to on the large display. The subjects review their experiences one by one. Subject eye-tracking data and their descriptions are verified. Special eye movement phenomena are queried and corrected.

4. ANALYSIS AND RESULTS

By combining data collected from the above two experiments, this study determines similarities and differences between spatial behaviors exhibited by different groups of people. Analysis of accumulated data clearly indicates that the subjects’ attention was concentrated in only part of the street space. There are 3384 objects in this scene. For each, we have a detailed description of spatial location and an independently collected point-cloud database of viewing behavior, that is, the spatial locations of the gaze of each viewer. Each time the surface of the 3-D object is viewed, the visual thermal value of that particular point increases; this value increases with viewing duration. After superimposing the data of spatial behaviors of all subjects, if the value of certain part exceeds a threshold, it will be shown in red in the thermal map and can be regarded as a hotspot. According to the thermal map, there were only 82 hotspots, most of which were cultural elements of southern Fujian architecture and nostalgic objects. These hotspots represent a very small number of the total 3384 objects of the model in the scene. In other words, the subjects are only interested in 2.4% of objects in the VR space, ignoring
the influence of occlusion. This observation reveals some patterns of the people’s spatial behavior in this street space, as shown in Figure 3.

According to their background in architecture, the subjects are classified into two types: professional users (architects) and naive users (non-architects). Among them, 27 subjects are professional users (17 local architects and 10 non-local architects), 54 subjects are naive users. Data analysis of the eye-tracking thermal map reveals considerable difference between the two groups.

The relative elevation of viewed objects can be obtained from a regression analysis of point-cloud data of viewed locations. Spatial behavior of the samples produces a mass data. Two minutes of viewing on the part of one user generates about 60,000 spatial point-cloud data arrays. 82 samples will generate approximately 4,920,000 groups of data. Therefore, it is a major task to analyze the mass of accumulated data. In this paper, we used SQL (Structured Query Language) to extract elapsed viewing time and the height of viewed point from the MDB database. Regression analysis of the eye-tracking data is then performed, with elapsed viewing time on the X axis and height on the Y axis. Analysis results indicate that the average height of the viewed point is 3.82 m for professional subjects, 1.83 m higher than 1.99 m calculated for inexperienced subjects. At the same time, the regression lines of the two datasets are approximately horizontal, as shown in Figure 4.

- Subjects with a background in architecture. Professional users’ eye movement spatial behavior summary thermograms show a total of 45 hotspots, the eye-tracking data shows that their attention is primarily concentrated on the wooden bucket arch, the cornice, the window, the portico and the memorial arch. Based on the height of the point of view of the scattered regression analysis, the samples with a background in architecture have a high sight point, and the height amounts to 3.83 m in average. And they are more interested in the second-floor features, as shown in Figure 4(a). According to protocol analysis, their attention is drawn to the red brick, the roof, the window, and
the memorial arch. The samples with a background in architecture paid more attention to the second-floor features. And they extracted more features in average. This observation is especially true for the local samples with a background in architecture, compared with the non-local subjects. The eye-tracking data of two expert samples indicates that over 73% of their time was spent on the second-floor objects. According to the protocol analysis results, the expert samples think that the features of southern Fujian street space are concentrated in the top of roof and chapter.

- Subjects without an architectural background. The eye-tracking data shows that their attention is primarily concentrated on the colonnade, the wooden bucket arch, the lantern, the memorial arch and nostalgic objects. According to protocol analysis, their attention is mostly focused on red brick, colored items, the arcade and the memorial arch. The samples without a background in architecture has a low sight point, and the height only amounts to 1.98 m in average, as shown in Figure 4(b). Most of these subjects concentrate their attention on the first floor and are very interested in nostalgic objects used in daily life. They can quickly identify architectural features in the first floor but pay little attention to the features in the second floor and above.

The samples with a background in architecture pay more attention to the second-floor architectural features than the samples without a background in architecture and they have little interest in the first floor. This finding can be attributed to the fact that those with professional training are more likely to
have their attention drawn to spatial combinations and aesthetic elements of the street space. For example, a series of gridded windows on the second floor characterized by conciseness and appropriate scale attracted the attention of 75.3% of the subjects. Attention distribution of local subjects is almost the same as that of non-local subjects.

Samples without a background in architecture demonstrate more interest in the first-floor features and nostalgic objects of daily use; they appear to have a taste for basic characteristics of the traditional street space in southern Fujian. This pattern is best explained by the fact that the traditional architecture of southern Fujian is easy to relate to. Although they cannot appreciate the profound aspects of architecture, the general public can properly understand the basic elements. For example, the samples without a background in architecture have little interest in the second-floor features and tend to be attracted by nonconforming factors like awnings and water pipes in disorder. But they understand the first-floor architectural features as accurately and quickly as samples with a background in architecture. They also pay a lot of attention to red brick, the colonnade, and the bucket arch. Meanwhile, the nostalgic objects associated with daily life do arouse the people’s homesickness. Many of the samples without a background in architecture view the nostalgic objects on the first floor (bike, bamboo chair and couplet), associating them with personal experience.

The understanding of street space varies between architects and the general public, especially when it comes to the historical street scene. For architects to tune into the public’s aesthetic appreciation is essential to design and conservation of street space. In the Production of Space, Lefebvre (1974) distinguished between architectural space and the space of architects. In his opinion, architectural space benefits from the people’s experience and it is one of the ways to produce social space. The space of architects refers to the space manipulated by the professional practice of architects. According to Lefebvre, the space of architects is neither neutral nor innocent, because the graph plot training and all of the other training that architects receive privilege visual perception over that of other sensory organs. In this context, the dominance of the visual sensation is maintained for extended periods of time and the extraordinary scene becomes a substitute for reality, separate from the aesthetic appreciation of the general public (Lefebvre, 1991). Hence, it is important to understand conservation and design of elements of spatial characteristics in the traditional street space of southern Fujian from the perspective of both architects and non-architects; eye-tracking and protocol analysis technologies potentially provide these different perspectives.

5. CONCLUSION

In summary, the paper reports on the implementation of a visual thinking method that relied on 3-D eye-tracking technology integrated in the VR headset. Using this tool, the experimental study of 82 subjects identified the visual components and patterns that capture people’s attention in the context of traditional preservation site, a virtual rendering of the Hong Kong Street in ancient Zhangzhou (Fujian, China). This analysis was supplemented by a protocol analysis that interrogated subjects as to which features they focused their gaze on and why. Experimental
subjects included architects, non-architects and the people with different cultural backgrounds, who perceived different elements of the virtual street space. The architects are found to differ greatly from non-architects in terms of their understanding of the simulated traditional street in southern Fujian. The former tends to focus their attention on the second-floor space and understands the formation of streetscape from the perspective of aesthetics. The latter, the non-architects, are mostly interested in the first-floor space. In addition to the first-floor features, they are also very interested in the space for daily life, especially the nostalgic elements associated with daily life, which arouse their homesickness and memories of home.

Therefore, conservation and redesign of streetscape in southern Fujian does not merely involve rebuilding the surface and reusing the space. Instead, it should fully and systematically accommodate the recognition and understanding of different groups. While embodying the fundamental architectural features, space and material, it should also strive to protect local culture and living space. The findings provide better references for VR-informed design, preservation and development of renovation of historic urban districts in architectural practice. As this paper only shows an example of Hong Kong Street, more work needs to be done for comprehensive simulation and analysis of the entire historical region. The types of traditional street space obtained from this paper are limited, but the analytic methodology and conclusion lay the foundation for simulation and experimentation across a wider region. Meanwhile, the methods and technologies used in the experimental study can be extended to support the integrated design of urban space in the future.

References
Jacob, R.J. and Karn, K.S.: 2003, Eye tracking in human-computer interaction and usability research: Ready to deliver the promises, Mind, 2, 4.