

Visual Presentation and Interactivity of Spatial Cognition in Dougong Structures: A Study Based on Eye Movement Characteristics and Cognitive Load

Qi Xia^{1*}, Xinyue Zhang¹

¹ School of Architecture and Art Design, Inner Mongolia University of Science and Technology, Baotou, China

*Correspondence: Qi Xia

Received: 21 April 2026; Accepted: 16 May 2026; Published: 15 June 2026

Abstract:

The Mental rotation is a key cognitive process in understanding complex geometric structures. The Dougong (bracket set) in traditional Chinese architecture, with its high spatial complexity, serves as an ideal subject for research in this field. This study investigates the effects of three visual presentation methods—photographs, 2D renderings, and 3D interactive models—on eye movement characteristics and cognitive load during the comprehension of Dougong spatial structures. Fourteen participants were recruited, and a Tobii Pro desktop eye tracker (sampling rate 120 Hz) was used to record their saccade duration, saccade amplitude, and pupil diameter under different presentation conditions. Statistical analysis was performed using a one-way within-subjects design and repeated-measures general linear model (GLM). The results are as follows: (1) Presentation method had an extremely significant main effect on pupil diameter ($F(2,26) = 135.330$, $p < 0.001$). Pupil diameter under the 3D interactive model condition was significantly smaller than that under 2D renderings, indicating that interactivity effectively reduced extraneous cognitive load. (2) Saccade amplitude showed a significant linear effect ($F(1,13) = 7.338$, $p = 0.018$). Digital models induced larger saccade amplitudes than photographs, reflecting more structure-oriented visual exploration. (3) The 3D interactive model elicited the longest fixation duration. Combined with the lower pupil diameter, this suggested that the process was not due to processing difficulty but deep cognitive processing driven by a closed-loop of operation–feedback–verification. In summary, 3D interactive models optimize the spatial cognitive efficiency of complex structures through the dual mechanisms of reducing cognitive load and promoting deep processing, providing empirical references for digital education and cultural heritage communication of ancient architecture.

Keywords:

Spatial Structure Comprehension, Dougong Structure, Visual Presentation, Eye Movement Characteristics, Interactivity, Cognitive Load

1. Introduction

Spatial cognitive ability is a core cognitive function for humans to process, understand, and memorize environmental information, playing an important role in architecture, engineering design, virtual reality, education, and other fields [1]. Effective spatial structure comprehension and accurate encoding of complex 3D objects are among the core challenges in spatial cognition research. As a sophisticated mortise-and-tenon structural system in traditional Chinese architecture, Dougong is composed of interlocking components such as Dou, Sheng, Ang, and Qiao, featuring high geometric complexity and rigorous 3D spatial logic [2]. Understanding Dougong construction involves not only static geometric shape recognition but also complex mental rotation and spatial logic reconstruction in the mind, making it an ideal subject for studying spatial comprehension of complex structures.

Mental rotation (MR) is an important aspect of visuospatial information processing, referring to the mental rotation of objects to judge whether images from different perspectives match [3]. Analyzing, locating, and encoding the complex spatial relationships of Dougong essentially relies heavily on this cognitive ability. Existing studies have confirmed that angular differences and dimensional features of stimuli affect the speed and accuracy of mental rotation [4]. However, most current research focuses on comparisons between static 2D or 3D stimuli, with relatively insufficient empirical investigation into the role of interactivity in complex structure cognition [5].

As a non-invasive cognitive measurement tool, eye-tracking reflects cognitive processing in real time and objectively [6]. Saccade duration is related to cognitive inhibition during spatial information processing; longer saccades usually imply higher task complexity [7]. Saccade amplitude reflects the spatial scope of visual search and indicates the influence of object dimensionality on cognition [8]. Pupil diameter is a sensitive physiological indicator of cognitive workload (CWL), with dilation associated with increased task complexity [9]. Cognitive Load Theory (CLT) states that learning material design should minimize extraneous cognitive load to free working memory resources for deep processing [10]. Therefore, optimizing presentation methods to reduce cognitive load and improve efficiency has long been a focus in human–computer interaction and digital education.

Although 3D digital models are widely used, empirical support for how interactivity shapes individual cognitive paths remains limited. According to CLT [10], inappropriate presentation generates ineffective extraneous cognitive load. This study compares photographs, 2D renderings, and 3D interactive models to examine whether interactivity can effectively externalize mental rotation and optimize cognitive efficiency. Using five groups of Dougong as experimental materials under three presentation conditions, this study combines eye-tracking to investigate the effects of presentation and interactivity on Dougong spatial cognition from three dimensions: cognitive resource consumption–visual search strategy–depth of information processing. The following hypotheses are proposed:

H1: Presentation mode significantly affects cognitive load; 3D interactive models reduce extraneous load more effectively than 2D renderings.

H2: Digital models (2D renderings and 3D models) induce more structure-oriented visual search strategies than photographs.

H3: 3D interactive models elicit deeper cognitive processing through an operation–feedback–verification mechanism.

2. Materials and Methods

Fourteen healthy graduate students (7 males, 7 females; mean age = 24 years) from the School of Architecture and Art Design participated in this study. All participants had no history of mental disorders and had normal or corrected-to-normal vision. Written informed consent was obtained prior to the experiment, and all participants received financial compensation after the test.

The experiment was conducted in a quiet laboratory with stable lighting and ventilation. Binocular eye movements were recorded at 120 Hz using a Tobii Pro desktop eye tracker. Visual stimuli were presented on a 24-inch monitor with 1920×1080 resolution. Participants sat comfortably at a distance of approximately 65 cm from the eye tracker and monitor and were instructed to view freely. Before formal testing, basic information registration and a nine-point calibration were performed to ensure the accuracy of eye-tracking data.

2.1. Procedure

Before the formal experiment, instructions were explained in detail, and practice trials with non-experimental Dougong structures were provided to familiarize participants with the eye tracker and task interface. The formal task included five sets of Dougong stimuli, each presented as photographs, 2D renderings, and 3D interactive models in sequence as shown in Figure 1. For the 3D condition, participants could freely rotate and zoom using mouse and scroll wheel to assist component identification and spatial relationship understanding. No real-time response was required. An event-marking system precisely marked stimulus onset and offset times, ensuring accurate synchronization between eye-tracking data and stimulus grouping for subsequent analysis.



Figure 1. 2D renderings and 3D interactive models.

2.2. Measures

In this study, eye-tracking data were recorded using an eye-tracking system to analyze participants' visual exploration strategies and cognitive processing depth during the spatial cognition of Dougong structures. The eye-tracking metrics collected included: Saccade duration: the time required to complete one saccade, reflecting the inhibitory period in spatial information processing. Saccade amplitude: the angular distance between two consecutive fixations, indicating the scanning range in a single eye movement and the integrity of information encoding. Pupil size: quantified as pupil diameter, a sensitive physiological indicator of cognitive load and task complexity [11]. These metrics provide objective data support for revealing the effects of different presentation modes on the spatial cognition of Dougong.

3. Results of Pupil Diameter Analysis

A one-way within-subjects design was adopted with presentation mode as the independent variable and pupil diameter as the dependent variable, analyzed using repeated-measures GLM. Mauchly's test of sphericity indicated that the assumption of sphericity was satisfied ($p = 0.721 > 0.05$); thus, unadjusted F-statistics were reported for within-subjects effects. Multivariate tests showed an extremely significant combined effect of presentation mode on pupil diameter (Pillai's trace = 0.964, $F(2,12) = 159.803$, $p < 0.001$), explaining 96% of variance and indicating a large effect size. Within-subjects tests further revealed extremely significant differences in pupil diameter across conditions ($F(2,26) = 135.330$, $p < 0.001$). Within-subjects contrasts showed no significant linear trend ($F(1,13) = 4.075$, $p = 0.065 > 0.05$) but a highly significant quadratic effect ($F(1,13) = 344.240$, $p < 0.001$) as shown in Table 1.

Table 1. General Linear Model of Pupil Diameter.

	Value	Significanc	F
Mauchly's Test of Sphericity		0.721	
Multivariate Tests	0.964	<.001	159.803b
Tests of Within-Subjects Effects		<.001	135.330
Tests of Within-Subjects Contrasts		0.065	4.075
Tests of Between-Subjects Effects		<0.001	344.240

From a visual cognitive perspective, the quadratic effect stems from differences in information complexity and depth cues. Photographs contain redundant texture, lighting, and background details, demanding more visual working memory and resulting in the highest cognitive load and largest pupil dilation [12]. Although 3D models provide stereoscopic depth, interactive manipulation moderately increases processing demands. 2D renderings lack depth and redundant details, yielding the lowest visual load. The pupil diameter gradient photograph > 3D model > 2D rendering corresponds to the nonlinear quadratic pattern[13]. While 3D interactive models were not the lowest in load, they outperformed photographs, supporting the hypothesis that interactivity reduces extraneous load.

3.1. Fixation Duration Results

Saccade amplitude showed a significant linear effect ($p = 0.018$) as shown in Table , driven by gradients in spatial information density. Photographs have dispersed layouts and rich details requiring larger scanning amplitudes, while 3D and 2D conditions have progressively simplified information and smaller amplitudes. The divergent effect patterns reflect separate influences of information redundancy and spatial layout density on visual processing (Table 2).

Table 2. General Linear Model of Saccade Amplitude.

	W	Significanc	F
Mauchly's Test of Sphericity	0.891	0.500	
Multivariate Tests		0.370	3.525b
Tests of Within-Subjects Effects		<.050	4.060
Tests of Within-Subjects Contrasts		0.018	7.338
Tests of Between-Subjects Effects		<.001	1459.932

From the perspective of complementary cognitive mechanisms, the quadratic effect of pupil diameter explains how presentation mode affects resource usage, while the linear effect of saccade amplitude explains how it shapes spatial search strategy. Photographs combined high load and large amplitude, indicating increased both

processing cost and spatial coverage demand. 2D renderings showed both low load and small amplitude, confirming low redundancy and high spatial focus. This dual-index pattern overcomes limitations of single-metric studies and systematically clarifies the combined cognitive-behavioral effects of visual presentation.

3.2. Fixation Duration Results

Fixation duration, a core indicator of processing depth, was analyzed using repeated-measures GLM. Mauchly's test indicated violation of sphericity ($W = 0.320$, $p = 0.001 < 0.05$); thus, Greenhouse–Geisser corrected degrees of freedom ($\varepsilon = 0.595$) were used. Multivariate tests showed a marginally significant combined effect (Pillai's trace $F(2,12) = 3.463$, $p = 0.065$). Within-subjects contrasts revealed a significant linear effect ($F(1,13) = 7.425$, $p = 0.017$) and a nearly significant quadratic effect ($F(1,13) = 5.769$, $p = 0.032$), indicating dual characteristics of information-complexity-driven monotonic gradient and processing-depth deviation under specific modes. as shown in Table 3.

Table 3. All tables are to be numbered using Arabic numerals.

	W	ε	Significance	F
Mauchly's Test of Sphericity	0.320		0.001	
Multivariate Tests			<.065	3.463
Tests of Within-Subjects Effects		0.595	<.001	135.330
Tests of Within-Subjects Contrasts			0.017	7.425
Tests of Between-Subjects Effects			0.032	5.769

Combined with pupil and saccade results, the three eye-tracking metrics establish a hierarchical cognitive model: information complexity (photograph > 3D model > 2D rendering) regulates cognitive load via redundancy, shapes scanning range via spatial density, and determines processing granularity via detail demands. Photographs exhibited the highest load, largest amplitude, and longest duration, supporting a coherent chain: high redundancy → high resource occupation → wide spatial search → deep detail processing. Deviations in 3D and 2D gradients reflect modulatory effects of stereoscopic depth cues. This multi-dimensional framework advances single-metric research and provides empirical support for linking visual stimulus design to cognitive mechanisms.

4. Effects of Presentation Mode on Spatial Cognitive Load of Dougong

As a physiological indicator of cognitive load, pupil diameter directly reflects information-processing difficulty. Presentation mode had a significant main effect on mean pupil diameter, with 3D interactive models showing significantly smaller diameters than photographs. According to Sweller's CLT, although 2D renderings provide visual detail, their fixed perspective forces intensive mental rotation and spatial completion to interpret complex mortise-and-tenon relationships, imposing high extraneous load. In contrast, 3D interactive models allow free rotation and zooming, providing dynamic motion parallax. Self-controlled perspective breaks spatial relationships into continuous logical slices, reducing the processing cost of 3D mental model construction and decreasing pupil diameter [14]. The slightly higher load of 3D models compared with 2D renderings stems from operation–feedback–verification interactivity, which constitutes germane cognitive load that enhances deep

spatial encoding rather than impeding it—supported by the longest fixation durations in the 3D condition.

4.1. Similarities and Differences in Visual Exploration Patterns

Saccade amplitude showed no significant difference between 3D and 2D conditions, suggesting bottom-up consistency in visual search for homogeneous digital stimuli. Due to the high repetitiveness and symmetry of Dougong components, gaze jumps concentrated on core elements (Dou, Sheng, Ang, Qiao) in both interactive and static digital viewing.

Photographs contain rich environmental and textural details, whereas digital models eliminate background interference and focus attention on structure. This structure-focusing effect leads to more frequent and wide-ranging saccades to capture geometric relationships in digital shading, demonstrating digital media's advantage in guiding structural logic comprehension.

4.2. Cognitive Significance of Deep Fixation and Interactive Feedback

The 3D model elicited the longest mean fixation duration, clearly differing from 2D and photograph conditions. According to attentional allocation theory, longer fixations correspond to deeper information uptake and processing. In the 3D interactive environment [15], extended fixations reflect not difficulty but a closed-loop operation–feedback–verification cognitive process.

When rotating the model, the brain processes real-time visual feedback of structural changes, matching dynamic visual input with existing architectural component knowledge. Compared with the superficial processing of 2D renderings, 3D interactivity induces deep immersive processing, which supports more robust long-term memory representations—highly valuable for science education and research on logically rigorous ancient architecture.

4.3. Summary: Balancing Cognitive Efficiency and Deep Exploration

Integrating eye-tracking metrics, the 3D interactive model acts as a cognitive mediator in digital heritage display: it reduces instantaneous load to enable understanding and elicits deep fixation to promote internalization. This confirms that interactivity in cultural heritage digital communication is not merely functional enhancement but cognitive empowerment, reshaping attention allocation to efficiently convey the spatial logic of complex wooden structures.

5. Conclusions

This experimental study compared eye-movement behaviors when viewing photographs, 2D renderings, and 3D interactive models of Dougong, examining how digital presentation affects spatial cognition.

(1) 3D interaction alleviates cognitive pressure in complex structure observation. Pupil diameter was significantly smaller in the 3D condition than in 2D renderings. Although 2D images appear intuitive, interpreting layered mortise-and-tenon structures requires intensive mental effort for spatial inference. Free perspective switching in 3D models clarifies component logic and reduces comprehension difficulty.

(2) Interactive manipulation guides deep structural inspection. 3D models induced longer fixation durations. Combined with lower pupil diameter, this confirms high-

quality deep processing via operation–verification rather than processing difficulty. Active engagement moves beyond passive browsing to genuine verification of spatial relationships, achieving higher-quality cognitive participation.

(3) Digital models are more structure-oriented than photographs. Saccade analyses show more frequent gaze jumps between components in digital conditions, demonstrating that digital media reduce environmental distraction, strengthen geometric contours, and direct attention to structure—highly practical for ancient architecture popularization.

Based on a correlation analysis of multiple eye-tracking indicators concerning cognitive efficiency and in-depth exploration, 3D interactive models serve as a critical cognitive mediator in the digital display of ancient architecture. On the one hand, through dynamic motion parallax, 3D models visually externalize complex spatial logic, effectively reducing the instantaneous cognitive load on viewers when comprehending structures such as mortise-and-tenon joints, thereby enabling lightweight and efficient cognitive processing. On the other hand, supported by the interactive closed loop of “operation–feedback–verification”, these models encourage users to sustain longer durations of focused gaze, prompting them to actively integrate visual information with architectural component knowledge and facilitating the deep internalization of such knowledge. Interactivity represents more than mere functional optimization; it reshapes patterns of attention allocation, shifting visual processing from superficial scanning to in-depth engagement with structural logic, thus efficiently conveying the intricate spatial relationships of traditional wooden architecture. Accordingly, the 3D interactive approach achieves a favorable balance between cognitive efficiency and exploratory depth, providing empirical support for the transformation of digital communication of cultural heritage from function-driven to cognition-enabled.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Author Contributions

Conceptualization: Qi Xia, Xinyue Zhang; Methodology: Qi Xia; Software: Qi Xia; Validation: Qi Xia, Xinyue Zhang; Formal analysis: Qi Xia; Investigation: Xinyue Zhang; Resources: Qi Xia, Xinyue Zhang; Data Curation: Qi Xia; Writing – original draft preparation: Qi Xia, Xinyue Zhang; Writing – review and editing: Qi Xia; Visualization: Xinyue Zhang; Supervision: Qi Xia, Xinyue Zhang.; Project administration: Xinyue Zhang; Funding acquisition: Qi Xia.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Acknowledgments

The author would like to acknowledge Professor Miao Xiu for providing support with the eye-tracking equipment, and Professor Li Yuanhe for his educational guidance and assistance throughout the research.

References

- [1] Hegarty, M.; David, W. A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 2004, 32(2), 175-191.
- [2] Liang, S.C. History of Chinese Architecture. Tianjin: Baihua Literature and Art Publishing House, 2005.
- [3] Shepard, R.N., and Jacqueline Metzler. Mental rotation of three-dimensional objects. *Science*, 1971, 171(3972), 701-703.
- [4] Parsons, L.M. Imagined spatial transformations of one's hands and feet. *Cognitive psychology*, 1987, 19(2), 178-241.
- [5] Jang, S.S, et al. Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, 2017, 106, 150-165.
- [6] Kennedy, A. Book review: Eye tracking: A comprehensive guide to methods and measures. 2016, 607-609.
- [7] Irwin, D.E. Information integration across saccadic eye movements. *Cognitive psychology*, 1991, 23(3), 420-456.
- [8] Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 1998, 124(3), 372.
- [9] Beatty, J. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological bulletin*, 1982, 91(2), 276.
- [10] Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 1994, 4(4), 295-312.
- [11] Rogers, B.; Maureen, G. Motion parallax as an independent cue for depth perception. *Perception*, 1979, 8(2), 125-134.
- [12] Mayer, R.E. Multimedia learning. Psychology of learning and motivation. *Academic Press*, 2002, 41, 85-139.
- [13] Just, M.A.; Patricia, A. Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 1980, 87(4), 329.
- [14] Stull, A.T.; Mary, H. Model manipulation and learning: Fostering representational competence with virtual and concrete models. *Journal of Educational Psychology*, 2016, 108(4), 509.
- [15] Craik, F.I.M.; Robert, S.L. Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 1972, 11(6), 671-684.



© 2026 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)