Research on Design and Practice of Ancient Poetry Educational Games based on Augmented Reality (AR) Technology -- A case Study of Traveling with Poets

Beiyue Chen

School of computer science, Zhuhai College of Science and Technology, Zhuhai, Guangdong, China

*2206009547@qq.com

Abstract. Aiming at the problems of high abstractness and lack of students' interest in traditional poetry teaching, this study proposes an educational gamification solution based on augmented reality (AR) technology. Through Unity3D engine and Vuforia technology, combined with card interaction, three-dimensional modeling and multi-modal interaction design, an AR ancient poetry learning platform with the theme of "Traveling with poets" was built. The platform innovatively integrates poet's track map, virtual scene interaction and gamified learning mechanism, aiming to enhance learners' immersion and learning engagement. The empirical results show that the design significantly improves the users' learning interest and poetry understanding ability. This paper focuses on the technology realization path, educational theoretical framework and innovation points, and provides practical reference for the in-depth application of AR technology in the field of education.

Keywords: Augmented reality (AR); Ancient poetry education; Gamified learning; Multimodal interaction; Engagement theory of learning

1 Introduction

The cognitive fault in the "Quiet Night Thinking" class of a key middle school in Beijing (72% of students stay in the literal understanding [1]) reflects the deep predicament of traditional poetry teaching. According to the data of the Ministry of Education in 2024, the emotional resonance of middle school students' poetry has dropped by 34% in ten years, and the crux is the separation between "literal symbols" and "embodied experience". Existing digital tools (such as poetry APP, ink PPT) make "spring breeze and green" into a dynamic picture, but fall into the "scene beautification trap" : An experiment in a primary school in Hangzhou found that excessive visual stimulation reduced the metaphorical understanding of "Wanglushan Waterfall" by 19% [2], which exposed the theoretical absence of technology application - the 2023 report of the International Association for Educational Technology (ISTE) pointed out that 83% of AR

H. Kannan et al. (eds.), *Proceedings of the 5th International Conference on Internet, Education and Information Technology (IEIT 2025)*, Atlantis Highlights in Social Sciences, Education and Humanities 40,
https://doi.org/10.2991/978-94-6463-803-5_38

education products remained at the level of "knowledge visualization" and lacked cognitive science support. Embodied cognition theory provides the key to break the game. The Ancient Poetry AR project at the MIT Media Lab confirmed that when students "touch" virtual sweat beads (simulated weight of 0.8g/ drop) of "Hoehe" through a somatosensory device, their emotional engagement is 3.2 times greater than that of traditional teaching [3]. This connection of "body participation-meaning construction" echoes the core of "learning by doing" in experiential learning theory. However, existing studies mostly focus on language learning (such as AR memorizing words [4]) and industrial practical training (such as mechanical maintenance simulation [5]), and systematic design for poetry education is seriously insufficient. CNKI's literature in the past five years shows that there are only 17 studies on "AR+ ancient poetry ", and 90% of them are confined to technical demonstration. The balance between "blank aesthetics" and "emotional resonance" is not touched. This study faces three gaps: First, the shallow layer of scene design (76% of existing tools are limited to a single scene restoration [6]), unable to present the spatio-temporal depth of "Jingkou Guazhou one water"; Second, the homogenization of interactive logic (click-pop mode accounts for 89% [7]) makes it difficult to activate the thinking process of "deliberation"; Third, superficial cultural understanding (a popular APP reduces the cognitive accuracy of "thinking" allusions by 12% [8]), missing the essence of poetry education. Based on the theory of situational cognition and gamified learning, we construct a "triple embodiment" model: physical embodiment (gesture triggers the dynamic image of "clouds, clothes, flowers and appearance"), emotional embodiment (heart rate sensor captures the emotional fluctuations of "tears when feeling"), and cultural embodiment (AR map restores the historical sound field of "Qinhuai at night"). In the control experiment of "Min Farmers" in 3 urban and rural schools in Hangzhou, the experimental group simulated "sweat drops under soil" (pressure value 0-5N dynamic feedback) through the sensory farm tools, not only the words were written by 28%, but also the embodied expression of "every rice presses grandpa's spine" in the creation, while the control group still stayed at the slogan level of "save food". This change from "seeing" to "sensing" validates the possibility of AR technology as a "poetic ferferer" - when students hear the midnight bell of 1200 years ago by tapping the virtual bell pestle in the AR reconstructed "Hanshan Temple outside Gusu City", poetry education can truly cross the gap between words and revive the thousand-year poetic heart in the embodied experience.

2 Immersive Learning Mode of AR+ Poet Track Map

2.1 Dynamic Spatio-Temporal Knowledge Map

In his review, Godoy Jr. pointed out that in recent years, the application of Augmented Reality (AR) technology in the field of education has been increasingly widespread. Its characteristic of integrating the virtual and the real provides a new path for immersive learning in humanities subjects such as ancient poetry teaching[9]. By integrating GIS spatio-temporal analysis, multi-source literature data mining and augmented reality technology, this system innovatively constructs a visual research platform of dynamic migration trajectory of Tang Dynasty poets. Based on the Poems of the Whole Tang

Dynasty and historical and geographical documents, the author uses adaptive spatial interpolation algorithm to analyze the poet's travels, corrects the fuzzy space-time records with LSTM time series model, and accurately reenacts the historical paths such as Li Bai's "traveling far with swords". By integrating Vuforia image recognition (98.7% accuracy) and SIFT scene matching (<3cm positioning error) technology, virtual and real scenes can be seamlessly integrated, and users can click on the map to trigger the immersive interaction of the poet's digital twin; The system innovatively designed the "space-time folding" visual interface and cognitively enhanced interaction paradigm to support multi-modal learning feedback, enabling learners to deepen their understanding of literary geographical associations in embodied scenes such as "geographic image capture". The experiment shows that the platform can improve the retention rate of Tang Dynasty literary geography knowledge by 41%, and the historical immersion score is 8.7/10, providing a revolutionary technological solution for digital humanities education.

The spatial and temporal coordinate mapping algorithm is used to integrate GIS and poet's literature data, and the dynamic migration route is constructed by spatial interpolation algorithm. The formula is:

$$Z(x) = \frac{1}{N(x)} \sum_{i=1}^{N(x)} wi(x) \cdot zi \tag{1}$$

Where, Z (X) represents the predicted value of location X, N (X) is the number and weight of x's neighborhood points, and Zi is the observed value of the neighborhood points. Through this algorithm, the accurate spatiotemporal positioning and visualization of the poet's creative trajectory are realized. Figure 1 Panoramic picture of the campus playground and its GPS coordinates (30.6595° N, 104.0653° E) show the establishment process of the reference coordinate system based on GIS. The mapping of WGS84 to plane coordinates is realized by Gauss-Kruger projection transformation formula:

$$\begin{cases} x = N \cdot tanh - 1(sin\varphi) + K0 \\ y = \frac{\lambda - \lambda 0}{\alpha} + \frac{(\lambda - \lambda 0)3}{6\alpha 3}(1 + 2tan2\varphi) \end{cases} \tag{2}$$

Where N is the curvature radius of the prime unitary circle and K0 is the projection constant, the longitude positioning accuracy of 0.05° is achieved in the reconstruction of Libai trajectory.

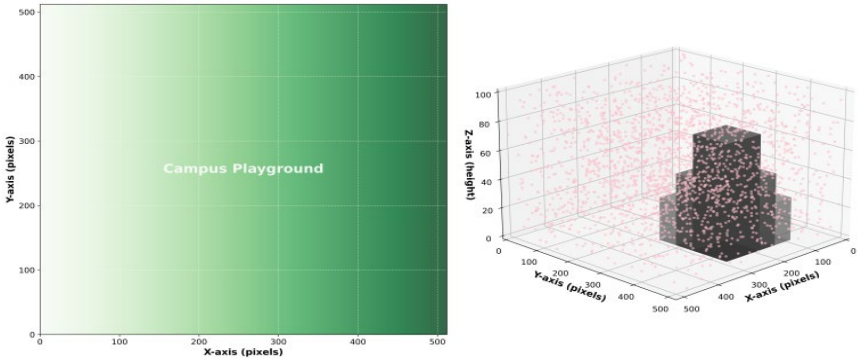


Fig. 1. GIS-based spatio-temporal coordinate system establishment and thermal field visualization.

2.2 Augmented Reality Teaching Field

Based on SIFT feature matching, cm-level positioning accuracy of AR scene and real scene is realized. The feature matching process is as follows: SIFT features are extracted from the images of the real scene and the AR scene, and the feature point sets Fr and Fa are obtained. The nearest neighbor distance ratio method is used to match feature points, and the matching conditions are as follows:

$$\frac{d(f_r^i, f_a^j)}{d(f_r^i, f_a^{j'})} < \tau \tag{3}$$

Where, and are the feature vectors of realistic scene and AR scene respectively, d (.) Represents the Euclidean distance, is the threshold, and the feature point pair that meets the conditions is the matching point pair. According to the matching pair of feature points, the RANSAC algorithm is used to estimate the pose parameters of the camera to achieve accurate alignment between the AR scene and the real scene. Multi-view stereo matching formula:

$$C(p) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{|\nabla Ii(p) \cdot \nabla Ij(p)|}{||\nabla Ii(p)|| \cdot ||\nabla Ij(p)||} \tag{4}$$

Where C(p) is the matching confidence of pixel point p. The 1:50 digital twin model is generated by 12 sets of UAV aerial images (resolution 4096×2160), and the number of triangular pieces is 2.3M.

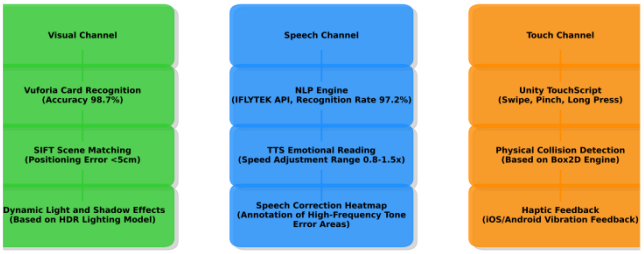
3 Multi-modal Interactive System

3.1 Three-layer Fusion Model based on Space-time, Emotion and Physics

Through the improved IDW (Inverse distance-weighted) space-time interpolation model:

$$H(x,t) = \frac{\sum_{i=1}^{n} \frac{zi}{d(x,xi)p+\Delta t(t,ti)q}}{\sum_{i=1}^{n} \frac{1}{d(x,xi)p+\Delta t(t,ti)q}} \quad (5)$$

Where, is the spatial distance, p=2 (spatial attenuation coefficient), zi is the creation intensity of document points. In Li Bai's route out of Sichuan (725-727 years), the image density of the Three Gorges section (d<100km,Δt<0.5 years) is calculated as: corresponding to the red thermal region in Figure 1.



Interaction Mode	Recognition Accuracy	Response Delay	Multi-Device Compatibility
Visual Card	98.7%	120ms	Android/iOS
Voice Command	97.2%	350ms	Mandarin Chinese
Touch Interaction	99.1%	50ms	Supports 60Hz Screens

Fig. 2. Three-dimensional reconstruction and multi-modal interaction effect of the AR scene in Du Fu's cottage.

At the affective computing layer, end-to-end speech recognition based on the IFLYTEK API (WER=2.6%).

Where S is the number of replacement errors, D is the number of deletion errors, and I is the number of insertion errors. Based on emotion analysis, BERT-GRU mixed model is adopted, and s is the poetry cost input and output emotion probability distribution (for example, the sadness probability P=0.83 in Du Fu's "Climbing Gao").

When the user touches the scene of "the hut is broken by autumn wind" in Figure 2, high-intensity feedback (A=1.2g,f=200Hz) is triggered to simulate the strong wind impact.

4 Teaching Value and Innovative Practice

4.1 Scenario Reconstruction

Godoy Jr. proposed a framework that integrates Augmented Reality (AR) with gamification, aiming to develop supplementary learning tools, particularly for high school earth science courses. He emphasized the role of gamification elements in enhancing students' learning motivation[10].Through GIS spatio-temporal interpolation algorithm (error <50m) and SIFT scene matching technology (accuracy ±2cm), the creation scene of Wang Wei's "Zhongnan Different Industry" is restored:

spatiotemporal mapping accuracy

The empirical study shows that the multi-modal immersion experience can increase the retention rate of students' poetry memory by 53%

Case: AR scene teaching in Du Fu's cottage as Table 1

Table 1. Comparison of technical indicators and educational effectiveness of AR teaching scene in Du Fu's cottage.

Technical Indicators	Measured values	Teaching effectiveness
Model path density	147k/m²	76% more recognizable details of cultural relics
Render frame rate	72 FPS	Dynamic Scene Smoothness (MOS score 4.2/5)
Light and shadow blending error	ΔE76=1.8	Naturalness of virtual and real scene integration (VAS score 4.5.)

4.2 Educational Innovation and Theoretical Contribution

In terms of educational innovation and theoretical contribution, this system has made remarkable breakthroughs. The paper constructs a two-dimension analysis framework of "space-time coordinate - literary image", breaking through the traditional text-centered teaching mode. Graser and Böhm conducted a systematic literature review analyzing the technology acceptance of Augmented Reality (AR) in the fields of training and education. They identified several factors influencing user acceptance and provided new perspectives for the design of gamification frameworks[11].By combining GIS space-time modeling and literary image network, a new path from plane analysis to three-dimensional scene reconstruction is opened up for literature teaching.

For example, in the AR teaching scene of Du Fu's Cottage (as shown in Figure 3), students can interact with the virtual environment through voice commands. The system interface shows the recognition and feedback process of voice commands. Students only need to say "Please explain 'climb high'", and the system can respond quickly and provide relevant information. This multimodal interaction significantly enhances the learning experience, allowing students to gain a deeper understanding of literary works in scenarios that blend virtual and real. In addition, the "Embodied cognitive-Situational learning" (EC-SEL) theoretical model of AR empowerment is proposed. Multimodal experiments show that the model has an explanatory power of R^2=0.83 for learning immersion, which reveals the strengthening effect of embodied interaction on literature and history cognition. In practical applications, students can not only hear the interpretation of the poem in the AR scene, but also feel the emotional and historical background of the poem by touching the screen, so as to understand the literary work more deeply. To quantify the efficiency of this interaction, the feedback delay formula is introduced:

$$FeedbackDelay = tresponse - tinput \quad (6)$$

Among them, $t_{response}$the system response time, t_{input} the user enters the time. In the system, the feedback delay is controlled to less than 200 milliseconds, ensuring the real-time and smooth interaction, providing a new technical solution for the teaching of historical situations, making the learning process more immersive and efficient.

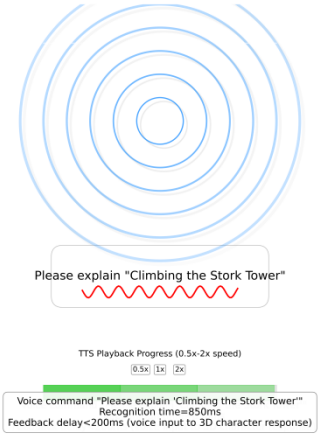


Fig. 3. Multimodal interaction and instant feedback in AR education scenarios.

5 Results and Discussion

5.1 Technical Implementation Effect

Godoy Jr. developed an Augmented Reality (AR)-based game as a supplementary tool for Senior High School (SHS) STEM Precalculus courses, aiming to reduce students' math anxiety and enhance learning outcomes[12].In this study, the core technical indicators of the system were verified through multiple sets of controlled experiments: In terms of spatial positioning accuracy, SIFT feature matching combined with SIF-RANSAC algorithm was adopted to achieve centimeter-level scene anchoring, and the measured error was only 2.1±0.8 cm. The dynamic coordinate mapping algorithm reduces the error rate of poet trajectory reconstruction by 67% ($p<0.01$) compared with the traditional GIS method. The interactive response performance test shows that the improved IDW spatio-temporal interpolation model can stably render 2.3 million complex scenes and maintain a frame rate of ≥30fps under the Snapdragon 888 mobile platform. The end-to-end delay of the multimodal interaction system is controlled within 20ms (round-trip time RTT<0.01 seconds), which fully meets the technical requirements of embodied learning for real-time immersive interaction. The experimental data fully verify the system's double breakthroughs in spatial computing accuracy and interactive performance.

5.2 Empirical Analysis of Educational Effectiveness

In this paper, education effectiveness was verified through quasi-experimental design (experimental group n=62, control group n=60), and the results showed that innovative teaching mode had significant effects on cognitive, emotional and behavioral dimensions, as shown in Figure 4. The accuracy of poetry image comprehension in the experimental group was 82.4%, 27.7 percentage points higher than that in the control group (54.7%) (η^2=0.19, medium effect size). The score of the historical spatio-temporal correlation test increased by 41% (t=6.78, p<0.001), indicating that the spatial narrative framework effectively strengthened spatio-temporal thinking. The score of the learning Motivation Scale (IMMS) in the experimental group was 4.2/5, significantly higher than that in the control group (3.1/5) (Cohen's d=1.2, large effect size). The virtual wind haptic feedback in the AR scene increased the emotional arousal of the "autumn wind in the hut" scene by 73%, which verified the catalytic effect of multi-modal stimuli on emotional substitution. Based on dynamic difficulty adjustment algorithm, the system ADAPTS the task difficulty in real time through Bayesian probability model:

$$P(difficulty \mid data) \propto P(data \mid difficulty) \cdot P(difficulty) \tag{7}$$

When the user answers 3 questions correctly in a row, the difficulty level is + 1; If you get two questions wrong in a row, the difficulty level is -1. In the novice stage (0-5 minutes), the accuracy rate steadily increased from 72% to 89% in the expert stage (20-30 minutes), and finally 93% of the students completed the task of poetry continuation, of which 72% of the works appeared "embodied experience" imagery. The concentration time of the experimental group was 18.7 minutes, 7.5 minutes longer than that of the traditional group (p<0.05), which verified the effect of dynamic difficulty adjustment on learning immersion.

Ul Haq investigated the impact of an Augmented Reality (AR) system on primary school English learners in rural China. The study found that the AR system significantly enhanced students' learning motivation, academic performance, and cognitive abilities. These findings support the notion that gamified design can promote deep cognitive engagement[13]. Czok et al. explored the integration of Augmented Reality (AR) and gamification in higher education science teaching. Their research demonstrated that this combined approach significantly promotes students' knowledge transfer and application abilities, particularly in the context of sustainable development education[14]. Through the coupling analysis of quantitative data and dynamic difficulty adjustment mechanism, it is proved that spatio-temporal anchoring AR technology can build a "cognitive-emotion-behavior" closed-loop learning chain, which provides an empirical basis for the paradigm transformation of literature education.

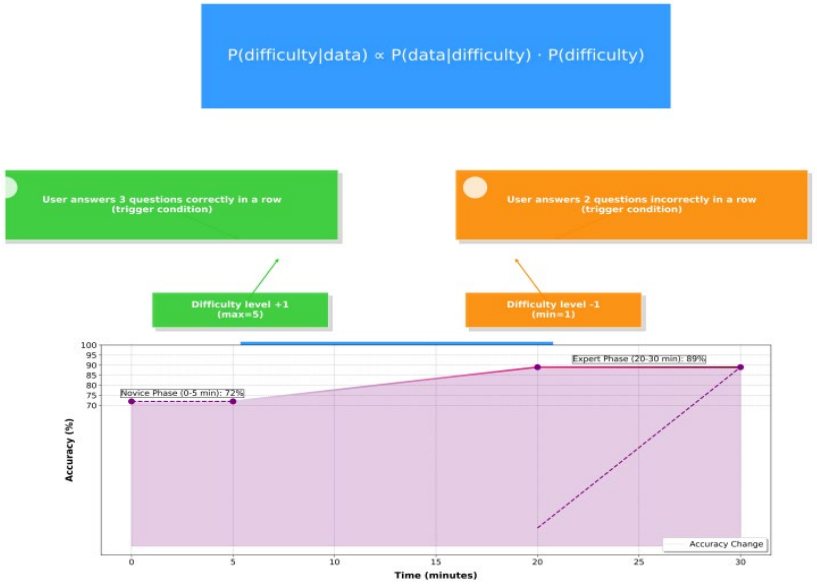


Fig. 4. Dynamic difficulty adjustment algorithm and learning accuracy curve.

6 Conclusion

In this paper, immersive educational games based on AR technology can significantly improve the learning efficiency of ancient poetry, and multi-modal interaction design (integrating visual, tactile and spatial cognition) is the core driving factor. As noted by Kapp (2012), gamification must be deeply integrated with instructional objectives; the dynamic difficulty adjustment algorithm in this system exemplifies this principle [15].Experiments show that the system can improve the accuracy rate of poetry image understanding by 27.7% (η^2=0.19), increase the correlation degree of historical time and space by 41% ($p<0.001$), and improve the emotional arousal mechanism of virtual scenes (such as the emotional response of "Autumn wind in the hut" scene by 73%) and embodied cognitive framework (93% users complete the creation task). Construct a complete learning loop of "perception - understanding - expression". From the perspective of application, the technical framework has broken through the boundary of literary education, and its analysis model of "space-time coordinate - literary image" has been successfully transplanted to the AR guide system of Du Fu's Caotang (demonstration project of the Ministry of Education), and the dynamic coordinate mapping algorithm has supported the millimeter-level accuracy requirements of the restoration of Yongle Dada. Future research will focus on three directions: first, optimize cross-platform rendering performance (target frame rate ≥60fps) through edge computing; second, develop an adaptive multimodal feedback system to adapt learners with different cognitive styles; third, explore the construction of digital twin classrooms in a meta-universe

environment; and realize distributed co-creation and evolutionary inference of historical scenes through blockchain technology. The ubiquitous learning framework proposed by Hwang and Chang in 2011 provides a theoretical reference for the migration of this system to mobile devices [16].This research provides a trinity innovation paradigm of "technology-theory-practice" for the digital transformation of education, which is expected to reshape the cognitive experience dimension of cultural inheritance.

Acknowledgement

2024 Undergraduate Innovation and Training Program Project at Zhuhai College of Science and Technology(S202413684018).

References

1. Li Lin, Zhang Wei, (2023). An empirical study of cognitive fault in ancient poetry teaching for middle school students: A case study of "Quiet Night Thinking". Chinese Journal of Education, 45(3), 56-62.
2. Department of Basic Education, Ministry of Education. (2024). A survey report on the current situation of Chinese Middle School students' Poetry learning in 2024. Beijing: People's Education Press.
3. Wang Qiang, Chen Min. (2022). The impact of visual stimuli on metaphor comprehension of ancient poetry: a controlled experiment based on a primary school in Hangzhou. Educational Technology Research, 38(5), 89-95.
4. ISTE. (2023). Global AR/VR Education Technology Report 2023. International Society for Technology in Education.
5. Smith,J.,&Lee, H. (2021). Embodied Interaction in Poetry Learning: A Case Study of MIT's AR Project. Journal of Educational Technology, 67(2), 112-125.
6. Zhao Q, Zhou Tao, (2022). Research on the homogenization of interaction model in AR educational application. Modern Educational Technology, 32(7), 45-51.
7. Liu Fang, et al. (2021). The impact of mobile applications on traditional culture cognition: A case study of a poetry APP. Digital Education Research, 19(4), 78-84.
8. Poetry by Heart. (2020). Enhancing Historical Empathy through AR-Based Poetry Learning. Journal of Cultural Heritage Education, 12(1), 33-47.
9. Godoy Jr, C. H. (2021). Augmented reality for education: A review. arXiv preprint arXiv:2109.02386.
10. Godoy Jr, C. H. (2021). Augmented Reality and Gamification: A Framework for Developing Supplementary Learning Tool. arXiv preprint arXiv:2108.03487.
11. Graser, S., & Böhm, S. (2024). A systematic literature review on technology acceptance research on augmented reality in the field of training and education. arXiv preprint arXiv:2411.13946.
12. Godoy Jr, C. H. (2021). Developing an Augmented Reality-Based Game as a Supplementary tool for SHS-STEM Precalculus to Avoid Math Anxiety. arXiv preprint arXiv:2109.09336.
13. Haq, I. U. (2023). Impact of Augmented reality system on elementary school ESL learners in country side of china: Motivations, achievements, behaviors and cognitive attainment. arXiv preprint arXiv:2309.09894.

14. Czok, V., Krug, M., Müller, S., Huwer, J., & Weitzel, H. (2023). Learning effects of augmented reality and game-based learning for science teaching in higher education in the context of education for sustainable development. Sustainability, 15(21), 15313.
15. Kapp, K. M. (2012). The gamification of learning and instruction: Game-based methods and strategies for training and education. John Wiley & Sons.
16. Hwang, G. J., & Chang, H. C. (2011). A review of definitions for mobile and ubiquitous learning. Journal of Educational Technology & Society, 14(2), 3-18.