Multimodal Technologies and Interaction

MDPI

Article

AI-Enhanced Motion Capture for Multimodal Interaction in Chinese Shadow Puppetry Heritage

Gaihua Wang [1,2,*], Hengchao Yun [1,2,*], Lixin Yang [1,2], Qingyuan Zheng [2] and Tianmuran Liu [2]

[1] School of Digital Arts, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; lxy_@stu.xupt.edu.cn
[2] Shaanxi Provincial Key Laboratory of Intelligent Media, Xi'an 710121, China
* Correspondence: ghw@xupt.edu.cn (G.W.); yunhengchao@stu.xupt.edu.cn (H.Y.)

Abstract

This study examines how AI-enhanced motion capture (AI-MoCap) mediates the preservation, transmission, and re-creation of Chinese shadow puppetry as performative intangible cultural heritage. Through a state-of-the-art review and comparative analysis of three representative application models—technology-driven, culturally integrated, and entertainment-oriented—the paper explores how AI-MoCap supports the digitization of performative techniques while reshaping modes of cultural presentation and interaction. Cross-case comparison highlights recurring tensions between technical standardization and cultural authenticity while also indicating possibilities for symbolic reconstruction, contextual continuity, and ethically grounded design. Based on this comparison, the paper develops a dual-channel inheritance framework—"perception–symbol" and "design–performance"—and treats cultural resolution and digital ethics as analytical and normative principles for resisting algorithmic homogenization. Rather than functioning only as a digitization tool, AI-MoCap can be understood as a mediating mechanism whose cultural value depends on how it remains embedded in community-based performative logics, symbolic systems, and ethical boundaries. The resulting framework offers transferable guidance for future research, curation, training, and policy discussion in the digital safeguarding of performance-based heritage.

Keywords: multimodal interaction; motion capture; Chinese shadow puppetry; intangible cultural heritage; digital heritage

1. Introduction

Under the dual impact of globalization and digital-intelligent transformation, China's intangible cultural heritage (ICH)—such as Cantonese Opera, Nuo Opera, and Shadow Puppetry—is facing a severe survival crisis [1,2]. Challenges include a shrinking audience base, generational discontinuity in craftsmanship transmission, and a growing disconnection between cultural forms and the intelligent society. As one of the vital carriers of traditional Chinese culture, shadow puppetry possesses significant artistic value and cultural connotation [3–5]. An urgent need is to establish a multimodal and dynamic inheritance system where the virtual and the real coexist, through a deep integration of emerging technologies and traditional craftsmanship [6–8].

The rapid development of Artificial Intelligence (AI) and Motion Capture technologies offers new possibilities for the digital preservation and dissemination of ICH. Motion

Check for updates

Academic Editor: Qianling Jiang

Received: 12 March 2026
Revised: 17 April 2026
Accepted: 23 April 2026
Published: 28 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

capture technology records the performer's skeletal point data in real time, converting traditional techniques into storable, repeatable, and interactive digital assets. This process injects technical momentum into cross-media communication, educational transmission, and creative reinterpretation of ICH [3,4,6,8].

Looking at the historical trajectory of ICH digital preservation, the field has transitioned from static image documentation (e.g., UNESCO's Memory of the World Programme) and database construction to paradigms involving dynamic capture and virtual reality technologies [6–8]. Researchers have made significant progress in the technical dimension [5,7,9]. For instance, they proposed a theoretical framework for building a motion capture-based ICH movement database and empirically validated the technological efficiency gains through digitalization projects. These efforts advance the research frontier toward high-precision motion reproduction and algorithmic optimization [10].

Recent studies have highlighted a growing trend toward interdisciplinary integration, making the intersection of heritage studies with computer science and communication studies a research hotspot [2,7,8]. However, beneath the radical narrative of technology empowerment lie two emerging dilemmas:

1. The ongoing value conflict between technological adaptability and cultural authenticity. Proponents emphasize the advantages of transcending spatiotemporal limitations, while critics warn against the risks of performative stylization resulting from over-technologization.
2. Ethical dilemmas in technological implementation are becoming increasingly prominent, including the lack of standardized data protocols, varying levels of technological acceptance among cultural bearers, and instances of cultural appropriation in commercial applications.

These dual dilemmas expose a theoretical blind spot in current research: how to protect cultural subjectivity and construct sustainable development mechanisms. At their core, these debates point to a fundamental question: how can a dynamic balance be built between instrumental rationality and value rationality in the irreversible process of digital-intelligent transformation [1,2,7,11]? This question forms the central proposition and logical thread of this review.

Building on this concern, the present study does not simply ask whether AI-enhanced motion capture can be applied to the digitalization of intangible cultural heritage. Rather, it asks how AI-enhanced motion capture mediates the preservation, transmission, and recreation of Chinese shadow puppetry as a form of performative intangible cultural heritage; what kinds of opportunities and risks emerge across different application pathways; and what analytical dimensions and practical directions these cases may offer for future research. Accordingly, this paper is positioned as a state-of-the-art review combined with comparative case analysis and theoretical synthesis, rather than as a fully validated original empirical study. By comparing three representative application models—technology-driven, culturally integrated, and entertainment-oriented—the paper aims to clarify key issues concerning cultural fidelity, technical simplification, contextual continuity, and ethical boundaries, and develop an extensible analytical framework for the digital safeguarding of performance-based heritage. The overall analytical workflow of the study is summarized in Figure 1.

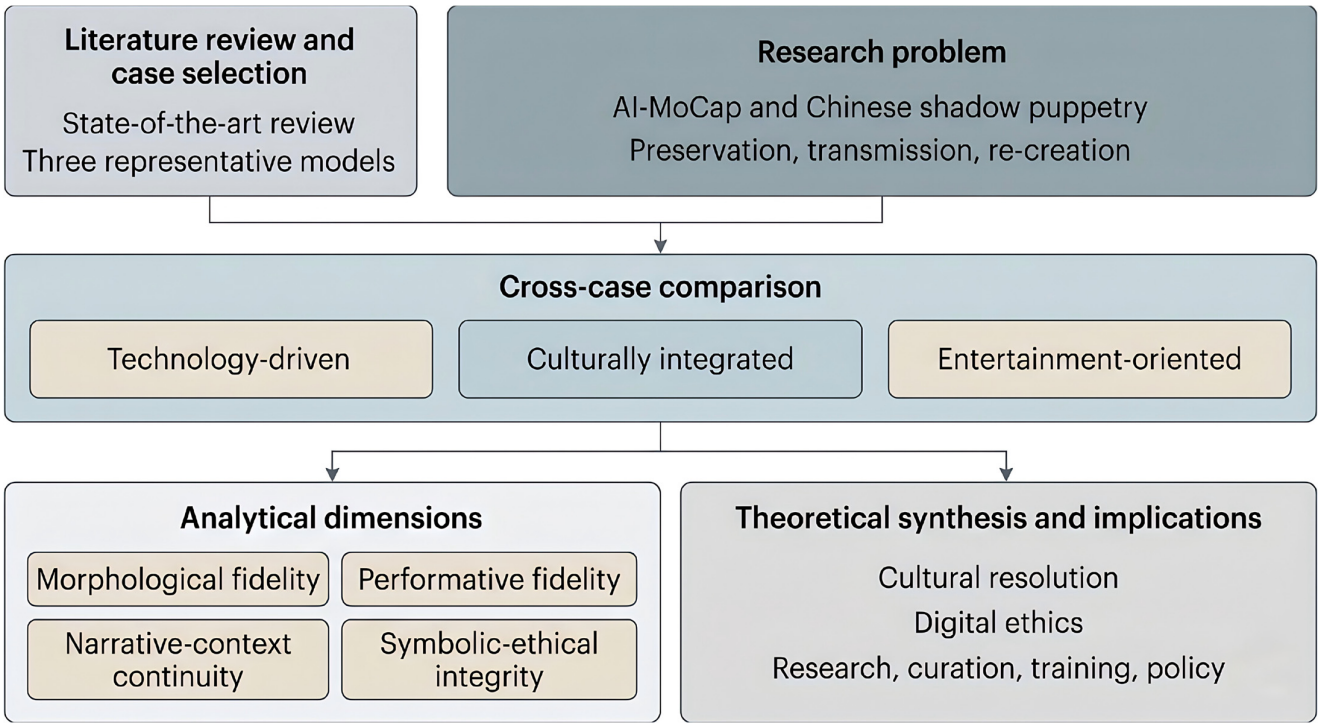


Figure 1. Conceptual and analytical workflow of the study.

2. Overview of AI-Enhanced Motion Capture Technology

2.1. Technical Principles and Classification

AI-Enhanced Motion Capture (AI-MoCap) is a composite technology that integrates artificial intelligence algorithms with motion capture systems [7,9,12]. The system aims to digitally model and dynamically reconstruct human motion through high-precision data acquisition and intelligent analysis. Its core principles can be divided into three components:

First, motion data acquisition:

The system uses multimodal sensors (optical cameras, inertial measurement units, and depth sensors) to capture and analyze motion data [10,12,13]. The system captures the performer's skeletal trajectories, joint angles, and muscle dynamics in real time.

Optical Motion Capture reproduces millimeter motion through infrared markers and multi-camera collaboration [14,15].

Inertial Motion Capture relies on wearable IMU devices and is suitable for flexible data collection in markerless environments [12,13,16].

Markerless Motion Capture, based on computer vision, directly extracts human posture from video streams using deep learning algorithms (e.g., Open Pose, Media Pipe), significantly reducing equipment dependency and operational thresholds.

Second, AI algorithmic intervention:

Convolutional Neural Networks (CNNs) eliminate environmental noise and anomalous data, enhancing the smoothness of motion trajectories [17].

Researchers then apply temporal models (e.g., LSTM, Transformer) to identify culturally meaningful motion units within the sequence, such as core techniques in shadow puppetry like "thread manipulation" (挑线) and "shadow transition" (转影)—thus enabling semantic archiving of the motion library [17,18].

Finally, the captured physical motion data is transformed into control instructions for virtual avatars or mechanical devices, enabling real-time interaction in mixed reality environments [5,19].

Third, practical application scenarios:

The incorporation of motion capture technology and AI algorithms enables more diverse presentations and interactive experiences of the digitized motion data:

1. Application in virtual performances within VR/AR environments (e.g., the Palace Museum's "Digital Shadow Puppetry Theatre" project) [8,20].
2. Enhancing user engagement through immersive and intuitive motion-capture-based interactive experiences that offer real-time emotional feedback to participants [8,21].
3. Enabling multi-user interactive scenarios powered by embedded AI algorithms expands the expressive boundaries of intangible cultural heritage (ICH) art forms.

As shown in Table 1, AI-MoCap technology can be classified into three categories according to different ICH craftsmanship characteristics and preservation needs [22].

Table 1. Classification of Motion Capture Technologies and Their Applicability to Intangible Cultural Heritage (ICH) Scenarios.

Type	Core Technology	Advantages	Limitations	Application Scenarios
Optical Motion Capture	Infrared markers + multi-camera tracking	High precision, low error rate, strong data integrity	High equipment cost, requires a fixed setup	Immersive experience and public-oriented teaching of Cantonese Opera
Inertial Motion Capture	Wearable IMU sensors	High portability, supports outdoor dynamic capture	Accumulated errors over time require periodic calibration	Field recording of ritual dances such as Nuo Opera
Vision-based Motion Capture	Deep learning + mono-/multi-view vision	Low cost, non-intrusive, easily scalable	Sensitive to lighting, complex movements may lose detail	Construction of shadow puppetry motion libraries

2.2. Technological Evolution and Key Breakthroughs

2.2.1. Transformation from Film Entertainment to Cultural Heritage Preservation

Researchers and engineers initially developed motion capture technology (MoCap) for use in modern entertainment industries such as film and video games, which served to create realistic animated characters and special effects [14]. In 1973, the predecessor of the British company Vicon—Oxford Metrics—developed optical motion capture technology for remote sensing and control in the Royal Navy. Building on this, MIT developed a system in 1983 called the "graphic marionette," which used LEDs and multiple cameras to track luminous markers, laying the technological foundation for optical MoCap.

In 1984, Vicon officially established itself and released its first commercial optical motion capture system. The film industry later adopted this technology. In 1995, the film Titanic used Vicon's system to capture extras' movements to generate large-scale crowd animations. In 2009, the film Avatar employed head-mounted cameras and real-time motion capture systems to transform actors' performances into the virtual "Na'vi" characters, ushering in a technological revolution in cinema. Since then, MoCap applications have gradually expanded from the entertainment industry into intangible cultural heritage (ICH), marking a significant transformation over the past decade.

2.2.2. Integration of AI Algorithms

Traditional MoCap systems have relied heavily on hardware precision and manual post-processing, leading to high costs and limited flexibility. Integrating AI algorithms has restructured the technical implementation logic and introduced notable breakthroughs:

First, enhanced accuracy and naturalness. Traditional optical MoCap systems depend on marker recognition and often struggle to accurately capture intricate movements such as thread manipulation, shadow vibration, thread kneading, and thread looping (挑线、抖影、揉线、绕线). Embedded AI algorithms now employ convolutional neural networks (CNNs) and spatio-temporal attention models (such as Transformers) to recognize unmarked body movements, thereby reducing errors and improving fidelity.

Second, reduced dependence on specialized hardware. With AI integration, MoCap systems leverage computer vision and edge computing to optimize recognition algorithms. Even standard RGB cameras can perform real-time motion capture. Compared to traditional hardware setups costing hundreds of thousands of RMB, AI-based MoCap significantly reduces implementation costs, greatly enhancing accessibility. This democratization of technology enables a broader range of users to engage with the Chinese ICH through immersive experiences.

Finally, interactive innovation through AI. By incorporating reinforcement learning (RL) and generative adversarial networks (GANs), AI-powered MoCap systems can go beyond replication to enable creative interactions and novel motion generation, thus expanding the expressive possibilities of traditional art forms.

3. Current Applications of AI-Based Motion Capture in Intangible Cultural Heritage (ICH)

AI-enhanced motion capture technology has offered a potential solution to the crisis of "lost skills due to the loss of masters" in ICH preservation. However, the inherent contradiction between technical standardization and the unique characteristics of traditional art forms gives rise to a paradox: the deconstruction of craftsmanship and the superficialization of culture. Three major application pathways—technology-driven, culturally integrated, and entertainment-oriented—respectively reflect ongoing efforts in data homogenization, the codification of aesthetic principles, and the enhancement of communication efficacy.

While the technological approach faces constraints from algorithmic simplification, the cultural paradigm focuses on the digital transcreation of traditional skills, and the entertainment pathway, though effective in audience expansion, risks diluting cultural depth. This chapter explores the tension between ontological preservation and innovation in ICH digitization through critical case studies, aiming to uncover the underlying logic of cultural continuity in the age of intelligent technologies.

To make the concept of cultural resolution more analytically usable, the following case comparison is interpreted through four dimensions: morphological fidelity, referring to the extent to which traditional component structure, character configuration, and visual style are preserved; performative fidelity, referring to the preservation of manipulation logic, movement language, and performance rhythm; narrative-context continuity, referring to whether the cultural background, story matrix, and use context remain sustained; and symbolic-ethical integrity, referring to whether symbolic meaning, value structure, and cultural boundaries are maintained under technological translation and commercialization. Taken together, these dimensions provide the comparative framework through which the three application models are discussed in the following sections.

3.1. Case Studies and Critical Analysis

3.1.1. Technology-Driven Approach

Among notable applications of motion capture in ICH preservation, the "Shadow Story" project—jointly initiated in 2011 by the Intelligent Engineering Laboratory of the Institute of Software, Chinese Academy of Sciences (ISCAS), Microsoft Research Cambridge, Pennsylvania State University, and the National Key Laboratory of Computer Science at ISCAS—has attracted widespread attention. The research team designed the project in response to the dual threats of globalization and urbanization, particularly the disappearance of local cultures and the growing disconnection of children from their cultural heritage. By developing interactive technologies inspired by traditional arts, "Shadow Story" enables children to engage with and learn about regional cultural heritage through playful experiences, opening new avenues for ICH preservation.

"Shadow Story" focuses on bringing traditional Chinese shadow puppetry into the digital age. By leveraging motion capture technologies and algorithmic processing, performers' movements are mapped onto virtual shadow-puppet characters, offering users an interactive experience of shadow play. However, the project also reveals significant limitations: its simplified seven-component model (head, torso, four limbs, and hands) compresses the multi-component articulated structure of traditional shadow puppetry into a much more schematic configuration. As shown in Figure 2, a traditional Chinese shadow puppet is composed of multiple interconnected parts rather than a simple flat silhouette. Such structural reduction increases the risk of formulaic and stylized performances and weakens the performative fidelity of the form. Moreover, the lack of a standardized motion database contributes to algorithmic homogenization, eroding stylistic differences across shadow-puppet traditions and reducing cultural diversity and performative nuance in the captured motions.

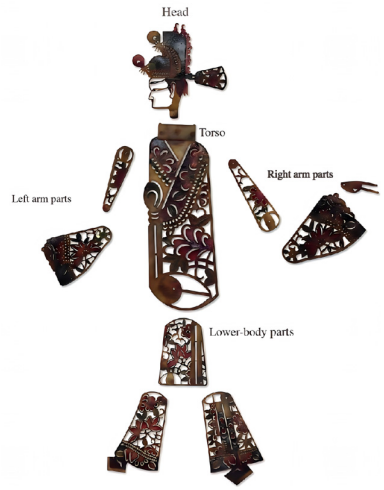


Figure 2. Decomposition diagram of a traditional Chinese shadow puppet.

Additionally, the system requires users to interact through handheld controllers, which hinders natural interaction and prevents users from fully immersing themselves in the performance. This compromises the intimate connection between the puppeteer and the puppet, central to traditional shadow play.

From a cultural perspective, although "Shadow Story" successfully introduces shadow puppetry to the digital domain, it remains at the level of surface representation. It fails to deeply explore the rich folklore, symbolic meanings, and regional characteristics embedded in the art of shadow puppetry, leading to a shallow presentation of cultural content. In terms of narrative structure, although basic interactive storytelling is supported,

the plot design is simplistic, character motivations are vague, and the integration of cultural context appears forced. It lacks the intricate story arcs and refined character development typical of traditional shadow plays, making it difficult for users to fully grasp the genre's cultural essence and artistic value.

Nevertheless, "Shadow Story" remains a useful reference case for understanding both the enabling potential and the simplifying risks of technology-mediated heritage interaction.

3.1.2. Cultural Integration Model

At a university in southern China, the AI Shadow Puppet Robot Project embodies the cultural tension between "Dao" and "Qi" (principle and technique) beneath its technical surface. Drawing inspiration from the aesthetic doctrine of "regularity in motion and stillness" (from the Book of Rites), the project team encoded the traditional shadow puppetry philosophy of "60% carving, 40% performance" into algorithmic rules. By employing Long Short-Term Memory (LSTM) models to analyze temporal sequences of techniques such as "lifting," "twisting," and "shaking," robotic arm movements follow the narrative rhythm of qi-cheng-zhuan-he (introduction, development, turn, and conclusion). This effort to "embody Dao through technique" is exemplified in the new adaptation of Nezha Conquers the Dragon King, where the AI inherits the Chaozhou shadow puppetry tradition of "performing civil scenes with martial skills." Specifically, the robot arm deliberately reduces its movement speed during fight scenes in order to preserve the aesthetic beauty of the "warrior-horseman" silhouette.

However, the lack of agility revealed by the MIT laboratory strikes at the core dilemma in the inheritance of Chinese intangible cultural heritage (ICH): when technology deconstructs "body rhythm" (as in the "suspended string pulse" of Quanzhou marionettes) into skeletal coordinate points, how can the intuitive knowledge or "Xinfa" (heart-method)—where the master's hand follows the mind—be digitized? The Suzhou Intangible Cultural Heritage Pavilion offers a solution rooted in Eastern wisdom: its VR puppet system incorporates a "qi-yun" interaction value intended to simulate vitality and expressive rhythm. Users must regulate their operating rhythm via a breathing sensor, and only when their operating rhythm approaches that of the master performer can they unlock more advanced techniques. This interaction design, which advances from technique to principle, aligns with the ancient Chinese craft philosophy from Kaogong Ji: "Heaven has its timing, Earth has its energy, materials have their beauty, and craftsmanship has its ingenuity."

The digital museum of Lufeng shadow puppetry illustrates the broader conflict between "form" and "function" in the digital era. While its 4K exhibits accurately replicate the craftsmanship of "hundred-window carving" (form), they fail to reconstruct the ritualistic context of "ancestor worship in the dancing shadows of the window" (function). This disconnect is especially prominent in southern Fujian's clan-based culture, where shadow puppetry traditionally served as a ritual instrument for winter solstice ancestral rites. The Quanzhou project achieved a breakthrough by integrating blockchain technology to link genealogical data with digital puppets: when users access the Chen San and Five Daughters performance, the system automatically connects to 3D scans of the Chen clan ancestral hall in Quanzhou, and full viewing is unlocked only when "ancestral worship points" meet the threshold. This construction of "digital ritual order" not only sustains the function of "art as a vehicle for morality," but also appears to encourage greater youth participation in digital heritage contexts.

More profoundly, this cultural fusion extends to the ethical dimension. When Baidu's Wenxin Yiyan AI generates new scripts for Nezha, the system enforces that a substantial

proportion of lines retain tonal patterns from Chao Opera's "double rhyme couplets," and emotional computation ensures that the AI-generated worldview of loyalty and filial piety remains broadly aligned with the ethical framework of Zhuzi Jia Li. This "algorithmic philology" practice is actively reshaping cultural continuity in the digital age.

Such in-depth cultural integration demonstrates China's unique wisdom in the digital preservation of ICH and offers new perspectives and methods for global heritage protection. These projects safeguard heritage while promoting cultural innovation by combining traditional philosophy with modern technology. The ability to balance tradition and modernity, technology and culture, constitutes the core competitive edge of China's ICH digital preservation efforts.

3.1.3. Entertainment-Driven Integration Model

Among cases involving motion capture (MoCap) technology in ICH preservation, the adventure puzzle game Projection: First Light, developed by Canadian studio Shadowplay Studios and published by Blowfish Studios, stands out. Released in 2020 across Steam, GOG, Xbox, PlayStation, and Nintendo Switch, the game aims to introduce shadow puppetry to players through interactive entertainment, enabling them to understand and appreciate this ancient art form during gameplay, paving a novel path for ICH preservation.

Projection: First Light integrates traditional shadow puppetry with modern game technology, showcasing puppet traditions from Indonesia, China, Turkey, and 19th-century England. By leveraging advanced MoCap technology and algorithms, the movements of puppet characters are accurately captured and converted into virtual animations, offering players an immersive experience of shadow play across diverse cultures. However, the game also reveals significant limitations: simplified character models cannot faithfully replicate the intricate structure of traditional puppets, increasing the risk of formulaic movements. The absence of a motion database leads to homogenization of distinct styles, resulting in reduced cultural diversity within the motion library. Furthermore, the game requires continuous handheld controller interaction, which limits natural user engagement and undermines the authentic connection between performer and puppet that defines traditional shadow puppetry.

In terms of cultural content, although Projection: First Light successfully brings multinational shadow puppet art into the digital domain, it largely remains at the level of surface representation. The game fails to deeply explore the rich folklore, symbolism, and regional identities behind the art, resulting in a superficial portrayal of its cultural significance. From a narrative perspective, while the game supports basic interactive storytelling, it features simplistic plotlines, vague character motivations, and awkward integration of cultural backgrounds. The lack of cohesive plots and nuanced character development typical of traditional shadow puppetry limits the user's ability to grasp the essence and artistic value of the form.

Nevertheless, Projection: First Light contributes valuable experience to the digital preservation of ICH by attracting younger audiences to shadow puppetry. It lays a potential foundation for dual-mode dissemination, combining entertainment and education. As UNESCO's White Paper on Digital Intangible Cultural Heritage asserts, when technological mediation becomes irreversible, the key lies in building "innovation channels with controllable cultural loss."

4. Innovative Practices of AI Motion Capture Technology in Intangible Cultural Heritage

4.1. Specific Pathways of Technological Empowerment

In preservation of intangible cultural heritage (ICH), AI-based motion capture (MoCap) technology has opened up new pathways for capturing and digitally modeling traditional techniques.

In motion capture, multimodal sensors enable the precise acquisition of movement data from various parts of the performer's body. Optical motion capture, known for its high accuracy, utilizes infrared markers and multi-camera setups to accurately reconstruct the complex trajectories of movements such as "thread lifting" and "shadow turning" performed by shadow puppetry artists. Researchers can transform these subtle manual gestures used in traditional puppetry into digital motion data with high fidelity. Visual motion capture, empowered by deep learning algorithms, can extract human postures from ordinary video footage, reducing dependence on specialized hardware. The simplified data collection process makes it easier for more folk artists to participate in motion data collection, especially in crafts such as shadow puppetry. It facilitates broader inclusion in preserving intangible cultural heritage (ICH).

Before motion capture can be applied, digital modeling is required as a foundational step in technological empowerment. Modeling must be grounded in in-depth research into the ICH technique, including its movement features, performance styles, and embedded cultural meanings. This analytical foundation informs the model architecture and parameter configuration. Shadow puppetry serves as a representative example for demonstrating the modeling process. The modeling process must fully consider the component structure of shadow figures, joint range of motion, and manipulation methods. The goal is to construct virtual models that accurately reflect the performance characteristics of shadow puppetry. Figure 3 shows the stepwise digital reconstruction of a shadow-puppet character from decomposed traditional parts.

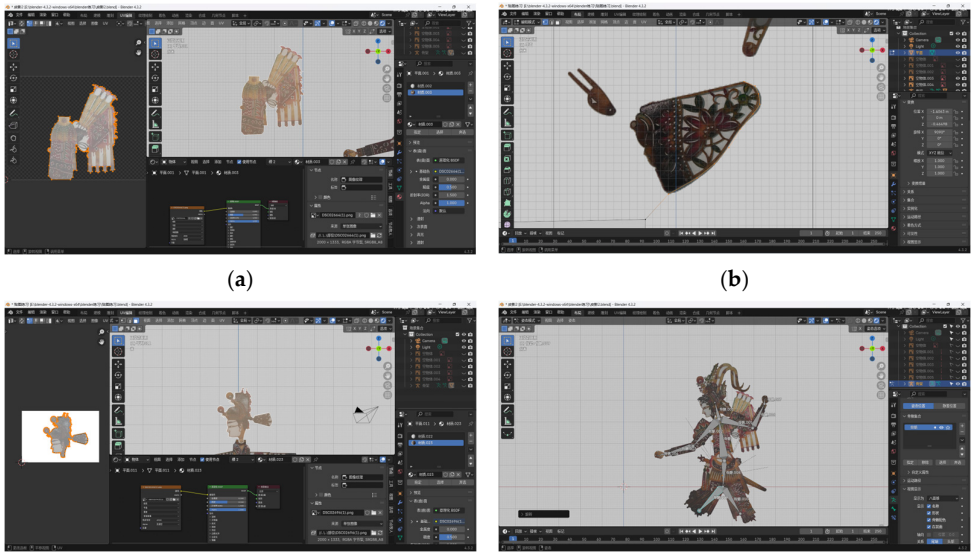


Figure 3. Stepwise digital reconstruction of a shadow-puppet character from decomposed traditional parts. (**a**) Initial import and alignment, (**b**) component processing, (**c**) structural assembly, (**d**) completed character model for later engine import and animation. Non-English interface labels appearing in the software screenshots are standard software menu and interface terms and do not affect the interpretation of the reconstruction process.

https://doi.org/10.3390/mti10050046

As shown in Figure 3, the modeling process begins with the digital import of decomposed puppet parts and proceeds through component-level adjustment toward the structural reconstruction of a complete digital character. This stepwise process is important because it links the component structure of traditional shadow figures to later considerations of joint range of motion, manipulation logic, and system-driven animation. After this modeling stage, the reconstructed character is imported into the engine, where captured motion data are transformed into usable digital assets. AI algorithms then play a critical role in analyzing, processing, and optimizing vast and often noisy datasets. Convolutional Neural Networks (CNNs) filter out environmental noise and anomalous data, smoothing the motion trajectories and ensuring that gestures appear natural and continuous, avoiding distortions or interruptions caused by external disturbances. Temporal models such as LSTM and Transformer are applied to mine the motion sequences deeply, identifying embedded cultural semantic units. These identified units are then archived into a motion library, enabling the systematic and digital representation of intangible cultural heritage (ICH) techniques, thus laying a solid foundation for their preservation and dissemination. In this sense, Figure 3 serves not merely as a technical workflow, but as a visual explanation of how traditional shadow-puppet structure is translated into a digitally operable character model.

This technical approach allows for the precise recording of detailed motion elements in ICH practices and converts them into digital formats that are easy to store, transmit, and recreate. In doing so, ancient intangible cultural heritage gains renewed vitality in the digital age, offering immense potential for future protection, transmission, and innovative development. The following section turns to representative cases in order to examine how these technological pathways have been interpreted, applied, and discussed in shadow puppetry heritage contexts.

4.2. Comparative Analysis of Technology-Enabled Shadow Puppetry Cases

4.2.1. The Shadow Story Digital Narrative System: A Case Analysis of Wireless Sensor Interaction

This case centers on shadow puppetry as a form of ICH and involves developing the digital narrative system Shadow Story. Motion mapping is realized through wireless handheld directional sensors (WiTilt v3.0; SparkFun Electronics, Boulder, CO, USA), reducing traditional puppet control to simple wrist movements. The system adopts a dual-mode framework of "Design" and "Performance": in Design Mode, users simulate the shadow puppet carving process via a tablet interface using virtual engraving tools, digital brushes, and traditional pattern stamps, thus lowering the barrier to creative entry. In Performance Mode, coordinated projectors and multiple sensors enable real-time collaborative manipulation of digital characters. The system includes a built-in video library of traditional shadow puppet performances, forming a closed creative loop of "imitation–reconstruction–innovation."

The source study reported a 7-day embedded field trial at an elementary school in Beijing, involving 36 children aged 7–9 [4]. Rather than being treated here as a fully standardized empirical validation, the case is used as published evidence showing how digitally mediated shadow puppetry interaction may support children's participation, creative engagement, and initial cultural familiarity. The available case materials mainly include field observation, video records, semi-structured interviews, and documented creative outputs such as stories, characters, props, and backdrops. For the purposes of this paper, these materials are interpreted comparatively as indications of how technology can lower participatory barriers while reconfiguring the relationship between interaction design and cultural transmission.

From a comparative perspective, the Shadow Story case suggests that digital interaction can lower the operational threshold of shadow puppetry participation while retaining part of its "design–performance" logic in a simplified interface. Rather than preserving the full cultural and performative complexity of traditional shadow puppetry, the system appears to work more effectively as an introductory and engagement-oriented medium through which children can encounter selected symbolic elements, collaborative storytelling processes, and basic performative structures. In this sense, the case is analytically useful not because it provides definitive proof of transmission effectiveness, but because it reveals both the enabling potential and the simplifying risks of technology-mediated heritage interaction.

More cautiously understood, this case suggests a possible pathway through which digital interaction may extend the visibility and accessibility of shadow puppetry in new media contexts. However, such extension should not be equated with the full preservation of cultural ontology; rather, it remains a partial and selective form of transmission shaped by the trade-off between accessibility, structural simplification, and cultural fidelity.

4.2.2. The Kinect-Based Digital Shadow Play Interaction System: A Case Analysis

This case focuses on the intangible cultural heritage of shadow puppetry and examines the "Virtual Puppet" digital experience system [23]. The Microsoft Kinect V2 (Microsoft Corporation, Redmond, WA, USA) device is used to transform traditional shadow puppetry manipulation into a motion-sensing interactive experience. The system adopts a "single-player–multi-player" dual-mode framework: in single-player mode, the user controls a virtual puppet independently, whereas in multi-player mode, several users can participate simultaneously by controlling different puppet characters. The system includes twelve virtual puppets based on classic *Journey to the West* characters, creating an "interaction–creation" experience.

The system uses Microsoft Kinect V2 together with the official 2.0 SDK to process motion data and map it within a Unity 2D environment, thereby driving 2D virtual puppet movement and enabling human–computer interaction. It also introduces Visual Gesture Builder to construct a gesture database for improved interactivity. Data are recorded using Kinect Studio v2.0 (Microsoft Corporation, Redmond, WA, USA), and the resulting gesture database supports the recognition of more complex user actions within the interactive system.

The source study reported a two-stage process at the Fo Guang Shan Foundation in Philadelphia, combining user play-testing with subsequent expert review. In the present paper, this case is not treated as a fully standardized validation study; rather, it is used as a published example showing how Kinect-based interaction can reframe shadow puppetry as a participatory digital experience in educational and display contexts. The available materials are most usefully understood as play-testing observations, system descriptions, and expert-panel interpretations, which together provide indicative rather than fully generalizable evidence for cross-case comparison.

In addition to play-testing, the source case also incorporated expert interpretation from the perspective of human–computer interaction and heritage transmission. For the purposes of this review, the value of this case lies less in any single metric than in the broader pattern it reveals: Kinect-based interaction can make shadow puppetry more accessible, collaborative, and physically engaging in public-facing settings, but it may also simplify cultural depth, narrative context, and traditional performative logic. From a comparative perspective, this case is therefore useful as an example of how technology may broaden participation while simultaneously raising questions about contextual continuity, symbolic fidelity, and the limits of interaction-driven reinterpretation.

4.3. Parametric Modeling-Driven Digital Shadow Puppetry Interaction System: A Conceptual and Technical Analysis

This section presents a parametric-modeling-based digital shadow puppetry interaction system as an illustrative technical example for discussing how structural reconstruction and motion constraints may support the digital translation of shadow puppetry [24]. Rather than being treated as an independently validated empirical case, it is used here to clarify the technical logic and conceptual implications of parametric modeling for heritage-oriented interaction design. Its technical process can be summarized through a "Classify–Control–Simulate" framework:

Character Classification and Template Engine: The digitally modeled shadow puppets are classified into four prototype categories (human-shaped, winged biped, quadruped, and serpentine animals), and a parametric geometric model library is established. A dynamic texture mapping algorithm (D-TMA) is developed to support real-time adaptation of artist-input materials and component size precision adjustments.

Gesture Control System: Leap Motion devices are used to capture hand movements at 200 Hz, with three control point mapping rules defined as shown in Table 2:

Table 2. Control-Point Mapping Rules for Different Shadow Puppet Types.

Shadow Puppet Type	Number of Control Points	Mapping Strategy
Human	5	Palm movement → torso, finger bending → head/limbs
Winged Bipedal Animal	4	Palm tilt → wing extension, finger spacing → leg movement
Serpentine Animal	3	Hand trajectory → spine curve fluctuation

Through the inverse kinematics optimization algorithm (IK-Opt), hand movements are converted into the joint movement angles of the shadow puppetry figure.

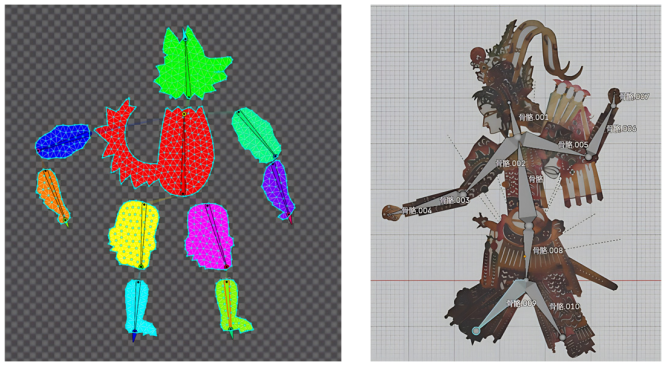
Physical Constraint Simulation Engine: An inertial dynamics model is constructed in Unreal Engine 4 to simulate the swing damping of the shadow puppet components. At the same time, digital boundary conditions for baffles are introduced to restore the physical constraints of leg movements in traditional performances.

From an analytical perspective, this example is useful for illustrating a "deconstruction–recon-struction" logic in the digital translation of shadow puppetry. Parametric modeling, control-point mapping, and physical-constraint simulation make it possible to preserve certain structural relations, articulation rules, and motion-control correspondences of shadow puppetry within a programmable environment. At the same time, such technical reconstruction should not be equated with the full preservation of cultural ontology. Ornamentation, regional style, symbolic density, and narrative context may still be simplified in the process of digital abstraction. For this reason, the value of this example in the present paper lies less in demonstrating completed heritage transmission than in clarifying the technical possibilities and cultural limits of parametric reconstruction.

More cautiously understood, this example is most valuable as a technical illustration of how parametric modeling may contribute to morphological fidelity and, to some extent, performative fidelity in the digital representation of shadow puppetry. By contrast, narrative-context continuity and symbolic-ethical integrity cannot be secured by modeling and control mechanisms alone, but require broader cultural design, contextual framing, and interpretive guidance. In this sense, the example helps clarify both the potential and the limits of technical reconstruction within the broader framework of cultural resolution.

To further clarify the first dimension of cultural resolution—morphological fidelity—Figure 4 provides an illustrative comparison between a traditional shadow puppet and a digitally reconstructed character from our self-built motion-capture shadow puppetry system. The comparison is intended not as an additional empirical case, but as a

visual explanation of how certain articulated organizational principles may be retained in digital translation while ornamentation, formal refinement, and regional stylistic detail may still be simplified.



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

Figure 4. Digitally reconstructed shadow-puppet-like character from the authors' self-built motion-capture shadow puppetry system: (**a**) decomposed digital character model; (**b**) articulated rigging structure. In panel (**a**), different colors distinguish segmented puppet components for assembly and motion mapping; in panel (**b**), lines and node symbols indicate the skeleton/rigging connections and joint-control relationships. Non-English interface labels appearing in the software screenshots are standard software menu/interface terms and do not affect the interpretation of the reconstruction and rigging process.

5. Challenges, Solutions, and Future Outlook

At the technical level, the comparative review suggests that current applications of AI-enhanced motion capture in intangible cultural heritage continue to face recurring tensions between accuracy, accessibility, portability, and interpretability [25–27]. On the one hand, highly instrumented capture systems may support more detailed motion reconstruction, but they often depend on fixed environments, costly hardware, and specialized workflows. On the other hand, lower-cost and more scalable approaches—especially vision-based systems—may broaden access, yet they can also simplify subtle movement detail, weaken stylistic differentiation, and complicate the semantic organization of heritage motion data. Across the cases discussed in this paper, a persistent technical challenge is therefore not only how to capture movement, but also how to organize, annotate, and interpret culturally meaningful motion units without collapsing regional and performative differences [17,21,22].

In light of these challenges, several technical directions may be identified for future exploration:

1. Hybrid capture strategies that combine the portability of inertial sensing with the lower-threshold advantages of vision-based motion capture may offer a more balanced approach to documentation across different heritage settings [13].
2. Greater attention should be given to ontology-informed metadata frameworks capable of describing not only spatiotemporal parameters, but also culturally meaningful motion semantics, so as to reduce the homogenization of cross-school and cross-regional performance data [17,21,22].
3. Future systems may further explore lightweight temporal models and edge-based interaction architectures in order to improve responsiveness in real-time heritage interaction, while remaining attentive to interpretability, deployment conditions, and cultural use contexts.

Cultural Challenges

The essence of the conflict in the digital protection of intangible cultural heritage lies in the dialectical relationship between technological innovation and the preservation of artistic authenticity [25–28]. When the granularity of technological deconstruction exceeds the self-consistency threshold of the cultural system, it can easily lead to the alienation of cultural genes and structural ruptures in the inheritance ecosystem. Based on comparative research, this paper summarizes the following cultural challenges that urgently need to be addressed:

1. While parametric modeling improves digital conversion efficiency, it may also contribute to the algorithmic dissolution of regional artistic characteristics [7,22].
2. Commercialization may lead to the decontextualization of intangible cultural heritage symbols. For example, simplifying Chinese shadow puppet patterns in the game "Projection: First Light" leads to a gap in user recognition.
3. The intergenerational digital divide may create a "digital breakpoint" in the inheritance chain, manifested by insufficient digital literacy in elderly artists and superficial cultural understanding among young practitioners.

In response to these cultural challenges, several future-oriented directions may be proposed:

1. A "heritage practitioner–technician–user" collaborative platform may provide a useful direction for future work. Rather than treating the technical intervention threshold as a fixed technical limit, future research may explore it as a culturally negotiated boundary—concerning what should be preserved, what may be moderately adapted, and what should not be altered solely for the sake of novelty—through sustained dialogue with heritage practitioners and comparative case discussion.
2. Future work may also explore the use of smart-contract or traceability-oriented frameworks to better coordinate digital records, performance parameters, and rights-related governance in heritage transmission contexts [29].
3. Another promising direction is the development of training-oriented digital systems that combine gesture recognition, multimodal feedback, and practitioner-informed verbal guidance in order to support apprenticeship, movement correction, and the transmission of embodied know-how.

6. Conclusions

Through a comparative reading of representative cases, this paper highlights the double-edged role of AI-enhanced motion capture in the digital safeguarding of intangible cultural heritage, especially in relation to both technological empowerment and the risk of cultural simplification [25,26]. Rather than offering direct empirical proof, the study develops "cultural resolution" and "digital ethics" as analytical lenses for interpreting how technological intervention may preserve, reshape, or weaken different dimensions of shadow puppetry heritage. From this perspective, three main comparative insights emerge:

1. Cross-case comparison suggests that technological intervention should remain responsive to the aesthetic density and performative complexity of the heritage form itself [7,25]. When digital systems over-simplify component structure, movement logic, or stylistic detail—as in the often-cited reduction in shadow-puppet configuration in Shadow Story—the result may be increased accessibility at the cost of cultural and performative fidelity.
2. The comparison also indicates that symbol embedding and context reconstruction may provide a promising pathway for renewing cultural engagement, especially when digital systems retain narrative background, symbolic coherence, and recog-

nizable cultural cues [21]. At the same time, such approaches remain vulnerable to commercialization, contextual detachment, and the risk of reducing cultural meaning to a surface-level aesthetic resource.

3. The comparison further highlights the importance of digital ethics as a governance-oriented dimension of heritage digitization. Issues such as consent, cultural boundaries, interpretive authority, and responsible reuse cannot be resolved by technical optimization alone but require localized ethical frameworks that remain sensitive to community knowledge, symbolic value, and the conditions of cultural transmission.

The study also points out current research limitations:

1. The case studies primarily focus on performance-related intangible cultural heritage, with insufficient exploration of the technical adaptability for craft-based intangible cultural heritage.
2. The application of neural-symbolic systems to the digital representation of tacit principles, embodied know-how, or "mental principles" remains largely conceptual and requires further case-based exploration.
3. The broader transferability of the digital ethics framework across different cultural traditions, heritage categories, and regional contexts still requires further comparative discussion.

Future research may further explore three directions:

1. Developing a "Perception–Symbol" dual-channel motion capture system combining LSTM and knowledge graphs to address the digitization challenges of "body rhythm" [17,18].
2. Building a global intangible cultural heritage action semantic network using multimodal fusion technologies to achieve cross-school technique translation [18,21,22].
3. Establishing a "Digital Cultural Community" based on the metaverse, exploring an "each beauty in its way" inheritance ecology in a decentralized structure.

The ultimate goal of technological empowerment is not to replace tradition but to reconstruct the vitality of intangible cultural heritage through digital civilization, allowing millennia-old techniques to breathe in algorithms and dance in code [26].

Author Contributions: Conceptualization, G.W. and H.Y.; methodology, H.Y.; software, H.Y.; validation, G.W. and L.Y.; formal analysis, Q.Z.; investigation, L.Y. and T.L.; data curation, L.Y.; writing—original draft preparation, G.W.; writing—review and editing, H.Y., L.Y., Q.Z. and T.L.; visualization, T.L.; funding acquisition, G.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Social Science Fund of China (Grant No. Z2024020) and the Xi'an University of Posts and Telecommunications Postgraduate Innovation Fund (Grant No. CXJJZW2024010). The APC was funded by the authors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article. The source materials supporting the analysis are publicly available in the references.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AI-MoCap	AI-enhanced Motion Capture
ICH	Intangible Cultural Heritage
MoCap	Motion Capture
VR/AR	Virtual Reality/Augmented Reality

References

1. Li, J.; Krishnamurthy, S.; Pereira Roders, A.; van Wesemael, P. Community Participation in Cultural Heritage Management: A Systematic Literature Review Comparing Chinese and International Practices. *Cities* **2020**, *96*, 102476. [CrossRef]
2. Yang, C.; Tengku Wook, T.S.M.; Rosdi, F. Advancing Cultural Heritage: A Decadal Review of Digital Transformation in Chinese Museums. *npj Herit. Sci.* **2025**, *13*, 189. [CrossRef]
3. He, Y. ShadowPlayVR: Understanding Traditional Shadow Puppetry Performance Techniques Through Non-Intuitive Embodied Interactions. In Proceedings of the VRST '23: Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology, Christchurch, New Zealand, 9–11 October 2023. [CrossRef]
4. Lu, F.; Tian, F.; Jiang, Y.; Cao, X.; Luo, W.; Li, G.; Zhang, X.; Dai, G.; Wang, H. ShadowStory: Creative and Collaborative Digital Storytelling Inspired by Cultural Heritage. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011), Vancouver, BC, Canada, 7–12 May 2011; pp. 1919–1928. [CrossRef]
5. Hu, Z.; Lin, M.; Liu, S.; Wang, M.; Hong, R.; Yan, S. eHeritage of Shadow Puppetry: Creation and Manipulation. In Proceedings of the 21st ACM International Conference on Multimedia (MM '13), Barcelona, Spain, 21–25 October 2013. [CrossRef]
6. Selmanović, E.; Rizvić, S.; Harvey, C.; Bosković, D.; Hulusić, V.; Chmeliar, J.; Okanović, V. Improving Accessibility to Intangible Cultural Heritage Preservation Using Virtual Reality. *J. Comput. Cult. Herit.* **2020**, *13*, 13. [CrossRef] [PubMed]
7. Hou, Y.; Kenderdine, S.; Picca, D.; Egloff, M.; Adamou, A. Digitizing Intangible Cultural Heritage Embodied: State of the Art. *J. Comput. Cult. Herit.* **2022**, *15*, 55. [CrossRef]
8. Yu, J.; Wang, Z.; Cao, Y.; Cui, H.; Zeng, W. Centennial Drama Reimagined: An Immersive Experience of Intangible Cultural Heritage through Contextual Storytelling in Virtual Reality. *J. Comput. Cult. Herit.* **2025**, *18*, 11. [CrossRef]
9. Mourot, L.; Hoyet, L.; Le Clerc, F.; Schnitzler, F.; Hellier, P. A Survey on Deep Learning for Skeleton-Based Human Animation. *Comput. Graph. Forum* **2022**, *41*, 122–157. [CrossRef]
10. Jia, W.; Wang, H.; Chen, Q.; Bao, T.; Sun, Y. Analysis of Kinect-Based Human Motion Capture Accuracy Using Skeletal Cosine Similarity Metrics. *Sensors* **2025**, *25*, 1047. [CrossRef]
11. Plemons, A.M.; Spiros, M.C. Toward Ethical Digital Practices: Guidelines for Consent, Accountability, and Transparency in Anthropology. *Am. J. Biol. Anthropol.* **2025**, *186*, e70044. [CrossRef] [PubMed]
12. Jiang, X.; Ibrahim, Z.; Jiang, J.; Liu, G. Motion Capture as an Immersive Learning Technology: A Systematic Review of Its Applications in Computer Animation Training. *Multimodal Technol. Interact.* **2026**, *10*, 1. [CrossRef]
13. Pan, S.; Ma, Q.; Yi, X.; Hu, W.; Wang, X.; Zhou, X.; Li, J.; Xu, F. Fusing Monocular Images and Sparse IMU Signals for Real-Time Human Motion Capture. In Proceedings of the SIGGRAPH Asia 2023 Conference Papers, Sydney, Australia, 12–15 December 2023; pp. 1–11. [CrossRef]
14. Guo, Y.; Zhong, C. Motion Capture Technology and Its Applications in Film and Television Animation. *Adv. Multimed.* **2022**, *2022*, 6392168. [CrossRef]
15. Hu, Z.; Tang, J.; Li, L.; Hou, J.; Xin, H.; Yu, X.; Bu, J. MarkerNet: A Divide-and-Conquer Solution to Motion Capture Solving from Raw Markers. *Comput. Animat. Virtual Worlds* **2024**, *35*, e2228. [CrossRef]
16. McInnes, M.; Chadwick, E.; Blana, D.; Starkey, A. Improving the Accuracy and Reliability of Upper Limb Inertial Motion Capture without Increasing Calibration Complexity. *Gait Posture* **2024**, *113*, 144–145. [CrossRef]
17. Zabulis, X.; Partarakis, N.; Manikaki, V.; Demeridou, I.; Dubois, A.; Moreno, I.; Bartalesi, V.; Pratelli, N.; Meghini, C.; Manitsaris, S.; et al. A Digitally Enhanced Ethnography for Craft Action and Process Understanding. *Appl. Sci.* **2025**, *15*, 5408. [CrossRef]
18. Du, D.; Ding, J.; Liu, Y. Knowledge Graph Construction of Chinese Embroidery Evolution Based on Associating Cultural Space and Critical Incidents under Intangible Cultural Heritage. *Electron. Libr.* **2025**, *43*, 283–302. [CrossRef]
19. Zabulis, X.; Stamou, A.; Demeridou, I.; Koutlemanis, P.; Karamaounas, P.; Papageridis, V.; Partarakis, N. Simulation and Visualisation of Traditional Craft Actions. *Heritage* **2024**, *7*, 7083–7114. [CrossRef]
20. Shen, X.S.; Gao, J.; Li, M.; Zhou, C.; Hu, S.; He, M.; Zhuang, W. Toward Immersive Communications in 6G. *Front. Comput. Sci.* **2023**, *4*, 1068478. [CrossRef]
21. Zabulis, X.; Partarakis, N.; Bartalesi, V.; Pratelli, N.; Meghini, C.; Dubois, A.; Moreno, I.; Manitsaris, S. Multimodal Dictionaries for Traditional Craft Education. *Multimodal Technol. Interact.* **2024**, *8*, 63. [CrossRef]

22. Zabulis, X.; Meghini, C.; Dubois, A.; Doulgeraki, P.; Partarakis, N.; Adami, I.; Karuzaki, E.; Carre, A.-L.; Patsiouras, N.; Kaplanidi, D.; et al. Digitisation of Traditional Craft Processes. *J. Comput. Cult. Herit.* **2022**, *15*, 53. [CrossRef]
23. Zhang, M.; Chen, X.; Pan, Z. Interaction Study of Digital Shadow Play Using the Kinect. In Proceedings of the SIGGRAPH Asia 2014 Posters, Shenzhen, China, 3–6 December 2014. [CrossRef]
24. Li, T.; Cao, W. Research on a Method of Creating Digital Shadow Puppets Based on Parameterized Templates. *Multimed. Tools Appl.* **2021**, *80*, 20403–20422. [CrossRef]
25. Wagner, A.; de Clippele, M.-S. Safeguarding Cultural Heritage in the Digital Era—A Critical Challenge. *Int. J. Semiot. Law—Rev. Int. Sémiot. Jurid.* **2023**, *36*, 1915–1923. [CrossRef]
26. UNESCO. Report of the Independent Expert Group on Artificial Intelligence and Culture. 2025. Available online: https://www.unesco.org/sites/default/files/medias/fichiers/2025/09/CULTAI_Report%20of%20the%20Independent%20Expert%20Group%20on%20Artificial%20Intelligence%20and%20Culture%20(final%20online%20version)%201.pdf (accessed on 22 April 2026).
27. UNESCO. Recommendation on the Ethics of Artificial Intelligence. 2021. Available online: https://unesdoc.unesco.org/ark:/48223/pf0000380455 (accessed on 22 April 2026).
28. Marek, H.M. Navigating Intellectual Property in the Landscape of Digital Cultural Heritage Sites. *Int. J. Cult. Prop.* **2022**, *29*, 1–21. [CrossRef]
29. Liu, X.; Dong, F.; Shui, W.; Geng, G. Blockchain in Digital Cultural Heritage Resources: Technological Integration, Consensus Mechanisms, and Future Directions. *npj Herit. Sci.* **2025**, *13*, 235. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.