PriMera Scientific Engineering
Volume 7 Issue 2 August 2025
ISSN: 2834-2550

PriMera
Scientific
Publications

Application of Mixed Reality (MR) with Gesture Recognition for Teaching and Training Repetitive Movements in Intangible Cultural Heritage Crafts

Type: Research Article
Received: June 24, 2025
Published: July 29, 2025

Citation:
Weili Yang. "Application of Mixed Reality (MR) with Gesture Recognition for Teaching and Training Repetitive Movements in Intangible Cultural Heritage Crafts". PriMera Scientific Engineering 7.2 (2025): 03-12.

Weili Yang*

School of Communication Hong Kong Baptist University, Lee Shau Kee Communication and Visual Arts Building 5 Hereford Road, Kowloon Tong, Kowloon, Hong Kong, China

***Corresponding Author:** Weili Yang, School of Communication Hong Kong Baptist University, Lee Shau Kee Communication and Visual Arts Building 5 Hereford Road, Kowloon Tong, Kowloon, Hong Kong, China.

Abstract

The preservation and transmission of Intangible Cultural Heritage (ICH) face significant challenges in today's globalised world, mainly when teaching complex traditional handicraft techniques that involve repetitive and precise hand movements. This study explores the application of Mixed Reality (MR) technology combined with gesture recognition to enhance the teaching and training of these repetitive movements, focusing on the thread-winding process in Yunnan-style kite making as a case study. MR enables learners to interact with virtual materials in a simulated environment, while gesture recognition, implemented using Google's MediaPipe, captures and evaluates hand movements in real time.

By comparing learners' gestures with the movements of skilled artisans, the MR system provides immediate feedback to correct technique and improve accuracy. The research demonstrates that MR-assisted gesture recognition significantly enhances the standardisation of repetitive movements, improves teaching quality, and reduces training time and material costs. Importantly, repetitive actions—essential to mastering traditional handcrafts—benefit from the immersive, interactive feedback that MR provides, helping learners refine their movements through continuous practice.

This study contributes to the growing field of digital heritage education by showcasing how MR can modernise the transmission of traditional craft skills. In addition, it highlights the potential of gesture recognition technology in advancing the teaching and training of intricate motor skills in ICH practices. The experimental results suggest that MR-based systems could be valuable in preserving and promoting craftsmanship by offering scalable, cost-effective training solutions.

Keywords: Mixed Reality; Gesture Recognition; MediaPipe; Intangible Cultural Heritage; Traditional Handicrafts

Introduction

As defined by UNESCO, Intangible Cultural Heritage (ICH) includes practices, expressions, knowledge, and skills passed down through generations. However, preserving and transmitting ICH faces significant challenges in an increasingly globalised and modernised world (Yang, 2019). Bamboo handicrafts, such as Yunnan-style kites, represent one such traditional practice at risk due to the complexity of the techniques and the lengthy process of hands-on learning (Liu & Zhang, 2020). With the added challenges of high material costs and geographical dispersion of artisans, traditional methods of teaching these crafts often fail to meet the needs of modern learners (Chen & Hu, 2021). Additionally, conventional techniques, such as video-based tutorials, fall short in replicating the 3D intricacies of crafts and do not provide the interactive experience necessary for mastering fine motor skills (Chen et al., 2021).

Technological innovations, particularly Mixed Reality (MR), have shown promise in addressing these challenges in recent years. MR allows for an immersive, interactive learning experience where users can manipulate virtual objects and practice complex tasks in 3D (Azuma, 1997; Billinghurst & Kato, 2002). This makes it especially useful in teaching bamboo crafts, where students can engage with virtual materials and receive real-time feedback on their performance. Studies have demonstrated that MR enhances learners' understanding of intricate processes and enables better visualisation of tasks (Billinghurst & Kato, 2002). Despite this, current MR systems often struggle with accurately recognising complex gestures, particularly in crafts like Yunnan-style kite making, where precision in hand movements is crucial (Tatzgern et al., 2015).

To address the limitations of MR in capturing intricate hand gestures, we explore gesture recognition technology. MediaPipe, a machine-learning-based tool developed by Google, is employed in this study to detect hand landmarks in real-time. By integrating MediaPipe into the MR environment, learners can practice the thread-winding process in Yunnan-style kite making, and their gestures are compared with the pre-recorded actions of skilled artisans. The system provides immediate feedback on the accuracy of the gestures, allowing learners to adjust their hand movements and improve their technique (Lugaresi et al., 2019; Bronstein et al., 2008).

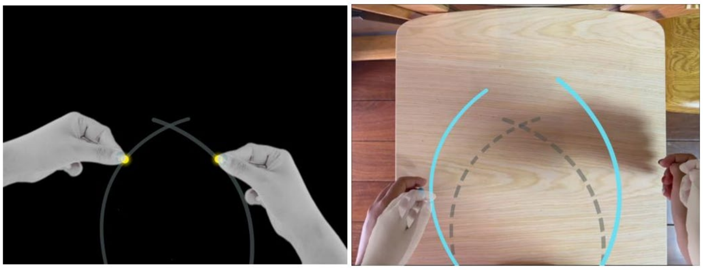
This paper investigates the effectiveness of combining MR with gesture recognition technology for teaching traditional bamboo handicrafts, using the thread-winding process in Yunnan-style kite making as a case study. Through this research, we explore how MR can enhance learning by offering a more interactive and immersive experience, reducing the time and cost associated with learning complex craft skills. The experiment focuses on gesture recognition and evaluates its impact on the standardisation and accuracy of learners' movements, to determine the effectiveness of this approach in enhancing traditional craft teaching.

Method

Mixed Reality (MR) Technology in Handcraft Training

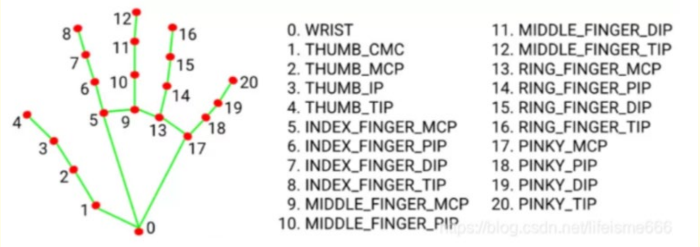
Mixed Reality (MR) technology seamlessly integrates virtual and physical environments, allowing real-time interaction between users and digital objects. In the context of handcraft training, MR enhances the learning experience by providing a more immersive and intuitive environment where users can practice complex motor skills with real-time feedback (Azuma, 1997). For example, traditional handcrafts, such as kite making, require precision and dexterity. MR systems enable learners to interact with virtual representations of the materials and tools, improving understanding and retention (Billinghurst & Kato, 2002). The integration of MR into training offers a novel approach to teaching skills that require a deep knowledge of hand movements and spatial manipulation, making it an ideal medium for replicating such delicate tasks.

MR applications in education have shown significant potential. Studies on using MR for practical training suggest improving user engagement and learning efficiency by providing a contextualised learning environment where users can visualise and practice tasks in a near-real scenario (Chen et al., 2021). In this project, the MR system utilised a head-mounted display (HMD) that tracked head and hand movements, dynamically adjusting the user's view of the virtual objects, which included kites and various thread-winding tools (Vogt et al., 2021). This setup enabled learners to interact with these objects while receiving real-time feedback on their gestures, fostering an interactive and engaging learning experience.



Gesture Recognition in MR

Gesture recognition in MR systems is crucial for accurately simulating the intricate hand movements necessary for traditional crafts (Tatzgern et al., 2015). This study utilises Google's MediaPipe framework for real-time hand tracking, which can detect 21 critical hand landmarks to model the hand's skeletal structure and gestures in detail (Lugaresi et al., 2019).



MediaPipe's machine-learning algorithm processes real-time video inputs, identifies hand gestures, and matches them to pre-defined models. This provides a robust foundation for comparing learners' gestures with expert models, facilitating learning by offering immediate corrective feedback (Bronstein et al., 2008).

MediaPipe Hand Tracking Model

The MediaPipe hand-tracking model used in this research detects 21 hand landmarks, including critical points such as the wrist, knuckles, finger joints, and fingertips. These points form a detailed skeletal map of the hand, which is used to analyse hand movements and gestures (Lugaresi et al., 2019). MediaPipe processes video frames in real-time, extracting the coordinates of each landmark.

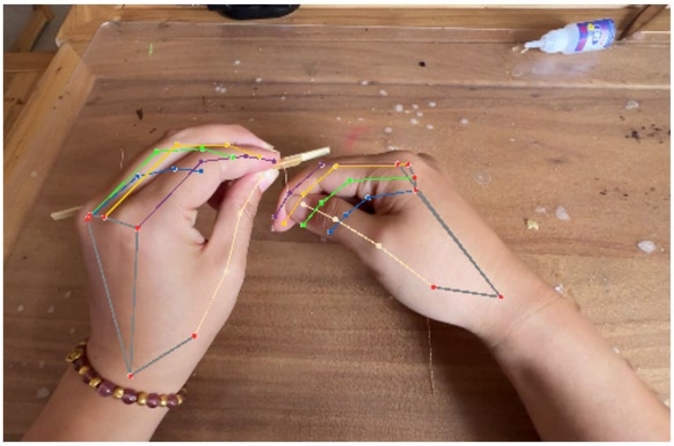
The system continuously monitors the learner's gestures, comparing them against a pre-recorded expert model to assess accuracy. This process enables learners to receive feedback on their actions, allowing them to correct their movements and improve their skills (Bronstein et al., 2008).

Video Info:	over. mp4
Frame Count:	59.73210548346589
Total Number of Frame:	1427.0
Total Screen Time:	23.89 seconds
Video Info:	front.mp4
Frame Count:	60.0198347107438
Total Number of Frame:	1513.0
Total Screen Time:	25.21 seconds

The gesture recognition process involves several key steps:

Hand Detection: The system uses a deep learning model to detect hand regions within video frames.

Landmark Localization: MediaPipe identifies and localizes 21 critical hand landmarks, which serve as the basis for gesture recognition.



Gesture Comparison: The learner's hand gestures are compared against the expert reference model using a similarity measure based on Euclidean distance.

Real-Time Feedback: The system calculates the similarity score between learner and expert gestures. If the score is below a certain threshold, corrective suggestions are given.

Gesture Matching: Euclidean Distance

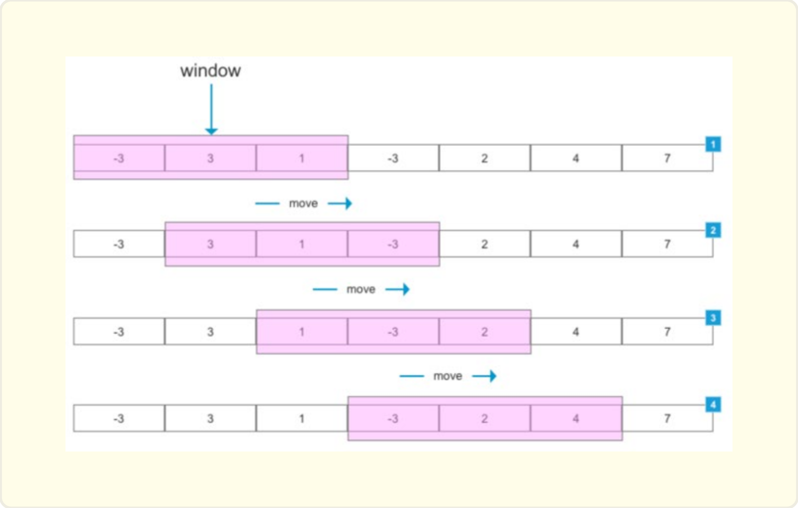
The key to assessing gesture accuracy is measuring the difference between the learner's hand position and the expert model. The Euclidean distance formula is used to quantify this difference. For two points in three-dimensional space, the Euclidean distance is calculated as:

$$\text{similarity} = \frac{1}{1 + \sum_{i=1}^{n} \sqrt{(x_i^{\text{std}} - x_i^{\text{cur}})^2 + (y_i^{\text{std}} - y_i^{\text{cur}})^2 + (z_i^{\text{std}} - z_i^{\text{cur}})^2}}$$

Where $\{(x_i^{std}, y_i^{std}, z_i^{std})\}$ represents the coordinates of a landmark in the expert model, and$\{(x_i^{cur}, y_i^{cur}, z_i^{cur})\}$ represents the corresponding landmark in the learner's hand. A smaller distance indicates a closer match between the learner's and the expert's gestures, while more considerable distances indicate more significant discrepancies. This approach allows the system to provide real-time feedback, enabling users to adjust their movements to better align with the desired gestures (Bronstein et al., 2008).

Smoothing Mechanism: Sliding Window Technique

A sliding window technique is employed to improve the consistency of gesture recognition and mitigate the impact of rapid hand movements or jitter. This technique averages the similarity scores over 30 frames, reducing the effect of sudden fluctuations in hand position (Chen et al., 2021). By calculating a smoothed score, the system ensures that feedback is based on the overall performance of a gesture rather than reacting to minor, momentary deviations. This approach enhances the reliability of the gesture recognition system, particularly in tasks requiring high precision, such as thread winding in kite making (Tatzgern et al., 2015).



Experimental Setup: Two Testing Methods

The experiment employed two methods for assessing gesture recognition in the MR environment: direct video input fitting and live camera fitting—both methods aimed to evaluate the accuracy and reliability of the gesture recognition system during the thread-winding process.

Direct Video Input Fitting: This method used a pre-recorded video of an expert performing the thread-winding process. MediaPipe extracted hand landmarks from each video frame, creating a reference model. The learners' gestures were then compared to this reference model in real-time. This method was expected to produce accurate results, as the controlled environment of the pre-recorded video minimised variations in lighting and camera angles.

Live Camera Fitting: The second method involved tracking the learner's hand movements in real-time using the MR system's built-in camera. This method provided immediate feedback on the learners' gestures, allowing them to adjust their movements dynamically. However, the real-time nature of this approach introduced additional challenges, such as varying lighting conditions and rapid hand movements, which could affect the accuracy of the gesture tracking.

Both methods provided valuable insights into the strengths and limitations of MR-based gesture recognition systems.

Experiment and Results

Experimental Setup

The experiment focused on evaluating the effectiveness of the MR-based gesture recognition system for teaching traditional thread-winding techniques. Two gesture recognition methods were tested: direct video input and live camera fitting. In both cases, the learners' hand gestures were compared to a reference model based on expert performances, and the system provided real-time feedback to guide learners in adjusting their movements (Chen et al., 2021).

Parameters and Evaluation Metrics

Several vital parameters were evaluated to measure the effectiveness of gesture recognition, including gesture accuracy, feedback responsiveness, and system robustness under different conditions. Gesture accuracy was quantified using the Euclidean distance between the learner's and expert's hand landmarks, with a similarity score closer to 1 indicating a high match (Bronstein et al., 2008). Additionally, system responsiveness was assessed by measuring the time to provide feedback after a gesture was performed. Finally, robustness was evaluated by observing how well the system maintained accuracy under different lighting conditions and varying hand movement speeds (Vogt et al., 2021).

Results

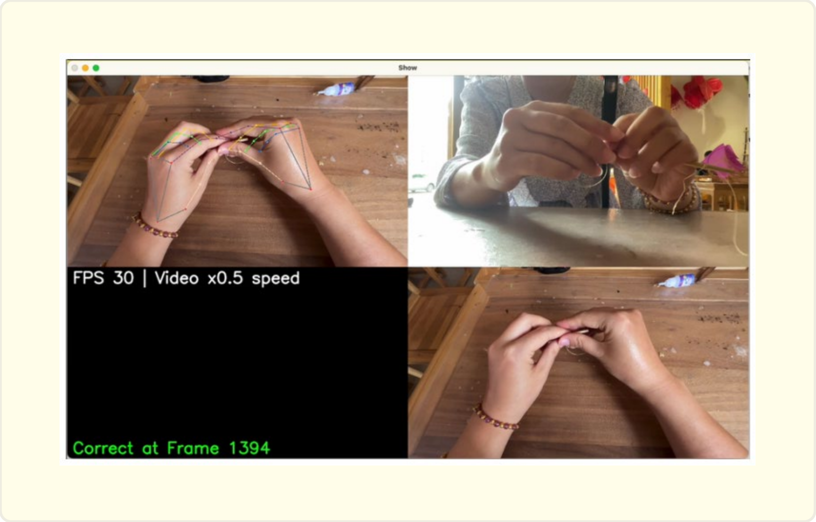
Overall Performance

Direct Video Input Fitting: This method yielded highly accurate results, with learners' gestures closely matching the reference model due to the controlled environment of the pre-recorded video. The similarity scores consistently approached values near 1, indicating high accuracy (Chen et al., 2021). However, this method required careful calibration to ensure that the video conditions (e.g., lighting, camera angle) matched the real-time environment, which could be challenging in some cases (Xu et al., 2022).

Live Camera Fitting: While this method provided more interactive feedback, it was more sensitive to external factors such as lighting conditions and hand movement speed. For example, when learners performed rapid gestures, the system sometimes struggled to track hand movements accurately, resulting in lower similarity scores and less precise feedback (Tatzgern et al., 2015). Additionally, variations in lighting could cause temporary tracking errors, although these issues were partially mitigated through a sliding window smoothing mechanism (Chen & Xu, 2021). Nevertheless, the Live Camera Fitting method better simulates the scenarios encountered when learners wear MR devices for training. Therefore, despite its susceptibility to external variables, the experimental results obtained from Live Camera Fitting are more representative of the system's performance in real-world applications (Vogt et al., 2021).

Data and Screenshots

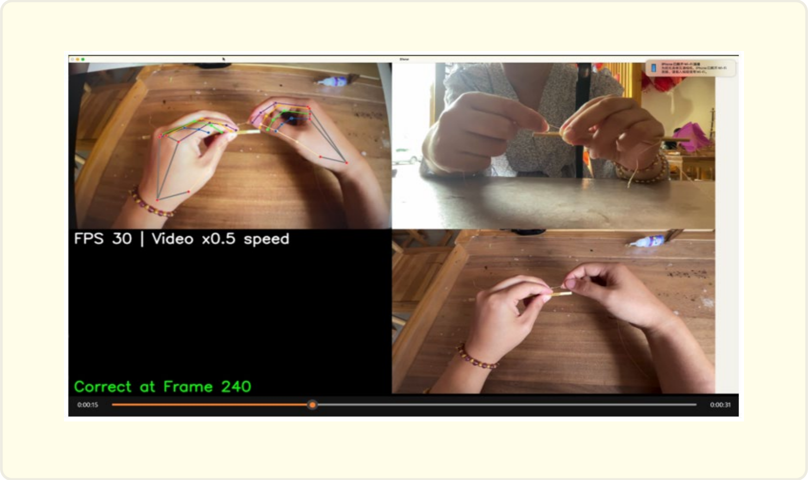
During the experiment, screenshots were captured at specific time points to illustrate the gesture recognition system's performance. For the Direct Video Input Fitting method, which uses pre-recorded standard videos created by expert artisans, the fitting results consistently displayed "Correct at Frame" due to the lack of external interference, validating the system's capability to recognise gestures accurately in a controlled environment.



FPS 30 | Video x0.5 speed

Correct at Frame 1394

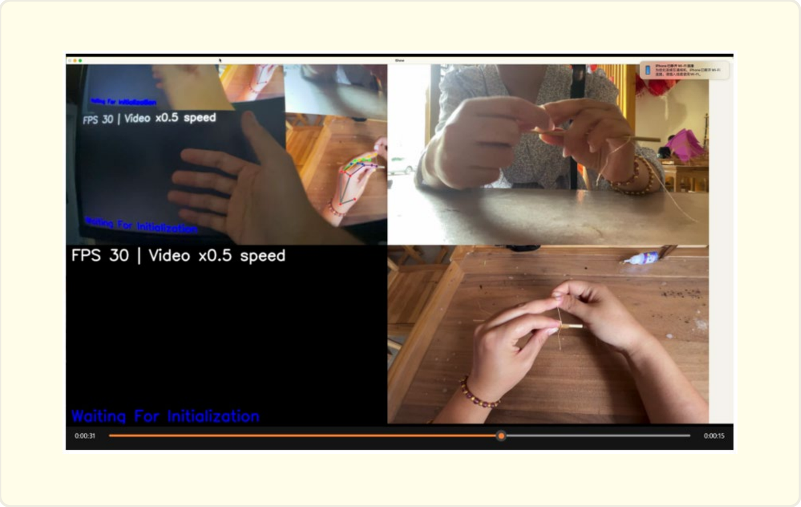
In contrast, the Live Camera Fitting method, designed to emulate the perspective of MR headset cameras, showed varying performance under different conditions. At timestamp 00:15, the gesture fitting result was labelled "Correct at Frame," indicating a high similarity to the expert model.



FPS 30 | Video x0.5 speed

Correct at Frame 240

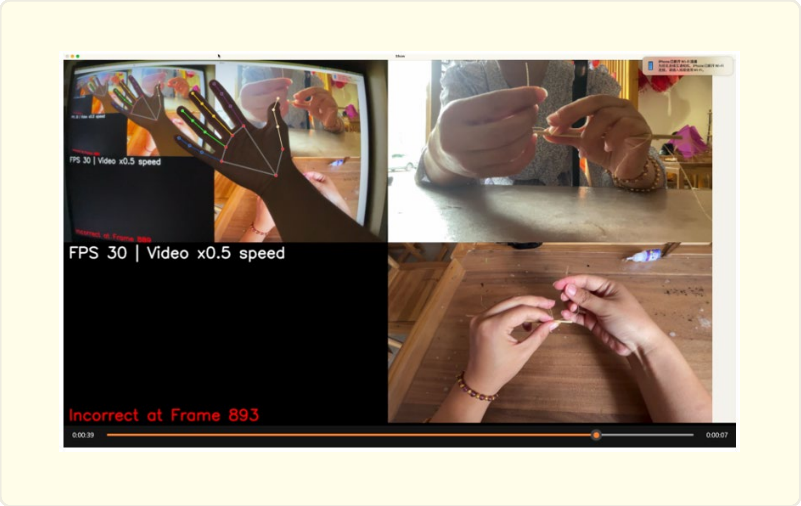
However, at timestamp 00:31, a rapid motion in the captured video caused the fitting result to display "Waiting for Initialization," highlighting the system's sensitivity to motion speed. Furthermore.



FPS 30 | Video x0.5 speed

Waiting For Initialization

At timestamp 00:39, an incorrect gesture was intentionally introduced, and the system correctly identified it by displaying "Incorrect at Frame," demonstrating the system's ability to discern between correct and incorrect gestures in a more dynamic and realistic setting.



FPS 30 | Video x0.5 speed

Incorrect at Frame 893

Advantages and Challenges

The two gesture recognition methods—direct video input and live camera fitting—presented distinct advantages and challenges.

Advantages: Due to its controlled environment, the direct video input method provided highly accurate results. In contrast, the live camera method enhanced interactivity by allowing learners to adjust their gestures in real time. Both methods offered real-time feedback, enabling users to refine their hand movements and improve their performance.

Challenges: The direct video input method required precise calibration between the reference video and the real-time environment. In contrast, the live camera method was sensitive to external factors such as lighting and hand movement speed. These challenges suggest areas for further improvement in the system, such as enhancing the camera's ability to adjust to varying light conditions and refining the algorithms to handle rapid gestures better.

Discussion and Future Directions

The experiment's results demonstrate the potential of MR-based gesture recognition for training in traditional handicrafts. Combining MR and machine learning technologies provides an interactive, real-time training environment that can significantly enhance learning outcomes. However, the system's sensitivity to external conditions, particularly in live camera fitting, suggests that future work should improve the robustness of gesture tracking under varying environmental conditions.

One possible enhancement could involve integrating EMG (electromyography) technology to improve the precision of gesture recognition by capturing muscle signals. This could mitigate the limitations of optical tracking, especially in cases where lighting or movement speed affects the accuracy of the hand-tracking system. Research shows combining visual and EMG signals can significantly improve gesture recognition accuracy (Zhang et al., 2020). Incorporating these advancements into MR systems could further enhance their capability to teach complex, delicate motor skills.

Conclusion

This study demonstrates that integrating Mixed Reality (MR) technology with gesture recognition systems, such as MediaPipe, can significantly enhance the teaching and preservation of complex traditional bamboo handicrafts like Yunnan-style kite making. The results indicate that MR environments provide an immersive, interactive platform for learners to develop intricate motor skills, such as the thread-winding technique. The combination of MR and gesture recognition enables real-time feedback, allowing learners to refine their gestures and improve accuracy, thus shortening the learning curve and reducing material waste. By employing a gesture comparison method based on Euclidean distance and improving gesture consistency using a sliding window technique, the system successfully addresses the challenge of replicating fine motor skills in virtual training environments.

This research fills a gap in the field of Intangible Cultural Heritage (ICH) education, particularly in the application of MR and gesture recognition to traditional crafts. The key contributions of this work are the demonstration of how MR can standardise and accelerate the learning process in crafts that require high precision and the introduction of gesture recognition as a tool to provide immediate corrective feedback. This approach can benefit both educational institutions and artisans by providing a scalable solution to preserving and transmitting delicate craftsmanship skills. Moreover, the methodology presented here can extend to other domains that require precise hand movements, such as medical training, surgery simulations, and vocational education.

Looking ahead, future research should focus on overcoming the limitations of the current system, particularly in terms of its sensitivity to lighting conditions and rapid hand movements in live camera fitting scenarios. Further development could explore integrating electromyography (EMG) technology to enhance the accuracy of gesture recognition by analysing muscle signals, thus mitigating the challenges associated with optical tracking systems. Additionally, expanding the application of MR and gesture recognition technology to other ICH practices could lead to a more comprehensive digital archive of traditional crafts, contributing to their preservation in the digital age. The next steps for this field include improving the robustness of MR-based training systems and exploring multimodal human-computer interaction techniques that combine visual, tactile, and muscle sensing technologies to create a more immersive and intuitive learning experience.

References

1. Azuma RT. "A survey of augmented reality". Presence: Teleoperators & Virtual Environments 6.4 (1997): 355-385.
2. Billinghurst M and Kato H. "Collaborative mixed reality. In Proceedings of the First International Symposium on Mixed and Augmented Reality". IEEE (2002): 2-9.
3. Bronstein MM., et al. "Data fusion through Euclidean and non-Euclidean manifold modeling: Applications to sensor networks". IEEE Transactions on Signal Processing 56.7 (2008): 2678-2689.
4. Chen H and Hu W. "The impact of online learning platforms on traditional craft transmission". Journal of Cultural Heritage Studies 12.3 (2021): 85-97.

5. Chen CH and Wang JJ. "Virtual Reality in Education: Mixed Reality Applications for Vocational Training". Journal of Educational Technology 38.2 (2021): 125-145.
6. Chen HT and Xu J. "Gesture recognition using multi-modal sensor fusion: A study in handcraft preservation". Journal of Human-Computer Interaction 37.9 (2021): 1132-1145.
7. Liu Z and Zhang L. "Challenges in the preservation of intangible cultural heritage in the digital age". International Journal of Cultural Studies 23.2 (2020): 176-189.
8. Lugaresi C., et al. "MediaPipe: A framework for building perception pipelines". arXiv preprint arXiv:1906.08172 (2019).
9. Tatzgern M., et al. "Adaptive and robust hand gesture recognition for interactive systems". IEEE Transactions on Visualization and Computer Graphics 21.5 (2015): 659-668.
10. Taylor AM and Kumar S. "Preserving cultural heritage through immersive technologies: A mixed reality approach to craft education". Virtual Reality & Intelligent Hardware 3.2 (2021): 145-157.
11. Xu Z, Chen P and Qian S. "Advances in gesture recognition systems: Applications to vocational training and traditional handcrafts". International Journal of Human-Computer Studies 157 (2022): 102770.
12. Yang J. "The challenges of globalisation in preserving local cultural practices: A case study of bamboo crafts". International Journal of Heritage Studies 25.1 (2019): 55-72.
13. Zhang Y, Wang Z and Chen X. "Multimodal Gesture Recognition: Combining EMG and Computer Vision". Journal of Neuro Engineering and Rehabilitation 17.5 (2020): 178-190.