Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 2, No. 1, 2025

Tai Chi Assisted Teaching System Based on Kinect and Unity3D

Zixuan Lu, Weibo Yang *

College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China
* **Corresponding author**: Weibo Yang

Abstract: With the rapid development of computer technology, the deep integration of traditional sports and cutting-edge technologies has gradually become a reality, and motion capture technology is increasingly being applied in sports education. This paper designs and implements an intelligent Tai Chi teaching system based on Kinect motion capture devices and the Unity3D platform, aiming to address the lack of interactivity and feedback in traditional teaching methods. The system captures user poses and compares them in real time with standard Tai Chi movements, offering diverse learning and evaluation features. In the beginner mode, the system provides instructional videos of standard movements, allowing users to learn the essentials through repeated viewing. Users can then compare their movements against the annotated key points in the colored data streams for targeted practice. In the advanced mode, the system introduces a motion scoring algorithm that evaluates the user's movements based on the number of matching key points, helping identify areas for improvement and optimize performance. To achieve these functionalities, the system integrates OpenCV and Mediapipe technologies to extract human key points and joint angles from videos, providing robust data support for algorithm design. This paper overcomes the limitations of traditional mouse-and-keyboard interactions by leveraging Kinect's skeletal tracking technology and Unity3D's virtual reality capabilities to create a novel motion-sensing interaction method. This enables users to intuitively control virtual interfaces, enhancing the immersive and engaging nature of the learning process. Experimental results demonstrate that the system can accurately recognize user poses and quantitatively assess learning outcomes. In the future, this system holds significant potential for promoting Tai Chi instruction, enriching sports education models, and contributing to the digital preservation of traditional culture.

Keywords: Kinect, Unity3D, Joint Comparison, Mediapipe, Tai Chi.

1. Introduction

Tai Chi is the treasure of Chinese traditional culture, with its unique characteristics of internal and external training and the philosophical concept of rigidity and flexibility, is loved by people at home and abroad [1]. It is not only a martial art, but also a comprehensive exercise to harmonize the body and mind and strengthen the body through slow-motion exercises. Tai Chi can improve flexibility, coordination and balance, while effectively relieving psychological pressure in modern society [2], and its low-intensity and soothing characteristics make it an ideal choice for people of all ages to pursue health and release pressure [3].

With the promotion of national fitness movement, Tai Chi has been widely spread around the world, and many colleges and universities have incorporated it into their physical education curriculum system [4, 5], helping students to improve their physical fitness and cultural cultivation. In China, Tai Chi is not only a representative of traditional sports, but also carries cultural identity and national pride [6]. However, the traditional teaching mode of tai chi still faces many challenges, especially in large-scale teaching, due to the limited energy of teachers, it is difficult to pay attention to the learning progress of each student, resulting in unsatisfactory teaching results [7]. In addition, the traditional teaching method is relatively single, lack of interactivity and interest, which is difficult to effectively attract the interest of the younger generation in learning. How to improve the teaching effect of tai chi with the help of modern technology has become a key topic of concern in the academic and practical fields. In recent years, with the continuous maturity of motion capture technology [8], the combination of traditional sports teaching and modern technology has become possible. Motion capture technology can capture human movement data in real time and transform it into visualized information to provide more accurate feedback for teaching. The Kinect device introduced by Microsoft has become a revolutionary product in the field of motion capture with its excellent depth sensing and bone tracking functions.Kinect can capture human skeletal data and apply it to real-time interaction and motion recognition [9], which is widely used in the fields of sports, gaming and medical treatment [10].

The application of Kinect in tai chi has attracted extensive attention from scholars. For example, Li et al. developed a physical rehabilitation system based on the Kinect sensor, which utilizes skeletal tracking to identify and verify each movement of the patient, and helps patients with movement disorders to perform tai chi training at home [11]; He et al. used the Kinect camera to capture the movements of the experts and the students, and delivered them to the remote students in multimodal He et al. used Kinect camera to capture the movements of experts and students, and delivered them to remote students in a multimodal way to help them learn tai chi in an immersive way [12]; Cao Shufen et al. utilized Kinect and mobile edge computing technology to collect a large amount of tai chi movement data, and developed a tai chi movement recognition system to improve the effectiveness of Tai Chi teaching and training [13] and Chen Yongsen et al. utilized the FDTW algorithm [14] and provided the students with movement scores through Kinect, which helped them to detect movement deviations and correct them in a timely manner. movement deviations and correct them in time. These studies show that motion capture technology not only improves the accuracy of Tai Chi teaching, but also enhances learning interest through real-time feedback.

However, the currently available Tai Chi assisted teaching systems still have problems such as insufficient interactivity, insufficiently precise scoring criteria, and difficulty in providing personalized training programs. For this reason, it is of great practical significance and broad application prospect to design an intelligent tai chi movement recognition and scoring system based on Kinect and Unity3D, which combines deep learning and motion capture technology to provide a personalized and interesting learning platform for students.

This study aims to design and implement a tai chi movement recognition and scoring system based on Kinect and Unity3D. The system captures the user's movements in real time and analyzes the data with the scoring algorithm to provide users with detailed movement feedback, which helps students to correct movement deviations and improve learning efficiency. At the same time, the system incorporates somatosensory interaction and gamification design to make the learning process more vivid and interesting, and stimulate students' interest in participation. The ultimate goal of the research is to realize the intelligence of tai chi teaching, which not only provides effective assistance for daily teaching, but also promotes the global dissemination of tai chi and the inheritance of Chinese traditional culture.

2. Related Work

2.1. Introduction of Kinect

Kinect V2.0 is an advanced 3D somatosensory camera from Microsoft that integrates multiple functions such as motion capture, image discrimination, and microphone input, and its appearance is shown in Fig. 1. The device mainly consists of an RGB color camera, a depth camera, an IR infrared camera, and a quadratic linear microphone array with signal lights to enhance the indication of the device [15]. Compared with the previous generation Kinect V1, Kinect V2.0 realizes significant improvements in several key technologies, including the extension of the target detection range from 0.5 m to 4.5 m, the increase of the number of detectable persons to six, and the significant increase in resolution, which makes it more competitive in terms of accuracy and application scope.

The core architecture of Kinect can be summarized into three aspects: data acquisition, data processing and feature extraction, and human posture recognition. In data acquisition, the RGB color camera provides high-quality color image data and retains the functions of a traditional camera; the depth camera further collects depth images and human bone data. These data constitute a complete combination of color images, depth images, and skeletal images, laying the foundation for subsequent analysis, as shown in Fig. 2. In addition, the Kinect for Windows SDK 2.0 has a built-in skeletal tracking tool [16] that collects real-time data on the position of human joints in relation to the Kinect device itself, and displays detailed joint information in each frame.

It is worth mentioning that the skeletal tracking function of Kinect provides powerful support for human posture recognition, and its performance has been verified by several studies. For example, Wenlu Yang et al. conducted a systematic review of Kinect-based human posture recognition methods, comparing the performance of multiple algorithms in terms of accuracy in detail, and providing scientific and reasonable suggestions for different research needs [17]. In addition, Kinect's data acquisition and processing capabilities not only provide innovative applications in sports and entertainment, but also are widely used in medical rehabilitation and human-computer interaction scenarios.

To summarize, Kinect V2.0 has become a benchmark device in the field of motion capture by virtue of its versatility, precision and compatibility. Through real-time acquisition and processing of human movement data, Kinect provides technical support for a variety of application scenarios, including tai chi teaching, and its wide applicability makes it show great potential in sports teaching, health monitoring, virtual reality and other fields.

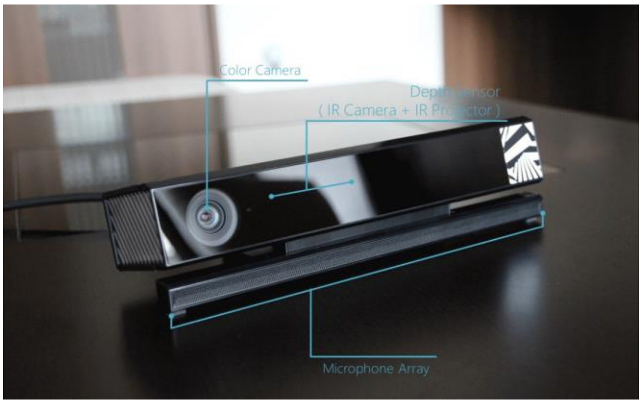

Figure 1. Kinect device demonstration

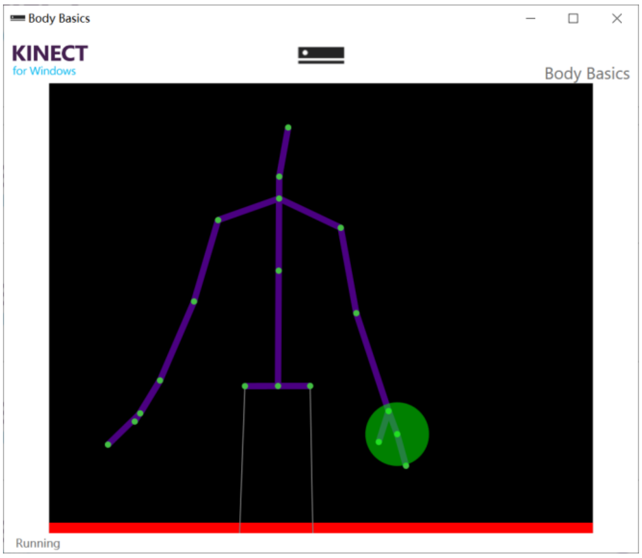

Figure 2. Skeletal images captured by Kinect

2.2. Introduction to Unity3D

Unity3D is a powerful cross-platform game development engine developed by Unity Technologies, Inc. and its excellent flexibility and wide range of applications have made it the tool of choice for many developers and researchers [18]. Unity3D makes it easy to create 3D games, real-time movies, animations, virtual reality (VR), augmented reality (AR) and many types of interactive content. In addition, its excellence in visualization and real-time interaction design makes it widely used in education, healthcare, entertainment, and motion analysis. As an integrated development environment (IDE), Unity3D provides a rich set of development tools covering all aspects from graphics rendering, physics engine to scripting, which fully satisfy the diverse needs of developers [19].

The strong compatibility with multiple resource formats, high-quality rendering effects and real-time interactive features give it a significant advantage in the development of virtual reality, augmented reality, and motion capture applications [20]. Compared with other development tools such as Blender, Virtools, and Torque Game Builder, Unity3D significantly lowers the development threshold with its powerful visualization and interactivity. In addition, Unity3D provides a large number of development resources and plug-in support, which facilitates developers to realize complex functions.

In the field of motion capture and motion analysis, the advanced features and flexible environment make it an ideal platform for technology development. Especially when combined with motion capture devices such as Kinect, Unity3D is able to efficiently process real-time captured data into interactive digital information [21]. For example, Kinect, as a depth-sensing device, can accurately capture human skeletal data. On the Unity3D platform, developers can write C# scripts to control the Kinect device in real-time, extract and process the human body movement data, and generate accurate movement analysis and feedback. the combination of Unity3D and Kinect not only enhances the potential of motion capture technology in the fields of physical education, gaming and entertainment, and health monitoring, but also significantly optimizes the development efficiency and system performance. The combination of Unity3D and Kinect not only enhances the potential of motion capture technology in sports teaching, games and entertainment, and health monitoring, but also optimizes development efficiency and system performance.

Unity3D's development environment provides a modular design to help developers manage and organize their projects efficiently. Its navigation menu includes general functions such as file management, editing, resource import, and project settings. Through the Assets module, developers can easily import external plug-ins and resource files to provide necessary support for project development. In this study, the related plug-in of Kinect is successfully integrated with Unity3D by importing Kinect v2.9. unitypackage, which is used for the construction of motion capture and analysis system. In addition, the toolbar of Unity3D provides basic operations such as panning, rotating, and scaling; the scene view and game view help developers to adjust the objects and verify the design effects, respectively; and the hierarchy view, project view, and monitoring view manage the object hierarchy, resource storage, and attribute information, respectively, which provide all-around support for project development and debugging.

In the field of motion capture, Unity3D's powerful rendering engine and script control capabilities support high-precision motion evaluation and feedback. For example, Kinect-based research systems widely use Unity3D as the core platform to provide users with dynamic motion analysis and real-time feedback through its powerful graphic processing and interactive features. Combining virtual reality and augmented reality technologies, Unity3D platform can enhance the interactivity and fun of motion capture technology in sports teaching and rehabilitation training, making the learning and training process more vivid. In addition, Unity3D enables real-time motion tracking and feedback suggestions, which greatly enhances the user experience and helps students optimize their performance and improve learning efficiency.

In conclusion, Unity3D, as a full-featured and easy-to-expand development platform, provides strong support for the research and application of motion capture technology [22]. Its excellent performance in enhancing interactivity and fun provides a feasible path for the in-depth application of virtual reality and motion analysis technology. By combining with Kinect and other devices, Unity3D has been widely used in sports training, health monitoring, and rehabilitation therapy. In the future, with the further integration of Unity3D and motion capture technology, this platform will continue to play the role of core technology support in innovative applications, and promote technological innovation and application breakthroughs in many fields.

2.3. Unity3D Related Components

In Unity3D, video playback is one of the key components for building interactive applications and virtual environments. To fulfill developers' needs for flexible video playback, Unity3D provides several video playback solutions, mainly including the powerful AVProVideo plug-in and the built-in VideoPlayer component.

AVProVideo is a professional plugin designed for high-performance needs, supporting HD and 360-degree panoramic video playback, with excellent compatibility and efficient rendering capabilities that make it especially good for mobile applications. Compared with Unity's MoveTexture and UGUI RawImage components, AVProVideo can effectively avoid the latter's compatibility issues on some devices, while providing higher playback quality. AVProVideo also supports customized playback rate and progress bar, allowing developers to adjust the video playback rhythm according to specific needs, bringing users AVProVideo also supports customizable playback rates and progress bars, allowing developers to adjust the pace of video playback according to specific needs, providing users with a smoother, more immersive viewing experience. AVProVideo also supports customizable playback rates and progressions, allowing developers to adjust the pace of video playback according to their specific needs, providing a smoother, more immersive viewing experience. In addition, its support for a wide range of video formats and low-latency performance makes it ideal for use in scenarios that require high-quality video playback, such as Virtual Reality (VR) presentations and Augmented Reality (AR) teaching.

VideoPlayer is a built-in video playback component introduced in Unity 5.6 to provide developers with richer and easier-to-use video playback controls. VideoPlayer is designed to provide developers with richer and easier-to-use video playback controls. It supports not only calling local video files directly, but also loading remote video sources through network streams, which makes it very convenient for developers to build online content. VideoPlayer's core strength lies in its flexibility, which allows developers to finely control video playback behaviors such as adjusting playback speed, transparency, and looping through the playback via scripts. In addition, VideoPlayer supports automatic playback when scripts are activated, providing an efficient solution for scenarios that require seamless video playback.

Together, these two video playback components provide powerful multimedia support for Unity3D, meeting the diverse needs for video quality and interactivity in different application scenarios. For example, interactive courseware in education can utilize the flexibility of VideoPlayer to play instructional videos, while AVProVideo is commonly used in high-end entertainment and VR environments to provide immersive video experiences. By combining these video playback capabilities, developers can build more engaging and interactive content on the Unity3D platform, providing users with diverse experiences. Unity3D provides video playback capabilities that meet the complex needs of different scenarios in terms of both quality and performance. AVProVideo's high performance and VideoPlayer's flexibility complement each other, providing strong technical support for developers in the fields of education, entertainment, and virtual reality. As Unity3D technology continues to evolve, these components will play an even more important role in the future of multimedia interaction and content creation.

2.4. Mediapipe

MediaPipe is an open-source cross-platform multimedia machine learning modeling framework developed by Google Research that focuses on tasks such as efficient target detection, keypoint detection, and face recognition [23]. Its powerful performance and flexible compatibility allow it to be widely applicable to servers, mobile devices such as Android and iOS, and embedded device Raspberry Pi. The main advantages of the framework are its support for real-time processing, multi-task parallelism, and cross-platform deployment, which provides strong support for efficient processing of multimedia data.

MediaPipe's core framework is implemented in C++, while Java and Python are supported, making it easy for developers to choose the appropriate development language according to the application scenario. The architectural design is based on five concepts: Stream, Packet, Calculator, Graph and Subgraph. Among them, Calculator is the core of the framework, which is used to perform different stages of data processing tasks. Based on their functions, computational units can be categorized as follows. Preprocessing computing units: to process input image and multimedia data, such as image transformation and tensorization Image Transform and image to tensor. Inference Computation Units: used to locally integrate TensorFlow or TensorFlow Lite models to perform machine learning inference tasks. Post-processing computing unit: responsible for processing the detection results, including segmentation, classification and related feature extraction. Tools computing unit: used to support image annotation and other functions to improve data visualization.

In recent years, MediaPipe has performed well in the field of human gesture recognition, and its multitasking capability has been widely verified. For example, Indriani et al. used MediaPipe to realize gesture recognition, which effectively enhanced the user interactivity of the application [24]. In addition, the modular design of the framework facilitates users to combine functions according to their needs, which balances development efficiency and functionality scalability. In order to facilitate developers to get started quickly, MediaPipe provides a wealth of open source examples and pre-built solutions covering a wide range of areas such as face detection, iris tracking, gesture recognition, 3D target detection, etc. These examples provide developers with the ability to realize complex applications. These examples provide a solid foundation for developers to implement complex tasks, and the framework also supports building highly customized solutions. In Python, developers can install the pre-built packages via pip install mediapipe and quickly invoke their functionality. When implementing MediaPipe, OpenCV is an essential dependency tool for processing input image and video data [25]. OpenCV is a lightweight cross-platform computer vision library that supports multiple language interfaces include Python and MATLAB, and its functions cover image color space conversion, pixel matrix operations, etc. [26]. In the task of posture recognition, the combination of MediaPipe and OpenCV realizes the annotation of key points of the human body and the accurate calculation of joint angles, which provides efficient and reliable data support for motion analysis.

From a comprehensive point of view, MediaPipe is a multimedia processing framework with high efficiency and flexibility, and it shows great potential in the fields of gesture recognition, target detection and human-computer interaction. By combining with OpenCV and other tools, its performance in real-time motion capture and data processing is especially outstanding, which provides a solid technical foundation for related research.

3. Method

3.1. Overall System Analysis

In recent years, Tai Chi, as a treasure of Chinese traditional culture, has attracted many foreign friends to come to learn it, and its promotion in the world has shown a booming trend. However, due to the limitation of time and region, many tai chi enthusiasts are unable to learn systematically under the guidance of famous masters, and rely on books, teaching websites and other ways for self-study. However, there are many limitations in these learning methods, as students can not grasp the essentials of movements intuitively through books, and the video quality of many teaching websites is low and the interface design is not friendly enough, which makes it difficult for students to find out the deficiencies in their own movements during the learning process, and the learning effect is also greatly reduced.

In order to solve the above problems, this study designs and realizes a set of intelligent system to assist the learning of tai chi based on Kinect device and C# programming language in Unity3D development environment. The system uses Kinect SDK as the transmission platform to realize data acquisition and delivery; builds multiple interactive interfaces through the UGUI module of Unity3D; and develops action recognition and scoring algorithms using Visual Studio. The system provides students with an efficient and convenient tai chi learning experience, and at the same time increases the fun and immersion of the learning process through somatosensory interaction.

The overall functional structure of the system is relatively simple, mainly including two modules: learning module and challenge module. The learning module is designed for beginners, with a simple and intuitive interface, and a standard tai chi movement video embedded in the scene, so that students can follow the video to learn and imitate the movements. The Challenge Module is designed to enhance students' motivation through the addition of gamification. The interface also contains videos of standardized movements, which are used to help users recall forgotten movements. In the designated area, students show complete Tai chi movements, which are recognized and scored by the system in real time, and the scores are displayed in the upper-right corner of the interface. Students can challenge the system several times in order to refresh their personal maximum scores. The functional design flowchart of the system is shown in Fig. 3.

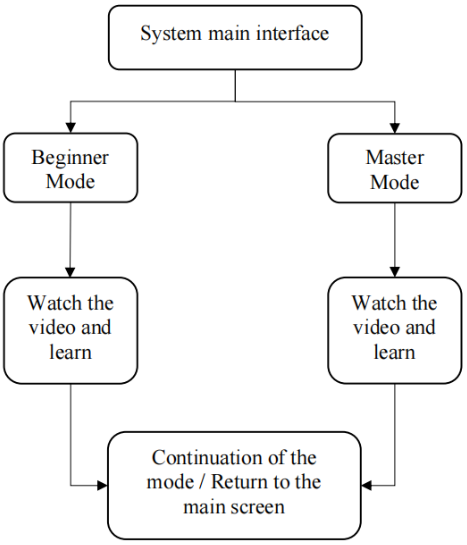

Figure 3. System Functional Design Flowchart

3.2. Parameter Setting

Before detecting the human skeleton, you have to build the detection model. There are related solutions under mediapipe. solutions in the official website, in which drawing_utils is used for drawing, drawing_styles is used for rendering styles, face_detection is used for face detection, face_mesh is used for drawing human face mesh, hands is used for hand detection. holistic is the overall solution (including human face, human skeleton, hands), pose is to recognize pose, objectron is for target detection, selfie_segmentation is selfie segmentation. With Mediapipe 33 joints of the human body can be captured as shown in Fig. 4.

In this study, we use the Mediapipe framework for data feature extraction. Mediapipe provides real-time keypoint detection capability to efficiently extract human skeleton data from input images. In order to optimize the accuracy and efficiency of feature extraction, some key parameters need to be set in the algorithm. Mediapipe implements the function of real-time data extraction in Python, which provides some application programming interfaces for data extraction, and the parameters need to be set in the algorithm when using it. The key parameters are shown in Table 1. Google based MediaPipe framework to achieve a new hand perception method, the use of machine learning to infer a single frame within the 21 3D hand key points, when access to the key points and their coordinates, you can start calculating the angle between the joints. The derivation process is relatively simple, with the help of trigonometric arcs to find the tangent angle of the two straight lines, and then subtracted to get the degree of the angle.

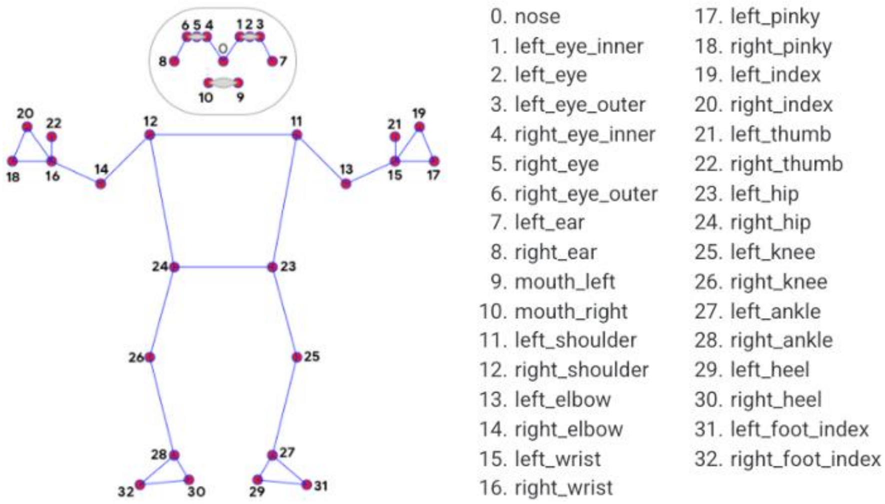

Figure 4. Diagram of Human Joints

Table 1. Parameter Setting

Parameter name	Setting	Function
STATIC_IMAGE_MODE	False	Set to true to treat the input image as a video stream. Attempts are made to detect the most prominent person in the first image and further localize pose landmarks after successful detection. Simply track those landmarks in subsequent images without invoking another detection, reducing computation and latency.
MODEL_COMPLEXITY	0,1,2	0, 1, or 2 refers to the complexity of the pose landmark model, and landmark accuracy and inference latency typically increase with model complexity.
SMOOTH_LANDMARKS	True	Filtering of pose landmarks on different input images to minimize jittering.
UPPER_BODY_ONLY	33	Setting it to 33 means to keep track of all 33 landmarks pose landmarks.

3.3. User Interface Design

The system interface is designed and developed on Unity3D platform, which mainly contains three parts: the main interface, the novice learning interface and the master challenge interface. The functions of each interface are clear, and the layout is intuitive, which provides a good interactive experience for users.

The main interface introduces the basic functions of the system to users in the form of text introduction and provides the entrance for mode selection. Users can independently choose to enter the novice learning mode or master challenge mode according to their own needs. The main interface adopts background pictures closely related to tai chi culture and adds soft background music to create a strong atmosphere of tai chi culture and bring users an immersive visual and auditory experience. The novice learning interface is designed to help users master tai chi movements. Users can watch the embedded video of standard tai chi movements by clicking the video play button in the interface to learn and imitate the movements. During the learning process, the interface displays real-time footage of the user's movements to help users compare and adjust the movements. After the video has finished playing, the user can choose to replay the video to strengthen his/her memory; if he/she does not want to continue learning, he/she can return to the main interface at any time. The design of the interface is simple and intuitive, and the repetition and self-selection functions enhance the fun and flexibility of learning.

4. Experiment

4.1. System Design

The UI interface in this system is designed in a virtual environment developed by Visual Unity, which contains three interfaces: the main interface, the novice interface, and the master interface, of which the design of the main interface is shown in Fig. 5.

Figure 5. Main Interface

Take the main interface as an example to introduce the process of making interface in Unity3D. Firstly, the latest interactive system of Unity, UGUI, is used in designing the UI interface, which has the characteristics of flexible use, beautiful interface, and can be personalized and formulated, and the UGUI system of Unity3D produces the menu, buttons, and scene UI interfaces of the virtual tai chi system, which is simple and convenient to operate, with beautiful images and good user experience [19].

First of all, create a canvas in the Hierarchy view, when the canvas is generated, it will automatically generate an EventSystem for event triggering and invocation. Create a Raw Image in the canvas canvas and add the modified background image to the Texuture of the Raw Image to change the background image. Next, create a Button button in the hierarchical view, the UI contains a Button with an Image component and a Button component; to change the appearance of the Button, you only need to change the Source Image in the Image component. The operation of adding images to Unity is very simple, just select the local image you want to add and drag and drop it to Unity3D. It should be noted that you need to change the Texture Type attribute of the image you want to add independently, and you can use the image as you like only if you change it to Sprite attribute. Add event responses to the corresponding buttons and create a script, name it ClickButton. cs, write your own OnClick function and bind the script to the corresponding buttons to complete a button click event. Place the buttons in order and write the corresponding click events.

4.2. Gesture Trigger Implementation

Before realizing the interaction between the user and the system through the hand, it is necessary to know how to control the mouse with gestures through Kinect. Kinect 2.0 acquires the user's right hand coordinates by collecting the depth data, analyzing the collected depth data, and projecting the coordinates to the XY plane to get the X-axis and Y-axis coordinates of the right hand relative to the Kinect device. The acquired X-axis and Y-axis coordinates are converted to planar coordinates, and then the converted coordinates are judged whether they are out of the interface range. If not, the converted coordinates are assigned to the gesture image, so that the added gesture image follows the user's right hand to move.

4.3. Implementation of GUI Interface Switching

This system mainly consists of three interfaces, the main interface, the novice learning interface, and the master challenge interface, and to realize the user's independent choice of different interfaces, it is necessary to do a good job of switching between the interfaces. All the interface switching in this system is triggered by the button click event, so we need to write the switching function, add the written script to the button properties, and then select the corresponding switching function to add to the OnClick event.

4.4. Implementation of Action Recognition

There are two basic recognition methods: static recognition and dynamic recognition, both of which have different advantages and disadvantages. Static recognition is to analyze static images to obtain human joint images, the advantage is that the algorithm is mature, high stability, but for the movement state of the human body posture recognition accuracy is low, the application of occasions is limited; dynamic recognition is through the continuous image frames, the analysis of the human body's continuous action, the advantage is that the application of occasions is extensive, making it easier to carry out the interaction between humans and computers, but due to the complexity and instability of the dynamic recognition algorithm and the user's action posture has a highly personalized characteristics, the recognition of the advantages and disadvantages of the two methods. However, due to the complexity and instability of the dynamic recognition algorithm and the highly personalized nature of the user's action posture, the recognition error rate is relatively high [20].

In the design of this system, due to the limitation of human and material resources, time disallowance, and the success rate of pose recognition [21]. Finally, the method of simulating dynamic recognition with static recognition is adopted for the realization of system functions. Generally, the whole process of static recognition is to first select some action poses from a standard action video, vectorize the selected poses, project the 3D vectors onto a 2D plane, select suitable joints as key points, calculate the vectors between the joints, and then calculate the angle between the vectors, and then use the angle between the joints as the key feature to recognize the action by comparing it with the preset pose angle. The action is recognized by comparing it with the preset pose angle. For the simplified human skeletal map captured by Kinect, we can view it as a model of multiple skeletal points connected together [22]. For the scoring of user actions is this thesis is based on the number of key points corresponding to the number of key points.

4.5. Testing

During the development of the system, the connection of the Kinect device was first tested. By writing a self-test script in Unity3D, it was verified that the Kinect device was able to correctly recognize the user and capture images. The test results showed that Kinect was able to successfully detect the user and pass the image to the system for further processing. This testing process is as expected and ensures that the Kinect device can work stably.

This system realizes natural interaction through human bone tracking technology, where the user controls the virtual interface through hand gestures. Tests show that the system can accurately recognize the user's fist gesture and trigger the intended action. When the user makes a fist gesture in front of the Kinect, the system successfully recognizes and executes the corresponding command, and the console displays “complete fist gesture”, which verifies the accuracy and responsiveness of gesture recognition.

The system's interface includes a main interface, a novice interface and a master interface. In the test, the size of the human body image captured by Kinect was adjusted, while the font and color settings in the interface were optimized. The test found that there is a certain delay in the system, especially when the sensitivity of gesture recognition is low, resulting in user actions not being able to respond in time. By optimizing the sensitivity settings, some of the problems were alleviated. However, the system still appeared to move the mouse too fast. Eventually, on a better configured computer, the system performance was significantly improved and the loading speed was also effectively improved.

In the process of system development and testing, some technical problems were encountered, mainly including the following:

(1) Kinect SDK version compatibility problem: different versions of Unity3D need to be accompanied by different versions of Kinect SDK, otherwise the device cannot be recognized correctly.

(2) File path problem: When using Mediapipe and OpenCV for data extraction, Chinese characters in the path name will lead to reading failure and affect the normal operation of the system.

(3) Inconsistent naming of scripts: Some Unity scripts run incorrectly, mainly due to inconsistency between the names of script files and the names referenced in the code, especially spaces and case issues.

(4) Null Reference Error: In some scenarios, the script does not instantiate the object, resulting in a “NullReferenceException” error. This problem can be solved by remounting the script object.

In terms of device detection, gesture recognition and interface functions, the system basically meets the expected requirements. Although there are certain performance bottlenecks and compatibility problems, the stability and user experience of the system have been effectively improved through optimization and adjustment. The test results show that the system function implementation meets the design objectives, and will continue to optimize the performance and solve the remaining problems.

5. Conclusion

In this paper, an intelligent tai chi teaching system is developed based on Kinect and Unity3D. By combining somatosensory interaction technology and gamification design, the system innovatively realizes the function of motion capture and scoring, which significantly improves the interactivity and fun of tai chi learning. With the core of real-time capturing user movements and comparing them with standard movements, the system meets the needs of different learners through multiple modes, and initially realizes the intellectualization of tai chi teaching. Although the system has achieved some success in functional design and user experience, there are still deficiencies in the accuracy of movement recognition, real-time performance and personalized training programs. Future research will be devoted to introducing deep learning to optimize the scoring algorithm and enhance the recognition ability of complex movements, as well as expanding the support for more movement patterns and optimizing the system performance. In addition, combining cloud computing and big data analysis technology, we will explore the possibility of multi-device linkage and remote teaching to provide users with personalized training solutions; and through the integration of virtual reality and augmented reality technology, we will further enhance the immersion of the system and the extensiveness of the application scenarios, so as to inject new vitality into the traditional sports teaching.

References

[1] Wang, Chenchen, Jean Paul Collet, and Joseph Lau. "The effect of Tai Chi on health outcomes in patients with chronic conditions: a systematic review." Archives of internal medicine 164.5 (2004): 493-501.

[2] Lan, Ching, et al. "Tai chi chuan in medicine and health promotion." Evidence-Based Complementary and Alternative Medicine 2013.1 (2013): 502131.

[3] Caldwell, Karen, et al. "Effect of Pilates and taiji quan training on self-efficacy, sleep quality, mood, and physical performance of college students." Journal of bodywork and movement therapies 13.2 (2009): 155-163.

[4] Webster, Craig S., et al. "A systematic review of the health benefits of Tai Chi for students in higher education." Preventive medicine reports 3 (2016): 103-112.

[5] Xu, Rong. "Integrating the philosophy and practice of Tai Ji Quan in French university's curricula: reducing stress and improving well-being of students." (2018).

[6] Heng, Zou, and Pimurai Limpapath. "Intention to Use Innovative Instructional Methods for Teaching Tai Chi in Guangdong Province: Investigating Micro-Lecture and PAD Class." Library Progress International 44.3 (2024): 28033-28043.

[7] Wang, Zhen-Dong, and Feng-Yan Wang. "Ternary Taiji models of the traditional Chinese self: centered on Confucian, Taoist, and Buddhist cultures." Journal of Humanistic Psychology (2021): 00221678211016957.

[8] Menolotto, Matteo, et al. "Motion capture technology in industrial applications: A systematic review." Sensors 20.19 (2020): 5687.

[9] Lekova, Anna, D. Ryan, and Reggie Davidrajuh. "Fingers and gesture recognition with kinect v2 sensor." Information Technologies and Control 14.3 (2016): 24-30.

[10] Jawaid, Iqra, and Jawad Khalid Qureshi. "Advancements in medical imaging through Kinect: a review." 2017 International

Symposium on Wireless Systems and Networks (ISWSN). IEEE, 2017.

[11] Lee, Jiann-Der, Chung-Hung Hsieh, and Ting-Yang Lin. "A Kinect-based Tai Chi exercises evaluation system for physical rehabilitation." 2014 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2014.

[12] He, Tianyu, et al. "Immersive and collaborative Taichi motion learning in various VR environments." 2017 IEEE Virtual Reality (VR). IEEE, 2017.

[13] Cao, Shufang, and Xiaozhou Huang. "The 24-Type Tai Chi Classic Motion System Based on Kinect Motion Capture in Mobile Edge Computing." IEEE Access (2024).

[14] Pang, Xun, and Bin Liang. "Motion recognition based on Kinect for human-computer intelligent interaction." Journal of Physics: Conference Series. Vol. 1187. No. 3. IOP Publishing, 2019.

[15] Shao, Ling, et al. "Computer vision for RGB-D sensors: Kinect and its applications [special issue intro.]." IEEE transactions on cybernetics 43.5 (2013): 1314-1317.

[16] Webb, Jarrett, and James Ashley. Beginning kinect programming with the microsoft kinect SDK. Apress, 2012.

[17] Wenlu Yang, Ying Peng, Hong Xie. Action Recognition Based on Kinect Deep Learning [J]. Journal of Electronics and Information Science, 2022, 7(1).

[18] Wang, Sa, et al. "A new method of virtual reality based on Unity3D." 2010 18th international conference on Geoinformatics. IEEE, 2010.

[19] Ren, Ran, et al. "Vehicle crash simulations for safety: Introduction of connected and automated vehicles on the roadways." Accident Analysis & Prevention 186 (2023): 107021.

[20] Papakostas, Christos, et al. "Exploration of augmented reality in spatial abilities training: a systematic literature review for the last decade." Informatics in Education 20.1 (2021): 107-130.

[21] Luckykumar Dwarkadas, Agrawal, et al. "A review of the application of virtual and augmented reality in physical and occupational therapy." Software: Practice and Experience 54.8 (2024): 1378-1407.

[22] Kaur, Deepti Prit, Narinder Pal Singh, and Bonny Banerjee. "A review of platforms for simulating embodied agents in 3D virtual environments." Artificial Intelligence Review 56.4 (2023): 3711-3753.

[23] Chauhan, Rahul, Ikshit Dhyani, and Himadri Vaidya. "A Review on Human Pose Estimation Using Mediapipe." 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT). IEEE, 2023.

[24] Harris, Moh, and Ali Suryaperdana Agoes. "Applying hand gesture recognition for user guide application using MediaPipe." 2nd International Seminar of Science and Applied Technology (ISSAT 2021). Atlantis Press, 2021.

[25] Marengoni, Maurício, and Stringhini Stringhini. "Tutorial: Introdução à visão computacional usando opencv." Revista de Informática Teórica e Aplicada 16.1 (2009): 125-160.

[26] Kavitha, D., et al. "Multiple Object Recognition Using OpenCV." REVISTA GEINTEC-GESTAO INOVACAO E TECNOLOGIAS 11.2 (2021): 1736-1747.