國立臺灣大學電機資訊學院資訊網路與多媒體研究所

碩士論文

Graduate Institute of Networking and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

# 應用肢體引導於擴增實境遠距協作-以太極拳學習為例

# On-Body Pose Guidance in AR Remote Collaboration: Tai-Chi Chuan Learning as an Example

陳洞杋

Chiung-Fan Chen

指導教授：洪一平 博士

Advisor: Yi-Ping Hung, Ph.D.

中華民國 112 年 08 月

August 2023

doi:10.6342/NTU202302869

# 謝辭

能完成這篇論文，要感謝的人有好多。首先是洪一平老師，從研究方向、系統架構、實作細節到問卷設計等等大小事情都提供了許多建議與指導。再來要感謝宗翰，隨著越來越接近口試，在我壓力最大的時候，第一線的忍受我各種鬧脾氣，也是最常聽我講話、安慰我的人。實驗室的大家也都是同一艘船上的好夥伴，平常在實驗室一起嘴砲、一起吃飯、一起抱怨，都是支撐我下去的動力，讓我不用單打獨鬥。尤其是敬媛、冠廷、琪琪、芭樂哥，一路上的研究你們都陪我討論很多論文的問題，也是一起度過困難的前鋒部隊。在接近口試的幾個月，同爲碩二的宜儒、筱晴也陪我聊了很多心事，也互相勉勵了很多。也要謝謝博班的昇達，像是個小主管一樣幫我想各種方法與提供各種資源。還有任勞任怨的奕霆，在需要人力的時候總是都在。謝謝系上的心理師嘉琳，這是我第一次的心理諮商，覺得比我想像中的有效很多。半夜接電話的曜福、來做user study的品文、遞給我衛生紙的呂靜、臨時被我叫來幫忙user study的培凱、一樣遇過研究困難的媽媽，接受過好多大家的忙，每次都會覺得心裡暖暖的。另外，Migue的貼圖、好樂團的音樂、優酪乳、益生菌、維他命、大陸解說幹片都是處理焦慮的好夥伴。總之，只有我一個人肯定是不夠的，也希望我能以此作爲提醒，要做個溫暖、樂於幫助他人的人。

祝看到這篇文的你，研究順利。

# 摘要

太極拳是一項包含著連續肢體動作的中國傳統武術。在傳統上來說，太極拳的學習會要求指導者與學習者在同一個空間內。但是，再遠距離教學的需求下，現有的替代方案像是視訊教學與太極拳自學系統都有著各自的缺點。在本研究中，我們提出了一套結合了擴增實境與WIMP (視窗、圖標、選單、指標) 的遠距離協作系統。此系統包含了提供沉浸式環境的自學子系統與作爲溝通工具的指導子系統。在指導子系統中，我們設計了位於虛擬鏡子上的2D的視覺引導標示與位於學習者肢體上的3D視覺引導標示作爲肢體引導的媒介。在兩項使用者研究中，我們各自比較了指導者與學習者對於2D與3D視覺引導標示的想法。使用者在兩種視覺引導標示都給予了高評價，但使用者認爲3D的視覺引導標示比起2D版本，提供學習者更多的指導資訊，並提供指導者更多的引導功能。

關鍵字：擴增實境、遠距協作、太極拳、混合實境、肢體引導

doi:10.6342/NTU202302869

# Abstract

Tai-Chi Chuan (TCC) is a famous Chinese exercise with body movement sequences, which traditionally requires the instructor and the learner to be collocated. However, in situations where the instructor and the learner are not in the same place, alternative methods like video conferencing and TCC self-learning exist disadvantages. In this paper, we propose a remote collaboration system combining Augmented Reality (AR) and Windows-Icon-Menu-Pointer (WIMP) for TCC learning. It contains a self-learning subsystem providing an immersive environment and a guidance subsystem as a communication tool. For pose guidance, we design the 2D annotations on the augmented mirror in the virtual environment and the 3D on-body annotations imposed on the learner's joints. 2 user studies were conducted to compare the two types of annotations from the perspective of the instructor and the learner. The result shows that the 2D and 3D annotations both receive high reviews. The 3D annotation performs better by including more instructing information for the learner and more annotation features for the instructor.

Keywords: augmented reality, remote collaboration, Tai-Chi Chuan, mixed reality, pose guidance

doi:10.6342/NTU202302869

# Contents

謝辭 i

摘要 ii

Abstract iii

List of Figures vii

List of Tables viii

1 Introduction 1

2 Related Work 3

2.1 Tai-Chi Chuan Learning System . . . . . . . . . . . . . . . . . . . . 3

2.2 Remote Collaboration Combining AR and WIMP . . . . . . . . . 4

2.3 Visual Guidance for Body Movement . . . . . . . . . . . . . . . . 5

3 System Design 7

3.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.2 Self-Learning Subsystem . . . . . . . . . . . . . . . . . . . . . 8

3.2.1 Virtual Coach . . . . . . . . . . . . . . . . . . . . . . . . 8

3.2.2 Augmented Mirror . . . . . . . . . . . . . . . . . . . . . . 8

3.2.3 User Interface of Learner . . . . . . . . . . . . . . . . . . 8

3.3 Guidance Subsystem . . . . . . . . . . . . . . . . . . . . . . . 11

3.3.1 Receiving Body Data . . . . . . . . . . . . . . . . . . . . 11

doi:10.6342/NTU202302869

3.3.2 Providing Annotations . . . . . . 11
3.3.3 Receiving Annotations . . . . . . 11
3.3.4 Adjusting Body Posture . . . . . . 12

4 Implementation 14
4.1 Sensor Side . . . . . . 15
4.2 Server Side . . . . . . 15
4.3 Instructor Side . . . . . . 15
4.4 Learner Side . . . . . . 17

5 User Study 19
5.1 User Study 1 . . . . . . 19
5.1.1 Participant . . . . . . 19
5.1.2 Experimental Condition . . . . . . 20
5.1.3 Scale . . . . . . 20
5.1.4 Procedure . . . . . . 20
5.2 User Study 2 . . . . . . 21
5.2.1 Participant . . . . . . 21
5.2.2 Experimental Condition . . . . . . 21
5.2.3 Scale . . . . . . 21
5.2.4 Procedure . . . . . . 21

6 Result and Discussion 22
6.1 User Study 1 . . . . . . 22
6.2 User Study 2 . . . . . . 24

7 Conclusion and Future Work 26

References 28

A Questionnaire 34
A.1 System Usability Scale (SUS) . . . . . . 34

doi:10.6342/NTU202302869

A.2 NASA Task Load Index (NASA-TLX) . . . . . . . . . . . . . . . . . . 35
A.3 Others . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

doi:10.6342/NTU202302869

# List of Figures

3.1 Virtual coach. . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.2 Augmented mirror. . . . . . . . . . . . . . . . . . . . . . . . . . 10

3.3 User Interface for learner (Hand panel). . . . . . . . . . . . . . . . 10

3.4 User Interface for Instructor. . . . . . . . . . . . . . . . . . . . . . 12

3.5 Annotations in virtual environment. . . . . . . . . . . . . . . . . . 13

4.1 System architecture. The red arrows stand for large data transferring (images) while the pink arrows represent small data transferring. 14

4.2 Vicon Setup. There are 6 Vicon cameras in the area. . . . . . . . . 16

4.3 Sports braces and specific joints. . . . . . . . . . . . . . . . . . . 18

A.1 SUS questionnaire (5-point Likert scales). . . . . . . . . . . . . . 34

A.2 NASA-TLX questionnaire (5-point Likert scales). . . . . . . . . . 35

A.3 Questionnaire for others (5-point Likert scales). . . . . . . . . . . 35

# List of Tables

6.1 Result of User Study 1. . . . . . . . . . . . . . . . . . . . . . . . . . 23
6.2 Result of User Study 1 (Preference). . . . . . . . . . . . . . . . . . . 23
6.3 Result of User Study 2. . . . . . . . . . . . . . . . . . . . . . . . . . 24
6.4 Result of User Study 2 (Preference). . . . . . . . . . . . . . . . . . . 24

doi:10.6342/NTU202302869

# Chapter 1

# Introduction

Tai-Chi Chuan (TCC) is a famous Chinese exercise with body movement sequences. There are key poses in each sequence with pithy formulas, which are called ”Forms”. When practicing TCC, not only the accuracy of the forms themselves are important, but also the smooth changes between them. Hence, it requires attentive cooperation and manipulation of each body part to perform the precise movement.

Traditionally, the class of TCC requires the instructor and learner to be collocated. While the instructor is performing and explaining the movement, the learner is able to observe and mimic. When the learner practices the movements, the instructor is also able to provide feedback immediately. However, in situations where the instructor and the learner are not in the same place, these interactions are hard to be achieved.

There are researchers providing alternative methods trying to deal with the situation. Video Conferencing allows the instructor to teach remotely, which provides collaborative communication. However, the 2D display like a PC screen brings two problems. Firstly, the 2D display shows body movements from a limited perspective, which is difficult for the learner to observe. Secondly, in TCC Forms, there are lots of movements containing head rotation, letting the learner hard to observe and mimic the Forms simultaneously. Augmented Reality (AR) and Virtual Reality (VR) self-learning systems are other methods, which create

doi:10.6342/NTU202302869

immersive experiences for the learner and solve the issues of 2D display. Previous studies [1, 2, 3, 4, 5, 6] propose systems providing virtual coach avatar with TCC movement animation, automatic visual guidance, and automatic assessment. Nonetheless, these features can not completely replace the role of the instructor, who is able to give personal guidance depending on the learning situation and provide teaching material that is not pre-recorded.

In this paper, we propose a remote collaboration system combining AR and Windows-Icon-Menu-Pointer (WIMP) for TCC learning. On one hand, it includes features of a self-learning system, which provides the learner with an AR environment and virtual coach for mimicking. On the other hand, it also contains a tool for collaborative communication. It captures the body data of the learner and streams them to the remote instructor. According to these data, the instructor is able to make 2D/ 3D annotations as the pose guidance. Then, the learner can see the augmented 2D and 3D annotations displayed in the virtual environment.

doi:10.6342/NTU202302869

# Chapter 2

# Related Work

## 2.1 Tai-Chi Chuan Learning System

There are researchers developing TCC training systems in VR and AR. Chua et al. [1] and Iwaanaguchi et al. [2] build VR self-training systems. In the VR environment, there is a virtual coach containing the body motion data of TCC. While the former system displays the virtual coach in front of the user, the latter provides automatic changes in the angle of view to improve the experience. Han et al. [3, 4] propose an AR learning system with several drones shooting user's motion from different angles. These videos are streamed to the AR environment as augmented mirrors for users to modify their posture. Chen et al. [5] build a remote learning system for an instructor and multiple learners with several motion quality assessment methods. The system provides 3 different environments: a Head-Mounted Display (HMD), a PC, and the Cave Automatic Virtual Environment (CAVE). Jan et al. [6] propose a self-learning system in the AR environment with an augmented mirror, virtual coaches, and automatic pose evaluation driven by four cameras and a human pose estimation machine learning model.

Apart from the ones focusing on full-body motion, there are researchers aiming at specific parts of the body. Kao et al. [7] propose an AR system that monitors learners' weight distribution when striking a pose and provides visual hints for correction. Jan [8] pays attention to the footwork and designs a system that tracks the motion of the lower body and shows the augmented guidance.

Based on previous work, Li [9] proposes a remote collaborative learning system by adding a communication tool in the AR self-learning environment. The system provides a live stream video of the learner and allows the instructor to make 2D annotations on the augmented mirror and 3D annotations on the virtual coaches. The work solves the issues of video conferencing and AR/ VR self-learning systems. Inspired by this novel idea, we design our system and improve the guidance tool to provide more instructing information for the learner and more annotation features for the instructor.

## 2.2 Remote Collaboration Combining AR and WIMP

For remote collaboration on physical tasks, AR is an indispensable technology [10]. Cidota et al. [11] develop an AR framework to support visual communication. It allows the remote instructor to see the first-person perspective of the local learner and to generate 2D icons on AR HMD as guidance. It also provides automatic visual and audio notifications to support workspace awareness. Sun et al. [12, 13] present OptoBridge, a teaching platform based on AR, which is designed to assist skill acquisition. Through AR HMD, the learner can be indicated by virtual hands and augmented annotations controlled by the remote teacher's gestures. Wang et al. [14] develop a telepresence application using AR for remote medical training. In the application, the instructor's hand gestures are captured and virtually displayed to the remote learner. Gupta et al. [15] present an AR tele-assistance framework with a hand gestural interaction method to localize the ROI in a first-person view. Günther et al. [16] build a prototypical application implementing audio, visual, and tactile cues in order to compare different communication channels. Kim et al. [17, 18] develop a video conferencing system, which allows the remote user to draw augmented annotations in the workspace. The annotations can be seen through the AR HMD worn by the local user. The works also compare the effects of different collaboration styles, view independence, methods of drawing annotation, and visual notifications.

## 2.3 Visual Guidance for Body Movement

In motor skill training, visual cues play important roles in the guidance. LightGuide [19] is a hand movement guidance system using light to project annotations directly on the user's body. The study provides design guidelines for visual annotations. Following the guidelines, they create four different types of annotations to improve the accuracy of hand movement in mid-air. Anderson et al. [20], Velloso et al. [21] and Sieluzycki et al. [22] use full-body skeleton as the tool of providing guidance. Anderson et al. propose YouMove, a system that allows users to record and learn physical movement sequences. The specially designed AR mirror displays the skeleton of the user and the annotations (e.g. red circles and green ribbons) on it. MotionMA (Motion Modelling and Analysis) by Velloso et al. is a training system that transfers the demonstration exercise by one user into a skeleton model and provides real-time feedback for other users trying to mimic. Sieluzycki et al. present a trainer system for high-precision techniques in judo. Physio@Home [23] and SleeveAR [24] explore the guidance methods of rehabilitation exercise. Both of the two systems generate guidance according to pre-recorded exercise data. The former shows annotations such as Movement Arc, Directional Arrow, and Nearest Arm on the 2D screen with multi-camera views. The latter, on the other hand, provides projection-based annotation. Ribeiro et al. [25] provides a VR annotator that allows the user to visualize and annotate point cloud and skeleton data in a 3D virtual environment, which becomes the guidance for movement suggestion. FuturePose [26], a martial arts training system, uses deep learning-based real-time human pose forecasting to generate motion data and display it as a red skeleton. Since the system shows the future action in advance, it earns more time for the user to react. Lee et al. [27] conduct a comparative study of visual instructions in a VR environment. The annotation-based instruction and virtual tutor guidance are compared on three tasks, maze escape, stretching exercise, and crane manipulation.

doi:10.6342/NTU202302869

# Chapter 3

# System Design

## 3.1 Overview

In most cases of AR-based remote collaboration on physical tasks, the remote instructor has most of the experience while the local learner has a better overall view of the workspace [10]. According to the scenario, we design our AR TCC remote collaboration system to be a one-to-one, instructor-to-learner (remote expert) model. For the learner side, we choose an Optical See-Through Head-Mounted Display (OST-HMD) in order to solve the 2D display issues. Although VR HMD and Video See-Through Head-Mounted Display (VST-HMD) are also suitable, it is better to provide real-world views to the learner in motor skill training like TCC. As for the instructor side, we use WIMP as the interface since we choose 2D video streaming as the main communication media. 3D human reconstruction avatar in a virtual environment seems to be another feasible way, but the joint position error of 3D real-time reconstruction is too large to be ignored in our application [28, 29]. Even though there is work using multiple cameras to increase accuracy [30], the jittering joint data is another unresolved problem.

In the following sections, we will introduce the two subsystems contained in our work: the self-learning subsystem and the guidance subsystem.

doi:10.6342/NTU202302869

## 3.2 Self-Learning Subsystem

The self-learning subsystem aims to provide an AR environment with imposed virtual objects for the learner to self-practice TCC. In this part of our system, We follow the previous studies [3, 4, 6, 8, 9], which provide well-designed features.

### 3.2.1 Virtual Coach

The virtual coach is a 3D human model with pre-recorded body motion animation (Figure 3.1). The motion data is demonstrated by two Yang-Style TCC masters and is recorded by a motion capture system (Vicon). We place 8 virtual coaches around the learner in different directions. When these coaches perform the standard TCC movements, the learner is able to observe them from different angles.

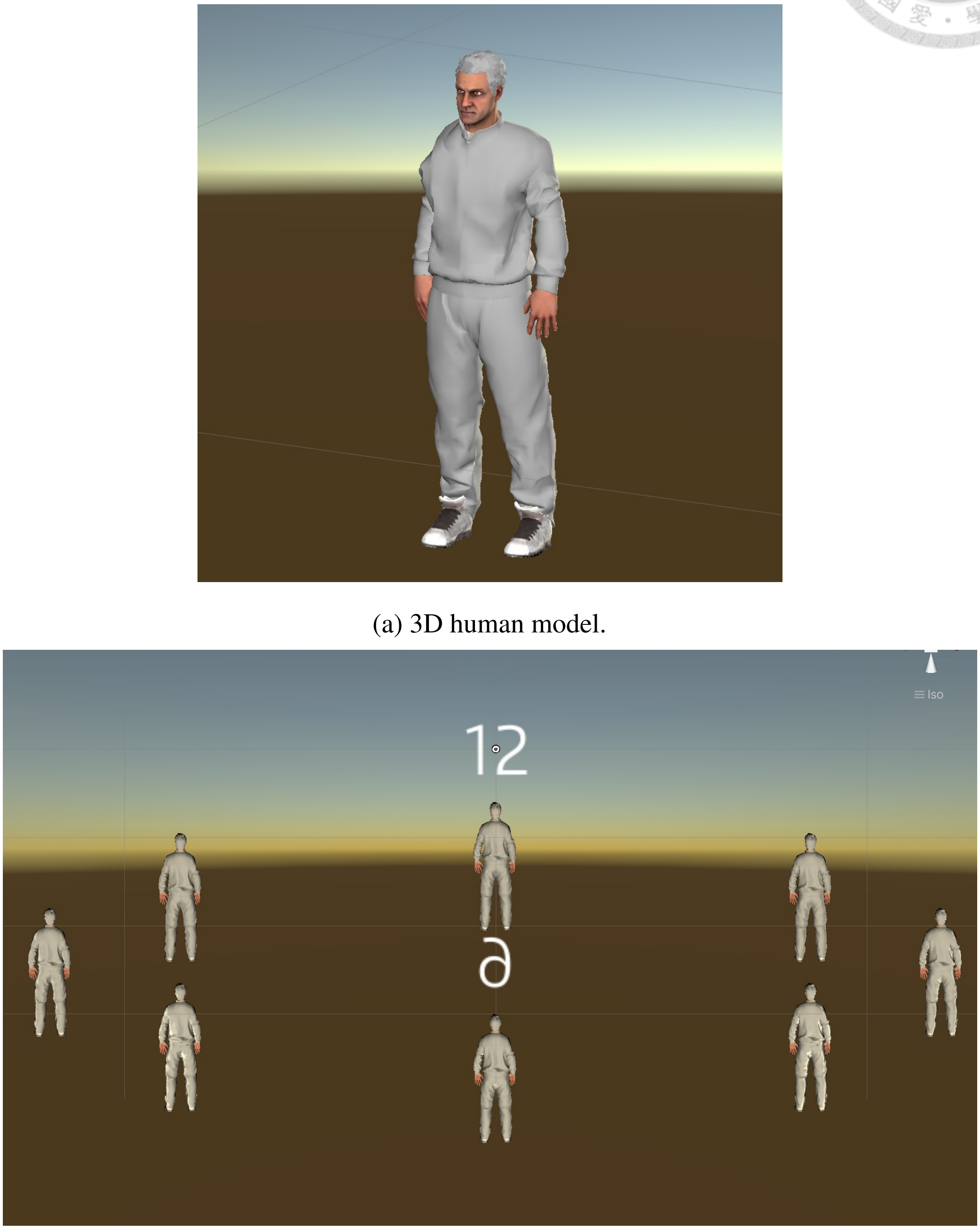
### 3.2.2 Augmented Mirror

When practicing motor skills like TCC, a mirror plays an important role for a learner to enhance spatial awareness. [24] In previous works  [20, 23, 24], the augmented mirror is proposed to provide further information like human pose joints and visual guidance. Thus, we follow the idea and put a virtual mirror in front of the learner in the virtual environment (Figure 3.2).

### 3.2.3 User Interface of Learner

For the learner to control the subsystem, we provide a hand panel (Figure 3.3) in the AR environment, which could be called out by the special gesture. Through the panel, the learner is able to adjust the mode of learning, TCC Forms, the play settings, and the position of virtual coaches.

doi:10.6342/NTU202302869



(a) 3D human model.

(b) Virtual coach layout.

Figure 3.1: Virtual coach.

doi:10.6342/NTU202302869

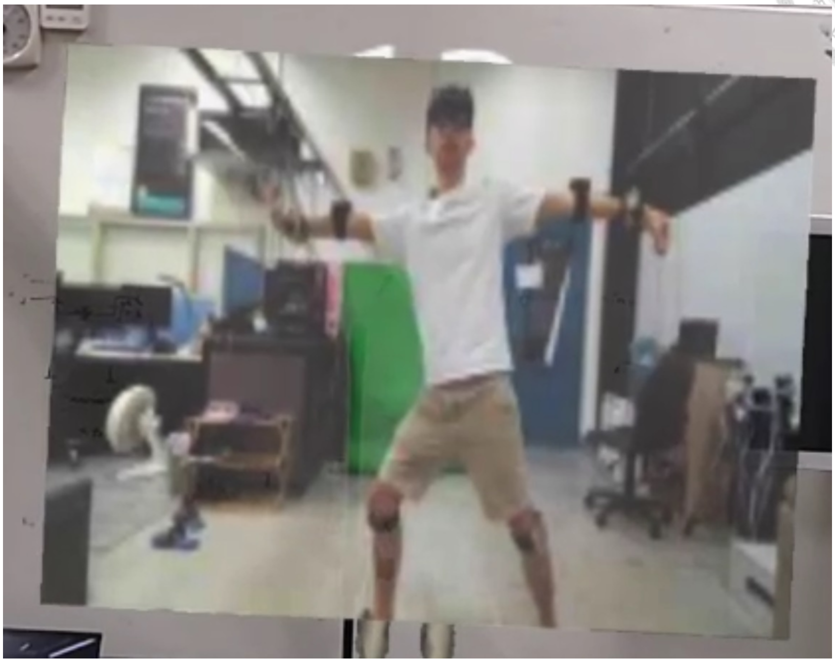
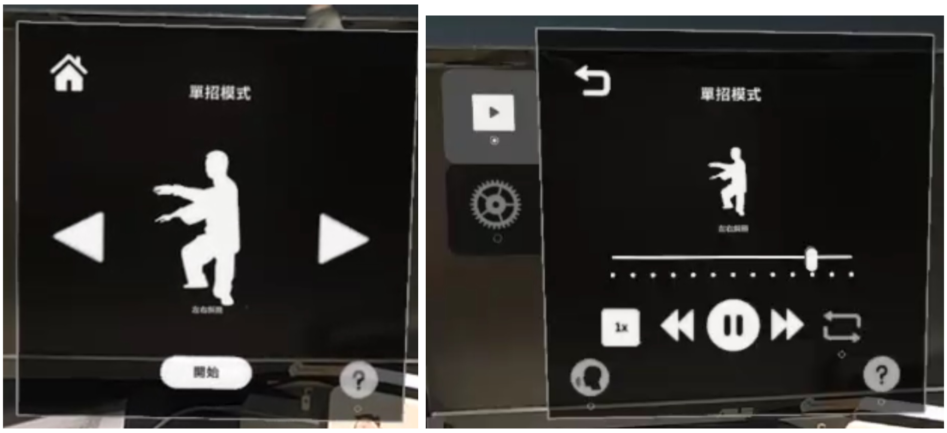


Figure 3.2: Augmented mirror.



(a) Teaching material setting.

(b) Play setting.

Figure 3.3: User Interface for learner (Hand panel).

doi:10.6342/NTU202302869

## 3.3 Guidance Subsystem

Before we design the guidance subsystem, we analyze the remote guiding routine of TCC and disassemble it into 4 steps: receiving body data, providing annotations, receiving annotations, and adjusting body posture. The first 2 steps are from the perspective of the instructor while the last 2 are from the perspective of the learner. The whole procedure keeps looping until the posture of the learner is correct.

### 3.3.1 Receiving Body Data

The first step of the guidance subsystem is to capture the body data of the learner. We follow previous works about body movement guiding [20, 21, 22] and use the full-body skeleton as the target to put annotations on.

### 3.3.2 Providing Annotations

The next step is to display the body data and provide a GUI for the instructor to make annotations as guidance. The layout is shown in Figure 3.4. On the left of the layout is the 2D video stream of the learner with 2D joint position data. On the right is the 3D joint position data, which supports yaw rotation. When the instructor selects a joint as the target point (interesting point), the color of the joint point will change to yellow. Additionally, a gray point followed by an axis icon appears at the same time, which stands for the destination point. The instructor can move the destination point (gray point) by the axis icon. This action tells the learner to move the target point (yellow point) to the destination point (gray point). There is also a gray line between the two points as the nearest path.

### 3.3.3 Receiving Annotations

Thirdly, the annotations are sent to the learner's HMD and shown in the virtual environment (Figure 3.5). Two types of annotations are generated in the view of the learner. For 2D annotations, we refer to Li's work [9] and make the target point

doi:10.6342/NTU202302869

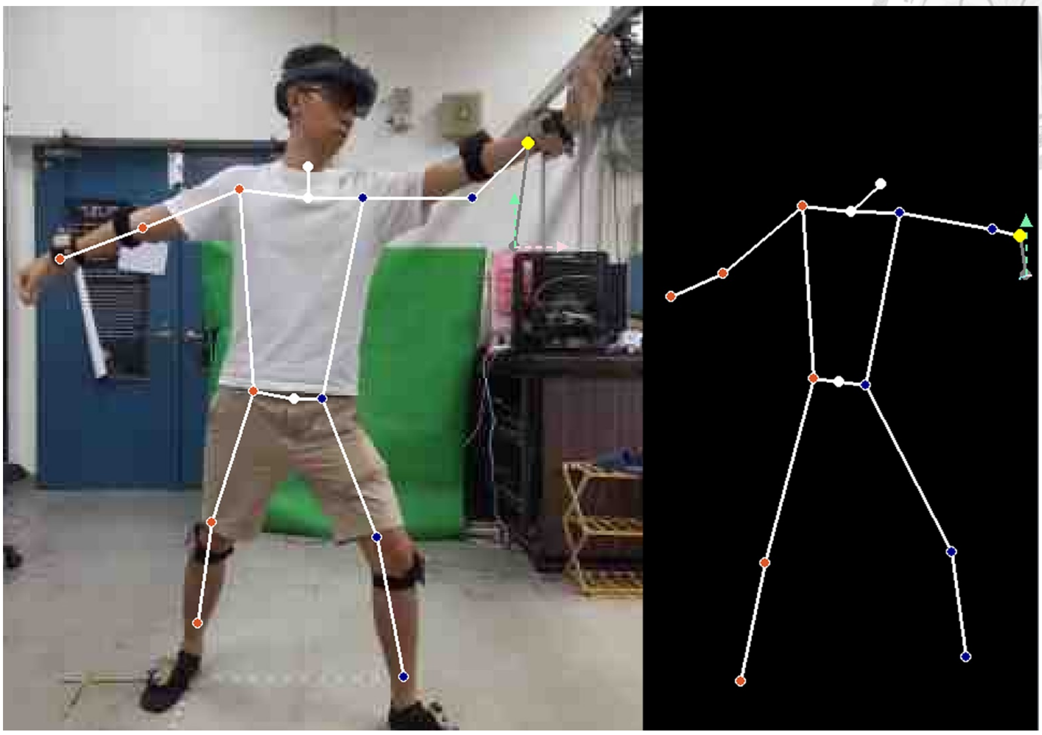


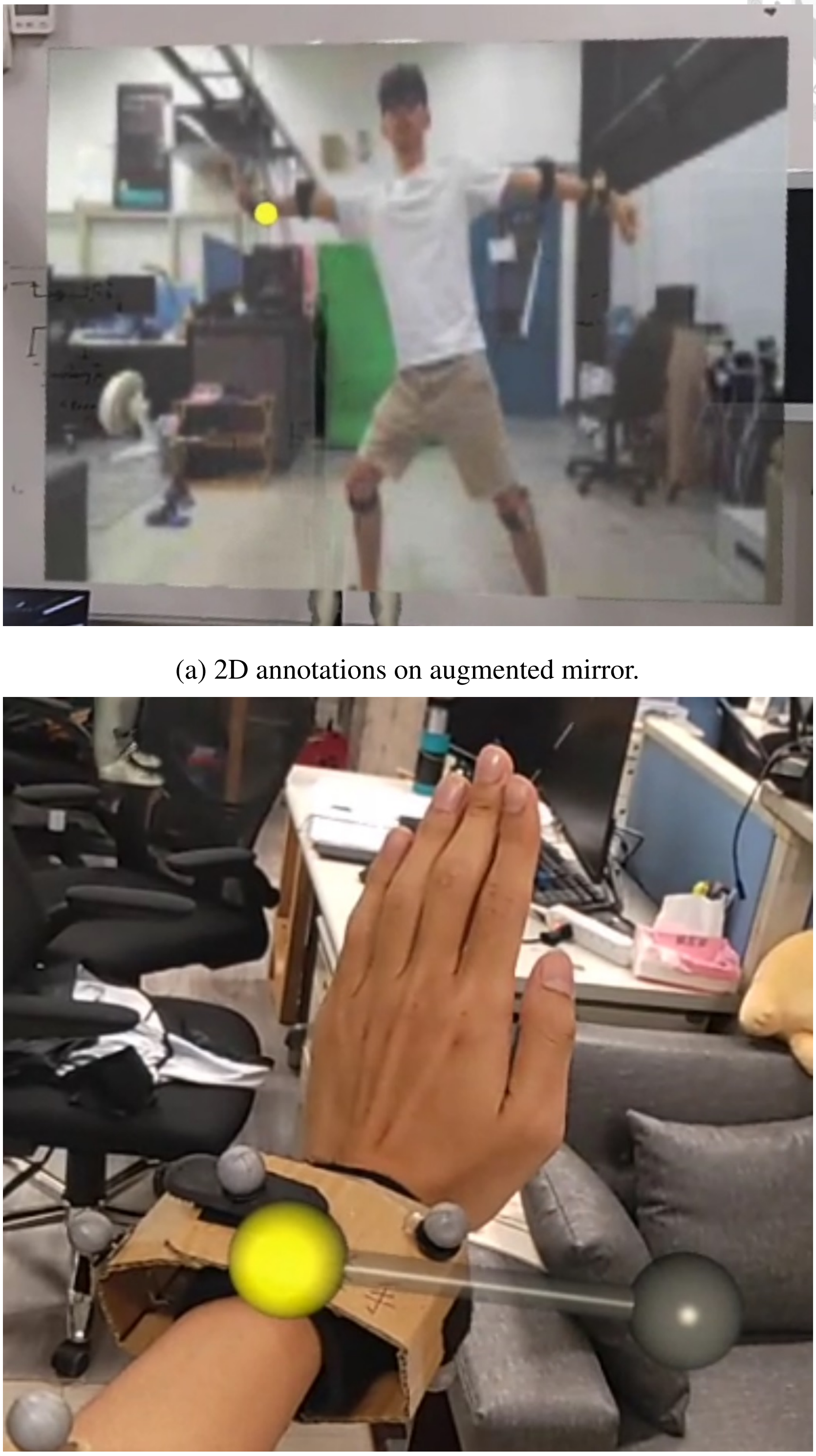
Figure 3.4: User Interface for Instructor.

(yellow point) displayed on the augmented mirror. However, we also improve it so that the 2D annotations will follow the target joint in real-time. It avoids the annotation position differing from the moving joint when the learner notices it. As for the 3D on-body annotations, the learner is able to see the yellow sphere (target point), the grey sphere (destination point), and the grey cylinder (nearest path) imposed on the body. Moreover, in previous work [11, 31, 20], audio notification is known to provide awareness of the instructor's action. Thus, we add a sound effect in the system that is played when the annotations appear.

### 3.3.4 Adjusting Body Posture

In the final step, the learner is able to adjust the posture according to the 2D and 3D annotations in the virtual environment.

doi:10.6342/NTU202302869



(a) 2D annotations on augmented mirror.

(b) 3D on-body annotations.

Figure 3.5: Annotations in virtual environment.

doi:10.6342/NTU202302869

# Chapter 4

# Implementation

The system consists of 4 parts: the sensor side, server side, instructor side, and learner side. Wi-Fi 5 (802.11ac) (Tx Rate: 200/200 Mbps) is used to connect each part. Figure 4.1 shows the system architecture.

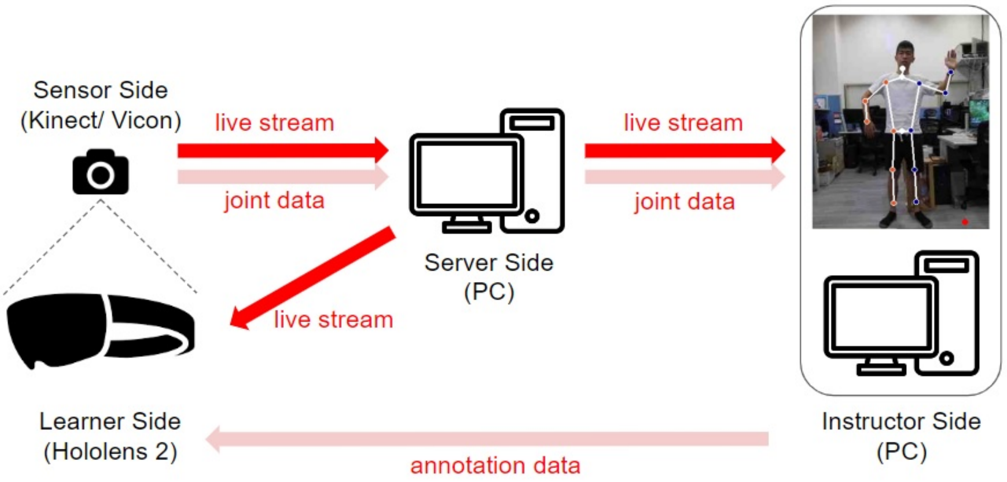


Figure 4.1: System architecture. The red arrows stand for large data transferring (images) while the pink arrows represent small data transferring.

doi:10.6342/NTU202302869

## 4.1 Sensor Side

The sensor side is responsible for capturing the body data of the learner. Our application requires high accuracy and stability since the learner has to keep looking at the augmented annotation when adjusting the posture. We have tried Mediapipe [32], a machine learning framework supporting 3D full-body joint detection which relies on a single RGB camera. However, the depth data of the joint point exists unacceptable error. Then, we tried Azure Kinect, an RGB-D camera with a built-in 3D full-body joint detection model. It comes out relatively accurate 3D position, but the data jitters a lot. Hence, we use Vicon Tracker (Figure 4.2), a motion capture system, as a temporal supporting method. In the final version, we apply one single Azure Kinect to capture the overall 2D images and 2D/ 3D joint position. For partial joints' 3D position, we use Vicon Tracker to get more accurate and stable results. Openpose-related methods like [33, 34] are also feasible, which provide accurate 3D joint data. These methods need multiple but low-cost cameras, which can be a replacement for the role of Kinect in our system.

## 4.2 Server Side

On the server side, we use a laptop (Intel i7-12700H, 32GB RAM, GeForce RTX 3070Ti) and build the application with Python (3.10.9). The application collects the body data from the sensor side. Then, it transfers the live camera image and 2D/ 3D joint data to the instructor side and the learner side.

## 4.3 Instructor Side

On the instructor side, we use a PC (Intel i7-10700K, 32GB RAM, GeForce RTX 2080Ti). We build the application with Python (3.10.9), which receives the live camera image and 2D/ 3D joint data. We also generate a GUI by Pygame module

doi:10.6342/NTU202302869

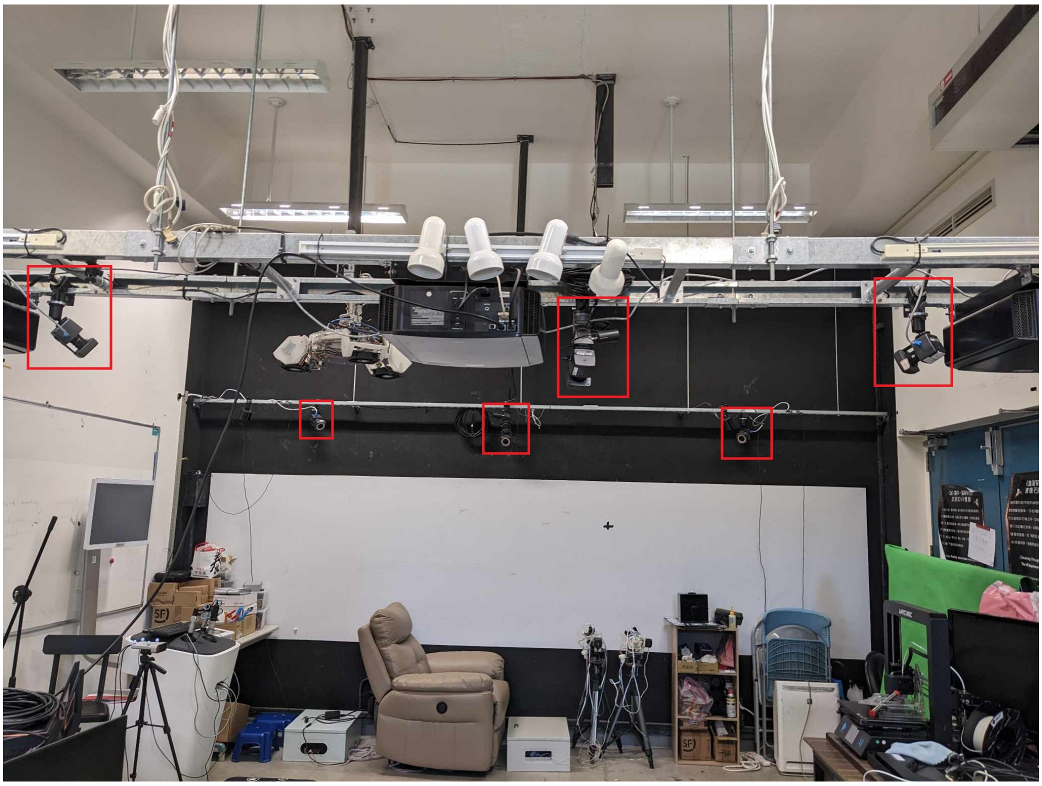


Figure 4.2: Vicon Setup. There are 6 Vicon cameras in the area.

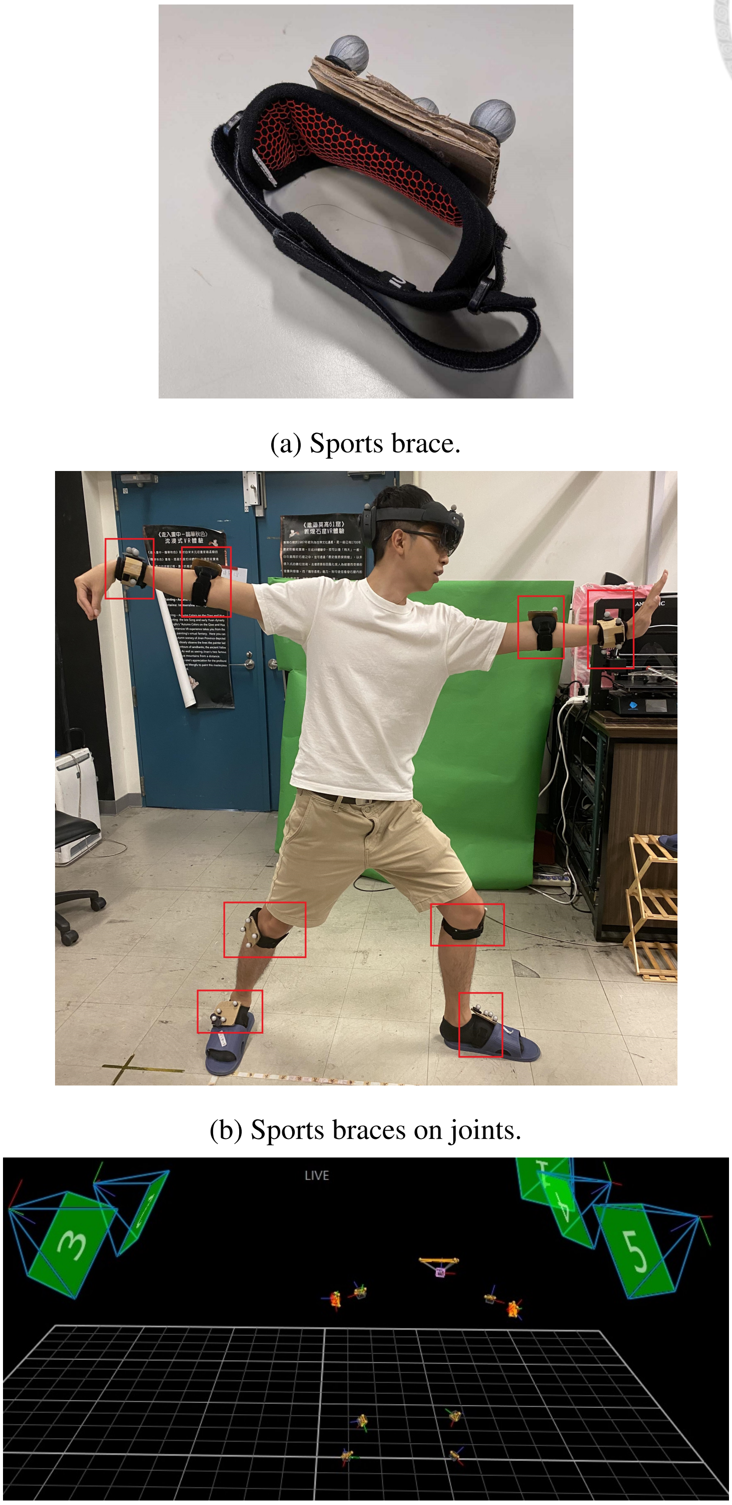
doi:10.6342/NTU202302869

for the instructor to make annotations. The annotation data is then sent to the learner side.

## 4.4 Learner Side

On the learner side, we use a Hololens 2 as the AR HMD. We develop the application with the Unity engine (2019.4.10f1) and Mixed Reality Toolkit (MRTK) to build the virtual environment for augmented annotations. The application receives the image from the server side and the annotation data from the instructor side. The former is made into the virtual mirror and the latter is used to locate the annotations. In order to get accurate and stable 3D joint data, we use Vicon Tracker as the supportive method. The learner then needs to wear sports braces (Figure 4.3) with Vicon markers on the wrists, elbows, knees, and ankles. The reason we choose the 8 joints is that they are the relatively visible part of the body through the limited Field of View (FOV) of Hololens 2. For other joints like shoulders, even though there are 3D on-body annotations on them, it is hard for the learner to observe.

doi:10.6342/NTU202302869



(a) Sports brace.

(b) Sports braces on joints.

(c) Sports braces detected by Vicon Tracker.

Figure 4.3: Sports braces and specific joints.

doi:10.6342/NTU202302869

# Chapter 5

# User Study

In our work, there are 2 subsystems included: the self-learning subsystem and the guidance subsystem. Since the self-learning subsystem follows the works of previous research, the effectiveness is guaranteed. Therefore, in the user study, we focused on the guidance subsystem as a communication tool for pose guidance. We conducted 2 user studies to make separate evaluations from the perspective of the instructor and the learner.

## 5.1 User Study 1

In user study 1, we compared the teaching experience of the instructor when using 2 different types of annotation in the guidance subsystem. We selected 4 Forms in TCC (左斜飛 Left Diagonal Flying, 白鶴亮翅 White Crane Spreads its Wings, 翻身單鞭 Single Whip, and 十字手 Cross Hands) for the instructor to provide guidance.

### 5.1.1 Participant

We recruited 1 TCC teacher (male, aged 78) as a participant, who has taught TCC for many years.

doi:10.6342/NTU202302869

### 5.1.2 Experimental Condition

We tested two different annotation types as our experimental condition: *2D annotation* and *3D annotation*. The *2D annotation* contained the target point (yellow point) on the GUI of the instructor and on the augmented mirror in view of the learner. In contrast, the *3D annotation* contained the target point (yellow point), the destination point (gray point), and the nearest path (gray line) on the GUI of the instructor and the augmented mirror in view of the learner. The on-body 3D spheres imposed on the joint of the learner were also included.

### 5.1.3 Scale

We designed the questionnaire with System Usability Scale (SUS) ( A.1) [35], NASA Task Load Index (NASA-TLX) ( A.2) [36], a question asking about perceived message understanding from Harms and Biocca's work [37], helpfulness of instructing the movement, helpfulness of teaching TCC, and preference ( A.3).

### 5.1.4 Procedure

Before the study started, we asked the learner to put on the Hololens 2 and our 8 sports braces with Vicon marker on the 8 joints (wrists, elbows, knees, and ankles). We then taught the participant to use the GUI and gave him 5 minutes to familiarize the two experimental conditions. After the participant and the learner were ready, we played the animation of the virtual coaches. The learner followed the motion until they stopped at the specific Forms. The participant then started to provide guidance on the key pose through the GUI. During the whole process, the participant and the learner were allowed to speak to each other. For each experimental condition, the participant was asked to guide two Forms and fill out the questionnaire.

doi:10.6342/NTU202302869

## 5.2 User Study 2

In user study 2, we on the other hand compared the learning experience of the learner when using 2 different types of annotation in the guidance subsystem. The same 4 Forms were selected for the learner to mimic and receive guidance.

### 5.2.1 Participant

We recruited 12 people (6 males and 6 females) aged from 19 to 28 ($Mean = 21.33$, $SD = 2.5$) as participants. All of them had the experience of learning TCC.

### 5.2.2 Experimental Condition

Same as user study 1, we tested two different annotation types as our experimental condition: *2D annotation* and *3D annotation*.

### 5.2.3 Scale

We designed the questionnaire in contrast to user study 1, which included SUS, NASA-TLX, perceived message understanding, helpfulness of adjusting the movement, helpfulness of learning TCC, and preference.

### 5.2.4 Procedure

The procedure was similar to user study 1, but the role of the participant was switched to the learner.

doi:10.6342/NTU202302869

# Chapter 6

# Result and Discussion

Through the user study, we compare the 2 experimental conditions with 5 different scales and preferences. The system usability scale (SUS) ranges from 0 to 100, where higher numbers indicate positive feedback. The task load index (raw-TLX) ranges from 0 to 100, where a smaller number indicates positive feedback. The following three scales (perceived message understanding, helpfulness of instructing/ adjusting the movement, and helpfulness of teaching/ learning TCC) are 5-Likert scale, where 5 indicates positive feedback.

## 6.1 User Study 1

In user study 1, *2D* and *3D annotation* are compared from the instructor side. Table 6.1 and 6.2 show the overall result of the questionnaire. The participant rated both *2D* and *3D annotation* positively on all scales. Though there does not exist a big difference, the score of the *3D annotation* is slightly higher than the *2D annotation* and becomes the preferable one. In feedback from open-ended questions, the participant mentioned that he preferred the *3D annotation* because it provided more information about the learner's body data and more features of annotations. The participant also gave us some advice about the whole system. Firstly, more information on specific body parts is needed, such as the middle finger, which is

doi:10.6342/NTU202302869

Table 6.1: Result of User Study 1.

| | 2D Annotation | 3D Annotation |
|---|---|---|
| SUS | 72.5 | **80** |
| raw-TLX | 8.33 | 8.33 |
| Perceived Message Understanding | 4 | **5** |
| Helpfulness of Instructing the Movement | 4 | **5** |
| Helpfulness of Teaching TCC | 4 | 4 |

Table 6.2: Result of User Study 1 (Preference).

| | 2D Annotation | 3D Annotation |
|---|---|---|
| Preference | 0 | **1** |

doi:10.6342/NTU202302869

also important in TCC Forms. Secondly, the GUI on a 2D display (PC) is hard to represent the 3D joint position. Finally, in the opinion of a TCC teacher, he believed that the system is more suitable for students who have learned TCC before. For beginners, the teaching lesson should include some haptic feedback, which is not supported in our system. We followed this suggestion to find experienced learners when conducting the user study 2.

## 6.2 User Study 2

Table 6.3: Result of User Study 2.

| | 2D Annotation | 3D Annotation | |
|---|---|---|---|
| | Mean (Stdev) | Mean (Stdev) | p-Value |
| SUS | 69.38 (19.66) | **78.33 (10.41)** | 0.08 |
| raw-TLX | 17.01 (19.66) | **7.99 (7.63)** | 0.07 |
| Perceived Message Understanding | 4.08 (1.08) | **4.83 (0.4)** | 0.02 |
| Helpfulness of Adjusting the Movement | 4 (0.6) | **4.5 (0.8)** | 0.06 |
| Helpfulness of Learning TCC | 3.83 (0.83) | **4.5 (0.67)** | 0.01 |

Table 6.4: Result of User Study 2 (Preference).

| | 2D Annotation | 3D Annotation |
|---|---|---|
| Preference | 5 | **7** |

In user study 2, *2D* and *3D annotation* are compared from the learner side. Table 6.3 and 6.4 show the result. For the scores of SUS and raw-TLX, our

doi:10.6342/NTU202302869

system received high reviews on both *2D* and *3D annotation*. Though there was no significant difference, the participants rated the *3D annotation* slightly higher. However, the standard deviation is also high in each condition. We then analyzed the feedback from open-ended questions and found that the opinions contained various aspects of our system. The participant gave feedback on the observing angle of the annotations (3), perceived direction (1), numbers of voice prompts needed (2), issues of small FOV (6), amount of information contained (5), and annotations being understandable in a different level (5). The feedback inferred that there were lots of changes between *2D* and *3D annotation*. Therefore, when the participants focused on different parts of the changes, their perceived score of the SUS and raw-TLX would exist gaps. We also considered it as the reason why there is no huge difference between the numbers of preferences. Thus, in future studies, variables in the *2D* and *3D annotation* should be considered and designed more carefully. In addition, we believe that the *3D annotation* will be more acceptable along with the new AR HMD with bigger FOV appearing in the future. It is because the most mentioned negative feedback comes from the small FOV (6), which makes the participants frequently rotate their heads.

As for the result of the helpfulness of adjusting the movement, there is no significant difference, which faces the same issue of the SUS and raw-TLX. Nonetheless, the participants agreed that the *3D annotation* provides more information for them to understand the instructions ($p < 0.05$), which is also mentioned by 5 participants in the interview. Finally, the participants think the *3D annotation* is more helpful ($p < 0.05$) for learning TCC.

doi:10.6342/NTU202302869

# Chapter 7

# Conclusion and Future Work

For TCC remote learning, we design a collaboration system combining AR and WIMP, which includes the advantage of video conferencing and TCC self-learning systems. It contains the AR self-learning subsystem and the guidance subsystem with 2D annotations on the augmented mirror and 3D on-body annotations. To evaluate our work, we conduct 2 user studies to compare the 2 types of annotation from the perspective of the instructor and the learner. The result shows that the participants are satisfied with our system with both types of annotations. The 3D annotation receives a slightly higher rating and performs better in several parts. For the instructor, it provides more body data and instruction features. For the learner, it provides more instruction information and more helpfulness for TCC learning.

One limitation of our work is that we focus on the still poses (Forms) in TCC. The assessment of changes between each Form is a feature that we can add in the future. In this scenario, the recording feature may be needed so that the instructor and the learner are able to replay the movement for teaching and learning. Another limitation is that the learner has to wear gadgets for body detection. It is due to the human pose estimation model being not accurate enough to support our application. However, along with the model becoming more stable and precise, our system may be more feasible and acceptable. Another interesting point is that if the 3D real-time human reconstruction becomes mature, it is also possible to replace the

doi:10.6342/NTU202302869

PC with VR HMD on the instructor side to provide a more intuitive way of making guidance. In future studies, we can also apply our system to other motor skill learning (e.g. yoga, dancing, and workout) that also focuses on pose correction.

doi:10.6342/NTU202302869

# References

[1] P. T. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, T. Camill, J. Hodgins, and R. Pausch, "Training for physical tasks in virtual environments: Tai chi," in *IEEE Virtual Reality, 2003. Proceedings.* IEEE, 2003, pp. 87–94. 2, 3

[2] T. Iwaanaguchi, M. Shinya, S. Nakajima, and M. Shiraishi, "Cyber tai chi-cg-based video materials for tai chi chuan self-study," in *2015 International Conference on Cyberworlds (CW).* IEEE, 2015, pp. 365–368. 2, 3

[3] P.-H. Han, Y. Zhong, H.-L. Wang, and Y.-P. Hung, "Augmented visualization for practicing tai-chi chuan with drone-enhanced approach," in *Oz CHI*, 2016. 2, 3, 8

[4] P.-H. Han, Y.-S. Chen, Y. Zhong, H.-L. Wang, and Y.-P. Hung, "My tai-chi coaches: an augmented-learning tool for practicing tai-chi chuan," in *Proceedings of the 8th Augmented Human International Conference*, 2017, pp. 1–4. 2, 3, 8

[5] X. Chen, Z. Chen, Y. Li, T. He, J. Hou, S. Liu, and Y. He, "Immertai: Immersive motion learning in vr environments," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 416–427, 2019. 2, 3

[6] Y.-F. Jan, K.-W. Tseng, P.-Y. Kao, and Y.-P. Hung, "Augmented tai-chi chuan practice tool with pose evaluation," in *2021 IEEE 4th International*

doi:10.6342/NTU202302869

*Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2021, pp. 35–41. 2, 3, 8

[7] P.-Y. Kao, P.-H. Han, Y. F. Jan, Z. Yang, C.-H. Li, and Y.-P. Hung, "On learning weight distribution of tai chi chuan using pressure sensing insoles and mr-hmd." in *VR*, 2019, pp. 1457–1464. 3

[8] Y.-F. Jan, "Tai-chi chuan footstep guiding for augmented reality assisted learning system," Master's thesis, National Taiwan University, 2022. 4, 8

[9] W.-X. Li, "Remote collaborative tai-chi chuan le arning with mixed reality," Master's thesis, National Taiwan University, 2022. 4, 8, 11

[10] P. Wang, X. Bai, M. Billinghurst, S. Zhang, X. Zhang, S. Wang, W. He, Y. Yan, and H. Ji, "Ar/mr remote collaboration on physical tasks: A review," *Robotics and Computer-Integrated Manufacturing*, vol. 72, p. 102071, 2021. 4, 7

[11] M. Cidota, S. Lukosch, D. Datcu, and H. Lukosch, "Workspace awareness in collaborative ar using hmds: a user study comparing audio and visual notifications," in *Proceedings of the 7th Augmented Human International Conference 2016*, 2016, pp. 1–8. 4, 12

[12] H. Sun, Z. Zhang, Y. Liu, and H. B. Duh, "Optobridge: assisting skill acquisition in the remote experimental collaboration," in *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, 2016, pp. 195–199. 4

[13] H. Sun, Y. Liu, Z. Zhang, X. Liu, and Y. Wang, "Employing different viewpoints for remote guidance in a collaborative augmented environment," in *Proceedings of the sixth international symposium of Chinese CHI*, 2018, pp. 64–70. 4

doi:10.6342/NTU202302869

[14] S. Wang, M. Parsons, J. Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. Meruvia-Pastor, M. Gong, and A. Smith, “Augmented reality as a telemedicine platform for remote procedural training,” *Sensors*, vol. 17, no. 10, p. 2294, 2017. 4

[15] A. Gupta, S. Mohatta, J. Maurya, R. Perla, R. Hebbalaguppe, and E. Hassan, “Hand gesture based region marking for tele-support using wearables,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 69–75. 4

[16] S. Günther, S. Kratz, D. Avrahami, and M. Mühlhäuser, “Exploring audio, visual, and tactile cues for synchronous remote assistance,” in *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 2018, pp. 339–344. 4

[17] S. Kim, M. Billinghurst, C. Lee, and G. Lee, “Using freeze frame and visual notifications in an annotation drawing interface for remote collaboration,” *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 12, no. 12, pp. 6034–6056, 2018. 4

[18] S. Kim, M. Billinghurst, and G. Lee, “The effect of collaboration styles and view independence on video-mediated remote collaboration,” *Computer Supported Cooperative Work (CSCW)*, vol. 27, pp. 569–607, 2018. 4

[19] R. Sodhi, H. Benko, and A. Wilson, “Lightguide: projected visualizations for hand movement guidance,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 179–188. 5

[20] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, “Youmove: enhancing movement training with an augmented reality mirror,” in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, 2013, pp. 311–320. 5, 8, 11, 12

doi:10.6342/NTU202302869

[21] E. Velloso, A. Bulling, and H. Gellersen, "Motionma: motion modelling and analysis by demonstration," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 1309–1318. 5, 11

[22] C. Sieluzycki, P. Kaczmarczyk, J. Sobecki, K. Witkowski, J. Maśliński, and W. Cieśliński, "Microsoft kinect as a tool to support training in professional sports: augmented reality application to tachi-waza techniques in judo," in *2016 Third European Network Intelligence Conference (ENIC)*. IEEE, 2016, pp. 153–158. 5, 11

[23] R. Tang, X.-D. Yang, S. Bateman, J. Jorge, and A. Tang, "Physio@ home: Exploring visual guidance and feedback techniques for physiotherapy exercises," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 4123–4132. 5, 8

[24] M. Sousa, J. Vieira, D. Medeiros, A. Arsenio, and J. Jorge, "Sleevear: Augmented reality for rehabilitation using realtime feedback," in *Proceedings of the 21st international conference on intelligent user interfaces*, 2016, pp. 175–185. 5, 8

[25] C. Ribeiro, R. Kuffner, and C. Fernandes, "Virtual reality annotator: A tool to annotate dancers in a virtual environment," in *Digital Cultural Heritage: Final Conference of the Marie Skłodowska-Curie Initial Training Network for Digital Cultural Heritage, ITN-DCH 2017, Olimje, Slovenia, May 23–25, 2017, Revised Selected Papers*. Springer, 2018, pp. 257–266. 5

[26] E. Wu and H. Koike, "Futurepose-mixed reality martial arts training using real-time 3d human pose forecasting with a rgb camera," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2019, pp. 1384–1392. 5

[27] H. Lee, H. Kim, D. V. Monteiro, Y. Goh, D. Han, H.-N. Liang, H. S. Yang, and J. Jung, "Annotation vs. virtual tutor: Comparative analysis on the

doi:10.6342/NTU202302869

effectiveness of visual instructions in immersive virtual reality," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2019, pp. 318–327. 6

[28] Y. Lu, H. Yu, W. Ni, and L. Song, "3d real-time human reconstruction with a single rgbd camera," *Applied Intelligence*, vol. 53, no. 8, pp. 8735–8745, 2023. 7

[29] Y. Zhou, M. Habermann, I. Habibie, A. Tewari, C. Theobalt, and F. Xu, "Monocular real-time full body capture with inter-part correlations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4811–4822. 7

[30] L. Xu, Z. Su, L. Han, T. Yu, Y. Liu, and L. Fang, "Unstructuredfusion: realtime 4d geometry and texture reconstruction using commercial rgbd cameras," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 10, pp. 2508–2522, 2019. 7

[31] Y. Kim, S. Hong, and G. J. Kim, "Augmented reality-based remote coaching for fast-paced physical task," *Virtual Reality*, vol. 22, pp. 25–36, 2018. 12

[32] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019. 15

[33] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, "Evaluation of 3d markerless motion capture accuracy using openpose with multiple video cameras," *Frontiers in sports and active living*, vol. 2, p. 50, 2020. 15

[34] M. Zago, M. Luzzago, T. Marangoni, M. De Cecco, M. Tarabini, and M. Galli, "3d tracking of human motion using visual skeletonization and stereoscopic

doi:10.6342/NTU202302869

vision," *Frontiers in bioengineering and biotechnology*, vol. 8, p. 181, 2020. 15

[35] J. Brooke, "Sus: a "quick and dirty'usability," *Usability evaluation in industry*, vol. 189, no. 3, pp. 189–194, 1996. 20

[36] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908. 20

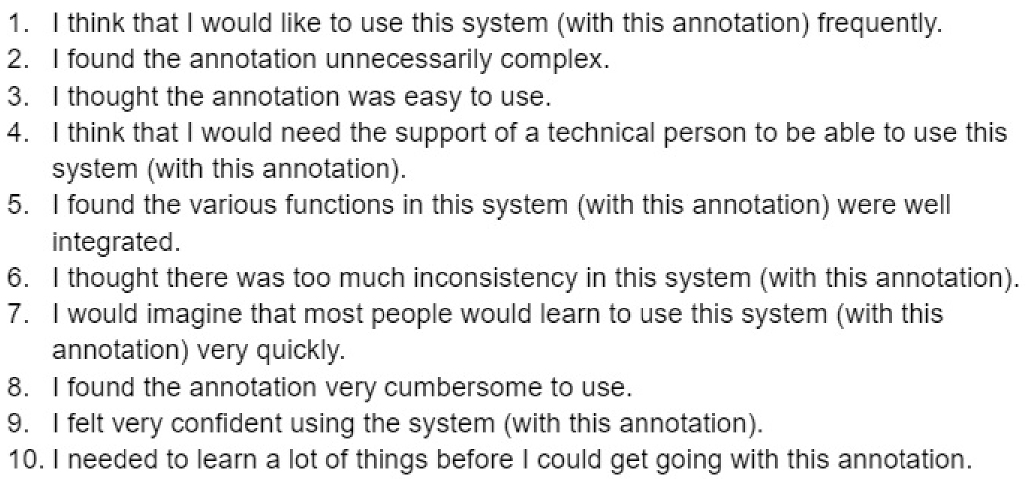
[37] C. Harms and F. Biocca, "Internal consistency and reliability of the networked minds measure of social presence," in *Seventh annual international workshop: Presence*, vol. 2004. Universidad Politecnica de Valencia Valencia, Spain, 2004. 20

doi:10.6342/NTU202302869

# Appendix A

# Questionnaire

## A.1 System Usability Scale (SUS)

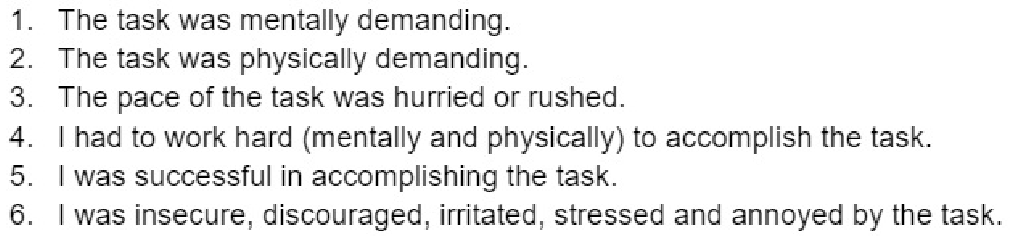


1. I think that I would like to use this system (with this annotation) frequently.
2. I found the annotation unnecessarily complex.
3. I thought the annotation was easy to use.
4. I think that I would need the support of a technical person to be able to use this system (with this annotation).
5. I found the various functions in this system (with this annotation) were well integrated.
6. I thought there was too much inconsistency in this system (with this annotation).
7. I would imagine that most people would learn to use this system (with this annotation) very quickly.
8. I found the annotation very cumbersome to use.
9. I felt very confident using the system (with this annotation).
10. I needed to learn a lot of things before I could get going with this annotation.

Figure A.1: SUS questionnaire (5-point Likert scales).

doi:10.6342/NTU202302869

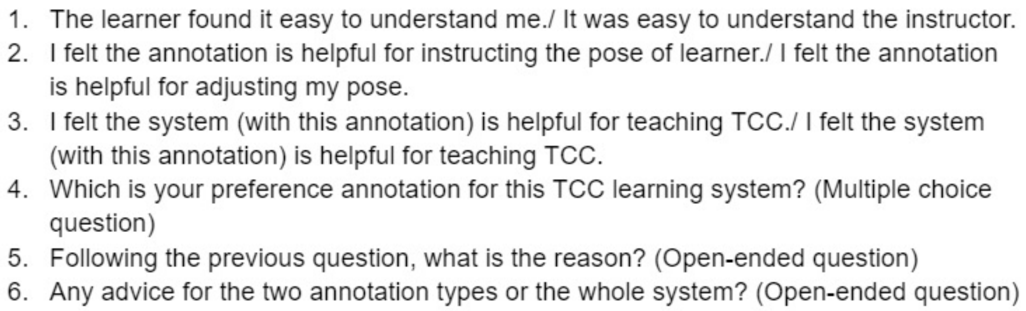
## A.2 NASA Task Load Index (NASA-TLX)



1. The task was mentally demanding.
2. The task was physically demanding.
3. The pace of the task was hurried or rushed.
4. I had to work hard (mentally and physically) to accomplish the task.
5. I was successful in accomplishing the task.
6. I was insecure, discouraged, irritated, stressed and annoyed by the task.

Figure A.2: NASA-TLX questionnaire (5-point Likert scales).

## A.3 Others



1. The learner found it easy to understand me./ It was easy to understand the instructor.
2. I felt the annotation is helpful for instructing the pose of learner./ I felt the annotation is helpful for adjusting my pose.
3. I felt the system (with this annotation) is helpful for teaching TCC./ I felt the system (with this annotation) is helpful for teaching TCC.
4. Which is your preference annotation for this TCC learning system? (Multiple choice question)
5. Following the previous question, what is the reason? (Open-ended question)
6. Any advice for the two annotation types or the whole system? (Open-ended question)

Figure A.3: Questionnaire for others (5-point Likert scales).