

Unity 結合影像辨識之元宇宙場景構建

專題編號：112-CSIE-S004-MID

執行期限：111 年第 1 學期至 112 年第 1 學期

指導教授：謝東儒

專題參與人員： 109590024 廖祉諺

109590028 吳凱印

109590046 李宗穎

109590056 溫立賢

一、摘要

我們這組的專題是使用 **OpenCV** 與 **Mediapipe** 建立動作捕捉，並將其資料傳導入 **Unity**，並透過 **Unity** 去實現元宇宙技術開發；報告內容大致是緣由與目的、研究方法及進度報告與參考文獻；緣由與目的的內容主要是我們這組為何發想這個主題，以及想要達成的目標結果；研究方法內容是我們在專題製作的過程中所使用的工具說明、技術說明、研究步驟及目前進度；參考文獻是在尋找資料時所引用的各式資源來源。

二、緣由與目的

我們這組的專題是透過 **Unity** 去實現元宇宙技術開發，起初的發想是組員們都有在玩遊戲，對最近熱門的元宇宙話題也有著興趣，而選擇的指導教授也契合這個主題，於是便開始了企劃。

對於元宇宙，最直觀的理解就是透過 VR 虛擬實境在遊戲場景中的互動，但是由於 VR 設備昂貴，而對於我們的專題製作上我們希望是不要花費到金錢，所以我們所發想的便是不透過 VR 設備，而是單純依靠鏡頭畫面捕捉對虛擬遊戲環境進行操作，目標便是能夠建立起一個遊戲環境，並製作自己的動作捕捉程式，將兩者結合，透過傳導自己的動作進入元宇宙環境中進行操作，並依照實際建立起的元宇宙環境與操作成果，分析其可行性和與 VR 設備對比的差別。

三、研究方法

工具說明：電腦、攝像頭。

研究範圍：強調透過 Python 實現捕捉動態姿勢及其獲取之數據與 Unity 環境建立並與之結合方法。

使用技術方法：

- **OpenCV**

是一個跨平台的電腦視覺庫。OpenCV 是由英特爾公司發起並參與開發，可以在商業和研究領域中免費使用。OpenCV 可用於開發即時的圖像處理、電腦視覺以及圖型識別程式。

- **Mediapipe**

是 Google Research 所開發的多媒體機器學習模型應用框架，支援 JavaScript、Python、C++ 等程式語言，可以運行在嵌入式平臺(例如樹莓派等)、移動設備(iOS 或 Android)或後端伺服器，使用 Python 語言進行開發，MediaPipe 支援下列幾種辨識功能：手部追蹤、人臉網格、手掌偵測、全身偵測、姿勢偵測、物體偵測、人物去背等功能。

- **Unity**

是一款跨平台的 2D 和 3D 遊戲引擎，由 Unity Technologies 研發，可開發跨平台的電動遊戲。除可以用於研發電子遊戲之外，Unity 還廣泛用作建築視覺化、即時三維動畫等類型互動內容的綜合型創作工具。

研究流程：

決定研究方向：與組員們討論並規劃想要實現的目標，並分析其可行性以及是否契合指導老師之領域項目主題。

規劃時程：由於組員們的時間安排皆不同，於是先共同討論出一個共識，規劃一個共同接受的時間段進行討論。

文獻探討與分析：分工蒐集相關文獻與資料，透過討論初步篩選後進行歸納整理，以做為本研究之立論基礎。

建立研究架構：研究定案後，即深入分析整理文獻資料，討論並排除不需要的資料，而後進行資料統整，據以建立研究架構。

專題實作：先安排好分工，並擬定進度表，規劃在時程內該完成的進度。

透過分析資料及數據進行實作，並在過程中時刻討論每一環節的問題與解決方案。

報告撰寫與回報：在專題的製作過程中調整報告簡報的內容，並推選一人作為和指導老師的對接窗口，負責連絡 meeting 及詢問專題相關事項。

研究結果及結論：

根據我們的研究成果，成功地創建了一個 Unity 元宇宙場景，並成功導入手部追蹤技術，以手勢判斷的方式對音樂進行切換。然而，我們也注意到一些挑戰和限制，其一是手部抖動影響精確度導致直接對實體物件操作困難，其二是真實物理互動(如觸摸、抓取、拖動等)的不足。

總的來說，儘管我們在虛擬世界中實現了一定程度的互動性，但要實現對實體物件的高度接觸和互動操作，仍然需要更先進的技術，例如虛擬實境(VR)。這對於需要更多物理參與的應用場景可能更為適合。

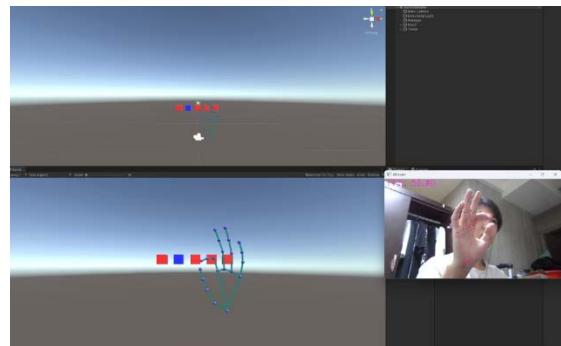


圖 1. Unity 導入姿勢追蹤

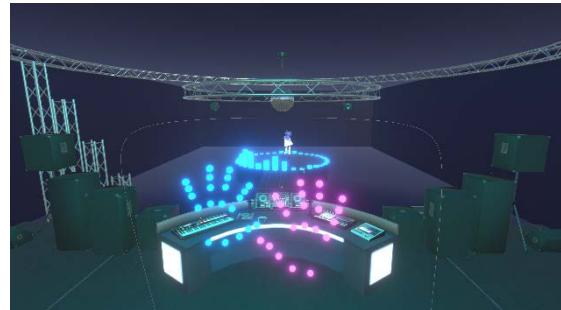


圖 2. 成果場景圖

參考文獻

- [1] Liuhan Ge, Hui Liang, Junsong Yuan, and Daniel Thalmann. Robust 3d hand pose estimation in single depth images: from single-view cnn to multi-view cnns. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3593–3601, 2016.
- [2] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenko, George Sung, Chuo-Ling Chang, Matthias Grundmann, MediaPipe Hands: On-device Real-time Hand Tracking, Google Research.
- [3] Liuhan Ge, Zhou Ren, Yuncheng Li, Zehao Xue, Yingying Wang, Jianfei Cai, and Junsong Yuan. 3d hand shape and pose estimation from a single rgb image. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 10833–10842, 2019.
- [4] George Sung Kanstantsin Sokal Esha Ubweja Valentin Bazarevsky Jonathan Baccash Eduard Gabriel Bazavan Chuo-Ling Chang Matthias Grundmann, On-device Real-time Hand Gesture Recognition, Google Research