

利用IMU改善純影像技術

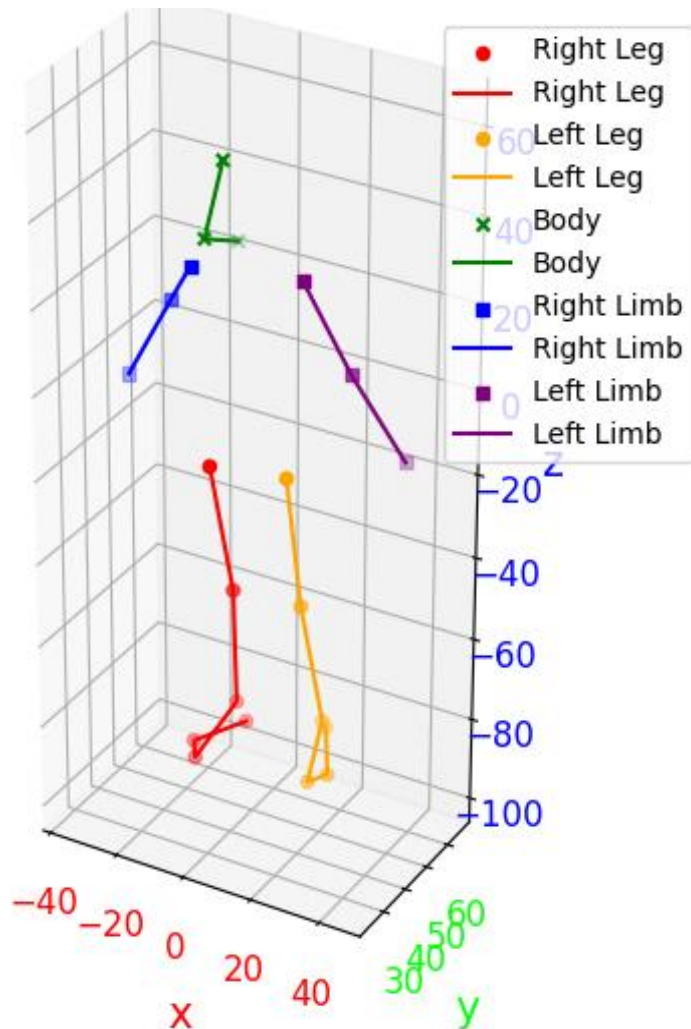
在人體動作重建中的表現

Using IMU to Improve the Performance
of Imaging Technology
in Human Motion Reconstruction

學生：陳珮甄

指導教授：詹魁元 博士

2024/7/30



National Taiwan University
System Optimization Laboratory

大綱

一

背景介紹

二

研究動機與目的

三

文獻回顧

四

探討減少相機使用數量
對於人體姿態估計精準度的影響

五

感測器融合方法及流程

六

實驗結果

七

結論及未來工作

背景介紹

	光標記 動作捕捉	無光標記 動作捕捉	慣性感測器 動作捕捉	多感測器融合 動作捕捉
特點	<ul style="list-style-type: none"> - 量測準確度高 - 易受環境光影響 - 設備價格高昂 - 設備架設不易 	<ul style="list-style-type: none"> - 不受環境光影響 - 設備架設容易 - 辨識模型集資料集影響辨識正確度 - 易受遮擋 	<ul style="list-style-type: none"> - 體積小、重量輕、易攜帶 - 無法直接量得位置 - 易飄移 (drift) 	<ul style="list-style-type: none"> - 可相互補足感測器間的缺點 - 資料處理程序繁雜
誤差	0.15 mm ^[1]	受相機數量影響， 誤差範圍較廣 19.2 ~ 30 mm	經運動學模型轉換 平均誤差 35 mm ^[2]	受感測器數量影響 24 ~ 26 mm
環境	室內	室內、室外	室內、室外	室內、室外
感測器	Vicon、OptiTrack	RGB 相機、 深度相機	慣性感測器 (IMU)	相機、IMU

[1] Merriault, P., Dupuis, Y., Boutteau, R., Vasseur, P., & Savatier, X. (2017). A study of vicon system positioning performance. *Sensors*, 17(7), 1591.

[2] Filippeschi, A., Schmitz, N., Miezal, M., Bleser, G., Ruffaldi, E., & Stricker, D. (2017). Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion. *Sensors*, 17(6), 1257.

研究動機與目的

IMU
輔助

證實加上 IMU 資訊可有效提升重建成功率

減少
量測設備需求量

減少量測時需使用的相機數量

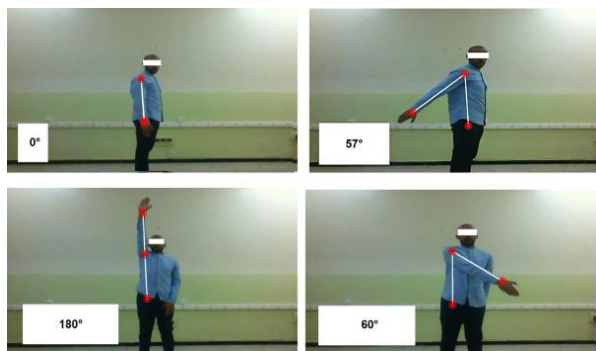
可在
不受限的任意環境
重建動作

解除對光標記動作捕捉系統的依賴

文獻回顧

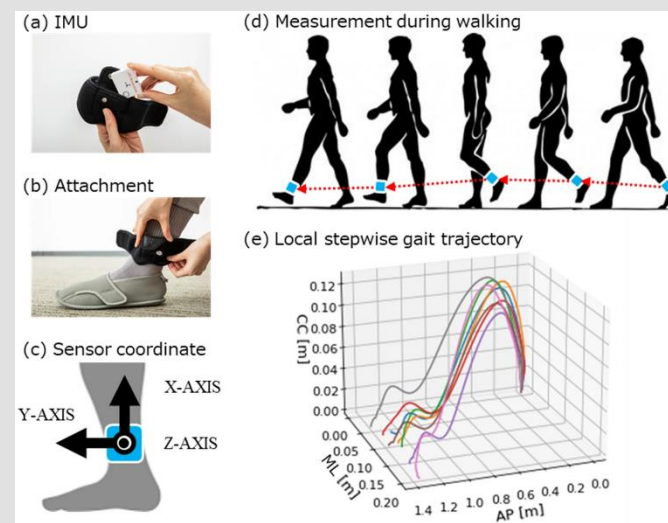
無標記動作捕捉系統

- 重建人體姿態
- 估計關節角度 [3]



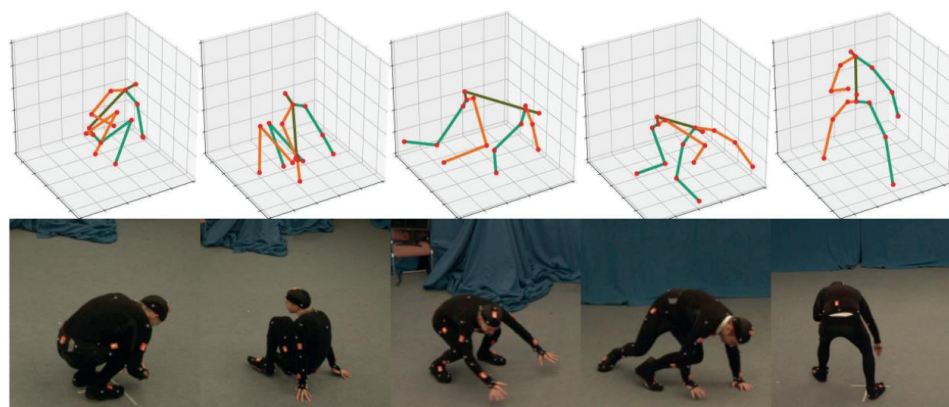
慣性動作捕捉系統

- 基於骨骼結構模型重建人體姿態
- 步態分析，用於評估步態障礙患者的康復情況 [4]



多感測器融合動作捕捉系統

- 使用4台攝影機，及8個IMU資料融合，重建人體姿態 [5]



[3] Latreche, A., Kelaiaia, R., Chemori, A., & Kerboua, A. (2023). Reliability and validity analysis of MediaPipe-based measurement system for some human rehabilitation motions. *Measurement*, 214, 112826.

[4] Uchitomi, H., Hirobe, Y., & Miyake, Y. (2022). Three-dimensional continuous gait trajectory estimation using single shank-worn inertial measurement units and clinical walk test application. *Scientific Reports*, 12(1), 5368.

[5] Zhang, Z., Wang, C., Qin, W., & Zeng, W. (2020). Fusing wearable imus with multi-view images for human pose estimation: A geometric approach. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2200-2209).

探討減少相機使用數量

對於人體姿態估計精準度的影響 – 方法

使用感測器融合方法重建人體姿態可選擇要使用的硬體數量，假設使用的 IMU 數量固定，相機數量及擺放位置則影響人體姿態重建的準確度，因此本研究試圖探討相機數量對於人體姿態估計精準度的影響，以增加實驗方便性

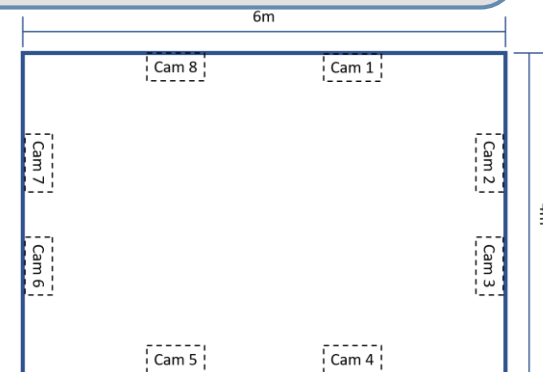
實驗工具

資料集：

TotalCapture Dataset [6]，提供八台相機資料、13 個 IMU 資料、Vicon 資料

感測器融合工具：

學者 Zhe Zhang 等人提出的感測器融合方法



實驗方法

IMU：固定使用資料集中的左右上臂、前臂、大腿、小腿，共 8 個 IMU
相機：

任選 x 台相機

與 IMU 融合

計算 MPJPE
(mean per joint
position error)

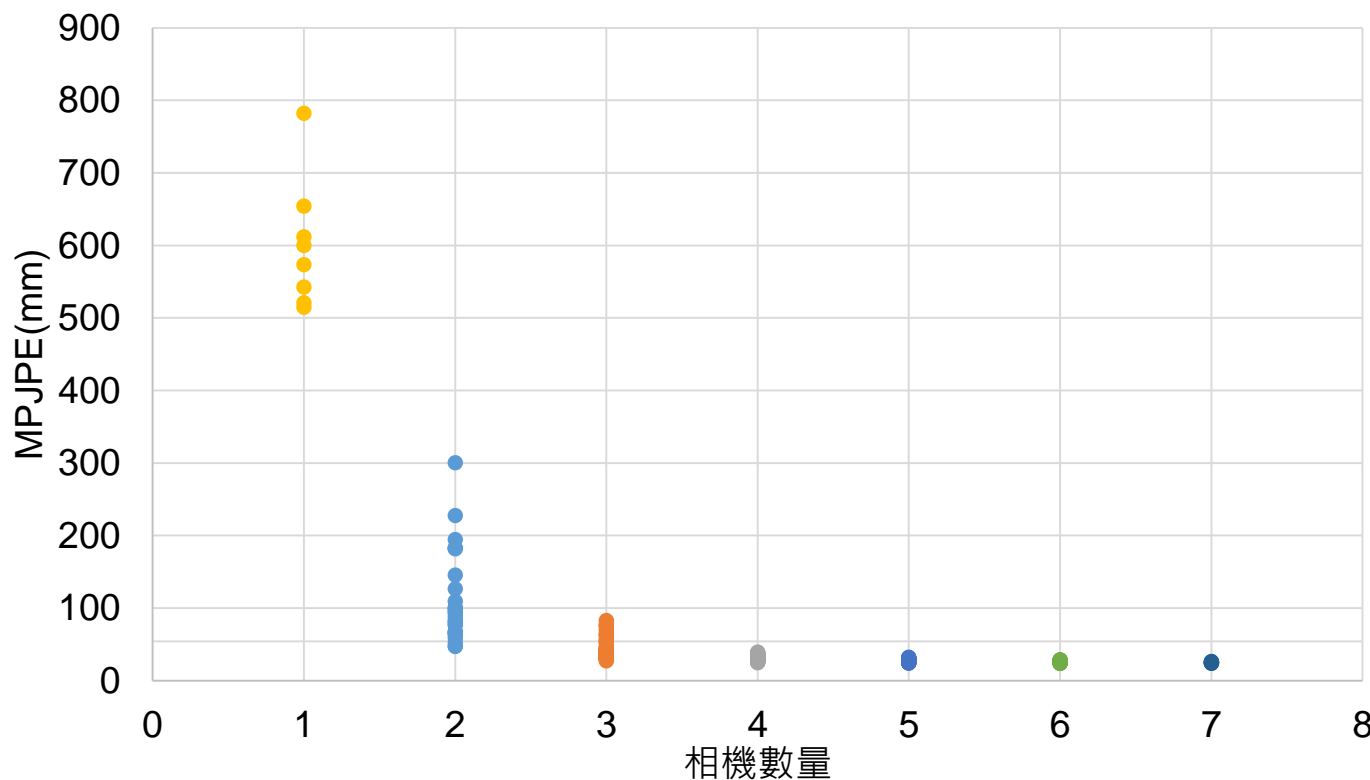
計算平均、
標準差

[6] Trumble, M., Gilbert, A., Malleson, C., Hilton, A., & Collomosse, J. P. (2017, September). Total capture: 3D human pose estimation fusing video and inertial sensors. In *BMVC* (Vol. 2, No. 5, pp. 1-13).

探討減少相機使用數量

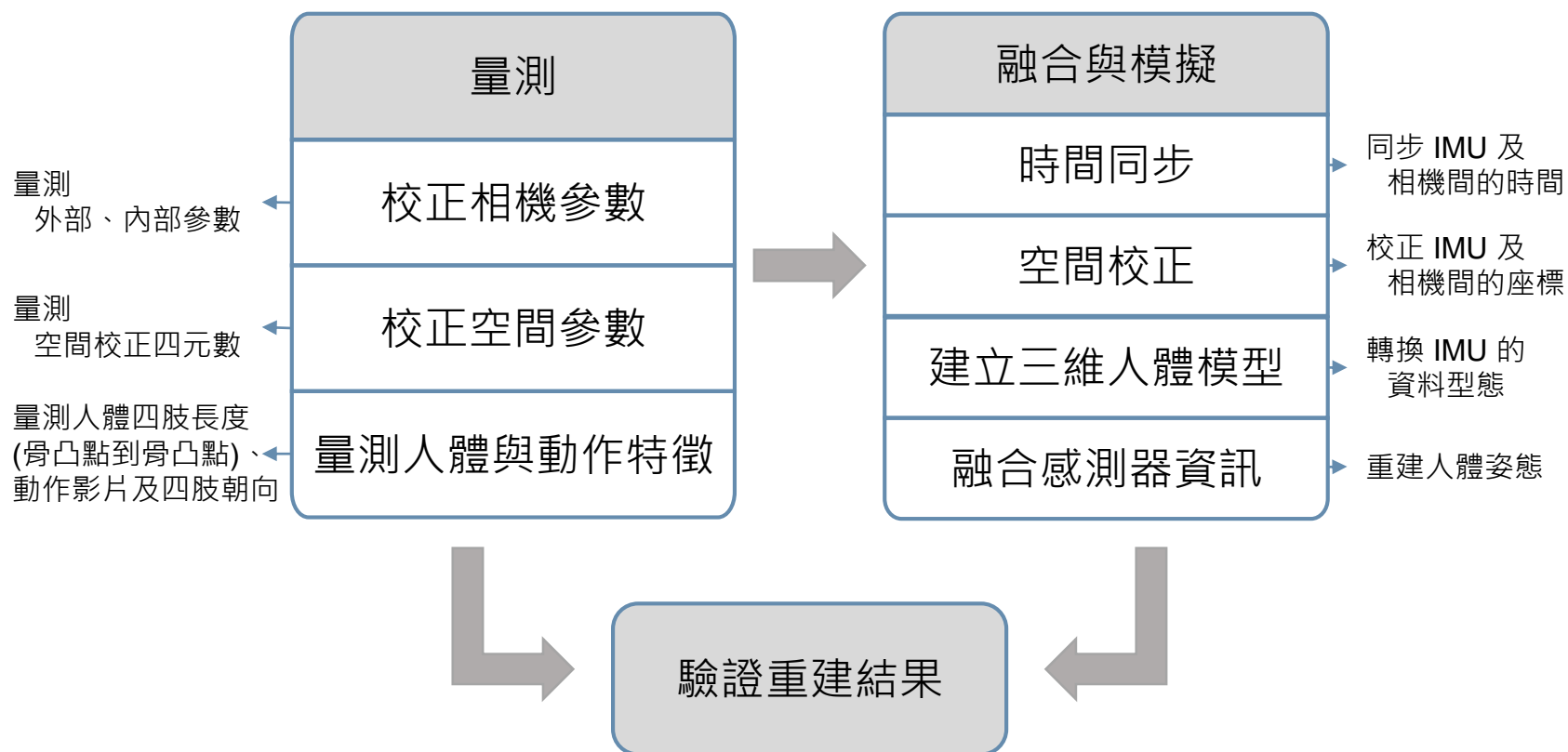
對於人體姿態估計精準度的影響 – 結果

	一台相機	二台相機	三台相機	四台相機	五台相機	六台相機	七台相機
平均 (mm)	599.89	109.87	42.90	31.01	27.66	26.00	24.93
標準差 (mm)	87.60	61.79	13.84	3.69	2.00	1.33	0.85



1. 相機數量增加，**MPJPE**明顯下降，穩定度提升
2. 一到四台時，精準度明顯提升，四台之後，精準度漸趨平緩
3. 為增加實驗架設方便性，減少相機數量是可考慮的選擇
4. 僅用兩台相機的情境中，面向受試者的相機1、8表現最佳

感測器融合方法及流程



實驗場地及實驗設備

相機

鏡頭規格：iPhone 15 Pro (24mm, f/1.78)
 iPhone XR (26mm, f/1.8)
 畫面規格：1920 x 1080
 幀率：60 Hz

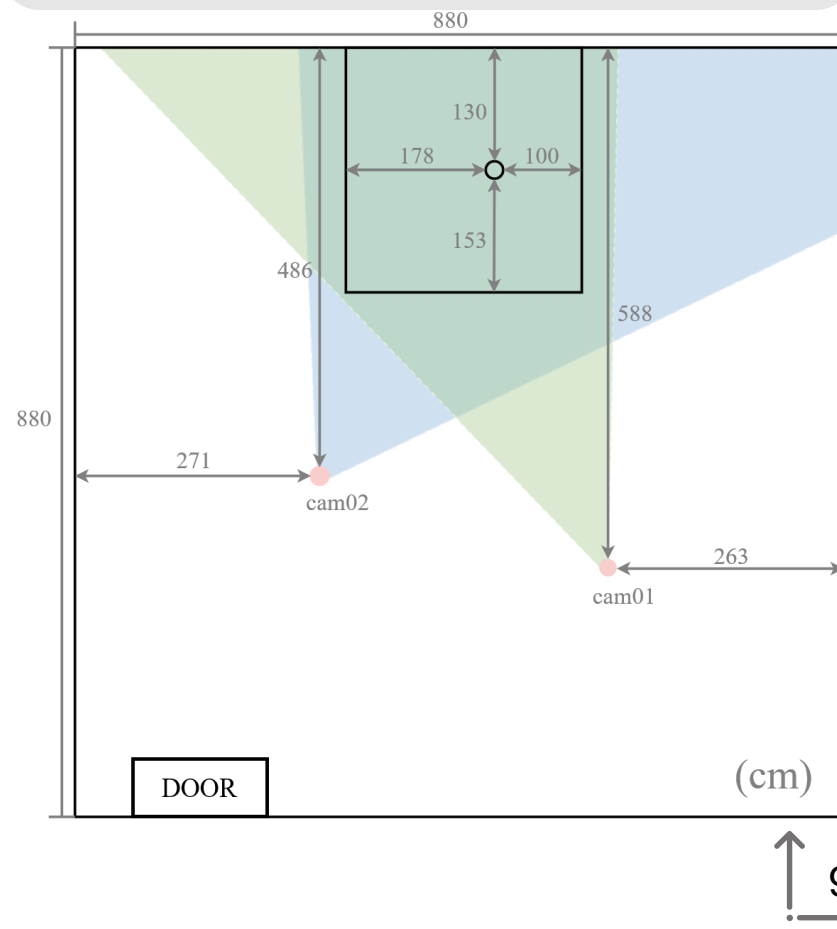
慣性感測器 (Inertial Measurement Units, IMUs)

型號：Awinda
 取樣率：60 Hz
 使用數量：10 + 1 (人體 + 校正板)



場地

整體空間：880 x 880 (cm)
 活動空間：278 x 283 (cm)
 相機位置：圖中紅點處



時間同步方法

預定義事件

T pose → 快速合併手掌 → 維持合掌姿勢

事件特徵

相機端：可在剪輯軟體音軌中找到明顯的拍手聲音，藉此同步相機間的時間

IMU 端：快速合併手掌並維持合掌姿勢會產生明顯的加速度及減速度，藉此找到合掌的時間點

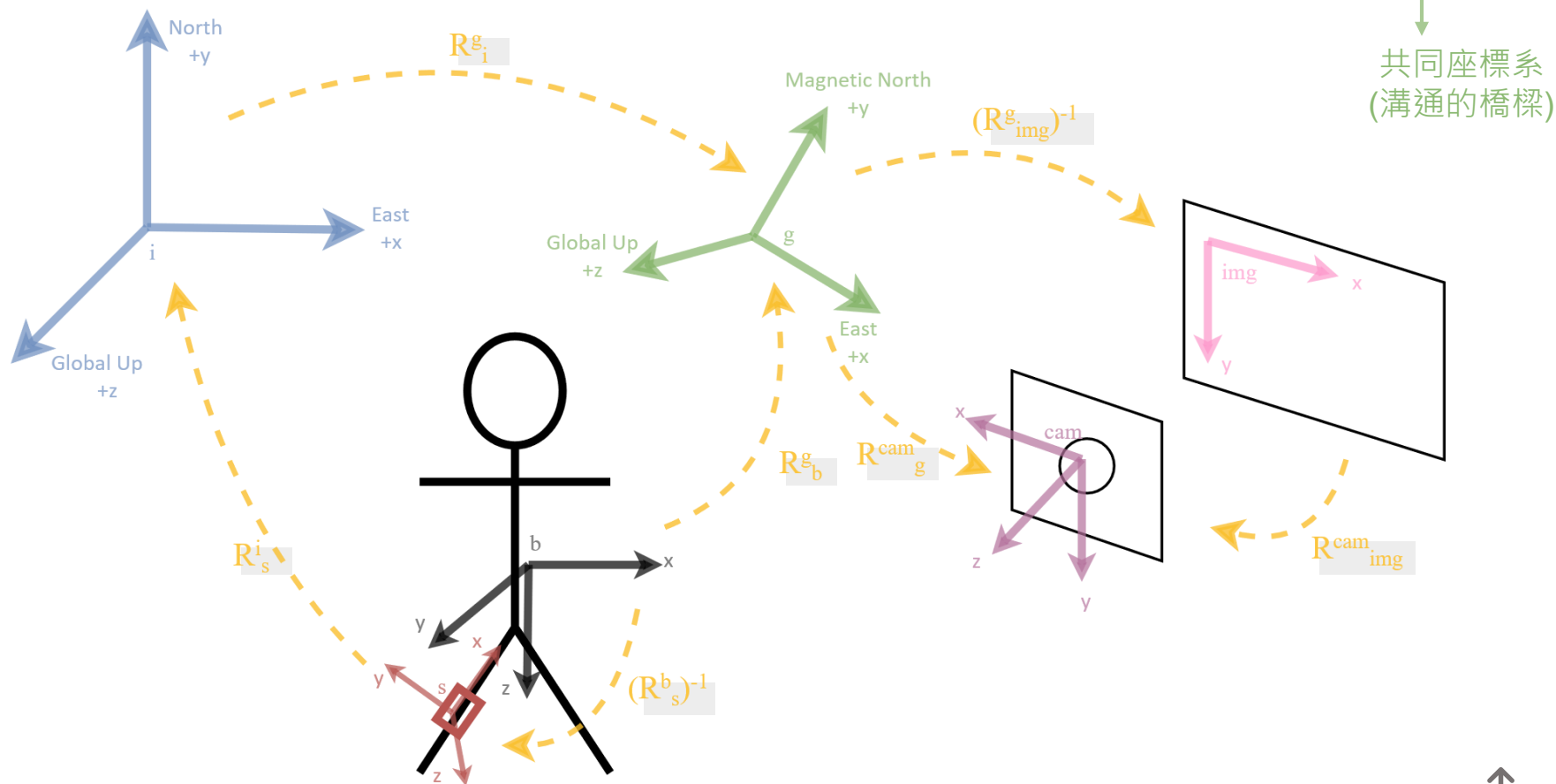
同步方法

影片及 IMU 同步方法：IMU 中明顯的加速度發生於拍手前一幀，因此 IMU 的最大加速度時刻與影片中的拍手前一幀同步



空間校正概念

人體模型座標系 (bone coordinate system, b)	相機座標系 (camera coordinate system, cam)
感測器座標系 (sensor coordinate system, s)	影像座標系 (image coordinate system, img)
IMU local 座標系 (imu local coordinate system, i)	全域座標系 (global coordinate system, g)

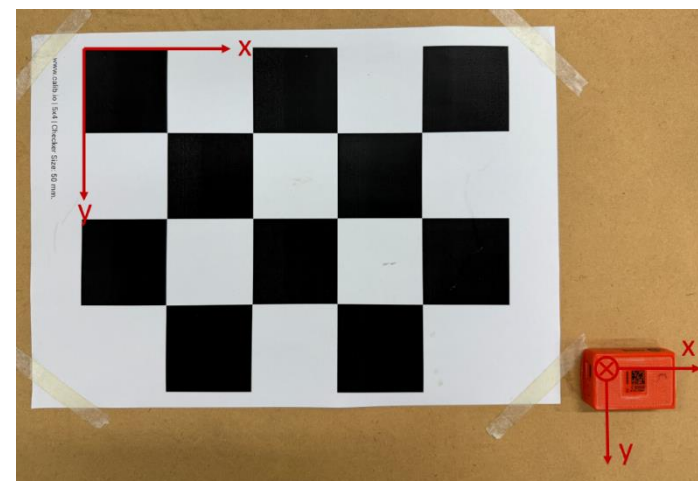
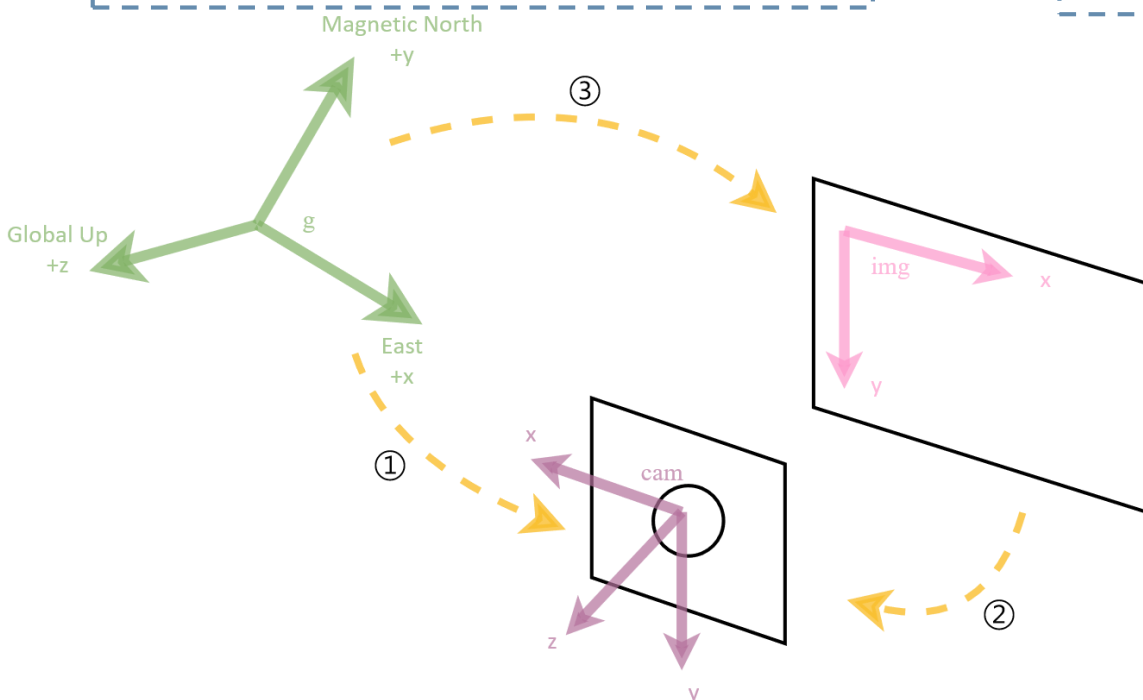


空間校正 – 影像系統

$$\underbrace{R_g^{cam}}_{\textcircled{1}} = \underbrace{R_{img}^{cam}}_{\textcircled{2}} \left(\underbrace{R_{img}^g}_{\textcircled{3}} \right)^{-1}$$

- 棋盤格相機校正方法 [7]
- 使用 Pose2Sim 計算相機參數 [8]
- 外部參數即為 R_{img}^{cam}

- IMU 量得
- 將校正板的座標與 IMU 座標對齊
- $R_s^g = R_{img}^g$

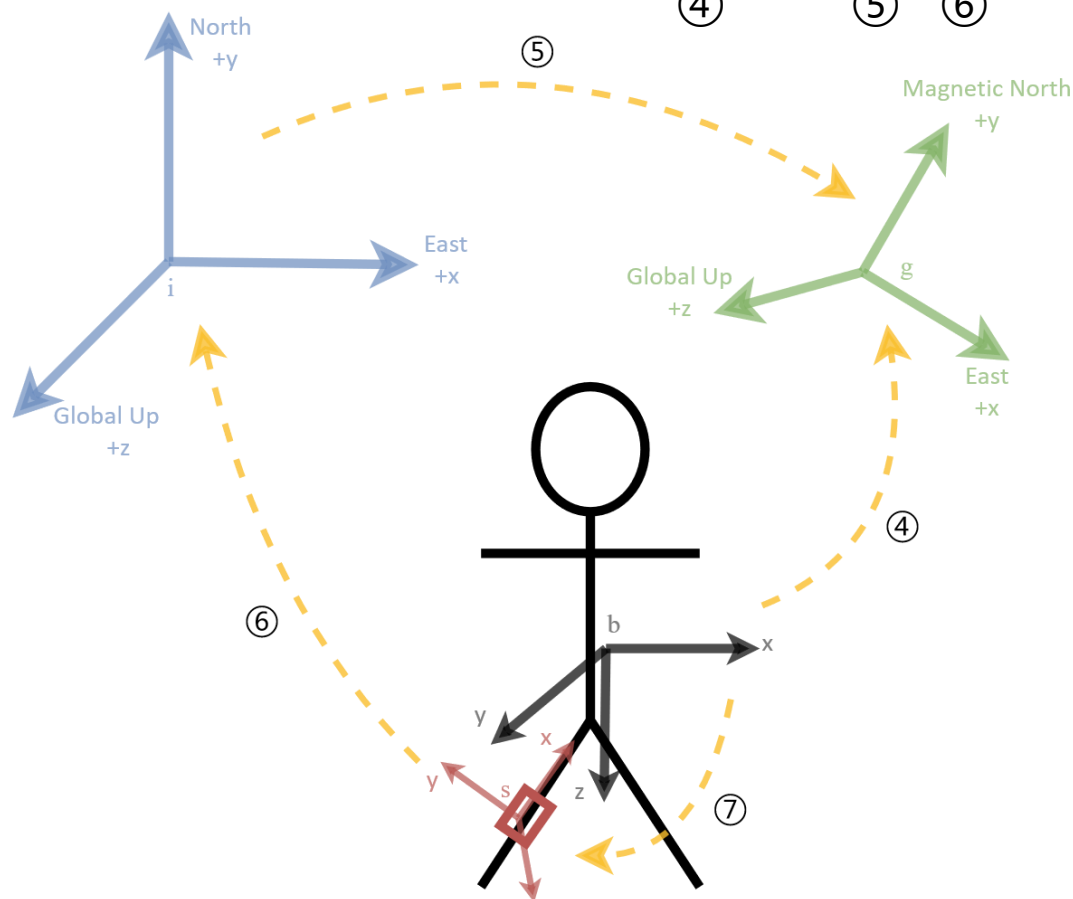


[7] Zhang, Z. (1999, September). Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the seventh IEEE international conference on computer vision* (Vol. 1, pp. 666-673). IEEE.

[8] Pagnon, D., Domalain, M., & Reveret, L. (2022). Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*, 7(77), 4362.

空間校正 – IMU 感測器系統

$$\underbrace{R_b^g}_{(4)} = \underbrace{R_i^g}_{(5)} \underbrace{R_s^i}_{(6)} \underbrace{(R_s^b)^{-1}}_{(7)} [9]$$



⑦

- 於 t_0 假設一個與 IMU local 座標系對齊的人體模型座標系統
- 量得的 $R_s^i = R_s^b$

⑥

- 在量測過程中由 IMU 量得

⑤

- 對 x 軸旋轉 -90°

[9] Malleon, C., Gilbert, A., Trumble, M., Collomosse, J., Hilton, A., & Volino, M. (2017, October). Real-time full-body motion capture from video and imus. In *2017 international conference on 3D vision (3DV)* (pp. 449-457). IEEE.

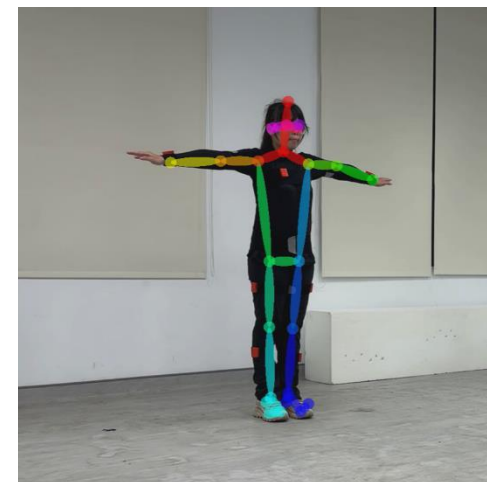
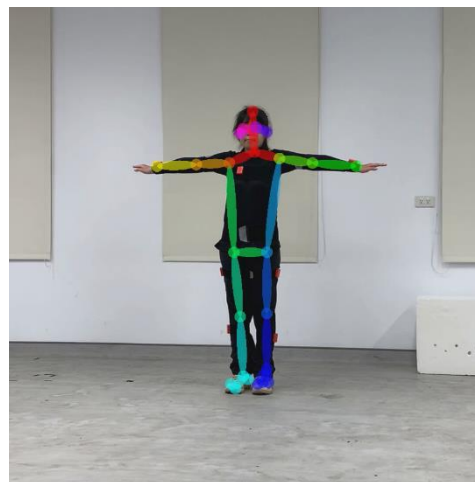
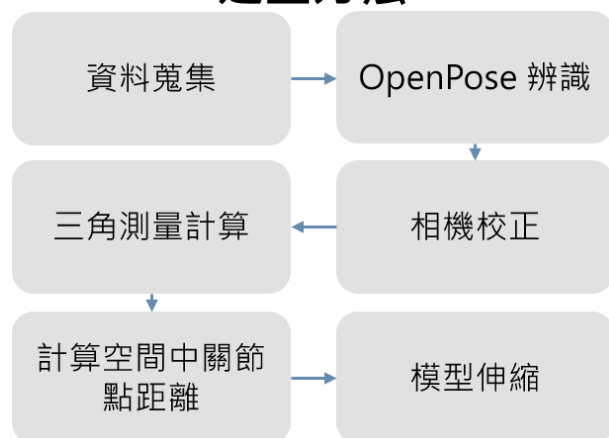
建立三維人體模型

將 IMU 的朝向資訊藉由三維人體模型轉換為以位置表示的向量資訊，
方便後續進行感測器融合

前人使用 Vicon 量得受試者的肢段長度，並對運動學模型進行伸縮

本研究使用影像辨識及三角測量計算得受試者的肢段長度，並對運動學模型進行伸縮

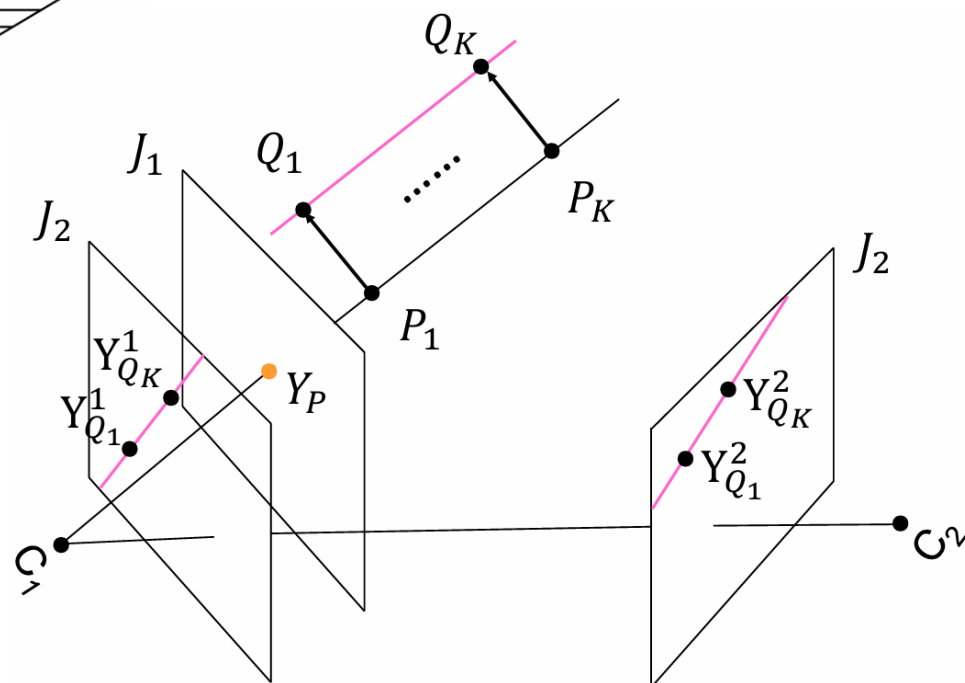
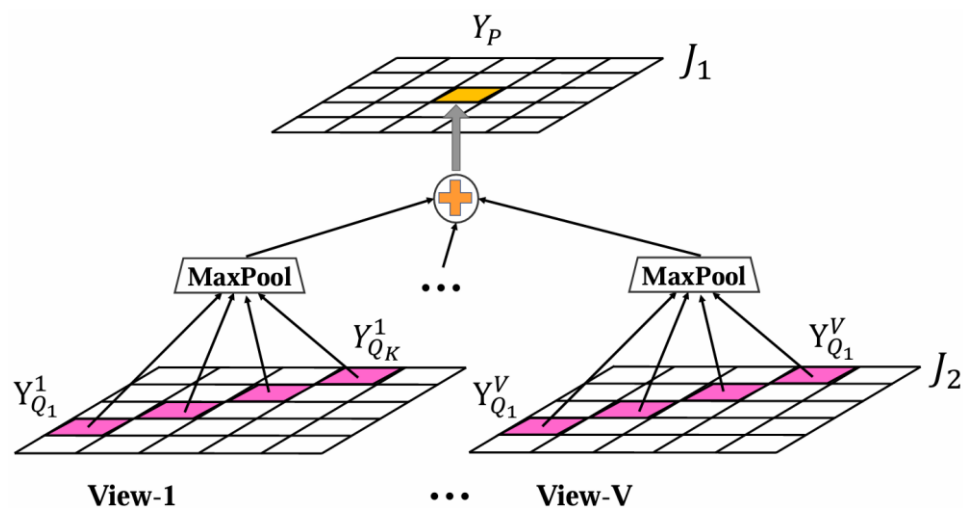
建立方法



驗證

取用 TotalCapture Dataset 中相機 1、8 的影像資料作為輸入，建立三維人體模型，並計算每一肢段長度，再與資料集 Vicon 肢段長度相減取絕對值，平均誤差約為 25.75 (mm)

感測器融合方法



姿態估計結果評估方法

估計結果每 20 幀取樣一次，並對樣本進行以下評估

檢查結果與人體是否相似

人工目視檢查重建結果是否符合人體姿態

計算誤差

$$error = \frac{\sum(\text{估計肢段長度} - \text{實際肢段長度})}{8}$$

決定是否估計成功

若平均差值小於 10 公分，則將該樣本點視為估計成功，同時計入估計成功次數，反之若平均差值大於 10 公分，則將該樣本視為估計失敗，不計入估計成功次數

計算成功率

$$success\ rate = \frac{\text{成功次數}}{\text{樣本數}}$$

動作估計結果

動作相似

誤差範圍內

成功

誤差範圍外

失敗

動作不相似

誤差範圍內

失敗

誤差範圍外

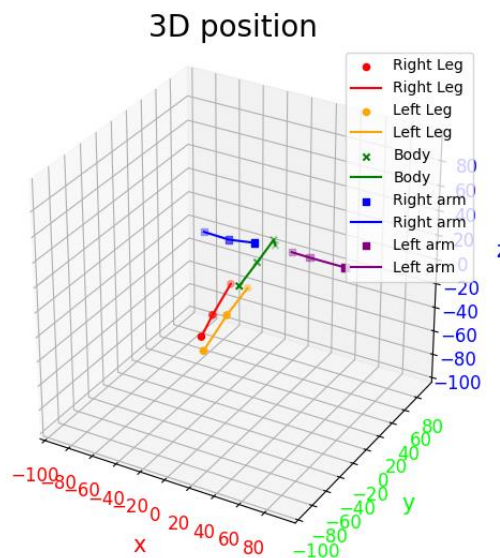
失敗

實驗結果 – T-pose

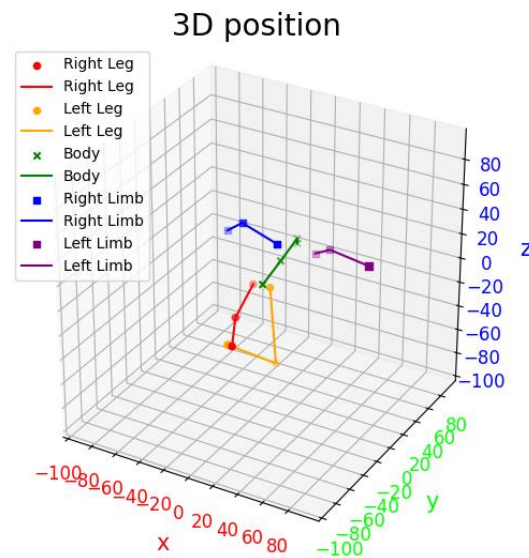
無任何動作變化，且無任何關節點被遮擋



Cam01 真實影像



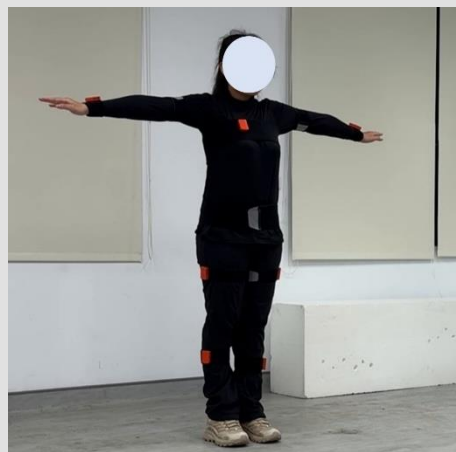
影像辨識融合 IMU 結果



影像辨識結果

小結

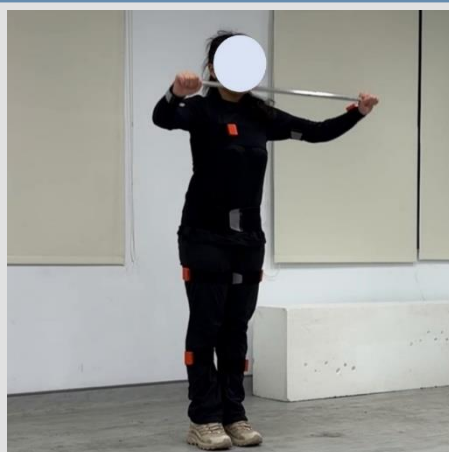
僅使用影像辨識建立出來的結果關節點位置會亂飄，影像辨識融合IMU朝向資訊的人體重建結果較為接近真實人體姿態



Cam02 真實影像

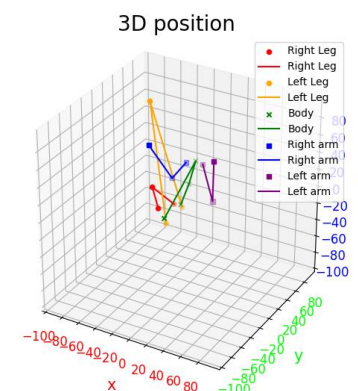
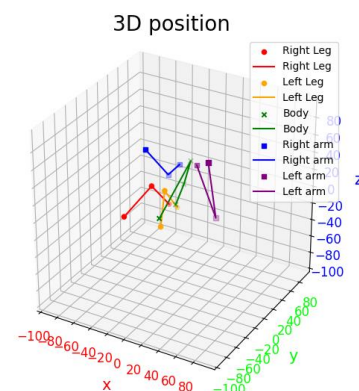
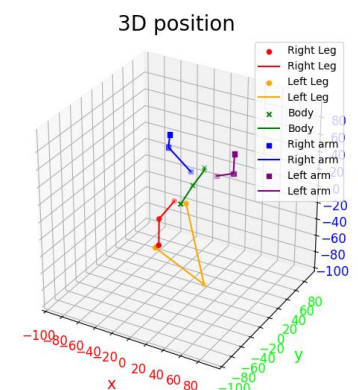
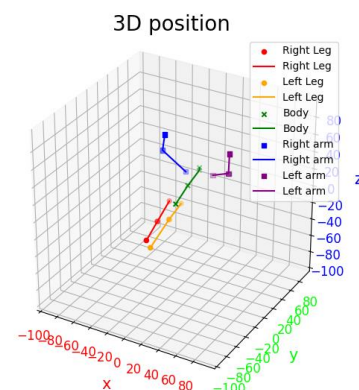
實驗結果 – 蹲站

動作有變化，在蹲下時關節點會受到遮擋



Cam01 真實影像

Cam02 真實影像



影像辨識融合 IMU 結果

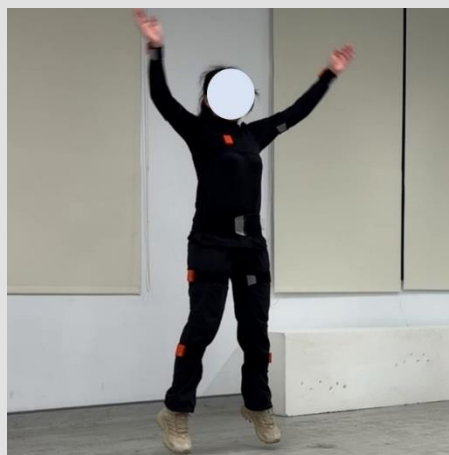
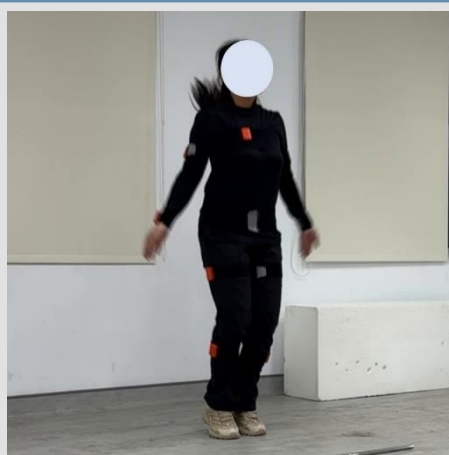
影像辨識結果

小結

動作過程中，關節點受到遮擋，融合 IMU 後仍可以成功重建姿態

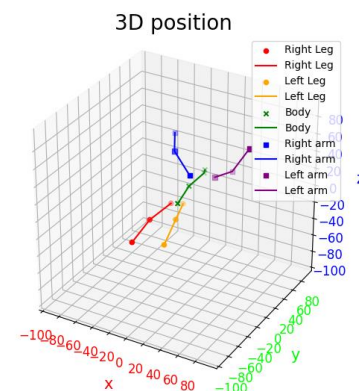
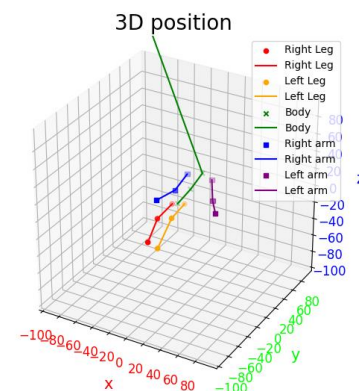
實驗結果 – 開合跳

動作變化快速，但無任何關節點受到遮擋

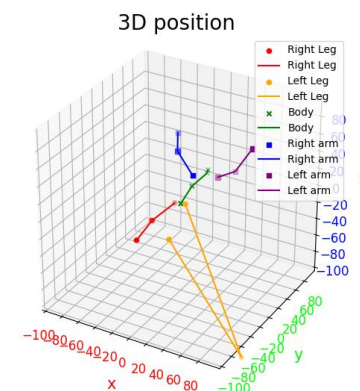
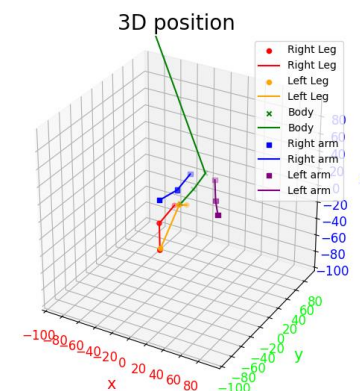


Cam01 真實影像

Cam02 真實影像



影像辨識融合 IMU 結果



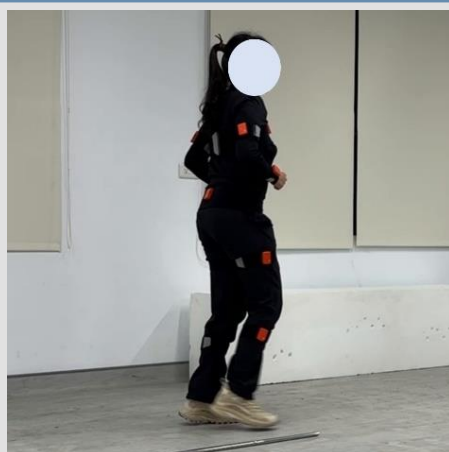
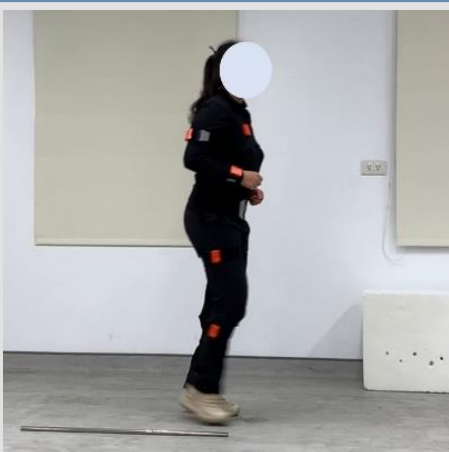
影像辨識結果

小結

變化快速的動作易造成影像動態模糊問題，融合 IMU 後仍可以成功重建姿態

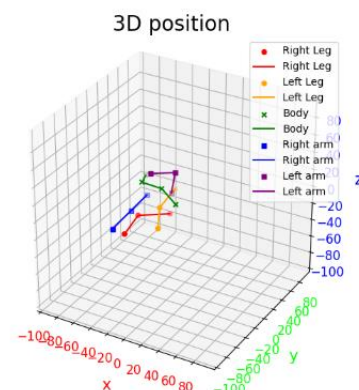
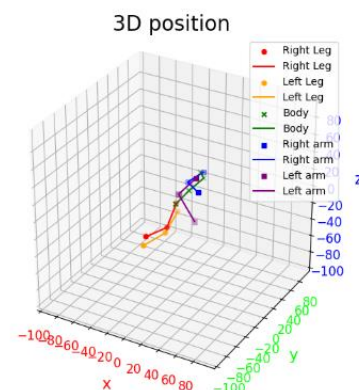
實驗結果 – 折返跑

動作有變化，且多關節點受到遮擋

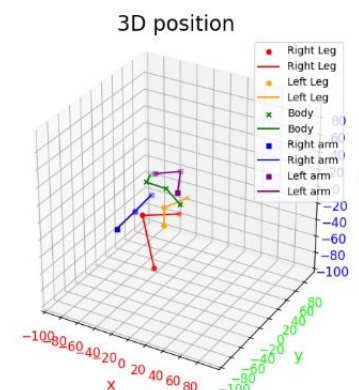
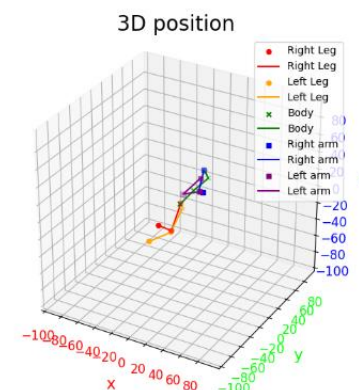


Cam01 真實影像

Cam02 真實影像



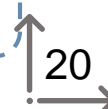
影像辨識融合 IMU 結果



影像辨識結果

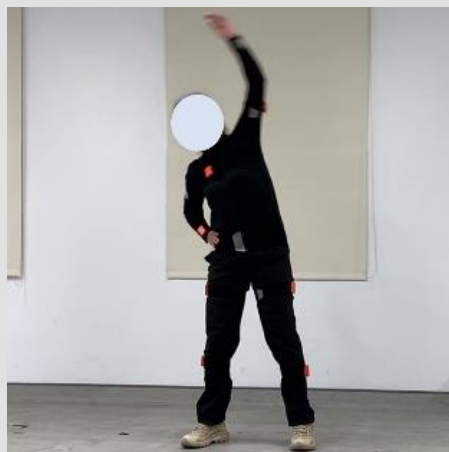
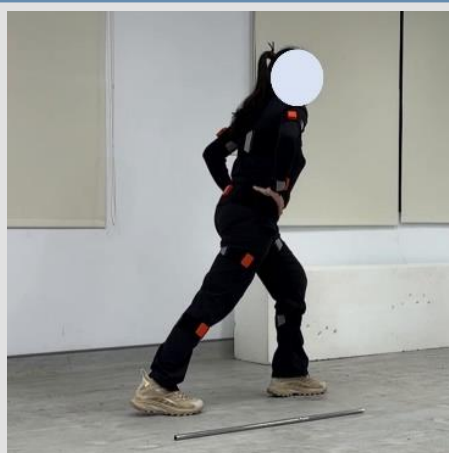
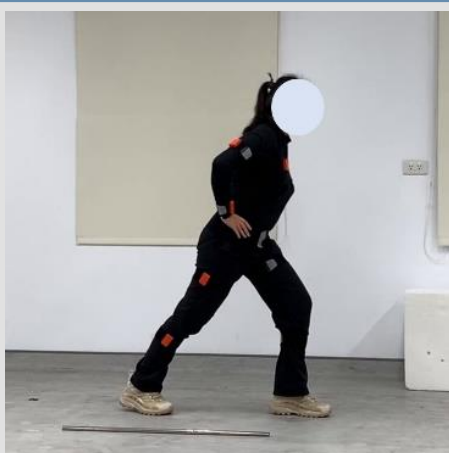
小結

關節點被遮擋時，影像辨識結果容易誤判，融合 IMU 後仍可以成功重建姿態



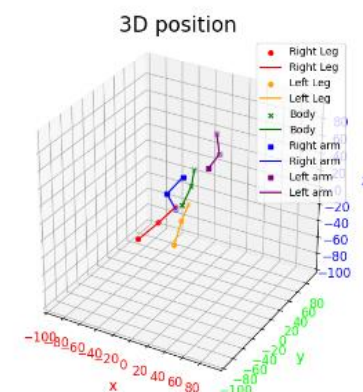
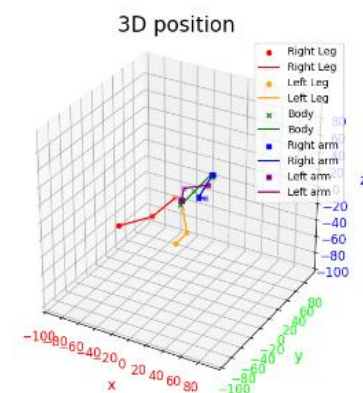
實驗結果 — 熱身運動-1

包含正面、側面及背面的動作，且關節點會受到遮擋

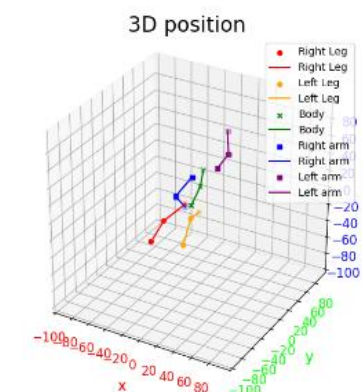
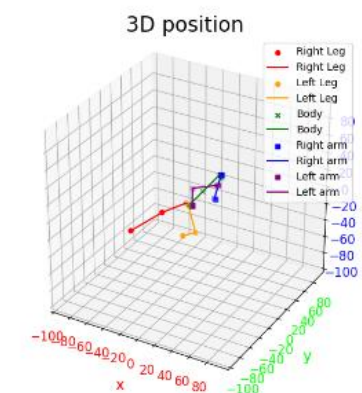


Cam01 真實影像

Cam02 真實影像



影像辨識融合 IMU 結果



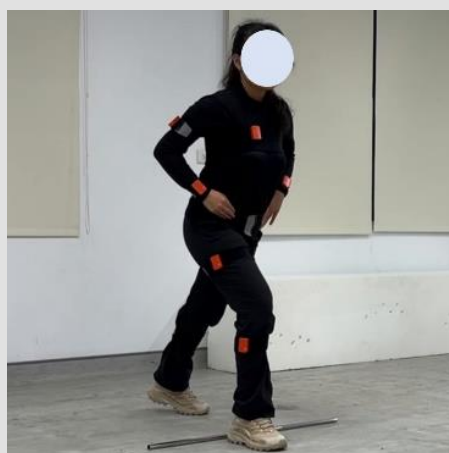
影像辨識結果

小結

在正面及有受到遮擋的側面都可以成功重建姿態

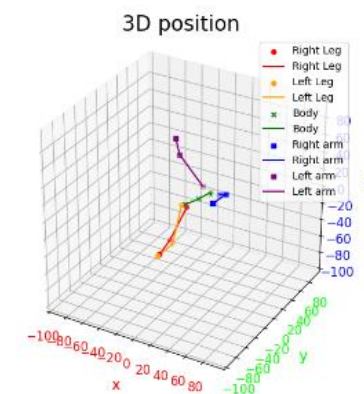
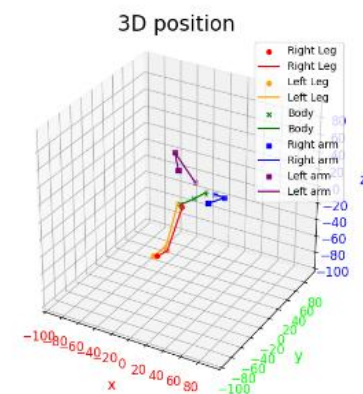
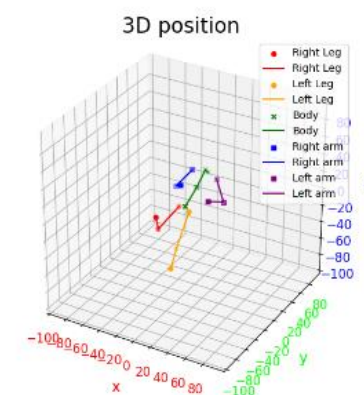
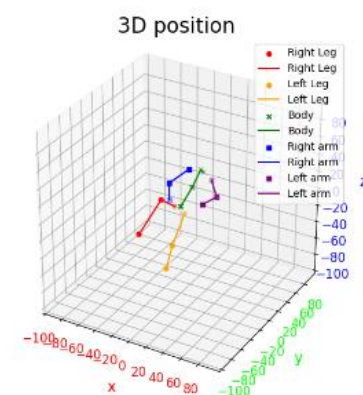
實驗結果 – 熱身運動-2

包含正面、側面及背面的動作，且關節點會受到遮擋



Cam01 真實影像

Cam02 真實影像



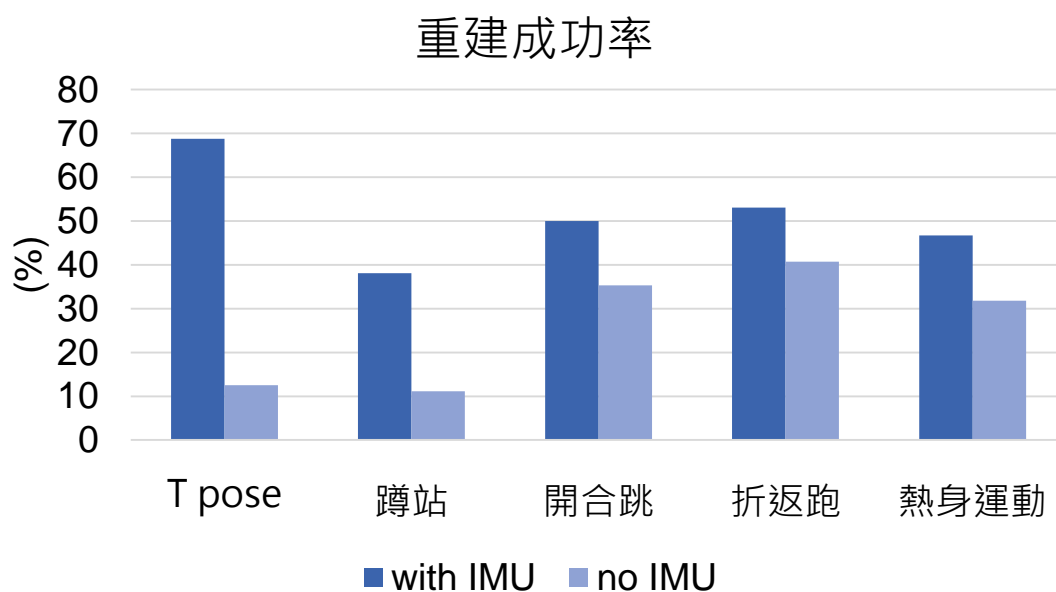
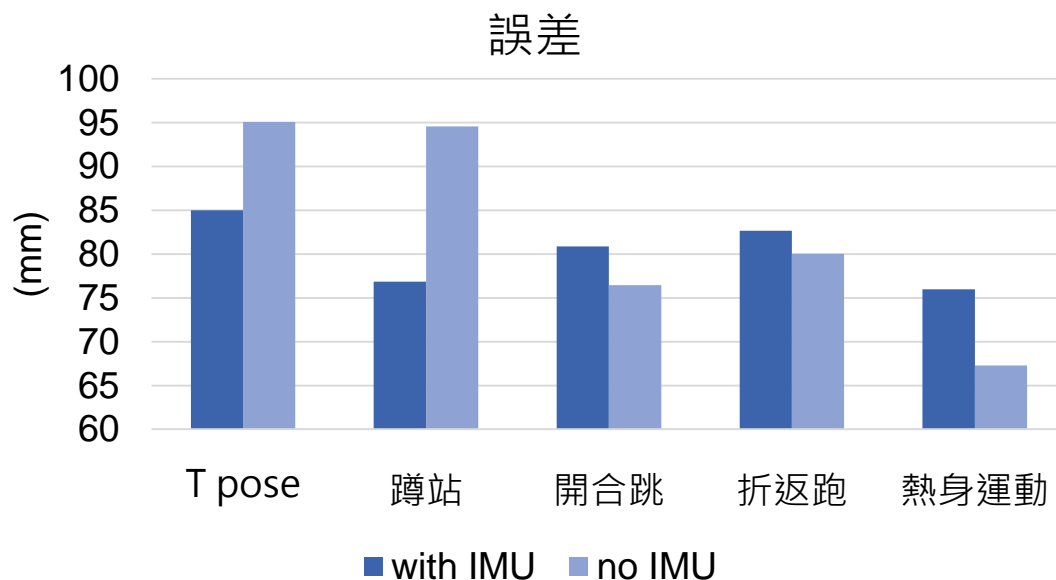
影像辨識融合 IMU 結果

影像辨識結果

小結

相似的姿勢在正面時可以成功重建，但在背面的時候就會失敗，推測可能因為影像辨識的模型沒有訓練背面，因此造成失敗

實驗結果



1. 成功率平均約提升 23.8%
2. 在 T-pose 及蹲站實驗中，誤差改善約 10 (mm)
3. 在開合跳、折返跑、熱身運動實驗中，誤差提升約 5 ~ 10 (mm)，推斷造成此現象的原因可能是這三組動作較複雜，使得辨識結果較差，而IMU朝向資訊雖然可以提供方向資訊，但是在這三組動作中，受試者的動作較快速，且量測時間較長，因此在IMU整合朝向資訊時產生誤差。

結論

- 探討減少相機使用數量對於人體姿態估計精準度的影響
- 應用影像辨識及三角測量計算建立三維人體模型
- 建立用於輸入感測器融合程式的資料前處理流程
- 驗證影像辨識融合IMU資訊可以有效提高人體姿態重建的成功率

未來工作

- 解除對背景、衣著的限制
- 探討相機擺放角度對人體姿態重建的影響
- 量化重建姿態與原始姿態的相似程度

利用IMU改善純影像技術

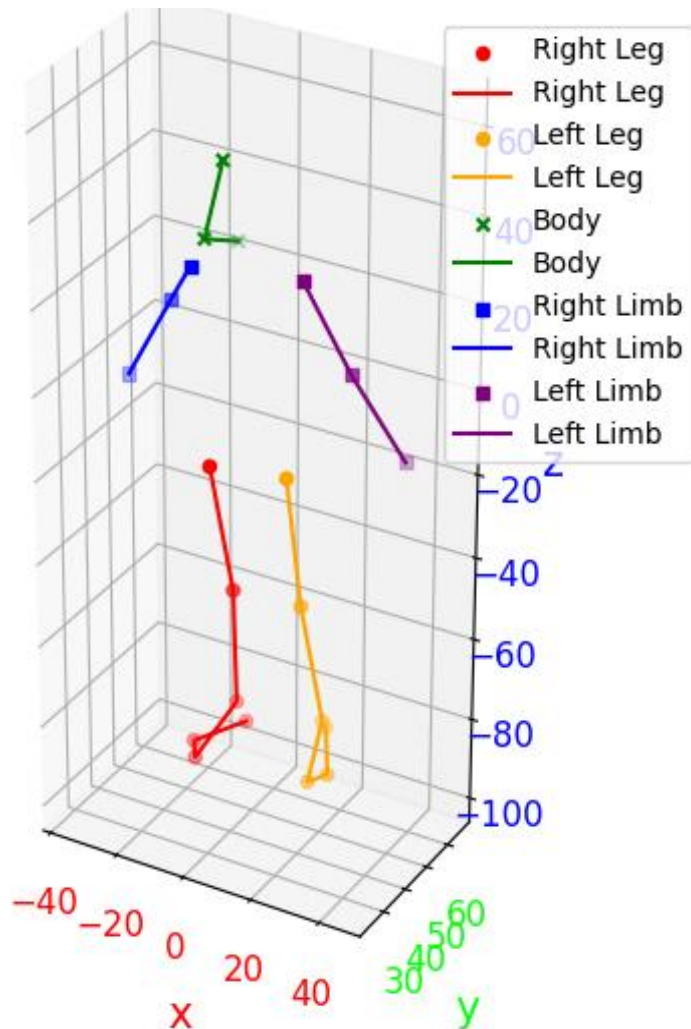
在人體動作重建中的表現

Using IMU to Improve the Performance
of Imaging Technology
in Human Motion Reconstruction

學生：陳珮甄

指導教授：詹魁元 博士

2024/7/30



National Taiwan University
System Optimization Laboratory