

高爾夫球姿態辨識 Golf Posture Recognition

指導老師：韓欽銓 教授

作者姓名：劉治緯、蔡睿宸、劉冠煒、王宏睿、韓德諺

國立聯合大學 資訊工程學系

苗栗市南勢里聯大 2 號

cchan@nuu.edu.tw

摘要

本研究建置一套高爾夫球姿態辨識系統，整合正面一般攝影機、側面一般攝影機、高速攝影機與可穿戴式感測器，將使用者揮桿過程轉換為可視覺化的 3D 骨架、桿頭軌跡、角度變化與文字建議。系統以網頁作為操作介面，後端負責準備姿勢偵測、自動化錄影、感測器接收、OpenPose 與 HybrIK 姿態分析、多來源資料融合及歷史紀錄保存。為降低單一資料來源在遮蔽、模糊或漂移情況下造成的誤差，本研究採用時間戳同步、座標基準統一、資料品質檢查與權重式融合策略。實作結果顯示，本系統可完成從錄影到結果呈現的完整流程，並提供後續姿勢修正與教學輔助的基礎。

關鍵詞：高爾夫球姿態辨識、人體姿態估計、穿戴式感測器、3D 視覺化

Abstract

This project develops a golf posture recognition system that integrates front-view cameras, side-view cameras, a high-speed camera, and wearable sensors. The system converts golf swing data into 3D skeletons, clubhead trajectories, angle charts, and textual feedback. The frontend provides recording, progress tracking, visualization, and history review, while the backend performs address-pose detection, automatic recording, pose estimation, sensor processing, data fusion, and result management. To reduce errors caused by occlusion, motion blur, sensor packet loss, and coordinate inconsistency, the proposed workflow uses timestamp alignment, unified coordinates,

data quality checks, and weighted fusion. The prototype demonstrates a complete pipeline from data acquisition to visualized analysis and can serve as a basis for golf training assistance and future motion analysis applications.

Keywords: Golf posture recognition, human pose estimation, wearable sensor, 3D visualization

一、前言

高爾夫揮桿是一項高速且連續的全身協調動作，身體穩定度、髖部旋轉、肩膀帶動、手腕控制與球桿釋放皆會影響擊球方向與穩定性。傳統錄影回放或教練觀察雖然直覺，但容易受到拍攝角度與觀察經驗限制，專業動作捕捉設備則有成本高、環境受限等問題。因此本專題建置一套低成本的高爾夫球姿態辨識系統，整合正面攝影機、側面攝影機、高速攝影機與可穿戴式感測器，透過網頁介面進行操作，並由後端完成資料擷取、姿態分析、資料融合、結果視覺化與歷史紀錄保存。使用者完成揮桿後，系統可產生 3D 骨架、桿頭軌跡、角度圖表與 AI 姿勢建議，作為後續姿勢修正與訓練輔助參考。

二、系統架構與相關技術

本系統以多來源資料整合為核心，整體流程可分為資料擷取、姿態分析、資料融合、結果輸出與歷史保存等階段。系統首先透過正面攝影機、側面攝影機、高速攝影機與可穿戴式感測器取得揮桿資料，再由後端進行人體姿態估計、感測器資料處理、時間同步與座標統一。完成多來源資料融合後，系統會產生 3D 骨架、桿頭

軌跡、角度圖表與 AI 姿勢建議，並將分析結果提供前端呈現，同時保存紀錄以供後續查詢。



圖 1 系統整體流程

2.1 資料來源分工

資料來源分工可整理為四類：正面攝影機負責左右平衡、肩膀高度與手臂位置；側面攝影機補充前後深度、身體前傾與重心移動；高速攝影機補強下桿、擊球瞬間與手腕快速變化；感測器提供旋轉、角速度、加速度。四者各有優勢與限制，因此系統必須依照不同動作階段調整資料可信度。

2.2 OpenPose

OpenPose 為一種基於深度學習之 2D 人體姿態估測模型，主要用於從影像中擷取人體關鍵點位置。該模型透過卷積神經網路（CNN）進行特徵萃取，並輸出人

體各關節之二維座標資訊。其核心技術包含關節熱度圖與部位關聯場，可有效建立關節間之連接關係，進而組成完整人體骨架[2]。

2.3 VideoPose3D

VideoPose3D 為一種基於深度學習之 3D 人體姿態估測模型，主要用於將二維人體關鍵點序列提升至三維空間座標表示。該模型以連續影格之 2D 關鍵點作為輸入，並透過時間卷積神經網路(TCN)學習人體動作之时序特徵，以推估對應之三維人體骨架結構[3]。

2.4 HybrIK

HybrIK 為一種基於深度學習之三維人體姿態與人體形狀估測方法。其核心概念是結合神經網路與逆向運動學，先估計人體三維關節位置，再將位置轉換為各身體部分的相對旋轉參數，並套用於 SMPL 人體模型，以產生較符合人體骨架結構與關節連動關係的三維姿態結果。相較於單純輸出 3D keypoints 的方法，HybrIK 可透過人體模型限制降低骨架比例不合理或關節旋轉不連續的情況，因此適合用於人體三維骨架重建與姿態分析。[4][5]

2.5 前後端整合與資料儲存技術

本系統前端以 Next.js 與 React 建置，負責錄影、分析結果、歷史紀錄與檢查頁面之呈現，並將攝影機畫面、流程狀態、3D 骨架與角度圖表等功能模組化。後端以 FastAPI 負責流程控制，將自動化錄影、感測器接收、姿態分析、資料融合與結果查詢封裝為 API，供前端呼叫。由於影像分析與 3D 姿態重建耗時較長，系統透過狀態紀錄與輪詢機制回傳進度。資料儲存方面，系統使用 Supabase 保存分析紀錄與檔案路徑，影片、3D 骨架 JSON、感測器資料與融合結果則存放於 Supabase Storage，以支援歷史紀錄查詢與結果重新載入。

2.6 雙向長短期記憶網路 BiLSTM

雙向長短期記憶網路（Bidirectional Long Short-Term Memory, BiLSTM）是一種用於處理序列資料的深度學習模型，其

基礎來自長短期記憶網路。Graves 與 Schmidhuber 於 2005 年將雙向 LSTM 應用於逐幀序列分類任務，並指出雙向架構能同時利用序列前後文資訊，以提升序列分類效果[6]。

傳統循環神經網路（Recurrent Neural Network, RNN）雖然能處理時間序列資料，但在序列較長時，容易因反向傳播過程中的梯度消失或梯度爆炸問題，導致模型難以保留長距離的時序資訊。LSTM 透過細胞狀態（Cell State）與閘門機制改善此問題，其中遺忘閘負責決定保留或捨棄舊資訊，輸入閘負責寫入新的重要資訊，輸出閘則決定當前時間點要輸出的隱藏狀態。透過這些機制，LSTM 能較有效地記憶長時間序列中的關鍵特徵。

BiLSTM 則同時使用正向 LSTM 與反向 LSTM，分別從序列的前後兩個方向擷取特徵，並將兩個方向的隱藏狀態進行串接，因此模型採用 BiLSTM 進行逐幀動作階段分類，藉由其在每一個時間點都能同時參考過去與未來的上下文資訊的能力，以提升模型對整體時序脈絡的理解與階段分割穩定性。

三、研究方法與系統設計

3.1 自動化模組流程

系統以狀態機控制錄影與分析流程。使用者進入錄影頁面後，前端顯示正面、側面與高速攝影機畫面，並確認感測器連線狀態。按下開始後，系統依序執行準備姿勢偵測、校正、同步錄影、資料保存與姿態分析。準備姿勢偵測用於確認使用者完整入鏡、雙手位置與身體姿態穩定；校正階段則建立影像骨架、感測器資料與球桿初始位置的共同基準。錄影期間，系統持續監測手腕、身體旋轉與桿頭動作，並於判斷揮桿結束後停止錄影，接續進入資料融合與結果分析流程。語音提示則用於提醒使用者維持姿勢、開始揮桿與錄影完成，降低單人操作時查看螢幕的需求。

3.2 一般攝影機架設與校正

本研究採用正面與側面雙攝影機同步擷取揮桿影像，以取得不同視角之人體姿

態資訊。為降低鏡頭畸變影響，先以棋盤格標定板進行內參校正，修正影像畸變。接著再以棋盤格作為參考基準，調整兩台攝影機之相對位置與拍攝方向，使兩視角夾角接近 90 度，以降低遮擋與視角誤差對姿態估測結果的影響。

3.3 感測器資料擷取與 T-pose 校正

本系統使用可穿戴式感測器輔助影像分析，感測器可配置於手腕、胸部、骨盆、腿部、頭部與球桿桿頭，用於擷取旋轉、角速度與加速度資料，以補充影像在遮蔽、模糊或高速揮桿時較難判斷的動作資訊[8]。正式錄影前，系統透過 T-pose 取得各 tracker 的初始四元數，作為本次穿戴狀態的基準姿態。後續感測器資料會轉換為相對於此基準的旋轉變化，並與影像骨架資料進行同步與融合。

3.4 時間同步與座標統一

由於一般攝影機、高速攝影機與感測器的取樣頻率不同，若直接以影格編號對齊，容易造成不同時間點的動作被誤判為同一時刻。因此，本系統以時間戳記為基準進行資料對齊，並將各來源資料轉換至一致的骨架座標系[9]。座標統一方面，系統以第一幀可靠骨盆位置作為原點，X 軸表示左右方向，Y 軸表示高度方向，Z 軸表示前後深度方向，以保留身體重心與桿頭軌跡的連續變化。

3.5 多來源資料融合

本系統不採用單純平均，而是依資料品質、關節部位與揮桿階段進行加權融合。正面攝影機適合觀察左右平衡，側面攝影機補足前後深度，高速攝影機捕捉快速手部變化，感測器則補充遮蔽或模糊時的旋轉趨勢[9]。

$$P_{\text{fused}}(j,t) = \Sigma(w_i(j,t) \times P_i(j,t)) / \Sigma w_i(j,t)$$

其中， $P_i(j,t)$ 表示第 i 個資料來源在時間 t 對關節 j 的位置估計， $w_i(j,t)$ 為其權重。權重會依 OpenPose 關鍵點信心度、資料缺失、異常跳動、動作階段、感測器動作強度與關節部位調整；例如手腕與手肘在高速揮桿階段提高高速攝影機與手部感測

器權重，桿頭則結合手腕位置、球桿長度與桿頭感測器姿態共同推估。

3.6 資料集建置

本研究以 GolfDB 作為高爾夫揮桿資料來源。GolfDB 包含 1,400 部高品質揮桿影片，並標註 Address、Toe-up、Mid-backswing、Top、Mid-downswing、Impact、Mid-follow-through 與 Finish 等八個關鍵事件，可作為揮桿階段分析與模型訓練依據 [7]。由於 GolfDB 僅提供單視角 RGB 影片，本研究先利用 OpenPose 擷取 2D 骨架，經信心度篩選、缺失補值與平滑處理後，再輸入 VideoPose3D 重建 3D 骨架，並透過後處理提升骨架穩定性與合理性[2][3]。最後透過視覺化檢查與品質篩選移除重建不佳樣本，保留約 800 餘部資料，並進行速度調整、三維視角轉換與冗長區段裁切等操作，建立後續模型訓練與分析所需之三維骨架資料集。

3.7 骨架辨識模型

系統會以融合後的 3D 骨架，計算每一幀的速度與加速度，模型會將 3D 座標、速度與加速度合併為每一幀的輸入特徵。每一幀輸入 17 個關節點各包含 3 維座標，合併位置、速度與加速度，因此單一特徵維度為 $17 \times 3 \times 3 = 153$ 。

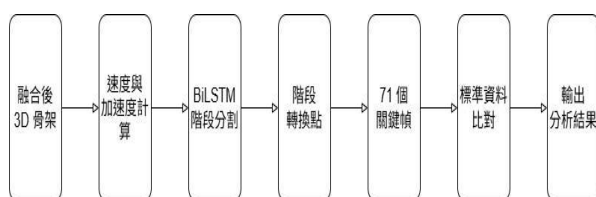


圖 2 動作姿態分析整體流程圖

如:圖 2 顯示，階段分割模型採用雙向長短期記憶網路（Bidirectional Long Short-Term Memory, BiLSTM）作為動作階段分割模型，輸出為該幀對應的動作階段類別，藉由分析後類別對應的分佈，將動作進行 71 關鍵幀的重取樣，以更好的與標準進行比較，為了避免用戶肢體長度所造成的差異影響判斷，系統以關節角度與旋轉特徵作為分析，例如左右膝、左右腕、左右手肘、左右肩、脊椎彎曲、肩胸角度、骨盆角度、胸口旋轉與腕部旋轉等。

如:圖 3 系統會將使用者資料與標準資料進行比對，若數值高於或低於標準範圍，系統會記錄偏差方向與偏差量，並將結果回傳給後續 AI 分析模型。

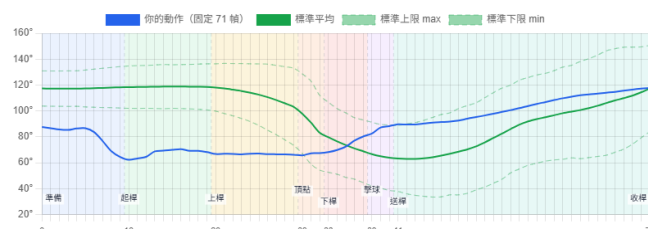


圖 3 用戶姿態比較圖

3.8 模型評估

受限於高爾夫揮桿資料取得不易，資料集規模相對有限，若直接以原始資料訓練模型，可能導致模型對不同揮桿速度、角度變化與少數階段的泛化能力不足。為了提升動作階段分割模型的辨識效果，藉由比較不同資料增強方式與類別權重設定對模型表現的影響。資料版本包含原始資料、時間拉伸角度不變，以及時間拉伸角度變化三種；每一種資料版本皆分別測試有無類別權重，以觀察資料增強與類別不平衡處理對模型辨識能力的影響。

資料版本	權重使用	Accuracy	Macro-F1	Impact F1-score
原始資料	否	88.28%	77.08%	71.71%
原始資料	是	89.03%	87.47%	77.10%
時間拉伸	否	91.64%	90.59%	83.06%
時間拉伸	是	90.86%	90.93%	82.68%
時間拉伸角度變化	否	92.53%	92.52%	86.51%
時間拉伸角度變化	是	91.91%	91.91%	85.59%

圖 4 模型版本比較圖

如:圖 4 中顯示，時間拉伸與角度變化皆能使模型學習不同的揮桿速度與姿態，進而大幅提升泛化能力與整體表現。以無權重版本為例，原始資料的 Accuracy、Macro-F1 與 Impact F1-score 分別低於時間拉伸後；進一步加入角度變化後更達到 92.53%、92.52% 與 86.51%。然而，當資

料經過增強後，類別權重所帶來的效益便大幅減少，在「時間拉伸+角度變化」版本中，加入權重反使 Macro-F1 由 92.52% 略降至 91.91%。

綜合比較後，考量整體階段切分穩定性、Macro-F1 表現，以及 Impact 關鍵擊球階段的辨識效果，本研究最終選擇「時間拉伸角度變化+無權重」作為系統實際使用的動作階段分割模型。

3.9 AI 姿勢建議生成

為了讓使用者不需直接解讀大量數值資料，本系統會在姿態分析完成後產生文字建議。系統先將揮桿資料依動作階段整理，並針對各關節、部位，比對是否位於標準範圍內。AI 建議生成時會優先參考連續異常、最大偏差與資料品質，避免只依單一幀或短暫跳動來判斷，使文字回饋更適合作為姿勢修正與訓練輔助參考。

四、系統實作與成果展示

4.1 實驗環境與穿戴配置

系統測試前會先確認使用者完整入鏡，並依照分析需求配置感測器。感測器可裝於左手、右手、胸部、骨盆、腿部、頭部與球桿桿頭，用於輔助判斷身體旋轉、手部動作與桿頭運動狀態。感測器資料並不直接取代影像定位，而是作為姿態趨勢、動作強度與球桿姿態的補充依據[8]。



圖 7 正面拍攝與感測器穿戴示意

4.2 錄影與分析功能

錄影頁面整合正面、側面與高速攝影機畫面，並顯示系統流程狀態與感測器連線資訊。使用者可在錄影前確認身體是否完整入鏡、拍攝角度是否正確。按下開始後，系統會依序執行準備姿勢偵測、校正、開始與結束錄影，前端則即時更新目前進度。

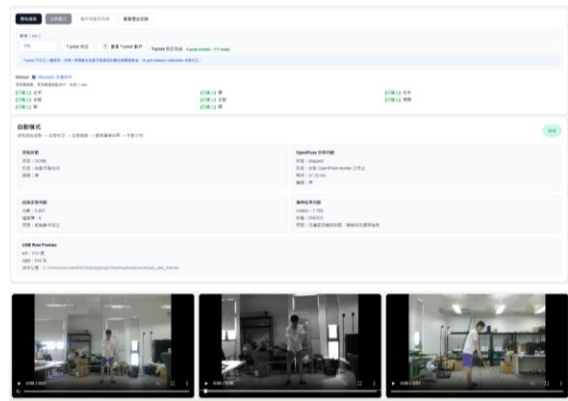


圖 8 錄影頁面與多攝影機畫面

使用者點擊分析鍵之後分析結果頁面會呈現標註影片、角度圖表與 AI 姿勢建議。使用者可透過影片觀察揮桿姿態，並透過角度曲線與標準範圍比較，判斷各關節在不同階段是否出現偏差。

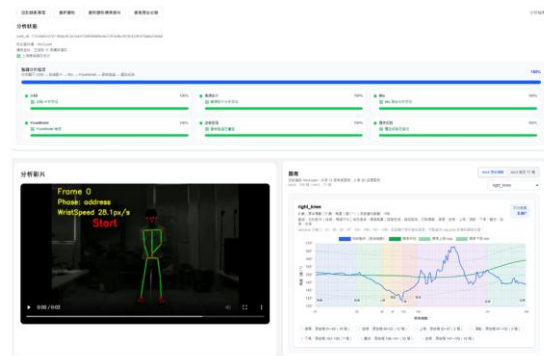


圖 9 分析結果、影片與角度圖表

融合後的 3D 骨架視覺化畫面可呈現人體姿態與桿頭位置，使使用者能從三維空間觀察身體、手部位置與球桿軌跡的變化。

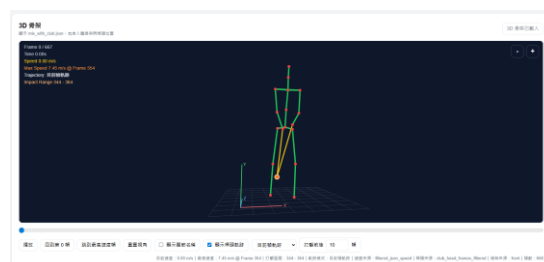


圖 10 融合後 3D 骨架視覺化畫面

4.3 歷史紀錄

歷史紀錄頁面用於保存與查詢過去的揮桿分析結果。系統保存建立時間、工作資料夾、影片路徑、3D 骨架路徑、融合結果與姿勢建議等索引資訊，使用者可依紀錄重新載入過去結果，方便查看不同次揮桿表現。



圖 11 歷史紀錄或結果查詢頁面

五、結論

本專題完成一套以多來源資料為核心的高爾夫球姿態辨識系統，整合一般攝影機、高速攝影機與可穿戴式感測器，並透過前後端分離式架構串接錄影、校正、姿態分析、資料融合、結果視覺化與歷史紀錄保存等流程。系統可將揮桿過程轉換為 3D 骨架、桿頭軌跡、角度變化與文字建議，使使用者能更直觀地觀察動作問題。整體而言，本系統已完成從資料擷取到視覺化分析的基本流程，具備作為高爾夫訓練輔助與動作分析系統之應用基礎。

六、參考文獻

- [1] P. Patel, B. Bhatt, and B. Patel, “Human body posture recognition — A survey,” 2017 International Conference on Innovative Mechanisms for Industry Applications, 2017.
- [2] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE TPAMI, 2021.
- [3] D. Pavlo, C. Feichtenhofer, D. Grangier, and M. Auli, “3D human pose estimation in video with temporal convolutions and semi-supervised training,” CVPR, 2019.
- [4] J. Li, C. Xu, Z. Chen, S. Bian, L. Yang, and C. Lu, “HybriK: A Hybrid Analytical-Neural Inverse Kinematics Solution for 3D Human Pose and Shape Estimation,” CVPR, 2021.
- [5] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: A Skinned Multi-Person Linear Model,” ACM Transactions on Graphics, 2015.
- [6] A. Graves and J. Schmidhuber, “Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures,” Neural Networks, vol. 18, no. 5–6, pp. 602–610, 2005.
- [7] W. McNally, K. Vats, T. Pinto, C. Dulhanty, J. McPhee, and A. Wong, “GolfDB: A Video Database for Golf Swing Sequencing,” CVPR Workshops, 2019.
- [8] K. Shirota, K. Watanabe, and Y. Kurihara, “Measurement and analysis of golf swing using 3-D acceleration and gyro sensor,” SICE Annual Conference, 2012.
- [9] W. Fang, L. Zheng, and H. Deng, “A motion tracking method by combining the IMU and camera in mobile devices,” International Conference on Sensing Technology, 2016.
- [10] R. Dong and S. Ikuno, “Biomechanical analysis of golf swing motion using Hilbert-Huang Transform,” Sensors, 2023.
- [11] OpenCV Documentation, “Camera Calibration and 3D Reconstruct