

國立清華大學 電機工程學系

實作專題研究成果摘要

Cocktail Pose Recommendation and Evaluation System

Based on Camera and Machine Learning Methods

基於相機與機器學習方法的調酒姿態建議與評價系統建構

專題領域：系統領域

組 別：B375

指導教授：黃朝宗

組員姓名：張智鈞、戴瑞傑

研究期間：2023 年 2 月至 2023 年 11 月止，共10個月

摘要

在調酒領域中，搖盪技巧的多樣性直接影響酒液的混和程度，進而對調酒的口感和風味產生深遠的影響。這個領域一直以來都是大師級調酒師們為人稱道的地方，其優雅而精湛的動作姿勢成為許多初學者學習和模仿的對象。基於這樣的背景，本專題著眼於提升調酒師的技藝，以訓練和姿態建議為指導原則，致力於建立一套高效的輔助系統，協助調酒師在調酒過程中進行動作姿態的精煉和改進。

為了不依賴於在雪克杯上安裝傳感器或要求調酒師戴手套進行追蹤，本專題選擇遠距離的單一相機鏡頭作為解決方案，以實時追蹤雪克杯的姿態。通過標記技術的運用，我們成功實現了對雪克杯的準確追蹤，並基於這些二維圖像的連續數據建構了一個六軸系統，進一步通過演算法找出適當的特徵參數，以達到精確追蹤的目的。

進一步結合機器學習的技術，我們建構了三種不同的分析模型，並進行了相關的分析和比較。通過計算與混合均勻程度相關的混亂度 (Entropy)，我們能夠更全面地了解不同搖盪方式對酒液的影響。同時，透過我們建立的建議系統，我們不僅演算出了最佳的搖盪路徑，還將其與原先使用者的搖盪路徑進行對比分析，為調酒師提供更具體的改進建議。

使用者能夠輕鬆地透過我們建構的系統軟體介面進行操作，該介面不僅顯示使用者當前的速度、加速度和角度等相關數據，還透過簡單而直觀的圖示提示，引導使用者朝著更為優越的搖盪方式邁進。

透過本系統，使用者不僅可以提升其調酒技巧，更能透過系統提供的相關姿態建議，不斷改進並強化調酒師的動作姿態，使其在調酒的藝術中更上一層。

背景與動機 Background/Motivation

在調酒領域中，搖盪技巧的多樣性直接影響酒液的混和程度，進而對調酒的口感和風味產生深遠的影響。而調酒的學習要一步一步的學習、模仿大師的動作。為了讓調酒師在學習過程中能夠獲得更有效率的進步，我們決定建立一個輔助系統，使調酒師能夠獨立進行動作姿態的改進，而無需依賴他人的協助。幫助調酒師能夠透過自己來進行動作姿態方面的改進。

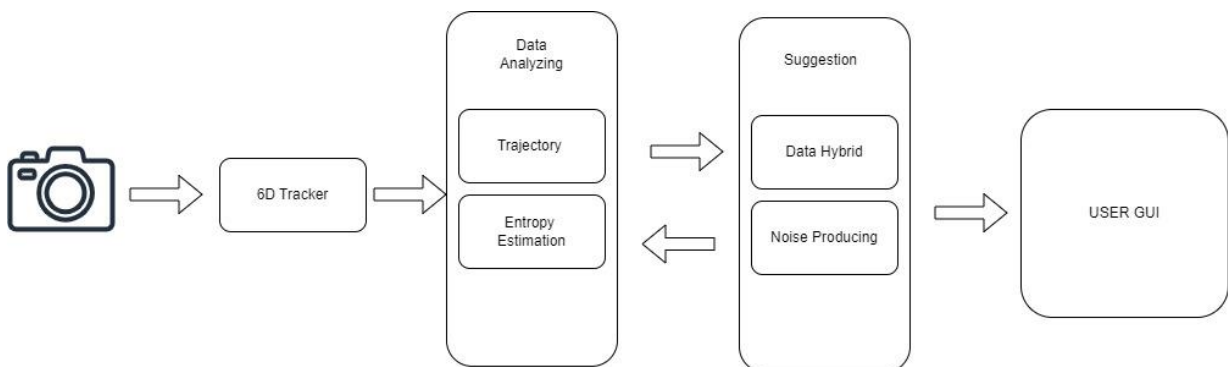
研究目的 Purpose

HomeCourt、Swing Vision 等軟體充分發揮人工智慧技術，透過 iPhone 鏡頭實時分析比賽和訓練影片，為運動員提供即時的姿勢和技巧分析。基於這樣的理念，我們計劃能構建一套專為調酒師提供姿態訓練的系統，旨在訓練出高水準的搖盪技巧，同時能夠確保穩定的調酒品質。本專題利用單相機捕捉畫面，結合機器學習，打造出一套全新的評分與建議系統，我們希望透過機器學習來判斷使用者的搖盪技巧分數，讓使用者透過系統有效地提升搖盪技巧。

研究方法 Method

主系統架構

研究設計的系統主要由相機、六軸追蹤系統(6D Tracker)，數據分析系統(Data analyzing)、建議系統(Suggestion)與使用者介面(User GUI)組成。下方圖1為總系統 Block Diagram。



圖表1: 總系統 Block Diagram

我們透過單一相機鏡頭捕捉使用者的搖盪過程，並將其儲存為影像，以進行後續的深入分析。這不僅提供了非侵入性的解決方案，同時也確保了搖盪過程的自然流暢性。

透過對影片每一幀(Frame)的分析，我們使用標記貼紙的方式來追蹤 Shaker 的位置。結合適當的演算法，我們不再僅侷限於傳統的 X-Y 向量數據，而是透過評估貼紙的大小和角度差異來評估 Shaker 的6軸運動狀態。這種基於2D 影像分析的方法為我們提供了更豐富的6軸資訊，進一步提高了對 Shaker 運動狀況的準確性和分析效率。

我們對取得的六軸資料進行前處理，通過回歸分析將其壓縮和標準化，然後運用不同類型的神經網路進行機器學習並給予評分。結合建議系統，我們生成了優化的虛擬路徑，並與原先的路徑進行比對與評估。通過分析角度、速度等差異，我們為使用者提供了優化調酒技巧的具體指引方向。

最終，使用者可以透過直觀的軟體介面進行操作。該介面不僅呈現了搖盪路徑的相關參數，還提供了改善的建議方向。透過動畫形式，使用者清晰地看到了改進的方向，同時還可以透過回放等功能不斷地進行練習，進一步改善調酒的姿態和技巧。這樣的設計旨在使使用者更直觀地理解和應用系統提供的優化建議，提高他們的調酒技藝。

相機與六軸追蹤系統

這部份我們利用 Aruco Marker 來協助判斷 shaker 的位置以及角度，位置的資訊用 X、Y、Z 來儲存，而角度的資訊用3*3旋轉矩陣表示。透過分析影片的每一幀，最後可以量化 Marker 的位置和角度。並交由之後的機器學習得出路徑和搖盪好壞的關係。

在使用相機讀取 Aruco Marker 之前我們必須先校正相機，防止拍攝的 Marker 經過鏡頭變形，進而影響準確性。在拍攝時，我們也要確保 shaker 不會因為快速搖晃而導致動態模糊因此我們選用有高速攝影模式的相機，並固定他的焦距以防在高速搖盪 shaker 時突然變焦導致模糊，並且經過測試發現快門時間為1/750能夠拍攝出清晰的 Marker 圖案，並且曝光時間不會太短導致圖片太暗。

混亂度

混亂程度的評估部分，我們透過計算鄰近像素的離散資訊熵 (Discrete Information Entropy) 來量化混亂程度，這是一種用於衡量系統混亂程度或不確定性的數學概念。離散資訊熵在信息理論和統計學中被廣泛應用，其目的是描述系統中的信息量。具體而言，離散資訊熵衡量的是在一個離散概率分布中，得到一個具體的值所需的平均信息量。

在搖盪前，我們特意將黑豆和白米分層放置到 Shaker 中，形成大範圍單調的色塊。我們設定在搖盪均勻時，鄰近像素中應該出現黑豆和白米的交錯，這被定義為高混亂度。相反，如果鄰近的像素相似，即黑豆附近是黑豆，或白米附近是白米，則被定義為低混亂度。

基於這一概念，如果鄰近範圍的像素越相似，帶來的信息量就越低，因為這個事件相對預測可能性較大。反之，如果事件發生的機會較低，即搖晃的結果呈現均勻的交錯分布，帶來的信息量就越高，因為這個事件發生的概率相對較為驚奇和意外。

此專題中我們先將圖片灰階化處理後，透過設定固定大小的窗(Window)後掃描圖片，統計每個窗內的像素強度種類並定義為「事件 k 」，同時也定義事件 k 在該窗發生的機率為 p_k ，若該窗內的像素們都相似且單調則我們會獲得較小的離散資訊熵，反之複雜且多變化我們則可獲的較大的數值，最後統計整張圖片中超過閾值(Threshold)的所有窗，並與全部窗數量進行比較後給予混亂程度值(Entropy)。

路徑

由於我們會記錄每一幀的位置以及角度，這會導致最後的資料量非常龐大，加上 3×3 旋轉矩陣其實只有 3 個自由度，導致機器學習做出來的效果不佳，因此我們決定將資料做簡化與壓縮。由於在拍攝時 shaker 的位置和角度都比較固定，所以我們先把原始路徑做初始化將初始點都設為零，然後使用回歸分析來壓縮數據，主要想法為將每次 shaker 的「來回」看成一條路徑。角度方面則是轉換成球座標系中的 $\theta(\theta)$ 、 $\phi(\phi)$ ，利用這段時間的路徑點和角度來回歸成三條路徑多項式(X 、 Y 、 Z)和兩條角度多項式(θ 、 ϕ)，並且考慮 X 、 Y 、 Z 的變化度(Z 不太隨時間變動，而 Y 變動很大)，我們也加上了原始路徑的平均值、標準差，來讓不太做隨時間變動的軸分析。

此外，由於未限制使用者搖盪秒速以因應現實狀況，我們萃取的資料會有長度不同的影響，而因為我們所使用的機器學習模型架構，我們使用了零填充(Zero-Padding)保持特定維度的長度。以確保神經網路能夠正確運作並有效地處理數據。

機器學習模型

由於資料的時間特性，路徑的分析可能與前面幾幀的六軸資訊有關，經過前處理後的資料在以特徵的方式表示一段時間內的資訊。基於此，我們須選擇多特徵輸入的時間序列評估模型。我們挑選了 GRU、LSTM、Transformer 架構進行改造。

建議系統

透過上述的機器學習模型，我們能夠評估一段影片中 shaker 的搖盪方式的優劣。我們期望協助使用者在搖動過程中取得更佳表現，因此提供了相應的建議。首先，我們將使用者生成的路徑引入一些噪點，然後透過評分系統挑選分數較高的虛擬路徑。最終，我們提取虛擬路徑的特徵值（例如平均速度、順時速度、最大角度等），並與使用者的實際路徑進行比較，以提供有針對性的建議。

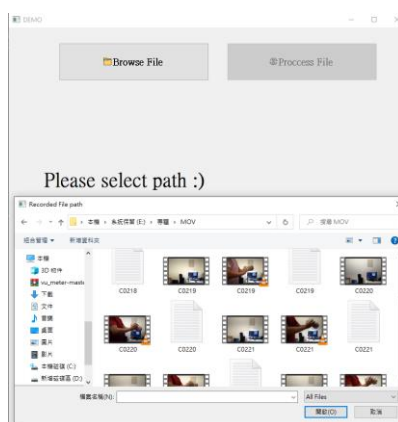
這個建議系統的獨特之處在於，它以使用者實際的搖動路徑為基礎生成更為優越的建議，而非僅僅參考資料庫中的最佳路徑。這種方式更貼合實際操作，使得建議更加切合實際情境，提升了學習的實用性。此外，這個系統還可以透過虛擬的模擬進行多次演算，進一步加強了使用者在不同情境下的搖動技巧。

使用者介面

為了方便使用者透過直觀的圖形界面進行操作，我們採用了 QT 進行使用者介面的開發。我們設計的介面不僅允許使用者輸入影片檔案進行進一步的分析，而且在分析完成後，介面會呈現推薦的路徑和姿態修正建議供使用者參考。同時，介面還會顯示角度、速度等相關資訊。為了使使用者更清晰地理解動作的建議方向，推薦改善姿態的方式將以生動的動畫形式展現。這使得使用者可以同時觀看原始輸入的影片，按照建議路徑重新練習，並實際看到六軸定位，從而確認和調整自己的動作姿態，進行更有效的修正。這樣的介面設計旨在提供使用者直觀而有力的學習工具，使其能夠快速理解並應用改進建議。

研究結果 Results

使用者介面操作



圖表 2: GUI 展示-選取路徑，使用者選擇欲分析的影片檔案

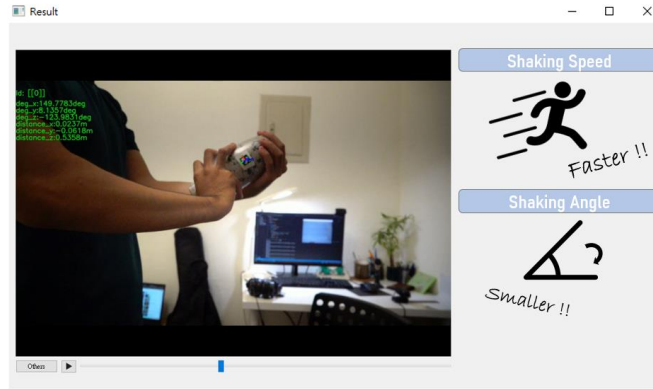


圖 3: GUI 展示-分析結果，使用者能夠重複回放並且練習搖盪技巧，也能選擇顯示速度等其他功能

資料探勘

在資料前處理我們將資料簡化來壓縮數據，最後結果將3267行的資料(27秒的影片)壓縮成171行的資料，最後結果相當於壓縮了約95%的資料量。並且在多項式方面，我們採用 X 、 Y 為3次多項式， Z 、 θ 、 ϕ 為2次多項式，並且加入了原始路徑的平均值、標準差，來讓不太做隨時間變動的軸分析(如 Z 軸)。

建議系統

而在建議系統部分，我們給出了兩種建議方向，一種是比較廣泛的建議方向(如瓶口要盡量拿正)，另一種則是根據使用者所搖盪的路徑來給出建議(如速度要加快、角度要大一點)。



圖 4: 初步提供的建議方向



圖 5: 根據使用者所搖盪的路徑來給出量化後的建議

混亂度與機器學習

機器學習方面，因我們專題的資料蒐集難度較困難，每次的蒐集約需2分鐘，包含重新分類與裝填豆子、拍攝影片與混亂度照片、擷取圖片等，因此我們雖花費大量時間蒐集數據仍舊未能達到我們所預期的結果，技術的部分我們也嘗試過數據增強 (Data Augmentation)，透過轉換座標位置嘗試複製相似路徑並擴充數據數量，但仍舊因為數據的特性較難有好的成效與結果。

由於混亂度是一項回歸任務，我們透過回歸指標(Regression metrics)進行模型好壞的評估，而此專題中我們採用了平均均方誤差(Mean Squared Error, MSE)進行評估。而我們比較的對象是平均模型 (Mean Model)，該模型簡單地使用訓練集中所有目標變量的平均值作為預測。若我們所佈署的模型能夠超越此簡單模型，代表該模型有可能能夠更精準的預測結果。

Model	測試集資料平均 MSE	模型效能(Mean Model 為基準)
Mean Model	198.2685852	1
GRU	146.2649719	1.2622
LSTM	162.9964508	1.1779
Transformer	225.7154846	0.8616

上表中的模型效能我們定義為:

$$Performance = 1 - \left(\frac{MSE_{Model} - MSE_{Meanmodel}}{MSE_{Meanmodel}} \right)$$

從圖中我們可以看到 GRU 的泛化能力較適合我們的專題，由於其構造簡單且由於我們的資料數不夠龐大，GRU 能夠較快速地訓練，測試集中的資料能夠有較少的誤差，而反觀 Transformer 結構較複雜，雖然在訓練集中能夠有效地減少誤差但在測試集中的效能未能夠比 Mean Model 來的優秀，因此我們選擇了 GRU 作為我們系統的評估模型。

總結 Conclusion

我們選擇了一個有趣且具有挑戰性的題目作為我們的實作專題。由於過去缺乏相關的文獻和參考，我們必須參考類似的路徑追蹤、運動姿勢評估等領域的論文，以建立我們的專題。在這個過程中，我們面臨了多個障礙，例如如何有效地捕捉對稱的 Shaker。我們曾考慮使用數位影像訊號處理或機器學習等方法來進行分析，但由於系統開發流程和時間的限制，六軸定位的演算法需要更多的深入研究才能穩定運作。因此，我們最終選擇了標籤的方案。

此外，我們也遇到了動態模糊的問題，需要研究相機快門速度的相關性，了解動態模糊的成因，並尋找可能的解決方案。最後，通過與教授和同伴的討論，我們成功地利用相機解決了這個問題。

對於模型的部分，我們也遇到了許多挑戰，包括資料的蒐集和優化。我們必須不斷探索如何優化模型架構並改進現有的方法。

總結來說，這次的專題研究讓我更深入地了解了如何有效地規劃一個系統。在每個層面，我們都學到了許多新知識，並討論出了相對應的解決方案。這個專題對我們來說是一個非常具有挑戰性的項目，因為我們必須從無到有地建立一個完整的系統，而且在缺乏參考來源的情況下，我們必須細心地規劃系統的開發步驟和流程。我們不斷討論問題，修正錯誤，並尋找更多解決問題的可能性。這些挑戰讓我們學到了許多課堂中所未學到的知識，並提升了我們的研究解決能力。未來，我將更自信地面對未知的挑戰和困難，因為我們知道我們有能力不斷嘗試和找到答案。

心得感想

張智鈞：

首先得先感謝黃朝宗教授的指導，給了我和瑞傑許多指引的方向，由於我們選擇的題目是前人未曾做過類似的題目，我們必須花費更多時間去尋找解決方案和提前假設可能會遇到的問題，以六軸追蹤來說，我們曾經想過使用3色塊標記的方法，透過 DSP 進行處理，又或者是直接訓練另一個模型去辨識 Shaker 即時定位與追蹤，但考量到我們時間有限且須開發出完整的系統，我們最後選擇了 Aruco Marker 的方案，較實際也能夠確保我們題目能夠順利進行。動態模糊也困擾我們，也花了許多時間去思考成因與解決方案。我認為我從這次的專題學到了許多課堂上學不到的知識，那些更實際且更需要考量到現實的部分，遇到問題時沒有完美的解答，而是需要不斷地搜索，嘗試實驗後才能夠逐步找到目標方向，這方面也很感謝教授的指導，我和瑞傑在過去這方面經驗的不足經由教授的指示後能更快尋找到方向並解決問題，也很感謝我的隊友瑞傑，問題討論與專題分工的部分我們都合作的十分愉快，也很樂意一起切磋並共同成長，感謝這份專題讓我們我們能見識到課堂外不同的新知，還有獲得那些問題探索與解決能力的寶貴經驗。

戴瑞傑：

在本次專題我們要開發出一個應用程式，在抽象概念到具體實現的過程中，不論是專題題目、系統架構、各個系統的實行方法都是我和智鈞一步一步討論出來的，這也是在大學期間第一次碰到可以完全依照自己想法來實作的一門課。而本次專題的領域比較偏向演算法，這也是我們在大學期間比較少接觸過的東西，不過這也正好讓我們練習碰到完全不懂的領域時要去如何找資料。而在實作的過程中，我們也經常碰到需要因為某些問題導致之前的系統需要重頭改寫的問題，以六軸追蹤為例，當時我們只有用路徑點當作資料來儲存，到了機器學習的部分，才發現這樣訓練的效果並不好，所以後來又回頭加上初始化以及迴歸分析的部分。雖然在重寫系統的時候花費了很多時間，但這也讓我們獲得了一些寶貴經驗，在設計整個系統架構時需要有更細節的思考，考慮每個系統之間的關係再來下手實作，而不是埋頭苦幹，最後只能重新改寫。最後我也要感謝黃朝宗教授在我們遇到想法瓶頸時給我們指導方向，也要感謝我的隊友智鈞在做專題時的相互幫助，讓我們能分工合作完成這次專題。