

國立台北大學資訊工程學系專題報告
應用在無人機上使用 YOLOv4 和 Deep SORT 的圖像追蹤技術
Image tracking using YOLOv4 and Deep SORT apply on UAV

專題編號：NTPUCSIE-109-005

執行期間：2020 年 09 月至 2021 年 05 月

一、摘要

隨著科技的進步與網路的發展，物聯網在各領域的應用上不斷推陳出新。在無人飛行機逐漸普及的現代，其所具備的功能也因使用者的喜好迅速發展。本專題整合電腦 server 端、用樹莓派控制的追蹤系統和無人機。在電腦 server 端選擇追蹤的對象，透過影像識別和機器學習技術，判斷追蹤對象的位置是否改變，並告知無人機採取對應的動作，使追蹤對象維持在鏡頭畫面中，以達到目標追蹤的目的。

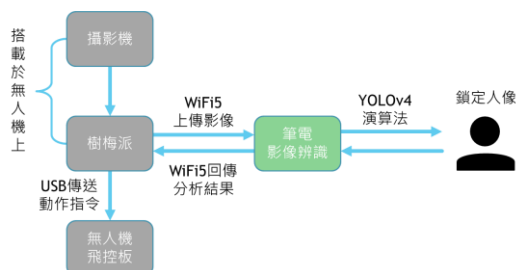
二、簡介

無人飛行機的傳統控制方式，是由使用者手動操縱遙控器，透過無線電波來與飛行機連線並控制其飛行。而初學者要學會操縱遙控器往往需要花費相當長的一段時間。本專題的目標在於，設計出一個以軟體來控制無人飛行機飛行的選擇，將無人飛行機所拍攝到的鏡頭，傳回地面的 server 端，透過影像辨識篩選出可進行追蹤的人物，經使用者選擇後，以 server 運算的結果，對無人飛行機下達對應的飛行指令，將欲追蹤的人像維持在鏡頭中央，達到人物追蹤的效果。

三、專題進行方式

(1)系統分析與設計摘要

系統流程：影像傳輸->影像辨識->人物特徵追蹤->追蹤人物選取->調整無人機位置



設計摘要

(2)實作平台與技術

1.平台與工具

此系統主要跨足在樹莓派與 Windows 電腦間，使用 python 作為開發環境，結合 opencv、github、dronekit-python 等套件與相關工具。硬體設備包含無人飛行機(含飛控板)、樹莓派、USB 攝像頭、windows 電腦。

2.影像傳輸

開始拍攝時，透過樹莓派以 TCP socket 的形式，把影像傳輸到 server 端進行影像的辨識。

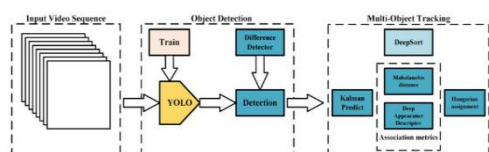
3.影像辨識

透過 opencv 函式庫中的 VideoCapture 將影像一幀幀地讀取，使用開源程式碼 YOLOv4 作為人物辨識的物件檢測器，將影像讀取作為輸入後，輸出該幀影像中所含有人物定界框 (bounding box, bbox)。

4.人物特徵追蹤

使用開源程式碼 Deep SORT 的特徵追蹤技術，將辨識出的人物作特徵的追蹤。其中包含卡爾曼濾波 (Kalman filter) 和匈牙利

(Hungarian) 演 算 法 。



人物特徵的提取是透過將動作矩陣 (motion matrix) 和 外觀矩陣 (appearance matrix) 結合在一起，得到新的成本矩陣。

(i: 追蹤目標定界框的 id;

j: 檢测定界框的 id)

動作矩陣是將追蹤目標做空間上的投影，透過馬氏距離 (Mahalanobis distance) 算出追蹤目標定界框與檢测定界框的距離。

$$d^{(1)}(i, j) = (d_j - y_i)^T S_i^{-1} (d_j - y_i)$$

Motion matrix

外觀矩陣是把檢測的定界框，透過 CNN 網絡計算出對應的 128 維特徵向量 r_j

Name	Patch Size/Stride	Output Size
Conv 1	3 × 3/1	32 × 128 × 64
Conv 2	3 × 3/1	32 × 128 × 64
Max Pool 3	3 × 3/2	32 × 64 × 32
Residual 4	3 × 3/1	32 × 64 × 32
Residual 5	3 × 3/1	32 × 64 × 32
Residual 6	3 × 3/2	64 × 32 × 16
Residual 7	3 × 3/1	64 × 32 × 16
Residual 8	3 × 3/2	128 × 16 × 8
Residual 9	3 × 3/1	128 × 16 × 8
Dense 10		128
Batch and ℓ_2 normalization		128

對於追蹤目標 i，為它創建一個集合 R_i ，儲存過去所有比對成功的 k 個特徵向量集合。在比較完集合中所有已知的外觀特徵後，求出與編號 j 定界框的最小餘弦距離。

$$d^{(2)}(i, j) = \min\{1 - r_j^T r_k^{(i)} | r_k^{(i)} \in R_i\}$$

Appearance matrix

最後，藉由加權公式將動作矩陣

與外觀矩陣結合在一起，得到成本矩陣。

$$c_{i,j} = \lambda d^{(1)}(i, j) + (1 - \lambda) d^{(2)}(i, j)$$

Cost matrix

($0 \leq \lambda \leq 1$)

卡爾曼濾波負責以影像前一幀目標的位置，預測目標下一幀的位置。透過深度學習取得目標的外觀特徵向量，結合動作向量轉換成成本矩陣，最後由匈牙利演算法進行定界框的比對，完成目標的追蹤。

5. 追蹤人物選取

透過 opencv 函式庫中的 mouse event，將滑鼠座標點選的位置與人物定界框匹配，追蹤特定 id 的人物。

6. 調整無人機位置

透過 dronekit-python 函式庫，在樹莓派上將接收到的運算結果，轉為對應的飛行指令，透過 USB 傳給無人飛行機的飛控板，藉此進行改變其位置，使人物維持在鏡頭畫面中央。

四、主要成果與評估

(1) 主要成果

使用上，能直接透過滑鼠左鍵來選取要的人物，並透過右鍵來取消追蹤，如果人物未做大幅度的動作，無人機就能達到長時間穩定追蹤。

(2) 評估

無人機追蹤時常被機身的穩定性直接影響，比如四軸飛行器水平飛行時需要傾斜機身，人物容易離開鏡頭。再者，追蹤的運算是在筆電上進行，缺乏穩定的網路也是明顯的降低了影像的 FPS，直接導致了追蹤人物瞬移並導

致誤判。因此，還是建議使用更加穩定的機身並配合優良的網路或是直接在機體上進行用算。

五、結語與展望

(1)結語

此次的研究已能達到基本的追蹤能力，準確率也達到一定水準。YOLOv4 及 Deep Sort 都只是辨識及追蹤的其中一種演算法，未來希望能找出更好的組合。

(2)未來展望

在避障方面，我們還尚未著手。目前可用的有紅外線感測及聲納，兩者都是不錯的感測設備，但都有各自的明顯缺點，兩者的結合應用應可成為日後的新研究目標。

六、銘謝

非常感謝專題教授每周的用心指導和提供無人機的相關設備，以及實驗室學長在無人機上的相關經驗，使我們在唯有二人的情況下還能在期限內完成這次不簡單的專題。

七、參考文獻

- [1] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934
- [2] Nicolai Wojke, Alex Bewley, Dietrich Paulus. Simple Online and Realtime Tracking with a Deep Association Metric. arXiv:1703.07402