

Using Hierarchical DBSCAN to construct the LIDAR data clustering method

學生:黃廷睿 指導老師: 陳世芳、郭彥甫 博士

一、前言

現代農業人口逐漸老化，為緩解農業勞力缺乏的困境，使用影像進行作物監測的技術應用十分廣泛，尤其如蘆筍這般密集度高、生長迅速、易受環境影響的作物，更需要人力頻繁巡田以保障作物品質及產量，因此我們已開發一台田間行動載具，以進行場域作物即時生長影像監測，本專題期間希望藉此機會改善現有的自走算法，利用合適的分群方法技巧，透過光學雷達找出行走車道資訊，建立導航輔助線以取得車行誤差作為修正依據，未來將其結果用以輔助車體自動行駛於兩畦之間走道進行蘆筍生長資訊之取得。

二、材料與方法

(1)載具介紹

本研究已開發了一台田間監測載具(如下圖 1)，行動載具車主要使用樹莓派控制器(Raspberry Pi 4 Model B, Raspberry Pi Foundation, Cambridge, UK)，配合底部安置四顆高扭力馬達(IG-42GM-02TYPE-24V, 祥儀, 桃園, 臺灣)，及四顆具高摩擦力的越野橡膠輪胎，以利於農田中行走。車體前方配置一光學雷達(hokuyo ust-05lx)，用於補足純視覺導航無法克服的場景。

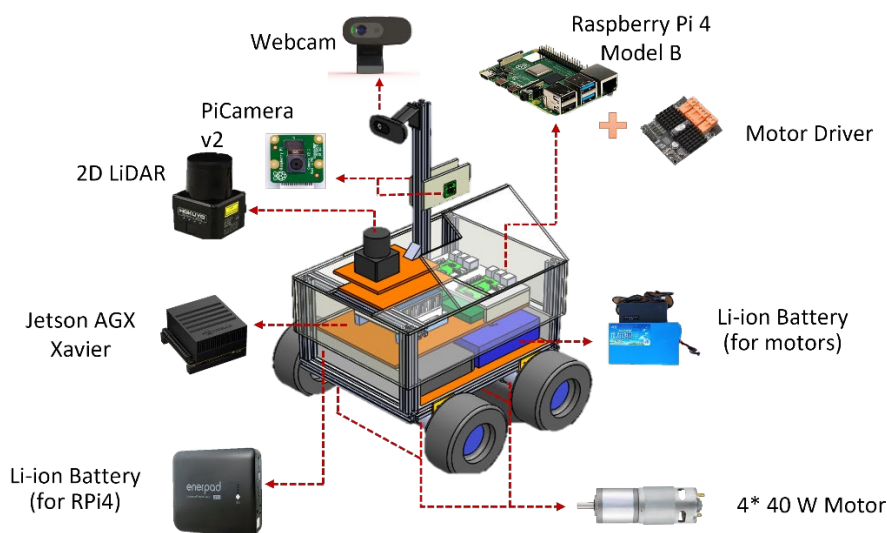


圖 1 載具元件配置圖

(2)點雲分群方法

原先在點雲分群的方法選用(Density-Based Spatial Clustering of Applications with Noise, DBSCAN) 算法 [1]，此方法特色便是不用決定最終分群數量，算法會根據設定的最小搜尋半徑與數量點自動決定分群數量，並且可以排除雜訊影響。然而此方法對被目標進行一假設：被分群的群集都具有相同的密度分布，這假設在實際蘆筍溫室場域中並不能很好的成立，因蘆筍的分布疏密程度變化很大，由圖2之掃描結果中可發現不同位置的密度分布具有相當大的差異，因此若是採用 DBSCAN 這類將密度假設為相同規模的算法容易將群集分成過多小區域，不利後續取得參考線。

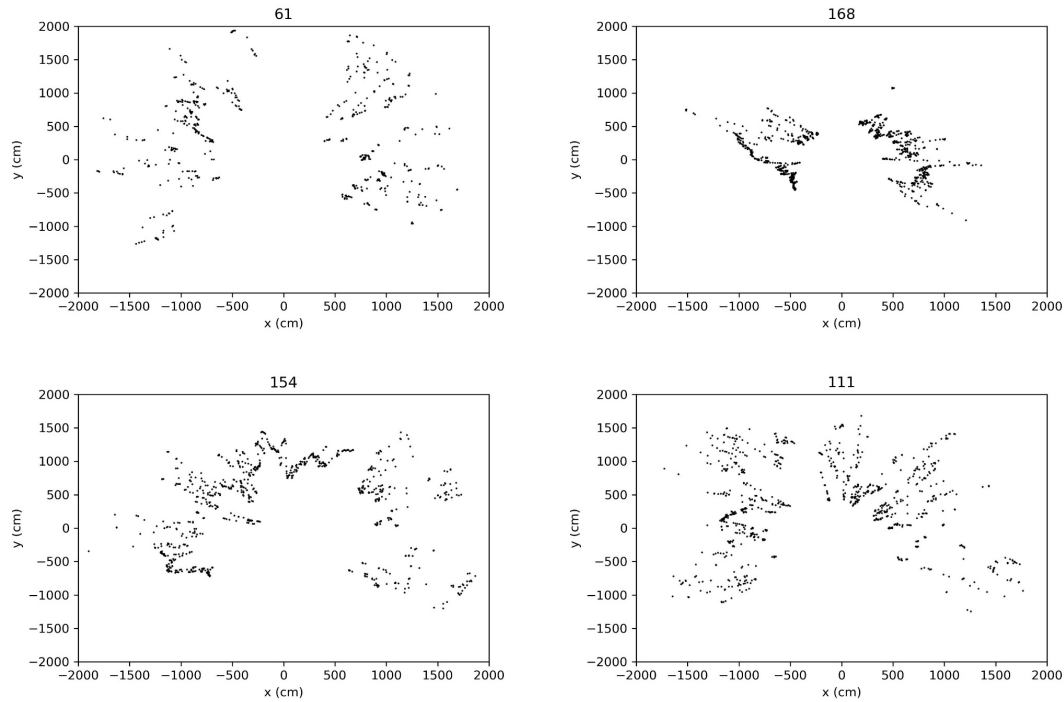


圖 2 蘆筍場域點雲掃描結果

因此本次研究採用了一種基於 DBSCAN 的改良方法，稱為(Hierarchical DBSCAN, HDBSCAN) [2]，此算法主要利用各點間的距離建立了最小生成樹，並藉此建立階層結構圖 3，透過設定的單一群集最小數量點自動修剪合適的分支結果，以達到較為彈性的密度選擇

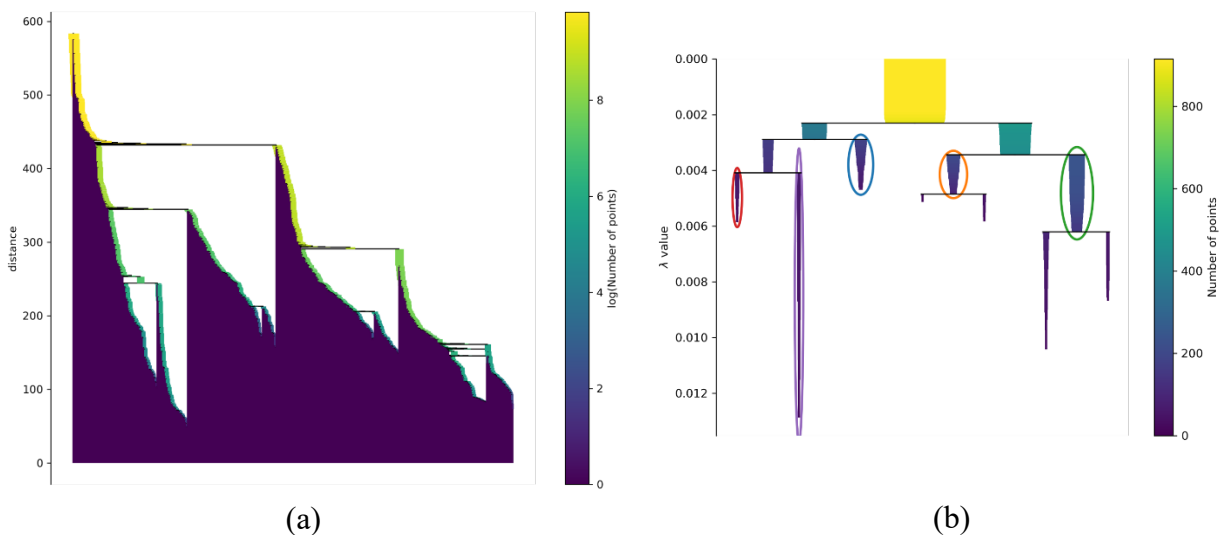
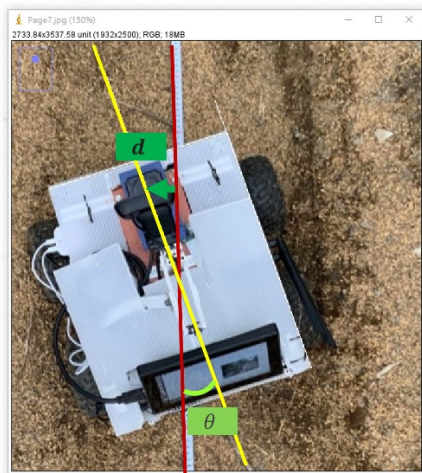


圖 3 HDBSCAN 建立的階層結構 (a) 原始結構、(b) 經篩選後的分群結果

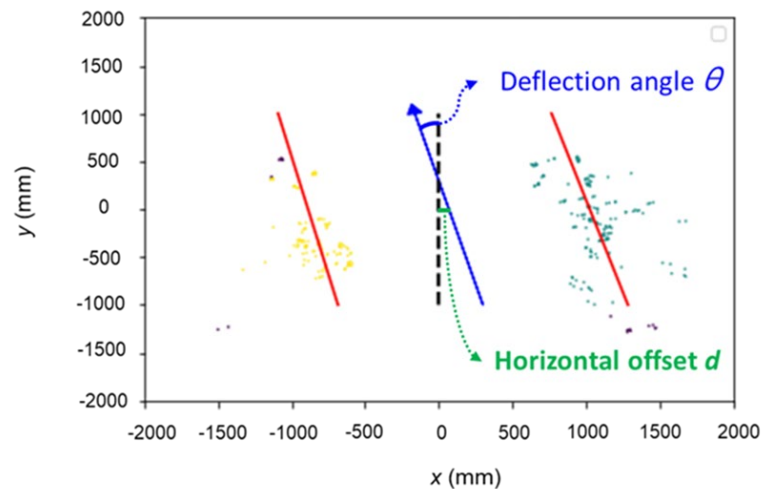
(3)導航參考線建立方法

本研究自台南區農業改良場義竹工作站蒐集共 63 筆資料以作為數據集，包含此刻的車輛俯視影像以及光達掃描的原始數據，利用 ImageJ 軟體進行標記載具與車道實際中心(畫面中白色捲尺)的平移誤差(Horizontal offset)、偏向角度(Deflection angle)作為本次實驗的 Ground truth，如圖 4(a)。並且由分群方法分出的群集選擇最靠近畫面中心左右兩側的目標進行迴歸分析，

以取得分群的迴歸線，作為車輛左(右)側的邊界，並將左右邊界進行平均即可得到利用此刻分群結果的導航參考線，並和影像資料一樣計算與中心的平移誤差(Horizontal offset)、偏向角度(Deflection angle)，如圖 4(b)，兩筆資料利用均方根誤差(root-mean-square error, RMSE)進行評估。



(a)



(b)

圖 4 導航參考線之建立 (a) ImageJ 軟體標記 (b) 由分群結果取得的資訊

三、結果與討論

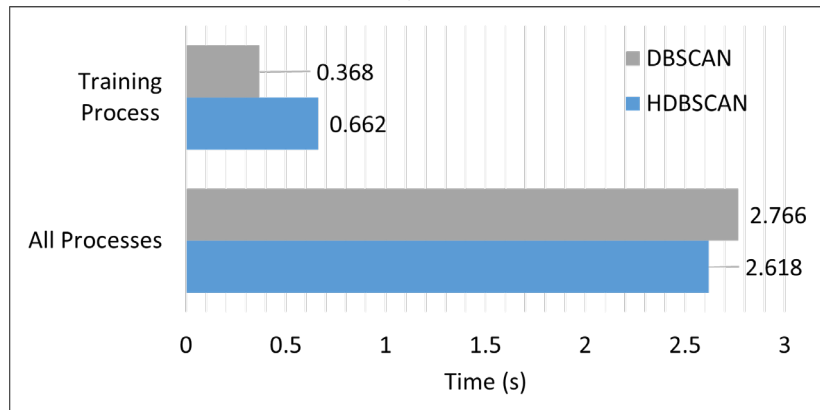
將 63 張有標記資訊的資料進行分群分析，並且與 Ground truth 比較，得到導航評估結果如表 1 所示，HDBSCAN 在偏移角度上比起 DBSCAN 準確率提升了 32.9 %，在平移誤差上更是提升了 116.8 %，此結果表示這項分群技術確實有可見的效果提升。

表 1 參考線結果比較

	Deflection angle (deg)	Horizontal offset (mm)
DBSCAN	13.19	313.64
HDBSCAN	9.92	143.67

在執行時間上同樣也進行測試，將 63 張照片依序進行分群並將時間加總，比較兩時間依據，一為單純 HDBSCAN 分群時所花費的時間，目的是為了判斷此方法是否需要較高的計算量，二為計算執行全部流程所需的時間，包括分群迴歸、將結果可視化等等流程，結果如表 2 所示，可發現在分群階段確實需要較多的運算時間，但在整理流程下兩者的時間近乎相同，新的 HDBSCAN 算法甚至花費了較低的時間。考量到因車輛實際在場域運作有一顯示螢幕即時觀看行走與辨識狀況，因此對於整體運行效率應並無拖累的影響。

表 2 執行分群花費時間比較



實際辨識結果如圖 5 所示，經目測觀察分群結果，HDBSCAN 確實表現較 DBSCAN 優秀的多，可以較好的修正 DBSCAN 會分類出過多群集，造成迴歸線取得效果不好代表當下情況的缺失，並且在無法分群的狀況也減少許多。

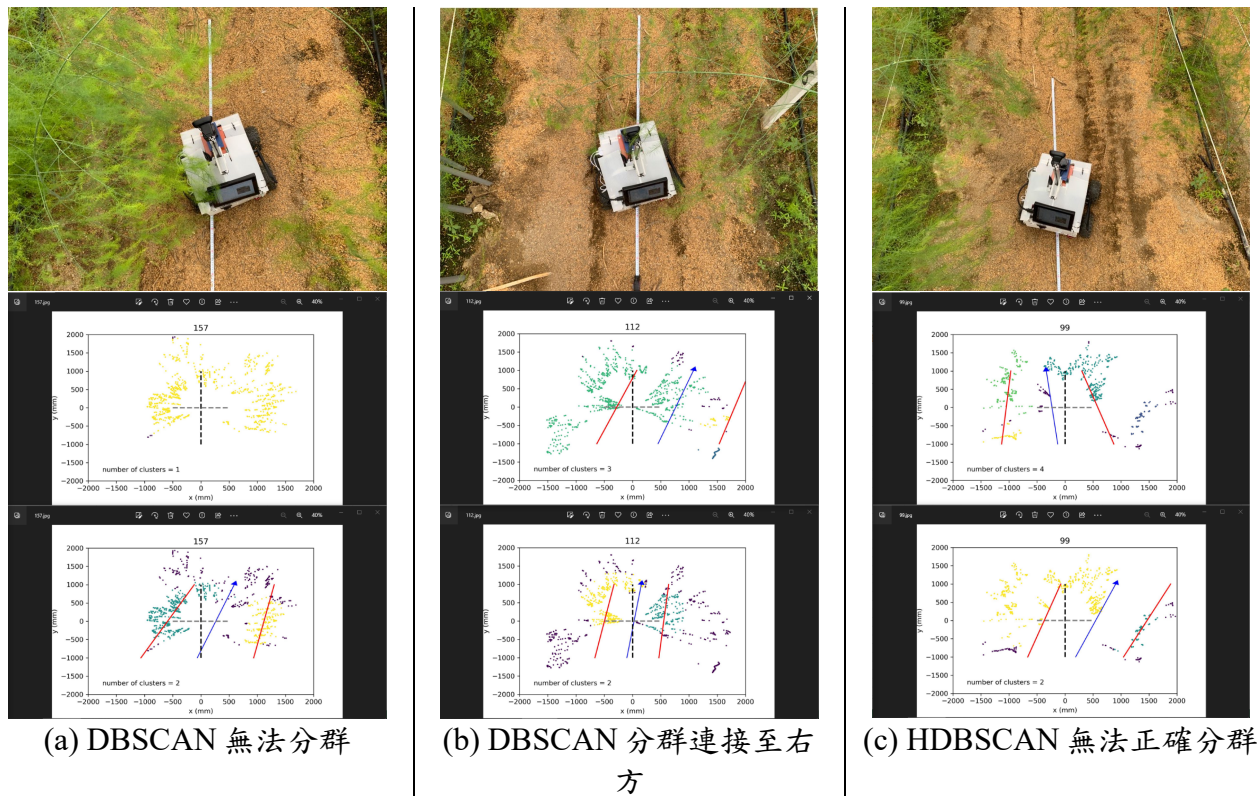


圖 5 分群方法與實際俯視圖比較(上:DBSCAN, 下:HDBSCAN)

三、結論

本研究以目前辨識結果看起來此方式是可行的，但由於能夠與實際結果比較的資料集目前還太小，因此未來需要擴增各種情況及其他場域的資料進入資料集當中，並能利用較適當的實驗方法以取得適合的模型參數，並且也需要讓車輛實際在溫室場域進行實地驗證，才能證實此方法確實可以運用在場域，才能確定其運行效果如何。

四、參考資料

- [1] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).
- [2] Campello, R. J., Moulavi, D., & Sander, J. (2013, April). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160-172). Springer, Berlin, Heidelberg.