

以單眼視覺 SLAM 建立場景三維特徵地圖輔助相機 定位之定位精度與適應性分析

林緯程^{1*} 饒見有²

摘要

基於影像特徵點之視覺 SLAM (Visual Simultaneous Localization and Mapping, V-SLAM)，以重複的三維特徵地圖建立、影像特徵至三維特徵地圖的匹配和空間後方交會過程定位相機。若其建立之三維特徵地圖能做場景先驗控制重複使用，則可解決無可靠 GNSS 訊號環境之定位問題。然環境光照條件常影響影像特徵點萃取和匹配結果，因此值得進一步探討不同光照條件下建立之三維特徵地圖，其相機定位成果對光照條件改變之適應性。本研究輸入不同光照條件下建立之三維特徵地圖至 ORB-SLAM 系統，藉其地圖再利用功能輔助相機定位。研究成果表明，三維特徵地圖輔助之相機定位，能獲接近參考地圖精度的相機定位精度，為無可靠 GNSS 訊號環境之定位問題的可能解決方案。

關鍵詞：ORB-SLAM、三維特徵地圖、坐標轉換

1. 研究動機與目的

1.1 無 GNSS 訊號環境之定位

隨全球衛星導航系統 (Global Navigation Satellite System, GNSS) 技術及導航演算法的發展，在有可靠 GNSS 訊號的環境中，移動載具可以透過 GNSS 訊號的接收，進行直接地理定位。以無人飛行載具 (Unmanned Aerial Vehicle, UAV) 為例，其靈活、高機動性、低拍攝環境限制等的特性，使其可以深入建築構造物或危險環境中，進行過往之載人航空攝影因其飛行限制而難以達成之拍攝任務。攝影測量的角度而言，此一進展更能使載具按照預先規劃的航線，進行自主導航，獲得穩定的影像重疊比例，確保立體製圖品質。

然而實際應用狀況中，仍會不可避免地碰到室內或 GNSS 訊號受阻隔的室外環境之應用需求，此屬無可靠 GNSS 訊號之環境，將使移動載具之定位能力受限，難依預先設定之路線進行移動或導航，

因此須尋求額外手段解決定位問題。

近年來研究的目光逐漸轉向透過即時定位與製圖 (Simultaneous Localization and Mapping, SLAM) 方法來解決無可靠 GNSS 訊號環境之定位問題。Taheri and Xia (2021) 將 SLAM 概念定義為在缺乏對場景的先驗認知下，建立環境地圖，並定位感測裝置的技術。現今多數移動載具皆有搭配相機，使基於影像的視覺 SLAM (Visual-SLAM, V-SLAM) 成為熱門的解決方案。

1.2 場景地圖輔助相機定位

鑒於基於特徵點法的 V-SLAM，是萃取影像特徵點，並和建立之三維特徵地圖進行特徵匹配，達成於未知環境的定位目的。倘若此一建立之三維特徵地圖能被保存，作為對場景的先驗控制，供後續定位目的重複使用，則移動載具面臨的無 GNSS 訊號環境定位問題將能解決。

此場景地圖輔助相機定位概念，為一在無可靠 GNSS 訊號時輔助定位的可能手段，在 GNSS 訊號

¹ 國立成功大學測量及空間資訊學系 碩士

² 國立成功大學測量及空間資訊學系 教授

* 通訊作者, E-mail: p66094133@gs.ncku.edu.tw

收到日期：民國 111 年 01 月 21 日

修改日期：民國 111 年 02 月 09 日

接受日期：民國 111 年 03 月 07 日

受限時，持續提供移動載具定位坐標。Nobis *et al.* (2020) 在 ORB-SLAM (Mur-Artal *et al.*, 2015) 中添加了地圖保存的功能，並藉 ZED 立體相機拍攝影像，建立三維特徵點雲地圖，做為自動駕駛的定位輔助工具。其分別透過公開測試資料集和自行拍攝之測試資料進行測試。其研究成果顯示，隨載具移動速度加快，相機定位誤差有增長的趨勢，但在低速應用的環境，仍不失為一可行的輔助定位手段。

1.3 坐標轉換與光照條件改變

現今提出的視覺 V-SLAM 系統中，已有許多不同種類的相機被採納。主流的 V-SLAM 演算法常採用單眼相機 (Monocular)、雙目立體相機和結構光相機。室內環境應用中，雙目立體和結構光相機是最受歡迎的選擇。其不同於單眼相機，可以直接於觀測中獲取並使用帶有尺度資訊的成果，無須再經過坐標轉換。然而室外場景的應用而言，雙目立體相機受限於其兩鏡頭間的基線長度，隨相機和被拍攝物距離增加，其對深度的解析能力會逐漸下降，最後退化至和單眼相機並無二致；基於結構光的相機，其仰賴於相機本身主動投射的圖案獲取尺度資訊，在室外場景中易受強光和拍攝距離影響而無法使用。因此室外環境中，最能廣泛應用的選項仍屬單眼相機。

然而使用單眼相機作為感測裝置的最大問題，是其僅有二維的影像觀測，無法直接從影像獲取第三維度資訊。即便透過多張影像光束交會建構三維資訊，其成果仍位於尺度未知的獨立坐標系中，需再增加標定控制點等額外手段的觀測以進行坐標轉換，方能獲得帶有尺度資訊的觀測成果，供後續應用。

此外，影像特徵萃取和匹配技術是基於影像之灰度值計算而得，其成果常受到環境光照條件影響。Wang *et al.* (2020) 以 13 個不同特徵點萃取演算法，進行了光照條件改變對特徵萃取成果影響的研究，其實驗成果表明，在多數特徵點萃取演算法中，光照方向的改變對特徵萃取、匹配並不會有明顯影響，而是曝光值的影響較為顯著。Ranganathan *et al.* (2013) 亦表示光照條件改變會影響特徵描述子的

建構，進而影響匹配結果，同時其文獻中的前導實驗成果亦表明，在相同位置但不同時間拍攝的影像，其特徵匹配數會隨時間推移而發生明顯改變。

對室外場景應用而言，環境光照條件改變明顯，因此對基於影像特徵萃取技術建立的三維特徵地圖，以其輔助的相機定位對環境光照條件之適應性值得進一步探討，其適應性包含 (1) 光照條件改變對 V-SLAM 建圖精度之影響，(2) 以一光照條件下拍攝的影像建立之三維特徵地圖，輔助不同光照條件下之定位目的使用時，影像與地圖的光照條件不匹配，對定位結果造成之影響。

1.4 研究目的

針對以上所列之問題，以及本節所述之內容，本研究認為，若以預先建立之三維特徵地圖做為拍攝場景先驗控制，則移動載具於無 GNSS 訊號環境所面臨的定位問題則可獲得解決，其中又以單眼相機之適用範圍最廣。然而目前鮮有以三維特徵地圖輔助相機定位之相關研究，因此在實際應用於室外環境前，需先對其能提供之相機定位能力進行評估。此外，影像特徵萃取、匹配成果容易被環境光照條件所影響，因此其對相機定位之適應性亦值得進一步測試。本研究作為實際應用三維特徵地圖輔助相機定位概念於實際使用場景前的先行研究，主要著重於三維特徵地圖輔助相機定位之定位精度分析，檢驗此概念所提供之相機定位精度。條列本研究之目的如下：

- (1) 提出一套適用於單眼 V-SLAM、且不仰賴影像序列參考軌跡的坐標轉換模式，求解獨立坐標系和物空間坐標系間的轉換參數，供影像序列軌跡及所建地圖進行坐標轉換。
- (2) 以提出之坐標轉換模式，評估測試用之 V-SLAM 系統在不同光照條件下拍攝之影像序列之定位精度表現，並探討其誤差來源。
- (3) 進行不同光照條件之影像序列和三維特徵地圖的交叉定位測試，以定位精度差異分析相機定位對不同光照條件之三維特徵地圖的適應性。

2. 研究方法與流程

此章節分別會介紹本研究所選用之 SLAM 系統、影片資料收集、參考軌跡回復、坐標轉換模式和不同三維特徵地圖之相機定位適應性測試。

2.1 ORB-SLAM

本研究選擇 ORB-SLAM (Mur-Artal *et al.*, 2015、Mur-Artal and Tardos, 2017) 做為測試用 V-SLAM 系統。此系統透過對輸入系統之所有影像，萃取 ORB 特徵點 (Rublee *et al.*, 2011)，並對萃取之 ORB 特徵點進行空間前方交會，建立三維特徵點雲地圖，並透過將影像特徵匹配至三維特徵地圖的方式，進行空間後方交會、即時相機定位及建立場景的三維特徵地圖。

ORB-SLAM 對原始影像序列取樣，並稱取樣之影像為關鍵影像 (Keyframe)，系統設計上僅讓關鍵影像參與建立三維特徵地圖，維持系統計算效率。其輸出之相機軌跡亦為關鍵影像之定位軌跡，含獨立坐標系中的三維坐標及三軸旋轉角。本研究對 ORB-SLAM 程式進行修改，增加了三維特徵地圖的保存/載入功能，使其在 SLAM 處理過程中產生之三維特徵地圖可以被保存，並於 ORB-SLAM 的 Localizaiton 模式下再次讀入系統，做為場景的控制，進行僅定位相機而不建立場景三維特徵地圖的「三維特徵地圖輔助相機定位」。

2.2 影片資料收集

本研究目的為在室外環境中，測試三維特徵地圖輔助之相機定位對不同光照條件的適應性。考量實際載具移動過程，除三維特徵地圖輔助之相機定位的定位誤差外，尚有載具機身震動、路線規劃等外部因素可能對相機定位精度或 SLAM 成果造成影響。為更精確評估三維特徵地圖輔助相機定位之精度，本研究將不使用實際移動載具拍攝之資料進行測試，而以手持單眼相機的方式，穩定地步行環繞測試區，拍攝測試場景之影片資料，模擬載具移動和資料拍攝，免除外部因素對相機定位成果之影響。

對於「不同光照條件」，考量對於未來延伸至實際應用之狀況，本研究參考一般外業工作時段，分別選擇陽光傾斜入射之早上或下午時段、陽光垂直入射之中午時段、以及陽光均勻入射的陰天環境，拍攝不同光照條件影片資料，並用以建立地圖，彼此互為對照組進行後續實驗分析。

影片資料拍攝部分，鑑於現今之 V-SLAM 技術仍十分仰賴 Loop Closure 糾正整體尺度飄移 (Scale Drift) 以及定位軌跡，因此本研究在所有影片資料中，皆會確保影像序列軌跡之重疊，使 Loop Closure 能夠被執行。

2.3 相機參考軌跡回復

定位精度分析前須獲得相機之參考軌跡，方得和 ORB-SLAM 所獲得之定位成果進行比較。過往各種 SLAM 系統間之對比，皆是透過 KITTI (Geiger *et al.*, 2013) 等以高精度 GPS/INS 建立之公開資料集進行。然而現今之公開資料集中，並沒有符合本研究目的之「相同場景」，但「不同光照條件」所拍攝之資料，因此需自行收集之場景影片資料進行測試。而受限於儀器，本研究亦難以獲得其使用之高精度 GPS/INS 系統，需額外使用其他方法取得參考相機軌跡。

考量本研究僅有影像資料的收集，本研究採運動恢復結構 (Structure from Motion, SfM) 技術進行相機軌跡回復。和特徵點法 V-SLAM 之方法相比，SfM 方法同樣是將特徵匹配成果作為影像連結點，進行影像相對方位的解算，並重建場景之稀疏點雲。然而和 SfM 不同的是，SLAM 更著重定位的即時性，因此會為了處理效率而在計算上做妥協，而 SfM 則更追求計算結果的正確程度，對效率較不要求，可以在計算成果的精度與可靠度上進行更完整的評量。對比一般移動載具多採用網路即時動態定位技術 (Real-Time Kinematic, RTK) 的 GNSS 訊號做為導航的坐標來源，其定位精度為公分等級。本研究認為以攝影測量商用軟體，透過 SfM 方法回復，並進行地面控制點標定之相機軌跡，其公分級之精度足作為影像之參考軌跡 (Ground Truth Trajectory) 使用，因此選擇以 SfM 方法回復之相機軌跡為參考

軌跡使用。此比較成果亦可視為以 SfM 方法回復之相機軌跡，和以 SLAM 方法回復之相機軌跡的差異。

2.4 坐標轉換模式

本研究之資料為單眼相機拍攝之影片，因此輸出資料將位於缺乏尺度資訊的獨立坐標系中，又其輸出為稀疏點雲，無法精確標定控制點進行坐標轉換，使精度評估或實際輔助相機定位之應用難以進行。SLAM 領域中最常使用之坐標轉換方式為最小化兩軌跡套合誤差之 Umeyama’s Method (Umeyama, 1991)。除此之外，本研究亦提出一套坐標轉換模式，透過三維正形轉換和最小二乘平差，可以分別基於相機軌跡和自行在獨立坐標系中建立之控制點進行坐標轉換。

2.4.1 Umeyama’s Method

Umeyama’s Method 主要被用於處理兩個點集合的套合或定向的問題 (Umeyama, 1991)。此亦是現今許多 SLAM 系統在評估定位精度時，用於套合 SLAM 定位軌跡和參考軌跡，以進行後續精度評估之演算法。

Umeyama 在其提出之演算法，透過奇異值分解 (Singular Value Decomposition, SVD) 最小化目標函數式(1)，尋找能夠最佳套合兩個點集合，並使兩個點集合中對應各點的距離間均方誤差最小的轉換參數。式(1) 中共有n個點對，其中第i個待轉換點 x_i

會透過尺度S、旋轉R和平移T的調整，和參考點集合中對應的第i個點 y_i 進行套合，轉換後的坐標均方誤差則以 e^2 表示。

$$e^2(R, T, c) = \frac{1}{n} \sum_{i=1}^n \|y_i - (SRx_i + T)\|^2 \dots (1)$$

2.4.2 三維正形轉換

ORB-SLAM 輸出之三維特徵地圖為稀疏點雲，難以精確標定控制點進行坐標轉換，因此本研究提出於建立獨立坐標系控制點的方法：以 ORB-SLAM 輸出之相機外方位參數，配合手動影像量測，進行空間前方交會，求解控制點於獨立坐標系的三維坐標，並以最小二乘法，分別輸入控制點於獨立坐標系和物空間坐標系之坐標，求解坐標轉換參數，其概念示意如 圖 1 所示。

啟發自 Umeyama’s Method，相同最小二乘平差模型亦可以 ORB-SLAM 相機定位軌跡和其參考軌跡作輸入，最小二乘求解轉換參數，其概念示意圖如圖 2 所示。

三維正形轉換形式如式(2)所示，此公式表示 xyz 坐標系中的點 (x_0, y_0, z_0) ，經過旋轉矩陣 R 旋轉，尺度 S 調整，以及 (T_x, T_y, T_z) 平移後，會被轉換至 XYZ 坐標系中的坐標點 (X_0, Y_0, Z_0) 上。

$$\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} = S R \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \dots \dots \dots (2)$$

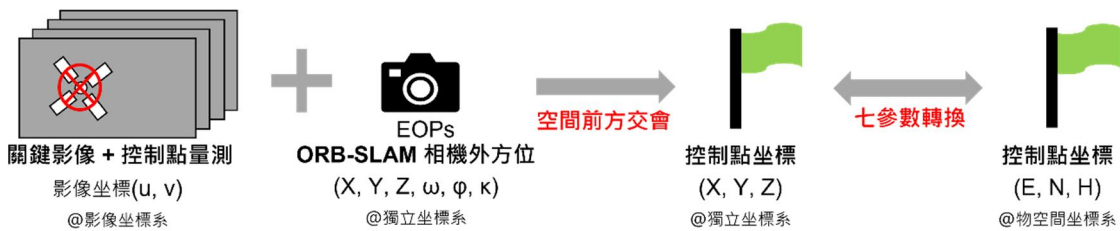


圖 1 基於控制點之坐標轉換



圖 2 基於軌跡之坐標轉換

在本研究中，位於物空間坐標系之參考軌跡被視為定值，在平差過程不應被修正，不需賦予改正數；而在坐標轉換模型中，待轉換的 ORB-SLAM 坐標則視為觀測，在其後方附加參數，以吸收觀測值所含之誤差，避免觀測誤差於函數模型內傳播，造成平差成果無法收斂，因此式(2) 之三維正形轉換模型，加入附加參數的考量之後修正如式(3)。

$$Ax + Bv = w \dots\dots\dots (3)$$

其中 A 為將三維正形轉換模型線性化後而得的未知數設計矩陣式(4)，其中 m 為待轉換點個數。 x 為未知數的修正數矩陣式(5)。 B 為線性化三維正形轉換模型中的坐標觀測而得之附加參數設計矩陣式(6)。 v 則為附加參數矩陣式(7)。 w 則為不符值矩陣式(8)，其中 $f(T_x, T_y, T_z, \omega, \varphi, \kappa, S)$ 為將求解之轉換參數帶入式(2) 計算得之值。

其中式(3) 為非線性函數，在進行最小二乘求解時需仰賴起始值得給予，本研究參考 Luhmann *et al.*(2013) 介紹的法向量方法，透過三個控制點建立之中間坐標系 uvw ，求取 XYZ 和 xyz 坐標系間的轉換參數近似值。

$$A_{3m \times 7} = \begin{bmatrix} \frac{\partial f_E}{\partial T_x} & \frac{\partial f_E}{\partial T_y} & \frac{\partial f_E}{\partial T_z} & \frac{\partial f_E}{\partial \omega} & \frac{\partial f_E}{\partial \varphi} & \frac{\partial f_E}{\partial \kappa} & \frac{\partial f_E}{\partial S} \\ \frac{\partial f_N}{\partial T_x} & \frac{\partial f_N}{\partial T_y} & \frac{\partial f_N}{\partial T_z} & \frac{\partial f_N}{\partial \omega} & \frac{\partial f_N}{\partial \varphi} & \frac{\partial f_N}{\partial \kappa} & \frac{\partial f_N}{\partial S} \\ \frac{\partial f_H}{\partial T_x} & \frac{\partial f_H}{\partial T_y} & \frac{\partial f_H}{\partial T_z} & \frac{\partial f_H}{\partial \omega} & \frac{\partial f_H}{\partial \varphi} & \frac{\partial f_H}{\partial \kappa} & \frac{\partial f_H}{\partial S} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \dots\dots\dots (4)$$

$$x_{7 \times 1} = \begin{bmatrix} \Delta T_x \\ \Delta T_y \\ \Delta T_z \\ \Delta \omega \\ \Delta \varphi \\ \Delta \kappa \\ \Delta S \end{bmatrix} \dots\dots\dots (5)$$

$$B_{3m \times 3m} = \begin{bmatrix} \frac{\partial f_E}{\partial x} & \frac{\partial f_E}{\partial y} & \frac{\partial f_E}{\partial z} & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \frac{\partial f_N}{\partial x} & \frac{\partial f_N}{\partial y} & \frac{\partial f_N}{\partial z} & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\partial f_H}{\partial x} & \frac{\partial f_H}{\partial y} & \frac{\partial f_H}{\partial z} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \dots\dots\dots (6)$$

$$v_{3m \times 1} = \begin{bmatrix} v_x \\ v_y \\ v_z \\ \vdots \end{bmatrix} \dots\dots\dots (7)$$

$$w_{3m \times 1} = \begin{bmatrix} E - f_E(T_x, T_y, T_z, \omega, \varphi, \kappa, S) \\ N - f_N(T_x, T_y, T_z, \omega, \varphi, \kappa, S) \\ H - f_H(T_x, T_y, T_z, \omega, \varphi, \kappa, S) \\ \vdots \end{bmatrix} \dots\dots\dots (8)$$

2.5 不同光照條件之三維特徵地圖定位測試

本研究收集不同光照條件下的場景影片作為測試資料。再將各光照條件下的場景影片，輸入 ORB-SLAM 進行 SLAM 處理，分別建立屬於該光照條件之三維特徵地圖，並同時回覆該場景影片之 ORB-SLAM 軌跡。

對於前述步驟中產生之不同光照條件之三維特徵地圖，本研究再透過 ORB-SLAM 之三維特徵地圖讀取功能，配合不同光照條件的影片輸入，在 ORB-SLAM 之 Localization 模式下，進行僅定位相機而不建圖之三維特徵地圖輔助相機定位。進行 Localization 模式處理後，其僅輸出三維特徵地圖輔助相機定位之軌跡。本研究亦會將此相機定位軌跡，和其相機定位之參考軌跡相比較，以相機定位精度分析其對光照條件之適應性。

3. 研究成果與分析

本章節將會介紹使用之相機，並對兩實驗區之相機定位成果，和定位成果對光照條件之適應性進行分析。

3.1 相機資訊和資料處理環境

本研究所使用之相機為日本富士相機公司推出之 X-T20 相機，為基於 CMOS 感光元件的 APS-C 相機。其亦提供錄影功能，可以支援解析度 1920 × 1080 像素的 Full HD 影片錄製，輸出之影片資料格式為.MOV。影片資料皆以 18 mm 焦距和自動曝光拍攝。

資料處理環境部分，本研究使用 Intel i7-7700 電腦進行資料處理，配備之記憶體為 32 GB，顯示卡為 NVIDIA GeForce GTX 1050 Ti。本研究透過 VirtualBox 軟體，於 Windows 10 作業系統中建立虛擬主機，另外安裝 Ubuntu 作業系統作為 ORB-

SLAM 之安裝及作業環境，其中 Ubuntu 作業系統之版本為 18.04.4。

3.2 成功大學歸仁校區測試區

本測試區位於成功大學歸仁校區西南角之率定場，為一 40 m × 60 m 大小的矩形場域，四周為建築物環繞，並有一車道貫穿，其俯視圖及近照如圖 3 所示。



圖 3 成功大學歸仁校區俯視圖(左)及近照(右)

3.2.1 影片收集及 ORB-SLAM 處理

本研究於 2021/03/04 前往成功大學歸仁校區率定場，以鏡頭外朝建築物方式，步行環繞測試區拍攝建築物影片資料。其影片拍攝資訊及 ORB-SLAM 處理成果紀錄如表 1 所示，影片內容圖樣則如圖 4、圖 5、圖 6 所示。

表 1 歸仁測試區之影片及 ORB-SLAM 處理紀錄

影片資料	早上	中午	陰天
影片長度(秒)	85	72	64
影片長度(幀)	2550	2160	1920
處理時長(秒)	1285	1097	857
平均處理時間(毫秒/幀)	440	439	387
閉環發生	是	是	是
建立地圖點數	15923	16901	16334
關鍵影像數	146	136	146



圖 4 成大歸仁校區早上影片



圖 5 成大歸仁校區中午影片



圖 6 成大歸仁校區陰天影片

3.2.2 參考軌跡回復

將前述步驟獲得之關鍵影像序列，輸入 Agisoft Metashape 軟體進行 SfM 處理，回復影像相對方位，並於影像上進行控制點量測，將影像軌跡回復成果轉換至絕對坐標系。三組資料之回復成果如表 2 所示，其成果顯示三組資料回復之影像軌跡，其整體的反投影誤差約為 0.7 個像元，控制點、檢核點誤差則約落在 5 cm 左右。

表 2 歸仁測試區影片之 SfM 軌跡回復精度

影片	反投影誤差 (Pixel)	控制點誤差 (m)	檢核點誤差 (m)
早上	0.734	0.029	0.040
中午	0.697	0.049	0.048
陰天	0.659	0.054	0.055

3.2.3 坐標轉換及定位精度分析

由於本研究使用單眼相機拍攝影片資料，因此 3.2.1 節中輸出之 ORB-SLAM 相機定位軌跡，位於缺乏尺度資訊之獨立坐標系中，需再行坐標轉換，將轉換獨立坐標系中的相機定位軌跡至絕對坐標系，方得進行定位精度分析及後續應用。

又 ORB-SLAM 中，影像的定位來自於對於場景點雲的匹配，而場景點雲的產生，亦是來自於影像的前方交會成果。鑒於兩者相應而生，因此對於 ORB-SLAM 而言，影像定位軌跡和場景之三維特徵地圖，不應被視為兩個獨立物件看待，而應是兩不可分割的存在，因此對於相機定位精度的評估，亦同時可視為環境建圖品質的評估。經 2.4.2 節中所述之方法進行坐標轉換後，之相機定位軌跡如表 3 所示，相機定位精度則以敘述性統計的方式描述，其中 RMSE 為均方根誤差、Med 為中位數、Max 為最大值。

表 3 成大歸仁校區率定場測試區：各影片資料坐標轉換成果 (單位：m)

影片資料	坐標轉換	RMSE	Med	Max
早上	Umeyama	0.177	0.150	0.306
	基於軌跡	0.177	0.150	0.306
	基於控制點	0.294	0.247	0.455
中午	Umeyama	0.088	0.067	0.481
	基於軌跡	0.088	0.067	0.481
	基於控制點	0.303	0.292	0.668
陰天	Umeyama	0.177	0.150	0.306
	基於軌跡	0.177	0.150	0.306
	基於控制點	0.254	0.238	0.390

自表 3 中的成果可以看到，三組影像資料定位精度之 RMSE 皆約莫位於 10-20 cm 的範圍。其中亦可看到在各組資料或精度指標中，Umeyama's Method 和基於軌跡的最小二乘平差的精度表現皆相同，而基於控制點的最小二乘平差所獲得的精度則較差於另外兩者。

以中午之影片資料為例，將坐標轉換後之 ORB-SLAM 點雲和 3.2.2 節中產生之參考相機軌跡及參考場景點雲進行套合，其成果如圖 7、圖 8 所示。其中由於以 Umeyama's Method 和基於軌跡進行之最小二乘平差坐標轉換，獲得之坐標轉換成果精度相同，因此在此套合成果中視為相同資料，以基於軌跡之坐標轉換表示之。

自圖 7 至圖 8 中之成果可以看到，本測試資料中，基於軌跡進行之坐標轉換，其產生之場景點雲和參考點雲能夠較好的套合；做為對比，基於控制點進行之坐標轉換，其產生之場景點雲和參考場景

點雲相比，則明顯有局部偏移的狀況發生。歸納各測試資料之成果，此現象較常發生於相機拍攝到場景中距離較遠的部分時，如測試區右側和測試區左上角，如圖 9 所示，其對應的部分為兩筆直車道，被攝物和相機的距離較遠，且所拍攝之場景較開闊而缺乏特徵，產生較弱之拍攝幾何以及不穩定之特徵匹配成果，進而影響 ORB-SLAM 對相機外方位參數的回復，因此在以相機外方位參數，透過前方交會方法建立獨立坐標系中之控制點時產生了較大的誤差。

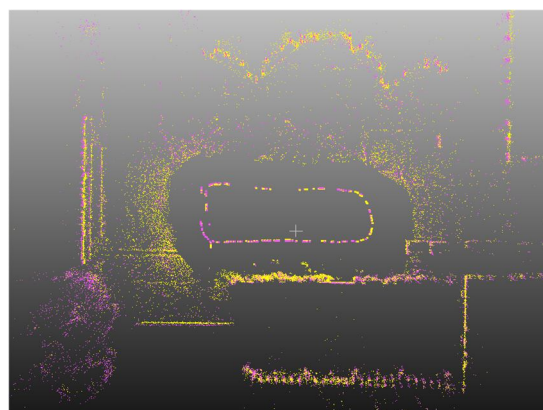


圖 7 以軌跡坐標轉換之點雲(紅)和參考點雲(黃)

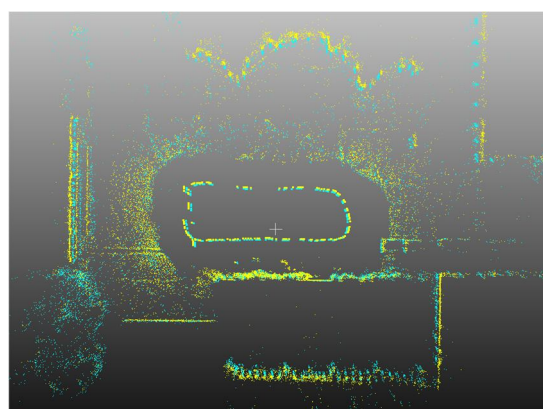


圖 8 以控制點坐標轉換之點雲(紅)和參考點雲(黃)



圖 9 歸仁測試區之車道

3.2.4 交叉定位測試

本節中將透過 ORB-SLAM 之 Localization 模

式，輸入前一節中建立之場景三維特徵地圖至 ORB-SLAM 做為場景控制，分別對在本場景拍攝之三組不同光照條件影片資料，進行只跟蹤而不建圖的純定位模式，檢驗 ORB-SLAM 建立之三維特徵地圖之做為場景控制的輔助定位能力，並基於相機定位精度分析對不同光照條件的適應性。

交叉定位測試成果如表 4 所示，其說明了不同的影片資料，透過以不同方式進行坐標轉換之地圖，進行 ORB-SLAM Localization 模式處理的成果，並以敘述性統計呈現。

自表 4 中可以看到，以控制點進行坐標轉換之三維特徵地圖，其所獲得之定位精度明顯差於以定位軌跡進行坐標轉換者，且其均方根誤差亦不會小於其所參考之三維特徵地圖之精度。然在表 4 中亦可看到，以不同光照條件下建立之三維特徵地圖輔助之相機定位之成果中，其定位精度未因光照條件改變而有明顯差異，反而和其所參考之三維特徵地圖精度關聯性較高。其所參考之三維特徵地圖精度較佳者，往往在定位精度上有較佳表現；相反的，所參考之三維特徵地圖精度較差者定位精度則較差。

表 4 交叉定位測試成果表 (單位：m)

坐標轉換	影片資料	參考地圖	地圖精度	RMSE	Med	Max	
基於控制點	早上	早上	0.247	<u>0.295</u>	0.255	0.456	
		中午	<u>0.292</u>	0.293	<u>0.279</u>	0.394	
		陰天	0.238	0.267	0.243	0.398	
	中午	早上	0.247	<u>0.311</u>	0.271	0.501	
		中午	<u>0.292</u>	0.301	<u>0.290</u>	<u>0.668</u>	
		陰天	0.238	0.276	0.245	0.431	
	陰天	早上	0.247	0.277	0.239	<u>0.454</u>	
		中午	<u>0.292</u>	<u>0.305</u>	<u>0.280</u>	0.396	
		陰天	0.238	0.254	0.238	0.391	
	基於軌跡	早上	早上	<u>0.150</u>	<u>0.178</u>	<u>0.151</u>	<u>0.303</u>
			中午	0.067	0.116	0.097	0.253
			陰天	0.121	0.135	0.116	0.248
中午		早上	<u>0.150</u>	<u>0.161</u>	<u>0.162</u>	0.253	
		中午	0.067	0.088	0.065	<u>0.474</u>	
		陰天	0.121	0.122	0.109	0.211	
陰天	早上	<u>0.150</u>	<u>0.168</u>	<u>0.163</u>	<u>0.243</u>		
	中午	0.067	0.110	0.096	0.187		
		陰天	0.121	0.124	0.120	0.179	

註：粗體為以不同地圖定位該影片資料獲得之誤差中表現最佳者，底線為表現最差者

中午影片資料之相機定位軌跡如圖 10、圖 11 所示。和 SLAM 處理過程中獲得之離散的關鍵影像軌跡不同，此模式進行的是逐影像的後方交會處理，因此其輸出的相機軌跡為密集的序列。歸納定位成果可以看到，縱使是透過不同影片和坐標轉換方式建立的地圖，ORB-SLAM 仍可以藉輸入之三維特徵地圖輔助，回復待定位影片之相機軌跡。此一以三維特徵地圖輔助之相機定位，能夠獲得接近於該影片在 SLAM 模式回復之軌跡的軌跡，其中以影片和該地圖之光照條件相符者之成果最為接近(例如以中午地圖定位中午影片)，然而此不等於和其參考軌跡的接近程度，仍決定於該 SLAM 成果之精度。且 SLAM 成果所含之瑕疵，亦會透過地圖傳播至定位軌跡。如此測試區中，中午地圖在左上角有較大的控制點誤差，而在圖 10、圖 11，此一誤差也被反應在該部分之局部定位軌跡，和該處之參考軌跡產生較大偏離。

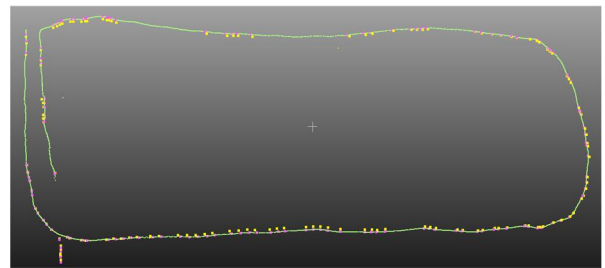


圖 10 基於軌跡坐標轉換之定位軌跡

註：黃：SfM 軌跡，紅：基於軌跡轉換之 SLAM 軌跡，綠：Localization 模式軌跡

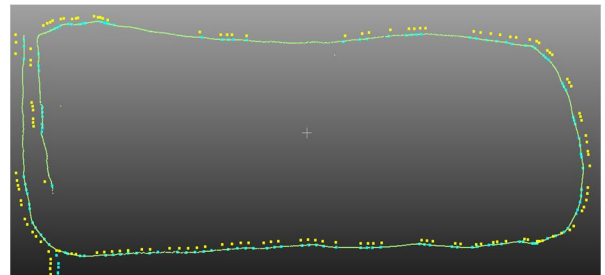


圖 11 基於控制點進行坐標轉換之定位軌跡

註：黃：SfM 軌跡，淺藍：基於控制點轉換之 SLAM 軌跡，綠：Localization 模式軌跡

在資料處理時長方面，由於本節所進行之測試是在 ORB-SLAM 之 Localization 模式下進行，僅進行影像特徵萃取、匹配和空間後方交會，未有費時的三維特徵地圖建立的過程，因此各組資料在一幀

影像的平均處理時長上，僅需約 270 至 320 毫秒，和 SLAM 模式下的 400 至 440 毫秒之平均處理時長相比，約節省 30% 的時間。

3.3 成大測量系副館測試區

本測試場位於台南市東區的國立成功大學成功校區中央，數學系館之北側。此測試場為一 30 m × 10 m 大小的矩形建築物，其俯視圖及近照如圖 12 所示。



圖 12 成大測量系副館俯視圖(左)及近照(右)



圖 14 成大測量系副館下午影片



圖 15 成大測量系副館陰天影片

3.3.1 影片收集及 ORB-SLAM 處理

本研究於 2021/09/06 和 2021/09/07 前往成功大學測量系副館測試區，以鏡頭內朝建築物方式，步行環繞測試區拍攝建築物影片資料。其影片拍攝資訊如表 5 所示，影片內容圖樣則如圖 13、圖 14、圖 15 所示。

表 5 測量系副館之影片及 ORB-SLAM 處理紀錄

影片資料	中午	下午	陰天
影片長度(秒)	151	152	141
影片長度(幀)	4530	4560	4230
處理時長(秒)	1889	1831	1711
平均處理時間 (毫秒/幀)	344	356	344
閉環發生	是	是	是
建立地圖點數	36490	35333	33491
關鍵影像數	353	344	346



圖 13 成大測量系副館中午影片

3.3.2 參考軌跡回復

將前述步驟獲得之關鍵影像序列，輸入 Agisoft Metashape 軟體進行 SfM 處理，回復影像相對方位，並於影像上進行控制點量測，將影像軌跡回復成果轉換至絕對坐標系。三組資料之回復成果如表 6 所示，其成果顯示三組資料回復之影像軌跡，其整體的反投影誤差約為 1 個像元，控制點、檢核點誤差則約落在 1.5 cm 左右。

表 6 測量系副館影片之 SfM 軌跡回復精度

影片	反投影誤差 (Pixel)	控制點誤差 (m)	檢核點誤差 (m)
中午	1.089	0.014	0.015
	0.857	0.146	0.054
下午	1.085	0.011	0.015
	0.866	0.129	0.050
陰天	1.149	0.012	0.020
	0.919	0.115	0.026

3.3.3 坐標轉換及定位精度分析

本測試區之資料中，三組影片之 ORB-SLAM 定位精度分析以述性統計資料則整理於表 7。

自表 7 中的成果可以看到，三組影像資料定位精度之均方根誤差皆約位於 5 cm 內的範圍。其中

不論在何組資料或精度指標中，Umeyama's Method 和基於軌跡的最小二乘平差之坐標轉換精度表現皆相同，而基於控制點的最小二乘平差，雖然所獲得的精度較差，但和基於軌跡的方式相比之成果仍屬接近，均方根誤差僅有 1 至 2 cm 的差距。

表 7 成大測量系副館測試區：各影片資料坐標轉換成果(單位：cm)

影片資料	坐標轉換	RMSE	Med	Max
中午	Umeyama	2.6	2.2	5.6
	基於軌跡	2.6	2.2	5.6
	基於控制點	4.2	4.1	8
下午	Umeyama	3.1	3	5.8
	基於軌跡	3.1	3	5.8
	基於控制點	4.7	4.1	8.4
陰天	Umeyama	2.2	2	6.2
	基於軌跡	2.2	2	6.2
	基於控制點	4.3	3.6	8.2

以中午之影片資料為例，對 3.3.1 節輸出之 ORB-SLAM 相機軌跡及產生之場景點雲進行坐標轉換，並和 3.3.2 節中產生之參考相機軌跡及場景點雲進行套合，進行視覺比較，其成果如圖 16、圖 17 所示。其中由於以 Umeyama's Method 和基於軌跡進行之最小二乘平差坐標轉換，獲得之坐標轉換成果精度相同，因此在圖 16、圖 17 中，皆視為相同資料，以基於軌跡之坐標轉換表示之。

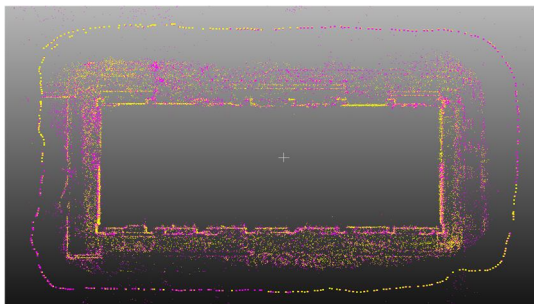


圖 16 以軌跡坐標轉換之點雲(紅)和參考點雲(黃)

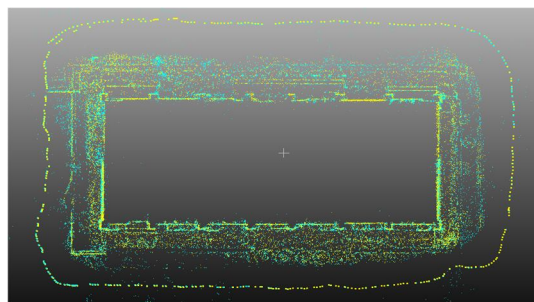


圖 17 以控制點坐標轉換之點雲(紅)和參考點雲(黃)

自圖 16、圖 17 中之成果可以看到，由於拍攝過程中，相機和場景間的拍攝幾何較為穩定，無明顯尺度變化，因此在本測試資料中，並沒有看到如成大歸仁校區測試區所見之尺度誤差。不論以何種坐標轉換方式進行之何種影片，其所獲得之相機定位軌跡和三維特徵地圖，在經過坐標轉換後，皆和黃色的 SfM 軌跡和點雲十分套合，自圖面上之視覺比較，難以分辨明顯差異。

3.3.4 交叉定位測試

交叉定位測試成果如表 8 所示，其說明不同三維特徵地圖輔助之相機定位成果，其定位精度以敘述性統計呈現於表 8 中。

表 8 交叉定位測試成果表(單位：m)

坐標轉換	影片資料	參考地圖	地圖精度	RMSE	Med	Max	
基於控制點	中午	中午	<u>4.1</u>	4.2	4.1	7.9	
		下午	<u>4.1</u>	<u>5.3</u>	<u>5.2</u>	13	
		陰天	3.6	4.2	4	11	
	下午	中午	<u>4.1</u>	<u>5</u>	<u>5</u>	10.3	
		下午	<u>4.1</u>	4.6	4	8.4	
		陰天	3.6	4.1	3.8	8.3	
	陰天	中午	<u>4.1</u>	<u>5.5</u>	<u>5</u>	10.1	
		下午	<u>4.1</u>	5.4	4.4	9.3	
		陰天	3.6	4.3	3.6	8.1	
	基於軌跡	中午	中午	2.2	2.6	2.2	5.6
			下午	<u>3.0</u>	<u>4</u>	<u>3.7</u>	11.8
			陰天	2.0	3.4	3	8
下午		中午	2.2	<u>4.3</u>	<u>4.1</u>	8.9	
		下午	<u>3.0</u>	3.1	3	5.3	
		陰天	2.0	3.6	3	11.4	
陰天		中午	2.2	<u>4</u>	<u>3.7</u>	7.7	
		下午	<u>3.0</u>	3.2	2.4	6.7	
		陰天	2.0	2.1	0.8	5.9	

註：粗體為以不同地圖定位該影片資料獲得之誤差中表現最佳者，底線為表現最差者

自表 8 中可以看到，和成大歸仁校區之測試資料相比，雖然以基於軌跡之坐標轉換，其所獲得之精度仍略優於基於控制點者，然兩種坐標轉換模式所獲得精度差異，和成大歸仁校區之成果相比較不明顯。在表 8 中亦可看到，以不同光照條件下建立之三維特徵地圖輔助之相機定位，其獲得之定位精度

未因光照條件改變而有明顯差異，反而和其所參考之三維特徵地圖精度之關聯性較高。其所參考之三維特徵地圖精度較佳者，常在均方根誤差上有較佳表現；相反地，所參考之三維特徵地圖精度較差者，其往往有較差的表現。

中午影片資料之三維特徵地圖輔助相機定位軌跡如圖 18、圖 19 所示。圖中可以看到，ORB-SLAM 可藉由輸入系統內之三維特徵地圖的輔助，回復待定位影片之相機軌跡。且其所獲得之相機軌跡和其 SfM 軌跡或其 SLAM 軌跡十分接近，無肉眼可明顯辨別之定位軌跡差異。

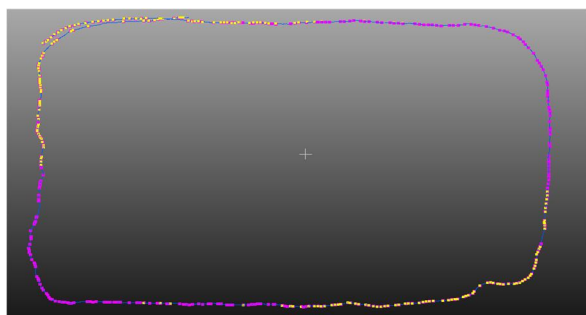


圖 18 基於軌跡坐標轉換之定位軌跡

註：黃：SfM 軌跡，紅：基於軌跡轉換之 SLAM 軌跡，綠：Localization 模式軌跡

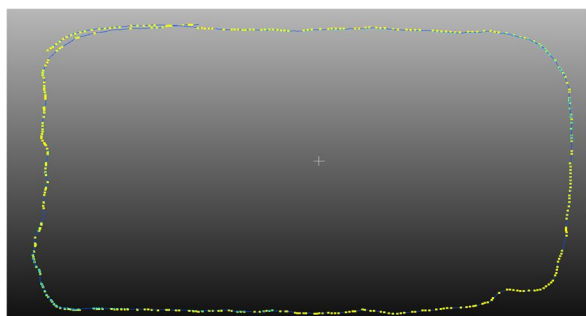


圖 19 基於控制點進行坐標轉換之定位軌跡

註：黃：SfM 軌跡，淺藍：基於控制點轉換之 SLAM 軌跡，綠：Localization 模式軌跡

在資料處理時長方面，由於本節所進行之測試是在 ORB-SLAM 之 Localization 模式下進行，僅進行影像特徵萃取、匹配和空間後方交會，未有費時的三維特徵地圖建立的過程，因此各組資料在一幀影像的平均處理時長上，僅需約 230 至 260 毫秒，和 SLAM 模式下的 400 至 440 毫秒之平均處理時長相比，約節省 30% 的時間。

4. 結論

自本研究之實驗成果可以看到，ORB-SLAM 的相機定位精度，和其光照條件並無明顯關聯，主要受其坐標轉換精度和場景之拍攝幾何影響。

在成大歸仁校區的案例中，場景深度變化較大且拍攝距離較遠，產生了較弱的拍攝幾何，因此該測試區部分的局部建圖成果較不可靠，也連帶地影響該處的坐標轉換和相機定位精度。從基於控制點的坐標轉換成果中亦可以發現，局部建圖成果不佳的部分，其前方交會獲得的控制點亦較不可靠，使得在歸仁校區案例中，基於控制點進行的坐標轉換，往往含有局部的尺度誤差。

在成大測量系副館的案例中，由於場景拍攝距離短且拍攝幾何較穩定，因此各種坐標轉換模式之成果均未見明顯差別，能獲得穩定的相機定位成果，定位精度皆為公分級。

三維特徵地圖輔助之相機定位部分，本研究之實驗成果顯示，其定位精度和光照條件亦無明顯關聯，主要影響之條件為其所參考之三維特徵地圖的精度，倘若其所參考之三維特徵地圖精度越高，則以三維特徵地圖輔助之相機定位則越能有可靠的表現。

本研究所採用之坐標轉換模式部分，Umeyama's Method 和基於軌跡的最小二乘平差方法，兩者總是能獲得相同的坐標轉換成果，顯示兩者在數學上應是透過不同方法，進行了相同意義的處理，惟仍需後續研究驗證。本研究所提出之基於控制點的最小二乘方法，由於其建立過程隱含了 ORB-SLAM 相機外方位參數回復的誤差，因此其在兩測試區之案例中，其精度表現皆遜於基於軌跡進行的坐標轉換。在成大歸仁校區的案例中，受限於場景拍攝，其精度明顯遜於基於軌跡進行之坐標轉換；然而在成大測量系副館的案例中，其成果顯示基於控制點進行的坐標轉換，其能夠獲得和基於軌跡的坐標轉換接近之成果。考量在實際即時定位的應用目的中，可靠的相機參考軌跡難以獲取，且會使 SLAM 應用的必要性降低，本研究認為在妥善的拍攝條件配置下，基於控制點的坐標轉換不失為在

實際 SLAM 應用中獲取尺度資訊的手段之一。

本研究於成功大學測量系副館測試區之成果表明，以距被攝物 2 m 距離拍攝之影像序列建立三維特徵地圖，並藉其輔助，定位相同距離拍攝之影像，透過本研究之坐標轉換模式，能獲得 5 cm 左右的絕對定位精度。據此，對於無 GNSS 訊號環境定位問題，根據所獲得之定位精度，本研究認為在妥善場景、拍攝條件配置下，本研究測試的三維特徵地圖輔助相機定位，有望成為一在無 GNSS 訊號環境之中定位問題的解決方案，且環境光照條件改變並不會對三維特徵地圖輔助之相機定位成果產生明顯影響。

5. 未來展望

針對載具於無 GNSS 訊號時的定位導航問題，本研究提出可能的解決方案，也就是預先針對目標環境四周製作三維特徵地圖，並進行坐標轉換到物空間坐標系統。接著，在實際以移動載具拍攝高解析影像時，以 ORB-SLAM 的 Localization 模式進行即時定位導航。因此本研究之主要目的為針對不同光照條件下拍攝相同場景之即時相機定位，進行了適應性與精度分析。然而實際應用在移動載具，做為無 GNSS 訊號環境之輔助定位手段前，本研究仍有部分限制，尚待未來研究中突破。

首先，ORB-SLAM 為基於影像特徵點的 SLAM 系統，需在富有紋理的環境才能運行。且本研究所拍攝之測試資料，皆在影片首尾相連，透過閉環進行全域最佳化的條件下進行，但在實際應用案例中，閉環的條件會造成拍攝航線規劃上的限制，因此未來系統設計上，需再另尋閉環以外的方案，解決尺度、軌跡飄移的問題。

其二，本研究在現有之軟硬體配置和參數設定下，仍需十餘倍於影片資料長度的資料處理時間，尚無法達成即時定位的目標，仍須仰賴未來研究中，針對不同定位目的需求，進行影像大小或影像取樣頻率的調整，取得資料處理效率和定位需求間的平衡，以在未來成為移動載具在無 GNSS 訊號環境下定位問題的解決方案。

最後，本研究僅透過手持相機方式，對低速之載具移動進行模擬，測試其定位能力。實際載具移動過程中，機械運轉造成之高頻振動，或是強風吹襲等氣流擾動對定位成果之影響，仍待後續研究以實際之移動載具資料進行測試。

致謝

本研究承蒙科技部專題研究計畫編號 MOST110-2121-M006-003 經費補助，得以順利完成，謹此致謝。

參考文獻

- Geiger, A., Lenz, P., Stiller, C., and Urtasun, R., 2013. Vision meets robotics: The KITTI dataset, *The International Journal of Robotics Research*, 32: 1231-1237.
- Luhmann, T., Robson, S., Kyle, S., and Boehm, J., 2013. *Close-Range Photogrammetry and 3D Imaging*, Berlin, Germany: De Gruyter.
- Mur-Artal, R., Montiel, J.M.M., and Tardos, J.D., 2015. ORB-SLAM: A versatile and accurate monocular SLAM system, *IEEE Transactions on Robotics*, 31: 1147-1163.
- Mur-Artal, R., and Tardos, J.D., 2017. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras, *IEEE Transactions on Robotics*, 33: 1255-1262.
- Nobis, F., Papanikolaou, O., Betz, J., and Lienkamp, M., 2020. Persistent map saving for visual localization for autonomous vehicles: An ORB-SLAM 2 extension, *Proceedings of the 2020 Fifteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*, Monte-Carlo, Monaco, pp. 1-9.
- Ranganathan, A., Matsumoto, S., and Ilstrup, D., 2013. Towards illumination invariance for visual localization, *Proceedings of the 2013 IEEE International Conference on Robotics and*

- Automation, Karlsruhe, Germany, pp. 3791-3798.
- Rublee, E., Rabaud, V., Konolige, K., and Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF, Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, pp. 2564-2571.
- Taheri, H., and Xia, Z.C., 2021. SLAM; definition and evolution, Engineering Applications of Artificial Intelligence, 97: 104032.
- Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns, IEEE Transactions on Pattern Analysis & Machine Intelligence, 13: 376-380.
- Wang, R., Zeng, L., Wu, S., Cao, W., and Wong, K., 2020. Illumination-invariant feature point detection based on neighborhood information, Sensors, 20(22): 6630.

Feasibility and Accuracy Analysis of Camera Localization Aided by a 3D Feature Map Established by Monocular Visual SLAM

Wei-Cheng Lin ^{1*} Jiann-Yeou Rau ²

Abstract

The feature-based Visual Simultaneous Localization and Mapping (V-SLAM) systems localize cameras by repeatedly establishing the 3D feature maps, matching the image features to the 3D feature maps and conducting space resection. By reusing the 3D feature maps established during SLAM processing, it can be considered as a control field to solve the GNSS-denied localization problem. However, the different lighting condition in the outdoor environments can possibly affect the results of image feature extraction and matching, which will lead to different localization results. This study applies ORB-SLAM to establish the 3D feature maps, and the map-aided camera localization is conducted under ORB-SLAM Localization Mode. The experimental results demonstrate the camera localization aided by a 3D feature map can reach close camera localization accuracy as the accuracy of its' reference 3D feature maps. As a result, the camera localization aided by a 3D feature map can become a potential solution of GNSS-denied localization problem.

Keywords: ORB-SLAM, 3D Feature Map, Coordinate Transformation

¹ Master, Department of Geomatics, National Cheng Kung University

² Professor, Department of Geomatics, National Cheng Kung University

* Corresponding Author, E-mail: p66094133@gs.ncku.edu.tw

Received Date: Jan. 21, 2022

Revised Date: Feb. 09, 2022

Accepted Date: Mar. 07, 2022