# 語意閉環檢測

# Semantic Loop Closure in Simultaneous Localization and Mapping Systems

楊哲宇、張育晟、陳毓琇、黃志煒
Che-Yu Yang, Yu-Cheng Zhang, Yu-Shiu Chen, Chih-Wei Huang

同時定位與地圖建置 (simultaneous localization and mapping, SLAM) 演算法，是利用移動中的光學感測器取得 3D 環境資訊的核心技術之一，廣泛應用在自動駕駛車、自走車 (automated guided vehicles, AGV)、和家用機器等。有時候，SLAM 演算法的效益很大程度上依賴於光學設備，像是相機、LIDAR 等，取決於輸入的信號品質。而在 SLAM 演算法中的其中一個步驟為閉環檢測，這個步驟為的是要檢測機器人是否到達曾經到過的位置，進而消除在建立地圖時所累積的誤差，成為 SLAM 中的重要步驟。一般藉由幾何特徵描述來判斷照片裡的場景是否為相似，但是有時候在接近相同的場景時，幾何特徵的方法仍然沒有效。因此，我們可以將物件辨識以及時間－空間序列這兩個特徵去做比較，並把它們整合到 SLAM 的流程中，可以呈現更準確的 3D 空間資訊。在這篇文章當中，首先我們會提供 SLAM 的概述，再來是介紹物件辨識和時間－空間序列的比較方法，這兩個方法為的是在 SLAM 中更可以比較照片中所出現的場景是否相似，經由辨識物件像是地標或標誌甚至是物件，我們可以更好的分類相似的場景，並且改善室內 3D 建圖的結果。

Simultaneous localization and mapping (SLAM) algorithm is one of the core technologies to build 3D environment data using optical sensors. The technology is widely utilized on machines such as automated guided vehicles (AGV) and domestic robots. The performance of SLAM algorithms is highly depending on both the software and the quality of optical sensors, like cameras, LIDAR, etc. Loop closure, one of the crucial SLAM function component to highlight, is responsible for detecting visited locations and correcting accumulated errors. Conventionally, loop Closure calculates the similarity of scenes by comparing geometric features, but in scenarios where different scenes appears nearly identical, the performance of feature-based methods degrade significantly. Therefore, instead of using geometric feature, we introduce object recognition combining with time-spatial sequences to evaluate the similarity and improve the SLAM process. In this article, we first overview the SLAM process, and then give the introduction with object recognition and time-spatial comparing schemes. By identifying landmarks objects or signs, we can better classify similar scenes and improve 3D indoor mapping results.

## 一、背景介紹

同時定位與地圖建置 (simultaneous localization and mapping, SLAM) 方法，是建立 3D 環境圖資以及機器人導航技術中的一項重要議題，它的精隨在於可以同時對相機所看到的環境產生 2D 或 3D 地圖，並且可以估計出相機和機器人位於地圖中的位置。在執行 SLAM時，機器人可以在環境中導航至目標位置，但是機器人必須知道兩件事情：(1) 機器人本身在哪裡，也就是定位問題，(2) 瞭解周圍環境的地圖建置問題。這意味著機器人要同時確定自己的位置和對周圍環境的了解，例如知道哪裡有障礙物。整個 SLAM 問題可將其分為前端和後端兩部份所組成，如圖 1 所示的視覺 SLAM 架構，包含：

1. 特徵提取 (feature extraction)：在視覺 SLAM 中，主要透過照片來提取特徵。
2. 視覺里程計(visual odometry, VO)：估計連續照片之間的關聯性，例如：追蹤特徵點的位移或旋轉。
3. 優化 (optimization)：結合視覺里程計與閉環檢測這兩個步驟來進行優化。
4. 閉環檢測 (loop closure)：用來判斷機器人是否到達曾經到過的地點，如果閉環發生，資訊將被納入後端處理。
5. 地圖建置：根據優化過的影像及軌跡資訊建置環境，依據需要可為 3D 或 2D。

常用於 SLAM 的光學感測器可分為三種類型：(1) 單目相機 (monocular)[1, 2]、(2) 立體相機 (stereo camera)、和 (3) 彩色深度相機 (RGB-D)[3]。通常單目相機因為只有一個鏡頭，需要額外使用幾何的方法進行處理，來得到 SLAM 所需的深度訊息，透過相機的移動資訊，估計連續影像之間的差異，即可估計在場景裡面代理人距離物體的遠近和大小。比起其他種類的相機，更加簡單和實用。立體相機使用兩個鏡頭進行 SLAM 處理，可根據兩個鏡頭之間的基線估計每個像素的空間位置，我們可以假想成人的眼睛，但仍需要很大量的運算來獲得深度資訊。彩色深度相機經由結構化紅外光或測量飛行時間來收集相機與場景之間的距離，仍然存在深度準確率和可靠性的問題。

在這篇文章中，提供了以下的技術重點：

1. 把時間－空間影像序列納入閉環檢測步驟，有了位置訊息可以讓閉環檢測更準確。
2. 利用語意物件辨識的輔助來增強 SLAM 中的照片相似性的判斷。
3. 提出我們的機率評分模型，結合物件辨識和詞袋模型方法來檢測是否發生閉環 (Loop closure)。

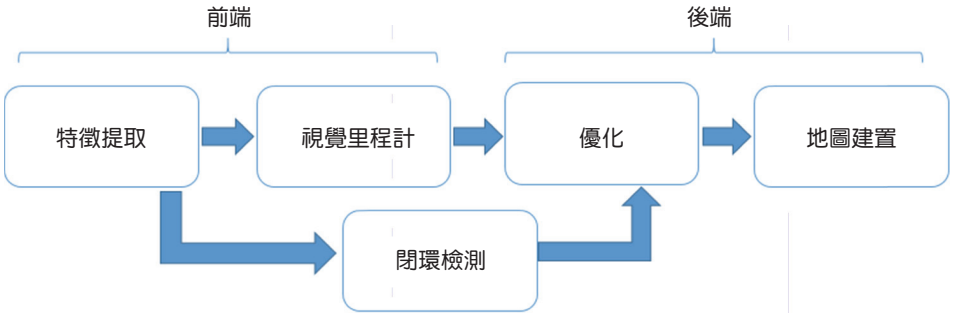


圖 1. 視覺 SLAM 架構。

我們所提出的語意閉環檢測[4] 可以更好的分類相似的場景，進而改善 SLAM 中的建置地圖表現，接下來的第二節我們會介紹何謂閉環檢測和閉環檢測的方法等等，第三節將會講解結合了單張照片比較，透過現有方法 (1) 單張比較方法用詞袋模型 (bag of word)，(2) 物件辨識和 (3) 多張照片時間－空間序列的比較方法來實現閉環檢測，第四節則是實驗設置以及我們所提出方法的結果，第五節說明該研究未來的發展。

## 二、閉環檢測

### 1. 閉環檢測的目標

閉環檢測是為了解決以下問題：機器人執行 SLAM 時，誤差會隨著時間累積，而使得地圖產生漂移，我們可以把 SLAM 看做是一個個優化問題，所以需要透過新的約束條件來修正誤差，從圖 2(a) 可知，來自視覺測距的約束點為相鄰的節點，像是 $X_1-X_2$，$X_2-X_3$，$X_3-X_4$，$X_5-X_6$。如果當 $X_4$ 發生漂移，$X_4$ 會受到影響，相對的 $X_5$，$X_6$ 也會跟著發生錯誤，我們可以把它想成是六張照片，所以這些錯誤累積了一段時間後，優化結果會變得不準確，由於目前的約束都是由上一張照片所建立的，但是發現 $X_6$ 不一定只能靠著 $X_5$ 估計，而是可以靠著第二幀 $X_2$ 和第三幀 $X_3$ 來估計位置，如圖 2(b)，這樣加入新的約束的方式就叫做閉環檢測，並且可以減少誤差的累積。

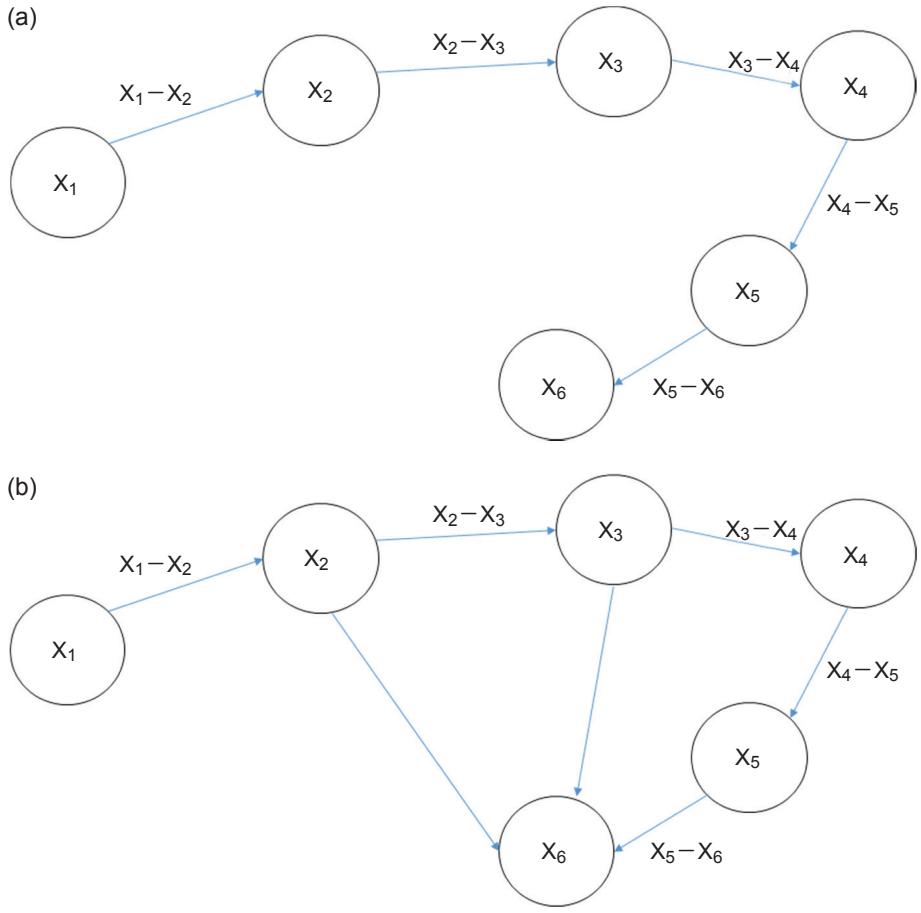


圖 2. (a) 未做閉環檢測之前。(b) 執行閉環檢測。

閉環檢測的定義，就是識別機器人是否有看到自己曾經到過的地方，有很多方法能夠幫助機器人識別場景，像是加入一些明顯的地標，但是這個方法有時會受到環境限制，比如說某個環境中各個地方都很相似，沒有明顯地標。但是通常機器人在執行閉環檢測時只能依靠感測器，比如透過兩張影像的相似度，來決定是否發生閉環，所以常見閉環檢測的兩個問題如圖 3(a)、(b)。

## 2. 閉環檢測的基本方法

最簡單的閉環檢測，是對任兩張圖的特徵點做擷取，在透過特徵點匹配檢查兩張圖片之間的相關性，這種方法是圖片匹配的最簡單的方法，任兩張圖片都會被檢測到，但是計算成本太高，如果總共有 $N$ 張圖片，就必須檢查 $C_2^n$ 次是否彼此相似，複雜度是 $O(N^2)$，所以在 SLAM 裡不是很實用。另一種做法，是隨機搜尋兩張照片，然後比較它們之間的特徵點，這種方法可以降低圖片的比對次數，但是隨著資料庫(地圖)越來越大，可以檢測到閉環的機率很低，所以我們希望依據哪裡較有可能發生閉環的照片再去做搜尋的動作，而非盲目地搜尋是否發生閉環 (loop closure)。

閉環檢測的方法有兩類，一種是依據里程計 (odometry)。里程計的方法是假設已知機器人的確切位置，像是相機的座標位置，如果相機的座標位置接近一樣，我們就可以當作相同的地方，可是通常相機的位置不會很準確，會有偏差[5, 6]。

另一種是根據感測器所拍的照片來比較之間的相似性 (appearance-based)，並以此判斷是否發生閉環，此方法的中心思想，是比較兩張照片的相似性，像是特徵點等等，並且希望

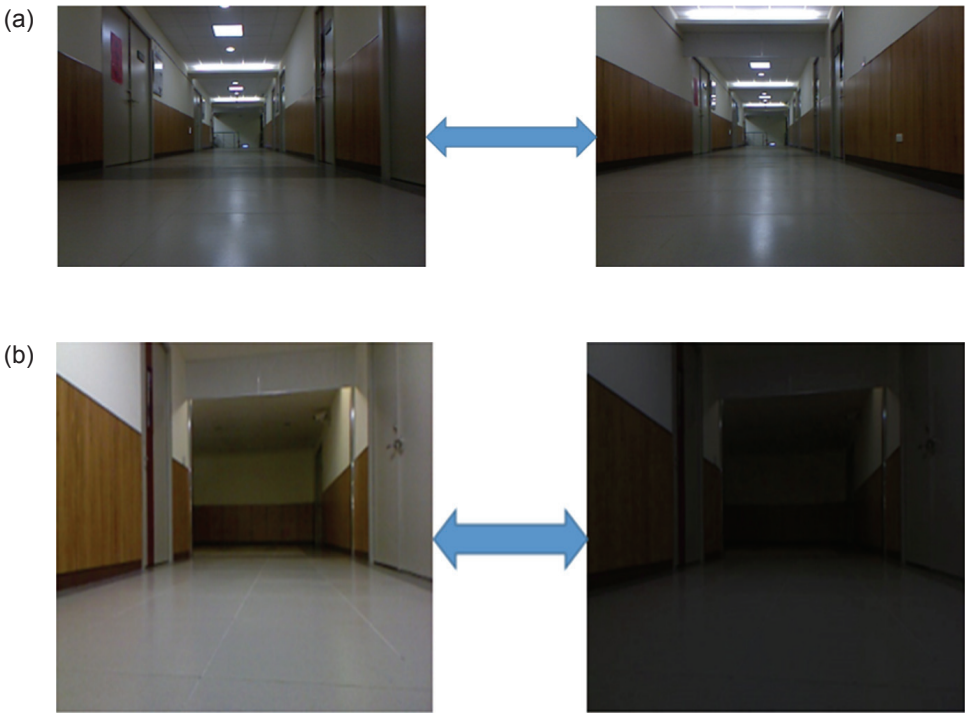


圖 3. (a) 這裡的問題是兩個不同的位置，被認為是一樣的。(b) 這裡的問題是因為光源變化，同一個位置看起來不同。

表 1. 閉環檢測的分類。

| | | 真實場景 (是否發生閉環) | |
|---|---|---|---|
| | | 是 | 不是 |
| 預測 (是否發生閉環) | 是 | true positive | false positive |
| | 不是 | false negative | true negative |

比較的結果盡量不受外在因素影響。舉例來說，有兩張照片，$A$ 和 $B$，如果比較兩者的相似性，那麼在不同的光源環境下，照片 $A$ 和 $B$ 的相似性會得到不同於真實的結果。

## 3. 判斷標準

有好幾種方法可以計算兩張圖片的相似度，所以需要一套標準來判別這些方法的好壞。我們可以把經過閉環檢測後的數據整理成表 1 的形式：

我們希望 TP (true positive) 和 TN (true negative) 可以高一點，FP (false positive) 和 FN (false negative) 低一點，之後根據得到的 TP、TN、FP 和 FN 計算出精準率 (precision) 和召回率 (recall)。

## 4. 詞袋模型

我們使用 ORB (Oriented FAST and Rotated BRIEF) 演算法[19] 提取照片的特徵點，ORB 由兩個部分組成，分別是特徵點提取和特徵點描述，特徵點提取是根據 FAST (features from accelerated segment test, FAST) 演算法，而特徵點描述是根據 BRIEF 演算法 (binary robust independent elementary features, BRIEF) 描述。由於所有圖片題取出來的特徵很多，需要一個好的方法對這些特徵點做分類整理，也就是詞袋模型 (Bag of word)[7]，用 Bow 去描述 ORB 特徵，具有維度的稀疏性，可以更有效率的描述圖片，舉例來說，一張照片有一輛車子、一棟房子，另一張有兩輛車子、一隻狗，我們可以描述第一張照片是由一輛車子、一棟房子組成，第二張照片也是如此。根據這些描述，可以計算兩張圖片之間的相似性詞袋模型，步驟如下：

1. 做特徵分類，像是汽車的特徵、狗的特徵、人的特徵，並以這些特徵組成對應類別的單字，例如狗的特徵包含哪些，汽車的特徵包含哪些，而狗和汽車可以分別把它想成是一個單字，但是這些單字不見得是一個物件，之後，所有單字就會生成一個字典。
2. 檢查圖片中是否出現字典裡的某些單字，並將圖片用向量描述。
3. 比較兩張圖片所對應的向量的相似性。

舉例來說，首先選取一個字典，裡面包含很多單字，例如貓、狗、椅子，並定義為 $w_1$、$w_2$、$w_3$，對於圖片 $A$，可將之表達為：

$$I_A = 1 \cdot w_1 + 0 \cdot w_2 + 1 \cdot w_3$$

此處可以使用向量 $[1,0,1]^T$ 來描述照片 $A$。值得注意的是，因為這個向量描述的是根據特徵數量，而不是照片中哪個位置出現這個特徵，同理，圖片 $B$ 用 $[1,2,2]^T$，根據這兩個向量，可以設計出一個方法來計算兩張照片的相似分數， $a$、$b \in R^W$：

$$S(a,b)=1-\frac{1}{w}|a-b|$$

由於一個單字是由同類型的特徵組成，所以我們可以把建立字典可看作是一個分群問題。分群問題可以自動搜尋數據的規則，例如，我們希望創建一個包含 $k$ 個單字的字典，每個單字由一組相鄰近特徵組成，可以用 K-means 來處理這個問題將會非常有效的。例如，有 $N$ 個數據點，要把它分成 $k$ 個群集，步驟如下：

1. 隨機選取 $k$ 個數據中心點：$C_1,C_2,C_3$ .... $C_k$
2. 對於 $N$ 個數據點，我們計算出數據點 $N$ 與 $k$ 個數據中心點的距離，選擇最近的中心點作為群集。
3. 重新計算每個群集的中心點。
4. 如果新的中心點和原中心點距離夠小，就結束，如果沒有則返回第二個步驟。

除了 K-means 之外，還有許多聚類算法，如 K-means++[8]。現在的問題是如何根據圖片中的特徵，搜尋具有相同類型的單字。計算字典有多少單字不是有效的做法，因數量太多。最簡單的方法是用樹狀結構，實際上，也可以使用更複雜的數據結構，像是 Fabmap[9-11] 使用了Chou-Liu tree[12]。

在 (12) 用了 k-means tree 如圖 4 建立字典，假設有 $N$ 個特徵點，希望建構層是 d，每個分叉是 $k$，如以下步驟所示：

1. 在根節點上，使用 k-means 使將 $N$ 個特徵點分為 $k$ 類，然後完成第一層。
2. 對於第一層中的每個節點，再將其劃分為 $k$ 個類別，之後再獲得下一層。
3. 持續一層接著一層，直到最後一層，這層就是單字。

## 三、物件辨識和序列比較

為了建構一個更好的閉環檢測，使用以下三種方法：(1) 將原始的 ORB 特徵結合詞袋模型用於單張圖片比較，(2) 使用物件辨識來提高場景識別的穩定性，(3) 用時間－空間序列做多張照片的相似性比較。最後，將三種方法的結果融合在一個機率模型裡。

我們先簡單介紹時間－空間序列比較的概念，假設資料庫裡的照片 $I_A$ 與當前所拍的照片 $I_B$ 外觀相似，找跟 $I_A$ 在現實環境中 $x$、$y$ 座標距離最近的兩張照片，和 $I_A$ 一起形成一個

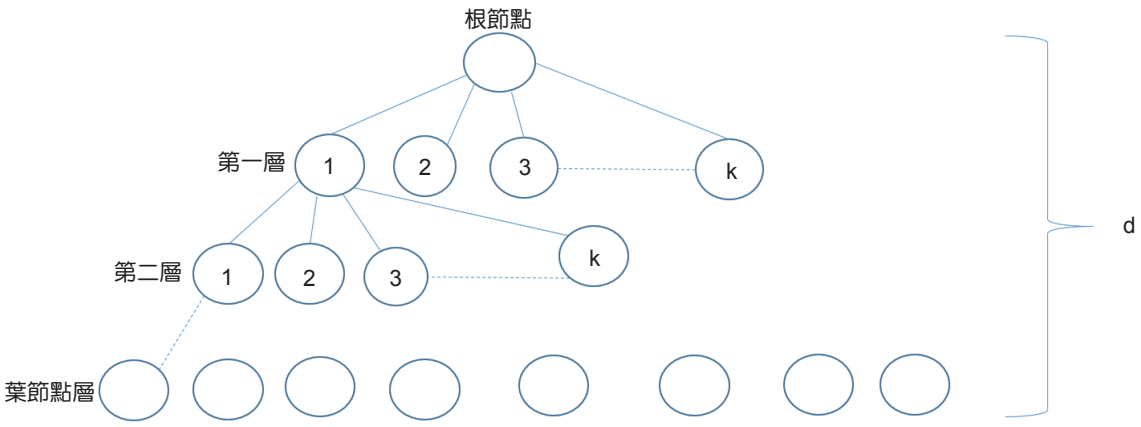


圖 4. K-meanns tree 架構。

空間序列，這個序列會有三張照片，此外，假設照片 $I_B$ 為在 t 時刻所拍攝，我們把它和上一刻 t – 1 與下一刻 t + 1 所拍的照片，形成另一個時間序列，我們拿這兩個序列去比較，這兩個序列分別都是數個三張照片所組成，如果兩個序列的相似性夠高，檢測出來的閉環就更精確。

## 1. 物件辨識

影像識別中一個主要的問題是，極端的光源變化會影響特徵檢測的結果，然而，物件辨識可以使用門、椅子等語意標籤來取得圖片中更高階的訊息，所以可以提高影像識別的穩定性。

首先，將物件分類，把每個物件當作一個單字放至詞袋模型當中，當照片完成物件辨識時，就可以獲得描述照片的向量。根據照片 $I_A$ 的物件，比如椅子、門，可以將之描述為：

$$A = 1 \cdot b_1 + 1 \cdot b_2 + 0 \cdot b_3$$

所以這張照片得到物件向量是 $[1,1,0]^T$，下一步將計算 time-frequency (TF) 和 inverse document frequency (IDF)，有點像是於分類後物件的權重，例如:如果每張照片都有某個物件出現的話，可能這個物件權重就會低一些。

考慮到不同物件的重要性，需要給予物件不同的權重，使得物件的信息更可靠，有一種方法稱作 TF-IDF[13, 14]。計算物件的 TF，就是計算某個物件在一張圖像中出現的次數，使用 IDF 可以減少資料庫中頻繁出現的物件，並提高很少出現的物件的權重。如果在所有照片中很少出現的物件，則相對應 IDF 的值會很高，因為它比其他物件鑑別度更高。假設總物件類別是 $M_{obj}$，在圖片 $I_A$ 物件 $b_i$ 出現 $m_i$ 次，可以寫下 TF-IDF 的定義：

$$TF_{obj,i} = \frac{m_i}{M_{obj,i}} \ , \ IDF_{obj,i} = IDF_{obj,i} \ , \ \log \frac{M_{obj,i}}{m_i}$$

物件辨識 $b_i$ 的權重是 $W_{obj}$，$W_{obj,i} = TF_{obj,i} \times IDF_{obj,i}$，為全部照片以及全部單字的 TF-IDF 權重矩陣。根據照片 $A$ 和照片 $B$ 的 物件，假設兩張的照片向量分別是 $b_A$，$b_B$，我們定義兩張的相似分數如下：

$$S_{obj}(b_A, b_B) = 1 - \frac{1}{2} \left| (2 - 2w_{obj}) \left( \frac{b_A}{|b_B|} - \frac{b_B}{|b_B|} \right) \right|$$

算出來的結果 $S_{obj}(b_A, b_B)$ 的值將會介於 0 到 1 之間。

## 2. 照片序列比較

如果只比較單張照片是否發生閉環，可能會把兩個相同外觀但位置不同的場景視為同一個，Milford 和 Wyeth 提出的 SeqSLAM[15] 通過時間序列的比較可以避免這樣的問題，這個方法適用於鐵路，公路等固定軌跡的情況。對於室內場景，機器人的軌跡可能不固定，所以我們透過地圖位置也就是 $x$、$y$ 座標的訊息，構造空間序列，再跟時間序列做比較。首先，考慮單張照片的比較，計算兩張照片之間的相似度，當隨機給出特徵 $f_i$，搜尋字典中的每一層，最後，得到相對應的單字 $W_i$，假設從一張照片中抓取 $N$ 個特徵，接著從字典找相關的

單字，以收集此照片的單字分布，因為這些單字會被當作是向量的基底，所得到的數個基底可以用來表示這張照片。

由於每個單字的重要程度不同，這裡同樣也需要用到 TF-IDF，為了計算 IDF，首先統計葉節點 $W_i$ 中的特徵數量，它們的比例與特徵總數相關，假設特徵總數為 $n$，$W_i$ 的數量為 $n_i$，那麼這個單字的 IDF 是：

$$IDF_i = \log \frac{n}{n_i}$$

而 TF 則表示在一張照片中的特徵出現的頻率，假設照片 $I_A$ 具有單字 $W_i$，並且單字出現 $n_i$ 次，在字典中，這個單字總共出現了 $n$ 次，那麼這個單字的 TF 就是：

$$TF_i = \frac{n_i}{n}$$

$W_i$ 的權重就是 $TF_i \times IDF_i$，此處用 $\eta_i$ 來代表 $W_i$ 的權重，當考慮權重時，對於照片 $I_A$，它的特徵可以對應於許多單字，結合它們的詞袋描述子：

$$I_A = \{(W_1,\eta_1),(W_2,\eta_2),\ldots,(W_N,\eta_N)\} \triangleq V_A$$

$S(V_A,V_B)$ 的值在 [0,1] 中，我們使用這個分數並設定一個門檻來檢查圖片是否是閉環，公式如下：

$$L(I_A,I_B) = \begin{cases} 1, S(V_A,V_B) \geq T \\ 0, S(V_A,V_B) < T \end{cases}$$

$L(I_A,I_B)$ 表示圖片 $I_A$ 和 $I_B$ 之間是否存在閉環，$T$ 是詞袋的門檻。

如果兩張照片之間的相似性計算出來為 $L(I_A,I_B) = 1$，則視為閉環，然後從資料庫收集照片的位置資訊，構造時間序列和空間序列。此處採用 KNN 演算法收集每個節點的最近鄰居，

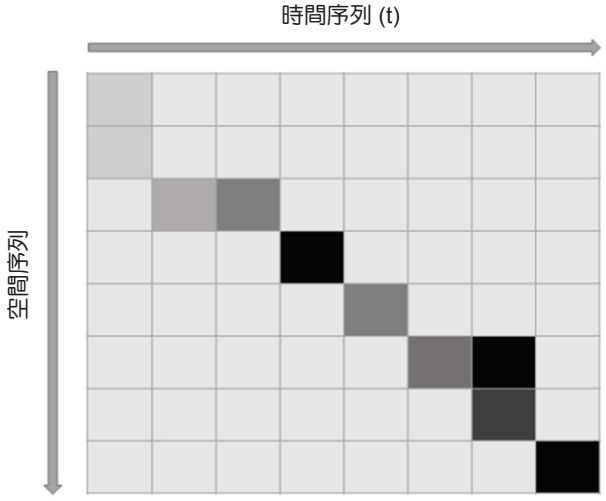


圖 5. 時間－空間序列比較。

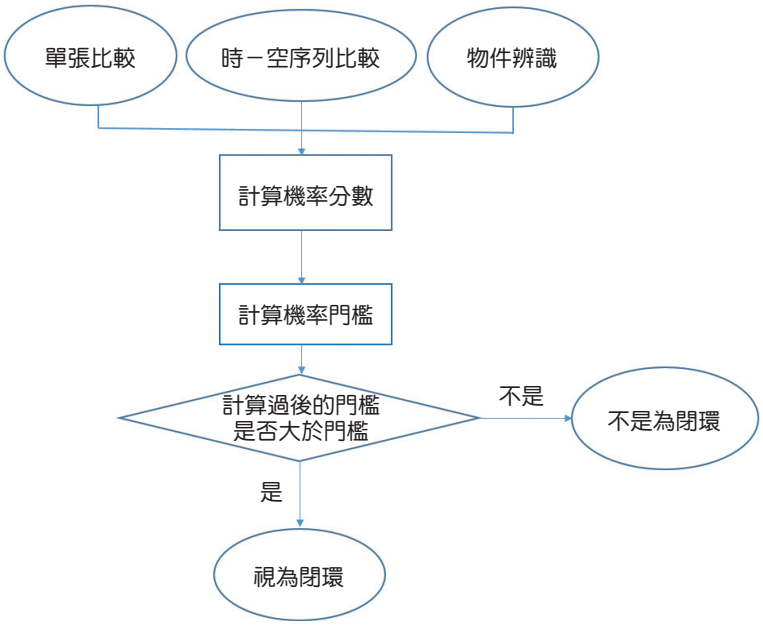


圖 6. 機率評分模型流程圖。

並設定 $K = 3$，也就是按歐式距離收集位置最近的三個照片，然後提取每個圖片的描述子，組合成同一個群集的特徵描述。假設群集由空間序列 $I_A$ 組成，$i$ 是照片索引，則可表示成：

$$C_A = \{I_{A,1}, I_{A,2}, I_{A,3}\}$$

同理， 假設空間序列 $I_A$ 的描述子是 $d_{A,i}$，可以表示為：

$$D_{spatial,i} = d_{A,1} + d_{A,2} + d_{A,3} \triangleq D_A$$

同理，時間的序列 $I_B$ 得到相應的描述子 $D_{time,B}$， 然後使用下列式子計算時間序列和空間序列之間的相似性：

$$S_{time-spatial}(D_A, D_B) = 1 - \frac{1}{2}\left|(2-2\eta)\left(\frac{D_A}{|D_A|} - \frac{D_B}{|D_B|}\right)\right|$$

時間和空間序列比較的效果如圖 5，圖 5 中的時間序列是查詢圖片，空間序列是對應的圖片，較暗的陰影代表時間序列和空間序列之間比較相似，當我們得到 $S_{time-spatial}(D_A, D_B)$，配合物件辨識，就可以從照片中收集更多資訊來比較照片的相似性。

## 3. 三種比較方法的機率模型

在這一個部分中，我們將結合三個分別從 (1) 單張照片比較、(2) 物件辨識、(3) 時間和空間序列 (多張照片) 比較的結果，並根據機率評分來判斷是否是閉環，架構如圖 6 所示：

首先：從三種方法中分別得到三個分數，並按照以下步驟進行：

1. 調整三種比較方法的權值 $\alpha$、$\beta$、$\gamma$。
2. 計算閉環發生的機率和期望值。

表 2. 機率評分模型參數設置。

| | |
|---|---|
| $S(I_i, I_j)$ | 圖片 $I_i$ 和圖片 $I_j$ 的相似性通過單幀比較，取值在 [0,1] 之間。 |
| $S_{time-spatial}(C_A, C_B)$ | 時間序列 $C_A$ 和空間序列 $C_B$ 的相似性，取值在 [0,1] 之間。 |
| $S_{obj}(I_i, I_j)$ | 圖片 $I_i$ 和圖片 $I_j$ 出現物件的相似度，取值在 [0,1] 之間。 |
| $\alpha$ | 單幀比較的分數權重，取值在 [0,1] 之間。 |
| $\beta$ | 時間得分的權重和空間序列的比較，取值在 [0,1] 之間。 |
| $\gamma$ | 物件辨識方法的權重，$\gamma = 1 - \alpha - \beta$。 |
| $X_i$ | 隨機變量，取值在 [0,1] 之間，代表圖片 $I_i$ 單幀比較是否發生閉環。 |
| $X_{(obj,i)}$ | 隨機變量，取值在 [0,1] 之間，這意味著圖片是否偵測到物體。 |
| $X_{(spatial,i)}$ | 隨機變量，取值在 [0,1] 之間，這意味著圖片 $I_i$ 時間－空間序列的比對是否發生閉環。 |

3. 計算有辨識到物件和沒有辨識到物件的分數和門檻值。
4. 判斷照片是否發生閉環。

除此之外，還要設置一些參數和符號，如表 2 所示：

相似性評分與門檻值　這裡使用 (0,1) 之間的參數 $\alpha$、$\beta$ 和 $\gamma$ 計算相似性分數的加權平均，然後再對參數進行調整。評分模型的定義如下：

$$S_{prob} = \begin{cases} \alpha \times S(I_i, I_j) + \beta \times S_{time-spatial} + \gamma \times S_{obj}(I_i, I_j) \text{ , } S_{obj}(I_i, I_j) > 0 \\ \dfrac{\alpha}{\alpha+\beta} \times S(I_i, I_j) + \dfrac{\beta}{\alpha+\beta} \times S_{time-spatial} \text{ , } S_{obj}(I_i, I_j) < 0 \\ \alpha+\beta+\gamma=1 \end{cases}$$

當 $S_{obj}(I_i, I_j) = 0$ 時，用 $\alpha/\alpha+\beta$ 和 $\beta/\alpha+\beta$ 做為 $S(I_i, I_j)$ 和 $S_{time-spatial}(I_i, I_j)$ 的權重，即使沒有辨識到物件，單張圖片和時間－空間序列比較比較方法的權重仍有同樣比例。

計算相似性分數之後，還需要給定門檻值，當分數超過門檻值，對這個方法來說，才算有檢測到閉環，為選取合適的門檻值，此處考慮三個方法各別檢測到閉環的期望值，因為期望值代表一種平均，相似性分數應該要超過平均才視為有檢測到閉環。另外，如果沒有辨識到物件，也會給出另一個期望值，最後將這兩個期望與不同的權重結合起來，整體的門檻值就是：

$$T_{prob} = (\alpha+\beta) \cdot E\left[X \cdot X_{time-spatial} \middle| X_{obj} = 0\right] + \gamma \cdot E\left[X \cdot X_{time-spatial} \middle| X_{obj} = 1\right]$$

現在，透過相似性評分與門檻值，我們可以著手閉環檢測的實驗。

## 四、實驗環境與實驗結果

### 1. 環境設置

首先，要獲取環境訊息，這裡使用 RGB-D 傳感器 Kinect v1 搭載機器人 TurleBot2 來獲取資料集，並且用 RTABMAP[16] 軟體來掃描環境，以及 C++ 程式語言和 SQLite, OpenCV, DBoW3[17], YOLO libraries。

## 2. 單張照片 (詞袋模型) 與多張照片 (時間－空間序列) 方法比較

為了測試有效性，在室內環境的應用如圖 7(a)、(b) 所示：

由於空間－時間序列需要調整門檻值，如圖 8 所示，不同門檻會有不同表現， 當召回率低時，不同門檻的準確率非常接近。但當門檻設置為 0.01 時，並且召回率比較高時，準確率會比其它門檻高，當時間－空間序列門檻設置的更嚴格時，曲線會向左移動，由於門檻過高，檢測到的閉環數比較少。

## 3. 分析和總體比較結果

我們需要測試不同權重 $\alpha$、$\beta$ 和 $\gamma$，相應的方法分別是 (1) 單張圖片比較、(2) 時間－空間序列比較和 (3) 物件辨識，如圖 9 和圖 10 所示，因為 $\gamma = 1 - (\alpha + \beta)$，只需要調整兩個參數，由於 $\alpha$、$\beta$ 是小數點以下一位的實數，可以用貪婪搜索法來搜尋所有可能的 $\alpha$ 和 $\beta$。

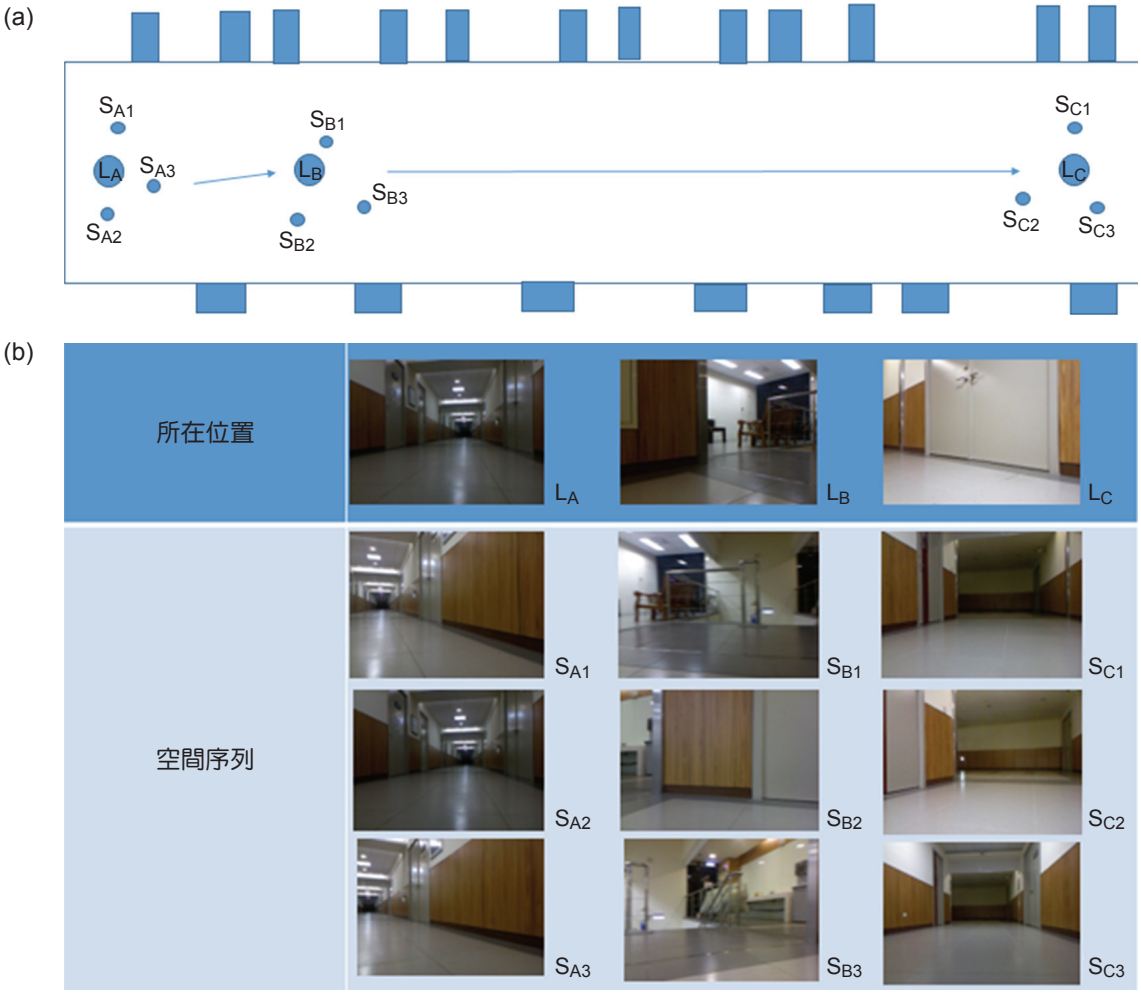


圖 7. (a) $L_A$，$L_B$，$L_C$ 是機器人當前所在位置，並收集資料庫裡歐式距離最近的關鍵幀 (keyframe) $S_{Ai}$，$S_{Bi}$ ，$S_{Ci}$，$i = 1$，2，3，以構造空間序列。(b) 最上列為當前位置的圖片，在第一列的下方為相對應的空間序列，序列匹配可提高閉環檢測的準確率，結合物件辨識，檢測結果會更佳穩健。

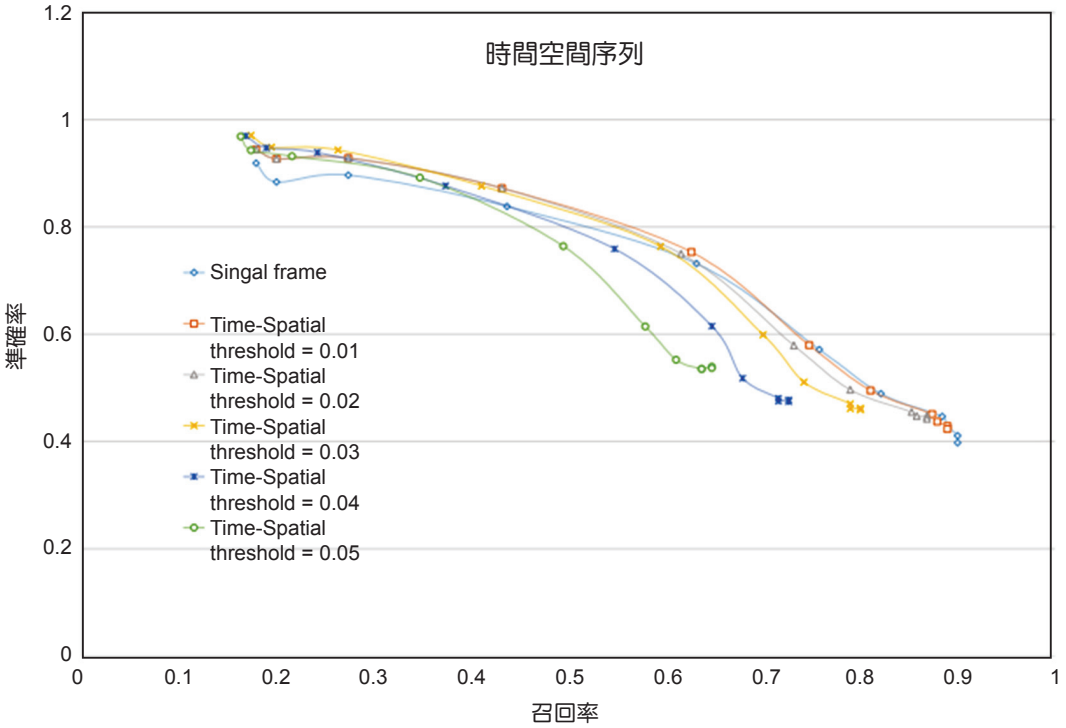


圖 8. 空間－時間序列比較不同門檻設定的表現。

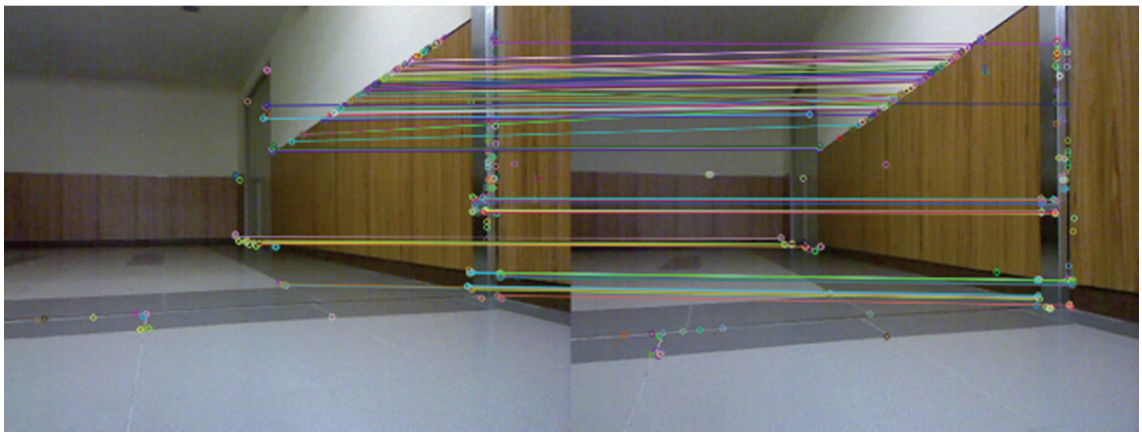

圖 9. 用 ORB 特徵結合詞袋模型方法進行單張照片比較。

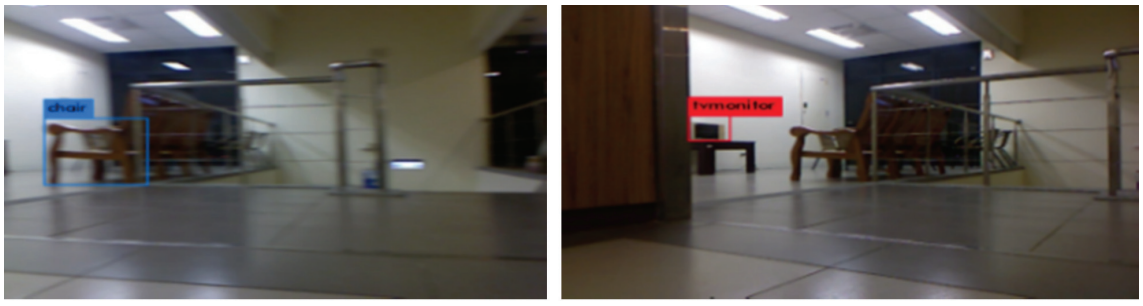


圖 10. 室內場景用 Yolo 做物件辨識。

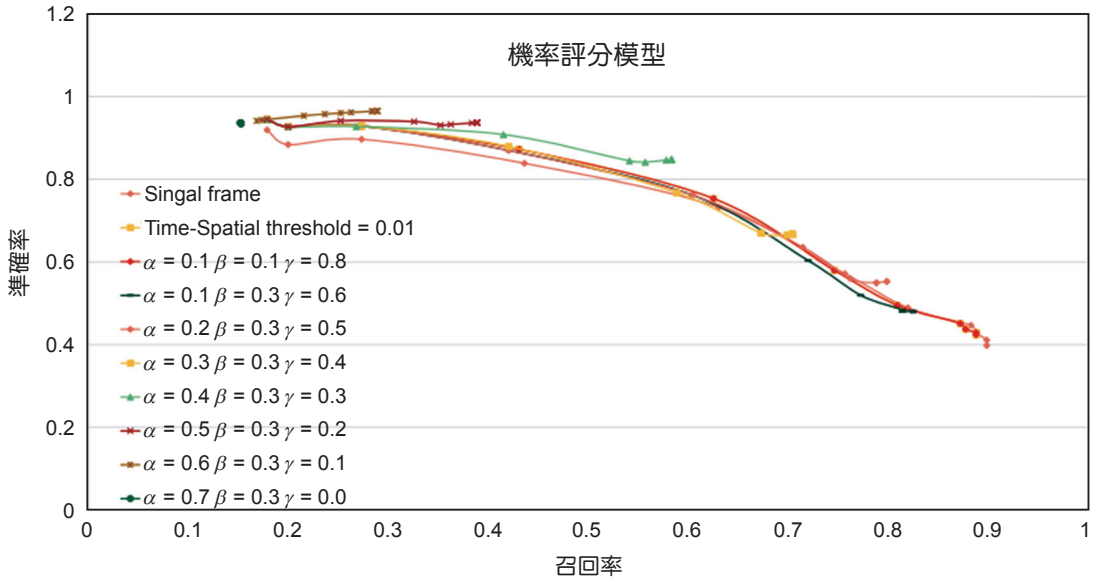


圖 11. 機率模型比較不同門檻設定的表現。

一般來說，透過 ROC 曲線[18] 可以選出表現最佳的參數，繪製曲線後，我們發現 $\alpha + \beta = 0.7$ 或 $\alpha + \beta = 0.8$ 可能是比較好的選擇。

由於在 SLAM 中準確率比召回率更重要，考慮 $\beta$ 為 0.3，測試結果如圖 11，當 $\alpha = 0.6$ 和 $\gamma = 0.1$，精度可以超過 0.95。如果放鬆準確率的限制，並增加物件辨識的權重 $\gamma$，也就是令 $\alpha = 0.4$ 和 $\gamma = 0.3$，可以發現 $\alpha + \beta$ 越高，曲線會往左上方移動，因為 $\alpha$、$\beta$ 可以調整門檻和機率分數，並且影響結果，如果 $\alpha$、$\beta$ 選擇越高，這會讓門檻變嚴苛，所以準確會越來越高。另一方面，如果 $\gamma$ 選擇較高的數值，準確率會下降，但是召回率上升，因此，$\alpha$、$\beta$ 的選擇是閉環檢測中平衡精度與召回率的因子。

## 五、結論

在這篇文章中，我們整合了比較照片間的特徵用詞袋模型的方法，以及空間位置、連續時間資訊和物件辨識的結果，以提高閉環檢測的準確性。我們的機率評分模型能夠評估多類特徵，能夠適應不同的環境，並且可以透過參數調整使得精準度－召回率滿足 SLAM 的要求，將來可以研究更多不同的特徵類型以及結合類神經網路，以進一步優化 SLAM 中的閉環檢測。

### 參考文獻


1. Engel, J., Sch ¨ ops, T., and Cremers, D., “LSD-SLAM: Large-scale direct monocular SLAM”, *European Conference on Computer Vision*, Springer (2014).
2. Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D., *IEEE Transactions on Robotics*, **31** (5), 1147 (2015).
3. Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., and Burgard, W., “An evaluation of the RGB-D SLAM system,” *IEEE International Conference on Robotics and Automation (ICRA)*, 14-18 May (2012).
4. Che-Yu Yang, Yu-Cheng Zhang, Yu-Hsiu Chen, and Chih-Wei Huang, “Toward semantic loop closure in simultaneous localization and mapping systems”, Proc. SPIE 10745, *Current Developments in Lens Design and Optical Engineering XIX*, 17 September (2018)

5. Lowe, D. G., *International journal of computer vision*, **60** (2), 91 (2004).
6. Dirk Hahnel, Wolfram Burgard, Dieter Fox, and Sebastian Thrun, “An efficient fastslam algorithm for generating maps of large-scale cyclic environments from raw laser range measurements”, *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, **1**, 206 (2003).
7. G´alvez-L´opez, D. and Tardos, J. D., *IEEE Transactions on Robotics*, **28** (5), 1188 (2012).
8. David Arthur and Sergei Vassilvitskii, “k-means++: The advantages of careful seeding”, *In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, Society for Industrial and Applied Mathematics, 2007.
9. Mark Cummins and Paul Newman, *The International Journal of Robotics Research*, **27** (6), 647 (2008).
10. Mark Cummins and Paul Newman, *The International Journal of Robotics Research*, **30** (9), 100 (2011).
11. Mark Cummins and Paul Newman, *IEEE Transactions on Robotics*, **26** (6), 1042 (2010).
12. C Chow and Cong Liu, *IEEE transactions on Information Theory*, **14** (3), 462(1968).
13. Sivic, J. and Zisserman, A., “Video google: A text retrieval approach to object matching in videos”, *Proceedings of Ninth IEEE International Conference on Computer Vision*, (2003).
14. Robertson, S., *Journal of documentation*, **60** (5), 503 (2004).
15. Milford, M. J. and Wyeth, G. F., “SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights”, *IEEE International Conference on Robotics and Automation (ICRA)*, (2012).
16. Labb´e, M. and Michaud, F., *Autonomous Robots*, **42**, 1133 (2018).
17. DBoW3 Library : https://github.com/rmsalinas/DBoW2 .
18. Fawcett, T., *Pattern recognition letters*, **27** (8), 861 (2006).
19. E. Rublee, V. Rabaud, K. Konolige and G. Bradski, “ORB: An efficient alternative to SIFT or SURF”, *2011 International Conference on Computer Vision*, (2011).

## 作者簡介

楊哲宇先生為國立中央大學通訊所碩士畢業生。

Che-Yu Yang received his M.S. in Communication Engineering from National Central University.

張育晟先生現為國立中央大學數學所博士生。

Yu-Cheng Zhang is currently a Ph.D. student in the Department of Mathematics at National Central University

陳毓琇小姐現為國立中央大學通訊所碩士生。

Yu-Hsiu Chen is currently a M.S. student in the Department of Communication Engineering at National Central University

黃志煒先生為美國華盛頓大學電機系博士，現為國立中央大學通訊系副教授。

Chih-Wei Huang received his Ph.D. in Electrical Engineering from University of Washington, USA. He is currently an associate professor in the Department of Communication Engineering at National Central University.