

中華民國第 59 屆中小學科學展覽會 作品說明書

高級中等學校組 電腦與資訊學科

第二名

052509

結合 SLAM 及 YOLO 建構 AR 空間記憶眼鏡

學校名稱：臺北市立內湖高級工業職業學校

作者： 職二 廖子游 職二 邵明諒 職二 林禹丞	指導老師： 黃明超 張俊雄
---	-----------------------------

關鍵詞：即時定位與地圖構建、物體偵測、擴增實境

摘要

本研究結合 SLAM 即時定位與地圖構建及 YOLO 即時物體偵測，實現環境物品辨識、定位、導航的 AR 記憶眼鏡。透過 SLAM 平面偵測建構空間中平面資訊，作為 AR 物件座標及物理資訊的依據參考，再經過機械學習訓練後的 YOLO 模型達到辨識物品，結合兩者建立空間與物品的依存關係。以 VR 眼鏡結合手機模擬 AR 眼鏡，採用 iOS 系統及 Swift 語言設計程式功能及介面，實現走動過程持續即時辨識空間與物品，可運用音量鈕選定要尋找物品，再透過眼鏡畫面方向及距離的指示，導航到該物品位置。本作品還可運用在工作場合，做到紀錄與提醒環境的變化，如航警可疑物盤查、公司財務貨物盤點等，成為幫助人腦記憶的輔助裝置。

壹、研究動機

現代社會生活步調快速，事情多且繁雜，做事常常需要一心多用，造成忘東忘西的現象，再加上人口高齡化，老化引發的記憶力衰退、健忘症與失智症人口，也都在不斷的增加。上述症狀者都會有忘記物品放置位置或忘記曾經做過的事情這類問題，或許只需要給予一些提醒就能使其回想起來。如果我們能將所有看過的物品資訊都記錄在電腦中，要尋找物品時，再使用電腦中的搜尋功能來尋找物品，提醒或告知使用者物品位置，就能讓使用者更有效率的去尋找物品。

藉由計算機概論介紹的熱門技術，運用程式設計課程學的程式邏輯，思考結合空間定位與影像辨識運用於 AR 眼鏡，以大腦記憶空間事物的方式，製作幫助人腦記憶的裝置。在選修課程學習演算法的原理，應用程式開發與專題實作課程製作程式內容，課後閱讀書籍論文，實現以手機與 VR 眼鏡模擬 AR 眼鏡，並測試實際使用狀況。

比較市面上現有記憶物品的解決辦法，將物品裝上藍芽發送裝置，用裝置取得物品位置，或外裝多台攝影機交叉比對影像定位的方式，與本作品的方法相比之下，本作品在記憶、運用的方式上，更符合人性化的使用，應用上更具有實用性，對其他領域都能有很大的貢獻。

貳、研究目的

- 一、蒐集並瞭解關於人類大腦的記憶與現有 AR 眼鏡的資料。
- 二、研究運用 SLAM 系統建構三維空間的原理，並用於建構環境座標系，與定位使用者當下的位置與視角。
- 三、研究物體偵測技術，運用 YOLO 方法，達成準確辨識物品種類、得知物品在圖片中的位置等資訊。
- 四、結合三維空間建模與物體偵測，以達到在座標系中定位與記錄物品資訊。
- 五、撰寫 AR 眼鏡搜尋與導航系統，讓使用者藉由搜尋與導航功能得知物品位置。

參、研究設備及器材

名稱	版本及型號	用途
(AR 眼鏡)	因市面上 AR 眼鏡價格昂貴，且開發功能難以使用，所以使用手機和 VR 眼鏡代替 AR 眼鏡的功能	
iPhone 手機	iPhone XR iOS 系統 12.1.2 iPhone 6s Plus iOS 系統 12.1.2	APP 安裝與執行的設備
VR 頭套	VRBOX 5PLUS	將手機固定在頭上，達到使用手機取代 AR 眼鏡
Mac 電腦	macOS 系統 10.14.2 Beta 版	iOS 系統手機的開發和測試
Xcode 開發軟體	10.1 版	Swift 程式語言的整合開發環境
Swift 程式語言	4.2 版	用於撰寫 APP
ARKit 框架	2 版	撰寫 AR 處理功能
SceneKit 物理引擎	1 版	撰寫 3D 物件與碰撞環境
Core ML 框架	2 版	影像辨識模型能在手機上使用
Anaconda	4.4.10 版	Python 程式語言的整合開發環境
Python 程式語言	3.6.6 版	用於撰寫 TensorFlow
TensorFlow 軟體	1.9 版	製作影像辨識的物件追蹤模型

肆、研究過程或方法

一、資料探討

(一) 人類的記憶

1. 記憶事情

人類記憶事情主要分成三個不同階段，從感官（如：視覺）接收到訊息，傳入下視丘形成短期記憶（持續約三十秒）的編碼過程；到海馬迴將部分短期記憶轉換為長期記憶（持續數天以上），在大腦皮層進行紀錄的儲存過程；以及提取儲存內容的檢索過程。因此，只要其中一部分未完成，就可能無法記住事情。

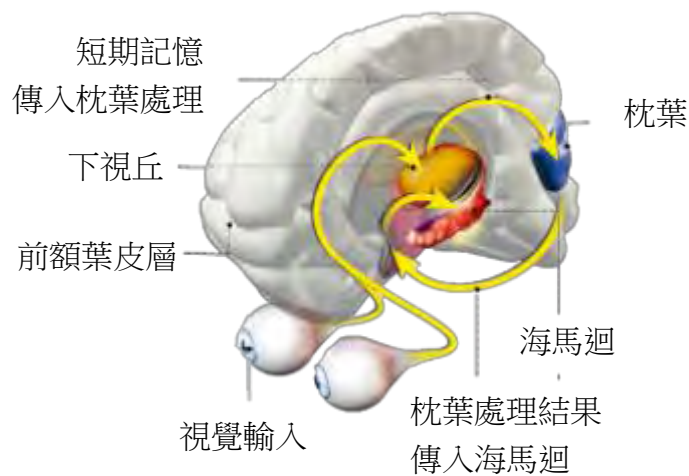


圖1. 大腦的記憶流程

2. 記憶空間與認知位置

大腦記憶環境空間的方式，是網格細胞（Grid Cells）在腦中，以六角形網格的方式做到建立座標系統、建構認知地圖。透過不同區域網格細胞的反應，使位置細胞能瞭解自身在認知地圖中的位置。人透過移動雙眼得知不同邊界的前後距離，使網格細胞建構認知地圖，運用位置細胞確定要如何移動。



圖2. 大腦對於環境的記憶

位置細胞場域

網格細胞場域

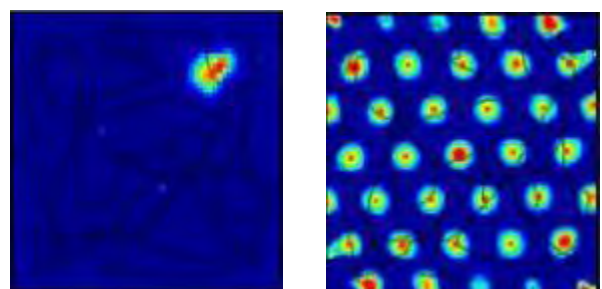


圖3. 位置細胞與網格細胞

(二) AR 眼鏡 (AR Glasses)

AR 最廣泛的定義為一九九七年由羅納德·阿祖瑪 (Ronald Azuma) 提出，主要說明 AR 的功能包含：結合虛擬與現實、即時互動與 3D 定位。也就是將電腦的資訊疊合到現實世界，讓使用者看現實空間時，能看到虛擬物件的資訊或物件。

1. AR 眼鏡介紹

AR 眼鏡的發展早在 1968 年，科學家 Ivan Sutherland 開發了第一套頭戴式 AR 顯示裝置，當時的裝置只能在現實中顯示單純的框線圖。多年之後隨著硬體設備與計算能力的提升，2013 年 Google 公司推出 AR 頭戴式顯示器 Google Glass，用於顯示各種即時資訊給使用者得知。現今 AR 最常應用於手機上，因為手機的研發、作業系統、以及生態系等，相較於 AR 眼鏡更完整，但手機對於 AR 可能只是一個過渡期，未來 AR 眼鏡很有可能取代手機成為 AR 常用的攜帶式裝置。根據 IDC 的 AR 市場營收分析，可見未來 AR 在市場中是很有發展價值的一種技術。



圖4. Google Glass 示意圖

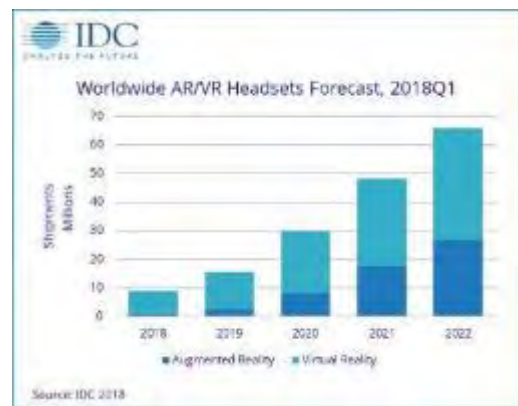


圖5. IDC AR/VR 分析圖

2. AR 眼鏡功能

近年來許多公司積極研發 AR 眼鏡，因為 AR 眼鏡攜帶與顯示的方式，改善了使用裝置的便利性；AR 眼鏡上的感測元件與使用者的位置與視角同步，能夠取得使用者的頭部動作與視角；配合處理元件，計算與分析資訊，讓使用者透過鏡片看現實景象時，能看到虛擬資訊與現實影像結合後的結果，達到融合現實提供更準確有用的資訊。

二、作品架構

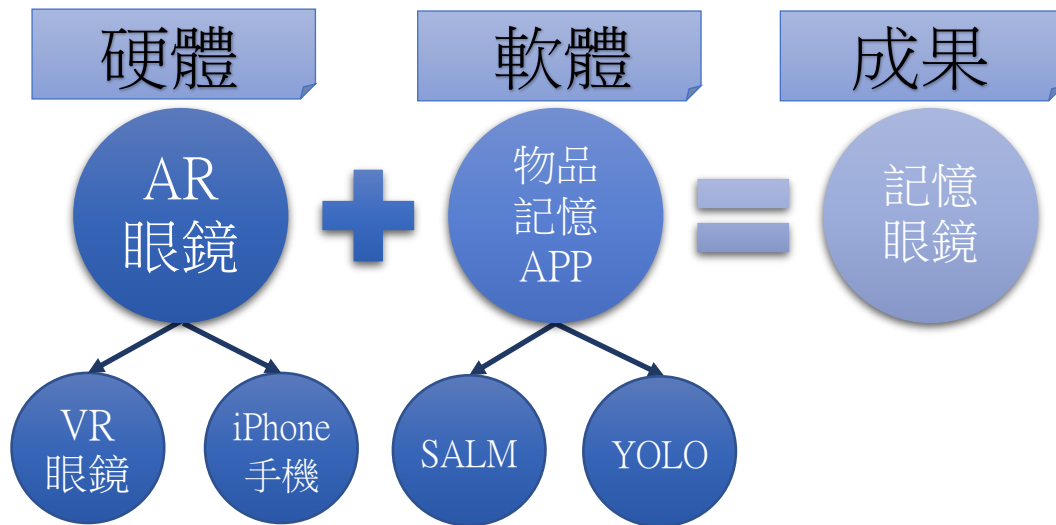


圖6. 作品軟硬體架構圖

(一) 硬體設備

1. AR 眼鏡的現有問題

現在市售的 AR 眼鏡價格普遍偏高，售價（三萬到十萬台幣不等），且硬體規格與作業系統不統一，造成開發者開發 AR 眼鏡 APP 較為困難，因此 AR 眼鏡的附加功能較少，整體性價比不如手機，市場普及率低。

2. AR 眼鏡替代方案

本研究 AR 眼鏡設備採用 VR 眼鏡及 iPhone 手機替代，達到符合 AR 眼鏡攜帶、顯示、運算的功能，價格比起市售 AR 眼鏡，可降到 1000 元以下。且市面上可執行 AR 程式功能的手機，相當普遍，開發資源也較為廣泛。

3. 替代方案存在的問題

- (1). 視野大小：使用時畫面只顯示視角六分之一的影像。
- (2). 重量：重量超過四百公克，以眼鏡的方式攜帶在頭上，難以長時間舒適的使用。

- (3). 3D 暈眩：視角快速改變，虛擬影像與真實影像未能即時同步，會造成暈眩問題。



圖7. 使用者攜帶樣式

(二) 軟體開發

使用蘋果公司的 Mac 電腦，在作業系統 macOS 中安裝 Anaconda 與 Xcode 開發環境，Anaconda 下載 Python 程式語言執行 Tensorflow 套件，將輸入的訓練資料經過機器學習 (Machine Learning) 訓練出影像辨識模型，用 Xcode 撰寫 Swift 程式語言，使用 ARKit、SceneKit、Core ML 套件，並加入影像辨識模型，以開發 APP (application 應用程式) 的方式來製作作品的軟體部分，此方法在開發上相較於 AR 眼鏡的開發環境，能有更多的學習資源，在相容上也較完整。使用者使用時，只需擁有符合的手機就能安裝 APP 來使用，能非常方便實現普及。

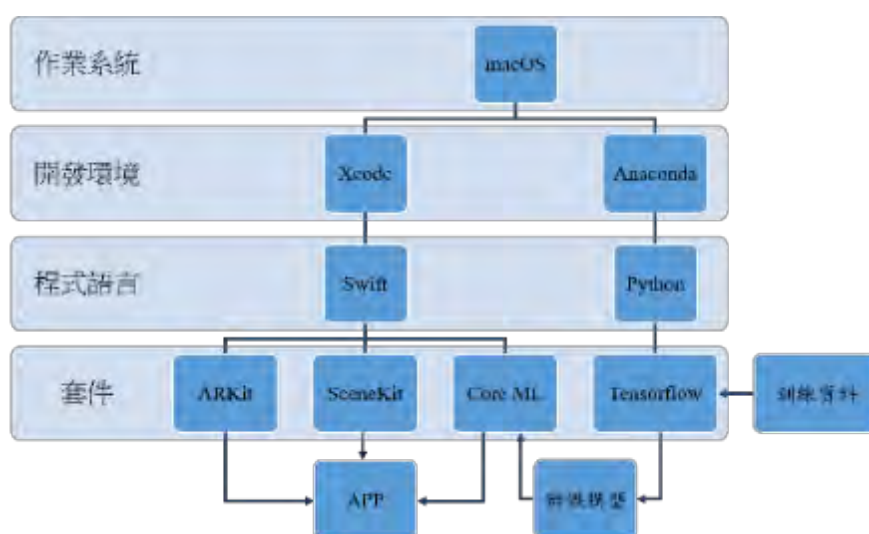


圖8. 軟體開發架構

(三) 系統運作流程

1. 辨識：使用者平時攜帶此作品在環境中移動時，就能因為視角與位置的轉變，給予辨識環境所需要的資訊，進而辨識出使用者在空間中的資訊、空間平面資訊與圖中物品資訊。
2. 記錄：將辨識到的平面資訊與圖中物品資訊統整並以座標化的方式記錄於裝置中，以便於使用者搜尋物品時，能取得空間中物品位置與資訊以及最後的影像。



圖9. 簡易運作方向

3. 搜尋：使用者操作耳機的音量按鍵選擇要搜尋的物品種類，就能檢索記錄中種類符合

的資訊，並將訊息顯示在 AR 眼鏡上，讓使用者可以去尋找所選的物品，或是依據其他條件提醒使用者去尋找。

4. 導航：取得搜尋資料後，分析記錄中物品的空間位置，根據使用者當下位置及視角指示到達物品位置的距離及方向，並將資訊顯示在眼鏡上，讓使用者能依據資訊去尋找物品。



圖10.AR 空間記憶眼鏡的運作流程圖

三、系統設計及規劃

系統的運作方式模仿大腦記憶事物的方式，依循以辨識、記錄、搜尋及導航四大流程，達到大腦記憶事物的編碼、處存、檢索的功能，針對運用的各項技術及演算法說明如下。

- (一) 辨識：辨識包含兩個主要方向，「辨識空間」與「辨識物品」，流程如圖 11 所示。

辨識空間以 SLAM（即時定位與地圖構建 Simultaneous localization and mapping）系統當作辨識方法，SLAM 系統的感測器分為用 ToF（飛時測距 Time-of-Flight）的雷射感測方式與用視覺的單目（單鏡頭相機）或雙目（多鏡頭相機）感測方式，比較資料如表 1。

辨識物品資訊是運用影像辨識，根據對物品的聚焦程度與實際需要，依序分為三個層次：從確定圖像內容的分類（Classification），將圖片依事先確定好的類別做分類；到框

出圖像物品位置的物體偵測 (Object Detection)，依序對不同位置做分類；以及能畫出物品輪廓的圖像分割 (Image Segmentation)，增加對邊緣與前後景的處理，如圖 12。

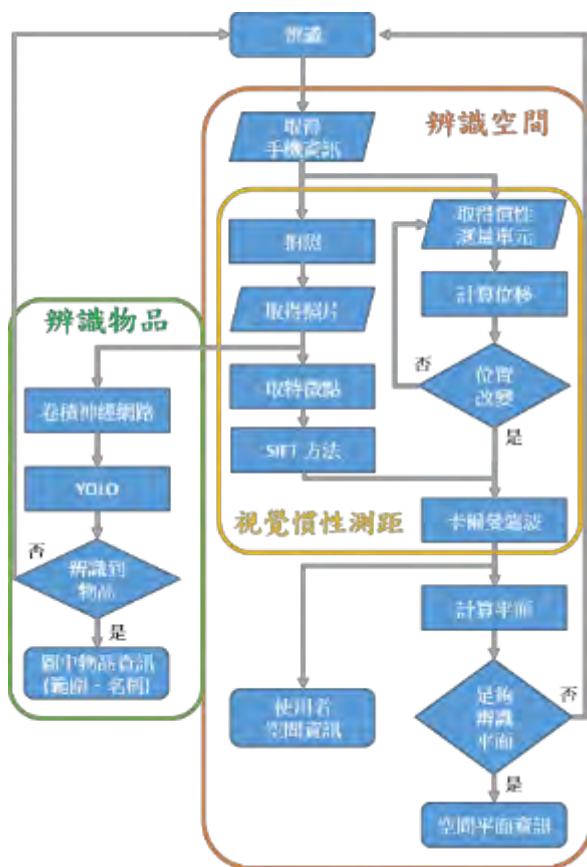
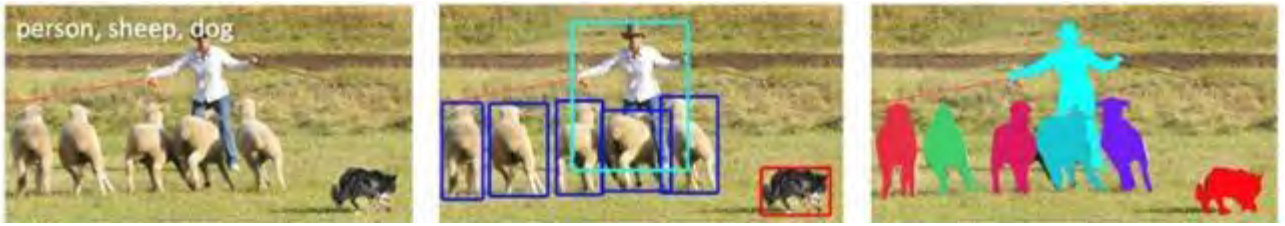


圖11.辨識流程圖

表1. SLAM 系統感測器比較表 來源：組員整理

感測器類型	雷射	視覺	
相機	飛時測距 ToF	單目 (單鏡頭)	雙目 (多鏡頭)
原理	以自身與環境障礙物間的距離作為感測依據。	以多張影像與移動訊息作為感測依據。	多個相機的單張影像與相機之間的距離作為感測依據。
優點	精準度高，環境影響小，速度快，計算量小。	結構簡單，成本低，技術資源多。	易取得深度信息，識別環境大。
缺點	價格昂貴，缺少影像資訊，技術較少。	易受環境影響，誤差大，缺少深度資訊。	高計算資源，速度慢，體積大。

由表 1 得知，視覺 SLAM 包含影像資訊，可直接用於物體偵測，以手機的單鏡頭與內建 IMU (慣性測量單元 Inertial measurement unit) 設備，就能做到單目 SLAM 系統，不須額外加裝其他設備，且單目 SLAM 系統只需較少的處理器計算資源，就能達到辨識空間資訊，符合小型設備應用。



分類 (Classification) 物體偵測 (Object Detection) 圖像分割 (Image Segmentation)
圖12.辨識物品資訊三個層次示意圖

取得相機拍攝的影像後，以手機處理器達到即時辨識物品，且需要有圖像中物品的位置做到定位物品在空間中的位置，考量處理速度與實際需求，選擇以物體偵測作為辨識物品的方式。

1. 辨識空間

與人類的空間感知模式相似，使用影像與位移達到辨識空間、定位自身位置，但本研究單目 SLAM 系統地圖架構以單鏡頭拍攝影像並加入 IMU 數據建構環境地圖，運用 VIO (視覺慣性測距 Visual Inertial Odometry) 技術測得自身方向與位移量，並以座標系的方式紀錄環境及自身的位置。簡要說明如下：

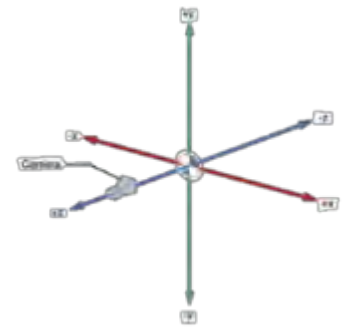


圖13. 座標系圖示

- (1) 單目 SLAM 系統：SLAM 系統主要功能為確定自身的位置與建構環境地圖。單目 SLAM 運用 VIO 技術以單鏡頭透過 SIFT (尺度不變特徵轉換 Scale-Invariant Feature Transform) 取得影像位移預估值，並與 IMU 的位移預估值經過卡爾曼濾波 (Kalman Filter) 互相校正以取得較精準的設備運動資訊。將空間資訊匯入 SceneKit 物理引擎的座標系中 (圖 13)，以右手座標系繪製環境地圖。當使用者開啟 APP，眼鏡就會將當下的位置設為地圖世界座標系的原點，移動時持續繪製環境地圖並定位。

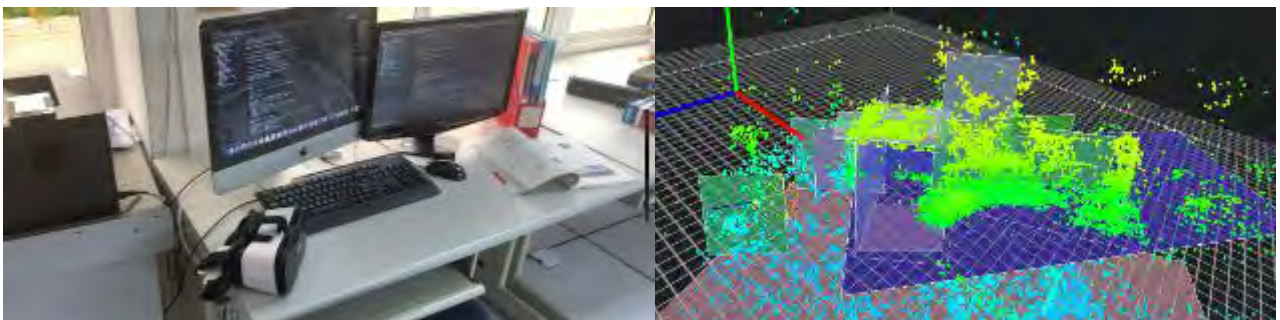


圖14.單目 SLAM 系統地圖的特徵點分布圖

(2) VIO 技術：VIO 技術是整合影像資訊與 IMU 資訊，做到測量自身移動量，VIO 技術有兩種整合方式，緊耦合（tightly-coupled）與鬆耦合（loosely-coupled），兩種方式都使用影像與 IMU 的位移預估資訊，兩種方法分析如（表 2），因為資料穩定度低且需要快速計算，採用鬆耦合方式取得位移。



圖15.VIO 方法示意圖

表2. VIO 技術位移預估方法 來源：組員整理

方法	緊耦合（tightly-coupled）	鬆耦合（loosely-coupled）
原理	同時優化影像與 IMU 的位移預估，建立統一位移預估值。	將影像的位移預估值與 IMU 的位移預估作為單獨模組，最後統整出位移預估值。
計算複雜度	高	中
優點	精度高	演算法簡單、誤差容忍度大

鬆耦合 VIO 是將鏡頭拍攝到的每幀畫面經過 SIFT 方法，根據特徵點（feature point）在每幀畫面間位置的變化與 IMU 等資訊，進行卡爾曼濾波比較與運算，最終將兩種不準確的預測結果，統整出較為準確的設備運動資訊。

(3) SIFT 方法：SIFT 方法的用途是從兩張影像取得自身的影像位移預估，先將圖片灰階後，取圖元（pixel）點值與周圍差距較大的或邊緣的交會點作為特徵點，之後匹配兩張圖片的特徵點，以取得多個點的向量，整合多個向量並去除誤差，最後計算出預估位移結果。

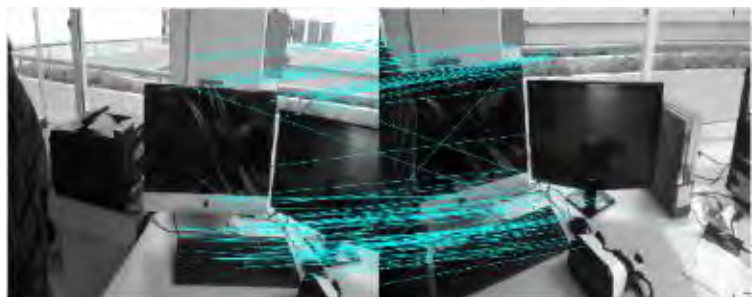


圖16. SIFT 方法 示意圖

(4) 卡爾曼濾波：卡爾曼濾波是一種遞歸濾波器，能從一系列不完整且包含雜訊的測量結果，估計出動態系統的狀態。卡爾曼濾波主要用於將影像的預估位移結果，與 IMU 的預估位移結果經過公式運算 如圖 16，最終得出較精準的設備運動資訊。

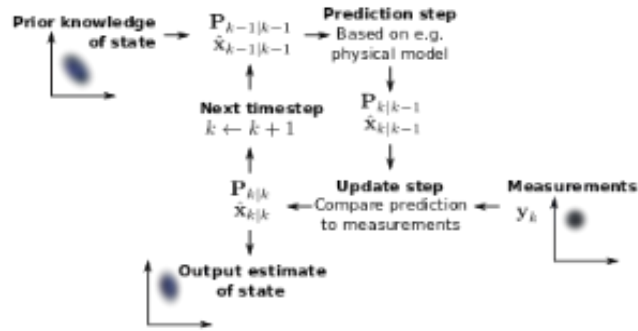


圖17.卡爾曼濾波流程與公式圖

(5) 空間平面資訊：根據尋找出的多個特徵點，再依據尺度、方向和位置建構平面，並判斷符合的結果，就可以得出多個空間平面的位置資訊。

2. 辨識物品

加入物體偵測模型，用以辨識物品在圖中的位置，現有的物體偵測方法是用 CNN（卷積神經網路 Convolutional Neural Networks）演算法辨識影像做到分類，並搭配深度學習模型來確定物品位置。

(1) CNN：CNN 的過程分為卷積層 (Convolution)、池化層 (Pooling)、平坦層 (Flatten)、全連接層(Fully Connected Layer)、輸出層 (Output)這五層，流程如下：

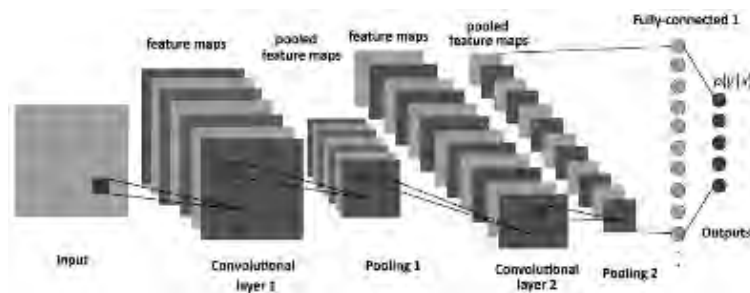
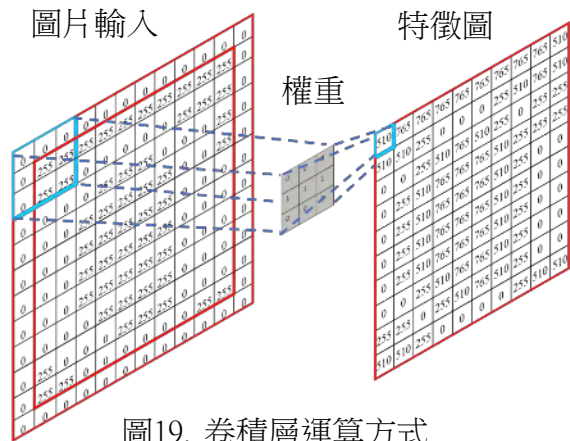
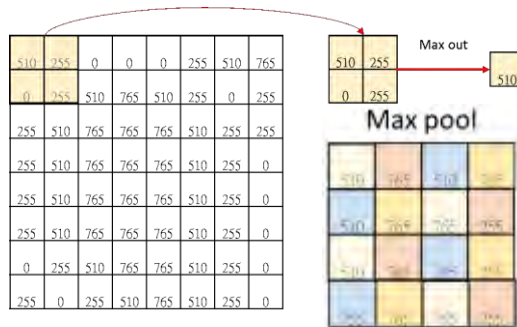


圖18. CNN 架構示意圖

a. 卷積層：先將圖片二值化，為了保持圖片輸出與原有大小相同，在圖片的最外圍補上 0，之後依序取輸入與權重大小相同的矩陣，進行矩陣相乘運算，輸出特徵圖。特徵圖中將保留輸入圖片明顯的特徵。

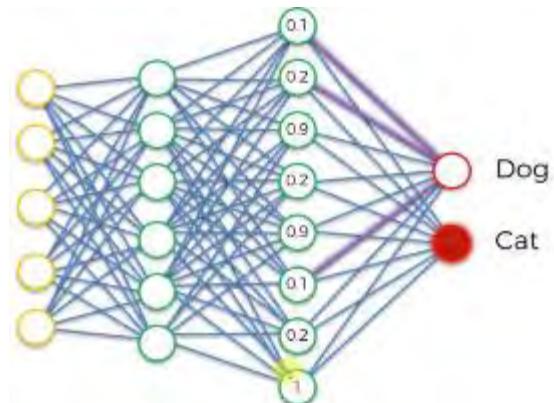
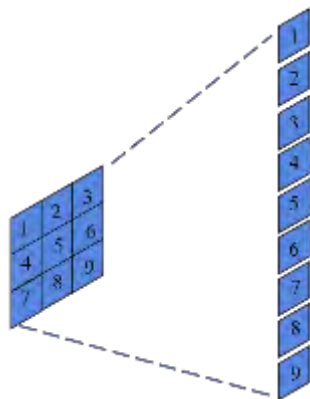


- b. 池化層：池化目的在於減少資料量、保留明顯特徵、減少過擬合（overfitting）的發生，將圖片分成多個小塊，保留每塊的最大值輸出，輸出將使特徵圖降維。



- c. 平坦層：為了連接全連接層，將特徵圖轉為一維陣列。

- d. 全連接層與輸出層：將平坦後的結果連接至神經網路，進行運算並輸出可以得到辨識的權重。



(2) 深度學習模型：常見的深度學習模型大致上分為兩種，運用預先篩選（Region Proposal）的 RCNN(Regions with CNN features)與運用迴歸分析(Regression Analysis)的 YOLO（You only look once），經過表 3 比較兩種類型的模型，最終採用 YOLO 模型，因 YOLO 比起 RCNN 辨識更快速且辨識準確度並不差，使用的效能較少，因此較適合使用 YOLO 模型進行辨識。

表3. RCNN 與 YOLO 比較 來源：組員整理

比較項目	RCNN	YOLO
速度	慢	快
準確率	中	中
訓練時間	中	多
同時偵測物品的數量	多	中
效能使用	多	少
演算法原理	分割出幾千塊可能是物品的範圍，每個範圍進行多次的 CNN，最後透過線性回歸的方式確定範圍。	將物體分類與定位的預測框直接輸出，運用機率的方式篩選可能的區塊，最後以 NMS 與 IOU 框出物品位置。

(3) YOLO 模型：在 YOLO 模型中只有卷積層與池化層，移出了全連接層，能夠預訓練不同大小的圖片。先是將圖片分成 S 乘 S 方格，判斷結果包含類別、位置與機率。計算邊界框對於每一個類別出現的可能性，機率高的方格會以中心每小格用 5 個不同尺寸的 achors box 進行預測。

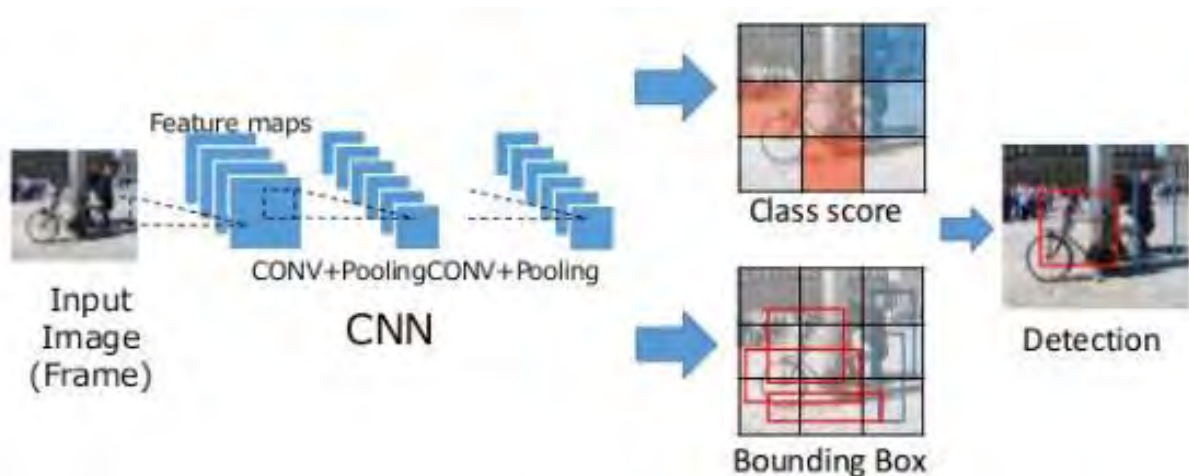


圖23. YOLO 方法示意圖

(4) 預測框：對應格子 Cell，左上角的距離為 (C_x, C_y) ， σ 定義為 sigmoid 啟動函數，處理後，將函數值約束在 0~1 的範圍，讓預測中心在 Cell 中，不會偏離 cell。預測框 Anchor 的寬高為 (P_w, P_h) ，預測位置 (b_w, b_h) 是相對於 Anchor 的寬高，並乘上係數。

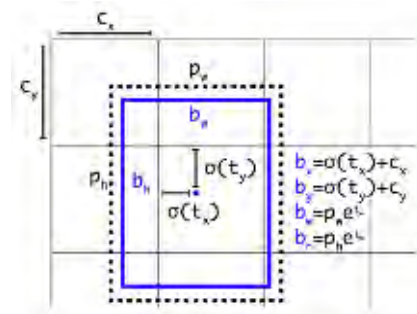


圖24. YOLO 預測框計算

(5) 訓練模型的方式：將幾百張物品經過標記的圖片，丟給電腦去學習不同物品的特徵，讓電腦以減少誤差為目標去修改權重，達到更準確的辨識物品。誤差主要分析的是邊界框的內容，包含五個預測值，中心座標 (x,y) 、長寬 (w,h) 和準確度 IoU（檢測評價函數 Intersection over Union）。

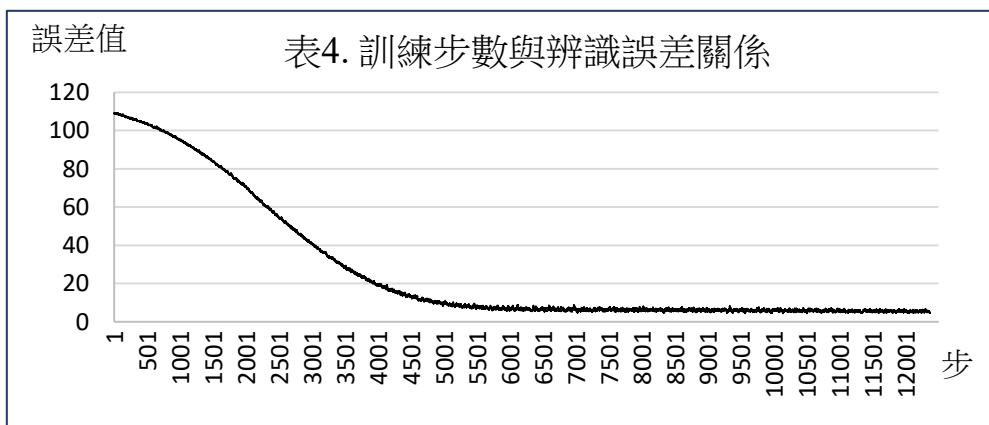
$$\begin{aligned}
 loss_t = & \sum_{l=0}^W \sum_{j=0}^H \sum_{k=0}^A 1_{\text{Max IOU} < \text{Thresh}} \lambda_{\text{obj}} * (-b_{ijk}^x)^2 \\
 & + 1_{t < 12800} \lambda_{\text{prior}} * \sum_{r \in (x,y,w,h)} (\text{prior}_k^r - b_{ijk}^r)^2 \\
 & + 1_k^{\text{truth}} (\lambda_{\text{coord}} + \sum_{r \in (x,y,w,h)} (\text{truth}^r - b_{ijk}^r)^2 \\
 & + \lambda_{\text{obj}} * (\text{IOU}_{\text{truth}}^k - b_{ijk}^k)^2 \\
 & + \lambda_{\text{class}} + (\sum_{c=1}^C (\text{truth}^c - b_{ijk}^c)^2)
 \end{aligned}$$

$$\text{IoU} = \frac{\text{預測範圍} \cap \text{標記範圍}}{\text{預測範圍} \cup \text{標記範圍}}$$

公式2. IoU 計算

公式1. YOLO 辨識誤差數學式

下表是在做模型訓練時，誤差與訓練筆數的關係，可知在訓練時隨著訓練筆數增加，辨識誤差會逐漸下降，辨識也會越精準。



(二) 記錄

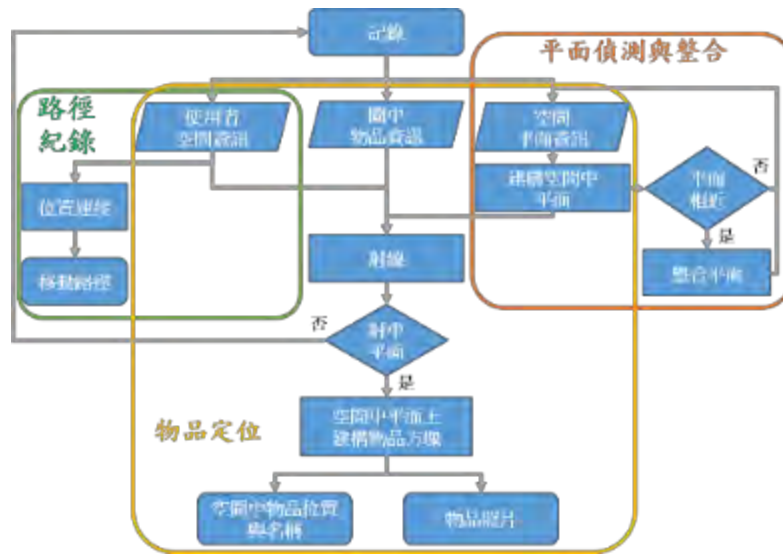


圖25. 記錄流程圖

1. 平面偵測與整合

使用辨識所得到的空間平面資訊，將其建構於 SLAM 系統的地圖中，因此地圖中會有多個高低大小不一的平面，藉由判斷新辨識到的平面位置與已建構平面位置相距小於 5 公分，將兩平面以新平面為基準，合併為一個大平面，如果平面位置差距較大則會保留，以便區分出高低不同的桌面、地板或櫃子的平面。

2. 路徑記錄

將辨識到的使用者空間資訊，先後位置點連接，用以完整記錄使用者在空間中的移動軌跡。

3. 物品定位

(1) 射線：將圖片中辨識到的物品範圍，依最低位置的中央為射線的起點，眼鏡的旋轉角度為方向，在虛擬空間中射出一條射線，當射線碰到已建構出的平面，就將此位置定為物品於空間中的位置。

(2) 空間中平面上建構物品方塊：以物品於空間中的位置為方塊底面中心，在虛擬空間建構一個十立方公分的方塊，以此表示物品的範圍。

(3) 新增：當第一次看見這個物品時，經過射線與空間中平面上建構物品方塊後，添加一個虛擬方塊，虛擬方塊內包含物品的名稱狀態等資訊。

(4) 更新：當第二次看到此物品時，再新增一個新的虛擬方塊，因為物品位置不變，所以兩次虛擬方塊添加的座標系位置都會相同，因此當名稱相同的虛擬方塊在座標系添加後發生重疊碰撞，就表示第二次看到此物品時位置還是與第一次位置相同，不論第二次看物品的視角是否與第一次的視角相同，都能因為添加虛擬方塊在座標系上的位置相同而確定物品位置。

4. 物品照片

根據物品定位後的位置結果，在虛擬方塊上方添加一張最後辨識到此物品的照片，讓使用者在尋找物品時，能看到最後拍攝的影像，用於幫助使用者回想當時狀況。

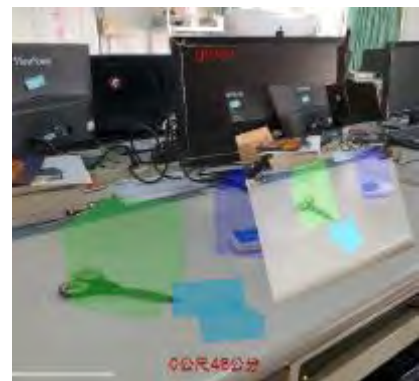


圖26. 物品照片圖示

(三) 搜尋

1. 主動搜尋

主動提供選定物品的資訊，使用耳機的選擇，得知使用者想尋找的物品，並用 SLAM 系統及平面檢測的結果，搜尋物品與自身的距離及方向，以及最後記錄的影像，讓使用者選擇搜尋結果，選定後能給予到達物品的導航。

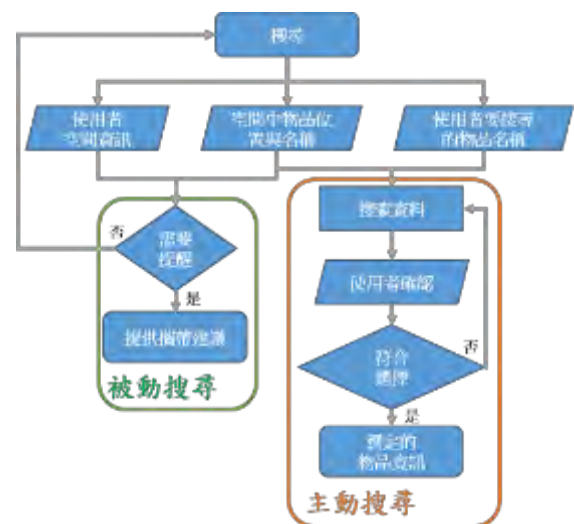


圖27. 搜尋流程圖

2. 被動搜尋

本模式提供攜帶者建議的方式呈現，蒐集使用者例行攜帶的物品，或是當下時間、天氣等資訊，以提供物品攜帶和整理建議等功能，讓使用者能因提醒而選擇要尋找的物品。

(四) 導航

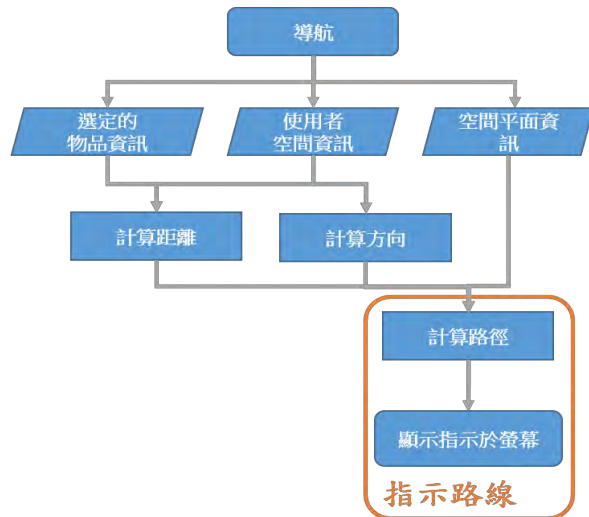
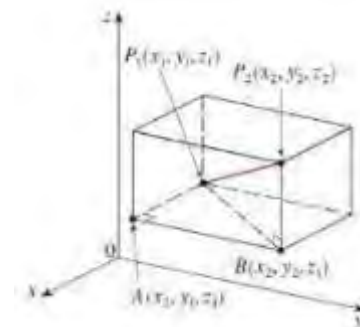


圖28. 導航流程圖

1. 計算距離

依具虛擬方塊與使用者在座標系中的兩點，用畢氏定理計算兩點在空間中的距離，並在視角上方即時顯示使用者與物品的三維距離。



$$|P_1P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

圖29. 計算距離方法

2. 計算方向

用三角函數求得座標系中物品相對於使用者的角度，加上使用者的偏轉視角計算出指示箭頭的旋轉角度。

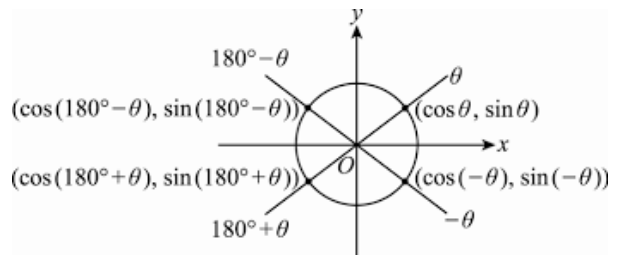


圖30. 計算方向方法

3. 指示路線

從已選定要尋找的物品資料中，計算且判斷最快能到達的位置，將直線距離與旋轉後的指示箭頭顯示在畫面中，讓使用者能依據指示與導航尋找到符合搜尋結果且距離最近的物品位置。

伍、研究結果

一、SLAM 空間辨識

(一) 特徵點取得

取得影像中灰階後顏色相差較大的點或明顯邊緣點作為特徵點。如：圖 31



圖31.特徵點(黃色圓點)與辨識平面

(二) 平面紀錄與擴充

依據特徵點的分布繪製平面，並整合多個相近平面，達到擴充與彙整空間資訊。如：圖 32



之前



之後



圖32.平面擴充過程

二、YOLO 物體偵測

(一) 辨識物品種類

將拍攝到的影像經過 YOLO 物體偵測模型辨識物品種類與可信度，只顯示可信度高於 50 的物品。如：圖 33

(二) 框出辨識物範圍

以模型辨識到的兩點位置，繪製辨識方框，以確定物品於視角中的範圍。如：圖 33



圖33.YOLO 辨識結果

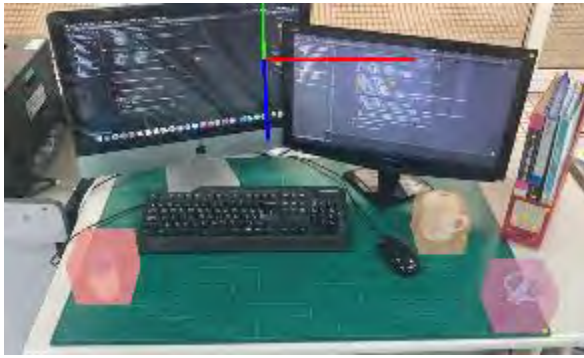
三、結合空間與影像辨識

(一) 建構地圖

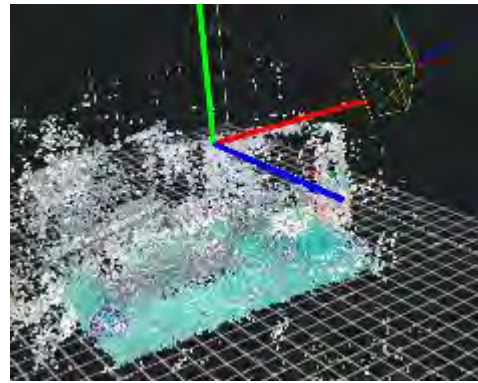
能夠依據特徵點在空間中的分布，建構空間地圖，如：圖 34

(二) 標記物品位置

定位辨識到的物品於空間中的位置，標記於地圖座標位置上，以紀錄物品的位置。如：圖 34



視角所見影像



特徵點建構空間地圖

圖34.空間辨識地圖

四、製作 AR 空間記憶眼鏡

(一) 製作 APP

以手機 APP 的方式撰寫程式。如：圖 35

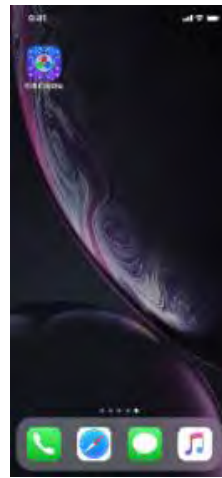


圖35.安裝 APP

(二) 自製 AR 空間記憶眼鏡

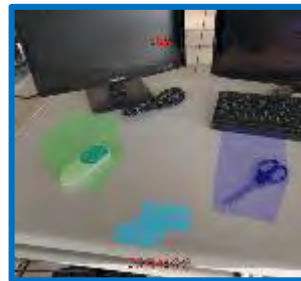
結合手機與 VR 眼鏡模擬 AR 眼鏡，使用相機影像與 IMU 資訊達到紀錄空間、定位自身，以螢幕顯示資訊來達成 AR 眼鏡功能。如：圖 36



圖36.使用者攜帶樣式

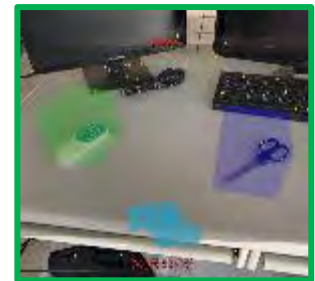
(三) 物品搜尋選擇

使用音量鈕選擇不同物品名稱，搜尋選擇的物品位置並給予導航。如：圖 37



(箭頭指向藍色方塊)

搜尋剪刀



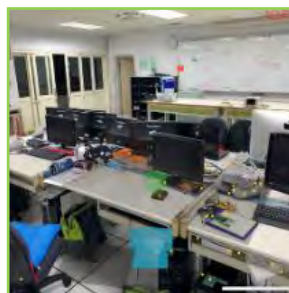
(箭頭指向綠色方塊)

搜尋遙控器

圖37.指示過程

(四) 導航物品位置

依據自身與物品的位置，箭頭與距離標示都可以正確指引使用者到達物品位置。如：圖 38



近距離面對物品



中距離背對物品



遠距離面對物品

圖38.導航過程

實驗一 實際距離與預估距離誤差

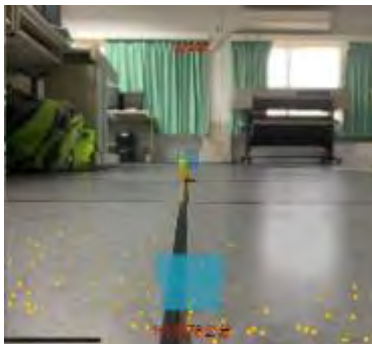
實際 60 公分
測得 58 公分



實際 120 公分
測得 118 公分



實際 180 公分
測得 178 公分



實際 240 公分
測得 237 公分



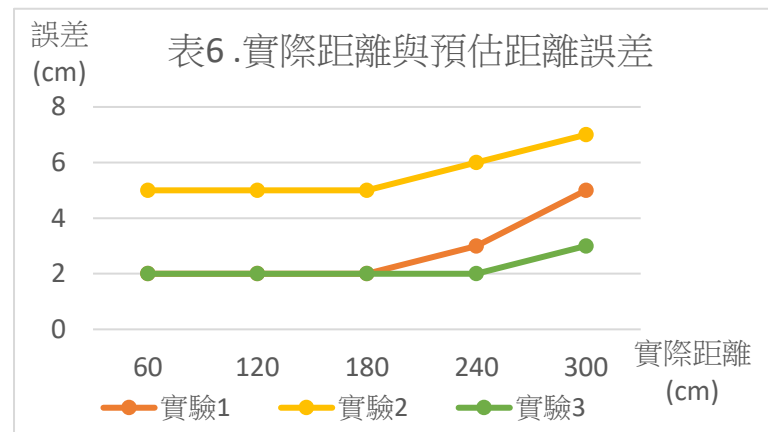
實際 300 公分
測得 295 公分



圖39.不同距離測試圖片

表5. 移動與距離誤差

實際距離 (公分)	預估距離 (公分)	差距 (公分)	誤差增量 (公分)	總誤差 (公分)
實驗 1				
60	58	2	0	0
120	118	2	0	0
180	178	2	0	0
240	237	3	1	1
300	295	5	2	3
實驗 2				
60	55	5	0	0
120	115	5	0	0
180	175	5	1	0
240	234	6	1	1
300	293	7	1	2
實驗 3				
60	58	2	0	0
120	118	2	0	0
180	178	2	0	0
240	238	2	0	0
300	297	3	1	1



實驗過程與結果

以每塊 60 公分的地板為基準，測量真實距離與程式預估距離的誤差，距離 200 公分內都保持相同誤差，此誤差是因為定位點與測量點的偏差，但到了 200 公分後，誤差明顯有增加，表示超過 200 公分後系統才會產生大於 1 公分的測量誤差。

實驗二 移動時物品辨識能力

表7. 物品不同方向
最大可辨識距離

方向(度)	距離(公分)
0	165
45	187
90	173
135	130
180	150
225	162
270	157
315	163
辨識：可信度高於 50 物品離地 70(公分) 相機離地 150(公分)	

表8. 辨識距離相等
不同方向的可信度

距離(公分)	方向(度)	可信度(%)
50	0	86.1
	45	92.0
	90	89.2
	135	86.8
	180	87.4
100	0	71.4
	45	66.8
	90	75.1
	135	66.9
	180	74.7



距離：1 公尺
辨識：是



距離：1.5 公尺
辨識：是



距離：2 公尺
辨識：否

圖40.不同距離辨識

分別測試不同方向與距離，不同物品的辨識準確度，經過實驗得知在不同方向辨識且距離 1.5 公尺以內，多數情況下辨識可信度皆高於 50。

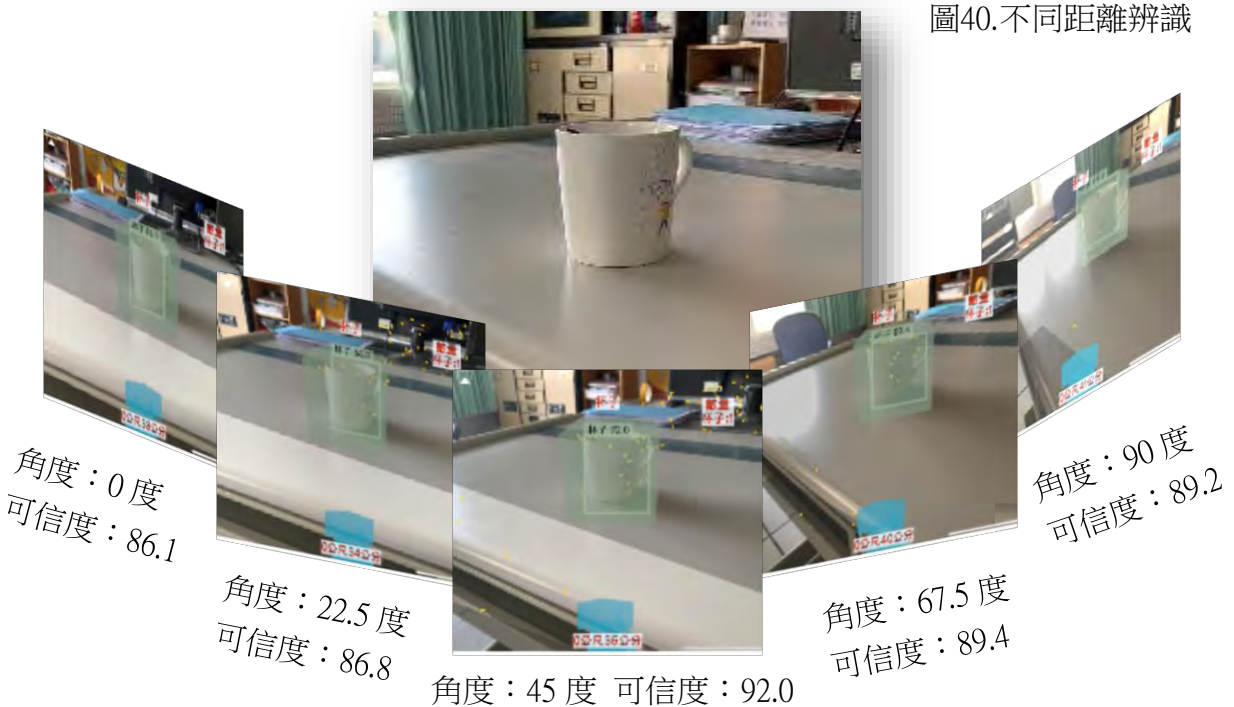


圖41.不同方向辨識

實驗三 測試遠離再回到定位點的偏移誤差

表10.移動過程紀錄演示

表 9.移動遠近與定位誤差

直線長度 (公分)	誤差距離 (公分)	每公分誤差 百分比(%)
480	2	0.42
	3	0.63
	2	0.42
	1	0.21
	1	0.21
	3	0.63
	2	0.42
	1	0.21
	3	0.63
1	0.21	
總平均誤差	1.9	0.33
600	2	0.50
	3	0.33
	2	0.50
	3	0.50
	3	0.17
	1	0.17
	1	0.33
	2	0.33
	2	0.17
1	0.42	
總平均誤差	2	0.63

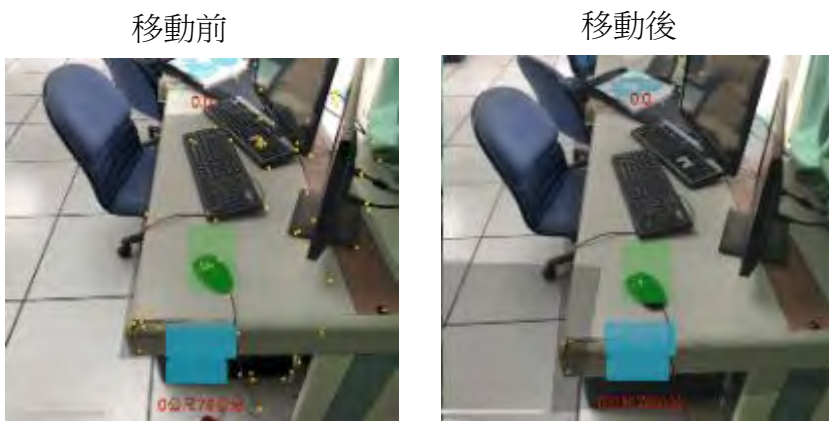
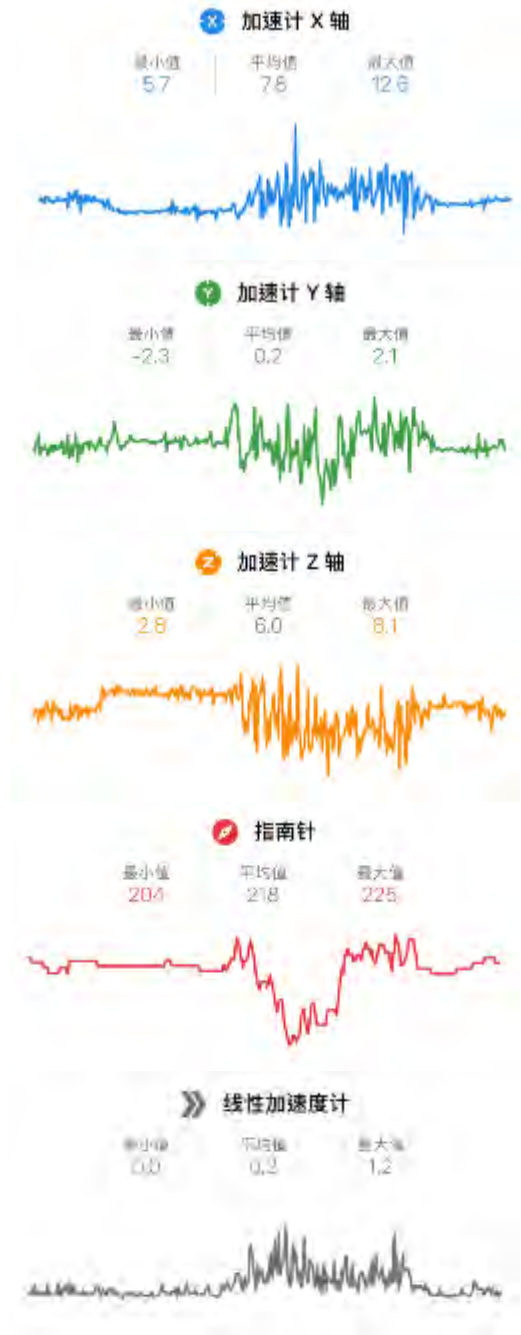


圖42.移動前與移動後定位誤差演示圖

實驗過程與結果

測試遠離再回到定位點後，定位點偏移的距離，以距離定位點 480 公分與 600 公分，分別測試 10 次，平均結果得出，約產生 0.35%的誤差。

陸、討論

- 一、運用 VR 眼鏡與手機模擬 AR 眼鏡的方法，可以讓使用看到真實與虛擬的資訊，此方法所存在的問題同樣是現有 AR 眼鏡主要的問題，但在價格與相容性上都有較大的優勢。
- 二、以 SLAM 系統辨識環境，YOLO 方法辨識物品達到紀錄物品位置的成果，雖然準確度與穩定性還有改善空間，但確實能達到移動時定位物品位置。
- 三、使用者能透過程式，方便的記錄與得知物品的擺放位置，並且能根據導航的指引到達物品放置的位置。
- 四、YOLO 方法在現有手機上還無法做到如電腦上執行一樣即時且準確的物體偵測，因為手機的效能達不到準確即時資訊處理能力。

柒、結論

本作品利用 SLAM、YOLO 等技術，建構一個可做到環境建構、記憶物品位置與指示物品資訊的 AR 記憶眼鏡，達到輔助人腦記憶物品，完成的功能如下：

- 一、運用 SLAM 達到辨識環境、YOLO 達到物品辨識，完成空間中定位物品的位置。
- 二、使用者能透過選擇物品，檢索到物品所在位置。
- 三、使用者能根據 AR 的指示資訊，正確且快速尋找到物品所在位置。

未來展望：近期目標是優化影像辨識，達到更方便且實用，遠期目標將整理不同使用者的辨識數據，達到使用者共用辨識資料，實現物品空間資訊的互通與傳遞，目標規劃如下：

- 一、近期：修改與優化影像辨識模型，達到更快速且精確的辨識物品。
- 二、遠期：將不同使用者的辨識數據透過網路儲存於資料庫，做到整合不同辨識資料。

發展空間：此作品可運用的範圍除了個人在家中尋找忘記的物品外，還可用於工廠管理與清查貨品，管理人員用看的就可以準確的清點貨品，不需要再翻單子打勾；另外機場巡邏的航警也可運用此作品幫助記憶與比對放置長久的物品，查找疑似爆裂物的物品；其他領域如：人與機器人的互動、現實物品空間資訊的傳遞與分析等，都有很多能應用的方向。

捌、參考資料及其他

一、論文書籍：

江翊璋 (2018)。基於 SLAM 之擴增實境技術於史蹟導覽之應用與創作。國立臺北科技大學互動設計系碩士論文，臺北市。取自 <https://hdl.handle.net/11296/76v76n>

高翔,張濤 (2017)。視覺 SLAM 十四講：從理論到實踐。出版商：電子工業出版社

陳玉琨,湯曉鷗 (主編) (2018)。人工智慧基礎 (高中版)。出版商：華東師範大學出版社

葉暘、何政勳 (2017)。用於機器人空間建模的仿生認知系統。臺灣國際科展

David G. Lowe, **Distinctive image features from scale-invariant keypoints**. In IJCV, 2004.

Ian Goodfellow, Yoshua Bengio, & Aaron Courville (2016) . **Deep Learning**. MIT Press.

J. Redmon, A. Farhadi. **YOLO9000: Better, Faster, Stronger**. In CVPR, 2016.

J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi. **You only look once: Unified, real-time object detection**. In CVPR, 2016.

Ole Kiehn & Hans Forssberg (2014) , **The Brain’ s Navigational Place and Grid Cell System**.

Sudeep Pillai and John J. Leonard (2015) ,**Monocular SLAM Supported Object Recognition**.

二、網路資源：

Apple Develop Document, **ARKit**, 引用自 <https://developer.apple.com/documentation/arkit>

Apple Develop Document, **Working with Core ML Models**,

引用自 <https://developer.apple.com/machine-learning/build-run-models/>

Steven Shen, **YOLO with CoreML**,

引用自 <https://medium.com/@syshen/yolo-with-coreml-819799789c11>

Vlas Voloshin,**Visualizing ARKit Sessions**,

引用自 <https://www.ittybittyapps.com/blog/posts/2018/12/visualizing-arkit-sessions/>

三、圖片來源：

圖 1：《How It Works 知識大圖解》 2015 年 9 月號第 12 期

圖 2：<https://case.ntu.edu.tw/blog/?p=19357>

圖 3：http://scimonth.blogspot.com/2014/12/blog-post_9.html

- 圖 4 : <https://zh.wikipedia.org/wiki/Google%E7%9C%BC%E9%95%9C>
- 圖 5 : <https://www.idc.com/getdoc.jsp?containerId=prUS44001618>
- 圖 12 : <https://code.fb.com/ml-applications/segmenting-and-refining-images-with-sharpmask>
- 圖 13 : <https://www.jianshu.com/p/7faa4a3af589>
- 圖 15 : <https://arxiv.org/pdf/1708.03852.pdf>
- 圖 17 : <https://goo.gl/TRBqpQ>
- 圖 18 : <https://tinyurl.com/ybshnf2j>
- 圖 19、20、21、22 : <https://medium.com/@chih.sheng.huang821>
- 圖 23 : <https://tinyurl.com/y3jxhkom>
- 圖 24、公式 1 : YOLO 論文
- 圖 29 : <https://goo.gl/TNCAUG>
- 圖 30 : <https://goo.gl/4Vg7WT>

【評語】 052509

此作品結合 SLAM 即時定位與地圖構建及 YOLO 即時物體偵測來實現環境物品辨識、定位、導航的 AR 記憶眼鏡。透過 SLAM 平面偵測建構空間中平面資訊，作為 AR 物件座標及物理資訊的依據參考，再經過機器學習訓練後的 YOLO 模型達到辨識物品，結合兩者建立空物理資訊的依據參考。

此應用研究主題很有趣。未來若繼續發展，建議能加入語音介面功能，讓使用者更容易輸入要找尋的物品。

摘要

本研究結合SLAM即時定位與地圖構建系統及YOLO即時物體偵測方法，實現環境物品辨識、定位、導航的AR記憶眼鏡。透過SLAM平面偵測建構空間中平面資訊，作為AR物件座標及物理資訊的依據參考，再經過機械學習訓練後的YOLO模型達到辨識物品，結合兩者建立空間與物品的依存關係。以VR眼鏡結合手機模擬AR眼鏡，採用iOS系統及Swift語言設計程式功能及介面，實現走動過程持續即時辨識空間與物品，可運用音量鈕選定要尋找物品，再透過眼鏡畫面上方向及距離的指示，導航到該物品位置。本作品還可運用在工作場合，做到紀錄與提醒環境的變化，如航警可疑物盤查、公司財務貨物盤點等，成為幫助人腦記憶的輔助裝置。

研究動機

現代社會生活步調快速，事情多且繁雜，做事常常需要一心多用，造成忘東忘西的現象，再加上人口高齡化，老化引發的記憶力衰退、健忘症與失智症人口，也都在不斷的增加。上述症狀者都會有忘記物品放置位置或忘記曾經做過的事情這類問題，或許只需要給予一些提醒就能使其回想起來。如果我們能將所有看過的物品資訊都記錄在電腦中，要尋找物品時，再使用電腦中的搜尋功能來尋找物品，提醒或告知使用者物品位置，就能讓使用者更有效率的去尋找物品。藉由計算機概論介紹的熱門技術，運用程式設計課程學的程式邏輯，思考結合空間定位與影像辨識運用於AR眼鏡，以大腦記憶空間事物的方式，製作幫助人腦記憶的裝置。在選修課程學習演算法的原理，應用程式開發與專題實作課程製作程式內容，課後閱讀書籍論文，實現以手機與VR眼鏡模擬AR眼鏡，並測試實際使用狀況。比較市面上現有記憶物品的解決辦法，將物品裝上藍芽發送裝置，用裝置取得物品位置，或外裝多台攝影機交叉比對影像定位的方式，與本作品的方法相比之下，本作品在記憶、運用的方式上，更符合人性化的使用，應用上更具有實用性，對其他領域都能有很大的貢獻。



尋找空間中的...

搜索中...

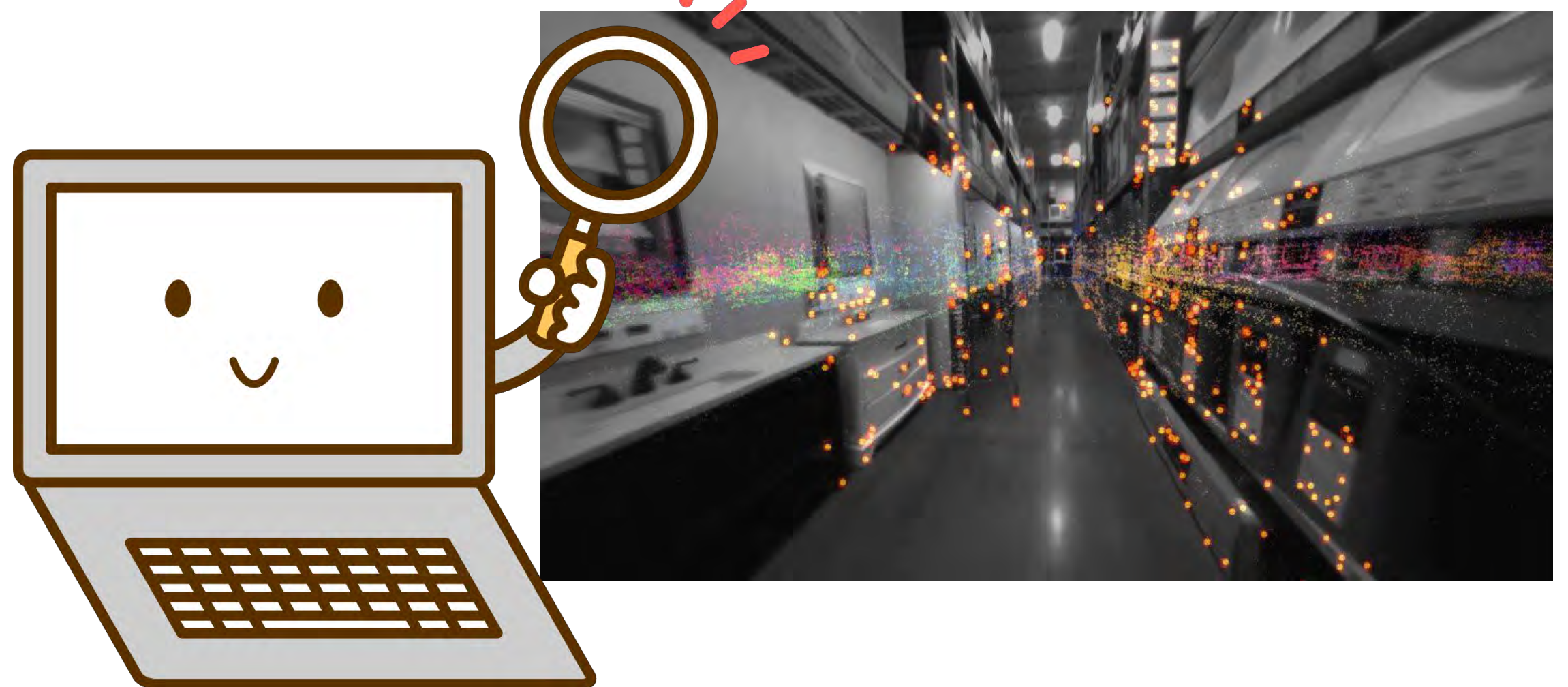


圖1 作品製作動機示意圖

研究目的

- 一、蒐集並瞭解關於人類大腦的記憶與現有AR眼鏡的相關資料。
- 二、研究運用SLAM系統建構三維空間的原理，並用於建構環境座標系，與定位使用者當下的位置與視角。

- 三、研究物體偵測技術，運用YOLO方法，達成準確辨識物品種類、得知物品在圖片中的位置等資訊。
- 四、結合三維空間建模與物體偵測，以達到在座標系中定位與記錄物品資訊。
- 五、撰寫AR眼鏡搜尋與導航系統，讓使用者藉由搜尋與導航功能得知物品位置。

研究過程或方法

作品架構

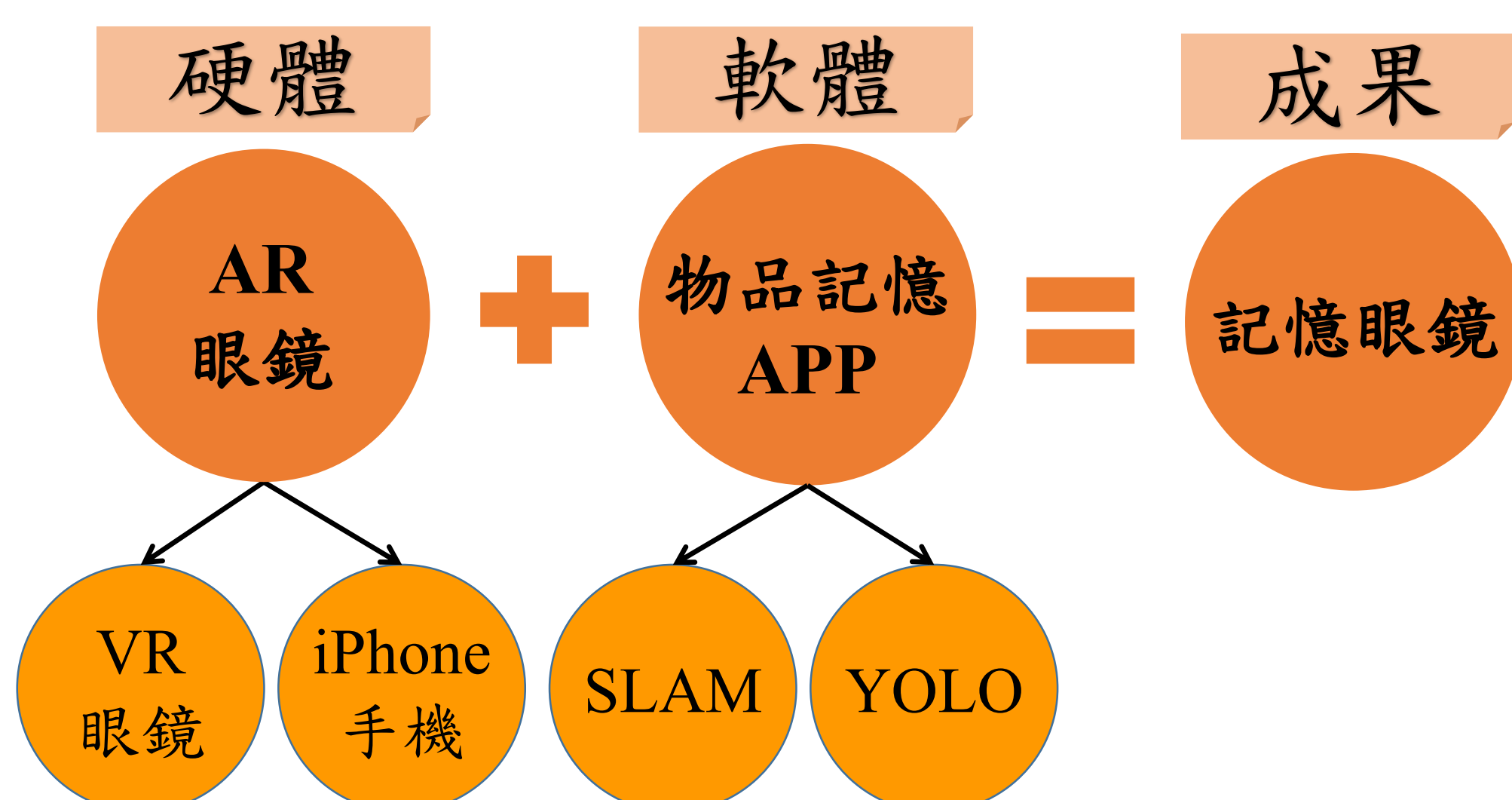


圖2 物品記憶眼鏡的軟硬體區分

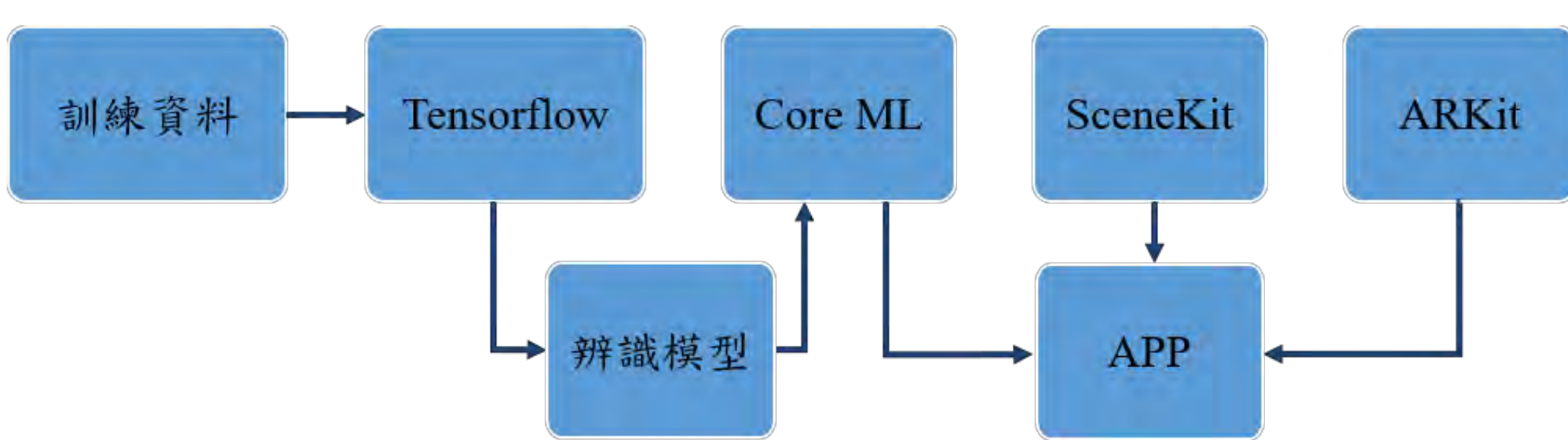
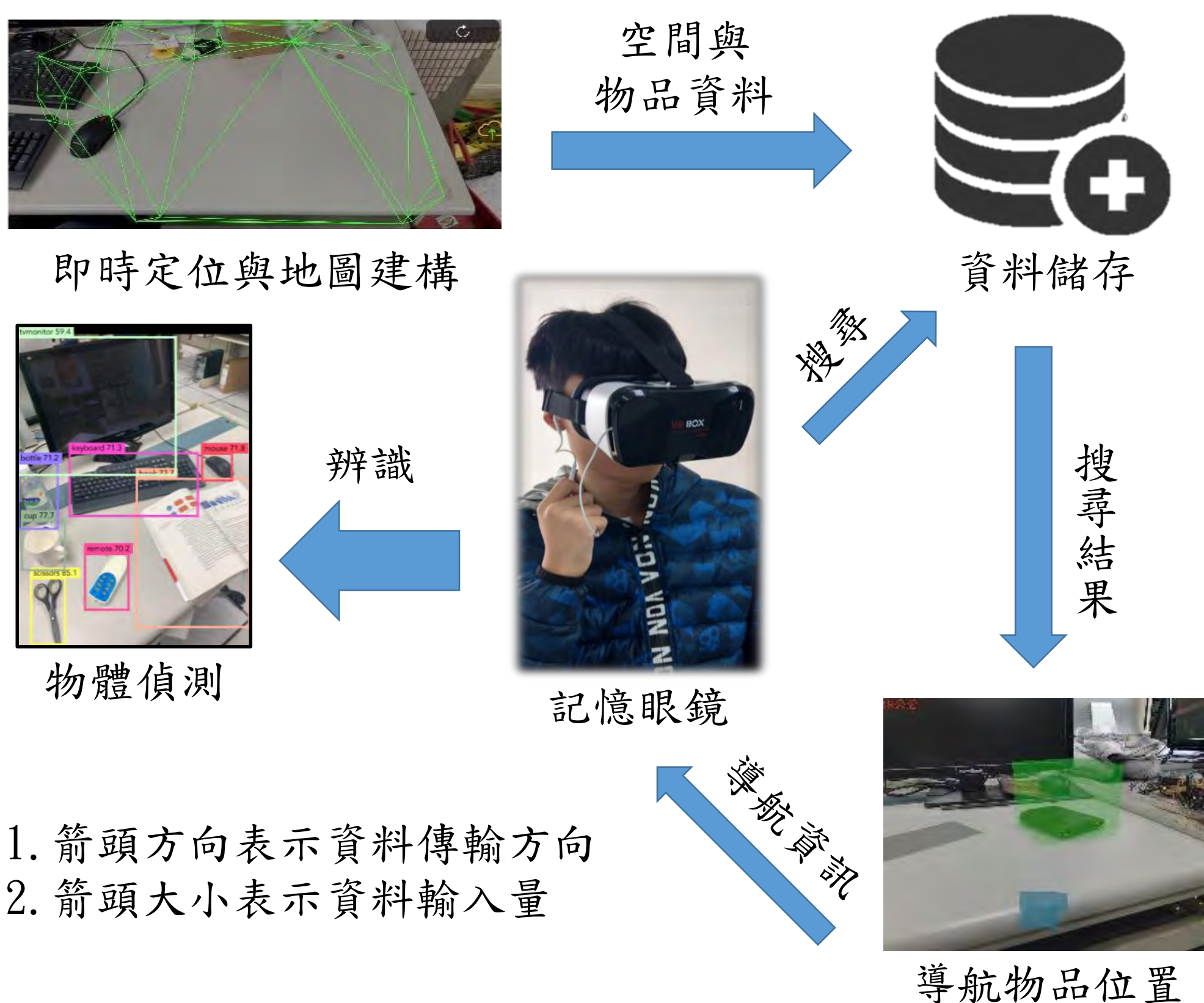


圖3 物品記憶眼鏡的軟體製作流程



1. 箭頭方向表示資料傳輸方向
2. 箭頭大小表示資料輸入量

圖4 物品記憶眼鏡的運作流程圖

記憶探討

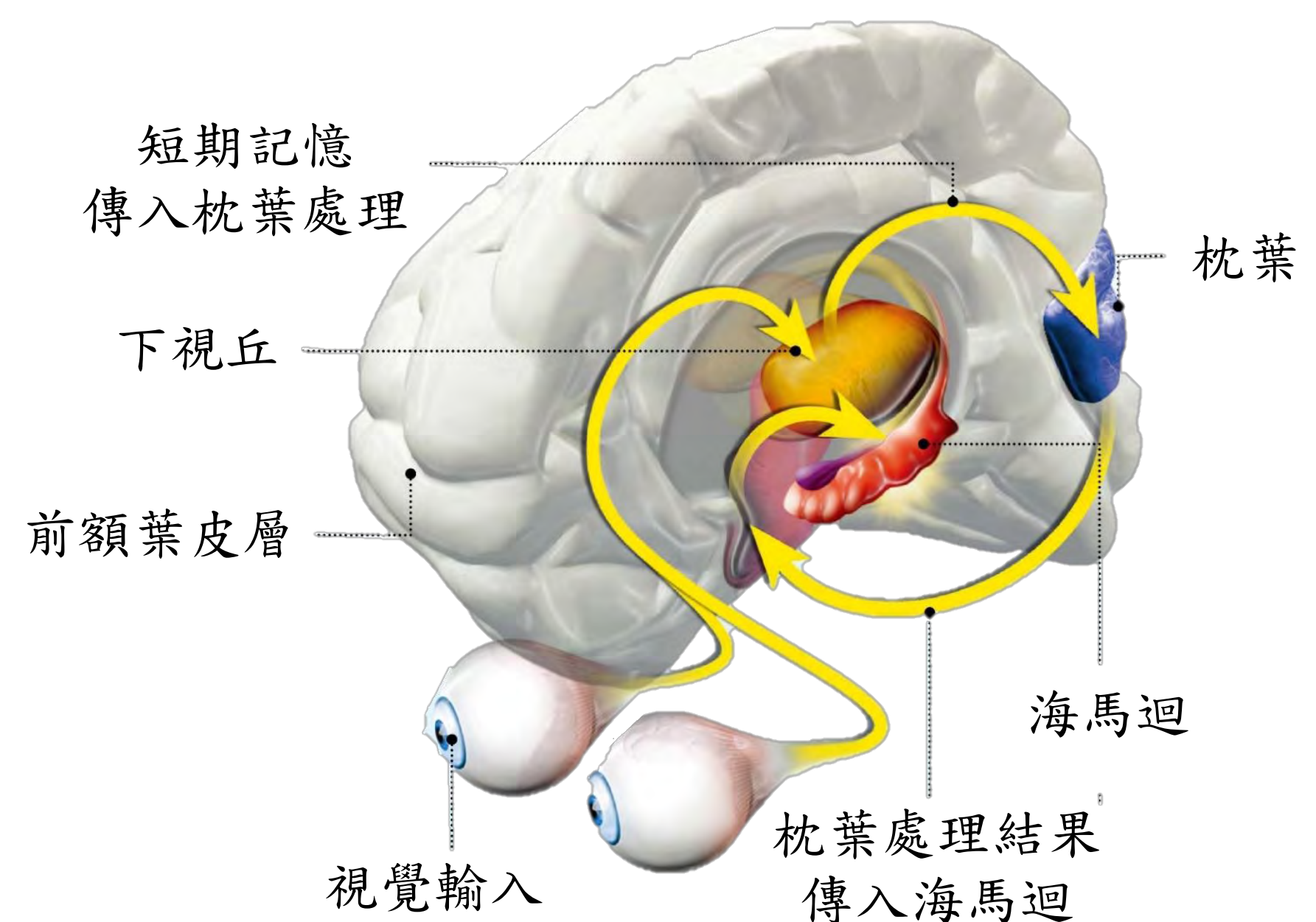
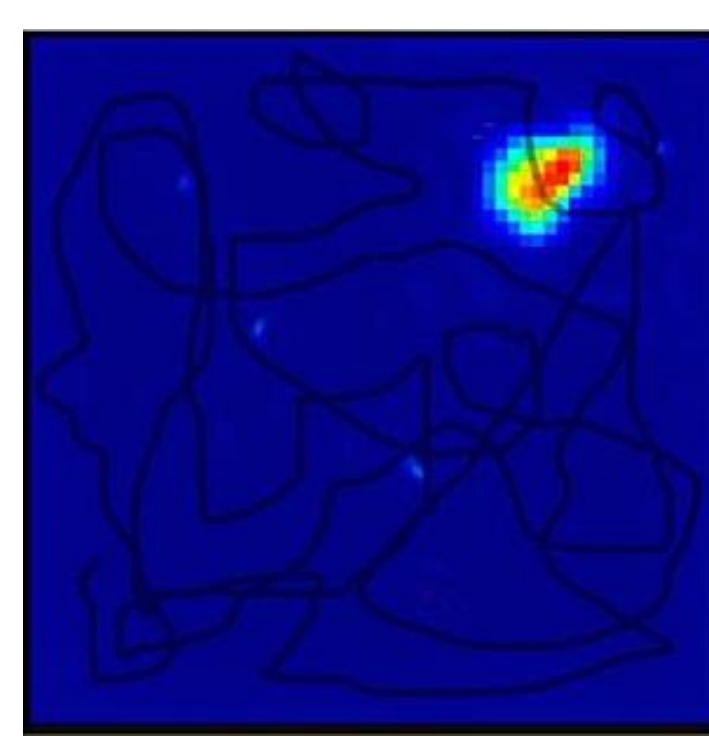
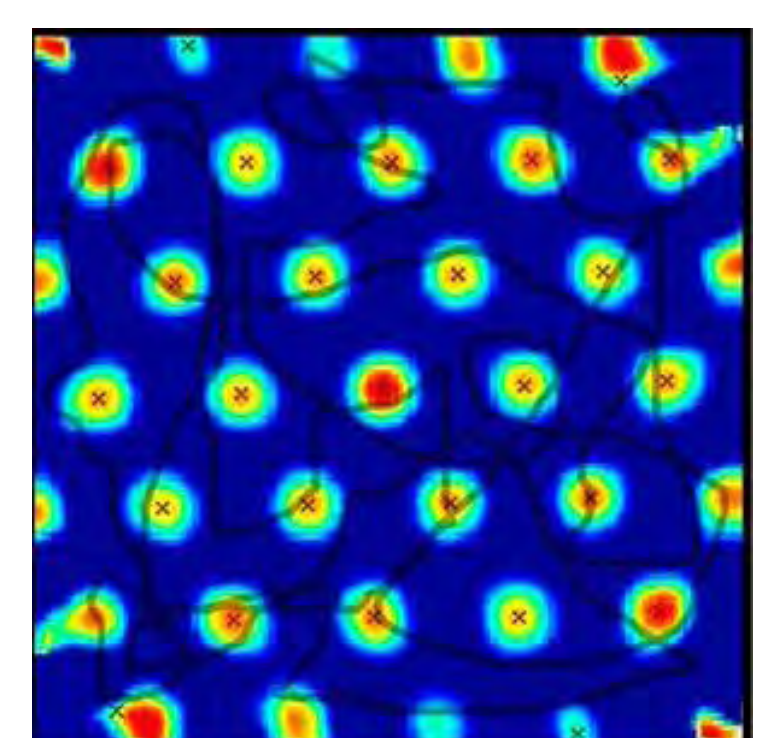


圖5 人類大腦記憶方式

記憶從感覺作為輸入到海馬迴，海馬迴將部分短期記憶轉換為長期記憶儲存在大腦皮層，回憶時從大腦皮層提取記憶。以此原理，製作系統辨識、儲存與搜尋的架構。



位置細胞場域



網格細胞場域

圖6 人類大腦記憶空間方式

人類的環境記憶，在位置細胞中建構認知地圖確定自身位置，以六角形網格的方式在網格細胞中建構座標系統。程式以SLAM定位自身位置與繪製地圖的方式，達到電腦記憶空間。

AR眼鏡探討

近年來許多公司積極研發AR眼鏡，AR眼鏡攜帶方便。且攝影機與使用者視角同步，所以可以取得使用者所看到的影像做資料分析。虛擬資訊可以顯示在鏡片上，讓使用者透過鏡片看現實景象，能看到虛擬影像與現實影像結合後的結果。



圖7 AR眼鏡 Google glasses

2014年Google公司推出的AR眼鏡，可以顯示即時資訊給使用者得知。

研究過程或方法

辨識

辨識空間

辨識空間需要做到確定自身的位置與建構環境地圖，比較SLAM感測方法，運用單目SLAM辨識空間。單目SLAM以鏡頭影像與IMU數據得知自身的位移方向與位移量，用SIFT方法從兩張影像取得自身的影像位移預估，在用卡爾曼濾波整合IMU預估與影像預估，統整出更為準確的運動資訊。

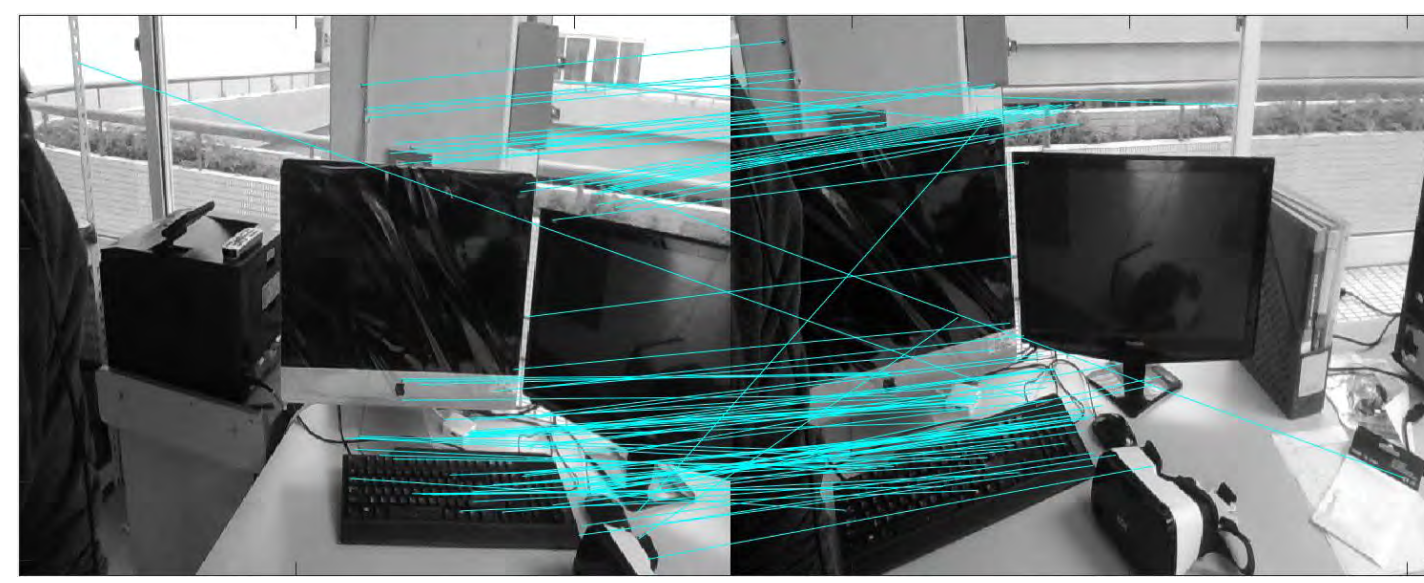
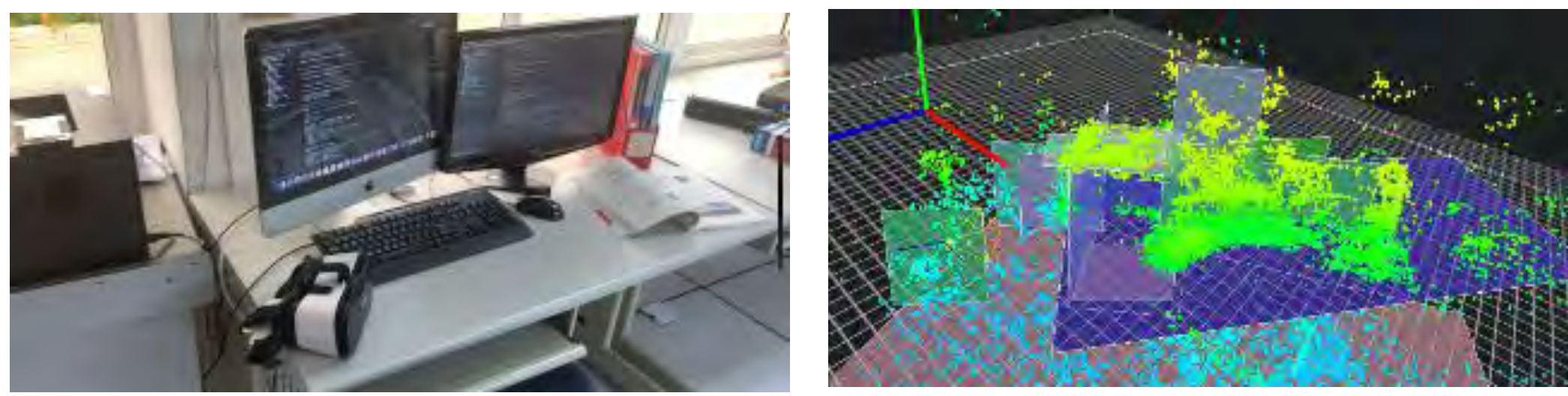
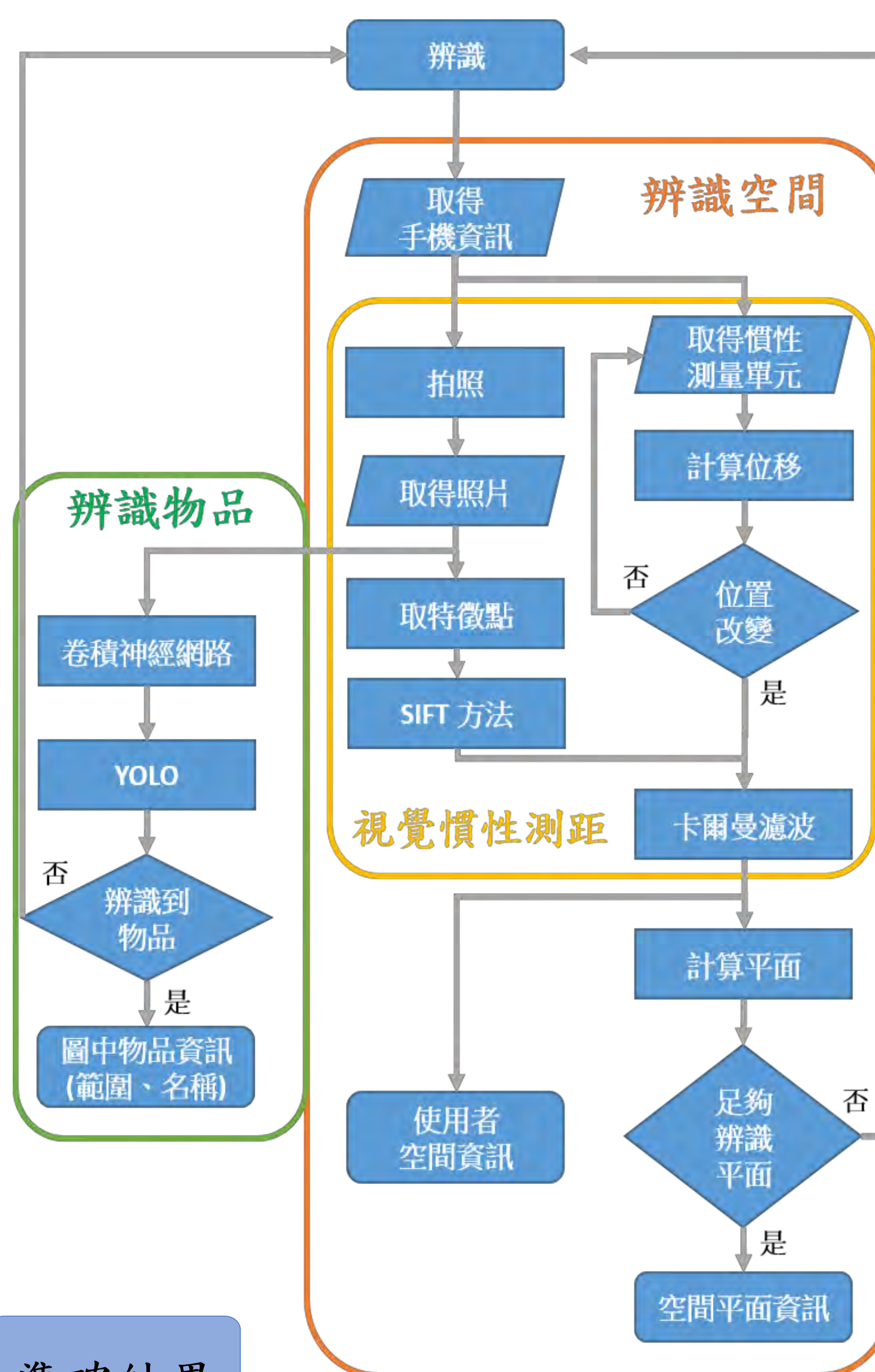
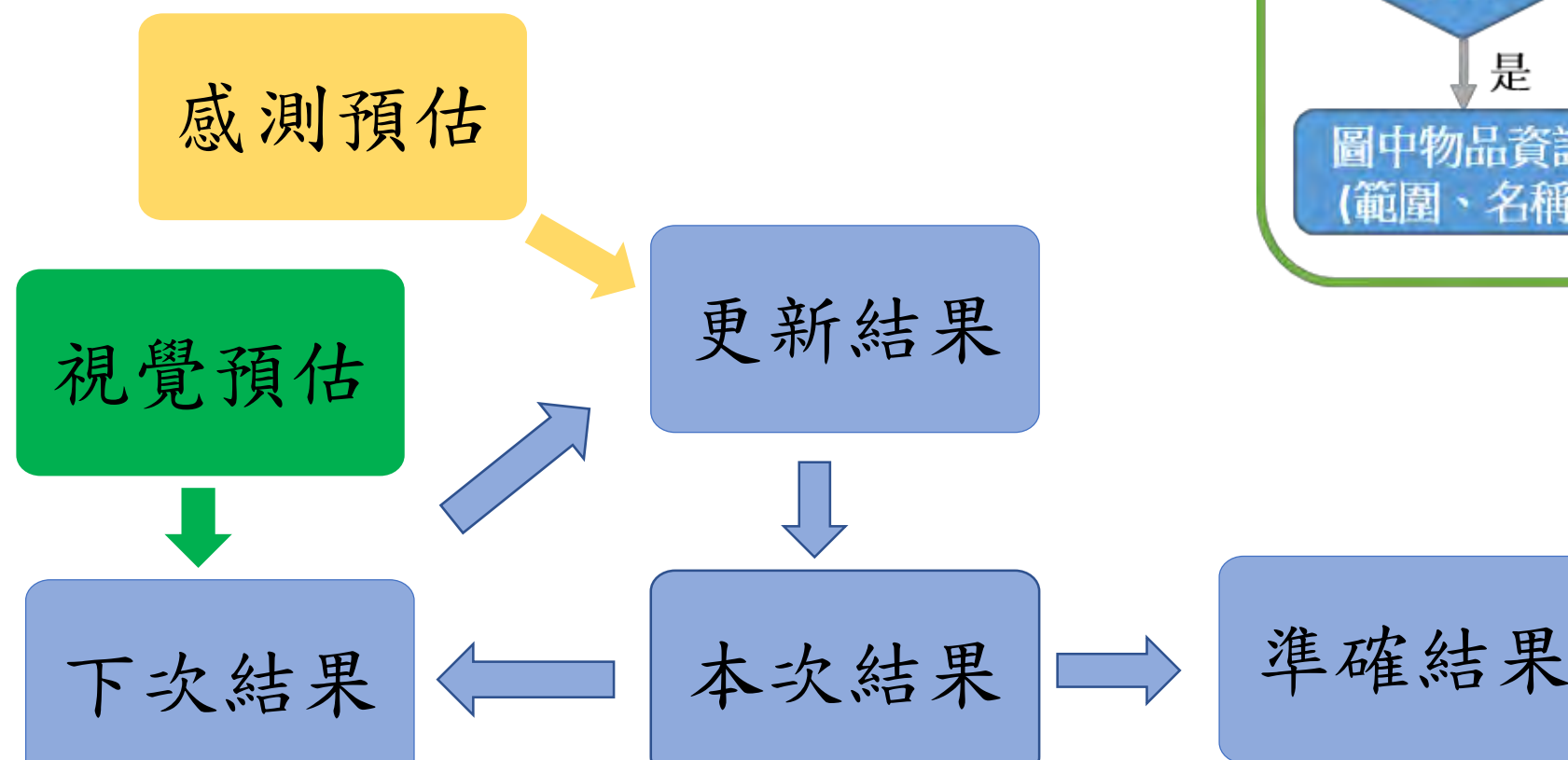


表1 SLAM系統感測器比較表

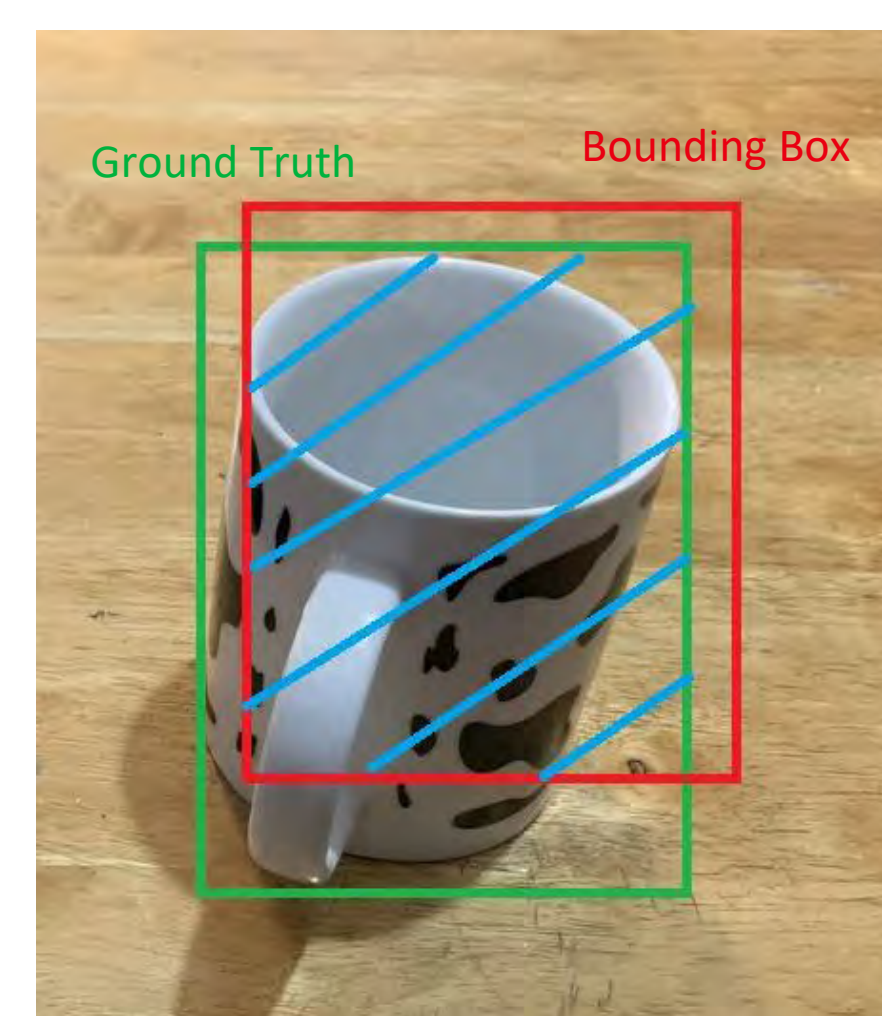
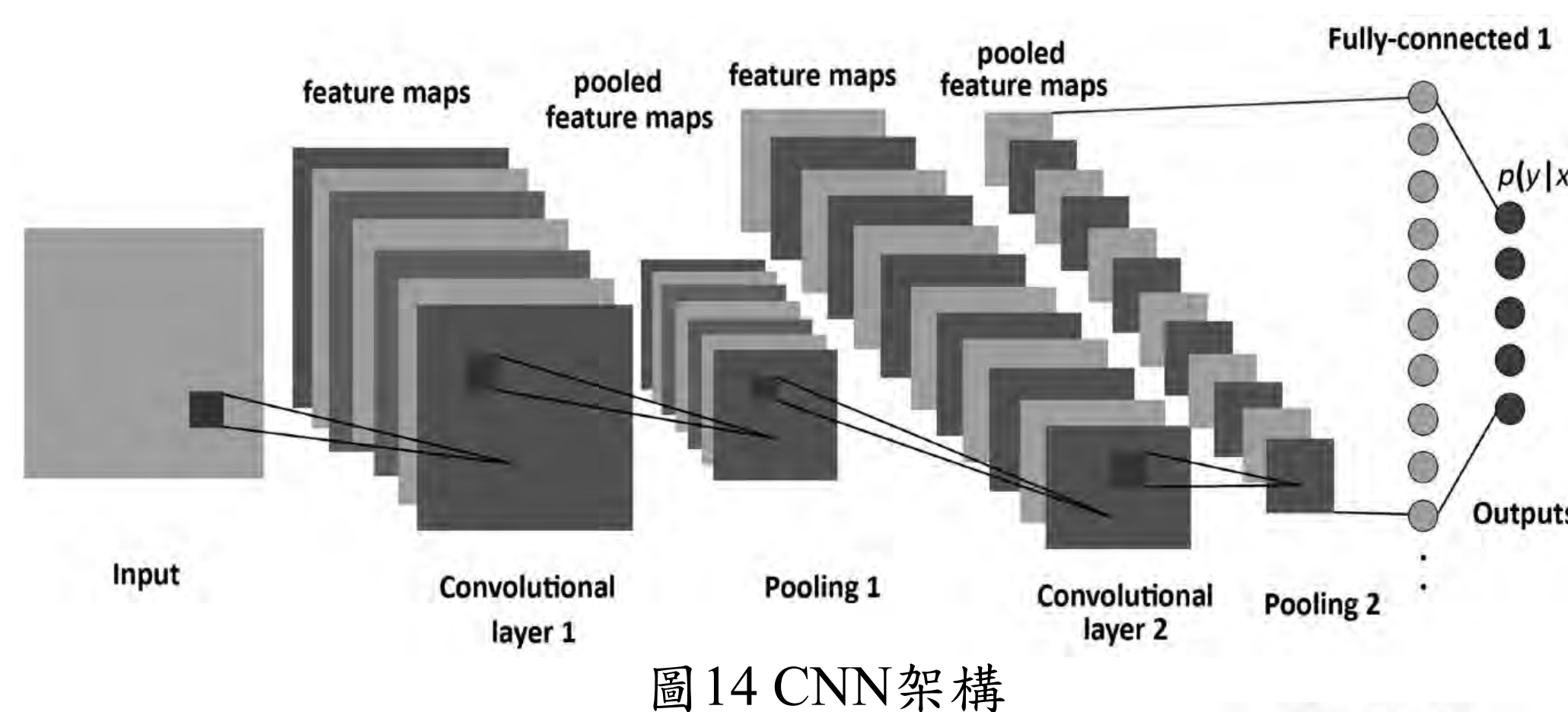
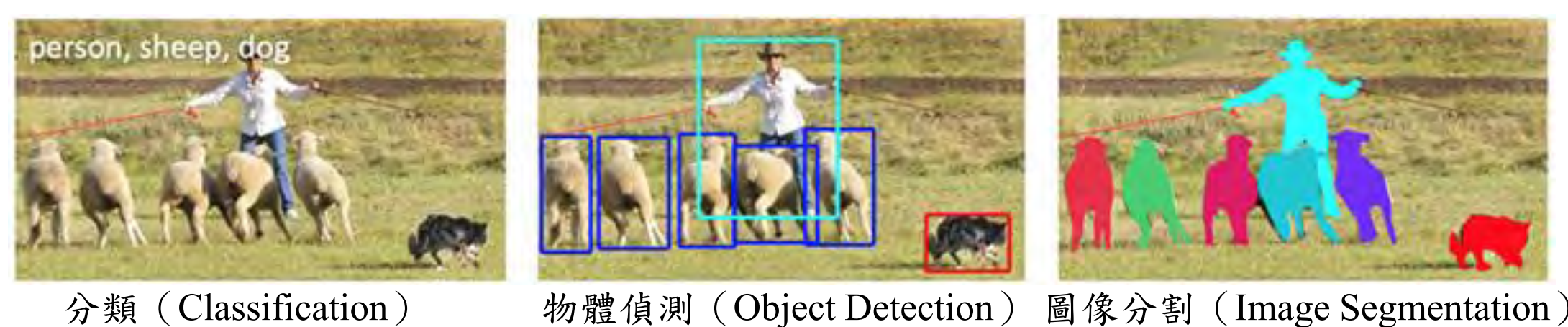
感測器類型	雷射	視覺	
相機	飛時測距 ToF	單目(單鏡頭)	雙目(多鏡頭)
原理	以自身與環境障礙物間的距離作為感測依據	以多張影像與移動訊息作為感測依據	多個相機的單張影像與相機之間的距離作為感測依據
優點	環境影響小，速度快，計算量小	結構簡單，成本低，技術資源多	精準度高，識別環境大
缺點	價格昂貴，缺少影像資訊，技術較少	易受環境影響，缺少深度資訊	高計算資源，速度慢



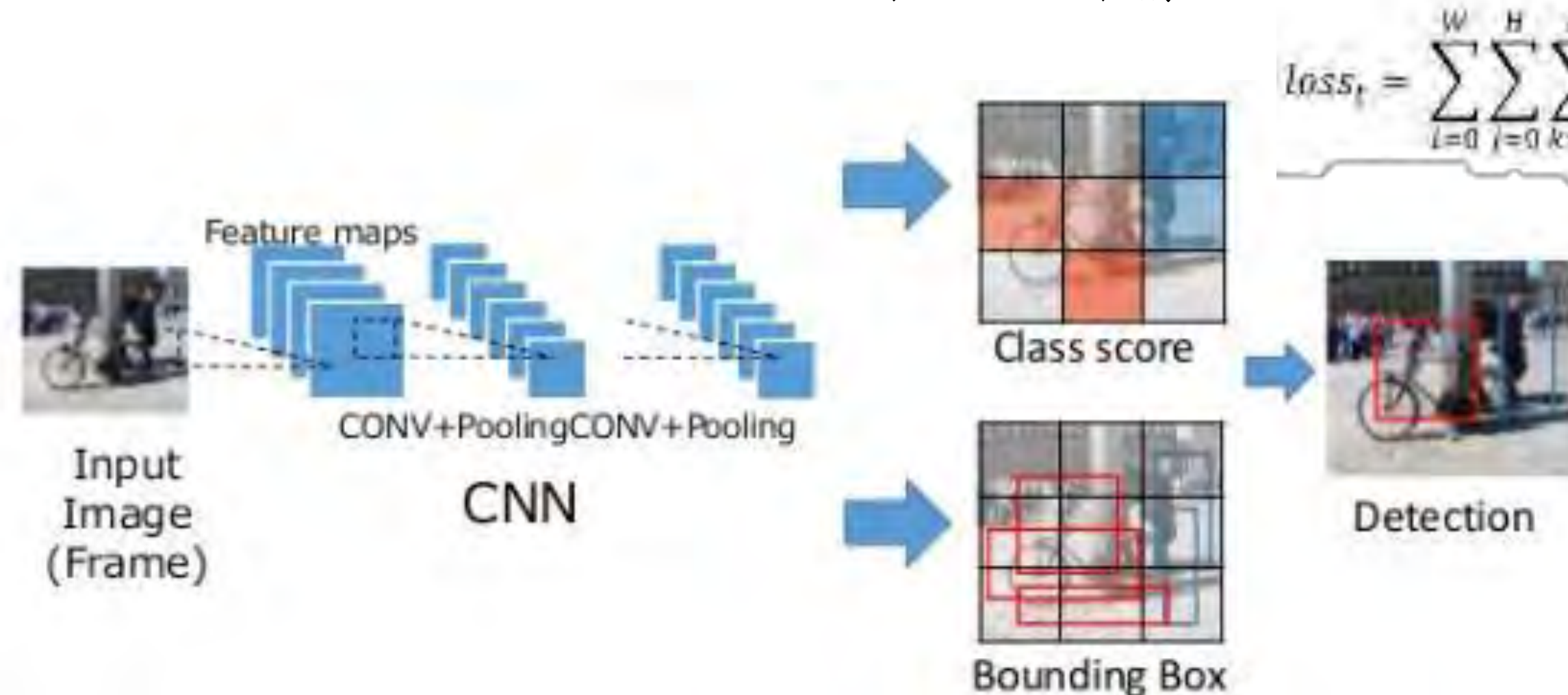
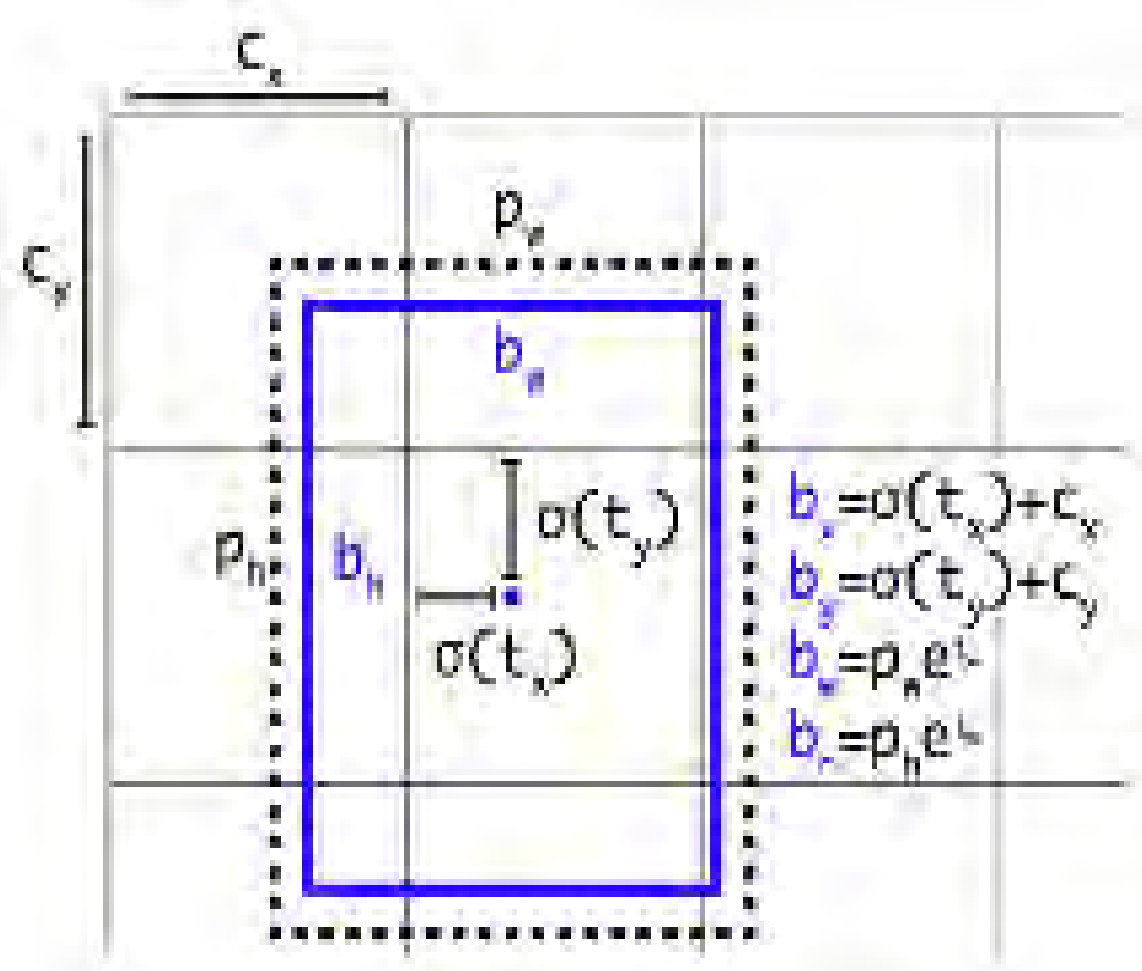
辨識物品

物品辨識為了取得物品的種類與區域，達到快速且低消耗的辨識物品，因此採用物體偵測方法，比較RCNN與YOLO方法，選擇YOLO方法辨識物品。

YOLO v2方法是先將圖片切成S乘S個方格，依序對方格做一次CNN，在YOLOv2模型中只有卷積層與池化層，移出了全連接層，計算邊框對於每一個類別出現的可能性，機率高的方格會以中心每小格用5個不同尺寸的anchors box 進行預測。分析包含五個預測值，中心座標(x,y)、邊框長寬(w,h)和準確度IoU。



$$IoU = \frac{\text{預測範圍} \cap \text{標記範圍}}{\text{預測範圍} \cup \text{標記範圍}}$$



$$loss_l = \sum_{l=0}^W \sum_{j=0}^H \sum_{k=0}^A$$

$$1_{\text{Max IOU} < \text{Thresh}} \lambda_{\text{noobj}} = (-b_{j/k}^0)^2$$

$$+ 1_{L < 12000} \lambda_{\text{prior}} = \sum_{r \in (x,y,w,h)} (\text{prior}_r^x - b_{j/k}^x)^2$$

$$+ 1_k^{\text{truth}} (\lambda_{\text{coord}} + \sum_{r \in (x,y,w,h)} (\text{truth}_r^x - b_{j/k}^x)^2$$

$$+ \lambda_{\text{obj}} \times (\text{IOU}_{\text{truth}}^k - b_{j/k}^0)^2$$

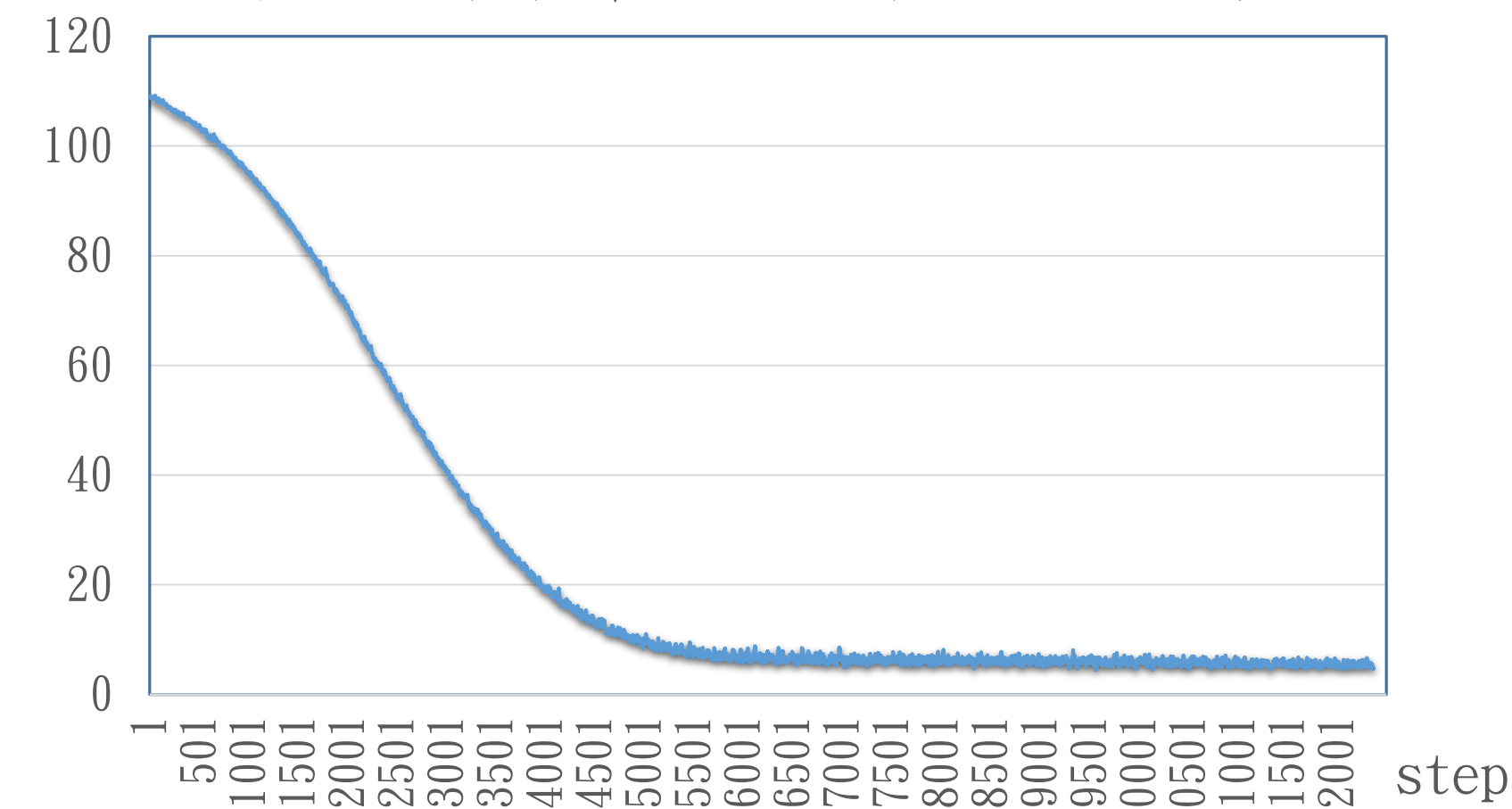
$$+ \lambda_{\text{class}} \times (\sum_{c=1}^C (\text{truth}_c^k - b_{j/k}^c)^2)$$

圖18 YOLO 損失函數公式

表2 物體偵測方法 RCNN 與 YOLO 比較

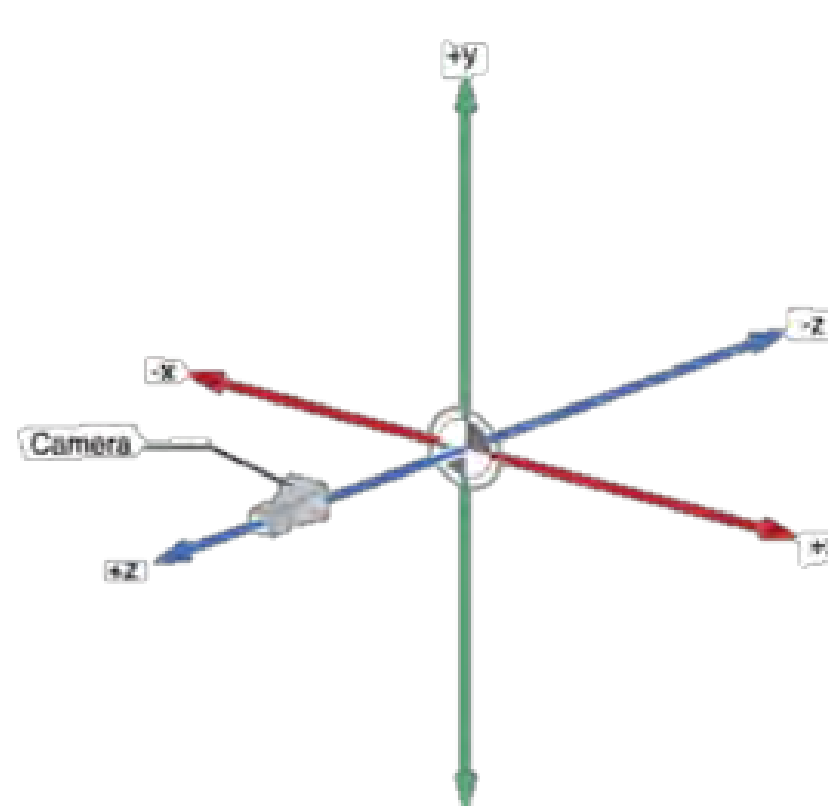
比較項目	RCNN	YOLO V2
速度 (fps)	6 (慢)	67 (快)
準確率 (mAP)	53.5 (較低)	76.8 (較高)
訓練時間 (h)	40	30
同時偵測數量	少	多
支系	multi-stage	one-stage
演算法差別	先提出多個候選框，針對候選框，進行多次的CNN，擷取特徵，透過線性回歸校正偵測框。	YOLO採用end to end的單一網路設計YOLO的CNN訓練與偵測是看整張圖片，因此錯誤率比RCNN還低。

表3 訓練筆數與辨識誤差關係



用標註過的圖片訓練影像辨識模型，隨著訓練筆數增加，辨識誤差逐漸下降，達到更為精準的辨識結果。

記錄



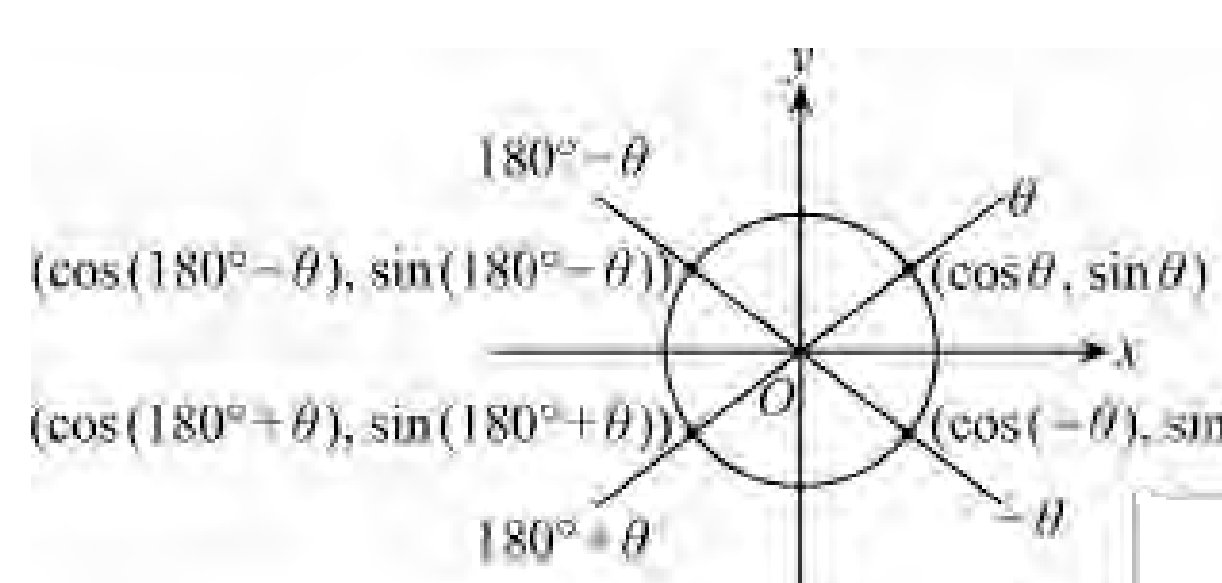
在地圖中以坐標系方式，記錄物品的位置，方向依據笛卡兒坐標系建構。



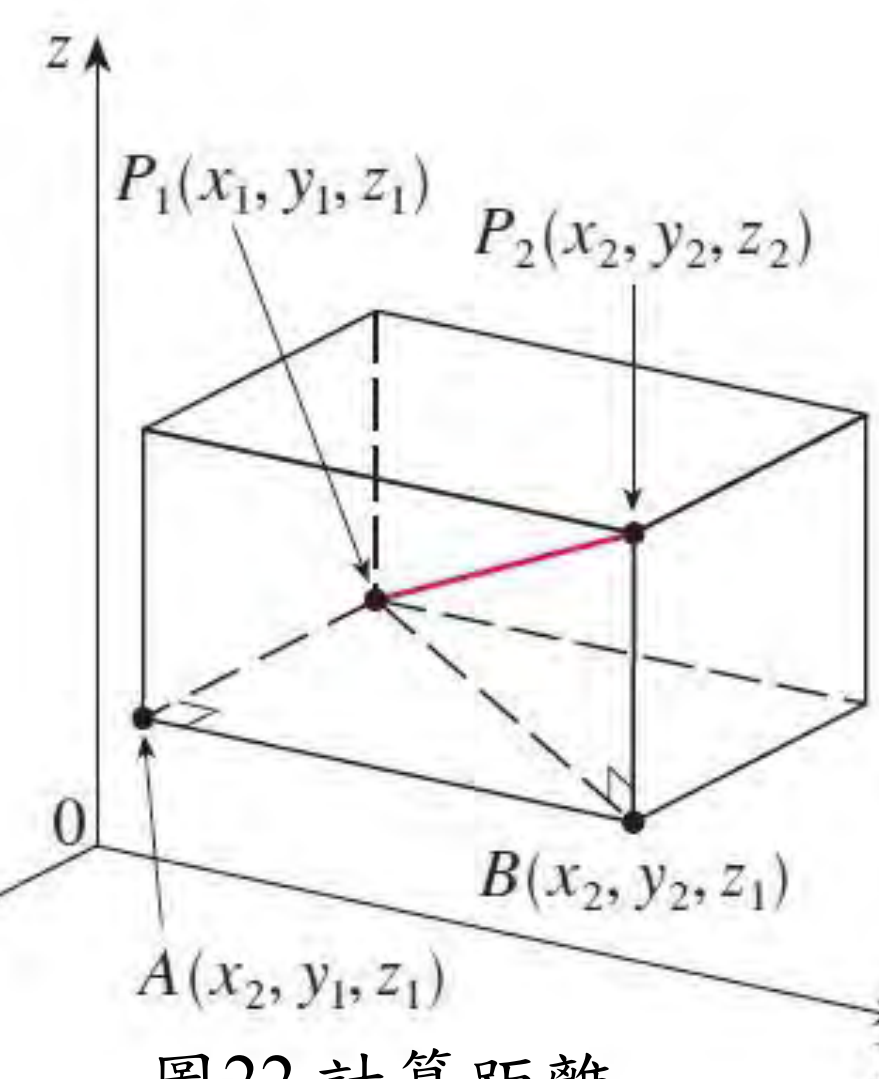
在最後辨識到此物品的位置，放置物品照片，讓使用者尋找物品時，能看到最後的影像。

導航

- 三角函數計算加上視角方向的目标位置方向角度。
- 畢氏定理計算自身與目标位置的空間距離。



$$|P_1 P_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$



研究結果

辨識功能



辨識到多個平面



整合多個平面為一個平面
圖23 辨識平面與整合平面

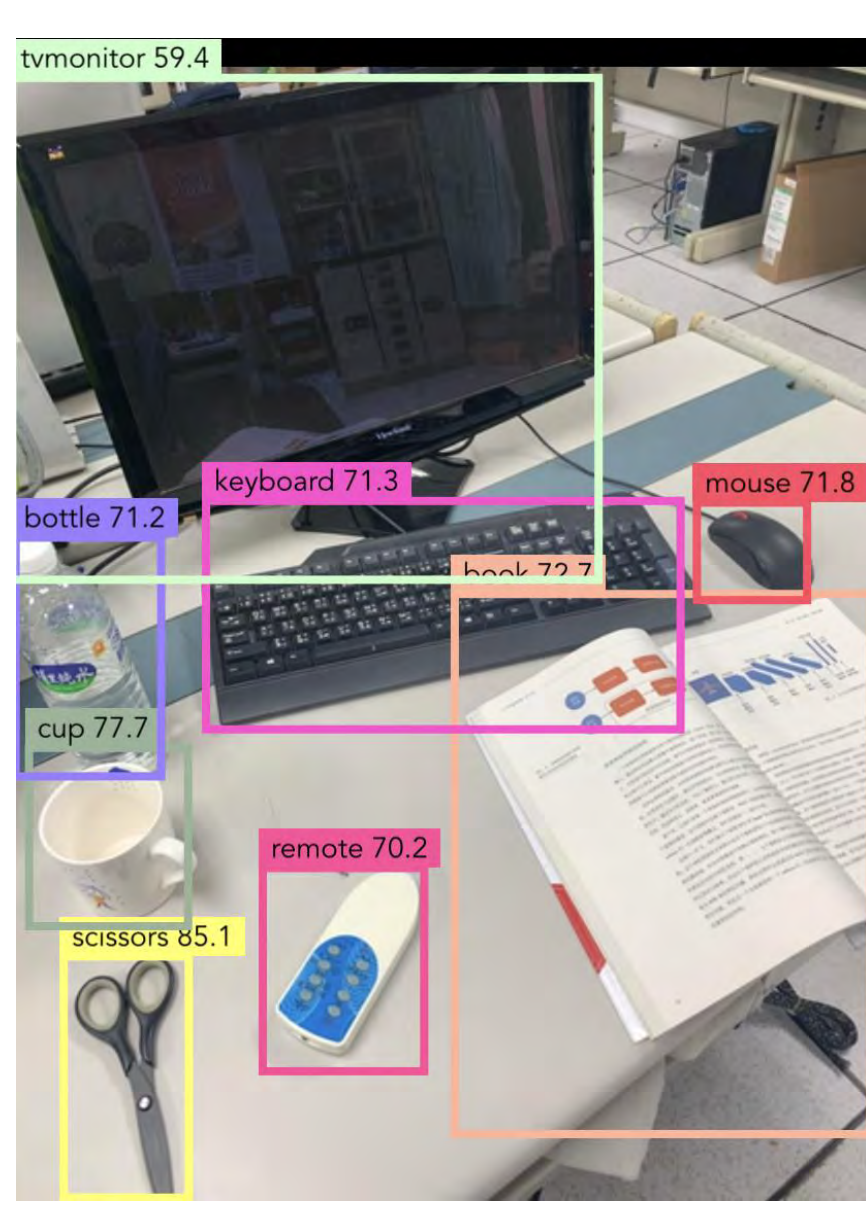
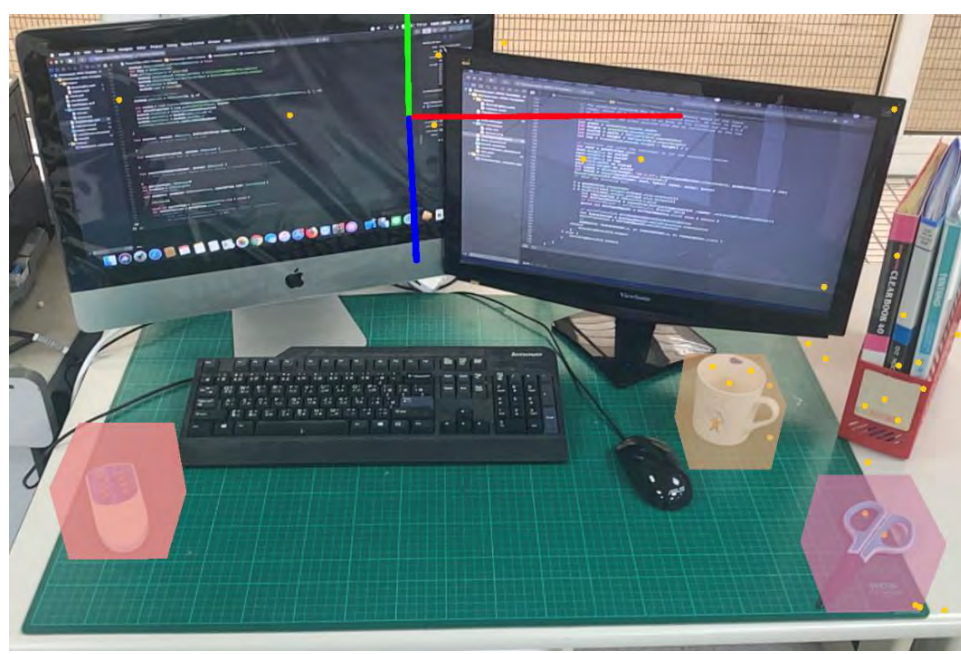


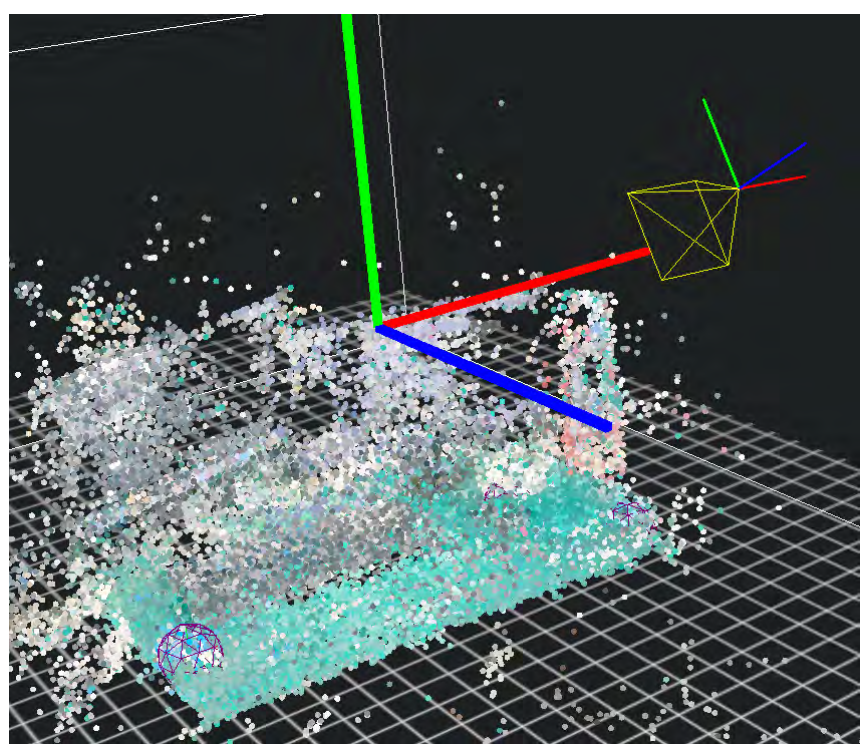
圖24 YOLO影像辨識

依據特徵點分布確認平面位置，整合相近平面。能夠辨識物品種類、位置。

紀錄功能



視角所見影像

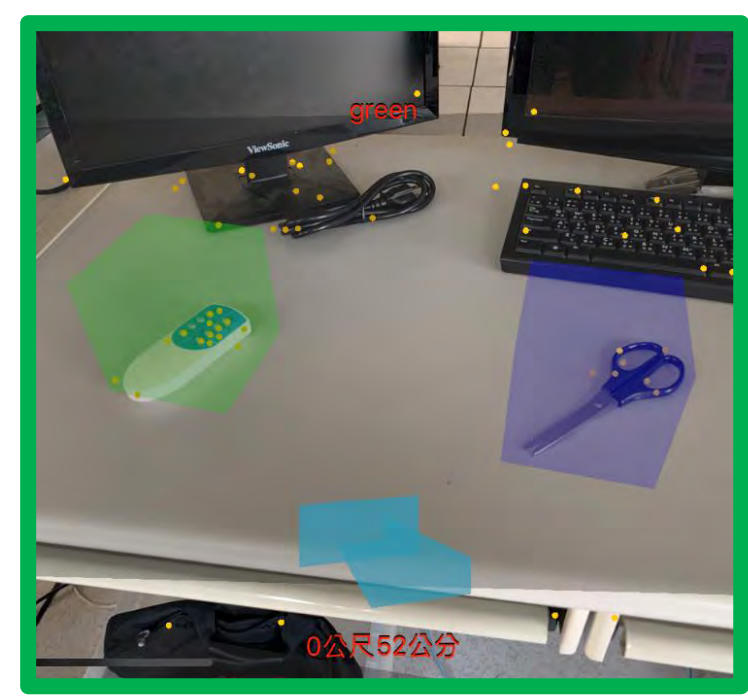


特徵點建構空間地圖

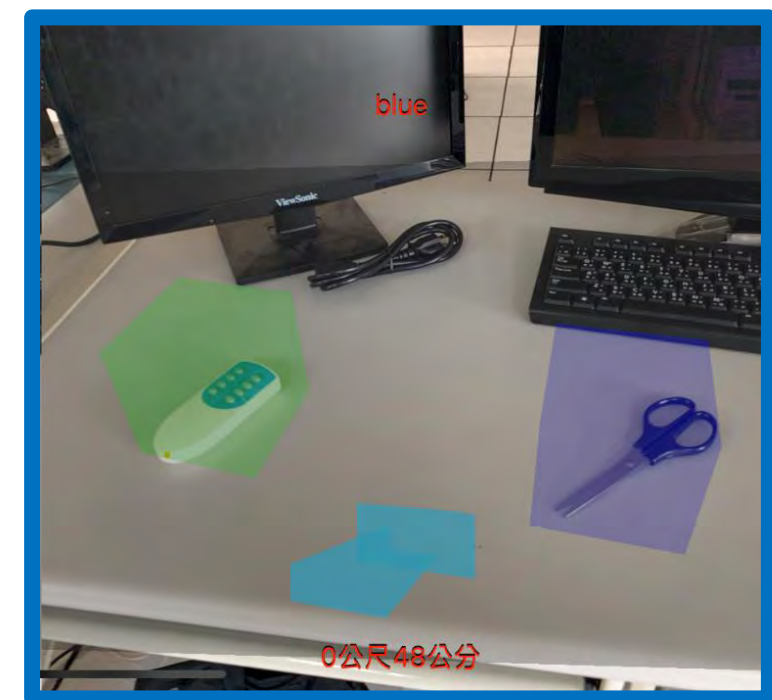
圖25 空間地圖

以特徵點紀錄空間地圖，標記物品位置於地圖中。

搜尋功能



(箭頭指向綠色方塊)
搜尋遙控器



(箭頭指向藍色方塊)
搜尋剪刀

圖26 搜尋不同物品

能依據選擇不同的物品 指向正確的物品位置

導航功能



近距離 面對物品 中距離 背對物品 遠距離 面對物品
圖27 不同距離與不同視角，箭頭導航物品

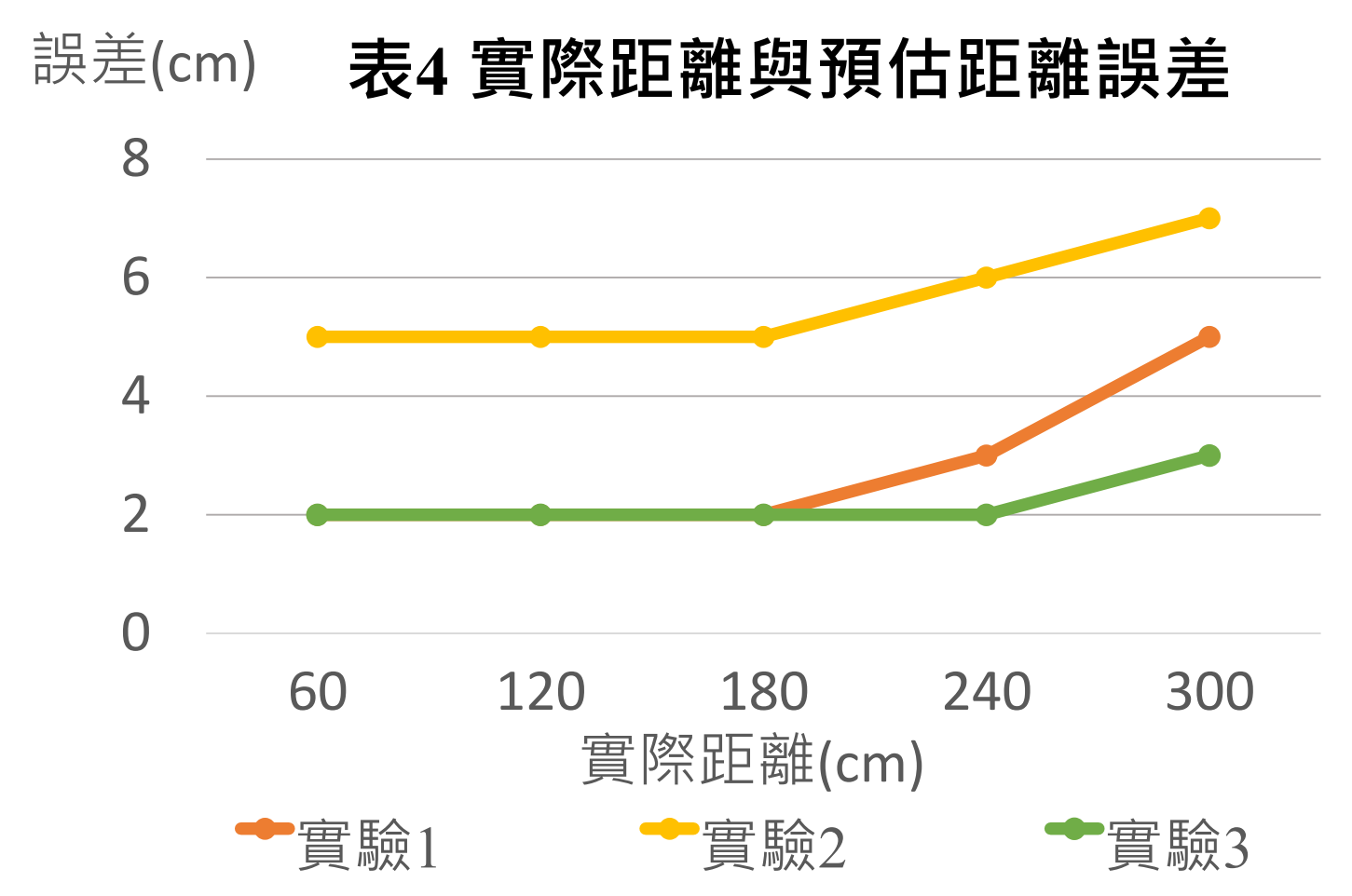
使用者在空間中移動，顯示自身到物品位置的直線距離。藍色箭頭會即時指示到達物品位置的方向。

實驗

實際距離與預估距離誤差



圖28 測量過程紀錄



以每塊60公分地板為基準，測量時實際距離與程式預估距離的誤差，200公分後誤差明顯增加，表示超過200公分後累積誤差才會大於1公分。

移動時物品辨識能力

表5 辨識距離相等不同方向的可信度

距離 (公分)	方向(度)	可信度(%)
50	0	86.1
	45	92.0
	90	89.2
	135	86.8
	180	87.4
100	0	71.4
	45	66.8
	90	75.1
	135	66.9
	180	74.7

表6 物品不同方向最大可辨識距離

方向(度)	距離(公分)
0	165
45	187
90	173
135	130
180	150
225	162
270	157
315	163

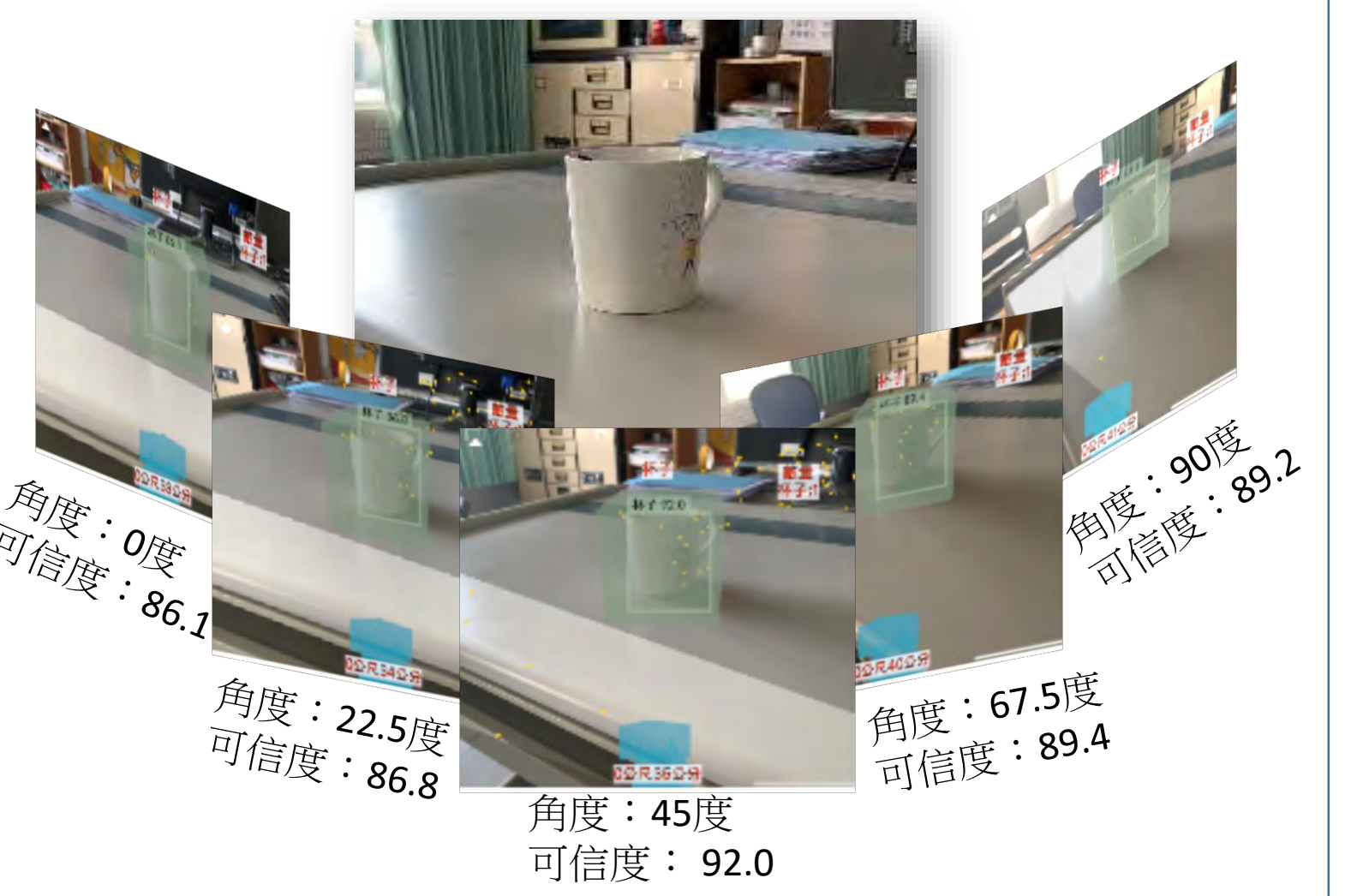
辨識：杯子
物品離地70(公分)
相機離地150(公分)



距離：1公尺 辨識：是
距離：1.5公尺 辨識：是
距離：2公尺 辨識：否

圖29 不同距離辨識

分別測試不同方向與距離，不同物品的辨識準確度，經過實驗得知在不同方向辨識且距離1.5公尺以內，可以辨識到物品。



角度：0度 可信度：86.1
角度：22.5度 可信度：86.8
角度：45度 可信度：92.0
角度：67.5度 可信度：89.4
角度：90度 可信度：89.2

圖30 不同方向辨識

計算離開後回到定位點的定位偏移誤差

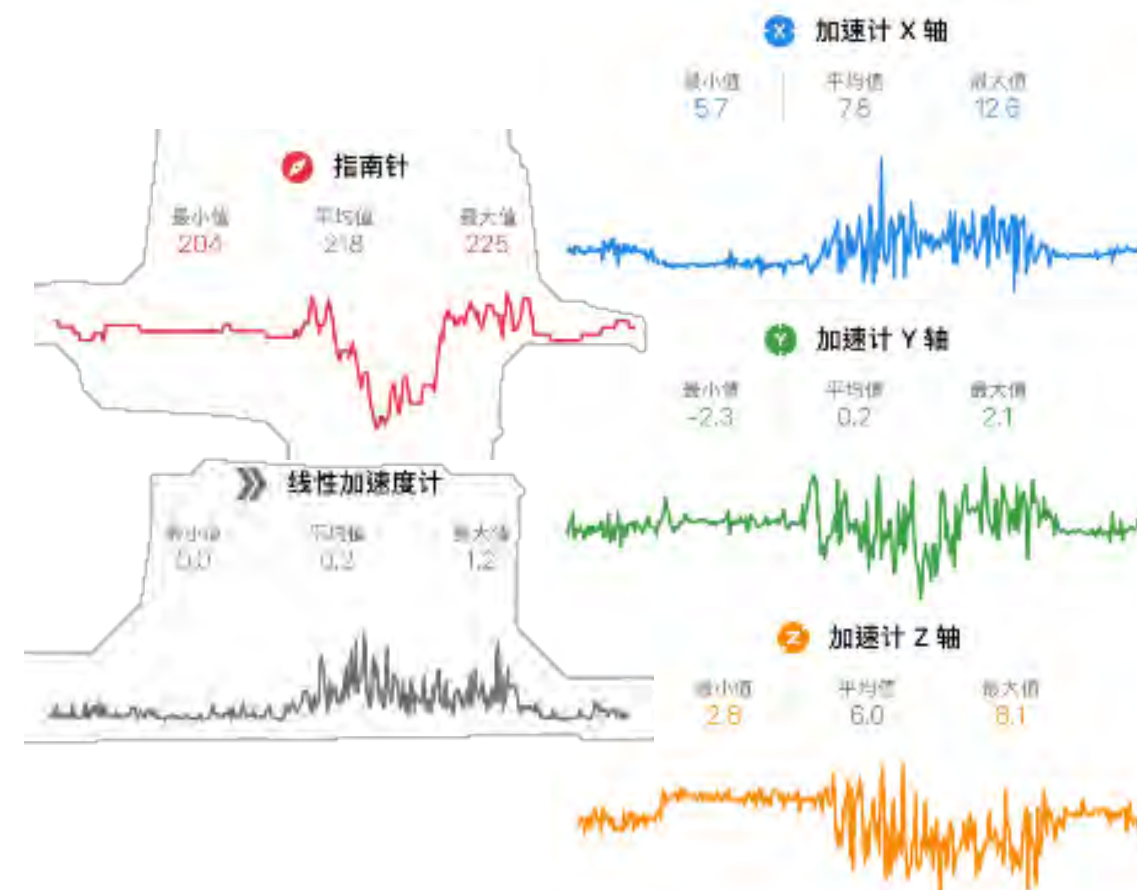
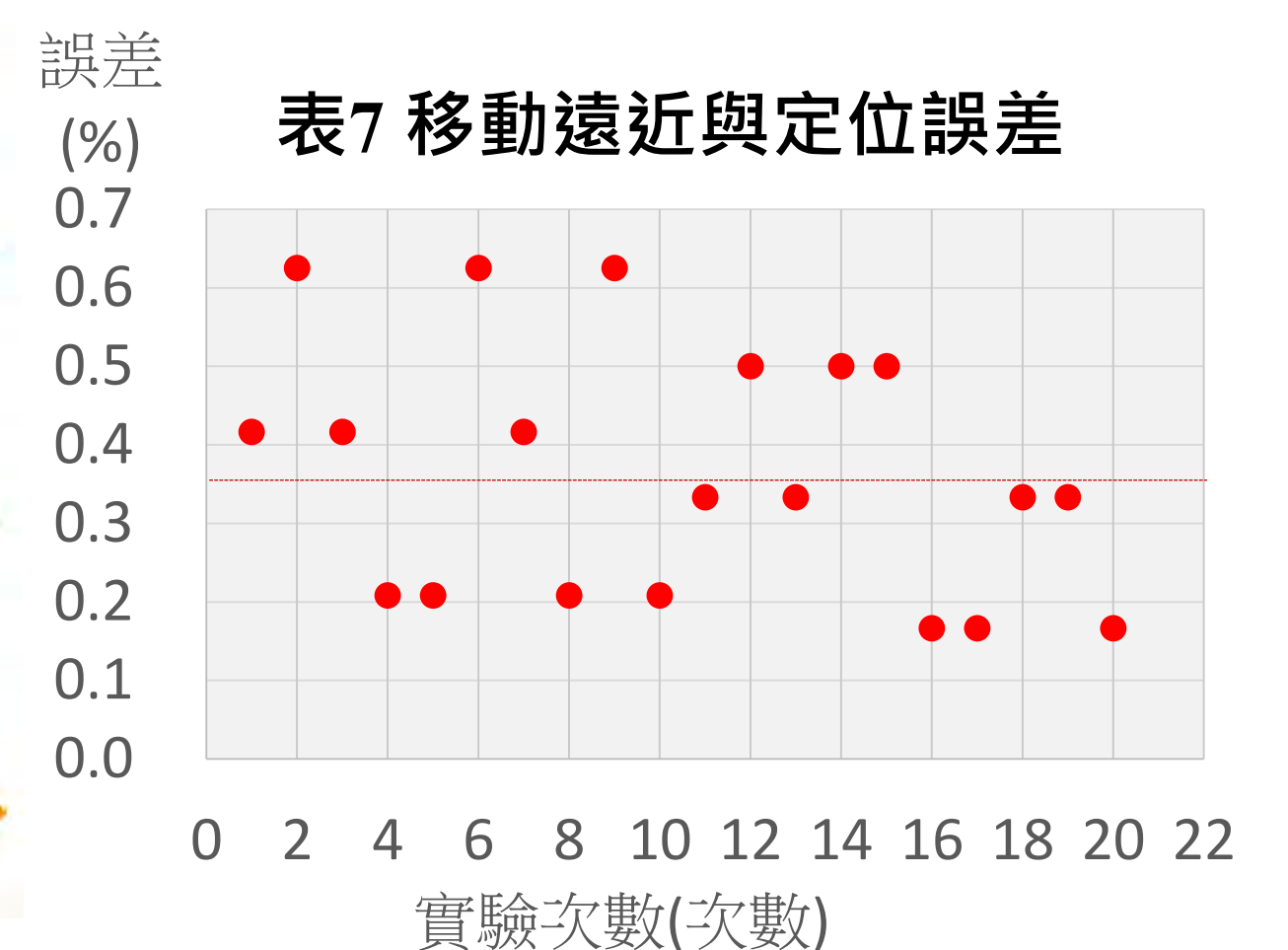


圖31 移動過程紀錄



圖32 移動前後的圖片



測試遠離定位點後再回到定位點，定位點偏移原定位點的距離，測試20次，平均結果得出，約產生0.35%的誤差。

討論

1. 系統架構流程仿造人類大腦記憶的方式，達到空間認知與物品辨識，輔助人記憶物品位置。
2. 相較於防丟器產品，本作品能快速記憶大量物品位置，明確的導引方向和距離，帶領使用者找到物品位置。
3. 本作品能即時建構空間地圖，自動辨識和記錄物品在空間中位置，相似於Google Map的室內物品地圖，擴展使用者的搜尋至現實的空間和物品。

結論

本作品利用SLAM、YOLO等技術，建構一個可做到環境建構、記憶物品位置與指示物品資訊的「AR記憶眼鏡」，並能呈現檢索結果及位置方向資訊，達到平時輔助記憶物品，完成的功能如下：

1. 運用SLAM達到辨識環境、YOLO達到物品辨識，完成定位物品的空間位置。
2. 使用者能透過選擇物品種類，檢索到物品所在位置。
3. 使用者能依照AR的指示資訊，正確且快速尋找到物品所在位置。

未來展望

近期目標優化影像辨識，讓作品更方便且實用，遠期目標將整理不同使用者的辨識數據，達到使用者共享辨識資料，實現物品空間資訊的互通與傳遞，目標如下：

1. 近期：修改與優化影像辨識模型，達到更快速且精確的辨識物品。
2. 遠期：將不同使用者的辨識數據透過網路儲存於資料庫，做到整合不同辨識資料。

參考資料及其他

論文書籍：
 Ian Goodfellow, Yoshua Bengio, & Aaron Courville (2016). **Deep Learning**. MIT Press.
 J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi.
You only look once: Unified, real-time object detection. In CVPR, 2016.
 Ole Kiehn & Hans Forssberg (2014),
The Brain's Navigational Place and Grid Cell System.

Sudeep Pillai and John J. Leonard (2015) ,**Monocular SLAM Supported Object Recognition**.
 高翔,張濤(2017)。視覺SLAM十四講：從理論到實踐。
 出版商：電子工業出版社
 陳玉琨,湯曉鷗(主編)(2018)。人工智慧基礎(高中版)。
 出版商：華東師範大學出版社