

運用 Google-Cloud-Vision-Based 半 自動照片分類檢索技術之 POI 製作系統

學生：呂兆凱、魏湧致、江羿賢、陳維竑

指導教授：黃崇明

目錄

壹、前言3
貳、主要技術議題4
參、如何解決技術議題5
肆、平台架構8
伍、系統發展與實行10
陸、使用及範例10
柒、結論11

壹、前言：

隨著手機相機品質的提升與行動網路的蓬勃發展，我們隨時隨地都能拿起手機拍照留下紀念，甚至上傳至社群平台與朋友分享，然而隨著照片日積月累的增加，若沒有好好的分類整理，等到某天想起曾經拍下的照片想要回味時就很難快速的找到，而就算手動的將所拍的照片加註解與分類，也需要花費大量的時間。

為了解決繁冗的分類過程，常見的方法包含了依據拍照地點、拍照時間等來分類。而我們更進一步地運用了 AI 來協助我們為照片加上許多特徵標籤，我們使用了 Google Cloud Vision 這項採用機器學習技術的影像辨識軟體，透過預先定義的標籤來分類圖片，也能夠識別熱門的地標與標誌，其優點是運用了 REST (REpresentational State Transfer)與 RPC (Remote Procedure Call) 架構，使用者只需安裝我們的 APP 即可享受這些服務。使用者在 APP 可以對每一張照片建立一些基本屬性(如標題、描述、關鍵字等)，APP 也會自動的對拍照的地點定位，最重要的是能夠選擇 Cloud Vision 分析出的特徵。

使用者將照片與各項屬性上傳至照片資料庫後，我們也提供 POI (Point of Interest)製作平台，讓使用者檢索出想要的照片來製作成各式各樣的景點介紹，並列表於首頁，使我們也能看到別人的精心製作，讓大家來一同欣賞。

貳、主要技術議題：

在現代手機普及的情況下，人們往往會使用手機來記錄身旁的人事物，使得手機相簿裡的照片每日是有增無減，而其中人們出遊的景點照更是佔有著一定的分量。但眾多照片在缺乏整理及分類的情況下，若日後想要尋找特定一張照片時是相當的困難，如我們想要找出某時間在台南安平古堡拍攝的照片，其實是一件大海撈針的事情，而我們主要想要解決的，就是利用半自動的方式，引導使用者在拍攝完當下就將照片做分類及標記，以方便其日後找尋。以下為相關技術問題。

(A)要如何為每一張照片快速建立一個容易被識別的方式，以及在實際開發的技術層面上會碰到哪些問題？

- (一)要使用手動註解還是 AI 辨識？
- (二)如果使用 AI 辨識，要怎麼處理精確度問題？
- (三)照片要存放哪些項目，讓使用者方便使用？
- (四)要如何將資料從 APP 傳輸至資料庫？
- (五)Android 以及 iOS 的版本眾多，要如何選擇支援的版本？

當拍攝完照片並進行標記處理之後，我們所需要的是一個易於使用且快速方便的搜尋照片系統。這個搜尋系統的目的是盡可能的讓使用者在輸入簡單的關鍵字的情況下，快速並且精準的找到欲搜尋之照片。以下為相關技術問題。

(B)要以何種的方式來對照片做搜尋，以及在實際開發的技術層面上會碰到哪些問題？

- (一)用以圖搜圖還是用文字來做搜尋？
- (二)搜尋時要以甚麼方式來限定範圍？
- (三)如果以文字搜尋，可能會難以完全吻合字詞？
- (四)當同時有多位使用者時，可能會造成衝突？
- (五) Google Cloud Vision 能產生的標籤有數萬個字詞，使用者在不記得確切 label 名稱的情況下要如何才能正確輸入以找到照片？

有了照片標記以及照片搜尋系統之後，我們也提供了 POI 景點製作平台，而這

個平台連結了以上兩個子系統，使得人人都可以透過簡易的步驟，選取照片後試著建立自己有興趣或是值得他人認識的 POI 景點，以豐富大家的觀光視野與推廣在地旅遊。以下為相關技術問題。

(C)要以何種方式來使照片做成 POI (Point of Interest)

(一)POI 要有包含哪些欄位才能清楚呈現出來

(二)要如何設計網頁 UI 才能夠讓使用者輕鬆點選查看

參、如何解決技術議題：

(A)要如何為每一張照片快速建立一個容易被辨識的方式，以及在實際開發的技術層面上會碰到哪些問題？

(一)要使用手動註解還是 AI 辨識？

我們認為，比較好的方式是部分讓使用者手動填寫，如標題、描述、關鍵字等等，拍照當下的時間與地點讓手機自動抓取，而在照片內容的部分則使用 Google Cloud Vision 來做分析，可以快速地取得照片的特徵以及辨識那張照片是哪個景點，利用使用者手動與 AI 辨識兩者來達到為照片半自動下標籤的功能

(二)如果使用 AI 辨識，要怎麼處理精確度問題？

在照片上傳給 Google Cloud Vision 分析時，回傳的結果通常會有很多項，但是有部分會有錯誤或冗餘的問題發生，所以我們會先透過信心程度，來篩選出比較高信心程度的標籤來給使用者做選擇，其中再讓使用者挑出錯誤的標籤或是不需要的標籤，來使最後出來的結果較為精確

(三)照片要存放哪些項目，讓使用者方便使用？

我們將 MySQL 資料庫分為 photo_tags 與 vision_api 兩個資料表，photo_tags 存放手動輸入與手機定位取得的標籤，有：標題、日期、經度、緯度、高度、方位、方位角、範疇、關鍵字、描述、參考資料、同伴、喜好程度、貢獻者，vision_api 存放由 Google Cloud Vision 所產生標籤，分成 label 和 landmark 兩大類。利用這樣的照片存放與資料庫建立方法讓這些資料更方便的存取，也能透過這些資訊知道哪張照片是對應到哪個資料欄位，解決了存放方式的問題。

(四)要如何將資料從 APP 傳輸至資料庫?

首先，我們先將照片傳至 HFS 伺服器，並依據上傳的照片順序來進行命名，這樣之後的存取會更加有效率，方便資料庫的資料對應到照片。接著，手動、手機定位及 Google Cloud Vision 所添加的標籤傳至 MySQL 資料庫，傳送的方式是先在手機 APP 上將那些標籤資料包成一個 JSON 檔再經由 node.js server parse 後傳至 MySQL 資料庫。

(五)Android 以及 iOS 的版本眾多，要如何選擇支援的版本?

根據 2020 年 Google 的 Android 系統版本使用情況報告，Android7.0 以上的使用者已達 50%以上，而使用 Android4.0 以上的使用者更是超過了 99%，為了要使大多數的人都可以正常使用這個 APP，我們採用的 Android 開發版本是可以使 Android4.0 以上的使用者正常的使用基本功能，而如果是 Android7.0 以上的使用者則可以享受到更完整的外觀設計以及使用者體驗。而在 iOS 的部分，2021 年 2 月的 iOS 系統使用情況指出，iPhone 上的 iOS14 有 86%的使用者，iOS13 則有 12%的使用者，其他 2%為更早之前的版本，所以我們 iOS 的開發版本是可以正確無誤的支援 iOS13 以上的版本，可以涵蓋到超過 98%的使用者。這樣子的版本選擇可使 98%以上的手機使用者正常的使用該系統。

(B)要以何種的方式來對照片做搜尋，以及在實際開發的技術層面上會碰到哪些問題?

(一)用以圖搜圖還是用文字來做搜尋?

我們認為用文字來搜尋是比較恰當的一個方式。一個是因為以圖搜圖只能針對照片內容來搜尋，無法比對使用者手動輸入的標籤，若想根據自己所下的標籤來搜尋準確率會較低。另一個是因為如果使用以圖搜圖的話勢必要一個一個比對目前在資料庫的所有照片，當資料量一多時，對於照片的分析會耗時很久，這樣就比較不符合手機使用者想要達到的方便快捷了。而使用文字搜尋的優勢是可以依照自己所下或所選 Google Cloud Vision 產生的標籤來做搜尋，這樣可以比較精確的知道自己要搜尋的照片可能會有甚麼樣的資料。另一個優勢是字串的比對速度相較於圖片快上非常多，所以能讓使用者快速的享受到這項服務。

(二)搜尋時要以甚麼方式來限定範圍？

我們挑出了一些使用者較容易搜尋的項目來讓其搜尋，有標題、日期、關鍵字、描述、參考資料、景點製作貢獻者、範疇、喜好程度、label、landmark，去除掉一些比較難以去搜尋的項目。透過建構這些欄位，可以引導使用者放入更精確的字詞，以便在日後的搜尋上，能夠更正確有效率的找到欲搜尋之照片。

(三)如果以文字搜尋，可能會難以完全吻合字詞？

在搜尋照片的時候，可能會因為輸入的時候少了一個字而因此搜尋不到，或者是想搜尋相同性質照片的時候，會不知道要下甚麼樣的搜尋關鍵字會比較好，所以我們想了以下方法改善此問題。我們改善了資料庫的搜尋型態，從一般字串完全相同的比對搜尋方式，利用了 SQL REGEXP 的語法改成使用模糊搜尋(fuzzy search)，透過這個方式，可以用更加簡單、分散的字詞，來搜尋到更全面且相關的資料，讓使用者輸入錯誤時也能有相對應的搜尋結果，順利改善了這項問題。

(四)當同時有多位使用者時，可能會造成衝突？

當搜尋系統有多位使用者時，可能會導致系統錯亂，進而導致使用者可能會搜尋到不正確的結果，這邊我們會要求使用者在使用此系統時，先進行輸入使用者名稱來判斷是哪位使用者在使用，如此一來後端就能正確地將搜尋結果回傳給對應的使用者。

(五) Google Cloud Vision 能產生的標籤有數萬個字詞，使用者在不記得確切 label 名稱的情況下要如何才能正確輸入以找到照片？

我們將資料庫中所存有的 label 和 landmark 蒐集起來，在使用者使用 label 或 landmark 做搜尋時，會根據其當前所輸入的字來跟資料庫的所有字詞的 prefix 來做比對，若有符合的就會成為推薦字詞。如想搜尋具有赤崁樓物件的照片時，輸入“chi”，推薦字詞列表中就會出現“chihkan tower”可供使用者選擇輸入，以避免使用者必須完全記得相關字詞，或僅僅可能因錯了一個英文字而造成搜尋不到結果的情況。

(C)要以何種方式來使照片做成 POI (Point of Interest)

(一)POI 要有包含哪些欄位才能清楚呈現出來

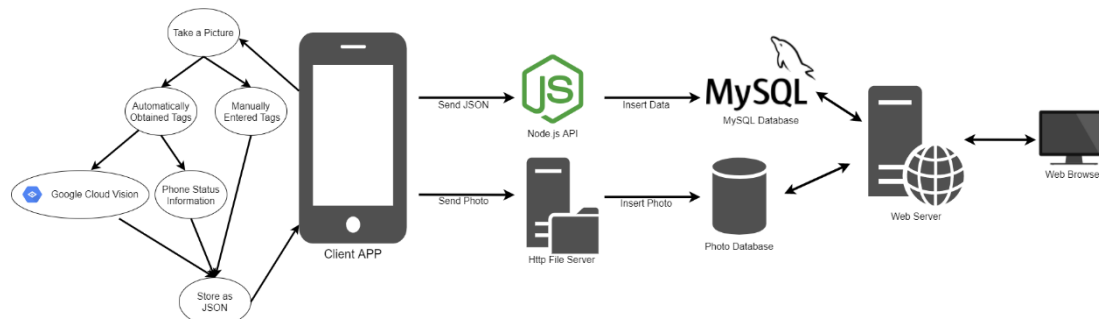
這邊我們選取照片的方式是利用在前述有提到過的搜尋系統來添加照片至 POI，點選進照片按添加按鈕做添加，而其中我們利用標題.描述.範疇.景點製作貢獻者來做為要添加的欄位，可以利用標題來輕鬆知道這裡大概是哪裡，而透過範疇可以知道這個 POI 是屬於什麼類型，而描述跟景點製作貢獻者可以寫下介紹並且知道作者是誰，利用這四個欄位就可以獲得該 POI 很全面的資訊，且容易使人產生興趣並觀賞。

(二)要如何設計網頁 UI 才能夠讓使用者輕鬆點選查看

這邊我們將整個系統的首頁做成一個 POI 列表，讓其他人能夠一進去網站中就能夠欣賞先前的人製作的成品，再來將製作的按鈕放在網頁的上方，先讓使用者產生興趣後，才會比較有意願去製作屬於自己的 POI，透過此方式可以讓這個網站的使用者使用體驗較佳，並且吸引更多的使用者透過此網站來製作 POI。

肆、平台架構：

結合網站、APP 與 Cloud Vision 的標籤結果及後端資料庫。前端所使用的是 HTML、CSS、swift(iOS)、java(Android)，後端使用 mySQL 資料庫，搭配上 node.js 的 express 架構和 HTTP File Server。利用 Google Cloud Vision 分析出該照片的特徵，由使用者選擇完後，上傳至資料庫與伺服器中，在製作 POI 時由資料庫提供資料。



資料庫架構:

photo_tags :

儲存基本照片屬性，分為手機自動抓取(如：日期、經緯度)與使用者手動添加

(如：關鍵字、描述)。

vision_api：

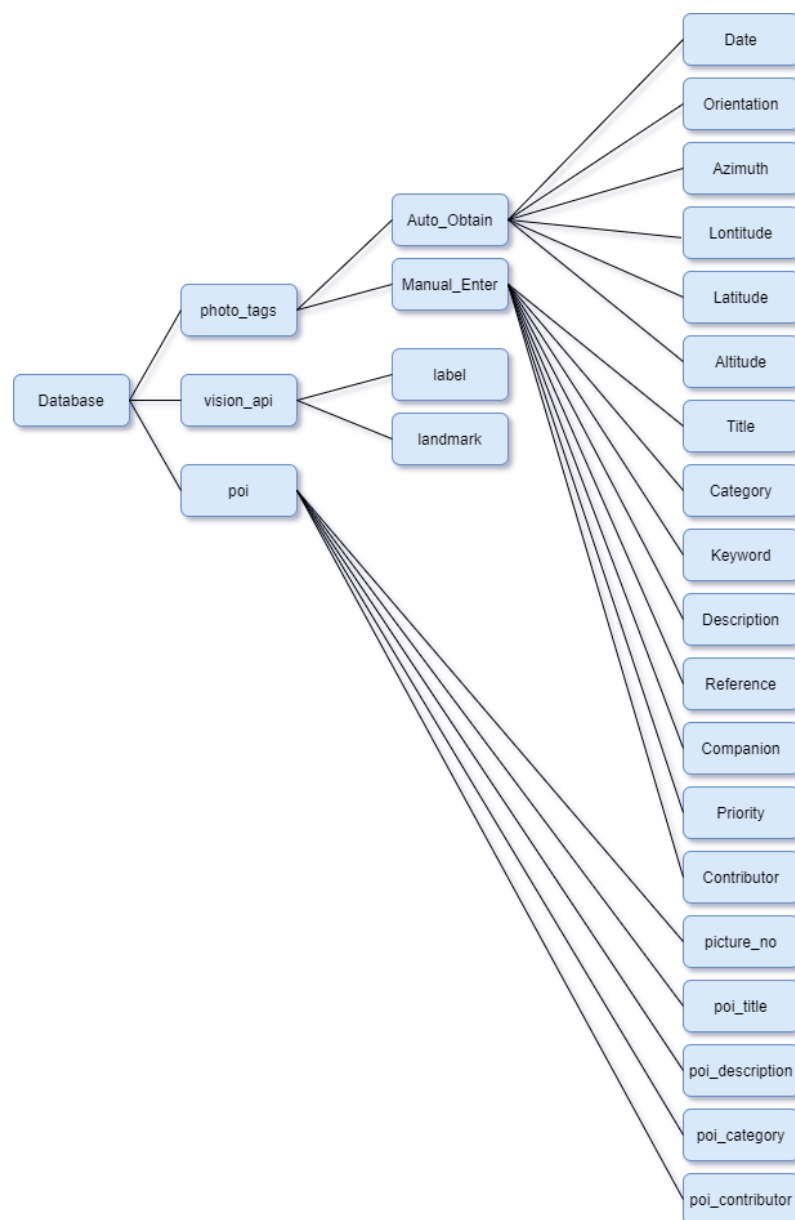
儲存使用者所選來自 Google Cloud Vision 分析的照片特徵。

poi：

儲存製作出來的 POI 資料。

Photo Server：

利用 HFS (HTTP File Server) 架構，將照片上傳至此 file server 做儲存。



伍、系統發展與實行：

網站端:搜尋照片、製作 POI、檢視目前已存在的 POI

使用者可以輕鬆使用網站來去製作 POI 以及搜尋照片，製作完之後可以用手機

或是電腦上網查看成品。

APP 端:獲取 Google Cloud Vision 分析之資訊、手動添加屬性、APP 自動抓取位置與時間、上傳照片以及相關資料至伺服器及資料庫
透過手機來拍照並且分析照片資訊，並手動添加一些標籤後就製作好一張照片的基本資料，可以用來製作成 POI

陸、使用及範例：

例如，一般的旅行者或甚至是解說導覽員都能使用這項服務來讓自己有更輕鬆、愉快的體驗。

使用者能利用手機 APP 去拍各式的景點，如:安平古堡、赤崁樓，拍完照後透過 APP 來對照片進行分析後選出最符合的一些特徵，以及手動添加一些標籤之後就可以上傳至伺服器以及資料庫。

等到拍完照之後，可以至網頁上的製作 POI 頁面來進行製作，可以挑選一些使用者認為合適的照片加入其中來製作出景點介紹。也可以利用此系統結合成功大學的文史脈流網站(DEH)，來方便使用者的創作。

以下是 POI 的製作流程:

1. 利用 APP 拍照並標記後上傳圖片至伺服器以及資料庫。
2. 使用照片搜尋系統來尋找想要的照片，並輸入景點標題、範疇、描述等製作 POI。
3. 透過網頁端上傳此筆 POI 資訊至資料庫。
4. 在手機或電腦上的首頁景點列表查看各個 POI。

柒、結論：

我們製作的這一套系統，提供了使用者一套非常方便的平台來進行照片的分類檢索，以及可以在網站上為這些照片製作成 POI 來欣賞，並且分享給大

家看。

這項半自動照片分類檢索技術結合 AI 分析照片及手動添加屬性，不用再自己花費大量的時間來記錄每張照片有甚麼特徵或地景，輕鬆了分擔的使用者分類照片的時間

只要有手機就可以輕鬆便利的使用這一套系統，且跨平台(Android/iOS)都適用，不管身上有多少裝置想要一起整合的，也都可以透過這套軟體輕鬆達成目標。

無論是想要為自己的旅程做個紀錄，亦或是想要將這些好看的照片蒐集分享給他人一同欣賞，甚至是製作出 POI 便於介紹，這套方便的系統一定是使用者不二的選擇。