

第三章 系統簡介

本章將說明研究系統之架構與主要流程。本研究系統共可分為兩個模組，分別是表單手寫欄位擷取與表單手寫資料擷取。如圖 3.1(a)為第一階段空白表單文件之手寫欄位擷取流程圖，(b)為第二階段已填寫之表單文件手寫資料擷取流程圖。

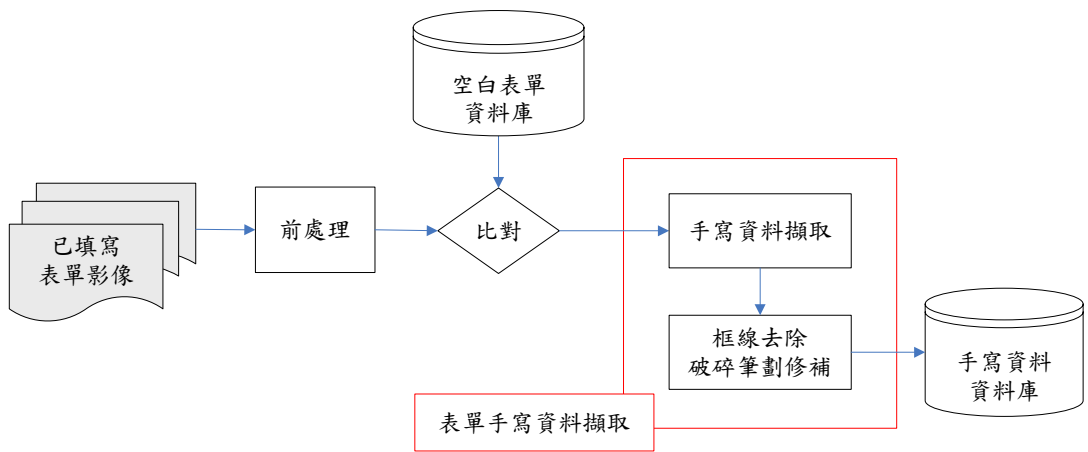
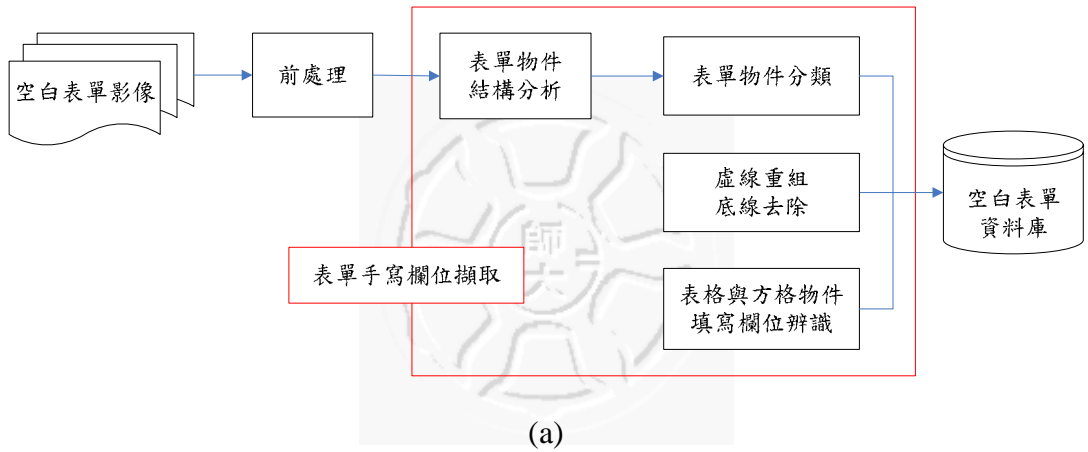


圖 3.1 系統流程圖

(a)空白表單手寫欄位擷取流程

(b)已填寫之表單手寫資料擷取流程圖

一、表單手寫欄位擷取

本模組主要目標在擷取出空白表單影像中，可能為使用者需要填寫的所有填寫欄位。工作流程可分為前處理、表單物件分類與填寫欄位抽取等三個主要步驟，最後再將求得之資料填寫欄位的相關資訊，儲存至表單資料庫中，供下一階段與填寫表單比對之用。

前處理包括空白表單影像的二元化、校正表單影像的傾斜角度等處理工作。

前處理之後，即進行表單資料物件的分類。表單資料物件分類的步驟，主要目的為分析出表單中屬於手寫資料欄位的資料物件與非填寫資料欄位物件。本研究根據常用的表單格式，歸類出四類屬於手寫資料欄位的基本物件類別，分別為橫實線類、表格類、方格類及點物件類等，加上非填寫欄位物件類別，如印刷字、圖形與符號等共五類資料物件分類。

由於橫實線類物件包括屬於手寫資料欄位物件的橫直線，與非手寫資料欄位物件的說明文字之底線，故需進一步針對說明文字之底線進行分析並去除之。

點物件為構成虛線之基本組成元素。連續多個點物件可構成可能為手寫資料欄位之橫虛線，因此亦需分析點物件之分佈情況，進行虛線之重組。

由於表格物件的結構，通常為包含一個以上資料欄位的複雜結構，故在資料物件分類之後，需要針對複雜表格物件進行細部欄位的擷取。此外，由於表格欄位與方格物件皆可能同時包含填寫欄及說明欄，且部分填寫欄中亦含有印刷文字，非一般的空白填寫欄。因此針對表格欄位及方格物件，必須進行更仔細的分

類。

透過上述資料物件之分析與分類，本研究即由五大類資料物件中，擷取出橫直線、橫虛線、核對框、長方框及表格填寫欄位等六類手寫欄位物件。圖 3.2 顯示手寫欄位擷取之過程與結果範例。其中(a)為空白表單文件輸入樣本，(b)為經表單資料物件分類之結果。(c)則為經過點物件重組、表格與方格說明欄位與手寫欄位切割分類後，所擷取出之手寫欄位之結果。

二、表單手寫資料擷取

表單手寫資料的擷取模組，其目的為針對未知且已填入資料之表單，進行表單比對、手寫資料擷取、框線去除及破碎字修補等處理。

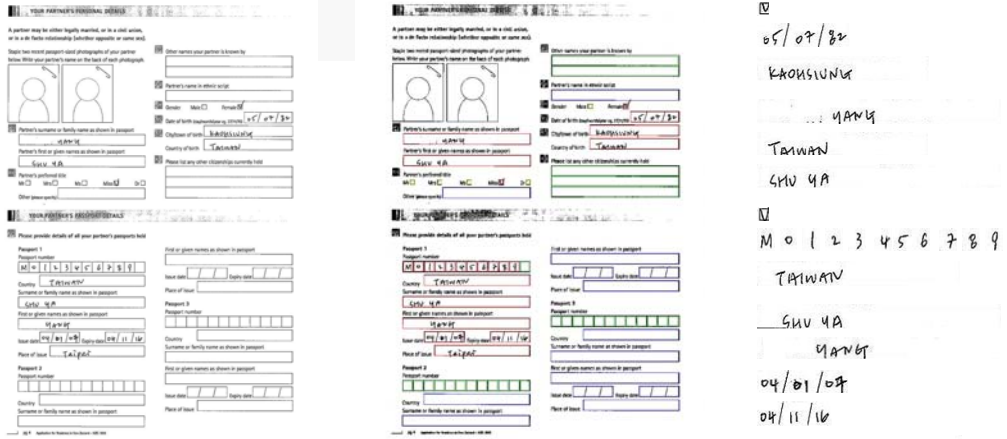
在表單比對的步驟中，輸入已填寫之未知表單影像，利用表單資料庫中所儲存之同類型空白表單進行比對。其後即利用表單資料庫中同類空白表單之手寫資料欄位資訊，擷取出所有手寫欄位中的手寫資料。由於一般人習慣將文字書寫在橫直線欄位上，造成部分手寫筆劃與欄位框線相連，此情況會直接影響到手寫資料辨識的辨識率。因此去除框線，並修補框線與原筆劃相交處所去除之像素，亦為一必要的處理步驟。圖 3.3 顯示手寫資料擷取之過程與結果範例。其中(a)為以填寫之表單文件輸入影像樣本，(b)為與空白表單比對後，所擷取出之已填寫之手寫欄位，圖中紅色方框處即為辨識出已填寫之手寫欄位區域，(c)則針對擷取出之手寫資料經框線去除及破碎字修補所得之結果。



(a) (b) (c)

圖 3.2 空白表單手寫欄位分類擷取過程與結果範例

- (a) 空白表單輸入樣本
- (b) 經表單資料物件分類與擷取結果
- (c) 表格說明欄位與填寫欄位辨識結果



(a) (b) (c)

圖 3.3 已填寫表單手寫資料擷取過程與結果範例

- (a) 已填寫表單輸入樣本
- (b) 經表單比對後所得之已填寫區域，以紅色方框表式
- (c) 經框線去除與破碎字修補之手寫資料影像範例