

以影像為基礎的書店自動化盤點系統

An Image-based Bookstore Inventory Check System

劉震昌

國立暨南國際大學資訊工程學系

Email: jcliu@ncnu.edu.tw

周秉宜

國立暨南國際大學資訊工程學系

Email: s97321515@ncnu.edu.tw

王健駿

國立暨南國際大學資訊工程學系

Email: s97321508@ncnu.edu.tw

摘要—書店在盤點書籍時，需要耗費大量的人力與時間對書店架上書籍逐一進行清點比對，因此盤點周期通常較長，店家無法準確掌握庫存資訊。

本論文提出一個以影像為基礎的圖書管理方法，利用大量佈建在書架附近的網路攝影機，拍攝書架中的書籍影像。這些書架影像經過一個以有限狀態機為基礎的方法將架上書籍的書背影像切割出來，然後擷取其 SIFT 特徵，並與資料庫預先儲存的圖書資料利用影像搜尋技術進行比對，得到目前仍在架上的圖書資訊。最後，根據盤點的結果立即更新資料庫，快速的掌握目前書店庫存的情形。

實驗後整體的準確率可以高達九成以上，證明此方法的可行性。比對錯誤大部分都出現在同系列不同集數的套書上，因為區分書本集數的影像特徵數量太少所導致。

關鍵詞—書店管理，盤點，影像搜尋

Abstract—To check the inventory in a bookstore, it always takes a lot of manpower and time to count the books one by one on the bookshelves. The high cost results in longer inventory check period, such that the bookstore managers can not know the accurate stock information anytime.

This paper proposes an image-based approach for bookstore management. A number of web cameras are uniformly distributed around the bookstore, and then the bookshelves' images are captured in inventory check. A finite state machine is applied to segment the images into individual book spine images, in which the SIFT features are extracted to search for the identities of the books on the shelves. After that, we can get the information about the books which are still on the book shelves. The database is updated immediately according to inventory check results to reflect the latest stock information in time.

The experiment results show that the book spine recognition rate is above 90%, which proves the feasibility of the proposed image-based bookstore inventory check system. Most errors come from the book series because the distinguishable features in them are fewer.

Keywords—bookstore management, inventory check, image search.

一、研究簡介

書店在營業的過程中一定會有進貨、售出、下架或遺失等情形發生，經過一段時間後書本數量必定與上架時不同，所以需要透過大規模的盤點來管理庫存數量 (inventory checking)。雖然大部分的書店都已經透過電腦化管理庫存，但還是需要人力清點總數再和電腦的統計結果比對，才能得到正確的統計數據。傳統的書店盤點方式幾乎完全依賴人力，為了確實執行盤點工作，在盤點期間書店通常會暫停營業，投入所有的人力進行盤點。對管理者而言，除了需要花費大量的時間與人力外，還損失了盤點期間的收入。為了減少損失，書店盤點的週期可能延長為半年一次甚至一年一次。減少盤點次數將導致書店無法掌控庫存的數量，當發現書本數量錯誤或遺失時，已經無法追蹤流向。

其他書店庫存管理技術包括使用 RFID (radio frequency identification) [5]，但是 RFID 標籤與相關硬體設備目前的成本過高，除非出版業上、下游都能配合，一般書店難以自行使用 RFID 進行庫存管理。

在本論文中提出一個以影像為基礎的自動化書店盤點系統，利用大量佈建在書架附近的網路攝影機，拍攝書架中的書籍影像，這些書架影像經過一個以有限狀態機 (finite state machine) 為基礎的方法將架上書籍的書背 (book spine) 影像切割出來，然後擷取其 SIFT 特徵 (scale invariant feature transform) [1]，並與資料庫預先儲存的圖書資料利用影像搜尋技術進行比對，得到目前仍在架上的圖書資訊，幫助書店能快速且準確地完成盤點工作。

首先，我們必須先設定相關參數並完成資料庫的初始化。所有參數的設定都在上架時完成，包括書架的層數 (row)、書架每一層分割成幾個欄位 (column) 來拍攝書架影像、PTZ 參數與控制點 (control point) 位置，這些參數將儲存在資料庫中提供盤點時使用。書架影像是利用網路攝影機根據資料庫中的 PTZ 參數拍攝取得，我們必須將書架影像中的書背影像個別獨立分割出來並儲存到資料庫中，提供盤點比對時使用。在拍攝書架影像時，因為網路攝影機拍攝角度的關係使得書架影像中的書本排列會呈現歪斜的現象，造成書背邊緣辨識錯誤而影響切割結果。因此，我們必須先將書架影像校正，讓書背影像在書架影像中水平排列整齊，提高影像切割的正確率。影像切割利用有限狀態機的概念，將一個完整的書背由左而右分成左邊界、左書脊、標題、右書脊、右邊界等五個部份，依照此順序判斷出書背的位置，最後再將書背從書架影像中分割出來並儲存於資料庫中。至此，整個系統的參數與資料庫的初始化都完成。

書店盤點時將重新拍攝書架影像，經過校正與切割後得到的書背影像即是查詢影像 (query image)。盤點時，給定系統一張書背查詢影像，系統會將此查詢影像對原資料庫中的書背影像做搜尋比對，目的是找出與查詢影像相同的書背影像。我們使用以圖找圖的方式來進行影像搜尋。搜尋的方法是採用 SIFT 特徵比對，此方法

可以克服影像雜訊、旋轉、放大、縮小等影響，我們希望藉由此強健的比對方法從資料庫中找出相似度最高的書背影像。實驗後，影像搜尋的正確率高達 98.7%，證明使用 SIFT 來作為影像搜尋的特徵是可行的。

根據搜尋比對的結果，我們能很快的得知書本目前的狀態。我們將書本庫存狀態依照搜尋的結果分成三類：1.架上、2.遺失或售出、3.新增書籍。表 1 為這三種分類的數學定義，其中 $S = \{s_1, s_2, s_3, \dots, s_n\}$ 為查詢影像集 (query image set)， $D = \{d_1, d_2, d_3, \dots, d_m\}$ 為資料庫的書背影像集 (database image set)。 s_i 與 d_j 皆為書籍的唯一編號 (unique ID)，例如 ISBN。

表 1 書本庫存狀態數學定義

狀 態	定 義
架 上	$\forall s_i \in S \cap D, 1 \leq i \leq n$
新增書籍	$\forall s_i \in S - D, 1 \leq i \leq n$
遺失或售出	$\forall d_j \in D - S, 1 \leq j \leq m$

這三種書本庫存狀態之間的關係整理如圖 1。

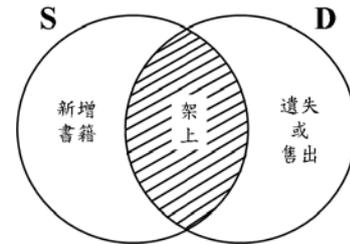


圖 1 書本庫存狀態關係圖

二、文獻探討

從書架影像中辨識書背影像常見的作法都是初步以邊緣偵測 (edge detection) 判斷書背的位置，再針對每個書背影像作文字偵測與辨識 (optical character recognition, OCR)，以文字來代表書背影像的內容，如論文 [8]。書背切割通常先偵測兩本書籍的邊界，使用邊緣偵測再用霍夫變換 (Hough transform) 偵測長的直線邊界為書

背影像的邊界，這種方法容易受到雜訊與光源影響。我們採用論文 [2] 提出的方法，以有限狀態機嚴謹的定義了書架影像中的元件，並加入書籍顏色的考慮，以動態規劃 (dynamic programming) 來求出整體最佳解，得到精確的書背位置。雖然動態規劃需花較久的時間來計算，但一個自動的盤點系統，必須偵測出大部分的書背，否則進行盤點時會因偵測失誤而影響盤點結果。

書背辨識方面可以對書背影像作文字偵測與辨識，再以文字搜尋的技術來盤點。這種方法受限於影像的解析度，若書背影像上的字不清晰就會造成文字辨識的困難；另外現在的書背樣式五花八門，甚至有些書背影像上是以圖片或是出版社的商標為主要特徵。本論文以整張書背影像為基礎，利用 SIFT 特徵來描述每個書背影像，並以影像搜尋來確定書籍的唯一編號。因為 SIFT 特徵的對於各種雜訊、尺度縮放、旋轉等環境變化的強健性，適合來描述這些細部的影像特徵，使得搜尋相似書背影像能更精確。

三、書店自動化盤點系統

本論文提出的圖書自動化盤點系統可分成上架與盤點兩部份，流程圖如圖 2。

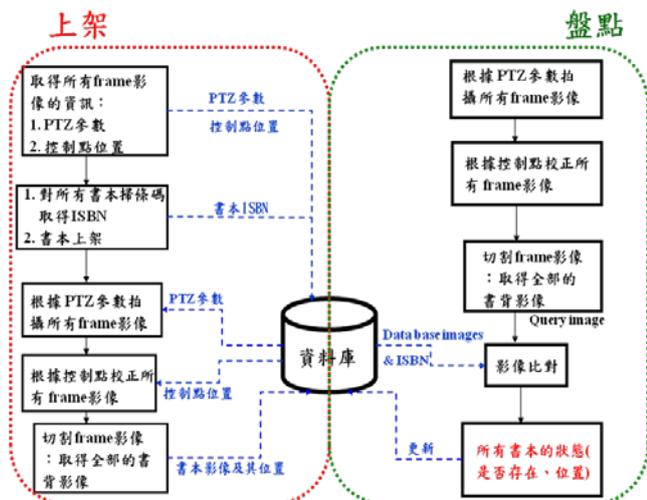


圖 2 書店自動盤點流程圖

(一) 名稱定義

- **書架影像 (frame):** 網路攝影機拍攝書架得到的每張影像都稱為書架影像。
- **row:** 書架的層數。
- **column:** 書架每一層分割的欄位數。
- **PTZ 參數:** 書架影像的 PTZ 參數定義了拍攝時網路攝影機的水平轉動 (pan)、垂直轉動 (tilt) 和縮放 (zoom) 參數。
- **控制點:** 在拍攝書架影像時，因為網路攝影機拍攝角度的關係使得書架影像中的書本排列會呈現歪斜的現象，所以必須取得各個書架影像的控制點，以利後續作幾何校正 (geometric rectification)。控制點是透過人工的方式選取，如圖 3，詳細流程請參考 (三) 影像校正。



圖 3 控制點位置 (紅點處)

(二) 攝影機參數設定與書本上架

上架的第一個步驟必須完成攝影機參數的設定與資料庫的初始化。對於一台 PTZ 網路攝影機所涵蓋的拍攝範圍，使用者必須自行設定書架的 row 和 column 兩個參數，這兩個參數會將拍攝範圍劃分成 row×column 個書架影像。假設所有書架影像是共平面，我們可以利用四個角落的書架影像的攝影機參數利用內插法來得到其他位置的書架影像攝影機參數。在書本尚未上架之前，先移動網路攝影機鏡頭至書架四個角落的書架影像位置，記錄這四個角落的 PTZ 參數與控制點位置，其它區塊的 PTZ 參數與控制點位置則是利用內插法 (1) 計算求得。

$$p(i) = p(1) + \frac{p(n) - p(1)}{n-1} \times (i-1), \quad 1 < i < n \quad (1)$$

上式中 $p(i)$ 為攝影機參數， $p(1)$ 及 $p(n)$ 為已知兩個角落的攝影機參數。以圖 4 為例，書架四個角落的影像參數分別以 $T(1)$ 、 $T(2)$ 、 $T(3)$ 、 $T(4)$ 表示。首先，對這四個參數作縱向內插運算（步驟 1），求得 column 1 與 column 4 所有書架影像的參數，接著再做橫向內插運算（步驟 2）即可求出所有書架影像的參數。

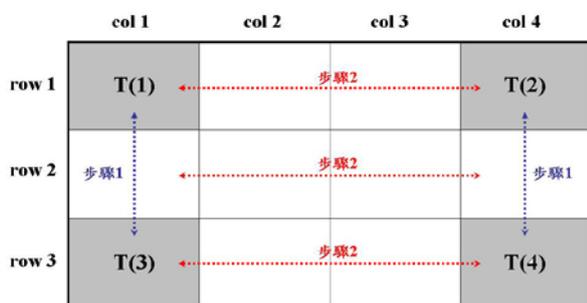


圖 4 內插法執行說明

原則上參數設定只需要執行一次即可，這些參數都會儲存在資料庫中提供以後盤點時使用，除非書架或攝影機的位置被移動過才需要重新設定參數。參數是此管理系統中相當重要的元件，若是參數選擇不當，很可能會影響最後的執行結果。

參數設定完成後即可將書本上架，在書本上架的同時也要將書本相關的資訊儲存到資料庫中，例如書本的 ISBN。書本的 ISBN 具有唯一性，可用來配合書店的 POS 系統作庫存管理，所以每本書在上架時都必須先用條碼機掃描書本條碼，取得書本 ISBN 後才將書本上架。所有的書本上架後，使用網路攝影機讀取資料庫中的 PTZ 參數，攝影機會根據 PTZ 參數移動鏡頭拍攝書架影像，如圖 5。書本上架後，必須將每本書的書背影像及特徵記錄在資料庫中以利後續利用書背影像來搜尋。擷取書背影像與特徵的方法如後面小節所述。

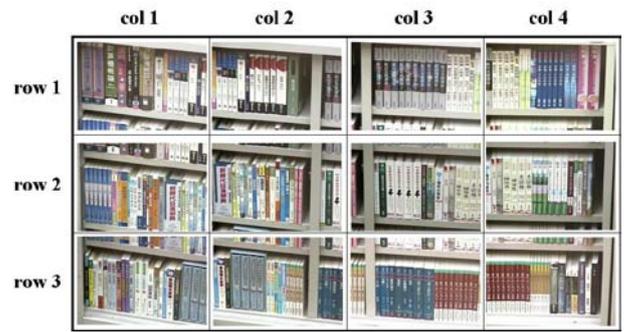


圖 5 書架影像

(三) 影像校正

拍攝書架影像時，因為攝影機拍攝位置與角度的關係使得書本的排列在書架影像中呈現歪斜的現象。為了校正書背影像，我們利用控制點對應得到影像幾何轉換矩陣 T 。幾何轉換就是根據影像中像素之間的空間關係，將影像從原本的座標系統轉換到新的座標系統中，此轉換通式可表示成 (2)：

$$(x, y) = T\{(w, z)\} \quad (2)$$

如圖 3 所示，我們將一個矩形的盒子放置在空書架的四個角落中，利用此盒子向外那面的四個角當作控制點的位置，讓使用者在影像中選取控制點時有個依據。因為這四個角對應到真實座標系統的矩形區域，可以測量得知其相對座標，便於計算從影像座標系統轉換到真實座標系統的幾何校正所需轉換矩陣 T 。最後，我們從校正後的書架影像中擷取書背的部份，目的是排除其他背景（例如校正後的黑色區塊，書架的框架）的影響，藉此提高書背切割的準確率，流程如圖 6。



圖 6 校正流程 (T 為轉換矩陣)

(四) 影像切割

校正後的書架影像必須經過切割才可以得到一本本獨立的書背影像。上架時每一個書背影

像都對應到一個 ISBN；盤點時則將切割好的書背影像與儲存於資料庫的書籍書背影像作比對，比對完再更新資料庫。

1. 書背偵測

本論文的書背偵測方法是參考 [2] 所提出的演算法並加以改進，其中使用到有限狀態機與動態規劃 (dynamic programming) [7]。根據觀察，在自然的書架影像中每個像素可以對應到六種物件類別之一，包括：左、右書籍邊緣 (left, right boundary)、左、右書脊 (left, right spine)、標題 (title) 及背景 (background)，所以我們在有限狀態機亦定義了六種狀態 $Q = \{q_1, \dots, q_6\}$ ，如圖 7。根據定義的狀態轉換模型 (state transition diagram)，我們將影像中的書背切割問題描述成序列狀態 (sequence $s_1, \dots, s_i, \dots, s_M$ where $s_i \in Q$) 的最佳化問題。如圖 8， s_i 代表影像中央水平線上每個位置點 i 的狀態，而序列 p_i 則是描述每個位置 i 對應到影像中書籍擺放結構的最可能頂端位置，位移範圍由 $-W \sim +W$ (sequence $p_1, \dots, p_i, \dots, p_m$ where $p_i \in [i-W, i+W]$)，所以 p_i 與 i 的連線決定書本的角度，然而要如何決定 p_i 的位置，就必須由此連線是否正確對應到影像中的物件類別來判斷。

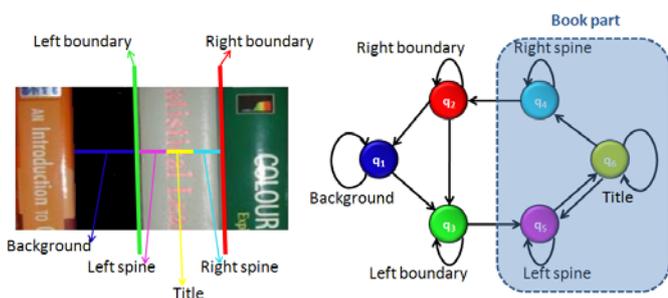


圖 7 書架影像內的物件與狀態轉換模型

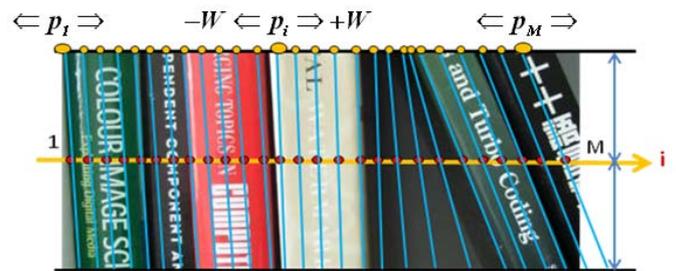


圖 8 串列狀態轉換

觀察書架影像中六種物件類別的邊緣特性是不同的。利用 Canny 邊緣偵測 [4] 將書架灰階影像作水平與垂直的邊緣偵測，如圖 9，可發現每種物件不同的邊緣特性，如書本左右的邊界上有很多垂直邊但沒有水平邊，我們將書架物件類別之邊緣的特性整理成表 2。

表 2 書架物件類別之邊緣特性

垂直邊緣	多		少	
	多	少	多	少
書籍邊緣		○		
標題	○		○	
書脊與背景				○



(a)



(b)

(c)

圖 9 (a) 書架影像；(b) 書架影像水平邊緣；(c) 書架影像垂直邊緣

所以根據表 2 的書架物件特性可以定義三個函數，來代表三種物件：

$$\begin{cases} d_N(q_1, p_i) = d_N(q_4, p_i) = d_N(q_5, p_i) \\ \quad = (2 \times \text{edgeMax} \times \text{numPixel} - \text{hStr} - \text{vStr}) / \text{numPixel} \\ d_N(q_6, p_i) = (\text{hStr}) / \text{numPixel} \\ d_N(q_2, p_i) = d_N(q_3, p_i) \\ \quad = (\text{vStr} + \text{edgeMax} \times \text{numPixel} - \text{hStr}) / (2 \times \text{numPixel}) \end{cases} \quad (3)$$

其中，由於邊緣偵測後會做二值化的動作，所以書架影像邊緣偵測後的像素值皆為 0 或 1，故 edgeMax 即代表最大的像素值為 1。代入不同的狀態 s_i 與書頂位置 p_i 後，統計該連線所經過的像素個數 (numPixel) 及連線上的水平邊緣 (hStr) 與垂直邊緣 (vStr)，依據這些數值特性分辨出邊界、標題、書脊與背景，使其中一函數得到最大值，則此狀態很有可能是該連線最有可能的狀態。

只觀察一個水平位置 i 上的函數 $d_N(s_i, p_i)$ ，找到 s_i 與 p_i 使這個函數得到最大值，只能得到局部最佳解。我們必須對整張書架影像的每個水平位置，整體求出最佳解。利用動態規劃演算法 [6] 並配合定義好的有限狀態轉換機，在所有的位置點 i 上找出一條具有最大值的路徑，我們可以寫成公式 (4)。

$$\max \sum_{i=1}^M d_N(s_i, p_i) \quad (4)$$

另外，依據常理當路徑進入書籍本體部分 (book part) 後，其書本可傾斜的角度是固定的，所以頂端 p_i 可根據上一個位置 p_{i-1} 來決定，此限制條件可以寫成公式 (5)：

$$p_i = \begin{cases} p_{i-1} + 1, & \text{if } s_i \in \{q_4, q_5, q_6\} \text{ or } s_{i-1} \in \{q_4, q_5, q_6\} \\ p_{i-1} + \{0, 1, 2\}, & \text{otherwise} \end{cases} \quad (5)$$

根據上述的定義設計屬於此問題的動態規劃演算法示意圖如圖 10，追蹤一條擁有最大權重的路徑，此路徑上每個位置的狀態就是書櫃影像中各點位置所代表的元件，之後再利用書本的左右邊界來切出一本本獨立的書背影像。

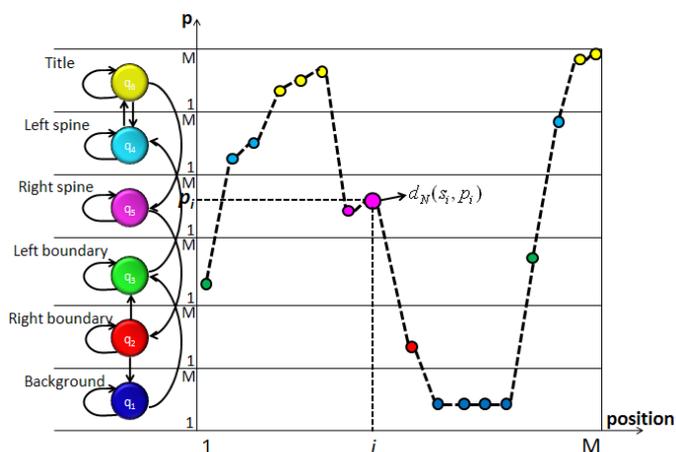


圖 10 動態規劃演算法示意圖

2. 色相的考慮

根據上一節的方法來偵測書架影像中的物件會有無法判斷顏色差異的現象，例如兩本顏色不同的書，程式會判斷為同一本書的不合理現象，如圖 11(a)，這是因為在作邊緣偵測時只有用到影像的明暗度，若兩本相鄰的書顏色不同，但是亮度值相似，就無法偵測出在灰階影像下的書籍邊界。

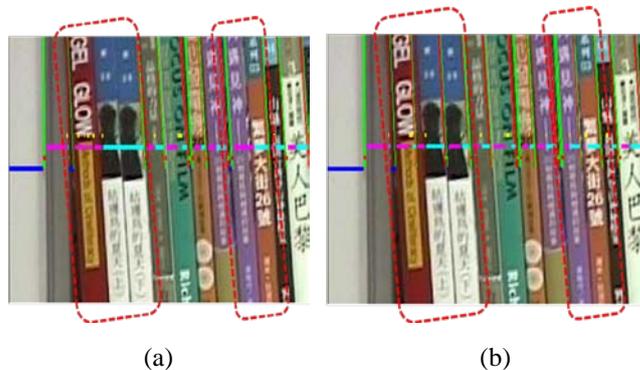


圖 11 (a) 書背邊緣判斷錯誤；(b) 加入色相修正

為了改進這個問題，我們必須考慮顏色的差異。由於 RGB 顏色空間並不是對應到人類感官均勻分布 (perceptually uniform) 的距離空間，所以將書架影像的顏色空間由 RGB 轉換成 HSV，如圖 12，並拿出色相層 (hue) 來作為色差計算。色相值是由角度來表示不同的顏色，我們定義了色相值之間的差距公式 (6)：

$$d_H(H_i, H_{i-1}) = 1 - \cos(2 \times \pi \times (H_i - H_{i-1})) \quad (6)$$

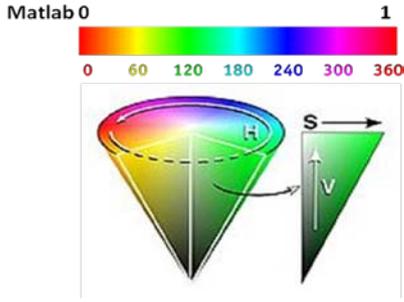


圖 12 HSV 色相值分佈

計算書架影像每個位置點 i 與其對應之書頂 p_i 連線上的平均色相值 H_i ，相鄰位置點 i ， $i-1$ 的平均色相值變化的大小就可以當作動態規劃演算法的權重，加入此相鄰位置色差權重後的動態規劃演算法如下 (7)。

$$\max \sum_{i=1}^M d_N(s_i, p_i) + d_T[(s_i, p_i) | (s_{i-1}, p_{i-1})] \quad (7)$$

$d_T[(s_i, p_i) | (s_{i-1}, p_{i-1})]$ 的定義如公式 (8)。書本右邊界轉換到左邊界的情況，直接將色相差當作權重，差值越大越有可能是兩本書的邊界。如果是在書籍本體狀態內轉換，可能因為標題的關係而有些許色相差。如果在書籍本體部分與非書籍本體部分 (non-book part)，狀態沒有改變，則給固定的權重值。在合理的情況下，書本部分因為是在同書背的顏色下，所以顏色變化較小，權重值也較小；非書本部分包含了較多的雜訊，所以顏色變化較大，權重值也較大。

$$d_T[(s_i, p_i) | (s_{i-1}, p_{i-1})] = \begin{cases} d_H(H_i, H_{i-1}), & \text{if } s_i = q_3 \text{ and } s_{i-1} = q_2 \\ 1 - d_H(H_i, H_{i-1}), & \text{if } s_i = s_{i-1}, \text{ and } s_i \in \{q_4, q_5, q_6\} \\ 1 - \cos(5^\circ), & \text{if } s_i \neq s_{i-1}, \text{ and } s_i, s_{i-1} \in \{q_4, q_5, q_6\} \\ 1 - \cos(15^\circ), & \text{if } s_i \neq s_{i-1}, \text{ and } s_i, s_{i-1} \in \{q_1, q_2, q_3\} \end{cases} \quad (8)$$

圖 13 中，將色相值的變化量加入動態規劃所經路徑的權重時，可以追蹤出一條更符合書架影像內容的路徑。根據上述的改良，圖 11 (a) 中無法被偵測出的書本邊界便可以正確偵測出來，如圖 11 (b)。

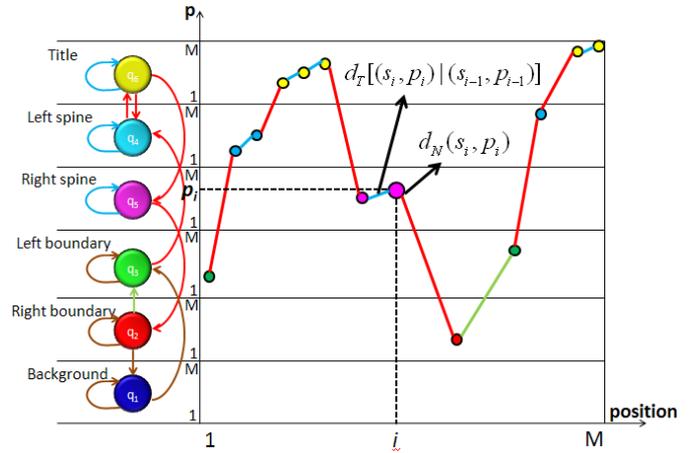


圖 13 加入色相的動態規劃路徑示意圖

3. 書背切割

根據以上方法偵測出的書背左右邊界，我們可以切割出一本本獨立的書背影像，但並非所有書本都是擺放整齊，部分的書本可能為傾斜擺放，所以初步以左右邊界切出的影像必須作幾何校正。利用左右邊界的兩條線方程式就可以算出書背影像的四個頂點位置，再以此四點作為控制點，對應到真實座標上的一個矩形四個角落的座標點，便可以利用這些控制點對應關係計算出轉換矩陣，利用 (三) 影像校正的幾何校正方式就可以讓傾斜的書背校正，如圖 14。

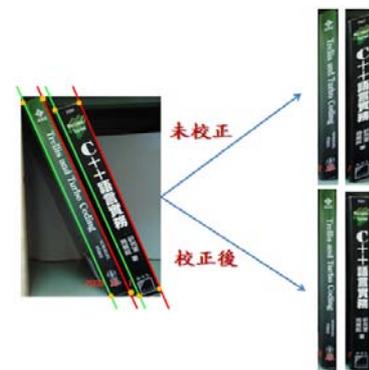


圖 14 書背校正

由於書背偵測仍會有其他誤判情形，必須在初步切割後在做額外的修正，我們整理了以下須作後處理的誤判情形，配合圖 15 作解釋，並且一一的敘述方法。

- i. 非書背誤判為書背 (如：書之間的縫

隙、書櫃隔層)。

- ii. 書背影像的標題部分水平邊太少而誤判為背景。
- iii. 多本書誤判為一本，常出現在同系列或書背顏色相同的書本擺放在一起之情形。



圖 15 誤判情況

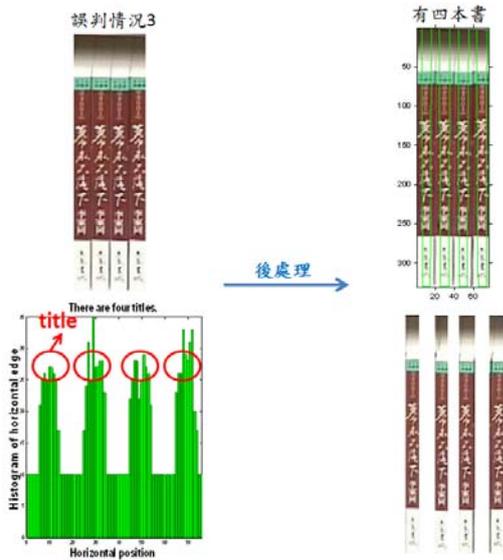


圖 16 書背後處理

首先，我們使用 Sobel 邊緣偵測 [3] 對切割後的書背影像作水平邊緣偵測，再將邊緣化的數值作垂直投影，此時我們訂定門檻值，水平投影的數值若大於門檻值則表示標題的部位，小於則代表沒有標題，如此一來就可以找出在切割出來的書背影像中有幾個標題，也就代表有幾本書，利用此方法來作為後處理核心：

- i. 非書背影像 (如：書縫) 的寬度小於真實書本的正常厚度，或非書本影像 (如：書櫃夾層) 的水平邊緣經過水平投

影後皆小於門檻值，故這些誤判的書背影像都可被排除。

- ii. 誤判的背景影像經過後處理的核心方法，找出影像內含有標題，即可認定為書背影像。
- iii. 多本書認定為一本的書背影像經過後處理後也可以判斷出影像內有多少標題即為幾本書，判斷後再依每本書標題的寬度切割書背。如圖 16。

由以上的作法我們將其整理成決策樹 (decision tree)，如圖 17，使得在書背切割的後處理部分可以更加嚴謹。

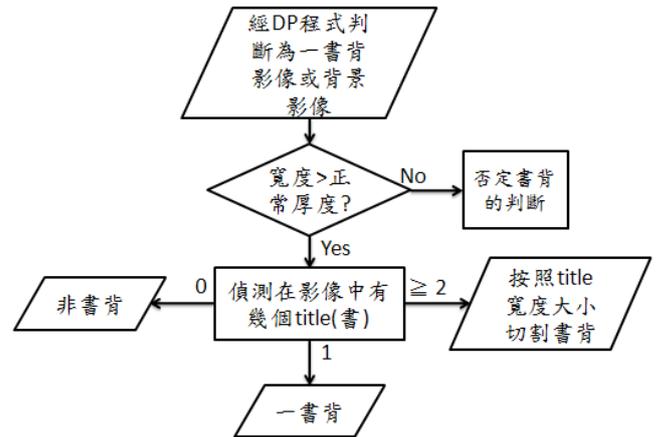


圖 17 書背後處理決策樹

4. 書背比對

本論文在影像比對採用 SIFT 特徵，使用的原因是因為 SIFT 特徵相當強健，是具有尺度不變性的局部特徵描述子 (local descriptor)，可以適用於 PTZ 攝影機所拍攝出書背影像較低的解析度，另外由於書背影像的內容大部分都以字、logo 或是出版社的商標為主要特徵，因此很適合用 SIFT 來描述細微部分所代表的特徵。

SIFT 特徵是使用雙高斯差函數 (difference of Gaussian) 找尋影像中感興趣的關鍵點 (keypoint)，以關鍵點為中心對周圍切割成 4×4 個的矩形方塊，再統計每個方塊在每個方向上邊的個數，將 360 度的方向量化成 8 個，所以每一

個關鍵點會有 128 維向量的區域描述子，根據這些邊方向的統計值找到一個擁有最大統計值的方向，而此最大統計值就是關鍵點的大小 (scale)，其方向就是關鍵點的方向 (orientation)。每張圖會有許多關鍵點，而每個關鍵點會有自己的區域描述子，其形式如圖 18：

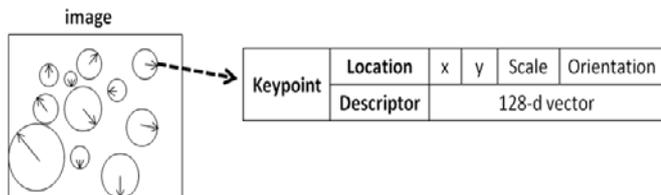


圖 18 SIFT 特徵的區域描述子

兩張圖片在比對時，是以關鍵點根據 128 維的描述子在作配對，關鍵點描述子距離相近的就可能配對成功，而兩張圖如果有越多的關鍵點配對成功，代表圖片內容越相似。而配對成功的法則是一張圖片的其中一個關鍵點與另外一張圖的所有關鍵點計算距離，距離最近的與第二近的比值必須小於 0.6，否則即使距離再近也都不能算是配對成功。

比對時查詢影像與資料庫裡的每一筆影像逐步作特徵比對，此時我們必須先定一個門檻值，兩張影像的特徵配對數必須大於此門檻值才認為是相同，因為書背比對是用於盤點系統，必須找出一模一樣而非相似的書背影像，所以影像特徵若沒有大於門檻值，系統就認定為找不到，代表此書沒有在資料庫裡，可能為上架後的新增書籍。在比對之前必須對儲存於資料庫的上架書背影像作特徵擷取，再以盤點時書背影像的特徵查詢在資料庫內的特徵。

在特徵配對時常常會發生一張查詢影像的多個特徵配對到一張資料庫影像的同一個特徵；或是相反的，一張搜尋影像的一個特徵配對到一張資料庫影像的多個特徵。如此一來，會造成比對上的錯誤，為了排除這個問題，在特徵配對時，無論是多對一或一對多的特徵配對到，我

們都只算一個配對。

同一個 ISBN 編號的書籍在上架時可能會放置多本書，因此資料庫中會有多個書背影像是對應到同一個 ISBN 編號。我們可以讓一個 ISBN 編號的書籍只保留一個書背影像作比對，如此可以增加搜尋速度；另外一種作法是保留同一個 ISBN 編號的多個書背影像，因為保留這些不同拍攝情況下的書背影像作比對，可以增加搜尋的準確性。當查詢影像配對到資料庫某一個最相似的書背影像後，會轉換到其 ISBN 編號，以作庫存管理。

四、實驗

(一) 實驗環境

我們以三個鐵製書櫃作為測試環境，約可放置 200 本左右的書籍。為了測試不同解析度的書架影像對於效能的影像，我們會使用的解析度較低的 PTZ 網路攝影機以及解析度較高的數位攝影機作為取像裝置，硬體規格如表 3 所示。

表 3 攝影機規格

型號	AXIS 213 PTZ	SONY W5
解析度	704×480	2592×1944
亮度	40	自動測光

(二) 實驗結果

這部分主要有兩個重點來評估本方法的好壞，一個為書背切割，另一個為書背比對，書背切割的效能會影響比對的結果，以下分為兩小節來探討。

1. 書背切割

本論文提出的自動化盤點系統是基於佈建在書店的網路攝影機，所以自動書背切割以 PTZ 攝影機拍攝的書架影像為主。我們以 Recall 及 Precision 來評估切割的效能，先定義好 true positive、true negative、false positive、false negative

如表 4。

表 4 效能名稱定義

true positive (TP)	正確切出書背
true negative (TN)	無
false positive (FP)	切出非書背部分 (如：書櫃或書與書夾縫)
false negative (FN)	不正確切割 (如：兩本以上被切成一本)

再將切割好的影像以此四種屬性來算出 Recall 及 Precision，如公式 (9)。

$$\begin{cases} \text{Recall} = \frac{TP}{TP + FN} \\ \text{Precision} = \frac{TP}{TP + FP} \end{cases} \quad (9)$$

接著，利用 Recall 與 Precision 算出 F1 (10)：

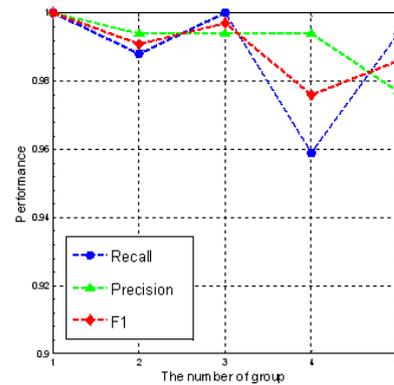
$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (10)$$

本論文的主要貢獻是加入色相值於有限狀態機中以及書背偵測完的後處理，所以我們必須比較原始演算法與改善後的效能，如表 5。

表 5 改善前後效能比較

	原始方法[2]	本論文方法
Recall	0.757	0.977
Precision	0.942	1.0

由以上的實驗結果可以證明本論文提出方法的可行性，而且 Recall 與 Precision 都接近 1，如此一來可以確保大部分的書可以被自動切割出來，比對時不會因為切割的效能不好而受影響。接著，為了驗證此方法可以符合許多情況，我們將書櫃上的書隨機擺放並使用 PTZ 攝影機拍攝與校正後作自動書背切割，實驗五次來計算平均數據，圖 19 為五次的效能數據，橫軸為組別，縱軸為效能數值。



	平均值
Recall	0.988
Precision	0.992
F1	0.990
Time(s)	537.1

圖 19 五次的效能數據

實驗數據說明，在五次的切割中，無論是 Recall、Precision 還是 F1 都在九成以上，而且 F1 也達到 99%，證實了本論文提出的切割方法可符合各種書本擺放的情況。

2. 書背比對

由於是將影像比對使用在盤點的功能上，我們必須模擬書店盤點時書本的狀況，所以上架完成後，將書籍重新擺放，並且從中抽走書籍及加入新書籍，完整的實驗流程如圖 20；本次實驗上架時的書本有 173 本，書架上的書本經過弄亂後，盤點時的書本有 172 本，PTZ 攝影機拍攝並切割後，上架時的書背影像皆成功切割，盤點時成功切割出 165 本。

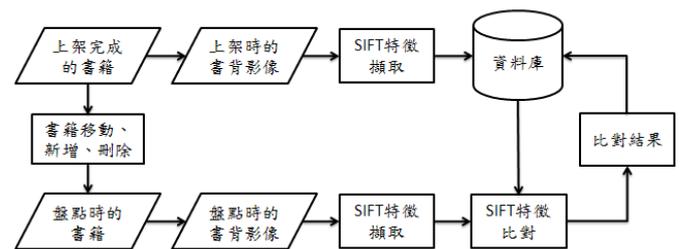


圖 20 書背比對實驗流程圖

為了證明 SIFT 特徵能適用在 PTZ 攝影機較低的解析度下，我們亦使用高像素的相機拍攝書背當作對照組，並使用手動書背切割來排除自動切割會有誤差的情形而影響比對的結果；將 PTZ 攝影機與相機拍攝的書背影像作 SIFT 特徵擷取後的相關統計整理成表 6，我們可發現相機

拍攝的每張書背影像的平均維度為 1602×132，遠大於 PTZ 攝影機拍攝的 341×23，所以相對的，相機在 SIFT 平均特徵數有 2000 個左右，遠多於 PTZ 攝影機拍攝的書背，只有接近 100 個，在如此少的特徵數下，SIFT 特徵是否可以比對成功？這也是我們此實驗設計的目的，更是此系統可行性的依據。

表 6 SIFT 特徵數量統計表

	AXIS 213 PTZ		SONY W5	
書背影像 平均維度(/張)	341×23		1602×132	
SIFT 特徵數量	上架	盤點	上架	盤點
平均數目(/張)	95	97	2213	2034
總數	16372	16027	383005	332385

在比對的效能評估方面，由於搜尋結果必須是正確的書籍書背影像而不是相似書籍的書背影像，所以系統只需要對影像搜尋結果的第一名來判斷比對是否正確，因此評估的方法我們只取 Top1 的結果來計算 Recall、Precision 及 F1。我們的實驗環境中有多本同樣的書籍在架上，所以在搜尋到最相似的書背影像後，會轉換到其 ISBN 編號，來做庫存數量的統計。表 7 為書背比對的效能，搜尋時間如表 8。

表 7 書背比對的效能

	AXIS 213 PTZ	SONY W5
Recall	0.955	0.988
Precision	0.987	0.994
F1	0.971	0.991

表 8 搜尋時間比較

	AXIS 213 PTZ	SONY W5
平均搜尋時間	每個書背影像： 0.4510 s	每個書背影像： 282.5 s
	每個 SIFT 特徵： 0.0052 s	每個 SIFT 特徵： 0.1238 s

由實驗數據得知，相較於高像素的相機下，

PTZ 攝影機雖然解析度較低，但使用 SIFT 特徵比對的結果都在九成以上，不過還是有些微的落後，主要原因是相機拍攝的影像比對是在完美的情況下，包括了高像素及手動切割，然而本論文所強調的是一套全自動的盤點系統，所以我們必須使用 PTZ 攝影機及自動切割來達成，在搜尋時間上，因為 PTZ 攝影拍攝的書背影像 SIFT 特徵數遠少於相機拍攝的，相對的，其搜尋所需時間也遠少於相機拍攝的書背影像。整合了效能與搜尋時間的評估，此實驗證明了本論文的可行性。

五、結論

本論文提出了一套以影像為基礎的書店自動化盤點系統，此系統利用佈建在書店的攝影機，來拍攝架上圖書進行自動辨識與盤點。主要的影像處理包括書背切割與書背辨識。在書背切割部分改善了論文 [2] 所提出的方法，加入了色相值的參考與後處理，使得書背切割相較於原始方法能更精確的切割書背影像，可以達到九成以上的正確率。雖然書背切割還未達到百分之百的準確率，有少部分多本書背影像可能被切在一起，但是這部分的錯誤可以在書背比對時解決，因為 SIFT 特徵並非是比對影像的整體，而是注重影像中每個局部關鍵點的特徵，所以書背影像中只要含有相同的特徵就可能被比對出來。例如：使用兩個未被分割的書背影像當作查詢影像，在使用 SIFT 特徵比對時，還是可以個別比對出來資料庫的這兩本書籍的書背影像。然而這會造成一張盤點時的查詢影像對應到多張資料庫的書背影像，此部分就必須利用書籍的一些資訊，例如書本頁數或厚度來判斷，再由使用者回饋 (feedback) 資訊給系統。

在書背比對的部分，我們使用 SIFT 特徵比對作為影像搜尋的排序，不管是高解析度或低解析度的影像，都可以達到九成以上的正確率。我們發現比對部分發生的錯誤，通常是因為同系列

但不同集的書背影像，這些書籍的書背影像差異往往只有差在集數的部分，如圖 21。

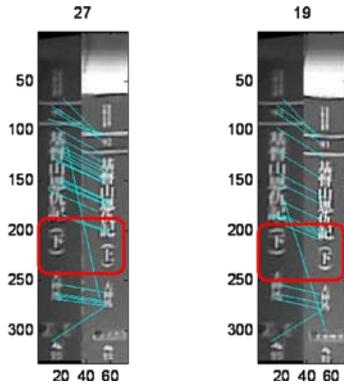


圖 21 集數上的特徵

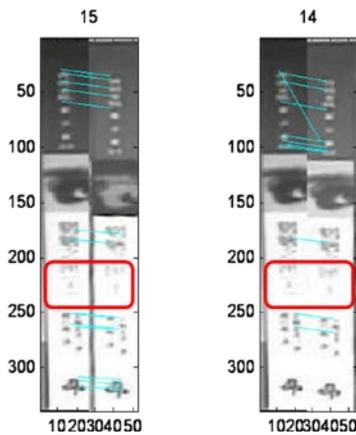


圖 22 模糊的集數

而在這些集數上可能沒有特徵，或是即使有特徵也無法配對，或甚至有些集數的特徵在影像中是很不清晰的，如圖 22，所以導致比對時的錯誤，而未來我們將對這些影像作後續的處理，如可以針對集數的區域作其他特徵的比對，或是

相同地可以透過使用者回饋資訊等方法來改善。

六、參考文獻

- [1] D. G. Lowe, "Object recognition from local scale-invariant features", Proceedings of the Seventh International Conference on Computer Vision, vol. 2, pp. 1150-1157, Sept. 1999.
- [2] E. Taira, S. Uchida, H. Sakoe, "A model-based book boundary detection technique for bookshelf image analysis", Kyushu University, Japan.
- [3] I. Sobel, G. Feldman, "A 3x3 isotropic gradient operator for image processing", presentation for Stanford Artificial Project, 1968.
- [4] J. Canny, "A computational approach to edge detection", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-8, pp. 679-698, Nov. 1986.
- [5] J. T. Lin, J. L. Hou, W. C. Chen, C. H. Huang, "An RFID application model for the publication industry: a Taiwan perspective", International Journal of Electronic Business Management, vol. 3, No.2, 2004.
- [6] J. R. Deller, J. G. Proakis, J. H. Hansen, "Discrete-Time Processing of Speech Signals", Macmillan, 1993.
- [7] R. Bellman, S. Dreyfus, "Applied Dynamic Programming", Princeton University Press, 1962.
- [8] X. Chen, J. Yang, J. Zhang; A. Waibel, "Automatic detection and recognition of signs from natural scenes", IEEE Transactions on Image Processing, vol. 13, pp. 87-99, Jan. 2004.