

印章真偽辨識系統實作

指導老師：曾建誠 教授

參賽組員：周佑儒、童昊煜、林世民、張健鋒

摘要

本專題實作出一套可辨識印章真偽的系統，最典型的應用便是銀行的存簿提款，客戶須在提款單中蓋印章，然後由銀行人員用目視比對印章的真偽，只有當印章為真時，才可提款，此工作不僅耗時，而且容易出錯，若能實作出一套可辨認印章真偽的自動化系統，不僅速度快，而且不容易因疲勞而出錯。所發展出來的辨識系統，包含下列四大部分，第一部分是利用掃描器將提款單轉成數位影像，完成取像的工作。第二部分是影像前處理，其功能包括印章的定位與切割，旋轉校正，雜訊消除，和大小正規化等。第三部分是利用互斥或(XOR)運算子來計算提款單輸入印章和存簿印章的相似度，當相似度高過事先預設的門檻值，便可視輸入印章為真，不過此部分須事先知道客戶的帳號或身份証字號，方可調出存簿印章來進行相似度比對，為改善此缺失，我們進一步發展印章辨識功能來自動調出存簿印章。第四部分是利用傅立葉轉換或互斥或來自動辨識輸入印章的身份，而非用人工方式來輸入客戶的身份資料。整套系統以視窗的圖型使用者界面與外界溝通，銀行人員只須輕按一下鍵(one-touch)，便可完成提款單中印章真偽的判斷工作，操作方便又快速。

前言

1.1 研究動機

在現今的社會中，人們身上往往都會有一些生活中必需的小東西，比如像身分證、健保卡、駕照和印章等等。由於現今詐騙集團行騙手法日新月異，像身分證跟印章等重要個人文件的辨識，就顯得非常重要。

由於中華民國國民身分證在製作的過程中，將使用者本人的照片蓋印在上面，所以在使用的過程中，只要比對照片就很容易辨別出來是否為本人使用。相對之下，印章無法辨別為使用者本人使用，當印章在刻印時，店家並不會去確認是否為當事人，但是據分析結果得知，每家印章店刻印出來的字型並不會完全相同，即使是同一間店、同一台機器所刻出來的印章，也不會一模一樣。因此印章跟指紋一樣，有自己的獨立性，由機器的痕跡跟材料的痕跡可以辨識出不一樣的地方。

可惜的是，人們無法精確辨認出微小的差異性，於是將問題託諸於精細度高的電腦來辨識，所以需研發出一套軟體，事先將印章的資料庫建立起來，然後再將所要辨識的印章輸入到電腦裡，利用文獻中數位影像處理的技術 [1]-[10]，就可辨識出是否是同一個印章，最典型的應用便是銀行的存簿提款，便須要印章真偽辨識技術。

1.2 研究目的

在現今的社會中，人們慢慢注重理財的觀念。講到理財，有很多種方法，不過，最基本的就是郵政儲金或銀行儲金。在存款之前，必須要先在當地的郵政或銀行開戶。一般個人開戶時，只需攜帶國民身分證、第二證件(健保卡或駕照)及印鑑即可辦理。

由於郵政儲金或銀行儲金在現今的社會中非常的普遍，所以在存簿的印章辨識中就顯得特別重要。當需要申請金融卡及使用提款單提款時，就必須使用到印章。此時，就是要辨識使用者所使用的印章是否為開戶印章(註:開戶印章就是指當初開戶時所使用的印章)，必須使用開戶印章才可辦理提款手續。

因此，此專題的目的就是要用電腦來辨識提款單上所蓋的印章跟開戶印章的相似度有多高，如果辨識出來的結果，相似度高於九成以上，即可判定是相同的印章，便可以辦理提款手續，其方塊流程圖如圖 1.1 所示。次要目的就是提高辨識的速度，使用上的方便性和正確性，提升金融業者在數位化之後，不只管理上更加方便，連處理速度也比以前更加快速、方便。

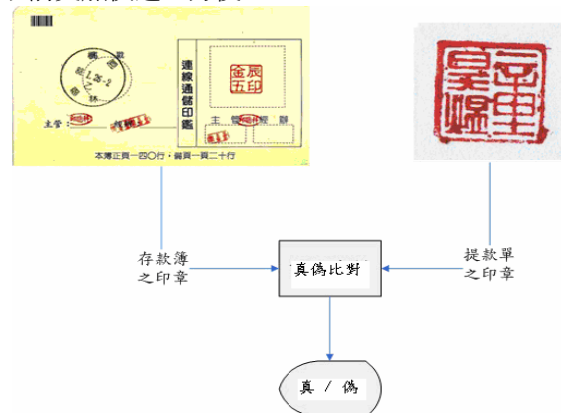


圖 1.1 提款時所須之印章真偽辨比對工作

本專題所提自動辨識提款單印章真偽的系統流程圖如圖 1.2 所示，共包括取像，前處理，印章辨識，相似度計算，和真偽判斷等工作，其中取像工作由掃描器完成，真偽判斷只要相似度高於 9 成，便視為真，故底下各小節只對前處理，相似度計算，印章辨識等工作做說明，值得一提的是存簿印章已被事先掃描成電子檔，存在資料庫內，只要有客戶之帳號資料，便可調閱，目前中國信託商業銀行便是以此方式來檢視印章真偽，不過，為了自動化，本專題利用印章辨識來調出存簿印章，來取代以手動方式輸入客戶帳號獲得存簿印章，而達到完全自動化的目的。

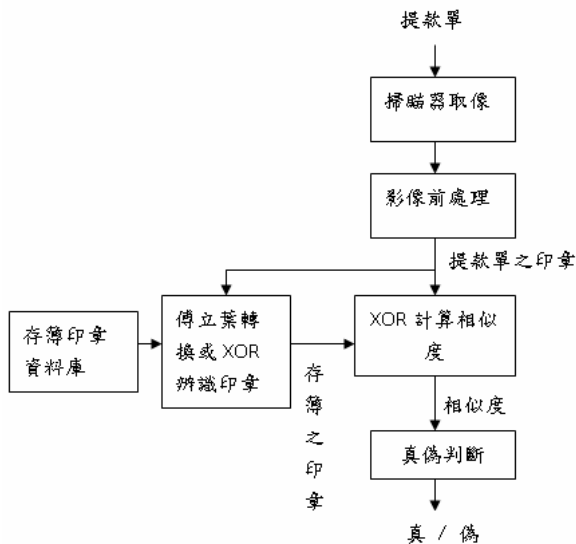


圖 1.2 提款單真偽自動辨識流程圖

2. 影像前處理

影像前處理共分為二值化、雜訊抑制、印章定位、印章轉正及大小正規化等五個部分，其流程圖如圖 2.1 所示，現分別說明如下：

為了加快程式運作效能，須先將輸入影像做二值化，而且輸入影像可能因人為操作導致紙面上的污點，或者因轉正過程中產生的鋸齒雜訊，故須以邊緣偵測和膨脹侵蝕的方式來減少雜訊對於印章定位、辨識的不利影響。

印章定位主要在影像上尋找出印章所在位置，因為印章為影像中最大的物件，所以取出最大的物件，就能找出印章。

在確定印章正確位置後，所要處理的是印章旋轉問題，處理方式是先對二值化的印章影像作霍氏轉換 (Hough transform) 然後挑選出長度大於一定值的線段，再由各線段計算出各段的角度之後作群集處理，即能得到所要旋轉的角度。

由於印章圖像可能在切割後尺寸大小不一，為了要讓每張經由切割後的圖片有統一的尺寸大小，以便讓辨識處理上能有較高的效率。本專題對於印章的像素以寬高：100×100 正方形的大小作正規化。

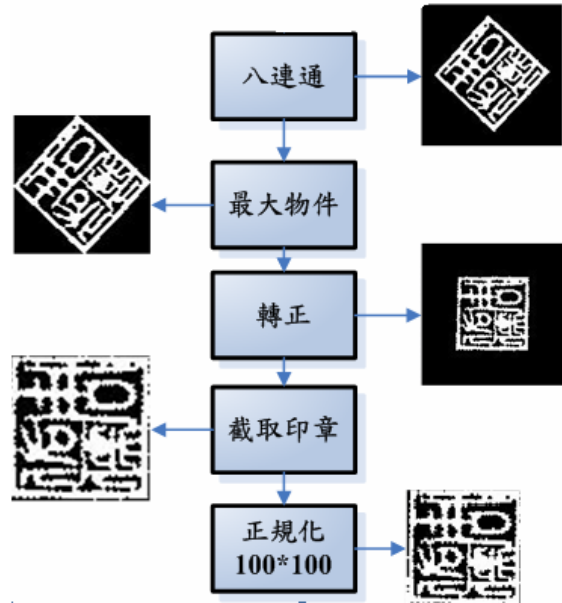
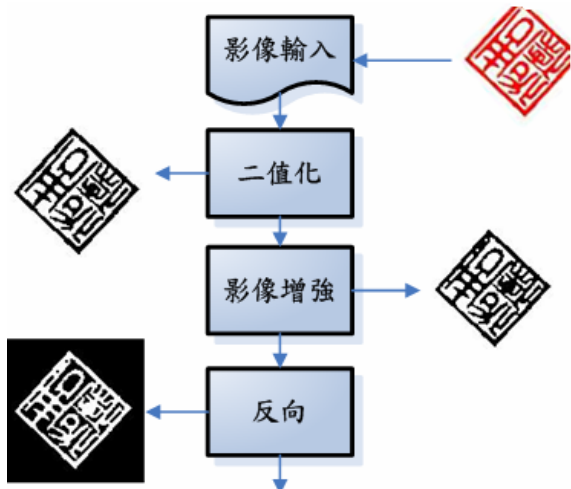


圖 2-1 前處理的流程圖

3. 使用 XOR 計算印章相似度

在掃描完提款單，經過定位切割正規化等前處理之後，因影像還是處於二值化的影像，所以計算相似度的方法，最快且直覺的方式便是 XOR，因為 XOR 就是不同的位元像素值才會被標記出來，下面討論 XOR 計算做模板比對的原理。

3.1 模板比對

模板比對的方式是採用邏輯運算上的 XOR (Exclusive-OR) 來完成，XOR 的計算規則是比對兩個像素值內容是否相同，相同則結果為邏輯上的 0 (false)，反之則為 1 (true)。在本專題的印章比對上，運算的方式是將前處理完之後的取款單印章影像矩陣 $A(m,n)$ 與資料庫中紀錄的存款簿印章影像 $B(m,n)$ 來做 XOR 運算 (存款簿印章也經事先處理成二值化影像矩陣)，兩矩陣實行 XOR 運算比對，運算完畢後，就會產生新的矩陣，新的矩陣內容即為相異的像素值，如果影像愈相似，則新矩陣的內容就愈多 0，故計算相異像素值的總數目 S ，其方程式如下：

$$S = \sum_{m=1}^{100} \sum_{n=1}^{100} XOR(A(m,n), B(m,n)) \quad (1)$$

3.2 相似度計算

因為影像正規化之後，成為 100×100 的矩陣，故執行 XOR 比對之後，相同像素值的總數目 S 至多為 10000，故經過下列公式：

$$p = \left(1 - \frac{S}{10000}\right) \times 100\% \quad (2)$$

計算後，即可取得相似度百分比 p ，當百分比 p 大於 90% 便將輸入之存款單之印章視為真，否則為

假, 整個使用 XOR 運算計算相似度流程圖如圖 3.1 所示。

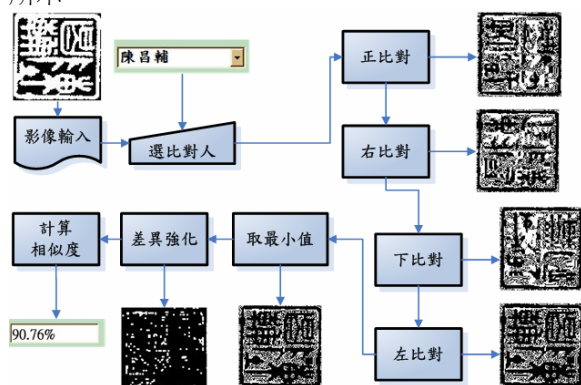


圖 3-1 使用 XOR 運算計算相似度流程圖

4. 使用傅立葉轉換特徵辨識印章

4.1 影像的傅立葉轉換

完成印章影像的初步處理後, 基於影像上原定義的表示方式較不易或無法解決辨識問題, 需將影像圖片做適當的轉換求得適合表示印章影像的表示方式, 進而解決問題, 對於轉換法的選擇, 需能夠讓各影像的特徵有顯著的區別性, 利用各種不同的轉換方式可將影像做不同種的定義, 並依據各轉換法特性找出對影像無關之性質進行特徵抽取, 以對後級辨識做區分性的比對來找出相似度高的印章資料, 本專題所用的一維離散傅立葉轉換(DFT)定義如下:

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{-\frac{2\pi kn}{N}} \quad (3)$$

二維離散傅立葉轉換定義如下:

$$X[r,s] = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x[m,n] e^{-\frac{2\pi mr}{M}} e^{-\frac{2\pi ns}{N}} \quad (4)$$

由於傅立葉轉換所得的頻譜為數位影像的基本, 也是重要特徵之一, 所以對於有關影像解析或影像辨識上, 它在許多地方是需要被使用到的。原影像之轉換所得值, 其表空間頻譜, 此為原影像中畫素的層階值做直接計算的特徵量, 例如, 濃度梯度、層階值平均值、標準偏差值、力矩等, 可反映影像之特徵。

利用傅立葉轉換法將影像以頻譜的方式呈現, 以進行影像粗糙、細緻、方向等紋理組織的特徵估測, 由於影像圖章與頻譜分佈有所關係, 較高頻的區域代表著影像上較明亮細緻之區域; 反之較低頻的區域代表著影像上較為陰暗粗糙的區域。另外由最主要的頻譜峰值的位置可看出紋理組織圖樣的主要方向。

4.2 特徵抽取

在圖樣辨識的問題中, 我們常需要在眾多可利用的特徵中選取適當特徵以供分類運用。特徵選取的問題已在許多文獻廣泛討論但並沒有明確的答案。印章圖像可以視為類似像生活中的符號、表示式、號誌、等等的圖樣, 而圖樣的辨識必須先找出圖樣之特徵(feature), 再根據這些特徵來區分出不同的圖樣類型。圖樣特徵可以用定量或定性的方式來描述, 甚至也可用向量或矩陣來表示, 一般而言, 圖樣是由若干影像描述子所形成, 例如:

$$\Phi = [\phi_1, \phi_2 \cdots \phi_L]^T \quad (5)$$

上述向量稱為特徵向量(feature vector)。對不同物體形狀的圖樣, 其特徵向量通常有顯著的不同。但對同一形狀物體放大、縮小的特徵向量須大致相近, 這就是分類的一個依據。本專題利用先前傅立葉轉換法後的影像圖片, 取出每一列的平均值(mean)與正規化標準差(normalized standard deviation)值, 並將擺入一個陣列當中來當特徵向量, 每一列的平均值可用下列公式計算:

$$\mu_r = \frac{1}{N} \sum_{s=0}^{N-1} X[r,s] \quad (6)$$

即計算影像矩陣內每列的總和再取其平均值, 每一列的正規化標準差值可用下列公式計算:

$$\sigma_r = \sqrt{\frac{1}{N} \sum_{s=0}^{N-1} (X[r,s] - \mu_r)^2} \quad (7)$$

最後, 平均值與正規化標準差值交錯合成一特徵向量如下:

$$\Phi = [\mu_0 \sigma_0 \mu_1 \sigma_1 \cdots \mu_{M-1} \sigma_{M-1}] \quad (8)$$

圖 4.1 是整個傅立葉轉換特徵抽取的示意圖。

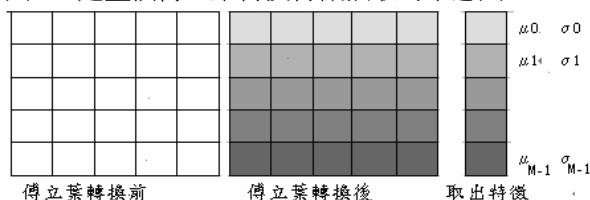


圖 4.1: 傅立葉轉換特徵抽取的示意圖

4.3 印章辨識

當抽取印章的特徵向量, 便可做辨識的動作。首先計算輸入提款單印章之特徵向量和資料庫中各個存簿印章之特徵向量間的誤差向量, 然後計算各個誤差向量的歐幾里得範數(Euclidean norm), 最後比較歐幾里得範數的大小, 擁有最小範數的存簿印章, 便是最終所須的辨識結果。至於如何計算

歐幾里得範數，現以圖 4.2 中之三維歐幾得空間 (Euclidean space) 中的向量為例來說明，其中 a, b, c 分別為沿著 x, y, z 軸的距離。而此向量的長度，也就是座標(0,0,0)到座標(a, b, c)之間的距離，便是此向量的歐幾里得範數。同樣地，對於 n 維的向量 $X=[x_1, x_2, \dots, x_n]$ ，其歐幾里得範數可計算如下：

$$\|X\|_e = \sqrt{\sum_{i=1}^n X_i^2} \quad (9)$$

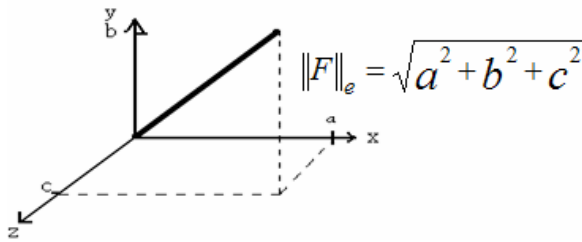


圖 4-2 三度歐幾里得空間中的範數計算說明

利用歐幾里得距離求出最小差距來做辨認的步驟，其方法簡單，也可有效地辨認出結果。圖 4.3 是整個傅立葉轉換特徵的辨識流程圖，本專題我們使用快速傅立葉轉換(FFT)法來計算特徵向量，節省所須的辨認時間。

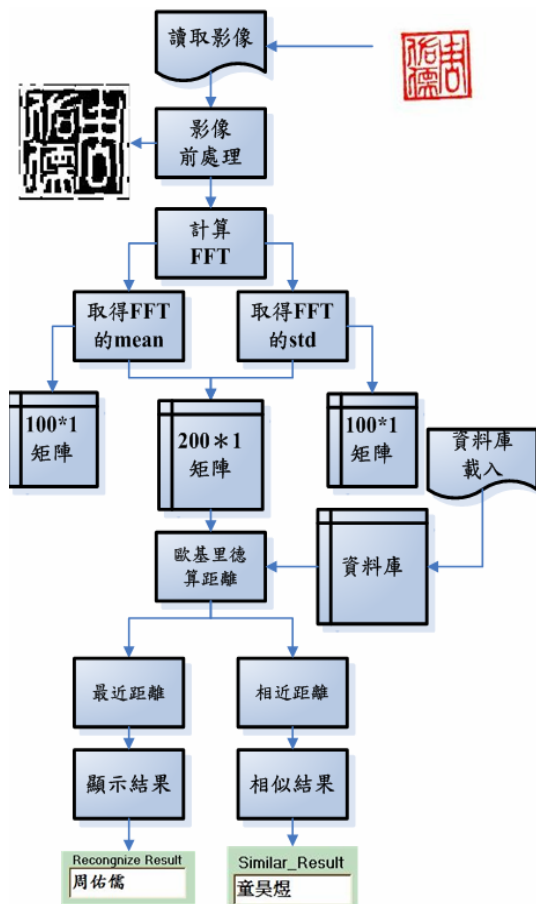


圖 4-3 傅立葉轉換特徵的辨識流程圖

5. 辨識結果

5.1 輸入測試影像

輸入影像為 22 人，每個人 70 張，共計 1540 張影像，再抽取每個人各五十張切割完整的印章，總共拿來判斷辨識率的共有 1100 張影像。

5.2 印章辨識結果

FFT 影像成功辨識的條件為：影像 FFT 取 mean 跟 std 算歐基里德最短距離，最接近資料庫中哪個存簿印章就判為那顆印章，下表沒有列入第二相近的結果。

XOR 影像成功辨識的條件為：算出相似度最高者，下表沒有列入第二相近的結果。

圖片	印章	資料庫	FFT	XOR
img01	陳明瑤	data1.dat	48	50
img02	李俊昇	data2.dat	21	49
img03	盧志遠	data3.dat	38	50
img04	邱彥勳	data4.dat	35	46
img05	曾建誠	data5.dat	32	49
img06	吳文揚	data6.dat	34	49
img07	吳家銘	data7.dat	27	48
img08	許伯誠	data8.dat	42	47
img09	田克群	data9.dat	33	49
img10	王斯弘	data10.dat	50	50
img11	吳博文	data11.dat	32	48
img12	蔡煜基	data12.dat	38	34
img13	小叮噹	data13.dat	32	50
img14	王蒲韋	data14.dat	32	50
img15	吳仕仁	data15.dat	27	49
img16	林世民	data16.dat	37	49
img17	孫卓成	data17.dat	33	50
img18	張建鋒	data18.dat	35	50
img19	童春木	data19.dat	24	50
img20	周佑儒	data20.dat	50	50
img21	陳雅琴	data21.dat	50	50
img22	童昊煜	data22.dat	38	50
總和	788 / 1006	總辨識率	71.6%	97%

表 5-1 印章辨識結果統計表

6. 程式實作介面

6.1 使用者介面

在研究實體的使用介面 (User Interface) 部分是以 Borland C++ Builder 6.0 (Build 10.157) 編寫開發而成，核心部分則是用 Matlab 7.0.4 SP2 (R12) 內 M 檔編輯器 (M-file Editor) 編寫完成。另外關於掃描器的控制，是採用 EZTwain 的 lib。

在使用者介面 (User Interface) 系統上兼具以下特色：

1. 人性化互動式對談介面，完全只需要操作電腦滑鼠即可。

2. 藉由介面互動可瞭解每個細部的過程實作結果 例：(定位、二值化、截圖、轉正)。

3. 掃瞄時不用去點掃瞄器上的快捷鍵，且掃瞄完的影像會即時顯示在使用者介面。

4. 擁有 One Touch(Auto)功能，方便操作，直接按下此按鈕，即可完成掃瞄和辨識。

圖 6.1 至圖 6.6 是各個銀行提款單的印章真偽辨識範例。



圖 6-1 按下 Auto 前範例

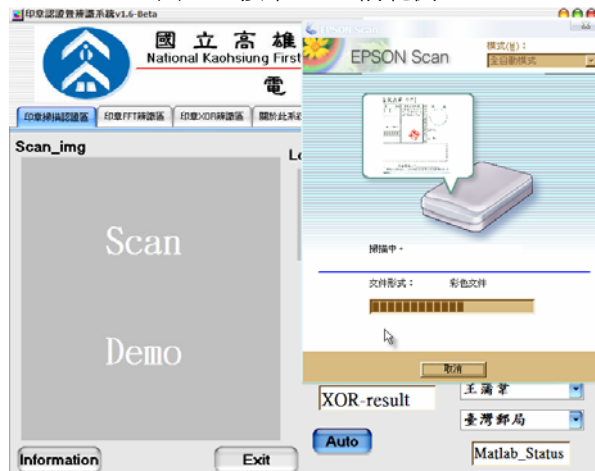


圖 6-2 掃描提款單範例

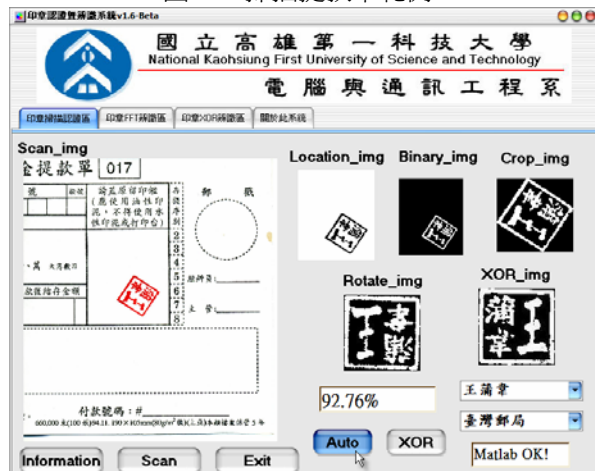


圖 6-3 臺灣郵局提款單認證範例

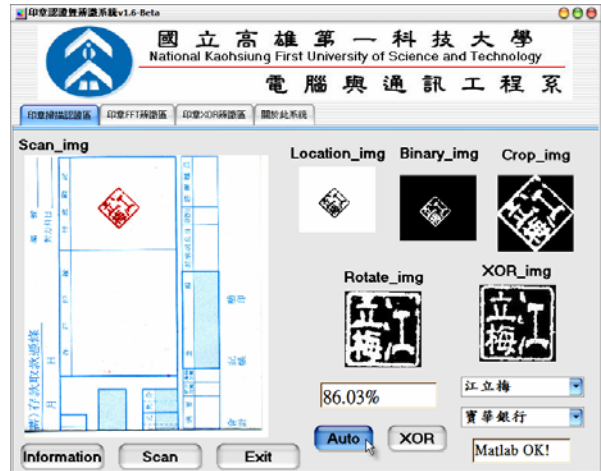


圖 6-4 寶華銀行提款單認證範例

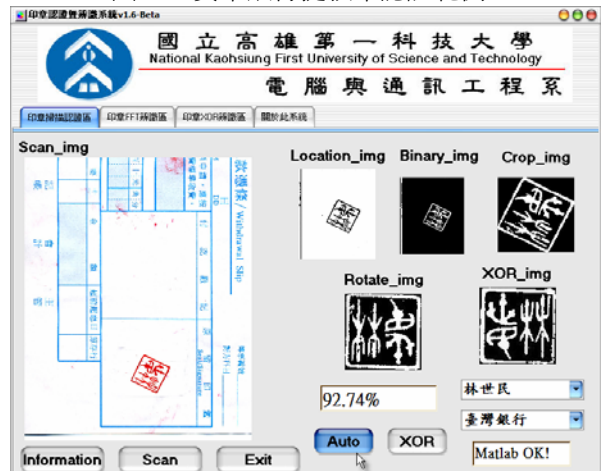


圖 6-5 臺灣銀行提款單認證範例



圖 6-6 印章 XOR 模板比對辨識範例

7. 結論與未來研究

現今在銀行提款的方式主要有兩種：一種由提款卡領取，另一種是由存簿領取。使用提款卡領取現金，是現階段人們最常使用的方法，主要的原因為它是最方便、最快速，可是相對的危險性就比較高，因為只要擁有提款卡跟密碼，任何人都可以領取現金，完全無法判別是否為當事人本身。可是在存簿提款方面，除了需要存款簿跟密碼之外，還必

須要有開戶印章(註:開戶印章就是指當初開戶時所使用的印章)。因此只要在印章方面做個辨識,就可以降低危險性。

目前銀行在印章辨識方面,還是採取人工的方式,用肉眼來辨認。此方法在效率跟準確度上,有一定的上限,畢竟有人為疏失的時候。因此,我們選擇電腦來幫忙做辨識,除了可以提高效率之外,在準確度上也會有大大的提升。

我們此研究的目的,就是在於提高印章辨識的效率,只需將印章掃描到電腦裡,再輸入此印章上的人名,就可以很快速的辨認出相似度有多高,如果相似度高達九成以上,就可表示此印章跟此人名是相同的。可是在使用此系統的時候,還是會有一些問題存在,再此提出一些探討跟說明,以作為未來相關研究時可做為參考。

影像旋轉:

印章影像做完擷取後,由於建立印章影像時並未蓋正,需要做轉正的動作,而轉正的動作是為了能讓傅立葉轉換時能夠以固定的四個方向來做比對與辨識,因為傅立葉轉換不抗旋轉的特性,所以印章旋轉角度大於 20 度時,會使得抽取特徵值時產生錯誤,故未來能以較佳的演算法來提升影像轉正的成功率以提升整體辨識率。

特徵抽取:

因為經轉換過後的矩陣,是以頻譜能量的成份儲存,特徵的抽取是根據此矩陣觀察到的的差異性來做分類,但在解析多個人的平均求得的特徵值,在做分類時重疊的類別太多,無法有效區分出各個類別,所以辨識的結果會變得很差,針對此問題特徵的抽取需以更具區別性的性質來當做特徵,例如說字型特性或是分部切割的方式等,都是未來值得去探討的方向。

8. 參考文獻

- [1] 楊武智, 影像處理與辨認, 全華科技 (1994)
- [2] 葉秋城、王文豪, 影像處理與分析, 碁峰資訊 (1995)
- [3] 林仲芬, 影像辨認技術, 全華科技 (1995)
- [4] 繆紹綱, 數位影像處理-(活用Matlab), 全華科技 (2004)
- [5] 鍾國亮, 影像處理與電腦視覺, 東華書局, (2004)
- [6] 張智星, Matlab程式設計與應用, 清蔚科技 (2000)
- [7] 李顯宏, Matlab 7.x介面開發與編譯技巧, 文魁資訊 (2005)
- [8] R.C. Gonzalez and R.E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall (2002)
- [9] R.C. Gonzalez, R.E. Woods and S.L. Eddins, Digital Image Processing Using Matlab, Prentice Hall (2004)
- [10] W.K. Pratt, Digital Image Processing, 3rd Edition, John Wiley & Sons (2001)
- [11] S.K. Mitra, Digital Signal Processing: A

Computer-Based Approach, 3rd Edition, McGraw Hill (2006)

誌謝

在這邊特別感謝學長同學老師們提供的印章,讓專題資料庫更加的充實,並感謝所有在專題上幫助過我們的人。

指導老師(姓名)

參賽組員(姓名)¹, 參賽組員(姓名)², 參賽組員(姓名)³, 參賽組員(姓名)⁴

材料成本

項次	項目	規格	數量	單價(元)	金額(元)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10	其他零配件及耗材				
	合計				

設備折舊 (使用設備折舊金額以購買價格的 1/5 計算)

項次	項目	規格	數量	單價(元)	金額(元)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
	合計				

人力成本

項次	工作項目	說明	工時	單價(元)	金額(元)
1					
2					
3					
4					
5					
6					
7					
8					
9					
10	雜項工作				
	合計				

組員貢獻度 (全部組員貢獻度合計 100%)

學號	姓名	主要工作項目	貢獻度(%)