

第十三屆旺宏科學獎

成果報告書

參賽編號：SA13-236

作品名稱：以手寫模式推演個人化字型之研究

A User-Feature Based Approach to Dynamically
Building Personal Calligraphy Fonts

姓名：程肯

關鍵字：模式識別、動態組字、末級部件

摘要

電腦字體近來流行許多近似手寫字體的字形，像是少女體、硬筆書法體等等。這些字形都是由書法家一字字寫入電腦再編輯完成。這樣的製作方法不僅讓個人用戶無法製作自己的個人化字形，也讓專業的類手寫字體制作極為費時。結合型態學演算以及模式識別、中文字資料庫等等，我製作了利用使用者少數部件的重組，模擬該使用者書寫其他中文字的系統。此系統不僅可以用在製作個人化字形，亦可以用在文化資產的保存與研究，凡如佚漏古籍之校定及重製。

我的系統從使用者的部件擷取到字庫分析及重新組字，皆經過調校已符合中文字的特性，可利用三百個中文字部件組合出超過 3500 個常用中文字，涵蓋超過 95% 的中文字使用率。

Abstract

Though user-feature based font is easy to obtain in phonetic symbol languages like English, creating a similar font in Chinese is not an easy task. With over thirty thousand characters, user-feature based font is limited to only professional font designing companies with sky-high cost. My system is a whole new approach to creating personal font. While only requesting a little amount of user features, we can analyze and capture the information and components needed to simulate other characters not written by the user. My system not only provides a way for individual users to create personal font, it is also a tool to preserve Chinese calligraphy culture. When certain characters by a calligrapher are missing, my system provides the ability to reproduce these characters. After carefully manipulating algorithms and parameters, only 350+ characters are needed to produce more than 3500+ Chinese characters, covering more than 95% of everyday usage.

目錄

壹、研究動機	1
貳、研究目的	1
參、研究設備及器材	1
肆、研究過程與方法	2
一、名詞解釋	2
二、筆畫組字	3
三、部件組字	5
四、書法組字	16
伍、研究結果	16
一、系統使用流程	16
二、系統介面整合與呈現	17
三、動態組字結果之範例	19
四、組字範例分析	20
五、書法組字結果之範例	21
陸、討論與結論	22
柒、參考資料及其他	23
捌、附錄	I
一、末級部件表	I
二、漢字字表	II
三、漢字部件拆分表(節錄第 1~100 個字)	V
四、使用貪婪演算法製作之使用者輸入字表	VII
五、研究使用字體例圖	VII
六、程式畫面	VIII

壹、研究動機

一般在日常生活當中，我們在各個地方時常要簽署各式各樣的文件，大至動輒數十萬的交易契約書，小至父母幫小孩簽名的註冊通知單，既然文件內容可以電子化，那簽名為何不能也電子化呢？所以就產生了一個研究的想法，可以利用資訊課所學的 C++ 程式語言，根據一個人的寫字樣本，利用影像辨識技術製作出與樣本提供者手寫字相同的電子字型，再將其以向量圖形的格式輸出，以增加安全性，防止被他人盜用複製的可能性。這個想法可以應用在更多方面，像是在 Blog 上使用個人化字型發表文章、使用個人字型和朋友在通訊軟體上聊天等。使用這種個人化字型可以使身處在資訊化世代的我們有更方便的認證方式，還可以讓過去只是呆版的電腦文字更有個人風格。故本研究希望能製作出一套製作個人化字型的流程。

貳、研究目的

利用影像辨識技術，以單一使用者提供的手寫字樣本，進行字體構造的拆解與重組，製作出具個人書寫風格的手寫模式推演字型。探討使用何種方法製作個人化字型能有最佳的效果。

- 一、利用手寫板的書寫功能記錄使用者的寫字特徵
- 二、以最少使用者輸入字數涵蓋組字需要的所有部件
- 三、將中文字庫進行拆分以得到部件配置資料
- 四、以使用者手寫部件重新組字
- 五、組字系統及程式之介面整合

參、研究設備及器材

一、筆記型電腦

(一) 硬體

1. CPU：Intel(R) Core(TM) i5-3210 CPU @ 2.50GHz
2. RAM：8.00GB

(二) 軟體

1. Microsoft Visual Studio 2012
2. OpenCV 2.4.5
3. Qt 5.1.1

4. potrace-1.11

二、Genius F350 手寫板、Wacom CTL-480 繪圖板、AIPTEK Ultimate XL

三、計算紙

四、筆

肆、研究過程與方法

一、名詞解釋

(一) 末級部件

末級部件是構成漢字最基本的部件，通常是由部首為單位。例：「草」可拆分成「艹」和「早」，而「早」又可拆成「日」與「十」，「艹」、「日」與「十」就是末級部件，但「日」不可拆分成「一」與「口」，因為「日」就是一個基本文字單位，再進行拆分會變成只有形，沒有意的筆畫。將這些末級部件彙編成末級部件表(參見附錄)。

(二) 漢字字表

由葉健欣教授所組織的剎那工坊，提供了中文字庫及其拆分部件，但不包含末級部件，因為末級部件無法再進行拆分。字庫內容包含 69650 個字，並按照 Unicode 排序，但此字庫內容包含了過多的罕用字，故將此字表和教育部公佈之常用字表整合，挑選出共 3749 個常用字作為漢字字表(參見附錄)。

(三) OpenCV

OpenCV 的全稱是 Open Source Computer Vision Library，是一個跨平台的電腦視覺庫，以 BSD 授權條款授權發行，可以在商業和研究領域中免費使用。OpenCV 可用於開發即時的影像處理、電腦視覺以及模式識別程式。本研究使用的為 OpenCV 2.4.5 版本。

(四) Qt

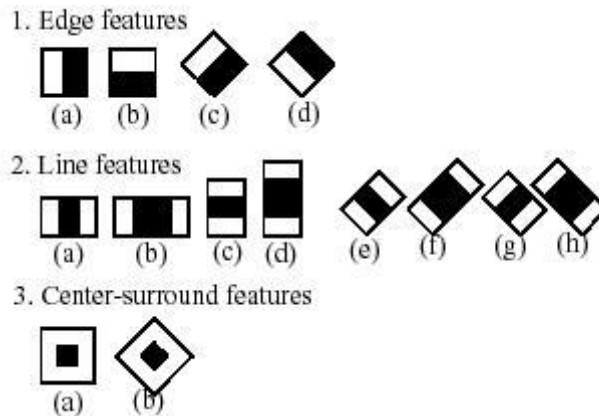
Qt 是一個跨平台的 C++ 應用程式開發框架，是自由且開放原始碼的軟體，在 GNU 較寬鬆公共許可證條款下發布。廣泛用於開發 GUI 程式，Qt 使用標準的 C++ 和特殊的代碼生成擴充功能以及一些巨集。本研究使用的為 Qt 5.1.1 版本。

(五) 模板匹配

模板匹配為在一影像上尋找一已知區塊是否存在的方法，多半使用類相關係數進行實作。

(六) 梯級式級聯分類器

Haar 特徵檢測是 OpenCV 開發中的程式，其方法大致為，以機器學習的方法找出正負樣本中 Haar 特徵的規律，用以辨識未知的圖片。OpenCV 的 HaarTraining 算法以 Friedman, J. H 等人的「Additive Logistic Regression: a Statistical View of Boosting」為基礎，實現其中 2 類分類問題的四種 Boost 算法：Discrete AdaBoost、Real AdaBoost、LogitBoost 和 Gentle AdaBoost。Friedman, J. H 等人只描述了如何訓練一個強分類器，對於訓練級聯的強分類器(Cascade of Classifiers)，OpenCV 採用的是 Paul Viola 等人的「Robust Real-Time Face Detection」中所述方法。常見的 Haar 特徵(Haar-like feature)即如下圖所示：

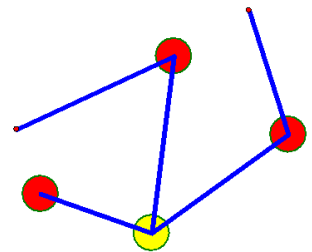


圖(一) Haar-like Feature

利用 Haar 特徵檢測，以正負樣本訓練，生成的分類器，可即時地偵測影像裡的目標。

(七) 覆蓋

在圖論中，當選定一個點時，與其相連的點及邊即為被覆蓋的點(邊)。此研究中主要使用點的覆蓋。



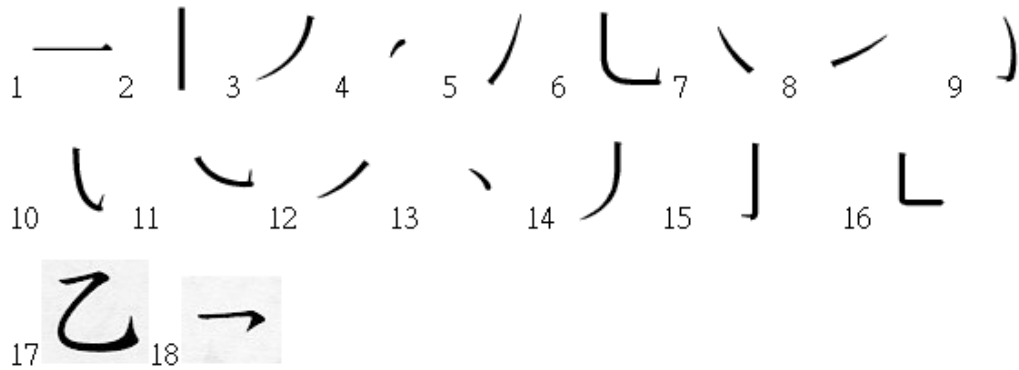
圖(二) 覆蓋示意圖

二、筆畫組字

最初使用「筆畫」為單位組字，其方法為：首先，以日本網站 kage 歸納出的漢字基本組成筆畫，挑選出手寫漢字的基本筆畫。接著建置一個記錄筆畫位置的資料庫，其內容包含筆畫數、筆畫種類、筆畫位置、筆畫之長寬及任兩筆畫之交叉點位置。再利用 OpenCV 編寫出能夠即時讀取使用者的寫字樣本之程式，並以筆畫為單位儲存成圖片，以達到使用者的模式學習。最後根據上述步驟所獲取之資訊，將所有使用者提供的基本筆畫進行方向與大小上的調整置入圖中，即可獲得此方法建立之模擬字型。

(一) 挑選基本筆畫

筆畫是手寫漢字時最基本的單位，因此可以按照順序分別擷取使用者手寫漢字的筆畫，搭配之後製作的筆劃位置資料庫，經過線性變換處理來達到研究目的。以下是參考日本網站 kage 歸納出的漢字基本組成筆畫，並進行編號，方便接下來的研究：



圖(三) 漢字基本筆畫表

(二) 建置筆畫位置資料庫

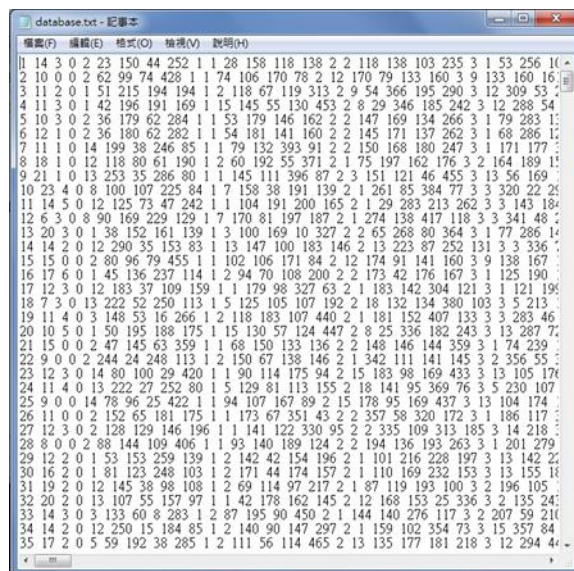
為了達到準確的筆畫組合，我建置一個記錄筆畫之各類相關資訊的資料庫，採用半人工的方式進行記錄(根據設計能用點選的方式告訴程式每一筆畫的起始點、結束點，程式會自動計算並記錄其他資訊)。資料庫內容包括了筆畫數、筆畫種類、筆畫位置、筆畫之長寬及任兩筆畫之交叉點位置。

```
1 struct Cchar{
2     int stroke_cnt,sn,cross_cnt;
3     stroke s_data[40];
4     cross c_data[20];
5 };
6
7 struct point{
8     int sn,x,y;
9 };
10
11 struct stroke{
12     int sn,type;
13     point start,end;
14     double m,k;
15 };
16
17 struct cross{
18     int sn,stroke_sn1,stroke_sn2;
19     point s;
20     float s1_p,s2_p;
21 };
```

圖(四) 筆畫資料庫結構



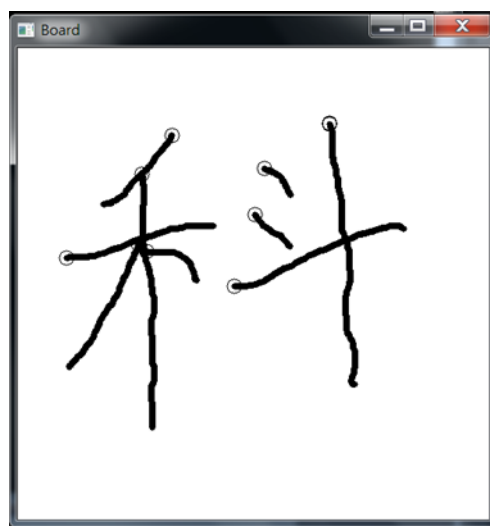
圖(五) 建置資料庫畫面



圖(六) 資料庫內容

(三) 使用者寫字模式學習

為了蒐集使用者寫字模式並學習，使用 OpenCV 編寫出能夠即時讀取使用者提供的寫字樣本之程式，由於目前是以筆畫為單位進行組字，所以將其讀取到的寫字樣本以筆畫為單位儲存成圖片格式，方便程式進行組字利用之。



圖(七) 使用者模式學習

(四) 進行組字

運用上述資料庫中的筆畫各項相關資訊，能從使用者擷取到的寫字樣本集取出需要的筆畫種類，並按照筆畫位置資訊將其放置在正確的位置，再將其按照筆畫的長寬資訊，透過 OpenCV 中的 cvResize 函數進行有效的縮放，並以長寬比例判斷筆畫的斜率，進行適度的旋轉，將所有筆畫依此方式置入圖中，再經由全域特徵(傾斜度、字元間距)調整後，即可獲得模擬的字型。

三、部件組字

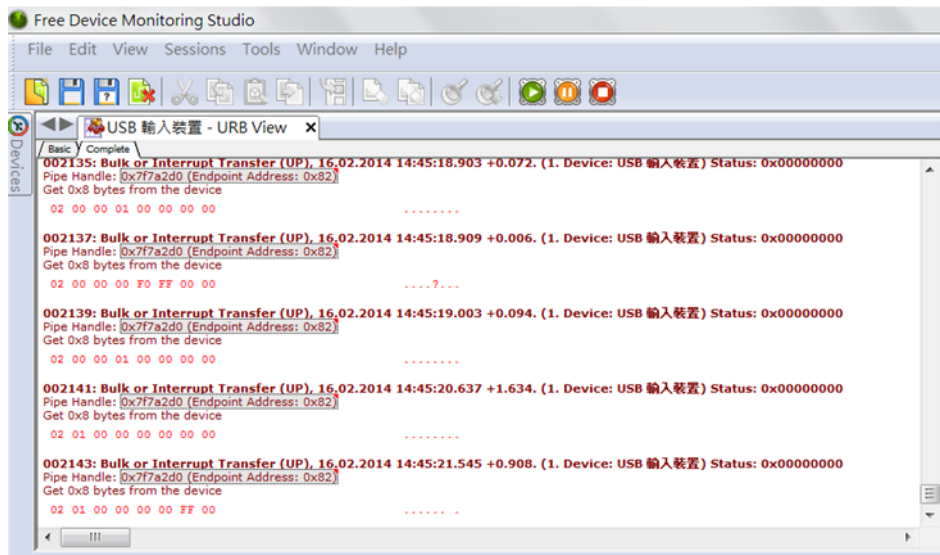
(一) 利用手寫板的書寫功能記錄使用者的寫字特徵

1. RawInput 手寫資料讀取

利用手寫板即時讀取，且使用具有深度感應的筆順擷取，並將其轉換為圖片，進行末級部件的拆分，取得所需的末級部件。讀取資料格式為 OpenCV 內建之圖像

結構 IplImage。惟因 OpenCV 內建之滑鼠呼叫程式並不支援壓力資訊讀取，我使用封包監聽的方法，讀取 INTERRUPT_TRANSFER 的資訊。經過比對與分析之後，得到手寫板資訊傳輸的格式，為 8 byte 的封包，其中包含了按鍵狀態、座標、壓力數據等，以 16 進位呈現。接著要讓讀取手寫資訊的程式也能接收壓力資訊，使用 <Windows.h> 中的 RawInput API 讀取手寫板的資料流，解讀出壓力數據改變筆畫粗細。

在 RawInput 定義的 RawInput Object 中，必須使用 UsagePage=0x0d, Page=0x02，鎖定非滑鼠、鍵盤的 HID (Human Interface Device) 裝置，再檢查 Product ID 及 Vendor ID，確認為手寫板的資料流。以預先宣告的 Buffer 儲存資料，再進行進制轉換，以 extern 變數的方式傳送到 OpenCV 的繪圖函數。以此方法，可以巧妙避開擷取 USB Pipe 必須使用 Windows DDK (Driver Development Kit) 開發，以及 RawInput API 必須要在 WinMain 介面上執行的限制。



圖(八) 手寫板封包監聽畫面

數據說明：10 XX_{觸控狀態} XX_{X座標} XX_{X座標} XX_{Y座標} XX_{Y座標} XX_{壓力座標} XX_{壓力座標}

```

1 struct pendata{
2     BYTE head;
3     BYTE button_flag;
4     BYTE x1;
5     BYTE x2;
6     BYTE y1;
7     BYTE y2;
8     BYTE p1;
9     BYTE p2;
10 };

```

圖(九) Raw Input 轉型後之結構

2. Wintab API 手寫資料讀取

為了更精確的讀取使用者手寫資訊，嘗試使用 Wacom 的繪圖板讀取手寫資訊，但是原先使用的方法並不支援 Wacom 之繪圖板，因為 URB Packet 中的 INTERRUPT_TRANSFER 並不包含壓力及位置資訊，且 Windows 內建的 HID Driver 亦不能支援 Wacom，使 RawInput API 失效。

Wacom 使用的驅動程式是其自行開發的 Wintab，也有提供 C++ 的 API。將原先使用 RawInput 的程式改寫，引入 Wintab API。使用 Wintab 可以做到滑順的寫入且更貼近紙筆的觸感。

3. AIPTEK 手寫資料讀取

為了進一步模擬紙筆的書寫習慣，採購了 AIPTEK 的 Ultimate XL 手寫板。此手寫板同 Wacom 手寫板，不需安裝電池於筆桿中即可運作，使筆桿之配重更貼近真實。此手寫板之面積也更大，方便使用者書寫，也具有筆觸傾斜度感測功能，在未來的研究中或許能納入使用者模式學習計算。

此手寫板不需使用 Wintab API，RawInput 即可讀取，惟因功能更加複雜，數據流編碼也較複雜，但使用第一點的方法，修改編碼、Product ID、Vendor ID 後，同樣能讀取使用者的手寫資訊。

(二) 以最少使用者輸入字數涵蓋組字需要的所有部件

1. 挑選適合的末級部件

根據教授提供的漢字部件拆分表（見附錄），得到每一個中文字以「最少」的部件數拆分方法，見範例：

於=木於

梱=木固

梱=木困

棟=木東

棠=尚木

這並不是需要的格式。在系統進行動態組字時，需要的資料是組成該中文字的所有「末級部件」清單，並不是最少部件清單，因為這些部件可能可以進一步拆分。我使用遞迴的方法，以 DFS (Depth First Search) 找到所有漢字的末級部件拆分方法，一共有 1195 個末級部件。

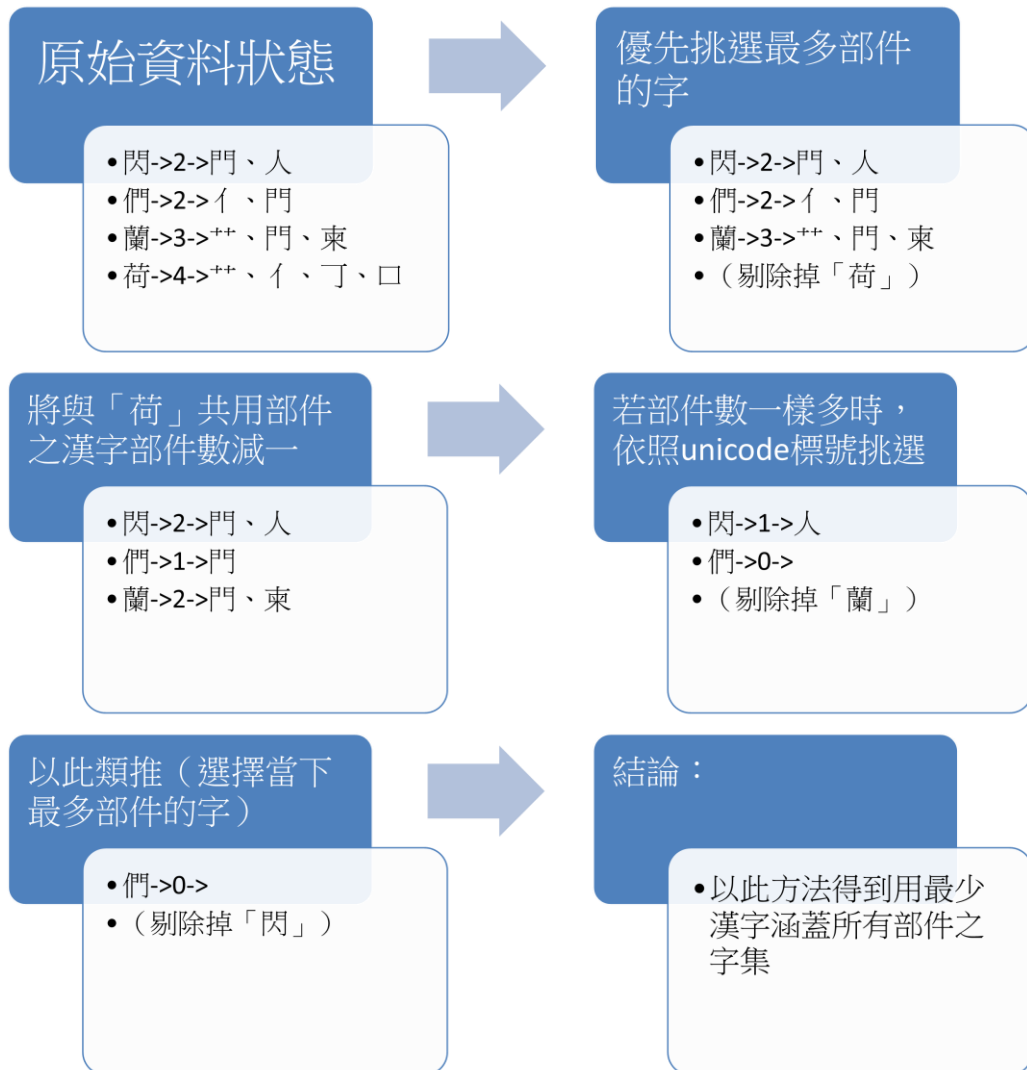
但事實上，有許多部件都是特殊的異形字才使用到，甚至有些特殊部件只在一兩個漢字中出現。因此透過程式篩選的方式，比照教育部公布的常用字表，去除掉甚少使用的末級部件。運用約五百個部件就可以組成大部份的漢字(參見附錄)。

2. 使用挑選出的末級部件製作出需使用者提供的漢字字表

得到組成中文字必須的部件清單後，系統將需要使用者書寫這些部件。以使用者的部件替換在中文字中的相對位置，就可以輸出個人化特徵的字形。

雖然從一千多部件減少至五百多個部件是很大幅度的縮減，但是讓使用者寫五百多字還有可減少的空間。而且寫單純的部件必會與寫出完整的漢字有所誤差，因此我想出一個方法，以最少的漢字來涵蓋這些末級部件，找出所有需使用者寫出的漢字。要求使用者以手寫板寫入這些漢字，就可以獲得組字必須的所有部件。

貪婪演算法：



圖(十) 使用貪婪演算法之流程圖

透過此方法所得到的使用者輸入字表詳見附錄。

貪婪演算法的優化：

有些部件只有在一個漢字中被使用，所以包含所有部件的漢字字集必須包含這些字。預先選取出這些字，再進行貪婪演算法處理，可以減少運算時間，也可優化選字結果，縮小必須讓使用者寫入的字數。

雖然這個方法確實減少了使用者需寫的字數，但過程中卻有一個盲點，那就是在第四步驟。當部件數相同時，並無法確定選取哪一個字較佳，而只能按照 unicode 挑選。這也使得使用者輸入字表雖然有優化效果，但卻不是最佳解，這也是未來希望改善的地方。

(三) 將中文字庫進行拆分以得到部件配置資料

利用矩陣線性變換、型態學變換等影像處理技術，搭配 Cascade Classifier 和模版匹配將漢字常用字分解成多個末級部件的組合，並將其組合以及當中各個末級部件的位置記錄到資料庫，以便後續製作個人化字型。

```
1 struct c_character{
2     int sn;
3     int component_cnt;
4     set<int> components;
5     IplImage* graph;
6 };
7
8 struct component{
9     int sn;
10    IplImage* graph;
11 };
```

圖(十一) 漢字與部件之資料結構

將漢字拆分的目的，在將使用者的部件帶入時，需要知道將哪個部件放置哪個位置的資訊。而將漢字的影像解析成各個部件的合成，牽涉到影像辨識和許多型態處理。我嘗試了以下方法：

1. 型態學切分

觀察發現，大部分的漢字皆由數個連通區間構成，經常一個連通區間就是一個部件。使用這個方法可以輕鬆得到許多易於拆分的漢字之部件位置，但是此方法很難解決兩部件接觸成為一個連通區間的狀況。舉例來說：

共有三個連通區間，分別是⁺、日+大、巾。若使用型態學切分法，就只能得到三個部件，要將日+大分開，需要進一步的處理。我嘗試使用侵蝕、擴張、細化和許多選擇性的像素腐蝕(如下圖)，斷開連通區間中可能為兩部件交接觸的部分。實驗後發



現，因漢字部件間可能的組合方式太多，使用固定的區間切割方式，必會影響到其他部件內的筆畫連接。之後嘗試使用其他的部件辨識方法。



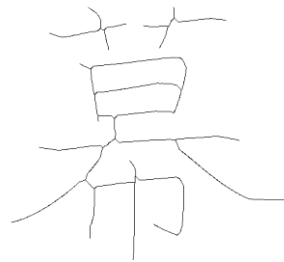
圖(十二) source



圖(十三) skel



圖(十四) skel+thin



圖(十五) thin



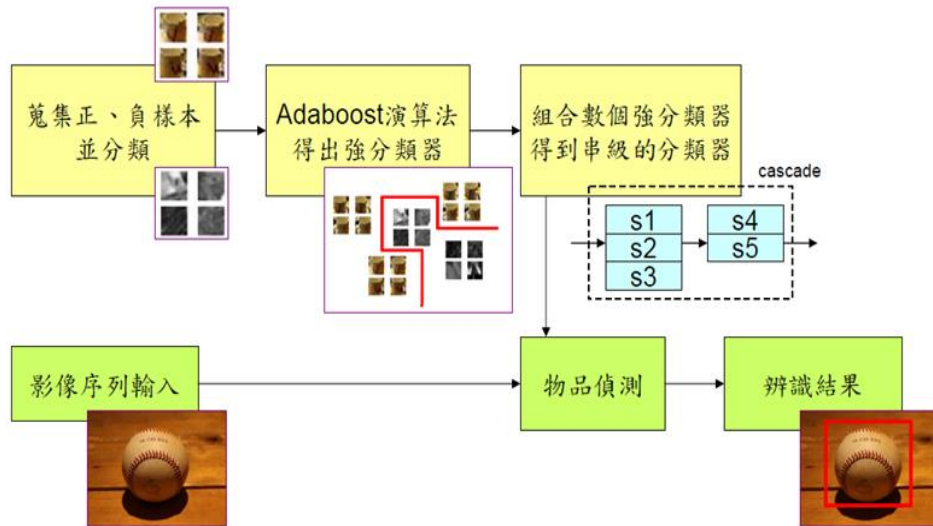
圖(十六) Expand

2. Cascade Classifier 辨識

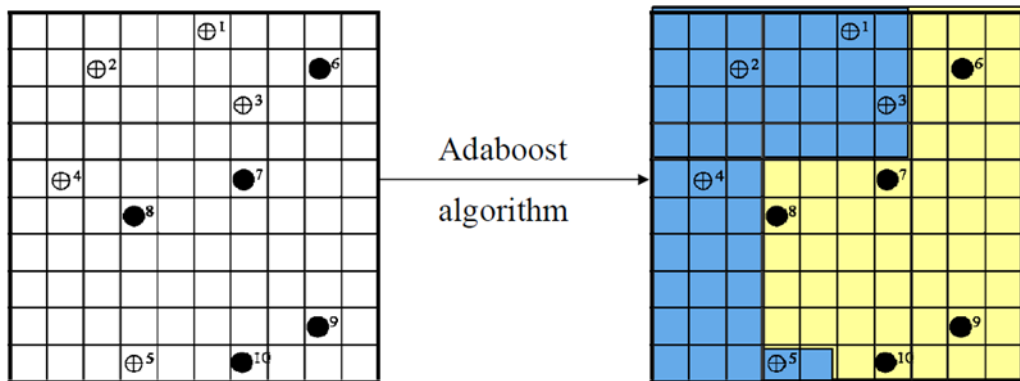
此方法不將漢字中的連通區間拆分，直接在全字影像中辨識部件。辨識部件的方法，我嘗試使用業界進行人臉辨識速度最快、穩定性最高的 Cascade Classifier 搭配 Ada Boost 進行部件辨識。

此方法基於分類器的訓練以及疊加。分類器為數百個弱分類器的合成，單個弱分類器由 Haar Feature 進行判斷，為容錯率高但準確度低的分類。經過大量樣本的訓練，將測試標的輸入後，經過一層層的弱分類器辨識，得到結果。

實測發現，Cascade Classifier 需要數千個正樣本及負樣本，而漢字的樣本數不夠進行分類器的訓練。Cascade Classifier 較適合用於有各種不同的正樣本模式、大小、光源等等。漢字的樣本中部件出現的形式較為單一，正樣本的變化性不足，反而不適合使用。



圖(十七) Cascade Classifier 流程



圖(十八) Adaboost 分類器優化

3. Template Matching

模版匹配 (Template Matching) 的方法較適用於辨識標的和尋找目標間關係變化性不大的狀況，正是漢字的特徵。一個部件在漢字中，只會具有一定限度的大小、長寬比變化，將模版進行型態學轉換後，及可以容許部件在漢字內的變化性。

常用的模版匹配有五種方法：SQDIFF、SQDIFF NORMED、TM CCORR、TM CCORR NORMED、TM COEFF、TM COEFF NORMED。

a. `method=CV_TM_SQDIFF`

$$R(x, y) = \sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2$$

b. `method=CV_TM_SQDIFF_NORMED`

$$R(x, y) = \frac{\sum_{x', y'} (T(x', y') - I(x + x', y + y'))^2}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}}$$

c. `method=CV_TM_CCORR`

$$R(x, y) = \sum_{x', y'} (T(x', y') \cdot I(x + x', y + y'))$$

d. `method=CV_TM_CCORR_NORMED`

$$R(x, y) = \frac{\sum_{x', y'} (T(x', y') \cdot I(x + x', y + y'))}{\sqrt{\sum_{x', y'} T(x', y')^2 \cdot \sum_{x', y'} I(x + x', y + y')^2}}$$

e. `method=CV_TM_CCOEFF`

$$R(x, y) = \sum_{x', y'} (T'(x', y') \cdot I(x + x', y + y'))$$

where

$$T'(x', y') = T(x', y') - 1/(w \cdot h) \cdot \sum_{x'', y''} T(x'', y'')$$

$$I'(x + x', y + y') = I(x + x', y + y') - 1/(w \cdot h) \cdot \sum_{x'', y''} I(x + x'', y + y'')$$

f. `method=CV_TM_CCOEFF_NORMED`

$$R(x, y) = \frac{\sum_{x', y'} (T'(x', y') \cdot I'(x + x', y + y'))}{\sqrt{\sum_{x', y'} T'(x', y')^2 \cdot \sum_{x', y'} I'(x + x', y + y')^2}}$$

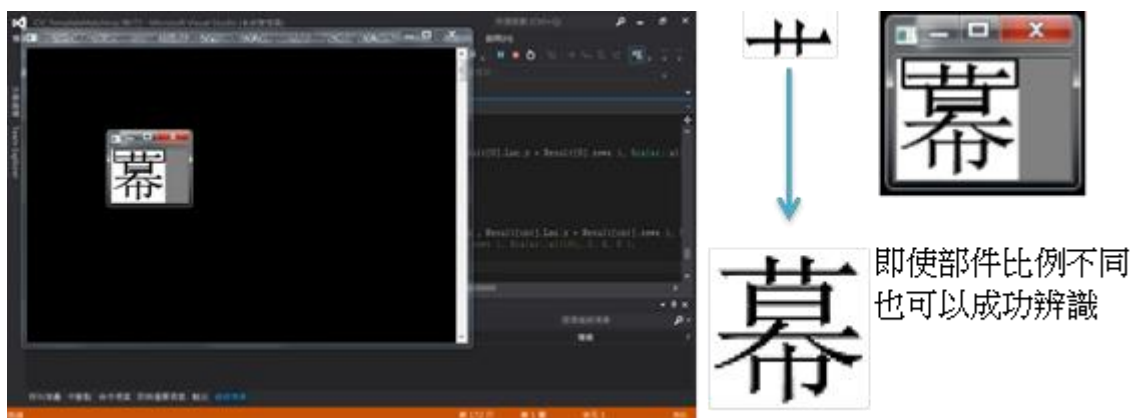


圖(十九) 模板匹配示意圖



圖(二十) 模板匹配結果示意圖

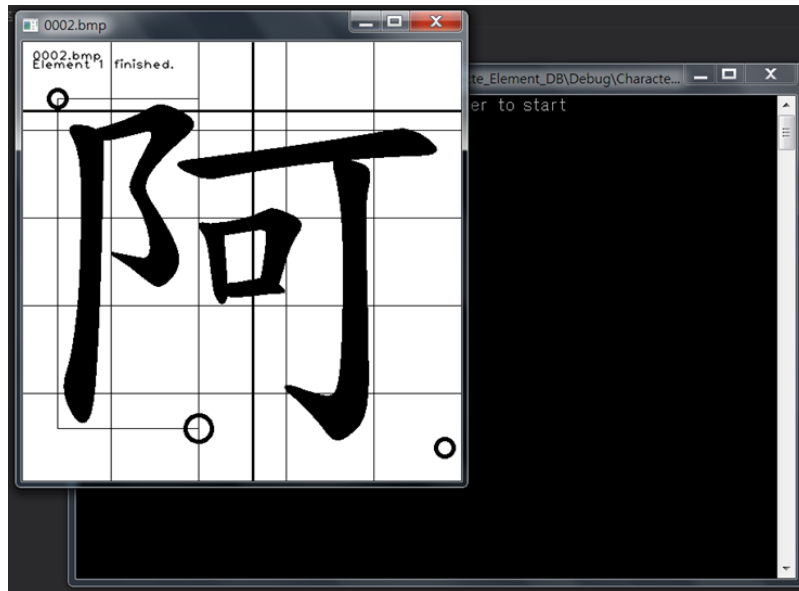
辨識系統會調整模版（部件）的大小及常寬比，在漢字圖像中比對，找到相關係數最高的位置。統計部件在漢字中最大程度的變形，即可得到辨識成功率最高的參數設置。經過測試，這是最穩定將漢字拆分出部件位置之合成的方法。範例如下：



圖(二十一) 部件辨識範例

4.備用部件切分法

在某些特殊中文字，template matching 無法準確辨識部件位置，就必須使用手動切分的方式。同樣使用 OpenCV 編寫自動記錄部件位置的工具。調校部件匹配的參數後，九成以上的漢字部件都可以準確辨識，但是諸如：丿、等部件，亦同時被許多末級部件包含，無法使用部件匹配的方式，即需要手動切分。

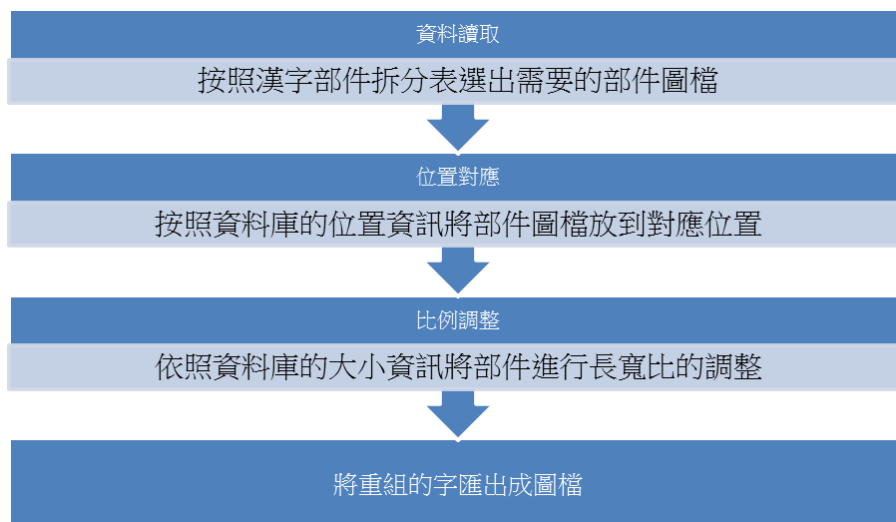


圖(二十二) 手動部件切分

(四) 以使用者手寫部件重新組字

1. 使用部件重新組字

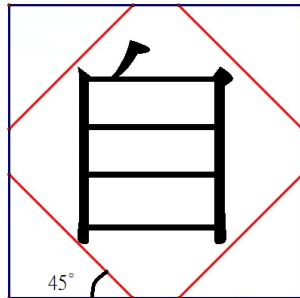
依照部件位置資料庫所記錄的位置、大小等資訊，將使用者寫出的末級部件加以組合。



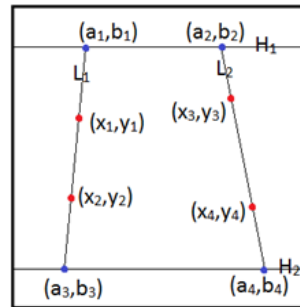
2. 全域特徵調整

經過動態組字重新以使用者的部件取代原始字符後，考慮到使用者的寫字習慣，在使用手寫板輸入資訊的同時會記錄全域特徵，以在最後輸出文字時加入適當的特徵調整。

(1) 字元內全域特徵-字元傾斜度



圖(二十三) 掃描線



圖(二十四) 傾斜度示意圖

上圖為示意圖，尺寸為 500x500 像素，圖中梯形內部為使用者所提供的漢字，其梯形的取法首先得找出漢字的最高點座標與最低點座標，以其 Y 值且與 X 軸平行的方程式分別設為梯形的上底與下底；另外再以四條傾斜四十五度的掃描線，得到字的四個端點，如圖(二十五)中紅點，並以其四點分別作出左右斜邊，因此可以得出下列方程式：

$$L_1 : y = \frac{y_1 - y_2}{x_1 - x_2} x + \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2}$$

$$L_2 : y = \frac{y_3 - y_4}{x_3 - x_4} x + \frac{x_3 y_4 - x_4 y_3}{x_3 - x_4}$$

$$H_1 : y = b_1 = b_2 \quad H_2 : y = b_3 = b_4$$

根據上述方程式可以解出 L1、L2 分別與 H1、H2 的四個交點，為(a1,b1)、(a2,b2)、(a3,b3)、(a4,b4)四個座標，由此訂出梯形的四個頂點。將所有使用者樣本所計算出的四個頂點平均，得到一以四個座標點表示的平均傾斜度，依照此數值將整個梯形放大對應到 500x500 的正方形中，產生出的新座標再使用 OpenCV 內建的函數進行字體的扭曲變換。

(2) 字元間全域特徵-字元間距

字元間距的計算是利用掃描線由左到右進行掃描，如右圖中標示的綠線，測

量出每個字的左端線及右端線，計算右端線與下一個字的左端線之距離，如圖中藍色區塊所示，並以得出知各距離平均作為使用者字元間距值。



圖(二十五) 字元間距示意圖

(五) 字元輸出

完成組字的字元在輸出給使用者時必須考慮到傳送及後續處理的方便性，包括檔案的大小，排版的方便性，及旋轉、扭曲、放大縮小時失真的情況。為了解決字元輸出後，放大縮小造成的點陣圖失真，以及點陣圖檔案過大的問題，系統必須將組字完成的圖檔轉為向量檔。將 Raster Image 向量化成為 Vector Image 的過程稱為 tracing (描邊)，而我的系統使用 Potrace 及 Autotrace 這兩個開放原始碼的 tracing engine (根據 GNU GPL 發佈) 進行描邊。



圖 (二十六) 左：原圖 中：Autotrace 右：Potrace

由上圖的測試中可以看出 Potrace 之轉換效果較 Autotrace 好，所以後續研究皆使用 Potrace 進行描邊。Potrace 為 Polygon Tracing Algorithm 之縮寫，為一將 Raster Image 快速且精確的向量化之演算法。利用多個多邊形先近似輸入影像中的所有同色區間，再利用貝茲曲線(Bezier Curve) 擬和多邊形之外框，達到向量化的目的。

在將字元向量化輸出成 SVG 之後，可大幅縮小字型所使用的容量、放大時不影響解析度，且因貝茲曲線近似的限制，同時能將使用者輸入部件時因手寫板讀取不穩定造成的鋸齒狀邊緣消去。

將多個 SVG 影像按照使用者指定的順序嵌入一個 HTML 文件之後，即可使用網頁瀏覽器開啟，利用網頁語言的排版，即可生成能跨平台傳送的個人化字型文章。

多字元輸出測試

圖（二十七）在網頁瀏覽器中顯示的向量化組字結果

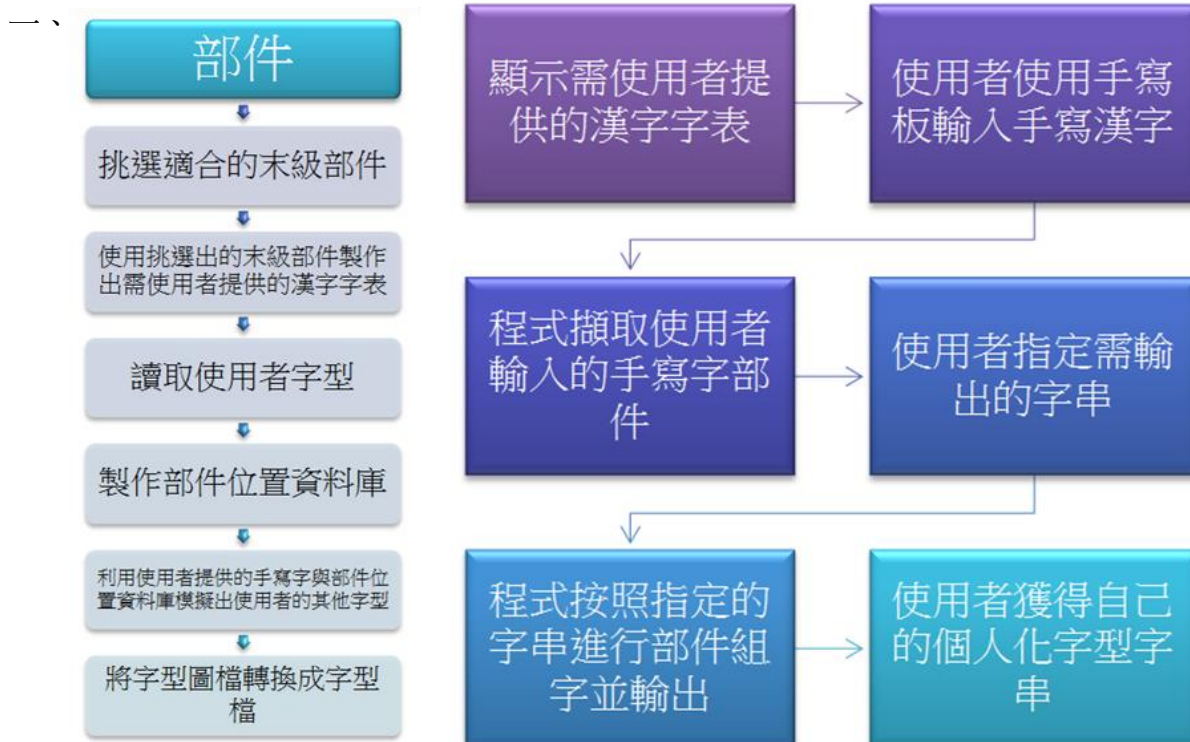
(六)組字系統及程式之介面整合

最後，整合所有程式中所使用到的函式庫(library)，並增加 Qt 函式庫進行 GUI (Graphic User Interface)程式的編寫，利用 Qt Design 編輯程式專案之.ui 檔中的元件(Widget)以建立圖像介面，並建立信號(Signal)與信號槽(Slot)之關聯(QObject::connect)，連結所有程式碼的功能，製作成一套流程完善的圖像使用者介面程式，以便使用者擁有更加簡明化、方便的操作。

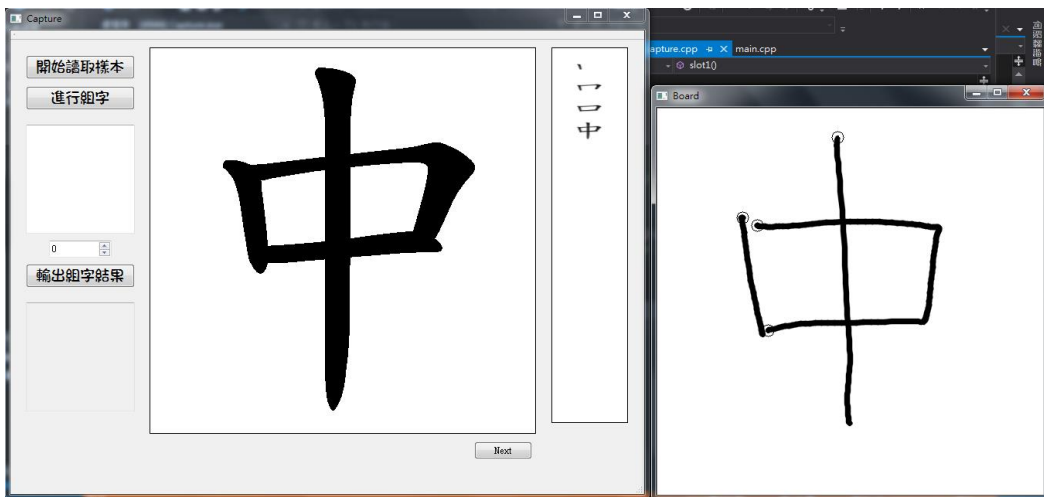
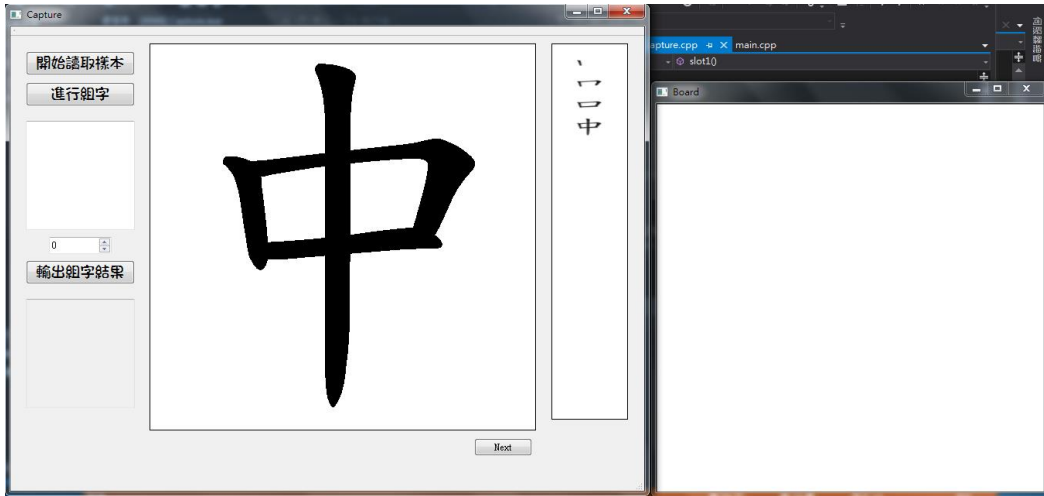
四、書法組字

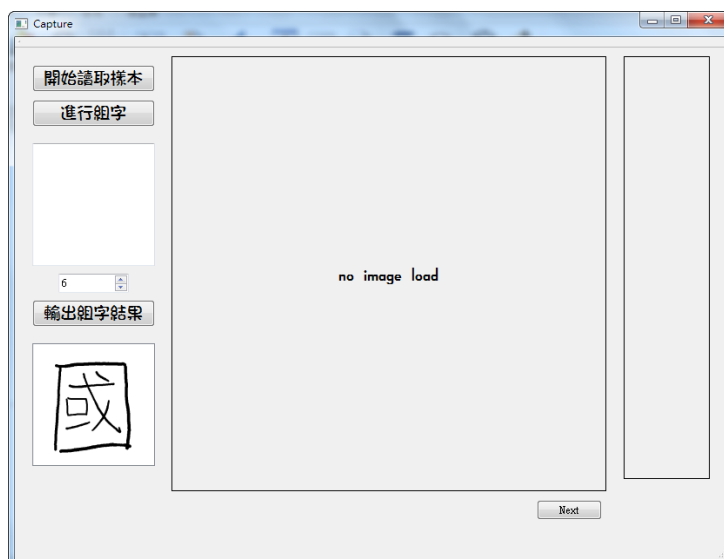
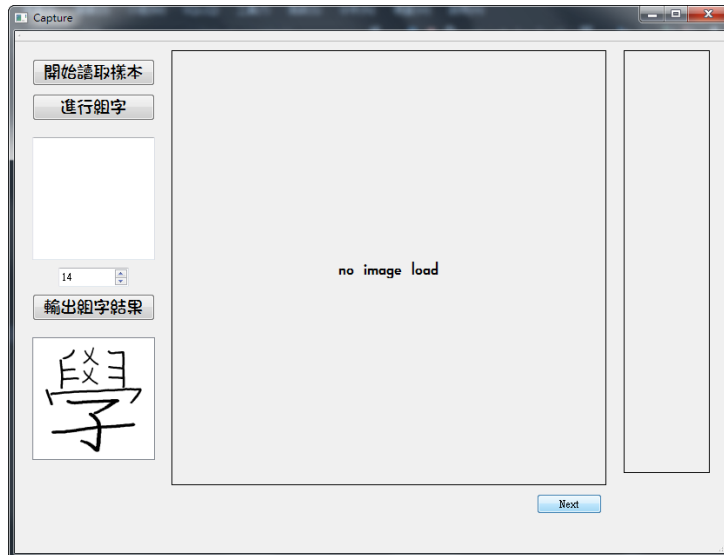
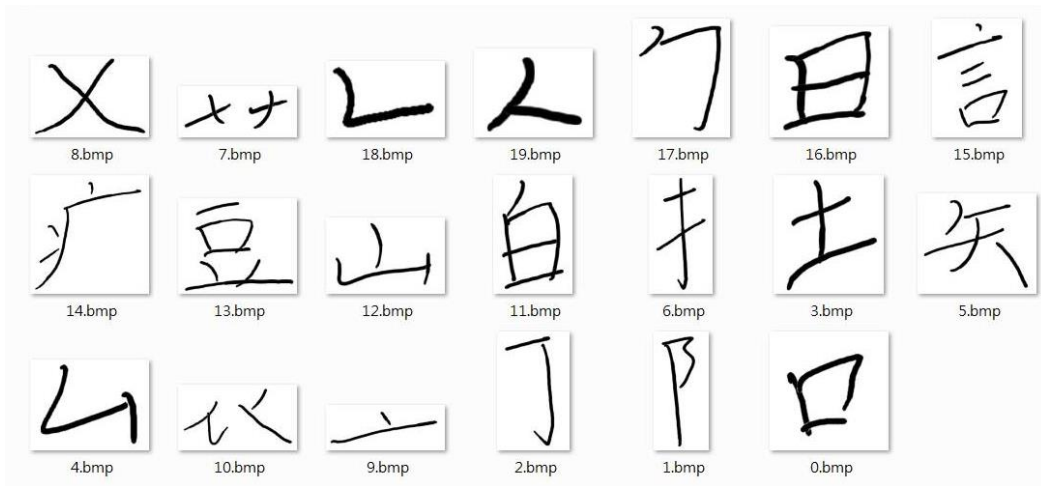
書法組字因為使用者輸入讀取，並不是單筆畫分別輸入，故無法自動將部件切分，但可利用手動的部件切分方法得到書法字之部件，帶入系統，即可做到自動組字。

伍、研究結果



二、系統介面整合與呈現





三、動態組字結果之範例

(一) 使用者提供樣本

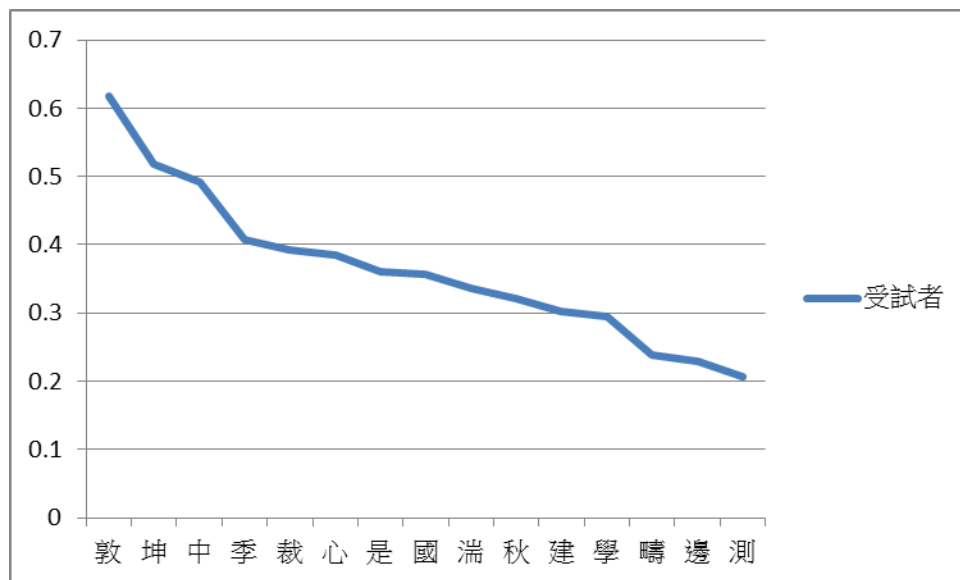
山	中	心	申	仿
而	衣	囡	廷	沖
或	牧	則	哉	臭
郭	堤	窖	筆	愁
遍	壽	覺		

(二) 組字結果

組字結果	實際寫字	組字結果	實際寫字	組字結果	實際寫字
中	中	心	心	坤	坤
季	季	建	建	是	是

秋	秋	國	國	敦	敦
測	測	湍	湍	裁	裁
學	學	壽	壽	邊	邊

四、組字範例分析



表(一) 組字與實際寫字相關係數折線圖

上圖的縱軸為使用 TM_COEFF_NORMED 相關係數公式所計算出的，使用者實際寫字結果與利用部件組字的相關係數數值，橫軸為按照部件數排列的範例字元，根據計算公式的特性，將理想值界定在大於 0.4。

可以清楚的看到整體的數值有從左到右下降的趨勢，判斷造成此結果的原因為部件數的多寡，若一個字的部件數越多，那麼程式組字的結果就較容易失真。

對於「心」部件數不多相關係數卻較低，推測是這個字的空白區域太多，部件落點的變異性過大，導致相關係數算法較難表現其相似度數值。未來盼能尋找更精準表現中文字相似度的方法。

五、書法組字結果之範例

數字為模版匹配中的 Coeff_Normed 相關係數計算結果。

店	店
0.6547	
娃	娃
0.7444	
省	省
0.8603	
忠	忠
0.6264	
星	星
0.7378	

陸、討論與結論

一、筆畫在中文字中的變化空間過大，故筆畫組字無法充分顯露出使用者的寫字特徵，不適宜用於建立個人化字型

二、使用者的末級部件在不同漢字中，只有少許的比例改扁，使用末級部件作為使用者寫字特徵學習的單位，較能表現出寫字特徵

三、經過周全的證明，可以利用貪婪演算法及深度優先搜索，製作出以最少的使用者輸入漢字包含組字所需的末級部件之列表

四、在讀取手寫資料時，利用手寫板的封包監聽，在電腦上重現漢字的筆劃粗細，可以增加重新組字的相似度

五、在自動辨識末級部件位置所使用的模板匹配中，TM_COEFF_NORMED 對於辨識部件是最有成效的。針對一個漢字中含有多個相同部件的問題，可以利用排序演算法及調校後的部件挑選解決。

六、和相關研究的比較

在此列出和此研究目的相似的論文之成果，以及現在市面上之個人化字型製作服務和我的系統的比較。

	本系統	其他論文之類似系統	客製字型服務
處理方式	在線	離線	離線
組字過程	動態部件的替換	手動	
支援字集	大	小	1 : 1
模式學習	部件基 →完整複製特徵	筆畫推演部件 →未考慮筆畫變異	
個人化構成	完整	部分筆畫不改變	最完整
全域特徵	有	無	只有字元內特徵有
部件切割	完整切割	存在許多特例 →無法處理複雜字	

對於現有的研究成果，我希望能在未來實現幾項目標，第一，為使手寫板所讀取之樣本更加準確，對手寫板進行硬體改良，以模擬紙筆寫字之手感；第二，為了讓系統組出的字更像使用者真正手寫的樣子，應再深入進行算法、參數的調整，使組字結果與實際寫字結果擁有更高的相關係數數值；第三，為了減少使用者須輸入字，希望能改善我所使用的演算法。相信在完善的發展後，我的研究可以為這個資訊化的時代增添更加個人化的特質。

柒、參考資料及其他

Viola, P., & Jones, M. (2004). Robust Real-Time Face Detection. *International Journal of Computer Vision* 57(2), 137 – 154.

Zhang, T., & Suen, C. (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Communications of the ACM* 27(3), 236-239.

Lam, L., et al. (1992). Thinning Methodologies—A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(9), 869-885.

Lienhart, R., & Maydt, J. (2002). An Extended Set of Haar-like Features for Rapid Object Detection. *International Conference on Image Processing*.

Kuranov, A., & Lienhart, R. & Pisarevsky, V. (2002). An Empirical Analysis of Boosting Algorithms for Rapid Objects With an Extended Set of Haar-like Features. Intel Technical Report MRL-TR.

Adolf, F. (2003). How-to Build a Cascade of Boosted Classifiers Based on Haar-like Features.

Cv 模式识别。OpenCV 中文網站。取自: <http://wiki.opencv.org>.

cn/index.php/Cv%E6%A8%A1%E5%BC%8F%E8%AF%86%E5%88%AB

直方圖均衡化。維基百科。取自:

<http://zh.wikipedia.org/zh-tw/%E7%9B%B4%E6%96%B9%E5%9B%BE%E5%9D%87%E8%A1%A1%E5%8C%96>

Briechele, K., & Hanebeck, U. Template Matching using Fast Normalized Cross Correlation. Unpublished doctoral dissertation, Universität München, Germany

Template matching. Wikipedia. From http://en.wikipedia.org/wiki/Template_matching

剎那工坊。取自: <http://www.ksana.tw/>

日本網站 kage。取自: <http://fonts.jp/kage/>

日本網站 glyphwiki: <http://glyphwiki.org/wiki/GlyphWiki:%E3%81%A9%E3%81%86%E3%82%84%E3%81%A3%E3%81%A6%E4%BD%BF%E3%81%86%E3%81%AE%E3%81%8B>

捌、附錄

一、末級部件表

0 口	37 大	74 貝	111 勺	148 冊	185 米	222 東	259 亦
1 卩	38 电	75 手	112 魚	149 乏	186 尸	223 見	260 且
2 丁	39 日	76 丰	113 火	150 繼	187 月	224 牙	261 玄
3 土	40 立	77 巾	114 不	151 卞	188 臣	225 冉	262 角
4 厶	41 厂	78 王	115 非	152 言	189 生	226 戊	263 攴
5 矢	42 干	79 文	116 丩	153 糸	190 曲	227 丁	264 而
6 扌	43 肉	80 丿	117 車	154 西	191 束	228 壬	265 川
7 艹	44 木	81 舟	118 彡	155 示	192 冊	229 承	266 豕
8 乂	45 几	82 又	119 井	156 虎	193 尸	230 由	267 重
9 宀	46 匚	83 刀	120 用	157 儿	194 囧	231 亏	268 串
10 衣	47 冂	84 頁	121 皮	158 彡	195 千	232 平	269 囪
11 白	48 央	85 片	122 升	159 主	196 白	233 乙	270 欠
12 山	49 凹	86 半	123 本	160 弋	197 又	234 疋	271 垂
13 豆	50 士	87 辛	124 水	161 止	198 宀	235 也	272 屯
14 疒	51 方	88 瓜	125 辶	162 少	199 艹	236 米	273 卜
15 言	52 攴	89 力	126 冂	163 丘	200 工	237 牛	274 二
16 日	53 灬	90 糸	127 田	164 彡	201 毛	238 耳	275 幺
17 勺	54 十	91 卩	128 鼻	165 内	202 斤	239 齒	276 宀
18 乚	55 羽	92 寸	129 匕	166 秉	203 兔	240 夕	277 肉
19 人	56 衤	93 巾	130 尙	167 业	204 兔	241 尺	278 瓦
20 禾	57 奧	94 丿	131 聿	168 采	205 早	242 赤	279 窓
21 女	58 亠	95 夫	132 彡	169 彡	206 里	243 支	280 彡
22 石	59 彡	96 丰	133 白	170 弓	207 生	244 去	281 业
23 匕	60 巴	97 虫	134 乂	171 子	208 刀	245 中	282 鼠
24 丩	61 冂	98 金	135 尙	172 馬	209 丿	246 龍	283 夕
25 疋	62 力	99 巳	136 畢	173 爻	210 冂	247 酉	284 巳
26 冂	63 丿	100 彡	137 歹	174 自	211 長	248 州	285 日
27 一	64 艹	101 冰	138 必	175 宀	212 申	249 一	286 坐
28 心	65 攴	102 甫	139 广	176 夕	213 止	250 冂	287 羊
29 攴	66 足	103 人	140 丌	177 乂	214 少	251 九	288 丁
30 立	67 乘	104 艮	141 門	178 青	215 彡	252 鬼	289 卮
31 八	68 雨	105 彡	142 才	179 戈	216 果	253 犬	290 弋
32 皿	69 月	106 冂	143 目	180 目	217 犮	254 中	291 隶
33 革	70 皿	107 羊	144 更	181 步	218 手	255 彡	292 丹
34 宀	71 育	108 又	145 自	182 艮	219 辰	256 且	293 辶
35 气	72 父	109 止	146 穴	183 无	220 鹿	257 冂	294 疋
36 亻	73 一	110 彡	147 戶	184 斤	221 宀	258 隹	295 黑

296 鳥	329 亞	362 丹	395 幾	428 支	461 兜	494 冂	527 西
297 至	330 爾	363 百	396 尤	429 莧	462 面	495 舌	528 下
298 丷	331 凡	364 呂	397 小	430 束	463 戍	496 升	529 冂
299 首	332 母	365 卯	398 臣	431 毘	464 末	497 黽	530 缶
300 尋	333 飛	366 兀	399 齊	432 來	465 冂	498 冂	531 冂
301 氏	334 仁	367 卅	400 火	433 良	466 內	499 史	532 血
302 丨	335 市	368 卅	401 夾	434 良	467 乃	500 事	533 熏
303 弟	336 弗	369 业	402 甲	435 勹	468 巾	501 正	534 彥
304 厂	337 風	370 黃	403 冂	436 囟	469 冂	502 朱	535 永
305 皿	338 几	371 龜	404 冂	437 曲	470 年	503 予	536 畚
306 丸	339 夫	372 五	405 兼	438 吏	471 比	504 黍	537 父
307 电	340 冂	373 中	406 東	439 雨	472 丑	505 日	538 曳
308 失	341 冂	374 丿	407 声	440 禺	473 巳	506 毛	539 夷
309 世	342 卅	375 彙	408 与	441 兩	474 爪	507 長	540 已
310 鼎	343 五	376 萸	409 金	442 汧	475 兵	508 四	541 肩
311 丷	344 卜	377 豕	410 夭	443 支	476 尤	509 冂	542 冂
312 重	345 冂	378 毛	411 冂	444 冂	477 并	510 肅	543 玉
313 冂	346 百	379 雀	412 冂	445 王	478 兵	511 系	544 冂
314 斗	347 夕	380 了	413 彗	446 巫	479 聿	512 曷	545 肅
315 門	348 己	381 行	414 切	447 切	480 七	513 太	546 鼎
316 上	349 丂	382 冂	415 冂	448 荒	481 巢	514 聿	547 戊
317 小	350 甘	383 冂	416 皇	449 串	482 冂	515 兆	548 乍
318 毋	351 丂	384 乎	417 圭	450 囟	483 巫	516 戴	549 丈
319 牛	352 羊	385 亞	418 尹	451 反	484 禹	517 凸	550 承
320 廿	353 鬲	386 畢	419 尹	452 率	485 入	518 主	551 之
321 冂	354 井	387 圭	420 久	453 麥	486 頁	519 夕	552 竹
322 兌	355 夬	388 冂	421 求	454 底	487 刃	520 為	553 豕
323 豕	356 小	389 冂	422 巨	455 肅	488 三	521 冂	
324 冂	357 呂	390 夬	423 鞮	456 民	489 又	522 乡	
325 盲	358 井	391 亡	424 夬	457 矛	490 申	523 鳥	
326 身	359 鼓	392 川	425 勻	458 力	491 色	524 于	
327 我	360 骨	393 及	426 甘	459 尸	492 尢	525 五	
328 鳥	361 谷	394 甘	427 冂	460 未	493 册	526 吳	

二、漢字字表

啊阿埃挨哎唉哀皚癘藹
矮艾礙愛隘鞍氨安俺按
暗岸胺案肮昂盎凹敖熬
翱襖傲奧澳芭捌扒叭
吧筈八疤巴拔跋靶把耙
壩霸罷爸白柏百擺佰敗
拜裨斑班搬扳般頒板版
扮拌伴瓣半辦絆邦幫梆

北輩背貝銀倍狽備億焙
被奔笨本笨崩繡甬泵躡
迸逼鼻比鄙筆彼碧蓖蔽
畢斃戔幣庇痹閉敝弊必
辟壁臂避陛鞭邊編貶扁
便變卞辨辯辯遍標彪膘
表鬱慙別瘍彬斌瀕濱賓
擯兵冰柄丙秉餅炳病並

擦猜裁材才財睬睬采彩
菜蔡餐參蠶殘慚慘燦蒼
艙倉滄藏操糙槽曹草廁
策側冊測層贈插叉荏茶
查碴搽察岔差詫拆柴豺
攙摻蟬饑讒纏鎗產聞顛
昌猖場嘗常長償腸廠敞
暢唱倡超抄鈔朝嘲潮巢

榜膀綁棒磅蚌鏘傍滂苞
胞包褒剝薄雹堡飽寶
抱報暴豹鮑爆杯碑悲卑
逞騁杼吃癡持匙池遲弛
馳恥齒侈尺赤翅斥熾充
沖蟲崇寵抽酬壽壽稠愁
籌仇綢瞅醜臭初出櫛廚
躊鋤雛除除楚礎儲矗播
觸處揣川穿椽傳船喘串
瘡窗幢床闌創吹炊捶錘
垂春椿醇唇淳純轟戳綽
疵茨磁雌辭慈瓷詞此刺
賜次聰蔥囟匆從叢湊粗
醋簇促躡篡竄摧崔催脆
瘁粹淬翠村存寸磋撮搓
措挫錯搭達答瘡打大呆
歹傣戴帶殆代貸袋待逮
怠耽擔丹單鄴揮膽旦氮
但憚淡誕彈蛋當擋黨蕩
檔刀搗蹈倒鳥禱導到稻
悼道盜德得的蹬燈登等
瞪凳鄧堤低滴迪敵笛狄
滌翟嫡抵底地蒂第帝弟
遞締顛掂滇碘點典靛墊
電佃甸店惦奠澱殿碉刁
雕凋刁掉吊釣調跌參碟
蝶迭諜疊丁叮叮釘頂鼎
錠定訂丟東冬董懂動棟
侗恫凍洞兜抖鬥陡豆逗
痘都督毒犢獨讀堵睹賭
杜鏞肚度渡妒端短鍛段
斷緞堆兌隊對墩噸敦敦
頓囤鈍盾遁撥多多奪垛
躲朵蹂舵剝情墮娥峨鵝
俄額訛娥惡厄扼遏鄂餓
恩而兒耳爾餌洱二貳發
罰筏伐乏闊法珞藩帆番
翻樊讐釁繁凡煩反返範
販犯飯泛坊芳方肪房防
妨仿訪紡放菲非啡飛肥
匪誹吠肺廢沸費芬酚吩
氛紛紛墳焚汾粉奮份忿
憤冀豐封楓峰鋒風瘋
烽逢馮縫諷奉鳳佛否夫
敷膚孵扶拂幅幅氟符伏

玻菠播撥鉢波博勃搏鉞
箔伯帛舶脖膊勃泊駁捕
蔔哺補埠不布步簿部怖
盪溉幹甘杆柑竿肝趕感
稈敢贛岡剛鋼缸缸綱崗
港杠篙皋高膏羔糕搞鎬
稿告哥歌攔戈鴿貉疙割
革葛格蛤閤隔絡個各給
根跟耕更庚羹埂耿梗工
攻功恭龔供躬公宮弓鞏
汞拱貢共鈎勾溝苟狗垢
構購夠辜菇咕箍估沽孤
姑鼓古壘骨谷股故顧固
雇刮瓜劄寡掛褂乖拐怪
棺關官冠觀管館罐慣灌
貫光廣逛瑰規圭矽歸龜
閨軌鬼詭癸桂櫃跪貴劊
輓滾棍鍋郭國果裹過哈
骸孩海氦亥害駭酣憨邯
韓含涵寒函喊罕翰撼捍
旱憾悍焊汗漢秀杭航壕
壕豪毫郝好耗號浩呵喝
荷荷核禾和何合盒貉閻
河涸赫褐鶴賀嘿黑痕很
狠恨哼亨橫衡恒轟哄烘
虹鴻洪宏弘紅喉侯猴吼
厚候後呼乎忽瑚壺葫胡
瑚狐糊湖弧虎唬護互滬
戶花嘩華猾滑畫劃化話
槐徊懷淮壞歡環桓還緩
換患喚煥煥煥煥宦幻荒
慌黃磺蝗簧皇凰惶煌晃
幌恍諱灰揮輝徽恢蛔回
毀悔慧卉惠晦賄穢會燴
匯諱誨繪葦昏婚魂渾混
豁活夥火獲或惑霍貨禍
擊圾基機畸稽積箕肌饑
跡激譏雞姬績緝吉極棘
輯籍集及急疾汲即嫉級
擠幾脊己薊技冀季伎祭
齋季濟寄寂計記既忌際
妓繼紀嘉伽夾佳家加莢
頰賈甲鉀假稼價架駕嫁
殲監堅尖箋間煎兼肩艱
奸緘爾檢東城揀檢間儉

吵炒車扯撤掣徹澈郴臣
辰塵晨忱沉陳趁襯撐稱
城橙成呈乘程懲澄誠承
驕嬌嚼攪絞矯僥腳狡角
餃繳絞剿教酵轆較叫窖
揭接皆秸街階截劫節桔
傑捷睫竭潔結解姐戒藉
芥界借介疥誠屆巾筋斤
金今津襟緊錦僅謹進靳
晉禁近儘浸盡勁荊兢莖
睛晶鯨京驚精縷經井警
景頸靜境敬鏡徑瘞靖竟
競爭炯窘揪究糾玖韭久
灸九酒廐救舊臼鼻咎就
疚鞠拘狙疽居駒菊局咀
矩舉沮聚拒據巨具距踞
鋸俱句懼炬劇搗鵲娟倦
眷卷絹掇攪抉掘倔爵覺
決訣絕均菌鈞軍君峻俊
竣浚郡駿喀咖卡咯開揩
楷凱慨刊堪勘坎砍看康
慷糠扛抗亢炕考拷烤靠
珂苛柯棵磕顆科殼咳可
渴克刻客課肯啃壘懇坑
吭空恐孔控摳口扣寇枯
哭窟苦酷庫褲誇垮垮跨
胯塊袂會快寬款匡匡狂
框礦眶曠況虧盔歸窺葵
奎魁傀饋愧潰坤昆困困
括擴廓闊垃拉喇臘臘辣
啦萊來賴藍婪欄欄藍蘭
蘭瀾瀾覽覽懶纜爛濼琅
榔狼廊郎朗浪撈勞牢老
佬姥酪烙勞勒樂雷鐸蕾
磊累儘壘播肋類淚稜楞
冷厘梨犁黎籬狸離漓理
李裏鯉禮莉荔吏栗麗厲
勵礫曆利僕例俐痢立粒
瀝隸力璃哩倆聯蓮連鎌
廉憐漣簾斂臉鏈戀煉練
糧涼梁梁良兩輛量晾亮
諒撩聊僚療療寥遼潦了
撻鐸廖料列裂烈劣獵琳
林磷霖臨鄰鱗淋凜賃吝
拎玲菱零齡鈴伶玲凌靈

俘服浮涪福袱弗甫撫輔
俯釜斧脯腑腐赴副覆
賦複傅付阜父腹負富訃
附婦縛咐噶嘎該改概鈔
麓碌露路賂鹿潞祿錄陸
戮驢呂鋤侶旅履屢縷慮
氣律率瀟綠戀孿孿樂卵
亂掠咯掄輪倫侖淪綸論
蘿螺羅邏羅羅裸落洛
駱絡媽麻瑪碼嗎馬罵嘛
嗎埋買麥賣邁脈瞞饅蠻
滿蔓曼慢漫謾芒茫盲氓
忙莽貓茅貓毛矛鉤卯茂
冒帽貌貿麼玫枚梅酶徽
煤沒眉媒鎡每美味寐妹
媚門悶們萌蒙檬盟猛猛
夢孟眯醜靡糜迷謎彌米
秘覓泌蜜密幕棉眠綿冕
免勉媿緬面苗描瞄藐秒
渺廟妙蔑滅民抵皿敏憫
閻明螟鳴銘名命謬摸摹
磨模膜磨摩魔抹末莫墨
默沫漠寞陌謀牟某毋牡
畝姆母墓暮幕慕木目
睦牧穆拿哪訥那娜納
氛乃奶耐奈南男難囊撓
腦惱鬧淖呢餒內嫩能妮
霓倪泥尼擬你匿膩逆溺
蕪拈年碾擲燃念娘醜烏
尿捏聶孽齧錫錫涅您檸
獐凝甯寧寧牛扭鈕紐膿
濃農弄奴努怒女暖虐瘧
挪儒糯諾哦歐鷗毆藕區
偶溫溫叭爬怕怕琶拍排
牌排湃派攀潘盤磐盼畔
判叛兵龐旁榜胖拋咆刨
炮袍跑泡坯胚培裴陪陪
配佩沛噴盆評抨烹澎彭
蓬棚棚篷膨朋鵬捧碰坯
砒霹批披劈髒毗啤脾疲
皮匹痞僻屁譬篇偏片騙
飄漂瓢票撇瞥拚頻貧品
聘乒坪蘋萍平憑瓶評屏
坡潑頗婆破魄迫柏剖撲
鋪僕莆葡菩蒲埔僕圃普

剪減薦檻鑿踐賤見鍵箭
件健艦劍饒漸濺澗建僵
姜將漿江疆蔣漿獎講匠
醬降蕉椒礁焦膠交郊澆
謙乾黔錢鉗前潛遣淺譏
塹嵌欠歉槍槍腔羌牆蕎
強槍橈鞦敲悄橋瞧喬僑
巧鞘撬翹峭倘覈切茄且
怯竊欽侵親秦琴勤芹擒
禽寢沁青輕氳傾卿清擎
晴氛情頃請慶瓊窮秋丘
邱球求囚酋囚趨區蛆曲
驅屈驅渠取娶齣趣去圈
顰權醞泉全痊拳犬券勸
缺缺癩卻鵲權確雀裙群
然燃冉染瓢壤攘嚷讓饒
擾繞惹熱壬仁人忍韌任
認刃妊紉扔仍日戎茸蓉
榮融熔溶容絨冗揉柔肉
茹孺儒孺如辱乳汝入褥
軟阮蕊瑞銳閩潤若弱撒
灑薩腮腮塞賽三三傘散
桑嗓喪搔騷掃嫂瑟色澀
森僧莎砂殺剝少紗傻啥
煞篩曬珊苔杉山刪煽衫
閃陝擅贍膳善汕扇繕墻
傷商賞晌上尚裳梢梢稍
燒芍勺韶少哨邵紹奢賒
蛇舌舍赦攝射攝涉社設
呻申呻伸身深娠紳神沈
審審甚賢慎滲聲生甥牲
升繩省盛剩勝聖師失獅
施濕詩屍虱十石拾時什
食蝕實識史矢使屎駛始
式示士世柿事拭誓逝勢
是嗜噬適仕侍釋飾氏市
侍室視試收手首守壽授
售受瘦獸蔬樞梳殊杼輸
叔舒淑疏書賸孰熟薯暑
曙署蜀黍鼠屬術述樹束
戍豎墅庶數漱恕刷耍摔
衰甩帥拴拴霜雙爽誰水
睡稅吮瞬順舜說碩朔爍
斯嘶嘶思私司絲死肆寺
嗣四伺似飼已松聳慫頌

陵嶺領另令溜琉榴硫留
留劉瘰流柳六龍聾隴籠
隆隆壘攏隴樓婁樓婁漏
陋蘆盧慮慮爐摑鹵虜魯
撻蹋踏胎苔抬台泰酖太
態汰坍攤貪攤灘壇檀痰
潭譚談坦毯袒碳探歎炭
湯塘塘堂棠膛唐糖倘躺
淌趙漫掏濤滔條苟桃逃
淘陶討套特藤騰疼騰梯
剔踢錫提題蹄啼體替嚏
惕涕剃屨天添填田甜恬
舔腆挑條迢眺跳貼鐵帖
廳聽煙丁廷停亭庭挺艇
通桐桐瞳同銅彤童桶捅
筒統痛偷投頭透凸禿突
圖徒途塗屠土吐兔湍團
推頹腿蛻褪退吞屯臀拖
托脫駝陀馱駝橢妥拓唾
挖哇蛙窪娃瓦襪歪外腕
彎灣玩頑丸煇完碗挽晚
皖惋宛婉萬腕汪王亡枉
網往旺望忘妄威巍微危
韋違椽圍唯唯為維維葦
葦委偉偽尾緯未蔚味畏
胃喂魏位渭謂尉慰衛瘟
溫蚊文聞紋吻穩紊問噲
翁甕搗螞渦窩我鞦臥握
沃巫鳴鷄烏于誣屋無蕪
梧吾吳毋武五捂午舞伍
侮塢戊霧晤物勿務悟誤
昔熙析西晒晰嘻吸錫犧
稀息希悉膝夕惜熄熾溪
汐犀檄襲席習媳喜銑洗
系隙戲細瞎蝦匣霞轄暇
峽俠狹下廈夏嚇掀鋏先
仙鮮纖鹹賢銜舷閑涎弦
嫌顯險現獻縣腺餽羨憲
陷限線相廂鑲香箱襄湘
鄉翔祥詳想響享項巷橡
像向象蕭硝霄削哮囂銷
消窩瀉曉小孝校尚嘯笑
效楔些歇蠍鞋協挾攜邪
斜齋諧寫械卸蟹懈泄瀉
謝屑薪芯錚欣辛新忻心

浦譜曝瀑期欺稜戚妻七
 淒漆柒沏其棋奇歧畦崎
 臍齊旗祈祁騎起豈乞企
 啟契砌器氣迄棄汽泣訖
 掐洽牽地鈎鉛千遷簽仟
 眩絢靴薛學穴雪血勳熏
 循旬詢尋馴巡殉汛訊訊
 遜迅壓押鴉鴨呀芽牙蚜
 崖衙涯雅啞亞訝焉咽閹
 煙淹鹽嚴研蜒岩延言顏
 閻炎沿奄掩眼衍演豔堰
 燕厭硯雁唁彥焰宴諺驗
 殃央鴛秧楊揚佯瘍羊洋
 陽氧仰癢養樣漾邀腰妖
 瑤搖堯遙窰謠姚咬啣藥
 要耀柳噎耶爺野冶也頁
 掖業葉曳腋夜液一壹醫
 揖鉞依伊衣頤夷遺移儀
 臆疑沂宜姨彝倚蟻倚已
 乙矣以藝抑易呂屹億役
 臆逸肄疫亦裔意毅憶義
 益溢詣議誼譯異翼翌繹
 茵蔭因殷音陰姻吟銀淫
 寅飲尹引隱印英櫻嬰鷹
 應纓瑩螢營熒迎贏盈
 影穎硬映約擁傭癯庸

送宋訟誦搜艘撒嗽蘇酥
 俗素速粟塑溯宿訴肅酸
 蒜筭雖隨隨綏髓碎歲穗
 遂隧崇孫損筍蓑梭唆縮
 瑣索鎖所塌他它她塔獺
 幽優悠憂尤由郵鈿猶油
 游酉有友右佑紬誘又幼
 迂淤于孟榆虞愚與余俞
 逾魚愉愉漁隅予娛雨與
 嶼禹宇語羽玉域芋郁籲
 遇喻峪禦愈欲獄育譽浴
 寓裕預豫馭鴛淵冤元垣
 袁原援轅園員圓猿源緣
 遠苑願怨院曰約越躍鑰
 嶽粵月悅閱耘雲鄧勻隕
 允運蘊醞暈韻孕匝砸雜
 栽哉災宰載再在咱攢暫
 贊贓躄葬遭糟鑿藻棗早
 澡蚤躁噪造皂灶燥責擇
 則澤賊怎增憎曾贈紮喳
 渣劄軋鋤閤眨柵榨乍乍
 炸詐摘齋宅窄債寨瞻氈
 詹粘沾盞斬輾嶄展蘸棧
 占戰站湛綻樟章彰漳張
 掌漲杖丈帳賬仗脹瘴障
 招昭找沼趙照罩兆肇召

信響星腥猩惺與刑型形
 邢行醒幸杏性姓兄凶胸
 匈洵雄熊休修羞朽嗅鏽
 秀袖繡墟戌需虛噓須徐
 許蓄酗敘旭序畜恤絮婿
 珍斟真甄砧臻貞針偵枕
 疹診震振鎮陣蒸掙掙征
 爭爭怔整拯正政幘症鄭
 證芝枝支吱蚰知肢脂汁
 之織職直植殖執值侄址
 指止趾只旨紙志摯擲至
 致置幟峙制智秩稚質炙
 痔滯治窒中盅忠鐘衷終
 種腫重仲眾舟周州洲謫
 粥軸肘帚咒皺宙晝驟珠
 株蛛朱豬諸誅逐竹燭煮
 拄矚囑主著柱助蛀貯鑄
 築住注祝駐抓爪拽專磚
 轉撰賺篆椿莊裝妝撞壯
 狀椎錐追贅墜綴諄准捉
 拙卓桌琢茁酌啄灼濁茲
 咨資姿滋淄孜紫仔籽滓
 子自漬字鬢棕蹤宗綜總
 縱鄒走奏揍租足卒族祖
 詛阻組鑽纂嘴醉最罪尊
 遵昨左佐柞做作坐座

三、漢字部件拆分表(節錄第 1~100 個字)

1 啊
 4 口, 讠, 丁, 口
 2 阿
 3 讠, 丁, 口
 3 埃
 3 土, 厶, 矢
 4 挨
 3 扌, 厶, 矢
 5 哎
 3 口, 讠, 义
 6 唉
 3 口, 厶, 矢
 7 哀
 3 宀, 衣, 口
 8 皑
 3 白, 山, 豆
 9 癌
 5 疒, 口, 口, 口, 山

12 艾
 2 艹, 义
 13 礙
 5 石, 匕, 矢, 丿, 正
 14 愛
 4 忄, 宀, 心, 爻
 15 隘
 4 阝, 立, 儿, 皿
 16 鞍
 3 革, 宀, 女
 17 氨
 3 气, 宀, 女
 18 安
 2 宀, 女
 19 俺
 3 亻, 大, 电
 20 按
 3 扌, 宀, 女

23 胺
 3 肉, 宀, 女
 24 案
 3 宀, 女, 木
 25 肮
 3 肉, 宀, 几
 26 昂
 3 日, 匕, 冂
 27 盎
 2 央, 皿
 28 凹
 1 凹
 29 敖
 3 土, 方, 爻
 30 熬
 4 土, 方, 爻, 灬
 31 翱
 4 白, 大, 十, 羽

10 藹
6 艹,言,日,勺,匕,人
11 矮
3 矢,禾,女
34 奧
1 奧
35 懊
2 忄,奧
36 澳
2 氵,奧
37 芭
2 艹,巴
38 扒
2 扌,八
39 叭
2 口,八
40 吧
2 口,巴
41 笆
2 竹,巴
42 八
1 八
43 疤
2 疒,巴
44 巴
1 巴
45 拔
2 扌,扌
46 跋
2 足,扌
47 靶
2 革,巴
48 把
2 扌,巴
49 耙
2 耒,巴
50 壩
4 土,雨,革,月
51 霸
3 雨,革,月
52 罷
4 皿,育,匕,匕
53 爸
2 父,巴
54 白
1 白

21 暗
3 日,立,日
22 岸
3 山,厂,干
57 擺
5 扌,皿,育,匕,匕
58 佰
3 亻,一,白
59 敗
2 貝,攴
60 拜
3 扌,一,丰
61 稗
3 禾,卑,十
62 斑
3 王,文,王
63 班
3 王,丿,王
64 搬
4 扌,舟,几,又
65 扳
3 扌,厂,又
66 般
3 舟,几,又
67 頒
3 八,刀,頁
68 板
3 木,厂,又
69 版
3 片,厂,又
70 扮
3 扌,八,刀
71 拌
2 扌,半
72 伴
2 亻,半
73 瓣
3 辛,瓜,辛
74 半
1 半
75 辦
3 辛,力,辛
76 絆
2 糸,半
77 邦
2 丰,卩

32 襖
2 衤,奧
33 傲
4 亻,土,方,攴
80 榜
5 木,宀,灬,宀,方
81 膀
5 肉,宀,灬,宀,方
82 綁
3 糸,丰,卩
83 棒
3 木,夫,丰
84 磅
5 石,宀,灬,宀,方
85 蚌
2 虫,丰
86 鏘
5 金,宀,灬,宀,方
87 傍
5 亻,宀,灬,宀,方
88 謗
5 言,宀,灬,宀,方
89 苞
3 艹,勺,巳
90 胞
3 肉,勺,巳
91 包
2 勺,巳
92 褒
5 宀,衣,亻,口,木
93 剝
3 立,水,丩
94 薄
4 艹,彳,甫,寸
95 雹
3 雨,勺,巳
96 保
3 亻,口,木
97 堡
4 亻,口,木,土
98 飽
5 人,一,貝,勺,巳
99 寶
6 宀,王,艹,十,口,貝
100 抱
3 扌,勺,巳

55 柏
2 木,白
56 百
2 一,白

78 幫
5 土,土,寸,白,巾
79 柳
3 木,丰,卩

四、使用貪婪演算法製作之使用者輸入字表

躊矮抱殿窺愛掣檢欽啥
聽享彬產諷扶淮妓兼琳
淋拿稀像星藻柄詞竄燈
固官監劍坎錄批縣旬氨
盍熬傲壩磅暴鮑億幣柄
病捕參慘摻唱徹稱橙馳
斥彈墊犢俄符斧副苟拐
圭跪孩函煥慌稼檻姐謹
枯傀厲掠妹免錫濃徘漂
汽簽錢然任剝樞庶帥寺
恬析腺朽涯浴詹宙阿按
腓凹敖扒跋爸稗搬伴半
電碑本鼻比畢斃閉敞邊

扁卞辨瀕瀆冰餅坡步殘
艙槽曹側測蹭荏茶拆柴
饑償鈔吵沉趁逞秤吃癡
赤翅抽疇綢櫛鋤雞除觸
川穿船瘡幢錘蠱蔥從篡
錯大殆貸怠單蹈鳥到的
滴帝點電佃參蝶錠懂陡
毒段隊對朵厄餌番凡肥
廢佛輔該概蓋簪更羹龔
弓古股寡掛關灌廣逛閨
夯賀亨侯乎葫華劃基機
畸積激幾己佳鉀肩城嬌
窘睫界浸境韭舊舉掘菌

款蠟來榔籬狸栗勵聯梁
聊僚吝陵琉婁虜履濾賣
瞞饅忙錨眉媒面滅末吶
鈉乃撓匿碾攆瘡帕龐咆
頻坪戚妻薈趣泉繞三騷
瑟澀苦上繩省十矢拭抒
鼠甩伺飼酸他太搪逃帖
禿旺忘濼紊錫誣毋晒厦
蟹絮熏閭漾窳業腋遺乙
音英螢盈臙冤躍增丈掙
枝燭茁

五、研究使用字體例圖



圖(二十六) normal



圖(二十七) Simsun

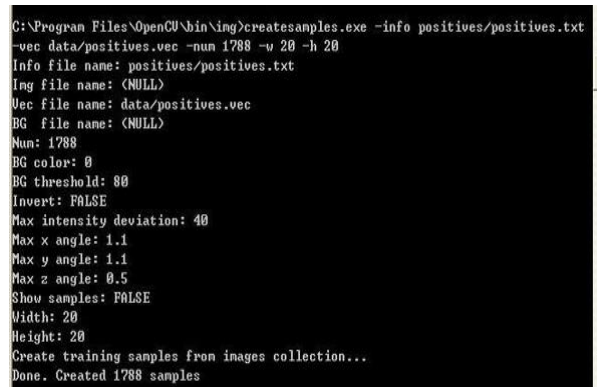


圖(二十八) tensentye

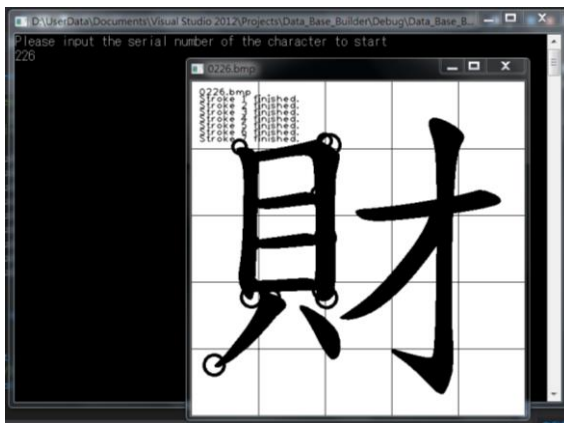
六、程式畫面



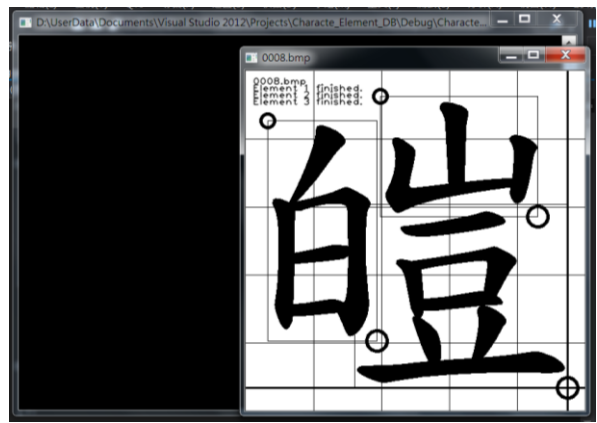
圖(二十九) 使用者模式學習



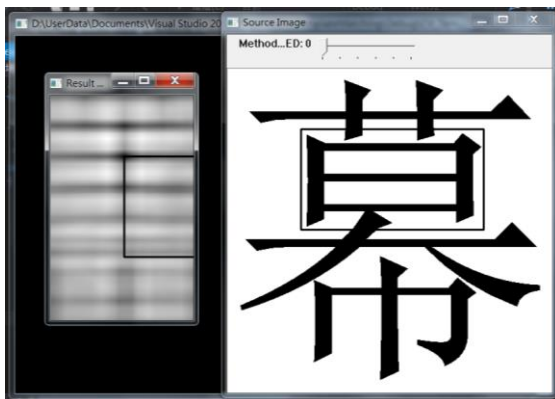
圖(三十) 訓練樣本製作



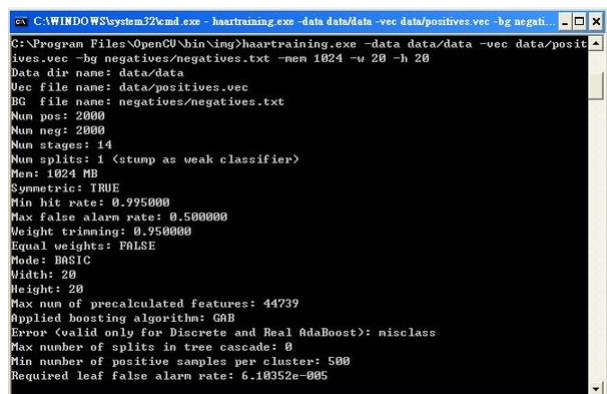
圖(三十一) 筆畫資料庫



圖(三十二) 部件資料庫



圖(三十三) 部件匹配



圖(三十四) 分類器訓練