

國立中央大學

資訊工程研究所

碩士論文

三維圓柱曲面上的文字偵測與校正

Text Detection and Deskew on 3D Cylinders

指導教授：范國清 博士

溫敏淦 博士

研究生：謝宗廷

中華民國九十八年七月



國立中央大學圖書館

碩博士論文電子檔授權書

(98 年 4 月最新修正版)

本授權書所授權之論文全文電子檔(不包含紙本、詳備註 1 說明)，為本人於國立中央大學，撰寫之碩/博士學位論文。(以下請擇一勾選)

☒ **同意** (立即開放)

☐ **同意** (一年後開放)，原因是：_____

☐ **同意** (二年後開放)，原因是：_____

☐ **同意** (三年後開放)，原因是：_____

☐ **不同意**，原因是：_____

以非專屬、無償授權國立中央大學圖書館與國家圖書館，基於推動「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、公開陳列、與發行，或再授權他人以各種方法重製與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名： 謝宗廷 學號： 955202039

論文名稱： 三維圓柱曲面上的文字偵測與校正

指導教授姓名： 范國清

系所： 資訊工程 所 ☐ 博士班 ☒ 碩士班

日期：民國 98 年 7 月 30 日

備註：

1. 本授權書之授權範圍僅限電子檔，紙本論文部分依著作權法第 15 條第 3 款之規定，採推定原則即預設同意圖書館得公開上架閱覽，如您有申請專利或投稿等考量，不同意紙本上架陳列，須另行加填聲明書，詳細說明與紙本聲明書請至 <http://thesis.lib.ncu.edu.tw/> 下載。
2. 本授權書請填寫並**親筆**簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
3. 請加印一份單張之授權書，填寫並親筆簽名後，於辦理離校時交圖書館（以統一代轉寄給國家圖書館）。
4. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

摘要

隨著科技的日新月異造成的進步，電子產品的功能也越來越強大。在過去需要透過掃描機才能取得的高解析度文字影像，在現代卻往往只要透過一個普通的數位相機便能夠取得，相較起來更具有可攜性與便利性。受此影響，文字辨識的範圍也不再侷限於過往的平面文字影像，而延伸到了更生活化的三維立體影像中。

而在文字影像不再屬於單純的平面以後，對於文字影像辨識也帶來了更多問題。其中最主要的問題是文字所屬平面帶來的文字變形。在圓柱影像中，位於圓柱兩側的文字在影像中會因相機並非完全水平面對文字而呈現傾斜，除了文字變形外，也造成原來水平的文字行變形成一條曲線，也進而導致了文字辨識的困難。

本研究提供了一個有效正確的圓柱文字影像校正方法，利用影像前處理包含全域二值化、文字區塊標記來擷取欲分析之影像資料。其次將利用本研究中所提出之連通元件將文字字元進行正確的外型標記。而後將連通元件連結成曲線文字行，在以迴歸分析分析出曲線方程式後，將曲線文字行校正為直線，並可將此正確之校正結果提供後續的切割與辨識系統之用。

Abstract

Due to the rapid development of scientific technology, electronic products becomes more advanced functions. It used to use scanner to get high resolution 2D image, but now we can do that by using ordinary digital cameras. Portability and convenience of using digital cameras. Character recognition is no longer limited by recognizing 2D characters. It can also be extended to recognize 3D characters because of the inherent characteristic of digital cameras.

Since the images and texts are no longer planes, the recognition of them has brings in many problems. The major problem is that the plane to which the texts belong causes the sphere change of texts. When it formed the images of cylinder, the texts locate on both sides of the cylinder will slant because the camera is not totally horizontal in the images. In addition to the change of texts, the texts horizontal will be transformed into a curve which increases the difficulty in recognizing the texts.

In this thesis, we present an effective method to correct the text images of cylinder sphere. Firstly, image preprocessing is performed including global binarization and connected-component labeling to extract the image information. Next, a modified connected-component labeling is employed to will correct the labeling characters, and link these components to a curved text line. After the using of regression analysis to analyze the curve function, we will correct the curved text line to a horizontal line. The result can be used to facilitate later segmentation and recognition system.

誌謝

三年的研究所生活一下子就過去了，在圖形識別實驗室的這三年中，我學到了很多事情，也認識了許多可貴的朋友。首先要感謝的是我的老師 范國清教授與 溫敏淦教授，從一開始研究方向的擬定、研究方法的討論，以及後續論文、口試投影片的修改，老師無不盡心指導。並且在我遇到一些困難和挫折時，老師仍能耐心的給予包容，並且提供我各種資源去解決問題，在此致上最誠摯的感謝之意。也要感謝口試委員李錫堅教授、范欽雄教授，以及黃仲陵教授在口試時不吝給予我論文上的寶貴意見與指教，使我的論文得以更臻完善。

感謝圖形識別實驗室的博士班學長，在我有課業上的問題時總能給予適時的幫助，尤其是呂信德 博士，從一開始訂定研究方向後就不斷的給予許多指導，並且在程式與論文上都很有耐心的幫助我解決許多難題。感謝與我同一屆，常常拉我出門運動的少鵬、映宇，與看似虛弱卻外向好動的柏棟；主動找我聊天的學弟妹力中、建業、霏嫻；以及更小一屆，卻積極進取的學弟宗湧。這三年來你們真的帶給了我許許多多的回憶。因為有你們，這三年的生活才能如此多采多姿。

最後，要將最真摯的感謝，獻給我親愛的家人，他們經常關心我研究所的生活，並且給我鼓勵與支持，尤其是我的父親 謝景棠先生和母親 徐美貞女士，感謝你們這三年來的支持。如果沒有你們不斷的給予幫助和關懷，這本論文也無法順利的完成。

付梓在即，心中充滿歡喜，再次感謝曾經幫助過我的每一個人，衷心謝謝大家。

目錄

Abstract.....	i
目錄.....	iii
第一章 緒論.....	1
1.1 研究動機與目的.....	1
1.2 相關研究.....	3
1.3 系統流程.....	5
1.4 論文架構.....	8
第二章 前處理.....	9
2.1 彩色影像轉換灰階影像.....	11
2.2 全域二值化.....	12
2.3 連通元件偵測.....	14
2.4 雜訊去除與圖文分離.....	17
2.5 文字行串連.....	17
2.6 區域二值化.....	19
第三章 曲線文字行繪製.....	21
3.1 連通元件修正.....	21
3.1.1 文字分類.....	29
3.1.2 連通元件繪製.....	29

3.2 連結曲線文字行	29
3.2.1 橢圓逼近與拋物線逼近	29
3.2.2 拋物線回歸分析	30
第四章 曲線文字校正	36
4.1 曲線 Typoline 繪製	37
4.1.1 K-means 演算法介紹	29
4.1.2 Typoline 演算法介紹	38
4.1.3 以 Typoline 修正連通元件	39
4.2 曲線文字行校正	40
4.3 文字寬度校正	41
第五章 實驗結果與討論	44
5.1 曲線文字行偵測結果	44
5.2 曲線文字行校正及其結果	48
5.3 圓柱寬度校正及其結果	52
5.4 討論	57
第六章 結論與未來工作	62
6.1 結論	62
6.2 未來工作	62
參考文獻	64

附圖目錄

圖 1.1：圓柱影像示意圖。	2
圖 1.2：圓柱影像圖。	2
圖 1.3：Hiroyuki Hase 分類範例圖。(a)類型一(b)類型二(c)類型三(d) 類型四	5
圖 1.4：系統流程圖。	7
圖 2.1：前處理流程圖。	10
圖 2.2：RGB 色彩空間圖。	11
圖 2.3：灰階值統計直方圖。	12
圖 2.4：四連通、八連通示意圖。	14
圖 2.5：八連通演算法示意圖。	15
圖 2.6：連通元件偵測結果。(a)(b)原始影像(c)(d)灰階影像(e)(f)二值化 影像(g)(h)連通元件偵測影像	16
圖 2.7：文字行串連流程圖。	18
圖 2.8：文字區塊串連結果。	19
圖 2.9：區域二值化流程圖。	20
圖 2.10：區域二值化結果圖。	20
圖 3.1：以直線逼近圓柱文字的文字行。	21
圖 3.2：以矩形進行連通元件標記。(a) 水平文字影像(b)圓柱文字影像	

.....	22
圖 3.3：文字分類依據示意圖。	23
圖 3.4：類型一連通元件。	23
圖 3.5：類型二連通元件。	24
圖 3.6：類型三連通元件。	24
圖 3.7：類型四連通元件。	24
圖 3.8：四種文字類型的平行四邊形連通元件取得法。(a)分類一(b)分類二(c)分類三(d)分類四	28
圖 3.9：Typoline 與連通元件分佈示意圖。	30
圖 3.10：Typoline 示意圖 1。	31
圖 3.11：Typoline 示意圖 2。	31
圖 4.1：圓柱曲面取像側面圖。	36
圖 4.2：橢圓文字影像寬度縮短示意圖。	37
圖 4.3：曲線 Typoline 圖。	39
圖 4.4：Typoline 連通元件修正圖。	40
圖 4.5：曲線文字行校正結果。	41
圖 4.6：寬度校正示意圖 1。	41
圖 4.7：寬度校正示意圖 2。	42
圖 4.8：曲線文字行校正結果。	43
圖 5.1：六百毫升寶特瓶曲線文字行偵測結果。(a) 原圖(b)方程式逼近	

結果.....	45
圖 5.2：四百公克果醬瓶曲線文字行偵測結果。(a) 原圖(b)方程式逼近結果.....	46
圖 5.3：四百一十公克玉米罐曲線文字行偵測結果。(a) 原圖(b)方程式逼近結果	47
圖 5.4：六百毫升寶特瓶曲線文字行校正圖。(a)曲線 Typoline(b)校正結果(c)校正前辨識結果(d)校正後辨識結果.....	49
圖 5.5：四百一十公克玉米罐曲線文字行校正圖。(a)曲線 Typoline(b)校正結果(c)校正前辨識結果(d)校正後辨識結果。	50
圖 5.6：四百公克果醬瓶曲面文字行校正圖。(a)曲線 Typoline(b)校正結果(c)校正前辨識結果(d)校正後辨識結果.....	51
圖 5.7：六百毫升寶特瓶校正圖。(a)曲線 Typoline 校正結果(b)寬度校正結果(c)校正前辨識結果(d)校正後辨識結果	53
圖 5.8：四百一十公克玉米罐文字影像校正圖。(a)曲線 Typoline 校正結果(b)寬度校正結果(c)校正前辨識結果(d)校正後辨識結果.....	54
圖 5.9：四百公克果醬瓶文字影像校正圖。(a)曲線 Typoline 校正結果(b)寬度校正結果(c)校正前辨識結果(d)校正後辨識結果	55
圖 5.10：複雜背景影像圖及其校正與辨識結果。	58
圖 5.11：二值化造成的文字破碎影像圖及其校正與辨識結果。	58

圖 5.13：連字與藝術字影像圖及其校正與辨識結果。	61
----------------------------------	----

附表目錄

表 3.1：數字連通元件分類。	25
表 3.2：大小英文字母連通元件分類。	25
表 3.3：小寫英文字母連通元件分類。	25
表 3.4：連通元件上方端點分佈表。	32
表 3.5：連通元件下方端點分佈表。	33
表 4.1：寬度校正比例表。	43
表 5.1：Fujifilm FinePix 700 規格。	44
表 5.2：實驗影像的平均均方根誤差與 R 平方係數。	48
表 5.3：實驗影像曲線校正前與校正後的辨識率比較。	52
表 5.4：60 張實驗影像曲線校正前與校正後的辨識率比較。	52
表 5.5：實驗影像寬度校正前與校正後的辨識率比較。	56
表 5.6：60 張實驗影像寬度校正前與校正後的辨識率比較。	56
表 5.7：實驗影像及校正結果比較。	60
表 5.8：演算法辨識率比較。	60

第一章

緒論

1.1 研究動機與目的

在影像處理技術中，文字辨識是受到廣泛研究的一塊區域。而隨著科技技術的研究發展，文字辨識也已逐漸從過往的高解析度文件掃描影像，往日常生活中隨處可見的數位相機(digital camera)或是網路攝影器(webcam)取得影像。而隨著取像設備的改變，文字影像也不再侷限於過往的平面文件影像，而會延展到我們生活空間中各式各樣的物體上。

圓柱體，如鋁罐、鐵罐、寶特瓶上的文字，由於文字影像多半只能採用數位相機或網路攝影器取得，影像內的文字除了可能不如使用掃描器取像來的完整、清晰外，往往有多種顏色混雜其中，增加了影像複雜度；也由於圓柱體形狀造成的三維立體影響，導致文字排列不如過往平面文字影像行那般屬於一條水平的長直線，更因此造成了位於圓柱體兩旁，非垂直取向光軸文字的扭曲與變形。要將文字串列以提供後端 Optical Character Recognition(OCR)軟體辨識，已經超出過去使用如 xy cut 等簡易文字擷取方式的適用範圍。參考下圖 1-1，其中 A 為相機位置。原來 B、C、D 三點連線於平面上將會是一條水平直線，但卻因為圓柱曲面的影像而導致 B、C、D 三點連線成為曲線。圖 1-2 則為現實生活上的實際例子。

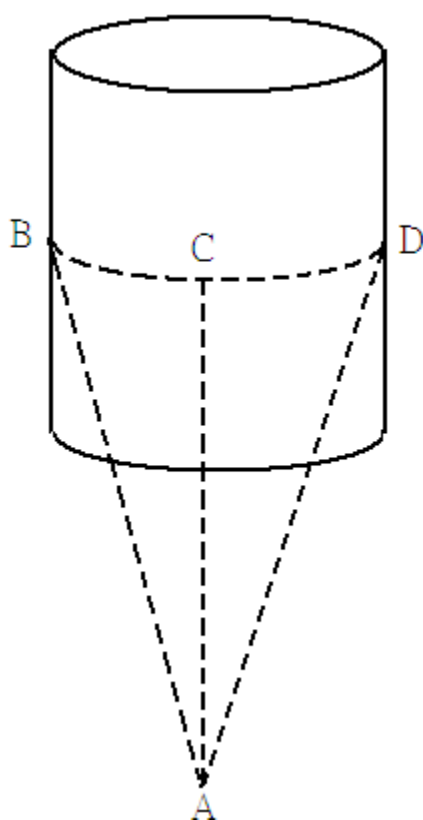


圖 1-1 圓柱影像示意圖



圖 1-2 圓柱影像圖

為了正確將圓柱造成的文字扭曲進行校正，本篇論文提出以串連弧形文字行的方法，建立一個文字校正系統，經由偵測並分析曲線文字行，進而將曲面文字進行校正。我們採用的影像是文字在圓柱上呈現水平分佈，並在取像範圍中包含圓柱邊界的影像，在此並不對圓柱

體上傾斜分佈的文字行進行考慮。以此做為前題，實驗證明本篇論文提出的方法可以有效應用在圓柱體的文字影像校正辨識上，提升文字分析、辨識的正確率，增加文字辨識的應用性。

1.2 相關研究

以下將簡述與本篇圓柱曲線文字校正系統相關研究中，近來一些較為重要的，或是與本論文所討論之重點有高度相關的文獻內容。由於在資料搜尋中並沒有找到針對圓柱文字的校正方式，相關研究的內容著重在平面文字影像的曲線文字校正上。

Hong Yan [1] 提出了 Fuzzy Curve-Tracing 演算法，對圖像資料進行分群，接著以曲線連接各個群聚中心，最後將連接出來的曲線進行平滑化。他並將此演算法應用在曲線文字影像上[2]。

Benson S. Y Lum [3] 等人改進了 Fuzzy curve-tracing 演算法。他們運用一個分類器使得目標圖像能被分割成多個更具代表性的群聚中心，也改善了 Fuzzy curve-tracing 演算法在對銳角曲線擬合與曲線交錯時容易出錯的問題。

Godfried Toussaint [4] 提出了 Rotating Caliper 這個快速而方便的區域框記工具。利用兩條平行線夾住目標區域進行旋轉，以獲得目標區域在各個角度上的資訊。

Abdelwahab Zramdini [5] 等人提出了以 Typoline 分析文字的概念。此概念以水平的四條線將文字分割成多種類型，並以這些特徵加強文字辨識的功能。

Takuma Yamaguchi [6] 等人針對招牌上的電話號碼提出了一個數位分類系統。其中以 Hough transform 計算出各文字的傾斜角，在對傾斜文字進行校正。

Hideaki Goto[7]等人提出一個偵測文件方向的方法。將文件分成多個小區塊判斷方向，利用行間距大於字間距的特性進行文字行連接，最後抽取部分文字辨識進行確認。

Hideaki Goto [8] 等人提出一個曲線文字行的抽取系統。此方法首先對整張文件做六個方向的文字方塊偵測，以取出包覆其中各個文字的最小方塊，接著利用上段提出的方法[7]將文件分割成多個小區塊分別進行連接，將文字方塊組合成單字以後，藉著判斷可能的連接角度進行各單字的連接，完成最後的曲線文字行。

Zheng Zhang[9] 等人完成了一個厚書本影像中書本中央處文字扭曲的校正系統。他們將文件分成正常部分與扭曲部分，對扭曲文字利用字母間不會瞬間出現過大角度變化的特性畫出文字區塊，再用最小平方差整理出文字行的曲線參數，進行最後的校正。

對於平面影像的歪曲文字行，Hiroyuki Hase [10] 等人提出了一個校正系統。首先以矩形連通元件對文字進行標記，再以兩條拋物線分別對文字連通元件的上緣與下緣進行逼近，並以逼近結果判斷出字元方向與文字行方向的傾斜角。Hiroyuki Hase 利用連通元件的分佈與文字傾斜角將文字行分為四種類別。類別一的連通元件分佈為一條直線，文字傾斜角呈現統一的垂直，如圖 1-3(a)；類型二的文字行分佈同樣為一條直線，文字傾斜角則呈現同角度的傾斜，如圖 1-3(b)；類型三的字元分佈為一條曲線，其文字傾斜角為垂直，如圖 1-3(c)；類型四的字元分佈同樣為曲線，字元角度則依各個文字在文字行中的位置而不同，如圖 1-3(d)。其中以類型四最接近本論文的目标。



ABCDE

ABCDE



圖 1-3 Hiroyuki Hase 分類範例圖

(a)類型一(b)類型二

(c)類型三(d)類型四

在校正上，Hiroyuki Hase 對於四種類型的文字分別採用不同的校正法。針對類型一，會直接旋轉整串文字至水平方向；對於類型二，則會在計算斜率後給與傾斜校正；對於類型三，會針對各個字元進行旋轉校正後再依照相對位置重新排列成水平文字行；對於類型四，則會在計算曲線文字行方程式的微分系數後，對應各個字元的位置進行校正。

1.3 系統流程

本曲面文字校正系統可以分成三個步驟：影像前處理，曲線文字行逼近與曲線文字行校正。圖像資料採用數位相機取像，其中前處理包含了二值化以及文字區塊偵測，連通元件抽取與文字行偵測，則為本曲面文字校正系統的核心部分。

本論文的系統流程如圖 1-3 所示。系統第一個階段為前處理，在影像前處理部分，對影像進行灰階處理，並使用全域二值化讓前景與背景區分開來，之後對前景物進行連通元件偵測，以標記文字區塊並對文字區塊擷取相關資訊使用。

完成初步的文字連通元件後，第二階段將針對連通元件進行修正，使之更能符合曲線文字行的需求。之後便可將圓柱上的連通元件連結成曲線文字行，於第三階段進行此曲線的方程式逼近。

在串連曲線文字行後可得知在該文字行中共有哪些文字，依照此橢圓方程式進行平移，藉由對齊文字邊界以可出該文字行的 TypoLine。最後的第四階段，會在整理上述方程式資訊後以仿射轉換做曲面文字校正。校正完畢後的文字將送進後續的 OCR 系統進行辨識。

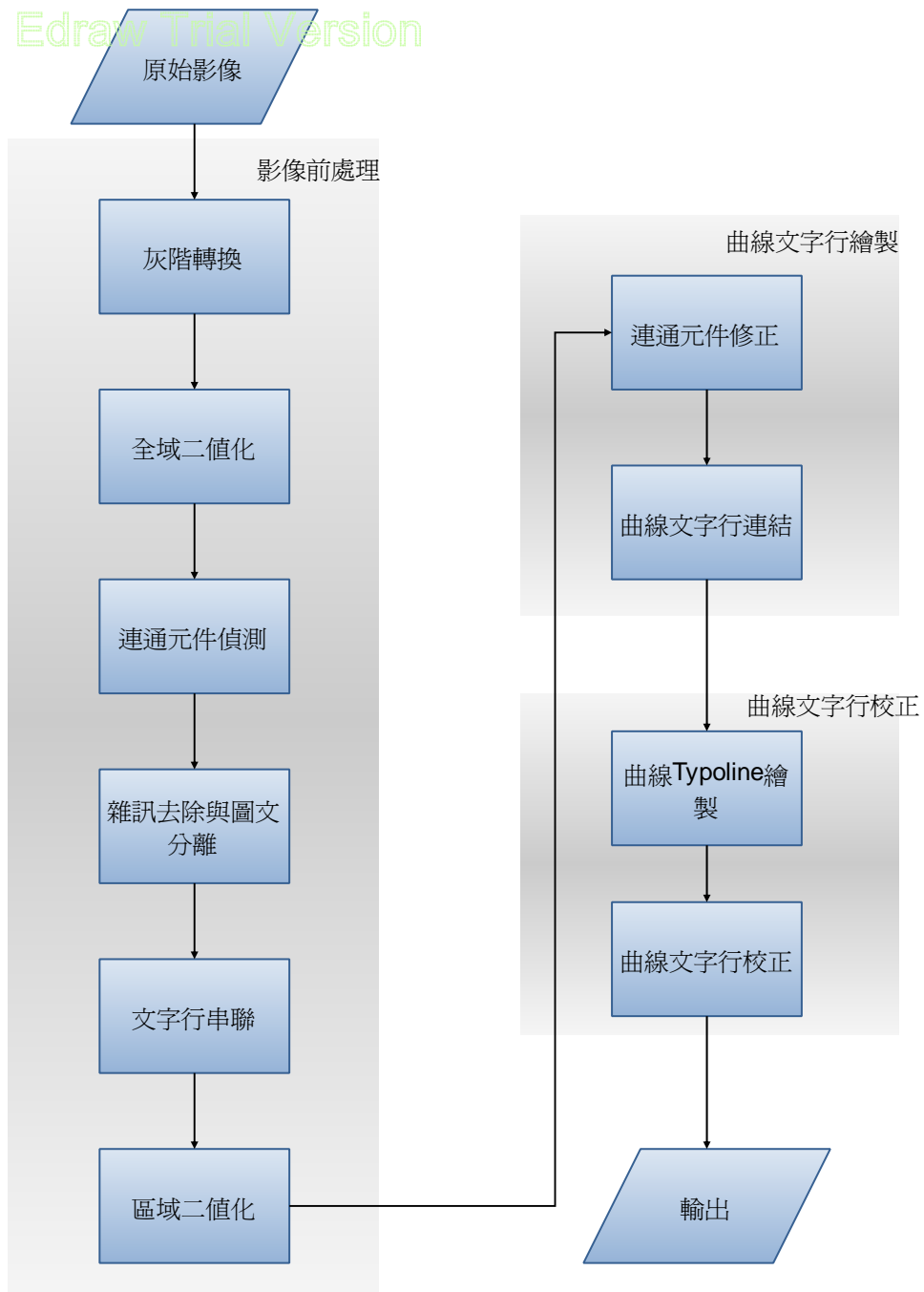


圖 1-3：系統流程圖

1.4 論文架構

本論文包含六個章節，各章節內容簡述如下：

- 第一章 說明研究動機、目的以及實驗系統的流程與架構。
- 第二章 敘述對輸入影像所做的各項前處理步驟，包含灰階處理、二值化，以及初步的連通元件偵測。
- 第三章 詳述本論文使用的連通元件特徵，以及一個新的曲線文字行抽取方法。
- 第四章 說明如何利用曲線文字行資訊對扭曲文字進行正確校正。
- 第五章 實驗的設計與討論。
- 第六章 本論文之結論及未來可再進行之研究方向。

第二章

前處理

不管是任何的文字校正系統，前處理都是相當重要的一個步驟。經由數位器材所拍攝的影像，並不能立刻被我們使用；數位影像必須經過幾個處理步驟，才能將有用的資訊抽取出來，這些步驟簡稱前處理。在本論文中，數位影像可能是由數位相機或是網路攝影器所拍攝的彩色文件影像；而有用的資訊，則泛指影像內的所有中文與英文文字。

前處理的第一步即是把輸入的彩色影像轉成灰階影像，再進一步轉換成二值化影像。之後經由連通元件偵測取得並記錄二值化影像中所有連通物件的資訊：包括每個連通元件在原始影像中的位置、連通物件的寬、高...等；並在雜訊去除與圖文分離階段中設定過濾機制，刪除一些影像中圖片、面積過大不可能是文字、或是面積過小無法辨識的連通物件。此階段完成後，留下來的連通物件都將視為文字。各步驟詳細說明於如下小節。以下是前處理流程圖：

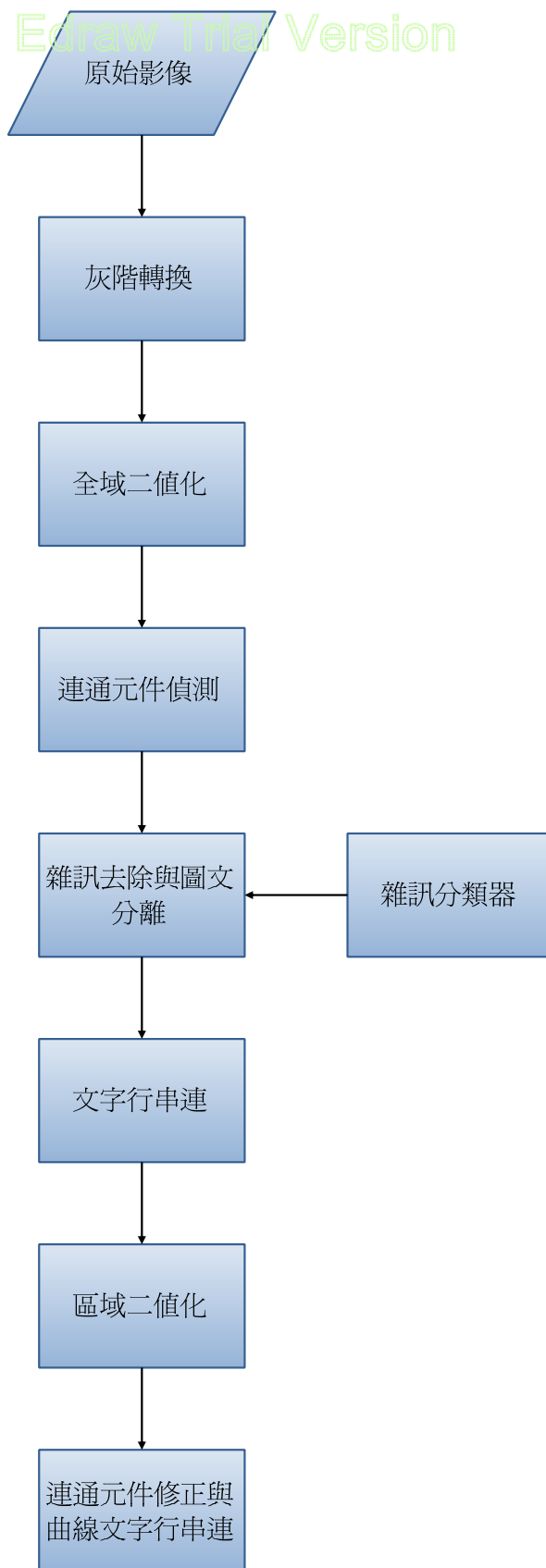


圖 2-1 前處理流程圖

2.1 彩色影像轉換灰階影像

在本研究中所取得的數位影像為 RGB 色彩空間，一般數位相機或網路攝影機在擷取影像之後通常都將其儲存成 RGB(紅、綠、藍)色彩空間格式，但 RGB 色彩空間具有以下缺點：

一、屬於硬體導向的色彩空間。

二、使用 R、G、B 三原色合成出來的影像飽和度偏低，色調變化亦不大。

三、人眼無法直接感受到 R、G、B 三原色的比例。

舉例來說，人們會以暗紅色或亮紅色來形容一個物體，而不會用其在 R、G、B 三原色上的比例來形容的該物體的顏色。圖 2-2 為 RGB 色彩空間圖。

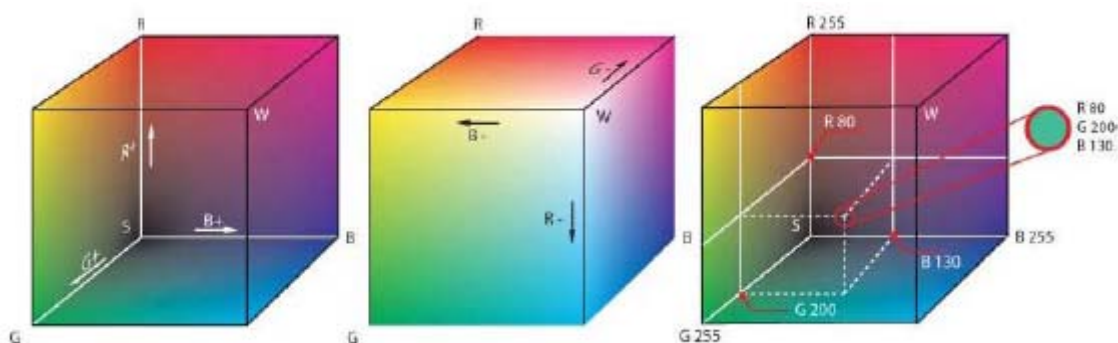


圖 2-2 RGB 色彩空間圖

因此在本研究中，前處理一開始即將 RGB 色彩空間轉換成灰階影像，轉換公式如下所示：

$$Gray = \frac{1}{3}(R + G + B) \quad (\text{公式 2.1})$$

灰階影像代表了該影像的亮度、色彩的強度，能夠呈現物體表面相對的明暗特性，符合本研究所需的特性，而且將 RGB 色彩空間轉換成灰階影像還能減少影像資料量，加快程式處理速度。

2.2 全域二值化

想要得到文字區塊，先知道前景是必要的。在本研究中，前景也就是圓柱影像中的文字範圍所在。為了得到前景，必須針對所欲分析的資料進行前景與背景的分割，而在影像處理上，二值化則是最常使用的方法。所謂的二值化即是將欲分析的彩色影像或灰階影像，轉換成只由黑及白所組成的黑白影像，將影像轉換成二值化影像具有以下兩個優點：

1. 減少所需處理的影像資料量。
2. 加快後續處理影像的速度。

一般而言，所謂的二值化即是由灰階影像中定義一個臨界閾值，當影像灰階值大於此臨界閾值時，則將該影像像素值設為 255，即將影像像素設為白色；反之則設為 0，也就是將影像像素設為黑色。如下圖所示：

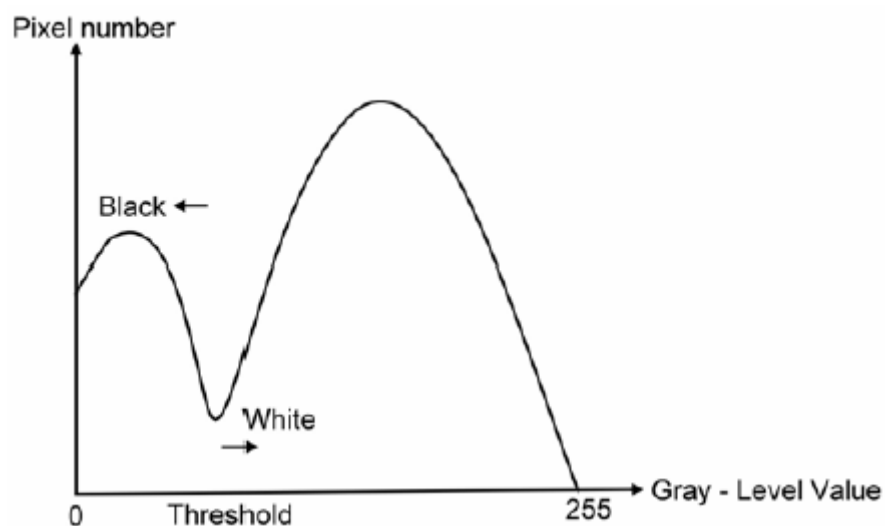


圖 2-3 灰階值統計直方圖

本論文採用 Otsu[12]所提出的二值化方法，能自動由影像中選取

一個最佳的閾值將灰階影像中所有像素點分成兩群：前景資訊與背景資訊，同時使得群內變異最小而群間變異達到最大，完成整張影像的二值化。Otsu 的二值化演算法概略介紹如下：

若灰階影像其像素值分佈為 $[1, 2, 3, \dots, L]$ ，可以計算出不同灰階值其分佈機率值 p_i 。而下式也就成立：

$$p_i = \frac{n_i}{N} \quad (\text{公式 2.2})$$

其中 n_i 代表灰階值為 i 的個數， $N = n_1 + n_2 + \dots + n_L$ 是不同灰階值個數的總和。其中任何一個 p_i 皆大於零($p_i \geq 0$)、且所有的 p_i 加總和為一($\sum_{i=1}^L p_i = 1$)。

二值化的目的在於選擇一個臨界值 T ，以將灰階影像像素分成 C_0 與 C_1 兩個群集，在本研究中個別代表了圓柱上的文字與背景。 C_0 的灰階值分佈為 $[1, 2, 3, \dots, T]$ ，而 C_1 的灰階值分佈範圍則是 $[T+1, T+2, T+3, \dots, L]$ 。並可藉此計算出各群集的機率分佈 w_0 與 w_1 。公式如下：

$$w_0 = \sum_{i=1}^T p_i \quad (\text{公式 2.3})$$

$$w_1 = \sum_{i=T+1}^L p_i \quad (\text{公式 2.4})$$

同時可計算出各群集平均值 m_0 、 m_1 。計算公式如下：

$$m_0 = \sum_{i=1}^T i * p_i \quad (\text{公式 2.5})$$

$$m_1 = \sum_{i=T+1}^L i * p_i \quad (\text{公式 2.6})$$

利用上述公式可計算出各群集變異數 σ_0 、 σ_1 如下式：

$$\sigma_0^2 = \sum_{i=1}^T (i - m_0)^2 * p_i \quad (\text{公式 2.7})$$

$$\sigma_1^2 = \sum_{i=T+1}^L (i - m_1)^2 * p_i \quad (\text{公式 2.8})$$

所以各群集變異數加權總和 $\sigma_w^2(T)$ 為：

$$\sigma_w^2(T) = w_0 * \sigma_0^2(T) + w_1 * \sigma_1^2(T) \quad (\text{公式 2.9})$$

Otsu 演算法的目的在找出一個 T 值，使得變異數加權總和 $\sigma_w^2(T)$ 為最小。能達到此條件的 T 即為二值化的臨界閾值。

2.3 連通元件偵測

經過上述二值化之後，我們可以把影像中文字與背景區域做出區分，其中當然還包含了一些雜訊，對於留下來的前景部分，我們使用連通元件進行前景物的標記，藉以尋找出文字區塊以提供後續處理之用。

該標記方法是利用二值化之後的影像，由上而下、由左至右找尋影像中每一個屬於同一個獨立區塊的部分，再將該區塊標記成同一個標號。連通元件的分析法一般分為兩種，即四連通 (4-ways connected-components) 和八連通 (8-ways connected-components) 兩種。其各自的檢查方向表示如圖 2-4 所示。

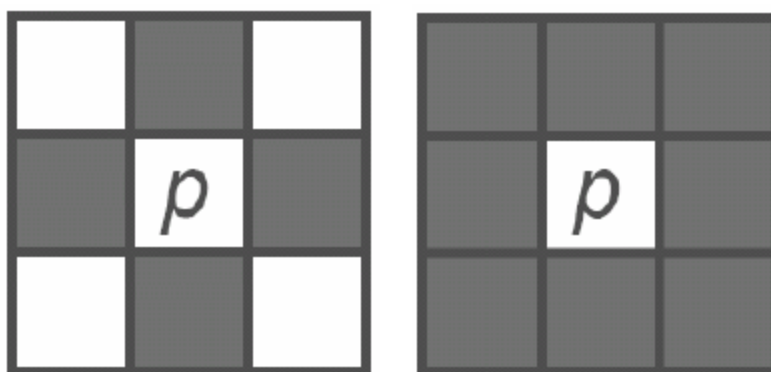


圖 2-4 四連通、八連通示意圖

由圖 2-4 可以得知八連通比四連通多檢查了四個方向，即左上、右上、左下、右下，本研究中採用八連通作為標記各自獨立區塊的方法，現將該演算法說明如下：

8-ways connected-components 演算法

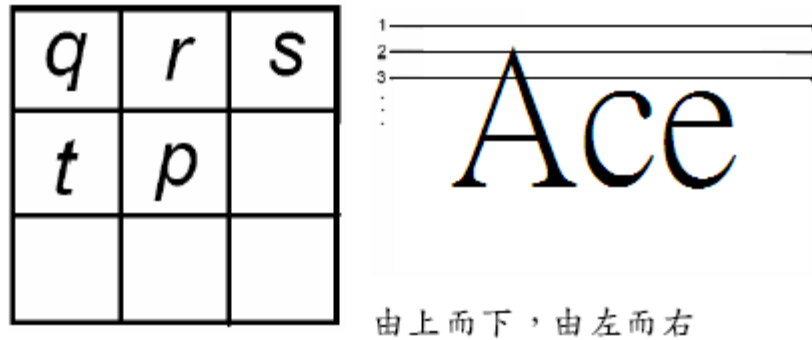


圖 2-5 八連通演算法示意圖

階段一：從左至右、上至下掃描影像中每個像素點。

1. 若 p 為 255，掃描下一個像素點。
2. 若 p 為 0，則執行步驟 3、4、5。
3. 若 $\{q, r, s, t\}$ 為 255，則設給 p 一個新的標籤。
4. 若 $\{q, r, s, t\}$ 其中一個像素點為 0，則設 p 的標籤和 x 的標籤一致，其中 x 為 0 且 $x \in \{q, r, s, t\}$ 。
5. 若 $\{q, r, s, t\}$ 有兩個以上為 0，設 p 的標籤和 x 的標籤一致，其中 x 為 0 且 $x \in \{q, r, s, t\}$ ，並且標註 y 的標籤和 x 的標籤一致，且針對所有 y 為 0 且 $y \in \{q, r, s, t\}$ 。

階段二：合併所有具有相等標籤的區塊。

在本研究中除了利用八連通標記各個獨立區塊外，同時以此繪製出該獨立區塊的邊界。由灰階處理一直到連通元件偵測的結果如圖 2-6 所示。



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)

圖 2-6 連通元件偵測結果

(a) (b) 原始影像

(c) (d)灰階影像

(e)(f) 二值化影像

(g)(h)連通元件偵測影像

2.4 雜訊去除與圖文分離

經過連通物件分析後，我們就可以取得二值化影像中所有前景物的資訊，但是並非所有前景物都是我們有興趣的。影像中前景物除了包含我們需要的中、英文文字外，也可能包含了一些圖片、雜點、甚至是因為陰影導致二值化後產生的大片黑色區塊。為了不讓雜訊影響後續的處理，我們用一個雜訊去除的分類器來達到淨化影像的目的。

雜訊分類器

由觀察可知，大多數雜訊都具有黑色像素點數量遠大於白色像素點的特性，故論文中使用連通物件原始影像中各像素的二值化資訊做特徵，訓練了一個多層感知機的類神經網路做為雜訊分類器。該分類器的輸入層維度為 30×30 ，即所有連通物件影像需正規化至 30×30 的大小，同時包含一階隱藏層與輸出層。若分類器把輸入影像視為雜訊，則系統將會把該塊區完全忽略，因此我們可以視雜訊去除後剩下的連通物件全部為文字連通物件，以進行後續處理。

2.5 文字行串連

經過上述步驟之後，我們可以得到一張去除雜訊的文字影像。不過由於文字在整張圓柱影像中所佔比例通常並不大，單純以全域二值化進行影像的二值化會容易出現二值化閾值上的誤差，造成文字容易出現雜訊。為了減少這種情形並且提高文字區塊的特徵，在本研究中使用文字行串連，以求在接下來的步驟中針對此串連區域進行區域二

值化。

文字行串連的目的如下：

- 一、 可將位於同一行之文字區塊串連在一起，以利後續辨識輸出之用。
- 二、 以行為基準，進行 2.6 節的區域二值化。

文字行串連的方法如圖 2-7 所示。

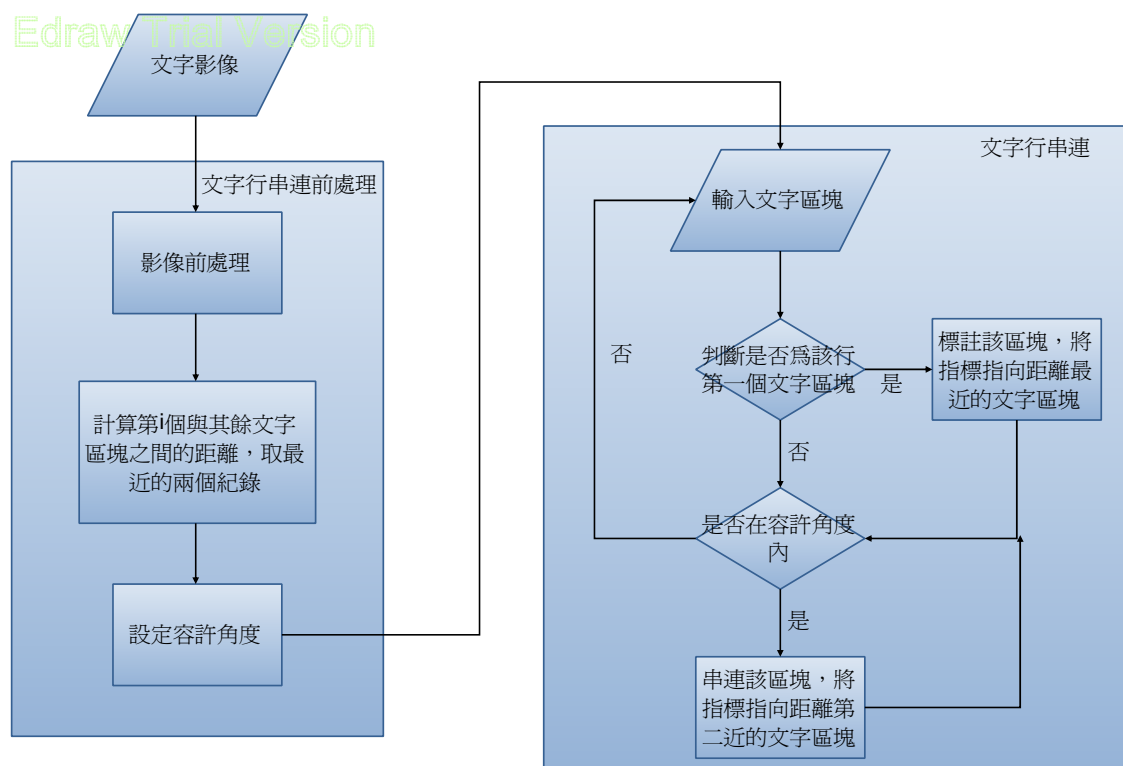


圖 2-7 文字行串連流程圖

經由文字區塊標記，再經由圖 2-7 文字行串連流程圖的步驟後，我們就可得到文字行的範圍。首先是文字行串連的前處理。我們計算，並記錄每個文字區塊與最近的兩個文字區塊距離。其次設定在水平串連文字行時文字區塊間可容許的角度。由於文字影像中，每一行文字的各字母間幾乎不可能出現過大夾角，因此本研究中此容許角度固

定設為 30° 。

再來開始進行文字行串連。在串連文字過程中，只判斷距離該區塊最近的兩個文字區塊，利用容許角度資訊判斷文字區塊間的角度差異是否位在正負 30° 之內，若有，在此視為可以串連，反之則視為是另一條文字行的開頭。圖 2-8 即為做完文字區塊串連的結果，藍色連線所橫跨過的文字區塊，則代表屬於同一行文字，在本研究中稱此藍色連線為文字行，此連線結果將有助於下節所要介紹說明的區域二值化。

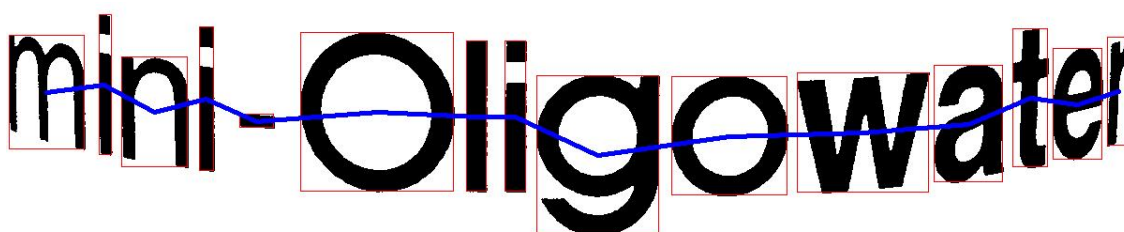


圖 2-8 文字區塊串連結果

2.6 區域二值化

在二值化演算法之中，二值化範圍內所包含的影像複雜度與二值化效果有著非常直接的關係。這通常也代表著，越能夠適當的縮小二值化範圍，二值化的效果就越好。

為了提升二值化品質，我們使用 2.5 節找出的文字行區域做為二值化的範圍，以此二值化結果來取代原來以整張影像為單位進行的二值化影像。藉此減少因相機取像時光影不均所引起的雜訊干擾，並強化文字邊緣特徵以利後續文字辨識的進行。圖 2-9 為區域二值化的流程圖。

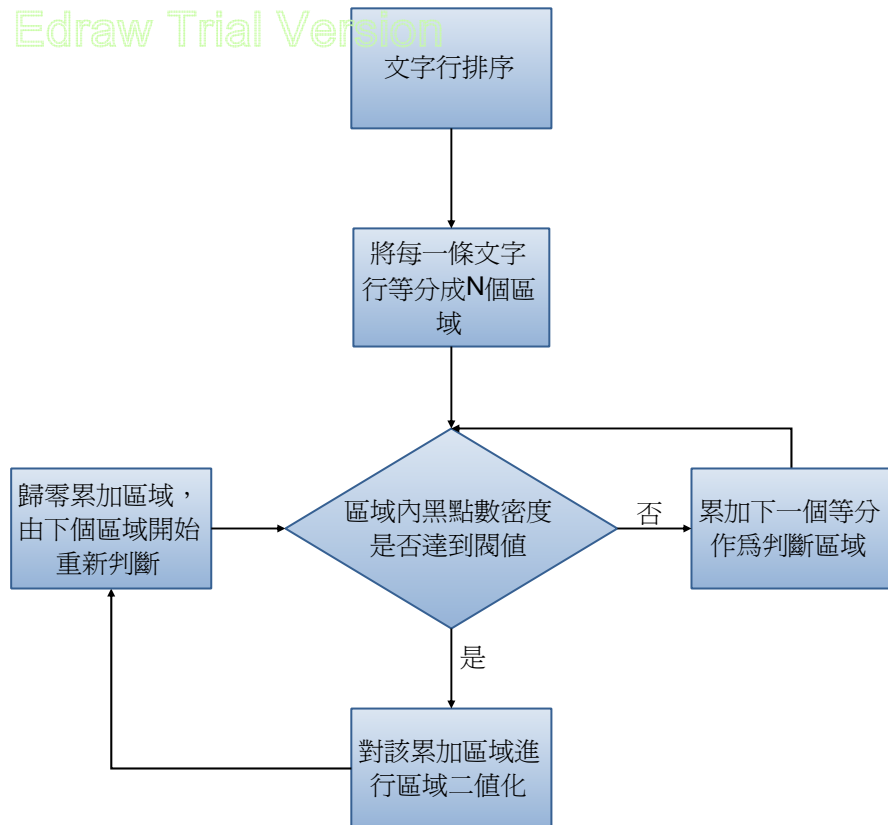


圖 2-9 區域二值化流程圖

我們將要做區域二值化的文字行區分為 N 等份，計算每一等份內的黑點密度，若達到閾值則對這一等份的區域進行 Otsu 演算法，若沒達到則累加下個區域直到超越閾值，便以此累加區域範圍進行 Otsu 演算法直到到達文字行尾端。

利用黑點數密度(pixel density)為判斷依據，是因為當密度過低時，經由 Otsu 演算法後反而會發生字元破碎的情形。因此若密度不足閾值時，該區域並不適合再次進行二值化。圖 2-10 是區域二值化的結果。



圖 2-10 區域二值化結果圖

第三章

曲線文字行繪製

為了正確將圓柱中因文字貼於圓柱平面上而造成的文字扭曲進行校正，在前處理完畢並繪製出各個連通元件的矩形區塊後，我們需要將文字部分加以連線以建立文字行。但在圓柱影像中，由於文字分布已經不再如平面文字影像中呈現直線分佈，導致文字行已經無法由計算影像中連通元件最短連線之直線方程式進行描述，如圖 3-1。因此我們將這些文字區塊連接成曲線文字行，以在後續步驟中將曲線文字行校正為直線。

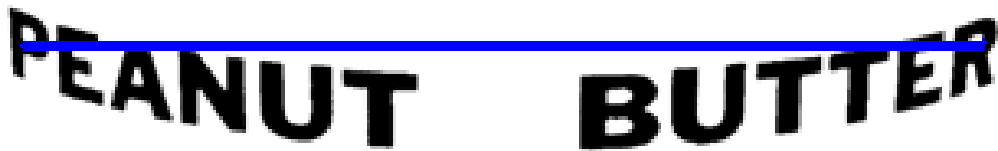


圖 3-1 以直線逼近圓柱文字的文字行

為了從影像中連接出曲線文字行，我們首先將原來以矩形標記的連通元件進行修改，依照連通元件內部的文字外型分類，各別以左斜向平行四邊形、右斜向平行四邊形、以及矩形進行第二次的連通元件標記。接下來以這些平行四邊形連通元件的底邊連線繪出一道弧形，再以拋物線方程式進行此弧線的方程式逼近。

以下將針對各步驟進行詳細介紹：

3.1 連通元件修正

在文字區塊的標記上，連通元件是常常被使用的技術，能夠簡單而確實的標記出目標區域。但由於所標記的目標物件形狀往往是不規則圖形，所以在運用時多半會使用一個固定形狀，對目標物件進行包覆，以做為該連通物件的外形標記。尤其在進行文字影像的連通元件標記時，由於文字連通元件有著大小穩定與排列整齊的特點，是以多半使用矩形外框進行文字影像的連通元件標記，如圖 3-2(a)。然而當文字位於三維援助平面時，由於圓柱造成影像文字與文字排列皆出現變形，導致矩形連通元件儘管能夠框出文字範圍，卻未能夠保有足夠的資訊以判斷文字特徵，如圖 3-2(b)。



(a)



(b)

圖 3-2 以矩形進行連通元件標記

(a) 水平文字影像 (b) 圓柱文字影像

在本論文中，圓柱影像的文字連通元件中，一個連通元件一般就代表一個字母。是以本論文從連通元件中追求的主要資訊為該文字的位置與上下左右四方向的邊界資訊。當影像為一般平面文字影像時，矩形連通元件能夠很直接的切齊文字邊緣，使得連通元件的邊界資訊能夠簡單而正確的被解讀出來。但是若在圓柱影像中使用矩形連通元件，則會因為視角不同造成文字變形導致連通元件的上下左右四邊沒有辦法貼齊於真正變型前的文字邊緣，也直接影像了後續邊界資訊的

取得。為使後續校正獲得良好的文字變形資訊以做改善，本論文更換使用單一矩形連通元件的作法，加入平行四邊形做為圓柱影像中繪製連通元件的基礎圖形。

3.1.1 文字分類

觀察文字影像在三維圓柱平面上不同位置時的變形，我們可以針對英文大小寫與數字傾斜文字進行分類，並以此分類使用不同方法將文字連通元件由矩形調整為平行四邊形。分類依據為文字在矩形連通元件中左右兩側最高點與最低點之位置，如圖 3-3 的 A、B、C、D。分別為由矩形左側向右找出的最高點、最低點，以及由矩形右側向左找出的最高點與最低點。

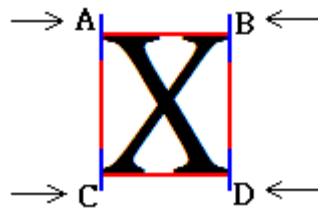


圖 3-3 文字分類依據示意圖

我們利用 A、B、C、D 的位置，將英文與數字連通元件分為四種類型。

分類一：此類型的 A、B、C、D 四個端點皆分佈於連通元件中央。此類連通元件通常有著圓弧形的上下底邊，連通元件的整體圖形接近圓形。以文字 e、O 為其典型，如圖 3-4。

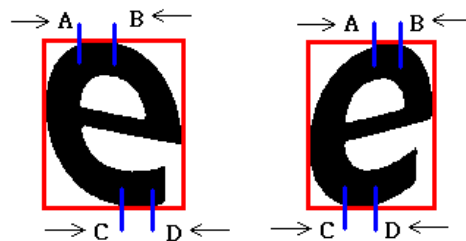


圖 3-4 類型一連通元件

分類二：此類型的 A、B 頂點分佈於連通元件中央，C、D 頂點則分佈於連通元件兩側。整體圖形接近正向三角形。以文字 A 為其典型，如圖 3-5。

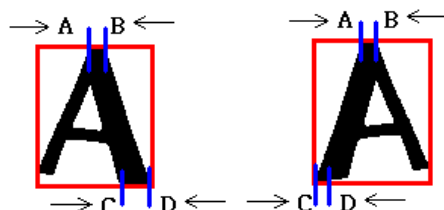


圖 3-5 類型二連通元件

分類三：此類型的 A、B 頂點分布於連通元件兩側，C、D 頂點則分佈於連通元件中央，與類型二上下顛倒。整體圖形接近倒三角形，典型文字則為 Y，如圖 3-6。

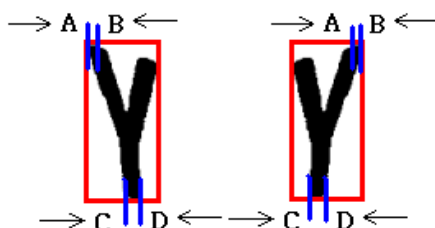


圖 3-6 類型三連通元件

分類四：我們將上述以外的文字都分到此類型。此類型的四個頂點中會至少有 A、C 或 B、D 的一組頂點分佈於連通元件兩側。這些連通元件所接近的圖形則有兩種，分別為矩形與直角三角形。以文字 E 與 L 為其典型，如圖 3-7。

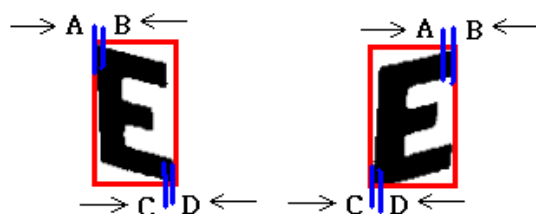


圖 3-7 類型四連通元件

我們以此分類將受到圓柱影響而傾斜的文字分類成四個類別，如表 3-1、表 3-2、表 3-3。

表 3-1 數字連通元件分類

分類	文字
分類一	3、4、6、8、9、0
分類二	1、2
分類三	5、7
分類四	無

表 3-2 大小英文字母連通元件分類

分類	文字
分類一	C、G、O、S
分類二	A、Q
分類三	J、T、U、V、W、Y
分類四	B、D、E、F、H、I、K、L、M、N、P、R、X、Z

表 3-3 小寫英文字母連通元件分類

分類	文字
分類一	c、e、o、s
分類二	a、t
分類三	b、g、v、w
分類四	d、f、h、i、j、k、l、m、n、p、q、r、u、x、y、z

分類一由於其上下邊皆為弧形，故難以直接由現有的邊界資訊直接判斷該文字是否有產生傾斜。

分類二的文字由於其外型，當受到圓柱影像使文字傾斜時將導致原來切齊於下邊的點產生偏移，變成文字只有在左右其中一側才貼齊邊界。是以我們可由下邊的變化判斷該文字是否有產生傾斜。

分類三與分類二上下顛倒，當圓柱影像使文字傾斜將導致原來切齊於上邊的點產生偏移，變成文字只有在左右其中一側才貼齊邊界。是以我們可由上邊的變化判斷該文字是否有產生傾斜。

被分類到分類四的文字於上下邊皆有點接近兩側邊緣，可以同時從上下兩邊進行傾斜判斷，在各自進行判斷後再選擇合適者。藉由上述四種分類，我們得以在後續步驟進行平行四邊形連通元件的繪製。

因為分類一為延後處理，而分類四連通元件的處理法有包含了分類二與分類三的處理方法在內，故當面對雜訊造成分類錯誤時，原屬於分類二、分類三、分類四的連通原件並不會因被分類到分類一或分類四而產生影響。也由於分類二與分類三在外型特徵上的差異，此兩種分類並不會出現分到對方類別的情形。

由此可知在面對雜訊影響時，本論文的連通元件分類也能達到一定程度的抗雜訊效果。

3.1.2 連通元件繪製

平行四邊形連通元件是由原來的矩形連通元件變化而來。矩形連通元件的左右兩邊維持為平行四邊形連通元件的左右兩邊，藉由修改上下兩邊以完成平行四邊形連通元件。繪製平行四邊形連通元件的演算法如下：

Step1：由前處理取得矩形連通元件。

Step2：取得上邊最高點與下邊最低點的分佈資訊，並以此判斷該

連通元件的分類。

Step3：如圖 3-8，依照 3.1.1 的四種不同分類，個別以不同方式尋找其平行四邊形連通元件的上邊或下邊。

分類一：沿用原來的矩形連通元件做為平行四邊形連通元件。

分類二：由下邊頂點為中心向左右傾斜，尋找出能貼齊文字，讓左右兩邊頂點都在連通元件兩側的最小角度。

分類三：由上邊頂點為中心向左右傾斜，尋找出能貼齊文字，讓左右兩邊頂點都在連通元件兩側的最小角度。

分類四：由上下兩邊各別尋找能貼齊文字，讓左右兩邊頂點都在連通元件兩側的最小角度。再由兩者中取角度較小者。

Step4：針對步驟三找出上邊或下邊進行平行推移，找出對稱的另一對邊。

Step5：串連四個邊，完成平行四邊形連通元件。



(a)



(b)



(c)



(d)

圖 3-8：四種文字類型的平行四邊形連通元件取得法

(a)分類一(b)分類二(c)分類三(d)分類四

3.2 連結曲線文字行

在 3.1 節中，已抽取出各個文字的平行四邊形連通元件。但是光使用這些資訊還不足以將整行曲線文字校正回水平直線，想要達到這結果，還需要關於該曲線的更多資訊，以將該文字曲線的方程式計算出來。

為了完整表現文字在圓柱上的曲線變化，我們使用統計學中的回歸分析來進行圓柱影像的曲線方程式逼近。回歸分析主要用在尋找兩個或多個變數之間是否存在相關、以及其相關方向與強度，並建立數學模型以便觀察其相關變化。

3.2.1 橢圓逼近與拋物線逼近

對於圓柱影像中的曲線文字行，由於到圓柱的切面為橢圓形狀，文字在圓柱上的變化也會傾向橢圓。考慮到此，以橢圓方程式做為曲線擬合似乎是較為合理的。但以橢圓方程式逼近曲線文字行卻也有著下列問題：

- 一、 在回歸分析中，斜橢圓逼近屬於二元多項式線性回歸，比起一般多項式回歸更為複雜，且需要更多處理時間。
- 二、 圓柱影像中的曲線文字行最多只能達到半個橢圓，並沒有一個完整的橢圓輪廓。這使得影像在橢圓逼近時可能會有多個大小不同的橢圓都能達到逼近條件，而橢圓大小不同也就代表了這些滿足逼近條件的橢圓在曲率、圓心、周長等特徵上都有所不同，導致難以從中取捨出唯一解。
- 三、 橢圓方程式逼近需要較完整的參考點才能在逼近上獲得好的結果。

相對的若使用拋物線進行曲線文字行逼近，除了由於拋物線屬於簡單的二次方程式，使得在執行上較為簡潔快速，也較為容易外；以拋物線進行逼近在結果上也較為穩定，而拋物線方程式為 $y = a_0 + a_1x + a_2x^2$ ，所需要計算的參數也較少，更方便應用。

因此，我們在此決定使用拋物線針對圓柱影像文字行進行方程式逼近。

3.2.2 拋物線回歸分析

我們選用拋物線函數進行方程式逼近曲線文字行。考慮到參考點的選擇與逼近結果有著最直接的關係，我們必須盡可能選到符合目標拋物線的有效參考點。

Typoline[5]由 Abdelwahab Zramdini 於 1998 年提出，是一種主要針對英文與數字的文字分佈資訊。在非刻意排列的情形下，水平排列的英文與數字都會依照五條 Typoline 進行排列。五條 Typoline 由上至下分別是 topline、upperline、centerline、baseline 與 bottomline，其中 centerline 並非任何英文與數字的分佈線，而是屬於連接線等標點符號的分佈線。如圖 3-4。

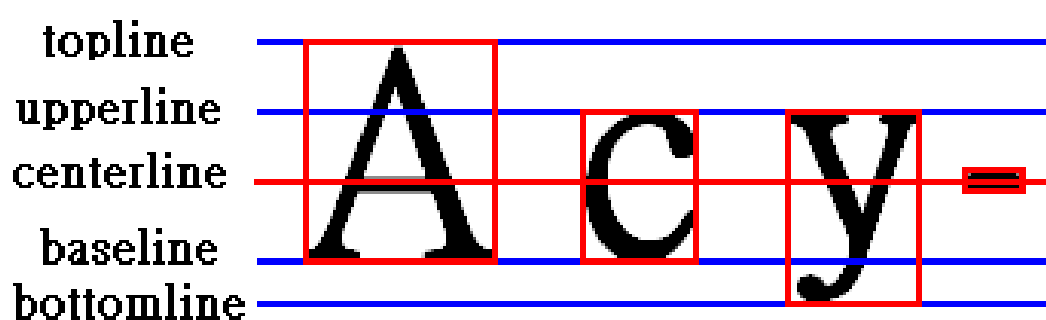


圖 3-9 Typoline 與連通元件分佈示意圖

經由連通元件於 Typoline 上的分佈，可將文字行內的區域分成三個。分別為 Upper Zone、Center Zone 以及 Lower Zone，如圖 3-9。在本研究中依據文字區塊所橫跨過的區域不同，可將文字區塊區分為 7

種 Type，如圖 3-10。分別為：

Type 0: Full Type，橫跨 Upper Zone、Center Zone 及 Lower Zone 三區。

Type 1：High Type，橫跨 Upper Zone、Center Zone 兩區。

Type 2：Deep Type，橫跨 Center Zone、Lower Zone 兩區。

Type 3：Short Type，只在 Center Zone 區。

Type 3：Super Type，位於 Upper Zone 或橫跨 Upper Zone、Center Zone 兩區且其下緣較靠近 upperline。

Type 5：SupScript Type，位於 Lower Zone 或橫跨 Center Zone、Lower Zone 兩區且其上緣較靠近 baseline。

Type 6：Center Type，位於 Center Zone 區且上下緣都較靠近 centerline。

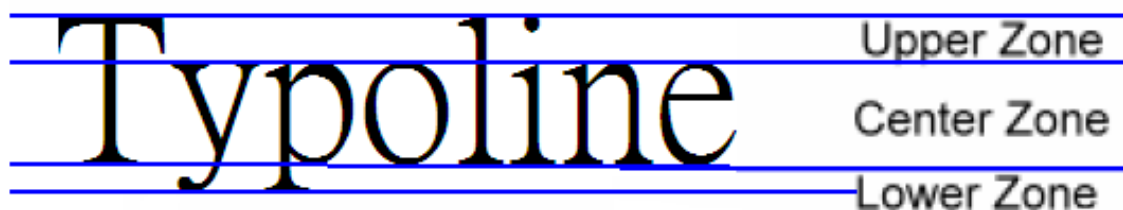


圖 3-10 Typoline 示意圖 1



圖 3-11 Typoline 示意圖 2

依照 Typoline 資訊，我們將英文文字中大小寫與數字連通元件的端點的分佈列於表 3-4 與 3-5。由表中可見到，英文與數字連通元件上

方端點分佈於 topline 與 upperline 的數量分別為 42 與 20，比例約是二比一；然而下方端點於 baseline 與 bottomline 的分佈數量則是 57 與 5，比例將近十一比一。此數據代表在一般文字行中，以文字下緣進行方程式逼近較為穩定。是以我們會個別對連通元件的上緣與下緣進行逼近，再以回歸分析中可信度較高者為該文字行的方程式。

在系統執行中，針對文字行中的每一個連通元件，我們依照 3.1.1 中的文字分類擷取不同的點進行逼近。屬於類型一的文字採用下緣中央的點，其餘分類則採取左下、右下兩點進行拋物線方程式逼近。

表 3-4 連通元件上方端點分佈表

分佈線	文字	數量
topline	ABCDEFGHIJKLMNOPQRSTUVWXYZ bdfhkl 0123456789	42
upperline	Acegijmnopqrstuvwxyz	20
baseline	無	0
bottomline	無	0

表 3-5 連通元件下方端點分佈表

分佈線	文字	數量
topline	無	0
upperline	無	0
baseline	ABCDEFGHIJKLMNOPQRSTUVWXYZ abcdefghijklmnopqrstuvwxyz 0123456789	57
bottomline	Gjppqy	5

由於選用二次拋物線，所以我們假設所需的數學模型為：

$$y = f(x_i; a_0, a_1, a_2) = a_0 + a_1 x_i + a_2 x_i^2 \quad (\text{公式 3.1})$$

其中 y 為此模型的應變數， x 為此模型的自變數， a_0 、 a_1 及 a_2 則為此模型的參數。平方誤差為 $[y_i - f(x_i)]^2$ ，而總平方誤差為下式：

$$E = \sum_i [y_i - f(x_i)]^2 = \sum_i [y_i - (a_0 + a_1 x_i + a_2 x_i^2)]^2 \quad (\text{公式 3.2})$$

藉由求出最小誤差，我們可找出相對的逼近方程式。從另一個方面來看，假設參與逼近的 m 個點都通過所逼近的拋物線，則下列式子可成立：

$$a_0 + a_1 x_1 + a_2 x_1^2 = y_1$$

$$a_0 + a_1 x_2 + a_2 x_2^2 = y_2$$

⋮

$$a_0 + a_1 x_m + a_2 x_m^2 = y_m \quad (\text{公式 3.3})$$

亦可寫成：

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$X \qquad A \qquad Y$

(公式 3.4)

其中 X 、 Y 為已知， A 為未知向量。我們能夠找到一組 θ 使等號兩邊差異為最小，此差異也就是前述的總平方誤差 E ，而完成拋物線逼近方程式。

首先我們對總平方誤差 E 做偏微分，可獲得標準方程式組如下：

$$\begin{bmatrix} \sum_{i=1}^n y_i = na_0 + a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i y_i = a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 y_i = a_0 \sum_{i=1}^n x_i^2 + a_1 \sum_{i=1}^n x_i^3 + a_2 \sum_{i=1}^n x_i^4 \end{bmatrix} \quad (\text{公式 3.5})$$

通過方程式求解，可得到 a_0 、 a_1 、 a_2 的值，然後開始進行相關

性檢驗。由於拋物線回歸屬於二元線性回歸，我們利用可決係數 R^2 進行擬合度檢驗，公式如下。 R^2 介於 0 到 1 之間，該系數越接近於 1，回歸模型與輸入資料的擬合度就越高。一般統計中， $R^2 < 0.3$ 被稱作擬合度低， $R^2 > 0.8$ 則被稱作擬合度高。

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{公式 3.6})$$

第四章

曲線文字校正

在第三章分析出圓柱文字影像中的曲線文字行，並逼近出方程式後，接下來所要做的就是將這條曲線文字行校正為水平直線。文字由一般平面被貼上圓柱後所產生的變化主要分為兩個，其一是文字扭曲；其二則是文字寬度與間距的縮短。

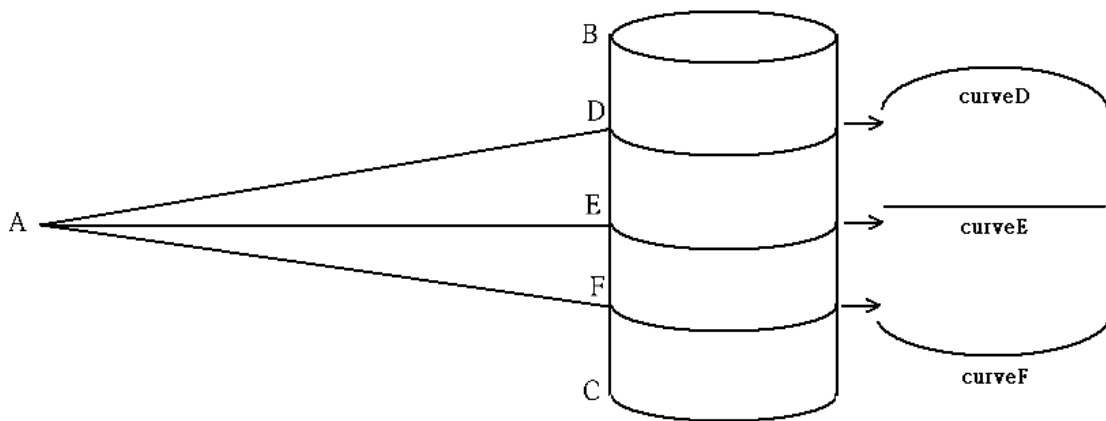


圖 4-1 圓柱曲面取像側面圖

文字扭曲主要來自相機對圓柱上文字的上下斜角。以圖 4-5 為說明，其中 A 為相機取像點，B 與 C 則是圓柱的頂端與底端，D、E、F 為假設的圓柱上文字線。若相機與文字水平，如 A 與 E 的連線，取像出來的文字行也會呈現水平。若相機仰視文字，如 A 與 D 連線，取像出來的文字則會呈現一開口向下的弧形。若相機俯視文字，如 A 與 F 連線，取像出來的文字則會呈現一開口向上的弧形。

我們將針對此弧形變化進行校正。首先是文字校正後的分佈位置，在這裡我們使用先前逼近出來的拋物線文字行，以擷取 Typoline

資訊來取得；另一部分是文字的傾斜校正，我們則會靠著將取得的曲線 Typoline 區域拉直為水平直線來完成。

圓柱影像另一個帶來的影響是文字寬度與間距的縮短，如圖 4-2 所示。

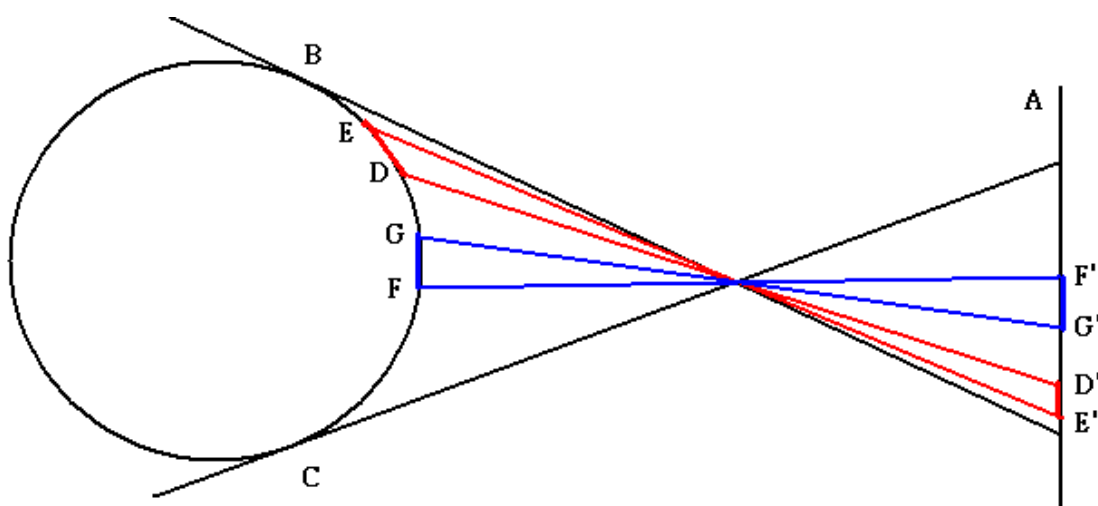


圖 4-2 橢圓文字影像寬度縮短示意圖

上圖中，B、C 為相機所取得的圓柱邊界，D、E 與 F、G 是文字於圓柱上的位置，D'、E'與 F'、G'則是圓柱經過相機取像投影於平面 A 上的位置。長度 \overline{DE} 原來與 \overline{FG} 相同，但由於投影距離的不同，造成投影過後的 $\overline{D'E'}$ 比起 $\overline{F'G'}$ 有著一定比例的縮短。影響比例在越靠近圓柱邊緣的文字時顯得越大。

在完成文字扭曲校正，讓文字校正為一條水平直線後，我們將文字行垂直分為多個區塊，個別依照計算出來的比例進行延長，以達到校正目的。

4.1 曲線 Typoline 繪製

我們已經於前一章節取得了曲線文字行的拋物線方程式。藉由此資訊，並配合群聚分析演算法，我們可以取得曲線 Typoline 的資訊。

4.1.1 K-means 演算法介紹

K-Means[13]為非監督式(unsupervised)的群聚分析演算法。在輸入資料沒有預先的標籤資訊時，也能夠將輸入資料指定分為若干群，同時滿足每群資料與該群聚中心的距離總和越小越好。在本論文中，K-Means 演算法將在 Typoline 繪製時進行應用。

K-Means 演算法流程如下(假設將輸入資料分成 K 群)：

輸入： $\{x_1, x_2, \dots, x_n\}$ 、 K ：預設分群數。

輸出： K 個群集。

Step1：選擇 K 個初始群聚中心 $z_1(1), z_2(1), \dots, z_K(1)$ ，通常是由輸入資料中隨機挑選。其中 $z_i(t)$ 代表在第 t 次訓練時第 i 群的群聚中心。

Step2：在第 t 次訓練時，將輸入資料 $\{x\}$ 依照下式關係分配到 K 個群集中。 $x \in S_j(t)$ 如果 $\|x - z_j(t)\| < \|x - z_i(t)\|$ 對所有 $i=1, 2, \dots, K, i \neq j$ ，其中 $S_j(t)$ 代表以 $z_j(t)$ 為群聚中心的樣本集合。

Step3：由步驟 2 得到的結果計算新的群聚中心： $z_j(t+1), j=1, 2, \dots, K$ ，如下：

$$z_j(t+1) = \frac{1}{N_j} \sum_{x \in S_j(t)} x, j=1, 2, \dots, K$$

其中 N_j 為 $S_j(t)$ 集合內的樣本數。

Step4：如果 $z_j(t+1) = z_j(t), j=1, 2, \dots, K$ ，則演算法終收斂並且停止；否則回到步驟 2 繼續執行。

4.1.2 Typoline 演算法介紹

以下是 Typoline 的演算法：

Step1：將 3.2.2 逼近出的拋物線方程式參數固定，常數項設為變數，

讓拋物線方程式能在圖片上進行平移。對位於同一條文字行的文字區塊，以平行四邊形連通元件上緣座標為特徵，以 K-mean 演算法分成兩群， a_1 、 a_2 。

Step2：以位置較高的一群 a_1 做為拋物線方程式常數項，可得到 Typoline 中的 topline。

Step3：利用位置較低之一群 a_2 作為拋物線方程式常數項，可得出第二條 Typoline，即 upperline。

Step4：以平行四邊形連通元件下緣座標為特徵，以 K-mean 演算法分成兩群， a_3 、 a_4 。

Step5：以位置較高的一群 a_3 做為拋物線方程式常數項，可得到 Typoline 中的 baseline。

Step6：利用位置較低之一群 a_4 作為拋物線方程式常數項，可得出第四條 Typoline，即 bottomline。

Step7：利用 upperline 及 baseline 可以得出一條位於兩者之間的中線稱為 centerline。

Step8：判斷 topline 與 upperline，baseline 與 bottomline 是否過近，是則將該兩條 Typoline 給予合併。

完成後的曲線 Typoline 如下圖 4-4 所示。



圖 4.3 曲線 Typoline

4.1.3 以 Typoline 修正連通元件

平行四邊形連通元件已經於 3.1 節完成，但其中仍有如 O、e 等在表 3-1、3-2、3-3 中被分類為第一類型的文字，這些字儘管可能因圓柱影響而導致文字傾斜，但由於直接從文字影像特徵難以判斷，其連通元件仍然維持原來的矩形。

在取得 Typoline 特徵之後，我們可以針對這些連通元件進行修正。以連通元件左右兩邊進行延伸，找出其與四條 Typoline 的八個交點，各取與原來矩形連通元件頂點距離最短的交點，作為平行四邊形連通元件的頂點，以完成 3.1 節第一類型文字的平行四邊形連通元件。但由於我們的 Typoline 是以逼近方式產生，無法保證其是完全正確的對齊文字，是以計算產生的連通元件頂點比原來頂點要更小時，我們會配合將連通元件的上下邊修改，成為上下邊與 Typoline 平行，而能包覆整個文字範圍的平行四邊形連通元件。完成圖如下，圖中屬於第一類型字體的連通元件皆已修正為平行四邊形。



圖 4-4 Typoline 連通元件修正圖

4.2 曲線文字行校正

完成曲線 Typoline 分類後，將正式進行曲線文字行校正。由於已經計算出所有 Typoline 的拋物線方程式，我們在此直接進行將拋物線拉直的動作。演算法如下：

Step1：從該拋物線最左側開始向右取三點 $[x, x+1, x+2]$ 。

Step2：對此三點進行逼近。

Step3：若此三點呈直線，以直線進行逼近。

Step4：若此三點非為直線，以拋物線進行逼近。

Step5：依照逼近結果，將對應點的像素值進行輸出。

Step6：由 bottomline 開始直到 topline 為止，在每一行拋物線重複以上步驟。

曲線文字行校正結果如下圖：

mini-Oligowater

圖 4-5 曲線文字行校正結果

4.3 文字寬度校正

在完成曲線文字行校正之後，我們有了一條水平的文字行。是以接下來將針對圓柱造成的寬度縮減進行校正。若將上節校正過的影像垂直切為多段，我們需要求得的是各段影像該橫向延長的校正比例。

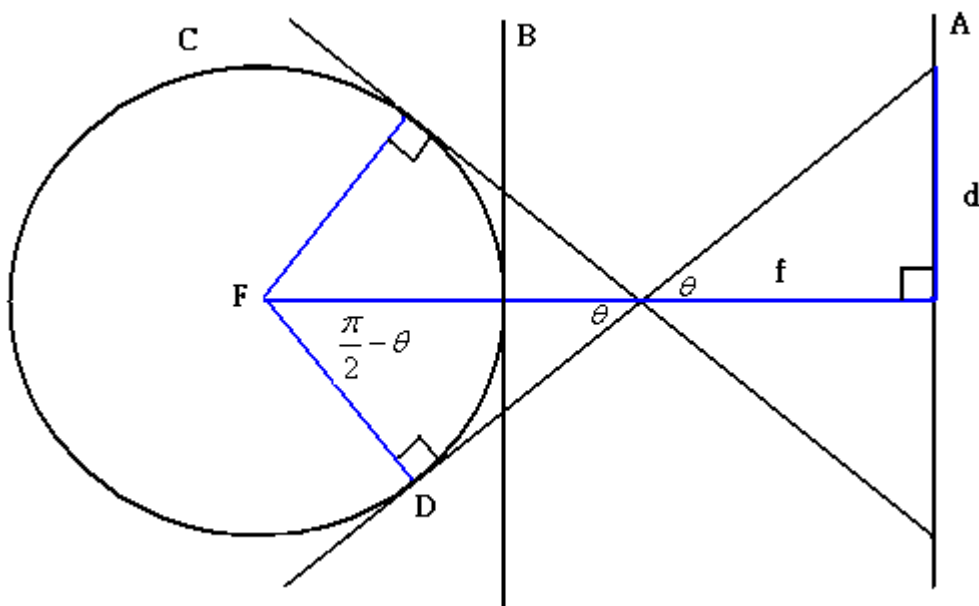


圖 4-6 寬度校正示意圖 1

上圖為相機對圓柱取像的俯瞰示意圖。假設取像時我們會將圓柱對齊畫面中央。由於取相時是對畫面中央的主要景物在取焦點，藉由相機取像平面 A 對圓柱 C 所取得的影像即為平面 B。平面 B 與平面 A 平行，並與圓柱 C 相切。平面 A 則以兩條切面與圓柱 C 相切，兩條切線在 A 上的距離即為影像中圓柱的寬度。

令取像平面中的 d 為影像中圓柱寬度的一半，取像平面至焦點的距離為 f ，夾角 θ 即為 $\tan(f/d)$ 。由於切線與圓柱呈九十度，是以圓柱圓心 F 對切點 D 與中心線的夾角為 $\pi/2 - \theta$ 。

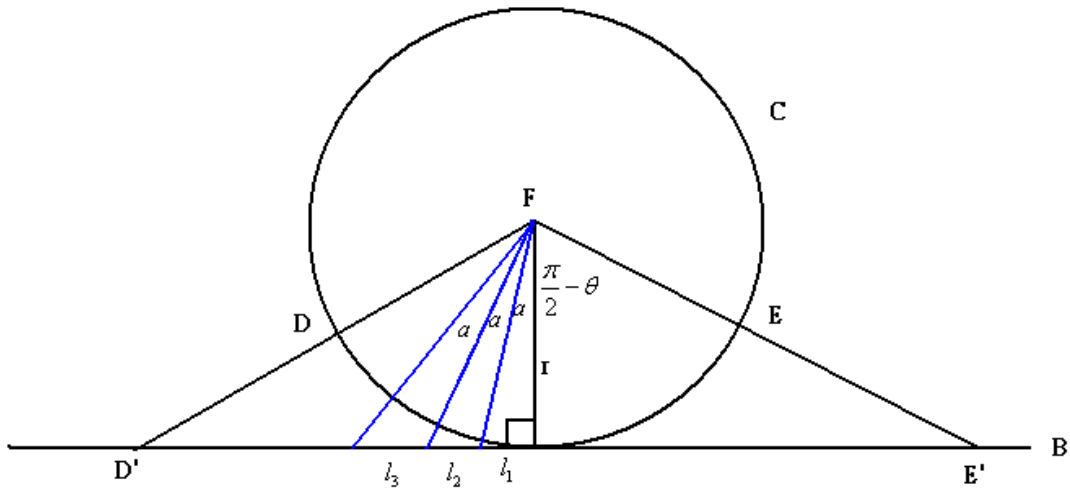


圖 4-7 寬度校正示意圖 2

我們將夾角 $\pi/2 - \theta$ 以角度 a 做 i 段的等分切割。由於圓柱半徑固定且平面 B 與圓柱 C 相切，可知 $l_1 = r \tan a$ ， $l_1 + l_2 = r \tan 2a$ 。是以 l_2 與 l_1 的比值計算如下：

$$\frac{l_2}{l_1} = \frac{l_1 + l_2 - l_1}{l_1} = \frac{r \tan 2a - r \tan a}{r \tan a} = \frac{\tan 2a - \tan a}{\tan a} \quad (\text{公式 4-1})$$

可依序推得 l_3 與 l_1 、 l_4 與 l_1 、直到 l_i 與 l_1 的比值如下：

$$\frac{l_3}{l_1} = \frac{l_1 + l_2 + l_3 - (l_1 + l_2)}{l_1} = \frac{r \tan 3a - r \tan 2a}{r \tan a} = \frac{\tan 3a - \tan 2a}{\tan a}$$

(公式 4-2)

$$\frac{l_4}{l_1} = \frac{l_1 + l_2 + l_3 + l_4 - (l_1 + l_2 + l_3)}{l_1} = \frac{r \tan 4a - r \tan 3a}{r \tan a} = \frac{\tan 4a - \tan 3a}{\tan a}$$

(公式 4-3)

$$\frac{l_i}{l_1} = \frac{\tan ia - \tan(i-1)a}{\tan a}$$

(公式 4-4)

公式 4-4 的比值即為圓柱上影像轉換至平面上所需的延長比例，也就是我們影像中文字區塊的校正比例。

我們令 $a = 0.0174533$ ，轉為角度後約為 1 度，也就是每 1 度進行一次寬度校正。我們將計算出來的校正比例列於下表：

表 4-1 寬度校正比例表

角度(度)	10	20	30	40	50
校正比例	1	1.06418	1.21014	1.48445	2
角度(度)	60	70	80	90	
校正比例	3.06418	5.75877	16.58172	211655296.44805	

在實作上，我們首先用將影像中的圓柱長度等比例進行分割，再分段進行不同比例的放大。校正結果如下圖：

mini-Oligowater

圖 4-8 圓柱影像寬度校正結果

第五章

實驗結果與討論

在本研究中針對曲線文字行的偵測與校正結果分別以實驗說明，並呈現本研究所提方法之完整執行結果。本研究使用數位相機，取得解析度為 2382x2128 的影像進行分析。共使用 60 張圓柱影像進行分析，總共有 769 個英文字元。使用的數位相機為 Fujifilm FinePix 700，其詳細規格請看表 5.1。

表 5.1Fujifilm FinePix 700 之規格

感光元件總像數	314 萬像素
感光元件有效像素	314 萬像素
感光元件尺寸	1/1.7 吋
變焦能力	3X
光圈範圍	F/2.8
最高解析度	2048x1536；2382x2128

5.1 曲線文字行偵測結果

下面將針對曲線文字行偵測結果做呈現。圖 5-1、圖 5-2、圖 5-3 為三張實驗原始影像以及對影像進行文字行偵測與對連通元件上緣參考點與下緣參考點分別進行方程式逼近後的結果。我們將所偵測的文字行以紅框畫出，並附上回歸分析的均方根誤差 (RMSE, Root Mean Square Error) 與 R 平方係數(R-square)。其中均方根誤差越低越代表逼近的正確性；R 平方係數則是越接近 1 代表資料與逼近方程式的相關性越高。



(a)



上緣逼近方程式： $y = 0.00005433x^2 - 0.09081x + 271.9$

下緣逼近方程式： $y = 0.0001717x^2 - 0.2367x + 162.9$

(b)

圖 5-1 六百毫升寶特瓶曲線文字行偵測結果

(a)原圖(b)方程式逼近結果



(a)



上緣逼近方程式： $y = 0.0002462x^2 - 0.3378x + 234$

下緣逼近方程式： $y = 0.0002899x^2 - 0.3962x + 173.6$

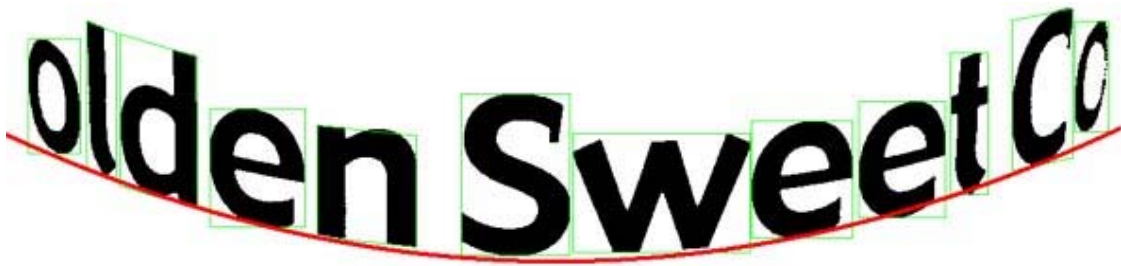
(b)

圖 5-2 四百公克果醬瓶曲線文字行偵測結果

(a)原圖(b)方程式逼近結果



(a)



上緣逼近方程式： $y = 0.0004206x^2 - 0.5069x + 283.1$

下緣逼近方程式： $y = 0.0004341x^2 - 0.5171x + 158.4$

(b)

圖 5-3 四百一十公克玉米罐曲線文字行偵測結果

(a)原圖(b)方程式逼近結果

表 5-2 實驗影像的平均均方根誤差與 R 平方係數

	圖 5.1	圖 5.2	圖 5.3	60 張實驗影像平均
RMSE(上緣 逼近)	18.68	2.129	19.57	10.67855
R-square(上 緣逼近)	0.2304	0.99	0.784	0.67591
RMSE(下緣 逼近)	5.236	2.783	4.995	2.88309
R-square(下 緣逼近)	0.9426	0.9877	0.9848	0.93799

由以上結果可以看出，由連通元件上緣進行拋物線逼近雖然通常效果不好，但依照文字不同有時也會出現比下緣更適合拿來作逼近結果的可能，我們可選擇效果較好者給予採用。以上拋物線對曲線文字行逼近的 R-square 參數都能夠維持在 0.93 以上，代表了逼近的曲線方程式與文字座標點的相關性非常高。

5.2 曲線文字行校正及其結果

針對曲線 Typoline 描繪與曲線文字行校正結果將在本節中進行呈現與討論。圖 5-4、圖 5-5、圖 5-6 三個範例為對影像進行 Typoline 描

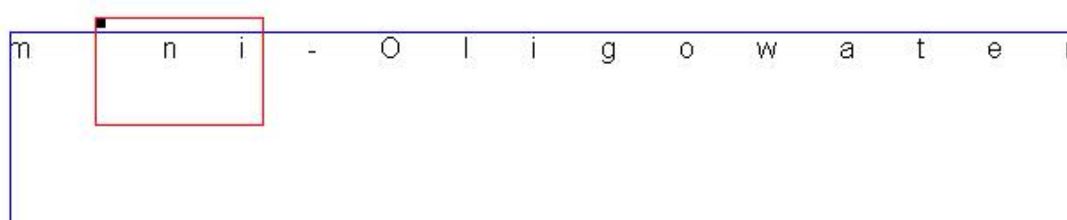
繪與曲線文字行校正的成效，並附上校正前與校正後的辨識結果，辨識程式為商用文字辨識軟體丹青。我們顯示實驗圖於校正前後的辨識正確率於表 5-3，另將六十張實驗影像的辨識率於表 5-4。由表中可看出，我們的曲線文字行校正系統能提升 10.16% 的文字辨識率。



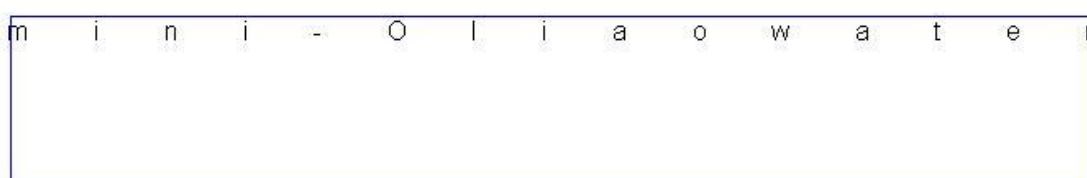
(a)

mini-Oligowater

(b)



(c)

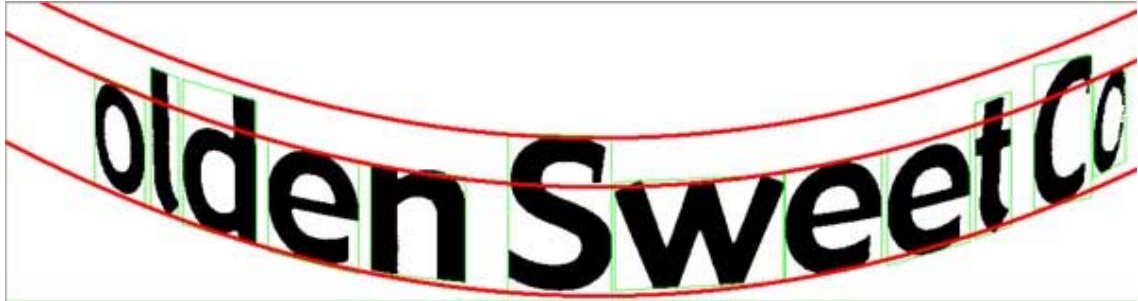


(d)

圖 5-4 六百毫升寶特瓶曲線校正圖

(a)曲線 Typoline(b)校正結果

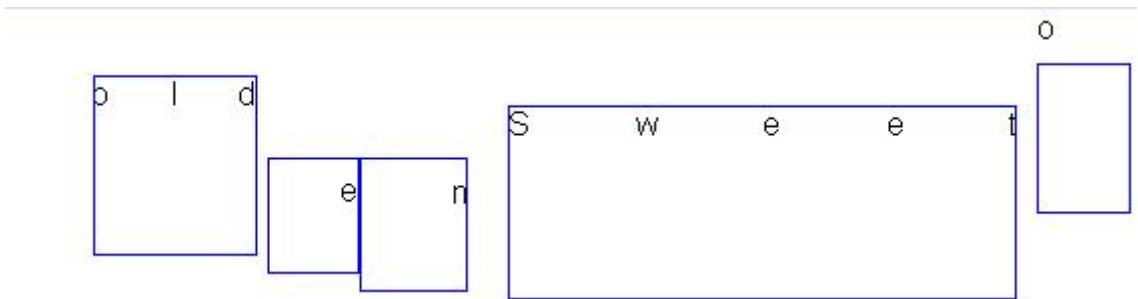
(c)校正前辨識結果 (d)校正後辨識結果



(a)

olden Sweet Co

(b)



(c)

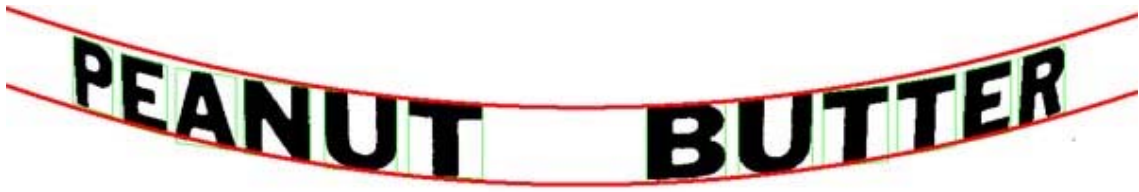


(d)

圖 5-5 四百一十公克玉米罐曲線文字影像校正圖

(a)曲線 Typoline(b)校正結果

(c)校正前辨識結果(d)校正後辨識結果



(a)

PEANUT BUTTER

(b)



(c)



(d)

圖 5-6 四百公克果醬瓶曲線文字影像校正圖

(a)曲線 Typoline(b)校正結果

(c)校正前辨識結果(d)校正後辨識結果

表 5-3 實驗影像曲線校正前與校正後的辨識率比較

	字 元 數	校正前錯誤 字元數	校正前辨識 率(%)	校正後錯誤 字元數	校正後辨識 率(%)
圖 5-4	14	1	92.857	0	100
圖 5-5	12	2	83.333	1	91.667
圖 5-6	12	3	75	0	100

表 5-4 60 張實驗影像曲線校正前與校正後的辨識率比較

	字 元 數	校正前錯誤 字元數	校正前辨 識率(%)	校正後錯誤 字元數	校正後辨 識率(%)
實驗影 像平均	769	111	85.566	31	95.968

5.3 圓柱寬度校正及其結果

針對圓柱寬度校正，將在本節中進行結果的呈現與討論。圖 5-7、圖 5-8、圖 5-9 三個範例為對影像進行寬度校正的成效，並附上校正前與校正後的辨識結果，辨識程式同上節，為商用文字辨識軟體丹青。

我們顯示實驗圖於校正前後的辨識正確率於表 5-4，另將六十張實驗影像的辨識率顯示於表 5-5。

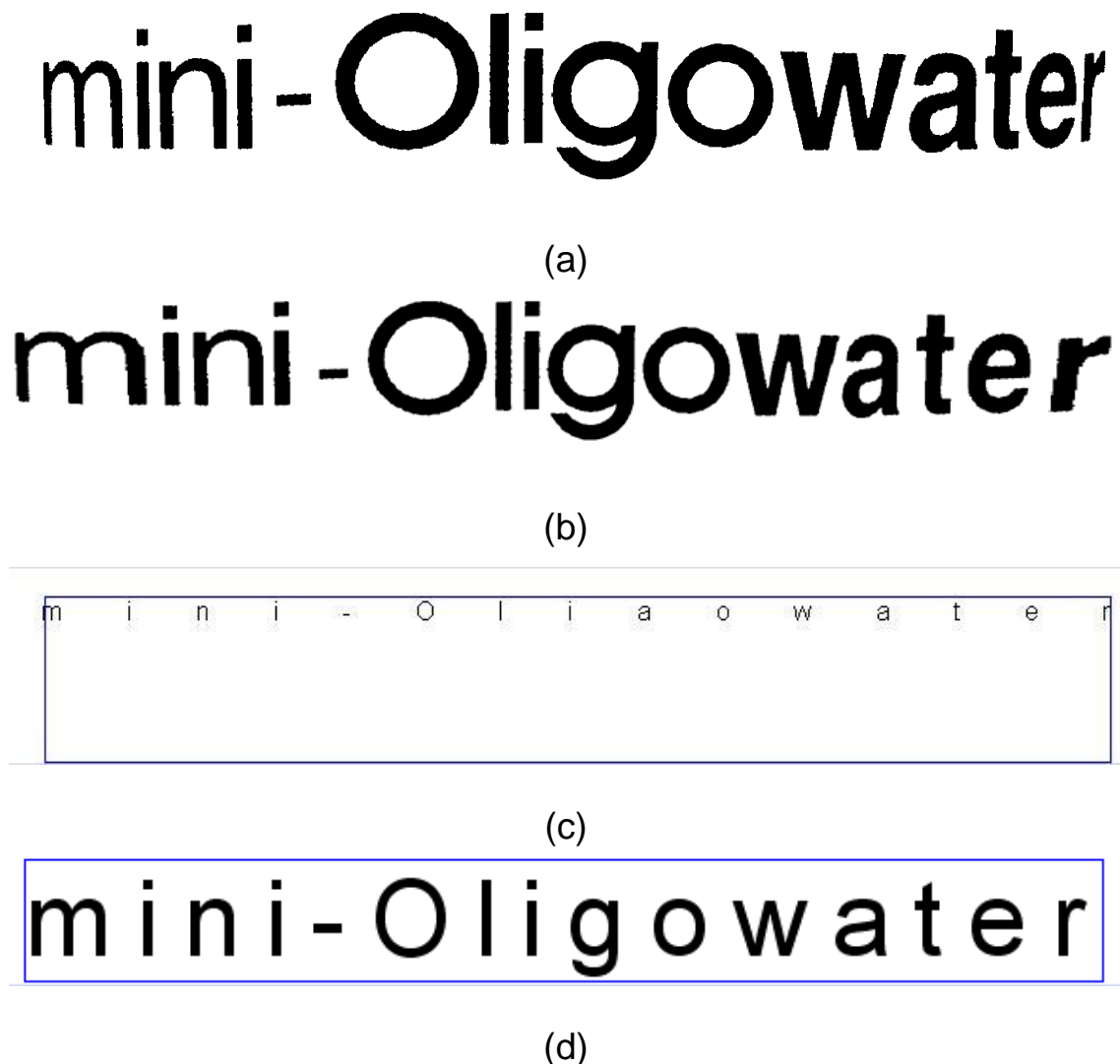


圖 5-7 六百毫升寶特瓶校正圖

(a)曲線 Typoline 校正結果(b)寬度校正結果

(c)校正前辨識結果 (d)校正後辨識結果

olden Sweet Co

(a)

olden Sweet Co

(b)

o l d e n S w e e t C c

(c)

o l d e n S w e e t C c

(d)

圖 5-8 四百一十公克玉米罐文字影像校正圖

(a)曲線 Typoline 校正結果(b)寬度校正結果

(c)校正前辨識結果(d)校正後辨識結果

PEANUT BUTTER

(a)

PEANUT BUTTER

(b)

P E A N U T

B U T T E R

(c)

P E A N U T B U T T E R

(d)

圖 5-9 四百公克果醬瓶文字影像校正圖

(a)曲線 Typoline 校正結果(b)寬度校正結果

(c)校正前辨識結果(d)校正後辨識結果

表 5-5 實驗影像寬度校正前與校正後的辨識率比較

	字 元 數	校正前錯誤 字元數	校正前辨識 率(%)	校正後錯誤 字元數	校正後辨識 率(%)
圖 5-4	14	0	100	0	100
圖 5-5	12	1	91.667	1	91.667
圖 5-6	12	0	100	0	100

表 5-6 60 張實驗影像寬度校正前與校正後的辨識率比較

	字 元 數	校正前錯誤 字元數	校正前辨 識率(%)	校正後錯誤 字元數	校正後辨 識率(%)
實驗影 像平均	769	31	95.968	28	96.358

5.4 討論

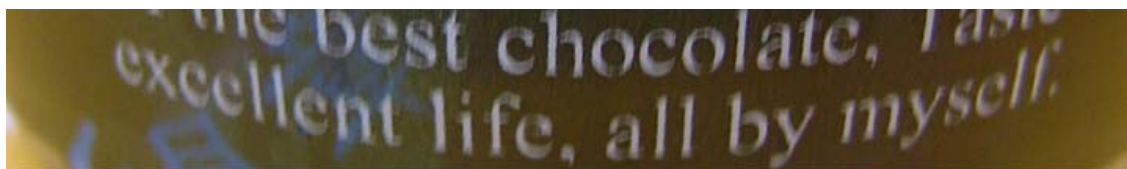
在經過本論文提出的校正後，對圓柱文字的辨識率仍然未能達到百分之百。出錯的理由大致有幾種：

首先是文字受到圓柱邊界切割導致變形。如圖 5-5 與 5-8 的最後一個字，在原來影像中 o 就有受到邊緣切割，所以儘管有保留資訊進行辨識，但仍不夠完整，讓辨識結果產生錯誤。

其次是二值化問題。圓柱影像中也包含著有複雜背景的文字影像，這些影像在經過二值化後很容易文字上產生雜訊或破碎，使得辨識軟體難以從中辨識出正確結果，而將其判斷為圖形。又或者原來在圓柱上的文字就偏小，導致二值化造成的結果不如預期得好，影響到辨識結果，如圖 5-10、5-11 所示。

最後則是連字與藝術字的問題。影像中包含著少數原來就相黏的連字，以及靠近圓柱邊界，二值化後相黏在一起的連字，這些連字讓辨識軟體在辨識上產生誤差，造成辨識上的錯誤，如圖 5-12。也由於圓柱影像的文字多屬於標題或標語，其中使用粗體字、斜體字等藝術字體的文字也提升了辨識上的困難。連字問題在經過寬度校正之後由於文字比例恢復，可以在一定程度上達到解決，但藝術字問題仍會對辨識造成影響。

總和所有錯誤辨識結果，在經過曲線校正後的 31 個錯誤中，因圓柱邊界切割造成的錯誤有 1 個；因二值化結果不良而導致錯誤的有 19 個；因連字與藝術字問題而造成的錯誤則有 11 個；而在經過寬度校正之後的影像中，能夠經由讓文字比例恢復，使得連字問題其中的 3 個獲得解決，如圖 5-13。

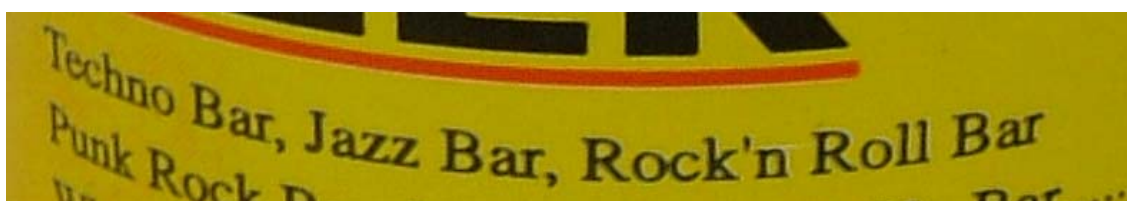


excellent life all by myself

e x c e l l e n t l i f e a l l b

y m y s c l F

圖 5-10 複雜背景影像圖及其校正與辨識結果



Techno Bar, Jazz Bar, Rock'n Roll Bar

T e c h n o B a r , J a z z B a r ,

R o c k ' n R o l l B a r

圖 5-11 二值化造成的文字破碎影像圖及其校正與辨識結果



Cream Style

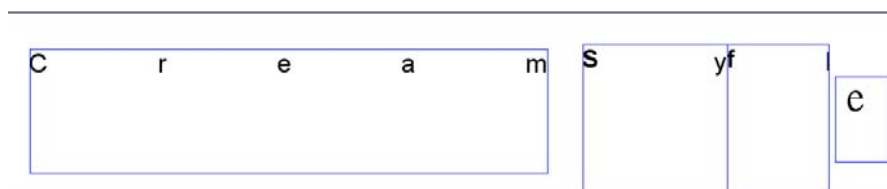


圖 5-12 連字與藝術字影像圖及其校正與辨識結果

Cream Style
Cream Style

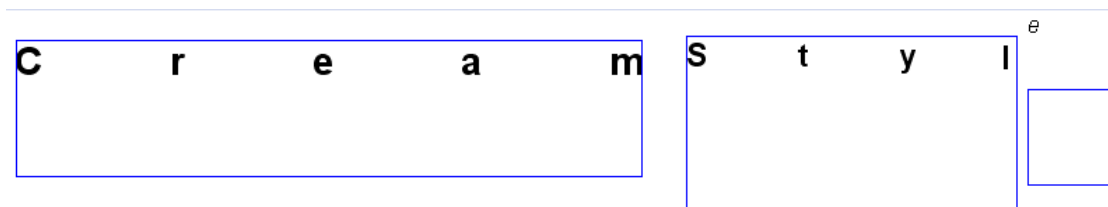


圖 5-13 寬度校正使文字比例恢復結果圖

表 5-7 實驗影像及校正結果比較





原始影像		
Hiroyuki Hase	0938-099-668	A B O U T W A T E R
Our algorism	0938-099-668	A B O U T W A T E R
原始影像		
Hiroyuki Hase	<u>Premium Black Tea</u>	<u>BLACK</u>
Our algorism	Premium Black Tea	BLACK

表 5-8 演算法辨識率比較

	Hiroyuki Hase	Our algorism
總字元數	769	769
錯誤字元數	32	28
辨識率(%)	95.838	96.358

以上是 Hiroyuki Hase 的曲線文字行校正法，與我們的校正方式進行的比較。由於 Hiroyuki Hase 的演算法僅針對文字的歪斜排列進行校正，而沒有對圓柱造成的文字寬度壓縮進行處理，導致在圓柱邊緣的文字校正上不夠完全。經由以上表可得知，在面對圓柱影像時，我們的演算法能取得比較好的效果。

第六章

結論與未來工作

6.1 結論

隨著科技不斷進步，數位相機的取像能力越來越好，也越來越普及，文字影像也不再被限制於文件影像，而文字所屬的平面也不再被限制為單一平面。本論文建立了一個三維圓柱曲面的文字校正系統，能對圓柱平面造成的文字扭曲進行校正。為了避免傳統矩形連通元件造成的資訊流失，我們以一種新的連通元件外型取代傳統的矩形對文字進行包覆；並能夠以回歸分析找出該文字行的逼近線。

在文字校正的部分，我們以文字逼近線平行畫出四條曲線 Typoline 做為校正資訊，並以此給予曲線文字行的校正。經由本論文的實驗結果證明，經過校正後的圓柱文字影像，在後續文字辨識的辨識率上確實有所提升。而同時本論文中所建立的 Typoline 特徵資訊也可延續到校正後，達到輔助文字辨識的功能。更重要的是，此實作可以真正應用在內建於手機、PDA、筆記型電腦等行動運算裝置的取像工具上。

6.2 未來工作

在未來工作上，我們將朝下列方向進行努力，以求能讓系統達到更好的效果。

首先可以考慮利用不同的前處理技術，如直接使用灰階影像來進行前景與背景的區分，可以減少雜訊的干擾，或直接由灰階影像資訊擷取字元特徵。而對文字偵測部分，為了能夠同時處理多條文字行，在連結文字連通元件時除了利用距離條件給予分析，也可參考其餘影像

特徵進行處理。而當影像不再局限於文件影像後，文字也不再完全以印刷字體為主，圓柱影像中的文字除了多出許多傾斜字體，也出現了更多五花八門的藝術字，造成文字辨識上的困難度。可以針對藝術字中的共通特徵進行處理，並透過細線化等特徵擷取方式以提升藝術字的辨識率。

最後，目前我們的研究僅針對圓柱影像進行處理，而我們將來會希望能將處理範圍擴展到其餘曲面，甚至不規則曲面上。為了完成此目的，需要能夠對更多變的文字行進行處理，並能夠以方程式描述文字行，以進行後續的校正與辨識。

參考文獻

- [1]. H. Yan, "Fuzzy Curve-Tracing Algorithm", *IEEE Transactions on Systems, Msn, and Cybernetics-PartB: CyberNetics*, Vol.31, No.5, October 2001.
- [2]. H. Yan, "Detection of Curved Text Path Based on the Fuzzy Curve-tracing (FCT) Algorithm", *International Journal on Document Analysis and Recognition*, 2001.
- [3]. B. S. Y Lum and H. Yan, "Complex Curve Tracing Based on a Minimum Spanning Tree Model and Regulrized Fuzzy Clustering", *International Conference on Image Processing*, 2004.
- [4] G. Toussaint, "Solving Geometric Problems with the Rotating Calipers", *IEEE proceedings In MELECON*, 1983.
- [5]. A. Zramdini and R. Ingold, "Optical Font Recognition Using Typographical Features", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 8, August 1998.
- [6]. T. Yamaguchi, Y. Nakano, M. Maruyama, H. Miyao and T. Hananoi, "Digit Classification on Signboards for Telephone Number Recognition", *conference on the Seventh International Document Analysis and Recognition*, 2003.
- [7]. H. Goto, K. Aoba, H. Aso, "A Framework for Detecting and Selecting Text Line Candidates of Correct Orientation", *IEEE conference on 14th International Pattern Recognition, icpr*, vol. 2, pp.1074, 1998.
- [8]. H. Goto, H. Aso, "Extracting curved text lines using local linearity of the text line", *International Journal on Document Analysis and Recognition*, 1999.
- [9]. Z. Zhang, C. L. Tan, "Correcting Document Image Warping Based on

Regression of Curve Text Lines ", *IEEE Conference on the Seventh International Document Analysis and Recognition*, 2003.

[10]. H. Hase, M. Yoneda, T. Shinokawa, C. Y. Suen, "Alignment of Free Layout Color Texts for Character Recognition", *Sixth International Conference on Document Analysis and Recognition*, pp.0932, 2001.

[11]. Din-Chang Tseng, "Image Processing" , *Institute of Compute. Sciences & Information Engineering National Central University*.

[12]. N. Otsu, "A Threshold Selection Method from gray-level Histograms." *IEEE International Conference on System, Man, and Cybernetics. vol. 9, pp. 62–66*, 1979..

[13]. J. T. Tou and R. C. Gonzalez, "Pattern Recognition Principles." *Addision-Wesley Publishing Company*, 1974.

[14]. S. S. Khan and A. Ahmad, "Cluster center initialization algorithm for K-means Clustering." *IEEE Pattern Recognition Letters, vol. 25, pp. 1293-1302*, 2004..