

行政院國家科學委員會專題研究計畫 成果報告

複雜背景影像中多尺度取向文字偵測技術及其應用之研究

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-003-011-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣師範大學圖文傳播系(所)

計畫主持人：周遵儒

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 11 月 1 日

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

複雜背景影像中多尺度取向文字偵測技術及其應用之研究

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC 92-2213-E-003-011

執行期間：92年08月01日至93年07月31日

計畫主持人：周遵儒 博士

共同主持人：陳舜德 博士

計畫參與人員：楊婉琳

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、
列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：國立台灣師範大學圖文傳播學系

中 華 民 國 93 年 10 月 31 日

摘 要

複雜背景文字偵測與切割是近年來重要的研究課題，許多相關應用均已此作為其核心技術，例如：排版複雜文件的光學文字辨識、以內容為基礎之影片媒體檢索等，本計畫我們針對這個主題進行一系列的探討，主要研究內容是：一、色彩學觀點之複雜背景文件影像與具字幕影片的色彩特性分析，二、利用色彩量化的原理進行彩色影像文字的偵測與切割，藉由這兩項關鍵技術的深入研究，可作為進一步應用研究與系統研發的理論基礎。

關鍵詞： 文字偵測、文字切割、文字辨識、彩色影像、色彩量化、以內容為基礎之影片檢索

Abstract

Text Detection and character segmentation from images with complex background are very important research issues in recent years. They are the kernel techniques of many related applications, for examples, optical character recognition for documents of complex layout, content-based video retrieval, etc. In this project, we devoted our attention to two basic research topics; including (1) color characteristic analysis of document images of complex background and video images with subtitles, and (2) color quantization-based text detection and character segmentation from color images. Both researches offer the theoretic basis and are much helpful for the further application studies and system development.

Keywords : text detection, character segmentation, character recognition, color image, color quantization, content-based video retrieval

一、背景

具字幕影片與複雜文件影像的文字偵測與切割是一個困難度相當高的分類課題，本質上是一種在所謂複雜背景影像中偵測文字的存在與否，並且將文字正確地從複雜背景影像中分離出來，最後將所分離的文字加以進一步辨識處理，而透過辨識後的文字正可以在多媒體影片的內容檢索中扮演具有關鍵性的角色，於是，實用系統的研發也是一項值得投入的工作，因此，在這整個處理流程中，我們所面臨的幾項研究課題分別是：複雜背景影像文字偵測技術的研究、低解析度文字影像辨識技術的研究、離形應用系統的系統整合與實作，這三項是本計畫必須突破的研究重點。

複雜背景影像通常是文字偵測與切割失敗的一大原因，雖然，已經有非常多的文獻探討文件版面切割 (Page Segmentation)、文件分析 (Document Analysis)、文字切割 (Character Segmentation) 等問題[5][6][7][8][9][10][11][12][13][14]，但是，先期的初步研究卻發現，在文件影像中背景底紋若與文字本身重疊出現時，在此類影像中，傳統的圖文分離技術主要利用二元影像中的幾何特徵 (Geometric Features) [5][6][7][8][9][10][11] 或者文字紋路特徵 (Texture Feature) [11][12][13][14] 並不能有效地區別二者，導致無法取得正確的文字影像資訊，而辨認失敗乃是必然的結果。

低解析度影像同樣是影響文字辨識效果的一項關鍵因素，在一些掃描速度要求非常高、記憶體要求非常低的問題中，這樣低品質影像會導致非常嚴重的結果，特別是近年來利用影片中文字作為影片檢索依據的探討愈來愈受到重視[15][16][17][18][19][20]，而目前常用的影片格式如 MPEG-1、MPEG-2 等，單一的影格畫面影像甚至低至 320x240，在這種情形下往往表現每一單一文字的影像大約只有 15x15 像素，再加上背景底紋的複雜性與多色彩的干擾，偵測文字的所在位置、圖文切割、文字辨識等等都是目前尚待解決的研究課題。

二、研究方法與成果

本計畫的研究重點有兩項：1、色彩學觀點之複雜背景文件影像與具字幕影片的色彩特性分析，2、利用色彩量化的原理進行彩色影像文字的偵測與切割，以下分就這兩部分加以說明。

1、色彩學觀點之複雜背景文件影像與具字幕影片的色彩特性分析

印刷品中文字影像會因為設計者而有眾多的變化，例如：重疊、手寫字、變形、文字影像色彩變化等等...。而印刷品中文字影像難以分割及辨識的原因就在於無法以一個文字作為分割的標準。印刷品文字影像的特點在於：(1)背景顏色複雜、(2)字的顏色多變且複雜或是反白字、(3)排版複雜、(4)印出的文件品質差，例如：傳真或是影印過的文件、(5)中英文夾雜、(6)特殊的字體。

印刷品文字影像解析度會隨著辨識器材與品質的高地而影響分割及辨識的好壞。一般處理印刷品中的文字影像會將掃描品質提高，若是掃描品質低或是文件品質低劣則會造成文字模糊或是無法區隔文字邊界的情形。

對影片媒體而言，一般人常用的影片儲存媒體主要有 DVD 與 VCD 兩種形式，畫面格

式與字幕附加方式略有不同，在 VCD 影片格式中影片佔滿所有影像尺寸，字幕壓在影片上，形成複雜背景的字幕型態，且 VCD 影片格式中的字幕顏色以白色、黃色為主，字體多為標楷體、粗明體、粗圓體居多；字的邊緣加寬一黑色邊界作為與複雜背景之間的區隔。另一方面，DVD 影片格式的影像呈現 3：7 的畫面比例，字幕多放置在畫面下方黑色部分，字幕顏色以白色為主，字體多為粗明體、標楷體；字的邊緣亦是加寬一黑色邊作為字與背景的區隔，此外，影片字幕顏色皆為單一色彩，不同於文件影像的色彩變化複雜情況。

	VCD 影片格式	DVD 影片格式
影片範圍	全螢幕方式呈現	3：7 比例呈現
字幕位置	壓在影片上	大多在影片下 部分壓在影片上
字幕邊緣加黑邊	有	有
字幕字形	標楷體、粗明體、 粗圓體	粗明體、標楷體
字幕顏色	白、黃色	白色

表 1、VCD 與 DVD 影片格式字幕狀態比較

DVD 與 VCD 影片的影像解析度也是不同規格的，以影像元素為基準來記錄或顯示影像表現的能力，在數位影像中是以像素(pixel)為單位，像素越高其解析度越高表示其影像品質表現越好。以影片解析度的高低而論，DVD 影片的解析度最高(640*480 pixel)，其次是 VCD 影片(320*240 pixel)，而網路視訊影片則是所有格式中解析度最低、影像品質最差的影片。然而，影片在轉製過程中容易經由不同層次、不同次數的檔案壓縮或者尺寸大小的更動而影響影片中的影像品質，這造成影像品質產生變化使得文字破損、筆畫斷裂。

本研究觀察色彩的概念在低解析度的文字影像上有可發揮的空間[26][27]。雖然觀察發現文字顏色經解析度的變換而有降低及損失的表現，但是在大部分影像中的文字觀察結果中，文字在影像中相鄰的色彩皆有彩度、色相上的明顯差異，利用色彩量化技術劃分文字影像色彩與其他非文字影像色彩，可獲得明顯成效並且有助於減少文字切割步驟。

觀察影片中低解析度文字影像的破壞情形，可由圖 1 及圖 2 來說明。圖 1 為未改變解析度前的文字影像，圖 2 為改變過解析度的文字影像。

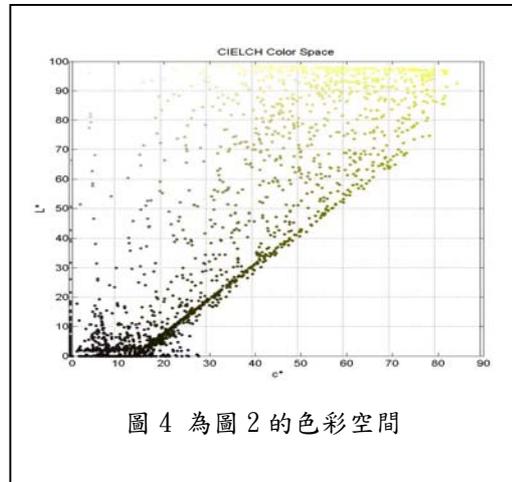
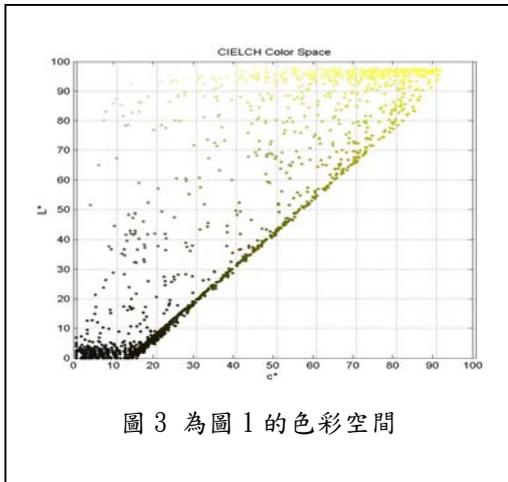


圖 1 改變解析度前的文字影像



圖 2 改變過解析度的文字區塊影像

比較圖 3 及圖 4 兩個影像，發現經改變過解析度的影像其黃色的彩度值由 90 降至 80；色彩擴散的情形改變了影像的明度值而使影像的彩度降低，色彩飽和度降低；黑色與黃色之間的混合色增加造成文字色彩模糊，筆劃結構鬆散。



2、利用色彩量化的原理進行彩色影像文字的偵測與切割

本研究利用 K-Means 群聚演算法作為執行色彩量化的技術，K-Means 群聚演算法進行在群聚方法中是被廣泛使用的一項技術。所謂的 K-means 群聚演算法非常簡單，假設整張影像一共有 N 個像素，且整張影像將被分割成 K 個影像分割，每一個像素的色彩向量值為 $\mathbf{p}_i = (R_i, G_i, B_i)^T$ 為 $i = 1, 2, \dots, N$ ，演算法如以下所示：

[K-means Clustering Algorithm] [28]

```

begin initialize  $N, K, \mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K$ 
    do          classify  $N$  pixels  $\mathbf{p}_i$  according to the nearest  $\mathbf{c}_i$ 
                recompute  $\mathbf{c}_i$ 
    until      no change in  $\mathbf{c}_i$ 
    return     $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_K$ 
end

```

經由 K-means 群聚分析後，整張影像 I 即被分割為 $I_{c_1}, I_{c_2}, \dots, I_{c_K}$ 等 K 個子影像，而且， $I_{c_i} \cap I_{c_j} = \text{empty set}$ ，其中 $i, j = 1, 2, \dots, K$ 且 $i \neq j$ ，同時， $I = I_{c_1} \cup I_{c_2} \cup \dots \cup I_{c_K}$ ，圖 5 所示是一張擷取自影片之影像框 (Image Frame) 範例，圖 6-1 與圖 6-2 是圖 5 透過 K-means 群聚演算法做影像分割後的結果，其中文字部分可以在數個影像分割中清楚呈現出來，特別是 $I_{c_{14}}$ ，此為文字主體顏色像素所在之分割。



圖 5、影片影像範例（取自『貓狗大戰』電影畫面）

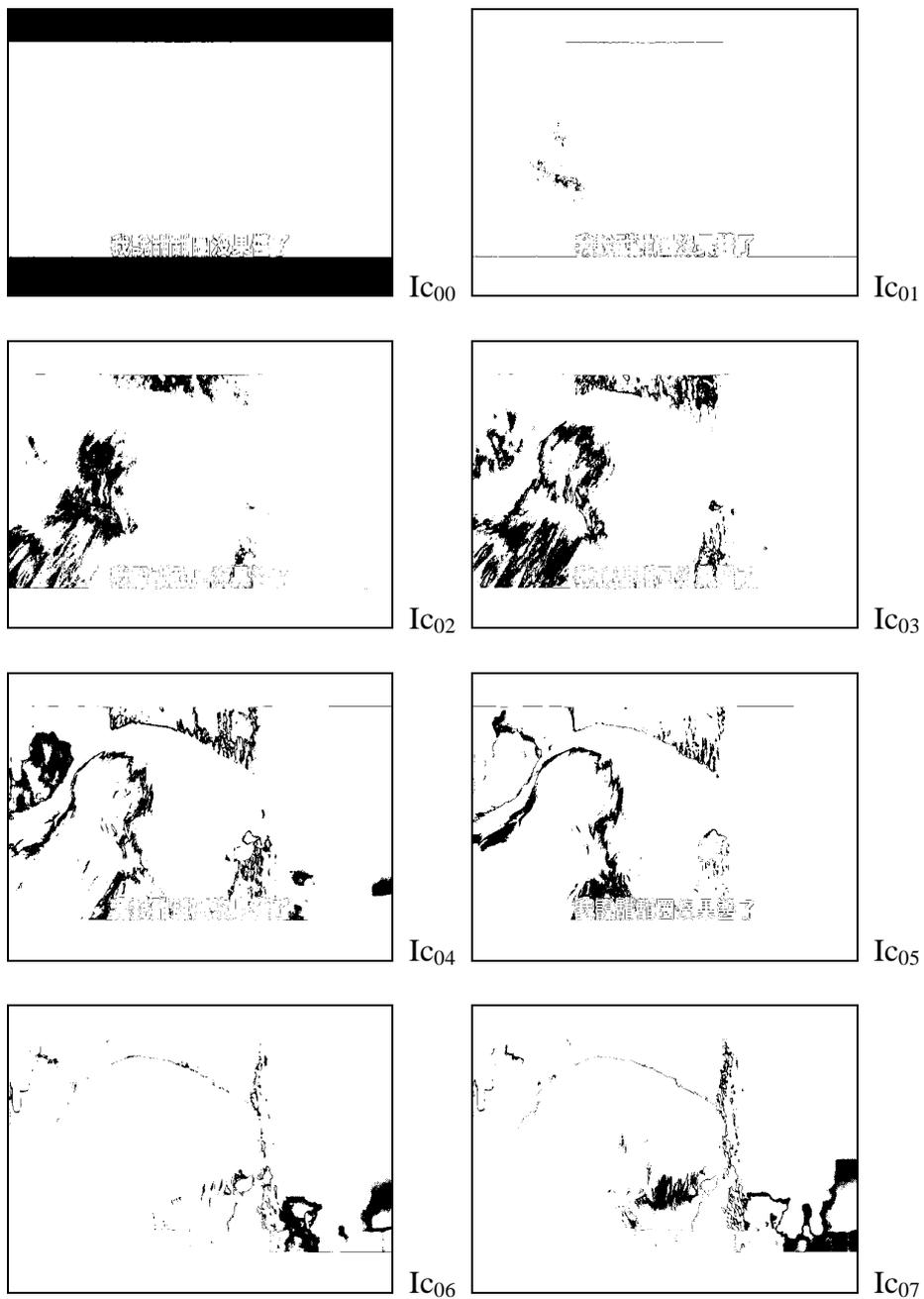


圖 6-1、影片框影像不同色彩區域之影像切割範例（色彩編號 C00 至 C07）

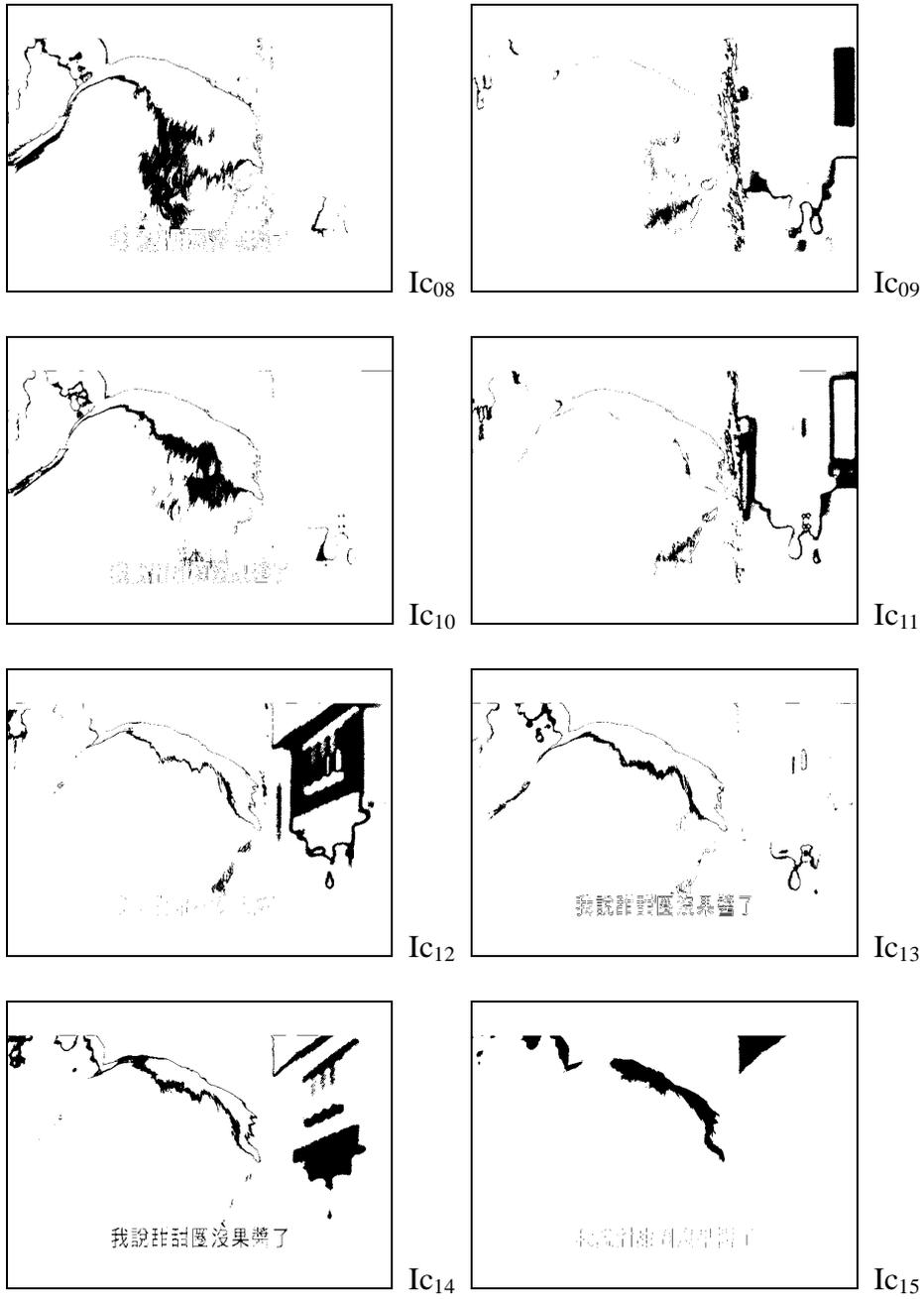
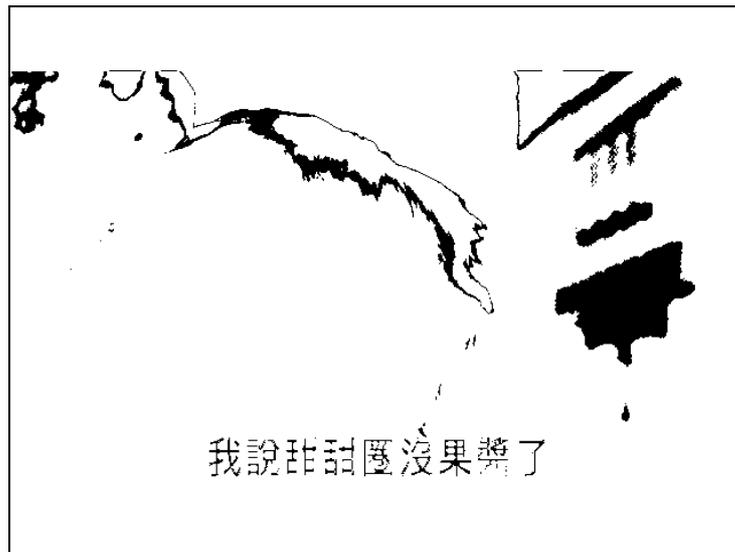
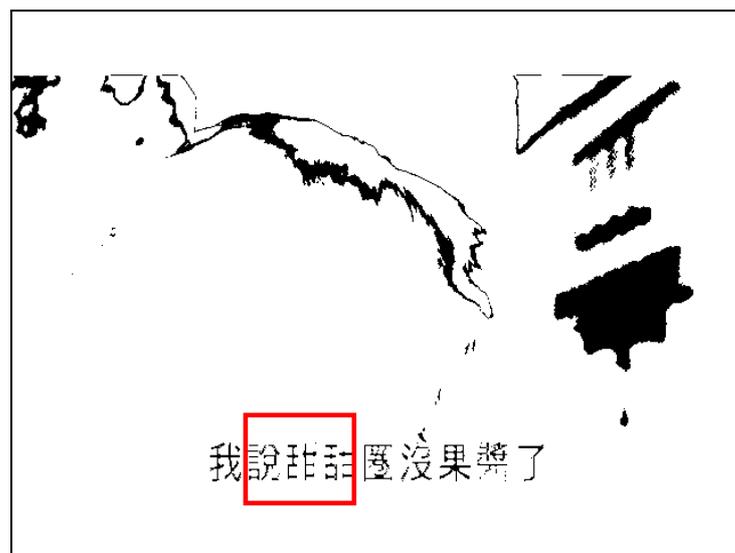


圖 6-2、影片框影像不同色彩區域之影像切割範例（色彩編號 C08 至 C15）



(a)



(b)

圖 7、(a)為色彩編號 c_{14} 的影像 $I_{c_{14}}$ ，(b)紅線框取部分為 $I_{c_{14}}$ 含有文字的區域

針對文字主體顏色像素所在之分割影像 $I_{c_{14}}$ ，運用文字與背景所展現的不同影像特性，即可分割影片中的字幕影像與非字幕影像(如圖 7 所示)，完成影片中字幕的切割研究。

圖 8 至圖 10 所示是典型的彩色文件影像範例，經由一系列色彩量化影像切割的結果，本質上所表現出的現象與影片影像框是類似的，並沒有顯著的差異，因此，在某整程度上也說明了利用色彩量化做切割的方式，比較不受的影像來源的影響，能夠透過簡潔而系統化的程序得到相當一致的結果。

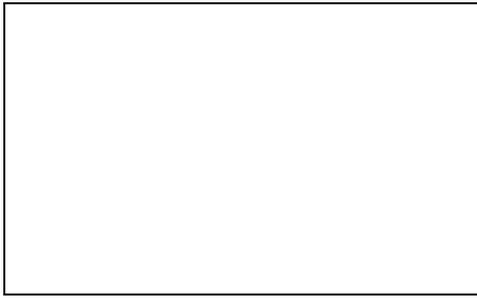


The advertisement features a dark red background with a person in a blue shirt and black pants. A large white speech bubble with the word "NEW" in red is positioned in the upper right. The main headline reads "輕鬆上手，建構電腦安全完整防護" (Easy to use, building a complete computer security protection). Below this, three key features are listed with icons and descriptions:

- 告別感染新病毒的夢魘 主動式病毒預警通知 (Outbreak Alert)**
全新研發的主動式病毒預警通知功能，會搶先在新病毒肆虐之前，主動向您發出病毒疫情警告，幫您及時提升防護戰備，讓病毒無可趁之機。
- 無線網路讓您更安心 Wi-Fi無線安全防護 (Wi-Fi Protection)**
隨著無線網路的快速發展，透過行動式個人電腦於公開場合進行資料傳輸機率與日遽增，安全防護也越顯重要。藉由趨勢科技全球首創的Wi-Fi無線安全防護機制的啟動，不論何時何地悠遊無線網路，都能讓您更加安心自在，不必擔心個人機密資料外洩。
- 防範網路恐怖攻擊 主控式個人防火牆 (Enhanced Personal Firewall)**
全新研發的主控式個人防火牆功能，除具備駭客入侵偵測功能外，還能即時監控、過濾與追蹤任何形式的網路資料傳輸或交換，讓所有駭客無可遁形。主控式個人防火牆還能避免您的電腦遭到來自網站的惡性程式扶持，還能為您阻絕特洛伊木馬程式的襲擊，避免你成為病毒的幫兇。

On the right side, there is an image of the PC-cillin 2003 software box, which is red and white with a blue pill graphic. Below the box, the website www.trendmicro.com.tw and the customer service number 02-2378-3666 are provided.

圖 8、複雜背景文件影像影像範例（取自趨勢公司廣告資料）



Ic00



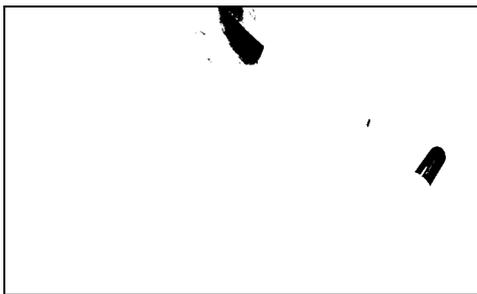
Ic01



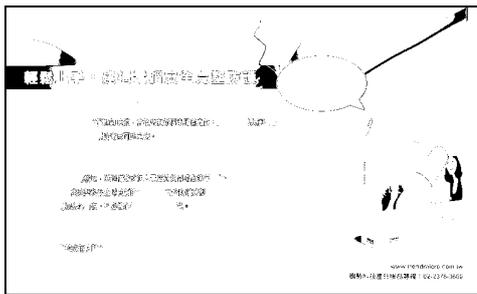
Ic02



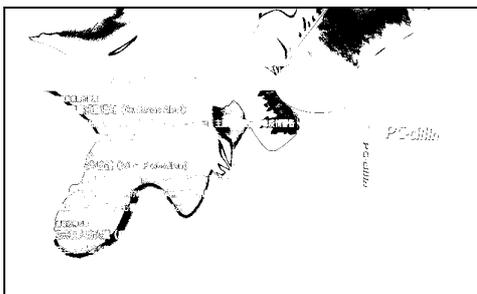
Ic03



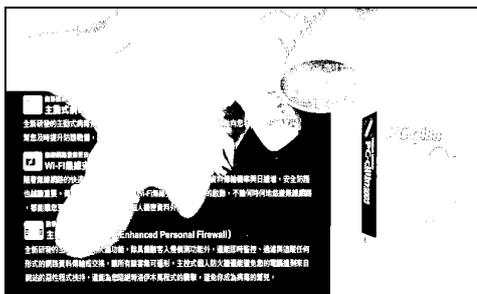
Ic04



Ic05



Ic06

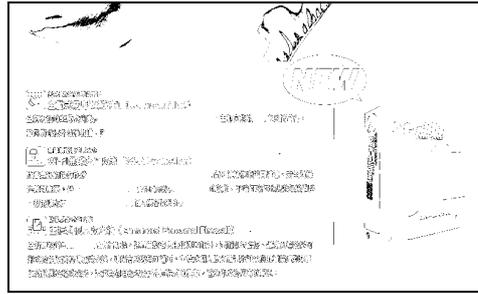


Ic07

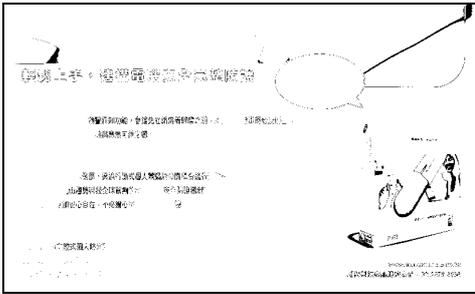
圖 9-1、文件影像不同色彩區域之影像切割範例（色彩編號 c00 至 c07）



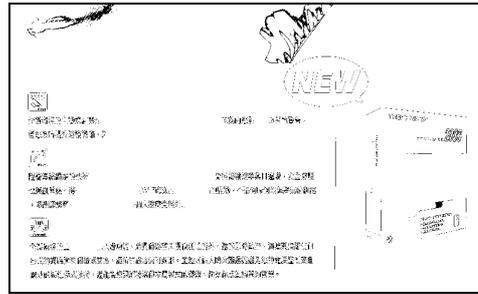
Ic08



Ic09



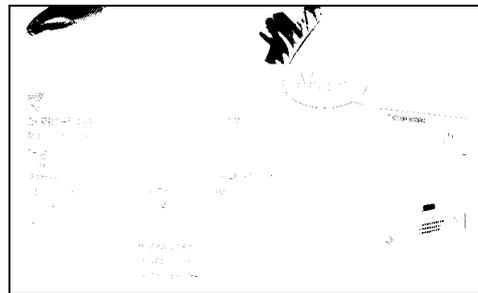
Ic10



Ic11



Ic12



Ic13



Ic14

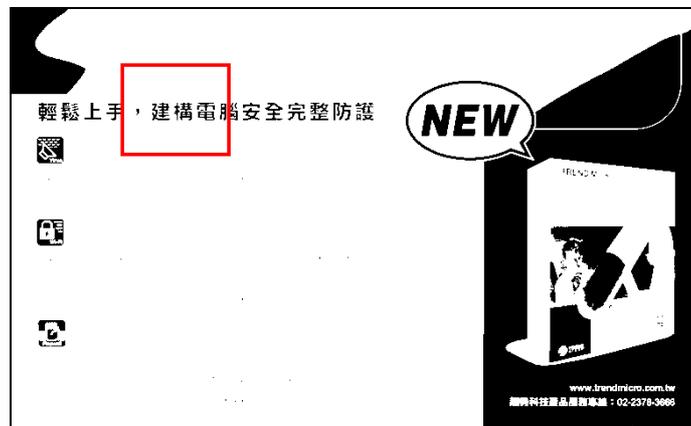


Ic15

圖 9-2、文件影像不同色彩區域之影像切割範例 (色彩編號 c08 至 c15)



(a)



(b)

圖 10、(a)為色彩編號 c_{15} 的影像 $I_{c_{15}}$ ，(b)紅線框取部分為 $I_{c_{15}}$ 含有文字的區域

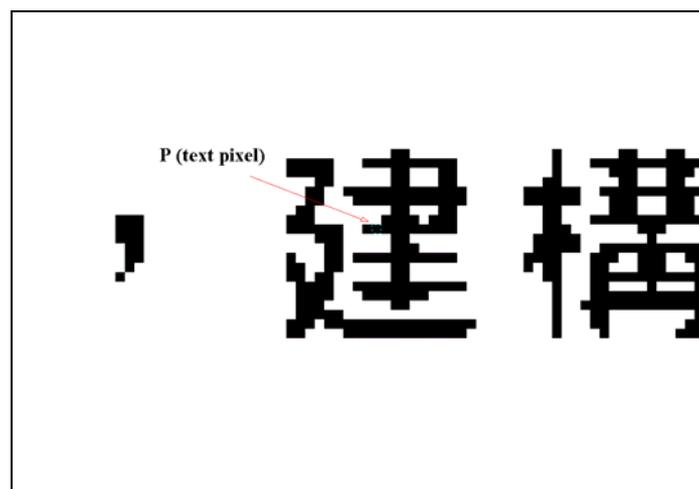


圖 11、圖 10(b)之紅線框取部分放大，其中P為 $I_{c_{14}}$ 中的文字像素

三、結論與討論

本計畫中完成了以下的研究工作：

1. 從色彩學觀點分析文字影像的特性
2. 建立數學模式用以分析描述文件影像中文字與背景底紋的差異性。
3. 開發完成影片字幕文字切割軟體。
4. 建立低解析度影片字幕文字的變形數學模式。
5. 訓練參與研究者系統研發與實作的能力。

目前此技術有效率的應用在相關研究中，並且與相關研究學者進行合作，協助跨平台多功能的影像檢索式的教學平台之研究。

參考文獻

- [1] F. Chang, et al., "A Document Analysis and Recognition System," *Inter. Conf. Document Analysis and Recognition*, Ulm, Germany, 1997.
- [2] F. Chang and T. R. Chou, "Problems and Solutions for a Content Capture and Information Retrieval System," *4th IAPR International Workshop on Document Analysis Systems*, 2000, Rio de Janeiro, Brazil.
- [3] F. Chang, "Retrieving Information from Document Images: Problems and Solutions," *Intern. J. Document Analysis and Recognition, Special Issues on Document Analysis for Office Systems*, vol. 4, no. 1, pp. 46-55, 2001.
- [4] T. R. Chou and F. Chang, "Optical Chinese Character Recognition for Low-Quality Document Images," *Inter. Conf. Document Analysis and Recognition*, Ulm, Germany, 1997.
- [5] M. Nadler, "A Survey of Document Segmentation and Coding Techniques," *CVGIP*, vol. 28, pp. 240-262, 1984.
- [6] T. Pavlidis and J. Zhou, "Page Segmentation and Classification," *CVGIP: Graphical Models and Image Processing*, vol. 7, no. 2, pp.367-377, 1998.
- [7] D. Wang and S.N. Srihari, "Classification of Newspaper Image Blocks Using texture Analysis," *CVGIP*, vol. 47, pp. 327-352, 1989.
- [8] F.M. Wahl, K.Y. Wong, and R.G. Casey, "Block Segmentation and Text Extraction in Mixed Text/Image Documents," *CVGIP*, vol. 20. pp. 375-390, 1982.
- [9] S.W. Lee and D.S. Ryu, "Parameter-Free Geometric Document Layout Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, Issue: 11, pp. 1240-1256, 2001.
- [10] Y. Xiao and H. Yan, "Text Region Extraction in a Document Image Based on the Delaunay Tessellation," *Pattern Recognition*, vol. 36, pp. 799-908, 2003.
- [11] A. Jain and S. Bhattacharjee, "Text Segmentation Using Gabor Filters for Automatic Document Processing," *Machine Visual Application*, vol. 5, pp. 169-184, 1992.
- [12] A. Jain and B. Yu, "Document Representation and Its Application to Page Decomposition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20 pp. 294-308, 1998.

- [13] K. Etemad, D. Doermann, and R. Chellappa, "Page Segmentation Using Decion Integration and Wavelet Packets," *Proc. of 12th IAPR Intern. Conf. on Pattern Recognition*, vol. 2, pp.345-349, 1994.
- [14] V. Wu, R. Manmatha and E.M. Riseman, "TextFinder: An Automatic System to Detect and Recognize Text in Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1224-1229, 1999.
- [15] A. Jain and B. Yu, "Automatic Text Location in Images and Video frames," *Pattern Recognition*, vol.31, no. 12, pp. 2055-2076, 1998.
- [16] D. Lopresti and J. Zhou, "Locating and Recognizing Text in WWW Images," *Information Retrieval*, vol. 2, pp. 177-206, 2000.
- [17] H. Li, D. Doermann and O. Kia, "Automatic Text Text Detection and Tracking in Digital Video," *IEEE Trans. on Image Processing*, vol. 9, no. 1, Jan. 2000.
- [18] K. I. Kim, K. Jung, S.H. Park, and H.J. Kim, "Support Vector Machine-Based Text Detection in Digital Video," *Pattern Recognition*, vol. 34, pp. 527-529, 2001.
- [19] X. Tang, X. Gao, J. Liu, and H. Zhang, "A Spatial-Temporal Approach for Video Caption Detection and Recognition," *IEEE Trans. on Neural Networks*, vol. 13, no. 4, 2002.
- [20] R. Lienhart and A. Wernicke, "Localizing and Segmenting Text in Images and Videos," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 4, April 2002.
- [21] S. Amari, A. Cichocki, and H. Yang, "A New Learning Algorithm for Blind Signal Separation," *Advances in Neural Information Processing Systems 8*, pp. 757-763, 1996.
- [22] S. Amari and J.F. Cardoso, "Blind Source Separation – Semiparametric Statistical Approach," *IEEE Trans. on Signal Processing*, vol. 45, no. 11, pp. 2,692-2,700, 1997.
- [23] T.W. Lee, M.S. Lewicki, and T.J. Sejnowski, "ICA Mixture Models for Unsupervised Classification of Non-Gaussian Classes and Automatic Context Switching in Blind Signal Separation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, Oct. 2000.
- [24] A. Cichocki and S. Amari, "Adaptive Blind Signal and Image Processing – Learning Algorithms and Applications," John Wiley & Sons, LTD., 2002.
- [25] T-R Chou, "Deformation Models of On-Line Cursive Chinese Characters and Their Applications," *Ph.D. Thesis*, National Tsing Hua University, (1995).
- [26] 大田 登原著，陳鴻興 陳君彥編譯，基礎色彩再現工程，全華科技圖書股份有限公司，2003。
- [27] Mark D. Fairchild, *Color appearance models*, Mass: Addison-Wesley, 2003.
- [28] Richard O. Duda, Peter E. Hart, and David G. Stork, "Pattern Classification," Second Edition, John Wiley & Sons, Inc., 2001.
- [29] A. Yoshitaka, T. Ichikawa , "A survey on content-based retrieval for multimedia databases," *IEEE Transactions on Knowledge and Data Engineering*, vol.11, pp.81-93, 1999.