

影像文字自動擷取演算法

蔡賢亮
義守大學
資訊管理學系
助理教授

jim@isu.edu.tw

林君諺
義守大學
資訊管理學系

m9422020@
stmail.isu.edu.tw

蔡佩君
義守大學
資訊管理學系

m9522005@
stmail.isu.edu.tw

周煜書
義守大學
資訊管理學系

m9522021@
stmail.isu.edu.tw

摘要

本研究旨在發展並改良其於連續影片中自動擷取影像文字的演算法。影像文字的擷取分為兩部份：一、影像文字區塊之偵測；二、影像文字字元之切割。本研究中，對區塊偵測採用混合統計式暨類神經網路之平行偵測演算法，其前置處理上運用高斯濾波器（Gaussian Filter），有助於使文字區塊中破碎字元予以完整化，並對於非文字的複雜背景作初步過濾，方便後續邊緣偵測和紋理分析的階段。字元切割部分，是將前部份偵測出的文字區塊作直方圖等化，搭配垂直投影量之特性，即能將文字字元分別切割。本研究能有效地擷取影片中的文字字元，往後將資訊儲存至資料庫中以供查詢。此影片資訊管理系統之影片文字擷取架構，對於龐大的影片管理有極其貢獻。

關鍵字：文字擷取、文字偵測、字元切割、高斯低通濾波、數位典藏

the image characters block detection, we use a hybrid automatic character blocks detection. It includes both neural-based and statistics-based detecting algorithms. And it also uses the Gaussian filter to increase image quality before detecting. For instance, the characters block incompleteness, image noise and complicated background. In the experiment, we find Gaussian filter is a good way to improve characters block detection. It can improve some problems in characters detection. In the image characters segmentation, we first apply histogram equalization to characters block. Then we use the vertical projection method to segment every characters in each block. Our algorithm can help us extract characters information in numerous film automatically and efficiently. It will help us to construct a digital archival video system automatically.

Keywords: character extraction, character detection, character segmentation, Gaussian filter, digital archives.

Abstract

We develop and improve an algorithm which can extract characters automatically in image among continuous film. The algorithm of automatic image characters extraction can extract characters information in image among continuous film, and store the character information to database for searching. This function has made very great contribution to huge miscellaneous film management. The character extraction of the image can be divided into two parts: the image characters block detection, and image characters segmentation. In

1. 導論

新聞影像數位典藏涵蓋的技術非常廣泛，其中的「數位內容自動建構機制」便是一個相當重要的部份。此部份是用來自動擷取出新聞影片中重要資訊（例如，新聞影像中的標題文字便是最重要的資訊），我們的目標是希望能夠在新聞節目的內容中找到標題文字出現的位置，再將標題文字擷取出來以便進一步做辨識。

再利用此資訊作為新聞影像分類的依據，最後將新聞分門別類地典藏在資料庫中，以方便往後使用者檢索新聞。新聞影像

的數位內容自動建構機制可粗略分為三大流程：一、影像文字區塊偵測與擷取；二、影像文字辨識及關鍵字辨識；三、分類器與資料庫的建立機制。本研究沿用了[1]影像文字區塊偵測演算法，以類神經網路為基礎，應用於新聞影像中的標題文字偵測。此方法包含了兩大階段：第一階段為邊緣偵測的部份。我們使用 $n \times n$ 輸入視窗當成類神經網路的輸入，經過計算後會留下影像中物件的邊緣部分，將影像中物件的輪廓標示出來。第二階段：針對影像的邊緣部分使用紋理分析的方法找到標題文字的區域。最後判斷區塊尺寸大小，若區塊太小則此區塊可視為非標題文字部分即可將此區塊去除，保留下的區塊即可視為標題文字的區域。加入高斯濾波器（Gaussian Filter）來修正影像的雜訊及背景的干擾，文字字元的分割採用垂直投影量的特性來切割字元。本研究提出的演算法可以自動從新聞影像中擷取文字區塊，並將文字區塊作字元分割，達成新聞影像自動化的文字擷取，對於新聞影像數位典藏的資料庫建置有很大的貢獻。

新聞影像文字擷取其效能好壞，將關係到整個新聞影像數位典藏的優劣，因為文字區塊的誤判將會造成後序龐大的額外處理，也會造成重要資訊的遺失，導致整個新聞影像數位典藏的品質低落及效率不佳。

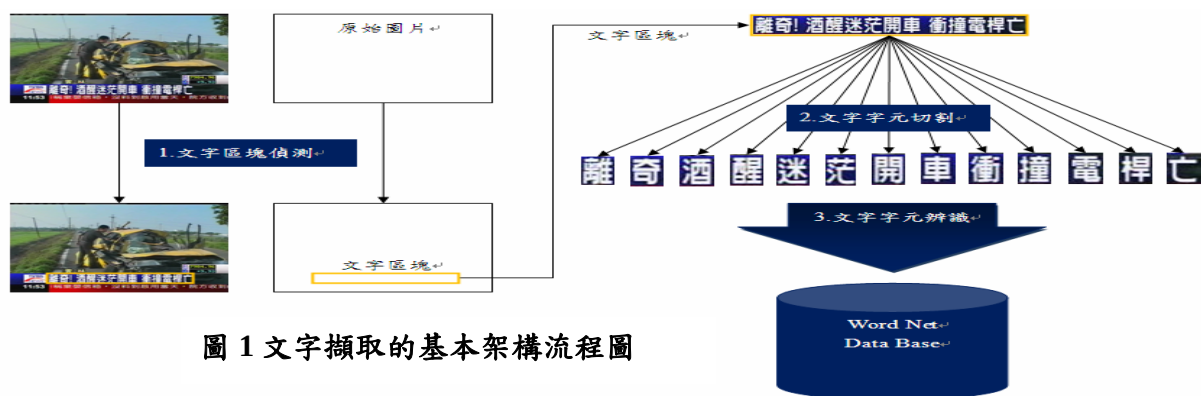


圖 1 文字擷取的基本架構流程圖

一個好的新聞影像數位典藏系統需要一個有效能的文字自動擷取機制，這也是本研究主要的目標。

2. 文字擷取演算法基本介紹

傳統影像的檢索系統，是以文字為基礎的模式。在資料庫中，每一張影像都必須經過人工的註解，利用使用者輸入的文字與每張影像的註解作比對，找出可能的資訊。而人工的註解不但費時費力且成本高，若可以利用既有影像資料中的資訊直接作為檢索值，如此便不需要對每張影像作註解而可以直接對影像資料庫搜尋，將可以節省許多影像前置處理的成本。過去關於視訊內容分析的許多研究中，常會看到利用視訊檔案中視訊格（Frames）裡的各種特徵資訊來對整個視訊檔案內容作索引，其中自動判斷其內容的技術是近年來研究的重心，而最直接的方法即是判斷出現在畫面中的字幕，例如：新聞節目的標題字幕（標題字幕中可能包含了對此則新聞的重要內容）；而戲劇節目或電影中的字幕則是整個故事的內容、片名、以及演員的名字。對於體育節目中的字幕，包含了比賽中的重要訊息，國內也有許多關於體育節目中字幕擷取的研究[2][3]，並分類各種精彩的畫面。影像中的文字擷取演算法可以粗略分成兩個部份：一、影像文字區塊之偵測；二、影像文字字元的切割。此二部分為連續性的處理過程，前階段的處理效果會大大影響到後續階段的處理結果的好壞，若前階段的效果並不理想，便會造成接下來的階段有極大的問題產生。舉例來說，在第一部份文字區塊偵測的處理過程中，若不能精

確的找到文字區塊的位置，則在第二部份文字字元分割時，便容易產生處理上的困難，如字元分割錯誤，字元分割不完全等，若偵測的文字區塊是錯誤的區塊，更會造成第二階段文字字元無法分割，甚至造成後續研究

無法辨識的大問題。圖 1 為文字擷取的基本架構流程圖（以新聞影像為例）。

3.影像文字區塊偵測

文字偵測的主要目的在於找出文字位於畫面中的位置，也就是文字區塊的所在，偵測的方法則是針對文字在影像中所呈現的特徵來進行設計。一般而言有下列幾種方法：

3.1 以紋理（Texture）為基礎的演算法

根據文字有相似的方向和頻率的資訊，因此文字可被視為是一種特殊的材質。文字區塊的紋理特徵，與背景間通常有較明顯的對比存在。文字與文字之間通常也會留有一定的間距，所以利用這些區域的密度變化，可以找出一些排列變化。因此，可以透過紋理的特徵將文字區域找出來。

3.2 以顏色相似度（Color Similarity）為基礎的方法

此種方法的前提，在於擷取的字幕文字需有相似的顏色，因此處理背景僅能為全黑或全白的字幕文字，對於場景文字以及位於複雜背景中的字幕文字，將無法正確的被偵測出來。若依顏色相似度偵測出位置後，再使用連接元件（Connected Component）方法將大小相似、具水平排列的可能文字連接成一個字串。如 A. K. Jain[8]處理彩色影像的方法以及 R. Lienhart[4]。

3.3 以邊緣（Edge）為基礎的方法

在檢測數位影像中有三種基本灰階不連續類型：點、線、以及邊緣的技術，而根據文字具有多筆畫（即多邊緣）以及邊緣聚集的特性，若能偵測出邊緣的資訊（Edge Map），再利用各種合併方法如連接元件法來連接相近的文字像素點，或者利用如型態學運算（Morphological Operations）來連接相近的文字區塊，便能處理出不論背景單純或複雜的字幕文字；而場景文字方面，僅能處理高對比度且文字周圍為單純背景的情形，不過此種方法容易將一群小物件誤認為是文字區塊，這是因為很多聚集的小物件也會呈現多邊緣且邊緣叢聚的特色，容易造成高錯誤偵測率的結果。

4.影像文字字元切割

做字體辨識的時候，通常都是一個字元做一次辨識，而不是一整串文字一起進去做辨識，所以必須要有文字字元切割的動作。在此切割文字字元所使用的方法有下列幾種：

4.1 區塊化（Labeling algorithm）分割演算法

區塊化[5]顧名思義就是將影像中連結在一起的像素視為一個區塊，對影像區分成數個區塊。越複雜的影像，區塊可能越多，而這樣的處理，在一般的影像處理中，是針對影像的前景來作區塊化的處理。連結情形的判斷基本上有兩種方式，一種是四連結（4-connect），另一種是八連結（8-connect），四連結的方式，即像素與其上下左右四個方向的相鄰像素才稱為鄰居。而八連結的方式，是除了四連結的鄰居之外，在加上左上、左下、右上、右下等斜對角四個方向的相鄰像素。而區塊化演算法步驟有三：

- 1.獨立搜尋影像每個水平像素的連結情形，記錄每個水平區塊的起始位置和終止位置。
- 2.以記錄的區塊為主，逐次由影像的上端往下搜尋，比較每個區塊與其上一水平的所有區塊的相鄰情形，若符合四連結的相鄰情形，這將兩相鄰區塊標為同一區塊。
- 3.當影像搜尋完後，整理記錄，即可分出多個區塊，並為每個區塊作編號。

4.2 投影量法

通常與背景形成對比的文字，在其上面邊緣與下面邊緣應該會有大的垂直密度變化，因此我們預期水平文字行會有高的垂直頻率。而針對每個可能的單一文字字元區域，必須確定有足夠強烈明顯且密集的邊緣資訊，若為真正的單一文字字元區域，則在每個字的列都具有高的投影量，而且也應該有強烈明顯且密集的邊緣資訊，字元與字元間會產生峰谷的現象，因此我們使用水平投影量的方法來偵測峰谷現象，進而分割文字字元。以車牌字元切割為例[6]，利用車牌垂直列的黑色點分佈來判斷字元的分割位置。以車牌的特性而言，字元與字元之間會有一些空白，而這些空白在垂直投影下會出現最

少黑色點，利用此特性可找出字元間的分界。如下圖 2 所示：

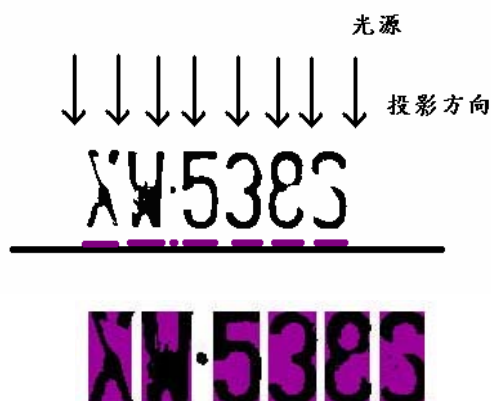
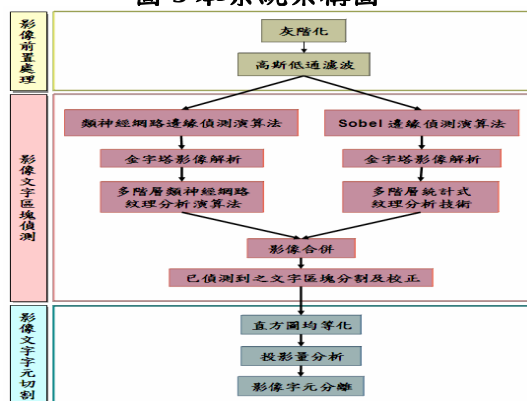


圖 2 車牌字元投影分割示意圖

5. 實驗架構

本研究為發展一套可以自動擷取出新聞影片片段中重要之資訊（例如，新聞影像中的標題文字便是最重要的資訊）的系統，再利用此資訊作為新聞影像分類的依據，最後將新聞分門別類地典藏在資料庫中，以方便往後使用者檢索新聞。而本系統分為兩階段架構，下列分述於第 6 與第 7 章節，本系統架構如圖 3 所示。

圖 3 本系統架構圖



6. 文字偵測階段

文字偵測的演算法在過去已有許多的方法被提出，且有不錯的成果。然而影像中的文字表現是多樣性的，如文字型態、大小，角度的不同，都會影響偵測的結果。為了降低文字偵測所出現的誤判區塊，本研究提出

混合式的演算法，針對演算法與應用兩個層面來探討影像文字偵測技術，在演算法的部分之中，對於目前已發表的方法，可以將其分為兩大類：統計式演算法（例如：Sobel 邊緣偵測演算法、離散小波轉換、傅利葉轉換等）及人工智慧演算法（類神經網路演算法、模糊理論及資料探勘等）。表 1 與表 2 為兩者方法的比較說明。下頁為本系統文字區塊偵測架構圖。文字偵測處理粗略可分以下四個階段：6.1 前置處理。6.2 邊緣偵測與紋理分析。6.3 形態學處理。6.4 文字區塊切割。

6.1 前置處理

前置影像處理包含灰階轉換及高斯低通濾波，目的是將原始彩色影像轉成方便後續處理的灰階影像，並加入高斯濾波器，去除

表 2 統計式與類神經網路比較

系統 特色	統計式文字偵測系統	類神經文字偵測系統
優點	- 統計式文字偵測系統會偵測到較多的文字區塊，所以分割出的文字區塊較完整。 - 執行速度較快，且不需額外的訓練時間。	類神經網路文字偵測系統，對雜訊的容忍度高，較不會將雜訊誤判為文字區塊。
缺點	- 統計式文字偵測系統可偵測到較多的文字區塊，但是也造成較高的誤判機率（將非文字區塊錯判為車牌區塊）。 - 對雜訊容忍度差。若雜訊增多時，執行成果會變的很差。	- 能成功被分割出的文字區塊較不完整，但會有破碎細小不相連的區塊。 - 執行速度慢，相較於傳統式，還額外需要較耗時的訓練過程。

影像中較為複雜且容易誤判成文字區塊的背景。

6.2 邊緣偵測與紋理分析

在上述前置處理階段所處理過後的影像，是一灰階化且經過高斯低通濾波所產生的高斯模糊灰階影像，此影像因高斯濾波遮罩具有對稱性，且非常適合邊緣偵測的特點，故下列我們透過邊緣偵測以及紋理分析

來初步找出類似文字區塊的紋理表現之位置。

表 1 統計式與類神經網路演算法特性比較

能力 \ 演算法	統計式演算法	類神經演算法
處理速度	較快	較慢
容錯能力	較差	較佳
自我學習能力	無	有
訓練時間	無	費時
雜訊影響	較大	較小

本研究影像邊緣偵測分別使用 Sobel 迴旋積分方法與類神經網路方法。邊緣是影像中最基本的特徵，所謂邊緣就是指周圍色度強度有反差變化之像素集合，其為影像分割，紋理分析和影像識別等後序處理的重要基礎。邊緣偵測的方法的種類非常的多，其中 Soble 方法為最多人使用，因為 Soble 是採用影像色度階級跳躍變化的微分運算元，其優點在於邊緣偵測速度快，此方法是將處理後的灰階影像利用 Soble 的方式對像素值作加權後再進行微分的運算，並利用水平濾波器與垂直濾波器分別對應影像像素位置進行分量運算。另外一個模組則是利用類神經網路來達到邊緣偵測的目的，其作法是使用尺寸的輸入視窗當成類神經網路的輸入，經過計算後會留下影像的邊緣。其訓練方法即是先將欲處理的影像經過前置處理得到「訓練影像」與「目標影像」，並設定網路的架構與其它參數，此時網路為「未學習網路」。執行網路訓練過程需設定一收斂值與訓練次數，訓練完後利用若干測試樣本測試網路的執行效果，當網路其精確性不令人滿意時，可以重新調整網路參數或重新訓練網路，以改進網路的精確性。當建立準確度高且令人滿意之網路後，為了進一步確認系統的可靠性，可利用更多的「測試樣本」來測試訓練完成的網路，這些測試樣本既有輸入向量與輸出向量。欲測試一個網路的正確性即是將「測試樣本」的輸入向量輸入到已訓練之網路並計算出網路推論輸出值，再與測試樣本的目標輸出值作比較，以確認系統的可靠

性。當建立準確度令人滿意的網路後，即可應用此網路於「未知資料」的處理。所謂的「未知資料」是指只有輸入向量資料，而無輸出向量資料。將未知資料的輸入向量輸入到已訓練的網路得到其推論輸出值，達到影像邊緣偵測之目的。在統計式文字偵測系統部分，偵測文字時會得到較多的文字區塊，分割文字區塊也較完整，但相較於類神經網路有較高的錯誤偵測。而類神經網路文字偵測系統，偵測文字區塊準確度較高，不易偵測到雜訊，但在分割方面有時會發生分割不完整的情形。為了提高文字偵測的準確率與分割的完整，我們結合了兩個系統的方法以期達到更好的結果。

針對影像高頻的部分也就是邊緣處理後的結果，我們希望利用紋理分析的方法找到文字的區塊，紋理是一種描述物件區塊的重要方法。此階段分為統計式紋理分析與類神經網路紋理分析作討論，兩種方法皆使用了不同大小尺寸的紋理視窗，經由紋理視窗水平掃描影像，將每一個掃描到的區塊像素值做直接的運算或當成類神經網路的輸入值。為了能夠找到影像中不同大小的文字如：場景文字及標題文字，故將原始影像分解成數個不同大小的解析度，此即金字塔紋理分析技術。金字塔紋理分析的處理方式是產生許多低解析度影像，在金字塔紋理分析影像後，最下層影像為原影像尺寸，越上層解析度越低，其降低解析度方式為每上升一層長寬各減為一半，如圖 4 所示。其金字塔上層的影像越容易尋找出大字，相反的金字塔下層之影像越容易尋找出小字。

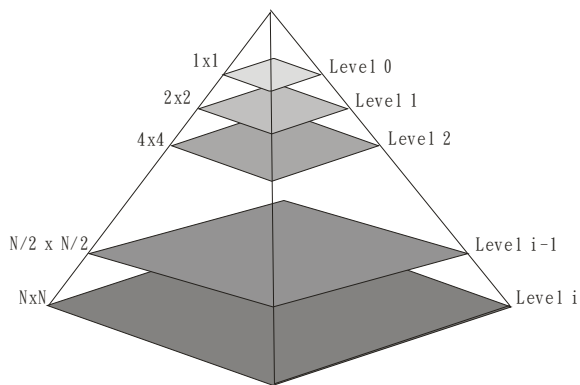


圖 4 金字塔紋理分析架構圖

由於類神經網路紋理分析結合金字塔紋理分析方法流程，是利用金字塔紋理分析後將不同尺寸的影像利用紋理視窗掃描並且個別對應至個別不同架構的類神經網路，接著紋理視窗掃描到的每一個像素值分別作為類神經網路的輸入向量。視窗則以滑動的模式由左至右做水平的掃描影像，視窗大小則分別使用 5×5 、 7×7 、 9×9 像素點三種不同的尺寸作為紋理視窗，使用紋理視窗結合金字塔紋理分析的方法可使影像中大小不一的字體更容易的被偵測到。如圖 5。

不同的紋理經過網路的計算則會有不同的效果，故在紋理視窗的選擇將會使得後續的分割有不同的結果。統計式紋理分析作法則直接對紋理視窗掃描到的像素點作運算，並設一門檻值來衡量是否有文字存在此位置，而紋理視窗的不同對統計式紋理分析也造成影響。本研究採用「十」字型紋理進行研究，此紋理字型在對於新聞影像的處理有較好的文字偵測結果[1]。

6.3 形態學處理

至前一個步驟為止，經過金字塔影像解析處理與紋理視窗比對之後，在不同大小的金字塔紋理分析中會產生許多可能的候選文字區塊，若我們將這些不同大小的候選文字區塊以 AND 或 OR 合併會發現，使用 OR 方法合併後會得到較多的候選文字區塊。然而候選區塊越多表示所找到不同大小的文字越多，但後選區塊越多，相對的錯誤率可能也會隨之提高。相較於 AND 方法作為合併基礎，候選文字區塊影像合併後則會得到較少的文字區塊，由於使用 AND 與 OR 方法合併，兩者結果差異性太大，所呈現出來的效果並不使人滿意，故我們轉而使用三分之二法則，作為第一次合併的方式。

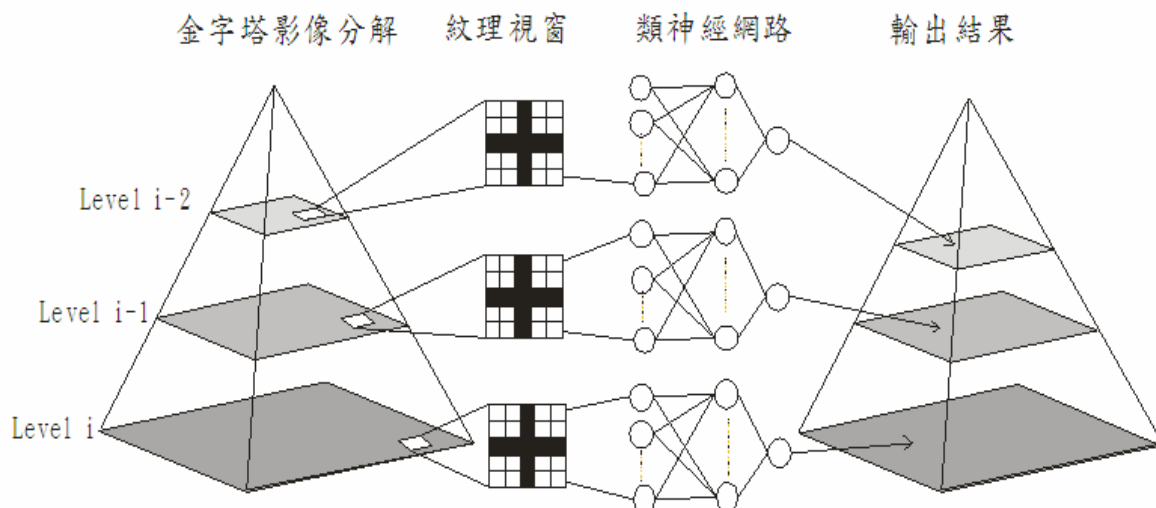


圖 5 金字塔紋理分析示意圖

第一次合併之後的統計式和類神經網路兩種不同演算法中，產生兩種各有其特性的區塊，如表 1。使用傳統式演算法會產生容易

受雜訊影響的區塊，故我們對其作侵蝕處理，去除一些因雜訊而產生的小區塊。而類神經演算法具有強大的容錯能力，不易受雜訊所干擾，但也因其訓練方式較為嚴格，有可能將文字區塊所忽略，且偵測出來的區塊較不連續，故我們對其作膨脹處理，將其區塊完整化。經過侵蝕與膨脹處理後的區塊，在第二次的影像合併中使用 OR 方法，影像合併後再進行下一階段的分割文字區塊。

6.4 文字區塊切割

文字偵測的最後階段為文字區塊切割，此階段是將經過形態學處理並且透過 OR 合併後的二值化影像，利用像素標記法為基礎，分割出影像中包含文字的區塊 [7]。見圖 6。



圖 6 像素標記法文字區塊切割示意圖

7. 文字字元分割

在上述文字偵測系統中已經可以將影像文字區塊偵測出來，我們將在這個部分把文字區塊中的各字元作字元切割的動作。本系統文字字元切割流程分兩階段：7.1 高斯低通濾波處理、7.2 投影量分析。以下對此架構作詳細說明。

7.1 直方圖均等化處理

在文字切割第一階段中，首先將文字偵測系統所偵測出來的文字區塊加入直方圖均等化處理，如此可將已偵測出來的區塊對比

度提高，有助於我們後續使用投影量分析來作字元切割的處理。

7.2 投影量分析

在已偵測的文字區塊中，此研究分割字元的方式是以投影量為分割的基礎。我們先對上階段經過高斯低通濾波處理後的文字區塊進行投影量分析，圖 7 投影量分析圖中 X 軸代表受分析的文字區塊 X 座標位置，Y 軸代表在受分析的文字區塊 X 座標中，像素點灰階值的總合，故在文字呈現亮度較高的情況之下，是文字的部分，其 X 座標的投影量必定較高，而在字元與字元分隔的位置則會呈現投影量極低，甚至為 0 的情況，我們把此位置稱為投影量分析圖上的「峰谷」，以峰谷作為文字切割的切割點。

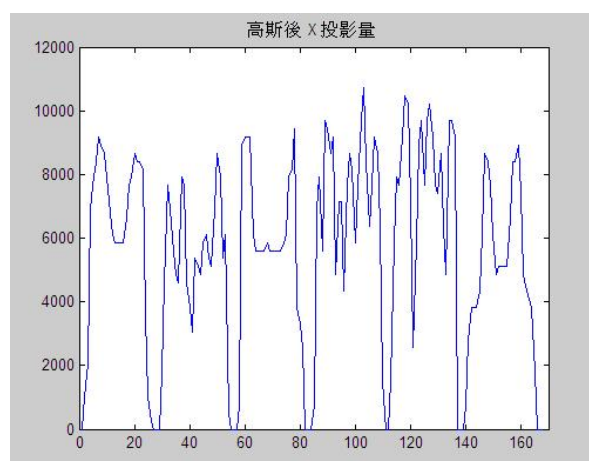


圖 7 投影量分析圖

8. 實驗結果與分析

本研究實驗中最主要的部份為：文字擷取系統兩部分階段實驗。本章節即以細述說明此實驗系統，並對實驗的結果進行討論。此系統根據本研究所提出之架構來撰寫，共分兩個部分，8.1 可自動偵測影像中的文字區塊、8.2 可將偵測出的文字區塊切割其各字元。



圖 8 系統文字偵測介面圖



圖 9 文字偵測結果圖

9. 實驗樣本類型

實驗樣本圖片分為三類型：10(1)單一標題文字新聞影像；10(2)單一標題文字+跑馬燈標題文字新聞影；10(3)大量標題文字新聞影像。見圖 10。

本研究對不同類型的實驗樣本個別作不同臨界值的實驗，以及比較加入高斯低通濾波器後與加入高斯低通濾波器之前的數據，發現在加入高斯低通濾波器後確實能改善本系統之偵測效果。



圖 10 三種不同類型樣本：
左上(1)、右上(2)、下(3)

實驗分類如下：9.1 對單一標題文字新聞影像之實驗，9.2 對單一標題文字+跑馬燈文字新聞影像之實驗，9.3 大量標題文字新聞影像之實驗，9.4 加入高斯低通濾波器後系統改善狀況。

9.1 單一標題文字新聞影像實驗

此樣本屬於偵測難易度較為簡單之實驗樣本，在偵測的實驗樣本群中，並未發現大問題產生，在偵測率、錯誤率以及遺失率都有不錯的結果。表 3 為此實驗數據：

表 3 單一標題文字新聞影像實驗結果

臨 界 值	候 選 文 字 區 塊	正 確 文 字 區 塊	遺 失 文 字 區 塊	錯 誤 文 字 區 塊	實 際 文 字 區 塊	偵 測 率	錯 誤 率	遺 失 率
500	125	110	0	15	110	1.00	0.12	0.00
750	110	107	3	3	110	0.97	0.02	0.02
1000	100	99	11	1	110	0.90	0.01	0.10

在單一標題文字新聞影像的實驗中，各項臨界值的表現都很優異，偵測率都可以達到九成以上，在錯誤率方面，臨界值為 500 時錯誤率較其他偏高，但若考慮到資訊不遺漏的情況之下，臨界值 500 為第一首選，因為其遺失率最低且幾乎可以偵測到所有的文字區塊。



圖 11 單一標題文字新聞影像實驗結果實例

9.2 單一標題文字+跑馬燈文字新聞影像實驗

在此類型研究樣本實驗中，錯誤率較第一類型樣本高，但在偵測率的表現中，各臨界值的效果都不錯，偵測率都在九成六以上。見表 4。

表 4 單一標題文字+跑馬燈文字新聞影像實驗結果

臨 界 值	候 選 文 字 區 塊	正 確 文 字 區 塊	遺 失 文 字 區 塊	錯 誤 文 字 區 塊	實 際 文 字 區 塊	偵 測 率	錯 誤 率	遺 失 率
500	48	25	0	23	25	1.00	0.47	0.00
750	36	24	1	12	25	0.96	0.33	0.04
1000	28	24	1	4	25	0.96	0.14	0.04

在此類型的樣本實驗中，於偵測過程中，會發生文字區塊過大的問題，因標題文字與跑馬燈文字過於接近，促使系統認為成同一區塊，此問題有待於後續文字字元切割階段，再進行修正。見圖 12。



圖 12 單一標題文字+跑馬燈文字新聞影像實驗結果實例

9.3 大量標題文字新聞影像之實驗

此類型樣本為偵測度難易度較高之樣本，樣本影像標題文字大小不一，且偵測結果容易受複雜的背景干擾。

此樣本實驗中，臨界值對偵測率的影響甚大，臨界值過高時偵測率的效果非常不好，雖然錯誤率非常低，但文字區塊的遺失率太高，造成資訊的遺漏，於文字擷取系統的功能來說，並非是一個良好的表現。

故臨界值的考量上，我們選擇臨界值較低的數值 150 為我們系統的設定值，雖然錯誤率有一成五，但是在偵測率與遺失率的表現上是最佳的。見表 4；見圖 12。

表 4 大量標題文字新聞影像之實驗結果

臨 界 值	候 選 文 字 區 塊	正 確 文 字 區 塊	遺 失 文 字 區 塊	錯 誤 文 字 區 塊	實 際 文 字 區 塊	偵 測 率	錯 誤 率	遺 失 率
150	203	172	4	31	176	0.98	0.15	0.02
300	176	164	12	12	176	0.93	0.07	0.07
500	143	140	36	3	176	0.80	0.02	0.20
750	112	111	65	1	176	0.63	0.01	0.37
1000	87	87	89	0	176	0.49	0	0.51

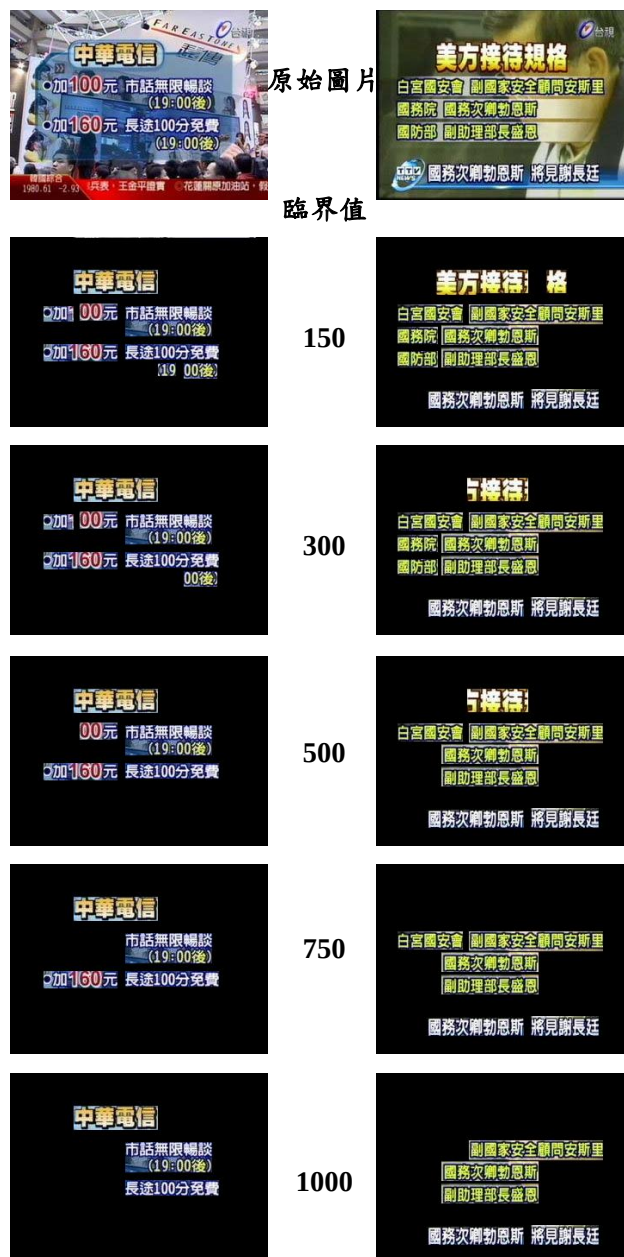


圖 13 大量標題文字新聞影像之實驗結果
實例

未加入高斯低通濾波



臨界值

150

加入高斯低通濾波



300



500



750

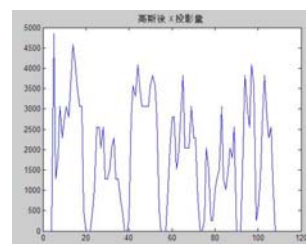


圖 14 高斯低通濾波改善狀況

9.4 加入高斯低通濾波器後系統改善狀況

在此部分中，將以實例說明再加入高斯低通濾波器後，在文字偵測系統的改善效果。

由圖 14 可以觀察出，在未加入高斯低通濾波處理前，文字偵測所切割出來的效果並不理想，偵測出來的區塊很容易受到背景的干擾，連結成一大片的偵測區塊，此結果在後續文字切割的進行中，將會產生很多問題。而在加入高斯低通濾波處理之後，偵測區塊易受背景干擾此問題，便得到大幅的改善。見圖 14。



陳女異常冷靜

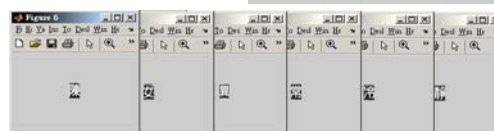


圖 15 文字字元分割結果(1)

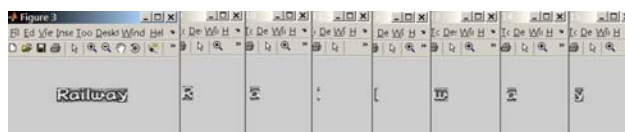
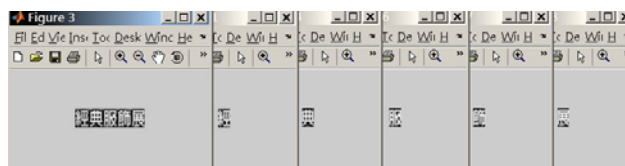
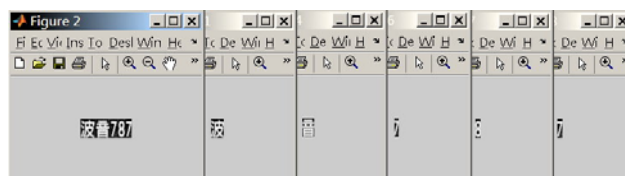


圖 16 文字字元分割結果(2)

10. 文字字元分割實驗

垂直投影量法對於文字切割具有不錯的效果，對於英文字串也能正確的分割出來，但是在某些特別的中文字，如：“位”、“川”，等容易因為垂直投影量的特性而將此種類型的中文字分成兩部份或三部份，如：“位”切割成“人字旁”與“立”等問題，我們就必須加入其他演算法來使字元分割系統更完整。

見圖 15、圖 16。

11. 非文字區塊投影量分析

在本系統第一階段文字偵測的過程中，會偵測到許多非文字的區塊，此類區塊在第二階段的字元切割中，經過投影量分析後的圖形會與正確的文字區塊大不相同。正確的文字區塊具有規律性，峰頂之間的高度並沒有很大的差別，而在非文字區塊的投影量圖中則呈現不規律性，且各峰頂之間的差異很大，依照這些特性，未來我們可以構思出一個演算法來過濾非文字區塊，使我們的系統更完善。見圖 17。

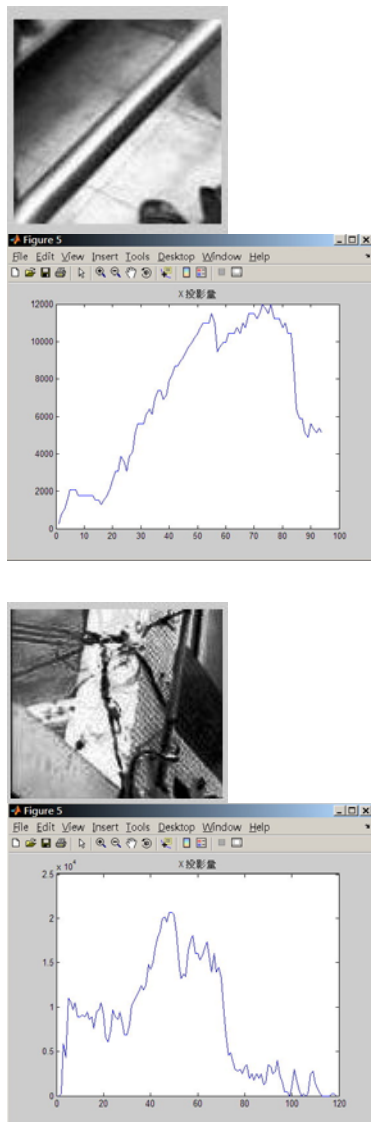


圖 17 非文字區塊投影量圖

12. 結論與未來研究

本研究意在發展一種能從連續影片中自動擷取影像文字的演算法。自動擷取影像文字的演算法可以在連續的影片中擷取影像中文字的資訊，進而將資訊儲存至資料庫中以供查詢。此項功能對於龐大繁雜的影片管理有很大的貢獻，該方法使我們在眾多的影片中能有效率地搜尋到我們需要的影片片段，省下許多搜尋的時間，預估不單只是對資訊搜尋的使用者有很大的助益，對於新聞節目的管理更具改革式的進步。在文字偵測階段，遺失率為文字偵測系統性能好壞最重要的一個指標。遺失率過高會造成此系統無法優化發展，理想狀態中，每一文字區塊都必須被偵測到，才能確保資訊的滴水不漏。能將遺失率控制到最低，對於後續資訊系統資料庫的完整性是相當重要的一個關鍵。在研究中發現，在文字偵測部份加入高斯低通濾波器，改善了複雜背景的干擾，對於破碎的文字字元也能有修復的作用，在實驗中也確實證明了在加入高斯低通濾波器後，對部份類型的樣本有改善文字區塊偵測效果。而在文字分割部份中，未來我們希望能夠尋找到更適合的演算法，藉由投影量字元分割的方法來過濾掉一些錯誤偵測的文字區塊，如此便可在錯誤率高但遺失率低的情況之下，保留更多的重要資訊。本研究發展一套可以自動偵測文字區塊，並進行文字字元分割的文字擷取系統，未來若能與文字辨識系統及資料庫結合，便可以推進影像數位典藏技術的發展。

13. 致謝

本論文承蒙九十五年度國科會研究計畫 NSC95-2221-E-214-065-經費補助，特此致謝。

參考文獻

- [1] J. H. Liu, "A hybrid system for automatic text extraction in images using neural- and statistic-based information processing" *A Thesis Submitted to Institute of Information Management I-Shou University in Partial Fulfillment of the Requirements for the Master Degree in Information Management*, June, 2006.

- [2] S. J. Liang, "The Caption-based Highlight Event Detection and Classification for Baseball Videos" **A Thesis Submitted to Institute of Information Engineering I-Shou University in Partial Fulfillment of the Requirements for the Master Degree in Information Engineering**, June. 2006.
- [3] K. S. Lin, "Caption Extraction and Recognition for Sports Videos" **A Thesis Submitted to Institute of Information Engineering I-Shou University in Partial Fulfillment of the Requirements for the Master Degree in Information Engineering**, June. 2005.
- [4] R. Lienhart, and F. Stuber, "Automatic Text Recognition in Digital Video", in **Proceedings of ACM Multimedia**, pp.11-20,1996.
- [5] R. M. Haralick, and L. G. Shapiro, "Computer and Robot Vision", **Addison-Wesley Publishing Company**, Ch 2.3, pp.28-48, 1992.
- [6] C. K. Chen, "Segmenting License Plate in Color Images", **Thesis for Department of Mechanical & Electro-Mechanical Engineering of National Sun Yat Sen University**, July, 2005
- [7] Q. Zhang, J. Y. Xiang, X. M. Zhao, and B. Sun, "An Approach of Object Description for Image Segmentation", **Infrared Technology**, Nov. 2004.
- [8] 蔡賢亮、劉榕軒,"倒傳遞類神經網路應用於影像文字自動偵測," **電子商務與數位生活研討會**, February 2006。
- [9] 蔡賢亮、林君諺,"類神經網路暨統計式處理之混合式自動影像車牌偵測演算法," **資通技術管理與應用會議**, 2007。