

第一章 緒論

近幾年來，由於影像內容資料量大幅度的增長，如何自動化的來處理這些龐大資料量成為重要的課題。本論文將利用文字萃取的技術來增加影像內容處理的效率，因此於本章節中將介紹本研究相關之研究背景與動機以及研究目的。

第一節 研究背景與動機

對於影片檢索而言，最主要的目標是能方便且快速的找出使用者想要的影片內容。傳統上主要是以人工的方式對影片加入註解以達到進階的搜尋與存取的目的，但這樣的方式不但費時而且非常沒有效率。因此許多的研究及技術不斷被提出以解決這些問題[1]。

就影片內容而言，通常包含影像、聲音及文字資訊，在這其中文字資訊包含了許多關於影像內容意義的關鍵資訊，因此文字資訊萃取技術(Text Information Extraction, TIE)逐漸受到重視，其中包含了三個原因：[2] (1) 影像中的文字區域與非文字區域是很容易被區分的，(2) 文字通常包含了影像中很重要的資訊，(3) 相較於現存的語音分析技術(Speech Analysis Techniques)及物件分析技術(Visual Object Analysis Techniques)，光學文字辨識技術(Optical Character Recognition, OCR)是發展較成熟的技術。

所謂的文字資訊萃取技術主要是將文件分離出文字區及非文字區的一種程序[3]，其中包含了文字偵測(Text Detection)，文字定位(Text Localization)、文字追蹤(Text Tracking)、文字萃取(Text Extraction)以及文字辨認(Recognition)這幾個範疇。在文字抽離方面常會遇到幾個基本的問題，例如：複雜的背景、未知的文字顏色、影像品質不佳以及不同的語言、字型都是造成文字不易從背景中分離的原因[4, 5]。

為了解決上述的問題，在本文章中提出了以獨立成份分析(Independent Component Analysis, ICA)的技術從影像中抽取文字資訊。這項技術是將混合訊號經

由線性或非線性轉換，利用統計學上獨立的原理找出呈現最非高斯分佈的獨立成份[6]。ICA 這個演算法最常被應用於語音訊號分離[7]，它可以成功的從混合的聲音訊號中分離出獨立的聲源，因此本研究依據這個原理，將 ICA 推廣應用於影片中的文字萃取。在論文當中，假設影片中每張影像皆是一組混合訊號，主要是由文字及非文字訊號依照某種線性關係混合而成，本研究欲藉由 ICA 來分析這些訊號，進而估計出影像中的文字區以及非文字區。

第二節 研究目的

本研究希望提出了一個系統化的方法，並且利用此方法處理大部分的影像特徵，達到一個自動化的文字萃取機制，解決傳統上費時且沒效率的人工方式，並且降低一般 TIE 技術中複雜的處理程序。本論文欲以 ICA 方法來分析影像，並萃取出影片中的字幕，即使影片的解析度不佳，或是背景複雜，還是在文字字型、尺寸、顏色、及排列方式不同的情況下，都可以成功的將影像中的文字從背景中分離，因此本研究最主要的目的為：

1. 發展出一個系統化的文字萃取架構。
2. 降低一般文字資訊萃取技術中複雜的處理程序。
3. 建構強韌性佳、具有效率並且適合影片字幕的文字萃取方法。