

## 以獨立成份分析法進行影片字幕粹取

Using Independent Component Analysis for Video Subtitle Extraction

\* 周遵儒 (Tzren-Ru Chou), \*\* 謝珮琳 (Pe-Lin Hsieh)

### 摘要

在本論文中，我們針對影片中的字幕提出一個創新的文字粹取方法，在此，我們利用獨立成份分析法(Independent Component Analysis, ICA)來分離影像中的文字區及非文字區。為了符合ICA的處理程序，我們提出了兩種資料的表示方法，分別為「片段特徵表示方式」及「單張影像特徵表示方式」。為了分析本方法的強韌性(robustness)，我們改變了字幕中文字的字型、顏色、尺寸以及排列方式。本研究採用影像的灰階強度做為特徵，實驗結果顯示本研究所提出的方法比傳統的方法能更有效率且更正確分離出影像中的文字區域及非文字區域。

### Abstract

In this paper, a new text extraction method is proposed for digital video. The text in the video frames is extracted by means of the independent component analysis (ICA). Two representations of the image frames are introduced to fit the need of the ICA procedure. The font, color, scale, and orientation of the subtitles in the sample videos are changed to evaluate the performance of our method. We use the gray values as the features. The experimental results show that the method we developed in this research can extract the texts in digital video more efficiently and accurately than the conventional work, without any influence on the variations of font, color, scale, and orientation of texts.

關鍵詞：文字粹取、獨立成份分析法、影片檢索

Keywords: text extraction, independent component analysis, video indexing

---

\*周遵儒：國立台灣師範大學圖文傳播學系副教授/ National Taiwan Normal University, Department of Graphic Arts Communication, Associate Professor

\*\*謝珮琳：國立台灣師範大學圖文傳播學系碩士研究生/ National Taiwan Normal University, Department of Graphic Arts Communication, Graduate Student

# 目 錄

壹、緒論.....	60
貳、獨立成份分析.....	60
參、影片資料表示方法.....	62
肆、實驗結果.....	63
伍、結論.....	66

## 壹、緒論

近幾年來，由於影像內容資料量大幅度的增長，如何更有效率且自動化處理這些龐大資料量成為重要的課題。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，因此文字資訊粹取技術(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。

所謂的文字資訊粹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)，文字的定位(Text Location)、文字的追蹤(Text Tracking)以及文字的粹取(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]。根據上述的問題，在本文章中提出了獨立成份分析(Independent

Component Analysis, ICA)的技術從影像中抽取文字資訊。這項技術是一種線性轉換的方式，主要是由統計學上獨立的原理找出非高斯分佈的獨立成分[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。在本研究中，我們假設影片中每張影像皆是由文字及非文字訊號依照某種線性關係混合而成，並且嘗試以 ICA 的技術來抽離影像中之文字。

## 貳、獨立成份分析(Independent Component Analysis, ICA)

獨立成份分析(以下稱之為 ICA)是一種統計及計算的技術，主要是用來找出隨機變數或訊號中的隱藏因子(hidden factors)。在 ICA 的模型中(ICA model)，假設觀察的多變量資料是由某些未知的潛在變數(latent variable)經由線性或非線性的方法混合而成[8]。這些潛在的變數被稱之為觀察資料中的獨立成份(Independent Component)，並且假設他們為非高斯分佈且互相獨立。而 ICA 就是利用分析的方法找出這些獨立成份。以聲音訊號分離為例，在公式(1)中有  $m$  支麥克風，從空間中接收到  $m$  個混合的訊號  $x_1, x_2, x_3, \dots, x_m$ ，假設這  $m$  個混合的訊號是由  $n$  個統計上獨立的聲源訊號  $s_1, s_2, s_3, \dots, s_n$  經由線性關係所混合而成的。

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = AS \quad (1)$$

其中 A 為混合矩陣， $a_{ij}$  代表第 j 個聲源在第 i 個麥克風的權重。ICA 技術就是在只有混合訊號 X 已知的情況下，去找一個矩陣 W (demixing matrix) 經過轉換後，得到原來的聲源訊號(見公式(2))。

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \approx S \quad (2)$$

利用訊號與訊號之間獨立的關係，經由 Y 中每個訊號是否獨立來決定 W，如果此 W 轉換出來的 Y 是呈現最不高斯分佈的，就是原來彼此獨立的訊號源，這就是 ICA 的基本觀念，下圖 1 為 ICA 的基本處理程序。

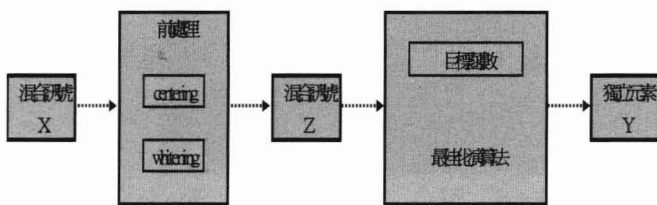


圖 1 ICA 演算法之架構

在 ICA 的演算法中，通常會先做 centering(公式(3)) 與 whitening(公式(4))兩個前處理，此兩個前處理的動作可以使接收到的混合訊號與獨立聲源的平均為零(zero-mean)，而經由 whitening matrix V 的線性轉換之後使訊號變的更 uncorrelated 如此一來可大大簡化 ICA 演算法。其中  $E\{x\}$  為混合訊號的平均值。

$$X_{center} = x - E\{x\} \quad (3)$$

$$Z = VX \quad (4)$$

接下來，根據 Aapo Hyvärinen 所提出之 FastICA 演算法，使用 neg-entropy 為目標函數(object function)，藉此做為 non-Gaussian 的量測，假設量測之  $y = w^T z$ ，從公式(5)中，若  $J(w)$  越大，代表所分析出來的 y 越 non-Gaussian [6]。

$$J(w) \propto [E\{G(w^T z)\} - E\{G(v)\}]^2 \quad (5)$$

因此，我們若是要找出最 non-Gaussian 的分佈，可藉由尋找  $E\{G(w^T z)\}$  之極值。其中 G 為 non-quadratic function 有以下三種選擇(公式(6), (7), (8))，而 v 為 Gaussian random variable。

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \quad (6)$$

$$G_2(y) = -\exp(-y^2 / 2) \quad (7)$$

$$G_3(y) = y^4 \quad (8)$$

其中  $1 \leq a_i \leq 2$ 。

最後根據 fixed-point iteration(公式(9)) [6]，經過幾次疊代後，我們就可以取得使 y 呈現最非高斯分佈的 w，藉此得到獨立元素 (Independent Component)。

$$\begin{aligned} w &\leftarrow E\{zg(w^T z)\} / E\{g'(w^T z)\}w \\ w &\leftarrow w / \|w\| \end{aligned} \quad (9)$$

本研究最主要是希望藉由 ICA 之技術來達到自動化的文字粹取，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，即使影片在文字字型、尺寸、顏色、及排列方式的變化下都可以成功的將影像中的文字從背景中分離。

在本研究中，我們假設所有的影片畫面皆是由文字  $s_{ik}$  及非文字訊號  $s_{nk}$  經由混合矩陣 A 依照某種線性關係混合而成，如公式(10)所示。其中  $x_{ij}$  為第 i 個 image frame 的第 j 個灰階值。本計畫

欲以公式(10)為基本構想，利用 ICA 方法分離影像中的文字區及非文字區。

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1\gamma} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2\gamma} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3\gamma} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{N\gamma} \end{bmatrix}_{N \times \gamma} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ \vdots & \vdots \\ a_{N1} & a_{N2} \end{bmatrix}_{N \times 2} \begin{bmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1\gamma} \\ s_{21} & s_{22} & s_{23} & \cdots & s_{2\gamma} \\ s_{31} & s_{32} & s_{33} & \cdots & s_{3\gamma} \\ \vdots & \vdots & \vdots & & \vdots \\ s_{N1} & s_{N2} & s_{N3} & \cdots & s_{N\gamma} \end{bmatrix}_{2 \times \gamma} \quad (10)$$

### 參、影片資料表示方法

在本研究中，我們預計以 TIE 技術利用 ICA 的方式來處理影片片段以及影片中的單一影像。為了符合 ICA 的處理程序，在此我們提出了兩種資料的表示方法，分別為「片段特徵表示方式」及「單張影像特徵表示方式」。

#### 一、片段特徵表示方式

在這個方法中，我們首先會選取一段有相同字幕的影片片段做為測試圖片，然後將此段影片切割成一張張的影像(image frame)。再從每張影像中抽取出其灰階值，經過 ICA 的運算後分離出字幕與影像兩個成份。在此先把彩色影像轉為灰階影像，再將 2 維的影像訊號轉為 1 維訊號，如公式(11)所示

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1\gamma} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2\gamma} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3\gamma} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{N\gamma} \end{bmatrix}_{N \times \gamma} \quad (11)$$

在公式(11)中有 N 張 image frame，而這 N 張影像擁有相同的字幕內容，我們假設此 N 張影像是由「字幕成分」及「影像成分」根據不同的權重，線性混合而成，其中  $x_{ij}$  是第 i 張影像中第 j 個灰階值，而  $\gamma$  是一張影像中像數(pixel)的總數。

#### 二、單張影像特徵表示方式

在 ICA 這個方法中有一個非常重要的假設，就是混合的成分(mixed component) — 在本研究中指的是一張張的 image frame，其數目不得小於欲求的獨立成分，但在本實驗中，我們至少要找出兩種獨立成分，分別為文字成份以及非文字成份，換句話說就是我們無法將 ICA 這個方法用在單張的影像上。為了解決這個問題，我們必需將輸入的單一變數轉換成多變數的輸出。以色彩特徵而言，一張影像中的每一個 pixel 都可以被視為由 R, G, B 三種訊號依據不同權重混合而成，因此我們可以將影像的 R, G, B 三層視為三張灰階影像(見圖 2)。如此一來我們可以得出單張影像的表示方法，如公式(12)所示，其中  $X_{Rk}, X_{Gk}, X_{Bk}$ ，分別表示影像 R, G, B 三層中的第 k 個特徵。

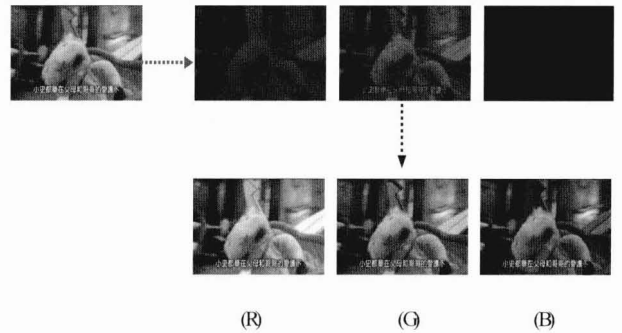


圖 2 將彩色影像視為 R, G, B 三張灰階影像的輸出

$$X = \begin{bmatrix} x_{R1} & x_{R2} & x_{R3} & \cdots & x_{R\gamma} \\ x_{G1} & x_{G2} & x_{G3} & \cdots & x_{G\gamma} \\ x_{B1} & x_{B2} & x_{B3} & \cdots & x_{B\gamma} \end{bmatrix}_{3 \times \gamma} \quad (12)$$

經由上述兩種對於影像的表示方式(公式(11), (12))，我們可以根據公式(10)，經由觀察影像 X 中的變數來估計混合矩陣 A，再計算出兩獨立成分 — 影像中的文字區域及非文字區域

## 肆、實驗結果

在此實驗中我們準備如圖 3 中具有相同字幕的一段影片。再從其中隨機取出 2, 3, 4, 5, 10, 20, 30, 40, 50 張影像當作測試的樣本。然後再使用

ICA 的方法來執行文字的粹取，我們將粹取的結果與簡易的 coefficient of variation (CV) 做比較(見圖 4)。



圖 3 含有相同字幕的測試影片片段


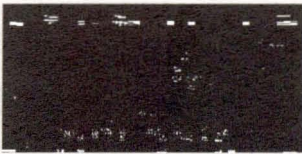


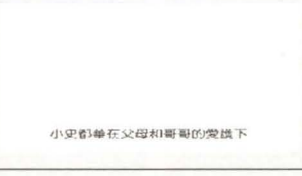

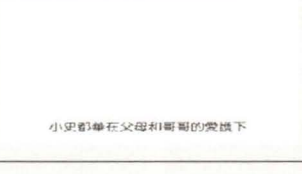

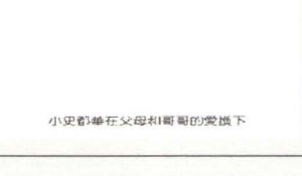
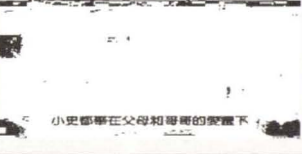
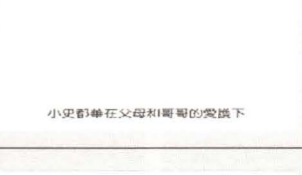
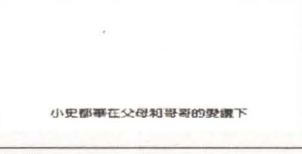
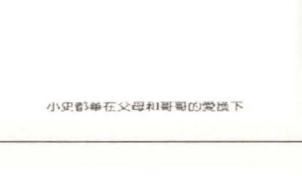
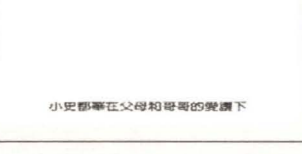
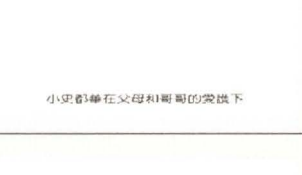
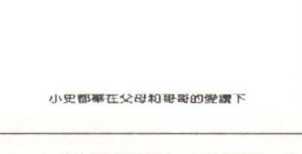
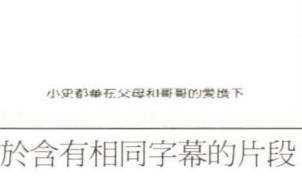
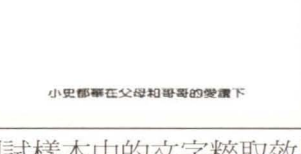
測試樣本數	ICA 方法	CV 方法
2		
3		
4		
5		
10		
20		
30		
40		
50		

圖 4 CV 與 ICA 對於含有相同字幕的片段，在不同測試樣本中的文字粹取效果

從圖 4 中我們看到 CV 方法對於文字粹取的效果取決於影像的樣本數目，在樣本數夠多時 CV 的確有很好的效果，但若是樣本數不夠多的情況下，CV 的效果並不理想。相較之下，ICA 縱使在樣本數很少的情況下，還是能成功的抽離背景中的文字。

再來我們將嘗試從單張影像中粹取出文字。我們從圖 3 中任意抽出一張 image frame，然

後將此張 RGB 的彩色影像輸出成三張灰階影像並根據公式(10)的方式帶入 ICA 公式中計算。爲了分析我們提出的 ICA 方法是否夠強韌 (robustness) 我們改變字幕的字型、顏色、尺寸以及排列方式，如圖 5 所示。

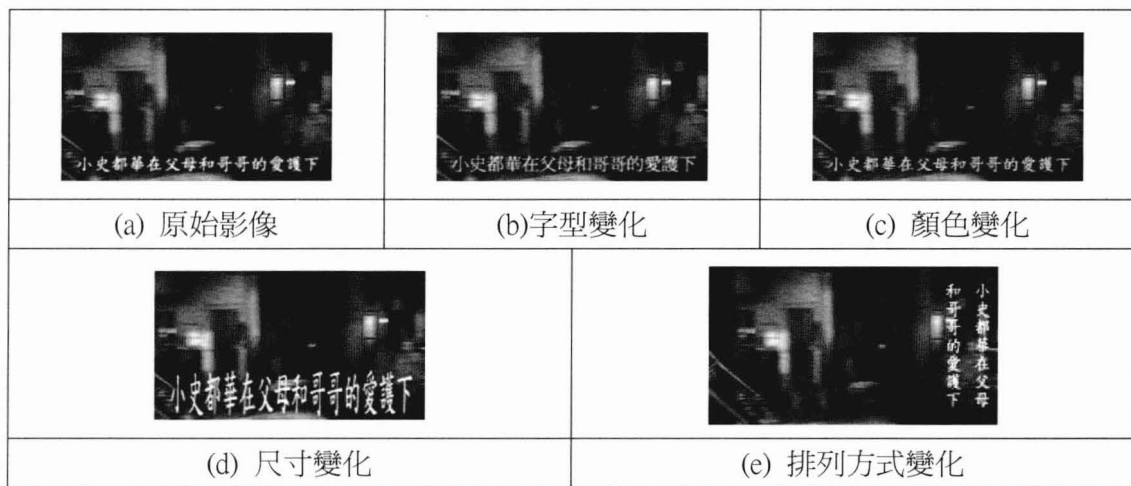


圖 5 變化各種不同字幕以測試本系統的強韌程度

從圖 6 中，我們可以看到 CV 方法無法很清楚地分離文字及背景區域，而其對於文字顏色的變化

是非常敏感的。相對於 CV，ICA 的方法在各種變化之下還是能從背景中抽取出文字。

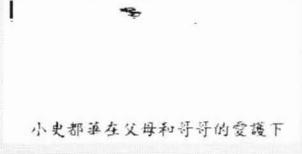
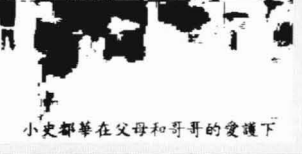

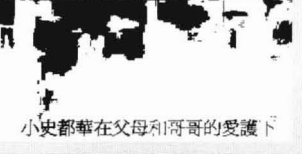
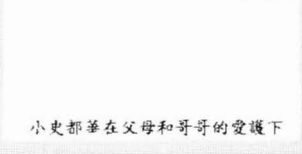

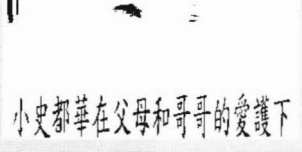
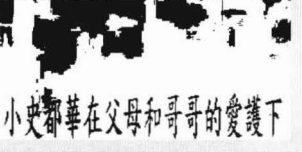


methods changes	The ICA method	The CV method
Original	 小史都華在父母和哥哥的愛護下	 小史都華在父母和哥哥的愛護下
Font	 小史都華在父母和哥哥的愛護下	 小史都華在父母和哥哥的愛護下
Color	 小史都華在父母和哥哥的愛護下	 小史都華在父母和哥哥的愛護下
Scale	 小史都華在父母和哥哥的愛護下	 小史都華在父母和哥哥的愛護下
Orientation	 小史都華在父母和哥哥的愛護下	 小史都華在父母和哥哥的愛護下

圖 6 CV 與 ICA 對於單張影像，在不同文字變化中的文字粹取效果

## 伍、結論

在本研究中，我們利用 ICA 的方法進行影片中字幕的粹取。為符合 ICA 的處理程序，在此提出兩種資料表示的表示方法，分別為「片段特徵表示方式」及「單張影像特徵表示方式」。實驗結果顯示，本研究所提出之方法即使在 image frame 數目很少時甚至是單張影像上，對於字幕粹取都有很好的效果。相較於傳統的文字粹取方式，本方法不需要很繁複的處理程序，並且在字幕字型、顏色、尺寸或排列方式的變化下都可以成功的分離文字區及非文字區。

### 參考文獻

1. J. C. Shim, C. Dorai, and R. Bolle, "Automatic

Text Extraction from Video for Content-Based Annotation and Retrieval," Proceedings of International Conference on Pattern Recognition, Vol. 1, pp. 618-620, 1998.

2. M. R. Lyu, J. Song, and M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction," IEEE transactions on circuits and systems for video technology, Vol. 15, No. 2, pp. 243, Feb. 2005.

3. K. Jung, K. I. Kim, and A. K. Jain, "Text Information Extraction in Images and Video: A Survey", Pattern Recognition, Vol. 37, pp. 977-997, 2004.

4. M. Cai, J. Song, and M. R. Lyu, "A New Approach for Video Text Detection," Proceeding 2002 International Conference, Vol. 1, pp. 22-25, Sept. 2002.
5. Q. Ye, W. Gao, W. Wang, W. Zeng, "A Robust Text Detection Algorithm in Images and Video Frames," Proceedings of the 2003 Joint Conference of the Fourth International Conference, Vol. 2, pp. 15-18, Dec. 2003.
6. A. Hyvärinen and E. Oja, "Independent Component Analysis: Algorithm and Applications," Neural Networks, Vol. 13, pp. 411-430, 2000.
7. J. Cardoso, "Informax and Maximum Likelihood for Source Separation," IEEE Letters on Signal Processing, Vol. 4, No. 4, pp. 112-114, Apr. 1997
8. P. Comon, "Independent Component Analysis - A New Concept," Signal Processing, Vol. 36, No. 3, pp. 287-314, 1994