

第三章 ICA 文字萃取方法

在上述的文獻探討中，說明了 ICA 相關的演算法則，而此章節中，將介紹本研究如何把 ICA 的演算法則運用於影像當中，並且利用 ICA 可以純化混合影像的能力，進階的萃取出影像中的字幕區域。

第一節 符合 ICA 方法的影像假設與特徵表示法

在本研究中，假設在一連串的畫面中，每一張影片畫面(image frame)就是一組混合訊號 x ，此訊號是經由文字成份 s_T 及非文字成份 s_n 經由線性關係混合而成(見圖 3.1)。



圖 3.1 假設所有的影片畫面都是一組混合訊號 x ，經由文字成份 s_T 及非文字成份 s_n 混合而成

有了上述的假設，便可建構出適用於文字影像萃取的 ICA 模型，如公式 3.1 所示。公式中說明每張影片畫面中的每個特徵值 x_{ij} 皆是由文字成份 s_{Tj} 及非文字成份 s_{nj} 經由混合權重 a_{ik} 線性混合而成。其中 x_{ij} 為第 i 張影像的第 j 個特徵，共有 N 張影像，每張影像各有 γ 個特徵值。

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1\gamma} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2\gamma} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3\gamma} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{N\gamma} \end{bmatrix}_{N \times \gamma} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ \vdots & \vdots \\ a_{N1} & a_{N2} \end{bmatrix}_{N \times 2} \begin{bmatrix} s_{T1} & s_{T2} & s_{T3} & \cdots & s_{T\gamma} \\ \vdots & \vdots & \vdots & & \vdots \\ s_{n1} & s_{n2} & s_{n3} & \cdots & s_{n\gamma} \end{bmatrix}_{2 \times \gamma} \quad (3.1)$$

從文獻當中可以得知，ICA 主要是在只有混合訊號 \mathbf{x} 已知的情況下，去找一個矩陣 \mathbf{w} ，經過轉換後可得到原本獨立的訊號。換句話說，只要能夠將每張影片畫面的每個特徵值表示的像公式 3.1 中 x_{ij} 的形式，便可透過 ICA 處理程序計算出文字成份及非文字成份。因此在下一個段落中，將依據測試影片畫面的特性，發展出 3 種不同的特徵表示法。

壹、針對 ICA 的影像特徵表示法

在本研究中，預計利用 ICA 的方式來處理影片中的連續畫面以及影片中的單一影像。在此，提出了 3 種資料的表示方法，其中有針對連續影像的「序列特徵表示法」。另外，若是輸入值為單張影像，就必須要製作多張影像的輸出，如此一來才能符合 ICA 的處理程序，而此部份的細節將於「rgb 特徵表示法」與「高次特徵表示法」兩個段落中在做說明。根據上述說明，本研究為 ICA 文字萃取系統設計出如圖 3.2 的處理流程。

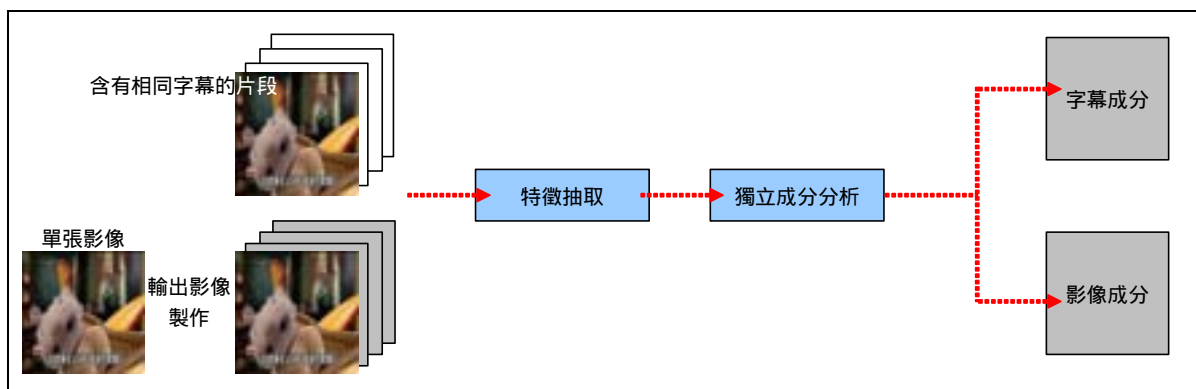


圖 3.2 ICA 文字萃取系統處理流程架構圖

一、 序列特徵表示法

在這個方法中，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片畫面 I_o 。接下來，如公式 3.2 所示，將每張彩色影像轉為灰階影像 I_{gray} 。

$$\begin{aligned}
 & k = 1, 2, \dots, N \\
 \mathbf{I}_{gray}^{(k)}(i, j) &= (\mathbf{I}_r^{(k)}(i, j) + \mathbf{I}_g^{(k)}(i, j) + \mathbf{I}_b^{(k)}(i, j)) / 3 & i = 1, 2, \dots, h & \quad (3.2) \\
 & j = 1, 2, \dots, w
 \end{aligned}$$

其中有 N 張影像，而 $\mathbf{I}_r^{(k)}(i, j)$, $\mathbf{I}_g^{(k)}(i, j)$, $\mathbf{I}_b^{(k)}(i, j)$ 分別為第 k 張影像 $\mathbf{I}_o^{(k)}(i, j)$ 的 r, g, b 三個色頻， h 與 w 分別為影像的高度及寬度。最後把 2 維的影像訊號 $\mathbf{I}_{gray}^{(k)}$ 轉為 1 維訊號，如此一來便可得到混合的訊號 \mathbf{x}_{cont} (見公式 3.3)。

$$\mathbf{x}_{cont} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1\gamma} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2\gamma} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3\gamma} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{N\gamma} \end{bmatrix}_{N \times \gamma} \quad (3.3)$$

在公式 3.3 中有 N 張影像，而這 N 張影像擁有相同的字幕內容，其中 x_{ij} 是第 i 張影像中第 j 個畫素的灰階強度值，而 γ 是一張影像中畫素的總數。若影像的尺寸為 128×128 pixel，則 $\gamma = 16384$ 。

二、 rgb 影像特徵表示法

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影片畫面 I_o ，其數目不得小於欲求的獨立成份(詳見文獻探討：ICA 之限制)，但在本實驗中，至少要找出兩種獨立成份，分別為文字成份以及非文字成份，換句話說就是無法將 ICA 這個方法用在單張的影像上。為

了解決這個問題，就必需將輸入的單一影像轉換成多張影像的輸出。

以色彩特徵而言，一張影像中的每一個畫素值都可以被視為由 r, g, b 三種訊號依據不同權重混合而成，因此可以將影像 I_0 的 I_r, I_g, I_b 三個色頻視為三張灰階影像 (如圖 3.3)。同樣的將 2 維的影像訊號 I_r, I_g, I_b 轉換成 1 維訊號，就可得到如公式 3.4 的混合訊號 x_{rgb} ，其中 x_{rk}, x_{gk}, x_{bk} 分別表示影像 I_r, I_g, I_b 中的第 k 個畫素的灰階值，每張影像總共有 γ 個畫素。

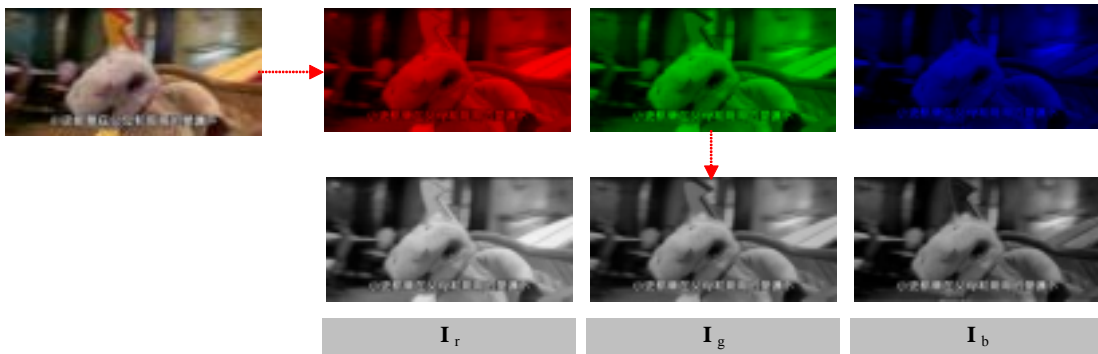


圖 3.3 影像的 r, g, b 三層視可被視為三張灰階影像

$$\mathbf{x}_{rgb} = \begin{bmatrix} x_{r1} & x_{r2} & x_{r3} & \cdots & x_{r\gamma} \\ x_{g1} & x_{g2} & x_{g3} & \cdots & x_{g\gamma} \\ x_{b1} & x_{b2} & x_{b3} & \cdots & x_{b\gamma} \end{bmatrix}_{3 \times \gamma} \quad (3.4)$$

三、 高次影像特徵表示法

在本研究中，希望能儘可能的模擬由文字成份及非文字成份所混合而成的混合影像，在高次特徵表示法中，除了 I_r, I_g, I_b 三種影像之外，本方法更進階的考慮影像中每個畫素與它周圍畫素間的關係，以上、下、左、右的差異再製造出 4 張混合影像 I_T, I_B, I_L, I_R 。為了能從單張混合影像中製造出更多的輸出影像，本研究發展出影像多次的混合法則，如公式 3.5 與公式 3.6 分別表示 2 次與 3 次的混合影像。

$$\mathbf{I}_{s,t}^{(2)}(i, j) = \mathbf{I}_s(i, j) * \mathbf{I}_t(i, j) \quad \begin{array}{l} \text{for } s, t \in \{r, g, b, T, B, L, R\}, \\ \text{and } s \neq t \end{array} \quad (3.5)$$

$$\mathbf{I}_{s,t,u}^{(3)}(i, j) = \mathbf{I}_s(i, j) * \mathbf{I}_t(i, j) * \mathbf{I}_u(i, j) \quad \begin{array}{l} \text{for } s, t, u \in \{r, g, b, T, B, L, R\}, \\ \text{and } s \neq t \neq u \end{array} \quad (3.6)$$

同樣的 $\mathbf{I}_r, \mathbf{I}_g, \mathbf{I}_b$ 分別為影像中三個色頻， $\mathbf{I}_T, \mathbf{I}_B, \mathbf{I}_L, \mathbf{I}_R$ 如以下公式 3.7，其中而 \mathbf{I}_{gray} 如公式 3.2 所示。

$$\begin{aligned} \mathbf{I}_T(i, j) &= \mathbf{I}_{\text{gray}}(i, j) - \mathbf{I}_{\text{gray}}(i-1, j) \\ \mathbf{I}_B(i, j) &= \mathbf{I}_{\text{gray}}(i, j) - \mathbf{I}_{\text{gray}}(i+1, j) \\ \mathbf{I}_L(i, j) &= \mathbf{I}_{\text{gray}}(i, j) - \mathbf{I}_{\text{gray}}(i, j-1) \\ \mathbf{I}_R(i, j) &= \mathbf{I}_{\text{gray}}(i, j) - \mathbf{I}_{\text{gray}}(i, j+1) \end{aligned} \quad i = 2, 3, \dots, h-1, j = 2, 3, \dots, w-1 \quad (3.7)$$

得到高次的混合影像之後，再利用公式 3.8 將訊號正規化至 0 到 255 之間，其中 $\mathbf{I}^{(m)}$ 表示 m 次的混合影像。

$$\mathbf{I}' = [\mathbf{I}^{(m)} - \min(\mathbf{I}^{(m)}) / \max(\mathbf{I}^{(m)}) - \min(\mathbf{I}^{(m)})] * 255 \quad (3.8)$$

在本研究中，共採用 2 次及 3 次正規化後的混合影像 \mathbf{I}' ，因此總共能得到 56 張的混合影像，之後將 2 維訊號轉換至 1 維訊號便可得到如公式 3.9 混合訊號 \mathbf{x}_{high} 。

$$\mathbf{x}_{\text{high}} = \begin{bmatrix} x_{rg,1} & x_{rg,2} & x_{rg,3} & \cdots & x_{rg,k} \\ x_{gb,1} & x_{gb,2} & x_{gb,3} & \cdots & x_{gb,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{LR,1} & x_{LR,2} & x_{LR,3} & \cdots & x_{LR,k} \\ x_{rgb,1} & x_{rgb,2} & x_{rgb,3} & \cdots & x_{rgb,k} \\ x_{gbT,1} & x_{gbT,2} & x_{gbT,3} & \cdots & x_{gbT,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{BLR,1} & x_{BLR,2} & x_{BLR,3} & \cdots & x_{BLR,k} \end{bmatrix}_{56 \times k} \quad (3.9)$$

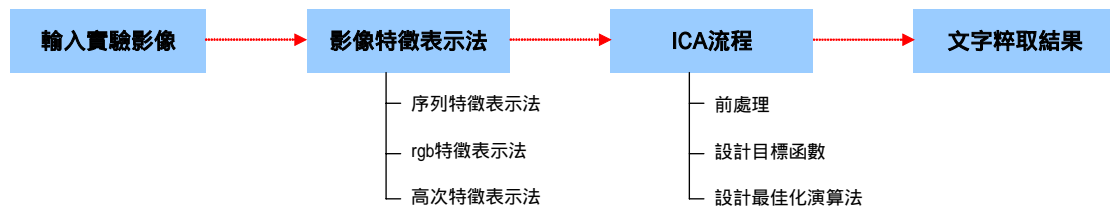
經由上述 3 種對於影像的表示方式，可以根據公式 2.34，經由觀察混合影像 \mathbf{x} 中的變數來估計混合矩陣 \mathbf{w} ，再計算出兩獨立成份 文字成份及非文字成份。

第二節 針對文字萃取的 ICA 方法與傳統方法簡介

本研究的結果將與前述文獻[14]中，由 Lienhart 等人所提出的 MoCA 方法，以及統計學上的變異係數法(coefficient of variation, CV)做為實驗效果的比較對象。以下將簡單介紹 ICA、MoCA 及 CV 三種方法的基本處理程序。

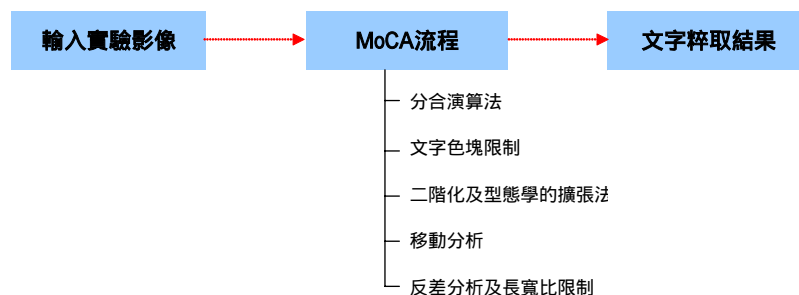
壹、ICA 方法

針對文字萃取實驗，ICA 的處理流程如下圖 3.4 所示：



貳、MoCA 方法

從文獻中可得知 MoCA 方法是經由結合很多技術以達到文字萃取的效果，其中包括：分合演算法對影像做分割、文字色塊大小限制、二階化與型態學的擴張法、移動分析結果以及反差分析與長寬比限制，下圖 3.5 中，顯示 MoCA 方法的處理程序。



參、CV 方法

變異係數的方法，是 Karl Pearson 所發展出來的一個衡量的方法[41]。主要是以標準差 除以平均數 μ ，如公式 3.10。一但變異係數 CV 小於某個門檻值，此特徵值便被歸類為文字區，反之，若是 CV 大於，則被歸類為非文字區。

$$CV_{i,j} = \frac{\sigma_{i,j}}{\mu_{i,j}}$$

$$\sigma_{i,j} = \sqrt{\frac{\sum_{k=1}^N (x_{i,j}^{(k)} - \mu_{i,j})^2}{N}}$$

$$\mu_{i,j} = \frac{\sum_{k=1}^N x_{i,j}^{(k)}}{N}$$

$$\begin{aligned} & \text{if } CV_{ij} < \tau, \quad x_{ij} = 1 \\ & \text{if } CV_{ij} > \tau, \quad x_{ij} = 0 \end{aligned} \quad (3.10)$$

公式 3.10 中顯示有 N 張測試的影像，其中 $x_{ij}^{(k)}$ 表示第 k 張影像中(i, j)位置的特徵值。下圖 3.6 中顯示 CV 方法的文字萃取程序。

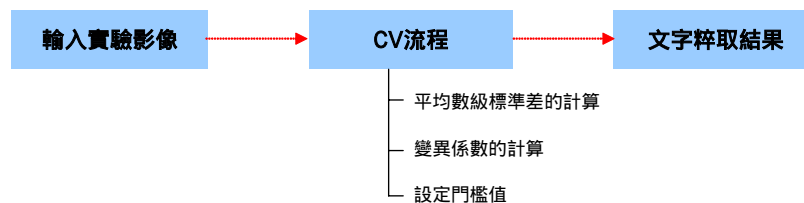


圖 3.6 CV 方法的處理流程