

文件影像檔的高精確度偏斜估算

葉長茂¹ 邱創標² 鄭合志³

¹ 中州科技大學 yeh120@dragon.ccut.edu.tw

² 中州科技大學 ccbiau@dragon.ccut.edu.tw

³ 中州科技大學 hccheng@dragon.ccut.edu.tw

摘要

本文提出了一個新的文件影像檔的傾斜角度估計算法，文件影像檔基於 FCRM（模糊 c 回歸模型）的聚類方法，目的是解決現有方法的低精度和低強健性的缺點。該演算法包括以下三個步驟。第一步驟為將所述輸入圖像通過圖像分析，建立平行的直線。第二個步驟選擇重要區域，加速執行時間及使用快速插值的圖像放大，提高準確度的偏斜角估計。最後，使用 FCRM 方法估計偏斜角。使用 184 張的不同類型的文件影像檔的測試，測量該演算法的性能。實驗結果證實該方法在不同的類型文件具有很高的精度度，它能夠準確地估計範圍從-89°到+89°之間的傾斜角度。

關鍵詞:FCRM、影像處理



壹、緒論

因基於紙張的文件管理，歸檔，檢索和許多其他應用佔用相當大的空間，難以應對各種複雜的數據挖掘，所以紙張文件轉換成電子記錄文件相形重要。在紙張文件轉換成電子文件系統有以下階段，掃描，區域分割，文字識別，圖像增強，偏差的估計和校正。其中在掃描過程中，可能因人為或機器的因素，產生圖像傾斜角度，從而降低了精度分析或 OCR 文件。因此，歪斜文件分析及估計是一個非常重要的階段。

在過去的幾十年中已提出了許多傾斜角度估計方法，這些方法可分為四種類型：輪廓投影法[1]-[5]，鄰近聚類法[6]-[8]，霍夫變換法[9]-[12]，和互相關法[13]-[14]。postl [1]提出一種計算文件檔的輪廓投影的傅立葉變換，然後從傅立葉空間中的密度獲得偏斜角。Hashizume 等人描述鄰近聚類方法的歪斜估計[6]，發現所有鄰近的連接組件，在直方圖中累積所有同樣的方向的向量，直方圖中的峰值中到估測文件歪斜，這兩種方法的可識別角度範圍受到很大的限制。Hinds and Colleagues [9]提出了霍夫變換的偏移估計方法，從圖像的灰度建立垂直文，然後確定圖像的 Hough 變換，為了提高 Hough 變換方法的準確度[12]，平滑和定位的黑白的強化文件檔，這種方法具有較高的計算複雜性和較高計算成本。Yan [13]提出了隔行掃描的交叉關聯整個文件檔的互相關法，這種方法也非常費時，文件檔包含的圖形和字符的內部筆劃常造成不是很準確。最近，Amin 等人[15]提出另一種偏移的估計方法，通過以下方式獲得旋轉具有一定程度的原始圖像的子圖像的邊界框之內的背景像素的總數。偏斜角計算的背景像素的原始圖像的旋轉角度的像素的總數。這種偏斜檢測方法，可以為圖形或圖像區域可能沒有檢測的傾斜角度（例如，一個正五邊形或圓）。這些方法能有效估計傾斜角度，但在實際的應用中，常受到雜訊而影響偏斜角估計的準確性。

本研究運用的模糊 c 回歸模型 (FCRM) 以改善歪斜的估計精度。偏斜角估計算法包括以下三個步驟，第一步驟為將所述輸入圖像通過圖像分析，建立平行的直線，以產生平行的直線的圖像分析模塊，所形成的線圖案是很重要的，因為它顯示在基於 FCRM 的偏斜估計的分佈。第二個步驟選擇重要區域，加速執行時間及使用快速插值的圖像放大，提高準確度的偏斜角估計，從圖像分析階段的輸出圖像的最小的有意義的區域被選擇為降低計算複雜度。然後，將所得到的圖像送到快速插值模塊，將圖像放大來提高估計精度。在最後的過程中，使用 FCRM 聚類算法[16]，[17]方法將數據分類成幾個群並計算集群參數，從兩個或多個不同的數據建立模型，利用此模型估計其偏斜角。這種方法具有很高的傾斜角度估計精度。所提出的方法使用 184 種不同的文件檔當測試集。實驗結果顯示該方法具有高精度估測文件檔的傾斜角度。

本文的結構如下，在第二段提出基於 FCRM 的偏移估算法，第三段落介紹實驗和測試結果，最後在第四段落敘述結論。



貳、基於 FCRM 的偏移估計方法

該演算法包括以下三個步驟。第一步驟為將所述輸入圖像通過圖像分析，建立平行的直線。第二個步驟選擇重要區域，加速執行時間及使用快速插值的圖像放大，提高準確度的偏斜角估計。最後，使用 FCRM 方法估計偏斜角。圖 1 顯示基於 FCRM 的偏斜角估測法的流程圖。我們首先介紹 FCRM 聚類方法，然後每個模組的功能。

A. 模糊 C-回歸模型 (FCRM)

FCRM 從兩個或多個不同的數據建立模型，有賴於實現一個不可觀測的隨機離散變量的組合。設 $S = \{(x_1, y_1), \dots, (x_k, y_k), \dots, (x_N, y_N)\}$ 是一組數據，其中每個獨立的觀察變數 $x_k \in \mathbb{R}^S$ 都有一個對應的依賴觀察變數 $y_k \in \mathbb{R}^t$ 。 (x_k, y_k) 表示第 k 個特徵向量。因此該數據來自不同的模糊 c -回歸模型：

$$y = f_i(\mathbf{x}_k, \beta_i) + \varepsilon_i, \quad i=1, L, c \quad (1)$$

其中， x_k 屬於 $\{x_1, \dots, x_k, \dots, x_N\}$ 表示第 k 個採樣的數據和 β_i 是網絡參數， c 是數據的數量。為了得到合適的參數，模糊 c 回歸模型的目標函數定義

$$J_m(\mathbf{U}, \beta_1, L, \beta_c) = \sum_{k=1}^N \sum_{i=1}^c u_{ik}^m E_{ik}(\beta_i) \quad (2)$$

其中 $m \in (1, \infty)$ 是一個加權指數，阻止模糊性所產生的集群，在 U 元素 u_{ik} 的每一個值被當作 (x_k, y_k) 個模糊子集（集群） S ， $(c \times N)$ 矩陣 $U = \{u_{ik}\}$ 陣列值 $\{u_{ik}\}$ 被稱為模糊 c 分區矩陣，其中 $\{u_{ik}\}$ 部分或全部滿足以下條件：

$$0 \leq u_{ik} \leq 1, \quad \forall i, k \quad (3)$$

$$\sum_{i=1}^c u_{ik} = 1, \quad k=1, L, N \quad (4)$$

和

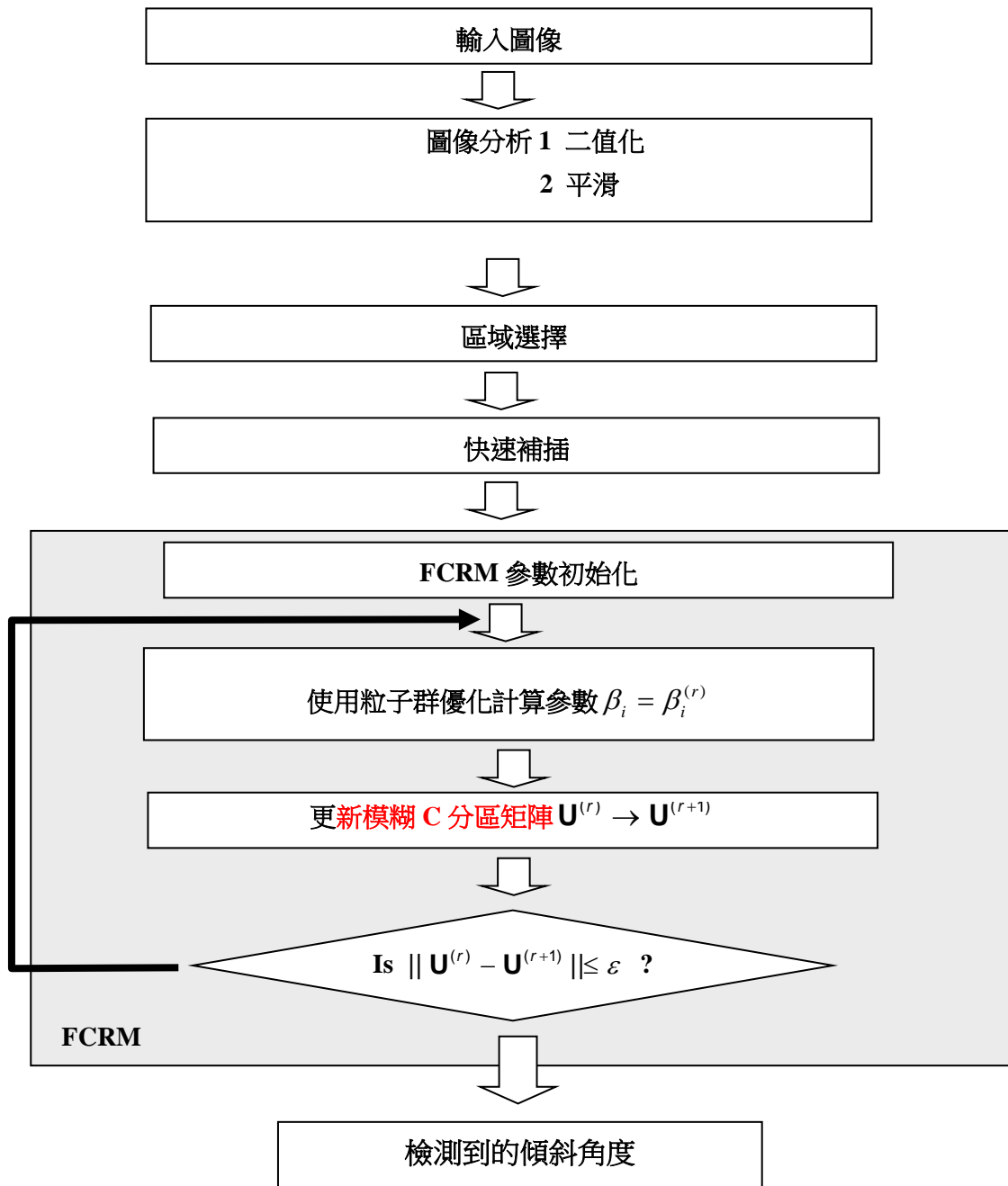
$$0 < \sum_{k=1}^N u_{ik} < N, \quad \forall i=1, L, c \quad (5)$$

(4) 指沒有存在同一類。 $E_{ik}(\beta_i)$ 是衡量在一個特定的模型的預測值的誤差。採取這種方法最常見的例子是所定義的向量的範數的平方：

$$E_{ik}(\beta_i) = \|y_k - f_i(\mathbf{x}_k, \beta_i)\|^2 \quad i=1, L, c \quad (6)$$

FCRM 提供了一個有效的方法來估計 $\{\beta_1, \dots, \beta_c\}$ ，它定義了最適合的模糊 c 回歸模型算法 [16]，[17]。





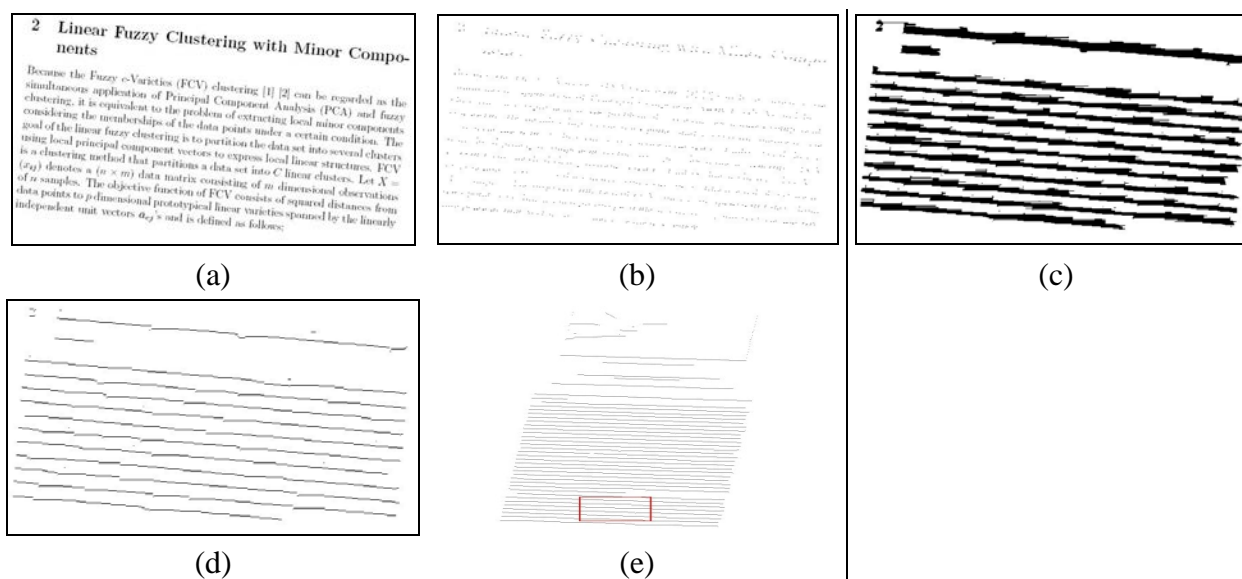


圖2.偏差估計方法的一個例子：(a) 該傾斜的文字圖像；(b) 該脈衝串的圖像，(c) 使用水平運算長度平滑過程的結果(d)定位垂直黑白轉換的結果(e) 操作窗口選擇的結果。

B.基於FCRM的偏移估計算法

偏斜角估測演算法包括以下三個步驟。第一步驟為將所述輸入圖像通過圖像分析，建立平行的直線，以產生平行的直線的圖像分析模塊，所形成的線圖案是很重要的，因為它顯示在基於FCRM的偏斜估計的分佈。第二個步驟選擇重要區域，加速執行時間及使用快速插值的圖像放大，提高準確度的偏斜角估計，從圖像分析階段的輸出圖像的最小的有意義的區域被選擇為降低計算複雜度。然後，將所得到的圖像送到快速插值模塊，將圖像放大來提高估計精度。在最後的過程中，使用FCRM聚類算法將數據分類成幾個群並計算集群參數，從兩個或多個不同的數據建立模型，利用此模型估計其偏斜角。這種方法具有很高的傾斜角度估計精度。

其中圖像分析部分使用了三個階段處理：二值化，長度平滑和黑白。當原始圖像進入系統圖2(a)，輸入圖像的格式可以是RGB或灰度，將黑色像素表示1，並且將白色像素記為0，橫向執行長度平滑過程平滑黑色實行聯繫起來，如果間隔的白色空間的長度小於給定的閾值水平，最內層的圖像平滑。圖2(c)表示施加水平實行長度的平滑過程在圖後得到的圖像。平滑後的圖像重新取樣，以提高基線和減少的數據量，平滑後的圖像的第一垂直掃描一列一列，找到每個黑白在保持其值設為1和更換其他黑色像素的值與0的黑色像素，由此產生的圖像強調的文字的基線，圖2(d)示出通過圖中縱向定位的黑白轉換的結果。為了取得操作區域的尺寸和計算成本之間的折衷，該區域的大小一般是相當小的，在大約100x100像素。值得注意的是，幾個本文行是足夠的估測偏斜角，因此，使用一整頁是沒有必要的。當前的

工作區域中的選擇方法，因為這個方法可以用來在執行長度平滑過程，而不是在原始圖像和黑白處理與正在處理的圖像，將仍然可以工作得很好，即使期間被減少到基準線與圖形或圖片先前的預處理階段。圖 2 (e) 顯示出的操作區域中的選擇結果的圖。圖 2 (c) 中，已被處理的執行長度的平滑和黑白色的處理程序。圖 2 (e) 中，尺寸為 100x100 像素，為中心的突起，其水平和垂直直方圖的最大值的像素。

在下一階段，該圖像以快速插值方法被放大到原來的四倍大小，以提高估測精度。許多方法，如線性，二次和三次插值函數，可用於在快速插值增強圖像。然而，這些方法有些的缺陷，如模糊的邊緣，周圍邊緣振鈴和紋理損失。一種新的快速算法已被用於提高圖像分辨率，任意一個因素的基礎上自適應地選擇不同的內插濾波器。FCRM 聚類演算法是用來估測圖像文件的偏斜角，產生許多平行直線，畫像進入基於 FCRM 的偏斜角估測演算法時，是用直線表示：

$$f_i : y_i = \beta_{i1} + \beta_{i2}x \quad i=1,L,c \quad (7)$$

方程組的模型參數 $\beta_i = \beta_i(r)$ 在步驟 2 計算的值：

$$\beta_{i1} = ((\sum_{k=1}^N (u_{ik})^2)(\sum_{k=1}^N (u_{ik})^2 x_k y_k) - (\sum_{k=1}^N (u_{ik})^2 x_k)(\sum_{k=1}^N (u_{ik})^2 y_k)) / K_i \quad (8)$$

$$\beta_{i2} = ((\sum_{k=1}^N (u_{ik} x_k)^2)(\sum_{k=1}^N (u_{ik})^2 y_k) - (\sum_{k=1}^N (u_{ik})^2 x_k)(\sum_{k=1}^N (u_{ik})^2 x_k y_k)) / K_i \quad (9)$$

$$K_i = (\sum_{k=1}^N (u_{ik})^2)(\sum_{k=1}^N (u_{ik} x_k)^2) - (\sum_{k=1}^N (u_{ik})^2 x_k)(\sum_{k=1}^N (u_{ik})^2 x_k) \quad i=1,L,c \quad (10)$$

每個數學模型的參數反覆調整與模糊 c 分區矩陣。最後，如果 $\|U(r)-U(r+1)\| \leq \epsilon$ ，那麼該過程將停止，否則，執行新的迭代。 C 的模型相當於這些行的斜坡，參數 β_{i2} 是這些線的角度。

參、實驗結果

本論文所提出的演算法使用 184 張影像作測試，這 184 張影像由以下 7 種不同類型的文件檔組成，包括：(1) 一般的文本，(2) 不同大小的字體，(3) 兩欄文件，(4) 方程式 (5)，(6) 圖片和 (7) 名片，如期刊論文和雜誌。圖 3 說明了一些測試圖像。每種類型有 23 個圖像，形成 184 個測試圖像的圖像集。掃描條件是在一個 300 dpi 的解析度和色彩深度為 24 位或 8 位或灰度歪斜文件，由 Adobe 公司的 Photoshop 建立。估計範圍設置為 $[-89^\circ, +89^\circ]$ 和角解析度為 0.1° 。圖像在每個類別進行測試，來衡量演算法的表現，所有這些演算法使用 C 語言實現。

表 I 紀錄 FCRM 的偏移估計演算法和行間互相關法[14]所示的平均誤差和計算時間，兩種演算法比較。從表 1 中 FCRM 方法比隔行互相關法更準確。對於大多數文件 FCRM 方法

平均誤差是低於 0.2° ，不過 FCRM 的計算時間遠長於行間互相關的方法。因為 FCRM 必須計算目標函數的最小值，直到計算出模糊 c-回歸數學模型的最佳參數 β_i ，隔行互相關法計算互相關函數的二進制圖像，FCRM 的計算成本高於行間互相關的方法。由於 FCRM 必須調整每一個數學模型的參數，從而決定了傾斜角度。這兩種方法測試名片時，錯誤率增加，因為名片通常包含的信息較少，這是不夠傾斜角度的估測。圖 4 顯示了不同類型文件的平均誤差圖，由於對每一種類型的文件具有最小精度誤差，FCRM 方法表現最好。

圖 5 顯示演算法計算時間，平均偏移估計的計算時間的互相關的方法是 0.449 秒，自互相關法的複雜性比 FCRM 是較低，FCRM 需較長的計算時間，平均 8.414 秒，但在大多數情況下是可以接受的。



(a) 一般的文本

(b) 不同大小的字體

(c) 兩欄文件

(d) 方程式

(e) 不同的語言

(f) 圖片

(g) 名片

圖 3 測試文件範例



Image Type	Cross-Correlation		FCRM	
	Average Error (deg)	Computational Time (sec)	Average Error (deg)	Computational Time (sec)
(1) General Text	0.70	0.428	0.09	16.1
(2) Different Sizes	0.31	0.440	0.1	15.5
(3) Two Columns	0.36	0.447	0.1	13.9
(4) Equations	0.31	0.436	0.22	3.6
(5) Different Languages	0.36	0.469	0.12	6.7
(6) Pictures	0.74	0.509	0.2	12.5
(7) Name Card	0.68	0.412	0.4	4.6
Average	0.5	0.45	0.15	11.6

表 1 不同類型文件的平均誤差和計算時間

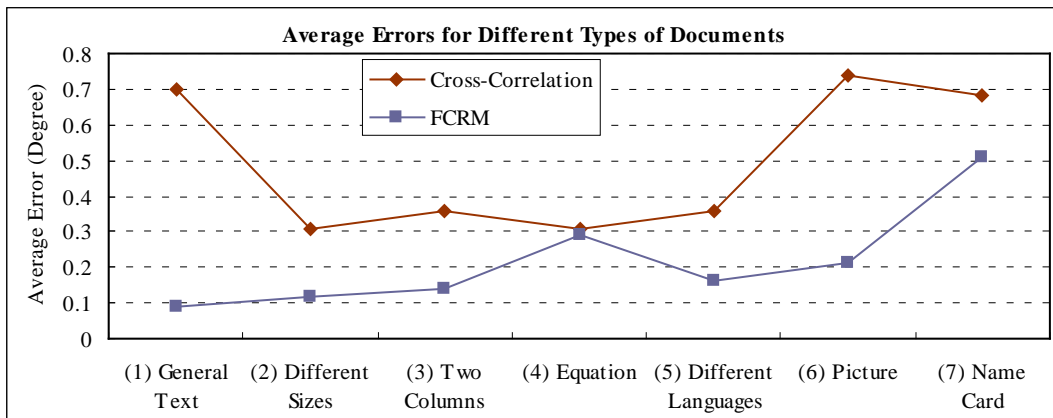


圖 4 不同類型文件的平均誤差，其中的菱形符號表示互相關的方法和方形符號表示 FCRM 的方法

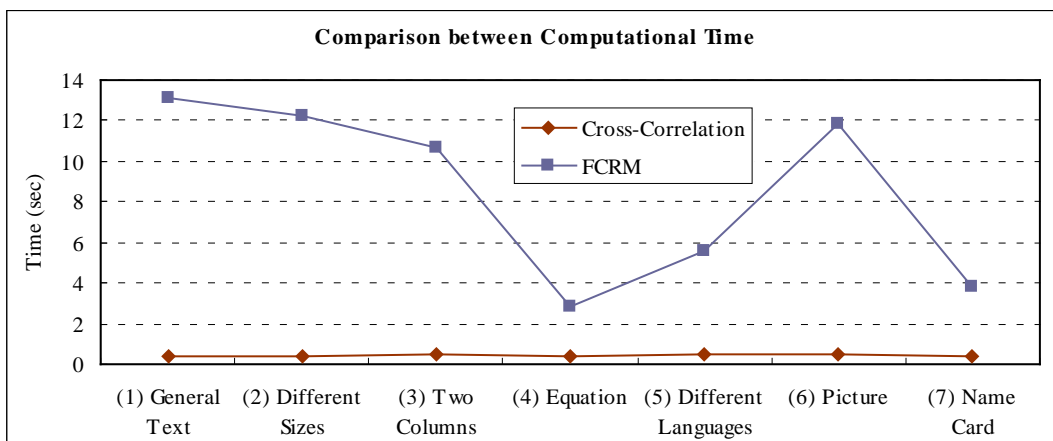


圖 5 不同類型文件的平均計算時間，其中的菱形符號表示的互相關的方法和方形符號表示在 FCRM 方法

肆、結論

本研究採用模糊 C-回歸的算法 (FCRM) 估計文件的傾斜角，實驗結果顯示該算法比行間互相關法具有更好的精確度。歪斜的估計能適用於不同類別的掃描或複印文件包括一般文件，雜誌，報紙，其平均誤差為 0.2° 內，模糊 C-回歸的算法 (FCRM) 演算法具有較高的精確率，合理的計算成本為。

參考文獻

- [1] W. Postl, "Detection of linear oblique structures and skew scan in digitized documents," in Proc. 8th International Conference Pattern Recognition, 1986, pp. 739-743.
- [2] G. Ciadiella et al., "An experimental system for office document handling and text recognition," in Proc. 9th International Conference Pattern Recognition, 1988, pp. 739-743.
- [3] D. S. Bloomberg, G. E. Copec, and L. Dasari, "Measuring document image skew and orientation," in Proc. SPIE, San Joes, CA, Feb. 1995, vol. 2422, pp. 302-316.
- [4] H. S. Baird, "The skew angle of printed documents," in Proc. of the Conference of the Society of Photographic Scientists and Engineers, 1987, pp. 21-24.
- [5] E. Kavallieratou, N. Dromazou, N. Fakotakis, and G. Kokkinakis, "An integrated system for handwritten document image processing," International Journal of Pattern Recognition, Vol. 17, pp. 617-636, June 2003.
- [6] A. Hashizume, P. S. Yeh, and A. Rosenfeld, "A method of detecting the orientation of aligned components," Pattern Recognition Letters, Vol. 4, pp. 125-132, 1986.
- [7] L. O'Gorman, "The document spectrum for page layout analysis," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 15, pp. 1162-1173, Nov. 1993.
- [8] Y. Lu and C. L. Tan, "A nearest-neighbor chain based approach to skew estimation in document images," Pattern Recognition Letters, Vol. 24, pp. 2315-2323, Oct. 2003.
- [9] S. C. Hinds, J. L. Fisher, and D. P. D'Amato, "A document skew detection method using run-length encoding and the Hough transform," in Proc. of the Tenth International Conference on Pattern Recognition, 1990, pp. 464-468.
- [10] D. S. Le, G. R. Thoma, and H. Wechsler, "Automated page orientation and skew angle detection for binary document images," Pattern Recognition, Vol. 27, pp. 1325-1344, Oct. 1994.
- [11] B. Yu and A. K. Jain, "A robust and fast skew detection algorithm for generic documents," Pattern Recognition, Vol. 29, pp. 1599-1629, Oct. 1996.
- [12] P. Y. Yin, "Skew detection and block classification of printed documents," Image and Vision Computing, Vol. 19, pp. 567-579, May 2001.
- [13] H. Yan, "Skew correction of document images using interline cross-correlation," in CVGIP



- Graphic Models Image Processing, Vol. 55, pp. 538–543, Nov. 1993.
- [14] Avanindra and S. Chaudhuri, “Robust detection of skew in document images,” IEEE Transactions on Image Processing, Vol. 6, pp. 344–349, Feb. 1997.
- [15] A. Amin and S. Wu, “A robust system for thresholding and skew detection in mixed text/graphics documents,” International Journal of Image and Graphics, Vol. 5, pp. 247-265, April 2005.
- [16] R. J. Hathaway and James C. Bezdek, “Switching regression models and fuzzy clustering,” IEEE Transactions on Fuzzy Systems, Vol. 1, pp. 195-204, Aug. 1993.
- [17] C. C. Kung and C. C. Lin, “Fuzzy C-regression model with a new cluster validity criterion,” in Proc.of the 2002 IEEE International Conference on Fuzzy Systems, Vol. 2, 2002.