

論 文

映像處理에 의한 한글印章의 自動識別에 關한 研究

正會員 李 基 敦* 正會員 全 炳 敏** 正會員 金 商 雲***

A Study on the Automatic Identification of HAN-GEUL Seal by using the Image Processing.

Ki Don LEE*, Byoung Min JUN** and Sang Woon KIM***, Regular Members

要 約 印章識別은 平滑化, 回轉 및 移動, 細線化整合技法의 順序로 수행된다. CCTV카메라에 入力된 印章映像은 A/D 變換器와 8 bit microcomputer에 의하여 256×256畫素의 2進映像으로 변환한다. 두 映像을 正位置로 回轉시킨 後, 細線化過程을 거쳐 文字의 골격선을 찾는다. 畫素거리에 따라 Weight를 정하고 distance-weighted Correlation, C^k 를 계산한다. 계산된 C^k 와 결정상수 C_{YES} , C_{NO} 를 비교하여 印章을 識別한 結果 95%의 높은 識別率을 얻었고 한개의 對象印章을 識別하는데 걸린 시간은 3분이내였다.

ABSTRACT The proposed seal identification procedure consists of the smoothing, rotation, thinning, and matching techniques. The seal images which are scanned by CCTV are thresholded into the binary pictures of 256×256 pixels through A/D converter and 6502 microcomputer. After the sample and target images are rotated into an identical orientation, a thinning process is used to extract the skeletons of the character strokes. The weighted map is constructed by distance weight from which the distance weighted correlation C^k is computed. The C^k is compared with the decision constant C_{YES} or C_{NO} for the purpose of seal identification. The identification rate is 95% and the total CPU time is less than 3 minutes for each identification in the experiment.

1. 序 論

西歐에서 署名(signature)과 같은 역할을 하는 것이 우리나라에서는 印章(seal)이다. 印章은 各種文書와 銀行手票등에 널리 사용되고 있다. 이러한 印章의 사용은 模造되어서는 안된다는 敬을 전제로 하지만 만약, 印章을 정교하게 모

조하여 犯罪目的으로 사용한다면 現在의 識別方法으로는 이런 犯罪를 예방할 수 없다.

本 論文에서는 우리나라 印章의 自動識別(automatic indentification)에 關한 시스템의 開發과 處理技法을 研究하였다. 印章識別은 두 印章의 文字가 동일해야 함은 물론 文字의 크기, 폭, 길이 등의 상대적인 위치가 같아야 한다는 특징을 갖는다.

文字認識⁽¹⁾(character recognition)은 文字의 형태만을 비교하고 署名識別⁽²⁾(signature identification)은 형태와 모양(Style)만을 비교한다. 한편 指紋識別⁽³⁾(fingerprint identification)은

*, **, *** 延世大學校 工科大學 電子工學科
Dept. of Electronic Engineering, Yonsei
University Seoul. 120 Korea.
論文番號 : 85 - 09 (接受 1984. 11. 5)

印章識別과 성질상 비슷하지만 人爲的으로 가짜 指紋을 만들지는 않는다.

印章識別의 문제점은 비록 동일한 印章이라도 捺印할 때 가하는 힘과 인주색에 따라 형상이 달라질 수 있으며 印章의 부분적인 汚損이나 손흔들림에 의한 중복현상이 誤識別의 주요 원인이 된다.

本 論文에서는 印章識別을 쉽게 하기 위해서 다음의 세 조건을 만족시키는 印章을 택하였다.

- ① 타원형의 한글印章
- ② 捺印된 형상이 90° 以內로 기울어야 한다.
- ③ 捺印된 형상에서 테두리의 두 끝부분이 손실되어서는 안된다.

그림 1-(a)는 識別가능한 捺印형상이고 그림 1-(b)는 識別곤란한 捺印형상이다.



그림 1 捺印형상 Seal imprints.

2. 시스템 構成

印章識別 시스템은 그림 2와 같다.

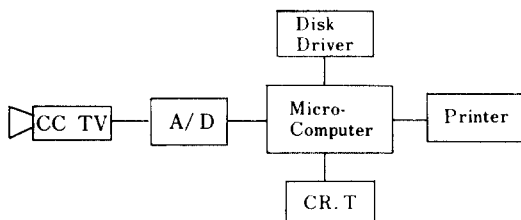


그림 2 印章識別 시스템 Seal identification system.

C. C. TV (closed circuit TV) camera에 入力된 印章映像은 A/D變換器와 6502-microcomputer의 프로그램제어로 256×256 해상도를 갖는 2進映像으로 變換한다. 畫素의 水平, 垂直位置를 표시하는데는 8 bit ($2^8=256$)가 필요하며 A/D變換기에 대한 블럭도는 그림 3과 같다.

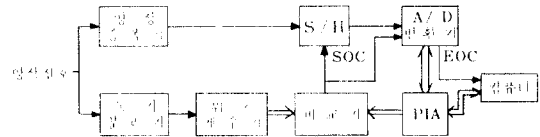
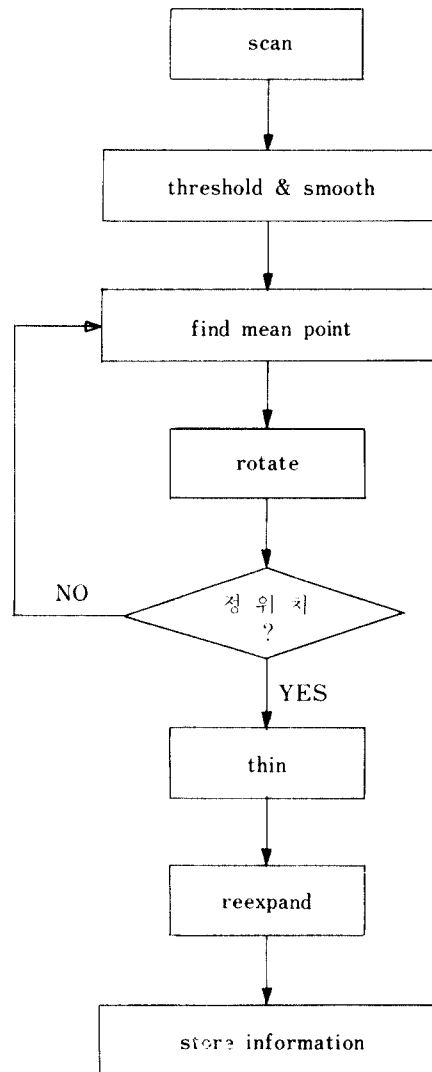


그림 3 A/D變換기의 블럭도. A/D converter block diagram.

標本印章 (sample seal)과 識別하고자하는 對象印章 (target seal)의 映像處理過程을 그림 4에 flow chart로 나타냈고 다음절에서 상세히 論한다.



(a)

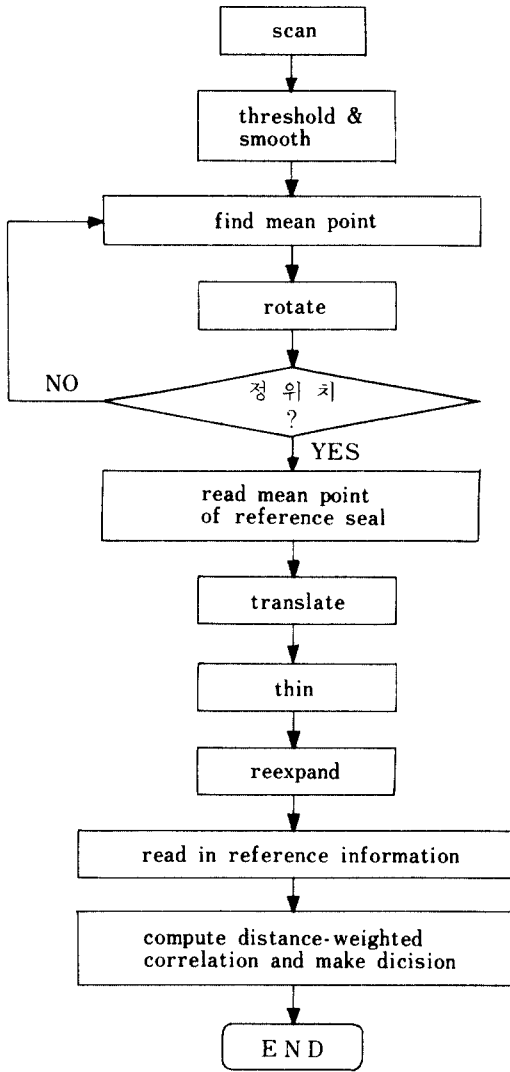


그림 4 (a) 標本印章의 映像處理 흐름도
Flow diagram for the operations on sample seal images.
(b) 對象印章의 映像處理 흐름도
Flow diagram for the operations on target seal images.

3. 處理 技法

3 - 1. 平滑化⁽⁴⁾ (Smoothing)

2進으로 變換된 印章映像에서 잡음을 제거하고 最適의 映像을 얻기 위하여 다음의 여러가지 平滑化技法을 수행한다.

① 3 - 이웃平滑化 (3 - neighborhood smoothing)

그림 5의 畫素배열(pixel array)에서, 畫素 P_1 은 式(1)에 의하여 P'_1 로 變換한다.

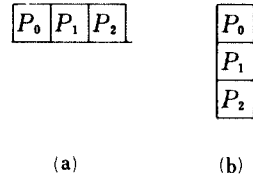


그림 5 3 - 화소배열
3 - pixel array.

$$P'_1 = P_0 P_1 + P_1 P_2 + P_0 P_2 \quad (1)$$

② 4 - 이웃平滑化

①의 平滑化技法을 수행한 후, 그림 6의 畫素배열에서 P_1 과 P_2 는 式(2)와 (3)에 의하여 P'_1 와 P'_2 로 變換한다.

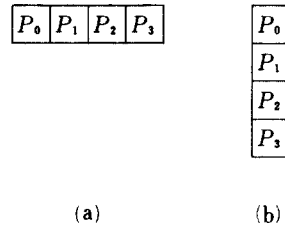


그림 6 4 - 화소배열
4 - pixel array.

$$P'_1 = P_0 P_1 + P_0 P_3 + P_1 P_3 \quad (2)$$

$$P'_2 = P_2 P_3 + P_0 P_3 + P_0 P_2 \quad (3)$$

③ 8 - 이웃平滑化

그림 7의 畫素배열에서 임의의 값, P_x 는 式(4)로 구한다.

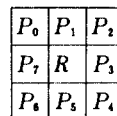


그림 7 8 - 화소배열
8 - pixel array.

$$S_x = P_0 + P_1 + P_2 + P_3 + P_4 + P_5 + P_6 + P_7$$

$$= \sum_{i=0}^7 P_i$$

$$\left. \begin{aligned} P_x = 1; S_x \geq T \\ = 0; 0 \leq S_x < T \end{aligned} \right\} \quad (4)$$

여기서, T 는 臨界值 (threshold value) 를 의미한다. (本 論文에서는 T 의 값으로 4 를 선택) 위의 3 가지平滑化技法을 반복 수행하는 프로그램을 그림 1-(a)에 적용한 결과, 그림 8 과 같은 印章映像을 얻었다.

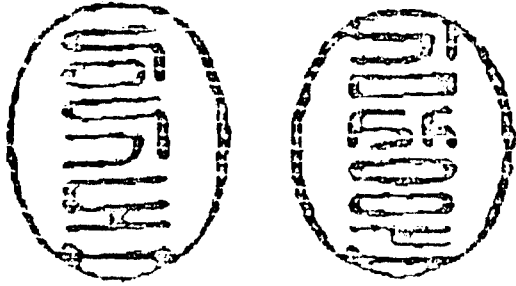


그림 8 그림 1-(a)를 平滑化한 印章映像
Smoothing seal images of Fig 1-(a).

3-2. 回轉⁽⁵⁾ (Rotation)

標本印章 (sample seal) 의映像, $S = \{P_s(i, j) | i, j = 1, \dots, n\}$ 와 對象印章 (target seal) 의映像, $T = \{P_T(i, j) | i, j = 1, \dots, n\}$ 를 대응시키기 위하여 기하학적 中心點 (mean-point) 과 回轉角을 구해야 한다. 시계방향으로 Q 만큼 기울어진 그림 9의 映像座標에서 中心點 (X_0, Y_0) 과 회전각 Q 는 式(5), (6)으로 구한다.

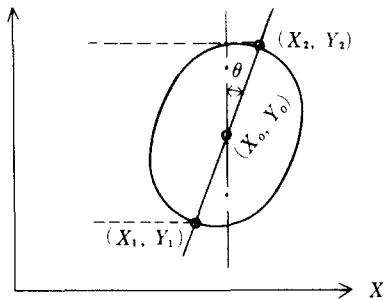


그림 9 回轉된 映像座標
Rotated image coordinate.

$$\left. \begin{aligned} X_0 = X_2 - X_1 \\ Y_0 = Y_2 - Y_1 \end{aligned} \right\} \quad (5)$$

$$Q = \tan^{-1} \frac{X_2 - X_0}{Y_2 - Y_0} \quad (6)$$

中心點 (X_0, Y_0) 를 기준으로 映像을 반시계방향으로 Q 만큼 회전시켜 正位置에 놓기 위하여 式(7)의 回轉行列 (rotation matrix) 을 사용한다.

$$\begin{pmatrix} \cos \theta & \sin \theta & 0 \\ \sin \theta & -\cos \theta & 0 \\ -X_0 \cos \theta + Y_0 \sin \theta + X_0 & -X_0 \sin \theta - Y_0 \cos \theta + Y_0 & 0 \end{pmatrix} \quad (7)$$

3-3. 細線化 (Thinning)

비교식별할 두 印章映像의 데이터량을 줄이고 文字의 골격선 (skeleton) 을 찾기 위하여 細線化를 수행한다. 細線化技法에는 여러가지가 있지만 實驗結果, 한글印章의 細線化에 가장 적합한 다음의 방법을 이용했다.

수평방향으로, 연속적인 黑點 (畫素값이 1 인點) 의 중심만을 추출한 映像의 골격선과 수직방향으로, 연속적인 黑點의 중심만을 추출한 映像의 골격선을 더한다. 그림 10는 그림 8를 細線化한 결과이다.

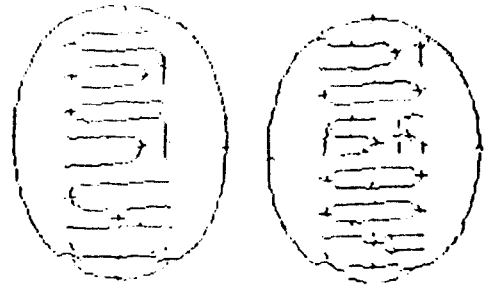


그림 10 그림 7을 細線化한 印章映像
Thinning seal images of Fig. 7.

3-4. 整合 (Matching)

細線化技法으로 字劃 (stroke) 이 얇아진 두 映像의 類似度 (similarity) 를 계산하여 眞偽을 판별하여야 한다.

특히, 두 印章映像을 整合시켰을 때 標本印章

에 대한 對象印章의 類似度를 측정할 수 있는 합수는 **product correlation**이 아닌 **distance-weighted correlation**임을 증명하자.

예를 들면, 두 映像을 整合시켰을 때 畫素배열이 그림11과 같다.

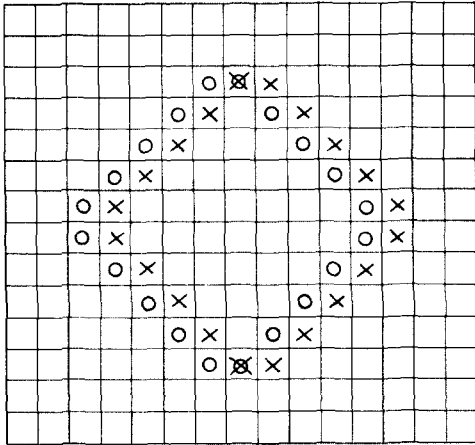


그림11 整合例
Example of matching.

여기서 “0”표시는 映像 S 안의 흑점(畫素값이 1)이고 “X”표시는 映像 T 안의 흑점이며 표시가 없는 畫素는 畫素값이 0이다.

映像 S를 映像 T에 整合시켰을 경우, **Correlation** 계수는

$$C(S, T) = \frac{\sum_i P_i(S, T)}{\sum_i P_i(S)} \quad (8)$$

이다. 단, i 는 映像 S 안의 흑점총수(그림11에서 “0”표시의 총갯수)이고 j 는 映像 T의 흑점과 겹치는 映像 S 안의 흑점총수(그림11에서 “X”표시의 총갯수)이다.

映像 T를 映像 S에 整合시켰을 경우, **Correlation** 계수는

$$C(T, S) = \frac{\sum_j P_j(T, S)}{\sum_j P_j(T)} \quad (9)$$

이다. 단, i' 는 映像 T 안의 흑점총수(그림11에서 “X”표시의 총갯수)이고 j' 는 映像 S 안의 흑점과 겹치는 映像 T 안의 흑점총수(그림11에서

“X”표시의 총갯수)이다.

式(8)과 式(9)를 곱한 **Product Correlation**은

$$C = \frac{\sum_j P_j(S, T) \sum_i P_i(T, S)}{\sum_i P_i(S) \sum_j P_j(T)} \quad (10)$$

$$= \frac{(\sum_i P_i(S, T))^2}{\sum_i P_i(S) \sum_j P_j(T)} \quad (11)$$

이다. 式(10)에서 두 映像이 겹친 畫素의 갯수는 동일하기 때문에 式(11)이 된다.

그림11에 式(11)의 **Product Correlation**을 적용하면,

$$C = \frac{4^2}{20 \times 20} = 0.01$$

두 映像의 類似度가 0.01의 아주 작은 값이 얻어진다. 이 결과는 두 映像이 전혀 다르게 판별된다. 그러나 실제의 印章映像을 整合하면 동일한 印章에 의할지라도 그림11과 같이 완전히 겹치지 않는 상대공간위치가 비슷한 경우가 많다. 그래서 印章의 類似度를 계산하는데 **Product Correlation**은 부적합하다.

상대공간위치까지 고려할 수 있도록, 거리에 따라 **Weight**를 갖는 계산방법을 도입하여야 한다. 두 映像을 整合시켰을 때,

① 완전히 겹치는 흑점의 **weight**는 1로 정한다.

② 비교할 흑점사이의 거리가 멀수록 0에 가깝도록 **weight**를 정한다.

위 조건을 만족시킬 수 있는 거리에 따른 비교 흑점의 **weight**에 대한 式은

$$W_p^k = \left. \begin{aligned} & \frac{1}{d_p^2 + 1} ; 0 \leq d_p \leq K \\ & = 0 ; d_p > K \end{aligned} \right\} \quad (12)$$

이다. 여기서 d_p 는 畫素사이의 **Euclidean**거리이고, 계산이 간편하도록 **weight**의 범위를 K 로 제한한다. 本論文에서는 K 의 값으로 $2\sqrt{2}$, 즉 5×5 畫素를 선택하였다.

그림12-(a)의 畫素배열을 式(12)를 이용하여 거리에 따른 **weight**로 **reexpansion**한 **weighted map**은 그림12-(b)와 같다.

0 0 0 0 0 0	0 0 0 0 0 0
0 0 0 0 0 0	0.11 0.17 0.2 0.17 0.11 0
0 0 0 0 0 0	0.17 0.33 0.5 0.33 0.17 0
0 0 1 0 0 0	0.2 0.5 1 0.5 0.2 0
0 0 0 0 0 0	0.17 0.33 0.5 0.33 0.17 0
0 0 0 0 0 0	0.11 0.17 0.2 0.17 0.11 0
(a)	(b)

그림12 (a) reexpansion 前 (b) reexpansion 後
Before reexpansion. After reexpansion.

비교映像 S와 T안의 모든 畫素에 대하여 각각 그림12-(b)와 같이 weighted-map을 작성한다. 이 weighted-map을 이용하여 상대공간 위치도 고려한 두映像의 類似度는 다음의 distance-weighted correlation⁽⁶⁾으로 계산할 수 있다.

$$C^k(S, T) = \frac{1}{2} \left(\frac{\sum_i W_i^k(T)}{\sum_i P_i(S)} + \frac{\sum_{i'} W_{i'}^k(S)}{\sum_{i'} P_{i'}(T)} \right) \quad (13)$$

여기서 i 와 i' 는 각映像안의 焦點(weight가1) 총수이고 $W_i^k(T)$ 는映像 S안의 焦點과 대응하는映像 T안의 weight이며 $W_{i'}^k(S)$ 는映像 T안의 焦點과 대응하는映像 S안의 weight이다.

distance-weighted correlation, C^k 에 대한 식(13)을 그림11에 적용하여 類似度를 계산하면,

$$C^k(S, T) = \frac{1}{2} \left(\frac{1 + 1 + 0.5 \times 18}{20} + \frac{1 + 1 + 0.5 \times 18}{20} \right) = 0.55$$

앞의 product correlation으로 계산한 0.01보다 훨씬 큰 값을 얻었으며 두映像이 類似하다고 말할 수 있다.

式(13)에서 두映像이 완전히 동일한 경우만 $C^k = 1$ 이고 그 외는 $0 \leq C^k < 1$ 이다. 완전히 겹쳐지는 焦點뿐만 아니라 약간 떨어진 焦點까지 고려하여 類似度를 계산할 수 있는 함수가 바로 distance-weighted correlation, C^k 임을 알 수 있다.

계산된 C^k 로 비교映像의 眞偽를 판별하기 위하여 결정상수 C_{YES} 와 C_{NO} 를 정하여야 한다. 본

論文에서는 $C_{YES} = 0.76$, $C_{NO} = 0.74$ 를 선택하였다. 이 결정상수를 利用하여 對象映像을 세 유형으로 識別한다.

- ① $C^k \geq C_{YES}$; 標本映像과 동일
- ② $C^k \leq C_{NO}$; 標本映像과 다름
- ③ $C_{YES} < C^k < C_{NO}$; 識別 곤란

4. 實驗 結果

實驗對象으로 다음과 같은 印章의 種類를 사용하였다.

- ① 4인이 한개씩 標本印章製作.
- ② ①과 同一인이 標本印章을 보고 다시 한개씩 製作.
- ③ 다른 3인이 標本印章을 보고 4개씩 製作 위의 標本印章 4개(LS, KS, JS, BS)에 대하여 對象印章20개(標本印章포함)의 C^k 값을 계산한 결과는 表1이다.

표1 계산된 C^k 값
Computed C^k value.

SAMPLE	TARGET & C^k VALUE				
	LS	LT 1	LT 2	LT 3	LT 4
LS	0.84	0.68	0.70	0.71	0.69
	KS	KT 1	KT 2	KT 3	KT 4
KS	0.87	0.70	0.71	0.74	0.70
	JS	JT 1	JT 2	JT 3	JT 4
JS	0.76	0.70	0.77	0.72	0.70
	BS	BT 1	BT 2	BT 3	BT 4
BS	0.80	0.72	0.68	0.73	0.69

表1에서, 標本印章을 對象印章(LS, KS, JS, BS)으로 했을 경우에 C^k 값이 제일 크고 ②의 對象印章(LT3, KT3, JT3, BT3)의 평균 C^k 값이 두 번째로 큼을 알 수 있다.

20개의 對象印章중에서 標本印章을 對象印章으로 했을 경우만 "眞(YES)", 나머지 16개의 對象印章은 "偽(NO)"로 판별되어야 타당하기 때문에 본 實驗에서는 결정상수의 값으로 $C_{YES} = 0.76$, $C_{NO} = 0.74$ 를 선택하였다. 表1에서 계산된 C^k 와 결정상수를 비교하면 원으로 표시한 對象印章, JT 2만이 誤識別印章임을 안다. 그래서

실험 결과 印章의 識別率은

$$\begin{aligned} \text{識別率} &= \frac{\text{對象印章數} - \text{誤識別印章數}}{\text{對象印章數}} \times 100 [\%] \\ &= \frac{20 - 1}{20} \times 100 [\%] = 95 [\%] \end{aligned}$$

이다.

5. 結 論

8 bit microcomputer (Apple II) 를 이용하여 回轉技法만 제외하고 다른 處理過程을 기계어로 수행시킨 결과 한개의 對象印章을 識別하는데 걸린 시간이 3 분이내였다. 그리고 높은 識別率 (95%) 을 얻은 것은 本 論文의 印章識別處理技法이 타당함을 입증한다.

앞으로의 연구과제는, 보편화될 16bit micro-

computer 를 이용하여 어떤 형상의 印章도 빠른 시간에 識別이 가능한 시스템과 알고리즘을 개발하여 實用化시켜야 한다.

參 考 文 獻

- (1) W. Stallings, "Approaches to chinese character recognition," Pattern Recognition 8, 1976, pp. 87-98.
- (2) R. N. Nagel and A. Rosenfeld, "Steps toward handwritten Signature verification," Proceedings, First International Joint Conference on Pattern Recognition, Washington D. C, 1973.
- (3) K. Rao and K. Balck, "Type classification of fingerprints A syntatic approach," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2, pp. 223-231, 1980.
- (4) A. Rosenfeld and A. C. Kak, "Digital picture processing," Academic, 1976, pp. 192-200.
- (5) T. Pavlidis, "Algorithms for graphics and Image Processing," Computer Science, 1982, pp. 318-322.
- (6) T. J. Fan and W. H. Tsai, "Automatic chinese seal identification," Computer Vision Graphics Images Processing 25, pp. 311-330, 1984.



李 基 敦 (Ki Don LEE) 正會員
 1953年 11月 5日生
 1973年~1976年: 韓國航空大學電子工學科卒業
 1977年~1979年: 延世大學校大學院電子工學科(工學碩士)
 現 在: 延世大學校大學院博士過程在學中
 1985年 3月~現 在: 湖西大學電算科專任講師



全 炳 敏 (Byoung Min Jun) 正會員
 1954年 1月 9日生
 1972年 3月~1976年 2月: 韓國航空大學電子工學科卒業
 1976年 3月~1978年 2月: 延世大學校大學院電子工學科(工學碩士)
 1982年 8月~現在: 延世大學校大學院博士過程在學中



金 商 雲 (Sang Woon KIM) 正會員
 1956年 3月 13日生
 1974年~1978年: 韓國航空大學通信工學科卒業
 1978年~1980年: 延世大學校大學院電子工學科(工學碩士)
 現 在: 延世大學校大學院博士過程在學中
 1984年~現 在: 韓國放送通信大學電子計算學科專任講師

1978年 8月~1982年 3月: 空軍士官學校電子工學科 專任講師
 1982年~現在: 東洋工業專門大學專任講師.