

멀티-폰트 印刷數字 認識

Recognition of Multi-Font Printed Numeric Characters

安 錫 出*

ABSTRACT

This paper describes an algorithm recognizing multi-font printed numeric characters.

In order to extract feature selection of printed numeric characters, this paper describes an algorithm using stroke density function. Printed numeric characters are recognized by using the set of stroke-density feature vectors.

I. 序 論

近年 韓國語情報處理 system의 開發에 關心이 높아지고 있다. 이러한 system이 發展하기 위해서는 文字入力の 高速化, 肖力化가 큰 課題로 되어있다. 文字入力에는 key bord를 利用한 直接入力方法이 있으나 肖力化를 위해 文字OCR에 기대하는 것이 크다.

文字OCR은 크게 나누어 印刷文字OCR과 필기체文字 OCR가 있다. 그러나 그 問題點은 Category數가 많다는 量的問題, 對象字形이 複雜하다는 質的인 問題로 集約된다.

印刷文字 OCR에 對해서 여러가지 方法이 있으나^{1~3)} 本論文에서는 文字認識時 細線化 作業을 거치지 않고 文字의 周邊分布 特徵을 軸出하여 그 特徵 Vector의 集合을 利用해서 印刷數字 認識을 했다.

對象文字의 畫像은 가로80點×세로 80點으로 正規化된 2進데이터를 使用했으며 그 認識結果는 96.7%였다.

이로서 字認識時 細線化 作業을 하지않고, 周邊分布特徵을 利用하면 category數가 많은 印刷된 한글文字 認識에도 利用할수 있으리라 사료된다.

II. 畫像의 自動認識

人間이 對象物을 눈으로 보고 認識 할때는, 먼저 對象物로부터 빛이 눈의 水晶體를 通하고, 網膜上의 光感受性의 細胞에 의해 光의 應答이 神經을 通해서 大腦에 傳達되어 여기서 綜合的인 處理를 하여 對象物을 認識하고 있다.

이와 마찬가지로 機械를 利用하여 自動的으로 對像物을 認識시키고져할 경우에도 人間과 같은 構成이 必要하다.

즉 人間의 눈에 相當하는 畫像入力裝置와 人間의 大腦의 作用을 畫像處理用 計算機가 이를 담당한다.

패튼(pattern)認識의 課題는 크게 3가지로 分類된다.

- ① 사전처리
- ② 特徵軸出
- ③ 分類

II-1. 사전처리(pre processing)

一般的으로 畫像入力裝置를 통해서 入力된 畫像은 階調가 있는 畫像이고, 인물사진이나 풍경사진의 경우는 32~64단계의 階調가 必要하지만 이들 畫像도 각각의 畫素당 24단계의 階調만 있으며 印刷物 만큼의 階調再現이 가능하다.

한편 文字나 圖形의 경우는 2단계 階調로 變化하여 2진 데이터의 畫像으로 變換한다.

다음은 文字畫像의 크기와 位置의 正規化 作業으로서 對像文字畫像을 M×M畫素로 入力시킨 후 文字의 重心(Weight point)을 中心으로 하여 N×N畫素로 強制的인 制約을 加한다.

特徵軸出은 이 正方領域內에서 실시하고, 畫像의 幅n值는 本來 必要로하는 스토로크(stroke)情 報를 잃지 않는 範圍로 決定해야만 한다.

II-2. 스토로크 密度函數에 의한 特徵軸出

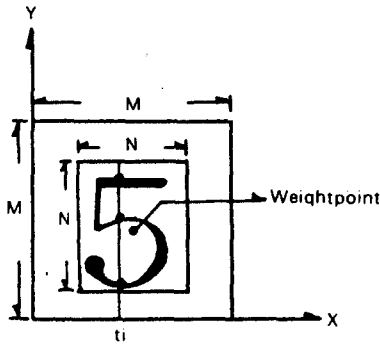


Fig. 1 Stroke density function

本 論文에서는 패튼의 分析 및 認識을 爲한 특징으 로서 stroke density 特徵을 利用했다.

使用한 畫像데이터의 규격은 N=80인 30個 文字을 中心으로 實驗했다. 그림 1의 文字圖形은 水平또는 垂直 方向을 座標軸으로하여 座標軸上 各點으로 文字圖形 의 行또는 列方間으로 切斷線을 만든다.

切斷線이 가로짜르는 文字線을 計數하고 이를 stroke density function(以下SDF)라 부른다.

그림 1의 座標軸 X上的 임의의 點 Xi 點에서 SDF는

$$F(X_i)=4 \dots\dots\dots (1)$$

로 된다.

같은 方法으로 垂直方向을 座標軸으로한 SDF를 구하여 한文字의 特徵으로 한다.

n=80인 文字圖形 data를 利用하면 상세한 情報를 내포하고 있으나 文字의 位置에 민감하므로 位置變化에 對한 安定된 SDF를 求하기 위하여 20點 近似를 採用했다.

20點 近似를 採用한 경우 N=80인 경우의 SDF 패튼 Fp는 다음式으로 주어진다.

$$F_p = \left\{ \sum_{i=1}^4 F(x_i)/4, \sum_{i=1}^4 F(x_i)/4, \dots\dots\dots \sum_{i=77}^80 F(x_i)/4, \right. \\ \left. \sum_{i=1}^4 F(y_i)/4, \sum_{i=1}^4 F(y_i)/4, \dots\dots\dots \sum_{i=77}^80 F(y_i)/4, \dots\dots\dots \right\} (2)$$

同一 Category의 N個 文字의 平均特徵 \tilde{F} 는

$$\tilde{F} = \frac{1}{N} \sum F_p \dots\dots\dots (3)$$

이다. 여기서 F_p 는 特定書體 文字로부터 求한 特徵을 말하고 P는 同一 category內의 各各의 文字를 表示한다.

III. 判別函數(Discriminant function)

III-1. 3-class 最小거리 함수

2次元 特徵空間에서 임의의 F패튼이 C_1, C_2 또는 C_3 中 어디에 屬할 것인가는 유사성에 의해서 分類된다.

C_1, C_2 와 C_3 群의 各패튼의 集合 $\{F_{ij}\}$ 의 平均 Vector를 \tilde{F}_i 라 하면

$$\tilde{F}_i = \frac{1}{N} \sum_{j=1}^N F_{ij} \quad i=1, 2, 3 \quad (4)$$

만약 임의의 패튼 Vector F가 \tilde{F}_i 에 가장 가깝다면 패튼 F는 C_i 에 속한다.

D_i 를 \tilde{F}_i 로부터 F까지의 거리로 두고, D_i 가 最小인 函數를 利用하여 패튼을 認識한다.

이때 使用하는 函數를 最小거리함수(minimum-distance function)이라한다.

$$D_i^2 = \|F - \tilde{F}_i\|^2 = \|F\|^2 - 2\{\tilde{F}_i' F - \frac{1}{2} \|\tilde{F}_i\|^2\} \quad i=1, 2, 3 \quad (5)$$

式(5)로부터 D_i^2 가 最小가 되기 爲해서는 $\{\tilde{F}_i' F - \frac{1}{2} \|\tilde{F}_i\|^2\}$ 가 最大가 되어야 한다. 따라서 D_i^2 가 最小가 되는 패튼을 分類하는것 보다 $\{\tilde{F}_i' F - \frac{1}{2} \|\tilde{F}_i\|^2\}$ 가 最小가 되는 패튼을 認識하는 편이 간편하다. 따라서 이 式을 判別함수(discriminant function)이라하고 式(6)과 같다.

$$g_i(F) = \tilde{F}_i' F - \frac{1}{2} \|\tilde{F}_i\|^2 \quad i=1, 2, 3 \quad (6)$$

II-2. 10-class minimum distance function.

使用한 認識對像文字는 書體가 다른 3종류의 숫자 30個를 sample로하여 그 特徵은 20點 近似에 의한 stroke density를 利用했다.

同一 Category內의 文字數는 3이므로, 文字의 特徵點에 對한 平均 stroke density를 \tilde{F}_i 라하면 式(6)과式(7)로 表現된다.

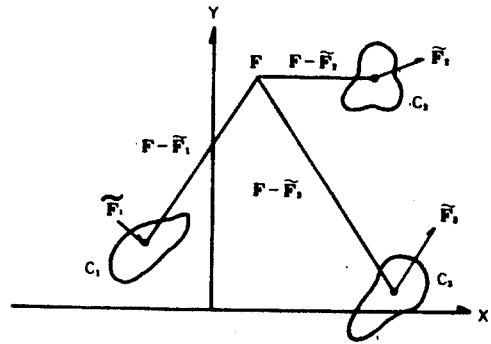


Fig. 2 2 dimensional feature space associated with C_1, C_2 , and C_3 .

$$f_{iN} = [\tilde{F}_{i0}, \tilde{F}_{i1}, \tilde{F}_{i2}, \dots, \tilde{F}_{i39}]' \quad i = 0, 1, 2, \dots, 9 \quad (6)$$

$$f_p = [F_{p0}, F_{p1}, F_{p2}, \dots, F_{p39}]' \quad (7)$$

式으로부터 discriminant function은 다음과 같다.

$$g_i(f_p) = \tilde{f}_{iN} f_p - \frac{1}{2} \|\tilde{f}_{iN}\|^2 \quad \begin{matrix} i = 0, 1, 2, \dots, 9 \\ N = 0, 1, 2, \dots, 39 \end{matrix} \quad (8)$$

IV. 實 驗

IV-1. 實驗에 使用한 文字

- (1) 對像書體 : 명조체, 太명조체, 太고딕체
- (2) 對像文字 : 各書體의 0~9의 30文字
- (3) 畫像크기 : 正規化된 80×80 表示

0 1 2 3 4 5 6 7 8 9

(가) 명조체

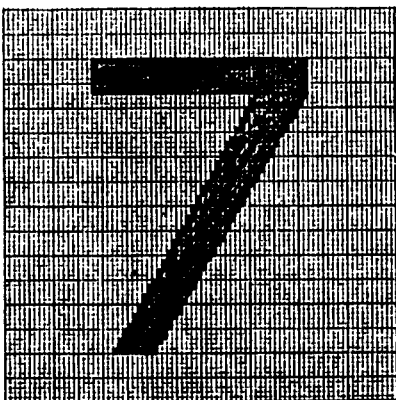
0 1 2 3 4 5 6 7 8 9

(나) 太명조체

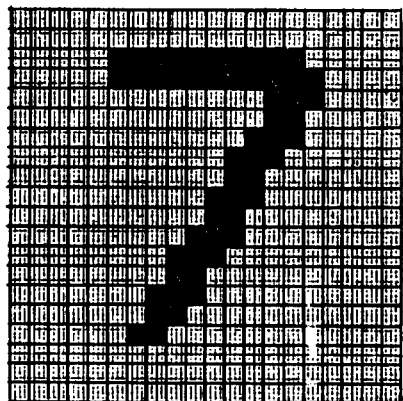
0 1 2 3 4 5 6 7 8 9

(다) 太고딕체

그림 3. 사용서체 및 對像文字



(가) 80×80 画像



(나) 20×20 画像

그림 3. 画像의 形狀

IV-2. stroke density 軸出

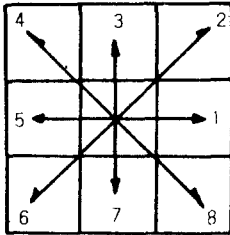


그림 4. 8方面

80×80의 畫素로 形成된 畫像을 對像物을 “1”로 背景을 “0”로 바꾸어 正方形 Matrix 형태로 재구성한후 出發點을 選定하여 X軸방향과 Y軸方向의 2方向에 對해서 추적했다.

그림4의 8方向에서 “1”方向과 “7”方向만 利用하여 X軸의 stroke density와 Y軸의 strok density를 計수했다.

⊙한개의 stroke density의 존재조건

- 010
- 00...010...0
- 01.....10

以上の 方法에 의해 計수한 stroke density 數를 20近似點으로 하기爲해 平均하여 낸 값이 표1-3과 같다.

그리고 표4는 같은 Category 內에서의 strake density의 값을 平均하여 소수점 以下 2째자리까지 구한 값이 表1-2와 같다.

표 1 서체 : 명조체

| 구분 | “1” 方向추적 Stroke density 數 | | | | | | | | | | | | | | | | | | | | |
|----|---------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|---|
| | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ | |
| 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 2 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 4 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 1 | 1 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 4 | 4 | 3 | 3 | 4 | 4 | 4 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 구분 | “7” 方向 Stroke density 數 | | | | | | | | | | | | | | | | | | | | |
| | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ | |
| 1 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 3 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 |
| 4 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 5 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 6 | 0 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 7 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 8 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 0 |
| 9 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 1 | 2 | 2 | 2 | 1 | 0 | 0 |

표 2 서체 : 太명조체

| 구분 | 방향 | "1" 方向추적 Stroke density 數 | | | | | | | | | | | | | | | | | | | |
|----|----|---------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ |
| 0 | | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| 1 | | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 3 | 4 | 3 | 2 | 1 | 0 | 0 | 0 | 0 |
| 4 | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 1 | 1 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |
| 5 | | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 6 | | 0 | 0 | 0 | 0 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 0 | 0 | 0 | 0 |
| 7 | | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 8 | | 0 | 0 | 0 | 1 | 2 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 3 | 2 | 1 | 0 | 0 | 0 |
| 9 | | 0 | 0 | 0 | 0 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 구분 | 방향 | "7" 方向 Stroke density 數 | | | | | | | | | | | | | | | | | | | |
| | | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ |
| 0 | | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 |
| 1 | | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 2 | | 0 | 0 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 0 | 0 | 0 |
| 3 | | 0 | 0 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 0 | 0 | 0 |
| 4 | | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 5 | | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 0 | 0 | 0 |
| 6 | | 0 | 0 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 7 | | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| 8 | | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 0 | 0 |
| 9 | | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 0 | 0 |

표 3 서체 : 太고딕

| 구분 | 방향 | "1" 方向추적 Stroke density 數 | | | | | | | | | | | | | | | | | | | |
|----|----|---------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ |
| 0 | | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 3 | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| 4 | | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| 6 | | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 8 | | 0 | 0 | 0 | 0 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |
| 9 | | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| 구분 | 방향 | "7" 方向 Stroke density 數 | | | | | | | | | | | | | | | | | | | |
| | | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ |
| 1 | | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 |
| 1 | | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 2 | | 0 | 0 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 3 | | 0 | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 0 | 0 | 0 |
| 4 | | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 5 | | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 0 | 0 | 0 |
| 6 | | 0 | 0 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 |
| 7 | | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 8 | | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| 9 | | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |

표 4 各書體의 평균 stroke數

| 방향 | "1" 方向추적 Stroke density 數 | | | | | | | | | | | | | | | | | | | |
|----|---------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ |
| 0 | 0 | 0 | 0 | 0.7 | 1.00 | 1.07 | 1.67 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.33 | 1.00 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0. | 0. | 1.33 | 2.00 | 2.00 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0. | 1.07 | 3.00 | 2.67 | 2.67 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.67 | 1.00 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0.33 | 1.33 | 2.00 | 2.00 | 2.33 | 3.00 | 3.00 | 3.33 | 3.33 | 3.00 | 3.00 | 2.00 | 1.33 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0.07 | 1.33 | 1.07 | 2.60 | 2.00 | 2.00 | 2.67 | 2.67 | 1.00 | 1.00 | 1.67 | 1.67 | 0.33 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0. | 1.00 | 2.00 | 2.67 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.33 | 1.67 | 0.33 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0.33 | 1.00 | 2.33 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 1.67 | 1.33 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0. | 0.07 | 1.33 | 1.33 | 1.67 | 2.00 | 2.00 | 2.00 | 1.67 | 1.67 | 1.00 | 1.00 | 0.67 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0.33 | 2.00 | 3.00 | 2.67 | 3.67 | 3.33 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | 0.67 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0. | 1.33 | 2.33 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | 1.33 | 1.00 | 0 | 0 | 0 | 0 |
| 방향 | "7" 方向 Stroke density 數 | | | | | | | | | | | | | | | | | | | |
| | F ₁ | F ₂ | F ₃ | F ₄ | F ₅ | F ₆ | F ₇ | F ₈ | F ₉ | F ₁₀ | F ₁₁ | F ₁₂ | F ₁₃ | F ₁₄ | F ₁₅ | F ₁₆ | F ₁₇ | F ₁₈ | F ₁₉ | F ₂₀ |
| 0 | 0.33 | 1.33 | 1.67 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.67 | 1.33 | 0.67 | 0.33 | 0 |
| 1 | 0 | 0.33 | 0.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.33 | 0 |
| 2 | 0 | 0 | 1.00 | 1.67 | 2.00 | 2.00 | 1.67 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 | 1.67 | 1.67 | 1.33 | 0.33 | 0.33 | 0 |
| 3 | 0 | 1.00 | 1.67 | 2.00 | 1.67 | 1.33 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 | 1.67 | 1.33 | 0.33 | 0 | 0 |
| 4 | 0 | 1.00 | 1.00 | 1.33 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.67 | 1.33 | 1.33 | 1.00 | 1.00 | 1.00 | 0.33 | 0.33 | 0 |
| 5 | 0 | 0 | 0.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.33 | 1.67 | 2.67 | 1.33 | 1.00 | 1.00 | 1.67 | 2.00 | 1.67 | 1.33 | 0.67 | 0.33 | 0 |
| 6 | 0 | 0.33 | 1.33 | 1.67 | 2.00 | 1.33 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.67 | 2.00 | 2.00 | 2.00 | 1.67 | 1.67 | 1.00 | 0.33 | 0 |
| 7 | 0 | 0 | 0.67 | 1.00 | 1.33 | 1.67 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.67 | 0.33 | 0 |
| 8 | 0 | 0.33 | 1.33 | 1.67 | 2.00 | 2.00 | 2.00 | 2.00 | 1.67 | 1.00 | 1.67 | 1.00 | 2.00 | 2.00 | 2.33 | 1.67 | 1.67 | 1.00 | 0.33 | 0 |
| 9 | 0 | 0 | 1.00 | 1.67 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.67 | 2.33 | 1.67 | 1.67 | 1.33 | 1.33 | 1.67 | 1.67 | 1.00 | 0.33 | 0 |

IV-3. 認識

認識對象 文字에 對한 特徵 Vector 를 垂直方向을 座標軸으로한 경우만을 取했을때, 水平方向을 座標軸으로한 경우만을 取했을때, 그리고 水平및 垂直方向 모두를 取했을 때의 구분 認識을 했다.

표 5

| 서체 | 명 조 체 | | | 太 명 조 체 | | | 太 고 덕 체 | | | 계 |
|-------------------------------|-------|------|--------|---------|------|--------|---------|------|--------|-------|
| | 인식율 | 오인식율 | 인식못한문자 | 인식율 | 오인식율 | 인식못한문자 | 인식율 | 오인식율 | 인식못한문자 | |
| "1" 方向 Stroke 특징사용 | 80% | 20% | 5 | 90% | 10% | 5 | 60% | 40% | 0.2 | 76.7% |
| "7" 方向 Stroke 특징사용 | 100% | | | 100% | | | 70% | 30% | 2 | 90% |
| "1", "7" 方向 Stroke 특징사용 | 100% | | | 100% | | | 90% | 10% | 2 | 96.7% |

V. 結 論

가로 80點×세로 80點의 正規化된 2進畫像데이터를 이용하여 文字細線化를 하지 않고 8方向中에서 "1" 방향과 "7" 방향만의 stroke density 를 계수하여 그 값을 特徵Vector 로하여 숫자 認識을

하였다.

숫자文字에 대한 stroke density 특성은 “1”方向보다는 “7”方向이 많은 情報를 갖고 있음을 알 수 있고, 各各의 方向 特徵보다는 2方向의 特徵이 文字自體의 特徵情報를 많이 함유하고 있음이 판명되었다.

그리고 stroke density 特徵으로 國漢文 文字認識의 大分類에 有效하게 應用 할 수 있으며 次後 보다 精確한 正規化에 關한 研究와 畫像入力 裝置의 한글 文字物 認識에 對해 研究가 要求된다.

참 고 문 헌

- (1) 中野, 中田: 週邊分析 spectral에 의한 漢字의 認識” 數學論(D), 56-D, 3, pp.146-153
- (2) 山本, 中田: 階層的 Pattern 정합에 의한 文字認識의 기초” 數學論(D), 56-D, 3, pp. 365-372
- (3) H.Tamura “A Comparison of line Thinning Algorithms” Proceeding of the forth international Joint Covference on Pattern Recoqnition.
- (4) N.Ahmed · K.R.Rao : “Orthogonal Trans forms for Digital Signal Processing” Springer-Verlag Berlin · Heidelberg · New York 1975 pp.225~253.
- (5) 安居院 猛 · 中嶋正之: “Compwter 畫像處理” 電子科學시리즈, 1984. pp. 55-88.