

# 임계 획 밀도를 이용한 한글, 한자, 영문구분 (A Distinction of the Korean Character, Chinese Character and English Character using the Threshold Stroke Density)

원 남식\*  
(Nam-Sik Won)

**요 약** 다중 문자 환경의 문서인식 시스템에서 문자를 인식하기 전에 문자의 종류를 먼저 구분하는 것은 인식률의 향상에 중요한 요인이 된다. 각 나라의 문자는 그 문자마다 고유의 구성상의 다양한 특징을 가진다. 본 연구에서는, 문자를 구분하기 위한 방법으로 획 밀도 값을 이용하였고, 대상 문자는 한글, 영문과 한자로 하였다. 다양한 형태의 활자가 사용되는 문서에 적용하기 위해 입력 데이터는 정규화 과정을 거친 후 처리되었다. 제안된 방법은 80% 이상의 높은 확률로 구분이 가능함을 실험 결과로써 입증하였다.

**Abstract** It is an important factor to distinguish the kind of the character for increasing recognition rate before the character recognition in the document recognition system composed of the multi-font and multi-letters. All the letters of each country have a various unique characteristic in the each composition. In this paper, we used the stroke density as a method to distinguish the letter, and it has been adopted Korean, English and Chinese character. Input data is processed by the normalization to adopt multi-font document. Proposed method has been proved by the results of experiment the fact that the distinction probability of the Korean and English is more than 80%.

## 1. 서 론

국내외에서 문서인식에 관한 연구는 매우 활발하게 이루어지고 있으나, 대부분이 하나의 특정 문자에 국한된 연구이다. 그러나 국내에서 사용되는 일반적인 문서 환경은 한문이나 영문과의 혼용이 불가피한 상태이므로, 이러한 혼용된 문서를 인식하기 위한 연구는 중요한 과제가 된다. 현재 한글, 한자, 영문 구분에 관한 연구는 국내외에서 거의 수행되지 않고 있는 실정이나 실제의 문서인식 시스템에서는 이를 자동으로 구분하여 인식하는 기능은 매우 요구되고 있다.

각 나라의 문자는 그 문자마다 고유의 구성상의 다양한

특징을 가지므로, 다중 문자 환경의 문서인식 시스템에서 문자를 인식하기 전에 각 문자의 특징을 이용하여 문자의 종류를 먼저 구분하는 것은 인식률과 인식 속도의 향상에 중요한 요인이 된다.

문서인식을 위한 문서 구조 분석은 연결 요소 분석 방법 [1]과 인접 선분 밀도 분석 방법 [2,3], 런 길이 평활화 방법, 투영 윤곽 분석 방법, Fourier 변환에 의한 방법, 원형 정합에 의한 방법, 교차 횟수 분석 방법 등이 있다 [4,5,6]. 문서인식을 위한 전처리 과정으로서 정규화 과정이 있다. 일반적인 정규화의 목적은 외접 다각형 또는 모멘트 [7] 등을 이용하여 영상을 일정한 위치, 크기, 기울기 등을 갖는 영상으로 변환하는데 있다. 정규화 방법에는 선형 변환과 비선형 변환이 있다. 선형 변환은 선형성을 보존하는 특성이 있으며, 수학적 표현이 쉬운 반면, 불규칙적이고 부분적으로 발생하는 형태 왜곡을 보정하기 위해 다양한 비선형 형태 정규화 방법이 제안되었다 [6,8,9,10,11,12].

\*경일대학교 컴퓨터공학과

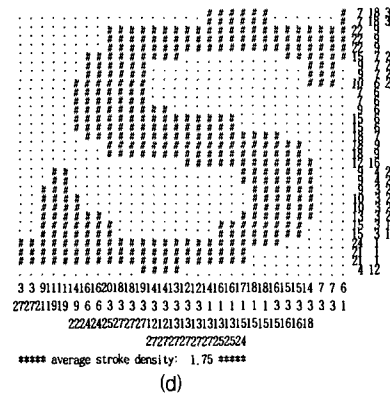
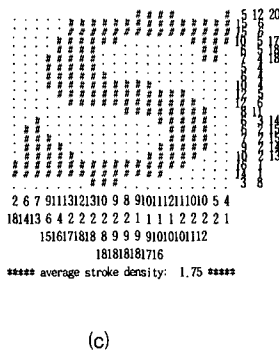
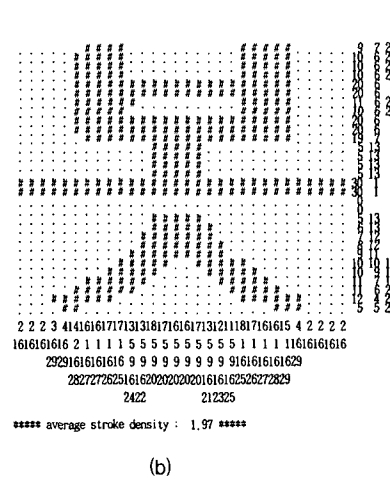
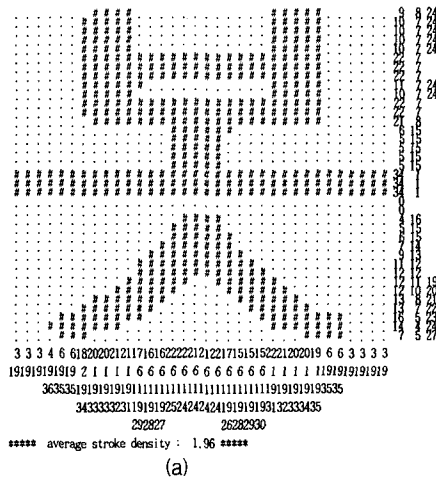
실제의 문서 인식 시스템에서 문자의 인식율을 향상하기 위해서는 먼저 인식 대상 문자의 종류를 알아야만 빠르고 정확한 인식이 가능해진다. 현재 국내에서 사용되는 대부분의 문서들은 국문과 영문 그리고 한문이 혼용되어 사용되고 있으므로, 이들 문자를 먼저 구분한 후 인식함으로써 인식율과 인식 속도를 보다 향상시킬 수 있다. 한글, 영문, 한문은 각각 그 글자의 구성상의 특징에 의해 각 문자마다 획 밀도의 분포가 다르게 나타나므로 획 밀도를 분석하면 각 문자의 종류를 높은 확률로 구분할 수 있다. 본 연구에서는 정규화 과정을 거친 후 획 밀도를 분석하여 각 문자를 구분하는 방법이다. 모든 문자는 매우 다양한 형태의 활자로 구성되어 사용되므로 이를 완벽하게 구분할 수 있는 방법은 처음부터 존재할 수 없다. 그러므로, 일정한 수준의 실용성 있는 범위에서의 구분이 가능하면 된다. 또한 구분을 위한 알고리즘을 수행하는데 소요되는 시간은 인식 시간에 대비하여 상대적으로 무시될 수 있는 범위라

야 한다. 본 연구에서는 높은 확률로 구분이 가능하고, 적은 경비로 처리할 수 있는 획 밀도를 이용한 한글, 영문, 한자를 구분하기 위한 방법 개발을 연구 목적으로 한다.

실험 방법은 많이 사용되고 있는 한글 폰트 5종에 대해서 각각 한글 완성형 2350자에 대하여 획 밀도 측정을 수행하였고, 영문 폰트는 일반성 있는 5종의 대문자, 소문자 폰트에 대해 News Week지에 게재된 영문을 이용하여 획 밀도 측정을 수행하였으며, 한문은 상용 1,800자를 명조체에 대하여 수행하였다.

## 2. 정규화

형태 정규화 방법은 선형 형태 정규화 방법과 비선형 형태 정규화 방법으로 분류되고[13], 비선형 형태 정규화 방



<그림 1> 입력 데이터의 정규화 (a), (c): 입력 데이터 (b), (d): 정규화 데이터

<Fig-1> Normalization of the input data (a), (c): input data (b), (d): normalization data

법은 투영하는 특징의 종류에 따라 점밀도[12]를 이용하는 방법과 획 밀도를 이용하는 방법으로 구분될 수 있다. 획 밀도를 이용하는 방법은 교차 횟수(crossing count)에 바탕을 둔 획 밀도[10]를 이용한 방법, 획 간격(line interval)에 바탕을 둔 획 밀도[9]를 이용하는 방법, 그리고 내접원(inscribed circle)에 바탕을 둔 획 밀도[11]를 이용하는 방법으로 구분된다. 본 논문에서의 정규화는 선형 형태 정규화 방법에 의해 구현하였다.

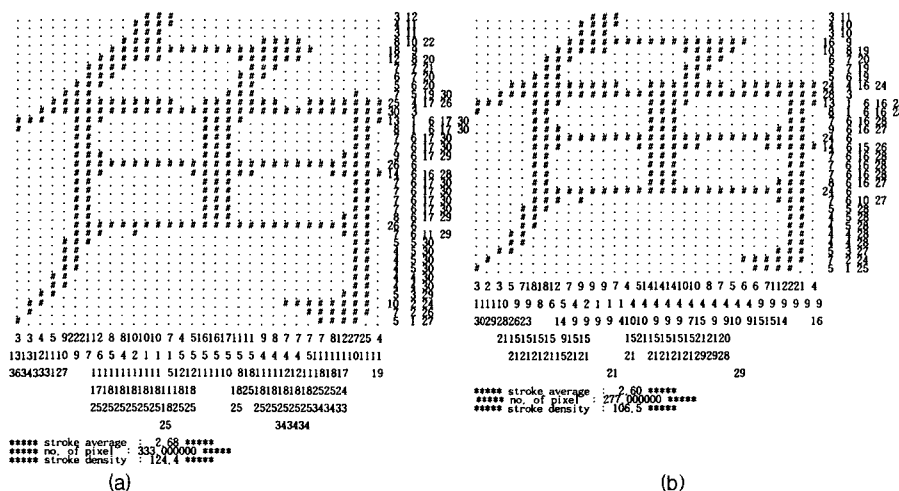
대부분의 인쇄소에서 사용되는 활자들은 동일한 글씨체라도 서로 미세한 차이를 나타낼 수 있으므로, 이러한 활자들에 의해 인쇄된 문서를 인식하는데 일정한 통계 치의 적용은 오차의 범위를 더 크게 하는 결과가 될 수도 있다. 그러므로 문자 구성의 특징을 추출하기 위한 방법으로서, 통계적인 수치에 의해 처리하기 위한 방법에서는 정규화 과정이 필요하게 된다. 본 연구에서는 입력된 데이터에서 특징을 추출하는데 적당한 크기의 데이터로 처리하기 위해 정규화 과정에서 데이터의 크기를 가변으로 조절하여 정규화 된 데이터를 구하도록 하였다.

정규화 된 데이터의 크기에 관련된 것으로서 데이터가 크면 특징 데이터의 추출에는 유리하나 처리 시간이 문제가 되고, 크기가 적으면 처리 시간은 단축되나 정규화 과정에서 입력된 데이터의 일부가 소멸될 수도 있다. 문자의 구성 획이 복잡한 경우 획의 일부가 사라질 수도 있으므로, 정규화 된 데이터의 크기는 특징 데이터를 얻기 위한 적절한 크기로 결정되어야 한다. <그림 1>은 한글과 영문에서의 정규화된 결과를 보이고, <그림 2>는 한자의 정규화된 결과를 보인다. 획 밀도를 얻기 위한 정규화 데이터의 크기는 30×30으로 하였다. 300dpi의 해상도로 데이터가 입

력될 때, <그림 1>과 <그림 2>에서와 같이 정규화 된 결과에서 한글과 한자는 축소되었고, 영문은 확대되었음을 알 수 있다. 그림에서 우측면과 밑면에 보이는 수치들의 의미는, 첫째 줄에 나타나는 수치는 수평 방향과 수직 방향으로 투영된 화소의 숫자를 나타내고, 두 번째 줄부터는 수평, 수직 방향에서 각각 획이 시작되는 위치를 순서대로 나타낸다. 획이 시작되는 위치는 화소가 백화소에서 흑화소로 바뀌는 점이 된다. 그리고 평균 획 밀도를 나타내었다.

### 3. 획 밀도 측정 방법

교차 횟수( $N_c$ )는 문서의 수평 또는 수직의 주사 방향에서, 백화소에서 흑화소로 바뀌는 점의 수 또는 흑화소에서 백화소로 바뀌는 점의 수를 각 주사선 별로 계산한 것이다. 본 논문에서의 교차 횟수에 대한 정의는 백화소에서 흑화소로 바뀌는 점의 수로 정의하였다. 그러므로 임의의 화소점에서의 수평 방향의 교차 횟수는 (식1)과 같고, 수직 방향의 교차 횟수는 (식2)와 같이 된다. 획 밀도는 교차 횟수를 수평 방향과 수직 방향으로 투영하여 이들 교차 횟수의 합을 주사한 화소 수로 나눈 값을 획 밀도( $\rho$ )로 정의하였다. 그러므로 수평 방향 획 밀도  $\rho_h(i)$ 는 (식3)과 같고, 수직 방향 획 밀도  $\rho_v(j)$ 는 (식4)와 같이 표현된다. 평균 획 밀도  $\rho_a$ 는 수평 방향 획 밀도와 수직 방향 획 밀도의 합을 평균한 값으로 (식5)와 같으며, 정규화 이전 영상에서의 평균 획 밀도  $\rho_{na}$ 는 (식6)과 같다. 임의 획 밀



<그림 2> 입력 데이터의 정규화 (a) 입력 데이터 (b) 정규화 데이터  
 <Fig-2> Normalization of the input data (a) input data (b) normalization data

도의 기준값을 임계 획 밀도  $\rho_t$ 라 하고, 이 값은 문자의 종류를 구분하기 위한 기준 값으로 사용될 수 있다. 문자의 종류를 구분하기 위한 임계 획 밀도 값은 많은 획 밀도 측정 데이터의 결과로서 결정할 수 있다.

$f(i, j)$ :  $I \times J$  크기의 입력 이진 영상( $i=1 \dots I, j=1 \dots J$ )

$f_n(m, n)$ :  $m \times n$  크기의 정규화 이진 영상

$N_c$ : 교차 횟수(crossing count)

$\rho$ : 획 밀도(stroke density)

$\rho_a$ : 평균 획 밀도(average stroke density)

$\rho_t$ : 임계 획 밀도(threshold stroke density)

$\rho_{na}$ : 정규화 이진 영상에서의 평균 획 밀도

임의 화소점에서 수평 방향 교차 횟수:

$$N_{hc}(i) = \sum_j f(i, j) \cdot f(i, j+1) \quad \dots\dots(1)$$

임의 화소점에서 수직 방향 교차 횟수:

$$N_{vc}(j) = \sum_i f(i, j) \cdot f(i+1, j) \quad \dots\dots(2)$$

수평 방향 획 밀도:

$$\rho_h(i) = \frac{\sum_j N_{hc}(i)}{I} \quad \dots\dots(3)$$

수직 방향 획 밀도:

$$\rho_v(j) = \frac{\sum_i N_{vc}(j)}{J} \quad \dots\dots(4)$$

평균 획 밀도:

$$\rho_a = \frac{\rho_h(i) + \rho_v(j)}{2} \quad \dots\dots(5)$$

정규화 된 이진 영상에서의 평균 획 밀도:

$$\rho_{na} = \frac{\rho_h(m) + \rho_v(n)}{2} \quad \dots\dots(6)$$

#### 4. 획 밀도 측정

획 밀도 측정을 위한 객관성 있는 데이터로서 한글은 한글 완성형 2350자를 명조, 신명조, 고딕, 궁서, 샘물체등의 5가지 폰트에 대하여 각각 실험하였고, 영문은 News Week지에서 문장을 발췌하여 이를 명조, 신명조, 고딕, 산세리프, 이탤릭체에 대하여 각 2500자에 대해 대문자와 소문자를 별도로 획 밀도를 측정하였으며, 한자에 대해서는 상용한자 1800자를 명조체에 대하여 측정하였다.

모든 문자들은 각 각 독특한 획 구성 형식을 가지므로

각 문자의 획 밀도 특성은 서로 다르게 나타나므로, 이러한 획 밀도 특성을 이용함으로써 문자의 구분이 가능하게 된다. 한글은 획의 구성이 영문에 비해 복잡하고, 한문에 비해서는 간단한 편이므로, 각 문자의 획밀도 특성이 서로 다르게 나타남을 알 수 있다.

한글에서는 대부분의 글자의 획 밀도( $\rho$ )는  $\rho_{na}$ 값이 1.9이상의 높은 값에 분포되어 있으므로, <표 1>에서는 임계 획 밀도값  $\rho_t$ 보다  $\rho_{na}$ 값이 적은 글자 수를 나타내었다. 영문에서는 대부분의 글자의  $\rho_{na}$ 값의 분포가 1.8미만의 낮은 값에 분포되어 있으므로 <표 3>와 <표 6>에서는 임계 획 밀도값  $\rho_t$ 보다  $\rho_{na}$ 값이 큰 글자 수를 나타내었다. 한문에서의 획밀도 분포는 <표 7>와 <표 10>에 나타내었고, 여기서는 획 밀도가 2.5 이상의 높은 값에 분포함을 알 수 있다.

한글은 <표 1>에서 Kf1은 명조, Kf2는 신 명조, Kf3은 고딕, Kf4는 궁서, Kf5는 샘물체 등의 5가지 폰트에 대하여 각각 실험하였고, <표 2>는 영문 대문자로서 Euf1은 명조, Euf2는 신 명조, Euf3은 고딕, Euf4는 산세리프, Euf5는 이탤릭체에 대하여 각각 실험하였고, <표 3>은 영문 소문자로서 Euf1은 명조, Euf2는 신 명조, Euf3은 고딕, Euf4는 산세리프, Euf5는 이탤릭체에 대하여 각각 실험하였다.

#### 4.1 한글의 획 밀도 분포

<표 1> 한글 2350자에 대한 획 밀도 분포( $\rho > \rho_{na}$ )

<Table 1> The stroke density distribution of 2350 characters in the Korean ( $\rho > \rho_{na}$ )

$\rho_t$ font	1.6	1.7	1.8	1.9	2.0	2.1
Kf1	18	41	92	165	302	497
Kf2	14	27	55	98	194	305
Kf3	9	23	56	103	205	356
Kf4	40	107	241	456	746	1014
Kf5	110	206	354	574	857	1156

<표 2> 한글 임계 획 밀도에 대한 구분율(%)

<Table 2> Distinction percentage by the threshold stroke density in the Korean character(%)

$\rho_t$ font	1.6	1.7	1.8	1.9	2.0	2.1
Kf1	99.2	98.2	96.1	93.0	87.2	76.5
Kf2	99.5	98.8	97.7	95.8	91.7	87.1
Kf3	99.6	99.0	97.6	95.6	91.3	84.9
Kf4	98.3	95.4	89.7	80.6	68.3	56.9
Kf5	95.3	91.2	84.9	75.6	63.5	50.8

## 4.2 영문의 획 밀도 분포

<표3> 영문 대문자 2500자에 대한 획 밀도 분포( $\rho_f < \rho_{na}$ )  
 <Table3> The stroke density distribution of the upper case 2500 characters in the English( $\rho_f < \rho_{na}$ )

$\rho_f$ font	1.7	1.8	1.9	2.0	2.1	2.2
Euf1	1168	661	397	91	56	14
Euf2	1220	990	782	211	60	37
Euf3	1115	701	145	2	0	0
Euf4	463	171	8	0	0	0
Euf5	610	150	73	59	27	2

<표4> 영문 소문자 2500자에 대한 획 밀도 분포( $\rho_f < \rho_{na}$ )  
 <Table4> The stroke density distribution of the lower case 2500 characters in the English( $\rho_f < \rho_{na}$ )

$\rho_f$ font	1.8	1.9	2.0	2.1	2.2	2.3
Elf1	630	339	47	0	0	0
Elf2	668	398	61	36	9	0
Elf3	326	166	13	4	0	0
Elf4	231	24	2	0	0	0
Elf5	512	191	38	8	0	0

<표5> 영문 대문자에서의 임계 획 밀도에 대한 구분율(%)  
 <Table5> Distinction percentage by the threshold stroke density in the upper case English character(%)

$\rho_f$ font	1.7	1.8	1.9	2.0	2.1	2.2
Euf1	53.3	73.6	84.1	96.4	97.7	99.4
Euf2	51.2	60.4	68.7	91.5	97.6	98.5
Euf3	55.4	72.0	94.2	99.9	100	100
Euf4	81.5	93.2	99.7	100	100	100
Euf5	75.6	94.0	97.1	97.6	98.9	99.9

<표6> 영문 소문자에서의 임계 획 밀도에 대한 구분율(%)  
 <Table6> Distinction percentage by the threshold stroke density in the lower case English character(%)

$\rho_f$ font	1.8	1.9	2.0	2.1	2.2	2.3
Elf1	74.8	86.4	98.1	100	100	100
Elf2	73.3	83.0	97.6	98.6	99.6	100
Elf3	87.0	93.4	99.5	99.8	100	100
Elf4	91.8	99.3	99.9	100	100	100
Elf5	79.5	92.4	98.5	99.7	100	100

## 4.3 한자의 획 밀도 분포

<표7> 상용 한자 1,800자에 대한 획 밀도 분포( $\rho_f=1.0\sim1.9$ )  
 <Table7> The stroke density distribution of 1,800 characters in the Chinese character( $\rho_f=1.0\sim1.9$ )

$\rho_f$	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$\rho_f < \rho_{na}$ character count	1799	1795	1792	1789	1785	1782	1770	1763	1753	1731

<표8> 상용 한자 1,800자에 대한 획 밀도 분포( $\rho_f=2.0\sim2.9$ )  
 <Table8> The stroke density distribution of 1,800 characters in the Chinese character( $\rho_f=2.0\sim2.9$ )

$\rho_f$	2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
$\rho_f < \rho_{na}$ character count	1702	1663	1632	1599	1547	1490	1412	1326	1249	1171

<표9> 상용 한자 1,800자에 대한 획 밀도 분포( $\rho_f=3.0\sim3.9$ )  
 <Table9> The stroke density distribution of 1,800 characters in the Chinese character( $\rho_f=3.0\sim3.9$ )

$\rho_f$	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
$\rho_f < \rho_{na}$ character count	1033	936	800	666	536	422	299	205	158	98

<표10> 상용 한자 1,800자에 대한 획 밀도 분포( $\rho_f=4.0\sim4.9$ )  
 <Table10> The stroke density distribution of 1,800 characters in the Chinese character( $\rho_f=4.0\sim4.9$ )

$\rho_f$	4.0	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
$\rho_f < \rho_{na}$ character count	62	35	28	7	3	1	0	0	0	0

## 4.4 획 밀도 고찰

서로 다른 종류의 문자는 획 밀도 값에 의해서 구분될 수 있다. 이를 구분하기 위한 획 밀도 값을 임계 획 밀도( $\rho_f$ )라 하고, 입력 데이터의 평균 획 밀도는 정규화 한 이진 영상에서의 평균 획 밀도( $\rho_{na}$ )로 표시한다. 한글, 한문과 영문의 획 밀도 특성은 한문이 가장 높게 나타나고, 한글은 중간 영역이고, 영문은 가장 낮게 나타나므로, 구분율이 가장 높은 임계 획 밀도 값을 적절히 설정하여 각 문자를 구분할 수 있다.

실험 결과로서 글자체에 따라서 획 밀도 값이 다르게 나타남을 알 수 있다. 한글에서는 <표 1>에서와 같이 명조, 신명조, 고딕체에서 획 밀도 특성이 양호하게 나타나고, 궁서체와 샘물체에서는 획 밀도 특성에 의한 구분율이 떨어짐을 알 수 있다. 그러나 이러한 글씨체는 일반적인 문서에서는 거의 사용되지 않고 있다. 명조, 신명조, 고딕체에서는 <표 2>에서와 같이 임계 획 밀도가 1.9에서 90%이상의 높은 확률로 구분할 수 있음을 알 수 있다. 영문에서는 <표 3>와 <표 4>에서와 같이 획 밀도가 1.9보다 적은

값에 대부분이 분포하고 있으며, 글씨체에 따라서 획 밀도 값의 변화가 심하게 나타남을 알 수 있다. 영문에서는 고딕체, 산세리프체, 이탤릭체가 획 밀도 분포 특성이 우수하게 나타났으며, 명조체와 신 명조체에서 획 밀도 분포 특성이 비교적 나쁘게 나타남을 알 수 있다. 획 밀도 측정에 의한 실험 결과 한글의 획밀도 분포는 1.9에서 2.6정도 나타났으며, 영문에서는 1.2에서 1.9정도의 범위로 나타났다. 한자에서는 획 밀도 분포는 1.5에서 4.5에 걸쳐서 나타나지만 대부분의 글자는 2.5에서 4.2에 분포함을 알 수 있다. 그러므로 한글, 한문, 영문의 획 밀도 특성 값이 서로 다른 영역에 80%에서 90% 이상이 분포하므로 이러한 획 밀도 분포 특성에 의해 한글, 한자, 영문의 구분이 80% 이상의 확률로 구분 가능함을 알 수 있다.

## 5. 결 론

혼용된 문서 인식 시스템에서 문자를 인식하기 위해서는 인식 대상 문자의 종류를 먼저 파악한 후, 적합한 알고리즘에 의해서 인식하는 것은 문자의 인식률과 인식 속도 향상에 큰 영향을 가져올 수 있다. 그러나 이러한 문자 구분을 위한 알고리즘의 수행에 많은 경비를 치르게 되면 문자 인식 속도에 영향을 미치게 되므로, 적은 경비에 의해 경제적인 범위의 확률로 판단되면 실용성 있는 방법이 될 수 있다. 본 연구에서는 문자의 종류를 구분하기 위한 방법으로 획 밀도 값을 이용하였고, 대상 문자는 한글, 한문과 영문으로 한정하였다. 그리고 다양한 형태의 활자가 사용되는 문서에 적용하기 위해 입력 데이터는 정규화 과정을 거친 후 처리되었다. 문자인식의 전 처리 과정에서 적절한 임계값 밀도 값의 범위를 설정한 후 해당 문자의 획밀도 값의 범위와 비교하여 문자의 종류를 구분할 수 있다. 그러나 이는 완전한 구분이 되는 것이 아니고, 문자의 획밀도 대역이 서로 겹쳐지는 부분에서는 정확한 구분이 어려우므로, 일정한 확률로 구분이 가능해진다. 그러므로, 문자가 서로 중복되는 대역의 범위가 적으면 문자 구분을 위한 경제성이 있는 방법으로 생각된다. 문자인식에서 문자 종류의 구분은 필연적이기도, 매우 어려운 문제이다. 그러므로, 어느 정도 높은 확률로 구분되면 실용 가능성 있는 방법으로 생각된다. 제안된 방법은 빠른 수행시간으로 처리되며, 80% 이상의 높은 확률로 한글과 한자와 영문의 구분이 가능함을 실험 결과로서 입증하였다. 추후 연구과제로서는, 문자를 구분하는데 획 밀도값 외에 구분 확률을 높이기 위한 보조적인 수단으로서, 각 문자의 독특한 구성상의 특징을 이용하여 획 밀도 측정에서 잘못 구분되는 경우를 극복할 수 있는 연구가 수행되어야 한다.

## 참 고 문 헌

- [1] L. A. Fetcher and R. K. Kusutri, "A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 10, No.6, 1988, pp.910-918.
- [2] O. Iwaki, H. Arakawa, "A Segmentation Method Based on Office Document Hierarchical Structure," Proc. 1987 IEEE Int. Conf. on Systems, Man and Cybernetics, 1987, pp.759-763.
- [3] K. Kubota, O. Iwaki and H. Arakawa, "Document Understanding System," Proc. 7th Int. Conf. on Pattern Recognition, 1984, pp.612-614
- [4] T. Akiyama and N. Hagita, "Automated Entry System for Printed Documents," Pattern Recognition, Vol.23, No.11, 1990, pp.1141-1153.
- [5] J.-C. Oriot, D. Barba and J.-C. Salome, "Address Block Location Method Based on Transition Analysis Approach: Design and Evaluation on Flats Objects," Proc. 1st Int. Conf. on Document Analysis and Recognition, Saint-Malo, France, Sep. 1991, pp.665-673.
- [6] 이성환, "문자 인식," 1994, pp.89-113
- [7] R. G. Casey, "Moment Normalization of Handprinted Character," IBM Journal of Research and Development, Vol. 14, Sep. 1970, pp.548-557.
- [8] Z. C. Li, T. D. Bui, Y. Y. Tang and C. Y. Suen, Computer Transformation of Digital Images and Patterns, World Scientific Press, 1989.
- [9] J. Tsukumo and H. Tanaka, "Classification of handprinted Chinese Characters Using Nonlinear Normalization Methods," Proc. 9th Int. Conf. on Pattern Recognition, Rome, Italy, Nov. 1988, pp.168-171.
- [10] H. Yamata, T. Saito and K. Yamamoto, "Line Density Equalization - A Nonlinear Normalization for

Correlation Method," Trans. IECE of Japan, Vol. J67-D, No.11, Nov. 1984, pp.1379-1383.

[11] H. Yamata, K. Yamamoto and T. Saito, "A Nonlinear Normalization for Handprinted Kanji Character Recognition - Line Density Equalization," Pattern Recognition, Vol.23, No.9, pp.1023-1029.

[12] Y. Yamashita, K. Higuchi, Y. Yamata and Y. Haga, "Classification of Handprinted Kanji Characters by the Structured Segment Matching Method," Pattern Recognition Letters, Vol.1, No.5, pp.475-479.

[13] 이성환, 박정선, "대용량 필기체 문자 인식을 위한 비선형 형태 정규화 방법의 정량적 평가," 대한전자공학 회논문지, 제30권 B편 제9호, 1993년 9월, pp.896-905.



#### 원 남 식

1974년 인하공대 전자과 졸업(학사)  
1984년 영남대학교 공대 전자과(석사)  
1986년 영남대학교 공대 전산공학과(박사)  
1976년-1977년 한국과학기술연구소  
1978년-1981년 한국전자기술연구소  
1981년-현재 경일대학교 컴퓨터공학과  
관심분야 : 문자인식, 세션화 알고리즘, 컴퓨터 그래픽스