

# 표준 문자 패턴과의 미적 평가를 통한 한글 문자 익히기 시스템

한 군 희 \*\*, 조 동 욱 \*, 전 병 민 \*\*\*

\*\* : 대천 대학 \* : 독립 충북과학대학교 \*\*\* : 충북대학교

e-mail : ducho@ctech.ac.kr

## Hanguel Character Learning System by Beauty Evaluation from Standard Character Pattern

K. H. Han \*\* D. U. Cho \* & B. M. Jun \*\*\*

\* : Chungbuk Provincial University of Science & Technology

\*\* : Daechon College

\*\*\* : Chungbuk University

### 요 약

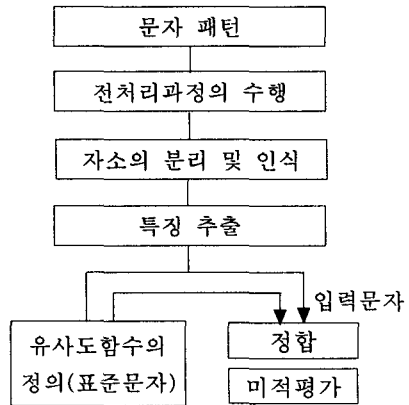
필기체 서체 인식은 온라인 문자 인식 시스템에서 주로 사용되는 시스템이다. 또한 오프라인 문자 인식 시스템은 문자 인식에만 초점이 맞추어져 있는 상황이다. 본 논문에서는 오프라인 방식으로 기초의 문자 인식에만 머물던 시스템을 문자 익히기까지 행할 수 있는 시스템으로 확장하는 방법을 제안 하고자 한다. 이를 위해 신명조체 80포인트에 대한 표준문자 패턴을 생성하고, 유사도함수를 정의하며 이를 통해 입력 문자 패턴과의 유사성을 계산하여 문자 익히기를 행하고자 한다.

### 1. 서론

패턴인식은 계산기의 급속한 처리속도 향상으로 그 적용 분야가 대단히 확대된 분야이다[1]~[3]. 이중 입력 문서의 자동화를 위한 문자 인식은 인쇄체 문자를 뛰어넘어 실시간으로 필기체를 인식하는 수준에 이르고 있는 실정이다. 또한 문자 인식은 그 입력 방식에 따라 오프라인 인식과 온라인 인식으로 나뉘어 진다. 그러나 오프라인 인식이나 온라인 인식 모두 해당 문자의 인식에만 초점이 맞추어져 있다. 근래 계산기(주로 PC)를 통한 학습 프로그램의 개발은 사회적 요구사항이다. 특히 한글 문자의 자동 익히기 프로그램은 어린 학생이나 외국인들을 위해 대단히 필요한 분야이다. 이를 위해 본 논문에서는 한글의 특정 서체를 익히고 시스템을 제안하고자 한다. 이는 추후 익히고자 하는 특정 서체 익히기와 서명 인식에도 확대 적용 할 수 있는 분야로 여겨지며 실험에 의해 본 논문의 유용성을 입증하고자 한다.

### 2. 본 시스템 구성

본 시스템의 구성은 아래 [그림 1]과 같다.



[그림1] 본 시스템의 구성

### 3. 전처리과정의 수행

전처리과정은 잡음 제거, 경계선 추출 그리고 세션화(thinning)과정으로 나뉘어 진다. 이중 잡음 제거와 경계선 추출은 Cho의 방법[4]으로 행하였으며 세션화는 Wakayama [5]의 방법으로 행하였다.

### 4. 히스토그램의 projection profile

특성에 의한 자소의 분리 및 인식

세션화된 문자 패턴에 대한 X축, Y축으로 Projection profile을 구한다. 즉, X축, Y축 Projection profile에 의해 자소를 분리해 내고 이의 특성을 이용하여 <표 1> 과 같은 방법으로 자소의 인식을 행한다. X축, Y축 projection profile의 수식은 식(1), 식(2)와 같다.

$$H_X(j) = \sum_{i=1}^n Y(i) \quad (\text{식 1})$$

$$H_Y(i) = \sum_{j=1}^m X(j) \quad (\text{식 2})$$

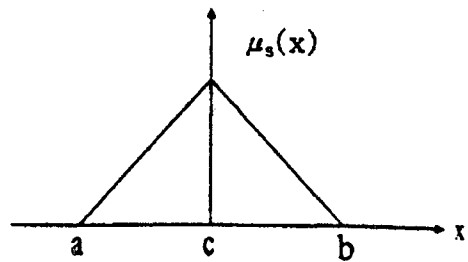
<표 1> 자소의 인식 예

X 히스토그램	Y 히스토그램	대상 장소
0	0	ㅅ, ㅇ
0	1	ㅈ, ㅡ
0	2	ㅎ, ㅊ
1	0	ㅇ, ㅣ
1	1	ㄱ, ㄴ, ㄷ, ㄹ, ㅌ, ㅍ
1	2	ㄷ, ㅋ, ㅌ
1	3	ㅌ
2	1	ㄱ, ㅅ, ㅈ, ㅊ
2	2	ㅊ, ㅌ, ㅍ, ㅎ, ㅋ
2	3	ㄷ

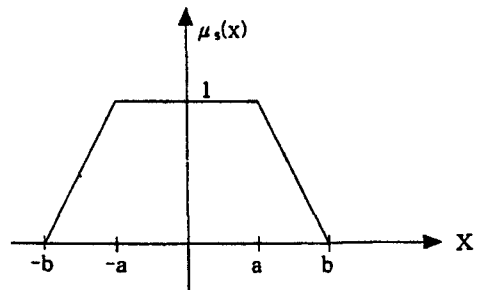
### 5. 특징 추출 및 유사도 함수의 계산

이제 특징 추출을 해야하는데 특징 추출은 획 사이의 특징 추출과 획들 사이의 특징 추출로 나누어 진다. 또한 문자 패턴에 대한 특징 추출 값을 바탕으로 유사도 함수를 정의하여 입력문자 패턴과의 유사도를 측정하여 미적 평가 (beauty evaluation)를 행한다. 획에 대한 특징

추출 요소는 획의 시작점과 끝점 사이의 각으로 정한다. 획들 사이에 대한 특징 추출은 획들 사이의 이루는 각(intersection angle)과 길이비(length ratio)로 한다. 마지막으로 모음의 경우는 교차점의 위치에 대한 특징 추출 요소값을 선정한다. 교차점의 위치에 대한 유사도함수는 삼각 퍼지수의 형태로, 나머지 특징값에 대한 유사도함수는 사다리꼴 퍼지수의 형태로 정의하며 이를 아래 그림으로 나타내었다.



[그림 2] 교차점에 대한 유사도함수



[그림 3] 자음에 대한 유사도함수

그림2에서 a와 b는 교차 획에 대한 시작점과 끝점을, c는 중간점을 나타낸다(1교차점의 경우).

### 6. 실험 및 고찰

본 논문에서의 실험은 IBM-PC상에서 C언어를 사용하여 구현하였다. 실험에 사용한 표준 문자 패턴은 신명조체 80포인트이다. 입력 문자 패턴은 흰 종이 위에 싸인펜으로 쓴 문자를 스캐너로 읽어서 처리하였다. [그림 4]가 표준 문

자 영상을 <표 2>가 표준 문자 영상에 대한 특징 추출 결과를 나타낸다. 표준 문자 영상에 대한 유사도함수 정의 예를 'ㄱ'에 대하여 <표 3>에 나타내었다. 또한 [그림 5]에 다양한 입력 문자 영상의 예를 <표 4>에 그 중 'ㄱ'에 대한 특징 추출 결과를 나타내었다. 최종적으로 <표 5>, <표 6>, <표 7>에 각각 'ㄱ', 'ㅂ', 'ㄴ'에 대한 미적 평가 결과를 나타내었다. 지금까지의 실험 결과는 우리가 의도한대로 결과치가 나옴을 확인할 수 있었다. 그러나 앞으로 다양한 표준 문자 패턴에 대한 실험 수행과 제안한 방법을 상용화하기 위한 노력이 지속적으로 경주되어야 하리라 여겨진다.



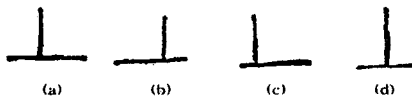
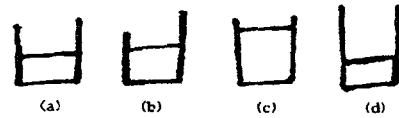
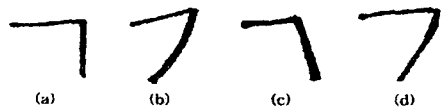
[그림 4] 표준 문자 영상

<표 2> 표준 문자 영상에 대한 특징 추출 결과

대상 문자	획의 특징추출	획들 사이의 특징 추출	
ㄱ	0, 90	Intersection Angle	90
		Length Ratio	1.0
ㅂ	90, 0	Intersection Angle	90, 180, 90 90, 180, 90
	90, 0	Length Ratio	0.882, 1.117, 0.882 1.266, 1.0, 0.789
ㄴ	90, 0	Intersection Angle	90
		Length Ratio	0.5
		Cross Point	0.5

<표 3> 'ㄱ'에 대한 유사도 함수

1	0 for $x \leq -20$ or $x \geq 20$ , $\frac{-20-x}{-10}$ for $-20 \leq x \leq -5$ ,
	1 for $-5 \leq x \leq 5$ , $\frac{x-20}{-15}$ for $5 \leq x \leq 20$
2	0 for $x \leq 70$ or $x \geq 110$ , $\frac{70-x}{-15}$ for $70 \leq x \leq 85$ ,
	1 for $85 \leq x \leq 95$ , $\frac{x-110}{-15}$ for $95 \leq x \leq 110$
3	0 for $x \leq 75$ or $x \geq 105$ , $\frac{75-x}{-10}$ for $75 \leq x \leq 85$ ,
	1 for $85 \leq x \leq 95$ , $\frac{x-105}{-10}$ for $95 \leq x \leq 105$
4	0 for $x \leq 0.6$ or $x \geq 1.4$ , $\frac{0.6-x}{-0.2}$ for $0.6 \leq x \leq 0.8$ ,
	1 for $0.8 \leq x \leq 1.2$ , $\frac{x-1.4}{-0.2}$ for $1.2 \leq x \leq 1.4$



[그림 5] 입력 문자 영상

<표 4> 'ㄱ' 입력 문자 영상에 대한 특징 추출 결과

대상 문자	획의 특징추출	획들 사이의 특징 추출	
(a)	90.6, 3.2	Intersection Angle	87.4
		Length Ratio	0.6744
		Cross Point	0.419
(b)	91.3, 5.6	Intersection Angle	85.7
		Length Ratio	0.515
		Cross Point	0.666
(c)	93.7, 3.2	Intersection Angle	90.5
		Length Ratio	0.533
		Cross Point	0.26
(d)	92.8, 6.2	Intersection Angle	86.6
		Length Ratio	1.353
		Cross Point	0.471

<표 5> ‘ㄱ’에 대한 미적 평가 결과

대상 문자	미적 평가 결과
(a)	1.0, 1.0, 1.0, 1.0 BE = 1.0
(b)	0.553, 0.0, 0.0, 0.5 BE = 0.25
(c)	1.0, 0.0, 0.0, 1.0 BE = 0.5
(d)	0.918, 0.0, 0.0, 1.0 BE = 0.4795

<표 6> ‘ㅂ’에 대한 미적 평가 결과

대상 문자	미적 평가 결과
(a)	1.0, 0.98, 1.0, 0.88, 1.0, 1.0, 0.99, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0 BE = 0.99
(b)	1.0, 0.88, 1.0, 1.0, 0.94, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.365, 1.0, 0.0, 0.0, 0.455 BE = 0.79
(c)	0.847, 1.0, 1.0, 1.0, 1.0, 0.753, 1.0, 0.98, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0 BE = 0.973
(d)	1.0, 0.78, 1.0, 0.485, 1.0, 1.0, 1.0, 1.0, 1.0, 0.468, 0.7, 1.0, 0.66, 0.0, 1.0, 0.975 BE = 0.817

<표 7> ‘ㄴ’에 대한 미적 평가 결과

대상 문자	미적 평가 결과
(a)	1.0, 1.0, 1.0, 1.0, 0.838 BE = 0.967
(b)	1.0, 0.96, 1.0, 1.0, 0.668 BE = 0.925
(c)	1.0, 0.0, 1.0, 1.0, 0.52 BE = 0.904
(d)	1.0, 0.92, 1.0, 0.0, 0.942 BE = 0.772

## 7. 결론

본 논문에서는 기존의 패턴 인식에서 문자 인식의 경우 문자 인식에만 머물렀던 적용의 한계성을 문자 익히기까지 행할 수 있는 방법론을 제안하였다. 이를 위해 전처리과정의 수행, 히스토그램의 Projection profile을 이용한 자소의 분리 및 인식 그리고 퍼지수를 이용한 유사도 함수의 정의에 대해 제안하였다. 실험 결과 주관적인 평가와 객관적인 실험치가 일치함을 확인할 수 있었다. 그러나 개발한 시스템을 상용화 하고자 하는 것과 다양한 문자 패턴에 대한 알고리즘 개발과 실험이 지속적으로 행해져야 하리라 여겨진다.

끝으로 본 논문 작성에 자료 정리등을 도와준 안은영양에게 감사하는 바이다.

## 참고문헌

- [1] “인식 시스템”, 한국정보처리학회 특집호, Vol.6, no-4, 1999
- [2] 이상범, “한글 처리-문자 중심 인식 기술 고찰, ”한국정보처리학회지 ,Vol.5, No.5 1998
- [3] 이인동, “신문 영상의 구조 해석 관한 연구”, 충남 대학교 박사학위 논문, 1991
- [4] D.U.Cho, “Noise Removal and Edge Detection of Image by Image Structure Understanding,”Transaction of the Korea Information Processing Society, Vol.4, No.7, 1997
- [5] Wakayama, “Skeleton Tracing Based on Maximal SqUARE Moving, ”IEC Technical Report, PRL78 ~ 87, 1978