

초등학생과 외국인을 위한 한글 문자 익히기 시스템의 개발 (A Development of Hanguel Learning System for Elementary School Students and Foreigners)

조 동 옥*
(Dong-Uk Cho)

요 약

본 논문에서는 초등학교 학생이나 외국인을 위해 한글을 익히는 교육용 시스템을 개발하고자 한다. 모델이 될수 있는 표준 문자 패턴을 선정하고 이의 특징을 추출하여 데이터베이스화한다. 이를 위해 환경에 불변인 전처리과정의 수행과 특징 추출 그리고 유사도 함수를 정의한다. 최종적으로 초등학생이나 외국인이 쓴 문자에 대해 표준 문자 패턴과의 유사도를 계산하여 필기체 문자의 미적 평가를 행한다. 본 시스템은 특정 서체에 대해서도 표준 문자 패턴을 선정하여 서체 익히기까지 행할 수 있도록 확정이 가능한 시스템이며 실험에 의해 본 논문의 유용성을 입증하고자 한다.

ABSTRACT

This Paper develops the Hanguel character learning system for elementary school students and foreigners. Standard character pattern is selected and DB is constructed for model by feature extraction. For this, performance of pre-processing independent of environments, feature extraction and simility functions are defined. Finally, beauty evaluation is done by matching between input character pattern written by elementary school students or foreigners and standard character pattern. It is possible for this system to extend the specific character font learning from selecting the specific standard character pattern. Also the effectiveness of this paper is demonstrated by several experiments.

1. 서론

최근 개인용 컴퓨터의 급속한 보급으로 개인용 컴퓨터에서 사용할 수 있는 입력 문자의 자동화를 위한 문자 인식기가 꾸준히 개발되며 보급되고 있다 [1]~ [3]. 1990년대 초반부터 상용화 제품이 나오기 시작하여 ReadEx, NeuroOct, 글눈, 아르미 등으로부터 AnyPage에 이르기까지 상당히 많은 글자 인식기가 개발되었다. 현재 아르미는 한글, 한자, 영어를 초당 300자씩 고속으로 인식하고 있는 상황이다.

그러나 이 같은 패턴 인식 기술 등은 주로 인쇄체 문자 인식에 초점이 맞추어져 있으며 문자 인식 [4], [5] 외에 지능형 교통 시스템(ITS : Intelligent Transportation System) [6]~[8]에서 자동요금부과(ETC : Electronic Toll Collection)도 모두 인쇄체 문자 인식에 해당되는 경우이다. 그러나 입력문자의 자동화는 인쇄체 문자 인식 보다 필기체 문자 인식에 보다 많은 사회적 요구가 있는 분야로 고려되고 있다.

* 정회원 : 충북과학대학 정보통신공학과 교수

또한 문자 인식 시스템의 경우 거의 모든 시스템이 문자 인식에만 초점이 맞춰져 있고 문자 익히기 (character learning)의 기능은 없는 것이 현재의 실정이다. 그러나 초등학교 저학년 생이나 외국인들이 처음으로 한글문자에 대한 쓰기와 읽기를 접할 시 숙련된 교사의 도움을 필요로 하고 있기 때문에 이를 대체할 수 있는 프로그램의 개발은 중요한 일이 아닐 수 없다.

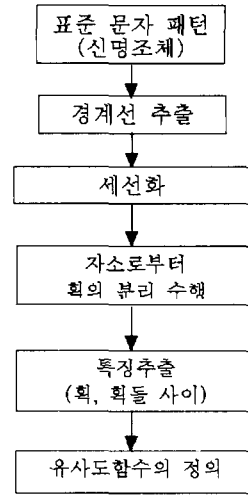
따라서 이를 위해 문자 익히기를 행할 수 있는 시스템의 개발이 사회적으로 요구되고 있다. 본 논문에서는 이 같은 작업을 수행하기 위해 한글의 표준 문자 패턴(예를 들어 신명조체)에 대해 초등학교생이나 외국인들이 한글 문자를 익힐 수 있는 시스템을 개발하고자 한다. 이는 문자 익히기에도 사용할 수 있지만 성인의 경우 다양한 서체에 대해 표준 문자 패턴을 선정하여 이를 익히는데 까지 확장 가능한 시스템이 될 수 있을 것으로 여겨진다.

본 논문에서는 우선적으로 어린 학생이나 외국인들을 대상으로 표준 문자 패턴을 신명조체로 선정하여 이를 익힐 수 있는 시스템을 개발하고자 하며 여러 가지 다양한 실험에 의해 본 논문의 유용성을 입증하고 한다.

2. 전체 시스템의 구성

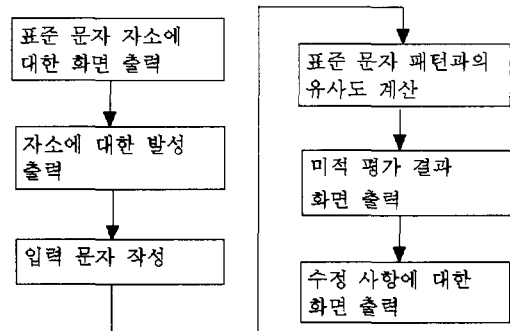
구축하고자하는 시스템은 크게 표준 문자 생성부와 실행부로 나뉘어진다. 아래 [그림 1]에 표준 문자 생성부를 [그림 2]에 실행부를 나타내었다. 표준 문자 패턴으로는 신명조체를 선정하였으며 이에 대한 처리과정을 [그림 1]에 나타내었다.

또한 실행부는 컴퓨터 화면상에 익히고자 하는 자소에 대한 출력과 이에 대한 발음을 나타내주며 이를 펜라이트나 종이 위에 쓰고 입력과정을 행하게 된다. 이후 이에 대한 전처리과정(pre-processing)의 수행과 특징추출 그리고 표준 문자 패턴의 유사도 함수와의 정합과정을 걸쳐 미적 평가 (beauty evaluation)를 행하게 된다. 최종적으로 미적 평가 결과를 화면상에 출력해 주며 이와 함께 교정 사항도 동시에 출력이 된다.



[그림 1] 표준 문자 생성부

[Fig.1] Generation Part for Standard Characters Font



[그림 2] 실행부

[Fig.2] Practice Part

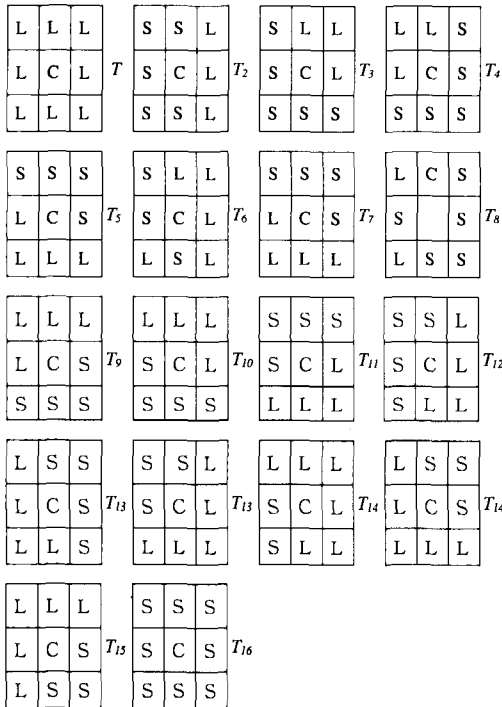
3. 표준 문자 패턴의 생성

본 논문에서는 표준 문자 폰트로 신명조체를 선정하였다. 이제 표준 문자 패턴의 생성을 위해 전처리과정을 수행하고 자소의 분리를 행하여 특징을 추출해야 한다.

최종적으로 획 그리고 획들간의 특징 벡터에 대해 유사도 함수를 정의해야 한다.

3.1 전처리과정의 수행

전처리과정은 크게 잡음제거와 경계추출 그리고 세션화과정으로 나뉘어진다. 이때 잡음제거와 경계추출은 아래 [그림 3]과 같은 영상의 구조를 정의하여 행한다.



[그림 3] 영상의 구조

[Fig.3] Image Structures

또한 'S' 와 'L'의 계산을 아래 (식1) 과 (식2)에 의해 계산한다.

$$\psi_s(X) = -\log\left[\frac{-(X-225)}{255}\right] \quad (식1)$$

여기서 $X = |C - S|$

$$\psi_L(X) = -\log\left[\frac{X}{255}\right] \quad (식2)$$

여기서 $X = |C - L|$

최종적으로 세션화는 Wakayama[11]의 알고리즘을 이용하여 행한다.

3.2 특징 추출

표준 문자 폰트에 대한 특징 추출은 자음과 모음으로 나누어 행한다.

3.2.1 자음의 특징 추출

자음에 대해 획의 특징과 획들 사이의 특징을 추출해야 한다. 우선 획의 분리는 인식을 위해 기억된 히스트그램에 대해 최대값과 최소값의 좌표(X,Y)를 구하면 획의 분리 및 획의 시작점과 끝점을 구할 수 있다. 이중 획에 대한 특징 벡터는 획의 시작점과 끝점 사이의 이루는 각으로 정하며 획들 사이에 대한 특징 벡터는 획들 사이의 이루는 각과 길이 비로 선정하였다. 또한 'ㅇ' 과 'ㅎ'은 이 같은 방법으로 특징 벡터를 선정할 수 없기 때문에 이를 따로 추출해야 한다. 우선 'ㅇ'의 특징 벡터는 원의 둥근 정도(circleness)로 선정하였는데 이는 원의 반지름에 대한 전체 반지름의 누적 배열로 정한다. 또한 'ㅎ'은 획의 각도와 'ㅇ'에 대한 원의 둥근 정도와의 산술 평균을 통해 획의 특징 벡터를 선정한다. 최종적으로 'ㅎ' '획의 경우 획들 사이의 특징 벡터는 이루는 각과 길이 비 그리고 직선 성분의 획과 'ㅇ'획 사이의 최소 거리값을 구하여 계산한다.

3.2.2 자음에 대한 유사도 함수의 정의

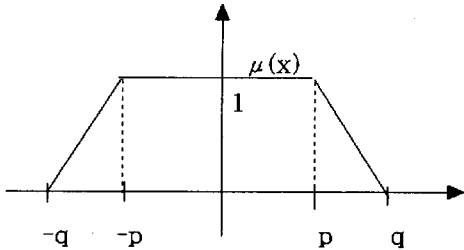
자음은 아래 <표 1>과 같이 크게 두 개의 특징 벡터로 이루어져 있다. 이제 이에 대한 유사도 함수를 정의해야 한다.

<표 1> 자음의 특징 벡터

<Table 1> Feature Vectors of the Consonant

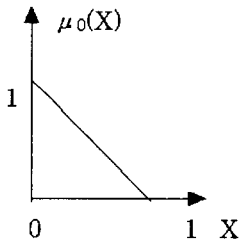
자음의 특징 벡터	
획 추출 요소	획의 각도
획들 사이의 추출 요소	이루는 각 (Intersection Angle)
	길이 비 (Length Ratio)

이를 아래 [그림 4]와 같은 사다리꼴 함수를 통해 유사도 함수를 정의한다.



[그림 4] 유사도 함수 계산을 위한 사다리꼴 함수의 정의
[Fig.4] Definition of Trapezoidal Function for Similarity Measure Calculation

최종적으로 'o'에 대한 유사도 함수는 아래 [그림 5]와 같은 삼각형 함수를 통해 정의한다.



[그림 5] 'o'에 대한 유사도 함수의 정의
[Fig.5] Definition of Similarity Measure Function for 'o'

3.2.3 모음에 대한 특징 추출 및 유사도 함수의 정의

모음에 대한 특징 벡터는 크게 획과 획들 사이의 요소로 나뉘어 진다. 우선 획에 대한 특징 벡터는 획의 시작점과 끝점 사이의 각으로 정한다. 획들 사이의 특징 요소는 크게 세 가지로 구분되어 진다. 첫째가 획들 사이의 이루는 각이고 둘째가 획들 사이의 길이 비 그리고 셋째가 획과 획들 사이의 교차점 위치이다. 이중 첫째 요소와 둘째 요소는 자음의 경우에 적용한 식을 그대로 반복 적용하면 된다. 그러나 셋째 요소는 교차점 위치에 대한 유사도 함수식을

새로이 정의해 주어야 한다. 획들 사이의 교점 위치도 크게 1교차점 (예를 들어 'ㄱ', 'ㄴ' 등), 2교차점 (예를 들어 '교' 등) 그리고 4교차점 (예를 들어 'ㅁ' 등) 등으로 나눌 수 있다. 예로서 1교차점에 대한 유사도 함수를 정의해 보자. 모음 교차점이란 세로로 긴 획과 가로로 획과의 만나는 위치를 뜻하는데 이를 위해 세로로 긴 획에 대한 정규화를 수행하여 교차점에 대한 유사도 함수를 정의한다. 즉,

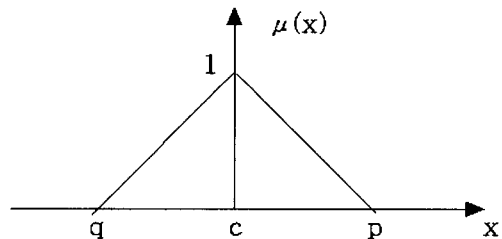
$$D_i = \sqrt{(X_S - X_E)^2 + (Y_S - Y_E)^2} \quad (\text{식3})$$

$$D_j = \sqrt{(X_S - X_M)^2 + (Y_S - Y_M)^2} \quad (\text{식4})$$

여기서 (식 3) 은 세로 획의 길이를 구하는 식으로서, (X_E, Y_E) 는 끝점을 나타낸다. 또한 (식 4)는 교차점의 위치를 구하는 식으로 (X_M, Y_M) 은 세로획과 가로획간의 교차점의 위치를 나타낸다. 교차점에 대한 미적 평가는 (식 5) 와 같이 정의한다.

$$\text{교차점의 미적 평가} = \frac{D_i}{D_j} \quad (\text{식5})$$

최종적으로 유사도 함수식은 아래 [그림 6]과 같다. [그림 6]에서 p와 q는 교차점에 있어 긴 획의 시작점과 끝점을 나타낸다. 먼저 1교차점은 중간점 즉, $c = (p + q) / 2$ 이다. 2교차점에 대해서는 1교차점을 확대 적용하면 된다.



[그림 6] 1교차점에 대한 유사도 함수
[Fig.6] Similarity Measure Function for 1 Cross Point

아래 <표 2>에 모음의 특징 벡터에 대해 정리를 하였다.

<표 2> 모음의 특징 벡터
 <Table2> Feature Vectors for Vowel

특 징		미적 평가 준 값	
획 추출 요소	획의 각도	자음과 동일	
획 사이의 추출 요소	이루는 각	자음과 동일	
	길이 비	자음과 동일	
교차점	1교차점		$c=(p+q)/2$
	2교차점	위 교차점	$p=p$
			$c=p/3$
		$q=2p/3$	
	아래 교차점	$p=2p/3$	
		$c=p/3$	
	$q=q$		
4교차점		2교차점의 앞·뒤 반복적용	

$$M = \text{length of profile [i] with O} \quad (\text{식6})$$

$$N = \text{length of profile [i] with non - zero} \quad (\text{식7})$$

만일 입력문자 패턴이 경사 상태가 되면 행과 열에 대한 투사 프로파일에 대해 (식8)과 같은 조건이 성립한다.

$$M = O \quad (\text{식8})$$

또 입력상태가 기울어져 있지 않은 문자 패턴이 횡서인 경우 행에 대한 투사 프로파일에 대해 같은 조건이 성립된다.

$$M/(M+N) > K, K=0.3 \leq K \leq 0.8 \quad (\text{식9})$$

열에 대한 투사 프로파일에 대해서는 (식10)의 조건이 성립한다.

$$N/(M+N) < K, K=0.3 \quad (\text{식10})$$

4. 입력 문자의 처리

지금까지는 표준 문자 패턴의 생성부에 대해 기술하였다. 이제 손으로 쓴 입력 문자에 대한 처리를 행하여야 한다. 이 같은 입력 문자에 대한 처리는 입력 자소에 대한 처리와 경사 교정 그리고 특징 추출과 정합으로 나뉘어진다. 이 중 특징 추출과 정합 과정은 표준 문자 패턴과 같은 처리 과정을 거치게 되며, 입력 자소에 대한 처리와 경사 교정이 필요하게 된다. 입력 자소에 대한 처리는 공간 좌표상의 일정 범위 내에 투사 프로파일 존재하게 되면 이를 균집화를 수행하고 균집화된 요소들에 대해 중심점을 선택해 줌으로써 입력 자소에 대한 획의 분리와 인식이 가능하게 된다.

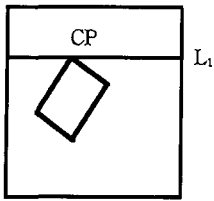
또한 입력 문자가 비뚤어져 있다면 이를 교정하여야 한다. 이를 위해 히스토그램의 투사 프로파일을 사용한다. 투사 프로파일은 행(X축) 또는 열(Y축)에 대해 1로 고정된 화소 개수를 기록한 화일이며, 이는 배열로써 구현된다. 즉, Projection[i] 는 i 번째 행과 열의 화소수가 된다. 이때 M과 N을 하식과 같이 정의한다.

마찬가지로 입력 상태가 정상인 데이터가 종서인 경우는 횡의 경우와 같다. 이제 경사 상태 교정을 수행하기 위해 하식을 정의해 보자.

$$NP = \sum P_i, P_i \in \text{선분}, P_i = 0 \text{ 또는 } 1 \quad (\text{식11})$$

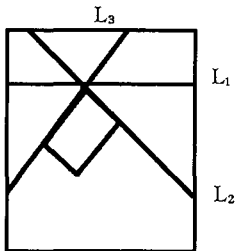
P_i 는 선분 위에 있는 i번째 화소의 값을 나타낸다. 만약 위의 단계에서 입력 데이터가 경사 상태라는 것으로 판정되면 [그림 7]과 같이 입력된 문자에 대해 행에 평행한 직선을 긋는다. 이때 NP가 다음 조건을 만족할때까지 선분을 아래로 평행 이동시킨다. 이 선분을 L1이라 한다.

$$NP > K, K \text{는 잡음 임계치} \quad (\text{식12})$$



[그림 7] 경사 상태 교정 과정 (1)

[Fig.7] Procedure of Skew Normalization(1)



[그림 8] 경사 상태 교정 과정 (2)

[Fig.8] Procedure of Skew Normalization(2)

위의 [그림 8]과 같이 선분 상에 있는 화소 중에 값이 1인 최 좌측점(또는 최 우측점, 이를 CP라 칭함)을 중심으로 시계 방향(시계 반대 방향)으로 NP 값을 계산하면서 하식을 만족할때까지 선분을 회전시킨다. 이때 선분을 L2라 하자.

$$NP > K, K \text{는 Text 임계치} \quad (\text{식}13)$$

이때 L2와 수직이며 CP를 지나는 직선을 L3라 하자. 이 L3에 평행한 모든 직선 위에 있는 화소들을 투사를 실시하면 경사 상태가 교정이 된다. 이 같이 입력문자 패턴에 대해 경사 교정을 수행한 후에 특징 추출을 행하여 표준 문자 패턴의 유사도 함수와 정합을 통하면 미적 평가가 가능하게 된다.

5. 실험 및 고찰

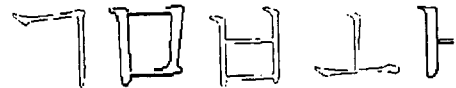
본 논문에서의 실험은 IBM-PC 상에서 C언어로 행하였다. 아래 [그림 9]가 표준 문자 패턴의 생성을 위한 영상이며 [그림 10]과 [그림 11]이 각각 경계선 추출 결과와 세션화 결과가 된다. 또한 <표 3>에 표

준 문자 패턴에 대한 특징 추출 결과를 나타내었다. 아울러 <표 4>에 'ㄱ'에 대해 <표 5>에 'ㅁ'에 대해 그리고 <표 6>에 'ㄴ'에 대한 유사도 함수를 나타내었다. 이제 입력 문자 패턴에 대한 실험 결과를 나타내고자 한다. [그림 12]가 'ㄱ'에 대한 입력 문자 영상 [그림 13]이 전처리과정 수행 결과 [그림 14]가 세션화 결과이다. <표 7>이 'ㄱ'의 입력문자 패턴에 대한 특징 추출 결과 그리고 <표 8>이 미적 평가 결과이다. 그리고 <표 9>가 영상처리에 의한 유사도 결과 및 100명의 사람을 상대로 설문 평가한 결과 비교를 나타낸다. <표 9>에서 알 수 있듯이 본 논문에서 제안한 방법과 100명의 사람이 눈으로 검사한 결과가 거의 일치함을 파악할 수 있었다. 또 다른 실험 영상인 'ㅁ'에 대한 실험결과를 살펴보자. [그림 15]가 입력 문자 영상, [그림 16]이 전처리과정 수행 결과 그리고 [그림 17]이 세션화 결과이다.



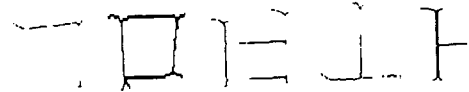
[그림 9] 표준 문자 패턴

[Fig.9] Standard Character Pattern



[그림 10] 전처리과정 결과

[Fig.10] Results of Pre-processing



[그림 11] 세션화 결과

[Fig.11] Thinned Results

또한 <표 10>이 'ㅁ'에 대한 특징 추출 결과, <표 11>이 이에 대한 미적 평가 결과이다. 그리고 <표 12>에 영상처리에 의한 결과와 100명의 사람이 눈으로 판단한 결과에 대한 비교표를 나타내었다. 이제 모음에 대한 실험 결과를 살펴보면 [그림 18], [그림 19], [그림 20]이 각각 입력 문자 영상, 경계선 추출 결과 그리고 세션화 결과를 나타낸다.

<표 3> 표준 문자 패턴에 대한 특징 추출 결과
 <Table.3> Feature Extracted Results for Standard Characters

대상문자	획의특징 추출	획들 사이의 특징 추출	
ㄱ	0, 90	각	90
		길이 비	1.0
모	90, 0 90, 0	각	90, 180, 90 90, 180, 90
		길이 비	0.615, 1, 0.8 1.625, 1.3, 0.8
ㅂ	90, 0 90, 0	각	90, 180, 90 90, 180, 90
		길이 비	0.882, 1.117, 0.882 1.266, 1.0, 0.789
ㅌ	90, 0	각	90
		길이 비	0.1875
		교차점	0.5
ㄴ	90, 0	각	90
		길이 비	0.5
		교차점	0.5

<표 4> 'ㄱ'에 대한 유사도 함수

<Table.4> Similarity Measure Function for 'ㄱ'

획의 각	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 $5 \leq x \leq 20$,
획의 각	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 $95 \leq x \leq 110$
획 사이 각	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 $95 \leq x \leq 105$
길이 비	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 $1.2 \leq x \leq 1.4$

아울러 <표 13>이 입력 문자 'ㄴ'에 대한 특징 추출 결과, <표 14>가 이에 대한 미적 평가 결과 그리고 <표 15>가 제안한 방법과 사람이 행한 결과와의 비교표를 나타낸다.

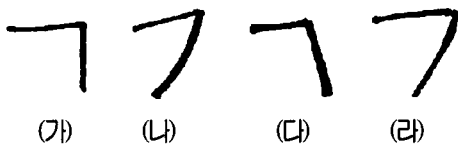
<표 5> 'ㅂ'에 대한 유사도 함수
 <Table 5> Similarity Measure Function for 'ㅂ'

획의 각	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 $95 \leq x \leq 110$
획의 각	0 for $x \leq -20$ or $x \geq 20$, $\frac{-20-x}{-15}$ for $-20 \leq x \leq 5$, 1 for $-5 \leq x \leq 5$, $\frac{x-20}{-15}$ for $5 \leq x \leq 20$
획의 각	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 $95 \leq x \leq 110$
획의 각	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 $5 \leq x \leq 20$
획사이 각	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 $95 \leq x \leq 105$
획사이 각	0 for $x \leq 160$ or $x \geq 200$, $\frac{160-x}{-15}$ for $160 \leq x \leq 175$, 1 for $175 \leq x \leq 185$, $\frac{x-200}{-15}$ for $185 \leq x \leq 200$
획사이 각	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 $95 \leq x \leq 105$
획사이 각	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 $95 \leq x \leq 105$
획사이 각	0 for $x \leq 160$ or $x \geq 200$, $\frac{160-x}{-15}$ for $160 \leq x \leq 175$, 1 for $175 \leq x \leq 185$, $\frac{x-200}{-15}$ for $185 \leq x \leq 200$
획사이 각	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 $95 \leq x \leq 105$
길이 비	0 for $x \leq 0.482$ or $x \geq 2.082$, $\frac{0.682-x}{-0.2}$ for $0.482 \leq x \leq 0.682$, 1 for $0.682 \leq x \leq 0.82$, $\frac{x-1.282}{-0.2}$ for $0.82 \leq x \leq 1.282$
길이 비	0 for $x \leq 0.717$ or $x \geq 1.572$, $\frac{0.717-x}{-0.2}$ for $0.717 \leq x \leq 0.917$, 1 for $0.917 \leq x \leq 1.317$, $\frac{x-1.517}{-0.2}$ for $1.082 \leq x \leq 1.282$
길이 비	0 for $x \leq 0.482$ or $x \geq 2.082$, $\frac{0.682-x}{-0.2}$ for $0.482 \leq x \leq 0.682$, 1 for $0.682 \leq x \leq 0.82$, $\frac{x-1.282}{-0.2}$ for $0.82 \leq x \leq 1.282$
길이 비	0 for $x \leq 0.866$ or $x \geq 1.666$, $\frac{1.066-x}{-0.2}$ for $0.866 \leq x \leq 1.066$, 1 for $1.066 \leq x \leq 1.466$, $\frac{x-1.666}{-0.2}$ for $1.466 \leq x \leq 1.666$
길이 비	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 $1.2 \leq x \leq 1.4$
길이 비	0 for $x \leq 0.389$ or $x \geq 1.189$, $\frac{0.389-x}{-0.2}$ for $0.389 \leq x \leq 0.589$, 1 for $0.589 \leq x \leq 0.989$, $\frac{x-1.189}{-0.2}$ for $0.989 \leq x \leq 1.189$

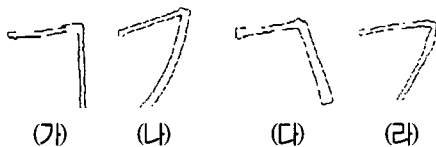
이상의 실험 결과에서 알 수 있듯이 본 논문에서 제시한 방법이 실험적으로 타당함을 입증할 수 있었으며 또한 설문 데이터와의 비교에서도 유사함을 확인 할 수 있었다. 이는 본 논문이 유용한 방법임을 확인할 수 있는 결과가 된다. 그러나 현재의 실험 결과는 본 연구에서 제시한 특징 벡터만을 기준으로 한 결과이다. 따라서 언어학을 전공한 전문가에게 의뢰하여 어떤 벡터가 특징 벡터 값으로 더 추가되어야 할지 그리고 현재 정의한 유사도 함수의 수식 범위가 적절한지에 대한 추가 연구가 지속적으로 수행되어야 하리라 여겨진다.

<표 6> 'ㄱ'에 대한 유사도 함수
<Table 6> Similarity Measure Function for 'ㄱ'

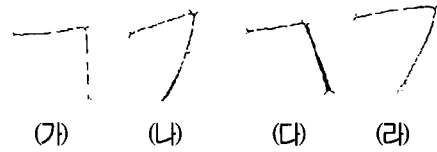
획의 각	0 for $x \leq 70$ or $x \geq 110$, $\frac{70-x}{-15}$ for $70 \leq x \leq 85$, 1 for $85 \leq x \leq 95$, $\frac{x-110}{-15}$ for $95 \leq x \leq 110$.
획의 각	0 for $x \leq -20$ or $x \geq 20$, $\frac{-20-x}{-10}$ for $-20 \leq x \leq -5$, 1 for $-5 \leq x \leq 5$, $\frac{x-20}{-10}$ for $-20 \leq x \leq -5$.
획 사이 각	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 $95 \leq x \leq 105$
길이비	0 for $x \leq 0.3875$, $\frac{0.5875-x}{-0.2}$ for $0.3875 \leq x \leq 0.5785$, 1 for $x \geq 0.5785$
교차점	$\frac{x}{0.5}$ for $0 \leq x \leq 0.5$, $\frac{1.0-x}{0.5}$ for $0.5 \leq x \leq 1.0$



[그림 12] 'ㄱ'에 대한 입력 문자
[Fig.12] Input Character for 'ㄱ'



[그림 13] 전처리과정 수행 결과
[Fig.13] Results of Pre-processing



[그림 14] 세선화 결과
[Fig.14] Thinned Results

<표 7> 'ㄱ'입력 문자에 대한 특징 추출 결과
<Table7> Feature Extraction Results for 'ㄱ'
Input Characters

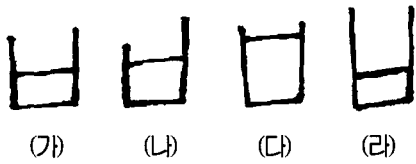
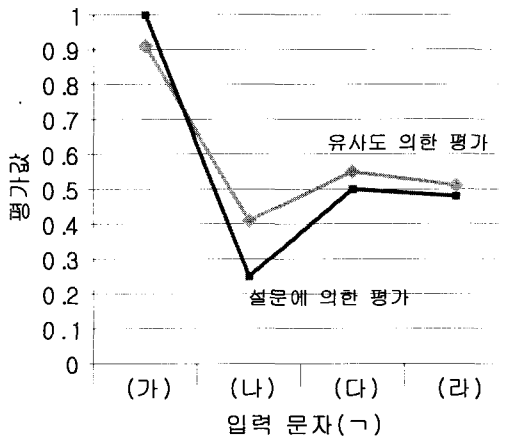
대상문자	획의 특징 추출	획들 사이의 특징 추출	
		각	길이비
ㄱ	-3.5, 92.5	각	91.3
		길이비	0.8
ㄱ	12.1, 44.6	각	43.5
		길이비	1.3
ㄱ	2.8, 124.2	각	124.2
		길이비	1.125
ㄱ	6.23, 40.82	각	55.41
		길이비	1.1

<표 8> 'ㄱ'에 대한 미적 평가 결과
<Table 8> Results of Beauty Evaluation for 'ㄱ'

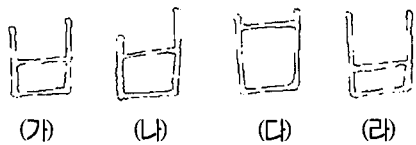
대상문자	유사도에 의한 미적 평가	설문에 의한 미적평가	미적 평가 결과		
			획	획들 사이	
ㄱ	1.0	0.92	1.0, 1.0	각	1.0
				길이비	1.0
ㄱ	0.25	0.42	0.553, 0	각	0.0
				길이비	0.5
ㄱ	0.5	0.54	1.0, 0.0	각	0.0
				길이비	1.0
ㄱ	0.4795	0.51	0.918, 0.0	각	0.0
				길이비	1.0

<표 9> 유사도 및 설문에 의한 평가 결과 비교(ㄱ)

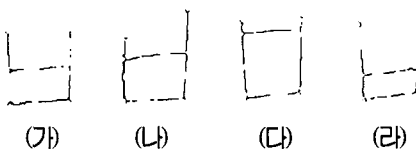
<Table 9> Comparison Between Evaluation Results for Similarity Measure and Questionnaire (ㄱ)



[그림 15] 'ㄱ'에 대한 입력 문자
[Fig.15] 'ㄱ' Input Characters



[그림 16] 전처리과정 수행 결과
[Fig.16] Results of Pre-processing



[그림 17] 세션화 결과
[Fig.17] Thinned Results

<표 10> 'ㄱ'입력 문자에 대한 특징 추출 결과

<Table 10> Feature Extracted Results for 'ㄱ' Input Characters

대상 문자	유사도에 의한 미적 평가	설문에 의한 미적 평가	미적 평가 결과		
			획	획들 사이	
ㄱ	0.99	0.94	1.0, 0.98 1.0, 0.88	각	1.0, 1.0, 0.99 1.0, 1.0, 1.0
				길이 비	1.0, 1.0, 1.0 1.0, 1.0, 1.0
ㄱ	0.79	0.75	1.0, 0.88 1.0, 1.0	각	0.94, 1.0, 1.0 1.0, 1.0, 1.0
				길이 비	1.0, 0.365, 1.0 0.0, 0.0, 0.455
ㄱ	0.973	0.45	0.847, 1.0 1.0, 1.0	각	1.0, 1.0, 1.0 0.98, 1.0, 1.0
				길이 비	1.0, 1.0, 1.0 0.98, 1.0, 1.0
ㄱ	0.817	0.83	1.0, 0.78 1.0, 0.485	각	1.0, 1.0, 1.0 1.0, 1.0, 0.468
				길이 비	0.7, 1.0, 0.66 0.0, 1.0, 0.975

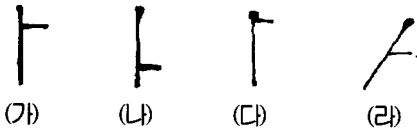
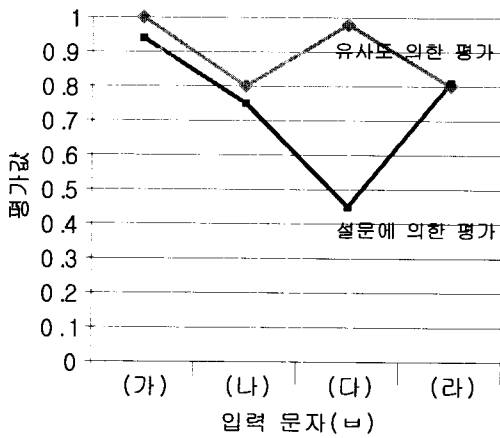
<표 11> 'ㄱ'에 대한 미적 평가 결과

<Table 11> Results of Beauty Evaluation for 'ㄱ'

대상 문자	획의 특징 추출	획들 사이의 특징 추출	
		각	획들 사이의 특징 추출
ㄱ	91.7, 5.3 92.6, 6.8	각	86.4, 179.1, 84.9 87.3, 178.5, 85.8
		길이 비	0.764, 1.117, 0.764 1.461, 1.0, 0.684
ㄱ	91.2, 6.8 93.6, 3.4	각	84.4, 177.6, 87.8 86.8, 176.6, 90.2
		길이 비	0.722, 1.444, 0.694 2.0, 0.961, 0.480
ㄱ	97.3, 3.8 88.6, 3.2	각	93.5, 171.3, 94.1 84.8, 179.4, 85.4
		길이 비	0.789, 1.027, 0.686 1.301, 0.869, 0.668
ㄱ	93.76, 8.3 92.4, 12.72	각	85.46, 178.64, 81.04 84.1, 175.58, 79.68
		길이 비	0.542, 0.942, 0.55 1.738, 1.015, 0.584

<표 12> 유사도 및 설문에 의한 평가 결과 비교(부)

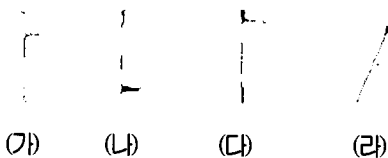
<Table 12> Comparison Between Evaluation Results for Similarity Measure and Questionnaire(부)



[그림 18] 입력 문자 (ㅏ)
[Fig.18] Input Characters for 'ㅏ'



[그림 19] 전처리과정 수행 결과
[Fig.19] Results of Pre-processing



[그림 20] 세션화 결과
[Fig.20] Thinned Results

<표 13> 'ㅏ' 입력 문자에 대한 특징 추출 결과

<Table 13> Feature Extracted Results for 'ㅏ'
Input Characters

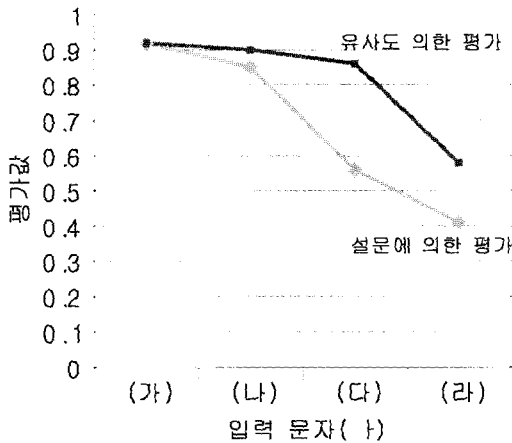
대상문자	유사도에 의한 미적 평가	설문에 의한 미적 평가	미적 평가 결과		
			획	획들 사이	
ㅏ	0.905	0.91	1.0, 1.0	각	1.0
				길이비	0.1
				교차점	0.526
ㅑ	0.90	0.85	1.0, 1.0	각	1.0
				길이비	1.0
				교차점	0.5
ㅓ	0.848	0.56	1.0, 0.96	각	0.77
				길이비	1.0
				교차점	0.512
ㅕ	0.569	0.41	0.0, 0.913	각	0.0
				길이비	1.0
				교차점	0.932

<표 14> 'ㅏ'에 대한 미적 평가 결과

<Table 14> Results of Beauty Evaluation for 'ㅏ'

대상문자	획의 특징추출	획들 사이의 특징 추출	
		각	88.7
ㅏ	89.8, 2.1	길이비	0.263
		교차점	0.263
		각	90.7
ㅑ	92.3, 1.6	길이비	0.156
		교차점	0.75
		각	97.3
ㅓ	91.7, -5.6	길이비	0.256
		교차점	0.256
		각	59.1
ㅕ	65.4, 6.3	길이비	0.195
		교차점	0.536

[표 15] 유사도 및 설문에 의한 평가 결과 비교(%)
 <Table 15> Comparison Between Evaluation Results
 for Similarity Measure and Questionnaire



6. 결론

본 논문에서는 초등학생이나 외국인들이 한글을 익히기 위해 표준 문자와 입력문자의 획에 대한 특징 추출 및 유사도 함수를 정의하여 표준문자와 입력문자가 얼마나 유사한가를 평가할 수 있는 알고리즘을 제안하였다. 이를 위해 임계치 선정 없는 전처리 과정의 수행 방법의 획의 특징 추출과 유사도 함수 정의 방법을 제안하였다. 또한 실험에 의해 제안한 방법의 유용성을 입증하였다. 이는 제안한 방법에 의한 실험 결과와 일반 사람들이 눈으로 행한 실험 결과가 유사하게 일치한다는 결론을 얻을 수 있었기 때문이다. 그러나 앞으로 어떠한 요소들을 특징 벡터로 더 선정해야 하는지 와 완성 문자에 대한 특징 추출 및 유사도 함수의 정의 및 입력 문자 패턴에 대한 연구가 지속적으로 수행되어야 하리라 여겨진다. 또한 필자 인식과 다양한 서체 익히기에도 적용될 수 있도록 하기 위한 시스템의 확장에 대해서도 연구가 계속되어야 하리라 사려된다.

※ 참고문헌

- [1] 이인동, 문자 인식 기술의 현황과 전망, 한국 정보산업 연합회지 특별기고, 통권 제144호
- [2] 이성환, 문자 인식-이론과 실제, 홍릉과학출판사, 1994
- [3] S.Liang et al, "Segmentation of Touching Characters in Printed Document Recognition, Pattern Recognition, Vol.27, No.6, 1994
- [4] J.M Westall and M.S Narasimha, "Vertex Directed Segmentation of Handwritten Numerals, Pattern Recognition, Vol.26, No.10, 1993
- [5] 안창, 이상범, "한글처리-문자 중심 인식 기술 고찰," 한국정보처리학회지, Vol.5, No.5, 1998
- [6] 조동욱의 2인, "영상처리에 의한 버스 전용차로 위반 차량의 검출," 한국정보처리학회 충북지부 학술대회, 2001
- [7] 국가 ITS 아키텍처 및 기술 표준화에 관한 세미나, 국토개발 연구원, 1998
- [8] 조형기의 2인, "영상 처리 기반의 voting technique을 이용한 대기 길이 예측 알고리즘의 개발," 대한 교통학회지, Vol.16, No.3, 1998
- [9] 조동욱의 2인, "정규화에 의한 2차원 물체 인식 알고리즘의 제시," 한국정보처리학회 논문지, Vol.5, No.8, 1998
- [10] 조동욱의 1인, "영상 구조 파악에 의한 경계선 추출," 한국정보처리학회 논문지, Vol.6, No.9, 1999
- [11] Wakayama,, "Skeleton Tracing Based on Maximal Square Moving," IECE Technical Report PRL 78-87, 1978

조동욱



1979.3~1983.2 : 한양대 공대
전자공학과 (공학사)
1983.9~1985.8 : 한양대 대학원
전자공학과 (공학석사)
1985.9~1989.2 : 한양대 대학원
전자통신학과(공학박사)

1982~1983 : (주) 신도리코 기술연구소
연구원 겸 장학생
1989~1990 : 한양대학교 Post Doc. 연구원
1989.9~1991.2 : 동양공전 전자통신과 조교수
1991.3~2000.2 : 서원대학교 정보통신공학과
부교수
1999.3~1999.5 : 미국 오레건주립대학교 교환교수
2000.3~현재 : 충북과학대학 정보통신공학과 교수
1996.10 : 동양일보 선정, “충청을 이끌 50인”
1997.1 : 중부매일 선정, “21세기의 주역 5인”
1997.12 : 중앙일보 선정, “한국을 움직이는 인물들”
1996.11 : 한국통신학회, 한국통신학회 충북지부
학술상 수상
1997.7 : 한국통신학회 공로상 수상
1999.11 : 한국통신학회, 한국통신학회 충북지부
학술상 수상
1999 : 한국통신학회 편집위원
1999.1~현재 : 한국정보처리학회 논문지
편집위원
2000.5~현재 : 충북농업정보화 지원센터
협의회장
1998.3~현재 : 한국통신학회 충북지부 부회장
2000.3~현재 : 한국정보처리학회 충북지부 부회장
2000.8~현재: 한국전자통신연구원 초빙연구원