

論 文

거리변환법에 의한 한글패턴의 특징분류

正會員 高 燦* 正會員 李 大 寧**

Feature Classification of Hanguel Pattens by Distance Transformation method

Chan KOH*, Dai Young LEE** *Regular Members*

要 約 본 논문에서는 한글문자패턴의 새로운 특징추출 및 분류 알고리즘을 제안하였다. 입력된 패턴을 한글기본 6형식으로 분류하고 자소분리를 시행한 후 각 자소별 위치에 따른 굴곡특징점을 추출하였다. 이 특징점에 의해 입력문자의 내용을 정의하고 이를 색인-순차 파일로 구성하였다. 이 파일과 표준사전파일과의 검색으로 인식처리토록 하였다. 간단한 알고리즘으로 인한 처리시간의 단축과 소프트웨어 작성이 용이함을 보였다. 실험의 결과는 입력패턴의 특징추출과 분류의 결과를 나타내준다. 제안된 알고리즘은 문자를 이루는 최소 4각형 안에서 거리변환을 시켜 굴곡특성을 추출하여 이들이 갖고 있는 상대 위치 정보를 이용한 것이 특징으로 실험을 통해 97%의 인식율을 나타내었다.

ABSTRACT In this paper, a new algorithm for feature extraction and classification of recognizing Hanguel patterns is proposed. Inputed patterns classify into six basic formal patterns and divided into subregion of Hanguel phoneme and extract the crook feature from position information of the each subregion. Hanguel patterns are defined and are made of the indexed-sequence file using these crook feature points. Hanguel patterns are recognized by retrieving these two files such as feature indexed-sequence file and standard dictionary file. This paper show that the algorithm is very simple and easily construct the software system. Experimental result presents the output of feature extraction and grouping of input patterns. Proposed algorithm extract the crooked feature using distance transformation method within the rectangle of enclosure the characters. That uses the information of relative position feature. It represents the 97% of recognition ratio.

I. 서 론

정보화 사회의 발전에 따라 한글의 자동인식분야는 정보처리 자동화에 중요한 분야이다. 한글 정보처리의 자동화는 모든 문서화된 지식을 대량으로 직접 자동입력처리되어야 한다. 이로써

한글정보의 자동입력 전문가 시스템의 개발, 데이터 뱅크의 활용등이 원활히 이루어질 수 있다. 더우기 인공지능을 지닌 한글처리시스템의 등장을 기대할 수 있다.

한글문자인식의 연구는 여러방법이 연구되어 왔다. 한글이 갖는 구조적 특성을 이용⁽¹⁾⁽³⁾하거나 형상의 정점 정보이용⁽⁴⁾, 한글의 문법적 특성을 이용⁽⁵⁾하는 방법이 많이 연구되었다. 한글은 모아쓰기의 형태로 되어있고 각 자소가 모아졌을 때 그 모양이 변형된다. 또 한글의 자소가 위치

* 서울産業大學 電子計算學科
Seoul National Polytechnic University.
Dept. of Computer Science.

** 慶熙大學校 電子工學科
Dept. of Electronics Eng. Kyung Hee University.
論文番號 : 89-63(接受1989. 9. 2)

하는 일정한 규칙성을 갖고 있다. 구조적 특성을 이용하기 위해 세선화 처리를 할 때는 자획폭을 가진 한글문자패턴으로부터 골격선을 추출하게 되어 많은 처리시간을 요하며 사소한 잡음으로 인해 오인식이 되는 변형을 자주 초래하게 된다. 이런 단점을 시정하기 위하여 한글의 두께 정보를 가지고 형상전체를 하나의 대상패턴으로 놓고 인식처리를 하고자 한다. 한글의 자소는 주로 직선성분이고 돌기부분이나 가로, 세로줄기나 가지의 시작과 끝부분의 인식에서 불필요한 부분을 제거하면 일정한 표준패턴의 형상으로 정규화할 수 있다. 본 논문에서는 이런점에 착안하여 세선화하지 않고 한글입력패턴을 6형식에 기준으로 하여 초성, 중성, 종성의 부분을 분리하고 각 분리된 부분을 다시 자소영역의 특징 부분으로 분리하여 특징을 추출하였다. 한글의 입력시 변형이 발생하여도 굴곡부분의 상대위치와 갯수는 변함이 없음을 유의하여 위치별 굴곡점을 특징점으로 하였다. 기존 6형식에 따른 존재가능 자소의 종류¹⁾와 위치별 굴곡특징점의 내용을 테이블로 작성하여 이들로 하나의 입력패턴에 대한 레코드를 구성하였다. 이 레코드와 기존 표준사전 화일을 준비하여두고 이에 대한 검색으로 인식처리를 하였다. 본 논문의 출력으로 각 레코드의 성질에 따른 한글패턴의 특징 추출과 분류한 결과를 출력하였다.

한글 문자의 인식은 많은 연구가 진행되고 있으며, 관련 논문도 많이 발표되고 있으나 실용화 시스템은 등장하지 않고 있다. 외국의 경우는 영문자와 숫자를 인식하는 시스템이 사용되고 있으며, 현재는 필기체 문서의 인식에 관한 연구가 진행중이다. 한글 문자 인식을 위해 외국의 시스템을 활용한 것이 있으나 고가이므로 실용화에는 문제점을 안고 있다. 한글문자인식에서의 문제점은 다음과 같다. 첫째는 문자의 수가 방대한 점이다. 대략 2000자 정도의 많은 글자를 대상으로 해야 하기 때문이다. 둘째는 문자와 문자간의 유사성이 심한점이다. 유사성은 문자간의 형태가 거의 같으며 그 모양의 차이가 적은 것을 말한다. 한글은 모음의 구조가 단조롭고

같은 글자내에 같은 자음이 반복해서 나타나기 때문이다. 자음과 모음이 서로 모일 때 모양이 변하면서 다른 자소에 끝이 서로 접촉하기도 하고 겹치기도 한다. 여러가지 서체에 따른 복잡성도 있다.

일반적으로 문자의 종류가 많게 되면 식별이 복잡해지고 인식시스템이 크게되어 실용화에 문제점을 갖게 된다. 한글인식의 문제를 한번에 해결할 수는 없으며 단계적으로 해결되리라고 예상된다. 본 논문에서는 이러한 문제점 중에서 문자의 유사성을 제거하는데 유효한 알고리즘을 제안하고자 한다.

유사성의 제거에는 여러가지 방법이 있겠으나 본 논문에서는 한글이 가지는 형상구조의 굴곡특성을 중심으로 하였다. 한글에서는 문자패턴의 왜곡된 스캐너 입력시나 서로다른 필기의 습관을 가진 경우에도 한글의 굴곡점이 갖는 상대적 위치는 거의 같음을 알 수 있다. 문자를 이루는 최소 4각형안에서 문자를 정규화시켜 굴곡특성을 추출하여 대분류하고 다시 소분류로 찾아갈 수 있도록 하였다. 이렇게 대분류후 소분류로 특성 자료를 구성하였으므로 처리시간의 단축과 인식율을 향상할 수 있었다. 거리변환법을 이용하여 굴곡특성을 추출하였으므로 알고리즘이 매우 간결하다. 이로써 간단하게 전체적인 특징을 추출할 수 있으며 또한 알고리즘의 내용이 간단하므로 이의 실현을 위한 소프트웨어 개발이 비교적 쉽고 처리시간의 단축을 가져올 수 있었다.

그림 1. 이 전체의 특징점 추출 및 인식처리 흐름도이다. I.서론에 이어 II.에 본 논문에서 제안한 굴곡특성에 따른 특징점 검출 알고리즘을 설명하였고 III.에 한글문자인식처리시 이용될 수 있도록 특징점 화일의 구성과 표준사전화일의 구성에 대해 설명하였다. 여기에서의 화일은 검색시간의 단축을 위해 색인-순차 화일로 구성하였다. IV.에 실험 및 결과 고찰에 대하여 설명하였고 V.에 결론을 논하였다.

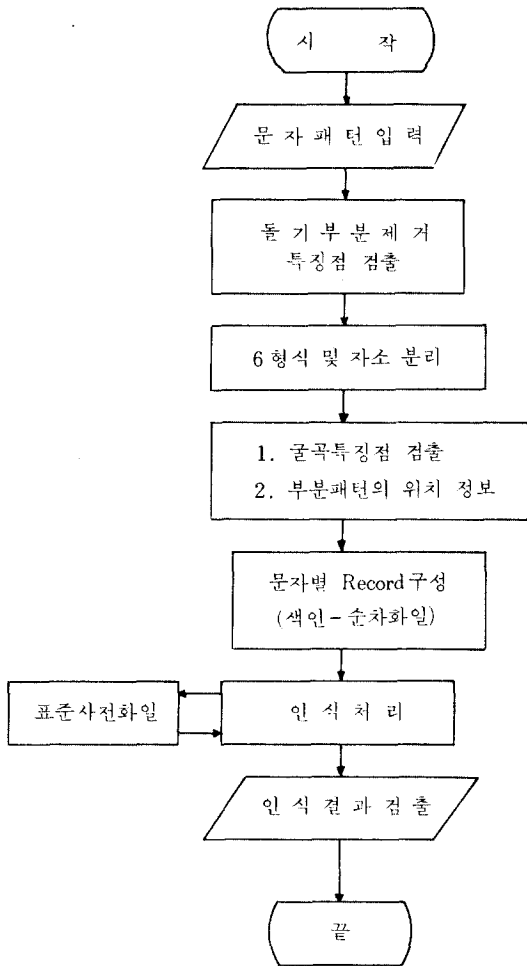


그림 1. 처리 순서도
Block diagram of process

II. 특징점 검출 알고리즘

가. 거리변환 및 방향코딩

한글인식에서의 문제점인 유사성의 제거를 위해 굴곡특징점들을 중심으로 분자포함 최소 4각형 내에서 정규화된 상대위치 정보를 활용하게 된다. 이를 위하여 한글 문자 형성의 중첩운곽형상을 구성하여 골격에 가까운 정도로 접근하여 골격정보를 추출하도록 거리변환 처리를 하였다. 거리변환 처리에서는 외곽부터 내부골격으로

거리에 따라 값을 부여하므로 가장 깊은 길이를 나타내는 굴곡부분은 큰 값을 갖게 된다. 가장 큰 값부터 역으로 검사하면서 굴곡특징을 찾게 된다.

한글형상은 자획과 자획이 연결되는 부분에서 직선성분보다 외곽으로부터 내부로의 거리가 긴 것을 알 수 있다. 또한 입력 스캐너의 조작에서 변형이 생기거나 필기체의 인식시의 변형된 문자라도 연결부위가 나타나는 수나 상대위치는 같게 나타난다. 본 논문에서는 이의 특징을 이용하여 특징점을 추출하고 인식에 이용한다. 그림 2에 자획의 연결부위의 굴곡특성을 보여준다. 입력문자 패턴을 70×70화소로 간이스캐너로 입력 받아 “*”, “.”으로 표시한 것이 그림 3이다.

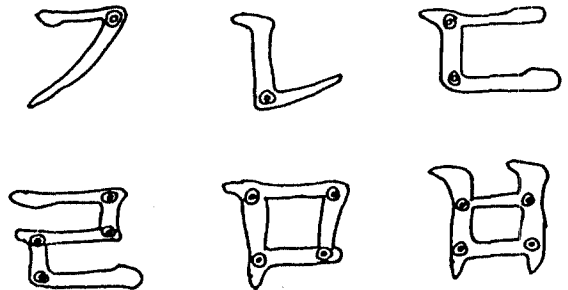


그림 2. 한글 자소의 굴곡특징
Crooked feature of Hangeul phoneme

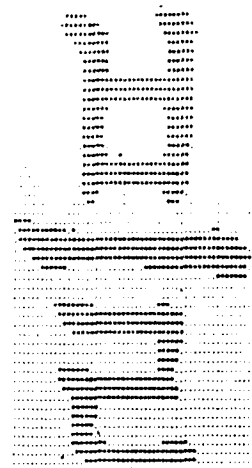


그림 3. 입력문자 패턴
Input character pattern

입력문자에서 문자가 차지하는 최소 직사각형의 화소를 중심으로 처리한다. 사각형의 네 모서리를 변수 t,b,r,l로 하고 문자영역은 s로 한다. 네 변수 t,b,r,l의 가로, 세로 좌표값은 각각 t1,t2,b1,b2,r1,r2,l1,l2로 하여 패턴처리는 문자영역 s의 (i=t1:i<b1; i++)와 (j=l2; j<=r2; j++)의 반복처리로 하였다. 한글의 모음의 수직선분에는 그 시작 부분에 돌기부분이 존재한다. 이는 인식처리지 오인식의 원인을 초래하므로 제거한다. 또한 수평선분에는 시작부분과 끝부분에 같은 모양이 있다. 이의 제거를 위해 수직선분의 돌기부분과 만나는 점을 찾아 제거한다. 이 특징점을 찾기위해 거리변환을 시행하고 다시 방향코딩 한다.

(1) 거리변환 처리

거리변환법은 패턴의 내면으로의 거리를 정하여 중첩되는 관계를 나타내주기위해 사용된다. 한 임의의 “*”화소의 그 배경과 또 다른 “*”화소와의 연결관계를 표시하고 그 배경에서의 최소거리를 구하여 그 점의 값으로 한다. 이는 점 p(x,y)의 거리변환된 값 T(x,y)는 다음식으로 주어진다. \bar{s} 는 배경화소를 표시한다.

$$T(x,y)=\min\{d(p(x,y),\bar{s})\} \quad (1)$$

여기에서 거리 d는 연속평면이아니므로 이산평면상의 거리를 적당히 정의되어 사용된다. 거리가 (x1, y1), (x2, y2)간의 거리일 때 다음식으로 된다.

$$(x1,y1;x2,y2)=|x1-x2|+|y1-y2| \quad (2)$$

$$(x1,y1;x2,y2)=\max(|x1-x2|, |y1-y2|) \quad (3)$$

$$(x1,y1;x2,y2)=\sqrt{(x1-x2)^2 + (y1-y2)^2} \quad (4)$$

(2)식은 4근방 거리변환에, (3)식은 8근방 거리변환에, (4)식은 유클리드거리변환에 사용된다. 이는 두점간을 각각 근방관계를 유지하며 처리될 때 거리관계이다. 4근방 화소 연결상태로 계산된 거리변환과 8근방 화소연결 상태로 거리변환,

그리고 4근방과 8근방을 교대로 적용하여 계산된 거리변환이 있다. 한글 문자 패턴에서 돌기부분을 제거하기 위한 특징점검출을 위해 위의 세가지 변환법을 이용해 실험한 결과 4근방 화소가 유용함을 알았다. 4근방 화소거리변환이 8근방 화소거리변환이나 4-8근방 화소거리변환보다 문자의 자획이 방향을 변하게 될때와 연결이 이루어질 때 그 골곡상태를 상대적으로 크게 나타내주어 특징추출에 유용하다. 그림 4에 4근방 거리변환된 문자패턴의 예를 나타내준다. 거리 변환된 패턴을 이용하여 방향코딩처리를 한다.

(2) 방향코딩 처리

방향코딩은 문자패턴을 구성하는 자획에 대하여 각 자획의 진행방향에 해당 방향코드를 부여한다. 본 연구에서의 방향코딩은 수직선분은 7로 수평선분은 9로 하였다. 4근방 거리변환에서 1-6까지의 코드를 사용하였고 같은 패턴에 이용코자 7과9를 사용하였다.

(가) 수직방향의 코드부여

- 1) 문자패턴의 i=r2에서 l2까지 검사(i=t1; i<=b1; i++)
- 2) (j-1)행, j행, (j+1)행... {p(i,j)=1의 끝} 근방요소검사 (일차원배열로 검사진행) 계속 p(i,j)의 내용변환→ 코드7부여 (1-6의 거리변환값중 증가후 감소특성검사) p(i,j)의 내용불변→ 1)로 반복
- 3) p(i,j)의 값이 p(i,j-1) < p(i,j)하고 p(i,j) < p(i,j+1)하면 코드 7부여
- 4) 거리변환값 1부터 6까지 변화되는가 여부 반복처리

(나) 수평방향의 코드부여

- 1) 문자패턴의 j=t1에서 b1까지 검사(j=l2 ; j<=r2; j++)
- 2) (i-1)행, i행, (i+1)행... {p(i,j)=1의 끝} 근방요소검사 (일차원배열로 검사진행) 계속 p(i,j)의 내용변환→코드9부여(1-6의 거리변화 값중 증가후 감소특성검사) p(i,j)의 내용불변→1)로 반복

- 3) $p(i,j)$ 의 값이 $p(i-1,j) < p(i,j) < p(i+1,j)$ 이면 코드 9의 부여
 - 4) 거리 변환값 1부터 6까지 변화되는가 여부 반복처리
- 그림 5에 방향코딩 처리한 후의 문자패턴예를 나타내준다.

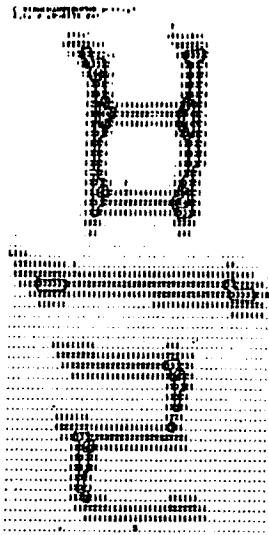


그림 4. 4군방 거리변환 패턴예
Example of 4-distance transformed pattern

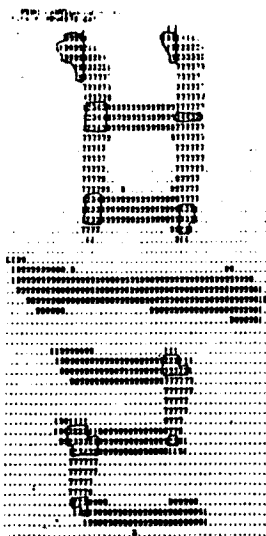


그림 5. 방향코딩 처리 패턴예
Example of directional coding pattern

(3) 돌기부분 제거점 결정

수직방향선분 코드가 계속 $p(i,j)$ 에서 i 가 증가 (1-6거리변환 값 갯수만큼)되고 j 가 (거리변환갯수만큼) 계속될 때

- 1) 수직성분 코드 7의 값 상단에 (i,j 방향검사) 굴곡특징구역 존재
- 2) 좌측에 수평성분코드 9의 값
- 3) 수평선분 끝부분에 최외 윤곽 변환값인 1존재 상기 3개항을 만족시 연결성이 수평에서 수직으로 변하는 곳에서 제거점 검출한다. 이와같이하여 필요시 굴곡부분, 돌기부분, 상투부분, 보선시작 끝부분 등의 제거시 사용한다.

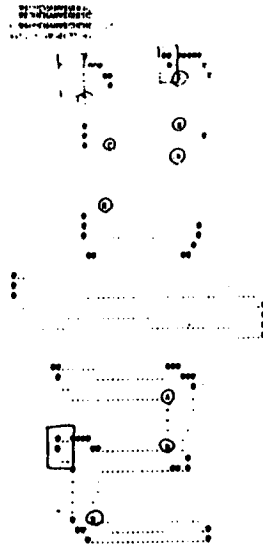


그림 6. 돌기부분 제거점에
Example of Remove point

나. 굴곡특징점 결정

(1) 굴곡특징점 검출

거리변환처리와 방향코딩처리한 패턴에서 수직방향성분(코드 7부여)와 수평방향성분(코드9부여)을 제거하면 거리변환처리에서 계산된 값중 7,9로 바뀌지 않은 나머지가 남게 된다. 이 나머

지의 거리변환값은 상대적으로 큰 값이 남게된다. 이것이 또한 거리 변환값이 크다는 것은 주로 문자패턴에서 굴절부분이거나 모서리부분이다. 한글의 형상에서 굴곡부분, 돌기부분, 상투부분, 보선의 시작과 끝부분이 다른 자획부분보다 굵기 때문이다. 이를 이용하여 굴곡특징점을 구한다. 다음은 굴곡특징점 검출 알고리즘의 설명이다.

- 1) 패턴에서 남은 1-6데이타값의 영역으로 번호 부여. $N=1,2,3\cdots M$
- 2) 남은 화소중 최좌측 하단점을 기준으로 하여 반시계방향으로 번호부여 $K=1,2,3\cdots L$
- 3) 각 영역($N=1,2,3\cdots M$)에서 각 화소점($k=1,2,3\cdots L$)중 기준점 $P_k=1$ 에서 시작하여 반시계방향으로 외곽점 제거 최후의 점이 특징점으로 선정(PL) 외곽점 제거시 반평면 검사 알고리즘의 이용(17)
- 4) $P_k=(a,b)$, $P_{k+1}=(e,f)$, $P_{k+2}=(c,d)$ 라 놓고 다음을 계산한다.

$$\vec{u}=(e-a,f-b,0), \vec{v}=(c-a,d-b,0) \quad (5)$$

$$\vec{u} \times \vec{v} = \{0,0,(e-a)(d-b)-(f-b)(c-a)\} \quad (6)$$

$$(e-a)(d-b) > (f-b)(c-a) \quad (7)$$

$$(e-a)(d-b) < (f-b)(c-a) \quad (8)$$

여기서 외적 $\vec{u} \times \vec{v}$ 가 (0,0,1)의 양의 배수가 될 조건은 식(7)을 만족하는 것이다.

- 5) 식(7)을 만족하면 P_{k+1} 점은 내부점
 식(8)을 만족하면 P_{k+1} 점은 외부점이 된다. 왜냐하면 기준 화소점에서 검사되는 점이 반시계 방향으로 진행하는 것에 대하여 우측에 존재하므로 이점은 외부점이 되고 또한 본 연구에서는 제거되는 점이 된다. 그러므로 식(7)만족시는 계속진행
 식(8)만족시는 P_{k+1} 점 삭제

- 6) P_k, P_{k+1}, P_{k+2} 를 반시계 방향으로 계속 부여하여 1개 화소가 남을때까지 계속처리한다.

그림 7에 굴곡특징점 검출패턴 예를 보여준다.

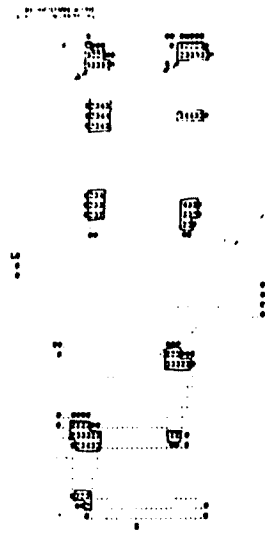


그림 7. 굴곡특징점 검출 예
 Example of detected crook feature point

(2) 굴곡특징점의 위치값 부여

굴곡특징점을 구한점이 $P(i,j)$ 이다. 이 $P(i,j)$ 값이 각 문자별 6형식과 부분패턴 분리구역에서 어떤 상대위치값을 가지는지 계산한다.

위치값 계산을 위하여 문자패턴의 이진값 행과 열의 합과 네끝점을 이용한다.(9)

먼저 문자유형분리를 위해 모음의 형태와 받침유무를 검사하여 분리한다. 모음의 인식후 자음의 영역을 분리한다. 여기에서 이용되는 문자패턴은 돌기부분의 제거를 실시한 영상으로 한다. 6형식으로 구분하기 위해 종모음을 기준으로 6형식을 분류토록하는 (10)의 알고리즘을 이용하였다. 이 알고리즘은 우선 기본 6형식을 종모음이 있는 1형식, 3형식, 4형식, 6형식과 종모음이 없는 2형식, 5형식으로 구분하고 한글문자패턴에서 이상 $2/3 \times 2/3$ 의 영역을 영상에서 좌하로 검사하여 수직방향성분이 있으면 이는 종모음이 있는 것으로 판정한다. 우선 두가지로 분류한 후 t,b,r,l의 네모서리점과 2진 영상의 행과 열의 합으로 분류한다. 원래의 영상을 $F(x,y)$ 로 하고 2진영상을 $G(x,y)$ 로 한다. 스캐너 입력처리시 스캐너에서는 임계치 T를 정하여 히스토그램에서 T값에

의해 0,1로 변환한다. 이의 식은 다음과 같다.

$$F(x,y) \geq T \text{ 일때 } G(x,y) = 1 \quad (9)$$

$$F(x,y) < T \text{ 일때 } G(x,y) = 0 \quad (10)$$

종모음이 있는 형식과 없는 형식은 그림 8과 같다.

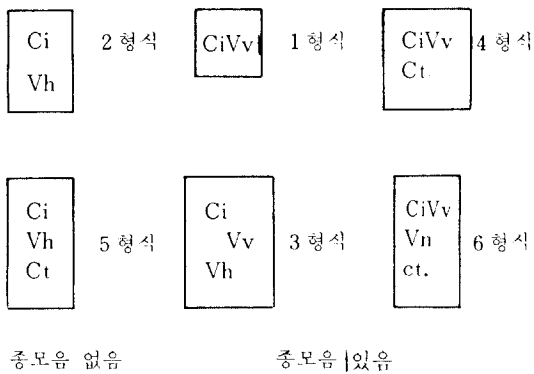


그림 8. 한글의 6가지 유형
Standard 6-type of Hanguel

자소 분리의 시행 후 각 자소마다 윈도우를 설치하여 자소의 't,b,r,l'을 계산하고 이를 다시 그림 9의 부분영역으로 나눈다.

각 분리된 자소를 그림 9의 부분위치에 굴곡점이 존재하는가 검사하고 이를 특징점의 위치 정보로 한다.

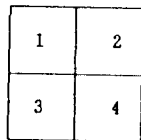


그림 9. 부분 패턴의 영역
Region of subpattern

Ⅲ. 특징분류 화일의 구성

가. 굴곡특징화일의 구성

한글 문자패턴의 인식을 위한 특징화일을 구성한다. 먼저 한글의 기본 6형식중 어느 것인가인 특징변수를 택한다. 각 형식별로 자소의 수가 결정된다. 1형식, 2형식은 자소가 두개, 3형식, 4형식, 5형식은 자소가 세개, 그리고 6형식은 4개이다. 각 자소의 수만큼 필드를 잡게 된다. 각 필드마다 자소의 굴곡수, 굴곡수에 따라 각 굴곡점마다 위치정보를 갖는다. 각 위치는 그림 9의 부분패턴의 영역으로 주어지는 값을 취한다. 같은 "ㄱ"이라도 기본 형식에 따라 그 모양을 달리한다. 그러나 구부러진 곳이 부분영역중 "2"에 존재하는 것은 같으므로 본 논문의 알고리즘이 성립함을 증명할 수 있다. 실험을 통해 이를 확인하여 보았다. 그림 10에 각 형식별 존재 가능자소의 종류를 나타내고 있다. 이는 초성, 중성, 종성에 따른 자소의 종류이다.

| | 가 능 형 식 | 자 소 |
|----|----------------------------------|---|
| 초성 | 형식1, 형식2, 형식3 형식4, 형식5, 형식6 | ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㆁ ㅅ ㅈ ㅊ ㆁ 기 ㅂ ㅅ ㅈ ㅊ ㆁ |
| 중성 | 형식1, 형식4 형식2, 형식5 형식3, 형식6 | ㅂ ㅅ ㅈ ㅊ ㆁ ㅅ ㅈ ㅊ ㆁ ㅇ ㅂ ㅅ ㅈ ㅊ ㆁ 기 ㅂ ㅅ ㅈ ㅊ ㆁ |
| 종성 | 형식4, 형식5 형식6 | ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㆁ ㅅ ㅈ ㅊ ㆁ 기 ㅂ ㅅ ㅈ ㅊ ㆁ |

그림 10. 기본 형식의 자소
Phoneme of basic forms

나. 표준사전화일의 구성

표준사전화일의 구성은 자음, 모음의 표준형상을 기준으로 작성한다. 한글의 자모를 그림 9의 영역 코드값으로 굴곡위치정보가 주어진다. 또 여기에 각 자모의 굴곡갯수 정보가 추가되어 기록된다.

기본 자모의 집합을 $G=\{G_i\}$, $i=1,2,3\cdots 24$ 인데 아래의 순서로 구성한다.
 자음 $C=\{C_i | C \in G, C_i \text{ 자음}\}$, $i=1,2,3\cdots, 14$
 $=\{\text{ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ}\}$
 모음 $V=\{V_i | V \in EG, V_i \text{ 모음}\}$, $i=1,2, 3\cdots, 10$
 $=\{\text{ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ}\}$
 한문자 $G=\{G_i | G_i \in f(c,v)\}$
 $f(c,v)=\{PQR\}$ P: 초성, Q: 중성, R: 종성
 자음 14자, 모음 10자의 표준 특징점들의 내용은 그림 11이다. (ㅇ, ㅎ, 쌍받침 제외)

| 글꼴수 | 초성, 중성 | | | | | | 중 성 | | | | | |
|-----|--------|-----|-----|-----|---|---|-----|-----|-----|-----|---|---|
| 0 | ㅇ ㅎ | | | | | | ㅡ ㅣ | | | | | |
| 1 | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| | | ㄱ | ㅅ | | ㄴ | | ㅌ | | ㅊ | ㅋ | ㅌ | |
| 2 | 1,5 | 2,3 | 4,6 | 2,4 | | | 3,5 | 4,6 | 5,6 | 1,2 | | |
| | ㄷ | ㅈ | ㅊ | ㅋ | | | ㅌ | ㅋ | ㅛ | ㅠ | | |
| 3 | ㅌ | | | | | | | | | | | |
| 4 | ㅍ | | | | | | | | | | | |

그림 11. 표준특징점
Standard feature points

IV. 실험 및 결과고찰

입력패턴에 따른 1차원 코드열을 가지고 준비된 표준패턴의 1차원 코드열과 2진검색에 의한 다. 이때 일치되는 경우 그 내용을 출력하게 된다. 실험에 사용된 문자는 무작위로 추출된 160자의 명조체이다. 이는 책제목, 논문제목, 논문목차들 중에서 임의로 스캐너로 입력하였다. 크기는 1글자당 8mm×8mm 정도이다. 입력장치는 GFI사의 HS-2000 간이스캐너(Handy scanner)를 사용하였다. 200DPI의 흑백데이터 입력이고 1회 입력버는 약 50자 정도이다. 처리프로

그램은 마이크로소프트 C5.0을 사용하였고 IBM PC/AT기종에 실현하였다. (10MHZ 속도)하나의 입력데이터를 읽어 코드부여후 표준화일과 매칭 후 출력까지의 처리가 약 5초가 소요되었다. 무작위로 추출된 문자 160자를 인식 실험한 결과 97%의 인식율을 보였다. 실험에 사용된 형식별 문자의 수와 인식율은 다음과 같다.

1형식의 실험문자는 32자이고 인식문자가 32자로 100% 인식율, 2형식의 실험문자는 36자이고 인식문자가 35자로 97% 인식율, 3형식의 실험문자는 33자이고 인식문자가 33자로 100%의 인식율, 4형식의 실험문자는 28자이고 인식문자가 28자로 100% 인식율, 5형식의 실험문자는 18자이고 인식문자가 16자로 88% 인식율, 6형식의 실험문자는 13자이고 인식문자가 11자로 85% 인식율을 나타내었다. 전체적으로 실험문자는 160자이고 인식문자가 155자로 97%의 인식율을 얻었다. 실패의 경우는 입력시 핸드스캐너 조작미숙으로 입력패턴이 많은 변형을 가져온 경우가 대부분이다. 실패시의 경우는 자소와 자소가 접촉되거나 중첩된 문자패턴에서 발생됨을 알 수 있었다. 이를 위해 자소 접촉 부근에서의 불필요한 요소의 제거가 원활히 처리될 수 있는 방법의 연구가 요구된다. 실패했을 경우의 패턴예를 그림 13에 보였다. 같은 자소가 위치에 따라 모양을 달리하더라도 본 논문을 알고리즘으로 같은 자소로 인식됨을 증명하였다.

그림 12.에 인식결과 예를 표시하였다. 특징점이 차지하고 있는 위치정보를 이용하기 위해 이들이 포함하고 있는 위치의 공간을 재귀적으로 분할하여 이를 중심으로 나타내었다. 공간분할시 전체를 4분할 하지 않고 자소를 기준으로 분할하였다. 트리형태로 자료를 구성할 때 특징점들이 트리내에서의 위치를 정확히 나타내기 위하여 이의 순서를 항상 일정하게 유지할 필요가 있으므로 그림 12에 특징분할과 함께 자료 순회를 표시하였다. 여기서 사용된 중위 순회 프로시저는 다음과 같다.

C:\WKOHC\ANWKOHPHD>prtfig6a
data # aWsd070.dat

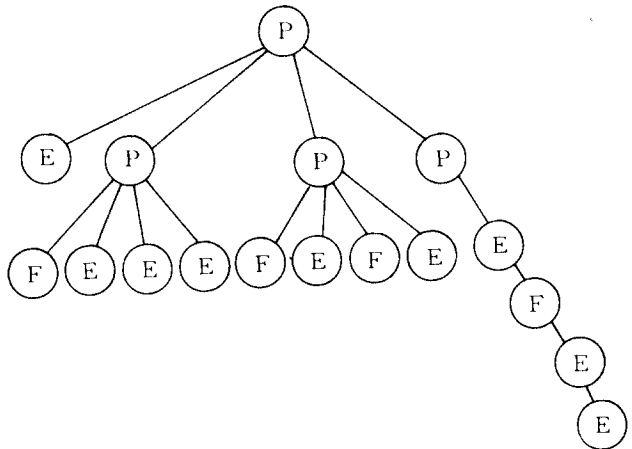
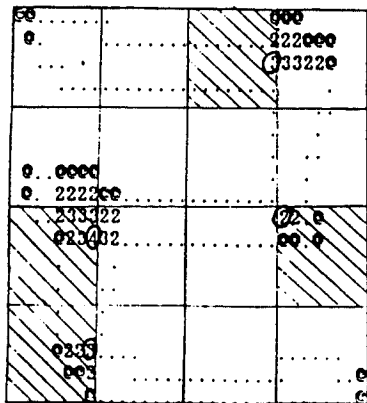
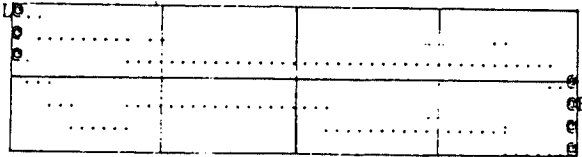
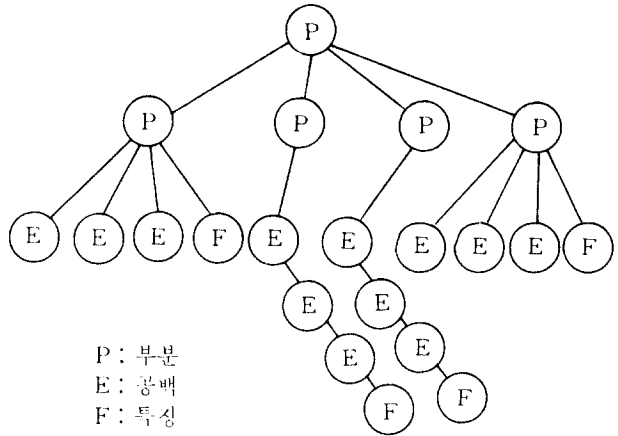
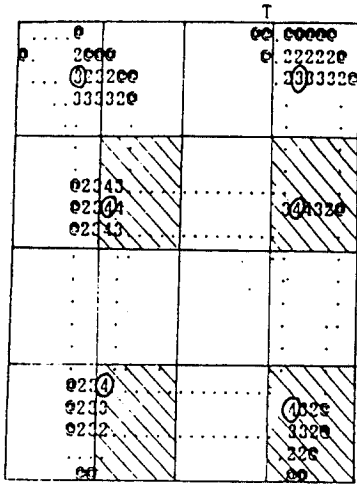


그림 12a. 인식처리 결과 예
Example of Recognition Results

C:\WKOCHAN\WKOHPHD>prtfi6a
data # aWsd071.dat

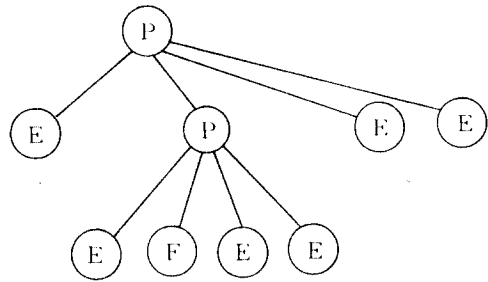
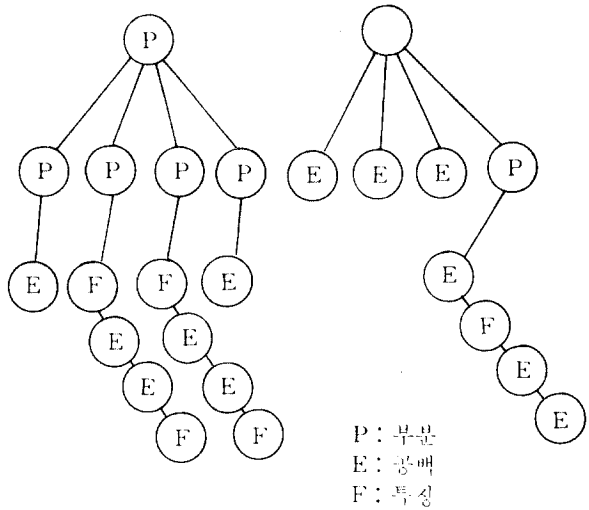
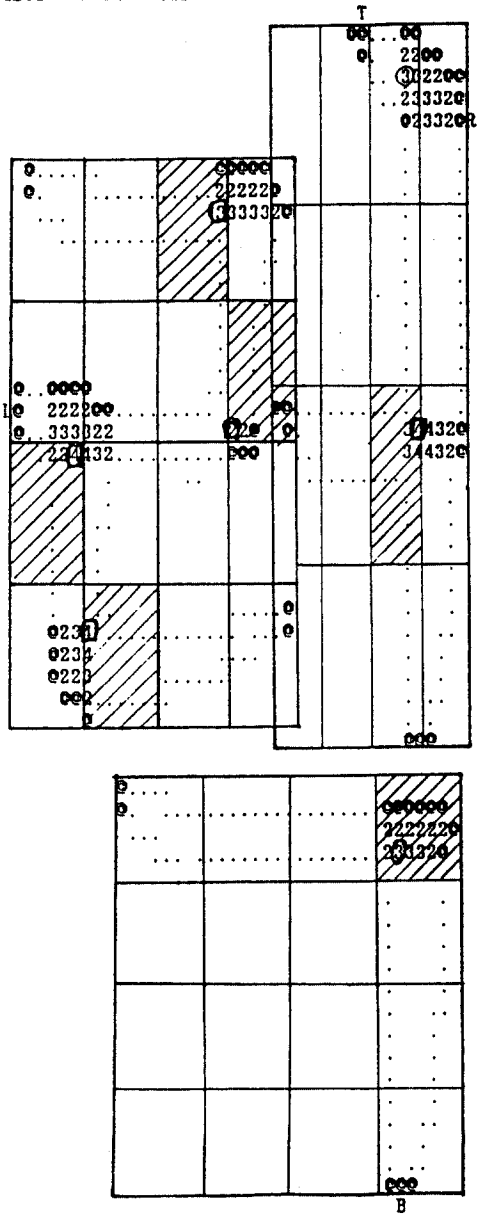
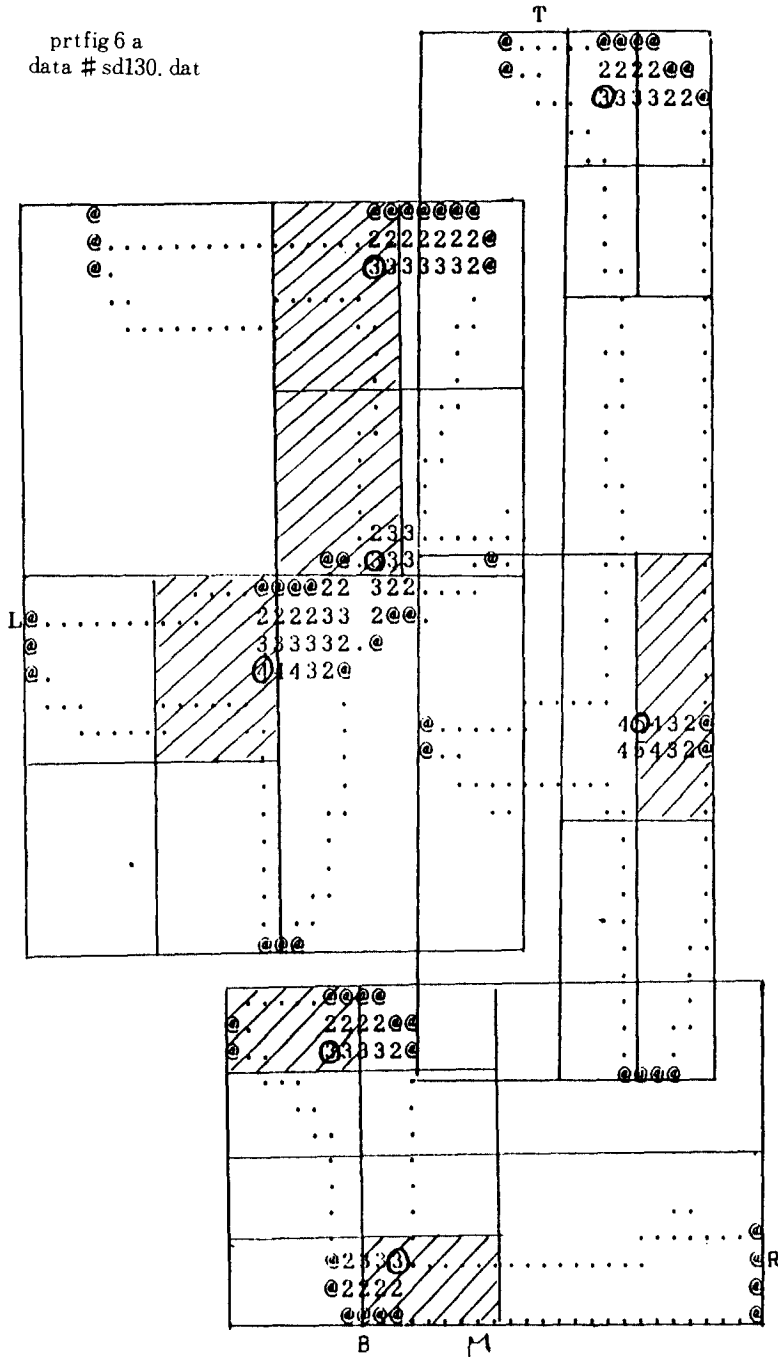


그림 12b. 인식처리 결과 예
Example of Recongnition Results



“권”패턴중 “쿠”부분이
“ㄱ”과 “ㄷ”의 자소로
구분인식이 되지않아
인식처리가 실패되었음.

그림 13. 실패 인식 결과 예
Fig 13. Example of Failure

```

procedure inorder(currentnode: treepointer);
begin {inorder}
  if currentnode< >nil
  then
    begin
inorder (currentnode, Leftchild);
write (currentnode, data);
inorder (currentnode, rightchild);
end end
    
```

V. 결 론

본 논문은 거리변환법을 이용한 한글 인식의 방법을 제안한 것이다. 거리 변환한후 방향 코딩하고 이를 제거하여 남은 내용을 이용하여 인식특성점을 검출하였다. 입력패턴의 돌기부이거나 이와 동등한 부분을 제거하여 오인식율을 감소시켰다. 6형식과 부분패턴에서의 굴곡갯수, 위치정보등을 이용한 일차원 코드열을 만들고 이를 표준사전코드열과 이진검색에 의하여 검색 후 출력토록 하였다. 저렴한 핸드스캐너를 실험에 이용하였으며 세선화를 실시하지 않고 인식처리를 하였으므로 세선화 할 때의 단점을 시정하였다. 이론이 간단하여 소프트웨어 작성이 용이하였고 처리시간도 단축되었다. 본 논문에서는 인쇄체문자로만 실험을 하였으나 한글의 굴곡점의 갯나 위치는 상대적으로 어떤 범위내에 존재하므로 이런 성질을 이용하여 필기체 문자로의 적용도 실험해 보고자 한다. 한글의 특징중에 하나인 같은 자소라도 형식이나 위치에 따라 그 모양을 달리 하는 특징을 같은 방법으로 인식하도록 시도하여 보았다. 그 형상은 위치에 따라 달리 하더라도 굴곡되는 부분의 위치나 갯수가 같음에 착안하여 알고리즘을 구성하였으며 이의 실험으로 증명하였다. 앞으로는 0이나 ㅎ 또 s이나 z자소의 원활한 인식이 되도록 알고리즘

이 보완되어야 하겠고 두개의 자소가 합하여 받침을 이루는 경우의 오인식을 줄이도록 보완되어야 하겠다.

參 考 文 獻

1. 남궁재찬, "Index-window 알고리즘에 의한 한글 pattern의 부분분리와 인식에 관한 연구," 인하대 대학원 박사학위 청구논문, 1982.2
2. 이승호, 김진형, "한글의 구조적 인식을 위한 자획 추출에 관한 연구," 한국정보과학회 추계 학술발표회 논문집, pp.151-154, 1987.10.
3. 유황빈, "구조정보를 이용한 한글문자패턴 인식에 관한연구," 광운대학 논문집, 제15집, pp93-100, 1986.
4. 박종욱, "도형패턴의 벡텍스 검출법과 필기체 한글 세이프 패턴의 인식," 인하대 대학원 박사학위 청구논문, 1985.2
5. 전병민, "배열문법을 이용한 필기체 한글문자의 인식에 관한 연구," 연세대 대학원 박사학위 청구논문, 1988.6
6. 심원태, 김진형, "혼합형 제어전략을 사용한 인쇄체 한글문자의 인식," 한국정보과학회 가을 학술발표논문집, vol14, no.2, 1987.
7. 이성환, 강희중, 김형훈, 박진규, 심원태, 이승호, 김승호, 김진형, "문서인식 및 검색을 위한 전처리 시스템의 설계 및 구현," 한국정보과학회 가을 학술발표논문집 1986년 10월
8. 고찬, 이대영, "형상분석에 의한 한글패턴의 특징추출," 한국통신학회 하계종합학술대회 논문집, pp. 143-148, 1989.8
9. 김진형외 7인, "문서인식 및 처리기의 개발에 관한 연구," 과학기술처, 1988.3.
10. 金錫泰, 堤喜代可, 平井宏, 弓場芳治, "形狀コード列の整合によるハングルの認識", 電子情報通信學會 論文誌, Vol.J71~D.No.I pp.387-394, 1988년 2월
11. L.G Shapiro, "A Structure model of shape", IEEE Trans. on Pptern analysis and Machine Intelligence, vol. PAMI-2, No.2, Mar, 1980.

12. K.Wall and P.E Danielsson, "A Fast sequential method for polygonal approximation", computer vision, Graphics and image processing, vol. 28, pp220-227, 1984.
13. Shungi Mori, Kazuhio Yamamoto, "Research on Machine Recognition of Hand-Printed Characters," IEEE Trans. vol. PAMI-6. no4, July 1 984.
14. K.Yamamoto and A.Rosenfeld, "Recognition of Hand-Printed KANJI Characters by a Relaxation method," proceedings of 6th ICPR 1982, pp 395-398.
15. 長尾直, "畫像認識論," コロナ社, 1983.
16. 森俊二, "文字. 圖形認識技術の基礎," オム社, 1982.
17. 고찬, "컴퓨터 그래픽스," 서울산업대학 출판부, pp153-157, 1986.3.



高 燦(Chan KOH) 正會員
1951年 9月14日生
1974年 2月:慶熙大學校 工科學 卒業
1987年 8月:慶熙大學校 大學院電子工
學科 博士課程 修了
1987年:情報處理 技術士 資格取得
1974年10月~1978年 2月:海軍大學電算
教官(海軍將校 服務)
1982年 2月~1989年 現在:서울産業大學
電算學科 助教授



李大寧(Dai Young LEE) 正會員
1940年 3月18日生
1968年 9月~1970年 3月:캘리포니아洲
立大學院(工學碩士)
1976年 9月~1979年 9月:延世大學校大
學院電子工學科(工學博士)
1971年 9月:慶熙大學校工科學電子工
學科 助教授
1977年 3月:慶熙大學校工科學電子工
學科 副教授
1982年 3月~現在:慶熙大學校工科學電子工學科 教授
1988年 1月~現在:本 學會理事