

자모 결합 유형을 이용한 적은 어휘에서의 필기 한글 단어 인식

(Handwritten Hangul Word Recognition from Small
Vocabulary using Grapheme Combination Type)

진 유 호 † 김 호 연 †† 김 인 중 ††† 김 진 형 †††

(Yu Ho Jin)(Ho Yeun Kim) (In Jung Kim) (Jin Hyung Kim)

요 약 필기 단어 인식 방법에는 낱자별 분할 및 낱자 단위 인식을 통해 인식하는 방법과 단어 사전
을 이용하여 단어와 영상을 직접 비교하는 방법이 있다. 이 중 후자는 인식 대상이 되는 단어들이 작은
수의 어휘로 제한되었을 때 매우 효과적이다. 본 논문에서는 입력 영상이 주어졌을 때 자모를 순차적으로
탐색하고 그 결과의 최적 조합을 찾아 인식하는 사전을 이용한 필기 한글 단어 인식 방법을 제안한다.

입력 영상은 사전의 각 단어와의 매칭을 통해 인식된다. 단어는 필기 순서로 정렬된 자모열로 표현하고
입력 영상은 획들의 집합으로 표현한다. 단어의 자모들은 입력 영상으로부터 추출된 획들의 집합으로부터
단계적으로 탐색된다. 각 단계에서는 전 단계까지의 매칭 상태와 탐색하려는 자모의 형태로부터 자모가 존
재할 것이라고 기대되는 정합 기대 영역을 설정한 후 그 안에서 자모 탐색기를 이용해 자모를 찾는다. 자
모 탐색기는 획들의 집합으로 이루어진 복수의 자모 후보와 그 점수를 출력한다. 각 단계마다 생성된 자
모 후보들은 최적의 단어 매칭을 찾기 위한 탐색 공간을 이룬다. 본 연구에서는 단어 사전을 trie로 구성
하고, 탐색 과정에서 dynamic programming을 이용하여 효과적으로 탐색을 수행하였다. 또한 인식 속도
를 향상시키기 위해 사전 축소, 탐색 공간 축소 등 다양한 지식을 이용하였다. 제안하는 방법은 무제약으
로 쓰여진 필기 단어도 인식할 수 있을 뿐 아니라, 동적 사전을 이용하기 때문에 사전의 내용이 변하는
환경에서도 적용할 수 있다.

인식 실험에서는 39개의 단어로 이루어진 사전에 대하여 613개의 단어 영상에 대해 실험한 결과
98.54%의 높은 인식률을 보임으로써 제안하는 방법이 매우 효과적임을 확인하였다.

Abstract There are two kinds of approaches in handwritten word recognition. One is recognition
by character segmentation and character recognition, while the other is lexicon-driven approach that
compares the input image with word models directly. The latter is effective especially when we want
to recognize words from a small vocabulary. This paper proposes a lexicon-driven handwritten Hangul
word recognition method, which recognizes the input image by sequential grapheme spotting and
finding the optimal combination of the graphemes.

The input image is recognized by matching with every word in the lexicon. A word is represented
by a sequence of graphemes in writing order, while an image is represented by a set of strokes. The
graphemes of a word are searched from the set of strokes, one by one. At every step, an expectation
region for the target grapheme is estimated from the previous matching state and the type of the
target grapheme. The grapheme is searched only within the expectation region by the grapheme
spotter, which gives a number of grapheme hypotheses composed of groups of strokes and their

† 비 회 원 : 한국 오라클 연구원

jinyh@konantech.com

†† 비 회 원 : 한국전자통신연구원 연구원

hoyon@etri.re.kr

††† 비 회 원 : 한국과학기술원 전산학과

ijkim@ai.kaist.ac.kr

†††† 종신회원 : 한국과학기술원 전산학과 교수

jkim@cs.kaist.ac.kr

논문접수 : 1999년 11월 19일

심사완료 : 2000년 11월 9일

efficiency, we organized the lexicon into trie and adopted dynamic programming technique to find the optimal word matching. We also proposed some heuristics to reduce the lexicon and the search space in order to increase the recognition speed. The proposed method can recognize unconstrained handwritten words. Moreover, it is applicable to the environment in which the lexicon may change, because it uses dynamic lexicon.

The efficiency of the proposed method was confirmed by the experiment, in which the recognition rate was 98.54% for 613 word images from 39 classes.

1. 서론

단어 인식은 단어 영상이 주어졌을 때 어떠한 단어인지를 알아내는 기술이다. 단어는 여러 개의 날자로 구성되기 때문에 날자를 인식하는 문자 인식 기술에 비해 복잡도가 더 높다. 반면, 단어는 임의의 문자열이 아니라 사람이 사용하는 어휘이기 때문에 어휘 사전이나 문자쌍들이 서로 연속해서 나타나는 확률 등 문맥 정보를 이용할 수 있는 이점이 있다.

단어를 인식하는 방법은 크게 두 가지로 나눌 수 있다[9,18]. 첫 번째는 날자별 분할과 날자 단위 인식을 통해 인식하는 방법이다. 이 방법은 먼저 분할한 후 인식하는 방법과 분할과 인식을 연동하여 동시에 수행하는 방법으로 다시 나눌 수 있다. 두 번째 방법은 단어 사전을 이용하여 분할을 수행하지 않고 단어 단위로 직접 인식하는 방법이다. 단어 사전은 미리 정해져서 변경할 수 없는 정적 사전과 임의로 변경할 수 있는 동적 사전으로 나눌 수 있다. 정적 사전보다는 동적 사전을 이용할 수 있는 기술이 더 유용하며 높은 수준의 기술을 요구한다.

단어를 분할하여 날자 인식기로 인식할 경우 완성형 한글 2350자에 의한 임의의 조합을 모두 인식할 수 있는 장점이 있다. 따라서 어휘의 수가 매우 많거나 사람의 이름처럼 사전을 구성하기 어려운 경우에 적합하다. 그러나 모든 날자의 조합에 대한 가능성을 고려하기 때문에 복잡도가 매우 높아지는 단점이 있다.

단어를 분할하지 않고 사전을 이용하여 직접 인식할 경우는 사전에 존재하는 단어만을 인식할 수 있는 제한이 있으나, 제한된 어휘의 문자 조합만을 고려하기 때문에 훨씬 효과적으로 인식할 수 있다. 따라서 주소 인식, 형식문서 자동 입력 시스템에서의 필드 인식기 등 제한된 어휘만을 인식하는 응용분야에 적합하다. 한 예로 주소의 특별시/광역시/도에는 여러 변형을 고려하더라도 39개의 단어가 사용된다. 따라서 39가지의 가능성만 고려하면 충분한 반면, 날자 분할을 통해 인식할 경우 가능한 모든 문자 조합을 고려하여야 한다.

본 논문에서는 제한된 수의 단어를 사전을 이용하여 효과적으로 인식할 수 있는 방법을 제안한다. 단어를 문자별로 분할하지 않고 사전 내의 각 단어와의 비교를 통하여 인식한다. 단어와 영상의 비교는 단어의 자모를 영상에서 필기 순서를 따라 단계적으로 탐색함으로써 이루어진다. 각 단계에서는 앞 단계에서 탐색된 자모의 위치와 탐색하려는 자모의 형태에 의해 다음 자모의 존재 영역을 예상하고 그 영역 내에서만 탐색을 수행함으로써 탐색 효율을 높인다. 모든 단어에 대하여 자모를 탐색한 후 영상과 최적의 대응을 이루는 단어를 선택함으로써 인식 결과를 출력한다. 제한하는 시스템은 동적 사전을 이용하기 때문에 사전의 내용이 변하는 환경에도 적용할 수 있다.

2. 관련 연구

2.1 한글 문자별 분할 및 필기 한글 인식

권재욱은 인쇄 한글 문서를 인식하기 위하여 신경망을 사용하였다[13]. 문서에서 문자를 추출하기 수평 투영을 통하여 문자열을 추출한 후 수직 투영으로 WR(white run)과 BR(black run)으로 나눈 후 문자 분리 신경망과 문자 결합 신경망을 이용하여 수직 투영 결과의 분리와 결합을 통하여 문자를 추출해 낸다. 문자 인식은 유형 분류 신경망으로 문자를 6가지 유형의 하나로 판별한 후 이에 따라 유형별 문자인식 신경망이 문자를 자모별로 인식한다.

황순자는 한글의 구조적 특성에 기반을 둔 자모 분할 방법을 제안하였다[14]. 수직 투영을 통한 WRC(white run column), 연결성분(connected component), 수평 투영으로 일정한 두께를 가진 검정 띠들을 묶은 획을 특징으로 사용하였다. 검출한 획을 한글 자모 구성 법칙과 경험적 지식을 이용하여 초성, 수직 모음, 수평 모음과 종성의 4개로 분류하여 분할 결과를 얻었다. 하지만 분할 과정에 많은 경험적 지식이 이용되는 규칙 기반(rule-based) 방법이므로 사용되는 지식의 정당성 문제가 발생할 수 있고, 인식과는 개별적으로 분할이 수행되어 인식 단계에서 분할 오류를 복구할 수 없다는 단점

이 있다.

필기 한글 문자 인식은 다양한 필기 형태와 많은 클래스로 인하여 계속적으로 많은 연구가 이루어지고 있다. Hwang의 연구에서는 영상의 외곽선 추적을 통하여 많은 세그먼트로 나눈 후에, 한글지식을 반영하고 있고 세그먼트의 partition을 만들 수 있는 제약 그래프(constraint satisfying graph)를 만든다[15]. 제약 그래프 내에서 세그먼트를 결합하여 신경망을 이용한 자모 인식기를 통하여 최적의 경로를 인식 결과로 한다.

김호연은 영상에서 획을 추출한 후, 획을 조합하면서 그래프 형태의 계층적 모델과 상향식으로 정합하는 방법을 사용하였다[10]. 김호연의 자모 탐색기는 인식 과정에서 자모의 분할 정보를 내줄 수 있는 반면 Hwang은 미리 분할된 세그먼트에 대하여만 연결을 해 줄 수 있다. 이층식은 단어 영상을 연결 성분(connected component)단위로 나눈 후 김호연의 문자 인식기를 이용하여 최적의 조합을 탐색함으로써 필기 한글 단어를 인식하였다[20].

2.2 무제한 필기 한글 인식기 KAIST-HR

본 논문에서는 김호연의 한글 인식기에 사용되는 한글 자모 탐색기를 이용하여 단어를 인식한다.

2.2.1 계층적 낱자 인식기[10]

한글은 자모로 이루어져 있으며 자모는 획으로 이루어져 있다. KAIST-HR에서는 이같은 구성 원리를 그림 1과 같이 계층적 랜덤 그래프로 표현한다.

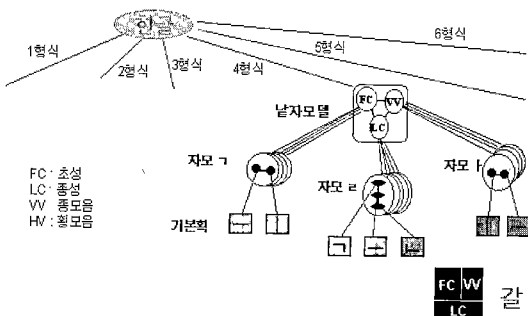


그림 1 KAIST-HR의 계층적 랜덤 그래프

계층적 랜덤 그래프의 가장 아래 계층에서는 기본획을 표현한다. 각 node에서는 기본획의 방향성 및 길이의 확률 분포를 갖는다. 두 번째 계층에서는 기본획의 조합에 의해 자모를 표현한다. 각 node는 하나의 자모를 표현하는데, 각 자모는 첫 번째 계층에서의 서브 그래프(subgraph)에 해당한다. 각 서브 그래프에서 node는 하나의 기본 획을, edge는 기본 획 사이의 관계를

확률적으로 표현한다.

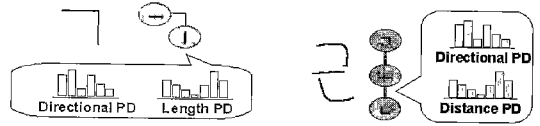


그림 2 기본획의 형태 및 기본획 간의 관계성의 확률적 표현

한글은 중성의 종류 및 받침의 유무에 따라 그림 2와 같이 6가지 형태를 가질 수 있다. 계층적 그래프의 가장 위의 계층은 각 형식에 따른 자모의 조합 규칙 표현하는데, 초성, 중성, 종성의 위치를 각 node에 의해, 자모의 상호 관계를 edge에 의해 확률적으로 표현한다.

입력 영상이 주어졌을 때 매칭 과정은 우선 영상으로부터 획을 추출한 후 획들을 조합하여 기본획의 모델들과 매칭시킨다. 기본획들이 발견되면 이들을 조합하여 자모를 매칭하고, 다시 그 결과를 조합하여 글자를 매칭한다. 이 때 각 매칭 단계에서는 가능한 모든 조합에 대하여 매칭을 시도하기 때문에 무제한으로 쓰여진 필기 문자까지도 잘 인식하지만, 인식 속도가 느린 단점이 있다. 자세한 내용은 [10]을 참조하기 바란다.

본 논문에서는 KAIST-HR의 계층적 그래프 중 기본 획 계층과 자모 계층만을 이용하여 자모를 표현하고, 자모 계층까지만 매칭을 수행하여 자모를 탐색한다.

2.2.2 KAIST-HR을 이용한 필기 한글 단어 인식기 [20]

KAIST-HR을 이용하여 단어 인식을 수행한 예로는 본 논문 이외에도 이층식의 연구가 있다[20]. 우선 입력 영상을 연결 화소(connected component)단위로 나눈 후 x좌표 순서에 의해 나열한다. 각 연결 화소를 node로 표현하고 글자를 이룰 수 있는 연결 화소들의 조합을 edge로 표현하면 그림 3과 같이 그래프를 구성할 수 있다. Node a로부터 b의 edge는 각각 b-a개의 node a+1, a+2,...,b를 조합하여 한 글자를 이룰 수 있음

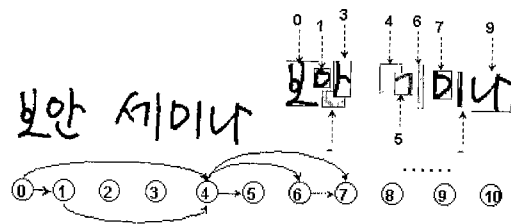


그림 3 층식의 단어인식기

을 나타낸다. 각 edge에는 각 조합을 KAIST-HR에 입력하였을 때의 매칭 점수를 배정한 후 첫 node부터 마지막 node까지의 최적 경로를 탐색하면 최적의 문자 분할 및 인식 결과를 알 수 있다.

이 방법은 사전 없이도 인식할 수 있는 장점이 있으나 복잡도가 높고, 글자간에 접촉이 있을 경우 인식하지 못하는 문제점이 있다.

3. 자모의 결합 유형을 이용한 한글 단어 인식

3.1 단어 정합 개요

한글 단어는 문자로 구성되며 각 문자는 자모로 구성된다. 문자는 중성의 형태와 중성의 존재 유무에 따라 그림 4와 같은 6가지 형식으로 분류된다[13]. 사람이 한글을 쓸 때에는 문자의 형식에 따라 정해진 순서로 자모를 필기한다. 문자 안의 자모 필기 순서는 왼쪽에서 오른쪽으로, 위에서 아래로의 진행이다. 따라서 임의의 단어가 주어지면 그 단어를 구성하는 자모의 필기 순서에 따라 나열하면 $s / l / o / t / r / s / l$ 이 된다. 여기에서 각 자모의 위치는 문자의 6가지 형식에 따라서 대략적으로 추정할 수 있다.

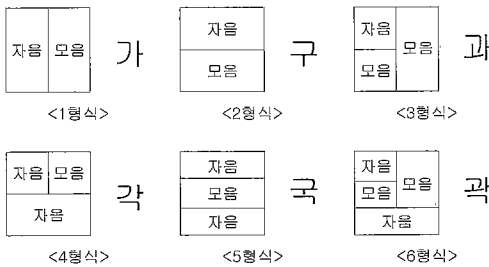


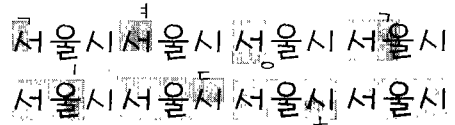
그림 4 한글 문자의 형식 분류

본 논문에서는 영상을 주어진 단어의 자모 필기 순서와 각 자모의 위치 정보로부터 영상에서 자모들을 순차적으로 탐색해 나가는 단어 정합 방법을 제안한다. 자모를 순차적으로 탐색해 나가며 각 단계에서 탐색할 자모의 위치를 이전까지 탐색된 상태로부터 추정한다. 이렇게 추정된 영역을 정합 기대 영역이라 한다. 각 자모는 정합 기대 영역 내에서만 탐색된다. 각 단계까지 정합된 모든 자모의 영역을 정합 완료 영역이라 하며 현재까지의 매칭 상태를 표현한다. 정합 기대 영역에서 자모를 발견하면 이를 이용하여 정합 완료 영역을 확장하며 매칭을 진행한다. 추정된 위치에서 여러 번 자모가 탐색에 실패했을 경우 영상은 주어진 단어에 해당하지 않는다

고 판단할 수 있다.



(a) 영상(서울시)과 단어(서울시)가 일치할 경우



(b) 영상(서울시)과 단어(경기도)가 일치하지 않을 경우
그림 5 영상과 단어의 비교

그림 5는 입력 영상이 주어졌을 때 서울시와 경기도 단어에 대하여 정합을 확장해 나가는 과정이다. (a)는 추정된 위치에서 자모 탐색이 성공적으로 수행되어 영상과 단어가 일치하지만, (b)는 자모 탐색에 실패하기 때문에 영상과 단어가 일치하지 않는 것을 알 수 있다.

정합의 순서도는 그림 6과 같다. 입력 영상으로부터 추출한 획 그래프를 사전의 단어와 순차적으로 탐색한다. 각 단계에서는 목표 자모에 대하여 정합 기대 영역을 설정하고 그 안에서 자모를 탐색(spotting)한다. 자모가 발견되면 이를 이용하여 정합완료 영역을 설정하

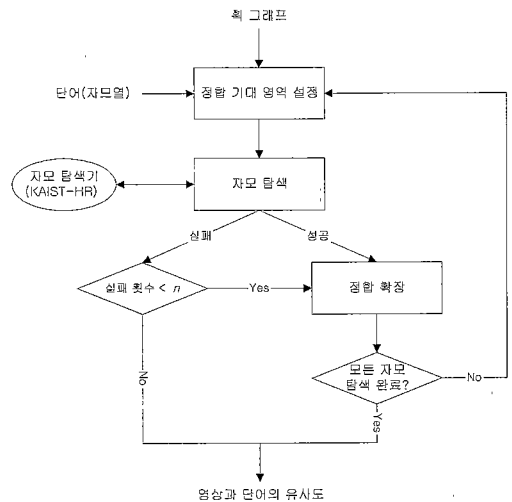


그림 6 영상과 단어 비교 순서도

고, 발견되지 않았을 경우는 그 위치를 추정하여 정합 완료 영역을 설정한다. 자모탐색이 미리 정해진 횟수(n) 이상 실패하였을 경우는 입력 영상과 해당 단어가 일치하지 않는다고 판단하여 정합을 중지한다. 단어내의 모든 자모의 탐색이 끝난 후 최종 정합 완료 영역 중 최적 매칭의 점수가 가장 높은 단어를 인식 결과로 출력한다.

3.2 정합 완료 영역

단어의 자모를 필기 순서로 탐색하면 각 단계에서는 현재까지의 정합 상태를 알 수 있다. 이렇게 영상에서 이미 정합된 자모들에 대응하는 영역을 정합 완료 영역이라 한다. 정합 완료 영역은 마지막에 정합된 자모 위치에 의하여 그림 7(a)와 같은 두 가지 형태로 표현된다. 완전사각형은 정합된 자모가 완전한 문자를 만들거나 1형식 초성과 3형식 첫 번째 모음과 같이 다음에 정합 할 자모가 상하로 겹치지 않는 경우를 의미한다. 불완전사각형은 나머지 자모의 경우로 아래의 오목한 부분에 다음에 정합 할 자모가 존재할 수 있는 것을 의미한다. 그림 7의 (b)는 영상을 단어 “대전”과 비교할 때 정합 단계에 따른 정합 완료 영역의 예를 나타낸다.



(a) 정합 완료영역 유형



(b) 낱자 ‘전’ 자모의 정합 완료영역

그림 7 정합의 진행에 따른 정합 완료 영역의 형태

3.3 정합 기대 영역

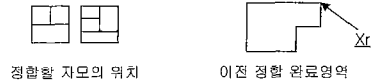
정합 기대 영역은 현 단계에서 정합 할 자모가 존재할 것으로 기대되는 영상에서의 영역이다. 정합 할 자모는 이전 단계의 정합 완료 영역과 인접한 위치에 존재한다. 또, 정합 할 자모의 위치로부터 정보를 얻을 수 있으므로 이들을 이용하여 정합 기대 영역을 설정한다. 정합 기대 영역은 사각형 형태로 표현하고 사각형의 네 좌표($Xmin, Xmax, Ymin, Ymax$)를 결정함으로써 정합 기대 영역을 설정한다.

왼쪽 x좌표($Xmin$)는 이전 단계의 정합 완료 영역의 한 꼭지점 x좌표(Xr)와, 정합에 참가하지 않은 선분들의 끝점의 최소 x좌표(Xs) 중 작은 값으로 설정한다. 그림 8은 정합 할 자모의 위치에 따라 이전 단계의 정

합 완료 영역의 한 꼭지점을 정하는 것을 나타낸다. 첫째, 이전 단계의 정합 완료 영역이 완전 사각형일 경우 Xr 은 그 사각형의 오른쪽 x좌표로 정한다. 둘째, 이전 단계의 정합 완료 영역이 불완전 사각형일 경우는 그림 8와 같이 정한다. 자모의 위치가 오른쪽 아래 오목한 부분에 기대되는 경우는 그림 8(a)와 같이 정한다. 반대로 오른쪽 위 볼록한 부분의 오른쪽에 기대되는 경우는 그림 8(b)와 같이 정한다.

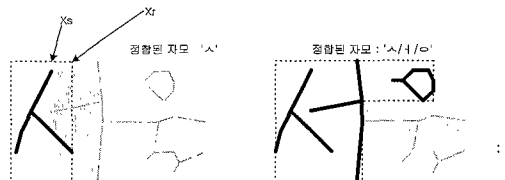


(a) 불완전 사각형의 오목한 부분에 자모가 기대됨



(b) 불완전 사각형의 볼록한 부분에 자모가 기대됨
그림 8 기대 영역의 왼쪽 x좌표 설정

Xr 이 결정되면 정합 되지 않은 선분들의 끝점 중 최소 x좌표(Xs)와 비교하여 더 작은 값이 자모 기대 영역의 왼쪽 x좌표($Xmin$)가 된다. 그림 9(a)에서는 Xs 가 Xr 보다 작으므로 Xs 가 $Xmin$ 으로 선택되었다.



(a) ‘기’ 기대 영역



(b) ‘우’ 기대영역



(c) ‘우’ 기대 영역

그림 9 설정한 정합 기대 영역

정합 기대 영역의 오른쪽 x좌표($Xmax$)는 위와 같이 얻은 왼쪽 x좌표의 값에 추정한 문자 폭을 더한 값이다. 문자 폭은 획의 분포와 단어의 길이에 추정한다. 자모 기대 영역의 위 y좌표($Ymin$)는 $Xmin$ 과 $Xmax$ 사이에

있으며 정합에 참가하지 않은 선분들의 끝점 중 최소 y 좌표이다. 자모 기대 영역의 아래 y 좌표(Y_{max})는 전체 영상에서 가장 아래에 있는 검은 화소의 y 좌표 값이다. 이 값은 전처리 과정에서 세선화 영상의 수평 투영(horizontal projection)을 통하여 얻어진다. 그림 106은 이렇게 설정한 정합 기대 영역의 몇 가지 예를 보여준다.

정합 기대 영역을 설정한 후 그 안에 있는 선분으로부터 자모 탐색기를 이용하여 자모를 이루는 획들의 조합들을 찾는다. 자모 탐색기는 지정된 영역으로부터 복수의 자모 후보 및 그 점수를 출력한다. 이전 정합 완료 영역과 발견된 자모영역을 합쳐 새로운 정합 완료 영역을 만든다.

3.4 정합 확장

영상에서 단어의 자모를 필기 순서로 탐색하는 것은 발견된 자모에 대하여 정합 완료 영역을 확장해 나가는 것과 같다. 각 단계에서 자모 탐색기는 여러 개의 후보를 생성하기 때문에 정합 완료 영역도 여러 개의 후보가 생성되어 전체적으로는 트리 형태의 탐색 공간이 형성되는데, 동일한 정합 완료 영역들을 통합하면 그림 10과 같은 탐색 공간이 형성된다. 각 경로에서 정합 완료 영역의 점수는 이전 단계의 정합 완료 영역의 점수와 발견된 자모영역의 점수를 더한 값이 된다. 이 때 각 경로에 대한 확장 방법은 다음과 같다.

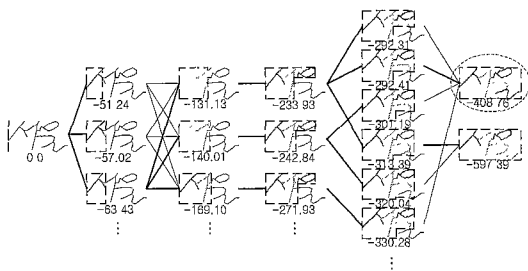


그림 10 “서울”에 대한 탐색 공간

3.4.1 확장된 정합 완료 영역 설정

자모 탐색이 성공하면 출력된 각 자모 후보에 대하여 이전 정합 완료 영역과 자모 후보가 나타내는 자모영역을 더하여 정합 완료 영역을 확장한다. 자모 탐색이 실패하면 자모 영역을 추정하여 설정한 후 그것을 더하여 정합 완료 영역을 확장한다. 자모 영역의 추정은 자모의 위치와 이전 정합 완료 영역과 추정된 문자 너비를 이용하여 추정한다. 문자를 사각형 형태로 추정하여 그 위에서 자모가 차지하는 부분을 구한다. 그림 10은 “ㅇ”의

영상이 왜곡되어 직선 획처럼 추출되었기 때문에 자모 탐색에 실패한 예이다. 이 때 “ㅇ”은 5형식 초성이라는 정보를 이용하여 자모 영역을 추정한다. 자모 탐색이 성공하였을 때와 실패하였을 때 정합 완료 영역의 확장 예는 그림 11과 같다.

정합 완료 영역은 그림 12와 같이 2-3개의 꼭지점으로 그 영역의 외곽 선을 결정할 수 있다. 탐색된 자모 영역은 사각형 형태로 표현된다. 확장된 정합 완료 영역의 형태는 마지막에 정합 된 자모의 위치에 따라 결정하고, 그 영역은 이전 단계 정합 완료 영역과 자모 영역을 포함하는 최소 면적을 갖도록 결정한다.

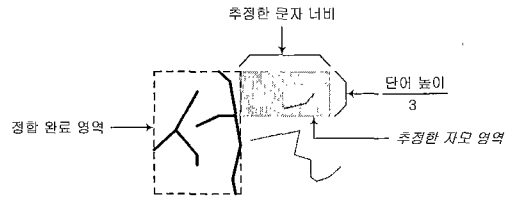


그림 11 자모 영역 추정

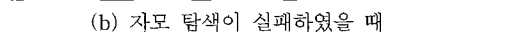
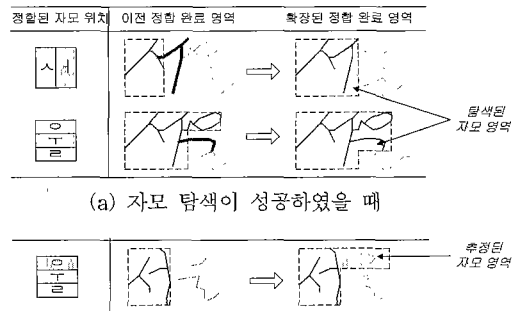


그림 12 정합 완료 영역 확장

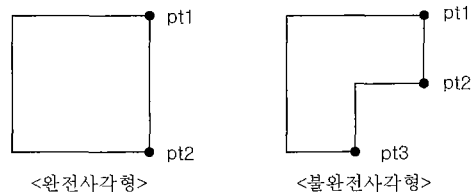


그림 13 정합 완료 영역 표현

3.4.2 정합 완료 영역 점수 계산

정합 완료 영역 점수는 정합 된 자모와 영상과의 유사도를 나타내는 확률 값이다. 정합 완료 영역의 점수는

자모 탐색기로 찾은 각 자모의 확률 값들의 곱이다. 또한 영역 내에 있지만 정합에 사용하지 않은 선분에 대해서는 penalty를 가한다. 본 논문에서는 정합 되지 않은 선분의 길이의 상수배로 penalty를 지정하였다. 자모의 탐색이 끝나면 마지막 정합 완료 영역들 중 가장 높은 값을 가진 것의 점수를 단어와 영상의 유사도로 사용한다. 정합 완료 영역의 점수를 계산하기 위해 다음과 같은 값들을 정의한다.

L : 단어의 자모 개수

$R(l)$: l 번째 자모까지의 정합 완료 영역 후보 집합($1 \leq l \leq L$)

$$R(l) = \{r_1^l, r_2^l, \dots, r_{j_l}^l, \dots, r_{J_l}^l\} \quad (1 \leq j_l \leq J_l)$$

$C(k)$: k 번째 자모의 자모 후보들의 집합($1 \leq k \leq L$)

$$C(k) = \{c_1^k, c_2^k, \dots, c_{j_k}^k, \dots, c_{J_k}^k\} \quad (1 \leq j_k \leq J_k)$$

$p(r)$: 정합 완료 영역 r 의 점수

$p(c)$: 자모 후보 c 의 점수

그림 14는 $(l-1)$ 번째 자모까지의 정합 완료 영역에서 l 번째 자모까지의 정합 완료 영역으로 확장하는 모습을 보여준다. $(l-1)$ 번째 자모까지의 정합 완료 영역에서 l 번째의 자모 영역으로의 확장표는 설정된 정합 기대 영역에서 발견된 자모를 의미한다. 확장된 l 번째 자모까지의 정합 완료 영역은 발견된 자모영역의 수만큼 만들어진다. 정합 완료 영역의 확장 알고리즘은 dynamic programming으로 그림 14와 같다. Iteration에서 연산 자 \cap 는 정합 완료 영역과 자모 영역간의 공통 확을 의미한다. 또한 \oplus 는 정합 완료 영역의 확장을 나타내며 이전 단계의 정합 완료 영역과 자모영역을 합하여 새로이 확장된 정합 완료 영역을 생성하는 연산자이다.

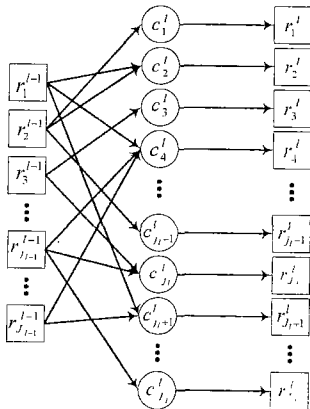


그림 14 정합 완료 영역 확장

Initialization:

$$R(0) = \{r_1^0\} \quad p(r_1^0) = 1$$

Iteration: 1부터 L 까지

$$p(r_h^l) = \max_{j_{l-1}} \{K \cdot p(r_{j_{l-1}}^{l-1}) p(c_{j_l}^l)\} \quad K = \begin{cases} 1 & \text{if } r_{j_{l-1}}^{l-1} \cap c_{j_l}^l = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

$$j_{l-1}^l = \arg \max_{j_{l-1}} \{K \cdot p(r_{j_{l-1}}^{l-1}) p(c_{j_l}^l)\}$$

$$r_h^l = r_{j_{l-1}}^{l-1} \oplus c_{j_l}^l$$

Termination:

$$p(r^l) = \max_{j_l} \{p(r_{j_l}^l)\}$$

그림 15 단어와 영상의 유사도 계산

정합 후에 정합에 사용되지 않은 확들이 남아 있을 때 총 길이에 대한 penalty값과 정합 되지 않은 자모 개수에 대한 penalty값 중 더 큰 값을 마지막 구한 정합 완료 영역의 점수에 더한다. 이와 같이 정합을 진행하여 마지막 자모를 적합한 후의 정합 완료 영역 중 점수가 가장 높은 것을 유사도로 선택한다.

4. 지식을 이용한 인식 속도 향상

본 단어 인식기는 사전으로부터의 문맥 정보를 효과적으로 이용하여 변형이 많은 필기에 대하여도 매우 높은 수준의 인식 성능을 보인다. 그러나, 테이타에 나타날 수 있는 다양한 변형을 거의 모두 고려하기 때문에 매우 상세한 비교를 수행하고, 또한 탐색 공간이 크다. 하지만, 단어 인식에 사용 가능한 다양한 지식을 이용한다면 훨씬 빠른 인식 속도를 얻을 수 있다. 본 단어 인식기의 인식 속도는 탐색 공간의 크기와 자모탐색기의 호출 횟수에 의해 영향을 받는다. 이 같은 요소들에 대하여 지식을 이용함으로써 인식속도를 향상시켰다.

4.1 탐색 공간의 축소

입력 영상을 사전의 단어들과 비교하는 과정에서 자모 탐색기 HR이 발견한 자모 후보들은 그림 15와 같은 탐색 공간을 이룬다. 이를 축소할수록 더 빠른 인식 속도를 얻을 수 있다.

4.1.1 정합 완료 영역 내에서 정합에 사용하지 않은 확에 의한 pruning

정합 완료 영역내의 자모 정합에 사용하지 않은 확이 많으면 잘못 탐색된 자모일 가능성이 높다. 그 기준은 정합 완료 영역의 마지막 자모 정합에 사용한 확의 총 길이와의 비율을 사용한다. 그 비율이 정한 임계상수(D)

이상일 때에 그 정합 완료 영역을 pruning한다. D 를 조절함으로써 pruning 정도를 조절할 수 있다.

l_1 : 정합 완료 영역 내의 정합되지 않은 획의 총 길이

l_2 : 마지막 정합된 자모에 사용한 획들의 총 길이

$$r = \frac{l_1}{l_2}$$

if ($r > D$) do pruning

l_1 값을 계산할 때 부분적으로 정합 완료 영역에 포함되는 획들은 영역에 포함된 만큼만 더하였다.

그림 16는 $D = 1.0$ 이고 “ㄱ”의 자모 후보에 대하여 각각의 정합 완료 영역에 대하여 pruning을 적용한 예이다. 아래의 정합 완료 영역은 실제 “ㄱ”의 획으로 자모 “ㄱ”를 얻었기 때문에 실제 “ㄱ”의 획에 의하여 r 값이 1이상이 되어 pruning 된다.

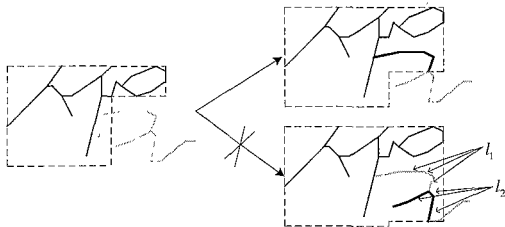


그림 16 정합되지 않은 획 길이 제한에 의한 pruning

4.1.2 단계별 정합 완료 영역 개수 제한

자모를 정합하고 생긴 정합 완료 영역 중 점수 순으로 상위 n 개만을 확장에 이용한다. 매칭을 이같이 진행함으로써 각 단계에서 확장되는 탐색 공간의 크기를 제한할 수 있다. 이는 트리 구조에서 child 노드의 수를 최대 n 개로 제한하는 것과 동일하다.

4.1.3 사전 축소

본 단어 인식기는 인식의 대상이 되는 사전이 적을수록 탐색 공간이 작아진다. 따라서 매칭을 시작하기 전 지식을 이용하여 사전의 크기를 축소하여 인식 속도를 더 향상하였다. 본 연구에서는 입력 영상으로부터 추출한 획들의 총 길이를 이용하여 사전을 축소하였다.

입력 영상으로부터 추출한 획들의 길이를 모두 더한 후 글씨의 크기에 의한 영향을 제거하기 위해서 영상의 높이로 정규화 한다. 이 값을 l_{input} 라고 하자. 사전의 단어 w 마다 가능한 획의 총 길이의 최대(l_{max}^w), 최소값(l_{min}^w)을 훈련 데이터로부터 얻는다. 매칭을 시작하기 전 다음 조건을 만족하는 단어에 대해서만 탐색을

수행한다.

$$l_{max}^w \leq l_{input} \leq l_{min}^w$$

4.1.4 Heuristic을 이용한 pruning

매칭의 각 단계에서 정합 완료 영역으로부터 확장될 수 있는 매칭의 score를 추정함으로써 탐색 공간 축소에 사용한다. 각 정합 완료 영역 x 에 대하여 score 함수는 다음과 같이 나타낼 수 있다.

$$f(x) = g(x) + h(x)$$

$g(x)$: 현 단계까지 정합 완료된 자모의 score합

$$g(x) = \sum score(c_i)$$

$score(c_i)$: 정합된 자모 c_i 의 score

$h(x)$: 아직 정합되지 않은 획들에 대한 추정값

본 연구에서는 $h(x)$ 를 다음과 같이 정의하여 사용하였다.

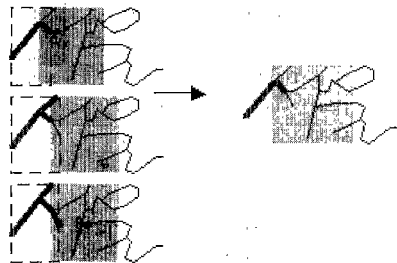
$$h(x) = \sum l_j * k$$

l_j : 정합되지 않은 획 j 의 길이

k : 상수

4.2 자모 탐색기 이용의 효율화

본 인식기에서 HR의 호출 횟수는 전체 인식 시간에 큰 영향을 미친다. 탐색과정 동안 매우 많은 수의 자모 기대 영역이 설정되어 자모 탐색기를 호출한다. 그 중에는 유사한 자모 기대 영역이 많은데, 이들 각각에 대하여 자모 탐색기를 호출하는 것보다는 유사한 영역은 통합하여 한 번만 호출하게 하면 자모 탐색기의 호출 횟수를 크게 줄일 수 있다. 그림 17에서 왼쪽 세 개의 자모 기대 영역을 통합하여 오른쪽과 같은 영역으로부터 자모 탐색기를 호출한다. 그 후 탐색된 자모 후보들 중 각 자모 기대 영역에 속하는 것들을 이용한다.



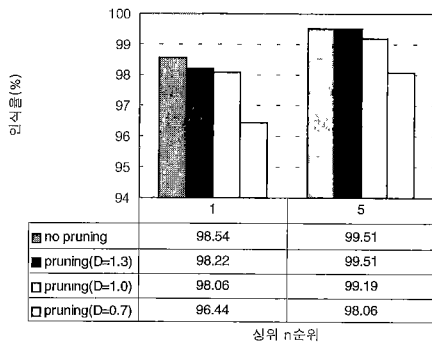
각각의 기대 영역 통합된 기대 영역

그림 17 기대 영역의 통합

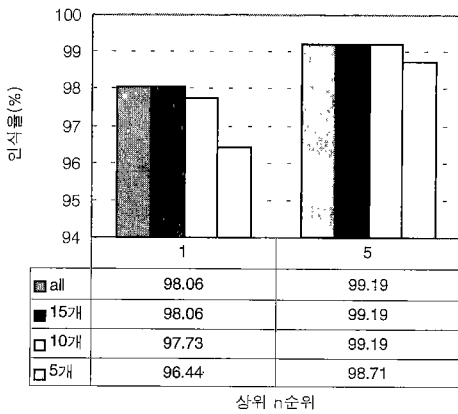
5. 실험

5.1 실험 환경 및 결과

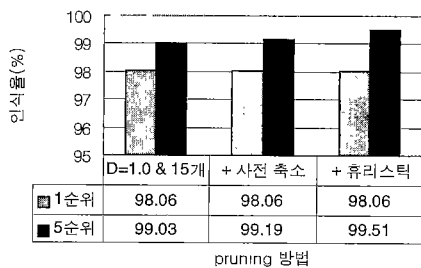
본 논문에서 제안된 방법론은 Pentium II Celeron



(a) 임계상수(D)에 따른 실험



(b) 정합 완료 영역 개수 제한 실험



(c) 사전 축소 및 휴리스틱 pruning에 대한 인식률
그림 18 인식률

450Mhz PC에서 Visual C++을 이용해 구현, 분석을 하였다. 획 추출 과정에서는 강경원의 방법을 사용하였다 [11]. 단어 영상 데이터는 약 170여명의 필기자들이 우편 주소를 주고 그대로 규격 편지 봉투와 데이터 용지에 쓴 것을 스캐너를 통하여 얻었다. 봉투 영상의 경우

는 발신인과 수취인의 두 주소를 모두 실험에 사용하였다. 우편 주소는 계층적인 지역 구분명으로 구성되어 있다. 계층의 가장 첫 단계인 특별시/광역시/도의 지역구분에 해당하는 단어영상을 전체 주소 영상에서 수작업으로 분할하여 만들었으며 실험에 사용한 단어 영상 데이터는 300dpi로 획득하였으며 총 개수는 613개이다.

사전은 우편 주소의 특별시/광역시/도의 지역명이며 실제 지역명은 13개뿐이나 같은 지역이라도 지역구분명을 쓰지 않거나 줄여서 쓰는 경우의 지역명 단어도 포함한다. 예를 들어 “서울특별시”의 경우 “서울”, “서울시”와 같이 쓰는 경우가 있게 되기 때문에 사전은 총 39개의 단어를 가진다.

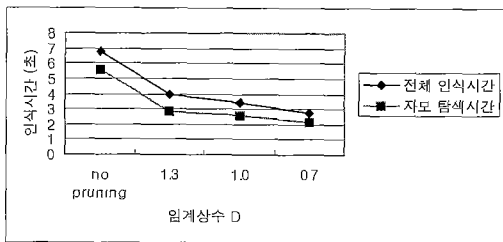
인식률 실험은 탐색 공간의 축소에 대하여 그림 18과 같다. 정합 되지 않은 획에 의한 pruning 임계상수(D)가 작아지거나 단계별 확장할 정합 완료 영역 개수를 줄이면 인식률이 떨어지게 된다. 이는 올바른 정합 완료 영역이 pruning될 가능성이 높아지기 때문이다. Heuristic pruning에서 $k = -12$ 를 사용하였다. 사전 축소 및 heuristic pruning을 하였을 때에는 1순위 인식률이 감소되지 않았고, 5순위 인식률은 오히려 증가하였다. 5순위 인식률이 증가한 것은 정인식 결과와 경쟁이 되는 후보가 미리 제거되었기 때문이다.

인식 속도는 단어당 1.28초이다. 그림 19는 다양한 지식이 인식 속도에 미치는 영향을 보여준다. 많은 지식을 사용할수록 불필요한 계산을 생략하기 때문에 인식 속도를 속도 향상시킬 수 있다. 사전 축소 실험의 경우 평균적으로 39개의 사전 내 단어를 14.1개로 축소하여 63.8%의 사전 축소율을 보였다.

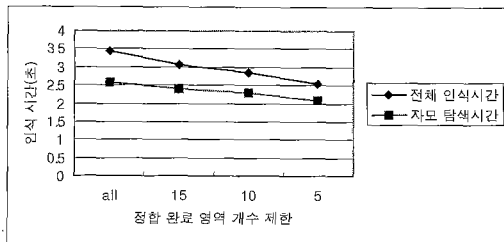
2절에서 설명한 이층식의 방법과 실험 결과를 비교하였다[20]. 이층식의 방법은 문자간 접촉이 있는 경우 인식할 수 없기 때문에 접촉이 있는 영상 데이터를 제외한 486개의 데이터에 대하여만 실험을 하였다. 그 결과 인식률은 80.45%와 평균 인식 속도는 4.1초를 보였다.

그림 19에서 알 수 있듯이 인식 시간의 대부분은 자모 탐색기에서 소모된다. 이는 HR이 무제한으로 흘러쓴 글씨를 인식하기 위해 매우 다양한 획의 조합을 고려하기 때문이다. 그러나, 본 시스템과 같이 HR을 이용하여 흘러쓴 글씨를 인식할 수 있는 이층식의 방법과 비교하면 상대적으로 빠른 것을 알 수 있다. 만약 필기 형태에 제약을 가하거나 인식률을 약간 희생시킬 경우 더 빠른 자모 탐색기를 구현할 수도 있을 것이다.

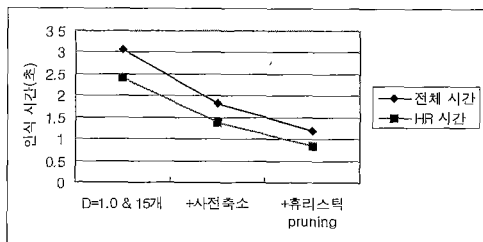
그림 20는 정인식한 영상의 예이다. 자모간에 심하게 붙은 “서울”과 “부산광역시”의 영상도 인식하였다. 또한, 첫 번째의 “서울특별시” 영상의 경우에는 “울”자의 “ㄹ”



(a) 정합 되지 않은 획의 길이에 의한 pruning



(b) 단계별 정합 완료 영역의 개수 제한에 의한 pruning (D=1.0)



(c) 사전 축소 및 휴리스틱에 의한 pruning (D=1.0, 정합 완료 영역 개수 15개)
그림 19 인식 속도

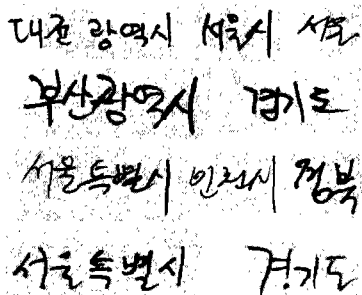


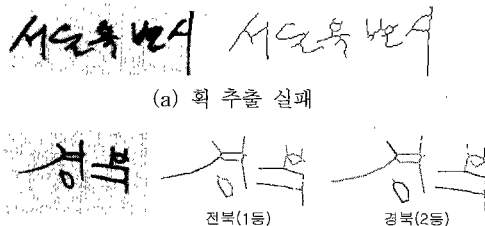
그림 20 정인식 영상

과 "특"자의 "一"가 부드럽게 연결되어 있음에도 불구하고 두 자모를 모두 잘 분리하여 인식을 하였다. 이 경우는 미리 분할점을 찾는 경우 이와 같이 부드럽게 붙어

있는 획들은 분리하기가 어렵기 때문에 인식 오류가 나타난다. 두 번째의 "서울특별시"와 "경기도" 영상은 "ㅇ" 획이 접촉되어 직선형태의 획으로 나타나서 자모 탐색이 실패하지만 자모영역을 추정하여 올바르게 인식한 것을 보여준다.

5.2 오류 분석

단어 인식의 오류는 크게 획 추출 실패와 자모 탐색 오류로 분류될 수 있다. 몇 개의 획 추출 실패로 인한 자모 탐색 실패는 본 논문에서 자모영역 추정으로 해결할 수 있지만 그림 21 (a)와 같은 경우는 획 추출이 어려워 자모 탐색 실패와 잘못된 획 추출이 오히려 잘못된 자모탐색을 유발하여 오인식되었다. 그림 21 (b)는 "ㄱ" 수직획이 매우 길어서 부분적으로만 찾고 나머지는 모두 penalty가 되었다. 반면 전복과 정합에서는 "스"과 "ㄴ"에 해당되는 획을 잘못 찾아내었지만 전체적으로 전복보다는 높은 점수를 얻어 오인식되었다. pruning을 하지 않은 인식을 98.54%의 실험결과에 대하여 오인식 영상을 분류하면 표 1과 같다.



(b) 자모 탐색 오류

그림 21 오인식 영상

표 1 오인식 분류

원인	개수	비율(%)
획 추출	3	33
자모 탐색	6	67

6. 결론

본 논문에서는 제한된 어휘의 필기 한글 단어 인식 방법을 제안하였다. 입력 단어를 사전의 각 단어들과 비교함으로써 인식을 수행한다. 단어와 입력 영상의 비교는 단어의 자모들과 영상에서 순차적으로 탐색함으로써 이루어진다. 각 단계에서 정합된 자모에 대응하는 영역인 정합 완료 영역을 설정하고, 이전 단계의 정합 상태와 탐색하려는 자모의 형태로부터 자모가 예측되는 영역인 정합 기대 영역을 추정하였다. 정합은 김호연의 자모탐색기를 이용하여 정합 기대 영역에서만 이루어지며

자모가 탐색될 때마다 정합 완료 영역의 확장으로써 정합을 진행한다. 자모 탐색기에서 출력된 자모 후보를 dynamic programming을 이용하여 최적으로 결합하는 단어를 인식 결과로 출력한다. 사전을 trie로 구성하고 사전 축소, 탐색공간 축소 등 지식을 이용함으로써 인식 속도를 높일 수 있었다. 제안하는 방법은 제한된 어휘 내에서 단어를 인식하는 경우에 효과적이며, 무제약으로 쓰여진 필기 한글 단어를 인식할 수 있을 뿐 아니라 동적 사전을 사용하기 때문에 사전이 변하는 환경에도 적용할 수 있다.

실험에 의하여 본 방법은 우편 주소의 특별시/광역시/도의 지역명 단어 영상과 39개의 단어 사전에서 98.54%의 높은 인식률을 보였다. 실험에서는 인식 속도가 느린 것으로 나타났다. 이는 실험에 사용된 자모 탐색기가 무제약 필기 인식을 위해 매우 다양한 형태의 정합을 시도하기 때문이다. 필기 형태에 약간의 제약을 가하거나 자모 탐색기의 속도를 개선한다면 더 빠른 인식 속도를 얻을 수 있을 것이다.

참 고 문 헌

- [1] G. Kim and V. Govindaraju, A Lexicon Driven Approach to Handwritten Word Recognition for Real-Time Applications, *IEEE Transaction on Pattern Recognition and Machine Intelligence*, Vol. 19, No. 4, April 1997
- [2] R. Bozinovic and S. Srihari, Off-Line Cursive Script Word Recognition, *IEEE Transaction on Pattern Recognition and Machine Intelligence* Vol. 11, No. 1 pp. 68-83 January 1989
- [3] M. Leroux, J. Salome and J. Badard, Recognition of Cursive Script Words in a Small Lexicon, *1st International Conference on Document Analysis and Recognition*, pp.774-782, 1991
- [4] M. Cote, E. Lecolinet, M. Cheriet and C. Suen, Automatic Reading of Cursive Scripts Using a Reading Model and Perceptual Concepts, *First Issue of International Journal of Document Analysis and Recognition*, 1997
- [5] R. Powalka, N. Sherkat and R. Whitrow, Word Shape Analysis for a Hybrid Recognition System, *Pattern Recognition*, Vol.30, No. 3, pp.421-445, 1997
- [6] D. Elliman and I. Lancaster, A Review of Segmentation and Contextual Analysis Techniques for Text Recognition, *Pattern Recognition*, Vol.23, No.3/4 pp.337-346, 1990
- [7] J. Simon, Off-Line Cursive Word Recognition, *Proc. of IEEE*, Vol. 80, No. 7, July, pp.1150-1161, 1992
- [8] A. Senior and A. Robinson An Off-Line Cursive Handwriting Recognition System, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 3, pp.309-321, March 1998
- [9] R. Casey and E. Lecolinet, A Survey of Methods and Strategies in Character Segmentation, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.18, No. 7, pp.690-706, July 1996
- [10] H. Kim and J. Kim, Handwritten Korean Character Recognition Based on Hierarchical Random Graph Modeling, *6th International Workshop on Frontiers in Handwriting Recognition*, pp.577-578, 1998
- [11] E.Fredkin, Trie memory, *Communications of ACM*, Vol. 3., No. 9, pp.490-500, 1960
- [12] K. Kang, J. Suh and J. Kim Skeletonization of Grayscale Character Images Using Pixel Superiority Index, *IAPR Workshop on Document Analysis Systems*, pp.326-335, 1998
- [13] 권재욱, 계층적 신경망을 이용한 다중 크기의 다중 활자체 한글문서 인식, *KAIST 석사학위 논문*, 1991
- [14] 황순자 and 김문현, 자소 클래스 인식에 의한 off-line 필기체 한글 문자 분할, *한국정보처리학회 논문지 제3권 제4호*, pp.1002-1013, 1996
- [15] Y.S. Hwang and S.Y. Bang, Recognition of a Handwritten Korean Character By Combining Segments Using Constraint Satisfying Graph, *6th International Workshop on Frontiers in Handwriting Recognition*, pp.515-526, 1998
- [16] M. Koga, R. Mine, H. Sako, and H. Fujisawa, Lexical Search Approach for Character-String Recognition *IAPR Workshop on Document Analysis Systems*, pp.237-251, 1998
- [17] R. Bozinovic and S. Srihari, A String Correction Algorithm for Cursive Script Recognition, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.4, No.6, November 1982
- [18] E. Lecolinet and O. Baref, Cursive Words Recognition: Methods and Strategies, *Fundamentals in Handwriting Recognition*, Springer-Verlag, pp.235-263, 1993
- [19] S. Madhavanath and V. Govindaraju, Holistic Lexicon Reduction, *3rd Workshop on Frontiers in Handwriting Recognition*, pp.71-81, 1993
- [20] 이충식, Handwritten Hangul Word Segmentation, personal communication
- [21] J.T. Favata, S.N. Srihari and V. Govindaraju, "Off-Line Handwritten Sentence Recognition," *Proc. IWFHR'96*, pp.171-176, Essex, England, 1996.



진 유 호

1991년 3월 ~ 1995년 8월 한국과학기술원 전산학과 학사. 1997년 3월 ~ 1999년 2월 한국과학기술원 전산학과 석사. 1999년 4월 ~ 2000년 10월 한국 오라클 전자상거래 컨설턴트. 2000년 10월 ~ 현재 (주)코난 테크놀로지 연구원



김 호 연

1992년 연세대학교 전산학과 학사. 1994년 한국과학기술원 전산학과 석사. 1997년 1월 ~ 2월 일본 NHK放送技術研究所 방문연구원. 1999년 한국과학기술원 전산학과 박사. 1999년 9월 ~ 11월 독일 SIEMENS ElectroCom 방문연구원. 1999년 ~ 현재 한국전자통신연구원 선임연구원. 관심분야는 패턴인식, 문자인식, HMM, 영상처리, 신경망, 기계학습



김 인 중

1994년 한국과학기술원 전산학과 졸업. 1995년 한국과학기술원 전산학과 석사학위 취득. 1995년 ~ 현재 한국과학기술원 전산학과 박사과정. 관심분야는 패턴인식, 인공지능, 문자인식



김 진 형

1971년 서울대학교 공과대학 졸업. 1979년 UCLA 전산학 석사학위 취득. 1983년 UCLA 전산학 박사학위 취득. 1973년 ~ 1976년 KIST 전산실 연구원. 1981년 ~ 1985년 미국 Hughes 인공지능센터 선임연구원. 1986년 ~ 1988년 한국정보과학회 산하 인공지능연구회 위원장. 1989년 ~ 1990년 미국 와트슨 연구소 초빙 과학자. 1985년 ~ 현재 한국과학기술원 전산학과 교수. 1991년 ~ 1999년 인공지능연구센터 부소장. 1995년 ~ 1999년 연구개발정보센터 소장. 1997년 ~ 현재 한국공립한림원 회원. 1999년 ~ 현재 한국과학기술한림원 회원. 관심분야는 패턴인식, 인공지능