

문자의 구조적 제약과 동적 격자 탐색을 이용한 필기 한글 문자 인식

강경원^o 김진형
한국과학기술원 전자전산학과
{kwkang, jkim}@ai.kaist.ac.kr

Handwritten Hangul Recognition by Dynamic Lattice Search with Structural Constraints

Kyung-Won Kang^o Jin-Hyung Kim
Dept. of EECS, KAIST

요 약

필기 한글문자 인식은 다양한 필기 변형, 자모 간의 접촉과 같은 문제들을 내포하고 있다. 최근 이를 해결하기 위한 방법으로 랜덤 그래프를 이용한 필기 한글 모델링이 제안되었으나, 상향식 정보처리의 한계인 시간 복잡도 문제를 겪고 있다. 영어 단어인식에 관한 인지과학적 연구에서는 하향식 정보처리의 주요한 역할 중 하나로 인식 과정에서의 계산 중복을 없애는 필터링의 역할을 들고 있다. 본 논문에서는 랜덤 그래프를 이용한 필기 한글 모델링을 기반으로 하여 필기체에 나타나는 다양한 변형을 흡수하며, 시간 복잡도를 해결하기 위한 한글 문자의 구조에 바탕을 둔 하향식 정보처리 방법을 제안한다. 제안하는 방법은 모델 발화를 이용한 자모 후보 추출, DP 정합과 동적 격자 탐색을 이용한 문자 후보 탐색, 그리고 문자의 구조적 제약을 이용한 후보 제거 기법을 포함한다. 필기 한글 데이터베이스인 SERI-DB에 대한 예비 실험 결과, 제안한 방법은 인식률의 큰 저하 없이 상향식 정보 처리에 바탕을 둔 기존 방법에 비해 높은 속도 향상을 가져 왔다.

1. 서론

문자인식은 인간의 인지능력을 대표하는 패턴인식 문제로서 상업적 응용이 활발해지고 있으나 현재는 방법론적인 한계에 도달하였다. 구조분석이나 패턴정합, 신경망, Hidden Markov Model 등 패턴인식 분야의 다양한 기술과 인공지능 기법들은 대부분 통계적 인식방법을 적용한 상향식 정보처리에 의존하고 있다. 그러나, 현재의 기술로는 제한적인 환경에서만 동작할 뿐, 복합적인 문제가 얽혀있는 실제 환경에서의 적용은 어려운 것이 사실이다.

필기 한글문자 인식은 많은 어려운 문제를 내포하고 있다. 먼저, 필기 한글문자는 작은 획에 의하여 그 의미가 변할 수 있으며(나와 ·의 차이), 각 획의 길이 및 방향 등이 다양한 형태의 필기 변형을 갖는다. 게다가 한글문자는 초성, 중성, 종성이 2 차원 공간에서 배치됨으로써 획들을 조합하는 순서에 따라 여러 문자로

해석이 가능하다 (그림 1). 따라서 올바른 인식결과를 갖기 위해서는 이러한 해석들을 모두 고려하여야 한다.

이러한 문제 복잡도로 인하여 필기 한글 인식 문제는 고난이도의 문제로 간주되었으며 최근에 들어서야 기본적인 연구들이 수행되었다. 대표적인 것으로 필기체에서 나타나는 다양한 변형을 흡수하기 위한 방법으로 랜덤 그래프를 이용한 필기 한글 모델링 및 확률계산에 근거한 문자인식 방법론이 제안되었다 [8]. 이 방법은 전적으로 상향식 정보처리에 기반을 두고 있다 (그림 2). 입력 영상으로부터 subgraph isomorphism 을 통하여 자모 후보들을 추출하고, 다음으로 추출된 자모 후보들을 조합하여 문자 후보를 생성하였다. 마지막으로 생성된 문자 후보들과 문자 랜덤 그래프를 정합 함으로써 최적의 인식결과를 도출하였다. 이 방법론은 필기 한글에서 나타나는 다양한 형태의 변형을 통계적 처리를 통하여 흡수할 수 있었으나, 이로 인하여 자모 후보 및 문자 후보 개수가 기하급수적으로 증가하는 문제점을 보였다.



그림 1. 필기 한글 인식의 어려움

이는 상향식 정보처리의 어쩔 수 없는 한계이며 인간의 인식 메커니즘을 적용한 하향식 정보처리에 관한 연구가 필요하다. 특히, 복잡한 배경에서 중요한 특징만을 선택적으로 주의 집중하는 인간의 능력은 인식에 있어 매우 중요한 것으로 평가된다.

최근 들어 신호 레벨에서의 인간의 시각정보 처리 및 이를 응용한 시각 정보처리 시스템에 관한 연구가 상당수 수행되고 있다. 반면, 인식과 같은 상위 정보처리 과정은 인지심리학적 실험을 통하여 조금씩 규명되고 있으나 아직은 문자인식 및 물체인식 등에 적용될 만큼 충분히 밝혀지지는 않았다.

문자인식 분야를 살펴 보면, 인지심리학적 연구에 기반한 대표적인 영어 단어인식 모델로는 Interactive Activation Model (IAM)이 있으며, 최근 이를 응용한 영어 단어인식 시스템이 개발되었다 [5]. 하지만, 이 시스템은 특징 추출 및 인간의 하향식 정보처리가 1 차원 공간으로 한정되어 있어 필기 한글에 적용하기에는 문제가 있다.

Wickens 는 여러 심리학 연구를 토대로 영어 단어인식에서의 하향식 정보처리의 역할을 두 가지로 정리하였다 [6]. 첫번째로 단어인식 과정에서의 계산 중복(redundancy)을 없애는 필터링의 역할을 수행한다. 즉, 특정 특징, 문자, 단어들을 주변 특징이나 문자들에 의하여 예상함으로써 하향식 정보처리가 상향식 정보처리만을 수행하는 경우보다 효율적일 수 있다는 것이다. 다른 하나는 많은 문헌들에서 밝히고 있는 단어우월효과(word superiority effect)이다. 최근 한글 문자인식에서도

이러한 단어우월효과와 유사한 문자우월효과가 존재한다는 연구가 보고되었다 [7]. 하지만 이를 실생활 필기 한글 문자인식에 적용하기에는 아직 부족함이 있다. 이는 초성, 중성, 종성에 나타날 수 있는 자모의 수가 매우 많아 문맥정보를 이루기 어렵고, 이로 인하여 효율적인 시스템을 구축할 수 없기 때문이다. 따라서 본 연구에서는 하향식 정보처리의 주된 역할을 상향식 정보처리에서 발생하는 계산 중복의 제거로 본다.

본 연구는 랜덤 그래프를 이용한 필기 한글 모델링을 기반으로 필기 문자에서의 변형을 흡수하고, 여기서 발생하는 시간 복잡도 문제를 선택적 주의집중과 인지 과정에 대한 연구를 바탕으로 해결하고자 한다. 본 논문은 다음과 같이 구성되어 있다. 먼저 2 절에서 문자 정합의 기본이 되는 랜덤 그래프(random graph)를 이용한 필기 한글 모델링에 관하여 설명하고, 3 절에서는 동적 격자 탐색 및 문자의 구조적 지식을 이용한 필기 한글 인식 방법론을 제안한다. 마지막으로 4 절에서 제안한 방법의 효율성을 실험을 통하여 증명하고 결론을 맺도록 한다.

2. 랜덤 그래프를 이용한 필기 한글 모델링 [8]

문자인식 방법론은 전처리의 관점에서 크게 구조적인 방법과 통계적인 방법으로 분류된다. 구조적인 방법은 문자 영상으로부터 획을 추출한 후 획들간의 관계를 분석하여 인식하는 방법이며, 통계적인 방법은 영상으로부터 직접 특징을 추출하고 신경망 등의 분류기를 통하여 인식하는 방법이다. 구조적인 방법은 획이라는 개념적인 단위를 이용함으로써 좀더 세밀한 분석이 가능한 반면, 필기 문자에서의 다양한 변형을 해결하기 어렵다는 단점이 있다. 이에 반하여 통계적인 방법은 특징 추출과 분류는 용이한 반면, 어느 특징을 어떻게 이용할 것인가에 관한 문제점을 가지고 있다.

최근에는 구조적인 방법과 통계적인 방법을 결합하여 각 방법론의 문제점을 해결하고 있다. 그 중의 하나로 랜덤 그래프에 기반을 둔 방법이 있으며 필기 문자 인식 문제에 성공적으로 적용되었다. 랜덤 그래프란 vertex 와 arc 들이 확률적 랜덤 변수인 그래프를 말하며, 그래프의 위상(topology)을 통하여 문자 구조를 표현하며 각 랜덤 변수의 확률분포를 이용하여 문자 구조의 변형을 확률적으로 모델링 한다. 따라서 구조적인 방법인면서도 통계적인 방법의 장점을 갖는다.

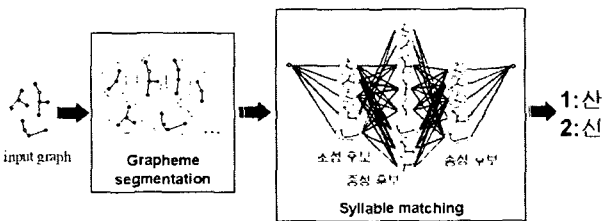


그림 2. 상향식 정보처리에 의한 한글인식 방법론

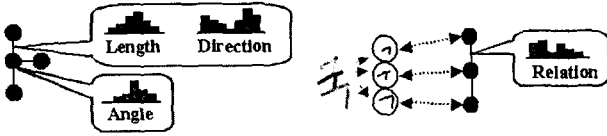


그림 3. Random graph를 이용한 필기 한글 모델링

한글은 자모가 모여 문자를 이루는 구조를 가지고 있으며 이 사실을 이용하여 문자를 계층적으로 표현할 수 있다 (그림 3). 자모 모델의 경우, 각 vertex 는 획의 끝점이나 교차점 등의 특징점을 모델링하며 각 arc 는 특징점을 연결하는 획을 모델링한다. 그리고 문자 모델의 경우, 각 vertex 는 자모 모델의 발현을 모델링하며 각 arc 는 이러한 발현들 사이의 관계를 모델링한다. 이때 인식의 문제는 아래의 확률식을 최대화 하는 입력 그래프로부터 문자 모델 그래프로의 사상 혹은 정합으로 정의된다 (X: 입력 문자 그래프, G: 문자 랜덤 그래프 모델, M: X로부터 G로의 사상).

$$\arg \max_{(M,G)} P_M(G|X) = \arg \max_{(M,G)} P_M(X|G)P(G)$$

랜덤 그래프를 이용한 한글 인식 방법론은 높은 시간 복잡도 문제를 내포하고 있다. 직관적으로 볼 때 가능한 모든 사상 중에서 가장 높은 확률을 갖는 사상을 찾아야 하므로 NP-complete 문제라 볼 수 있다. 구체적으로 살펴 보면 랜덤 변수를 이용하여 확률적으로 모델링 함으로써 입력 그래프에서 발현되는 자모 후보들의 개수가 많아지게 되며, 이로 인하여 문자 후보의 개수가 기하급수적으로 증가한다. 이는 전체 프로세스가 상향식 정보처리에만 바탕을 둬 기인한다. 다음 절에서는 이러한 시간 복잡도를 줄이기 위한 동적 격자 탐색 및 후보 제거 기법(candidate pruning technique)에 관하여 설명한다.

3. 동적 격자 탐색 및 문자의 구조적 지식을 이용한 필기 한글 인식

제안하는 시스템은 그림 4 와 같이 구성되어 있다. 문자 영상은 전처리, 획 추출, 그래프 생성을 거쳐 한글 6 형식에 따른 6 개의 격자 탐색기로 입력된다. 각각의 격자 탐색기는 각 형식에 맞는 문자 후보들을 생성하며 이 중 가장 높은 인식 확률을 갖는 후보가 최종 인식 결과로 선택된다.

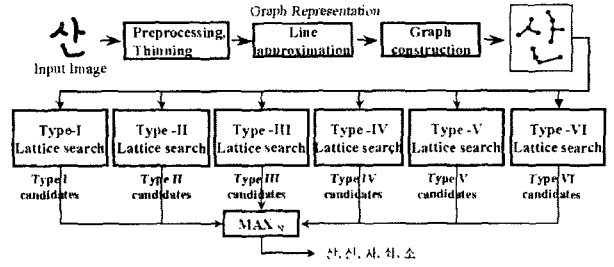


그림 4. 제안하는 시스템의 구조

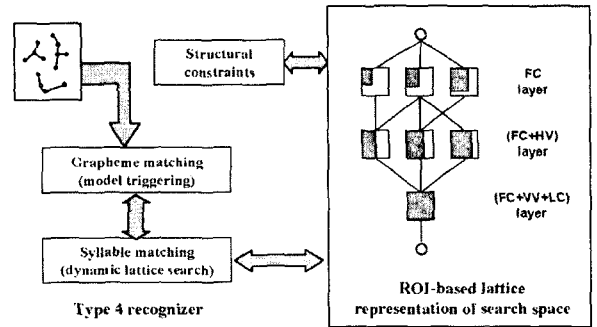


그림 5. 한글 6형식에 따른 형식 인식기

각 격자 탐색기는 모델 발화 기법을 이용한 자모 후보 추출과 격자 탐색을 이용한 문자 후보 정합, 그리고 문자의 구조적 제약을 이용한 후보제거 모듈로 구성되며, 자모 후보 추출이 격자 탐색 도중 필요시에만 수행되는 대화식 구조를 갖는다 (그림 5). 격자 탐색은 초성, 중성, 종성 순으로 자모 후보들을 추출하며, best-first search 를 이용한다. 그리고 격자 탐색 도중 문자의 구조적 제약에 비추어 부적절한 부분 문자 후보들은 제거된다.

3.1. 동적 격자 탐색을 이용한 문자 후보 정합

문자 후보 생성 단계의 복잡도는 발현된 자모 후보 개수에 기하급수적으로 증가한다. 이는 문자 후보가 자모 후보들을 조합하여 생성되기 때문이다. 이때 주시할 점은 상향식 정합에 기반을 두는 경우, 문자 후보의 생성 이전에 모든 자모 후보가 발현되어야 하므로 불필요한 자모 후보 및 문자 후보가 생성될 수 있다는 사실이다. 먼저 동일한 위치에 같은 클래스의 자모 후보들이 다수 발현될 수 있다. 그림 2 에서 볼 수 있듯이 초성의 위치에 비슷한 형태의 'ㅅ' 후보들이 생성되었다. 다음

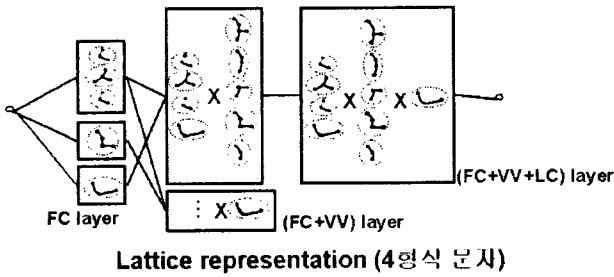


그림 6. 격자 구조를 통한 문자 후보 생성의 탐색 공간 표현

으로 자모 추출 시 자모의 위치를 고려하지 않으므로 적당하지 않은 위치에서 자모 후보들이 발견되고 있다. 즉, 초성 위치에서 중성, 종성 후보들이 발견될 수 있다. 이러한 문제들은 그림 1 과 같은 다양한 해석을 모두 고려해야 한다는 것에 기인한 것이다.

첫번째 문제를 해결하기 위하여 DP matching (dynamic programming matching) 기법을 사용한다. ROI(region of interest)를 단위로 하여 문자 후보 생성의 탐색 공간을 격자 구조로 표현하였으며(그림 6), 이 격자 구조 상에서 DP matching 및 best-first search 에 기반한 격자 탐색을 수행하여 최적의 문자 후보를 탐색한다. 여기서 ROI 는 부분 문자 후보의 bounding box 로 정의하였으며 격자 구조의 각 노드는 같은 ROI 를 갖는 부분 문자 후보들을, 각 링크는 부분 문자 후보가 생성되는 순서를 가리킨다. Best-first search 의 기준이 되는 부분 문자 후보의 점수는 2 절에서 설명한 문자 랜덤 그래프 모델과 부분 문자 후보와의 정합 확률값을 사용하였다. 따라서 중복된 문자 후보가 생성되는 것을 막을 수 있으며, 각 노드에 저장되는 부분 문자 후보의 개수를 상위 n 개로 제한하여 그 효율을 높일 수 있다.

두번째 문제를 해결하기 위하여 자모 유형에 따른 동적 격자 탐색을 제안한다. 한글의 자모는 유형에 따라 초성, 중성, 종성으로 구분된다. 즉, 격자 탐색 도중에 필요한 유형의 자모 후보를 예상되는 위치에서만 동적으로 추출함으로써 불필요한 위치에서 자모 후보가 생성되는 것을 억제하였다.

3.2. 모델 발화를 이용한 자모 후보 추출

Subgraph isomorphism 에 의한 자모 후보 추출 방법은 입력된 문자 그래프의 부분 그래프(subgraph)들과 자

모 모델들을 정합하여 자모 후보들을 추출한다. 이때 각 자모 모델들을 상호독립적으로 정합 함으로써 중복된 계산이 수행된다. 또한 자모의 위치, 즉 초성, 중성, 종성의 문자 전체에서의 위치를 고려하고 있지 않으며 따라서 무의미한 자모 후보들이 생성될 수 있다.

본 연구에서는 이러한 문제점을 해결하기 위하여 예상과 검증에 기반을 둔 자모후보 추출 기법을 제안한다. 특정 자모 모델의 존재 여부는 영상 내에서 그 자모를 구성하는 특징들의 존재 여부와 동일한 문제로 볼 수 있다. 즉, 자모를 구성하는 특징의 일부분이 영상에서 발견될 때 그 자모가 영상 내에 존재할 것이라는 예상이 가능하며, 그 자모를 구성하는 나머지 특징들이 발견되었을 때 실제 그 자모가 존재한다는 결론을 내릴 수 있다.

한글 자모의 경우에는 ‘ㄴ’과 ‘ㄷ’에서 볼 수 있는 것처럼 자모들 간에 특징을 공유하고 있다. 이 사실을 이용하여 특징들과 모델들 간의 관계를 그림 7 과 같이 계층적으로 표현할 수 있다. 특징의 종류는 원, 수직획, 수평획, 사선획, 역사선획의 다섯 유형으로 정의하였으며, 필기체에 나타나는 각 특징의 변형을 흡수할 수 있도록 퍼지(fuzzy) 하게 표현하였다. 그림에서 실선은 모델의 발화(model triggering), 다시 말하여 예상을 가리키며, 점선은 모델의 검증을 가리킨다. 예를 들어 영상에 수평획이 존재하는 경우 ‘ㄱ’ 모델이 발화되며, 그 주변에 수직획이 존재하는 경우 ‘ㄱ’ 자모후보가 검증, 추출된다. 레벨 2 에서는 레벨 1 에서 ‘ㄱ’ 자모후보가 검증 되었으므로 ‘ㅋ’, ‘ㄱ’ 모델들의 존재가 예상되며, 이러한 예상과 검증의 반복에 의하여 존재하는 모든 자모후보들이 추출된다. 계층적 모델 발화 구조의 각 노드는 발화 특징과 검증 특징, 그리고 그들 사이의 관계로 표현된다. 특징들 간의 관계는 한글에서 나타날 수 있는 17

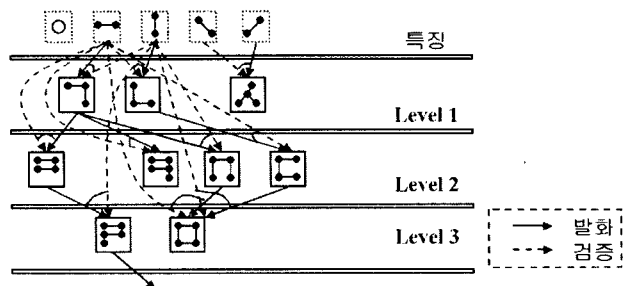


그림 7. 예상과 검증에 의한 모델 발화 및 검증

개의 관계를 정의하여 이를 이용하였다. 다음은 레벨 1 과 레벨 2 에서 노드의 발화 규칙의 예를 보여 준다.

```
MODEL          '7'
TRIGGERED_BY   '1'
VERIFIED_BY     '1'
RELATION       RIGHT_CLOCKWISE
```

```
MODEL          'ㄷ'
TRIGGERED_BY model 'T'
VERIFIED_BY     '1'
RELATION       DOWN_CNT_CLOCKWISE WITH 2nd PRIMITIVE
SUP_RELATION   PARALLEL WITH 1st PRIMITIVE
```

Subgraph isomorphism 에 기반한 자모후보 추출이 $O(M \times S \times C)$ 의 시간복잡도를 갖는 반면 (M: 자모 모델의 개수, S: 입력그래프의 부분그래프의 개수, C: 모델과 부분그래프의 정합에 요구되는 시간), 예상과 검증에 기반한 추출 방법은 특징 주도의 정합(feature-directed matching)을 통하여 M 과 S 의 값을 감소시키는 효과를 가져온다.

3.3. 문자의 구조적 제약을 이용한 후보 제거

상향식 정보처리에 기반한 한글 문자인식 방법은 이전 단계에서 추출된 자모 후보들을 조합하여 문자 후보들을 생성하고, 이 중 최적의 정합 확률을 갖는 후보를 최종적으로 선택한다. 이 과정 중 한글문자로서는 무의미한 부분 문자 후보들이 생성되며, 이로 인하여 시간 복잡도가 증가하게 된다.

본 연구에서는 부분 문자 후보의 생성시에 한글문자의 구조적 제약을 이용하여 한글문자로 부적합한 후보들을 제거한다. 한글문자의 구조적 제약은 세가지로 정의한다 (그림 8). 첫째, 후보의 내부에 후보에 해당하는 획들을 가로막는 획이 존재하지 말아야 한다 (topological constraint). 둘째, 후보의 영상 전체에서의 상대적 위치가 한글문자에 적합하여야 한다 (positional constraint). 마지막으로 각 후보들은 적은 양의 잡영만을 포함하여야 한다 (noise constraint). Topological constraint 는 영상을 위상기하학적으로 분석함으로써 쉽게 적용이 가능하다. 그리고 나머지 두 제약들은 부분 문자후보의 ROI 의 전체 영상에서의 상대적인 위치 및 ROI 내에 존재하는 잡영의 양을 이용하여 통계적으로 정의된다 (그림 9).



(1) Topological constraint (2) Positional constraint (3) Noise constraint
부분 문자후보 'ㄱ' 초성 후보 'ㄴ' 중성 후보 'ㅏ'

그림 8. 한글 문자의 구조적 제약

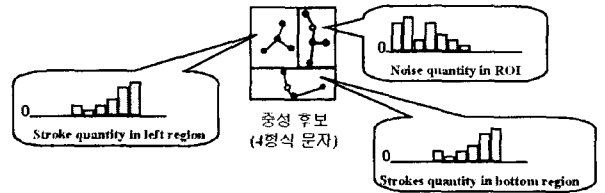


그림 9. 통계치를 이용한 문자의 구조적 제약의 정의

4. 실험 및 결론

520 자의 클래스로 이루어진 필기 한글 문자영상 DB 인 SERI-DB 를 이용하여 제안한 방법과 상향식 처리에 기반한 방법[8] 을 비교하였다. 랜덤 그래프 및 문자의 구조적 제약의 학습을 위해서 200 set, 104,000 자의 문자영상을 사용하였으며, 테스트를 위해서 1 set, 520 자의 문자영상을 사용하였다.

먼저, 제안한 방법과 상향식 처리에 기반한 방법의 효율성을 문자 영상 당 문자 정합 시에 생성된 부분 문자 후보의 평균 개수를 이용하여 비교하였다 (표 1). 이 수치는 문자 후보 탐색 및 정합 시에 생성되는 후보의 개수를 가리키며, 이 수치가 낮을수록 시스템의 성능이 좋다는 것을 뜻한다. 그 결과 제안한 방법이 상향식 처리보다 약 90%의 속도 향상을 보였다. 실제 수행 시간에 있어서도 제안한 방법이 상향식 처리에 기반한 방법에 비하여 2 배 이상의 속도를 보였다.

제안한 방법의 문자인식률은 상향식 처리에 기반한 방법에 비교하여 0.4 % 낮은 성능을 보이고 있다. 그림 10 과 11 은 각각 인식된 데이터와 오인식된 데이터의 예를 보여준다. 오류의 대부분은 자모 후보 추출과 정합 확률 계산 과정에서 발생하였다. 먼저, 계층적 모델 발화 구조에 정의되지 않은 형태의 'ㅁ'자가 모델 발화 오류의 대부분을 차지하였다. 또한 정합 확률 계산 시 이용되는 랜덤 그래프 모델들이 subgraph isomorphism 에 의한 자모 후보 추출에 바탕을 두고 학습되어 정합 확률 계산시에 오류가 발생하였다. 향후에는 모델 발화의 조율 및 제안한 자모 후보 추출에 기반을 둔 학습 방법

에 관한 연구가 필요하다. 이외에도 대용량의 데이터를 이용하여 제안한 방법의 성능을 검증할 필요성이 있다.

결론으로, 본 연구는 인식에서의 하향식 정보처리의 주된 역할을 인식 과정에 발생하는 중복된 연산의 제거라 보고, 모델 발화, 동적 격자 탐색 및 문자의 구조적 지식을 이용한 필기 한글 인식 방법을 제안하였다. 예비 실험 결과 제안한 방법은 기존 상향식 정보처리만을 이용한 방법에 비하여 문자 인식률의 저하 없이 높은 속도 향상을 가져 왔다.

	상향식 처리	제안한 방법
문자영상 당 생성된 부분 문자후보의 개수	6671.3 개	747.2 개
문자 인식률	91.2 %	90.8 %

표 1. 제안한 방법과 기존 방법[8]과의 성능비교

5. 참고 문헌

[1]. Jayanta Basak and Sankar K. Pal, "PsyCOP - A Psychologically Motivated Connectionist System for Object Perception," IEEE Trans. Neural Networks, Vol. 6, No. 6, pp. 1337-1354, 1995

[2]. Ethem Alpaydin, "Selective Attention for Handwritten Digit Recognition," Advances in Neural Information Processing Systems, Vol. 8, pp. 771-777, The MIT Press, 1996

[3]. Pierre Van De Laar et al., "Task-Dependent Learning of Attention," Neural Networks, Vol. 10, No. 6, pp. 981-992, 1997

[4]. Ruggero Milanese et al., "Integration of Bottom-Up and Top-Down Cues for Visual Attention Using Non-Linear Relaxation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1994

[5]. M. Cote, E. Lecolinet, M. Cheriet, and C. Y. Suen, "Automatic Reading of Cursive Scripts Using a Reading Model and Perceptual Concepts," International Journal of Document Analysis and Recognition, No. 1, 1998

[6]. C. D. Wickens, Engineering Psychology and Human Performance, 2nd Ed., Harper Collins Publishers, 1992

[7]. C. S. Park and S. Y. Bang, "Modeling Character Superiority Effect in Korean Characters by Using IAM, Biologically Motivated Computer Vision (Lecture Notes in Computer

Science 1811), Springer, May 2000

[8]. H. Y. Kim and J. Kim, "Hierarchical Random Graph Representation of Handwritten Characters and its Application to Hangul Recognition," Pattern Recognition, Vol. 34, No. 2, 2000



그림 10. 올바르게 인식된 문자의 예



그림 11. 오인식된 문자의 예