

다중 활자체 한글 문자 인식을 위한 유형 분류

김민기, 권영빈
중앙대학교 컴퓨터공학과

The Type Clustering for the Multi-Font Hangul Character Recognition

Min-Ki Kim and Young-Bin Kwon
Dept. of Computer Science and Engineering, Chungang University

요 약

본 논문에서는 글꼴의 변화와 잡영을 흡수할 수 있도록 자소의 탐색 영역을 정의하였으며 이 영역에 나타나는 횡모음과 종모음의 주획을 추출하는 방법을 기술하였다. 종모음 영역에서 추출한 수직획들과 횡모음 영역에서 추출한 수평획들을 각각 종모음과 횡모음의 주획이 될 수 있는 후보들로서 이들로 부터 종모음과 횡모음의 존재를 파악하는 것이 한글 유형 분류의 주된 내용이다. 그러나 다양한 글꼴에 나타나는 수평획들로부터 곧바로 횡모음의 존재를 파악하는 것은 쉬운 문제가 아니다. 본 논문에서는 기존의 트리 분류기를 확장하여 복잡하고 다양한 특징을 단계별로 단순화시키고 트리 분류기의 상위 노드에서 결정된 정보와 제약 조건을 이용하여 유형을 분류하는 방법을 제안하였다. 제안된 방법은 한글 상위 빈도 1405 자, 3 가지 글꼴에 대하여 99.8 %의 유형 분류율을 보이고 있다.

1. 서 론

컴퓨터는 종이에 나타난 문자를 곧바로 처리할 수 없기 때문에 키보드에 의한 수작업이 필수적이었으며 아직까지도 이러한 수작업은 크게 줄지 않았다. 일상 생활에서 사람들이 다루는 대부분의 정보는 종이위에 문자의 형태로 존재하기 때문에 일찍부터 광학 문자 인식에 대한 연구가 이루어 졌다[1]. 인쇄체 문자는 형태의 변형이 적어 필기체에 비해 상대적으로 인식이 용이하여 문자 인식 연구의 초기부터 관심의 대상이 되어 왔으며 현재 상용화된 제품도 나오고 있다. 그러나 인쇄술의 발달과 컴퓨터에 의한 문서 편집이 보편화되면서 다양한 크기와 글꼴이 출현하게 되어 인쇄체 문자를 완벽하게 인식하지 못하고 있다. 특히 한글은 영어나 일본어와는 달리 가능한 문자의 수가 방대하고, 한자에 비해 글자의 구별이 부분적인 획의 차이로 나타나기 때문에 각 문자 사이의 유사성이 커서 상대적으로 인식하기가 어려운 문자이다.

한글 인식에 대한 연구는 지난 25 여년간 꾸준한 연구가 이루어져 왔는데[2], 이를 인식 단위에 따라 크게 둘로 나누면 자소 단위 인식과 음절 단위 인식으로

구분할 수 있다. 자소 단위 인식[3, 4, 5, 6]은 정합할 부류의 수가 40 여자로 줄어 인식에 소요되는 시간을 크게 줄일 수 있으며 한글로 표현할 수 있는 모든 글자를 인식할 수 있는 장점이 있으나, 자소 분할에 따른 부담이 존재하며 여기서 파생되는 오류가 인식 성능에 큰 영향을 미치게 된다. 음절 단위 인식[7, 8, 9]은 자소 분할 과정이 불필요하나 인식할 부류의 수가 11,172 자에 이르기 때문에 이를 해결할 수 있는 방안이 모색되어야 한다. 음절 단위의 인식을 하는 대부분의 연구들은 상위빈도수에 의한 제한된 문자만을 인식 대상으로 하거나 완성형 2,350 자를 대상으로 인식을 수행하기 때문에 완전하게 한글을 인식한다고 볼 수 없다.

한글의 유형 분류는 한글을 음절 단위로 인식할 때는 대분류의 효과를 갖게 되며 자소 단위로 인식할 때는 자소 분할을 효과적으로 수행하기 위한 전단계의 역할을 하게 된다. 본 논문에서는 대분류 효과를 갖고 자소 분할에 도움을 줄 수 있도록 한글을 6 가지 유형으로 분류하는 방법을 제안한다. 논문의 구성은 다음과 같다. 2 절에서는 유형 분류에 관련된 기존의 연구들을 살펴보고, 3 절에서는 본 논문에서 제안하는 방법을 기술한다. 그리고 4 절에서는 제안한 방법에 대한 실험

및 결과를 보이고 5 절에서 결론을 맺도록 한다.

2. 기존의 유형 분류 방법

한글의 형태 분류에 관련된 기존의 연구들을 크게 구조적인 방법[4, 10, 11, 12]과 신경망을 이용한 방법[3, 6, 13]으로 나누어 볼 수 있다. 구조적인 방법은 횡모음과 종모음을 구성하는 수평선과 수직선을 기본적인 특징으로 이용하며, 신경망을 이용하는 방법은 문자 영역 전체 혹은 특정 영역에 나타나는 흑화소의 정보를 기본적인 특징으로 이용하고 있다.

구조적인 방법을 사용한 도정인[10]의 연구는 수평, 수직, 사선, 역사선의 4 방향에 대한 모든 런을 구한 후 문자 영상에 나타나는 극좌점과 극상점을 수평획과 수직획을 검출하기 위한 기본 위치로 설정하고 있다. 이 방법은 구조적인 방법으로 접근하는 여타의 연구에 비해 매우 높은 분류율을 보이고 있으나 잡영이나 글꼴에 민감한 극점을 기본 특징으로 이용하고 있기 때문에 실제적인 응용이나 확장성에 문제가 발생할 수 있다. 이근수[11]는 유형 분류를 위한 특징으로 최대 길이 투영(MRLP: Maximum Run Length Projection) 방법을 제안하였다. 최대 길이 투영법은 각 행 또는 각 열에 존재하는 연속된 런(run)들 중에서 최대 길이를 갖는 런만을 투영하여 히스토그램을 생성한다. 이렇게 형성된 수평, 수직 히스토그램에 나타난 정점(peak)의 크기와 위치 등 11 개의 특징을 정의하고, 트리 분류기와 퍼지 추론을 이용하여 형태 분류를 수행하였다.

신경망을 이용한 이동현[13]의 연구는 문자 패턴을 유형에 따라 17 종류의 부분 영역으로 나누어, 각 영역 내의 화소를 입력으로 하는 신경망을 통해 한글의 유형을 분류하는 방법을 제안하였다. 조성배와 이진수는 이동현이 제안한 신경망의 기본 구조를 이용하여 한글의 유형을 분류하고 인식하였다. 조성배[6]는 1,600 개의 입력 노드를 갖는 입력층과 17 개의 노드를 갖는 하나의 중간층, 그리고 6 개의 노드를 갖는 출력층으로 신경망을 구성하고 학습 자료에 변형을 첨가하는 방법으로 신경망의 일반화 능력을 향상시켰다.

3. 제안하는 유형 분류 방법

일반적으로 한글의 유형은 그림 1 과 같이 모음의 형태와 받침의 유무에 따라 6 가지로 나누고 있다. 본 논문에서는 각 유형에 대한 직관적인 이해를 돕기 위

하여 <가>유형, <각>유형, <고>유형, <곡>유형, <과>유형, <괘>유형으로 표기하였다. 그림 2는 한글 활자체에서 모음의 기본 줄기와 명칭[15]을 도시한 것이다.

	종모음	횡모음	종모음 횡모음
받침없음	가	고	과
받침존재	각	곡	괘

그림 1. 한글의 유형 정의

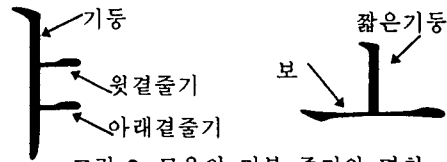


그림 2. 모음의 기본 줄기와 명칭

3.1 자소의 탐색 영역

합자해[16]의 부서법에 정의된 음절 구성 원칙에서 알 수 있듯이, 한글의 유형을 결정하는데 중요한 특징이 되는 모음의 상대적인 위치가 정해져 있고, 특히 인쇄체의 경우 다양한 글꼴에 따른 위치 변화가 크지 않기 때문에 여러 글꼴에 대한 글자의 통계적 특징으로부터 자음과 모음이 나타나는 영역을 정의할 수 있다. 본 연구에서는 문자 영상을 인쇄체 글꼴을 디자인할 때 사용하고 있는 32×32 격자[15]를 가로와 세로로 두 배 확장한 64×64의 크기를 갖는 영상으로 정규화한 후 누적한 실험 결과를 토대로 그림 3 과 같은 자소의 영역을 정의하였다.

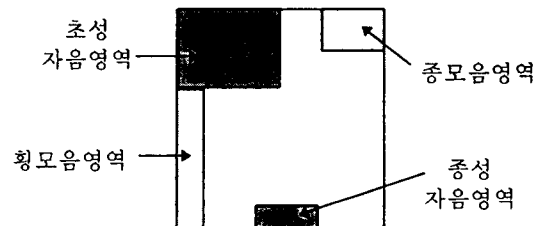


그림 3. 자소별 탐색 영역

자소 탐색 영역은 해당 자소를 포함하는 영역이 아니라 해당 자소의 일부분이 나타나는 영역을 정의한

것이다. 초성 자음 영역에는 초성의 일부가 반드시 나타나게 되며, 횡모음이나 종모음이 존재할 경우 해당 영역에 모음의 일부가 나타나게 된다. 그리고 종성이 존재할 경우 ‘ㄱ’과 ‘ㅋ’을 제외한 종성의 일부가 종성 자음 영역에 나타나게 된다. 따라서 우리는 문자 영상에서 찾고자 하는 자소를 해당 영역내에 나타나는 연결요소를 조사하여 보다 쉽게 추출할 수 있다.

3.2 수평획과 수직획 검출

모음의 형태와 위치는 한글의 유형을 결정짓는 중요한 특징이다. 그리고 모음은 자음에 비하여 전체적인 모양이 단순할 뿐 아니라 횡모음의 주획이 되는 보와 종모음의 주획이 되는 기둥은 다양한 글꼴에서 안정된 특징을 나타내고 있기 때문에 모음의 주획을 찾는 일은 매우 중요하다. 본 논문에서는 모음을 검출하기 위하여 횡모음과 종모음의 주획이 항상 포함되는 수평획과 수직획의 집합을 선정하고 트리 분류기에서 횡모음이나 종모음이 될 수 없는 후보들은 제거하는 방식으로 접근하고 있다. 따라서 트리 분류기의 전단계에서 수평획과 수직획 집합을 최소로 유지하면서 횡모음과 종모음의 주획을 반드시 포함되도록 하는 것이 효율적이다. 횡모음은 한글의 유형이나 글꼴에 따라 우측 상단으로 약간 기울어져 나타나기 때문에 단순히 일정 길이 이상의 수평선을 횡모음의 주획으로 판단하는 것은 효과적이지 못하며 종모음의 경우에도 단순히 수직선을 주획으로 판단할 경우 미세한 기울어짐이 발생할 경우 이를 흡수할 수 없다.

본 연구에서는 획의 최대 두께보다 큰 수평선과 수직선을 구하고 이웃한 런들을 병합하여 수평획과 수직획을 추출하였다. 이때 단순히 단일 임계값을 사용할 경우 기울어진 획이나 획의 접촉이 나타날 때 원하지 않는 결과가 발생할 수 있기 때문에 런의 시작위치에 따라 다른 임계값을 부여하는 방법을 사용하였다. 이 방법은 기울어짐이나 잡영에 의한 수평획과 수직획의 변형을 효과적으로 흡수할 수 있다. 그러나 다중 임계치를 사용하더라도 획이 접촉된 모든 형태를 처리할 수는 없으므로 이를 분리하는 방법이 필요하다. 획을 분리하는 기본적인 정보는 획의 두께 변화와 분기가 발생하는지의 여부를 감지하여 병합된 획을 분할하게 된다.

수평획과 수직획이 추출되면 그림 3에 도시한 횡모음 영역과 종모음 영역에 나타나는 수평획과 수직획을 각각 횡모음과 종모음의 주획이 되는 후보로 선정한다. 그림 4는 정규화된 입력 문자 영상으로부터 수

평획과 수직획을 추출한 후 횡모음 영역과 종모음 영역에 나타난 후보 획들을 ‘O’로 표시하고 있다.

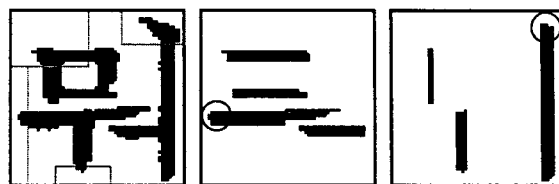


그림 4. 수평획과 수직획 추출

3.3 트리 분류기

트리 분류기는 복잡하고 다양한 특징을 단계별로 나누어 단순화 시키는 방식으로 상대적으로 안정된 특징을 먼저 이용하여 유형을 군집화하고, 정해진 군집내에서 다시 유형을 세분하게 된다.

종모음 영역은 횡모음 영역에 비하여 영역의 크기가 작기 때문에 종모음 영역에 나타나는 수직획은 매우 제한적이고, 종성에 나타나는 여러 특징에 비하여 수직획의 특징은 단순하다. 또한 수직획이 존재하지 않는 <고>유형과 <곡>유형은 문자 영상의 최소외접 사각형의 좌측 끝에서 우측 끝에 이르는 긴 수평획을 갖고 있으므로 수직획과 긴 수평획에 의하여 유형을 분류하는 것은 매우 안정적이다. 그러나 횡모음 영역은 상하로 길게 늘어져 있고 횡모음 영역에 다양한 형태의 수평획들이 나타나기 때문에 이들로 부터 횡모음을 추출하는 것은 간단한 문제가 아니다. 본 연구에서는 한글의 구조적 특징에 충실한 기존의 트리 분류기[10]를 확장하여 그림 5의 트리 분류기를 제안하였다.

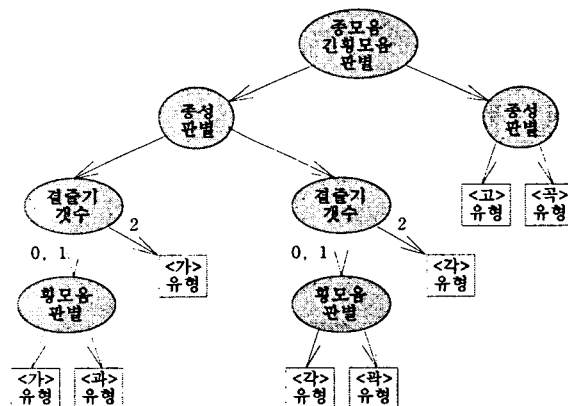


그림 5. 제안하는 트리 분류기

제안하는 트리 분류기는 횡모음의 판별을 트리의 최하위 노드에 둬므로써 횡모음이 나타날 수 있는 영역과 수직획에 나타나는 결줄기의 갯수에 따른 제약 조건을 이용하여 보다 정확하게 횡모음의 유무를 판별할 수 있다.

중모음과 긴 횡모음의 판별: 한글의 중모음을 찾는 문제는 중모음 영역에서 추출된 수직획의 집합으로부터 중모음의 특성에 부합되는 획을 찾아내는 것으로 이것은 중모음의 후보가 되는 수직획들 중에서 중모음이 아닌 것을 제거하는 문제로 생각할 수 있다. 중모음 영역은 우측 상단에 위치하기 때문에 추출된 수직획이 중모음이 아닌 경우는 모두 초성 자음에서 발생하게 된다. ‘ㄱ’, ‘ㄴ’, ‘ㄷ’, ‘ㄹ’, ‘ㅁ’, ‘ㅂ’, ‘ㅅ’, ‘ㅇ’ 등은 우측에 수직획이 존재하므로 이것이 중모음 영역에 나타날 경우 이것을 제거해야 한다. ‘ㄱ’, ‘ㄴ’, ‘ㄷ’, ‘ㄹ’은 상단에 나타나는 수평획 정보를 이용하여 제거할 수 있으며, ‘ㅁ’과 ‘ㅂ’은 좌측에 발생하는 수직획과 하단의 수평획 정보를 활용하여 제거할 수 있다. 물론 긴 횡모음이 존재할 경우에 수직획은 존재할 수 없으므로 긴 횡모음의 존재한다는 특징이 명확하다면 중모음의 후보들을 모두 제거할 수 있다.

긴 횡모음의 검출은 횡모음의 위치에 따라 두가지로 나누어 볼 수 있다. 첫째는 중간부에 나타나는 것으로 수평획의 길이나 기울어진 정도 등을 이용하여 판별할 수 있다. 둘째는 하단부에 나타나는 것으로 이것은 수평획의 최대 y 좌표값과 길이만으로도 식별이 가능하다.

중성의 판별: 받침이 되는 중성의 판별은 그림 5의 트리 분류기에서 볼 수 있듯이 모음의 유형에 따라 두가지로 나누어 볼 수 있다. 중모음이 존재할 경우에는 해당 수직획의 최하단점이 중요한 특징을 나타낸다. 수직획의 최하단점이 정해진 임계치보다 작은 경우는 중성이 존재한다고 판단할 수 있으며 임계치보다 큰 경우는 수직획에서 파생되는 좌측 가지의 특성에 따라 중성의 존재 여부를 판단하게 된다. ‘각’, ‘감’, ‘갑’ 등과 같이 받침에 ‘ㄱ’, ‘ㄷ’, ‘ㄹ’이 나타나는 수직획이 중모음의 기둥과 접촉될 경우 중모음 영역에서 추출된 수직획이 길어지기 때문이다. 특히 ‘의’와 ‘익’은 매우 유사하기 때문에 본 연구에서는 중성에 나타나는 ‘ㄱ’의 수평획 기울기, 수평획과 수직획의 비 등을 이용하여 중성이 ‘ㄱ’인지를 판별하여 중성의 존재 여부를 결정하게 된다.

긴 횡모음이 하단에 나타날 경우에는 받침이 존재

할 수 없으며 중간부에 위치할 경우에 받침이 존재할 수 있다. 긴 횡모음이 중간부에 위치하고 받침이 없는 경우는 횡모음이 ‘ㄷ’이거나 ‘ㅌ’이므로 횡모음에서 나온 짧은 기둥의 특징을 추출하면 쉽게 받침의 유무를 판별할 수 있다.

중모음 결줄기의 갯수와 방향: 중모음 결줄기의 갯수와 방향은 횡모음의 조합 가능성과 횡모음의 형태를 제약하는 구문적 특징이 된다. 횡모음 ‘ㄷ’는 중모음 ‘ㅏ’와 조합될 수 없으며, 횡모음 ‘ㅓ’는 중모음 ‘ㅑ’와 조합될 수 없다. 그리고 횡모음은 중모음의 결줄기가 두개인 것과는 조합될 수 없다. 표 1은 횡모음과 중모음의 조합 규칙을 나타낸 것으로 ‘O’표한 것은 조합 가능함을 나타내며 ‘X’표는 조합이 불가능함을 나타낸다.

표 1. 횡모음과 중모음의 조합 규칙

중모음 \ 횡모음	결줄기 1 개	결줄기 2 개		결줄기 3 개
		우향	좌향	
중모음	1	ㅏ, ㅓ	ㅑ, ㅕ	ㅗ, ㅛ, ㅜ, ㅠ
ㅏ	O	O	X	X
ㅓ	O	X	O	X
ㅑ	O	X	X	X
ㅕ, ㅠ	X	X	X	X

표 1에서 보듯이 횡모음과 중모음의 가능한 조합은 5 가지 밖에 없으므로 중모음에 나타난 결줄기의 갯수와 방향은 횡모음의 형태를 제약하는 중요한 정보가 된다. 그러나 중모음의 수직획에서 발생하는 결줄기의 갯수를 정확하게 추출하는 것은 그리 간단한 문제가 아니다. 왜냐하면 초성이나 중성의 일부, 또는 횡모음의 수평획이 중모음의 수직획에 접촉되기 때문이다. 만약 중모음의 결줄기 갯수가 잘못 판단될 경우 뒤에 이어지는 수평획의 검출 작업에 치명적인 오류를 발생시킬 수 있으므로 본 논문에서는 염세적인 감소 전략(pessimistic minus strategy)으로 결줄기의 갯수를 추정하는 방법을 제안한다. 염세적인 감소 전략이란 결줄기의 갯수가 n 개로 추정되었을 때, 접촉 가능성이 존재할 경우 n-1 개로 판단하는 것을 말한다. 이렇게 하는 이유는 표 1에 나타난 바와 같이 결줄기의 갯수를 감소시키므로써 조합 가능한 수평획의 제약 조건을 완화시키기 위한 것이다.

횡모음의 판별: 횡모음은 종모음에 비하여 한글의 유형에 따라 나타나는 위치의 변이가 크고 초성이나 종성의 일부에 나타나는 수평획이 횡모음과 유사한 특징을 나타내기 때문에 횡모음의 유무를 정확하게 식별하는 것은 매우 어려운 문제이다. 특히 횡모음의 보나 수평획에서 나온 짧은 기둥이 초성이나 종성에 접촉될 경우 횡모음이 갖는 일반적인 특징에 변형이 발생하여 횡모음의 식별을 더욱 어렵게 하고 있다. 이러한 문제를 해결하기 위하여 본 연구에서는 횡모음의 판별을 트리 분류기의 최하위에 둠으로써 상위 노드에서 발생한 한글 유형의 군집화 정보를 이용하여 횡모음이 나타나는 영역을 최소화하고 종모음의 기둥에서 나온 결줄기의 방향과 갯수를 이용한 제약 조건을 이용하여 잡영이나 접촉에 의한 횡모음의 변형을 흡수할 수 있도록 하였다.

횡모음 보의 후보가 되는 수평획의 좌측에 수직획의 존재 유무를 결정하여 횡모음의 보가 될 수 있는 후보의 갯수를 줄이고 최종적으로 남은 수평획을 우측상단으로 추적하여 수평획의 길이를 추정한다. 수평획의 길이를 알고 있고 트리 분류기의 상단에서 종모음 결줄기의 방향과 갯수를 알고 있기 때문에 수평획의 중앙부에서 상단이나 하단에 짧은 기둥의 존재를 확인하여 횡모음을 결정하게 된다.

유형의 검증: 트리 분류기에서 종모음, 종성, 횡모음의 존재 여부가 파악되면 한글의 유형이 결정된다. 그러나 트리 분류기에서는 <가>유형과 <각>유형의 초성에 'ㅋ'이나 'ㅍ'이 나타날 경우 수평획이 횡모음의 보로 잘못 판단되는 것을 해결할 수 없다. 예를 들면 <가>유형인 '커', '파'의 수평획이 <과>유형인 '귀', '괘'에 나타나는 횡모음의 보로 판단되거나, <각>유형인 '컬', '팔', 의 수평획이 <곽>유형인 '귤', '괘'에 나타나는 횡모음의 보로 판단되는 경우이다.

본 연구에서는 이러한 문제를 해결하기 위하여 트리 분류기에서 <가>유형이나 <각>유형으로 판단된 경우 초성이 'ㅋ'이나 'ㅍ'인지를 검증하는 단계를 두었다. 'ㅋ'은 수평획과 사선이 이루는 각도가 주된 특징이 되며 'ㅍ'은 상하 두개의 수평획 사이에 존재하는 수직획의 x 좌표값과 접촉 형태를 분석하여 결정하게 된다.

4. 실험 및 결과 분석

본 논문에서 제안된 한글의 유형 분류 방법은 SUN sparc station 상의 Motif 환경하에서 C 언어로 구현

하였다. 실험에 사용된 문자 영상은 한글 상위빈도[17] 1405 자를 한글 워드프로세서로 편집하여 레이저 프린터로 출력한 후 이것을 400 dpi로 스캔하여 얻었다. 실험에 사용된 자료는 명조, 고딕, 신명조, 공작, 중고딕으로 일반적으로 널리 사용되고 있는 글꼴을 선택하였으며, 각 글꼴에 대하여 9, 10, 12의 크기를 갖는 15벌의 자료를 사용하였다. 인쇄 상태는 대부분 양호한 편이나 신명조의 가로획, 공작체의 세로획의 끊어짐이 발생하고, 글꼴의 크기가 9인 경우 획이 붙는 경우가 발생하고 있다.

표 2는 글꼴의 크기가 10인 고딕, 명조, 신명조 글꼴에 대한 실험 결과이다. 괄호안에 표시된 것은 '파'와 '과', '팔'과 '괘'를 구분하지 않았을때의 결과를 나타낸다.

표 2. 고딕, 명조, 신명조 글꼴에 대한 유형 분류율

	고딕	명조	신명조	전체
상위	100%	99.60%	99.80%	99.80%
500 자	(100%)	(99.80%)	(99.80%)	(99.87%)
상위	100%	99.78%	99.78%	99.85%
900 자	(100%)	(99.89%)	(99.89%)	(99.93%)
상위	100%	99.86%	99.57%	99.81%
1405 자	(100%)	(99.93%)	(99.86%)	(99.93%)

그림 6은 상위빈도 900 자에 대하여 글꼴의 크기 변화에 따른 유형 분류율을 나타내고 있다. 막대그래프 쌍의 우측에 나타난 것은 '파'와 '과', '팔'과 '괘'를 구분하지 않은 결과이다. 그림 7은 글꼴의 크기가 10인 다양한 형태의 글꼴에 대한 적응성을 실험한 결과로 제안된 방법이 확장성이 있음을 확인할 수 있다.

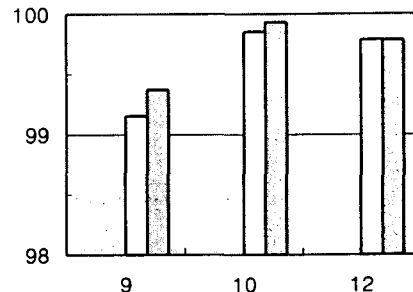


그림 6. 글꼴의 크기 변화에 따른 유형 분류율

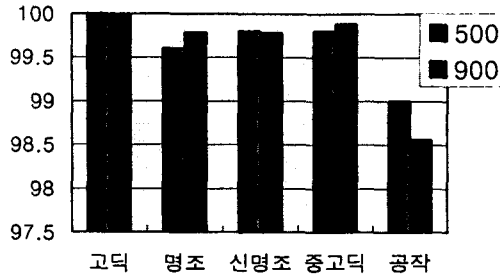


그림 7. 다양한 글꼴에 대한 유형 분류율

5. 결 론

수평획 정보만으로 획모음을 올바르게 검출하는 것은 매우 어려운 문제이다. 본 논문에서는 획모음과 종모음이 조합될 수 있는 구문적 정보를 이용하기 위하여 결줄기의 갯수를 추정하는 방법을 제안하였다. 이때 결줄기의 갯수를 잘못 추정하여 유형 분류의 오류로 이어지는 것을 방지하기 위해 염세적 감소 전략을 사용하였다.

이와 같이 획모음을 결정하는 노드를 트리 분류기의 최하위 노드에 둠으로써 상위 노드에서 발생하는 여러 제약 조건을 이용하여 획모음의 위치와 형태를 미리 추정할 수 있었으며 실험을 통해 제안된 방법이 효과적임을 알 수 있었다. 또한 ‘ㄱ’과 ‘ㅇ’의 수평획이 획모음의 보로 오인식 되는 경우를 검증 단계를 두어 해결할 수 있었다. 그러나 획의 두께가 얇거나 끊어지는 경우가 오분류의 주된 원인이 되기 때문에 획의 두께를 계산하여 획이 미세하거나 끊어짐이 발생할 수 있는 경우에도 안정적으로 획을 추출할 수 있도록 획 연결성 유지에 대한 연구가 필요하다.

참고문헌

[1] Shunji Mori, Ching Y. Suen, Kazuhiko Yamamoto, "Historical Review of OCR Research and Development", Proc. of IEEE, Vol 80, No. 7, 1992.
 [2] 이성환, 문자인식(이론과 실제), 홍릉과학출판사, 1994.
 [3] 이진수, 권오준, 방승양, "개선된 자소 인식 방법을 통한 고인식을 인쇄체 한글 인식", 한국정보과학회 논문지, 제 23 권, 제 8 호, pp. 841-851, 1996.

[4] 김민석, 손한웅, 최완수, 김수원, "자소 추출 방법을 이용한 고속 한글인식 시스템의 구현", 전자공학회 논문집, B 편, 제 6 호, pp. 418-424, 1992.
 [5] 함경수, 표창우, 원유현, "모음 우선 인식에 의한 한글 인쇄 문자의 분리와 인식", 한국정보과학회 추계 학술발표논문집, 제 18 권, 제 2 호, pp. 251-254, 1991.
 [6] 조성배, 김진형, "인쇄체 한글문자의 인식을 위한 계층적 신경망", 한국정보과학회 논문지, 제 17 권, 제 3 호, pp. 306-316, 1990.
 [7] 김상우, 전운호, 최종호, "복합신경회로망을 이용한 인쇄체 한글 문자의 인식", 전자공학회 논문지, 제 27 권, 제 4 호, pp. 520-526, 1990.
 [8] 최원호, 최동혁, 이병래, 박규태, "한글 인식을 위한 신경망 분류기의 응용", 전자공학회 논문지, 제 27 권, 제 8 호, pp. 1239-1249, 1990.
 [9] 이승호, 조창제, 김일영, 오세창, 조성배, 김진형, "한글 문서 인식 시스템 SILNOON", 한글 및 한국어정보처리 학술발표논문집, pp. 132-136, 1989.
 [10] 도정인, "인쇄체 한글 문자의 인식을 위한 자소 분리에 관한 연구", 한국정보과학회 추계 학술발표논문집, 제 17 권, 제 2 호, pp. 175-178, 1990.
 [11] 이근수, 최형일, "퍼지 추론을 이용한 한글 문자 유형 분류", 제 5 회 영상처리 및 이해에 관한 워크샵 발표논문집, pp. 164-172, 1993.
 [12] 이광호, "다중 활자체 한글 인식을 위한 자모의 분리", 한국과학기술원 석사학위 논문, 1988.
 [13] 이동현, 조원규, 양현승, 김진형, "신경망 모델을 사용한 한글 문자의 형태 분류", 한국정보과학회 춘계 학술발표논문집, 제 16 권, 제 1 호, pp. 215-218, 1989.
 [14] 김우성, 함영국, 박래홍, "Hidden Markov Model 을 이용한 인쇄체 한글의 유형분류 및 인식", 제 5 회 영상처리 및 이해에 관한 워크샵 발표논문집, pp. 178-184, 1993.
 [15] 김진평, 한글의 글자표현, 미진사, 1995.
 [16] 김석환, 한글문건(수정판), 한맥, 1995.
 [17] 이강섭, 이기성, "한글 출력코드 체계에 관한 기초 연구", 제 4 회 한글 및 한국어정보처리 학술발표논문집, pp. 285-291, 1992.