

# 모음 우선 인식에 의한 줄단위 필기체 한글의 인식

함 경수  
동국대학교 전자계산학과

## Recognition of Handprinted Hangul Line using Vowel Pre-Recognition Method

*Kyungsoo Ham*  
Computer Science Dept., Dongguk University

### 요 약

본 논문에서는 글자 구분선 없이 자유로이 쓰여진 필기체 한글의 인식 방안을 보인다. 줄단위의 한글 입력 영상에서 글자의 골격선을 추출하는 새로운 방법과 골격선들 간의 접촉점과 끝점을 그래프의 노드로 표현하고, 획은 그래프의 가지로 표현하는 방안을 보인다. 한글의 글자 구성 원리는 모음을 중심으로 모아쓰므로, 그래프로 표현된 줄단위의 한글에서 모음의 시작위치 및 속성을 가지는 노드로부터 한글의 모음을 가장 먼저 유도하여 인식하고, 우측 글자 및 자소 끼리의 접촉을 분리하여 초성 자음 및 중성 자음을 인식하여, 좌에서 우의 방향으로 한 문자씩 인식해 나간다.

본 논문에서의 자유로이 필기된 한글의 인식 실험은 우리나라의 주소 50개를 서로 다른 25인이 필기한 영상 데이터를 사용하였고 한글 문자의 인식율은 89%이다.

### 1 서론

우리 인간들이 책, 잡지, 신문 등의 인쇄물을 읽거나 편지, 노트 등의 필기물을 읽는 동안, 우리의 눈과 두뇌는 문자를 인식하는 작업을 한다. 이러한 인간의 문자 해독 능력은 인간이 가진 기본적인 패턴 처리 능력과 태어나면서부터 장기간에 걸쳐서 이루어진 학습의 결과라고 할 수 있다[Sizov91]. 지나간 수십 년간 우리는 인간이 읽을 수 있는 문자 패턴을 기계가 읽을 수 있는 코드로 변환하기 위한 문자 인식 연구를 지속적으로 해 왔으나, 이 일은 보기보다 대단히 어려운 작업이어서, 사람도 정서체로 필기한 문자를 문맥적인 정보가 없이 해독할 때 약 4%의 오독률을 가진다고 알려져 있다[Suen91].

한글 인식 연구는 지난 20여년간 꾸준히 이루어져서, 인쇄체 문자의 오프라인 인식의 경우 몇몇 상용화된 인식 시스템이 발표되어 있는등 인쇄체 한글 인식의 연구가 좋은 성과를 보임에 따라 몇몇 연구자들은 필기체 한글을 인식하기 위한 연구를 시작하였다[김태균88, 류승필89, 함경수92, 이성환92]. 그러나 필기 문자는 인쇄체 문자에 비하여 문자의 패턴에 수많은 변형이 있고, 필기구의 다양성, 표준 데이터의 부재 등의 어려

움에 직면해 있다. 이러한 어려움을 피해 가는 방법으로 대부분의 필기 한글 연구에서는 입력 문자의 필기에 제약을 두는데, 예를 들면 필기 박스 안에 정서체로 필기하게 하거나, 인식하기 용이한 특정 문자체로 필기하게 하거나, 또는 자소간 분리하여 필기하게 하는 등의 제약을 가한다. 이러한 필기에 대한 제약은 필기한 글 인식기의 실용화를 가로막는 장애중의 하나이다.

본 논문은 한글 문서 인식 연구의 일환으로서, 모음 우선 인식 방법에 의하여 문자 구획선이 없이 정서체로 자유로이 필기된 한글의 줄단위 오프라인(off-line) 인식에 관한 연구로서, 한글의 문자의 유형이 모음에 의하여 정하여지는 점에 착안하여, 정서체로 자유로이 필기된 한글의 입력으로부터 문자를 분리하고, 분리된 한 문자 내에서 자소 분리를 이룬 다음, 분리된 각 자소를 초성자, 중성자의 순서로 인식한다. 한글 문자에는 수직 또는 수평 모음이 반드시 존재하고, 이들 모음의 상대적인 위치는 고정되어 있으며, 모음의 획은 수직획 또는 수평획으로 이루어져 있다. 이와 같은 모음을 이루는 획들의 특징을 이용하여, 입력된 영상으로부터 모음획을 찾아서 모음 구성 규칙에 따라 인식하고, 문자간의 접촉 분리와 문자 내의 자소 분리를 행하는 방안을 제안한다.

입력된 줄단위 문자 영상으로부터 모음을 찾아내기 위하여는 문자를 구성하는 획들의 특징 추출을 위한 전처리가 필수적이다. 본 연구에서는 입력 영상의 전처리로서 간단한 개념의 영상 골격화 방안을 제시한다. 특징추출로서는 문자의 세선화된 영상을 이용하여 문자 획을 추출하고, 문자획의 끝점(end point)과 문자 획간의 접촉점(contact point)을 찾아내어 그래프 패턴화 하는 방안을 제시한다.

본 논문에서는 전처리 과정에서 추출된 그래프 패턴에서 모음을 가장 먼저 인식하는 방법에 의하여 입력된 필기 문자의 문자 분리와 자소 분리를 행하는 알고리즘을 제안한다.

본 논문은 다음과 같이 구성되었다. 제2장에서는 한글의 구조적인 특성과 필기 한글 문자 인식에 쓰이는 특징들을 기술하였다. 제3장에서는 특징 추출을 위한 전처리 과정으로써, 스캐너로 입력된 영상의 표현 방안과 골격선 추출 방법을 기술하고, 추출된 골격선에서 문자의 특징을 그래프 패턴화 하는 방안을 논의하였다. 제4장은 그래프 패턴에서 모음 우선 인식에 의하여 한글 문자를 분리하고 인식하는 알고리즘을 기술하고, 제5장은 실험 및 실험 결과에 대한 논의, 끝으로 6장에서는 결론을 기술하였다.

## 2 필기체 한글 문자의 특징

### 2.1 한글의 구조적 특징

한글의 기본 자모는 자음자(子音字) 14자와 모음자(母音字) 10자이나, 현대 한글의 자모는 복자음을 포함한 자음 30자와 복모음을 포함한 모음 21자가 쓰인다.

한글의 자모를 이루는 것은 주로 직선획인 'ㅣ', 'ㅡ', '／', '＼', 'ㅇ' 등의 단순한 획들이다. 각 자모는 이들 5가지의 단순한 획에 다른 단순획이 추가되어서 복잡한 구조의 자모를 이룬다.

음절(Syllable)을 이루는 모든 한글 문자는 중성, 즉 모음자를 중심으로 모아 쓰여진다. 모음은 수직 모음, 수평 모음, 수평-수직 혼합 모음으로 분류되는데 쓰여진 모음자의 종류와 중성 자음자의 존재 유무에 따라 잘 알려져 있는 6가지 유형으로 구분되며, 여기에서 중요한 것은 각 자모가 한 문자 내에서 어느 위치에 놓여져야 하는지 미리 정해져 있다는 것이다.

한글의 모아쓰기 규칙은 구조 해석 상의 중요한 단서를 제공한다. 즉, 모음자는 한글 문자의 중심 또는 기둥이고, 수직 또는 수평 모음중 어느 것이 쓰였는지와 수평 또는 수직 모음이 복모음으로 첨가되었는가에 따라 한글의 6가지 유형중의 하나로 분류할 수 있다. 따라서 한글 인식에서 구조상의 특징이 명확한 모음을 먼저 구분하고 인식하여 문자의 유형을 구분할 수 있는 알고리즘의 고안이 요구된다.

한글의 자음의 경우도 초성자와 중성자로 쓰일 수

있는 자소의 종류가 제한되어 있으며, 초성자의 쓰여지는 위치가 문자의 상단 및 좌측 편에 있고 중성자의 위치는 문자의 하단으로 정하여져 있으므로 인식을 위한 탐색 공간을 줄일 수 있다.

### 2.2 필기체 한글의 특징

사람이 손으로 쓴 문자의 패턴은 그 가짓수가 필기하는 사람의 수와 같다고 말할 수 있다. 본 절에서는 이들 정서체로 써진 한글문자열에서 문자간접촉과 자소간접촉에 따른 모음의 특징들을 살펴보기로 한다.

우리 나라 사람들이 한글을 가로 쓰기로 필기할 문서에서 문자간 접촉을 어느 정도 유발시키려는지는 조사된 바가 없으나, 본 연구에서 사용한 필기 데이터의 경우는 다음 문자와의 접촉이 약 16%로 나타났다. 인쇄체 문서의 경우는 수직 투영의 결과로 접촉되는 유형이 6가지[이군하89]로 알려져 있는데 [그림 1]과 같고, 이와 비슷한 문자간 접촉 문제가 필기 한글 문자 열의 영상에서도 나타난다.

## 시회 돕는 이 아이 놓아 이름

(a) (b) (c) (d) (e) (f)

[그림 1] 한글 인쇄 문서에서의 접촉 유형 6가지

[그림 1]의 접촉 유형중 투영적 접촉인 (a)와 하나의 문자가 두개로 분리된 (c)의 경우는 문자 인식의 전처리 과정에서 개별 획들로 분리시키거나 인식 과정에서 합자 처리하는 방법 등으로 해결할 수 있다. 직접 접촉인 (b),(d),(e),(f)의 경우는 접촉 지점의 식별과 강제적인 분리가 필요하며, 이들은 모두 모음 획이 다음 문자의 구성 요소와 접촉하고 있다.

필기 문자열은 문자간 접촉의 유형도 다양해서 모든 경우의 문자간 접촉이 있다고 보아야 한다. 그러나 대부분의 한글 필기 문자의 문자간 접촉에서는 한 문자가 우측 다음 문자와 직접 접촉하는 부위는 모음 획이다. 그러므로 다음과 같은 두가지의 접촉 문자의 특징, 즉 수직 모음이 접촉된 경우와 수평 모음이 접촉된 경우로 나눌 수 있다.

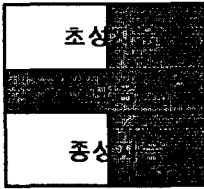
1) 수직 모음의 접촉 지점은 가장 우측 획의 우측 끝에서 이루어진다.

2) 수평 모음의 접촉 지점은 수평획의 우측 끝이다.

한글 문자 내에서 모음과 자음들의 상대적 위치는 [그림 2]와 같다. 보통 정서체의 한글 필체에서 초성 자음은 모음의 좌측 상단에 놓이고, 중성 자음은 모음의 아래쪽에 놓인다. 따라서 문자 내의 자소간의 접촉은 모두 모음과의 접촉이며 다음과 같다.

1) 초성자와 수직 또는 수평 모음과의 접촉

2) 종성자와 수직 또는 수평 모음과의 접촉



[그림 2] 한글 자모의 상대적 위치

제 3 장 전처리와 특징 추출

문자 인식 과정은 대량의 영상 데이터를 단계적으로 축소하여 하나의 문자 코드로 만드는 방안을 추구하는 것이다. 본 연구에서 제안하는 영상전처리 및 특징추출 방법은 개념적으로 간단하고 속도가 빠른 영상의 요약 표현 방법[함경수92], 골격선 선택, 획 추출 기법 및 문자 패턴의 특징 표현 방안[Ham93] 등이다.

3.1 영상의 연속 길이 표현

광학 스캐너로 입력한 문서 영상은 이치화(binary) 되어 영상 버퍼에 2차원으로 저장되며, 임의의 영상 라인은 0또는 1로 이어지는 백화소 연속 길이(nunlength)와 흑화소 연속 길이의 열이므로, 이들중 흑화소의 연속 길이만을 각각  $(c_1, r_1), (c_2, r_2), \dots, (c_n, r_n)$ 로 사상시켜 이들 각각을 연속 길이 표현이라 한다. 한 쌍의 연속 길이 표현에서  $c_n$ 는 연속 길이의 시작 위치이며,  $r_n$ 는 흑화소의 연속 길이이다.

연속 길이 표현들은 문서영상에서의 라인 별로 구하여져서 2차원 배열로 구성되며, 이를 영상 요약(Binary Image Descriptor) 테이블이라 하고 약어로 BID라 부르자. 이때 수평 투영에 의하여 얻어진 영상 요약 테이블(BID)을 R(Row major)BID라고 하며, 상하 방향(up to down)으로 좌에서 우(left to right)로 투영하여 얻어진 영상 요약 테이블을 C(Column major)BID라 한다. RBID와 CBID는 한 페이지의 필기 데이터의 입력 영상이 주어졌을 때 한글의 필기영상을 표현하게 된다.

3.2 영상의 골격화 및 획추출

문자 영상에서 획을 추출하기 위한 다양한 세션화 방안이 소개되어 있고, 가장자리 부식(edge erosion)기법이 가장 많이 쓰이나, 속도가 얼마나 빠른가와 세션화된 획들이 왜곡되지 않았는가가 문제가 된다.

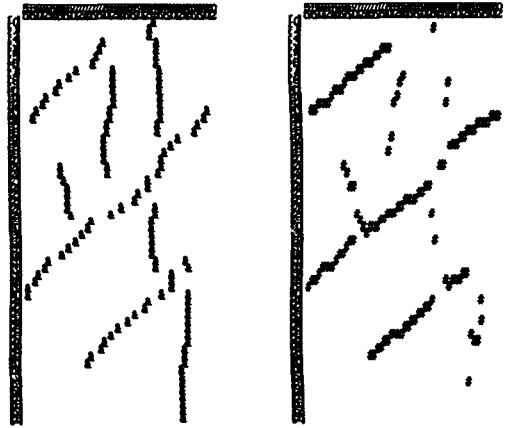
한글 문자를 이루는 획들은 직선 획들로 이루어지고 획의 방향도 대개 수평 방향이나 수직 방향이다. 세션화는 문자획의 대표선을 얻는것이 목적이다.

앞절에서 구해진 RBID와 CBID를 이용하여 획을

구성하는 연속 길이들에서 대표 화소의 위치를 취하여 연결하면 골격선이 된다. 대표 화소의 위치  $R_i$ 는 연속 길이 표현  $(c_i, r_i)$ 로부터 다음과 같이 유도될 수 있다.

$$R_i = c_i + r_i / 2$$

여기에서  $c_i$ 는 흑화소 연속 길이가 시작되는 열 위치를 나타내고,  $r_i$ 는 흑화소의 연속 길이이다. 대표 화소의 열 위치들은 골격선이 되는 점들의 위치를 표현하고, 이 점들을 연결하여 보면 수직 골격표의 점들은 수직 획의 골격을 대표하여 표현하며, 수평 골격표의 점들은 수평획의 골격을 대표하는 선이 된다. 이들의 위치를 같은 크기의 이진 영상 평면에 그려보면 [그림3]과 같다.



[그림3] '과'의 수직 및 수평골격선

3.3 획 추출

수직 골격선들과 수평 골격선들로 부터 각각 별개의 획 집합이 얻어지게 되는데, 수직 골격표로부터 얻어진 획들은 원 문자 영상에서 수직획들을 주로 표현하게 되고, 수평 골격표로부터 얻어진 획들은 문자 영상의 수평획을 주로 표현하며, 기울어진 획은 수직 골격이나 수평 골격이 거의 비슷한 기울기를 가지는 골격선을 표현하게 된다.

정확한 하나의 획 집합을 얻기 위하여는 이들 두가지의 획집합을 병합하여 대표획의 집합을 계산하여야 한다. 획의 표현은 획의 위치, 획의 방향, 획의 길이 등으로 표현한다. 획의 위치와 획의 길이는 획의 양 끝점에 대한 영상의 좌표로서 표현된다. 획의 방향은 인식 대상 문자의 난이도에 따라서 결정하여야 하는데, 한글을 구성하는 자모의 획은 'ㅇ'과 'ㅎ'을 구성하는 획 이외는 직선획으로 구성되며, 주로 수직 방향의 획과 수평 방향의 획들로 이루어지므로, 본 연구에서는

필기 한글의 획은 '丨', '一', '丿', '㇇', 'ㅇ' 등의 직선획 4가지와 'ㅇ'을 구성하는 획만을 가지게 하였다.

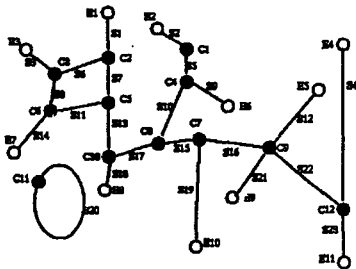
### 3.4 획의 그래프 표현

문자의 패턴을 문자획의 접촉점(contact point)을 중심으로 보면, 문자는 획의 접촉점과 획의 끝점(end point) 또는 접촉점과 접촉점 사이의 획의 연결로 구성된다고 볼 수 있고, 한글은 기본 자소의 패턴이 대부분 직선획으로 구성되어 있어서, 획 해석은 접촉점과 끝점, 접촉점과 접촉점 사이의 직선획으로 해석하게 하여 준다.

본 연구에서는 한글을 구성하는 획의 끝점과 접촉점을 그래프의 노드로 보고, 획은 노드와 노드를 연결하는 가지(branch)로 표현한다.

획과 획의 접촉점과 획의 끝점을 찾아서 그래프 패턴화 하는 일이 영상 전처리의 마지막 단계이다. 먼저, 모든 획의 양끝점을 찾아서 그래프 노드로 표시한다. 이어서 열과 행의 순서로 분류된 인접한 끝점들에 대하여, 각 획의 연장선을 계산하여 획들의 연장선이 교차하는 위치를 산출하고 RBID를 참조하여 흑화소 영역이면 이 위치를 접촉점으로 한다. 접촉점과 끝점은 확인 될 때마다 레이블로 일련 번호를 붙여 그래프의 노드로 하고, 노드들 간의 연결 획을 가지로 하여 그래프 표현으로 바꾼다.

이상과 같은 전처리 과정은 자유 필기 한글의 줄단 위 입력 영상을 전처리 하여, [그림4]와 같이 한글의 영상에 대한 획의 집합과 이들 획과 획간의 연결 관계인 그래프로 표현 할수 있고, 실제로는 그래프 매트릭스로 표현하였다.



[그림4] 문자열 '경주시'를 표현한 그래프

## 4. 문자 인식

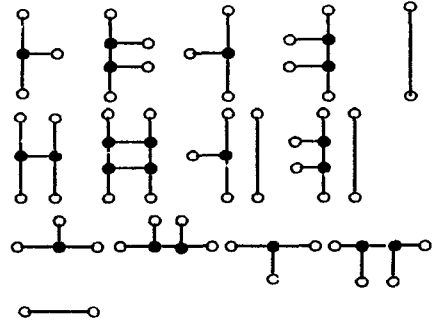
본 연구에서는 필기 칸이 없이 자유로이 필기된 한글의 정서체 한글을 인식한다. 본 장에서는 그래프 매트릭스로부터 한글의 모음 획을 찾아 인식하면서 문자 분리와 자소 분리를 하는 모음 우선 인식 방법과 자소 분리된 자음을 인식하는 과정이 소개된다.

## 4.1 모음의 분리 및 인식

### 1) 한글 모음의 패턴

그래프 표현으로부터 인식할 문자를 유도하려면 문자획의 기본 패턴이 정의되어야 하고, 본 논문에서는 모음을 우선적으로 인식하여 입력 라인으로부터 한 문자씩을 분리해내어 한글 문자를 인식하려 하므로 한글의 모음을 구성하는 획을 먼저 정의한다. 또한, 획의 해석 방법은 모음을 구성하는 획을 모두 직선획으로 해석하기 때문에 간단한 기본 패턴을 정의 하더라도 기본 패턴들의 조합을 유도하는 데는 어려움이 없다.

한글의 모음은 기본 모음자가 10자 이나, 수직 모음에는 수직획을 하나씩 더하여 생성된 복모음이 4자가 더 있으므로 이 14자소를 기초로 모음의 기본 패턴을 [그림5]와 같이 수직 모음 패턴 9가지와 수평 모음 패턴 5가지를 정한다. 수평 모음과 수직 모음의 복모음으로 '나', '내', '니', '니', '네', '니', '니' 등 7자모가 있으나 이들은 인식 과정에서 수평 모음과 수직 모음으로 각각 인식한후 합쳐하여 복모음으로 나타낸다. 기본 패턴을 간단히 하는 것은 패턴의 해석 알고리즘은 복잡해지더라도 보다 정교한 해석을 할 수 있기 때문이다.



[그림5] 수직, 수평모음의 기본 패턴

이들 모음의 기본 패턴은 모음 유도 과정에서 알고리즘으로 표현되고, 알고리즘이 수행됨에 따라 모음의 기본 패턴은 입력 영상에서 추출된 그래프 패턴과 비교되는 것이다.

### 2) 모음의 인식 및 분리

모음 우선인식 방법은 모음의 특징을 이용하여 그래프 매트릭스에서 모음을 먼저 인식해 내며, 한 문자의 인식 범위를 분리하고, 접촉 문자를 식별하며, 한 문자내의 자소 분리를 이룬다. 인식된 모음은 그래프 매트릭스에서 제거되어, 자음 인식은 간단한 자음획 추적을 통해서 쉽게 이룬다.

모음 인식에 실패하면 인식 범위의 문자는 부호 또는 영숫자, 혹은 영상 잡음으로 간주한다. 그래프 매트릭스는 문서 영상에서 획과 획의 끝점 및 접촉점과

의 연결 관계를 나타낸 것이므로, 모음 인식 과정은 그래프 패턴에서 수직 모음의 유도 및 인식, 수평 모음의 유도 및 인식의 순서로 이루어지며 동시에 접촉 식별 및 분리를 한다.

#### ■ 수직 모음의 유도 및 인식

그래프 패턴에서 가장 먼저 찾아야 할 획은 수직 모음의 시작획이다. 수직 모음의 시작은 문자 라인의 그래프 패턴에서 보면, 가장 좌측의 끝점 노드중 수직 획으로 연결된 노드로부터 시작된다. 따라서, 그래프 매트릭스에 표현된 끝점 노드들을 좌측 우선과 높이 순서로 배열한 후, 수직획으로 연결된 노드로부터 유도를 시작한다.

유도 과정중 접촉점을 만났을 때 접촉된 가지획을 점검하는 순서는 우측 가지획을 먼저 보고, 이어서 좌측 가지획을 점검하며, 이어서 연결획으로 간주되는 수직획 순서로 유도한다. 접촉점은 유도 과정중 후에 설명할 접촉 식별 과정에서 접촉점이 확인될 때 접촉을 강제로 끊어 내는 지점으로 이용된다. 유도 과정중 획들의 접촉점 이후에 나타나는 획중 같은 방향의 획은 이어지는 획으로 간주해야 한다.

선택된 노드로부터 수직 모음의 유도가 실패로 끝나면 다음 순서의 가장 좌측의 끝점 노드중 수직 획으로 연결된 노드를 선택하여 수직 모음 유도를 재시행한다. 이때 끝점 노드의 위치는 문자 시작 위치로부터 문자 폭이 문자 높이의 1.3배 이내의 위치에 있는 것을 대상으로 하고, 이 범위내에 수직획이 연결된 끝점 노드가 존재하지 않으면 수직모음이 없는 문자로 간주하고 수평모음의 인식이 이어진다.

#### ■ 수평 모음의 유도 및 인식

수평 모음을 인식하여야 하는 시점은 이미 첫 문자의 수직 모음에 대한 탐색이 이루어져서, 수직 모음의 존재가 알려져 있고 인식된후 이거나 수직 모음이 존재하지 않음을 확인한 후이다. 수평 모음의 유도에서도 가장 먼저 찾아야 할 획은 수평 모음의 시작 획이다. 수평 모음의 시작은 문자 라인의 그래프 패턴에서 보면 가장 좌측의 끝점 노드중 수평 획으로 연결된 노드로부터 시작된다. 따라서, 그래프 매트릭스에 표현된 끝점 노드들을 좌측 우선과 높이 순서로 배열한 후, 수평 획으로 연결된 노드로부터 유도를 시작한다.

수평획의 유도 과정 중에도 접촉점을 만났을 때 접촉된 가지획을 점검하는 순서는 위쪽 가지획을 먼저 보고, 이어서 아래 쪽 가지획을 점검하며, 이어서 연결획으로 간주되는 수평획의 순서로 유도한다. 접촉점은 유도 과정 중에서도 접촉 여부를 검사하는 시작점이 되고, 후에 설명할 접촉 식별 과정에서 접촉점이 확인될 때 접촉을 강제로 끊는 지점으로 이용된다. 여기에서도 수직획 유도 과정과 마찬가지로 유도 과정중 획들의 접촉점 이후에 나타나는 획중 같은 방향의 획은

이어지는 획으로 간주해야 한다.

#### ■ 접촉 식별 및 분리

한글의 인식에서 한 문자 내의 자소간 접촉이나 다른 문자와의 문자간 접촉을 탐지하고 접촉을 분리해 내는 문제는 대단히 어려운 문제이다.

[최필용93]에서는 미리 조사된 접촉의 양상을 표현한 패턴과 문자의 전처리에서 얻어진 국소 패턴과를 비교하여 매칭되는 것이 없는 경우 접촉을 분리 하였으나, 모든 경우의 접촉 양상을 수집 하였다고는 볼 수 없고, 필기 문자의 모든 자소의 접촉 양상을 수집 정리 한다는 것은 불가능에 가깝다. 본 연구에서는 문자간 접촉이나 문자 내의 자소간 접촉을 같은 접촉 문제로 보고, 모음획의 유도 과정 중에서 획간의 접촉점을 만나면 모음의 기본 패턴에 어긋나는 접촉된 가지획을 시작으로 자소 유도를 하여, 접촉된 지점을 탐지하고 접촉을 강제 분리한다.

모음의 유도 과정에서 나타나는 접촉의 양상은 대체적으로 세가지로 정리할 수 있다.

첫째, 모음과 초성 자음과의 접촉,

둘째, 모음과 중성 자음과의 접촉,

셋째, 모음과 우측의 다음 문자와의 접촉

등이다. 필기 문자에서는 이 이외의 여러 가지 접촉 양상이 나타날 수 있지만, 본 연구에서는 우선 정서체 필기 한글에서의 양상만을 고려하였다.

## 6. 실험결과

본장에서는 인식 실험 사용된 실제 필기 데이터에 대하여 먼저 살펴보고, 주소 인식 실험 결과에 대하여 인식 결과의 통계치와 오인식의 원인을 중심으로 살펴 본다.

인식실험을 위하여 우리 나라 각 지역의 주소를 여러 사람에게 필기케하여 실험 데이터로 사용하였는데, 이들 데이터는 필기 칸이 없이 자유로이 필기하는 방법으로 필기되었으며, 인치당 200화소의 해상도로 Umax300 스캐너에 의해 이진 영상(binary image)으로 입력되다. 이 데이터는 우편 번호부에서 무작위로 뽑아 낸 전국의 250개 동면 단위의 행정 구역명에 알맞는 상위 행정 구역명을 추가한 동면 단위까지 표시한 주소이다. 필기는 25사람이 행정 구역명 10개씩을 필기 하게 하였고, 필기를 위한 가이드 라인은 단지 횡선을 한줄 그어서 좌에서 우로 줄을 맞추어 정서체로 필기 하도록 하였으며 별다른 필기 제한을 두지 않았고 2257자의 필기 한글로 구성되었다. [그림6]은 입력 데이터의 예이다.

# 경북 경주시 석장동

# 서울 특별시 서초구

[그림6] 행정구역을 필기한 데이터

본 연구의 필기체 한글 인식기의 인식 속도는 대략 초당 25자이며, 문자 분리율은 92%이고, 문자 인식율은 86%이다. 인식기의 자소별 인식율은 89%이나 이를 한글 문자 단위의 인식율로 따지면 인식율이 하강하게 된다.

필기체 한글 인식기에서 모음 인식의 오류는 아주 적은 편이나, 대부분의 모음 인식 오류는 문자 분리의 과정에서 일어났다. 문자 분리의 오류는 전처리에서 획의 접촉 판정의 잘못이나 획의 종류를 구분하지 못하는데 기인한다. 이 부분의 오류는 필기 문자가 정자체가 아닌 경우 심각한 인식 오류를 야기시키므로, 획의 속성을 더욱 정밀히 표현할 수 있는 방안이 요구된다.

모음 인식에 실패한 예로서는 수직 모음의 시작획 부분에 장식선이 과도하게 쓰여져 'ㅣ'의 상단 끝점이 수평획으로 이어져서 수직 모음의 인식에 실패하는 경우가 발생 하였으며, '합'자의 경우 'ㅎ'의 상단 첫 획과 'ㅣ'의 시작획이 장식선을 씌워서 접촉되어 수직 모음의 인식에 실패하였다. 또한, '중구'와 같이 수평 모음이 문자간 접촉을 일으킨 경우 인데 필기체 문자를 분리 하는데는 문자 폭의 정보를 이용하기 어려우므로 수평 모음 끼리의 접촉을 분리하지 못하였다. 이와 같은 수평 모음끼리의 접촉은 초서체의 글씨를 쓰는 필기자들의 데이터에서 가끔 출현되는데 이들을 분리할 수 있는 별도의 알고리즘의 고안이 요구된다.

## 6. 결론

본 논문에서는 한글의 문자 패턴이 모음에 의하여 결정되는 점에 착안하여 자유 필기된 필기 한글의 입력 영상으로부터 문자를 분리하고 분리된 문자내에서 자소 분리를 이루어 각자소를 인식하였다. 한글 문자는 수직 또는 수평 모음이 반드시 존재하고 이들의 상대적인 위치는 고정되어 있으며, 모음의 획은 수직획 또는 수평획으로 시작하게 되어있다. 본 논문에서는 입력된 필기 한글 라인으로부터 모음획을 찾아서 모음 구성 규칙에 따라 인식하고, 문자간의 접촉 분리와 문자내의 자소 분리를 행하는 방안을 제안하였다.

본 연구의 한글 인식기는 모음을 우선적으로 인식

하여 문자 분리 및 자소 분리를 행하므로 모음의 인식 여부가 전체 문자 인식률에 영향을 미치는 단점을 가지는데 이를 개선할수 있는 전략이 필요하다.

필기 한글의 다양한 변형에 보다 잘 적응하기 위하여는 보다 세심한 전처리 기술과 접촉의 패턴을 2차원적으로 표현하는 기법이 앞으로 더 연구 되어야할 과제이다. 한글 필기 데이터를 확보하는 문제 또한 문자 인식 연구에 중요한 일이라 사료된다.

[Ham93]K.S.Ham, C.W.Pyo and Y.H.Won, "A vowel Pre-Recognition Method for Segmentation/Recognition on Freely Hand-printed Hangeul Lines", Proc. Infoscience'93, Oct 21-22, seoul, 1993, pp45-50

[Sizov91]Sizov K. A, "Recognition of Symbols and Words Written by Hand", proc. ICDAR91 volum 2, p. 692-700, 1991

[Suen91].C.Y.Suen,"Character & Handwriting Recognition: Expanding Frontiers",World Scientific series in CS vol.30, preface.

[Sur91]A. Sur, A.K. Datta,"Single-Pass Pragmatic Algorithm for Thinning for Syllabic Script",ICDAR91, Saint\_Malo, France, 1991.pp.559-567

[김태균88]김태균, T. Agui, M. Nakajima,"Stroke조합에 의한 필기체 한글의 표현과 인식", 대한전자공학회 논문지, 제25권 제1호, 1988년 1월,pp18-26

[류승필89]류승필,김태균,"속성 문법을 이용한 필기체 한글 문서내의 자모 인식", 대한 전자공학회 논문지, 제26권 제3호,1989년 3월,pp85-94

[이균하89]이균하 외,"한글 문서에서의 낱자 분리 알고리즘",한국인지과학회 1989년도 한글날 기념 학술대회 발표 논문집, 1989, pp203-208

[함경수92]함 경수, 임 재걸,"OCR 한글 필기 인식기 개발", 1차 연구 최종 보고서, 삼성전관 주식회사, 1992.6.30

[함경수93]함경수,표창우,원유헌,"필기 한글 주소 인식을 위한 오류 후처리 방법" 한국 정보과학회 '93봄 학술 발표논문집,제20권 1호, pp829-832,1993

[이성환92]이성환,박정선,"통계적 특징 추출 방법을 이용한 샘플체 필기 한글의 오프라인 인식," 제4회 한글 및 한국어 정보처리 학술 발표회 논문집, 서울, 1992년 10월, pp555-566