

Shape Pattern에 의한 필기체의 한글 인식 (Recognition of Handwritten-Hangeul by Shape Pattern)

朴 鍾 旭*, 李 柱 根**

(Jong Wook Park and Joo Keun Lee)

要 約

이 논문은 필기체의 한글 세이프 패턴을 자모패턴으로 분리하여 인식하는 방법을 제안한다. 본 방법은 세이프 패턴의 윤곽선에 대한 특징 버텍스를 추출하고, 특징 버텍스간의 접촉링크와 연결링크에 의한 위상구조로 표현하고, 이 위상구조로 부터 한글을 44개의 자모패턴으로 분리한다. 그리고 자모패턴에 대한 종합속성과 국부적속성을 검출하고 이들 속성을 사전과 매칭함으로써 인식한다. 본 방법은 패턴의 선분을 직접대상으로 하지 않기 때문에 필기체 패턴의 변형에 유연하고 효과적이다. 실험 결과 종래의 세선화 방법에 비하여 인식시간이 향상되고 인식율은 99%가 되었다.

Abstract

In this paper, a new methods which decomposes the handwritten-Hangeul shape patterns into subpatterns and recognizes the decomposed subpatterns are proposed.

The feature vertices are detected by searching boundary of the shape pattern and a topological structure is represented by a bridge links and contact links between the feature vertices. From the topological structure, Hangeul shape patterns are decomposed into the subpatterns of 44-Korean alphabet. The global and the local attributes are extracted from the subpatterns and the subpatterns are recognized by matching those attributes with the dictionary.

It is assured that this method is more effect and reasonable for deformed handwritten Hangeul shape patterns. Experimental results show that recognition rate is 99(%) and recognition time is also reduced as those using the thinning process.

I. 序 論

한글패턴에 대한 인식에서 인쇄체의 경우는 많은 연구가 발표되었다.⁽¹⁾⁻⁽⁶⁾

그러나 필기체의 한글 인식에 대해서는 극히 저조한 실정이다. 그런데 필기체의 한글 인식에 있어서 가장 문제가 되는 점은 1) 한글은 본질적으로 유사성이 심하고, 기본자모가 극히 단조롭기 때문에 약간의 변형에도 그 영향이 크다. 2) 대부분의 기본자모는 직선 구조이지만 필기체의 경우 필기자의 습성에 따라 다양한 곡선으로 표현되고, 자모의 연결부분이 매우 애매해진다. 3) 글자의 수가 방대하다. 따라서 인식에 있어서 대단히 어려운 문제에 부딪치게 된다.

일반적으로 패턴인식은 전처리 과정에서 세선화 방

*正會員, 圓光大學校 電子工學科
(Dept. of Electronics Eng., Won Kwang Univ.)

**正會員, 仁荷大學校 電子工學科
(Dept. of Electronic Eng., Inha Univ.)

接受日字: 1984年 10月 25日

법을 채용하고 있으나, 세선화 처리는 패턴의 변형을 가져오고 또 처리속도가 떨어진다.¹⁾

이에 대한 대안으로 윤곽선에 의한 세이프패턴의 인식 방법이 발표되었다.^{2)~4)} 세이프패턴의 인식 방법에는 프리미티브(primitive)의 속성에 의한 인식 방법^{2)~4)}과 프리미티브간의 관계로서 인식하는 방법⁵⁾ 등이 있다. 전자는 선분 사이의 관계가 주어지지 않고, 후자는 기하학적 표현은 우수하나, 프리미티브로서의 단순부분(simple part)을 추출하기 어렵고, 또 단순부분과 인트루전(intrusion) 사이의 관계를 구하기가 매우 어렵다.

본 논문에서는 세선화 처리를 하지 않고, 필기체의 한글에 대한 세이프패턴을 44개의 부분패턴으로 분리하여 인식하는 새로운 방법을 제안한다.

본 방법은 저자들이 앞서 발표한 도형패턴의 벡터스 추출 방법⁷⁾을 문자패턴에 도입하고, 필기체의 한글 세이프패턴을 1차원의 그래프로 변환한다. 그리고 패턴의 변형에 대한 영향을 제거하기 위하여 그래프상에서 연결링크와 접촉링크를 추출하고, 한글의 6형식(form)¹¹⁾을 기준으로 하여 링크에 의한 위상구조(topological structure)를 식별하여 44개의 자모패턴으로 분리한다.

또한 이들 자모패턴의 종합속성은 홀(hole)의 수와 연결링크의 수로써 구성하고, 국부적속성은 연결링크에 인접된 링크의 관계와 이들 방향속성으로 한다. 이들 속성에 의한 부분패턴의 디스크립터(descriptor)를 유도하여 사전을 작성하고, 입력패턴으로부터 이들 디스크립터를 검출하여 사전과 매칭시킴으로써 인식하는 방법이다.

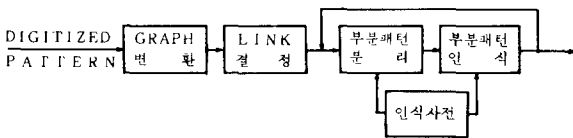


그림 1. 본 방법의 구성도

Fig. 1. Block diagram of this method.

본 방법에 대한 구성도는 그림 1과 같다. 그 결과 문자선분을 직접 인식 대상으로 하지 않고 종합속성과 국부적속성으로써 인식하므로 필기자의 관습에 의한 변형에 영향이 적고 또 인식 파라메타가 감소되었다. 또한 조합문자 패턴을 자모패턴의 부분영역으로 분리 인식하므로 한글의 본질적인 유사성이 제거되고 소수의 기본자모를 식별함으로써 다종류의 조합문자를 인식할 수 있었다. 그리고 세선화 처리를 하지

않으므로 패턴의 변형이 없고 처리속도가 빠르다.

5명이 쓴 500자의 한글을 표본으로 하여 실험한 결과 인식율은 99%로써 높은 인식율을 가져오는 성과를 확인하였다. 필기체가 이와 같이 높은 인식율을 가지는 이유는 변형의 요인이 되는 선분을 직접 인식대상으로 하지 않기 때문이다.

II. 세이프패턴의 링크 추출과 디스크립터

도형에 대한 종래의 벡터스 추출 방법은 픽셀(pixel) 하나 하나를 대상으로 하였으나, 저자들은 앞서 수개의 픽셀을 묶어서 하나의 라인세그먼트(line segment)로 규정하고, 또 몇개의 라인세그먼트를 연결 상태에 따라 하나의 호(arc)로써 결정한 후, 내부벡터스와 외부벡터스를 구분하는 방법을 발표한 바 있다.⁷⁾

종래의 문자인식 방법은 문자의 선분을 직접 인식대상으로 하기 때문에 다양하게 변형되는 필기체에서는 인식에 대한 일반성이 없다. 본 논문에서는 변형패턴의 선분을 직접 인식대상으로 하지 않는 방법을 모색하였다.

이를 위하여 논문⁷⁾의 호와 벡터스의 추출방법을 문자패턴에 도입하면 그래프 이론에서 그래프는 벡터스와 호의 집합으로 (1)식과 같이 표시된다.

$$G = \{V_i, A_j | i, j \in N\} \tag{1}$$

(단, V_i : 벡터스, A_j : 호, N : 자연수)

그런데 세이프패턴을 윤곽선으로 표현할 때 벡터스와 호의 관계는 1차원의 체인(chain) 그래프로 표현되므로 그림 2-(a), (b), (c)의 문자패턴은 그림 3-(a), (b), (c)와 같다.

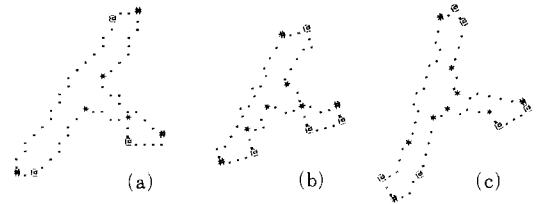


그림 2. 변형된 한글패턴의 예

Fig. 2. Example of deformed pattern.

그림 2에서 “@”는 외부벡터스, “*”는 내부벡터스, “#”는 최외곽점이며 외부벡터스가 된다. 여기서 그래프 G는 n개의 호가 체인으로 연결되므로 (2)식으로 표시할 수 있다.

$$G = C_n^1, A_1 = A, A, \dots, A_n \tag{2}$$

그러나 그래프를 식별하기 위해서는 벡터스간의 호에 대한 속성이 주어져야 한다. 따라서 호에 대한 속성의 집합은 (3)식으로 정의한다.

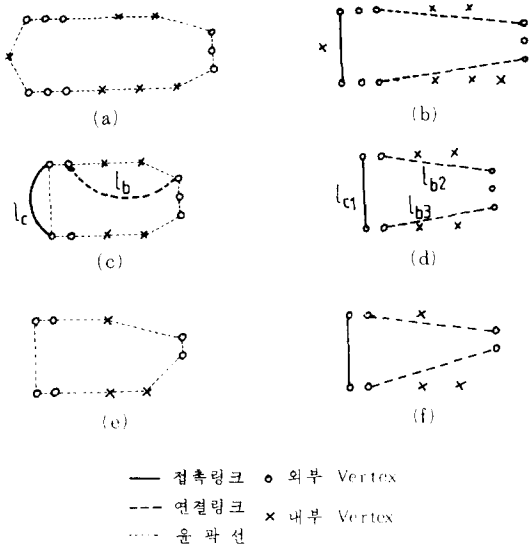


그림 3. 그림 2에 대한 graph 변환 및 link 결정 (a), (b), (c) graph 변환, (d), (e), (f) link 결정

Fig. 3. Graph converting and extracted link for Fig2. (a), (b), (c) graph converting, (d), (e), (f) extracted link.

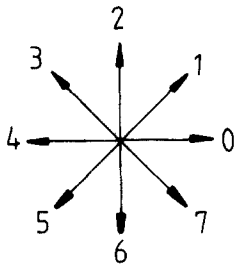


그림 4. 8방향 양자화 코드

Fig. 4. The codes of 8-quantized direction.

$$A_{hi} = \{A_r, A_x, A_y, A_v, A_c\} \quad (3)$$

여기서 A_{hi} 는 h번째 패턴의 i번째 호, h와 i는 부분 패턴의 레이블과 호의 레이블(label)이고, A_r 는 A의 길이로써 픽셀과 픽셀 사이의 x축, y축 간격을 단위로 하여 계산한다. $A_x, A_y \in \{1, 2, \dots, \dots\}$ 이 되는 x축, y축의 좌표이며, A_v 는 A의 시작점에 대한 버텍스의 종류로써 내부버텍스는 "4", 외부버텍스는 "5"의 값으로 정한다. A_c 는 그림 4의 8방향 코드이다.

그런데 필기체 문자패턴은 그림 2-(a), (b), (c)와 같이 다양하게 변형된다. 따라서 본 논문에서는 이들 변형에 대한 문자의 선분을 직접 인식대상으로 하

지 않고, 패턴의 외형적 구조를 이루는 연결링크와 접촉링크, 내부 홀(hole)에 주목한다.

그림 3-(b)에서 r개의 내부버텍스(x)를 포함하는 두개의 외부버텍스(o)를 연결할 때 하나의 루우프(loop)를 형성한다. 또 내부버텍스를 포함하지 않는 두개의 외부버텍스를 연결한 선분 l_c 는 윤곽선에 접촉된다. 따라서 내부버텍스를 포함하는 연결선분(l_b)을 연결링크 L_b 로 하고, 내부버텍스를 포함하지 않는 연결선분(l_c)을 접촉링크 L_c 로 한다.

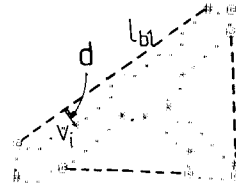


그림 5. 연결링크에 대한 거리

Fig. 5. Distance of bridge link.

그러나 그림 5에서 선분 l_b (점선)은 내부버텍스 V_i 를 포함하므로 연결링크가 되지만 접촉링크에 가깝다. 따라서 l_b 과 내부버텍스 V_i 간의 수직거리 d를 계산¹¹⁾하고, 기준치를 d_r 라 할 때 $d \leq d_r$ 이면 접촉링크로 판정한다. 또한 접촉링크 L_c 에 대하여는 선분의 길이 l , 선분길이의 기준치 l_r 라 할 때 $l < l_r$ 이면 무시한다. 이와 같은 방법으로 그림 3-(a), (b), (c)에 대한 그래프를 링크집합으로 표시하면, 각각 그림 3-(d), (e), (f)와 같이 된다.

이와 같이 패턴을 연결링크와 접촉링크에 의한 위상 구조로 표현하면, 그림 2-(a), (b), (c)와 같이 각각 다르게 변형된 패턴들을 그림 6과 같이 동일한 하나의 위상구조로 표현할 수 있다. 따라서 패턴의 변형과는 무관한 위상구조를 추출할 수 있다.

위에서 기술한 링크들의 속성을 (4)식과 같은 집합으로 정의한다.

$$L_{hj} = \{L_b, V_s, A_s, V_r, A_r, L_c\} \quad (4)$$

여기서 L_{hj} 는 h번째 부분패턴의 j번째 링크이며, h, j는 각각 부분패턴과 링크의 레이블, L_b 는 8방향 코드이며, V_s 와 V_r 는 링크의 시작버텍스와 끝버텍스의 레이블이며, A_s 는 시작되는 호의 방향속성, A_r 는 끝나는 호의 방향속성으로 호 A_{hi} 의 속성을 h_{xi} 의 매트릭스로 표현할 때 $A_s = A_c(h, V_s)$, $A_r = A_c(h, V_r - 1)$ 이 된다. L_c 는 L_b 와 L_c 를 구분하기 위한 속성으로써 L_c 는 "5", L_b 는 "4"의 값으로 한다. 또한 이들 속성값을 h_{xi} 의

마트릭스로 표현할 때 $L_D(h, j)$, $V_S(h, j)$, ...으로 각각 표시한다.

이들 속성에 대한 예는 그림 2-(a)의 패턴 "ㅅ"에 대한 링크 집합은 표 1.과 같다. 따라서 그림 6과 표 1에서 링크 l_{b_2} 는 $l_{b_2} = \{6, 2, 5, 5, 7, 4\}$ 로 표현된다.

표 1. 그림 2-(a)에 대한 링크집합
Table 1. Link set of the Fig. 2-(a).

< H = 1 >
< LINK LIST >

j	L_D	V_S	A_S	V_T	A_T	L_A
1	1	13	1	15	1	13
2	6	2	5	5	7	4
3	4	7	2	11	5	4

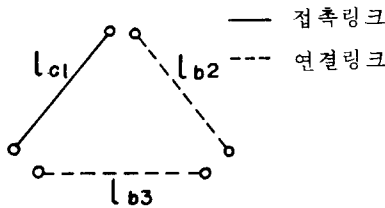


그림 6. 링크에 의한 위상구조.
Fig. 6. Topological structure by link set.

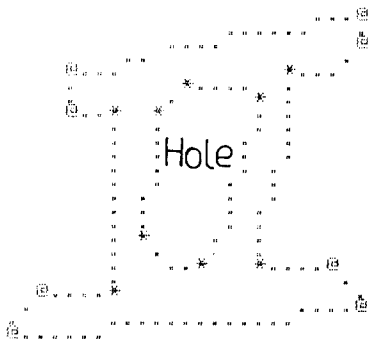


그림 7. 내부 홀을 갖는 자모패턴
Fig. 7. Subpattern with inner hole.

다음에 그림 7에서 내부 홀은 모두 내부버텍스로써 하나의 루우프를 형성한다. 그러므로 내부버텍스만으로 구성되는 부분패턴을 내부 홀로 정한다.

따라서 (4) 식으로 표현되는 링크와 내부 홀에 의한 위상구조를 가지는 조합문자 패턴의 디스크립터 $D(P)$ 는 (5) 식과 같다.

$$D(P) = \{H_s, n_s, S_p, L, O_H\} \quad (5)$$

여기서 H_s 는 한글패턴의 레이블, n_s 는 부분패턴의 수, S_p 는 부분패턴 집합, L 는 링크집합, O_H 는 내부 홀 집합이다. 예를 들면 그림 8-(a)의 패턴 "뿔"자의 경우 (6) 식과 같이 된다.

$$D(P) = \{\text{"뿔"}, 1, S_{p_1}, L\} \quad (6)$$

$$L = \{l_{11}, l_{12}, l_{13}, \dots\}, L \in S_{p_1}, O_H = \emptyset$$

다음에 한글패턴은 2개 이상의 기본자모가 상, 하, 좌, 우의 위치관계를 가지므로 한글의 6-형식⁽¹⁾을 기준으로 할 때 각 기본 자모의 위치관계로부터 6-형식에 대한 디스크립터를 $D(H)$ 라 하면 (7) 식과 같다.

$$D(H) = \{H_j, n_j, S_b, R_j\} \quad (7)$$

$$R_j = \{(S_{b1}, R_p, S_{b2}) \mid i, j \in N\}$$

$$R_p \in \{\gamma_u, \gamma_d, \gamma_r, \gamma_l\}$$

H_j : 6형식의 레이블

n_j : 자모패턴의 수

S_b : 자모패턴의 집합

R_j : 자모패턴간의 인접관계

R_p : 인접된 자모패턴간의 위치관계

$\gamma_u, \gamma_d, \gamma_r, \gamma_l$: S_{b1}, S_{b2} 가 각각 상, 하, 좌, 우에 위치할 때의 관계

예를 들면 그림 8의 "뿔"자의 경우는 (8) 식과 같이 표현된다.

$$D(H) = \{\text{"뿔"}, 4, \{c, \downarrow, \uparrow, r\}, \{R_j\}\} \quad (8)$$

$$R_j = \{(c, \gamma_u, \downarrow), (\downarrow, \gamma_d, c), (\downarrow, \gamma_l, \uparrow),$$

$$(l, \gamma_u, c), (\downarrow, \gamma_u, c)\}$$

III. 한글 세이프패턴 인식

1. 자모패턴의 분리

한글의 인식문제는 6개로, 형식화하고, 그 형식에 따라 자음과 모음을 각각 분리하면 한글 특유의 유사성이 제거되고, 방대한 수의 문자를 소수의 자모패턴으로 분리하여 식별할 수 있다는 것은 이미 발표되었다.⁽¹⁾⁻⁽⁴⁾

그러나 자모패턴으로 분리하는 문제는 쉬운 일이다.

본 장에서는 조합된 한글의 세이프패턴을 자모패턴의 부분영역으로 분리하여 인식하는 방법을 보인다.

세이프패턴을 자모패턴으로 분리하는 것은 그래프를 부분그래프(subgraph)로 분리하는 것으로 귀착된다.

따라서 세이프패턴을 자모패턴으로 분리하기 위하여 다음과 같은 규칙과 정의를 설정한다.

[규칙] (9-a) 식을 만족하는 벡터 V^{b1} , V_{b1} 를 연결하여 부분영역 S_{p_1} , S_{p_2} 를 분리할 때 (9-b), (9-c) 식을 만족하도록 분리한다.

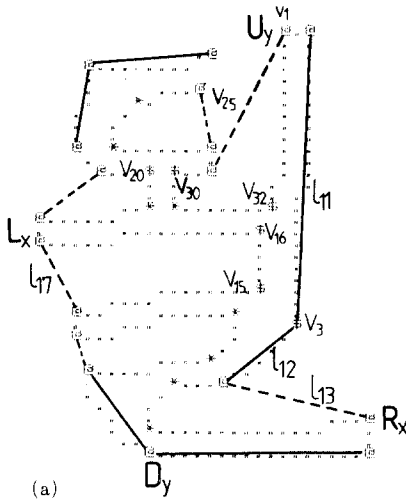
$$V_{b1}, V_{b2} \in Sp_1 \cap Sp_2 \quad (9-a)$$

$$(Sp_1 \cup Sp_2 = Sp) \wedge (Sp_1 \cap Sp_2 = \emptyset) \quad (9-b)$$

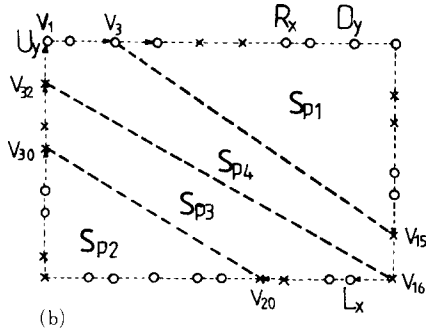
$$Sp_i, Sp_j \in \{\text{한글 자모패턴}\} \quad (9-c)$$

[정의] (9-a), (9-b), (9-c)식을 만족하는 버텍스 V_{b1}, V_{b2} 를 한쌍의 경계버텍스라 한다.

예를 들면 그림 8-(b)의 경우 V_3 에서 V_{18} 까지의 부분그래프는 받침 “ㄹ”을 형성하므로 한쌍의 경계버텍스 V_{b1}, V_{b2} 가 되어, V_3 와 V_{18} 를 연결하므로 받침 “ㄹ”과 “되”로 분리할 수 있다. 이와 같은 방법으로 Sp_2, Sp_3, Sp_4 를 점차로 분리하면 “ㄷ”, “ㄱ”, “ㅣ”를 각각 구할 수 있다.



(a)



(b)

그림 8. 한글 패턴에 대한 자모패턴 분리 (a) 한글 패턴 (b) 자모패턴 분리

Fig. 8. Decomposition hangeul pattern into sub-patterns. (a) Hangeul pattern. (b) Decomposition for hangeul pattern.

여기서 그림 8-(b)와 같은 그래프를 좀 더 효율적이고, 체계적으로 경계버텍스를 결정하기 위하여 그림

8-(a)와 같이 최외곽점 U_y, D_y, R_x, L_x 를 검출하여 $U_y-R_x-D_y-L_x$ 의 구간을 설정한다. 이들 최외곽점은 외부버텍스로서 결정되며, (3)식에서 속성 A_x, A_y 를 비교함으로써 추출할 수 있다.

그러므로 이들 구간에 대하여 (5), (7)식으로부터 링크들의 속성을 조사함으로써 자모패턴의 영역에 대한 경계버텍스를 추출하여 분리할 수 있다.

예: 그림 8-(a)의 경우 U_y-D_y 구간에서의 링크 l_{11} 을 조사할 때, 조건식 $(L_x(h, j)=5) \wedge ((A_y(h, j_1) + l_1) < D_y), j_1=V_T(h, j)$ 을 만족하므로 수직모음 하단에 받침이 연결되었음을 판정하고, 이때 경계버텍스는 V_3 와 D_y-U_y 구간에서 V_{18} 로 결정할 수 있다.

따라서 받침의 자모패턴을 분리하여 인식한 후, 수평모음과 초성자음이 수직모음에 연결되었는가를 판정하여 점차로 자모패턴을 각각 분리하여 인식할 수 있다. 이와 같은 방법으로 경계버텍스가 추출된 예로써 그림 8-(a)에 “\$”로 표시 하였다.

그러므로 이들 링크에 대한 속성과 인접된 링크들의 속성을 조사하여 자모패턴의 구조적특징을 추출함으로써 자모패턴간의 경계버텍스를 결정하고 분리한다.

그러나 그림 9(a), (b)의 “귀”(ㄱ, ㅍ, ㅌ)와 “커”(ㅋ, ㆁ)는 위상구조가 거의 동일하므로 오분리의 원인이 된다. 이것은 버텍스 X_1 과 X_2 에 대한 X축좌표의 위치를 비교함으로써 해결되지만 완전한 방법은 못된다. 그 밖에 “의”와 “익”에서 “익”의 모음 “ㅣ”에 받침 “ㄱ”의 수직부분이 직선으로 접속될 때는 분리하기 어렵다.

또한 그림 9-(c)와 같이 기본 자모 패턴이 연결되어 내부 홀을 포함하는 문자들은 (9)식의 조건에 의하여 V_{b1}, V_{b2} 와 내부 홀이 갖는 내부 버텍스간의 최소 거리를 계산하여 결정한다. 즉 그림 9-(c)에서 내부 홀 H_{01} 을 구성하는 내부 버텍스 $V_{n1}, V_{n2}, V_{n3}, V_{n4}$ 중에서 V_{b1}, V_{b2} 와 최소 거리가 되는 버텍스는 V_{n1} 과 V_{n2} 이고, 이때 경계버텍스는 $(V_{b1}, V_{n1}), (V_{b2}, V_{n2})$ 로써 2개의 쌍이 된다.

2. 자모패턴의 디스크립터

앞에서의 디스크립터는 조합문자를 자모패턴으로 분리하기 위한 것이다. 이 절에서는 자모패턴(44개)을 인식하기 위한 디스크립터를 유도한다.

필기체에서 선분은 다양하게 변하지만 홀의 수와 연결링크의 수는 본질적으로 변하지 않는다는 것은 그림 2와 그림 6에서 보였다. 따라서 자모패턴이 가지는 연결링크의 수와 윤곽선 내에서 포출되는 홀 수를 종합속성으로 한다.

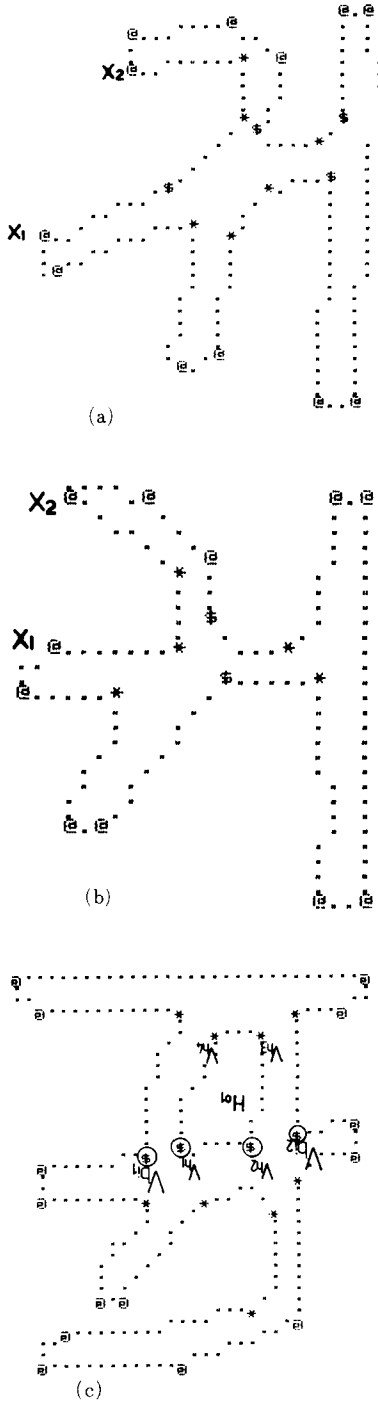


그림 9. 한글패턴에 대한 분리예 (a) “귀” (b) “커” (c) “표”

Fig. 9. Example of decomposition hangeul patterns into subpatterns. (a) “Gui”. (b) Ke. (c) Poy.

또한 이들 링크의 속성을 자모패턴의 국부적 속성으로 한다. 그러므로 이들 종합속성과 국부적속성으로써 디스크립터를 정의하면 (10) 식과 같다.

$$D(S_p) = \{n_n, n_b, L_b, L_c, R_j\} \quad (10)$$

$$R_i = \{L_b(h, j-1), l_{h_j}, L_b(h, j+1)\} \mid L_A(h, j) = 4 \mid R_i \in \{1, 2, 3\}$$

여기서 n_n : 자모패턴의 내부에 존재하는 홀 수, n_b : 연결링크 수, R_i 는 연결링크 l_{h_j} 에 인접된 두개의 링크간 인접관계로써

$$L_A(h, j-1) = L_A(h, j+1) = 5 \text{ 이면 } R_j = 1$$

$$L_A(h, j-1) = L_A(h, j+1) = 4 \text{ 이면 } R_j = 3,$$

$L_A(h, j-1) \neq L_A(h, j+1)$ 이면 $R_j = 2$ 로 정의한다. 예를 들면 그림 2의 “ㅅ”에 대한 디스크립터는 그림 6과 표 1에 의하여 (11) 식과 같이 된다.

$$D(S_p) = (0, 2, \{l_{b_2}, l_{b_3}\}, \{l_{c_1}\}, R_j) \quad (11)$$

$$R_j = \{R_1, R_2\} \mid R_1, R_2 \in 2\}, R_1 = (1, l_{b_2}, 4), R_2 = (6, l_{b_3}, 1)$$

3. 사전의 작성

44개의 자모패턴을 인식하기 위하여 표준 데이터로서의 디스크립터를 사전에 수록할 때 수록방법에 따라 효율적인 시스템이 될 수 있다.

그리고 디스크립터 $D(S_p)$ 에 대한 n_n 와 n_b 에 의하여 자모패턴의 성격별로 분류하고 R_j 에 따라 트리(tree) 구조를 갖도록 사전을 작성한다.

또한 오인식을 막고 인식시간을 절약하기 위하여 표 2와 같이 목록과 내용으로 구분하여 수록한다.

목록은 종합속성 n_n 와 n_b 로써 수색방향의 순번을 지정하여 지정된 순번에 의하여 사전을 수색토록 한다.

또 사전 내용은 기술함수 $D(S_p)$ 로부터 연결링크 l_{h_j} 에 대한 인접관계 R_j 를 수록한다. 그리고 매칭에 성공할 때와 실패할 때의 수색방향을 설정하여 사전을 자동으로 수색토록 하여 인식한다.

이상의 사전에 대한 데이터 베이스의 일부를 표 2에 보인다.

표 2에서 n_b 는 (10) 식의 연결링크 수, n_n 는 홀 수이다. 또 “NUM”은 순번이며, “RE”는 (10) 식에서의 관계 R_j 가 된다. 또한 “FD”와 “BD”는 $L_b(h, j-1)$ 과 $L_b(h, j+1)$ 의 값이다. 또한 “TY”는 연결링크의 내부 구조 형태이며, “SUC”와 “FAIL”은 성공필드와 실패필드를 각각 표시한다.

예를 들면 그림 2와 같은 자모패턴 “ㅅ”의 경우 (11) 식의 디스크립터로부터; $n_n=0$, $n_b=2$ 이고 링크 l_{b_1} , l_{b_2} 에 대한 $\{R_j, L_b(h, j-1), L_b(h, j+1)\}$ 는 $\{(2, 1, 4), (2, 6, 1)\}$ 이 된다. 따라서 목록에 의하여 25번째부터

표 2. 사전의 작성

Table 2. The compose of the dictionary.

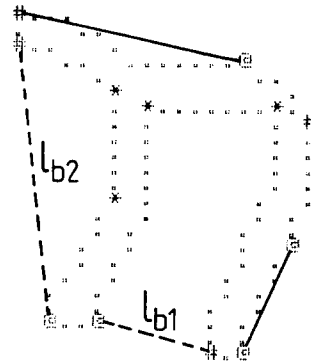
DICTIONARY CONTENT

$n_h \backslash n_b$	1	2	3	4	5	6
0	5	25	55	80	95	180
1	105	120	125	140	180	180
2	165	170	180	180	180	180

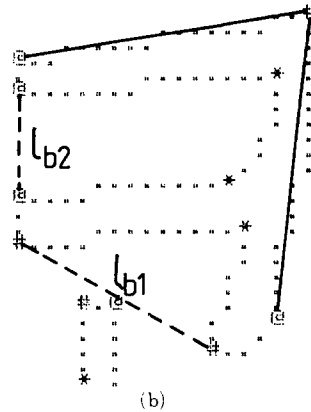
DICTIONARY DATA

NUM.	RE.	FD.	TY.	BD.	SUC.	FAIL.
25	1	4	1	7	16	1
26	1	4	2	0	16	1
27	1	0	2	4	5	1
28	1	2	2	0	6	1
29	1	1	2	0	5	1
30	1	3	2	0	4	1
31	1	3	2	1	3	7
32	1	6	3	4	48	1
33	1	5	3	4	48	16
34	1	6	2	4	26	1
35	1	7	2	4	26	1
36	1	5	2	4	26	1
37	1	6	3	4	47	1
38	2	1	1	4	3	1
39	2	0	2	6	25	1
40	2	0	2	7	25	3
41	2	6	1	1	29	1
42	2	7	1	1	29	16
43	2	4	1	0	42	1
44	2	3	2	0	22	1
45	2	2	2	0	22	1
46	2	2	1	0	42	1
47	2	3	1	0	42	200
48	0	0	0	0	0	0
49	0	0	0	0	0	0
50	0	0	0	0	0	0

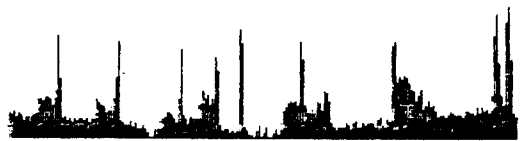
매칭하기 시작하여 실패하므로 "Fail"의 정보 "1"에 의하여 다음 순번을 순차로 수색한다. 이 때 31번째에서 실패하므로 정보 "7"에 의하여 38번째로 이동하여 수색할 때 (2, 1, 4)가 매칭된다. 따라서 "Suc"의 정보 "3"에 의하여 41번째에서 다음 데이터로써 (2, 6, 1)을 매칭할 때 성공되므로 "Suc"의 정보 "29"에 의하여 다음 "스"임을 판정한다.



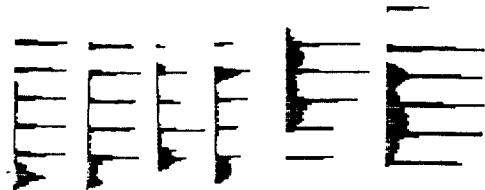
(a)



(b)



(c)



(d)

그림10. 부분패턴에 대한 위상구조의 유사성.

Fig. 10. Similarity of topological structure for discription function.

그러나 그림 10 의 "ㄱ"과 "ㄱ"은 동일한 위상구조를 가지므로, 이때는 연결링크의 내부구조형태를 부여한다. 즉 연결링크의 속성 $A_s(h, j)$ 와 $A_r(h, j)$ 와의 차 $[A_s(h, j) - A_r(h, j)]$ 의 값을 구하여 "TY" 필드와 매칭

하여 식별한다. 또한 “o”과 “口”은 $n_h = 1, n_o = 0, L_o = \phi, R_j = \phi$ 이므로 접촉링크 L_c 의 방향속성 $L_o(h_j, j)$ 를 조사하여 판정한다.

지금까지 고찰해 본 시스템의 흐름선도는 그림 11과 같다.

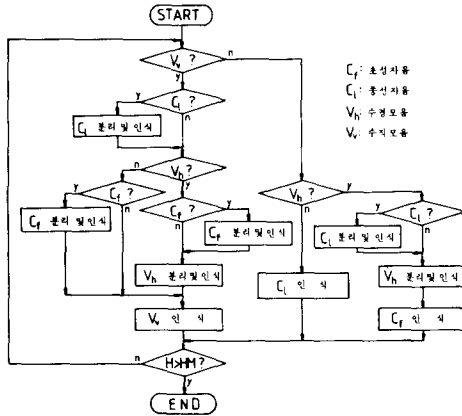


그림 11. 본 방법의 흐름선도
Fig. 11. Flowchart of this Method.

IV. 실험 결과 및 검토

한글을 6개 form으로 분류하여 10명의 학생이 쓴 1000자를 표본으로 하고 이들 표본을 중심으로 사전을 작성하였다. 그리고 정상적인 필기자로 인정되는 5명이 작성한 500자를 실험대상으로 하였다. [부록 2]

본 방법에 대하여 프로그램을 작성하여 실험한 결과는 표 3과 같다.

표 3. 실험 결과
Table 3. Recognition result of the hangeul pattern.

한글형식	인식대상문자수	인식율(%)
C V	100	100
C V C	100	99
C V	100	99
C V C	100	100
⌘ V	60	98.3
⌘ V c	40	97.5
계	500	99(%)

부록 실험에 사용한 표본 Data. 예
Appendix 2. Example of sample data being used in experiment.

기까나 더 다래며 버네서 기까나 더 다래며 버네서
씨에예적차따카타피씨 씨에예적차기타파피허
각강본달뒤만만밀남샷 각강본달당잔만밀남샷
깃알의권잠찾검인딱한 있표의령잠찾검인딱한
고꾸보도도뚜류로보보소 고꾸보도도뚜류로보보소
쓰(⊙)조크추코트프호포 쓰유조크추코트프호포
근근눈눈몸못본볼승웃 근근근눈눈몸못본볼승웃
본웃웃주춤본볼볼동몰 본웃웃주춤본볼볼동몰
켜과되뒤뒤쉬켜의의켜 췌과되뒤뒤쉬켜의의켜
취화악될길갈랜왔원된 취화악될길갈랜왔원된

실험 결과 패턴의 선분을 직접 식별대상으로 하지 않고, 홀수, 연결링크 수, 연결링크와의 인접된 링크와의 관계와 이들 방향속성을 인식 파라메타로 하여 99%의 높은 인식율을 가졌다. 또한 부분패턴의 상대적인 위치관계를 이용하므로 자음과 모음이 동일한 위상구조를 갖는 패턴에 대한 오인식이 없었다.

그리고 세선화 처리를 하지 않으므로 세선화에 의한 패턴의 변형이 없고 처리속도가 빠르기 때문에 실용성이 강하다.

여기서 인식되지 않은 것은 시각으로 판별하기 어려운 문자에 한한다. 즉 “o”과 “口”이 아주 유사한 경우(부록 “o”표)에 만 오인식되는 경우가 나타났고, “익”과 “의”등과 같은 형에서 “o, l, 7”인지 “o, 一, l”인지 시각으로 판별하기 어렵다.

V. 結 論

이상의 연구에 대한 성과는 다음 같이 요약된다.

1. 변형이 있는 부분패턴에 대하여 동일한 위상구조를 갖고, 또한 패턴의 선분을 직접 식별대상으로 하지 않고, 종합속성과 국부적속성을 이용하므로 인식율이 높다.

2. 본 방식의 최대 이점은 부분패턴에 대하여 연결 가능한 기본자음의 부분패턴을 분리함과 동시에 사전에 의한 자동인식하므로 잘못 분리된 패턴에 대하여 경계 버텍스를 다시 결정하여 분리하고 인식하므로 케환처리가 가능하다.

3. 세선화 처리를 하지 않으므로, 세선화에 의한 패턴의 변형이 없고 처리속도가 향상된다.

4. 일반 도형과 문자를 동일 선상에서 처리할 수 있으므로 문자, 도형의 전송에 유리하리라 생각된다.
5. 그러나 시스템이 다소 복잡하다는 단점은 있으나, 반도체 소자의 발달로 인하여 해결 가능하리라 본다.

参 考 文 献

- [1] J.K. Lee, "Recognition of printed korean characters"(I), *AEU*. vol. III, no. 3, 1970. Aug.
- [2] J.K. Lee, *Recognition of Printed Korean Characters (II)*. Proc. I.S.C, Sept. 1970.
- [3] J.K. Lee, *Korean Characters Recognition by Decomposition Method and Display by Variable Combination*. Ph.D. dissertation in Keio Univ. Japan, 1972.
- [4] T. Agui, Makajma, T.K. Kim & E.T. Takhashi, "A Method of recognition and representation of korean character by tree grammars", *IEEE Trans. P.A.M.I.*, vol. PAMI-1, no. 3, pp. 254-251, July, 1979.
- [5] 최병욱 외 2인, "한글인식에 있어서의 자소 추출," 대한전자공학회지, 제18권, 제2호, PP. 36~43, 1984. 4.
- [6] 이주근, 남궁 재찬, "한글 패턴에서 subpattern 분리와 인식에 관한 연구," 대한전자공학회지, 제18권, 3호, PP.1~8, 1981. 6.
- [7] 이주근, 박종욱, "가변 링크에 의한 도형패턴의 모서리 결정 방법," 대한전자공학회지, 제20권, 3호, PP.13~20, 1983. 5.
- [8] L.G. Shapiro, "A structure model of shape", *IEEE Trans. P.A.M.I.*, vol. PAMI-2, no. 2, pp. 111-126, Mar. 1980.
- [9] K. Yamamoto, "Recognition of hand-printed Hiragana characters by concave & covex feature & automatically merging the dictionary", *JIECE*, vol. J65-D, no. 6, pp. 774-781, June 1982.
- [10] L.G. Shapiro & R.M. Haralick, "Decomposition of two-dimensional shape by graph-theoretic clustering", *IEEE Trans. P.A.M.I.*, vol. PAMI-1, pp. 10-20, Jan 1979.
- [11] L.S. Davis, "Shape matching using relaxation techniques", *IEEE Trans. P.A.M.I.*, vol. PAMI-1, no.-1, pp. 60-72, 1979.