

印刷體 한글문자패턴의 變形文字패턴 發生

안 석 출*
부산개방대학 인쇄공학과

김 명 기
동아대학교 전자공학과

A Method of Distorted Character Pattern Generation for the Printed Hanguk Character Pattern

Suk Chul AHN
Busan Open University

Myung Ki KIM
Dong-A University

ABSTRACT

An effective and simple method for the distorted pattern generation based on pattern transform is presented to save time and work for the accumulation of hand-written Hanguk patterns.

Normalized patterns can be made as desired distorted patterns if an appropriate transformation is performed on each input pattern.

The transformation can be done by coordinate and elastic transforms.

Computer simulation results show that the desired distorted patterns are closely similar to hand-written Hanguk patterns.

I. 서론

문자의 변형과 종류가 비교적 적은 數字나 英字는 OCR(Optical Character Reader)로 여러분야에서 거의 실용화되고 있다.

그러나 한글의 필기체 문자는 글을 쓰는 사람의 습성이나 사용하는 도구등으로 문자의 변형이 심하고 문자의 종류가 많기 때문에 학습용 문자패턴의 데이터를 수집 하는데 많은 작업량과 시간이 소요되어 필기체 한글용 OCR 개발에 문제점이 많다.

그렇지만 최근에는 한글의 자동인식이 정보처리 자동화의 한 분야로서 필기체 한글의 자동인식이 요망되고 있다. 이러한 문제점들을 해결하기 위해서 데이터 수집의 보조 또는 대체수단으로 變形文字패턴 발생을 이용할 수 있다.

변형문자발생의 인공적 발생방법으로서는 筆順정보 또는 필기정보를 알고 세션화된 패턴에 습동을 주어서 발생시키는 방법과 필기체문자를 인식하기 위해서 표준패턴을 신속변환하여 변형된 필기체 문자패턴에 인공적으로 변형시킨 변형패턴을 정합시켜 필기체 문자를 인식하는 방법이 보고된 바 있다. (1), (2)

그러나 습동변환은 필순정보나 필기정보를 알고 세션화된 패턴이어야 하고, 신속 변환은 문자패턴의 방향성 Segment의 독립적인 회전변형을 흡수하지 못한다.

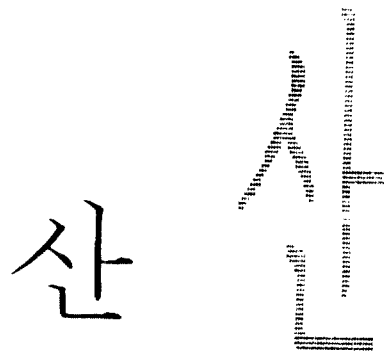
따라서 본 논문에서는 필자의 습성에 따라 변형이 심한 한글문자를 화상으로도 입력시켜 세션화처리를 하지 않고, 문자의 방향성 Segment의 독립적인 회전변형이 추가된 신속변환을 시켜 변형패턴을 발생시키는 한 방법을 제시하고 Compute Simulation을 통해서 제시된 변형패턴 발생방법의 타당성과 유용성을 확인했다.

2. 패턴변환

2-1 패턴의 선형변환

그림 (1-a)와 같은 인쇄체한글문자를 ITV카메라와 Image

Digitizer로 구성된 화상 입력장치를 통해서 문자화상-입력시킨 후 K×L 크기의 2值화상으로 변환된 화상 G를 문자패턴이라 한다.



(a) Printed Character (b) Dot Pattern
Fig 1. Printed Character and Dot Pattern

$$G = \{(X', Y', g(X', Y')) \mid g(X', Y') = 0, 1; X' = 1, 2, 3, \dots, K; Y' = 1, 2, 3, \dots, L\}$$

여기서 g(X', Y')는 좌표(X', Y')의 화소농도이고, K와 L은 X'와 Y'의 최대값이며 K×L은 문자패턴(이하 패턴이라 칭함)의 크기이다.

그리고 g(X', Y')=0이면 화상의 배경, g(X', Y')=1이면 문자 Stroke를 구성하는 화소이다. 화상G에 선형변환을 시켜 화상을 얻었다면 화상 F는 다음과 같다.

$$F = [A][G]$$

여기서 A는 2×2 변환 Matrix로서 다음과 같다.

$$[A] = \begin{bmatrix} A11 & A12 \\ A21 & A22 \end{bmatrix}$$

패턴 G에 좌표변환을 이용하여 패턴 F를 구하면 다음과 같이 된다.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = [A] \begin{bmatrix} X' \\ Y' \end{bmatrix}$$

식 (3)에서 A11과 A12는 X축 Y축의 비율을 나타내고 비율이 1이고 A12 = 0 일때 A21의 경우는 Sin⁻¹(A21)만큼 X축 방향의 Segment가 회전을 하고, A21 = 0 일때 Sin⁻¹(A12) 만큼 y축 방향의 Segment가 회전변형을 한다. 회전변형된 패턴을 Fno 라하고 그림(2)와 같다.

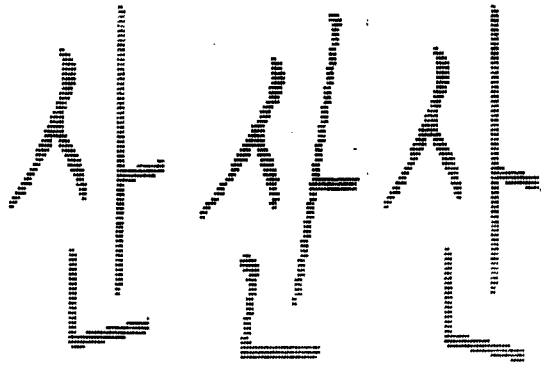


Fig 2. Linear Transform Pattern

2-2 신속변환

$K \times L$ 의 화상크기로 정규화된 2値패턴 F_{no} 는 식 (5) 와 같이 표시 할 수, 있다.

$$F_{no} = \{ (X, Y, f(X, Y)) \mid f(X, Y) = 0, 1; X = 1, 2, 3, \dots, K \\ Y = 1, 2, 3, \dots, L \} \quad (5)$$

G 를 신속변환한 화면을 F_m 으로 하고 다음식으로 표시한다.

$$F_m = \{ (X_k, Y_\ell, f(X_k, Y_\ell)) \mid 0 \leq f(X_k, Y_\ell) \leq 1 \\ ; k = 1, 2, 3, \dots, K; \ell = 1, 2, 3, \dots, L \} \quad (6)$$

여기서 X_1, X_2, \dots, X_k 와 Y_1, Y_2, \dots, Y_ℓ 은 샘플링점의 좌표이고 $X_1 = 1, Y_1 = 1, X_k = K, Y_\ell = L$ 이다.

식 (5) 의 샘플링점 (X_k, Y_ℓ) 에 있어서의 농도값 $f(X_k, Y_\ell)$ 은 2변수보간법에 의해 구한다. (5)

2-3 신속함수

신속함수는 식 (6) 의 샘플링점의 간격을 결정하는 함수이다. 따라서 필기체 문자의 변형에 대한 이론 및 통계적인 해석을 한 후, 거기에 적합한 여러가지의 신속함수를 적용해야만 한다. 그러나 현 단계에서는 필기체문자의 변형이론 및 통계적인 해석이 조사되지 않아서 가로 및 세로의 신속을 수행하는 5가지의 신속함수 $E_n(u)$ 를 예를 든다.

$$E1(u) = 1 \quad (7)$$

$$E2(u) = 2bu + 1 - b \quad (8)$$

$$E3(u) = -2bu + 1 + b \quad (9)$$

$$E4(u) = 4b | u - 1/2 | + 1 - b \quad (10)$$

$$E5(u) = -4b | u - 1/2 | + 1 + b \quad (11)$$

여기서 $E_n(u)$ 는 정의역 $0 \leq u \leq 1$ 에 정규화 되어 있다. 그리고 b 는 신속율을 나타내는 신속도이다. $K \times L$ 크기의 F 를 $i \times j$ 크기의 화상 F_m 으로 신속변환 시킬때 위에서 열거한 신속함수 $E_n(u)$ 를 이용하여 식 (8) 의 샘플링점의 위치 X_i 와 Y_j 는 다음식으로 주어진다.

$$X_i = X_1 + (K-1/i \times j - 1) \sum E_{n1} (k-1/K-2) \quad (12)$$

$$\text{단 } i = 2, 3, \dots, K-1$$

$$n1 = 1, 2, 3, 4, 5$$

$$Y_j = Y_1 + (L-1/j \times j - 1) \sum E_{n2} (\ell-1/L-2) \quad (13)$$

$$\text{단 } j = 2, 3, \dots, L-1$$

$$n2 = 1, 2, 3, 4, 5$$

3. 패턴변형

식 (4) 와 식 (7) ~ (11) 에서 $E_n(u)$ 를 X, Y 방향에 적용하면 선형변환 no ($no=1, 2, \dots, p$) 에 대해 25종류의 신속변환이 가능하고 식 (6) 의 점자 m 은 변환모드라 한다. 따라서 전체변환종류는 $25 \times p$ 개 이다.

표 (1) 은 각 변환모드의 관계를 나타낸 것이다.

표1 신속모드와 $no, n1, n2$ 의 관계

m	1	2	3	4	5	6	7	8	9	p
no	1	1	1	1	1	1	1	1	1	5
$n1$	1	1	1	1	1	2	3	4	5	5
$n2$	1	2	3	4	5	1	1	1	1	1

4. 실험 및 고찰

본 실험은 화상입력장치를 통해서 입력된 문자화상을 2値화상으로 변환시킨후 76×80 의 화상크기로 정규화 된 패턴을 테이타로 사용 하였다.

회전변환모드 no 를 3가지로 하고, 신속변환모드 $n1$ 과 $n2$ 는 각각 9가지로 하여 변환모드 m 을 27로 하였다.

신속변환의 신속도 b 의 크기를 변환후 문자 Stroke 폭의 변화가 크지않은 범위를 택하여 0.35를 선택하였다. 그림 (4) 는 각 모드별 변형된 패턴이다.

선형변환이 된 패턴 F_{no} 는 no 모드에 따라 가로방향 또는 세로방향의 Segment가 기울어진 변형이 발생하고 no 의 크기에 따라 기울어짐의 정도가 달라진다. 그리고 $n1$ 과 $n2$ 의 모드에 따라 국부적 신속율이 달라지므로 일반적인 필기체문자패턴과 유사한 변형패턴을 얻을 수 있었다.

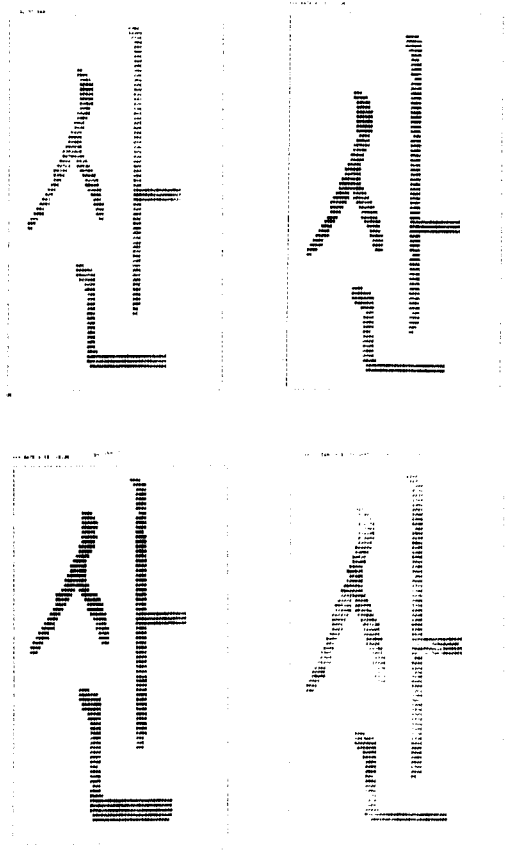


Fig 3. Distorted Character Pattern

5. 결론

인쇄체 한글문자를 입력하여 모드선택에 의한 서로 다른 변형패턴을 발생시켰다. 형성된 변형패턴은 입력패턴의 전면에 대해 국부적인 신축변형과 방향Segment의 독립적인 회전변형으로 필기체 문자패턴과 유사하다.

본 논문에서 제시한 방식은 글을 쓰는 사람의 습성에 따라 변형이 심한 필기체 문자데이터를 수집하는 대신에 정규화된 문자 패턴을 이용하여 변형패턴을 발생 시켜서 필기체 문자패턴으로 대체할 수 있다. 따라서 데이터수집에 소요되는 시간을 절약할 수 있고 데이터작성이 손쉽다.

<참고문헌>

1. Kenichiro Ishii, "Generation of Distorted Characters and Its Application", IECE. Vol.J66-D, pp1270-1277, No.11,83
2. Mitsuru, "Recognition of Handprinted Characters by Simplified Elastic Matching Method" IECE, Vol.J64-D, pp387-393, No.5,81
3. Moriyuki, "Numerical Porcessing Using Two-Dimensional Linear System Functions and Its Application to Picture Processing" IECE. Vol.J62-D, pp491-498, No.7,79
4. Suk Chul AHN, Boo Gui Choi, "A Study on Inertial Moment Normalization of the Pattern" 한국통신학회 춘계학술발표논문집, Vol.5, pp99-101, No.1,86
5. Mitsuru SHIONO, "A Method of Size Transformation and Shape Adjustment for Dotted KANJI Patterns" IECE, Vol.J63-D, pp.557-564, No.7,80