

〈論 文〉

전자제산기에 의한 필기체 한글 인식에 관한 연구

A study on the Automatic Recognition of Hand Printed
Hangeul patterns by the Computer남궁재찬*·김 영 건**
(J.C. Namkung) (Y.K. Kim)

(접수일자 80. 12. 15)

ABSTRACT

This paper proposes a method of the automatic recognition of the handprinted Hangeul patterns.

A certain pattern oriented basic letters is normalized to the prototype Hangeul pattern by the linking compensation and normalization algorithm.

Tree grammar is used in recognition process. Compared with the previous method, automata processing is simplified and the error is reduced by the new parsing method.

This method shows the effectiveness for the constrained pattern.

We expect that the new parsing method is very useful for the on-line pattern recognition.

요 약

본 논문에서는 필기체 한글인식을 위한 한 방법을 제안했다.

기본 자모를 대상으로 하였으며, 임의의 Pattern에 대하여 접합보상및 정형 Algorithm을 제안함으로써 본래의 표준 한글 pattern으로 정형화하였다. 인식에는 Tree grammar를 사용하였으며, 새로운 Parsing 방법을 제안함으로써 종래의 방법보다 처리를 간단화시켰으며 error를 감소시켰다. 제안된 필기체에 대하여는 매우 효과적이었으며 on line 필기체 인식에도 유용성이 있음을 보였다.

I. 서 론

Pattern 인식에 관한 연구는 인공 지능에 관련된 문제로서 컴퓨터의 발달과 더불어 비약적인 발전을 하여 경제, 군사, 의학, 기상학, 통신등의 여러 분야에 응용되고 있다. 또한 각종 산업 시스템은 컴퓨터에 의한 자동화의 방향으로 진행됨에 따라 Pattern 인식은 중요한 위치를 차지하고 있다. 이러한 추세에 따라 문자및 도형등의 Pattern 인식에 관한 연구는 꾸준한 발전을

하여 현재 영문자, 숫자등의 소수 문자에 관해서는 실용화가 되고 있다.

한글의 경우 문자의 수가 대단히 방대하며, 한 문자에 동일한 기본 자모가 반복 조합되기 때문에 많은 문자들이 거의 같은 유사성을 갖는 난점이 있으나, 계속하여 좋은 성과가 발표되고 있다.

인체체의 경우 논문(1,5,6)에 의하면 한글을 6개의 form으로 조직화하고 기본 자모를 분리할 수 있는 algorithm을 제시함으로써 방대한 수의 문자를 기본 자모의 식별만으로서 인식 가능하게 하였으며, 동시에

* 광운공대 전임강사, 正會員

** 인하대 대학원생, 正會員

유사성도 제거할 수 있었다.

논문(11)에서는 필기체 기본 자모를 대상으로 다상 구조에 의하여 인식하는 방법을 제안한 바 있다. 그러나, 필기체의 경우 조합문자를 대상으로 하였을 때, 기본 자모의 분리가 대단히 어려워지므로 본 논문에서도 기본 자모를 대상으로 하였다.

필기체는 개인의 특성 및 주위 환경에 따라 많은 변형을 가져오나, 사람이 필기체를 인식할 경우 자신이 알고 있는 기준 Pattern 을 연상시킴으로서 정형하여 인식한다는 점에 착안하여 변형된 필기체 Pattern 을 원형 Pattern 으로 정형시키기 위하여 접합보상 및 정형 알고리즘을 제안하였다. 인식에는 Tree grammar 를 사용하였으며 새로운 Parsing 방법 및 Primitive 를 제안함으로써 종래의 방법¹²⁾에 비교하여 처리를 간단화 하였으며 내부구조의 변형에 의한 error 를 감소시켰다. 또한 on line 필기체에서 운동단위로 처리할 경우에 분기점은 분리되어 입력되므로 본 Parsing 방법이 유용함을 알 수 있다.

II. 본 론

1. 전처리

입력의 영상패턴(image Pattern)을 양자화하여 $P(i, j) = v (1 \leq i \leq m, 1 \leq j \leq n, v \in \{0, 1\})$ 의 $m = n = 16$ bit 인 $m \times n$ matrix 를 구성하고 문자 부분은 $v = 1$, 그 외 부분은 $v = 0$ 인 2치 Pattern 으로 하였다. 2치 Pattern 에서 각 $P(i, j)$ 를 Pixel로 정하고 Pattern 내의 각 Pixel($v = 1$)의 topological 성질에 따라 전체 Pattern 의 local operation 을 주어 세선화 한다.⁽¹¹⁾

입력과정시 noise 로 인하여 발생된 한개 또는 두 개의 고립 Pixel 은 제거하였으며 세선화된 Pattern graph 에서 단점, 끝꼭점, 분기점을 그림 1과 같이 T, N, B 로 정의하였다. Pattern 의 선분방향은 8방향으로 하였으며 T, N, B 간의 선분방향은 기울기 및 Next Searching

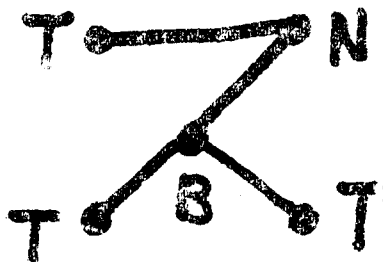


그림 1. T, N, B 의 예

Decision 방식에 의하여 선분방향을 결정되었다.

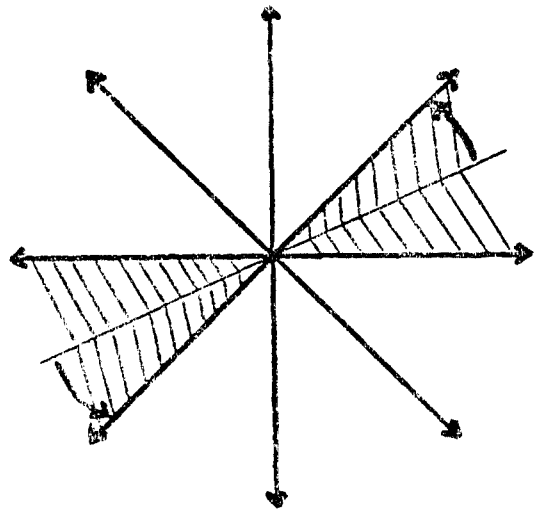


그림 2. 선분결정 구역

2. 접합보상 Algorithm

필기체인 경우 Pattern 의 요소들이 떨어진 경우가 많으므로 이를 보상하기 위하여 단점 T에 threshold 값을 정하여 연장시키고 이들이 서로 접합하는 경우 그림 3과 같이 node 를 잡아주고, 접합되지 않거나 접합 시 생기는 불요부분은 제거시킨다. 그림 3은 8방향중 2방향이상의 선분이 접합됨으로서 node 가 발생하는 것을 보였다. 그 실제 처리에는 그림 4에 보였다. 이 때, 접합에 의해서 생겨지는 noise 인 불을 제거하였다.

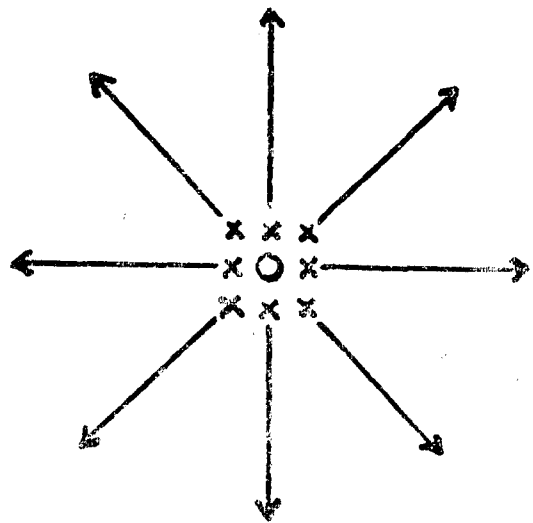


그림 3. 접합보상

3. 정형 Algorithm

필기체의 경우 분기점 B는 그다지 큰 변형을 일으키

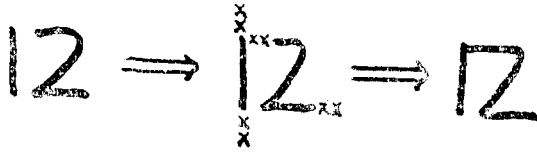


그림 4 처리예

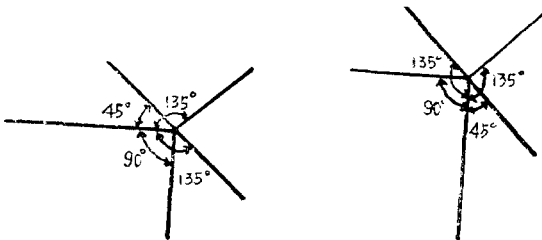


그림 5. 방향각

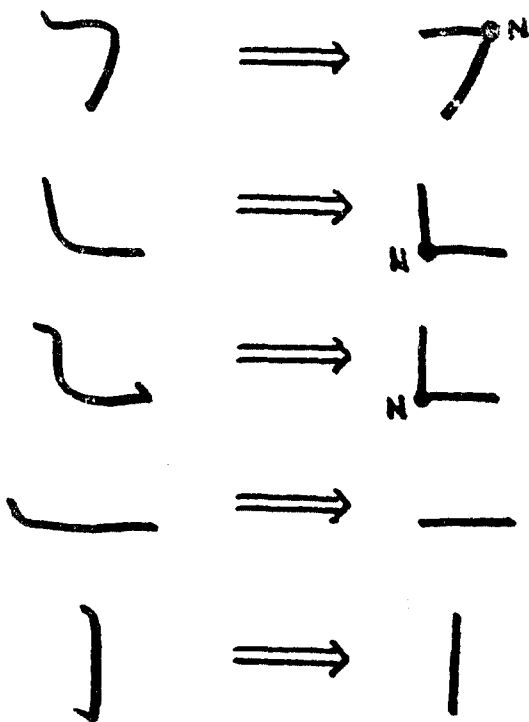


그림 6. 처리예

지 않으나 끝꼭점 X는 많은 변형이 발생된다. 표준 Pattern 의 끝꼭점은 방향각이 90°내지 45°이므로 필기체의 경우도 90°이내의 방향각은 끝꼭점으로 인정한다. 그러나, 135°의 경우는 패턴의 고유 특징 및 주위 여건을 고려하여 결정한다.

방향각의 예는 그림 5와 같고 습관상으로 발생하는 뿔은 제거시킨다. 또한, 방향각이 존재하나 그 선분길이 Threshold 값이 안될 때는 인식대상에서 무시하였다.

4. 인식 Algorithm

본 논문에서는 분기점 B는 진행 방향을 기준으로 하여 분리시켜 Parsing 하는 방법을 제안하였다. Tree를 위한 Primitive는 그림 7과 같다. 이를 종합한 tree는 표 1에 보였다. 표 1에서 모음의 경우에는 동일 tree가 있으므로 이들을 식별하기 위해서 분기점에서 재차 Parsing 하므로써 식별이 가능케 하였다.

이는 종래의 tree grammar⁹⁾를 이용한 방법에 비하여 처리과정을 간단화시켰으며 frontier⁶⁾만을 이용한 grammar에서 일어날 수 있는 내부구조의 변형에 의한 error를 감소시켰다.

Parsing 예는 그림 8과 같다.

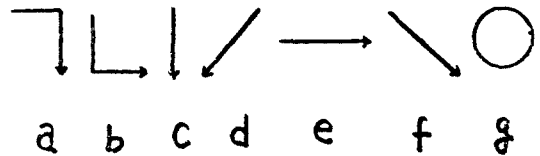


그림 7. Primitives

자음	순서 primitive
ㄱ	a
ㄴ	b
ㄷ	cb
ㄹ	acb
ㅁ	ab(eac)
ㅂ	bec(eecc)
ㅅ	df
ㅇ	g
ㅈ	af
ㅊ	caf(eaf)
ㅋ	ac
ㅌ	ccb(cba)
ㅍ	ceec
ㅎ	ccg

모음	1차 순서 primitive	2차 순서 primitive
ㅏ	ec	ce
ㅑ	ecc	cec
ㅓ	ec	e
ㅕ	ecc	e
ㅗ	ec	c
ㅛ	eec	
ㅜ	ce	
ㅠ	cee	
ㅡ	e	
ㅣ	e	

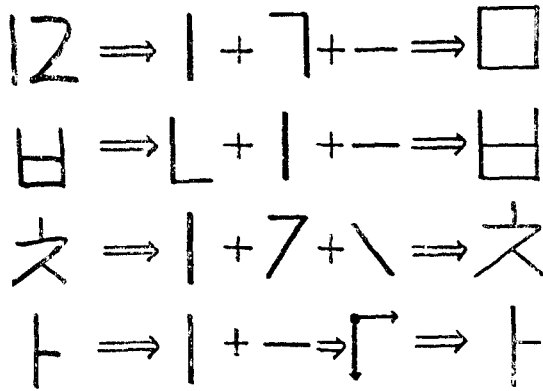


그림 8. 처리예

Ⅲ. 결 론

필기체에 대하여 집합보상 및 정형 Algorithm을 적용하여 효과적으로 원형패턴으로 정형시켰으며 새로운 Parsing 방법을 제안함으로써 종래의 tree grammar에 비하여 처리 과정을 간단화시켰고 내부구조의 변형으로 발생할 수 있는 error를 감소시켰다. 또한 제한된 필기체에 대하여는 조합문자도 논문 (6)에 의한 분리방법을 적용시켜부분 분리한 후 인식할 수 있는 가능성을 제시하였다. 또한 본 Parsing 방법은 분기점이 분리되므로 운동단위로 입력처리되는 on line 필기체 인식에 유용성이 있음을 밝혔다.

문제점으로는 모음의 구조 단순성에 의해서 동일 tree가 발생되는 점이 있으나 이들은 Primitive 선정 및 Parsing 방법의 향상으로 해결될 수 있을 것으로 전망된다.

REFERENCES

1. Lee, J.K. "Korean Character Display by Variable Combination and Its Recognition by Decomposition Method," Ph. D. dissertation in Keio University, Japan.
2. Lee, J.K. "A Method for the Recognition of Printed Korean Characters," J. KIEE, Vol. 7, No. 4, pp. 198-209, Dec. 1969.
3. Lee, J.K. "Recognition of Printed Korean Characters (II)," J. KIEE, Vol. 7, No. 1/3, pp. 130-136, Nov. 1970.
4. Lee, J.K. "Mathematical Analysis of the Structure of Korean Characters," J. KIEE, Vol. 9, No. 4, pp. 501-508, Sep. 1972.
5. 이주근, 김영진, 남궁재찬, "Character Pattern 위부분분리에 의한 인식 Algorithm," 정보과학회 학술발표 논문집 pp. 43-50, Apr. 1980.
6. 이주근, 남궁재찬, 김영진 "Window-Cutting Algorithm에 의한 Character Form 식별 및 Subpattern 분리" 전자공학회 학술발표 논문집, Vol. 3, No. 1, pp. 31-35, 1980.
7. Kim, T.K. and T. Agui, "A Study on the Pattern Recognition of Korean Characters," J. KIEE, Vol. 14, No. 5, pp. 15-21, Dec. 1977.
8. Williams, K.L. "A Multidimensional Approach to Syntactic Pattern Recognition," Pattern Recognition, pp. 125-137. Sep. 1974.
9. Fu K.S. and Bhargava, B.K. "Tree System for Syntactic Pattern Recognition," IEEE Trans. on Computers, Vol. C-22, No. 12, pp. 1087-1099, Dec. 1973.
10. Lee, J.K. H. Choo, "Algebraic Structure for the Recognition of the Korean Characters," J. KIEE, Vol. 12, No. 2, pp. 11-17, Apr. 1975.
11. Lee, J.K. and Kim, H.K. "Automatic Recognition of Handwritten Hangeul by the Phase Rotation," J. KIEE, Vol. 13, No. 1, pp. 23-30, Mar. 1976.
12. Joo, B.G. "A Study on the Recognition of Hangeul Character Patterns," MS. dissertation in KAIS, 1980.

13. Kim, M.H. and Park, C.M. "Algorithms for Machine Recognition, Noise Cleaning and Thining of the Korean Characters," "Proceedings of the International Computer Symposium, Vol. 2, pp.513—519. 1975.
14. Fu, K.S. and Rosenfeld, "Pattern Recognition and Image Processing," IEEE Trans. on Computers, Vol. C-25, No. 12, pp.1336—1346, Dec. 1976.
15. Kanel, L. "Patterns in Pattern Recognition: 1968—1974," IEEE Trans. on Information Theory, Vol, IT-20, No. 6, pp.697—722, Nov. 1974.
16. Nakata, K. "Recognition of Handprinted Alphanumeric and Special Charcters," IECE, Vol. 58-D. 8, pp.442—449, Aug. 1975.
17. Yamamoto K. and T. Mori and S. Mori, "Machine Recognition of Handprinted Japanese Character KATAKANA and Numericals-Topological Line Segment Method and Automatic Formation of Mask-," IECE. Vol. J59-D, No. 6, pp.414—421, July, 1976.
18. Fu, K.S. "Syntactic Methods in Pattern Recognition" New York: Academic Press, 1974.
19. Tou J.T. and Gonzalez, R.C. "Pattern Recognition Principles" Addison-Wesley Publishing Company, 1974.
20. _____, "Recognition and Translation of Syntactic Structures," in Syntactic Pattern Recognition An Introduction, R.C. Gonzalez and M.G. Thomason, Ed. Addison-Wesley Publishing Co. 1978.
21. _____, "Symbolic Image Description," in Digital Image Processing, W.K. Pratt, Ed. New Work: John Wiley & Sons, Inc. 1973.
22. You K.C. and Fu, K.S. "A Syntactic Approach to Shape Recognition Using Attributed Grammars," IEEE Trans. on SMC, Vol SMC-9 No. 6 Jun. 1979.
23. Sherman H.: A quasi-topological method for the recognition of line patterns. Information processing. Proceedings of a Unesco conference, 1960.
24. Stallings W.: Recognition of a printed Chinese characters by automatic pattern analysis. Computer Graphics and Image Procesing, Vol. 1, No. 1. Apr. 1972.
25. Morie S.: Algebraic Structure of Characters. Journal Electronics, Apr. 1974 and Sept. 1974.
26. Masuda I.: Machine Recognition of Hand Printed Japanese Characters, Kadakana. J. ECE, Vol. 55-d, No. 10, Oct. 1972.
27. Suen C.Y. and MARC BERTHOD, SHUNJI Mori "Automatic Recognition of Handprinted characters-The State of the Art," Proc. of the IEEE, Vol. 68, No. 4, April, 1980. pp. 469—487.
28. SUZUKI S. "Four Properties of a Structured Model in Pattern Recopnition and their Applications" J. IECE Vol. J 60-D No. 9. Sept. 1977 pp. 710—717.
29. ITO M.R. and Chui T.L. "On-Line Computer recognition of proposed standard ANSI(USAST) Hardprinters" IEEE P.R. Vol. 10, pp. 341—349 Feb. 1978.
30. Shapiro L.G. "A Structural model of Shape" IEEE Trans. on pattern analysis and Machine in telligence, Vol. PAMI-2, No. 2. March 1980.