

論文99-36C-7-4

신경회로망을 이용한 제약 없이 쓰여진 필기체 문자열로부터 단어 분리 방법

(Segmentation of Words from the Lines of Unconstrained
Handwritten Text using Neural Networks)

金 敬 煥 *

(Gyeonghwan Kim)

요 약

필기서술의 인식과 관련된 연구는 인식대상 영상이 바르게 분리된 인식단위를 포함한다는 전제로 진행되어 왔다. 그러나 실제적인 필기인식 시스템의 설계에 있어서, 다양한 필기방식으로 인해, 인식단위로의 분리가 선결되어야 할 문제이다. 본 논문에서는 제한없이 쓰여진 필기 문자열로부터 인식의 도움없이 독립된 단어를 분리하는 방법을 제안한다. 구성요소간 물리적인 거리에 의존하는 종래의 방법과 달리, 필기서술 자체로부터 필기자의 띄어쓰기와 관련된 특징들을 적극적으로 추출하고 이를 신경회로망을 사용하여 해석한다. 띄어쓰기와 관련된 정보는 문자 분리과정을 통해 분리된 문자 세그먼트의 높이와 세그먼트 중심선 사이의 간격들을 정규화하여 구한다. 연결요소간의 거리에 기반한 방법들과의 비교실험을 통해 제한한 방법의 유용성을 입증하였다.

Abstract

Researches on the recognition of handwritten script have been conducted under the assumption that the isolated recognition units are provided as inputs. However, in practical recognition system designs, providing the isolated recognition unit is an challenge due to various writing styles. This paper proposes an approach for segmenting words from lines of unconstrained handwritten text, without help of recognition. In contrast to the conventional approaches which are based on physical gaps between connected components, clues that reflect the author's writing style, in terms of spacing, are extracted and utilized for the segmentation using a simple neural network. The clues are from character segments and include normalized heights and intervals of the segments. Effectiveness of the proposed approach compared with the conventional connected component based approaches in terms of word segmentation performance was evaluated by experiments.

I. 서 론

필기서술의 인식과 관련된 연구의 대부분은 분리되

어 독립된 인식단위(문자/단어)들의 제공을 전제로 진행되어 왔다. 예를 들어, 필기문자 인식의 경우에는 분리된 문자가 입력으로 제공된다고 전제하며, 필기단어 인식의 경우에는 입력 영상이 분리된 하나의 단어를 포함한다고 가정한다.^[1] 그러나, 실제 인식을 위해서는 그 분리된 인식단위의 제공 자체가 선결되어야 할 문제임을 쉽게 발견하게 된다. 필기서술을 포함하는 대부분의 문서영상은 하나 이상의 인식단위를 포함하게 되며, 필기자에게 인위적인 제약을 가하지 않는 한

* 正會員, 西江大學校 電子工學科

(Dept. of Electronic Engineering, Sogang University)

※ 본 연구는 1998년도 서강대학교 교내 연구비 지원에 의하여 이루어졌음

接受日字:1999年3月8日, 수정완료일:1999年6月30日

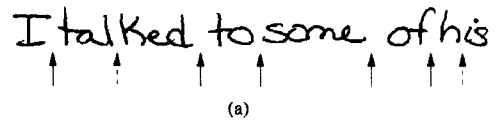
문서영상에 포함된 필기서술로부터 인식시스템이 전제로 하는 독립된 인식단위를 자동으로 분리하는 과정이 반드시 필요하기 때문이다.

이 문제를 완화시키기 위하여 인위적인 제한을 가하는 경우를 주변에서 쉽게 발견할 수 있다. 편지봉투에서 그려진 우편번호 기입란은 6자리의 우편번호를 각각의 숫자로 쉽게 분리하기 위함이다. 은행의 지로용지와 세금계산서 등의 금액을 적기 위해 제공되는 칸들도 이 문제를 완화시키기 위한 인위적인 제한으로 해석할 수 있다. 많은 수의 문서를 컴퓨터를 이용하여 자동으로 인식·처리하기 위하여 스캐너에 의해 감지되지 않는 잉크(dropout ink)를 사용한 양식을 제공하는 경우가 점차 증가하는 것 또한 인식단위의 분리 문제를 완화하려는 노력의 일환으로 해석할 수 있다. 이와 같은 인위적인 노력은 인식을 통해 처리하려는 영역의 내용이 고정된 형식을 갖거나 최소한 그 범위라도 정해진 경우에만 가능하게 된다. 예를 들어, 우편번호는 6자리이고, 주민번호는 13자리이며, 직업을 표시하는 칸에는 대개의 경우 1-3개의 단어가 기록되는 식의 정보가 주어지게 된다. 만일, 인식대상 필기서술이 포함하는 문자 또는 단어의 수(또는 수의 범위)가 알려지지 않거나 제한되지 않으면, 발생 가능한 모든 경우를 고려해야하기 때문에 인식관련 모듈의 계산 복잡도가 크게 증가하게 되어 설계하는 시스템의 효용성이 처리속도와 인식신뢰도 측면에서 크게 저하된다.

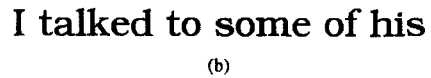
본 논문에서는, 단어의 수가 알려져 있지 않은 제약 없이 쓰여진 일반적인 필기문자열로부터 인식의 도움 없이 구성단어를 효율적으로 분리할 수 있는 방법을 제안한다. 즉, 그림1(a)처럼 필기된 문자열 영상에서, 화살표로 표시된 연결요소간 간격들 중 실선으로 표시된 간격을 구분하는 것이 목적이다. 필기서술을 구성하는 요소들 사이의 물리적인 거리를 계산하고 이들의 통계적 특성에 따라 단어를 분리하는 기존의 방법들^[2,3,4]과는 달리 필기자의 필기습관으로부터 띄어쓰기와 관련된 단서를 추출하여 신경회로망을 통해 해석하는 방법을 소개한다.

인식대상은 제약 없이 쓰여진 필기 영문 문자열이며, 이 문자열은 다양한 필기형식을 포함한다. 각 문자열은 한 개 이상의 단어를 포함한다. 제안하는 방법은 인식기의 도움 없이 문장을 구성하는 단어를 정확하게 분리하도록 설계되었기 때문에, 인식과정에서 단어분리

가설을 모두 점검하기 위해 수행되어야 하는 경우의 수를 크게 감소시킬 수 있다. 인식과정의 도움을 받지 않고, 제약 없이 쓰여진 문자열을 그 구성단어들로 완벽하게 분리하는 것은 거의 불가능하지만, 가능한 효율적으로 단어들을 분리하면 인식시스템에서 가장 많은 처리시간을 사용하는 인식관련 모듈의 작동 빈도를 감소시켜 실용적인 시스템의 구현을 가능하게 한다. 구성요소간 물리적인 거리를 사용하는 종래의 방법들과, 본 논문에서 제안하는 방법들의 단어 분리 능력에 임의의 수의 단어들을 포함하는 문자열 영상을 사용하여 실험을 통해 비교한다.



(a)



(b)

그림 1. 문자열로부터 단어분리: (a) 필기서술의 경우, (b) 인쇄된 문자열
Fig. 1. Segmenting words from a line of text.

II. 단어분리의 일반적인 문제

필기서술에서 인식단위의 분리문제는 필기형태의 다양성에서 출발한다. 인쇄된 문자들은 글자간 간격이 일정하거나 최소한 예측할 수 있고, 대부분의 경우 단어간 간격이 문자간 간격에 비해 크기 때문에 문자 및 단어의 분리가 문제되지 않는다(그림1(b)). 반면, 필기서술의 경우에, 흘려쓰는 서체(cursive)에서는 문자사이의 간격을 기대하기 어려우며, 인쇄체(discrete) 서술에서는 필기환경 및 필기자의 습관에 따라서 단어간의 간격이 문자간의 간격보다 클 것이라는 일반적인 전제를 벗어나는 경우를 쉽게 발견할 수 있다(그림 1(a)). 특히 영문 필기서술에서는 시작하는 문자가 화려하게 쓰여지는 경우가 많고 단어의 끝에 쓰여지는 문자의 꼬리가 과장되는 경우가 많아 단순히 물리적인 간격의 해석으로는 완전한 단어의 분리가 어렵게 된다.

문자열로부터 독립된 단어를 분리하는 문제는 그 문자열을 처리해야하는 전체 시스템의 성능과 직접적으로 연관된다. 정확한 단어의 분리가 전제되지 않으면, 단어 인식기는 여러 단어분리 가설에 따른 조합 가능한 경우의 수만큼의 인식을 수행하여 최적의 인식결과

를 내는 경우를 찾아 내어야 한다.^[4] 일반적인 인식시스템에서 인식과정이 가장 많은 계산시간을 소비하고 조합의 수가 증가할수록 오인식의 가능성도 함께 증가하여 전체 시스템의 속도와 정확도가 함께 낮아지기 때문이다. 이러한 문제를 제거하기 위해, 실험 단계에서는, 필기영상 취득 시 인위적인 제약을 두거나,^[5] 단어분리가 문제되지 않을 정도의 간격을 단어간에 유지하고 단어를 구성하는 문자를 모두 연결하여 쓰도록 제한하는 경우^[6]도 있다.

주어진 문자열에 포함된 단어의 수에 관한 정보 또한 단어분리 문제와 연관되어 있다. 만일 한 문자열에 포함된 단어의 수가 고정되어 있다면, 단어분리의 문제를 문자들 사이의 간격을 해석하여 정해진 단어의 수로 분리하는 작업으로 해석할 수 있다. 단어 수의 범위가 정해진 경우에도, 정해진 범위에 따른 수만큼의 단어분리 가설을 통해 단어를 분리하고 각 가설을 인식기와 후처리과정의 도움을 받아 검증하는 방법으로 문제를 해석할 수 있다.^[4] 그러나 일반적인 문장처럼 하나의 문자열이 포함할 수 있는 단어의 수에 제한이 없을 때, 설정할 수 있는 가설의 수가 급증하게 되어 처리 속도 및 정확도 측면에서 실용성을 잃게 된다.

1. 인식을 위한 분리 - 분리를 위한 인식

일반적인 top-down 문서해석 방식을 고려할 때, 문자열을 해석하기 위해 한 문자열은 구성하는 단어들로 분리되어야 하고, 단어인식을 위해 한 단어는 구성하는 문자들로 분리되어야 한다. 필기서술에서 문자 또는 단어의 경계를 정확하게 구분하는 일이 앞서 언급한 것처럼 쉽지 않기 때문에, 인식대상을 초과 분할하고 인식기의 도움을 받아 초과 분할된 세그먼트의 조합과 인식을 반복하며 최선의 인식결과를 내는 조합을 선택하는 방식을 일반적으로 취한다. 결국 인식단위의 분리가 인식을 위함이지만, 정확한 분리를 위해 인식기의 도움이 필요하다는 역설^[7]을 마주하게 된다.

인식의 도움이 없이 인식단위를 가능한 한 정확하게 분리할 수 있다면 시스템의 전반적인 성능을 크게 향상시킬 수 있게 된다. 정확한 단어분리를 위해, 간단한 특징을 문자 세그먼트로부터 추출하여 활용하는 적극적인 단어분리 방식이 본 논문에서 제안된다.

입력된 영상에 포함된 연결요소 (connected components)간의 기하학적인 간격 해석에 의존하는 기존의 접근방법들과 입력영상으로부터 피어쓰기와 관

련된 필기특징을 찾아 해석하는 방법으로 구분하여 설명한다.

2. 연결 요소에 기반한 접근 방법

단어간의 간격이 문자간의 간격에 비해 크다는 일반적으로 직관적인 사실에 근거하여 제안된 방법이다. 입력영상에 존재하는 모든 연결요소들 사이의 수평거리를 측정하고 그 통계적 분포에 따라 입력영상이 포함하는 단어의 수를 추정하여 분리하는 방법이다. 기존의 방법들은 연결요소간 간격을 정의하는 방식에 따라 분류될 수 있으며 정리하면 다음과 같다.^[2,3]

● **Bounding box**: 연결요소를 포함하는 최소사각형을 정의하고 한 최소사각형의 우측변과 다음 최소사각형의 좌측변 사이의 수평거리를 연결요소간 간격으로 정의하여 사용한다.

● **Run-length**: 최소 사각형에 의해 정의된 간격이 과장된 획들에 의해 쉽게 영향을 받는 단점을 개선하기 위해 제안된 방법으로 영상이 run-length 방법으로 표현되었을 때 좌측 연결요소의 run들과 우측 연결요소의 run들 사이의 최소거리를 찾아 연결요소간 간격으로 정의하여 사용한다.

● **Convex hull**: 한 연결요소를 감싸는 최소 사각형에 의해 경계를 정하는 대신, convex hull 방법은 연결요소를 감싸는 최소 다각형으로 경계를 정한다. 이때, 연결요소간 간격은 여러 형태로 정의 될 수 있는데, 일반적으로 인접한 두 최소 다각형의 무게 중심을 연결하였을 때, 이 선상에서 측정한 두 교점 사이의 거리를 연결요소간 간격으로 정의하여 사용한다.

이상의 내용을 그림으로 정리하여 그림2에 나타내었다.

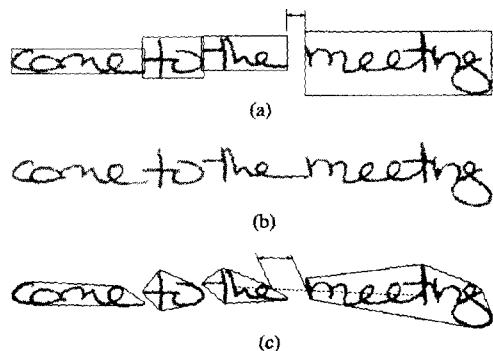


그림 2. 연결요소간 간격의 정의: (a) bounding box, (b) run-length, (c) convex hull
Fig. 2. Distance measure between connected components.

3. 필기특징에 기반한 접근 방법

필기서술이 아무런 제약 없이 자연스럽게 쓰여졌다고 가정할 때, 필기자의 띄어쓰기와 관련된 정보를 입력 필기서술 자체에서 추출하여 활용하는 방법이다.^[8,9] 즉, 문자간 간격을 특징짓는 정보를 추출하여 이를 단어간 간격을 찾아내는데 활용한다. 대부분의 사람들이 글씨 쓰는 방법을 처음배울 때, 글자의 균형을 유지하기 위해 기본적인 손동작을 익히게되는데 이를 통해 익힌 방법은 정한 글자의 크기와 지면의 여백에 따라 필기자가 문자간 간격을 조절할 수 있게 한다. 글자의 크기가 변하거나 지면의 여백에 제한을 받더라도 단어들 이 구분될 수 있도록 최소한의 간격을 확보하기 위해 조절되는 문자내의 특징을 찾아 활용하는 것이 이 접근방법의 기본적인 출발점이다.

이러한 특성을 반영하는 특징으로 쉽게 활용할 수 있는 것 중의 하나가 위에서 아래로 향한 주요 획(세로획)들이다. 세로획들은 필기자의 필기습관 중 pitch에 관련된 정보를 포함하고 있으며 이 pitch 정보를 이용하여 띄어쓰기와 관련된 단서를 추출하여 단어분리에 활용한다. 세로획의 분리를 위해서 주어진 입력 영상으로부터 특징들을 추출하는데, 경우에 따라서는 하나의 연결요소를 하나 이상의 세그먼트로 분리하는 적극적인 특징 추출과정이 필요하게 된다. 이러한 적극적인 특징 추출과정은 영상의 표현방식에 따라서 간단하게 수행될 수 있다. 실용적인 시스템에서 단어의 인식을 전제로 단어를 분리할 경우, 단어인식단계에서 수행되는 문자분리 결과가 위에서 언급한 특징을 중심으로 수행되는 경우가 많기 때문에, 문자분리 단계를 미리 수행하여 단어분리 단계에서 활용한 후 분리된 단어를 인식하는데 다시 사용할 수 있어 추가적인 계산·시간상의 부담을 최소화 할 수 있다.

III. 신경망을 이용한 방법

신경회로망은 다수의 처리단위를 다층으로 배열하고 그 사이를 학습이 가능하도록 연결된 구조를 갖고, 알려진 패턴에 대해 학습된 후 비슷한 양상을 지닌 미지의 패턴에 대해 적응적으로 반응하는 특성 때문에 패턴의 인식 및 분류에 많이 활용된다. 일반적인 신경회로망은 그 구조적인 특성상 정적인(static) 패턴의 인식·분류에 주로 활용되고 있다. 반면, 음성을 비롯한 시간에 따라 특성이 변하는 신호의 해석을 위해 고안

된 신경회로망이 시간지연 신경망이다.^[10] 단일 문자의 인식에는 일반적인 신경회로망의 적용이 가능하지만, 연속적으로 쓰여진 하나 이상의 문자를 포함하는 필기서술의 해석에는 음성신호 해석을 위해 개발된 알고리즘들이 많이 적용되는데, 시간의 흐름에 따른 음성신호의 변화에 대한 해석 기법을, 좌측에서 우측으로 쓰여지는 필기서술의 해석에 적용가능하기 때문이다. 본 논문에서 다루는 문자열로부터 단어를 분리하는 문제도 시간의 흐름에 따른 신호의 해석으로 간주할 수 있기 때문에 시간 지연 신경망의 개념을 적용한다.

1. 시간 지연 신경회로망

인지와 관련된 연구에서 시간 개념의 도입은 컴퓨터 비전 및 음성신호처리 분야의 연구에서 중요한 역할을 하는데, 문제는 시간을 어떻게 표현·해석하는가 하는 것이다. 특히 일반적인 다층 신경회로망의 구현에 있어서 어떤 보안을 통해 시간개념이 포함된 신호를 처리할 것인가 하는 것이 중요한 문제이다. 이에 대한 한가지 해결 방안으로 제안된 것이 시간지연을 신경망에 도입하여 신경망이 시간에 따라 변화하는 신호에 대해 동적으로 동작할 수 있는 특성을 갖게 하는 것이다. 이러한 접근 방법은 신경생물학적으로 볼 때 동물 신경계의 정보처리 과정과 부합하는 것으로 알려져 있다.^[11] 시간지연 신경회로망은 다층 순방향전파 신경망 (multi-layer feed-forward neural network)의 하나로서 그 은닉층과 출력층의 처리단위가 시간에 대해 반복되는 구조를 갖는다.

2. 제안하는 단어분리 방법

문자열로부터 추출된 띄어쓰기와 관련된 특징을 시간 지연 신경회로망의 개념을 적용하여 제약 없이 쓰여진 문자열로부터 단어를 분리하는 방법을 소개한다. 전체 구성은 그림3에 나타난 것과 같고, 각 기능모듈을 중심으로 설명한다.

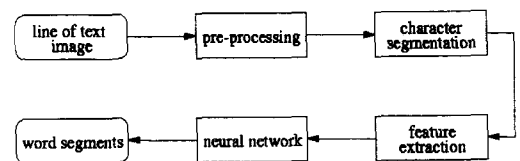


그림 3. 단어분리 절차

Fig. 3. Word segmentation procedure.

1) 영상처리 및 문자분리

입력 영상의 표현 방식 변화와 기울기 보정 및 잡영 제거 등의 과정과 문자열을 문자단위로 분할 (character segmentation)하는 과정은 별도의 논문^[12]에 설명되어 있다.

2) 특징의 정의 및 추출

신경회로망의 입력으로 사용될 특징은 필기 입력영상의 세로획으로부터 추출된다. 단어 인식을 위한 초과 문자 분할방식을 적용하면 각 문자세그먼트는 대부분 주요 세로획을 포함하게된다.^[12, 13] 따라서, 문자 분할방법을 적용하여 얻게되는 문자세그먼트로부터 특징을 추출한다. 그 과정을 단계별로 설명하면 다음과 같다.

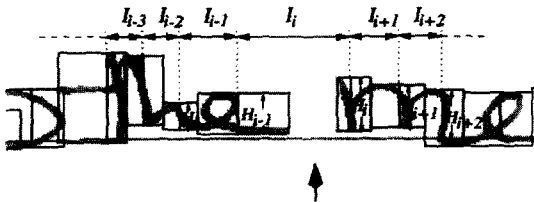


그림 4. 단어분리를 위한 특징 정의
Fig. 4. Features for the word segmentation.

- 각 문자 세그먼트를 포함하는 최소 사각형(bounding box)을 구한다.
- 주요 세그먼트를 기준으로 특징 높이 위에 존재하는 획들은 일정 규칙에 의해 합병되며, 합병된 획들을 모두 포함하는 새로운 최소 사각형을 다시 구한다. 합병 후, i 번째 최소 사각형을 B_i 로 정의한다.
- 문자열의 국소 최소점들을 curve-fitting하여 기준선(Baseline)을 구한다.
- 문자 세그먼트 내에 포함되는 획의 화소 분포에 따른 무게중심을 구하여 중심선을 구한다. B_i 의 중심선을 C_i 로 정의한다.
- 중심선 사이의 수평간격을 문자 세그먼트간 간격으로 정의하여 신경회로망의 입력으로 사용한다. i 번째 문자 세그먼트간 간격 I_i 는 다음과 같이 정의한다.

$$I_i = X(C_{i+1}) - X(C_i)$$

여기서 $X(C_i)$ 는 C_i 의 x -좌표를 의미한다.

- 기준선으로부터 각 문자 세그먼트의 상단까지의 거리를 문자 세그먼트의 높이로 정하여 신경회로망의 입력으로 사용한다. B_i 의 높이는 다음과 같이 정의한다.

$$H_i = \begin{cases} Y_u(C_i) - \text{Baseline}(X(C_i)) & \text{if } \text{Baseline}(X(C_i)) \geq Y_b(C_i) \\ Y_u(C_i) - Y_b(C_i) & \text{otherwise} \end{cases}$$

여기서 $Y_u(C_i)$, 와 $Y_b(C_i)$ 는 각각 C_i 의 상단과 하단의 y -좌표를 의미한다. $\text{Baseline}(C_i)$ 은 기준선에서 C_i 의 y -좌표를 의미한다. 그림4에서 볼 수 있는 것처럼, 기준선 아래 부분의 중심선은 문자세그먼트의 높이계산에서 제외되는데 이는 f, g 와 같은 문자에서 볼 수 있는 아랫부분(descenders)의 영향을 최소화하기 위함이다.

여기서 신경회로망의 입력으로 문자 세그먼트간 간격뿐만이 아니라 높이도 공급하는 이유는, 일반적으로 단어의 시작 부분이 대문자로 시작되거나, 또는 나머지 부분에 비해 크게 쓰여지는 경향을 특징으로 반영하기 위함이다. 가로획에 의한 간섭을 최소화하기 위해 획을 구성하는 화소의 분포를 통해 중심선을 구한다. 그림4에, 위에서 정의한 특징들을 예를 통해 나타내었다. 그림에서, 화살표로 표시된 부분이 단어간 간격인지 여부를 판단한다고 가정할 때, 이를 중심으로 신경회로망의 입력으로 사용될 전후의 중심선간 간격들과 높이를 표시하였다.

이러한 과정을 통해 구해진 H_i 와 I_i 는 신경회로망의 입력으로 사용되기 위해 정규화 과정을 거치게 되는데, 이 때 지나치게 과장되어 표현된 단어간 간격이나 크기가 큰 문자가 있을 경우 나머지 높이와 간격들의 정규화에 영향을 주게된다. 이러한 영향을 최소화하기 위해 다음과 같은 과정을 통해 정규화를 수행한다.

- H_i 와 I_i 의 각각의 평균과 표준편차를 구한다.
- 만일 H_i 와 I_i 각각의 최대치가 각각의 평균 + $2.0 \times$ 표준편차 보다 크면 평균 + $2.0 \times$ 표준편차로 H_i 와 I_i 들을 정규화하고,
- 그렇지 않으면 각각의 최대치로 H_i 와 I_i 를 정규화 한다.

3) 신경회로망의 설계

시간지연을 갖는 신경회로망은 SNNS^[14]를 사용하여 구현한다. 시험중인 두 문자 세그먼트의 간격을 중심으로 2개의 선행 세그먼트 간격 및 2개의 후행 세그먼트 간격과 2개 선행 세그먼트의 높이 및 3개의 후행 세그먼트 높이를 입력으로 받아 해석할 수 있도록 설계되었다. 이상의 신경망 입력들이 지역적(local) 특징을 반영한다. 전역적(global) 특징의 반영을 위해 마지막 특징쌍에 입력 문자열의 평균 간격과 높이를 추가하였다. 따라서 설계된 신경망은 그림5에 나타난 것처럼 간단한 구조를 갖는데, 입력층에서 전체 지연 길이는 5이며, 은닉층과 출력층에서의 지연 폭은 4이다. (그림5에서 은닉층의 각 노드에 대하여 반복적인

구조를 갖는데, 그림에서는 첫 노드에 대해서만 표시한다.)

영상에서 추출된 특징들은 단어간 간격여부의 결정이 진행되면서 한 문자세그먼트씩 순환하며 사용된다. 설계된 신경회로망의 구조 때문에 선행 및 후행 세그먼트들의 정보가 필요하게되는데, 문자열의 시작 부분과 끝 부분에서 해당 정보들을 공급하기 위해 입력 문자열의 처음 부분과 마지막 부분을 띠처럼 연결된 구조로 해석한다. 즉, 마지막 문자 세그먼트와 첫 문자 세그먼트 사이의 간격은 최대값(1.0)을 갖는다고 가정하고 신경망 입력 정보를 추출하여 첫 문자 세그먼트 간격부터 마지막 문자 세그먼트 간격을 모두 확인하여 단어분리 과정을 수행한다.

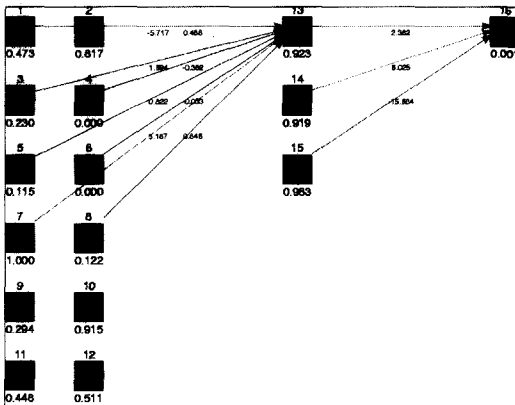


그림 5. 단어분리를 위해 설계된 신경회로망
Fig. 5. The neural network for the word segmentation.

IV. 실험 결과 및 성능 비교

본 절에서는, 단어분리와 관련하여 이미 발표된 방법들과의 비교를 통해 본 논문에서 제안하는 방법의 성능을 평가한다. 기존의 방법들은 이미 발표된 문헌에 근거하여 구현되었으며 방법들 간의 공정한 비교를 위해 필요한 최소한의 가정들을 추가하여 실험을 수행하였다.

1. 실험환경

단어분리와 관련된 방법은 주로 C와 C++ 언어를 사용하여 구현하였으며, Pentium II 350Mhz 프로세서와 128Mbyte의 메모리를 탑재하고 Linux 2.0을 운영체제로 갖는 컴퓨터에서 수행하였다. 학습 및 성능측정에 사용된 문자열 영상은 일반적인 필기 영어문

장을 포함하는 페이지 영상에서 문자열 분리를 통해 수집하였고, 학습에 사용된 영상이 성능측정에 사용되지 않도록 구분하였다. 실험에 사용된 문자열 영상들은 육안검사에 의해 단어 분리점의 위치를 표시하여 학습을 위한 특징추출과 성능검사를 프로그램을 통해 할 수 있도록 하였다. 여기서 주목할 점은, 필기문장을 수집할 때 필기를 요청한 문장들을 필기자에게 보여주지 않고 별도의 녹음 시설을 통해 청취하며 필기하도록 함으로서 필기자가 보여지는 문장예제에 의해 필기의 서체나 여백할당 등을 인위적으로 조절하지 않고 자신의 고유필체로 제한 없이 필기할 수 있도록 실제적인 환경을 설정하였다. 표1에 실험에 사용된 문자열 영상과 관련한 통계를 필기 형태별로 정리하였다.

표 1. 실험에 사용된 문자열 영상 통계
Table 1. Number of images used for the experiment.

구분		cursive	discrete	mixed	계
문자열 영상 수	training	345	119	206	700
	testing	868	438	541	1,847
평균단어 수/영상		7.1			

2. 연결요소에 기반한 접근 방식의 성능

제2절에서 소개한 연결요소에 기반한 접근 방식 중, bounding-box에 의한 방법과 이의 단점을 보완하기 위해 제안된 convex-hull에 의한 방법을 구현하여 실험을 수행하였다. Mahadevan과 Nagabushnam의 논문^[3]에서는, 구현된 알고리즘의 성능평가를 위해 문자열이 포함하는 단어의 수에는 관계없이 구한 간격을 큰 것부터 정렬한 후, 상위 (단어 수 - 1)개의 간격이 실제 영상에서 단어간 간격을 모두 포함하면 구성 단어를 바르게 분리한 것으로 간주하는 방식을 사용하였다. 이 방식을 사용할 경우, 상위 몇 개의 간격을 취할 것인지에 대한 명확한 기준이 없기 때문에 실제 존재하는 단어의 수보다 많은 경우에 대해 인식과 분리작업을 반복하는 가설검증 과정이 필요하게된다. 반면, 본 논문에서 제안하는 방법은 단어의 분리 작업 자체가 가설검증을 전제로 하지 않기 때문에 모아진 간격에 대한 정보의 분석을 통해 문자열에 포함된 단어의 수를 정하고 그에 따라 분리결과가 정확한지 여부를 판단하도록 평가방법을 보완하였다. Bounding-box와 convex-hull방법에서 단어간 간격여부 판별을 위한 임계치는 다음과 같은 과정을 통해 결정하였다.

- a. 구해진 연결요소간 간격으로부터 평균값(Avg_{gap})과 표준편차($StdDev_{gap}$)를 구한다.
- b. 임계치를 다음 식으로 계산한 뒤 임계치보다 큰 간격을 단어간 간격으로 정하여 단어를 분리한다.

$$Threshold = Avg_{gap} + \alpha \times StdDev_{gap}$$

위 식에서 사용되는 상수들은 학습용 문자열들을 이용하여 가장 최적의 단어 분리를 내도록 조절되었다. 여기서, α 값이 감소하면 초과 분할되는 정도가 증가하고, 따라서 구성 단어들의 분리 정도 또한 증가할 수 있게된다. 본 논문에서 제안하는 방법들과의 실험의 형평성을 유지하기 위해 초과 분할되는 정도를, 학습용 영상을 사용하여 α 를 통해 조절하였다. 표2의 상단에 두 방법(BB와 CH)에 의한 단어분리 능력을 정리하였다.

3. 제안하는 방식의 성능

제3절에서 설명한 방법을 통해 구현한 단어분리 방법들의 성능을 평가하기 위한 실험을 수행한다. 별도의 논문^[9]에서 발표된 방법과 비교해 볼 때, 그 논문에서는 두 개의 인접한 간격 및 높이가 사용되었으나, 본 논문에서 제안하는 방법은 시간지연 신경망의 특징을 활용하기 위해 그 비교 대상이 확대되었다. 실험대상 영상 또한, 우편물에서 취득된 주소영상 대신 이보다 일반적인 필기문장을 사용한다. 별도의 논문^[9]에서 제안하는 방법을 NN1이라 하고 본 논문에서 제안

하는 방법을 NN2라하여 두 방법의 성능을 각 서체별로 비교하여 정리하면 표2의 하단부와 같다.

표2에서 분리 능력을 단어와 문장으로 구분하여 측정하였다. 문장의 완전분리는 입력된 문장영상을 구성 단어수로 정확하게 분리한 경우이며 문장의 초과분리는 구성단어 수보다 많은 수로 분리한 경우이다. 초과 분리된 경우는 해당 분리 알고리즘이 제안하는 단어간 간격들이 실제 단어간 간격을 모두 포함하고 일부 단어들에 초과분리 된 경우이다. 단어분리 능력은 시험대상 문자열 영상이 포함하고 있는 모든 단어 수에 대해 바르게 분리된 단어 수의 비를 표시한다.

4. 결과 분석

표1에 나타난 것처럼 실험에 사용된 문자열 영상이 포함하고 있는 평균 단어 수는 7.1이다. 따라서, 문자열을 구성하는 단어들에 모두 완전히 분리되어야 하는 문장 분리능력과 단순히 단어의 분리 여부만을 계산하는 단어 분리능력에는 표에 나타난 것과 같이 차이가 있다.

초과분할의 경우들은 인식과정의 도움을 통해 완전히 해석될 수 있는 가능성이 있다. 이 경우에 초과분리 정도가 초과 분리능력과 증가되는 계산량에 영향을 미치게 된다. 예를 들어, 극단의 경우 모든 문자 세그먼트를 해석 가능한 단어간 간격으로 간주하면 초과 분리능력은 완전히 가깝게 증가하지만 인식과정에서 모든 조합들을 해석해야 하기 때문에 계산량은 급격히 증가하게 된다. 따라서 본 논문에서 수행된 실험에서 사용된 각 단어분리 방법들은 단어의 완전분리에 초점을 맞추어 초과분리정도가 1.02 ~ 1.03가 되도록 조절되었다.

연결요소간 간격의 해석에 의존하는 BB및 CH방법에 비해 문자 세그먼트들로부터 띄어쓰기의 단서가 될 수 있는 특징을 추출하여 신경회로망을 통해 해석하는 NN1 및 NN2방법의 단어 분리능력이 매우 우수함을 볼 수 있다. 특히, 문자 세그먼트로부터 특징들을 추출하는 적극적인 단어분리방법의 도입으로, 필기서술에서 쉽게 발견할 수 있는 과장된 꼬리와 같은 획에 의해 연결된 단어간 간격을 해석할 수 있게된다.

그림6에는 별도의 논문^[9]에서 제안하는 방법(NN1)으로 완전 분리할 수 없었으나 본 논문에서 제안하는 방법(NN2)으로 분리할 수 있었던 두 경우를 보여주고 있다.

표 2. 단어분리 능력비교

Table 2. Word segmentation performances.

방법	구분	cursive	discrete	mixed	평균
BB	문장 완전분리	23.56	67.70	46.52	40.84
	문장 초과분리	26.50	68.66	47.32	42.68
	단어 완전분리	38.50	84.27	69.20	58.86
	단어 초과분리	44.90	87.60	73.90	64.01
CH	문장 완전분리	24.54	69.30	48.51	42.25
	문장 초과분리	27.36	70.02	49.30	43.98
	단어 완전분리	38.60	85.21	69.63	59.26
	단어 초과분리	44.67	88.03	74.12	64.13
NN1	문장 완전분리	59.98	63.57	70.24	65.53
	문장 초과분리	83.05	83.13	87.43	84.39
	단어 완전분리	83.68	92.16	90.58	87.78
	단어 초과분리	85.92	94.26	92.29	89.82
NN2	문장 완전분리	59.63	76.20	73.57	67.64
	문장 초과분리	83.16	87.41	89.65	86.07
	단어 완전분리	84.54	93.01	91.64	88.73
	단어 초과분리	86.46	94.81	92.85	90.43

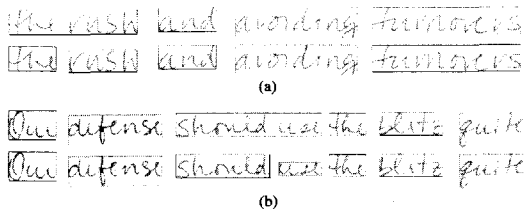


그림 6. 단어분리 결과 비교

Fig. 6. Word segmentation results

V. 요약 및 결론

제약 없이 쓰여진 영문 필기문자열로부터 인식의 도움 없이 단어를 분리하는 방법을 본 논문에서 제안하였다. 연결요소들 간의 기하학적인 거리 정보를 이용하여 단어를 분리하는 일반적인 접근방법 대신, 단어 분리를 위한 시각적인 단서를 문자열 영상으로부터 적극적으로 추출하여 시간 지연 개념을 갖는 신경회로망을 통해 해석하는 방법을 제시하였다. 이 방법은 연결요소간의 기하학적 거리정보만을 이용하는 종래의 방법에 대해 뛰어난 단어분리 능력을 보여줄 뿐 아니라, 이전의 단순한 신경회로망에 의한 방법에 대해서도 개선된 성능을 보여주고 있다.

컴퓨터의 처리 용이성 위주로 특징들을 추출하고 해석하던 종래의 방식과 달리, 인간의 시각적 해석 방식을 모방하려는 노력을 관련된 분야에서 쉽게 발견할 수 있다.^[15] 이와 같은 노력은 컴퓨터와 관련 기술의 발전에 따라 더욱 가속될 것으로 판단되며, 그 결과 이전에 구현하기 어려웠던 단서들을 찾아 여러 응용분야에의 적용을 가능하게 해준다. 본 논문에서 제안하는 단어분리 방법은, 전통적인 접근 방법에서 벗어나 시각적 해석방법을 도입하여 실제적인 문자열 영상에 만족할 만한 정확도를 내도록 설계되었다.

문자분리 단계의 결과를 단어분리를 위해 활용하였는데, 단어분리의 목적이 단어의 바른 인식을 위함이라고 생각해 볼 때, 단어의 인식을 위해 필요한 과정인 문자분리 단계를 미리 수행하여 단어분리와 인식에 모두 사용함으로써 추가적인 계산과 이에 따른 처리시간을 최소화 할 수 있다. 단어인식을 위한 문자분리가 주요 세로획을 기준으로 수행되는 것이 일반적인 접근 방식임으로, 제안하는 방법은 대다수의 영어·유럽언어권의 문자열 처리에 일반적으로 사용될 수 있다.

참 고 문 헌

- [1] M.-Y. Chen, A. Kundu, and J. Zhou, "Off-Line Handwritten Word Recognition Using a Hidden Markov Model Type Stochastic Network," *IEEE Trans. on Pat. Anal. and Machine Intel.*, vol. 16, no. 5, pp. 481-496, May 1994.
- [2] G. Seni and E. Cohen, "External Word Segmentation of Off-Line Handwritten Text Lines," *Pattern Recognition*, vol. 27, no. 1, pp. 41-52, 1994.
- [3] U. Mahadevan and R. C. Nagabushnam, "Gap Metrics for Word Separation in Handwritten Lines," in *3rd Int. Conf. on Document Anal. and Recog.*, Montreal, Canada, pp. 124-127, 1995.
- [4] U. Mahadevan and S. N. Srihari, "Hypotheses Generation for Word Separation in Handwritten Lines," in *5th Int. Workshop on Frontiers in Handwriting Rec.*, UK, pp. 453-456, Sep. 1996.
- [5] S. H. Kim, "Handwritten Korean address recognition by combining minimum distance classification and dictionary look-up," in *Proc. of Sixth Int. Workshop on Frontiers in Handwriting Recog.*, Taejon, Korea, pp. 436-448, Aug. 1998.
- [6] U.-V. Marti and H. Bunke, "Towards General Cursive Script Recognition," in *Proc. of Sixth Int. Workshop on Frontiers in Handwriting Recog.*, Taejon, Korea, pp. 379-388, Aug. 1998.
- [7] K. M. Sayre, "Machine Recognition of Handwritten Words: A Project Report," *Pattern Recognition*, vol. 5, pp. 213-228, 1973.
- [8] J. C. Simon, O. Baret, and N. Gorski, "A System for the Recognition of Handwritten Literal Amounts of Checks," in *Proc. of the Conf. on Document Anal. Sys.*, Kaiserslautern, Germany, pp. 135-155, 1994.
- [9] G. Kim and V. Govindaraju, "Handwritten Phrase Recognition as Applied to Street

- Name Images," *Pattern Recognition*, vol. 31, no. 1, pp. 41-51, 1998.
- [10] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. 37, no. 3, pp. 328-339, 1989.
- [11] S. Haykin, *Neural Networks - A Comprehensive Foundation*. New Jersey: Macmillan College Pub. Co., Inc., 1994.
- [12] G. Kim and V. Govindaraju, "Efficient Chain Code Based Image Manipulation for Handwritten Word Recognition," in *Proc. of the SPIE symposium on elect. imaging sci. and tech.*, San Jose, CA, vol. 2660, pp. 262-272, Feb. 1996.
- [13] C. Farouz, M. Gilloux, and J.-M. Bertille, "Handwritten Word Recognition with Contextual Hidden Markov Models," in *Proc. of Sixth Int. Workshop on Frontiers in Handwriting Recog.*, Taejon, Korea, pp. 133-142, Aug. 1998.
- [14] IPVR, University of Stuttgart, *Stuttgart Neural Network Simulator v4.1*, 1995.
- [15] G. Lorette, "Handwriting recognition or reading? Situation at the dawn of the 3rd millennium," in *Proc. of Sixth Int. Workshop on Frontiers in Handwriting Recog.*, Taejon, Korea, pp. 1-13, Aug. 1998.

 저자소개

金敬煥(正會員) 第36卷 C編 第4號 參照

현재 서강대학교 전자공학과 조교수