

문자 인식을 위한 신경망컴퓨터에 관한 연구

(A Study on the Neural Network for the Character Recognition)

李 昌 基,* 田 炳 實*

(Chang Ki Lee and Byoung Sil Chon)

要 約

본 논문은 필기체 문자 패턴의 학습 및 인식을 위한 신경망 컴퓨터에 관한 연구를 하고자 한다. 입력문자의 전처리는 복합 CELL의 방향성 필터링을 통해서 문자의 윤곽만을 추출하며 이를 정규화 한 후 패턴분류기인 ART에 입력된다. ART는 하향패턴 매칭에 의해 스스로 안정된 학습을 할 수 있는 신경망이며 안정된 학습이 이루어진 후에는 입력에 대한 인식작용이 자동으로 이루어지며 복잡한 구조의 새로운 입력패턴에도 그 인식시간은 증가되지 않는다. 또한 학습된 패턴과 새로 입력된 패턴과 유사도의 정도를 나타내는 Vigilance값을 변화시켜서 좀더 유연한 문자인식을 하도록 하였으며 실제로 쓴 문자들을 입력시키고 여기에 대한 인식을 컴퓨터 시뮬레이션을 통해서 확인하였다.

Abstract

This paper proposed a neural computer architecture for the learning of script character pattern recognition categories. Oriented filter with complex cells preprocess about the input script character, abstracts contour from the character. This contour normalized and inputed to the ART. Top-down attentional and matching mechanisms are critical in self-stabilizing of the code learning process. The architecture embodies a parallel search scheme that updates itself adaptively as the learning process unfolds. After learning ART self-stabilizes, recognition time does not grow as a function of code complexity. Vigilance level shows the similarity between learned patterns and new input patterns. This character recognition system is designed to adaptable. The simulation of this system showed satisfied result in the recognition of the hand written characters.

I. 序 論

기존의 문자인식 시스템은 입력된 문자들의 특성 벡터를 추출한 다음 판별함수 등을 이용하여 그 특성 벡터가 통계적인 분포상에서 어느 영역에 속하는

가를 알아냄으로써 입력문자를 인식하는 방법이다. 그러나 활자체의 종류와 크기에 따라 모델문자의 특성 벡터가 달라져야 하므로 여러가지 활자체로 구성된 문서의 문자인식에는 적합하지 않다. 또한 현재의 문자인식 시스템은 폰노이만형 컴퓨터의 프로그램 등에 의해 구성되므로 유연성이 결여되고, 대량의 복잡한 문자처리가 요구될 때 안정된 실시간 처리가 곤란하다. 이러한 문제들을 해결하며 고도의 지능을 갖는 문자인식용 신경회로 컴퓨터를 생각할 수 있는데 문자인식의 대표적인 신경망 구조는 후쿠시마의 다층신경망 구조인 Neocognitron^[1]으로서 필기체 숫

*正會員, 全北大學校 電子工學科

(Dept. of Elec. Eng., Chonbuk Nat'l Univ.)

接受日字: 1992년 4월 10일

※ 이 논문은 1990년도 教育部 學術研究助成費에 의하여 研究되었음)

자 인식에서는 높은 인식률을 보이고 있으나 실용화에는 아직까지 많은 어려움이 있다. KAIST, 포항공대, 연세대 전자과등 많은 대학에서 인쇄체 및 필기체 인식 연구가 활발히 진행되고 있으며 최근에는 이를 상용화한 제품이 나오고 있는 실정이다.

본 논문에서는 문자인식 분야에서 많이 시도 되지 않고있는 Grossberg의 ART(adaptive resonance theory)^{[2],[3],[4]}를 최종 패턴 분류기로 사용하며 전처리(preprocessing)는 Carpenter가 제안한 방향성 필터 알고리즘^[11]을 변형하여 입력 패턴의 잡음과 뒤틀림을 제거하고 정규화시키는 문자인식 시스템을 설계하고 이를 시뮬레이션 하고자 한다.

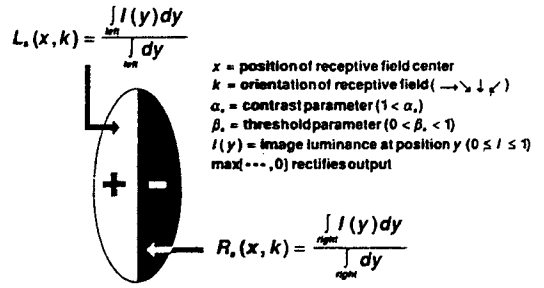


그림 1. 단순 cell
Fig. 1. Simple cell.

II. 本 論

1. 경계선 추출 및 정규화

입력 패턴의 윤곽을 추출하기 위해서 특정 방향에만 민감하게 반응하며 병렬로 동작하는 방향성 필터의 단위 cell의 출력은 다음과 같은 식으로 나타낼 수 있다.

$$\max[L_s(x, k) - \alpha_s R_s(x, k) - \beta_s, 0] \quad (1)$$

여기에서 x는 입력 cell의 중앙 위치(position)이며 k는 cell의 방향인데 이는 0°, 45°, 90°, 135°이다. α_s ($1 < \alpha_s$)는 contrast parameter, β_s ($0 < \beta_s < 1$)는 threshold parameter, $L_s(x, k)$ 는 입력 cell의 좌반부 activation, $R_s(x, k)$ 는 우반부 activation값을 의미하고 $I(y)$ 는 cell내부의 위치 y에서 image luminance 값을 의미하며 좌, 우반의 평균 luminance값은 다음 식과 같으며 그림 1에 simple cell의 모습을 보였다.

$$L_s(x, k) = \frac{\int_{left\ half} I(y) dy}{\int_{left\ half} dy} \quad (2)$$

$$R_s(x, k) = \frac{\int_{right\ half} I(y) dy}{\int_{right\ half} dy} \quad (3)$$

다음으로 좌반부 및 우반부 activation값을 합성한 복잡 cell을 생각해 보자

$$C_s(x, k) = \max[L_s(x, k) - \alpha_s R_s(x, k) - \beta_s, 0] + \max[R_s(x, k) - \alpha_s L_s(x, k) - \beta_s, 0] \quad (4)$$

식 (4)에 의해서 잡음이 섞인 입력 패턴은 잡음이 제거되고 각 방향에서 복잡 cell의 출력이 가장큰 값 방향만이 activation되고 이들을 연결하면 윤곽만을 추출 가능하게 되며 complex cell의 모습을 그림 2에

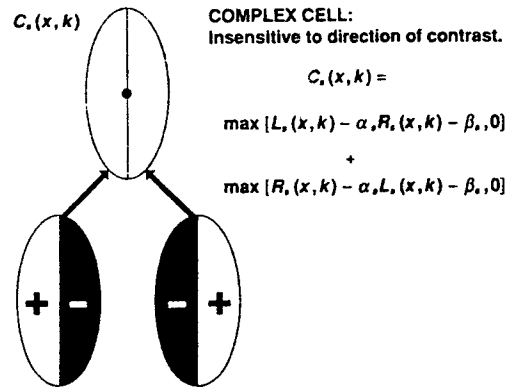


그림 2. 복합 cell
Fig. 2. Complex cell.

보였다.

입력 패턴은 그 크기와 모멘트의 방향이 다양하게 입력되므로 이들의 윤곽선 추출후 정규화 작업을 해야한다. 하나의 표준 window를 설정하고 입력문자의 무게 분포, 최대폭과 최소폭을 계산한 후 설정한 무게가 큰쪽이 아래를 향하도록 해서 설정한 window 크기에 맞게 확대하거나 축소한다.

2. 문자의 분류

입력문자의 분류에 적용하는 신경회로 모델은 Grossberg가 제안한 비지도 학습(unsupervisor learning) 모델이다. ART의 구조는 그림 3과 같다.

ART는 입력 layer인 F_1 과 coding layer인 F_2 로 구성되며 두개의 층 사이에는 지속기억열(LTM: long term memory)이라고 하는 adaptive filter가 존재한다. F_1 층에는 1에서 M까지, F_2 층에는 M+1에서 N까지의 노드가 존재한다. 이들 노드들은 신경세포 하

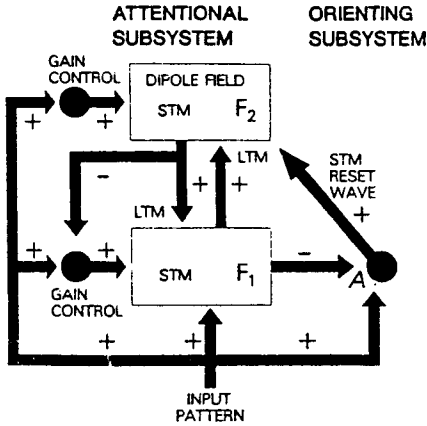


그림 3. ART 구조
Fig. 3. Structure of ART.

나하나에 해당하며 이들을 단기 기억열(STM : short term memory) x_i, x_i 라고 부르고 방정식은 각각(5), (6) 과 같다.

$$\epsilon \frac{d}{dt} x_i = -x_i + (1 - A_1 x_i) J_i^* - (B_1 + C_1 x_i) J_i \quad (5)$$

$$\epsilon \frac{d}{dt} x_i = -x_i + (1 - A_2 x_i) J_i^* - (B_2 + C_2 x_i) J_i \quad (6)$$

F_1 층의 i 번째 노드 v_i 에 가해지는 자극신호 J_i^* 는 필터링과 정규화 과정을 거친 상향 입력 패턴 I_i 와 학습된 하향 기대 패턴 V_i 와의 합이다.

$$V_i = D_1 \sum_j f(x_j) z_{ji} \quad (7)$$

$$J_i^* = I_i + V \quad (8)$$

여기에서 $f(x_i)$ 는 F_2 층 v_i 노드의 활성값 x_i 에 의해 발생하는 신호이고 z_{ji} 는 v_j 에서 v_i 로 가는 하향 지속기억열 값이다.

억제 신호 J_i 는 비교회로의 이득(attentional gain control)을 제어한다.

$$J_i = \sum_k f(x_k) \quad (9)$$

즉 F_2 층의 노드가 비활성일 때 $J_i = 0$ 이며 활성일 때는 $J_i > 0$ 이다. 따라서 식 (6)의 J_i 는 F_1 층의 모든 단기 기억열에 일정량의 억제신호를 가한다. F_2 층의 단기 기억열 방정식에 의해 F_1 층으로부터 가장 큰 입력을 받은 노드가 승자(winner)로 선택되며 F_2 층의 그 노드 v_i 에 입력되는 J_i^* 와 J_i 는 다음과 같다. 입력 J_i^* 는 v_i 로부터 양(positive)의 값으로 자기 자신에게 되돌려지는 신호 $g(x_i)$ 와 상향 적응필터를 통해

출력된 값 T_i 의 합이다.

$$T_i = D_2 \sum_l h(x_l) z_{li} \quad (10)$$

$$J_i^* = g(x_i) + T_i \quad (11)$$

여기에서 $h(x_i)$ 는 F_1 의 노드 v_i 에 의해 발생하는 신호이고 z_{li} 는 v_l 에서 v_i 로 가는 상향 지속기억열 값이다. 억제신호 J_i 는 F_2 층의 노드들에 의해서 v_i 노드에 음(negative)의 값으로 되돌려지는 신호 $g(x_k)$ 의 합이다.

$$J_i = \sum_{k=i} g(x_k) \quad (12)$$

즉 식 (11)에서 양의 되돌림신호 $g(x_i)$ 와 식 (12)에서 음의 되돌림 신호 J_i 는 입력 패턴 T 에 반응해서 F_2 의 단기 기억열에 기억된 패턴과의 대비효과를 증가시키는 작용을 한다. F_2 의 매개변수들이 적절하게 선택되었을 때 이러한 대비효과는 상승해서 가장 큰 T_i 입력신호를 받은 v_i 노드만이 선택될 수 있게 해준다.

F_1 층의 v_i 노드와 F_2 층의 v_i 노드 사이의 상향 지속 기억열 z_{ii} 와 하향 지속기억열 z_{ii} 의 방정식은 식(13), (14)와 같이 쓸 수 있으며 어떻게 학습과 인식에 관계하는지 살펴보자.

$$\frac{d}{dt} z_{ii} = K_2 f(x_i) [x_i [-E_{ii} z_{ii} + h(x_i)]] \quad (13)$$

$$\frac{d}{dt} z_{ii} = K f(x_i) [(1 - z_{ii}) L h(x_i) - z_{ii} \sum_{k=1} h(x_k)] \quad (14)$$

식 (14)에서 F_2 층의 노드 v_i 에 의한 신호 $f(x_i)$ 의 값이 양이고 F_1 층의 노드 v_i 로부터 양의 신호를 받을 경우 지속기억열 z_{ii} 의 연산은 $(1 - z_{ii}) L_n(x_i) K f(x_i)$ 의 비율로 변화한다. 노드 v_i 에 동시에 입력되는 신호들($h(x_i), k=i$)은 경쟁수식 항($-z_{ii} h(x_i) K f(x_i)$)들의 작용을 통해 이미 학습에 참여한 적이 있는 z_{ii} 들 간의 경쟁을 유도한다. 달리 말하면 학습에 참여한 적이 있는 z_{ii} 들은 총경쟁입력($-\sum h(x_k), k=1$)과 F_2 층의 노드 v_i 에 의한 신호이득 $K f(x_i)$ 로 이루어지는 식($-z_{ii} \sum h(x_k) K f(x_i), k=i$)에 비례해서 감소한다.

만약 입력패턴과 여기에 의해 활성화(activation)된 node에 기억된 패턴과의 일치 정도는 식 (15)에 의해 주어지며 식(15)를 만족하면 그림 3의 A에 의해 현재의 입력패턴을 버리고 다시 새로운 패턴을 입력받아 학습과 인식을 위한 탐색작업을 시작한다.

$$\frac{|X|}{|I|} < \rho \quad (15)$$

여기에서 I는 현재의 입력 패턴이고 |X|는 그 입력 패턴에 따라 F₁층에서 활성화된 x_i들 중 x_i > 0인 노드들의 갯수이다.

Ⅲ. 實驗方法 및 考察

1. 문자인식 시스템의 구성

신경회로망의 시뮬레이션은 PC-386 상에서 이루어졌으며 입력패턴은 Video camera와 Matrox 사의 MVP/AT(COLOR FRAME GRABBER)를 사용하여 손으로 쓴 문자의 gray영상을 얻는다. 시스템의 구성도를 그림 4에서 보이고 있다.

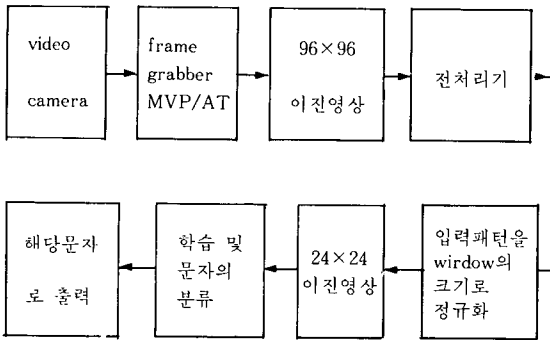


그림 4. 시스템 구성도
Fig. 4. Block diagram of the system.

입력문자는 대학원생 21명의 필기체 문자 영상을 기초로 하고 복합 cell을 사용해서 잡음을 제거한 문자의 윤곽을 추출한 후 입력문자의 중심을 계산하고 이를 96×96의 Window 중앙에 위치 시킨다. 입력된 문자는 입력할 때 마다 그 크기, 기울기, 가로, 세로 비율이 다를 것이므로 문자의 폭과 높이를 산출하고 여기에 비례하는 값을 주어서 문자를 확대한 다음 Window안에 완전히 들어가도록 한다.

시뮬레이션에서 사용된 F₁층의 노드수는 24×24=576개이므로 이 정규화된 패턴을 24×24의 격자구간으로 나누어 이진패턴을 얻었다. F₂층의 노드 수는 75개로 했다. 이렇게 정규화된 패턴이 인식과 학습을 위한 패턴이 된다.

그림 5는 방향성 필터를 사용해서 대문자 A와 소문자 a를 필터링한 모습이다.



그림 5. 추출된 윤곽패턴
Fig. 5. Extracted contour pattern

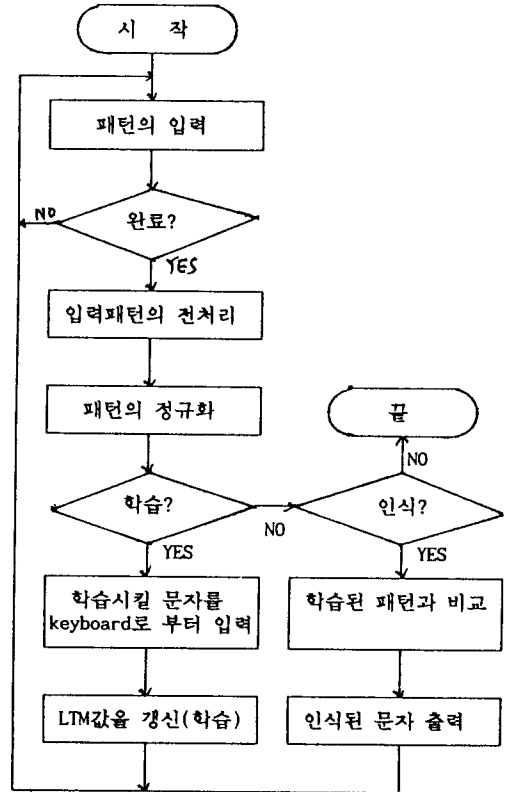
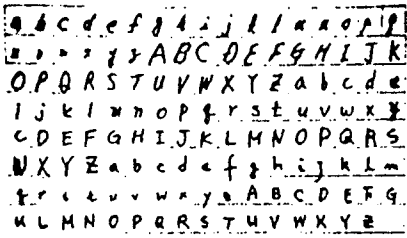


그림 6. 시뮬레이션의 흐름도
Fig. 6. Flowchart for the simulation.

2. 입력문자의 학습 시뮬레이션

Camera로부터 입력된 gray level의 frame을 모니터에 display한 후 학습시키고자 하는 문자를 마우스로 선택하며 이때 mask window는 96×96이다.

시뮬레이터는 모든 전처리 과정을 수행한 후 user에게 방금 입력한 문자를 어떤 문자로 학습시킬 것인지 물어오고 이때 해당문자를 keyboard로 부터 입력한다. 그림 6은 시뮬레이션의 흐름도를 보여준다. 여기에서 LTM은 F₁층 노드와 F₂층 노드 사이의 상



(a) 입력 패턴



(b) $\rho=0.5$ 일때 인식 패턴



(c) $\rho=0.8$ 일때 인식 패턴



(d) $\rho=0.95$ 일때 인식 패턴

그림 7. ρ 값에 대한 인식을

Fig. 7. Recognition rate for the ρ value.

향지속기억열 z_{it} 와 하향 지속기억열 z_i 의 값의 의미한다.

입력된 문자가 어떻게 상위 노드에 학습되는지 알아보자. 학습패턴의 입력전에 신경망의 각 LTM 및 STM은 초기화 되어 있으며 최초의 입력패턴에 반응하는 상위 노드들의 출력 값은 동일하다. 따라서 상위의 첫 노드 v_M 을 학습 노드로 선정하고 여기에 입력패턴을 학습시킨다. 학습규칙에 따라서 상향 지속 기억열은 $z_{iM} = Lh(x_i) / [(Lh(x_i) + \sum Lh(x_i))] : k=i$ 의

값으로 변한다. 즉 x_i 의 활성노드 수가 많으면 z_{iM} 의 값은 여기에 반비례 하므로 스스로 normalization이 이루어 진다. 이때 하향 지속 기억열은 $z_{M1} = h(x_i)$ 값으로 대치되어 입력된 패턴의 모습을 저장하고 있게 된다. 두번째 문자가 입력되면 각 노드들은 입력에 따라 자신들의 상향 LTM값에 따라서 각기 다른 값으로 활성화 될 것이다. 여기에서 학습에 참가했던 F_2 층의 v_M 노드는 가장 큰값을 출력할 것이 자명하며 최대 출력을 내는 v_M 노드는 입력패턴과 자신이 기억하고 있는 패턴과의 matching관계를 확인하기 위해서 하향 지속기억열이 가지고 있는 패턴을 F_1 층으로 내려 보낸다. F_1 층은 현재의 입력패턴과 F_2 층에서 전달된 입력 패턴중 서로 일치하는 부분들만을 활성화 시켜 비교기의 ρ 값에 따라 동작하도록 한다. 즉 $\rho=1.0$ 일때 두개의 패턴이 100%일치할 때만 비교기 A가 F_2 층에 reset signal을 보내지 않게 된다. 두번째 문자 패턴은 첫번째 패턴과 다르며 따라서 v_M 은 reset되고 v_{M+1} 이 활성화 된다. v_{M+1} 은 처음으로 활성화 되었고 학습되지 않은 노드이므로 현재의 입력패턴을 그대로 여기에 저장한다. 이와 같이 비교값 ρ 에 따라서 입력패턴이 분류되는 종류는 변한다. $\rho=0.95, 0.8, 0.5$ 의 세가지 경우에 있어서 입력된 문자가 어떻게 인식되는지 알아보자. 그림 7 a)는 입력패턴 그림 6 b), c), d)는 $\rho=0.95, 0.8, 0.5$ 일때 각 경우의 인식패턴을 보여주고 있다.

ρ 값의 변화는 분류도를 변화시키고 있기 때문에 세밀한 분류를 필요로 할 때는 ρ 값을 크게하면 된다. 이때의 문제점은 입력패턴 분류를 위해서 더 많은 출력노드를 요구한다. 때문에 상황에 따라 적절한 ρ 값을 선정할 필요가 있다. 물론 같은 종류의 입력 문자가 각각 다른 노드에 분류된다. 하더라도 각 노드는 같은 종류의 문자를 인식문자로서 출력하거나 학습하기 때문에 문제되지 않는다.

본 실험의 결과는 표 1에서 잘 보여주고 있으며 대문자(A, B, ... Z) 26종 546자를 입력해서 학습시키면 $\rho=0.5$ 일때 32종으로 분류한다.

표 1. ρ 값의 변화에 따른 분류도의 변화

Table 1. Variation of the classification rate for ρ value.

| ρ | $\rho=0.5$ | $\rho=0.8$ | $\rho=0.95$ |
|----------------|------------|------------|-------------|
| 입력문자 | | | |
| 대문자(26종, 546자) | 32종 | 48종 | 62종 |
| 소문자(26종, 546자) | 36종 | 53종 | 68종 |
| 숫자(10종, 210자) | 12종 | 21종 | 33종 |

인식은 이렇게 학습된 패턴을 가지고 이루어지며 숫자의 경우 $\rho=0.8$ 에서 94% 정도의 인식률을 보이며 결과는 표 2에서 보였다.

표 2. ρ 값의 변화에 따른 인식률 변화

Table 2. Variation of the recognition rate for ρ value.

| 인식률 입력문자 | $\rho=0.5$ 인식률 | $\rho=0.8$ 인식률 | $\rho=0.95$ 인식률 |
|----------------|-------------------|-------------------|--------------------|
| 대문자(26종, 546자) | 85% | 93% | 89% |
| 소문자(26종, 546자) | 83% | 92% | 87% |
| 숫자(10종, 210자) | 87% | 94% | 90% |

IV. 結 論

본 연구에서는 지금까지 패턴 인식에 있어서 자율경쟁 신경회로망 모델의 가장 큰 문제점이었던 적응성과 안전성 문제를 해결하기 위해 ART모델을 이용하고 이에 적절한 전처리를 통해서 필기체 문자 인식을 수행한 결과, 패턴의 처리과정에서 사람의 경험에 의한 판단과 결정과정이 개입되지 않고 자율적으로 패턴을 처리하는 시스템을 구현할 수 있음을 확인하였다. 또한 특징의 수가 많고 패턴의 종류가 많을 때 발생할 수 있는 시스템의 불안정성을 상당히 줄일 수 있고 분류된 클러스터의 증가에 따르는 인식률의 저하가 거의 없음을 확인할 수 있었다. 현재의 시스템 상에서 이루어진 시뮬레이션은 상당한 효율을 보이고 있지만 사용가능 메모리의 한계 때문에 입력과 출력단에 있어서 한정된 노드수로 이루어 졌고 또한 처리속도상의 문제 때문에 충분한 시뮬레이션의 어려움이 있었다. 위에서의 문제를 해결할 수 있는 고성능 컴퓨터의 사용이 필요하다 하겠다.

본 연구에서의 패턴 인식 시스템은 시각 패턴 뿐만 아니라, 음성의 스펙트럼 정보를 기초로 하는 패턴의 인식 즉 음성 인식등에 응용할 수 있으리라 기대되며 신경망 이론을 이용한 실용적인 패턴인식 시

스템이 되기 위해서는 보다 신경세포와 유사한 배열로 그 칩, 즉 신경망 컴퓨터의 하드웨어에 관한 연구가 병행해서 이루어져야 할 것이다.

參 考 文 獻

- [1] K. Fukushima and S. Miyake "A neural network model for a mechanism of visual pattern recognition" *IEEE Transactions on Systems, Man, and Cybernetics* 13(5), pp 826-834, 1983.
- [2] S. Grossberg, editor: *Neural Networks and Natural Intelligence*, MIT Press, 1988.
- [3] S. Grossberg : *Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors* *Biological Cybernetics* 23:121-134, 1976.
- [4] S. Grossberg : *How does abrain build a conitive code?* *Psycho logical Review* 87:1-51, 1980.
- [5] Anderson and Rosenfeld deitors: *Neurocomputings*, MIT Press, 1988.
- [6] V. Verma, *Neural Networks: Artificial Neural Networks: theoretical Concepts*, Computer Society Press, 1988.
- [7] Gupta and M.R. Saych, "Neural networks for plane shape classification," *ICASSP*, vol. 2, pp. 936-939, 1988.
- [8] R. Lippman "An introduction to computing with neural net," *IEEE ASSP Mag.* pp. 422, Apr. 1987.
- [9] R.F. Thompson. "The Brain," W.H. Freeman and Co., U. S. A. 1985.
- [10] Gail A. Carpenter "Neural network models for pattern recognition and assoiative memory" *neural networks*, vol. 2, pp. 243-257, 1989.
- [11] Gail A. Carpenter "Invariant recognition of cluttered scenes by a self-organizing ART architecture: CORT-X boundary segmentation" *Neural Networks*, vol. 2, pp. 169-181, 1989.

著 者 紹 介



李 昌 基 (正會員)

1961年 7月 15日生. 1988年 2月
전북대학교 전자공학과 졸업.
1990年 2月 전북대학교 대학원
전자공학과(석사) 1991年~현재
전북대학교 대학원 전자공학과
박사과정

田 炳 實 (正會員) 第27卷 第10號 參照

현재 전북대학교 전자공학과
교수