

논문 2009-5-29

순차 램 기반 누적 신경망을 이용한 수화 인식

Sign Language recognition Using Sequential Ram-based Cumulative Neural Networks

이동형*, 강만모**, 김영기***, 이수동**

Dong-Hyung Lee, Man-Mo Kang, Young-Kee Kim, Soo-Dong Lee

요 약 가중치가 없는 램 기반 신경망은 가중치를 재조정하는 기존 신경망에 비해 계산량 및 인식 시간이 적은 장점을 가지고 있다. 특히 연속적인 연관성을 갖는 제스처와 같은 행위 정보는 각각의 정보들이 시계열적 상관관계를 갖는다. 이와 같은 행위 정보를 인식하려면 일반적으로 많은 계산량과 처리 시간이 요구된다. 이런 문제점을 해결하기 위해 일반적으로 전처리 과정의 삽입 및 하드웨어 인터페이스 활용 등을 이용한다. 본 논문에서는 이와 같은 추가적인 방법 없이 순차 램 기반 누적 신경망으로 연속적인 행위 정보인 한글 복합어 수화 인식 시스템을 구현하였다. 제안된 모델의 성능을 검증하기 위하여 카메라로부터 입력받은 연속적인 복합어 수화 영상을 최소한의 이미지 처리인 경계선 검출만으로 수화 인식을 실험하였다. 경계선 검출 후 이진 영상을 전처리 과정 없이 제안된 순차 램 기반 누적 신경망 시스템으로 처리된 결과는 93%의 인식률을 얻었다.

Abstract The Weightless Neural Network(WNN) has the advantage of the processing speed, less computability than weighted neural network which readjusts the weight. Especially, The behavior information such as sequential gesture has many serial correlation. So, It is required the high computability and processing time to recognize. To solve these problem, Many algorithms used that added preprocessing and hardware interface device to reduce the computability and speed. In this paper, we proposed the Ram based Sequential Cumulative Neural Network(SCNN) model which is sign language recognition system without preprocessing and hardware interface. We experimented with using compound words in continuous korean sign language which was input binary image with edge detection from camera. The recognition system of sign language without preprocessing got 93% recognition rate.

Key Words : SCNN, CSWNN, Compound Words of Recognition, Sign Language Recognition

I. 서 론

미래 고도정보화사회는 제스처 및 얼굴/음성 등의 행위 정보를 처리하는 기술이 인간의 편리함을 위하여 많은 각광을 받을 것이다. 이를 위하여 인체가 나타내는 행위 정보를 인식 및 처리 등을 하기 위하여 정보를 보다

자연스럽고, 유연하게 조작하여 원하는 신호로 표현되어야 하며, 또한 가장 큰 특징인 순차적인 흐름도 같이 표현되어야 한다.

생체 인식 분야는 대표적인 영상 패턴 인식 분야에 속하며, 여러 가지 종류의 응용 시스템이 개발되었다. 최근에는 지문 및 홍채 등을 인식하는 기기들이 상용화되어 시판되었으며, 또한 청각 장애인을 위한 수화를 인식하는 정보통신기기 개발에 활발하게 연구가 진행되고 있다. 지문 및 홍채와 같은 정지영상은 이상적이면서 단편적인

*정회원, 한국폴리텍 VII대학 울산캠퍼스 정보통신시스템과

**정회원, 울산대학교 컴퓨터정보통신공학부

***정회원, 울산과학기술대학교 컴퓨터정보학부

접수일자 2009.9.12, 수정일자 2009.10.8

원형 패턴을 이용하여 유사 패턴들을 인식한다. 그러나 시계열성을 갖는 제스처 및 수화과 같은 동영상은 순차적인 행위 정보 패턴으로 인식되어야 한다. 이는 하나의 정지 영상 패턴을 인식하는 것과 여러 장의 연속된 정지 영상을 한 묶음 단위로 인식하는 차이와 같다. 연속된 여러 장의 정지 영상 패턴 단위로 학습 및 인식을 하면 시간과 계산량 증가를 초래할 것이다. 만약 일반적인 패턴 인식으로 행위 정보를 분해하여 처리된다면 각각의 영상 패턴에 대한 연관성을 찾아 연결하여 처리되어야 하며, 이는 별도의 다른 프로세서가 추가되어야 한다.

패턴 인식 분야에서 대표적인 모델로는 원형 정합을 이용한 템플릿 매칭 방법, 확률 통계적인 방법, 구조적 방법, 신경회로망 등이 있으며^[1], 그 중에서 신경회로망이 성능이 뛰어나다고 알려져 있다. 연결 강도인 가중치를 가지는 기존 신경망(Weighted Neural Network)은 망에 입력되기 전에 복잡한 전처리 과정에 의해 많은 변형이 이루어지고, 또한 가중치 재조정에 의해 반복 계산이 많이 이루어진다. 이는 동영상과 같은 연속적인 입력 패턴들을 실시간으로 처리하기 위해서는 비효율적이다. 이런 기존의 신경회로망의 문제점을 해결하기 위해서 Aleksander와 Stonham이 제안한 램 기반의 WNN(Weightless Neural Net, BNN)^[2]는 하드웨어 구현이 용이하면서 복잡한 전처리 과정이 없이 한 번의 교육으로 학습이 이루어지는 모델을 제안하였다. WNN의 이런 장점에도 불구하고 메모리 포화 문제 및 추가 반복 학습과 일반화 패턴의 특징점 추출 어려움 등의 단점을 가지고 있었다. 추가 반복 학습 및 용이한 일반화 패턴 추출을 할 수 있는 경험유관 축적 신경망(Experience Sensitive Cumulative Neural Network : ESCNN)^{[3][4]}과 동적 3차원 뉴로 시스템(3-D Neuro System: 3DNS)^[5] 모델이 제안되었다. 그리고 연속적인 교육 패턴을 입력할 수 있는 램 기반 신경망인 Single Layer Sequential RAM-based neural network(SWNN)^{[6][7][8]}과 Cascaded sequential weightless neural network(CSWNN)^[6]으로 형식언어 인식을 위한 모델이 제안되었다.

본 논문에서는 3DNS를 개선하여 연속적인 교육 패턴을 인식하면서 추가 및 반복 학습 그리고 일반화 특징점을 추출할 수 있는 램 기반 신경망(Sequential Ram-based Cumulative Neural Network : SCNN) 모델을 제안한다. SCNN은 연속적인 입력 패턴을 처리하기 위하여 카메라로부터 입력 받은 영상을 최소한의 전처리

과정인 에지 검사로 한글 복합어 수화를 인식하였다.

제안된 시스템을 평가하기 위하여 복합어 수화를 임의로 100개 추출^[3]하여 학습과 인식에 사용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 램 기반 신경망인 3DNS와 순차적인 신경망인 SWNN, CSWNN에 대하여 설명하고, 3장에서는 제안된 순차적 누적 신경망인 SCNN 설명하면서 4장은 제안한 모델을 한국어 복합어 수화 인식으로 실험 및 성능을 측정하였다. 마지막으로 5장에서는 결론 및 향후 방향에 관하여 설명한다.

II. 3DNS 및 CSWNN

1. 동적 3차원 뉴로 시스템(3DNS)

가. 3DNS 신경세포

동적 3차원 뉴로 시스템은(3DNS)의 신경세포는 그림 1과 같이 N Bit 주소 정보로 2^N bit 크기의 이진 신경세포^[3]를 Byte 단위로 확장하여 2^N Byte 크기인 RAM으로 구성되어 있다. 출력층에서 기억층으로 귀환 회로를 사용하여 신경세포 내부에 기억된 정보를 귀환하여 연산할 수 있도록 하였다.^[4] 이런 구조는 기존의 교육된 정보를 유지하면서 새로운 정보에 대하여 추가 교육이 가능하고 동일한 범주에 대한 반복 교육을 가능하다. 그리고 추가 및 반복 교육을 통하여 일반화 패턴을 생성할 수 있다.

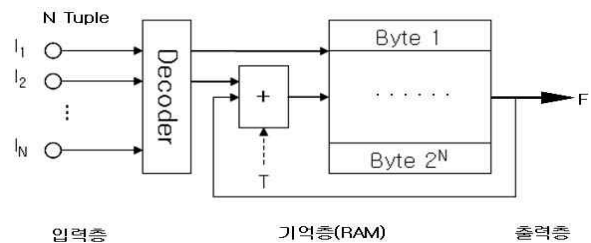


그림 1. 3DNS 신경세포.
Fig. 1. The Cell of 3DNS Neuron

학습 단계는 0과 1로 이진화된 입력 패턴으로부터 N-tuple 기법을 사용하여 N개의 노드를 램덤하게 샘플링하여 디코더기에 식 (1)과 같은 이진수 주소 정보 형태로 입력된다.^{[3][4]} 선택된 주소는 원형 패턴 값인 T값을 저장한다. T값은 원형 패턴의 0과 1의 픽셀값으로 누적한다.^[5]

$$\text{Address} = 2^N + 2^{N-1} + \dots + 2^1 \quad (1)$$

인식 단계는 학습 과정과 같이 N-tuple기법을 사용하며, 학습 때 사용된 램덤 정보를 이용하여 디코더기에 의해 만들어진 주소의 해당 램 값들을 가져온다. 가져온 램 값들을 원형 패턴과 매칭하여 값을 생성한다. 생성된 값이 일정한 값 즉 임계값 이하이면 다시 귀환 연산을 사용하여 만족할 만한 결론이 얻을 때까지 반복 수행한다.^[5]

나. 3차원 다중 판별자 및 MRD

한 개의 범주로 학습된 정보를 저장하는 신경세포들은 하나의 범주로 입력되면서 램으로 구성된 한 개의 판별자를 만든다. 다양한 범주에서 인식하기 위해서는 다중 판별자를 사용 하며, 이를 사용한 예로는 WISARD(Wilkie, Stonnen and Aleksander's Recognition Device)^[10] 시스템과 경험 유관 누적 이진신경망^[3] 등이 있다. 그림 2는 3DNS의 다중 판별자를 보여주고 있다. 학습 단계에서 원형 이미지와 학습 이미지를 N-tuple 방법에 의해 각 범주별 학습이 이루어지면서 원형 이미지의 픽셀 값들을 판별자의 램에 기억한다. 인식 단계에서는 인식 이미지의 각 범주별의 판별자의 출력값인 MRD(Maximum Response Detector)를 이용하여 임계값을 구하여 패턴 인식이 이루어진다.^[5]

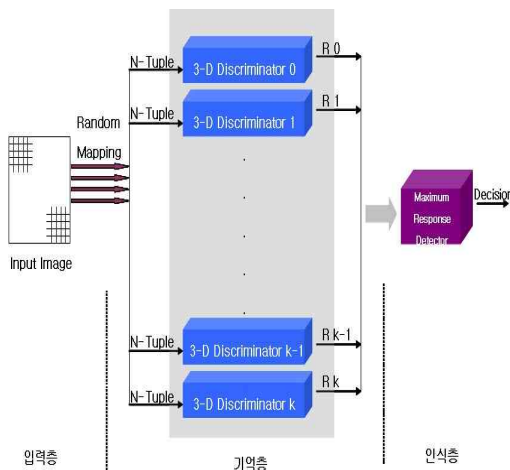


그림 2. 다중 판별자
Fig. 2. A Multi Discriminator

각 판별자별 임계값은 식 (2)와 같이 생성 이미지 (GenImage)의 생성된 유효 픽셀 합과 원형 이미지 (PrototypeImage) 픽셀 합에 의해 임계값이 생성된다.

식 2는 MRD에 의해 생성된 임계값이 정해진 값보다 크다면 인식된 패턴으로 처리한다.^[5]

$$\text{Threshold} = \frac{\sum_i \sum_j \text{GenImage}(i,j)}{\sum_i \sum_j \text{PrototypeImage}(i,j)} \times 100(\%) \quad (2)$$

2. CSWNN

가. SWNN

SWNN(Single layer Sequential Weightless Neural Network)은 동기 회귀 신경망이다. 그림 2에서 I(t)는 이진 형태의 입력 공간 차원을 가지는 입력 벡터, F(t)는 귀환 벡터, R(t)는 응답 벡터, D(t)는 원하는 응답 벡터이다. 귀환 벡터는 지연 단위인 d번째 후의 벡터로 F(t)=R(t-1)이다. 응답 벡터는 노드들의 현재 주소화된 위치의 값 또는 확률적인 방법을 적용한 결과물들을 포함한다.^[6]

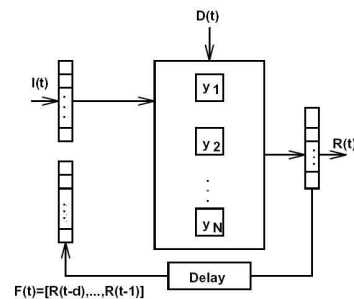


그림 3. SWNN
Fig. 3. A Single layer Sequential Weightless Neural Network

SWNN의 교육은 현재의 입력과 연속적인 현재의 전 단계 벡터를 이용하여 처리하는 간단한 쉬프트 레지스터 프레임워크와 같다. Aleksander's^[11]는 SWNN을 이용하여 단순한 형태의 정규 언어를 학습할 수 있는 것을 보여주었다.

나. CSWNN

Silva와 Ludermir는 정규 언어의 문제점을 해결하기 위하여 SWNN을 개선한 모델인 CSWNN^[12]을 제안하였다. CSWNN은 그림 4와 같이 바로 전단계의 지연 단위를 이용하여 상호 연결된 SWNN의 그룹으로 구성되어 있으며, 문장 기호의 맨 끝에 도달하기 전까지는 바로 전

단계의 응답값을 이용하여 학습을 처리하고 맨 마지막의 응답은 처음 시작의 지연 단위를 연결하여 처리하는 시스템으로 이는 앞 뒤 전후 관계를 연동할 수 있는 문맥 학습(context learning)이 가능 하도록 되었다.^[6]

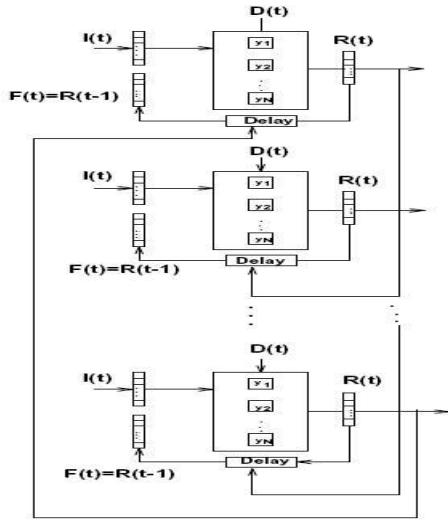


그림 4. CSWNN
Fig. 4. A Cascaded sequential weightless neural network

III. SCNN

1. SCNN

CSWNN는 연속적인 입력 패턴을 학습하여 연속적으로 기억하여 처리하는 모델로 정형화된 패턴에 대하여 안정적 성능을 나타낸다.^[6] 이를 위하여 제한적이면서 정규화된 연속적인 입력 패턴은 가능하지만 사람마다 다양한 형태 및 위치 변화가 많은 수화와 같은 행위 정보에 대한 연속적인 입력 패턴에 대하여는 성능이 낮아질 수 있다. 시계열적인 연속적인 패턴에 대하여 추가 및 반복 학습이 가능하면서 일반화 특징점 패턴 추출이 가능한 순차적인 누적 신경망을 제안한다. 3DNS를 연속적 입력 패턴으로 학습할 수 있도록 개선하여 행위 정보와 같은 원형 패턴의 변형이 많은 정보를 전처리 과정 없이 빠른 시간 안에 최소한의 계산량으로 처리 가능한 모델을 제안한다. 그림 5는 연속적인 입력 패턴들을 하나의 범주인 판별자에 학습하는 것이다. 한 범주의 판별자에 연속적인 입력 패턴 프레임인 1 ~ k까지 패턴을 교육한다. 교육 단계는 첫 번째와 마지막 교육 패턴은 Blank 입력으로 시작과 끝을 마무리 한다.

인식 단계는 교육된 패턴을 그림 6에서와 같이 각 프레임별 범주 판별자의 MRD를 찾아 일정 임계값 이상일 경우 그 범주로 선택하고 이하일 경우 오인식으로 처리한다. 그러나 각 프레임당 MRD값의 결정값을 취합하여 전체적으로 가장 많은 발생한 값으로 결정 범주를 선택하여 인식한다.

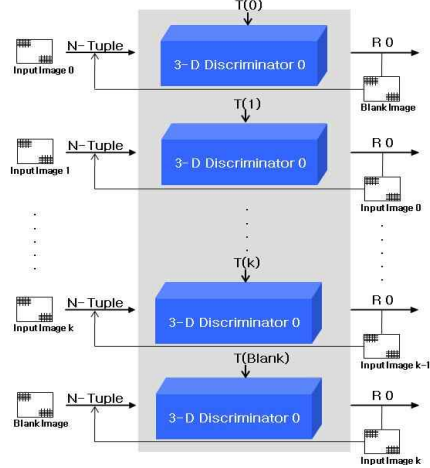


그림 5. SCNN
Fig. 5. A Sequential Cumulative neural network

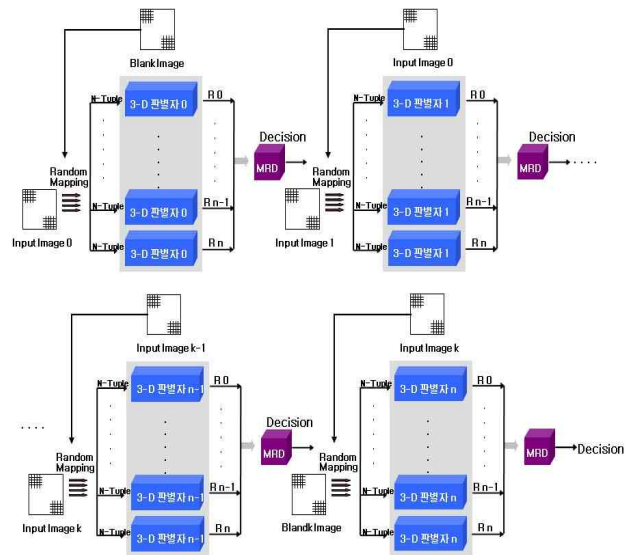


그림 6. SCNN 인식 구성도
Fig. 6. A Sequential Cumulative neural network of recognition diagram

IV. 수화 인식 시스템 실험 및 결과

1. 수화인식 시스템

수화는 인간의 행위 정보 가운데 제스처를 이용하여 의사 소통을 하는 언어로 비장애인의 언어와는 다른 체계를 가지고 있다. 수화는 시각적인 전달 수단을 목표로 하는 언어체계로 여러 시각적인 요소를 동시에 혼합하여 신호를 전달 할 수 언어이다. 구어를 바탕으로 한 독립적인 수화의 신호에는 손 구성(손이 취하는 모양), 조음위치(손모양이 이루어지는 장소 또는 위치), 그리고 동작(손이 움직이는 모습 및 방향) 등의 세 가지 기본적인 특성이 있다. 세 가지 기본 특성을 이용하여 수화를 자동적으로 인식하는 시스템 개발에 연구가 활발히 진행되고 있다. 수화를 구성하는 단어는 하나의 단일한 동작의 단일어와 여러 개의 단일어를 조합한 복합어^[14]가 있다. 본 논문에서는 수화의 기본적인 특성을 연속적인 수화 행위 정보를 입력 받아 복합어의 통사론적으로 인식 가능한 시스템을 제안한다. 카메라에서 입력 받은 이미지를 초당 10 프레임으로 캡처하여 경계선을 검출한 뒤 이진화 처리 후 연속적으로 SCNN에 입력하여 학습한다. 인식 단계에서는 입력 받은 정보를 이용하여 SCNN의 연속적인 MRD 값을 이용하여 인식한다.

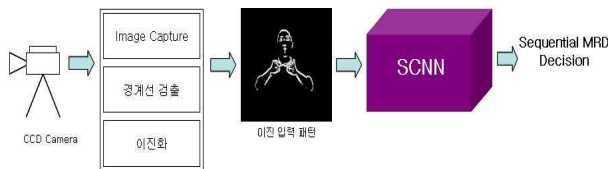


그림 7. 수화 인식 시스템
Fig. 7. The recognition system of sign language

2. 실험 결과

연속적인 수화 이미지 패턴들을 SCNN 모델에 연속적으로 입력하여 학습하였다. 입력 방법은 카메라를 통하여 초당 10프레임으로 영상을 입력 받아 에지 검사를 거쳐 후 이진 영상으로 변환하여 64X64의 패턴으로 학습망에 입력된다. 실험으로 사용된 수화는 한글 복합어 수화 100개의 단어로 한 글자 20개, 두 글자 60개, 세 글자 14개, 네 글자 5개, 다섯 글자 1개를 이용하여 실험하였다. 수화 인식 성공률은 93%로 나타난다.



그림 8. 10 프레임 단위의 “고맙다” 수화 영상
Fig. 8. A "thank" sign language image per 10 frame



그림 9. "물" 수화 영상
Fig. 9. A "Water" sign language image



그림 10. "얼음" 수화 영상
Fig. 10. A "Ice" sign language image

그림 8은 카메라로 입력된 연속적인 “고맙다” 수화 이미지를 학습망에 사용하기 위하여 경계선 검출 후의 이진 영상이미지들이다.

표 1과 같은 오인식이 발생하는 경우는 교육 영상과 오인식 영상의 유사도가 매우 높은 경우와 학습된 영상의 특징점들이 적은 경우이며 이는 손 모양 영역 검출 등의 전처리 과정을 삽입한다면 오인식을 줄이는 방법이 될 것이다.

표.1. 오인식 단어들

Table. 1. The false recognition words

단어	오인식 단어	최종 오인식 단어	오류 프레임
물	눈, 얼음, 산	얼음	10
아기	공군, 사기꾼, 청년	공군	9
공기	얼음	얼음	3
안개	눈, 비	Don't care	3
모두	공군, 단체	Don't care	5
무지개	공기, 눈	Don't care	5
어린이	얼음, 공기	Don't care	5

V. 결론

일련의 연속적인 사건 및 지식을 인식하기 위하여 기호 및 의미 등으로 오토마타 및 인공지능에 많은 방법으로 연구가 진행되어 왔다. 그러나 이들 대부분은 많은 전처리 과정 및 교육/인식 시간이 필요로 하는 문제점을 가지고 있었다. 이런 문제점을 해결하기 위하여 램 기반의 모델 및 회귀적 또는 통계적인 방법의 모델 등이 제안되었다.

본 논문에서는 연결성을 가지면서 연속적으로 입력되는 지식을 인식하기 위하여 순차적 램 기반 신경망을 개선한 순차적인 누적 신경망을 제안하여 수화 인식에 적용하여 보았다. 한글 수화는 단일어와 복합어로 구성되어 있는데 대부분의 연구는 단일어에 집중되어 있다. 단일어의 결합 형태인 복합어를 분석하여 보면 단일어의 연속적인 행위 정보들의 결합 형태이다. 카메라에 입력되는 영상을 경계선 검출 및 이진화를 통하여 입력 영상을 만들고 이를 전처리 과정 없이 램 기반의 순차적 누적 신경망에 적용하여 인식하여 보았다. 인식률은 93%로 얻었다.

향후 영상에서 손 모양 영역 검출 및 신경망의 입력 단계에서의 전처리 과정의 추가 그리고 다양한 응용 분야에 적용 활용에 연구가 진행되어야 할 것이라고 본다.

참 고 문 헌

- [1] 이성환 “오프라인 필기체 문자인식 기술의 현황”, 한국정보과학회지, 제11권 제5호, 1993년 10월.
- [2] I. Aleksander and H. Morton, “An Introduction to Neural computing”, Chapman & Hall, 1990
- [3] 권영철, “경험 유관이진신경망”, 울산대학교, 1995
- [4] 김성진, 권영철, 이수동, “RAM을 이용한 경험유관 축적신경망모델”, 대한전자공학회, 제41권, CI편, 제2호, pp. 98~102, 2004년 3월
- [5] 김성진, 이동형, 권영철, 이수동, “다중판별자를 가지는 동적 삼차원 뉴로 시스템”, 한국정보과학회, Vol.34, pp.585-594, No. 7. 2007
- [6] Marcílio Carlos Pereira de Souto, Paulo J. L. Adeodato, “Learnability in sequential RAM-based neural networks“, SBRN 1998: pp.20-25
- [7] P. J. L. Adeodato. “Sequential digital neural networks.“, Technical report, King’s College, London, UK, 1994
- [8] I. Aleksander. “Microcircuit learning computers.“, Mills and Boon Ltd., 1971
- [9] I. Aleksander. Neural systems engineering: towards a unified design discipline? IEE Journal of Computing & Control, 1(6):pp.259-265, 1990
- [10] Aleksander, I., Thomas, W. V., and Bowden, P. A. 1984 “WISARD a radical step forward in image recognition,” Sensor Review, pp. 120-124, July 1984
- [11] I. Aleksander. “Iconic learning in networks of logical neurons.“, In Proc. of ICES96, Tokyo, 1996. Invited paper.
- [12] S. Silva and T. B., Ludermir. “Implementing finite state automata in recurrent RAM-based networks.“, In Proc. ICANN95, pp.461-466, 1995
- [13] 전북도교육청, DCU 특교과 스티디, <http://www.cein21.net/suhwoa/>
- [14] 김정배, 김대진, 장원, 변증남, 김성권, “연속된 한글 수화에서의 복합어 인식”, HCI 2000, 한국정보과학회, pp.24-26, 2000년. 1월

저자 소개

이 동 형(정회원)



- 1996 울산대학교 컴퓨터공학과학사 졸업.
- 1998 울산대학교 컴퓨터공학과석사 졸업.
- 2000 울산대학교 컴퓨터공학과 박사 수료.
- 2001~현재 한국폴리텍VII대학 울산 캠퍼스 정보통신시스템과 부교수

<주관심분야 : 네트워크, 신경망, 지능형 로봇>

강 만 모(정회원)



- 1998 울산대학교 전자계산학과 학사 졸업.
- 2000 울산대학교 컴퓨터공학과 석사 졸업.
- 2003 울산대학교 컴퓨터공학과 박사 수료
- 2006~현재 울산대학교 객원교수

<주관심분야 : 전자상거래, 멀티에이전트 시스템>

김 영 기(정회원)



- 1975. 인하대학교 조선공학과.
- 1985 부산대학교 조선공학과 석사 졸업.
- 1996 울산대학교 컴퓨터공학과 박사 졸업.
- 1979 ~ 현재 울산과학대학 컴퓨터정보학부 교수

<주관심분야 : 인공지능, 신경망, 멀티미디어 >

이 수 동(정회원)



- 1969 서울대학교 전자공학과 석사 졸업.
- 1980 Brunel University 석사 졸업.
- 1984 Brunel University 박사 졸업..
- 1976~2004 울산대학교 컴퓨터 정보통신 공학부 교수.
- 2005~현재 울산과학대학 학장

<주관심분야 : 인공지능, 신경망, 소프트 컴퓨팅>