

시간 지연 신경망을 이용한 동작 분석

Motion Analysis with Time Delay Neural Network

장 동 식, 이 만 희, 이 종 원
(Dong-Sik Jang, Man-Hee Lee, and Jong-Won Lee)

Abstract : A novel motion analysis system is presented in this paper. The proposed system is inspired by processing functions observed in the fly visual system, which detects changes in input light intensities, determines motion on both the local and the wide-field levels. The system has several differences from conventional motion analysis systems. First, conventional systems usually focused on matching similar feature or optical flow, but neural network is applied in this system. Back propagation is used by learning method, and Time Delay Neural Network (TDNN) is also used as analysis method. Second, while conventional systems usually limited on only two frames of sequence, the proposed system accepts multiple frames of sequence. The experimental results showed a 94.7% correct rate with a speed of 71.47 milli seconds for real and synthetic images.

Keywords : time delay, neural network, motion analysis

I. 서론

약간의 시차를 두고 연속적으로 입력되는 동영상(dynamic image)에서 동작(motion)을 분석하여 실제세계에 대한 정보를 추론하는 작업은 영상 처리의 여러 분야에서 매우 중요한 일이다. 특히, 동작 추정은 시각적 정보에서 잉여정보를 줄이는 방법이고, MPEG과 같은 최근의 많은 동영상 압축 알고리즘에 매우 중요하게 사용되고 있다.

원래의 3차원 장면에서의 동작의 재구축은 잘못된 포즈(pose) 등의 이유로 항상 가능하지는 않다[1]. 따라서 많은 접근방법들은 3차원 속도들의 영상 평면으로의 사영인 영상 속도들, 또는 광류(optical flow)의 추정에 제한하여 왔다. 영상의 연속적인 데이터의 이용으로 광류 벡터 필드를 정의하기 위한 여러 가지 계산 방법들이 제안되었다[2]-[6].

움직이는 물체를 검출해내고 추적하는 문제는 더 크고 더 복잡한 문제인 동물의 동작 인식과 분석 능력을 모델링하는 문제의 기초이다. 동작 인식 문제는 시각 능력을 가진 동물에게 중요한 것으로, 동작 인식은 영상을 3차원 적으로 재구성하고, 영상을 분할하고, 눈의 움직임을 조정하고, 주의를 환기시키고, 자신의 움직임을 알아내고, 불변 크기의 중재, 움직이는 물체를 검출하는 등의 일을 해낸다[7]. 이러한 능력들을 흉내내기 위하여 차영상이나 광류 필드를 이용하여 인공적으로 동작을 검출해 내려는 여러 가지 접근 방법들이 연구되었으나 많은 이러한 노력에도 불구하고, 생물학적 시각 시스템과 같이 효과적으로 동작을 처리하는 인공적인 계산 방법들은 아직 없었다. 최근 발표된 일부의 연구에서 신경망 이론을 도입하거나 동물의 시각 시스템을 분석, 연구하여 그 시스템을 본 따는 시도를 하기 시작하였다[8].

본 논문에서는 동물의 시각 시스템을 참조로 하여 구

성한 시스템에 신경망 기법을 응용하여 동작 분석을 하는 새로운 접근 방법이 제시되었다. 참조할 시각 시스템은 현재까지 어느 정도 알려진 파리의 시각 시스템을 참조 모델로 하였다. 파리의 시각 시스템을 모델로 하는 이유는 비교적 작고, 모듈화 되어 있으며 알기 쉬운 시스템이면서도, 효율적으로 움직이는 물체를 쫓거나 방해물을 피할 수 있기 때문이다. 그리고 이 시스템에 시간 지연 신경망(time delay neural network)을 도입하였다. 신경망은 현재 많은 분야에 응용되고 있으며 특히 패턴 인식(pattern recognition) 문제에 성공적으로 응용되고 있다. 그것은 신경망이 재래식 노이만(Neumann)형 컴퓨터와는 달리 병렬처리, 학습 그리고 잡음에 강한 계산 기능을 갖고 있기 때문이다. 특히 화상, 음성인식의 문제는 오랫동안 인공지능 분야에서 연구되어 왔으나 주로 algorithmic, symbolic 및 논리적 방법을 사용함으로써 큰 진전을 보지 못하였다. 그 이유는 패턴 인식 문제가 근본적으로 병렬처리를 요구할 뿐 아니라 그 문제의 해결에 있어 체계적인 알고리즘을 발견하기 어려우며, 또 흔히 잡음이 긴 데이터를 다루는 데 비하여 재래식 컴퓨터나 인공지능에서 사용하는 symbolic 방법이 적절하지 못하기 때문이다. 현재 재래식 방법으로 해결하기 어려웠던 여러 가지 문제, 특히 패턴 인식과 로봇 등의 제어 그리고 최적화 문제 등에 이러한 신경망의 응용이 시도되고 있다.

본 논문의 2절에서는 동작 분석에 관한 전반적인 설명을 한 후, 현재까지 연구되었던 기존의 방법들을 특징에 기초한 방법(feature based method), 경사에 기초한 방법(gradient based method) 등 2가지로 나누어 이에 대해 설명을 한다. 3절에서는 본 논문의 시스템에 대한 설명으로써 이론적 배경을 설명하고, 시스템의 구성에 대해 설명을 한다. 4절은 본 논문의 시스템에 대한 실험으로써, 실험 방법의 설명과 실험 및 실험 결과에 대해 설명한다. 끝으로 5절은 결론으로써 본 논문의 내용을 종합하여 정리하고, 앞으로의 연구 과제에 대해 언급하

였다.

II. 기존의 동작 분석 알고리즘

동작 분석은 컴퓨터 비전 분야에서 오랫동안 많은 다른 접근 방법과 응용을 가진 대표적인 문제였다. 연속 장면에서 동작 모수(motion parameter)들을 추정하는 것도 풀어야 할 기본적인 문제들 중의 하나이다. 연속된 장면 중 한 영상에서 동작 요소가 검출되었을 때, 만약 그 속도벡터의 방향과 크기가 알려진다면, 단순한 추정 절차에 의해 연속된 장면 중 각 영상에서의 정확한 위치를 찾아낼 수 있다. 같은 방식으로 같은 물체의 두 연속적인 영상이 주어지고 이 밀착된 두 영상의 상관관계 함수가 최대화된다면, 변위는 추정되어질 수 있다. 연속된 장면 중 단지 두 영상만이 필요한 이 방법은 움직이는 물체의 속도가 일정하게 유지된다고 가정하면, 많은 연속된 영상으로 일반화되어 질 수 있다. 영상의 작은 영역 안의 모든 픽셀의 명암이 t 라는 시점에 완전히 정의되어 질 수 있다고 가정하고 한 쌍의 동작 인자들(크기와 방향)이 주어지면, 이 영역에서 짧은 시간 동안의 부분 픽셀의 연속적인 명암 값을 예측할 수 있고 실제 영상과 예측된 영상의 연속된 장면 사이의 상관관계가 평가된다. 만약 예측 인자들 집합을 위한 상관관계의 최대값이 찾아지면, 이 인자들이 그 물체의 동작을 잘 정의한 것이라 할 수 있다. 동작 인자들의 추정을 필요로 하는 많은 요소 연산들은 - 많은 테스트된 속도들, 많은 연속적인 영상들과 밀착된 영상의 크기들에 비례하여 매우 중요하다

1. 특징에 기초한 방법을 이용한 동작 분석 알고리즘

동작 분석을 위한 첫 번째 접근방법은 특징에 기초한 방법(feature based method)으로, 전후 영상에서의 움직임에 불변한 특징의 연관성으로부터 움직임을 계산한다. 이 방법은 연속적인 영상을 따라서 두드러진 물체의 특징을 추적하는 것으로 이 원칙을 기초로 하는 모든 방법들은 연속된 영상에서 그 물체를 나타내는 두드러진 특징과 동일성을 이용, 패턴 매칭 알고리즘을 적용하여 그 물체의 동작을 검출한다.

특징에 기초한 방법(feature based method)을 이용한 동작 분석을 간단히 설명하면 다음과 같다. 시간 순차적으로 얻어지는 2개 혹은 그 이상의 사영적으로 투영된 영상(perspective projected image)을 얻는다. 동작으로 인하여 얻어진 영상평면들에는 미리 결정되고 추출된 영상 특징들이 있다. 이 때 연속된 두 개의 영상평면 위의 특징들 사이의 대응 관계를 알 수 있어야 한다. 이러한 대응관계를 결정하는 문제를 정합문제(correspondence problem)라 하는데 실제로 매우 어려운 작업이다. 일단 대응 관계가 결정되었다면 해야 할 작업이 그 대응 관계로부터 운동인자들, 즉 회전량, 이동량, 물체에 대한 법선 벡터를 구해내는 것이다. 이 마지막 단계를 운동 문제(problem of structure from motion) 또는 전통적으로 사진 측량학에서 말하는 상대적 위치 문제(problem of relative orientation of camera)라고 한다.

특징에 기초한 방법을 이용한 동작 분석 알고리즘에는 몇 가지 문제점이 있다.

- ① 사용할 수 있는 정보의 비효율적 사용: 한번에 단 2개의 프레임만을 다루는 게 일반적이고, 그 이상의 프레임을 다룰 때에는 계산이 복잡해진다.
- ② 비자연스런 이동과 회전 모수들: 한번에 단 2개의 프레임만을 다룰 때 나타나는 문제이다.
- ③ 높은 영상 샘플링율: 연속된 프레임들에서 작은 동작들이 필요하다 (특히 회전 동작 시). 이것은 경사에 기초한 방법에서 특히 문제시된다.
- ④ 제한적 가정: 특징 점들이나 특징 동작들을 동일한 평면상에 있다고 제한해야 한다.
- ⑤ 방정식의 비선형성: 대부분의 방법에서 방정식은 비선형성을 갖는다.
- ⑥ 잡음 민감도: 입력할 때 발생한 작은 잡음이 동작 인자들에서 커다란 오류가 될 수 있다.
- ⑦ 순수 회전 동작과의 문제: 물체의 동작이 순수 회전 운동일 때 움직이는 물체의 구조가 복구 될 수 없을 수도 있다.
- ⑧ 좋은 초기 값의 필요성: 반복을 거듭함으로써 해답을 얻는 방법에 있어서 수렴을 위해 초기 값의 선택이 아주 중요하다.
- ⑨ 높은 복잡성: 몇몇 방법들은 너무 많은 수의 변수를 갖는 방정식으로 나타내어진다.

최근에는 기존의 특징에 기초한 방법에 신경망 기법을 도입하려는 여러 시도가 있었다. 이런 방법들은 대부분 특징이 되는 점, 선 등을 찾는데 신경망을 사용했다는 점에서 공통점을 찾을 수 있다[9].

2. 경사에 기초한 방법을 이용한 동작 분석 알고리즘

광류(optical flow)란 빛의 밝기의 이동 방향을 나타내는 벡터로 광류를 이용하면 물체의 움직임을 알아낼 수 있다[2][10][11]. 두 번째 접근방법인 경사에 기초한 방법 (gradient based method)에 있어서 우선적인 가정은 영상 명암의 변화는 전적으로 물체의 동작에 달려 있다고 가정을 한다. 이 가정 하에서 연속 영상에서의 움직이는 물체의 추출은 2단계로 이루어진다. 첫 단계는 여러 개의 영상으로부터 광류를 계산하는 단계로, 밝기의 공간적 변동(spatial variation)과 시간적인 변화(temporal difference)의 관계로부터 광류를 구한다. 두번째 단계는 계산된 광류를 분할하여 움직인 부분을 추출하는 단계로, 첫 번째 단계에서 구해진 2차원 속도 차원인 광류에 근거하여 3차원 동작 모수들인 방향, 속도, 위치 등을 구한다. 그림 1과 같이 일정한 명암을 갖고 있는 영역(물체의 중심 부분)과 동작의 방향에 평행한 에지(edge)는 동작을 추출하기 위한 정보를 제공하지 못한다. 그림 1에서 (a)는 움직이기 전의 영상이고 (b)는 움직인 후의 영상이다. 또한 (c)는 (a)영상에서 (b)를 뺀 영상이다.

오직 동작의 방향과 수직인 에지만이 동작의 정보를 나타낼 수 있다. 또한 그림 2와 같이 작은 틈새(aperture)안에서 이용할 정보가 너무 적어 전체 속도 벡터를 계산할 수 없을 때 이것을 aperture 문제라고 한다. 그

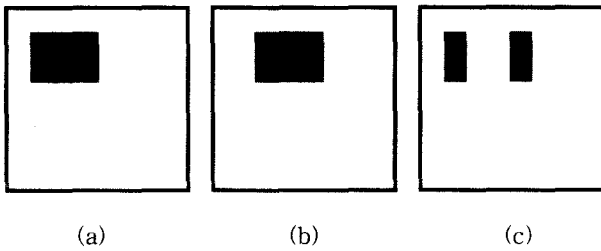


그림 1. 차 영상의 효과.
Fig. 1. The effect of difference image.

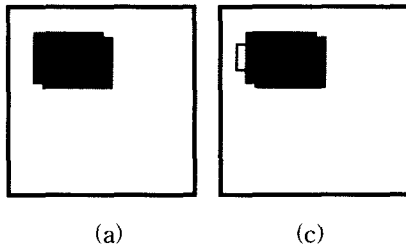


그림 2. Aperture 문제.
Fig. 2. Aperture problems.

림 2의 (a)에서 짙은 회색은 동작영역을 보여주며 (b)의 작은 틈새(검은 색 사각형)은 거의 동작을 보여주지 못한다.

이런 경사에 기초한 방법에도 신경망 기법을 도입하려는 시도가 여러 차례 있었다. 신경망은 잡음에 강하고 하드웨어로 구현하면 실시간 처리가 가능하기 때문이다 [12]-[14].

경사에 기초한 방법과 특징에 기초한 방법과는 다음과 같은 차이점이 있다. 특징에 기초한 방법은 사영된 치환 모델을 이용하는 반면에, 경사에 기초한 방법은 사영된 속도 모델을 이용한다. 그리고 특징에 기초한 방법에서는 두 영상간에 두드러진 특징이 되는 점, 선들의 집합과 이들의 매칭을 이용하는 반면, 경사에 기초한 방법은 영상 명암 기울기의 추정을 이용한다. 이러한 경사에 근거하여 광류를 이용한 동작 분석 알고리즘은 특징이 되는 점 또는 선 사이의 일치를 수립할 필요가 없고 영상에서 각 점들의 명암값으로 구하기 때문에 특징에 기초한 방법의 단점을 어느 정도 극복했다고 말할 수 있다. 하지만 추정된 광류 필드와 공간-시간적 영상 명암의 기울기에 많은 영향을 받는데다가, 수식이 복잡하고 계산 량이 많아 계산시간이 오래 소요되어 아직 실용적이지 못한 점이 있다

III. 시간 지연 신경망을 이용한 동작 분석 알고리즘

현재까지의 동작 추정을 위한 표준적 계산 방법들은 동물이나 사람의 시각 시스템과 비교해볼 때 매우 취약하다고 할 수 있다. 예를 들어 파리의 눈 구조는 그 단순함에도 불구하고 움직이는 물체를 쫓거나, 방해물을 피하는 능력이 뛰어나다. 그래서 여러 연구자들은 곤충과 같은 비교적 단순한 동물들의 시각 시스템에 대하여 연구해 왔다. 연구들의 목표는 시각 데이터를 처리하는

데 있어 생물학적 시각 시스템과 같이 효과적인 처리 구조를 정의하는데 있다.

최근, Texas주 Arlington에 위치한 텍사스 대학의 한 연구진이 동작 검출에 인공 신경망을 사용하는 방법을 제안하였다[15]. 그들이 제안한 네트워크의 구조는 아직 완전히 이해되지 않은 파리의 시각 시스템으로부터 기인하였다. 이 네트워크는 파리의 초기 시각 처리를 본 때 신호대 잡음비를 높이기 위해 3×3 평활화 필터를 이용했고, 파리의 시각 시스템을 구성하고 있는 On-Off와 지연단위의 활동을 모방, 입력 광량의 변화를 네트워크의 입력으로 사용했다. 또한 파리의 상관관계형 동작 검출기를 흉내 낸 medulla 층 구조를 제안하여 지역적인 수준과 전역적 수준의 동작을 결정한 후 동작 추적을 제어하기 위한 변위 정보를 출력값으로 결정했다. 본 논문에서 제시되는 새로운 동작 분석 알고리즘도 이러한 파리의 시각 시스템으로부터 시작하였다. 받아들인 데이터를 처리, 학습, 결과를 끌어내는 방법으로 시간 지연 신경망(TDNN : Time Delay Neural Network)을 이용하여 영상에서의 흑백 명암값 변화의 분석을 기반으로 한 동작 분석 방법이 제시된다.

1. 시간 지연 신경망(time delay neural network)

인간의 뇌는 대량의 복잡한 데이터를 효율적으로 병렬처리 할 수 있을 뿐만 아니라 학습능력이 있다. 이에 근거하여 새로운 계산방식으로서 신경망이 제안되었으며 여러 분야에 성공적으로 사용되고 있다. 신경망은 정보의 병렬처리와 복잡한 기능의 학습이 가능함으로써 다량의 정보처리가 요구되거나 복잡하고 어려운 문제들, 특히 algorithmic한 해결책을 찾기 어려운 문제들에 효율적으로 응용될 수 있다. 신경망은 재래식 계산방식에서는 특히 비효율적인 패턴 인식문제를 효율적으로 해결하는 데 매우 적합하다. 그 이유는 패턴 인식에 필요한 많은 양의 데이터를 병렬 처리할 수 있고, 패턴 인식과 같은 수학적 알고리즘의 적용이 곤란한 문제를 학습 방법에 의하여 효과적으로 해결 할 수 있으며, 패턴 인식에 있어 자주 발생하는 잡음이 많은 영상이나 퍼지(fuzzy)한 데이터를 효과적으로 처리할 수 있기 때문이다. 이러한 특징을 갖는 기존의 신경망은 시간의 개념이 없이 정적인 데이터만 이용할 수 있었기 때문에 컴퓨터 비전 부문에서는 문자인식, 특징인식등의 어떤 패턴을 인식하는 분야에 많이 이용되었다. 1988년 Waibel[16]은 기존의 신경망 연결에 시간 지연(time delay)을 고려한 새로운 신경망 이론을 제안하였다. 시간 지연 신경망이라 불리는 이 신경망은 시간 지연이라는 시간 개념을 넣어 시간에 따라 변하는 값(time-variance)도 고려할 수 있게 설계한 신경망이다. 이 시간 지연 신경망은 각 층의 연결에 특징 단위(feature unit)뿐만 아니라 지연 단위(delay unit)도 같이 구성되는데, 이 지연 단위는 따로 존재하는 것이 아니라, 각 특징 단위에 부속되어 각각의 지연 단위 값을 가지게 됨으로써 각 층이 연결되어 학습을 할 때 시간에 따라 변하는 값도 고려할 수 있게 된다. 그는 이 신경망을 음성인식 분야에 이용하기 위하여

고안하였으나 현재는 기존의 음성 인식 분야뿐만 아니라 증권 동향 분석, 기계 조절 등의 분야에도 성공적으로 이용되고 있다.

2. 파리의 시각 시스템

최근 생물학계의 여러 연구에 의하면 동물들이 색 구분 능력이나 binocular 시각 능력은 떨어지더라도 동작을 처리하는 능력은 모자라지 않다고 한다. 많은 연구들이 그 중에서도 파리의 시각 시스템에 주시하여 일반적인 시각 시스템을 이해하거나 특별한 동작 인식 능력을 이해하려고 노력했다. 다른 동물의 시각 시스템의 모델을 개발하는 것 보다 파리의 시각 시스템의 모델을 개발하는 것이 보다 쉬운데 그 이유는 파리의 시각 시스템은 비교적 작고, 모듈화 되어있으며, 알기 쉬운 시스템이기 때문이다.

파리의 시각 시스템의 일반적인 조직 구조는 여러 연구자들에 의해 조사, 연구되었다. 간단히 설명하자면 다음과 같다. 빛이 대략 3500개의 Ommatidia(OM)라 불리는 6각형 모양의 빛을 받아들이는 세포들로 이루어진 복잡한 2개의 눈으로 들어온다. 각 ommatidium은 R1에서 R8까지의 8개의 reticular 세포들로 이루어져 있는데, R1~R6은 동작 검출에 있어서 매우 중요하다. R1~R6의 출력값은 ommatidia와 같은 수를 갖는 cartridge에 의해 이루어진 lamina 신경 세포에 입력된다. 각 cartridge는 주변의 6 ommatidia로부터 하나의 입력 값을 갖는다. lamina cartridge에 입력되는 자극들은 같은 공간 점으로부터 얻어진 빛으로부터 기인된 것이다. ommatidia의 R1~R6에서 나온 출력값이 lamina를 자극시킨 후, medulla 세포로 들어간다. medulla의 출력값은 lobula와 lobula plate의 두 개의 신경세포로 복잡하게 구성된 lobula에 수렴된다.

파리의 시각 시스템을 정확하게 복사하려는 시도는 두 가지 이유에서 현재 실용적이지 않다. 첫 번째로 신경생리학자들이 파리의 시각 시스템의 뉴런들의 활동과 상호작용을 아직 완전하게 매핑하고 결정하지 못하였다. 두 번째로 설사 그 시스템이 완전히 매핑되고 이해됐을 지라도 그렇게 많은 수의 노드와 상호 연결을 가진 하드웨어를 디자인한다는 것은 매우 어렵고, 만약 가능하다면 가까운 미래에는 불가능하다. 하지만, 시스템의 주요 기능들을 신경망을 이용하여 모델링하는 방법은 동작 검출과 추적 하드웨어와 소프트웨어를 발전시키는데 매우 유망한 접근방법이다.

3. 시스템의 구조

1) 기본 가정과 데이터 전처리

본 논문에서 제안된 동작 분석 접근방법은 다음과 같은 가정을 전제로 한다:

- ① 화면에서 명암값의 변화는 동작에 의해서만 발생된다.
- ② 목표 물체의 명암 값은 배경과 다르다.
- ③ 동작은 관련된 연속 영상에 걸쳐 일정한 속도로 일어난다.
- ④ 동작은 이동 운동과 회전 운동의 요소들로 분해되어진다.

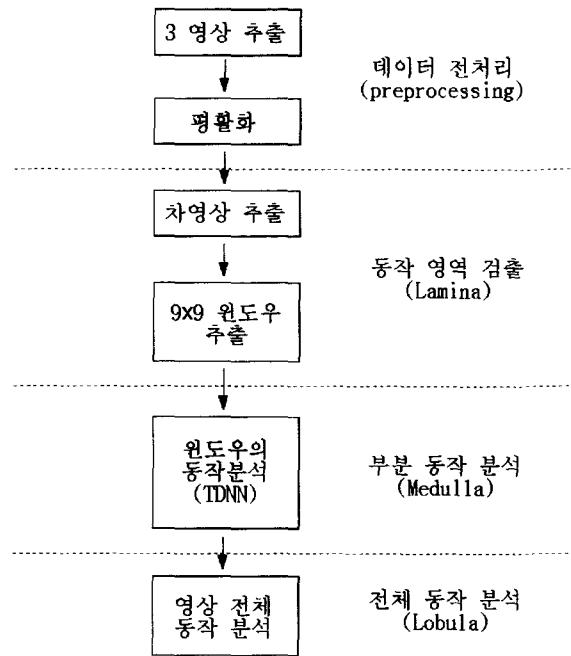


그림 3. 전체 시스템 구성도.
Fig. 3. Overview of the whole system.

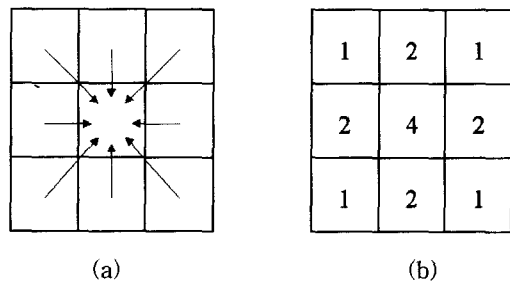


그림 4. (a) 평활화 방법, (b) 평활화 가중치 값.
Fig. 4. (a) Smoothing method, (b) The value of smoothing weight.

⑤ 카메라를 향해 오거나, 멀어지는 운동이 아니다.

이상의 가정은 다른 기존의 여러 동작 분석을 위한 알고리즘과 크게 다른 점이 없다.

그림 3에 본 논문에서 사용된 시스템의 구성도가 나타나 있다.

데이터 입력을 위한 전처리과정은 다음과 같다. 연속된 장면 중에서 이어지는 3개의 영상을 추출한다. 추출된 각 영상에 (1)과 같이 평활화(averaging)를 수행한다. 평활화를 수행함으로써 추출된 영상들의 잡음을 줄여 신호 대 잡음비(SNR : Signal to Noise Ratio)를 높일 수 있다. 그림 4의 (a)와 같이 주변 8근방의 명암값들을 이용하여 평활화를 수행하는데 평활화를 하는 과정은 파리의 시각 시스템에서 cartridge 부분이 초기에 필터링하는 일과 같은 역할을 한다.

$$I_A = \frac{\sum_{i=1}^9 w_i I_i}{\sum_{i=1}^9 w_i} \quad (1)$$

여기서 I_A : 평활화한 영상값

I_i : 평활화를 할 좌표와 8근방 좌표들의 명암값

w_i : 평활화를 할 좌표와 8근방 좌표들의 가중치값

2) 동작 영역 검출(lamina)

파리의 시각 시스템에서 lamina 부분이 하는 일은 받은 영상을 필터링하여 영상의 질을 높인 후, 동작 영역을 검출하는 일이다. 본 연구에서는 연속적인 동영상에서 추출해낸 평활화한 연속적인 3 프레임의 영상들을 각각 1 프레임 전의 영상의 명암값을 뺀 차영상을 구한다. 각 영상의 차영상을 구함으로써 각 영상에서 동작이 있는 부분만을 나타내고, 나머지 동작이 있지 않은 부분은 영상에서 제외된다.

$$D_{xy} = \begin{cases} 0 & \text{if } I_{xy}(t) - I_{xy}(t-1) < \Gamma \\ I_{xy}(t) - I_{xy}(t-1) & \text{else} \end{cases} \quad (2)$$

여기서 D_{xy} : 좌표 (x,y)에서의 차영상값

$I_{xy}(t)$: 시간 t에서 좌표 (x,y)에서의 명암값

Γ : 임계(threshold)값

(2)와 같이 영상에서 동작이 있는 부분은 0이 아닌 값이 할당되고, 그 이외의 동작이 없는 부분은 0의 값이 할당된다. 이때 임계값 Γ 에 의해서, 잡음에 의하여 생긴 명암의 변화, 무시해도 좋을 만큼 매우 작은 크기의 동작으로 생긴 명암의 변화, 또는 카메라의 진동이나, 조명에 있어서의 약간의 변화 등으로 인해 생긴 명암의 변화 등이 제거된다. 구해진 차영상에서 값이 0이 아닌 부분, 즉 동작이 있는 부분만을 추출하여 부분 동작 분석을 위한 신경망에 집어넣을 수 있는 형태로 만드는 작업이 그림 5와 같이 수행된다. 영상에서 스캔라인을 따라 값이 0이 아닌 부분을 찾아 그 좌표를 중심으로 왼쪽으로 4, 오른쪽으로 4, 위쪽으로 4, 아래쪽으로 4 픽셀씩을 취하여 9×9 윈도우를 추출한다. 그 부분은 다음에 다시 동작 영역을 찾을 때 중복되지 않도록 표시를 하고 다음 동작 윈도우를 추출한다. 이러한 과정을 전체 영상에 대해 적용하여 모든 동작 영역을 9×9 윈도우로 추출한다. 이때 그림 5의 (a)와 같이 기준이 되는 영상은 현재의 영상에서 1 프레임전의 영상을 뺀 차영상이 되고, (b)와 같이 1 프레임 전의 영상에서 2 프레임 전의 영상을 뺀 차영상에서도 기준 영상에서 추출된 윈도우와 같은 좌표를 갖는 9×9 윈도우를 추출한다. 이렇게 추출된, 같은 좌표를 갖는 두 9×9 윈도우를 묶어 하나의 입력값으로 만든다.

3) 부분 동작 분석(Medulla)

앞서 동작 영역 검출부에서 구한 9×9 윈도우들을 시간 지연 신경망을 이용해 구현한 네트워크에 입력한다. 이 네트워크의 구조는 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)의 3부분으로 구성되었다. 입력층 각 뉴런들의 입력값은 9×9 윈도우에서의 명

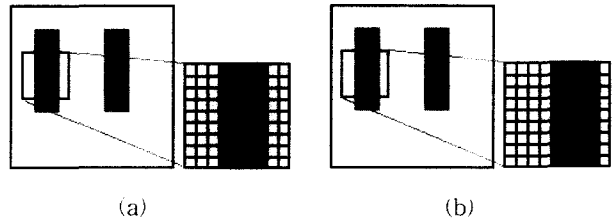


그림 5. 9×9 윈도우 추출 예.
Fig. 5. 9×9's window extraction.

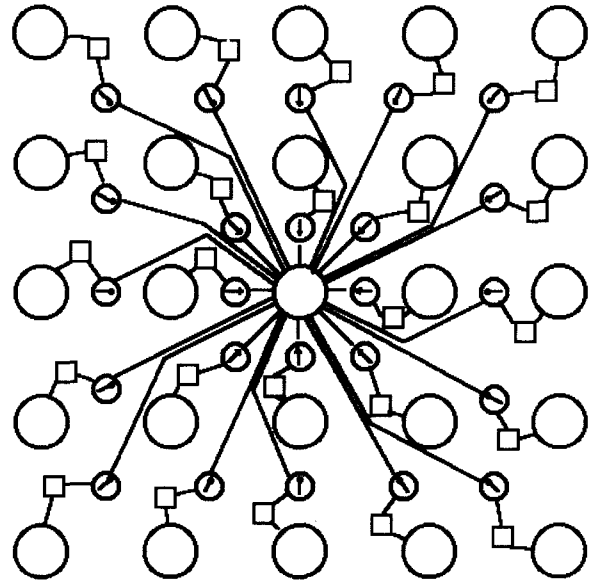


그림 6. m 노드의 구조.
Fig. 6. Structure of m node.

암값들이다. 입력층의 뉴런들의 출력값들은 지연을 갖고 주변 뉴런의 출력값들과 함께 은닉층에 연결되고, 은닉층의 출력값들이 출력층에 총합된다. 이 부분은 파리의 시각 시스템에서 medulla 부분에 해당한다.

연속된 두 차영상으로부터 얻어진 두 9×9 윈도우의 명암값들이 입력층의 각 노드들의 입력값이 된다. 입력층의 노드의 수는 모두 9×9×2하여 162개이다. 입력층 각 노드들의 출력값은 지연치(delay)와 가중치(weight)를 갖고 은닉층으로 각각 연결이 된다.

은닉층에 입력되는 각 값들은 그 좌표에서 동작의 유무를 나타낸다. 즉, 값이 0이 아닌 값을 가지면 동작이 있는 것이고 0이면 동작이 없는 것이다. 이러한 각 좌표를 입력층의 하나의 노드라 가정하고, 그 노드들 주위에 그림 6과 같이 각각 24개의 m 이라는 은닉층의 노드가 있다. 이 노드들의 연결은 기준이 되는 좌표(그림 6의 중심)와 지연된(1 프레임 전의) 주변의 좌표로 연결이 된다. 이 노드에 입력되는 값은 (3)과 같다.

$$x_m = [f(x_{adi})x_{ctr}]^+ \quad (3)$$

여기서 x_{adi} : 주변값, x_{ctr} : 중심값,
 $f(x)$: 지연된(delay) 값,
 $[x]^+$: $\max(x, 0)$

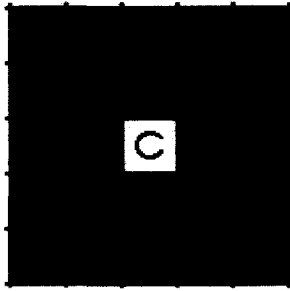


그림 7. 25개의 출력 노드들(C: 중심).
Fig. 7. 25 output nodes (C: Center).

그림 6에서 큰 원은 입력층의 각 노드를 나타내고, 작은 원은 m 노드를 나타낸다. 또한 사각형은 지연(delay)을 나타낸다.

연결되어있는 m 노드의 역할은 움직임의 방향을 감지하는 요소의 역할을 한다. 이 노드가 감지할 수 있는 방향은 인접한 좌표로부터 중심으로의 방향이다(그림 6에서 각 노드 안의 화살표 방향). 인접한 좌표의 값과 중심 좌표의 값의 곱이 0보다 크면 이 방향의 동작 요소를 갖는 것이다. 곱의 값이 0이면 이 방향의 동작 요소가 없는 것이고, 곱의 값이 0보다 작으면 중심 좌표의 동작 요소와 인접한 좌표의 동작 요소가 반대 방향이므로 역시 무시하여도 된다. 이런 m 노드를 9×9 윈도우의 각 좌표들 중에서 가장자리로부터 2 픽셀씩을 건너뛰고 중심 5×5 좌표들에 대해 각각 24개의 m 노드를 만든다. 가장자리 2픽셀씩을 건너뛰는 이유는 중심 좌표에 대해 상, 하, 좌, 우로 2픽셀씩이 있어야 하기 때문이다. 이렇게 만들어진 600개의 m 노드들이 은닉층의 각 노드들이 된다.

출력층 구조는 다음과 같다. 출력층은 중심점을 기준으로 0 pixel/frame의 속도 벡터와 1 pixel/frame의 속도를 갖고 8개의 방향값을 갖는 속도 벡터, 그리고 2 pixel/frame의 속도를 갖고 16개의 방향값을 갖는 속도 벡터까지 모두 25개의 노드를 갖는다. 그림 7에서 밝은 회색 부분이 1 pixel/frame의 속도와 중심으로의 방향을 갖는 8개의 출력 노드들이고 짙은 회색 부분이 2 pixel/frame의 속도와 중심으로의 방향을 갖는 16개의 출력 노드들이다. 0 pixel/frame의 속도를 갖는다는 것은 움직이지 않는다는 것이고, 1 pixel/frame의 속도를 갖는 8개의 방향값은 360° 를 8개의 부분으로 나눈 $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ 의 값들을 갖는다. 그리고 2 pixel/frame의 속도를 갖는 16개의 방향값은 360° 를 16개의 부분으로 나눈 $0^\circ, 27^\circ, 45^\circ, 63^\circ, 90^\circ, 117^\circ, 135^\circ, 153^\circ, 180^\circ, 207^\circ, 225^\circ, 243^\circ, 270^\circ, 297^\circ, 315^\circ$ 의 값들을 갖는다.

4) 전체 동작 분석 (Lobula)

마지막 처리단계는 부분 동작 분석에서 나온 결과를 이용해 전체 영상의 동작을 분석하는 것이다. 이 부분은 파리의 시각 시스템에서 lobula 부분에 해당한다. 구성요소는 부분 동작 분석 부분(medulla)들로 이루어져 있다. 전체 차영상에서 얻어진 9×9 윈도우 모두를 순서대로

부분 동작 분석을 위한 시간 지연 신경망에 입력하여 나온 값들이 이 마지막 단계의 입력값이 된다. (4)와 같이 입력된 속도 벡터값들을 전부 총합한 후 평균을 내어 전체 동작 분석 값을 얻는다.

$$x_{lob} = \sum_i x_{med_i} \quad (4)$$

여기서 x_{lob} : lobula 노드의 출력값,

x_{med} : medulla 노드의 출력값,

med_i : i 번째 medulla 노드.

IV. 실험 및 결과

본 논문에서 제시된 방법을 실험하기 위해 사용된 하드웨어는 영상 획득부와 영상 처리부로 나누어져 있다. 영상 획득부는 CCD 흑백 카메라가 이용되었고 영상 처리부로는 펜티엄 100Mhz CPU와 16Mbyte 메모리를 갖고, BitFlow사의 프레임그래버인 Raptor가 설치된 PC가 이용되었다. 영상 처리부에서는 Watcom C/C++ Ver. 10.0과 Raptor 라이브러리를 이용하여 알고리즘을 구현하였다.

본 논문에서의 실험 방법은 영상을 획득하는 부분과, 합성된 영상으로 학습을 하는 부분, 그리고 학습하지 않은 합성된 영상과 실제의 영상을 넣어 실험하는 부분의 3부분으로 나뉜다. 영상의 구조는 128×128 크기를 가지고, 256 그레이 값을 갖는 헤더 없는 연속영상이다.

영상 학습 부분에서는 합성된 영상으로 학습을 시행하였다. 실제의 영상으로 학습을 하지 않고 합성된 영상으로 학습을 하는 이유는 다음과 같다. 첫째로, 합성된 영상은 정확한 속도 벡터를 갖도록 만들 수 있다. 실제의 영상을 획득하여 사용하는 경우, 이 영상의 정확한 속도 벡터를 알기가 어려우며, 또한 학습 조건에 맞는 여러 개의 영상을 정확하게 획득하기가 매우 어렵다. 둘째로, 합성된 영상을 이용할 경우 원하는 품질의 영상을 만들기가 쉽기 때문에 잡음이 없는 영상뿐만 아니라 임의의 잡음이 첨가된 영상도 학습에 이용할 수 있다.

영상 학습 부분의 절차는 다음과 같다. 우선 임의의 물체가 있는 영상을 만들어 낸다. 그 다음 그 이미지를 1 픽셀, 또는 2 픽셀씩 여러 방향으로 이동시켜 학습 영상을 만들어 낸다. 이 과정을 계속 반복하여 연속 영상을 만들어 낸다. 영상의 종류는 속도가 1 pixel/frame인 경우 360° 를 8방향으로 나눈 $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ 의 8가지와, 속도가 2 pixel/frame인 경우 $0^\circ, 27^\circ, 45^\circ, 63^\circ, 90^\circ, 117^\circ, 135^\circ, 153^\circ, 180^\circ, 207^\circ, 225^\circ, 243^\circ, 270^\circ, 297^\circ, 315^\circ$ 의 16가지가 있다. 그림 8에 제시된 학습영상은 180° 방향으로 1 pixel/frame의 속도를 갖고 있는 합성 영상이다.

학습을 하게 될 신경망의 구조는 다음과 같다. 우선 학습 방법으로는 back propagation 방법을 사용하였다. back propagation 방법은 다층의 구조를 갖는 복잡한 신경망 학습 방법으로 최급하강법(gradient descent method)을 기본으로 한 방법이다. 각 층의 노드들의 연결에는 각각의 가중치들이 있다. 각 뉴런은 연결된 모든

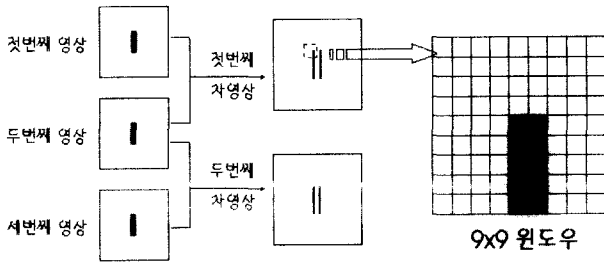


그림 8. 학습 영상의 예.

Fig. 8. An example of image for learning.

입력단위(input unit)의 신호값과 가중치의 곱의 합을 계산한 후, 활성화 함수에 의해 임계치와 비교한다. 이때 많이 사용되는 활성화 함수는 시그모이드(sigmoid) 함수이다

신경망은 입력층, 은닉층, 출력층의 3층으로 이루어졌다. 입력층의 노드 수는 앞서 3.3.3.에서 기술한 바와 같이 9×9×2로 모두 162개의 노드를 갖고, 은닉층의 노드 수는 앞서 말한 m노드들의 총 개수인 600개를 갖는다. 출력층은 모두 25개의 노드를 가지며 각각의 노드는 정지, 1pixel/frame의 이동 방향 8가지, 2pixel/frame의 이동 방향 16가지를 나타낸다.

학습이 끝난 신경망을 대상으로 다음의 3가지 영상들로 나누어 실험을 실시하였다.

- ① 합성된 연속 영상.
- ② 실제 영상을 합성시켜 만든 연속영상.
- ③ 실제 연속 영상.

위 3가지 경우에 대해 25가지의 결과 값들 중의 하나의 결과를 갖는 128×128 크기의 100개의 영상을 획득하였다. ①번 영상의 경우는 사각형, 원, 그리고 사각형과 원의 조합으로 생긴 도형을 만들어 이 영상들을 실험하였다. ②번 영상의 경우는 책과 사람의 얼굴이 들어 있는 영상을 인공적으로 동작이 있게 합성시킨 연속 영상들을 실험하였다. ③번 영상의 경우는 책과 사람 얼굴을 움직여 얻은 실제 연속 영상들로 실험하였다.

실험의 결과는 표 1과 같다. 3가지의 평균 적중률이 94.7%로 산출되었고 한 영상에 대해 평균 수행 시간은 71.47 msec이 소요되었다. 합성시켜 만든 연속 영상이 실제 연속 영상보다 좋은 결과를 보여주었다. 실제 연속 영상의 경우, 속도와 방향에 대한 정확한 값을 측정하기가 어렵고 실제로 얻어진 영상에 대한 잡음과 손실, 보다 랜덤하고 자유로운 물체의 윤곽선, 그리고 3차원적인 평행이동과 회전이동에 의하여 목표 물체의 변형이 가능(non-rigidity)하기 때문에 오차가 비교적 많이 발생한다. 이 표의 결과를 다른 방법의 결과와 비교하기는 매우 어렵다. 특징에 기초한 방법은 그 특징의 복잡성 정도에 따라 정확도와 계산 시간이 좌우된다. 경사에 근거한 방법은 가정이 어떻게 주어지느냐에 따라 정확한 결과를 얻을 수도 있으나, 일반적으로 광류 계산 시간이 다른 두 방법에 비해 매우 오래 걸린다. 하지만 이 표의 결과가 말해주는 본 논문의 장점은 다른 두 방법에 비해 시간이 적게 걸리면서도 원하는 결과에 근접하다는 것

표 1. 3가지 경우에 대한 실험 결과.

Table 1. Experiment Result from 3 cases.

	적 중 륜	평균 계산 시간 (msec)
합성시켜 만든 연속 영상	98 %	71.46
실제의 영상을 합성시켜 만든 연속 영상	96 %	71.44
실제의 연속 영상	90 %	71.50

이다. 하드웨어가 점점 발전하고 있으므로 본 논문의 방법을 더욱 발전시키면 실시간에 가까운 분석 속도를 얻을 수 있을 것이다.

V. 결론

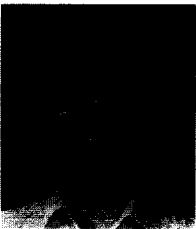
본 논문에서는 연속된 영상에서 동작 분석의 방법에 있어서 기존의 방법들과는 다른, 생물의 시각 시스템을 본뜬 시간 지연 신경망을 이용하는 동작 분석 알고리즘이 제안되었다. 시스템의 전체 구조는 파리의 시각 시스템을 참조하여 설계하였다. 파리의 시스템은 간단하고 모듈화 되었으며, 목표물 추적에 효과적인 구조를 갖고 있다. 제안된 시스템은 기존의 신경망에 시간 지연이라는 시간 특성을 띠는 요소를 집어넣은 신경망을 이용함으로써 어느 순간의 정적인 데이터가 아닌 동적인 데이터가 될 수 있었으며, 동적인 동작 분석을 할 수 있었다. 이런 시간 지연을 이용한 본 논문은 특징에 기초한 방법이나 경사에 기초한 방법에 비해 복잡한 수식이나 많은 계산량이 필요 없다. 단지 이전에 행한 학습만으로 다른 동작 분석 알고리즘에 비해 빠른 속도로 원하는 결과에 근접하는 결과를 얻을 수 있다. 본 논문에서 제시한 동작 분석 방법은 무인 이송 로봇의 장애물 회피, 물체 추적 등에 응용될 수 있다.

앞으로 해결해야 할 향후 과제는 다음과 같다. 첫 번째는 샘플링 비율이다. 본 논문에서 동작 검출을 위해서는 연속되는 영상에서 목표 물체의 위치가 겹쳐야만 하기 때문에 샘플링 비율이 충분히 커야 한다. 하지만 아직 하드웨어적인 제한이 있기 때문에 일정 샘플링 비율 이상 크게 할 수 없으므로 어느 정도의 속도 이상은 검출이 되지 않는다. 하지만 이 점은 현재 영상 획득 분야의 하드웨어와 마이크로 프로세서의 급속한 발전으로 최대 샘플링 비율이 증가되면 신경망의 입력층도 증가되어 해결이 될 것이다. 두 번째는 한 영상 안에서 움직이는 물체는 하나여야만 한다는 점이다. 이 점은 분할(segmentation)기법을 도입하면 해결될 것이라 본다. 세 번째는 회전이동의 검출이 쉽지 않다는 것이다. 이것은 부분 동작 검출 부분에서 나온 결과들을 총합하는 과정에서, Grouping과 총합 방법의 개선으로 어느 정도 나아지리라 본다. 마지막으로 학습 방법을 들 수 있다. 본 논문의 학습 방법인 back propagation 방법은 앞서 언급한 몇 가지의 문제점이 발생할 수 있다. 이 학습 방법을 RBF(Radial Basis Function)를 이용한다거나, 또는 전체 시스템을 ART(Adaptive Resonance Theory)의 구조로 만

들 수 있다면 더욱 향상된 효과를 볼 수 있을 것이다. 특히 이 방법들은 학습이 되지 않은 패턴에 대해서는 결과를 내지 못하는 back propagation 방법에 비해 미지의 패턴이 들어오면 스스로 학습을 해나가는 것이 가능하기 때문에 좀 더 생물학적 시스템에 접근하게 될 것이다.

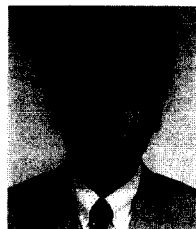
참고문헌

- [1] T. Poggio, "Computer vision," *Image Pattern Recognition*, vol. 755, pp. 54-62, 1987.
- [2] B. K. P. Horn and B. G. Schunk, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
- [3] A. Verri and T. Poggio, "Motion field and optical flow qualitative properties," *Patt. Anal. Machine Intell.*, vol. 11, pp. 490-498, May 1989.
- [4] H. Shariat and K. E. Price, "Motion estimation with more than two frames," *Patt. Anal. Machine Intell.*, vol. 12, pp. 417-434, 1990.
- [5] D. J. Fleet and K. Langley, "Recursive filters for optical flow," *Patt. Anal. Machine Intell.*, vol. 17, no. 1, pp. 61-67, January 1995.
- [6] P. Nesi, A. del Bimbo, and D. Bentzvi, "A robust algorithm for optical flow estimation," *Computer Vision and Image Understanding*, vol. 62, no. 1, pp. 59-68, 1995.
- [7] K. Nakayama, "Biological image motion processing : A review," *Vision Research*, 25, pp. 625-660, 1984.
- [8] T. Chen, W.-C. Lin and C.-T. Chen, "Artificial neural networks for 3-D motion analysis-part I: rigid motion," *Neural Networks*, vol. 6, no. 6, Nov., 1995.
- [9] T. Chen and W.-C. Lin, "Artificial neural networks for 3-D motion analysis-part II: nonrigid motion," *Neural Networks*, vol. 6, no. 6, Nov., 1995.
- [10] B. K. P. Horn, *Robot Vision*, McGraw-Hill, 1987.
- [11] A. Singh, *Optic Flow Computation*, IEEE Computer Society Press, 1991.
- [12] W. C. Lin, E. C. Tsao, and C. T. Chen, "Constraint satisfaction neural networks for image segmentation," *Pattern Recognition*, vol. 25, no. 7, pp. 679-693, 1992.
- [13] R. Hartley, "Segmentation of optical flow fields by pyramid linking," *Pattern Recognition Letters*, vol. 3, pp. 253-262, 1985.
- [14] A. Yamane, N. Ohnishi, and N. Sugie, "Robust extraction of moving objects through grouping edged along with optical flow," *Proc. of the IJCNN*, pp. 1267-1272, 1993.
- [15] J. M. Missler and F. A. Kamangar, "A neural network for pursuit tracking inspired by the fly visual system," *Neural Networks*, vol. 8, no. 3, pp. 463-480, 1995.
- [16] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang. "Phoneme recognition using time-delay neural networks," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no. 3, March, 1989.



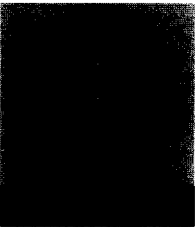
장 동 식

1979년 고려대학교 산업공학과 졸업, 텍사스대 석사, 텍사스 A&M 대학교 산업공학과 박사(1988), 1989년-현재 고려대학교 산업공학과 교수, 관심분야는 컴퓨터 비전 시스템, 패턴인식, 그리고 공장자동화.



이 만 희

1993년 고려대학교 산업공학과 졸업, 동대학원 석사, 동대학원 박사 과정 수료(1997), 1997년-현재 고려대학교 공학기술연구소 연구원, 관심분야는 패턴 인식, 영상 검색, 머신 비전.



이 종 원

1996년 고려대학교 산업공학과 졸업, 동대학원 석사(1998), 현재 (주)넥스트웨이브, 관심분야는 신경망이론, 데이터베이스, 패턴 인식.