

# 신경회로망을 이용한 동적 손 제스처 인식에 관한 연구

論文

53D-1-4

## A Study on Dynamic Hand Gesture Recognition Using Neural Networks

曹寅碩\* · 朴晉玄\*\* · 崔榮奎\*\*\*

(In-Seok Cho · Jin-Hyun Park · Young-Kiu Choi)

**Abstract** - This paper deals with the dynamic hand gesture recognition based on computer vision using neural networks. This paper proposes a global search method and a local search method to recognize the hand gesture. The global search recognizes a hand among the hand candidates through the entire image search, and the local search recognizes and tracks only the hand through the block search. Dynamic hand gesture recognition method is based on the skin-color and shape analysis with the invariant moment and direction information. Starting point and ending point of the dynamic hand gesture are obtained from hand shape. Experiments have been conducted for hand extraction, hand recognition and dynamic hand gesture recognition. Experimental results show the validity of the proposed method.

**Key Words** : Hand Gesture, Neural Network, Pattern Recognition

### 1. 서론

오늘날 현대 사회는 산업자동화 및 정보기술의 급진적인 발전으로 일상 생활 전반에 걸쳐 컴퓨터화된 시스템이 급격히 증가하고 있다. 이로 인해 사람과 컴퓨터 상호간의 정보 교환 수단으로 기존의 인터페이스 방법인 마우스나 키보드가 가지는 의사소통의 제약성을 극복하고 사람과 사람간의 의사소통과 같은 보다 자연스럽게 지적인 인터페이스를 구현하기 위한 연구가 활발히 진행되고 있다[1]. 사람과 컴퓨터간의 상호 의사소통을 구현하기 위한 방법으로 음성인식, 손동작 인식, 얼굴표정 인식 등이 있는데, 사람이 실생활에서 대부분을 손을 통해 사물을 사용하고 있고 손동작이 가지는 빠른 의사전달의 특징과 함축적 의미의 동작을 통해 많은 정보를 전달할 수 있는 특징이 있어, 손동작 인식을 이용한 사람과 컴퓨터간의 인터페이스에 관한 연구가 활발히 진행되고 있다. 손동작 인식은 크게 글러브 데이터(glove-data)에 기반한 방법과 비전(vision)에 기반한 방법으로 나눌 수 있다. 전자의 방법에서 손의 각 부위에 센서를 장치한 장갑을 착용하여, 그 센서의 출력 값을 이용하는 방법으로 3차원 공간 손동작 해석은 가능하지만, 고가의 장비 사용과 센서와 컴퓨터를 연결하기 위한 선이 필요하여 사용자가 불편함을 느낀다. 또한 손동작의 행동반경이 제한되는 등 여러 가지 제약 조건을 수반한다[2]. 후자의 방법은 카메라를 통한 입력 영상을 사용하는 방법으로 3차원 공간 손동작을 해석할 경우, 하나 이상의

카메라와 별도의 보상 알고리즘이 필요하지만, 전자의 방법보다 장비가 간단하며, 행동반경이 자유롭고, 사용자가 불편함을 느끼지 않으면서 자연스러운 손동작이 가능하다.

손동작은 손의 고정된 모양으로 나타내는 정적인 자세(static gesture)와 시간적인 움직임에 의해 나타나는 동적인 자세(dynamic gesture)로 나타낼 수 있다. 이러한 두가지 자세는 영상에서 관심영역인 손모양을 얼마나 잘 추출하고, 그 움직임의 궤적을 얼마나 잘 추적, 인식하는 것이 관건이라 할 수 있다. 동적인 자세 시 움직임의 궤적을 추적하는 방법으로는 움직임만을 배경에서 분리하는 방법과 형태인식을 통해 분리하는 방법이 있다. 첫 번째 방식은 크기 모양에 상관없이 물체를 추적하는 방식으로 연속된 두 영상의 차를 이용하는 차영상(difference image)방법[3]과 영상에서의 밝기패턴의 가시적인 운동으로 정의되는 광류(optical flow)방법[4], 정합 기법에 기초한 블록 정합(block correlation match)방법 등으로 나뉜다[5]. 이러한 방법은 동적 제스처(dynamic gesture)인식 및 추적 시 원하는 관심영역인 손모양만을 추적하기 어렵고, 정확한 손영역의 중심궤적을 구하기가 어려운 단점이 있다. 두 번째 방식은 좋은 인식률과 추적속도가 핵심이다. 최근 하드웨어의 발달로 컬러정보를 이용한 추적에 관한 방법들이 많이 연구되고 있다. 히스토그램 매칭(histogram matching) 기법을 이용하여 컬러영상에서 물체의 색상을 인식하고 추적하는 방법[6], 능동 윤곽선 모델(active contour model)의 한 기법인 snake 방법을 추적 기법에 적용한 방법 등이 제시되었다[7]. 이러한 방법들 또한 동적 자세(dynamic gesture)인식 및 추적 시 원하는 관심영역인 손영역의 색상정보만을 이용할 경우, 얼굴 등의 동일한 색상이 존재하므로 추적에 어려움을 겪게 된다.

본 논문에서는 동적인 손동작 인식을 글러브 데이터에 기반한 방법에 비하여 보다 간단하며, 행동반경에 제약이 없고,

\* 非會員 : 釜山大 電氣工學科 · 工學碩士

\*\* 正會員 : 晉州産業大 메카트로닉스工學科 助敎授 · 工博

\*\*\* 正會員 : 釜山大 電子電氣情報컴퓨터工學部 敎授 · 工博

接受日字 : 2003年 8月 26日

最終完了 : 2003年 11月 21日

사용자가 자연스러운 제스처를 취할 수 있는 비전에 기반한 방법을 사용하였다. 색상정보만을 이용한 손영역의 인식 및 추적의 문제점을 해결하기 위하여, 색상 정보뿐만 아니라 손 모양의 형태인식을 통하여 문제를 해결하였다. 그리고 빠른 실시간 추적을 위하여 손모양을 전역탐색으로 인식한 후, 지역탐색방법을 대치하여 사용하였다. 이상의 손영역의 색상, 형태인식 및 동작 자세의 궤적인식 분류기(classifier)로 일반화 기능과 비선형 맵핑 기능이 우수한 신경회로망을 사용하였다. 제안된 손 제스처 인식 알고리즘의 성능 평가를 위하여, 다양한 동적 제스처로 소형로봇의 주행명령을 구현하고, 이를 실험하였다.

2. 신경회로망을 이용한 손동작 인식

2.1 신경회로망을 이용한 손후보영역 추출

하드웨어 성능의 급진적인 발전으로 인해 이진(binary) 및 그레이 스케일(gray scale) 영상처리뿐만 아니라, 계산시간이 비교적 많이 소요되는 컬러 영상처리방법들이 많이 사용되고 있다. 손인식과정에서 손의 색상정보는 복잡한 배경으로부터 손을 보다 빨리 분리할 수 있어 인식시간을 상당히 줄일 수 있는 장점이 있다. 일반적으로 color clustering 방법이 색상 추출기법으로 많이 사용되고 있다[13]. 그러나, 이 방법은 컬러의 min-max 값에 의해 영상에서 대상을 추출하는 간단한 방법이지만, 컬러 영상은 조명변화에 영향을 많이 받으므로 단순한 min-max clustering에 의해 좋은 결과를 얻을 수 없다. 최근에 이러한 조명효과를 줄이기 위해 HSI방법, Gaussian 분포와 같은 확률적인 방법, 그리고 정규화 RGB(normalized RGB) 컬러공간으로의 컬러 변환 방법 등이 많이 연구되어지고 있다[14-16].

본 연구에서는 비선형 맵핑능력이 우수한 신경회로망을 이용하여 컬러영상에서 손후보영역을 추출하고자 한다. 16bit RGB 공간상에서 피부색상에 해당되는 부분을 학습데이터로 준비하여 각 층의 노드수가 3-5-1의 형태인 신경회로망을 구성하였다. 신경회로망의 학습은 오차역전파학습(error backpropagation learning)방법을 사용하였고, 은닉층에는 바이어스를 추가하였으며, hyperbolic tangent 함수를 활성화 함수로 사용하였다. 학습에 사용된 데이터의 컬러공간은 정규화 RGB(normalized RGB) 공간으로서 식(1)과 같이 표현된다.

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \frac{1}{(R+G+B)} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

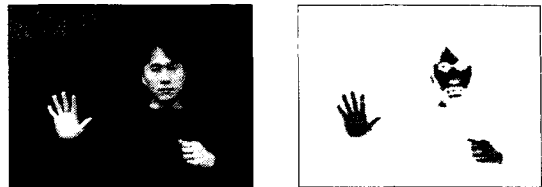
위에서 R, G, B는 original RGB이고, r, g, b는 normalized RGB이다.

그림 1은 컬러 카메라의 입력영상으로부터 신경회로망을 이용하여 손후보 영역을 추출한 이미지를 나타내고 있다.

2.2 영역 레이블링(labeling)과 탐색 영역 지정

그림 1에서 보듯이 영상에는 손과 얼굴 등 동일한 색상을 가진 후보 영역들이 존재하게 된다. 이러한 영상에서 손의

후보 영역을 찾기 위하여 레이블링 알고리즘(labeling algorithm)을 이용하여 각 영역을 구별하였다. 또한 구별된 영역에서 특징정보를 추출하기 위하여 최적의 탐색 영역을 찾아야 한다.



(a) 컬러 영상 (a) Color image  
(b) 손후보 영상 (b) Hand candidate image

그림 1 손후보 영역 추출 영상  
Fig. 1 Extracted image of hand candidates

레이블링 알고리즘은 이진영상 내에서 이웃 화소(pixel)로 뭉쳐있는 영역을 찾아내고, 이러한 영역에 각각의 라벨을 붙임으로써 영역을 구분하는 방법이다[17]. 또한 원하는 크기보다 작은 영역은 제거함으로써 노이즈 제거의 효과도 얻을 수 있다. 그림 2는 그림 1-(b)인 이진영상에서 레이블링 알고리즘과 사이즈 필터링으로 노이즈를 제거한 이미지를 나타내었다.

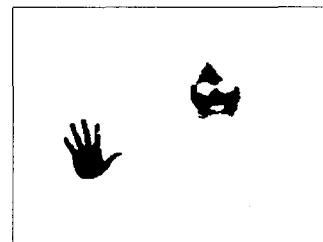


그림 2 레이블링 및 잡음제거 이미지  
Fig. 2 Labeling and noise filtered image

본 논문에서는 입력컬러영상을 신경회로망에 의해 이진영상으로 손후보영역을 추출하고, 이진영상의 손후보영역을 레이블링 알고리즘과 사이즈 필터링을 통해서 손후보영역을 추출하였다.

레이블링 과정을 통해서 원하는 크기의 후보영역을 추출한 후, 각 영역에 대한 적절한 탐색영역을 지정해야 한다. 탐색영역의 지정은 손후보영역들을 정확한 손영역과 손동작 추출을 위한 탐색구간을 미리 결정하여야 한다. 탐색영역을 지정하는 방법으로는 관심물체가 원형이나 정사각형에 가까울 경우 원의 방정식에 의해 간단히 탐색영역을 지정할 수 있으며, 만약 얼굴과 같이 타원형일 경우에는 영역의 2차 모멘트에 의한 영역의 장, 단축 길이정보를 이용하여 탐색영역을 지정할 수 있다. 본 연구에서는 원의 방정식을 이용한 탐색 방법이 영역의 장, 단축 길이 정보를 이용한 방법에 비하여 간단하며 계산시간이 적으므로 탐색영역을 지정하는 방법으로 사용하였다.

아래 식 (2)와 (3)은 각 영역의 중심위치에 대해서 간단히 원의 방정식을 이용하여 반경 r에 대한 탐색영역의 중심위치

와 탐색영역의 크기를 지정하는 식이다.

$$x_c = \frac{1}{A} \sum_{(i,j) \in R} i, \quad y_c = \frac{1}{A} \sum_{(i,j) \in R} j \quad (2)$$

$$r = \sqrt{\frac{A}{\pi}}, \quad \text{Search Area} = (2*(r+10))^2 \quad (3)$$

단, R : region, A : region의 총 pixel수, xc, yc : region center

입력영상에 맺히는 손의 크기는 불규칙적으로 조금씩 변하기 때문에 식(3)에서 보듯이 +10만큼 여유를 주어 탐색영역을 지정하였다.

### 2.3 신경회로망을 이용한 손동작 추출

손동작을 인식하기 위해서 손에 대한 특징정보를 알아야 한다. 따라서 각 영역에 대한 탐색영역이 정해지면 영역 탐색을 통해서 특징정보를 추출하는 작업이 필요하다. 일반적으로 컬러영상에서 손의 형태는 움직임에 따라 크기, 회전, 이동변화가 있으므로 손의 변하지 않는 특징정보를 추출해야 한다. 본 논문에서는 손동작 인식을 통한 실시간 추적을 위해 이동, 회전, 크기 변화에 강인하고 특징정보 수가 적으면서도 손특징을 잘 표현할 수 있는 invariant moment를 이용하여 특징정보를 추출하였다[19]. 입력영상에서 각각의 후보영역에 따른 탐색영역을 binary image  $f(x,y)$ 로 표현할 때  $(p+q)$  차 모멘트는 식(4)과 같이 다시 쓸 수 있다.

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (4)$$

단,  $p, q = 0, 1, 2, \dots$

그리고 식(4)로부터 central moment는 아래와 같은 식으로 나타나게 된다.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (5)$$

단,  $\bar{x} = \frac{m_{10}}{m_{00}}$  and  $\bar{y} = \frac{m_{01}}{m_{00}}$

위의 식을 3차까지의 크기(scale)에 불변하는 moment를 구하기 위하여, 다음 식과 같이 정규화(normalize) 시켰다.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r}, \quad r = \frac{p+q}{2} + 1, \quad p + q = 2, 3, \dots \quad (6)$$

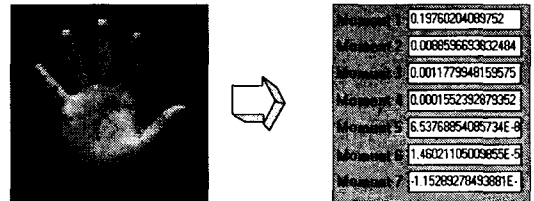
마지막으로 회전에 불변하는 moment는 식(7)과 같이 7개의 함수로 나타낼 수 있다.

각각의 탐색영역으로부터 특징정보를 추출하고, 추출된 특징정보를 인식과 분류하여 손동작을 찾았다. 본 논문에서는 비선형 매핑능력이 좋은 신경회로망을 이용하여 손동작을 인식하였다. 손의 형태적인 특징정보를 학습데이터로 준비하여 오프라인 상태에서 학습시켰다. 손동작 인식을 위한 신경회로망의 구조는 각 층의 노드 수를 7-9-3의 형태로 구성하고,

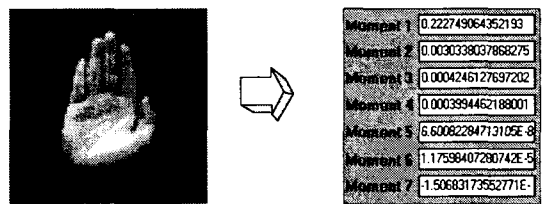
은닉층에는 바이어스를 추가하였으며, hyperbolic tangent 함수를 활성화 함수로 사용하였다.

$$\begin{aligned} \Phi_1 &= \eta_{20} + \eta_{02} \\ \Phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ \Phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \Phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \Phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - (3\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)] \\ \Phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \Phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (7)$$

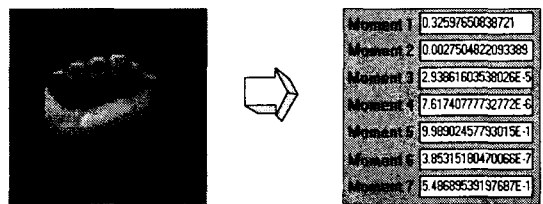
그리고 학습 알고리즘은 역전파 알고리즘을 사용하였다. 출력층의  $y_1, y_2, y_3$ 는 그림 3과 같은 3가지 손동작을 나타낸다. 손동작 1,2는 동적 손 제스처 인식시 시작과 끝을 구분하며, 손동작 3은 소형로봇의 정지명령에 사용된다. 그림 3에서는 이러한 3가지 손동작에 대한 invariant moment를 나타내었다.



(a) 손동작 1  
(a) Hand shape 1



(b) 손동작 2  
(b) Hand shape 2



(c) 손동작 3  
(c) Hand shape 3

그림 3 손동작의 특징정보  
Fig. 3 Feature information of hand shape

### 3. 신경회로망을 이용한 동적 손 제스처 인식

동적 손 제스처 인식은 움직이는 손동작의 궤적을 추적, 인식하는 것이 관건이다. 결국 이러한 추적은 영상으로부터 손과 배경을 분리하는 것으로 볼 수 있다. 손을 추적하는 방법에는 크게 손의 움직임만을 배경으로부터 분리하는 방법과 다른 하나는 손인식의 수정된 형태로서 손을 연속적으로 인식하면서 추적하는 방법이다. 전자의 방법은 영상에서 손이 외에 다른 움직임이 없어야 한다는 제약조건이 따르고 손의 중심을 찾기 어려워 정확한 제스처 궤적을 나타내기가 어렵다는 단점이 있다. 본 연구에서는 후자의 방법을 사용하여 손의 색상뿐만 아니라 형태까지 인식하여 제스처의 시작과 끝을 명확히 하고자 한다. 실시간 추적이 가능하기 위해서는 손을 인식하는 시간이 빨라야 하므로 전역탐색을 통해 여러 후보 중 관심대상을 인식하고, 탐색영역 안에 있는 관심대상만을 인식하는 지역탐색을 통해 추적하게끔 하였다. 또한 동적인 손동작을 인식하기 위한 특징추출방법으로 8-direction code를 사용하였다[8, 9]. 따라서, 동적인 손동작 인식을 위하여 손동작의 움직임을 연속적으로 추적하고, 각 프레임마다 이동한 방향정보값을 신경회로망으로 오프라인 학습하여 동적 손 제스처를 인식하였다. 신경회로망으로 인식된 제스처는 소형로봇에 주행명령을 주도록 하였다.

#### 3.1 전역탐색 및 지역탐색

여러 손후보영역 중 손을 인식하고 난 후, 중심위치를 찾아서 손을 추적해야 한다. 본 논문에서는 비교적 영상처리 시간이 오래 걸리는 전역탐색을 통해서 손은 인식하고, 손이 발견되면 처리시간이 얼마 걸리지 않는 지역탐색을 통해 손을 인식한 후 중심위치를 찾아서 실시간 추적하게끔 하였다. 그림 4에서 첫 번째와 두 번째 이미지는 전역탐색을 통하여 후보영역 추출과 탐색 영역을 지정하고, 인식 분류기를 통해 손을 인식한 것을 보여주고 있다. 그리고 세 번째 이미지는 지역탐색을 통해 손을 인식하여 추적하는 것을 보여주고 있는데, 여기서 손의 좀 더 빠른 추적을 위해 이전 프레임에서 물체의 위치정보를 이용하여 식(8), (9)과 같은 선형식 손 통해 물체의 다음 위치를 미리 예측하여 탐색하게끔 하였다[18].

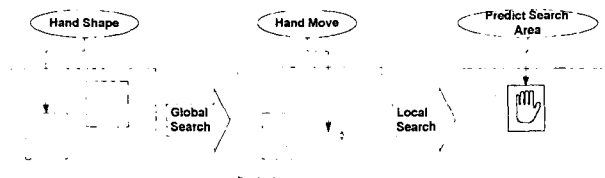


그림 4 전역탐색 및 지역탐색  
Fig. 4 Global search and local search

$$x^p(k) = 2 * x(k-1) - x(k-2) \quad (8)$$

$$y^p(k) = 2 * y(k-1) - y(k-2) \quad (9)$$

여기서,  $x^p(k)$ ,  $y^p(k)$ 는 현재 프레임에서의 손의 예측 위치를 나타내며,  $x(k-1)$ ,  $y(k-1)$ 는 바로 이전 프레임에서의

손의 위치이며,  $x(k-2)$ ,  $y(k-2)$ 는 현재 프레임 보다 2 프레임 전의 손의 위치를 나타낸다.

만약, 지역탐색을 통하여 손을 추적하는 중에 손이 지역탐색영역을 부분적으로 벗어나서 인식이 안 될 경우, 손과 같은 색상에 의해 겹쳐지거나 안 보일 경우, 손의 색상정보만을 이용하여 손영역을 찾은 후 다시 지역탐색을 통해 손을 인식 추적하게끔 하였다. 그러나 손이 탐색영역을 많이 벗어나거나 완전히 사라진 경우에는 다시 전역탐색을 실시하였다.

#### 3.2 신경회로망을 이용한 동적 손 제스처 추출

동적 손 제스처를 인식하기 위하여 8-direction code를 사용해 방향정보의 특징값을 추출하였다. 8-direction code는 실시간 추적을 통해 나타나는 손 중심점 위치의 이동 방향을 코드로 나타낸 것이다. 제스처의 이동 방향각을 8방향으로 구분하여, 현재 프레임과 이전 프레임의 이동 방향각을 구한 것이다. 또한, 정규화 된 히스토그램을 사용하여 손의 순간적인 변화보다 전체적인 변화를 나타내어 손의 순간적인 움직임에 대한 오차가 줄어들도록 하였다. 방향각은 그림 5와 같이 0°에서부터 45°씩 더하여 360°까지 8개의 기준 방향값을 정하고 기준 방향값에서 다시 ±22.5°씩 더한 영역을 최종 방향값으로 설정하였다.

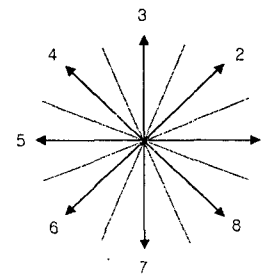


그림 5 8-방향 코드  
Fig. 5 8-direction code

본 논문에서 사용된 동적 손 제스처 인식 신경회로망의 구조는 그림 6과 같다. 각 층의 노드수는 8-15-12의 형태로 구성되어 있으며, 은닉층에는 바이어스를 추가하였고, hyperbolic tangent 함수를 활성화 함수로 사용하였다. 그리고 학습 알고리즘은 역전파 학습 알고리즘을 사용하였다. 출력층의 G1~G12는 본 논문에서 인식하는 12가지 동적 제스처를 나타낸다

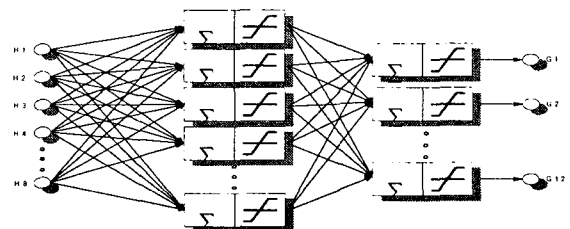


그림 6 동적 손 제스처 인식을 위한 신경회로망  
Fig. 6 Neural networks for dynamic hand gesture recognition

4. 실험 및 검토

4.1 전체 시스템 개요

그림 7은 본 연구에서 제안하는 동적 손 제스처 인식의 전체 구성도이다.

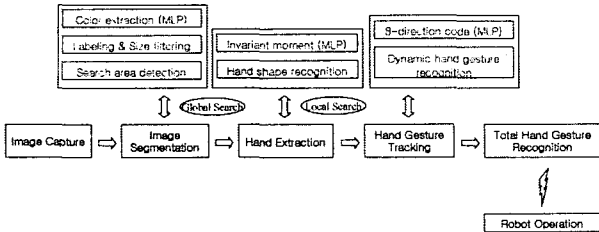


그림 7 전체 시스템 개념도

Fig. 7 Conceptual diagram of the total system

제안하는 방법은 탐색영역, 손영역 추출, 손동작인식(정적 및 동적)으로 크게 3부분으로 나뉘어진다. 손의 색상정보와 형태정보 그리고 방향정보를 이용하여 보다 빠르고 정확한 동적 제스처 인식을 적용하였다. 또한, 인식된 동적 제스처에 따라 소형 로봇을 원격으로 제어하여 실험하고자 한다. 본 논문에서는 2차원의 손모양을 인식하고, 추적하는 것을 목적으로 하기 때문에 손은 항상 카메라의 정면에 위치하여 실험하였다.

4.2 실험 장치 및 환경

본 실험에서 사용한 CCD 칼라 카메라와 이미지 캡처 보드의 간략한 사양은 표 1과 같다.

표 1 실험의 사양

Table 1 Specification in experiments

종류	사양 및 특징
Camera	<ul style="list-style-type: none"> <li>· 41만 화소 CCD 칼라 카메라</li> <li>· 유효화소수(H×V) : NTSC : 768×494, PAL : 712×582</li> <li>· 최저조도 : 0.5 Lux(F : 1.2)</li> <li>· 전원 소비전력 : 12V ± 10% DC, 220mA</li> </ul>
Image capture board	<ul style="list-style-type: none"> <li>· 30 frame/sec capture rate(OnAir 16ch capture board)</li> <li>· 16 채널 캡처 기능</li> <li>· 다양한 image resolution, NTSC, PAL 등 비디오 포맷 지원</li> </ul>
기타 실험 환경	<ul style="list-style-type: none"> <li>· pentium IV 1.7GHz, RAM 256MB의 main PC</li> <li>· Window 98 SE, Dephi 5의 제어 환경 및 언어</li> </ul>

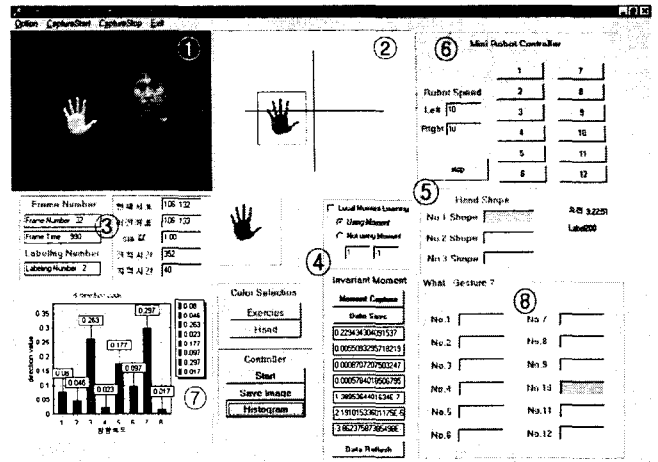


그림 8 시스템 제어 창

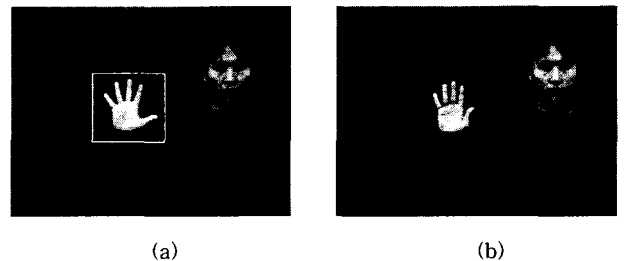
Fig. 8 System control window

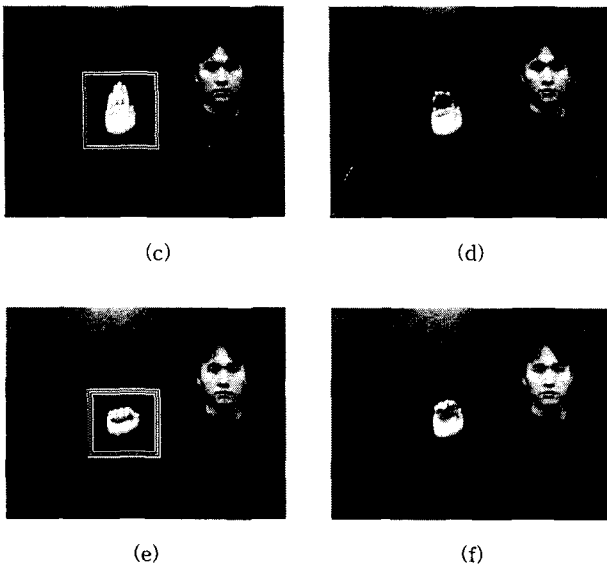
또한 본 실험을 위한 영상 제어 창은 그림 8과 같다. 그림에서 ①은 캡처된 원 이미지를 display하며, ②는 전역 탐색 및 지역 탐색을 통한 손영역만을 인식, 추적하여 표시하였다. ④는 손동작을 인식하기 위한 형태 특징정보인 invariant moment를 실시간으로 계산하여 나타내며, 이렇게 인식된 손동작은 ⑤에 나타내었다. ③은 레이블링 수, 전역탐색과 지역탐색에 걸린 시간 및 프레임 시간 등을 나타내며, ⑦은 동적 손 제스처 특징값인 정규화된 방향값을 히스토그램으로 나타내었다. 이러한 동적 손 제스처는 ⑧에서 최종적으로 표시하였다. ⑥은 소형 로봇을 움직이기 위한 제어창으로, 실제 실험에선 ⑧에서 인식된 제스처로 로봇에 주행 명령을 구현하였다.

4.3 실험 결과

4.3.1 손동작 인식 실험

그림 9는 정적인 3가지 손동작 인식실험 결과이다. 그림 9의 (a), (c), (e)는 정적인 3가지 손동작을 정확히 인식하며, 나머지 (b), (d), (f)는 인식하지 않음을 알 수 있다. 손동작 1은 그림 9의 (a)에서와 같이 1줄의 사각형으로, 손동작 2는 그림 9의 (c)에서와 같이 2줄의 사각형으로, 손동작 3은 그림 9의 (e)에서와 같이 3줄의 사각형으로 표시하였다.





(c) (d)  
(e) (f)

그림 9 손동작 인식  
Fig. 9 Hand shape recognition

4.3.2 동적 손 제스처 인식 실험

동적 손 제스처 인식을 통하여 소형로봇에 주행 명령을 내 려 얼마나 잘 인식하는가를 실험하고자 한다. 소형 로봇의 주행 명령은 표 2와 같이 설정하였다. 그림 10은 소형 로봇 의 주행 명령을 위한 동적 손 제스처의 특징정보인 정규화된 방향값을 히스토그램으로 나타내었다.

표 2 동적 손 제스처 집합(a)  
Table 2 Dynamic hand geture set(a)

No.	Gesture	Command	No.	Gesture	Command
①	→	우회전	⑦	⊥	3방향으로 가기
②	←	좌회전	⑧	M	전·후진하기
③	↑	전진하기	⑨	↷	제자리 턴
④	↓	후진하기	⑩	✕	빨리 가기
⑤	↙	지그재그로 가기	⑪	↶	좌회전 후 전진
⑥	△	원형으로 가기	⑫	↷	우회전후 전진

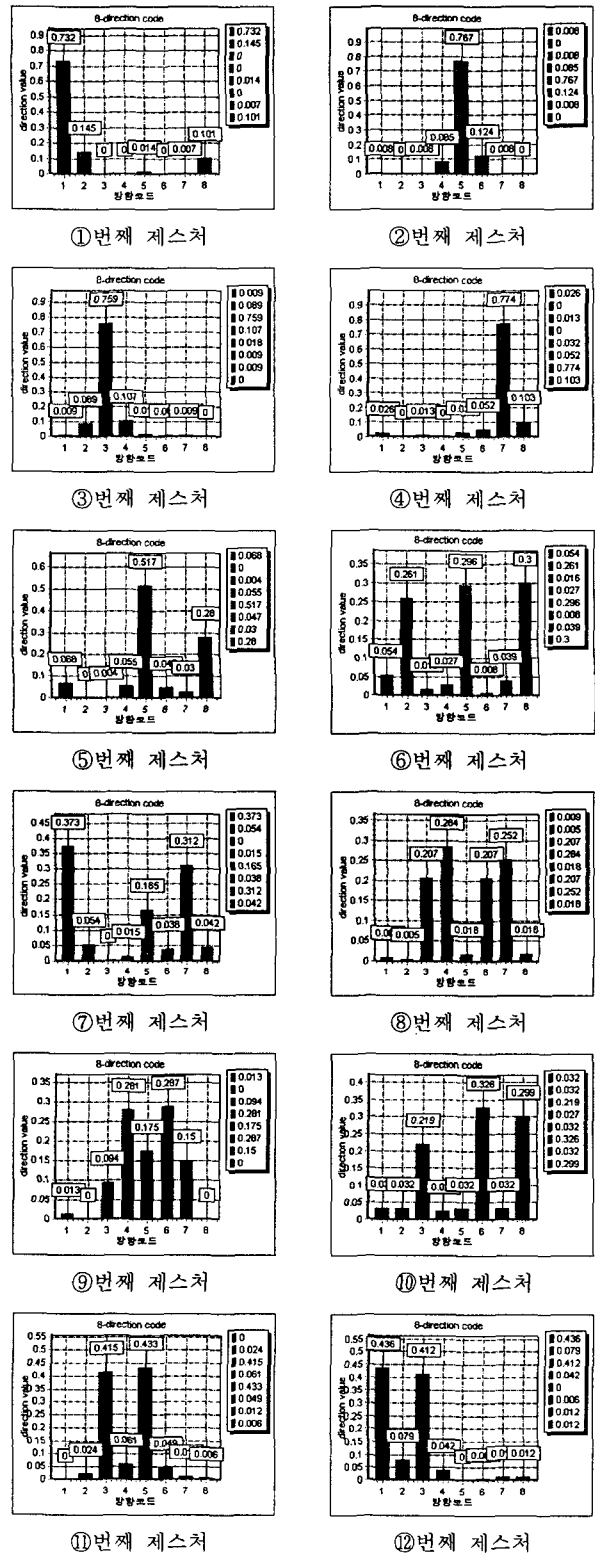
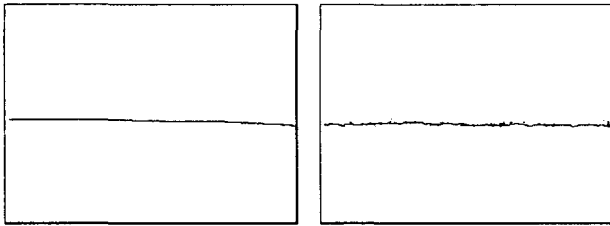


그림 10 동적 손 제스처의 특징정보  
Fig. 10 Feature information of dynamic hand gesture

본 연구에서 표 2의 손동작의 궤적을 정확하게 추적하는 것이 인식률을 높이고, 동적 손 제스처 인식을 위한 신경망의 학습시간을 단축시킬 수 있다. 그러므로 손동작 궤적 추적을 비교하기 위하여, 제안된 색상 및 형태를 학습시킨 신경망과

일반적인 차영상 방법을 사용하여 비교하였다. 비교 시 정의된 제스처를 사람이 수행할 경우, 항상 동일하게 행할 수 없으므로 항상 일정한 속도로 움직이는 축구로봇을 사용하여 좌에서 우로 움직이도록 하였다. 그림 11은 등속도 운동을 하는 축구로봇의 매 프레임에서의 중심좌표를 도시하였다. 그림 11의 (a)는 본 논문에서 제안한 방법을 사용하였으며, 그림 11의 (b)는 연속된 두영상의 차를 이용한 차영상방법을 사용한 결과이다. 그림에서 알 수 있듯이 제안된 방법이 축구로봇의 궤적을 보다 정확히 추적함을 보이며, 이는 제안된 방법이 보다 정확한 특정정보를 얻을 수 있음을 알 수 있다.



(a) 색상 및 형태를 학습한 경우 (b) 차영상방법의 경우  
 (a) The case of learning color and shape (b) The case of difference image method

그림 11 물체중심의 추적점  
 Fig. 11 The tracking point of object center

아래 그림들은 12개의 동적 손 제스처 중 4가지의 결과를 나타내었다. 그림 12와 13은 1번째 제스처의 인식과 인식된 명령에 대한 소형 로봇의 주행 영상을 나타내었다. 그림 14과 15는 4번째 제스처에 대한 결과이며, 그림 16과 17는 8번째 제스처, 그림 18와 19은 12번째 제스처의 인식과 소형로봇의 주행 영상을 나타내었다.

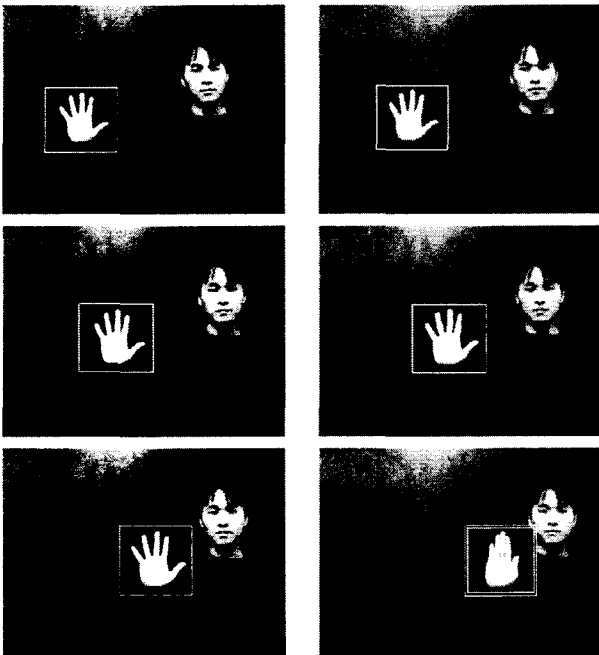


그림 12 첫 번째 동적 손 제스처 인식  
 Fig. 12 The 1st dynamic hand gesture recognition

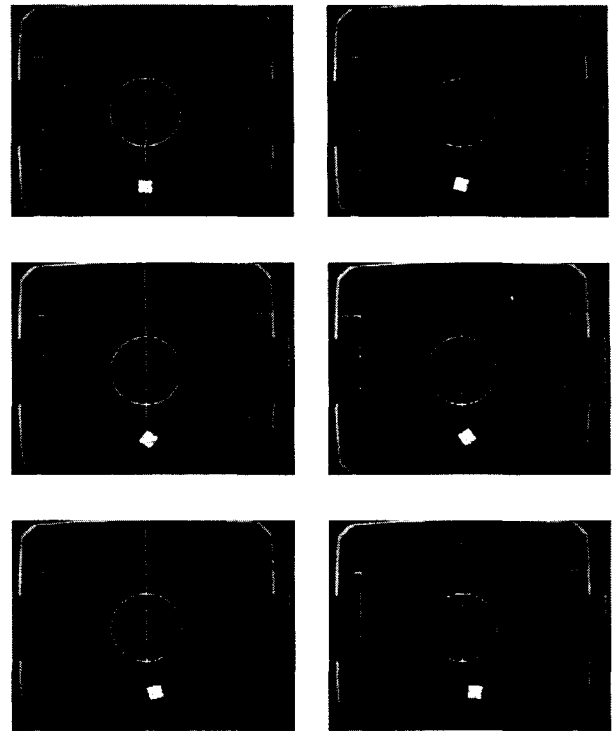


그림 13 첫 번째 제스처의 소형로봇 주행  
 Fig. 13 The mini-robot navigation of the 1st gesture

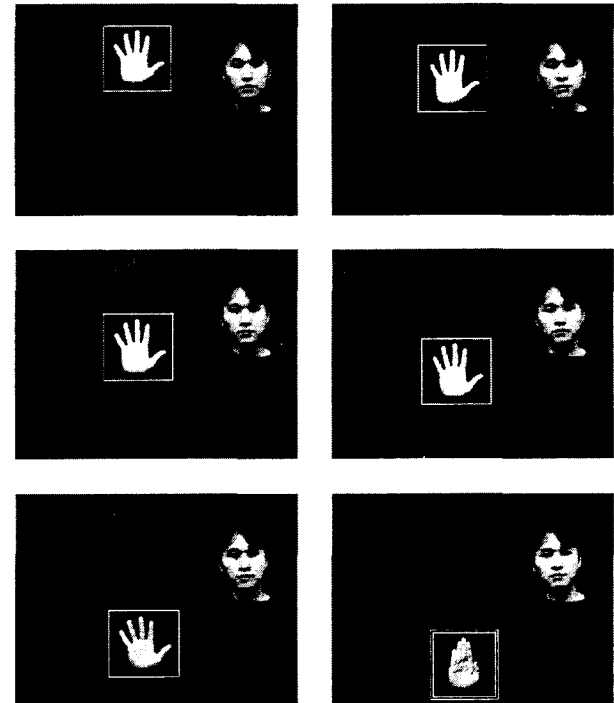


그림 14 네 번째 동적 손 제스처 인식  
 Fig. 14 The 4th dynamic hand gesture recognition

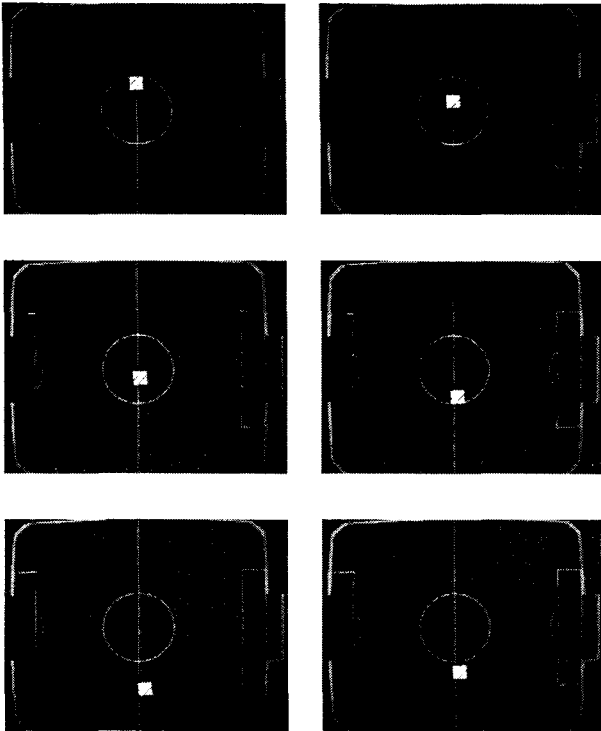


그림 15 네 번째 제스처의 소형로봇 주행  
 Fig. 15 The mini-robot navigation of the 4th gesture

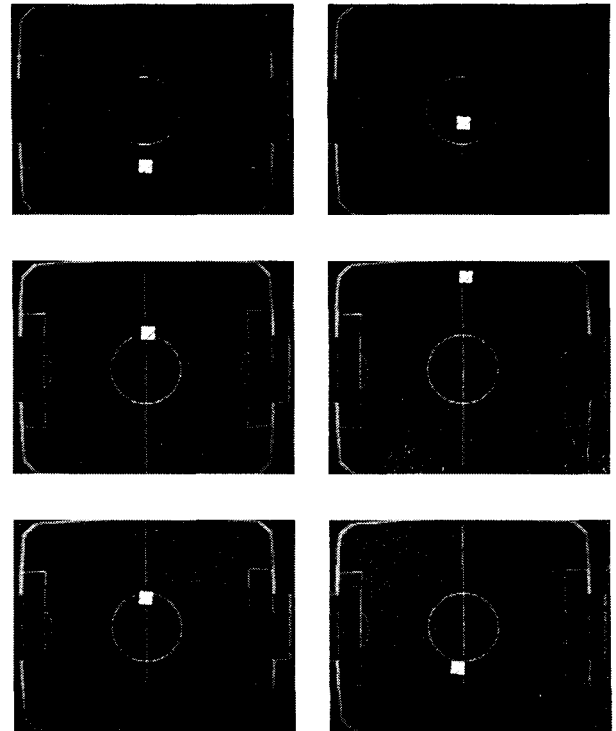


그림 17 여덟 번째 제스처의 소형로봇 주행  
 Fig. 17 The mini-robot navigation of the 8th gesture

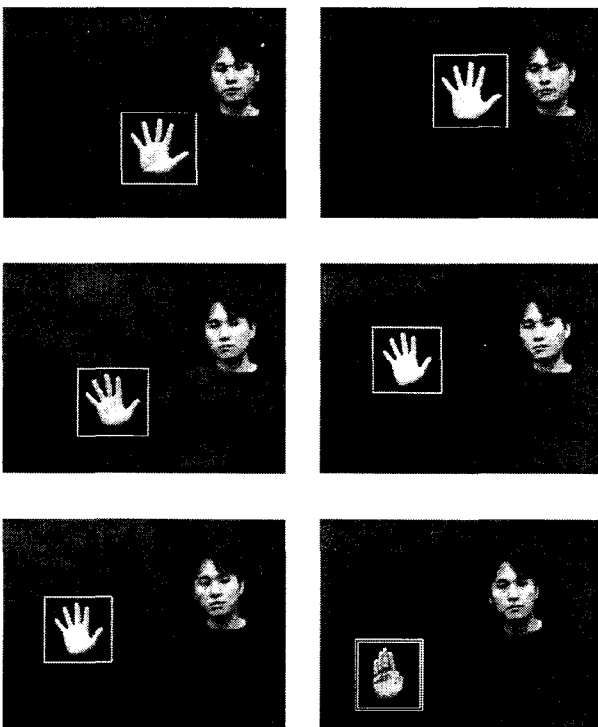


그림 16 여덟 번째 동적 손 제스처 인식  
 Fig. 16 The 8th dynamic hand gesture recognition

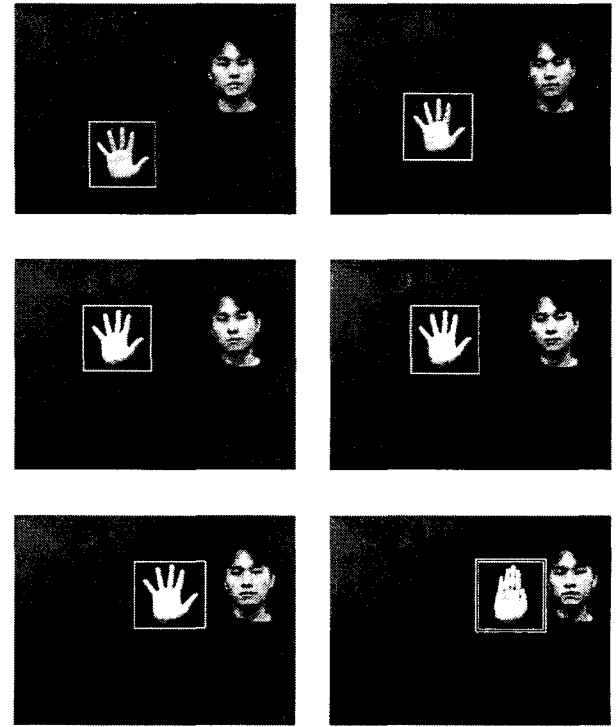


그림 18 열 두 번째 동적 손 제스처 인식  
 Fig. 18 The 12th dynamic hand gesture recognition



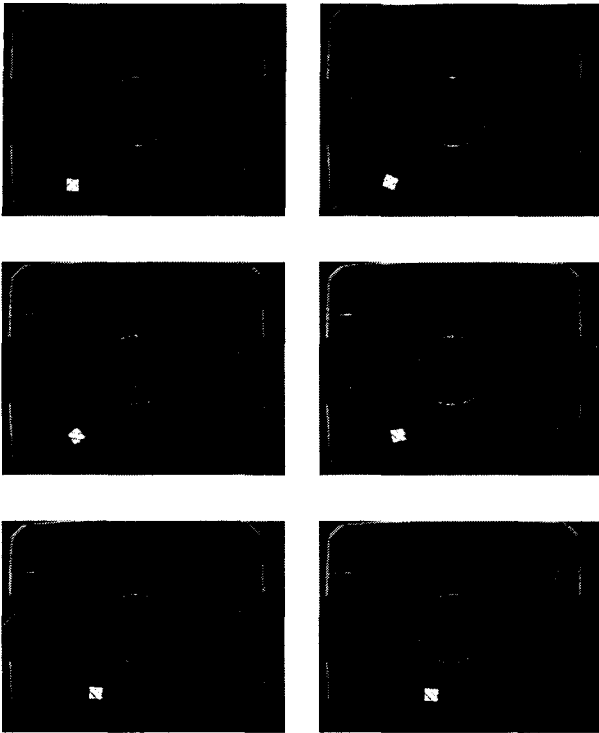


그림 19 열 두 번째 제스처의 소형로봇 주행  
 Fig. 19 The mini-robot navigation of the 12th gesture

동적 손 제스처의 인식은 얼마나 빨리 정적인 손동작을 인식하고, 정확히 이동 궤적을 추적하느냐이다. 손을 인식하는데 걸린 시간은 전역탐색일 경우에는 약 400ms 정도 소요되며, 지역탐색일 경우 약 80ms 정도 소요되었다. 또한, 실험에서 100% 손동작 인식률을 보이지 못하였으며, 이는 그림자 영역의 존재와 배경 속에 피부색상이 많은 경우 등으로 확인되었다. 그러므로 실험 시간대에 따른 적절한 색상 학습값을 신경회로망으로 학습시켜 이를 보정하였으며, 고정된 실험 환경 하에서 90% 이상의 인식률을 보였다.

### 5. 결 론

본 논문에서는 빠른 손동작 인식을 통한 동적 손 제스처 인식 방법을 제안하였다. 본 논문의 손동작 인식 과정은, 손의 색상정보를 이용하여 후보영역을 추출하고, 레이블링과 탐색영역 지정 방법을 통해 각 후보영역에 대하여 적절한 탐색영역을 찾아 형태 특징정보를 추출하며, 인식 분류기를 통하여 손동작을 인식하는 것으로 되어 있다. 이러한 인식과정은 전역탐색에 해당되며 지역탐색은 전역탐색을 통해 인식된 손의 크기에 따른 탐색영역을 통해 특징정보를 추출하여 손동작을 인식하게 하였다. 이렇게 인식된 손동작은 초당 30프레임의 속도로 동적 제스처의 특징정보를 추출하여, 비선형 매핑능력이 우수한 신경회로망을 이용함으로써 조명변화에도 강인한 색상추출성능을 얻을 수 있었다. 또한 손동작의 invariant moment 특징정보와 동적인 제스처의 8-방향 코드

특징정보 및 신경회로망을 통하여 인식률을 높일 수 있었다. 그리고 간단한 선형 예측식을 통해 보다 빨리 손동작을 추적할 수 있도록 설계하였다. 최종적으로 동적 제스처를 통한 다양한 의미의 표현들을 소형로봇 주행명령에 적용하였다.

본 연구는 2차원 형상만을 인식하는 것을 목적으로 하기 때문에 좀더 많은 의미의 의사 전달이나 행동 표현을 하기 위해서, 두 개 이상의 카메라를 사용하여 3차원의 형상 및 동적인 제스처를 인식하는 방법의 연구가 필요하다.

### 참 고 문 헌

- [1] D. M. Gavrila, "The visual analysis of human movement: a survey." *IDEAL Computer Vision and Image Understanding*, vol. 73, No. 1, pp. 82-98, January 1999.
- [2] D. J. Sturman, "A survey of glove-based input," *IEEE Computer Graphics & Applications*, pp. 30-39, Jan. 1994.
- [3] E. Y. Cha and M. H. Kang, "Multiple target tracking in clutter backgrounds using self-organizing map," *IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 1162-1165, 1998.
- [4] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1, 2, Addison Wesley, 1993.
- [5] P. A. Laplante and A. D. Stoyenko, *Realtime Imaging : Theory, Techniques and Application*, IEEE Press, 1996.
- [6] T. M. Chen, Ren C. Luo and T. H. Hsiao, "Visual tracking using adaptive color histogram model," *Proceedings of IECON*, vol. 3, pp. 1336-1341, 1999.
- [7] W. Kim and J. J. Lee, "Visual tracking using snake algorithm based on optical flow information," *Proceedings of the 14th KACC*, pp. E13-E16, 1999.
- [8] A. McAulay, A. Coker and K. Saruhan, "Effect of noise in moment invariant neural network aircraft classification," *Proceedings of NAECON*, pp. 743-749, 1991.
- [9] H. K. Lee and J. H. Kim, "An HMM-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, No. 10, October 1999.
- [10] 오창석, 뉴로 컴퓨터, 지성 출판사, pp. 1-13, 1996.
- [11] J. S. R. Jang, C. T. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1997.
- [12] 임영도, 퍼지·신경망·유전진화, 인솔미디어, pp. 107-144, 1999.
- [13] R. Kjeldsen and J. Kender, "Finding skin in color images," *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pp. 312-317, 1996.
- [14] Y. Kanai, "Images segmentation using intensity and color information" *Proceedings of the IEEE Conference*

color information” *Proceedings of the IEEE Conference on Visual Communications and Image Processing*, 1997.

- [15] L. Jordao, M. Perrone and J.P. Costeira, “Active face and feature tracking,” *Proceedings of International Conference on Image Analysis and Processing*, pp. 572-576, 1999.
- [16] J. C. Terrillon, M. David and S. Akamatsu, “Automatic detection of human faces in natural scene images by use of skin color model and of invariant moment,” *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 112-117, 1998.
- [17] R. M Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1, Addison Wesley, 1993.
- [18] S. Birchfield, “An elliptical head tracker,” *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp.1710-1714, 1997.
- [19] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison Wesley, pp. 518-522, 1993.

저 자 소 개



**조 인 석 (曹寅碩)**

1974년 4월 20일생. 2001년 동아대 전기공학과 졸업. 2003년 부산대 대학원 전기공학과 졸업(석사).  
Tel : 051 - 510 -1445  
E-mail :happycho@hananet.net



**박진현 (朴晋玄)**

1967년 3월 2일생. 1992년 부산대학교 전기공학과 졸업. 1994년 동대학교 대학원 전기공학과 졸업(석사). 1997년 동대학교 대학원 전기공학과 졸업(공학박). 1997년~1999년 부산정보대학 전기전자계열 전임강사. 1999년~현재 진주산업대학교 메카트로닉스공학과 조교수.  
Tel : (055) 751-3386, Fax : (055) 751-3389  
E-mail : jh-park@jinju.ac.kr



**최영규 (崔榮奎)**

1957년 6월 18일생. 1980년 서울대학교 전기공학과 졸업. 1982년 한국과학기술원 전기 및 전자공학과 졸업(석사). 1987년 동 대학원 전기 및 전자공학과 졸업(공학박). 1990~1991년 CALTECH Visiting Scholar. 1998~1999년 Univ. of Southwestern Louisiana Visiting Scholar. 현재 부산대 공대 전자전기정보컴퓨터공학부 교수.  
Tel : (051) 510-2371, Fax : (051) 513-0212  
E-mail :ykichoi@pusan.ac.kr