

스테레오 비전에서 대응문제 해결을 위한 알고리즘의 개발

임 혁 진*, 권 대 갑**

Development of an Algorithm for Solving Correspondence Problem in Stereo Vision

Hyuck-Jin Im*, Dae-Gab Gweon**

ABSTRACT

In this paper, we propose a stereo vision system to solve correspondence problem with large disparity and sudden change in environment which result from small distance between camera and working objects. First of all, a specific feature is divided by predefined elementary feature. And then these are combined to obtain coded data for solving correspondence problem. We use Neural Network to extract elementary features from specific feature and to have adaptability to noise and some change of the shape. Fourier transformation and Log-polar mapping are used for obtaining appropriate Neural Network input data which has a shift, scale, and rotation invariability. Finally, we use associative memory to obtain coded data of the specific feature from the combination of elementary features. In spite of specific feature with some variation in shapes, we could obtain satisfactory 3-dimensional data from corresponded codes.

Key Words : Stereo Vision(스테레오 비전), Neural network(신경회로망), pattern recognition(패턴인식), associative memory(연상 기억 방법), threshold learning(역치 학습)

1. 서 론

공장 자동화 즉 인간을 대체할 수 있는 기술의 개발에 있어서 가장 기초가 되는 것은 인간이 느낄 수 있는 것을 얼마나 유사하게 기계가 느낄 수 있도록 하는가 하는데 있다. 특히 시각기술에 의한 3차원 정보의 획득은 공장 자동화에 있어 작업환경에 대한 적응성을 증대시킬 수 있어 그 중요성이 커지고 있다. 본 연구에서는 정상적인 조명, 간단한 장비를 사용하여 3차원 정보를 추출할 수 있는 스테레오 비전 방법(1, 2, 3, 4, 7, 8)으로 시각 시스템을 구축하고자 한다. 그러나 일반적으로

로 스테레오 비전의 적용시 가장 큰 문제인 대응문제의 해결 없이는 처리시간 등에서의 제한때문에 실제 자동화에 적용하기 어렵게 된다. 일반적으로 대응문제의 해결을 위한 방법으로 양쪽 화면에 대한 상관도(correlation)를 이용한 방법(area-based method)과 양쪽 화면에서의 특징에 대한 속성을 비교하여 대응점을 찾는 방법(feature-based method)이 있다. 본 연구에서 구현하고자 하는 시스템은 로봇의 팔에 부착되어 작업물체의 3차원 정보를 얻을 수 있도록 되어야 한다. 이런 시스템에서 일반적으로 제기될 수 있는 문제점은 카메라와 작업물체간의 거리가 가까기 때문에 양쪽 카메라

* 한국과학기술원 정밀공학과

** 한국과학기술원 정밀공학과 (정회원)

라의 화면에서 대응점의 변위가 크게 되어 대응점 선정이 어렵다는 것과 카메라의 위치가 계속 변하므로 카메라와 작업물체간의 보정이 계속적으로 요구되어 진다는 것이다. 이런 요구 상황에서 기존의 스테레오 대응방법 중 상관도에 근거한 방법은 화상특징(feature)에 정향된 전처리 과정이 수행되는 것이 아니라 전체 화면영역에 대한 동일한 필터 연산을 수행하게 되므로 좌우측 영상에서의 대응점 결정에 있어서 불필요한 연산 또는 제약점을 갖게 된다(5, 6, 9). 또한 특징점에 근간을 둔 방법은 좌우 화면에 대해 대응시켜야 할 점의 변위(disparity)가 클 경우 해당 대응점을 쉽게 결정할 수 없게 되며 또한 피쳐(feature)의 변위가 커질수록 피쳐의 형상 변화가 심해져서 올바른 대응인지 판단이 어렵게 된다. 결국 본 연구에서는 빠른 수행속도를 위해 화상특징에 근거하여 대응점이 찾아졌으며, 변위에 따른 피쳐의 형상 변화를 극복하기 위해서는 환경의 변화에 적응성을 갖는 신경회로망이 이용되었다. 또한 본 시스템에 적합한 즉, 새로운 대상(작업 물체)에 대한 빠른 학습과 기존의 학습된 결과에 대한 재학습이 필요없이 추가가 용이하도록 한 신경회로망 구조가 제안되었다.

2. 3차원 영상 인식 시스템의 구성

스테레오 비전에서 가장 큰 문제인 대응문제 - 특히 로봇의 팔에서 처럼 작업환경의 거리가 가깝고 화면의 변화가 심한 시스템에 대한 문제 - 를 해결하기 위해서 신경회로망의 방법이 사용된 전체 시스템을 Fig.1에 나타내었다. 그림에서 처럼 카메라로부터 받은 영상에서 주위 환경등의 영향에 의한 노이즈를 줄이기 위한 전처리 과정과 신경회로망에 적합한 입력 형태인 회전, 이동, 크기에 독립적인 영상을 얻는 시스템이 설계되었다. 또한 여기에 피쳐의 특징에 따라서 분류

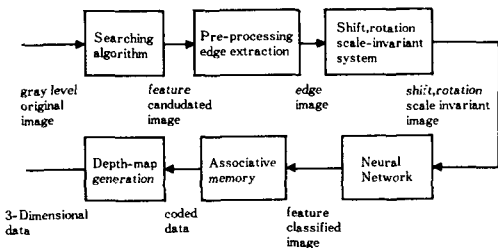


Fig.1 Blockdiagram of the entire stereo vision system

(classification)를 해주는 신경회로망 구조와 이 분류된 결과에 대해 특징값을 인가해 줄 수 있는 연상기억방법(associative memory)으로 시스템을 구성하였다. 최종적으로 이 대응점을 이용하여 삼각법에 의한 3차원 거리정보(depth map)를 얻을 수 있게 하였다.

2.1 전처리 과정

이 과정은 원래의 영상에서 오차 요소로 작용하는 빛의 강도(intensity), 그림자 등의 영향을 줄이기 위한 것으로 Marr과 Hildreth에 의한 Laplacian 필터를 사용하여 에지(edge)를 추출하도록 하였다. 이것의 원리는 원래의 영상에 대역통과필터(band pass filter)를 씌워서 특정 주파수 영역의 특징 피쳐를 추출하는 것이다. 우선 Gaussian 형태를 갖는 저역통과필터(low pass filter)는 다음과 같이 된다. 여기서 $g(x, y)$ 는 (x, y) 좌표에서의 빛의 강도(intensity)를 의미한다.

$$g(x,y) = \exp[-(x^2+y^2)/2\pi\sigma^2] \text{ ----- [1]}$$

이것의 Fourier 변환 형태는 다음과 같이 된다.

$$G(\Omega_x, \Omega_y) = 2\pi^2\sigma^2 \exp[-\pi\sigma^2(\Omega_x^2 + \Omega_y^2)/2] \text{ ---- [2]}$$

여기서 σ 에 의해서 저역통과필터의 차단 주파수(cutoff frequency)를 결정할 수 있다. 이것의 부분이차미분(partial second derivative) $\nabla^2(g(x, y))$ 를 구하면 대역통과필터가 된다. 원래이 영상 $f(x, y)$ 와 ∇^2g 와의 포션(convolution)연산은 선형 연산(linear operation)이므로 다음과 같이 된다.

$$\begin{aligned} \nabla^2(f(x,y) ** g(x,y)) &= f(x,y) ** \nabla^2g(x,y) \\ &= f(x,y) ** \left(\frac{\partial^2 g}{\partial x^2} \right. \\ &\quad \left. + \frac{\partial^2 g}{\partial y^2} \right) \text{ ----- [3]} \end{aligned}$$

여기서

$$\nabla^2g(x,y) = \frac{(x^2+y^2-2\pi\sigma^2)}{(\pi\sigma^2)^2} \exp[-(x^2+y^2)/(2\pi\sigma^2)] \text{ [4]}$$

이 되고 이것의 Fourier 변환을 구하면 다음의 대역통과필터가 된다.

$$F[\nabla^2g(x,y)] = -2\pi^2\sigma^2 (\Omega_x^2 + \Omega_y^2) \exp[-\pi\sigma^2(\Omega_x^2 + \Omega_y^2)/2] \text{ ----- [5]}$$

여기서 σ^2 의 값에 의해서 대역통과필터의 폭을 결정할 수 있게 된다. 이 방법을 사용하면 모서리(edge)가 되는 영점통과위치(zero crossing point)와 빛의 강도(intensity) 변화의 방향등의 정보를 얻을 수 있게 된다. 여기서는 경계(edge) 부분에서의 왜곡을 줄이기 위해 지역적 분산도(local variance)를 구해 이것이 일정 역치와 비교해 클 경우 경계로 확정하게 된다. 이 방법에 의한 전체 전처리 알고리즘을 Fig.2에 나타내었다(13) (16).

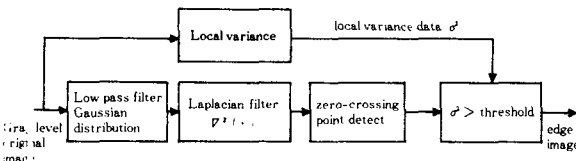


Fig.2 Blockdiagram of preprocessing process for edge extraction

2.2 크기, 이동, 회전에 대한 불변양 발생 시스템

신경회로망에 적합한 입력형태를 얻기 위해서는 크기, 이동, 회전에 불변이 되도록 하는 시스템을 구성하여야 한다. 이 부분의 작동에 관한 블록 선도는 Fig.3과 같다. 이 시스템의 입력은 앞의 전처리 과정을 통과한 이치화된 값이므로 $N_1 \times N_2$ 의 영상에 대해서 다음과 같이 된다.

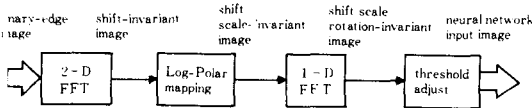


Fig.3 Blockdiagram of the invariant system-shift, scale and rotation

$$f(n_1, n_2) = \begin{cases} 1, & \text{경계점} \\ 0, & \text{경계점 이외의 점} \end{cases} \quad [6]$$

단 $0 \leq n_1 \leq N_1$
 $0 \leq n_2 \leq N_2$

이 2차원 영상으로부터 이동불변인 변환 영상을 얻기 위해 우리는 원래의 영상을 주기적 확장(periodic extention) 하였으며 이것을 위한 방법으로 Fourier 변환을 사용하였다. 즉 이 2차원 영상에 대한 DFT (Discrete Fourier Transform)를 구하면 다음과 같이 된다.

$$F(k_1, k_2) = \begin{cases} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f(n_1, n_2) \exp(-j \frac{2\pi}{N_1} k_1 n_1) \\ \times \exp(-j \frac{2\pi}{N_2} k_2 n_2) \\ \text{단 } 0 \leq n_1 \leq N_1-1 \\ \quad 0 \leq n_2 \leq N_2-1 \\ 0 \quad \text{위 영역 이외의 경우} \end{cases} \quad [7]$$

다시 여기서 크기(scale)에 독립적인 영상을 얻기 위해 아래의 식처럼 크기와 모양을 각각 독립적인 축으로 분리시킬 수 있는 로그-극좌표 대응(log-polar mapping)을 하고 여기서 얻은 결과에 대해 크기(scale) 축방향으로의 정규화를 수행하였다.

$$F^*(\ln r, \theta) = F(k_1, k_2) \quad \text{단 } r = 1 + \sqrt{k_1^2 + k_2^2}, \quad \theta = \tan^{-1}(k_2/k_1) \quad [8]$$

마지막으로 회전에 독립적인 영상을 얻기 위해서 앞에서 로그-극좌표 대응된 영상에 대해서 각 r에 대하여 각도 방향으로 주기적 확장(periodic extension)을 위한 1차원 DFT를 하면 다음과 같이 된다.

$$Z(\ln r, w) = \begin{cases} \sum_{n_3=0}^{N_3-1} F^*(\ln r, n_3) \exp(-j \frac{2\pi}{N_3} w n_3) \\ \text{단 } 0 \leq n_3 \leq N_3-1 \\ 0 \quad \text{위영역 이외의 경우} \end{cases} \quad [9]$$

이렇게 하여 최종적으로 얻어진 $Z(\ln r, w)$ 는 회전, 크기, 이동에 대해서 불변인 값일 뿐만 아니라 근본적으로 Fourier 변환 자체가 일원변환(unitary transform)이므로 각각의 출력단은 직교(orthogonal) 즉 서로 독립적인 값이 된다. 그리고 Fourier 변환의 형태는 여러 형상의 물체가 한 영역내에 들어오더라도 각각 독립적으로 처리해서 더한 것과 그 결과가 같게 되므로 위치나 회전에 상관없이 한 영역내에 여러 형상이 들어와도 분류가 가능하게 된다(14) (15) (18).

2.3 피쳐 인식을 위한 신경회로망의 적용

2.3.1 신경회로망의 문제점

현재의 물리적 접근 방법으로는 한계에 이른 영상 처리 기법의 해결을 위해 고도의 병렬 처리구조, 적용

형 학습능력, 결합 극복능력을 갖는 신경회로망의 기법을 본 시스템에 도입하였다. 신경회로망은 현재의 입력과 기존에 학습된 가중치(weight)와의 내적에 의하여 그 크기를 구하고 이것을 역치와 비교하여 그 결과를 출력하는 형태이다. 이런 기본적인 뉴론(neuron)을 병렬로 연결하면 복잡한 형태의 선형적 결합이 가능하게 되며 이것을 직렬로 연결할 경우에는 각 뉴론 끝에 있는 역치(threshold) 함수에 의해서 비선형적인 시스템의 구현이 가능하게 된다. 결국 신경회로망은 이런 기본적인 단위를 시스템의 특성에 따라 어떻게 구성할 것인가 하는 문제와 어떤 방식으로 가중치를 결정할 것인가 하는 것이 시스템 성능의 관건이 된다고 할 수 있다. 신경회로망은 그 구조적인 특성에 의해 영상처리에 적용시 기존의 영상처리 모델에 비해 많은 장점을 갖고 있음에도 불구하고 기존의 신경회로망으로 시각 시스템을 구현(10, 11)했을 경우 대상체의 이동, 회전, 비틀림 그리고 크기 변화에 적용할 수 있는 시스템 구현의 어려움이 있다(12).

2.3.2 신경회로망의 설계와 실제 시스템에서의 역할

일반적인 신경회로망이 가중치를 학습하고 역치를 고정시키는데 반하여 본 시스템은 특정 분류 위주가 되므로 특정 작업 물체를 그 피처에 따른 기본요소로 효율적으로 분해시킬 수 있고 빠른 학습속도 등을 얻기 위해서 미리 가중치를 직교(orthogonal) 형태로 주로 역치를 학습하는 형태로 신경회로망의 구조를 설계하였다. 여기서 직교 가중치 벡터들을 결정하는데 있어 최적의 방법은 입력 벡터간의 상호관계를 고려하여 구해져야 하지만 여기서는 계산의 복잡성을 피하기 위해서, 실수, 대칭, 직교하고 단지 ±1의 이원(binary)의 값만을 사용하므로 승산등의 계산이 필요하지 않아 빠른 변환 알고리즘이 가능하게 된 Hadamard 변환 방법에 의해서 고정된 직교(orthogonal)한 가중치를 사용하였다. 길이가 N인 입력 영상을 u(m)이라고 할 때 이것의 1차원 Hadamard 변환은 다음과 같다.

$$\begin{aligned}
 \text{출력 영상 } v(k) &= \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} u(m)(-1)^{b(k,m)} \\
 , \quad 0 \leq k \leq N-1 & \quad [10] \\
 \text{단, } b(k,m) &= \sum_{i=0}^{n-1} k_i m_i \quad , \quad k_i, m_i = 0,1
 \end{aligned}$$

여기서 k_i, m_i 는 k 와 m 의 이원표현(binary representation)으로 다음과 같다.

$$k = k_0 + 2k_1 + \dots + 2^{n-1}k_{n-1} \quad [11]$$

$$m = m_0 + 2m_1 + \dots + 2^{n-1}m_{n-1} \quad [12]$$

다음에는 앞에서 처럼 결정된 가중치와 각 학습할 영상 입력의 내적에 대한 값의 역치를 결정하는 알고리즘이 필요하게 된다. 여기서 역치를 결정한다는 것은 들어오는 입력신호를 그 분포 상태에 따라 효율적으로 분리시킬 수 있는 값을 결정하는 것이다. 내적에 의해 구해진 값에 대한 시제열적으로 일련의 내적값 크기에 따른 분포도(histogram)가 각 특징 주변에 Gaussian 분포를 나타내게 될 것이라 가정한다. 이들 Gaussian 분포를 실제 시제열에 의한 분포값과의 차이를 오차로 정의하여 반복적(iterative)인 방법으로 역치를 결정하도록 하였다. 이때 분포를 Gaussian 형태로 가정한다면 초기 상태의 추정값과의 차이가 큰 입력 시제열 값은 학습에 대한 영향이 적으므로 초반의 학습 속도가 느리게 된다. 이것을 방지하기 위해 Gaussian 분포보다 넓은 분포를 갖는 분포 함수를 아래와 같이 정의하였다.

$$g(\theta_i) = \frac{1}{1 + (\theta_i - I_i)^2} \quad [13]$$

단, θ_i : i번째 반복(iteration)에서 추정(seed point)값
 I_i : i번째 반복에서 입력 값

i번째에서의 오차함수(error function)의 값은 아래와 같이 정의할 수 있다.

$$\begin{aligned}
 f(\theta_i) &= g(I_i) - g(\theta_i) \\
 &= \frac{(\theta_i - I_i)^2}{1 + (\theta_i - I_i)^2} \quad [14]
 \end{aligned}$$

여기서 오차함수 $f(\theta_i)$ 가 이중 미분가능하고 콘벡스 함수(convex function)이므로 스티피스트 디센트 방법(steepest descent method)를 사용하여 반복 알고리즘을 결정한다. 여기에 수렴 속도의 향상을 위해 새로이 분포함수를 정리하면 아래와 같이 된다.

$$g(\theta_i) = \frac{t_i^2}{t_i^2 + (\theta_i - I_i)^2} \quad [15]$$

단, t_i : 감쇄값

이에 따른 오차함수(error function)는 아래와 같이 된다.

$$f(\theta_i) = g(I_i) - g(\theta_i) = \frac{(\theta_i - I_i)^2}{t_i^2 + (\theta_i - I_i)^2} \quad [16]$$

이것을 스티피스트 디센트 방법(steepest descent method)에 대입하면 아래와 같이 된다.

$$\begin{aligned} \theta_{i+1} &= \theta_i + \alpha_i \nabla f(\theta_i) \\ &= \theta_i + \alpha_i \frac{\partial f(\theta_i)}{\partial \theta_i} \\ &= \theta_i + \alpha_i \frac{\partial}{\partial \theta_i} \left(\frac{(\theta_i - I_i)^2}{t_i^2 + (\theta_i - I_i)^2} \right) \\ &= \theta_i + \alpha_i \frac{g(\theta_i)}{t_i^2} (\theta_i - I_i) \end{aligned} \quad [17]$$

앞에서 제안한 것처럼 신경회로망의 구조는 가중치를 학습하는 것이 아니라 역치를 학습하게 하여 여러 형상이 대한 학습을 빠르게 수행 할 수 있게 하였다. 그리고 특징 분류에 대한 비선형적인 특성을 부여하기 위해 다층으로 구성해야 하는데 이것을 위해서 처음의 층(layer)은 Fourier 변환에 의한 직교 단(orthogonal node)에 대한 역치값의 조정으로 나타내었고 두번째 층은 Hadamard 매트릭스에 의한 직교 벡터(orthogonal vector)에 대한 역치를 조정하는 형태로 구성하였다.

앞의 학습 알고리즘에 의해서 처음 층에서는 역치가 $\tau^{(1)}$ 로 두번째 층에서는 역치가 $\tau^{(2)}$ 로 결정 되었다고 하면 실제 작업 과정에서 2.2에서 보인 불변양 시스템(invariant system)을 통과한 영상에 대한 인식과정은 다음과 같다.

불변양 시스템을 통과한 i 번째 단의 값을 z_i 라 할 때 Hadamard 매트릭스에 의한 직교 가중치를 위한 입력 x_i 의 형태는 Hadamard 매트릭스의 연산을 단순화하기 위해 ± 1 이 되어야 하므로 다음 형태로 된다.

$$x_i = \begin{cases} 1 & \text{단, } z_i \geq \tau_i^{(1)} \\ -1 & \text{위 영역이외의 경우} \end{cases} \quad [18]$$

Hadamard 변환을 수행한뒤 얻어진 j 번째 단의 값을 y_j 라 하면 두번째 역치를 통과한 출력 p_j 는 다음과 같이 구해진다.

여기서 k 값은 비선형성을 주는 시그모이드(sigmoid) 함수의 기울기를 결정할 수 있게 한다. 피쳐인식을 위해 제안된 신경회로망의 구성을 Fig. 4에 나타내었다(12) (16) (17).

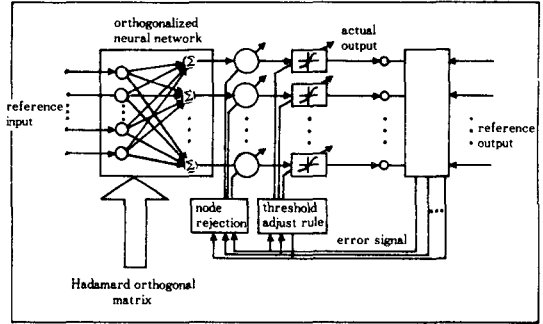


Fig. 4 The structure of the newly designed neural network for feature extraction (Pre-defined Orthogonal net)

2.4 연상기억에 의한 코드화

앞의 신경회로망에 의해서 얻어진 특징 피쳐에 독립적인 출력을 실제 스테레오 비전 시스템에 규격화된 형태로 적용하기 위해서는 특징 코드 값으로의 변환이 필요하게 된다. 우선 신경회로망을 통과한 각 단에서의 출력이 의미를 갖기 위해서는 특징 피쳐에 대한 각각의 출력 단에서의 값이 +1이나 0에 집중된 분포형태를 나타내야 한다. 여기서 분포된 정도에 따라 그 출력단의 특징 피쳐에 대한 중요도를 알 수 있으므로 넓게 분포하는 단은 제거하고 좁게 분포하는 경우는 1을 연상기억장소(associative memory)에 저장시킨다. 이렇게 함으로서 연상기억장소의 가중치는 특정한 한 피쳐에 대해서 현재 입력으로 들어온 피쳐가 같은지를 판별할 수 있게 해준다. 만일 여러개의 피쳐가 동시에 입력으로 들어오면 각각 해당 연상기억장소의 출력단을 1로 하므로 이것을 통해 여러 기본 형상의 조합인 특징 피쳐인식이 가능하게 된다(17).

2.5 스테레오 알고리즘에 의한 3차원 정보 추출

양쪽 카메라에 의해서 얻어진 영상에 대해 앞의 과정을 거쳐 각 피쳐에 대한 특징 코드값을 얻은 뒤 이 코드값의 대응을 통해 3차원 정보를 얻을 수 있게 된다. 일반적으로 거리(depth)를 구하는 계산의 단순화를 위하여 양쪽 카메라의 수평 스캔라인(scan line)이 배

이스라인(base line)과 평행하다(epipolar constraint)는 가정하에서 거리를 결정하도록 하였다. 이렇게 되면 수직방향의 변위는 고려하지 않고 단지 수평방향의 변위만을 고려하면 되므로 계산이 단순화하게 된다. 그런뒤 보간을 통하여 대응점 주변의 대응되지 못한 점의 좌표를 보간에 의해서 전체 영상에 대한 3차원 거리정보(depth map)를 완성하게 된다.

3. 실험 및 검토

본 시스템은 한 화면 내에서 여러 형상이 섞여 있거나 겹쳐져 있는 경우 그 각 형상을 구성하는 피처를 분류한뒤 이 피처의 조합으로 구성될 수 있는 형상을 연상기억방법(associative memory)을 사용하여 인식하도록 하였다. 그래서 실험은 특징 기능에 중점을 두었으며 최종적으로 양쪽의 카메라에서 얻은 두 영상에 대해 특징점을 코드화 시키고 이것으로 대응점을 찾아 3차원 정보를 얻는 형태로 실험을 진행하였다. 본 연구를 위한 전체 시스템의 하드웨어 구조는 Fig. 5와 같다. 처리시간의 감소를 위해서 특히 불변량 발생 시스템에서 소요되는 시간의 감소를 위해 처리 영역의 제한이 필요하므로 본 시스템의 처음에 대상점의 선택을 위한 알고리즘을 추가한다.

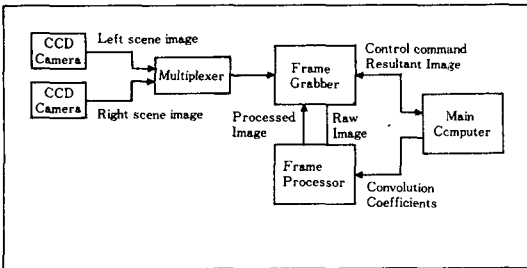


Fig.5 Configuration of stereo vision system

3.1 전처리 과정의 실험

단순한 형태의 영상인 Fig. 6에 대해서 앞의 2.1에서 제시한 Laplacian of Gaussian 방법에 의해 얻어진 에지정보를 Fig. 7에 나타내었다. 지역분산(local variance) 값에 특정 역치값과 비교하여 경계선을 결정해서 Fig. 7에서 볼 수 있는 것처럼 한 선에 의해서 에지가 표현되므로 피처 추출에 있어 왜란의 영향을 줄일 수 있음을 알 수 있다.

3.2 회전, 이동, 크기 불변량 발생 시스템의 실험

이동, 회전 크기에 독립적인 출력을 얻기 위한 시스템에 대한 성능을 알기 위해 원과 사각형을 각각 이동시키고 회전시킨 경우에 대해서 2차원 FFT(fast fourier transform)와 로그-극좌표 대응(log-polar mapping)한 결과에 대해서 Fig. 8에서 Fig. 10까지 나타내었다. 그림에서 볼 수 있는 것처럼 대상현상이 한 화면 내에서 이동이나 회전에 상관없이 독립적인 피처의 중첩(superposition)에 의해 표현됨을 알 수 있다.

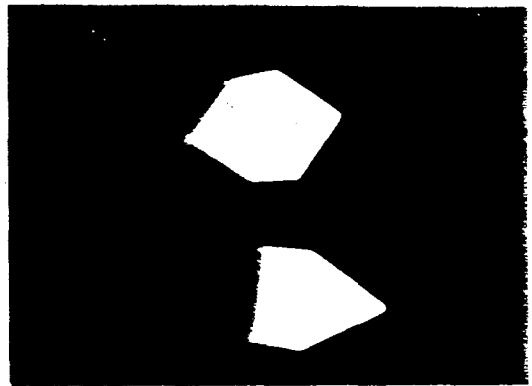


Fig.6 Original image for edge extraction

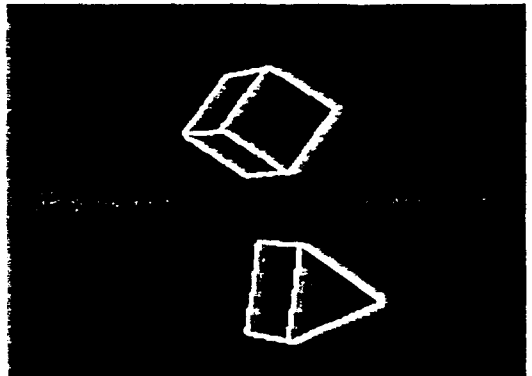


Fig.7 Edge image of Fig.6 by Laplacian of Gaussian filtering

3.3 신경회로망의 학습능력

신경회로망에서 역치를 학습하는 알고리즘이 잘 적용되는 지를 알아보기 위해서 두개의 임의의 값을 중심으로 히스토그램이 Gaussian 분포를 갖는 일련의 신호를 주고 신경회로망에 의한 역치값과 최적역치 결정 방법에

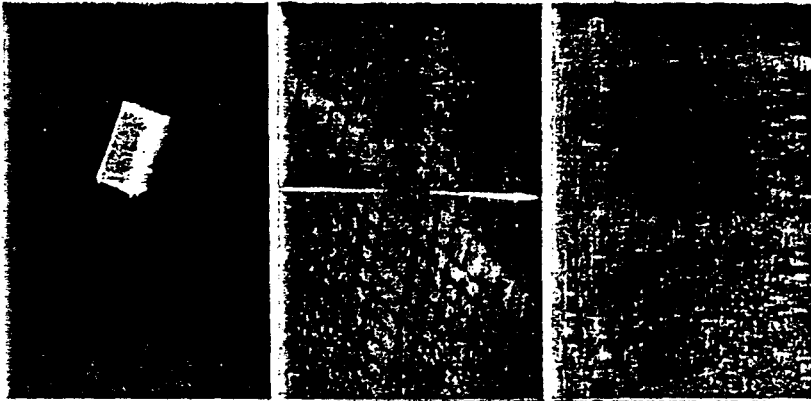


Fig.8 Transformed images by invariant system for rectangular shape
1) original image, 2) Fourier transformed image and
3) log-polar mapping image

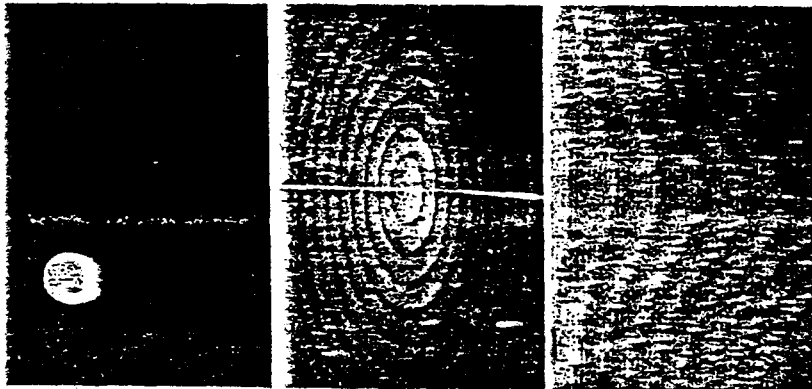


Fig.9 Transformed images by invariant system for circular shape
1) original image, 2) Fourier transformed image and
3) log-polar mapping image



Fig.10 Transformed images by invariant system for rectangular shape and circular shape
1) original image, 2) Fourier transformed image and
3) log-polar mapping image

의한 값과 비교하였다. Fig. 11에서 볼 수 있는 것처럼 Threshold Inputdata의 그림은 두개의 Gaussian 분포를 갖는 일련신호의 히스토그램을 나타낸 것이고 아래의 Seed Distribution은 학습에 의해 두 분포의 대표값을 찾아낸 것을 나타낸다. 그림에서 알 수 있듯이 통계적 계산에 의한 최적역치값 44.7518과 학습에 의한 역치값 44.508이 거의 일치함을 알 수 있다. 실제 구분하기 어려운 경우에 대해서도 잘 분류되는 것을 볼 수 있다(19).

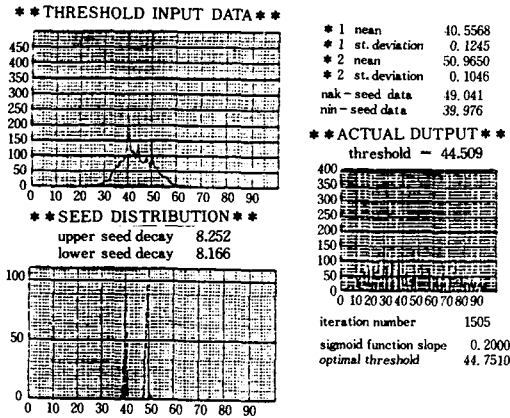


Fig. 11 Performance simulation of neural network for classifier Two input sequences with Gaussian distribution are divided as shown figure by neural classifier

3.4 학습 영상에 의한 기본 피쳐의 인식

Fig. 12는 신경회로망 학습을 위한 기본 예지의 형상과 불변량 발생 시스템을 통과한 후의 각각의 변환된 영상을 겹쳐서 주파수 축에 따라 나열된 것을 나타낸다. 각각 3개의 패턴에 대한 변환 결과를 옆의 도표에 겹쳐서 나타냈음에도 불구하고 거의 일치됨을 볼 수 있다. 이것에 의해 학습되어진 결과로서의 역치를 Fig. 13에 나타내었다. 여기서 처음의 역치는 Hadamard 매트릭스의 직교 가중치를 통과하기 전에 불변량 발생 시스템의 출력 영상에 대한 비선형성을 주기 위한 것이고 두번째의 역치는 Hadamard 매트릭스의 직교 가중치를 통과한 후의 변환 영상에 대해 비선형성을 주기 위한 것이다. Fig. 14에서 일반적인 예지 형상에 대한 회전과 이동을 수반했을 때의 분류 결과를 Fig. 15에 나타내었다. 여기에서 볼 수 있는 것처럼 직선 예지, 3개의 선에 의한 코너, 2개의 선에 의한 코너등의 80개의

입력에 대한 리콜(recall) 결과는 1개의 오류를 ; 외하고는 잘 분류되는 것을 볼 수 있다.

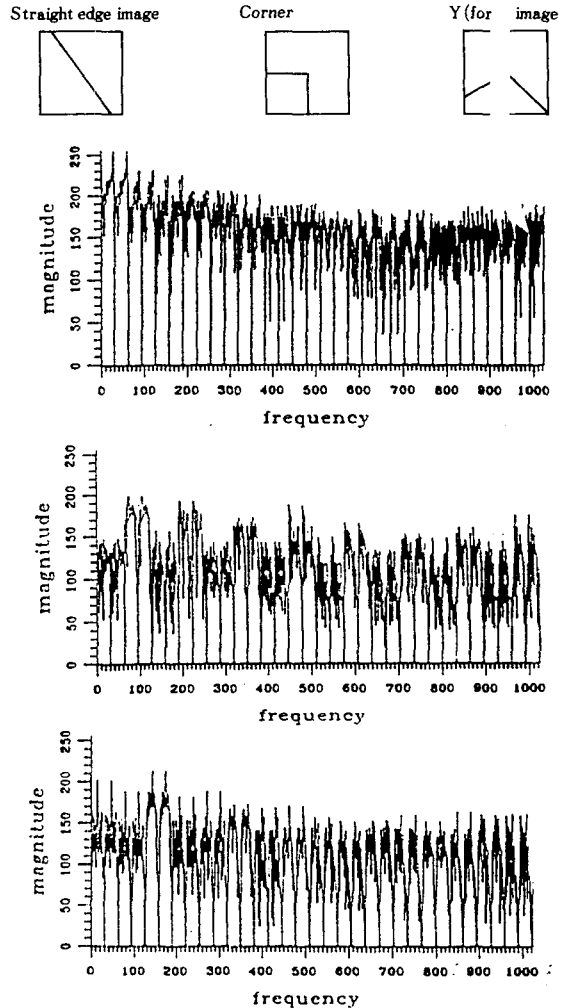


Fig. 12 1) Learning patterns for edge extraction
2) Results by invariant system for some of the straight edge
3) Results by invariant system for some of the corner
4) Results by invariant system for some of the fork

3.5 스테레오 비전 시스템에의 적용

앞의 3.4절에서 얻은 실험 결과를 갖고 좌우 카메라에 의한 입력 영상과 이것을 신경회로망에 의한 처리결과를 Fig. 16에 나타내었다. 처리된 결과의 명확성을 위

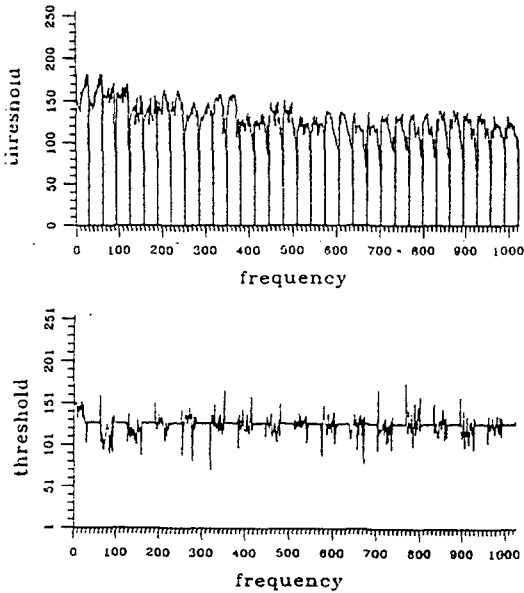


Fig.13 Learning results of the neural network

- 1) First stage threshold for input of orthogonal weight
- 2) Second stage threshold for output of orthogonal weight

히서 원래의 영상과 코드화 된 영상을 겹쳐서 나타내었다. 처리된 영상에서 밝은 점(code=255)은 3선이 만나는 코너, 그리고 그 다음으로 밝은 점(code=192)은 2선이 만나는 코너를 나타낸다. 실제 작업 물체와 카메라

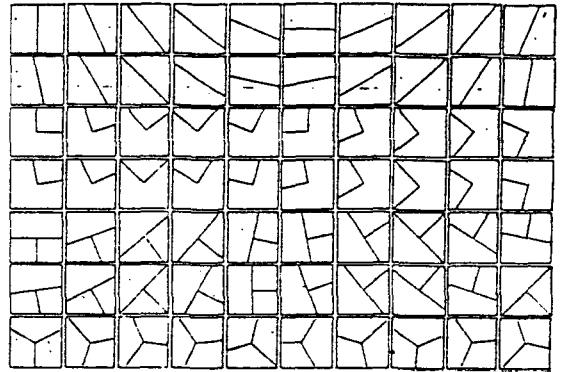


Fig.14 Input data for test of classification performance

라의 거리가 0.6m 밖에 되지 않아 화면내의 변위가 커져서 대응에 어려움이 있음에도 불구하고 Fig.17에서 처럼 3차원 거리정보가 구해졌다. 이와같이 3차원 데이터를 얻는데는 약 30초 정도의 처리시간이 소요되었으며 이중 대부분의 시간은 전처리 과정인 불변량 발생 시스템에서 사용되었다. 이와같은 전처리 과정을 전용 하드웨어로 처리한다면 전체 신호처리 시간은 대폭 단축될 수 있을 것으로 여겨진다.

4. 결 론

본 논문에서는 로봇 작업에 적합한 스테레오 비전 시스템의 개발을 위해 즉 일반적으로 변위가 크고 카메라

== stereo vision feature position ==


number	x-pixel	y-pixel	code
1	59	127	192
2	84	97	255
3	93	148	255
4	116	116	255
5	140	119	255
6	105	98	192
7	115	159	192
8	152	190	255
9	157	143	255
10	220	169	255
11	204	171	255
12	174	143	192

== stereo vision feature position ==

number	x-pixel	y-pixel	code
1	59	72	192
2	84	42	255
3	93	93	255
4	116	61	255
5	140	67	255
6	115	107	192
7	152	135	255
8	157	88	255
9	220	115	255
10	204	116	255
11	174	91	192
12	172	136	192

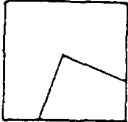
Fig.16 Processing result of images by stereo vision system

*** neural network recall result

straight edge	32	
corner	64	
fork	128	

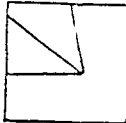
	s. edge	corner	fork	code
1	95.271	76.870	84.382	32
2	94.424	70.865	76.602	32
3	94.134	69.078	79.786	32
4	94.855	74.814	80.450	32
5	95.372	67.302	76.828	32
6	63.275	44.284	66.753	128
7	95.451	69.606	78.663	32
8	95.824	79.406	80.355	32
9	95.394	71.759	83.798	32
10	92.237	64.423	77.956	32
11	95.271	76.870	84.382	32
12	94.723	73.541	75.599	32
13	95.695	63.217	80.806	32
14	93.531	75.176	82.383	32
15	95.647	70.465	78.867	32
16	95.271	76.870	84.382	32
17	95.240	71.885	79.929	32
18	91.864	73.906	80.551	32
19	94.281	68.430	79.751	32
20	94.899	70.837	78.945	32

*** neural network recall result

straight edge	32	
corner	64	
fork	128	

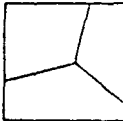
	s. edge	corner	fork	code
1	40.799	74.203	30.272	64
2	29.715	78.643	26.072	64
3	32.284	73.483	26.218	64
4	69.321	81.856	61.206	64
5	55.023	60.840	60.821	64
6	66.154	72.741	53.291	64
7	48.419	61.146	57.092	64
8	48.632	82.836	54.612	64
9	43.873	82.790	33.711	64
10	33.685	77.963	26.744	64
11	40.799	74.203	30.272	64
12	36.073	81.904	27.220	64
13	41.554	79.768	28.013	64
14	71.316	83.773	58.550	64
15	42.388	67.347	41.897	64
16	43.589	77.364	38.569	64
17	52.060	63.057	54.711	64
18	52.147	82.662	59.108	64
19	40.276	81.681	28.232	64
20	44.532	81.303	31.011	64

*** neural network recall result

straight edge	32	
corner	64	
fork	128	

	s. edge	corner	fork	code
1	48.690	41.178	84.388	128
2	68.192	49.807	79.176	128
3	47.237	39.870	79.992	128
4	44.717	46.421	64.067	128
5	49.349	49.158	76.978	128
6	43.943	42.926	71.804	128
7	57.209	51.400	74.247	128
8	41.737	47.727	66.061	128
9	49.333	38.652	82.945	128
10	62.710	54.442	81.753	128
11	35.406	31.358	73.889	128
12	47.037	37.481	81.846	128
13	41.563	46.661	76.066	128
14	40.782	42.870	61.200	128
15	57.413	54.116	73.885	128
16	45.668	39.814	74.475	128
17	41.883	39.616	72.481	128
18	41.409	38.876	65.692	128
19	69.030	58.984	86.028	128
20	44.009	33.324	77.457	128

*** neural network recall result

straight edge	32	
corner	64	
fork	128	

	s. edge	corner	fork	code
1	55.229	44.488	86.995	128
2	54.692	42.549	85.067	128
3	44.940	41.256	63.236	128
4	45.105	48.600	75.562	128
5	48.790	31.879	83.181	128
6	47.486	29.785	84.829	128
7	43.200	32.967	85.000	128
8	44.458	49.980	70.194	128
9	53.748	38.325	80.285	128
10	52.713	49.727	87.298	128
11	51.326	37.190	85.383	128
12	56.685	52.112	89.513	128
13	46.464	51.043	64.788	128
14	29.091	34.489	65.782	128
15	61.061	48.451	85.747	128
16	63.619	59.289	86.516	128
17	50.768	48.428	76.263	128
18	49.691	31.376	86.261	128
19	69.534	66.091	83.529	128
20	44.965	38.636	62.549	128

Fig. 15 Classification results

```

== stereo vision depth map ==

* camera focal length : 0.016000
* camera baseline width : 0.042000

number of feature point : 11
number      x-coord      y-coord      z_coord      disparity
1           -0.0435      -0.0488      0.6015      0.0022
2           -0.0893      -0.0310      0.6015      0.0022
3           -0.0115      -0.0247      0.6015      0.0022
4           -0.0603      -0.0085      0.6015      0.0022
5           -0.0565      0.0089      0.6362      0.0021
6           0.0081      -0.0097      0.6362      0.0021
7           0.0527      0.0169      0.6015      0.0022
8           -0.0191      0.0204      0.6015      0.0022
9           0.0218      0.0661      0.6126      0.0022
10          0.0237      0.0536      0.6015      0.0022
11          -0.0178      0.0343      0.6362      0.0021

stereo vision - depth map making completed
    
```

Fig.17 Coded feature values and depth map by stereo vision system

리로부터 얻은 영상의 변화가 심한 경우에 적합하도록 신경회로망을 도입 설계하였다. 여기서 적용한 신경회로망은 가중치를 미리 적고 형태로 주고 역치를 학습하는 형태로 구성하여 학습 시간을 단축하고 성능을 향상 시켰다. 또한 신경회로망에 적합한 입력을 얻기 위해 이동, 회전, 크기에 독립적인 시스템을 설계하였다. 결과적으로 앞의 과정을 통해 얻은 시스템으로 실제 작업환경에서 스테레오 비전을 적용할 때 접할 수 있는 문제들 - 큰 변위로 인한 대응점 결정의 어려움 등 - 을 단순한 피쳐의 대응이 아니라 특정 코드화 된 피쳐의 대응을 해결하였다.

참고문헌

1. Arking, A., Lo, C., Rosenfeld, A., "A Fourier approach to cloud motion estimation", American Meteorological Soc., Journal of Applied meteorology, 1978, 17, pp.735-744
2. Haass, U. L., and Brubaker, T. A., "Estimation of cloud motion from satellite pictures", Proc. IEEE Int. Conf. Acoustics, speech and signal proc., ICASSP, Denver, Colorado, 1980, pp.422-425
3. Rinehart, R. E., and Garvey, E. T., "Three dimensional storm motion detection by conventional weather radar", Nature, 1978, 273, pp.287-289
4. Y. C. Kim and J. K. Aggarwal, "positioning Three-dimensional Objects using stereo images", IEEE Trans. Journal of Robotics and Automation, Vol. R. A-3, No. 0. 4, pp. 361-373, Aug. 1987
5. W. E. L. Grimson, "Computational Experiments with a Feature based stereo algorithm", IEEE Trans. on Pattern Analysis and Machine intelligence, Vol. PAMI-7, pp. 17-34, Jan. 1985
6. J. R. Jordan III and A. C. Bovik, "Computational stereo vision using Color", IEEE control system magazine, pp.31-36, Jun. 1988
7. Stephen T. Barnard and Martin A. Fischler, "Computational stereo", Computing surveys, Vol. 14, No. 4, December 1982
8. S. B. Pollard, et al., "Geometrical Modeling from Multiple Stereo Views", The international journal of robotics research, vol. 8, no. 4, August 1989
9. J. F. Vega-Riveros, K. Jabbour, "Review of

- motion analysis techniques", IEEE Proceedings, Vol.136, Pt.I, No.6, December 1989
10. Yoh-Han Pao, "Adaptive Pattern Recognition and Neural Networks", Addison-Wesley Publishing Company, Inc., 1989
 11. D.G.Elliman and R.N.Banks, "Shift invariant neural net for machine vision", IEEE Proceedings, Vol.137, Pt.I, No.3, June 1990
 12. Philip D. Wasserman, "Neural computing-theory and practice", Van Nostr and Reinhold, 1989
 13. Anil K. Jain, "Fundamentals of digital image processing", Prentice-Hall, 1989
 14. Jack D. Gaskill, "Linear systems, Fourier transforms, and Optics" John Wiley & Sons, 1978
 15. Edward R. Dougherty and Charles R. Giardina, "Image processing-Continuous to discrete volume I", Prentice-Hall, 1987
 16. Jae. S. Lim, "Two-dimensional Signal and Image processing", Prentice-Hall, 1990
 17. David E. Rumelhart, James L. McClelland and the PDP Research Group, "Parallel Distributed Processing Vol.1, Vol.2", the MIT press, 1986
 18. Dan E. Dudgeon and Russell M. Mersereau, "Multidimensional Digital Signal Processing", Prentice-Hall, 1984
 19. Friedrich M. Wahl, "Digital Image Signal Processing", Artech House, 1987