

## 스테레오 영상 융합을 위한 신경망회로

正會員 李熙滿\*

### A Neural Network for Stereo Image Fusion

Hee Man Lee\* Regular Members

#### 要 約

스테레오 비전의 목적은 2개의 신경망회로부터 3차원 공간정보를 얻는데 있으며 스테레오 비전에서 가장 어려운 과정은 매칭 포인트를 찾는 데 있다. 본 논문에서는 매칭율을 향상시키기 위해서 웨이블릿변환과 신경망회로를 함께 사용하였다. MRA 웨이블릿 변환은 저해상도에서 고해상도 매칭을 위한 유용한 도구이며 또한 웨이블릿 기저함수의 위치이동에 의해 신호파형의 이동 및 변형을 용이하게 할 수가 있다. 본 논문에서 제안한 스테레오 신경망회로는 웨이블릿 변환계수를 입력신호로 사용하여, 두 파형의 융합(fusion)을 위해 지역매칭 위치까지 파형을 이동시키되 전체 영상파형의 RMSE값이 최소가 되도록 고안되었다. 이를 위해 기저함수의 위치를 정규화시켜 신경망회로의 가중치(weight)로 표현하여 가중치변화에 따라 파형의 일부 또는 전체를 이동시킬 수 있도록 하였다. 본 논문에서 제안한 스테레오 신경망회로는 인간의 융합 메카니즘을 설명할 수 있는 한 모델이다.

#### ABSTRACT

The purpose of stereo machine vision is to determine the three-dimensional location of objects from two two-dimensional images. The most difficult task in stereo vision applications is the determination of matching points in the two image planes, called correspondence problem. This paper concentrates on the utilization of the wavelet transform and neural network architecture jointly for improving this matching capability. The wavelet transform is used to facilitate a coarse-to-fine matching approach by using its Multi Resolution Analysis(MRA) properties. The wavelet transform also gives an easy way to shift or modifies signal waveforms by changing there basis locations. The wavelet transform coefficients are used in the stereo neural network, which is a hierarchical structure accommodating the coarse-to-fine matching strategy. The artificial neural network achieving good performance

\*삼성중공업 중앙연구소

Samsung Heavy Industries Co., LTD. Deaduk R  
& D Center

論文番號 : 94262-0929

接受日字 : 1994年 9月 29日

through dense interconnection of non linear simple computational elements changes weights, a normalized basis location, for stereo image fusion. The proposed stereo neural network is a good model for simulating human fusion mechanism which is the process of minimizing the global Root Mean Square Error (RMSE) of the two signals with shifting operations.

## [. 서 론

본 논문의 연구목적은 스테레오비전 시스템에 있어 2개의 영상을 융합하는 신경망회로를 구현하는데 있다. 스테레오비전에서는 3차원 공간을 2차원 공간으로 투영 변환을 행하는 카메라를 2대 이용하여 카메라의 기하학적인 배치로부터 잃어버린 3차원공간을 재구현함에 있다. 이때 가장 문제점으로 대두되고 있는 것은 2개의 영상, 즉 왼쪽 영상과 오른쪽 영상 사이에서 서로 대응되는 점을 찾는 일이다. 2대의 카메라는 서로 위치가 다르므로 왼쪽 영상에서의 한점에 대응하는 오른쪽 영상의 점은 그 위치가 다르며(그위치의 다른 정도에 따라 3차원 공간을 재구현을 할수 있음), 또한 어느 점은 서로 대응되는 점이 없는 경우도 있다.

기존의 스테레오 영상의 매칭방법은 크게 3가지로 분류할 수 있다. 그 첫번째는 영상계조매칭방법으로 영상계조가 같은 점을 찾는 것이다. 즉 대응되는 점이란 에피폴라(epipolar) 선상에 영상계조가 연속적으로 일치되는 이웃수가 가장 많은 점을 찾는 것이다. 그러나 에피폴라 선상의 영상계조 파형은 단순히 위치만 이동되지는 않는다. 카메라의 각도가 다르기 때문에 에피폴라 선상의 파형의 모양은 수축/확장되며 부분적으로 다른 모양을 하고 있다. 또한 대응점 영상계조는 정확히 일치하지 않으며, 반사등에 의해 때로는 매우 다른 값을 가질 수도 있으므로 일반적으로 영상계조 매칭방법은 많은 문제점이 있다. 두번째 방법은 두 영상간 상관도를 이용하는 방법이다. 영상 위의 한 점(m,n)의 상관도는 다음과 같이 계산된다.

$$c(m, n) = \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} W(i, j) I(m+i, n+j)}{\sqrt{\sum_{i=0}^{M-1} w^{2(i, j)} \sqrt{\sum_{j=0}^{N-1} I^{2(m+i, n+j)}}}} \quad (1)$$

영상 위의 한 점(m,n)에 대해 상관도를 계산한 후 다른 영상 평면상의 에피폴라 선상의 각점마다 상관도를 구한 후 가장 큰 상관도값을 갖는 점을 매칭 포인트로

가정된다. Barnard<sup>(1)</sup>는 서로 이웃하는 영상 위의 점들이 서로 유사한 변위정도(disparity)를 갖는다는 점을 이용하여 릴렉세이션 라벨링방법으로 매칭하는 위치를 찾고 있다. 그러나 이 방법은 윈도우크기를 얼마로 하여야 하는가가 문제가 되며, 상관도가 가장 큰점이 반드시 대응하는 점이라고 말할수 있는가? 또한 복수개의 매칭점들을 어떻게 처리하여야 하는가 등의 문제점이다.

세번째 방법은 특징정보추출(feature-based)에 의한 매칭방법으로 정보의 양을 줄이고 프로세싱을 하므로 계산속도가 빠르고 비교적 신뢰성 있는 변위정도를 구할 수 있으므로 가장 널리 사용되고 있다. Marr & Poggio<sup>(2)</sup>는 인간 시각계의 스트레오 프로세싱에 대한 이론을 발표하였으며 Grimson<sup>(3)</sup>은 그 이론을 컴퓨터에 실현하였다. 그 이론은 2개의 눈에 비친 영상에서 매칭 포인트를 구하고 3차원의 정보를 얻기 위해 다음의 5단계 과정을 제시하고 있다. (1)각기 다른 크기 및 방향의 마스크를 통해 특징을 추출하여 정보의 양을 줄인다. (2)각각 다른 크기의 마스크에서 얻어진 영상에서 제로 교차점을 찾는다. (3)매칭포인트를 찾는 과정으로 매칭되는 점은 같은 부호와, 같은 방향성을 갖어야 한다. (4)눈동자 움직임은 저해상도의 시신경 경로에 의해 이루어진다. 즉, 인간의 시각계는 복수개의 해상도 채널을 갖고 있으며 매칭과정은 저해상도에서 시작하여 고해상도로 옮겨가 진다. (5)매칭 프로세스가 끝났을 때 그 결과를 다이나믹 버퍼에 보관한다. Marr는 특징추출 방법으로 라플라시안 오프 가우시안(Laplacian of Gaussian) 연산자를 추천하였다. 그 연산자는 아래와 같다.

$$\nabla G = \left[ \frac{x^2 + y^2}{\sigma^2} - 2 \right] \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2)$$

복수개의 매칭점이나 비매칭점은 복수개의 해상도를 사용함으로써 어느정도 해소 될 수 있다. 즉 저해상도의 특징추출 영상에서 얻어진 결과는 고해상도의 매칭포인트를 구할 때 참조된다. 그러나 이 방법은 이산 변위정도 정보만을 얻을 수 있어 특징정보가 없는 부분은 인터

플레이션 연산을 할수 밖에 없으므로, 사람의 융합 메카니즘 즉, 왼쪽, 오른쪽 눈에서 얻어진 영상을 이중상없는 하나의 선명한 영상으로 융합하며, 아울러 거리(3차원정보)를 인지할 수 있는 메카니즘에 대해 설명하지 못한다. 이에대해 본 연구는 기존방법과는 달리 웨이브렛을 이용하여 신호를 분해하고 이를 신경망회로에 입력하여 지역매칭 정보에 따라 원신호를 신축, 신장 및 이동에 의해 스테레오신호가 융합되도록하며, 아울러 원위치에서 최종 이동한 거리가 변위정보가 되므로 이로부터 3차원 정보를 얻을 수 있도록 하는 방법에 관한 것이다.

## II. 웨이브렛 변환

본 논문에서는 웨이브렛 변환을 스테레오 융합의 도구로서 사용하고 있으므로 이에 대해 간략하게 설명하고자 한다.

푸리에 변환은 시간축에 대해 구간간 적분하여야 함으로 실제 공학적 응용에는 많은 어려움이 있었다. 웨이브렛 변환은 원신호를 주파수영역과 시간영역에서 모두 지역특성이 좋은 기저함수의 이동(translation)과 신축(dilation)에 의해 합성하므로 신호처리나 컴퓨터비전 등 그 응용은 점차 확대되고 있다.

웨이브렛 변환은 Morlet<sup>(44)</sup>에 의해 처음 제안되었으며 Meyer는 컴팩트 서포트 직교 기저함수를 제안하였다. 그러나 Meyer의 기저함수는 기저함수 표현에는 많은 상수들을 필요로 하므로 좋은 지역 특성을 갖지 못했다. Battle & Lemare<sup>(45)</sup>은 각기 독립적으로 좋은 지역특성을 갖는 직교 기저함수를 제안하였다. Meyer & Mallat은 웨이브렛변환에 MRA(multi resolution analysis)개념을 도입하였다. Mallet<sup>(46)</sup>은 Battle-Lemare 기저함수를 사용하여 신호를 복수해상도에서 분해 및 재조합하는 알고리즘을 개발하였다. Chui & Wang<sup>(47)</sup>은 직교기저함수가 아닌 준직교기저함수를 사용하는 웨이브렛변환을 연구하였다. 본 논문에서는 chui & Wang의 B-spline 기저함수를 사용하였다.

MRA개념을 사용하는 웨이브렛 변환에 대해 고찰해보자. 원신호가 공간  $V_n$ 에 존재하고, 기저함  $\phi(t)$ 는 다음의 특성을 만족한다고 가정하자.

1.  $V_n = Clos_{L^2} \langle \phi_{n,k}; k=0, \pm 1, \pm 2, \dots \rangle$   
 $\dots V_{n-1} \subset V_n \subset V_{n+1} \dots$
2.  $Clos_{L^2} \langle \cup_{n=-\infty}^{\infty} V_n \rangle = L^2(-\infty, +\infty)$

$$3. \cap_{n=-\infty}^{\infty} V_n = 0$$

$$4. \phi(x-k), k=0, \pm 1, \pm 2, \dots$$

Riesz Basis

$$5. f(x) \in V_n \Leftrightarrow f(2x) \in V_{n+1}$$

그러면  $\forall f(x) \in L^2(-\infty, \infty)$ 는 기저함수  $\phi(t)$ 의 이동과 신축에 의해 근사된다.

$$\text{즉 } f(x) \rightarrow f_n(x) = \sum_{l=-\infty}^{\infty} \alpha_l \phi_{n,k}(2^n x - l).$$

$V_n$ 의 직교상보공간을  $W_n = Clos_{L^2} \langle \psi_{n,k}; k=0, \pm 1, \pm 2, \dots \rangle$ 라고하면 두 공간 기저함수들의 내적은  $\langle \phi_{n,k}(t), \psi_{n,k}(t) \rangle = 0$ 되며, 공간  $W_n$ 를 웨이브렛 공간이라 한다. 이때 공간  $V_n$ 는 2개의 부공간으로 아래와 같이 분해된다.

$$\begin{aligned} V_n &= V_{n-1} \otimes W_{n-1} \\ &= V_{n-2} \otimes W_{n-2} \otimes W_n \\ &= V_{n-3} \otimes W_{n-3} \otimes W_{n-2} \otimes W_n \\ &= \cup_{k=-\infty}^{n-1} W_k \end{aligned} \tag{3}$$

즉 함수  $f(x)$ 는 웨이브렛 공간으로 투영된다.

$$\begin{aligned} f(x) \rightarrow f_n(x) &= \sum_l C_l \phi(2^n x - l) \\ &= \sum_k C^{n-1}_k \phi(2^{n-1} x - k) \\ &= \sum_k D^{n-1}_k \psi(2^{n-1} x - k) \\ &= \sum_{n=-\infty}^{\infty} g_n(x), g_n(x) \in W_n \\ g_n &= \sum_{k=-\infty}^{\infty} D^n_k \psi(2^n x - k) \end{aligned} \tag{4}$$

패턴매칭은 고해상도 보다는 저해상도가 유리할 때가 많다. 그 이유는 저해상도의 프로세싱처리 단가는 고해상도보다 낮으며, 또한 구체적 미세한 매칭은 전체 모양의 매칭에 있어 중요하지 않기 때문이다. 그러므로 저해상도에서 매칭을 하고 이를 바탕으로 고해상도에서 매칭하는 방법(coarse-to-fine matching)은 자연스러우며 유용하다. MRA 웨이브렛 변환의 공간  $V_n$ 는 그레이 레벨 피라미드 데이터 구조와 유사하며, 공간  $W_n$ 는 Grimson<sup>(48)</sup>이 Mar & Poggio의 이론을 컴퓨터에 실현한 특징정보 영역의 멀티채널과 흡사하다. MRA 웨이브렛변환은 이들 2가지 정보를 모두 제공할 수 있다. 또한 웨이브렛 변환은 원신호를 지역 특성이 좋은 기저함수의 위치변화에 따라 원신호파형을 변형 시킬 수 있다. 즉 모양의 어느 한 부분만을 확대 및 축소가 가능하다.

며 또한 자리가동도 가능하다. 스테레오 영상의 에피폴라 선상 파형은 서로 유사하지만 그 위치가 서로 다르고 부분적으로 어느 한 부분은 확대 및 축소되어 있어 이들 2개의 파형을 서로 이동 및 일부 변형하면서 매칭시키는 방법에 웨이브렛은 매우 유용한 도구이며 이를 이용한 그 응용 분야도 매우 넓다고 사료된다. 본 논문에서는 융합 메카니즘을 실험결과로 보여주기 위해 웨이브렛 변환 중 공간  $V_n$ 만을 이용하였으나 공간  $W_n$ 를 사용하는 경우도 매우 좋은 결과를 기대할 수 있을 것이다.

### Ⅲ. 신경망회로

생물학적인 신경망회로를 모방하여 만든 인공 신경망회로는 비선형 분산병렬 프로세싱 소자 사이의 밀도깊은 연결을 통해 좋은 성능을 얻기위한 Data flow의 한 특수한 정보처리 구조이다. 인공 신경망회로(이하 신경망회로라 칭함)는 폰 노이만방식의 한계에 대한 대체를 찾기 위한 시도에서 연구가 시작되어 인식분야 및 비선형 매핑분야에 있어 비교적 성공한 연구분야이다.

신경망회로의 프로세싱 소자는 생물학적인 뉴런의 단순모델로서 복수개의 입력이 내부의 복수개의 가중치(weight)와 상호 작용하여 한개의 출력을 얻고 있다. 즉 출력  $y$ 와 입력  $x$  및 가중치  $w$ 와의 관계식은 다음과 같다.

$$y = f\left(\sum_{i=0}^{N-1} w_i x_i - \theta\right) \quad (5)$$

$\theta$  : threshold

신경망회로의 가장 성공적인 응용분야는 패턴분류일 것이다. 생물학적인 모델과 거리가 먼 BPN은 가장 성공적인 네트워크로 비선형 입력과 출력의 매핑, 즉 비선형 좌표변환에 널리 응용된다.

스테레오 매칭에 있어 신경망회로 적용장점은 대응점을 찾는 데 있어 계산상의 복잡성을 피할수 있다는 점이다<sup>9)</sup>. Sun<sup>10)</sup>은 Hebbian 학습론을 이용하여 스테레오 랜덤도트 영상의 대응점을 찾는 알고리즘을 제안하였다. 신경망회로의 학습과정이란 외부입력에 따라 원하는 출력이나 반응을 하도록 내부의 가중치 ( $w_n$ )를 변화시키는 과정을 말한다<sup>11) 12) 13)</sup>. Hebbian 학습은 아래와 같다.

$$w_i(t+1) = w_i(t) + \alpha r_i(t)$$

$$r_{ij}(t) = y(t)x_j(t) \quad (6)$$

$\alpha$  : 학습율,  $x$  : 입력,  $y$  : 출력

Nasrabadi<sup>9)</sup>는 2차 배열의 바이너리 Hopfield를 이용하여 스테레오 매칭방법을 제안하였다. 그는 Moravec 윤곽추출 연산자<sup>14)</sup>를 사용하여 영상에서 특징을 추출하여 얻고 Hopfield 네트워크의 에너지 함수를 최소화 시키면서 대응점을 찾고 있다. Hopfield의 에너지 함수는 다음과 같다.

$$H(x) = - \sum_{i=1}^m \sum_{j=1}^m w_{ij} x_i x_j + 2 \sum_{i=1}^m T_i x_i \quad (7)$$

Grossberg & Marshall<sup>15)</sup>은 인간의 시각피질(visual cortex)을 모델링하는 복수 해상도 피이드백 신경망회로를 제안하였다. 그들은 Kaufman stereogram<sup>16)</sup>이나 kulikowski stereogram<sup>17)</sup>에서 알 수 있듯이 인간 시각계의 융합과 경합현상이 동시에 존재함을 신경망회로의 모델로 설명하려고 시도하였다.

### Ⅳ. 스테레오 신경망회로

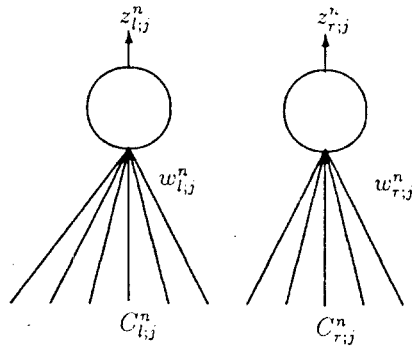
인간은 왼쪽과 오른쪽으로 받아들인 영상을 매우 빠르게 이중상이 생기지 않도록 선명한 하나의 영상으로 만들어 3차원 공간을 인지할 수 있다. 결과적으로 융합 과정에서 3차원 정보인 변위정도정보가 작성되는 셈이다. Marr & Poggio<sup>23)</sup>가 제안한 휴먼 스테레오 프로세싱 이론은 특징정보를 추출하여 매칭하므로 결과적으로 특징정보가 없는 부분은 거리 정보를 얻을 수 없는 단점(이산 변위정도 정보이기 때문)이 있으며 사람의 융합 메카니즘의 설명에는 어려움이 있다. 한편, Julesz<sup>18)</sup>의 스테레오 랜덤도트영상에서 알 수 있듯이 인간은 물체의 인식없이도 융합이 가능하다. 즉 물체에 대한 인식이 없어도 로우레벨에서 영상의 융합이 이루어 짐을 의미한다. 본 논문에서 제안한 스테레오 신경망회로는 로우레벨에서 영상의 인식이나 특징정보를 추출하지않고 융합시키므로 이산변위정보가 아닌 연속변위정보를 얻을수 있어 휴먼스테레오 프로세싱에 대한 설명이 가능하다.

본 논문의 스테레오 융합 메카니즘은 복수개의 해상도(multi channel)를 갖도록 오른쪽 및 왼쪽 영상을 웨이브렛변환하며 이를 신경망회로에 입력하고 저해상도 영상에서 부터 융합 프로세싱을 시작하여 고해상도에 이르는 처리방식으로 두 영상 위의 파형은 서로가 서로를

답아가도록 웨이브렛의 기저함수의 위치를 신경망회로의 가중치에 의해 지역 매칭위치정보에 의해 전체 신호의 오차를 최소화하도록 융합한다. 융합이란 의미는 두 신호가 한개의 신호가 되는 과정이 아니며 각각의 신호는 각 해상도에서 서로가 독립적으로 존재하여 Marshall<sup>[15]</sup>의 “어느 해상도에서는 융합이 일어나고 어느 해상도에서는

경합현상이 일어난다”는 제안에 대해 경합현상은 두뇌가 융합 과정을 끝낸후 오른쪽과 왼쪽 영상신호를 선택적으로 교환하면서 인식하는 과정에서 발생되는 현상이라고 해석한다.

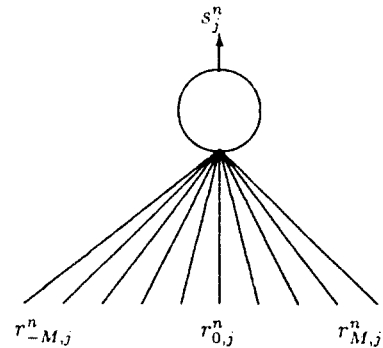
즉 융합과정은 두개의 영상을 중첩했을 때 이중상이 생기지 않도록 만드는 과정이며 이에 대한 전체매칭기준



$$z_{l;j}^n = \sum_k C_{l;k}^{n-1} \phi(2^{n-1} w_{l;j}^n - k)$$

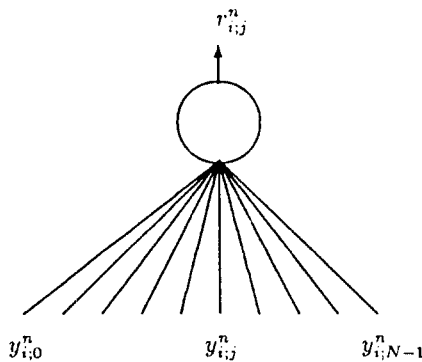
$$z_{r;j}^n = \sum_k C_{r;k}^{n-1} \phi(2^{n-1} w_{r;j}^n - k)$$

(a) Neuron of Fusion Layer



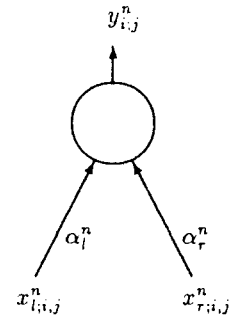
$$s_j^n = \begin{cases} i & \text{one winner with } i > 0 \\ & \text{one winner with } i < 0 \\ 0 & \text{otherwise} \end{cases}$$

(b) Neuron of Tristate Layer



$$r_{i;j}^n = f_t \left( \sum_{k=0}^{N-1} \beta_{i,k} y_{i;k}^n - V_s \right)$$

(c) Neuron of Rivalry Layer



$$y_{i;j}^n = \begin{cases} 1/N & \text{if } G_{max} \\ & -Abs(\alpha_r^n x_{r;i;j}^n - \alpha_l^n x_{l;i;j}^n) > Vth \\ 0 & \text{otherwise} \end{cases}$$

(d) Neuron of Affinity Layer

그림 1. 스테레오 신경망회로

(global criteria)은 두 영상 파형의 RMSE값을 이용하며 이 값이 최소가 되도록 각각의 파형을 좌우로 이동시킨다. 3차원의 정보는 각 융합과정이 끝난 후, 파형의 초기위치에서 최후 위치까지 이동한 변화량에 따라 계산된다.

본 연구에서의 스테레오 신경망회로는 공간  $V_n$ 정보, 즉 영상계조 값을 매칭 정보만으로 사용하고 있으나 공간  $W_n$ 정보를 사용한다면 특징정보(feature)를 이용할 수 있다. 스테레오 신경망회로는 계층구조로 MRA 웨이브렛 변환의 각 해상도마다 4개의 서브 레이어가 있다. 저 해상도의 4개의 층에서 융합이 끝나면 이 결과를 이용하여 한차원 높은 해상도에서 융합 과정이 시작된다. 지역매칭을 위한 검색공간은 계산의 효율을 위해 제한되어 있으며  $n-1$  해상도에서의 검색공간  $s_n$ 는 실질적으로  $n$  해상도의  $2s_n$ 에 해당된다. 각 해상도에 있는 4개 층은 각각 fusion layer, affinity layer, rivalry layer, tristate layer로 명명하였다(그림 1. 참조).

(그림 2.)는 한 해상도의 스테레오 신경망회로의 층별 연결도를 보인 것이다. Fusion layer는 왼쪽 카메라 및 오른쪽 카메라에서 얻어진 영상이 서로가 닮아가도록 전체 영상파형의 RMSE값을 최소화하면서 융합되

는 층이다. 이층은 가중치 값이  $0 \leq w \leq 1$  사이에서 변함에 따라, 실질적으로 기저함수의 위치를 변화시키어 이로인해 에피플라 선상의 파형을 변화시키는 과정이다. 바꾸어 말하면 융합과정은 2개의 영상이 한개로 합쳐지는 과정이 아니라 각각이 독립적으로 동화되는 과정이라 할 수 있다. Fusion layer의 가중치는 tristate layer의 출력값에 따라 변화되는데 tristate layer의 출력값은 에피플라 선상의 한점에서 로컬매칭 스코어가 가장높은 위치를 가르킨다.

융합과정시 왼쪽 영상과 오른쪽 영상은 카메라의 기하학적 특성상 서로 반대방향으로 움직이게 된다. 해상도  $n$ , 위치  $j$ 의 fusion layer 출력값  $z_j^n$ 은 다음과 같이 계산된다.

$$\begin{aligned} z_{l,j}^n &= \sum_k \phi(2^{n-1}w_{l,j}^n - k) C_{lk}^{n-1} \\ z_{r,j}^n &= \sum_k \phi(2^{n-1}w_{r,j}^n - k) C_{rk}^{n-1} \end{aligned} \quad (8)$$

$C_k^{n-1}$ 는 wavelet 변환계수

여기서  $l$ 은 왼쪽,  $r$ 은 오른쪽,  $w_j^n$ 은 가중치를 의미한다.

tristate layer의 출력값은 지역매칭 위치를 찾는 것

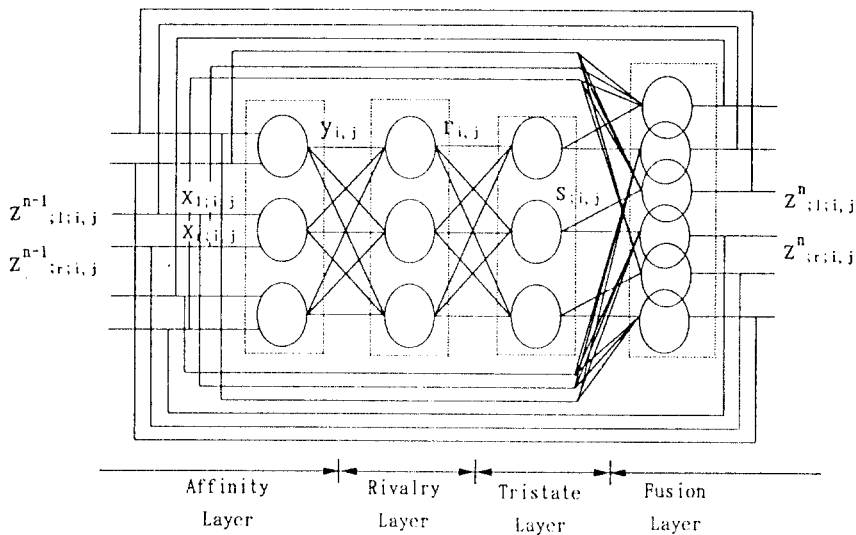


그림 2. 스테레오 신경망회로 전체구성도

으로 승자의 위치값을 출력한다. 즉 출력  $s_j^n$ 은 아래와 같다.

$$s_j^n = \begin{cases} i \text{ winner 위치 \#} \\ 0 \text{ otherwise} \end{cases} \quad (9)$$

복수개의 승자가 존재하는 경우는 승자가 없는 것으로 간주된다.

fusion layer의 가중치는 다음과 같은 갱신법칙을 따른다.

1.  $s_j^n > 0$  인 경우

$$\begin{aligned} w_{r,j+k}^n &= w_{r,j+k}^n + \frac{1}{2N_n} (|s_j^n| - |k|) \\ \text{where } -s_j^n &\leq k \leq s_j^n \\ w_{r,j+k}^n &= \max(w_{r,j+k}^n, w_{r,j+k}^n) \\ \text{if } j+k &= j+k \\ w_{l,j+k}^n &= w_{l,j+k}^n - \frac{1}{2N_n} (|s_j^n| - |k|) \\ \text{where } -s_j^n &\leq k \leq s_j^n \\ w_{l,j+k}^n &= \min(w_{l,j+k}^n, w_{l,j+k}^n) \\ \text{if } j+k &= j+k \end{aligned} \quad (10)$$

2.  $s_j^n < 0$  인 경우

$$\begin{aligned} w_{r,j+k}^n &= w_{r,j+k}^n - \frac{1}{2N_n} (|s_j^n| - |k|) \\ \text{where } -s_j^n &\leq k \leq s_j^n \\ w_{r,j+k}^n &= \min(w_{r,j+k}^n, w_{r,j+k}^n) \\ \text{if } j+k &= j+k \\ w_{l,j+k}^n &= w_{l,j+k}^n + \frac{1}{2N_n} (|s_j^n| - |k|) \\ \text{where } -s_j^n &\leq k \leq s_j^n \\ w_{l,j+k}^n &= \max(w_{l,j+k}^n, w_{l,j+k}^n) \\ \text{if } j+k &= j+k \end{aligned} \quad (11)$$

위의 가중치 갱신법칙은 형상연속(figural continuity) 제한조건을 만족시키기 위함이며 영상표면연속(surface continuity) 제한조건에 따라  $j < j'$ 일때  $w_j^n < w_{j'}^n$  조건을 만족시켜야 한다. 변경된 가중치 값으로 RMSE 값을 계산 했을 때 이 값이 종전보다 큰 값인 경우 매칭포인트를 잘못 선정할 경우이므로 갱신한 가중치 값을 취소한다. tristate layer는 rivalry layer 출력값에서 가장 큰 값을 선정한다. rivalry layer는 지역 매칭점수를 계산한다. rivalry layer 출

력값  $r_{i,j}^n$ 은 다음과 같이 계산된다.

$$\begin{aligned} r_{i,j}^n &= f_i \left( \sum_{k=0}^{N-1} \beta_{i,k} v_{i,k}^n - v_{s, \min} \right) \\ \beta_{i,k} &= \begin{cases} 1 & \text{all } y_{i,i}^n > 0 \quad j \leq l \leq k \\ 1 & \text{all } y_{i,i}^n > 0 \quad k \leq l \leq j \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (12)$$

$v_{s, \min}$ 은 임계값(threshold)로 한 물체의 정규화된 최소값이다.  $y_{i,k}^n$ 값은 왼쪽과 오른쪽 영상을 비교하여 서로가 매칭되었다고 인정될 때 영이 아닌 숫자를 출력하는 affinity layer의 출력값이다.

$$y_{i,j}^n = \begin{cases} \frac{1}{N_n} & \text{if } G_m - |a_r^n x_{r,i,j}^n - a_l^n x_{l,i,j}^n| > V_{th} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

$N_n$ 는 해상도  $n$ 에서의 정규화 요소이며,  $a^n$ 는 감쇄계수.  $x_n$ 은 fusion layer에서 피이드백 된 값이다. Fusion layer의 출력값  $z^n$ 과  $x^n$ 의 피이드백 연결관계식은 서로 이웃 픽셀과의 순차적 비교를 위해 다음과 같이 연결한다.

$$\begin{aligned} x_{r,i,j}^n &= z_{r,j-1/2}^n \\ x_{l,i,j}^n &= z_{l,j-1/2}^n \end{aligned} \quad (14)$$

## V. 실험결과

컴퓨터 시뮬레이션은 Unix Sys V3.0 MIPS 3230/3240 컴퓨터를 이용하였다. 스테레오 알고리즘의 성능 평가에 대한 정규화된 방법이 없으므로 알고리즘의 성능 분석에는 사실 어려움이 많다. 본 논문에서는 인공 스테레오 영상에 가우시안 잡음을 부가하여 정확히 매칭한 점과 부정확히 매칭한 점을 계산하는 수치적 측정방식을 택하였다. 자연 영상을 성능평가에 이용하지 않은 이유는 인공 스테레오 영상은 매칭포인트를 정확히 알고 있지만 자연 영상은 물체표면이 천천히 변화기 때문에 정확한 매칭포인트를 수적으로 계산하기에는 어려움이 많다. 에피폴라 선이 영상 평면상에 수평축이 되도록 평행 광축 카메라 모델을 사용하였다.

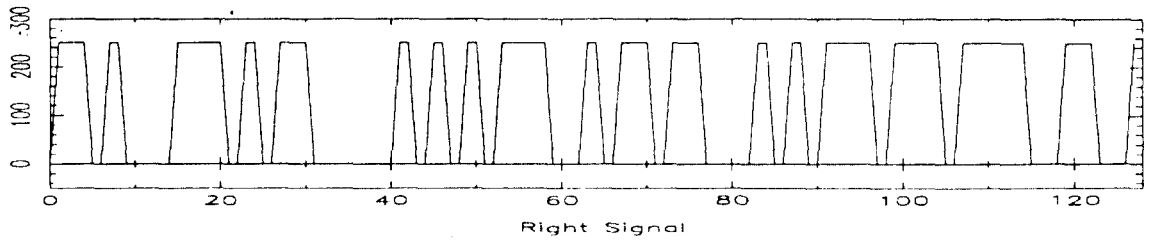
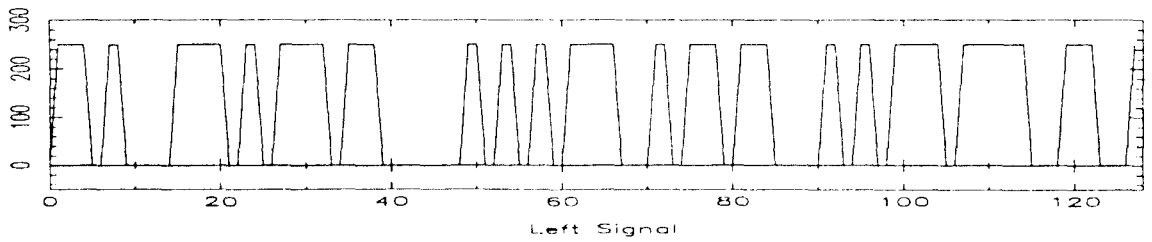
랜덤도트 스테레오 영상(그림 3.)은 128×128 픽셀로 50%의 검은색과 흰색의 장방형 모양을 하고 있어 아무런 특징정보를 갖고 있지 않다. 흑색은 계조값 1로, 흰색은 계조값 250로 하였으며, 2개의 영상은 중간 위만 서로의 픽셀 위치가 다르다. 웨이브렛의 최상위 해상



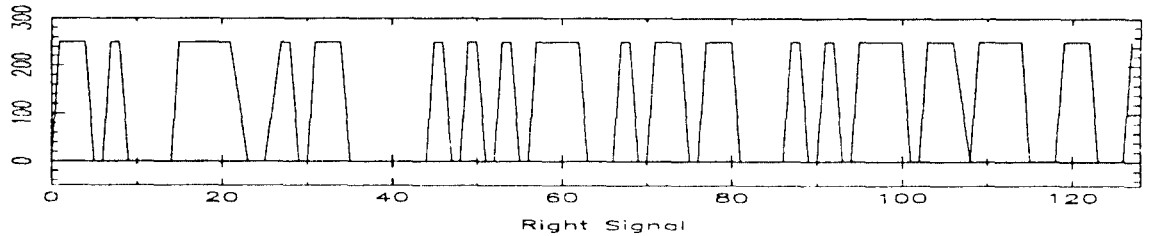
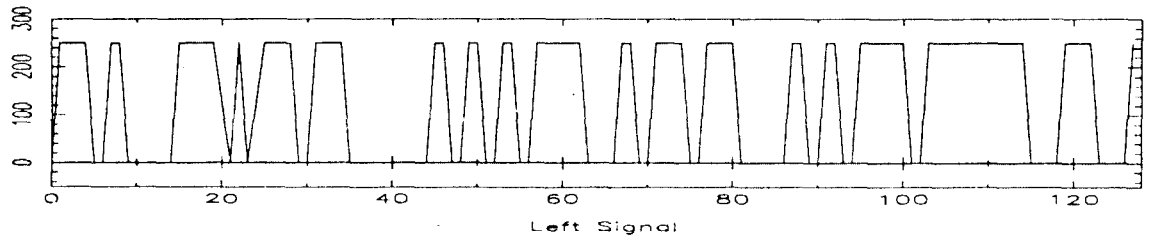
그림 3. 랜덤도트 스테레오 영상

도는  $n = 6$ 로 하였다. 그 이유는 기저함수( $\phi(128x-l)$ ,  $0 \leq x \leq 1$ )는 원신호를  $0 \leq l < 128$ 로 표현할 수 있기 때문이다.

랜덤도트 영상의 융합과정은 스테레오 신경망회로의 기본동작원리를 보여 주기 위해 해상도  $n = 6$ 에서만 하였다. (그림 4.)는 fusion layer의 에피플라 선상의 파형을 보인 것이며 (a)는 초기상태를 (b)는 융합 프로세싱후의 파형을 보였으며 아울러 3차원 정보인 변위정



(a) Initial State (RMSE=11.0)



(b) Iteration 7 (RMSE=3.51)

그림 4. 신경망회로의 반응결과



도정보를 보였다. 초기상태 영상에서 알 수 있듯이 왼쪽 영상과 오른쪽 영상은 단순히 자리 이동된 모습이 아니며 부분적으로 파형은 매우 다른 모습을 보이고 있다. 자연 영상인 경우에도 2대의 카메라의 위치가 서로 다르기 때문에 이러한 현상이 많이 발생한다. 그러나 융합 프로세싱에서는 이러한 비매칭점이 존재하여도 반응 결과에서 알 수 있듯이 이를 융합가능하며 그 결과로 거리정보인 변위정도를 얻을 수 있다. 그러므로 융합과정이라 함은 2개의 영상이 한개로 융합되는 현상이라기보다는 각각의 영상이 서로 닮아가는 과정이라 말할 수 있다.

스테레오 신경망회로의 성능을 분석하기 위하여 전통적인 방법인 상관도에 의한 매칭 및 특징정보매칭방법과 비교하기로 하였다. 전통적인 방법은 물론 융합 기능이 없지만 여기서는 변위정도 산출의 정확성을 성능비교의 기준으로 설정하였으며 다음과 같은 수식을 사용하였다.

$$E_{error} = \sqrt{\frac{\sum_{i=0}^{N-1} (r_i - d_i)^2}{N_2}} \quad (15)$$

상기 수식에서  $N$ 은 에피폴라 line상의 픽셀갯수이고  $r_i$ 는 이미 알고 있는 변위정도값,  $d_i$ 는 각 방법에서 산

출한 변위정도값이다. 가우시안 잡음은 다음과 같이 첨가되었다.

$$I' = I(1+rg(\cdot)) \quad (16)$$

$I$ 는 원 신호의 그레이 레벨값이며  $I'$ 는 잡음이 첨가된 값,  $r$ 은 양의 상수,  $g(\cdot)$ 는 정규 가우시안 램덤함수이다. (그림 5.)는 각각의 방법에 대한 잡음 특성을 보였다. 그림에서 알 수 있듯이 스테레오 신경망회로 방식이 잡음에 보다 더 강한 면을 보이고 있다. 상관도에 의한 방법은 윈도우 크기  $28 \times 1$ , 임계값은 0.92를 사용하였다. 특징정보에 의한 매칭방법은 라플라시안 오프 가우시안(laplacian of gaussian) 연산자를 사용하였으며 윈도우 크기는  $2 \times 1$ 을 사용하였다. 성능의 비교에는 물론 어려운 점이 따른다. 왜냐하면 상관도에 의한 방법과 신경망회로에 의한 방법은 연속형 거리정보를 계산하지만 특징정보 추출에 의한 방법은 이산형 거리정보를 생성하기 때문에 매칭 표준도 서로 달라 질수 밖에 없다. 그러므로 숫적인 비교는 상대적일 뿐이며 다만 잡음에 대한 반응특성을 알수 있을 뿐이다.

다음은 자연 영상에 대한 실험을 설명한다. (그림 6.)는 가정용 캠코더 소니 TR81을 사용하여 카메라를

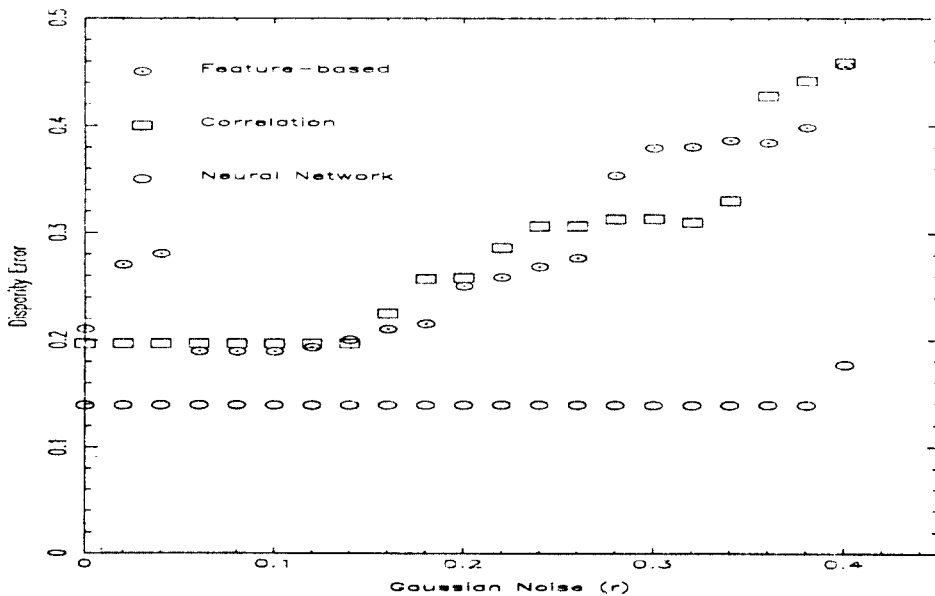
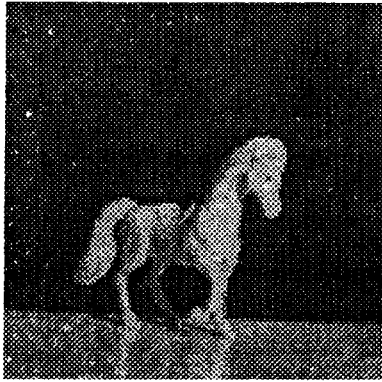
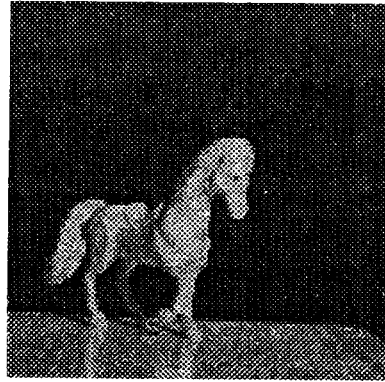


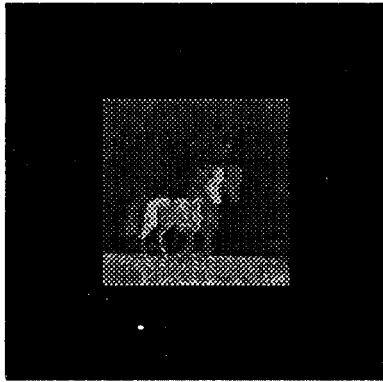
그림 5. 잡음 환경에서의 성능비교



(a) Left image



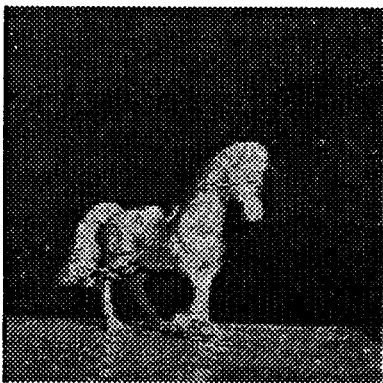
(b) Right image



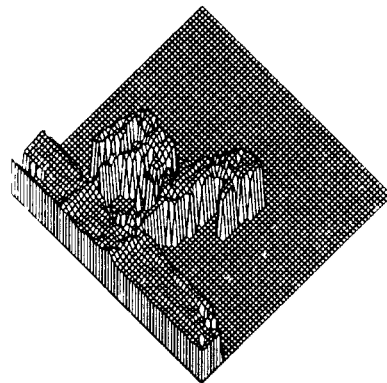
(c) Initial image(level 6)



(d) Initial image(level 7)



(e) Fused image



(f) Disparity map

그림 6. 실내 자연영상

슬라이드상에서 이동시키면서 각각 찍은 영상을 Data Translatron Inc의 DT286T-LC(60Hz) 디지털타이저를 이용하여 얻은 것이다. 이때 카메라 초점거리는 0.78cm로 고정했고, 디지털타이저는 half-gain, 2400 offset를 사용하였다. 640×640 디지털타이저 버퍼의 중심에서 256×256영상을 얻었다.

(그림 6.)의 초기영상(a, b)을 각해상도에서 중첩한 영상이 (c, d)이며 영상(e)는 저해상도 융합 후 고해상도 융합 프로세싱이 끝난후 두 영상을 중첩하여 얻은 영상이며, 변위정도정보 (f)는 신경망회로에서 산출한 3차원 정보이다. 자연 영상은 랜덤도트 영상과 같은 인공스테레오 영상과는 달리 카메라의 투영왜곡 때문에 물체의 크기가 부분적으로 다르며, 또한 반사광이 입사각에 따라 달라지므로 휘도발광변화(photometric luminance variation)가 발생한다. 본 스테레오 신경망회로는 웨이브렛 기저함수의 위치를 변경함으로써 지역 매칭위치까지 파형을 이동할 수 있으므로 투영왜곡에 대처 가능하며 아울러 전체매칭기준인 RMSE값이 최소가 되도록 수렴시킴으로서 (그림 5.)에서 알 수 있듯이 휘도발광변화에 민감하지 않다.

표 1. 실험 파라미터

	image size	Vs:min	user time	system time
그림 2	128	15	2.1 sec	.01 sec
그림 4 level6	128	15	4.3 sec	0.0 sec
그림 4 level7	256	15	10.6 sec	0.8 sec

(표 1.)은 2개의 실험 영상에 대한 프로세싱 파라미터이며 프로세서 사용시간(CPU time)은 Unix환경의 복수사용자 환경하에서 한개 에피폴라 선상의 처리 시간임으로 사용자 이용사용시간(user time)은 주어진 여건에 따라 다소 가변적이 될 수있다.

## VI. 결 론

본 논문은 스테레오 융합 및 3차원 정보를 추출하는 스테레오 신경망회로의 구조를 제안하였으며 인공 스테

레오 영상인 랜덤도트영상을 이용하여 전통적인 방법과 함께 성능을 분석하였다. 본 논문에서 제안한 신경망회로는 잡음환경과 투영왜곡 또한 휘도발광변화에도 강한 특성을 보였다.

카메라 모델은 에피폴라 선이 영상 평면상에 수평축이 되도록 평행광축 모델을 사용하였다. 피이드백 타입의 스테레오 신경망회로는 웨이브렛 변환한 계수를 사용하고 있으며, 본 논문에서는 연속 거리정보를 얻기 위해 공간  $V_n$  만 이용하였고 융합과 3차원 공간정보를 동시에 얻을 수 있는 인간의 시각계에 대한 한 모델로서 적절하다. MRA 웨이브렛 변환은 복수개의 해상도를 이용할 수 있도록 하며 저해상도에서 융합 단계를 거친후 고해상도로 프로세싱을 옮겨가는 방식은 저해상도에서 변위정도가 큰 영상도 작은 검색공간의 윈도우로도 처리할 수 있으므로 처리효율을 향상할 수 있는 잇점이 있다.

융합 메카니즘은 웨이브렛 변환된 기저함수 함수들의 정규화된 위치를 전체 영상신호의 RMSE 값이 작아지도록 이동시키는 과정이다. 스테레오 영상에서 공간  $W_n$  를 공간  $V_n$  와 함께 이용한다면 더욱 좋은 결과가 산출될 것이다. 왜냐하면 공간  $W_n$  는  $V_n$  와 보상공간 관계이기 때문에 상호보완적일 수 있다. 본 신경망회로는 음성 인식이나, 패턴매칭 등 인식분야에 적용 응용할 수 있을 것으로 사료된다.

## 참고문헌

1. S.T. Barnard, M.A. Fishler, "Computaional stereo" *ACM comput. surveys*, vol. 14, no. 4, pp.195-210, 1982.
2. D. Marr, Poggio, "A computaional theory of human stereo vision," *Proc. Roy. Soc. London*, Ser. B204, pp.301-328, 1979.
3. W.E.L. Grimson, "A computer implementation of a theory of human stereo vision," *Philos. Trans. Roy. Soc. London*, Ser. B292, pp.217-253, 1981.
4. J. Morlet, G. Arens, "Wave propagation and sampling theory," *Geophys.*, vol. 47, pp.203-236, 1982.
5. G. Battle, "A block spin construction of

- ondelettes, Part I : Lemarie functions." *Commun. Math. Phys.*, vol. 110, pp.601-615, 1987.
6. S.G. Mallat, "A theory for multi resolution signal decomposition : the Wavelet representation," *IEEE Trans. Pattern. Anal. Mach. Intelligence*, PAMI-11, no. 7, July, pp.674-693, 1989.
  7. Charles K. Chui, *An Introduction to Wavelets*, New York: Academic Press, 1992.
  8. W.E.L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Mach Intelligence*, PAMI-7, no. 1, pp.17-34, 1985.
  9. N.M. Nasrabadi, "Hopfiled network for stereo vision correspondence," *IEEE Trans. Neural Networks*, vol. 3, no. 1, January, pp.5-14, 1992.
  10. G.Z. Sun, H.H. Chen, Y.C. Lee, "Learning stereopsis with Neural Network," *IEEE First Int. Conference on Neural Networks*, San Diego, C.A., June, pp.345-355, 1987.
  11. Robert Hecht-Nielsen, *Neuro Computing*, New York : Addison-Wesley Publishing Company Inc., 1990.
  12. Marilyn McCord Nelson, W.T. Illirgworth, *A Practical Guide to Neural Nets*, New York : Addison-Wesley Publishing Company Inc., 1991.
  13. Bart Kosko, *Neural Networks and Fuzzy Systems*, Englewood Cliffs, N.J.: Prentice-Hall Inc., 1992.
  14. H.Moravec, "Robot Rover visual navigation," Ann Arbor, M.I. : U.M.I. Research Press, 1981.
  15. S.Grossberg, J. Marshall, "Stereo boundary fusion by cortical complex cells : A system of maps, filters, and feedback networks for multiplexing distributed data," *Neural Networks*, vol. 2, pp.29-51, 1989.
  16. L. Kaufman, *Sight and Mind : An Introduction to Visual Perception*, New York : Oxford University Press, 1972.
  17. J.J. Kulikowski, "Limit of single vision in stereopsis depend on counter sharpness," *Nature*, vol. 275, pp.126-127, 1978.
  18. Peter Burt, Bela Julesz, "A disparity gradient limit for binocular fusion," *Science*, vol. 208, 9 May, pp.615-617, 1980.



李熙滿(Hee Man Lee) 성희원

1961년 8월 5일생

1984년 : 고려대학교 전자공학과(학사)

1986년 : 한국과학기술원 전기 및 전자공학과(석사)

1994년 : 미국 Texas A & M Univ. Electrical Eng. Dept. (박사)

1986년 1월~1990년 7월 : 산업연구원 연구원

1994년 5월~현재 : 삼성중공업 중앙연구소 선임연구원