

# 3차원 로봇트 비전과 신경망 기법의 이용

정 동 석

(인하대학교 전자공학과)

## ■ 차 례 ■

- I. 3차원 정보 취득 기법
- II. 3차원 형태의 서술 방법
- III. 3차원 물체 인식기법

- IV. 신경망을 이용한 3차원 물체인식
- V. 보행 로봇트의 자동 주행을 위한 Computer Vision 기법

## 요 약

입력된 영상정보로부터 3차원 물체를 찾아내거나 인식하는 것은 컴퓨터에게 인간과 같은 능력을 부여하기 위한 가장 기본적인 과제라 할 수 있다. 본 글에서는 우선 3차원 정보를 취득하는 다양한 방법을 간단히 살펴보고 이어서 받아들이는 3차원 정보를 가지고 어떻게 3차원 물체의 형태를 서술하는가를 기존에 발표된 많은 문헌에 근거해설 요약해 보았다. 3차원 물체의 인식기법을 살펴보고 신경망을 이용한 3차원 물체인식 특히 Kohonen Map을 통한 3차원 인식을 중점적으로 다루었다. 마지막으로 보행 로봇트의 자동주행에 필요한 3차원 정보 취득 및 그 이용 방법을 제시하였다.

### I. 3차원 정보 취득 기법

본 장에서는 3차원 정보를 획득하는 다양한 기법에 대해서 살펴보고자 한다. 크게 분류하면 수동적 기법, 능동적 기법, 그리고 연속 영상에서 3차원 정보를 획득하는 방법등 세가지로 나눌 수 있다. 수동적 기법은 이미 획득한 2차원 영상 정보를 근거로 하여 3차원 자료를 계산해 내는데 비해 능동적 기법은 직접 3차원 정보를 획득하는 것을 말한다. 연속영상에서 3차원 정보를 추출하는 방법은 엄밀하게는 수동적 기법의 범주에 속하지만 다른 특징을 많이 포함하기 때문에 따로 구분해 보았다.

#### I-1. 수동적 기법

#### I-1-1. Stereo Vision

공간상의 한점을 서로 다른 두 위치에서 보게되면 그 점의 3차원적 정보를 추정할 수 있다. 이 원리를 이용한 stereo vision에는 한 장면에 대해 두개의 lens를 통해 얻은 두장의 영상을 해석하여 해당 장면의 3차원 정보를 구해 내게 된다. 거리 정보 도출 과정은 우선 동일한 점을 알아내어서(이를 대응점 이라고 한다) 동일한 점이 양 영상에서 가지는 변위(이를 disparity map이라 한다)를 계산하고, 변위량과 두 lens간의 기하학적인 거리를 이용하여 거리정보를 구하게 된다. 변위(disparity)란 3차원 공간상의 한점에 대한 오른쪽 영상과 왼쪽 영상에서의 상대적인 차이를 의미한다. 아래에 거리도출 구성을 도시하였다[1].

식에서 보듯이 거리 Z는 양 영상에서의 변위에 반

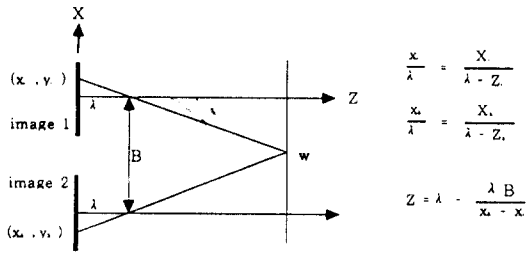


그림 1. Stereo Vision에서의 거리 측정 구성도.

비례한다. 일반적으로 변위가 클수록 두 영상 내에서의 대응점을 찾아내기가 어려워지므로, 해당 피사체까지의 거리를 충분히 할 필요가 있다. 또한 카메라 간의 기하학적인 배치를 정확히 해야하는 번거로움이 있다.

1-1-2. Shape from Shading

어떤 물체에 대한 밝기 데이터(계조도)로부터 해당 물체의 3차원 정보를 추출해 내는 방법이다. 3차원 물체의 영상은 물체의 형상, 표면의 반사 특성, 조명의 배치등에 따라 다르게 나타난다. 대상 물체 표면의 반사 특성은 아래 그림에서 알 수 있듯이 보는 방향에 따른 밝기  $\theta_e, \varphi_e$ 와 조명(또는 광원) 방향에 따른 밝기  $(\theta_i, \varphi_i)$ 와 관련되는 BRDF 함수(Bidirectional Reflectance Distribution Function)  $f(\theta_e, \varphi_e; \theta_i, \varphi_i)$ 에 의해 설명되어진다[2].

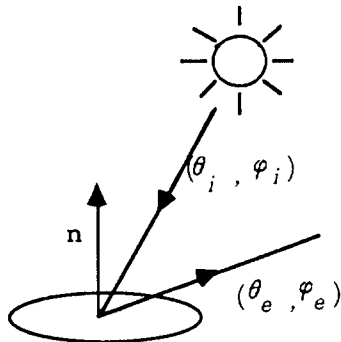


그림 2. Shape from shading 계산 구성도.

BRDF가 표면의 반사특성을 이해하기 위해 필요한 것이라면 반사지도(reflectance map)는 영상 형성을

이해하기 위해 필요한 개념이라 할 수 있다. Reflectance map은 1975년 Horn에 의해 제안된 것으로 대상 물체표면의 방향과 밝기와의 관계를 명확하게 보여준다[3]. Reflectance map  $R(p, q)$ 는 gradient space에서 정의된다. 즉 관측 방향을 Z-axis로 하면 3-D대상체의 형상은 xy평면에서의 높이 z로 나타내어지게 된다. 이때 z의 x와 y에 대한 일차 편미방이 p, q가 된다.

즉,  $p = dz/dx$   
 $q = dz/dy$

에 대한  $R(p, q) = E(x, y)$ 인 irradiation equation을 풀어서 shading 정보로부터 3-D정보를 유도해 내게 된다. 위 식에서  $E(x, y)$ 는  $(x, y)$ 점에서의 밝기를 나타낸다.

1-1-3. Shape from Texture

Texture에 기초해서 표면 정보를 찾아내려는 첫 시도는 1950년 Gibson[4]에 의해서이다. Gibson은 texture density의 gradient에서 피사체 표면의 방향 정보를 유도하였는데 평면성(planarity)과 균등 분포라는 가정을 하였다.

그는 아래와 같은 다섯가지 texture measure를 정의하고 각 measure에 대해 두 방향으로의 gradient를 구하였다. 그가 정의한 measure는 texture 소자의 1. width(w), 2. height(h), 3. eccentricity( $e = h/w$ ), 4. area( $a = hw$ )와 5. density( $\rho$ )등으로서 이들을 measure로 삼았다.

Stevens[5]는 1979년 그의 논문에서 locally planar와 locally uniform이라는 가정하에 normalized texture gradient를 제안하여 임의의 표면에 대한 국부적인 기울기를 계산하였다. Stevens가 제안한 normalized texture gradient는 아래와 같다.

$$\Delta w/w = -\tan^{\sigma}$$

$$\Delta h/h = -2 \tan^{\sigma}$$

$$\Delta e/e = -\tan^{\sigma}$$

$$\Delta a/a = -3 \tan^{\sigma}$$

$$\Delta \rho/\rho = 3 \tan^{\sigma}$$

여기서  $\sigma$ 는 기울어진 각도(slant)를 의미한다.

1-2. Active Technique(능동기법)

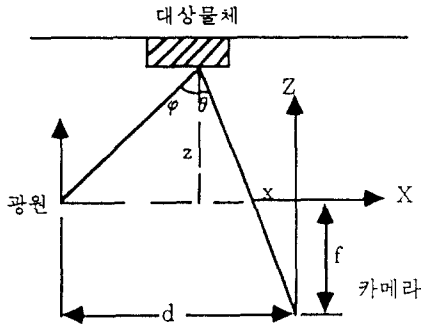
직접 range image를 획득하는 기법을 말하며 앞의 수동기법에서와 같이 주어진 영상으로부터 3차원 정보를 계산해내는 과정을 수반하지 않는다.

1-2-1. Structured Light 이용법

점, 선 또는 그리드(grid)등의 조명 형상을 피사체에 투사하고 이로인해 생긴 영상의 적합한 점들을 연관지어서, 다시 피사체 공간에서의 필요 제위 특히, 거리 정보를 구하는 방법이다.

조명원으로는 slit projector를 사용하며, 아래 그림과 같이 정해진 위치에 있는 projector의 수직 slit을 통해 빛을 피사체에 투사한다. Projector를 회전시킴에 따라 피사체의 영상이 미리 정해진 시간 간격을 두고 얻어진다.

따라서 영상내에서의 slit의 위치(아래 그림상의 x)가 결정되면 삼각법에 의해서 3-D에서의 위치(그림에서 z)도 구할 수 있다. 아래 그림에서



- d: 투사광원과 카메라 간격
- $\phi$ : 투사각도
- $\theta$ : 반사각도
- x: 영상점

$$\text{거리 } z = \frac{d - x}{\tan(\phi) + x / f}$$

그림 3. 광 투사를 이용한 거리정보 측정 개념도.

즉, object까지의 거리를 구할 수 있다. 이와같은 구조적 조명을 이용하여 거리정보를 구할 때의 문제점은 시간이 많이 걸린다는 점이다. 즉, 피사체상의 모든 면에 골고루 빛을 쏘아가면서 하나하나 영상을 구하는데 많은 시간이 소요된다. 보통 5~10분 정도의 acquisition time이 필요하다.

1-2-2. Moire Shadow 이용법

아래 그림의 배치와 같이 선형 광원을 optical grading

을 통해서 물체에 보내고 이를 camera에서 살피면, 물체에 비친 grading의 그림자들은 물체표면에서 contour curve를 이룬다. 이때 특정 contour를 이루는 각각의 element들은 camera로부터 같은 거리(equal range)상에 있다는 성질을 이용한다[6].

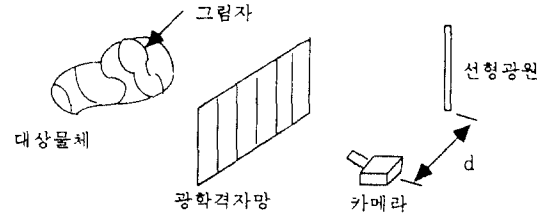


그림 4. 광격자 이용법의 구성도.

이때 카메라와 물체 사이의 거리(=range)를 계산하는 과정은 아래와 같다.

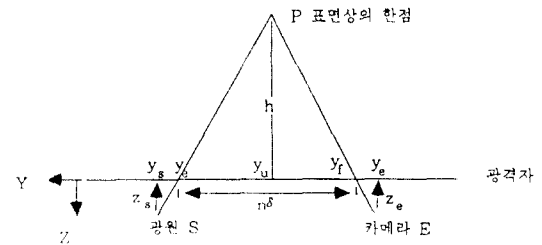


그림 5. 카메라와 물체사이의 거리 계산과정.

그림에서  $y_s, y_e$ : 광원 S와 카메라 E의 위치

P: surface 상의 한 point

$z_s, z_e$ : source S 및 camera E와 광격자간의 거리

삼각형의 닮은꼴 성질을 이용하면  $y_e$ 를 y축의 원점이라 할때

$$\frac{h}{z_e} = \frac{V_u - V_e}{V_e}$$

$$\frac{h}{z_s} = \frac{V_p - V_u}{y_s - V_p}$$

$$V_p = V_e + n\delta$$

여기서 n은 물체표면에 나타난 fringe의 순서를 말한

다. 이상을 정리하면

$$h_n = \frac{z_c z_s n \delta}{(z_s - z_c)V_r + z_c(y_s - n\delta)}$$

만일  $z_c = z_s$ 면, 즉 gridn부터 같은 거리에 있다면

$$h_n = \frac{z n \delta}{y_s - n \delta}$$

여기서  $y_s$ 는 카메라와 광원간의 거리가 되며,  $\delta$ 는 optical grating의 grid간격이다.

이상의 식을 실제 상황에 적용시킬 때의 문제점은 바로 각 fringe의 order  $n$ 을 찾아내는 일이다. 이를 해결하기 위해 기준선(reference line)을 이용한 기준 fringe를 인위적으로 구해서 이용하는 법이 제안되었다[6].

### 1-2-3. Time-of-Flight(TOF) 이용방법

Ultrasonic pulse나 beam을 이용해서 피사체까지의 왕복시간을 고려한 3차원 정보획득 방법이다. 초음파 pulse를 사용한 경우  $2d = v t$ 라는 관계가 성립해서  $d = v t/2$ 로 물체까지의 거리를 구하게 된다. 여기서  $t$ 는 pulse가 송출되어 되돌아 올 때까지의 시간이고  $v$ 는 소리의 전달속도이다. 문제는  $v$ 가 기압에 따라 변하므로 이의 보상을 고려해야 한다는 점과 ultrasonic beam의 초점을 정확히 맞추기가 어렵다는 점이다. 이와 같은 이유로 laser beam을 이용한 TOF range finder가 개발되었다[7].

Laser beam은 초음파 빔에 비해 초점을 잘 맞출 수 있으므로 보다 나은 해상도를 갖는 range image를 구할 수 있다. 그 구성도는 아래와 같다.

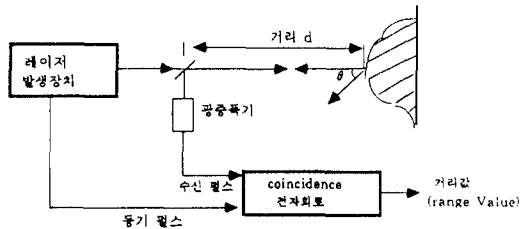


그림 6. 레이저 비행시간을 이용한 거리측정 개념도.

한쪽면이 투과되는 거울을 이용하여 pulse를 투사하면 object에 반사된 pulse가 이번에는 같은 거울에 의해 출력되어 photo multiplier로 들어가게 된다. 이

때 발생하는 arrival pulse와 이미 와있던 sync pulse간의 시간차( $\Delta t$ )를 측정하는 것이 coincidence electronics이다. 이와같이 송출하고 되돌아 오는데 소요된 시간은 거리에 비례하게 된다. 이때의  $t$ 값은 매우 적은 것이 보통이다. 3m range라고 할 경우  $\Delta t = 20ns$ 가 되는데  $\pm 0.5cm$ 의 정확도를 요할 경우  $\Delta t$ 는  $\pm 20ps$ 의 정확도로 측정되어야 한다. 이와같은 정확도를 얻기 위해서는 매우 빠른 electronic coincidence system을 필요로 하게 된다는 단점이 있다. 또한 물체 표면에서 반사되는 값이 매우 적을 경우에는 photo multiplier에 의해 noise가 돌보일 가능성도 있다.

### 1-3. Image Sequence로부터의 3차원 정보획득

3차원 공간에서 움직이는 물체의 표면상의 점들이 갖는 순간적인 외관속도의 장을 optical flow라고 한다. 이러한 optical flow가 인간이나 동물들의 시각적 지각계통에서 중요한 요소라는 것은 이미 임상학적으로 널리 알려진 사실이다. 이러한 optical flow를 이용하여 3차원 정보를 구하고자 하는 시도는 Prazdry [8]에 의해 처음 제안되었다. 그는 object point, P를 지나가는 투사선의 각속도(angular velocity)들의 분포를 optical flow로 정의하고 관계식을 전개하였다. 각속도  $a_s$ 는 수평운동성분인  $a_x$ 와 회전성분인  $a_r$ 의 합으로 구성된다. 여기서 회전성분인  $a_r$ 은 이동체가 갖는 성분으로서 object상의 위치에 따라 변하지 않는다. 수평운동 성분인  $a_x$ 는 투사중심위치 (O)와 object상의 점간의 거리값에 따라 영향을 받는다.

$$a_s = a_r + a_x = a_r + \frac{TP}{\sin \beta}$$

$$|a_x| = \frac{T \sin \beta}{S}$$

여기서 T는 3차원 수평이동 vector  
 P는 점 P의 위치 vector  
 S는 점 P와 원점간의 거리  
 $\beta$ 는 P와 T간의 각

이 식의 의미는 가령 어떤 object상의 모든 점이 같은 값 T만큼의 수평이동을 했어도 실제 motion으로 감지되는 부분은 원점과 point간의 거리와 수평이동의 방향에 따라 좌우된다는 것이다. 따라서 object 상에 두점을 정해서 각 점간의 각속도  $a_s$ 에서 수평이동성분  $a_x$  간의 차이를 가지고 두 점간의 상대적인 깊이를 계산할 수 있게 된다.

$$\text{즉, } \frac{(a_{x1}-a_{x2})P1}{(a_{x1}-a_{x2})P2} = \frac{S_1}{S_2}$$

여기서  $a_{x1}$ ,  $a_{x2}$ 는 점  $P_1$ 과 점  $P_2$ 에서의 optical flow를 의미하고  $P_1$ ,  $P_2$ 는  $P_1$ 과  $P_2$  방향으로의 unit vector, 그리고  $S_1$ ,  $S_2$ 는 원점에서 점  $P_1$ ,  $P_2$ 까지의 거리를 의미한다.

Optical flow와 깊이정보를 이용하여 2차원 물체의 운동을 추정하고자 하는 연구가 Ballard와 Kimball[9]에 의해 제안되었다. 3차원 운동을 3가지 유형의 계수로 표현할 수 있는데 이들은 Global 기준 좌표계에 대해 Local 좌표계가 같은 상대적인 면이를 나타내 주는 위치계수와 local 좌표계의 원점이 갖는 직선운동속도를 나타내 주는 계수와 회전운동속도를 나타내주는 계수 등이다. 이러한 3차원 속도는 2-D optical field와 깊이지도(depth map)를 이용하여 계산한다. 3차원 속도  $V = (u, v, V_z)$ 에서 z방향 성분인  $V_z$ 는 깊이지도  $Z(X, Y)$ 를 이용하여 아래와 같이 구한다.

$$V_z = \frac{1}{\Delta t} [Z(X+u \Delta t, Y+v \Delta t, t+\Delta t) - Z(X, Y, t)]$$

$u, v$ 는 2-D optical flow 속도이다.

### II. 3차원 형태의 서술 방법

3차원 비전, 즉 3차원 형태를 인식하는 과정은 2차원 형태 인식보다 훨씬 복잡한 많은 어려운 문제들을 포함하게 된다. 이와같은 3차원 비전의 전단계로서 연구해야 할 과제가 3차원 형태를 어떻게 서술하는가 하는 문제이다[10, 11]. 이제까지 많은 연구자에 의해서 많은 이론이 제기되어 왔으나 여기서는 이들 이론들을 3가지로 분류해서 살펴보고자 한다. 즉, 3차원 물체를 이들의 표면이나 각 면의 가장자리들을 이용해서 서술하는 방법, Nevatia 및 Binford에 의해 제안된 generalized cylinder 방법, 그리고 체적(volume)을 이용한 서술방법등을 일차로 살펴보고 그외에 최근에 제안된 서술방법을 몇가지 추가로 고찰하고자 한다.

#### II-1. 표면(Surface)을 이용한 서술

Requich[12]에 따르면 간단한 형태를 갖는 3차원 물체라면 그 표면(surface)이나 표면을 둘러싸는 가장자리 만으로도 3차원 형태를 어려움 없이 서술할 수 있어야 한다. 3차원 물체에서 우리 눈에 들어오는 것은 표면(surface)이기 때문에, 면의 표현방법은 3차원

vision에서 중요한 과정중의 하나가 된다. 초기 연구는 주로 다면체(planar polyhedra)를 대상으로 했으나 후에는 sculpture surface로 확장되었다. 표면(surface)을 구성하는 기초단위인 면(face)은 어떻게 정의할 것인가? 면의 물체를 구성하는 가장자리(boundary)의 부분집합이라고 할 수 있다. 따라서 모든 면들의 합집합은 물체의 가장자리나 물체를 구성하게 된다. 이상의 정의에 또 다른 사항을 덧붙이는 것은 일반적이지 못하고 각 응용시 필요에 의해 추가될 수 있다.

면(face, F)과 경계선(edge, E) 및 꼭지점(vertex, V) 간의 관계식은 Euler의 공식으로 주어진다. 즉,  $V - E + F = 2$ . 이를 여러 개의 물체로 확장시키면  $V - E + F = 2(B - H)$ . 여기서 B는 연결되지 않은 물체의 수이고, H는 구멍(hole)의 수이다.

한가지 예로서 winged edge 다면체 표기법(surface-edge vertex 표기법)을 이용하여 4면체를 표기하는 방법을 살펴보자. 다음 그림에 보듯이 vertex, edge, face의 3종류의 노드가 있으며 각 노드는 link에 의해 연결되어진다.

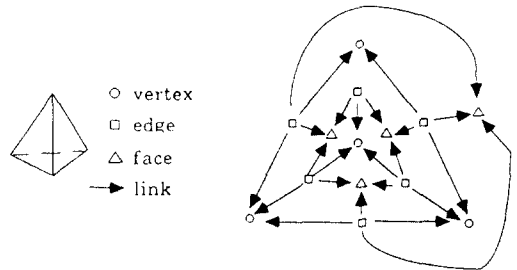


그림 7. 4면체에 대한 winged edge 다면체 표기법의 예.

각 vertex는 edge에 의해 연결되며 3개의 edge들이 한개의 face를 이루는 것을 알 수 있다. 각 node는 나름대로의 data를 저장하게 된다. 즉, 각 vertex는 3차원 공간상에서의 좌표를, face는 표면의 여러가지 특성들을, edge는 관련된 face나 vertex들과의 위상학적인 상호관계등의 정보를 attribute로 보관하고 있어서 인식 과정에서 이용할 수 있게 한다.

이외에 3차원 물체를 약 320개의 삼각형태의 면들의 합성으로 설명하려는 시도[13]와 구형(sphere) 표면조각들의 합으로 구성하여 설명코자 하는 시도가 있었다[14].

II-2. 원통(Cylinder)을 이용한 서술

Generalized Cylinder는 3차원 물체를 서술하고자 하는 원시단위(primitive)로서 축과 축상의 각점을 포함하는 단면들로써 구성되어진다. 여기서 단면은 축을 따라서 진행하면서 3차원 형태를 만들게 되는데 원통형의 경우 축이 직선이고 축상의 단면은 동일한 반지름을 갖는 원으로 이루어진다. 원뿔형은 축은 직선이고 단면은 원형이되 그 반지름은 0에서 어떤 값까지 일정하게 증가하는 형태가 된다. Torus는 축이 원이고 단면은 일정한 면적을 갖는 원이 된다. 이와 같은 원통의 단위를 통해 3차원 물체를 서술하고자 하는 방법은 1971년 Binford에 의해 처음 제안되었으며 이후 많은 발전을 계속해 왔다. Nevatia와 Binford[15], Marr와 Nishihara[16], Mohan과 Nevatia[17] 등으로 이어지며 3차원 인식 시스템으로 유명한 ACRONYM에 까지 응용되었다.

II-3. 체적(Volume)을 이용한 서술

체적을 기초로한 서술법은 표면을 기초로한 알고리즘과는 사뭇 다른 형태를 갖는다. 가장 쉬운 방법은 volume occupancy array로 표기하는 방법이다.

이것은 2차원에서의 spatial occupancy array 표기법을 단순 확장시켜서 구할 수 있다. Spatial occupancy array는 runlength방법으로 부호화할 수 있으며 만일 기본 block의 크기가 가변일 경우는 특히 cell-decomposition 방법이라 부른다. Cell decomposition에서는 cell의 형태가 다소 복잡한 것도 허용이 되지만 각 volume cell은 서로 disjoint 되어야 하며 cell 간의 허용 operation은 glue(붙임) 하나로서 충분하다. 각 cell은 그 내부에 hole을 갖는 것이 허용되지 않는 특징이 있다. 아래에 decomposition의 예를 들어 보았다.

Cell decomposition을 이용한 3차원 표기법의 대표적인 것으로 oct-tree 표기법[18]과 K-d tree[19]표기법이 있다. 이들은 전체 volume을 recursive 하게 나누어 가는 것으로서 2차원 이진영상을 표기하는 pyramid 데이터구조의 확장이라고 할 수 있다[20].

Oct-tree표기법은 하나의 부모 노드가 8개의 자식 노드를 가지는 형태로서 8-ary tree구조를 갖는다. Tree에서의 각 node는 8조각낸 6면체의 한 조각을 나타낸다. 각 node는 full(F), empty(E), partial(P)이라는 정보를 갖는데 F는 서술하고자 하는 object의 해당 subcube에 대응하는 체적이 온전히 있을 때이고 E의 경우는 비어 있는 E인 node는 그 밑에 자식 node를 갖지 못한다. 2차원의 quad-tree 표기법과 비교해 보면 black이 E, gray가 P, white가 F에 해당된다고 볼 수 있다. 아래에 8진 트리(oct-tree)를 이용한 표기법의 예를 들어보았다.

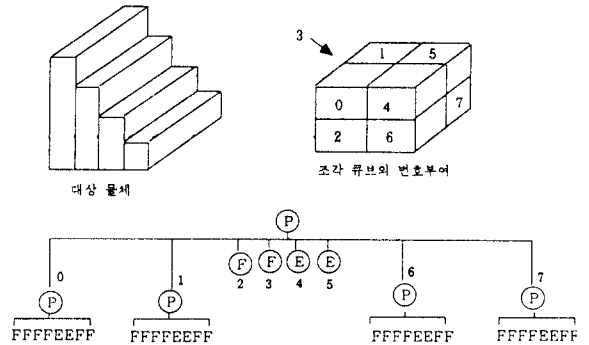


그림 9. 8진 트리를 이용한 3차원 물체 서술의 예

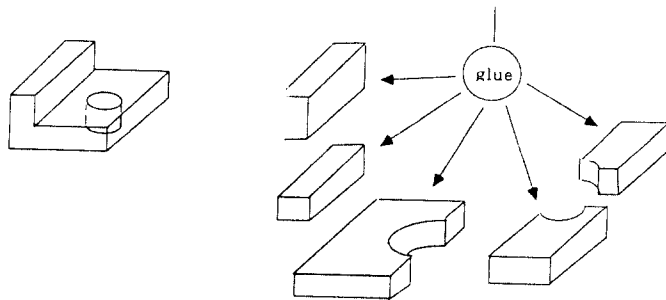


그림 8. cell decomposition의 예.

### II-4. 가우시안 구(Gaussian Sphere)를 이용한 표현

1979년 Horn[21]에 의해 처음 제안된 이래 많은 개선이 진행되어 오고 있는 3차원 서술방법이 가우시안 구 이용법이다. 3차원 물체 표면의 각 점에 대한 surface normal(표면 법선)로서 3차원 형태를 서술한다는 것이 기본개념이다. 만일 표면이 평면이라면 모든 점에서 법선방향은 동일할 것이며, 만일 볼록한 표면이라면 모든 점에서 다른 방향으로서 법선 방향을 갖게 된다. 이러한 표면 법선 벡터의 꼬리를 단위구(unit sphere) 또는 Gaussian구의 중심 원점에 놓이게 하고 법선 벡터를 외부로 연장하여 Gaussian 구의 표면과 만나는 점들의 집합을 해당 3차원 물체 표면에 대한 Gaussian 영상이라고 한다. 평면 표면을 갖는 물체의 경우는 Gaussian영상으로부터 본래 물체형상으로 되돌릴 수 없으므로 이 경우에는 법선의 방향과 함께 표면의 면적을 해당 법선의 weight로서 같이 고려해 주어야 한다. 아래에 Gaussian 영상의 개념을 보여주는 예를 들어 보았다.

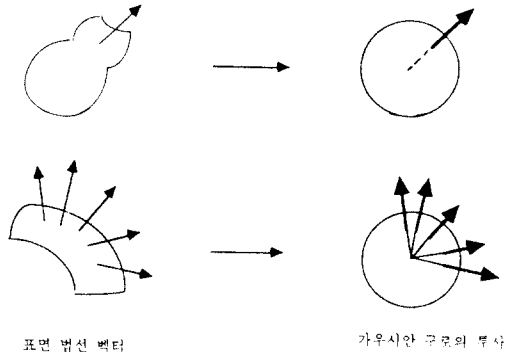


그림 10. Gaussian 영상의 예.

### III. 3차원 물체 인식기법

주어진 입력 영상으로부터 원하는 3차원 물체를 인식해내는 일은 machine vision에서 가장 중요한 과제 중의 하나라고 할 수 있다. 3차원 물체인식의 난이도는 몇가지 인자에 의해 결정된다. 주어진 영상에 몇가지의 물체가 있으며 그 물체의 복잡도는 어느정도인가? Model database에는 얼마나 많은 물체가 수록되어 있는가? 주어진 영상에 대해 알고있는 사전 정보

는 많이 있는가? 등등이 대표적인 인자가 된다. 물체를 인식하기 위해서는 우선 인식대상 물체로부터 중요 특징들을 찾아내고 그 특징들간의 관련성을 잘 이용해야 한다.

#### III-1. Model-based 기법

3차원 물체인식에서 보편적으로 쓰이는 방법이 model based 기법이다. 물체는 여기서 사용되는 model이 어떤 형태인가에 따라 다양한 알고리즘이 가능하다는 것이다. 수없이 많은 연구 사례가 발표되어 왔는데 그중에서 비교적 최근에야 관심을 갖게 된 것이 range data를 입력으로 하는 3차원물체 인식 분야이다. 입력 data가 range data이므로 이에 상응되는 model database를 구축해야 되고 또한 인식해야 된다. 주지하듯이 3차원 물체는 바라보는 방향에 따라 수없이 많은 형태를 가지므로 이를 모두 model base에 기록해 두기는 불가능하다. Fan et al.[22]은 symbolic description을 제안하여 각 물체에 대해 4~6가지의 symbolic description으로 model base를 구성한 바가 있다.

Computer vision에서 두 물체 또는 두 장면의 대응관계를 확인하는대는 특징점을 서술하는 단계별로 trade off가 존재한다. 특징 서술의 단계가 낮을수록, 즉 lower level의 경우일수록, 특징들의 계산은 쉽게 할 수 있는 반면 이러한 특징들은 모든 각도에 따른 변화가 크고 물체의 가림(occlusion)에 큰 영향을 받을 수 있는 등 여러가지 약점을 보유하고 있다. 반면 특징서술의 단계가 높을수록, 즉 high level일 경우에는, 특징을 추출해 내기가 어렵거나 일반적인 기법을 사용하긴 힘들긴 하지만 인식환경의 변화에 큰 영향을 받지 않는다는 등 특징으로서의 여러 강점을 가지고 있다. 따라서 3-D 물체인식에서는 이러한 특징 서술단계를 어느정도 level에서 잡을 것이냐 하는 것이 중요한 고려사항이 된다.

예를 들어 Fan et al.[22]은 object 서술에 물체표면(surface)정보를 이용하였는데 표면을 표면조각(surface patch)들로 분할하고 각 조각들의 상세한 서술과 각 조각간의 상호 관계로서 대상물체의 표면의 특징을 기록하였다. 이는 각 조각(patch)을 node로 조각간의 관계를 link로 한 속성도(attributed graph)로 나타낼 수 있음을 보인 것이다. 표면 서술에는 곡률(curvature) 정보가 이용되었다. 이와 같이 표면(surface)을 이용한 서술방법은 경계선(edge) 또는 점(point)을 통한 서술보다는 high level이라 할 수 있지만 체적(volume)을 이용한 서술보다는 낮은 level의 경우라고 할 수 있

다. 물체 인식(object recognition)은 2가지 단계를 필요한다. 첫단계는 Model 구축이고 두번째 단계는 입력영상에서 물체를 인식해내는 것이다. 주어진 입력 영상으로부터 자동으로 Model을 추출해 낼 수 있다면 더욱 바람직한 인식시스템이라 할 수 있을 것이다. 아래에 이제까지 발표된 3차원 물체인식 알고리즘 중 대표적인 것 몇가지를 간략히 서술한다.

### III-2. 3차원 물체 인식 시스템

#### 1. EGI(Extended Gaussian Image) 이용법 :

Horn et al.[21]은 1984년에 3-D물체인식에 multiview EGI 모델을 도입하였다. 각 object에 대한 model은 object를 Gaussian sphere에 투사하여 표기했으며 database 상의 model과 장면(scene)상의 물체를 서로 비교할때는 최소 EGI mass inertia를 갖도록 배열한 다음에 정합 정도를 비교하여 인식여부를 결정한다. EGI계산을 쉽게 할수 있다는 장점이 있는 반면에 물체가 가려지거나 여러물체가 한 장면내에 있는 경우에는 인식과정이 어려워진다는 단점이 있다.

#### 2. Augmented CAD Model 이용법(3DPO System) :

Bolles et al.[23]은 1984년에 range data를 받아들여서 3-D 형태를 갖는 부품을 찾아내는 3DPO system을 제안하였다. 시스템은 두 부분으로 이루어 졌는데, 그 하나는 경계선, 표면, 꼭지점과 각각의 연결성에 대한 정보를 갖는 CAD형 모델 부분이고 또 하나는 특징점들을 형태나 크기별로 분류해 주는 부분이다. 미지의 물체를 인식하기 위해서는 D.B.상에 있는 Model들의 특징들과 일치되는 특징들을 인식 대상 물체에서 찾아낸 다음에, 인식 대상물체가 D.B.상의 특정 물체를 가정을 하고 주어진 가설에 대한 검증 단계를 거친다. 본 방법은 상당히 복잡한 장면에서도 물체를 정확히 인식해내는 것으로 확인이 되었으나, 인식대상물체의 수가 적어야 하며 물체의 모양도 상당히 제한된다는 단점이 있다.

#### 3. Planar Face 이용법 :

Grimson et al.[24]은 1984년 3-D position과 surface normal을 이용한 3차원 물체인식에 대해 발표하였다. 본 방법에서는 planar face를 갖는 polyhedral 물체만을 대상으로 하였다. 3차원 물체에 대한 range data를 입력으로 하여 이로부터 각 표면 및 그 방정식을 추출하고 표면간의 관계를 계산한다. 이와같이 계산된 특징들을 가지고 정합과정을 시작하게 된다.

정합과정은 2단계로 나누어 볼 수 있다. 일단계에서는 계산된 특징값들과 database내 model들의 각 표면과의 일치정도를 비교하여 정합가능성이 있는 model들을 골라내고, 이단계에서는 변환시행 과정을 통해 진위여부를 확인한다. 변환시행이란 입력된 물체의 표면을 구성하는 특징점들이 model 표면상의 특징점들로 변화되어 지는가를 시험하는 것이다. 평면 표면을 여러개 가지고 있는 다면체 만을 인식대상으로 한다는 약점이 있다.

#### 4. Interpretation Tree 이용법 :

Ikeuchi[25]는 1987년 공장의 생산라인에서 부품의 방향과 위치를 찾아내는 연구를 하였다. 한가지 물체만을 인식대상으로 하여 여러 각도에서 관측된 형상에 대한 표면상의 특징점들을 기억시켰다가 interpretation tree를 통하여 인식시에 이용한다. 이 시스템은 엄밀하게 얘기해서 인식 시스템이라기 보다는 3차원물체를 주어진 장면상에서 확인해내는 시스템이라고 할 수 있다.

#### 5. Rigidity Constraints 이용법 :

Faugeras et al.[26]은 1986년 3차원 공간에서 rigid object를 인식하고 찾아내는(locate) 시스템을 제안하였다. Model 물체는 선형특징들 즉 점, 선, 면등을 이용하여 표기되고 인식대상 입력영상이 들어오면 영상내의 물체에서도 중요한 점, 선, 면등의 특징들을 추출한다. 정합과정에서는 rigidity constraint가 이용된다. 우선 model 물체와 인식 물체들간에 특징값들을 근거리 해서 가능한 짝들을 찾아내고 rigidity constraints를 이용하여 맞는 물체로 인식해낸다. 발표된 바로는 상당히 복잡한 형태의 물체도 잘 인식하는 것으로 알려졌는데 물체가 일부 가려졌을 경우에는 적용할 수 없다는 약점을 갖고 있다.

#### 6. Relational-Feature Graph 이용법 :

Oshima et al.[27]은 1979년 평면 또는 곡면을 갖는 물체를 인식하는 model-based인식 시스템을 발표하였다. 각 model은 평면 또는 곡면을 node로 하고 이웃하는 면 간의 연결 관계를 link로 하는 relational-feature graph로 나타내진다. 인식과정은 우선 입력 물체에서 가장 대표적인 면을 기준으로 잡아서 이 기준면과 유사한 면을 갖고 있는 model을 database 내에서 전부 찾아내어 인식 후보로 선택한다. 후보로 선택된 대상 물체에 대해서만 나머지 표면과의 관계를 추가로



확인하여 최종적으로 결론을 맺게 된다. 이 방법의 문제점은 동일한 물체라도 보는 방향에 따라 각각의 relational-feature graph를 만들어야 된다는 점과 인식 과정에서 이들 각 graph가 별도의 물체로 다루어진다는 비효율성 그리고 가려진 표면은 인식할 수 없다는 점 등이다.

7. Generalized Cone 이용법 :

Nevatia et al.[28]은 1977년 generalized cone을 이용한 3차원 인식 시스템을 제안하였다. Range data가 입력으로 사용됐으며 model의 표기나 인식대상 물체의 서술을 자동으로 획득할 수 있다는 특징이 있다. 인식 과정에서 indexing 기법을 사용하여 모든 database를 검색하지 않고 일부만을 검색할 수 있는 효율성을 가지고 있으나 복잡한 형태를 갖는 물체는 generalized cone으로 표현하기가 힘들다는 약점을 갖고 있다.

8. Generalized Cylinder이용법 :

Brooks [29]는 1981년 3차원 물체를 서술하는데 GC(Generalized Cylinder)를 사용하는 방법을 제안하였다. Database에 수록되는 model물체는 GC를 이용한 기본 체적들의 계층도 (hierachical graph)로 표기된다. 물체의 각 부분에 대한 GC를 node로 하고 부분간의 관계를 link로 하는 tree도 구성한다.

이 외에도 인식 대상 물체를 크게 분류해서 인식대상을 좁혀주는 restriciton graph를 구성해야 한다. 인식은 이 단계로 이루어지는데 일단계에서는 국부적인 특징들만을 이용하여 정합정도를 확인하며 이 단계에서는 전체적으로도 인식된 결과가 일관성을 갖고 있는지를 확인한다. GC이용법의 문제점은 model을 구성하는 수동작업이라는 것과 인식가능 물체의 형상이나 주위 영상이 제한적이라는 점이다.

9. Surface Description이용법 :

복잡한 형태의 면을 갖는 물체가 입력되면, 첫째로 주어진 면을 면편(surface patch)으로 분리해 내야 한다. 입력 data가 3-D이므로 각 면의 curvature를 계산하여 최대값과 zero-crossing 들을 이용하여 경계선을 검출해 내고 이로부터 면편을 구한다[22]. 즉, 각 surface는 면편으로 분할되어 표기되고, 이후 각 object는 attributed graph로 서술이 되어진다. 면편을 나타내는 node의 attribute로는 면적, 방향, 곡률의 평균, 중심축, 가림의 정도 등이 포함되고, link의 attribute로는 이웃되는 형태, 연결 확률 등이 이용된다.

주어진 물체와 database상의 모델과의 정합은 3단계 과정을 거친다. 첫단계인 screener에서는 database 상의 model view 중 가장 정합 가능성이 높은 상위 5개를 골라낸다. 다음단계인 graph matcher에서는 선정된 후보들과 인식 대상물체와의 정합정도를 기하학적인 변형(geometric transform)을 기친 다음 점수로 계산해 낸다. 3단계에서는 2단계에서 정합으로 판정된 영상내의 여러 표면들끼리의 상호관계를 파악하여 잘못된 부분을 정정하는 것으로 되어진다. 본 방법은 순수하게 data-driven 알고리즘으로 사전 정보가 필요없다는 장점이 있고 그 외에 일부 가려워진 물체의 인식에도 적용할 수 있다.

VI. 신경망을 이용한 3차원 물체인식

VI-1. 신경회로망을 이용한 3차원 인식 개념

Digital image로 부터 3차원 물체를 인식해 내는 것은 컴퓨터에게 인간과 같은 능력을 부여하기 위한 가장 기본적인 과제라 할 수 있다. 이차원 영상으로부터 3차원 물체를 찾아내야 (identify)하는데, 이는 인간의 시각을 이용하여 물체를 인식해내는 것과 유사한 과정을 취하게 된다. 따라서 사전에 대상 물체에 대한 정보, 예를 들어 여러 방향에서 본 이차원 형상 등이 존재하고 있어야만 인식이 가능하다. 회전, 이동, 또는 크기의 변화(scaling)에도 불구하고 해당 물체를 영상으로부터 찾아내기 위해서는 각 대상 물체(object)에 대한 view-independent model의 구축이 필요하다.

종래의 3차원 물체 자동분류(classification) 시스템은 각 물체에 대한 막대한 사전정보를 필요로 하고 있다. 인식률을 높이기 위해서는 해당 물체에 대한 보다 많은 이차원 정보를 저장해 놓고 이를 근거로 해서 해당물체의 존재 여부를 판단하므로 막대한 기억공간이 필요하고 database로부터 정보를 끄집어 낼때의 복잡성 등의 문제점을 안고 있었다.

신경회로망(neural network)기법을 사용하면 기존의 분류기(classifier)와는 달리 제한된 학습영상만을 이용한 학습을 통하여 이차원 영상 정보로부터 인식대상 물체의 정확한 소속(class)을 찾아낼 수 있는 분별함수(discriminating function)를 생성해 내므로 기존의 3차원 물체 인식 시스템이 갖는 복잡성의 문제가 많이 사라지게 된다. 현재까지 발표된 신경망을 이용한 3차원 인식대상은 주로 열쇠나 기계부품, 나뭇잎 등 이차원에 가까운 평면형의 물체[30]였으며 물체

와 물체가 겹쳐지지 않은 것을 대상으로 하고 있다. 신경망의 학습을 위한 입력 데이터는 해당물체의 특정 모습(view)에 대한 외곽선(boundary)을 기초로한 centroidal, cumulative angular, curvature 값[31] 등이 사용되었다.

Lin et. al. (1991. Jan.)은 여러개의 2-D view로 부터 3차원 물체를 인식하는데 Hopfield 신경망을 사용하였다. 그들은 2단계 인식과정을 제안하였는데 우선 입력 영상과 database내의 model 간의 surface matching score를 계산하여 인식 대상 model의 범위를 좁히고 2 단계로는 1단계에서 선택된 model들과 입력 영상의 vertex correspondance를 Hopfield 신경망을 이용하여 계산한다. 여기서 최고 점수를 받은 model을 인식된 model로 결정하게 된다.

Poggio et al.[32]은 대상 물체에 대한 임의의 view로부터 표준 view로 변환해주는 vector-valued object-specific function을 제안하였다. 얼마 후 Poggio et al. [33]은 generalized radial basis function(GRBF)를 제안하였다. GRBF의 계산은 입력력의 노드수가 동수인 two layer의 신경망을 이용한다. 학습은 컴퓨터 그래픽을 이용한 wire frame planar 3-D object model을 입력으로 하여 수행한다. 여기서 제안된 신경망은 학습을 시킨 한가지 대상 물체만을 인식하는 object-specific한 신경망이라는 특성을 가진다. 다만 여러개의 신경망을 결합하면 다수 물체의 인식도 가능해진다.

Bimbo et al.[34]은 Kohonen신경망을 이용한 3차원 물체 인식(분류)방법을 제안하였다. 인식 대상은 planar-faced 3차원 물체이며 한 신경망이 다수의 물체를 인식할 수 있는 특성을 갖는다. 각 Kohonen 신경망은 modular 형태로 결합되어 인식 대상 물체의 종류를 확대해 갈 수 있다[35]. 신경망의 입력으로는 shape descriptor의 계수가 사용되며 이것은 인식 대상 물체의 축척(scale) 변화에도 불변하는 특성을 갖는다. 따라서 system에 갖추어야 할 view의 숫자가 감소되는 효과를 갖는다.

#### VI-2. 신경회로망을 이용한 3차원 인식 연구

신경회로망을 사용해서 기하학적인 변형, 즉, 회전, 이동, 크기 등의 변화에 무관하게 물체를 인식코자 하는데는 3가지 접근방법이 있다. 첫번째가 신경회로망의 구조적 특성을 이용해서 불변성을 찾아내는 것으로서 한가지 입력영상에 대한 여러가지 변형된 입력영상을 나타내는 neuron들을 한가지 출력을 나타내는 동일한 neuron에 연결시키는 방법이다. 이 방법의 문

제점은 많은 복잡한 연결성을 갖게 된다는 점으로서 복잡한 형상의 물체의 경우에는 매우 심각한 문제로서 현실적으로 적용하기가 힘들게 된다. 두번째 접근 방법은 입력 data에 변형이 있어도 변하지 않는 특징(feature)을 추출해서 사용하는 방법이다. 이 방법의 불리한 점은 신경망에 적용시키기 전에 필요한 feature들을 미리 계산해 내야 한다는 점이다. 세번째 접근 방법은 학습을 통하여 물체의 변형에 무관하게 해당물체를 인식하는 것이다. 이 경우 학습에 이용되지 않은 perspective view가 신경망에 입력되어도 물체를 인식해 낼 수 있어야 한다. 이 방법의 문제점은 인식해내야 할 대상물체의 여러가지 다른 모습에 대하여 학습을 수행해야 한다는 점이다.

Bimbo et. al.[34]은 feature vector로서 3차원 물체에 대한 영상의 직선성분을 feature로 정하고 각 직선성분마다 3개의 feature vector를 부여하였다. 3가지 성분은  $\theta$ ,  $\rho$  및 직선의 길이  $l$ 이다. 여기서  $\theta$ ,  $\rho$ 는 Hough transform에 나타나는  $\theta$ ,  $\rho$ 를 의미한다. 그런데 이러한 feature들은 기하학적인 변형에 대해 역시 변화를 일으키므로, 동일 물체에 대해 여러가지 다른 각도에서 취득한 영상을 가지고 신경망을 학습시켜야 한다. 예를 들어 축척이 변하게 되면  $\theta$ 는 불변이지만  $\rho$ 나  $l$  등은 변하게 된다. 축척의 변화에 대한 민감도를 줄이려면 각 직선성분의 길이를 다른 직선성분과의 상대적인 비율로 나타내는 방법이 있다.

Noise로 인해 성분이 중도에 끊어져서 나타나는 것은 큰 문제가 되지 않지만 직선이 아닌 것이 직선으로 감지되는 경우에는 문제가 된다. 각각의 직선이 검출될 경우 이의 진위여부에 대한 판단과정을 threshold를 이용하여 기치는 것이 바람직하다.

$\theta$ ,  $\rho$ 의 정량화 단계가 너무 적을 경우에도 가짜 검출될 수 있으므로( $\theta$ ,  $\rho$ 와 비슷한 값을 가진 픽셀들이 직선 형태로 군집돼 있을 경우) 이들 parameter의 정량화 단계를 세분화 시킬 필요가 있다. 따라서 어떤 영상내의 직선성분이  $n$ 개 검출되었다면, 주어진 영상은 집합  $S$ 로 표기되어진다.

$$S = \{t_i, i = 1, \dots, n\}$$

단  $t_i$ 는 ordered triplet( $\theta$ ,  $\rho$ ,  $l$ )로서 직선성분  $S_i$ 를 나타낸다.

#### VI-3. Kohonen 신경망의 구성

비슷한 형태를 갖는 입력영상들을 출력 공간(output space)에서 같은 group으로 묶어 주는 것이 Kohonen

의 self-organizing map이나, 한개의 map내에서 여러개의 object를 나타낼 수가 있다. 즉 3-D data가 2-D space로 mapping 되면서 3-D공간에서의 유사함의 정도가 Kohonen map에서의 거리로서 표시되게 된다. 여러방향에서의 투사도는 Kohonen층에서 neuron의 활성화 정도로서 나타내진다. Kohonen망의 기본 idea는 입력 data간의 상하관계(order)가 출력 map에서는 유지될 수 있도록 하는 학습 알고리즘에 있다. 즉, 비슷한 형태의 입력끼리는 출력공간(2차원공간)에서도 이웃하는 위치에 오도록 투사(map)시킨다. 또 한가지 바람직한 특성으로는, 입력형태의 출현 빈도수(sampling density)가 출력공간에서는 차지하는 영역의 대소로서 대응되어 나타난다는 점이다.

Kohonen망은 입력공간 X에서 M개로 이루어진 neuron 열에의 투사(map)로 이루어진다[35]. 각각의 입력  $\mathbf{x} \in X$ 는  $x_1, x_2, \dots, x_d$ 의 activity vector로 표기되어지는데 여기서 d는 입력공간 X의 차원을 의미한다. 각각의 neuron은 그 위치  $r \in M$ 에 따라 명명되는데, 모든 입력과 연결되었으며 그 연결 강도를  $m_{ri}$ 로서 표시한다. (단,  $1 \leq r \leq M, 1 \leq i \leq d$ ).

Vector  $\mathbf{m}_r = \{m_{r1}, m_{r2}, \dots, m_{rd}\}$ 로서 입력들로부터 neuron r로서 연결강도를 나타낸다면, 입력  $\mathbf{x}$ 는 neuron r에게  $h(\mathbf{x} \cdot \mathbf{m}_r)$ 만큼의 여기(excitation)를 유발한 것으로 본다. 여기서  $h(\cdot)$ 는 sigmoidal형태의 함수로서  $[0, 1]$  사이의 값을 갖는다. 입력  $\mathbf{x}$ 와 neuron r의 여기값  $h(\mathbf{x} \cdot \mathbf{m}_r)$ 이 일치될 경우에는 연결강도  $\mathbf{m}_r$ 을 아래와 같이 조정하게 된다.

$$\delta \mathbf{m}_r = h(\mathbf{x} \cdot \mathbf{m}_r) (\mathbf{x} - \mathbf{m}_r)$$

실제로는 계산의 효율성을 높이기 위해, 위와 같은 연결강도 조정은 입력  $\mathbf{x}$ 에 의해 가장 많은 여기를 받게 되는 neuron C의 인접에 있는 neuron에 대해서만 행하는 것이 좋다. 즉,

$$\delta \mathbf{m}_r = \varepsilon h_{rc} (\mathbf{x} - \mathbf{m}_r) \quad 0 \leq h_{rc} \leq$$

여기서  $h_{rc}$ 는 대표 neuron C와의 상대적인 거리를 감안한 weight값으로서 c = r에서 최대값을 갖고  $\|r\|$ 가 증가함에 따라 0으로 가게된다. 또한 neighborhood set,  $N_c$ 를 정하여  $N_c$ 가 아닌 neuron에 대해서는 h를 0으로 하는 방법도 있다. Neuron C의 결정은

$$\|\mathbf{m}_c - \mathbf{x}\| = \min_{r \in M} d(\mathbf{x} \cdot \mathbf{m}_r)$$

여기서  $d(\cdot)$ 는 입력공간의 유사성을 감안하기 위한 distance measure이다. Training vector  $\mathbf{x}$ 를 이용하여 weight

vector  $\mathbf{m}_r$ 가 학습되는 과정은 아래 공식에 의한다. 만일  $\mathbf{x}$ 가 정확하게 분류되었다면

$$\mathbf{m}_r(t+1) = \mathbf{m}_r(t) + \varepsilon(t) [\mathbf{x}(t) - \mathbf{m}_r(t)]$$

만일  $\mathbf{x}$ 의 분류가 잘못되었을 경우

$$\mathbf{m}_r(t+1) = \mathbf{m}_r(t) - \varepsilon(t)$$

기타  $i \neq C$ 인 경우

$$\mathbf{m}_r(t+1) = \mathbf{m}_r(t)$$

$\varepsilon(t)$ 는 scalar gain ( $0 < \varepsilon(t) < 1$ )으로서 시간이 증가함에 따라 단속 감소하는 값을 갖는다. 보통  $\varepsilon(0)$ 는 0.01이나 0.02등의 작은 값을 갖고 시작하되, 계속 값을 줄여서 0까지 계속 하는데 100,000 step 정도의 학습을 시키면 좋은 결과를 얻을 수 있다.

#### VI-4. Kohonen Map을 이용한 3차원 물체의 인식

c개의 object에 대한 n개의 조감도(perspective view)로 이루어진 집합 X가 있다고 하자. 이 집합 X를  $X_1, X_2, \dots, X_c$ 의 disjoint subset으로 분할을 하되 각 subset 내의 sample은 서로 유사성을 많이 갖고 있도록 분류해야 한다. 각각의 object에 대응해서 한개의 neuron이 필요하되 서로 다른 object라도 입력 feature만으로서서는 구분이 불가능한 경우에는 한개의 neuron에 다수의 object를 대응시킬 수 있다. 유사성의 척도(similarity measure)로서는

$$\frac{x_1^t \cdot x_2^t}{\|x_1^t\| \|x_2^t\|}$$

$\mathbf{x}_1, \mathbf{x}_2$ 는 두개의 feature vector이다.

위와 같은 normalized inner product(정규화 내적)은 두 vector간의 cosine 값을 계산하는 것으로서 서로간의 correlation 척도로 볼 수 있다. 따라서, 최대로 여기되는 neuron은

$$\mathbf{x}^t \cdot \mathbf{m}_r = \max_r (\mathbf{x}^t \cdot \mathbf{m}_r) \quad \text{이 되도록}$$

학습방정식은

$$\mathbf{m}_r(t+1) = \begin{cases} \mathbf{m}(t) + \varepsilon(t) \mathbf{x}(t) \\ \text{-----}, & r \in N_c(t) \\ \|\mathbf{m}_r(t) + \varepsilon(t) \mathbf{x}(t)\| \\ \mathbf{m}_r(t), & r \notin N_c(t) \end{cases}$$

예를 들어 아래와 같이 정사면체에 대한 두가지 투사도가 있을 수 있다[34].

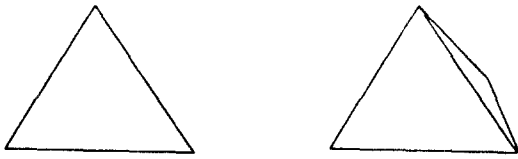


그림 11. 사면체 물체에 대한 연속적인 view의 예

왼쪽의 경우 object view가 9차원(선3개 X 각 선당 3개의 계수)에 속하는데 비해 오른쪽과 같은 view는 15차원을 갖게된다. 물론  $I$  계수의 경우 0에서 서서히 변하게 되지만  $\rho, \theta$  계수는 급한 변화율을 갖게된다. 즉  $R^9 \rightarrow R^{15}$ 로의 비연속적인 변이를 갖게 되는 것이다. 이와 같은 경우를 위해 각 object는 최대 characteristic domain의 수보다 적은 수의 dimension을 갖는 것이 허용된다.

## V. 보행 로봇의 자동 주행을 위한 Computer Vision 기법

### V-1. 자동주행 로봇의 주행 기법

보행로봇의 주위 장면은 연속적으로 변화하기 때문에 이러한 장면에 대한 해석이나 설명이 종합적으로 잘 연결될 수 있어야 한다. 따라서 장면의 통합(scene integration) 기법이 주행로봇의 컴퓨터시각 설계에서 가장 중요한 관건이라고 할 수 있다[9]. Tou[36]는 3가지 상황으로 나누어 주행상태를 설명했는데 초기화의 경우, 보행중의 경우, 경고상황의 경우가 그것이다.

#### 1. 초기화 상황:

로봇이 정지상태에서 출발을 시작하려면, 우선 자신 주변에 대한 수장의 영상을 채취하여 그에 대한 해석과정을 거쳐야 한다. 영상해석 결과가 나올때까지 제자리에 정지한 채로 대기하고 있다가 영상해석 결과 안전진행경로가 제시되면 그 방향으로 출발하는 것이다. 영상은 로봇의 정면, 좌측, 우측에 대한 석장이 필요하며, 거리영상이 바람직하다. 해석결과는 "전방으로 X 거리만큼 진행", "Y 거리만큼 진행뒤 좌회전"등의 식으로 전달된다.

#### 2. 보행 상황:

초기화 이후 로봇이 보행을 시작하면 시각해석 시스템은 연속적으로 진행방향의 영상을 취득 하고

해석을 한다. 실시간 영상해석을 위해서는 다양한 방법이 구현될 수 있을 것이나 쉬운 방법은 매 영상의 새로운 입력부분, 즉, 앞쪽의 수 열(row)에 대한 변화만을 찾아내는 방법을 생각해 볼 수 있다. 이렇게 해서 로봇이 매 보행을 할 때 마다 진행방향이 순탄한지 또는 장애물이 나타나는지를 판단해서 주행시스템에 알려주어야 한다.

### 3. 경고 상황:

주행 예정경로상에 장애물이 나타났을 경우 경고 신호를 주행시스템에 보내주게 된다. 경고 신호를 받으면 장애상황을 해석하고 새로운 명령을 내릴 수 있는 시간을 제공한다. 비전시스템은 주변장면을 해석해서 새로운 진행방향을 제시할 수도 있고, 또는 장애물의 해석을 통하여 그것이 단순한 글미사라든가 낮은 높이를 갖는 장애물이라든가 하는 판단을 내릴 수도 있다.

## V-2. 자동주행 로봇의 Vision System

보행 로봇의 유도조정을 위해서 기본적으로 고려해야 할 사항으로는 ① 진행 방향 유도(directional guidance) 능력 ② 장애물 회피능력 ③ 방향별 정보 결정능력 ④ 거리정보 측정능력 ⑤ 물체 확인 능력 등이다.

진행방향 유도는 안전경로 계획(safety path planning) 기법을 이용할 수 있으며, 장애물 회피 능력을 주기 위해서 필요한 것은 그림자를 구분해 낼 수 있는 알고리즘 및 직립형 물체인지 납작한 물체인지등을 구분할 수 있는 방법이 필요하다. 방향별 정보를 획득하기 위해서는 시야내의 배치물 분포 기억도를 작성해야 한다. 거리정보는 range finder를 이용해서 보행로봇의 현 위치에서 주요지점까지의 거리를 직접 측정할 수 있으며 주행로상의 장애물 형태를 확인해야 한다. 주행환경을 분석하거나 이해하기 위해서는 인공지능기법(AI)과 지식기반(KB)의 이용이 바람직하다.

자동주행로봇의 경로를 정정해 주기 위해서는 먼저 Camera의 입력영상에 대한 영상처리, 즉 전처리를 해주고 그 처리결과를 이용하여 영상의 해석이 이루어진다.

#### 1. 방향 유도 기능:

로봇 주행의 초기화 상황에서 필수적인 기능이 다. 기본 가정은 로봇의 출발점은 안전한 지점이라는 것이다. 즉, 안전한 장소에 정지해 있다가 출발을 시작한다고 가정을 한다. 전면방향의 입력영상을 일

정한 크기의 격자로 나누어서 각 격자내의 pixel값들의 평균  $G$ 를 계산한다. 로봇트가 주행하는데 문제가 없을 정도의 넓이를 갖는 공간상의 영역을  $R$ 이라고 하고  $R$ 내에는  $N_1$  개의 격자가 있다고 가정하여,  $R$ 내의 모든 격자들에 대한 평균  $M_1$ 을 1차로 구한다.

$$M_1 = \frac{1}{M_1} \sum_k \sum_T G(k, T)$$

다음에는 1차로 구한 평균과 각 격자들의 평균을 비교하여 그 차이가 크지 않은 격자들을 선택한다.

$$M_1 - T \leq G(k, T) \leq M_1 + T$$

선택된 격자들만을 대상으로 2번째의 평균  $M_2$ 번째의 평균  $M_2$ 를 구한다.

$$M_2 = \frac{1}{M_2} \sum_k \sum_T G(k, T)$$

$N_2$ 는  $N_1$ 에서  $G(k, T)$ 과 큰 차이가 나는 격자들을 제외한 격자수를 뜻한다. 다시  $M_2$ 와 큰 차이가 나는 격자들을 제거한 뒤 나머지 격자들만의 평균을 구한다. 이 과정은 평균계산에 사용된 모든 격자가 최종 평균값과 일정한 차이 이내에 들어올 때까지 계속된다. 이와 같이 계산된 최종 평균값  $M$ 은 로봇트 주행 예정 방향의 공간상의 영역  $R$ 의 평균 greylevel을 대표한다고 할 수 있다. 이상의 계산 정보를 가지고 유도 방향을 결정하되, 다음의 4가지 서로다른 경우를 예상할 수 있다.

**전방으로의 진행 경우**

주어진 영상에서 설정한 공간 영역  $R$ 내에서 구한 전체 평균  $M$ 의 계산에 사용된 격자가 전방으로, 즉  $Y$  방향으로 연속해서 존재할 경우에는 전방으로의 진행에 아무 문제가 없다고 판단한다. 만일  $Y$  방향으로 3개 이상의 비정상적인 평균값을 갖는 격자가 연속으로 존재하면 장애물이 있다고 판단하고 경고 신호를 발하고 전방주행을 중단한다.

**방향을 전환해야 하는 경우**

장애물의 전방에 있는 수 행의 격자가 평균계산에 사용된 정상적인 격자이고 장애물 좌우의 격자도 정상적인 것이라면 방향 전환 결정을 내린다. 명령은 "전방으로  $X$ 보 주행 후 우측으로 방향을 바꿀 것" 식

이 될 것이다.

**막다른 골목인 경우**

만일 장애물 좌우의 격자들도 비정상적인 값을 갖고 있다면 로봇트는 더 이상 진행을 할 수 없는 막다른 곳(dead end)에 있다고 판단한다.

**방향을 전환하고 싶은 경우의 점검**

이 경우는 로봇트가 진행을 하고 있다가 스스로 진행길로의 방향을 바꾸고 싶은 경우 이의 허용 여부를 결정해 주는 것이다. 점검은 좌측 또는 우측 전방방향이 정상적인 격자들의 연속으로 되어 있는가를 판단해서 한다.

**2. 장애물 회피 기능 :**

로봇트 비전 시스템이 해석해야 할 가장 어려운 임무중의 하나가 장애물을 판단하는 일이다. 앞의 격자별 평균 grey level 접근 방법으로 볼 때에 장애물로 판단된 뒤에도 확인 과정이 필요하다. 왜냐하면 그림자가 겹다거나 물이 좁고여있거나 하는 경우에는 장애물로 판단될 수 있기 때문이다.

그림자 여부 판단에는 texture의 성질을 이용할 수 있다[36]. 즉, texture의 성질은 그림자 내에서도 그대로 유지된다라는 것을 이용하는 방법으로 histogram이나 정합(correlation)을 이용한다.

우선 분석해야 할 영상 중 문제가 되는 부분을 일정한 subregion(또는 window)으로 분할한다. 윈도우가 모두 정상적인 영역내에 속하면  $W_1$ , 장애물로 예상되는 영역내에 속하면  $W_0$ , 두 영역에 걸쳐서 윈도우가 존재하면  $W_n$ 로 한다.  $W_1$ 에 대한 histogram과  $W_0$ 에 대한 histogram을 구했을 때 전체적인 모양은 유지하면서 좌우로 값만 shift 됐다면 장애물이 아니고 그림자로 판단한다. 정합도를 판단의 근거로 사용할 경우, 아래식과 같이 두 윈도우간의 정합도를 거리를 변화시켜 가면서 구해보고 그 최대값을 두 윈도우간의 정합도  $R$ 로 본다.

$$R_{W_1, W_2}(r, c) = \frac{\sum_x \sum_y W_1(x, y) W_2(x, y)}{\sum_x \sum_y W_1^2(x, y)}$$

$$R = \text{Max}_{r, c} [R_{W_1, W_2}(r, c)]$$

두 윈도우간의 판단지수  $\alpha(W_1, W_2) = |1 - \text{Max}(R(r, c))|$ 라 하여  $W_1$ 와  $W_{r0}$ , 그리고  $W_0$ 와  $W_{r0}$ 간의  $\alpha$  값을

구하여 아래와 같이 판단한다.

- (i)  $\alpha(W_r, W_{r0}) \leq 0.2$ 이고  $\alpha(W_{r0}, W_o) \leq 0.2$  이면 : 장애물이 아닌 그림자이다.
- (ii)  $\alpha(W_r, W_{r0}) \leq 0.4$ 이고  $\alpha(W_{r0}, W_{r1}) \leq 0.4$  이면 : 장애물이고 그림자가 아니다.

- (iii) 위의 두가지 경우에 속하지 않으면 : 판단이 불가하므로 장애물로 본다.

Tou(36)가 제시한 주행 로봇의 vision system 개념은 아래와 같다.

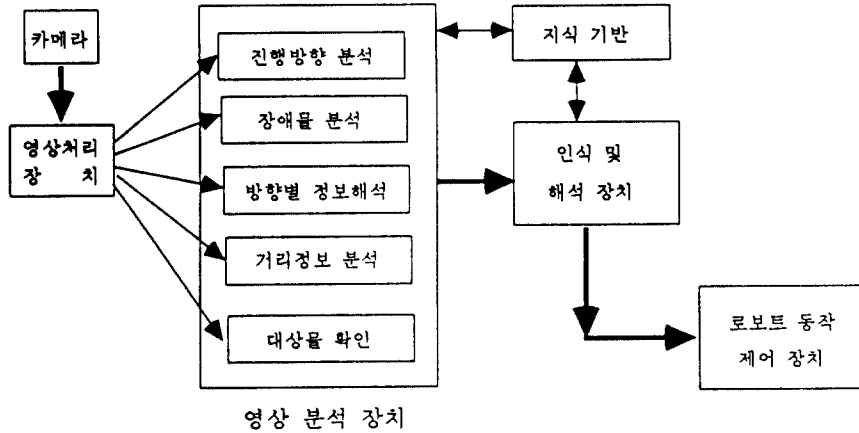


그림 12. 주행 로봇의 비전시스템 개념

### V-3. Robot의 안전주행을 위한 3차원 정보구축 기법

로봇이 주행하면서 주행경로 주위의 장면을 이해해야 한다는 것은 가장 기본적인 요구조건이다. 하지만 실제로는 카메라의 시야(field of view)가 제한적일 수밖에 없으므로 2단계에 걸친 이해 과정이 필요하다. 로봇의 현재 위치에서 가까운 곳에 있는 특징을 찾아내기 위해서는 한 방향으로 근접해서 살펴야 하는 반면, 로봇이 속해있는 전체공간을 크게 이해하기 위해서는 주위를 글로루 살펴야 한다. 후자에 관한 연구는 많이 발표되어 왔는데 88년 Zheng[37]에 의해 발표된 Panorama적 표기법이 대표적이다.

Tsujii[38]는 로봇 주행경로 양편으로 시각 기준점을 두고 이를 이용하여 로봇 주행에 필요한 정보를 추출해 내는 알고리즘을 발표했다. Aloimmos[39]는 시각 기준점을 설정하고 로봇에 장착된 카메라의 시야 정중앙에 기준점이 오도록 고정시키면서 주위를 살필 경우 주위 환경의 3차원 정보를 추출해 낼 수 있음을 보였다. Ballard[40]도 시각 기준점을 사용하되 카메라 2대를 사용하여 stereo를 이용한 3차원

정보를 추정하였다.

로봇이 사진에 정해진 주행경로를 진행한다고 가정하되 카메라의 움직임은 수평면에서의 회전만을 허용할 경우를 가정해 보자. Bruss[41]에 의하면 X, Y, Z 좌표를 camera에 고정시키고, Z 축을 광축(optical axis)에 일치시키고 카메라의 수평운동  $T = (U, V, W)$ , 회전 운동  $R = (A, B, C)$ 로 할 때, 임의의 공간상의 점  $P = (X, Y, Z)$ 의 영상내에서의 순간속도(u, v)는 아래와 같다.

$$u = \frac{f X'}{Z} - \frac{f X Z'}{Z^2}$$

$$v = \frac{f Y'}{Z} - \frac{f Y Z'}{Z^2}$$

여기서 f는 렌즈의 초점거리이다.

Tsujii[38]에 의하면 로봇이 일정하게 움직이면서 경로상의 고정점을 화면의 중앙에 오게 할 때 화면상의 중요 특징점들의 고정점에 대한 상대적인 위치를 추정할 수 있다. 아래 그림과 같이 고정점과 특징점 간의 거리를 L, 각도를  $\theta$ 라 하고, Y를 표면으로부터

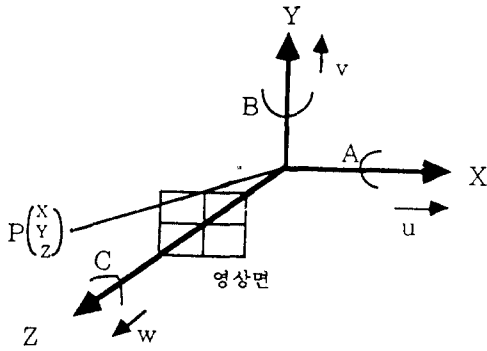


그림 13. XYZ 공간 배치도.

의 높이라 하면 원통좌표계(L, θ, Y)로 3차원 점을 표기하게 된다. 실제로 Y는 고정된 값이 되므로 고려할 필요가 없다. 카메라의 수평 동작을 나타내는 계수 U, V, W와 회전 동작을 나타내는 A, B, C는 아래와 같다.

$$\begin{aligned} U &= U_0 \cos \varphi & A &= 0 \\ V &= 0 & B &= \omega \\ W &= U_0 \sin \varphi & C &= 0 \end{aligned}$$

여기서 U<sub>0</sub>는 로봇의 속도를 나타내며 ω는 카메라의 회전각속도이다. 카메라는 panning만 허용된다고 가정하므로 A = C = 0이 된다.

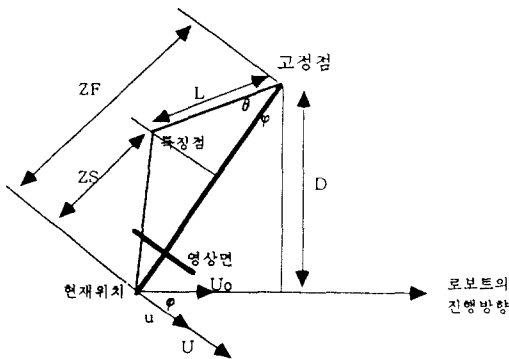


그림 14. 특징점에 대한 L, θ의 계산과정도

카메라 중앙에서 고정점까지의 거리 ZF는

$$ZF = \frac{U_0 \cos \varphi}{\omega} = \frac{D}{\cos \varphi}$$

카메라 중앙에서 특징점까지의 거리 ZS는

$$ZS = \frac{U_0(x \sin \varphi - f \cos \varphi)}{u + f \omega + \frac{\omega x^2}{f}}$$

따라서

$$\frac{ZS}{ZF} = \frac{f - x \tan \varphi}{u + f + \frac{x^2}{f}} = K$$

특정점의 위치를 결정해 주는 L, θ는

$$L = ZF \sqrt{(1 - K)^2 + \frac{K^2 x^2}{f^2}}$$

$$\theta = \tan^{-1} \frac{x k}{f(1 - k)}$$

아래 그림과 같이 로봇트가 한 지점에서 또 다른 지점으로 직선으로 이동하면, 고정점을 기준으로 해서 특징점의 3차원 정보를 계산할 수 있다.

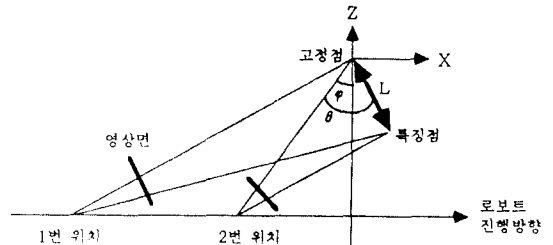


그림 15. 위치이동에 의한 3차원 정보 계산원리

특정점의 X, Z 좌표는 아래식으로 계산된다[38]. Y 좌표는 지상으로 부터의 높이로서 고정된 값이다.

$$X = L \cos(\theta - \varphi - \pi/2)$$

$$Y = \text{고정된 높이}$$

$$Z = L \sin(\theta - \varphi - \pi/2)$$

로봇트가 반경 R인 원주를 따라서 주행하는 원형 운동을 하는 경우에는 카메라 동작을 아래와 같이 나타낼 수 있다.

$$U = U_0 \quad A = 0$$

$$V = 0 \quad B = \omega$$

$$W = 0 \quad C = 0$$

직선 운동과 같은 원리로 K를 구하면

$$K = \frac{ZS}{ZF} = \frac{f}{\frac{u}{\omega} + f + \frac{x^2}{f}}$$

$\theta$ , L은 직선운동의 경우와 동일한 식으로 구하되 K는 윗식을 이용한다.

### 참 고 문 헌

1. R.C.Gonzalez & R.E.Woods, Digital Image Processing, A-W, 1992.
2. B.K.P.Horn, Robot Vision, MIT Press, 1986.
3. B.K.P Horn, "Obtaining Shape from Shading Information", Psychology of Computer Vision, M-H, N. Y., 1975.
4. J.J.Gobson, The Perception of the Visual World, Houghton Mifflin, Boston, Mass, 1950.
5. K.Stevens, Surface Perception from local analysis of texture and contour, Ph.D.thesis, MIT, 1979.
6. J.P.Duncan, D.P.Dean, and G.C.Pate, "Moire Contourgraphy and Computer aided replication of human anatomy", Engineering in Medicine, 9, 1, 29-36, 1980.
7. R.A.Jarvis, "A Perspective on range finding techniques for computer vision", IEEE Trans. on PAMI, 5, 2, 122-139, 1983.
8. K.Prazdny, "On the information on optical flows", CVGIP, 22, 2, 239-259, 1983.
9. B.H.Ballard & O.A.Kimball, "Rigid body motion from depth and optical flow", CVGIP, 22, 1, 95-115, 1983.
10. D.H.Ballard and C.M.Brown, Computer Vision, P-H, 1982.
11. J.L.C.Sanz, Advanced in Machine Vision, Springer-Verleg, New-York, 1980.
12. A.A.Requicha, "Perception of rigid objects", Computer Surveys, 12, 4, Dec, 1980.
13. J.D.Clinton, "Advanced structural geometry studies", NASA, CR-1734, Sep. 1971.
14. R.B.Shudy and D.H.Ballard, "Towards an anatomical model of heart motion as seen in 4-d cardiac ultrasound data", Proc. of 6th conference on Computer Applications in Radiology, June 1979.
15. T.O.Binford, "Visual perception by Computer", IEEE Conference on Systems and Control, Miami, Dec. 1971.
16. D.Marr and H.K.Nishihara, "Representation and recognition of the spatial organization of three dimensional shapes", Proc. of Royal Society of London, B, 200, 269-294, 1978.
17. R.Mohan and R.Nevatia, "Using perceptual Organization to extract 3D structure", IEEE Trans. on PAMI, 11, 11, 1121-1139, 1989.
18. C.L.Jackins and S.L.Tamimoto, "Oct-tree and their use in representing three-dimensional objects", CGIP, 14, 3, 249-270, Nov. 1980.
19. J.L.Bentley, "Multidimensional search trees used for associative searching", Comm.ACM, 18, 9, 505-517, Sept. 1975.
20. R.M.Haralick and L.G.Shapiro, "Computer and Robot Vision", A-W, 1993.
21. B.K.P.Horn, "Extened Gaussian Images", Processdings of the IEEE, 72, 1671-1686, 1984.
22. T.J.Fan, G.Medion, and R.Nevatia, "Recognizing 3-D objects using surface descriptions", IEEE Trans, PAMI, 11, 11, 1140-1157, Nov. 1989.
23. P.Horand and R.C.Bolles, "3DOP's strategy for matching 3-demensional objects in range data", Proc. Int. Conf. Robotics, Atlanta, GA, 78-85, Mar. 1985.
24. W.E.L.Grimson and T.perez, "Model-based recognition and localization from sparse range or tactile data", Int.J.Robotics Res, 3, 3, 3-35, Fall, 1984.
25. K. Ikeuchi, "Precompiling a geometrical model int an interpretation tree for object recognition in bin-picking tasks", Proc. DARPA Image Understanding Workshop, 321-339, Feb. 1987.
26. O.D. Faugeras and M.Herbert, "The representation, recognition and locating of 3-D objects", Int.Robotics Res., 5, 3, 27-52, Fall 1986.
27. M. Ohima and Y.shirai, "A scene description method using three-dimensional information", Pattern Recognition, 11, 9-17, 1979.
28. R. Nevatia and T.O.Binford, "Description and recognition of complex-curved object", Artificial Intell., vol. 8, pp.77-98, 1977.
29. R.A.Brooks, "Symbolic reasoning among 3-D



models and 2-D images", Artificial Intell, vol.17, 285-348, 1981.

30. W.Wechsler & C.L.Zimmerman, "2-D invariant object recognition using distributed associative memory", IEEE Trans. PAMI, 10, 811-821, Nov. 1988.
31. W-C Lin, F-H Liao and T.Lingutla, "A hierachical multiple-view approach to three-dimensional object recognition", IEEE Trans. Neural Network, 2, 84-92, Jan. 1991.
32. T.Poggio and E.Girosi, "Networks for approximations and learning," Proc. IEEE, 78, 1881-1497, 1990.
33. T. Poggio and S.Edelmav "A network that learns to recognize three-dimensional objects", Nature, 343, 263-266, J an. 1990.
34. A.D.Bimbo, L.Landi, and S.Santni, "Three-dimensional panor-faced object classification with Kohonen maps", Optical Engineering, 32, 6, 1222-1234, June 1993.
35. T.Kohonen, "The self-organizing map", Proc. IEEE, 78, 1464-1480, Sept 1990.
36. J.T.Tou, "Computer vision for the guidance of roving fobots", SPIE, vol.595, 192-200, 1985.
37. J.Y.Zheng, M.Asada and S. Tsuji, "color-based panaramic representation of outdoor environment for a mobile robot", Proc. 9th Int. Conf. on Pattern Recognition, 801-803, 1988.
38. J.Y.Zheng and S.Tsuji, "panoramc representation of scenes for routeunderstanding", Proc. 9th ICPR., 151-167, 1990.
39. J.Aloimonos and A.Bandyopadhyay, "Active Vision", Proc. Image Understanding Workshop, 552-573, 1987.
40. D.H.Ballard, "Reference frame for animate vision", Proc. 11th Inter. Joint Conf. Artificial Intelligence, 1635-1541, 1989.
41. A.R.Bruss and B.K.P.Horn, "Passive navigation", CVGIP, 21, 3-20, 1983.

정 동 석 (Dong Seok Jeong)

• 1954년 4월 29일생

- 학 력

- 1977년 2월 : 서울대학교 전자공학과
- 1985년 6월 : 미국 Virginia주립대학교(공학석사) 전자공학
- 1988년 2월 : 미국 Virginia주립대학교(공학박사) 전자공학

- 경 력

- 1977년 3월 ~ 1982년 12월 : 국방과학연구소 연구원
- 전공 : 신호 영상처리, 컴퓨터비전, 패턴인식

- 학위논문

- Nonlinear Image Restoration Using a Segmentation-Oriented Expert System.

주요논문

- 윤곽선방향의 히스토그램과 샘플된 점일치를 이용한 이치형상의 인식 알고리즘 전자공학회 논문지-B, Vol.29B-7, pp.219-226, 1992.
- A Study on the Half toning Algorithm Using Hopfield Neural Network, Signal Processing Applications and Technology, pp.721-726, September, 1993.

주요저서

- 디지털화상처리, 청문각, 1989.
- 전자회로, 광문각, 1992.
- 학회 : 대한전자공학회, 한국통신학회, IEEE