

□ 특 집 □

## 3차원 장면분석을 위한 영상이해 기술의 현황

한국과학기술원 양 현 승\*

● 목	차 ●
I. 서 론	3.1.2 클릭의 정의
II. 장면인식을 위한 영상이해 시스템의 사례	3.1.3 Site값의 결정
2.1 VISIONS	3.1.4 최적값의 결정
2.2 ACRONYM	3.2 영상 레이블링에의 응용
2.3 Nagao와 Matsuyama의 항공사진 분석 시스템	3.2.1 단일노드 클릭함수의 정의
2.4 SPAM	3.2.2 복수노드 클릭함수의 정의
III. Markov Random Field 모델을 이용한 장면인식	3.3 물체인식에의 응용
3.1 MRF 모델이란 무엇인가	IV. 신경회로망을 이용한 영상이해 기술
3.1.1 MRF의 정의	4.1 다층 퍼셉트론을 이용한 물체인식
	4.2 스테레오 비전과 광흐름
	4.2 Hopfield Net을 이용한 화상 레이블링
	V 결 론

### I. 서 론

영상이해란 입력된 영상으로부터 의미 있는 영상구조(선 혹은 영역)를 추출하고 이들을 결합하여 이를 사전에 미리 저장된 물체모델중 하나로 인식하는 작업이다. 이러한 영상이해에 대한 연구는 1970년대 후반부터 1980년 후반 사이에 매우 활발히 전개되어 몇가지 대표적인 시스템들이 개발되어 지면에 발표되었다. 그러나 이들 대부분의 시스템들은 규칙을 기반으로 한 전문가 시스템 기법을 기초로 한것이고 이러한 기호논리에 바탕을 둔 인공지능의 연구가 최근 신경회로망, 퍼지이론 등의 대두와 함께 수그러들기 시작하면서 아울러 이들 영상이해 시스템의 개발에 대한 연구도 주춤거리기 시작하였다. 그

리하여 최근에는 신경회로망을 이용한 영상인식 및 이해기술을 비롯 신경회로망을 닮은 Markov Random Field이론 등을 적용한 영상이해 기법이 활발히 연구되고 있으나 아직은 하나의 시스템으로 개발되기에는 미흡한 단계이다. 따라서 본 논문에서는 과거 10여년간 연구개발되어 왔던 몇가지 대표적인 시스템과 그들의 특성을 알아보고 최근 영상인식 분야에서 각광을 받고 있는 Markov Random Field모델과 신경회로망을 이용한 영상인식 및 이해기술에 대하여 살펴보기로 하겠다.

### II. 장면인식을 위한 영상이해 시스템의 사례

#### 2.1 VISIONS[1]

Hanson과 Riseman은 1978년 VISIONS(Vi-

\* 증신회원

sual Integration by Semantic Interpretation of Natural Scenes)라 명명한 지식기반형 영상이해 시스템을 개발 발표하였다. 이 시스템에서는 주어진 장면에 대한 정보와 지식베이스에 저장되어 있는 정보를 계층적 단계로 표현하였다. 그리고 각각의 독립된 지식원(knowledge sources)은 이들 단계간의 정보변환을 담당함으로써 영상의 화소단위에서부터 물체에 이르기까지의 인식작업을 담당하도록 하였다. 이 VISIONS 시스템은 선언적(declarative) 형태와 프로시듀어(procedural) 형태의 지식 모두를 사용하여 주어진 영상을 분석하고 이에 대한 가정을 설정 및 검증하는 하향식(top-down) 및 상향식(bottom-up)의 인식전략을 수행하였다. VISIONS는 자연야외영상을 이해하는 문제에 적용되었으며 따라서 장면에서 나타날 수 있는 물체를 제한하여 이들 영상을 분석하고 이해하려고 하였다.

## 2.2 ACRONYM[2]

ACRONYM은 3차원 물체를 발견하기 위한 영상이해 시스템으로서 이 시스템에서의 지식표현 기법을 분석하여 보면 다음과 같다.

1. 기호적 지식 표현(Symbolic Knowledge Representation): ACRONYM에서 인식하고자 하는 물체의 모델은 프레임[3]을 사용하여 기호적으로 표현된다. 하나의 프레임은 물체에 대한 구조적 지식을 표현하기 위한 데이터 구조로서 여러개의 슬롯으로 구성되며 이들 슬롯에는 물체의 성질, 다른 물체들과의 관계 및 성질값을 결정하기 위한 프로시듀어와 같은 다양한 정보가 저장된다. 이러한 프레임은 많은 영상이해 시스템에서 널리 이용되어 왔다[1,4]

2. 3차원 물체모델(3D Object Model): ACRONYM은 3차원 물체를 표현하기 위해서 일반화된 실린더(GC: generalized cylinders)모델을 사용하였다. GC는 영상이해에서 3차원 물체의 모양을 모델링하기 위해 종종 사용되어 있다. 즉, 축(axis)을 가진 3차원 물체는 GC를 사용하여 자연스럽게 기술이 가능하다[5]. 그리고 이렇게 해서 표현된 3차원 물체는 어떤 위치에서 물체를 보아도 인식가능하다.

3. 구조적 지식의 구성(Structural Knowledge Organization): ACRONYM에서의 환경모델은 두가지 유형의 계층으로 구성된다. 하나는 Part-of 관계에 바탕을 둔 구성(composition)계층이며 다른 하나는 A-Kind-Of 관계에 바탕을 둔 일반화/세분화 계층이다. 전자는 물체의 기하학적인 구조를 표현하기 위하여 사용되어지며 후자는 추상화의 단계를 바탕으로 물체를 구성하는데 사용된다.

비록 ACRONYM이 일반적 3차원 영상의 이해를 위한 전형적인 틀로서 제시되었지만 인식 결과는 그리 만족할 만하지 못했다. 이에 대한 주요 이유는 (1) 영상특징을 추출하기 위한 영상분할 과정이 완전하지 못하다. (2) 인식전략은 단지 상향식(bottom-up)으로만 진행된다. 따라서 물체의 미발견된 부분들을 발견하기 위한 하향식(top-down)분석이 필요하다.

## 2.3 Nagao와 Matsuyama의 항공사진 분석 시스템[6]

Nagao와 Matsuyama는 블랙보드 모델에 기반을 둔 항공사진 분석 시스템을 개발하였다. 블랙보드는 영상의 영역들의 성질 및 그들간의 관계 등에 대한 모든 정보 및 인식하고자하는 물체에 대한 지식이 저장되어 있는 데이터베이스로서 일단의 분석모듈들은 분석을 위한 데이터를 검증하거나 또는 분석된 결과를 일관된 방법으로 블랙보드로부터 획득 또는 저장한다. 각각의 물체인식 부 시스템(object-detection subsystem)으로부터의 모든 인식된 결과를 블랙보드에 저장함으로써 각각의 부 시스템들은 새로운 물체를 발견하기 위해 이들 결과를 활용할 수 있다. 예를 들어 도로를 발견하는 부 시스템에 의해 도로가 발견되어지면, 자동차를 발견하는 부 시스템은 이를 이용하여 자동차에 대한 후보 영역(candidate regions)을 선택할 수가 있다. 즉, 이미 발견된 도로영역 위의 작은 영역을 자동차에 대한 후보 영역으로 선택할 수가 있게 된다. 시스템의 통제부(controller)는 항상 블랙보드의 내용을 주시하면서 각각의 서로 독립된 물체발견 부 시스템에 의해 얻어진 결과중 서로 모순된

결과를 해결하거나 또는 이들 부 시스템에 의해 분석된 결과로부터 잘못된 초기 영역분할 상태를 복구한다.

이 시스템은 매우 복잡한 항공사진을 분석할 수 있다. 그리고 이 시스템이 나름대로 성공할 수 있었던 것은 블랙모드 모델에 바탕을 둔 개별단위의 시스템 구성(modular system organization) 및 융통성있는 통제 구조에 기인한다고 볼 수 있다. 그러나 영상이해 시스템의 관점에서 볼 때 이 시스템의 성능은 제한되어 있다. 즉,

1. 물체에 대한 모든 지식이 프로그램에 코드화 되어 있다. 따라서 어떠한 지식이 사용되었는가가 명확하지 않으며 이를 수정할 경우 매우 복잡한 프로그래밍 과정을 요한다.

2. 모든 분석은 물체의 2차원적인 모양을 기초로 이루어진다. 때문에 영상단계와 장면단계에서의 정보들간에 구별이 명확하지 않다.

3. 물체의 발견을 위해서 사용된 대부분의 지식은 물체자체의 특징에 관한 것으로서 물체들간의 공간적 관계에 대한 지식의 사용이 미흡하다.

## 2.4 SPAM[7]

SPAM(System for Photo interpretation of Airport using MAPS)은 McKeown 등에 의해 Carnegie Mellon 대학에서 개발된 항공사진분석 시스템으로 이 시스템에서는 영상의 해석을 위한 기본요소들을 영역(regions), 단편(fragment), 기능영역(functional areas) 및 모델의 내단계로 나누어 표현하고 있다. SPAM은 주어진 영상을 분할함으로써 생성된 영역을 단편으로 해석하고 다시 이 단편들을 서로 관련된 것끼리 모음으로써 기능 영역을 생성한 다음 마지막으로 이들 기능 영역들중 일부를 선택하여 장면의 모델을 생성한다.

## III. Markov Random Field 모델을 이용한 장면인식

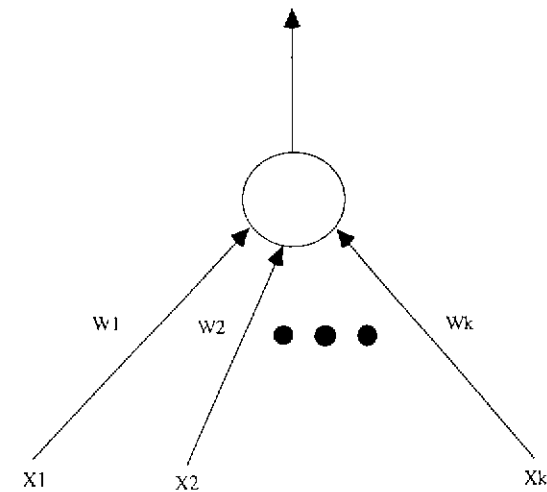
Markov Random Field 모델은 기존의 3차원 장면인식 시스템이 규칙을 기반으로 한 기호적인

접근방법으로 문제를 해결하려 했던 것과는 달리 장면인식 문제를 수리적인 모델로 정형화하여 해결하려 한다는 점에서 크게 다르다. Markov Random Field 모델은 신경회로망이나 퍼지이론처럼 다양하고 광범위한 응용분야를 가지고 있지는 않으나 그 특성상 영상복원/분할[8-13], 영상레이블링[14,15], 양안시각[16], 초기시각을 위한 센서정보의 융합[17,18] 등 주로 컴퓨터 시각분야와 관련된 문제에 유용하게 적용될 수가 있다. 다음 절부터는 Markov Random Field에 대한 기본적인 개념과 함께 이 모델이 어떻게 장면인식 및 물체인식과 같은 영상이해 문제에 적용될 수 있는지 알아보도록 하겠다.

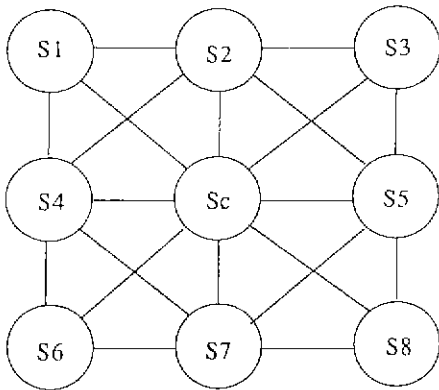
### 3.1 MRF 모델이란 무엇인가

MRF 모델의 가장 큰 특징은 국부성(locality)에 있다. 신경회로망에서 한 노드의 값은 그 노드에 직접으로 연결된 링크를 통하여 들어오는 입력값과 그 연결선들의 가중치에 의하여 결정된다. 즉 다시 말하면, 한 노드의 값은 그 노드와 직접적으로 연결된 노드들의 현재상태 및 그들과의 연결강도를 국부적으로 관찰함으로써 결정될 수 있다. MRF 모델에서도 이와 마찬가지로 한 site(신경회로망에서의 한 노드는 MRF 모델에서 한 site에 해당한다.)가 어떤 값을 가질 것인가 하는 것은 그 site와 공간적 인접관계를 형성하고 있는 국부적인 site들의 값에 의해서만 좌우된다. 그런데 신경회로망에서는 한 노드의 값은 주위에 연결된 노드들의 현재상태 및 연결선의 가중치에 대해서 deterministic하게 결정되지만 MRF 모델에서는 한 site의 값은 주위의 site들의 값에 대해서 확률적으로 정의된다. (그림 1)과 (그림 2)에 신경회로망의 단위구조와 MRF 모델의 기본골격을 예로 나타내었다. (그림 1)을 보면 신경회로망에서의 각 노드의 출력값은 입력연결선의 각 입력값과 연결선가중치의 곱의 합에 의하여 결정된다. 그러면 이번에는 MRF 모델의 주어진 site  $S_i$ 의 값이 어떻게 확률적으로 정의되는지를 알아보기로 하겠다.

#### 3.1.1 MRF의 정의



(그림 1) 신경회로망의 단위구조



(그림 2) MRF 모델에서의 기본골격의 예

(그림 2)를 예로 들어 설명하기로 한다. 그림에서 각 노드들은 임의의 site를 나타내고 이들을 집합  $S = \{S_1, S_2, \dots, S_8\}$ 로 나타내기로 하자. 그리고 이들간에 연결된 고리는 이들 site들간의 공간적 인접성을 나타낸다. 주어진 임의의 site  $s$ 가 가질 수 있는 값(혹은 레이블)의 집합을  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_M\}$ 로 하고 이 값을 변수  $X_s$ 에 의해 표현하기로 하자. 그러면 이들 변수들의 집합은  $X = \{X_{S_1}, X_{S_2}, \dots, X_{S_8}\}$ 로 나타낼 수가 있으며 각각의 변수들에 임의의 값이 부여된 임의의 configuration을  $\omega$ 로 나타내기로 하면  $\{X = \omega\} = \{X_s = \omega_s | X_s \in X, \omega_s \in \Lambda\}$ 가 된다. 그리고  $N_s$ 는 site  $s$ 에 인접해 있는 site들의 집합을 나타낸다고 하자. 즉, (그림 2)를 예로 들면  $N_{S_c} = \{S_2, S_4, S_5, S_7\}$ 가 된다. 그러면

주어진 변수들의 집합  $X$ 는 다음의 두가지 성질을 만족할 때 Markov Random Field라고 한다[8].

Positivity:

$$P(X = \omega) > 0 \text{ for all } \omega$$

Locality:

$$P(X_s = \omega_s | X_r = \omega_r, r \in S, r \neq s) =$$

$$P(X_s = \omega_s | X_r = \omega_r, r \in N_s)$$

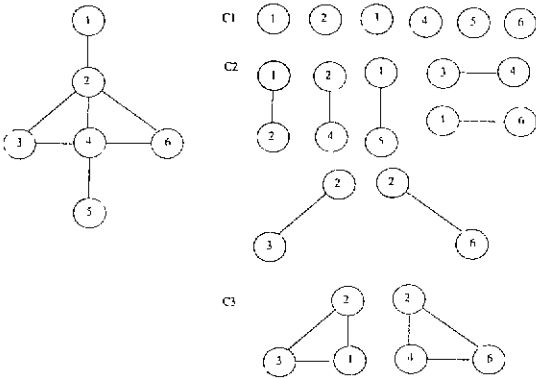
위의 두 성질중 첫번째 성질은 각각의 site들이 가질 수 있는 값들은 사전에 미리 정의되어야 함을 의미한다. 즉, 각각의 site를 영상에서의 각 화소라고 가정하고 이들은 각각 그 화소에 해당하는 명암값을 가진다고 할 때 그 명암값의 범위는 사전에 미리 정의되어야 하며 정의되지 않은 범위내에서의 명암값은 허용되지 않음을 의미한다. 두번째 성질은 MRF 모델을 특징화하는 가장 중요한 성질로서의 임의의 site  $s$ 가 어떤 값  $\omega_s$ 를 가질 확률은  $N_s$ 에 의하여 정의되는  $s$ 에 인접한 site들이 어떤 값들을 갖느냐에만 의존함을 나타낸다.

### 3.1.2 클릭의 정의

각 site들이 특정한 값을 가질 확률을 정의하기에 앞서 이들의 각 site들은 서로 공간적으로 인접한 site들로부터 영향을 받기 때문에 우선적으로 이들 서로 영향을 주고 받을 수 있는 공간적 인접구조를 정의하여야 한다. 이러한 공간적 인접구조를 클릭이라고 하며 앞서 (그림 2)에서의 MRF 모델의 기본구조를 양방향 그래프(undirected graph)라고 하면 클릭은 이러한 그래프의 부그래프(subgraph)로서 임의의 클릭에 속한 모든 노드들은 서로 fully connected되어 있게 된다. 이러한 클릭을 그 클릭에 속한 노드의 갯수에 따라 1개인 경우, 2개인 경우, 3개인 경우를 각각  $C_1, C_2, C_3$ 로 분류하기로 하자. 그러면 임의의 그래프에 대하여  $C_1, C_2, C_3$ 는 (그림 3)과 같이 분류될 수가 있다.

### 3.1.3 site 값의 확률적 정의

이제 각 site가  $\Lambda$ 에 속한 임의의 값을 레이블로 가질 확률을 정의하기로 하자. site 각각에 대하여



(그림 3) 임의의 그래프에 대한 클릭분류의 예

임의의 값이 부여된 configuration  $\omega$ 를 고려하면  $P(X=\omega)$ 는 다음과 같이 정의된다.

$$P(X=\omega) = Z^{-1} e^{-U(\omega)/T}$$

여기서는  $U(\omega)$ 는 에너지 함수로서 주어진 문제에 대한 입력 데이터로부터 구성된 그래프내에 존재하는 모든 클릭함수의 합으로서 정의된다. 즉, 다음과 같이

$$U(\omega) = \sum_{c \in C} V_c(\omega)$$

로 정의될 수 있다. 이러한 클릭함수는 문제의 성격에 따라 매우 다양하게 정의될 수 있으며 이들 클릭함수를 적절하게 정의함으로써 많은 문제에 MRF 모델을 적용할 수가 있게 된다.

### 3.1.4 최적값의 결정

신경회로망에서는 문제의 해결과정이 연결선의 가중치를 변화시키면서 각각의 노드의 출력값이 변하지 않을때까지 또는 출력층의 노드들의 출력값과 바람직한 출력값과의 차이가 어느정도 이상으로 줄어들때까지 학습시킴으로써 이루어진다. 이에 대해서 MRF 모델에서는 각각의 site에 임의의 값을 부여하였을때 확률이 최대화되는 혹은 에너지 함수의 값이 최소화되는 configuration을 구하는 것이 문제의 해결과정이다. 이러한 configuration을 구하는 과정은 site의 수 및 각 site가 가질 수 있는 값의 수가 많아짐에 따라 기하급수적으로 증가하기 때문에 단순한 탐색알고리즘으로 해결하는 것이 매우 비효율적이다. 따라서 이러한 문제점을 해결하기

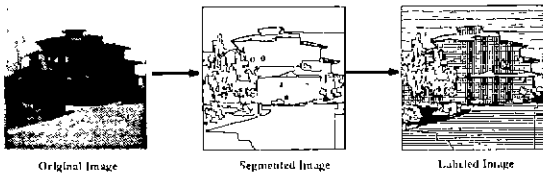
위해서 Simulated annealing[19] 등과 같은 최적화방법 등이 MRF 모델을 적용한 문제의 해결에 많이 이용되어오고 있다.

MRF 모델은 모델자체가 공간적 인접성을 기본으로 하는 것이기 때문에 주로 영상을 다루는 컴퓨터시각 관련분야에 폭넓게 적용되어 왔다. MRF 모델을 문제에 적용하기 위해서 제일 먼저 고려해야 할 것은 주어진 문제위에 MRF 모델을 어떻게 정의할 것인가 하는 것이다. 즉, 무엇을 site로 정의하고 각각의 site들은 어떤 값을 가질 수 있는가를 먼저 정의하여야 한다. 그리고 나서 가장 중요한 것은 우리가 원하는 올바른 해답을 얻기 위해 클릭함수를 적절하게 정의하여 주는 것이다. 이러한 클릭함수는 달리 말하면 문제에 대한 사전지식(a priori knowledge)또는 가능한 해의 가짓수를 줄이기 위한 제한조건(constraints)으로 해석될 수 있다. 즉, 각각의 site들은 모든 값들을 가질 수가 있는데 이중 주어진 제한조건을 만족하는 것들만을 얻기 위해서 제한조건을 클릭함수를 통하여 부여하는 것이다.

이러한 클릭함수는 앞장에서 예로 든것과 같이 site의 수가 한개, 두개 혹은 세개 정도의 경우만을 고려하면 대체적으로 문제해결이 가능하기 때문에 주어진 문제의 국부적 성질만을 고려함으로써 전체적으로 바람직한 결과를 얻을 수 있게 되는 것이다. 그러면 몇가지의 응용사례를 살펴봄으로써 MRF 모델이 어떻게 실제문제에 적용될 수 있는가를 알아보기로 하겠다.

### 3.2 영상레이블링의 응용

본 절에서는 MRF 모델을 영상레이블링에 적용한 예를 설명하기로 한다[14,15]. 영상레이블링이란 입력데이터로서 들어온 영상을 영상분할 알고리즘에 의하여 성질이 균일한 다수의 영역으로 분리한 후 각각의 분리된 영역의 성질 및 영역들간의 공간적 인접관계를 고려하여 주어진 영역을 하나의 물체로 명명하는 것이다. (그림 4)에 영상레이블링을 간단한 예를 통하여 개념적으로 나타내었다.  $N$ 을 영상분할기법에 의해서 결과된 분할된 영역의 수라 하자. 그리고  $M$ 을 각 영역에 부여가능한 레이블의 수라 하자. 이제



(그림 4) 영상레이블링의 개념적 처리과정

각 영역을 실제영상에서 나타날 수 있는 물체에 대한 레이블중 하나로 명명하고자 한다. 분할된 각 영역을 site로 하고 물체에 대한 레이블을 각 site가 가질 수 있는 값들이라 하면 문제는 다음과 같이 정의될 수 있다.

$$X = \{X_1, X_2, \dots, X_N\}$$

$\Lambda = \{\text{sky, foliage, road, grass, wall, roadline, window}\}$  such that  $X_i \in \Lambda$  for all  $i$ .

클릭함수를 정의하는데 있어서 본 논문에서는 두가지 유형의 클릭 즉, 단일노드로 구성된 클릭과 두개의 인접한 노드로 구성된 클릭만을 고려하기로 한다. 전자의 것은 영역위에서의 1차 지식(unary constraints)을 규정하기 위한 것이고 후자의 것은 영역들간의 2차지식(binary constraints)을 규정하기 위한 것이다. 각각을 간단히 정의하면 다음과 같다.

### 3.2.1 단일노드 클릭함수의 정의

우선 단일노드로 이루어진 클릭에 대한 함수를 고려하기로 하자.  $c$ 를  $R$ 에 속하는 임의의 단일 영역  $R_i$ 로 이루어진 클릭이라 하자. 그러면 이에 대한 클릭함수는 다음과 같이 정의할 수 있다.

$$V_c(X_i = \lambda_i | F_i) = \alpha$$

여기에서  $\lambda_i$ 는 영역  $R_i$ 에 부여된 레이블이고  $F_i$ 는 영역  $R_i$ 에 대한 특징을 나타내는 함수이며  $\alpha$ 는 파라미터로서 클릭함수의 값을 나타낸다. 일반적으로 최적의 영상레이블링결과는 영역위에서 측정된 특정값과 그 영역에 부여된 레이블에 대한 모델지식이 가장 부합되는 것이어야 한다. 따라서 파라미터의 값을 정의하기 위한 일반적인 원리는 다음과 같이 말할 수 있다. 즉, 클릭에 속한 영역의 특징이 그 영역에 부여된 레이블에 대한 모델지식과 부합되면 클릭함수의 값을 감

소시키고 이는 곧 에너지 함수값을 감소시킨다. 이와 반대의 경우에는 클릭함수의 값을 증가시키므로써 에너지함수의 값을 증가시키도록 클릭함수를 정의하면 에너지함수의 최소화를 통하여 올바른 레이블링 결과를 얻을 수 있다. 이 원리에 바탕을 둔 파라미터 값은 통계적인 방법이나 또는 신경회로망과 같은 기법을 이용하여 결정할 수가 있으며 이에 대한 논의는 접어두기로 한다.

### 3.2.2 복수노드 클릭함수의 정의

이제 복수노드에 대한 클릭함수를 정의하기로 하자.  $c$ 를 두개의 인접한 영역  $R_i$ 와  $R_j$ 로 이루어진 클릭이라 하자. 그러면 이에 대한 클릭함수는 다음과 같이 정의될 수 있다.

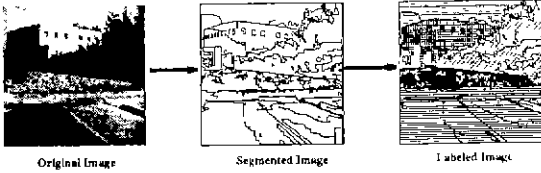
$$V_c(X_i = \lambda_i, X_j = \lambda_j | M_{ij}) = \beta$$

여기에서  $\lambda_i$  및  $\lambda_j$ 은 각각 영역  $R_i$  및  $R_j$ 에 부여된 레이블이다. 그리고  $M_{ij}$ 는 두영역간에 측정된 공간적관계 혹은 2차특징을 나타낸다. 앞서와 마찬가지로 파라미터  $\beta$ 의 값은 두영역간에 측정된  $M_{ij}$ 의 값과 두 레이블  $\lambda_i$ 와  $\lambda_j$ 간에 나타날 수 있는 2차특징에 관한 모델지식이 얼마나 유사한가에 따라 단일노드 클릭함수의 파라미터값의 결정원리와 같이 정의된다.

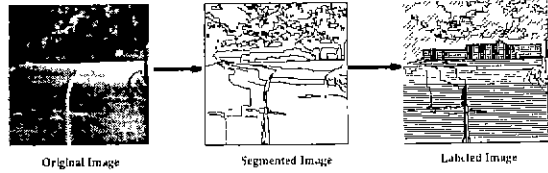
이와같이 파라미터값을 결정하고 나면 최적 레이블은 각 영역에 임의의 레이블을 부여하였을 경우 단일노드 클릭함수의 값과 복수노드 클릭함수의 값의 합, 즉 에너지함수의 값을 최소화 하는 레이블의 조합이 되며 이는 Simulated annealing이나 혹은 다른 일반적인 에너지 최소화 알고리즘에 의하여 구할 수 있다. (그림 5)와 (그림 6)은 이에 대한 실험결과를 나타낸다.

### 3.3 물체인식에의 응용

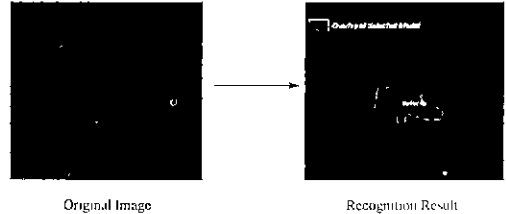
MRF를 이용한 또다른 응용사례로서 CAD나 기타 기하학적인 primitive로 표현하기 어려운 물체의 인식기법을 소개하고 한다[20]. 이 방법에서는 (그림 7)과 같이 물체의 윤곽(contour)을 물체의 중심점으로부터 일정한 각도 간격으로 뻗어나온 반경들의 길이로서 물체를 표현한다. 이를 바탕으로 문제를 정의하면 다음과 같이 될



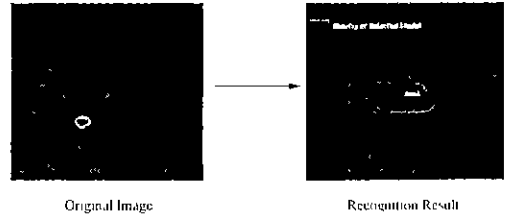
(그림 5) MRF 모델에 이용한 영상레이블링의 실험예



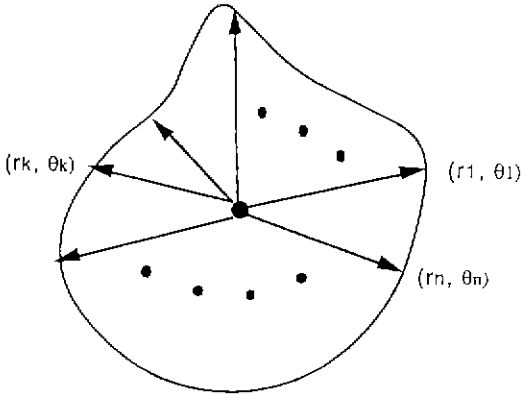
(그림 6) MRF 모델에 이용한 영상레이블링의 실험예



(그림 8) 실제영상에 대한 물체인식 실험결과



(그림 9) 실제영상에 대한 물체인식 실험결과



(그림 7) 극좌표에 의한 물체윤곽의 표현

수 있다. 여기에서  $\theta_i$ 는 일정한 각도만큼 사이를 두고 있으며 시계 반대방향으로 순서가 매겨져 있다고 가정한다.

$$X = \{r_1, r_2, \dots, r_n\}$$

$$\Lambda = \{1, r, \dots, \text{maxradius}\} \text{ such that } r_i \in \Lambda \text{ for all } i.$$

이러한 물체의 표현은 극방향으로 순환구조를 이루고 있기 때문에 1차원적인 배열로서 표현할 수가 있고 각 반경들의 이웃관계가 명확하기 때문에 물체윤곽의 유연성 등에 대한 제한조건과 더불어 윤곽위에서의 에지의 강도 및 물체모델과의 정합 정도 등을 정의하면 배경으로부터 물체의 분할과 인식의 과정을 하나의 처리과정으로 수행할 수 있다. 이 방법은 지금까지 물체의 분할을 선행결과로서 얻고 물체인식은 이에 대한

후처리로서 수행하던 기존의 방법이 가지고 있는 단점에 대한 해결책을 제시해 줄 수가 있으므로 매우 유용하다. 이 문제에서 site는 중심으로부터 물체의 윤곽까지 뻗어나온 반경들이 되며 각 site의 값은 이 반경의 길이가 된다. 클릭함수는 물체의 윤곽은 유연하다는 제한조건과 물체의 윤곽선위에서는 강한 에지가 존재한다는 가정 및 모델의 극좌표 표현과의 정합정도를 반영하도록 정의함으로써 영상에 나타난 물체를 올바른 물체로 인식할 수가 있다. (그림 8)과 (그림 9)는 실제 영상에 대한 실험결과를 나타내고 있다.

#### IV. 신경회로망을 이용한 영상이해 기술

본 장에서는 최근들어 활발하게 연구되고 있는 분야로서 신경회로망을 이용한 화상인식 및 영상이해 기술에 대해서 설명한다.

신경회로망은 기존의 화상처리 기술이 갖는 여러가지 문제점을 보완할 수 있게 하는데, 우선 그중 대표적인 것으로 충분한 속도 및 융통성의 결여와 형상표현의 비유연성 등을 들 수 있다. 신경회로망은 간단한 기능을 갖는 다수의 처리기들이 인접한 처리기들과 정보를 주고 받으면서 동작하는 형태로 이루어지므로, 병렬 하드웨어를 구현하기에 적합하다. 이러한 특성은 컴퓨터비전에서 광대한 양의 화상 데이터를 처리함에 있

어 연산을 분산 병렬 처리할 수 있게 함으로써 처리 시간을 크게 단축시킬 수 있다. 두번째의 유용한 특성으로 신경회로망의 적응적 학습능력(adaptive learning capability)을 들 수 있다. 다시 말해서 사람이 직접 문제해결을 위한 해법을 구체적으로 프로그래밍하는 것이 아니고 시스템으로 하여금 주어진 예로부터 연결선의 가중치를 조정함으로써 스스로 문제의 해법을 발견하게 한다. 이러한 특성은 화상처리 시스템 구현에서 범용성과 융통성을 제공한다. 다음으로 결합 극복능력을 들 수 있는데, 신경회로망은 가중치나 처리기의 출력값이 어떤 고장에 의해 변경되거나 소실되더라도 전체 성능이 급격히 저하되지 않는다는 특징을 갖는다. 그러므로 신경회로망은 지식의 내재적인 표현기능을 갖는데 이는 문제에 관한 정보를 사람이 직접적으로 표현할 필요가 없고, 연결선의 가중치에 의하여 내부적으로 표현됨으로써 복잡한 화상처리 시스템 구현에서 유연성을 제공한다.

다음 절부터는 신경회로망을 이용한 화상인식 및 영상이해의 대표적인 몇가지 응용사례들을 간략하게 소개하기로 한다.

#### 4.1 다층 퍼셉트론을 이용한 물체 인식

전통적인 패턴 인식과정은 일반적으로 크게 특징 추출(feature extraction)과정과 분류(classification)과정의 두 단계로 이루어진다. 특징 추출과정은 입력패턴으로부터 서로 다른 부류에 속하는 패턴들을 효과적으로 구분지어 줄 수 있는 특징 벡터를 구성하게 된다. 특징은 패턴의 구조 등으로부터 휴리스틱(heuristic)에 의하거나, 또는 수학적방법에 의해 추출될 수 있다. 영상에 대하여 수학적방법에 의해 추출되는 특징의 예로서 모멘트(moment)나 FT(Fourier Transform) 등을 들 수 있다.

신경회로망을 이용하여 패턴인식 시스템을 구현하는 데는 신경회로망입력의 형태에 따라 두가지로 분류될 수 있다. 첫째는 특정 추출단계를 거치지 않고 원래의 패턴을 직접 신경회로망의 입력으로 사용하는 경우이고, 또다른 방법은 원래의 패턴인식 시스템에서 분류기에 해당하는

부분만을 신경회로망으로 구성하는 방법이다. 즉 입력패턴은 특정 추출과정을 거쳐 생성된 특징들을 신경회로망의 입력으로 사용하게 된다.

다층 퍼셉트론 신경회로망은 실제로 여러 응용분야에서 가장 많이 사용되는 신경회로망이라 할 수 있다. 최근까지 국내에서도 한글 문자 인식 등 여러 응용에서 다층 퍼셉트론을 사용하는 연구가 활발하게 시도되고 있다. 다층 퍼셉트론을 사용한 화상인식 기법의 예로서, 위에서 제시한 두번째의 방법을 채택한 경우로 모멘트(moment)를 이용한 화상인식기를 들 수 있다[21,26]. 이 방법은 모멘트로부터 유도된 모멘트 불변량(MI: Moment Invariant)을 사용하는데, 모멘트 불변량 자체가 물체의 위치, 회전, 크기 변화에 무관한 성질을 가지고 있기 때문에 신경회로망 분류기의 비선형적 의사 결정 기능과 잡음에 강한 특성과 더불어 좋은 결과를 얻을 수 있게 된다. 모멘트 불변량은 화상의 기하학적인 모멘트로부터 유도된 6개의 연속치를 갖는 함수이다. 이 6개의 모멘트 불변량을 신경회로망의 입력으로하여 신경회로망을 학습시킴으로써 화상 인식기를 구현할 수 있다. 이러한 방법은 원화상을 입력으로 선택한 경우에 비하여 신경회로망 노드의 수를 획기적으로 감소시킴으로써 경제적인 측면 뿐만 아니라 학습 효율 등의 측면에서 유용하다. 또한 모멘트 불변량이 제공하는 장점이 그대로 반영되어 불변인식(invariant recognition)을 가능하게 한다.

#### 4.2 스테레오 비전과 광 흐름(Optical Flow)

앞 절에서 기술한 내용 외에도 신경회로망을 이용한 컴퓨터 비전에 관한 연구들이 여러가지가 있는데 그중에는 스테레오 비전에 관한 연구와 광 흐름 계산에 관한 연구를 들 수 있다. Sun은 스테레오 비전의 연속성과 불변성을 자동적으로 학습할 수 있는 고차의 순환적(recursive) 신경회로망 구조를 제안하였다[22]. 제안된 신경회로망에서 연결선의 가중치는 Hebbian 학습법칙에 의해 학습되는데, Hebbian 법칙은 선형적이고 지역적이기 때문에 분석적인 계산이 가능하다. 또한 가중치를 대칭적으로 학습시킴으



로써 항상 수렴이 가능하며, 한두번의 반복만으로 불균형의 접합을 조정할 수 있다. 또한 Stewart 등은 계층적 신경회로망을 스테레오 비전에 사용하였는데, 이 신경회로망은 불변성, 상제 접합 다중근, 연속성 등의 제한조건을 병렬적으로 처리함으로써 단일 제한조건을 사용한 모델보다 성능이 우수함을 보였다[23]. 제한된 신경회로망에서 각 노드는 해당 접합을 나타내고, 후보 접합 노드 사이의 가중치에 제한조건을 포함시키며, 후보 접합을 지원하는데만 제한조건을 사용하였다. 일반적으로 알려진 광 흐름 계산 기법은 화상의 밝기를 기반으로 한 것과 모서리나 선형 특성 같은 특징을 기반으로 하는 방법이 있는데, 이러한 경우 모서리의 회귀성, 잡음에 의한 찌그러짐, 회전 등에 민감하다는 문제가 있다. 광 흐름 계산에 관한 신경회로망 연구로는 Zhou 등의 연구를 들 수 있는데, 그는 이러한 문제를 해결하기 위하여 불연속적인 영상에서 연속적인 화상의 밝기와 주곡률을 구하기 위한 2차원 다각형 전개(2-D polynomial expansion)를 사용하였다[24].

### 4.3 Hopfield Net을 이용한 영상 레이블링

앞 장에서 기술한 MRF를 이용한 영상이해 기법과 유사하게 영상내의 각 영역에 대한 제약 조건으로부터 에너지함수를 정의하고 이를 최소화하는 과정으로 Hopfield Net을 이용할 수 있다[25]. 주어진 영상을 MRF의 특성을 만족한다는 가정하에서 각종 제약조건(constraints)을 에너지함수로서 모델링한다. 이어서 영상내의 가능한 모든 영역과 레이블의 쌍에 대하여 이항관계(binary relation)을 의미하는 2차원 행렬로서 이른바 호환성 행렬(compatibility matrix)을 구성한다. 이 행렬에 대하여 에너지함수를 적용하여 전체 에너지 값이 감소하는 방향으로 레이블링을 변경시켜 나간다. 최소의 에너지 값으로 수렴하면 그때의 매칭이 최적의 레이블링이 된다.

## V. 결 론

본 고에서는 3차원 장면인식을 위한 영상이해

기술의 현황에 대해서 알아보았다. 우선 이 분야에서 특이할만한 것은 적어도 80년대 후반까지는 몇가지의 대표적인 시스템들만이 나름대로의 응용범위에서 좋은 성능을 발휘하고자 했을 뿐 기술적인 측면에서 어떤 흐름을 찾아가 매우 어려웠다는 것이다. 이것은 결국 기존의 개발된 시스템은 그 시스템 자체로서만이 의미를 갖고 이것이 차후의 다른 시스템의 개발에 결정적인 기여를 하는 예가 극히 드물었다고 볼 수 있다. 그러나 다행히도 최근들어 신경회로망이나 Markov Random Field와 같은 모델이 주목을 받기 시작하면서부터 영상이해에 관한 분야에서도 보다 일반적이고 체계적인 방법론에 관한 연구가 활기를 띠어가고 있다는 것은 매우 흥미로운 일이라 하겠다. 따라서 본 고에서도 기존에 개발되었던 대표적인 시스템 몇가지를 살펴보고 아울러 Markov Random Field 모델과 신경회로망을 이용한 화상인식 및 영상이해에 관하여 보다 상세히 다루었다. 특히 아직은 영상이해 분야에서 생소하지만 충분한 가능성을 가지고 있는 Markov Random Field 모델을 이용한 응용사례에서는 실제 영상을 통한 실험결과도 소개하였다.

## 참 고 문 헌

1. A. R. hanson and E. M. Riseman, "VISIONS: A Computer System for Interpreting Scenes", in *Computer Vision Systems*, Academic Press, New York, 1978, pp. 303~334.
2. R. A. Brooks, "Symbolic Reasoning Among 3-D Models and 2-D Images", *Artificial Intelligence*, Vol. 17, 1981, pp. 285~348.
3. M. Minsky, "A Framework for Representing Knowledge", in *The Psychology of Computer Vision*, McGraw-Hill, New York, 1975, pp. 211~277.
4. V. S. S. Hwang, L. S. Davis and T. Matsuyama, "Hypothesis Integration in Image Understanding", *Computer Vision, Graphics and Image Processing*, Vol. 36, 1986, pp. 321~371.
5. D. marr, *Vision*, Freeman, 1982.
6. M. Nagao, T. Matsuyama and H. Mori, *A Structural Analysis of Complex Aerial Photograph*, Plenum Press, New York, 1980.

7. D. M. McKeown, W. A. Harvey and J. McDermott, "Rule-based Interpretation of Aerial Imagery", *Proc. of IEEE Workshop on Principles of Knowledge-based Systems*, Denver, Colorado, Dec. 1984.
8. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images", *IEEE Trans. Pattern Analysis Mach. Intell.*, Vol. 6 1984, pp. 721~741.
9. H. Derin, H. Elliott and D. Geman, "Bayes smoothing algorithms for segmentation of binary images modeled by Markov Random Fields". *IEEE Trans. Pattern Analysis Mach. Intell.*, Vol. 6, 1984, pp. 707~720.
10. S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs Random Fields using simulated annealing", *IEEE Trans. Pattern Analysis Mach. Intell.*, Vol. 11, 1989, pp. 799~813.
11. R. C. Dubes, A. K. Jain, S. G. Nadabar and C. C. Chen, "MRF model based algorithms for image segmentation", *Proc. Int. Conf. on Pattern Recognition*, 1990, pp. 808~814.
12. A. K. Jain and S. G. Nadabar, "MRF model-Based Segmentation of Range images," *Proc. Int. Conf. on Pattern Recognition*, 1990, pp. 667~671.
13. M. J. Daily, "Color Image Segmentation Using Markov Random Field", *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, 1989, pp. 304~312
14. J. W. Modestino and J. Zhang, "A Markov random field model based approach to image interpretation", *IEEE Trans. Pattern Analysis Mach. Intell.*, Vol. 14, 1992, pp. 606~615.
15. I. Y. Kim and H. S. Yang, "Markov Random Field-based Image labeling with Parameter Estimation by Error Backpropagation", *IEICE Trans. on Information and Systems*, Vol. E74, 1991, pp. 3513~3521.
16. S. P. Clifford and N. M. Nasrabadi, "Integration of Stereo Vision and Optical Flow Using Markov Random Field", *Proc. ICNN*. 1988. pp. 577~584.
17. E. B. Gamble, D. Geiger and T. Poggio, "Integration of Vision modules and labeling of Surface Discontinuities", *IEEE Trans. Syst. Man Cybern.*, Vol. 19, 1989, pp. 1576~1581.
18. P. B. Chou and C. M. Brown, "The Theory and Practice of Bayesian Image Labeling", *Int. J. Computer Vision*, Vol. 4, 1990, pp. 185~210.
19. S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimization by simulated annealing", *Science*, Vol. 220, 1983, pp. 671~680.
20. N. S. Friedland and A. Rosenfeld, "Compact Object Recognition Using Energy-Function-Based Optimization", *IEEE Trans. Pattern Analysis Mach. Intell.*, Vol. 14, No. 7, July 1992. pp. 770~777.
21. A. Khotanzad and Y. H. Hong, "Invariant Image Recognition by Zernike moments", *IEEE Trans. on Pattern Anal. and Mach. Intel.*, Vol. 12, May 1990, pp. 489~497.
22. G. Z. Sun, H. H. Chen and Y. C. Lee, "Learning Stereopsis with Neural Networks", *Proceeding of the IEEE First Int. Conf. on Neural Networks*, Vol. 4, June 1987, pp. 345~355.
23. C. V. Stewatr and C. R. Dyer, "A Connectionist Model for Stereo Vision", *Proceeding of the IEEE First Int. Conf. on Neural Networks*, Vol. 4, June, 1987, pp. 215~223.
24. Y. T. Zhou and R. Chellappa, "Computation of Optical Flow Using a Neural Network", *IEEE Int. Conf. on Neural Networks*, Vol. 2, July 1988, pp. 71~78.
25. B. K. Ko and H. S. Yong, "Hopfield Net Based Image Labeling with MRF-Based Energy Function", *Proc. of First Korea-Japan Joint Conf. on Computer Vision*, Oct. 1991, pp. 100~106.
26. 고병기, 양현승, 김진형, "신경망을 이용한 2차원 형상인식", 정보과학회 추계 학술발표회 논문집, Vol. 16, No. 2, 1989, pp. 327~330.

**양 현 승**



1976 서울대학교 전자공학과 졸업  
 1981 ~ 1982 미국 Pennsylvania주립대학 Electro-Optics Lab 연구원  
 1983 미국 Purdue대학교 전 기공학과 석사  
 1986 미국 Purdue대학교 전 기공학과 박사  
 1983 ~ 1986 미국 Purdue대학교 Robot Vision Lab 연구원

1986~1988 미국 Iowa대학교 전기 및 컴퓨터공학과 조교수  
 1988~현재 한국과학기술원 전산학과 부교수  
 1992~현재 한국정보과학회 인공지능연구회 운영위원장  
 관심분야 : 컴퓨터 비전, 영상처리, 신경회로망, 지능형 로보틱스