

# 지능형 로봇 스테레오 비전 처리 기술

Stereo Vision System for Intelligent Robot

## 지능형 로봇 특집

조재일 (J.I. Cho)	H/W컴포넌트연구팀 팀장
최승민 (S.M. Choi)	H/W컴포넌트연구팀 연구원
장지호 (J.H. Chang)	H/W컴포넌트연구팀 연구원
황대환 (D.H. Hwang)	인지로보틱스연구그룹 그룹장

## 목 차

- .....
- I. 서론
  - II. 국내외 기술개발 동향
  - III. 스테레오 비전 처리 기술
  - IV. 결론

지능형 로봇에 사용되는 필수적인 여러 센서들 중 시각센서는 로봇 지각의 중추적인 역할을 담당한다. 로봇 시장의 확대와 함께 이러한 시각 센서를 기반으로 인간과 로봇과의 자연스러운 상호작용에 대한 연구가 활발히 진행되고 있으며 최근에 학교와 연구소를 중심으로 대상체의 삼차원 거리 영상 정보를 함께 얻을 수 있는 스테레오 비전 처리 기술 개발이 활발히 진행되고 있다. 본 고에서는 스테레오 비전 처리에 대한 국내외 연구 개발 동향에 대해 살펴보고 ETRI에서 개발중인 지능형 로봇의 스테레오 비전 처리 기술에 대하여 소개한다.

## I. 서론

최근 로봇 산업은 산업용 로봇에서 출발하여 군 사용이나 과학기술 용도를 벗어나 가정용 로봇으로 까지 그 활동 영역을 넓혀나가고 있다. 네트워크 기반 지능형 로봇 분야는 고성능 서버를 통해 로봇의 지능과 능력을 외부로 분산시켜 “언제 어디서나 나와 함께 하면서 나에게 필요한 서비스를 제공하는 로봇”이란 개념으로 연구 개발되고 있다.

지능형 로봇 개발에는 여러 종류의 센서 기술이 필요하다. 로봇의 충돌 회피, 장애물 감지 등을 위한 거리 센싱 시스템, 로봇의 위치 파악을 위한 위치 추정 시스템 그리고 로봇의 시각 정보 획득을 위한 시각 센싱 시스템 등이 필요하게 된다. 그 중에서도 시각 센서는 변화하는 주위 환경에 융통성 있게 대처할 수 있는 정보를 제공해주는 로봇 시각의 중추적인 역할을 담당한다.

시각 센서는 주로 카메라를 이용하여 실시간으로 영상을 취득하고, 취득된 방대한 양의 영상 데이터를 실시간으로 처리하여 영상으로부터 물체의 크기, 위치, 색깔 등의 정보를 분석하여 물체를 인지하며 이를 통해 로봇의 자율주행 분야, 사용자 의도 및 행위 인식 분야에까지 이용될 수 있다.

최근에는 이러한 시각 센서를 두 개 이상 사용하여 물체의 실시간 3차원 영상 정보를 추출하는 연구가 학계 및 연구소를 중심으로 진행중에 있다. 기술적 측면에서 볼 때 실시간 거리 영상의 생성 기술은 로봇 자율 주행 및 장애물 인식 등의 지능형 로봇 분야의 응용뿐만 아니라 영화, 애니메이션, 방송, 가상 현실 등의 여러 분야에서 요소기술로서 응용될 수 있으므로 그 기술적 파급 효과가 크다고 할 수 있다.

실시간 거리 영상 취득 시스템은 1980년대 말부터 유수의 연구 기관(CMU, 동경공대, INRIA, Teleos Research)에서 연구되어 왔고, 카네기 멜론 대학의 로봇 공학 연구소는 1990년대 초반부터 실시간 거리 영상 취득을 위한 알고리즘과 하드웨어 장치를 개발해 오고 있다. 2~6개의 카메라 영상으로부터 실시간으로 거리 영상을 생성하며, 이를 이용하여

여러 가지 응용 시스템을 개발하고 있다. 국내에서 3차원 거리 영상 취득을 위한 연구는 주로 알고리즘 위주로 여러 대학과 연구소에서 연구되고 있다. 2003년 KIST에서 3차원 복원이 가능한 병렬 처리형 고속 영상처리 하드웨어/소프트웨어 시스템의 개발이 발표된 바 있다[1].

본 고에서는 지능형 로봇에 적용될 수 있는 스테레오 비전 처리에 대한 국내외 연구 개발 동향에 대해 살펴보고 ETRI에서 개발중인 URC 로봇의 스테레오 비전 처리 기술개발에 대해 소개하고자 한다.

## II. 국내외 기술개발 동향

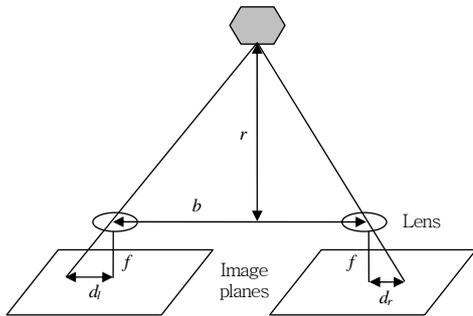
### 1. 스테레오 정합 개념

인간은 두 눈과 뇌에서의 시각피질을 이용하여 입체 공간을 인식할 수 있듯이 이와 유사하게 한 쌍의 카메라로부터 들어오는 영상을 이용하여 관측 공간상의 거리 및 관측 물체의 삼차원 형태를 감지할 수 있다. 그 원리는 한 쪽 카메라의 영상상의 임의의 패턴을 삼차원 공간상의 같은 지점으로 매핑되는 다른 카메라 영상상에서 찾아내는 것이다. 이러한 패턴의 대응관계와 삼각기법을 사용하여 역으로 삼차원 공간상의 위치를 알아낼 수 있다. 여기서 패턴은 픽셀, 원도블록, feature(edge, 꺾적) 등 여러 가지를 사용한 방법들이 시도되어 왔다.

두 대의 카메라를 이용한 스테레오 매칭 기법의

#### ● 용어해설 ●

URC (Ubiquitous Robotic Companion): “언제 어디서나 나와 함께 하면서 나에게 필요한 서비스를 제공하는 로봇”이란 개념으로 2004년부터 정보통신부 주관으로 개발되고 있는 IT 기반 지능형 서비스 로봇으로, 모든 기능을 로봇 내에서 해결하는 기존 로봇 개념 대신 네트워크를 통해 외부 서버에 주요 기능을 분담시켜 낮은 가격으로 로봇을 제작할 수 있게 하고 동시에 세계적으로 기술 우위에 있는 우리의 IT 인프라를 최대한 활용하는 IT 인프라와 로봇의 융합 개념이다.



(그림 1) 스테레오 매칭 시스템의 기하학적 구조

원리를 (그림 1)을 이용해 간략히 설명하면 다음과 같다. 임의의 base line( $b$ )의 거리를 두고 배치된 두 대의 카메라를 통해 촬영된 두 개 영상에서의 피사체 간 거리차(변위, disparity)를 구하고 렌즈의 초점거리( $f$ )를 알면 카메라로부터 피사체 사이의 거리는 (그림 1)의 기하학적 구조를 통해 다음 식을 이용하여 추출할 수 있다.

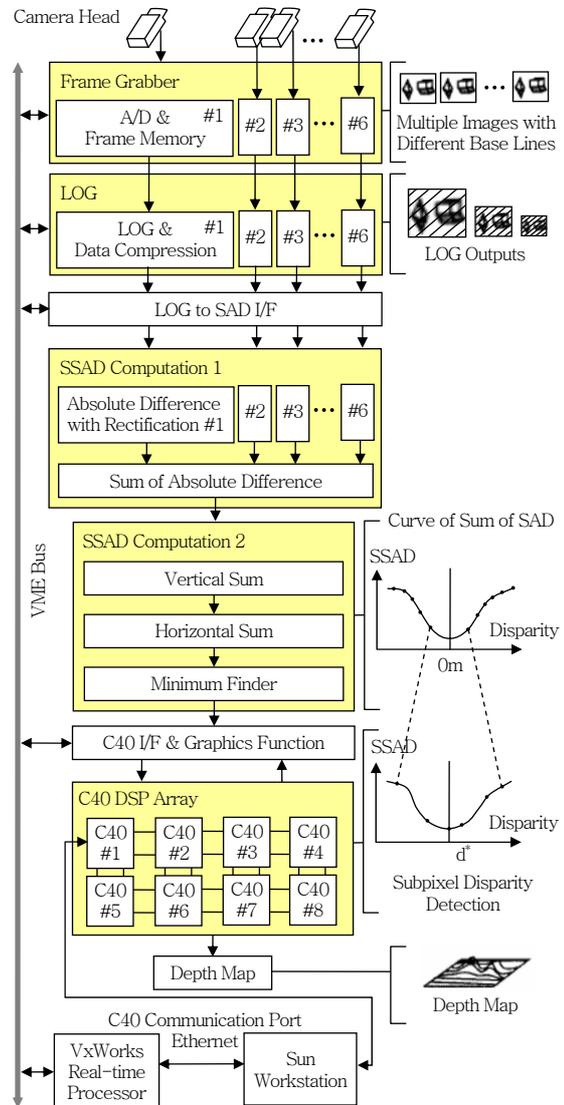
$$r = f \frac{b}{d}$$

여기서,  $r$ 은 피사체와 카메라와의 거리,  $b$ 는 두 대 카메라 사이의 base line 거리,  $f$ 는 카메라 렌즈의 초점 거리이고,  $d$ 는 두 개 영상에 맺힌 동일 피사체 사이의 거리차로서  $d_l - d_r$ 이다.

## 2. 학계 연구 동향

스테레오 카메라를 이용한 실시간 거리 측정 장치의 초기 연구로서 CMU에서는 (그림 2)의 블록도와 같이 여섯 개의 카메라를 동시에 사용하고 FPGA를 이용하고 VME bus에 기초한 실시간 매칭 시스템을 개발하였다[2].

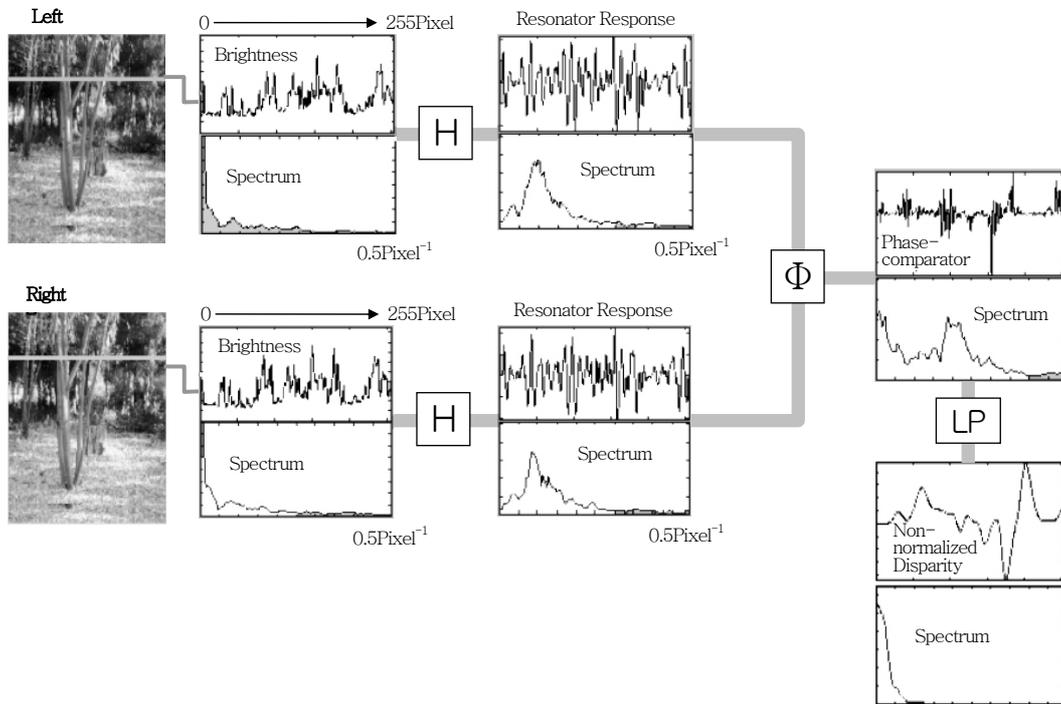
Ruhr University에서는 (그림 3)의 블록도와 같이 스테레오 영상 간의 phase 정보를 이용하여 실시간으로 매칭하는 시스템을 개발하였다[3]. 그러나 이 기법은 영상의 에지 부분의 거리 정보는 정확하게 추출할 수 있는 반면 영상의 에지 부분이 아닌 경우 거리 정보를 추출하지 못하는 단점이 존재한다.



(그림 2) CMU 스테레오 매칭 시스템의 블록도

캐나다의 토론토 대학에서 개발된 스테레오 매칭 시스템의 경우  $360 \times 256$  크기의 영상을 초당 30프레임 속도로 처리 가능하며 비교적 작은 크기의 하드웨어로 구성하였다[4].

국내에서는 포항공과대학교를 중심으로 시스템릭 배열(systolic array) 기반 DP 방식에 기반을 둔 실시간 병렬 처리 스테레오 비전 칩을 개발한 바 있으나 상용까지는 성능상의 제약이 많이 남아 있는 상황이다[5],[6].



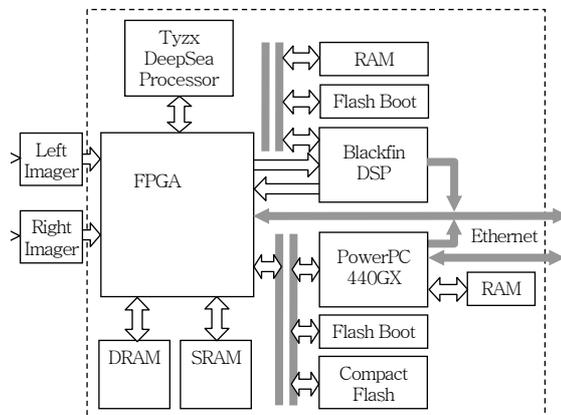
(그림 3) Ruhr 스테레오 매칭 시스템의 블록도

### 3. 제품 개발 동향

#### 가. 미국 TYZX

미국 TYZX사에서 개발된 DeepSea G2라는 스테레오 매칭 시스템은 VGA급 CMOS 카메라를 사용하고 있으며 census transform을 사용하여 계산량을 줄이면서 스테레오 매칭을 초당 200 프레임 정도의 고속으로 처리할 수 있는 기능을 가지고 있다. (그림 4)는 DeepSea G2의 하드웨어 구성과 스테레오 카메라의 사진을 나타내며 임베디드프로세서(Power PC), Tyzx DeepSea 칩, DSP 그리고 FPGA로 구성된다. (그림 5)는 DeepSea G2 시스템을 이용한 실제 영상과 스테레오 매칭 결과에 의한 실시간 depth 이미지를 나타낸다.

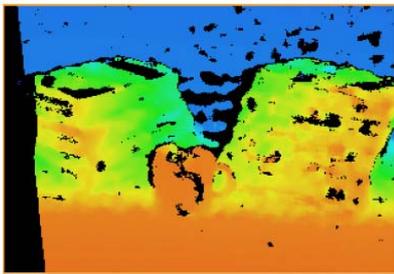
또한 TYZX사에서는 스테레오 비전 시스템의 응용 소프트웨어 패키지로 사람추적과 장애물 검출 및 제스처 인식 소프트웨어를 함께 제공하고 있다[7].



(그림 4) TYZX사의 스테레오 매칭 시스템



(a) Rectified Intensity Image

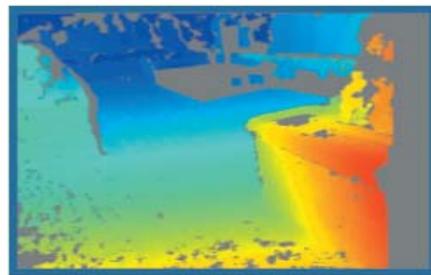


(b) Real-time Depth Image

(그림 5) TYZX사의 스테레오 영상 결과



(a) Raw Images



(b) Depth Image

(그림 7) Point Grey사의 스테레오 영상 결과

#### 나. 캐나다 Point Grey Research

캐나다 Point Grey Research사의 bumblebee 라는 스테레오 비전 카메라 시스템은 스테레오 비전 분야에서 현존하는 최고의 선도기술로 평가되고 있다. Bumblebee는 1/3" 고성능 CCD 카메라를 사용하고 있고, QVGA급에서 초당 2~3 프레임을 처리하고 있으며 (그림 6)과 같은 스테레오 카메라를 사용하고 있다. 카메라 캘리브레이션 오차는 0.05 픽셀 이하로 상당히 정교한 카메라 캘리브레이션 성능을 가지고 있다.

교정된 좌/우 영상을 IEEE1394를 통해 PC에서 수집한 후 PC에서 SAD 알고리즘을 이용해 스테레오 정합을 소프트웨어로 구현하고 있다. Maximum



bumblebee™

(그림 6) Point Grey사의 스테레오 카메라 사진

disparity는 64~230까지 지원하며 (그림 7)은 bumblebee 시스템을 이용한 실제 영상과 스테레오 매칭 결과에 의한 3차원 depth 이미지를 나타낸다.

Point Grey Research사에서도 스테레오 매칭 결과를 이용해 사람추적, 제스처 인식 등의 다양한 응용 소프트웨어를 제공하고 있다[8].

#### 다. 미국 Videre Design

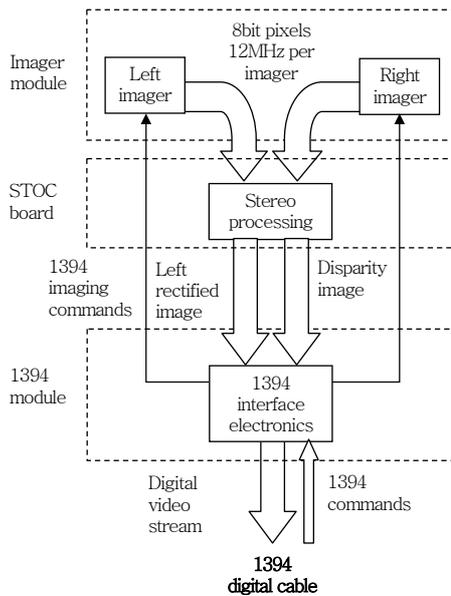
상기 Point Grey Research사에서 제작 공급하고 있는 스테레오 비전 카메라 시스템과 비슷한 저가형 제품으로 미국의 Videre Design사의 STH-MDCS 제품이 있다. 저가격화를 위해 30만 화소의 CMOS 카메라를 채택하고 있고 QVGA급에서 초당 16~17 프레임을 처리하고 있으며 두 대 카메라 사이의 거리(base line)를 가변적으로 조절할 수 있는 제품을 출시하고 있다. (그림 8)은 STH-MDCS 제품 사진이며 Point Grey Research사와 마찬가지로 IEEE1394 I/F를 통해 PC로 한 쌍의 영상을 전송하고 PC에서 스테레오 정합 기능을 구현하고 있다. (그림 9)는 STH-MDCS 시스템을 이용한 실제영상



(그림 8) Videre Design사 스테레오 카메라 사진



(그림 9) Videre Design사의 스테레오 영상 결과



(그림 10) Videre Design사 STOC 제품 구성도

과 disparity image 결과이다.

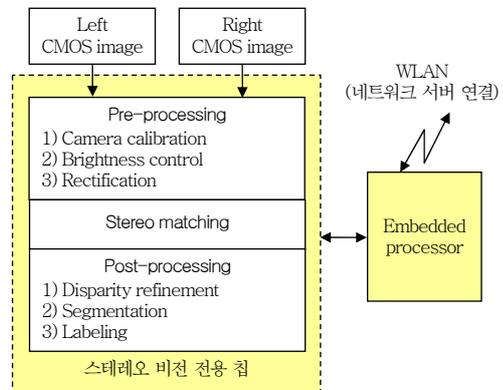
또한 Videre Design사에서는 기존 제품과는 달리 STOC라는 모듈이 장착된 (그림 10)과 같은 블록의 제품을 출시하고 있는데, 이는 disparity image 생성을 비교적 작은 용량의 FPGA를 이용해 하드웨어로 구현한 제품이다[9].

### Ⅲ. 스테레오 비전 처리 기술

다음은 ETRI에서 개발중인 스테레오 비전 처리 기술을 소개한다. (그림 11)은 현재 개발중인 스테레오 비전 처리부의 구성도를 나타낸 것으로 좌/우 영상을 취득하기 위한 두 대의 카메라와 전처리부, 스테레오 매칭부 그리고 후처리부로 구성된다.

가정용 서비스 로봇에 장착을 목표로 하고 있고 저가로 구현하기 위해 시각 센서로는 CMOS 이미지 센서를 사용하고 있으며 스테레오 정합을 포함한 영상처리 전체를 하드웨어로 처리하며 하나의 칩으로 집적하여 개발하고 있다. 따라서, 로봇에 장착된 상태에서 로봇의 주행과 장애물 회피에 직접 사용될 수 있도록 스테레오 정합을 통해 피사체와의 거리를 실시간으로 제공하게 된다.

또한 로봇 제어용 임베디드 프로세서와 함께 연동되어 로봇 클라이언트로부터의 다양한 영상 정보를 무선네트워크를 이용하여 서버로 전송하는 동시에 서버의 처리 결과를 수신할 수도 있다.



(그림 11) ETRI 스테레오 비전 시스템 구성도

## 1. 전처리부

전처리부는 rectification과 좌우 영상의 독립적인 밝기 조절 등의 기능을 통해 카메라의 왜곡을 최소화 함으로써 스테레오 정합의 성능을 향상시킬 수 있다.

스테레오 영상의 geometry에서 한 쪽 영상의 한 점은 다른 영상에서 라인으로 대응되는데, 이 대응될 수 있는 라인을 epipolar line이라 한다. 이러한 epipolar 제한 조건을 이용하여 찾고자 하는 대응 점을 영상 전체에서 찾는 것이 아니라 해당 epipolar line으로 범위를 한정시킬 수 있다[10]. ETRI에서 개발중인 스테레오 비전 처리 기술도 이러한 epipolar 제한조건을 전제로 하여 알고리즘을 구성하였기 때문에 스테레오 정합부에 입력되기 전에 좌우 영상의 epipolar line을 정확하게 맞추어 입력시켜 주는 것이 성능에 영향을 많이 주게 된다.

이러한 epipolar line을 맞추기 위해서 스테레오 카메라 전방에 좌우 카메라의 베이스라인 벡터와 평행하도록 스테레오 카메라 앞에 rectification용 패턴을 이용하여 바로 rectification parameter를 추출한다. 이렇게 미리 계산되어진 변수들을 전처리부의 레지스터 파일에 저장하여 두고 초기화 작업시 위의 parameter들을 이용하여 epipolar line을 수평이 되게 함으로써 영상을 rectification 하게 된다 [11]. (그림 12), (그림 13)은 이런 방식을 통해서



(그림 12) 입력된 영상



(그림 13) Rectification 후의 영상

epipolar line이 정렬된 영상을 보여주고 있다.

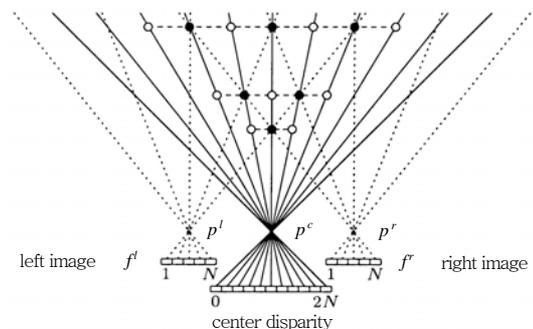
또한, 좌우 카메라의 물리적인 특성 차이 또는 한 쪽 카메라에게만 센 빛이 입력되는 현상 등으로 인해 발생하는 좌우 영상의 밝기 차이를 보정하기 위한 밝기보정(brightness control) 기능이 구현되어 있다.

## 2. 스테레오 정합부

스테레오 정합부는 전처리부에서 처리된 좌우의 영상 입력으로부터 disparity를 계산하여 이를 밝기 정보로서 나타내어 주는 부분이다. 다음은 스테레오 정합부에서 실제 구현을 위한 알고리즘이다.

일반적으로 스테레오 비전에서 3차원 영상을 나타내는 기준 영상을 정할 때 좌, 우 영상 중 하나를 택하는 경우와 이들 사이의 중심에 가상의 영상을 정의하여 이 영상을 사용하는 경우가 있다. Disparity를 이용하여 영상의 depth 정보를 표현할 때, 어느 한 쪽의 이미지만 가지고 정보를 나타내는 경우는 다른 한 쪽의 불연속에 대한 이미지 손실이 많을 수 있는 반면, center referenced disparity를 사용하면 이러한 표현에 있어서의 균형을 이룰 수 있다. 또한 불연속한 공간을 해석하고 솔루션을 찾는 데 있어서 많은 장점을 가지고 있다.

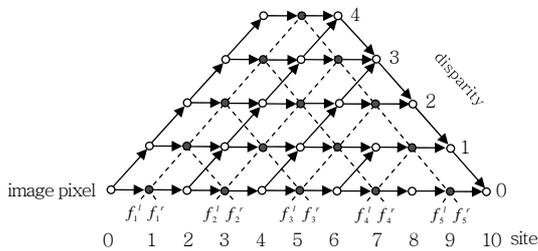
(그림 14)는 스테레오 정합부에서 사용된 알고리즘의 기반이 된 center referenced disparity가 어떻게 형성되는지를 나타내고 있다[5]. Center referenced disparity의 좌표는 좌, 우 두 영상이 생성



(그림 14) 평행 광축 카메라의 중앙 양안차 모델

할 수 있는 영역을 모두 포함하고 있으며, 각각의 영상이 N 라인으로 구성된다면, 중앙 양안차 영상은  $2N+1$ 개의 인덱스를 갖는 것을 확인할 수 있다.

(그림 15)는 평행 광축 시스템의 center referenced disparity 영상에서  $N=5$ 일 때 생성되어지는 center referenced disparity 격자 구조와 이러한 격자 구조에서 dynamic programming을 이용하여 공간 내에서 최적의 경로를 찾는 과정을 보여 준다 [12],[13].



(그림 15) 평행 광축 카메라의 양안차 격자 구조

본 스테레오 정합부에서는 병렬 구조를 가지며 비터비(viterbi) 알고리즘을 근간으로 하여 최적의 경로를 찾아내는데  $O(N)$ 의 시간 복잡도를 가짐으로써 빠른 속도의 실시간 계산이 가능하다.

(그림 16)은 실제 구성된 시스템을 이용하여 얻어진 (a) 좌측 카메라 영상, (b) 우측 카메라 영상, (c) 스테레오 정합 결과 영상이다. 스테레오 정합 결과 영상에서는 가까운 물체는 밝게 하고, 멀어질수



(그림 16) 스테레오 정합 결과 영상

록 어둡게 표시하여 대상체의 거리정보를 표시하게 된다.

### 3. 후처리부

후처리부는 스테레오 정합부에서 얻어진 disparity map에서 노이즈를 제거하고, 밝기 정보에 따라 대상체를 구분하고 이를 각각 출력할 수 있도록 설계되어 있다. 연구자료에서 많이 활용하는 tsukuba 영상을 이용하여 후처리부의 실제 처리 과정을 (그림 17)~(그림 19)에 단계별로 나타내었다.

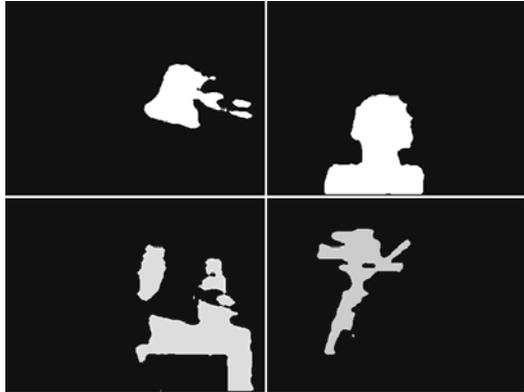
(그림 17)은 tsukuba 영상의 스테레오 정합부의 결과인 disparity map이다. 이 map의 disparity 값은 영상에서 밝기 정보로 매핑되어 카메라로부터 가까운 물체는 밝게 나타나고, 멀어질수록 어둡게 표시된다. 그러나, disparity map은 구현상의 다양한 이유로 인해 실제 상황에서는 (그림 17)에서와 같은 노이즈 성분을 가지게 되며, 이것은 map을 활용한 응용의 성능에 영향을 미치게 된다. 따라서, 후처리부에서는 refinement 과정을 통해 노이즈를 제거하여 (그림 18)과 같이 정제된 disparity map을 출력



(그림 17) Raw Disparity Map



(그림 18) 정제된 Disparity Map



(그림 19) 각 객체의 거리 및 영역 정보 출력

한다.

그리고, 마지막으로 정제된 disparity map을 이용하여 (그림 19)에서 보듯이 밝기 값이 비슷한 점들을 그룹화하여 4개의 객체로 구분하며, 이 객체들을 독립적으로 출력하는 기능과 객체들의 거리, 폭, 높이 정보를 외부에서 읽기 가능한 register로 제공하는 기능을 수행한다.

이와 같이, 후처리부는 스테레오 정합의 결과를 가공하여 대상체를 구분하여 출력하고 동시에 객체 정보를 제공해 줌으로써 함께 연동되는 임베디드 프로세서나 서버에서 수행될 응용프로그램의 복잡한 연산 부하를 덜어줄 수 있도록 설계되었다.

## IV. 결론

지능형 로봇 개발에는 여러 종류의 인지 기술이 필요하며 그 중에서도 시각 처리 관련 기술은 변화하는 주위 환경에 융통성 있게 대처할 수 있는 정보를 제공해주는 시각의 중추적인 역할을 담당한다.

본 고에서는 지능형 로봇에 적용될 수 있는 시각 처리 분야 중 스테레오 비전 처리에 대한 국내외 연구 개발 동향에 대해서 살펴보았고 ETRI에서 개발 중인 URC 로봇의 스테레오 비전 처리 기술개발에 대해 소개하였다.

ETRI에서 개발 중인 스테레오 비전 처리 시스템은 가정용 서비스 로봇에 장착을 목표로 하고 있으

며 저가의 CMOS 이미지 센서를 사용하고 스테레오 정합을 포함한 영상처리 전체를 하드웨어로 처리하는 하나의 칩으로 집적하여 개발하고 있다.

## 약 어 정 리

CCD	Charge Coupled Device
CMOS	Complementary Metal-Oxide Semiconductor
DP	Dynamic Programming
DSP	Digital Signal Processor
FPGA	Field-Programmable Gate Array
SAD	Sum of Absolute Differences
STOC	Stereo-on-a-Chip
URC	Ubiquitous Robotics Companion
VGA	Video Graphics Array
VME	VERSA Module Eurocard

## 참 고 문 헌

- [1] 조동일, "지능형 로봇 센서," 주간기술동향, 통권 1183호, 2005. 2., pp.31-41.
- [2] Takeo Kanade, Atsushi Yoshida, Kazuo Oda, Hiroshi Kano, and Masaya Tanaka, "A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications," *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, 18-20 June 1996, pp.196-220.
- [3] Bernd Porr, Alex Cozzi, and Florentin Wörgötter, "How to "Hear" Visual Disparities: Real-time Stereoscopic Spatial Depth Analysis Using Temporal Resonance," *Biological Cybernetics*, Vol.78, No.5, June 1998, pp.329-336.
- [4] A. Darabiha, J. Rose, and W.J. MacLean, "Video-rate Depth Measurement on Programmable Hardware," *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'03)*, June 2003, pp.203-210.
- [5] Hong Jeong and Yuns Oh, "A Parallel Real Time Implementation of Stereo Matching," *Proc. of 15th Int'l Symp. on Parallel and Distributed Processing*, Apr. 2001, pp.23-27.

- [6] H. Jeong and S.C. Park, "Generalized Trellis Stereo Matching with Systolic Array," *In Lecture Notes in Computer Science*, Vol.3358, 2004, pp.263-267.
- [7] John I. Woodfill, Gaile Gordon, Dave Jurasek, Terrance Brown, and Ron Buck, "The Tyzx DeepSea G2 Vision System, A Taskable, Embedded Stereo Camera," *Proc. of IEEE Computer Society Workshop on Embedded Computer Vision, Conf. on Computer Vision and Pattern Recognition*, June 2006.
- [8] [http://www.ptgrey.com/products/bumblebee2/bumblebee2\\_xb3\\_datasheet.pdf](http://www.ptgrey.com/products/bumblebee2/bumblebee2_xb3_datasheet.pdf), "Bumblebee Stereo Vision Camera Systems," Point Grey Research, Jan. 2007.
- [9] <http://www.videredesign.com/docs/stoc-1.2.pdf>, Stereo-on-a-Chip [STOC] Stereo Head User Manual 1.2, Videre Design, Aug. 2006.
- [10] Stuart Geman and Donald Geman, "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Trans. On PAMI*, PAMI, Vol.6, No.6, 1984, pp.721-741.
- [11] Charles Loop and Zhengyou Zhan, "Computing Rectifying Homographies for Stereo Vision," *Proc. IEEE Computer Vision and Pattern Recognition*, Vol.1, June 23-25, 1999, pp.125-131.
- [12] H.H. Baker and T.O. Binford, "Depth from Edge and Intensity Based Stereo," *Proc. of the Int'l Joint Conf. on Artificial Intelligence*, 1981, pp.631-636.
- [13] Y. Ohta and T. Kanade, "Stereo by Intra- and Inter-scanline Search," *IEEE Trans. PAMI*, PAMI, Vol.7, No.2, 1985, pp.139-154.