

논문 2013-50-1-10

# 다중 베이스라인 기반 질감 적응적 신뢰도 전파 스테레오 정합 기법

(Multibaseline based Stereo Matching Using Texture adaptive Belief  
Propagation Technique)

김진형\*, 고윤호\*\*

(JinHyung Kim and Yun Ho Ko)

## 요약

스테레오 비전을 사용하여 거리정보를 획득하기 위해서는 스테레오 영상 쌍에서 대응점을 찾는 스테레오 매칭이 이루어져야 한다. 기존의 스테레오 비전은 주로 두 대의 카메라를 사용해서 양안시차를 획득하였다. 따라서 기존의 방법들은 베이스라인의 길이에 따른 정확도와 정밀도 사이의 이율배반적 관계를 해소하지 못하였다. 또한 근래에 사용되는 신뢰도 전파 기법의 경우 고정된 가중치  $\lambda$ 에 의하여 스테레오 정합 성능이 크게 좌우되는 문제점이 있었다. 본 논문에서는 이율배반적 문제를 해소하기 위해 다중 베이스라인 스테레오 비전에 기반을 둔 개선된 신뢰도 전파 스테레오 정합 기법을 제안한다. 제안된 기법은 국부-명백성 함수로 EMAD(Extended mean of absolute difference)를 계산한다. 그리고 제안된 기법은 가중치  $\lambda$ 를 지역적 질감 정보에 적응적으로 결정한다. 제안된 기법은 기존 기법보다 더 높은 정합 성능을 보여주며 적은 반복을 통해 최적해에 도달하였다. 정합 성능이 PSNR수치상 약 4.85dB만큼 증가하였다.

## Abstract

To acquire depth information using stereo vision, it is required to find correspondence points between stereo image pair. Conventional stereo vision systems usually use two cameras to get disparity data. Therefore, conventional stereo matching methods cannot resolve the tradeoff problem between accuracy and precision with respect to the length of baseline. Besides, belief propagation method, which is being used recently, has a problem that matching performance is dependent on the fixed weight parameter  $\lambda$ . In this paper, we propose a modified belief propagation stereo matching technique based on multi-baseline stereo vision to solve the tradeoff problem. The proposed method calculates EMAD(extended mean of absolute differences) as local evidence. And proposed method decides weight parameter  $\lambda$  adaptively to local texture information. The proposed method shows higher initial matching performance than conventional methods and reached optimum solution in less iteration. The matching performance is increased about 4.85 dB in PSNR.

**Keywords:** 스테레오 정합, 다중 베이스라인 스테레오, 신뢰도 전파, 적응적  $\lambda$  추정

\* 학생회원, \*\* 평생회원, 충남대학교 메카트로닉스공학과

(Department of Mechatronics Engineering,  
Chungnam National University)

※ 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2011-0013923)

접수일자: 2012년6월3일, 수정완료일: 2013년1월3일

## I. 서론

최근 로봇의 패러다임이 노동 대체 수단으로서의 전통적 산업 로봇에서 인간 친화적인 지능형 로봇으로 변화하고 있다. 이러한 지능형 로봇은 주변 환경을 스스로 인지하고, 이를 바탕으로 스스로 판단하여 사용자에

게 유용한 고수준의 서비스를 수행한다<sup>[1]</sup>. 지능형 로봇들이 적절한 서비스를 제공하기 위해서는 근방의 구조물에 대한 위치를 정확하게 파악하여 이에 근거한 경로 계획을 통해 이동을 해야 하는 과정이 필수적으로 요구된다. 하지만 주변 환경을 정밀하게 인지하지 못하는 문제로 인하여 서비스를 위한 로봇의 이동이 완벽하지 못한 실정이다.

지능형 로봇이 주변 환경을 인지하기 위해서 주로 사용하는 방법은 카메라, 초음파, 적외선, 레이저 방법 등이 이용되고 있다. 이 중에서 카메라를 이용한 방법은 레이저 센서에 비해 가격이 저렴하고, 적외선과 초음파 센서에 비하여 높은 정확도를 가진다. 그리고 소프트웨어를 통한 기능 추가가 용이하며, 영상을 통해 다양한 부가 정보를 얻을 수 있다<sup>[2,3]</sup>. 또한 장치의 크기가 레이저에 비해 작기 때문에 주행기능이 요구되는 지능형 로봇에 적용하기가 용이하며, 한 번에 면 단위의 정보를 획득하기 때문에 로봇이 이동하지 않아도 영상정보만을 처리해서 주변 환경을 인지할 수 있다. 카메라를 이용한 3차원 정보의 획득은 스테레오, 움직임, 명암, 질감 등을 이용하는 기법들이 있다. 이 중에서 동일 장면에 대한 다시점 영상을 이용하는 스테레오 비전 기법이란 인간의 시각 기관을 기계적 시스템으로 구현하려 한 것으로 물체의 형상을 두 대의 카메라로 촬영하여 2차원 평면에 대응시켜서 물체에 대한 제반 정보를 3차원으로 해석하는 것이다. 스테레오 비전에서 거리정보를 획득하기 위해서는 스테레오 정합이 필수적이다. 스테레오 정합기법에는 지역적(local) 기법과 전역적(global) 기법이 있는데, 최근에는 지역적 최소점 문제에 빠질 우려가 없고 병렬화가 쉬우며 높은 정합 성능을 보여주는 신뢰도 전파(belief propagation) 기법이 주로 이용된다.

디지털 영상을 이용한 스테레오 비전에서 카메라의 광축 사이 거리를 베이스라인(baseline)이라 한다. 이 길이는 거리 정보를 획득하는데 있어서 매우 중요한 변수로, 스테레오 정합에 있어서 정확도와 정밀도를 결정한다. 일반적으로 베이스라인의 길이와 정밀도는 비례적인 관계를 가진다. 하지만 정확도는 베이스라인의 길이에 반비례의 관계를 가진다. 이는 단일 베이스라인 기반 스테레오 정합기법으로는 정확도와 정밀도를 동시에 만족할 수 없음을 의미한다. 스테레오 정합은 밝기 정보를 기반으로 하기 때문에 질감정보가 적은 영역에서 정확한 정합이 이루어지지 않을 가능성이 높다. 특

히, 근래에 많이 사용되는 신뢰도 전파 알고리즘을 사용하는 정합 기법의 경우 지역적 정합 비용과 주변 관계에 따른 비용의 비중을 결정하는 가중치  $\lambda$ 에 따라 정합성능이 크게 좌우된다. 따라서 영상에 따라 지역적으로  $\lambda$ 를 추정해야할 필요성이 대두된다.

이러한 배경으로부터 본 논문에서는 신뢰도 전파 알고리즘에 기반을 둔 다중 베이스라인 스테레오 정합 기법의 성능을 높이기 위한 두 가지 기법을 제안한다. 첫째로 모든 베이스라인에서의 후보 깊이의 표본의 합집합을 사용하여 국부-명백성 함수를 계산하는 다중 베이스라인 기반 EMAD기법을 제안한다. 그리고 영상 내 질감 정보에 적응적으로  $\lambda$ 를 추정하는 기법을 제안한다. 영상 내 질감 정보에 적응적으로  $\lambda$ 를 추정하는 기법은 다중 베이스라인을 사용하지 않는 다양한 신뢰도 전파 알고리즘에 기반을 둔 스테레오 정합 기법에도 적용 가능하다.

본 논문의 구성은 다음과 같다. II장에서는 베이스라인 길이에 따른 정확도와 정밀도의 이율배반성에 대하여 고찰하고, 기존의 스테레오 정합 알고리즘인 다중 베이스라인 기반 SSSD기법과 신뢰도 전파 기법에 대하여 설명한다. III장에서는 제안하는 알고리즘의 상관도의 측정방법, 후보깊이의 결정과정, 질감도에 적응적인 지역정보의 사용에 대하여 설명한다. IV장에서는 Middlebury에서 제공한 실험영상을 이용하여 기존 기법과 제안하는 기법의 스테레오 정합 결과를 비교 및 분석하여 평가한다. 마지막으로 V장에서는 결론을 맺고 향후 연구 진행방향에 대하여 논의한다.

## II. 기존 스테레오 정합 알고리즘

그림 1과 같이 광축이 평행한 동일한 렌즈를 갖는 양안 스테레오 비전에서 등극선 기하 조건(epipolar constraint)을 만족하는 경우 영상에 맺힌 임의의 상의 거리  $z$ 는 식(1)과 같이 구한다.

$$z = \frac{bf}{(x_l - x_r)} \quad (1)$$

여기서  $b$ 는 두 카메라의 광축간 거리이고, 베이스라인이(baseline)라 부르며,  $f$ 는 카메라 렌즈의 초점거리이다. 그리고  $x_l$ 과  $x_r$ 은 각각 임의의 상이 영상평면에 맺힌 위치이며, 이 둘의 차이를 시차변이(disparity)로

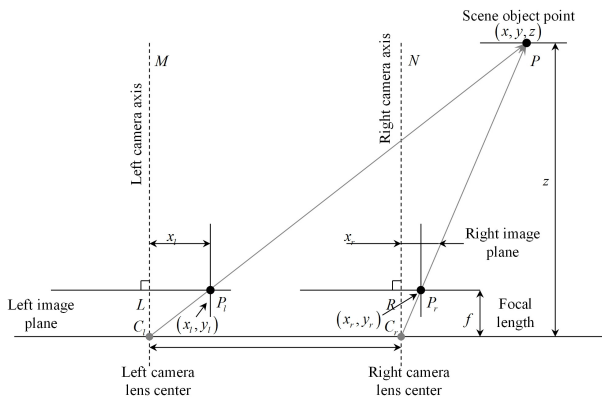


그림 1. 평행 광축 양안 스테레오 비전 모델  
Fig. 1. Parallel optical axis stereo vision model.

정의하고 있다<sup>[4~5]</sup>.

디지털 스테레오 영상에서 시차변이는 일반적으로 정수형이다. 따라서 시차변이는 베이스라인의 거리가 증가할수록 측정거리에 대한 해상도는 높아져서 추정된 거리 값에 대한 정밀도가 높아진다. 하지만 그만큼 탐색범위가 넓어지기 때문에 지역 최소점 문제에 빠질 가능성이 높아지며, 동시에 연산이 많아지는 단점이 있다.

### 1. 다중 베이스라인 기반 SSSD 기법

기존의 연구들은 주로 두 대의 카메라를 사용하여 작업에 맞는 작업거리를 설정하여 해당 범위에서 사용자의 요구에 따라 적절한 베이스라인의 길이를 설정하여 사용해왔다. 단일 베이스라인에 대한 스테레오 정합의 경우 반복되는 무늬 혹은 모호한 영역에서 오정합이 일어난다. 이러한 문제를 해결하기 위해 Kanade는 다중 베이스라인 기반 SSSD(sum of sums of squared difference)에 따른 스테레오 정합 기법을 제안하였다<sup>[6]</sup>. SSSD는 최장 베이스라인을 기준으로 후보 시차변이를 구하고, 해당 후보 시차변이들에 대한 길이가 다른 각 베이스라인에서의 SSD(sum of squared difference)의 합을 비용함수로 사용하여 스테레오 정합 알고리즘의 모호성을 낮춰서 오정합률을 낮추었다. SSSD기법은 기존의 정합 알고리즘과 다르게 추정 대상으로 시차변이가 아닌 깊이 정보의 역수인  $\zeta$ 를 기준으로 상관도를 계산하였다. 임의의  $i$ 번째 베이스라인 기준의 시차변이  $d_{(i)}$ 는 다음 식 (2)와 같이 정리된다.

$$d_{(i)} = b_i f \zeta \quad (2)$$

여기서  $i$ 는 베이스라인의 순번을 의미한다. 각 후보 시차변이에 대한  $i$ 번째 베이스라인 기준 스테레오 영상 쌍에서의 SSD는 아래 식 (3)과 같이 구할 수 있다.

$$e_{\zeta(i)}(x, y, \zeta) = \sum_{y=-(M/2)}^{M/2} \sum_{x=-(N/2)}^{N/2} (I_l(x, y) - I_r(i)(x - b_i f \zeta, y))^2 \quad (3)$$

위 식에서  $M, N$ 은 각각 각 축방향의 마스크 크기이다.

SSSD는 각 베이스라인에서의  $\zeta$ 에 대한 SSD를 누적하여 계산된다.  $n+1$ 개의 카메라로 영상을 입력받을 때 식 (4)를 통해 기준 영상의 임의의 한 위치에서의 SSSD를 계산할 수 있다.

$$SSSD(x, y, \zeta) = \sum_{i=1}^n e_{\zeta(i)}(x, y, \zeta) \quad (4)$$

SSSD가 구해지면, 해당 위치에서 SSSD가 최소가 되는  $\zeta$ 의 역수를 거리 정보로 추정한다. 계산하기 위해 다중 베이스라인에 대한 추가적인 영상 정보를 사용하기 때문에 잡음과 반복에 의한 모호성이 감소하는 장점이 있다.

### 2. 신뢰도 전파를 사용한 스테레오 정합

최근 스테레오 정합 알고리즘은 다양한 사전확률정보를 적용할 수 있는 장점을 가진 베이시안(bayesian) 추론 기법을 통해 전역 최적의 해를 찾는 방법이 많이 거론되고 있다<sup>[7~8]</sup>.

#### 가. 스테레오 정합 문제의 확률 모델

주어진 영상 정보를  $Y$ , 구하고자 하는 시차변이의 정보를  $X$ 라 하면, 사후결합확률은 베イズ 정리에 의해 식 (5)와 같이 표현된다. 여기서  $P(Y)$ 는 영상의 밝기 값 정보로 항상 양의 확률을 갖기 때문에, 사후확률은 우도함수(likelihood)와 사전확률(prior)의 곱에 비례하게 된다.

$$P(X|Y) \propto P(Y|X)P(X) \quad (5)$$

여기서 영상 정보  $Y$ 에 대한 센서 잡음을 독립적으로 동일한 분포(I.I.D. : independent identical distribution)을 따르는 가우시안으로 가정하면 우도함수와 사전확률은 다음 식 (6)과 같이 표현될 수 있다.

$$P(Y|X)P(X) = \prod_s \exp(-F(s, d_s)) \prod_{s} \prod_{t \in N(s)} \exp(-V(d_s, d_t)) \quad (6)$$

여기서  $s$ 는 화소의 위치를 의미한다.  $F(s, d_s)$ 는 위치  $s$ 에서 시차변이  $d_s$ 를 가질 때의 정합 비용이다. 일반적으로 정합 비용은 AD(absolute difference)가 많이 사용된다. 사전확률은 인접한 위치에서의 시차변이정보들의 사전제약조건들이며 일반적으로 인접한 화소의 시차변이는 급격하게 변하지 않는다는 조건을 사용한다. 여기서  $N(s)$ 는 노드  $s$ 에 인접한 노드의 집합이다. 사전확률 분포에 사용되는 제약조건 함수  $V$ 는 Potts모델, 선형 모델, 강인(robust) 모델, 절단된 선형(truncated linear) 모델 등 다양한 모델이 존재한다. 이 분포 모델들은 기본적으로 근접한 화소 사이의 시차변이 사이의 차에 대한 비용으로, 시차변이의 차이가 커질수록 결합 확률은 낮아지는 경향을 갖는다.

스테레오 정합 문제를 Pair-wise MRF(Markov Random Field)로 모델링하면 식 (7)과 같이 나타낼 수 있다. 이때 각 노드들의 집합인  $\{x_s\}$ 는 은닉 노드이고,  $\{y_s\}$ 는 관측 노드를 나타낸다. 스테레오 정합에 있어서 추정하고자 하는 은닉 노드는 시차변이를 의미하며 앞서 설명한 사전확률분포를 따른다.

$$P(X|Y) \propto \prod_s \psi_s(x_s, y_s) \prod_s \prod_{t \in N(s)} \psi_{st}(x_s, x_t) \quad (7)$$

여기서  $\psi_s$ 는 국부-명백성 함수(local evidence)라 하고 식 (6)에서 우도함수와 관련되며,  $\psi_{st}$ 는 국부-적합성 함수(compatibility matrix)라 하며 식 (6)의 사전확률과 관련된다. 따라서 식(6)에서의 최대사후확률을 찾는 것은 결국 MRF에서 최대사후확률을 찾는 것과 같게 된다<sup>[8]</sup>.

#### 나. 신뢰도 전파(belief propagation)

앞서 제시한 확률모델에서 전역적으로 결합 확률이 최대가 되도록 하는 해를 찾는 일은 매우 복잡하고, 경우의 수가 무한대에 가깝다. 따라서 정확한 해를 찾는 것은 불가능하다. 근래에는 이러한 문제의 최적의 해를 근사적으로 구할 수 있는 방법들이 많이 제안되고 있다. 그 중에서 널리 사용되는 알고리즘이 Pearl에 의해 제안된 신뢰도 전파 알고리즘이다<sup>[9]</sup>. 신뢰도 전파 알고리즘은 비방향성 그래프에 적용하는 국소적 메시지 전

파 알고리즘으로, 베이지안 추론을 효과적으로 풀기 위해 제안되었다. 그리고 Sun은 신뢰도 전파 알고리즘을 사용해서 스테레오 정합 문제의 해를 구하는 방법을 제시하였다<sup>[8]</sup>. 신뢰도 전파 알고리즘에서 메시지는 각 노드가 가지는 모든 상태에 대한 확률을 포함하는 벡터이다. Pair-wise MRF 모델에서 메시지는 관측 노드에서 은닉노드로 보내는 메시지와 인접한 은닉노드사이에서 전달하는 메시지로 구분된다. 전자의 경우 앞서 설명한 우도함수와 관계가 있으며, 후자의 경우 인접한 위치에서의 시차변이는 부드럽게 이어진다는 사전 제약조건과 관계가 있다. 각 노드에서의 주변 확률을 높여서 전체 노드 사이의 결합 확률을 높이는 방법으로 메시지를 갱신하는 max-product 방법은 식 (8)을 사용해서 메시지를 갱신한다.

$$m_{st}^{i+1}(x_t) \leftarrow \kappa \max_{x_s} \left[ \psi_{st}(x_s, x_t) m_s(x_s) \prod_{x_k \in N(x_s)|x_t} m_{ks}^i(x_s) \right] \quad (8)$$

여기서  $N(x_s)|x_t$ 는  $x_t$ 를 제외한  $x_s$ 의 이웃 노드이며  $\kappa$ 는 정규화 상수이다.  $i+1$ 번째 반복일 때  $s$ 에서  $t$ 로 보내는 메시지는 두 위치에서의 시차변이의 차이에 대한 메시지  $\psi_{st}$ 와  $s$ 위치 자신이 가지는 메시지  $m_s$ , 그리고  $s$ 위치에서 보내는 방향을 제외한 나머지 이웃들의 메시지의 곱으로 구성된다.  $m_s$ 는 식(7)에서  $\psi_s$ 와 같다. 일반적으로  $\psi_{st}$ 와  $m_s$ 사이의 불균형을 조절하기 위해  $m_s$ 의 내부의 국부-명백성 함수에 가중치  $\lambda$ 를 곱해준다. 가중치  $\lambda$ 는 영상에 따라 실험적으로 구하는 것이 일반적이다. 이 메시지는  $t$ 위치에서 시차변이  $x_t$ 를 가질 때  $s$ 가  $x_s$ 라는 시차변이를 가질 확률을 내포하고 있다. 그림 2는 위에서 설명한 메시지 갱신을 도시한

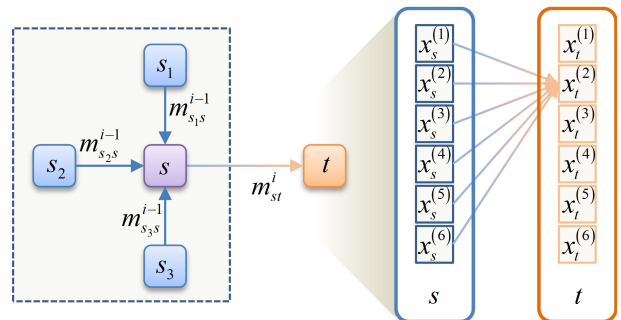


그림 2. Max-product 방법에서의 메시지 갱신방법  
Fig. 2. Message update rule of max-product method.

예이다.

Max-product방법을 통해 메시지 갱신을 반복적으로 수행한 후에 각 위치에서의 시차변이를 구하기 위해 해당 위치의 신뢰도를 식 (9)를 사용해서 구한다. 신뢰도는 임의의 위치  $s$ 에서 시차변이  $x_s$ 를 가질 주변 확률로 해당 노드의 값과 인접한 모든 위치에서 보내는 메시지의 곱으로 구해진다.

$$b_s(x_s) \leftarrow \kappa m_s(x_s) \prod_{x_k \in N(x_s)} m_{ks}(x_s) \quad (9)$$

메시지 전달의 충분한 반복을 통해 임의의 위치  $s$ 에서 신뢰도가 수렴하면, 각 후보시차변이 중에서 가장 높은 신뢰도를 가지는  $x_s$ 를 식 (10)과 같이 시차변이  $d_s^{MAP}$ 로 결정한다.

$$d_s^{MAP} = \operatorname{argmax}_{x_s} b_s(x_s) \quad (10)$$

앞서 살펴본 신뢰도 전파 알고리즘은 지수함수의 곱으로 이루어져 있기 때문에 계산 및 구현이 용이하지 않다. 따라서 실제 구현에 있어서는 음의 자연로그를 취하여 식을 간편화한다. 식 (11)은 변환된 메시지 갱신 수식이며, 식 (12)는 변환된 신뢰도 계산식이다. 변환된 수식은 결과적으로 결합 확률을 최대화 하는 문제로부터 전역 에너지를 최소화 하는 문제로 전환되며, 식 (13)을 통해 임의의 위치  $s$ 에서 시차변이를 구한다.

$$M_{st}^{i+1}(x_t) \leftarrow \min_{x_s} \left[ \phi_c(x_s, x_t) + M_s(x_s) + \sum_{x_k \in N(x_s)|x_t} M_{ks}^i(x_s) \right] \quad (11)$$

$$B_s(x_s) \leftarrow c M_s(x_s) + \sum_{x_k \in N(x_s)|x_t} M_{ks}(x_s) \quad (12)$$

$$D_s^{MAP} = \operatorname{argmin}_{x_k} B_s(x_k) \quad (13)$$

#### 다. 기존 기법의 한계

SSSD 기법은 후보 추정  $\zeta$ 에 대한 정합 상관도를 계산하기 위해 다중 베이스라인에 대한 추가적인 영상 정보를 사용하기 때문에 부분적 시점에서 발생하는 잡음과 반복되는 패턴에 의한 모호성이 감소하는 장점이 있어서 베이스라인이 길어져서 발생하는 오정합률을 낮출

수 있었다. 그러나 SSSD 기법의 경우 단 하나의 가장 긴 베이스라인에 대한 후보 시차변이를 기준으로 다중 베이스라인에서의  $\zeta$ 에 대한 SSSD를 계산하기 때문에, 실질적으로 기존 단일 베이스라인 기반 정합 기법에서 베이스라인의 길이만 늘려서 정합을 수행하는 방식과 동일한 깊이 정보 해상력을 가지는 단점이 존재한다.

전역적 방법인 신뢰도 전파 알고리즘의 경우 스테레오 정합 문제를 MRF기반 확률 모델로 가정하여 근접한 화소 사이에서 반복적으로 메시지를 주고받으면서 오정합된 정보를 버리고 올바른 시차변이를 선택할 확률을 높여 스테레오 정합의 정확도와 품질을 크게 높였다. 하지만 국부-명백성 함수와 국부-적합성 함수 사이의 비중을 결정하는  $\lambda$ 에 따라 그 성능이 크게 변하는 단점이 있다. 따라서 보다 고해상력을 가지는 동시에 반복횟수를 줄일 수 있는 개선된 정합 방법이 요구된다.

### III. 제안 알고리즘

베이스라인 길이에 따른 정밀도와 정확도의 이율배반적 관계에 따른 문제점을 개선하고 더 정확한 거리 정보를 획득하기 위해 본 논문에서는 모든 베이스라인에서의 후보 시차변이에 대한 정합을 수행하는 신뢰도 전파 기법을 적용한 다중 베이스라인 기반 스테레오 정합 알고리즘을 제안한다.

그림 3은 제안하는 기법의 개략적인 흐름을 보여준다. 제안하는 기법은 입력된 스테레오 영상으로부터 후보 깊이를 정의하고, 각 후보 깊이에 대한 지역적 정합 비용을 제안된 EMAD를 사용하여 계산하며 지역적 질

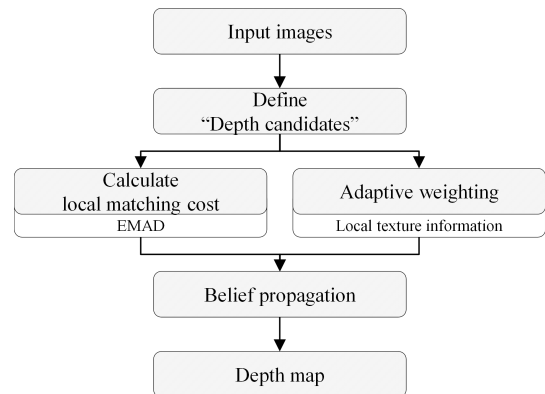


그림 3. 제안하는 기법의 흐름도  
Fig. 3. Flow chart of proposed method.

감 정보에 적응적으로 메시지의 가중치를 결정하여 신뢰도 전파 기법을 적용하여 깊이 정보를 출력한다.

가. 지역적 정합 비용의 계산

기존의 다중 베이스라인 기반 SSSD 정합 기법은 최장 베이스라인에서의 시차변이만을 후보 시차변이로 사용하기 때문에 깊이 정보의 정확도는 높아지지만, 정밀도는 개선되지 않았다. 이러한 문제를 해결하기 위해 본 논문에서는 추정 후보 깊이로 모든 베이스라인에 대한 영상 쌍에서의 후보 시차변이를 사용하는 EMAD (extended mean of absolute difference)를 제안한다. EMAD는 다음 식 (14)와 같이 계산된다.

$$EMAD(x,y,z,i) = \frac{1}{n} \sum_{i=1}^n \left| I_0(x,y) - I_i \left( x - \frac{b_i f}{z}, y \right) \right| \tag{14}$$

여기서  $n$ 은 베이스라인의 총 개수이며,  $I_0$ 는 기준영상,  $I_i$ 는  $i$ 번째 베이스라인에 해당되는 비교영상이다. EMAD를 사용한 국부-명백성 함수는 식 (15)와 같이 계산된다.

$$m_s(x_s) = \exp(-\lambda EMAD(x,y,z,i)) \tag{15}$$

위 식에서  $x$ 와  $y$ 는 임의의 위치에서의 노드  $s$ 의 영상에서의 좌표이고,  $\lambda$ 는 지역정보의 사용에 대한 가중치이다.

EMAD는 후보깊이  $z$ 에 대한 모든 베이스라인에서의 정합 비용의 평균을 의미한다. 즉, 하나의 후보깊이  $z$ 에 대하여 모든 베이스라인 기반 스테레오 영상 쌍에서의 시차변이를 계산하고, 해당 시차변이에 대한 대응점 사이의 밝기값 차이의 절댓값에 대한 평균을 사용하여 계산한다. 따라서 카메라 센서에 의한 잡음과 카메라 위치에 따라 달라지는 대상체의 밝기값 변화에 강인하다.

나. 추정 후보깊이  $z$ 의 선정

앞서 설명한 EMAD를 신뢰도 전파 알고리즘의 지역적 정보로 사용하기 위해서는 후보깊이  $z$ 가 결정되어야 한다. 연속적인 값을 신뢰도 전파 알고리즘의 추정 후보로 사용할 수 없으므로 본 논문에서는 모든 베이스라인 기반 스테레오 영상 쌍에서의 각 시차변이에 대응되는 거리정보를 후보깊이  $z$ 로 사용한다. 이때 각 베



그림 4. 후보깊이 선정의 예  
Fig. 4. An example of depth candidates decision.

이스라인 기반 스테레오 영상 쌍에서의 시차변이들은 거리에 따른 정렬이 되어 있지 않기 때문에 대응되는 거리값에 따라서 오름차순 정리를 수행해서 전체 베이스라인에 대한 가능한 모든 시차변이를 추정 후보에 포함시키도록 한다.

그림 4는 추정 후보깊이 선정의 한 예로 세 가지 베이스라인에 대한 각각의 후보깊이와 제안하는 방법의 후보깊이를 보여준다. 그림 4(a)에는 각 베이스라인에서 후보 양안시차가 0~6의 값을 가질 때의 해당 후보양안시차에 대응되는 후보깊이가 표기되어 있다. 기존의 SSSD기법을 적용하여 스테레오 정합을 수행하는 경우,  $3b$ 의 길이를 가지는 베이스라인이 최장 베이스라인이 되므로, 이보다 더 짧은 베이스라인에 대한 후보양안시차는 제외되어 깊이 정보의 해상도는 높아지지 않는다. 그림 4(b)는 제안하는 방법에 있어서의 후보깊이의 선정을 표현한 것이다. 제안하는 방법은 모든 베이스라인의 후보양안시차에 대응되는 깊이값을 후보깊이로 정하기 때문에, 깊이 정보의 해상도가 베이스라인의 개수에 따라 증가하게 된다.

추정하고자 하는 값이 시차변이가 아니라 시차변이에 해당되는 거리이기 때문에, 임의의 위치  $s$ 에서 은닉 노드의 값  $x_s$ 는 해당 위치에서의 거리정보  $z_s$ 가 되며, 거리정보는 시차정보의 변화량과 반비례 관계가 있기 때문에 기존의 방법에서 사용하는 인접한 후보시차의 차이의 절댓값과 배율이 크게 달라진다. 본 논문에서는 커지는 배율을 보완하기 위해 인접한 거리정보의 차이의 절댓값을 후보거리 구간의 크기로 정규화하여 국부-적합성 함수를 식 (16)과 같이 모델링하였다.

$$\psi_{st}(x_s, x_t) = \exp\left(-\frac{|z_s - z_t|}{z_{\max} - z_{\min}}\right) \quad (16)$$

위에서  $z_{\max}$  와  $z_{\min}$  은 각각 최대 후보깊이와 최소 후보깊이를 의미하며, 후보거리 구간의 크기를 구하기 위해 사용되었다.

#### 다. 적응적 지역 정보의 사용

일반적으로 신뢰도 전파 알고리즘은 국부-명백성 함수를 메시지에 포함함에 있어서 국부-적합성 함수와의 비중을 상수  $\lambda$ 를 곱해주는 것으로 조절한다. 스테레오 정합에 있어서는 이 상수에 따라 정합 성능이 크게 좌우된다.  $\lambda$ 값은 장면에 매우 의존적이므로 입력되는 영상에 따라 실험적으로 구해야 하는 특성이 있다. 그리고 영상에 따라 실험적으로 구해진  $\lambda$ 라 할지라도 영상 내 영역에 따라 국부-명백성 정보가 더 중요한 곳과 국부-적합성 정보가 더 중요한 곳이 각각 다를 수 있다. 모호한 영역이 넓고 많을수록, 최적의 해를 구하기 위한 신뢰도 전파의 반복횟수는 늘어난다. 최악의 경우 넓은 영역에 걸쳐 오정합이 일어날 경우 신뢰도 전파를 통해 잘못된 정보가 전파될 가능성도 있다. 표 1은 Middlebury<sup>[10]</sup>에서 제공한 스테레오 영상 Rocks와 Woods 영상에 대한 가중치 변화에 따른 PSNR의 추이를 보여준다. 그림 5는 실험에 사용된 Rocks와 Woods 영상이다. 상대적으로 거친 부분이 많은 Rocks 영상의 경우 낮은 가중치에서 가장 높은 정합 성능이 나오는

표 1. 실험영상의  $\lambda$ 변화에 따른 PSNR<sup>[11]</sup>  
Table 1. PSNRs of experimental images for varying  $\lambda$ .

$\lambda$	0.2	0.02	0.002	0.0002
Rocks	16.05	19.99	20.64	20.93
Woods	14.84	18.64	18.33	18.15

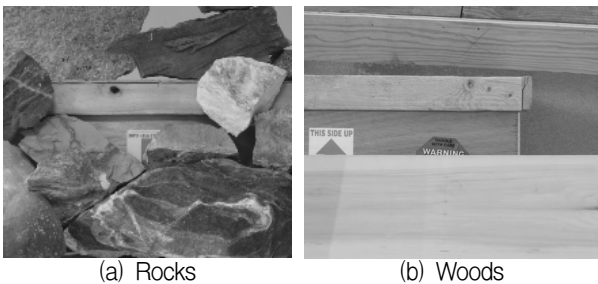


그림 5. Rocks 영상과 Woods 영상  
Fig. 5. Rocks and Woods images.

것을 알 수 있고, 전역적으로 평탄한 영역이 많이 포함된 Woods영상의 경우 비교적 높은 가중치를 주었을 때 가장 높은 정합 성능을 보여준다. 질감도가 높은 영상은 임의의 한 위치에서 국부-명백성 함수에 의한 깊이 정보의 추정이 보다 정확하게 이루어지기 때문에 신뢰도를 전파함으로써 인접한 영역의 정합정도가 더 정확해지고 확고해진다. 이에 반해 질감도가 낮은 영상의 경우 지역적으로 국부-명백성 함수에 의한 깊이정보의 추정이 정확하게 이루어지지 않기 때문에 신뢰도 전파를 통해 부정확한 추정 깊이 정보가 전파되어 영상 전반에 걸쳐 부정확한 깊이 추정이 이루어진다. 이와 같은 관찰로부터 신뢰도 전파 기법을 통해 보다 높은 정합성능을 이끌어내기 위해서는 질감도에 반비례하는  $\lambda$ 를 사용해야 한다.

본 논문에서는 신뢰도 전파 알고리즘을 사용한 스테레오 정합 성능을 높이기 위해 앞서와 같은 관계를 가지는 가중치  $\lambda$ 를 지역적 질감도를 반영하여 결정하는 방법을 제안한다. 질감도는 마스크 내에서의 표준편차( $\sigma$ )를 사용해서 측정하였다. 앞서의 실험에 따르면 질감도와  $\lambda$ 는 반비례 관계를 갖는다. 이러한 특성을 반영하는 새로운  $\lambda_v$ 는 식(17)과 같이 정의된다.

$$\lambda_v = \frac{\lambda}{\rho\sigma/\sigma_{\max} + 1} \quad (17)$$

여기서  $\sigma_{\max}$ 는  $N \times N$ 크기의 마스크에서 가능한 최대의 표준편차이며 이는 식 (18)과 같이 계산된다.

$$\sigma_{\max} = \left[ \sqrt{f_{med}^2 \times \frac{(N^2 - 1)}{N^2}} \right] \quad (18)$$

위 식에서  $f_{med}$ 는 회색계조(greyscale)로 표현 가능한 밝기값들의 중간값으로 8bit 회색계조(greyscale) 영상에 있어서는 127이 된다.  $\rho$ 는  $\lambda_v$ 의 범위를 결정해주는 배율이다.

#### IV. 실험결과 비교 및 평가

제안한 알고리즘의 성능을 평가하기 위하여 Middlebury에서 제공한 Aloe 영상과 Baby2 영상에 표 2에 제시된 네 가지 기법을 적용하여 실제 거리값과 추정 거리값 사이의 PSNR을 계산하여 성능을 평가하였다. 두 영상은 평탄한 영역과 질감도가 높은 영역, 그리



고 반복되는 패턴이 존재하는 영역이 골고루 존재하기 때문에 실험 영상으로 적합하다. 성능 실험은 과도한 신뢰도 전파로 인해 정합 성능이 저하되어 정합 결과의 PSNR이 이전 정합 결과의 PSNR보다 낮아지는 순간 신뢰도 전파를 정지하였다. 다중 베이스라인 기반 기법의 효용을 살펴보기 위해 기존의 SAD를 국부-명백성 함수로 사용한 단일 베이스라인 기반 신뢰도 전파 기법과 SSSD를 국부-명백성 함수로 사용한 다중 베이스라인 기반 신뢰도 전파 기법을 적용 및 비교하였고, EMAD를 국부-명백성 함수로 사용한 다중 베이스라인

기반 신뢰도 전파 기법을 적용하여 제안한 기법의 정합 성능을 평가하였다. 마지막으로 EMAD를 국부-명백성 함수로 사용하면서 지역정보에 적응적으로 가중치  $\lambda_r$ 를 결정하여 신뢰도 전파 기법을 적용하여 적응적으로 지역정보를 사용하는 기법의 성능을 평가하였다. 표 2에는 실험에 사용된 정합 기법과 그에 따른 실험내용이 표기되어 있다.

그림 6은 Aloe 영상과 실제 거리정보, 그리고 Aloe 영상에 표 2의 네 가지 정합 방법을 적용한 결과이다. 여기서 그림 6(b)는 True disparity map을 실제 깊이 정보로 변환하여 8bit 계조영상으로 표현한 것으로, 어두운 색일수록 가까운 위치에 있다. 단일 베이스라인에서의 정합결과인 그림 6(c)의 경우 오정합률이 매우 높은 것을 확인할 수 있다. 부분적으로 잘못된 지역정보로 인하여 부정확한 정합이 이루어진다. SSSD를 국부-명백성 함수로 사용한 MBP를 적용한 정합 결과인 그림 6(d)에서, SBP에 비해 오정합률이 낮아진 것을 확인할 수 있다. 다음으로 제안한 알고리즘인 EMAD MBP

표 2. 정합기법의 약어에 따른 실험내용  
Table 2. Test method with respect to acronym.

정합기법	설명
SBP	단일 베이스라인 신뢰도 전파
MBP	다중 베이스라인 신뢰도 전파
EMAD MBP	EMAD기반 다중 베이스라인 신뢰도 전파
EMAD-V MBP	EMAD MBP기법에 지역정보를 적용

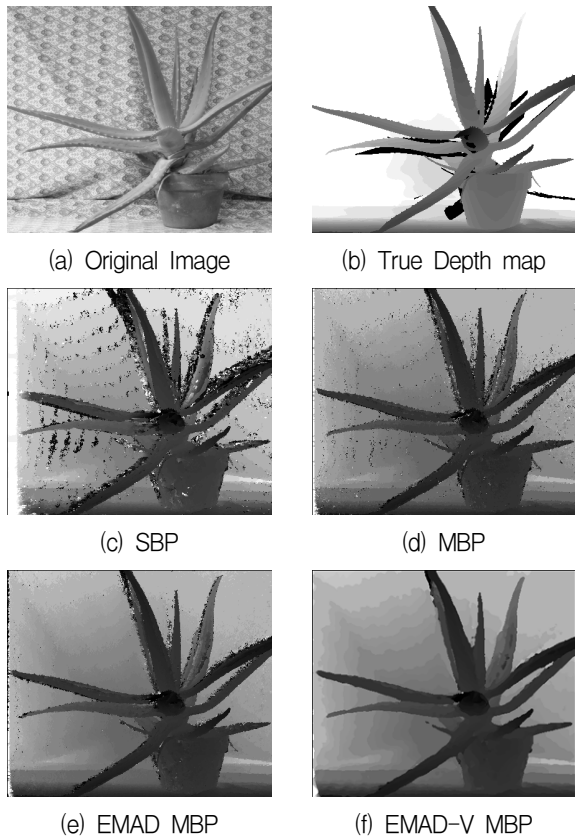


그림 6. Aloe 영상의 정합 결과  
Fig. 6. Stereo matching results of Aloe.

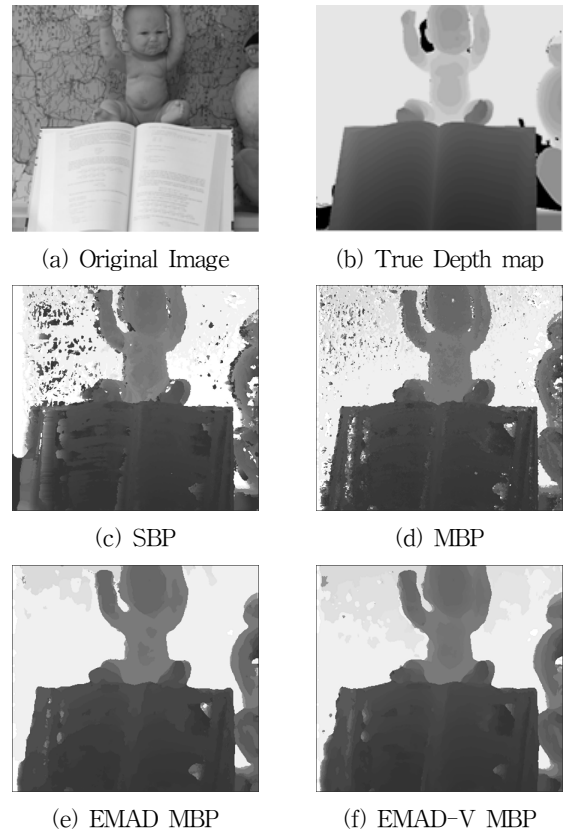


그림 7. Baby2 영상의 정합 결과  
Fig. 7. Stereo matching results of Baby2.



를 적용한 그림 6(e)의 경우, 기존 기법의 결과 영상의 왼쪽에서 보이던 폐색영역이 매우 줄어든 것을 확인할 수 있으며, 가까이 있는 물체의 폐색영역도 줄어드는 현상이 관찰된다. 또한 상대적으로 먼 거리에 있는 배경부분이 기존의 기법에 비해 더 세밀한 단계로 표현된다. 마지막으로 EMAD의 값에 지역적으로 적응적 가중치를 곱해준 방법인 EMAD-V MBP를 적용한 결과인 그림 6(f)에서는 이전 단계에서 보이던 잘못된 지역정보로 인해 객체의 경계부분에서 나타나던 오정합되는 부분이 현저하게 줄어든 것을 관찰할 수 있다.

그림 7은 Baby2 영상과 실제 거리정보, 그리고 해당 영상에 표 2의 네 가지 방법을 적용한 결과이다. 그림 7(c)의 경우 책의 글씨 부분, 배경에 있는 지도 부분에서 오정합이 과도하게 일어나는 것을 확인할 수 있다. 책의 글씨 부분은 영상 내에서 비교적 평탄한 부분이며, 배경의 지도 부분은 부분적으로 반복되는 패턴이 존재하기 때문이다. 또한 각 객체의 경계부분에서 오정합이 발생하는 것을 확인할 수 있다. 그림 7(d)의 경우 배경의 지도 부분에서 그림 7(c)에 비해 높은 정합 성능을 보여준다. 또한 각 객체의 좌측 경계부분에서 발생되던 오정합점이 줄어든 것을 확인할 수 있다. 다음으로 그림 7(e)에서는 기존에 발생하던 폐색영역이 현저히 줄어들고, 배경의 지도 부분에서 빈번히 일어나던 오정합점이 현저히 줄어든 것을 확인할 수 있다. 특히 책의 글씨 부분에서 발생하던 오정합점이 많이 줄어든 것을 확인할 수 있다. 하지만, 지역적으로 과한 신뢰도 전파로 인해 인형의 얼굴 부분, 배경부분, 책의 경사면

부분에서 거리 정보의 정밀도가 낮아지는 현상이 발생하였다. 그림 7(f)에서는 앞서 그림 7(e)에서 과도한 신뢰도 전파로 인한 거리 정보의 정밀도 하락이 완화된 것을 확인할 수 있다. 특히 책의 경사 부분과 인형의 얼굴 부분에서 거리 정보가 완만하게 변하는 것을 관찰할 수 있다. 이상에서 살펴본 주관적 화질면에서 제안한 방법이 가장 우수한 정합 성능을 보이는 것을 확인하였다.

다음으로 실험결과에 대한 객관적 분석을 토론했다. 그림 8은 표에 제시된 네 방법에 대한 메시지 갱신 반복 횟수에 따른 Aloe 영상의 PSNR의 추이를 보여준다. SBP의 경우 제한된 지역정보의 사용으로 인해 최종적으로도 낮은 PSNR을 보여주며, 최대 PSNR 까지 도달하는데 더 많은 메시지 반복이 필요한 것을 알 수 있다. 다중 베이스라인에 대한 영상 정보를 사용하는 MBP 기법의 경우 기존 기법에 비해 더 정확한 거리 정보를 추정하는 것을 확인할 수 있다. 제안하는 방법인 EMAD를 적용한 경우, 앞서의 두 기존 기법에 비해 초기 거리 추정이 더욱 정확하게 이루어지는 것을 확인할 수 있으며, EMAD-V MBP를 적용한 경우 메시지 갱신 횟수도 7번으로 가장 적으며 PSNR값에 있어서도 20.58dB로 가장 높은 성능을 보여주었다.

다음으로 Baby2 영상에 대한 실험 결과인 그림 9를 분석한다. 먼저 기존의 단일 베이스라인 기반 신뢰도 전파 알고리즘을 적용했을 경우이다. 최대 PSNR인 15.23dB에 도달하기까지 42번의 반복을 수행해야 한다. 다음으로 기존의 다중 베이스라인 기반 신뢰도 전파 알

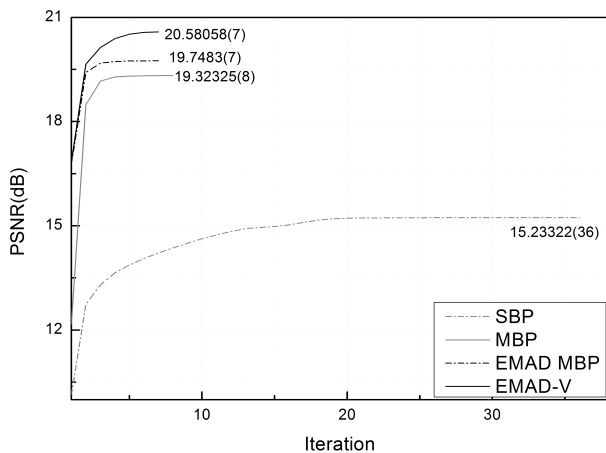


그림 8. Aloe 영상의 반복횟수에 따른 PSNR 추이  
Fig. 8. PSNR comparison results for Aloe.

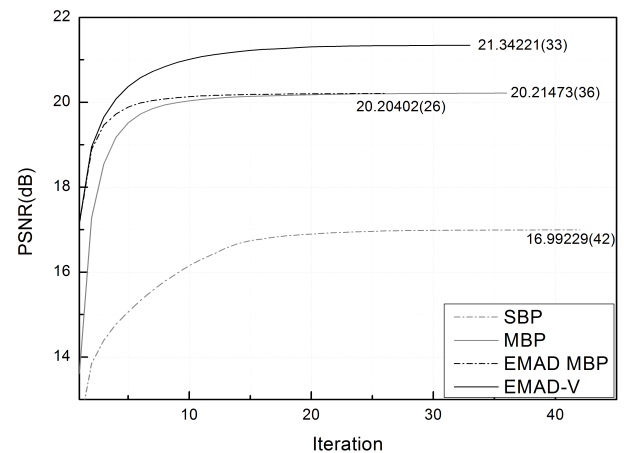


그림 9. Baby2 영상의 반복횟수에 따른 PSNR 추이  
Fig. 9. PSNR comparison results for Baby2.

고리즘을 적용한 경우이다. 최대 PSNR인 20.21dB에 도달하기까지 36번의 반복을 수행하였다. 단일 베이스라인 기반 기법에 비해 4.98dB의 성능 향상을 보였다. 제안하는 EMAD를 국부-명백성 함수로 사용한 다중 베이스라인 기반 신뢰도 전파 기법을 적용한 경우 최대 PSNR인 20.20dB에 도달하기까지 26번의 반복을 수행하였다. 수렴하는 곡선을 살펴보면, 기존 기법에 비해 초기 추정에 의한 PSNR이 매우 높은 것을 확인할 수 있다. 마지막으로 제안한 두가지 기법을 모두 사용한 경우, 초기 추정에 의한 PSNR이 약 17.3dB로 EMAD MBP의 경우와 같은 것을 확인할 수 있다. 반복 수행이 늘어남에 따라 최대 PSNR이 21.34dB로 수렴하는 것을 관찰할 수 있다. 최대 PSNR에 이르기까지 33번 반복을 하였다. 이는 지역정보에 가변적으로  $\lambda$ 를 적용하기 이전에 비해 7번이나 반복 수행한 것이다. 이러한 현상은 가변적인  $\lambda$ 에 따라 신뢰도 전파를 통해 정합성능이 더 향상되는 것을 보여준다.

## V. 결 론

본 논문에서는 기존의 단일 베이스라인 기반 스테레오 정합 알고리즘의 한계인 베이스라인의 길이에 따른 정확도와 정밀도의 이율배반적 관계를 해소하고자 다중 베이스라인에 기반하는 질감 적응적 신뢰도 전파 스테레오 정합 기법을 제안하였다. 제안한 기법은 기존 기법보다 조밀한 깊이 정보를 획득하기 위해 모든 베이스라인으로부터 후보깊이를 정의하며, 여기에 제안된 EMAD를 지역적 정합비용으로 계산한다. 또한 정합 성능을 높이기 위해 지역적 질감도에 적응적으로 가중치를 결정하여 신뢰도 전파를 수행한다. 제안한 기법의 성능을 평가한 결과 시차변이에 기반하지 않고 거리정보를 추정 대상으로 하는 EMAD기법에 의한 효과로 초기 추정 성능은 기존기법 대비 평균 5.6dB만큼 증가하였고, 이것으로 인해 기존 기법보다 더

빠르게 해에 근접하는 효과가 있었다. 또한 EMAD는 다중 베이스라인에 기인한 모든 정보를 사용하기 때문에 추정 거리의 정확도뿐만 아니라 추정 거리 정보의 분해능도 높아져서 더 세밀한 결과를 얻을 수 있었다. 결론적으로 평균 4.85dB만큼의 정합 성능의 향상이 있었다.

## 참 고 문 헌

- [1] 김곤우, 이상무, “지능형 로봇을 위한 디지털 융합 기술,” 전자공학회지, 제35권, 제2호, 30-39쪽, 2008년 2월
- [2] 정연구, 이재연, 소정, 김계경, 조영주, “지능형 로봇의 시각인식기술과 활용,” 전자공학회지, 제32권, 제 1호, 70-80쪽, 2005년 1월
- [3] 고정환, “지능형 로봇제어를 위한 3D 비전 기술동향,” 전자공학회지, 제38권, 제 12호, 72-78쪽, 2011년 12월
- [4] 최우영, 박래홍, “Stereo vision 및 응용,” 전자공학회지, 제21권, 제 10호, 36-46쪽, 1994년 10월
- [5] U. R. Dhond and J. K. Aggarwal, “Structure from stereo—a review,” Systems, Man and Cybernetics, IEEE Transactions on, vol. 19, pp. 1489-1510, 1989.
- [6] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 15, pp. 353-363, 1993.
- [7] 변영기, 한동엽, 김용일, “MRF 모델과 분할 영상을 이용한 영상정합에 관한 연구,” 한국지형공간정보학회, 2004 GIS /RS 공동 춘계학술대회 논문집, 511-516쪽, 2004년 3월
- [8] J. Sun, N. N. Zheng, and H. Y. Shum, “Stereo matching using belief propagation,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, pp. 787-800, 2003.
- [9] J. Pearl, Reverend Bayes on inference engines: A distributed hierarchical approach: Cognitive Systems Laboratory, School of Engineering and Applied Science, University of California, Los Angeles, 1982.
- [10] <http://vision.middlebury.edu/stereo/>
- [11] [http://en.wikipedia.org/wiki/Peak\\_signal-to-noise\\_ratio](http://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio)

저 자 소 개



김진형(학생회원)  
2010년 충남대학교 메카트로닉스  
공학과 학사 졸업.  
2012년 충남대학교 메카트로닉스  
공학과 석사 졸업.  
2012년~현재 충남대학교 메카트  
로닉스공학과 박사 과정.

<주관심분야 : 영상처리, 스테레오비전>



고윤호(평생회원)-교신저자  
1995년 부산대학교 전자공학과  
학사 졸업.  
1997년 한국과학기술원 전기및  
전자공학과 석사 졸업.  
2002년 한국과학기술원 전기및  
전자공학과 박사 졸업.

2002년~2004년 하이닉스반도체 선임연구원  
2004년~현재 충남대학 메카트로닉스공학과  
부교수

<주관심분야 : 영상신호처리, 컴퓨터비전, 영상인  
식, 영상압축>