

# 다양한 스테레오 카메라 배열을 위한 효율적인 깊이 지도 생성 방법

정회원 장 우 석\*, 이 천\*, 종신회원 호 요 성\*\*

## Efficient Depth Map Generation for Various Stereo Camera Arrangements

Woo-Seok Jang\*, Cheon Lee\* *Regular Members*, Yo-Sung Ho\*\* *Lifelong Member*

### 요 약

본 논문에서는 평행형 카메라 배열 뿐 아니라 수렴형 카메라 배열에서도 깊이 지도를 직접적으로 구할 수 있는 방법을 제안한다. 기존의 스테레오 정합 방법들은 복잡도를 줄이고 정확도를 향상시키기 위해서 영상 정렬화를 수행한다. 하지만 수렴형 배열에서의 영상 정렬화는 원치 않는 결과를 발생시킨다. 따라서 제안하는 방법은 문제가 되는 영상 정렬화 과정을 생략하고, 영상 자체의 에피폴라 제약 사항을 이용하여 깊이 값을 직접적으로 추출한다. 깊이 지도의 정확도를 보다 향상시키기 위해서 폐색 영역을 탐지하고 처리하는 과정을 추가적으로 수행한다. 탐지한 폐색 영역은 보이는 주변 영역의 화소들과의 거리와 색상차를 고려하여 적절한 깊이 값으로 채워진다. 실험을 통해 제안한 방법이 기존의 방법에 비해 카메라 배열에 대한 제약이 적고, 안정적으로 깊이 지도를 생성할 수 있음을 확인했다.

**Key Words** : epipolar constraint, image rectification, occlusion handling, stereo matching

### ABSTRACT

In this paper, we propose a direct depth map acquisition method for the convergence camera array as well as the parallel camera array. The conventional methods perform image rectification to reduce complexity and improve accuracy. However, image rectification may lead to unwanted consequences for the convergence camera array. Thus, the proposed method excludes image rectification and directly extracts depth values using the epipolar constraint. In order to acquire a more accurate depth map, occlusion detection and handling processes are added. Reasonable depth values are assigned to the obtained occlusion region by the distance and color differences from neighboring pixels. Experimental results show that the proposed method has fewer limitations than the conventional methods and generates more accurate depth maps stably.

### I. 서 론

3차원 영상 획득을 위해서는 3차원 장면의 거리 정보를 표현하는 깊이 지도를 획득해야한다. 깊이 정보를 이용하면 실제로 획득하지 못한 시점의 영상을 깊이 영상 기반 렌더링(Depth Image Based Rendering)을 통해서 합성할 수 있다<sup>[1]</sup>.

깊이 정보는 능동적 깊이 센서 방식(active depth sensors), 수동적 깊이 센서 방식(passive depth sensors)으로 획득할 수 있다. 능동형 센서 방식은 물리적 센서 장치를 이용하여 깊이 정보를 직접 획득하는 반면에, 수동적 센서 방식은 두 대 이상의 카메라로 얻은 영상으로부터 깊이 정보를 계산해서 획득한다<sup>[2,3]</sup>.

\* 광주과학기술원 정보통신공학과({jws, leecheon, hoyo}@gist.ac.kr) (° : 교신저자)

논문번호 : KICS2012-02-094, 접수일자 : 2012년 2월 28일, 최종논문접수일자 : 2012년 5월 27일

수동형 센서 방식의 하나인 스테레오 정합은 서로 다른 시점으로부터 얻어진 동일한 장면의 두 영상으로부터 한 영상의 화소와 일치하는 화소를 다른 영상에서 찾음으로써 3차원 데이터를 얻는다<sup>[4]</sup>. 지금까지의 스테레오 정합 방법들은 주로 평행형 배열(parallel array)로 촬영되었다는 가정 하에 개발되었다. 하지만 최근 3차원 영상을 촬영 시 두 카메라의 시야 범위를 비슷하게 하기 위해서 수렴형 카메라 배열(convergence array)이 많이 이용되기 때문에 이전 평행형 배열에서의 방법들을 적용하면 문제가 발생한다.

기존의 스테레오 정합은 보통 정확도와 속도 향상을 위해 두 영상을 정렬화하여 동일한 스캔 라인 위에서 정합 과정을 수행한다. 하지만 이러한 정렬화는 평행형 배열에서는 유용하지만, 수렴형 배열에서는 영상 왜곡을 발생시킨다. 이는 평행형 배열은 두 영상이 거의 동일한 평면에 놓여있지만, 수렴형 배열은 수렴각이 존재하기 때문이다.

본 논문에서는 기존 스테레오 정합 방법들에서 전제로 하고 있는 정렬화의 문제점을 지적하고, 이를 극복하기 위해서 에피폴라 제약 사항을 이용한 직접적인 깊이 값 추출 방법에 대해서 제안한다. 또한 깊이 지도의 질을 향상시키기 위해서 폐색 영역의 탐지와 처리 방법을 제시한다. 이는 기존의 방법들에 비해서 제약이 적고, 좀 더 유연한 깊이 지도 생성에 기여한다.

## II. 영상 정렬화를 통한 3차원 정보 획득의 문제점

영상 정렬화는 임의의 두 시점에서 촬영된 영상을 2차원 변환을 통하여 모든 에피폴라 선(epipolar line)이 평행하도록 만드는 작업으로 보통 스테레오 정합에서 필수적이다. 정렬화된 두 영상은 한 평면에 위치한 각각의 영상 평면을 가지며, 동일 선상에 위치한 두 카메라에서 촬영된 영상의 특성을 가진다. 두 영상의 대응점들은 동일한 수직 좌표를 가지며, 결과적으로 두 영상 사이에는 수평 방향의 변위만이 존재하게 된다<sup>[5]</sup>. 그림 1은 수렴형 카메라에서의 영상 정렬화 과정을 보여준다. 서로 다른 평면에 존재하던 두 카메라는 동일한 평면으로 옮겨지고, 대응점에 대한 에피폴라 선은 일치하게 된다.

그림 1에서  $p_L$ 과  $p_R$ 은 각각 좌우 영상에서의 일치하는 임의의 한 점을 나타낸다. 두 일치점이 실제 3차원 공간에 있는 위치를  $P$ 로 나타냈다.  $l_L$ 과  $l_R$ 은 각 좌우 영상의 에피폴라 선이다.

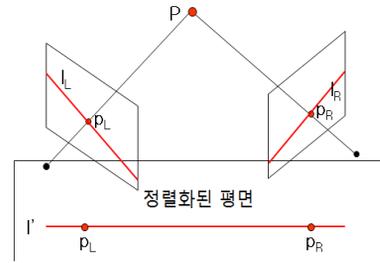


그림 1. 수렴형 카메라의 정렬화  
Fig. 1. Image rectification in the arc camera array

수렴형 배열에서 영상 정렬화의 원본 영상 훼손 문제는 수렴각을 갖는 두 평면을 동일 평면으로 옮겨서 발생한다. 그 예로 원본 영상에서 서로 다른 영역에 있는 여러 화소가, 정렬화된 평면의 한 화소로 옮겨지거나, 옮겨진 화소가 영상 평면의 밖에 있게 되는 경우이다. 그림 2는 수렴형 배열의 카메라로 획득된 두 장의 영상에 정렬화 과정을 수행한 결과와 왜곡이 발생함을 보여준다. 결과 영상의 스캔 라인은 거의 일치하지만, 원본 영상의 일부가 정렬화를 거치면서 소실되었다. 이와 같이 영상 정렬화를 이용해서는 수렴형 배열의 카메라로 획득된 영상을 보존하면서 3차원 정보를 얻기가 어렵다.



그림 2. 실제 영상에서의 정렬화 과정을 거친 후 상태  
Fig. 2. Rectified result of left and right images

## III. 에피폴라 제약 사항을 이용한 깊이 지도 획득

### 3.1 스테레오 영상에서의 에피폴라 제약 사항

영상 정렬화를 통해서 수직 방향의 변위를 제거하는 방법은 영상의 왜곡 때문에 모든 화소의 변위를 구하기 어렵다. 그렇다고 대응점을 찾기 위해서 수평 방향 뿐 아니라 수직 방향까지 스캔하는 방법은 너무 소모적일 뿐 아니라 정확도 또한 떨어진다. 우리는 에피폴라 제약 사항(epipolar constraint)을 사용하여 이를 해결한다. 에피폴라 선은 좌시점에서 카메라 중심과 영상 평면에 맺힌 화소를 연결한 직선을 3차원 공간으로 보내고, 이를 우영상에 투영하여 획득한다. 그림 3에서 에피폴라 선 위에, 좌영상에 존재하는 화소  $p_L$ 의 대응점인  $p'_R$ 이 놓여있다. 보통 1차원 변위는 쉽게 깊이 정보로 변환할 수 있

지만, 제안한 방법으로 변위를 구할 경우 2차원 변위가 발생하고 음의 변위도 발생할 수 있기 때문에, 변위로 객체가 카메라에서 얼마나 떨어져 있는지 판단할 수 없다. 따라서 본 논문에서는 변위 값을 추출하지 않고 직접 깊이 값을 구한다.

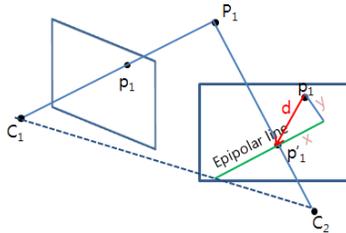


그림 3. 에피폴라 제약 사항에 따른 변위  
Fig. 3. Disparity according to epipolar constraint

### 3.2. 직접적인 깊이 값 추출

본 논문에서는 깊이 값을 추출하기 위해서 마르코프 랜덤 필드(Markov Random Field, MRF)에 의한 에너지 함수를 정의하고 최적화한다<sup>[6]</sup>. 정합을 위한 에너지 함수는 다음과 같이 정의된다.

$$E(x, y, d) = E_{data}(x, y, d) + E_{smooth}(x, y, d) \quad (1)$$

여기서  $x$ 와  $y$ 는 각각 좌영상의 수평, 수직 좌표를 나타내고,  $d$ 는 깊이 값을 나타낸다. 에너지  $E(x, y, d)$ 는 데이터 항(data term)과 평활화 항(smoothness term)으로 구성된다. 데이터 항과 평활화 항은 다음과 같이 정의된다.

$$E_{data}(x, y, d) = U\{I_L(x, y), I_R(x, y, d)\} \quad (2)$$

$$E_{smooth}(x, y, d) = \sum_{(p, q) \in N} W(p, q) \quad (3)$$

$U\{\cdot\}$ 는 좌·우영상의 화소값 차이,  $N$ 은 주변 화소의 집합이고,  $W(p, q)$ 는 주변 화소들의 변위 차이,  $I_L(x, y)$ 는 좌영상의 좌표가  $(x, y)$ 일 때의 화소값,  $I_R(x, y, d)$ 는 좌영상의 좌표  $(x, y)$ 가 깊이 값이  $d$ 일 때 우영상에 매칭되는 화소값을 나타낸다.  $I_R(x, y, d)$ 를 구하기 위해서 워핑을 이용한다. 이를 위해 먼저 깊이의 범위를 정한다. 깊이 범위는 그림 4에 제시된 것처럼 최소·최대 깊이 위치에 있는 두 영상의 화소를 3차원 공간으로 보내서 만나는 점을 찾는 삼각화(Triangulation) 방법을 이용하여 구한다. 만나는 3차원 점을 알면 그 위치의 깊이를 알 수 있다.

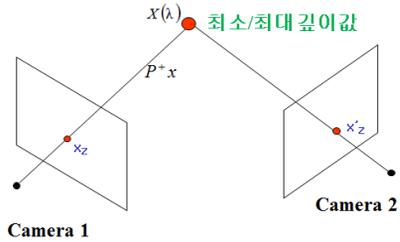


그림 4. 깊이 범위를 구하기 위한 삼각화 방법  
Fig. 4. Triangulation for range of depth

깊이 범위를 구한 후, 깊이 값에 맞게 좌영상의 화소를 3차원 워핑 기술을 이용하여 우영상으로 투영한다<sup>[7]</sup>. 각 깊이 값의 우영상에서의 대응 좌표가 정해지고, 이들은  $I_R(x, y, d)$ 의 화소값을 결정한다. 최종 에너지, 식 (1)은 시간 복잡도를 고려한 상수 공간 신뢰 확산(Constance Space Belief Propagation, CSBP)에 의해 최적화된다<sup>[8]</sup>. 얻어진 깊이 지도는 정확한 깊이 값을 위한 초기값으로 사용된다.

### 3.3. 폐색 영역 탐지 및 처리

스테레오 영상은 한 영상에서만 보이는 폐색 영역이 존재하기 때문에 3.2에서 정의한 깊이 값 추출 에너지 함수로만으로는 정확한 깊이 지도를 생성하기가 어렵다<sup>[3]</sup>. 폐색 영역을 탐지하기 위해서 3.2에서 구한 초기 깊이 지도를 이용하여 좌영상의 모든 화소를 우영상에 투영한다. 여기서 두 가지 종류의 폐색 영역 후보를 구할 수 있다. 첫 번째로 좌영상에서의 화소가 우영상의 영상 평면의 범위 밖으로 투영되는 화소인데, 이들은 우영상과 매칭되지 않는 화소이기 때문에 폐색 영역 후보로 간주할 수 있다. 에너지 함수는 다음과 같이 정의한다.

$$E_O(d_{u,v}) = \sum_{u,v} |o_{u,v} - G_O(u, v; d_{u,v})| \quad (4)$$

$G_O(u, v; d_{u,v})$ 은 우영상 평면 범위 밖으로 화소가 투영되는지를 나타내는 이진 지도이다. 우영상의 평면 범위 안으로 투영이 된다면 '0', 그렇지 않으면 '1'로 세팅한다.  $o_{u,v}$ 는 추정에 의한 폐색 영역 값이다. 화소가 폐색 영역으로 가정되면 '1'로 세팅된다. 즉, 정확히 예측되면 에너지 값이 작아지게 된다.

두 번째는 좌영상의 여러 화소가 우영상의 동일한 화소에 투영되는 경우이다. 이러한 화소 중 일부는 실제로 3차원 공간에서 다른 화소에 가려진 화소이다. 이에 관한 에너지 함수는 다음과 같다.

$$E_W(d_{u,v}) = \sum_{u,v} w_w |o_{u,v} - G_W(u,v;d_{u,v})| \quad (5)$$

$G_W(u,v;d_{u,v})$ 에서는 우영상의 동일한 화소에 투영된 좌영상의 다중 화소를 ‘1’로 세팅한다. 확률적으로 다중 매칭 화소 중 가장 깊이 값이 큰 화소가 보이는 부분일 가능성이 가장 크기 때문에  $w_w$ 를 이용하여 가중치를 달리 적용한다. 그림 5는 투영을 사용한 두 종류 폐색 영역의 검출 방법을 설명한다.

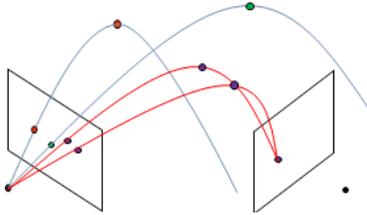


그림 5. 투영을 사용한 폐색 영역의 탐지  
Fig. 5. Occlusion detection using projection

폐색 영역 탐지를 위한 다른 방법은 대응점이라고 판단되는 두 화소들이 실제로 3차원 공간에서 같은 위치에 존재 하는지를 확인하는 것이다. 특정 화소가 폐색 영역이 아니라면, 두 영상에서 3차원 공간으로 보낸 위치는 동일해야한다. 3차원 공간에서의 폐색 영역 탐지 에너지 는 다음과 같다.

$$E_C(d_{u,v},d_{u',v'}) = \sum_{u,v} |o_{u,v} - G_C(u,v;d_{u,v},d_{u',v'})| \quad (6)$$

$G_C(u,v;d_{u,v},d_{u',v'})$ 는 두 영상에서 3차원 공간으로 보낸 위치가 동일하면 ‘0’, 아니면 ‘1’의 값을 갖는다. 그림 6은 3차원 공간에서의 폐색 영역을 탐지 하는 방법에 대해서 나타낸다.

폐색 영역을 탐지 최종 에너지 함수는 상기한 에너지 함수의 합으로 다음과 같이 나타낸다.

$$E_{OD} = \sum_{u,v} (1 - o_{u,v}) E_{data}(x,y,d_{u,v}) + \lambda_a o_{u,v} + \lambda_o E_O(d_{u,v}) + \lambda_w E_w(d_{u,v}) + \lambda_c E_C(d_{u,v},d_{u',v'}) + \sum_{p,q \in N} \lambda_s |o_p - o_q| \quad (7)$$

식 (7)은 두 에너지 함수 외에도 두 대응점의 휘도 성분 차이도 포함한다. 이는 폐색 영역 탐지 조건들에서 보이는 부분이라고 인식된 화소라도 영상 간 휘도 차이가 크다면 잘못된 정합일 수 있다는 가정에 의한 것이다. 탐지된 폐색 영역에는 주변의 보이는 영역에 있는 화소 중 폐색 영역과 비슷한 영역이라고 판단되는 화소를 이용하여 깊이 값이

채워진다<sup>31</sup>. 이와 같이 얻어진 깊이 지도는 후처리를 거쳐서 깊이 지도의 질을 향상시킨다<sup>9</sup>.

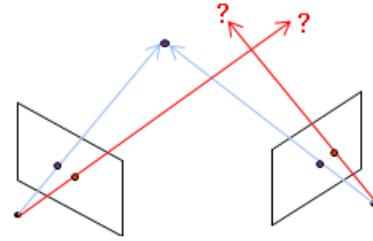


그림 6. 3차원 공간에서의 폐색 영역 탐지  
Fig. 6. Occlusion detection in 3D space

#### IV. 실험 결과

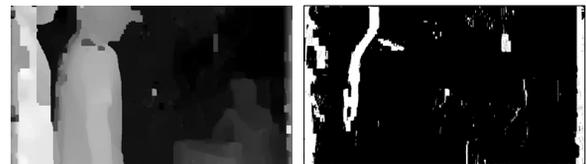
표 1은 각 에너지 항의 균형을 맞추기 위해서 실험적으로 구해진 매개변수 값들을 나타낸다. 그림 7은 수렴형 배열에서 초기 깊이 지도와 폐색 영역 지도를 보여주고, 그림 8은 평행형 배열에서 얻은 초기 지도 및 폐색 영역 탐지 결과를 보여준다.

표 1. 실험에 사용된 매개변수  
Table. 1. Experimental parameters

	$\lambda_a$	$\lambda_o$	$\lambda_w$	$\lambda_c$	$\lambda_s$	$\sigma$
값	7.5	12	3	12	4.2	7



(a) 원본영상



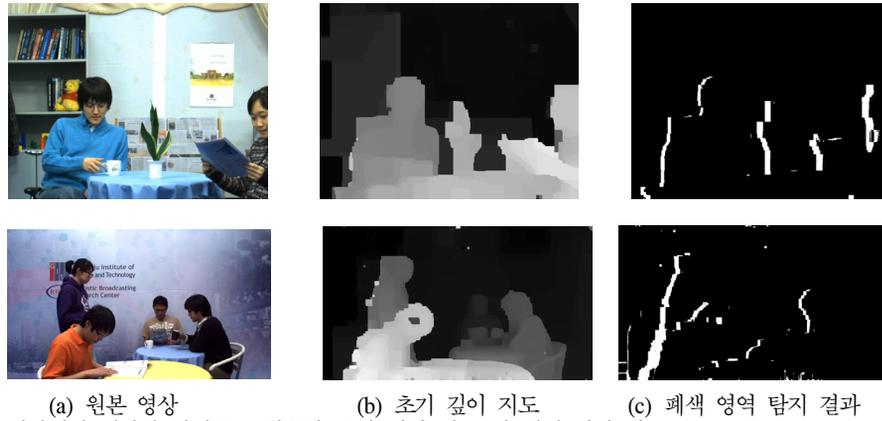
(b) 초기 깊이 지도

(c) 폐색 영역 탐지

그림 7. 수렴형 배열 영상에서 제안한 방법으로 획득한 초기 깊이 지도 및 폐색 영역 지도

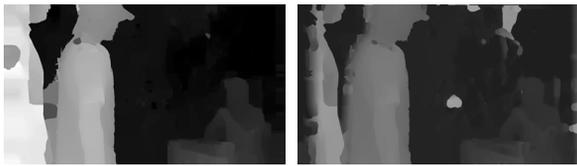
Fig. 7. Initial depth and occlusion maps obtained by the proposed method in convergence camera array

그림 9는 제안하는 최종 깊이 지도와, 영상 정렬 화를 거쳐서 얻은 변위 지도를 깊이 지도로 변환한 영상을 나타낸다. 그림 9(b)는 상수 공간 신뢰 확산으로 초기 변위 지도를 구하고 후처리를 적용했다<sup>18,91</sup>. 그림 9(b)를 보면 영상의 상단 부분이 정렬화 과정에 의해 우영상에서 사라져 정확한 깊이 값을 얻을 수 없었던 반면, 그림 9(a)에서는 모든 영역에



(a) 원본 영상 (b) 초기 깊이 지도 (c) 폐쇄 영역 탐지 결과  
 그림 8. 평행형 배열 영상에서 제안한 방법으로 획득한 초기 깊이 지도 및 폐쇄 영역 지도  
 Fig. 8. Initial depth and occlusion maps obtained by the proposed method in a parallel array

서 깊이 값을 얻을 수 있었다.

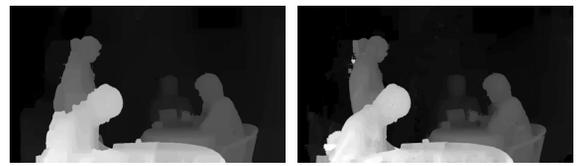


(a) 제안하는 방법 (b) 정렬화 과정 적용  
 그림 9. 수렴형 배열에서 깊이 지도 결과 비교  
 Fig. 9. Depth map comparison in an arc camera array

그림 10과 그림 11은 평행형 카메라 배열에서 제안한 방법으로 구한 깊이 지도와 영상 정렬화 과정을 거쳐서 얻은 깊이 지도의 결과를 보여준다.



(a) 제안하는 방법 (b) 정렬화 과정을 적용  
 그림 10. 'Newspaper' 영상에서의 깊이 지도 결과 비교  
 Fig. 10. Depth map comparison on 'Newspaper'



(a) 제안하는 방법 (b) 정렬화 과정을 적용  
 그림 11. 'Cafe' 영상에서의 깊이 지도 결과 비교  
 Fig. 11. Depth map comparison on 'Cafe'

평행형 배열에서는 영상 정렬화를 이용한 기존 방법과 제안한 방법이 모두 정확한 깊이 값을 예측했지만, 수렴형 카메라 배열에서는 제안한 방법이 기존 방법보다 우수한 결과를 보였다. 실험 결과를

객관적으로 평가하기 위해서 새로운 시점을 합성하였다. 시점 합성의 정확도를 계산하기 위해서는 세 개 이상의 시점이 필요하기 때문에 스테레오 영상인 수렴형 배열 영상은 객관적 평가에서 제외하였다. 표 2는 제안한 방법과 기존 방법과의 PSNR 차를 비교한 것이다. 그림 12와 그림 13은 두 방법의 시점 합성 결과를 보여준다.

표 2. 실시간 합성의 PSNR 비교  
 Table. 2. PSNR comparison of view synthesis

영상	제안하는 방법	기존 방법
Newspaper	29.64 dB	26.18 dB
Cafe	33.46 dB	32.16 dB



(a) 제안하는 방법 (b) 기존 방법  
 그림 12. 'Newspaper' 영상에서의 시점 합성 결과  
 Fig. 12. Result of view synthesis on 'Newspaper'



(a) 제안하는 방법 (b) 기존 방법  
 그림 13. 'Cafe' 영상에서의 시점 합성 결과  
 Fig. 13. Result of view synthesis on 'Cafe'

위 실험 결과로부터 제안하는 방법은 카메라 배열에 상관없이 정확한 깊이 값을 얻어내는 것을 알 수 있다. 따라서 제안하는 방법은 영상 촬영 전 카메라의 설치 시 제약 사항을 줄여주는 효과를 낸다.

## VI. 결 론

본 논문에서는 다양한 카메라 배열에서 별다른 제약없이 스테레오 영상의 깊이 지도를 획득하는 방법을 제시했다. 스테레오 정합에서 일반적으로 사용되는 영상 정렬화는 평행형 카메라 배열에서는 효율적이나, 수렴형 배열에서는 정확한 영상 정렬이 어려울 뿐만 아니라 원본 영상이 훼손될 수 있다. 따라서 본 논문은 에피폴라 제약 사항에 의해서 영상 정렬화 없이 깊이 값을 구하는 방법을 제안했다. 이 방법은 기존 방법에 비해 영상 전체에 대해서 깊이 값을 구할 수 있는 확률을 높였다. 또한 서로 다른 위치에서 획득되어 생긴 폐색 영역을 탐지하고, 주변 화소와의 관계를 이용하여 적절한 깊이 값을 할당하여 깊이 지도의 정확도를 높였다. 제안하는 방법은 기존의 방법들에 비해 다양한 카메라 배열에서 보다 안정적으로 깊이 지도를 생성했다.

## 참 고 문 헌

[1] L. Zhang and W.J. Tam, "Stereoscopic image generation based on depth images for 3DTV," IEEE Trans. Broadcast., vol. 51, no. 2, pp. 191-199, June 2005.

[2] S.Y. Kim, J.H. Cho, and A. Koschan, "3D video generation and service based on a TOF depth sensor in MPEG-4 multimedia framework," IEEE Trans. Consum. Electron., vol. 56, no. 3, pp. 1730-1738, Aug. 2010.

[3] W.S. Jang, Y.S. Ho, "Efficient Disparity Map Estimation Using Occlusion Handling for Various 3D Multimedia Applications," IEEE Trans. Consum. Electron., vol. 57, no. 4, pp. 1937-1943, Nov. 2011.

[4] D. Sharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," IEEE Workshop on Stereo and Multi-Baseline Vision, pp. 131-140, 2001.

[5] Y.S. Kang and Y.S. Ho, "An Efficient Image Rectification Method for Parallel Multi-Camera Arrangement," IEEE Trans. Consum. Electron., vol. 57, no. 3, pp. 1041-1048, Aug. 2011.

[6] S.Z. Li, Markov Random Field Modeling in

Image Analysis, 2nd ed. New York: Springer-Verlag, 2001.

[7] C. Lee and Y.S. Ho, "Boundary filtering on synthesized views of 3D video," International Symposium on Signal Processing, Image Processing and Pattern Recognition, pp. 15-18, 2008.

[8] Q. Yang, L. Wang, and N. Ahuja, "A constant-space belief propagation algorithm for stereo matching," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1458-1465, 2010.

[9] Q. Yang, C. Engels, and A. Akbarzadeh. "Near real-time stereo for weakly-textured scenes," British Machine Vision Conference, pp. 80 - 87, 2008.

장 우 석 (Woo-Seok Jang)

정회원



2007년 전남대학교 전자공학  
과졸업(학사)

2009년 광주과학기술원 정보  
통신공학과 졸업 (석사)

2009년~현재 광주과학기술원  
정보통신공학과 박사과정  
<관심분야> 디지털 영상처리,

3D영상처리, 실감방송

이 천 (Cheon Lee)

정회원



한국통신학회 논문지  
제37A권, 제4호 참조

호 요 성 (Yo-Sung Ho)

중신회원



한국통신학회 논문지  
제37C권, 제4호 참조