

# 스테레오 영상에서 깊이정보 획득 및 중간시점 영상합성 방법

장용준, 호요성  
광주과학기술원 정보통신공학부  
e-mail : [yjchang@gist.ac.kr](mailto:yjchang@gist.ac.kr), [hoyo@gist.ac.kr](mailto:hoyo@gist.ac.kr)

## Depth Estimation and View Synthesis for Stereo Images

Yong-Jun Chang, Yo-Sung Ho  
School of Information and Communications  
Gwangju Institute of Science and Technology (GIST)

### 요약

다시점 영상은 시청자에게 단일시점 영상보다 자유로운 시점의 장면을 보여주며, 영상에 입체감을 더해 더욱 실감나는 영상을 제공한다. 다시점 영상을 획득하는 방법으로, 서로 다른 두 시점의 깊이정보를 이용해 중간시점 영상을 합성하는 방법이 사용되는데, 깊이정보를 얻기가 쉽지 않다. 이 논문에서는 스테레오 영상에서 깊이정보를 획득하기 위해, 스테레오 정합만을 이용하는 방법, ToF(time of flight) 깊이 카메라로 얻은 초기 깊이정보를 이용하여 스테레오 정합을 하는 방법, 스테레오 정합 없이 ToF 깊이 카메라의 정보만을 직접 이용하는 방법을 고찰했다. 또한, 획득된 깊이정보의 정확도를 살펴보기 위해 각 방법으로 구한 깊이정보를 이용하여 합성한 중간시점 영상을 비교하여 분석했다.

### 1. 서론

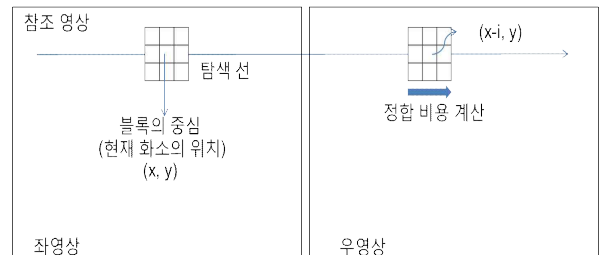
최근 상영하고 있는 대부분의 3차원 영화는 스테레오 영상의 양안 시차를 이용한 방법이다. 이 방법은 사람의 양쪽 눈에 서로 다른 시점의 영상을 보여주며 영상에 입체감을 더한다. 하지만 스테레오 영상을 이용한 3차원 영상은 시점이 다양하지 못한 단점이 있다. 이를 보완하기 위해, 여러 시점에서 촬영한 장면을 이용하여 영상이 입체적으로 보이게 해주는 연구도 진행되어왔다. 다시점 영상은 여러 대의 카메라로 다양한 시점의 장면을 촬영해 획득할 수 있다. 그러나 이 방법은 많은 수의 카메라가 들어가기 힘든 좁은 공간에선 사용하기 힘들다.

적은 수의 카메라를 이용하여 중간시점 영상을 합성하는 방법은 공간적인 제약을 적게 받으면서 효율적으로 다시점 영상을 획득할 수 있다. 이러한 중간시점 영상합성에 필요한 정보 중 하나가 바로 깊이정보이며, 깊이정보를 구하기 위해 여러 가지 방법이 사용된다. 이 논문은 스테레오 영상을 이용하여 중간시점 영상합성을 했으며, 합성에 필요한 깊이정보를 얻기 위해 세 가지 방법으로 실험을 진행했다. 서로 다른 방법으로 얻은 깊이정보를 이용하여 중간시점 영상을 합성한 후, 실제 중간시점 위치에서 촬영한 영상과 합성된 영상 사이의 PSNR을 구해 결과를 비교하고 분석했다.

### 2. 스테레오 영상을 이용한 깊이정보 획득

#### 2.1 지역기반 스테레오 정합을 이용한 깊이정보 획득

스테레오 영상을 이용한 깊이정보 획득 방법으로 스테레오 정합이 있다. 스테레오 정합은 좌영상과 우영상 사이의 변위 차이를 이용하여 깊이정보를 획득한다. 이때 물체는 변위 차이가 클수록 카메라와 가까이 있으며 작을수록 멀리 있다.



(그림 1) 좌영상 기준의 스테레오 정합 과정

그림 1은 좌영상의 깊이정보를 얻기 위한 스테레오 정합 과정을 보여준다. 이 그림에서  $x$ 와  $y$ 는 화소의 위치를 뜻하며,  $i$ 는 좌영상의 블록 위치를 기준으로 우영상의 블록이 이동한 변위정보를 나타낸다. 한 화소에 대한 스테레오 정합은 같은 탐색선에서 수행된다. 이때 최소 정합 비용을 갖는 블록의 위치와 참조 영상의 블록 위치 차이로 해당 화소의 최종 변위값을 예측한다.

정합 비용은 참조블록과 비교블록의 유사도 측정 방법으로 구할 수 있다. 유사도 측정에 사용되는 식으로 SAD(sum of absolute differences), SSD(sum of squared differences), NCC(normalized cross correlation) 등이 있

지만, 이 논문은 SAD 방법을 사용하여 유사도를 측정하였다. 블록 정합에 사용되는 SAD 수식은 식 (1)과 같다.

$$C_L(x, y, d) = \sum_u \sum_v |I_L(u, v) - I_R(u - d, v)| \quad (1)$$

식 (1)은 좌영상을 기준으로 스테레오 정합을 할 때 발생하는 정합 비용을 수식으로 나타낸 것이다.  $x, y$ 는 변위값을 구하려는 현재 화소의 위치를 뜻하며,  $u$ 와  $v$ 는 블록 안에 있는 화소의 좌표값을 나타낸다.  $d$ 는 변위값으로  $d$ 에 따라 우영상의 블록을 움직이며 정합 비용  $C_L$ 을 계산한다.  $I_L$ 과  $I_R$ 은 각각 좌영상과 우영상의 화소값이다.

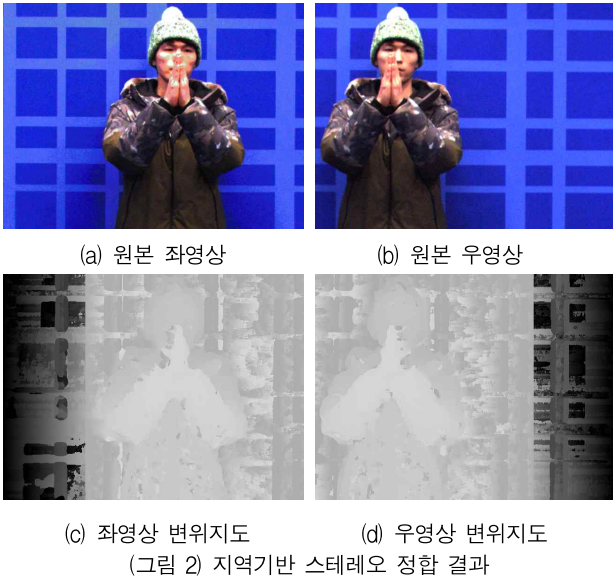


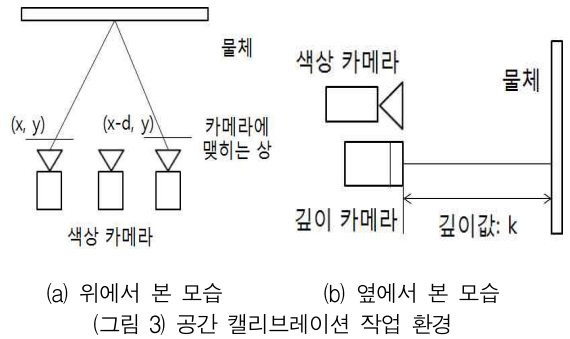
그림 2는 11×11 블록 크기로 스테레오 정합을 수행한 결과를 보여준다. 결과 영상에서 좌영상 변위지도와 우영상 변위지도를 보면, 잘못된 변위값이 발생한 영역을 확인할 수 있다. 이 영역은 각 시점에서 서로 보이지 않는 폐색영역을 나타낸다. 따라서 기존의 스테레오 정합은 별도의 후처리 작업을 없이 이러한 폐색영역을 채우기 힘들다.

### 2.2 ToF 깊이 카메라의 깊이정보를 이용한 스테레오 정합

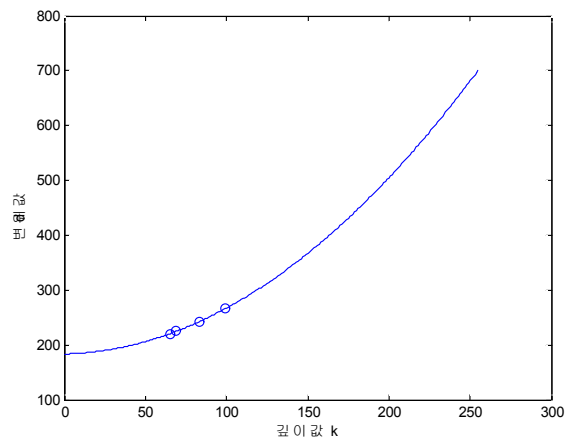
지역기반 스테레오 정합의 단점은 변위값 탐색범위와 블록의 크기에 따라 변위 예측의 정확도와 수행 시간이 달라지는 것이다. 일반적으로 정합에 사용하는 블록의 크기가 클수록 정합 비용 계산에 필요한 연산량이 많아져 수행 시간이 길어지며, 변위값 탐색범위 또한 넓을수록 스테레오 정합에 걸리는 시간이 늘어나게 된다. 이러한 문제점을 개선하기 위해, 색상 카메라와 ToF 깊이 카메라를 함께 사용하여 깊이 영상의 정보로 변위값 탐색범위를 제한하는 하이브리드(hybrid) 방법이 제안되었다 [1].

깊이 카메라로 촬영한 깊이값은 적외선으로 측정된 물체의 깊이를 8 비트로 표현하기 때문에, 스테레오 정합을 통해 구한 변위값과 다르다. 따라서 깊이 영상의 정보를

스테레오 정합에 사용하기 위해 깊이값을 변위값으로 변환하는 작업이 필요한데, 우리는 이 작업을 공간 캘리브레이션(calibration)을 통해 수행할 수 있다.



공간 캘리브레이션은 색상 카메라와 깊이 카메라를 모두 사용하며 진행된다. 우선 물체의 위치를 카메라로부터의 거리에 따라 바뀌가며 여러 장 촬영한다. 촬영된 여러 장의 영상 중 특정 거리에서 물체를 촬영한 모습을 그림 3에 나타냈다. 촬영된 물체는 깊이 카메라에 의해 깊이값  $k$ 가 측정된다. 이때 깊이 영상과 동시에 촬영된 스테레오 영상의 변위값  $d$  또한 얻을 수 있다. 이렇게 구한 값들을 이용하여 깊이와 변위값의 좌표를 만든다. 같은 방법으로 다른 거리에서 촬영한 영상들에 대해서도 깊이와 변위값 사이의 좌표를 만들 수 있다. 이렇게 만들어진 여러 개의 좌표를 통해 다항식 근사(polynomial approximation)를 해주면 그림 4와 같이 깊이값과 변위값 사이의 관계식을 정의할 수 있다.



(그림 4) 깊이값과 변위값의 관계

깊이 카메라로 촬영한 깊이 영상은 일반적으로 색상 카메라로 촬영한 영상보다 해상도가 작다. 따라서 깊이 영상의 해상도를 색상 영상의 해상도와 맞추기 위한 전처리 과정으로 워핑(warping)을 수행한다. 변위값으로 변환된 깊이 영상을 색상 영상으로 워핑하면 색상 영상의 화소마다 초기 변위정보를 얻을 수 있다. 그림 5는 깊이 영상의

변위값을 색상 영상으로 워핑한 결과다.

워핑된 변위값은 색상 영상에서 스테레오 정합을 할 때 변위 탐색범위를 제한한다. 각 화소의 초기 변위정보는 스테레오 정합에 사용되는 블록과 같은 크기의 블록을 사용하여 결정한다. 초기 변위정보 탐색을 위해, 워핑된 변위지도에서 변위값을 구하려는 화소 위치에 블록을 입힌다. 이때 블록의 중심은 초기 변위정보를 구하려는 화소의 위치이다. 화소의 초기 변위정보를 결정하는 방법은 세 가지 경우로 나누어 고려했다.



(a) 원본 촬영상 (b) 워핑된 변위값  
(그림 5) 깊이 영상의 변위값 워핑

먼저 블록의 중심에 초기 변위정보가 있을 경우, 그 값을 현재 화소의 초기 변위정보로 이용한다. 두 번째로 블록의 중심이 아닌 블록 내부의 다른 위치에 초기 변위정보가 있다면, 블록 안에 있는 변위정보의 평균값을 초기 변위정보로 정한다. 마지막으로 블록 안에 어떤 초기 변위정보도 포함되지 않을 경우, 기존의 지역기반 스테레오 정합에 사용되는 변위 탐색범위를 이용한다.

화소의 초기 변위정보가 결정되면 각 화소의 초기 변위정보에 10의 오차범위를 더해 최소 변위값과 최대 변위값을 결정하며, 이 범위를 그 화소의 최종 탐색범위로 한다.



(a) 촬영상 변위지도 (b) 우영상 변위지도  
(그림 6) 하이브리드 스테레오 정합 결과

그림 6은 깊이 영상을 이용하여 변위 탐색범위를 제한한 스테레오 정합 결과이다. 원본 영상과 스테레오 정합에 사용한 블록 크기는 그림 2에서 사용한 것과 같다. 그림 2의 변위지도와 그림 6의 변위지도를 비교하면, 하이브리드 알고리즘으로 구한 변위지도가 폐색영역의 깊이정보 획득에 더 강인하며 영상의 배경 부분도 조금 더 깔끔하게 나온 것을 확인할 수 있다.

### 2.3 ToF 깊이 영상을 이용한 깊이정보 획득

ToF 깊이 카메라로 촬영한 깊이 영상은 2.2절의 방법처럼 색상 영상으로 변위값을 워핑할 수 있다. 이때 워핑된 변위값은 스테레오 정합에서 변위 탐색범위를 제한하는 역할도 하지만 워핑된 변위값만으로 각 시점의 깊이정보를 얻을 수 있다 [2].

워핑된 변위지도는 그림 5(b)처럼 변위값 사이에 홀이 생기게 되는데, 홀을 채우기 위해 결합형 양방향 필터(joint bilateral filter)를 이용한 업샘플링(upsampling) 방법을 사용한다.

$$D_o(x, y) = \frac{\sum_u \sum_v W(u, v) D_i(u, v)}{\sum_u \sum_v W(u, v)} \quad (2)$$

$$W(u, v) = \exp\left\{-\frac{\|I(x, y), I(u, v)\|^2}{2\sigma_c^2}\right\} \cdot \exp\left\{-\frac{(x-u)^2 + (y-v)^2}{2\sigma_d^2}\right\} \quad (3)$$

식 (2)에서  $D_o$ 는 출력되는 변위값이며  $D_i$ 는 입력 변위값이다.  $(x, y)$ 는 필터 커널(kernel)의 중심 화소 좌표,  $(u, v)$ 는 중심 화소 주변의 화소 위치를 나타낸다. 따라서 식 (2)는 입력 변위값에 필터 가중치  $W$ 를 곱한 결합형 양방향 필터 수식을 나타내고, 식 (3)은 필터의 가중치 수식을 나타낸다.  $I$ 는 화소의 밝기값을 뜻하며, 가중치 값은 밝기값의 유클리드 거리(Euclidean distance)와 화소의 위치 차이를 이용하여 결정한다. 그림 7은 워핑된 깊이정보에 결합형 양방향 필터를 적용하여 만든 변위지도이다.



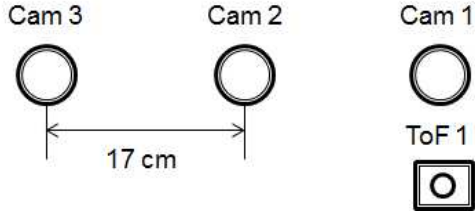
(a) 촬영상 변위지도 (b) 우영상 변위지도  
(그림 7) 워핑 + 업샘플링 결과

### 3. 실험 결과

2장에서 소개한 세 가지 깊이정보 획득 방법을 실험하기 위해, 그림 8처럼 색상 카메라 3대와 ToF 깊이 카메라 1대로 실험 환경을 구성했다. 이때 스테레오 정합에 사용되는 모든 블록 크기는 11×11로 고정했으며, 색상 영상의 해상도는 640×480, 그리고 깊이 영상의 해상도는 176×144이다. 색상 카메라 사이의 거리는 17 cm로 설치했다.

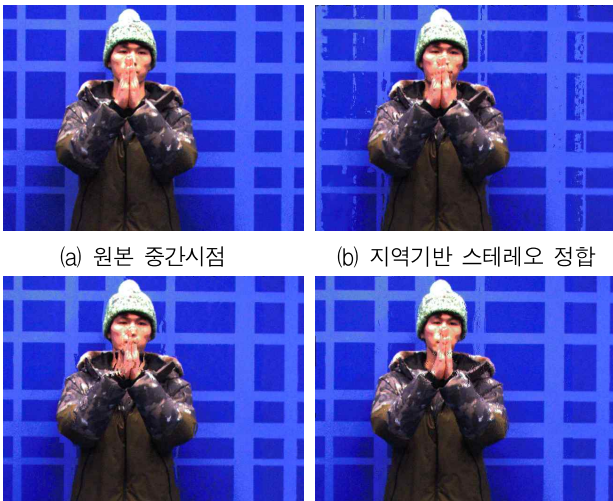
실험 목적은 세 가지 방법으로 획득한 깊이정보의 정확도와 수행 시간을 비교하는 것이다. 정확도 비교를 위해, 1번 색상 카메라와 3번 색상 카메라의 깊이정보를 구한

후 2번 카메라 위치로 중간시점 영상을 합성했다. 이때 합성한 중간시점 영상을 2번 색상 카메라로 촬영한 영상과 비교해 깊이정보의 정확도를 확인했다. 초기 변위정보는 1번 ToF 카메라에서 촬영한 깊이정보를 3번과 1번 색상 카메라로 워핑시켜 얻는다.



(그림 8) 깊이정보 획득 실험 환경

그림 9는 실제 중간시점 위치에서 촬영한 원본 영상과 세 가지 깊이정보 획득 방법으로 중간시점 영상을 합성한 결과를 보여준다.



(a) 원본 중간시점 (b) 지역기반 스테레오 정합  
(c) 하이브리드 (d) 워핑 + 업샘플링  
(그림 9) 중간시점 합성 결과

먼저, 지역기반 스테레오 정합으로 구한 깊이정보로 중간시점 영상을 합성하면 영상의 배경이 많이 왜곡된 것을 볼 수 있다. 반면에 하이브리드 방법의 깊이정보로 합성한 중간시점 영상은 배경 영역이 원본 영상과 거의 흡사한 것을 확인할 수 있다. 하이브리드 방법은 배경의 깊이정보를 ToF 깊이 영상을 참조하기 때문에 지역기반 스테레오 정합 방법보다 더 정확한 값을 얻는다. 마지막으로, 워핑과 업샘플링으로 깊이정보를 구한 후 중간시점 영상을 합성하면, 육안으로 봤을 때 전체적으로 큰 왜곡 없이 깔끔한 결과가 나오는 것을 볼 수 있다.

합성한 중간시점 영상들의 객관적인 품질 비교를 위해 원본 중간시점 영상과 실험 결과로 얻은 중간시점 영상 사이의 PSNR을 구했다. 식 (4)는 PSNR 값을 구하는 수식을 나타낸다.

$$PSNR = 10 \log \left( \frac{255^2}{MSE} \right) \quad (4)$$

식 (4)에서 MSE(mean square error)는 원본 영상과 합성한 중간시점 영상 사이의 평균 제곱 오차를 의미하며, 이를 식 (5)에 정리했다.

$$MSE = \frac{\sum_{x \in M} \sum_{y \in N} (I(x, y) - \hat{I}(x, y))^2}{MN} \quad (5)$$

식 (5)에서  $M$ 과  $N$ 은 각각 영상의 너비와 높이를 뜻하며,  $x$ 와  $y$ 는 영상의 화소 위치를 나타낸다.  $I$ 는 원본 영상의 밝기값이며  $\hat{I}$ 는 합성한 영상의 밝기값이다.

각 알고리즘의 전체적인 성능을 비교하기 위해 PSNR뿐만 아니라 깊이정보를 구하는 데 걸리는 시간을 측정하여 표 1에 정리했다. 이때 워핑과 업샘플링을 이용한 방법의 수행 시간은 워핑 및 결합형 양방향 필터를 이용한 시간도 포함했다.

(표 1) 중간시점 영상 결과 비교

실험 방법 \ 실험 결과	PSNR	수행 시간
	dB	초
스테레오 정합	23.09	19.80
하이브리드	25.44	5.94
워핑 + 업샘플링	22.35	0.30

#### 4. 결론

이 논문에서는 스테레오 영상의 깊이정보를 세 가지 다른 방법으로 구했으며, 획득한 깊이정보의 정확도 비교를 위해 중간시점 영상을 합성했다. 그 결과, 하이브리드 방법이 일반 스테레오 정합 방법보다 더 높은 PSNR 값과 2배 이상 빠른 수행 시간을 보였다. 워핑 및 업샘플링을 이용한 방법은 가장 빠른 수행 시간을 보였으나, PSNR 값이 가장 낮게 나왔다. 따라서 하이브리드 방법을 이용한 깊이정보 획득 및 중간시점 영상 합성이 비교적 빠른 시간에 정확한 깊이정보를 획득한 것을 확인할 수 있었다.

#### 감사의 글

본 연구는 미래창조과학부 ‘범부처 Giga KOREA 사업’의 일환으로 수행하였음. [GK13C0100, 기가급 대용량 양방향 실감 콘텐즈 기술 개발]

#### 참고문헌

- [1] 강윤석, 호요성, “깊이 카메라를 이용한 고해상도 변위 정보 획득 방법,” Telecommunications Review, 제21권, 제4호, pp. 599-608, 2011.
- [2] 고은상, 호요성, “다중 GPU 기반의 고속 다시점 깊이 맵 생성 방법,” 한국방송공학회 추계학술대회, B4-1, pp. 1-4, 2014.