

한 시점 깊이맵과 카메라 사이의 기하정보를 이용한 다시점 깊이맵 생성 방법

이은경, 윤승욱, 호요성
광주과학기술원 정보통신공학과
{eklee78, suyoon, hoyo}@gist.ac.kr

Generation of Multiview Depth Maps from a Single Depth Image Based on Camera Geometry

Eun-Kyung Lee, Seung-Uk Yoon, and Yo-Sung Ho
Gwangju Institute of Science and Technology

요 약

최근 고품질의 영상 서비스에 대한 수요가 증가함에 따라 다양한 멀티미디어 응용분야에서 깊이맵을 이용한 장면 생성 연구가 활발히 진행되고 있다. 평행한 카메라 배치에서는 깊이맵을 이용하면 임의 시점의 영상을 간단한 화소이동 (pixel shifting) 함수를 사용하여 쉽게 생성할 수 있다. 일반적으로 스테레오 정합을 이용해 깊이맵을 생성하지만, 이 방법은 시간이 많이 걸릴 뿐만 아니라 정확도도 떨어지는 단점이 있다. 본 논문에서는 한 시점의 깊이맵으로부터 다시점 카메라 사이의 기하정보를 이용하여 다시점 깊이맵을 생성하는 방법을 제안한다. 우선 주어진 한 장의 깊이맵으로부터 주변 카메라와의 기하관계를 고려해 다른 시점의 초기 깊이맵과 색상영상을 생성한다. 생성된 다시점 영상은 원본 영상과 위치가 정확히 일치하지 않으므로 이를 보정하고, 물체 사이의 서로 다른 시차로 인해 발생하는 공백(hole)을 카메라의 위치정보를 이용해 채운다. 결과적으로 기존의 다시점 카메라에서 일반적인 스테레오 매칭 방법을 이용하는 경우보다 빠른 시간에 다시점 깊이맵을 생성할 뿐만 아니라 향상된 화질의 깊이맵을 생성한다.

1. 서론

고품질 영상 서비스에 대한 수요가 증가함에 따라, 사용자들은 자유시점 TV, 3차원 TV, 실감방송과 같은 응용 분야에서 다양한 형태의 멀티미디어를 이용해 실감나는 서비스를 즐길 수 있게 되었다[1][2]. 그 중에서도 깊이맵을 통한 장면 생성 방법이 폭넓게 사용되고 있다. 최근 ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group(MPEG)에서 깊이맵과 다시점 색상정보를 이용한 자유시점 TV에 대한 표준화 활동이 활발하다[3][4].

깊이정보를 이용하면 중간 시점의 영상을 생성할 수 있어, 사용자가 원하는 시점의 영상을 제공할 수 있는 장점이 있다. 특히, 다시점 깊이맵 생성은 사용자에게 현실감 있는 장면을 제공하기 위해 반드시 필요하기 때문에 멀티미디어 응용 분야에서 가장 중요한 연구 분야 중의 하나로 주목을 받고 있다[5].

깊이맵을 생성하는 방법은 크게 센서 기술을 이용한 깊이맵 획득 방법과, 일반적인 스테레오 정합 방법을 통한 깊이맵 생성 방법으로 나뉜다. 깊이정보를 획득하는 센서 기술들이 발전함에 따라, 깊이 카메라를 이용하여 실제 장면에서 직접 깊이정보를 실시간으로 얻을 수 있게 되었다. 하지만 센서를 이용할 경우, 한 번에 한 장의 깊이맵만을 획득할 수 있다는 단점이 있다. 한편, 인간의 시각 체계는 서로 다른 위치에서 획득한 두 영상을 적절히 정

합함으로써 깊이정보를 얻을 수 있다. 이 특징을 이용한 방법이 일반적인 스테레오 정합 방법이다. 스테레오 정합 방법은 다시점 또는 스테레오 카메라로부터 획득한 좌, 우 영상의 대응점을 찾아 모든 화소마다 변이(disparity)를 찾는다. 그리고 찾은 변이를 깊이 값으로 변환하여 깊이맵을 생성한다. 그러나 스테레오 정합 방법을 통해 깊이맵을 생성할 경우, 시간이 많이 걸릴 뿐만 아니라 깊이맵에 포함된 삼차원 깊이정보의 정확도도 떨어진다.

이러한 단점을 극복하기 위해 본 논문에서는 한 시점의 깊이맵으로부터 다시점 카메라 사이의 기하정보를 이용하여 다시점 깊이맵을 생성하는 방법을 제안한다. 우선 주어진 한 장의 깊이맵으로부터 주변 카메라와의 기하관계를 고려해 다른 시점의 초기 깊이맵과 색상영상을 생성한다.

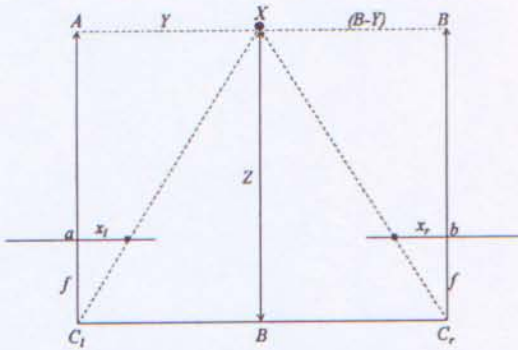
그러나 생성된 다시점 영상은 원본 영상과 위치가 정확히 일치하지 않으므로 이를 보정하고, 물체 사이의 서로 다른 시차로 인해 발생하는 공백(hole)을 카메라의 위치정보를 이용해 채운다. 일반적인 스테레오 정합에서는 찾은 변이를 3차원 깊이정보로 변환하여 깊이맵을 얻는다. 역으로 깊이맵에서 획득한 깊이정보를 초기 변이값으로 미리 알고 있다면 다시점 깊이맵을 빠르게 생성할 수 있다.

본 논문에서는 한 시점의 깊이맵을 이용하여 초기 변이값을 얻고, 이를 색상정보를 기반으로 보정하여 다시점 깊이맵을 고속으로 생성하는 방법을 제안한다.

2. 스테레오 영상의 깊이정보

인간의 시각 체계는 서로 다른 위치에서 획득된 두 영상을 적절히 정합하여 거리 정보를 얻는다. 스테레오 정합은 인간 시각 체계의 거리 추출 능력을 자동화하기 위한 컴퓨터 시각처리 분야 중 하나이다. 이 방법은 초음파와 레이저를 광원으로 사용하여 빛의 비행 시간과 속도의 함수로 거리를 측정하는 것보다 효과적이고 실제 응용시 환경의 제약을 적게 받는다는 장점 때문에 의료 영상, 공장 자동화 및 지도 제작에 널리 이용되고 있다.

일반적으로 스테레오 카메라의 기하학적인 모델은 두 대의 카메라가 하나의 물체를 바라보고 있는 구조를 가지며, 응용분야에 따라 다양한 형태를 나타낸다. 그림 1은 두 카메라가 평행하게 배치되어 있는 구조이다.



(그림 462) 스테레오 카메라의 기하학적 모델

스테레오 영상으로부터 실제 깊이정보를 얻는 방법은 님은꼴 삼각형의 특징을 이용하며, 유도식은 아래와 같다.

$$Z/f = Y/x_l \quad (1)$$

$$Z/f = (Y-B)/x_r \quad (2)$$

여기서 Z 는 실제 깊이정보를 나타내고, f 는 초점거리, B 는 두 카메라 간의 거리를 나타낸다. x_l 과 x_r 은 3차원 공간상의 점인 X 가 각각 오른쪽과 왼쪽 카메라의 영상에 맺힌 영상좌표를 나타낸다. 그림 1은 크게 두 개의 님은꼴 삼각형 구조를 포함하고 있다. 삼각형(X, A, C_l)과 삼각형(x_l, a, C_l)이 하나의 님은꼴 삼각형이고, 삼각형(X, B, C_r)과 삼각형(x_r, b, C_r)이 또 다른 님은꼴 삼각형이다. 모든 대응하는 선의 비율이 같은 님은꼴 삼각형의 성질을 이용해 식 (1)과 식 (2)를 유도할 수 있으며, 두 식에 존재하는 인자인 Z 에 의해 식을 정리하면 식 (3)을 구할 수 있다.

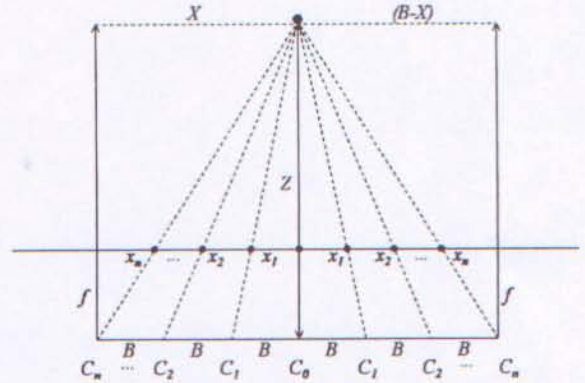
$$Z = fB/(x_l - x_r) = fB/d \quad (3)$$

여기서 d 는 두 영상의 영상 좌표 차이값인 변이값을 나타낸다. 최종적으로 스테레오 카메라의 기하학적 특성을 통해 변이값 d 를 찾을 수 있고, 실제 깊이값은 식 (3)을 이용해 계산할 수 있다. 반대로 깊이값을 알고 있을 경우 변이값 d 를 찾을 수 있다[6].

3. 다시점 깊이맵 생성 방법

3. 1 다시점 영상에서의 깊이정보

깊이맵을 생성하기 위한 가장 일반적인 구조는 스테레오 카메라를 이용하는 방법이다. 본 논문에서는 이러한 스테레오 카메라의 기하학적 구조를 다시점 카메라 구조로 확장하여 살펴본다. 그림 2는 다시점 카메라가 평행하게 배치되어 있는 구조이다.



(그림 463) 다시점 카메라의 기하학적 모델

카메라 C_0 는 참조 카메라이고, 총 $(2n+1)$ 개의 카메라가 존재하며, 모든 카메라는 동일한 간격으로 배치되어 있다. 참조 카메라로부터의 카메라 간의 거리가 증가함에 따라 n 의 값도 증가한다. 즉, 참조 카메라 C_0 에 가장 가까운 카메라는 C_1 카메라이다. 그림 2의 인자는 그림 1의 인자와 동일하다. 식 (1)부터 식 (3)은 유도한 식을 다시점 카메라 구조에서도 동일하게 적용될 수 있다. 단, 각 카메라 간의 거리가 멀어짐에 따라 카메라 간의 거리를 나타내는 B 에 떨어진 거리만큼의 값인 n 이 곱해진다. 즉, 참조 카메라로부터의 거리가 증가함에 따라 변이값이 n 배만큼 증가하게 된다.

$$Z/f = X/nx_n \quad (4)$$

$$Z/f = (X - nB)/x_n \quad (5)$$

식 (4)와 식 (5)는 스테레오 카메라 구조에서 유도한 식을 다시점 카메라 구조에 맞게 확장한 님은꼴 삼각형의 유도식이다. 각 카메라의 위치에 따라 각각 두 개의 님은꼴 삼각형 비례식을 유도할 수 있다.

$$d_n = fnB/Z \quad (6)$$

식 (6)은 동일한 3차원 공간상의 점 X 의 깊이정보 Z 에 대한 모든 다시점 카메라의 변이값 d_n 을 구하는 식이다. 이 변이값은 참조 카메라에서 획득한 영상을 기준으로 각 카메라의 화소 이동 (pixel shifting) 값을 계산한다.

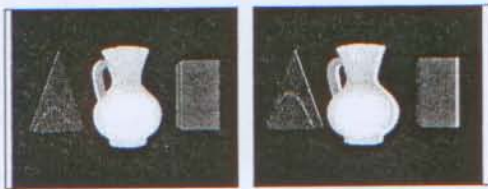
3. 2 다시점 깊이맵 생성

본 논문에서는 주어진 한 장의 깊이맵으로부터 주변 카메라와의 기하학적인 관계를 이용해 다시점 깊이맵을 생성하기 위해 식 (4)부터 식 (6)을 이용한다. 먼저 깊이

맵을 식 (6)에 적용하여 각 영상의 화소 이동 값인 변이를 계산한다. 이 변이값은 현재의 깊이값과 색상영상을 얼마만큼 이동할 것을 나타내는 정보이다. 현재 구하고자 하는 깊이맵의 카메라 위치가 참조 카메라 위치를 기준으로 왼쪽인지 오른쪽인지에 따라 깊이값과 색상정보를 왼쪽으로 이동할 것인지 오른쪽으로 이동할 것인지를 결정할 수 있다. 제안한 방법은 모든 카메라 배열이 평행하다는 가정 하에 수행되기 때문에, 모든 좌표이동은 단지 수평 방향으로만 이동하며, 수직방향에 대해서는 고려하지 않는다. 모든 시점의 깊이맵을 생성하기 위해 변이값을 찾고, 그 변이값만큼 이동하여 초기 깊이맵과 색상영상을 생성한다. 그러나 생성한 다시점 영상은 다시점 카메라를 통해 획득한 원본 색상영상과 위치가 정확히 일치하지 않는다. 이를 보정하기 참조 카메라의 현재 깊이값과 동일한 위치의 색상값 주변 영역을 블록으로 지정하고, 이 블록을 실제 원본 색상영상에서 찾는다. 그리고 찾은 위치의 좌표값을 최종 변이값으로 결정하고, 이 값을 다시 깊이값으로 변환하여 최종 깊이맵을 생성한다.

3. 3 공백 채우기 및 보간방법

깊이정보를 이용해 다시점 깊이맵을 생성하는데 있어 큰 문제점은 공백을 채우는 문제이다. 그림 3은 깊이맵을 이용해 생성한 초기 다시점 깊이맵 결과이다. 그림에서 보는 바와 같이 물체 내부영역과 주변영역에 공백이 생긴다. 공백의 발생 원인은 크게 두 가지로 나누어 생각할 수 있다. 하나는 변이값을 계산할 때 실수의 계산결과를 정수로 변환하면서 생기는 오차값에 의해 생기는 공백이고, 다른 하나는 카메라의 위치가 변하면서 보이지 않던 영역이 보이면서 생기는 공백이다.



(그림 3) 생성한 깊이맵과 공백

먼저 정수 계산에 의해 생긴 공백은 주변영역의 값을 이용해 보간(interpolation)하여 공백은 채운다. 그리고 카메라 위치 변화에 따라 생긴 공백은 카메라 위치 정보를 이용해 채운다. 즉, 현재 카메라가 참조 카메라의 오른쪽에 위치하면 공백은 물체의 왼쪽 영역에 나타난다. 반대로, 현재 카메라가 참조 카메라의 왼쪽에 위치하면 공백은 물체의 오른쪽에 나타난다. 이 공백은 배경영역에 해당하기 때문에 배경의 가장 바깥쪽 색상정보로 공백을 채운다. 식 (7)은 주변 색상정보를 이용해 공백을 채우는 식이다.

$$I_s(x, y) = \frac{1}{k} \sum_{i=0}^w \sum_{j=0}^w I(R_{(i, j)}) \quad (7)$$

4. 실험 결과 및 분석

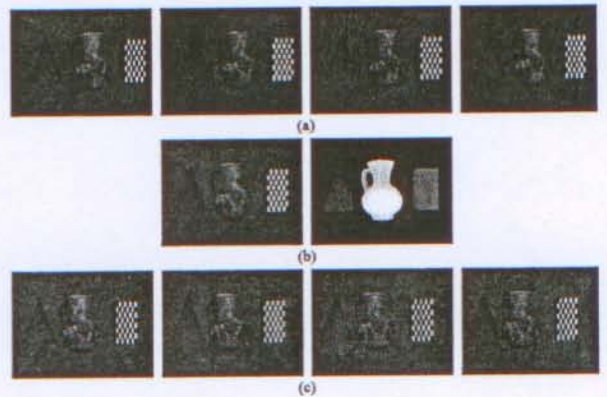
제안한 방법의 성능을 평가하기 위해, 기준 변이맵(ground truth)을 획득할 수 있는 컴퓨터 그래픽스 데이터를 사용하여 색상영상과 기준 깊이맵을 생성하여 실험하였다. 제안하는 다시점 깊이맵 방법은 Scharstein과 Szeliski이 제시한 품질 측정 방법을 통해 검증하였다[7]. 품질 측정은 기준 변이맵이 있는 테스트 영상을 사용하여 제안한 방법으로 생성한 변이맵과 기준 변이맵의 차이를 비교함으로써 이루어진다. 폐색영역(occlusion)에 대한 변이는 현재까지 신뢰할 만한 값을 생성해 낼 방법이 없기 때문에, 비폐색 영역에서만 변이를 비교한다.

$$B_O = \sum_{s \in O} (|d(s) - d_T(s)| > \delta_d) \quad (4)$$

$$B_T = \sum_{s \in T} (|d(s) - d_T(s)| > \delta_d) \quad (5)$$

$$B_D = \sum_{s \in D} (|d(s) - d_T(s)| > \delta_d) \quad (6)$$

식 (4)부터 식(6)은 기준 변이값과 한 화소 이상 떨어져 있는 화소들의 전체 영상에 대한 비율을 수치적으로 나타낸 것으로, B_T 는 텍스처가 없는 영역에서의 성능을 나타낸다. B_D 는 불연속 영역에서의 성능을 나타내고, B_O 는 스테레오 정합 방법의 전반적 성능을 나타낸다.

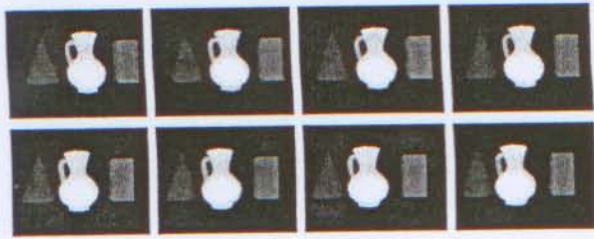


(그림 4) 입력영상: 다시점 색상영상과 한 장의 깊이맵

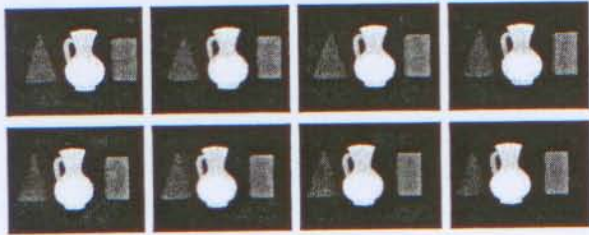
그림 4는 9대의 카메라에서 획득한 색상영상과 참조 카메라인 4번 카메라에서 획득한 깊이맵 한 장이다. 4번 카메라에서 획득한 하나의 깊이맵과 다시점 색상영상을 제안한 방법을 통해 9대 카메라의 각 시점에서의 깊이맵을 고속으로 생성한다.



(그림 5) 초기 다시점 깊이맵



(그림 6) 보간한 다시점 깊이맵



(그림 7) 원본 깊이맵



(그림 8) 최종 다시점 깊이맵

그림 5는 제안한 깊이맵 생성 방법을 통해 생성한 초기 깊이맵 결과이다. 카메라의 위치 이동과 정수 연산에 의해 물체 내부와 물체 외부영역에 공백이 발생하였다. 그림 6은 발생한 공백을 제안한 보간방법을 통해 생성한 다시점 깊이맵 결과이다. 그러나 생성한 깊이맵이 기준 변이맵과 정확히 일치하지 않는다. 생성한 색상영상과 원본 다시점 영상이 일치하지 않는 것을 보정하기 위해 참조 카메라의 현재 깊이값과 동일한 위치의 색상값 주변 영역을 블록으로 지정하고, 이 블록을 실제 원본 색상영상에서 찾아 최종 깊이맵을 생성하였다. 그림 7은 기준 다시점 깊이맵 영상이고, 그림 8은 최종 다시점 깊이맵 결과이다. 결과에서 볼 수 있듯이 생성된 결과와 기준 변이맵이 거의 같음을 확인 할 수 있다.

<표 1> 성능 평가

	B_O	B_T	B_D	PSNR
카메라 0	1.25	1.62	6.68	32.51
카메라 1	1.21	1.75	6.39	32.92
카메라 2	1.11	1.37	5.79	33.16
카메라 3	0.88	1.29	4.76	33.81
카메라 5	0.97	1.75	5.09	33.25
카메라 6	1.29	1.71	6.83	32.92
카메라 7	1.30	1.57	6.92	32.37
카메라 8	1.39	1.64	6.85	32.12
평균	1.18	1.59	6.16	32.88

표 1은 제안한 방법을 통해 생성한 다시점 깊이맵의 성능 평가와 기준 변이맵의 PSNR 결과를 보여준다. 이 표는 Scharstein과 Szeliski의 참고 결과 비교표에 근거한 것으로, 결과를 통해 알 수 있듯이 비교적 정확한 다시점 깊이맵을 생성하였으며, PSNR도 평균 32.88 dB의 좋은 결과를 보여주고 있다.

5. 결론

본 논문에서는 한 시점의 깊이맵으로부터 다시점 카메라 사이의 기하정보를 이용하여 다시점 깊이맵을 생성하는 방법을 제안하였다. 주어진 한 장의 깊이맵으로부터 주변 카메라와의 기하관계를 고려해 다른 시점의 초기 깊이맵과 색상영상을 생성하였다. 생성된 다시점 영상은 원본 영상과 위치가 정확히 일치하지 않으므로 이를 보정하고, 물체 사이의 서로 다른 시차로 인해 발생하는 공백을 카메라의 위치정보를 이용해 채워서 다시점 깊이맵을 생성하였다. 결과적으로, 빠른 시간에 다시점 깊이맵을 생성할 뿐 아니라 향상된 화질의 깊이맵을 생성할 수 있었다.

감사의 글

본 연구는 광주과학기술원 (GIST) 실감방송연구센터 (RBRC)를 통한 정보통신부 대학 IT연구센터(ITRC)와 교육인적자원부 두뇌한국21 (BK21) 정보기술사업의 지원에 의한 것입니다.

참고문헌

- [1] C. Fehn, R. de la Barre, and S. Pastoor, "Interactive 3DTV Concepts and Key Technologies," Proceedings of the IEEE, vol. 94, no. 3, pp. 524-538, March 2006.
- [2] ISO/IEC JTC1/SC29/WG11 N6051, "Call for Comments on 3DAV," October 2003.
- [3] ISO/IEC JTC1/SC29/WG11 N8944, "Preliminary FTV Model and Requirements," April 2007.
- [4] S. Kim, S. Yoon, and Y. Ho, "Realistic Broadcasting Using Multi-Modal Immersive Media," Lecture Notes in Computer Science (LNCS), vol. 3768, pp. 164-175, November 2005.
- [5] M. Okutomi and T. Kanade, "A Multiple-baseline Stereo," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 4, pp. 353-363, April 1993.
- [6] Y. Wang, J. Ostermann, and Y. Zhang, Video Processing and Communications, Prentice Hall, 2002.
- [7] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," International Journal of Computer Vision, vol. 47, no 1, April-June 2002.