

다시점 데이터의 계층적 표현에 기반한 가상 시점 생성 방법

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

Virtual View Generation Based on a Hierarchical Representation of Multiview Data

Seung-Uk Yoon and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)

요 약

최근 차세대 서비스인 삼차원 TV, 자유시점 TV, 실감방송 등에 사용될 중요한 실감형 멀티미디어로 다시점 데이터의 표현과 처리 및 부호화에 대한 연구가 활발하다. 이런 응용들의 주요 기능 중의 하나는 사용자에게 자유로운 시점의 변화를 제공하는 것이다. 이를 위해서는 임의의 카메라 배치에서 새로운 시점의 영상을 생성하는 기술이 필수적이다. 하지만 기존의 시점 생성 방식들은 규칙적이고 밀집된 카메라 배치에서만 잘 동작하는 문제점이 있다. 본 논문에서는 다시점 데이터의 계층적 표현에 기반해 규칙적이고 밀집한 카메라 배치뿐 아니라 불규칙적인 카메라 구성에서도 자유롭게 가상 시점을 생성하는 방법을 제안한다. 제안 방식은 다시점 색상 및 깊이 데이터를 영상기반 렌더링 기법의 하나인 계층적 깊이영상 형태로 표현한다. 우선 다시점 카메라 시스템에서 각 카메라 위치를 삼차원 공간상의 꼭지점으로 해석하고, 이를 여러 집합으로 분할한다. 각 집합별로 계층적 깊이영상을 생성한 후 새로운 시점이 선택되면, 이 시점에 가장 가까운 계층적 깊이영상으로부터 삼차원 워핑을 이용해 선택된 시점을 생성한다. 따라서 제안한 방식은 다양한 카메라 배치에서 기존 방식보다 손쉽게 가상 시점을 생성할 수 있다.

1. 서론

최근 고화질의 음향 및 영상 서비스에 대한 사용자의 요구가 증가하면서, 삼차원 TV (three-dimensional TV, 3DTV), 자유시점 TV (free-viewpoint TV, FTV), 실감방송 등 차세대 멀티미디어 응용들에 대한 관심이 높다. 이런 다양한 응용들은 사용자에게 실재감과 입체효과를 제공하는 것을 목적으로 하므로, 자유로운 시점의 변화를 제공하는 것이 중요한 기능 중의 하나이다. 이를 위해서는 중간 시점 복원 (intermediate view reconstruction, IVR), 시점 합성 (view synthesis) 또는 가상 시점 생성 (virtual view generation)에 관한 연구가 필수적이다.

기존의 가상 시점 생성 방법들은 규칙적이고 밀집된 카메라 배치에 크게 의존한다. 하지만 삼차원 TV나 자유시점 TV 등의 응용을 위해서는 규칙적인 카메라 구성뿐 아니라 불규칙적인 배치에서도 원하는 시점의 영상을 생성하는 기법이 필요하다.

일반적으로 카메라 사이의 거리가 가까울수록 고화질의 중간 시점을 생성할 수 있다. 이는 카메라 간격이 좁을수록 다시점 카메라에서 획득된 영상간의 선형적인 연관관계가 높기 때문이다. 하지만 자유시점을 제공하기 위해 사용자가 원하는 모든 시점에 카메라를 배치하는 것은

현실적으로 어렵다. 카메라의 개수가 증가함에 따라 데이터의 양도 선형적으로 증가하며, 카메라의 가격과 공간의 효율적인 활용도 고려해야 하기 때문이다. 따라서 이러한 요구조건을 고려해 가능한 적은 수의 카메라로부터 시각적인 오류가 적은 가상 시점을 생성하는 것이 중요하다.

본 논문에서는 이러한 문제를 해결하기 위해 다시점 색상과 깊이 데이터를 다수의 계층적 깊이영상(layered depth image, LDI)[1]으로 표현하고, 이로부터 가상 시점을 생성하는 방법을 제안한다. 계층적 깊이영상은 영상기반 렌더링(image-based rendering, IBR)[2] 기법의 하나로, 삼차원 데이터인 깊이정보를 이용해 워핑(warping)을 수행함으로써 새로운 시점의 영상을 손쉽게 생성할 수 있는 장점이 있다. 삼차원 워핑은 영상 사이의 선형적인 연관성 보다는 카메라 매개변수와 깊이정보에 의존하므로 불규칙한 카메라 배치에 더 적합하다. 또한, 다시점 영상으로부터 가상 시점을 생성하는 경우, 삼차원 기하정보를 사용하면 출력 영상의 화질을 높이고 입력 영상의 수를 줄일 수 있다[3].

따라서 제안 방식은 카메라 배치가 선형적이고 밀집된 다시점 시스템에서만 잘 동작하는 기존 방식보다 불규칙한 카메라 구성에서도 쉽게 가상 시점을 생성할 수 있다.

2. 영상기반 렌더링과 가상 시점 생성 기법

영상기반 렌더링은 다시점 영상을 입력으로 이용하여, 삼차원 공간의 임의 시점에서 영상을 생성하는 방법이다. 영상기반 렌더링 기술은 가상 시점의 영상을 생성할 때 사용하는 기하정보의 양에 따라 기하정보를 이용하지 않는 렌더링 기법, 화소간 대응 관계를 이용하는 방법, 그리고 삼차원 기하정보를 사용하는 방식으로 나뉜다[2].

기하정보를 사용하지 않는 렌더링 방법은 임의 위치에서의 빛의 세기를 매개변수를 통해 이산적인 함수로 정의하는 플레넵틱 함수(plenopic function)의 특성을 이용해 새로운 시점의 장면을 재현한다. 대표적으로 라이트 필드 렌더링(light field rendering)이 여기에 속한다. 많은 수의 입력 영상이 필수적이며, 이에 따라 처리해야 할 데이터의 양도 입력에 비례하여 증가한다.

화소간 대응관계를 이용하는 렌더링 방법은 새로운 시점의 장면을 재현하기 위해 입력 영상들의 위치적 관련성을 이용하며, 삼차원 기하정보가 직접적으로 렌더링에 이용되지는 않는다. 서로 다른 두 영상 사이의 자연스런 변화를 만들어 내는 시점 모핑(view morphing)[4] 기법이 대표적인 예이다.

삼차원 기하정보를 사용하는 렌더링 방법은 렌더링을 수행할 때 깊이정보와 같은 삼차원 정보를 사용하며, 본 논문에서 기술하는 계층적 깊이영상이나 영상기반 모델링을 통한 텍스처 매핑 방법이 이 범주에 속한다.

기존의 다양한 방법들 중에서 대표적인 가상 시점 생성 방법을 소개하면 다음과 같다. Chen[5]는 두 장의 입력 영상에 대해 화소간 대응관계를 계산하여 중간 시점을 생성하였다. 이 방식은 공간적인 상관도가 매우 높고 큰 대안 가까이 위치한 영상에 적합하다. 하지만, 합성 영상에 대한 실험결과만이 발표되어 실사 영상에 대한 검증이 필요하다. Seitz[4]가 제안한 시점 모핑은 두 영상의 대응관계를 선을 이용한 특징을 사용해 기술하였으나, 사전에 수작업을 통해 이를 설정하는 과정이 필요하다.

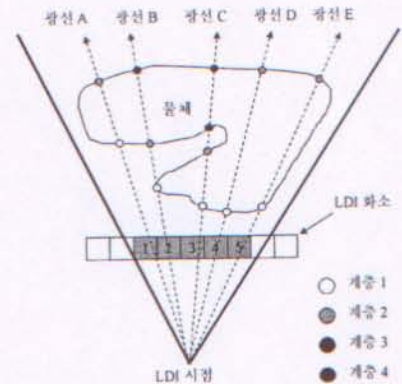
Droese[6]는 블록 매칭을 통해 계산한 변이(disparity) 정보를 이용하여 중간 시점을 생성했으나, 고정된 블록 크기를 사용하므로 물체의 윤곽선과 같이 자세한 부분에서 오류가 발생하는 단점이 있다. Lee[7]는 이 방법을 개선하여 가변 블록 크기를 사용해 화소 단위의 변이를 예측하였는데, 제안된 방법은 평행한 카메라 배치에만 적용이 가능한 문제점이 있다.

최근에 모든 다시점 데이터를 하나의 기준 시점으로 모아 계층적 깊이영상을 생성함으로써 다시점 색상 및 깊이정보를 부호화하는 방법이 제안되었다[8]. 하지만 이 방식은 가상 시점 생성보다는 원본 시점 복원과 부호화를 목적으로 한다. 또한, 하나의 계층적 깊이영상으로부터 생성할 수 있는 시야가 제한적이고, 기준 시점에서 멀어질수록 복원된 영상의 화질이 저하되는 문제가 있다. 따라서 본 논문에서는 다수의 계층적 깊이영상을 생성하고, 이를 바탕으로 선택된 시점을 생성하는 방법을 사용한다.

3. 다시점 데이터의 계층적 표현에 기반한 가상 시점 생성 방법

본 장에서는 제안한 다시점 데이터의 계층적 표현에 기반한 가상 시점 생성 기법에 대해 기술한다. 제안 방식은 계층적 깊이영상을 이용하므로 우선 이의 개념과 다시점 데이터로부터 계층적 깊이영상을 생성하는 과정을 언급한다. 그리고 제안 방식을 통해 새로운 시점을 생성하는 절차를 구체적으로 설명한다.

계층적 깊이영상은 다시점에서 획득된 색상과 깊이정보를 한 시점(LDI 시점)으로 모은 후, 이를 깊이 값에 따라 정렬해 계층적으로 표현하는 기법이다. 그림 1은 계층적 깊이영상의 개념을 나타낸다[8]



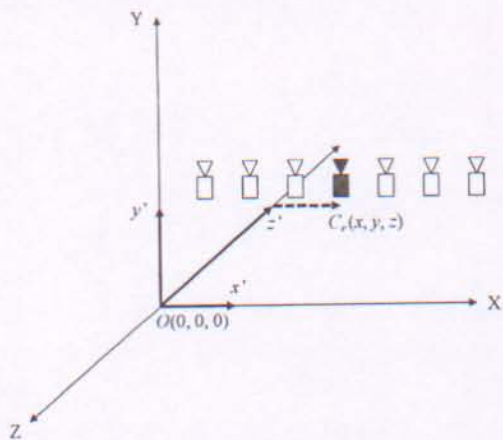
(그림 1) 계층적 깊이영상의 개념도

광선이 LDI 시점으로부터 물체를 향해 발산된다고 가정하면, 각 화소 위치에 광선과 물체가 교차하는 점의 색상과 기준 시점에서부터 교점까지의 거리를 저장할 수 있다. 하나의 광선이 물체와 만나는 교점의 수는 물체의 모양에 따라 불규칙적으로 변하므로, 각 화소 위치에 서로 다른 수의 교점이 저장된다. 각 화소는 계층적 깊이 화소(layered depth pixel, LDP)로 명명하며, 각 LDP는 서로 다른 개수의 깊이 화소(depth pixel, DP)를 포함한다. 이때, DP는 색상과 깊이정보를 모두 포함하는 포맷이다. 하지만 그림 1에 나타난 개념은 광선이 물체를 투과해야 하므로 삼차원 그래픽스 객체가 아닌 실세계의 물체에 대해서는 적용이 불가능하다.

실사 다시점 데이터로부터 계층적 깊이영상을 생성하기 위해서는 삼차원 워핑을 이용해 다시점 정보를 기준 시점으로 모으는 과정을 수행한다[8]. 이때, 삼차원 워핑은 삼차원 좌표계 및 카메라 매개변수를 정의하는 방식에 따라 다양한 식으로 수행할 수 있다. 모아진 깊이 화소들을 깊이 값에 따라 정렬하여 계층적으로 표현하면 그림 1과 같은 구조의 계층적 깊이영상을 생성할 수 있다.

앞서 설명한 계층적 깊이영상의 개념을 바탕으로 제안한 방식을 이용해 가상 시점을 생성하는 절차를 알아보자. 제안한 방법은 우선 각 카메라를 삼차원 꼭지점으로 해석하여 월드 공간에서 꼭지점의 삼차원 좌표를 계산한

다. 카메라의 외부 매개변수(extrinsic parameters)는 기준 카메라로부터 각 카메라의 상대적인 이동과 회전정보를 담고 있으므로, 삼차원 공간상의 월드 좌표계에서 기준 카메라의 위치를 설정하면 다른 카메라의 절대적인 위치도 계산할 수 있다. 그림 2는 월드 좌표계에서 기준 카메라의 위치를 정하는 예를 나타낸다.



(그림 2) 월드 좌표계에서 기준 카메라의 위치

그림 2에서 기준 카메라 C_i 의 삼차원 공간상의 위치는 다음 식으로 정의된다.

$$C(x, y, z) = O(0, 0, 0) + D(x', y', z') \quad (1)$$

여기서 O 는 월드 좌표계의 원점이며 D 는 기준 카메라가 원점 O 에서 얼마나 떨어져 있는지를 나타낸다. 이제 다른 카메라의 위치도 XYZ 좌표계 상에서 새롭게 계산할 수 있다.

다음으로 삼차원 꼭지점들을 클러스터링 (clustering) 기법을 사용해 몇 개의 집합으로 분할한다. 이 때 각 집합에 포함되는 카메라 또는 꼭지점의 개수 M 은 다음 식으로 정의된다.

$$N_{Min} < M < N_{Max} \quad (2)$$

여기서 N_{Min} 은 계층적 깊이영상을 생성하기 위해 필요한 최소의 카메라 개수이며, N_{Max} 는 총 카메라의 개수이다. 본 논문에서는 N_{Min} 을 결정하기 위해 계층적 깊이영상 생성시 몇 번째 계층까지가 유효한지를 실험하였다. 표 1은 8개의 카메라로부터 획득된 "Ballet" 데이터[9]로부터 하나의 계층적 깊이영상을 생성했을 경우, 계층별 화소의 분포를 나타낸다.

<표 1> 계층별 화소의 분포

계층	화소 분포 [%]	계층	화소 분포 [%]
1	100.0	5	15.8
2	99.0	6	5.0
3	68.2	7	1.1
4	38.4	8	0.1

표 1을 통해 세 번째 계층 이후에 화소의 분포가 50% 이하로 급격히 감소하는 것을 확인할 수 있다. 이와 같이 계층적 깊이영상으로부터 새로운 시점을 생성할 때 사용할

수 있는 유효한 계층은 앞의 몇 계층으로 제한된다. 이는 카메라의 간격과 물체에서 카메라까지의 거리에도 영향을 받는다. 또한, 다시점 데이터로부터 계층적 깊이영상을 생성하기 위해서는 적어도 세 시점의 데이터가 필요하다. 따라서 본 논문에서는 이 두 가지 요소를 고려하여 실험에 사용된 테스트 데이터에 대해 N_{Min} 을 2로 정하였다. 그리고 적어도 두 개 이상의 기준 시점에서 계층적 깊이영상을 생성하기 위해 M 은 N_{Max} 보다 작도록 정하였다.

이렇게 꼭지점들을 여러 집합으로 분할한 후에는 집합별로 다시점 색상 및 깊이정보를 모아 계층적 깊이영상을 생성한다[8]. 다음은 새로운 시점이 선택되었을 때 이 시점의 영상을 생성하기 위해, 선택된 시점의 카메라 매개변수를 계산해야 한다. 이는 계층적 깊이영상으로부터 가상 시점을 생성하는 기초인 삼차원 워핑이 카메라 매개변수와 깊이정보를 이용하기 때문이다. 우선 카메라의 내부 매개변수(intrinsic parameters)는 집합에 속한 카메라들의 매개변수를 평균하여 계산한다. 새로운 시점에서 카메라의 위치 성분(translational component)은 삼차원 좌표계 상에서 직접 확인 후, 기준 카메라로부터의 상대적인 위치를 계산한다. 그리고 회전 성분(rotational component)은 다음 식에 의해 기준 카메라와 새로운 시점에서의 카메라가 이루는 각을 계산한다[4].

$$N_n = N_r \times N_s, \theta = \cos^{-1}(N_r \cdot N_s) \quad (3)$$

여기서 N_r 은 기준 카메라의 영상 평면에 대한 법선 벡터이며, N_s 는 가상 카메라의 영상 평면에 대한 법선 벡터이다. θ 는 두 벡터가 이루는 각을 나타낸다.

마지막으로 선택된 시점으로부터 주변 꼭지점까지의 거리를 계산하여 가장 가까운 점이 속한 집합을 찾는다. 그 후, 선택된 집합에서 생성된 계층적 깊이영상으로부터 최종 시점의 영상을 렌더링하게 된다[8].

4. 실험 결과 및 분석

본 논문에서는 제안한 방법의 성능을 검증하기 위해 두 종류의 다시점 테스트 데이터를 이용하였다. 한 시퀀스는 일차원 원호 형태로 규칙적으로 배치한 8개의 카메라에서 획득한 데이터이다[9]. 다른 하나는 불규칙적으로 구성된 16개의 카메라에서 획득한 데이터이다[10].

그림 3은 규칙적인 카메라 배열에서 생성된 카메라 0번과 1번 사이의 가상 시점 영상을 나타낸다. 그림 3(a)는 기존의 방식인 하나의 계층적 깊이영상으로부터 생성한 결과이며 특히 손 부분에서 시각적인 오류가 발생했다. 그림 3(b)는 제안한 방식인 다중 계층적 깊이영상을 이용해 생성한 결과로 기존 방식에 비해 향상된 화질을 보여준다.

그림 4는 기존의 방식과 제안한 방식을 이용해 생성한 시점별 영상의 객관적인 화질을 비교한 결과이다. 하나의 계층적 깊이영상을 사용한 경우는 기준 시점인 시점 5에서 멀어질수록 복원된 영상의 PSNR이 감소함을 알 수 있다. 반면, 제안한 방식은 모든 카메라를 세 개의 집합으로 분할하고 각 집합별로 계층적 깊이영상을 생성했기 때문

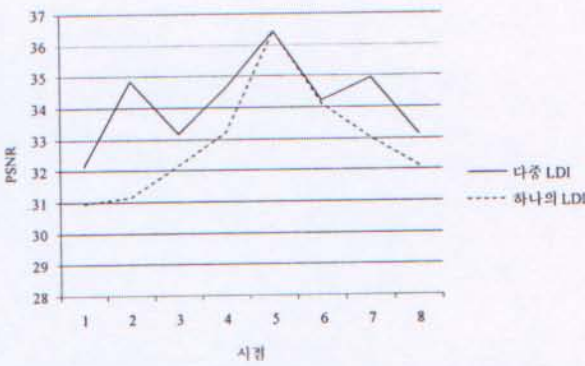
에 다른 시점에서도 비교적 높은 PSNR을 나타내었다.



(a) 하나의 계층적 깊이영상에서 생성된 중간 시점 영상



(b) 다수의 계층적 깊이영상에서 생성된 중간 시점 영상
(그림 3) 규칙적인 카메라 배치에서 생성된 가상 시점



(그림 4) 기존 방식과 제안 방식의 시점별 영상 화질 비교

마지막으로 불규칙적인 카메라 배치에서 생성한 가상 시점의 영상을 그림 5에 나타내었다. 화질의 비교를 위해 세대의 카메라(카메라 6, 7, 8)에서 계층적 깊이영상을 생성한 후, 카메라 6번과 8번의 원본 영상과 생성된 영상을 함께 배치한 결과이다. 첫 번째와 세 번째 영상이 원본이며, 나머지가 제안한 방식으로 생성한 영상이다. 원본에 비해 약간의 번짐 (blurring) 현상이 발생했지만, 주관적인 화질의 열화는 크게 없음을 알 수 있다. PSNR은 두 번째 영상의 경우 31.23 dB, 네 번째 영상은 30.52 dB이다.



(그림 5) 불규칙적인 카메라 배치에서 생성된 영상

5. 결론

본 논문에서는 다시점 색상 및 깊이 데이터를 영상기반 렌더링 기법 중 하나인 계층적 깊이영상 형태로 표현하여 가상 시점을 생성하는 방법을 제안하였다. 우선 다시점 카메라 시스템에서 각 카메라 위치를 삼차원 공간상

의 꼭지점으로 해석하고, 이를 여러 집합으로 분할한다. 각 집합별로 계층적 깊이영상을 생성한 후 새로운 시점이 선택되면, 이 시점에 가장 가까운 계층적 깊이영상으로부터 삼차원 워핑을 이용해 선택된 시점을 생성한다. 실험 결과에서 확인하였듯이 제안한 방식은 불규칙적인 카메라 배치에서도 손쉽게 가상 시점을 생성할 수 있다. 또한, 하나의 계층적 깊이영상을 이용하는 기존의 방식에 비해 향상된 화질의 가상 시점 영상을 생성하였다.

감사의 글

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

참고문헌

- [1] J. Shade, S. Gortler, L. He, and R. Szeliski, "Layered Depth Images," Proc. ACM SIGGRAPH'98, pp. 231-242, July 1998.
- [2] H. Shum, S. Kang, and S. Chan, "Survey of Image-based Representations and Compression Techniques," IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 11, pp. 1020-1037, November 2003.
- [3] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, "Unstructured Lumigraph Rendering," Proc. ACM SIGGRAPH'01, pp. 425-432, August 2001.
- [4] S. Seitz and C. Dyer, "View Morphing," Proc. ACM SIGGRAPH'96, pp. 21-30, August 1996.
- [5] S. Chen and L. Williams, "View Interpolation for Image Synthesis," Proc. ACM SIGGRAPH'93, pp. 279-288, August 1993.
- [6] M. Droege, T. Fujii, and M. Tanimoto, "Ray-space Interpolation based on Filtering in Disparity Domain," Proc. 3D Conf., pp. 213-216, June 2004.
- [7] C. Lee, K. Oh, S. Kim, and Y. Ho, "An Efficient View Interpolation Scheme and Coding Method for Multiview Video Coding," Proc. IWSSIP Conf., pp. 107-110, June 2007.
- [8] S. Yoon and Y. Ho, "Multiple Color and Depth Video Coding Using a Hierarchical Representation," IEEE Tran. Circuits Syst. Video Technol., vol. 17, no. 11, pp. 1450-1460, November 2007.
- [9] Interactive Visual Media Group at Microsoft Research, <http://research.microsoft.com/vision/InteractiveVisualMediaGroup/3DVideoDownload/>
- [10] 3D Video at ETH Zurich, <http://graphics.ethz.ch/research/3Dvideo/main.php>