

계층적 깊이맵 분해법을 이용한 3차원 장면의 표현 및 부호화 방법

김성열, 호요성
광주과학기술원 정보통신공학과
e-mail: {sykim75, hoyo}@gist.ac.kr

Representation and Compression of 3-D Scenes Using Hierarchical Decomposition of Depth Maps

Sung-Yeol Kim, Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)

요 약

본 논문은 계층적 깊이맵 분해법(hierarchical decomposition of depth maps)을 이용하여 3차원 장면을 표현하고 부호화하는 방법을 제안한다. 기존의 메쉬를 이용한 깊이맵 표현법(depth map representation using meshes)은 깊이맵으로부터 특징점(feature point)을 불규칙하게 선택하였기 때문에, 불규칙한 특징점 정보를 부호화하기 위해 새로운 3차원 정보 부호화기가 필요하였다. 결과적으로 기존의 방법은 현재의 2차원 비디오 시스템과 호환되지 않아, 3차원 장면을 사용자에게 서비스하기 위해 추가적인 비용이 부가된다. 제안하는 계층적 깊이맵 분해법은 깊이맵으로부터 특징점을 규칙적으로 추출하고, 특징점 정보를 규칙 메쉬 영상, 경계 영상, 특징점 영상으로 변환한다. 또한, 제안하는 깊이맵 분해법은 3차원 장면을 기술하기 위해 계층의 수 영상을 생성한다. 깊이맵으로부터 얻어진 분해영상들은 H.264/AVC를 이용하여 부호화되고, 3차원 장면을 메쉬로 표현한다. 제안한 방법은 규칙적인 메쉬 구조를 이용하여 3차원 장면을 표현하고 실시간으로 렌더링할 뿐 만 아니라, 기존의 깊이맵 표현법과 달리 현재의 2차원 비디오 시스템과도 호환된다.

1. 서론

몰입형 멀티미디어 서비스가 장소와 시간에 관계없이 가능해짐에 따라, 고품질 영상으로서 3차원 비디오에 대한 관심이 높아지고 있다. 깊이 영상 기반 렌더링 기법(depth image-based rendering)은 3차원 비디오를 동기화된 색상 영상과 깊이맵을 이용하여 효과적으로 표현한다[1]. 특히, 메쉬를 이용한 깊이맵 표현법(depth map representation using meshes)은 메쉬 구조를 사용하여 3차원 장면을 실시간으로 재현한다[2].



(그림 451) 메쉬를 이용한 깊이맵 표현

메쉬를 이용한 깊이맵 표현법은 깊이맵으로부터 특징점(feature point)을 추출하여 3차원 장면을 생성한다. 특징점은 3차원 표면 형태에 영향을 주는 깊이맵의 화소이다. 그림 1에서 보듯이, 메쉬를 이용한 깊이맵 표현법은 추출한 특징점 정보를 이용하여 3차원 표면을 만들고, 색

상 영상을 텍스처 정보로 간주하여 특징점에 의해 생성된 3차원 표면에 입혀 3차원 장면을 만든다.

메쉬 기반 깊이맵 표현법의 장점은 사용자가 실시간으로 3차원 비디오를 즐길 수 있게 고속으로 3차원 장면을 렌더링하는 데 있다. 하지만, 기존의 깊이맵 표현법은 특징점을 불규칙하게 선택하였기 때문에, 3차원 장면의 특징점 정보를 표현하고 부호화하는 것이 매우 어렵다.

불규칙한 특징점 정보를 다루기 위해서, Grewatsch는 깊이맵에서 추출한 특징점을 MPEG-4 3DMC를 이용하여 부호화하였다[3]. 또한, Chai는 실시간으로 3차원 장면을 렌더링하기 위해 깊이맵의 전경과 배경을 분리하여 처리하였다[4]. 최근에, Morvan은 쿼드트리 분해와 평면 근사법을 이용하여 3차원 장면 정보를 표현하였다[5].

기존의 깊이맵 표현법들은 불규칙한 특징점 정보를 다루기 위해 특별한 3차원 정보 부호화기를 설계하였다. 결과적으로, 기존의 방법은 새로운 부호화기를 현재의 비디오 시스템에 결합해야 하기 때문에, 3차원 비디오를 사용자에게 서비스하기 위해서 추가적인 비용이 필요하였다.

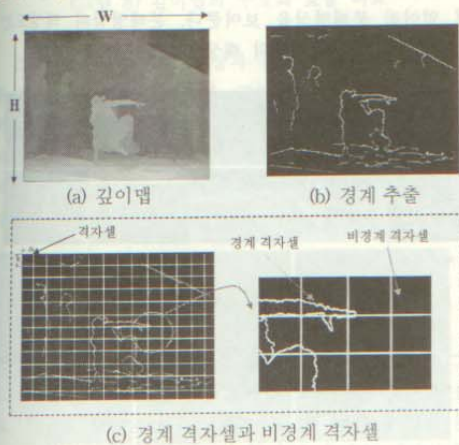
본 논문은 3차원 장면의 품질을 유지하면서 깊이맵으로부터 특징점을 규칙적으로 선택하는 계층적 깊이맵 분해법(hierarchical decomposition of depth maps)을 제

한다. 제안하는 방법은 깊이맵으로부터 특징점을 규칙적으로 선택하기 때문에, MPEG-2와 H.264/AVC와 같은 2차원 비디오 부호화기를 그대로 이용하여 3차원 장면 정보를 부호화한다.

2. 계층적 깊이맵 분해법

계층적 깊이맵 분해법은 깊이맵으로부터 특징점을 규칙적으로 추출하고, 특징점 정보를 규칙 메쉬 영상(regular mesh image), 경계 영상(boundary image), 특징점 영상(feature point image)으로 변환한다. 또한, 제안하는 깊이맵 분해법은 3차원 장면을 기술하기 위해 계층의 수 영상(number-of-layer image)을 생성한다.

계층적 깊이맵 분해법은 격자셀(grid cell)을 기본 단위로 한다. 격자셀은 2차원 비디오 부호화에서 매크로블록과 같은 영상블록으로, 격자셀의 크기는 $2^m \times 2^n$ 으로 정의한다. 제안하는 깊이맵 분해법은 우선 Sobel 필터를 깊이맵에 수직수평방향으로 적용하여 경계를 추출한다. 그런 다음, 깊이맵 영역을 격자셀의 크기에 따라 균일하게 나눈다. 그림 2는 깊이맵 영역을 분할한 모습을 보여준다.



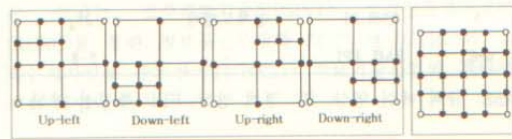
(그림 452) 격자셀에 의한 깊이맵 영역 분할

그림 2(b)는 그림 2(a)의 깊이맵에 Sobel 필터를 적용하여 경계를 추출한 결과를 보여준다. 또한, 그림 2(c)에서 보듯이, 경계 화소를 포함하는 격자셀은 경계 격자셀(edge-grid cell)이라 하고, 경계 화소를 포함하지 않는 격자셀을 비경계 격자셀(no-edge-grid cell)이라 한다.

규칙 메쉬 영상은 상응하는 깊이맵을 다운샘플링하여 얻는다. 격자셀의 크기가 $p \times q$ 일 때, 규칙 메쉬 영상은 깊이맵을 수평축 샘플링을 p 와 수직축 샘플링을 q 로 다운샘플링하여 얻는다. 다시 말하면, 규칙 메쉬 영상은 격자셀의 모서리에 위치한 4개의 화소를 모아 만든다.

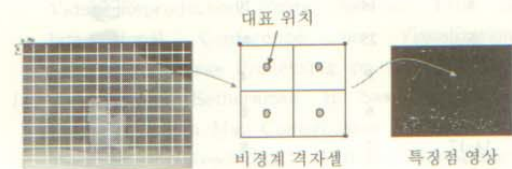
경계 영상은 경계 격자셀의 특징점을 포함한다. 경계 격자셀에서 특징점을 규칙적으로 선택하기 위해, 4개의 쿼드트리 모드와 하나의 전체모델링 모드를 사용한다. 경

계 격자셀은 수직수평방향으로 4개의 부격자셀(sub-grid cell)로 나눌 때, 경계 화소를 포함하는 부격자셀의 개수가 2개 이상이면 전체모델링 모드를 사용한다. 그렇지 않으면, 경계 화소를 포함하는 부격자셀의 위치에 따라 4개의 쿼드트리 모드 중 하나를 선택한다. 그림 3은 4개의 쿼드트리 모드와 전체모델링 모드를 보여준다. 그림 3에서의 검은 점 위치는 추출할 특징점 위치이다. 쿼드트리 모드는 10개의 특징점을 추출하고, 전체모델링 모드는 21개의 특징점을 추출한다. 경계 영상은 모드에 따라 추출한 특징점을 레스터 스캐닝 순서에 따라 저장하여 얻어진다.



(그림 453) 쿼드트리와 전체모델링 모드

특징점 영상은 비경계 격자셀의 특징점을 포함한다. 비경계 격자셀에서 영향력 있는 특징점을 선택하기 위해, 비경계 격자셀에 있는 모든 화소에 대해 최대 거리 알고리즘을 적용하여 점수를 매긴다. 모든 비경계 격자셀에서 점수가 가장 높은 화소는 첫 번째 특징점 영상에 저장된다. 또한, 모든 비경계 격자셀에서 두 번째로 점수가 높은 화소는 두 번째 특징점 영상에 저장된다. 이와 같은 과정을 반복하여 특징점 영상을 만든다.



(그림 454) 특징점 영상 생성

그림 4는 특징점 영상을 얻는 과정을 보여준다. 특징점 영상은 규칙 메쉬 영상과 경계 영상과 달리, 추출한 특징점의 x축, y축 위치 좌표를 명시해야 한다. 본 논문은 부격자셀의 중간위치를 4개의 특징점 대표위치로 정한다. 그런 다음, 비경계 격자셀에서 추출한 특징점을 4개의 대표위치 중에 하나에 대응시킨다. 결과적으로 특징점 영상은 깊이맵마다 최대 4개까지 가질 수 있다.

계층의 수 영상은 규칙 메쉬 영상, 경계 영상, 특징점 영상의 특징점 정보로 표현될 3차원 장면을 기술한다. 계층의 수 영상의 화소값은 필요한 분해영상과 격자셀의 모드 정보를 나타낸다. 표 1에서 보듯이, 계층의 수 1부터 4는 4개의 쿼드트리 모드를 나타내고, 계층의 수 5는 전체모델링 모드를 나타낸다. 또한, 계층의 수 6부터 21은 특징점 영상의 개수와 특징점의 대표위치를 나타낸다. 특

히, 계층의 수 0은 해당 비경계 격자셀을 오직 규칙 메쉬 영상에 있는 4개의 특징점만으로 표현함을 의미한다.

<표 1> 계층의 수 영상의 정보

계층의 수	분해영상	사용모드	특징점 수
0	RMI	-	4
1	RMI, BI	Up-left 쿼드 트리	14
2	RMI, BI	Down-left 쿼드 트리	14
3	RMI, BI	Up-right 쿼드 트리	14
4	RMI, BI	Down-right 쿼드 트리	14
5	RMI, BI	전체모텔링	25
6-21	RMI, FPI	-	5-8

*RMI: 규칙 메쉬 영상, BI: 경계 영상, FPI: 특징점 영상

3. 분해영상에 의한 3차원 장면 표현

계층적으로 깊이맵을 분해한 후, 3차원 장면은 격자셀 단위로 표현된다. 표 2는 계층의 수에 따른 격자셀의 3차원 표현 형태를 보여준다. 예를 들어, 계층의 수가 5이면, 해당 격자셀은 25개의 특징점과 44개의 삼각형을 이용하여 표 2의 표현 형태와 같은 3차원 표현을 생성한다.

<표 2> 격자셀의 3차원 표현 형태

계층의 수	특징점 수	삼각형 수	표면 형태
0	4	2	
1	14	20	
2	14	20	
3	14	20	
4	14	20	
5	25	44	
6-9	5	4	
10-13	6	6	
14-17	7	8	
18-21	8	10	



(a) 생성된 3차원 표면 (b) 생성된 3차원 장면

(그림 5) 최종 3차원 장면 생성

그림 5는 생성된 3차원 장면을 보여준다. 생성된 3차원 장면에서 볼 수 있는 흥미로운 사실은 3차원 표면의 각 꼭지점의 차수가 거의 6이다. 다시 말하면, 생성된 3차원 장면은 준규칙 메쉬(semi-regular mesh)의 특성을 갖

는다. 또 다른 흥미로운 사실은 규칙 메쉬 영상과 경계 영상에 저대역 필터를 적용하면, 사용자에게 넓은 3차원 시야각을 제공할 수 있다. 일반적으로 메쉬를 이용하여 3차원 장면을 생성하면, 폐색된 경계 영역에서 심각한 왜곡이 일어난다. 우리는 이를 러버 시트(rubber-sheet) 왜곡이라 부른다. 결과적으로 계층적으로 깊이맵을 분해한 후, 저대역 필터를 적용하면 러버 시트 왜곡을 완화한다.

제안한 방법으로 생성된 분해영상은 H.264/AVC를 이용하여 부호화한다. 기존의 메쉬를 이용한 깊이맵 표현법은 특징점을 불규칙적으로 선택하였기 때문에, 새로운 3차원 정보 부호화가 필요하였다. 반면에 제안한 계층적 깊이맵 분해법은 현재의 비디오 시스템과 호환된다.

분해영상을 H.264/AVC로 부호화하기 전에, 특징점 영상에 존재하는 흠이진 화소를 수평방향으로 모은다[6]. 또한, 계층의 수 영상, 규칙 메쉬 영상, 경계 영상, 특징점 영상을 수평방향으로 붙여 하나의 결합영상을 만든다. 그런 다음, 생성한 결합영상을 H.264/AVC 부호화기의 입력으로 하여 3차원 장면 정보를 부호화한다.

4. 실험 결과

제안한 방법을 평가하기 위해, Breakdancers를 실험 영상으로 선택하였다[7]. Breakdancers는 총 100 프레임이고, 영상 해상도는 1024×768이다. 본 실험에서 격자셀의 크기를 8×8로 정하였다. 그림 6은 Breakdancers의 10번째와 20번째 프레임에 계층적 깊이맵 분해법을 적용하여 얻어진 분해영상을 보여준다. 분해영상의 해상도는 128×96이었고, 결합영상의 해상도는 640×96이었다.



(그림 6) 계층적 깊이맵 분해 결과

그림 7은 계층적 깊이맵 분해법으로 생성한 3차원 장면을 보여준다. 격자셀 단위로 3차원 표면을 생성하고, 3차원 표면에 색상 정보를 입혀 3차원 장면을 성공적으로 재현하였다. 계층의 수에 따라 표 2의 일정한 3차원 표면 패턴을 적용하여 3차원 장면을 렌더링하였기 때문에, 초당 10.21 프레임을 재현하였다. 결과적으로, 제안한 방법은 실시간으로 3차원 비디오를 사용자에게 서비스할 수 있다.

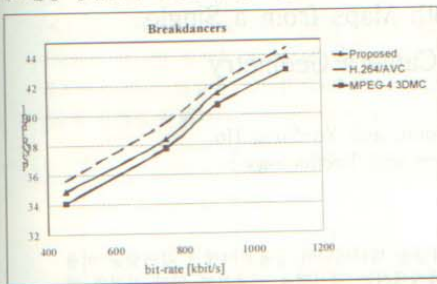


7번째 프레임 13번째 프레임 21번째 프레임

(그림 7) 3차원 장면 렌더링

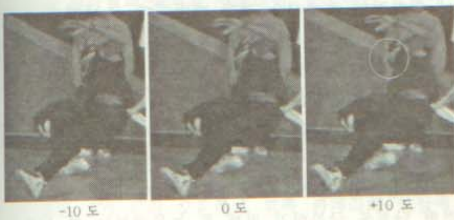
3차원 특징점 정보에 대한 부호화 효율을 평가하기 위해, 본 논문에서는 제안한 방법을 MPEG-4 3DMC를 이용한 Grewatsch 방법과 깊이맵의 모든 화소를 특징점으로

로 간주한 전체모델링 방법과 비교하였다. 전체모델링 방법은 깊이맵을 그대로 H.264/AVC를 이용하여 부호화한다. 그림 8에서 보듯이, 제안한 방법이 MPEG-4 3DMC 방법보다 약 1 dB 정도 부호화 효율이 높았다. 하지만, 제안한 방법은 전체모델링 방법에 비해 부호화 효율이 낮았다. 그 이유는 원 깊이맵의 공간적 시간적 상관도가 제안한 방법으로 생성한 분해영상보다 더 높았기 때문이다. 그러나 제안한 방법은 3차원 장면에 대한 렌더링 속도가 전체모델링 방법보다 약 20배 정도 빨랐다.

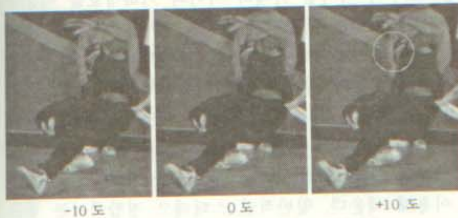


(그림 8) 깊이맵의 부호화 효율 비교

그림 9는 규칙 메쉬 영상과 경계 영상에 저대역 필터를 적용한 결과를 보여준다. 그림 9(a)에서 보듯이, 저대역 필터를 적용하기 전에는 러버 슈트 왜곡 때문에 사용자에 넓은 시야각을 제공하기 어렵다. 하지만, 그림 9(b)와 같이, 저대역 필터를 깊이맵의 경계 영역에 적용하여 러버 슈트 왜곡을 완화하였다.



(a) 저대역 필터 적용 전



(b) 저대역 필터 적용 후

(그림 9) 다시점 영상 생성

5. 결론

본 논문은 계층적 깊이맵 분해법을 이용하여 3차원 장

면을 표현하고 부호화하는 방법을 제안하였다. 기존의 메쉬 기반 깊이맵 표현법과 달리, 제안한 방법은 깊이맵으로부터 추출한 특징점 정보를 부호화하기 위해 H.264/AVC를 이용하였고, 기존의 메쉬 기반 깊이맵 표현법에 비해 약 1 dB 정도의 부호화 효율을 높였다. 또한, 메쉬 기반 표현법을 이용하여 연속적인 3차원 장면을 실시간으로 렌더링할 수 있었다. 제안한 방법은 3차원 TV와 같은 3차원 멀티미디어 응용에 활용될 것으로 기대된다.

감사의 글

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

참고문헌

- [1] A. Ignatenko and A. Konushin, "A Framework for Depth Image-based Modeling and Rendering," Proc. of International Conference on Computer Graphics and Vision, pp. 169-172, 2003.
- [2] D. Farin, R. Peerlings, and P.H.N. de With, "Depth-image Representation Employing Meshes for Intermediate-view Rendering and Coding," Proc. of 3DTV Conference, paper no. 116, 2007.
- [3] S. Grewatsch and E. Müller, "Fast Mesh-Based Coding of Depth Map Sequences for Efficient 3D Video Reproduction Using OpenGL," Proc. of International Conference on Visualization, Imaging and Image Processing, pp. 66-71, 2005.
- [4] B. Chai, S. Sethuraman, H. Sawhey, and P. Hatrack, "Depth Map Compression for Real-time View-based Rendering," Pattern Recognition Letters, vol. 25, no. 7, pp. 755-766, 2004.
- [5] Y. Morvan, D. Farin, and P.H.N. de With, "Novel Coding Technique for Depth Images Using Quadtree Decomposition and Plane Approximation," Proc. of Visual Communications and Image Processing, vol. 5960, pp. 1187-1194, 2005.
- [6] S.Y. Kim and Y.S. Ho, "Mesh-based Depth Coding for 3D Video Using Hierarchical Decomposition of Depth Maps," Proc. of International Conference on Image Processing, pp. 117-120, 2007.
- [7] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality Video View Interpolation Using a Layered Representation," SIGGRAPH, pp. 600-608, 2004.