

# 깊이정보를 포함한 다시점 비디오로부터 계층적 깊이영상 생성 및 부호화 기법

윤승욱\*, 이은경\*, 김성열\*, 호요성\*, 윤국진\*\*, 김대회\*\*, 허남호\*\*, 이수인\*\*  
\*광주과학기술원 정보통신공학과  
\*\*한국전자통신연구원

## Generation and Coding of Layered Depth Images for Multi-view Video Representation with Depth Information

Seung-Uk Yoon\*, Eun-Kyung Lee\*, Sung-Yeol Kim\*, Yo-Sung Ho\*,  
Kugjin Yun\*\*, Daehee Kim\*\*, Namho Hur\*\*, and Soo-In Lee\*\*  
\*Gwanju Institute of Science and Technology (GIST)  
E-mail : {suyoon, ekle78, sykim75, hoyo}@gist.ac.kr  
\*\*Electronics and Telecommunications Research Institute (ETRI)  
E-mail : {kjyun, daeheckim, namho, silee}@etri.re.kr

### Abstract

The multi-view video is a collection of multiple videos capturing the same scene at different viewpoints. The multi-view video can be used in various applications, including free viewpoint TV and three-dimensional TV. Since the data size of the multi-view video linearly increases as the number of cameras, it is necessary to compress multi-view video data for efficient storage and transmission. The multi-view video can be coded using the concept of the layered depth image (LDI). In this paper, we describe a procedure to generate LDI from the natural multi-view video and present a method to encode multi-view video using the concept of LDI.

### I. 서론

다시점 비디오는 한 장면을 여러 시점에서 다수의 카메라로 촬영한 것으로, 3차원 TV (3DTV), 자유시점 TV (Free viewpoint TV, FTV) 등의 응용분야에서 사용될 수 있다. 다시점 비디오는 여러 시점에서 한 장면을 획득하고 이를 이용하여 사용자에게 원하는 시점의 영상을 제공하는 것을 목적으로 한다. 그러나 카메라 수만

큼의 비디오가 존재하므로 데이터양이 매우 많아서 이를 효과적으로 저장하고 전송하기 위해서는 부호화가 필수적이다. 기존의 시공간상의 상관도를 이용하는 부호화 기법과는 달리 영상기반 렌더링 기법의 하나인 계층적 깊이영상(Layered Depth Image, LDI)의 개념을 이용하면 다시점 비디오를 효과적으로 부호화할 수 있다.

영상기반 렌더링은 여러 시점의 2차원 영상을 이용하여 3차원 공간의 임의 시점에서의 영상을 생성한다. 이러한 접근 방식은 2차원 영상을 입력으로 사용하므로 생성하려는 영상의 복잡도와 무관하다. 또한, 원하는 장면을 생성하기 위해 복잡한 3차원 모델을 만드는 것보다 영상이나 사진을 얻는 것이 더 쉽다. 최근에는 이러한 영상기반 렌더링 기법 중에서 계층적 깊이영상에 관한 연구가 주목을 받고 있다. 계층적 깊이영상은 여러 시점의 깊이영상을 합성하여 하나의 데이터 구조로 만든 것으로, 다중 깊이정보와 워핑 함수를 사용하여 임의 시점의 영상을 생성할 수 있다. 이러한 기능은 다시점 비디오의 목적과 유사하므로 영상기반 렌더링 기법을 다시점 비디오 부호화에 적용할 수 있다.

본 논문에서는 영상기반 렌더링 기법인 계층적 깊이영상의 개념을 이용하여 3차원 합성 모델이 아닌 실사영상으로부터 계층적 깊이영상을 생성하고, 이를 이용하여 다시점 비디오를 부호화하는 방법을 제안한다.

본 연구는 광주과학기술원(GIST) 실감방송연구센터(RBRC)를 통한 정보통신부대학 IT 연구센터(ITRC)와 교육인적자원부 두뇌한국 21(BK21) 정보기술사업의 지원으로 이루어졌음.

## II. 다시점 비디오 부호화

다시점 비디오는 여러 대의 카메라를 이용하여 한 장면을 찍은 다중 비디오들을 말한다. 다시점 비디오를 획득하면 사용자의 시점에 맞는 영상을 서비스 할 수 있다. 이러한 다시점 비디오를 이용할 수 있는 응용 분야로는 자유시점 TV, 3DTV, 감시 카메라 영상 분야 (surveillance), 파노라믹 영상 응용 분야 등이 있다.

이와 관련하여 ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group(MPEG)에서는 다시점 비디오 부호화의 필요성을 인식하여, 2001년 12월부터 새로운 3DAV 부호화 기술의 표준화 작업을 준비해 오고 있다. 최근에는 이 분야의 연구가 더욱 활발히 진행되어 2004년 10월에 다시점 비디오 부호화 기술에 대한 Call for Evidence(CfE)[1]가 발행되었고, 이에 대한 응답으로 2005년 1월에는 다양한 다시점 비디오 부호화 알고리즘에 대한 결과가 수집되어 분석되었다. 2005년 7월에 CfP[2]가 배포되었다. CfE에 선택된 테스트 시퀀스의 특징을 표1에 기술했다[2].

표 1. CfP 를 위해 선택된 테스트 시퀀스의 특성

제공	시퀀스	특성	카메라 개수	카메라 구성
MERL	Ballroom	VGA 25fps	8	1 차원 평행
	Exit	VGA 25fps	8	1 차원 평행
KDDI	Race1	VGA 25fps	8	1 차원 평행
	Flamenco2	VGA 25fps	5	2 차원 평행
HHI	Uli	VGA 25fps	8	1 차원 평행, 수렴
MSR	Breakdancers	VGA 25fps	8	1 차원 원호
Nagoya Univ.	Rena	VGA 25fps	100	1 차원 평행
	Akko&Kayo	VGA 25fps	100	2 차원 배열

지금까지 제안된 다시점 비디오 부호화 방법을 살펴보면 시공간상의 상관도를 이용하는 방법이 대부분이다. 시공간상의 상관도를 이용하는 알고리즘은 Group of GOP(GoGOP)와 공용 움직임 벡터 메모리를 이용한 방법[3], 2차원 직접 모드 및 조명 보상을 이용한 방법[4] 등이 있다. 이러한 방식들의 공통점은 기존의 2차원 비디오 부호화 방식을 확장하여, 시간 축뿐만 아니라 각

카메라의 영상 사이에서도 상관관계를 고려하였다는 점이다. 이와는 달리 본 논문에서 제안한 계층적 깊이영상 기반의 다시점 비디오 부호화 방식은 영상기반 렌더링의 특징을 이용하여 하나의 통합된 LDI 영상을 만들어 이를 부호화하고, 복원한 뒤에 새로운 시점의 영상을 손쉽게 생성할 수 있는 장점이 있다.

## III. 계층적 깊이영상의 개념 및 생성

### 3.1 계층적 깊이영상의 개념

계층적 깊이영상은 복잡한 기하정보를 갖는 3차원 물체나 장면을 영상기반 렌더링 기법을 이용하여 표현하는 방법 중 하나이다. 다각형 메쉬를 사용해서 모델을 표현하는 방식과 달리, 계층적 깊이영상은 여러 시점에서 얻은 다수의 깊이영상을 합성하여 하나의 데이터 구조를 생성한다. 따라서 2차원 영상과는 달리 각 계층적 깊이영상 화소는 색상정보 외에 화소와 카메라 사이의 거리를 나타내는 깊이정보와 렌더링을 지원하는 추가적인 특성정보를 담고 있다. 또한, 계층적 깊이영상은 각 화소의 위치마다 다수의 계층을 가지므로, 이 정보를 이용하면 새로운 카메라 위치에서 3차원 물체를 효과적으로 렌더링할 수 있다[5].

### 3.2 실사영상으로부터 계층적 깊이영상 생성

실사영상으로부터 계층적 깊이영상을 생성하기 위해서는 다수의 깊이영상을 워핑하는 방식을 사용한다 [5][6]. 일반적으로 실사 다시점 비디오로부터 계층적 깊이영상을 생성하기 위해서 변이 (disparity) 정보를 예측하고 이로부터 깊이영상을 획득하는 과정이 필요하다. 그러나 변이예측을 통한 깊이영상 생성은 정확도 측면에서 부족한 점이 많다. 이러한 점을 고려하여 본 논문에서는 MSR에서 제공하는 깊이영상과 카메라 매개변수를 이용하여 계층적 깊이영상을 생성하였다.

3차원 모델로부터 계층적 깊이영상을 생성할 때 사용되는 카메라 행렬은 다음 식에 의해 계산된다[5].

$$C_1 = V_1 \cdot P_1 \cdot A_1, \quad C_2 = V_2 \cdot P_2 \cdot A_2, \quad T_{1,2} = C_2 \cdot C_1^{-1} \quad (1)$$

여기서  $C_1$ 과  $C_2$ 는 시점 1과 2에서의 카메라 행렬이며,  $V$ 는 viewport 행렬,  $P$ 는 projection 행렬,  $A$ 는 affine 행렬을 나타낸다. 이때, 에서 바라본 영상 평면의 한 점을

의 위치로 이동하는 변환행렬이 이다. 따라서 다음 식에 의해 카메라 시점 1에서 본 영상을 카메라 시점 2에서 보는 영상으로 워핑할 수 있다[5].

$$T_{1,2} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_2 \cdot w_2 \\ y_2 \cdot w_2 \\ z_2 \cdot w_2 \\ w_2 \end{bmatrix} = T_{1,2} \begin{bmatrix} x_1 \\ y_1 \\ 0 \\ 1 \end{bmatrix} + z_1 \cdot T_{1,2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = start + z_1 \cdot depth \quad (2)$$

실사 영상에 식 (1)을 적용할 경우에는 V와 P, 그리고 A 행렬을 유추해야 하나, 그 의미를 정확히 파악하기가 어렵다. 따라서 본 논문에서는 카메라 행렬을 각 구성요소 행렬로부터 유추하지 않고, 주어진 카메라 매개변수로부터 새롭게 계산하였다. MSR 데이터는 3x4 affine 행렬을 포함하며, 3차원 워핑에 사용되는 카메라 행렬  $\dot{C}$ 는 다음 식에 의해 계산된다.

$$\dot{C}_1 = \dot{A}_1 \cdot \dot{E}_1, \quad \dot{C}_2 = \dot{A}_2 \cdot \dot{E}_2, \quad \dot{T}_{1,2} = \dot{C}_2 \cdot \dot{C}_1^{-1} \quad (3)$$

$$\dot{A} = \begin{bmatrix} -f_{s_x} & \theta & t_x \\ 0 & -f_{s_y} & t_y \\ 0 & 0 & 1 \end{bmatrix}, \quad \dot{E} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \end{bmatrix} \quad (4)$$

여기서  $f_s$ 는 초점거리,  $S_x, S_y$ 는 크기 성분,  $t_x, t_y$ 는 카메라 중심의 위치,  $\theta$ 는 비틀어짐 (skew) 정도, R은 회전행렬, T는 이동행렬이다. 이때,  $\dot{C}$ 은  $\dot{A} \cdot \dot{E}$ 에 [0 0 0 1]인 4번째 행을 추가하여 생성된다.

#### IV. 계층적 깊이영상을 이용한 다시점 비디오 부호화 구조

다시점 비디오는 여러 시점에서 한 장면을 획득하고 이를 이용하여 사용자에게 원하는 시점의 영상을 제공하는 것을 목적으로 한다. 이러한 개념은 영상기반 렌더링 기법이 제공하는 중요한 기능 중의 하나인 새로운 시점 생성과 유사하다. 따라서 본 논문에서는 영상기반 렌더링 기법 중의 하나인 계층적 깊이영상을 이용한 다시점 비디오 부호화 구조를 사용한다[6].

우선 8대의 카메라로부터 획득한 깊이영상을 기준이 되는 계층적 깊이영상 시점으로 워핑한다. 각 카메라에서 획득한 비디오의 첫 프레임을 따로 살펴보면, 각 프레임은 색상과 깊이정보를 포함하고 있다. 따라서 첫 번째 프레임의 색상과 깊이정보를 이용하여 워핑을 수행하면 계층적 깊이영상 시퀀스의 첫 번째 프레임을 생성할 수 있다. 두 번째 색상과 깊이정보를 이용하여 같은 방식으로 워핑하면 역시 두 번째 계층적 깊이영상

프레임을 생성할 수 있다. 이때, 깊이 값을 비교하여 계층적 깊이영상을 생성하는 중간 단계에서 정보 손실이 발생하기 때문에 이를 보상해 주는 부분이 그림에 포함되어 있다.

부호화를 수행할 때는 생성된 계층적 깊이영상 프레임이 일반적인 2차원 영상형태가 아닌 중간에 화소가 비는 형태를 취하기 때문에, H.264/AVC를 그대로 적용할 수가 없다. 따라서 데이터 모으기 및 패딩을 통해 H.264/AVC에 지원하는 형태로 생성된 LDI 프레임을 변환해 주는 과정이 필요하다.

한편, 계층적 깊이영상이 제공하는 기본 기능 중의 하나가 자유로운 임의 시점 영상을 생성하는 것이므로, 생성된 계층적 깊이영상으로부터 역으로 워핑을 수행하면 원래의 다시점 데이터를 다시 생성할 수 있다. 이와 같이 다시점 비디오를 계층적 깊이영상으로, 반대로 계층적 깊이영상을 다시점 비디오로 변환할 수 있으므로, 계층적 깊이영상의 개념을 이용하면 새로운 방식으로 다시점 비디오를 부호화할 수 있다.

#### V. 실험 결과 및 분석

본 논문에서는 Microsoft Research(MSR)에서 제공하는 8시점 다시점 비디오 테스트 데이터를 사용하였다[7]. MSR 다시점 비디오 데이터는 카메라별로 색상 및 깊이영상을 제공하며, 각 카메라에 대한 카메라 매개변수도 포함한다. 제안한 방식으로 3차원 워핑을 수행한 하여 다시점 비디오의 프레임별로 계층적 깊이영상 프레임을 생성하였다. 각 프레임은 색상과 깊이영상을 포함하고 있으므로, 첫 번째 8 프레임으로 계층적 깊이영상 프레임을 생성할 경우 8장의 색상영상과 8장의 깊이영상이 사용된다. 그림 1은 이렇게 생성된 계층적 깊이영상의 화소 분포를 각 계층별 영상으로 표현하고 있다[8][9]. 후미 계층으로 갈수록 화소가 급격히 감소하는 것을 확인할 수 있다.

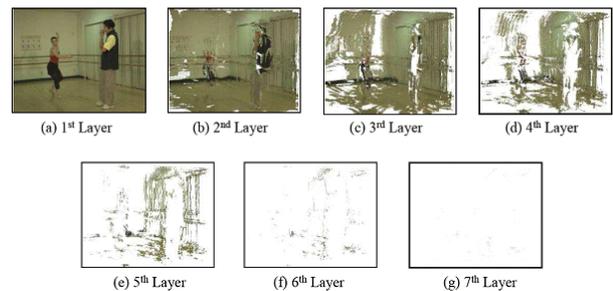


그림 1. 생성된 계층적 깊이영상의 화소 분포

표 2는 이렇게 변환한 계층적 깊이영상 프레임과 8 시점 다시점 영상 데이터를 비교한 결과이다. 표3에서 8 시점 영상의 합계는 색상영상 8장과 깊이영상 8장의 데이터의 합계를 나타낸다. 이와 같이 부호화 과정 없이도 다시점 비디오를 계층적 깊이영상으로 변환하면 데이터양이 많이 줄어드는 것을 확인하였다[6].

부호화를 위해 본 논문에서는 일차적으로 패딩을 수행하였다. 즉, 빈 공간을 영상에서 사용되지 않은 한 가지 색상으로 채운 뒤 이를 부호화 하였다. 그러나 이런 방식을 사용할 경우 계층적 깊이영상의 특성인 후미 계층으로 갈수록 화소가 없어지는 성질을 전혀 사용하지 못하게 되는 단점이 있다. 따라서 단순한 패딩은 부호화 효율을 향상시키지 못하고 오히려 더 많은 데이터가 부호화에 사용된다. 이의 해결을 위해서는 빈 데이터를 없애고 존재하는 화소들만 모아서 처리하는 데이터 모으기가 필수적이다. 본 논문에서는 수평방향으로 데이터 모으기를 수행하였다. 데이터 모으기를 수행한 후에도 생기는 빈 곳은 마지막 화소 값으로 패딩을 수행한 후 H.264/AVC를 적용하여 부호화 하였다.

표 2. 다시점 영상과 계층적 깊이영상의 크기 비교 (kB)

Ballet	첫 번째 프레임	두 번째 프레임
8 시점 영상의 합계 (색상 + 깊이)	25,165.9	25,165.9
LDI 프레임	14,078.0	14,061.6
8 시점 영상을 H.264 부호화 (색상 + 깊이)	134.4	149.5
제안한 방식	159.7	168.7
Breakdancers	첫 번째 프레임	두 번째 프레임
8 시점 영상의 합계 (색상 + 깊이)	25,165.9	25,165.9
LDI 프레임	12,726.6	12,689.7
8 시점 영상을 H.264 부호화 (색상 + 깊이)	165.9	160.6
제안한 방식	155.3	151.9

표 2 에서 확인할 수 있듯이, Ballet 시퀀스의 경우는 데이터양이 조금 증가하였으며, Breakdancers 시퀀스의 경우는 제안한 방식을 사용할 경우 데이터양이 감소하였다. 이의 원인으로 Ballet 시퀀스의 깊이 값 분포가 Breakdancers 시퀀스보다 크다는 사실을 들 수 있으며, 깊이 값 분포와 부호화 효율에 관한 깊이 있는 연구가 수행되어야 할 필요가 있다.

## VI. 결론

본 논문에서는 계층적 깊이영상의 개념을 다시점 비디오 부호화에 적용하기 위하여 3차원 합성 모델이 아닌 실사 다시점 비디오로부터 계층적 깊이영상을 생성하였다. 또한, 다시점 비디오와 계층적 깊이영상간의 변환을 통해 기존의 시공간적 예측 기반 부호화 방식이 아닌 새로운 다시점 비디오 부호화 구조를 제안하였다.

## 참고문헌

- [1] ISO/IEC JTC 1/SC 29/WG 11/N6720, "Call for Evidence on Multi-view Video Coding," October 2004.
- [2] ISO/IEC JTC 1/SC 29/WG 11/N7327, "Call for Proposals on Multi-view Video Coding," July 2005.
- [3] ISO/IEC JTC1/SC29/WG11 m11570, "Multi-view Video Coding using Shared Picture Memory and Shared Vector Memory," January 2005.
- [4] ISO/IEC JTC1/SC29/WG11 m11588, "Multi-view Video Coding using Illumination Change-adaptive Motion Estimation/Motion Compensation and 2D Direct Mode," January 2005.
- [5] J. Shade, S.J. Gortler, and R. Szeliski, "Layered Depth Image," SIGGRAPH, pp. 291-298, July 1998.
- [6] 윤승욱, 이은경, 김성열, 호요성, "계층적 깊이영상 생성 및 이를 이용한 다시점 비디오 부호화 구조," 대한전자공학회 컴퓨터소사이어티 하계종합학술대회 논문집, 제28권, pp. 63-66, 2005.
- [7] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality Video View Interpolation using a Layered Representation," SIGGRAPH, pp. 600-608, August 2004.
- [8] ISO/IEC JTC1/SC29/WG11 m12278, "Intermediate Result on Multi-view Video Coding using Layered Depth Image," July 2005.
- [9] S.U. Yoon, E.K. Lee, S.Y. Kim, and Y.S. Ho, "A Framework for Multi-view Video Coding using Layered Depth Images", Lecture Notes in Computer Science (LNCS), 3767, pp. 431-442, 2005.