

# 다시점 동영상의 영상기반 표현, 처리 및 복원기법

윤승욱\* · 호요성\*

## 1. 서론

우리는 정보혁명과 디지털 시대로 대변되는 21 세기에 살고 있다. 이러한 디지털 시대의 핵심은 여러 가지 형태의 멀티미디어의 등장이라고 할 수 있다. 삼차원 컴퓨터 그래픽스 영상을 비롯해서 파노라마 영상, 삼차원 음향과 같은 멀티미디어의 출현이 우리의 삶의 질을 향상시키고 있다. 또한, 멀티미디어를 이용한 다양한 응용분야와 고품질 콘텐츠가 나타나게 되었다. 이러한 디지털 시대에 사용자의 시청 실감을 한층 더 높이기 위해 요구되는 기술 중의 하나가 삼차원 오디오 비주얼 (three-dimensional audio visual, 3DAV) 기술이다. 그 중에서도 다시점 동영상(multi-view video)의 표현, 처리, 압축, 재현은 삼차원 오디오 비주얼 기술의 핵심 요소로서 중요한 역할을 할 것이다.

다시점 동영상은 여러 시점에서 다수의 카메라로 한 장면을 촬영한 다중 동영상의 집합이며, 획득된 정보를 이용하여 사용자에게 원하는 시점의 영상을 제공하는 것을 주요 목적으로 한다. 응용 분야로는 자유시점 동영상 (free viewpoint video, FVV), 자유시점 TV (free viewpoint TV, FTV),

삼차원 TV, 실감형 방송, 감시 카메라 영상 (surveillance), 파노라믹 영상, 교육, 홈 엔터테인먼트 등이 있다. 그러나 다시점 동영상은 카메라 수에 비례하여 데이터양이 늘어나기 때문에, 이를 다양한 응용분야에서 효과적으로 사용하기 위해서는 다시점 동영상의 효율적인 표현, 처리, 부호화, 재생 기법을 개발하는 것이 필요하다.

한편, 영상기반 렌더링(image-based rendering)은 여러 시점의 이차원 영상을 이용하여 삼차원 공간의 임의 시점에서의 영상을 생성하는 기술이다. 이러한 접근 방식은 이차원 영상을 입력으로 사용하므로 생성하려는 화면의 복잡도와 무관하다. 또한, 많은 경우에 원하는 장면을 생성하기 위해 복잡한 삼차원 모델을 만드는 것보다 영상이나 사진을 얻는 것이 더 쉽다. 최근에는 이러한 영상기반 렌더링 기법을 이용하여 동영상으로부터 삼차원 정보를 추출하고, 새로운 시점의 영상을 생성하는 연구가 주목을 받고 있다. 그 중에서도 계층적 깊이영상 (layered depth image, LDI) 기법은 여러 시점에서 획득한 색상 및 깊이영상을 합성하여 하나의 데이터 구조로 만들고, 다중 깊이정보와 삼차원 워핑(warping) 함수를 사용하여 임의 시점의 영상을 손쉽게 생성할 수 있는 장점이 있다. 이러한 기능은 다시점 동영상의 목적과 유사하므로 영상기반 렌더링 기법을 이용하여 다시점 동영상을 표현할 수 있다.

\* 광주과학기술원

※ 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업 (IITA-2005-C1090-0502-0022)인 광주과학기술원 실감방송연구센터의 연구결과로 수행되었음.

본 논문에서는 영상기반 렌더링 기법 중 하나인 계층적 깊이영상(layered depth image, LDI)의 개념을 이용하여 실사 다시점 동영상을 효과적으로 표현하여 처리하고, 이를 이용하여 다시점 영상을 복원하는 방법을 제안한다.

본 논문의 2장에서는 다시점 동영상에 관련된 기술의 개발 현황을 소개하고, 3장에서는 영상기반 렌더링 기술을 살펴본다. 4장에서는 계층적 깊이영상의 개념을 이용한 다시점 동영상 표현, 부호화, 그리고 다시점 복원기법에 대하여 설명한다. 그리고, 5장에서는 제안한 방법을 이용하여 실험한 결과를 보여주고, 6장에서 본 논문의 결론을 맺는다.

## 2. 다시점 동영상 기술 개발 현황

다시점 동영상의 가장 현실성 있는 응용분야 중 하나가 삼차원 TV이므로, 관련 기술의 개발이 삼차원 TV 또는 입체 TV의 연구와 같이 이루어지는 경우가 많다. 유럽에서는 Advanced Three-Dimensional Television System Technologies (ATTEST) 과제를 통해 2002년 3월부터 2년 동안 삼차원 TV에 관한 기초기술을 연구하였으며, 2004년 9월부터는 20여개 기관이 컨소시엄을 구성하여 3DTV (Three-dimensional Television) 과제를 수행하고 있다. 3DTV 과제는 삼차원 장면의 획득, 표현, 부호화, 전송, 디스플레이까지를 모두 포함하고 있으며, 다시점 동영상을 획득하는 부분도 주요 연구 분야 중의 하나이다. 미국에서는 NASA에서 주로 삼차원 영상과 관련된 연구를 수행하고 있으며, MIT에서는 다시점 동영상 및 홀로그래픽 디스플레이 기술을 연구하고 있다. 일본에서도 다시점 동영상 과제를 진행하고 있으며, 초다시점 삼차원 TV 기술개발에 많은 투자를 하고 있다. 일본의 경우, 이미 FTV 시험방송을 할

정도로 다시점 동영상에 대한 관심이 매우 크다.

최근에는 다시점 동영상의 부호화에 대한 관심이 높고, 이와 관련하여 일본의 나고야 대학, NTT, 독일의Fraunhofer Heinrich-Hertz-Institut (HHI) 연구소, 미국에 위치한 MitsubishiElectricResearch Laboratories (MERL), Thomson 등에서 연구 개발이 활발하다. 특히, 일본의 나고야 대학에서는 100대의 카메라를 이용하여 다시점 동영상을 획득하는 세 가지 종류의 시스템을 개발했다. HHI, MERL, Microsoft Research 에서도 8대의 카메라로부터 다시점 동영상을 획득하는 시스템을 구축하고, 이를 이용하여 다시점 동영상 테스트 데이터를 제작하여 배포하고 있다.

국내의 경우, ETRI에서 2002년 월드컵에서 삼차원 TV 시범방송 서비스를 제공하였다. KIST는 다시점 디스플레이 장치를 연구하고 있고, 삼성전자와 LG전자에서도 삼차원 카메라, 삼차원 TV, 안경식 스테레오 LCD 모니터를 연구하고 있다. 그 밖에, 광주과학기술원, 강원대, 광운대, 세종대, 연세대 등 일부 대학교에서 다시점 영상 획득, 압축, 전송기술을 연구하고 있다. 이와 같이, 국내의 일부 대학, 연구소, 산업체에서 다시점 동영상에 대한 연구가 진행되고 있으며, 이에 대한 관심이 급속히 높아지고 있다.

이와 같은 기술적인 흐름에 발맞춰 ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group(MPEG)에서도 다시점 동영상 부호화의 필요성을 인정하여, 2001년 12월부터 새로운 3DAV 부호화 기술의 표준화 활동을 준비해 오고 있다. 최근에 이 분야의 연구가 활발히 진행되어 2004년 8월에는 기초적인 성능 실험을 위한 다시점 동영상 테스트 데이터가 제공되었으며, 2004년 10월에 다시점 동영상 부호화에 대한 Call for Evidence(CfE)가 발행되었다[1]. 2005년 1월에는 CfE에 대한 응답으로 다양한 다시점 동영상 부호

화 방식에 대한 결과를 수집하여 분석하였다[2]. 2005년 7월에 Call for Proposal(CfP)이 배포되었고[3], 2006년 1월에는 CfP에 대한 응답들이 평가되어[4] 다시점 동영상에 대한 기본 부호화 방식과 참조 소프트웨어가 선정되었다. 현재는 주요 부호화 기법들에 대한 Core Experiment(CE)가 진행 중이며[5], 향후 2년 내에 다시점 동영상 부호화의 표준화 작업이 완료될 예정이다.

### 3. 영상 렌더링 기술

#### 3.1 영상기반 렌더링 기법의 분류

영상기반 렌더링은 여러 시점의 이차원 영상을 이용하여 삼차원 공간의 임의 시점에서의 영상을 생성하는 방법이다. 이러한 접근 방식은 이차원 영상을 입력으로 사용하기 때문에 생성하려는 장면의 복잡도와 무관하다. 또한, 많은 경우에 원하는 장면을 생성하기 위해 복잡한 삼차원 모델을 만드는 것보다 영상이나 사진을 얻는 것이 더 쉽다. 이런 이유로 최근에 영상기반 렌더링은 고전적인 모델기반 렌더링의 대체 기술로 각광을 받고 있다.

영상기반 렌더링 기술은 기하학 정보의 사용 유무에 따라 크게 장면의 기하정보를 사용하지 않는 렌더링 기법, 화소간 대응관계를 이용하는 방법, 그리고 명확한 기하정보를 사용하는 방식으로 분류할 수 있다[6]. 기하정보를 사용하지 않는 렌더링 방법은 플레넵틱 함수(plenoptic function)의 특성을 이용하여 새로운 시점의 장면을 재현하는 것으로, 빛 정보와 텍스처 정보만을 이용한다. 대표적으로 루미그래프(lumigraph)[7]와 라이트 필드 렌더링(light field rendering)[8]이 여기에 속한다. 화소간 대응관계를 이용하는 렌더링 방법은 새로운 시점의 장면을 재현하기 위해 입력 영상들

의 위치적 관련성을 이용하며, 기하정보가 직접적으로 렌더링에 이용되지는 않는다. 이 방식은 투영행렬 (projection matrix) 계산을 통해 얻은 삼차원 위치정보를 새로운 장면 재현을 위해 사용한다. 시점 모핑 (view morphing) 기법[9]이 대표적인 예이다. 마지막으로, 명확한 기하정보가 있는 렌더링 방법은 직접적으로 깊이정보와 같은 삼차원 정보를 사용하며, 본 논문에서 기술하는 계층적 깊이영상[10]이나 영상기반 모델링을 통한 텍스처 매핑이 이 범주에 속한다.

본 장의 나머지 부분에서는 영상기반 렌더링 기법의 각 분류 중 대표적인 기하정보를 사용하지 않는 라이트 필드 렌더링, 화소간 대응관계를 이용하는 시점 모핑, 그리고 삼차원 깊이정보를 이용하는 계층적 깊이영상의 개념에 대해 소개하고, 각 방식의 특징들을 살펴본다.

#### 3.2 라이트 필드 렌더링

라이트 필드 렌더링[8]은 조명을 받은 물체 주위의 투명한 공간을 그 물체 표면에서 반사되는 빛으로 채워 실세계와 같은 장면을 연출한다. 이와 같은 라이트 필드는 여러 시점으로부터 얻어진 이차원 영상들을 규칙적인 격자로 이루어진 평면에 배치한다. 라이트 필드도 사차원 플레넵틱 함수를 사용하여 광선을 매개변수화하며,  $uv$ 와  $st$  평면에 의해 임의의 위치에 있는 화소 값을 결정한다.

보다 현실감 있는 렌더링을 위해서, 물체의 라이트 필드는 최대 시차를 만족시킬 수 있도록 충분히 얻어져야 한다. 만약 그렇지 않으면, 왜곡(aliasing) 현상이 생겨서 영상의 품질이 저하된다. 고품질의 영상을 재현하기 위해 필요한 영상의 수는 영상의 해상도에 의해 결정된다. 예를 들면,  $256 \times 256$  크기를 갖는 라이트 필드 영상들을 통해 실사와 같은 장면을 재현하려면 약 200,000

장의 영상이 필요하다. 따라서 획득된 라이트 필드는 항상 물리적으로 여러 개의 부분 표본화된 레벨로 나누어서 표현한다. 그러나 부분적으로 표본화된 라이트 필드(sub-sampled light field)라 하더라도 고품질의 영상을 재현하기 위해서는 수천 장이 필요하기 때문에, 데이터 압축은 라이트 필드 렌더링 응용에서 필수적이다.

라이트 필드 렌더링 기법과 같이 장면의 기하 정보를 이용하지 않는 방식의 장점은 많은 수의 고화질 영상을 획득하면 자연스러운 출력 영상을 얻을 수 있다는 것이지만, 이를 위해 엄청난 양의 데이터를 처리해야 하는 단점이 있다. 또한, 최근 라이트 필드의 개념을 동영상으로 확장하려는 시도가 있으나 아직은 연구가 미흡한 상황이다.

### 3.3 시점 모핑

영상 모핑 기법은 영상 사이에 이차원 적인 변화를 생성해 낸다. 그러나 물체의 위치나 시점의 차이로 인해 부자연스러운 왜곡이 발생하게 되므로, 이를 해결하여 자연스러운 중간 영상을 만드는 것이 영상 모핑의 주요 목적이다. 영상 모핑을 확장한 개념 중 대표적인 영상기반 렌더링 기법으로 시점 모핑이 있다. 시점 모핑[9] 기법은 투영 기하학(projective geometry)의 원리를 기반으로 삼차원 투영 카메라와 장면의 변환을 계산하여 최소한의 왜곡으로 자연스러운 변환 장면을 생성할 수 있는 방법이다.

우선 두 장의 영상을 전위핑 (prewarping) 시킨 후 시점 보간(interpolation)을 통해 중간 시점 혹은 가상 시점의 워핑된 영상을 계산한다. 그 후, 보간된 영상에 카메라의 위치와 시점을 고려한 후위핑(postwarping)을 적용하여 통해 가상 카메라의 위치에서 획득한 영상을 생성해 낸다. 시점 모핑은 라이트 필드 렌더링에 비해 매우 적은 영

상을 사용하고, 카메라의 기하 정보와 영상 모핑 및 시점 보간을 사용해 중간 영상을 생성한다. 하지만 동영상이 아닌 정지된 영상간 자연스러운 변환을 목적으로 하므로 다시점 동영상을 표현하는 데는 적합하지 않다.

### 3.4 계층적 깊이영상

계층적 깊이영상은 복잡한 기하정보를 갖는 삼차원 물체나 장면을 영상기반 렌더링 기법을 이용하여 표현하는 방법 중의 하나이다[10]. 계층적 깊이영상은 여러 시점에서 얻은 다수의 색상 및 깊이영상을 합성하여 하나의 데이터 구조를 생성한다. 따라서 각 계층적 깊이영상 화소는 색상정보 외에 화소와 카메라 사이의 거리를 나타내는 깊이 정보와 계층적 깊이영상의 렌더링을 지원하는 추가적인 특성정보를 가지고 있다. 또한, 계층적 깊이영상은 각 화소의 위치마다 다수의 계층을 가지므로, 이 정보를 이용하면 새로운 카메라 위치에서 삼차원 장면을 효과적으로 렌더링할 수 있다. 앞서 설명한 라이트 필드 렌더링이나 시점 모핑 기법과는 달리, 계층적 깊이영상은 기하정보를 직접 사용하는 영상기반 렌더링 범주에 속하기 때문에 상대적으로 적은 영상과 데이터양으로도 고화질의 장면을 재현할 수 있다.

계층적 깊이영상은 한 장의 영상 구조에 다수의 색상 및 깊이정보를 포함한다. 본 논문에서는 렌더링 속도 향상을 위해 기존의 데이터 구조[10]를 변형하였으며, 그림 1에 제안한 시스템에서 사용되는 계층적 깊이영상의 데이터 구조를 나타내었다. 앞서 언급한 바와 같이 광선이 물체를 투과할 때 교차점의 정보를 저장하는 개념에 맞게, 각 계층적 깊이영상 화소 위치에 계층별로 다양한 정보가 저장됨을 알 수 있다.

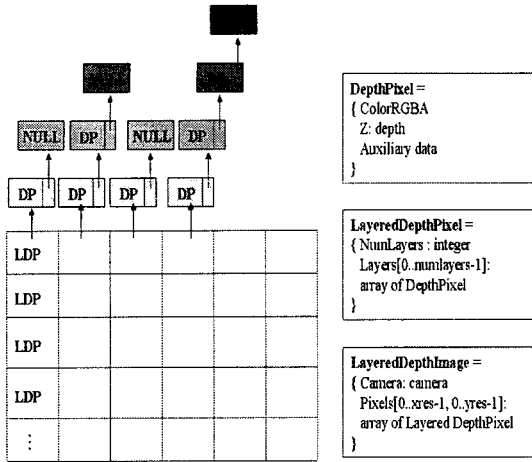


그림 1. 계층적 깊이영상의 데이터 구조

그림 1을 살펴보면, 계층적 깊이영상은 기존의 이차원 영상과 유사한 형태를 가지지만 각 화소가 색상 값만을 가지는 것이 아니라 추가적인 정보를 갖는다는 것이 가장 큰 차이점이다.

따라서 화소를 계층적으로 구분하여 Layered-DepthPixel(LDP)와 DepthPixel(DP)로 표기했으며, LDP는 서로 다른 개수의 DP를 포함하게 되고, 각 DP는 색상, 깊이, 렌더링에 사용되는 부가 정보를 가지게 된다. 또한, 각 계층별 화소의 존재 유무를 나타내는 정보도 추가적으로 필요하게 되며, 이와 같은 데이터 구조는 연결 리스트 형태로 구현된다.

계층적 깊이영상은 다시점 영상들이 가지는 다양한 정보를 한 시점으로 모아 구조적으로 표현이 가능하며, 삼차원 기하정보를 이용하기 때문에 라이트 필드에 비해 상대적으로 적은 수의 영상만으로도 고화질의 최종 영상을 생성해 낼 수 있다. 따라서 본 논문에서는 이러한 계층적 깊이영상의 개념을 기존의 정지된 영상이 아닌 동영상으로 확장하여, 다시점 동영상을 효과적으로 표현하고 처리해 원하는 시점을 재현할 수 있는 시스템을 제안한다.

## 4. 다시점 동영상의 영상기반 표현, 처리, 재현

### 4.1 계층적 깊이영상의 개념을 이용한 다시점 동영상의 표현

기존의 논문에서는 계층적 깊이영상을 주로 컴퓨터 그래픽스 모델로부터 생성했기 때문에 광선 투과기 등의 소프트웨어적 방법을 통해 앞서 설명한 개념을 적용할 수 있었다. 그러나 실제 물체에 대해서는 광선이 물체를 투과할 수 없기 때문에 새로운 방식이 필요하다. 따라서 실사영상으로부터 계층적 깊이영상을 생성하기 위해서는, 여러 시점에서 색상 및 깊이영상을 획득하여 이를 합성하는 방식을 사용한다[10,11].

계층적 깊이영상 생성의 가장 중요한 부분인 삼차원 워핑에 대해 구체적으로 알아보자. 우선 기존의 계층적 깊이영상이 주로 삼차원 컴퓨터 그래픽스 모델로부터 생성되었기 때문에, 이 때 사용되는 카메라 행렬을 살펴볼 필요가 있다. 삼차원 모델로부터 계층적 깊이영상을 생성할 때 사용되는 삼차원 워핑식은 다음과 같다[10].

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

여기서  $C_1$ 과  $C_2$ 는 시점 1과 2에서의 카메라 행렬이며,  $V$ 는 뷰포트(viewport) 행렬,  $P$ 는 투영(projection) 행렬,  $A$ 는 어파인(affine) 행렬을 나타낸다. 이때,  $C_1$ 에서 바라본 영상 평면의 한 점을  $C_2$ 의 위치로 이동하는 변환행렬이  $T_{1,2}$ 이다. 따라서 다음 식에 의해 카메라 시점 1에서 본 영상을 카메라 시점 2로 워핑할 수 있다[10]

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

(2)

실사 영상에 식(1)을 적용할 경우에는  $V$ 와  $P$ , 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.  $3 \times 4$  투영행렬이 존재할 경우에, 삼차원 위평에 사용되는 카메라 행렬  $\hat{C}$ 는 다음 식에 의해 계산된다.

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

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

여기서  $A$ 는 카메라의 내부 매개변수를 나타내는 행렬이며,  $f_s$ 는 초점거리,  $S_x, S_y$ 는 크기 성분,  $t_x, t_y$ 는 카메라 중심의 위치,  $\theta$ 는 비뚤어짐 (skew) 정도이다. 그리고  $E$ 는 외부 매개변수를 나타내는 행렬로서 구성요소인  $R$ 은 회전행렬을  $T$ 는 이동행렬을 나타낸다[12]. 이때,  $\hat{C}$ 는  $4 \times 4$  행렬이 되어야 하기 때문에  $\hat{A} \cdot \hat{E}$ 에  $[0 \ 0 \ 0 \ 1]$ 인 행을 추가하여 생성된다[12].

본 논문에서는 계층적 깊이영상의 개념을 이용해 다시점 동영상을 표현하기 위해, Microsoft Research(MSR)에서 배포한 다시점 동영상을 사용하여 실험을 수행하였다[13,14]. 현재 다양한 기관에서 다시점 동영상 테스트 데이터를 제작하여 배포하고 있지만, MSR에서 제공하는 데이터만이 다시점 깊이정보를 포함하고 있다. 깊이정보가 없는 다시점 동영상에 대해서는 스테레오 알고리즘을 사용하여 변이(disparity)정보를 예측할 수 있으나, 알고리즘의 종류와 성능이 매우 다양하여 적절한 기법의 선택에 어려움이 있다고 판단하여, 여기서는 직접 제공되는 데이터를 이용하였다. MSR 데이터는 일차원 원호형(arc)으로 배치된 8

대의 카메라로부터 획득되었으며, 카메라 매개변수가 제공된다.

#### 4.2 계층적 깊이영상의 개념을 이용한 다시점 동영상의 처리

다시점 동영상을 계층적 깊이영상의 형태로 표현하면 깊이값 정렬 및 임계화 과정에서 깊이값의 차이가 적은 화소들이 하나의 평균값을 가진 화소로 합쳐지므로, 상당 부분의 데이터를 줄일 수 있다. 하지만 매 시각( $T=0, 1, 2, \dots$ )마다 다시점 영상들로부터 하나의 계층적 깊이영상이 생성되므로, 다시점 동영상을 모두 계층적 깊이영상 기반 표현으로 변환하고 나면 계층적 깊이영상의 시퀀스가 생성되게 된다. 계층적 깊이영상의 개념을 이용해 다시점 동영상을 표현하고 처리하기 위해, 우선  $N$ 대의 카메라로부터 획득한 색상 및 깊이영상을 기준이 되는 계층적 깊이영상 시점으로 워핑한다. 이때, MSR에서 제공하는 데이터의 경우 카메라 사이의 간격이 20cm이기 때문에 같은 장면에 대해 카메라 별로 비폐색 (disocclusion) 영역이 발생하게 된다. 특히, 좌우측 끝에 위치한 카메라 일수록 보이지 않는 부분이 많아 최종 다시점 영상을 복원할 경우 필요한 정보가 부족하게 된다. 또한, 삼차원 워핑을 수행하고 깊이값을 비교하여 계층적 깊이영상을 생성하는 중간 단계에서도 같은 위치로 이동된 화소들이 통합되는 과정에서 어느 정도 정보 손실이 발생하기 때문에 부가 정보를 보내주고, 정보 손실을 보상해 주는 부분이 시스템에 포함된다[11,15].

부호화를 수행할 때는 생성된 계층적 깊이영상 프레임이 일반적인 이차원 영상형태가 아닌 중간에 화소가 비는 불규칙한 형태를 취하기 때문에, 임의영상 부호화를 지원하지 않는 H.264/AVC를 그대로 적용할 수가 없다. 따라서 데이터 모으기

및 패딩을 통해 H.264/AVC에 지원하는 형태로 계층적 깊이영상 프레임을 변환해 주는 전처리 과정이 필요하다[12]. 이와 같이 다시점 동영상을 계층적 깊이영상 기반 표현으로 나타내면, 삼차원 정보를 포함하는 계층적 깊이영상의 데이터 구조를 압축함으로써 효과적으로 데이터양을 줄일 수 있을 뿐만 아니라, 적은 수의 영상으로도 영상기반 렌더링 기법을 적용해 새로운 시점을 생성할 수 있는 장점이 있다.

### 4.3 계층적 깊이영상 기반 표현을 이용한 다시점 동영상 재현

계층적 깊이영상 표현으로부터 원래의 다시점 동영상을 재현하고 새로운 시점을 생성하는 과정은 4.1절에 기술한 계층적 깊이영상 생성 과정의 역으로 생각할 수 있다. 우선 복호된 계층적 깊이영상 프레임을 각 카메라 시점으로 역위평한다. 이 때, 이미 생성단계에서 발생한 폐색영역과 데이터 손실로 인해 역위평된 결과만으로는 복원이 불가능하다. 여기서 사용하는 것이 LDP의 후면 계층에 위치하고 있던 DP이며, DP가 담고 있는 다시점 색상 및 깊이정보를 사용하여 현재 프레임의 빈 화소 위치를 최대한 보상한다. 하지만 여전

히 폐색영역은 동영상 획득시부터 정보가 부족한 상황이었기 때문에 보상이 어렵다. 따라서 이 부분에 대해서는 원본 다시점 동영상으로부터 보상에 필요한 정보를 추가로 보내주어야 한다. 최종적으로 역위평된 결과에 DP에 담겨있던 다시점 색상 및 깊이정보와 원본으로부터 추출된 부가정보를 사용하면 다시점을 복원할 수 있다.

## 5. 실험 결과 및 분석

### 5.1 계층적 깊이영상 기반 표현을 이용한 다시점 동영상 표현

그림 2는 실험에 사용된 다시점 동영상 테스트 데이터[13,14]의 첫 번째 프레임을 나타낸다. 총 8대의 카메라에서 획득된 데이터이며, 영상의 해상도는 1024×768이고, 프레임율은 15fps이다. 카메라 4를 기준 카메라 위치로 사용하였다.

그림 2를 살펴보면 카메라 0번 위치에서는 좌측 끝부분에 벽이 보이지만, 카메라 번호가 증가할수록 벽부분이 점점 보이지 않게 되어 카메라 6번 위치부터는 아예 보이지 않는 것을 확인할 수 있다. 반대로 카메라 7번 위치에서는 우측 끝부분에 벽이 보이지만, 카메라 0번 위치에서는 보이

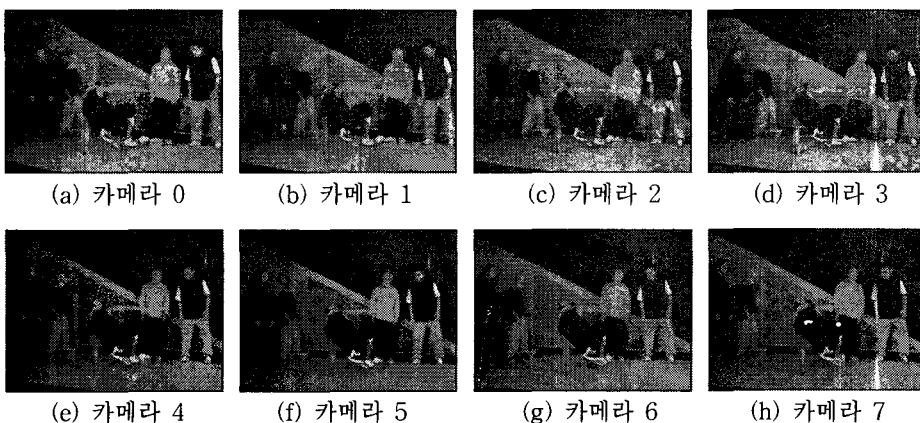


그림 2. 다시점 동영상 테스트 데이터: "Breakdancers"

지 않는다. 이런 부분들은 폐색영역으로서 획득시 부터 정보가 부족하기 때문에, 시스템에서는 복원 시 필요한 부가정보를 복호기단으로 보내주게 된다. 기준 시점인 4번 카메라 위치로 모든 시점의 영상을 워핑한 후에는 이동된 화소들을 정렬하고 깊이값을 비교하여 계층적 깊이영상을 생성하게 된다. 계층적 깊이영상시 생성되는 최대 계층의 수는 다시점 동영상의 최대 시점의 수와 동일하다. 깊이값 비교시 임계값 30을 사용하였을 때 생성된 계층적 깊이영상의 계층별 영상을 그림 3에 나타내었다. 화소별로 다른 수의 계층을 갖으며, 후위 계층으로 갈수록 화소의 수가 줄어드는 것을 확인할 수 있다.

5.2 계층적 깊이영상 기반 표현을 이용한 다시점 동영상 처리

본 논문에서는 생성되는 계층적 깊이영상 시퀀스의 데이터양을 줄임으로써 복호화 성능을 향상시키고, 결과적으로 다시점 재현시 렌더링 속도를 높이기 위해 계층적 깊이영상을 부호화하였다. 부호화를 위해 H.264/AVC를 사용하였으며, 임의형상 부호화를 데이터 모으기를 위해 그림 5에 나타낸 계층적 깊이영상의 각 계층별 영상에 각각 데

이터 모으기를 적용한다. 그 후, 데이터 모으기가 수행된 계층별 영상을 하나의 영상으로 합친 후, 다시 정해진 해상도의 영상으로 분할하여 시퀀스를 생성하게 된다. 같은 과정을 계층적 깊이영상에 대해서도 수행한 후 생성된 색상과 깊이영상 시퀀스를 H.264/AVC를 이용해 각각 부호화한다.

표 1은 원본 다시점 동영상과 계층적 깊이영상 표현을 부호화 후 데이터의 크기를 비교한 결과를 나타낸다. 표 1에서 8시점 영상의 합계는 BMP 형식의 원본 데이터양을 나타낸다. 계층적 깊이영상 표현은 이 다시점 영상을 계층적 깊이영상으로 표현하였을 경우 데이터양이며, 이때 깊이값 비교시 임계치는 50을 사용하였다. 실험 결과 임계치가

표 1. 다시점 영상과 계층적 깊이영상 표현의 데이터 크기 비교 (단위: kB)

| "Breakdancers"         | 첫 번째 8 프레임(T=0) | 두 번째 8 프레임(T=1) |
|------------------------|-----------------|-----------------|
| 8 시점 영상의 합계 (색상 + 깊이)  | 25,166          | 25,166          |
| 계층적 깊이영상 표현            | 13,808          | 13,723          |
| 8시점 영상을 H.264/AVC로 부호화 | 137.7           | 132.5           |
| 계층적 깊이영상 표현 부호화        | 91.7            | 93.0            |

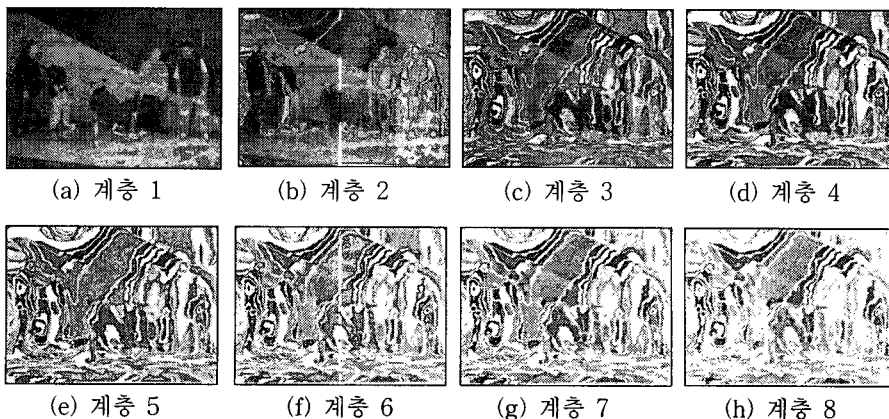


그림 3. 생성된 계층적 깊이영상의 계층별 색상영상



5.0 이상이 되면 데이터양의 감소가 거의 없어지게 된다. 다음은 8시점 영상을 H.264/AVC로 부호화한 후 데이터 크기를 나타낸다. 기존의 계층적 깊이영상을 압축하는 논문에서는 임의형상을 그대로 유지하여 웨이블릿 기반의 임의형상 코덱으로 압축하였으나, H.264/AVC를 사용하면 표 1에 나타난 바와 같이 압축률이 200배 정도로 기존 방식에 비해 훨씬 높다. 또한, 계층적 깊이영상 표현에 데이터 모으기를 적용한 후 H.264/AVC로 부호화한 경우 데이터 크기가 더 감소했음을 확인할 수 있다.

### 5.3 계층적 깊이영상 기반 표현을 이용한 다시점 동영상 복원

복호된 계층적 깊이영상으로부터 각 시점으로 역으로 삼차원 워핑을 수행한 결과를 그림 4에 나타내었다[16]. 가운데 위치한 영상이 기준 카메라로 다시 삼차원 워핑을 수행한 결과이다. 위쪽의 영상들은 각각 카메라 0과 카메라 1번 위치로, 아래쪽 영상들은 카메라 6과 카메라 7번 위치로 역워핑한 결과이다. 폐색과 깊이값 임계화를 통한 정보손실로 많은 부분이 빈화소로 표시됨을 관찰할 수 있다. 따라서 그림 4의 결과를 후위 계층에 있는 DP에 포함된 색상 및 깊이정보를 이용하여 보상해 주어야 한다. 또한, 획득시부터 존재했던

폐색영역부분과 물체의 움직임이 심한 부분도 보상이 필요하다. 이를 위해 원본에서 추출된 부가정보를 이용하여 최종 복원 영상을 생성하였다. 그림 5는 최종 다시점 재현 결과를 나타내며, 부자연스러운 부분이 있긴 하지만 대부분의 영역이 제대로 복원된 것을 알 수 있다[16]

## 6. 결론

다시점 동영상은 여러 시점에서 다수의 카메라로 한 장면을 촬영한 다중 동영상의 집합이며, 획득된 정보를 이용하여 사용자에게 원하는 시점의 영상을 제공할 수 있다. 따라서 사용자의 시청 실감을 높일 수 있는 다양한 응용분야에 사용될 중요한 실감형 멀티미디어이다. 그러나 다시점 동영상은 카메라 수에 비례하여 데이터양이 늘어나기 때문에, 이를 다양한 응용분야에서 효과적으로 사용하기 위해서는 다시점 동영상의 효율적인 표현, 처리, 부호화, 재생 기법을 개발하는 것이 필요하다. 본 논문에서는 영상기반 렌더링 기법 중 계층적 깊이영상의 개념을 이용하여 다시점 동영상을 효과적으로 표현, 처리, 복원하는 방법을 제안하였다. 제안한 방법은 다시점 동영상을 삼차원 정보가 포함된 계층적 깊이영상 표현으로 나타냄으로써 다시점 영상들이 가지는 다양한 정보를 구조

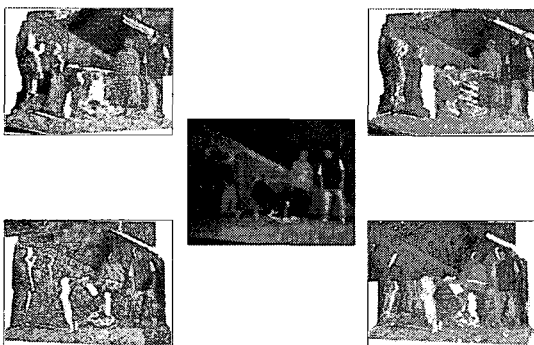


그림 4. 역 삼차원 워핑 수행 결과



그림 5. 최종 복원된 다시점 영상

적으로 표현하였다. 또한, 이를 압축함으로써 데이터양을 줄이고, 적은 수의 영상으로도 다시점 영상을 자연스럽게 복원하였다.

## 참 고 문 헌

- [1] ISO/IEC JTC1/SC29/WG11 N6720, "Call for Evidence on Multi-view Video Coding," Oct 2004
- [2] ISO/IEC JTC1/SC29/WG11 N6999, "Report of the Subjective Quality Evaluation for Multi-view Coding CfE," January 2005.
- [3] ISO/IEC JTC1/SC29/WG11 N7327, "Call for Proposals on Multi-view Video Coding," July 2005
- [4] ISO/IEC JTC1/SC29/WG11 N7779, "Subjective Test Results for the CfP on Multi-view Video Coding," Jan 2006
- [5] ISO/IEC JTC1/SC29/WG11 N8019, "Description of Core Experiments in MVC," Apr 2006
- [6] H. Shum and S. Kang, "Survey of Image-based Representations and Compression Techniques," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 13, No. 11, pp. 1020-1037, 2003
- [7] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The Lumigraph," Proc. of ACM SIGGRAPH, pp. 43-54, 1996
- [8] M. Levoy and P. Hanrahan, "Light Field Rendering," Proc. of ACM SIGGRAPH, pp. 31-42, 1996
- [9] S. Seitz and C. Dyer, "View Morphing," Proc. of ACM SIGGRAPH, pp. 21-30, 1996
- [10] J. Shade, S.J. Gortler, and R. Szeliski, "Layered Depth Image," Proc. of ACM SIGGRAPH, pp. 291-298, 1998
- [11] 호요성, 윤승욱, 이은경, 김성열, "계층적 깊이영상을 이용한 다시점 비디오 표현 및 부호화 기술", 전자공학회지, 제32권, 제9호, pp. 1070-1080, 2005. 9
- [12] ISO/IEC JTC1/SC29/WG11 m12485, "Generation and Coding of Layered Depth Images for Multi-view Video," Oct 2005
- [13] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality Video View Interpolation using a Layered Representation," Proc. of ACM SIGGRAPH, pp. 600-608, 2004
- [14] Interactive visual media group at Microsoft Research, <http://research.microsoft.com/vision/InteractiveVisualMediaGroup/3DVideoDownload/>
- [15] S.U. Yoon, S.Y. Kim, E.K. Lee, and Y.S. Ho, "A Framework for Multi-view Video Coding using Layered Depth Images," Lecture Notes in Computer Science (LNCS), vol. 3767, pp. 431-442, 2005
- [16] ISO/IEC JTC1/SC29/WG11 m12849, "Reconstruction of Multi-view Images from Layered Depth Images," Jan 2006



윤 승 욱

- 2000년 서강대학교 전자공학 학사
- 2002년 광주과학기술원 정보통신공학 석사
- 2002년~현재 광주과학기술원 정보통신공학과 박사과정
- 관심분야 : 다시점 동영상 부호화, 계층적 깊이영상 부호화 및 영상 기반 렌더링, 실감미디어 처리 및 실감방송



호 요 성

- 1981년 서울대학교 전자공학 학사
- 1983년 서울대학교 전자공학 석사
- 1990년 Univ. of California at Santa Barbara, 전기컴퓨터 공학 박사
- 1990~1993 미국 Philips 연구소 선임연구원
- 1983~1995 한국전자통신연구원 선임연구원
- 1995~현재 광주과학기술원 교수
- 관심분야 : 디지털 신호처리 및 방송, 영상 압축, 멀티미디어 시스템, 3-D TV 및 실감방송