

# 계층적 깊이영상을 이용한 다시점 비디오 표현 및 부호화 기술

호요성, 윤승욱, 이은경, 김성열 (광주과학기술원 정보통신공학과)

## I. 서론

우리는 정보혁명과 디지털 시대로 대변되는 21세기에 살고 있다. 이러한 디지털 시대의 핵심은 여러 가지 형태의 멀티미디어의 등장이라고 할 수 있다. 3차원 컴퓨터 그래픽스 영상을 비롯해서 파노라마 영상, 3차원 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 3차원 오디오 비주얼 (Three-dimensional Audio Visual, 3DAV) 기술이다. 그 중에서도 다시점 비디오 (multi-view video)에 관련된 기술은 차세대 영상 산업에서 중요한 역할을 할 것이다.

다시점 비디오는 똑같은 3차원 장면을 여러 시점에서 다수의 카메라로 촬영한 것으로, 획득된 정보를 이용하여 사용자에게 원하는 시점의 영상을 제공하는 것을 목적으로 한다. 주요 응용분야로는 자유시점 TV (Free viewpoint TV, FTV), 3차원 TV, 감시 카메라

라 영상 (surveillance), 교육, 홈 엔터테인먼트 등이 있다. 그러나 이러한 장점에도 불구하고 카메라 수만개의 비디오가 존재하기 때문에 데이터양이 매우 많아서 이를 효과적으로 저장하고 전송하기 위해서는 효율적인 부호화 기술을 적용하는 것이 필요하다.

본 논문에서는 영상기반 렌더링 기법의 하 나인 계층적 깊이영상을 이용한 다시점 비디오 부호화 방법에 대해 알아본다. 제II장에서는 다시점 비디오에 관련된 기술의 개발 현황을 소개하고, 제III장에서 영상기반 렌더링 기술을 살펴본다. 제IV장에서는 계층적 깊이 영상의 개념 및 이를 이용한 다시점 비디오 부호화 기술에 대하여 설명하고, 마지막으로 제V장에서 결론을 맺는다.

## II. 다시점 비디오 기술 개발현황

유럽에서는 ATTEST (Advanced Three-Dimensional Television System Technologies) 과제를 통해 2002년 3월부터 2년 동안 3차원 TV에 관한 기초 기술을 연구하였으며, 2004년 9월부터는 20여개 기관이 컨소시엄을 구

성하여 3DTV (Three-dimensional Television) 과제를 수행하고 있다. 3DTV 과제는 3차원 장면의 획득, 표현, 부호화, 전송, 디스플레이까지를 모두 포함하고 있으며, 다시점 비디오를 획득하는 부분도 주요 연구 분야 중 하나이다. 미국에서는 NASA에서 주로 3차원 영상과 관련된 연구를 수행하고 있으며, MIT에서는 다시점 비디오 및 홀로그래픽 디스플레이 기술을 연구하고 있다. 일본에서도 다시점 비디오 과제를 진행하고 있으며, 초다시점 3차원 TV 기술 개발에 많은 투자를 하고 있다. 일본의 경우, 이미 FTV 시험방송을 할 정도로 다시점 비디오에 대한 관심이 매우 크다.

다시점 비디오 부호화와 관련하여 일본의 나고야 대학, NTT, 독일의 Fraunhofer Heinrich-Hertz-Institut (HHI) 연구소, 미국에 위치한 Mitsubishi Electric Research Laboratories (MERL) 등에서 연구 개발이 활발하다. 특히, 일본의 나고야 대학에서는 자유시점 TV 시스템의 중간 단계로 광선 공간(ray space) 시스템을 연구하고 있으며, 100대의 카메라를 이용하여 다시점 비디오를 획득하는 시스템을 개발했다. HHI, MERL, Microsoft Research 등에서는 8대의 카메라로부터 다시점 비디오를 획득하였고, 이를 이용하여 다시점 비디오 테스트 데이터를 제작하여 배포하고 있다.

국내의 경우, ETRI에서 2002년 월드컵에서 3차원 TV 시범방송 서비스를 제공하였다. KIST는 다시점 디스플레이 장치를 연구하고 있고, 삼성전자와 LG전자에서도 3차원 카메라, 3차원 TV, 안경식 스테레오 LCD 모니터를 연구하고 있다. 그 밖에 광주과학기술원,

강원대, 광운대, 세종대, 연세대 등 일부 대학교에서 다시점 영상 압축과 전송 기술을 연구하고 있다. 이와 같이, 국내의 일부 대학, 연구소, 산업체에서 다시점 비디오에 대한 연구가 진행되고 있으며, 이에 대한 관심이 급속히 높아지고 있다.

이와 같은 기술적인 흐름에 발맞춰 전통적인 스테레오 영상의 압축 부호화 뿐만 아니라, 여러 형태의 다시점 비디오 데이터를 압축 부호화할 필요성이 대두되었다. ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group(MPEG)에서는 이미 MPEG-2(ISO/IEC 13818-2)와 MPEG-4(ISO/IEC 14496-2)에서 평행형 다시점 영상 부호화 기능과 3차원 메쉬(mesh) 압축 부호화 기능을 제공하였다. 여기에 추가적으로 다시점 비디오 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 비디오 테스트 데이터가 제공되었으며, 2004년 10월에 다시점 비디오 부호화에 대한 Call for Evidence(CfE)가 발행되었다.<sup>[1]</sup> 2005년 1월에는 CfE에 대한 응답으로 다양한 다시점 비디오 부호화 알고리즘에 대한 결과가 수집 및 분석되었으며, 2005년 7월에 Call for Proposal(CfP)이 배포되었다.<sup>[2]</sup>

지금까지 MPEG 3DAV 그룹에 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<sup>[3]</sup>, 2차원 직접 모

드 및 조명 보상을 이용한 방법<sup>[4]</sup> 등이 있다. 이러한 방식들의 공통점은 기존의 2차원 비디오 부호화 방식을 확장하여, 시간축 뿐만 아니라 각 카메라의 영상 사이에서도 상관관계를 고려한 점이다. 이와는 달리 최근에 제안된 계층적 깊이영상 기반의 다시점 비디오 부호화 방식은 영상기반 렌더링의 특징을 이용하여 하나의 통합된 계층적 깊이영상을 만들어 이를 부호화하고, 복원한 뒤에 새로운 시점의 영상을 손쉽게 생성할 수 있는 장점이 있다.<sup>[5]</sup> 따라서 본 논문에서는 계층적 깊이영상 기반의 다시점 비디오 표현 및 부호화 기술에 대해 구체적으로 살펴보고자 한다.

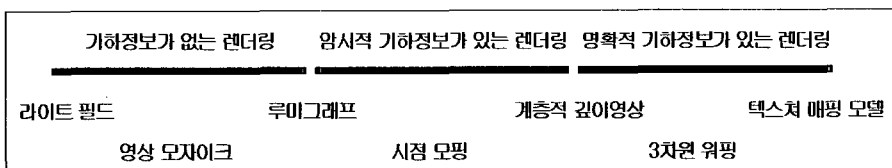
### III. 영상기반 렌더링 기술 동향

#### 1. 영상기반 렌더링 기법의 분류

영상기반 렌더링은 여러 시점의 2차원 영상을 이용하여 3차원 공간의 임의 시점에서의 영상을 생성하는 방법이다. 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 훨씬 쉽다. 최근에는 이러한 영상기반 렌더링 기법 중에서 계층적 깊이영상(Layered Depth Image, LDI)에 관한 연구가

주목을 받고 있다. 계층적 깊이영상은 여러 시점에서 생성된 깊이영상을 합성하여 하나의 데이터 구조로 만든 것으로, 다중 깊이정보를 이용하여 임의 시점의 영상을 생성할 수 있다. 이러한 기능은 다시점 비디오의 목적과 유사하므로 영상기반 렌더링 기법을 다시점 비디오 부호화에 적용할 수 있다.

영상기반 렌더링은 최근에 고전적인 모델기반 렌더링의 대체 기술로 각광을 받고 있다. 영상기반 렌더링 기술은 기하학 정보의 사용 유무에 따라 크게 기하정보가 없는 렌더링, 암시적 기하정보가 있는 렌더링, 그리고 명확한 기하정보가 있는 렌더링으로 분류할 수 있다.<sup>[6]</sup> 기하정보가 없는 렌더링 방법은 플렌옵틱 함수(plenoptic function)의 특성을 이용하여 새로운 시점의 장면을 재현하는 것으로, 빛 정보와 텍스처 정보만을 이용한다. 루미그래프(lumigraph)와 라이트 필드(light field) 렌더링이 이 범주에 속한다. 암시적 기하정보가 있는 렌더링 방법은 새로운 시점의 장면을 재현하기 위해 입력 영상들의 위치적 관련성을 이용하며, 암시적이란 용어를 사용한 이유는 기하정보가 직접적으로 렌더링에 이용되지 않기 때문이다. 이 방식은 3차원 위치정보를 투영행렬(projection matrix) 계산을 통해 얻고, 이 정보를 새로운 장면 재현을 위해 사용한다. 시점 모핑(view morphing) 기법이 이 범주에 속한다. 마지막으로, 명확



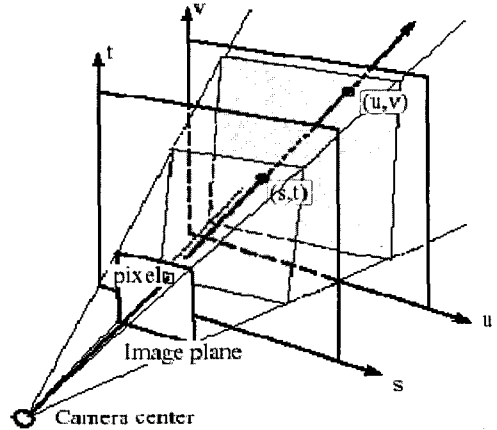
〈그림 1〉 영상기반 렌더링 기술의 분류

한 기하정보가 있는 렌더링 방법은 직접적으로 깊이정보와 같은 3차원 정보를 사용하며, 전통적인 텍스처 맵핑 방법이 이 범주에 속한다. 본 논문에서 기술하는 계층적 깊이영상도 여기에 속하며, 최근 이에 관한 연구가 활발하다. 그림 1은 다양한 영상기반 렌더링 기술의 분류를 나타낸다.

## 2. 루미그래프 (lumigraph)

루미그래프는 4차원 플레네펍틱 함수를 이용하여 복잡한 모델의 완벽한 형상을 부호화하고, 모델의 복잡도와는 독립적으로 객체를 표현한다.<sup>[7]</sup> 플레네펍틱 함수는 모든 3차원 공간 좌표에서 모든 2차원 방향으로 빛의 흐름을 나타내는 5차원적인 함수이다. 만약 공기가 투명하다고 가정하면 빈 공간을 통하는 광선의 광휘는 상수이다. 또한, 관심영역을 객체의 불록한 표면에 남아있는 빛으로 더 제한한다면, 객체를 둘러싼 표면에 대해서만 플레네펍틱 함수의 값을 표현하면 된다. 그러므로 객체에 대한 플레네펍틱 함수는 4차원으로 줄어들 수 있으며, 루미그래프에서는 이 4차원 함수를  $(s, t, u, v)$ 의 매개변수로 표현한다. 플레네펍틱 함수의 4차원 매개변수로 표현한 결과를 그림 2에 나타내었다. 그림 2에서 보듯이, 방향은  $st$ 평면에 평행한  $uv$  평면을 사용하여 매개변수화한다. 4차원 루미그래프에서 임의의 한 점의 위치는  $st$  평면의 한 점  $(s, t)$ 를 지나  $uv$  평면의 한 점  $(u, v)$ 와 만나는 광선의 좌표  $(s, t, u, v)$ 에 의해 인식된다. 평면에 수직인  $z$ 축을 지니고,  $uv$  평면의 중앙에 원점이 위치한다.  $st$  평면은  $z = 1$ 에 위치한다고 가정한다. 완전한 루미그래프

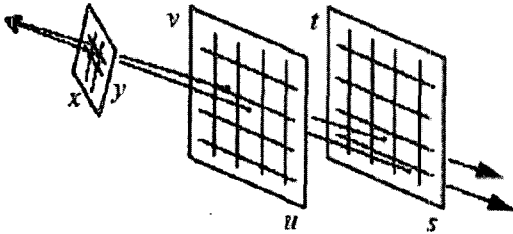
는  $x, -x, y, -y, z, -z$ 방향의 법선 벡터를 가지는 6개의 평면으로 구성된다.



〈그림 2〉 임의의 영상에서 화소와 루미그래프 사이의 관계

## 3. 라이트 필드 (light field) 렌더링

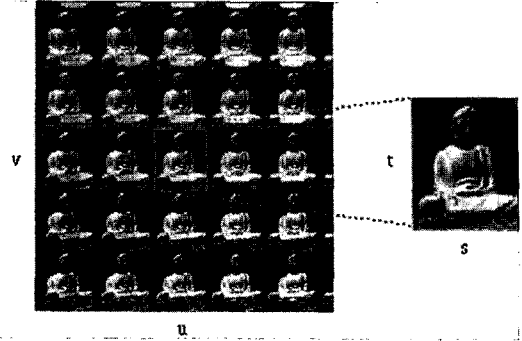
라이트 필드 렌더링은 조명을 받은 물체 주위의 투명한 공간을 그 물체 표면에서 반사되는 빛으로 채워 실세계와 같은 장면을 연출한다.<sup>[8]</sup> 이와 같은 라이트 필드는 여러 시점으로부터 얻어진 2차원 영상들을 규칙적인 격자로 이루어진 평면에 배치한다. 라이트 필드도 4차원 플레네펍틱 함수를 사용하여 값을 매개변수화하며,  $uv$ 와  $st$  평면에 의해 임의의 위치의 화소 값을 결정한다. 그림 3에 나타낸 바와 같이, 카메라 평면  $uv$ 와 영상 평면  $st$ 가 평행하게 존재하고, 그 두 평면 사이를 광선이 통과한다. 라이트 필드는 이 두 평면과 광선간의 교차점을 이용하여, 루미그래프에서와 마찬가지로, 라이트 필드 함수  $LF(u, v, s, t)$ 로 매개변수화할 수 있다. 보다 현실감 있는 렌더링을 위해서, 물체의 라이트 필드는 최대 시차를 만족시킬 수 있도록



〈그림 3〉 라이트 필드의 4차원 매개변수화

충분히 얻어져야 한다. 만약 그렇지 않으면, 왜곡 (aliasing) 현상이 생겨서 영상의 품질이 저하된다. 고품질의 영상을 획득하기 위해 필요한 영상의 수는 영상의 해상도에 의해 결정된다. 예를 들면, 256 x 256 크기를 갖는 라이트 필드 영상들을 통해 실사와 같은 장면을 재현하려면 약 200,000장의 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.

그림 4는 영상평면  $st$ 에서의 시점 영상과 그 시점에 해당하는 재표본화(resampling) 영상을 보여주고 있다.<sup>9)</sup> 이는 시점의 변화에 따라 각기 다른 영상을 생성한다. 라이트 필드는 이렇게 카메라 평면과 영상 평면간의 인텍싱 과정과 재표본화 과정을 통해 현재 관찰자가 원하는 시점의 장면을 생성하고 렌더링할 수 있다.

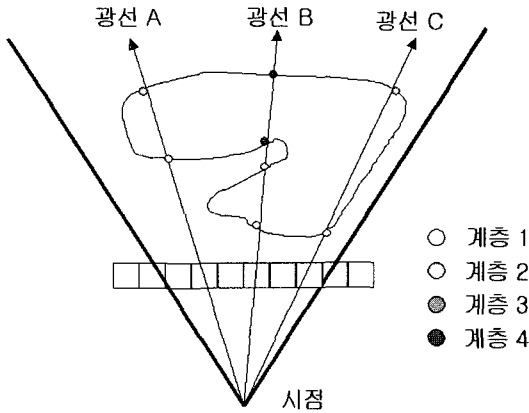


〈그림 3〉 영상평면  $st$ 와 카메라 평면  $uv$ 의 인텍싱 구조

## IV. 계층적 깊이영상을 이용한 다시점 비디오 부호화 기술

### 1. 계층적 깊이영상의 개념

계층적 깊이영상(Layered Depth Image, LDI)은 복잡한 기하정보를 갖는 3차원 물체나 장면을 영상기반 렌더링 기법을 이용하여 표현하는 방법 중의 하나이다.<sup>10)</sup> 다각형 메쉬를 사용해서 모델을 표현하는 방식과는 달리, 계층적 깊이영상은 여러 시점에서 얻은 다수의 깊이영상을 합성하여 하나의 데이터 구조를 생성한다. 따라서 각 계층적 깊이영상 화소는 색상정보 외에 화소와 카메라 사이의 거리를 나타내는 깊이정보와 계층적 깊이영상의 렌더링을 지원하는 추가적인 특성정보를 가지고 있다. 또한, 계층적 깊이영상은 각 화소의 위치마다 다수의 계층을 가지므로, 이 정보를 이용하면 새로운 카메라 위치에서 3차원 물체를 효과적으로 렌더링할 수 있다. 앞서 설명한 루미그래프나 라이트 필드와는 달리, 계층적 깊이영상은 기하정보를 직접 사용하는 영상기반 렌더링 범주에 속하기 때문에 수천 장의 영상을 필요로 하지 않는다.



〈그림 5〉 계층적 깊이영상 생성의 개념도

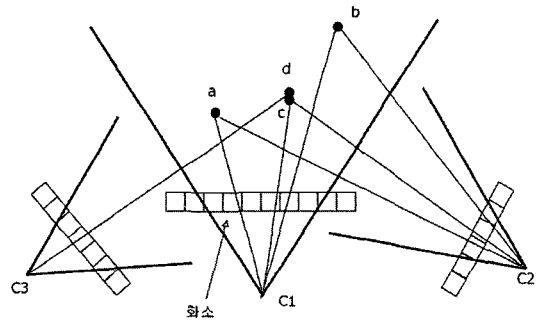
그림 5는 이러한 특징을 갖는 계층적 깊이 영상의 생성 방식을 나타내고 있다.<sup>9)</sup> 그림 5에서 보는 바와 같이, 계층적 깊이영상을 생성하기 위해 기준 시점에서 물체를 향해 광선을 투과시킨다. 이때, 광선 A는 물체와 두 곳에서 만나게 되며, 광선 B는 네 곳에서 만난다. 따라서 광선 A가 통과하는 계층적 깊이영상 화소에는 2개의 계층이 생성되며, 광선 B가 통과하는 화소에는 4개의 계층이 형성된다. 따라서 모든 화소 위치마다 같은 수의 계층이 형성되는 것이 아니라, 광선과 물체의 교차점의 개수에 따라 생성되는 계층의 수가 달라진다. 그리고 각 교차점에서 색상과 깊이정보를 모두 저장하여 하나의 구조로 만들면 계층적 깊이영상이 생성된다.

## 2. 다시점 비디오로부터 계층적 깊이영상 생성

컴퓨터 그래픽스를 이용해 생성된 모델이 아닌 실제 물체에 대해서는 광선이 실제 물체를 투과할 수 없기 때문에 앞서 언급한 방

식을 그대로 적용할 수가 없다. 따라서 실사 영상으로부터 계층적 깊이영상을 생성하기 위해서는, 앞서 설명한 방식이 아닌, 여러 시점에서 깊이영상을 획득하여 이를 합성하는 방식을 사용한다.<sup>10)</sup>

계층적 깊이영상과 관련하여 MPEG에서는 MPEG-4 Synthetic and Natural Hybrid Coding (SNHC) 그룹에서 영상기반 렌더링에 관한 내용(AFX CE 8.3)을 Animation Framework eXtension(AFX)에 포함시켜 이에 대한 표준화 작업을 진행하였다. 그러나 AFX에서는 단순히 3차원 물체를 표현하는 영상기반 렌더링 도구로서의 계층적 깊이영상의 기능에만 초점을 맞추었다. 최근 3DAV 그룹에서 제안된 계층적 깊이영상 기반 다시점 비디오 부호화 기법은 기존의 계층적 깊이영상의 개념을 다시점 비디오 부호화로 확장한 새로운 기술이다.



〈그림 6〉 다수의 깊이영상을 이용한 계층적 깊이영상의 생성

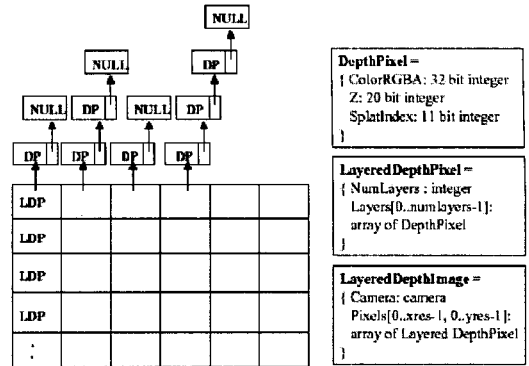
그림 6은 다수의 깊이영상을 이용한 계층적 깊이영상 생성 방식을 나타낸다. 카메라 시점 C2와 C3의 위치에 있는 깊이영상을 공통의 시점인 계층적 깊이영상 시점 (LDI view), C1로 워핑(warping)하여 계층적 깊이

영상을 생성할 수 있다. 입력 카메라 시점인 C2와 C3의 화소를 C1 시점으로 워핑할 때 2개 이상의 화소가 같은 위치로 맵핑(mapping)되면, 그 위치에서 맵핑된 화소들이 가지고 있는 원래의 깊이 값을 비교한다.

만약 깊이 값의 차이가 미리 정한 임계치보다 크면 현재 화소 위치에 새로운 계층이 추가된다. 그렇지 않고 깊이 값의 차이가 임계치보다 작으면, 두 화소의 평균값으로 현재 계층의 값을 갱신하고 새로운 계층은 추가하지 않는다. 그림 6에서 후자에 해당하는 경우가 공통의 카메라 시점으로 워핑되는 c와 d이다. c와 d 화소는 깊이 값이 거의 일치하므로 두 화소의 깊이 값의 평균을 취하고, 두 화소를 하나의 계층으로 통합한다.

이렇게 생성된 계층적 깊이영상은 한 장의 영상 구조에 다수의 깊이정보를 포함한다. 그림 7은 계층적 깊이영상의 데이터 구조를 나타내며, 각 계층적 깊이영상 화소 위치에서 서로 다른 수의 계층이 존재함을 알 수 있다.<sup>10)</sup> 이렇게 한 영상 구조에 모든 부가정보를 포함시키면 가려진 영역(occlusion region)을 처리할 때 기존의 우선순위 기반 화소선택법을 사용할 수 있어 상당히 유용하다.

그림 7을 살펴보면, 계층적 깊이영상은 기존의 2차원 영상과 유사한 형태를 가지지만 각 화소가 색상 값만을 가지는 것이 아니라 추가적인 정보를 갖는다는 것이 가장 큰 차이점이다. 따라서 화소를 계층적으로 구분하여 LDP와 DP로 표기했으며, LayeredDepthPixel(LDP)은 서로 다른 개수의 DepthPixel(DP)을 포함하게 되고, 각 DP는 색상, 깊이, 렌더링에 사용되는 부가정보인 SplatIndex를 가지게 된다.



〈그림 7〉 계층적 깊이영상의 데이터 구조

이러한 구조를 가지는 계층적 깊이영상을 실사영상으로부터 생성하기 위해서는 우선 사용이 가능한 다시점 비디오 테스트 데이터의 특성을 분석해 볼 필요가 있다. 현재 MPEG 3DAV 그룹에서는 다양한 다시점 비디오 테스트 데이터를 제공하고 있다. 몇몇 기관, 연구소, 회사에서 다양한 특성을 지닌 20여개의 테스트 시퀀스를 제공하고 있으며, 최근에 MPEG에서 다시점 비디오 부호화에 대한 Call for Proposals(CfP)을 배포하면서 8개의 테스트 시퀀스가 선택되었다[2]. CfP를 위한 테스트 시퀀스 선택의 기준은 사용된 카메라의 개수, 카메라 구성, 초당 프레임수, 영상의 해상도, 장면의 복잡도, 카메라의 움직임 등이다. 또한, 선택된 테스트 시퀀스는 모두 카메라 매개변수를 포함하고 있으며, 이에 대한 검증 작업이 현재 MPEG 3DAV 그룹 내에서 진행 중이다. 최근까지 MPEG 3DAV 그룹에서 제공한 다양한 형태의 다시점 비디오 데이터의 특성을 표 1과 표 2에 나타내었다.<sup>11)</sup> 표 2는 CfP를 위해 선택된 8개의 테스트 시퀀스를 나타낸다.

표 1과 표 2에서 보는 바와 같이, 카메라의 개수는 5대, 8대, 100대가 사용되었으며, 영

〈표 1〉 MPEG 3DAV 다시점 비디오 테스트 데이터의 특성

데이터 특성	KDDI	MERL	HHI	Nagoya Univ.	MSR
시퀀스 이름	Flamenco Race Golf Objects Crowd	Ballroom Exit	Jungle Uli	Rena Akiko Akko&Kayo	Ballet Breakdancers
카메라 개수	5/8	8	8	100	8
카메라 매개변수	있음	있음	있음	있음	있음
깊이 정보	없음	없음	없음	없음	없음

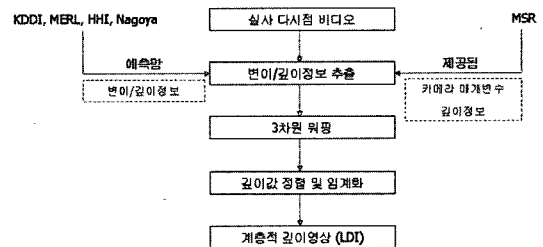
〈표 2〉 CfP를 위해 선택된 테스트 시퀀스의 특성

제공 기관	시퀀스 이름	영상 특성	카메라 개수	카메라 구성
MERL	Ballroom	VGA, 25fps	8	1차원 평행
MERL	Exit	VGA, 25fps	8	1차원 평행
KDDI	Race1	VGA, 30fps	8	1차원 평행
KDDI	Flamenco2	VGA, 30fps	5	2차원 평행, 직교
HHI	Uli	XVGA, 25fps	8	1차원 평행, 수렴
MSR	Breakdancers	XVGA, 15fps	8	1차원 원호
Nagoya Univ.	Rena	VGA, 30fps	100	1차원 평행
Nagoya Univ.	Akko&Kayo	VGA, 30fps	100	2차원 배열

상은 VGA (640 x 480) 크기와 XVGA (1024 x 768) 크기가, 초당 프레임 수는 15, 25, 30 이 사용되었다. 또한, 카메라 구성은 1차원 평행, 1차원 수렴, 1차원 원호, 2차원 평행, 2차원 배열이 사용되었다.

이렇게 다양한 다시점 비디오가 제공되고 있지만 이 중에서 계층적 깊이영상을 생성하는데 필요한 정보인 깊이정보를 제공하고 있는 데이터는 Microsoft Research(MSR)에서 배포한 데이터가 유일하다. MSR에서는 평행한 8대의 카메라로부터 획득한 다시점 비디오를 제공하고 있으며, MSR에서 배포하는 테스트 데이터는 카메라별 색상뿐 아니라 깊

이영상 및 카메라 매개변수를 포함하고 있다.<sup>[10]</sup> 이러한 실사 다시점 비디오 시퀀스로부터 계층적 깊이영상을 생성하는 전체적인 흐름도를 그림 8에 나타내었다.<sup>[5]</sup>



〈그림 8〉 실사 다시점 비디오로부터 계층적 깊이영상 생성



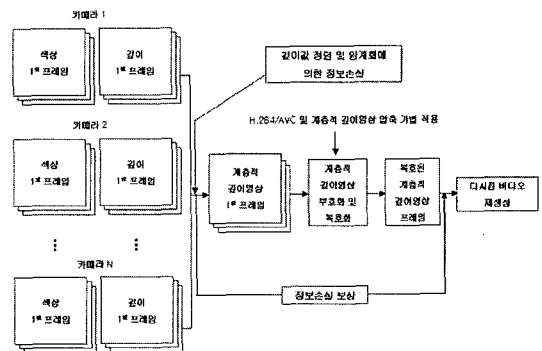
### 3. 계층적 깊이영상을 이용한 다시점 비디오 부호화 구조

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 영상기반 렌더링 기법 중의 하나인 계층적 깊이영상을 이용하면 다시점 비디오를 부호화할 수 있다.<sup>15)</sup>

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 그림 9에 나타낸 바와 같이, 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 그림 9에 포함된다.

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 임의시점 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.



〈그림 9〉 계층적 깊이영상 기반 다시점 비디오 부호화 구조

## V. 결론

본 논문에서는 현재 MPEG 3DAV 그룹에서 표준화가 진행되고 있는 다시점 비디오 부호화 기술 중 계층적 깊이영상의 개념을 이용한 다시점 비디오 표현 및 부호화 기술을 살펴보았다. 다시점 비디오는 한 장면을 여러 시점에서 다수의 카메라로 촬영한 것으로, 자유시점 TV (Free viewpoint TV, FTV), 3차원 TV, 감시 카메라 영상 (surveillance), 교육, 홈 엔터테인먼트 등에서 핵심기술로 사용된다. 그러나 카메라 수만큼의 비디오가 존재하기 때문에 데이터양이 매우 많아서 이를 효과적으로 저장하고 전송하기 위해서는

부호화가 필수적이다. 본 논문에서 소개한 계층적 깊이영상 기반의 다시점 비디오 부호화 기술은 기존 접근 방식과는 달리 영상기반 렌더링 기술을 이용하여 다시점 비디오를 부호화하는 새로운 접근 방식이다. 향후에는 구축된 프레임워크 상에서 효과적인 부호화 기법에 관한 연구가 진행될 예정이다.

---

### 참고 문헌

---

- [1] "Call for Evidence on Multi-view Video Coding," ISO/IEC JTC1/SC29/WG11 N6720, October 2004.
- [2] "Call for Proposals on Multi-view Video Coding," ISO/IEC JTC1/SC29/WG11 N7327, July 2005.
- [3] "Multi-view Video Coding using Shared Picture Memory and Shared Vector Memory," ISO/IEC JTC1/SC29/WG11 m11570, January 2005.
- [4] "Multi-view Video Coding using Illumination Change-adaptive MotionEstimation/Motion Compensation and 2D Direct Mode," ISO/IEC JTC1/SC29/WG11 m11588, January 2005.
- [5] "A Framework for Multi-view Video Coding using Layered Depth Image," ISO/IEC JTC1/SC29/WG11 m11582, January 2005.
- [6] H. Shum, S.B. Kang, and S. Chan, "Survey of Image-based Representations and Compression Techniques," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 11, pp. 1020-1037, November 2003.
- [7] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The Lumigraph," *Proc. of ACM SIGGRAPH*, pp. 43-54, August 1996.
- [8] M. Levoy and P. Hanrahan, "Light Field Rendering," *Proc. of ACM SIGGRAPH*, pp. 31-42, August 1996.
- [9] J. Shade, S. Gotler, L. He, and R. Szeliski, "Layered Depth Image," *Proc. of ACM SIGGRAPH*, pp. 231-242, July 1998.
- [10] 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, August 2004.

## 저자소개



### 호요성

1983년 3월 - 1995년 9월 한국전자통신연구소 선임 연구원  
 1990년 1월 - 1993년 5월 미국 Philips 연구소, Senior Research Member  
 1995년 9월 - 현재 광주과학기술원 정보통신공학과 교수  
 주관심 분야 디지털 신호처리, 영상신호 처리 및 압축, 초저속 영상통신, 디지털 TV와 고선명 TV 방식, 실감 방송/실감미디어 처리



### 김성열

2001년 2월 강원대학교 정보통신공학과 학사  
 2003년 2월 광주과학기술원 정보통신공학과 석사  
 2003년 3월 - 현재 광주과학기술원 정보통신공학과 박사과정  
 주관심 분야 3차원 메쉬 모델링 및 압축, 디지털 신호 처리 및 영상신호처리, 멀티미디어 전송



### 윤승욱

2000년 2월 서강대학교 전자공학과 학사  
 2002년 8월 광주과학기술원 정보통신공학과 석사  
 2002년 9월 - 현재 광주과학기술원 정보통신공학과 박사과정  
 주관심 분야 다시점 비디오 부호화, MPEG 3DAV, 영상기반 렌더링 데이터 부호화, 3차원 메쉬 모델링 및 압축



### 이은경

2002년 2월 호남대학교 컴퓨터공학과 학사  
 2004년 2월 호남대학교 컴퓨터공학과 석사  
 2004년 9월 - 현재 광주과학기술원 정보통신공학과 박사과정  
 주관심 분야 다시점 비디오 부호화, MPEG 3DAV, 라이트 필드 데이터 부호화