

+ 호요성, 이 천 · 광주과학기술원 실감방송연구센터

# TECH&TREND

## MVC 다시점 비디오 부호화 기술

### 1. 서론

바야흐로 3차원 영상시대가 다가오고 있다. 3차원 입체 영화가 연이어 개봉되고 있으며, 3차원 디스플레이 장치의 수요도 증가하고 있다. 이와 더불어, 여러 방송 사업자들은 3차원 방송을 준비하고 있고, 영화 제작사들도 3차원 영화 제작을 서두르고 있다. 이렇듯 3차원 영상에 대한 관심이 높아지면서, 3차원 영상 서비스에 대한 국제 표준을 제정하기 위한 작업도 활기를 띠고 있다[1][2].

현재 MPEG(moving picture experts group) 국제 표준화 그룹에서는 기존의 양안식(stereoscopic) 영상뿐만 아니라, 다시점(multi-view) 영상을 효율적으로 압축하여 전송하고 재생하는 기술의 표준화 작업을 진행하고 있다[3]. 이 과정에서 다수의 카메라로 획득한 다시점 영상과 3차원 거리 정보를 나타내는 깊이 영상을 압축 부호화한다. 기존의 다시점 영상에 비해, 다시점 영상은 시점 수에 따라 데이터양이 늘어나므로 이러한 단점을 보완하기 위해 새로운 부호화 기술이 필요하다[4]. 수신단에서는 전송된 신호를 복원하여 여러 시점의 입체 영상을 재현할 뿐만 아니라, 시청자가 원하는 임의 시점의 영상도 만들어 보여 줄 수 있어야 한다. 이러한 기술을 이용하면 자연스럽게 사실감 나는 3차원 영상을 안방에서 바로 시청할 수 있으며, 3차원 영상을 이용한 실감나는 게임이나 전시에도 효과적으로 적용할 수 있다[5].

MPEG 표준화 그룹에서는 2001년 말부터 3차원 영상 전송을 위한 새로운 표준화 대상을 모색하던 중, 2004년 9월부터 다시점 비디오 부호화 기술에 대한 표준화 작업을 시작했고, 최근에 다시점 비디오를 효과적으로 부호화하는 방법(multi-view video coding, MVC)을 국제 표준으로 제정했다[6].

본 논문에서는 3차원 비디오 시스템을 위한 다시점 비디오 부호화 기술을 설명하고, 현재 MPEG 그룹에서 작업하고 있는 3차원 비디오 부호화(3D Video Coding) 기술을 소개한다.

## 2. 3차원 오디오-비주얼 부호화 기술(3D Audio-visual, 3DAV)

동영상을 효율적으로 압축하는 기술을 개발해 온 MPEG 표준화 그룹에서는 단시점 영상에 대한 압축뿐만 아니라, 3차원 비디오에 대한 기술도 오랫동안 연구해 왔다. 이미 MPEG-2(ISO/IEC 13818-2)와 MPEG-4(ISO/IEC 14496-2)에서 평행형 다시점 비디오 부호화 기능과 3차원 메쉬(mesh) 압축 부호화 기능을 만들어 제공했고, 2001년 12월부터 새로운 3DAV 부호화 표준의 가능성에 대해 조사했다. 이어서 2002년부터 2004년까지 관련 기술에 대한 다양한 선행 실험(Exploration Experiments, EE)을 수행했다[7].

[그림 1]처럼 EE1에서는 360° 영상을 보여주는 전방향 비디오(omni-directional video)에 관한 실험을 수행했다. 이 실험을 위해 촬영된 전방향 비디오를 3차원 메쉬(mesh)를 이용하여 표현하고, 이를 MPEG-4 표준 기술로 부호화하는 방법이 제안되었다.



[그림 1] 전방향 비디오

EE2에서는 사용자가 시청하려는 시점을 자유롭게 선택할 수 있는 기능을 가진 자유시점 비디오 기술이 검토되었다. 이 시스템을 제안한 일본의 나고야 대학교에서는 광선 공간(ray space)을 이용한 중간시점 영상의 합성 방법을 MPEG-4 시스템에 추가할 것을 제안했다.

EE3에서는 MAC(multiple auxiliary component) 시스템을 기반으로 하는 양안식 비디오 부호화 방법을 검토했다. 이 기술은 인접한 두 시점 영상 사이의 변이(disparity)를 추정하고, 이때 얻은 변이 지도(disparity map)와 한 쪽 컬러 영상을 부호화하여 전송한다. 수신단에서는 부호화된 신호를 복원화하고, 복원된 변이 지도와 한 쪽 영상을 이용하여 다른 쪽 시점의 영상을 생성하여 양안식 비디오를 복원하는 기술이다.

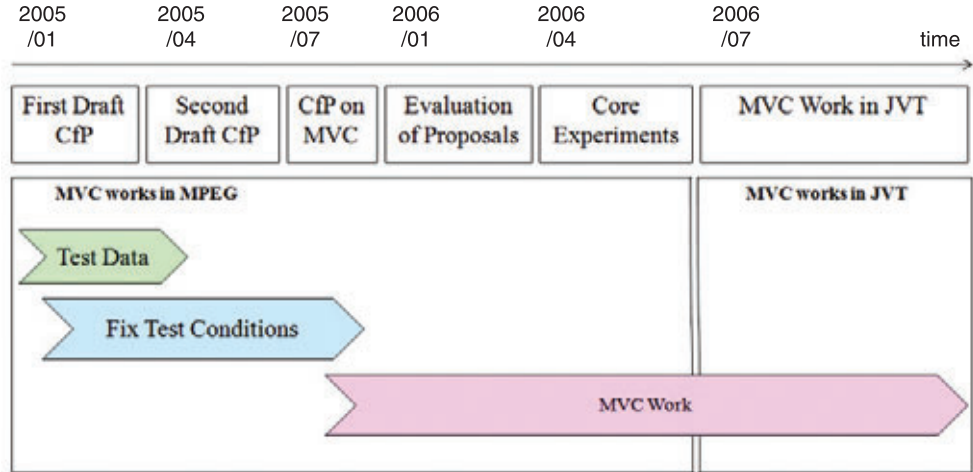
EE4에서는 깊이 또는 변이 정보를 부호화하기 위한 실험으로, 비디오와 깊이 영상을 MAC을 이용하여 부호화하는 방법이 주로 검토되었다. 이 실험에서는 깊이나 변이 정보를 다양한 필터를 이용하여 처리한 뒤에 부호화 결과를 비교했다. 디블록킹 필터(deblocking filter)나 중간 값 필터(median filter)를 이용했을 때, 부호화 비트량은 감소했고, 깊이 영상을 이용한 합성 영상의 화질은 개선되었다.

### 3. 다시점 비디오 부호화 기술

#### (Multi-view Video Coding, MVC)

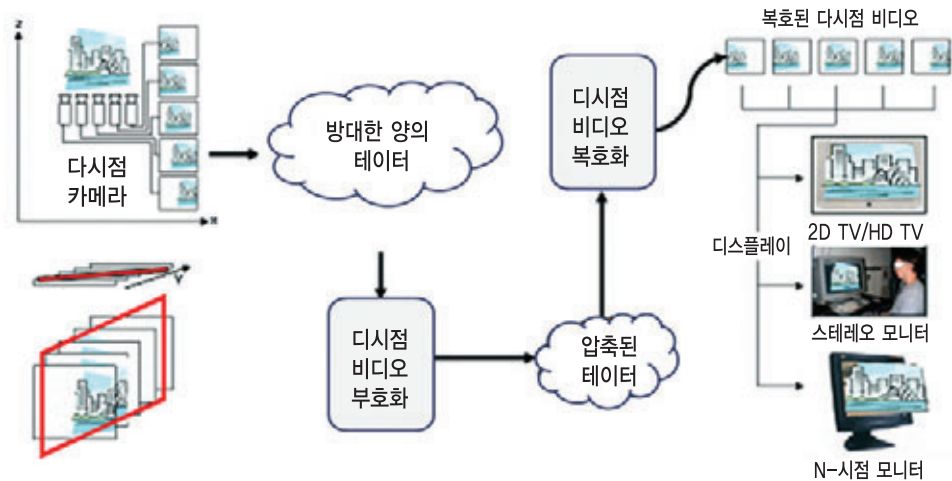
3차원 오디오 비주얼 부호화의 선행 실험을 통해 3차원 영상의 처리 방법을 검토했으나, 그 범위가 너무 넓어 표준화 항목을 정하는데 어려움이 많았다. 그러던 중 다시점 비디오 부호화의 필요성을 인지하여 표준화 논의가 구체화되었다.

이어 2004년 8월에 다시점 비디오 테스트 영상을 제공했고, 2005년 8월에 제안요청서(Call for Proposals)를 배포하여 여러 연구 기관에서 제안한 기술을 검토했다. 그로부터 2년 후인 2006년 7월부터 MVC에 관한 표준화 작업을 JVT(joint video team)로 이양하여 수행했다.



[그림 2] 다시점 비디오 부호화의 표준화 작업

MPEG 표준화 그룹에서는 3차원 비디오 시스템을 구성하기 위한 기본 입력 데이터로 다시점 비디오를 이용했다. [그림 3]과 같이 다시점 카메라로 획득한 영상을 효과적으로 압축하여 전송하고, 단말기에서 수신된 신호를 복원하여 재생장치에 적합한 시점을 선택하여 재현하는 시스템을 구성했다. 다시점 영상을 이용하면 양안식 재생 장치를 이용하여 시청자가 원하는 임의의 시점을 선택하여 감상하거나, 다시점 3차원 모니터에 필요한 만큼의 임의의 시점 영상을 생성하여 보다 자연스럽게 실감나는 3차원 영상 서비스를 제공할 수 있다.



[그림 3] 다시점 비디오 시스템

하지만, 이러한 시스템이 상용화되기 위해서는 시점 수의 증가에 따른 방대한 양의 데이터 처리가 필요하다. MPEG 그룹에서는 이러한 다시점 비디오 정보를 효율적으로 압축하여 부호화하는 기술을 표준화하는 작업을 수행했다[8].

기존의 비디오 부호화 방법에서는 단시점 영상의 공간적/시간적/통계적 상관도를 이용하여 중복 정보를 줄이거나 제거하지만, 다시점 비디오 부호화 방법에서는 기존의 방법에 시점간 상관도를 추가적으로 이용한다. 다시점 영상은 인접한 여러 대의 카메라를 이용하여 똑같은 장면을 동시에 촬영하기 때문에, 시차와 약간의 조명 차이를 제외하면 거의 같은 정보를 담고 있으므로 시점간 상관도가 매우 높다. MPEG 표준화 과정에서는 이러한 시점간 상관도를 이용한 다양한 기술이 제안되었다. 기본적으로는 다시점 비디오를 효과적으로 부호화하기 위한 예측 구조에 대한 다양한 방법이 논의되었다.

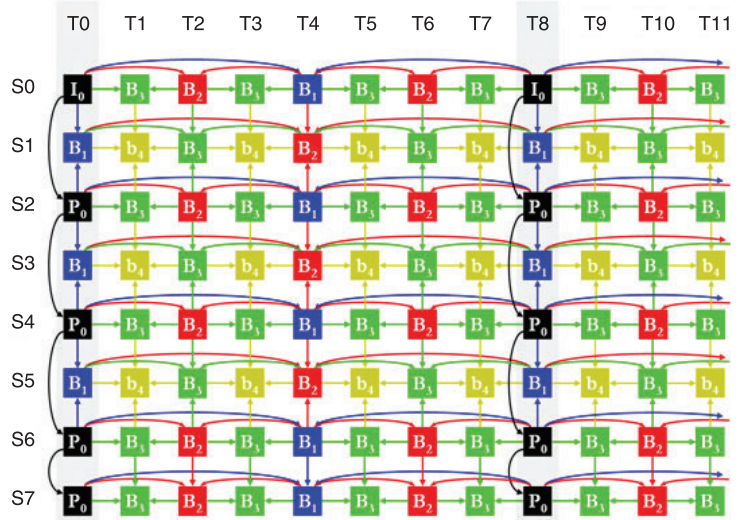
또한, 시점간 상관도를 이용하면 부호화할 시점의 영상을 생성할 수 있는데 이렇게 생성한 영상을 부호화에 다시 이용하는 시점 합성 예측(view synthesis prediction) 방법이 제안되었으며, 시점간 색상 차이를 보정하는 조명 보상 (illumination compensation) 등의 다양한 기술도 검토되었다.

### 3-1. 계층적 B 화면 부호화를 이용한 예측 구조

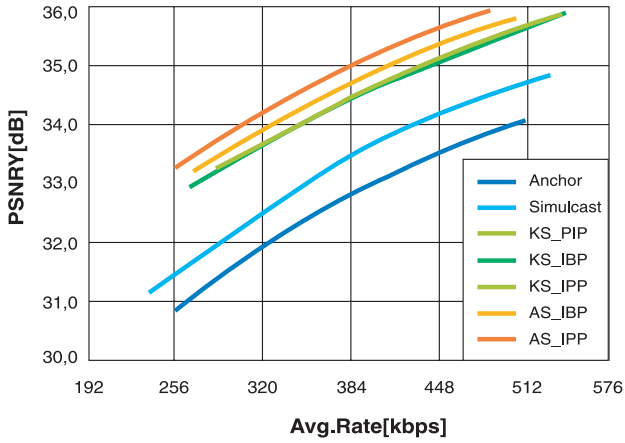
다시점 비디오 시스템에서는 두 대 이상의 카메라를 이용하여 획득한 영상을 효과적으로 압축 부호화한다. 이를 위한 기본적인 예측 구조는 [그림 4]와 같다. [그림 4]의  $S_m$ 은  $m$ 번째 시점의 카메라를 의미하고,  $T_n$ 은 시간적으로  $n$ 번째 화면을 의미한다. 화살표는 이웃하는 화면들 사이의 예측 참조 관계를 의미한다. 기존 시스템과의 호환성을 유지하기 위해 다른 시점과 상관없이 독립적으로 복원할 수 있는 시점을 I-시점이라 하며, 부호화가 끝난 하나의 시점만 참조하여 예측 부호화하는 시점을 P-시점, 양쪽의 두 시점을 참조하여 예측 부호화하는 시점을 B-시점이라고 말한다.

[그림 4]에서는 시점  $S_0$ 이 I-시점에 해당하고,  $S_2, S_4, S_6, S_7$  등이 P-시점,  $S_1, S_3, S_5$  등이 B-시점에 해당한다. 이렇게 정해진 예측 구조에서 I-시점을 가장 먼저 부호화하고, P-시점을 부호화한 후, 이어서 B-시점을 부호화한다. 즉,  $S_0-S_2-S_1-S_4-S_3-S_6-S_5-S_7$  순으로 부호화를 수행한다. 임의 접근(random access)을 위해 일정한 간격으로 기준 화면(anchor frame)을 둔다. 이 기준 화면은 오직 시점간 예측만을 이용하여 부호화한다.

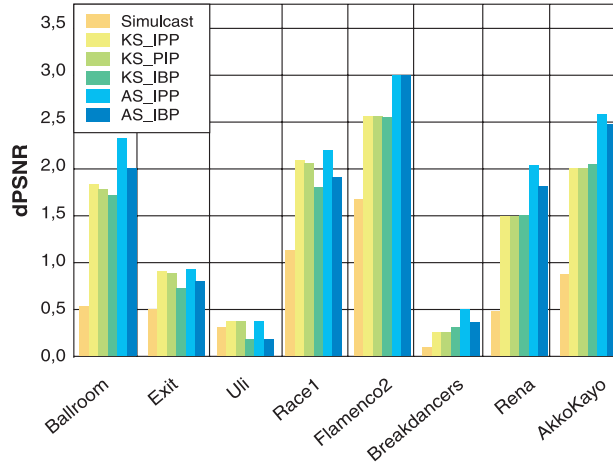
[그림 5]는 다시점 비디오를 부호화하기 위해 다양한 예측구조를 테스트한 결과이다[9]. [그림 5]에서 'Anchor'는 H.264/AVC를 이용하여 각 시점별로 따로따로 부호화한 결과이고, 'AS\_IBP'가 MVC에서 사용하고 있는 예측 구조를 이용한 결과이다. 이러한 부호화 구조를 이용하면 H.264/AVC로 독립적으로 부호화하는 것보다 약 50%의 비트수를 줄일 수 있다.



[그림 4] 계층적 B화면을 이용한 기본적인 예측 구조



〈비트율-왜곡 곡선〉



〈평균 부호화 이득 비교〉

[그림 5] 다양한 예측구조에 따른 부호화 결과 비교

### 3-2. 조명 보상(Illumination Compensation)

다시점 비디오는 여러 카메라로 촬영한 영상이기 때문에 동일한 장면을 촬영하더라도 카메라의 위치, 카메라 변수의 차이와 조리개 조절 등이 문제로 색상과 조명 값의 변화가 생길 수 있다. 이는 사람의 눈으로 그 차이를 크게 느낄 수 없으나, 적은 화소 값의 변화에도 민감한 비디오 부호화 과정에서는 약간 문제가 될 수 있다. 따라서 이 문제를 해결하기 위한 다양한 조명 보상 기술이 제안되었다[10].

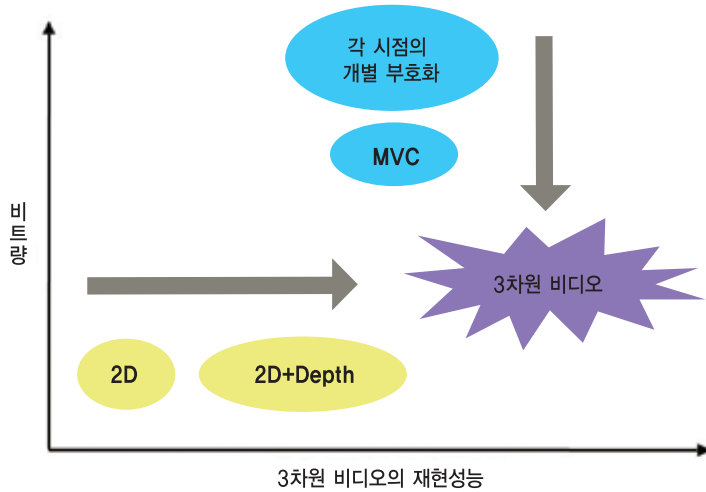
특히, 움직임 예측/보상 과정에서 매크로블록 단위로 조명을 보상해 주는 방법(ICA MC, illumination change adaptive motion compensation)이 참조 소프트웨어에 구현되었다. 이 방법은 기존의 다양한 매크로블록 모드에 조명을 보상한 16x16 모드를 추가하여 움직임 예측을 수행하고, 비트율-왜곡(rate-distortion) 관점에서 조명을 보상한 16x16 모드가 선택될 경우 DVIC(difference value of illumination change) 값을 전송하여 조명을 보상하는 방법이다.

### 3-3. 시점 합성 예측(View Synthesis Prediction)

다시점 비디오 부호화의 예측 구조를 이용하면, 이미 부호화가 끝난 인접 시점의 영상을 이용하여 부호화할 다른 시점의 영상을 합성할 수 있다. 이렇게 합성한 영상을 부호기의 참조 화면으로 이용하면 부호화 효율을 높일 수 있다[11]. 이와 같이 부호화할 시점의 영상을 합성할 때 가장 큰 영향을 주는 것은 시점간 변위를 정확하게 측정하는 것이다. 따라서, 인접한 시점 영상의 각 화소별 변위를 측정하고, 이를 이용하여 부호화할 시점의 중간시점 영상을 생성하는 기술이 제안되었다.

### 3-4. 변위 벡터 부호화(Disparity Vector Coding)

비디오 부호화에서 B-화면은 시간적으로 앞뒤에 놓인 두 참조 화면을 이용하여 부호화하므로, 복잡도는 증가하나 보다 적은 비트를 가지고 좋은 화질을 나타낼 수 있다. B-화면은 이전 참조 화면을 이용하는 경우와 이후 참조 화면만을 이용하는 경우, Direct 모드, 두 참조 화면을 모두 이용하는 경우로 나누어진다. 두 화면을 모두 이용하는 경우에는 한 블록의 움직임 벡터를 두 참조 화면으로부터 구할 수 있다. 이러한 경우에 움직임 벡터를 변위 벡터(disparity vector)라고 말하며, 이를 효율적으로 부호화하는 방법이 제안되었다[12].



[그림 6] 3차원 비디오 부호화의 개발 방향

#### 4. 3차원 비디오 부호화 기술

##### (3D Video Coding, 3DVC)

다시점 비디오 부호화 기술은 3차원 비디오 시스템의 기본 입력인 다시점 비디오를 압축하기 위해 개발된 국제 표준이다. 이러한 기술을 기반으로 3차원 비디오 시스템을 더욱 구체화한 기술이 3차원 비디오 시스템이다. 이 시스템은 각 시점의 깊이 영상 획득, 다시점 영상의 전처리, 3차원 데이터의 압축 및 전송, 그리고 최종 단말기에서 3차원 영상을 재현하는 단계까지 모두 포함한다.

[그림 6]은 MPEG 그룹에서 발표한 3차원 비디오 부호화 기술의 개발 방향을 나타낸 것으로, 3차원 비디오를 처리하기 위한 다양한 접근 방법을 제시하고 있다[6]. 현재 MPEG 그룹에서 3차원 비디오 부호화를 위해 고려하고 있는 접근 방법은 다시점 비디오와 다시점 깊이 영상을 동시에 부호화하는 방법이다. 기존의 다시점 비디오 부호화 방법을 기반으로 깊이 영상 부호화까지 포함하는 새로운 부호화 기술이 개발될 예정이다.

##### 4-1. 다시점 깊이 맵 추정 기술

깊이 맵(depth map)이란 영상 내에 존재하는 객체의 3차원 거리 정보를 나타내는 영상으로, 깊이 영상의 화소 값은 해당 화소의 깊이 정보를 나타낸다. 깊이 영상의 정확도는 합성된 중간시점 영상의 화질에 영향을 주기 때문에 정확한 깊이 영상을 구하는 작업이 매우 중요하다. 깊이 영상의 이해를 돕기 위하여 깊이 값의 개념을 다음과 같이 설명할 수 있다.



[그림 7] 변이와 깊이의 상관관계

[그림 7]과 같이 두 대의 카메라를 정확히 수평이 되도록 정렬하고, 카메라의 방향이 평행이 되도록 조정하자. 이러한 카메라 구성에서 어느 한 영상에 존재하는 특정한 화소가 다른 시점의 영상에도 존재할 때, 이 두 시점에서의 좌표의 수평 차이를 변이(disparity)라고 한다. 만약, 객체가 카메라에서 가까이 있을 경우에 변이는 큰 값을 가지며, 반대로 객체가 카메라에서 멀어질수록 변이는 작은 값을 가지게 된다. 시점간 색상의 상관도를 이용하면 각 화소의 변이 값을 구할 수 있다. 이렇게 구해진 값을 3차원 거리 값으로 변환한 것이 깊이 값(depth value)이다.

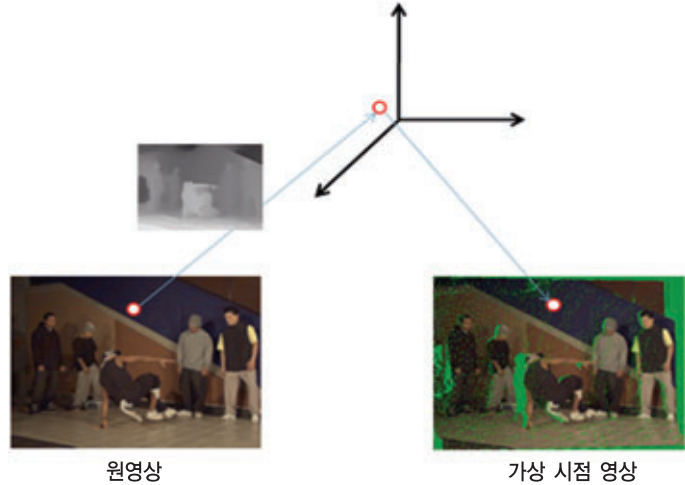




MPEG 그룹에서는 3차원 비디오 부호화 기술을 표준화하기 위해 다시점 비디오의 깊이 영상을 생성하는 방법을 개발했다. 2008년 4월에 깊이 추정 참조 소프트웨어(DERS)를 배포했고, 이후 다양한 기술이 참조 소프트웨어에 구현되었다. 현재는 DERS 5.1까지 만들었는데, MPEG 그룹에서 관리하는 SVN 서버에서 해당 소프트웨어를 얻을 수 있다[13].

#### 4-2. 가상시점 영상 합성 기술

MPEG 그룹에서 고려하고 있는 3차원 비디오 시스템에서는 자유시점 영상의 선택 기능과 다시점 3차원 모니터의 재생 기능을 위해 카메라가 존재하지 않는 가상시점의 영상을 합성하여 만드는 기술을 논의하고 있다. [그림 8]은 카메라 파라미터와 깊이 정보를 이용하여 임의 시점의 영상을 생성하는 과정을 설명한 것이다. 카메라로 획득한 원영상과 깊이 영상을 이용하면, 각 화소를 3차원 공간으로 투영할 수 있다. 이렇게 투영된 각 화소를 원하는 가상시점으로 다시 투영하면 가상시점의 영상을 생성할 수 있다. 이러한 방법을 3차원 워핑을 이용한 영상 합성이라고 한다.



가상시점의 영상을 합성할 때 시점 이동에 따라 새로 드러나는 물체의 뒤 부분에 빈 영역(hole region)이 발생한다. 이러한 겹친 영역의 대한 정보는 원영상에 존재하지 않기 때문에 빈 영역 주변의 정보를 이용해서 보완해야 한다. 기존의 양안식 데이터 형식과 같이 한 시점의 참조 영상과 깊이 맵을 이용할 경우에는, 인접한 화소 값을 이용하여 빈 영역을 채울 수 있다.

하지만, 다시점 비디오와 같은 환경에서는 합성하려는 중간시점 영상의 좌우로 참조 영상과 깊이 영상이 동시에 존재하므로, 다른 참조 시점의 정보를 취하여 중간시점의 빈 영역을 채운다. MPEG 표준화 그룹에서는 이러한 기술을 이용한 시점 합성 참조 소프트웨어(view synthesis reference software, VSRS)를 개발하여 배포하고 있다. 현재 VSRS 3.5까지 개발되었으며, MPEG SVN 서버에서 관리하고 있다[14].

#### 4-3. 3차원 비디오 부호화 기술의 표준화 동향

표는 현재까지 진행된 3차원 비디오 부호화 기술의 표준화 작업을 정리한 것이다. 2008년 1월의 MPEG 모임에서는 3차원 비디오 시스템을 개발하기 위한 기본 환경을 구축하는데 필요한 테스트 영상, 깊이 영상 생성 소프트웨어, 중간시점 영상 합성 소프트웨어를 포함한 요소 기술의 기고를 요청했다.

[3차원 비디오 부호화 기술의 표준화 과정]

표준화 회의	회의 내용
2001년 12월 ~ 2004년 7월	• 3DAV에 관한 표준화 진행
2004년 10월 ~ 2007년 4월	• MVC에 관한 표준화 진행
2007년 4월	• 자유시점 TV에 대한 표준화작업 요청
2008년 1월	• 3차원 비디오 부호화를 위한 테스트 데이터 요청
2008년 4월 ~ 현재	• 탐색 실험(exploration experiments) 수행
2009년 2월	• 3차원 비디오 부호화에 대한 비전 발표
2010년 1월 ~ 2010년 7월	• 테스트 데이터 선정
2010년 4월 ~ 2010년 7월	• 제안요청서(Call for Proposal) 작성
2010년 10월	• 제안요청서 배포 (예정)

2008년 4월의 MPEG 그룹 미팅에서는 4개 기관에서 제공한 10개의 테스트 영상과 두 가지 기준 소프트웨어를 수집했고, 이를 이용하여 선행 실험(exploration experiment)을 진행했다.

몇 번에 걸친 선행 실험을 통해 깊이 영상 생성 소프트웨어와 영상 합성 소프트웨어의 성능이 상당히 개선되었다. 하지만, 아직도 3차원 비디오 부호화를 본격적으로 시작할 수 있을 정도의 정확한 깊이 영상을 얻지 못해 2009년 1월에 만족할 만한 정확도를 가지는 깊이 영상을 제공해 달라고 요청했다. 이에 대한 응답으로 2009년 4월 MPEG 미팅에서는 반자동 깊이 추정방법을 포함하는 다양한 깊이 영상 추정 기술들이 제안되었다.

2010년 1월부터 4월까지 합성한 중간영상의 화질을 고려하여 적합한 테스트 영상을 선별하고 있다. 3차원 비디오 부호화의 표준화 작업을 위해 필요한 테스트 환경을 정하는 작업을 거친 후, 2010년 10월에 제안요청서를 배포할 예정이다. 3차원 비디오 부호화 기술의 제안요청서가 배포되면 여러 나라의 많은 연구기관에서 치열한 기술 경쟁을 벌일 것으로 전망된다.

#### 4. 결론

미래의 멀티미디어 서비스는 고화질 영상뿐만 아니라 3차원 비디오로 이루어 질 것이다. 3차원 영상은 2차원 영상보다 실감나고 현실감 있는 장면을 재현할 수 있기 때문에 큰 매력을 가지고 있다. 본 논문에서는 두 대 이상의 카메라로 획득한 영상을 시점간 상관도를 이용하여 효율적으로 부호화하는 기술을 소개했다. MPEG 표준화 그룹에서는 이를 바탕으로 다시점 비디오 부호화 기술을 포함하는 보다 포괄적이고 유용한 3차원 비디오 부호화 방법을 만들고 있다. 이러한 표준은 앞으로 다가올 3차원 영상 서비스 산업의 핵심 기술이 될 것으로 전망된다.

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었다.[NIPA-2010-(C1090-1011-0003)]

#### 참고 문헌

- [1] 호요성, 김성열, "3차원 TV와 실감방송", 방송과 기술, pp.130-143, 2008. 04.
- [2] 이봉호, 엄기문, 이현, 허남호, 김진웅, "3DTV 방송기술 동향", 방송공학회지, 제13권 제1호, pp.4-15, 2008. 03.
- [3] 호요성, 오관정, "다시점 비디오 기반의 3차원 실감방송", 방송과 기술, 2010. 05.
- [4] 호요성, 오관정, "3차원 비디오 부호화 기술", 한국통신학회지 정보와 통신, 제27권 제3호, pp.29-35, 2010. 03.
- [5] 호요성, 김성열, "3DTV 3차원 입체영상 정보처리", 두양사, 2010.
- [6] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video", N10357, Feb. 2009.
- [7] ISO/IEC JTC1/SC29/WG11, "Report on 3DAV Exploration", N5878, 2003. 07.
- [8] ISO/IEC JTC1/SC29/WG11, "Description of Core Experiments in MVC", N8019, July 2006.
- [9] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Multi-view Video Coding and 3DTV, vol. 17, no. 11, pp.1461-1473, Nov. 2007.
- [10] ISO/IEC MPEG&ITU-T VCEG, "CE11: Illumination Compensation", JVT-U052, Oct. 2006.
- [11] ISO/IEC MPEG&ITU-T VCEG, "CE10: Multi-view Video Coding using View Interpolation Method", JVT-V080, Jan. 2007.
- [12] ISO/IEC MPEG&ITU-T VCEG, "Disparity Vector Prediction Methods in MVC", JVT-V040, Oct. 2006.
- [13] [http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/depth\\_estimation/DERS/DERS](http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/depth_estimation/DERS/DERS)
- [14] [http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/view\\_synthesis/VSRS](http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/view_synthesis/VSRS)