

이질적인 협업환경을 위한 시공간적 계위를 이용한 스테레오 비디오 압축

(Stereo Video Coding with Spatio-Temporal Scalability for Heterogeneous Collaboration Environments)

오 세 찬 [†] 이 영 호 ^{**} 우 운 택 ^{***}
(Sehchan Oh) (Youngho Lee) (Woontack Woo)

요 약 본 논문에서는 MPEG-2에서 정의한 시공간적 계위를 확장하여 enhanced Access Grid (e-AG) 상의 서로 다른 디스플레이 시스템 및 네트워크 환경을 가진 수신 단말의 효율적인 디스플레이를 위한 새로운 3차원 영상의 부호화 방법을 제안한다. 제안된 부호화기는 3차원 영상에 대해서 시공간적 기본 해상도 및 고 해상도의 영상을 표현 가능하도록 여러 계층의 비트 열을 생성한다. 생성된 비트 열은 수신 단말의 네트워크 환경 및 처리속도, 디스플레이 환경에 따라 다른 시공간적 해상도를 표현할 수 있도록 선택적으로 전송된다. 제안된 시공간적 계위를 이용한 3차원 영상 부호화기는 이질적인 분산환경에서의 효율적인 3차원 비디오 전송 서비스에 활용 가능하다.

키워드 : e-AG, 이질적인 수신단말, 시공간적 계위, 3차원 영상 부호화

Abstract In this paper, we propose a new 3D video coding method for heterogeneous display systems and network infrastructure over enhanced Access Grid (e-AG) using spatio-temporal scalability defined in MPEG-2. The proposed encoder produces several bit-streams for providing temporally and spatially scalable 3D video service. The generated bit-streams can be delivered with proper spatio-temporal resolution according to network bandwidths and processing speeds, visualization capabilities of client systems. The functionality of proposed spatio-temporal scalability can be exploited for construction of highly scalable 3D video service in heterogeneous distributed environments.

Key words : e-AG, heterogeneous system, spatio-temporal scalability, 3D video coding

1. 서 론

최근, 인터넷 및 멀티미디어 서비스의 발전과 더불어 멀리 떨어져 있는 사용자들 사이에 몰입감을 제공하는 협업환경이 가능해 졌다. 일례로, e-AG는 사용자가 몰입감을 느낄 수 있는 협업환경을 제공하기 위해, 기존의 2D 영상 기반의 Access Grid (AG)에 3차원 및 파노라믹 디스플레이를 추가로 결합한 시스템이다[1,2]. 따라서 e-AG 사용자는 원격지에 있는 상대방의 얼굴을 마주

보면서 회의에 필요한 여러 자료들을 공유할 수 있을 뿐만 아니라, 3차원 혹은 파노라마 영상과 같은 고화질의 실감미디어를 추가로 제공받음으로써 더욱 협업환경에 몰입할 수 있다. 하지만 이러한 실감 영상 미디어의 가용성은 전적으로 네트워크 인프라 및 수신 측 디스플레이 환경에 달려있다. 예를 들어, 고화질의 비디오는 대용량의 허용 대역폭을 가지는 네트워크 환경과 고화질의 디스플레이 시스템을 가지는 수신 단에서만 서비스 제공이 가능하다. 또한 그림 1과 같이, 현재의 네트워크 인프라 및 수신 측의 3차원 디스플레이 장치들은 상당히 이질적이다. 따라서 다양한 네트워크 허용 대역폭 및 이질적인 수신 측 디스플레이 환경을 고려한 새로운 3차원 영상 압축기법이 필요하다.

이질적인 시스템과 네트워크 환경 하에서 스케러블 비디오(scalable video) 전송을 위한 기법으로는 계층적 부호화(layered coding) 또는 단계적 부호화(hierarch-

· 본 연구는 한국학술진흥재단(KRF-2002-003-D00221)의 지원 하에 수행됨

[†] 비 회 원 : 광주과학기술원 정보통신공학과
soh@gist.ac.kr

^{**} 비 회 원 : 광주과학기술원 정보통신공학과
ylee@gist.ac.kr

^{***} 종신회원 : 광주과학기술원 정보통신공학과 교수
wwoo@gist.ac.kr

논문접수 : 2003년 12월 29일

심사완료 : 2004년 7월 15일



그림 1 이질적인 분산 시스템과 실감 미디어의 서비스

ical coding) 기법이 있다. 이들 기법은 하나의 비디오 영상을 여러 다른 시각적 중요성을 가지는 성분으로 분할하여 영상 데이터 전송에 유연성(flexibility)을 제공한다.

MPEG-2에서 정의한 여러 계위(scalability)를 시작으로 현재까지 많은 3차원 영상 부호화와 관련된 연구의 대부분이 MPEG-2 멀티뷰 프로파일(multiview profile)에서 정의한 여러 계위들을 이용한 변형된 형태의 압축기법을 제시하고 있다[3-7]. 이들 기법에서는 주로 2차원 기반의 여러 계위들을 3차원 영상 부호화에 적합하도록 변형하고 좌우 영상의 상관관계(양안차)를 이용한 압축효율의 개선에 초점을 맞추고 있다. 특히, 최근에는 양안차 검색의 복잡도를 줄이기 위한 다 해상도 기반의 양안차 예측기법과, 예측의 정확도를 높이기 위한 움직임 벡터 예측에서와 같이 고정된 블록 크기를 이용하는 것 대신 가변블록 기반의 양안차 예측기법에 대한 연구들이 진행되고 있다[8-11]. 하지만 이질적인 네트워크 환경 및 수신측 디스플레이를 고려한 3차원 비디오 압축기법에 관한 연구는 아직까지 미흡한 실정이다.

본 논문에서는 e-AG상의 이질적인 수신 단말들의 효율적인 디스플레이 환경을 제공하도록, 시공간적 계위를 이용한 3차원 영상 부호화 방법을 제안한다. 효율적인 3차원 영상의 전송과 재현을 위해서, 제안된 시스템은 하

나의 기본계층(Base Layer: BL)과 여러 향상계층(Enhancement Layer: EL)을 제공한다. 기본계층은 입력된 기준영상(좌 영상)에 대한 기본 해상도를 위한 계층이며, 향상계층에서는 기본 해상도의 우 영상과 고 해상도의 좌우 영상을 제공하기 위한 추가적인 정보를 제공한다.

제안된 부호화 시스템에서는 전송할 3차원 비디오 영상을 여러 계층으로 나누어 부호화하며, 네트워크 허용 대역폭에 따라서 전송할 영상의 시공간적 해상도를 결정한다. 한편, 복호화기에서는 전송 받은 여러 계층의 비트 열을 선택적으로 결합함으로써, 자신의 디스플레이 환경에 맞도록 디스플레이 한다. 시간적 해상도는 다른 픽처의 복부호화에 참조가 되지 않는 B-픽처를 BL과 EL로 분할함으로써 구현 가능하다. 이러한 계층적 부호화는 이질적인 수신 단말의 처리 및 디스플레이 능력, 네트워크 환경을 고려하여 필요한 계층의 비트 열만 보냄으로써 효율적인 3차원 영상의 전송이 이루어지도록 할 수 있다. 또한, 일반적으로 공간적 계위에 따르는 오버헤드를 제안된 부호화기에서 양안차 예측을 수행함으로써 줄일 수 있음을 보였으며, 각 계층의 효율적인 비트율 분할 방법에 대해 실험을 통해 증명하였다.

본 논문의 구성은 다음과 같다. 2장에서 제안 시스템의 전반적인 개요 및 구현 사항들을 다룬다. 3장에서는 제안된 시공간적 계위를 이용한 3차원 영상 복부호화에 대해 설명한다. 4장에서는 제안된 부호화 시스템에 적합한 3차원 영상 전송 서비스 방법에 대해 언급한다. 5장에서는 실험을 통한 제안된 시스템을 분석하며, 6장에서는 개선 및 보완되어야 할 기술적인 문제점에 대해 간략하게 기술한다.

2. 스테레오 영상 획득, 처리 및 재현을 위한 시스템 구성

그림 2에서와 같이, 이질적인 분산협업환경에서의 효율적인 3차원 비디오 전송을 위해서 획득된 스테레오

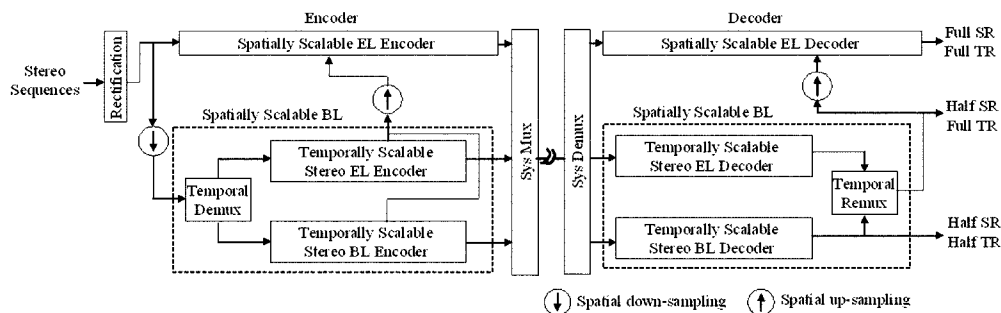


그림 2 제안된 스테레오 비디오 복부호화기의 구조, SR(Spatial Resoultion), TR(Temporal Resolution)

영상에 여러 시공간적 해상도를 지원 가능하도록 복부호화기를 디자인한다. 우선, 스테레오 영상을 3차원 카메라를 이용하여 획득한다. 획득된 영상은 카메라 렌즈 왜곡으로 인해서 정확한 양안차 예측이 난해하다. 이를 교정하기 위한 영상보정작업(rectification)이 필요하다. 실제 영상보정작업은 카메라로부터 3차원 영상 획득하는 일련의 과정에 포함된다. 가변적인 공간적 해상도를 위해서 입력받은 좌우 영상을 공간적 간축기(spatial decimator)를 이용하여 기본 해상도의 영상을 생성한다. 생성된 기본 해상도의 좌 영상은 기존의 비계위 MPEG-2 부호화기를 통해 부호화되며, 우 영상은 시간적으로 동일한 좌 영상과의 양안차를 이용하여 부호화된다. 또한, 고해상도의 좌(우) 영상을 제공하기 위한 추가적인 데이터를 위해서 이미 부호화 되어진 기본해상도의 영상을 다시 복호화한 후에 이를 다시 공간적 보간기(spatial interpolator)를 이용하여 본래의 해상도를 갖는 좌(우) 영상을 생성한다. 공간적 상향 부호화기에서는 원래의 좌(우)영상과 함께 공간적 보간기로부터 얻어진 좌(우)영상을 참조하여 고해상도 영상을 제공하기 위해 추가적으로 필요한 데이터만을 추출하여 부호화한다. 또한 가변적인 시간적 해상도를 위해서 B-픽처에 해당하는 비트 열을 각각 시간적 계위를 위한 BL과 EL로 분할하여 생성한다.

복호화기에서는 입력받은 비트 열로부터 우선 기본해상도의 좌 영상을 복호화 한다. 복호된 좌 영상은 우 영상을 복호화 하는데 참조된다. 또한, 공간적 계위 상향 계층 부호화기에 입력되어 고해상도의 좌 영상을 복호화 하는데 사용된다. 마찬가지로 복호화 된 기본 해상도의 우 영상은 공간적 상향계층에 입력되어 고해상도의 우 영상을 복호화 하는데 사용된다. 복호화기에서 이러한 계층간의 결합은 수신측 표현 가능한 시공간적 해상도에 따라서 적절히 선택적으로 이루어진다.

3. 시공간적 계위를 이용한 스테레오 비디오 복부호화 시스템

제안된 시스템의 3차원 영상 복부호화는 크게 (1) 공간적 계위 복부호화와 (2) 3차원 비디오 복부호화 (3) 시간적 계위 복부호화로 나눌 수 있다.

3.1 시공간적 계위의 구성

이질적인 디스플레이 환경을 가지는 시스템들 간에 효율적인 3차원 영상 전송을 위하여, 그림 3과 같이 하나의 BL 이외에 추가로 여러 EL을 정의한다. BL은 입력된 기준영상(좌 영상)의 기본 해상도에 해당하는 비트 열이며, EL1은 시간적으로 고해상도의 기준영상을 제공하기 위해 추가적인 비트 열이다. EL2는 기준영상에 대해 공간적 고해상도를 제공하기 위한 추가적인 비트 열

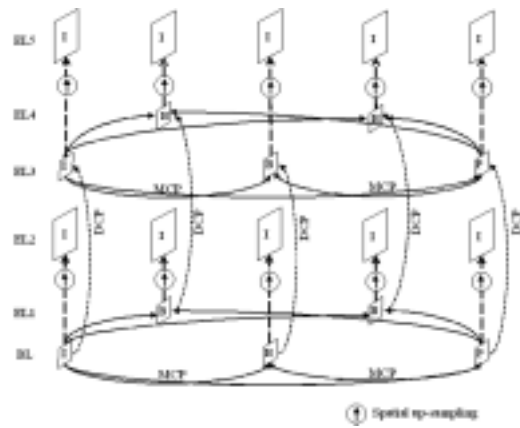


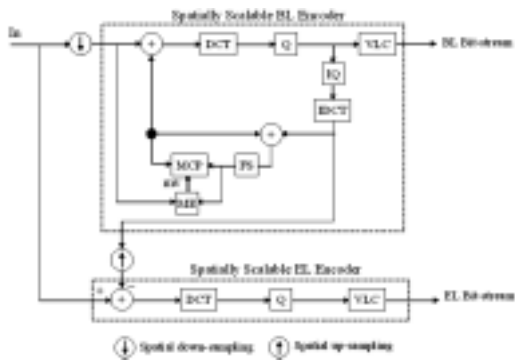
그림 3 시공간적 계위 구성

을 표현한다.

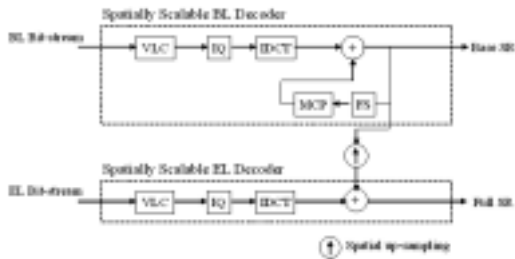
마찬가지로, EL3는 기본해상도에 해당하는 우 영상을 위한 비트열을 생성하며, EL3은 고화질의 우 영상을 위한 추가적인 비트열을 생성한다. BL에서는 시간적으로 근접한 자신의 계층에 존재하는 프레임으로부터 움직임 보상예측(Motion Compensation Prediction: MCP)을 기반으로 부호화가 수행되며, EL2는 고해상도의 좌 영상을 제공하기 위한 추가적인 데이터를 다른 예측과정 없이 인트라 부호화 과정을 수행한다. EL3에서는 보다 높은 압축 효율을 위해서 시간적으로 일치하는 BL의 영상으로부터 양안차보상예측(Disparity Compensation Prediction: DCP)과 함께, 시간적으로 근접한 자신의 계층에 존재하는 영상들로부터 MCP를 수행한다. 마찬가지로, EL4와 EL5에서는 각각 시간적 공간적 고해상도의 우 영상을 제공하기 위한 추가적인 데이터를 표현한다.

3.2 공간적 계위 복부호화

공간적 계위 부호화기는 그림 4(a)와 같이, 기본 해상도를 가지는 기준영상에 고해상도를 지원하기 위한 추가적인 비트 열을 생성한다. BL 부호화기에 입력된 원영상은 각각 수직 수평 방향으로 하향 샘플링 된 후, 기존의 MPEG-2 부호화기를 이용하여 부호화 된다. 한편, 향상계층은 원 영상과 복호화된 BL영상을 다시 원영상의 크기로 상향 샘플링한 영상과의 차분 영상(residual image)을 부호화 한다. 차분 영상은 일반적인 영상의 특성을 가지지 않는다. 따라서 움직임 예측이나 움직임 보상의 수행 없이 인트라 부호화를 수행한다. 차분 영상에서의 각 픽셀의 밝기 값 분포는 "0"의 값을 중심으로 라플라시안(Laplacian) 분포를 갖는다. 따라서 대부분의 블록마다 "0"을 중심으로 적은 변화량을 갖는다. 특정 문턱 값 보다 적은 변화량을 가지는 블록은 양자



(a) 부호화기(Encoder)



(b) 복호화기(Decoder)

그림 4 공간적 계위 복부호화기 구조

화 레벨에 따라서 부호화 없이 건너뛰므로써 보다 많은 변화량을 가지는 블록에 대해서 많은 비트 열을 할당하도록 한다.

공간적 계위 복호화 과정은 그림 4(b)에서 보여진 것처럼 부호화 과정의 반대이다. EL의 영상을 복호화를 위해서는 시간적으로 일치하는 BL 영상의 복호화가 선행되어야 한다. 복호화된 BL의 좌(우) 영상은 기본 해상도에 해당하는 디스플레이를 가지는 수신 단에서 그대로 재현이 가능하다. 예를 들어, 수신 단에서 스테레오 영상의 디스플레이를 위한 장비가 구비되어 있지 않고, 고해상도에 해당하는 영상을 위한 디스플레이 시스템이 지원하지 않는 경우, 기본 해상도의 좌(우) 영상만을 복원하여 디스플레이 하는 편이 효율적이다. 하지만, 고해상도의 디스플레이 시스템을 갖춘 수신측 디스플레이를 위해서는 이를 다시 원 영상의 크기로 복원한 후, 고화질의 영상을 위한 추가적인 정보인 상향계층의 영상과 결합하여 디스플레이 한다.

3.3 3차원 비디오 복부호화

제안된 시스템의 3차원 비디오 부호화기의 구조는 그림 5(a)과 같다. 3차원 비디오 부호화는 한 쌍의 스테레오 영상이 양안차를 따라 유사성이 있다는 점을 이용한다. 기준영상(좌 영상)은 MPEG-2 비계위 부호화기를 통해서 부호화 된다. 반면, 상향계층의 입력인 우 영상

의 압축을 위해서는 현재의 입력된 영상과 시간적으로 일치하는 참조 계층의 복호화 된 좌 영상을 참조하여 양안차 벡터를 구한다. 구한 움직임 벡터 또는 양안차 벡터 중 전송효율이 높은 쪽을 선택한 후, 그에 따라서 움직임 차이 정보(Displaced Frame Difference: DFD) 혹은 양안 차이정보(Displacement Compensated Difference: DCD)를 부호화 하여 전송한다.

해당 이미지에서 하나의 픽셀에 대한 DCD는 수식 (1)과 같이 정의된다.

$$DCD(x, y, dv) = I_R[x][y] - I_L[x + dv][y] \tag{1}$$

위 식에서 I_R 과 I_L 은 각각 우 영상과 좌 영상에서의 픽셀 밝기 (intensity) 값이며, dv 는 양안차 벡터이다. 양안차 벡터 dv 는 수식 (2)와 같이 정의된다.

$$dv(x, y) = \arg \min_{dv \in S} |DCD(x, y, dv)| \tag{2}$$

위식에서 S는 윈도우 탐색 영역을 의미한다. 실제 영상에서 양안차 벡터를 구하는 방법은 수식 (3)처럼 각 블록에 대해 SAD (Sum of Absolute Difference)를 구하여 SAD 값이 가장 작은 블록을 매칭 블록으로 정의한다. 양안차 벡터는 SAD 값이 최소가 되는 좌표(x, y)로 정의한다.

$$dv(x, y) = \min_{dv=dv_{\min}}^{dv_{\max}} \sum_{i=-S/2}^{S/2} \sum_{j=-S/2}^{S/2} |I_R[x+i][y+j] - I_L[x+i+dv][y+j]| \tag{3}$$

제안된 3차원 비디오 부호화기에서의 부호화 방법의 특징은 아래와 같다.

- 기준영상(좌 영상)은 일반적인 MPEG-2 비계위 부호화기를 통해서 부호화 된다.
- 우 영상에서의 I-픽처는 시간적으로 동일한 좌 영상으로부터 구한 양안차 벡터를 이용하여 예측된 프레임(predicted frame)을 이용하여 부호화 한다.
- 우 영상에서의 P-픽처는 동일한 시퀀스 내에서 순방향 예측(forward prediction)을 하여 얻어진 움직임 벡터와 시간적으로 동일한 좌 영상을 참조하여 얻어진 양안차 벡터, 그리고 이 둘을 보간하여 얻어진 보간 벡터(interpolated vector)를 이용하여 얻어진 각각의 SSD 중에서 최소값을 가지는 벡터를 선택하여 부호화 한다.
- 우 영상에서의 B-픽처는 그림 6에서와 같이 동일한 시퀀스 내에서 순방향 움직임 예측 및 역방향 움직임 예측을 하여 얻어진 프레임과 양안차 예측을 통해서 얻어진 프레임을 보간하여 얻어진 결과 영상을 이용하여 부호화 한다. 이때 SSD에 따라서 순방향 또는 역방향 혹은 양안차 예측인지 아니면 이들 벡터를 보간한 예측인지를 결정한다.

우 영상에서 I-픽처의 경우 각 매크로 블록마다 기준 블록과 가장 유사한 블록을 탐색하기 위해 SAD 뿐만 아니라 차의 제곱합(Sum of Squared Difference: SSD)을 구한다. 이때 구해진 SSD의 크기와 정해진 문턱값(threshold value)과 비교하여 인트라 부호화를 수행 할 것인지 양안차를 이용한 부호화를 수행 할 것인지를 판단하게 된다.

P-픽처의 경우 보간 벡터가 최소 SSD 값을 가지는 경우는 움직임 예측된 프레임(motion predicted frame)과, 양안차 예측된 프레임(disparity predicted frame)을 각각 50%의 가중치를 두어 보간 되어진 결과 영상을 이용하여 부호화 한다. 예측된 P-픽처는 수식 (4)와 같이 표현될 수 있다.

$$\hat{P}(v_f, v_d) = w_f \tilde{P}_f(v_f) + w_d \tilde{P}_d(v_d) \quad (4)$$

위 식에서 P, v, w 는 각각 매크로블록, 벡터, 가중치 팩터를 나타낸다. 그리고 인덱스 f, d 는 각각 순방향, 양안차를 나타내며, \tilde{P} 는 재 생성된(reconstructed) P-픽처에서의 매크로블록을 나타낸다.

B-픽처의 경우 보간벡터를 이용하여 예측된 픽처는 수식 (5)에서와 같이 표현 가능하다.

$$\hat{B}(v_f, v_b, v_d) = w_f \tilde{B}_f(v_f) + w_b \tilde{B}_b(v_b) + w_d \tilde{B}_d(v_d) \quad (5)$$

위 식에서 B, v, w 는 각각 매크로블록, 벡터, 가중치 팩터를 나타낸다. 인덱스 f, b, d 는 각각 순방향, 역방향, 양안차를 나타내며, \tilde{B} 는 재 생성된 B-픽처에서의 매크로블록을 나타낸다. 가중치 w_f, w_b, w_d 에 대한 값은 각각 25%, 25%, 50%로 할당하였다.

움직임 벡터의 경우와 마찬가지로 양안차 탐색영역이 실제 픽처 안에서의 양안차보다 작다면 양안차 보상작업은 비효과적이며 화질 또한 저하된다. 반대로 너무 넓은 양안차 탐색 영역을 지정한다면 양안차 벡터의 전송을 위한 비트의 할당이 과도하게 되고 결과적으로 픽처의 질 또한 낮아진다. 따라서 효율적인 양안차 탐색을 위해서는 실험에 사용된 스테레오 카메라 특성을 기반의 양안차 탐색 영역을 새로 지정해 줄 필요가 있다. 효율적인 양안차 예측을 위해서는 프레임에서 양안차 벡터 특성을 고려한 탐색영역(Region Of Disparity: ROD)을 설정해 주어야 한다. ROD는 수식 (6)을 이용하여 구할 수 있다.

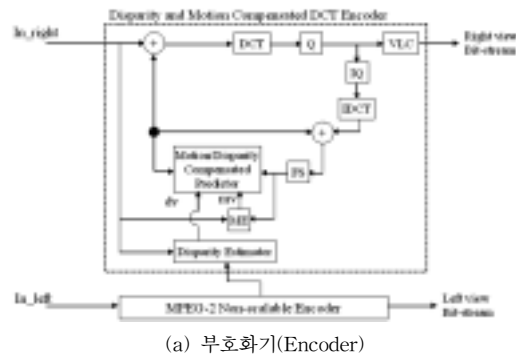
$$disparity = \frac{base_line \times focal_length}{depth} [cm] \quad (6)$$

일반적으로 양안차는 카메라 렌즈와 가까워질수록 증

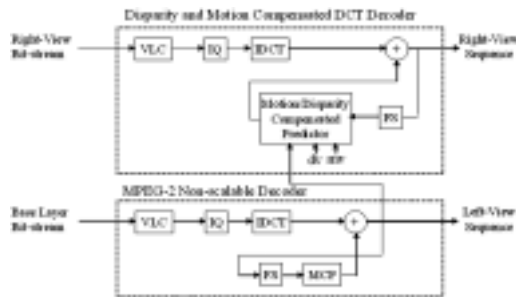
가하며, 반대로 멀어질수록 감소한다. 따라서 해당 깊이에 기반한 양안차를 정의 할 필요가 있다. 앞의 식에서 구한 양안차 값을 영상에 적용하려면 수식 (7)을 이용하여 픽셀단위로 변환한다.

$$pixel_size = \frac{CCD_array_width}{horizontal_resolution} [cm / pixel] \quad (7)$$

그림 5(b)는 3차원 비디오 복호화기 구조를 보여준다. EL2에서 우 영상은 시간적으로 일치하는 BL의 좌 영상을 참조하여 기본 해상도의 우 영상을 복호화 한다. 복호화기에서와 마찬가지로 양안차 벡터를 이용하여 복원된 좌 영상으로부터 양안차 보정을 수행하며, 움직임 벡터를 이용하여 시간적으로 이웃하는 우 영상으로부터 움직임 보정을 함께 수행함으로써 우 영상을 복원할 수 있다.



(a) 부호화기(Encoder)



(b) 복호화기(Decoder)

그림 5 3차원 비디오 복부호화기 구조

3.4 시간적 계위 복부호화

시간적 계위 부호화기는 해당 네트워크 허용 대역폭에 맞도록 프레임 레이트를 변화시킴으로써 때로는 부드러운 움직임 혹은 거친 움직임을 갖도록 할 수 있다. 일반적으로 시간적 계위에서 2 단계의 서로 다른 해상도 즉, BL과 하나의 EL을 제공하기 위해서는 I-픽처와

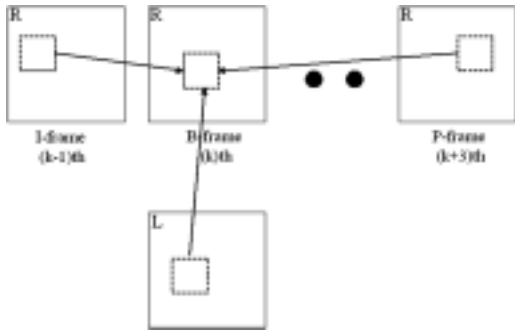


그림 6 우 영상에서의 B-frame의 예측, M=4

P-픽처 사이에 홀수개의 B-픽처가 필요하다[12]. 시간적 계위 부호화기는 그림 2와 그림 3에서와 같이, 복호화 하는데 비교적 많은 시간 복잡도가 요구되는 B-픽처를 시간적 디믹스(temporal demux)에서 기본계층과 향상계층으로 분할한다. 분할된 계층의 비트 열은 공간적 계위와 마찬가지로 3차원 비디오 서비스를 담당하는 서버로부터 네트워크 및 수신 단말의 상태에 따라서 적절히 제공받을 수 있다.

시간적 계위 복호화 과정은 그림 2와 같이 수신 단말기의 네트워크 및 디스플레이 환경에 따라 적절히 제공받은 시간적 계위 비트 열들을 시간적 리믹스(Temporal Remux)에서 적절히 결합한다. 시간적 리믹스 과정에서는 최종적으로 디스플레이 되는 3차원 영상의 시간적 해상도를 결정하며, 이는 다시 공간적 및 3차원 비디오 부호화기를 거쳐 디스플레이 환경에 맞는 3차원 영상의 공간적 해상도를 결정한다.

4. 3차원 비디오 전송 및 재현

제안된 압축 기법은 멀티캐스트 기반의 3차원 비디오 스트리밍 서비스를 목표로 한다. 만약 하나의 서버 S가 N개의 클라이언트에 대해서 3차원 비디오 서비스를 제공하는 경우, 다중 end-to-end 유니캐스트는 같은 비디오 스트리밍 데이터라 하더라도 전송 채널을 각각 할당하여 네트워크 허용 대역폭을 효율적으로 사용하지 못한다. 하지만 멀티캐스트의 경우에는 이러한 중복성을 없앨 수 있다. 멀티캐스트 중간노드는 등록된 클라이언트의 수신 상태를 미리 알고 있으며, 수신된 스트림의 최종 목적지 클라이언트의 개수 만큼 복사하여 전송한다.

그림 7에서는 3차원 비디오 스트리밍 서비스의 예를 보여준다. 비디오 서버는 스테레오 카메라로부터 3차원 비디오를 획득한 후, 부호화 하여 비디오 스트리밍 저장 공간에 저장한다. 해당 비디오 스트리밍에 대해서 서비스를 제공해야 할 경우, 각각의 계층의 비트열을 해당 클라이언트로 전송한다. 예를 들어, 클라이언트 A는 저

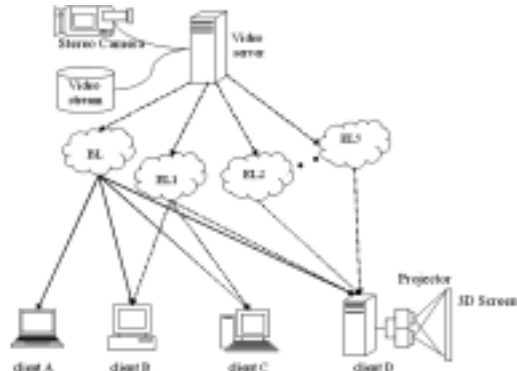


그림 7 3차원 비디오 스트리밍 서비스

해상도의 비디오 데이터만 필요한 경우이며, 기본 계층에 해당하는 비디오 비트 열을 제공한다. 반대로 클라이언트 D는 고해상도의 3차원 데이터가 필요한 경우이며, 이 경우 모든 계층의 비트 열을 제공한다. 따라서 클라이언트는 자신의 디스플레이 및 네트워크 환경에 맞는 비트 열을 제공받을 수 있으며, 클라이언트 환경에 따라서 하나의 3차원 비디오 소스에 대해서 모노 또는 여러 가지 스테레오 디스플레이가 가능하다.

5. 실험 및 비교

본 실험에서 테스트를 위한 스테레오 영상을 획득하기 위해서 스테레오 카메라인 *Bumblebee*TM[13]을 이용하여 최대 해상도 640×480의 좌우 영상을 획득한다. 표 1에서는 본 실험에서 사용한 스테레오 카메라의 자세한 사양을 보여준다. 이를 기반으로 양안차 예측에 필요한 벡터 탐색영역을 구할 수가 있다. 수식 (6), (7)을 이용하여 피사체가 카메라로부터 최대 5m, 최소 1m에 위치한다고 가정할 때, 양안차 값은 각각 11 픽셀, 54 픽셀임을 알 수 있다.

실험에 사용된 스테레오 영상은 순차방식의 640×480의 해상도를 가지는 컬러 영상이며, 초당 24프레임의 프레임 레이트를 가진다. 또한 YUV 비율은 4:2:0을 사용하였다. GOP는 16 프레임이며, B-픽처의 개수는 GOP(Group Of Picture)의 75%를 차지한다. 그림 8(a), (b)와 8(c), (d)는 각각 카메라로부터 얻은 테스트 시퀀스

표 1 스테레오 카메라 *Bumblebee*TM의 하드웨어 사양

Imaging Device	1/3" progressive scc4CDs
Frame rate	30, 15, 7.5, 3.75 frame/sec
Resolution	640 x 480
Focal Length	6mm
Base Line	12cm

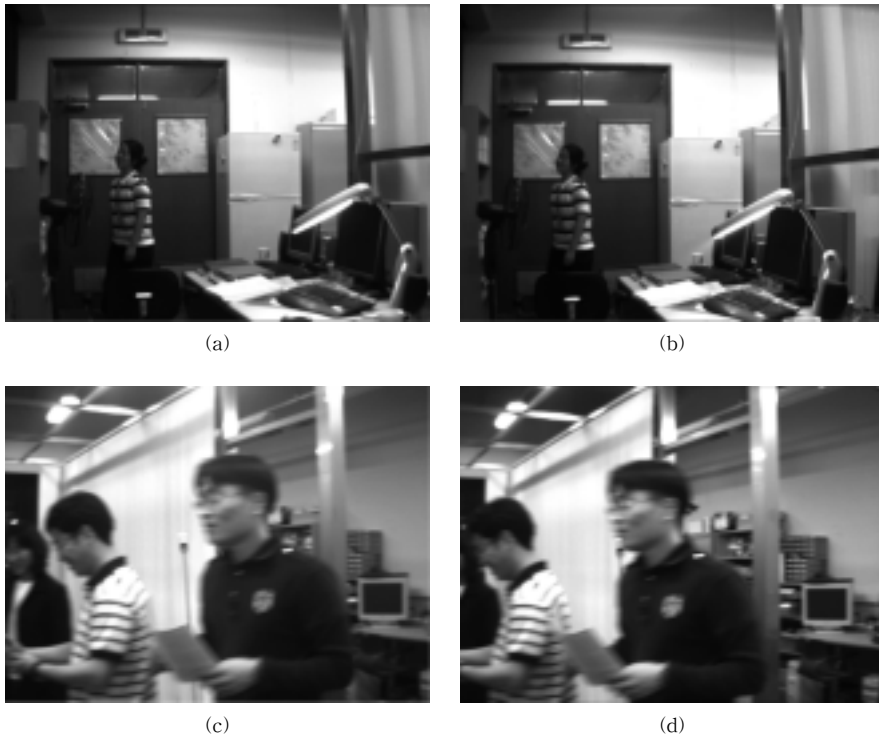


그림 8 테스트 시퀀스, (a),(b)Laboratory, (c),(d)UbiHome

인 Laboratory와 UbiHome 시퀀스 중에서 임의의 프레임에 해당하는 각각의 좌우 영상을 보여준다. Laboratory에 비해, UbiHome 시퀀스는 카메라 뷰의 이동을 비롯하여 움직임이 많다.

표 2에서는 제안된 스테레오 영상 복부호화기를 이용하여 부호화를 수행한 결과를 보여준다. 테스트 시퀀스는 Laboratory를 사용하였다. 하나의 GOP에 해당하는 각각의 좌우 영상의 밝기 성분에 대한 평균(Average) PSNR 값 및 좌우 영상 모두에 대한 Mean PSNR 값을 구하였다. 좌우 영상에 할당된 목표 비트율은 각각 1 Mbps이며, 좌우 기준 영상([BL+EL1]과 [EL3+EL4])의 목표 비트율은 각각의 좌우 영상에 할당된 목표 비트율의 40%이다. 복호화된 기본해상도의 영상을 본래의 해상도로 보간 한 후에 공간적 EL의 추가적인 정보를 더하여 고해상도의 영상을 얻을 수 있다.

실험에 사용된 테스트 시퀀스 Laboratory는 같은 목표 비트율을 가지고 각각 독립적으로 부호화를 수행할 지라도 좌우 영상의 차이 때문에, 좌 영상의 PSNR 값이 약 1 dB 정도 높다. 하지만 우 영상 부호화에 양안차 예측 및 보상을 수행하였을 때, 오히려 우 영상의 PSNR 값이 좌 영상 보다 약 1.3dB 정도 높음을 알 수 있다. 우 영상의 향상계측과 결합하여 생성된 영상

(EL3+EL4+EL5) 또한 양안차 예측을 수행함으로써 독립적으로 부호화 한 경우보다 약 1 dB정도 높은 화질의 고해상도 영상을 얻을 수 있다.

표 3과 표 4에서는 실험 결과를 비계위 MPEG-2 부호화기와 비교한 값을 보여준다. MPEG-2 비계위 부호

표 2 프로그램 수행 결과

	BL+EL1 (320x240)	BL+EL1+EL2 (640x480)	EL3+EL4 (320x240)	EL3+EL4+EL5 (640x480)
Average MSE for a GOP	9.31	7.06	6.96	7.38
Average PSNR for a GOP [dB]	38.30	39.84	38.71	39.45
Mean PSNR for left and right [dB]	39.34			

표 3 테스트 시퀀스 Laboratory에 대한 공간적 계위 오버헤드

		Left	Right
Single Layer Encoder (Non-scalable MPEG-2 Encoder)	Bitrate [Mb]	1.0	1.0
	Average PSNR [dB] for luminance	41.03	40.25
	Mean PSNR [dB]	40.62	
Proposed Scalable Stereo coder	Bitrate [Mb]	1.16	1.16
	Average PSNR [dB] for luminance	40.89	40.28
	Base layer bitrate [Mb]	0.46	0.46
	Base layer bitrate as percent of total bitrate [%]	39.7	39.7
	Spatial scalability overhead [%]	18	
	Mean PSNR [dB]	40.65	

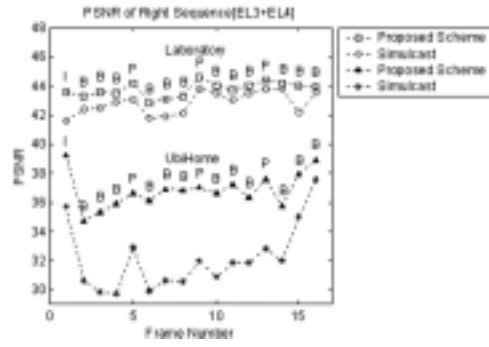
표 4 테스트 시퀀스 *UbiHome*에 대한 공간적 계위 오버헤드

		Left	Right
Single Layer Encoder	Bitrates [Mb]	1.0	1.8
	Average PSNR [dB] for luminance	36.50	36.33
	Mean PSNR [dB]	36.42	
Proposed Scalable Stereo Coder	Bitrates [Mb]	1.845	1.045
	Average PSNR [dB] for luminance	36.83	36.67
	Base layer bitrate [Mb]	0.41	0.41
	Base layer bitrate as percent of total bitrate [%]	39.2	39.2
	Spatial scalability overhead [%]	4.5	
	Mean PSNR [dB]	36.44	

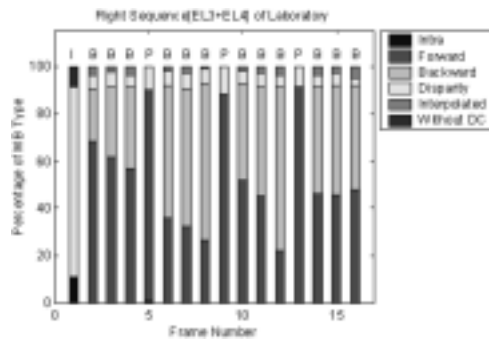
화기의 경우, 좌우 영상에 대한 비트 율은 각각 1Mbps로 고정 할당하였다. 공간적 계위에 대한 오버헤드는 제안된 부호화기를 통해 비계위 부호화기를 통해 얻어진 Mean PSNR 값과 같은 PSNR을 얻기 위해서 추가적으로 필요한 비트 율로 정의 할 수 있다. 제안된 부호화기는 3차원 비디오 부호화기에 공간적 계위를 결합하여 공간적 계위에 필요한 오버헤드를 줄일 수 있다. 즉, 좌영상을 참조함으로써 얻어지는 압축 효율이 결국 추가적으로 필요한 비트 율을 감소시킨다.

표 3의 경우는 움직임이 비교적 적은 *Laboratory*에 대한 공간적 계위 오버헤드를 보여준다. 이 경우 대부분의 블록마다 변화량이 적은 움직임 벡터를 선택하기 때문에 양안차 예측을 함으로써 얻어지는 압축 효율이 적다. 반면, 표 4에서 테스트 시퀀스 *UbiHome* 경우는 블록마다 움직임 변화량이 크기 때문에 움직임 벡터 대신, 양안차 벡터를 이용하여 부호화를 수행하게 된다. 따라서 양안차 벡터를 사용하지 않는 기존의 비계위 부호화기와 비교하였을 때, 얻어지는 압축 효율은 상당히 크며, 결과적으로 공간적 계위에 따르는 오버헤드도 크게 줄일 수 있음을 알 수가 있다.

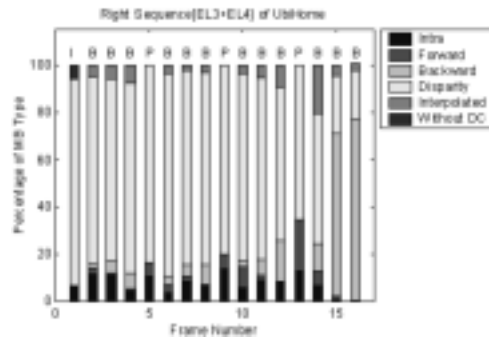
그림 9에서는 좌우 영상에서 공간적 계위 BL에서 움직임에 따르는 양안차 예측의 정도와 그에 비례하여 얻어지는 압축 효율성을 보여준다. 그림 9(a)는 하나의 GOP에 해당하는 오른쪽 영상을 각각 양안차 예측을 한 경우와 그렇지 않은 경우에 대한 PSNR 값을 *Laboratory*와 *UbiHome* 시퀀스에 대해서 수행한 결과를 보여준다. 그림 9(b)는 *Laboratory*에 각 픽처들의 블록이 어떠한 형태로 부호화를 수행하였는지 퍼센트로 나타내었으며, 그림 9(c)는 *UbiHome*에 대한 경우를 보여준다. *Laboratory*의 경우, 오른쪽 영상은 양안차 예측보다는 상대적으로 변화량이 적은 움직임 벡터를 통해 부호화가 수행되었다. 결과적으로 양안차 예측을 하지 않은 경우와 비교하여 볼 때 얻어지는 압축 효율도 적다. 반면, 움직임 변화량이 큰 *UbiHome*의 경우에는 대부분



(a)



(b)



(c)

그림 9 움직임 정도에 따른 양안차 예측 정도와 부호화 효율성, (a) 양안차 예측의 경우와 그렇지 않은 경우의 PSNR 비교, (b) *Laboratory*에 대한 매크로 블록 타입, (c) *UbiHome*에 대한 매크로 블록 타입

의 블록이 양안차 예측을 통해서 부호화 되었으며, 얻어지는 압축효율도 양안차 벡터를 참조한 블록의 수에 비례한다. 그림 9(b) 및 그림 9(c)의 Without DC의 경우는 DV가 0인 경우로 참조영상의 매크로블록과 같은 위치의 값을 그대로 사용하는 매크로 블록을 나타낸다.

하나의 입력 시퀀스를 여러 계위로 분할하여 부호화하는 경우, 비록 할당된 전체 비트율이 같더라도 각각의 계층에 할당되는 비트율의 변화에 따라서 부호화 효율은 달라지게 된다. 따라서 각각의 계층에 어떠한 비율로 제한된 비트율을 할당하는가에 대한 실험을 하였다.

표 5에서는 3차원 부호화기에서 좌우 영상에 할당되는 비트율의 변화에 따른 Mean PSNR 값을 보여준다. 이 경우 공간적 계위에 대한 영향을 고려하지 않는다. 따라서 좌우 영상에 대한 계층은 공간적 계위의 기본 계층인 [BL+EL1]과 [EL3+EL4]에 해당된다. 전체 좌우 영상에 할당된 비트율은 2 Mbps이다. 좌우 계층에 같은 비율의 비트율을 할당할 때보다는 우 영상의 부호화를 위해서 참조되는 좌 영상의 계층에 약 7:3 혹은 6:4 정도의 높은 비트율을 할당하는 경우에 더 높은 압축 효율을 얻을 수 있음을 알 수 있다.

표 6과 표 7에서는 공간적 계위 부호화기에서 각 계위에 할당되는 비트율의 변화에 따른 Mean PSNR 값을 보여준다. 표 6에서 전체 비트율이 1 Mbps일 경우 항상 계층보다는 공간적 계위 기본 계층에 약 7:3 정도로 높은 비트율을 할당하는 경우, 같은 비율로 비트율

을 할당하는 경우보다 더 높은 압축 효율을 얻을 수 있음을 알 수 있다. 반면, 표 7에서와 같이 전체 비트율이 6 Mbps인 경우, 오히려 향상계층에 더 높은 비트율을 할당하는 것이 가장 높은 압축 효율을 갖는다. 일반적으로 참조되는 영상에 더 높은 비트율을 할당하는 경우, 그렇지 않은 경우보다 높은 압축 효율을 얻는다. 하지만, 공간적 계위의 경우 기본 계층에 해당하는 데이터의 크기는 향상계층에서 표현하는 데이터 크기보다 작기 때문에 상대적으로 적은 비트율을 가지고도 상대적으로 높은 압축율을 가지며, 빨리 충전 상태에 이르게 된다. 따라서 전체 비트율이 낮은 경우 전체 Mean PSNR은 기본 계층에 영향을 받게 되지만, 전체 비트율이 충분히 높은 경우 기본 계층보다는 향상 계층에 높은 비트율을 할당 하는 것이 더 효율적임을 알 수 있다.

표 8에서는 각각의 계층에서 생성된 비트 열의 크기를 보여준다. 표에서는 전체 비트율이 4Mbps인 경우에 가장 효율적인 비트 분할을 이용하여 부호화 하였다.

그림 10은 제안된 스테레오 영상 압축기법을 적용한 수신 측 환경에서 복원 가능한 스테레오 영상의 종류를 보여준다. 그림 10(a),(b)는 수신 측에서 3차원 영상 디스플레이를 사용하지 않을 경우에 기본해상도 혹은 고해상도의 좌 영상에 해당하는 계층만을 복호화하여 얻어진 결과 있다. 그림 10(c),(d),(e)는 수신 측에서 계층간 결합 방법에 따라 복호화 가능한 3차원 영상의 종류를 보여준다. 따라서 수신 측에서는 단지 여러 계층간의 결합 방법을 통하여 자신의 디스플레이 환경에 맞는 스테레오 혹은 모노 영상의 재현이 가능하다.

표 5 3차원 부호화기 계층의 비트 율 분할에 따른 Mean PSNR 값의 변화

Left[BL+EL1]/Right[EL3+EL4]	1.4/0.6 Mbps	1.2/0.8 Mbps	1/1 Mbps	0.8/1.2 Mbps	0.6/1.4 Mbps
Sequence					
Laboratory	43.02	43.85	43.83	43.81	42.74
UrbHans	34.52	34.49	34.13	33.48	32.84

표 6 공간적 계위 계층의 비트 율 분할에 따른 Mean PSNR 값의 변화(전체 비트 율: 1Mbps)

Base[BL+EL1, EL3+EL4]/Enhance[EL2, EL5]	0.35/0.15 Mbps	0.5/0.2 Mbps	0.25/0.25 Mbps	0.2/0.3 Mbps	0.15/0.35 Mbps
Sequence					
Laboratory	39.38	39.68	34.69	33.91	33.23
UrbHans	32.38	32.32	32.34	32.36	32.32

표 7 공간적 계위 계층의 비트 율 분할에 따른 Mean PSNR 값의 변화(전체 비트 율: 6Mbps)

Base[BL+EL1, EL3+EL4]/Enhance[EL2, EL5]	4.2/1.8 Mbps	3.6/2.4 Mbps	3/3 Mbps	2.4/3.6 Mbps	1.8/4.2 Mbps
Sequence					
Laboratory	43.34	43.99	46.10	48.32	46.58
UrbHans	41.34	41.92	42.23	42.31	42.32

표 8 각 계층에 생성된 비트 열의 크기

Sequence	BL	EL1	EL2	EL3	EL4	EL5
Bit rate [Mbps]	0.7	1.4	0.5			1.4
Laboratory [Bytes]	48,650	79,595	255,750	66,128	35,418	255,866
UrbHans[Bytes]	63,560	66,491	252,827	58,285	33,744	255,272

6. 결론 및 추후연구

본 논문에서 제안한 시스템은 e-AG상에 존재하는 여러 이질적인 네트워크 환경 및 디스플레이 환경을 가지는 수신 단말들 사이에 효율적인 3차원 영상 서비스를 위해서 시공간적 계위를 이용한 3차원 영상 부호화 방법을 제안하였다. 또한, 다양한 실험을 통해서 제안된 시스템의 효율성을 보였으며, 구현된 부호화 및 복호화 시스템은 스테레오 영상에 대해서 여러 시간적, 공간적 해상도를 제공하여 이질적인 수신단말들 사이에 여러 형태의 3차원 영상 서비스가 가능하도록 하였다. 따라서 제안한 스테레오 영상 부호화 방법은 기존 e-AG상에서의 3차원 영상 서비스의 한계를 극복하는 데 기여할 수 있을 것으로 판단된다. 앞으로 공간적 계위 부호화기에 입력 영상인 차분 값을 위한 새로운 양자화기 개발과 더불어 보다 효율적인 양안차 예측의 개선이 이루어진다면 향상된 압축효율을 기대할 수 있다.

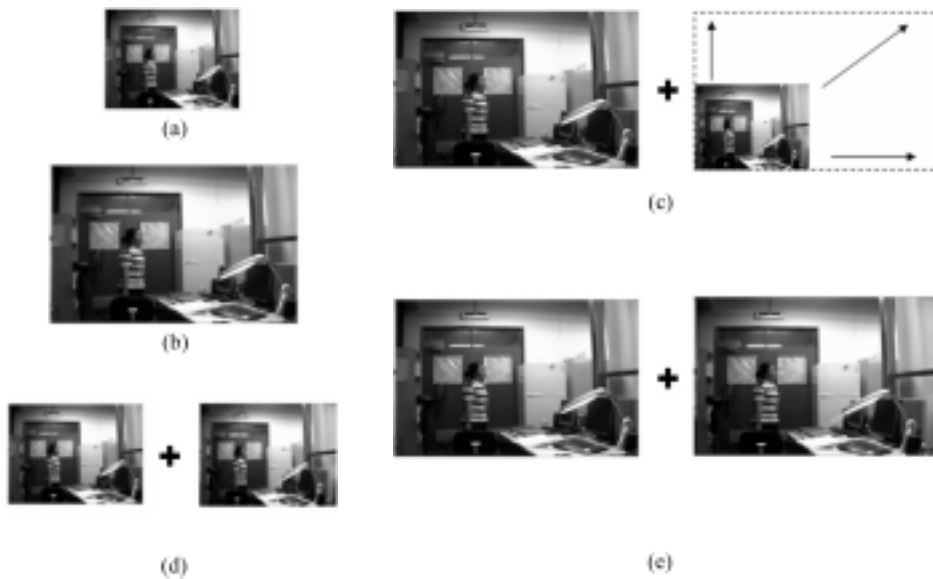


그림 10 재현 가능한 수신측 디스플레이. (a) 기본해상도의 좌 영상 (b) 고해상도의 좌 영상 (c) 고해상도의 좌 영상과 기본해상도의 우 영상 (d) 기본 해상도의 좌우 영상 (e) 고해상도의 좌우 영상

참 고 문 헌

[1] Y.Lee, S.Oh, Sh.Lee, W.Woo "Sharing 3D Media with Enhanced Access Grid(e-AG)," in Proc. of IEEK, Nov. 2003.

[2] Access Grid Homepage, <http://www-fp.mcs.anl.gov/fl/accessgrid/>.

[3] B. Haskell, A. Purl, A. N. Netravali, "Digital Video: An Introduction to MPEG-2," New York: Chapman & Hall, 1997.

[4] M. Domanski, A. Luczak, S. Mackowiak, "On Improving MPEG Spatial Scalability," ICIP Proc. 2000.

[5] A. Luczak, S.Mackowiak, M. Domanski, "Spatio-Temporal Scalability Using Modified MPEG-2 Predictive Video Coding," Proc. Signal Processing, 2000.

[6] Y. Song, "Improved Disparity Estimation Algorithm with MPEG-2s Scalability for Stereoscopic Sequences," IEEE Tr. on CE, vol. 42, no. 3, Aug. 1996.

[7] B. L. Tseng and D. Anastassiou, "Multiviewpoint Video Coding With MPEG-2 Compatibility," IEEE Tr. on CSVT, vol. 6, no. 4, Aug. 1996.

[8] S.Sethuraman, M.W. Siegel, A.G. Jordan, "A multiresolutional region based segmentation scheme for stereoscopic image compression," Proc. of the IS&T/SPIE Symp. on Electronic Imaging, Digital Video Compression- Algorithms and Technologies 1995, Vol. 2419, 1995.

[9] S. Sethuraman, M.W. Siegel, A.G. Jordan, "A

multiresolution framework for stereoscopic image sequence compression," Image Processing, 1994. Proceedings ICIP-94., IEEE International Conference, Volume: 2, 13-16 Nov. 1994, Page(s): 361-365 vol.2

[10] Sriram Sethuraman, M. W. Siegel, and Angel G. Jordan. "Segmentation based coding of stereoscopic image sequences," In Proceedings of the 1996 SPIE/IS+T Conference (San Jose), page TBD, Bellingham WA, January 1996.

[11] M. W. Siegel, S. Sethuraman, J. S. McVeigh, and A. G. Jordan. "Compression and interpolation of 3D-stereoscopic and multi-view video," In Proceedings of the SPIE, volume 3012, pages 227-238, Feb. 1997.

[12] M. Narroschke, "Functionalities and Costs of Scalable Video Coding for Streaming Services," Proc. of 36th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, Nov. 2002.

[13] PointGrey Homepage, <http://www.ptgrey.com/>.



오 세 찬

2002년 강원대학교 정보통신공학과(학사). 2002년~현재 광주과학기술원 정보통신공학과 석사과정. 관심분야는 신호처리, 3차원 비디오, HCI, 컴퓨터 비전, 혼합현실



이 영 호

1999년 한국과학기술원 수학과(학사).
2001년 광주과학기술원 정보통신공학과
(석사). 2002년~현재 광주과학기술원 정
보통신공학과 박사과정. 관심분야는 네트
워크 가상환경, 유비쿼터스 컴퓨팅, HCI
등

우 운 택

정보과학회논문지 : 소프트웨어 및 응용
제 31 권 제 2 호 참조