

H.264 기반 다시점 비디오 부호화를 위한 양자화 계수 결정 방법

준회원 박 필 규*, 정회원 호 요 성*

Quantization Parameter Selection Method For H.264-based Multi-view Video Coding

Pil-Kyu Park* Associate Member, Yo-Sung Ho* Regular Member

요 약

최근 다시점 비디오를 효율적으로 부호화하기 위해 다양한 시공간 예측 구조들이 제안되었다. 이 논문에서는 다시점 비디오 부호화 예측 구조에서 각 GOP의 첫 화면들에 삽입되는 B_{ANC} 화면을 위해 양자화 계수를 결정하는 방법을 제안한다. 각 시점의 영상 화질을 균일하게 유지하고 전체적인 부호화 효율을 높이기 위해 B_{ANC} 화면의 양자화 계수를 적응적으로 조절하며, 인접한 두 참조화면의 비트-왜곡 비용을 사용하여 상관도를 추정하고 이를 이용하여 B_{ANC} 화면의 양자화 계수를 결정한다. 이 논문에서 제안한 방법을 다시점 비디오 참조 예측 구조에 적용하여 부호화한 결과, 0.09~0.16 dB 정도의 PSNR 개선을 얻었다.

Key Words : Multi-view video coding, Inter-view prediction structure, JSVM

ABSTRACT

Recently various prediction structures have been proposed to exploit inter-view correlation among multi-view video sequences. In this paper, we propose a QP(quantization parameter) selection method for the B frame inserted in the first frames of each GOP(group of pictures), where we change QP for the B frame adaptively to achieve uniform picture quality and overall coding gain. Each B frame is coded with reference to two frames in its adjacent views. We calculate QP for the B frame based on the correlation between the two reference frames, calculated using their rate-distortion costs. By applying the proposed method to the MVC reference prediction structure, we have improved the coding gain by 0.09~0.16 dB.

I. 서 론

최근 인터넷을 통한 데이터 전송 기술이 발달함에 따라, 다양한 멀티미디어 서비스가 가능해졌으며, 대화형 콘텐츠, 실감 콘텐츠에 대한 사용자들의 요구가 증가하고 있다. 이러한 요구를 충족시키기 위한 대안으로 다시점 비디오(multi-view video)에 대

한 연구가 활발히 이루어지고 있다.

일반적인 비디오는 영상을 획득할 때 정해지는 한 방향에서의 시청만이 가능하지만, 다시점 비디오는 3차원 장면을 두 대 이상의 카메라로 촬영하여 획득하므로 임의의 방향에서의 시청이 가능하다. 다시점 비디오의 한 예인 파노라믹(panoramic) 비디오는 컴퓨터 비전, 컴퓨터 그래픽스와 같은 분야에

※ 본 연구는 광주과학기술원(GIST) 실감방송연구센터(RBRC)를 통한 정보통신부대학IT연구센터(ITRC)와 교육인적자원부 두뇌한국 21(BK21) 정보기술사업의 지원에 의한 것입니다.

* 광주과학기술원 정보통신공학과 (ppk79, hoyo@gist.ac.kr)

논문번호 : KICS2007-01-014, 접수일자 : 2007년 1월 15일, 최종논문접수일자 : 2007년 6월 15일

서 연구되고 있으며, 이러한 파노라마 비디오 기술은 항공사진 해석, 영상변화 감지, 비디오 압축 등의 다양한 분야에 응용되고 있다¹⁾.

컴퓨터 비전 기술을 사용하여 다시점 비디오를 구성하는 여러 개의 영상으로부터 시차(disparity)와 깊이맵(depth map)을 추출할 수 있으며, 두 개 이상의 영상과 3차원 디스플레이를 이용한 3차원 영상 서비스를 구현할 수 있다. 따라서, 다시점 비디오 부호화(multi-view video coding, MVC) 기술은 자유시점 TV(free viewpoint television, FTV), 3차원 TV(3D television)와 같은 응용 분야에 필수적이다. 그러나 다시점 비디오는 여러 개의 영상으로 구성되므로 처리할 데이터의 양 또한 단일 시점 비디오에 비해 많기 때문에, 데이터의 효율적인 전송과 저장을 위해 기존의 단일시점 비디오 부호화 기술보다 효율적인 부호화 기술이 요구된다²⁻⁴⁾.

다시점 비디오 부호화 기술의 표준화 작업은 MPEG (Moving Picture Experts Group)의 3DAV (3-D audio and visual) AHG (Ad Hoc Group)에서 시작되었다. 2005년 7월에 제안요청서(Call for Proposals)와 요구사항들이 발표되었으며, 시점 보간 예측(view interpolation prediction) 등에 대한 핵심 기술 성능평가 실험(core experiments)이 진행되었다^{5,6)}. 핵심기술 성능평가 실험에는 Fraunhofer-HHI에서 JSVM(Joint Scalable Video Model)을 기반으로 제작한 참조 소프트웨어가 사용된다.

보다 효율적인 다시점 비디오 부호화를 위해, 영상들 사이에 존재하는 공간 중복성을 제거해야 하며, 영상들의 화질을 균일하게 유지할 필요가 있다. 각 GOP(group of pictures)의 첫 번째 화면들로 구성된 앵커(anchor)의 효율적인 부호화를 위해, B 화면이 삽입된다. 이 앵커의 B 화면들에는 앵커의 I 화면과 P 화면보다 높은 양자화 계수(quantization parameter)가 사용된다. 따라서 앵커의 화면들은 균일한 화질로 부호화되지 못하고, 이는 각 영상 간의 화질 불균형을 발생시킨다. 이 논문에서는 영상 사이의 화질을 균일하게 유지하면서 전체적인 부호화 효율을 향상시키기 위해 B 화면 앵커의 양자화 계수를 결정하는 방법을 제안한다.

II. 다시점 비디오 부호화

2.1 다시점 비디오 부호화의 요구사항

다시점 비디오 부호화를 위해 제안되는 알고리즘들에 대한 몇 가지의 의무사항과 권장사항이 있다.

압축과 관련하여 다시점 비디오 부호화는 각 시점 영상을 독립적으로 부호화한 것보다 높은 부호화 효율을 제공해야 한다. 또한 오류에 강인하고 각 시점 영상 화질을 유연하게 조절할 수 있어야 한다.

시스템 지원과 관련하여, 각 시점 영상들은 시간적으로 정교하게 동기화되어야 하며, 효율적인 가상 시점 영상의 생성 기술이 권장된다⁷⁾.

2.2 다시점 비디오 부호화의 예측 구조

다시점 비디오는 똑같은 3차원 장면을 촬영한 여러 개의 영상으로 구성되므로, 인접한 영상 사이에 공간 중복성이 존재한다. 따라서, 추가적인 부호화 효율을 얻기 위해 시간 중복성과 공간 중복성을 동시에 제거해야 한다. 공간 중복성은 다른 시점의 영상을 참조하여 제거되므로, 보다 효율적인 시공간 예측을 위한 여러 가지 GOP (Group Of Pictures) 구조들이 제안되었다⁶⁾. 참조 소프트웨어에서 예측 구조의 구현은 설정(configuration) 파일 내의 SequenceFormatString 항목을 통해 이루어진다.

그림 1은 Fraunhofer-HHI에서 제시한 시공간 예측 구조이다. 이 구조는 다시점 비디오 부호화의 표준화를 위한 참조 예측 구조로 사용되고 있다⁸⁾.

이 구조는 시간 예측을 위해 계층적 B 화면 구조(hierarchical B picture structure)를 사용하며, 앵커에는 시간적 임의 접근과 오류의 전파 방지를 위해 0.5초 혹은 1초 간격으로 I 화면이 삽입된다. 앵커 화면들의 효율적인 부호화를 위해 S1, S3, S5 영상의 앵커 화면들은 B화면(B_{ANC} 화면)으로 부호화되며, 이들은 같은 시간상의 인접한 두 시점 영상을 참조한다. 이와 같이 앵커 화면이 B_{ANC} 화면으로 부호화되는 영상들은 시간 예측과 함께 공간 예측을 사용하여 부호화된다.

GOP 내의 각 계층에는 각기 다른 양자화 계수가 사용된다. 기본 계층인 앵커의 I 화면과 P 화

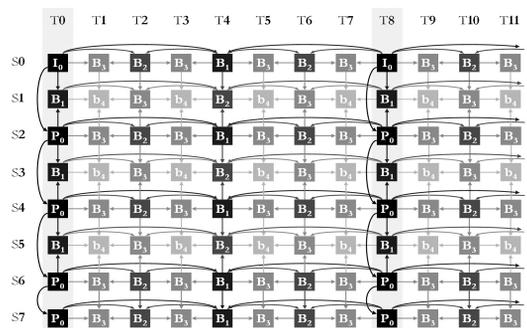


그림 1. Fraunhofer-HHI에서 제시한 예측 구조

표 1. 기본 양자화 계수

실험영상	기본 양자화 계수		
Ballroom	34	31	29
Exit	31	29	26
Uli	36	30	28
Race1	28	26	24
Flamenco2	34	30	28
Breakdancers	31	26	22
Rena	33	28	23
Akko&Kayo	36	29	24

표 2. 양자화 계수 차이값

DeltaLayer0Quant	0
DeltaLayer1Quant	3
DeltaLayer2Quant	4
DeltaLayer3Quant	5
DeltaLayer4Quant	6
DeltaLayer5Quant	7

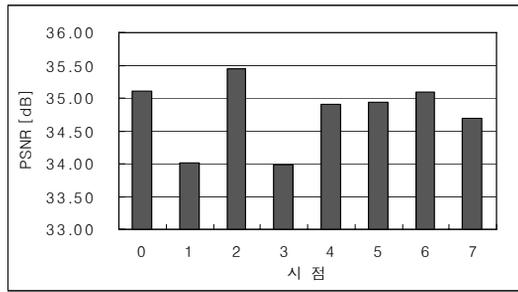


그림 2. “Ballroom” 영상의 부호화 결과

면에는 고정된 기본 양자화 계수가 사용되며, 나머지 상위 계층에는 기본 양자화 계수에 각 계층에 대한 양자화 계수 차이값(delta layer quant)을 더한 값이 사용된다. 현재 핵심기술 성능평가에 사용되는 실험영상들의 기본 양자화 계수값과 양자화 계수 차이값들을 표 1과 표 2에 제시했다.

앞에서 설명한 것처럼, 그림 1에서 I_0 화면과 P_0 화면에는 DeltaLayer0Quant, B_1 화면에는 DeltaLayer1Quant가 사용되는데, 이때 B_{ANC} 화면을 B_1 화면과 같은 계층으로 간주하여 동일한 양자화 계수 값을 할당한다. 따라서 B_{ANC} 화면은 I 화면이나 P 화면보다 낮은 화질로 부호화된다.

계층적 B 화면 구조에서는, 화면을 부호화할 때 참조하는 화면들의 화질이 낮을수록 현재 화면의 부호화 효율도 낮아진다. 따라서, 그림 1에서 S1, S3, S5 영상들과 같이 앵커 화면이 B_{ANC} 화면으로 부호화된 영상의 경우, B_{ANC} 화면을 참조하는 GOP

내부 화면들의 부호화 효율도 다른 영상의 화면들에 비해 낮아진다. 결과적으로, 영상들이 서로 다른 화질로 부호화된다.

그림 2는 “Ballroom” 영상을 기본 양자화 계수 31로 부호화한 후 각 시점별 PSNR 값을 나타낸 것이다. 1번, 3번, 5번 영상의 앵커 화면이 B_{ANC} 화면으로 부호화되었으며, PSNR 값들의 차이가 최악의 경우 1dB 이상인 것을 알 수 있다.

III. 제안한 양자화 계수 선택 방법

이 논문에서 제안한 방법은 B_{ANC} 화면에 사용되는 양자화 계수를 B_1 화면의 양자화 계수와 관계없이 독립적으로 결정하여 영상 간의 화질 차이를 줄이고 전체적인 부호화 효율을 개선시키고자 한다.

앵커의 화면들은 영상내 시간적 임의 접근을 위해 공간 예측만을 사용하여 부호화된다. 따라서 B_{ANC} 화면 역시 좌우 시점의 I 화면과 P 화면, 혹은 두 개의 P 화면을 참조하여 부호화된다.

표 3은 “Ballroom” 영상 중 3번 영상의 첫 번째 화면을 양자화 계수와 참조화면을 달리하여 부호화한 결과이다. 표 3의 상단은 3번 영상의 첫 화면을 2번과 4번 영상의 첫 화면을 참조하여 부호화한 결과이고, 표 3의 하단은 0번과 6번 영상의 첫 화면을 참조하여 부호화한 경우이다. 참조화면의 양자화 계수는 34이다. 카메라의 배치에 의해, 2번과 4번 영상은 0번과 6번 영상에 비해 서로 가까우므로 상관도 역시 높다. 이 결과를 토대로, 2번과 4번 영상을 참조할 때, 동일 양자화 계수에서 부호화 효율이 높고 더 큰 양자화 계수로 부호화해도 PSNR의 감소가 적음을 확인할 수 있다.

따라서, 참조화면 간의 상관도가 높을수록 B_{ANC} 화면의 부호화 효율이 증가하고, 화질이 양자화 계수에 덜 민감함을 알 수 있다. 반대로, 참조화면 간의 상관도가 낮은 경우, 중간 B_{ANC} 화면이 높은

표 3. “Ballroom” 영상의 실험 결과

시점	QP for $B_{ANC} = 35$		QP for $B_{ANC} = 36$	
	PSNR [dB]	Bits [bits]	PSNR [dB]	Bits [bits]
2	38.5436	58208	38.5436	58208
3	37.2056	21016	36.9255	17968
4	37.7067	42920	37.7067	42920
0	38.4674	57848	38.4674	57848
3	37.0546	33032	36.6693	28552
6	37.4065	76376	37.4065	76376

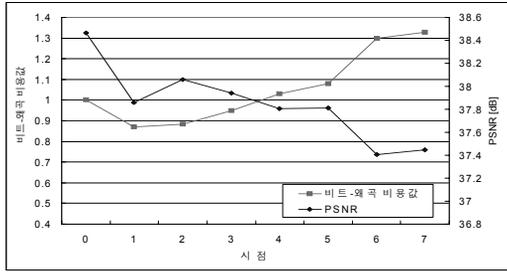


그림 3. "Exit" 영상의 부호화 결과

중요도를 가지므로 상대적으로 작은 양자화 계수를 사용하여 일정 수준 이상의 화질을 확보해야 한다. 이 논문에서 제안하는 방법은 두 참조화면 간의 상관도를 이용하여 B_{ANC} 화면의 중요도를 추정하고 이를 기반으로 양자화 계수를 결정한다.

그림 3은 "Exit" 영상의 첫 앵커 화면들을 기본 양자화 계수 31로 부호화한 결과이다. 0번 화면은 I 화면으로 부호화되었으며, 나머지 화면들은 모두 0번 화면을 참조하여 P 화면으로 부호화되었다. 각 화면의 비트-왜곡 비용값(rate-distortion cost)은 식 (1)을 사용하여 계산하는데 이들을 0번 화면의 비트-왜곡 비용값으로 나누어 정규화한 후 그림 3에 표시하였다.

$$RDCost = D + \lambda R \quad (1)$$

여기서 D는 부호화로 인해 발생한 화소당 평균 왜곡값으로, 식 (2)와 식 (3)을 통해 계산한다.

$$D = \frac{SSD(Y) + SSD(U) + SSD(V)}{width \times height \times 1.5} \quad (2)$$

$$SSD(A) = \sum_{x,y \in A} \{org(x,y) - rec(x,y)\}^2 \quad (3)$$

여기서 org(x,y)와 rec(x,y)는 각각 원본 화면과 부호화된 화면의 해당 화소값을 나타낸다. 식 (2)는 원본 화면과 부호화된 화면 간의 SSD(sum of squared difference)를 총 화소 수(Y, U, V 성분)로 나눈 값을 뜻한다.

식 (1)의 변수 R은 각 화면을 부호화한 후 발생한 비트량을 총 화소 수로 나눈 값이며, 식 (4)를 이용하여 구한다.

$$R = \frac{total\ bits}{width \times height \times 1.5} \quad (4)$$

식 (1)의 λ는 비트-왜곡에 대한 라그랑지안(Lagrangian)

계수를 나타내며, 식 (5)를 사용하여 구한다⁹⁾.

$$\lambda = 0.85 \times 2^{\min(52, QP)/3 - 4} \quad (5)$$

그림 3에서 볼 수 있듯이, 참조화면과의 거리가 멀어져 상관도가 감소할수록 PSNR이 대체로 낮아지고 비트-왜곡 비용값은 증가한다. 이를 토대로, 이 논문에서 제안한 방법은 비트-왜곡 비용값을 화면 간의 상관도를 추정하기 위한 기준으로 삼아 B_{ANC} 화면의 양자화 계수 결정에 이용한다. B_{ANC} 화면의 양자화 계수는 식 (6)을 이용하여 계산한다.

$$QP_{B_{ANC}} = \min[51, QP_{REF} + dQP] \quad (6)$$

여기서 QP_{B_{ANC}}는 B_{ANC} 화면에 사용될 양자화 계수로서, 51 이하의 값으로 제한된다. QP_{REF}는 I 화면과 P 화면에 사용될 기본 양자화 계수를 뜻하고, dQP는 B_{ANC} 화면에 대한 양자화 계수 차이값으로, 식 (7)과 식 (8)을 이용하여 얻는다.

$$dQP = \max[\lceil \sqrt{\alpha(RDCost_{ratio} - \beta)} \rceil, 0] \quad (7)$$

$$RDCost_{ratio} = RDCost_I / RDCost_P \quad (8)$$

여기서 RDCost_P는 B_{ANC} 화면의 두 참조화면 중 두 번째로 부호화되는 화면의 비트-왜곡 비용값이며, RDCost_I는 현재 B_{ANC} 화면이 포함된 앵커의 I 화면의 비트-왜곡 비용값이다. B_{ANC} 화면이 I 화면과 P 화면을 참조할 경우에는 α=2, β=1을 사용했고, 두 개의 P 화면을 참조할 경우, α=3, β=0.9를 사용했다. 식 (7)에서 α와 β의 값은 dQP가 주로 0과 3 사이의 값을 갖도록 조절되었다.

IV. 실험 결과 및 분석

이 논문에서 제안한 방법의 성능을 평가하기 위해 사용된 실험영상들의 특성이 표 4에 제시되어 있다¹⁰⁾. 기본 양자화 계수는 핵심기술 성능평가와 동일한 값들을 사용하였다.

표 5와 그림 4는 "Ballroom" 영상을 기본 양

표 4. 실험영상

실험영상	영상 특성	카메라 개수	카메라 배열
Ballroom, Exit	640×480 25fps	8	1D/ parallel
Race1	640×480 30fps	8	1D/ parallel

표 5. “Ballroom” 영상의 부호화 결과

양자화 계수	제안한 방법		참조 소프트웨어	
	비트율 [kbps]	PSNR [dB]	비트율 [kbps]	PSNR [dB]
34	272.58	33.24	269.83	33.17
31	397.37	34.87	392.61	34.78
29	511.65	35.88	503.34	35.77

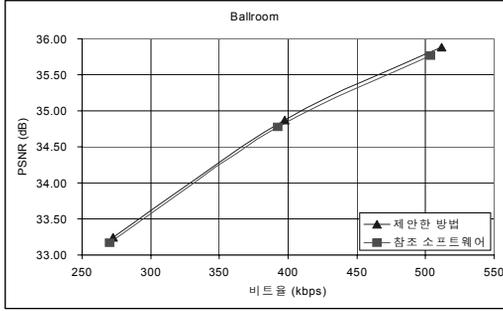


그림 4. “Ballroom” 영상의 비트-왜곡 곡선

표 6. “Exit” 영상의 부호화 결과

양자화 계수	제안한 방법		참조 소프트웨어	
	비트율 [kbps]	PSNR [dB]	비트율 [kbps]	PSNR [dB]
31	198.32	36.98	194.48	36.86
29	252.54	37.73	246.63	37.61
26	389.27	38.73	378.74	38.61

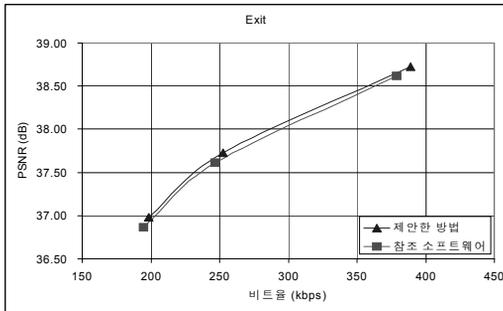


그림 5. “Exit” 영상의 비트-왜곡 곡선

자화 계수 34, 31, 29로 부호화한 결과를 나타낸다. 이 논문에서 제안한 방법이 참조 소프트웨어에 비해 유사한 비트율에서 0.09dB 정도의 PSNR 개선을 보였다.

표 6과 그림 5는 “Exit” 영상을 기본 양자화 계수 31, 29, 26으로 부호화한 결과를 나타낸다. 제안한 방법은 참조 소프트웨어에 비해 유사한 비트율에서 0.13dB 정도의 PSNR 개선을 보였다.

표 7. “Race1” 영상의 부호화 결과

양자화 계수	제안한 방법		참조 소프트웨어	
	비트율 [kbps]	PSNR [dB]	비트율 [kbps]	PSNR [dB]
28	412.57	37.61	405.12	37.46
26	548.07	38.62	536.05	38.45
24	740.26	39.65	724.22	39.50

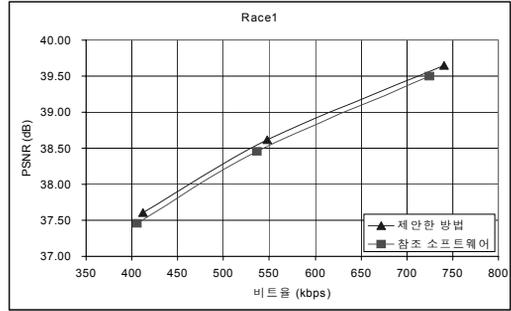


그림 6. “Race1” 영상의 비트-왜곡 곡선

표 8. 영상 별 PSNR값의 분산

실험영상	양자화 계수	제안한 방법	참조 소프트웨어
Ballroom	34	0.1606	0.2186
	31	0.1622	0.2634
	29	0.1515	0.2843
Exit	31	0.1136	0.1828
	29	0.0942	0.1590
	26	0.0766	0.1312
Race1	28	0.5100	0.5523
	26	0.4955	0.5311
	24	0.4612	0.4930

표 7과 그림 6은 “Race1” 영상을 기본 양자화 계수 28, 26, 24를 사용하여 부호화한 결과를 나타낸다. 이 논문에서 제안한 방법이 참조 소프트웨어에 비해 유사한 비트율에서 0.16dB 정도의 PSNR 개선을 보였다.

표 8은 각 실험영상에 대해 모든 시점 영상들의 PSNR 값의 분산(variance)을 나타낸다. 이 논문에서 제안한 방법은, 앵커 화면이 B_{ANC} 화면으로 부호화된 영상들에 대해, 앵커 화면을 더 고화질로 부호화하여 부호화 효율을 높인다. 따라서, 유사한 비트율에서 해당 영상들의 PSNR 값이 증가하므로, 시점 간의 화질 편차가 감소한다. 결과적으로, 이 논문에서 제안한 방법에 의해 PSNR 값들의 분산이 평균 33% 감소하였다.

V. 결론

이 논문에서는 다시점 비디오 부호화의 예측 구조에 사용된 계층적 B 화면 구조의 특성과 참조화면 사이의 상관도를 고려하여 B_{ANC} 화면의 양자화 계수를 조절하는 방법을 제안하였다. 우선 각 참조화면의 비트-왜곡 비용값을 이용하여 화면 간의 상관도를 추정하고, 이를 이용하여 B_{ANC} 화면의 양자화 계수를 결정한다. 이로써 B_{ANC} 화면의 화질을 개선하고 해당 시점 영상의 부호화 효율을 높인다. 결과적으로 이 논문에서 제안한 방법은 참조 소프트웨어를 사용한 부호화 결과에 비해 0.09~0.16 dB 정도의 PSNR 개선과 33%의 PSNR 분산 감소를 보였다.

참고 문헌

- [1] K.-T. Ng, S.-C. Chan and H.-Y. Shum, "The compression issues of panoramic video," in *Proc. International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 36-39, 2001.
- [2] ISO/IEC JTC1/SC29 WG11 N6909, "Survey of Algorithm used for Multi-view Video Coding (MVC)," Jan. 2005.
- [3] A. Smolic and P. Kauff, "Interactive 3D Video Representation and Coding Technologies," in *Proc. IEEE, Spatial Issue on Advances in Video Coding and Delivery*, vol. 93, pp. 98-110, Jan. 2005.
- [4] R.-S. Wang and Y. Wang, "Multiview Video Sequence Analysis, Compression, and Virtual Viewpoint Synthesis," *IEEE Transactions on Circuit and System for Video Technology*, vol. 10, pp. 397-410, April 2000.
- [5] ISO/IEC JTC1/SC29 WG11 N7327, "Call for Proposals on Multi-view Video coding," July 2005.
- [6] ISO/IEC JTC1/SC29 WG11 W8019, "Description of Core Experiments in MVC," April 2006.
- [7] ISO/IEC JTC1/SC29 WG11 N6501, "Requirements on Multi-view Video Coding," Oct. 2004.
- [8] ISO/IEC JTC1/SC29 WG11 M12945, "Multiview Coding using AVC," Jan. 2006.
- [9] T. Wiegand and B. Girod, "Lagrange Multiplier

Selection in Hybrid Video Coder Control," in *Proc. IEEE Int. Conf. Image Proc.*, pp. 542-545, Oct. 2001.

- [10] ISO/IEC MPEG & ITU-T VCEG U211, "Common Test condition for Multiview Video Coding," Oct. 2006.

박 필 규 (Pil-Kyu Park)

준회원



2005년 중앙대학교 전자전 기공
학부 졸업(학사)
2005년~현재 광주과학기술원 정
보통신공학과 석사 과정
<관심분야> 디지털 신호처리, 다
시점 비디오 부호화

호 요 성 (Yo-Sung Ho)

정회원



1981년 서울대학교 공과대학 전
자공학과 졸업(학사)
1983년 서울대학교 대학원 전자
공학과 졸업(석사)
1989년 Univ. of California, SB,
Dept. of Electrical and Computer
Engineering.(박사)

1983년~1995년 한국전자통신연구원
선임연구원
1990년~1993년 미국 Philips 연구소, Senior Research
Member
1995년~현재 광주과학기술원 정보통신공학과 교수
<관심분야> 디지털 신호처리, 영상 신호 처리 및 압축,
디지털 TV와 고선명 TV, 멀티미디어 시스템,
MPEG 표준, 3차원 TV, 실감방송