

H.264/AVC 스케일러블 비디오 코딩에서 빠른 부호화를 위한 단계적 모드 선택 기법

(Phase Mode Decision Scheme for Fast Encoding in H.264 SVC)

고 경 은 * 강 진 미 *
(Gyeongun Goh) (Jinmi Kang)

조 미 숙 ** 정 기 동 ***
(Misook Cho) (Kidong Chung)

요 약 멀티미디어 서비스의 다양한 재생장치와 네트워크 환경에 적용 가능한 멀티미디어 코덱을 위해 ISO/IEC MPEG와 ITU-T VCEG의 JVT팀에서는 계층구조를 가지는 H.264/AVC Scalable Video Coding(SVC)을 제안하였다. H.264/AVC SVC는 향상계층의 비트율과 왜곡에 관한 효율을 높이기 위해 계층 간 예측, 계층적 B 구조 등 다양한 기법이 적용하며 이로 인해 계산복잡도가 증가하여 부호화 시간이 길어지는 단점이 있다. 본 논문에서는 부호화 시간을 단축시키기 위해 예측모드를 분석하고 그 결과에 따라 기준 값을 선정하여, 단계적으로 예측모드를 선택하는 기법을 제안한다. 실험결과 제안하는 기법이 계산복잡도를 감소시켜 평균 38%의 부호화 시간을 단축시키며 화질의 저하는 평균 0.01dB로 무시할 수 있을 정도로 작음을 확인하였다.

키워드 : SVC, 스케일러블 코딩, 예측모드선택

- 이 논문은 2단계 두뇌한국21사업에 의하여 지원되었음
- 이 논문은 2008 한국컴퓨터종합학술대회에서 'H.264/AVC 스케일러블 비디오 코딩에서 빠른 부호화를 위한 단계적 모드 선택 기법'의 제목으로 발표된 논문을 확장한 것임

* 학생회원 : 부산대학교 컴퓨터공학과
shf7395@gmail.com
wolff98@gmail.com

** 학생회원 : 부산대학교 멀티미디어협동과정
mscho@pusan.ac.kr

*** 종신회원 : 부산대학교 컴퓨터공학과 교수
kdchung@pusan.ac.kr
논문접수 : 2008년 8월 28일
심사완료 : 2008년 10월 20일

Copyright©2008 한국정보과학회: 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제14권 제8호(2008.11)

Abstract To achieve flexible visual contents adaptation for multimedia communications, the ISO/IEC MPEG & ITU-T VCEG form the JVT to develop an SVC amendment for the H.264/AVC standard. JVT uses inter-layer prediction that can improve the rate-distortion efficiency of the enhancement layer. But inter-layer prediction causes computational complexity to be increased. In this paper, we propose a fast mode decision for inter frame coding. It makes use of the correlation between optimized prediction mode and its RD cost. Experimental results show that the proposed schemes save up to 38% of encoding time with a negligible coding loss and bit-rate increase.

Key words : SVC, Scalable Coding, Fast mode decision

1. 서론

최근 각종 디스플레이 장치의 등장과 다양한 환경의 유무선망의 등장에 따라 멀티미디어 서비스는 유무선망의 채널 용량이나 장치의 성능에 따라 다양한 스트림을 제공할 필요가 있다. 프레임율, 해상도, 화질 등이 다른 다양한 스트림을 제공하기 위한 방법은 크게 세 가지로 구분할 수 있다. 먼저 최고 상태의 스트림을 전송한 후 디코더나 라우터에서 down-sampling을 하게 하는 방법이 있다. 이 방법은 down-sampling 과정의 추가비용이 발생하므로 수신 측의 부담이 크고 채널이 낭비되는 문제가 있다. 두 번째는 다양한 상태의 스트림을 다수 전송하여 다양한 환경에 적합하게 하는 방법이 있다. 이 방법은 수신 측의 부담은 적으나 채널의 낭비가 심하고 실시간 전송의 경우 유무선망의 비트율에 적합하도록 지속적인 피드백이 필요하게 되는 문제점이 있다. 세 번째 방법은 다양한 상태를 하나의 스트림으로 구성하고 전송하는 방법으로 변환이나 실시간 인코딩 과정의 추가비용 없이 비트율에 맞출 수 있다[1]. 세 번째 방식을 지원하기 위해 ISO/IEC MPEG과 ITU-T VCEG의 JVT에서는 H.264/AVC를 확장한 SVC를 규정하였다.

SVC는 하나의 부호화된 비트스트림으로부터 다양한 해상도와 프레임율로 복호화할 수 있도록 입력영상을 계층적으로 부호화한다. 이러한 성질을 확장성이라 하며 SVC는 공간 확장성(Spatial Scalability), 시간 확장성(Temporal Scalability), 화질 확장성(Quality Scalability)을 지원한다. 공간 확장성과 화질 확장성은 하나의 압축된 비트스트림으로 각각 다양한 해상도 또는 다양한 화질을 지원하기 위해 하나의 기본계층(Base Layer)과 다수의 확장계층(Enhancement Layer)으로 구성된다. 기본계층은 H.264/AVC와의 호환을 위해 H.264/AVC의 부호화 과정을 따르며 확장계층의 부호화 과정 또한 H.264/AVC의 부호화 과정에 기반을 두고 있다. 시간 확장성은 다양한 프레임율을 지원하기 위해 계층

적 B 구조(Hierarchical B Structure)를 사용한다[2].

이러한 SVC는 다양한 상태의 스트림을 다수 전송하는 방법보다 코딩 효율(Coding Efficiency)을 높여서 보다 짧은 비트스트림으로 고품질의 영상을 제공한다. 하지만 다양한 기술을 적용함으로써 부호화 과정의 복잡도가 증가하게 되는 문제점이 발생하므로 계산복잡도(Calculate Complexity)를 감소시켜 부호화 시간을 줄이는 연구가 반드시 필요하다[5,6]. 본 논문에서는 SVC의 세 가지 확장성 중 공간 확장성을 부호화하는 과정에 소요되는 시간을 줄이는 방법에 초점을 두었다. 기존 부호화 속도를 개선하기 위한 기법으로 [3]은 기본계층과 향상계층의 QPD(Quantization Parameter Difference)가 작을수록 화질 확장성을 가지는 기본계층과 향상계층 사이의 유사도가 커지는 점을 이용하여 QPD에 따라 향상계층에서 기본계층의 예측모드를 사용하는 비율을 조절하고 기본계층의 예측모드에 따라 향상계층의 후보군을 줄이는 방법을 제시하였다. 하지만 [3]의 방법은 공간 확장성으로 계층별 영상의 해상도가 다를 경우, 양자화 계수의 차이 값만으로 계층의 유사도를 판별하기 어렵고 예측모드가 16×16이거나 SKIP인 경우 모든 모드에 대해 RDO(Rate Distortion Optimization)과정을 수행해야 하는 문제가 있다.

본 논문은 SVC의 참조 소프트웨어인 JSVM11[4]을 기반으로 부호화 과정 중 예측모드 선택의 복잡도를 감소시키는 방법을 제시한다. 2장에서는 SVC 향상계층에서의 모드 선택 방법을 소개하고 3장에서는 제안하는 기법에 대해 설명한다. 4장에서는 제안 기법을 이용한 시간 감소와 PSNR 변화량에 대한 실험결과를 제시하고 5장에서는 결론으로 논문을 끝맺는다.

2. SVC 부호화 과정의 예측모드 선택

SVC 부호화는 코딩 효율을 높이기 위해 계층적 B 구조, 계층 간 예측(Inter Layer Prediction) 등 다양한 기술이 추가되었다. 다양한 기술들에 의해 비트스트림의 압축률이 높으면서 화질을 유지하는 장점이 있지만 계산복잡도가 증가하고 부호화 시간이 길어지는 단점이 있다.

SVC 부호화 과정은 기존 H.264/AVC[2]에서 확장되었다. 먼저 원본 영상 크기와 해상도 등을 조절하여 계층별로 나누고 계층적 B 구조로 재배치한다. 각 계층별로 예측과정에 의해 잔여신호(Residual)를 추출하고 변환(Transform)과 양자화(Quantization)를 통해 비트스트림을 생성한다. 기본계층의 부호화는 최저의 화질과 해상도, 프레임율을 지원하며 H.264/AVC 부호화 방식과 동일하게 이루어진다. 향상계층의 경우 기존 H.264/AVC의 인터예측과 인트라예측뿐만 아니라 계층 간의 중복성을 줄이기 위한 계층 간 예측이 추가되었다.

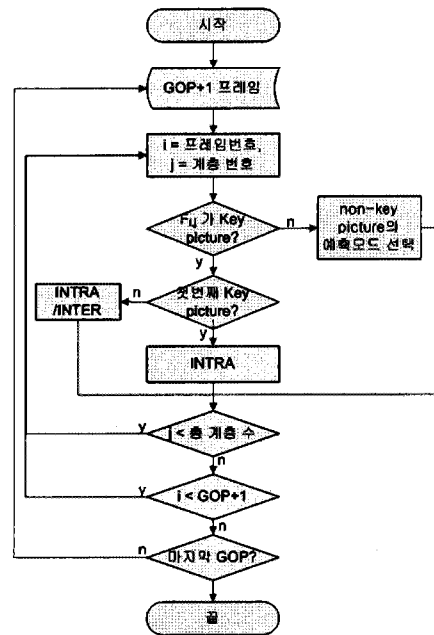


그림 1 GOP 단위 예측모드 선택

계층 간 예측은 계층 간 인트라예측, 계층 간 움직임 예측, 계층 간 잔여신호 예측으로 나누어진다. 첫 번째 계층 간 인트라예측은 부호화하려는 매크로블록에 하위 계층의 대응하는 블록이 인트라예측모드로 부호화된 경우 하위 계층의 대응 블록을 업샘플링(Up-sampling)하여 예측과정의 참조 블록으로 사용한다. 두 번째로 계층 간 움직임 예측은 하위 계층의 대응하는 블록의 움직임 벡터와 매크로블록 분할 모드를 업샘플링하여 사용하는 “Base layer mode”와 움직임 벡터를 업샘플링한 뒤 1/4 단위의 추가적인 움직임 보상과정을 거치는 “Quarter Per Refinement mode”가 있다. 세 번째 계층 간 잔여신호 예측은 잔여신호의 계층 간 유사성을 이용하는 기법으로 하위 계층에서 발생한 잔여신호를 변환한 뒤 업샘플링하여 잔여신호의 차를 부호화하는 방법이다.

부호화의 예측과정은 GOP단위로 이루어지며 프레임당 계층별로 기본계층부터 진행되고 해당 프레임이 Key picture인지 Non-Key picture인지의 여부와 계층의 종류에 따라 예측모드의 후보군이 달라진다(그림 1)[4]. 예측모드의 선택은 RDO를 계산하는 식 (1)을 이용하여 최소의 비용을 가지는 모드를 선택한다[2].

$$C = D + \lambda \times R \tag{1}$$

C: 비용
D: 예측된 왜곡 정도
R: 예측된 부호화 용
λ: Lagrange 계수

예측모드가 가장 다양한 향상계층의 Non-Key picture

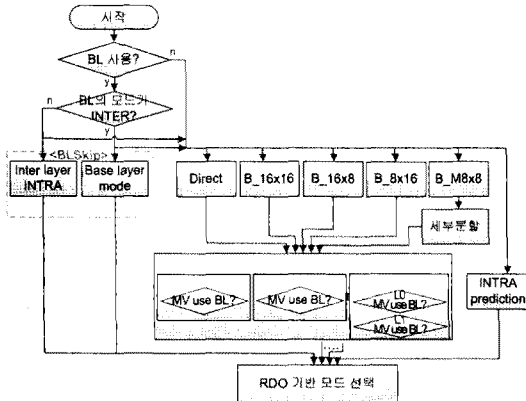


그림 2 Non-key picture의 예측모드 선택

의 예측과정은 그림 2와 같다. 부호화하려는 해당 프레임의 매크로블록과 대응하는 하위계층의 매크로블록이 존재하면 Base layer mode 또는 계층 간 인트라예측을 수행한다. 또한 같은 계층의 이웃한 움직임 벡터를 그대로 이용하는 Direct 예측, 같은 프레임 내의 픽셀을 이용하는 인트라예측, 계층적 B 구조에 따른 B 픽처 예측이 수행된다. B 픽처 예측은 매크로블록 분할 모드인 P_M16x16, P_M16x8, P_M8x16, P_M8x8 마다 움직임 벡터를 예측하기 위한 상세모드 선택을 수행한다. 상세모드 선택은 List0를 이용하는 P예측(L0), List1를 이용하는 P예측(L1), List0와 List1을 모두 사용하는 B예측(BI) 중에서 최적의 모드를 선택한다. 각 예측은 같은 계층 내에서 움직임 예측을 수행하여 움직임 벡터를 생성하는 방법과 하위 계층의 벡터를 활용함으로써 참조 범위를 줄인 움직임 예측으로 움직임 벡터를 생성하는 방법이 있다. 각 매크로블록마다 위와 같은 각 예측모드들을 수행한 뒤 RDO기반 모드 최적화를 통해 예측모드를 선정하는 방법은 계산복잡도가 높기 때문에 전체 부호화 과정 중에서 가장 많은 시간을 차지한다.

3. 제안하는 단계적 모드선택 기법

3.1 향상계층의 예측모드 분석

그림 3은 JSVM11을 기반으로 Non-Key picture인 향상계층이 하위계층의 정보를 사용할 수 있는 경우의 RD 비용별 매크로블록 예측모드의 분포를 나타낸 그래프이다. 네 가지 영상의 총 60프레임을 대상으로 GOP가 16인 시간확장성 구조와 기본계층이 QCIF, 향상계층이 CIF 해상도의 2개 계층의 공간 확장성을 사용하였다. 하위계층의 정보를 그대로 사용하는 Base layer mode와 계층 간 인트라예측을 BL_Skip으로 나타내었으며, 하위계층 정보를 사용하는 예측모드를 BL_‘블록 크기’, 하위계층 정보를 사용하지 않는 예측모드를

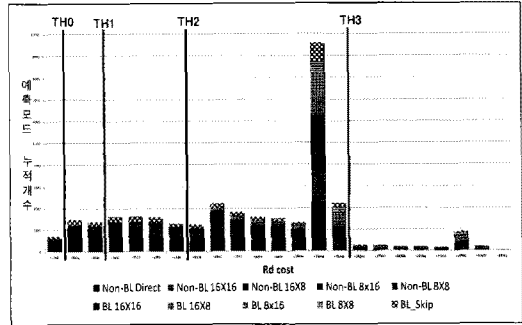


그림 3 RD 비용별 예측모드의 분포

NON-BL_‘블록 크기’로 표현하였다. 세로축은 같은 범위의 비용이 발생한 예측모드의 누적 개수를 의미하고 가로축은 각 예측모드의 RD 비용(Rate Distortion Cost)을 의미한다.

그림 3의 그래프는 RD 비용에 따라 발생하는 예측모드의 후보군을 나눌 수 있음을 보여준다. RD 비용이 기준치 TH0 이하에서는 BL_Skip 모드와 Direct 모드가 주로 발생하고 TH0와 TH1에서는 BL_Skip 모드와 Direct 모드, 하위계층을 사용하지 않는 16x16 매크로블록 단위의 예측(Non-BL 16x16)이 주로 일어나며 TH1과 TH2사이에서는 주로 BL_Skip 모드와 Direct 모드, Non-BL 16x16 모드, 하위계층을 사용하는 16x16 매크로블록 단위의 예측(BL_16x16)이 주로 선택되는 것을 알 수 있다. 또한 TH4를 초과하는 경우에는 Direct 모드의 발생빈도가 급감하는 것을 볼 수 있다.

부호화하는 현재계층(향상계층)에서 하위계층을 통해 계층간 예측 정보들을 이용하기 위하여 하위계층과 현재계층 간의 예측모드의 상관성을 살펴볼 필요가 있다. 표 1은 두 개의 공간계층을 지원하는 네 가지 영상별 Non-Key picture의 기본계층 인트라예측모드와 향상계층 인트라예측모드의 관계를 나타낸 것이다. 대상 영상이 Non-Key picture이고 큰 화면전환이 빈번히 발생하지 않는 영상으로 인트라예측의 발생 빈도가 2%이하이므로 표 1에서는 인트라예측만 나타내었다.

현재계층이 하위계층과 같은 블록 크기의 예측을 하는 경우는 약 68% 발생한다. 하위계층이 16x8, 8x16 블록 크기 예측을 수행한 경우에 현재계층이 같은 블록 크기의 예측을 하거나 더 작은 블록인 8x8 블록 크기 예측을 하는 경우는 80%가 넘고, 하위계층과 현재계층이 8x8 크기의 블록예측을 하는 경우는 약 78%의 확률을 보인다.

3.2 제안하는 단계적 모드 선택 기법

3.1장에서 RD 비용에 따라 후보군이 구분되고 하위계층의 모드와 향상계층의 모드의 상관성이 매우 높다는 것을 살펴보았다. 본 논문에서는 앞서 언급한 두 가지의

표 1 기본계층과 향상계층의 예측모드 상관성(단위 %)

영상	기본계층 모드	향상계층 모드			
		16×16 or Direct	16×8	8×16	8×8
Mobile	16×16 or Direct	68.56	11.94	11.62	7.82
	16×8	6.56	49.18	11.48	32.79
	8×16	6.06	10.10	53.54	30.30
	8×8	4.00	16.00	20.00	60.00
Habour	16×16 or Direct	72.05	5.57	12.95	9.43
	16×8	0.00	50.00	0.00	50.00
	8×16	6.67	3.33	76.67	13.33
	8×8	0.00	0.00	16.67	83.33
Foreman	16×16 or Direct	76.54	9.41	9.67	4.39
	16×8	11.11	66.67	11.11	11.11
	8×16	12.50	0.00	87.50	0.00
	8×8	0.00	0.00	0.00	100.00
Soccer	16×16 or Direct	80.17	6.56	8.75	4.51
	16×8	5.41	41.44	18.92	34.23
	8×16	10.28	10.28	53.27	26.17
	8×8	4.17	16.67	8.33	70.83

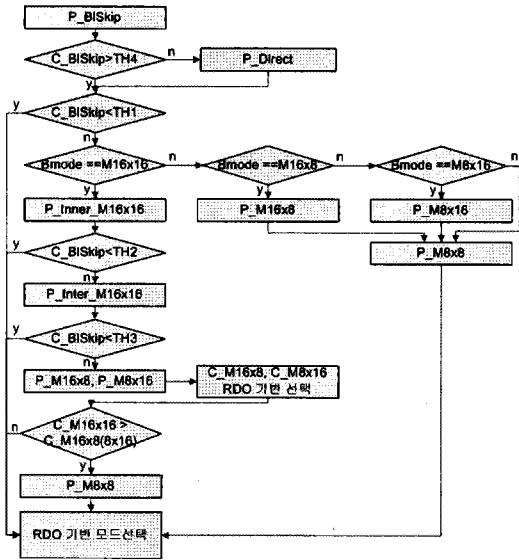


그림 4 제안하는 기법의 흐름도

특성을 이용하여 향상계층의 Non-Key picture의 예측모드를 제한하여 빠른 모드 선택 기법을 제안한다. 제안하는 기법은 다음과 같은 단계를 따른다.

• 1단계: 전체 예측모드 중 선택되는 빈도가 두 번째인 BL_Skip 모드 예측을 수행한다. BL_Skip 모드는 하위계층의 정보를 그대로 이용하는 모드이며 Base

layer mode와 계층 간 인트라예측을 포함한다.

- 2단계: BL_Skip 모드의 RD 비용(그림 4의 C_BLSkip)이 TH4 (단, TH1<TH2<TH3<TH4)보다 작은 경우 Direct 예측을 후보군에 추가한다.
- 3단계: BL_Skip 모드의 RD 비용이 TH1보다 큰 경우 하위계층의 예측모드(그림 4의 Bmode)에 따라 후보군에 예측모드를 추가한다. 표 1의 분석 결과에 따라 하위계층이 16×8(또는 8×16) 블록 크기 예측을 수행한 경우 현재계층의 부호화는 16×8(또는 8×16) 블록 크기 예측모드와 8×8 블록 크기 예측모드를 후보군에 추가하고, 하위계층이 8×8 블록 크기 예측을 수행한 경우 현재계층의 부호화는 8×8 블록 크기 예측모드를 후보군에 추가하고 7단계로 진행한다. 그 외의 경우에는 계층 내 16×16 블록 크기 예측을 후보군에 추가하고 다음 단계로 진행한다.
- 4단계: BL_Skip 모드의 RD 비용이 TH2 이상인 경우 16×16 블록 크기의 계층 간 인트라예측모드를 후보군에 추가한다.
- 5단계: BL_Skip 모드의 RD 비용이 TH3 이상인 경우 16×8, 8×16 블록 크기의 예측모드 중 RD 비용이 적은 예측모드를 후보군에 추가한다.
- 6단계: 5단계에서 선택된 예측모드가 16×16 블록 크기 예측모드 보다 RD 비용이 적은 경우 8×8 블록 크기의 예측모드를 후보군에 추가한다.
- 7단계: 후보군 중 가장 적은 RD 비용을 가지는 예측모드를 매크로블록의 가장 적합한 예측모드로 선정한다.

4. 실험결과

실험은 SVC의 참조 소프트웨어인 JSVM11을 사용하였다. 다양한 특성을 가지는 Soccer, Crew, Foreman, Ice의 네 가지 영상을 사용하였으며 각 실험환경은 표 2와 같다.

실험은 제안하는 기법에 의한 계산복잡도 감소 정도를 측정하기 위해 JSVM11과 비교하여 부호화 시간 감소량을 측정하고 코딩 효율과 화질의 감소 정도를 측정하기 위해 비트율과 PSNR 변화량을 측정하였다. 그림 5와 그림 6은 JSVM11과 제안하는 기법의 프레임별 RD 비용을 비교한 그래프이다. 비트율이 같을 때 PSNR

표 2 실험환경

공간 확장성	기본계층	QCIF
	향상계층	1개, CIF
시간 확장성	기본계층	GOP : 8
	향상계층	GOP : 16
움직임 벡터 정밀도		1/4 축소
움직임 벡터 추정범위		16
참조화면의 수		3
총 화면 수		50

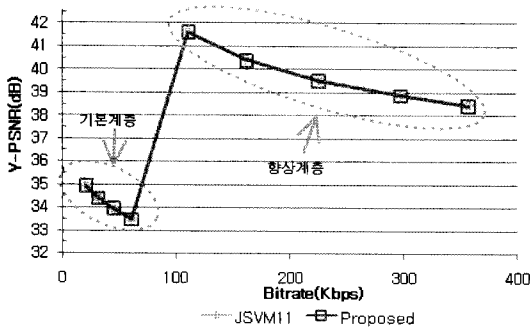


그림 5 RD 비용 그래프 (ICE)

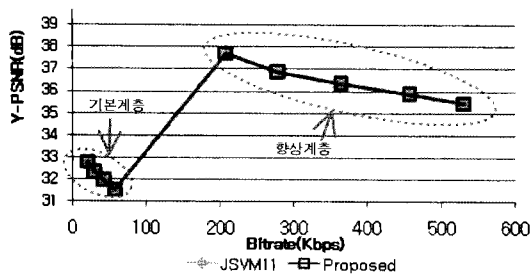


그림 6 RD 비용 그래프 (Soccer)

표 3 제안하는 기법의 결과

영상	부호화 시간 변화량(%)	비트율 변화량(%)	PSNR(dB)
foreman	-33.01	0.29	-0.02
crew	-34.97	0.08	-0.015
soccer	-25.84	0.05	-0.003
ice	-58.70	-0.08	-0.02

의 차이가 0.1dB 미만으로 화질의 차이가 거의 없음을 볼 수 있다.

표 3은 제안하는 기법을 적용한 향상계층의 Non-Key picture를 부호화할 때 계산복잡도 감소 및 RD 비용 변화이다. 부호화 시간은 최대 58.7%, 평균적으로 38%의 감소량을 보였으며 비트율과 PSNR은 각각 0.08%, 0.01dB의 차이를 보이고 있으므로 화질의 차이는 거의 없으면서 부호화 시간이 크게 줄어 든 것을 확인할 수 있다.

5. 결론

SVC는 코딩 효율을 증가시키기 위해 계층적 B 구조, 계층 간 예측 등 다양한 기법을 사용하며 이로 인해 계산복잡도가 증가하는 문제가 발생한다. 부호화 과정은 기존 H.264/AVC의 계층 내 인트라예측과 인터예측 이외에 계층 간 인트라예측, 계층 간 인터예측, 계층 간

잔여신호 예측이 추가되면서 최적 예측모드 선택에 소요되는 시간이 매우 증가하였으며 이는 곧 부호화 시간 증가에 큰 영향을 끼친다.

향상계층의 Non-Key picture의 경우 RD 비용에 따라 예측모드의 분포가 다르고 하위계층의 매크로블록 모드와 상관관계가 크다. 본 논문에서는 이러한 특성을 이용하여 부호화 시간을 줄이는 단계적인 RD 비용 기반 예측 모드 선택 기법을 적용하고 추가적으로 하위계층의 예측모드를 사용하여 예측모드의 후보군을 제한하는 기법을 제안하였다. 이를 통해 인트라 예측 부호화 시간은 약 38% 단축하면서 PSNR의 변화량이 0.01dB 정도로 원본과의 화질 차이가 거의 없는 실험 결과를 보였다.

참고 문헌

- [1] C. Andrew Segall, "Spatial Scalability Within the H.264/AVC Scalable Video Coding Extension," IEEE Trans.on CASVT, Vol.17, No.9, pp.1121-1135. September. 2007.
- [2] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), "Joint Scalable Video Model JSVM-11," JVT-X202, Geneva, July. 2007.
- [3] He Li, Z. G. Li, Changyun Wen and Shoulie Xie, "Fast Mode Decision For Coarse Granular Scalability via Switched Candidate Mode Set," IEEE International Conference on Multimedia & Expo (ICME 2007), July, 2007.
- [4] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), "JSVM 11 Software," JVT-X203, Geneva, July. 2007.
- [5] 이범식, 김문철, 함상진, 박창섭, 박근수, "공간적 스케일러블 비디오 부호화에서 계층간 모드 고속 결정 방법", 방송공학회논문지, 제 12권 4호, pp. 360-372, 2007년 7월.
- [6] 이금희, 임진영, 정경훈, 김기두, 강동욱, "H.264 스케일러블 확장 규격의 고속 부호화를 위한 향상계층의 RD-최적화된 탐색 모드 제한 기법", 방송공학회논문지, 제 13권 1호, pp. 152-161, 2008년.