

H.264/AVC 고속 모드 결정을 위해 비트율 왜곡값을 이용한 적응적인 임계값 선택 방법

황수진, 호요성
광주과학기술원 정보통신공학부
{soojin, hoyo}@gist.ac.kr

Adaptive Threshold Selection Technique using Rate-Distortion Cost for Fast Mode Decision in H.264/AVC

Hwang, Soo-Jin and Ho, Yo-Sung
Gwangju Institute of Science and Technology (GIST)

요약

H.264/AVC 부호화 표준은 영상의 특성을 반영하기 위해 16×16부터 4×4 크기의 가변적인 블록을 이용하여 부호화 효율을 높인다. 하지만 이로 인해 부호기의 복잡도가 증가된다. 부호기 복잡도를 증가시키는 여러 요인 중, H.264/AVC의 모드 결정은 부호기의 복잡도를 증가시키는 주요인이다. 본 논문에서는 IPPP구조에서 비트율 왜곡값을 이용하여 고속으로 매크로블록의 모드를 결정하는 방법을 제안한다. 인트라 화면에서의 인트라 4×4, 인트라 16×16의 비트율 왜곡 평균값으로 영상에 적응적인 최대 임계값과 최소 임계값을 결정한다. 다음, 16×16, 16×8, 8×16 인트라 모드의 비트율 왜곡값이 최대 임계값과 최소 임계값으로 분할한 범위 중 어느 곳에 해당하는지를 살펴보고, 이에 따라 인트라 모드 결정 단계를 선택적으로 결정한다. 제안하는 알고리즘은 기존의 H.264/AVC에 비해 부호화 효율의 큰 감소 없이 평균 27.42%의 부호화 시간을 감소시켰다.

1. 서론

H.264/AVC는 기존의 방법에서 부호화 효율을 높이기 위해, 4×4 블록 단위의 정수 변환 부호화, 가변 블록 움직임 보상, 방향성을 고려한 인트라 부호화 기법, 다중 참조 영상 부호화 기법, 그리고 움직임을 정확하게 예측하기 위한 1/4화소 단위의 움직임 벡터 추정 기법을 추가 도입하였다.

H.264/AVC는 그림 1과 같이 16×16에서부터 4×4까지 다양한 움직임 보상 블록을 제공한다. 이를 이용하여, 영상의 움직임 특성을 좀 더 정확하게 보상할 수 있다.

적절하게 매크로블록의 모드를 설정해야 압축 효율을 높일 수 있다. 다시 말해, 다양한 움직임 보상 블록 중에서 최적의 모드를 선택하는 것이 중요하다.

H.264/AVC에서 최적 모드를 결정하는 방법은 모든 모드에 대해 비트율 왜곡값을 계산하여 그 값이 최소인 모드를 선택하는 것으로, 이를 비트율 왜곡 최적화 기법(Rate-Distortion Optimization method)이라 한다. 비트율 왜곡 최적화 기법은 움직임 벡터 선택, 참조 영상 선택, 인트라-인트라 모드 결정에 사용하고 있다. 식 (1)은 비트율 왜곡값을 나타낸다.

$$J = \text{Distortion} + \lambda \cdot \text{Rate} \quad (1)$$

H.264/AVC는 부호화 효율을 높이기 위해 인트라 화면에서도 인트라 모드를 이용한 부호화를 수행한다. 영상의 시간적인 변화보다 공간적인 변화가 작을 경우, 인트라 모드를 사용하면 부호화 효율을 더 높일 수 있기 때문이다. 일반적으로 인트라 화면에서의 인트라의 발생 확률은 SKIP이나 인트라 모드에 비해서 아주 낮다. 하지만 인트라 화면에서 그림 1의 인트라 모드뿐만 아니라, 인트라 모드 총 13가지(인트라 4×4: 9가지, 인트라 16×16: 4가지) 모드의 대해 비트율 왜곡값을 계산하여 최적의 매크로블록 모드를 결정하면, 압축 효율을 높일 수 있다. 이로 인해 높은 부호화 효율은 얻을 수 있지만, 부호기의 복잡도가 높아진다.

본 논문에서는 비트율 왜곡값을 이용해 선택적으로 인트라 모드를 결정하는 방법을 제안한다. 인트라 화면에서 발생률이 낮은 인트라 모드의 비트율 왜곡값 특성을 이용하여 임계값을 정하고, 이를 기반으로 인트라 모드 비트율 왜곡값 계산 여부를 결정하여 모드 결정에 소요되는 시간을 줄일 수 있다.

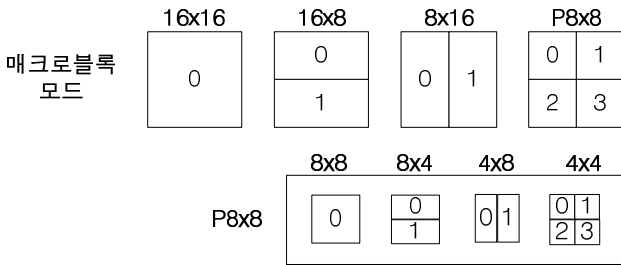


그림 1 움직임 보상 블록

압축하고자 하는 영상에서 배경과 같이 평활한 영역은 16×16과 같은 큰 블록으로 움직임을 보상하고, 상대적으로 움직임이 복잡한 영역은 4×4와 같은 작은 블록으로 움직임을 보상을 한다. 만약 큰 블록을 선택하면 보내야 할 움직임 벡터의 비트양은 줄어들지만, 잔여 데이터의 양은 늘어난다. 반대로 작은 블록을 선택하면 잔여 데이터의 양은 줄어들지만, 전송해야 할 움직임 벡터의 양이 늘어난다. 따라서 영상에

2. 비트율 왜곡(Rate distortion) 최적화 기법

H.264/AVC 매크로블록 모드는 SKIP, 16×16, 16×8, 8×16, P8×8(8×8, 8×4, 4×8, 4×4), 인트라 4×4, 인트라 16×16로 나뉜다. 이 중 SKIP, 16×16, 16×8, 8×16, P8×8 총 다섯 가지 모드는 인트라 모드에 해당하며, 인트라 4×4, 인트라 16×16 두 가지 모드는 인트라 모드에 해당한다.

인트라 모드는 시간적 상관성을 고려하여 화면 간 움직임 예측을 통해 매크로블록을 부호화한다. 그에 반해 인트라 모드는 공간적 상관관계를 이용하여 매크로블록을 부호화한다. 일반적으로 인트라 모드 중 복잡도가 높은 부분은 인트라 4×4를 이용하여 부호화하며, 이를 효과적으로 부호화하기 위해 9가지의 모드로 이루어져있다. 이것에 비해 변화가 적은 부분은 인트라 16×16로 부호화하며, 이는 4가지 모드로 이루어져있다.

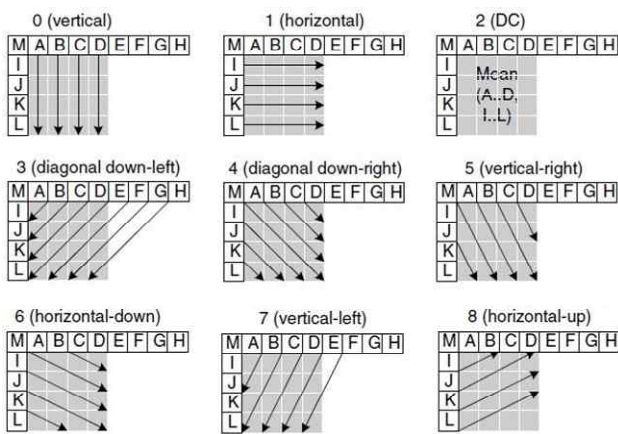


그림 2 인트라 4×4의 9가지 예측 모드

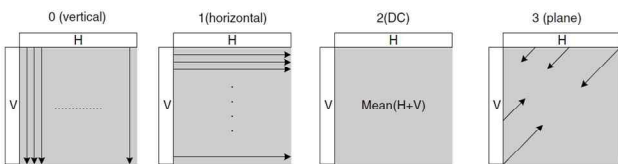


그림 3 인트라 16×16의 4가지 예측 모드

H.264/AVC는 부호화 효율이 가장 좋은 최적 모드를 선택하기 위해 각각의 모드에 대해 비트율 왜곡값을 계산한다. 최적 모드 결정을 위한 비트율 왜곡값 계산식은 식 (2)와 같다.

$$J_{mode}(s, r, M | \lambda_{mode}) = SSD(s, r, M) + \lambda_{mode} \cdot R(s, r, M) \quad (2)$$

where, $\lambda_{mode} = 0.85 \times 2^{(QP-12)/3}$,

$$SSD(s, r, M) = \sum_{x \in H, y \in V} |s(x, y) - r(x - m_x, y - m_y)|^2$$

여기서, J_{mode} 는 비트율 왜곡값을, M 은 모드, $R(s, r, M)$ 은 부호화할 때 발생하는 실제 비트량을 의미한다. SSD (Sum of Squared Difference)는 원본 블록(s)과 복원된 블록(r)의 차이값의 제곱의 합이다. 복원된 영상이란, 변환되고 양자화 된 후, 다시 역양자화 되고, 역변환된 영상을 의미한다. 비트율 왜곡값은 최적 모드를 결정하기 위해 모든 사용 가능한 모드에 대해 계산하므로, 결과적으로 부호화기의 복잡도가 높아진다.

3. 선택적 인트라 모드 생략법

IPPP 구조로 영상을 부호화할 때, 인트라 화면에서 인트라 모드 블록은 인트라 모드 블록보다 발생 빈도가 현저히 낮다. 그림 4는 “paris_cif”와 “akiyo_cif” 영상의 모드 분포를 보여준다. 그림 4에서 노란색 원은 현재 영상과 참조 영상의 정보가 같아서 SKIP 모드로 결정된 부분이고, 파란색 원은 시간적 상관도가 높은 인트라 모드로 결정된 부분을 나타낸다. 영상에서 빨간색 원은 인트라 모드로 결정된 부분을 표현한 것인데, 인트라 모드의 분포가 극히 일부분이라는 것을 확인할 수 있다.

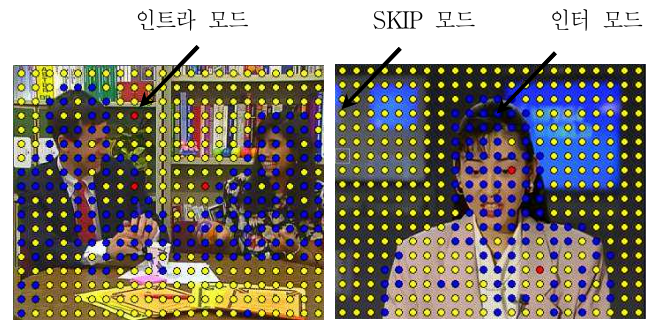


그림 4 인트라 화면에서의 모드 분포

H.264/AVC는 성능과 복잡도라는 상충되는 요건 중 성능에 중점을 두었다. 고속 부호화를 위해 인트라 화면에서 강제적으로 인트라 모드를 생략하면, 평균 0.18dB BDPSNR(Bjonteggard Delta PSNR) 감소와 3.5% BDBR(Bjonteggard Delta Bitrate) 증가가 발생하여 이에 따른 부호화 성능의 감소 정도는 매우 크다고 할 수 있다. BDPSNR, BDBR이란 JVT에서 채택한 객관적 부호화 효율 측정 방법이다 [2]. 즉, 인트라 화면에서 강제적으로 인트라 모드를 생략하는 것은 공간적 상관도가 큰 영상 일수록 부호화 효율을 떨어뜨리게 된다. 부호화 효율을 보장하면서 복잡도를 감소하기 위한 제안 방법은 아래와 같다.

인트라 화면이 부호화될 때, 인트라 4×4, 인트라 16×16 모드에 대한 평균 비트율 왜곡값을 저장한다. 실험적으로 인트라 16×16의 비트율 왜곡값은 인트라 4×4의 것에 비해 아주 작다는 것을 확인할 수 있었다. 이유는 크게 두 가지로 나눌 수 있다. 첫째, 인트라 16×16은 비교적 복잡하지 않은 영역에서 선택되는 모드가기 때문이다. 둘째, 인트라 16×16의 모드 결정 방법은 $\lambda \cdot Rate$ 를 고려하지 않는 식 (3)으로 계산하지만, 인트라 4×4는 식 (2)를 이용하여 계산하기 때문이다.

$$J_{mode}(s, r, M | \lambda_{mode}) = SATD(s, r, M) \quad (3)$$

where, $SATD = (\sum_{i,j} |DiffT(i, j)|) / 2$,

$$DiffT = H \times (Original(i, j) - Prediction(i, j)) \times H$$

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad \text{※ H: Hadamard 변환}$$

이를 이용하여 각 인트라 모드의 비트율 왜곡값을 구분하기 위해 최대 임계값과 최소 임계값, 총 2개의 임계값을 정한다. 고속 인트라 모드 결정에 관한 기존의 논문에서는 임계값을 실험적으로 정하였으나, 본 논

문에서는 인트라 화면에서 영상의 복잡도를 고려하여 모드별 평균 비트율 왜곡값으로 임계값을 정한다. 결정된 최대 임계값과 최소 임계값을 16×16, 16×8, 8×16 모드의 비트율 왜곡값과 비교하여 인트라 4×4와 인트라 16×16의 활성화 여부를 결정한다.

$RDCost_{mode} = \{RDCost_{16 \times 16}, RDCost_{16 \times 8}, RDCost_{8 \times 16}\}$
 조건 1) $RDCost_{mode} < \text{최소 임계값}$
 : 인트라 16×16, 인트라 4×4 모두 비활성화
 조건 2) $\text{최소 임계값} < RDCost_{mode} < \text{최대 임계값}$
 : 인트라 4×4모드 비활성화
 조건 3) $RDCost_{mode} > \text{최대 임계값}$
 : 인트라 16×16, 인트라 4×4 모두 활성화

그림 5는 제안한 알고리즘의 순서도를 보여준다.

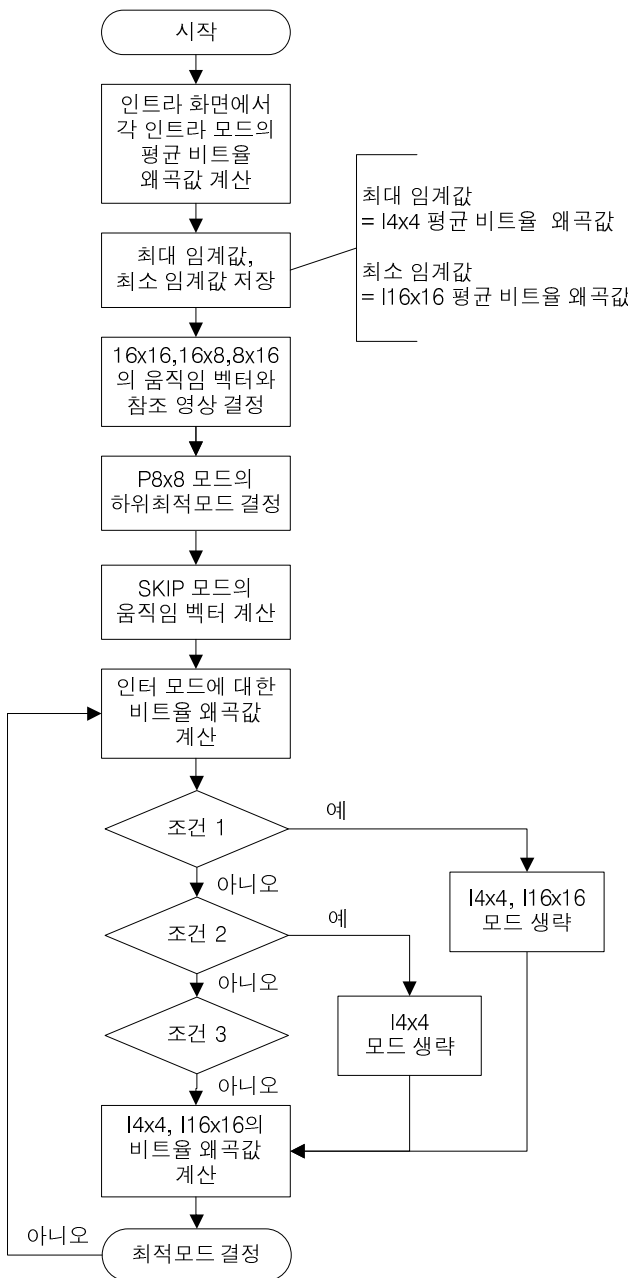


그림 5 제안 알고리즘 순서도

4. 실험 방법 및 결과

가. 실험 방법

본 논문에서는 H.264/AVC 참조 소프트웨어 JM 12.4에 제안한 알고리즘을 구현했다. 실험에 사용한 실험 영상 및 세부적인 환경 설정은 표 1과 같다.

표 2 실험 환경

영상 / 크기	Stefan_cif	352×288
	Mobile_cif	
	Paris_cif	
	Foreman_cif	
	BasketballDrill	832×480
	Night	1280×720
	City	1280×720
	Parkjoy	1920×1080
영상 수	100	
양자화 변수	16, 22, 27, 32, 37	
프로파일	메인 프로파일	
영상구조	IPPP 구조	
탐색영역	± 16	
엔트로피 코딩	CABAC	
참조 영상	1	

제안한 알고리즘의 평가를 위해 부호화 시간 변화율을 식 (4)에 정의했다.

$$\Delta T = \frac{\text{Proposed Time} - \text{Original Time}}{\text{Original Time}} \times 100(\%) \quad (4)$$

여기서, *Original Time*은 기존 H.264/AVC의 부호화 수행 시간이며, *Proposed Time*은 제안한 알고리즘의 부호화 시간을 의미한다. 제안한 고속 모드 결정 방법과 H.264/AVC의 모드 결정 방법의 수행시간을 비교하기 위해 ΔT 를 식 (4)와 같이 정의한다. 그 단위는 백분율로 나타냈다.

나. 실험 결과

다양한 크기의 영상에 대해 각각 실험하였으며, 표 2는 제안한 알고리즘의 양자화 변수에 따른 비트율 변화, PSNR의 변화, 시간의 변화를 평균적으로 나타냈다. 표 3은 영상 별 실험 결과를 나타낸다. 제안한 알고리즘을 적용했을 때, 성능의 큰 변화 없이 최대 36.28%, 최소 13.84% 부호화 수행 시간을 줄일 수 있었다.

표 2 양자화 변수에 따른 실험 결과

양자화 변수	Bit rate(%)	PSNR(dB)	Time(%)
16	0.37324	-0.04125	-30.15729
22	0.38073	-0.02375	-29.41495
27	0.57816	-0.02000	-28.78068
32	0.59638	-0.05250	-25.58504
37	0.66774	-0.08000	-23.15350
평균	0.51925	-0.04350	-27.41829

표 3 영상 별 실험 결과

	양자화 변수	Δ Bit rate	Δ PSNR	Δ T
Stefan	16	0.30891	-0.04	-36.27957
	22	0.42673	-0.02	-35.32141
	27	0.24619	-0.04	-31.69892
	32	0.24283	-0.06	-29.32976
	37	0.06386	-0.04	-28.53197
	평균	0.25771	-0.04000	-32.23233
Mobile	16	0.02732	-0.01	-33.46193
	22	0.07506	-0.01	-30.76885
	27	0.19906	0	-28.59878
	32	0.05967	-0.01	-24.35197
	37	-0.385658	-0.02	-13.84307
	평균	0.00491	-0.01000	-26.20492
Paris	16	0.18451	-0.01	-36.192
	22	0.05943	-0.01	-28.80146
	27	0.24142	-0.01	-28.91783
	32	0.14089	-0.03	-25.26675
	37	0.11578	-0.04	-25.98158
	평균	0.14840	-0.02000	-29.03192
Fore man	16	0.25281	0	-25.79386
	22	0.28564	-0.02	-27.93652
	27	0.4616	0	-27.17591
	32	0.66961	-0.02	-23.36491
	37	0.36573	-0.05	-21.79274
	평균	0.40708	-0.01800	-25.21279
Basket ball Drill	16	0.61646	-0.01	-33.50789
	22	1.16369	-0.01	-30.34718
	27	1.40367	0.01	-26.71916
	32	1.40687	-0.06	-23.53684
	37	1.76202	-0.14	-21.69978
	평균	1.27054	-0.042	-27.16217
Night	16	0.53465	-0.18	-20.86493
	22	0.2474	-0.05	-24.16355
	27	0.71722	-0.04	-26.87035
	32	0.84876	-0.07	-25.41505
	37	0.79336	-0.07	-23.37171
	평균	0.62828	-0.08200	-24.13712
City	16	-0.041813	-0.06	-26.05319
	22	0.13116	-0.01	-24.8619
	27	0.48456	-0.01	-28.72551
	32	0.50068	-0.05	-24.52724
	37	0.72486	-0.08	-23.60516
	평균	0.35989	-0.0420	-25.55460
Park joy	16	1.10304	-0.02	-29.10491
	22	0.65675	-0.06	-33.11871
	27	0.87153	-0.07	-31.53902
	32	0.90175	-0.12	-28.88775
	37	1.90194	-0.2	-26.40194
	평균	1.08700	-0.09400	-29.81047

그림 6은 표 2를 도식화한 것이다. 양자화 변수가 낮을수록 매크로블록이 인트라 모드로 결정되는 빈도가 낮으므로, 낮은 양자화 변수에서 제안한 알고리즘의 성능이 좋았다. 또한 복잡도가 높은 영상에서 인트라 모드로 예측되는 블록이 증가하므로, Foreman과 같은 움직임이 적은 단조로운 영상보다는 Stefan이나 Parkjoy와 같은 복잡한 영상에서 부호화 시간을 단축할 수 있었다. 전체 실험 영상에 대해 비트율은 약 0.52% 증가, PSNR은 0.04dB 감소, 시간은 27.42% 감소되었다.

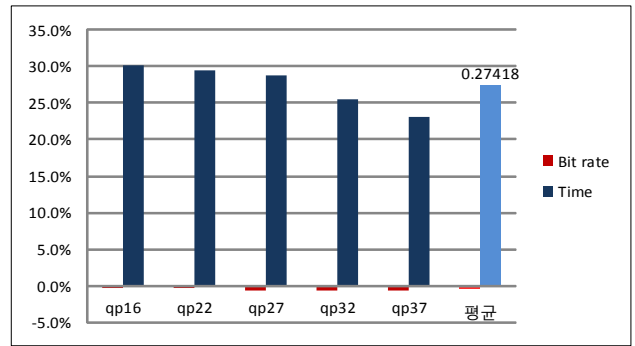


그림 6 양자화 변수에 따른 비트율 및 부호화 시간 변화를 비교

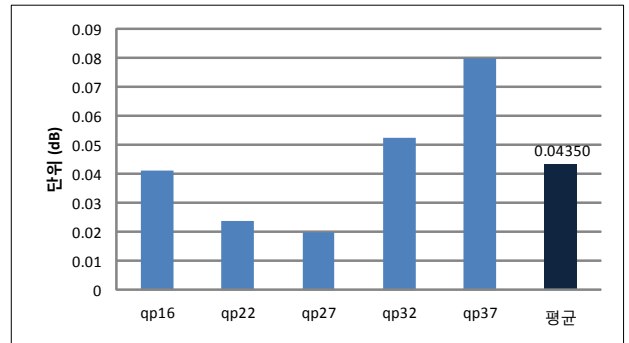


그림 7 양자화 변수에 따른 PSNR 변화

5. 결론

본 논문에서는 H.264/AVC 고속 부호화를 위한 모드 결정 방법을 제안했다. 인트라 화면의 모드 분포로부터 영상의 복잡도를 계산한 후, 이를 고려하여 최소 임계값과 최대 임계값을 결정한다. 결정된 두 임계값을 인트라 화면에서 16×16, 16×8, 8×16 모드의 비트율 왜곡값과 비교하여, 인트라 모드의 활성화 여부를 선택적으로 결정했다. 제안한 방법은 비트율과 PSNR 측면에서 큰 성능의 변화 없이 약 27.42% 부호화 시간을 단축했다.

ACKNOWLEDGMENT

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

참고문헌

- [1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no.7, July 2003.
- [2] 이제운, 전병우, "H.264 동영상 압축을 위한 고속 부호화 모드 결정 방법," 2004년 11월 전자공학회 논문지 제41권 제6호, pp. 209-219, 2004년 11월.
- [3] B. Kim, J. Kim, and C. Cho, "A Fast Intra Skip Detection Algorithm for H.264/AVC Video Encoding," *ETRI Journal*, vol. 28, no.6, Dec. 2006.
- [4] http://iphone.hhi.de/suehring/tml/download/old_jm/jm12.4.zip, Joint Video Team, Reference Software Version 12.4.