

H.264/AVC의 비트율 왜곡값과 시간 상관도를 이용한 고속 모드 결정 방법

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

Fast Mode Decision using Rate-Distortion Cost and Temporal Correlations in H.264/AVC

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

요약

가변적인 블록을 이용하는 H.264/AVC 인터 화면에서의 모드 결정 방법은 16×16부터 P8×8 모드까지 인터 모드에 대해 비트율 왜곡값을 계산 한 뒤 공간적 상관도를 고려하기 위해 인트라 4×4와 인트라 16×16 모드까지 비트율 왜곡값을 계산하고, 계산된 비트율 왜곡값 중 가장 작은 값을 갖는 모드를 최종 모드로 선택한다. 하지만 이러한 과정으로 부호기의 복잡도는 증가한다. 복잡도를 감소시키기 위해 본 논문에서는 IPPP구조에서의 인터 모드와 인트라 모드의 특징을 이용하여 고속 부호화를 제안한다. 제안하는 알고리즘은 인터 모드에 대해서 최적 모드를 결정하기 위한 비트율 왜곡값의 상관도와 시간 상관도를 이용하고, 인트라 모드에 대해서는 인트라 화면 내에서의 각 인트라 모드에 대한 최대·최소 임계값을 이용하여, 발생확률이 낮은 인터와 인트라 모드의 활성화 여부를 결정한다. 본 알고리즘의 적용 범위는 상대적으로 인터 화면에서 발생 빈도가 낮은 16×8, 8×16, P8×8와, 인트라 4×4, 인트라 16×16 모드이다. 제안하는 알고리즘은 기존의 H.264/AVC에 비해 1.36% 비트가 증가했고, PSNR은 0.06dB 감소했다. 즉, 부호화 효율의 큰 감소 없이 평균 42.58%의 부호화 시간을 감소시켰다.

1. 서론

H.264/AVC는 부호화 효율을 높이기 위해, 모드 결정 시 비트율 왜곡 최적화 기법(Rate-distortion optimization method, RDO)을 이용한다.[1] 비트율 왜곡 최적화 기법이란, H.264/AVC에서 한 매크로 블록의 모드를 결정하기 위해서 사용하는 기법으로, SKIP, 16×16, 16×8, 8×16, P8×8 (8×8, 8×4, 4×8, 4×4), 인트라 4×4, 인트라 16×16, 총 7가지의 모드의 비트율 왜곡값(J_{mode})을 계산하고, 이 중 가장 작은 비트율 왜곡값을 갖는 모드를 최적 모드로 결정하는 것을 말한다. 이러한 방법으로 부호화 효율을 기대할 수 있으나, 부호화 효율의 증가에 비해 복잡도가 많이 증가한다는 단점이 있다. 이는 실시간 부호화를 필요로 하는 애플리케이션에 적합하지 않기 때문에, 부호화 복잡도 감소를 위한 많은 알고리즘이 제안되고 있다.

제안된 대부분의 고속 부호화 알고리즘은 공간적·시간적 상관도, 최적 모드 결정을 위한 비트율 왜곡값과 부최적 모드 결정을 위한 비트율 왜곡값 등을 이용하였다.[2], [3] 그러나 제안된 대부분 논문들은 사전 실험을 통해 영상의 경향을 파악하고, 이를 통해 가장 적합한 조건들을 선택하였다.[3] 혹은 복잡하지 않은 영상에 대해서만 알고리즘을 적용한 경우가 많다. 보다 일반적인 부호화를 위해서는 이러한 방법 보다는 일반적인 영상에 적용할 수 있는 알고리즘이 필요하다.

본 논문에서는 인터 화면에서 고속 부호화를 위해 인터 모드 중 발생 빈도가 낮은 경향을 보이는 16×8, 8×16, P8×8의 인터 모드와 인트라 모드에 대해 최적 모드를 결정하기 위한 비트율 왜곡값(J_{mode})의 상관도와 인터 화면의 시간 상관도를 이용하고, 인트라 모드에 대해서는 인트라 화면 내에서의 각 인트라 모드에 대한 최대·최소 임계값을 이용하여 고속 모드 결정 방법을 제안한다.

2. 비트율 왜곡 최적화 기법을 이용한 모드 결정 방법

모드 결정을 위한 비트율 왜곡값 계산, 즉 비트율 왜곡 최적화 기법은 7가지 모든 모드에 대해 적용한다. H.264/AVC의 비트율 왜곡값은 아래의 식 (1)과 식 (2)의 2가지로 나눌 수 있다.

$$J_{motion}(MV, REF | \lambda_{motion}) = SAD(s, r(MV, REF)) + \lambda_{motion} \cdot R(MV, REF) \quad (1)$$

$$where \quad SAD(s, r(MV, REF)) = \sum_{x \in H, y \in V} |s(x, y) - r(x - m_x, y - m_y)|.$$

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

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

첫 번째는 J_{motion} 으로 움직임 벡터와 참조 영상을 결정하기 위한 보조 최적 모드 결정 비트율 왜곡값으로써 식 (1)과 같이 나타낼 수 있다. 식 (1)에서 MV 는 움직임 벡터, REF 는 참조 영상을 의미하고 $R(s, r(MV, REF))$ 은 움직임 벡터와 참조 영상을 부호화하는데 필요한 비트수로, 미리 정의한 표를 참조하여 얻어지는 값이다. SAD (Sum of Absolute Difference)는 원 영상(s)과 움직임 예측에 의해 움직임 보상된 영상(r)의 차이에 대한 절대값의 합을 나타낸다.

두 번째 비트율 왜곡값은 최적 모드를 결정하기 위한 J_{mode} 으로써, 식 (2)로 나타낼 수 있다. 여기서 J_{mode} 는 비트율 왜곡값을, M 은 모드,

$R(s, r, M)$ 은 부호화할 때 발생하는 실제 비트량을 의미한다. 여기서 $SSD(Sum\ of\ Squared\ Difference)$ 란 원본 블록(s)과 복원된 블록(r)의 차이값의 제곱의 합이다. 비트율 왜곡 최적화 기법을 통해 가장 작은 비트가 할당 되는 모드를 최적의 모드로 결정한다. 이는 움직임 벡터나 최적의 참조 영상을 결정하기 위한 J_{motion} 보다 이후에 계산되며, 보조 최적 모드 결정을 위한 비트율 왜곡값에 비해 복잡도가 높고 정확하다.

H.264/AVC는 영상의 시간적인 변화보다 공간적인 변화가 적을 경우 부호화 효율을 높이기 위해 인터 화면에서도 인트라 모드를 이용하여 부호화한다. 그림 1에서 볼 수 있듯이, 일반적으로 인터 화면에서 인트라 모드의 발생 확률은 SKIP이나 16x16 모드에 비해서 매우 낮다.

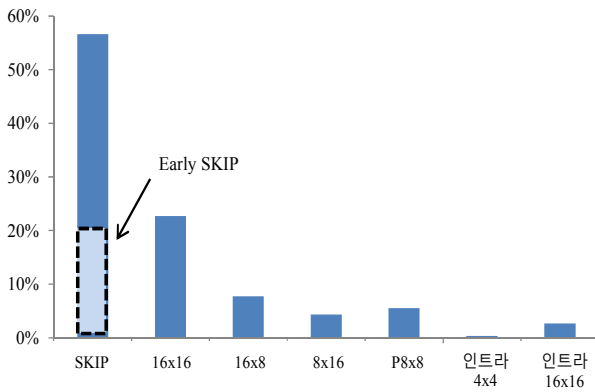


그림 1 모드 분포도

압축하고자 하는 영상의 복잡도나 양자화 계수에 따라 차이가 있지만 대부분의 영역에서 인터 모드 내에서도 16x8, 8x16은 빈도수가 낮다. P8x8은 복잡한 영역에서 빈도수가 높고, 그에 반해 배경과 같이 평활한 영역은 16x16과 같은 큰 블록으로 움직임을 보상되어, 상대적으로 P8x8의 빈도수는 낮다. 이러한 모드 결정의 특징을 이용하여 영상에 적절하게 매크로블록의 모드를 설정함으로써 압축 효율을 높일 수 있다. 다시 말해, 다양한 움직임 보상 블록 중에서 적합한 모드의 비트율 왜곡값을 계산하는 것이 복잡도 감소를 위해 중요하다.

3. 비트율 왜곡값의 상관도와 시간 상관도를 이용한 모드 결정 방법

가. 인터 모드의 선택적 모드 결정 방법

H.264/AVC 기준 소프트웨어(JM)에서 최적 모드를 결정하는 방법(RD Optimization)을 세 가지로 제공하고 있다. 이는 낮은 복잡도 모드(Low complexity mode)와 높은 복잡도 모드(High complexity mode), 고속의 높은 복잡도 모드(Fast high complexity mode)로 구성되어 있다. 낮은 복잡도 모드는 높은 복잡도 모드에 비해 적은 부호화 시간을 필요로 하지만, 실제 발생하는 비트수를 고려하지 않기 때문에 높은 복잡도 모드로 부호화 했을 때 보다 발생 비트수 측면에서 낮은 성능을 보인다.

고속의 높은 복잡도 모드는 높은 복잡도 모드에 초기 SKIP 모드 결정 과정을 추가하였다. 초기에 16x16 모드에 대한 움직임 예측을 수행하여 움직임 벡터와 참조 영상을 결정한 후, 비트율 왜곡 값을 계산한다. 이 계산 결과가 아래의 조건 2 ~ 4를 만족할 때에 SKIP 모드로

결정하고 이하의 불필요한 계산을 생략한다. 그림 2에서 Early SKIP으로 표시한 부분이 이에 해당한다. 다음과 같은 네 가지 조건을 모두 만족할 때 SKIP 모드로 결정되고 이를 Early SKIP이라 한다.

- 조건 1) 움직임 보상 블록: 16x16
- 조건 2) 참조 영상: 이전 영상
- 조건 3) 16x16 모드의 움직임 벡터 = SKIP 모드의 움직임 벡터
- 조건 4) 잔여 데이터의 비트수가 0 (CBP = 0)

기준 소프트웨어의 고속의 높은 복잡도 모드에서 Early SKIP을 통해 불필요한 계산 시간을 줄인 것처럼, 본 논문에서는 기준 소프트웨어에서 최적의 모드 결정을 위해 마지막 단계에서 비트율 왜곡값을 계산하는 것과 달리 각 모드에 대한 비트율 왜곡값을 각각 계산함으로써, 다른 모드에 대한 J_{mode} 의 상관도와 시간의 상관도를 이용하여 고속 부호화를 제안한다.

그림 2는 7가지 모든 모드에 대한 비트율 왜곡값의 분포도이다. 이를 통해 각 모드별로 비트율 왜곡값의 상관도가 높다는 것을 확인할 수 있다.

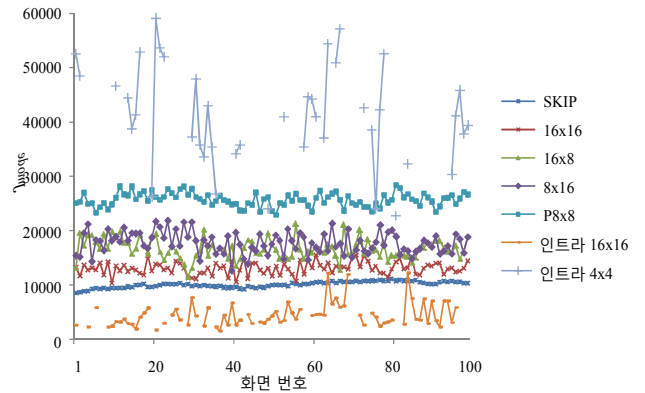


그림 2 비트율 왜곡값 분포도

인터 모드에 대해 제안하는 선택적 비트율 계산 방법의 적용범위는 그림 1에서 언급한 발생 빈도가 낮은 16x8, 8x16, P8x8 세 가지 모드이다. 그 중, 16x8, 8x16 모드에 대한 제안 조건은 다음과 같다.

- 조건 1) 현재 매크로블록에 대한 $\{J_{mode(SKIP)}, J_{mode(16x16)}\}$ 중 최소값 < 이전 화면의 16x8, 8x16 모드의 평균 J_{mode} 값
- 조건 2) 이전 영상의 동일 위치의 매크로블록 모드가 SKIP이나 16x16 모드일 때

대체로 SKIP과 16x16 모드의 발생 빈도가 높다. 그래서 현재 매크로블록의 SKIP과 16x16 모드의 J_{mode} 값 중, 최소값을 이용하여 이전 화면에서 16x8이나 8x16 모드로 결정된 매크로블록의 평균 J_{mode} 값을 비교한다. 즉, 조건 1을 만족하면 현재 매크로블록에서 16x8이나 8x16 모드의 발생 확률이 낮다.

또한 영상을 부호화 할 때 1초에 30장이나 60장 정도의 화면을 부호화하기 때문에, 인터 화면에서는 시간적 상관도가 높다. 이러한 특징을 이용하여 조건 2를 제시하였다. 이전 화면의 동일 위치의 매크로블록 모드가 SKIP이나 16x16 모드일 때, 16x8이나 8x16 모드의 발생 확

률이 낮다. 위의 두 가지 조건을 모두 만족할 때, 최적 비트율 왜곡값 결정을 위한 16×8과 8×16 모드의 J_{mode} 값은 계산하지 않는다.

복잡한 영역에서 주로 결정되는 P8×8 모드의 복잡도는 P8×8 모드를 제외한 다른 인트라 모드의 복잡도와 같다. P8×8 모드에 대한 비트율 왜곡의 선택적 계산을 위한 제안 조건은 다음과 같다.

- | |
|--|
| 조건 1) 현재 매크로블록에 대한 ($J_{mode(SKIP)}$, $J_{mode(16\times16)}$, $J_{mode(16\times8)}$, $J_{mode(8\times16)}$) 중 최소값 < 이전 화면의 P8×8 모드의 평균 J_{mode} 값
조건 2) 이전 영상의 동일 위치의 매크로블록 모드가 SKIP이나 16×16 모드일 때 |
|--|

SKIP, 16×16, 16×8과 8×16 모드의 비트율 왜곡값은 이미 계산되어 저장되어 있으므로 이 값들의 최소값과의 비교를 통해 선택적 비트율 왜곡값 계산을 수행할 수 있다. 네 가지 모드의 비트율 왜곡값 중 최소값이 이전 화면의 P8×8 모드의 평균 J_{mode} 값이 크다면, 현재 매크로블록이 P8×8로 결정될 확률이 낮다. 그러므로 위의 조건을 만족할 경우, P8×8 모드의 비트율 왜곡값 계산을 생략함으로써 인트라 모드의 최적 모드 결정을 위한 복잡도를 약 25% 줄일 수 있다.

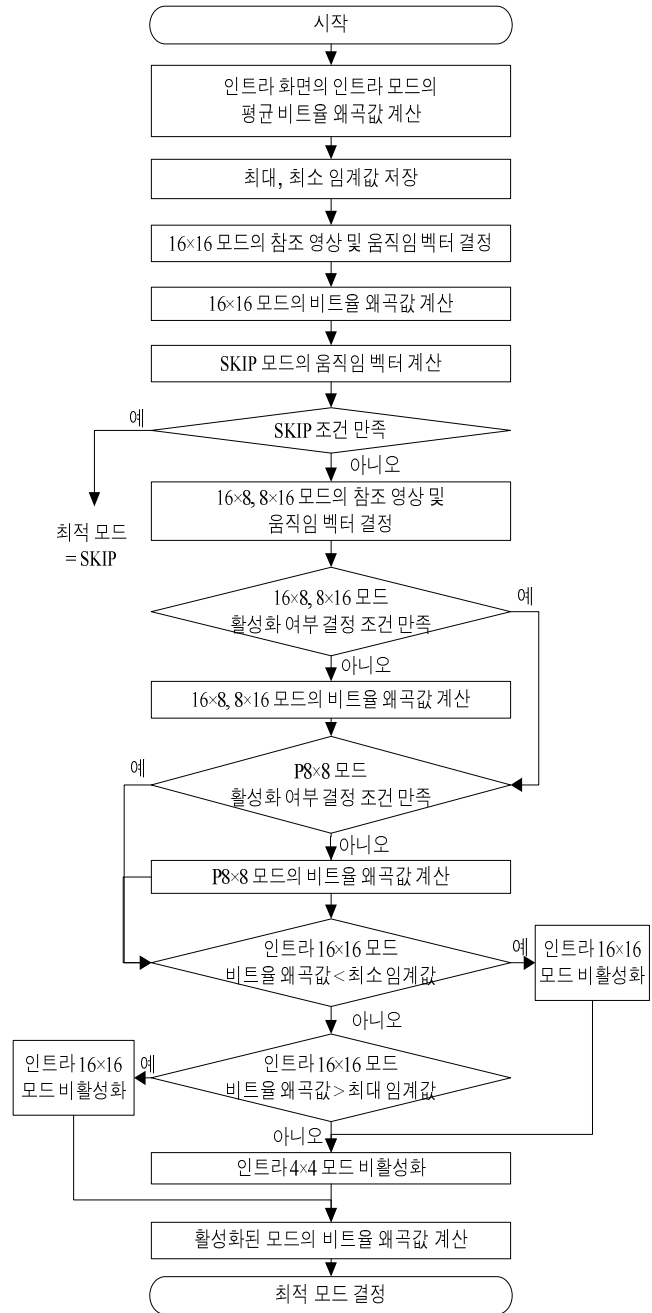
나. 인트라 모드의 선택적 모드 결정 방법

H.264/AVC는 성능과 복잡도라는 상충되는 요건 중 성능에 중점을 두었고 시간적 상관도를 이용한 인트라 화면을 부호화할 때에도 공간적 상관도까지 고려하여 위해 인트라 모드를 도입하였다. 인트라 화면은 시간적 상관도가 높은 화면이기 때문에 인트라 모드의 발생 빈도가 현저히 낮다. 하지만 인트라 화면에서 강제로 인트라 모드를 생략하는 것은 공간적 상관도가 큰 영상 일수록 부호화 효율을 떨어뜨리게 되므로, 부호화 효율을 보장하면서 복잡도를 감소하기 위한 인트라 모드에 대한 고속 모드 결정 방법을 제안한다.

GOP(Group of pictures)구조는 한 장의 인트라 화면을 부호화 한 뒤 인트라 화면을 부호화 하는 IPPP 구조를 이용한다. 인트라 화면이 부호화될 때, 인트라 4×4, 인트라 16×16 모드에 대한 평균 비트율 왜곡값을 저장한다. 평균 비트율 왜곡값으로 각 영상의 대략적인 인트라 모드의 비트율 왜곡값의 분포를 알 수 있다. 이를 이용하여 인트라 화면에서의 최대·최소 임계값을 계산하고, 이 임계값을 이용하여 각 인트라 모드의 적용여부를 결정한다. 그림 2에서 확인할 수 있듯, 인트라 16×16 모드의 비트율 왜곡값은 인트라 4×4 모드의 비트율 왜곡값에 비해 작은 값을 가진다. 즉, 본 논문에서 제안하는 알고리즘은 인트라 화면에서 영상의 복잡도를 고려하여 모드별 평균 비트율 왜곡값으로 임계값을 정하기 때문에 특정한 영상에 제한되는 것이 아닌, 일반적인 영상 각각의 특성을 반영한다는 것을 알 수 있다.

- | |
|--|
| $J_{mode(i)} = \{J_{mode(16\times16)}, J_{mode(16\times8)}, J_{mode(8\times16)}\}$
조건 1) $J_{mode(i)} < \text{최소 임계값}$
: 인트라 16×16, 인트라 4×4 모두 비활성화
조건 2) $\text{최소 임계값} < J_{mode(i)} < \text{최대 임계값}$
: 인트라 4×4모드 비활성화
조건 3) $J_{mode(i)} > \text{최대 임계값}$
: 인트라 16×16, 인트라 4×4 모두 활성화 |
|--|

그림 3은 제안한 알고리즘의 순서도이다.



4. 실험 방법 및 결과

가. 실험 방법

본 논문에서는 H.264/AVC 참조 소프트웨어 JM 12.4에 제안한 알고리즘을 구현했다.[4]

$$\Delta Time (\%) = \frac{Time_{proposed} - Time_{original}}{Time_{original}} \times 100 \quad (3)$$

$$\Delta Bitrates (\%) = \frac{Bitrates_{proposed} - Bitrates_{original}}{Bitrates_{original}} \times 100 \quad (4)$$

$$\Delta PSNR (dB) = PSNR_{proposed} - PSNR_{original} \quad (5)$$

표 2 실험 결과 (양자화 계수: 22, 27, 32, 37)

실험 영상	인터 모드			인트라 모드			인터 모드 + 인트라 모드		
	Δ Time (%)	Δ Bitrate (%)	Δ PSNR (dB)	Δ Time (%)	Δ Bitrate (%)	Δ PSNR (dB)	Δ Time (%)	Δ Bitrate (%)	Δ PSNR (dB)
Flowergarden	-16.25	0.61	-0.03	-15.70	0.03	-0.02	-31.79	1.12	-0.03
Football	-13.36	1.17	-0.05	-23.54	0.28	-0.01	-38.13	2.16	-0.06
Tempete	-25.76	1.35	-0.04	-17.18	0.41	-0.01	-42.00	1.53	-0.05
Paris	-31.61	2.22	-0.06	-17.73	0.21	-0.02	-49.20	2.28	-0.08
Akiyo	-27.36	0.43	-0.05	-14.92	0.03	-0.01	-42.02	0.49	-0.06
Hall	-34.47	0.68	-0.05	-16.52	-0.06	-0.02	-52.36	0.60	-0.08
평균	-24.80	1.08	-0.05	-17.60	0.15	-0.02	-42.58	1.36	-0.06

표 1 실험 환경

프로파일	Baseline
최적 모드 결정 방법	Fast high complexity mode
영상 크기	CIF (352×288)
영상 수	100
양자화 변수	22, 27, 32, 37
GOP 구조	IPPP 구조
탐색영역	± 16
엔트로피 코딩	CAVLC
참조 영상	5

제안한 알고리즘의 평가를 위해 식 (3)의 부호화 시간 변화율 (Δ 시간), 식 (4)의 비트의 변화율 (Δ 비트율), 식 (5)의 PSNR의 변화(Δ PSNR)를 통해 성능을 나타냈고, 시간 변화율과 비트율의 변화율은 백분율로 나타내었고, 화질의 변화인 PSNR의 단위는 dB로 나타냈다.[5] 실험에 사용한 실험 영상 및 세부적인 환경 설정은 표 1과 같다.

나. 실험 결과

영상의 복잡도에 따른 알고리즘의 성능을 측정하기 위해 다양한 복잡도의 영상으로 실험했다. 표 2는 제안한 알고리즘의 결과로 영상에 따른 비트율 변화, PSNR의 변화, 시간의 변화를 평균적으로 나타냈다. 제안한 알고리즘을 적용 했을 때, 성능의 큰 변화 없이 최대 52.36%, 최소 31.79% 부호화 수행 시간을 줄일 수 있었다. 그림 4는 CIF (352×288) 크기 영상인 Hall의 비트율 왜곡 곡선으로 성능 측면으로는 변화가 없음을 알 수 있다.

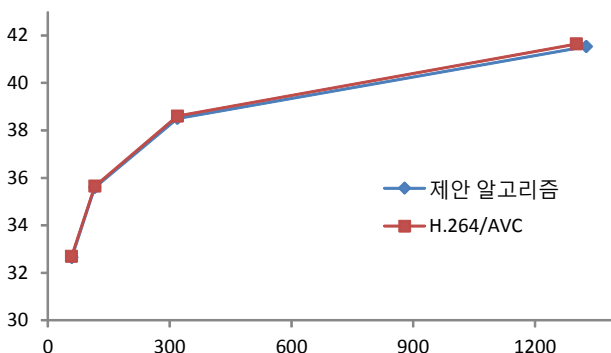


그림 4 비트율 왜곡 곡선

5. 결론

본 논문에서는 H.264/AVC 인터 화면 내의 고속 부호화를 위한 모드 결정 방법을 제안했다. 인터 모드의 경우, 비트율 왜곡값과 시간적 상관성을 이용한다. 적용범위는 모드 발생빈도가 상대적으로 낮은 16×8, 8×16, P8×8 모드이다. 미리 계산된 SKIP과 16×16 모드의 비트율 왜곡값과 이전 화면에서의 각 모드의 평균 비트율 왜곡값을 비교하여 각각의 모드의 활성화 여부를 결정한다. 인트라 모드의 경우, 인트라 화면으로부터, 모드 분포로부터 영상의 복잡도를 계산하고 이를 고려하여 최소 임계값과 최대 임계값을 결정한다. 결정된 최대 임계값과 최소 임계값을 16×16, 16×8, 8×16 모드의 비트율 왜곡값과 비교하여 인트라 4×4와 인트라 16×16의 활성화 여부를 선택적으로 결정했다. 제안한 방법은 비트율과 PSNR 측면에서 큰 성능의 변화 없이 약 42.58% 부호화 시간을 단축했다.

ACKNOWLEDGMENT

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2011-(C1090-1111-0003))

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, pp. 560-576, July 2003.
- [2] S. Ri, Y. Vatis, and J. Ostermann, "Fast Inter-Mode Decision in an H.264/AVC Encoder Using Mode and Lagrangian Cost Correlation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 19, no. 2, pp. 302-306, Feb. 2009.
- [3] L. Pan and Y. Ho, "Fast Mode Decision Algorithm for H.264 Inter Prediction," Electronics Letters, vol. 43, no. 24, pp. 1351-1353, Nov. 2007.
- [4] Reference Software JM 12.4. Available from: http://iphone/hhi.de/shehring/tml/download/old_jm/jm12.4.zip.
- [5] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves (VCEG-M33)," VCEG Meeting (ITU-T SG16 Q.6), April 2001.