

고화질 비디오 부호화를 위한 H.264/AVC 라인 기반 인트라 16×16 예측 방법

*최정아, **김낙우, **이병탁, *호요성

*광주과학기술원 정보기전공학부

**한국전자통신연구원

{jachoi, nwkim, bytelee, hoyo}@gist.ac.kr

Line Based Intra 16×16 Prediction in H.264/AVC for High Resolution Video Coding

*Jung-Ah Choi, **Nac-Woo Kim, **Byung-Tak Lee, and *Yo-Sung Ho

*Gwangju Institute of Science and Technology (GIST)

**Electronics and Telecommunications Research Institute (ETRI)

요약

기존의 H.264/AVC 비디오 표준은 고화질 비디오 부호화를 지원하지만 고해상도에 특화된 요소 기술이 도입되지 않아 만족할만한 성능을 보이지 못한다. 현존하는 동영상 압축 표준 중 가장 뛰어난 H.264/AVC 표준의 인트라 16×16 예측은 매크로블록에 인접한 최대 33개의 주변 화소를 이용하여 매크로블록에 속한 256개의 화소 값을 예측한다. 특히, 전체 예측 모드 중 수직과 수평 예측 모드에서는 16개의 수직 또는 수평 위치에 위치한 주변 화소로 전체 매크로블록 내의 화소 값을 예측하므로 매크로블록의 끝으로 갈수록 예측의 정확도가 떨어져 부호화 비트가 증가한다. 고화질 영상에서는 인트라 16×16 모드로 부호화되는 블록이 많으므로 수행되므로 인트라 16×16 예측의 정확도를 높일 수 있는 기술이 필요하다. 본 논문에서는 기존의 H.264/AVC의 예측 방법보다 예측 정확도가 높은 새로운 라인 기반 16×16 인트라 예측 방법을 제안한다. 일반적으로 편평한 특성을 보이는 인트라 16×16 블록이라도 좀 더 가까운 화소를 참조 화소로 사용하면 예측의 정확도를 높여 부호화 비트를 줄일 수 있다. 이를 이용하여 제안하는 알고리즘에서는 인트라 16×16 블록에서 16개 화소 한 줄을 단위로 예측 및 부호화를 수행한다. 1080p HD급 테스트 영상을 이용하여 실험한 결과, 기존의 H.264/AVC FExt High 프로파일에 비해 평균 약 6.92%의 부호화 비트를 감소시킬 수 있음을 보였다.

1. 서론

H.264/AVC는 ISO/IEC의 동화상 전문가 그룹(Moving Picture Expert Group, MPEG)과 ITU-T의 비디오 코딩 전문가 그룹(Video Coding Expert Group, VCEG)이 JVT(Joint Video Team)을 구성하여 공동 연구를 진행한 결과물로 완성한 동영상 압축 국제 표준이다 [1][2]. H.264/AVC는 기존의 MPEG-2 표준에 비해 2배 이상의 압축 효율을 가지는 현존하는 동영상 압축 표준 중 가장 뛰어난 코덱이다.

JVT는 HD급 영상을 위한 프로파일로 FExt(Fidelity Range Extension)라 불리는 High, High 10, High 4:2:2, High 4:4:4 프로파일을 추가하고 이를 위한 표준화 작업에 들어갔다 [3]. FExt로서 제정된 4개의 High 프로파일에서는 그동안의 취약점이었던 색상 공간의 보정 및 확장에 주안점을 두고 YUV 4:2:2 및 4:4:4 포맷 지원, 알파 혼합(alpha blending) 지원, 화소 당 10 비트 및 12 비트의 다양한 비트 깊이(bit depth) 지원, 무손실 코딩 및 스테레오 영상 코딩 지원, 정수 8×8 DCT 연산 추가, 무손실 코딩을 위한 새로운 인트라 예측 모드(I_PCM) 추가, 새로운 정수 색상 공간 지원(Y, Cg, Co), 선택적인 양자화 매트릭스 입력 등의 기술을 추가하였다.

이러한 추가적인 작업들을 하였음에도 불구하고, H.264/AVC는 아직까지 HD급 또는 그 이상의 해상도를 가지는 고화질 영상에서 만족할만한 성능을 보여주지 못하고 있다. 하지만, 최근 텔레비전 방송을

시청하기 위한 AV 기기들이 고도화되고 IPTV, DTV와 같은 상용 TV 서비스에서의 HDTV(High Definition Television) 콘텐츠들이 각광을 받게 됨에 따라, 고화질 비디오 부호화에도 적용할 수 있는 H.264/AVC 코덱의 필요성이 크게 대두되고 있다.

고화질 영상은 그 해상도의 크기로 인하여 매우 높은 비트레이트를 필요로 하고 있다. HD급 영상만 하더라도 SD급에 비해 720p의 경우 3.5배, 1080p의 경우 7배 가량 부호화 비트가 증가하기 때문에 이에 따른 어려움이 있다. 높은 압축률을 위해 양자화 계수를 높일 수 있지만 복호 후에도 고화질 그 자체의 특성을 그대로 전달하기 위해서는 너무 높은 양자화 계수는 피하는 것이 좋다. 따라서 고화질 영상의 화질을 떨어뜨리지 않으면서, 즉 낮은 양자화 계수로 부호화한 고비트율 환경에서 부호화 비트를 줄일 수 있는 효율적인 압축 방법이 필요하다.

H.264/AVC의 인트라 16×16 예측은 매크로블록의 화소들을 한번에 예측하는데, 현재 매크로블록에 인접한 최대 33개의 주변 화소를 이용하여 현재 매크로블록에 속한 256개의 화소 값을 예측한다. 특히, 전체 예측 모드 중 수직과 수평 예측 모드에서는 16개의 수직 또는 수평 위치에 위치한 주변 화소로 전체 매크로블록 내의 화소 값을 예측하므로 매크로블록의 끝으로 갈수록 예측의 정확도가 떨어져 부호화 비트가 증가한다.

본 논문에서는 기존의 H.264/AVC의 예측 방법보다 예측 정확도가 높은 새로운 라인 기반 16×16 인트라 예측 방법을 제안한다. 실험을 통

해, 고화질 영상에서 일반적으로 편평한 특성을 보이는 인트라 16×16 블록도 가까운 화소를 참조 화소로 사용하면 예측의 정확도를 높여 부호화 비트를 줄일 수 있음을 보인다. 또한 이러한 특성을 이용해 인트라 16×16 블록의 수직과 수평 예측에 대해, 16개 화소 한 줄을 단위로 예측을 수행하는 라인 기반 인트라 16×16 예측 방법을 제안한다.

2. H.264/AVC 인트라 16×16 예측

인트라 예측 모드는 H.264/AVC에서 처음으로 채택되어 사용되는 부분으로 인트라 화면의 부호화 성능 향상에 매우 중요한 역할을 한 부분이다. H.264/AVC는 영상의 공간적 중복성을 제거하기 위해서 인트라 4×4, 인트라 8×8, 인트라 16×16 예측을 사용한다.

인트라 16×16 예측에서는 그림 1과 같은 네 가지 모드를 제공한다. 수직 (모드 0), 수평 (모드 1), DC (모드 2), Plane (모드 3)으로 정의된 각 모드는 정해진 방향성을 가진다. 이 중 수직, 수평, DC 모드는 크기가 커졌을 뿐 인트라 4×4 모드에서의 방법과 같고, Plane 모드는 정수 연산만을 가지는 이중 선형 변환과 유사하다.

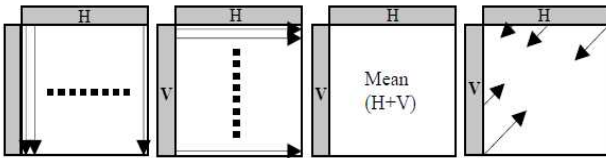


그림 1. 인트라 16×16 블록의 네 가지 예측 모드

인트라 16×16 예측에서 최적의 예측 모드는 다음 수식과 같은 SATD(Sum of Absolute Transformed Differences)를 사용하여 결정한다. 이 수식에서 $O(x,y)$ 는 원 영상의 블록, $P(x,y)$ 는 예측된 블록을 나타내며 $\mathcal{T}[\cdot]$ 는 직교변환, b_k 는 블록의 크기를 나타낸다. 즉, SATD는 원영상과 예측 영상 간의 차분치에 대해 직교 변환을 실행한 후 그 값에 절대 값을 취하여 모두 더한 값을 말한다.

$$SATD = \sum_{(x,y) \in b_k} |\mathcal{T}[O(x,y) - P(x,y)]| \quad (1)$$

SATD에서 직교 변환 $\mathcal{T}[\cdot]$ 을 수행할 때에는 부호화에 사용하는 DCT(Discrete Cosine Transform)가 아닌 다음 식과 같은 하다마드 변환을 사용한다. 이 식에서 D 는 차분 값이고, H 는 하다마드 변환이다.

$$SATD = HDH^T \quad (2)$$

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \quad (3)$$

이는 모든 4×4 블록에 대해 DCT를 수행하는 것보다 단순히 덧셈과 뺄셈으로만 이루어진 하다마드 변환을 사용하는 것이 연산량 및 복잡도 면에서 더 좋은 결과를 제공하기 때문이다.

모든 예측 모드에 대한 SATD 값을 모두 계산한 후, 그 중 값이 가장 작은 모드를 최적 모드로 선택한다. 예측 및 최적 모드 선택 후, 원 영상의 해당 블록에서 최적 모드를 사용하여 예측된 블록을 뺀 차분 값을 부호화한다.

3. 제안한 방법

가. 참조 화소와의 거리에 따른 예측 정확도 분석

먼저 본 연구의 동기가 된 참조 화소와 예측될 화소 사이의 거리에 따른 예측 정확도를 분석해본다. 인트라 16×16 예측은 한번에 현재 매크로블록 내의 모든 화소들을 예측하기 때문에 상대적으로 적은 참조 화소를 이용하여 화소 값을 예측한다. 따라서 수직 혹은 수평 예측의 경우 참조 화소에 인접한 화소들은 예측 값이 정확하지만, 참조 화소에서 멀리 떨어진 화소들은 예측 값이 부정확하다. 영상에는 공간적 상관도가 존재하므로 일반적으로 가까운 참조 화소를 이용하는 것이 더 나은 예측 결과를 제공하기 때문이다.

실험적으로 이러한 사실을 증명하기 위해 1080p HD급 실험 영상 중 하나인 BasketballDrive 영상 첫 프레임에 이용하여 양자화 계수가 24일 때, 다음과 같은 실험을 수행하였다. 참조 화소와의 거리에 민감한 인트라 16×16의 수직과 수평 모드에 대해, 기존의 인트라 16×16 예측 방법에서 사용하는 참조 화소의 위치를 0으로 두고 그보다 n 만큼 위의 화소를 거리가 n 만큼 떨어진 것으로 정의하였다. 참조 화소의 거리에 따른 비트량 변화를 확인하기 위해, 참조 화소의 위치를 0부터 8까지 옮기며 그에 따른 부호화 비트를 측정하였다.

표 1. 참조 화소와의 거리에 따른 비트율 비교

참조 화소와의 거리	PSNR (dB)	비트율 (kbits/s)	Δ비트 증가량 (kbits/s)
0	40.33 (동일)	27544.08	-
2		27614.4	+70.32
4		27576.24	+32.16
6		27588.96	+44.88
8		27640.56	+96.48

표 1은 그 실험 결과를 보여준다. Δ비트 증가량은 참조 화소와의 거리가 0일 때를 기준으로 비트의 증가량을 보여준다. 참조 화소의 거리와 예측의 정확도가 비례하지는 않지만, 참조 화소와의 거리가 기존의 거리보다 멀어지는 모든 경우에 대해서 예측 정확도가 떨어짐을 확인할 수 있었다. 모든 경우에서 동일 화질을 제공하기 때문에 더 확실한 비교를 진행할 수 있는데, 이는 동일 화질을 제공하기 위해 필요한 비트량이 거리가 멀어지면 더욱 늘어남을 뜻한다.

위의 실험 결과를 통해, 가능하면 가까운 참조 화소에서 대상 화소를 예측하는 것이 부호화 비트량 면에서 효율적임을 확인할 수 있다. 하지만, 현재의 인트라 16×16 예측 알고리즘에서는 한 매크로블록이 한번에 예측되기 때문에 이미 부호화된 참조 화소 중 가장 가까운 것이 매크로블록 경계에 인접한 33개의 화소이다. 이들 중 몇 개의 화소를 이용하여 전체 매크로블록을 예측하기 때문에 참조 화소와 거리가 먼 화소들에 대해서는 예측의 정확도가 감소한다.

나. 라인 기반 인트라 16×16 예측

앞 절에서 본 것과 같이 예측의 정확도는 참조 화소와의 거리에 영향을 많이 받는다. 현재 예측할 화소를 좀 더 가까운 참조 화소를 이용하여 예측하기 위해, 제안하는 알고리즘에서는 인트라 16×16 블록 예측을 한번에 수행하지 않고, 라인 단위로 나누어 예측한다.

그림 2는 제안하는 라인 기반 인트라 16×16 예측 방법의 흐름도를 보여준다. 매크로블록에 대해 먼저 기존의 H.264/AVC의 인트라 16×16

예측을 수행하여 네 개의 예측 모드에 대한 예측 블록을 만든다. 만들어진 네 개의 예측 블록에 대해 SATD 비용 값을 계산하여, 가장 작은 SATD 값을 가지는 예측 모드를 최적 모드로 결정한다.

결정된 최적 모드가 2보다 작은 경우, 즉 수직 (모드 0) 또는 수평 (모드 1)인 경우, 제안한 라인 기반 인트라 16×16 예측을 수행하고 그렇지 않은 경우, 즉 DC (모드 2) 또는 Plane (모드 3)인 경우에는 기존의 인트라 16×16 예측을 그대로 수행한다. 라인이라 함은 말 그대로 한 매크로블록을 수직 또는 수평 모드에 따라 1×16 또는 16×1을 한 라인으로 정의한 것이다. 즉, 한 매크로블록에 16개의 라인이 정의된다.

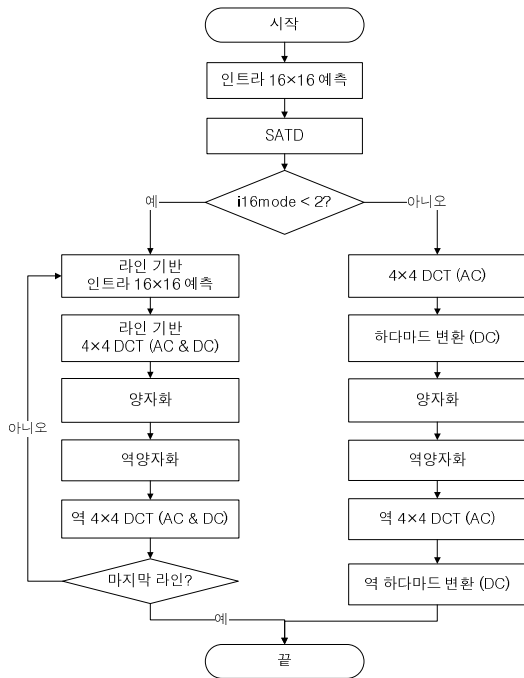


그림 2. 제안한 알고리즘의 흐름도

라인 기반 인트라 16×16 예측은 최적 모드의 방향성은 유지하면서 한 라인에 대해서만 예측을 수행한다. 그림 3은 최적 모드가 수직 모드인 경우 화소를 예측하는 방법을 보여준다. 기존의 인트라 예측에서는 한 매크로블록 전체에 대해서 예측을 수행하므로 한 매크로블록 전체에 대해서 참조 화소의 값을 복사하여 예측하였으나, 제안하는 예측 방법에서는 수직 모드일 경우 1×16을 한 라인으로 보고 한 라인에 대해서만 참조 화소의 값을 복사하여 예측한다.

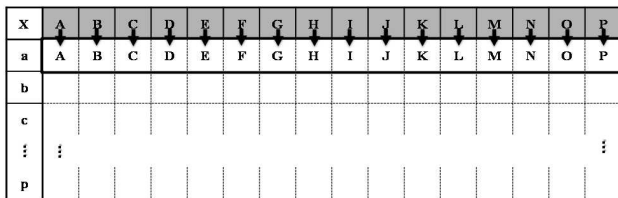


그림 3. 제안한 알고리즘의 흐름도

그 다음, 예측된 라인과 원 영상에서의 라인에 해당하는 화소 값의 차를 구해 잔여 영상을 구하고, DCT(Discrete Cosine Transform)를 수행한다. H.264/AVC 표준에서 제공하는 DCT는 4×4 정수 DCT이므로 호환성을 위해 잔여 영상의 한 라인을 그림 4와 같은 방법으로 재배열하여 4×4 블록으로 만든다.

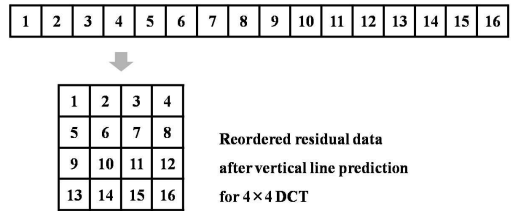


그림 4. 4×4 블록 만들기

이렇게 만들어진 4×4 블록은 4×4 DCT를 거쳐 양자화되어 비트스트림을 쓰는 단으로 넘어가게 되는데, 다음 예측을 위한 참조 화소 값을 생성하기 위해 복호도 함께 이루어진다. 현재 부호화한 라인을 복호한 후, 현재 라인이 매크로블록의 마지막 라인인지 확인한다.

마지막 라인이 아닌 경우에는 현재 부호화한 라인을 복호한 값을 참조 화소 값으로 사용하여 라인 단위 예측을 계속 수행한다. 현재 부호화한 라인이 마지막 16번째 라인인 경우에는 현재 매크로블록에 대한 라인 기반 예측을 종료한다.

제안한 알고리즘은 라인 기반 인트라 예측을 사용하기 위해 SATD로 최적 모드를 결정한 후, 라인 단위로 부호화 및 복호화를 반복하기 때문에 기존의 H.264/AVC 인트라 16×16 예측보다 복잡도는 높지만, 매크로블록의 모든 화소에 대해 바로 직전의 화소 값을 참조 화소로 사용하기 때문에 예측의 정확도가 높아져 부호화 비트를 감소시킬 수 있다.

4. 실험 결과 및 분석

본 논문에서는 H.264/AVC 기준 소프트웨어 JM 12.2에서 제안한 알고리즘을 구현하였다 [4]. 이 실험에서는 1920×1080 크기를 가지는 4개의 4:2:0 포맷의 실험 영상 (BasketballDrive, ChristmasTree, Cactus, 그리고 BQTerrace) 첫 프레임을 사용하였다. 실험에서 사용한 부호화 변수는 표 2와 같다.

표 2. 부호화 변수

<i>ProfileIDC</i>	100
<i>LevelIDC</i>	51
<i>QPISlice</i>	16, 20, 24, 28
<i>Transform8x8Mode</i>	1
<i>SymbolMode</i>	1

이 실험은 H.264/AVC 표준의 High 프로파일에서 수행하였다. 제안한 알고리즘과 H.264/AVC 표준의 High 프로파일의 부호화 비트 수 비교를 위해 High 프로파일의 추가 옵션인 인트라 8×8 예측을 사용하여 인트라 4×4, 인트라 8×8, 인트라 16×16 예측을 모두 사용하도록 설정하였다.

표 3은 H.264/AVC 표준과 제안한 알고리즘으로 부호화한 결과를 비교한 것이다. 부호화 성능을 평가하기 위한 객관적 지표로, BDBR(Bjonteggard Delta Bitrate)와 BDPSNR(Bjonteggard Delta PSNR)을 이용하였다 [5]. 이는 각각 비교하고자 하는 두 방법 간의 비트율과 PSNR 차이의 평균을 의미한다. BDBR에서의 (-) 부호와 BDPSNR에서의 (+) 부호는 부호화 성능의 이득을 나타낸다. 표에서 보는 바와 같이, 제안한 알고리즘은 H.264/AVC에 비해 BDBR로 평균 약 6.92%, BDPSNR로 평균 약 0.37dB의 부호화 이득을 제공한다.

표 3. 제안한 알고리즘의 성능

	BDBR (%)	BDPSNR (dB)
BasketballDrive	-14.90	+0.51
ChristmasTree	-4.00	+0.36
Cactus	-5.78	+0.29
BQTerrace	-2.98	+0.30
Average	-6.92	+0.37

그림 5는 여러 특성을 가진 실험 영상에 대한 비트율-왜곡 곡선을 나타낸다. 이 그림은 양자화 계수가 각각 16, 20, 24, 28인 경우에 대한 결과를 보여주는데, 제안한 알고리즘이 특히 양자화 계수가 낮은 고비트를 환경에서 더 좋은 성능을 보여주는 것을 확인할 수 있다.

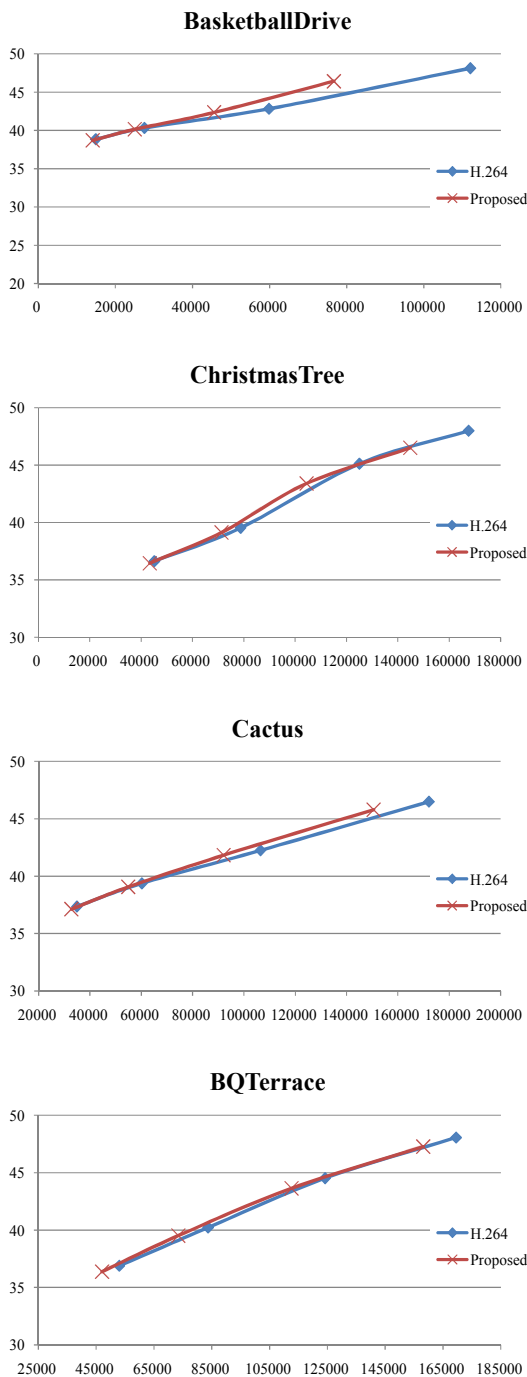


그림 5. 비트율-왜곡 곡선

표 4는 실험 영상 중 하나인 BasketballDrive 첫 프레임을 H.264/AVC와 제안한 알고리즘으로 부호화 한 후의 모드 분포를 보여 준다. 제안한 알고리즘을 사용함으로써, 기존에 인트라 4x4 모드로 부호화된 블록들이 인트라 16x16 모드로 부호화되는 것을 확인할 수 있다. 이는 제안한 인트라 16x16 예측 방법을 통해 인트라 16x16 예측의 정확도가 높아져 이것이 최적 모드로 결정되었기 때문이다.

표 4. 제안한 알고리즘의 성능

	인트라 4x4 (%)	인트라 16x16 (dB)
H.264/AVC	75.7	24.3
제안한 방법	47.0	53.0

인트라 4x4는 각 4x4 블록마다 부호화되는 최적 모드 정보가 전송되어야 하므로 인트라 16x16에 비해 부호화해야 할 비트가 많다. 따라서, 제안한 알고리즘을 사용하면 부호화해야 할 비트가 많이 발생하는 인트라 4x4 모드들이 인트라 16x16 모드로 부호화되어 인트라 부호화 시의 부호화 비트수를 상당량 줄일 수 있다.

5. 결론

본 논문에서는 H.264/AVC의 인트라 16x16 예측보다 좀더 정확한 라인 기반 인트라 16x16 예측 알고리즘을 제안하였다. 매크로블록 단위가 아니라 라인 단위로 예측 및 부호화를 수행함으로써 좀더 가까운 화소를 참조 화소로 사용할 수 있다. 이를 통해, 참조 화소와의 거리가 최대 16 화소인 기존의 방법보다 예측의 정확도를 높일 수 있다. 실험 결과로부터 제안한 라인 기반 인트라 16x16 예측 방법이 H.264/AVC의 High 프로파일에 비해 평균 약 6.92%의 비트 수를 감소시키는 것을 확인했다.

감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터의 지원과 ETRI의 광가입자망(FTTH) 서비스 개발 실험사업 과제에 의한 것입니다. (NIPA-2009-(C1090-0902-0017))

참고 문헌

- [1] T.Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits System Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.
- [2] I. Richardson, H.264 and MPEG-4 Video Compression. Wiley, 2003.
- [3] G. Sullivan, T. McMahon, T. Wiegand, and A. Luthra, Eds., Draft Text of H.264/AVC Fidelity Range Extensions Amendment to ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC, ISO/IEC JTC1/SC29/WG11 and ITU-T Q.6/SG16 Joint Video Team Document JVT-L047, July 2004.
- [4] http://iphome.hhi.de/suehring/tml/download/old_jm/jm12.2.zip, Joint Video Team, Reference Software Version 12.2.
- [5] G. Bjontegaard, "Improvements of the BD-PSNR model", ITU-T SG16/Q6, 35th VCEG Meeting, Berlin, Germany, 16th-18th July, 2008, Doc. VCEG-A111.