

움직임 정보를 이용한 H.264에서의 적응적 참조 화면 선택 기법

*권혁민, 최윤식
연세대학교 영상정보연구실
e-mail : KongMin@hitel.net

Adaptive Reference Picture Selection Method Using Motion Information in H.264

*Hyok-Min Kwon, Yoon-Sik Choe
Image Information Lab
Yonsei University

Abstract

본 논문은 동영상 압축 표준 H.264에서 화면간 예측 부호화시 부호화 효율을 높일 수 있는 참조 화면 선택에 대한 기법을 제안하였다. 선택에 대한 문제를 참조 화면 간 움직임 정보를 고려해서 가장 유사한 두 참조 화면을 구하는 것으로 해결하고자 하였고 움직임 정보에 대한 고려는 부호화시 계산되는 R-D 비용함수 값을 사용함으로써 부가적인 계산을 최소화 하고자 하였다. 실험결과 기존의 슬라이딩 윈도우 방식에 비해 Bit Rate 감소로 인한 부호화 효율개선을 얻을 수 있었다.

I. 서론

최근 표준화 과정에 있는 동영상 압축 표준 H.264는 기존의 표준들에 비해 부호화 효율이 비약적으로 증가하였고 그 구문 요소는 상대적으로 간단하고 단순하게 이루어져 있다[1]. 특히 4x4 정수 변환, 다양한 Intra/Inter 예측 모드, Deblocking 필터, 향상된 Entropy 부호화 등을 통해 그 부호화 효율을 높이고 있으며 어느 하나의 기법에 의한 큰 효율 개선 보다는 각각의 기법에 따른 작은 개선으로 전체적인 향상을 나타낸다. 이 중에서 가장 다양하게 응용될 수 있는 부분은 Intra/Inter 예측에 관한 부분이다. 공간적/시간적 중복성의 제거는 부호화 효율개선에 있어서 가장 많은 효과를 보이거나 부호화 복잡도 또는 시간과 trade-off 관계가 있기 때문에 각 응용 분야에 적합한 Intra/Inter 예측 모드의 사용 정도와 방법을 선택함으로써 그 효율과 복잡도의 균형을 이룰 수 있다. H.264에서는 화면내 부호화시 방향성을 고려한 다양한 Intra 예측 모드를 두었으며 그 속도 개선을 위해 Edge Histogram을 이용한 모드 예측 방법 등의 선택가능한 모드의 수를 줄임으로서 복잡도를 개선하는 연구가 되고 있다. 화면간 부호화시에는 한 장이 아닌 여러 장

의 복호된 화면들을 참조해서 16x16크기에서 4x4크기까지의 다양한 블록 단위 까지 1/4pel의 정확도로 움직임 추정/보상(ME/MC)을 할 수 있다. 기존의 표준들보다 다양한 예측 모드가 가능해졌으나 그 복잡도가 상당히 증가하였기 때문에 실제적인 응용을 목적으로 통계적인 관찰을 이용한 예측 모드의 간단화 또는 적절한 참조 화면 선택을 통한 빠른 움직임 추정/보상 등 화질의 면에서 약간의 손해를 감수하더라도 복잡도와 속도의 측면에서 많은 개선을 이룰 수 있는 방법에 대한 연구들이 많이 이루어지고 있다. 그러나 복잡도와 속도의 문제뿐만 아니라, H.264 표준에 채택되고 정의되어 있는 여러 부호화를 위한 도구들의 활용이 표준화 과정 중에 고려된 것 외에도 어떤 것들이 가능하고 어느 정도 까지 그 효율을 낼 수 있는지와 같은 활용과 효율에 관한 문제도 이제 막 표준화가 마무리되는 시점에서 상당히 중요할 것이다. 따라서 본 논문에서는 여러 장의 참조화면으로부터 Inter 예측 부호화시 복잡도와 시간의 증가 없이 부호화 효율을 개선하는 것을 목적으로 움직임 추정 후의 움직임 정보를 이용한 참조 화면 선택을 하는 기법을 제안하고자 한다.

II. 배경 지식

H.262/MPEG2, H.263, MPEG4 등의 기존 압축 규격에서는 Inter 예측 부호화시 P 화면일 경우 바로 전에 복호화된 가장 최근의 I 또는 P 화면 한 장을 참조해 부호화한다. B 화면일 경우 바로 전과 후 한 장씩, 모두 두장에서 참조해 부호화하지만 역시 시간적 전후관계에서 한 장씩만을 참조한다. 그러나 영상 신호는 각 화면 간 유사성이 시간적으로 길게 지속되거나 반복되어 나타날 수도 있고 짧은 기간 동안의 화면 간에서는 움직임에 따른 다양하고 세밀한 위치관계를 가진 유사성이 있을 수 있다. 따라서 참조 화면을 한 장이 아닌 여러 장을 들으므로 부호화 효율을 개선할 수 있으며 그 효율은 참조 화면 장수의 로그 함수와 거의 선형적인 관계가 있다[2]. 또한 입력 되는 영상의 특징에 따라서 그 효율의 개선 정도가 달라질 수 있으며 특히 움직임이 많은 영상에서 더 큰 효율을 얻을 수 있다. 이러한 여러 장의 참조 화면을 두는 것은 동영상 압축 규격으로서는 H.263++ Annex U에 정식으로 채택되었다. 이후 H.26L이 시작되면서 H.263++ Annex U의 구문을 이어받아 지금까지 표준화 과정을 거치며 수정되어 현재 H.264에 이르렀다.

H.264에서 참조 화면은 부호화기가 현재 화면을 부호화한 후 부호화기 내부의 복호화기에서 재 복호화 화면이 메모리 버퍼에 저장되었을때 그 화면에 대한 제어가 슬라이스 헤더 정보에 의해 이루어진다. 즉 슬라이스 헤더 정보에 의해 부호화기 또는 복호화기의 복호된 화면 버퍼에 있는 화면들에 대한 제어가 이루어지고 참조 화면으로서 선택과 버림, 재배치, 인덱싱이 이루어진다. 복호된 화면 버퍼에는 별도의 제어 신호 없이 복호된 최근 몇 장을 FIFO 방식으로 복호된 화면 버퍼에 가지고 있는 Short-Term Memory와 복호된 화면 중에서 임의로 참조 화면으로 지정하고 버려질 수 있는 상대적으로 시간적인 지속성이 긴 Long-Term Memory로 나누어진다. Short-Term과 Long-Term을 합친 수가 전체 참조 화면 수가 되고 Inter 예측 부호화 시 참조가 되는데 실제 활성화된 참조 화면 수는 각 슬라이스 마다 전체 참조 화면 이내에서 갱신될 수 있다.

슬라이스 헤더에서 참조 화면에 대한 제어 동작은 dec_ref_pic_marking(DRPM) 정보에 의해 이루어진다. DRPM 정보에 따라 각 화면은 참조 화면으로 비사용, Short-term으로 사용, Long-Term으로 사용 이렇게 세가지 중 하나로 선택되어 결정된다. 현재 화면이 IDR 부호화 된 슬라이스일 경우 이전까지의 참조 화면에 대한 초기화가 이루어지고 현재 IDR 화면을 Short-Term 또는 Long-Term 으로 할 것인지에 대한 선택이 가능하다. 한편 IDR 부호화 화면이 아닌 예측 부호화된 화면일 경우 adaptive ref pic marking mode

flag에 의해 참조 화면 선택에 대한 모드를 결정할 수 있게 된다. 이 flag가 0일 경우 슬라이딩 윈도우(Sliding window) 참조 화면 선택 모드로서 Short-Term 참조 화면들에 대해서 FIFO 방식으로 선택하고 버림이 이루어지는 모드이다. 반면 1일 경우 적용적인 참조 화면 선택 모드로서 참조 화면 제어를 위한 구문들이 제공되고 그 구문들을 통해 참조 화면으로서 버리거나 Long-Term으로의 지정 등이 가능하다. 이때 참조 화면 제어를 위한 구문 memory management control operation에는 모두 6가지가 있으며 그 동작은 표1과 같다.

표 1. Memory Management Control Operation

Value	MMCO
0	MMCO 무프롤 마침
1	Short-Term 화면을 비사용으로 지정
2	Long-Term 화면을 비사용으로 지정
3	Short-Term 화면을 Long-Term으로 지정
4	최대 Long-Term 화면 인덱스 지정
5	모든 참조 화면을 비사용으로 지정
6	현재 복호된 화면을 Long-Term으로 지정

각 MMCO에 대해 화면 번호나 인덱스 번호가 부가적으로 따라오며 MMCO가 이루어질때마다 참조 화면에 대한 정보를 가지고 있는 Ref Picture List 0 또는 Ref Picture List 1이 갱신되고 이 리스트에 의해 다음 화면의 Inter 예측 부호화시 참조 된다.

III. 제안한 방법

현재 JM에서는 adaptive ref pic marking mode가 항상 0으로 되어 있어 참조 화면 수만큼 최근 복호된 화면을 Short-Term 참조 화면으로 가지고 있고 슬라이딩 윈도우 방식으로 복호된 화면 버퍼에서 참조 화면 선택이 이루어지고 있다. 실험적으로 보면 영상 신호를 Inter 예측 부호화 하는 경우 시간적인 거리를 다양화 하여 참조 화면을 여러 장 두고 부호화 했을 때 그 시간적 거리가 짧은, 즉 가장 최근의 참조 화면일 수록 참조되는 빈도가 높다. 따라서 구현이 간단한 장점까지 가지고 있는 슬라이딩 윈도우 방식의 Short-Term 참조 화면 모드가 기본으로 되어 있다. 그러나 입력되는 영상 신호와 특징과 상관 없이 최근의 화면만을 참조 화면의 대상으로 하고 있는 현재의

방식에서는 Inter 예측 부호화시에 효율 개선에 상대적으로 크게 영향을 줄 수 있는 화면이 선택 되지 못하고 버려지는 경우가 생길 수 있다. 특히 예로 반복되는 영상의 경우 전환 되는 화면을 Long-Term으로 둘 경우 Short-Term보다 더 나은 효율을 보일 수 있다. 따라서 단순히 Short-Term만을 두고 있는 현재 방식에서는 영상 신호의 특징에 따라 그 부호화 효율이 개선될 여지가 있다. 결국 부호화기/복호화기의 복잡성과 비용, 메모리 대역폭의 제한에 따라 참조 화면의 수가 주어졌을때, 부호화 효율을 가장 높일 수 있는 참조 화면으로 어떤 것을 선택할 것인가에 대한 문제를 해결하고자 했으며 그 선택은 영상간 움직임 정보를 이용해 결정하였다.

가장 최적의 참조 화면 선택은 부호화기가 미리 영상 부호화 해보고 그중 가장 좋은 효율을 나타내는 참조 화면 조합을 선택하는 방법일 것이다. 그러나 이 방법은 이미 부호화기가 부호화할 영상을 알고 있어야 하고 부호화 시간의 제약이 없는 상황에서 가능하므로 현실적인 어려움이 많다. 따라서 참조 화면 선택시 그 이후의 화면을 미리 알 수 없는 온라인 상황을 가정하고 참조 화면 선택 과정에서 부가적인 계산을 최소화하고자 하였다.

부호화 할 다음 화면을 모르는 상황에서 현재 화면을 부호화한 후 예측 부호화시 참조 했던 여러 장의 참조 화면 중에서 하나 f를 선택하여 버리고 현재 부호화한 화면 F를 참조 화면으로 지정하는 상황을 가정하자. 그러면 다음 화면을 부호화 할 경우에는 현재의 참조 화면 중에서 f가 빠지고 F가 대신 참조 화면이 된 상태에서 예측 부호화를 수행할 것이다. 이때 다음 화면 예측 부호화시의 효율을 위해서는 f에서 참조되는 경우와 F에서 참조되는 경우의 효율차이가 최소가 되는 f를 선택하여야 한다. 즉 현재 화면 부호화시 참조했던 참조 화면 n개와 다음 화면 부호화시 참조 화면이 될 대상이 되는 F가 포함된 n+1개의 참조 화면의 상황에서 다음의 화면을 예측 부호화 할 때 참조 화면 한 장이 이 버려지는(n+1개 -> n개) 것에 의한 효율 저하 되는 부분을 최소화하기 위해서는 서로 가장 유사한, 즉 각각 화면에서부터 예측 부호화시 효율 차이가 최소가 되는 두 장중 한 장을 선택해 버려야 한다. 그러면 나머지 참조 화면들은 자연적으로 최대한 서로 다양한 움직임 위치 관계를 가지는 것들로 선택되어 남아있게 되고 이는 미리 알 수 없는 다음의 영상 화면을 부호화 할 때 부호화 효율을 증가 시킬 수 있다. 결국 n개의 참조 화면 중에서 최근 부호화된 화면 F가 추가 될 때 버려지는 화면 f로 어떤 것을 선택하느냐의 문제는 n개의 참조 화면 중에서 서로 유사성이 많이 비슷한 부호화 효율을 나타낼 수 있는 두

장의 참조 화면 중 한 장을 선택하는 것으로 해결할 수 있다. 이때 둘 중에 하나를 선택하는 경우에는 시간적으로 가장 오래전 화면을 f로 선택한다. 따라서 참조 화면 개수 n이 2 이하인 경우에는 항상 최근의 화면을 참조 화면으로 가지고 있게 되므로 슬라이딩 윈도우 방식과 같아진다. 그러므로 여기서 제한하고자 하는 기법은 참조 화면의 수가 3 이상일 경우를 대상으로 한다.

n개의 참조 화면 중에서 서로 유사한 두 장의 조합을 찾는 방법 중 가장 간단한 방법은 n가지의 가능한 각 조합마다 두 화면 서로를 화소대화소로 차이를 구하는 것이다. 하지만 이 방법은 블록 기반으로 움직임을 고려해 부호화 하는 현재 영상 부호화 표준에는 맞지 않다. 따라서 두 화면 간 유사도는 화면의 화소 값에만 의존하기 보다는 화면 간 움직임도 고려해야 할 것이다. 추가적인 계산 없이 화면 간 움직임을 고려한 유사도를 측정하기 위해 현재 화면을 Inter 예측 부호화하는 과정에서 나오는 정보들을 이용하고자 한다. 즉 두 장의 참조 화면을 서로 직접 비교해 유사도를 찾는 것이 아니라 부호화하는 화면의 기준에 대해서 유사도를 찾자 한다.

Inter 예측 부호화시 움직임 추정은 R-D 비용 함수에 따른다[3]. 이 비용 함수가 가장 적어지는 모드로 결정이 일어나고 Intra가 아닐 경우 움직임 벡터와 참조 화면의 번호, 차이 계수가 구해진다.(식 참조)

$$J(Rf|\lambda) = SAD(s, c(Rf, m(Rf))) + \lambda (R(m(Rf)) - p(Rf)) + R(Rf))$$

이 움직임 정보들을 통해 각 MB마다 어느 참조 화면과 얼마만큼 움직임이 일어났고 그 차이 계수가 어느 정도인지 알 수 있다. 그러나 MB간 다양한 모드가 존재하고 출력 비트의 양 또한 고려해야 하기 때문에 움직임 정보를 직접 비교하여 유사도를 찾기는 어렵다. 따라서 움직임 정보와 출력 비트, QP값이 고려되어 있는 R-D비용함수 결과 값을 이용해 그 유사도를 얻을 수 있다. 움직임 추정시 각 MB마다 n개의 참조 화면에 대해 R-D 비용함수 값이 나오는데 두 참조 화면에 대한 비용 함수 값의 차이가 적을 수록 두 화면은 유사하다고 할 수 있다. 따라서 각 MB마다 구해진 R-D 비용 함수값을 참조 화면에 따라서 합산하고 그 차이가 가장 적은 두 장이 가장 유사도가 높다고 할 수 있다. 이때 그 차이가 임계치 T보다 높다면 현재 참조 화면중에서는 유사하다고 볼 수 있는 참조 화면 쌍이 없는 것으로 보고 참조 화면 중에서 시간적으로 가장 오래 전 화면을 선택해 참조 화면에서 제외시킨다. 이

러한 적응적인 참조 화면 선택을 위해서 주어진 참조 화면 수 n 에서 가장 최근의 화면을 저장하는 Short-Term 으로 1장을 두고 Long-Term 으로 $n-1$ 장을 둔다. Short-Term 메모리에는 가장 최근에 부호화한 화면 이 저장되고 Long-Term 메모리에는 적응적인 MMCO 를 통해 선택적으로 참조 화면이 저장될 것이다.

IV. 실험 결과

위에서 제안한 참조 화면 선택에 대한 기법을 JM 7.3에 실험하였다. 실험 조건은 7가지 모든 Inter 예측 모드를 사용하였고 움직임 추정 범위는 $-16 \sim 16$ 화소로 하였다. 엔트로피 부호화는 CAVLC를 사용하였고 Hadamard 변환과 R-D 최적화를 사용하였다. B 슬라이스는 사용하지 않았고 IPPP구조로 고정 QP값으로 26을 주었다. Frame모드로 부호화하고 참조 화면의 수는 4장을 두었다. 실험에 사용된 영상은 foreman, akiyo를 사용하였고 1장씩 skip하여 총 20장을 부호화하였다.

표 2. 입력 영상이 foreman일때 부호화 효율

	PSNR	Bit Rate
슬라이딩 윈도우 (n=4)	37.80	158992
슬라이딩 윈도우 (n=5)	37.80	150208
제한한 방식 (n=4)	37.80	154560

표 3. 입력 영상이 akiyo일때 부호화 효율

	PSNR	Bit Rate
슬라이딩 윈도우 (n=4)	39.94	50080
슬라이딩 윈도우 (n=5)	39.92	49760
제한한 방식 (n=4)	39.94	49984

실험 결과를 살펴보면 제한한 방식을 사용했을때 기본의 슬라이딩 윈도우 방식보다 PSNR에서는 별 차이가 없으며 Bit Rate에서는 약간 감소한 것을 알 수 있다. 참조 화면에 따른 PSNR의 개선 정도는 입력 영상에 따라 많은 영향을 받고 특히 고정 QP값을 사용하였기 때문에 단기간 동안의 영상을 참조해 예측 부호화를 하면 거의 일정한 화질을 나타낸다. 따라서 참조 화면 수가 같으면 적응적으로 참조 화면을 선택한다 하더라도 시간적으로 거리가 거의 비슷한 화면들이 참조 화면이 되기 때문에 화질적인 측면에서는 개선 효과가 미비하다. 반면 Bit Rate가 약간 씩 감소하였는데 이는 기존 방식보다 상대적으로 더 다양한 움직임 관계의 화면들을 참조 화면으로 가지고 있음으로서 비슷한 화

질을 나타낼때 상대적으로 적은 Bit를 소모하기 때문이다. 그러나 한편으로는 적응적인 참조 화면 선택을 위한 MMCO 구문이 첨가 됴으로서 슬라이스 헤더의 Bit는 약간씩 증가하였다. foreman과 akiyo를 비교해보면 foreman의 개선 정도가 더 큰 것을 알 수 있는데 foreman처럼 akiyo보다 상대적으로 움직임이 큰 영상은 참조 화면 증가에 따른 효율 개선이 높고 제한한 기법 또한 참조 화면 증가와 비슷한 효과를 나타내기 때문이다. 전체적으로 화질면에서는 비슷하지만 Bit Rate 감소에 의한 부호화 효율 개선 효과가 있었으며 부호화기에 R-D 최적화에 따른 QP 제어를 한다면 화질 개선 효과도 기대할 수 있을 것이다.

V. 결론

본 논문에서는 다중 참조 화면을 이용해서 화면간 예측 부호화시 부호화 효율을 높일 수 있는 참조 화면 선택에 대한 기법을 제안하였다. 부호화 과정에서 R-D비용함수로 계산되는 움직임 정보를 고려해 부가적인 계산 없이 참조 화면을 선택함으로써 복잡도 증가를 최소화 하였고 실험 결과 Bit Rate 감소로 인한 효율 개선이 있었음을 확인하였다. 기존의 수동적인 참조 화면 선택보다는 단기간이래도 영상 화면 간 특징을 고려해 적응적으로 참조 화면을 선택함으로써 주어진 참조 화면 수내에서도 부호화 효율을 개선할 수 있음을 보인 것이다.

참고문헌

- [1] JVT of ITU-T and ISO/IEC JTC 1, "Draft ITU-T Recommendation and FDIS of Joint Video Specification", JVT-G050r1, May, 2003.
- [2] T.Wiegand X.Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction", IEEE Transaction on Circuits and Systems for Video Technology, Sep, 1998
- [3] JVT of ITU-T and ISO/IEC JTC 1, "JM Reference Encoding Methods and Decoding Concealment Methods", JVT-I049d0, Sep, 2003.