

MPEG-4 비디오 코딩을 위한 Scene Level Rate Control 알고리즘

김 용 욱*, 신 윤 식*, 허 도 근*
원광대학교 전자공학과

Scene Level Rate control Algorithm for MPEG-4 Video Coding

Kim Yong-Wook*, Shin Youn-Shic*, Huh Do-Geun*
Dept. of Electronic Eng. Wonkwang Univ.

요 약

MPEG-4 부호화기는 16×16 픽셀로 이루어진 매크로 블록을 기본 단위로 하는 비디오 객체에 대해 움직임 추정과 움직임 보상을 수행한다. 이때 이전 블록과 현재 블록간의 차이를 8×8 블록으로 DCT하여 압축한다. DCT는 영상 신호의 에너지를 낮은 대역으로 집중시켜 낮은 에너지를 갖는 고주파 성분을 제거하여 정보를 압축하므로 복원된 영상은 블록의 경계선이 표시되는 블록킹 현상이 발생한다.

본 논문에서는 4×4 블록과 8×8 블록을 이용하여 DCT 하는 가변 블록 변환 방식을 사용한다. 이는 작은 크기의 블록을 사용하므로 블록내 주파수의 분포를 비교적 좁은 영역으로 변환할 수 있어 블록킹 현상을 줄일 수 있다. 또한 INTRA 프레임의 DCT 과정에서 4개의 DCT 변환 블록중 하나를 이웃하는 블록의 DC 계수와 화소값들을 이용하여 그 값을 예측하여 DCT계수를 결정하고 INTER 프레임에서는 간략화된 정수 DCT의 사용과 작은 블록단위의 작업으로 인한 부호화의 부담을 줄인다.

I. 서 론

H.263이나 MPEG-4와 같은 동영상 부호화 방식은 16×16 픽셀로 이루어진 매크로블록을 기반으로 부호화를 수행하며 움직임 추정과 움직임 보상을 통해 얻어진 차신호를 8×8 DCT(Discrete Cosine Transform)의 블록

변환을 이용하여 압축한다. 변환에 사용되는 블록의 크기는 움직임 추정과 움직임 보상에 사용되는 블록의 최소 크기를 결정한다. DCT는 영상신호의 에너지를 낮은 주파수 대역으로 집중하여 낮은 에너지를 갖는 고주파 영역의 정보를 제거하여 신호의 정보를 압축한다. DCT는 높은 비트율의 환경에서는 좋은 성능을 나타내지만 고압축 동영상 부호화에서는 심한 블록화 현상으로 인해 급격한 화질 열화를 발생시키는 단점이 있다. 동영상 부호화에서 8×8 DCT를 이용할 경우 부호화될 프레임내의 움직임에 적절히 대응하지 못하는 경우가 발생한다. 따라서 작은 크기의 블록을 이용하여 차신호의 부호화를 행하는 것이 유용하다. 큰 변환 블록을 사용하는 경우는 높은 에너지의 압축을 얻을 수 있지만 자세한 움직임의 복원이 힘들다. 반면 작은 변환 블록을 사용하는 경우는 계산량의 증가를 가져오지만 자세한 움직임의 복원이 가능하다.

본 논문에서는 한 개의 프레임을 하나의 객체로 하는 MPEG-4 부호화기에서 동작하는 가변적인 블록크기에 대한 효율적인 변환방법을 제안한다. INTRA 프레임의 DCT 변환에 대해 공간적으로 이웃하는 블록간의 상관성이 가장 높으므로 4개의 DCT 변환 블록중 하나를 이웃하는 블록의 DC 계수와 화소값들을 이용하여 DCT 계수의 값을 예측하는 방법을 이용하고 INTER 프레임에 대해서는 연산량의 감소를 위해 간략화된 정수를 이용한 DCT(ICT : Integer DCT)방법을 사용하여 작아진 블록크기에 비례하여 증가하는 블록의 수로 인한 DCT 변환에 소요되는 연산을 감소시켜 부호화의 부담을 줄

인다. ICT는 실험을 통해 4×4와 8×8 크기의 블록에 대해서는 DCT와 거의 동일한 성능을 보인다.

II. INTRA 프레임에서 블록간의 상관성을 이용한 DCT 변환

임의의 한 블록은 자신과 공간적으로 이웃하는 블록들과 상관성이 가장 높다. 따라서 현재 블록의 DCT 계수는 이웃하는 블록의 DCT 계수로부터 예측이 가능하다. 현재 블록의 DCT 계수의 예측은 그림 1과 같다.

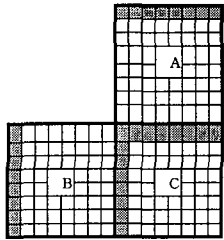


그림 1. 예측에 사용되는 DCT 블록의 공간적 구성

현재 DCT 계수를 예측하기 위한 블록은 C 블록이며 C 블록의 DC 계수의 예측은 식 (1)과 같다.

$$C(0, 0) = (A(0, 0) + B(0, 0)) / 2 \quad (1)$$

여기서 A(0,0)과 B(0,0)는 각각 A 블록과 B 블록의 DC 계수이다. C 블록의 가장 상측에 있는 AC 계수의 예측값은 식 (2)와 같고 C 블록의 가장 좌측에 있는 AC 계수의 예측값은 식 (3)과 같다.

$$C(u, 0) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (c(i, j) - a(i, j)) \cos(\frac{2\pi}{N} ui) \cos(\frac{2\pi}{N} vj) \\ = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (c(i, j) - a(i, j)) \cos(\frac{2\pi}{N} ui) \quad (2)$$

$$C(0, u) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (c(i, j) - a(i, j)) \cos(\frac{2\pi}{N} ui) \cos(\frac{2\pi}{N} vj) \\ = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (c(i, j) - a(i, j)) \cos(\frac{2\pi}{N} vj) \quad (3)$$

여기서 c(i, j)와 a(i, j)는 각각 C 블록과 A 블록의 화소값이다. 식 (2)에서 현재 블록의 위치 벡터는 I(x, y) = c(i, j) 이고 상측 블록의 위치 벡터는 I(x, y-N) = a(i, j) 이므로 C 블록과 A 블록간의 화소차는 c(i, j) - a(i, j) = I(x, y) - I(x, y-N) 이 된다. 즉 DCT 계수의 예측 방법은 8화소의 거리에 떨어져 있는 화소로 예측된다. 그러나 화소의 거리가 증가할수록 화소간의 상관성은 떨어지므로 식 (2)에서 사용된 상측블록의 화소를 상측블록의 가장 하단에 위치한 화소로 대체하여 화소간의 상관성을 높이며 이는 식 (4)와 같다.

$$C(u, 0) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (c(i, j) - a(i, N-1)) \cos(\frac{2\pi}{N} ui) \\ = c(u, 0) - N \sum_{i=0}^{N-1} a(i, N-1) \cos(\frac{2\pi}{N} ui) \quad (4)$$

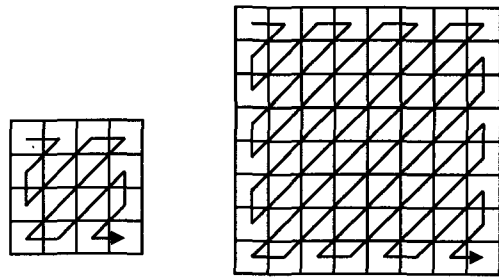
이러한 관계는 식 (3)에 대해서도 성립한다. 예측과정에서 제외된 영역의 DCT 계수는 예측된 AC 계수를 평균하여 일차원적으로 대입된다.

III. 가변적 변환 블록에서 INTER 프레임의 부호화

본 논문에서 사용된 변환 블록은 그림 2와 같이 4×4 크기의 블록과 8×8 블록을 사용한다.

그림 2와 같이 변환 블록의 크기는 4×4 블록을 기본으로 하여 블록을 확장하는 형태로 이루어진다. 이전 블록과 현재 블록의 차이를 구하면 우선 8×8 블록에서 움직임이 발생하지 않은 영역과 움직임이 발생한 영역을 나눈다. 움직임이 발생하지 않은 영역의 픽셀 값을 0으로 치환하고 움직임이 발생한 영역은 선택적으로 4×4 블록을 확장하여 적용한다. 매크로블록에 대한 움직임이 발생한 영역을 찾는 과정은 그림 3과 같다.

이때 부호화에 사용되는 블록의 수가 증가하여 계산을 복잡하게 하므로 실수값만을 사용하여 변환을 행하는 DCT를 ICT로 대체한다. 많은 간략화 방법들이 연구되었는데 본 논문에서는 Srinivasan과 Rao의 방법에 기반한 ICT를 사용한다.[1][2][3]



(a) 4×4 (b) 8×8
그림 2. 변환에 사용되는 가변 블록과 스캔방향

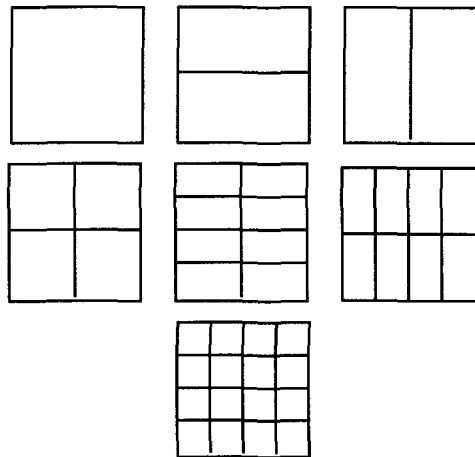


그림 3. 움직임 발생영역의 탐색

ICT는 4×4 Hadamard 변환을 기초로 변환 행렬을 유도한다. Hadamard 변환을 이용한 차신호의 부호화는 다음과 같다.

$$D_H = H_4 \cdot D_{mc} \cdot H_4^T \quad (5)$$

여기서 D_{mc}는 움직임 추정을 통해 구해진 이전 블록과의 차신호이다. 여기서 사용되는 H₄는 다음과 같다.

$$H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix} \quad (6)$$

식 (5)의 H_4 는 수직성분의 변환을 수행하고 H_4^T 는 수평방향의 변환을 수행한다. 움직임 추정을 위한 오류의 척도로 변환 영역내의 절대차의 합(SATD : Sum of Absolute differences in Transform Domain)이 사용된다. SATD는 왜곡에 대한 변환율을 나타낸다. 4x4 블록의 SATD는 다음과 같다.

$$SATD = \sum_{i=0}^3 \sum_{j=0}^3 |d_H(i, j)|, \quad D_H = \{d_H(i, j)\} \quad (7)$$

움직임 추정에서 SATD는 매크로블록의 모든 4x4 블록에서 계산되고 매크로블록의 움직임 벡터의 부호화를 위해 필요한 bit rate에 더해진다.

식 (6)의 Hadamard 변환 행렬을 기본으로하여 유도한 4x4ICT 변환 행렬은 다음과 같다.

$$T_4 = \begin{pmatrix} 13 & 13 & 13 & 13 \\ 17 & 7 & -7 & -17 \\ 13 & -13 & -13 & 13 \\ 7 & -17 & 17 & -7 \end{pmatrix} \quad (8)$$

4x8과 8x4 크기의 변환 행렬은 식 (8)의 변환 행렬을 확장하여 표시한다. 8x8 크기의 변환 행렬은 다음과 같이 구할 수 있다.

$$T_8 = \begin{pmatrix} 17 & 17 & 17 & 17 & 17 & 17 & 17 & 17 \\ 24 & 20 & 12 & 6 & -6 & -12 & -20 & -24 \\ 23 & 7 & -7 & -23 & -23 & -7 & 7 & 23 \\ 20 & -6 & -24 & -12 & 12 & 24 & 6 & -20 \\ 17 & -17 & -17 & 17 & 17 & -17 & -17 & 17 \\ 12 & -24 & -6 & 20 & -20 & -6 & 24 & -12 \\ 7 & -23 & 23 & -7 & -7 & 23 & -23 & 7 \\ 6 & -12 & 20 & -24 & 24 & -20 & 12 & -6 \end{pmatrix} \quad (9)$$

$n \times m$ 크기의 2차원 신호를 ICT하기 위한 식은 다음과 같다

$$C_{n, m} = T_v \cdot D_{n, m} \cdot T_h^T \quad (10)$$

여기서 $B_{n, m}$ 은 $n \times m$ 픽셀의 신호 블록을 표현하고 $C_{n, m}$ 는 변환된 계수를 나타낸다. T_v 와 T_h 는 각각 $m \times m$ 와 $n \times n$ 의 크기를 갖는 수직방향과 수평방향의 변환 행렬이다. 만일 T_v 와 T_h 가 직교변환일 경우

$$T_i \cdot T_i^T = T_i^T \cdot T_i = I \quad (i = h, v) \quad (11)$$

을 만족한다.

III. 양자화 및 부호화

본 논문에서 사용된 부호화기는 INTRA DC 계수를 위한 1개의 양자화기와 나머지 계수들(Non-INTRADC)을 위한 31개의 양자화기를 가진다. INTRADC 계수를 위한 양자화기의 양자화 간격(step size)을 8로하여 균일 양자화된다. 나머지 31개의 양자화기의 양자화 간격은 2에서 64까지의 짝수 값으로 균일 양자화된다. 매크로블

록 계층의 양자화 간격을 결정하는 QP(Quantization Parameter)를 변화시켜 버퍼에 유입되는 비트열의 양을 제어한다. 따라서 QP는 부호화된 비트열의 발생률에 따라 1에서 31까지의 값을 갖고, 양자화 간격은 2QP가 된다. 그리고 INTRA 프레임을 양자화하기 위한 초기 QP 값은 16으로 한다. 변환계수 $C(i, j)$ 로 양자화할 때 INTRADC 계수는 식 (12)을, INTRA 블록의 AC 계수는 식 (13)를 이용하여 양자화한다. 또한 INTER 블록의 DC 계수와 AC 계수는 식 (14)에 의해 양자화된다.

$$QC(0, 0) = C(0, 0)/8 \quad (12)$$

$$QC(i, j) = \text{sign}[C(i, j)] \times [|C(i, j)| / 2QP] \quad (13)$$

$$QC(i, j) = \text{sign}[C(i, j)] \times [\{ |C(i, j) - QP/2| / 2QP \}] \quad (14)$$

양자화기를 INTRA, INTER 블록으로 나누어서 수행하는 이유는 INTRA 블록일 때는 변환영역의 DC 계수에 에너지가 집중하기 때문에 양자화 간격을 작게 해야만 오차가 크게 발생하지 않는다. 반면 INTER 블록에 대해서는 보상된 차 신호를 보내기 때문에 DC 계수에 많은 에너지가 집중하지 않는다. $C(i, j)$ 의 절대값을 양자화해서 전송하므로 실제로 계수를 재생하는때는 계수가 양수인지 음수인지를 구별해주는 부호 비트가 1비트 더 필요하다. 여기서 DC 계수의 중요성은 충분히 고려가 되었으나 동일 매크로블록내의 모든 AC 계수는 동일한 양자화기를 사용하여 양자화하므로 영상의 에지 성분에 대한 손실을 많이 가져온다.

양자화 계수는 가변장 부호화(VLC : Variable Length Code)를 이용하여 부호화되며 이는 블록내의 0이 아닌 마지막 계수를 나타내는 LAST와 0의 값이 반복되는 횟수를 나타내는 RUN, 실제 양자화된 계수를 나타내는 LEVEL의 조합으로 구성되는 2차원 VLC 표를 참조하여 부호화한다.[4] VLC는 그림 3에 보였다. 각 코드들은 길이 l 과 $INFO = x_n \dots x_1 x_0$ 에 의해 표시된다. $INFO$ 의 길이는 $\lfloor l/2 \rfloor$ 이고, 여기서 $\lfloor \cdot \rfloor$ 는 정수부분만을 취함을 나타낸다. 양자화된 4x4 블록의 변환계수들은 그림 1에 보인 순서에 따라 스케닝된다. EOB(End of Block)은 첫 번째 VLC 코드인 "1"과 같다.

$$\begin{matrix} 1 \\ 0 \ x_0 \ 1 \\ 0 \ x_1 \ 0 \ x_0 \ 1 \\ 0 \ x_2 \ 0 \ x_1 \ 0 \ x_0 \ 1 \\ 0 \ x_3 \ 0 \ x_2 \ 0 \ x_1 \ 0 \ x_0 \ 1 \end{matrix}$$

그림 4. VLC의 기본적인 구조. x_k 는 0또는 1

IV. 모의실험 결과

INTRA 프레임의 블록을 제한된 DCT 계수의 예측방법을 이용하여 변환한 경우와 일반적인 DCT 방식에 의해 압축을 행한후 복원한 영상의 PSNR과 비교한 결과

는 표 1과 같다. 여기서 QP는 양자화 간격이다.

표1. 평균 PSNR을 이용한 성능 비교

QP	Foreman		Susie	
	기존방법	제안된방법	기존방법	제안된방법
4	38.45	38.45	39.78	39.89
5	37.56	37.55	38.65	38.79
7	35.33	35.39	35.21	35.34
10	31.26	31.53	30.22	30.45
15	29.78	29.65	29.65	29.44
25	28.15	28.20	28.30	28.20
평균	33.42	33.46	33.63	33.69

직교변환의 부호화 이득은 다음과 같이 정의된다.

$$G_T = 10 \log_{10} \frac{\sigma_x^2}{(\prod_{k=0}^{M-1} \sigma_{y_k}^2)^{\frac{1}{M}}} \quad (15)$$

여기서 M 은 4×4 크기의 subband의 수이고, σ_x^2 는 입력의 분산이다. 또한 $\sigma_{y_k}^2$ 는 k 번째 subband 신호의 분산이다. 표 2는 DCT와 ICT에서의 부호화 이득을 비교한다.

표 2. DCT와 ICT의 부호화 이득의 비교

크기	4×4	8×8	16×16
G_{DCT}	7.0963	8.2857	8.8817
G_{ICT}	7.0963	8.2401	7.7645

간략화된 변환 계수의 값 때문에 16×16 ICT의 부호화 이득은 감소됨을 알 수 있다. 하지만 4×4 블록이나 8×8 블록에 대한 ICT는 좋은 특성을 보인다.

부호화 환경내의 변환의 특성을 평가하기 위하여 DCT를 사용한 H.263 부호화기와 ICT를 사용한 H.263 부호화기를 비교한다.[5] 여기서는 QCIF 형태의 영상 신호에서 Y신호만을 고려한다. 두 경우에서 움직임 추정 은 움직임 벡터가 탐색영역의 원점에 대부분이 존재한다는 사실에 근거한 알고리즘을 사용한다. 이 방법은 3SS나 4SS방식에 비하여 움직임 추정을 위해 필요한 연산의 수는 줄어들면서 거의 동일한 성능을 보이는 움직임 추정 방식이다. 이 방식의 움직임 추정의 예를 그림 5에 보였다.

움직임 추정에 사용한 방법이 같기 때문에 DCT와 ICT를 직접적으로 비교한다. 그림 6에서 보인 것처럼 ICT의 특성은 DCT를 사용하였을 경우와 거의 동일하다.

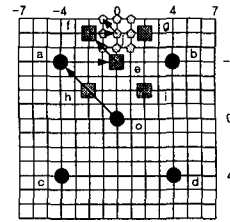
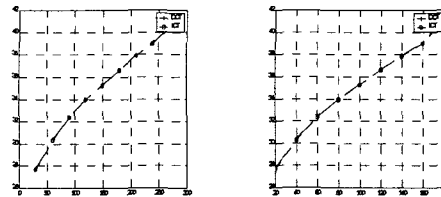


그림 5. 움직임 추정의 예



(a)FOREMAN (b)SUSIE
그림 6. DCT와 ICT의 H.263에서의 특성비교

V. 결론

본 논문에서는 8×8 DCT에서 발생하는 블록화 현상을 줄이고 자세한 움직임에 발생하는 오차를 줄이기 위하여 4×4 블록을 기본으로 차 신호에 대한 압축을 하였다. 또한 블록수의 증가에 따른 계산량의 감소를 위하여 INTRA 프레임의 경우 이웃하는 블록의 DC 계수와 화소값을 이용하여 현재 블록의 DCT 계수를 예측하고 간략화된 정수만을 사용하는 ICT를 사용하였다. 이는 8×8 DCT를 이용하여 부호화를 수행한 결과에 비교하여 영상의 화질을 저하시키지 않으면서 PSNR면에서도 거의 동일한 결과를 얻을 수 있었다.

참고 문헌

- [1] R. Srinivasan and K. R. Rao, "An approximation to the discrete cosine transform for $n=16$," Signal Processing5, pp.81~85, 1983
- [2] W. K. Cham and F. S. Wu, "On compatibility of order-8 integer cosine transforms and the discrete cosine transform," in IEEE Region 10 Conf. on Computer and Communication System, pp.447~449, Sept. 1990
- [3] W. K. Cham, "Development of integer cosine transforms by the principle of dyadic symmetry," Proc. IEEE136, pp.276~282, Aug. 1989
- [4] 한국통신학회, 대한통신학회지 "부호화기술," Vol. 15, no. 12, 1998
- [5] 허도근, 심우성, 이용식, 정상연, "H.263 성능 개선을 위한 DWT양자화기 설계," 대한전자공학회 하계발표 Vol. 22, No. 1