

3D-DCT 프레임간 확률 엔트로피에 의한 동영상 압축

박동춘, 최재윤, 서경호, 김태호
 경남대학교 전자공학과

E-mail : mir92@netian.com, hyo@hanma.kyungnam.ac.kr

Moving Picture Compression for the Probability Entropy in 3D-DCT Frames.

DongChun Park, JaeYoon Choi, KyungHo Seo, TaeHy0 Kim
 Dept. of Electronics Eng., Kyungnam Univ.

E-mail : mir92@netian.com, hyo@hanma.kyungnam.ac.kr

ABSTRACT

In this paper, we proposes Moving Picture Compression reducing bit numbers within the block by probability in 3D-DCT considered of temporal axis. From Delta Modification theory for frequency components, bit numbers within the block(4×4) were properly compressed by using the Probability Entropy.

At the result for the proposed algorithm, we could find that the values of PSNR for resolution and the compression rate were better than existing methods.

I. 서론

디지털 신호처리 분야에서의 눈부신 발전으로 디지털 영상 통신의 응용분야가 급속하게 확산되어 멀티미디어 분야와 동영상처리분야의 기술 발전이 더욱 가속화되어 가고 있다. 여전히 영상 신호를 다루는데는 많은 정보를 처리해야하므로 디지털 통신의 효율적인 영상 전송이나 저장을 위해서는 많은 제약 조건이 따르게 된다.

최근 정보통신과 멀티미디어, 인터넷이 통합되어가는 추세에 있고, 이에 따라서 영상 통신분야도 규격화 및 통합화에 접근하고 있다.[1]

동영상 처리에 가장 근접한 부호화, 복호화 표준으로는 H.261, H.263, MPEG 등이 있으며, CCITT 연구 그룹 XV 로부터 화상 전화용 동화상 압축 기법의 권고안인 H.261이 발표되었고, ISO/IEC JTC1/SC2/WG8에 의해 설립된 MPEG (Moving Picture Experts Group)에 의해 MPEG 동영상 압축 및 복원 표준안이 제정되어 CD (Compact Disk), DAT(Digital Audio Tape), HDTV 등의 응용 분야에 표준이 되고 있다.[2]

영상 압축의 단간이 되는 DCT, 특히 2D-DCT는 이미 지 압축과 Video frame 들의 인트라프레임(intra-frame) 압축에 적용할 수 있지만 시간에 대해서는 능동적이지 못하다. 반면 3D-DCT는 여기에 시간적인 개념이 추가 되어 인트라프레임(inter-frame) 화상 데이터 압축이 가능하다. 3D-DCT는 비디오 시퀀스에서 연속적인 프레임사

이에서 높은 상관도를 나타낸다.[3]

본 논문에서 제안하는 방법은 시간축에 능동적이지 못한 2D DCT을 이용한 압축에 비해 3D DCT의 특성을 이용하므로써, 즉 공간뿐만 아니라 시간축 상의 성분에 대해 저·고주파 성분으로 분리되며, 이에 대한 블록내의 확률적 엔트로피를 구하여 적용적 비트 할당으로 인해 큰 화질의 손상없이 높은 압축률을 나타낼 수 있다.

II. 본론

2.1 DCT(Discrete Cosine Transform)의 기초이론

직교변환 중 DCT와 KLT(Karhunen-Loeve Transform)는 저주파 영역에의 화상정보 집중도가 우수하다. 그 중에서도 통계적 최적 변환으로 알려진 KLT는 원리적으로 압축효율이 가장 높은 변환이지만 계산이 복잡하기 때문에 동화상과 같이 실시간 처리가 필요한 부호화 장치로는 실현이 곤란하다.

그러므로 현재 영상 부호화 방법에서 가장 많이 사용되는 DCT는 KLT에 가장 근접한 성능을 가지고 있으며 정보가 한곳으로 집중되는 효과가 크기 때문에 영상 및 음성 정보의 부호화에 많이 사용되고 있다.[4]

일반적으로 픽셀 $N_R \times N_C$ 블록(block)의 2D-DCT와 IDCT는 다음과 같이 정의되었다.[5]

Forward 2D-DCT :

$$S(y, u) = a_2 D(v, u) \sum_{y=0}^{N_C-1} \sum_{x=0}^{N_R-1} s(y, x) \cdot \cos\left(\frac{(2x+1)u\pi}{2N_C}\right) \cos\left(\frac{(2y+1)v\pi}{2N_R}\right) \quad (1)$$

Inverse 2D-DCT:

$$s(y, x) = \sum_{v=0}^{N_C-1} \sum_{u=0}^{N_R-1} a_2 D(v, u) s(y, x) S(y, u) \cdot \cos\left(\frac{(2x+1)u\pi}{2N_C}\right) \cos\left(\frac{(2y+1)v\pi}{2N_R}\right)$$

$$----- (2)$$

여기서

$$a_2 D(v, u) = \sqrt{\frac{2}{N_R} \frac{2}{N_C}} C(v)C(u)$$

그리고

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} & k=0 \\ 1 & \text{otherwise} \end{cases}$$

$s(y, x)$ 는 2차원 $N_R \times N_C$ 샘플 이미지 블록 안에서의 픽셀(pixel) 값이고, $S(v, u)$ 는 $N_R \times N_C$ DCT 블록에서의 DCT 계수들이다.

2.2 3D-DCT

만약 2차원공간에서 하나의 시간영역 변수가 더해진다면, $N_R \times N_C$ 블록은 $N_F \times N_R \times N_C$ 로 표현되는 육면체(cube)로 확장할 수 있다.

이때 3D-DCT 영역에서 육면체를 형성하는 $N_R \times N_C$ 블록의 연속적인 프레임은 N_F 이다.

일반적인 3D-DCT와 IDCT의 표현식은 다음과 같이 정의된다.

Forward 3D-DCT:

$$S(w, v, u) = a_3 D(w, v, u) \sum_{z=0}^{N_F-1} \sum_{y=0}^{N_R-1} \sum_{x=0}^{N_C-1} s(z, y, x) \cos(t_1) \cos(t_2) \cos(t_3) ----- (3)$$

Inverse 3D-DCT:

$$s(z, y, x) = \sum_{w=0}^{N_F-1} \sum_{v=0}^{N_R-1} \sum_{u=0}^{N_C-1} a_{3D}(w, v, u) \cdot S(w, v, u) \cos(t_1) \cos(t_2) \cos(t_3) ----- (4)$$

여기서

$$t_1 = \frac{(2x+1)u\pi}{2N_C}, t_2 = \frac{(2y+1)v\pi}{2N_R}, t_3 = \frac{(2z+1)w\pi}{2N_F}$$

$$a_3 D(w, v, u) = \sqrt{\frac{2}{N_F} \frac{2}{N_R} \frac{2}{N_C}} C(w)C(v)C(u)$$

그리고

$$C(k) = \begin{cases} \frac{1}{\sqrt{2}} & k=0 \\ 1 & \text{otherwise} \end{cases}$$

$s(z, y, x)$: 3D 샘플 픽셀값
(N_F 이미지 프레임 ($0 \leq z \leq N_F$))

$S(w, v, u)$: 3D DCT Coefficient
($N_F \times N_R \times N_C$ DCT 육면체)

w, v, u : 시간과 수평과 수직 주파수 변환영역에 대한 인덱스(Index)

$N_F \times N_R \times N_C$ DCT 육면체는 N_F 이미지 프레임들의 각각의 정보를 포함한다.

특히 DCT내의 육면체에서 frame0는 많은 정보를 포함하고 있는 것을 쉽게 알 수 있다. 한편 frame1에서 frame7까지의 프레임들의 대부분의 정보는 움직임 영역에 포함된다. 그러므로 frame0은 DC성분으로 구성된 프레임으로 생각될 수 있으며, 이미지 프레임들의 각각에 대한 중요한 정보를 포함하고 있음을 보여준다. 나머지 프레임들은 움직이는 고주파 성분의 정보를 포함하고 있으므로 원 영상의 시퀀스에서의 움직임과 일치한다.

그러므로 본 논문에서 연구하고자 하는 것은 3D-DCT 후 많은 정보를 포함하고 있는 frame0의 계수들을 각 프레임과의 상관관계를 이용함으로써 DCT 영역의 블록당 픽셀값들의 엔트로피를 계산하여 엔트로피의 확률에 의한 블록내 픽셀수를 적응적으로 고려하여 픽셀당 비트수를 계산함으로써 전송하는 데이터의 압축률과 복원 영상의 PSNR값을 평가하고자 한다.

DCT 영역에서 coding을 하기 위한 데이터를 보다 쉽게 분석하기 위하여 본 논문에서 사용된 이론으로 delta modification과 분리성(Separability)이 있다.

2.2.1 Delta Modification

DCT 영역에서 적은 움직임을 갖는 영역의 계수 수정은 분명하게 구별할 수 있다. (N_F 프레임의 영상 신호를 포함하는 범위내에서) $S_\delta(v, u)$ 는 zero 주위에서 보다 강력하게 그룹(Group)이 형성되므로 이것으로 인한 프레임의 감소를 가져온다.

그러므로 3D-DCT를 수행하면 상당히 큰 계수값들을 갖게 되는데, 이러한 계수값들에 델타 수정(delta modification)을 통해 3D-DCT 계수를 줄일 수 있다. 델타 수정의 수학적 표현을 나타내면,

$$S(0, v, u) = \frac{1}{\sqrt{N_F}} a_{2D}(v, u) \sum_{y=0}^{N_R-1} \sum_{x=0}^{N_C-1} s(z, y, x) \cos(t_1) \cos(t_2) ----- (5)$$

여기서 $s(x, y, z)$ 는 다음식으로 표현된다.

$$s(z, y, x) = s(0, y, x) + \delta(z, y, x) ----- (6)$$

따라서 식(5)와 식(6)을 정리하면,

$$S(0, v, u) = \sqrt{N_F} S_{\text{Frame0}}(v, u) + S_\delta(v, u) ----- (7)$$

여기서 식(7)의 우변 첫째항은 3D-DCT의 frame0를 나타내고, 두 번째항은 다음과 같이 표현된다.

$$S_\delta(v, u) = \frac{1}{\sqrt{N_F}} a_{2D}(v, u) \sum_{y=0}^{N_R-1} \sum_{x=0}^{N_C-1} \delta(z, y, x) \cos(t_1) \cos(t_2)$$

----- (8)

델타 수정에 대한 블록도를 그림 1과 그림 2에 나타내었다.

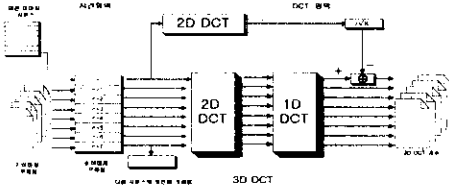


그림 1. 엔코더 : Delta Modification(델타 수정)

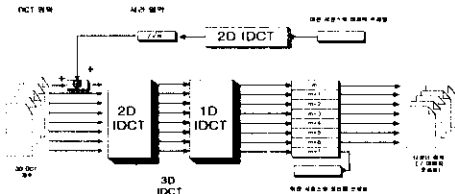


그림 2. 디코더 : Delta Modification(델타 수정)

2.3 Separability

3D-DCT(IDCT)을 수행할때 가장 간단한 방법은 위의 수식과 같은 이론적인 방정식을 충실히 따를 수 있다. (식3, 식4) 8×8×8 육면체에서 이것은 계수당 512개의 곱셈과 511의 덧셈 연산을 한다.[1]

그러나, DCT은 separable 변환이다. 이것은 다차원 (multi-dimension)의 DCT을 일차원 DCT의 시리즈로서 수행할 수 있다는 것을 암시한다. 일차원이거나 이차원에서 DCT나 IDCT을 수행하기 위해 수식적인 고속알고리즘이 존재한다. 따라서 N_F 프레임의 각각의 행과 열에 대하여 고속 알고리즘을 수행함으로써 2D-DCT을 할 수 있다.

이것을 다시 시간축을 따라 고속 1D-DCT을 수행 함으로써 3D-DCT을 수행 할 수 있게된다. 이것을 그림 3에 나타내었다.

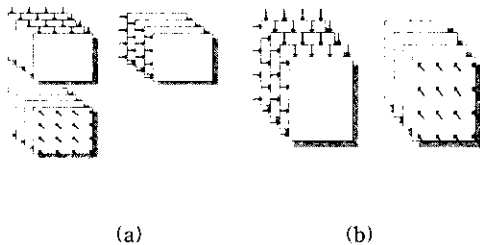


그림 3. a) 3D-DCT을 1D-DCT로 분리 계산
b) 3D-DCT을 2D-DCT한후 1D-DCT로 분리계산

(4X4X4인 경우)

2.4 제안한 확률적 엔트로피

3D-DCT한 각 블록의 값들을 아래 Table 2.1과 같이 양자화 테이블을 적용시켜서 식 9를 이용해서 양자화한다.

Table 2.1

$$Q(w, v, u) = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

$$F_q(w, v, u) = \text{round}\left(\frac{F(w, v, u)}{Q(w, v, u)}\right) \quad \text{----- (9)}$$

각 프레임에 대해 양자화된 값들의 전체 프레임에서 픽셀에 대한 엔트로피를 구해서 차지하는 비율이 높은 블록에 대해 많은 비트를 할당하고 상대적으로 차지하는 비율이 적은 블록에 대해서는 적은 비트를 할당함으로써 압축률과 해상도(resolution)를 높이고자 하였다.

각 프레임에 대한 엔트로피의 확률은 식 10을 이용해서 구할 수 있고 Table 2.2를 통해서 비트 할당을 한다. 각 프레임의 엔트로피 확률을 구하는 식은

$$P_q(w) = \frac{\sum_{v=0}^{N_F-1} \sum_{u=0}^{N_F-1} |F_q(w, v, u)|}{\sum_{w=0}^{N_F-1} \sum_{v=0}^{N_F-1} \sum_{u=0}^{N_F-1} |F_q(w, v, u)|} \quad \text{----- (10)}$$

와 같이 나타낼 수 있으며, 여기서 $w = 0, 1, 2, \dots, 7$ 이다.

Table 2.2

```

if  $P_{q_w} > 0.3$  then do
    ZigZag21( $F_q(w, v, u)$ )
else if  $P_{q_w} < 0.3$  and  $P_{q_w} > 0.2$  then do
    ZigZag15( $F_q(w, v, u)$ )
else if  $P_{q_w} < 0.2$  and  $P_{q_w} > 0.15$  then do
    ZigZag10( $F_q(w, v, u)$ )
else if  $P_{q_w} < 0.15$  and  $P_{q_w} > 0.1$  then do
    ZigZag6( $F_q(w, v, u)$ )
else if  $P_{q_w} < 0.1$  and  $P_{q_w} > 0.05$  then do
    ZigZag3( $F_q(w, v, u)$ )
else if  $P_{q_w} < 0.05$  then do
    ZigZag1( $F_q(w, v, u)$ )
end ALL
    
```

$F_g(w, v, u)$ 은 양자화 테이블을 이용한 양자화 된 값을 나타내고, P_{q_n} 은 전체 8 프레임 중 각 프레임이 차지하는 확률을 나타내며 $ZigZag_number$ 는 지그재그 스캔을 해서 데이터 비트를 취하는 개수를 의미한다.

III. 실험 결과 및 분석

실험 영상은 Miss America 와 Sales Man을 각각 40 프레임을 사용하였으며, 3D-DCT을 하기 위하여 8 프레임 을 한 그룹으로 사용하였으며, 각 프레임의 이미지는 8×8 블럭으로 나누어 정육면체($8 \times 8 \times 8$ cube)에 대해 양자화 한 후 지그재그 스캔, run length 부호화 하였다. 그림 4과 그림 5는 각 Miss America와 SalesMan의 이미지를 이용해서 3D-DCT 했을 때와 수정 3D-DCT을 했을 때의 PSNR을 비교한 것이다.

Miss America에서 평균 PSNR은 각각 35.87dB과 36.23dB로 수정된 PSNR값이 0.36dB정도 개선됨을 알

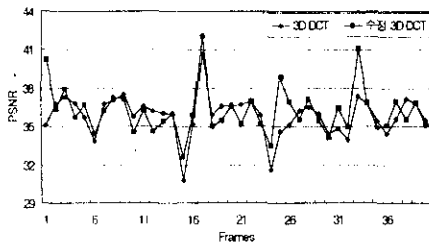


그림 4. PSNR 비교 데이터(Miss America)

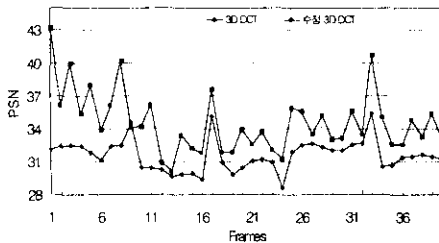


그림 5. PSNR 비교 데이터(Salesman)

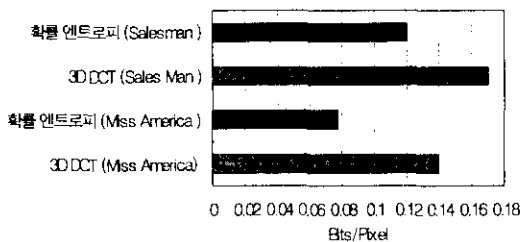


그림 6. 압축률 비교

수 있었으며, 움직임이 보다 많은 SalesMan에서는 각각 평균 PSNR이 31.54dB와 34.57dB로 움직임이 많은 이미지에 대해서는 약 3dB정도로 비교적 많이 개선됨을 알 수 있었다.

그림 5에서 보는 것과 같이 각각의 프레임으로 볼 때 개선되었음을 알 수 있다. 또한 그림 6에서는 H.261, H.263과 3D-DCT와 제안 알고리즘의 차이를 분석한 압축률로 비교해 보면 8×8 블럭당 데이터 압축률은 기존의 알고리즘보다 제안 알고리즘이 비교적 좋음을 보여 준다.

이 결과를 분석해보면, 제안된 알고리즘은 계수들이 '0'의 값에 비교적 많이 집중됨으로서 해상도와 압축률에서 향상된 결과 특히 움직임이 많은 동영상에 대하여 탁월한 결과를 얻을 수 있음을 알 수 있다.

IV 결론

본 논문은 3D-DCT에서 양자화한 각 프레임들에 대해 엔트로피 확률을 이용해서 적응적으로 비트를 할당함으로써 보다 높은 압축률과 PSNR을 얻을 수 있는 알고리즘을 제안하였다. 3D-DCT을 통해 얻은 계수들 중 frame0 부분의 계수에 많은 영상 정보가 포함되어 있고 이 계수를 줄임으로서 압축률을 높일 수 있었다. 이에 수정 알고리즘을 통해서 첫 번째 프레임(frame0)의 영상 정보를 현저히 줄일 수 있었고 또한 실험을 통해 얻은 결과 보다 움직임이 많은 영상에서 보다 좋은 압축률과 PSNR이 나타나는 것을 알 수 있었다. 또한 각 프레임이 8개의 프레임 중에서 블럭당 픽셀의 엔트로피가 차지하는 확률을 가지고 적응적으로 픽셀에 대한 비트수를 계산함으로써 Miss America는 0.36dB SalesMan은 3dB 보다 효율적으로 압축 할 수 있었다.

참고 문헌

- [1] 구준모, "동영상 압축표준, MPEG", 마이크로 소프트웨어 저널 164호, 1997.
- [2] Joan L., William B., Chad E., Didier J., " MPEG video : compression standard", Chapman & Hall, 1997.
- [3] M. Servais, G. De Jager, " Video Compression using the Three Dimensional Discrete Cosine Transform(3D-DCT)", IEEE.
- [4] 한국과학기술원, " 영상감축기술 개발연구", 1988.
- [5] K. Rao and P. Yip, "Discrete Cosine Transform: Algorithm, Advantages, Applications." Academic Press, Inc., 1990.