

# H.263 비디오 코덱의 변환 부호화 성능개선을 위한 최적화된 정수형 여현 변환

정희원 신경철\*, 이광형\*\*

## Optimized Integer Cosine Transform for Transformation Coding Performance Enhancement of H.263 Video Codec

Kyung-Cheol Shin\* , Kwang-Hyung Lee\*\* *Regular Members*

### 요 약

H.263은 64Kbps 이하 전송로에서 영상전화, 영상회의 등의 서비스를 가능하게 하는 ITU-T의 국제 표준권고안이다. H.263의 성능 평가에 사용된 TMN5는 변환부호화 방법으로 DCT를 권고하고 있다. 그러나 DCT 부호화는 계수가 실수이기 때문에 절삭으로 인한 오차가 발생하며, 부동소숫점 연산에 따른 하드웨어 구현의 복잡성과 연산 시간이 많이 걸리는 단점이 있다. 따라서 본 논문에서는 DCT의 변환 성능을 유지하면서 실수형 연산 문제를 보완한 OICT를 사용하여 성능평가를 비교하였다. OICT의 성능평가를 위해 평균 PSNR과 주관적인 화질비교를 하여 DCT와 비교한 결과 성능개선을 나타내었다. PSNR 측면에서 휘도 신호 Y는 OICT 변환방식이 약 0.02~0.21dB 정도의 이득을 보였으며, 색차신호 Cb, Cr은 0~0.5dB정도 성능향상을 보이는 것으로 나타났다.

### ABSTRACT

H.263 is an recommendation draft of ITU-T standard that can makes the service such as video phone, video conference in the transmission line less than 64Kbps. TMN5 used for the performance estimation of H.263 has fundamentally used DCT in transform coding method. But the DCT coding occurred the approximation-error by the real coefficients operation. And this have defects by the complexity of hardware implementation and the many times by calculation of floating point. So we compared the performance evaluation by using OICT resolved a calculation problem of a real number keeping in the transformation performance of DCT. we obtained improvement resulting from the comparison of subjective image quality and PSNR for performance evaluation of OICT compared to DCT. The results obtained benefit of about 0.02-0.21dB in luminance signal(Y) and performance improvement about 0-0.5dB in chrominance signal(Cb, Cr).

### I. 서 론

최근 문자, 음성 및 영상과 같은 다양한 정보들은 디지털 신호처리와 컴퓨터 등의 기술발전으로 인해 멀티미디어 형태로 저장 또는 전송되어진다. 이러한 정보의 표현이나 전송을 위한 국제 표준 규격들이

권고되고 있고, 이 중에서 64Kbps 이하의 전송속도를 갖는 저전송률 동영상 부호화에 관한 표준안으로 ITU-T/LBC 그룹에서 H.263 권고안이 제정되었다<sup>[1-3]</sup>. 동영상의 고압축 부호화를 위해서는 신호 성분, 통계적, 시간적, 공간적 중복성을 효과적으로 제거해야 한다. 이중 화면내에 존재하는 공간적 중복성은 변환 부호화 방법에 의해 제거할 수 있다. 변

\* 경남정보대학 전자정보과(keshim@nengjung.kit.ac.kr)

\*\* 숭실대학교 정보통신전자공학부(kylee@dsp.soongsil.ac.kr)

논문번호: 99039-0930

접수일자: 1999년 9월 30일

환부호화 과정은 H.263에서 이산 여현 변환 (Discrete Cosine Transform : DCT)을 권고하나 본 논문에서는 DCT의 부동소숫점 연산으로 인한 하드웨어 구현의 복잡성과 연산시간이 많이 걸리는 단점을 개선시키기 위해 적은 비트 구성, 하드웨어 구현의 용이성, 연산속도 개선 등의 장점이 있는 최적화된 정수형 여현변환(Optimized Integer Cosine Transform : OICT)을 사용하였다. 본 논문에서는 TMN5(Test Model Near 5)의 구조를 그대로 유지하면서 양자화기의 변환부호화 방법으로 DCT와 OICT에 대한 변환의 성능을 비교하였다. 양자화된 계수 값의 가변길이 부호화(VLC : Variable Length Coding)는 2차원 가변길이 부호화표를 이용하여 8×8 블록의 계수들을 지그재그(Zig-Zag) 주사한다. 그리고, 비트열 발생과정은 버퍼 제어를 통하여 부호화시 발생하는 비트율을 적절히 조절하여 넘침(Overflow)을 방지하는 TMN5의 버퍼 제어방법을 사용하였다<sup>[4]</sup>. 모의 실험에서는 ITU-LBC 그룹에서 권고되는 QCIF 형태의 Carphone 영상 100프레임을 실험영상으로 채택하였다. 논문의 구성은 제 2장에서는 DCT 기반 H.263의 테스트 모델인 TMN5에 관하여 설명하며 제 3장에서는 제안된 최적화된 정수형 여현 변환 방법을 설명한다. 제 4장에서는 DCT와 OICT 방법을 모의 실험을 통해 비교·분석하며 제 5장에서 결론을 맺는다.

## II. DCT 기반 H.263 : TMN5

### 2-1. TMN5의 전체 구성

TMN5는 H.263 권고안을 만족하는 영상전화 또는 영상회의의 시스템으로 변환부호화는 8×8 블록에 대하여 DCT를 수행한다. TMN5의 부호화기를 그림 1에 나타낸다<sup>[1][4]</sup>.

DCT를 기반으로 하고 있는 TMN5의 부호화 과정을 살펴보면 먼저 현재 프레임의 입력 영상이 ME1(Motion Estimator 1) 블록에서 프레임 메모리 M1에 저장되어 있는 이전 프레임의 원 영상과 MB 단위로 블록 정합방식을 이용하여 정수 단위의 움직임 벡터(Motion Vector : MV)를 추정한다. 만약 현재 MB가 INTRA 모드이면 ME/MC없이 바로 DCT하고, INTER 모드이면 프레임 메모리 M2에 저장되어 있는 이전에 재생된 영상을 기준으로 ME2 블록에서 반화소 단위의 정확한 MV를 추정한 후 PRED에서 MV에 해당하는 블록을 가져와서 현재 블록과의 차신호를 구하여 이를 DCT한다.

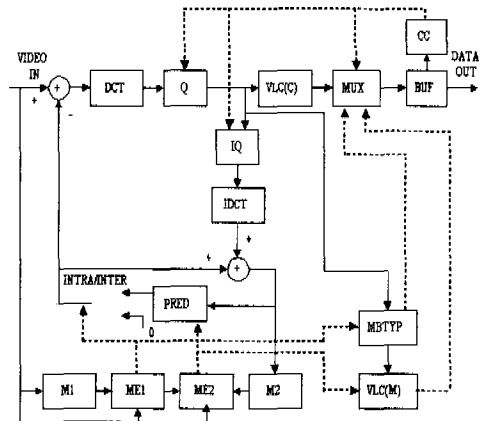


그림 1. DCT 기반 H.263 : TMN5

DCT 계수들은 양자화하여 가변장 부호화(VLC(C))를 하고 MB의 형태(MBTYP)와 MV도 가변장 부호화(VLC(M))를 하여 비트열을 발생시킨다. 변환 부호화, 양자화되어 비트열로 전송된 화질의 객관적 성능을 비교하기 위해 PSNR을 이용한다. 크기 N×N인 전송할 프레임과 전송하여 재구성된 프레임의 화소 X(i, j)와 Y(i, j)에 대하여 평균 자승오차 MSE는 식(1)로, PSNR은 식(2)와 같이 정의된다.

$$MSE = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [X(i, j) - Y(i, j)]^2 \quad (1)$$

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \quad (2)$$

### 2-2. DCT 변환 부호화

변환 부호화에서 1차원 입력 신호는 공간이나 시간영역으로부터 변환영역으로 매핑된다. 이 경우 에너지는 변환영역에서도 보존되며 원 신호는 역변환에 의해 복원된다. 또한 1차원적인 변환을 연속으로 처리함으로써 다차원 변환을 수행하게 된다. DCT 수행과정은 입력된 영상 데이터를 N×M의 블록으로 나누고 블록 내의 각 픽셀들에 대하여 식 (3)과 같이 DCT 연산을 행하며 그 역변환 과정은 식 (4)와 같다<sup>[5][7]</sup>.

$$X_{u,v}^c = \frac{4}{NM} C_u C_v \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} x_{n,m} \cos \left[ \frac{(2n+1)u\pi}{2N} \right] \cos \left[ \frac{(2m+1)v\pi}{2M} \right] \quad (3)$$

여기서,  $u=0,1,\dots,N-1$ ,  $v=0,1,\dots,M-1$

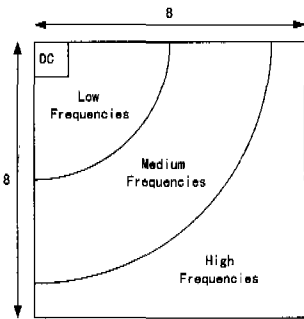
$$c_l = \begin{cases} 1/\sqrt{2} & l=0 \\ 1 & l \neq 0 \end{cases}$$

$$x_{n,m} = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} c_u c_v X_{u,v}^c \quad (4)$$

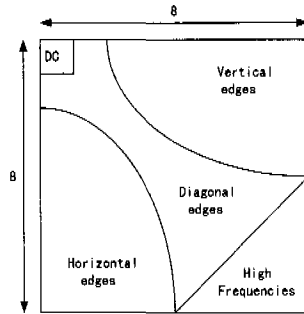
$$\cos\left[\frac{(2n+1)u\pi}{2N}\right] \cos\left[\frac{(2m+1)v\pi}{2M}\right]$$

여기서,  $n=0,1,\dots,N-1$ ,  $m=0,1,\dots,M-1$

DCT는 원 영상에 기저함수 코사인(cosine)을 곱하여 변환한다. 최저 주파수를 나타내는  $u=0$ ,  $v=0$ 인 경우를 DC 계수라 하고, 나머지 계수를 AC 계수라 하며  $u$ 와  $v$ 가 클수록 고주파 성분임을 나타낸다. DC 계수는 블록 내 화소의 평균값을 의미하고, AC 계수는 블록 내에 존재하는 에지 성분을 나타낸다. 그림 2는  $8 \times 8$  블록에 대한 DCT 변환영역의 주파수 분포와 각 방향에 따른 에지 성분의 분포를 보이고 있다<sup>[8]</sup>.



(a) Frequency distribution



(b) edge distribution

그림 2. DCT 영역의 주파수 및 에지 분포

식 (3)에 의하여 영상을 DCT 변환하면 변환계수들은 저주파 영역에서는 큰 값을 갖지만 고주파 영역에서는 작은 값을 갖는다. 따라서 영상의 대부분

의 에너지가 DC성분을 중심으로 하는 저주파 영역으로 모이는 DCT 계수 특성으로 인해 저주파 영역의 변환계수를 양자화하여 전송하고, 고주파 영역의 변환계수는 양자화하는 과정에서 버려지는 방법으로 영상을 압축하게 된다. 따라서 DCT는 계수가 실수이기 때문에 절삭으로 인한 오차가 발생하며, 부동소수점 연산에 따른 하드웨어 구현의 복잡성과 연산 시간이 많이 걸리는 단점이 있다<sup>[6][11]</sup>.

### III. 최적화된 정수형 여현 변환

DCT는 이산 직교변환으로써 표준 권고안의 영상 변환부호화 방법으로 많이 사용되고 있다. 그러나 DCT는 계수가 실수이기 때문에 절삭으로 인한 오차가 발생하며, 부동소수점 연산에 따른 하드웨어 구현의 복잡성과 연산 시간이 많이 걸리는 단점이 있다. 그러므로 DCT의 실시간 처리를 위해서 변환계수를 유한개의 비트로 표현하면 변환의 성능이 감소되지 않고 하드웨어 구조가 간단하면서 연산시간이 빨라진다. 본 논문에서는 DCT의 변환 성능을 유지하면서 하드웨어 구현이 간단하고 연산속도가 빠른 장점을 가지고 있는 OICT를 H.263의 테스트 모델인 TMN5에 적용하여 PSNR 및 화질을 비교한다. OICT의 방법은 DCT 행렬의 직교성과 크기순서를 갖는 특성을 이용한다. 그리고 정규화 계수가 단일 값을 갖도록 정방향 변환은 물론 역방향 변환에서도 변환 후에 정수형 연산만을 수행할 수 있도록 하기 위한 것이다. 정규화 계수로 인한 정규화 오차를 무시할 수 있는 범위에서 근사화한 식을 식 (5)에 나타낸다<sup>[10]</sup>.

$$\begin{aligned} 2^{-n-1} [t^2(0,0)]_{A_n} &\approx 2^{-n-i-1} \sum_{j=0}^i [d^2(0,j)] D_{N/2^{n-1}} \\ &\approx \sum_{j=0}^{(N/2)-1} [q^2(0,j)] Q_{N/2} \quad (5) \\ i &= 1, 2, \dots, n-2; \quad n = \log_2 N; \end{aligned}$$

$N=8$ 인 경우 OICT를 구하는 과정을 설명하면, 먼저 식 (5)는 식(6)과 같이 된다.

$$4t^2 \approx 2(d_1^2 + d_2^2) = a_1^2 + a_2^2 + a_3^2 + a_4^2 \quad (6)$$

정수 값을 갖는 OICT 행렬  $[A_8]$ 을 구하는데 있어서, 먼저 직교 조건과 크기 순서를 만족하는 정수 값  $a_1, a_2, a_3, a_4$ 를 구한 다음,

$$\sqrt{a_1^2 + a_2^2 + a_3^2 + a_4^2} / 2$$

를 구하여 반올림(round off)

했을 때의 정수 값을  $t$ 로 한다. 실제로 이러한 조건을 만족하는 정수 값의 집합은 여러 개가 존재하나 Hilbert-Schmidt norm 값이 작으면서 고속 알고리즘 구현이 가능한 정수 값을 선정하면  $t=91$ ,  $d1=119$ ,  $d2=49$ ,  $q1=125$ ,  $q2=108$ ,  $q3=72$ ,  $q4=25$ 인 7 비트로 구성된다. 그리고 이 정수 값을 이용하여 OICT 행렬  $[A_8]$ 을 구성하면 식 (7)과 같으며, 이 행렬을 정규 직교 행렬로 만드는 대각 행렬  $[K_8]$ 은 식 (8)과 같다<sup>[9][10]</sup>.

$$[A_8] = \begin{bmatrix} 91 & 91 & 91 & 91 & 91 & 91 & 91 & 91 \\ 125 & 108 & 72 & 25 & -25 & -72 & -108 & -125 \\ 119 & 49 & -49 & -119 & -119 & -49 & 49 & 119 \\ 108 & -25 & -125 & -72 & 72 & 125 & 25 & -108 \\ 91 & -91 & -91 & 91 & 91 & -91 & -91 & 91 \\ 72 & -125 & 25 & 108 & -108 & -25 & 125 & -72 \\ 49 & -119 & 119 & -49 & -49 & 119 & -119 & 49 \\ 25 & -72 & 108 & -125 & 125 & -108 & 72 & -25 \end{bmatrix} \quad (7)$$

$$[K_8] = \frac{1}{K} [I_8], k = \sqrt{66222} \quad (8)$$

OICT의 정규화 오차 정도는  $256 \times 256 \times 8$ 비트의 해상도를 갖는 임의의 실제 영상 10개를  $8 \times 8$ 의 블록으로 나눈 총 10240개의 블록에 대하여 정방향 변환을 수행한 다음, 역방향 변환을 수행하여 원 영상과 복원 영상에 대한 평균자승오차(MSE)를 측정 한 결과 MSE가 "0"에 가까웠다. 따라서 식 (7)의 OICT는 정규화 계수의 처리에서 발생하는 정규화 오차를 무시할 수 있다.

#### IV. 시뮬레이션 결과

OICT의 성능 평가를 위해 H.263의 테스트 영상인 Carphone 영상 100프레임을 사용하여 DCT와

OICT를 비교 분석한다. 전송속도 20kbps, 30kbps, 40kbps 각각에 대하여 영상의 변화에 따라 변화가 적은 1번째 프레임, 중간인 15번째 프레임, 변화가 큰 55번째 프레임을 DCT변환과 OICT 변환을 하여 양자화한 후 재생한 영상을 그림 3에 나타내었고 Y, Cb, Cr 신호에 대한 평균 PSNR 비교는 표 1과 같다.

표 1에서 휘도신호 Y는 OICT 변환방식이 약 0.02~0.21dB 정도의 이득을 보였으며, 색차신호 Cb, Cr은 변화가 큰 프레임 55에서는 약 0.06~0.17dB정도 OICT가 DCT보다 떨어지지만 나머지 프레임에서는 약 0~0.5dB 정도 성능 향상을 보이는 것으로 나타났다. 그림 3에서 프레임간 변화가 적은 때인 프레임 1번은 OICT 변환이 약간의 화질 개선을 보이지만 변화가 중간정도인 프레임 15의 재구성 영상에서는 주관적인 측면에서 확인할 수 있을 정도의 화질개선을 보인다.

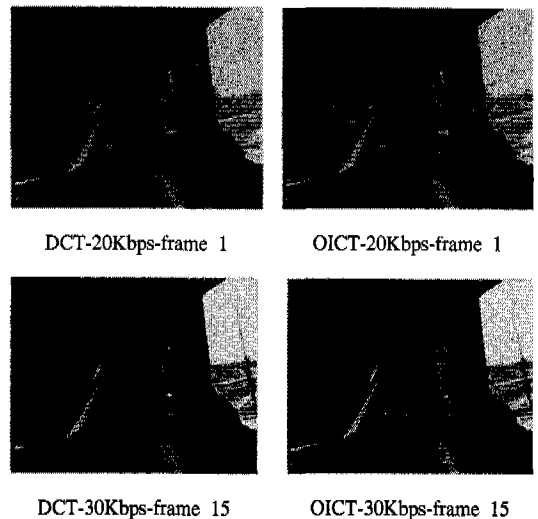
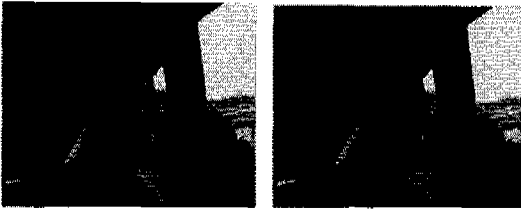


표 1. DCT와 OICT의 성능평가

부호화방법		전송속도(Kbps)					
		DCT			OICT		
구 분	PSNR (dB)	구 분					
		20	30	40	20	30	40
Frame-1	Y	32.9838	33.7021	34.3090	33.0307	33.8506	34.4418
	Cb	42.3227	42.3132	42.3917	42.3227	42.3974	42.4136
	Cr	43.0088	43.0850	43.1520	43.0088	43.1237	43.1555
Frame-15	Y	31.9487	32.5540	33.3265	31.9712	32.7660	33.3937
	Cb	41.5664	41.7191	41.8592	41.7850	41.7706	41.8681
	Cr	41.2658	41.3085	41.3973	41.3712	41.3335	41.4152
Frame-55	Y	32.6997	33.4962	34.2385	32.7410	33.6165	34.3439
	Cb	42.5293	42.5614	42.5626	42.5390	42.4961	42.5626
	Cr	40.6212	40.6358	40.7262	40.6212	40.4586	40.7387



DCT-40Kbps-frame 55      OICT-40Kbps-frame 55  
 그림 3. 각 프레임에 대한 DCT와 OICT의 복원 영상

### V. 결론

DCT대신 변환의 성능이 감소되지 않고 적은 비트 구성, 하드웨어 구현의 용이성, 연산속도가 개선되는 OICT를 사용하여 DCT를 사용한 경우와 비교 평가하였다. OICT의 성능평가를 위해 전송속도 20Kbps, 30Kbps, 40Kbps 각각에 대하여 영상 변화가 아주 적은 1번째 프레임, 중간인 15번째 프레임, 변화가 아주 큰 55번째 프레임을 양자화하여 재구성한 영상에 대하여 PSNR과 화질을 비교한 결과 PSNR 측면에서 휘도신호 Y는 OICT 변환방식이 약 0.02~0.21dB 정도의 이득을 보였으며, 색차신호 Cb, Cr은 변화가 큰 프레임 55에서는 약 0.06~0.17dB정도 OICT가 DCT보다 떨어지지만 나머지 프레임에서는 약 0~0.5dB 정도 성능 향상을 보이는 것으로 나타났다. 따라서 저전송률의 영상전화를 OICT 변환방식으로 구현할 경우 20Kbps 이하의 부호화에서도 화질이 우수함을 알 수 있었다.

### 참고 문헌

[1] ITU-T Recommendation H.263, "Video Coding for Low Bit Rate Communication," Mar. 1996.  
 [2] ITU-T Recommendation H.263 Version 2, "Video Coding for Low Bit Rate Communication," Jan. 1998.  
 [3] ITU-T Recommendation H.261, "Video Codec for Audio visual Services at P×64 Kbits," Mar. 1993.  
 [4] ITU-T SG 15, Working Party 15/1 Expert's Group on Very Low Bit rate Visual Telephony, "Video Codec Test Model TMN5," Jan. 1995.

[5] K. R. Rao, J. J. Hwang, "Techniques and Standards of Image, Video, and Audio Coding," Prentice Hall PTR, 1996.  
 [6] W. K. Cham, Y. T. Chan, "An order-16 integer cosine transform", IEEE Trans. on Signal Processing, vol. 39, pp. 1205-1208, May 1991.  
 [7] S. C. Chan and K. L. Ho, "A New Two-Dimensional Fast Cosine Transform Algorithm," IEEE Trans. Signal Process., vol. 39, pp. 481~485, Feb. 1991.  
 [8] N. I. Cho and S. U. Lee, "Fast Algorithm and Implementation of 2D Discrete Cosine Transform," IEEE Trans. Circuits Systems, vol. 38, pp. 297~305, Mar. 1991  
 [9] J. Y. Nam and K. R. Rao, "Image Coding Using a Classified DCT/VQ based on Two Channel Conjugate Vector Quantization," IEEE Trans. Circuits and System for Video Technology, vol. 1, pp. 325~336, Dec. 1991.  
 [10] O. G. Guleruz and M. T. Orchard "Optimized non-orthogonal transforms for image compression", IEEE Trans. Signal Processing, vol. SP-6, pp. 507~522, Apr 1997.  
 [11] W. K. Cham, "Development of integer cosine transforms by the principle of dyadic symmetry", IEEE Proc. Inst. Elect. Eng. Part I, pp. 276-282, Aug. 1989.

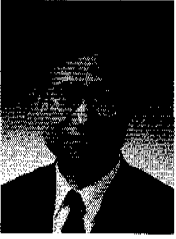
### 신경철(Kyung-Cheol Shin)



1987년 2월 : 원광대학교 전자공학(학사)  
 1990년 2월 : 송실대학교 대학원 전자공학과(석사)  
 1995년 2월 : 송실대학교 대학원 전자공학과 박사과정 수료  
 1994년~현재 : 경남정보대학 전자정보과 조교수

<주관심 분야> 멀티미디어, 영상신호처리, 영상압축

이 광 형(Kwang-Hyung Lee)



1968년 2월 : 서울대학교

전자공학과(학사)

1972년 2월 : 서울대학교

전자공학과(석사)

1987년 2월 : 중앙대학교

전자공학과(박사)

1973년~1982년 : 일본 동경대학 전자공학과 연구원

1993년~1995년 : 대한전자공학회 인공지능, 신경망,  
퍼지시스템 연구회 위원장.

1994년~현재 : 국제 전기전자학회

(IEEE Korea Council) 학술이사

1994년~1995년 : 대한 퍼지 및 지능 시스템 학회  
국제 담당이사

1994년~현재 : 대한전자공학회 평의원 이사

1982년~현재 : 숭실대학교 정보통신전자공학부 교수

<주관심 분야> 지능 신호처리, 퍼지신경망 음성인  
식 및 합성, 멀티미디어 통신처리,  
ATM 망