

시각특성과 DCT를 이용한 영상부호화에 관한 연구

正會員 南 承 珍* 正會員 崔 成 男* 正會員 全 重 滄** 正會員 朴 圭 泰*

A Study on Image Coding using the Human Visual System and DCT

Seung Jin Nam*, Sung Nam Choi*, Joong Nam Jeon**, Kyu Tae Park* *Regular Members*

要 約

본 논문에서는 인간의 시각특성을 반영한 적응 DCT 영상 부호화에 관하여 연구하였다. 인간의 시각은 특정한 중간 주파수 대역에 가장 민감하며, 저주파 대역과 고주파 대역에 대하여 상대적으로 덜 민감하다. 이러한 특성은 많은 심리시각적 실험을 통하여 MTF(Modulation transfer function)라는 수식으로 모델링되었다. DCT는 공간영역의 에너지를 주파수 영역으로 변환하므로 MTF를 효율적으로 이용할 수 있다. 또한 시각의 공간 마스크 특성에 의하여 변화가 많은 영역의 잡음은 변화가 적은 영역의 잡음에 비하여 인간의 시각에 덜 민감하다. 제안된 부호화 방법은 시각의 공간주파수 반응 특성을 양자화 행렬로 채택하였으며, 영상의 변화도에 따라 변환 블록을 차등양자화 함으로써 주관적 화질의 향상을 꾀하였다. 각 블록의 변화도를 측정하기 위한 방법으로 공간 영역에서 간단한 연산을 행하였으며, 이에 따라 변화가 적은 블록을 상대적으로 정확하게 양자화하였다. 실험결과 낮은 비트율에서 시각특성을 반영한 제안된 부호화 방법이 주관적으로 우수한 성능을 보임을 확인하였다.

ABSTRACT

In this paper, an adaptive cosine transform coding scheme which incorporate human visual properties into the coding scheme is investigated. Human vision is relatively sensitive to mid-frequency band, and insensitive to very low and very high frequency band. These property was mathematically modelled with MTF(Modulation Transfer Function) through many psychovisual experiment. DCT transforms energy in spatial domain into frequency domain, so can exploit the MTF very efficiently. Another well-known visual characteristics is spatial masking effect that visibility of noise is less in regions of high activity than in regions of low activity. Proposed coding scheme employs quantization matrix which represent the properties of these spatial frequency response of human vision, and adaptively quality of an image. To compute the activity index of an image block, simple operation is performed in spatial domain, and according to activity index, block of low activity re-

*延世大學校 電子工學科

Dept. of Electronics, Yonsei University

**忠北大學校 電子計算學科

Dept. of Computer Science, Choongbook University

論文番號 : 92-34 (接受1991. 7. 9)

gion is more exactly quantized relatively than that of high activity region. Results showed that, at low bit rate, the subjective quality of the reconstructed images by proposed coding scheme is acceptable than that of coding scheme without HVS properties.

I. 서 론

영상에 적용되는 압축기법은 크게 예측 부호화(Predictive coding) 방식과 변환 부호화(Transform coding) 방식의 2가지로 나눌 수 있다. 예측 부호화 방식은 영상 신호의 공간 및 시간상의 상관 관계를 이용한 것이며, 변환 부호화 방식은 적외 좌표를 이용하여 공간 영역에서의 에너지 분포를 부호화에 보다 적합한 형태로 변환하여 영상의 중복되는 성분을 제거함으로써 부호화할 데이터의 양을 줄이는 데 목적이 있다. 이러한 변환 부호화 방식에는 Fourier 변환, KLT(Karhunen loeve transform), WHT(Walsh-Hadamard transform), DCT(Discr cosine transform) 등이 있으며 KLT가 성능면에서 최적이라 알려져 있다. 그러나, KLT는 Kernel이 영상에 따라 바뀌므로 커널(Kernel) 산출을 위하여 많은 계산이 필요하며, 전송시 각 커널을 부가정보로 보내주어야 하므로 실제적인 적용이 어려운 단점이 있다. 따라서, KLT 다음으로 최적이라 알려져 있는 DCT가 영상의 변환 부호화에 널리 사용되고 있다^[1].

변환 부호화 방식에서는 보통 전체 영상을 8×8 또는 16×16 의 블록으로 분할하여 변환을 행하는데 영상의 특성상, 주변 화소들간의 상관관계로 인하여 16×16 블록이 최적이라 알려져 있다^[2]. 변환 부호화 방식에서는 데이터 압축을 위하여, 변환후 반드시 각 계수에 대한 양자화 과정이 뒤 따르게 된다. 이 때, 각 계수에 대한 양자화 오차는 변환된 블록 전체에 영향을 미치게 된다. 따라서 이러한 양자화 오차를 줄이는 것은 변환 부호화의 주된 연구과제의 하나이다. 최근까지는 동일한 비트율에서 SNR(Signal to noise ratio)을 최대화 하는 관점에서 연구가 진행되어 왔다. 즉, 각 계수의 통계적 성질에 따른 비트할당표를 작성하여, 블록의 각 계수에 동일한 비트를 할당하는 고정 비트 할당 부호화 방식에서는 에너지가 큰 계수에 상대적으로 많은 비트를 할당한다^[3]. 그러나, 이러한 방법은 전체적인 영상의 통계적 성질만 반영하고, 각 블록의 성질을 고려하지 않아, 많은 블록에 대하여 최적의 비트 할당이 이루어지지 않는 단

점이 있다. W.H.Chen과 C.H.Smith[7]는 각 블록의 AC 에너지에 따라 블록을 4개의 그룹으로 분류하여 변화도가 적은 그룹에는 적은 비트를 할당하고, 상대적으로 변화도가 큰 그룹에는 많은 비트를 할당하는 부호화 방법을 제안하였다. 그러나, 비트 할당표를 이용하는 위의 방식은 각 영상의 부호화때마다, 비트 할당과 그룹 분류를 위하여 많은 계산이 필요한 단점이 있다.

1984년 W.H.Chen과 W.K.Pratt^[8]는 비율버퍼를 이용하여 영상에 따라 임계치의 변환을 적응적으로 처리하는 방식을 사용하였다. 이 방식은 모든 블록에 대하여, 특정한 임계치가 넘는 계수를 모두 부호화하므로, 각 블록에 대하여 적응적으로 대처할 수 있는 장점이 있으며, 지그재그 스캔을 통한 제로 Run-Length 허프만 부호화기를 사용하여, 비트율과 SNR 면에서도 기존의 고정 비트 할당 부호화기보다 좋은 성능을 보여준다.

이와는 다른 접근으로서 1974년 D.J.Sakrison과 J.L.Mannos^[9]는 인간의 시각을 통한 실험적 결과를 토대로 하여, 영상의 주관적 평가를 수치로 나타내기 위한 연구를 행하였으며, 실험결과 시각 체계의 특성을 나타내는 MTF(Modulation transfer function)를 발표하였다. MTF는 공간 주파수에 대한 인간 시각의 상대적인 반응도를 나타내는 것으로서, 사람의 시각은 8cpd(cycles/degree) 부근에서 최대 반응감도를 나타내며, 고주파대역으로 갈수록 반응 감도가 떨어짐이 확인되었다.

1990년 CCITT/ISO의 산하단체인 JPEG(Joint photographic experts group)에서는 정지 영상 부호화에 대한 삼정적 표준안을 발표하였다^[10]. 여기에서는 Chen 방식에서와 같이 지그재그 스캔을 통한 제로 Run-Length 허프만 부호화기를 사용하였으며 8×8 DCT를 채택하였다. 또한 JPEG에서는 Chen 방식과는 달리 HVS(Human visual system)를 고려한 양자화 행렬을 발표하였는데, 이는 심리시각적인 실험을 통하여 얻어졌으며, 인간의 눈에 민감한 성분을 갖는 계수를 상대적으로 더 강조한 것이다

본 논문에서는 JPEG과 흐름을 같이 하면서 인간

의 시각에 좀더 적합한 영상부호화에 대하여 연구하였다. 이를 위하여 영상의 변환부호화에서 최적이라 알려져 있는 16×16 블록에 대한 양자화행렬을 구하였으며, 또한 블록의 변화도에 따라 각 블록을 시각 특성에 맞게 적응적으로 양자화하는 방법을 제안하였다.

II. DCT의 성질

DCT는 에너지 밀집도에서 최적변환방식인 KLT와 거의 동일한 성능을 나타내므로 영상의 변환 부호화에 널리 사용된다. 2차원 DCT는 순차함수 $f(m,n)$ 에 대하여 식(1)과 같으며^(11,12),

$$F(u,v) = \frac{4C(u)C(v)}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m,n) \cos\left[\frac{(2m+1)u\pi}{2N}\right] \cos\left[\frac{(2n+1)v\pi}{2N}\right]$$

단, $u,v=0,1,2,\dots,N-1$

$$C(w) = \begin{cases} 1/\sqrt{2} & \text{for } w=0 \\ 1 & \text{for } w=1,2,\dots,N-1 \end{cases} \quad (1)$$

역 변환은 식(2)와 같다.

$$f(m,n) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v) F(u,v) \cos\left[\frac{(2m+1)u\pi}{2N}\right] \cos\left[\frac{(2n+1)v\pi}{2N}\right]$$

단, $m,n=0,1,2,\dots,N-1$ (2)

위 식에서 N 은 변환 블록의 크기를 의미한다.

일반적으로 정현과 함수를 기본 벡터로 갖는 변환 방식은 공간 영역에서의 신호의 에너지를 주파수 영역으로 변환하는 특성이 있다. 특히 Sine과 Cosine 함수를 기본 벡터로 갖는 푸리에 변환은 이러한 특성을 잘 반영함으로 신호처리에 가장 많이 이용된다. DCT는 Cosine 함수를 기본 벡터로 사용하기 때문에 푸리에 변환과 많은 유사성이 있다. 그러나, DCT 영역의 공간 주파수는 푸리에 변환 영역의 공간 주파수와 정확히 일치하지 않는다. 따라서, DCT 영역에서의 주파수의 의미를 분석해 보고, 이를 신호처리에 이용하려는 연구가 이루어 졌다^(13,14).

필터의 주파수 반응 특성이 실수이며 우함수일때,

필터의 DFT 계수와 신호의 DCT 계수와와 곱은 공간영역에서 순환 컨볼루션한 것과 같으며, 식(3)으로 나타낼 수 있다.

$$\text{if } Y_c(k) = X_c(k) H_F(k) \quad k = -N, -N+1, \dots, N-1 \\ \text{then } y(n) = x'(n) * h(n) \quad n = 0, 1, 2, \dots, N-1 \quad (3)$$

식(3)은 다음의 과정을 통하여 증명된다. 먼저, 길이 N 의 1차원 DCT와 그의 역 변환은 식(4,5)와 같이 정의된다.

$$X_c(k) = \frac{2}{N} C(k) \sum_{n=0}^{N-1} x(n) \cos\left[\frac{(2n+1)\pi k}{2N}\right] \\ ; k = 0, 1, \dots, N-1 \quad (4)$$

$$x(n) = \sum_{k=0}^{N-1} C(k) X_c(k) \cos\left[\frac{(2n+1)\pi k}{2N}\right] \\ ; n = 0, 1, \dots, N-1 \quad (5)$$

$$\text{단, } C(k) = \begin{cases} 1/\sqrt{2} & k=0 \\ 1 & k=1,2,\dots,N-1 \end{cases}$$

$n=-1/2$ 에 대하여 대칭으로 확장된 신호 $x'(n)$ 은 식(6)과 같다.

$$x'(n) = \begin{cases} x(n) & n=0,1,2,\dots,N-1 \\ x(-1-n) & n=-N,-N+1,\dots,-1 \end{cases} \quad (6)$$

이 때, 대칭 확장된 신호 $x'(n)$ 의 푸리에 변환 계수 $X_{F'}(k)$ 와 $x(n)$ 의 DCT 계수 $X_c(k)$ 간에는 식(7)의 관계가 성립한다.

$$X_c(k) = 2C(k) \exp\left(-\frac{j\pi k}{2N}\right) X_{F'}(k) \quad k=0,1,\dots,N-1 \quad (7)$$

식(3)에서 Y_c 는 실수이며 우함수이므로, 그의 역 DCT는 식(8)과 같다.

$$y(n) = \frac{1}{2} \text{Re}\left\{ \sum_k C(k) X_c(k) H_F(k) \exp\left[\frac{j(2n+1)\pi k}{2N}\right] \right\} + \frac{1}{2} C(0) X_c(0) H_F(0) \quad (8)$$

식(8)을 식(7)과 결합하면 식(9)가 얻어진다.

$$y(n) = \text{Re}\left\{ \sum_{k=0}^{N-1} C^2(k) X_F'(k) H_F(k) \exp\left(\frac{j2\pi nk}{2N}\right) \right\} + C^2(0) X_F'(0) H_F(0) \quad (9)$$

식(9)에서 마지막 항은 실수이므로 이를 첫항의 Re 기호에 함께 묶으면 다음의 결과를 얻을 수 있다.

$$\begin{aligned} y(n) &= \text{Re}\left\{ \sum_{k=0}^{N-1} C^2(k) X_F'(k) H_F(k) \exp\left(\frac{j2\pi nk}{2N}\right) \right\} + C^2(0) X_F'(0) H_F(0) \\ &= \text{Re}\left\{ \sum_{k=0}^{N-1} X_F'(k) H_F(k) \exp\left(\frac{j2\pi nk}{2N}\right) \right\} \\ &= \text{Re}\{\text{IDFT}[X_F'(k) H_F(k)]\} \\ &= \text{Re}\{x'(n) * h(n)\} \\ &= x'(n) * h(n) \quad n=0,1,\dots,N-1 \quad (10) \end{aligned}$$

위의 결과로 부터 푸리에 영역에서 설계된 필터의 주파수 특성과 DCT변환 계수와와 곱은, 원 신호의 대칭 확장된 신호에 대한 푸리에 변환과 같음을 알 수 있다. 따라서, DCT는 원 신호를 정확하게 주파수 성분으로 분해하는 것이 아니라, 먼저 대칭 확장 후에 분해함으로 스펙트럼 분석등의 목적으로 DCT를 이용하기 위해서 변환된 각 계수에 대한 보정이 필요하다.

III. 인간의 시각 체계(HVS : Human Visual System)

HVS는 최근까지 꾸준히 연구되어온 분야이지만 하지만 정확한 시각모델이 없어 실제적으로 영상 부호화에 적용되는 데에 어려움이 있었다. HVS는 영상 부호화(압축)에 중요한 의미를 지니는 데, 그 이유는 대부분의 압축기법이 '정보 보존 형식'이 아니라, 단지 '정보 근사화'에 목적을 두고 있기 때문이다. 따라서, 재구성된 영상은 항상 어느 정도의 왜곡이 따르게 되며, 이러한 영상의 최종 판단자는 인간이므로 HVS를 영상 압축에 응용한다는 것은 타당한 근거를 지니게 된다^(5,15,16).

HVS를 모델링 하기 위해서는 일반적으로 다음 사항들을 고려해야 한다.

1. 빛의 세기에 대한 시각의 반응

2. 영상의 공간 세밀성에 대한 시각의 반응
3. 공간 주파수에 대한 시각의 반응
4. 색신호에 대한 시각의 반응

위 4가지중에서 1~3은 어느정도 정확하게 실험적으로 확인되었지만, 색신호에 대한 시각의 반응은 아직 정확한 결과가 얻어지지 않았다. 또한 4가지가 각각 독립적으로 작용하는 것이 아니라 서로 결합되어 반응을 나타내므로 HVS를 모델링하는데 더 큰 어려움을 준다. 그러나 흑백 정지영상에 대한 시각의 반응은 어느정도 연구가 되어 있으므로, HVS를 근사적으로 모델링할 수 있으며 이러한 HVS 모델링은 효율적인 영상 부호화에 응용될 수 있다. 그림 1은 시각체계를 근사적으로 모델링한 블록도이다.

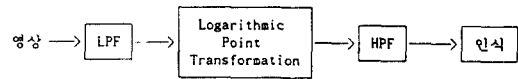


그림 1. 시각 체계의 모델링.

Fig.1. Modelling of Human Visual System.

시각 체계에 의해 다음과 같은 시각 현상이 발생한다.

- 빛의 세기에 대한 비선형적 반응 특성-웨버의 법칙으로 모델링
- 공간 주파수 특성-MIF로 모델링
- 공간 마스킹 특성(Spatial masking effect)-정확한 모델링이 아직 이루어지지 않았음.

이러한 시각 특성중 특히 공간 주파수에 대한 MTF가 활발히 연구되어 왔으며^(9,13,14) 공간 마스킹 특성과 함께 영상 부호화에 효과적으로 응용될 수 있다.

공간 주파수단 영상의 공간 세밀성을 나타내는 것으로서, 시각 1°에 대한 정현파 패턴의 변화율로써 나타내며, 단위는 cpd 즉, Cycles /degree를 사용한다. 인간의 시각 체계는 공간 주파수를 특정 주파수의 협대역 채널로 분리한 후, 각 채널의 출력을 독립적으로 처리하여 영상을 인지함이 알려졌다^[5]. 이러한 특성은 영상신호를 기본 벡터로 분리한 후 처리하는 변환 부호화 방식과 유사하다. 특히, DFT, DCT,

DST와 같이 기본 벡터로써 정현과 함수를 사용하는 변환 부호화에는 공간 주파수 특성이 매우 효율적으로 응용될 수 있다.

Sakrison은 식(11)로 표현되는 격자 모양의 정현파 패턴을 사용하여 심리시각적인 실험을 행하여 시각의 공간 주파수 반응 특성을 연구하였다.

$$u(x,y) = U_0 + \alpha p(x \cos \theta - y \sin \theta) \quad (11)$$

식(11)에서 U_0 는 영상의 평균 밝기를 나타내며, α 는 U_0 를 정점으로하여 진동하는 정현파 패턴의 최대 변화치이며, $p(\cdot)$ 는 주기함수를 나타낸다. 실험결과, 식(12)와 같은 공간 주파수 반응 특성이 얻어졌다.

$$\frac{U_0}{\alpha} \Big|_{\text{threshold}} \cong C \frac{f}{f_0} e^{-f/f_0} \quad (12)$$

fo = 최대 반응 주파수

수직, 수평 방향의 패턴에 대하여 같은 결과를 얻었으며, 다만 식(11)에서 $\theta=45^\circ$ 인 패턴에 대해서는 약 3dB 정도의 감도가 떨어짐이 관찰되었다.

위 결과를 Sakrison은 식(13)으로 모델링하였으며, 이러한 $H(f)$ 를 MFT(Modulation transfer function)라 한다.

$$H(f) = 2.6[0.0192 + 0.114f] \exp[-(0.114f)^{1.1}] \quad (13)$$

f : radial frequency = $(f_x^2 + f_y^2)^{1/2}$ (cpd)

많은 연구자들에 의해 인간의 시각은 4.7~8 (cpd)에서 최대 반응 감도를 나타내며, 0.1과 30~40 (cpd)에서 최대치의 약 3%의 반응을 나타냄이 알려졌다. 따라서, DCT등을 사용하여 영상을 변환한 후 MTF에 따라 차등적으로 양자화하면, 균일한 양자화를 행한 영상보다 같은 비트율에서 좀더 주관적으로 우수한 영상을 얻을 수 있다.

인간의 눈에 인지되는 밝기변화의 임계치는 근접한 Pixel의 밝기변화에 민감하다. 즉, 주변 영역에 밝기의 변화가 존재하지 않을 때의 임계치보다 존재할 때의 임계치가 더 높다. 이를 공간 마스크 특성이라 한다. 공간 마스크 특성을 이용하여 변화가 적은 블럭은 변화가 많은 블럭에 비하여 상대적으로 더 정확하게 양자화를 행함으로써 시각적으로 보다 우수한 부호화를 행할 수 있다. 즉, 인간의 눈은 고변화 영역

의 양자화 오차에 대하여 덜 민감하므로, 양자화 레벨을 영상의 변화도에 따라 차등화함으로써 같은 비트율에서 영상의 주관적인 품질을 향상시킬 수 있다⁽¹⁷⁾.

IV. HVS를 고려한 영상 부호화

HVS에 의한 시각 특성중 공간 주파수 반응 특성을 나타내는 MTF는 원 영상 신호를 공간 주파수로 분해하여 처리하는 변환 부호화에 효과적으로 응용될 수 있다. 즉, 변환된 각 계수에 대하여 MTF에 따른 가중치를 부여한 후, 부호화하면 시각에 좀더 적합한 영상을 얻을 수 있다. 일반적으로 영상의 변환 부호화에는 압축율이 뛰어난 DCT를 사용한다. 그러나, DCT는 원영상 신호를 정확한 공간 주파수로 분해하지 않는다^(13,18). 다음은 잘 알려진 DCT의 성질 중 하나이다. $g(x)$ 가 식(14)와 같이 우함수이며,

$$g(x) = f(-x) + f(x) \quad \text{where } f(-x) = f(x) \quad (14)$$

또, 실수값을 갖는 함수일때, $g(x)$ 의 푸리에 변환은 식(15)와 같이 정확히 $g(x)$ 의 DCT로 표현될 수 있다.

$$\begin{aligned} \int_{-\infty}^{\infty} g(x) \exp(-j2\pi fx) dx &= \int_{-\infty}^{\infty} g(x) \cos(2\pi fx) dx \\ &= 2 \int_0^{\infty} f(x) \cos(2\pi fx) dx \end{aligned} \quad (15)$$

따라서 DCT는 식(14,15)와 같이 원 신호를 먼저 대칭으로 확장한 후 푸리에 변환하는 것과 같은 효과를 나타낸다.

1985년 Ni11은 DCT영역에서 직접 MTF를 반영할 수 있도록 식(16)과 같은 연산자를 제안하였다⁽¹⁵⁾.

$$\begin{aligned} |A(f)| &= \left[\frac{1}{4} + \frac{1}{112} \right. \\ &\left. \left[\log_e \left[\frac{211}{\alpha} f + \left[\frac{411^2}{\alpha^2} f^2 + 1 \right]^{1/2} \right]^2 \right]^{1/2} \right] \quad (16) \end{aligned}$$

식(16)의 $A(f)$ 를 MTF에 곱함으로써, 시각의 공간주파수 반응 특성을 직접 DCT 영역에 반영할 수 있다.

변환 부호화에서는 영상의 변환후 데이터 압축을

위하여 반드시 양자화를 행해야한다. 본 논문에서는 DCT영역의 공간주파수에 따라 차등적인 양자화를 행하도록 하는 양자화 행렬을 제안하였다. 양자화 행렬은 변환된 각 DCT 계수의 양자화를 위한 양자화 단계(Step size)를 나타내며 양자화 단계가 증가할 수록 양자화 오차의 범위는 증가하게 된다. MTF는 위의 양자화 오차에 대한 시각의 반응특성을 나타내므로 MTF에 따라 각 계수에 대한 중요도를 결정하여 양자화 행렬을 구현할 수 있다. 본 논문에서는 Sakrison이 제안한 MTF를 근거로 하여 양자화 행렬을 구현하였다. 식(17)은 Sakrison이 제안한 MTF이며, 그림 2는 MTF를 DCT 영역에 적용하기 위하여, Nill의 변환 함수 A(f)와 곱한후의 결과를 나타낸다.

$$H(f) = 2.6[0.0192 + 0.114f] \exp[-(0.114f)^{1.4}]$$

$$f : \text{Radial frequency (cpd)} = (fx^2 + fy^2)^{1/2} \quad (17)$$

MTF의 공간주파수 f는 식(18)과 같이 나타낼 수 있다.

$$f(\text{Cycles/degree}) = f_n(\text{Cycles/pel}) f_s(\text{Pels/degree}) \quad (18)$$

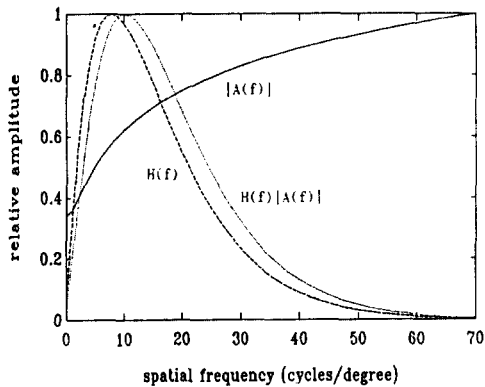


그림 2. DCT
Fig.2. Characteristics of spatial frequency in DCT area.

식(18)에서 f_n 은 정규화된 공간주파수를 의미하고, f_s 는 시각 1°안에 포함되는 화소의 수를 나타낸다. DCT의 첫번째 AC 계수가 1/2의 Cosine 주기를 가지므로, 식(18)의 f_n 은 식(19)와 같다.

$$f_n = k/2N \quad k=0,1,2,\dots,N-1$$

N : 변환 블록 크기 (19)

본 논문에서는 f_s 를 64(Pels/degree)로 설정하여 실험하였으며, 이것은 화면 크기의 4배 정도의 거리에서 영상을 관측하는 것과 같다.

위의 과정을 기저 DCT 영역에서의 공간주파수는 식(20)과 같이 샘플링된다.

$$f = \frac{(k+m)^{1/2}}{2 \times N} f_s \text{ CPD} \quad k,m=0,1,\dots,N-1 \quad (20)$$

식(4.7)에서 얻은 공간주파수 f를 식(4.4)의 MTF에 대입한 후, Nill의 변환함수 |A(f)|를 곱하여 식(21)과 같은 DCT 계수에 대한 상대적인 가중치를 구할 수 있다.

$$H(k,m) = H(f)|A(f)| \quad k,m=0,1,\dots,N-1 \quad (21)$$

식(4.8)의 H(k,m)은 변환계수 F(k,m)에 대한 상대적인 중요도를 의미하며, 이의 역수를 취함으로써, 식(22)와 같은 양자화 행렬(Quantization matrix)를 구할 수 있다.

$$QM(k,m) = 1/H(k,m) \quad k,m=0,1,\dots,N-1 \quad (22)$$

그림 3은 16×16 변환블록에 대한 양자화 행렬을 나타낸다.

40	19	13	11	10	10	11	11	13	15	17	20	24	28	34	42
19	16	12	11	10	10	11	12	13	15	17	20	24	28	34	42
13	12	11	10	10	10	11	12	13	15	17	20	24	29	35	43
11	11	10	10	10	10	11	12	14	16	18	21	25	30	36	45
10	10	10	10	10	11	12	13	14	16	19	22	26	32	38	47
10	10	10	10	11	12	12	14	15	18	20	24	28	34	41	50
11	11	11	11	12	12	14	15	17	19	22	26	31	36	44	53
11	12	12	12	13	14	15	17	19	21	24	28	34	40	48	58
13	13	14	14	15	17	19	21	24	27	32	37	44	53	64	
15	15	15	16	16	18	19	21	24	27	31	36	42	50	59	71
17	17	17	18	19	20	22	24	27	31	35	41	48	56	67	80
20	20	20	21	22	24	26	28	32	36	41	47	55	64	76	91
24	24	24	25	26	28	31	34	37	42	48	55	63	74	88	105
28	28	29	30	32	34	36	40	44	50	56	64	74	87	102	121
34	34	35	36	38	41	44	48	53	59	67	76	88	102	120	142
42	42	43	45	47	50	53	58	64	71	80	91	105	121	142	167

그림 3. MTF
Fig.3. Quantization matrix by MTF.

시각의 공간 마스크 특성에 의하여 사람의 시각은 변화가 심한 영역의 오차에 덜 민감하다. 따라서, 영

상의 변화도에 따라 차등적인 양자화를 하면, 부호화의 효율을 높일 수 있음을 예측할 수 있다. 이러한 시각 특성을 DCT 영상 부호화에 적용하기 위해 본 논문에서는 각 블록의 공간영역에서 블록의 변화도를 측정하기 위하여 다음의 연산을 수행하였다.

$$\begin{aligned}
 C_x &= C_x + 1 \quad \text{if } |f(i,j) - f(i,j+1)| > T \\
 &\quad i=0,1,\dots,N-1 \\
 &\quad j=0,1,\dots,N-2 \\
 C_y &= C_y + 1 \quad \text{if } |f(i,j) - f(i+1,j)| > T \\
 &\quad i=0,1,\dots,N-2 \\
 &\quad j=0,1,\dots,N-1 \\
 C &= C_x + C_y \quad (23)
 \end{aligned}$$

식(23)으로부터 각 블록에 대하여, 정해진 임계값 T를 넘는 변화의 갯수 C가 얻어지며, C는 영상의 변화도에 근사적으로 비례하는 값을 갖는다. 따라서 C에 따라 4.2절에서 구한 양자화 행렬에 가중치를 주어, 각 블록에 대하여 적응적인 양자화를 행할 수 있다.

DCT를 사용한 영상 부호화 방식에서는 영상전체를 한번에 변환하지 않고 블록(부영상)으로 나누어 처리를 행한다. 따라서 고압축율로 영상을 부호화할 경우, 블록간의 경계부분에 불연속적인 경계가 발생하는 블록화 현상이 발생한다. 본 논문에서는 블록화 현상을 줄이기 위한 후처리로서, 복원영상의 블록의 경계부분에 대하여 LPF(Low pass filter)를 취하는 방안을 사용하였다. 전체 영상에 대하여 LPF를 취하면 영상이 전체적으로 흐려지는 현상이 나타나지만, 블록의 경계부분에만 LPF를 취할 경우 그러한 현상은 크게 눈에 띄지 않는다. 블록의 경계부분에 대해서만 선택적인 LPF를 취함으로써 블록화 현상을 어느정도 줄일 수 있으며, 블록간의 경계값을 보정해줌으로써, SNR역시 높일 수 있다. 그림 4는 블록화 현상 보정을 위하여 사용한 LPF 마스크와 마스크가 적용되는 블록의 경계부분을 나타낸다.

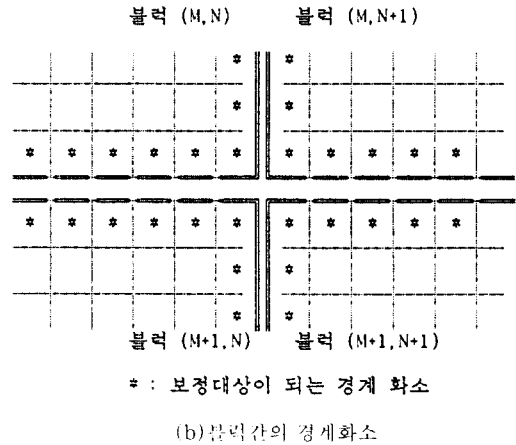
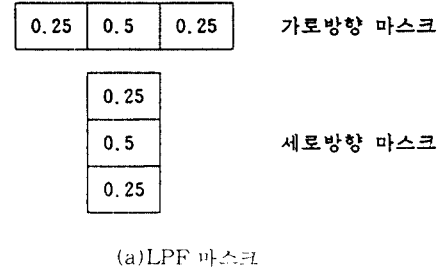


그림 4. 블록화 현상 보정을 위한 후처리.
Fig.4. Post processing for block effect compensation.

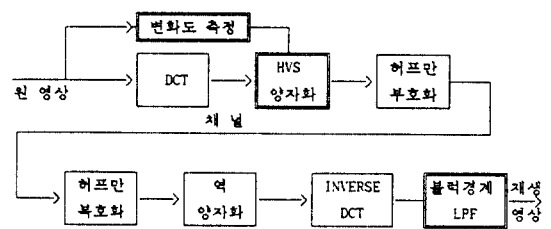


그림 5. 제안한 부호화 방식.
Fig.5. Proposed coding scheme.

V. 실험 및 결과 고찰

본 논문에서는 Chen 등이 제안한 부호화 방식에 HVS에 의한 적응 양자화 방법을 도입하여 실험을 행하였다.

그림 5는 제안한 부호화 방식의 전체 계통도이다.

Chen 방식에서는 식(24)와 같이 DCT된 계수 $f(u, v)$ 를 일률적인 정규화 인자 D 를 사용하여 나누어 다음 실수 계수를 정수로 변환함으로써 양자화를 행하였다.

$$F_x(u,v) = f(u,v) / D$$

$$\begin{aligned} \hat{F}_N(u,v) &= \text{정수 부분}[F_N + 0.5] \text{ if } F_N(u,v) > 0 \\ \hat{F}_N(u,v) &= \text{정수 부분}[F_N - 0.5] \text{ if } F_N(u,v) < 0 \end{aligned} \quad (24)$$

반면, 제안한 부호화 방식에서는 HVS를 고려하여 구한 양자화 행렬을 사용하여 계수를 나누어 줌으로써 블록내의 계수의 위치에 따라 차등적인 양자화를 행하였다. 또한 블록의 변화도를 측정후, 변화도가 큰 블록에 대해서는 양자화 간격을 크게 함으로써, 각 블록간에도 적응적인 양자화가 이루어지도록 하였다. 식(25)는 제안한 방식의 양자화 방법을 나타낸다.

$$\begin{aligned} F_N(u,v) &= f(u,v) / QM(u,v) \\ \hat{F}_N(u,v) &= \text{정수 부분}[F_N + 0.5] \text{ if } F_N(u,v) > 0 \\ \hat{F}_N(u,v) &= \text{정수 부분}[F_N - 0.5] \text{ if } F_N(u,v) < 0 \end{aligned} \quad (25)$$

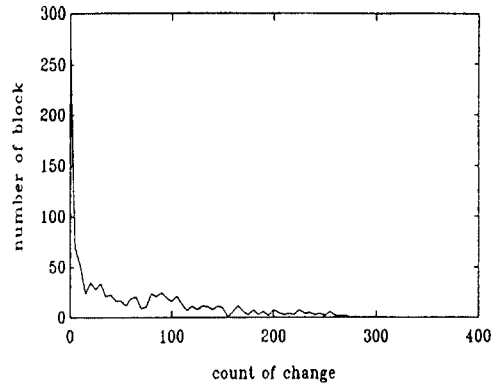
자연 영상을 DCT로 변환 하면 영상의 특성상 대부분의 에너지가 저주파 계수에 집중된다. 따라서 양자화를 행하게되면 저주파 계수를 제외한 대부분의 고주파 계수가 '0'이 된다. 이러한 변환블록을 지그재그 스캔하여 2차원 배열을 1차원 배열로 변환하면, 배열 뒷부분의 계수는 대부분이 '0'이 되므로 제로 Run-length 부호화를 하여 효율적인 엔트로피 부호화를 할 수 있다.

허프만 부호화기는 가변길이 부호화의 일종으로 입력되는 데이터의 확률분포에 따라 그에 준 최적인 코드를 발생하므로, 본 논문에서는 Chen 방식과 같은 허프만 부호화기를 사용하였다. 이때, 블록의 평균 계조도를 나타내는 DC 계수는 블록화 현상에 중요한 영향을 미침으로, 일률적으로 9 bit를 할당하였다.

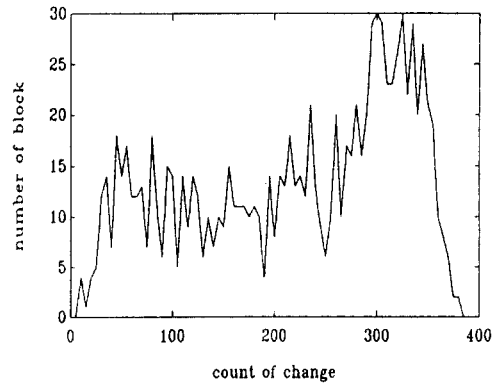
양자화 행렬은 한 블록 내에서 시각 특성에 따른 각 계수의 상대적인 가중치를 나타낸다. 즉, Sakrison의 실험에 의하여 MTF는 시각의 공간주파수 반응에 대한 상대적인 임계값을 나타내므로, 모든 주파수 성분에 대해 그값이 특정 임계값을 넘으면 시각에 영향을 주는 양자화 오차가 발생하게 된다. 이러한 양자화 오차는 변환 블록 전체에 영향을 미치는데, 특히 저변화 블록에 대해서는, 같은 양자화 오차가 더욱 심하게 눈에 띄게 된다. 그 이유는 고변화 공간영역에서는 원래의 신호가 주위에 발생한 양자화 오차를 마스킹하기 때문이다. 본 논문에서는 시각의 공간 마스킹 특성을 이용하기 위하여 각 블록을 식(23)의 변화도에 따라 차등적인 변환정수 SF를 부여

하였다.

변화의 기준이 되는 임계값 T는 실험을 통하여 '10'으로 설정하였다. 그림 6은 Lena 영상과 Baboon 영상에 대한 변화도 측정의 결과를 나타낸다.



(a) Lena 영상의 변화도



(b) Baboon 영상의 변화도

그림 6. 실험영상의 변화도.

Fig.6. Activity of experimented image.

그림 6에서 알 수 있듯이 Lena 영상은 대부분의 블록이 변화도가 50 이내인 단순한 영상이며, 반면 Baboon 영상은 변화도가 매우 심한 복잡한 영상임을 알 수 있다.

변화도에 따른 블록간의 차등적인 양자화 적용의 방법으로서, 각 블록을 변화도에 따라 표 1과 같이 4가지 클래스로 분류하였으며, 각 클래스에 대하여 4가지의 변환정수 SF를 설정하였다.

표 1. 변화도에 따른 변환정수.

블럭의 분류	변화도(C)	변환정수(SF)
클래스 1	$C < 100$	10.0
클래스 2	$100 \leq C < 200$	7.0
클래스 3	$200 \leq C < 300$	5.0
클래스 4	$300 \leq C$	2.0

각 블럭에 대하여 양자화를 행하기 전에 양자화 행렬(QM)을 클래스에 할당된 변환정수 SF를 사용하여, 다음과 같이 변환한다.

$$QM = QM / SF \quad (26)$$

부호화시 각 블럭은 변화도에 따라 4가지 클래스로 분류되며, 각 클래스에 따라 적응적으로 양자화 되게 된다. 단, 여러 비트율에서 성능을 측정하기 위하여 본 실험에서는 SF와 C를 선형적으로 변화시켜가며 실험을 행하였다.

제안한 부호화 방법의 성능을 Chen등이 제안한 단순양자화기를 사용한 기존의 방식[8]과 비교하였다. 그림 7은 실험에 사용한 원영상이며, 그림 8은 각 방식에 의하여 부호화한 후 재생된 영상이다.

영상의 평가방법에는 객관적 방법과 주관적 방법이 있다. 수치적으로 품질을 평가하는 객관적 방법에는 PSNR(Peak to peak signal to noise ratio)을 사용하였다.



(b) Baboon 영상

그림 7. 실험에 사용한 원영상.

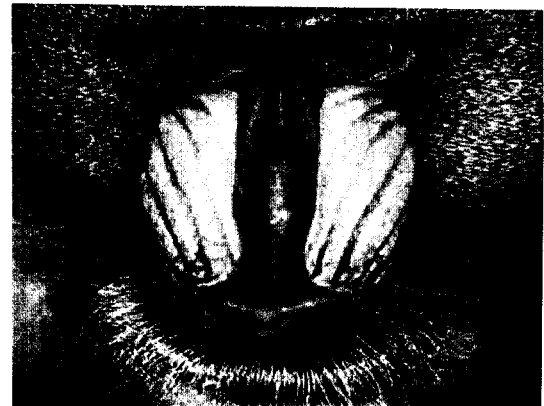
Fig.7. Original image used in this experiment.



(a) 기존 방식에 의해 재생된 Lena 영상



(a) Lena 영상



(b) 기존 방식에 의해 재생된 Baboon 영상



(c) 제안된 방식에 의해 재생된 Lena 영상



(d) 제안된 방식에 의해 재생된 Baboon 영상

그림 8. 각 방식에 의해 재생된 영상(0.5 bpp).

Fig.8. Reconstructed image by both coding schemes.

표 2. 객관적 평가 결과.

(a) Lena 영상

	0.25	0.5	0.7	1.0	bpp
기존방식	32.1	35.4	36.9	38.3	dB
제안한 방식	30.4	33.9	36.4	38.2	dB

(b) Baboon 영상

	0.25	0.5	0.7	1.0	bpp
기존방식	21.7	23.8	25.2	27.5	dB
제안한 방식	21.6	23.4	24.4	25.7	dB

표 2는 각 방식에 대한 객관적 성능 평가 결과를 나타낸다. PSNR을 사용하는 객관적 평가는 신호의 에너지에 대한 충실도는 효과적으로 반영하지만 사람의 시각특성은 정확하게 반영하지 못한다. 따라서, 시각적으로 더 우수한 영상이 PSNR은 더 낮은 경우가 발생할 수 있으며, 표 2로부터 제안한 방식이 PSNR면에서는 기존의 방식보다 낮은 결과를 보임을 알 수 있다.

주관적 평가는 영상처리에 경험이 있는 10명의 평가자를 대상으로 실시하였으며, 각 방식에 의하여 같은 비트율로 처리한 재생영상을 원영상과 함께 임의적인 순서로 10초씩 화면에 나타낸 후 표 3의 기준에 의하여 평가하게 하였다^{[12][19]}. 평가전에 전체 이미지를 5초씩 화면에 나타내어 평가자들이 평가 대상인 영상들을 미리 검색하도록 하여 영상의 화질의 비슷한 경우 상대적인 평가를 내리는 데 도움이 되도록 하였다. 표 4와 그림 9는 모든 평가자들의 평가결과를 종합한 후 평균을 취한 결과이다.

표 3. 주관적 평가 기준.

평가내용	등급
원영상과 식별되지 않음	7
원영상과 거의 식별되지 않음	6
원영상과 명확히 식별되지만 훼손정도가 가벼움	5
훼손됐지만 거부감이 들지 않음	4
이 정도 거부감이 생김	3
지나치게 거부감이 생김	1

표 4. 주관적 평가 결과.

(a) Lena 영상

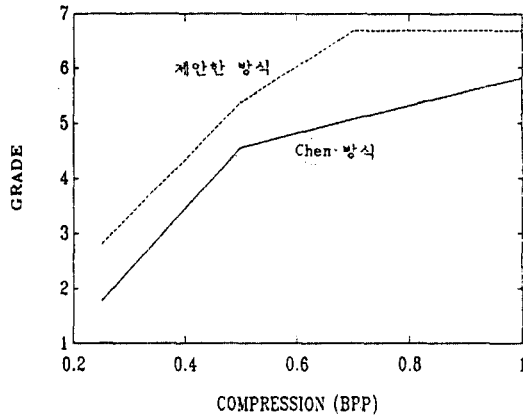
	0.25	0.5	0.7	1.0	bpp
기존방식	1.78	4.56	5.07	5.81	등급
제안한 방식	2.81	5.36	6.69	6.70	등급

(b) Baboon 영상

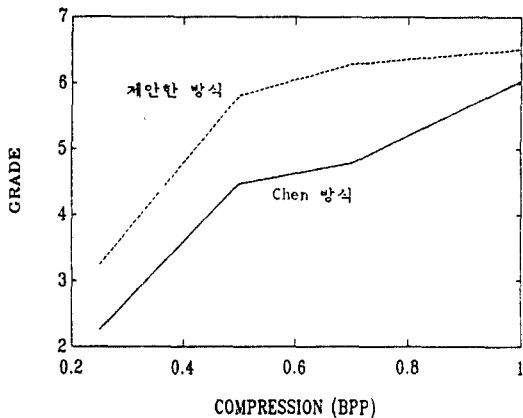
	0.25	0.5	0.7	1.0	bpp
기존방식	2.26	4.47	4.79	6.04	등급
제안한 방식	3.24	5.81	6.29	6.52	등급

주관적인 평가결과, 1.0 (bpp) 이상에서 3방식 모두 시각적으로 우수한 성능을 나타내었으며, 1.0 (bpp) 이하의 고압축 영상에 대해서는 제안한 방식이 상대적으로 보다 우수한 성능을 나타냄을 확인하였다. 그

이유는 Chen 방식이 영상 전체를 일률적으로 양자화하는 반면, 제안한 방식은 고변화 영역에서 비트를 절약하여, 시각에 민감한 저변화 영역을 상대적으로 강조함으로써, PSNR은 낮지만, 인간의 시각에 의한 주관적 평가에서는 보다 우수한 결과를 얻을 수 있었다.



(a) Lena 영상



(b) Baboon 영상

그림 9. 주관적 평가 결과.

Fig.9. Results of subjective evaluation.

정규화 인자로 전체 변환블럭을 양자화한 Chen 방식⁽⁶⁾(기존방식)보다 시각적으로 우수한 결과를 나타내었다. 반면, 객관적 수치인 PSNR은 최대 2.0 dB 정도 하락함이 관찰되었다. 그 이유는 시각에 민감한 저변화 블럭이 충실히 부호화되는 대신, 에너지가 많은 고변화 블럭이 상대적으로 많이 훼손됨으로써 전체적인 영상에서 MSE(Mean square error)가 증가하기 때문이다. 그러나 시각의 공간마스킹 특성에 의하여 고변화 영역에 나타나는 오차는 인간의 시각에 덜 민감하므로, 제안한 방식이 PSNR은 낮지만 주관적 평가에서는 더 우수한 결과를 보여주었다. 특히, Baboon 영상과 같이 변화도가 심한 영상에서는 변화도에 따라 양자화 행렬의 변환 정수인 SF를 효율적으로 설정함으로써, 시각에 민감한 저변화 영역을 고변화 영역보다 상대적으로 강조하여 HVS를 고려하지 않은 Chen 방식⁽⁶⁾보다 시각적으로 매우 우수한 품질의 영상을 얻을 수 있었다.

VI. 결 론

본 논문에서는 인간 시각체계의 공간주파수와 공간세밀성에 대한 특성을 영상압축에서 우수한 성능을 나타내는 변환부호화 방식인 DCT와 결합하여 주관적 화질이 우수한 부호화 방법을 연구하였다. 영상을 저장, 전송할 경우 영상 데이터의 방대한 양으로 인하여, 훼손이 따르는 영상압축이 불가피하다. 그러나 그 결과를 인간이 판단하므로 인간 시각체계를 영상 압축방법에 이용하면 저전송율에서 시각적으로 보다 우수한 화질을 나타내는 영상 부호화기를 구현할 수 있다.

인간의 시각은 입력되는 영상을 몇개의 협대역 주파수대로 분리한 후 개별적인 처리과정을 거쳐 영상을 인지함이 알려졌다. 이때 각각의 주파수대에 대한 시각의 반응 특성은 각기 다르며, 일반적으로 사람의 시각은 중간 주파수대에 가장 민감하고, 저주파와 고주파성분에 대해서는 반응감도가 떨어짐이 실험을 통하여 확인되었다. 이러한 시각특성은 MTF로 모델링되어 영상의 주관적 평가방식과 영상의 압축 부호화에 이용되어 왔다. 본 논문에서는 MTF를 근거로 하여 16×16의 DCT 변환블럭에 적용될 수 있는 양자화 행렬을 연구하였다. 이 행렬을 이용하여 DCT 계수를 양자화한 결과 시각에 민감하지 못한 대부분의 고주파성분이 제로가 되어 제로 Run-length 부호화

MTF를 반영한 양자화 행렬과 영상의 변화도에 따라 블럭간에 가중치를 주어 부호화한 방식이 일정한

의 효율을 높일 수 있으며, 상대적으로 시각에 민감한 저주파 성분을 강조함으로써, 같은 비트율에서 비교했을 때 영상의 시각적 품질을 높일 수 있었다.

인간의 시각이 변화가 심한 영역에서의 양자화 오차를 잘 감지하지 못하므로, 본 연구에서는 각 블록을 공간 변화도에 따라 분류한 후 차등적 양자화를 행하는 방법을 제안하여, 인간의 시각에 영향을 많이 주는 평탄한 영역의 부호화 효율을 증가시킬 수 있었다.

이상과 같은 부호화방법으로 컴퓨터 모의 실험한 결과 기존의 Chen방식⁽⁶⁾에 비하여 1.0 (bpp)이하의 저전송율에서 시각적으로 보다 우수한 품질의 영상을 얻을 수 있었다.

차후 연구과제로는 공간변화도에 대한 시각특성을 정확하게 수학적으로 모델링하는 연구가 필요하며, 이와 더불어 색신호의 시각특성에 대한 연구가 요구된다. 또한 각 비트율에서 변환계수의 통계적 분포가 다르므로 응용분야에 따라 각 비트율에 대한 엔트로피 부호화의 개선이 요망된다.

참 고 문 헌

1. T.J.Lynch, *Data Compression*, Life Learning Publication, 1985.
2. A.K.Jain, "Image data Compression: A Review," IEEE Proc., Vol.69, pp.349-389, 1981.
3. P.A.Wintz, "Transform Picture Coding," IEEE Proc., Vol.60, pp.809-820, 1972.
4. R.J.Clarke, "Relation between the Karhunen-Loeve and cosine transforms," IEEE Proc., Vol.128, pp.359-360, 1981.
5. R.J.Clark, *Transform Coding of Images*, Academic Press, 1985.
6. A.Habibi, "Survey of Adaptive Image Coding Techniques," IEEE Trans. Commu., Vol. COM-32, pp.225-232, 1984.
7. Wen-Hsiung Chen and C.Harrison Smith, "Adaptive Coding of Monochrome and Color Images," IEEE Trans. Comm., Vol. COM-25, pp.1285-1292, 1977.
8. Wen-Hsiung Chen and William K. Pratt, "Scene Adaptive Coder," IEEE Trans. Comm., Vol.COM-32, pp.225-232, 1984.
9. James L.Mannos and David J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," IEEE Trans. Info., Vol. IT-20, pp.525-536, 1974.
10. ISO /IEC JTC1/SC2/WG8, "Coded Representation of picture and Audio information," ISO /JTC1/SC2/WG8 N933 Draft Document, Jan. 1990.
11. N.Ahmed, T.Nartarajan, and K.R.Rao, "Discrete Cosine Transform," IEEE Trans. Comput., Vol.C-23, January, pp.90-93, 1974.
12. K.R.Rao, P.Yip, *Discret Cosine Transform*, Academic Press, 1990.
13. Norman B. Nill, "A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment," IEEE Trans. Comm., Vol.COM-33, pp.551-557, 1985.
14. Bowonkoon Chitprasert, and K. R. Rao, "Human Visual Weighted Progressive Image Transmission," IEEE Trans. Comm., Vol.38, pp.1040-1044, 1990.
15. King N. Ngan, Kin S. Leong, and H. Singh, "Adaptive Cosine Transform Coding of Images in Perceptual Domain," IEEE Trans. ASSP., Vol.37, pp.1743-1749, 1989.
16. T.G.Stockham, "Image processing in the context of a visual model," Proc. IEEE, Vol.60, pp.828-842, July 1972.
17. 박구만, "블록의 변화도를 이용한 적응 DCT 부호화," 연세대학교 대학원 박사학위논문, 1991. 2.
18. Jae S. Lim, *Two Dimensional Signal and Image Processing*, Prentice Hall, 1990.
19. G. Wallace, et al, "Subjective Testing Results for Still Picture Compression Algorithms for International Standardization," GLOBECOM 88, Vol.2, pp.1022-1027, 1988.



南承珍(Seung Jin Nam) 正會員
1966年3月22日生
1989年2月:延世大學校 工科學科 卒業
電子工學科 卒業
1991年8月:延世大學校 大學院 電
子工學科 卒業(工學
碩士)
1991年9月~現在:韓國放送公社
技術研究所 研究員

崔成男(Sung Nam Choi) 正會員
1965年8月25日生
1987年2月:亞洲大學校 電子工學科 卒業
1989年2月:延世大學校 大學院 電子工學科 卒業(工學
碩士)
1989年3月~現在:延世大學校 大學院 電子工學科 碩士
課程

全重滿(Joong Nam Jeon) 正會員
1959年2月19日生
1981年2月:延世大學校 電子工學科 卒業
1985年8月:延世大學校 大學院 電子工學科 卒業(工學
碩士)
1990年2月:延世大學校 大學院 電子工學科 卒業(工學
博士)

朴圭泰(Kyu Tae Park) 正會員
1933年6月11日生
1953年~1957年:延世大學校 電氣工學科 卒業
1957年~1964年:延世大學校 大學院 電子工學科 卒業
(工學碩士)
1962年~1964年:英國 LONDON UNIVERSITY, MSC
(工學碩士)
1967年~1969年:英國 SOUTHAMPTON UNIVERSITY,
Ph. D. (工學博士)
1970年~現在:延世大學校 教授