

DCT블록을 이용한 인간 시각시스템 기반의 영상 리사이징 알고리즘 개발

*양 대규, 정 상현, 서태석

KT플랫폼연구소

*yangdg@kt.com, shjeong@kt.com, morning@kt.com

Development of Image resizing algorithm based Human Visual System using DCT block

*Yang, Dae-Gyu, Jeong, Sang-hyeon, Seo Tae-Seok

KT Platform Laboratory Service Platform Development Department

요약

고객의 통신 니즈가 다양해지고 통신 인프라가 대용량, 고속화 되고 단말이 지능화됨에 따라 이를 기반으로 IPTV의 셋탑, 컴퓨터 모니터, 이동통신 핸드폰이 이르기 까지 다양한 화면 사이즈의 단말을 통하여 서비스를 제공하고 있다. 이를 위해 원영상을 화면의 크기에 적합하게 크기를 변환하는 영상 리사이징 기술을 이용하게 되며, 그 방식은 DCT 방식과 공간 방식으로 나눌 수 있다. DCT방식은 원영상의 고주파 영역의 조정하여 크기를 변화하는 방식으로 열화는 최소화되나 DCT 변환을 위한 계산량이 많고 고정 크기 블록을 사용하므로 유연성이 떨어진다. 공간영역은 객관적인 평가에서 품질이 열화되더라도 시각적으로 눈에 민감한 영역의 변화를 최소화하는 것이다. 본 논문은 인간시각시스템을 기반으로 눈의 민감도가 높은 망막 이심율이 작은 영역에서는 DCT블록의 크기를 변환을 최소화시키고 망막이심율이 큰 영역에서는 블록의 변화를 크게 하는 가변 블록방식을 제안한다. 이를 통해 우수한 영상 품질을 확보하면서도 시각적으로 뛰어난 리사이징 영상을 얻을 수 있다.

1. 서론

현재의 통신 인프라는 유비쿼터스 환경에서 통신, 방송, 인터넷이 융합된 품질보장형 광대역 멀티미디어 서비스를 언제 어디서나 끊김 없이(Seamlessly) 안전하게 이용할 수 있도록 진화를 거듭하고 있다. 이러한 인프라를 기반으로 서비스는 단순한 음성, Text 위주에서 영화, 음악, 애니메이션 등 다양한 멀티미디어 콘텐츠를 제공할 수 있게 되었다.

단말은 멀티미디어 콘텐츠를 사용자가 이용할 수 있도록 화면에 디스플레이하는 사용자 접점 장치로 크기나 성능에 있어서 다양한 스펙트럼을 가지고 있다. 즉 소형의 개인 휴대용 단말에서 가정용의 60~70 인치 이상의 대규모 사이즈 TV가 사용자의 이용 환경에 따라 용되어 사용되고 있다. 그러나 콘텐츠 측면에서는 단말의 크기나 성능에 맞춤형으로 콘텐츠의 제작은 경제적 비용, 시간적 손실이 발생하고 콘텐츠의 관리 측면에서도 복잡성이 증대되어 현실적인 어려움이 많다. 이를 위해 영상 리사이징 기술은 원본 영상을 필요에 따라 다양한 크기로 변환하는 기술로 컴퓨터 그래픽 분야나 동영상 압축의 SVC(Scalable Video Coding) 분야야 활용되고 있다.

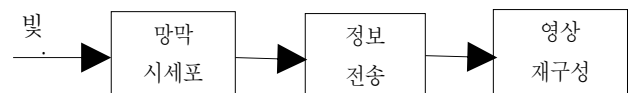
리사이징 방식은 크게 DCT영역 공간 영역으로 나뉘어지며 주파수 영역은 DCT 를 사용하며 에너지대가 높은 저주파수 영역은 보존하고 고주파수의 영역을 가변하여 크기를 변화시키는 방식으로 화질의 저하를 방지할 수 있는 장점이 있으나 블록단위로 가변 크기를 적용하기 위해서 계산량이 많은 단점이 있다. 공간 영역은 시각적으로 민감한 영역에서 변화를 작게하여 시각적으로 우수한 영상을 얻을 수 있으나 원 영상과 모양의 차이가 심해 심한 열화 현상을 가진다. 본 논문에서는 DCT 영역에서 인간 시각 특성을 기반으로 블록별 가변 리사이징

방식을 제안한다. 이 방식은 블록의 위치에 따라 망막 이심율에 기반하여 블록 대비(Contrast) 민감도를 구하고, 민감도에 따라 블록의 크기를 가변하는 방식으로 민감도가 높을 수록 크기 변화를 적게되어 주파수 영역의 장점과 공간 영역에서의 장점을 모두 가질 수 있는 방안이다. 논문의 구성은 다음과 같다. 본 논문의 2장에 인간 시각 시스템의 특징에 대하여 설명한다. 그 중에 특히 공간주파수 변화에 따른 민감도 전달 함수와 망막 이심율에 따른 민감도 변화에 대해 이야기 한다. 3장에서는 주파수 영역의 리사이징 방식과 공간 영역에서의 리사이징 방식에 대해 기술하였다. 그리고 4장에서는 이를 기반으로 디스플레이 화면에 적합한 인간 시각 기반의 영상 리사이징 알고리즘을 제안한다. 마지막으로 5장에서 결론을 맺는다.

2. 인간 시각 특성 분석

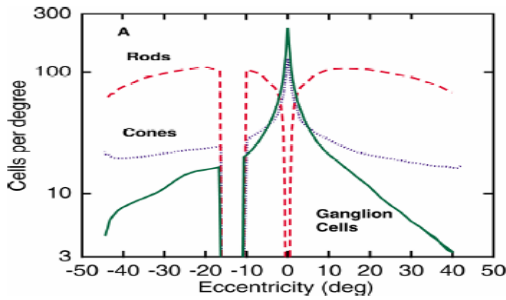
가. 인간 시각 특성

인간이 물체를 감지하는 하기 위해서 빛이 각막과 수정체를 통과하여 망막에 상이 맺히고 시세포(visaul cell)에 의해 감지가 된다. 이렇게 시세포에 의해 인식된 정보는 Gnnlion 셀을 통하여 두뇌로 전달이 되고 두뇌는 영상을 재구성하여 인간이 물체를 보게 되는 구조이다.



[그림 1. 영상 재구성]

여기서 영상의 민감도를 좌우하는 부분은 시세포에 의한 빛을 감지하는 시세포 영역이다. 시세포는 크게 나누어서 간상세포(rod cell)와 원추세포로 나누어지며 간상세포는 주로 어둠에서 명암과 형태 크기를 인식하는 세포이며 원추세포는 척추동물의 망막의 가장 안쪽에 있는 세포로 밝은 곳에서의 색깔을 인식한다.



[그림 2. 시세포 분포 특성]

간상세포와 원추세포는 망막의 foveation point 에 집중적으로 분포되어 있고 가장가지로 갈 수록 분포밀도가 낮아지고 있어 동일한 빛이 망막으로 입수되더라도 시각적 인식능률이 떨어지게 된다. 즉 시력(Visual Acuity), 움직임에 대한 반응, 대비민감도, Color 인식률, 크기 변화 인식율이 모두 감소한다. 그러나 이 감소율이 일정 비율을 가지는 것이 아니라 불규칙성을 가짐으로 이를 수학적 함수로 모델링하기 위해 연구가 지속되고 있다.

나. 공간 주파수 변환에 따른 민감도 변화

인간의 시각 기관은 영상을 공간 주파수 성분에 따라 각기 독립적으로 처리하는 구조를 가지고 있으며 각각의 공간 주파수 성분대 대한 시각 감도가 다른 것으로 알려져 있다. 일반적으로 인간 시각 시스템은 공간 주파수에 대한 감도 변화 특성 함수인 MTF로 모델링하였다. MTF 연구는 주는 신호 압축 분야에서 이루어졌는데 그 목적은 눈에 민감한 영역에서는 부호화 비트율은 높게 하고 둔감한 영역은 부호화 비트율을 낮게 하여 전체적으로 낮은 비트율에서도 시각적으로 뛰어난 화질을 얻을 수 있도록 MTF 를 이용하였다.

인간시각 특성은 시각 시스템의 광학적인 불완전성에 의한 저주파 대역 통과 필터링(low-pass filter)효과와 시각 신경에서의 측방 억제(lateral inhibition) 현상에 의한 고주파 대역 통과 필터(high-pass filter) 역할 등에 의하여 대조 함수는 약 $3 \sim 4.5$ (cycles/degree)에서 최대값을 나타내는 중간 대역 필터의 특성을 갖는 것으로 알려져 있다. Manno와 sakrison은 이러한 시각 특성에 의한 대조 감도 함수의 최대값이 1이 되도록 정규화하여 식(2)과 같은 인간 시각 시스템의 MTF를 제안하였다. 이후 Nill, Rao 등은 식 (3), 식(4)와 같은 MTF들을 제안하였다.

$$H(f) = 2.6(0.0192 + 0.114)\exp(- (0.114f)^{1.1}) \quad (2)$$

$$H(f) = (0.2 + 0.45)\exp(-0.18f) \quad (3)$$

$$H(f) = 2.46(0.1 + 0.25)\exp(-0.25f) \quad (4)$$

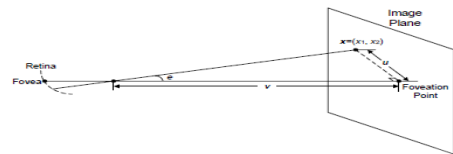
위의 식에서 저주파수에도 대비 민감도가 작고 15(cycle/degree) 이상의 고주파 영역에서 대비 민감도가 급격히 낮아지는 BPF(Band Pass Filter) 특성을 가지고 있다. Lena과 같은 인물영상에서 머리카락

과 같은 고주파 블록의 일부 변화나 배경과 같은 저주파수가 많은 영역에서 대비 변화에 감도가 낮다는 것을 알 수 있다. 그리고 예지 성분이 많은 영역에서 시각적으로 민감하다는 것을 이 함수를 통해 알 수 있다.

다. 망막이심율에 따른 민감도 변화

망막 이심율(Retinal Eccentricity)이란 망막상이 응시점으로 부터 떨어져있는 정도를 말하는 것으로 일반적으로 시각도(visual angle)로 나타낸다. 이심율의 크기는 식(5)로 표시할 수 있다. u 는 foveation point의 위치에서 화면상에서 특정까지의 거리로 $\|x(x_1, x_2) - x(0, 0)\|$ 이다. v 는 화면에서 눈까지의 거리를 나타낸다.

$$\theta = \tan^{-1}(U/V) \quad (1)$$



[그림 3. 망막 이심율]

눈의 구조에서 Fovea에 대다수의 시세포가 분포하므로 foveation point에서는 소량의 빛의 변화나 물체의 크기 변화가 적더라도 시각 민감도가 크므로 변화에 대한 감지율이 매우 높으며 이심율이 큰 영역에서는 감지율이 매우 떨어지게 된다.

W.S. Geisler and J. S. Prerry,는 망막이심율과 공간 주파수의 변화에 따라 시각 민감도 특성을 모델링하여 영상 압축에 이용하였다. 즉 같은 공간수라도 압축 대상이 되는 블록의 위치에 따라 이심율이 달라지게 되고 또한 민감도가 차이가 남을 이용하여 이심율이 큰 영역에서 같은 공간 주파수대라도 양자화 스텝 사이즈를 크게 하고 이심율이 적은 영역에서는 양자화 스텝 사이즈를 줄임으로써 작은 비트에서도 높은 화질로 압축할 수 있는 알고리즘을 제시하였다.

3. 영상 리사이징 알고리즘

가. DCT 영역에서의 리사이징

주파수 영역에서의 리사이징 방식은 먼저 영상을 일정 블록 크기로 분할한 다음 DCT를 수행하여 주파수 영역으로 변환한다. 그리고 리사이징을 행한 다음 IDCT를 수행함으로 리사이징된 영상 데이터를 확보할 수 있다. DCT 변환의 장점은 크기의 변환이 반복적으로 이루어진 경우라도 원 영상에 비해 영상 화질의 저하가 작고, 모양의 변형이 발생하지 않는 장점이 있다. 그러나 DCT의 계산량이 많고 블록단위로 DCT를 수행함에 따라 리사이징 비율이 제한성을 가지고 있다. 즉 임의의 비율로 축소 및 확대에 어려움을 가지고 있다.

Dugard 방식은 먼저 8*8 크기 블록으로 분할하고 각 블록을 8*8 DCT 한다. 크기가 8*8인 DCT계수에서 저 주파수 쪽의 계수 4*4만 떼어내서 4*4 IDCT를 취하면 바로 저역 통과되고 다운샘플링된 데이터를 얻을 수 있다. 크기가 줄어든 영상은 원영상의 모든 저 주파수 성분들을 그대로 가지게 되는 장점이 있다. 영상 크기를 확대할 때는 영상을 4*4 크기 블록으로 분할하고 각 블록을 4*4 DCT 를 한다. 이 블록을 저주파수 DCT계수로 하고 나머지는 영으로 채운 크기가 8*8 역

DCT를 취해주면 바로 업샘플링되고 저역 통과된 데이터 즉 확대된 영상을 얻을 수 있다.

Dugard 방식을 개선한 것이 Mukherjee방식이다. Mukherjee는 Jiang & Feng이 제안한 상위블럭과 하위블럭 사이의 상관계수를 이용하여 영상을 리사이징한 방식이다. 영상을 확대할 때 8*8 DCT 계수를 IDCT한 후 네개의 4*4 블럭으로 만든 후 각 블럭에 대하여 4*4 DCT를 취한다. 이 후 각 4*4 DCT 계수를 8*8 DCT 계수 블럭의 4*4 저주파수 계수로 사용하고 그 이외의 부분을 영으로 채운후 8*8 IDCT를 함으로써 원 영상 크기보다 가로 세로 각각 두배씩 큰 영상을 얻는다.

가장 최근의 리사이징 방식은 Park이 제안한 방식으로 리사이징 비율이 정수가 아닌 임의의 N/M 비율에 대해 Salazar가 제안한 방식을 기반으로 composite length DCT를 이용하여 최적의 DCT와 IDCT의 크기를 구하는 방식이다. 블럭의 크기를 구할 때 $N^*=M \times (8+q)$ 의 식에서 q의 해를 구하는 것으로 N^* 은 N과 M의 최소 공배수이면서 N과 M중에서 큰 값에 8을 곱한 값보다 크거나 같아야 한다.

나. 공간 영역에서의 리사이징

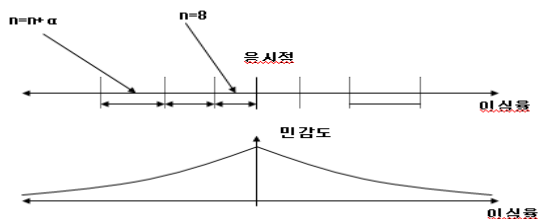
공간 영역에서 리사이징은 직접 영상의 내용을 변경하거나 공간 필터를 사용하여 크기를 변환하는 방식이다. 기존에는 영상을 공간 필터를 사용하여 직접 스케일링하거나 특정한 영역을 Crop하여 영상의 크기를 바꾸었다.

최신 방식으로는 영상 내용물을 기반으로 시각적으로 중요하지 않는 픽셀들을 선택적으로 제거함으로써 영상의 크기를 변경하는 방법이다. 이 방법[3]은 이스라엘 히브루대학 Shai Avidan and Ariel Shamir에 의해 개발되었으며 특정 픽셀의 중요도는 contrast 정보를 이용한다. 리사이징 방식을 단계별로 살펴보면 1단계에서는 이미지 에너지정보로부터 에너지 맵을 구한다. 에너지가 큰 영역은 이미니 details 를 많이 포함하고 있으므로 중요도가 높은 픽셀들이다. 2단계에서는 에너지맵에서 에너지가 가장 작은 영역을 가로지는 층을 찾는다. 만약 수평으로 줄이거나 늘일 경우 수평으로 연결된 최저의 에너지층 부근 영역을 줄이거나 늘이면 높은 에너지를 가진 영역에서는 원 영상의 변화없이 리사이징을 할 수 있게 된다.

IV. HSV 기반 영상 변환 알고리즘

가. 개요

시각 기반의 영상 리사이징 변환은 시각적으로 민감한 영역에서는 블럭 크기의 변화를 최소화하고 둔감한 영역에서 크기의 변화를 크게 하는 방식이다. 기존 DCT 리사이징 방식은 시각적인 중요도를 고려하지 않고 모든 블럭에 동일한 크기 변화를 함으로써 리사이징 이후에 모양의 변화가 크게 느껴지게 된다.



[그림 4. 리사이징 개념도]

이 방식은 위의 그림과 같이 응시점을 기준으로 한 상대적인 민감도 특성을 구하고 그 특성에 반비례로 블럭의 크기를 결정하는 방식이다. 즉 민감도가 큰 중심영역에서는 블럭크기를 원영상과 동일한 크기로 변환하고 민감도가 작은 가장자리 영역에서는 원영상 크기보다 큰 사이즈($n+a$, $a>0$)로 변환하게 된다.

나. 알고리즘

1) 망막 이심율에 따른 민감도 특성 함수

망막 이심율에 따른 contrast 민감도 함수는 많은 실험이 이루어졌으며 본 논문에서는 W.S. Geisler and J. S. Prerry에 의한 모델을 따른다. 이 모델은 공간주파수와 이심율에 따른 민감도 특성을 모델링한 함수이다. 식(5)에서 $CT(f,e)$ 는 시각적으로 contrast를 구분할 수 있는 임계치 특성 곡선이다. 이 임계치 이하의 영역에서는 contrast를 구분할 수 없다.

$$CT(f,e) = CT_0 \exp(\alpha \cdot f \cdot \frac{c+c_2}{c_2}) \quad (5)$$

f: 공간주파수

e: 이심율

CT_0 : minimal contrast 임계치

α : 공간주파수 감쇄 상수

C_2 : half-Resolution 이심율 상수

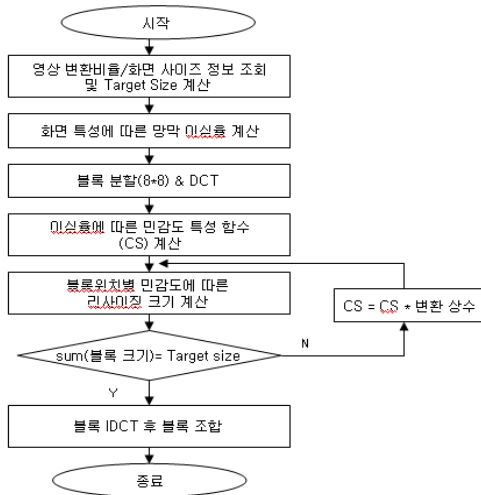
실제 블럭별 리사이징 크기를 찾기 위한 contrast 민감도 함수는 식(5)의 Inverse 함수로 식(6)과 같다. 식(6)은 응시점에서 높은 민감도 특성을 갖게 되며 이심율이 높아지수록 지수함수에 반비례하여 감쇄하게 된다. 감쇄함수 α 의 크기에 따라 감쇄율이 달라지게 되며 또한 공간주파수가 클수록 감쇄율이 높아져 이심율과 공간주파수에서 고주파일 경우 시각적인 영향이 매우 적다.

$$CS(f,e) = \frac{1}{CT(f,e)} \quad (6)$$

2) 리사이징을 위한 순서도

제안 방식은 1단계로 원영상을 변화 비율과 영상을 표시할 단말의 화면 크기 정보를 설정한다. 화면 크기를 통해 최대 이심율 크기를 구하고 블럭의 위치에 따른 민감도를 계산하기 위함이다. 2단계로는 원 영상을 8*8로 분할하고 이차원 DCT를 수행한다. 그리고 블럭별로 이심율을 계산한다. 3단계로 블럭별로 이심율에 따른 민감도를 계산하고 민감도의 역수로 블럭 사이즈를 결정한다. 4단계로 블럭별의 크기를 모두 합하여 최종 목표 타겟과 크기 비교를 한다. 만약 크기가 다르다면 민감도 함수를 재계산한다. 재계산 방법은 민감도에 변환 상수를 곱하여 민감도를 높이거나 낮추는 방법이다. 이를 통해 블럭별 크기가 재조정되고 재조정된 블럭의 크기가 최종 타겟 사이즈와 동일하다면 반복을 멈추게 된다.

마지막 5단계로 블럭별로 정해진 사이즈에 따라 크기가 $n>8$ 보다 크다면 제로 패딩을 하고 IDCT를 수행한다. 그리고 이 블럭들을 모두 재구성하여 리사이징된 영상을 구한다.

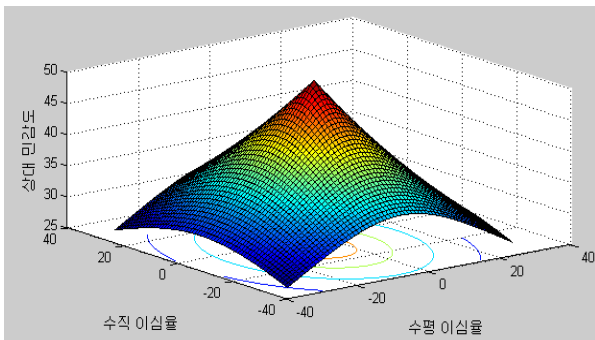


[그림 6. 리사이징 순서도]

다. 시뮬레이션 및 결과

1) 테스트 영상 및 파라미터

Test 영상은 Lena, Boat, Tank 영상을 이용하였으며 크기는 512*512 사이즈를 이용하였다. Lena 영상은 가장 보편적으로 시험이 이용되는 영상이고 Boat는 에지 성분이 많은 영상이고 Tank영상은 중심부의 탱크와 배경으로 구성된 영상이다.



[그림 6. 이심률에 따른 상대적 민감도 특성]

망막 이심률에 따른 특성함수의 계수는 원 논문에서 사용한 상수를 사용하였다. 상수의 값은 $CT_0=1/64$, $C2=2.3^\circ$, $\alpha=0.106$ 이다. 그리고 f 는 공간 주파수로 3 cycle/degree를 사용하였다. 그 이유는 MSF 함수에서 시각에 민감한 공간 주파수 대역이 3 ~ 5 cycle/degree이므로 그 중 Ngan과 Rao의 두 식에서 최대 민감도와 근사값을 선택하여 시뮬레이션을 하였다.

2) 영상 특성 비교

아래 영상은 영상 사이징을 전체 화면에 균일하게 적용한 방식과 제안한 방식의 영상을 비교하였다. 결과 영상은 Park 방식에 따라 전체 영상을 균일하게 확대한 화면과 나머지는 제안된 방식에 따라 시청 거리가 2m에서 4:3 비율의 화면을 가정하고 화면에 크기 변화에 따른 리사이징된 결과 영상이다.

화면의 크기가 적을 때는 이심률이 작아 블록별로 크기의 변화가 심하지 않으므로 전체적으로 균일하게 리사이징되어 Park의 결과 영상과 차이점이 없다. 그러나 화면이 커질수록 이심률에 의한 민감특성에 따라 블록의 리사이징 크기가 달라져 중심영역에서는 원영상의 크

기를 유지하고 주변 영역으로 갈수록 영상의 블록의 확대폭이 커짐을 알 수 있다.



[그림 7 리사이징 결과 영상]

3) 영상 화질 분석

영상의 객관적인 화질 분석을 위하여 원 영상을 특정 비율로 확대하였다가 축소된 영상을 원 영상과 비교하는 방식을 사용하였다. 일반적으로 DCT 화질 평가가 영상을 축소하였다가 다시 확대하는 방식이나 제안된 방식은 영상을 대형화면으로 확대하는 방식에 관한 알고리즘으로 리사이징시 화질의 열화가 발생하지 않음을 표1를 통해 알 수 있다.

영상	1:2 확대/축소
Lena512	47.0860
Boat512	41.3096
Tank512	42.4958

[표1 원영상과 확대/축소 후의 PSNR]

IV. 결론

본 논문에서 인간 시각 기반의 DCT 기반 영상 리사이징 방식을 제안하였다. 이 방식은 DCT를 기반으로 하여 리사이징시에도 원영상에 대해 우수한 품질을 유지하면서 시각적으로 민감한 영역의 변화율을 최소화하여 우수한 영상을 얻을 수 있었다. 향후 인간 시각특성을 고려하여 Multi foveation 영상, 움직임이 있는 동영상에도 시각적 특성을 고려한 알고리즘 개발을 위해 많은 연구가 필요하다.

[참고문헌]

1. W.S. Geisler and J. S. Prerry, "A real-time foveated multiresolution system for low-bandwidth video communication", Proc, SPIE, Vol 3299, 1998
2. Zhou Wang, Alan Conrad Bovik Embedded Foveation Image Coding IEEE Transaction on Image Processing VOL10. 2001
3. Shai Avidan and Ariel Shamir, "Seam Carving for Content-Aware Image Resizing", ACM Trans. Vol 26, 2007
4. 정용관, "인간의 시각 특성을 고려한 Sub-Band부호화에서 MRF 비교에 관한 연구", 전자공 Vol 27. 1990