시각 인지 기반의 영상 품질 측정 방법

이용헌, 서덕영
경희대학교
lyh1649@gmail.com, suh@khu.ac.kr

Human Vision Perception base Visual Quality Assessment

Yonghun Lee, Doug Young Suh
Kyung Hee University

요 약

본 논문에서는 인간의 시각 인지 능력(Human Visual System; HVS)을 기반으로 하는 영상 품질 측정 방법을 제안한다. 보다 구체적으로 제안하는 시각 인지 기반의 영상 품질 측정 방법은 기존의 객관적인 품질 평가 방법인 PSNR 품질 측정 방법에 1) HVS 에 기반한 foveated contrast sensitivity 을 이용하여 집중 영역으로부터 거리에 따른 인식률 저하를 반영하였으며, 이와 동시에 2) Just Noticeable Distortion(JND)을 통해 영상의 휘도 대비 차이에서 오는 인식률 차이를 반영하여 사용자가 인지하는 품질을 계산하였다. 이러한 연구 결과는 특히 대형 영상의 부호화 과정에서 양자화 및 bit allocation, 또는 기 부호화된 영상의 전송 과정에서 차등화된 오류 보호 기술 등에 활용할 수 있을 것이다.

1. 서론

최근 Ultra High Definition(UHD) 및 파노라마 영상과 같이, 인간이 인지할 수 있는 시야각을 넘는 크기의 대형 영상을 이용하여 사용자에게 몰입감 및 현장감을 제공하는 멀티미디어 서비스에 대한 연구가 활발하게 진행되고 있다. 본 연구에서는 이러한 대형 영상의 멀티미디어 컨텐츠에 대한 품질 측정에 있어 HVS 의 foveated contrast sensitivity 및 JND 를 활용함으로써 사용자가 체감하는 품질(Quality of Experience)을 측정하는 방법을 제안한다.

멀티미디어 영상의 품질 측정 방법은 객관적인 품질 측정 방법과 주관적인 품질 측정 방법으로 구분할 수 있다. 객관적인 품질 측정 방법은 품질을 측정하고자 하는 영상과 원 영상간의 차이를 이용한다. 원 영상의 활용 정도에 따라 원영상의 전체를 품질 측정에 사용하는 경우를 full reference metric 이라 하며 대표적으로 잘 알려진 Mean Square Error 및 Peak Signal to Noise Ratio 측정 방법이 있다. 주관적인 측정 방법은 다수의 사용자로부터 품질에 대한 등급을 피드백 받아 품질을 측정하는 Mean Opinion Score(MOS) 방법이 존재하며, [1]에서는 Double Stimulus Impairment Scale(DSIS), Single Stimulus Continuous Quality Evaluation(SSCQE)를 정의하고 있다. 객관적인 품질 측정 방법의 경우, 영상의 부호화 과정에서 발생되는 품질 열화 및 부호화된 영상의 전송 과정에서 발생되는 품질 열화를 품질 측정 과정에서 반영할 수 있으나, 사용자가 인지하는 품질을 반영하지 못하는 한계를 가진다. 이러한 한계를 극복하기 위한 연구로, [2]에서는 영상의 휘도 신호에 대한 HVS 의 JND 를 품질 측정에 활용하였다. JND 는 HVS 가 인지할 수 있는 최소의 픽셀 변화량을 의미한다

HVS 에서 영상을 구분하기 위해 중요한 상의 초점을 생성하는 부분은 중심와(fovea)이며, 중심와에서 최상의 화질을 만들 수 있는 시야각인 최적 시야각은 4~5 도로 연구되었다 [3]. [4]에서는 이를 바탕으로 사용자의 시청 거리가 정해졌을 때, foevated 영역으로부터 거리에 따른 영상의 대비에 대한 인식률을 정의하고, 이러한 특징을 계층화된 영상의 부호화에 활용하였다.

본 논문에서는 [2, 4]에서 연구된 시청 거리에 따른 HVS 에서의 foevated contrast sensitivity 및 JND 를 활용하는 품질 측정 방법을 제안하고자 한다. 본 논문의 구성은 다음과 같다. 2 절에서는 HVS 의 foevated contrast sensitivy 및 JND 에 대해 살펴본 후, 3 절에서는 제안하는 HVS 기반의 품질 측정 방법을 설명한다. 4 절에서는 제안하는 방법을 이용하는 품질 측정 결과를 실험을 통해 확인하며, 마지막으로 5 절에서는 본 논문에 대한 결론을 맺는다.

2. Human Visual System

본 절에서는 HVS 의 JND 및 foevation 기반의 HVS 모델에 대하여 자세히 살펴본다.

2.1 Just Noticeable Distortion

Weber-Fechner 의 법칙으로도 불리우는 JND 는 자극(stimuli)에 대한 차이를 느끼기 위한 최소의 변화량을 정의하고 있으며, 이는 기준 자극이 커질수록 차이를 느끼기 위해 필요한 자극의 크기 또한 커져야 하며, 기준 자극이 작을 경우 약간의 변화만으로도 그 차이를 느낄 수 있다는 것을

의미한다. [2]에서는 HVS 에서의 시각 자극에 대한 JND 를 다음과 같이 정리하였다.

HVS 는 영상의 휘도(luminance) 신호의 대비에 대한 민감도가 가장 높다. 특히 휘도 신호의 대비는 블록 기반의 DCT 를 이용하는 영상 압축 방법에서 각 블록의 평균 휘도 신호의 크기에 대한 비율로 정의되며, 평균 휘도 신호는 DCT 계수의 DC 컴포넌트 c_{DC} 의 값으로 대표된다. 그러므로 c_{DC} 에 따른 HVS 의 JND 는 다음과 같이 계산된다.

$$a_{lum}(c_{DC}) = \begin{cases} k_1\left(1-\dfrac{2c_{DC}}{GN}\right)^{\lambda_1}+1, \text{if } c_{DC} \le \dfrac{GN}{2} \\ k_2\left(\dfrac{2c_{DC}}{GN}-1\right)^{\lambda_2}+1, \text{otherwise} \end{cases} \quad (1)$$

k_1, k_2, λ_1, λ_2 는 상수 값이며, 각각 2, 0.8, 3, 2 이다. G 와 N 은 각각 Gray-level 의 수와 DCT 블록의 크기를 의미하는 값으로 각각 256, 8 로 한다. 수식 (1)을 이용하여 c_{DC} 에 따른 JND 를 나타내면 Fig. 1 과 같다.

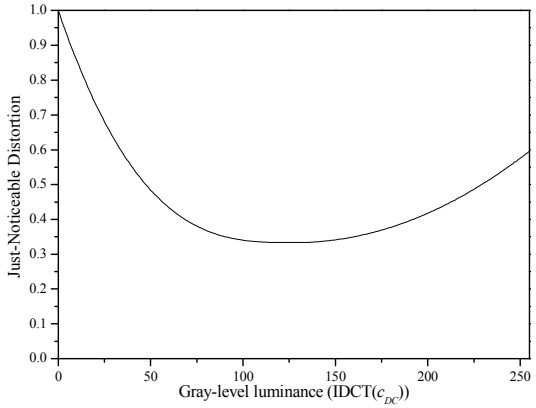


Fig. 1 Gray-level luminance 에 따른 HVS 의 JND 변화

Fig. 1 의 결과로부터 HVS 에서 휘도 신호에 대한 JND 는 평균 휘도 신호 c_{DC} 의 값이 작은 부분 및 큰 부분, 즉 블록의 평균 밝기가 어둡거나 밝을 때 커지고, 중간 밝기에서 가장 작은 것을 확인 할 수 있다. 이는 블록의 평균 밝기가 어둡거나 밝을 때는 해당 블록의 각 픽셀의 밝기 신호의 차이가 평균 밝기로부터 크지 않을 경우 인식하지 못하는 것을 의미하며, 평균 밝기가 중간일 경우 각 픽셀에서의 작은 차이도 민감하게 인식하는 것을 의미한다.

2.2 Foevation 기반의 HVS 모델

Foevation 기반의 HVS 모델은 시각-심리학적 실험을 통해 모델링된 망막에서의 편심률에 대한 대비 민감도 함수(contrast sensitivity function; CSF)를 기반으로 한다 [5]. [4]에서는 CSF 를 이용하여 망막의 편심률이 정해졌을 때, 이로부터 HVS 에서 인식할 수 있는 최대 spatial frequency 인 cutoff frequency f_c 를 다음과 같이 계산하였다.

$$f_c(e) = \frac{e_2 \ln\left(1/CT_0\right)}{\alpha(e+e_2)} \quad (2)$$

e_2, CT_0, α 는 상수 값이며, 각각 2.3, 1/64, 0.106 이다. e 는 영상의 위치에 따른 망막에서의 편심률을 의미하며 영상의 foevation 된 위치 (x_f, y_f) 로부터의 거리 d 와 영상의 크기 N 그리고 시청 거리 v 에 의해 다음과 같이 계산된다.

$$e = e(d,v,N) = \tan^{-1}\left(\frac{d}{Nv}\right), d = \sqrt{(x-x_f)^2+(y-y_f)^2} \quad (3)$$

수식 (2), (3)을 이용하여 HD 크기의 영상에 대한 spatial cutoff frequency 를 나타내면 다음의 Fig. 2 와 같다.

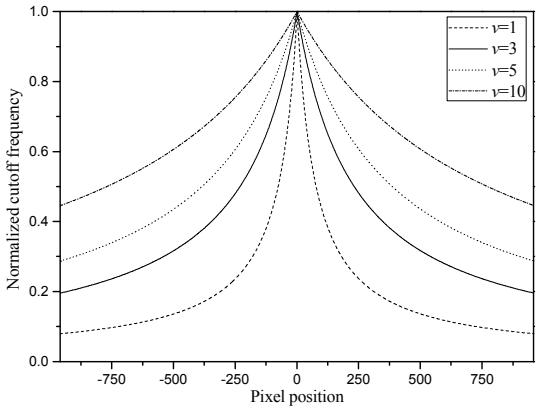


Fig. 2 시청 거리 및 픽셀 위치 변화에 따른 normalized cutoff frequency 의 변화

Fig. 2 의 결과에서 cutoff frequency 곡선의 아래 부분은 주어진 시청 환경에서 HVS 가 인지할 수 있는 공간 주파수 영역을 의미한다. 그러므로 시청거리가 고정되고, 사용자의 foevation 위치가 영상의 가운데 (e.g. HD 영상의 경우 (960, 540))인 경우에 HVS 에서 각 영상의 spatial cutoff frequency 는 foevation 위치로부터 멀어질수록 작아진다. 이는 앞서 서론에서 기술한 바와 같이 HVS 의 중심와(foevia)에서 최상의 화질을 만들 수 있는 각도에 제한이 있기 때문이며, 만약 시청 거리가 멀어질 경우(e.g. v =5, 10), 중심와에서 최적 시야각 범위 내에 맺히는 영상의 범위가 커지므로 foevation 위치로부터 멀어지더라도 cutoff frequency 가 큰 것을 확인 할 수 있다.

3. 제안하는 HVS 기반의 품질 측정 방법

본 논문에서 제안하는 HVS 기반의 품질 측정 방법은 먼저 제안하는 영상 자극 인식 모델에 대해 기술한다. 그리고 이를 바탕으로 시각 인지 기반의 품질 측정 방법을 기술한다.

3.1 제안하는 영상 자극 인식 모델

HVS 의 영상 자극 인식 모델은 앞서 2 절에서 설명한 HVS 의 두 가지 모델을 활용한다. 먼저 foevation 위치로부터의 거리에 따른 cutoff frequency 를 이용하여 거리에 따른 공간적 대비 값의 분해능(spatial contrast resolving power; SCRP)을 계산한다. SCRP 는 Fig. 2 로부터

영상의 대비에 대한 HVS 의 민감도가 foevation 위치로부터 멀어질수록 작아지는 특징을 이용한다. 휘도 신호의 대비에 대한 민감도가 작다는 의미는 해당 위치에서의 대비 값 변화를 인지하지 못하는 것을 의미한다. 반대로 휘도 신호의 대비에 대한 민감도가 가장 높은 foevation 위치에서는 대비 값의 작은 변화를 인지 할 수 있음을 의미한다. 따라서 SCRP 는 다음과 같이 계산된다. foevation 위치에서의 cutoff frequency 를 $f_{c,MAX}$ 로 하고, 이 지점에 대한 HVS 의 SCRP 를 1 로 하였을 때,

$$SCRP(e)=\left\lceil \frac{f_{c.MAX}}{f_c(e)} \right\rceil \tag{4}$$

이 때, $f_{c,MAX}=f_c(0)$ 로 계산된다. 그러므로 SCRP 를 이용하면 foevation 위치로부터의 거리에 따라 HVS 에서 구분 가능한 최소 휘도 신호의 차이를 계산할 수 있다. 다음으로 영상의 특징에 따른 HVS 의 영상 자극 인식률 차이를 반영하기 위하여 JND 를 반영한다. 앞서 2.2 절에서 기술한 영상의 블록 별 평균 밝기(c_{DC})에 따른 HVS 의 JND $a_{lum}(c_{DC})$ 를 수식 (4)에 적용함으로써 SCRP 는 다음과 같이 수정된다.

$$SCRP(e,c_{DC})=\left\lceil a_{lum}(c_{DC})\cdot\frac{f_{c.MAX}}{f_c(e)} \right\rceil \tag{5}$$

즉, 제안하는 HVS 의 영상 자극 인식 모델은 영상 신호와 관계 없이 최적 시야각으로부터 거리에 따른 영상의 대비에 대한 인식률 특성과 영상 자체의 특징(e.g. 고 대비)으로 인해 달라지는 HVS 의 인식률 차이를 동시에 반영한다.

3.2 시각 인지 기반의 품질 측정 방법

시각 인지 기반의 품질 측정을 위해 앞서 3.1 절에서 제안한 HVS 의 영상 자극 인식 모델을 활용한다. 먼저 품질 비교를 위한 원영상을 HVS 에서 인식하는 영상의 형태로 변화시키기 위해 다음과 같은 과정을 거친다.

$$l_{hvs}(n,x,y)=Round\left(\frac{l(n,x,y)}{SCRP(e,c_{DC,n})}\right)\cdot SCRP(e,c_{DC,n}) \tag{6}$$

$l(n,x,y)$ 는 영상의 휘도 신호 값을 의미한다. (n,x,y) 는 각각 n 번째 DCT 블록에 대한 인덱스, 그리고 해당 신호의 위치 좌표를 의미한다. $c_{DC,n}$ 은 n 번째 DCT 블록에 대한 DC 계수를 의미한다. 그러므로 수식 (6)을 이용하여 원 영상의 휘도 신호를 변환하면, 주어진 시청 거리에서 HVS 가 인식하는 영상을 얻을 수 있다. 다음으로 $l_{hvs}(n,x,y)$ 로 변환된 영상과 품질을 측정하고자 하는 영상과의 품질 비교는 수식 (5)의 SCRP 를 기반으로 하는 SCRP-MSE 를 활용한다. SCRP-MSE 는 다음과 같이 계산된다.

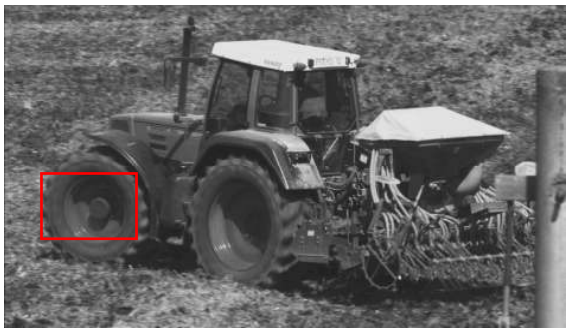
$$SCRP-MSE=\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left[\delta(i,j)\right]^2,$$

$$\delta(i,j)=\begin{cases}0, \text{ if } l_{hvs}(n,x,y)-l'(n,x,y)\le SCRP(e,c_{DC,n})\\ Round\left(\dfrac{l_{hvs}(n,x,y)-l'(n,x,y)}{SCRP(e,c_{DC,n})}\right), \text{ otherwise}\end{cases} \tag{7}$$

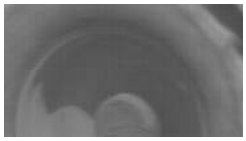
그러므로 수식 (7)의 SCRP-MSE 에서는 앞서 본 논문에서 제안하는 영상 자극 인식 모델을 반영하는 MSE 계산방법이라 할 수 있다.

4. 시뮬레이션

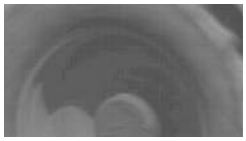
제안하는 시각 인지 기반의 품질 측정 방법에 대한 성능 검증을 위해 다음과 같은 시뮬레이션을 수행하였다. 시뮬레이션에서 가정하는 시청 환경은 시청 거리 v =3 으로 하며, 영상의 크기는 HD(1920x1080)으로 하였다. 실험 영상에 대하여 DCT 변환된 계수에 대하여 low-pass 필터를 적용하여 영상의 품질을 열화 시키고 각각 MSE 와 제안하는 SCRP-MSE 를 통해 품질을 측정하였다.



(a)



(b)



(c)

Fig. 3 (a)SCRP 를 적용한 영상; (b) SCRP 를 적용하지 않은 경우; (c)SCRP 를 적용한 경우(Tractor, 1080p)

Fig. 3 은 원본 영상에 대하여 수식 (6)을 적용한 영상을 보인다. Fig. (b), (c)와 같이 SCRP 를 적용한 경우, foevation 위치에서 벗어난 부분에서 원본 대비 품질이 저하된 것을 확인 할 수 있다.

Fig. 4 Foevation 위치 정보를 이용하여 품질을 열화시킨 영상

다음으로 실험 영상(Fig. 4)에 대하여 DCT 를 수행하고 변환된 계수에 대하여 제안하는 영상 자극 인식 모델에 기반하여 low-pass 필터를 적용하고, 이를 SCRP 를 적용한 원본 영상(Fig. 3 (a))과의 품질 비교를 위해 MSE 및 제안하는 SCRP-MSE 를 통해 품질을 측정하였다. 측정 결과 기존의 MSE 는 30.01dB 를 SCRP-MSE 를 50.17dB 를 나타내었다. 이러한 결과는 MSE 의 경우 영상의 모든 영역에 대해서 동일한 기준으로 품질을 측정하는 반면, 제안하는 SCRP-MSE 의 경우에는 foevation 위치로부터 멀어질수록 품질 열화에 대한 가중치가 작아지므로 전체적인 영상의 품질에 미치는 영향이 작아지기 때문이다.

5. 결론

본 논문에서는 대형 영상에 대한 HVS 의 시각 인지 기반 영상 품질 측정 방법을 제안하였다. 보다 구체적으로는 시청 환경에 따른 HVS 의 공간적 대비 값의 분해능 특징과 영상의 특징에 따라 달라지는 HVS 의 인식률 차이를 반영하는 영상 자극 인식 모델을 제안하였으며, 이를 이용하는 SCRP-MSE 품질 측정 방법을 제안하였다. 이를 통해 기존의 객관적인 품질 측정 방법에 제안하는 HVS 모델을 적용함으로써 대형 영상에 대한 사용자의 체감 영상 품질을 측정하는 방법을 제안하였다. 나아가 제안하는 영상 자극 인식 모델은 대형 영상의 부호화 과정에서의 양자화 계수 선택 및 bit-allocation 에 활용할 수 있을 것이며, 기 부호화된 영상의 전송 과정에서 차등화된 오류 보호 기술에도 활용될 수 있을 것이다.

감사의 글

연구는 지식경제부 및 정보통신산업진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2011-(C1090-1111-0001))

참고 논문

[1] ITU-R Rec. BT.500-11: “ Methodology for the subjective assessment of the quality of television pictures”, International Telecommunication Union, Geneva, Switzerland, 2002.

[2] N. Jayant, J. Johnston and R. Safranek, “ Signal compression based on models of human perception”, Proceedings of the IEEE, pp. 1385-1422, Oct. 1993.

[3] S. Winkler, “ Digital Video Quality: vision models and metrics”, Wiley, 2005.

[4] Z. Wang, L. Lu and C. Bovik, “ Foveation Scalable Video Coding With Automatic Fixation Selection”, IEEE Trans. on Image Processing, vol. 12, no. 2, Feb. 2003.

[5] W. S. Geisler and J. S. Perry, “ A real-time foveated multiresolution system for low-bandwidth video communication,” Proc. SPIE, vol. 3299, pp. 294-305, July 1998.