시선인식 집중도 기반의 영상 품질 측정 방법

*고정곤 **석주명 *서덕영
*경희대학교, **한국전자통신연구원
jgcool777@naver.com, jmseok@etri.re.kr., suh@khu.ac.kr

Video Quality Evaluation Method based on visual attention

*Junggon Ko **Seok Jumyung *Doug Young Suh
*Kyunghee University, **Electronics and Telecommunication Research Institute

요약

영상 서비스에서 사용자가 시청하는 비디오 화질을 측정하는 것은 사용자의 QoE(Quality of Experience)를 추정할 수 있는 중요한 작업 중 하나이다. 이를 위해 가장 널리 쓰이는 측정 방법 중 PSNR이 있다. PSNR은 원본영상과 시청영상간의 각 픽셀 값의 차이를 계산하여 화질을 측정하는 객관적인 평가 방법이다. 본 논문에서는 객관적 평가 방법인 PSNR에 시청자가 영상 측정 시 느낄 수 있는 시선 인식 집중도를 접목시킨 품질 측정방법을 제안한다. 이는 사용자의 시각적 특성이 고려되기 때문에 보다 사용자에게 맞는 영상 품질을 측정할 수 있게 된다.

1. 서론

PSNR(Peak Signal Ratio)은 영상의 품질을 측정하는 가장 널리 알려진 방법 중 하나이다. 이는 다음과 같은 식으로 표현된다.

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I(i,j)-K(i,j)]^2 \qquad (1)$$

$$PSNR = 10\times\log_{10}\left(\frac{255^2}{MSE}\right) \qquad (2)$$

식 (1)의 MSE(Mean Square Error)는 원본영상과 사용자의 시청 영상의 평균 제곱 오차로써, 영상 품질을 객관적으로 나타낸 값이라 할 수 있다. MSE가 크다는 것은 원본영상과의 오차가 커졌다는 것을 의미하며, 이는 그만큼 품질이 떨어진다는 것을 의미한다. 이 때 I(i, j)와 K(i, j)는 각각 원본 프레임과 시청 영상 프레임의 i x j번째 픽셀 값을 나타내며 M x N 이 영상 프레임의 크기가 된다. 이렇게 계산된 MSE 값을 로그화하는 방법이 식 (2)의 PSNR이다.

하지만 이 평가방법은 단순히 원본영상과 시청영상 간의 픽셀 값의 차이만을 고려한 평가방법으로 지극히 객관적일 수 밖에 없다. 하지만 사용자가 영상 시청 시 하나의 프레임에서도 중요하게 보여지는 지점이 생기게 되며 이 지점에서의 손실은 그렇지 않은 곳에서의 손실보다 더 큰 영향을 줄 수가 있게 된다.

시각인식 기반 FPSNR(Foveal Peak Signal Ratio)은 이러한 사람의 시각적 특성을 이용한 화질 측정 방법이다. 사람이 영상을 볼 때 하나의 부분에 초점을 맞추어 시청을 하게 되며 초점중심에 품질에 대한 민감도가 크게 된다는 사실을 고려하여 각 위치별 민감도에 따라 가중치를 주어 인식품질을 산출하는 화질측정방법이다.

본 논문은 좀더 나은 사용자 중심의 영상 서비스를 위해 인간의 시각적 특성을 고려한 보다 정확한 영상품질 측정 방법을 제안한다. 첫째, 사람의 시선 인식 집중도에 대해 설명한다. 둘째, 시선 인식 집중도 기반의 가중치를 구하고 이를 통해 FPSNR을 계산하는 과정에 대해 설명한다. 셋째, 이 방법을 통해 영상 품질을 측정한 simulation 결과에 대해 설명한다. 마지막으로 향후 연구과제를 도출한다.

2. 시선인식 집중도

시선인식 집중도라는 것은 '인간은 자신이 보려 하는 것만 본다'라는 인지심리학적인 특징으로 정해질 수 있다. 이는 1997년 하버드대의 젊은 심리학자 두 명이 '보이지 않는 고릴라' 실험을 통해 입증한 바 있다. 즉, 인지심리학 측면에서 보면 특정객체 혹은 물체 중심에 관심초점을 두고 있을 가능성이 높다는 것이다.

인간의 시각인식 측면에서는 객관적 품질 측정결과와 다르게 품질을 느낄수도 있다. 그림 1과 같이 동일한 PSNR이지만 인간이 느끼는 인식 품질 측면에서는 왼쪽의 이미지가 더 좋은 품질이라고 느낀다. 이러한 현상을 체감품질(Quality of Experience, QoE)이라 하며, IPTV 분야에서 계속 연구되고 있다. [1]

[그림 1] 동일한 PSNR을 갖는 다른 QoE 영상의 예

3. 제안 방법

시각인식기반 PSNR인 FPSNR(Foveal Peak Signal to Noise Ratio) 은 사람의 시각적 특성을 고려한 PSNR방법이다.[2,3] 인간의 시선인식 집중도에 따라 초점 중심에 품질에 대한 민감도가 크다는 사실을 고려하여 민감도에 따라 가중치를 주어 인식품질을 산출하는 화질 측정 방법이다.

본 논문에서는 FPSNR 측정을 위하여 필요한 사람의 시각적 특성 기반의 가중치를 다음과 같이 제안한다. 첫째, 기존의 방법으로 디코딩된 영상과 시선인식 집중도 기반으로 디코딩된 영상의 품질이 동일하여야 된다는 가정으로부터 시작한다. 이러한 가중치는 각 초점위치에 해당하는 슬라이스 크기와 인코딩된 스케일러비티 수에 따라 다를 수 있다. 시선인식 집중도에 따라 결정된 영상의 MSE를 MSE_F 라고 정의한다. MSEmin은 시선인식 집중도에 따라 선택되기 이전 최초로 디코딩된 영상의 MSE를 의미한다. MSEC는 디코딩된 현재 영상의 MSE를 의미한다. 결과적으로 식 (4)와 같이 FPSNR을 도출하기 위하여 식 (3)와 같이 FMSE를 구한다

$$\begin{cases} FMSE = MSE_{\min}, & MSE_C \le MSE_F \\ FMSE = MSE_C + |MSE_F + MSE_C|(1+w), & MSE_C > MSE_F \end{cases} \quad (3)$$

$$w = \frac{f_s(l)}{\sum_{l=0}^{l=N-1} fs(l)}, l = 0, 1, 2, \dots, N-1$$

$$FPSNR = 10 \times \log_{10}\left(\frac{255^2}{FMSE}\right) \quad (4)$$

w는 시선인식 집중도에 따른 가중치를 의미하며 최대 해상도의 공간주파수를 이용하여 결정한다. 현재 디코딩된 영상 품질을 시각 인식 기반으로 측정하기 위해서는 다음의 내용과 같이 MSE를 FMSE로 변환하는 절차를 갖는다. 이러한 변환 방식은 QoE에서 사용하는 MOS와 유사한 개념을 가지고 있다. [4]

기호	의미
MSE_F	i번째 slice에서 기준이 되는 품질의 MSE
MSE_C	디코딩된 현재영상의 MSE
MSE_{min}	최초로 디코딩된 영상의 MSE의 MSE (High Quality)
w	시선인식 집중도에 따른 가중치
$f_s(l)$	l번 계층의 최대 공간 주파수

[표 1] 각 수식별 기호 및 의미

① $FMSE = MSE_{\min}, \quad MSE_C \le MSE_F$

디코딩되어 보여지는 영상의 MSE(MSE_C)가 기준이 되는 품질의 MSE(MSE_F) 보다 작거나 같다는 것은 시각인식 집중도 측면에서 품질의 차이를 인식할 수 없는 상황이다. 즉, 디코딩 된 현재 영상의 품질이 기준 품질이상이라면 최고품질로 보여지는 것과 같은 만족감을 느끼므로 해당 위치 P(x, y) 에서의 MSEmin 가 FMSE가 된다. 여기서 P(x, y)는 픽셀단위 혹은 슬라이스 단위로 가정한다.

② $FMSE = M_{CRTi} + |M_{STDi} + M_{CRTi}|(1+w), MSE_C > MSE_F$

디코딩되어 보여지는 영상의 MSE(MSE_C)가 기준 품질의 MSE(MSE_F) 보다 크다는 것은 이 보다 품질보다 낮은 품질의 영상이 보여지고 있다는 것을 의미한다. 따라서 현재 보여지는 영상의 MSE(MSE_C)를 기반으로 FMSE가 계산되어야 한다. 참고적으로 $MSE_C > MSE_F$ 인 경우에는 동일한 양의 MSE가 발생하더라도 시선인식 집중도의 위치에 따라 사람이 느끼는 품질인식은 다르다. 시선인식 집중도가 낮은 위치의 MSE는 상대적으로 작은 값으로 생각하기 때문에 불만족감이 작다는 의미이다. 따라서 시선인식 집중도의 위치에 따라 가중치 w가 반영되어 FMSE가 계산되어야 한다. w는 스케일러빌리티에 의하여 선택 가능한 계층 l을 이용하며, $f_s(l)$은 각 l이 가지는 최대 공간주파수를 의미한다.

4. 시뮬레이션 과정 및 결과

시뮬레이션 과정은 다음과 같이 진행 하였다. 그림 1과 같이 1920 X 1080 영상을 48 X 1088 사의즈인 40개의 슬라이스로 나누어 인코딩한다. 사용자가 느끼는 시선인식 집중도는 영상의 중앙이 가장 크며 옆으로 이동할수록 집중도가 떨어진다고 가정하였다.

초점중심으로부터 멀어질수록 시선인식집중도가 줄어드는 상황변화에 비디오 품질을 적응하기 위하여 SVC를 기반으로 하는 3 계층의 공간 스케일러빌리티 인코딩을 활용하였다. 이를 위해 SVC reference software 중 가장 최신 버전인 JSVM 9.19를 사용하였다.

각 슬라이스 별로 적용한 해상도는 표 2와 같다. 시선인식집중도가 가장 큰 가운데 슬라이스는 최고 품질의 1920 X 1080 해상도를 적용하며, 중간에 위치한 슬라이스는 1280 X 720 해상도를 적용하고 끝쪽에 위치하는 슬라이스는 640 X 480 해상도를 적용하였다. 여기서 1/3로 축소된 영상은 HD, SD 크기의 영상으로, 현재 서비스되고 있는 환경과 유사하게 적용함으로써 실험결과에 대해 현실감을 높였다.

1280 X 720 영상과 640 X 480 영상의 품질 측정은 1920 X 1080 크기로 업샘플링 하여 측정하게 된다. 품질 비교를 위한 원본영상은 1920 X 1080 영상이 된다.

[그림 1] 영상의 품질적용을 위한 슬라이스 코딩의 예

적용한 해상도	슬라이스 넘버
1920 X 1080	19-22
1280 X 720	16-18, 23-25
640 X 480	1-15, 26-40

[표 2] 각 슬라이별 적용 해상도

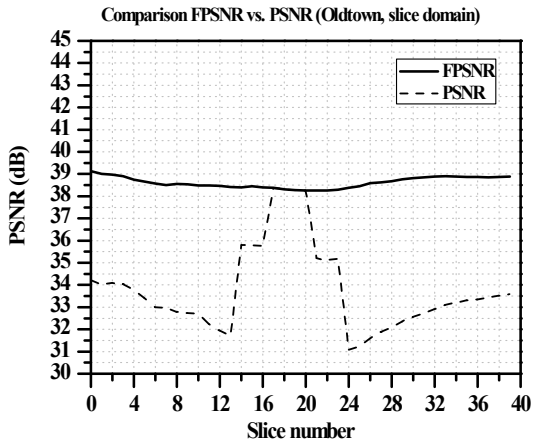
사용자는 슬라이스 19-22번의 위치는 1920 X 1080의 품질, 16-18, 23-25 번의 위치에서는 1280 X 720의 품질, 1-15, 26-40번의 위치에서는 640 X 48의 품질로 만족감을 느낄 수 있다고 가정하였다. 즉, 16-18번이나 1-15번 슬라이스가 1920 X 1080의 해상도로 서비스가

각 슬라이스 별로 적용한 해상도는 표 2와 같다. 시선인식집중도가 가장 큰 가운데 슬라이스는 최고 품질의 1920 X 1080 해상도를 적용하며, 중간에 위치한 슬라이스는 1280 X 720 해상도를 적용하고 끝쪽에 위치하는 슬라이스는 640 X 480 해상도를 적용하였다. 여기서 1/3로 축소된 영상은 HD, SD 크기의 영상으로, 현재 서비스되고 있는 환경과 유사하게 적용함으로써 실험결과에 대해 현실감을 높였다.

1280 X 720 영상과 640 X 480 영상의 품질 측정은 1920 X 1080 크기로 업샘플링 하여 측정하게 된다. 품질 비교를 위한 원본영상은 1920 X 1080 영상이 된다.

각 해상도 별 최대 공간주파수는 다음과 같이 적용하였다.

해상도 1080p가 갖는 공간주파수를 30cpd, 축소된 비율만큼 20 cpd, 10 cpd를 갖게 된다고 적용하였다.



[그림 3] 실험영상의 PSNR과 FPSNR의 비교 (슬라이스 도메인)

그림 3은 제안 방법에 따라 얻은 실험결과를 나타낸 그래프이다. PSNR측면에서 바라보면 양 사이드의 영상의 품질은 3-4 dB가 떨어지기 때문에 기존 영상보다 품질 열화가 있다고 판단할 것이다. 그러나 사람의 시선인식 집중도를 고려하여 FPSNR로 측정하게 되면 최고품질의 PSNR과 동일한 값으로 유지되는 것을 알 수 있다. 즉, 객관적 품질 평가와 주관적 품질 평가를 동시에 고려할 수 있다는 장점이 있다.

5. 결론 및 향후 연구 계획

시선 인식 집중도에 따라 각 슬라이스의 품질의 중요도를 고려하여 SVC(Scalable Video Coding)에 적용한다면, 기존의 최고 품질만으로 영상을 서비스하는 것보다 비트율은 절감되나 시청자 입장에서는 이를 크게 느끼지 못하는 영상을 전송할 수 있을 것이다. 비트율 절감은 영상을 서비스하는 시간을 그만큼 단축시킬 수 있다는 의미이며, 이는 영상을 서비스 하는데 생기는 delay를 줄일 수 있다는 의미이다. 향후 영상의 각 위치별 영상의 품질을 달리 적용한 비디오 서비스 시 본 논문의 품질 측정방법을 적용한다면 보다 사람의 시각에 맞는 영상품질을 측정할 수 있을 것이라 생각한다.

6. 감사의 글

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2011-(C1090-1111-0001)).

7. 참고 문헌

[1] E. P. Ong, W. Lin Z. Lu, S. Yao, and M.H. Loke, “PERCEPTUAL QUALITY METRIC FOR H.264 LOW BIT RATE VIDEOS”, IEEE, 2006

[2] S. Lee, M.S. Pattichis, A.C. Bovik, “Foveated video quality assessment,” IEEE Trans. on Multimedia, vol.4 , issue 1 , pp.129 - 132 , 2002

[3] Mario Vranješ, Snježana Rimac-Drlje, Ognjen Nemi, “Influence of Foveated Vision on Video Quality Perception”, 51st International Symposium ELMAR-2009, pp. 29-31, 2009

[4] T.Brandao and M.P.Queluz, “No-reference perceptual quality metric for H.264/AVC encoded video,” IEEE Trans. CSVT, vol.20, no.11, pp.1437-1447, Nov. 2010