

일반논문-07-12-2-11

멀티미디어 응용을 위한 주관적 동영상 품질평가 방법의 비교분석

최지환^{a)}, 정태욱^{a)}, 최현수^{a)}, 이은재^{b)}, 이상욱^{a)}, 이철희^{a)†}

Comparison of subjective video quality assessment methods for multimedia applications

Jihwan Choe^{a)}, Taeuk Jeong^{a)}, Hyun Soo Choi^{a)}, Eunjae Lee^{a)}, Sang Wook Lee^{a)}, and Chulhee Lee^{a)†}

요약

본 논문에서는 멀티미디어 응용을 위한 동영상의 주관적 품질평가방법으로 널리 사용되는 DSCQS(Double Stimulus Continuous Quality Scale method) 및 ACR(Absolute Category Rating)방법을 비교분석하였다. 두 가지 방법을 비교 실험한 결과 MOS(Mean Opinion Score) 값의 경우 DSCQS와 ACR의 평가치가 매우 높은 상관도를 보이는 것을 확인할 수 있었다. 또한 DMOS(Difference Mean Opinion Score)값의 경우에는 DSCQS와 ACR방법이 MOS값의 경우보다 약간 낮은 정도의 상관도를 보였다. 이 같은 결과는 ACR방식이 DSCQS에 필적하는 성능을 제공하며 많은 수의 영상을 평가할 수 있는 효율적인 주관적 품질평가 방법임을 보여주고 있다.

Abstract

In this paper, we compared two subjective assessment methods DSCQS(Double Stimulus Continuous Quality Scale method) and ACR(Absolute Category Rating). These methods are widely used in order to evaluate video quality for multimedia application. We performed subjective quality tests using DSCQS and ACR methods. The subjective scores obtained by the DSCQS and ACR methods show that these methods are highly correlated in terms of MOS(Mean Opinion Score) and have slightly lower correlation in terms of DMOS(Difference Mean Opinion Score). The results indicate that ACR method is an effective subjective quality assessment method, which shows compatible performance with DSCQS method and can evaluate a larger number of video sequences.

Keyword - video quality assessment, DSCQS, ACR, MOS, DMOS

I. 서론

최근 다양한 멀티미디어 서비스가 증가하면서 멀티미디어

서비스의 품질측정에 대한 요구가 증가하고 있다. 특히 동영상 품질평가는 멀티미디어 서비스의 핵심인 동영상 서비스의 품질 기준을 마련하는데 사용됨으로써 서비스의 차등화, 전송 매체의 최적화, 멀티미디어 기기의 품질 인증 등 다양한 분야에 응용될 수 있다. 이 같은 이유로 다양한 동영상의 품질을 측정하기 위한 객관적 화질평가모델 개발에 대한 연구가 널리 이루어지고 있다^[1-5].

a) 연세대학교 전기전자공학과
Dept. of Electrical and Electronic Engineering, Yonsei University
b) 삼성전자(주)
Samsung Electronics Co., Ltd.
† 교신저자 : 이철희(chulhee@yonsei.ac.kr)

현재 국제전기통신연합(ITU: International Telecommunication Union)과 VQEG(Video Quality Expert Group)를 중심으로 멀티미디어응용에 대한 객관적 화질평가 모델의 국제 표준화가 진행 중에 있다^{6, 9)}. 최근 이 과정에서 주관적 화질평가방법의 선정이 논점으로 부각되었다. VQEG Phase II Test에서는 DSCQS 방법을 사용하였으나¹³⁾, Multimedia Test에서는 많은 논의 후에 ACR-HRR 방법이 채택되었다⁶⁾. 주관적 화질평가 방법은 다수의 평가자가 직접 동영상을 보고 화질을 평가하는 방법으로서 인간의 화질 인지특성을 반영할 수 있는 가장 정확한 방법으로 널리 사용되고 있다³⁾. 또한 주관적 화질평가는 객관적 화질평가 모델의 검증에 위한 필수적인 방법으로서 객관적 화질평가 모델의 성능기준이 된다. 즉, 객관적 화질평가모델이 주관적 화질평가 결과를 정확히 예측할수록 객관적 화질평가모델의 성능이 우수하다고 볼 수 있다. 따라서 표준화 과정에서 인간의 화질인지를 효과적으로 반영할 수 있는 주관적 평가방법을 사용함으로써 이를 바탕으로 만들어지는 객관적 화질평가 모델의 성능을 담보할 수 있다.

주관적 화질평가는 그 방법 및 목적에 따라 다양한 방법이 존재한다. 먼저 원본영상과 처리영상의 상대적 화질을 측정하기 위한 주관적 화질평가 방법의 하나로서 DSCQS (Double Stimulus Continuous Quality Scale method)방법이 있다⁷⁾. 이 방법은 원본영상과 처리영상을 두 번 반복해서 보여준 후 평가자에게 화질의 좋고 나쁨을 묻는 방법으로서 매우 정확한 주관적 평가방법으로 알려져 있으며, 디지털 방송을 위한 객관적 화질평가 방법의 국제 표준화 과정에서도 사용되었다^{1,2,12,13)}. 그러나 DSCQS방법은 정확도가 매우 높은 반면 원본영상과 처리영상을 각각 2회 재생해야 하기 때문에 평가 시간이 길어지고 많은 수의 동영상

평가할 수 없다는 단점이 있다. 또 다른 주관적 화질 평가 방법으로서 ACR(Absolute Category Rating)이 있다⁸⁾. 이 방법에서는 원본영상과 처리영상을 무작위로 섞어서 각각의 영상을 한 번만 보여주며, 평가자는 각 영상에 대한 품질을 5가지(매우 좋음, 좋음, 보통, 떨어짐, 나쁨)중 하나로 평가한다. 즉 영상을 1회만 재생함으로써 짧은 시간 내에 많은 수의 동영상을 평가할 수 있다.

주관적 화질평가실험을 위해서는 다수의 평가자를 모집하고 다양한 종류의 동영상을 확보해야하는 과정이 필요하다. 또한 이 과정에서 많은 시간과 비용이 요구된다. 이 같은 이유로 DSCQS와 ACR방식의 장단점이 잘 알려져 있음에도 불구하고 최근 표준화 과정에서 두 방식의 채택에 대한 논란이 일어난 이후에 비로소 이들 방법의 상관도에 대한 구체적인 비교 연구가 시작되고 있다. 이에 본 논문은 DSCQS와 ACR방법의 비교실험을 통해서 멀티미디어 화질평가에 사용되는 ACR방법의 신뢰도와 효율성을 검토하고 국제 표준화 과정의 주관적 평가방법 선정에 실험적인 근거를 제공한다.

II. 실험 환경 및 절차

1. 주관적 평가방법

본 논문에서는 두 가지의 주관적 평가 방법을 비교한다. 먼저 DSCQS방법은 그림 1에서와 같이 평가자들에게 각각 원본영상과 처리영상을 연속해서 두 번씩 보여준다. 영상의 제시순서는 “원본영상(A)→처리영상(B)→원본영상(A*)→처리영상(B*)”이 되거나 “처리영상(A)→원본영상

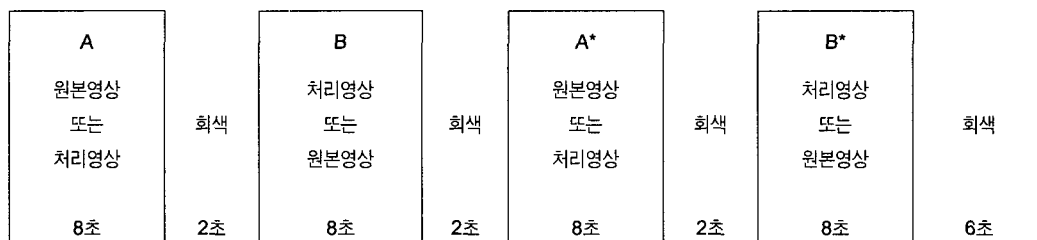


그림 1. DSCQS방법의 실험 동영상 제시 과정
Fig. 1. Presentation of video sequences for the DSCQS method

(B)→처리영상(A*)→원본영상(B*)”의 순서가 되는데 각각의 동영상에 대해 두 가지 방법 중의 하나를 무작위로 선택하여 사용한다. 또한 각 영상의 사이에는 회색영상이 2초간 나타나고, 마지막 영상이 제시된 후 6초간 회색 영상이 나타난다. 평가자는 두 번째 영상(B)의 제시가 끝난 후부터 두 영상에 대한 화질 평가 점수를 그림 2에 있는 5단계(매우 좋음, 좋음, 보통, 떨어짐, 나쁨)로 구분된 연속 측정 평가표(continuous grading scale)에 펜으로 직접 표시한다. 평가표에서 각 눈금은 각각 20단위로서 가장 아래에 표시된 선은 0점, 가장 위쪽의 선은 100점으로 환산된다.

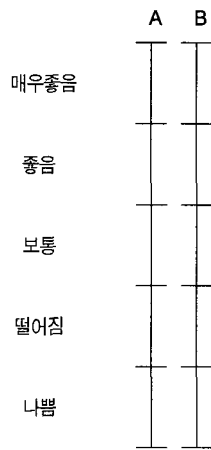


그림 2. DSCQS 연속 측정 평가표
Fig. 2. DSCQS continuous quality scale

반면 ACR방법은 원본영상과 처리영상을 구분하지 않고 모든 테스트영상을 무작위로 섞은 후 그림 3에서처럼 각 영상을 1번씩만 보여주고 각각에 대한 품질을 5단계(매우 좋음, 좋음, 보통, 떨어짐, 나쁨)로 평가하는 방식이다. 즉, ACR방법에서는 DSCQS와 달리 원본영상과 처리영상을 직접적으로 비교하는 과정이 없다. 또한 ACR은 DSCQS처럼 연속적인 값으로 동영상을 평가하지 않고 5단계의 불연속 값으로 영상을 평가한다. 즉, 평가자는 각각의 영상에 대한 품질을 그림 4와 같은 평가표에서 해당 빈칸에 표시하는 방식으로 영상을 평가한다.

일반적으로 DSCQS는 처리영상을 원본영상과 직접 비교함으로써 상대적인 품질 측정에 적합한 방법이고, ACR은 영상의 절대적 품질 측정에 적합하다. 그러나 ACR방식

을 적절히 설계함으로써 영상의 상대적 품질측정이 가능하다. 이러한 방식을 ACR-HRR(Hidden Reference Removal)이라고 한다. ACR-HRR과 ACR의 실험 방법은 동일하다. 그러나 ACR-HRR의 경우 처리영상과 그에 상응하는 원본영상을 무작위로 섞어 놓고, 원본영상 화질측정치와 처리영상 화질측정치의 차이를 구하여 상대적 품질을 측정할 수 있다.

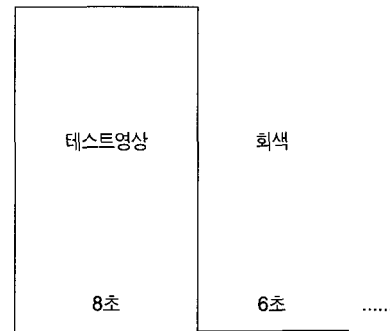


그림 3. ACR 테스트 비디오 재생 과정
Fig. 3. Presentation of video sequences for the ACR method

영상 번호	1	2	3	4	5
매우 좋음					
좋음					
보통					
떨어짐					
나쁨					

그림 4. ACR 방식의 주관적 평가표
Fig. 4. Score sheet for the ACR method

2. 실험실 설정

주관적 평가 실험은 그림 5에서처럼 다섯 명의 평가자가 각각에게 배정된 LCD 모니터를 통해서 동영상을 시청하도록 설계되었다. 실험에는 17인치 LCD 모니터를 사용하였고, 평가자들에게 독립된 환경을 제공하기 위해서 높이 1m의 칸막이로 각각을 구분하였다. 영상은 DVI(Digital

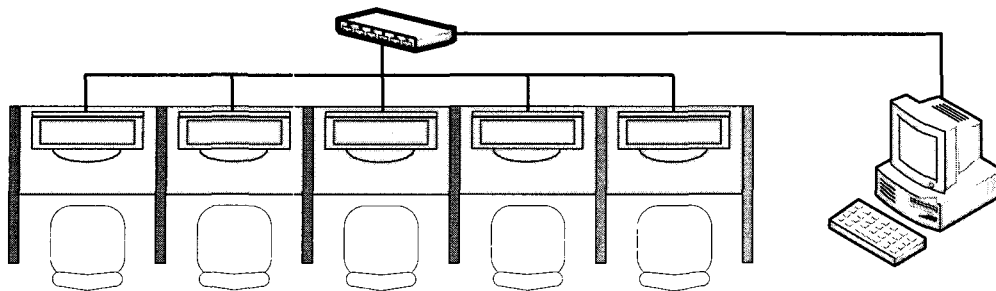


그림 5. 주관적 화질평가 실험실
Fig. 5. Test Environment

Visual Interface) 분배기를 통해 모든 모니터에서 동일하게 재생되었다. 평가자가 직접 보게 되는 화면은 그림 6과 같다. 영상은 화면의 중앙에 위치하고 영상의 배경화면은 회색(Red: 128, Green: 128, Blue: 128)으로 채워져 있다.

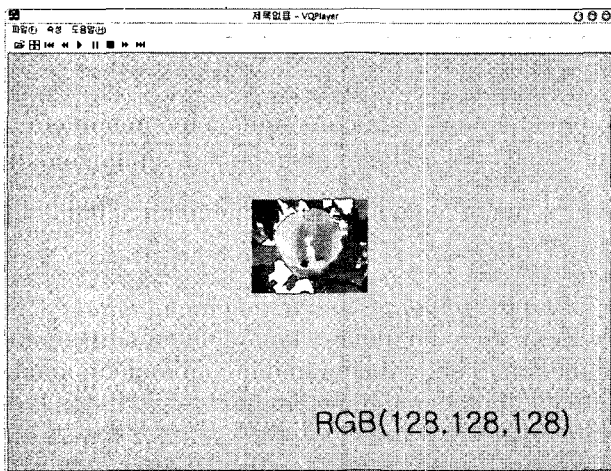


그림 6. LCD 모니터로 재생되는 화면
Fig. 6. Display of the test video in the LCD monitor

4. 실험에 사용된 영상

본 실험에서는 한국방송공사(KBS)에서 제공하는 다양한 종류의 영상을 대상으로 DSCQS 및 ACR방법으로 주관적 평가를 실시함으로써 각각의 주관적 평가방법을 비교하였다. 먼저 뉴스, 스포츠, 드라마 등 다양한 종류의 프로그램에서 화면의 움직임, 밝기, 복잡도 등 다양한 특성이 반영되도록 원본영상을 선택하였다. 이들 원본영상은 NTSC 형식

으로 기록되었으나, 본 실험에서는 QCIF(Quarter Common Intermediate Format) 형식으로 변환하였다. 변환된 원본영상에 대하여 MPEG-4와 H.264 부호화기를 적용해서 낮은 품질에서부터 높은 품질의 영상이 고르게 포함되도록 처리영상을 제작하였다^[10, 11]. 이들 원본영상과 처리영상을 합해서 총 180개의 영상을 대상으로 ACR방식의 주관적 평가를 실시하였다. 이들 180개의 영상 중에는 원본영상이 반드시 포함되도록 해서 HRR(Hidden Reference Removal) 기법을 이용한 상대적 화질측정이 가능하도록 하였다. 또한 ACR방법에서 사용된 180개의 영상 중에서 총 64개의 원본-처리영상 쌍(pair)을 무작위로 추출하여 DSCQS방식으로 주관적 평가실험을 실시하였다. 본 실험에서는 모니터 사이의 편차를 최소화하기 위해서 동일 제조사, 동일 모델의 LCD 모니터를 사용했다.

5. 평가자

본 실험에서는 DSCQS방법의 경우 19명, ACR방법의 경우 20명이 참여하였다. 이들은 평가자는 모두 비전문가들로 구성되었고, 교정시력이 모두 1.0이상이며 색맹이나 색약자는 제외되었다. 또한, 지시사항을 이해할 수 없거나 유효한 평가를 할 수 없는 사람은 배제하였다.

III. 실험 결과

각 동영상의 화질 점수는 MOS(Mean Opinion Score)라

는 수치로 환산된다. 이는 각 동영상에 대해 평가자가 부여한 평가 점수의 평균을 취한 것으로서 동영상의 절대적 화질을 나타내는 수치가 된다. 또 다른 화질 수치로서 원본영상과 처리영상의 상대적 화질측정에 사용되는 DMOS (Difference Mean Opinion Score)가 있다. 이는 평가자들이 부여한 동영상에 대한 점수들의 차이(Difference)값의 평균을 의미한다. 즉 원본영상의 MOS값(MOSr)에서 처리영상의 MOS값(MOSp)의 차이가 처리영상의 DMOS값이 된다.

$$DMOS = MOS_r - MOS_p$$

예를 들어 원본영상과 처리영상 사이에 화질차이가 거의 없다면 DMOS 값은 0 근처의 값을 갖게 되고, 화질 차이가 커질수록 DMOS 값도 커지게 된다. 본 실험에서는 DSCQS 와 ACR 방식을 비교하기 위하여 DMOS값과 MOS값을 모두 사용하였으며, 각 방식의 실험 결과를 합산하여 상호상관계수(Pearson Correlation Coefficient)를 구하였다. 두 변량 X, Y 사이의 상호상관계수 $\rho_{X, Y}$ 는 다음과 같이 정의된다.

$$\rho_{X, Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

여기서 $cov(X, Y)$ 는 X와 Y의 공분산을 의미하며, σ_X 와 σ_Y 는 각각 X와 Y의 표준편차를 의미한다. 상호상관계수는 두 변량 사이의 상관관계의 정도를 나타내는 수치로서 화질평가에서 평가측도로 사용된다^{8, 12, 13}. 상호상관계수가 1에 가까울수록 두 변량사이의 상관도가 높다는 것을 뜻하며, 본 실험결과에서 각각의 방식에서 산출된 자료 값들의 상관도가 1에 가깝다는 의미는 두 방식의 주관적 평가 결과가 매우 유사하다는 것을 의미한다.

먼저 절대적 화질측정치를 비교하기 위해서 DSCQS 방식의 실험결과에서 MOS값을 구한 후 ACR방식의 결과에서 구한 MOS값과 비교하였다. 그 결과 그림 7(a)에서처럼 0.96이상의 높은 상관계수를 얻을 수 있었다.

다음으로 두 방식의 상대적 화질측정 결과를 비교하기 위해서 HRR(Hidden Reference Removal)기법으로 DMOS 값을 구하였다. ACR-HRR 방식에서는 무작위로 섞인 영상들 사이에 반드시 원본영상이 포함되어있기 때문에 처리영상의 MOS값과 숨겨진 원본영상의 MOS값의 차이를 구함으로써 DMOS를 얻을 수 있다. 그림 7(b)는 이렇게 얻어진 ACR방식의 DMOS값과 DSCQS방식의 DMOS값의 분포를 보여주고 있다. 두 방식의 DMOS값을 비교했을 경우 상호상관계수는 0.924로 MOS값을 비교했을 경우에 비해서 상호상관도가 다소 낮음을 알 수 있다.

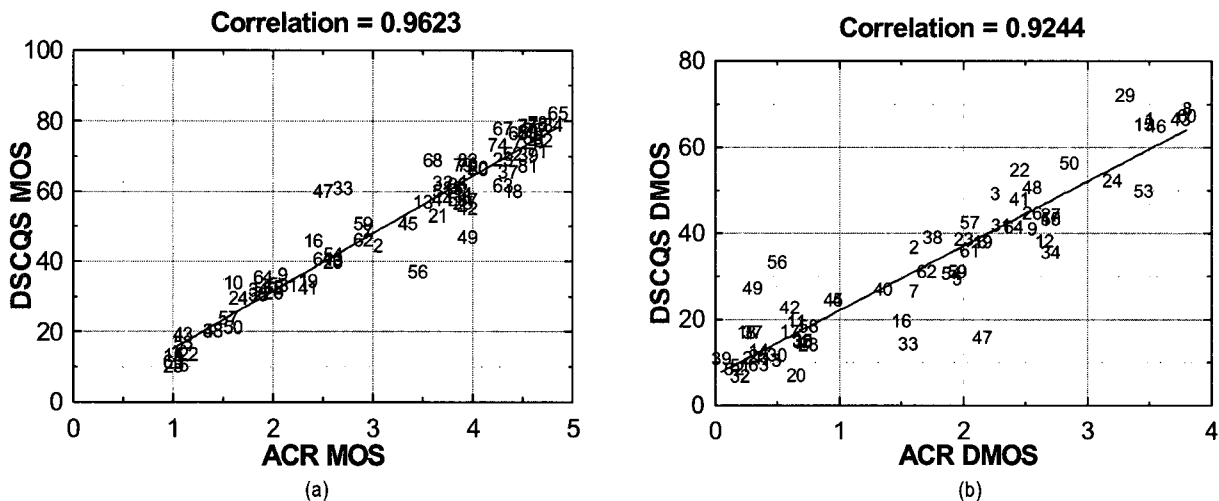


그림 7. ACR 방법과 DSCQS 방법의 화질측정치 비교. (a) MOS, (b) DMOS
 Fig. 7. Comparison between the ACR quality scores and the DSCQS quality scores. (a) MOS, (b) DMOS

또한 이처럼 상호상관도가 더 낮다는 것은 그림 7(b)에서 볼 수 있듯이 DMOS의 경우 선형회귀선에서 멀리 떨어져 있는 분포값이 MOS의 경우보다 많은 것을 의미한다. 즉, DSCQS 방식으로 얻은 화질측정값과 ACR-HRR방식에 의한 값들 사이에 차이가 존재한다. 예를 들어 47번 동영상의 경우 그림 7(a)에서 선형회귀라인의 위쪽에 위치한다. 이는 DSCQS방식으로 측정된 절대적 화질측정치(MOS)가 ACR방식으로 측정된 것 보다 더 높다는 것을 의미하며, DSCQS방식에서 화질이 더 좋다고 인식되었음을 보여준다. 또한 동일한 동영상이 그림 7(b)에서는 선형회귀라인의 아래쪽에 위치한다. 이는 DSCQS방식으로 측정된 상대적 화질측정치(DMOS)가 ACR방식에 의한 측정값보다 낮다는 것을 의미하며, DSCQS방식으로 측정할 경우 원본영상과 처리영상의 차이가 더 작게 인식되었다는 것을 보여준다. 그러나 7(b)에서의 분포 값이 7(a)의 경우보다 회귀라인에서 더 멀리 있음을 관찰할 수 있다. 29번 동영상의 경우 그림 7(a)에서는 회귀라인에 매우 근접하고 있으나 그림 7(b)에서는 회귀라인의 위쪽에 위치한다. 즉, MOS의 경우 두 방식의 화질측정이 유사한 결과를 보여주지만, DMOS의 경우 DSCQS방식이 ACR-HRR보다 같은 동영상에 대하여 화질이 더 나쁘다고 평가하고 있다. 이 같은 현상은 평가자가 해당 영상의 전에 위치하는 영상의 영향을 받기 때문으로 분석될 수 있다. 즉, HRR기법의 특성상 처리영상과 원본영상을 직접비교하지 않고 영상들의 사이에 원본영상을 숨겨둠으로써 앞에 오는 영상들의 화질에 따라 해당 영상의 평가가 영향을 받을 수 있다. 또한 원본영상과 처리영상의 직접 비교 시에 인지될 수 있는 열화요인이 처리영상이 원본영상과 멀리 떨어져서 제시되는 경우에는 인지되지 않는 현상이 발생할 수 있다. 그러나 이 같은 다양한 요인은 상호작용하며, 또한 평가자 개개인의 선호도의 편차가 존재하고 동일한 평가자의 경우에도 매 실험에서 다른 결과를 나타낸다는 점을 고려하면 각각의 요인들이 주관적 화질평가에 미치는 영향을 정량적으로 분석하는 것에는 한계가 있다.

IV. 결론

본 논문에서는 현재 널리 사용되고 있는 DSCQS 및 ACR방식의 주관적 화질평가방법을 비교분석하였다. DSCQS방식은 원본영상과 처리영상의 직접적인 비교를 통해 영상의 상대적 화질을 측정하는 방법으로서 정확도가 매우 높은 방법이다. 반면 ACR은 원본과의 비교과정 없이 영상의 절대적 화질을 측정하는 방법으로서 짧은 시간에 많은 수의 영상을 평가 할 수 있다. 본 논문에서는 ACR방법에 의한 화질측정 방법과 DSCQS방법을 직접 실험을 통해 비교 분석하였다. 그 결과 MOS값의 경우 두 가지 방법이 매우 높은 상관도를 보이는 것을 확인할 수 있었다. 반면 DMOS값의 경우 두 가지 방식의 상관도가 MOS의 경우보다 약간 떨어짐을 알 수 있다. 그러나 이 경우에도 두 방식 사이에 높은 수준의 상호상관도를 보여주고 있다. 즉, ACR방식의 경우 정확한 화질측정치를 제공하면서도 짧은 시간 동안 많은 수의 영상을 평가 할 수 있어서 매우 효율적이며 신뢰도가 높은 방법임을 알 수 있다. 이러한 결과는 ACR방법이 멀티미디어 화질평가의 국제 표준화 과정에서 적절히 사용될 수 있음을 실험적으로 보여주고 있다.

참고 문헌

- [1] ITU-R, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference," Recommendation ITU-R BT.1683, 2004.
- [2] ITU-T, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," Recommendation ITU-T J.144, 2004.
- [3] Z. Wang, HR Sheikh, and AC Bovik, Objective video quality assessment, in *The Handbook of Video Databases: Design and Applications* (B. Furht and O. Marques, eds.), CRC Press, 2003.
- [4] Video Quality Experts Group, "Video Quality Experts Group: Current Results and Future Directions," 2000.
- [5] S. Olsson, M. Stroppiana and J. Baina, "Objective methods for assessment of video quality: state of the art," *IEEE Trans. Broadcasting*, vol.43, no.4, pp.487-495, 1997.
- [6] Video Quality Experts Group, "Multimedia Group Test Plan," Ver. 1.16, 2007.
- [7] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Recommendation ITU-R BT.500-11, 2002.
- [8] ITU-T, "Subjective video quality assessment methods for multi-

media applications," Recommendation ITU-T P.910, 2002.

[9] <http://www.its.bldrdoc.gov/vqeg/>

[10] ITU-T, "Advanced video coding for generic audiovisual services," Recommendation ITU-T H.264, 2005.

[11] ISO/IEC, "Coding of Audio-Visual Objects: Visual," ISO/IEC Standard 14496-2, 1998.

[12] Video Quality Experts Group, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment," 2000.

[13] Video Quality Experts Group, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, Phase II," 2003.

저 자 소 개



최 지 환

- 1999년 2월 : 연세대학교 전자공학과 졸업 (학사)
- 2002년 2월 : 연세대학교 전기전자공학과 졸업 (석사)
- 2002년 ~ 현재 : 연세대학교 전기전자공학과 박사과정
- 주관심분야 : 영상신호처리, 비디오 화질평가



정 태 욱

- 1996년 2월 : 연세대학교 수학과 졸업 (학사)
- 1998년 2월 : 연세대학교 수학과 졸업 (석사)
- 2003년 8월 : 연세대학교 전기전자공학과 졸업 (석사)
- 2003년 ~ 현재 : 연세대학교 전기전자공학과 박사과정
- 주관심분야 : 영상신호처리, 비디오 화질평가



최 현 수

- 2004년 2월 : 연세대학교 전기전자공학과 졸업 (학사)
- 2006년 2월 : 연세대학교 전기전자공학과 졸업 (석사)
- 2006년 ~ 현재 : 연세대학교 전기전자공학과 박사과정
- 주관심분야 : 영상신호처리, 비디오 화질평가



이 은 재

- 2004년 2월 : 연세대학교 전기전자공학과 졸업 (학사)
- 2006년 2월 : 연세대학교 전기전자공학과 졸업 (석사)
- 현 삼성전자 연구원
- 주관심분야 : 영상신호처리

 저 자 소 개

**이 상 욱**

- 2004년 2월 : 연세대학교 전기전자공학과 졸업 (학사)
- 2006년 2월 : 연세대학교 전기전자공학과 졸업 (석사)
- 2006년 ~ 현재 : 연세대학교 전기전자공학과 박사과정
- 주관심분야 : 영상신호처리, 비디오 화질평가

**이 철 회**

- 1984년 2월 : 서울대학교 전자공학과 졸업 (학사)
- 1986년 2월 : 서울대학교 전자공학과 졸업 (석사)
- 1986년 ~ 1987년 : Technical University of Denmark (DTH), Denmark
- 1992년 12월 : Purdue University, Electrical Engineering (박사)
- 1993년 ~ 1996년 : National Institutes of Health, Maryland USA
- 현 연세대학교 전기전자공학부 교수
- 주관심분야 : 영상신호처리, 비디오 화질평가