특집논문-09-14-2-02

# 비쥬얼리듬을 이용한 IPTV Set-top Box 재생영상에 대한 PSNR 추정 기법

권 재 철<sup>a)‡</sup>, 서 창 렬<sup>a)</sup>

# A PSNR Estimation Method Exploiting the Visual Rhythm for Reconstructed Video Frames at IPTV Set-top Box

Jae Cheol Kwon<sup>a)‡</sup>, and Chang Ryul Suh<sup>a)</sup>

요 약

본 논문에서는 PSNR(Peak-to-peak Signal to Noise Ratio)이 다른 어떤 척도보다도 화질의 평가에 있어서 중요하다는 전제하에 비쥬얼리 등(VR: Visual Rhythm) 정보를 특징정보로 이용하여 원본영상이 가용하지 않은 일반 시청자들의 STB(Set-Top Box) 재생영상으로부터 PSNR을 추정하는 방법을 제안하였다. 핵심 아이디어는 VR정보는 근사적으로 2차원 영상프레임을 1차원으로 투영한 정보라 할 수 있기 때문에 수직성분의 화소값들로만 이루어져 있을지라도 화면특성에 관한 상당한 정보를 가지고 있기 때문에 송수신단의 VR정보만을 이용하여 PSNR값을 추정하는 것이 가능하다는 것이다. 모의실험을 통해 VR정보로부터 추정한 PSNR값이 매 프레임 별로 2차원 재생영상으로부터 구한 PSNR을 근사적으로 추정함으로써 시청자들이 평균적으로 어느 정도의 화질로 시청하고 있는지 비교적 정확하게 모니터링이 가능하다는 것을 확인하였다. 제안된 방법은 심각한 화면열화가 발생한 시간 위치와 지속시간, 발생횟수 등의 통계정보뿐 아니라 이들을 시각적으로 직접확인할 수 있으며, 자원이 부족한 STB의 연산부담을 최소화하면서 효율적으로 영상품질을 모니터링할 수 있는 장점이 있다.

#### Abstract

In this paper, we propose a PSNR(peak-to-peak signal to noise ratio) estimation method exploiting visual rhythm information for the reconstructed video frames at the customer's STB(Set-top Box). Key idea is that we can estimate the PSNR by using VR(visual rhythm) information even though a VR consists of the pixels in a vertical direction of a 2D(2-dimensional) video frame, because VR is the 1D projected version of a 2D video frame approximately. Simulation results show that the estimated PSNR from VR information is closely related to the PSNR from 2D video frames. The advantages of the proposed scheme includes that it can monitor the video quality efficiently while minimizing the computation load of STB, and show the location, duration and occurrence count of severe picture degradation.

Keyword: video quality measurement, video quality monitoring, visual rhythm; objective metric

# I. 서 론

통신사업자의 IPTV 진출로 방송통신융합이 가속화되고

a) KT 중앙연구소

Central R&D Lab., KT

‡ 교신저자 : 권재철(jckwon@kt.co.kr)

· 접수일(2009년1월14일), 수정일(1차:2009년2월27일, 2차:3월18일), 게재확정일(2009년3월23일)

있다. IPTV는 기존의 지상파방송이나 케이블방송과는 달리 공중통신망을 기반으로 하기 때문에 전송환경이 유리하다고 할 수는 없다. 이러한 전송환경은 결국 IPTV 시청자의 체감품질을 저하시키는 방향으로 작용하게 될 가능성이크다. IPTV사업자 입장에서는 시청자가 경험하는 주관적화질의 수준을 모니터링하고 그 결과를 컨텐츠 생성이나전송방식에 반영하여 고객의 만족도를 높이는 것이 경쟁력

의 원천이 될 것이다. 이를 위해 효과적으로 IPTV의 화질을 측정하고 평가하는 방법이 필요하다.

현재 ITU-T SG9와VQEG(video quality expert group)에서 주관적 화질의 객관적 평가를 위해 표준화 작업을 하고 있는데, 이들이 접근하는 방법은 다음과 같이 크게 세가지 방법이다<sup>[1-2]</sup>.

- FR(full reference): 원본영상과 재생영상간의 유사성 또는 왜곡차이를 계산하여 상대적인 화질을 측정한다. 가장 정확하게 재생화질의 수준을 판단할 수 있다. 그러나 수신측에서 송신측이 사용한 원본영상이 필요하기 때문에 사용이 매우 제한적이다.
- · RR(reduced reference): 원본영상으로부터 추출한 특 징정보와 재생영상으로부터 추출한 특징정보를 비교 하여 상대적인 화질을 측정한다. FR보다는 적은 계산 량으로 상대적인 화질을 측정할 수 있다.
- NR(no reference): 원본영상 자체나 특징정보가 존재 하지 않는 상태에서 재생영상만으로 화질을 측정한다. 기준영상이 없기 때문에 가장 부정확한 방법이다.

이 세 방법 중 FR 방법은 원본영상이 존재해야 하는 가정으로 인해 일반시청자가 경험하는 화질을 측정하는 목적으로는 사용할 수 없다. 원본영상이 가용한 실험실이나 특수한 환경에서만 적용할 수 있다. RR 방법은 원본영상이 필요없이 원본영상의 특징정보만을 이용하고, NR 방법은 아예 원본영상에 대한 어떠한 정보도 필요없는 방법이므로

일반 사용자를 위한 비디오 응용에 적용가능하다.

본 논문에서는 FR과 NR의 중간 정도의 복잡도와 정확도를 가지는 RR 기반의 화질측정 방법이 주요 관심대상이다. 일반적인 RR 방법의 개념도는 (그림 1)과 같다.

송신단에서는 원본영상으로부터 특징정보를 추출하여 이를 별도의 전송채널을 통해 수신단(STB)으로 전송한다. 수신단에서는 디코딩된 재생영상으로부터 송신단과 동일 한 특징추출 방법에 의해 추출된 특징정보를 전송되어온 송신단 특징정보와 비교하여 특정한 화질척도를 통해 화질 을 계산한다.

일반적으로 시청자 댁내의 STB(셋톱박스, Set-top Box)는 자원이 매우 제한된 임베디드 시스템으로 구현되기 때문에 특징정보의 추출과 화질계산에 많은 자원을 소모해서는 곤란하다. STB의 연산능력은 패킷을 처리하고, GUI (graphic user interface) 및 미들웨어(middleware), 다양한응용 프로그램을 작동시키는 등 STB 본연의 임무를 수행하는 것만도 벅차기 때문에 화질측정을 위해 할당할 여분의 연산능력이 넉넉하지 않은 게 현실이다. 그렇기 때문에 STB가 재생영상에 대한 특징정보를 추출하고 송신측에서 전송되어 온 특징정보와 비교하여 화질척도를 계산하는 것이 쉬운 일은 아니다. 따라서 화질측정을 위해 STB에서 실행되는 프로그램(이하 "STB 에이전트"하고 함)은 작고 가벼워야 한다.

기존의 RR 기반의 품질측정 기법들은 대부분 STB의 자원을 고려하지 않고 복잡한 계산을 해야만 얻을 수 있는 방법이다<sup>[3-8]</sup>. 이 중 [7]과 [8]은 [3]-[6]의 방법들과는 다르

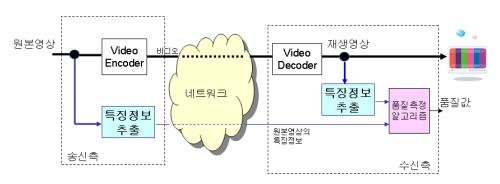


그림 1. 일반적인 감소기준법(RR)의 블록도 Fig.1 Typical block diagram of the reduced reference method

게 비교적 간단한 방법으로 특징정보를 추출하고, 이를 이 용하여 PSNR값을 추정하는 방법에 관한 내용을 다루고 있 다. [7]에서는 압력영상을 적당한 크기의 블록으로 나누어 대역확산(spread spectrum)과 직교변환(orthogonal transform)을 이용하여 주파수영역 또는 공간영역에서 영상데 이터를 무작위화(random)하여. 그 중 한 샘플을 특징정보 로 취하는 방법으로서 추정된 평균PSNR이 매우 정확하 다고 주장하고 있다. 그러나 이 방법은 특징정보를 추출 하기 위해 영상데이터를 PN(pseudo-noise) 시퀀스와 곱하 고 Walsh-Hadamard 변환을 해야 하는 연산이 따른다. 연산 량이 크게 많지는 않지만 모든 픽셀정보에 대해 수행되어 야하므로 여전히 STB에서 처리하기에는 부담스러우며 개 별 프레임 수준에서 추정된 PSNR값은 큰 차이가 날 수도 있다. 또한 720x480@30 fps영상의 경우 특징정보가 차지 하는 대역폭이 1,296 kbps나 필요하여 매우 큰 편이다. [8] 방법은 입력영상에서 "대표 휘도값"을 가지는 픽셀들의 위 치와 그 값을 특징정보로 전송하고 수신측에서 이 위치의 화소값들을 특징정보로 추출하여 비교함으로써 PSNR을 추정한다. 이 방식은 간단하긴 하나 "대표 휘도값"의 위치 정보가 많은 데이터량을 수반하며 실용화하기 위해서는 위 치정보를 압축하기 위한 별도의 이진(binary) 압축코덱이 필요하다.

본 논문은 STB가 가진 연산자원의 한계성을 고려하여 STB에 거의 연산부담을 주지 않으면서도, 특정정보의 전 송에 적은 양의 대역폭만을 필요로 하고, IPTV 사업자의 입장에서 구현가능한 RR 기반의 품질측정 방법을 제안한다. 제안된 방법에서는 비쥬얼리듬 정보를 특징정보로 이

용하여 원영상과 재생영상에서 동일한 방법으로 추출된 특징정보를 비교하여 PSNR을 추정하는 방법을 제안하고자한다. 제안된 방법은 연산자원이 매우 미약한 STB과 같은 임베디드 단말에서도 쉽게 특징정보를 추출할 수 있으며, 추출된 특징정보의 데이터량이 다른 방법에 비해 매우 작고, 비쥬얼리듬 정보를 통해 시각적으로 열화의 정도를 직관적으로 파악할 수 있다는 것이 특징이다. RR 방식의 경우 태생부터 정확도와 복잡도 간의 trade-off 에서 제안되었는데, 본 논문의 경우 PSNR추정의 정확도는 복잡도보다는 부차적인 문제이며, 추정된 PSNR의 평균이 큰 편차를 보이지 않는 수준이면 된다.

본 논문의 전개를 위해 다음과 같은 가정을 한다. 먼저 특징정보로서 계산량이 거의 요구되지 않는 비쥬얼리듬 정보를 사용하는데, 원본영상에 대한 특징정보는 사전에 추출되어 IPTV사업자가 관리하는 품질관리서버에 저장되고, 수신단으로는 전송되지 않는다. 수신단에서는 STB 재생영 상으로부터 송신단과 동일한 방법으로 특징정보를 추출하여 신뢰성 있는 채널을 통해 동일한 품질관리서버로 전송한다. 전송 전에 특징정보를 무손실압축을 하여 데이터량을 더 줄일 수도 있다. 품질관리서버는 송신단의 특징정보 와 수신단의 특징정보를 이용하여 본 논문에서 제시하는 추정방법을 사용하여 두 영상간의 PSNR을 추정하여 재생영상의 품질을 모니터링한다. (그림 2)는 본 논문에서 가정한 품질측정 방법의 블록도를 보여준다.

본 논문의 구성은 다음과 같다. 2장에서 비쥬얼리듬을 소개하고 3장에서 제안방법을 설명한다. 4장에서 모의 실험결과를 기술하고 5장에서 결론을 맺는다.

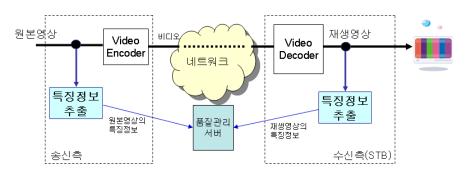


그림 2. 본 논문에서 가정한 감소기준법 기반 품질측정 방법의 블록도

Fig. 2 Block diagram of the proposed quality measurement scheme based on visual rhythm

# II. 비쥬얼리듬(Visual Rhythm) 정보

원래 비쥬얼리듬은 장면전환(scene change)이나 샷전환 (shot change) 검출 등 영상편집 용도로 제안되었으나, 본 발명에서는 RR 방식의 영상품질 측정을 위한 영상데이터의 특징정보로서 비쥬얼리듬 정보를 사용하다<sup>[9]</sup>.

VR은 2차원 비디오 프레임에서 수직, 대각선, 수평 등의 방향에 위치하는 픽셀들을 샘플링하여 1차원 정보로 구성한 다음, 시간축으로 연속적인 비디오 프레임들에 대해 동일한 위치에서 픽셀들을 샘플링하여 3차원 비디오 정보를 2차원 정보로 구성한 정보이다. 일반적으로 영상 프레임은 시간에 따라 그 변화속도가 매우 느리므로 비쥬얼리듬으로

만들어지는 화면은 (그림 3)처럼 같은 장면(scene) 내에서는 완만하게 변화하게 된다. 만일 같은 장면 내에서 화면이열화되는 경우에는 일시적인 불연속이나 이전 또는 이후 프레임들과는 다른 경계면 그림을 보여주게 된다.

비쥬얼리듬을 얻는 방법은 (그림 4)처럼 여러 가지가 있을 수 있다. 위 (그림 3)은 (가) 방법으로 화면의 중앙수직선 상에 있는 화소들을 모아놓은 것이다. 프레임의공간해상도가 W 화소 x H 라인이라고 했을 때 한 프레임에서 얻어지는 비쥬얼리듬 정보는 (라)방법을 제외한 방법들의 경우 H 바이트가 된다. 이 H 바이트의 정보가 한 프레임을 대표하는 특징정보로 사용된다.

SD 해상도(720x480) 비디오에 대해 STB이 품질관리서버

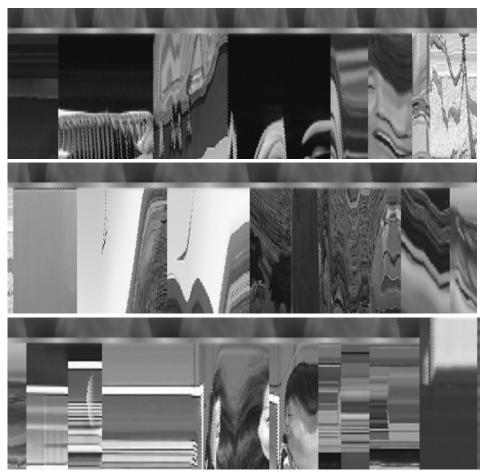


그림 3. 비쥬얼리듬의 예(스펀지-2008.2.16 방송분)

Fig. 3. Example of a visual rhythm.

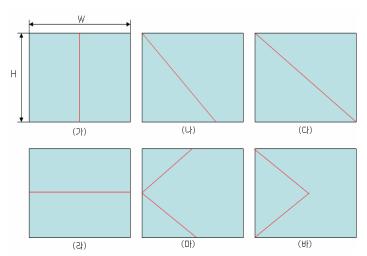


그림 4. 비쥬얼리듬 정보을 위한 화소추출 방법의 예

Fig. 4. Example of the pixel extraction method for visual rhythm.

로 재생영상에 대한 VR정보를 전송하는데 필요한 대역폭은 480 Byte /frame \* 8 bit/Byte \* 30 frame/ sec = 115,200 bit/sec 이다. 1시간짜리 비디오 컨텐츠에 대한 VR정보의 데 이터량은 약 52MB 정도가 된다. 전송대역폭이나 저장 데이 터량을 줄이기 위해서는 시간적 또는 공간적으로 픽셀을 적 절히 서브샘플링하면 되며, VR정보는 화소영역의 정보이기 때문에 적절한 무손실 압축기법으로 압축하여 전송/저장하 면 데이터량을 훨씬 줄일 수 있다. 비쥬얼리듬을 정보를 얻는 과정은 STB의 재생영상 프레임 메모리의 특정 어드레스 공 간의 값을 단순히 복사만 하면 되기 때문에 매우 적은 연산 만으로도 가능하며 실시간으로 처리하는데 문제가 없다. 추 출된 VR 정보는 컨텐츠의 재생이 끝날 때까지 STB의 하드 디스크와 같은 저장공간에 저장된 후 품질관리서버가 요청 할 때 일괄 전송될 수도 있고, 일정 시간단위마다 추출된 VR 정보를 주기적으로 전송할 수도 있을 것이다. 본 논문에서는 VR정보의 전송과 저장 방법에 대해서는 언급하지 않는다.

#### III. VR 정보를 이용한 PSNR 추정 기법

VR은 2차원 영상프레임을 1차원으로 투영(projection)한 정보라고 볼 수 있다. VR이 포함하는 공간방향의 정보는 원본 영상의 극히 일부분에 불과하지만 정보량 측면에서는 2차원

프레임정보의 상당부분을 반영함이 실험을 통해 확인되었다.

 $\mathbf{P}_{SRC}$ 를 원본영상의 픽셀값,  $\mathbf{P}_{STB}$  은 STB 재생영상의 픽셀값,  $\mathbf{P}_{ENC}$  은 인코딩 후 재생영상 프레임 내의 픽셀값을 나타내는 확률변수라 하자. STB에서 재생되는 영상의 총 왜곡량  $\mathbf{D}_{TOTAL}$ 을 다음과 같이 정의하자. 여기서  $\mathbf{E}[\mathbf{x}]$ 는 확률변수  $\mathbf{x}$ 의 기대값(즉 평균)이다.

 $D_{TOTAL}$ 

$$= E[(\mathbf{P}_{SRC} - \mathbf{P}_{STB})^2]$$

$$= E[\{(\mathbf{P}_{SRC} - \mathbf{P}_{ENC}) + (\mathbf{P}_{ENC} - \mathbf{P}_{STB})\}^2]$$

$$= E[\{(\mathbf{P}_{SRC} - \mathbf{P}_{ENC})^2 + (\mathbf{P}_{ENC} - \mathbf{P}_{STB})^2 +$$
 (1)

$$2(\mathbf{P}_{SRC} - \mathbf{P}_{ENC})(\mathbf{P}_{ENC} - \mathbf{P}_{STB})$$

$$= E[(\mathbf{P}_{SRC} - \mathbf{P}_{ENC})^{2}] + E[(\mathbf{P}_{ENC} - \mathbf{P}_{STB})^{2}]$$

$$\equiv D_{ENC} + D_{CHANNEL},$$

여기서 소스코딩의 양자화 왜곡과 관련이 있는 (  $\mathbf{P}_{SRC}$  -  $\mathbf{P}_{ENC}$  ) 정보와 채널오류와 관련된 (  $\mathbf{P}_{ENC}$  -  $\mathbf{P}_{STB}$  ) 성분간에는 상관성이 없다고 가정하였다. 이 중  $\mathbf{D}_{ENC}$ 는 인코딩 과정에서 발생한 왜곡량이며,  $\mathbf{D}_{CHANNEL}$ 은 전송과정 중 발생된 패킷손실에 의해 유입된 왜곡량이다. 그런데  $\mathbf{D}_{ENC}$ 는 인코딩 과정에서 알 수 있는 양이나  $\mathbf{D}_{CHANNEL}$ 은 알 수 없는 양이고 STB에서는 원본영상이 가용하지도 않으므로  $\mathbf{D}_{TOTAL}$ 은 STB이 알 수 없는 양이다.

본 논문에서는 이 양을 추정하기 위해 위에서 기술한 VR

정보를 이용한다.  $D_{CHANNEL}$ 의 추정치를  $D^E_{CHANNEL}$ , 인코딩 단계에서 얻어진 비트스트림의 재생영상에 대한 VR 정보를  $P^{VR}_{ENC}$ , STB에서 재생된 영상에 대한 VR 정보를  $P^{VR}_{STB}$ 이라 하면 총 왜곡량  $D_{TOTAL}$ 의 추정치  $D^E_{TOTAL}$ 는 다음과 같다.

$$\begin{split} \boldsymbol{D}^{E}_{TOTAL} &= \boldsymbol{D}_{ENC} + \boldsymbol{D}^{E}_{CHANNEL} \\ &= \boldsymbol{D}_{ENC} + \boldsymbol{E}[(\boldsymbol{P}^{VR}_{ENC} - \boldsymbol{P}^{VR}_{STB})^{2}] \end{split} \tag{2}$$

즉  $D_{ENC}$ 와 인코딩 과정에서 얻을 수 있는 재생영상으로 취한 VR 정보  $P^{VR}_{ENC}$ 와 STB재생영상에서 취한 VR 정보  $P^{VR}_{STB}$ 를 가지고 총 왜곡량을 추정한다. 전송과정 중에 발생된 오류에 의한 열화가 없다면 총 왜곡량은 인코딩 과정에서 발생된 왜곡량과 같다. 식 (2)로부터 우리는 전송오류에 의해 화질열화가 발생했는지 여부와 열화의 양은 얼마나 되는지에 대한 정보를 얻을 수 있다.

만일 품질관리서버 측에서  $\mathbf{P}^{VR}_{ENC}$  대신 원본영상에 대한  $\mathbf{P}^{VR}_{SRC}$  정보를 가지고 있다면 식 (2) 대신 식 (3)과 같이 더욱 간단하게 총 왜곡량을 추정할 수 있다. 총 왜곡 추정량

 $D^{E}_{TOTAL}$ 은 원본영상의 VR정보와 STB재생영상의 VR정보 로부터 얻어진다. 이 경우에는 인코딩 과정에서 발생한 왜 곡량  $D_{ENC}$ 를 이용하지 않아도 된다.

$$D_{\text{ENC}}^{E} + D_{\text{CHANNEL}}^{E}$$

$$= E[(\mathbf{P}_{\text{SRC}}^{VR} - \mathbf{P}_{\text{ENC}}^{VR})^{2}] + E[(\mathbf{P}_{\text{ENC}}^{VR} - \mathbf{P}_{\text{STB}}^{VR})^{2}]$$

$$= E[(\mathbf{P}_{\text{SRC}}^{VR} - \mathbf{P}_{\text{STB}}^{VR})^{2}]$$
(3)

총 왜곡 추정량에 대한 PSNR값은 다음과 같이 계산된다. 원본영상과 STB재생영상간의 2차원 프레임에 대한 PSNR 값은 식 (1)의 D<sub>TOTAL</sub> 로부터 식 (4)와 같이 구해진다.

$$PSNR_{2D} = 10 \log_{10} \frac{255 \times 255}{D_{TOTAL}}$$
 (4)

원본영상 VR정보와 STB재생영상 VR정보간의 PSNR 값, 즉 2차원 영상의 PSNR 추정치는 식 (3)의  $D^E_{TOTAL}$ 로부터 식 (5)와 같이 구해진다.

PSNR<sub>E</sub> = 
$$10 \log_{10} \frac{255 \times 255}{D_{\text{TOTAL}}^{E'}}$$
 (5)

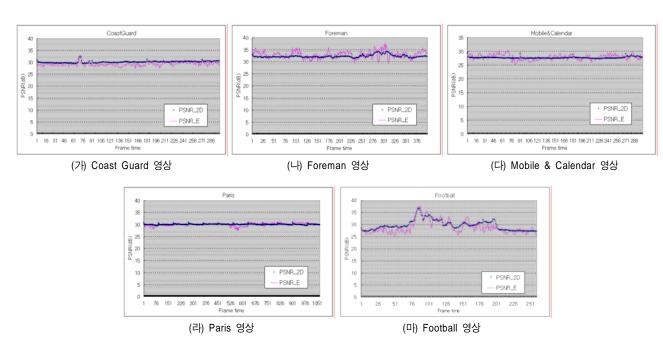


그림 5. 2차원 영상프레임간(PSNR\_2D) 및 이들의 VR정보간(PSNR\_E)에 대한 PSNR 비교(VR정보는 세로방향 픽셀들을 추출하였음) Fig. 5. PSNR comparison between 2D video frames and their visual rhythms.

(그림 5)는 다섯 개의 실험영상에 대해 2차원영상프레임을 H.263 TMN-8 코텍으로 양자화변수 Qp=16으로 인코딩한 후 얻어진 재생영상(PENC)과 원본영상(PSRC)간의 PSNR 값과 이들의 VR정보만으로 추정한 PSNR값을 보여주고 있다. 즉, 식 (1)과 식 (3)에서 채널왜곡량 DCHANNEL과 DECHANNEL이 0인 경우에 대해 총왜곡량을 PSNR로 환산한 값이다. (그림 6)은 양자화 변수 Qp만 달리하여 위와 동일한인코딩 방법으로 얻은 전 재생프레임들에 대한 PSNR값들의 평균을 나타낸 것이다. 여기서 VR 정보는 (그림 4) (가)방법(세로방향)으로 프레임의 한 가운데에 존재하는 수직픽셀들을 추출하여 얻었다.

VR정보로 구한 PSNR(PSNR\_E)과 2차원 영상프레임에 대한 PSNR(PSNR\_2D)이 영상의 종류에 따라 프레임별로 다소 변동은 있지만 평균적으로 상당히 근접함을 볼 수 있다. 단순한 영상일수록 추정이 좀더 정확하고, 화면이 복잡하고 움직임이 많은 영상일수록 좀 더 큰 폭으로 변동하는 경향이 있다. 이러한 성질을 이용하면 VR정보를 재생영상의 화질특성을 대표하는 정보로 활용할 수 있다. VR정보가 2차원 영상프레임의 특성을 반영하는 이유는 VR정보가 영

상프레임에서 시각적으로 중요한 화면 중간부분을 샘플링 한다는 것과 수평성분 없이 수직성분의 화소들만 가지고도 어느 정도 특성을 파악할 수 있기 때문이다.

일반적으로 인코딩 과정은 수평방향의 왜곡과 수직방향의 왜곡을 차별하지 않는다. 또 만일 전송오류에 의해 영상패킷이 손실되었을 때 이 손실에 의한 영향은 손실이 발생한 비트스트림이 포함하고 있는 공간방향 위치의 매크로블럭에서 시작되어 수평 및 수직방향으로 화면을 열화시키며 시간방향으로 열화가 전파되어 간다. 따라서 수직방향의 화소들을 샘플 링하는 VR 정보는 패킷손실의 효과를 충실히 반영하면서 시간방향으로 전파되는 왜곡까지 반영하므로 2차원 재생영상프레임의 특성을 어느 정도 반영한다고 볼 수 있다.

#### IV. 모의실험 결과

전송오류가 없는 환경에서 VR정보를 이용하여 PSNR을 추정한 (그림 5), (그림 6)의 결과가 패킷손실이 있는 실제 IPTV 전송환경에서도 적용될 수 있는 지 확인하기 위하여

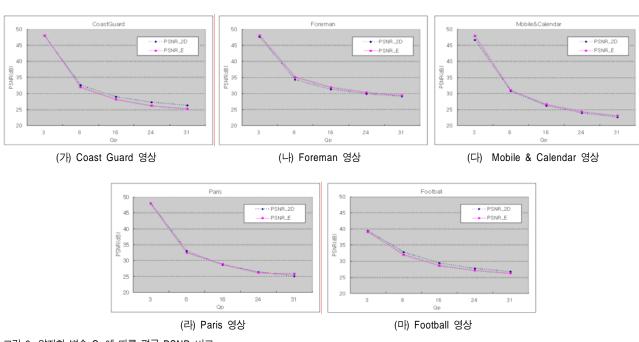


그림 6. 양자화 변수 Qp에 따른 평균 PSNR 비교 Fig. 6. Average PSNR comparison for different Qp's

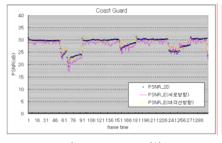
모의실험을 수행하였다. 실험에 사용된 영상은 CIF 규격의 Foreman, Coast Guard, Mobile & Calendar, Paris, Football 영상이다. 실험영상을 인코딩하는데 사용한 코덱은 H.263 TMN-8 코덱을 사용하였고, 실험영상을 양자화 계수 Qp=16으로 부호화하였다. 상용 IPTV에서는 H.264 코덱을 사용하지만 본 논문에서는 VR정보를 통해 원본영상과 재생 영상간의 PSNR을 추정하는 것이 목적이므로 어떤 코덱을 사용하였는지는 중요하지 않다.

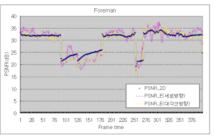
PSNR을 추정하기 위하여 원본영상과 디코더측에서 재생된 영상의 VR정보을 추출하였다. VR정보를 추출하는 방법이 추정성능에 미치는 영향을 알아보기 위하여 (그림 4) (가) 세로방향, (나) 대각선방향으로 화소들을 샘플링하여 VR정보를 추출하였다. 네트워크 전송 중에 발생된 패킷손실에 의해 재생화면이 열화되는 것을 에뮬레이션 하기 위하여 부호화된 H.263 압축 비트스트림을 1,500 바이트 단위로 세그먼트하여 패킷화하고, 패킷손실율 1%로 손실을 유발하였다. 패킷손실에 의한 화면열화가 계속 전파되는 것을 방지하기 위해 매 90 프레임마다 화면내(intra) 프레임을 삽입하였으며, 손실에 의해 열화된 화면의 주관적 화질 향상

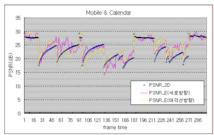
을 위해 디코더에서는 에러가 발생한 영역을 이전 프레임의 같은 영역으로 대체하는 에러감춤(error concealment) 기법 이 적용되었다. 실제 IPTV 전송환경에서는 하나의 GoP (Group of Picture)가 보통 15프레임으로 구성되어 패킷손 실이 발생하더라도 화면열화가 최대 15 프레임 이상을 전파하지는 않도록 되어 있는데, 이 실험에서는 패킷손실에 의한 화면열화가 진행되는 과정에서 PSNR 추정성능을 평가하기 위해 좀 더 열악한 손실 환경은 가정하였다.

실험영상을 가지고 패킷오류가 발생하는 상황에서 상기식 (1) 및 식 (3)을 통해 프레임 단위로 추정한 총 왜곡량을 PSNR로 환산한 결과는 (그림 7)과 같다. 여기서 PSNR<sub>2D</sub> (=PSNR\_2D)와 PSNR<sub>E</sub>(=PSNR\_E)는 각각 식 (4)와 식 (5)로 정의된 양으로서, 원본영상과 채널왜곡이 존재하는 STB 재생영상간의 2D영상과 그들의 VR정보에 대해 계산된 값이다. (그림 7)에서 2차원 영상에 대한 PSNR<sub>2D</sub>가 급격하게 감소한 곳은 패킷손실에 의해 재생영상이 현저히 열화된 경우이다.

Coast Guard 영상의 경우 모든 프레임에 걸쳐 비교적 높은 정확도로 PSNR을 추정하고 있으며, Foreman과 Mobile&



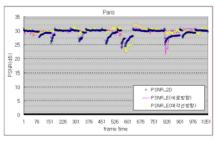


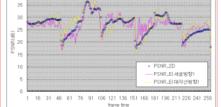


(가) Coast Guard 영상

(나) Foreman 영상

(다) Mobile & Calendar 영상





(라) Paris 영상

(마) Foorball 영상

그림 7. PSNR 추정 성능 비교

Fig. 7. Performance comparison on PSNR estimation accuracy

Calendar의 경우에는 패킷손실에 의한 화면열화가 없는 구간에서는 PSNR 추정이 비교적 잘 되다가 패킷손실 구간에서는 변동폭이 다소 컸다. PSNR<sub>2D</sub>가 급격하게 감소한 곳에서는 PSNR<sub>E</sub>도 급격하게 감소하여 VR정보가 패킷손실에 의한 화면열화를 충실히 반영함을 확인하였다. Paris영상의 경우에도 패킷손실이 없는 경우에는 전반적으로 추정이 잘이루어졌다. 그러나 첫 중반부분까지 패킷손실에 의해 PSNR<sub>2D</sub>는 떨어졌으나 PSNR<sub>E</sub>는 손실을 잘 검출하지 못했다. 이는 손실이 발생한 부분이 VR정보를 추출한 영역이외의 영역에서 발생했거나 영상의 특성상 VR정보를 추출한 화면중간 영역이 정적인 영역이라 에러감춤이 잘되어 VR상으로는 에러가 없는 것처럼 보였기 때문이다. Football 영상의 경우에는 화면도 복잡하고 움직임도 큰 프레임 구간에서 패킷손실로 화면이 심하게 열화되었기 때문에 이 구간에서 추정된 PSNR의 변동폭이 다소 컸다.

(그림 8)은 PSNR<sub>2D</sub>과 PSNR<sub>E</sub> 간의 상관성을 보여주는데 Paris 영상이 다소 낮은 상관성을 보여주는 것을 제외하면 네 영상에서 두 값이 상당히 밀접하게 연관되어 있음을 확인할 수 있다. Paris 영상의 다소 낮은 상관성은 영상의 특성 상 PSNR<sub>E</sub>가 패킷손실을 잘 반영하지 못했기 때문인 것으로 보인다. VR을 추출하는 방법에 따라 두 값들간의 상관도도 약간씩 다르긴 하지만 큰 차이는 없었다. (그림 9)는 VR로부터 추정된 평균 PSNR(PSNR<sub>E</sub>)과 2차원 재생영상의 평균PSNR(PSNR<sub>2D</sub>)을 비교한 것인데 영상의 종류에 따라 약간씩 다르긴 하지만 대체로 PSNR<sub>E</sub>가 PSNR<sub>2D</sub>에 근접한 추정성능을 보여주고 있다. 모든 실험영상에 대해 전체프레임에 대한 평균 PSNR 값은 개별 프레임의 경우와는 달리 1 dB 이내의 상당히 근접한 수준으로 추정하였음 볼수 있다. 한편 VR을 추출한 방향(세로방향, 대각선방향)에 따른 평균 PSNR추정성능도 실험영상에 따라 약 간 다르긴

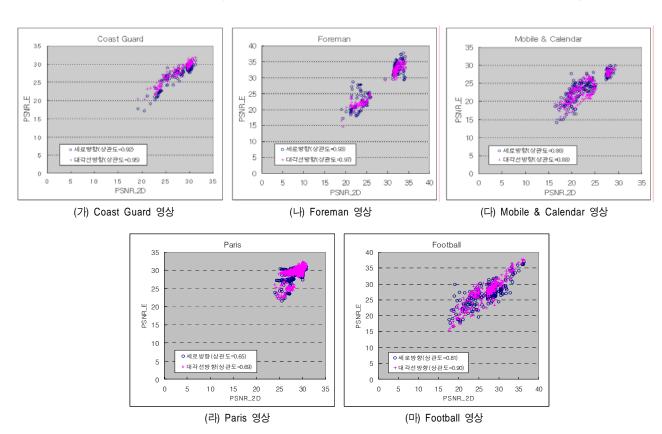


그림 8. PSNR<sub>2D</sub>와 PSNRE 간의 상관관계 Fig. 8. Correlation between PSNR<sub>2D</sub> and PSNRE

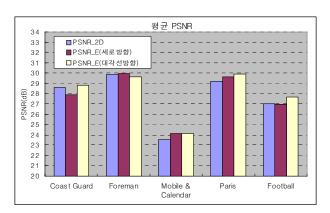


그림 9. PSNR<sub>2D</sub>와 PSNRE 의 평균값 비교

Fig. 9. Comparison of average PSNR<sub>2D</sub> and average PSNRE

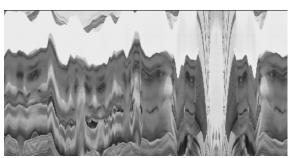
하지만 큰 차이는 없었다. 전반적으로 패킷손실이 없는 구간 보다도 손실에 의해 화면이 심하게 열화되는 경우 추정 성능 이 다소 떨어지는 경향을 보였지만, 평균 PSNR의 관점에서 는 여전히 의미있는 결과를 얻을 수 있었다.

STB 재생영상에 대한 PSNR값을 추정하는 것은 시청자가 어느 정도의 화질로 비디오 컨텐츠를 시청을 하였는지

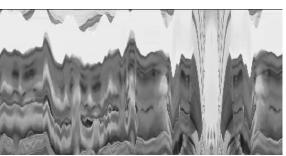
에 대한 통계정보를 얻기 위한 것이므로 매 프레임에 대해 IdB 이내의 정확도로 PSNR을 추정해야 할 필요는 사실 없다. 그보다는 오히려 컨텐츠의 어느 재생시각에서 화면이 심하게 열화되었는지, 열화가 어느 정도 지속되었는지, 심각한 열화가 몇 번이나 발생하였는지에 대한 정보가 더욱 중요한 통계정보이다.

실제 상용 IPTV시스템에서는 화면이 매우 복잡하거나 격렬한 움직임이 있는 장면 이외에는 인코딩 과정에서 발 생한 양자화 왜곡이 문제가 될 정도로 화질열화가 심하지 는 않다. 이러한 장면들의 경우에도 비트량을 좀더 할당하 는(즉, 전송율을 좀 더 높이는) 방법으로 얼마든지 해결이 가능하다. 만일 그렇지 않다면 고객들이 그 IPTV 사업자로 부터 이탈할 것이 분명하기 때문이다. 따라서 상용시스템 에서는 원본영상보다도 인코딩된 영상을 기준으로 STB 재 생영상의 화질을 평가하는 것이 더 효과적이다. 이 경우 STB 재생영상의 심각한 열화는 대부분 네트워크 전송과정 중에 발생한 패킷손실에 의해 나타난다.

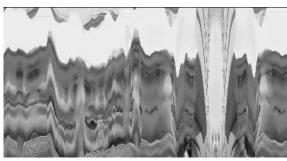
본 논문은 이러한 경우에도 매우 유용하게 적용된다. 인



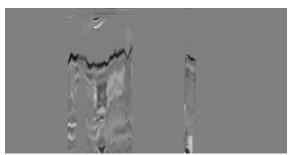
(기) 원본영상의 VR정보



(나) 인코딩 영상의 VR 정보



(다) STB 재생영상의 VR 정보

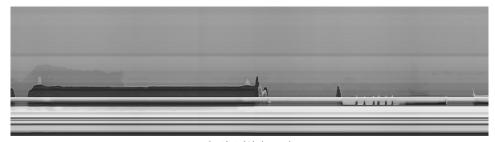


(라) (나)와 (다)의 차이영상

그림 10. Foreman 영상의 비쥬얼리듬(VR) 정보(총 400 프레임)

Fig. 10 Visual rhythm information of Foreman sequence (Total 400 frames)

코딩 영상과 STB 재생영상으로부터 각각 VR 정보를 추출 하여 이 두 정보를 단순히 비교하기만 하면 패킷손실로 야 기된 화질열화에 대한 각종 통계정보를 쉽게 얻을 수 있다. 제안된 PSNR추정방법은 화면열화가 발생했을 때 그 열화 가 어느 정도로 심각한 수준인가를 추정하는데 사용될 수 있다. 무엇보다도 (그림 10), (그림 11)과 같이 VR정보와 이들간의 차이를 확인함으로써 품질평가자가 직접 눈으로 도 패킷손실에 의한 화면열화의 위치와 지속시간, 횟수, 열



(가) 원본영상의 VR정보



(나) 인코딩 영상의 VR 정보



(다) STB 재생영상의 VR 정보



(라) (나)와 (다)의 차이영상

그림 11. Paris 영상의 비쥬얼리듬(VR) 정보(총 1062 프레임)

Fig. 11. Visual rhythm information of Paris sequence (Total 1062 frames)

화 양을 쉽게 확인할 수 있는 장점이 있다.

# V. 결 론

본 논문에서는 PSNR이 다른 어떤 척도보다도 화질의 평가에 있어서 중요하다는 전제하에 비쥬얼리듬(VR) 정보를 특징정보로 이용하여 원본영상이 가용하지 않은 일반 시청자들의 STB 재생영상으로부터 PSNR을 추정하는 방법을 제안하였다. 2차원 영상프레임으로부터 추출한 1차원 VR정보는 근사적으로 2차원 영상프레임을 1차원으로 투영한정보라 할 수 있기 때문에 비록 수직성분의 화소값들로만이루어져 있을지라도 화면특성에 관한 상당한 정보를 가지고 있다고 할 수 있다. 따라서 송수신단의 VR정보만을 이용하여 PSNR값을 추정하는 것이 가능하다.

모의실험을 통해 VR정보로부터 추정한 PSNR값이 매프레임 별로 2차원 재생영상으로부터 구한 PSNR을 근사적으로 추정함으로써 시청자들이 평균적으로 어느 정도의화질로 시청하고 있는지 근사적으로 모니터링이 가능하다는 것을 확인하였다. 패킷손실이 발생하여 화질이 급격하게 열화되는 구간의 경우에는 개별 프레임에 대해 추정된 PSNR이 실제 PSNR에 비해 비교적 큰 폭으로 변동하여 추정 성능이 떨어지기는 하지만 평균 PSNR의 관점에서는 여전히 의미있는 결과를 얻을 수 있었다. 또한 비쥬얼리듬 정보를 이용하면 심각한 화면열화가 발생한 시간 위치와 지속시간, 발생횟수 등의 통계정보뿐 아니라 시각적으로 직접 확인할 수 있는 장점이 있다.

본 논문에서 가정한 감소기준법(RR) 기반의 품질측정 시나리오를 따른다면 STB는 단순히 재생영상으로부터 VR 정보만 추출해서 품질관리서버로 전송만 하면 되고, 나머지 연산은 품질관리서버에서 수행하므로 자원이 부족한 STB의 연산부담을 경감하면서도 특징정보를 전송하는데 적은 상향 대역폭만을 사용할 수 있는 것이 본 논문의 특징이다. 비쥬얼리듬 정보를 이용한 감소기준법 기반의 품질측정 방법은 (그림 2)와 같은 시나리오 외에 STB의 능력에따라 다양한 시나리오가 가능하다.

본 논문에서는 원본영상과 STB 재생영상간의 PSNR을 추정하는 방법을 제안하였지만 실제 IPTV 사업 환경에서는 원본영상보다 인코딩된 영상과 STB 재생영상간의 품질이 더 중요할 수도 있다. 사실 원본영상과 인코딩된 영상간에는 인코딩 과정 중에 정확한 PSNR 값을 얻을 수 있기때문에 이 값을 이용할 수 있다면 전송과정 중에 발생한손실에 의해 야기된 STB 재생영상의 화질열화만이 문제로남기때문이다. 이 문제는 현재 연구 중에 있으며, 적절한품질측정 시나리오를 정립하고, 제안된 방법을 구현하기위해 VR정보를 가공, 저장 또는 전송하는 방법과 VR정보를 무기준법(NR)품질측정 방법에 활용할 가능성이 있는지타당성을 확인하는 연구도 진행 중이다.

#### 참 고 문 헌

- Recommendation J.144 Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference.
- [2] Video quality expert group, http://www.vqeg.org.
- [3] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," Human Vision and Electronic Imaging X, Proc. SPIE, vol. 5666, San Jose, CA, Jan. 2005.
- [4] Tobias Oelbaum and Klaus Diepold, "Building a Reduced Reference Video Quality Metric with Very Low Overhead using Multivariate Data Analysis", at http://www.iiisci.org/journal/CV\$/ sci/pdfs/I356ZP.pdf
- [5] Patrick Le Callet, Christian Viard-Gaudin, Stphane Pechard and milie Caillanult, "No Reference and Reduced Reference Video Quality Metrics for End to End QoS Monitoring", IEICE Transactions on Communications 2006 E89-B(2):289-296.
- [6] 김동오, 박래홍, 심동규, "에지 투영의 유사도를 이용한 압축된 영상에 대한 Reduced-Reference 화질 평가", 대한전자공학회 논문지, 제45권 SP편 제3호, pp. 221-229, 2008. 5.
- [7] Recommendation J.240 Framework for remote monitoring of transmitted picture signal-to-noise ratio using spread-spectrum and orthogonal transform.
- [8] T. Yamada, Y. Miyamoto, M. Serizawa, and H. Harasaki, "Reduced-reference based video quality metrics using representative-luminance values," in Proc. International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Ariz., USA, Jan. 2007.
- [9] 이진호 외, "화소샘플링을 이용한 시각율동 생성방법", 한국특허 10-1998-0056144, 2001.10.23

# ----- 저 자 소 개 <del>-</del>



# 권 재 철

- 1986년 : 한양대학교 전자공학과 졸업 - 1988년 : KAIST 전기및전자공학과 석사 - 2003년 : KAIST 전기및전자공학과 박사

- 1988년 ~ 현재 : KT 중앙연구소 연구전문 수석연구원

- 주관심분야 : 멀티미디어통신, 영상코덱, 패킷비디오, 영상품질측정



# 서 창 렬

- 1982년 : 경희대학교 전자공학과 졸업 - 1984년 : 연세대학교 전자공학과 석사 - 현재 : KT 중앙연구소 수석연구원

- 주관심분야 : 멀티미디어통신, 영상품질측정, 영상객체화 기술