

# 비주얼리듬 기반 IPTV Set-top Box 재생영상에 대한 PSNR 추정 기법

\*권재철, \*서창렬

\*KT 신사업부문 미래기술연구소

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

Jae Cheol Kwon<sup>o</sup>, Chang Ryul Suh

Advanced Technology Lab., New Biz Division, KT

jckwon@kt.com, crsuh@kt.com

### 요 약

본 논문에서는 PSNR 이 다른 어떤 척도보다도 화질의 평가에 있어서 중요하다는 전제하에 비주얼리듬(VR) 정보를 이용하여 원본영상이 가용하지 않은 일반 시청자들의 STB 재생영상으로부터 PSNR 을 추정하는 방법을 제안하였다. VR 로부터 추정한 PSNR 값이 매 프레임 별로 2 차원 재생영상으로부터 구한 PSNR 과 같지는 않으나 시청자들이 평균적으로 어느 정도의 화질로 시청하고 있는지 비교적 정확하게 모니터링이 가능하다. 제안된 방법은 감소기준법(RR) 영상품질 측정방법으로 사용될 수 있다. STB 은 단순히 재생영상으로부터 VR 정보만 추출해서 서버로 전송하는 역할만 수행하면 되고, 나머지 연산은 품질관리 서버에서 수행하므로 자원이 부족한 STB 의 부담을 경감하면서도 효율적으로 품질을 모니터링할 수 있는 장점이 있다.

### 1. 개 요

통신사업자의 IPTV 진출로 방송통신융합이 가속화되고 있다. IPTV 는 기존의 지상파나 케이블방송과는 달리 공중통신망을 기반으로 하기 때문에 전송환경이 유리하다고 할 수는 없다. 이러한 전송환경은 결국 IPTV 시청자의 체감품질을 저하시키는 방향으로 작용하게 될 가능성이 크다. 따라서 IPTV 사업자는 투자대비 효율적인 콘텐츠 전달을 위한 적절한 콘텐츠 인코딩 속도와 액세스망 자원의 할당을 위한 노력을 경주하여야 한다. IPTV 사업자 입장에서는 시청자가 경험하는 주관적 화질의 수준을 모니터링하고 그 결과를 콘텐츠 생성이나 전송에 반영하여 고객의 만족도를 높이는 것이 경쟁력의 원천이 될 것이다. 이를 위해 효과적으로 IPTV 의 화질을 측정하고 평가하는 방법이 필요하다.

현재 ITU-T 와 VQEG(video quality expert group)에서 주관적 화질의 객관적 평가를 위해 표준화 작업을 하고 있는데, 크게 FR(full reference), RR(reduced reference), NR(no reference) 등 세가지 방법으로 접근하고 있다[1-2]. 품질척도에 있어서는 그동안 새로운 많은 척도들이 주관적 화질을 객관적인 양으로 측정하기 위해 개발되어 왔지만, 아직까지 PSNR(peak-to-peak signal to noise ratio)만큼 정확하면서 널리 받아들여지고 있는 척도도 찾아보기 어렵다. PSNR 이 인간의 시각특성을 정확하게

반영하지 못하기 때문에 주관적 화질평가 측면에서는 절대적인 평가기준이 되지 못한다는 비판이 많기는 하지만 다른 어떤 척도보다 많은 정보를 제공함은 부인할 수 없다. 특히 시청자택내 STB(set-top box)에서는 원본영상을 이용할 수 없기 때문에 PSNR 은 화질평가 척도로서 고려대상이 될 수 없었다. 원본영상이 없는 이러한 환경에서는 NR 이나 RR 방법이 사용되는데, 이 방법에서 PSNR 값을 추정할 수 있다면 화질평가에 도움이 될 것이다.

본 논문에서는 상향 대역폭을 사용할 수 있는 IPTV 환경에서 일반 시청자 STB 이 재생하는 재생영상으로부터 특징정보를 추출하고, 상향대역폭을 이용하여 이 정보를 IPTV 사업자가 관리하는 품질관리서버로 전송하여, 그 특징정보로부터 서버가 PSNR 값을 추정함으로써 개별 시청자의 시청화질 수준을 추정하는 방법을 제안한다. 제안 방법은 RR 기법으로 분류될 수 있으며, 사용되는 특징정보로는 비주얼리듬(VR: visual rhythm) 정보를 사용한다.

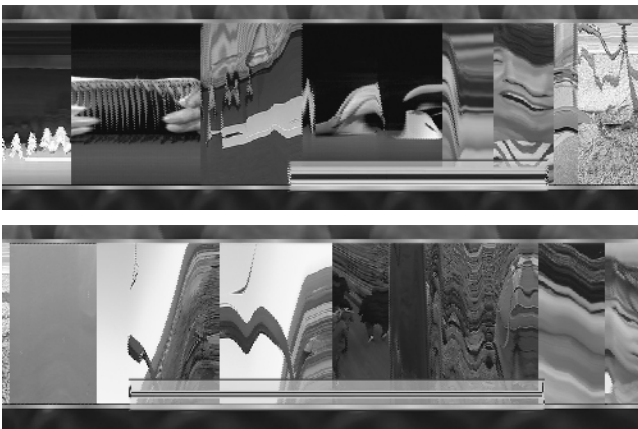
본 논문의 구성은 다음과 같다. 2 장에서 비주얼리듬을 소개하고 3 장에서 제안방법을 설명한다. 4 장에서 모의 실험 결과를 기술하고 5 장에서 결론을 맺는다..

### 2. 비주얼리듬(Visual Rhythm)

원래 비주얼리듬은 장면전환(scene change)이나 샷전환(shot change) 검출 등 영상편집 용도로

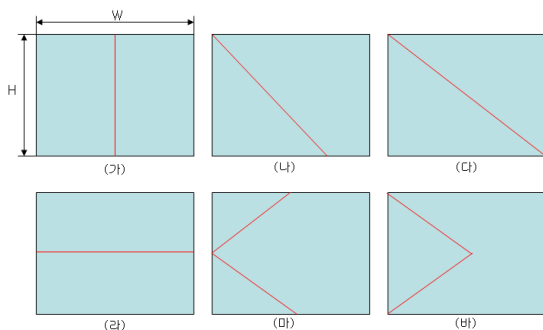
제안되었으나, 본 발명에서는 RR 방식의 영상품질 측정을 위한 영상데이터의 특징정보로서 비주얼운동 정보를 사용한다[3].

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 비주얼운동으로 만들어지는 화면은 <그림 1>처럼 같은 장면(scene) 내에서는 완만하게 변화하게 된다. 만일 같은 장면 내에서 화면이 열화되는 경우에는 일시적인 불연속이나 이전 또는 이후 프레임들과는 다른 경계면 그림을 보여지게 된다.



<그림 1> 비주얼운동의 예(스페인-2008.2.16 방송분)

비주얼운동을 얻는 방법은 <그림 2>처럼 여러가지가 있을 수 있다. 위 <그림 1>은 (가) 방법으로 화면의 중앙수직선상에 있는 화소들을 모아놓은 것이다. 프레임의 공간해상도가 W 화소 x H 라인이라고 했을 때 한 프레임에서 얻어지는 비주얼운동 정보는 (라)방법을 제외한 방법들의 경우 H 바이트가 된다. 이 H 바이트의 정보가 한 프레임을 대표하는 특징정보로 사용된다.



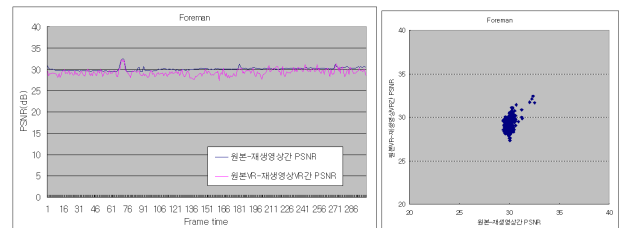
<그림 2> 비주얼운동을 위한 화소추출 방법의 예

SD 해상도(720x480) 비디오에 대한 VR 정보를 전송하는데 필요한 대역폭은  $480 \text{ Byte} \times 8 \text{ bit/Byte} \times 60 \text{ sec} = 230,400 \text{ bit/sec}$  이다. 대역폭을 줄이기 위해서는 시간적, 공간적으로 픽셀을 적절히 서브샘플링하면 된다. 비주얼운동을 얻

는 과정은 수신측에서 디스플레이되는 메모리의 값을 단순히 복사만 하면 되기 때문에 실시간 처리가 가능하다.

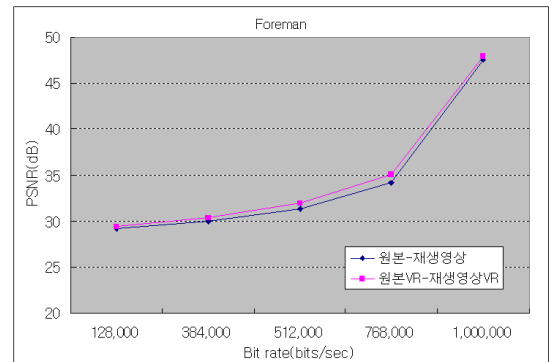
### 3. VR을 이용한 PSNR 추정 기법

VR은 2차원 영상프레임 정보를 1차원으로 투영(projection)한 정보라고 볼 수 있다. VR이 포함하는 공간 방향의 정보는 원본영상의 극히 일부에 불과하지만 정보량 측면에서는 2차원 프레임정보의 상당부분을 반영한다. 아래 <그림 3>은 2차원 영상프레임을 인코딩한 후 디코딩한 재생영상과 원본영상간의 PSNR 값과 이들의 VR 정보만으로 계산한 PSNR 값 및 이 PSNR 값들간의 상관성을 보여주고 있다. VR의 PSNR과 2차원 영상프레임에 대한 PSNR이 높은 상관성을 가짐을 볼 수 있다. 이러한 상관성을 이용하면 VR 정보를 재생영상의 특성을 대표하는 정보로 활용할 수 있다. VR이 한 프레임의 특성을 반영하는 이유는 화면의 중요한 중간부분을 샘플링한다는 것과 인코딩 및 전송 과정에서 영상의 공간적 상관성이 화질에 영향을 미치기 때문이다. 본 논문에서는 인코딩 과정에서 얻어진 재생영상에 대한 VR 정보는 사전에 품질관리서버에 저장된다고 가정하고, STB 재생영상에 대한 VR 정보는 시청자가 시청하는 동안 생성되어 일정 시간마다 또는 시청 후에 한꺼번에 품질관리서버로 전송된다고 가정한다.



(가)

(나)



(다)

<그림 3> 2차원 영상프레임과 이들의 VR에 대한 PSNR 비교(H.263 코덱 사용)

- (가) 원본영상과 재생영상, 이들의 VR들간의 PSNR 값
- (나) 원본-재생영상 PSNR과 원본 VR-재생영상 VR 간 PSNR 값의 상관성
- (다) 비트율별 PSNR 비교

STB에서 재생되는 영상의 총 왜곡량  $D_{TOTAL}$ 은 다음과 같다.  $P_{SRC}$ 는 원본영상,  $P_{STB}$ 은 STB 재생영상,  $P_{ENC}$ 은 인코딩 후 재생영상 프레임 내의 픽셀값이다.

$$D_{TOTAL} = E[(P_{SRC} - P_{STB})^2] = E\{[(P_{SRC} - P_{ENC}) + (P_{ENC} - P_{STB})]^2\}$$

$$\begin{aligned}
&= E[(\mathbf{P}_{\text{SRC}} - \mathbf{P}_{\text{ENC}})^2] + E[(\mathbf{P}_{\text{ENC}} - \mathbf{P}_{\text{STB}})^2] \\
&= D_{\text{ENC}} + D_{\text{CHANNEL}}, \\
&\text{if } D_{\text{ENC}} \text{ and } D_{\text{CHANNEL}} \text{ are uncorrelated.}
\end{aligned} \tag{1}$$

이 중  $D_{\text{ENC}}$  는 인코딩 단계에서 알 수 있는 양이나  $D_{\text{CHANNEL}}$  은 알 수 없는 양이고 STB 에서 원본영상이 가용하지도 않으므로  $D_{\text{TOTAL}}$  은 알 수 없는 양이다. 본 논문에서는 이 양을 추정하기 위해 위에서 기술한 VR 정보를 사용한다. 총 왜곡량의 추정치를  $D_{\text{TOTAL}}^E$  라고 하고,  $D_{\text{CHANNEL}}$  의 추정치를  $D_{\text{CHANNEL}}^E$ , 인코딩 단계에서 얻어진 비트스트림의 재생영상에 대한 VR 정보를  $\mathbf{P}_{\text{ENC}}^{\text{VR}}$ , STB 에서 재생된 영상에 대한 VR 정보를  $\mathbf{P}_{\text{STB}}^{\text{VR}}$  정보라 하면 다음과 같다.

$$\begin{aligned}
D_{\text{TOTAL}}^E &= D_{\text{ENC}} + D_{\text{CHANNEL}}^E \\
&= D_{\text{ENC}} + E[(\mathbf{P}_{\text{ENC}}^{\text{VR}} - \mathbf{P}_{\text{STB}}^{\text{VR}})^2]
\end{aligned} \tag{2}$$

즉 인코딩 과정에서 취한 VR 정보  $\mathbf{P}_{\text{ENC}}^{\text{VR}}$  와 STB 에서 취한 VR 정보  $\mathbf{P}_{\text{STB}}^{\text{VR}}$  를 가지고 총 왜곡량 추정이 가능하다. 만일 서버측에서  $\mathbf{P}_{\text{ENC}}^{\text{VR}}$  대신 원본영상에 대한  $\mathbf{P}_{\text{SRC}}^{\text{VR}}$  정보를 이용할 수 있다면 식 (2)는 식 (3)과 같이 더욱 간단하게 추정할 수 있다. 이 경우에는  $D_{\text{ENC}}$  를 이용하지 않아도 된다.

$$\begin{aligned}
D_{\text{TOTAL}}^E &= D_{\text{ENC}}^E + D_{\text{CHANNEL}}^E \\
&= E[(\mathbf{P}_{\text{SRC}}^{\text{VR}} - \mathbf{P}_{\text{ENC}}^{\text{VR}})^2] + E[(\mathbf{P}_{\text{ENC}}^{\text{VR}} - \mathbf{P}_{\text{STB}}^{\text{VR}})^2] \\
&= E[(\mathbf{P}_{\text{SRC}}^{\text{VR}} - \mathbf{P}_{\text{STB}}^{\text{VR}})^2]
\end{aligned} \tag{3}$$

PSNR 은 식(4)와 같이 계산된다.

$$\text{PSNR} = 10 \log_{10} \frac{255 \times 255}{D} \tag{4}$$

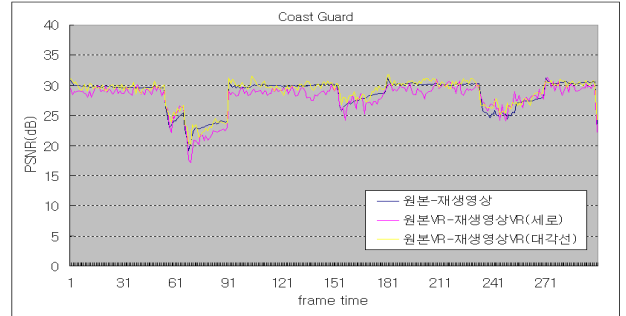
여기서  $D = \begin{cases} D_{\text{ENC}}, & \text{원본-재생영상간} \\ D_{\text{TOTAL}}^E, & \text{원본VR-재생영상VR간} \end{cases}$

#### 4. 모의실험 결과

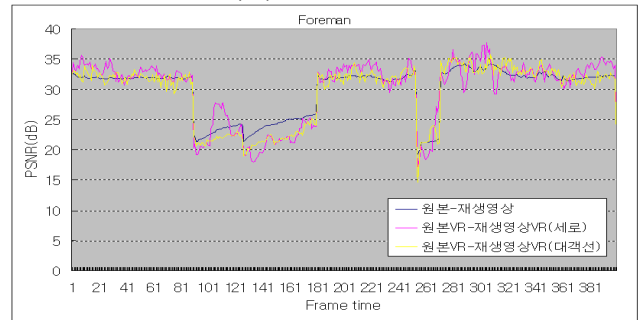
본 논문에서 제안하는 VR 정보를 이용한 PSNR 추정 성능을 평가하기 위해 모의실험을 수행하였다. 실험에 사용된 실험영상은 CIF 규격의 Foreman, Coast Guard, Mobile & Calendar, Paris 영상이다. 실험영상을 인코딩하는 데는 H.263@700kbps 을 사용하였고, 원본영상의 VR 과 디코더 측에서 재생된 영상의 VR 을 추출하였다. VR 을 추출하는 방법은 <그림 2>의 (가)세로방향, (나) 대각선방향의 화소를 각각 샘플링하였다. 네트워크 전송에러를 에뮬레이션하기 위해 H.263 압축 비트스트림을 1,500 바이트 단위로 세그먼트하여 패킷화하고, 패킷손실을 1%로 손실을 유발하였다. 패킷손실에 의한 화면열화가 계속 전파되는 것을 방지하기 위해 매 90 프레임마다 인트라 프레임을 삽입하였으며, 손실에 의해 열화된 화면의 주관적 화질 향상을 위해 디코더에서는 에러가 발생한 영역을 이전 프레임의 같은 영역으로 대체하는 에러감춤(error concealment) 기법이 적용되었다.

실험영상을 가지고 패킷에러가 발생하는 상황에서 상기 식(3)과 같이 추정한 총 왜곡량을 PSNR 로 환산한 결과와는 <그림 4>, <그림 5>와 같다. <그림 3>과 같이 여전히 높은 상관성을 보여주고 있다. <그림 6>은 VR 로부터 추정된 평균 PSNR 과 2 차원 재생영상의 평균 PSNR 을 비교한 것인데 평균 PSNR 상으로는 약 1 dB 이내의 차이를 보이는 정확도를 가짐을 확인할 수 있다.

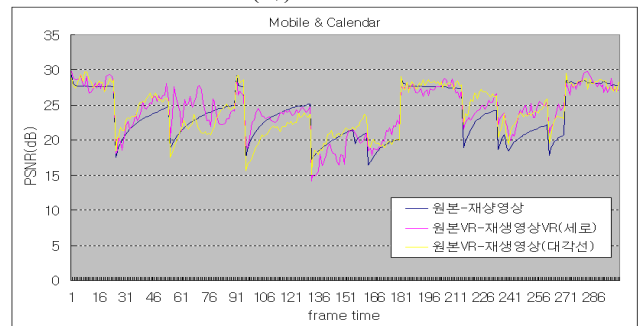
<그림 4> (라) Paris 의 경우 첫 중반부분까지 패킷손실에 의해 원본-재생영상간 PSNR 은 떨어졌으나 VR 은 손실을 잘 검출하지 못했다. 이는 손실이 발생한 부분이 VR 을 추출한 영역 이외의 영역에서 발생했거나 VR 을 추출한 화면중간 영역이 정적인 영역이라 에러감춤이 잘 됐기 때문이다. 한편 VR 을 추출한 방향(세로, 대각선)에 따른 추정 성능은 실험영상에 따라 약간 다르긴 하지만 큰 차이는 없었다.



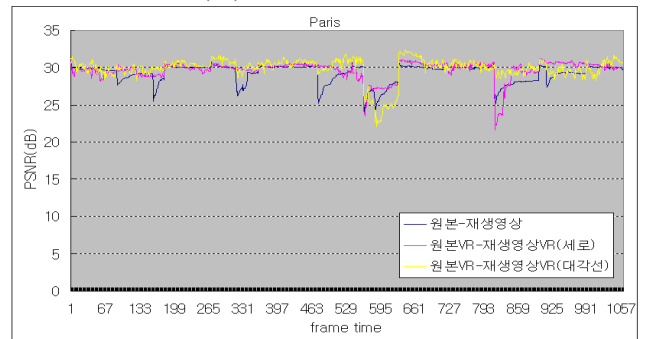
(가) Coast Guard



(나) Foreman

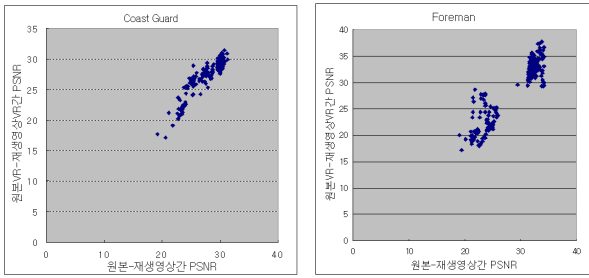


(다) Mobile & Calendar



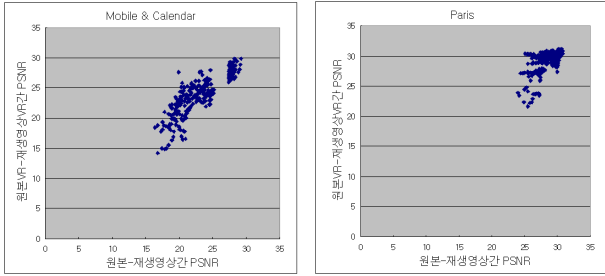
(라) Paris

<그림 4> 전송에러 발생시 원본-재생영상간 PSNR 과 VR 들간의 PSNR 비교



(가) Coast Guard

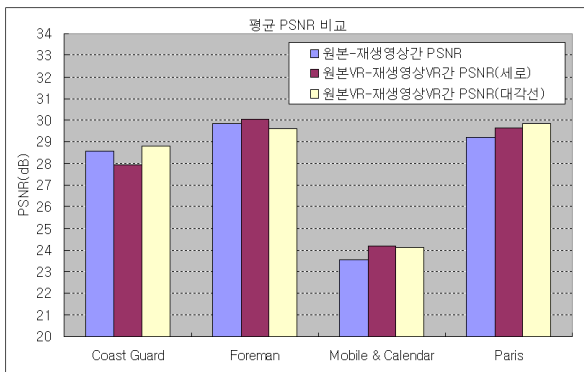
(나) Foreman



(다) Mobile & Calendar

(라) Paris

<그림 5> 원본-재생영상 PSNR 과 원본 VR-재생영상 VR PSNR 간의 상관성

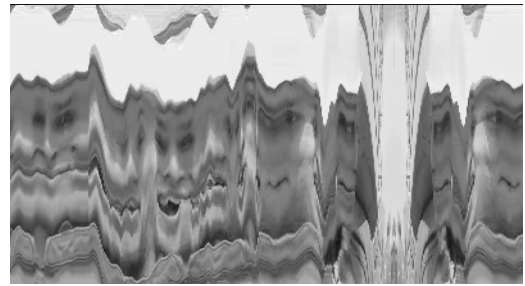


<그림 6> 평균 PSNR 의 비교

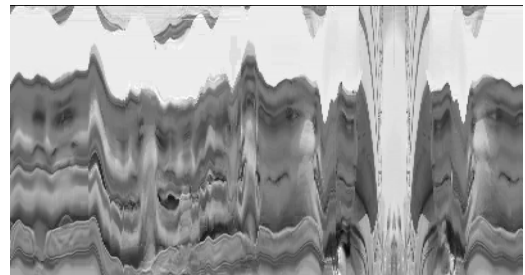
실제 IPTV 사업환경에서는 원본영상의 VR 정보보다는 인코딩된 영상의 VR 정보를 얻기가 보다 용이하다. 이 경우에는 식 (4)의 PSNR 을 직접 적용할 수 없고, 변형된 PSNR 척도를 도입하여야 한다. 제안된 방식은 PSNR 추정 뿐 아니라 식 (2)에 사용되었던  $P^{VR}_{ENC}$ ,  $P^{VR}_{STB}$  정보를 단순 비교하기만 하면 전송과정 중에 발생한 에러의 위치와 에러가 지속된 시간, 에러의 발생횟수 등에 대한 정보를 얻을 수 있으며, 무엇보다 <그림 7>과 같이 VR 들간의 차이를 확인함으로써 품질평가자가 직접 눈으로도 에러의 위치와 양을 확인할 수 있는 장점이 있다.

### 5. 결론

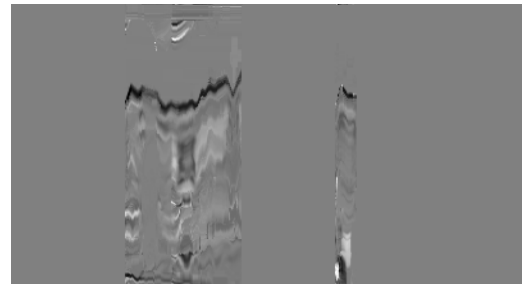
본 논문에서는 PSNR 이 다른 어떤 척도보다도 화질의 평가에 있어서 중요하다는 전제하에 비주얼리듬 정보를 이용하여 원본영상이 가용하지 않은 일반 시청자들의 STB 재생영상으로부터 PSNR 을 추정하는 방법을 제안하였다.



(가) 인코딩 재생영상의 VR



(나) 수신측 디코더 재생영상의 VR



(다) VR 차이영상

<그림 7> 인코딩 재생영상과 STB 재생영상의 비주얼리듬(VR)

VR 로부터 추정된 PSNR 값이 매 프레임 별로 2 차원 재생영상으로부터 구한 PSNR 과 정확하게 같지는 않으나 시청자들이 평균적으로 어느 정도의 화질로 시청하고 있는지 비교적 정확하게 모니터링이 가능하다. PSNR 이 커버하지 못하는 주관적인 품질양은 기존의 방법을 그대로 적용해도 된다. STB 은 단순히 재생영상으로부터 VR 정보만 추출해서 서버로 전송하는 역할만 수행하면 되고, 나머지 연산은 품질관리서버에서 수행하므로 자원이 부족한 STB 의 부담을 경감하면서도 효율적으로 품질을 모니터링할 수 있는 장점이 있다.

### 참고 문헌

- [1] 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, [www.vqeg.org](http://www.vqeg.org)
- [3] 이진호 외, “화소샘플링을 이용한 시각운동 생성방법”, 한국특허 10-1998-0056144, 2001.10.23