인터랙티브 파노라마 비디오 서비스에서 시공간 비디오 스트림 예측 필터 설계

*조용우 **석주명 *서덕영

*경희대학교, **한국전자통신연구원

yongwoo@khu.ac.kr, jmseok@etri.re.kr, suh@khu.ac.kr

Spatio-Temporal Prediction Filter Design in Interactive Panorama Video Service

*Yongwoo Cho **Joomyoung Seok *Doug Young Suh

*Kyung Hee University, **Electronics and Telecommunication Research Institute

요약

최근 방송과 통신의 융합으로 방송·통신 융합형 서비스가 활발해지고 있고, 사용자의 요구사항이 높아지고 있는 가운데 무선 채널을 이용한 인터랙티브 비디오 스트리밍 서비스는 가장 큰 서비스로 자리 잡고 있다. 인터랙티브 비디오 서비스중 하나인 파노라마 비디오는 기존의 고정적인 시청환경을 사용자가 능동적으로 선택할 수 있다는 측면에서 발전의 가능성이 큰 분야 중 하나이다. 하지만 넓은 시점을 가진 파노라마 비디오는 높은 대역폭이 요구된다는 단점이 있다. 이에 본 논문은 사용자가 파노라마 비디오 서비스를 받을 때 시청 시점을 변경시키면서 사용되는 비트율을 시공간적 필터를 사용하여 줄일 수 있는 방법을 제안한다. 이를 이용하여 고 대역폭 사용이 불가피한 파노라마 비디오 스트리밍 서비스의 요구 대역폭을 줄임으로서 인터랙티브 비디오의 스트리밍 서비스분야에서 효율적인 대역폭 사용을 위한 기술로 사용될 수 있음을 확인 할 수 있다.

1. 서론

인터랙티브 (Interactive) 비디오는 단방향의 평면적 콘텐츠가 제공하는 시점의 한계와 단일 오디오 장면을 벗어나 사용자에게 보다 많은 시점의 영상 및 오디오를 네비게이션 (navigation) 이 가능한 공간 형태로 제공한다. 그 결과 사용자는 원하는 시점과 청취점을 선택할 수 있게 된다. 첫째로는 제작과정에 따른 고정적인 시점의 시청환경에서 사용자가 관심이 있는 시점으로 선택시청이 가능한 맞춤전환시청 서비스가 가능할 것이다. 현재 멀티 앵글 서비스를 인터랙티브 비디오의 첫 번째 파일럿 서비스라고 할 수 있다.[1]

다음 단계의 비디오 서비스로는 다 수의 카메라로부터 얻은 영상을 눈(시각) 이 보는 시야에 동일하게 시공간적 연관성을 고려하여 하나의 영상 공간으로 만드는 파노라마 비디오 서비스를 예상할 수 있다.[2] 이러한 인터랙티브 비디오는 영상뿐만 아니라 영상과 관련된 정보들을 시공간적으로 결합하여 사용자에게 좀 더 사실적인 정보제공이 가능한 정보결합형 비디오로 나눌 수 있다. 이와 같이 다양한 실감 미디어 중 다수의 시점을 사실감 있게 시청할 수 있고, 고정적인 시청환경에서의 시점을 능동적으로 선택하여 볼 수 있는 인터랙티브 비디오 서비스로는 다시점인 파노라마 비디오가 사용자들의 요구를 충족시킬 수 있는 좋은 대안이 될 것이다.

그러나 현재 다시점을 가진 파노라마 비디오 시청환경은 여러 가지 제한적인 요소가 많다. 그 중에서도 대용량 비트율로 인한 고 대역폭의 필요성은 넓은 시점을 가진 파노라마 비디오를 제한된 소비환경에서 스트리밍 서비스할 때 큰 문제가 된다. 이러한 문제점을 해결하기 위해서는 인터랙티브 적인 방법을 활용한 접근이 필요하다.[3] 또한 파노라마 비디오 스트리밍 서비스를 받을 때에는 영상을 처음부터 끝까지 보는 단순한 시청 형태뿐만 아니라 고속감기(Fast Forward), 고속되감기(Fast Reverse Rewind), 일시정지(Pause), 배속재생, 임의시점 재생(random access) 등의 시간적 트릭모드(temporal trick mode) 기능을 이용하여 능동적인 시청형태가 가능하도록 해야 한다. 나아가 넓은 시청시점을 가진 파노라마 비디오의 경우, 제한된 스크린 크기와 한정된 대역폭을 효율적으로 극복하기 위한 시간적 트릭모드 뿐만 아니라 공간탐색, 확대 기능 등의 공간적 트릭모드(spatial trick mode)가 필요하다.

파노라마 비디오는 많은 시점을 가지고 있고, 고화질이므로 대용량 비트율로 인코딩 된다. 기존의 파노라마 이미지 서비스 방법과 같이 전체 시점을 모두 보내고 단말과 사용자간에 인터랙티브 트릭모드를 한다고 가정하면 동영상의 특성상 많은 대역폭 사용과 단말기의 리소스 낭비가 예상된다. 그러므로 파노라마 비디오의 공간적 트릭모드를 제공하기 위해서는 사용되는 리소스를 최소화 할 수 있는 방안에 대한 연구가 필요하다. 본 논문에서는 파노라마 비디오를 스트리밍할 때 사용자의 시점을 고려하여 시공간적으로 비디오 스트림을 예측하여 사용하는 대역폭을 감소시키는 시공간 비디오 스트림 예측 필터를 제안한다.

본 논문의 2장에서는 파노라마 비디오에 대해 기술 하고 3장에서 시공간 비디오 스트림 예측 필터를 제안한다. 4장에서는 제안하는 기술의 성능을 분석하고 5장에서 결론을 맺는다.

2. 파노라마 비디오

두 대 이상의 카메라를 통해 촬영된 영상물을 기하학적으로 교정하고 공간적으로 합성하여 여러 방향의 시점을 사용자에게 제공하는 3차원 영상처리의 한 분야인 다시점 비디오는 사용자에게 자유로운 시점을 제공할 수 있는 특징을 가진다.[4] 다시점 영상의 한 예인 파노라마 비디오 는 한 대의 카메라를 이동시키면서 획득한 여러 장의 이미지를 합성하여 하나의 영상으로 만드는 것으로서 파노라마 카메라를 이용하여 실린더 영상을 획득하거나 매우 긴 길이의 필름에 이미지를 녹화하여 하나의 영상을 만든다. 그림 1은 파노라마 비디오 서비스를 나타내고 있다.

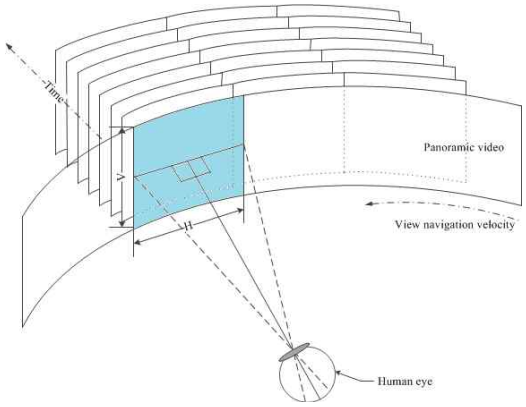


그림 1, 파노라믹 비디오 서비스

그림에서 볼 수 있듯이 파노라마 비디오 서비스는 제한된 스크린 때문에 시청영역 또한 제한적이지만 시청 시점을 자유로이 변경하여 다른 시점의 영상 또한 시청할 수 있게 된다. 시청 시점을 사용자가 임의로 자유로이 변경할 수 있게 하는 기능을 시점 네비게이션 트릭 모드라고 한다.

3. 비디오 스트림의 시공간 예측 필터 (STPF)

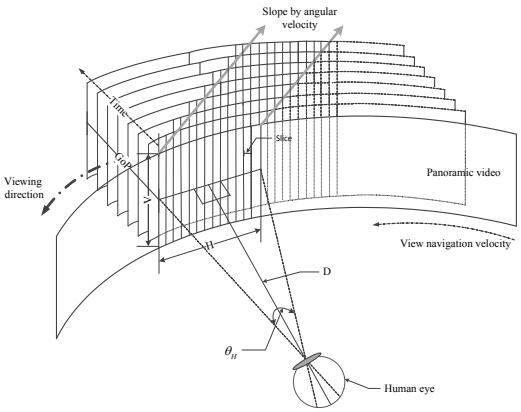
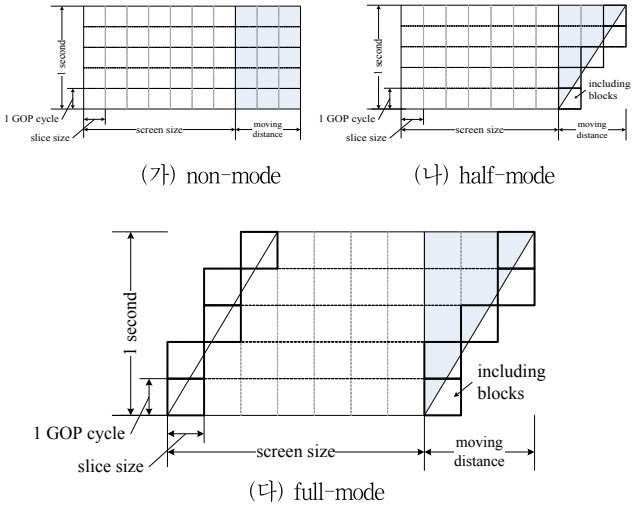


그림 2. 가변적 시점 네비게이션 속도에 따른 시공간 시청시점 예측

파노라마 비디오에서 시청 시점을 변경하는 시점 네비게이션 트릭 모드를 사용할 때, 그림 2와 같이 시점이 이동되는 속도 (V_H) 에 기인하여 시청시점 공간영역이 변하게 되는데 이때 대역폭의 절감을 위해 시점 네비게이션 트릭모드에 따른 비디오 스트림의 시공간 영역을 예측하여 해당 비디오 스트림을 요청할 수 있어야 한다. 이러한 기능을 시공간예측필터 (Spatio-Temporal Prediction Filter, STPF) 라고 정의하며 STPF 블록에서 수행한다. STPF 는 그림 3과 같이 세 가지 모드를 갖는다.



(가) non-mode (나) half-mode

(다) full-mode

그림 3. STPF 모드

그림 3의 한 블록은 슬라이스 크기와 1 GOP (Group of Pictures) 크기 [S, G] 로 구성된다. 이렇게 구성하는 이유는 복호화가 가능한 독립된 단위이기 때문이다. 따라서 그림 3은 1초 동안 필요한 $m \times n$ 개의 블록 [S_m, G_n]로 구성된다. 여기서 $m = 0, 1, 2, \ldots, h-1$ 과 $n = 1, 2, \ldots, \lceil G_{\max} \rceil ; G_{\max} = \frac{x\,fps(frame\,per\,\mathrm{second})}{yfpg(frame\,per\,GOP)}$ 것이며, 그림 3 (가)와 같이 비디오 품질 변화만을 고려하며 V_H 에 따라 변하는 시청시점 시공간 변화는 고려하지 않는 non-mode 가 있고, 그림 3 (나) 와 같이 V_H 로 인해 발생하는 시공간적인 시점이동에 대하여 새로운 시점이 나타나는 입력시점측면만 예측 필터링을 수행하는 half-mode 가 있다. 이미 지나간 시간시점의 비디오 스트림은 수신하지 않도록 필터링을 수행하는 것이다. 그리고 시점이동 방향인 출력시점 측면은 non-mode 형태와 같이 예측 필터링을 수행하지 않는다. 그로 인한 장점은 V_H 가 급속히 감소할 때 인터랙티브 지연으로 인한 끊김을 최소화할 수 있다. 그림 3 (다)와 같은 full-mode 는 속도에 따른 시점이동 변경에 대하여 입출력 모든 시점측면을 시공간적으로 예측 필터링을 수행하는 것으로 정의한다. 물론 full-mode 가 대역폭 낭비를 가장 최소화할 수 있는 방법이다. 이러한 V_H 에 따른 시공간 예측 비디오 스트림을 구성한다. 앞서 설명한 바와 같이 V_H 에 의해 발생하는 GOP 당 픽셀이동거리를 이용하여 GOP 주기별로 블록의 범위를 산출하게 된다.

제안하는 STPF 는 그림 4와 같이 V_H 에 의하여 새로운 시점이 입력되는 입력시점 영역과 이미 재생되어 사라지는 출력시점 영역을 GOP (G0<Gn Gmax) 주기별로 재생에 필요한 블록의 범위를 기준으로 필터링 하게 된다. GOP 주기별 블록의 범위가 정해지면 재생에 필요한 필수 블록들을 계산할 수 있으며, 이를 IPB(Indispensable play

blocks)라 정의한다.

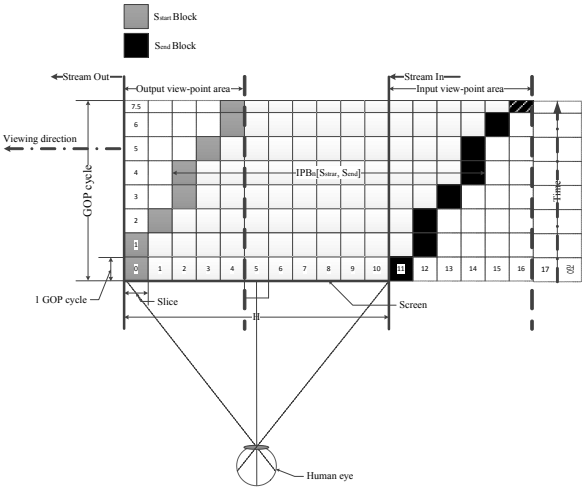


그림 4. STPF 개념

STPF의 non-mode 는 식 (1)과 같이 블록 범위를 정한다.

$$STPF[non-\mathrm{mode}] = \sum_{n=1}^{G_{\max}} IPB_n[S_{begin}, S_{end}] \begin{cases} begin = 0 \\ end = \frac{H}{S_h} + \left\lfloor \frac{MD_s}{S_h} \right\rfloor \end{cases} \quad (1)$$

여기서 H는 스크린 크기(픽셀)을 말하여 S_h 는 슬라이스 크기를 의미한다. $IPB_n[S_{begin}, S_{end}]$ 의 S_{begin} 은 n 번째 GOP 주기에서 재생에 사용되는 블록의 시작위치를 나타내며 S_{end} 는 같은 n 번째 GOP 주기에서 재생에 사용되는 블록의 끝 위치를 나타낸다. MD_S (moving distance per second)는 1 초당 움직인 픽셀거리를 의미한다. 따라서 이동이 가능한 예측거리만큼의 슬라이스를 추가하는 것으로 범위를 정할 수 있다.

식 (2)는 V_H 에 의하여 새로운 시점이 입력되는 입력시점 영역만을 필터링하는 half-mode 를 설명한다.

$$STPI_S[half-\mathrm{mode}] = \sum_{n=1}^{G_{\max}} IPB_n[S_{begin}, S_{end}] \begin{cases} begin = 0 \\ end = \frac{H}{S_h} + \left\lfloor \frac{MD_G \times G_n}{S_h} \right\rfloor \end{cases} \quad (2)$$

여기서 MD_G (moving distance per GOP)는 1 초당 움직인 픽셀거리를 의미한다. ⌊ ⌋ 는 내림(rounddown)을 의미한다. 만약 GOP 당 움직인 거리를 슬라이스로 나눈 결과가 정수가 아닌 경우는 그 해당 블록을 계속 사용해야 한다는 의미이므로 필요한 블록으로 처리한다. 여기서 ⌊ ⌋ 의 의미는 슬라이스 번호가 0부터 시작 (i.e. S_0) 하기 위한 것이다. 만약 슬라이스 번호를 1(i.e. S_1)부터 한다면 ⌈ ⌉ (roundup)으로 처리한다.

결과적으로 $IPB_n[S_{begin}, S_{end}]$ 은 n번째 GOP 주기에서의 재생에 필요한 블록의 개수를 나타내게 된다. 최종적으로 재생에 필요한 블록의 집합은 G1부터 Gmax까지 $IPB_n[S_{begin}, S_{end}]$ 의 총합으로 표현된다.

식 (3)은 V_H에 의하여 새로운 시점이 입력되는 입력시점 영역과 현재 시점이 사라지는 출력시점 영역을 모두 고려한 full-mode 방법이다. 식(1)과 (2)와 마찬가지로 GOP (G0<Gn Gmax)주기별로 블록 범위를 정하여 재생에 필요한 블록을 산출한다. 이때 출력시점 영역이 어느 블록부터 필요한지를 알기 위해서는 이전 GOP 주기 동안 이동한 거리를 고려하여야 한다. 만약 이전 GOP 주기에서 블록이 다 소비되지 않았다면, 그 블록은 현재 GOP 주기에서도 사용되어야 하기 때문에 포함하여야 한다.

$$STPI_S[full-\mathrm{mode}] = \sum_{n=1}^{G_{\max}} IPB_n[S_{begin}, S_{end}] \begin{cases} begin = 0 + \left\lfloor \frac{MD_G \times G_{n-1}}{S_h} \right\rfloor \\ end = \frac{H}{S_h} + \left\lfloor \frac{MD_G \times G_n}{S_h} \right\rfloor \end{cases} \quad (3)$$

4. STPF 성능 분석

표 1은 STPF 의 성능을 분석하기 위한 실험 시청환경에 대한 정보를 나타내고 있다.

			Roundup
Config	스크린 크기(pixels), H	704	
	V_H [deg/s]	3, 6, 12, 26	
	1deg/s 당 간거리 (pixels)	27.3	
	$S_{11} = 704/11$	64	
GOP 4	1GOP당 간거리(pixels): MD_G	43.6	44
	GOP 최대 수 ($G_{\max}$)	7.5	
	1초에 갈 예상거리: MD_S	327.3	328
	predicting additional spatial ranges : PSR	5.1	6

표1. 실험 시청환경

실험 시청환경으로 스크린 크기 H=704 pixels, V=512 pixels 로 가정하였다. GOP 구조는 IPPP 구조를 가지는 GOP=4 프레임(frames) 형태를 기준으로 인코딩 하였다. 다음 그림 5는 각 시점 네비게이션 속도 V_H 에 따라 STPF를 사용했을 경우와 사용하지 않았을 경우의 사용되는 비트율을 비교하여 보여주고 있다.

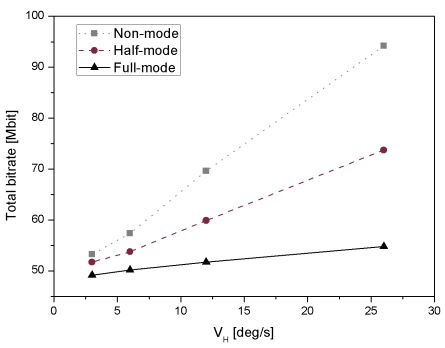


그림 5. 시점 네비게이션에 따른 각 STPF 모드별 비트율 비교

그림 5에서 볼 수 있듯이 본 논문이 제안하는 STPF를 사용하였을 때 가 사용하지 않은 non-mode 의 경우 보다 비트율을 더 적게 소모 하는 것을 알 수 있다. 시점 네비게이션 속도 V_H 가 커짐에 따라 더욱 빠르게 많은 데이터가 필요하기 때문에 사용되는 비트율이 늘어남을 알 수 있다. V_H가 26 deg/s 일 때, half-mode 의 경우에는 non-model 보다 약 22%를 절약할 수 있었고, full-mode일 경우에는 약 50%의 대역폭 이득을 얻을 수 있음을 알 수 있다.

5. 결론

가까운 미래에는 넓은 시청시야를 갖는 영상서비스의 확산으로 시점 네비게이션 트릭모드기반의 공간 트릭모드가 점점 중요하게 될 것이다. 본 논문이 제안한 STPF 는 시점 네비게이션 속도변화에 따라 필요한 비디오 스트림을 시공간적으로 예측하여 인터랙션을 요청하는 방식으로 사용자의 인식 품질의 저하 없이 파노라마 비디오 스트리밍 서비스의 사용 대역폭의 최소화를 실현 하였다. 결과적으로 본 논문에서는 파노라마 비디오 스트리밍 서비스 환경에서 불필요한 리소스 낭비를 약 50% 이상 줄일 수 있는 방법을 제안하였다. 이는 차후 파노라마 비디오 서비스 뿐만 아니라 다양한 인터렉티브 비디오의 스트리밍 서비스분야에서 효율적인 대역폭 사용을 위한 기술로 사용될 수 있음을 기대할 수 있다.

6. 감사의 글

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

7. 참고 문헌

[1] A.M. Tekalp, E. Kurutepe, M.R. Civanlar, “3DTV over IP,” IEEE Signal Process. Mag. Vol. 24 issue 6, pp.77 - 87, Apr. 2007.

[2] H. Kimata, S. Shimizu, Y. Kunita, M. Isogai, and Y. Ohtani, “Panorama video coding for user-driven interactive video application,” IEEE 13th International Symposium, Consumer Electronics, ISCE 2009, pp.112 - 114, 2009.

[3] M. Inoue, H. Kimata, K. Fukazawa, and N. Matsuura, “Interactive panoramic video streaming system over restricted bandwidth network,” ACM, Proc. of the int. conf. on Multimedia ’10, pp. 1191-1194, October 2010

[4] 호요성, 오관정, “MPEG 다시점 비디오 부호화 (Multi-view Video Coding),” pp.132-140. 2004

[5] Drive the City Streets of NYC, http://www.immersivemedia.com/demos/index.php