

MV-HEVC 와 SHVC 로 구성된 통합코덱에서 최적의 참조영상 선택방법

이재영¹, 강정원², 이하현³, 이진호⁴, 한중기⁵세종대학교^{1,5}, 한국전자통신연구원^{2,3,4}lly321456@gmail.com, jungwon@etri.re.kr, hanilee@etri.re.kr, jinosoul@etri.re.kr, hjk@sejong.edu

Optimal reference picture selection method in the combined codec of Multi-view HEVC and SHVC

Jae-Yung Lee¹, Jung-Won Kang², Ha-Hyun Lee³, Jin-ho Lee⁴, Jong-Ki Han⁵Sejong University^{1,5}, ETRI^{2,3,4}

요 약

실감형 디스플레이 장치들을 지원하기 위해 다시점 영상을 효율적으로 코딩하는 방법은 다시점 비디오 코딩 방법과 다양한 전송 환경과 다양한 종류의 단말들을 통합적으로 지원할 수 있는 스케일러블 비디오 코딩 방법은 존재하지만, 현재 유비쿼터스 환경에서 실감형 서비스를 지원하면서 동시에 다양한 환경을 지원할 수 있는 통합된 동영상 정보를 생성하는 방법은 존재하지 않는다. 따라서 본 논문에서는 통합된 동영상 정보를 생성하는 다중 확장 비디오 코딩의 개념과 동작에 대해 설명하고 최적의 inter layer 참조 영상 선택 방법에 대해 분석한다.

1. 서론¹

동영상 코딩 기술의 발전을 통해 사용자가 고화질, 고해상도의 동영상 정보를 접할 수 있게 되었지만 이들은 단지 2 차원 평면 디스플레이 장치를 지원하기 위한 방법일 뿐이며 사용자에게 입체감을 주거나 시점(View)을 자유롭게 선택할 수 있도록 하는 실감형 서비스(Realistic service)를 고려하지 않았다. 그리고 앞으로는 방송/통신의 융합 및 유/무선의 융합 등의 과정을 통해 다양한 전송 환경 속에서 다양한 종류의 단말들이 혼재하는 유비쿼터스(Ubiquitous) 환경에서 동영상 정보를 효율적으로 전달할 수 있어야 한다. 하지만 이들 동영상 코딩 방법은 각각의 응용에서 특정한 전송 환경과 특정한 단말을 대상으로 한정된 코딩을 수행하도록 설계가 되어 있기 때문에 효율적인 처리가 불가능하다. 만약 기존의 동영상 코딩 방법을 이용하여 하나의 동영상 정보를 서비스한다고 가정하면

다양한 전송환경과 단말들을 지원하기 위해 반복적으로 인코딩 과정을 수행해야하는 복잡하고 어려운 작업이 될 것임을 짐작할 수 있다. 이러한 환경의 변화 속에서 앞으로의 동영상 코딩은 스테레오스코픽(Stereoscopic) 영상 또는 다시점(Multi-view) 영상을 통해 사용자가 입체감을 느낄 수 있도록 하는 실감형 서비스를 지원할 수 있어야 함은 물론 다양한 전송 환경과 단말들을 통합적으로 지원하기 위한 동영상 정보를 생성하고 효율적으로 전달, 상호 교환할 수 있는 방법이 함께 실현되어야 한다.

2. 기존의 스케일러블 코덱

2.1 Scalable HEVC(SHVC)

SHVC[1]는 계층적 부호화를 위해 Base layer 와 Enhancement layer 가 존재하며 각 레이어를 구별하기 위해서 Layer ID 를 사용한다. SHVC 는 Enhancement layer 를 부호화 할 때, Base layer 의 영상을 확대하여 참조 영상으로 사용한다. Base layer 는 독립적인 HEVC[2]를 사용하여 부/복호화 하고, Enhancement Layer 는 이미 복호된 Base Layer 의 정보를 추가적으로 사용해서 부/복호화 한다.

2.2 Multi-view HEVC(MV-HEVC)

¹ 본 연구의 일부는 미래창조과학부 및 정보통신산업진흥원의 IT 융합 고급인력과정 지원사업(NIPA-2014-H0401-14-1006)의 연구결과로 수행되었고, 그리고 일부는 미래창조과학부가 지원하는 2014 년 정보통신·방송(ICT) 연구개발사업의 연구결과로 수행되었음.

MV-HEVC[3]는 다시점 부호화를 위해 Base view 와 Dependent view 가 존재하며 각 view 를 구별하기 위해 View ID 를 사용한다. Dependent view 의 영상을 부호화 할 때, Base view 의 영상을 참조 영상으로 사용한다. 참조 영상으로 사용하는 Base view 의 영상은 현재 부호화하는 Dependent view 의 POC 와 같은 POC 의 영상을 사용한다. Base view 는 독립적인 HEVC 를 사용하여 부/복호화 하고, Dependent view 는 이미 복호된 Base view 의 정보를 추가적으로

사용해서 부/복호화 한다.

3. MV-HEVC/SHVC 통합 코덱에서 최적의 참조 영상 방법

3.1 MV-HEVC/SHVC 통합 코덱의 구조

그림 1 은 MV-HEVC/SHVC 통합코덱의 부호화 구조이다. 각 레이어는 VPS 에 부호화된 scalability 타입과 종속성에

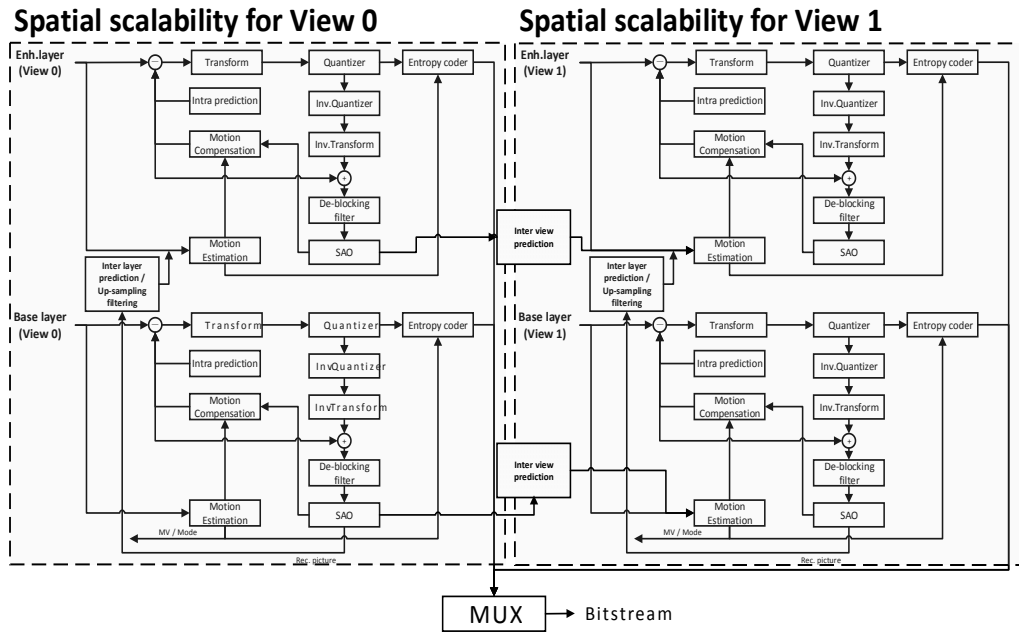


그림 1. The structure of MV-HEVC/SHVC combined codec

따라 Spatial scalability 또는 View scalability 를 사용하여 부호화 한다.

MV-HEVC/SHVC 통합코덱은 다른 레이어의 영상 중 이미 복호되고 POC 가 같은 영상의 정보를 현재 영상을 부호화 하기 위한 정보로 사용한다. 기존의 SHVC 에서는 Base layer 의 영상 정보를 사용하여 Enhance layer 을 부호화하였다면, MV-HEVC/SHVC 통합코덱은 Base layer 의 영상 정보뿐만 아니라 Base layer 가 아닌 이미 복호된 다른 시점의 영상 정보도 현재 영상을 부호화 하기위해 사용한다.

SHVC 처럼 inter layer reference 영상 중에서 현재 레이어의 DPB 의 저장해야하는 영상이 존재하는 경우, 복호기의 DPB(Decoded Picture Buffer) 크기가 커야하는 단점이 있다. 그러므로 통합 코덱의 Enhance layer 는 layer dependency 의 수에 상관없이 그림 3 과 같이 N 개의 제한된 참조 영상의 수를 사용해야 한다.

제한된 참조 영상의 수를 사용하는 Enhance layer 들은 참조하는 Inter layer reference 의 수만큼 적은 Temporal reference 영상을 참조 영상으로 사용한다.

3.2 MV-HEVC/SHVC 통합 코덱의 참조 영상 관리

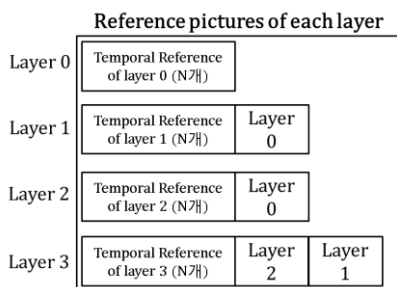


그림 2. MV-HEVC/SHVC 통합 코덱에서 참조 영상 목록

그림 2 와 같이 Enhance layer 의 layer dependency 수에 따라 사용 가능한 참조 영상의 수는 증가한다. 그러나, 참조 영상의 수가 증가하는 것은 부호기의 복잡도 증가하고

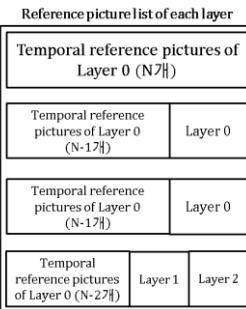


그림 3. 통합 코덱에서 제한된 수의 참조 영상 목록

3.3 최적의 inter layer 참조 영상 선택 방법

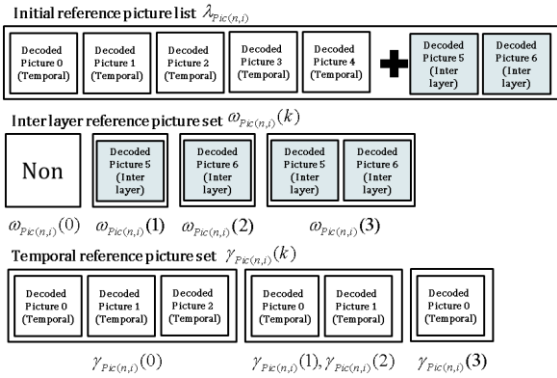


그림 4. 참조 영상 목록의 구성의 예

그림 4 와 같이 통합 코덱에서 n 번째 AU (Access Unit) 의 i 번째 레이어의 초기 참조 영상 목록은 $\lambda_{Pic(n,i)}$ 이고, 모든 inter layer 참조 영상에 대하여 조합 가능한 k 번째 inter layer 참조 영상 집합을 $\omega_{Pic(n,i)}(k)$ 라 할 때, 제한된 참조 영상의 수를 N 일 경우, 아래와 같이 temporal reference 영상의 수와 inter layer reference 영상의 수의 합은 항상 N 보다 작거나 같다.

$$|\gamma_{Pic(n,i)}(k) \cup \omega_{Pic(n,i)}(k)| \leq N \quad (1)$$

따라서, temporal 참조 영상의 집합 $\gamma_{Pic(n,i)}(k)$ 의 크기는 $\omega_{Pic(n,i)}(k)$ 의 크기에 따라 결정된다. 또한 $\lambda_{Pic(n,i)}$ 를 구성하는 j 번째 참조 영상을 $\hat{f}_{n,i,j}^T$ 이라 하면, $\gamma_{Pic(n,i)}(k)$ 의 구성은 아래와 같다.

$$\gamma_{Pic(n,i)}(k) = \bigcup_{j=0}^{N-|\omega_{Pic(n,i)}(k)|} \hat{f}_{n,i,j}^T \quad (2)$$

따라서, $\omega_{Pic(n,i)}(k)$ 에 따른 참조 영상 목록 $\psi_{Pic(n,i)}(k)$ 는 아래와 같이 정의된다.

$$\psi_{Pic(n,i)}(k) = \gamma_{Pic(n,i)}(k) \cup \omega_{Pic(n,i)}(k) \quad (3)$$

그러므로 최적의 inter layer 참조 영상 집합을 구하는 방법은 아래와 같다.

$$\omega_{Pic(n,i)}(k^*) = \arg \min_{\{\forall k\}} \{J_{Pic(n,i)}(\psi_{Pic(n,i)}(k))\} \quad (4)$$

위 식에서 $J_{Pic(n,i)}(\psi_{Pic(n,i)}(k))$ 는 n 번째 AU 의 i 번째 레이어의 픽처에서 참조 영상 목록 $\psi_{Pic(n,i)}(k)$ 를 사용했을 때 발생하는 RD-cost 를 나타낸다.

4. 실험 결과

통합 코덱의 구현은 SHVC 참조 소프트웨어 3.1 버전 [4] 에 구현했으며, 실험 조건은 아래 표 1 과 같다. 본 실험에서 총 레이어는 4 개이며, 2 개의 해상도와 2 개의 시점에 대해 실험하였다. MSVC 의 성능은 레이어 3 에 대해서 BD-rate [5] 를 이용하였으며, 실험 조건은 JCT-VC 와 JCT-3V 에서 관장하는 실험 조건 [6] [7] 중 Random access 경우만을 실험 하였다.

표 1. 실험 조건

Coding structure	Random access
QP setting	Base layer : 25/30/35/40 View dependent layer : Base layer QP + 2 Spatial dependent layer : Base layer QP+ 3
Intra period	32 (Refresh type: CRA picture)
Restricted reference num	GOP ID1 : 4 GOP ID2 : 2 GOP ID3 : 2 GOP ID4 : 2 GOP ID5 : 2 GOP ID6 : 2 GOP ID7 : 2
Number of frames	half sequence
Test Sequence	Balloons(1024x768) - Natural sequence Kendo(1024x768) - Natural sequence Newspaper(1024x768) - Natural sequence GTFly(1920x1088) - Graphic sequence Dancer(1920x1088) - Graphic sequence PoznanHall2(1920x1088) - Natural sequence PoznanStreet(1920x1088) - Natural sequence

아래 표 2 는 최적의 inter layer reference 영상 선택 방법의 성능을 나타내는 것으로 평균 22%의 부호화 효율 향상이 있었다. 그러나 부호화 복잡도는 370%로 증가했는데 이는 멀티 패스 코딩을 통해서 최적의 inter layer reference 를 결정하기 때문이다.

표 2. 최적의 interlayer reference 영상 선택 방법의 성능

Sequence	BD-rate		
	Y	U	V
1024x768	-21.6%	-21.6%	-26.0%
1920x1088	-32.3%	-25.0%	-38.2%
Graphics Sequence	-12.1%	-8.1%	-8.0%
Average	-22.0%	-21.6%	-26.00%
Enc Time [%]	370.1%		

5. 결론

최적의 inter layer reference 를 선택하는 방법은 높은 부호화 효율 향상을 보이지만 부호화의 복잡도가 많이 증가한다. 따라서, 복잡도를 감소하고 성능 저하가 적게 발생하는 빠른 inter layer reference 선택 방법이 필요하다.

참고문헌

[1] J. Chen, J. Boyce, Y. Ye, and M. M. Hannuksela, "High efficiency video coding (HEVC scalable extension draft 3," in Joint Collaborative Team on Video Coding of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JCTVC-N1008, Vienna, AT, Aug., 2013.

[2] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/ WG11, "High Efficiency Video Coding (HEVC) Text Specification Draft 10 (for FDIS & Consent)," JCTVC-L1003, Geneva, CH, Jan. 2013.

[3] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/ WG11,

- “ MV-HEVC Draft Text 5,” JCT3V-E1004, Vienna, AT, July. 2013.
- [4] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/ WG11, “ Scalable High Efficiency Video Coding Test Model 3 (SHM 3),” JCTVC-N1007, Vienna, AT, July. 2013.
- [5] G. Bjontegarrd, “ Calculation of average PSNR differences between RD curves,” in ITU-T SC16/Q6 13th VCEG meeting, No. VCEG-M33, Austin, TX, Apr. 2001.
- [6] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/ WG11, “ Common SHM test conditions and software reference configurations,” JCTVC-N1009, Vienna, AT, July. 2013.
- [7] ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/ WG11, “ Common Test Conditions of 3DV Core Experiments,” JCT3V-E1100, Vienna, AT, July. 2013.