

공간 웨이블릿 변환의 복잡도를 줄인 스케일러블 비디오 코딩에 관한 연구

Scalable Video Coding with Low Complex Wavelet Transform

*박 성 호, *김 원 하, **정세윤
 *Seong Ho Park, *Won Ha Kim, **Seyoon Jeong
 *경희대학교 전자정보대학
 pncc77@hotmail.com, wonha@khu.ac.kr
 ***한국 전자통신연구원 디지털방송연구단 방송미디어연구그룹
 jsy@etri.re.kr

Abstract - In the decoding process of interframe wavelet coding, the inverse wavelet transform requires huge computational complexity. However, the decoder may need to be used in various devices such as PDAs, notebooks, PCs or set-top Boxes. Therefore, the decoder's complexity should be adapted to the processor's computational power. A decoder designed in accordance with the processor's computational power would provide optimal services for such devices. So, it is natural that the complexity scalability and the low complexity codec are also listed in the requirements for scalable video coding. In this contribution, we develop a method of controlling and lowering the complexity of the spatial wavelet transform while sustaining almost the same coding efficiency as the conventional spatial wavelet transform. In addition, the proposed method may alleviate the ringing effect for certain video data.

Key Words : scalable video coding, wavelet based video coding, 스케일러블 동영상 압축, 웨이블릿 기반 동영상, 동영상 압축

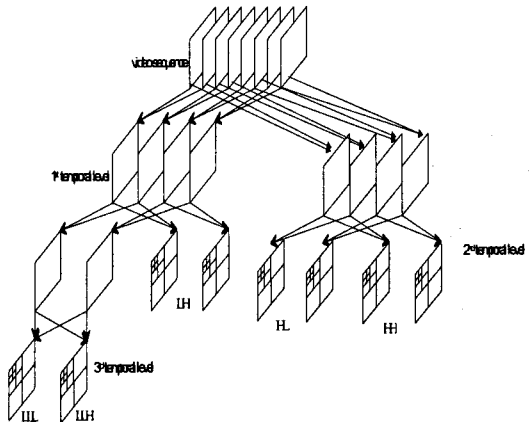
1. 개요

웨이블릿 기반 동영상 부호화에서 동영상 데이터들은 시간 축에서 블록 기반의 움직임 추정과 보상을 통해 시간적 연관성이 제거 된다. 또한 효율적인 공간적 scalability를 위해 Wavelet을 이용하여 공간적으로 분해한다. Wavelet 기반 비디오 코딩의 주된 장점은 시간적 중복성을 제거하는데 반복적인(recursive) 구조를 사용하지 않는다는 것이다. 이것은 시간적인 필터링과 공간 변환이 독립적으로 수행된다는 것을 의미한다.[1][2][3]

Wavelet 기반 Scalable 동영상 부호화에서는 움직임 보상 시간 필터링(MCTF : Motion Compensated Temporal Filtering)에서 두개 이상의 프레임들이 시간적으로 필터링된다. 두개의 프레임(A, B)만으로 필터링을 할 경우 각각에 의해 L(저주파 혹은 평균값)프레임과 H(고주파 혹은 차이값)프레임이 생성되며 이 과정을 L 프레임에 대해서만 반복적으로 수행하여 효율적으로 시간적 중복성을 제거하게 된다.[1][2] MCTF과정을 통과한 프레임들은 각각 웨이블릿 변환을 이용하여 공간적으로 분해되고[1][2][3], 각 프레임의 웨이블릿 변환 계수들이 양자화와 엔트로피 코딩을 거친다. 여기서 시간적 필터링과 공간 변환이 맞물려 수행되지 않고 순차적으로 수행된다.

다음 <그림 1>에서 인터 프레임 웨이블릿 부호화의 개념도를 나타내었다. 그림에서는 GOF 크기는 8이며, Haar 필터링

에 의한 3-level MCTF를 수행 하였다.

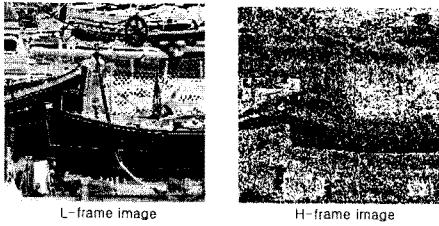


<그림 1> 인터프레임 웨이블릿 부호화의 개념도

2. Frame에 따른 영상 신호 분석

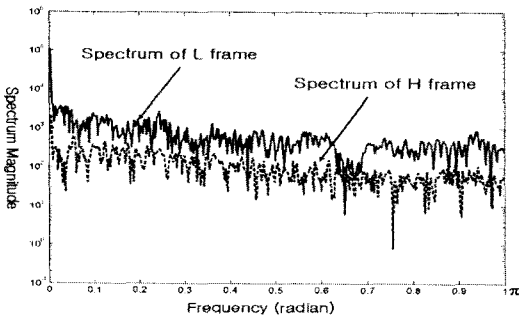
MCTF의 주된 목적은 영상 시퀀스 사이의 시간적 중복성을 제거하여 GOF의 영상 정보를 L프레임으로 모으는데 있다. 따라서 L 프레임의 영상 신호는 원래의 영상신호와 유사한 반면 H 프레임의 영상 신호는 잔여 영상 신호만을 포함하고 있다. <그림 2>에서 'Sailormen' 영상의 L프레임 영상

과 H프레임 영상을 비교하였다. 여기서 H 프레임의 신호 크기를 10배 하였다.



<그림 2> 실제 L-frame과 H-frame의 영상 신호

영상 신호의 스펙트럼도 L 프레임의 경우 원영상의 스펙트럼과 유사하지만 H 프레임의 경우 전 주파수 대역으로 퍼져있어 잡음과 비슷한 특성을 보인다. 따라서 H 프레임의 경우 신호의 에너지 밀집 주파수 대역을 선별하기 위해 길이가 긴 웨이블릿 필터를 사용할 필요가 없다. <그림3>은 L프레임과 H프레임의 스펙트럼을 비교하였다.



<그림 3> L-frame과 H-frame 영상신호의 스펙트럼

3. 프레임 적응적 공간 웨이블릿 변환

영상신호는 저주파 대역에 많은 에너지와 영상 정보를 포함하므로 원영상과 유사한 L프레임은 보다 나은 에너지 밀집(compaction)이 필요하다. 그러나 MCTF에 의해 고주파 영상신호를 갖게된 H프레임의 경우 주파수 대역에 따른 분해에 대한 이득이 적어 긴 길이의 필터를 사용하는 것과 짧은 길이의 필터를 사용하는 것에 별반의 차이가 없으며, 짧은 길이의 필터를 사용한다면 오히려 영상의 경계에서 'ringing' 현상을 줄이는 추가적인 이득을 얻을 수 있다.[5]

본 논문에서는 잡음과 비슷한 특성을 갖는 프레임에 대한 공간 웨이블릿 변환에 대해 기존의 9x7 필터를 대신하여 짧은 길이의 필터를 사용함으로써 영상의 화질에는 크게 영향을 미치지 않고 웨이블릿 변환의 복잡도를 줄이는 방법을 제안한다.

여기에 사용된 방법은 MPEG SVC의 MC-EZBC를 참조 S/W로 사용하여 프레임에 적응적으로 에너지가 높은 프레임은 9-7필터를, 에너지가 낮은 프레임에 대해서는 Haar 필터를 적용하여 공간 웨이블릿 변환을 수행하였다.

이 S/W에서는 unconnect 화소의 수에 따라 장면 전환을 판별하여 시간적 필터링을 적용하고 있다. 이는 장면 전환 시 시간적 필터링을 하지 않고 두 프레임을 그대로 다음 단계로 전달하기 때문에 H프레임에 L프레임과 같은 원영상이 전달된다. 이런 경우 공간 변환 시 길이가 긴 필터를 H프레임에 적용하여야 한다. 그러나 장면 전환이 아닌 시간적 필터링이 적용된 H프레임에 대해서는 이미 논한 바와 같이 잡음과 같은 특성을 지니기 때문에 짧은 길이의 필터를 사용하여 공간적 변환의 복잡도를 줄인다.

대부분의 연속되는 영상들은 시간적으로 밀접한 연관성을 지니기 때문에 대부분의 H 프레임에 대하여 길이가 짧은 필터를 적용할 수 있다.

YUV의 비율이 4:2:0의 형식을 갖는 영상에서 제안된 공간 웨이블릿 변환의 곱셈 수와 기존 방법의 곱셈 수는 다음과 같이 계산한다.

$$C_{traditional} = N \cdot \sum_{l=0}^L X \cdot Y \cdot \left(\frac{1}{4} + \frac{2}{4^{l+2}} \right) \cdot (l^{long} + h^{long})$$

$$C_{proposed} = (N - H) \cdot \sum_{l=0}^L X \cdot Y \cdot \left(\frac{1}{4} + \frac{2}{4^{l+2}} \right) \cdot (l^{long} + h^{long}) + H \cdot \sum_{l=0}^L X \cdot Y \cdot \left(\frac{1}{4} + \frac{2}{4^{l+2}} \right) \cdot (l^{short} + h^{short})$$

여기서,

N : Number of frames in the GOP that are temporally filtered together

L : Number of levels in spatial wavelet transform .

X : Horizontal frame size, Y : Vertical frame size

H : Number of frames containing residual signals. That is, H is the number of H-frames from slowly varying image sequences.

H is calculated during MCTF.

l^{long} : Length of low-pass wavelet filter used for L-frames (or A-frames) and H-frames from fast varying image sequences.

h^{long} : Length of high-pass wavelet filter used for L-frames (or A-frames) and H-frames from fast varying image sequences.

l^{short} : Length of low-pass wavelet filter used for H-frames containing residual signals.

h^{short} : Length of high-pass wavelet filter used for H-frames containing residual signals.

4. 실험 결과 및 고찰

기존의 모든 프레임에 9/7필터를 적용한 방법과 H-frame에 Haar 필터를 사용한 방법을 비교한다. 다음의 <표 1>과 <표 2>는 'CITY'와 'CREW'영상에서 Wavelet 변환의 복잡도를 비교한다.

Size	Frame rate	Conventional scheme		Proposed scheme		CI (%)
		# of multiplications	# of using (9-7) filter	# of multiplications	# of using (9-7) filter	
704x576	30	29080000	300	9970320	32	33.03
352x288	30	7206400	300	2392840	32	33.03

표 1 'CITY'영상의 웨이블릿 변환 복잡도

Size	Frame rate	Conventional scheme		Proposed scheme		CI (%)
		# of multiplications	# of using (9-7) filter	# of multiplications	# of using (9-7) filter	
704x576	30	232659200	240	690703200	15	336.84
352x288	30	581644800	240	172675800	15	336.03

표 2 'CREW'영상의 웨이블릿 변환 복잡도

여기서

$$CI = \frac{NMC}{NMP} \times 100 \quad (\%)$$

NMC : 기존방법의 곱셈 수, NMP : 제안방법의 곱셈 수

다음의 <표 3>과 <표 4>는 PSNR 결과를 비교한다.

Size	Frame rate (Hz)	tps (Kbit/s)	Conventional scheme			Proposed scheme		
			Y-SNR mean	U-SNR mean	V-SNR mean	Y-SNR mean	U-SNR mean	V-SNR mean
704x576	30	3000	36.18	44.35	46.16	36.14	44.61	46.50
704x576	30	1500	33.63	42.78	44.68	33.62	43.03	44.99
352x288	30	750	31.58	42.75	44.35	31.51	43.07	44.54
352x288	30	384	29.81	40.88	42.30	29.81	40.85	42.25

표 3 'CITY'영상의 PSNR 비교

Size	Frame rate (Hz)	tps (Kbit/s)	Conventional scheme			Proposed scheme		
			Y-SNR mean	U-SNR mean	V-SNR mean	Y-SNR mean	U-SNR mean	V-SNR mean
704x576	30	3000	36.16	39.20	39.15	35.89	39.81	39.77
704x576	30	1500	34.08	37.01	36.15	34.00	38.32	37.59
352x288	30	750	32.98	35.46	33.60	32.94	37.09	35.16
352x288	30	384	30.04	32.26	29.44	29.94	32.72	30.80

표 4 'CREW'영상의 PSNR 비교

<그림 4>는 같은 bit rate에서 영상의 화질을 비교한다.

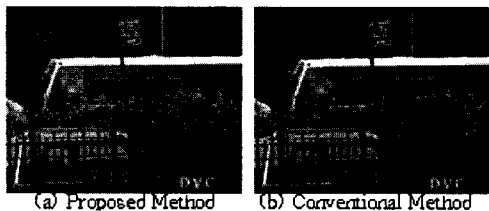


그림 4 동일 비트율에서 영상 화질 비교

위의 분석에서 보듯이 제안한 프레임에 적용적인 Wavelet 변환 기법은 대부분의 영상에서는 기존의 방법에 비하여 화질 차이 거의 내지 않으면서도 300%정도의 공간 웨이블릿 변환의 복잡도를 줄이고, 일부 영상들에('CITY', 'MOBILE') 대해서 'ringing' 현상을 줄이는 것을 확인하였다.

감사의 글

본 연구는 한국 과학재단 목적 기초연구 (R05-2003-000-11653-0) 지원으로 수행되었음.

참고 문헌

- [1] J.R. Ohm, "Complexity and Delay Analysis of MCTF Interframe Wavelet Structures", Contribution to the MPEG-meeting, July 2002.
- [2] S.-J. Choi and J. W. Woods, "Motion compensated 3-D subband coding of video," IEEE Trans. Image Proc., vol. 8, no. 2, February 1999.
- [3] M. Schaar and D. Turaga, "Wavelet coding for video streaming using new unconstrained motion compensated temporal filtering", Contribution MPEG-M8675, Kagenfurt, July 2002.
- [4] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, "Image coding using Wavelet transform", IEEE trans. Image Processing, vol. 11, pp674-693, July 1989.
- [5] G. Strang and Truong Nguyen, "Wavelets and Filter Banks", Wellesy-Cambridge Press, 1997
- [6] 3D MC-EZBC Software package "MPEG-CVS"
- [7] J. R. Ohm, "Three-dimensional subband coding with motion compensation," IEEE Trans. Image Proc., vol. 3, no. 5, September 1994.
- [8] S.-Ta Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," Signal Processing: Image Communication vol. 16, pp. 705-724, 2001