

비디오 코딩을 위한 다중 잔차신호 플리핑 및 변환 방법

김남욱, *강정원, *임성창, **이영렬

세종대학교, *한국전자통신연구원, **세종대학교

nukim@sju.ac.kr, jungwon@etri.co.kr, sclim@etri.co.kr, **yllee@sejong.ac.kr

Multiple Transform and Residual Flipping for Video Coding

Nam uk kim, Jungwon kang, Sung-Chang lim **Yung-Lyul Lee

sejong university *sejong university

요 약

비디오 압축에서, 변환은 데이터를 공간 영역에서 주파수 영역으로 변환 함으로써 에너지 압축에서 중요한 역할을 수행한다. HEVC(High Efficiency Video Coding)에서는 DCT-II(Discrete Cosine Transform type-II)를 사용하여 잔차신호 변환을 수행한다. DCT-II 는 픽셀간 상관도가 높은 신호일수록 높은 에너지 집중도를 보이지만, 픽셀간 상관도가 비교적 낮은 블록일수록 낮은 에너지 집중도를 보인다. 본 논문에서는 DST-VII(Discrete Sine Transform type-VII) 및 잔차신호 플리핑을 사용하여 다양한 변환 방법으로 영상을 부호화 및 복호화 하는 알고리즘에 대해 제안한다. 다양한 변환 방법은 부호화기에서 블록단위로 1 가지를 선택하여 비트스트림으로 선택된 방법에 대한 정보를 전송한다. 제안된 방법은 HEVC 대비 약 2.47%의 BD-rate 감소를 보인다.

1. 서론

HEVC(High Efficiency Video Coding) [1]은 가장 최신의 비디오 부호화 표준 기술로 ITU-T 와 VCEC 에서 공동으로 표준화를 진행하여 2013 년도 표준화가 완료되었다 HEVC 는 영상을 우선 CTU(Coding Tree Unit)라는 64×64 크기의 정사각형 블록으로 분할한다. CTU 들은 다시 쿼드트리 구조로 CU(Coding Unit)로 분할된다. CU 는 화면 내 예측을 사용할 것인지, 화면 간 예측을 사용할 것인지 정하는 최소단위이다. CU 는 다시 TU(Transform Unit)로 분할될 수 있다. TU 는 변환을 수행하는 기본 단위이다. CU 는 또한 PU(Prediction Unit)로 분할될 수 있다. PU 는 예측을 수행하는 기본단위이다. HEVC 의 압축과정은 예측, 변환, 양자화, 엔트로피 부호화 순서로 구성된다. 예측은 주변으로부터 현재 블록을 예측하는 과정으로 화면 내 예측, 화면 간 예측 두 가지 방법이 존재한다. 두 가지 예측 방법 중 하나를 선택해 예측신호를 생성한 후 원본신호로부터 빼주어 잔차신호를 생성한다. 이후 변환 양자화를 통해 주파수도메인에서 압축을 수행하며 변환 양자화과정에서 출력되는 변환계수들은 양자화되어 엔트로피 부호화를 통해 비트 스트림으로 변환된다. HEVC 는 화면 내 예측의 잔차신호에 대해서는 8×8 크기 이상의 TU 인 경우는 DCT-II(Dicrete Cosine type-II) [2], 4×4 크기의 TU 인 경우는 DST-VII(Discrete Sine Transform type-VII)를 사용하여 변환을 수행한다. 잔차신호들의 상관관계가 비교적 높은 큰 크기의 TU 에서는 DCT-II 변환이 효율적이며, 잔차신호들의 상관관계가 비교적 낮은 4×4 크기의 TU 에서는 DST-VII 변환이 효율적이다. 하지만 이러한 변환기저 선택은 평균모델을 기반으로 하였으며, 실제 잔차신호들은 임의적인 특성이 높은 편이기 때문에 아닌 경우가 많이 발생한다.

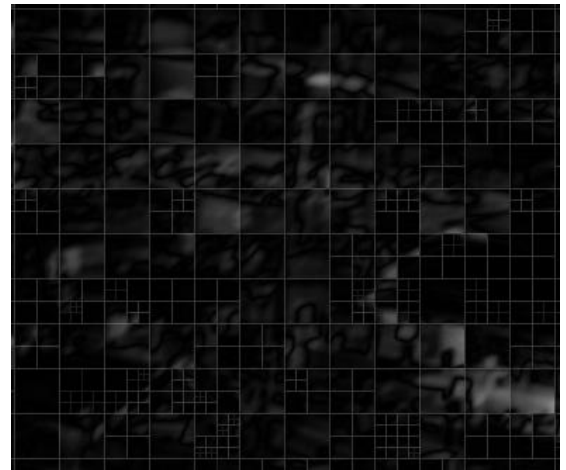


그림 1. RollerCoaster(2160p) 영상의 잔차신호

그림 1 은 3840×2160 해상도인 RollerCoaster 영상의 잔차신호의 일부이다. HEVC 로 부호화 되었으며 흰 선들은 CU 분할 구조를 표현하며 보다 얇은 흰 선들은 TU 분할구조를 표현한다. 잔차신호는 절대값이 클수록 밝은색으로 표현되며 절대값이 작을수록 어두운 색으로 표현된다. 그림에서 보면 잔차신호들이 CU, TU 경계의 특정한 모서리나, 꼭지점에 많이 몰려 있는 것을 확인 할 수 있다. 대부분 우측하단의 꼭지점에 많이 몰려 있지만 다른 방향에 몰려있는 경우도 발생하는 것을 확인 할 수 있다. 이러한 경우에는 DCT-II 를 사용하여 변환을 하는 것보다는 DST-VII 와 같은 비대칭 변환 기저를 사용하는

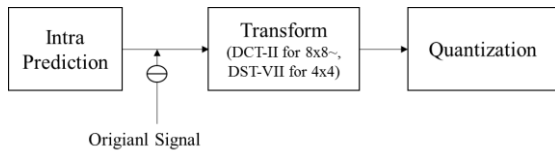


그림 2. HEVC 의 예측 부호화 과정

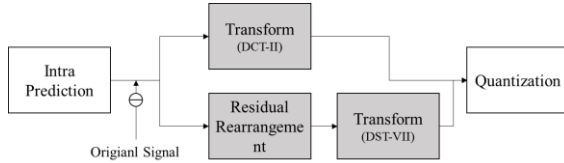


그림 3. 제안하는 예측 부호화 과정

것이 효율적이다. 본 논문에서는 잔차신호들이 한 곳으로 몰려 있는 특성을 고려하여 화면 내 예측된 TU 에 대해 변환을 수행할 때 DCT-II 와 DST-VII 중 하나를 선택하여 변환을 수행하는 방법을 제안한다. 그리고 잔차신호가 집중되는 위치를 맞추기 위해 DST-VII 로 변환을 수행하기 전에 잔차신호를 플리핑 하는 방법을 제안한다. 본 논문의 2 장에서는 제안하는 방법에 대해 설명하고, 3 장에서는 실험 결과 그리고 마지막 4 장에서는 결론으로 이 논문을 마친다.

2. 제안하는 방법

그림 2 는 기존의 HEVC 예측 및 부호화 과정을 보여준다. 예측 된 신호를 원본 신호로부터 빼 주어 잔차신호들을 생성하고, 그 잔차신호들에 대해 TU 단위로 변환을 수행하는데 8x8 크기 이상의 TU 에 대해서는 DCT-II 2 차원 변환을 수행하고, 화면 내 예측이며 4x4 크기의 TU 에 대해서는 DST-VII 2 차원 변환을 수행한다. 그리고 변환된 계수를 양자화와 엔트로피 부호화를 통해 비트 스트림으로 변환한다. 그림 3 은 제안하는 HEVC 예측 및 부호화 과정을 보여준다. 제안하는 방법은 음영된 블록으로 표시하였다. 변환을 수행할 때 TU 의 크기와 상관 없이 DCT-II 이차원 변환과 DST-VII 이차원 변환 중 하나를 선택하여 변환을 수행한다. 그리고 DST-VII 이차원 변환을 수행할 때는 변환 전에 잔차신호들의 재정렬을 수행한다. 잔차신호 재정렬은 잔차신호들을 신호의 세기가 큰 영역을 TU 의 오른쪽의 아래의 모서리로 이동시키기 위해 사용된다.

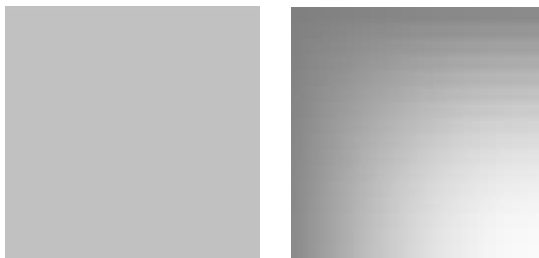


그림 4. DCT-II(왼쪽)와 DST-VII(오른쪽) 이차원 변환기저의 첫번째 기저

그림 4 는 DCT-II 이차원 변환기저의 첫번째 기저와 DST-VII 이차원 변환기저의 첫번째 기저를 보여준다. DCT-II 는 그림과 같이 평균적인 특성을 지니기 때문에 잔차신호들이 TU 영역에 전체적으로 고르게 분포할 경우 높은 성능을 보일 것이다.

반면에 DST-VII 은 잔차신호들이 TU 의 좌측과 위쪽에 적게 분포하고, 오른쪽과 아래로 갈수록 많이 분포할 경우 높은 압축 성능을 보일 것이다. HEVC 는 화면 내 예측에 참조하는 복원 신호들은 현재 블록의 왼쪽과 위쪽에 있기 때문에 좌측과 위쪽 부분의 예측오차의 확률이 작고 좌측과 위쪽으로부터 멀어질수록 오차의 확률이 점점 증가할 것이다. DST-VII 은 이러한 특성과 유사한 기저 모양을 가지기 때문에 이러한 경우 효율적인 변환 기저가 될 수 있다. 하지만 그림 1 에서 보인 예시와 같이 잔차신호가 집중되는 위치가 다른 경우도 발생할 수 있다. 이러한 경우 잔차신호가 집중되는 위치를 DST-VII 과 같은 우측하단으로 이동해 줌으로써 효율적으로 변환을 수행할 수 있다. 이러한 과정을 잔차신호 재배열이라고 하며 본 논문에서는 4 가지 재배열 방법을 제안한다. 4 가지 재배열 방법은 다음과 같다. 재배열 없음, 가로플립, 세로플립 그리고 가로세로플립 이다. 재배열 없음은 말 그대로 재배열을 수행하지 않는다. 가로 플립은 잔차신호를 가로 방향으로 반전시킨다. 세로 플립은 잔차신호들을 세로 방향으로 반전시킨다. 가로세로플립은 잔차신호들을 가로 및 세로 방향으로 반전시킨다. 각각 4 가지 재배열 방법들을 수식으로 표현하면 다음과 같다.

- 재배열 없음 : $p'(x, y) = p(x, y)$
- 가로 플립 : $p'(x, y) = p(nT - 1 - x, y)$
- 세로 플립 : $p'(x, y) = p(x, nT - 1 - y)$
- 가로세로 플립 : $p'(x, y) = p(nT - 1 - x, nT - 1 - y)$

$p(x, y)$ 는 재배열 이전의 잔차신호들이며 $p'(x, y)$ 는 재배열된 잔차신호들이다. nT 는 TU 의 너비 및 높이에 해당한다.

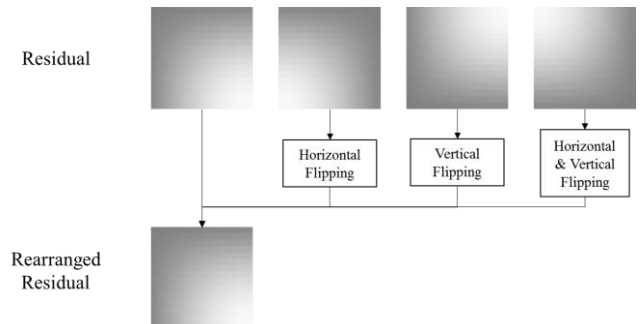


그림 5. 제안하는 4 가지 재배열 방법

그림 5 는 4 가지 재배열 방법을 보여준다. 4 가지 잔차신호들은 각각 4 가지 재배열 방법들의 이상적인 잔차신호 입력을 의미한다. 각각의 입력된 잔차신호들은 잔차신호 재배열 과정을 통해 우측 하단에 집중되게 된다. 그리고 재배열된 잔차신호들은 DST-VII 를 이용하여 변환을 수행한다. 부호화기에서 RD 비용 비교를 통해 4 가지 재배열 방법 중 한가지 방법이 선택되며, 선택 정보는 복호화기로 전송된다.

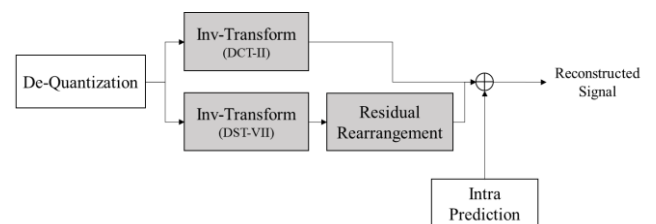


그림 6. 제안하는 복호화 과정

그림 6 는 제안하는 화면 내 예측 복호화 과정에 대한 도시이다. 역양자화 이후 부호화기로부터 변환정보를 획득하여 DCT-II 또는 DST-VII 으로 역변환을 수행한다. DST-VII 으로 역변환을 수행한 경우 잔차신호 재배열을 통해 잔차신호를 복원한다. 본 논문에서 제안한 재배열 방법은 재배열과 역재배열이 동일한 프로세스를 갖는다.

3. 실험 결과

제안하는 방법은 HEVC 참조 소프트웨어 HM16.6[3] 에서 구현되었으며, HEVC 공통 실험 조건[4]의 Intra, high efficiency, 10 bit 조건에서 실험하였다. 테스트에 사용한 영상은 JVET 테스트 영상들이며 표 1 에 클래스 별로 정리하였다. 모든 테스트 영상들은 8 프레임 간격으로 건너뛰며 부호화되었다. 양자화 파라미터는 22, 27, 32, 37 을 사용하였다.

표 1. 테스트 영상

클래스	시퀀스	해상도	비트심도	프레임율
A1	Tango	4096x2160	10	60 fps
A1	Drums100	3840x2160	10	100 fps
A1	CampfireParty	3840x2160	10	30 fps
A1	ToddlerFountain	4096x2160	10	60 fps
A2	CatRobot	3840x2160	10	60 fps
A2	TrafficFlow	3840x2160	10	30 fps
A2	DaylightRoad	3840x2160	10	60 fps
A2	Rollercoaster	4096x2160	10	60 fps
B	Kimono	1920x1080	8	24 fps
B	ParkScene	1920x1080	8	24 fps
B	Cactus	1920x1080	8	50 fps
B	BasketballDrive	1920x1080	8	50 fps
B	BQTerrace	1920x1080	8	60 fps
C	BasketballDrill	832x480	8	50 fps
C	BQMall	832x480	8	60 fps
C	PartyScene	832x480	8	50 fps
C	RaceHorses	832x480	8	30 fps
D	BasketballPass	416x240	8	50 fps
D	BQSquare	416x240	8	60 fps
D	BlowingBubbles	416x240	8	50 fps
D	RaceHorses	416x240	8	30 fps
E	FourPeople	1280x720	8	60 fps
E	Johnny	1280x720	8	60 fps
E	KristenAndSara	1280x720	8	60 fps
F	BasketballDrillText	832x480	8	50 fps
F	ChinaSpeed	1024x768	8	30 fps
F	SlideEditing	1280x720	8	30 fps

F	SlideShow	1280x720	8	30 fps
---	-----------	----------	---	--------

표 2. 실험 결과

	All Intra Main 10 Over HM-16.6			
	Y	U	V	EncT
Class A1	-3.19%	-0.57%	-0.83%	158%
Class A2	-3.17%	-1.45%	-1.46%	155%
Class B	-2.58%	-0.73%	-0.71%	163%
Class C	-2.29%	0.00%	-0.07%	169%
Class D	-1.96%	0.05%	0.30%	173%
Class E	-2.46%	-0.73%	-0.81%	156%
Overall	-2.61%	-0.57%	-0.59%	162%
Class F(optional)	-2.23%	-0.55%	-0.68%	163%

표 2 는 제안하는 방법의 실험 결과를 보인다. 제안하는 방법은 HM16.6 과 성능을 비교하였으며 BD-rate(Bjontegaard Distortion rate) [5]를 통해 성능 차이를 계산하였다. Y,U,V 는 각각 색 성분들의 BD-rate 를 의미하며 EncT 는 HM-16.6 대비 제안하는 알고리즘의 부호화 시간 비율, DecT 는 HM-16.6 대비 제안하는 알고리즘의 복호화 시간 비율을 의미한다. 고해상도 영상일수록 제안하는 알고리즘은 높은 성능을 보이며 해상도가 낮아질수록 제안하는 알고리즘의 성능 향상 폭이 줄어드는 것을 확인할 수 있다. 고해상도 영상에서는 픽셀간의 상관관계가 비교적 높기 때문에 큰 크기의 블록이 많이 발생한다. 그리고 제안하는 방법은 큰 블록일수록 성능 향상이 높게 나타나기 때문에 해상도에 따라 성능차이가 발생하는 것으로 보인다. 부호화 시간은 잔차신호 재배열의 RD 비용 계산량 증가로 인해 HEVC 대비 60%정도의 계산시간이 증가하였다.

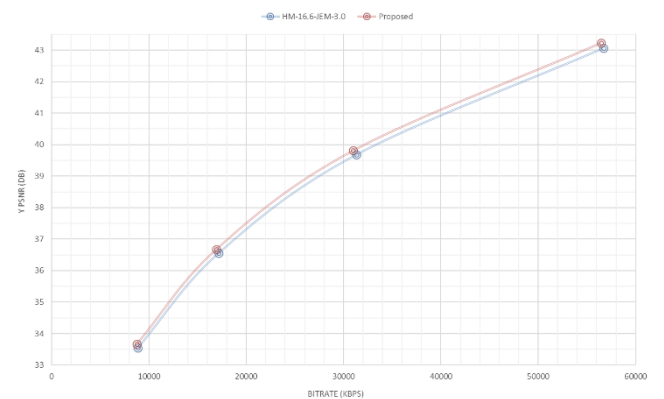


그림 7. ToddlerFountain 의 PSNR 곡선

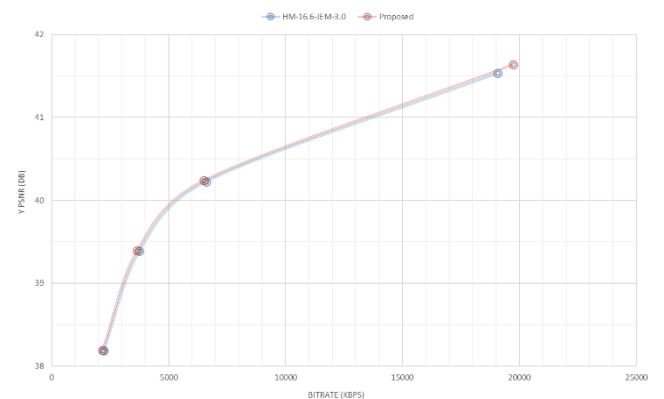


그림 8. Tango 의 PSNR 곡선

그림 7 과 그림 8 은 각각 ToddlerFountain 과 Tango 에서의

PSNR-BitRate 곡선을 보여준다. 제안하는 방법의 PSNR 곡선은 HM16.6 PSNR 곡선보다 항상 위에 있기 때문에 모든 비트율에서 제안하는 방법의 성능이 HEVC 보다 항상 높음을 확인할 수 있다. 그리고 모든 테스트 영상에서 성능 향상을 보이므로 제안하는 방법은 안정적인 성능향상을 얻어낼 수 있다고 볼 수 있다.

표 3. Class A0~E 의 변환 선택 비중 평균

TU	DCT-II	DST-VII (No Flip)	DST-VII (H-Flip)	DST-VII (V-Flip)	DST-VII (HV-Flip)
4x4	24.6%	36.3%	7.8%	27.6%	3.6%
8x8	34.6%	39.1%	8.8%	13.5%	4.0%
16x16	40.2%	37.2%	8.6%	10.8%	3.2%
32x32	62.4%	23.7%	5.8%	6.2%	2.0%

표 4. Class F 의 변환 선택 비중 평균

TU	DCT-II	DST-VII (No Flip)	DST-VII (H-Flip)	DST-VII (V-Flip)	DST-VII (HV-Flip)
4x4	62.0%	11.2%	2.8%	23.0%	1.0%
8x8	64.4%	21.2%	3.4%	10.1%	0.9%
16x16	63.1%	26.1%	3.1%	6.7%	0.9%
32x32	78.6%	14.4%	2.3%	4.1%	0.6%

표 3 은 Class A0 부터 Class E 까지의 영상을 제안하는 방법으로 부호화 하였을 경우 변환 종류가 선택되는 비중의 평균을 나타낸다. DCT-II 는 큰 블록일수록 선택되는 비중이 높아지는 특성을 보인다. 반면에 DST-VII 는 작은 블록일수록 선택되는 비중이 높아지는 것을 확인할 수 있다. 잔차신호를 재배열 방법은 재배열 하지 않음, V-Flip, H-Flip, HV-Flip 순으로 선택되는 확률이 높다. H-Flip 과 HV-Flip 의 선택 비중이 합쳐서 10% 내외로 적은 편이지만 해당 두 재배열 방법은 1% 정도의 BD-rate 성능향상을 가져온다.

표 4 는 스크린 콘텐츠인 Class F 에 대한 변환 종류 선택 비중의 평균을 나타낸다. 다른 클래스들과 다르게 DCT-II 의 선택 비중이 매우 높게 나타났다. 잔차신호 재배열의 선택확률은 다른 클래스들과 비슷하게 나타났다.

4. 결론

제안하는 DST-VII 과 잔차신호 재배열을 이용한 변환방법은 고해상도 영상에 대해 HEVC 대비 평균적으로 3.2% 그리고 전체 테스트 영상에 대해 평균 2.6%의 BD-rate 향상을 보인다.

감사의 글

이 논문은 일부 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2016-0-00572, 초고실감 미디어 서비스 실현을 위해 HEVC/3DA 대비 2 배 압축을 제공하는 5 세대 비디오/오디오 표준 핵심 기술 개발 및 표준화)

참 고 문 헌

- [1] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, High Efficiency Video Coding (HEVC) Text Specification Draft 10 (for FDIS & Consent), JCT-VC document JCTVC-L1003, vol. 1, 12thmeeting, Geneva, Switzerland, 2013.
- [2] V. Britanak, P. Yip, and K. R. Rao, "Discrete Cosine and Sine Transforms – General Properties, Fast Algorithms, and Integer Approximations," 1st Edition, Academic Press, Inc, San Diego, 2006.
- [3] HM reference software HM-16.6, https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware, Jun. 2015.
- [4] JCT-VC document, JCTVC-L1100, Common HM Test conditions and software reference configurations, Geneva, Switzerland, Jan. 2013
- [5] G. Bjøntegaard, Calculation of Average PSNR Differences Between RDCurves, ITU-T Q.6/SG16, Doc. VCEG-M33, 2001