

In Loop Filtering Tools in ECM Reference Software for Beyond VVC

□ 이민훈, 심동규 / 광운대학교

요약

JVET (Joint video experts team)은 VVC (Versatile video coding)보다 높은 압축 성능을 가지는 신호처리 기반 차세대 비디오 압축 표준 기술 개발을 위해 ECM (Enhanced compression model) 참조 소프트웨어를 배포하여 기술 탐색을 수행 중이다. ECM은 휘도 및 색차 성분 모두에서 VVC 대비 높은 부호화 성능을 보이며, ECM 내 기술은 VVC의 기술을 개선하거나 새로운 기술이 추가된 형태로 구성되어 있다. 그중, 인루프 필터링 기술은 VVC에 채택된 DF (Deblocking filter), SAO (Sample adaptive offset), ALF (Adaptive loop filter), CCALF (Cross-component ALF), LMCS (Luma mapping with chroma scaling)가 일부 수정 및 개선된 형태로 포함되어 있고, CCSAO (Cross-component SAO) 및 BIF (Bilateral filter) 기술이 새롭게 포함되어 있다. 본고에서는 2023년 7월에 진행된 143차 미팅 이후 배포된 ECM10.0에 포함되어 있는 인루프 필터링 기술에 대해 소개한다. 특히, VVC 대비 개선점이 많은 ALF 및 CCALF와 새롭게 추가된 기술인 CCSAO 및 BIF에 대해 자세히 설명하고자 한다.

I. 서론

JVET (Joint video experts team)은 ISO/IEC MPEG (Moving picture experts group)과 ITU-T VCEG (Video coding experts group) 표준 기구에 의해 차세대 비디오 압축 표준 기술의 제정을 위해 설립된 단체이며, 새로운 비디오 압축 표준 기술인 VVC (Versatile video coding, H.266)[1]의 표준화를 2020년 7월 완료하였다. 이후, JVET은 VVC보다 우수한 압축 성능을 갖는 신호처리 기반의 차세대 비디오 압축 표준 기술을 준

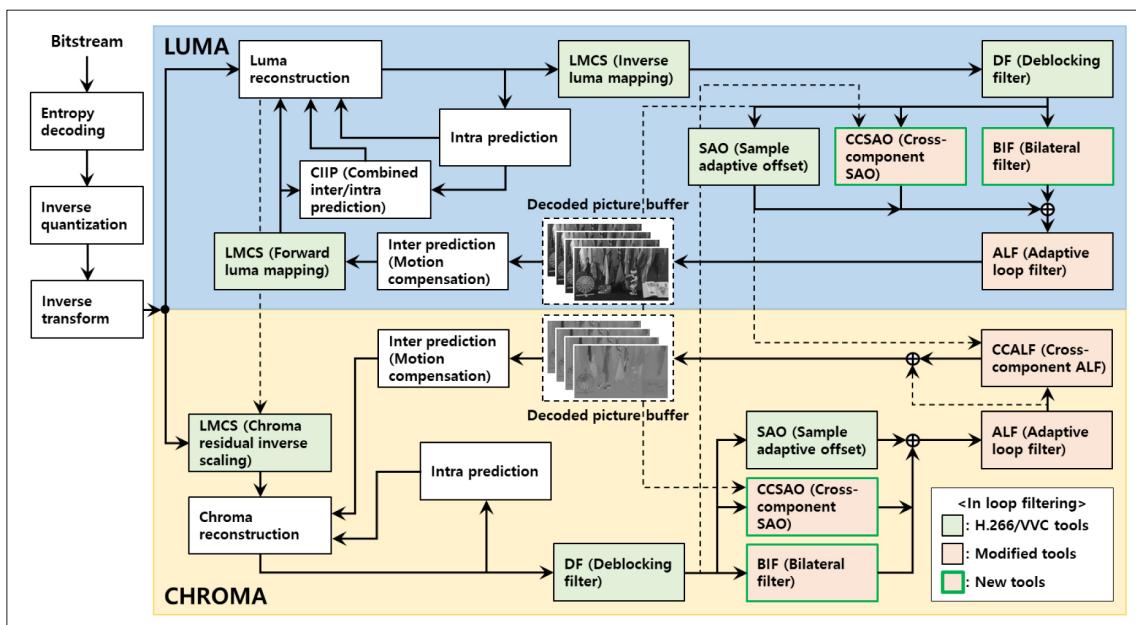
비하기 위한 논의를 시작하고, 기술 탐색과 제안된 기술들의 검증을 위해 ECM (Enhanced compression model) 참조 소프트웨어를 개발하고 채택된 기술을 통합하여 관리 중이다. 현재 ECM10.0[2]은 VVC 참조 소프트웨어인 VTM (VVC Test Model) 11.0 대비 AI (All Intra) 환경 및 RA (Random Access) 환경에서 휘도 성분에 대해 각각 -12.49%와 -22.23%의 부호화 효율을 얻는다[3].

ECM을 이용하여 영상을 부/복호화하는 과정은 최신 표준 기술인 VVC와 동일하게 블록 기반 하이브리드 방식을

사용하기 때문에 영상 분할 및 양자화 등의 과정에서 블록킹 열화 (Blocking artifact), 링잉 열화 (Ringing artifact), 텍스처/에지 스무딩 (Texture/edge smoothing) 등의 다양한 압축 열화가 발생할 수 있다. 이를 줄이고자 비디오의 부/복호화 과정 내 루프에서 필터링이 수행되며, 이러한 필터링 기술들은 예측 및 역변환 이후 복원 뷔퍼가 DPB (Decoded picture buffer)에 저장되기 전, 부/복호화 과정의 루프 내에서 수행되기 때문에 ‘인루프 (In loop)’ 형태로 정의된다.

본고에서 설명하고자 하는 인루프 필터링 기술을 구체적으로 도식화한 ECM 복호화기 개념도는 <그림 1>과 같다. ECM의 인루프 필터링 기술은 VVC에서 사용되었던 기술들과 추가로 채택된 기술로 구성되어 있다. 구체적으로는 블록킹 열화를 줄이기 위한 DF (Deblocking filter) 와 링잉 열화를 줄이고, 샘플 단위의 오프셋을 통해 왜곡을 보상하는 SAO (Sample adaptive offset) 및 부호화 효율을 향상시키기 위해 효율적인 예측 및 양자화를 수행할 수 있도록 신호의 범위를 재분배하는 LMCS (Luma

mapping with chroma scaling)는 VVC와 비교하여 큰 기술적인 변경 없이 사용된다. 그리고 인루프 필터링의 가장 마지막 단계에 적용되는 원본 신호와의 오차를 줄이기 위한 목적으로 사용되는 ALF (Adaptive loop filter)와 휘도와 색차 성분 간 상관관계를 통해 색차 신호를 보정하는 CCALF (Cross-component ALF)는 다양한 필터의 입력 사용 및 필터의 탭 수가 증가한 형태로 발전되어 높은 부호화 효율을 보여주고 있다. CCSAO (Cross-component SAO)와 BIF (Bilateral filter)는 ECM에 새롭게 추가된 기술이다. CCSAO는 복원 및 DF 수행 이후의 샘플들을 휘도와 색차 성분 간의 상관관계를 이용하여 다양한 카테고리로 분류하고 샘플 단위의 오프셋을 추가하여 왜곡을 줄이는 기술이며, CCALF와는 다르게 휘도 및 색차 성분에 모두 적용되는 기술이다. 또한, SAO와는 달리 한 성분의 CCSAO 수행 과정에 모든 성분의 복원 샘플을 사용한다. BIF는 SAO의 목적과 동일하게 양자화로 인해 발생하는 영상 내 에지 주변의 링잉 열화 현상을 줄이기 위해 적용되는 기술이다. ECM에서 SAO, CCSAO, BIF는 각 과정이



<그림 1> ECM 복호화기 개념도

휘도 및 색차 성분에서 각각 병렬적으로 수행되며, 세 과정의 출력값이 합산 및 클리핑 (Clipping)되어 ALF를 위한 입력으로 사용된다.

본고는 ECM10.0에 포함된 다양한 부호화 기술 가운데 인루프 필터링 기술을 중점적으로 설명하고, 그 중에서도 VVC 대비 개선점이 많은 ALF 및 CCALF와 새롭게 채택된 기술인 CCSAO 및 BIF에 대해 주요하게 설명한다[2][4].

II. ECM10.0 참조 소프트웨어의 인루프 필터링 기술

본 장에서는 ECM10.0의 인루프 필터링 기술인 LMCS, DF, SAO, CCSAO, BIF, ALF, CCALF 기술을 자세하게 살펴본다. 그중, LMCS, DF, SAO 기술은 VVC에서와 동일하게 적용되며, ECM에 채택된 다른 기술들로 인한 약간의 변경 사항만 존재한다. CCSAO 및 BIF 기술은 새롭게 채택된 기술로써 자세하게 살펴보고, ALF 및 CCALF는 관련 기술이 활발히 기고되고 있어, VVC 대비 변경된 사항이 다소 존재하여 변경된 사항을 위주로 살펴본다.

1. Luma mapping with chroma scaling (LMCS)

LMCS는 다른 인루프 필터링 기술과는 다르게 영상 신호의 동적 범위 (Dynamic range)에 대한 코드워드 (Codeword)를 재분배하여 보다 효율적인 샘플값들의 동적 범위에서 부호화를 수행하여 부호화 효율을 얻는 기술이다. LMCS는 휘도 맵핑 (Luma mapping)과 색차 스케일링 (Chroma scaling)으로 이루어져 있다. 휘도 샘플을 부호화 효율을 위해 재분배하고, 재분배로 인해 변화된 휘도 샘플들과의 차이를 보상하기 위해 색차 성분에 스케일링을 수행한다.

1) Luma mapping (LM)

일반적으로 10비트 영상의 경우, 휘도 성분에 대해 64에서 940의 값만 영상 신호에 허용되고, 0~63 및 941~1023의 값은 허용되지는 않지만 부호화 과정에서는 활용할 수 있다. LM의 기본 아이디어는 이러한 입력 영상 신호의 휘도값의 범위를 지정된 비트 심도 (Bit depth)에서 허용되는 범위에서 더욱 잘 활용하도록 재분배하여 예측 등을 효과적으로 수행하여 부호화 효율을 얻는 것이다.

LM은 크게 휘도 성분의 신호를 부호화에 효율적인 동적 범위로 맵핑시키는 순방향 맵핑 (Forward mapping)과 기존 동적 영역으로 다시 맵핑시키는 역방향 맵핑 (Inverse mapping)의 두 가지 과정으로 이루어져 있다.

순방향 맵핑 과정은 원본 샘플 도메인의 동적 영역을 동일한 16개의 구간으로 분할하여 휘도 코드값을 재분배 함으로써 맵핑된 도메인 (Mapped domain)에 맵핑하는 과정이다. 순방향 맵핑 과정에서는 10비트 영상에 대해 16개의 구간을 각 64개의 코드워드가 할당되도록 동일하게 분할하고, 영상의 특징을 분석하여 원본 샘플의 값을 구간 선형 모델 (Piece-wise linear model)을 이용하여 각 구간에 대해 맵핑을 수행한다. 예를 들어, 인지 시각에 상대적으로 큰 영향을 미치는 중간 범위의 구간은 맵핑된 구간의 길이가 더 길게 할당될 수 있고, 상대적으로 적은 영향을 미치는 크거나 작은 밝기값이 존재하는 양 끝 범위 구간의 길이는 짧게 할당될 수 있다. 순방향 맵핑을 위한 선형 모델 정보인 원본 샘플 도메인과 맵핑된 도메인 간의 관계에 대한 테이블은 LMCS APS (Adaptation parameter set)로 전송되며, 역방향 맵핑을 위한 선형 모델은 순방향 맵핑의 선형 모델로부터 유도하여 적용한다.

LM으로 인해 영상의 부호화 과정 및 <그림 1>의 복호화 과정의 각 단계는 원본 샘플 도메인 또는 맵핑된 도메인에서 수행된다. 휘도 성분에 대한 화면간 예측 (움직임 보상)은 원본 샘플 도메인에서 수행되며, 화면내 예측은 맵핑된 도메인에서 수행된다. 그리고 CIIP (Combined inter and intra prediction)로 예측이 수행되는 경우, 원본 샘플 도메인에서 생성한 화면간 예측 신호와 맵핑된 도메인에서 생성한 화면내 예측 신호를 가중합하여 최종 예

측 신호를 생성한다. 또한, 새롭게 채택되어 ECM에 포함된 OBMC (Overlapped block motion compensation) 기술은 주변 움직임 정보를 이용하여 현재 블록의 예측 신호를 보정하는 기술로, CIIP로 예측이 수행되면서 OBMC가 수행되는 경우, 원본 샘플 도메인에서 생성한 화면간 예측 신호에 OBMC를 통해 보정하고, 보정된 화면간 예측 신호를 맵핑된 도메인으로 변환한 후, 맵핑된 도메인에서 생성한 화면내 예측 신호를 가중합하여 최종 예측 신호를 생성한다[5].

2) Chroma scaling (CS)

색차 성분의 스케일링인 CS는 복원된 대응 위치의 주변 블록의 평균 휘도값에 따라 색차 성분의 차분값(Residual)을 스케일링하는 과정으로, 휘도 신호와 대응하는 색차 신호 사이의 관계에 따라 색차 신호를 보정하는 과정이며, 넓이가 4 이하인 색차 블록에 대해서는 적용되지 않는다.

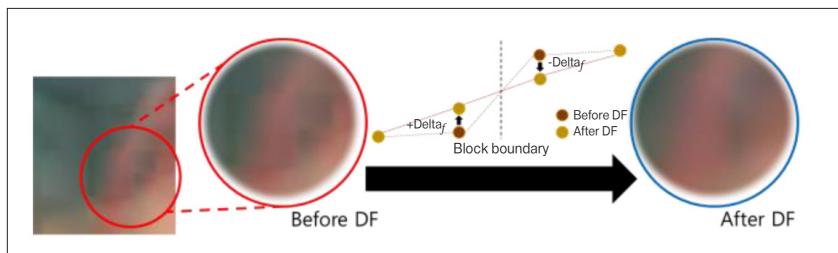
CS도 LM과 마찬가지로 순방향 및 역방향 과정으로 구성되어 있다. 순방향 스케일링 과정은 색차 성분의 차분값과 순방향 스케일 계수를 곱하여 스케일링된 색차 차분값을 얻는 과정이다. 이는 곧 색차 성분의 차분값에 역방향 스케일 계수를 나눈 것과 동일하며, 역방향 스케일 계수값을 이용하여 순방향 스케일링 과정을 수행한다. 부호화 과정에서는 복원되고 역방향 맵핑이 수행된 휘도 신호의 복원값을 이용하여 LM 과정에 사용된 원본 샘플 도메인과 맵핑된 도메인 간의 관계에 대한 테이블 정보 및 테

이를 내 모든 구간에서 동일하게 사용되는 오프셋 값에 기반하여 역방향 스케일 계수값을 구한다. 이때, 오프셋 값은 APS로 전송된다.

역방향 스케일링 과정은 역방향 스케일 계수와 복원된 스케일링된 차분값을 곱하여 수행된다. 역방향 스케일 계수는 휘도 신호의 복원값을 기반으로 APS로 전송된 휘도 성분의 원본 샘플 도메인과 맵핑된 도메인 간의 관계에 대한 테이블에서 구한 값과 APS로 전송된 오프셋 값에 의해 유도되어 결정된다. 이때 휘도 신호의 복원값은 VVC에서는 현재 색차 블록의 대응 위치에 해당하는 휘도 블록의 상단 및 좌측 VPDU (Virtual pipeline data unit)의 복원되고 역방향 맵핑이 수행된 휘도 샘플의 평균값이 사용된다[7]. ECM에서는 VPDU가 사용되지 않기 때문에 휘도 신호의 복원값은 현재 색차 블록의 대응 위치에 해당하는 휘도 블록의 상단 및 좌측 CU (Coding unit)의 복원되고 역방향 맵핑이 수행된 휘도 샘플의 평균값이 사용된다[2].

2. Deblocking filter (DF)

DF는 HEVC (High efficiency video coding)[6], VVC에 모두 채택된 기술이며, ECM에도 사용되는 기술로, 복원 신호에 가장 먼저 적용되는 인루프 필터링 기술이다. ECM은 HEVC, VVC와 같이 블록 기반 하이브리드 압축 방식을 사용하기 때문에 인접한 두 블록이 서로 다른 QP (Quantization parameter)로 부호화될 수 있다. 이에 따



<그림 2> DF 개념도

라 복원된 영상의 블록 경계에서 샘플값의 불연속성이 발생할 수 있다. 또한, 인접한 블록 간 QP값의 차가 커질수록 불연속성은 더욱 커질 수 있다. 이러한 불연속성은 블록킹 열화라고 하며, DF는 <그림 2>의 개념도와 같이 이를 제거하기 위한 목적으로 수행하는 기술이며, 별도의 플래그 및 인덱스의 전송 없이 부/복호화기에 동일하게 적용되는 기술이다.

<그림 3>은 ECM의 DF 블록도이다. ECM의 DF는 먼저, 인접한 두 블록의 경계가 변환 블록의 경계인지 서브 블록 경계인지 확인하고, 조건에 따라 경계를 기준으로 양쪽의 필터링을 수행할 샘플의 수를 {1, 2, 3, 5, 7} 중 하나로 결정한다. 이때, 블록 경계를 기준으로 양쪽은 비대칭 필터링이 수행될 수 있다. 이후 BS (Boundary strength)를 결정하며, 이는 두 블록의 예측 모드, 움직임의 유사도, 참조 픽처의 POC (Picture order count) 등 다양한 조건에 따라 {0, 1, 2} 중 하나로 분류된다. 휴드 성분에서는 1, 2인 경우에 DF가 수행될 수 있으며 색차 성분에서는 2인 경우에만 DF가 수행될 수 있다[6]. 이때, 두 블록 중 적어도 한 블록이 ECM에 포함되어 있는 기술 중 하나

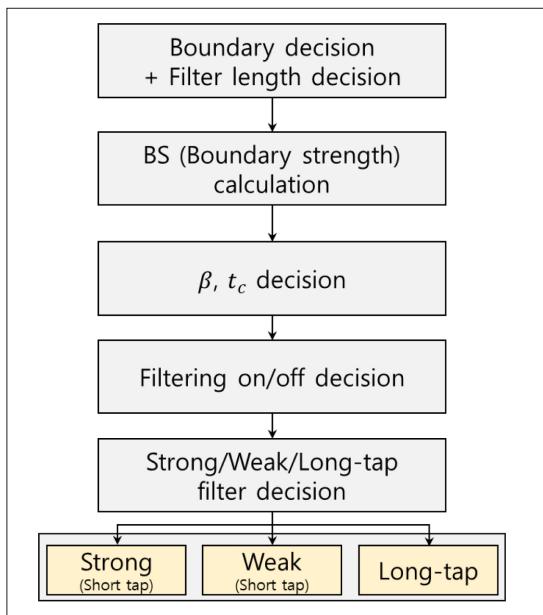
인 GPM (Geometric partitioning mode)으로 예측이 수행되고, 한 서브 영역은 화면내 예측, 다른 서브 영역은 화면간 예측으로 예측 신호를 생성하는 경우, BS는 2로 분류된다[8]. β 및 t_c 는 필터링 여부 판단 및 필터링 과정에 사용되는 임계값이며, 부/복호화기 간 각각의 LUT (Look up table) 형태로 정의되어 있다. 두 값은 각각 두 블록의 QP에 따라 테이블을 통해 결정된다. 이 과정에서 결정된 β 를 이용하여 두 블록 간 존재하는 불연속성이 원본 영상 내에인지, 혹은 블록킹 열화인지를 판단한다. 이후 β 및 t_c 를 이용하여 조건에 따라 필터의 종류를 결정한 후, DF를 수행한다[7].

3. Sample adaptive offset (SAO)

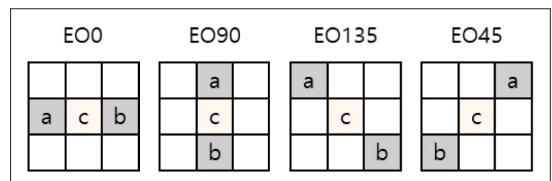
SAO는 DF와 마찬가지로 HEVC, VVC에 모두 채택된 기술이며, 원본 영상과 복원 영상 간의 왜곡을 왜곡의 정도가 다른 샘플을 구분하여 서로 다른 오프셋을 적용적으로 적용함으로써, 영상의 특성에 따라 부분적으로 다르게 발생하는 왜곡을 최소화하여 부호화 효율을 향상시키는 기술이다. SAO는 HEVC, VVC에서 동일하게 수행되며, ECM에서도 동일하게 두 종류의 오프셋 EO (Edge offset) 및 BO (Band offset) 중 하나가 수행되고 상단 또는 왼쪽 CTU (Coded tree unit)의 정보를 머지 (Merge)하여 수행된다[6].

1) Edge offset (EO)

EO는 <그림 4>와 같이, 1D Laplacian 수직, 수평, 두 대각 방향 (45 degree, 135 degree) 방향 패턴을 샘플 분



<그림 3> ECM의 DF 블록도



<그림 4> EO 클래스 종류

류에 사용하며, 패턴에 대한 클래스 정보는 전송된다. 이 때, ‘c’는 현재 샘플을 나타내며 ‘a’ 및 ‘b’는 두 개의 주변 샘플을 나타낸다. 전송받은 클래스에 따라, 현재 샘플 및 주변 두 샘플값을 비교하여 4개의 카테고리로 분류하고, 카테고리에 따라 음수 또는 양수의 오프셋을 구하여 4개의 오프셋 값을 전송한다.

2) Band offset (EBO)

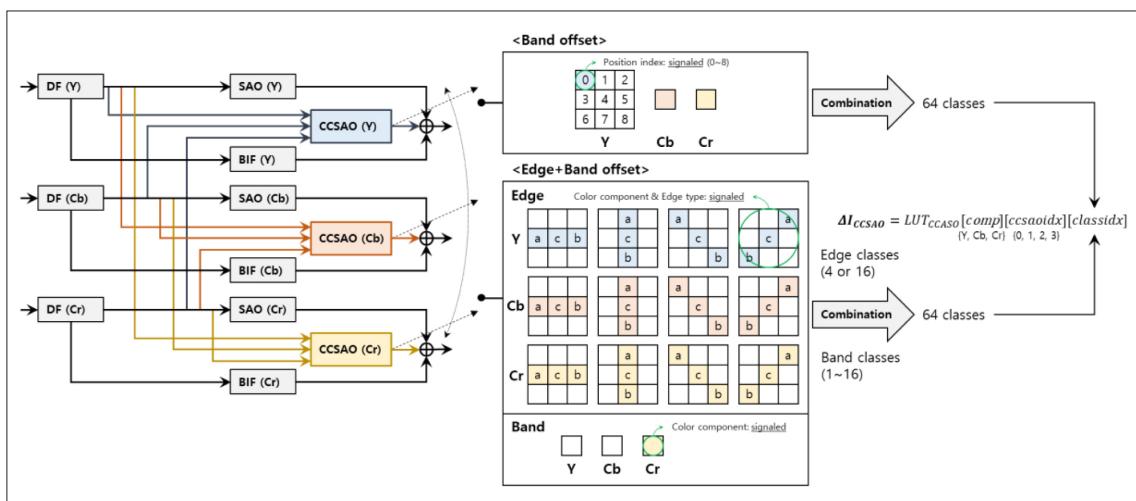
BO는 CTU 내 특정 대역의 밝기값을 갖는 샘플들을 보정한다. 밝기값 대역은 부호화되는 영상의 비트 심도를 32 구간으로 분할했을 때, 연속된 4개의 밴드 (Band)로 구성되며, 시작 밴드의 위치가 전송되고 4개의 오프셋 값이 전송된다. 따라서 4개의 오프셋은 시작 밴드로부터 연속되는 4개의 밴드에 대한 오프셋으로, 원본 샘플과 복원 샘플 간 차잇값의 평균값으로 구성된다.

VVC에서는 SAO를 이용하여 구한 오프셋 값 (ΔI_{SAO})을 이용하여 DF 수행 이후 샘플값에 더하여 필터링을 수행했지만, ECM에서는 이후 4장 및 5장에서 다룬 CCSAO의 오프셋 값 (ΔI_{CCSAO})과 BIF의 오프셋 값 (ΔI_{BIF})을 모두 합산하여 필터링을 수행함으로써 ALF의 입력으로 사용할 샘플값을 생성한다.

4. Cross-component SAO (CCSAO)

CCSAO는 새롭게 채택되어 ECM에 포함된 기술로써, SAO의 과정과 유사하게 복원된 휘도 및 색차 성분의 샘플을 다양한 카테고리로 분류하고, 각 카테고리에 대해 오프셋 값을 도출하여 적용함으로써 왜곡을 줄이고자 적용하는 기술이다. SAO는 휘도 및 색차 성분에 대해 독립적으로 수행되었던 반면, CCSAO는 휘도 및 색차의 세 가지 구성 요소를 모두 활용하여 적용한다. 또한, VVC 및 ECM의 기술 중 ‘CC (Cross-component)’의 명칭이 붙은 기술 (ex. CCLM (Cross-component linear model), CCALF 등)은 대부분 휘도와 색차 성분 간의 상관관계를 이용하여 색차 성분을 예측 및 보정하는 형태로 적용되었지만, CCSAO는 휘도 및 색차 성분에 모두 적용되는 특징을 갖는 기술이다.

CCSAO는 <그림 5>와 같이, 세 구성 요소 (Y, Cb, Cr)에 대해 BO (Band offset) 또는 EBO (Edge+Band offset)의 두 가지 종류 중 하나의 오프셋을 적용할 수 있다. 오프셋 값들 (LUT_{CCSAO})은 프레임 레벨 단위로 전송되고, 각 성분별로 최대 4개의 CCSAO 셋을 사용하며, 이는 CTU 레벨로 전송된다.



BO를 이용하는 경우, 모든 성분의 밴드값의 조합을 통해 클래스를 분류하며, Y 성분의 9개의 위치 중 하나의 샘플값과 Cb 및 Cr 성분의 샘플값을 이용하여 64개의 클래스 중 하나로 분류하고, 이를 이용하여 LUT 형태로 부/복호화기 간 정의되어 있는 오프셋 값 중 하나를 더해줌으로써 CCSAO를 수행한다. EBO를 이용하는 경우, 에지를 판별할 성분 및 에지 타입의 정보를 파싱하고 밴드를 확인할 성분 정보를 파싱하여 에지 클래스 및 밴드 클래스를 결정하고, 이를 이용하여 LUT 형태로 부/복호화기 간 정의되어 있는 오프셋 값 중 하나를 더해줌으로써 CCSAO를 수행한다. BO 및 EBO에 대한 내용은 4. 1) 및 4. 2) 절에서 자세히 다룬다.

1) Band offset (BO)

BO를 이용하여 CCSAO를 수행하기 위해, 모든 성분의 밴드값의 조합을 통해 한 성분의 클래스를 식(1)을 이용하여 결정하고, 이를 통해 오프셋 값을 결정한다.

$$\text{classIdx} = \text{band}_y \times (N_{Cb} \times N_{Cr}) + \text{band}_{Cb} \times N_{Cr} + \text{band}_{Cr} \quad (1)$$

이때, N_y , N_{Cb} , N_{Cr} 및 P_y 는 각 성분별로 프레임 레벨로 전송되는 변수이며, P_y 는 <그림 5>의 BO에서 Y 성분의 9개의 위치 (0~8) 중, 어떤 위치의 샘플값을 사용할지에 대한 정보이다. 또한, N_y , N_{Cb} , N_{Cr} 은 각 성분의 샘플값의 밴드 개수이며, $N_y \times N_{Cb} \times N_{Cr} \leq 64$ 를 만족하도록 N_y (0~15), N_{Cb} (0~3), N_{Cr} (0~3)의 값이 전송된다. 이를 각각 이용하여 식 (1)에 사용되는 밴드의 인덱스를 나타내는 band_y , band_{Cb} , band_{Cr} 값을 식(2)를 이용하여 계산한다. 이때, C는 각 성분을 의미하는 변수 ($C \in \{Y, Cb, Cr\}$)이고, I_C 는 각 성분의 샘플 밝기값을 의미하고, BD 는 입력 영상의 비트 심도를 의미한다.

$$\text{band}_C = (I_C \times N_C) \gg BD, \text{ where } C \in \{Y, Cb, Cr\} \quad (2)$$

classIdx 를 구한 이후, 각 성분에 따라 CTU 레벨 단위로

전송된 CCSAO 셋을 이용하여 최종 오프셋 값인 ΔI_{CCSAO} 를 구한다.

2) Edge+Band offset (EBO)

EBO를 이용하여 CCSAO를 수행하기 위해, 전송받은 에지 및 밴드에 대한 정보를 이용하여 한 성분의 클래스를 식(3)을 이용하여 결정하고, 이를 통해 오프셋 값을 결정한다.

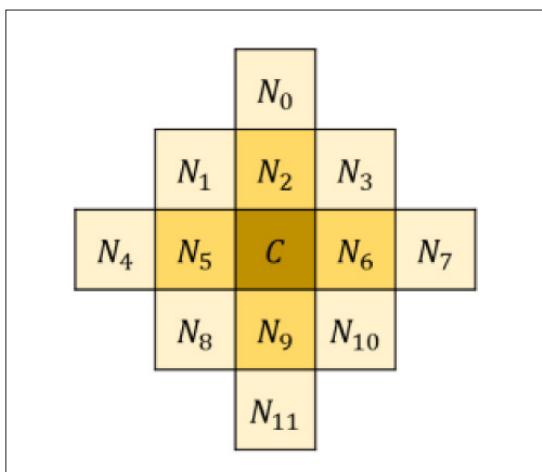
$$\text{classIdx} = \text{band}_{bc} \times N_E + \text{edge}_{cc} \quad (3)$$

이때, bc 는 프레임 레벨로 전송되는 EBO에서 밴드를 이용할 성분을 결정하는 변수 ($bc \in \{Y, Cb, Cr\}$)이고 어면 성분을 이용할지 (Y, Cb, Cr), 밴드의 수가 어떻게 되는지 (식(2)의 N_C , $C=bc$)를 16개의 조합 (ex. $bc=Y, N_C=14$)으로 부/복호화기 간 LUT 형태로 정의하여 인덱스를 전송한다. 전송받은 인덱스를 이용하여 해당 성분의 샘플값 (식(2)의 I_C , $C=bc$)을 구하고, 식(2)를 통해 밴드 클래스 (band_C , $C=bc$)를 결정한다. N_E 도 프레임 레벨로 전송되는 4 또는 16의 값을 가지는 변수로, 에지 클래스의 수를 나타내며 N_E 에 따라 4개 또는 16개의 에지 클래스를 활용할 수 있다. 세 성분 중, 어떤 성분의 샘플값을 이용하여 에지 클래스를 분류할지, <그림 5>의 EO0, EO90, EO135, EO45 중, 어떤 에지 타입을 이용할지에 대한 정보를 전송받고, 에지 클래스를 분류하기 위한 임곗값을 전송받아 N_E 의 값에 따라 a, b, c 위치에 해당하는 샘플값의 관계에 기반하여 최종적으로 4개 또는 16개 중 하나로 에지 클래스를 분류할 수 있다[2][4]. 이후 식(3)을 이용하여 classIdx 를 구한 이후, 각 성분에 따라 CTU 레벨 단위로 전송된 CCSAO 셋을 이용하여 최종 오프셋 값인 ΔI_{CCSAO} 를 구한다. 이후 CCSAO의 오프셋 값 (ΔI_{CCSAO})은 DF 수행 이후의 샘플에 SAO의 오프셋 값 (ΔI_{SAO}), BIF의 오프셋 값 (ΔI_{BIF})을 모두 합산하여 필터링을 수행하고, 이는 5 절에서 다룬다.

5. Bilateral filter (BIF)

BIF는 변환 이후 양자화로 인해 발생하는 영상 내 에지 주변의 링잉 열화를 제거하기 위한 샘플 간의 기하학적 거리와 샘플 간의 강도 차이를 고려한 비선형 에지 보존 필터링 기반 기술이며, 휘도 및 색차 성분에 각각 적용되는 기술이다. BIF는 CCSAO와 함께 새롭게 채택되어 ECM에 포함되어 있는 기술로, 복잡도를 고려하여 SAO, CCSAO, BIF는 각각 병렬적으로 수행된다.

BIF는 CTU 단위로 사용 여부를 결정할 수 있고, BIF 오프셋 값은 QP값, 예측 모드, TB (Transform block)의 크기, <그림 6>의 현재 (C) 및 주변 샘플 ($N_0 \sim N_{11}$)의 거리 및 밝깃값을 기반하여 결정된다. 또한, 화면간 예측이 수행된 블록에 대해서는 QP값이 17보다 크고, TB의 너비, 높이 중 큰 값이 128보다 작고, cbf (coded block flag)가 1인 경우에 적용할 수 있고, 화면내 예측이 수행된 블록에 대해서는 QP값이 17보다 크고, TB의 너비, 높이 중 작은 값이 32보다 작은 경우에 적용할 수 있다.



<그림 6> BIF에 사용되는 현재 및 주변 샘플의 위치

DF가 수행된 복원 샘플에 BIF가 수행되고, BIF 오프셋 값은 식(4)를 이용하여 구한다. 이때, c_i 는 공간적 거리에 대한 변수로, 식(5)를 통해 현재 샘플과 주변 샘플과의 공

간적 거리 ($dist_i$) 및 예측 모드와 TB의 크기에 따른 값의 스케일값 ($scale$)을 이용하여 계산된다. $dist_i$ 는 맨해튼 거리 (Manhattan distance)를 기반으로 결정되는 값으로, <그림 6>에서 N_2, N_5, N_6, N_9 에 대해서는 1, 나머지 주변 샘플에 대해서는 1/2의 값을 가진다.

$$\Delta_{BIF} = [\sum_{i=0}^{11} (c_i \times f_i)] + (32 \gg BIF strength_{pps})] \gg (6 - BIF strength_{pps}) \quad (4)$$

$$c_i = dist_i \times scale \quad (5)$$

식(5)의 $scale$ 은 <표 1>과 같이, 예측 모드와 조건에 따른 TB의 크기 (TB_{size})에 따라 1~3의 값을 가진다. 이때, <표 1>의 TB_{size} 는 블록 분할 구조에 따라 휘도 및 색차 블록에 대해 다르게 결정된다. 싱글 트리 (Single tree) 구조에서는 휘도 및 색차 성분 모두 TB_{size} 가 휘도 성분 TB의 너비와 높이 중 작은 값으로 결정되며, 듀얼 트리 (Dual tree) 구조에서는 휘도 성분에 대해서는 휘도 성분 TB의 너비와 높이 중 작은 값, 색차 성분에 대해서는 색차 TB의 너비와 높이 중 작은 값으로 결정된다. 또한, <표 1>의 임곗값인 T_0, T_1 은 <표 2>에 따라 결정된다.

<표 1> 예측 모드 및 TB_{size} 에 따른 $scale$ 값

| | $TB_{size} \leq T_0$ | $T_0 < TB_{size} < T_1$ | $T_1 \leq TB_{size}$ |
|-------|----------------------|-------------------------|----------------------|
| Intra | 3 | 2 | 1 |
| Inter | 2 | 2 | 1 |

<표 2> 분할 구조에 따른 각 성분의 T_0 및 T_1 임곗값

| | T_0 | T_1 |
|---|-------|-------|
| Single tree Luma & Dual tree Luma, Chroma | 4 | 128 |
| Single tree Chroma | 16 | 256 |

식(4)의 f_i 은 밝깃값 차이에 대한 변수이며, 식(6)을 통해 계산된다. $\text{sign}(C - N_i)$ 은 현재 샘플과 주변 샘플과의 밝깃값의 차이를 기반으로 결정되며, 현재 샘플이 더 큰 경우 1, 그렇지 않은 경우 -1의 값을 가진다. 식(6)의 LUT_{Diff} 은 부/복호화기 약속에 의해 휘도 및 색차 성분에

대해 각각 LUT 형태로 정의되어 있고, QP값 (BIF_{QP})과 현재 샘플과 주변 샘플의 밝기값의 차이 ($Diff$)를 기반하여 결정된다. 먼저, BIF_{QP} 는 17 및 42로 값이 클리핑되어 17~42 범위의 값을 가지며 $Diff$ 는 식(7)에 따라 0~15 범위의 값을 가지도록 계산된다. 식(4)의 $BIFstrength_{pp}$ 는 0~2의 값을 가질 수 있는 BIF 강도에 대한 변수이며, 현재 ECM10.0에서는 1로 고정하여 사용하고 있다. 최근 BIF의 scale 값의 범위를 더 확장하고, LUT 내 연속되는 두 값을 보간하여 사용함으로써 보다 다양한 오프셋 값을 결정할 수 있는 방법이 채택되어, 다음 버전의 ECM 참조 소프트웨어에 포함될 예정이다[9].

$$f_i = \text{sign}(C-N_i)LUT_{Diff}[\text{compID}][BIF_{QP}][Diff] \quad (6)$$

$$Diff = \text{clip3}(0, 15, (|C-N_i| + 4 \gg 3)) \quad (7)$$

DF 수행된 복원 샘플 (I)에 SAO, CCSAO, BIF를 각각 적용하여 세 종류의 오프셋 값 (ΔI_{SAO} , ΔI_{BIF} , ΔI_{CCSAO})을 구하고, 식(8)을 이용해 오프셋을 모두 적용하여 I' 를 구한 후, 이는 ALF로 입력된다. 이때, $clip1$ 은 비트 심도에 따라 0 및 최댓값 (ex. 10비트의 경우, 1023)으로 클리핑을 수행하는 함수이다.

$$I' = clip1(I + \Delta I_{SAO} + \Delta I_{BIF} + \Delta I_{CCSAO}) \quad (8)$$

6. Adaptive loop filter (ALF) and Cross-component ALF (CCALF)

ALF는 HEVC 표준화 단계에서도 논의되었던 기술로, HEVC에는 복잡도 문제로 채택되지 않았지만 VVC에 채택되어 우수한 부호화 효율을 보이는 기술 중 하나이며, CCALF는 색차 성분의 샘플값들을 추가로 보정하기 위해 채택된 기술이다. ECM의 ALF 및 CCALF는 VVC에서 와 마찬가지로 LLS (Linear least-square) 근사에 기초한 Wiener 필터 기반 필터링을 통해 원본 영상과 복원 영상

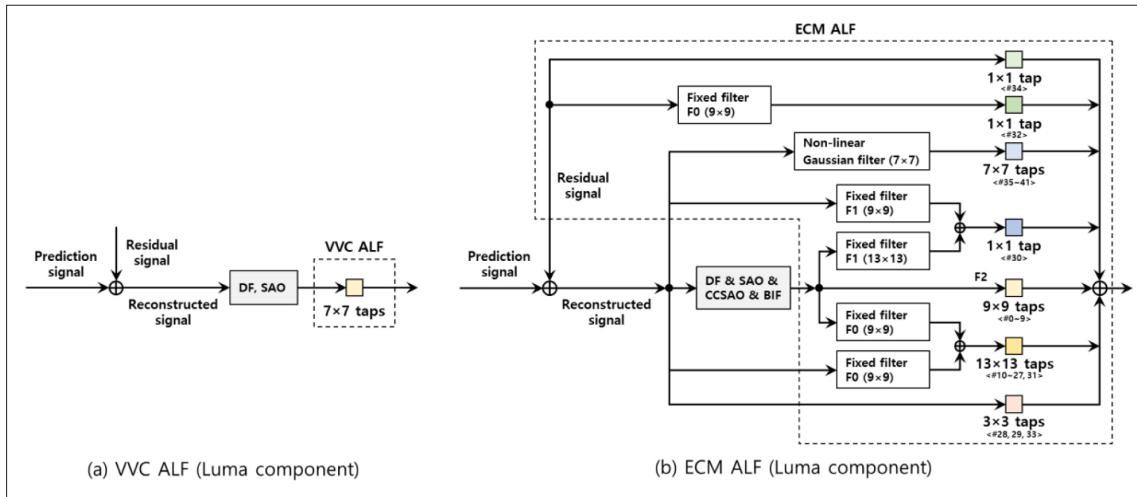
간의 왜곡을 직접 계산하여 부/복호화 손실을 보상해 주는 기술이며, 복호화 과정에서 가장 마지막에 수행되는 기술이다[5]. 또한, 부호화 복잡도는 높지만 부호화 효율은 매우 뛰어난 기술로, 이러한 성능을 바탕으로 현재 차세대 비디오 압축 표준 기술을 위한 ECM을 개발하는 과정에서도 꾸준히 관련 기술이 기고되고 있으며, ALF 관련 다양한 기술이 채택되어 더욱 정밀한 필터링을 수행하도록 하는 방향으로 발전되고 있다.

VVC의 ALF는 DF 및 SAO 수행 이후 샘플에 대해, 4×4 영역 단위로 25개의 클래스로 구별하고 클래스별로 부/복호화 손실을 최소화하는 필터를 생성하여 필터 셋을 전송하거나, 고정된 필터를 사용하여 필터링을 수행한다. 이때, 클래스와 필터에 대한 맵핑 정보는 APS (Adaptive parameter set)로 전송하고, 복호화기에서는 이를 파싱하여 사용한다[7].

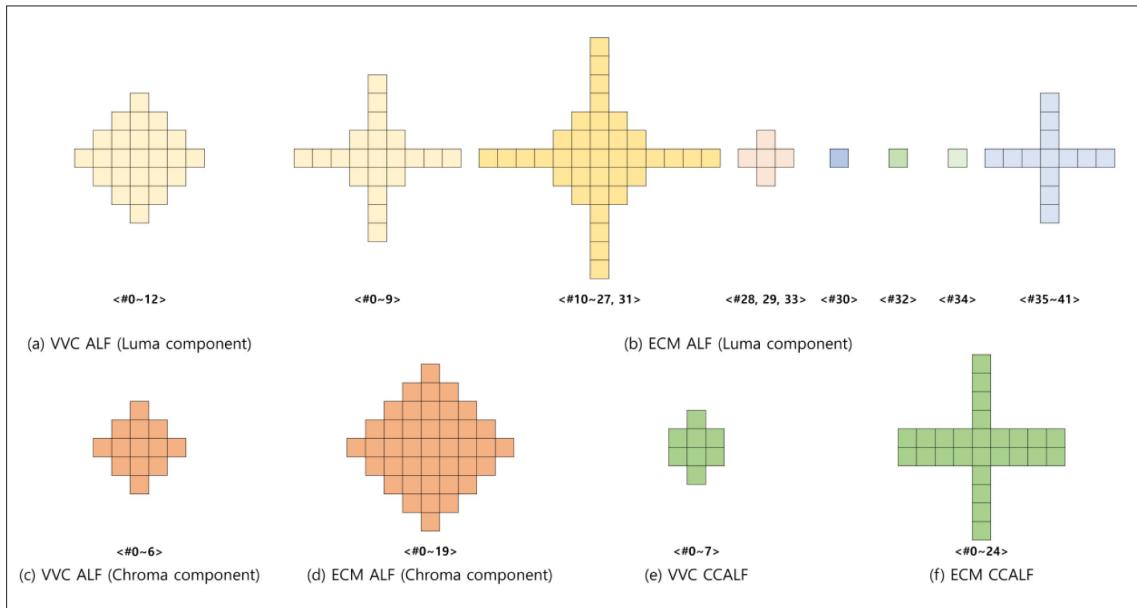
VVC의 ALF 대비 ECM의 ALF에서 크게 변화된 사항으로는 먼저, 다양한 입력 샘플을 활용하여 필터링을 수행하는 것이라고 볼 수 있다. <그림 7>과 같이, 휘도 성분을 기준으로 VVC에서는 SAO 이후의 샘플만을 사용했던 반면, ECM에서는 고정된 필터링 전/후의 차분값, 고정된 필터링 전/후의 DF 수행 이전의 복원값, 고정된 필터링 전/후의 DF, SAO, CCSAO 및 BIF 이후의 값을 모두 사용하여 ALF를 위한 필터링을 수행한다.

VVC에서는 4×4 영역 단위별로 방향성 (Directionality)과 활동성 (Activity)을 이용하여 클래스를 분류했지만, ECM에는 2×2 영역 단위별로 3개의 분류 (C0, C1, C2)를 수행한다. C0 및 C1은 <그림 7>의 고정된 필터 (Fixed filter) F0 및 F1을 결정하기 위한 과정이고 C2는 <그림 7>의 9×9 필터 (F2)를 적용하기 위한 클래스를 분류하는 과정으로, 3개의 분류기 (Classifier) 중 하나를 이용하여 25개의 클래스로 분류하며, 이는 6. 1) 절에서 자세히 다룬다.

또한, <그림 8>과 같이 필터의 모양이 변하고 필터 템수가 증가했다. 휘도 및 색차 성분에서 각각 다이아몬드 형태의 7×7 및 5×5 모양 필터를 사용했던 VVC의 ALF와는 달리, ECM에서는 각 입력 샘플마다 1×1 부터 13×13



<그림 7> VVC 및 ECM의 ALF 기술



<그림 8> VVC 및 ECM의 ALF 및 CCALF 필터 모양 및 탭 수

까지 다양한 모양의 필터를 사용한다.

CCALF는 색차 성분의 ALF 수행 이후 흐도와 색차 성분 간 상관관계를 활용하여 색차 성분에만 적용되는 기술로, 색차 성분의 샘플과 대응되는 흐도 샘플 주변의 샘플들을 이용하여 선형 필터링을 통해 ALF 수행 이후의 색차 성분

샘플을 보정하는 기술이다. 또한, ALF와 같이, Wiener 필터 기반 방식을 사용하여 MSE (Mean squared error) 감소를 위한 필터 계수를 구하여 Cb 및 Cr 성분에 독립적으로 적용될 수 있는 기술이다[6]. VVC의 CCALF 대비 ECM의 CCALF에서 변화된 사항은 <그림 8>(f)와 같이 필터

의 모양이 변하고, 텁 수가 증가하였다. 또한, VVC에서 CCALF의 필터는 APS로 Cb 및 Cr 각각 최대 4개를 전송했던 반면, ECM의 CCALF 필터는 APS를 통해 최대 16개의 필터를 전송할 수 있다[4][7]. 또한, 색차 성분의 ALF 및 CCALF에 추가로 휘도 성분의 차분값을 이용하여 필터링을 수행하는 방법이 채택되어 다음 버전의 ECM 참조 소프트웨어에 포함될 예정이다[10].

1) ALF classification

ECM의 ALF는 휘도 성분에 대해 3가지 분류 (C_0 , C_1 , C_2)를 2×2 영역 단위별로 수행한다. 이 중, C_0 및 C_1 은 각각 <그림 7>의 고정된 필터 F_0 , F_1 을 결정하기 위해 수행되며, C_2 는 VVC와 동일한 그래디언트 분류기, 밴드 분류기, 차분 분류기 중 하나를 적용하여 25개의 클래스로 분류하고, 결정된 클래스에 따른 필터 계수를 구하고, 필터 정보를 전송하여 <그림 7>의 9×9 필터 (F_2)를 적용하기 위해 수행된다. 또한, VVC에서는 4×4 영역 단위별로 10×10 윈도우에서 서브 샘플링하여 방향성을 구하였던 반면, ECM에서는 서브 샘플링 없이 모든 샘플에 대해 계산하는 과정을 수행한다.

F_0 및 F_1 은 9×9 또는 13×13 의 다이아몬드 모양 필터이고, VVC의 ALF 필터와 동일하게 대칭성을 가지고 있다. 또한, 각각 512개의 필터를 가지고 있는 8개의 필터 셋으로 구성되며, 필터 셋 정보는 슬라이스 단위로 전송된다. 그리고 각 필터는 C_0 및 C_1 분류의 결과를 통해 512개 중 하나로 결정된다. C_0 분류를 통해 2×2 영역 단위별로 4×4 윈도우의 방향성 및 활동성을 구하여 필터를 결정한다. 방향성을 구하기 위해 그래디언트 (Gradient)를 구하며, 이는 VVC에서와 동일하게 1D Laplacian 수직 (Vertical), 수평 (Horizontal), 두 대각 방향 (45 degree, 135 degree)을 구하고, 이를 이용하여 수직/수평 에지 강도 및 대각 에지 강도를 구한 후, 6개의 임곗값 {1.25, 1.5, 2, 3, 4.5, 8}과 조건에 따라 비교하여 56개 중 하나로 분류하며, 활동성은 수직 및 수평 그래디언트의 합으로 구하여 16개 중 하나의 값으로 분류한다[4]. C_1 분류는 2×2

영역 단위별로 12×12 윈도우의 방향성 및 활동성을 구하며, 두 결과에 따라 부/복호화기 간 정의되어 있는 필터를 결정하여 사용한다.

C_2 는 VVC와 동일한 그래디언트 분류기에 추가로 밴드 분류기, 차분 분류기로 구성되며, 분류기 인덱스를 APS로 전송하여 사용한다. 세 분류기 모두 25개의 클래스 중 하나로 결정하는 과정을 수행하며, 그래디언트 분류기를 통해 클래스를 결정하는 과정은 방향성 및 활동성을 이용한다. 이때, 방향성과 활동성을 구하는 과정은 C_0 및 C_1 분류에서의 과정과 동일하지만, VVC에서와 동일한 조건을 이용하여 5개의 활동성과 5개의 방향성으로 분류하여 25개의 클래스 중 하나로 분류한다. 밴드 분류기는 2×2 영역 단위별로 식(9)을 통해 25개의 클래스 중 하나 ($classIdx$)로 분류한다. 이때, $A \sim D$ 는 2×2 영역 내 각 샘플값을 의미한다. 차분 분류기는 2×2 영역 단위별로 이를 포함하는 8×8 윈도우 내 차분값들의 합 (sum)을 이용하여 식(10)을 통해 25개의 클래스 중 하나 ($classIdx$)로 분류하며, 차분 분류기는 인트라 슬라이스 (Intra slice)에는 적용되지 않는다. C_2 분류를 통해 <그림 7>의 9×9 필터 (F_2)를 적용하며, F_2 는 25개의 필터를 가지고 있는 2개의 오프라인 훈련된 (Offline trained) 필터 셋과 최대 8개의 온라인 훈련된 (Online trained) 필터 셋으로 구성되고, 필터 정보를 APS로 CTB (Coding tree block) 단위로 전송하여 사용한다.

$$classIdx = \{(A+B+C+D) \times 25\} \gg bitDepth + 2 \quad (9)$$

$$classIdx = \min(sum \gg (bitDepth - 4), 24) \quad (10)$$

색차 성분의 경우, VVC와 동일하게 APS로 전송되는 하나의 필터 셋과 8개의 필터 (7개의 필터 계수 및 클리핑 인덱스)를 통해 필터링을 수행하며, Cb 및 Cr 성분은 필터 셋을 공유하면서 서로 다른 필터를 사용할 수 있고, <그림 8>과 같이, 필터의 모양 및 텁 수가 증가하였다[10].

III. 결 론

본고에서는 JVET에서 VVC 이후 신호처리 기반 차세대 비디오 압축 표준 기술 개발을 위해 탐색 중인 ECM10.0에 포함된 인루프 필터링 기술에 대해 살펴보았다. LMCS, DF, SAO 및 ALF만을 수행하던 VVC 표준 기술의 인루프 필터링과 비교하여 LMCS, DF, SAO는 기술적인 변경은 없지만, 일부 사항이 ECM에 새롭게 채택된 기술을 고려하여 수정되었고, CCSAO 및 BIF 기술이 새롭게 채택되어 ECM에 포함되었다. 복잡도를 고려하여 SAO, CCSAO, BIF 기술은 서로 병렬적으로 수행되며, ALF 및 CCALF는

VVC 대비 더 넓은 범위의 샘플값 및 더욱 많은 입력 데이터를 사용하여 필터링을 수행함으로써 원본 영상과의 왜곡을 효과적으로 줄이도록 확장되었다. 새롭게 채택되거나 확장된 인루프 필터링 기술들로 인해 현재 ECM10.0은 VVC 대비 우수한 부호화 효율을 보여주고 있지만 높은 부/복호화 복잡도를 보인다. VVC 이후 차세대 비디오 압축 표준 기술의 표준화 과정에서 ECM에 포함된 인루프 필터링 기술이 반영될 가능성이 클 것으로 사료되며, 이에 따라 부호화 성능을 유지하면서 복잡도를 줄일 수 있는 연구가 필요할 것으로 보인다.

참 고 문 헌

- [1] M, Lee et al., "Overview of Versatile Video Coding (H.266/VVC) and Its Coding Performance Analysis," IEIE Trans. Smart Process. Comput., No. 12, Vol. 2, 122-154, Apr. 2023.
- [2] ECM10.0, https://vcgit.hhi.fraunhofer.de/ecm/ECM/-/tree/ECM-10.0?ref_type=tags
- [3] V. Seregin, J. Chen, R. Chernyak, F. L. Léannec, and K. Zhang, "JVET AHG report: ECM software development (AHG6)," 32nd JVET Meeting, Hannover, Doc. JVET-AF0006, Oct. 2023.
- [4] M. Coban, R.-L. Liao, K. Naser, J. Ström, and L. Zhang, "Algorithm description of Enhanced Compression Model 10 (ECM 10)," 31st JVET Meeting, Geneva, Doc. JVET-AE2025, Jul. 2023.
- [5] R.-L. Liao, X. Li, J. Chen, and Y. Ye, "Non-EE2: On combination of CIIP, OBMC and LMCS," 24th JVET Meeting, Teleconference, Doc. JVET-X0090, Oct. 2021.
- [6] 심동규, 조현호, "HEVC 표준 기술의 이해," 흥릉 과학출판사, 2014.
- [7] M. Karczewicz et al., "VVC In-Loop Filters," IEEE Trans. Circuits Syst. Video Technol., No. 31, Vol. 10, pp. 3907-3925, Oct. 2021.
- [8] Y. Kidani et al., "EE2-3.1: GPM with inter and intra prediction (JVET-X0166)," 25th JVET Meeting, Teleconference, Doc. JVET-Y0065, Jan. 2022.
- [9] V. Shchukin, P. Wennersten, and J. Ström, "EE2-5.1: Dynamic Scaling of Bilateral Filter (BIF)," 32nd JVET Meeting, Hannover, Doc. JVET-AF0112, Oct. 2023.
- [10] W. Yin, K. Zhang, Z. Deng, and L. Zhang, "EE2-5.2: Luma Residual Taps in Chroma-ALF and CCALF," 32nd JVET Meeting, Hannover, Doc. JVET-AF0197, Oct. 2023.

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과 (IITP-2023-RS-2023-00258639) 및 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (NRF-2021R1A2C2092848)의 연구결과 및 2023년도 광운대학교 교내 학술연구비 지원에 의해 수행되었음.

저자소개



이민훈

- 2019년 2월 : 광운대학교 수학과, 전자공학과 (복수전공) 학사
- 2021년 2월 : 광운대학교 전자공학과 석사
- 2021년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : <https://orcid.org/0000-0001-8165-5380>
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전



심동규

- 1993년 2월 : 서강대학교 전자공학과 공학사
- 1995년 2월 : 서강대학교 전자공학과 공학석사
- 1999년 2월 : 서강대학교 전자공학과 공학박사
- 1999년 3월 ~ 2000년 8월 : 현대전자 선임연구원
- 2000년 9월 ~ 2002년 3월 : 바로비전 선임연구원
- 2002년 4월 ~ 2005년 2월 : University of Washington Senior research engineer
- 2005년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 교수
- ORCID : <https://orcid.org/0000-0002-2794-9932>
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전