

ECM Inter prediction II

□ 최기호 / 경희대학교

요약

2020년 10월 국제표준화 단체 MPEG과 VCEG이 연합하여 만든 기구인 Joint video experts team (JVET)은 Versatile video coding (VVC)/H.266 표준을 완성하였다. JVET은 VVC/H.266 표준 이후를 준비하기 위한 스터디를 2021년부터 시작하였다. 하나의 방향은 신경망을 활용한 새로운 코딩 기술 연구이고, 다른 방향은 신호처리 기술을 활용한 압축 성능 향상 연구이다. 본고에서는 VVC/H.266 이후 차세대 비디오 코딩 표준의 기술 개발을 위해 탐색 실험 중인 신호처리 기반 비디오 부호화 기술 모델 Enhanced compression model (ECM) 중 화면간 예측의 대표적인 기술을 소개하고자 한다.

I. 서론

ISO/IEC JTC 1/SC 29 MPEG과 ITU-T Study Group 16 Question 6 VCEG 표준 기구에 의해 주도된 비디오 코딩 기술은 지난 반세기 동안 지속적인 연구와 협력을 통해 발전되어 왔다. 현재까지 이러한 노력으로 만들어진 표준 기술은 성공적으로 시장에서 채택되어 널리 사용되고 있으며, MPEG-2/H.262[1], MPEG-4 advance video coding (AVC)/H.264[2], MPEG-H high efficiency video coding (HEVC)/H.265[3] 등이 그 대표적인 예로

들 수 있다.

최근에는 MPEG과 VCEG가 공동으로 설립한 Joint video experts team (JVET)을 중심으로 MPEG-I Versatile video coding (VVC)/H.266[4] 표준이 완성되었다. VVC/H.266은 이전 세대 비디오 코딩 표준인 HEVC/H.265와 비교하여 약 2배의 압축률을 향상시켰으며, 이를 위해서 강력한 예측 모델, 향상된 움직임 보상, 개선된 변환 코딩 및 적응형 루프 필터링과 같은 고급 기술을 도입하였다. 이러한 향상된 성능은 4K 이상의 고해상도를 지원하는 최신 디스플레이와 다양한 응용 분야에

서 활용될 것으로 기대된다.

JVET은 2021년부터 VVC/H.266 이후 표준을 준비하기 위한 스터디를 시작하였다. 새로운 비디오 표준을 제정하기 위해서는 혁신적인 새로운 기술이 요구되는데, 이를 위해서 미리 표준에 적용할 수 있는 기술을 발굴하려는 것이 스터디의 목적이다. 현재 JVET은 두 가지 방향으로 후보 기술 탐색 실험을 진행하고 있다. 첫 번째는 신경망 기반의 영상 압축 기술 탐색 실험이고, 두 번째는 AVC/H.264, HEVC/H.265, VVC/H.266 규격 등 전통적인 블록 기반의 하이브리드 부호화 구조에서 압축률을 향상시키는 신호 처리 기반의 영상 압축 기술에 대한 연구이다.

신경망을 활용한 영상 압축 기술 탐색의 경우 최근 다양한 분야에서 사용되고 있는 신경망 기술을 영상 압축 기술에 적용하기 위한 시도이다. 신호 처리 기반의 영상 압축 기술 탐색의 경우 역사적으로 성공적으로 사용되어 왔던 영상 압축 요소 기술들을 확장 및 발전시켜 차세대 표준에 적용하려는 시도이다. 현재 JVET은 신호 처리 기반의 영상 압축 기술 탐색을 통해 도출된 결과물을 Enhanced compression model (ECM) SW에 채택하여 관리하고 있다. 본고에서는 JVET에서 차세대 비디오 코딩 표준의 기술 개발을 위해 탐색 실험 중인 ECM 기술 중 화면간 예측의 대표적인 기술을 소개하자 한다.

II. ECM 화면간 주요 기술

ECM SW에 들어있는 기술은 2021년 1월부터 기고를 받아 JVET에서 채택된 기술로 다음 표준 준비를 위해서 병합 및 관리되고 있다. 최신 ECM SW인 ECM-12.0 버전에서는 Random Access (RA) 시나리오 기준으로 VVC/H.266 표준 대비 약 22%의 성능 향상이 보고되었다[5]. 해당 수치인 VVC/H.266 대비 20% 이상의 성능 향상은 향후 표준을 시작하기 위해서 중요한 수치이고, 이는 다음 세대 비디오 코딩 표준을 ECM 기반으로 당장 시작해도 이상하지 않는 수치라 여겨진다.

ECM 활동을 통해서 획득한 코딩 성능은 다수의 기술 채택을 통해서 가능할 수 있었다. 특별히 화면간 예측 기술이 ECM SW에 많이 채택되었는데, 본고에서는 ECM SW에 채택된 화면간 예측 기술 중 Template Matching (TM)[6], Multi-pass decoder-side motion vector refinement (DMVR)[6][7], Adaptive decoder-side motion vector refinement[8], Sample-based bidirectional optical flow (BDOF)[6], Template matching based bi-prediction with CU-level weights (BCW) index derivation for merge mode[9], History-parameter-based affine model inheritance (HAMI)[10], Non-adjacent affine mode (NA-AFF)[10]을 중심으로 핵심 동작 아이디어를 설명하고자 한다.

1. Template Matching

TM은 디코더측 Motion vector (MV) 도출 방법 중 하나로, 현재 픽처 내의 현재 Coding unit (CU)의 상단 및/또는 좌측 이웃 블록의 템플릿과 참조 픽처 내의 블록 사이의 최적 일치 블록을 찾아 현재 CU의 MV를 개선하는 용도로 사용된다. TM 기술을 사용하면 디코더에서 정제된 MV를 찾기 때문에 인코더에서 전송하는 오버헤드를 줄일 수 있다.

구체적으로, Advanced motion vector prediction (AMVP) 모드에서는 템플릿 매칭 에러를 기반으로 Motion vector prediction (MVP) 후보를 결정하여, 현재 블록 템플릿과 참조 블록 템플릿의 최소 비용의 후보를 선택한 다음, MV의 정제 과정을 위해 선택된 후보에 대해서만 TM을 수행한다. TM은 16-point의 다이아몬드 검색을 사용하여 [-8, +8] 정수 픽셀 공간에서 정수 단위의 TM 작업을 수행하며, Adaptive motion vector resolution (AMVR) 모드에 따라 하프 픽셀 단위 및 쿼터 픽셀 단위의 교차 검색을 사용하여 추가적으로 모션을 정제하게 된다. 머지 모드에서는 인덱스가 나타내는 후보에 대해서도 AMVP와 유사한 검색 방식이 적용된다. AVMP와 다른 부

분은 머지 모드의 TM은 대체 보간 필터의 사용 여부에 따라 1/8 픽셀 단위의 정밀도까지 지원하거나 또는 하프 픽셀 단위의 정밀도 이상을 생략할 수 있다.

양방향 예측 블록에 TM을 적용하는 경우, 반복적인 프로세스를 통해 MV 정보가 정밀하게 업데이트된다. 구체적으로, L0 및 L1의 초기 움직임 벡터를 먼저 정제하고 L0 및 L1에 대해 각각의 비용인 Cost0, Cost1을 계산한다. Cost0이 Cost1보다 큰 경우, L1의 MV1의 정제된 움직임 벡터를 사용하여 L0의 MV0의 정제된 움직임 벡터를 도출한다. 이후 MV1은 MV0을 사용하여 추가적으로 MV를 정제한다. 유사하게, Cost0이 Cost1보다 크지 않은 경우, L0의 MV0의 정제된 움직임 벡터를 사용하여 L1의 MV1의 더 정제된 움직임 벡터를 도출하고 MV0은 MV1을 사용하여 더 정제된다. 또한, DMVR 조건이 만족되는 경우, 양방향 예측을 위한 TM이 활성화된다.

초기 제안 기고서인 [6]에 따르면 RA 실험 조건에서 TM의 성능은 VVC/H.265 코덱 대비 1.84%의 휘도 성능과 111%의 인코더 복잡도, 115%의 복잡도로 보고되었다.

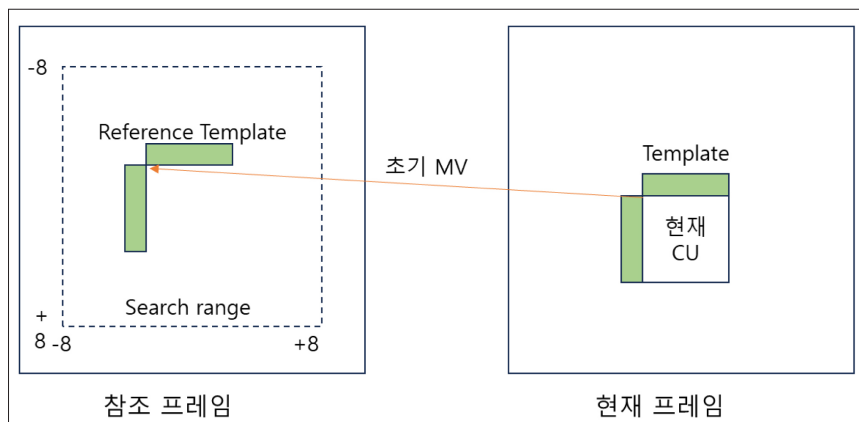
2. Multi-pass decoder-side motion vector refinement

DMVR 기술은 VVC/H.266 표준에 채택된 기술로,

ECM SW에 채택된 기술은 추가적인 성능 향상을 위해서 VVC/H.266의 DMVR 기술이 확장되었다. 구체적인 동작은 네 개의 패스를 통해서 MV를 정밀하게 보정하는 것이 특징인데, 핵심적으로 달라진 점은 적용하는 블록의 단위가 적응적으로 변경된 점과 BDof 기술이 DMVR과 연동되는 부분으로 볼 수 있다.

구체적으로, 첫 번째 패스에서는 블록 매칭을 CU에 적용하여 정제된 모션 MV를 도출한다. 이는 DMVR과 유사하게, 양방향 예측 동작에서 참조 픽처 리스트 L0 및 L1에서 두 개의 초기 MV0 및 MV1 주변에 정제된 MV를 검색한다. 정제된 MV(e.g., MV0_pass1 및 MV1_pass1)은 L0 및 L1에서 두 참조 블록들 사이의 최소 양방향 매칭 비용에 기초하여 유도한다. Bilateral matching (BM)은 deltaMV에서 정수 픽셀 정밀도를 도출하기 위해 로컬 검색을 수행한다. 로컬 검색은 3×3 크로스검색 패턴을 적용하여 수평 방향으로 [-sHor, sHor], 수직 방향으로 [-sVer, sVer] 검색하며, 여기서 sHor와 sVer의 값은 블록 크기에 의해 결정되며, sHor와 sVer의 최대값은 8이 될 수 있다. 첫 번째 패스 이후의 개선된 MV들은 다음과 같이 도출된다.

$$\begin{aligned} MV0_{pass1} &= MV0 + \text{deltaMV} \\ MV1_{pass1} &= MV1 - \text{deltaMV} \end{aligned} \quad (1)$$



<그림 1> 템플릿을 활용한 디코더에서 MV 정제 과정 개념도

두 번째 패스에서는 16×16 단위의 서브블록에 BM을 적용하여 정제된 MV를 도출한다. 각 서브블록에 대해 참조 픽처 리스트 L0, L1에서 제1 패스에서 획득된 두 개의 MV0_pass1, MV1_pass1를 중심으로 정제된 MV를 검색한다. 정제된 MV는 L0, L1에서 두 기준 서브블록 간의 최소 양방향 정합 비용에 기초하여 도출한다. 각 서브블록에 대하여 BM은 deltaMV에서 정수 샘플 정밀도를 도출하기 위해 전체 검색을 수행한다. 전체 검색은 수평 방향으로 [-sHor, sHor], 수직 방향으로 [-sVer, sVer]의 검색 범위를 가지며, 이때 sHor와 sVer의 값은 블록 크기에 의해 결정되며, 첫 번째 패스와 동일하게 sHor와 sVer의 최대값은 8이다. 검색 시 두 서브블록 사이의 비용은 Sum of absolute transformed differences (SATD) 비용에 cost factor를 적용하여 양자 매칭 비용을 계산한다. 검색 영역 $(2sHor + 1) * (2sVer + 1)$ 은 최대 5개의 다이아몬드 형상 검색 영역으로 구분되며, 각 검색 영역에는 각 검색 지점과 시작 MV 사이의 거리에 의해 결정되는 costFactor가 할당되며, 각 다이아몬드 영역은 검색 영역의 중심에서 시작하는 순서로 처리된다. 만약 현재 검색 영역 내의 최소 비용이 sbW * sbH와 동일한 임계 값 미만이면 정수 픽셀 단위 검색이 종료되고, 그렇지 않으면 모든 검색 지점을 검사할 때까지 다음 검색 영역으로 계속된다. 또한 반복에서 이전 최소 비용과 현재 최소 비용의 차이가 블록

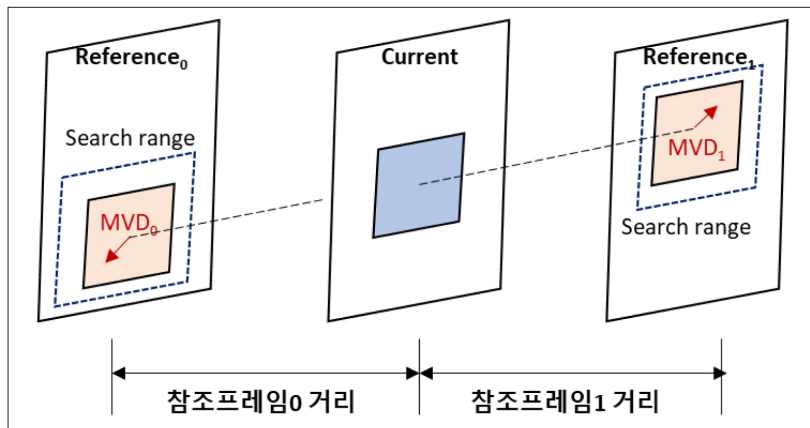
의 영역과 동일한 임계 값 미만이면 검색 프로세스를 종료한다. 두 번째 패스에서의 개선된 MV들은 다음과 같이 도출된다. 여기서 sbIdx2는 서브블록의 인덱스를 나타낸다.

$$\begin{aligned} MV0_pass2(sbIdx2) &= MV0_pass1 + \text{deltaMV}(sbIdx2) \\ MV1_pass2(sbIdx2) &= MV1_pass1 - \text{deltaMV}(sbIdx2) \end{aligned} \quad (2)$$

세 번째 패스에서는 BDOF를 8×8 단위 서브블록에 적용하여 정제된 MV를 도출한다. 각 8×8 서브블록에 대해, BDOF 정제가 적용되어 정제된 MV를 획득하기 위해 Vx 및 Vy를 도출한다. 유도된 Vx, Vy는 $1/16$ 샘플 정밀도로 반올림되고 -32와 32 사이에서 클리핑한다. 세 번째 패스에서의 정제된 MV(i.e., MV0_pass3(sbIdx3)) 및 MV1_pass3(sbIdx3)은 다음과 같이 된다. 여기서 bioMv는 BDOF를 통해서 추가적으로 정제된 MV 변화량을 나타낸다.

$$\begin{aligned} MV0_pass3(sbIdx3) &= MV0_pass2(sbIdx2) + \text{bioMv} \\ MV1_pass3(sbIdx3) &= MV0_pass2(sbIdx2) - \text{bioMv} \end{aligned} \quad (3)$$

네 번째 패스에서는 4×4 또는 8×8 단위의 서브블록에 BDOF를 적응적으로 적용하기 위해서, 블록이 1024 픽셀 보다 작은 경우 4×4 서브블록 단위로 BDOF를 적용하고,



<그림 2> Multi-pass DMVR 동작 개념도

그렇지 않으면 8×8 서브블록 단위로 BDOF를 적용한다. 각 서브블록의 MV 정제는 세 번째 패스에서 사용된 것과 동일한 방식으로 정제한다.

초기 제안 기고서인 [6]에 따르면 RA 실험 조건에서 TM과 Multi-pass DMVR을 사용하는 경우 VVC/H.265 코덱 대비 3.07%의 휘도 성능과 141%의 인코더 복잡도, 220%의 복잡도로 보고되었다.

3. Adaptive decoder-side motion vector refinement

Adaptive DMVR 방법은 DMVR 조건을 만족하는 머지 후보에 대한 양방향 예측의 한 방향(e.g., L0 또는 L1)으로만 MV를 정제하기 위해 두 개의 새로운 머지 모드로 구성된 DMVR 활용 기술이다. 기존 DMVR 프로세스는 모션 벡터를 정제하기 위해 선택된 머지 후보에 대해 적용되지만, Adaptive DMVR에서 MVD0 또는 MVD1 중 하나가 0으로 설정된다.

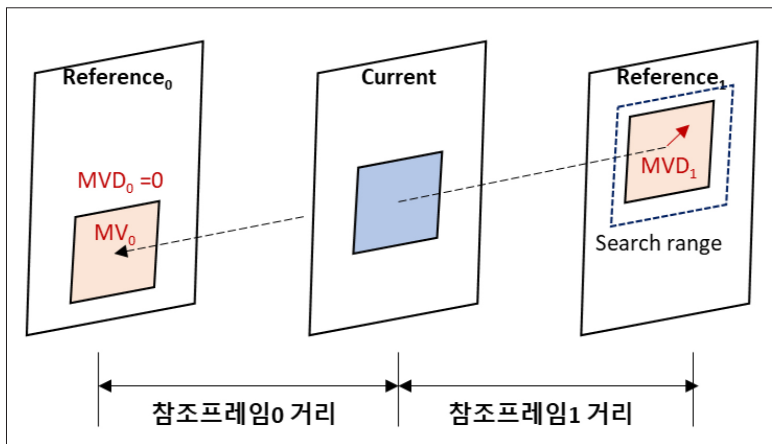
Adaptive DMVR을 활용하기 위한 새로운 머지 모드에 대한 후보는 일반적인 머지 모드에서와 유사하게 공간적 주변 블록, TMVP, 비인접 블록, HMVP, Pair-wise 후보에서 도출된다. Adaptive DMVR이 일반 머지 모드와 다

른 점은 DMVR 조건을 만족하는 것들만 후보 리스트에 추가된다는 부분이다. 만약 BM 후보 리스트에 상속된 BCW 가중치가 포함되어 있고, 가중치가 동일하지 않고 양방향 예측이 BCW 가중치로 부여된 경우 Mean removed sum of absolute differences (MRSAD) 또는 Mean removed sum of absolute transformed differences (MRSATD)를 사용하여 RD 계산을 사용하는 부분을 제외하고, 기존 DMVR 프로세스와 동일하다. 그리고, Adaptive DMVR을 위한 머지 인덱스 또한 일반 머지 모드에서와 동일하게 부호화한다.

초기 제안 기고서인 [8]에 따르면 RA 실험 조건에서 적응적 DMVR을 사용하는 경우 VVC/H.265 코덱 대비 0.84%의 휘도 성능과 106%의 인코더 복잡도, 97%의 복잡도로 보고되었다.

4. Sample-based bidirectional optical flow

BDOF 기술은 HEVC/H.265 표준 기술 개발 시 소개되어, VVC/H.266 표준에 채택되었다. ECM SW에 채택되어 있는 BDOF의 기술은 추가적인 성능 향상을 위해서 VVC/H.266의 BDOF 기술이 확장된 기술로 간주된



<그림 3> Adaptive DMVR 동작 개념도

다. 초기 BDOF로 제시된 기술의 경우, 양방향 예측 과정에서 광학 흐름을 사용하여 양방향 예측 값을 움직임의 변화량에 따라 픽셀단위로 정제하는 기술이 사용되었으나, VVC/H.266에 BDOF가 채택될 당시 디코더에서 복잡도를 줄이기 위해서 4×4 단위의 동일한 정제된 예측 값이 적용되는 형태로 변경되었다. 이러한 맥락에서 현 ECM SW에 채택되어 있는 BDOF의 기술은 초기 BDOF 기술로 다시 변경된 형태로 간주할 수 있다.

구체적으로, 샘플 기반 BDOF에서는 블록 단위로 움직임 미세화(i.e., V_x , V_y)를 도출하는 대신, 샘플별로 수행된다. 부호화 블록은 8×8 개의 서브블록으로 구분되며, 각 서브블록에 대하여 임계 값에 대한 두 기준 서브블록 사이의 Sum of absolute differences (SAD)를 확인하여 BDOF를 적용할지 여부를 결정한다. 서브블록에 BDOF를 적용하기로 결정된 경우, 서브블록의 모든 샘플에 대해 슬라이딩 5×5 윈도우를 사용하고, 슬라이딩 윈도우마다 기존의 BDOF 프로세스를 적용하여 V_x 와 V_y 를 도출한다. 도출된 V_x , V_y 를 적용하여 윈도우의 중심 샘플에 대한 양방향 예측 샘플 값을 조정한다.

이러한 샘플 기반 BDOF 기술은 움직임 추정의 정확도를 향상시키며, 샘플 수준에서의 정밀한 움직임 정보를 활용함으로써, BDOF는 동적인 시나리오에서의 효과적인

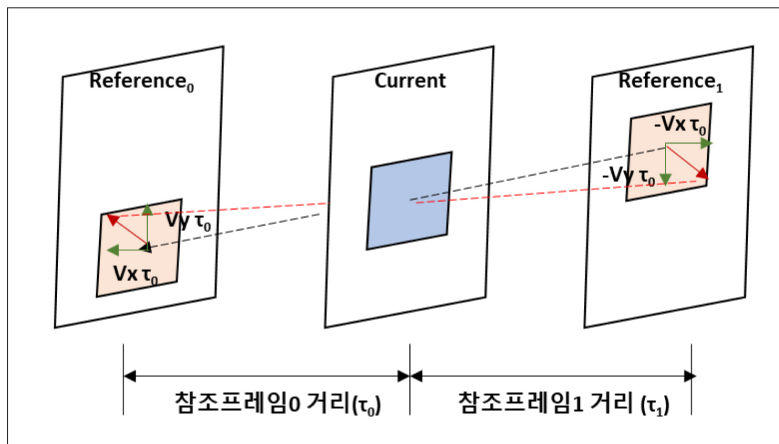
비디오 부호화를 지원할 수 있다.

ECM상에서 샘플 기반 BDOF 성능을 단독으로 제시한 기고는 없지만, VVC/H.266 개발 시 제안된 기고서에[11] 따르면 샘플 기반 BDOF를 사용하는 경우 약 1.41%의 휘도 성능과 103%의 인코더 복잡도, 119%의 복잡도가 나타남이 보고된 바 있다.

5. Template matching based BCW index derivation for merge mode

BCW의 기술에서 중요하게 여겨지는 부분 중 하나는 두 개의 예측 값에 가중치를 적응적으로 적용하여 더욱 정밀한 예측 값으로 도출하는 부분이다. 이때 예측 값에 적용되는 가중치는 준비된 테이블에 기초하여 최종적으로 하나가 선택되고, 선택된 가중치의 인덱스는 디코더로 전송되어 디코더에서 동일한 가중치가 적용될 수 있도록 한다. 전송되는 시그널의 오버헤드를 줄이고자 제시된 기술이 Template matching based BCW index derivation for merge mode이다.

구체적으로 머지 모드로 부호화된 CU들에 대한 BCW 인덱스는 템플릿 매칭 비용에 기초하여 유도된다. 선택된 머지 후보가 주어지면, TM 비용 값들은 서로 다른 양방향



<그림 4> 샘플 기반 BDOF 동작 개념도

예측 가중치들로 계산된 후, 최소 TM 비용 값을 갖는 양방향 예측 가중치가 머지 CU를 예측하는데 사용된다. 예를 들어, 예측 가중치에 대한 TM 비용을 계산할 때 다음과 같은 규칙이 적용된다.

- 상속된 양방향 예측 가중치는 다른 가중치에 비해 정확도가 높을 가능성이 있기 때문에 상속된 양방향 예측 가중치와 그 주변의 두 가중치(e.g., ± 1)만을 고려한다. 예를 들어, 상속된 양방향 예측 가중치가 4인 경우, TM 비용 계산에는 세 개의 가중치 {3, 4, 5}만 참여한다.
- 상속된 BCW 인덱스의 TM 비용은 0.90625를 곱한 값으로 적용하여 비용을 감소시킨다.
- 동일한 가중치를 갖는 CU에는 TM 비용에 0.90625를 곱해 해당 비용을 감소시킨다.

템플릿 매칭 기반 BCW 인덱스 도출은 일반 머지, 템플릿 매칭, Adaptive DMVR 및 Merge with motion vector difference (MMVD) 모드로 코딩된 CU에 적용된다. 또한, 머지 모드에 대한 양방향 예측 가중치는 {-2, 3, 4, 5, 10}에서 {1, 2, 3, 4, 5, 6, 7}로 변경하고, 머지 모드가 아닌 경우에 사용되던 {-2, 10}에 대한 음의 양방향 예측 가중치

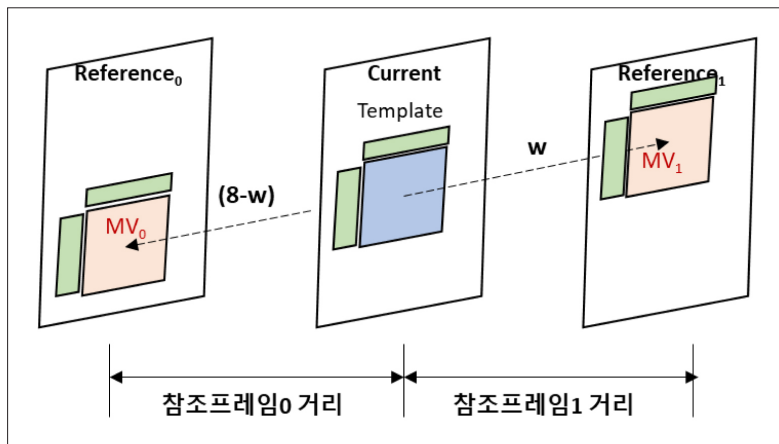
는 양의 가중치 {1, 7}로 대체한다.

제안 기고서인 [9]에 따르면 RA 실험 조건에서 제안 방법은 VVC/H.265 코덱 대비 0.11%의 휘도 성능과 100%의 인코더 복잡도, 100%의 복잡도로 보고되었다.

6. History-parameter-based affine model inheritance

HAMI은 현재 블록에 인접하지 않을 수 있는 이전의 Affine으로 부호화된 블록으로부터 Affine 모델을 상속할 수 있도록 한다. ECM에 채택된 일반 머지 모드와 유사하게, Affine에서도 비인접 모드를 도입하였다. 구체적으로 HMVP와 유사하게 이전 블록에서 Affine이 사용되었으면 해당 정보를 테이블에 저장하여 이후에 Affine 사용이 활용될 수 있도록 한다.

구체적으로, 첫 번째 history-parameter table(HPT)의 경우, 엔트리는 각각 16비트 부호화된 정수로 표현되는 Affine 파라미터 세트(i.e., a, b, c, d)를 저장한다. HPT 내의 엔트리들은 참조 리스트 및 참조 인덱스에 따라 분류되며, 5개의 참조 인덱스는 각각의 참조 리스트에 대해 지원한다. HPT의 카테고리는 다음과 같이 계산된다.



<그림 5> 템플릿 기반 BCW 인덱스 유도 동작 개념도

$$\text{HPTCat}(\text{RefList}, \text{RefIdx}) = 5 \times \text{RefList} + \min(\text{RefIdx}, 4) \quad (4)$$

여기서 RefList와 RefIdx는 각각 참조 픽처 리스트(e.g., L0 또는 L1)와 참조 인덱스를 나타낸다. 각 카테고리에 대해 최대 7개의 엔트리가 저장될 수 있으며, 따라서 HPT에는 총 70개의 엔트리가 있게 된다. 각 Coding tree unit (CTU) 행의 시작에서 각 카테고리에 대한 엔트리 수는 0으로 초기화시킨다. 참조 리스트 RefListcur와 RefIdxcur로 Affine으로 부호화된 CU를 디코딩한 후, Affine 파라미터들은 HMVP 테이블 업데이트와 유사한 방식으로 카테고리 HPTCat(RefListcur, RefIdxcur)의 엔트리들을 업데이트하는 데 사용된다.

History-affine-parameter-based candidate (HAPC)은 7개의 인접한 4×4 블록 중 하나 및 첫 번째 HPT의 대응하는 엔트리에 저장된 Affine 파라미터 세트로부터 유도된다. 인접한 4×4 블록의 MV는 베이스 MV로 가능하며, (x, y) 위치에서의 현재 블록의 MV는 다음과 같이 계산된다.

$$\begin{aligned} \text{mv}^h(x, y) &= a(x - x_{\text{base}}) + c(y - y_{\text{base}}) + \text{mv}_{\text{base}}^h \\ \text{mv}^v(x, y) &= b(x - x_{\text{base}}) + d(y - y_{\text{base}}) + \text{mv}_{\text{base}}^v \end{aligned} \quad (5)$$

여기서, $(\text{mv}_{\text{base}}^h, \text{mv}_{\text{base}}^v)$ 는 인접한 4×4 블록의 MV를 나타내고, $(x_{\text{base}}, y_{\text{base}})$ 는 인접한 4×4 블록의 중심 위치를

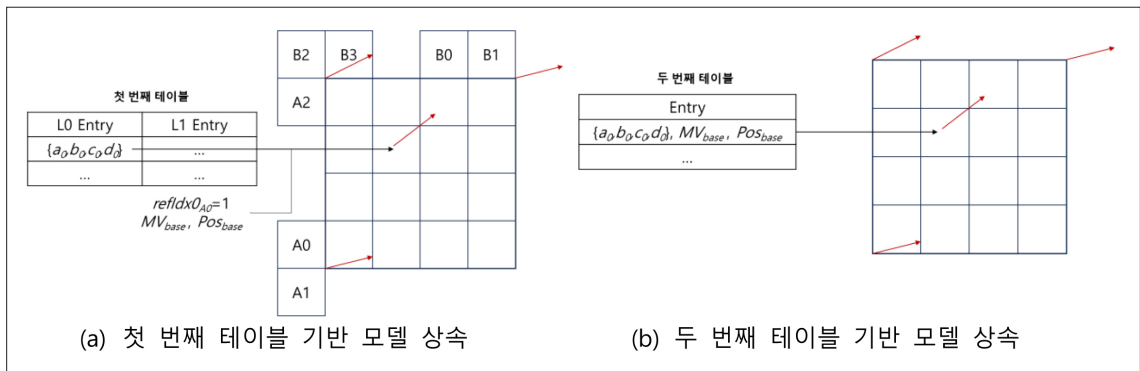
나타낸다. (x, y) 는 현재 블록에 대한 코너-위치 Control Point Motion Vector (CPMV)을 얻기 위해 현재 블록의 상-좌, 상-우 및 하-좌의 코너일 수도 있고, 또는 현재 블록에 대한 정규 MV를 얻기 위해 현재 블록의 중심일 수도 있다.

베이스 MV 정보를 갖는 두 번째 HPT도 추가되었다. 제 2 HPT에는 9개의 엔트리들이 존재하며, 엔트리는 베이스 MV, 참조 인덱스 및 각 참조 리스트에 대한 4개의 Affine 파라미터들 및 베이스 위치를 포함한다. 엔트리에 저장된 대응하는 베이스 MV 정보를 갖는 제2 HPT로부터 추가 머지 HAPC가 생성될 수 있다. Pair-wised Affine 머지 후보들은 히스토리 기반 혹은 비 히스토리 기반의 2개의 Affine 머지 후보들을 사용하여 생성된다. Pair-wised Affine 머지 후보들은 리스트 내의 기존 Affine 머지 후보의 CPMV들을 평균함으로써 생성된다.

제안 기고서인 [10]에 따르면 RA 실험 조건에서 제안 방법은 VVC/H.265 코덱 대비 0.27%의 회도 성능과 102%의 인코더 복잡도, 101%의 복잡도로 보고되었다.

7. Non-adjacent affine mode

NA-AFF는 공간적으로 인접하지 않은 블록들을 찾아 이를 바탕으로 Affine의 후보들을 생성하는 기술이다. 비 인접 일반 머지 후보와 동일하게, NA-AFF에서는 비인접



<그림 6> 히스토리 기반 Affine 모델 상속 개념도

공간적 주변 블록과 현재 코딩 블록 사이의 거리를 현재 CU의 폭과 높이를 기준으로 정의한다. 같은 프레임 내의 비인접 주변 블록의 움직임 정보는 추가적으로 상속되어 Affine 머지/AMVP 후보들을 생성하는 데 활용된다. 특히, 상속된 후보에 대해 CPMV들은 비인접 주변 블록으로부터 상속되지만, VVC/H.266과 동일하게 상속된 Affine 머지/AMVP 후보들의 유도 프로세스는 변경되지 않았다. 비인접 주변 블록은 현재 블록에 대한 거리를 기준으로 근접에서 원거리까지 탐색한다. 예를 들어, 좌측 및 상측으로부터 Affine 모드로 코딩된 첫 번째 이용 가능한 주변 블록이 상속된 후보로 포함된다. 좌측 및 상측의 블록들의 검사 순서는 각각 아래에서 위, 우에서 좌측으로 진행된다. 첫 번째 유형의 구성된 후보에 대해 1개의 좌측 및 상측 비인접 주변 블록의 위치를 먼저 독립적으로 결정하고, 그 후 상측-좌측 주변 블록을 결정하는 직사각형 가상 블록을 만든 다음, 3개의 비인접 이웃의 움직임 정보를 사용하여 가상 블록의 상측-좌측, 상측-우측, 및 하측-좌측에 CPMV를 형성하고, 이를 최종적으로 현재 CU에 투영하여 해당 구성된 후보를 생성한다. 최종적으로 NA-AFF

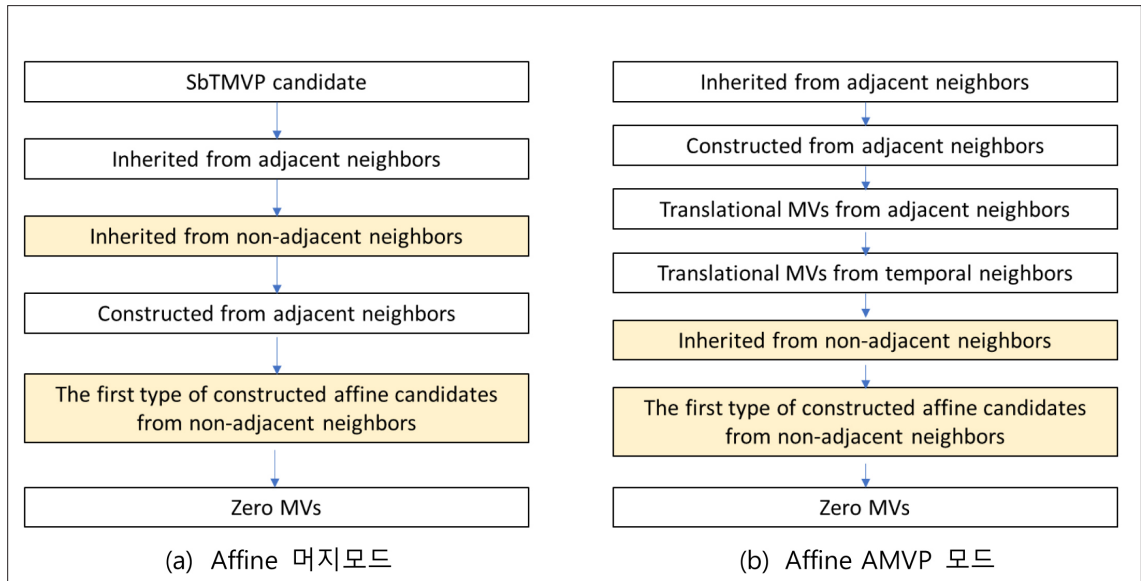
이 적용된 Affine 머지/AMVP 모드의 후보군 생성 순서는 <그림 7>과 같다.

제안 기고서인 [10]에 따르면 RA 실험 조건에서 제안 방법은 VVC/H.265 코덱 대비 0.30%의 휘도 성능과 101%의 인코더 복잡도, 101%의 복잡도로 보고되었다.

III. 결론

본고에서는 JVET에서 차세대 비디오 코딩 표준의 기술 개발을 위해 탐색 실험 중인 ECM 기술 중 화면간 예측의 대표적인 기술을 살펴보았다. 화면간 예측 기술의 경우 전통적으로 비디오 압축 성능을 향상시키는 것에 큰 영향을 끼쳐왔다. 현 ECM SW의 경우에도 상당 부분의 성능 향상이 다수의 화면간 예측 기술 채택으로 이루어졌고, 또 다음 세대 비디오 코딩 표준에도 ECM과 유사하게 다수의 화면간 예측 기술 채택이 있을 것이라 생각된다.

본고에서 살펴본 화면간 예측 기술의 경우 이전 세대의 비디오 코딩 표준과 비교하여 큰 특징이 있음이 확인



<그림 7> 비인접 정보를 활용한 Affine 후보군 생성 순서

된다. 해당 특징은 디코더에서 추가적인 처리를 통해서 전송 오버헤드를 줄이거나 또는 예측 값의 정밀도를 높여 비트레이트를 줄이는 것이다. 이러한 접근 방식은 디코더의 복잡성을 높이는 측면에서 이전 표준에서 회피해 온 기술 영역이다. 그러나 미래의 디바이스의 컴퓨팅 파워 향상을 전제로 하여 디코더에서 추가 계산을 수행하는 것은 한계

에 다른 압축 성능 향상을 위해 필수적인 부분이 될 것으로 예상된다.

현재 JVET에서 차세대 비디오 코딩 표준 준비를 위해서 진행 중인 ECM 활동의 추이를 살펴보고 다가올 차세대 비디오 코딩 표준이 어떠한 모습으로 만들어질지 관심이 있게 지켜볼 필요가 있다고 사료된다.

참 고 문 헌

- [1] Generic Coding of Moving Pictures and Associated Audio Information-Part 2: Video, ISO/IEC 13818-2 MPEG-2 and ITU-T Recommendation H.262, 1994.
- [2] Advanced Video Coding (AVC), ITU-T Recommendation H.264 and ISO/IEC 14496-10, May 2003.
- [3] High Efficient Video Coding (HEVC), ITU-T Recommendation H.265 and ISO/IEC 23008-2, Apr. 2013.
- [4] Versatile Video Coding (VVC), ITU-T Recommendation H.266 and ISO/IEC 23090-3, Jul. 2020.
- [5] V. Seregin, J. Chen, R. Chernyak, F. Le Léanec, K. Zhang, "JVET AHG report: ECM software development (AHG6)", JVET-AF0006, Oct. 2023.
- [6] Y.-J. Chang, C.-C. Chen, J. Chen, J. Dong, H. E. Egilmez, N. Hu, H. Huang, M. Karczewicz, J. Li, B. Ray, K. Reuze, V. Seregin, N. Shlyakhov, L. Pham Van, H. Wang, Y. Zhang, Z. Zhang, "Compression efficiency methods beyond VVC", JVET-U0100, Jan. 2021.
- [7] M. Salehifar, Y. He, K. Zhang, H. Liu, L. Zhang, J. Chen, R.-L. Liao, X. Li, Y. Ye, "EE2-3.5: Iterative BDOF pass in multi-pass DMVR", JVET-AE0065, Jul. 2023.
- [8] H. Huang, Z. Zhang, V. Seregin, W.-J. Chien, C.-C. Chen, M. Karczewicz, "EE2: Adaptive decoder side motion vector refinement (test 3.4)", JVET-X0049, Oct. 2021.
- [9] R.-L. Liao, J. Chen, Y. Ye, X. Li, "EE2-2.2: Template matching based BCW index derivation for merge mode", JVET-AB0079, Oct. 2022.
- [10] W. Chen, X. Xiu, H.-J. Jhu, C.-W. Kuo, X. Wang, K. Zhang, L. Zhang, Z. Deng, N. Zhang, Y. Wang, "EE2-2.7, 2.8, 2.9: History-parameter-based affine model inheritance and non-adjacent spatial neighbors for affine merge mode", JVET-Z0139, Apr. 2022.
- [11] X. Xiu, Y. He, Y. Ye, "CE9-related: Complexity reduction and bit-width control for bi-directional optical flow (BIO)", JVET-L0256, Sep. 2018.

저 자 소 개



최 기 호

- 2008년 : 한양대학교 정보통신대학 미디어공학과 학사
- 2012년 : 한양대학교 전자컴퓨터통신공학과 박사
- 2012년 ~ 2014년 : 한양대학교 부설연구소 Post Doc.
- 2014년 ~ 2021년 : 삼성전자 삼성리서치 책임연구원
- 2021년 ~ 2023년 : 가천대학교 AI·소프트웨어학부 조교수
- 2023년 ~ 현재 : 경희대학교 전자공학과 조교수
- 주관심분야 : 영상처리, 비디오압축, 딥러닝