

Overview of ECM

□ 전상훈, 임성창*, 강정원*, 김재곤 / 한국항공대학교, *ETRI

요약

ISO/IEC MPEG과 ITU-T VCEG이 공동으로 설립한 비디오 부호화 표준화 단체인 JVET(Joint Video Experts Team)은 2021년 1월부터 VVC 표준 이후의 새로운 비디오 부호화 표준(가칭: Beyond VVC)을 위한 후보 기술 탐색실험을 진행하고 있으며, 본고에서는 Beyond VVC의 기술 탐색실험을 위한 탐색모델 소프트웨어인 ECM(Enhanced Compression Model)의 개발 과정과 채택된 기술들의 주요 기술적 특징을 살펴보고 각 분야별 채택 기술을 간략히 소개한다.

I. 서론

비디오 부호화 국제 표준화 단체인 ISO/IEC JTC 1/SC 29 MPEG과 ITU-T SG 16 WP 3 VCEG은 공동 협력팀을 구성하여 비디오 부호화 국제 표준을 개발 및 관리해 오고 있으며, 2020년에 VVC(Versatile Video Coding)[1] 표준화를 완료하였다. VVC 표준은 더 커진 블록으로부터 다양한 블록분할 기술, 색성분 간 중복성 활용 기술, 간소화된 신경망 기반 화면내 예측 기술, 큰 블록으로부터 분할된 작은 블록에서 정보 유도 및 예측 수행 기술, 화면간 예

측 성능 향상을 위한 디코더 복잡도 부과 기술, 더 큰 변환/비정방형 변환/2차 변환 기술, 2가지의 새로운 인루프(in-loop) 필터 기술 등의 부호화 툴들을 채택하여 기존 HEVC(High Efficiency Video Coding) 표준보다 2배의 압축률을 달성하였다. 특히, VVC 표준은 그 명칭과 같이 4K/8K의 고해상도 영상뿐만 아니라 고화질 스트리밍, 몰입형(immersive) 가상현실 제공, 스크린 콘텐츠 지원까지 다양한 영상 응용 서비스에 유연하게 대응할 수 있다[2].

한편, JVET은 VVC 표준 이후 다음 세대의 새로운 비

디오 부호화 표준(Beyond VVC) 개발을 위한 다양한 기술을 검토하기 위해 두 가지의 AHG(Ad-Hoc Group)[3], [4]를 구성하였고, 이를 위해 두 기술 분야에 대한 탐색실험(EE: Exploration Experiment)을 진행하고 있다. 즉, 2020년 7월부터 신경망 기술을 종래의 블록 기반의 하이브리드 부호화 구조에 부호화 툴로써 적용하는 신경망 기반 부호화 툴 탐색실험(EE1: Neural Network-based Video Coding)[5]과 2021년 1월부터 종래의 신호처리 기반의 부호화 툴의 압축률 향상을 위한 탐색실험(EE2: Enhanced Compression beyond VVC Capability)[6]을 진행하고 있다. 그중 EE2에서는 탐색모델 소프트웨어인 ECM을 활용하여 성능검증 등의 탐색실험을 진행하고, 채택된 기술은 ECM에 통합하여 관리하고 있다. 현재 ECM에는 VVC 표준화 과정에서 구현 복잡도로 인해 채택되지 않은 기술, VVC 기술을 개선하는 기술, 새롭게 제안된 기술들을 포함하고 있으며, 향후 새로운 표준에 대한 표준화가 시작되면 ECM에 채택된 기술들 중 상당수가 새로운 표준에 채택될 가능성이 높을 것으로 예상된다.

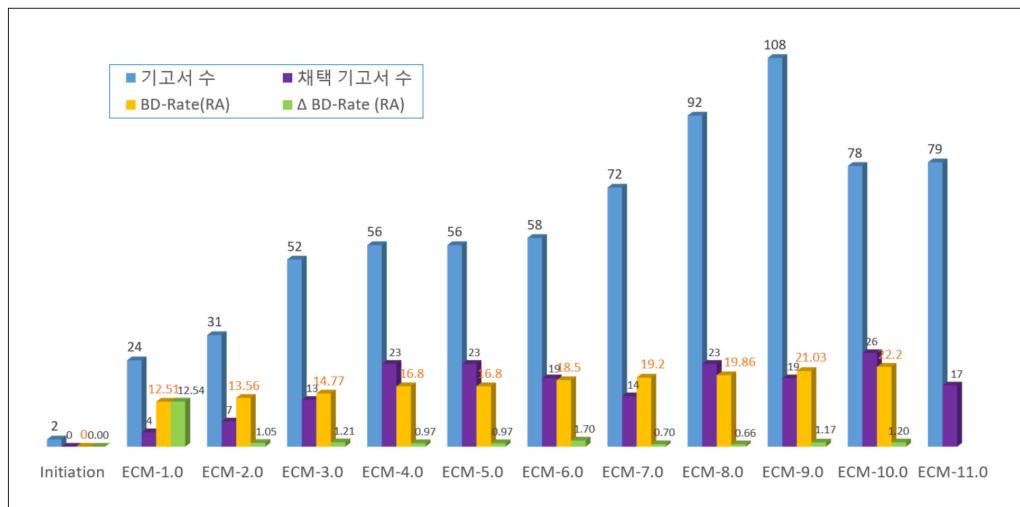
본고는 제31차 JVET 회의의 결과로 배포된 ECM-10.0[7],[8]에 채택된 주요 기술들을 소개한다. II장에서는 VVC 표준과 비교하여 각 기술 분야별로 ECM-10.0에서

개선되거나 추가된 부호화 툴들에 대하여 설명한다. III장에서는 VVC 표준의 탐색모델 소프트웨어인 VTM(VVC Test Model) 대비 ECM-10.0의 부호화 효율 및 부호화/복호화 시간을 살펴보고, IV장에서 결론을 맺는다.

II. ECM 개요 및 부호화 툴

1. ECM 개발 과정

JVET은 2021년 1월에 제안된 “Compression efficiency methods beyond VVC”[9] 기고서를 바탕으로 AHG12 (Enhanced compression beyond VVC capability) 및 EE2를 구성한 이후 ECM 개발과 동시에 꾸준히 차세대 비디오 부호화 표준 후보 기술에 대한 탐색실험을 진행해 오고 있다. 초창기 ECM 버전에는 VVC 표준화 과정에서 구현 복잡도로 인해 채택이 되지 않은 기술들이 대거 채택되었다. 이후 VVC 표준에 포함된 기술의 개선 기술들뿐만 아니라 새롭게 제안된 기술들 및 이러한 기술들로부터 파생된 기술들이 꾸준히 채택되고 있다. <그림 1>과 같이 매 JVET 표준화 회의마다 상당수의 기술들이 제안



<그림 1> ECM 개발 과정(회의별 기고서 현황 및 부호화 효율 향상)

되고 있으며, 부호화 효율 측면에서 검증된 기술들이 채택되면서 ECM 버전이 높아질수록 객관적 부호화 효율 측도인 BD(Bjøntegaard Delta)-Rate이 향상되고 있다[10].

2. ECM의 주요 기술적 특징

VVC 표준과 비교하여, ECM에 채택된 주요 부호화 툴들의 기술적 특징은 <표 1>과 같이 정리할 수 있다. ECM에는 기본적으로 VVC 표준에서 구현 복잡도 감소를 위해 포함된 기술들의 제거와 함께 VVC 표준화 과정에서 구현 복잡도로 인해 채택되지 않은 기술이 많이 포함되어 있다. 그리고, VVC 표준의 개선 기술 및 새로운 기술로는 템플릿(template)을 활용한 기술, 참조영역과 탐색영역을 확장하는 기술, 다수의 예측블록을 융합하는 기술, 합성곱 필터 기반 색성분 간 중복성 제거 기술, 다양한 후보 리스트 사용 기술, 디코더에서 움직임 정보 보정 단계 추가 기술, 학습된 변환 커널 및 필터 사용 기술 등이 있다. 한편, ECM에는 현재까지 블록분할 구조에 관한 새로운 기

술을 포함하지 않고 있으며, VVC 표준 대비 크기가 증가된 CTU(Coding Tree Unit)을 사용하고 있다. 각 기술 분야별 주요 특징에 대한 내용은 이어지는 절에서 보다 상세히 기술한다.

3. ECM의 부호화 툴

1) 개요

ECM-10.0에는 150개 이상의 기술들이 채택되어 있으며, 각 기술 분야별 채택 부호화 툴의 주요 기술적 특징은 <표 1>과 같다. 앞서 언급한대로 블록분할 구조는 VVC 표준과 거의 동일하나, 화면내 예측 및 스크린 콘텐츠 부호화 분야와 화면간 예측 분야에는 ECM에 채택된 부호화 툴의 수가 각각 67개와 56개로 가장 활발한 기술 개발이 이루어지고 있다. 변환 및 양자화 분야는 총 11개의 부호화 툴이 채택되어 있고, 인루프 필터 분야도 14개의 부호화 툴이 채택되어 있다. 그리고, 엔트로피 부호화 분야에는 전체 9개의 부호화 툴이 채택되어 있다.

<표 1> ECM 채택 기술의 주요 기술적 특징

기술 분야(채택 툴 수)	주요 기술적 특징
화면내 예측 및 스크린 콘텐츠 부호화(67)	<ul style="list-style-type: none"> 템플릿 활용 참조영역과 탐색영역의 확장 다수의 예측블록의 융합 합성곱 필터 기반 색성분 간 중복성 제거 고도화 다양한 후보 리스트 사용 스크린 콘텐츠 부호화 툴의 자연 영상 적용
화면간 예측(56)	<ul style="list-style-type: none"> 템플릿 활용 다수의 예측블록의 융합 디코더에서 움직임 정보를 보정하는 단계 추가 필터 탭 수 증가 다양한 후보 리스트 사용 샘플 단위 예측 수행
변환 및 양자화(11)	<ul style="list-style-type: none"> 변환 커널 크기 증가 비분리 변환 사용 학습(training)된 변환 커널의 적극 사용 다양한 변환 커널 추가
인루프 필터(14)	<ul style="list-style-type: none"> 학습된 필터의 적극 사용 필터 탭 수 증가 더 다양한 필터 입력 사용
엔트로피 부호화(9)	<ul style="list-style-type: none"> 템플릿의 샘플 값을 이용한 정보 유도 확률 예측 개선

2) 화면내 예측 및 스크린 콘텐츠 부호화

화면내 예측은 부호화하려는 현재 블록이 속한 영상 내에서 복원된 샘플들을 참조하여 현재 블록을 예측하는 기술이다. ECM에는 화면내 예측 기술로 기존 VVC 부호화 툴을 개선한 기술과 부호화 효율을 향상시키는 새로운 기술들이 채택되었다. 특히, 화면내 예측 분야에는 다양한 스크린 콘텐츠 부호화 기술뿐만 아니라 템플릿 활용 기술, 참조영역과 탐색영역의 확장 기술, 다수의 예측블록의 융합 기술, 합성곱 필터 기반 색성분 간 중복성 제거 고도화 기술, 다양한 후보 리스트 사용 기술 등이 포함되어 있다.

ECM의 주요 화면내 부호화 툴은 다음과 같다:

- MMLM(Multi-Model Linear Model)
- Gradient PDPC(Position Dependent Prediction Combination)
- Primary and Secondary MPM(Most Probable Mode)
- DIMD(Decoder-side Intra Mode Derivation)
- DIMD chroma mode
- IntraTMP(Intra Template Matching Prediction)
- TIMD(Template-based Intra Mode Derivation)
- Extended MRL(Multiple Reference Line) list
- TMRL(Template-based MRL)
- CCCM(Convolutional Cross-Component Model)
- GLM(Gradient Linear Model)
- CCCM using non-downsampled luma samples
- BVG-CCCM(Block-Vector Guided CCCM)
- GL-CCCM(Gradient and Location based Convolutional Cross-Component Model)
- CCCM with multiple downsampling filters
- LB-CCP(Local-Boosting Cross-Component Prediction)
- CCP(Cross-Component Prediction) merge
- SGPM(Spatial Geometry Partitioning Mode)
- Directional Planar mode

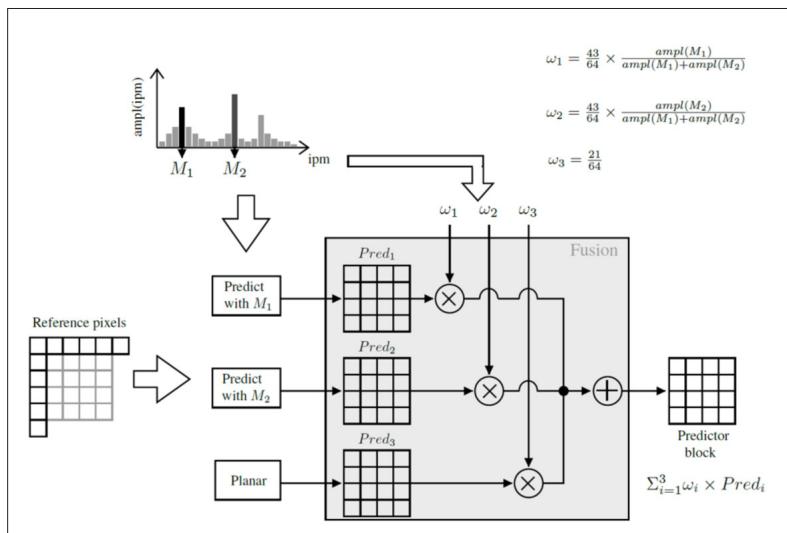
- DBV(Direct Block Vector) for chroma block
- IBC(Intra Block Copy) extended reference area
- Fractional-pel IBC
- IBC with TM(Template Matching)
- IBC-MBVD(IBC Merge mode with Block Vector Differences)
- IBC with GPM
- Combined intra block copy and intra prediction
- IBC with LIC(Local Illumination Compensation)
- Filtered IBC prediction
- RR-IBC(Reconstruction-Reordered IBC)

VVC 표준의 CCLM은 휘도성분과 색차성분의 중복성을 이용하는 예측 모드로 현재 블록의 복원된 휘도성분과 색차성분 간의 상관성을 선형 모델로 구하고 이 선형 모델을 통해 색차성분을 예측한다. ECM의 MMLM은 CCLM을 확장하여 복원된 이웃 샘플들의 평균값을 임계 값으로 사용하여 두 가지 클래스로 분류하고, 각 클래스에서 선형 모델을 유도하여 색차성분을 예측한다. CCCM은 합성곱(convolution) 필터를 사용하는 색차성분 예측 기술이며, 부호화 효율을 많이 향상시킴에 따라 CCCM과 관련된 많은 기술들을 파생시켰다. CCCM은 5-텝의 십자가 모양 공간적 필터와 비선형(non-linear) 항과 바이어스(bias) 항으로 구성된 7-텝 필터를 사용하여 복원된 휘도성분과 색차성분 간의 상관성을 모델링한다. 공간적 필터의 계수는 예측된 색차성분과 복원된 색차성분 간의 MSE(Mean Squared Error)를 최소화하는 방법으로 유도한다. CCLM과 유사하게, 하나의 모델을 사용하거나 다른 모델을 사용할 수 있다. CCCM이 지원하는 모드 중 하나로, GLM(Gradient Linear Model)은 휘도성분의 기울기를 모델 유도에 포함하는 기술이다. CCP 머지(merge) 모드는 화면간 예측의 머지 후보의 리스트 구성을 화면내 색차성분 예측에 적용한 기술이다. CCP 모드는 색성분 간 예측 방법인 CCLM, CCCM, GLM 모드 등을 후보로 포함할 수 있으며, 상기 모드로 부호화된 공간적 후보, 시

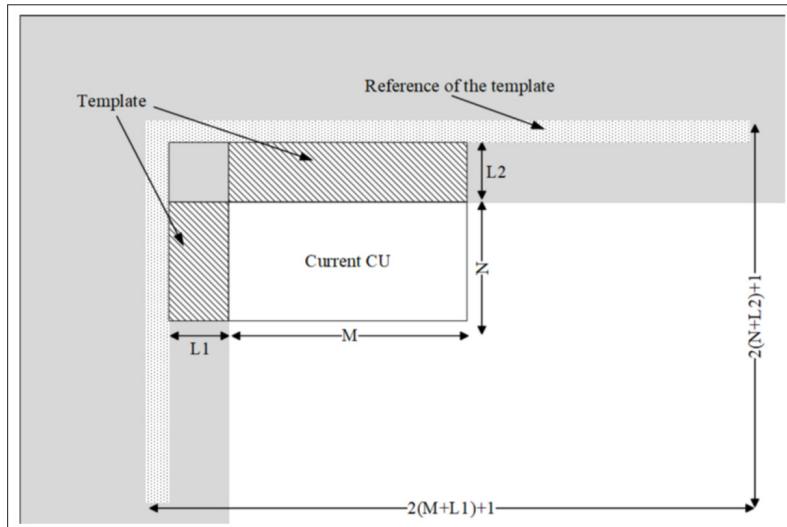
간접 후보, 비인접(non-adjacent) 공간적 후보, 히스토리(history) 후보와 기본(default) 후보 중 현재 블록의 색성분 간 예측에 사용할 최종 후보를 선택할 수 있다. CCP 메모리 모드를 사용함으로써 색차블록의 색성분 간 예측 방법을 효율적으로 시그널링(signaling)할 수 있다.

ECM은 템플릿(template)을 활용하여 부호화 효율을 향상시키는 새로운 기술들이 상당수 포함되어 있다.

DIMD와 TIMD는 템플릿을 사용해서 디코더에서 인코더와 동일한 동작을 수행해서 예측 모드를 유도함으로써 시그널링 오버헤드를 감소시킬 수 있게 하였다. DIMD와 TIMD는 템플릿을 사용하여 다수의 화면내 예측 모드를 유도할 수 있다. DIMD는 <그림 2>와 같이 현재 블록 주변의 템플릿에 소벨(Sobel) 필터를 이용한 기울기를 계산하고 기울기의 히스토그램을 구한다. 히스토그램에서



<그림 2> DIMD 개념도[11]



<그림 3> TIMD 개념도[12]

가장 높은 빈도의 다수 개의 기울기에 해당하는 화면내 예측 모드를 현재 블록의 화면내 예측 모드로 사용한다. TIMD는 MPM 리스트에 있는 화면내 예측 모드를 사용하여 템플릿 영역에서 예측 템플릿을 구하고, 템플릿과 예측 템플릿 간의 SATD(Sum of Absolute Transformed Difference) 비용이 가장 작은 한 가지 또는 두 가지의 화면내 예측 모드를 현재 블록의 화면내 예측 모드로 유도 한다. ECM 내 다른 기술에서 화면내 예측 모드 유도 시에 DIMD와 TIMD가 많이 활용된다.

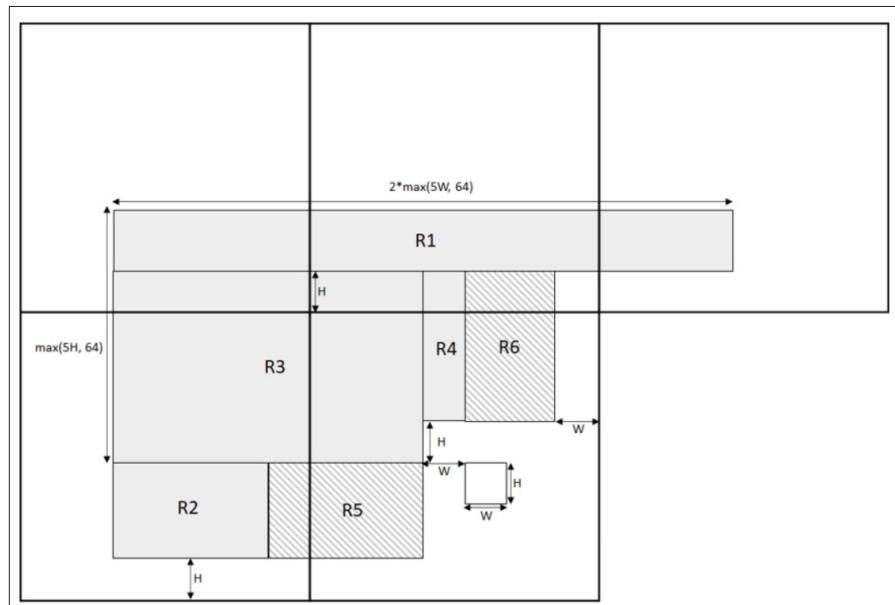
IntraTMP는 또 다른 화면내 예측 모드로 주변의 템플릿을 이용해 미리 설정된 탐색영역 안에서 가장 작은 매칭 비용을 갖는 참조 템플릿을 찾고, 그에 대응하는 참조블록을 현재블록의 예측블록으로 사용하는 기술이다. IntraTMP는 초기에 스크린 콘텐츠 부호화를 위한 기술로 채택되었으나, 이후 스크린 콘텐츠 뿐만 아닌 자연 영상에도 부호화 효율이 향상되는 기술이 적용되었고, 자연 영상에서 부호화 효율을 높이기 위한 다양한 기술들이 채택되었다.

DBV는 듀얼 트리(dual tree) 구조에서 적용되는 기술

로, 색차성분 블록에 대응하는 위치에 있는 복원된 휘도 성분 블록 내 5개 위치에 해당하는 블록이 IntraTMP 또는 IBC 모드로 부호화된 경우, 해당 블록의 블록 벡터(block vector)를 스케일링하여 색차블록의 블록 벡터로 사용한다.

SGPM은 VVC 표준의 화면간 예측에 사용되는 GPM을 화면내 예측을 위한 기술로 확장한 것이다. 현재 블록의 분할 각도는 26가지이며 각 분할된 블록에 3가지의 화면내 예측 모드 중 하나를 사용할 수 있다. 분할 각도와 화면내 예측 모드 조합으로 후보를 구성할 수 있으며, 최종적으로 선택된 후보의 인덱스가 시그널링된다.

ECM에서는 휘도블록과 색차블록의 예측에 각각 가중 융합(fusion) 예측 모드를 지원한다. 휘도성분의 융합 모드는 다음 세 가지로 나뉜다. 방향성 예측 모드가 사용된 경우, 보간 필터를 적용하는 방향성 화면내 예측 모드에 대해서 각 모드를 0, 1번째 참조라인에 적용하여 3:1의 가중치로 융합한다. TIMD의 경우, 2가지의 화면내 예측 모드를 유도할 때 2개의 모드를 각 모드의 SATD 비용의 비율에 따라 가중 융합한다. DIMD의 경우, DIMD를 통



<그림 4> IntraTMP의 참조영역[13]

해 5개의 화면내 예측 모드를 유도하고 비방향성 모드인 Planar 모드와 가중 융합을 하여 최종 예측블록으로 생성한다. 색차성분의 가중 융합 모드는 색성분 간 모델 사용하는 모드를 사용한 예측블록과 그 외 모드를 사용한 예측블록을 가중 융합하여 최종 예측블록을 생성한다.

VVC의 MRL은 0, 1, 2번째 참조샘플 라인을 이용해 화면내 예측을 수행하는 반면, ECM에서는 0, 1, 3, 5, 7, 12의 참조샘플 라인을 사용할 수 있도록 확장되었다. TIMD 적용 시에는 1, 3번째 라인만 추가로 사용 가능하며 TMRL이 채택된 후 TIMD가 아닌 MRL은 TMRL을 사용하도록 대체되었다. TMRL은 여러 참조샘플 라인을 사용하며, 예측 템플릿과 템플릿 간의 가장 낮은 비용을 가지는 참조샘플 라인과 화면내 예측 모드의 조합을 가지는 후보를 리스트 내 인덱스로 시그널링하는 기술이다. TMRL은 MRL의 0번째 참조샘플 라인이 템플릿으로 대체되며, 참조 가능한 샘플 라인 5개와 적용 가능한 화면내 예측 모드 10개의 조합으로 총 50개의 후보를 구성한다.

스크린 콘텐츠 부호화 기술인 IBC는 탐색영역이 크게 확장되었으며, 자연 영상에도 적용 가능하게 확장되면서 부화소 단위의 블록 벡터 정확도를 지원하는 기술이 ECM에 채택되었다. IBC 또한 TM을 적용하여 블록 벡터의 정확도를 개선할 수 있고, MMVD와 유사한 MBVD(Merge with Block Vector Difference), IBC 모드에서 GPM을 지원하는 IBC-GPM, CIIP의 화면간 예측 모드를 IBC로 대체하여 사용하는 기술 등이 ECM에 채택되었다.

3) 화면간 예측

화면간 예측 분야는 화면내 예측 및 스크린 콘텐츠 부호화 분야와 함께 가장 활발한 기술 제안 및 채택된 기술 분야이다. VVC 표준에 채택된 Affine motion model, AMVP(Advanced Motion Vector Prediction), BCW(Bi-prediction with CU-level Weights), DMVR(Decoder-side MV Refinement), MMVD(Merge with MV Difference), GPM(Geometric Partitioning Mode), CIIP(Combined Inter and Intra

Prediction), BDOF(Bi-Directional Optical Flow), RPR(Reference Picture Resampling)를 개선한 기술들이 많이 채택되었다.

VVC 표준에서 GPM은 ECM에서 블록에서 분할된 2개의 파티션 각각이 화면내, 화면간 예측 모드를 적용할 수 있게 되었고, TM 또한 적용할 수 있게 확장되었으며, MMVD 모드를 지원한다. GPM의 MMVD는 9개의 크기와 8개의 방향 후보의 조합으로 구성된다. 분할된 2개의 파티션 경계에 적응적 블렌딩(blending)의 너비를 사용할 수 있게 되었고, GPM의 머지 모드는 단방향 예측이 아닌 양방향 예측을 지원할 수 있게 개선되었다. 화면내 예측과 화면간 예측을 혼합해서 사용하는 CIIP 기술은 화면내 예측 모드로 Planar 모드를 사용하였는데, 이는 ECM에서 TIMD를 통해 유도되는 하나의 화면내 예측 모드를 사용하도록 확장되었다. MP(Multi-Pass)-DMVR은 DMVR을 확장한 것으로 DMVR의 단계를 추가하여 움직임 벡터(motion vector)를 보다 세밀하게 보정하는 기술이다. 첫 번째 단계에서 17x17 탐색영역을 3x3 사각형 탐색으로 정수 화소 움직임 벡터를 보정한 후, 부화소 단위의 보정을 수행한다. 두 번째 단계로 이전 단계에서 보정된 움직임 벡터에 기반하여 16x16 서브블록(sub-block) 단위로 양측 매칭(bilateral matching)을 수행한다. 이는 17x17의 탐색영역에서 마름모 영역의 정수 화소 움직임 벡터 보정을 수행한 후, 부화소 단위 보정을 수행한다. 마지막으로 8x8 서브블록 단위의 BDOF를 수행하며, 움직임 벡터의 보정 값은 [-32, 32]의 값을 가지도록 제한한다. 화면간 예측의 머지 모드를 위한 후보 구성 방법에서 비인접 공간적 후보를 사용할 수 있게 확장되었으며, 구성 가능한 후보의 최대 개수도 증가되었다.

ECM의 주요 화면내 부호화 툴들 중 VVC 표준에 포함되지 않았던 기술은 다음과 같다:

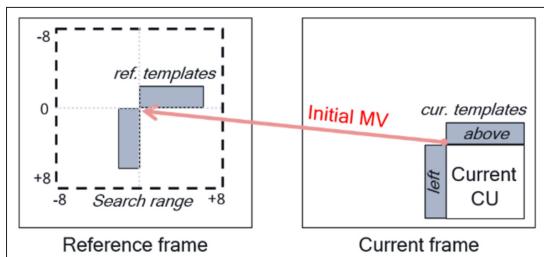
- TM(Template Matching)
- ARMC-TM(Adaptive Reordering of Merge Candidates with Template Matching)
- AMVP(Advanced Motion Vector Prediction)-

merge mode

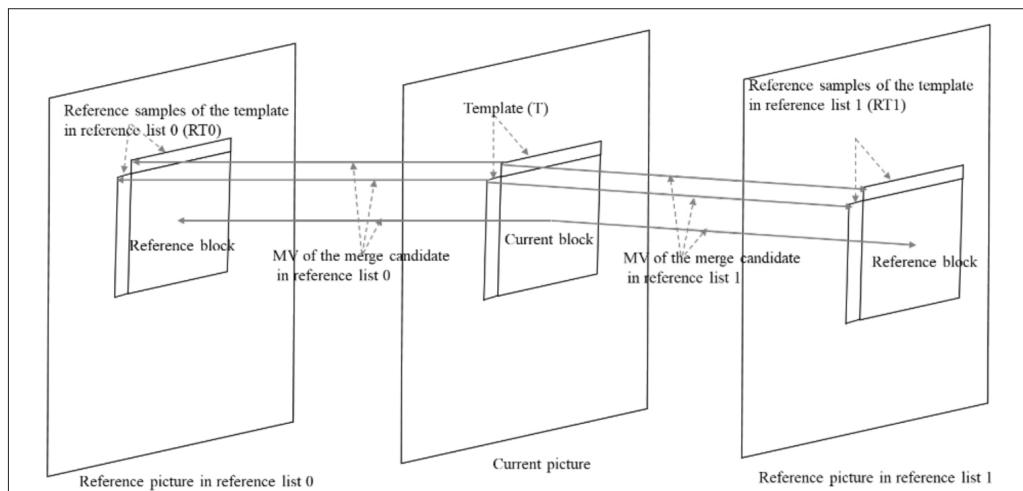
- MHP(Multi-Hypothesis Prediction)
- OBMC(Overlapped Block Motion Compensation)
- LIC(Local Illumination Compensation)
- MVD(Motion Vector Difference) prediction
- Inter CCCM
- Enhanced bi-directional motion compensation
- Block-level reference picture list reordering

ECM에 채택된 TM은 움직임 벡터에 대한 시그널링을 통해 해드 감소 및 정확도 향상을 위한 기술이다. TM은 디코더 측면의 움직임 벡터 보정 방법으로 현재 블록의 주변 영역인 템플릿과 참조영상 내 탐색영역([-8, +8]-pel)에서 참조 템플릿 간의 매칭 비용이 최소가 되는 위치로 현재 블록의 움직임 벡터를 보정하는 기술이다. AMVP 모드와 머지 모드에 적용되며, ECM 내 상당수의 기술들은 TM을 사용하여 부호화 효율을 향상시킨다. ARMC-TM은 머지 모드, TM 머지 모드, Affine 머지 모드에 적용될 수 있다. 머지 후보 리스트에 존재하는 머지 후보를 TM 비용을 기반으로 후보의 순서를 재정렬하고 재정렬된 후보 중 선택된 후보의 인덱스를 시그널링한다.

AMVP-merge 모드는 양측 매칭을 활용하여 움직임 정보의 시그널링을 감소시키는 기술이다. 단일 모드로 움직임 정보를 유도하는 VVC 표준과 달리 한쪽 방향의 움직임 정보는 AMVP 모드를 활용하고 다른 방향의 움직임 정보는 머지 모드로 활용한 후, 양측 매칭 비용을 기반으로 움직임 벡터를 보정하는 기술이다. MHP는 기존의 양예측(bi-prediction) 신호에 추가로 하나 이상의 움직임 보상 예측 신호를 사용하는 모드이며, 최종 예측의 결과는 샘플별 가중 합으로 얻을 수 있다. OBMC는 현재 블록의 상단 및 좌측 경계 샘플을 이웃 블록의 움직임 정보를 사용하여 가중 예측하여 블록 경계의 불연속성을 완화하는 기술이다. LIC는 현재 블록과 예측블록 간의 밝기 변화를 현재 블록의 템플릿과 참조블록 템플릿 간의 선형 모델로 모델링하여 예측블록의 정확도를 향상시키는 기술이다. MVD prediction은 MVD의 부호 2비트와 크기를 6비트로 표현하는 MVD 후보를 구성하고 구성된 후보는



<그림 5> Inter TM과 탐색영역[9]



<그림 6> ARMC-TM 개념도[14]

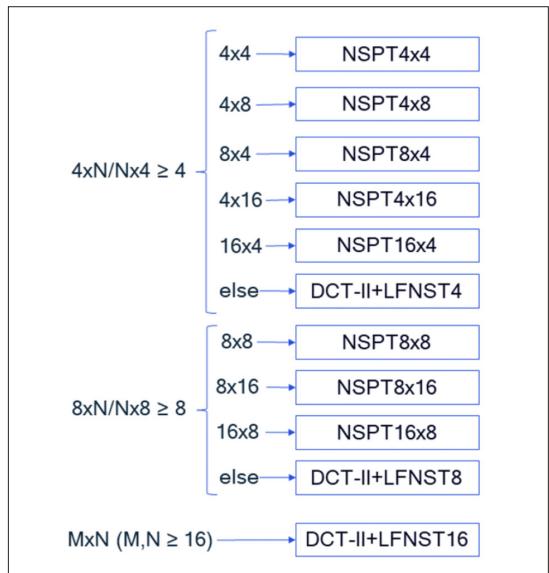
TM 비용에 따라 정렬된다. MVD의 부호와 크기를 예측 가능하게 함으로써 CABAC(Context-Adaptive Binary Arithmetic Coding)의 문맥 모델을 활용할 수 있다. 상기 MVD prediction은 AMVP, Affine AMVP, MMVD 및 Affine MMVD 모드에서 사용할 수 있다. Inter CCCM은 화면내 색차성분 예측 기술인 CCCM을 화면간 예측 또는 스크린 콘텐츠에 적용하는 기술이다. 6-텝 직사각형 모양의 공간적 필터와 비선형 항과 바이어스 항으로 구성된 8-텝 필터를 사용하여 휘도성분과 색차성분 간의 상관성을 모델링한다. 필터 계수는 가우시안 소거 방법을 통해 유도되며 복원된 휘도성분에 유도한 필터를 적용한 샘플값과 색차성분(Cb, Cr)의 샘플값을 3:1로 융합하여 최종 예측 블록으로 사용한다.

4) 변환 및 양자화

변환(transform)은 공간 영역의 잔차신호를 주파수 영역의 변환 계수로 변환함으로써 저주파 영역으로 에너지를 집중시키고, 양자화(quantization)는 고주파 신호에 둔감한 사람의 시각 특성(HVS)을 사용하여 변환 계수를 특정 값으로 나누어 주는 과정으로 변환 및 양자화는 잔차신호를 효과적으로 압축하는 방법이다. ECM에서는 VVC 표준의 1차 변환인 MTS(Multiple Transform Selection), 2차 변환인 LFNST(Low-Frequency Non-Separable Transform)의 개선 기술과 상대적으로 작은 크기의 블록에서 1차 변환과 2차 변환을 비분리(non-separable) 변환으로 대체하는 NSPT(Non-Separable Primary Transform), 8개 상태(state)를 지원하는 종속 양자화(Independent Quantization), 변환 계수의 부호 예측(Sign Prediction), 양자화 중심 값 이동 기술 등이 채택되었다.

VVC 표준의 MTS는 DST(Discrete Sine Transform)-7과 DCT(Discrete Cosine Transform)-8의 변환 커널(kernel)만을 화면내 블록 및 화면간 예측블록에서 사용하였다. ECM은 1차 변환에서 추가적으로 DCT-5, DST-4, DST-1 및 항등 변환(Identity Transform)의 변환 커널을 지원한다. MTS의 변환 세트는 블록 크기 및 화면내 예측

모드에 따라 결정되며, IntraTMP로 부호화된 블록의 변환 과정에는 DIMD를 이용해서 유도된 화면내 예측 모드를 사용하도록 확장되었다. 2차 변환인 LFNST는 LFNST 세트와 변환 커널 후보의 수가 각각 35, 3개로 확장되었다. ECM에 새롭게 채택된 NSPT는 최대 16x8의 크기의 블록에 적용 가능하며, 작은 블록에 대해서 MTS와 LFNST의 조합이 아닌 하나의 비분리 변환인 NSPT로 대체되었다. NSPT는 LFNST와 유사하게 35개의 세트와 3개의 변환 커널 후보를 지원한다.



<그림 7> 블록 크기에 따른 NSPT 수행 방법[15]

ECM은 양자화 기술로 2개의 기술이 채택되었다. 종속 양자화의 부호화 효율을 향상시키기 위해 상태의 수는 8개로 확장하였고, 양자화 인덱스에 오프셋을 더하여 0의 값으로부터 멀어지도록 변환 계수 값을 조절하는 양자화 중심 값 이동 기술이 새롭게 채택되었다.

변환 계수 부호화를 위해 ECM은 부호 예측을 채택하였다. 이 기술은 적용 가능한 변환 계수의 음수 및 양수의 부호 조합에 따라 복원된 잔차신호를 계산하고, 주변에 복원된 샘플들과 비교를 통해서 정의된 비용 함수를 최소화하

는 부호를 선택하여 부호화 효율을 향상시키는 기술이다.

5) 인루프 필터

VVC 표준의 인루프 필터(in-loop filter)는 HEVC 표준부터 존재했던 DF(Deblocking Filter), SAO(Sample Adaptive Offset)에 추가로 ALF(Adaptive Loop Filter)와 LMCS(Luma Mapping with Chroma Scaling)를 지원한다. DF는 블록 경계에서 발생하는 블록킹 열화(blocking artifact)를 개선하기 위한 기술이다. SAO는 DF가 수행된 영상에 대해 샘플 단위로 오프셋(offset)을 더해서 원본 영상과의 차이를 최소화하며, 영상의 에지(edge)에 대해 보정하는 에지 오프셋(edge offset)과 소정의 샘플 갯수 범위에 대해 보정하는 밴드 오프셋(band offset)을 지원한다. 영상의 경계 영역에서 발생할 수 있는 링잉(ringing) 열화를 개선하기 위해 SAO를 사용할 수 있다. ALF는 위너 필터(Wiener filter) 기반으로 원본과 복원 영상의 오차를 최소화하는 기술이다. LMCS는 휘도성분의 동적 범위(dynamic range)에 대한 코드워드(codeword)를 재분배하고, 매핑된 휘도성분과 색차성분 사이의 오차를 보상하는 기술이다.

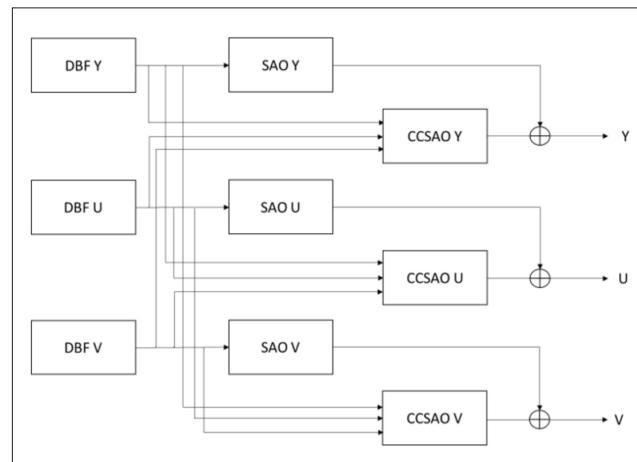
ECM의 인루프 필터는 VVC 표준의 SAO 및 ALF를 개선하는 기술들이 채택되었다. SAO의 개선 기술인 CCSAO(Cross-Component SAO)는 휘도성분과 색차성

분의 상관성을 이용하기 위해, 성분 단위로 오프셋을 결정하는 대신 휘도성분과 색차성분을 모두 이용하여 오프셋을 결정하도록 확장되었고, 히스토리 기반 오프셋과 새로운 에지 분류 방법이 추가되었다. ALF에서는 부호화 효율 향상을 위해 샘플 갯수 기반 필터 분류(classification), 필터 크기의 증가, 학습된 고정된 필터를 더 적극적으로 사용, 잔차신호 및 DF 수행 전 샘플을 필터링에 활용하는 기술 등이 채택되었다.

VVC 표준에 채택되지 못했던 BIF(Bilateral Filter)가 휘도성분과 색차성분의 부호화 효율 향상을 위해 새롭게 채택되었고, SAO와 병렬적으로 동작하도록 설계되었다.

6) 엔트로피 부호화

ECM의 엔트로피 부호화는 기존 VVC 표준의 CABAC을 기반으로 확장되었다. VVC 대비하여 2개의 확률 상태를 표현하는 정확도가 10비트에서 15비트로 확장되었고, 확장된 범위에 기반하여 LPS(Least Probable Symbol)의 범위를 개선하는 과정이 개선되었다. 슬라이스(slice) 타입을 기반으로 문맥 모델의 확률 상태를 업데이트하는 윈도우 크기가 정해졌다. 또한, 2개의 확률 상태를 이용하여 확률 상태를 예측할 때 적응적 가중치에 기반하도록 변경되었고, 이전 슬라이스의 문맥 모델로부터 현재 슬라이



<그림 8> CCSAO에서 휘도성분과 색차성분 활용

스의 문맥 모델을 초기화할 수 있는 방법이 채택되었다.

팀색모델 소프트웨어를 배포하고 있으며, 이전 ECM 대비 성능뿐만 아니라 VTM 대비 성능을 보고하고 있다[17]. 이때, 성능은 객관적 부호화 효율 척도인 BD-Rate과 구현 복잡도 중 인코딩 시간과 디코딩 시간 등을 포함하며, 공통실험조건(CTC: Common Test Condition)[18]에서 앵커(anchor)인 VTM-11.0이 비교 대상으로 사용된다.

JVET은 매 회마다 새로이 채택된 기술을 포함한 ECM

III. ECM의 성능

<표 2> ECM-10.0의 부호화 효율 및 인코딩/디코딩 시간 결과

All Intra Main 10					
	Over VTM-11.0ecm10			EncT	DecT
	Y	U	V	EncT	DecT
Class A1	-10.55%	-21.81%	-29.24%	915.8%	439.3%
Class A2	-16.68%	-28.60%	-31.54%	916.2%	458.0%
Class B	-11.02%	-26.58%	-24.82%	826.4%	455.0%
Class C	-11.07%	-16.52%	-17.06%	898.0%	452.7%
Class E	-14.59%	-24.40%	-22.89%	804.1%	467.8%
Overall	-12.49%	-23.52%	-24.63%	867.2%	454.4%
Class D	-9.17%	-14.38%	-14.35%	851.0%	445.1%
Class F	-26.21%	-35.26%	-35.17%	541.1%	452.7%
Class TGM	-39.28%	-47.12%	-46.45%	457.3%	496.2%
Random Access Main 10					
	Over VTM-11.0ecm10			EncT	DecT
	Y	U	V	EncT	DecT
Class A1	-22.25%	-27.01%	-36.07%	882.3%	784.9%
Class A2	-25.50%	-35.13%	-38.75%	791.3%	889.0%
Class B	-20.50%	-34.42%	-32.13%	673.3%	731.6%
Class C	-21.91%	-26.29%	-26.07%	708.2%	777.7%
Overall	-22.23%	-30.91%	-32.63%	744.0%	784.1%
Class D	-22.87%	-27.80%	-28.10%	738.6%	851.2%
Class F	-28.30%	-36.71%	-37.10%	602.2%	482.5%
Class TGM	-36.63%	-43.02%	-43.08%	570.5%	415.0%
Low delay B Main 10					
	Over VTM-11.0ecm10			EncT	DecT
	Y	U	V	EncT	DecT
Class A1					
Class A2					
Class B	-17.55%	-38.06%	-35.34%	589.9%	668.2%
Class C	-19.29%	-28.96%	-29.31%	648.1%	710.0%
Class E	-16.12%	-28.03%	-27.57%	642.1%	459.1%
Overall	-17.77%	-32.52%	-31.39%	621.7%	620.8%
Class D	-21.17%	-30.48%	-30.27%	656.7%	758.7%
Class F	-24.60%	-37.20%	-36.64%	588.5%	498.3%
Class TGM	-35.27%	-43.12%	-43.29%	560.3%	418.2%

<표 2>와 같이, ECM-10.0은 VTM-11.0과 비교하여, AI(All Intra), RA(Random Access) 및 LDB(Low Delay B) 환경에서 각각 -12.49%, -22.23%, -17.77%의 BD-Rate 결과를 보인다. 또한, 인코딩 시간(EncT)/디코딩 시간(DecT) 측면에서 AI, RA 및 LDB 환경에서 각각 8.6배/4.5배, 7.4배/7.8배, 6.2배/6.2배가 소요됨을 알 수 있다.

ECM에는 VVC 표준과 비교하여 구현 복잡도가 높은 부호화 툴들을 상당수 포함하고 있다. 이러한 구현 복잡도에 따라 부호화 툴들을 4개의 그룹으로 선별하여 ECM 대비 그룹 단위 부호화 툴 오프(Tool off) 실험을 진행하고 있다[19]. 4개의 그룹은 1) 화면간 예측의 TM 관련 기술, 2) Merge, AMVP, IBC 등의 리스트 유도와 화면내 예측/복원이 인터리빙(interleaving)되는 기술, 3) IntraTMP 및 IBC의 TM 관련 기술, 4) VVC보다 주변 복원샘플에 대한 프로세스가 더 많이 요구되는 기술로 나뉜다. 즉, 4번 그룹에서 1번 그룹으로 갈수록 ECM을 하드웨어로 구현하는데 더 많은 어려움이 존재한다. <표 3>은 ECM-10.0에서 각 그룹 단위의 부호화 툴 오프 결과를 보여준다. 특히, 1번 그룹에서 4번 그룹에 속하는 부호화 툴들을 모두 사용하지 않는 경우(-16.17%), 부호화 툴을 모두 사용하는 경우(-22.23%) 대비 RA 환경에서 약 6%의 부호화 효율 손실이 발생하는 것을 알 수 있다.

IV. 결론

본고에서는 현재 JVET에서 진행하고 있는 새로운 비디오 부호화 표준을 위한 기술 탐색실험에 사용되는 ECM에 채택된 기술들을 소개하였다. ECM에는 템플릿 활용 기술, 참조영역과 탐색영역의 확장, 다수의 예측블록의 융합, 합성곱 필터 기반 색성분 간 중복성 제거 고도화, 다양한 후보 리스트 사용, 스크린 콘텐츠 부호화 툴의 자연 영상 적용, 디코더에서 움직임 정보를 보정하는 단계 추가, 필터 템수 증가, 샘플 단위 예측 수행 기술, 변환 커널 크기 증가, 비분리 변환 사용, 학습된 변환 커널의 적극 사용, 다양한 변환 커널 추가, 학습된 필터의 적극 사용, 더 다양한 필터 입력 사용, 템플릿의 샘플 값을 이용한 정보 유도 및 엔트로피 부호화를 위한 확률 예측 개선 등 다양한 기술적 특징이 반영되어, 최신 비디오 부호화 국제 표준인 VVC 대비 높은 부호화 효율 향상과 더불어 높은 구현 복잡도를 보이고 있다. 향후 ECM의 압축률이 더 향상되면 JVET에서 새로운 비디오 부호화 표준을 위한 본격적인 논의가 진행될 것으로 예상되며, ECM에 채택된 기술들의 상당수가 새로운 비디오 부호화 표준에도 포함될 것으로 전망된다.

<표 3> ECM-10.0의 그룹 단위 부호화 툴 오프 결과

Group		BD-Rate		
		Y	U	V
1	AI	-12.50%	-23.53%	-24.63%
	RA	-19.12%	-28.05%	-29.76%
2	AI	-12.06%	-22.96%	-24.06%
	RA	-20.90%	-30.21%	-31.85%
3	AI	-10.89%	-21.12%	-22.06%
	RA	-21.48%	-29.24%	-30.60%
4	AI	-11.34%	-19.68%	-21.04%
	RA	-21.66%	-28.44%	-30.14%
1~4	AI	-8.96%	-16.42%	-17.59%
	RA	-16.17%	-22.89%	-23.89%

참 고 문 헌

- [1] ISO/IEC 23090-3:2021, Information technology - Coded representation of immersive media - Part 3: Versatile Video Coding, Feb, 2021.
- [2] 최기호, “Post VVC 표준을 위한 ECM”, 한국통신학회지(정보와 통신), 40권 9호, 2023년 9월.
- [3] A. Alshina, S. Liu, J. Pfaff, M. Wien, P. Wu, and Y. Ye, “JVET AHG report: Neural-network-based video coding (AHG11),” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-T0011, Oct, 2020.
- [4] M. Karczewicz, Y. Ye, B. Bross, X. Li, K. Naser, and H. Yang, “JVET AHG report: Enhanced compression beyond VVC capability (AHG12),” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-V0012, Apr, 2021.
- [5] E. Alshina, S. Liu, W. Chen, Y. Li, R.-L. Liao, Z. Ma, and H. Wang, “Description of Exploration Experiments on NN-based video coding,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-T2023, Oct, 2020.
- [6] V. Seregin, J. Chen, S. Esenlik, F. Le Léannec, L. Li, M. Winken, J. Ström, X. Xiu, and K. Zhang, “Exploration Experiment on Enhanced Compression beyond VVC capability,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-U2024, Jan, 2021.
- [7] M. Coban, R.-L. Liao, K. Naser, J. Ström, and L. Zhang, “Algorithm description of Enhanced compression Model 10,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-AE2025, Oct, 2023.
- [8] <https://vcgit.hhi.fraunhofer.de/ecm/ECM/-/tree/ECM-10.0>
- [9] 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. Shlyakhow, L. Pham Van, H. Wang, Y. Zhang, and Z. Zhang, “Compression efficiency methods beyond VVC,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-U0100, Jan, 2021.
- [10] 임성창, “ECM Overview”, MPEG뉴미디어포럼 추계심층기술워크숍, 2023년 11월.
- [11] M. Abdoli, T. Guionnet, E. Mora, M. Raulet, S. Blasi, A. Seixas Dias, and G. Kulupana, “Non-CE3: Decoder-side Intra Mode Derivation (DIMD) with prediction fusion using Planar,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-O0449, Jul, 2019.
- [12] K. Cao, N. Hu, V. Seregin, M. Karczewicz, Y. Wang, K. Zhang, and L. Zhang, “EE2-related: Fusion for template-based intra mode derivation,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-W0123, Jul, 2021.
- [13] Y. Yu, L. Zhang, F. Wang, J. Gan, H. Yu, L. Xu, Z. Xie, D. Wang, Y. Ma, H. Zhang, J. Huo, F. Yang, X. Gao, N. Wang, X. Xiu, N. Yan, C. Ma, H. Jhu, C. Kuo, W. Chen, and X. Wang, “EE2-2,9: Extended search areas for IntraTMP mode,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-AE0077, Jul, 2023.
- [14] N. Zhang, K. Zhang, L. Zhang, H. Liu, Z. Deng, and Y. Wang, “AHG12: Adaptive Reordering of Merge Candidates with Template Matching,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-V0099, Apr, 2021.
- [15] P. Garus, M. Coban, B. Ray, V. Seregin, M. Karczewicz, M. Koo, J. Zhao, J. Lim, and S. Kim, “EE2-4,2: Non-Separable Primary Transform for Intra Coding,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-AC0130, Jan, 2023.
- [16] C.-W. Kuo, X. Xiu, Y.-W. Chen, H.-J. Jhu, W. Chen, and X. Wang, “AHG12: Cross-component Sample Adaptive Offset,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-V0153, Apr, 2021.
- [17] V. Seregin, J. Chen, R. Chernyak, F. Le Léannec, and K. Zhang, “JVET AHG report: ECM software development (AHG6),” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-AF0006, Oct, 2023.
- [18] M. Karczewicz and Y. Ye, “Common Test Conditions and evaluation procedures for enhanced compression tool testing,” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-AE2017, Jul, 2023.
- [19] X. Li, L.-F. Chen, Z. Deng, J. Gan, E. François, H.-J. Jhu, X. Li, and H. Wang, “JVET AHG report: ECM tool assessment (AHG7),” Joint Video Experts Team of ITU-T and ISO/IEC, JVET-AF0007, Oct, 2023.

저자 소개

전상호



- 2023년 2월 : 한국항공대학교 정보통신학과 학사
- 2023년 3월 ~ 현재 : 한국항공대학교 항공전자정보공학과 석사과정
- ORCID : <https://orcid.org/0009-0009-7679-5026>
- 주관심분야 : 비디오 신호처리, 비디오 부호화

임성창



- 2006년 2월 : 세종대학교 컴퓨터공학과 학사
- 2008년 2월 : 세종대학교 컴퓨터공학과 석사
- 2022년 2월 : 세종대학교 컴퓨터공학과 박사
- 2008년 3월 ~ 현재 : 한국전자통신연구원 미디어부호화연구실 책임연구원
- ORCID : <https://orcid.org/0000-0002-0110-7489>
- 주관심분야 : 영상 압축, 영상 처리

강정원



- 1993년 : 한국항공대학교 항공전자공학과 학사
- 1995년 : 한국항공대학교 항공전자공학과 공학석사
- 2003년 : Georgia Institute of Tech, ECE 공학박사
- 2003년 ~ 현재 : 한국전자통신연구원 미디어부호화연구실 실장
- ORCID : <https://orcid.org/0000-0003-4003-4638>
- 주관심분야 : 2D 영상 압축, 입체 비디오 영상 압축, 영상신호처리

김재곤



- 1990년 : 경북대학교 전자공학과 학사
- 1992년 : KAIST 전기 및 전자공학과 석사
- 2005년 : KAIST 전자전산학과 박사
- 1992년 ~ 2007년 : 한국전자통신연구원(ETRI) 선임연구원/팀장
- 2001년 ~ 2002년 : Columbia University, NY, 연구원
- 2015년 ~ 2016년 : UC San Diego, Visiting Scholar
- 2007년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
- ORCID : <http://orcid.org/0000-0003-3686-4786>
- 주관심분야 : 비디오 신호처리, 비디오 부호화, 이미지브 비디오