

ECM 화면내 예측 기술 I: 휘도 화면 내 예측

□ 정현기, 김용성, 전병우 / 성균관대학교

요약

최근 영상 데이터 트래픽의 급격한 증가로 인해, 보다 높은 압축 효율을 제공하는 영상 부호화 기술에 대한 요구가 지속적으로 증가하고 있다. ISO/IEC MPEG과 ITU-T VCEG가 공동으로 구성한 Joint Video Experts Team (JVET)은 2020년 Versatile Video Coding (VVC) 표준을 완료한 이후, VVC를 넘어서는 추가적인 압축 성능 향상을 목표로 Enhanced Compression Model (ECM)을 만들어 가며 다양하고 새로운 부호화 기술 탐색을 진행하고 있다. 특히 화면 내 예측은 부호화 성능에 직접적인 영향을 미치는 핵심 요소로서, ECM에는 휘도 채널을 대상으로 한 다수의 신규 화면 내 예측 기술들이 반영되어 왔다. 본 글에서는 ECM에 구현된 다양한 휘도 채널 화면 내 예측 기술들을 중심으로, 해당 기술들의 기본 동작 원리를 체계적으로 정리하고 분석한다. 이를 위하여 ECM의 휘도 채널 화면 내 예측 기술들을 예측자 생성 방식에 따라 분류하고, 각 접근 방식이 가지는 특징을 분석한다. 이러한 분석을 통하여 VVC 이후 화면 내 예측 기술의 설계 흐름과 향후 Beyond VVC 단계에서 고려되어야 할 주요 방향을 제시하고자 한다.

I. 서 론

최근 영상 데이터 트래픽의 급격한 증가로 인해, 영상 데이터의 효율적인 전송 및 저장을 위한 고성능 비디오 압축 기술에 대한 요구가 지속적으로 증가하고 있다. 이러한 요구에 대응하기 위하여 ISO/IEC MPEG과 ITU-T VCEG가 공동으로 구성한 Joint Video Experts Team (JVET)은 2020년에 차세대 비디오 압축 표준인 Versatile Video Coding (VVC)을 완성하였다[1]. VVC는 기존 표준인 HEVC 대비 All intra (AI) 구성에서 Y, Cb 및 Cr 채널

에 대해 각각 -25.06%, -25.37%, -26.85%의 Bjøntegaard delta bitrate (BDBR) 성능을 달성하여, 다양한 응용 환경에서의 고효율 영상 압축 개선을 달성하였다[2].

VVC의 표준화 이후, JVET은 추가적인 압축 성능 향상을 목표로 VVC의 한계를 넘어서는 새로운 부호화 기술들의 개발을 위하여 beyond VVC capability라는 이름으로 새로운 기술들에 대한 다양한 탐색 연구개발을 진행하고 있다. 이 과정에서 다양한 기술들의 평가를 위한 참조 소프트웨어인 Enhanced Compression Model (ECM)을 만들고 있으며, ECM을 기반으로 다수의 신규 기술들이 성

능 및 복잡도 관점에서 비교되고 있다[3]. ECM-18.0 기준으로, All Intra (AI) 구성에서 VTM 대비 Y, Cb 및 Cr 채널에 대해 각각 -16.87%, -19.90%, -21.83%의 BD-BR 성능이 보고되었으며, 이는 VVC 표준화 이후에도 화면 내 예측을 포함한 새롭게 적용된 기술들을 통해 추가적인 압축 성능 향상을 가능함을 보여준다[4].

ECM에서는 특히 화면 내 예측 성능 향상을 위한 다양한 기술들이 제안되었으며, 디코더 측에서 예측자를 유도하거나 복수의 예측자를 결합하는 방식 등 기존 VVC의 인트라 예측 구조를 확장하는 방향으로 기술이 발전하고 있다. 이러한 인트라 예측 기술들은 압축 효율 측면에서 의미 있는 성능 향상을 제공하는 반면, 계산 복잡도가 증가하고 있다.

본 글에서는 ECM에 채택된 휘도 채널에서의 화면 내 예측 관련 기술들을 중심으로, 각 기술의 기본 동작 원리와 설계 특징을 정리하고 분석한다. 이를 통해 VVC 이후 인트라 예측 기술의 발전 방향과 주요 설계 흐름을 체계적으로 분석하고자 한다.

II. Beyond VVC에서의 화면 내 부호화 방법

VVC 표준화 이후, JVET은 추가적인 압축 성능 향상을 목표로 Enhanced Compression Model (ECM)을 중심으로 다양한 화면 내 예측 기술들을 채택하고 있다. ECM에 채택된 화면 내 예측 기술들은 기존 VVC에서의 화면 내 예측 기술을 단순히 확장하는 데 그치지 않고, 새로운 화면 내 예측 기술에 기반한 기법들 또한 함께 포함하고 있다. 본 절에서는 ECM-19.0 버전을 기준으로, 현재까지 채택된 휘도 채널 화면 내 예측 기술들을 중심으로 설명한다.

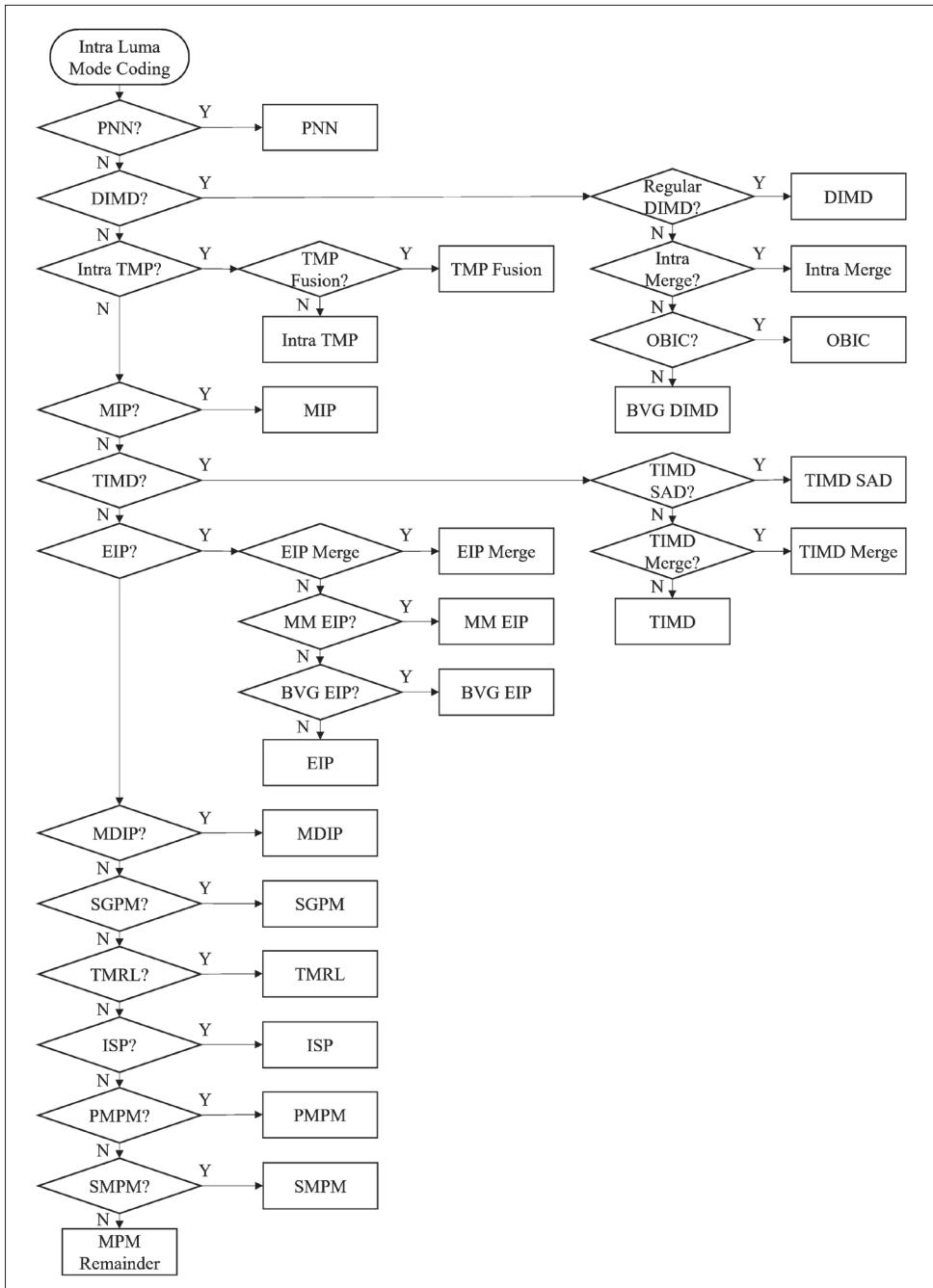
ECM 기반 화면 내 예측 기술들은 크게 세 가지 설계 흐름으로 분류할 수 있다. 첫 번째는 기존 VVC 인트라 예측 도구를 보다 정밀하게 개선하는 방식이다. 이는 참조 샘

플 범위의 확장이나 후보 리스트 크기의 조정을 통해 보다 정확한 예측자를 생성하는 기술들이다. 두 번째는 복호화기에서 예측자를 직접 유도하는 방식이다. 이는 현재 블록 예측에 사용되는 인트라 예측 모드에 대한 인덱스 신호를 줄이거나 제거함으로써 신호화 (Signaling) 오버헤드를 감소시키는 기술들이다. 이 경우, 예측 모드에 대한 명시적인 인덱스 신호 없이 복호화기가 주변 복원 샘플을 기반으로 동일한 예측자를 유도할 수 있도록 설계된다. 마지막으로, 사전 학습된 모델을 활용하여 예측자를 생성하는 방식이 있다.

이러한 기술개발 흐름에 따라, ECM에 채택된 휘도 채널 기반 화면 내 예측 기술들은 예측자를 생성하는 방식에 따라 구분될 수 있다. DIMD[5], OBIC[6], TIMD[7]와 같은 복호화기 기반 예측 모드 방법들은 현재 블록 주변의 복원된 샘플로부터 방향성 정보 또는 통계적 지표를 추출하고, 이를 기반으로 복수의 방향성 인트라 예측자를 생성한 후 가중 결합을 통해 현재 블록에 대한 최종 예측자를 생성한다. 이들 기법은 예측에 필요한 추가적인 정보가 비트스트림으로 신호되지 않으며, 부호화기와 복호화기가 동일한 절차를 수행함으로써 예측자가 재현되도록 설계되어 있다.

한편, MPDIP[8] 및 IntraNN[9] 모드는 사전 학습된 모델을 이용하여 현재 블록의 예측자를 생성한다는 점에서 기존의 방향성 기반 예측과 구별된다. 이러한 예측 모드에서는 복원된 주변 참조 샘플을 입력으로 하여, 사전 학습된 행렬 가중치 또는 신경망 모델을 통해 현재 블록에 대한 예측자를 생성한다. 이때 사용되는 모델 정보는 부호화기와 복호화기에 사전 정의되어 있으며, 추가적인 모델 파라미터에 대한 신호 없이 예측이 수행된다.

〈그림 1〉은 ECM-19.0 기준에서 휘도 채널 화면 내 예측 모드들이 어떠한 신호 구조를 통해 부호화되는지를 개략적으로 나타낸 것이다. 각 화면 내 예측 모드는 예측자 생성 방식에 따라 서로 다른 신호 구조를 가지며, 일부 모드는 인트라 예측 모드 인덱스를 명시적으로 신호하는 반면, 복호화기 유도형 예측 모드의 경우 예측 모드 인덱스 없이도 복호화기에서 예측자를 재구성할 수 있다.



<그림 1> ECM-19.0 기준 휘도 채널의 화면 내 예측 모드

이와 같이 ECM의 휘도 채널 화면 내 예측 모드들은 복호화기에서 예측자를 생성하는 방식, 인트라 예측에 사용되는 후보 리스트를 구성하는 방식, 그리고 사전 학습된

모델을 이용하여 예측자를 생성하는 방식으로 구분될 수 있으며, 각 방식은 예측 정확도, 신호화 오버헤드, 그리고 계산 복잡도 측면에서 상이한 특성을 가진다.

1. Decoder-side Intra Mode Derivation (DIMD) 기술

DIMD 기술은 복호화기에서 현재 블록 주변의 복원 샘플을 분석하여 인트라 예측자를 직접 유도하는 인트라 예측 기술로, 인트라 예측 모드에 대한 명시적인 시그널링 없이도 효율적인 예측을 가능하게 한다[5]. DIMD는 Sobel 필터 기반의 기울기 분석을 통해 IPM (Intra Prediction Mode)별 통계 정보를 생성하고, 해당 통계에서 두드러진 방향에 대응되는 다수의 예측자와 planar 예측자를 가중 결합하여 최종 예측자를 생성한다. DIMD의 전체 동작 과정은 다음과 같이 세 단계로 구성된다.

단계 1. 템플릿 영역 구성 (주변 참조 샘플 수집)

단계 1에서는 현재 블록의 상단 및 좌측 인접 복원 샘플을 이용하여 템플릿 (template) 영역을 구성한다. 해당 템플릿 영역은 Sobel 필터를 적용하기 위한 영역으로 정의되며, 현재 블록 주변의 공간적 구조와 방향성 정보를 반영하기 위한 입력 데이터로 사용된다. 이 과정에서 기본적으로 상단 및 좌측 인접 참조 영역이 사용되며, 우상단 및 좌하단 참조 영역은 복원된 샘플이 존재하는 경우에 한해 추가적으로 활용된다. 참조 샘플의 가용 여부에 따라 HoG 구성에 포함되는 영역이 결정되며, 이를 통해 주변 참조 정보의 유효한 범위 내에서 방향성 통계가 안정적으로 구성되도록 한다. <그림 2>에서 현재 블록 주변 템

플릿 영역은 HoG 구성을 위한 템플릿 영역을 의미한다.

단계 2. Sobel 기반 기울기 계산 및 HoG 구성

구성된 템플릿 영역에 대해 수평/수직 방향 Sobel 필터를 적용하여 G_{Hor} 및 G_{Ver} 기울기 성분을 계산한다. 이후 각 IPM 방향에 대한 기여도를 누적하여 HoG를 구성하며, 기울기 크기(또는 amplitude)는 연산 복잡도를 고려하여 다음과 같이 정의할 수 있다.

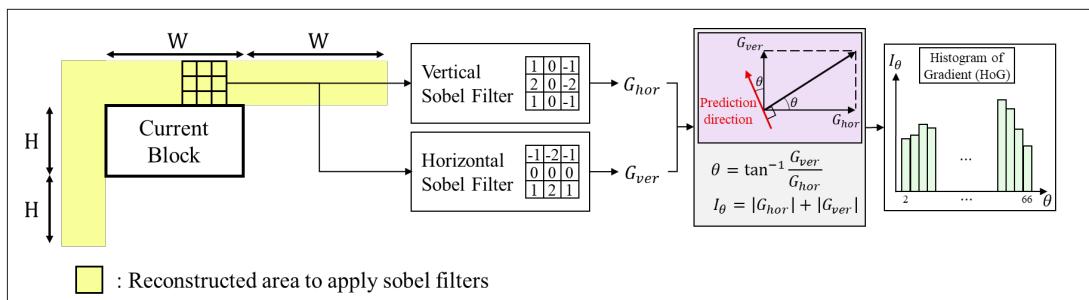
$$HoG \text{ amplitude} = |G_{Hor}| + |G_{Ver}|$$

HoG가 구성되면, IPM별 amplitude 값을 기준으로 예측 후보 모드들을 내림차순 정렬한다. <그림 2>는 인접 템플릿을 사용하여 HoG를 구성하는 과정을 도식화한 것이다.

단계 3. 상위 IPM 예측자 및 planar 예측자의 가중 결합

단계 3에서는 단계 2에서 구성된 HoG의 amplitude가 가장 큰 상위 다섯 개의 IPM으로부터 생성된 예측자 ($Pred_i, i = 1, \dots, 5$)와 planar 모드 또는 BV 정보를 사용하여 생성된 예측자 ($Pred_{planar/BV}$)를 가중합하여 최종 예측자를 생성한다. 최종 예측자 ($Pred_{DIMD}$)는 아래식 (1)과 같이 표현된다.

$$Pred_{DIMD} = \sum_{i=1}^5 w_i \cdot Pred_i + w_{planar/BV} \cdot Pred_{planar/BV} \quad (1)$$



<그림 2> DIMD 모드에서의 HoG를 구성하는 과정

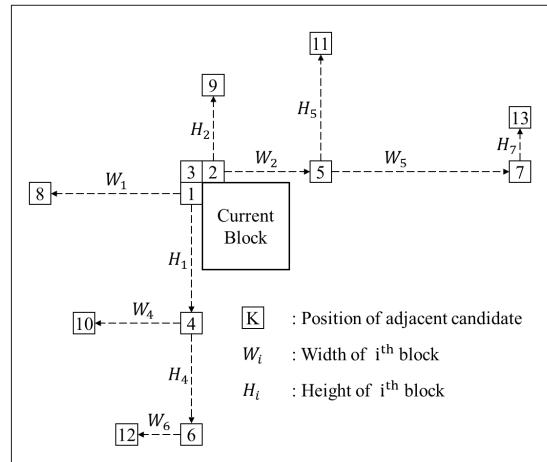
이때 각 방향성 예측자에 대한 가중치 w_i 는 해당 방향성 모드의 HoG amplitude에 비례하도록 설정하여, 전체 가중치 합이 48/64가 되도록 한다. 반면, $Pred_{planar/BV}$ 에 대한 가중치 $w_{planar/BV}$ 는 16/64로 설정한다. 여기서 $Pred_{planar/BV}$ 는 planar 모드와 BV 정보를 이용하여 각각 생성된 두 예측자 중 하나로 결정된다. 구체적으로, 단계 1에서 정의한 reference template 영역을 대상으로 각 예측자를 생성한 후, reference template 내의 복원된 샘플과의 SATD (sum of absolute transform difference) 비용을 계산한다. 두 예측자에 대한 SATD 비용을 비교하여, 더 낮은 비용을 갖는 예측자를 $Pred_{planar/BV}$ 로 선택한다. DIMD 모드는 VTM-11.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.45%, -0.26%, -0.30%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 113%와 110%로 측정되었다.

2. Occurrence-based Intra Coding (OBIC) 기술

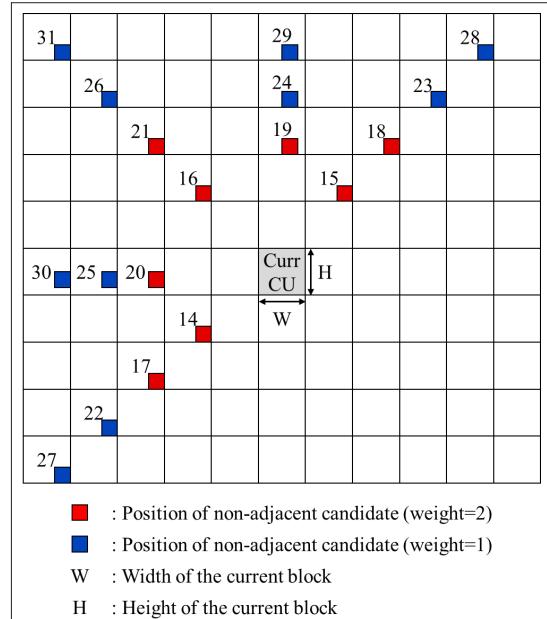
OBIC 기술은 현재 블록 주변에 위치한 블록들에서 예측에 사용된 인트라 예측 모드들의 발생 빈도를 Histogram of Occurrence (HoC) 형태로 분석하여, 복수 개의 방향성 인트라 예측자를 유도하고 이를 가중 결합함으로써 최종 예측자를 생성하는 화면 내 예측 기술이다[6]. OBIC의 전체 동작 과정은 다음과 같은 단계로 구성된다.

단계 1. 인접 및 비인접 후보 블록 수집

단계 1에서는 현재 블록 주변의 인접 (adjacent) 후보와 비인접 (non-adjacent) 후보로 OBIC 후보 리스트를 구성한다. 현재 블록에 대한 인접 후보 블록의 위치는 <그림 3>에 도시된 바와 같이 총 13개이며, 비인접 후보 블록은 <그림 4>에 도시된 바와 같이 총 18개의 위치에서 획득된다. 이를 통하여 최대 31개의 후보 블록으로 OBIC 후보 리스트가 구성된다.



<그림 3> OBIC 모드에서의 인접 후보 위치



<그림 4> OBIC 후보에서의 비인접 후보 위치와 비인접 후보 위치에 따른 weight 값

단계 2. 후보 정렬 및 HoC 계산

단계 2에서는 단계 1에서 수집된 총 31개의 OBIC 후보 블록에 대하여, 현재 블록과의 L1 distance를 계산하고 이를 기준으로 후보 블록들을 정렬한다. 정렬된 후보들 중 상위 20개의 후보 블록을 선정하여 HoC 계산에 사용한다. HoC는 IPM (Intra Prediction Mode)별 발생 빈도를

누적한 통계로 정의되며, 각 후보 블록이 HoC에 누적 기여하는 값은 해당 블록의 $width * height * weight$ 값이다. 이 때, $weight$ 는 후보 블록의 위치에 따라 기 설정된 값을 사용한다. 구체적으로, 인접 후보의 경우 $weight$ 를 4로, 비인접 후보의 경우 $weight$ 를 1 또는 2로 설정한다. <그림 4>는 비인접 후보 위치에 따른 $weight$ 값을 도식화한 것이다.

단계 3. 예측자 생성

단계 3에서는 단계 2에서 계산된 HoC를 기반으로 최종 예측자를 생성한다. 먼저 HoC 값이 가장 큰 상위 5 개의 IPM을 선택하고, 선택된 각 IPM에 대해 방향성 인트라 예측자 $Pred_i$ ($i = 1, \dots, 5$)를 생성한다. 이때 planar 모드는 HoC 통계 구성 및 상위 IPM 선택 과정에는 포함되지 않는다. 이후, 생성된 방향성 예측자들과 planar 모드 또는 BV 정보를 활용하여 생성한 예측자 $Pred_{planar/BV}$ 를 가중 결합하여 최종 예측자를 생성한다. Planar 모드와 BV 기반 예측자에 대해서는 현재 블록의 reference template 영역을 대상으로 각각 예측자를 생성한 후, reference template 내 복원된 샘플과의 SATD (sum of absolute transformed differences) 비용을 계산한다. 두 예측자 중 더 낮은 SATD 비용을 갖는 예측자가 $Pred_{planar/BV}$ 로 선택된다. 최종 예측자는 다음과 같이 표현된다.

$$Pred_{OBIC} = \sum_{i=1}^5 w_i * Pred_i + w_{planar/BV} * Pred_{planar/BV} \quad (2)$$

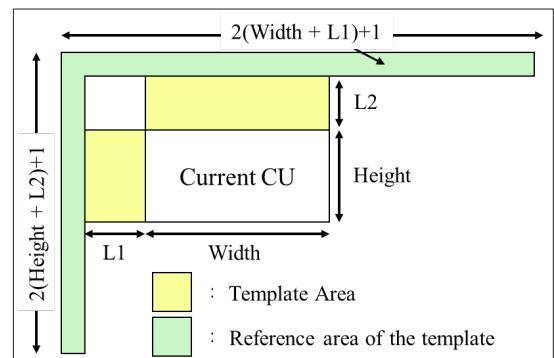
여기서 $w_{planar/BV}$ 는 16/64의 고정 가중치를 사용하며, 방향성 예측자에 대한 가중치 w_i 는 HoC 크기에 비례하도록 설정한 후 전체 합이 48/64가 되도록 한다. OBIC 모드는 ECM-12.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.10%, -0.04%, -0.04%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 100.5%와 99.2%로 측정되었다.

3. Template-based Intra Mode Derivation (TIMD) 기술

TIMD 모드는 복호화기에서 현재 블록 주변의 복원된 샘플을 이용하여 현재 블록 예측에 사용할 IPM을 결정하는 화면 내 예측 기술이다[7]. TIMD 모드는 IPM에 대한 명시적인 시그널링 없이, 템플릿 영역에서의 템플릿 비용을 기준으로 최적의 인트라 예측 모드를 선정함으로써 화면 내 예측을 수행한다. TIMD의 전체 동작 과정은 다음과 같이 구성된다.

단계 1. 템플릿 영역 구성

TIMD 모드에서는 <그림 5>와 같이 현재 블록의 상단 및 좌측에 위치한 복원된 샘플을 이용하여 템플릿 영역 (template area)과 템플릿의 참조 영역 (reference area of template)을 구성한다. 템플릿의 두께는 현재 블록의 너비 또는 높이가 8보다 큰 경우에는 좌측 또는 상단 템플릿의 두께를 4로 설정하고, 그렇지 않은 경우에는 템플릿의 두께를 2로 설정한다.



<그림 5> TIMD 모드의 대표도

단계 2. 템플릿 기반 SATD 계산 및 최적 모드 선정

단계 2에서는 MPM 리스트에 포함된 IPM들을 대상으로, 템플릿 참조 영역을 기반으로 템플릿 영역에 대한 예

측자를 생성한다. 이후 생성된 예측자와 템플릿 영역의 복원 샘플 간의 SATD 비용을 계산한다. 계산된 SATD 비용을 기준으로 가장 작은 비용을 갖는 상위 하나 또는 두 개의 IPM을 선정한다. 구체적으로, 두 번째로 작은 SATD 비용이 가장 작은 SATD 비용보다 두 배보다 크지 않은 경우에는 두 개의 IPM을 선정하며, 그 외의 경우에는 하나의 IPM만을 선정한다.

단계 3. 예측자 생성 및 가중 결합

최종 예측자는 다음과 같이 생성된다. 템플릿 영역에서의 SATD 비용을 기준으로 하나의 IPM만 선정된 경우, 해당 IPM으로 생성된 예측자를 최종 예측자로 사용한다. 반면, 두 개의 IPM이 선정된 경우, 각 IPM으로부터 생성된 현재 블록의 예측자를 SATD 비용에 기반한 가중치로 가중합하여 최종 예측자를 생성한다. 두 예측자를 가중합하는 경우의 최종 예측자는 식 (3)과 같이 표현된다.

여기서 $Pred_{IPM1}$ 과 $Pred_{IPM2}$ 는 각각 템플릿 SATD 비용이 가장 작은 $IPM1$ 과 두 번째로 작은 $IPM2$ 로부터 생성된 현재 블록 예측자를 의미한다. w_{IPM1}, w_{IPM2}

는 각 IPM에 대한 가중치를 의미하며, $SATD_{IPM1}$ 과 $SATD_{IPM2}$ 는 각 IPM에 대해 템플릿 영역에서 계산된 SATD 비용을 의미한다. TIMD 모드는 ECM-1.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.47%, -0.34%, -0.37%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 124%와 111%로 측정되었다.

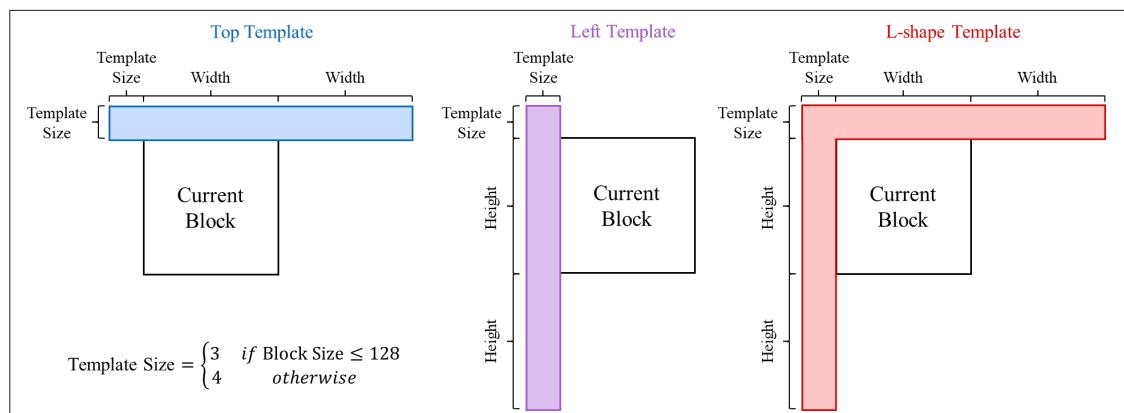
4. Most Dominant Intra Prediction (MDIP) 기술

MDIP 모드는 복호화기에서 현재 블록 주변의 복원된 샘플로부터 계산된 HoG를 기반으로 MDIP 후보 리스트를 구성하고, 구성된 후보들에 대해서 템플릿 매칭 비용을 이용하여 현재 블록 예측에 사용할 IPM을 유도하는 화면 내 예측 기술이다[10]. MDIP 모드는 블록 크기가 1024 이하인 경우에 적용되며, 전체 동작 과정은 다음과 같다.

단계 1. 템플릿 영역 구성 및 HoG 계산

MDIP 모드에서는 <그림 6>과 같이 현재 블록의 상단

$$Pred_{TIMD} = w_{IPM1} * Pred_{IPM1} + w_{IPM2} * Pred_{IPM2}, w_{IPM1} = \frac{SATD_{IPM2}}{SATD_{IPM1} + SATD_{IPM2}}, w_{IPM2} = 1 - w_{IPM1} \quad (3)$$



<그림 6> MDIP 모드에서의 템플릿 영역

및 좌측에 위치한 복원된 샘플을 이용하여 템플릿 영역(template area)을 구성하고, 해당 영역으로부터 HoG를 계산한다. HoG 계산 과정은 DIMD와 동일한 방식으로 수행되나, 템플릿 영역의 구성은 블록 크기에 따라 달라진다.

현재 블록의 샘플 수가 256 미만인 경우에는 3-line 템플릿 영역을 사용하며, 256 이상인 경우에는〈그림 6〉에 도시된 바와 같이 4-line 템플릿 영역을 사용한다. 계산된 HoG는 각 인트라 예측 모드에 대응되는 gradient amplitude 분포를 제공하며, 이후 후보 IPM 선별 과정에 활용된다.

단계 2. HoG 기반 MDIP 후보 리스트 구성

HoG 계산 결과를 기반으로 최대 8개의 IPM 후보를 도출한다. 구체적으로, 템플릿 영역별 HoG를 이용하여 다음과 같이 후보 IPM을 구성한다. 우선 L-shape 형태인 상단 및 좌측 템플릿 영역으로부터 HoG amplitude가 가장 높은 4개의 IPM을 MDIP 후보 리스트에 추가한다. 다음으로, 좌측 템플릿 영역과 상단 템플릿 영역으로부터 각각 HoG amplitude가 가장 높은 2개의 IPM 후보를 MDIP 후보 리스트에 추가한다. 이때, MDIP 후보 리스트 내에 이미 추가된 IPM은 중복 구성되지 않도록 한다. 상기 절차 이후에도 후보 리스트가 모두 구성되지 않는 경우, MDIP 후보 리스트에 구성된 IPM들에 대하여 $-1, +1, -2, \dots, +4$ 의 offset을 적용한 IPM을 MDIP 후보 리스트에 추가한다. 해당 과정은 MDIP 후보 리스트 내 첫 번째 후보부터 MDIP 후보 리스트가 모두 구성될 때까지 반복한다.

단계 3. 템플릿 매칭 비용 기반 최종 후보 선정

구성된 MDIP 후보 리스트 내 IPM 후보들에 대하여, 현재 블록에 대한 좌측 및 상단 1줄을 템플릿 매칭 비용을 계산하기 위한 템플릿으로 사용하여 템플릿 매칭 비용을 계산한다. 각 IPM에 대해 템플릿 영역을 예측한 결과와 실제 복원된 템플릿 샘플 간의 차이를 기반으로 비용을 산출하며, 가장 작은 템플릿 매칭 비용을 갖는 IPM을 최종 MDIP 모드로 선정한다.

MDIP 모드에서는 현재 블록 예측에 사용된 IPM과 HoG amplitude가 낮은 인트라 예측 모드를 PMPM, SMPM 및 Remainder 리스트에서 제외함으로써, 신호에 필요한 오버헤드를 감소시키고 부호화 효율을 향상 시킨다. 구체적으로, L-shape 템플릿 영역에 대하여 계산된 HoG 결과 중 amplitude가 가장 작은 20개의 IPM 을 PMPM, SMPM 및 Remainder 리스트에서 제외한다. MDIP 모드는 ECM-15.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.07% , -0.12% , -0.03% 의 BDDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 100.7%와 100.1%로 측정되었다.

5. Spatial Geometry Partitioning Mode (SGPM) 기술

SGPM 모드는 현재 블록을 기하학적 분할 구조로 분할하고, 분할된 각 영역에 서로 다른 IPM을 적용하여 예측을 수행하는 화면 내 예측 기술이다[11]. SGPM 모드는 하나의 분할 모드와 두 개의 IPM 조합으로 구성되며, 각 IPM 은 분할된 두 영역 중 하나에 각각 적용된다. SGPM은 블록의 너비와 높이가 각각 4 이상 64 이하이고, 블록의 종횡비가 1:8 보다 작으며 블록 크기가 32 이상인 경우에 적용된다. 이를 통해 지나치게 가늘거나 작은 블록에 SGPM이 적용되는 것을 방지한다.

SGPM 모드에서는 사전에 정의된 26개의 기하학적 분할 모드 중 하나를 사용하여 현재 블록을 두 영역으로 분할한다. 분할 경계에서는 두 영역에 대한 예측자가 블렌딩되며, 블렌딩 폭은 블록 크기에 따라 적응적으로 조절된다. 구체적으로, 블록의 최소 변 길이에 따라 블렌딩 폭을 조정함으로써, 작은 블록에서는 과도한 블렌딩을 방지하고 큰 블록에서는 보다 완만한 경계 처리가 가능하도록 한다. 이러한 분할 구조를 효율적으로 신호하기 위하여, SGPM 모드는 분할 모드와 IPM 조합을 직접 신호하는 대신 후보 리스트 기반의 신호화 방식을 사용한다. 즉, 하나의 분할 모드와 두 개의 IPM으로 구성된 조합들을 사전에 정의된 절

차에 따라 후보로 구성하고, 이를 중 하나를 선택하여 신호 함으로써 분할 및 예측 정보를 간결하게 표현한다. 각 분할 영역에 적용할 IPM은 제한된 후보 리스트로부터 선택된다. 분할 영역별 IPM 후보 리스트의 크기는 3으로 제한되며, 해당 후보 리스트는 inter-intra GPM에서 사용된 IPM 유도 방식을 재사용하되, TIMD 기반으로 유도된 모드는 제외된다. 이를 통해 고려해야 할 후보 수를 제한하면서도 예측 성능 저하를 최소화한다. SGPM에서는 현재 블록의 상단 1-line과 좌측 1-line으로 구성된 템플릿 영역을 이용하여 분할 모드와 IPM 조합의 적합도를 평가한다. 각 분할 모드와 두 IPM의 조합에 대하여, 분할 모드에 따른 블렌딩 가중치를 템플릿 영역까지 확장하여 템플릿 예측을 수행하고, 예측된 템플릿과 복원된 템플릿 샘플 간의 차이를 기반으로 SAD 비용을 계산한다. 계산된 SAD 비용을 기준으로 후보 조합들을 오름차순으로 정렬한 후, 상위 16개의 조합만을 SGPM 후보 리스트에 구성한다. SGPM 모드는 ECM-6.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.25%, -0.29%, -0.25%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 103%와 103%로 측정되었다.

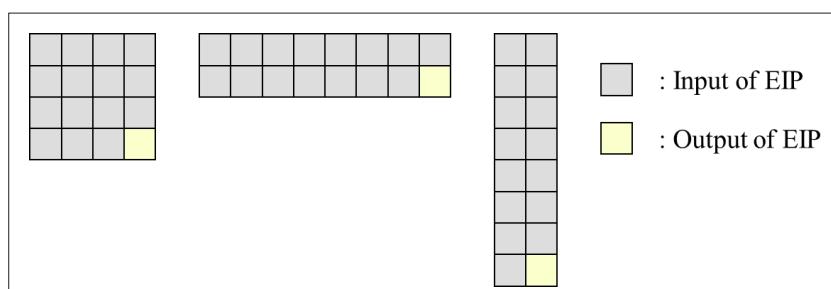
6. Extrapolation Filter-based Intra Prediction (EIP) 기술

EIP 모드는 현재 블록 주변의 복원된 샘플로부터 extrapolation filter를 직접 유도하거나, 이전 블록에서

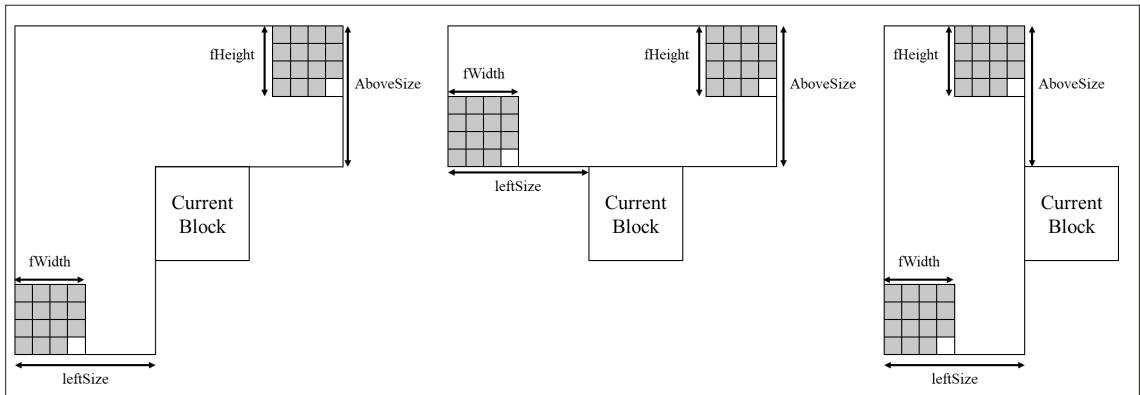
사용된 filter를 상속받아 현재 블록의 예측자를 생성하는 화면 내 예측 기술이다[12]. EIP 모드는 직사각형 혹은 정사각형 형태의 extrapolation filter를 사용하여 샘플 단위의 예측을 수행하며, 휘도 채널에 대해서만 적용된다. 또한, EIP 모드는 블록 크기가 32×32 이하이며 블록의 샘플 수가 64 이상인 경우에 한하여 사용된다. EIP 모드에서 사용되는 extrapolation filter는 하나의 샘플 값은 추정하기 위하여 15개의 입력 샘플을 참조하며, <그림 7>에 도시된 바와 같이 세 가지 서로 다른 filter shape 중 하나를 사용한다.

현재 블록 예측에 사용되는 filter 계수는 직접 유도되거나, 상속받아 사용할 수 있다. 직접 유도하는 방식의 경우에는 현재 블록의 참조 영역의 복원된 샘플을 이용하여 filter 계수를 추정한다. 이 경우, 복호화기는 신호받은 syntax를 통하여 참조 영역의 종류 (Left-Top, Left, Top)와 filter shape을 결정한다. 선택된 filter는 참조 영역 내에서 수평 또는 수직 방향으로 한 샘플씩 이동하면서 auto-correlation matrix와 cross-correlation vector를 구성하고, 이를 기반으로 filter 계수를 계산한다. 계수 계산에는 LDL decomposition method와 Gaussian back-substitution이 사용된다.

참조 영역의 크기는 현재 블록의 크기와 선택된 filter shape에 따라 결정된다. <그림 8>은 서로 다른 filter shape이 선택된 경우에 대하여, 현재 블록에 대한 참조 영역이 어떻게 구성되는지를 도식화한 것이다. 그림에 도



<그림 7> EIP 모드에서 사용되는 세 종류의 filter shape



<그림 8> Filter shape에 따른 EIP 모드에서의 참조 영역

$$aboveSize = \min(blockWidth, blockHeight) + fWidth - 1 \quad (4)$$

$$LeftSize = \min(blockWidth, blockHeight) + fHeight - 1 \quad (5)$$

$$Pred_{(x,y)} = \sum_{i=0}^{14} c_i * t_{(x-offsetX_i, y-offsetY_i)} \quad (6)$$

시된 바와 같이, 참조 영역의 가로 (left Size) 및 세로 크기 (above Size)는 각각 식 (4)와 (5) 같이 정의된다. 이 때, $fWidth$ 와 $fHeight$ 는 각각 선택된 extrapolation filter의 너비와 높이를 의미한다.

EIP 모드는 위와 같이 결정된 extrapolation filter를 이용하여 현재 블록의 예측 값을 생성한다. 예측은 현재 블록의 좌상단 위치부터 우하단 위치까지 대각선(diagonal) 순서로 수행되며, 각 위치의 예측 값은 식 (6)과 같이 계산된다.

여기서 $Pred_{(x,y)}$ 는 현재 블록 내 (x,y) 에서의 예측자이며, c_i 는 선택된 extrapolation filter의 i 번째 계수이다. 또한, $t_{()}$ 는 참조 영역의 복원된 샘플 값 또는 현재 블록 내에서 예측된 샘플 값을 의미하며, $x-offset X_i$ 와 $y-offset Y_i$ 는 filter shape에 따라 정의된 위치 오프셋이다. EIP 모드는 ECM-11.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.20%, -0.14%, -0.18%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 102.1%와 101.3%로 측정되었다.

7. Neural network-based Intra Prediction (IntraNN) 기술

IntraNN 모드는 복호화기에서 현재 블록 주변의 참조 샘플을 입력으로 하여, 사전에 학습된 신경망 모델을 통해 현재 블록에 대한 예측을 수행하는 화면 내 예측 기술이다[9]. 기존의 방향성 기반 인트라 예측 방식과 달리, 참조 샘플과 예측 블록 간의 상관관계를 간단한 신경망 모델을 통해 직접 모델링함으로써 보다 정확한 예측자를 생성한다. IntraNN 모드에는 복호화 복잡도를 줄이기 위한 여러 가지 최적화 기법이 적용되었다. 우선, 신경망 추론은 SADL (Simple Acceleration Deep Learning) 프레임워크를 이용하여 수행되며, 모델 가중치 및 중간 연산은 16비트 정수 정밀도로 표현되고 내부 누적 연산은 32비트 정수로 수행된다. 이를 통해 부동소수점 연산을 배제하고 복호화기 구현 복잡도를 낮춘다. 또한, 신경망 모델의 가중치 행렬에는 학습 단계에서부터 sparsity 제약이 강하게 적용되어, 대부분의 가중치가 0이 되도록 유도된다.

이러한 sparse 구조는 압축된 행 (row) 기반 저장 방식과 sparse 행렬 곱 연산을 통해 효율적으로 처리되며, 픽셀 당 MACs (Multiply-Accumulate Operations) 수와 디코딩 시간 복잡도를 효과적으로 감소시킨다. ECM IntraNN에서는 예측 성능을 유지하면서 모델 파라미터 수를 줄이기 위해, 블록 크기에 따라 서로 다른 6개의 신경망 모델을 사용한다. 각 모델은 4×4 , 8×4 , 16×4 , 8×8 , 16×8 , 16×16 크기의 휘도 블록을 직접 예측하도록 설계되어 있으며, 이외의 블록 크기에 대해서는 VVC의 MIP와 유사하게 참조 샘플의 전치 (transposition) 및 다운샘플링을 통해 입력을 구성한다. 전체 IntraNN 모드에서 사용되는 파라미터 수는 약 76만 개로, 기존 방식 대비 모델 규모가 크게 축소되었다.

신경망을 통해 생성된 예측 블록은 최종적으로 인트라 예측 모드 (IPM)를 직접 신호하지 않고, DIMD와 동일한 방식으로 예측 결과의 기울기 정보를 분석하여 등가의 IPM을 복호화기에서 유도한다. 이로 인해 TMP, EIP, MIP 와 마찬가지로 추가적인 인트라 모드 시그널링 없이 예측이 가능하다. 휘도 블록 크기가 64 샘플을 초과하는 경우에는, DIMD로 유도된 IPM과 함께 PLANAR 모드를 추가 후보로 사용하며, 이 중 선택된 IPM에 대한 CU-level 플래그만을 신호한다. 결과적으로, IntraNN 기반 예측은 신경망의 표현력을 활용하여 예측 정확도를 향상시키면서도, sparse 가중치, 정수 연산, 블록 크기별 모델 분리 및 복호화기 측 IPM 유도 구조를 통해 디코딩 복잡도를 효과적으로 제어한 ECM용 화면 내 예측 기술이라 할 수 있다. IntraNN 모드는 ECM-14.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.58%, -0.40%, -0.40%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 101.6%와 109.8%로 측정되었다.

8. Matrix-based Position Dependent Intra Prediction (MPDIP) 기술

MPDIP 모드는 기존의 방향성 기반 화면 내 예측 모드

와 사전 학습된 행렬과 참조 샘플을 이용해 현재 블록에 대한 예측을 수행하는 화면 내 예측 기술이다[8]. MPDIP 모드에서는 복원된 주변 참조 샘플을 입력으로 하여, 현재 블록 내 위치에 따라 선택되는 예측 필터 계수를 적용함으로써 현재 블록의 예측 샘플을 생성한다. 해당 필터 계수는 사전 학습 과정을 통해 도출되며, 부호화기와 복호화기가 동일한 필터를 사용하므로 추가적인 신호화는 필요하지 않다. MPDIP 모드는 휘도 채널에 대해 적용되며, 사전에 정의된 블록 크기 집합에 포함되는 경우에 한하여 활성화된다. 또한, 모든 인트라 예측 모드에 대해 MPDIP를 적용하는 것이 아니라, 일부 방향성 인트라 예측 모드를 PDP 기반 예측으로 대체하는 방식으로 동작한다. 블록 크기 및 선택된 인트라 예측 모드에 따라 참조 샘플의 길이와 필터 적용 방식이 달라지며, 이를 통해 다양한 공간적 특성을 효율적으로 반영한다. 상대적으로 큰 블록의 경우, MPDIP 예측은 16×16 위치에서 우선 수행되며, 이후 보간 기반의 업샘플링 과정을 통해 전체 블록에 대한 예측 값이 생성된다. 이러한 구조를 통해 MPDIP 모드는 기존 angular intra prediction 대비 향상된 예측 정확도를 제공하면서도, 복호화 복잡도를 제한된 범위 내에서 유지할 수 있다. MPDIP 모드는 ECM-12.0을 기준으로 Y, Cb, Cr 채널에서 각각 -0.31%, -0.33%, -0.34%의 BDBR 성능을 보였다. 인코딩 시간과 디코딩 시간의 경우 101.2%와 100.2%로 측정되었다.

III. 결 론

본 글에서 분석한 ECM 기반 휘도 채널 화면 내 예측 기술들은, 압축 성능 향상을 위해 서로 다른 방향을 취하고 있음을 확인할 수 있었다. 이들 기술은 복호화기 측에서 예측자를 생성하는 접근, 예측에 사용되는 후보 리스트를 재구성하는 접근, 그리고 사전 학습된 모델을 사용하여 예측자를 생성하는 접근으로 구분될 수 있으며, 각 방식은 예측 정확도, 신호 오버헤드, 그리고 계산 복잡도 간

의 상이한 절충점을 가진다. 특히 복호화기 측에서 예측자를 생성하는 화면 내 예측 기술들은 추가적인 예측 정보의 신호화 없이 예측자를 생성할 수 있다는 장점을 가지는 반면, 템플릿 영역에 대한 접근으로 인한 계산 복잡도 및 하드웨어 구현 측면에서의 부담이 함께 수반된다. 이는 Beyond VVC 단계의 화면 내 예측 기술 설계에서 압

축 성능뿐만 아니라, 하드웨어 친화성과 시스템 구현 가능성을 함께 고려할 필요가 있음을 시사한다. 이러한 관점에서 향후 화면 내 예측 기술은 단순한 예측 정확도 향상을 넘어, 예측 구조의 단순화, 템플릿 활용 범위의 제어, 그리고 복잡도 관리 측면에서의 설계가 중요한 방향으로 고려될 것으로 판단된다.

참 고 문 헌

- [1] B.Bross, et, al., “Versatile Video Coding Editorial Refinements on Draft 10”, Joint Video Experts Team (JVET), 20th Meeting, teleconference, JVET-T2001, 2020.
- [2] F. Bossen, et, al., “JVET AHG report: Test model software development (AHG3)”, Joint Video Experts Team (JVET), 20th Meeting, teleconference, JVET-T0003, 2020.
- [3] https://vcgit.hhi.fraunhofer.de/ecm/ECM/-/tree/ECM-19.0?ref_type=tags
- [4] V. Seregin, et, al., “JVET AHG report: ECM software development (AHG6)”, Joint Video Experts Team (JVET), 40th Meeting, Geneva, JVET-AN0006, 2025.
- [5] J. Zhao, et, al., “EE2-Related: Improvements of Decoder-Side Intra Mode Derivation”, Joint Video Experts Team (JVET), 22nd Meeting, teleconference, JVET-V0087, 2021.
- [6] R. G. Youvalari, et, al., “EE2-2,2: Occurrence-based intra coding (OBIC)”, Joint Video Experts Team (JVET), 34th Meeting, Rennes, JVET-AH0076, 2024.
- [7] K. Cao, et, al., “EE2-related: Fusion for template-based intra mode derivation”, Joint Video Experts Team (JVET), 23rd Meeting, teleconference, JVET-W0123, 2021.
- [8] B. Ray, et, al., “EE2-2,13: Matrix based intra prediction replacing conventional intra modes”, Joint Video Experts Team (JVET), 34th Meeting, Rennes, JVET-AH0209, 2024.
- [9] T. Dumas, et, al., “EE2-2,20_2,21: Neural network-based intra prediction with DIMD mode derivation”, Joint Video Experts Team (JVET), 36th Meeting, Kemer, JVET-AJ0249, 2024.
- [10] M. Hong, et, al., “EE2-1,2: Intra mode coding based on HoG of neighboring templates”, Joint Video Experts Team (JVET), 37th Meeting, Geneva, JVET-AK0059, 2025.
- [11] F. Wang, et, al., “EE2-1,6: Combination of spatial GPM tests”, Joint Video Experts Team (JVET), 28th Meeting, Mainz, JVET-AB0155, 2022.
- [12] L. Xu, et, al., “EE2-1,14: An extrapolation filter-based intra prediction mode”, Joint Video Experts Team (JVET), 33rd Meeting, teleconference, JVET-AG0058, 2024.

저자 소개



정현기

- 2022년 : 성균관대학교 수학과 졸업 (학사)
- 2024년 : 성균관대학교 전자전기컴퓨터공학과 졸업 (석사)
- 2024년 ~ 현재 : 성균관대학교 전자전기컴퓨터공학과 박사과정
- 주관심분야 : 멀티미디어, 영상압축



김용성

- 2024년 : 성균관대학교 전자전기공학부 졸업 (학사)
- 2024년 ~ 현재 : 성균관대학교 전자전기컴퓨터공학과 석박사통합과정
- 주관심분야 : 멀티미디어, 영상압축



전병우

- 1985년 : 서울대학교 전자공학과 (학사)
- 1987년 : 서울대학교 전자공학과 (석사)
- 1992년 : Purdue Univ. School of Elec. 졸업 (공학박사)
- 1993년 ~ 1997년 : 삼성전자 신호처리연구소 선임/수석연구원
- 1997년 ~ 현재 : 성균관대학교 전자전기공학부 교수
- 주관심분야 : 멀티미디어, 영상압축, 영상인식, 신호처리