

[招待講演] 次世代動画像符号化方式 VVC の 標準化動向と深層学習の適用

河村 圭^{1,a)}

概要：次世代動画像符号化方式 Versatile Video Coding (VVC) の国際標準化が 2018 年 4 月に開始され、2020 年 10 月標準化完了を目指して議論が続いている。本稿は、VVC の標準化動向として、2019 年 1 月会合までに採用された符号化ツールを概観し、VVC の特徴を紹介する。また発展的な話題として、深層学習のひとつである畳込みニューラルネットワーク (CNN) を動画像符号化方式に適用する事例と、深層学習で得られたモデル自体の符号化の標準化動向を紹介する。

VVC は汎用的に使える映像符号化方式を目指している。カメラで撮影した放送用映像に加えて、監視カメラ映像やユーザ生成映像、コンピュータ生成映像、近年話題となっていた 360 度映像 (Virtual Reality 映像) など、様々なアプリケーションの映像を対象としている。これは、標準化会合で利用されているテストシーケンスにも反映されている。目標とする符号化性能は、HEVC に対して 30% 以上である。これは、同一画質で比較をした時に、HEVC の帯域に対して VVC の帯域が 30% 削減 (= 70% の帯域) を意味する。

ブロック分割としては、HEVC の四分木構造から大きく拡張され、四分木・二分木・三分木構造が採用された。この変更だけで、HEVC に対して 8% (4K 映像では 10%) の性能改善を実現している。

イントラ予測では、角度単位が半分となり予測方向が倍増した。また、色差成分に特化した予測方式として、再構成した輝度信号から色差信号を線形予測する方式も採用され、色差予測性能が著しく向上した。

インター予測には、非常に多くの工夫が追加されている。HEVC の高度動きベクトル予測モードとマージモードは継承されたが、後者に動きベクトル予測差分を送る方法が追加された。アフィン変換に基づく動き補償として、対象ブロックを 4x4 サブブロックに分割し、サブブロックの動きベクトルのみをアフィン変換により導出し、補間画像を平行移動モデルで生成する。三角形パッチや、イントラ・インター融合の動き補償も採用された。さらには、動きベクトルをデコーダ側で導出する方式と、デコーダ側でオプティカルフローに基づく画素値補正方式が採用された。

変換は、HEVC の DCT-II/DST-IIIV に加えて、複数の DCT/DST タイプを追加することが検討されている。また、量子化もレベル値に依存した逆量子化が採用された。

インループフィルタには、これまでのデブロッキングフィルタとサンプル適応オフセットフィルタに加えて、適応ループフィルタが採用された。

このように、VVC にはこれまでの単純な改良に限定されない各種符号化ツールが採用されている。また、採用にあたってはハードウェア実装の実現可能性が入念に確認され、純粋な性能向上だけでなく実用化を前提とした議論がなされている。

目新しい動向として深層学習の活用が挙げられる。深層学習は演算量が非常に多いため性能改善も大きいですが、採用には慎重な姿勢が貫かれている。現在までのところ、深層学習のうち畳込みニューラルネットワークをインループフィルタとして利用する方法が複数提案されており、有意な性能改善が確認されている。一方で、必要な計算量に見合った性能向上であるのか、そもそもリアルタイム処理が可能なモデルサイズであるかが議論されている。認識分野におけるモデルと比較すると、そのレイヤ数は 1 桁前半であり、桁違いの小ささである。

さらに発展的な内容として、深層学習で導出されたモデルそのものの圧縮が検討されている。元々、静止画像や動画像の検索をアプリケーションとする、コンテンツ記述子の標準化が MPEG-7 として進められていた。その一環として、深層学習により抽出された記述子の性能が高いことが確認され、モデル自体の符号化が必要ではないかという議論になった。すでにモデルの交換フォーマットは様々な団体に議論されているが、圧縮にフォーカスした議論はあまりない。符号化されたモデル自体のファイルサイズと、モデル自体の劣化による検出性能低下とを定量的に裏付して、その符号化方式が検討される予定である。2018 年 10 月に提案募集が開始され、2019 年 3 月から標準化が開始される。

¹ KDDI 総合研究所
埼玉県ふじみ野市大原 2-1-15

a) kei@kddi-research.jp