6章 映像における次世代・次々世代符号化技術

杉本志織[†], 坂東幸浩[†](正会員), 高村誠之[†](正会員)

キーワード: Versatile Video Coding, MPEG, JPEG Pleno, 光線空間, 深層学習

1. まえがき

映像の広帯域化と符号化技術は共に成長を続けてきている。今後も、衛星経由で実用放送される4K8Kコンテンツを地上波で放送したり、スマートフォンなどで長時間撮影したり視聴する場合等で、現在よりも高い圧縮性能が求められる。またVR全天周(360°)映像や高密度多視点映像(ライトフィールド)、点群など従来の2D映像の範疇を超え、データ量が桁違いに増大する映像フォーマットも近年台頭している。したがってこれらへの対応を目的とした、高性能・高機能な映像符号化技術の研究開発が求められている。本稿では、現在標準化が進行中の次世代映像符号化規格Versatile Video Coding (VVC)の概要と、さらに先を見据えた取り組みである次々世代映像符号化の状況を述べる。

2. 次世代映像符号化規格VVC

2.1 JVET活動概要

ITU-T SG16 Video Experts Coding Group (VCEG)とISO/IEC JTC 1/SC 29/WG 11 Moving Picture Experts Group (MPEG)は、2015年10月のGeneva会合において次世代映像符号化規格の標準化を行う合同チームJoint Video Exploration Team (JVET)を設置した。同規格はH.265/HEVC比30%から50%のビットレート削減、2020年10月の標準化完了を目標としている¹⁾。

JVETは結成以後Standard Dynamic Range (SDR), High Dynamic Range (HDR), 360°の三つのテストカテゴリに対して、さまざまな候補技術の性能を確認する探索活動を行い、2017年4月にはCall for Evidence (CfE)²⁾を発行した。2017年10月Macau会合ではCall for Proposal (CfP)³⁾発行にあたり、チーム名をJoint Video Experts Teamに改称し、2018年4月のSan Diego会合におけるCfPレスポンスの評価を経て、規格名をVVCとした上でWD1⁴⁾

を発行し、標準化を開始した。

2.2 ターゲット

Versatile (多目的な)の名が示すとおり、VVCがターゲットとする映像は多岐にわたる。2018年7月時点でテストカテゴリはSDR, HDR, 360°, Screen Content Coding (SCC)の4種類である。SDR以外のカテゴリについて専用の符号化ツールを定義するか、プロファイルを切り分けるかどうかは現時点で未定である。H.264/AVC, H.265/HEVCはプロファイル/レベルを定義することで、アプリケーションや機器ごとにサポート機能を選択可能にしつつ互換性を確保していたが、VVCにおいて同様の仕組みを定義するかどうかはまだ定まっていない。現時点では符号化ツール間になるべく依存性を持たせず、必要に応じてツール別のOn/Offを切り替えることを意識して検討が進められている。

HDRはHEVCにおいては専用の符号化ツールが定義されず、Main 10プロファイルによる10bit符号化とSupplemental Enhancement Information (SEI)メッセージによるパラメータ伝送およびプリ・ポスト処理で対応されていたが、VVCにおいては専用符号化ツールの定義も検討対象とされている。

360°はHEVCではターゲットとしていなかったカテゴリであり、360°映像の平面展開映像を符号化対象とするものである。専用符号化ツールの他、プリ・ポスト処理用リファレンスソフトウェア360 Libを使用した符号化前の平面展開と復号後の再投影が検討対象となっている。MPEGでは同様の映像に対応するフォーマットOmnidirectional Media Application Format (OMAF)の標準化も進行中である。

SCCはHEVCにおいてはver. 4でプロファイルが設けられたが、VVCにおいても他の3カテゴリに少し遅れる形で2018年7月に検討が始まった。従来SCCのユースケースはPC画面の伝送などが中心に考えられていたが、VVCでは新たにeSportsの配信が挙げられている。専用符号化ツールはHEVCではPalette CodingやIntra Block Copyなどが存在しており、VVCにおいても同様のツールが検討対象に上がっている。

[†] NTTメディアインテリジェンス研究所
"Next-Generation Video Coding and the Next after the Next" by Shiori Sugimoto, Yukihiro Bando and Seishi Takamura (NTT Media Intelligence Laboratories, Yokosuka)

2.3 技術動向

探索活動ではHEVCのテストモデルであるHMをベースにVCEGが開発したHM KTAソフトウェアにさまざまな符号化ツールを追加して、探索モデルJoint Exploration Test Model (JEM)⁵⁾が開発された。CfP発行直前時点のJEM7.0は客観評価においてHM比29%程度のビットレート削減性能を示していた⁶⁾。また、CfPに対して寄せられたレスポンスの大半は主観評価においてHMおよびJEMと比較して良い性能を示し⁷⁾、客観評価においては最も良いものでHM比40%程度のビットレート削減性能を示した。

CfPレスポンスを承けて策定されたテストモデルVersatile Test Model (VTM) ver. 1 (VTM1)⁸⁾は、HEVCをベースとしつつ、後述するように一部のツールを除去し、ブロック分割構造としてQuad/Ternary/Binary Tree (QTTTBT)を採用した。VTM1から除かれたHEVCツールの一覧を以下に示す。

- ・ Mode Dependent Transform (モード依存変換)
- ・ Mode Dependent Scan (モード依存スキャン)
- ・ Strong Intra Smoothing (イントラ参照画素平滑化)
- ・ Sign Data Hiding (係数符号の暗黙的決定)
- ・ Tiles and Wavefront (タイル分割と時間差処理による並列化)
- ・ Quantization Weighting (量子化重み行列)

QTTTBTは木構造による再帰的分割法で、JEMのブロック分割構造として用いられていたQTBT (Quad/Binary Tree)にTernary分割を加えたものである。QTTTBTで取りうるブロック分割パターンを図1に、分割例を図2に示す。この分割構造は多くのCfPレスポンスで採用されており、性能向上において支配的な役割を果たしていた。HEVCでは分割したブロックをCoding Unit (CU), Prediction Unit (PU), Transform Unit (TU)と呼び分けてそれぞれの単位で予測、変換などを行っていたが、VVCにおいてはそうした分類は撤廃された。符号化対象画像は固定サイズのCoding Tree Unit (CTU)に分割された

図1 QTTTBTのブロック分割パターン

後にQTTTBTによってさらにCUに分割され、CU毎にすべての処理が行われる。ただし、CUサイズが最大の変換サイズよりも大きい場合は一つのCUが複数のTUに分割されて変換される。また、一部のツールではさらに小さなSub-CUと呼ばれる単位で動き補償やベクトル予測を行うこともある。

以上のように、テストモデルVTMはver. 1時点ではブロック分割以外の新規ツールを持たず、非常にシンプルな形でスタートを切った。一方で、別途コア実験 (Core Experiments: CE) のための評価用モデルとしてBenchmark Set (BMS)が定義され、そのver. 1 (BMS1)には探索活動やCfPレスポンスで性能が認められた主要な符号化ツール群が採用された。また、それらツールの多くは2018年7月のLjubljana会合でCEを経てVTM2に採用された。BMSは標準化プロセス上のステータスを持つものではないが、テストモデル採用までには課題を残しつつ有効性が期待されるツールについて、BMSを使用することで単独の性能のみならず他のツールとの相互作用も評価できると期待されている。

VTM/BMSに含まれる主要なツールの一覧を以下に示

図2 QTTTBTによる分割例

表1 HMに対するVTM1/BMS1性能比較⁹⁾(ランダムアクセス)

vs HM16.18	VTM	BMS
BD-Rate ¹⁰⁾ reduction	8.42%	22.84%
Decode time	81%	233%
Encode time	225%	927%

図3 VTM1に対するBMS1のツール別性能比較¹¹⁾(ランダムアクセス)

す。HEVCから継承されているツールは表記しない。

【BMS1】(*はVTM2にも採用)

- ・67 Intra Modes (方向刻みの細かい画面内予測)
- ・LM Chroma (輝度から色差への線形予測) *
- ・AMT (変換基底の適応的切り替え) *
- ・NSST (変換係数の再変換)
- ・ALF (ブロック適応ループフィルタ) *
- ・Coefficient Coding *
- ・Affine Motion (アフィン変換による動き補償) *
- ・AMVR (4, 1, 1/4画素精度の動きベクトル) *
- ・DMVR (テンプレートマッチングによる動き補正)
- ・ATMVP (サブブロック毎の動きベクトル導出) *

【VTM2】

- ・87 Intra Modes (より広角な方向を含む画面内予測)
- ・PDPC (画素位置依存の重み付き予測)
- ・Dependent Quantization (状態依存の量子化)
- ・Sign Data Hiding

【BMS2】

- ・Generalized Bi-prediction (ブロック別重み付き予測)
- ・Current Picture Referencing (ピクチャ内参照予測)
- ・BIO (オブティカルフローによる動き補正)

VTM1/BMS1の客観評価性能を表1に示す。ver. 2は本稿執筆時点で未リリースである。また、VTM1に対するBMS1のツール別の客観評価性能を図3に示す。VVCにおいてビットレート削減性能以外に重視されているのは、復号器の演算量をいかに抑えるかという点である。VTM1では復号時間はHM比0.8倍に抑えられていたものの、BMSの同バージョンでは2倍以上になる。Ljubljana会合では各

ツールの性能と演算量のトレードオフを考慮しつつ、使用するパラメータや制約を探る作業が多く行われた。

その他、今後注目される議論には以下のようなものがある。

- ・非矩形のブロック分割
- ・異なる予測モードを組み合わせた双方向予測
- ・オーバーラップ動き補償
- ・マージモードの改良
- ・各種フィルタ、周波数領域での演算
- ・深層学習の利用

また、HEVCと比較してブロック分割タイプ、符号化ツールの数や変換基底の数がそれぞれ増加したことにより、符号化最適化が難しくなっている。BMS1における符号化時間はHMに対して9倍にもなり、この中の多くのツールがVTM2に採用されたことから、こちらも相当に増加することが予想される。標準化作業の範囲外ではあるが、実用にあたっては Tool 選択を省略するか、または並列化するかなどが課題になると考えられる。

3. 次々世代映像符号化技術

3.1 Light field符号化の研究動向

次期VR/AR等の応用において高臨場感の映像表現を可能にする技術としてlight fieldに対する期待が高まっている。light fieldとは7次元の光線として空間中の光の信号を表現する方法である。光線の要素とは、空間位置を表す情報(3次元)、光線方向(2次元)、時間(1次元)、波長(1次元)である。なお波長の要素を固定とし、さらに光の直進性を仮定する場合が多い。その場合、静止シーンの輝度情報は4次元信号として表現され、4D light fieldと呼ばれる。以下、4D light fieldについて述べる。

4D light fieldは高方向分解能に基づく高臨場感表現を可能にする一方、その情報量増大への対応が課題となる。そこで、4D light fieldの効率的な伝送・蓄積を目的として、JPEG Plenoにおいて4D light fieldの符号化方式の標準化が進められている。以下、その標準化活動について概説する。

JPEG PlenoはISO/IEC JTC 1/SC 29/WG 1 (Joint Photographic Experts Group, JPEG)において、360°画像、点群、ホログラフィ、そしてlight fieldといった新たな画像形式の符号化を検討する標準化のフレームワークである^{12) 13)}。符号化方式とともに、メタデータ処理(metadata manipulation)、ランダムアクセス性、著作権保護、セキュリティ等をサポートする機能も検討する。2016年10月にCfPを発行し、light field符号化の標準化は表2のスケジュールで進められている。

提案された符号化ツールは、定められた共通実験条件に基づき評価される。評価に用いるコンテンツはレンズレット型カメラおよび高密度カメラアレイで撮影されたlight fieldデータから変換された多視点画像である。レンズレット型カメラとは主レンズと複数のマイクロレンズアレイの組合せで構成される撮像デバイスである。符号化性能の評

表2 JPEG Pleno : light field符号化の作業計画

作成対象	時期
作業原案	2018年4月(済)
委員会原案	2019年1月
標準規格原案	2019年7月
標準規格	2020年1月

価は次のような尺度に基づき行われる。

- ・Rate-distortion性能を評価するためのBjontegaard Delta Bitrate (BD-Rate), Bjontegaard Delta PSNR (BD-PSNR)¹⁰⁾
- ・復号画質を評価するためのPSNR, SSIM
- ・ランダムアクセス性を評価するための“random access penalty”

ここでrandom access penaltyとは興味領域 (Region of Interest: ROI) を復号するために必要な情報量が全情報量に占める割合であり, random access penaltyが大きな値をとることはランダムアクセス性の低いデータ構造であることを示す。例えば, random access penaltyが最大値1をとるのは, ROIの復号にすべての情報を必要とするデータ構造であることを意味する。

符号化ストリームは, 参照視点画像, 視差情報, その他の補助情報から構成される。参照視点画像以外の視点画像の生成処理は, 符号化器における処理とされ標準化の範囲外となる予定である。今後はCfPレスポンスの評価に必要な準備作業 (exploration studies) が進められる。2018年7月時点の作業項目としては, Verification Model (VM) と呼ばれる参照ソフトウェアの性能評価, VMにおけるランダムアクセス性に関する実験条件の設定, 評価用コンテンツに対する視差データの生成等があげられている。またコア実験として, 提案された符号化ツールの評価が行われる。2018年7月時点の評価項目としては視差データの符号化, ワーピング・予測処理, 参照視点画像の符号化, 残差視点画像の符号化が挙げられており, 2020年1月の規格策定を目指して活動が進められている。

3.2 深層学習応用・新たな挑戦

(1) 深層学習

深層学習はパターン認識を端緒とし爆発的にその応用範囲を広げつつあるが, こと画像・映像符号化についてはその参入は比較的緩やかであった。

符号化器の最適化を高速に行うため機械学習を用いる試みの改良としてまず適用がなされ (例えば¹⁴⁾), さらに従来の符号化ツールを補足・代替する試みとして, 小数動きベクトルを用いた動き補償における画像補間に深層学習を用いるもの¹⁵⁾や, ループ内フィルタや画面内予測に利用するものが現れた。これらはBMS1比で3~5%の符号量削減が確認され, 既存方式との親和性が比較的高いためVVC標準化でもその採用可否を検討中である¹⁶⁾。

またHEVC復号画に対するポストフィルタとして深層学習を応用し0.46~0.51 dBのPSNR改善を実現するもの¹⁷⁾や, 新しい符号化方式を志向したものとしてEnd-to-Endに深層学習符号化を行うもの (例えば¹⁸⁾), 映像の縮小と深層学習による超解像を組み合わせ符号化に適用した例¹⁹⁾や敵対的生成ネットワーク (Generative Adversarial Networks, GAN) をテキスト領域の補間に応用し主観画質を高める例²⁰⁾などが提案されている。

深層学習は現在着実に符号化への応用が進みつつあり, 今後の動向が注目される。

(2) 新たな挑戦

深層学習以外にも, 従来踏襲され続けてきた「動き補償+直交変換」の枠組みを超えようとする取り組みは存在しており, 遺伝的プログラミングにより新たなアルゴリズムを自動生成する進化的符号化²¹⁾や, 離散的な画素やフレームの集合としての画像・映像を離れ, 連続的な時間・空間中で色分布を持つモデル (Experts) の集合として画像・映像を表現する方法 (例えば²²⁾) などがその嚆矢と言える。

(3) 国際コンペティション

コンピュータビジョン分野の著名国際会議の一つであるCVPRでは, 2018年6月にWorkshopとしてChallenge on Learned Image Compression (CLIC) という画像符号化コンペティション²³⁾を実施した。初回は固定レートで静止画のMOS (主観評価値), PSNR, MS-SSIM (構造類似度に基づく客観画質評価値), 復号速度を競うものであり, 約20機関が参加した。うちPSNRを若干犠牲にしつつもMOSとMS-SSIMで首位となった方式 (tucodec²⁴⁾) は, Autoencoderを用いEnd-to-Endでの圧縮とポストフィルタを組み合わせた新規技術であった。CLICは2019年6月にもCVPR2019内のWorkshopとして開催される予定である。

また従来効率的な符号化が非常に困難であった水底映像を効率的に符号化する試み²⁵⁾を国際コンペティション化したものとして, 画像符号化・映像メディア処理シンポジウムPCSJ/IMPS2018内の企画 (WBVC2018²⁶⁾) が実施されており, CLIC2018に匹敵する参加機関数を集めているところである。

4. むすび

次世代映像符号化規格VVCの概要と, さらに先を見据えたライトフィールド, 深層学習, その他の挑戦的な取り組みに関する次々世代映像符号化の状況を述べた。映像符号化技術は, その歴史も長いが現在も大変活発に国際標準化や研究開発, コンペティションが行われている。今後の新技術の登場や新フォーマット対応, 飽和したと思われる符号化性能の一層の向上などから目を離せない状況である。

(2018年8月9日受付)

〔文 献〕

- 1) ISO/IEC JTC 1/SC 29/WG 11: "Requirements for a Future Video Coding Standard v5", N17074, 119th MPEG Meeting, Torino (2017)
- 2) M. Wien, V. Baroncini, J. Boyce, A. Segall, T. Suzuki: "Joint Call for Evidence on Video Compression with Capability beyond HEVC", JVET-F1002, 6th JVET Meeting (Apr. 2017)
- 3) A. Segall, V. Baroncini, J. Boyce, J. Chen, T. Suzuki: "Joint Call for Proposals on Video Compression with Capability beyond HEVC", JVET-H1002, 8th JVET Meeting (Oct. 2017)
- 4) B. Bross: "Versatile Video Coding (Draft 1) ", JVET-J1001, 10th JVET Meeting (Apr. 2018)
- 5) J. Chen, E. Alshina, G.J. Sullivan, J.-R. Ohm, J. Boyce: "Algorithm description of Joint Exploration Test Model 7 (JEM7) ", JVET-G1001, 7th JVET Meeting (July 2017)
- 6) X. Li, K. Suehring: "JVET AHG report: JEM software development (AHG3) ", JVET-H0003, 8th JVET Meeting (Oct. 2017)
- 7) V. Baroncini, J.-R. Ohm, G.J. Sullivan: "Report of results from the Call for Proposals on Video Compression with Capability beyond HEVC", JVET-J1003, 10th JVET Meeting (Apr. 2018)
- 8) J. Chen, E. Alshina: "Algorithm description for Versatile Video Coding and Test Model 1 (VTM 1) ", JVET-J1002, 10th JVET Meeting (Apr. 2018)
- 9) F. Bossen, X. Li, K. Suehring: "JVET AHG report: Test model software development (AHG3) ", JVET-K0003, 11th JVET Meeting (July 2018)
- 10) G. Bjøntegaard: "Calculation of average PSNR differences between RD-Curves", VCEG-M33, 13th VCEG Meeting (Apr. 2001)
- 11) W.-J. Chien, J. Boyce, E. Alshina, J. Chen, E. Francois, Y. He, Y.-W. Huang: "JVET AHG report: Tool reporting procedure (AHG13) ", JVET-K0013, 11st JVET Meeting (July 2018)
- 12) Overview of JPEG Pleno, <https://jpeg.org/jpegpleno/>
- 13) ISO/IEC JTC 1/SC 29/WG 1: "JPEG Pleno Call for Proposals on Light Field Coding", N73013 (Oct. 2016)
- 14) Y. Wang et al: "Fast CU partition decision using machine learning for screen content compression", OM-P3.5, in Proc. ICIP2015 (Oct. 2015)
- 15) N. Yan et al: "A Convolutional Neural Network Approach for Half-Pel Interpolation in Video Coding", ISCAS2017 (May 2017)
- 16) S. Liu et al., JVET-K0009 "JVET AHG report: Neural Networks in Video Coding (AHG9)"
- 17) R. Yang et al: "Multi-Frame Quality Enhancement for Compressed Video", CVPR2018 (June 2018)
- 18) J. Ballé et al: "End-to-end optimization of nonlinear transform codes for perceptual quality", in Proc. PCS2016 (Dec. 2016)
- 19) S. Umeda et al: "HDR Video Super-Resolution for Future Video Coding", IWAIT2018 (Jan. 2018)
- 20) 折橋ほか: "敵対的生成ネットワークを用いた画像補間に基づく画像符号化方式", 画像工学会, IE2018-27 (June 2018)
- 21) 高村: "進化的映像符号化", 映像学誌, 69, 2, pp.J38-J44 (Mar. 2015)
- 22) R. Verhack et al: "A universal image coding approach using sparse steered Mixture-of-Experts regression", in Proc. ICIP2016, pp.2142-2145 (Sep. 2016)
- 23) CLIC, <http://www.compression.cc/>
- 24) L. Zhou et al: "Variational Autoencoder for Low Bit-rate Image Compression", CVPR Workshop, pp.2617-2620 (June 2018)
- 25) 高村ほか: "水底映像符号化の効率改善に関する一検討", PCSJ2017, P-2-7 (Nov. 2018)
- 26) Water Bottom Video Challenge 2018, <https://sites.google.com/view/wbvc2018w>

杉本 志織 2010年、早稲田大学先進理工学研究科物理および応用物理学専攻修士課程修了。同年、NTTサイバースペース研究所入社。現在、研究主任として、NTTメディアインテリジェンス研究所にて、映像符号化の研究開発および国際標準化活動に従事。

坂東 幸浩 2002年、九州大学大学院システム情報科学研究科知能システム学専攻博士後期課程修了。同年、NTTサイバースペース研究所入社。現在、主任研究員(特別研究員)として、NTTメディアインテリジェンス研究所にて、映像符号化の研究開発に従事。博士(工学)。正会員。

高村 誠之 1996年、東京大学大学院工学系研究科電子工学専攻博士課程修了。同年、NTTヒューマンインタフェース研究所入社。2005年～2006年、スタンフォード大学客員研究員を経て、2008年、NTTメディアインテリジェンス研究所主幹研究員。2009年、特別研究員、2016年、上席特別研究員に就任し、現在に至る。映像符号化の研究開発と国際標準化に従事。博士(工学)。正会員。