

人と機械がともに観る時代の映像伝送 –スケーラブル符号化の高速化技術の開発–

Video Transmission for Both Human and Machine Vision: Fast Encoding Techniques for Scalable Video Coding



岩崎 裕江 (Hiroe IWASAKI, Ph. D.)

東京農工大学

工学研究院 知能情報システム工学部門 教授

(Tokyo University of Agriculture and Technology)

電子情報通信学会、情報処理学会、映像情報メディア学会、IEEE

受賞: IEEE DATE Design Contest Award(2003)、第 57 回前島密賞(2012)、映像情報メディア学会 ハイビジョン・次世代テレビ技術賞(2012)、文部科学大臣賞 科学技術賞(2013)、IEICE Communication Society Best Paper Award(2022)、The 23rd International Conference on Parallel and Distributed Computing, Applications and Technologies Best Paper Award(2022)

研究専門分野: メディア処理、ハードウェアアーキテクチャ

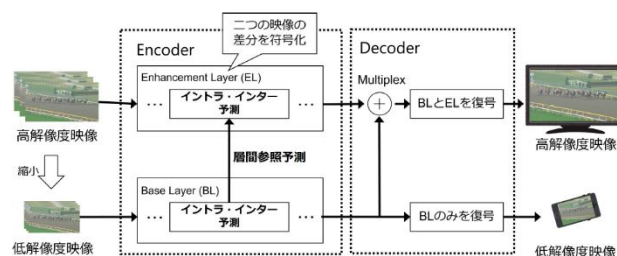
あらまし

映像の受け手は、いまや人だけではない。監視カメラ映像の AI 解析や自動運転支援など、機械が映像を「観て」認識・解析する用途が急速に拡大しており、同一の映像を人の視聴と機械の解析の双方へ効率よく届ける伝送技術が求められている。空間スケーラブル符号化は、低解像度映像（基本レイヤ）と高解像度映像（拡張レイヤ）を階層化して単一のビットストリームで伝送する技術であり、機械解析には基本レイヤのみを、人の視聴には両レイヤを復号するといった柔軟な配信を可能にする。しかし、最新の国際標準規格 VVC でこれを実現する場合、拡張レイヤの符号化に要する演算量が膨大となり、ライブ配信などのリアルタイム用途への適用が困難であった。本研究では、両レイヤのブロック分割構造に高い類似性が存在することに着目し、符号化済みの基本レイヤの情報を用いて拡張レイヤの分割探索を早期に打ち切り、さらに探索候補を絞り込む高速化手法を提案した。評価実験により、圧縮効率の低下を平均約 2%に抑えつつ符号化時間を平均約 50%削減できることを確認し、人と機械がともに観る時代の映像配信基盤としての実用可能性を示した。

1. 研究の目的

映像を観るのは、もはや人だけではなく、監視カメラ映像の自動解析、自動運転を支える周辺認識、工場や店舗における画像検査など、機械すなわち AI が映像を入力として認識・解析を行う用途が急速に広がっている。人が観る映像には高い解像度と画質が求められる一方、機械による解析には必ずしも最高解像度は必要でなく、用途に応じた解像度の映像が効率よく得られれば十分な場合が多い。つまり、これからの映像配信基盤には、同一の映像を「人向けの高品質な映像」と「機械向けの解析に適した映像」の両方の形で、無駄なく届ける仕組みが不可欠である。この要求に応える有力な技術が空間スケーラブル符号化である。これは、低解像度映像を基本レイヤ（Base Layer: BL）、高解像度映像を拡張レイヤ（Enhancement Layer: EL）として階層化し、単一のビットストリームに統合して伝送する方式である。受信側は BL のみを復号すれば低解像度映像を、BL と EL の両方を復号すれば高解像度映像を得られるため、機械解析には BL を、人の視聴には両レイヤを、という柔軟な使い分けが一つのストリームで実現できる（図 1）。しかし、最新の国際標準規格である Versatile Video Coding (VVC) で空間スケーラブル符号化を行うと、符号化に要する演算量が膨大となり、ライブ配信のようなリアルタイム用途への適用が困難になるという課題があった。本研究の目的は、可能な限り、圧縮効率を維持したままこの符号化時間を大幅に削減し、VVC 空間スケーラブル符号化を実際の配信サービスで使える技術へ昇華させることである。

図 1 スケーラブル符号化の概要



2. 研究の背景

インターネットトラフィックに占める動画の割合は増加を続けており、視聴端末もスマートフォン、タブ

人と機械がともに観る時代の映像伝送 –スケーラブル符号化の高速化技術の開発–

Video Transmission for Both Human and Machine Vision: Fast Encoding Techniques for Scalable Video Coding

レット、PC、大画面テレビ、車載機器へと多様化している[1]。従来、こうした多様な端末への配信は、解像度ごとに独立したストリームを複数用意する方式（サイマル配信）が一般的であったが、符号化・保存・配信のすべての工程で複数系列を管理する必要があり、帯域や運用の負荷が大きい。スケーラブル符号化はこの問題への解として古くから研究されており、H. 264/AVC の拡張である SVC、HEVC の拡張である SHVC として標準化されてきた[2]。2020 年に標準化された VVC は従来規格 HEVC に対して同一画質で主観評価にて約 40%程度のビットレート削減を達成し[3]、マルチレイアプロファイルとして空間スケーラビリティも規格に取り込まれている。

一方で VVC は、圧縮効率向上の代償として符号化演算量が大幅に増加しており、条件によっては HEVC の数倍から数十倍の符号化時間を要することが報告されている[3]。その主要因は、符号化の基本単位となるブロックの分割方法(ブロック分割)の自由度拡大にある。VVC では従来の四分分割に加えて二分分割・三分分割が導入され、映像の内容に適した細かな分割が可能になり、効率的に映像を符号化することができるようになった(図 2) 反面、符号化器は膨大な分割候補をレート歪み最適化により逐一評価する必要があり、この探索が符号化時間の大半を占める。空間スケーラブル符号化ではさらに、EL の符号化時に BL を参照する予測処理が加わるため、探索空間は一層拡大する。

符号化の高速化研究はこれまで SHVC 世代を中心に国内外で多数報告されており、下位レイヤの情報を利用した予測モードの早期決定や、ニューラルネットワークによる分割候補の推定などにより大幅な時間削減が示されてきた[4]。しかし、これらは比較的単純な分割構造を前提としており、分割の自由度が格段に高い VVC のスケーラブル符号化にそのまま適用することは難しい。VVC を対象とした空間スケーラブル符号化の高速化について、実装と評価を伴う報告はまだ少なく、知見の蓄積が求められている。また近年、MPEG では機械の解析を前提とした映像符号化 (Video Coding for Machines) の検討も進んでおり、人と機械の双方へ映像を届ける枠組みの重要性は国際的にも高まっている。



図 2 ブロック分割の例

3. 研究の方法

3. 1 高速化のアプローチ

アイデアの核は、「BL と EL は同じ映像を異なる解像度で符号化したものであるから、両者のブロック分割の構造には強い類似性があるはずだ」という点にある。空間スケーラブル符号化では、EL の符号化を始める時点で BL の符号化はすでに完了している。したがって、BL の分割結果や符号化情報を「事前知識」として活用できれば、EL で評価すべき分割候補をあらかじめ絞り込み、無駄な探索を省けると考えた。ただし、やみくもに探索を省けば圧縮効率が犠牲になる。そこで本研究では、(1) 両レイヤの分割にどの程度の類似性があるのかを定量的に確かめ、(2) 探索を省いても安全な条件をデータに基づいて抽出し、(3) その条件が成立する場面に限って探索を省略する、という三段階で手法を構築した。

3. 2 BL と EL の分割類似性の調査

標準動画 9 本[5, 6]を VVC の参照ソフトウェアであり、理想に近い符号化を実現するソフトウェア VTM により符号化し、BL と EL の分割結果をブロックサイズごとに比較する予備調査を行った。予備調査に用いた画像を図 3 に示す。EL の分割が BL と完全に一致する割合は、ブロックサイズにより大きく異なり、頻出する中間サイズのブロックでは 4 割以上が不一致となることが分かった。すなわち「BL の分割をそのまま EL に流用する」単純な方法では圧縮効率の低下が避けられない。一方で、BL で水平（または垂直）方向の分割が選ばれた領域において、EL でそれと直交する方向の分割が選ばれる割合は平均 16%程度と低く、分割の「方

人と機械がともに観る時代の映像伝送 –スケーラブル符号化の高速化技術の開発–

Video Transmission for Both Human and Machine Vision: Fast Encoding Techniques for Scalable Video Coding

向」には明確な類似性が存在することが確認された。この結果から、分割形状の完全な流用ではなく、「明らかに不要な候補の除外」に徹する方針が有効であると判断した。

3. 3 探索省略条件の抽出

探索を省略してよい場面を自動判定するための条件を抽出した。BL から得られる分割の深さ、動きベクトルの統計量、スキップモードの割合、および対象ブロックの輝度の平均・分散など 12 種類の特徴量を定義し、しきい値判定の組合せで表される候補条件を網羅的に生成・評価した。評価には判定の正確さを表す適合率と、取りこぼしの少なさを表す再現率、およびその調和平均である F1 スコアを用い、ブロックサイズごとに F1 スコアが最大となる条件を求めた上で、一定の基準 ($F1 \geq 0.85$) を満たす信頼性の高い条件のみを採用した。この枠組みは解釈可能なルールベース分類器の学習に相当し、どの特徴量がどう働いて探索が省かれるのかを人が確認できる点で、ブラックボックスになりがちな深層学習によるアプローチと異なる。

3. 4 提案手法：早期打ち切りと候補絞り込み

抽出した条件に基づき、提案手法では EL の分割探索に二つの機構を導入した。第一は「早期打ち切り」であり、条件を満たすブロックではそれ以上の分割探索をその場で終了する。第二は「候補絞り込み」であり、BL の分割方向と直交する方向の分割候補を探索対象から除外する。EL の各ブロックの処理時に、当該ブロックと対応する BL 領域から特徴量を算出し、打ち切り条件に合致すれば探索を終了、絞り込み条件に合致すれば直交候補を除外し、いずれにも合致しなければ従来どおりの探索を行う (図 4)。信頼性の高い条件が成立した場面に限って探索を省くことで、圧縮効率への影響を最小限に抑えつつ時間を削減する設計である。



図 3 予備実験に用いたシーケンス

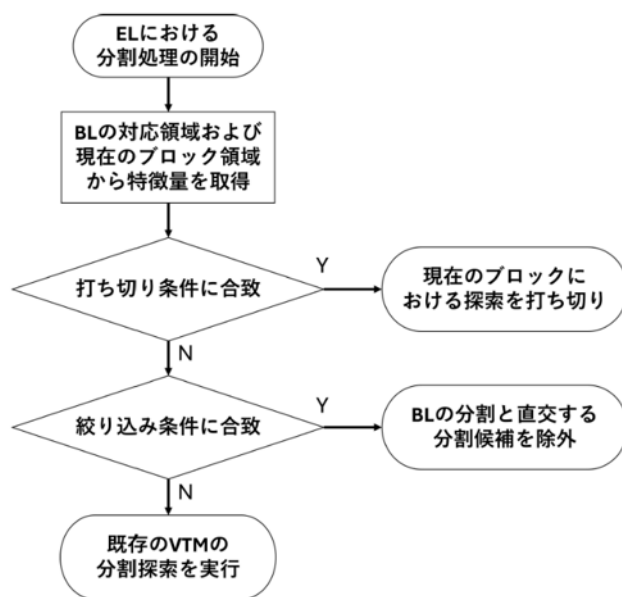


図 4 フローチャート

3. 5 評価結果

条件抽出に用いていない計 9 本[5, 6]の 4K 標準映像を、VTM と提案手法のそれぞれで符号化して比較した。評価に用いた画像を図 5 に示す。その結果、提案手法は圧縮効率の低下 (BD-rate) を平均約 2%に抑えつつ、EL の符号化時間を平均約 50%削減できることを確認した (表 1)。条件抽出に用いた映像群と未知の映像群とで削減率・効率低下に大きな差はなく、本手法が特定の映像に依存しない一定の汎用性を持つことが示され

人と機械がともに観る時代の映像伝送 –スケーラブル符号化の高速化技術の開発–

Video Transmission for Both Human and Machine Vision: Fast Encoding Techniques for Scalable Video Coding

た。一方、画面全体に複雑な模様が広がる映像では、両レイヤの最適な分割方向が一致しにくく効率低下がやや大きくなる傾向も確認しており、判定条件の汎化性能の向上が今後の課題である。本成果は査読付き論文誌および国際会議で発表済みである[7][8]。



図 5 評価に用いたシーケンス

表 1 評価結果

シーケンス	BD-rate[%]	EL における TS[%]
マラソン (フィックス)	2.52	44.3
マラソン (パンダウン)	2.83	41.4
水球 (ゴール)	1.54	42.6
水球 (横スーパー)	1.16	41.0
ドラマ (コーヒー)	1.42	49.8
ドラマ (実家)	1.62	49.7
桂川	3.30	79.1
舞妓	2.30	44.5
気動車	2.68	56.1
平均	2.15	49.8

4. 将来展望

本研究により、VVC 空間スケーラブル符号化の最大の障壁であった符号化時間を約半分にできることが示された。これは、スポーツ中継やイベントのライブ配信のように遅延が許されない場面、あるいはエッジサーバや車載機器のように計算資源が限られる環境において、スケーラブル符号化の採用を現実的な選択肢に引き上げるものである。

応用の観点で特に期待されるのが、冒頭に述べた人と機械の協調である。例えば都市の見守りシステムでは、

平常時は BL のみを伝送して AI が解析を行い、異常検知時には EL を加えて人が高解像度映像で確認する、といった運用が単一ストリームで可能になる。放送・配信分野では、多様な端末への同時配信の帯域と運用コストの削減に直結する。今後、機械向け映像符号化の標準化動向とも歩調を合わせ、機械の解析精度を意識したレイヤ設計へと発展させることも視野に入る。技術面の今後の課題としては、特徴量の拡充や機械学習の導入による判定条件の汎化性能向上、より多様な映像を用いた検証、そしてソフトウェアエンコーダやハードウェア (LSI・FPGA) への実装による実時間動作の実証が挙げられる。数年以内に配信システムへの組み込みを見据え、研究を継続していく所存である。

おわりに

本研究では、VVC に基づく空間スケーラブル符号化を対象として、基本レイヤと拡張レイヤの分割構造の類似性に着目した高速化手法を提案し、圧縮効率の低下を平均約 2%に抑えつつ符号化時間を平均約 50%削減できることを実証した。符号化時間の増大という空間スケーラブル符号化の長年の課題に対し、最新規格 VVC の下で実用的な解を示せたことは、本研究の大きな成果であると考えている。

映像が人と機械の双方に観られる時代は、すでに始まっている。一つのストリームが、ある端末では AI の眼に、別の端末では人の眼に届く—そうした映像配信基盤を支える要素技術として、本研究の成果が活かされることを期待している。本手法の汎化性能の向上と実装検証を進め、研究室で取り組んできた映像符号化 LSI 設計の知見とも結び付けながら、実用化に向けた研究を継続していきたい。

最後に、本研究の遂行にあたり研究助成を賜った一般財団法人テレコム先端技術研究支援センターに深く感謝したい。

参考文献

- [1] Leonardo Peroni and Sergey Gorinsky, “An end-to-end pipeline perspective on video streaming in best-effort networks: A survey and tutorial. ACM Computing Surveys,” Vol. 57, No.

人と機械がともに観る時代の映像伝送 –スケーラブル符号化の高速化技術の開発–

Video Transmission for Both Human and Machine Vision: Fast Encoding Techniques for Scalable Video Coding

- 12, pp. 322:1–322:47, 2025.
- [2] Jill M. Boyce, Yan Ye, Jianle Chen, and Adarsh K. Ramasubramonian, “Overview of shvc: Scalable extensions of the high efficiency video coding standard”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 26, No. 1, pp. 20–34, 2016.
- [3] Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J. Sullivan, and Jens Rainer Ohm, “Overview of the versatile video coding (vvc) standard and its applications. IEEE Transactions on Circuits and Systems for Video Technology,” Vol. 31, No. 10, pp. 3736–3764, 2021.
- [4] Liquan Shen and Guorui Feng, “Content-based adaptive shvc mode decision algorithm,” IEEE Transactions on Multimedia, Vol. 21, No. 11, pp. 2714–2725, 2019.
- [5] 映像情報メディア学会電波産業会, “超高精細・広色域標準動画像— A シリーズ,” 2016.
- [6] 映像情報メディア学会電波産業会, “超高精細・広色域標準動画像— B シリーズ,” 2017.
- [7] Yo Koura, Naoya Niwa, Hiroe Iwasaki, “Efficient Block Partitioning Method for Spatial Scalable Encoding in Versatile Video Coding,” International Journal of Technology, Vol. 17, No. 1, pp. 18-26, 2026 Doi: 10.14716/ijtech.v17i1.8218.
- [8] Yo Koura, Naoya Niwa, Hiroe Iwasaki, “Efficient Block Partitioning Method for Spatial Scalable Encoding in the Versatile Video Coding,” Quality in Research (QiR) 2025, Yogyakarta, Indonesia, October 2025.

この研究は、令和4年度SCAT研究助成の対象として採用され、令和5～7年度に実施されたものです。