

知っておきたいキーワード

VCM (Video Coding for Machines)

飯田 健太[†][†] 日本電気株式会社

"VCM (Video Coding for Machines)" by Kenta Iida (NEC Corporation, Kanagawa)

キーワード：映像符号化, VCM

まえがき

AI (Artificial Intelligence) 技術の急速な発展に伴い、映像はヒトが視聴するためだけでなく、機械による解析処理のためにも利用されるようになりました。映像はデータ量が大きいため、映像符号化技術を用いて圧縮された形式で伝送されるのが一般的です。しかし、従来の映像符号化方式は、ヒトの視覚特性に基づいて設計されているため、機械処理には冗長な情報が符号化

データに含まれます。例えば物体検出を行う時、検出対象の物体が含まれない背景領域の情報は、多くの場合で不要になります。また、物体追跡を行う際には、連続するピクチャ間で変化が少ない場合に、その一部のピクチャを破棄しても追跡精度には影響ないことが予想されます。このように、重要な情報を削減する最適化を行えば、機械による解析処理の精度を維持しつつ、伝送するデータ量を削減可能です。今後、機械処理用途での映像伝送の需

要が増加し、大きなネットワーク負荷となることが予想されるため、機械処理向けに設計された映像符号化方式が必要とされています。

このような背景のもと、現在MPEG (Moving Picture Experts Group) で標準化が進められている機械処理向けの新しい映像符号化方式がVCM (Video Coding for Machines)です。本稿では、MPEGにおけるVCMの標準化動向とVCM規格の概要について紹介します。

VCMのユースケースと特徴

MPEGにおける機械処理向けの映像符号化技術の検討は2019年7月より開始されており、図1に示すアーキテクチャを想定して検討が進められました¹⁾。このアーキテクチャでは、送信側は、映像そのものや映像から抽出した特徴マップの情報を圧縮符号化することにより、ビットストリームを生成および伝送します。一方、受信側ではその復号結果を使用して物体検出等の機械処理を実行します。また、ヒトが復号映像を視聴することも想定されます。

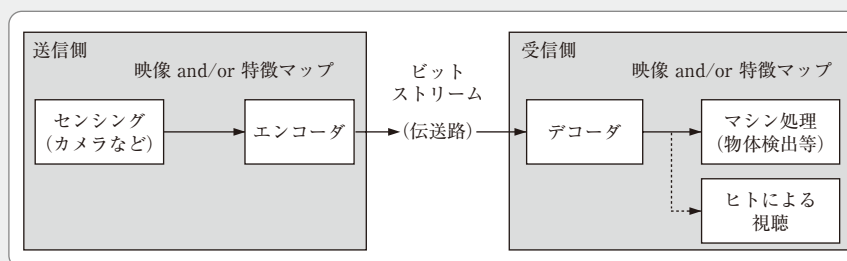


図1 機械処理向け映像符号化のアーキテクチャ

MPEGでは現在、映像そのものを符号化する方式をVCM、映像から抽出した特徴マップを符号化する方式をFCM (Feature Coding for Machines)

と方式を分けて、標準化が進められています。本稿では前者のVCMについて紹介します。📄

VCMのユースケースとしてはスマートシティやスマートインダストリなどが想定されており、例えば図2に示す、監視システムの例が挙げられています。このユースケースでは、多くのネットワークカメラから伝送されてきたビットストリームは、サーバで復号されて、機械による解析処理が復号映像を用いて実行されます。また、その処理結果の妥当性を検証するために、ヒトによる映像の確認が行われます。映像そのものを伝送することにより、特徴量のみを送った場合には実現できない「ヒトによる検証」を可能とすることが、VCMの特徴と言えます。

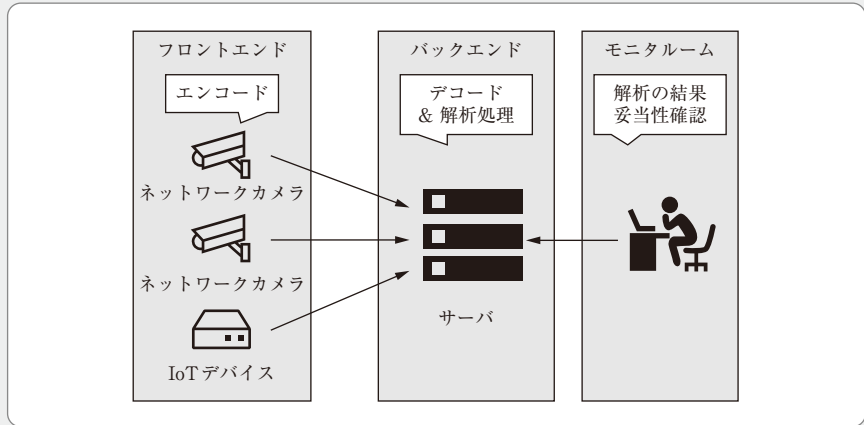


図2 ユースケースの例 (監視システム)

MPEGにおけるVCM規格の標準化動向

MPEGにおけるVCM規格は、2022年4月から3年以上かけてその標準化が進められ、2026年1月に国際規格案 (DIS: Draft International Standard) が発行されました。今後、国際規格の第1版が、2027年中にISO/IEC 23888-2として発行される見込みです²⁾。

VCM規格では、図1中のエンコーダが生成するビットストリームの構造 (シンタクス) や、デコーダにおける復号処理の手順が規定されます。併せて、このVCM規格に準拠したエンコーダおよびデコーダの参照ソフトウェア (VCMRS: VCM Reference Software) が開発されています。このVCMRSの性能は、共通テスト条件 (CTC:

Common Test Conditions)³⁾の下で評価されています。VCM規格は、機械処理向けの符号化方式であることから、従来の符号化方式のような復号映像の画質指標 (PSNR: Peak Signal-to-Noise Ratioなど) ではなく、復号映像を用いて機械処理を実行した際の解析精度 (mAP: mean Average Precision やMOTA: Multi-Object Tracking Accuracyなど) を主な指標として評価が行われます。表1は、従来の映像

符号化方式であるH.266/VVCの参照ソフトウェア (VTM)⁴⁾を用いた場合に対するVCMRSの性能を示します。ここで、各値は、同等のタスク処理の精度における符号量の削減率を示し、RAは多少の遅延を許容して符号化効率を高める条件、LDは最も少ない遅延を想定した条件です。この表から、どの条件においても60%から80%の符号量が削減されていることが確認できます。

表1 H.266/VVCの参照ソフトウェア (VTM) に対するVCMの参照ソフトウェア (VCMRS) の性能 (符号量削減率)

	RA (Random access)	LD (Low delay)
SFU-HW dataset (物体検出の評価用データセット)	- 67.81%	- 62.04%
TVD (Tencent Video Dataset) (物体追跡の評価用データセット)	- 81.62%	- 68.34%

VCMエンコーダ/デコーダの概要

図3に、VCMエンコーダおよびデコーダの構成を示します。VCMエンコーダは、入力映像から機械処理に冗長な情報を削除するための事前処理を適用して、事前処理後の映像を内部エンコーダで符号化します。一方、VCMデコーダは、VCMビットスト

リーム中の制御情報に基づいて、内部デコーダで復号された映像へ復元処理を適用します。エンコーダでの事前処理と統合した復号処理は、内部エンコーダでの符号化結果と復元処理の制御パラメータをビットストリームに多重化して伝送することにより、可能となります。なお、内部エンコーダおよびデコーダには、H.266/VVC等の既存の映像符号化方式が利用可能です。

VCM規格は、機械処理に必要な情報をVCMデコーダで復元できるようにしつつ、事前処理で冗長な情報を削減することで、既存の映像符号化方式よりも高い符号化性能を達成することが特徴です。そこで、その性能の実現に重要である、事前処理および復元処理の内容について以下で説明します。



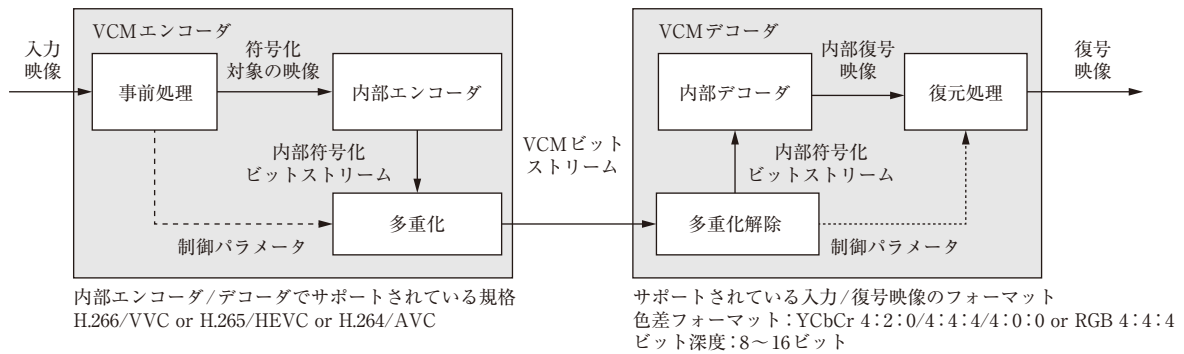


図3 VCMエンコーダおよびVCMデコーダの構成

エンコーダにおける事前処理

VCM規格では、エンコーダが生成するビットストリームの構造は規定されますが、それを生成するための処理内容は規定されません。これにより、エンコーダ実装の自由度を残しながら相互接続性を担保しています。ここでは、参照ソフトウェアであるVCMRSにおける事前処理の処理内容の例について紹介します(図4)。この事前処理は、以下の3つの処理に分類できます。

- (1) シーケンスに含まれるピクチャの枚数を削減する処理(時間情報を削減する処理): Temporal downsampling
- (2) ピクチャに含まれる画素数を削減する処理(空間情報を削減する処理): Spatial retargetingおよびSpatial resampling
- (3) 1画素の持つ情報を削減する処理(画素情報を削減する処理): Bit depth truncationおよびChroma removal

このうち、Spatial retargeting処理⁵⁾やTemporal downsampling処理⁶⁾、Bit depth truncation処理が、機械処

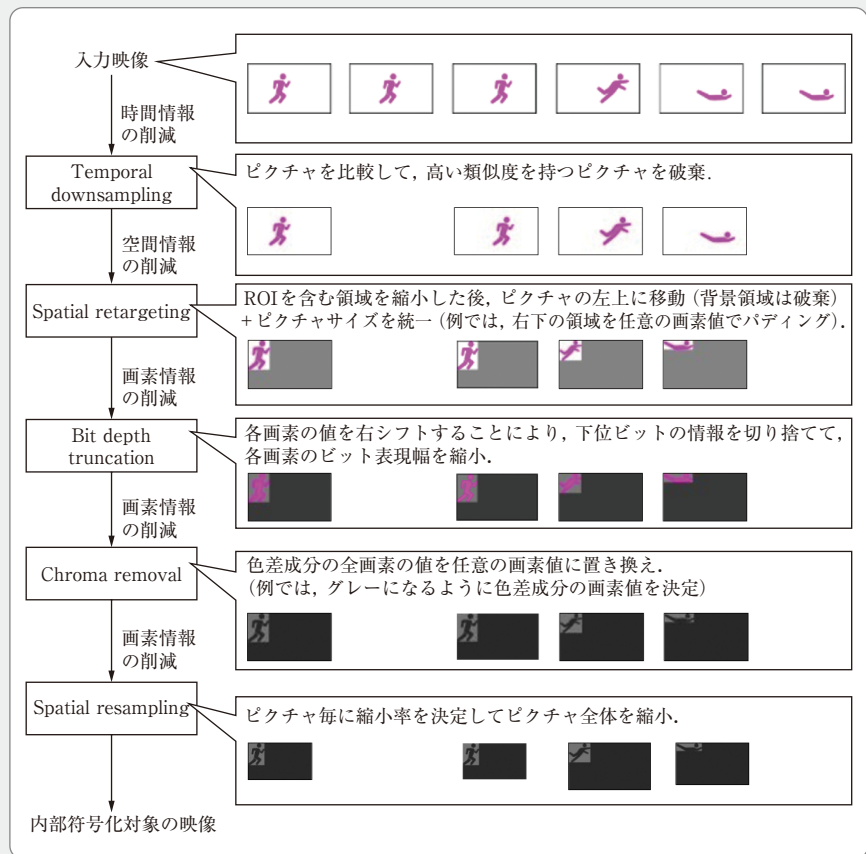


図4 VCMRSにおける事前処理のブロック図

理タスクの精度の低下を抑えつつ、符号量を大幅に削減させることが確認されています。

デコーダにおける復元処理

一方、デコーダにおける復元処理の手順は、VCM規格で規定されており、エンコーダから指定された映像フォーマットへ復元するための処理が実行されます(図5)。なお、性能評価の結果から、エンコーダでの事前処理と対応する復元処理は非対称になっています。

この復元処理では、ニューラルネットワーク(NN)を用いた3つの処理が規定されています。NNの構造(トポロジー)は規定された構造のみ使用可能ですが、重みなどのパラメータはユーザによって更新可能となる予定で、実装による改善の余地が残されています。

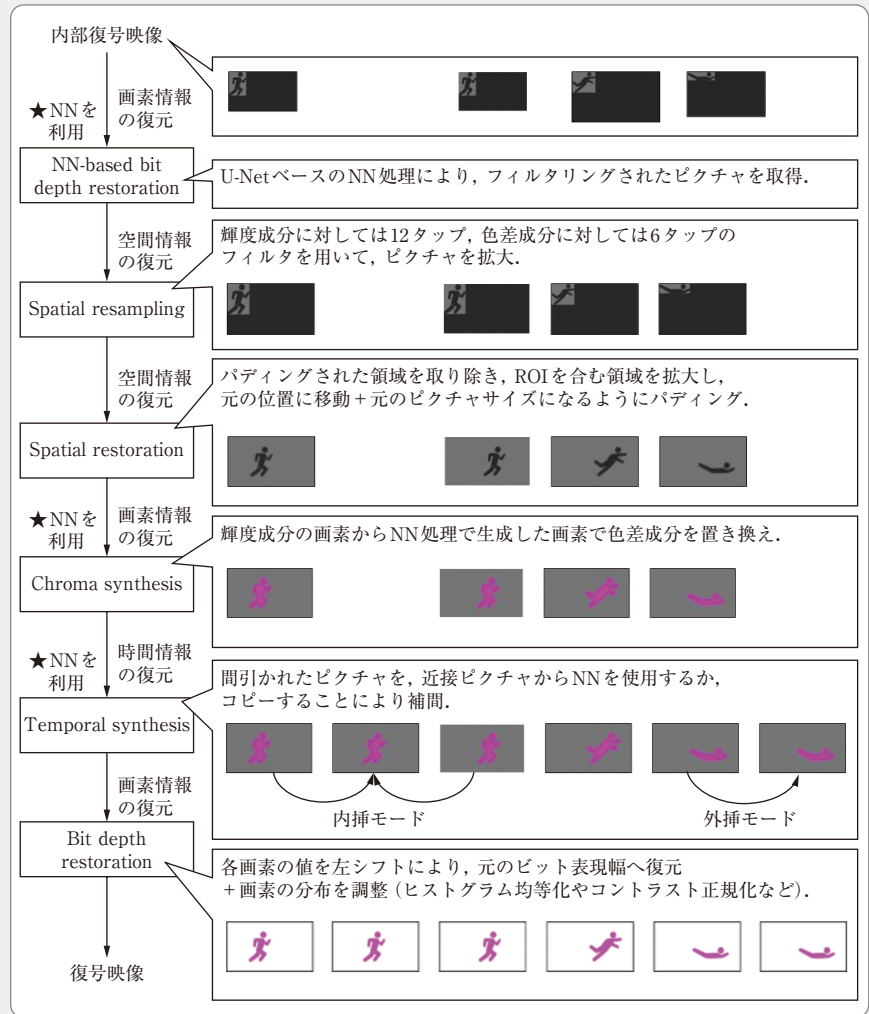


図5 VCM規格における復元処理のブロック図

むすび

本稿では、MPEGにおいて標準化が進められている機械向けの映像符号化方式VCMについて、その標準化動向

や技術の概要を紹介しました。現在、映像データはネットワークトラフィックの大部分を占めていますが、機械処理用途での映像伝送の需要が今後さらに増加することで、ネットワーク負荷

は一層高まると予想されます。こうした状況の中で、VCM規格は、機械処理に使用される映像データを効率的に伝送するための重要な技術として期待されています。(2026年2月13日受付)

参考文献

- 1) ISO/IEC JTC 1/SC 29/ WG 2: "Use cases and requirements for Video Coding for Machines", ISO/IEC JTC 1/SC 29/WG 02 N43 (2021)
- 2) ISO/IEC JTC 1/SC 29: "ISO/IEC DIS 23888-2 Information technology - Artificial intelligence for multimedia Part 2: Video coding for machines" (2026), <https://www.iso.org/standard/88879.html>
- 3) ISO/IEC JTC 1/SC 29/ WG 4: "Common test conditions for video coding for machines", ISO/IEC JTC 1/SC 29/ WG 04 N730 (2025), https://www.mpeg.org/wp-content/uploads/mpeg_meetings/152_Geneva/w25760.zip
- 4) "VVCSofware_VTM", https://vcgit.hhi.fraunhofer.de/jvet/VVCSofware_VTM
- 5) S. Rózek, et al.: "Video coding for machines using region-of-interest-based retargeting", EURASIP Journal on Image and Video Processing, pp.1-17 (2025)
- 6) Z. Huang, et al.: "Real-time intermediate flow estimation for video frame interpolation", ECCV 2022



飯田 健太

2018年、首都大学東京大学院システムデザイン研究科博士前期課程修了。同年、日本電気(株)入社。2022年、東京都立大学大学院システムデザイン研究科博士後期課程修了。日本電気(株)にて、ISO/IECおよびITU-Tの国際規格に準拠した映像コーデックのソフトウェア開発に従事するとともに、2023年より、MPEGにおける標準化活動に参画。工学博士。