

修士論文概要書

Master's Thesis Summary

Date of submission: 01/25/2021 (MM/DD/YYYY)

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	堀 隼也	指導員 Advisor	渡辺 裕 印 Seal
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	CD 5119F080-8-		
研究題目 Title	VVC 符号化映像の超解像を用いた解像度混在映像伝送方式 A Mixed-Resolution Video Transmission System Using Super-Resolution of VVC Coded Video				

1. まえがき

近年, 超解像を用いた映像伝送方式が提案されている. ビデオを一度ダウンスケールし, 符号化伝送し, 復号後に超解像などの補間技術を用いてビデオを元のサイズに戻す(図1). この手法を用いることで, 圧縮する情報量を削減し, 低ビットレート帯において, 符号化効率を改善できる. 一方, 単純にビデオを縮小し, 超解像処理を施すだけでは符号化ビデオの品質変動や時間的相関を利用できない. そこで, これらを活用するため, 本研究では解像度を意図的に変更して符号化を行う混在解像度映像伝送方式を提案する.

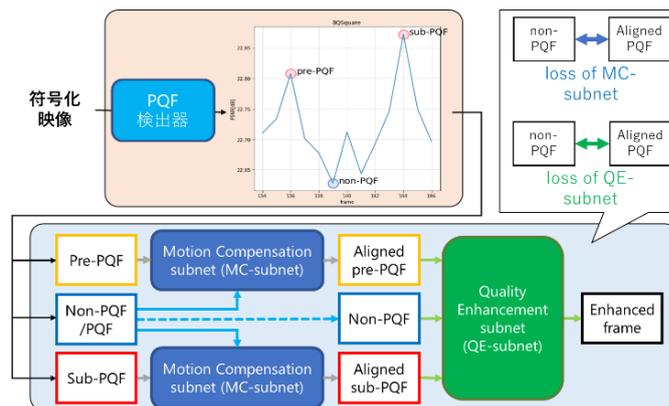


図 2 MFQE の構造

3. 提案手法

提案手法では, RPR を用いて PQF をオリジナルサイズ, non-PQF を半分のサイズで符号化する. 復号後, 元のサイズのフレームを参照し, 縮小された non-PQF の品質改善を MFQE で行う(図3). Non-PQF を縮小することで, 符号化する情報量を削減する. また, PQF をオリジナルサイズで保持し, MFQE の性能を向上させる. 以上のように, 解像度を意図的に変化させ, 符号化効率の改善を図る. 本研究では, 表 1 に示す三つのケースの RD 曲線を比較し, 提案手法を評価する.

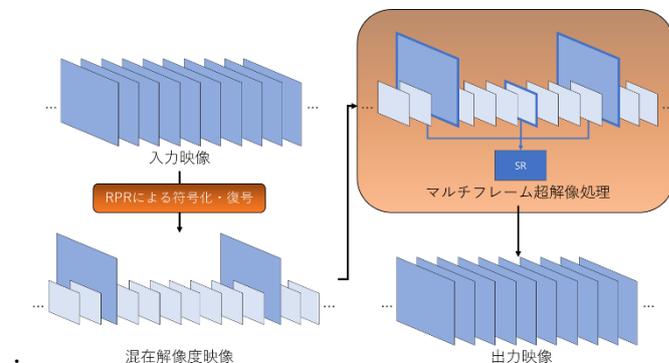


図 3 提案手法の概略図

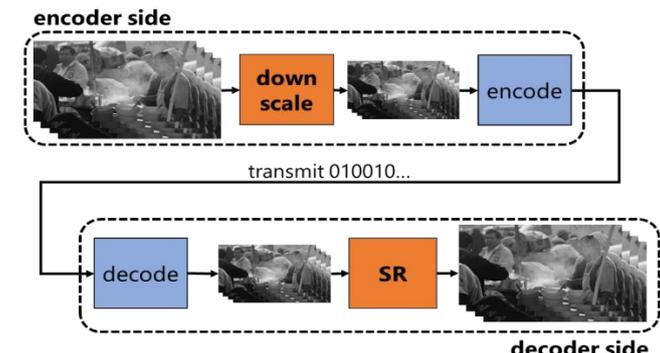


図 1 超解像を用いた映像伝送方式の概略図

2. 関連研究

2.1. Reference Picture Resampling (RPR)

最新動画圧縮規格 Versatile Video Coding (VVC) では異なる解像度間のフレーム参照を可能とする Reference Picture Resampling (RPR)を提供している. RPR により, シーケンス内でフレーム解像度を適宜変更させて符号化することができる.

2.2. Multi Frame Quality Enhancement (MFQE)

符号化映像では定期的に高品質フレームが出現する. この高品質フレームを参照し, 他フレームの品質を高める手法が Multi Frame Quality Enhancement (MFQE)である. 図 2 に MFQE の構造を示す. MFQE は符号化映像中の品質がピークとなるフレーム Peak Quality Frame(PQF)を検出する. その後, 動き補償サブネットで, PQF を, 品質向上を行う non-PQF に位置合わせする. そして, 位置合わせした前後の PQF と non-PQF を深度方向に結合し, 品質改善サブネットに入力し, non-PQF の品質改善を行う.

表 1 比較を行う三つのケース

ケース	符号化条件 (解像度)
Case1	オリジナルサイズ
Case2	縦横 2 分の 1 に縮小
Case3	PQF : オリジナルサイズ non-PQF : 縦横 2 分の 1 に縮小

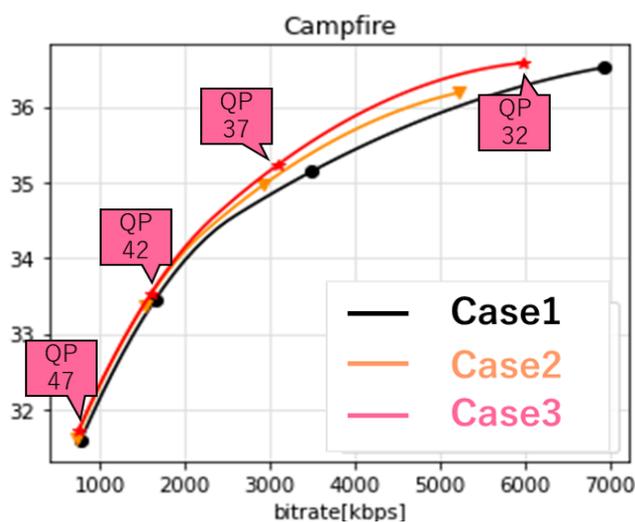
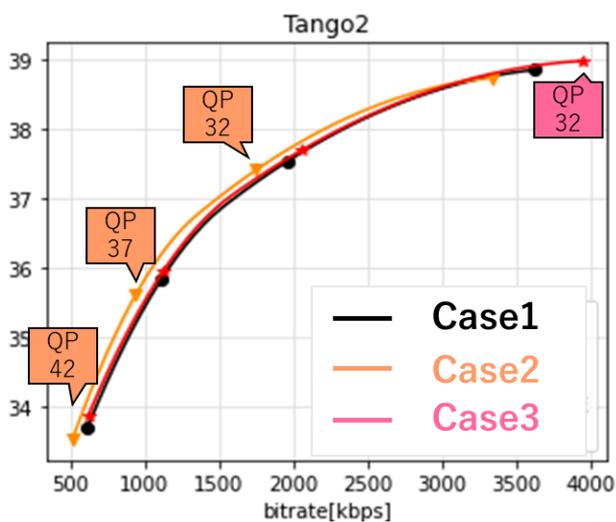


図 6 各ケースの RD 曲線の比較(“Tango2”, “Campfire”)

4. 評価実験

評価実験では、提案手法の有効性を確認するために、符号化効率の評価を行う。提案手法は 4K シーケンスに対し有効であることから、テストには JVET Common Test Conditions (CTC) for SDR[2]の内、4K 解像度の Class A1, A2 の六つのシーケンスを用いる。各シーケンスを表 1 の三つのケースで符号化し、RD 曲線を比較する。図 6 に各ケースの RD 曲線と、各測定点で最も効果が高かったケースを示す。RD 曲線の比較では、六つのシーケンスの内、“Tango2”, “FoodMarket4”, “Campfire”, “ParkRunning3”の 4 シーケンスで、Case3 は右側 2 点で Case1 よりも符号化効率が向上した。また、全シーケンスの超低ビットレート帯に対し、Case2 は Case1 より符号化効率が向上した。以上より、Case3 は低ビットレート帯で、Case2 は超低ビットレート帯で有効だといえる。

次に、Case2, 3 で符号化効率が改善した原因を考察する。Case3 では、“CatRobot1”, “DaylightRoad2”は劣化の影響が大きく、MFQE による復元が十分にできなかった。図 7 は Case1 と Case3 入力・出力画像の比較を示す。Case3 では、信号機の点灯部のディテールは失われ、MFQE による復元ができなかった。一方、焚火の映像ではエッジ部が残り、MFQE は有効に作用した。このことから、Case3 はシーケンスにエッジが残る場合に、元画像サイズの PQF を用いて、符号化効率を改善したと考えられる。Case2 ではブロックノイズを解消した。図 8 は Case1, Case2 出力画像の比較である。Case1 ではブロックノイズが生じたが、Case2 ではそれらが解消された。これは、低解像度で符号化することで、低周波成分に多くの情報を割り振れたためだと考えられる。

5. むすび

本論文では、異なる解像度間のフレーム参照を可能とする RPR と複数フレームを参照し符号化ビデオの品質改善を行う MFQE を組み合わせた VVC 符号化映像のための解像度混在映像伝送方式を提案した。RD

曲線を比較し、提案手法は 4K シーケンスに対し、通常サイズで符号化する場合よりも低ビットレート帯に対し符号化効率が改善することを確認した。また、全フレームを縮小する場合は超低ビットレート帯に対し、符号化効率が改善することを確認した。以上より、ビットレート帯に応じ、各手法を適用することが有効である。



図 7 Case1 と Case3 入力画像・出力画像の比較



図 8 Case1, Case2 出力画像の比較

参考文献

- [1]. Z. Guan, Q. Xing, M. Xu, R. Yang, T. Liu and Z. Wang: “MFQE 2.0: a new approach for multi-frame quality enhancement on compressed video”, Trans. TPAMI, 2019.
- [2]. F. Bossen, J. Boyce, K. Sühring, X. Li and V. Seregin: “JVET common test conditions and software reference configurations for SDR video” Joint Video Experts Team

2020 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

VVC 符号化映像の超解像を用いた解像度混在映像伝送方式

A Mixed-Resolution Video Transmission System Using Super-Resolution of VVC Coded Video

堀 隼也

(5119F080-8)

提出日：2021.01.25

指導教員：渡辺 裕 印

研究指導名：オーディオビジュアル情報処理研究

目次

第1章	序論	3
1.1	はじめに	3
1.2	研究目的	3
1.3	本論文の構成	4
第2章	関連研究	5
2.1	まえがき	5
2.2	Versatile Video Coding (VVC)	5
2.2.1	技術概要	5
2.2.2	Reference Picture Resampling (RPR)	5
2.3	超解像技術	6
2.3.1	超解像技術の概要	6
2.3.2	深層学習を用いた超解像技術	7
2.3.3	超解像技術の分類	8
2.3.3.1	単体画像超解像	8
2.3.3.2	マルチフレーム超解像	9
2.4	Multi Frame Quality Enhancement (MFQE)	9
2.4.1	技術概要	9
2.4.2	符号化ビデオの品質変動	9
2.4.3	構造	10
2.4.3.1	構造の概要	10
2.4.3.2	動き補償サブネットの構造	11
2.4.3.3	品質改善サブネットの構造	11
2.4.3.4	MFQE の End-to-End での学習	12
2.5	むすび	14
第3章	提案手法	15
3.1	概要	15
3.2	RPR と MFQE を組み合わせた映像伝送方式	16
3.3	提案手法と従来手法の比較	17
3.3.1	提案手法	17
3.3.2	全フレームを縮小する場合	17
3.3.3	提案手法と全フレームを縮小する場合の比較	17
3.4	まとめ	17
第4章	予備実験：VVC におけるダウンスケーリングを含む映像伝送方式の有効性	19

4.1	概要	19
4.2	評価指標	19
4.2.1	RD 曲線	19
4.2.2	PSNR	20
4.3	テストシーケンス	20
4.4	実験手順	22
4.5	結果および考察	22
4.6	まとめ	22
第 5 章	評価実験：超解像を含む解像度混在映像伝送方式の品質評価	24
5.1	概要	24
5.2	実験手順	24
5.3	MFQE の学習	25
5.3.1	学習の概要	25
5.3.2	学習用データセット	25
5.3.3	PQF の学習	25
5.4	符号化効率の評価	26
5.4.1	概要	26
5.4.2	結果	26
5.4.3	考察：Case3（提案手法）の特性	30
5.4.4	考察：Case2（全フレームを縮小する場合）の特性	32
5.4.5	Case2 と Case3 の組み合わせによる符号化効率の改善	33
5.5	MFQE 性能の評価	34
5.5.1	概要	34
5.5.2	結果 1 ・考察	34
5.5.3	結果 2 ・考察	36
5.6	まとめ	36
第 6 章	結論	38
6.1	結論	38
6.2	今後の課題	38
	謝辞	39
	参考文献	40
	図一覧	43
	表一覧	44
	研究業績	45

第1章 序論

1.1 はじめに

近年、4K (3840x2160 画素) や 8K (7680x4320 画素) といった高精細映像の普及が進んでいる。2018年12月には、BS、CS放送において4K・8K実用放送である「新4K8K衛星放送」が開始された[1]。また、各種動画配信サービスやPlayStation5などの家庭用ゲーム機、パソコンを使ったビデオゲームでも、4K・8K解像度への対応が進み、高精細映像が身近なものとなっている。

4K、8K解像度の映像は、Full High Definition (フルHD: 1920x1080画素) の映像に比べ、それぞれ4倍、16倍の画素数を持ち、被写体に立体感が生まれるなど映像に高い臨場感を持たせることが可能となる[2]。一方、これらの高精細映像では従来の映像と比べデータ量が増加する。8K映像の場合、空間解像度だけでもフルHD映像の16倍のデータ量となる。そのため、限られた伝送路で、大量のデータを送るには大幅なビデオの圧縮が必要とされる。

総務省では2015年より地上波での4K・8K放送を実現する地上放送高度化方式の技術的実現可能性について研究を進めている[3]。世界に先行して8K Ultra High Definition Television (8K UHDTV) での放送を開始したNHK BS8Kでは、最大100Mbpsの伝送容量を確保している[4]。一方、地上放送で8K映像を送信するには、最新の変調方式の利用、周波数帯域の割り当ての工夫をしても伝送速度を30.9Mbps程度しか確保できない[3]。したがって、最新の映像符号化方式であるVersatile Video Coding (VVC) [5]を使用しても十分な品質の映像を伝送することができない。また、帯域制限のあるインターネットでの映像配信なども想定し、低ビットレート下における映像伝送の効率化が必要である。

1.2 研究目的

映像伝送の効率化に対応する手法として、超解像を用いた映像伝送方式が提案されている[6][7][8]。超解像とは入力された低解像度画像を高解像度化する手法である。超解像を用いた映像伝送方式では、映像を縮小して符号化、伝送、復号する。そして超解像処理を用

いて、復号されたビデオを元のサイズに戻す。超解像処理により、映像品質を保ちつつ、元の解像度に復元することができれば、符号化による情報量を減らすことが可能となる。

一方で従来の超解像を用いた映像伝送方式には二つの問題点がある。一つ目の問題点は時間的相関を活用できない点である。従来の手法では現在フレームのみを超解像に使用しており、前後のフレームの時間的相関を用いることができない。二つ目の問題点は、符号化映像の品質変動を考慮できない点である。符号化された映像は階層参照構造により、定期的な品質変動が存在する。しかし、従来手法ではこの品質変動を考慮できない。そのため、時間的相関と符号化ビデオの品質を考慮した超解像を含む映像伝送方式が必要とされる。

本研究では、時間的相関と符号化ビデオの品質を考慮した映像伝送方式を実現するために、VVCの新機能である Reference Picture Resampling (RPR) とマルチフレーム超解像を用いた VVC 符号化映像のための解像度混在映像伝送方式を提案する。具体的には、RPR を用いて、定期的にフレームを元画像サイズで符号化、それ以外のフレームを縮小サイズで符号化し、復号後、さらにマルチフレーム超解像を用いて、元画像サイズのフレームを参照し、縮小されたフレーム品質を向上させる。本研究では、映像を通常サイズで符号化した場合、全てのフレームを縮小サイズで符号化した場合、提案手法で符号化した場合の三つを RD 曲線 (Rate-Distortion Curve) で評価し、提案手法の有効性を示す。また、元画像サイズフレームの参照有無による超解像性能を比較し、提案手法により元画像サイズフレームの情報が有効に利用されていることを確認する。

1.3 本論文の構成

本論文の構成を以下に示す。

- 第 1 章 本章であり、本研究の背景、目的について述べている。
- 第 2 章 本研究で用いる関連技術について述べる。
- 第 3 章 本研究の提案手法を述べる。
- 第 4 章 本研究の予備実験及び結果を述べる。
- 第 5 章 本研究の実験及び結果を述べる。
- 第 6 章 本研究の考察を述べる。

第2章 関連研究

2.1 まえがき

第2章では、関連技術である VVC, 超解像技術, MFQE について説明する。

2.2 Versatile Video Coding (VVC)

2.2.1 技術概要

Versatile Video Coding (VVC) は Joint Video Experts Team (JVET) によって策定された最新の動画圧縮規格の一つである。一般に動画は非圧縮の状態では非常に多くのデータ量を要する。そのため、圧縮をすることにより、データ量を減らす。その圧縮、展開に係るルールを決めているのが動画圧縮規格である。

2020年10月に最終バージョンが決定された VVC は 2013年に承認された High Efficiency Video Coding (HEVC) の後継にあたる。VVC は HEVC と比較して同等の視覚品質が得られる映像に対し、約 30~50%符号化効率を改善する。また、360° 映像や 16K までの高解像度映像に対応しており、用途の広さが特徴とされる。このように、様々な機能が追加された VVC だが、Reference Picture Resampling (RPR) というピクチャ参照での新機能もサポートされた。

2.2.2 Reference Picture Resampling (RPR)

RPR は参照されるフレームのサイズを予測フレームのサイズと一致するように変更するアルゴリズムである[9]。従来の動画圧縮規格では、フレーム間予測において、異なる解像度間でのフレーム参照を行うことができない。しかし、RPR を用いることにより、図 2.1 のような異なる解像度間でのフレーム参照が可能となる。また、異なるピクチャを参照可能なため、図 2.2 に示すように、シーケンス内でビデオの解像度を適宜変化させて符号化することが可能となる。

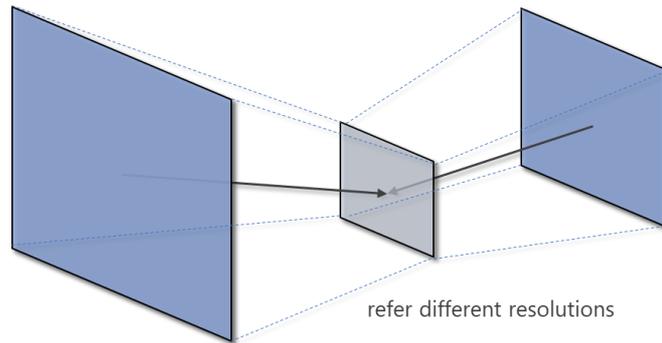


図 2.1 RPR による異なる解像度間でのピクチャ参照の例

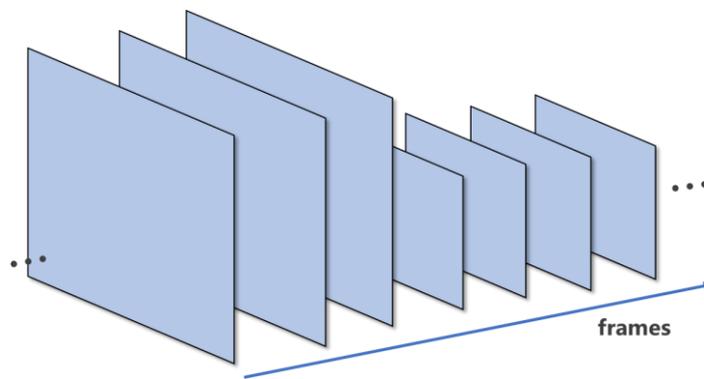


図 2.2 RPR による符号化時のフレーム解像度を切り替え

2.3 超解像技術

2.3.1 超解像技術の概要

超解像技術は入力された低解像度画像を高解像度化する画像補間技術の一種である。図 2.3 は超解像処理による低解像度画像の高解像度化の概念図である。従来の画像補間技術である BICUBIC 法や Lanczos 法などの線形補間法では、生成画素の画素値をその画素の周囲の画素値を線形的に計算することで求める。これらの線形補間手法は計算量が少ない反面、線形補間は周囲の画素値をサンプリングするローパスフィルタであるため、高周波成分を復元することができない。一方、超解像技術では、低解像度のパッチと高解像度のパッチ間の写像を学習する学習型（辞書型）超解像、そして、時系列画像や対象画像内から入力画像の各画素の対応点を探索・参照する再構成型超解像を用いることにより高周波成分を復元することができる。また、近年では、深層学習の発展に伴い、深層学習を用いた Convolutional Neural Network (CNN) ベースの超解像技術が高い性能を示している。

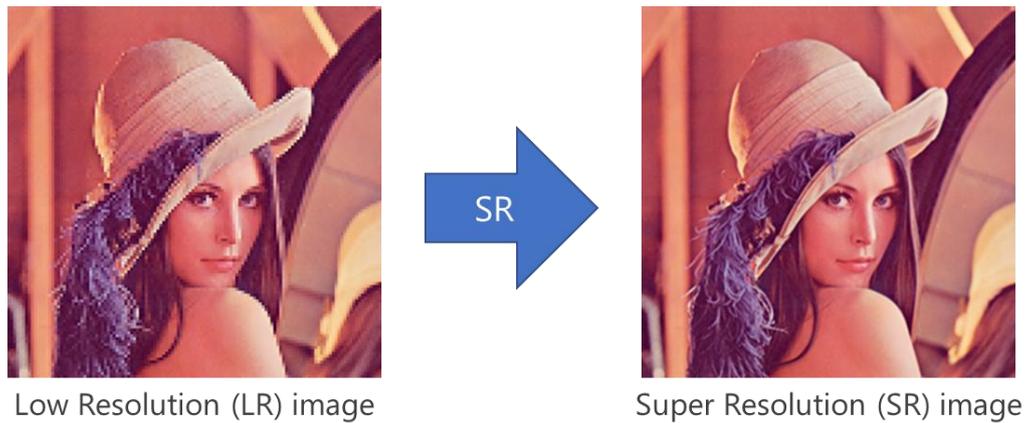


図 2.3 低解像画像（入力）と超解像画像の例

2.3.2 深層学習を用いた超解像技術

CNNは人間の神経モデルを模したニューラルネットワークというアルゴリズムに畳み込み層（Convolution Layer）を追加したアルゴリズムである。畳み込み層は図 2.4 のように、重みをもったフィルタを小領域ごとにスライドし、入力されたテンソルとフィルタとの内積を計算する。複数枚のフィルタを使い、内積を求め、特徴マップと呼ばれる 3 次元のテンソルを取得する。これが畳み込み層の役割である。この畳み込み層、非線形関数である活性化関数、条件を満たすベクトル値を一定領域ごとに一つサブサンプリングするプーリング層を図 2.5 に示すように重ねることにより、入力画像の特徴を詳細に抽出できる[10]。抽出した特徴量は全結合層などにより目的のデータサイズに合わせたサイズに出力される。

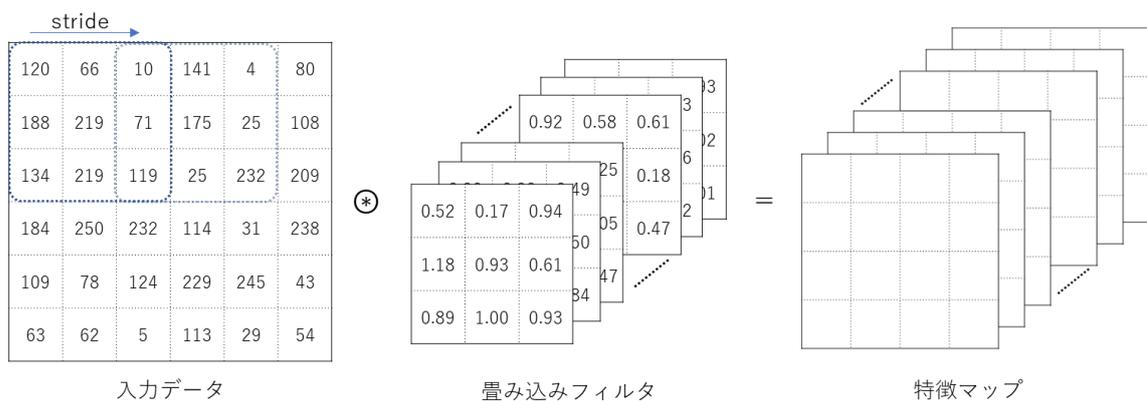


図 2.4 畳み込みによる特徴量の抽出のイメージ

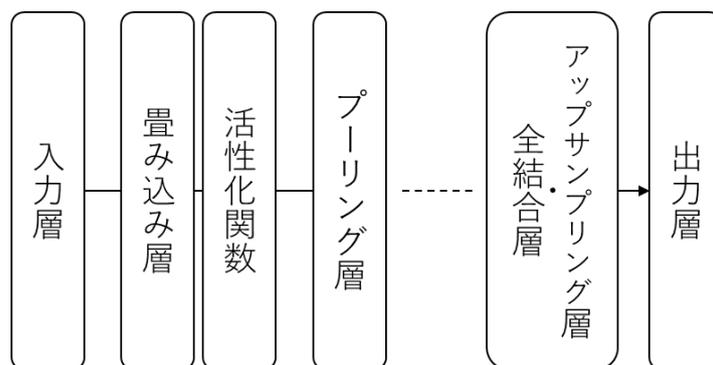


図 2.5 CNN の概要図

CNN の各フィルタのパラメタは、教師データと出力データの誤差を最小化するようにフィルタの各パラメタを変動させる誤差逆伝搬法を用いることで自動的に設計できる。具体的には、CNN により得られた出力データと、理想の出力データである教師データとの損失（超解像タスクの場合、二乗誤差や絶対誤差など）を計算する。その後、損失を最小化する方向に CNN 内のフィルタのパラメタを微分し、パラメタを微分方向に変化させる。この操作を損失値が収束するまで繰り返すことにより、教師画像に近い出力を得られるフィルタを自動で設計できる[11]。一般に、誤差逆伝搬によるフィルタの設計を学習と呼び、学習によって得られたフィルタ目的の出力を得る操作を推論と呼ぶ。

超解像タスクに CNN を用いる場合、抽出した特徴マップを転置畳み込みや Subpixel Convolution[12]による Upsampling 層を用いて出力サイズにマッピングを行い、超解像画像を生成する。超解像処理を行う CNN の学習には、大量の低解像度画像と高解像度画像のペアを用意し、低解像度から高解像度画像への写像を学習させる。以上の CNN を用いた超解像は従来の辞書ベースの超解像手法や自己合同性を利用した超解像手法と比較して、フィルタ設計を最適化できるため、高い性能を示している。

2.3.3 超解像技術の分類

超解像技術は使用するフレーム数によって単体画像超解像（SISR: Single Image Super-Resolution）とマルチフレーム超解像（MFSR: Multi Frame Super-Resolution）に分類される。

2.3.3.1 単体画像超解像

単体画像超解像は、一枚の低解像度画像から高解像度画像を生成する技術である。2015年に発表された SRCNN[13]では、CNN アプローチの超解像手法を提案し、3層の CNN を

用いた超解像処理に成功した。その後、入力や中間での出力値を後段に加算する残差構造を採用し、畳み込み層の数を数十に増やした VDSR[14]や EDSR[15]が発表された。残差構造をとることにより、層を深くすると学習が収束しなくなる勾配爆発、勾配消失の問題を解決し、高い超解像性能を発揮した。

2.3.3.2 マルチフレーム超解像

マルチフレーム超解像は、複数フレームを参照して、高解像度画像を生成する技術であり、主にビデオに対し適用される。TDAN[16]や EDVR[17]では前後のフレームを、超解像を行う対象フレームに位置合わせをする。その後、対象フレームと位置合わせを行った前後フレームの複数フレームを用いて超解像処理を行う。また、FRVSR[18]や TecoGAN[19]では再帰構造を採用している。前のフレームの超解像処理による出力画像を、超解像を行う対象フレームに位置合わせをする。その後、位置合わせされた前フレームの超解像処理画像と対象フレームを用いて超解像を行う。これを再帰的に繰り返すことで、各フレームの異なる標本点での画素値を積み上げることができ、入力されたビデオを高解像度化する。

2.4 Multi Frame Quality Enhancement (MFQE)

2.4.1 技術概要

符号化映像では定期的に高品質フレームが出現する。この高品質フレームを利用したフレーム品質強化手法が Multi Frame Quality Enhancement (MFQE) [20][21]である。MFQE は符号化ビデオ内の品質がピークとなるフレーム PQF (Peak Quality Frame) を参照し、高品質フレーム間のフレーム non-PQF の品質を向上させる。

2.4.2 符号化ビデオの品質変動

図 2.6 は VVC の参照用コーデックである VVC Test Model-10.0 (VTM-10.0) [22]を用いて Random Access モードで符号化したビデオのフレームごとの品質を示す。グラフからわかるように、符号化されたビデオの品質は変動する。また、図中で赤い丸印で示したように、高品質フレームは数フレームごとに、周期的に出現する。これは、符号化のフレーム間予測における、階層参照構造が原因である[23]。図 2.7 は Random Access モードにおける階層参照構造を示す。Random Access モードでは、0 フレーム目、8 フレーム目を参照し、4 フレーム目を予測、0 フレーム目と 4 フレーム目を参照し 2 フレーム目を予測…といった参照と予測を繰り返す。そのため、イントラフレームやそれを参照する上層の

フレーム群は品質が高く、参照を繰り返す下層のフレーム群ほど品質が低くなる傾向がある。この高品質な上層のフレームは8の倍数フレームごとに出現するため、定期的な品質変動が見られる。

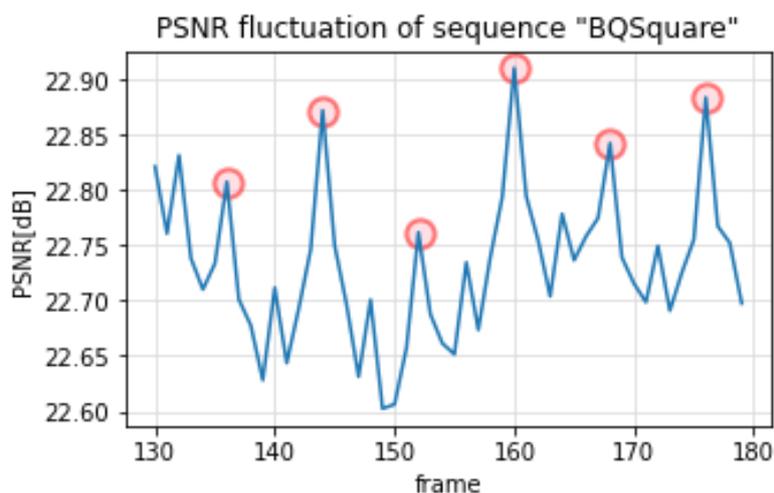


図 2.6 符号化ビデオのフレーム品質変動

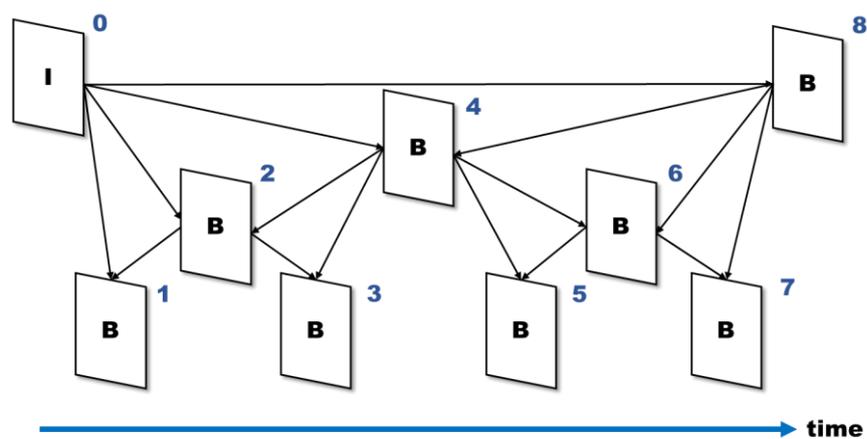


図 2.7 Random Access モードにおける階層参照構造

2.4.3 構造

2.4.3.1 構造の概要

MFQE の構造について説明する。図 2.8 は MFQE の概略図である。MFQE は Bi-LSTM [24] をベースとした検出器により PQF を検出する。その後、動き補償サブネット (MC-subnet : Motion Compensation subnet) を用いて、超解像処理を行う対象の non-

PQF に、その前後の PQF を位置合わせする。位置合わせされた前後の PQF は対象の non-PQF と深度方向に結合され、品質向上サブネット (QE-subnet: Quality Enhancement subnet) によって non-PQF の品質を向上する。

2.4.3.2 動き補償サブネットの構造

図 2.9 に動き補償サブネットの構造を示す。動き補償サブネットでは、PQF と non-PQF 間の動きベクトル (オプティカルフロー) を推定する。動きベクトルの推定はまず、低解像度での推定を行い、その結果を用いて高解像度での推定を行う。具体的には、4 倍にダウンスケーリングした PQF, non-PQF 間で動きベクトルを推定、PQF の動き補償を行う。次に、4 倍ダウンスケーリングで得られた動きベクトルと動き補償された PQF を 2 倍ダウンスケーリングでの動きベクトル推定の入力として用いる。同様に、等倍での動きベクトル推定にも 2 倍ダウンスケーリングの結果を用いる。以上のように、動きベクトルを低解像度から順に算出することにより、シーケンス中の大きな動きに対応する。

動き補償サブネットの学習には、動き補償後のフレーム F'_{PQF} と対象フレーム $F_{non-PQF}$ との平均二乗誤差を損失とする。学習時は動きベクトルの算出には、符号化映像を利用するが、損失の計算には非圧縮映像のフレーム $F'_{PQF}{}^{raw}$, $F_{non-PQF}{}^{raw}$ を用いる。そうすることにより、二乗誤差の計算において、符号化ノイズの影響を受けず、動きベクトルの推定を損失関数の目的とすることができる。以下に動き補償サブネットにおける損失関数 L_{MC} を示す。

$$L_{MC} = \|F'_{PQF}{}^{raw} - F_{non-PQF}{}^{raw}\|^2 \quad (2.1)$$

2.4.3.3 品質改善サブネットの構造

図 2.10 に品質改善サブネットの構造を示す。品質改善サブネットでは、3x3, 5x5, 7x7 と異なるフィルタサイズによる畳み込みを行う。異なるフィルタサイズ (受容野) による特徴抽出を行うことで、符号化映像の異なるスケールでの特徴を抽出できる。三つのサイズのフィルタを用いて、前後の PQF, 対象の non-PQF である $F_{non-PQF}$ の三枚からそれぞれ三つずつ各 32 チャンネルの特徴空間を抽出する。合計九つの特徴量は深度方向に合体され、dense connection [25] 構造を用いて、品質改善フレーム $F_{non-PQF}{}^{QE}$ を生成する。Dense connection 構造を用いることにより、特徴量の伝搬をより強固にし、勾配消失問題を解消する。

品質改善サブネットの学習には、対象の non-PQF である $F_{non-PQF}$ と生成フレーム $F_{non-PQF}^{QE}$ との平均二乗誤差を損失とする。以下にその損失関数を示す。

$$L_{QE} = \|F_{non-PQF} - F_{non-PQF}^{QE}\|^2 \quad (2.2)$$

2.4.3.4 MFQE の End-to-End での学習

MFQE では、 L_{MC} , L_{QE} の二つのサブネットワークの損失の和を最小化するように End-to-End で学習を行う。式 (2.3) に MFQE 全体での損失 L を示す。学習の初期段階では、後段の品質改善を収束させやすくするために、 L_{MC} を先に収束させる。そのため、 $a \gg b$ として、学習を進める。 L_{MC} の終息後、 $a < b$ とすることで、動き推定の精度を保ちつつ品質改善を学習する。

$$L = a \cdot L_{MC} + b \cdot L_{QE} \quad (2.3)$$

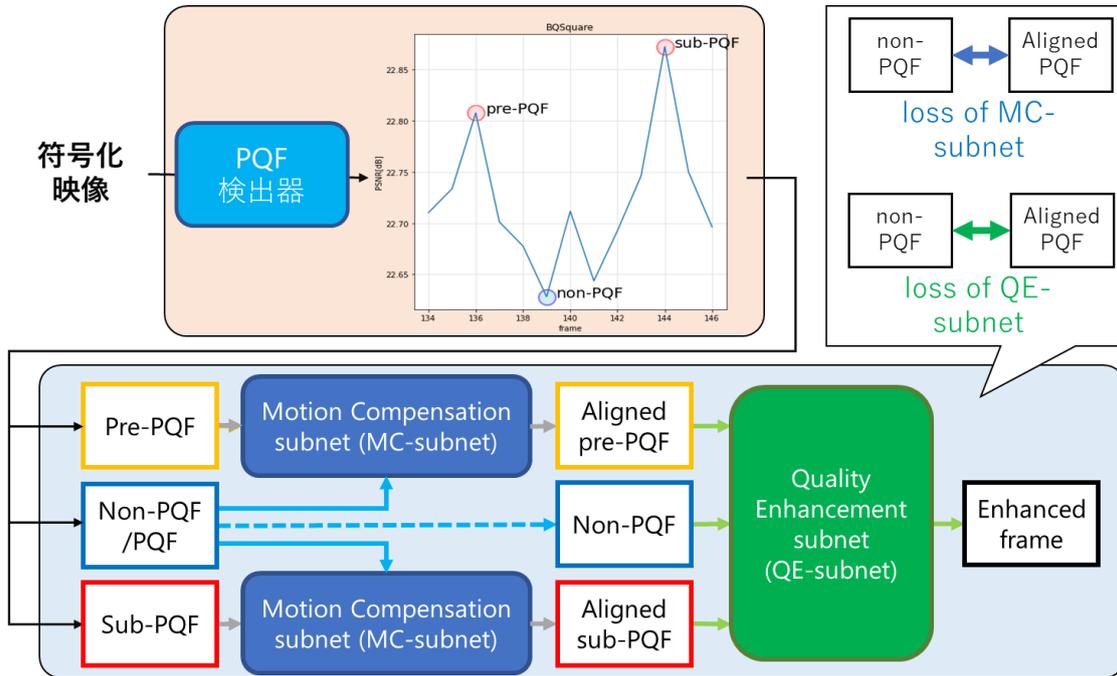


図 2.8 MFQE の構造の概略図

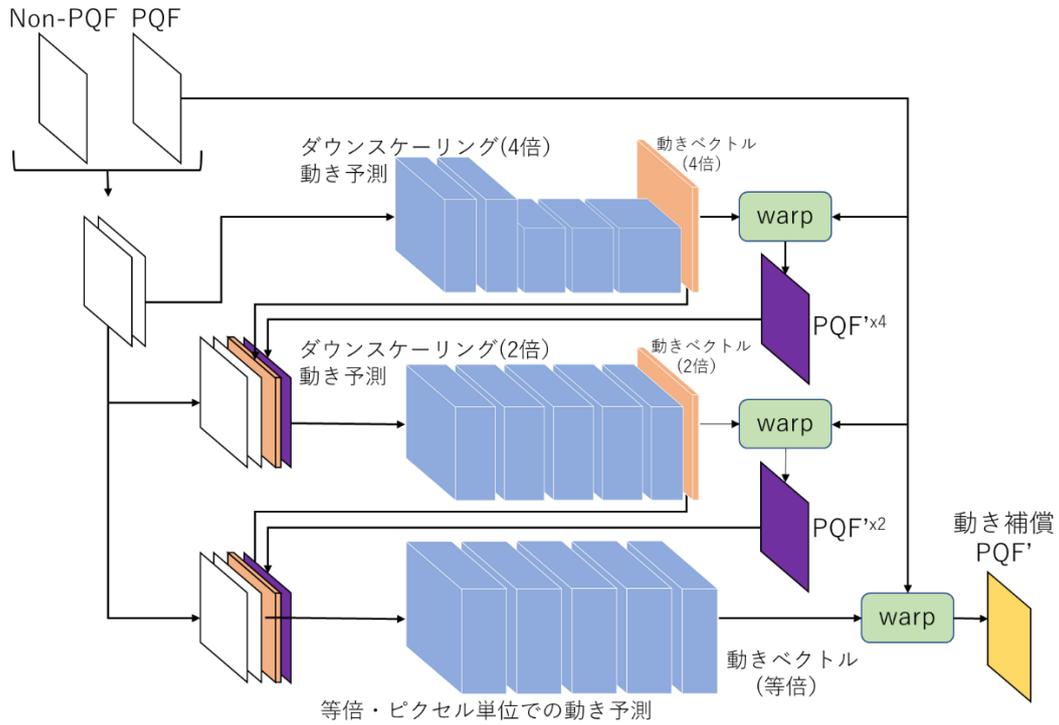


図 2.9 Motion Compensation subnet の構造

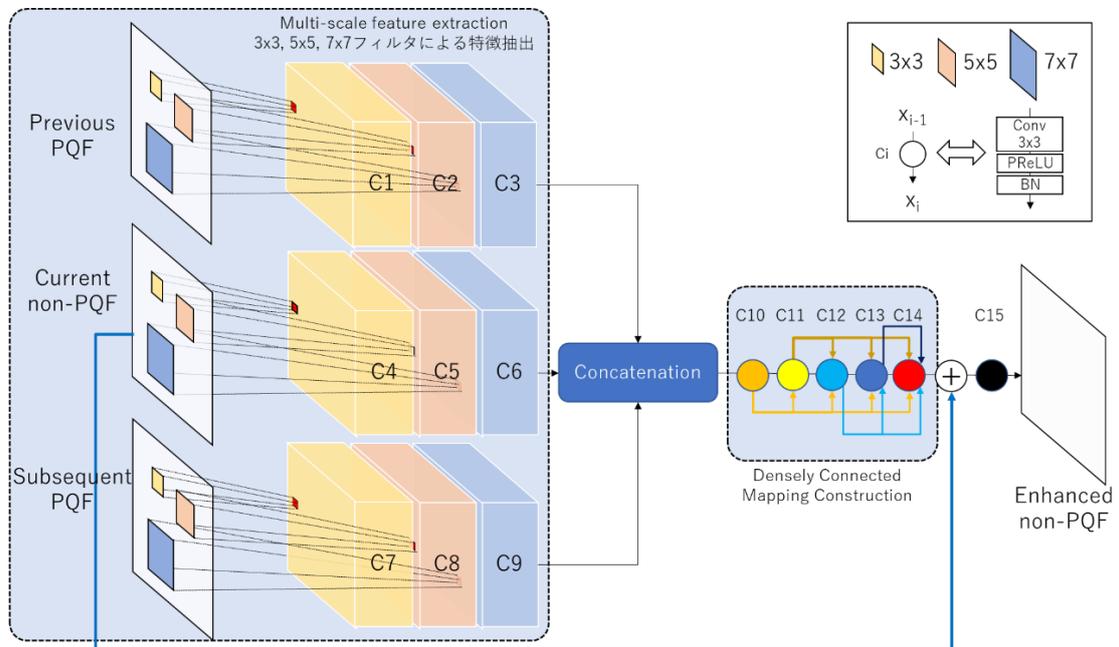


図 2.10 Quality Enhancement subnet の構造

2.5 むすび

第2章では、本研究を遂行するにあたり必要となる技術である VVC, 超解像技術および MFQE について述べた.

第3章 提案手法

3.1 概要

本章では、符号化効率改善を目的として、RPRとMFQEを組み合わせた解像度混在映像伝送方式を提案する。図3.1は超解像を用いた映像伝送方式の概略図である。この映像伝送方式では符号化前にビデオをダウンスケーリングし、符号化、伝送、復号をする。復号後、超解像などの補間技術を用いて映像を元のサイズに復元する。この手法を用いることにより、圧縮する情報を削減し、低ビットレートにおいて符号化効率が改善される [6][7][8]。しかし、単純にビデオを縮小し、超解像処理を施すだけでは符号化ビデオの品質変動や時間的相関を利用できない。そこで、品質変動や時間的相関を活用するために、RPRとMFQEを組み合わせた解像度混在映像伝送方式を提案する。提案手法の概略図を図3.2に示す。提案手法では、一部のフレームを元画像サイズで符号化、その他のフレームを縮小サイズで符号化を行う。元画像サイズ情報を一部のフレームに、元画像サイズのフレームを縮小フレームの復元に利用する。これにより、品質変動の周期性や時間的相関を活用する。

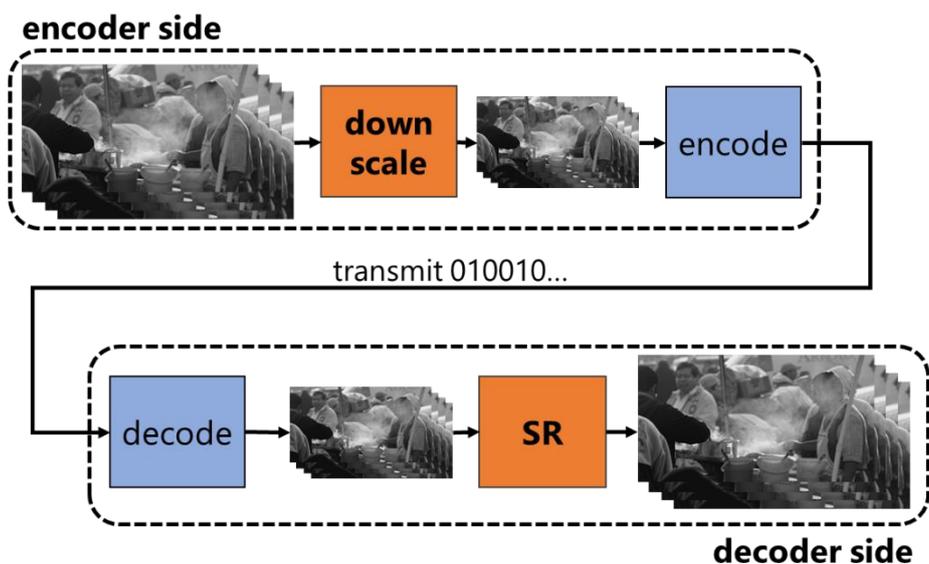


図 3.1 ダウンスケーリングと超解像を用いた映像伝送方式

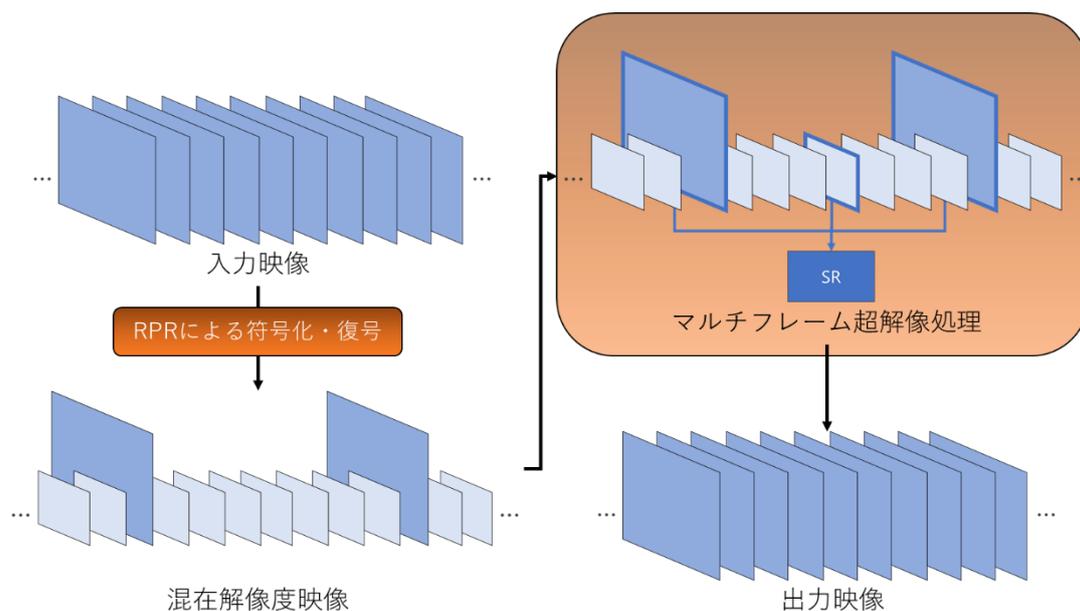


図 3.2 解像度混在映像伝送方式の概念図 (提案手法)

3.2 RPR と MFQE を組み合わせた映像伝送方式

本提案手法について具体的に述べる。MFQE は符号化映像中の高品質フレームを用いて低品質フレームの品質を高める手法である。そのため、高品質フレームに情報を集め、それ以外のフレームの情報を意図的に削減し符号化し、復号後、高品質フレームを用いて MFQE で他フレームの情報を復元することで、効率的な映像伝送が期待できる。そこで、提案手法では、RPR を用いて、PQF を元画像サイズ、non-PQF を縮小サイズで符号化し、意図的な品質制御を行う。

VVC Random Access モードで符号化した映像は 8 の倍数フレームの品質が高くなる。そのため、MFQE において、8 の倍数フレームを PQF と指定して品質向上処理を行うと効果的である。提案手法では、PQF を元画像サイズ、non-PQF を縮小して符号化することにより、通常サイズでの映像よりも圧縮する情報量を削減できる。また、PQF を元画像サイズで符号化することにより、元画像サイズの情報を用いて、縮小された non-PQF を復元できる。そのため、MFQE による十分な品質改善が期待できる。

以上のように、意図的に解像度を切り替えることにより、符号化する情報量の削減と MFQE の性能向上が期待できる。そして、これらの効果により符号化効率の改善を実現する。

3.3 提案手法と従来手法の比較

3.3.1 提案手法

図 3.3 は提案した映像伝送方式の概略図を示す。RPR により PQF を元画像サイズ、non-PQF を縦横それぞれ 2 分の 1 のサイズで符号化をする。そして復号後、縮小サイズの non-PQF を VTM に搭載されている線形補間フィルタを用いて、元画像サイズに補間をする。その後、元画像サイズで復号されたフレームを使用して、MFQE で補間されたフレームの品質改善を行う。ここで、MFQE は品質強化手法であり、超解像手法ではない。そのため、出力画像の解像度は入力画像の解像度と同じとなる。そこで、VTM の線形補間フィルタを用いて、アップスケーリングを行う。

3.3.2 全フレームを縮小する場合

図 3.4 は全フレームを縮小する場合での映像伝送方式を示している。すべてのフレームを単純に縦横それぞれ 2 分の 1 のサイズで符号化する。復号後、アップスケーリングをし、MFQE による品質改善を施す。

3.3.3 提案手法と全フレームを縮小する場合の比較

提案手法は全フレームを縮小する場合と比較して、PQF が元画像サイズで保持される。そのため、提案手法は元画像サイズの情報を活用でき、全フレームを縮小する場合よりも MFQE の性能が高まると予想される。一方で、提案手法は PQF を元画像サイズで保持するため、全フレームを縮小する場合よりも、多くの情報を符号化する必要がある。そのため、提案手法では、全フレームを縮小する場合よりも MFQE の性能が上がったとしても、符号化効率は悪化する可能性がある。そこで、本研究では提案手法と全フレームを縮小した場合での RD 曲線を比較し、本手法の有効性を評価する。

3.4 まとめ

本章では提案手法である RPR と MFQE を組み合わせた解像度混在映像伝送方式を提案した。また、単純に縮小処理と超解像処理を組み合わせた映像伝送方式との比較を示し、提案手法の利点と課題を示した。

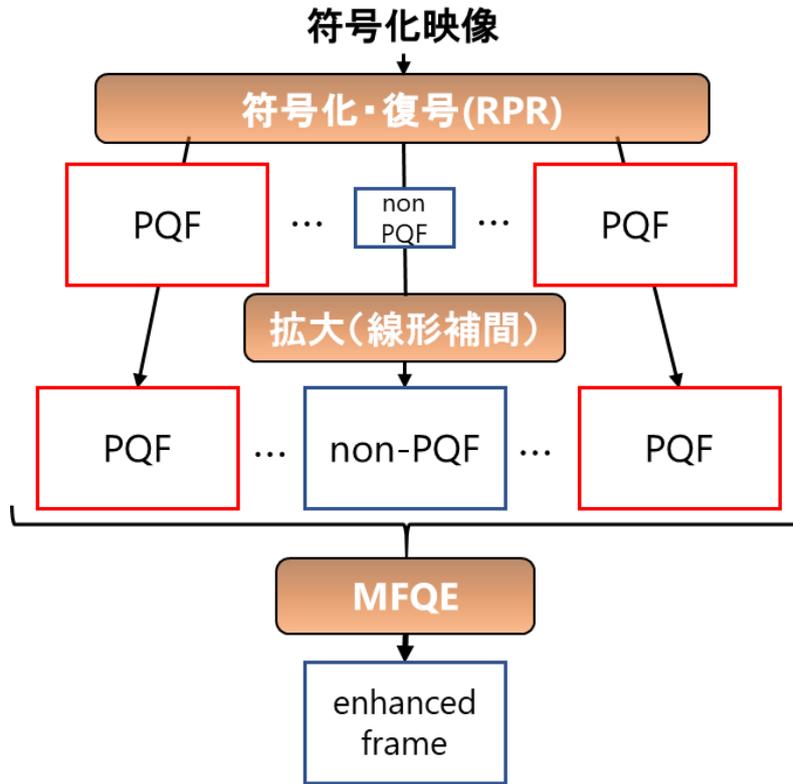


図 3.3 提案手法の概略図

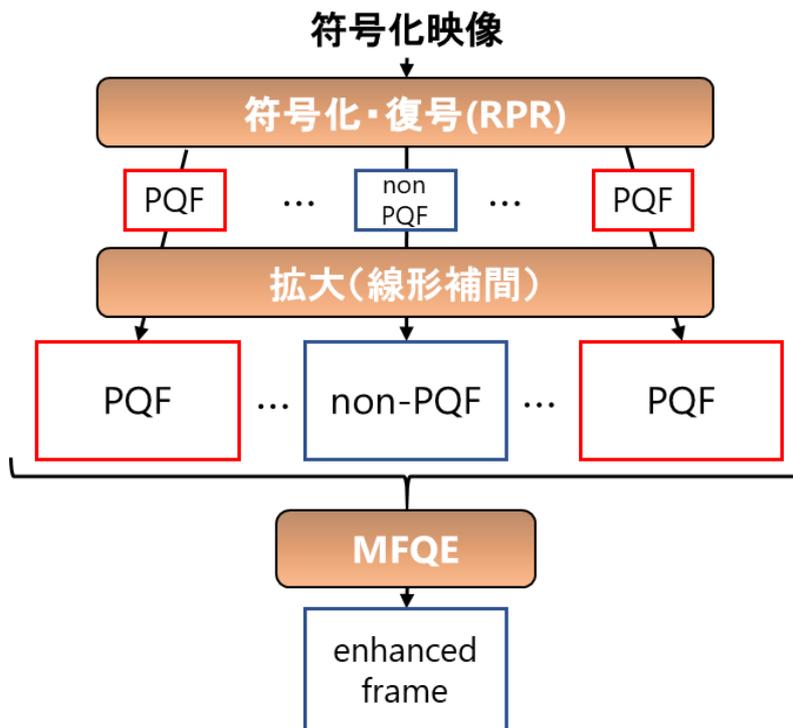


図 3.4 全フレームを縮小する場合の概略図

第4章 予備実験：VVC におけるダウンスケーリングを含む映像伝送方式の有効性

4.1 概要

本章では、VVC 符号化映像に対するダウンスケーリングを含む映像伝送方式の有効性を確認するために予備実験を行う。予備実験では、MFQE を用いず、表 1 に示す三つのケースに則り、テストシーケンスに対し符号化処理を行い、各ケースでの符号化効率の比較を行う。Case1 では全てのフレームを元画像サイズで符号化を行う。Case2 では全てのフレームを縦横二分の一のサイズで符号化を行い、復号後 VTM による補間フィルタで元のサイズに復元する。Case3 では、8 の倍数フレームである PQF を元画像サイズ、それ以外のフレームを縦横 2 分の 1 で符号化し、復号後、縮小サイズのフレームを VTM による補間フィルタで元のサイズに復元する。

表 1 比較する 3 つのケースの符号化条件

ケース	符号化条件（解像度）
Case1	元画像サイズ
Case2	縦横 2 分の 1 に縮小
Case3	PQF：元画像サイズ non-PQF：縦横 2 分の 1 に縮小

4.2 評価指標

4.2.1 RD 曲線

実験では、各ケースで符号化・復号したビデオの RD 曲線を比較し、符号化効率を比較する。RD 曲線とは横軸に符号化のビットレート、縦軸に画質評価指標の値をとったグラフにおいて、ビデオの符号化品質（ビットレート）を変化させながら、そのビットレートと画質評価指標の値をプロットして、各点を結んだ曲線である。RD 曲線を比較すること

により、各ケースの符号化品質を同一ビットレート上で比較ができる。本実験では、各ビデオに対し、4点プロットを施し、2次スプライン補間を用いて、二次関数の曲線で4点を補間する。

4.2.2 PSNR

画質評価指標には Peak Signal-to-Noise Ratio (PSNR:ピーク信号対雑音比)を用いる。PSNRは原画と入力画像の画素値がどの程度異なるかを評価する指標である。PSNRの計算式は式4.1, 4.2で表される。ここで、 MAX_I は対象画像のビット深度で表現できる画素値の最大値、I, Kは原画と入力画像、m, nはそれぞれ縦横の解像度を示す。PSNRの値が高いほど入力画像が原画に近いことを表す。

$$PSNR = 10 \log_{10} \frac{MAX_I^2}{MSE} \quad (4.1)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (4.2)$$

4.3 テストシーケンス

実験は JVET Common Test Conditions (CTC) for SDR [26]をテストシーケンスとして行う。CTCは Joint Video Experts Team (JVET)において共通条件下で符号化品質を評価するために使われるシーケンス集である。シーケンスは解像度やビット深度によって7種類のClassに分類されている。表2にCTCの構成を示す。本実験では、Random Accessに対応しているClass E以外のシーケンスをテストに用いる。また、RD曲線を描画するために、テストシーケンスを四つのビットレートでそれぞれ符号化する。異なるビットレート4点は各Caseに対し、量子化パラメタ (QP: Quantization Parameter)を変更して、決定した。

QPは量子化の粗さを決めるパラメタであり、QPの値が大きくなるほど、量子化が粗くなり、符号化映像の品質が低くなる。このQPを各Caseに対し、4点取る。表3にその詳細を示す。Case1ではQP=32, 37, 42, 47で符号化した、テストを行う。また、Case2はCase1を基準にしてオフセットとして5低いQPを用いる。Case3では、PQFはCase1と同じQPを用いて、non-PQFではオフセットとして6低いQPを用いる。各Caseで異な

るQPを用いるのは、各Caseのビットレートが近くなるように調整し、比較を容易にするためである。

表2 CTCの構成とシーケンスのプロパティ

クラス名	シーケンス名	解像度	フレーム数	フレームレート	ビット深度
A1	Tango2	3840x2160	294	60	10
A1	FoodMarket4	3840x2160	300	60	10
A1	Campfire	3840x2160	300	30	10
A2	CatRobot	3840x2160	300	60	10
A2	DaylightRoad2	3840x2160	300	60	10
A2	ParkRunning3	3840x2160	300	50	10
B	MarketPlace	1920x1080	600	60	10
B	RitualDance	1920x1080	600	60	10
B	Cactus	1920x1080	500	50	8
B	BasketballDrive	1920x1080	500	50	8
B	BQTerrace	1920x1080	600	60	8
C	RaceHorses	832x480	300	30	8
C	BQMall	832x480	600	60	8
C	PartyScene	832x480	500	50	8
C	BasketballDrill	832x480	500	50	8
D	RaceHorses	416x240	300	30	8
D	BQSquare	416x240	600	60	8
D	BlowingBubbles	416x240	500	50	8
D	BasketballPass	416x240	500	50	8
E	FourPeople	1280x720	600	60	8
E	Johnny	1280x720	600	60	8
E	KristenAndSara	1280x720	600	60	8
F	ArenaOfValor	832x480	600	60	8
F	BasketballDrillText	832x480	500	50	8
F	SlideEditing	1280x720	300	30	8
F	SlideShow	1280x720	500	20	8

表 3 Case1, 2, 3 符号化時の量子化パラメタ

設定		学習・テストの QP			
Case1 (元画像サイズ)		32	37	42	47
Case2 (全フレームを縮小)		27	32	37	42
Case3 (提案手法)	PQF	32	37	42	47
	non-PQF	26	31	36	41

4.4 実験手順

各ケースの符号化条件に従い四つの QP で符号化し、テストシーケンスを作成する。一つのビデオに対し、合計 12 のテストシーケンスが作成される。それぞれに対し、非圧縮のビデオとの PSNR を測定し、RD 曲線を描画、比較する。

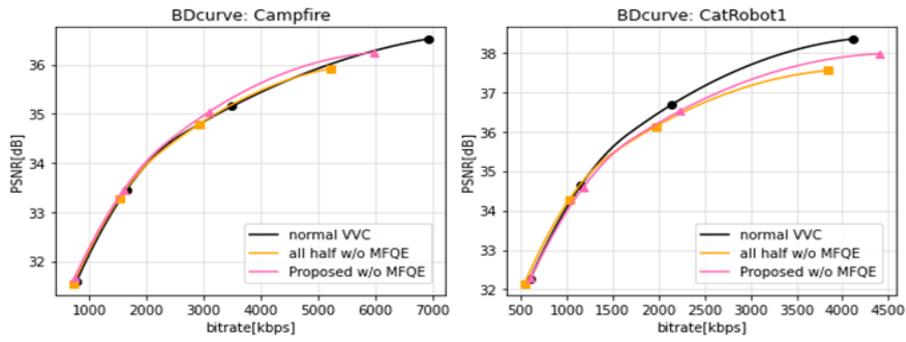
本論文内の実験では、学習・テストにはビデオの輝度成分のみを使用し、符号化・復号には VTM10.0 を用いる。

4.5 結果および考察

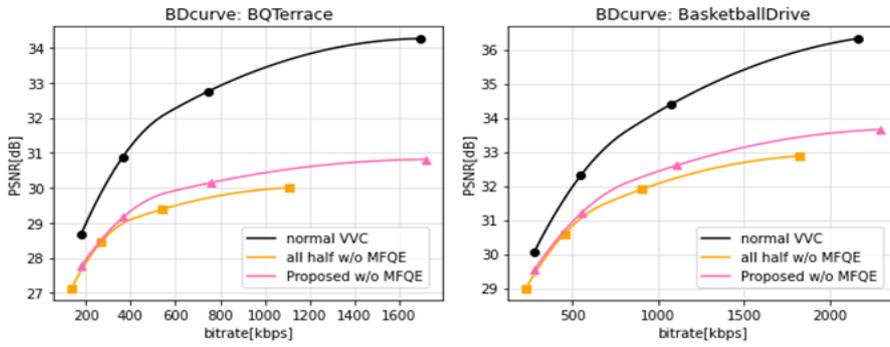
図 4.1 に Class A1 & A2, B, C, D のそれぞれのシーケンスに対する RD 曲線を示す。黒色の線が Case1 (通常の符号化)、オレンジ色の線が Case2 (全フレームを縮小)、ピンク色の線が Case3 (提案手法) を示す。Class B, C, D では、Case2, 3 の RD 曲線は Case1 と比較して低下している。一方で Class A1, A2 シーケンスに対しては、Case2, 3 の RD 曲線は Case1 に匹敵している。そのため、Class A1, A2 などの 4K シーケンスに対して、MFQE を用いることで Case1 よりも符号化効率を改善することが期待できる。

4.6 まとめ

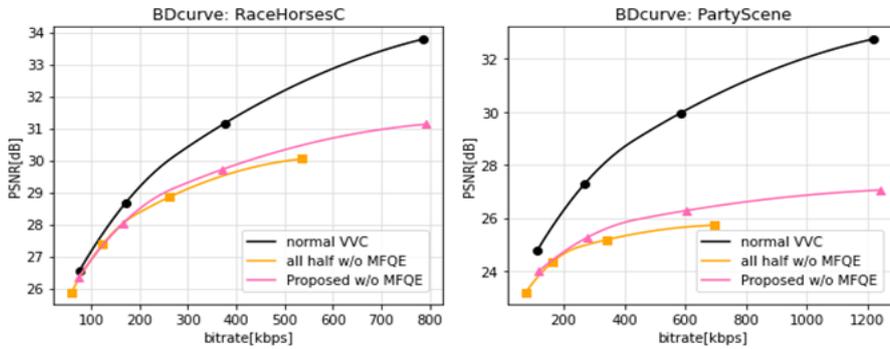
本章では、VVC 符号化映像に対するダウンスケーリングを含む映像伝送方式の有効性を評価した。通常サイズでの符号化、縮小サイズでの符号化、提案手法での符号化映像の RD 曲線を比較し、Class A1, A2 などの 4K シーケンスに対し、縮小サイズでの符号化、提案手法は通常サイズでの符号化に符号化効率が匹敵した。



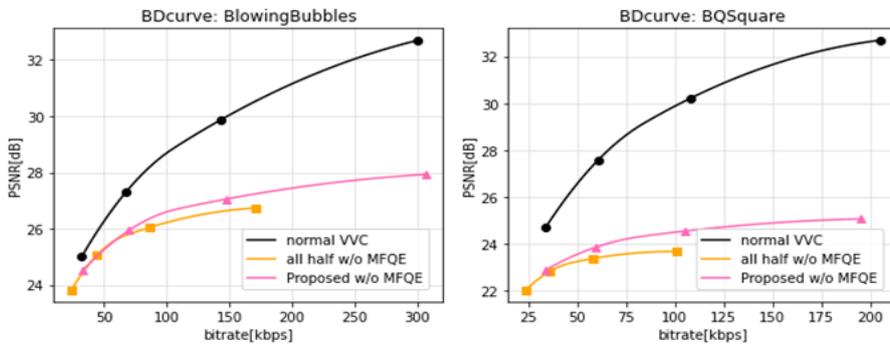
(a) Class A1, A2



(b) Class B



(c) Class C



(d) Class D

図 4.1 各 Class の予備実験条件下での RD 曲線

第5章 評価実験：超解像を含む解像度混在映像伝送方式の品質評価

5.1 概要

予備実験結果より，Class A1, A2 などの 4K シーケンスに対して提案手法が有効と考えられる．そこで，本章では提案手法の有効性を確認するために，提案手法の符号化効率の評価を行う．また，提案手法が元画像サイズのフレームを有効に参照していることを確認するために，MFQE 性能の評価を行う．最初に実験手順，MFQE の学習について述べ，その後，符号化効率改善の評価，MFQE 性能の評価について述べる．

5.2 実験手順

図 5.1 にテストに用いた Class A1, A2 の全六つのシーケンスを示す．これらのシーケンスを予備実験と同じく，Case1, 2, 3 に対し，表 3 の通り異なる四つの QP で符号化を行う．復号後，Case2, 3 では，MFQE による品質強化を行い，原画と出力画像との PSNR を測定し，RD 曲線を描画する．



図 5.1 Class A1, A2 シーケンスの画面内容

5.3 MFQE の学習

5.3.1 学習の概要

MFQE の学習について述べる。Case2 と Case3 では、フレーム品質が異なる。また、QP の値によってフレーム品質は変化する。一般に、ネットワークは汎化性能を高めると、特定の条件に対する特化性能が低下してしまう[27]。また、様々な入力画像特性を把握するためには、ネットワークサイズを大きくする必要がある。そのため、MFQE において、一つのネットワークで各 Case, 各 QP で符号化された映像を同時に学習するのは難しい。そこで、本研究では Case2 の QP ごと、Case3 の QP ごとに、合計 8 つの MFQE のモデルを学習する。また、学習には MFQE の著者らが公開している学習済みモデル[28]をファインチューニングして使用した。

5.3.2 学習用データセット

MFQE の学習には SJTU データセット[29]、UVG データセット[30]を用いた。SJTU、UVG データセットはそれぞれ 15 個、16 個の非圧縮の 4K 解像度のシーケンスによって構成されている。それぞれのシーケンスは 4:2:0 YUV フォーマットであり、ビット深度は 8bit である。学習データセット作成には最初の 64 フレームを使用した。

学習用データセットの作成のため、データセットを Case2 条件下での各 QP、Case3 条件下での各 QP、それぞれに対し符号化を行う。その後、非圧縮のデータセットと各条件下で圧縮したデータセットをペアにして学習を行う。学習には復号されたビデオの non-PQF とその前後の PQF の合計 3 枚と非圧縮ビデオのフレームの合計 4 枚を学習セットに用いる。PQF と non-PQF との時間方向の距離に偏りがないように、non-PQF は $8n+1, 8n+2, \dots, 8n+7$ 番目のフレームから学習セットごとに 1 枚ランダムに選択する。また、同一シーケンスを連続で学習すると、MFQE の性能がそのシーケンスに特化してしまうため、各シーケンスがランダムに入力されるように学習セットをシャッフルする。学習時には各フレームを 128×128 のパッチに分割して学習を行う。

5.3.3 PQF の学習

MFQE では non-PQF のほかに、PQF 自身も前後の PQF を用いて品質改善が可能である。Non-PQF と PQF では品質が異なる。そこで、PQF についても各 Case, QP ごとにモデルの学習を行う。

また、PQF において、I ピクチャは、ほかの PQF よりも PSNR が高い。そのため、I ピクチャがほかの PQF を参照して MFQE による品質強化処理を行うと、PQF を有効に使える。

ない場合がある。そこで、I ピクチャは、前後のフレームを用いずに、I ピクチャ単体で品質改善を行う。具体的には、MFQE の動き補償サブネットを用いず、品質改善サブネットのみを使用する。通常、品質改善サブネットは並び替えを行った前後の PQF と対象フレームの 3 チャンネルを入力とするが、これを 1 チャンネルに変更し、単体画像で品質強化を行う。

5.4 符号化効率の評価

5.4.1 概要

本節では Case1, 2, 3 の符号化効率を比較し、提案手法による符号化効率改善の評価をする。

5.4.2 結果

Case1, 2, 3 に MFQE を用いた際の RD 曲線を図 5.2 に示す。また、Case1 と Case3 の比較、Case1 と Case2 の比較をしたグラフをそれぞれ図 5.3, 図 5.4 に示す。各 Case の測定点のビットレートをそれぞれ表 4, 表 5, 表 6 に示す。Case1 と Case3 の比較では、“Tango2”, “FoodMarket4”, “Campfire”, “ParkRunning3” の 4 つのシーケンスに対し、Case3 は右側の測定点 2 点で Case1 よりも符号化効率が高くなった。Case1 と Case2 の比較では、全てのシーケンスの非常に低いビットレート帯に対し、Case2 は Case1 よりも符号化効率が高くなった。

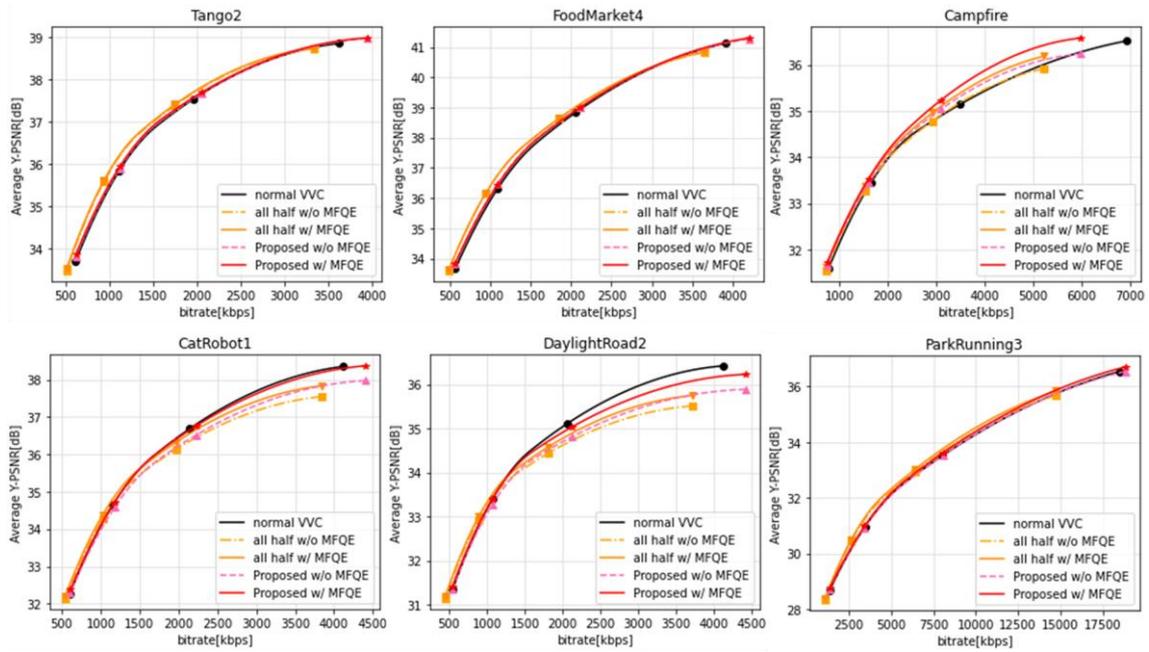


図 5.2 MFQE を用いた際の RD 曲線 (Case1, Case2, Case3)

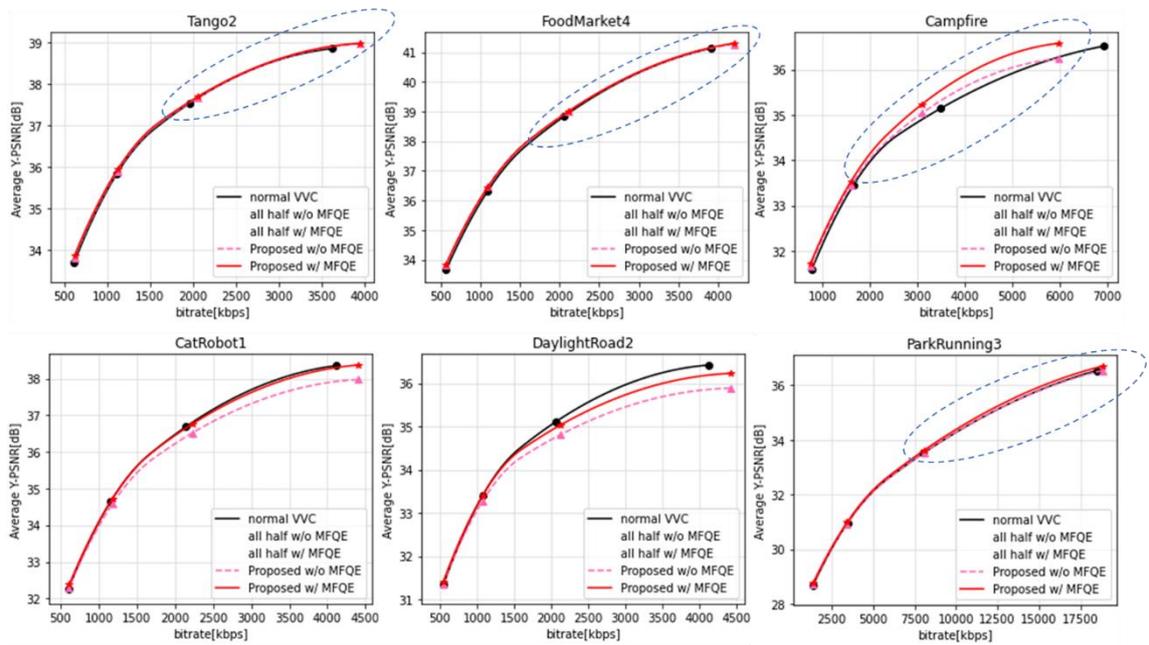


図 5.3 MFQE を用いた際の RD 曲線 (Case1, Case3)

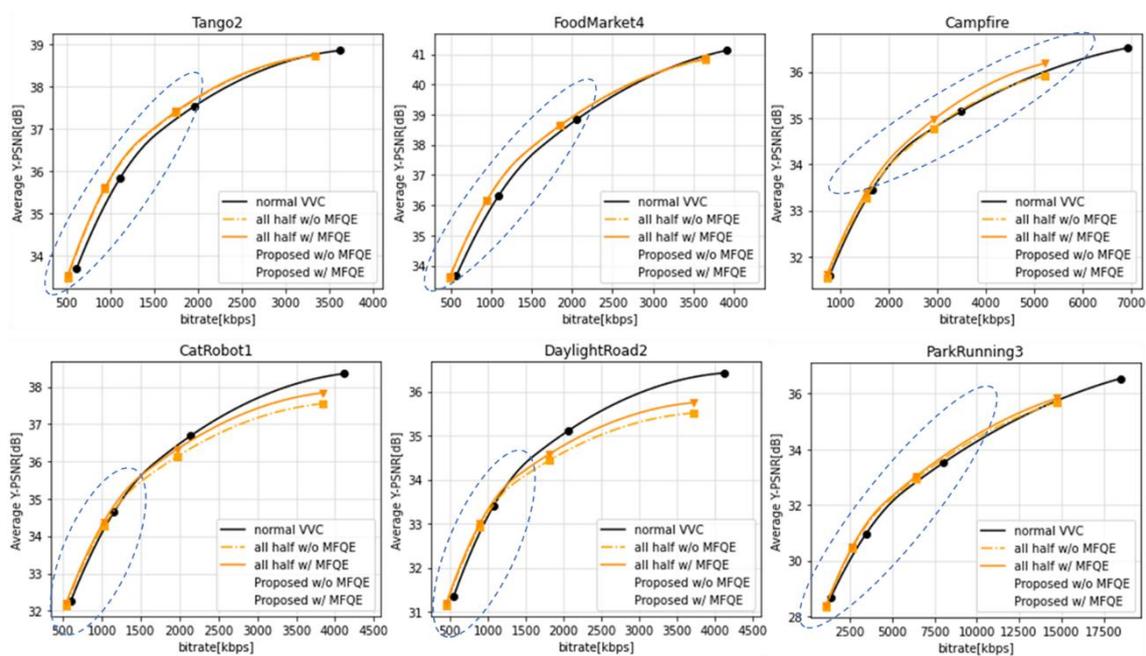


図 5.4 MFQE を用いた際の RD 曲線 (Case1, Case2)

表 4 Case1 の各 QP でのビットレートと PSNR の関係

Sequence	Tango2				FoodMarket4				Campfire			
QP	32	37	42	47	32	37	42	47	32	37	42	47
Bitrate [kbps]	3620.4	1968.8	1111.8	617.9	3910.1	2060.2	1095.1	570.2	6927.1	3500.5	1653.4	772.7
PSNR [dB]	38.859	37.549	35.829	33.697	41.124	38.840	36.326	33.669	36.520	35.151	33.456	31.584

Sequence	CatRobot1				DaylightRoad2				ParkRunning3			
QP	32	37	42	47	32	37	42	47	32	37	42	47
Bitrate [kbps]	4109.8	2144.8	1157.0	604.2	4122.3	2057.3	1072.8	547.6	18467.4	8015.0	3438.8	1361.8
PSNR [dB]	38.353	36.690	34.651	32.277	36.421	35.105	33.397	31.349	36.540	33.512	30.953	28.671

表 5 Case2 の各 QP でのビットレートと PSNR の関係

Sequence	Tango2				FoodMarket4				Campfire			
QP	27	32	37	42	27	32	37	42	27	32	37	42
Bitrate [kbps]	3338.7	1753.6	942.2	518.5	3654.2	1855.0	945.4	488.5	5233.7	2921.3	1522.4	725.5
w/o MFQE	38.733	37.393	35.583	33.491	40.826	38.640	36.158	33.615	35.921	34.785	33.276	31.556
w/ MFQE	38.745	37.418	35.615	33.537	40.848	38.645	36.159	33.616	36.190	34.965	33.380	31.624
Δ PSNR [dB]	0.011	0.024	0.032	0.047	0.022	0.006	0.001	0.001	0.270	0.179	0.105	0.068

Sequence	CatRobot1				DaylightRoad2				ParkRunning3			
QP	27	32	37	42	27	32	37	42	27	32	37	42
Bitrate [kbps]	3841.4	1965.4	1023.4	542.3	3723.7	1818.2	895.3	451.6	14730.0	6413.5	2643.3	1085.7
w/o MFQE	37.550	36.130	34.267	32.154	35.513	34.445	32.938	31.155	35.684	32.942	30.454	28.359
w/ MFQE	37.839	36.322	34.378	32.203	35.750	34.589	33.013	31.196	35.833	33.030	30.507	28.393
Δ PSNR [dB]	0.289	0.192	0.111	0.050	0.237	0.144	0.075	0.041	0.149	0.088	0.053	0.034

表 6 Case3 の各 QP でのビットレートと PSNR の関係

Sequence	Tango2				FoodMarket4				Campfire			
QP	32	37	42	47	32	37	42	47	32	37	42	47
Bitrate [kbps]	3941.5	2056.9	1122.7	622.8	4198.3	2115.1	1091.1	564.4	5988.9	3102.1	1605.0	748.2
w/o MFQE	38.976	37.675	35.906	33.819	41.272	39.009	36.457	33.826	36.242	35.041	33.453	31.677
w/ MFQE	38.987	37.699	35.942	33.870	41.298	39.010	36.435	33.822	36.584	35.240	33.540	31.725
Δ PSNR [dB]	0.012	0.024	0.036	0.051	0.026	0.001	-0.022	-0.004	0.342	0.199	0.087	0.048

Sequence	CatRobot1				DaylightRoad2				ParkRunning3			
QP	32	37	42	47	32	37	42	47	32	37	42	47
Bitrate [kbps]	4396.1	2227.9	1178.2	611.1	4414.4	2126.2	1073.2	544.2	18819.3	8092.7	3419.9	1362.5
w/o MFQE	37.975	36.520	34.583	32.300	35.887	34.813	33.276	31.363	36.543	33.544	30.923	28.707
w/ MFQE	38.371	36.765	34.728	32.375	36.229	35.036	33.402	31.414	36.702	33.632	30.979	28.746
Δ PSNR [dB]	0.396	0.245	0.145	0.075	0.341	0.223	0.126	0.051	0.159	0.089	0.055	0.039

5.4.3 考察：Case3（提案手法）の特性

Case3 では四つのシーケンスで Case1 対し、右の2点の低ビットレート帯で符号化効率の改善が見られた。一方で，“CatRobot1”，“DaylightRoad2”では全ての測定点において、Case3 は Case1 よりも符号化効率が低下した。

図 5.5 は “CatRobot1”，“DaylightRoad2”の RD 曲線である。図中の点線は MFQE を用いない場合での Case2, 3 の RD 曲線を示す。通常サイズで符号化した場合である Case1 と比較し、縮小拡大を含む Case2, 3 で MFQE を用いない場合、RD 曲線が他のシーケンスと比べ大きく低下する。

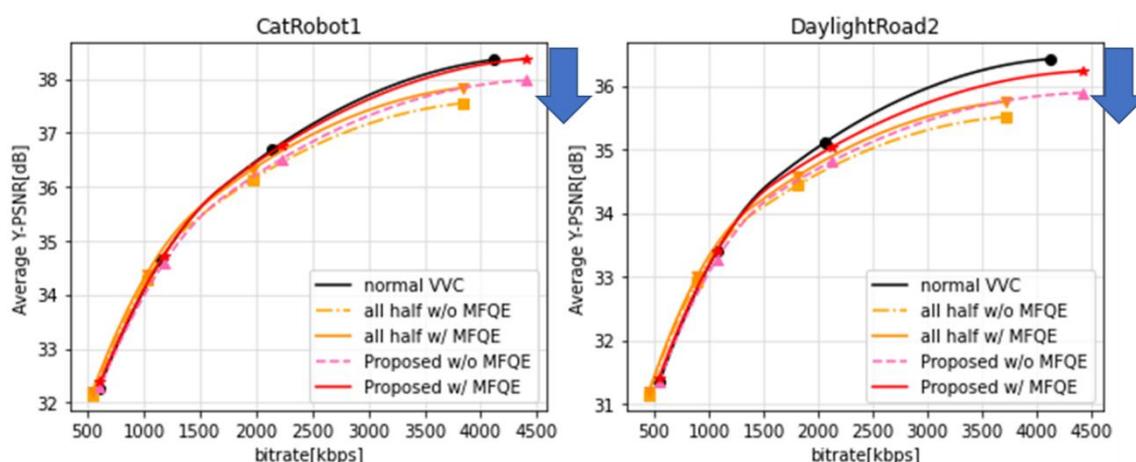


図 5.5 CatRobot1 と DaylightRoad2 の RD 曲線

これは，“CatRobot1”，“DaylightRoad2”に高周波成分が多く含まれるためと考えられる。Case2, 3 では、ビデオを縦横半分のサイズに縮小する。そのため、復元可能な最大周波数であるナイキスト周波数も半分となり、それ以上の周波数成分は失われる。ダウンサンプリングにより高周波成分が失われると、MFQE の手掛かりとなるエッジ部分も失われる。図 5.6 は Case1 の出力画像と Case3 の MFQE 出力画像、MFQE 入力画像の比較である。Case1 では信号機の点灯している部分のディテールが残されているが、Case3 の MFQE 入力画像では、縮小により、ディテールが消えている。そのため、MFQE による品質強化処理を用いても、前後の PQF を有効に利用できず、ディテールを再現できない。

以上より，“CatRobot1”，“DaylightRoad2”などの、縮小によって高周波成分が消えてしまうようなシーケンスでは、MFQE のヒントが消えてしまう。そのため、十分な復元ができず、符号化効率が通常サイズの符号化よりも低下すると考えられる。

一方で、エッジが残っているシーケンスに対して、MFQE は有効に作用する。図 5.7 は “Campfire”における Case1 の出力画像と Case3 の MFQE 出力画像、MFQE 入力画像を

示す。MFQE 入力画像でぼやけていた地面の草や、炎のエッジ部が MFQE 出力画像では復元され、先鋭化された。エッジが残っているシーケンスに対し、MFQE は有効に作用した。

また、Case3 は右 2 点、つまり PQF の QP が 32, 37 である点に対し、符号化効率が改善する傾向がある。一方で、QP=42, 47 に対し有効でないのは、以上の理由と同じく、符号化によりシーケンス内の高周波成分が失われるためだと考えられる。MFQE では入力画像の劣化が大きいと、エッジが失われ、入力画像に対して想定される出力が無数に存在する状態となる。そして、想定される出力画像を平均化するような出力となる。そのため、劣化の程度が激しい QP=42, 47 で符号化された映像では MFQE による品質改善が見られず、符号化効率が改善されなかったと考えられる。一方、劣化の影響が少ない QP=32, 37 では、MFQE のヒントとなるエッジが残り、縮小フレームを有効に復元したと考えられる。



図 5.6 Case1 の出力画像と Case3 の MFQE 出力画像，入力画像の比較 (DaylightLoad2)

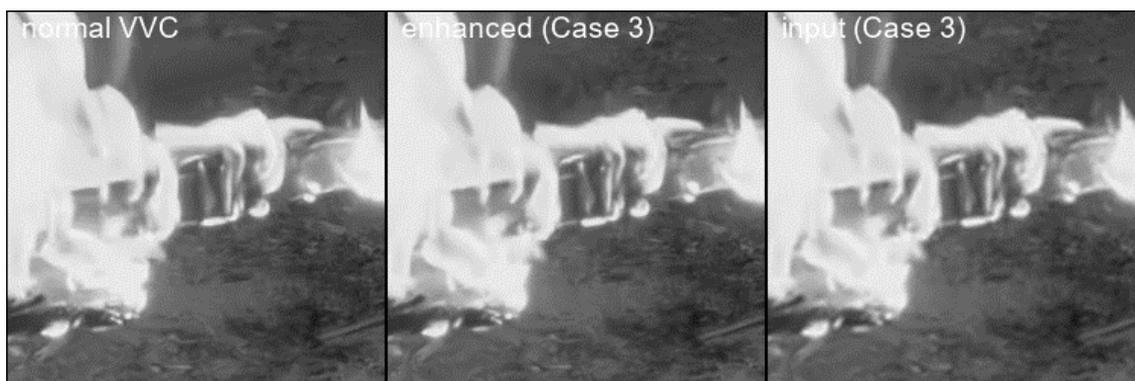


図 5.7 Case1 出力画像と Case3 の MFQE 出力画像，入力画像の比較 (Campfire)

5.4.4 考察：Case2（全フレームを縮小する場合）の特性

Case2 では全てのシーケンスで Case1 に対して、左側の非常に低いビットレート帯の測定点で符号化効率の改善が見られた。図 5.8 は元画像サイズで圧縮を行った Case1 と全フレームを半分のサイズで圧縮した Case2 の出力画像の比較である。Case1 では、ビットレートが非常に低いため、圧縮された映像は高周波成分が消えたり、ブロックノイズが生成されたりと画像劣化が顕著になる。しかし、Case2 ではブロックノイズが抑えられる傾向がある。これは、解像度を下げて符号化することにより、多くのビット量を縮小前の低周波にあたる成分に割り当てることができたためだと考えられる。ビットレートが非常に低い場合、符号化によって高周波成分のみならず、低周波成分の量子化数も大きく削減される。そのため、ブロックごとの輝度の連続性が失われブロックノイズが生じる。しかし、フレームサイズを縮小して符号化することにより、縮小前のビデオの低周波成分に集中してビット割り当てが可能となる。その結果、元画像サイズでは生成されていたアーティファクトが抑えられた。



図 5.8 Case1 の符号化画像と Case2 の MFQE 出力画像

5.4.5 Case2 と Case3 の組み合わせによる符号化効率の改善

Case2 は超低ビットレート帯で、Case3 は低ビットレート帯で符号化したビデオに対し、有効であった。図 5.9 には最も有効であった Case をビットレート帯ごとに示す。Case2, 3 のいずれかを用いることで、ほとんどの測定点で通常の符号化よりも高い符号化効率を得られる。このことから、ビットレート帯に応じて Case2, 3 を使い分ける手法が有効であると考えられる。表 7 は左 3 点で Case2, 右 1 点で Case3 を使用した際の Case1 に対する BD ゲインを示す。ビットレート帯に応じて、各手法を使い分けることにより、元画像サイズでの符号化映像よりも平均して 4.80% の BD ゲインが得られた。以上より、超解像を用いた映像伝送方式は有効であるといえる。

一方、Case2, 3 のいずれを各点に適用するべきかの判断はシーケンス内容に依存する。シーケンスにエッジが残り、PQF を有効に参照できるかはシーケンスの周波数特性や劣化の大きさによって変化する。具体的には、シーケンスを縦横半分のサイズにするとき、縮小したフレームのうち、元画像サイズのフレームのナイキスト周波数の 1/2 より大きな高周波成分が縮小による影響を受ける。PQF を有効に参照できるかは、これらの高周波成分が縮小で消失する程度、また、符号化による高周波成分の消失の程度によって変化すると考えられる。そのため、シーケンスの周波数特性と劣化による影響から提案手法の具体的な適用条件を導くのが今後の課題である。

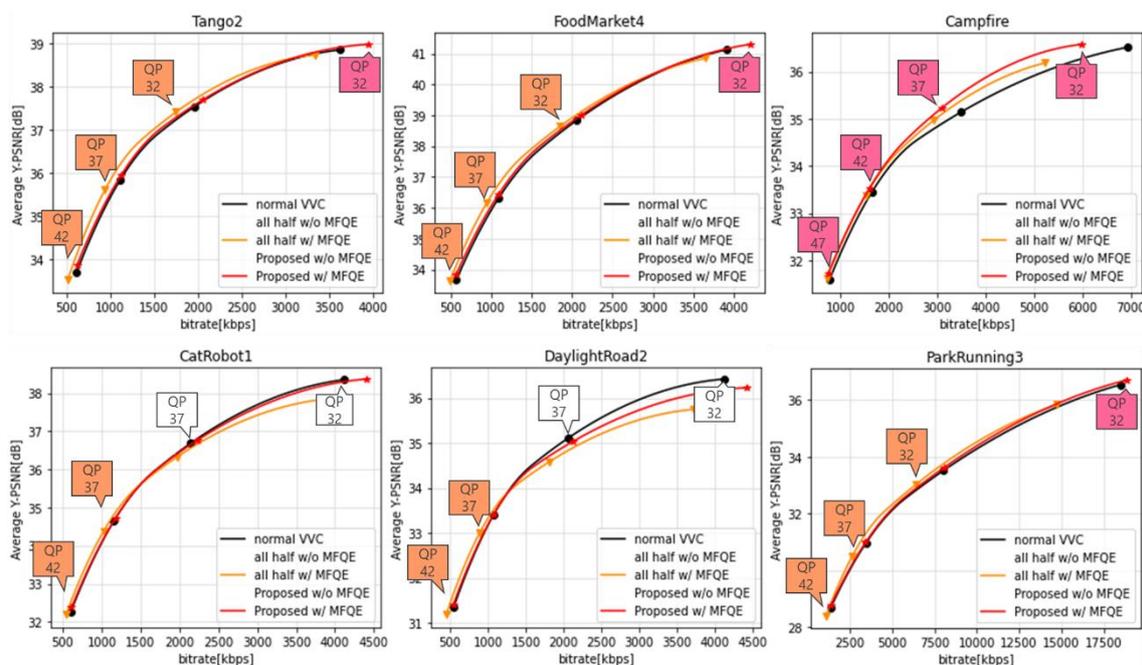


図 5.9 RD 曲線と各測定点で最も符号化効率が高い Case (Case1 : 白, Case2 : オレンジ, Case3 : ピンク)

表 7 Case2, 3 をビットレート帯に応じて適用した際の Case1 に対する BD ゲイン
(左 3 点 : Case2, 右 1 点 : Case3 を適用)

Over VTM-10.0 (QP 32,37,42,47)	
Sequence	Y (Left 3 points: Case2, Others: Case3)
Tango2	-7.93%
FoodMarket4	-7.29%
Campfire	-8.11%
CatRobot	-0.77%
DaylightRoad2	2.36%
ParkRunning3	-7.09%
Average	-4.80%

5.5 MFQE 性能の評価

5.5.1 概要

本節では, Case2, Case3 における MFQE の性能の評価を行い, 元画像サイズのフレームを参照する有効性を確認する. 結果 1 では MFQE の Case2 と Case3 での性能の比較を示す. 結果 2 では, Case3 でのフレームごとの品質改善を示す. これらの結果から, Case3 が有効に機能していることを示す.

5.5.2 結果 1 ・考察

図 5.10 は六つのテストシーケンスの各測定点での MFQE 前後の PSNR の平均値をそれぞれ Case2 と Case3 の場合について示している. Case2 に対し, Case3 の方が MFQE により PSNR が改善された値が大きいことがわかる.

表 8 は MFQE 出力画像の MFQE 入力画像に対する BD ゲインの比較を示す. Case2 は平均して 3.35%, Case3 は平均して 3.88% の BD ゲインを得られ, Case3 の方が平均して高いゲインが得られた. これらの BD ゲインは, シーケンス全体を符号化しビットレートを測定する必要があるため, non-PQF だけでなく, PQF も含めたゲインとなっている. Case3 では PQF が元画像サイズであり, PQF が縮小されている Case2 と比べて, PQF については PSNR の改善量が少なくなる. しかし, シーケンス全体の改善率を平均しても, Case3 の方が平均してゲインが高く, non-PQF の BD ゲインでは Case2 よりも Case3 の方が高いと

いえる。以上の定量的評価から、Case3では、元画像サイズのPQFを参照することにより、MFQEによる品質改善がCase2よりも効果的に機能しているといえる。

また、図5.11はCase2とCase3のMFQE出力画像の比較を示している。Case2では文字の部分がMFQEによる品質改善処理を施してもぼやけているが、Case3では元画像サイズPQFの文字を参照し品質改善処理を行うことで、はっきりとした文字を出力できている。視覚的評価からもCase3において、MFQEが元画像サイズのフレームを有効に参照できているといえる。

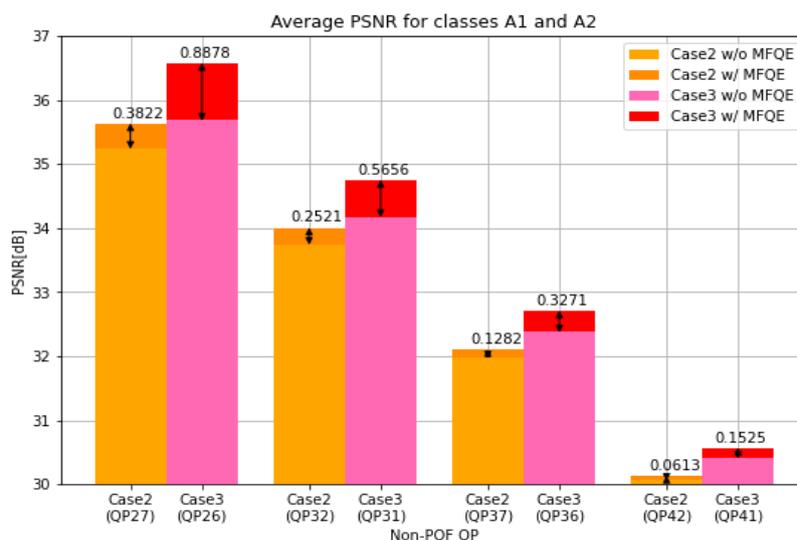


図 5.10 Case2 と Case3 の測定点ごとの MFQE 前後の PSNR の比較

表 8 Case2 と Case3 の MFQE 出力画像の MFQE 入力画像に対する BD ゲインの比較

Over w/o MFQE		
Sequence	Y (Case2)	Y (Case3)
Tango2	-0.99%	-1.06%
FoodMarket4	-0.15%	0.13%
Campfire	-5.93%	-6.06%
CatRobot	-5.25%	-6.41%
DaylightRoad2	-5.18%	-7.37%
ParkRunning3	-2.59%	-2.53%
Average	-3.35%	-3.88%

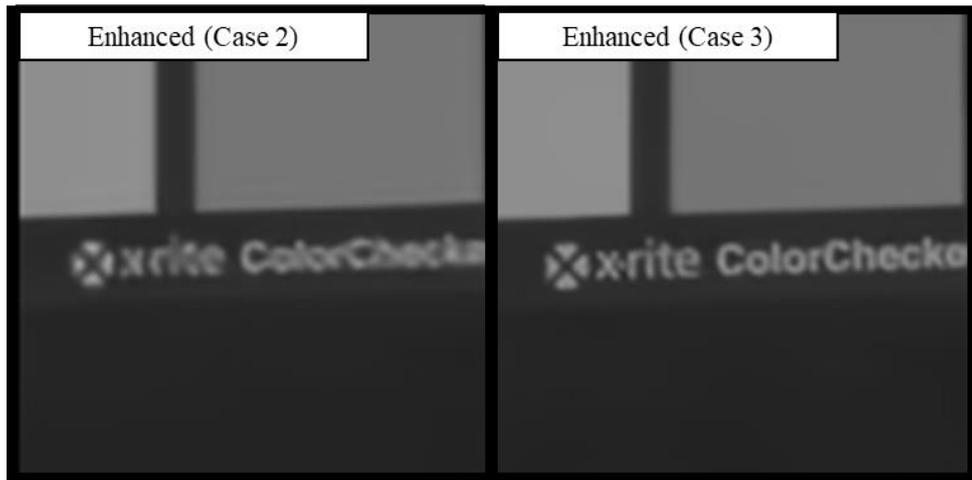


図 5.11 Case2 と Case3 の MFQE 出力画像の比較

5.5.3 結果 2 ・考察

Case3 において、MFQE を用いる場合、用いない場合のフレームごとの PSNR 変動を可視化したグラフを図 5.12 に示す。青線が MFQE を用いない場合、オレンジ色の線が MFQE を用いた場合である。PSNR が周期的に高くなっている点が PQF である。そして、その間の PSNR が低くなっている部分が non-PQF である。MFQE を用いない場合、PQF と non-PQF の品質の差が激しい。しかし、MFQE を用いる場合、PQF と non-PQF の差が縮まっていることがわかる。MFQE を用いることにより、品質変動によるフレーム間の品質差を埋めることができている。PQF を用いることにより、non-PQF の品質を PQF の品質に近づけることに成功しているといえる。

5.6 まとめ

本章では、4K 解像度のテストシーケンスに対し、各ケースでの符号化効率の比較、MFQE 性能の比較を行った。比較の結果から、Case2 は超低ビットレート帯に対し、Case3 は低ビットレート帯に対し有効であった。また、Case3 は元画像のサイズのフレームを参照することで Case2 よりも高い性能を示した。

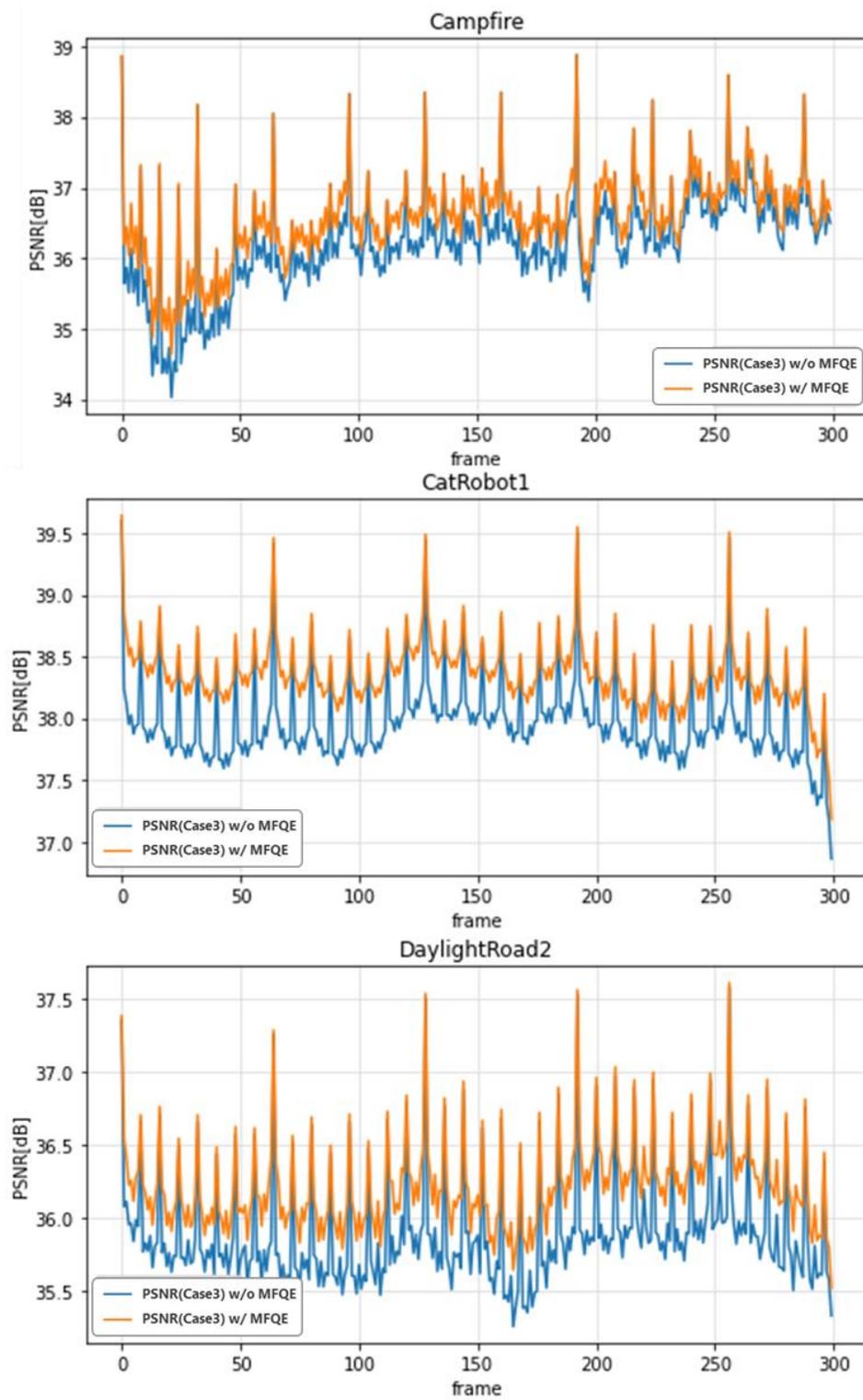


図 5.12 Case3 におけるフレームごとの PSNR

第6章 結論

6.1 結論

本論文では、異なる解像度間のフレーム参照を可能とする RPR と複数フレームを参照し符号化ビデオの品質改善を行う MFQE を組み合わせた VVC 符号化映像のための解像度混在映像伝送方式を提案した。

提案手法は、4K シーケンスに対し、通常サイズで符号化する場合よりも、低ビットレート帯に対し符号化効率が改善することを評価実験により確認した。また、全フレームを縮小する場合は超低ビットレート帯に対し、符号化効率が改善することも確認した。さらに、提案手法では MFQE が元画像サイズのフレームを有効に参照していることを評価実験より定量的、定性的に確認した。

6.2 今後の課題

今後の課題を2点挙げる。

1点目は、提案手法の適用範囲の探索である。提案手法は低ビットレート帯に対し有効であるが、その有効性はシーケンス内容に依存する(5.4.5項)。具体的には、エッジが残っているシーケンスに対して、提案手法は有効であるが、符号化や縮小により、エッジが消えたシーケンスには有効でない。そのため、シーケンスの周波数特性と符号化、縮小による影響を明確にする必要がある。また、本論文では符号化においてフレームを2倍にダウンスケールリングした。しかし、シーケンス内容によって縮小による劣化の影響が異なることから、シーケンスによって、最適な縮小倍率が存在すると考えられる。そのため、シーケンス内容に応じた品質制御手法を提案する必要がある。

2点目は、Pre-upsampling による計算コストの増大である。本提案手法では、縮小されたフレームを線形補間フィルタによりアップスケールリングした後、MFQE による品質改善処理を施す Pre-upsampling 手法を用いている。この手法では、入力画像のサイズが大きくなり、計算コストが増大する。そのため、各フレームを事前に補間せず、MFQE の後段で解像度を増やす Post-upsampling 手法を導入することが望ましい。

謝辞

本研究に際し、実験環境の用意をしてくださった渡辺裕教授，そして，研究プロジェクトを共に進めてくださったシャープ株式会社の伊藤典男様，猪飼知宏様，中條健様，佐々木瑛一様，鈴木拓矢様に心から感謝します。また，研究生活を支えてくれた研究室の皆さま，家族に感謝の意を表します。

参考文献

- [1]. 総務省, 4K放送・8K放送情報サイト (最終閲覧日: 2021年1月)
[http://www.soumu.go.jp/menu_seisaku/ictseisaku/housou_suishin.html](http://www.soumu.go.jp/menu_seisaku/ictseisaku/housou_suishin/4k8k_suishin.html)
- [2]. 小峯一晃, 對馬淑亮, 澤島康仁, 比留間伸行: “陰影画像の表示解像度と立体感の関係”, NHK技研R&D, No.151, pp.47-54, May 2015.
- [3]. “地上放送高度化実験方式について”, 情報通信審議会 情報通信技術分科会 放送システム委員会 (第73回), 資料73-6, May 2020.
- [4]. 筋誠久, 市ヶ谷 敦郎: “新4K8K衛星放送の特徴・情報源符号化と伝送技術”, 通信ソサイエティマガジンB-plus, 2019夏号No.49, pp.12-17, Jun. 2019.
- [5]. ITU-T Recommendation H.266 Versatile video coding, Aug. 2020.
- [6]. S. Umeda, N. Yano, H. Watanabe, T. Ikai, T. Chujoh and N. Ito: “HDR Video Super-Resolution for Future Video Coding”, IEEE Proc. International Workshop of Advanced Image Technology 2018, No.89, Jan. 2018.
- [7]. G. Georgis, G. Lentaris, and D. Reisis: “Reduced complexity superresolution for low-bitrate video compression”, IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 2, pp. 332–345, Feb. 2016
- [8]. Y. Li, D. Liu, H. Yang, H. Li and F. Wu: “Convolutional neural network-based block up-sampling for intra frame coding”, IEEE Trans. Circuits Syst. Video Technol., vol. 28, No. 9, pp. 2316–2330, Sep. 2018.
- [9]. Y. Huang, C. Huang and T. Huang, Apparatus for reference picture resampling generation and method thereof and video decoding system using the same, US pat. 8,644,381, 2014
- [10]. MathWorks, “畳み込みニューラルネットワーク(CNN)” (最終閲覧日: 2021年1月)
<https://jp.mathworks.com/discovery/convolutional-neural-network.html>
- [11]. しこあん: “モザイク除去から学ぶ最先端のディープラーニング技術”, 技術書典8, Feb. 2020.
- [12]. W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang: “Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network”, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1874–1883, Jun. 2016.
- [13]. C. Dong, C. C. Loy, K. He, and X. Tang: “Image Super-Resolution Using Deep Convolutional Networks”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.38, No.2, pp.1-14, Jul. 2015.

- [14]. J. Kim, J. K. Lee and K. M. Lee, “Accurate Image Super-Resolution Using Very Deep Convolutional Networks”, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1646-1654, Jun. 2016.
- [15]. B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee: ”Enhanced Deep Residual Networks for Single Image Super-Resolution”, New Trends in Image Restoration and Enhancement workshop and challenge on image super-resolution 2017 (NTIRE2017 Super-Resolution Challenge), Jul. 2017.
- [16]. Y. Tian, Y. Zhang, Y. Fu and C. Xu: “TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution”, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3360-3369, Jun. 2020.
- [17]. X. Wang, K. C. K. Chan, K. Yu, C. Dong and C. C. Loy: “EDVR: Video Restoration with Enhanced Deformable Convolutional Networks”, IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), pp.1954-1963, Jun. 2019.
- [18]. M. S. M. Sajjadi, R. Vemulapalli, and M. Brown: “Frame-Recurrent Video Super-Resolution”, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.6626-6634, Jun. 2018.
- [19]. M. Chu, Y. Xie, J. Mayer, L. Leal-Taixe, and N. Thuerey: “Learning temporal coherence via self-supervision for GAN-based video generation”, arXiv preprint arXiv:1811.09393, Nov. 2018.
- [20]. R. Yang, M. Xu, Z. Wang and T. Liu: “Multi-frame quality enhancement for compressed video”, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.6664–6673, Jun. 2018.
- [21]. Z. Guan, Q. Xing, M. Xu, R. Yang, T. Liu and Z. Wang: “MFQE 2.0: a new approach for multi-frame quality enhancement on compressed video”, IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI), Jun. 2019.
- [22]. VTM-10.0 (最終閲覧日 : 2021 年 1 月)
https://vcgit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM/-/releases/VTM-10.0
- [23]. 鈴木輝彦, “フレーム内符号化とフレーム間予測符号化”, 映像情報メディア学会誌, 67 卷, 7 号, pp. 537-540, 2013
- [24]. S. Hochreiter and M. C. Mozer: “A Discrete Probabilistic Memory Model for Discovering Dependencies in Time”, Springer Berlin Heidelberg, 2001.
- [25]. G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, “Densely connected convolutional networks”, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2261-2269, Jul. 2017.
- [26]. F. Bossen, J. Boyce, K. Sühring, X. Li and V. Seregin: “JVET common test conditions and software reference configurations for SDR video”, Joint Video Experts Team

Document, JVET-N1010, Geneva, 2019.

- [27]. T. Hori, Z. Gong, H. Watanabe, T. Ikai, T. Chujoh, E. Sasaki, N. Ito, "CNN-based super-resolution adapted to quantization parameters", Proc. International Workshop on Advanced Imaging Technology (IWAIT), Vol. 11515, Jun. 2020.
- [28]. Github, MFQEv2.0 (最終閲覧日 : 2021 年 1 月)
<https://github.com/RyanXingQL/MFQEv2.0>
- [29]. L. Song, X. Tang, W. Zhang, X. Yang and P. Xia, The SJTU 4K Video Sequence Dataset, the Fifth International Workshop on Quality of Multimedia Experience (QoMEX2013), Klagenfurt, Austria, 2013.
- [30]. A. Mercat, M. Viitanen, and J. Vanne: "UVG dataset: 50/120fps 4K sequences for video codec analysis and development", Proc. 11th ACM Multimedia Systems Conference, pp.297–302, 2020.

図一覧

図 2.1	RPR による異なる解像度間でのピクチャ参照の例	6
図 2.2	RPR による符号化時のフレーム解像度を切り替え	6
図 2.3	低解像画像（入力）と超解像画像の例	7
図 2.4	畳み込みによる特徴量の抽出のイメージ	7
図 2.5	CNN の概要図	8
図 2.6	符号化ビデオのフレーム品質変動	10
図 2.7	Random Access モードにおける階層参照構造	10
図 2.8	MFQE の構造の概略図	12
図 2.9	Motion Compensation subnet の構造	13
図 2.10	Quality Enhancement subnet の構造	13
図 3.1	ダウンスケーリングと超解像を用いた映像伝送方式	15
図 3.2	解像度混在映像伝送方式の概念図（提案手法）	16
図 3.3	提案手法の概略図	18
図 3.4	全フレームを縮小する場合の概略図	18
図 4.1	各 Class の予備実験条件下での RD 曲線	23
図 5.1	Class A1, A2 シーケンスの画面内容	24
図 5.2	MFQE を用いた際の RD 曲線 (Case1, Case2, Case3)	27
図 5.3	MFQE を用いた際の RD 曲線 (Case1, Case3)	27
図 5.4	MFQE を用いた際の RD 曲線 (Case1, Case2)	28
図 5.5	CatRobot1 と DaylightRoad2 の RD 曲線	30
図 5.6	Case1 の出力画像と Case3 の MFQE 出力画像, 入力画像の比較 (DaylightLoad2)	31
図 5.7	Case1 出力画像と Case3 の MFQE 出力画像, 入力画像の比較 (Campfire)	31
図 5.8	Case1 の符号化画像と Case2 の MFQE 出力画像	32
図 5.9	RD 曲線と各測定点で最も符号化効率が高い Case	33
図 5.10	Case2 と Case3 の測定点ごとの MFQE 前後の PSNR の比較	35
図 5.11	Case2 と Case3 の MFQE 出力画像の比較	36
図 5.12	Case3 におけるフレームごとの PSNR	37

表一覧

表 1	比較する 3 つのケースの符号化条件	19
表 2	CTC の構成とシーケンスのプロパティ	21
表 3	Case1, 2, 3 符号化時の量子化パラメタ	22
表 4	Case1 の各 QP でのビットレートと PSNR の関係	28
表 5	Case2 の各 QP でのビットレートと PSNR の関係	29
表 6	Case3 の各 QP でのビットレートと PSNR の関係	29
表 7	Case2, 3 をビットレート帯に応じて適用した際の Case1 に対する BD ゲイン	34
表 8	Case2 と Case3 の MFQE 出力画像の MFQE 入力画像に対する BD ゲインの比較	35

研究業績

国際学会

- [1]. T. Hori, R. Yano, H. Watanabe, T. Chujoh, T. Ikai, E. Sasaki, T. Suzuki and N. Ito: "Multi Frame Super-Resolution for Versatile Video Coding", Proc. International Workshop on Advanced Imaging Technology (IWAIT2021), No.45, Mar. 2021. (掲載決定)
- [2]. T. Hori, Z. Gong, H. Watanabe, T. Chujoh, T. Ikai, E. Sasaki, T. Suzuki and N. Ito: "CNN-based Super-Resolution Adapted to Quantization Parameters", Proc. International Workshop on Advanced Imaging Technology (IWAIT2020), No.42, Jun. 2020.
- [3]. Z. Gong, T. Hori, H. Watanabe, T. Chujoh, T. Ikai, E. Sasaki, T. Suzuki and N. Ito: "A Selective Fusion Module for Video Super Resolution with Recurrent Architecture", Proc. International Workshop on Advanced Imaging Technology (IWAIT2020), No.43, Jun. 2020.

国内学会

- [1]. 堀隼也, 糞子臣, 梅田聖也, 渡辺裕, 中條健, 猪飼知宏, 佐々木瑛一, 伊藤典男: “符号化ノイズを含む画像への PSNR 適応超解像 (PSNR Adaptive Super-resolution for Images with Coding Noise)”, IEICE 総合大会 D-11-33, Mar. 2020
- [2]. 堀隼也, 糞子臣, 梅田聖也, 渡辺裕, 中條健, 佐々木瑛一, 猪飼知宏, 伊藤典男: “縮小拡大と符号化によるノイズを含む画像の超解像に関する検討 (A Study on Super-resolution of Noisy Images by Scaling and Coding)”, 2019 年画像符号化シンポジウム・2019 年映像メディア処理シンポジウム (PCSJ/IMPS2019), P-4-12, Nov. 2019
- [3]. 堀, 加藤, 渡辺: “学習型超解像と再構成型超解像の異なる組み合わせ順による時系列画像の画質改善 (Image Quality Improvement of Sequential Images by Different Combination Orders of Learning-based and Reconstruction-based Super Resolution)”, 映像情報メディア学会年次大会, 32B-1, Aug. 2019
- [4]. 堀, 渡辺: “超解像を用いた時系列画像の画質改善 (Quality Improvement of Sequential Images Using Super Resolution)”, 電子情報通信学会総合大会, D-11-6, Mar. 2019