Vol.2021-DPS-187 No.11
Vol.2021-MBL-99 No.11
Vol.2021-ITS-85 No.11
2021/5/27

動画配信におけるフレームの特徴量に基づく映像の超解像処理手法の提案

大石 貴之[1]　後藤 佑介[1]

概要：動画配信サービスでコンテンツを視聴する利用者の満足度は，サーバとクライアントとの間の通信環境に大きく依存する．クライアントは，サーバとの通信状況が悪い場合，動画の再生中に中断が発生する可能性がある．この再生中断を減らすため，通信状況に応じて配信動画の品質を変更する手法が提案されているが，受信映像の解像度が低下すると，視聴品質も低下する．そこで，低品質の映像を受信した場合に受信映像の各フレームに対して解像度が向上する超解像処理技術を用いることで，クライアントは高品質の映像に変換して再生できる．しかし，クライアントの計算機において，CPU やメモリといった計算資源が十分でない場合，受信した映像を構成するすべてのフレームに対して，リアルタイムに超解像処理を行うことは難しい．また，特徴量の多少に関わらず，すべてのフレームに対して超解像処理を行うため，視覚的な映像品質の向上率は小さくなる．本研究では，低品質の映像受信時の特徴量を考慮して高品質の映像をリアルタイムで再生する超解像処理手法を提案する．提案手法では，クライアントが映像をバッファリングしながら再生する場合，バッファリング済みの映像のうち特徴量が多く視覚的な品質向上の効果が高いと予測されるフレームを優先して，再生開始までの間で超解像処理を行うことで，視覚的な映像品質は向上する．フレームの知覚的類似性を用いた提案手法における映像品質の評価では，特徴量に応じてフレームを選択しない手法，および超解像処理を行わない手法とそれぞれ比較して，視覚的な映像品質が向上することを示した．

1. はじめに

近年，動画配信サービスの普及により全世界のビデオトラフィックが急増しており [1]，通信環境の変化に適応した動画配信システムが必要となっている．サーバとクライアントとの間の通信状況が悪い場合，クライアントは動画の再生中に中断が発生する可能性がある．この再生中断を無くすため，Adaptive Bitrate [2,3] と呼ばれる配信方式が提案されており，多くの動画配信サービスで採用されている．Adaptive Bitrate では，クライアントは通信状況に応じて受信する映像の品質を切り替えることで再生中断の発生を抑制できるが，サーバとの通信状況が悪い場合，クライアントが受信する映像の解像度は低下し，視聴品質も低下する．

Adaptive Bitrate における問題を解決する方法として，低解像度で受信した映像を構成するすべてのフレームに対して解像度を向上できるフレームを予測して生成することで，視聴動画の品質を向上させる超解像処理技術が挙げられる．しかし，超解像処理においてフレームの予測精度を高くすると，計算量は増加する．このため，クライアント計算機において CPU やメモリといった計算資源が十分でない場合，受信した映像を構成するすべてのフレームに対してリアルタイムに超解像処理を行うことは難しい．

リアルタイムで超解像処理を行う場合，クライアントは，映像を構成するすべてのフレームに対して，バッファリングによる受信完了から再生開始までの間の処理時間を利用して可能な限りのフレームを高解像度に変換する．このとき，特徴量が少ないフレームに対する超解像処理では，視覚的な映像品質を大きく向上できない．

本研究では，低品質の映像受信時の特徴量を考慮して高品質の映像をリアルタイムで再生する超解像処理手法を提案する．提案手法では，クライアントが一定時間分の映像をバッファリングしながらリアルタイムで再生する場合，特徴量が多いフレームに対して優先的に超解像処理を行うことで，視覚的な映像品質を向上できる．

2. 画像の拡大技術

2.1 画素補間

画像を拡大して表示する場合，元画像を拡大した画像

[1] 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University

Vol.2021-DPS-187 No.11
Vol.2021-MBL-99 No.11
Vol.2021-ITS-85 No.11
2021/5/27

図 1 鳥の原画像

図 2 3 種類の手法による鳥の拡大画像

（以下，拡大画像）を生成する必要がある．これは，連続画像である映像の表示においても同様である．拡大画像は元画像に比べて多くの画素数をもつため，元画像では存在しない画素の値を補間する必要がある．画素の補間では，ニアレストネイバ法，バイリニア法，およびバイキュービック法 [4] といった手法が主に利用され，補間画素周辺の画素情報をもとに，補間画素の画素値を求める．

図 1 に示されている鳥を写した原画像の矩形領域に対して，ニアレストネイバ法，バイリニア法，およびバイキュービック法を用いてそれぞれ 4 倍に拡大した画像を**図 2** に示す．ニアレストネイバ法では，補間画素に最も近い位置に存在する画素値を補間画素の画素値に設定して補間する．ニアレストネイバ法による補間は，補間処理が容易であるとともに，原画像の画素値を失わない利点がある．しかし，周辺画素の画素値をこのまま補間画素として利用するため，エッジにジャギーが発生する．

バイリニア法では，補間画素の周辺 4 画素をもとに，縦横両方向から直線的に補間して画素値を求める．バイリニア法による補間は，周辺画素の平均化であるため，ニアレストネイバ法に比べてエッジは滑らかになる．一方で，バイリニア法では高周波成分を生成できず，画像にぼやけが発生する．

バイキュービック法では，補間画素の周辺 16 画素をもとに，縦横両方向から 3 次式で補間して画素値を求める．バイキュービック法による補間は，バイリニア法と同様に，エッジが滑らかになる．また，バイリニア法に比べて画像のぼやけの発生を抑制できる．しかし，補間が周辺画素の平均化である点はバイリニア法と同様であるため，バイキュービック法では高周波成分を生成できず，エッジを強調できない．

図 3 SRCNN による鳥の拡大画像

2.2 超解像

画像の拡大時に高周波成分を推定して高解像度化する超解像技術が研究されている．超解像では，2.1 節で挙げた一般的な画素補間による拡大と異なり，画像の特性をもとに画像の解像度を高くする．

主な超解像手法は，複数枚の類似画像をもとに 1 枚の高解像度の画像を生成する再構成型超解像，および学習用画像を用いて高画質画像と低画質画像の対応パターンを学習する学習型超解像の 2 種類に分類される．近年，学習型超解像では，畳み込みニューラルネットワーク（以下，CNN）を用いた手法が従来手法に比べて高精度で超解像を行うことができ，多くの学習モデルが提案されている．

図 1 に示す画像の矩形領域に対して，畳み込みニューラルネットワークを用いた超解像モデルである Super-Resolution CNN (SRCNN) [5] を用いて，4 倍に拡大した画像を**図 3** に示す．図 2 と図 3 を比較すると，SRCNN による拡大では，他の 3 種類の手法と比較してエッジが強調されている．SRCNN は 3 層の CNN モデルであり，従来の CNN を用いない学習型超解像手法と比べて高精度な超解像が可能である．最近では，CNN を用いた高精度な超解像モデルとして，SRCNN に比べて畳み込み層が多いモデル [6]，および敵対的生成ネットワークを用いたモデル [7] が提案されている．

映像は，連続した単一画像である複数のフレームで構成される．このため，映像を構成する各フレームに対して 2.2 節で示した単一画像の超解像手法を適用することで，映像に対する超解像（以下，映像超解像）が可能である．しかし，映像超解像の品質は，フレームごとの超解像精度だけでなく，フレーム間の動きに対する一貫性の維持が重要である．そこで，フレーム間の動きに対して一貫性を維持する映像超解像を行う手法 [8, 9] が提案されており，高精度な映像超解像が可能である．

2.3 動画配信における超解像の利用

動画配信サービスの利用時に低解像度映像を受信する場合，受信映像に対して超解像を適用することで，高解像度の映像を再生できる．低解像度映像を受信する状況として，サーバから低解像度映像のみが配信されている場合，および Adaptive Bitrate による配信において通信状況が悪い場

Vol.2021-DPS-187 No.11
Vol.2021-MBL-99 No.11
Vol.2021-ITS-85 No.11
2021/5/27

合が挙げられる．この場合，配信動画の再生中に超解像を行うため，受信する各フレームに対してリアルタイムに超解像を行う必要がある．このため，CPU やメモリといった計算資源が十分でない場合，単純にすべてのフレームに対してリアルタイムに映像超解像を行うことは難しい．そこで，映像の特性を利用してリアルタイムに映像超解像を行う手法が提案されている．

Zhang らは，映像の圧縮に着目した手法 [10] を提案している．この手法では，Group Of Picture (GOP) に含まれるキーフレームに対してのみ超解像を行うことで，超解像されたフレームを用いて復号される他のフレームに超解像の効果が伝播し，すべてのフレームに対して超解像による影響を与えることができる．しかし，この手法は映像の符号化に依存しており，MotionJPEG といったフレーム間予測を行わない符号化を用いる場合は利用できない．また，キーフレームのみに対して超解像を行うため，超解像されたキーフレームをもとに生成されるフレームにおける超解像の精度は，フレームに超解像を直接適用した場合に比べて低くなる．

Yeo らは，計算資源に応じて深度を変更可能な深層 CNN モデル [11] を提案している．この手法を用いることで，クライアントは計算資源に応じた深度によるモデルを構築し，リアルタイムに超解像を行いながら動画を再生できる．しかし，この手法では，軽量な CNN モデルにおいてリアルタイムに超解像処理が可能な状況を想定しており，大規模な CNN モデルでリアルタイムに超解像を行う状況を想定していない．

3. 特徴量と超解像精度の関連

3.1 特徴量検出

画像の特徴を数値化した特徴量の検出に関する研究では，Features from Accelerated Segment Test（FAST）[12] や Accelerated KAZE（A-KAZE）[13] といったコーナーの特徴を高速に検出する手法が提案されている．これらの手法は，顔認識や Simultaneous Localization and Mapping（SLAM） といった対象物の特徴をリアルタイムに抽出する処理に利用される [14, 15].

街を写した元画像，および街の画像に対して FAST を用いて検出したコーナーを描画した画像を**図 4** にそれぞれ示す．また，空を写した元画像，および空の画像に対して FAST を用いて検出したコーナーを描画した画像を**図 5** にそれぞれ示す．街の画像は，建物や車といった多くの対象物で構成された複雑な画像であり，検出されるコーナーは 5,407 個である．空の画像は，空と雲のみが写った単純な画像であり，検出されるコーナーは 30 個である．

3.2 特徴量と超解像の関連

図 4 および図 5 の元画像を 0.25 倍に縮小した後，バイキュービック法および SRCNN で 4 倍に復元した画像を**図 6**，**図 7** にそれぞれ示す．また，図 6 の 2 画像と図 4 の元画像との類似度，および図 7 の 2 画像と図 5 の元画像の類似度について，評価結果を**表 1** に示す．

表 1 より，街の画像に対して SRCNN による復元画像と元画像との類似度は，バイキュービック法による類似度に比べて，3 種類すべての評価指標で高い．一方で，空の画像では，SRCNN による復元画像と元画像との類似度は，バイキュービック法による類似度に比べて，LPIPS では高くなる一方で，PSNR および SSIM では低く，評価指標に応じて異なる．また，バイキュービック法と SRCNN で比べた場合，街の画像における LPIPS による類似度の差は 0.74 となる一方で，空の画像における差は 0.006 となり，小さい．

図 4 街の元画像およびコーナーを描画した画像

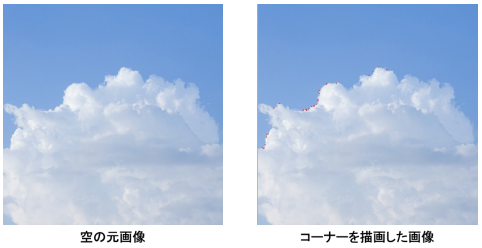

図 5 空の元画像およびコーナーを描画した画像

図 6 縮小した街の元画像に対する手法ごとの拡大画像

図 7 縮小した空の元画像に対する手法ごとの拡大画像

Vol.2021-DPS-187 No.11
Vol.2021-MBL-99 No.11
Vol.2021-ITS-85 No.11
2021/5/27

表 1 街および空の画像に対する手法ごとの復元画像と元画像の類似度

	街 (バイキュービック)	街 (SRCNN)	空 (バイキュービック)	空 (SRCNN)
PSNR	22.421	22.689	40.186	39.660
SSIM	0.615	0.628	0.954	0.948
LPIPS	0.561	0.487	0.188	0.182

図 8 図 6 の一部領域（赤枠）を拡大した画像

次に, 図 6 で示す街の画像のうち一部の領域を拡大した画像を**図 8** に示す．図 8 より，コーナーの部分では SRCNN を用いた超解像による強調効果が大きい．このため，コーナーが多い街の画像では，超解像による効果は大きい．

また，空の画像のようにコーナーが少ない場合，画素補間によるぼやけの発生が少なく，超解像による視覚的な品質向上の効果は小さい．この場合，LPIPS の差も小さくなる．

以上より，コーナーが多い画像では，超解像で拡大した場合におけるフレームの予測精度をより向上できる．

4. 提案手法

4.1 概要

本研究では，低品質の映像受信時に特徴量が多いフレームを優先して超解像処理を行いながら動画を再生する手法を提案する．提案手法では，リアルタイムに超解像処理を行うことができない環境を想定し，バッファリング済みの映像のうち，特徴量が多く超解像による視覚的な品質向上の効果が高いと予測されるフレームを優先して超解像処理を行うことで，再生する映像のフレームの視覚的な品質を向上する．本章では，提案手法において，バッファリング済みの映像に対する超解像処理の内容，およびすべてのフレームに対する超解像処理の順番の設定についてそれぞれ述べる．

4.2 バッファリング映像の超解像処理

多くの動画配信システムでは，クライアントは一定のデータを計算機にバッファリングしながら動画を再生することで，再生中断の発生を減らす．このため，リアルタイムによる動画配信では，クライアントはフレームに超解像を適用しながら映像を再生することは難しい．提案手法では，クライアントはバッファリング済みの映像に対して，特徴量が多く視覚的な品質向上の効果が高いと予測されるフレームを優先して，再生開始までの間で超解像処理を行うことで，視覚的な映像品質を向上させる．

提案手法の処理手順は，以下の通りである．

(1) バッファリング済みの複数のフレームをバッチ単位にまとめて分割
(2) 各バッチを構成するすべてのフレームでコーナーの合計を算出
(3) すべてのバッチをコーナー数が多い順番にソート

はじめに，バッファリング済みのフレームを N 枚ごとにバッチとしてまとめ，分割する．Adaptive Bitrate による動画配信では，再生映像に対する解像度の切り替え頻度が高い場合，視聴品質は低下する [16]．提案手法で用いる再生映像についても同様に，低解像度で受信した映像を構成するフレームごとに拡大させる手法が頻繁に変化すると，解像度が変化して視聴品質が低下する．そこで，提案手法では，拡大で用いる手法をバッチごとに決めることで，バッチを構成する N 枚のフレームすべてに対して同じ手法で拡大でき，視聴品質の低下を抑制できる．

次に, 各バッチを構成するすべてのフレームにおけるコーナーの合計を算出する．節で述べたように，コーナーの特徴が多いフレームは，超解像で拡大した場合に推定精度がより向上するため，超解像を行うフレームを選択する指標として用いる．提案手法では，コーナーを検出するアルゴリズムとして，コーナーの高速検出が可能な FAST [12] を用いる．

最後に，すべてのバッチをコーナー数が多い順番にソートする．提案手法では，コーナーの合計が多いバッチから順番に超解像を行う．

5. 実装

提案手法において，超解像を行いながら受信映像を再生するプレイヤの構成を**図 9** に示す．プレイヤは，受信モジュール，超解像モジュール，および再生モジュールの 3 種類で構成される．

受信モジュールでは，配信サーバから S 秒分の映像フレーム (以下，セグメント) を受信してバッファに保存する．また，2 個分のセグメントをバッファに保存すると受信を停止し，再生モジュールからの受信要求まで待機する．

超解像モジュールでは，超解像を行っているセグメントの再生が開始されると，このセグメントの超解像を終了し，バッファに保存済みで，かつ次に再生するセグメントに対して，4 章で述べた手法で超解像を行う．提案手法に

Vol.2021-DPS-187 No.11
Vol.2021-MBL-99 No.11
Vol.2021-ITS-85 No.11
2021/5/27

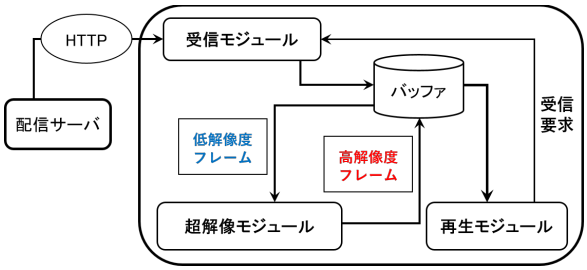

図 9 提案手法における映像再生プレイヤの構成

表 2 計算機の性能

Server	CPU	Intel®Pentium(R) CPU G4400 (3.30 GHz) × 2
	Memory	7.7 GBytes
	OS	Ubuntu 18.04.1 LTS
Client	CPU	Intel®CORE(TM) i5-7500 CPU (3.40 GHz)
	Memory	7.9 GBytes
	OS	Windows 10 Pro

おける超解像では，CNN を用いた学習型超解像モデルである FSRCNN [17] を用いて，映像フレームを 4 倍に拡大する．また，超解像を行うことができないフレームは，バイキュービック法で 4 倍に拡大する．

再生モジュールでは，1 個のセグメントの再生が終了した後に，新たなセグメントの受信を受信モジュールに要求する．受信モジュールは，再生済みのセグメントを棄却して，新たに受信したセグメントをバッファに保存することで，バッファに 2 個分のセグメントが常に保存された状態にする．受信モジュールと配信サーバの通信プロトコルは，Adaptive Bitrate 配信で主に利用される HTTP を用いる．

6. 評価

6.1 評価環境

5 章で示したプレイヤを実装した計算機と動画配信を行うサーバを Ethernet で接続し，提案手法を評価する．動画配信を行うサーバでは，ソフトウェアとして Apache HTTP Server [18] を用いた．評価に用いた計算機の性能を**表 2** に示す．クライアントとサーバは，映像の再生に十分な速度で通信可能である．また，クライアントは，映像の再生を開始すると最後まで再生する．

6.2 評価に用いる映像

評価に用いる 3 種類の映像を**表 3** に示す．すべての映像は，開始から 10 分間の映像データをトリミングして用いる．Tears of Steel [19] は，実写と CG が混在し，フレームの時間的変化が大きい映像である．また，他の 2 種類の映像とアスペクト比を揃えるため，左右を切り取りアスペクト比を 16 : 9 にクロップした映像を用いる．Big Buck Bunny [20] は，アニメーション映像である．Herzmark Homestead [21] は，ドローンで森を空中から映し続けた映像であり，フレームの時間的変化が小さい．

表 3 評価に用いる映像の構成

タイトル	再生時間	解像度 (pixel)
Tears of Stee [19]	10 min.	144 x 256 (144p) 180 x 320 (180p) 270 x 480 (270p)
Big Buck Bunny [20]	10 min.	144 x 256 (144p) 180 x 320 (180p) 270 x 480 (270p)
Herzmark Homestead [21]	10 min.	144 x 256 (144p) 180 x 320 (180p) 270 x 480 (270p)

6.3 映像の種類による映像品質への影響

提案手法，映像の再生時間が早いフレームから優先して超解像処理を行う手法 (以下，単純手法)，およびすべてのフレームに対して画素をバイキュービック法で補間して拡大する手法 (以下，BiC 手法) の 3 種類について，再生する映像を構成するすべてのフレームに対する品質の平均値を評価する．評価に用いる映像のフレームレートは 24 fps，各セグメントの映像時間は 20 秒，各バッチのフレーム数は 10 枚とする．評価項目は，画像の圧縮において変換後の画像が元の画像からどの程度劣化したかを客観的に評価する指標の一つである Peak Signal-to-Noise Ratio (PSNR)，評価に用いる画素の類似度を示す Structural Similarity Index Measure (SSIM) [22]，および人の知覚的類似性を学習させたニューラルネットワークによる評価値である Learned Perceptual Image Patch Similarity (LPIPS) [23] の 3 種類の各平均値である．

9 種類の映像で評価を行った結果を**表 4** に示す．表 4 より，Tears of Steel および Big Buck Bunny の場合，すべての解像度の映像に対して，提案手法はすべての評価項目で最も高い．一方で，Helzmark Homestead の場合，180p の映像における平均 SSIM 以外の評価項目について，単純手法が最も高い．また，BiC 手法は，すべての評価項目について，提案手法に比べて低い．一方で，Tears of Steel の 270p の映像，および Big Buck Bunny の 144p と 180p の映像では，BiC 手法は単純手法に比べて高い．

6.4 映像の解像度による超解像フレーム数への影響

提案手法と単純手法において，超解像処理が行われたフレーム数を評価する．評価に用いる映像のフレームレートは 24 fps であり，各セグメントの映像時間を 20 秒，各バッチのフレーム数は 10 枚とする．

3 種類の映像で評価した結果を**表 5** に示す．表 5 より，提案手法および単純手法において，受信する映像の解像度

表 4　各視聴映像の品質評価

受信映像		提案手法			単純手法			BiC 手法		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Tears of Steel	144p	**29.851**	**0.843**	**0.222**	29.799	0.841	0.227	29.695	0.837	0.272
	180p	**31.214**	**0.861**	**0.210**	31.144	0.859	0.217	31.122	0.857	0.240
	270p	**32.730**	**0.885**	**0.187**	32.660	0.884	0.193	32.702	0.884	0.198
Big Buck Bunny	144p	**28.152**	**0.799**	**0.271**	27.452	0.795	0.279	28.101	0.790	0.321
	180p	**29.075**	**0.818**	**0.257**	28.203	0.815	0.268	29.045	0.811	0.290
	270p	**30.176**	**0.850**	**0.223**	30.152	0.848	0.231	30.151	0.848	0.238
Helzmark Homestead	144p	20.478	0.439	0.561	**20.484**	**0.439**	**0.557**	20.377	0.407	0.608
	180p	20.278	**0.433**	0.570	**20.308**	0.432	**0.567**	20.216	0.412	0.596
	270p	20.013	0.434	0.560	**20.039**	**0.435**	**0.559**	19.982	0.425	0.572

表 5　視聴映像の超解像フレーム数

受信映像		超解像フレーム数	
		提案手法	単純手法
Tears of Steel	144p	10,388	**10,778**
	180p	6,409	**6,541**
	270p	2,659	**2,757**
Big Buck Bunny	144p	10,101	**10,907**
	180p	6,360	**6,658**
	270p	2,597	**2,728**
Herzmark Homestead	144p	9,568	**10,150**
	180p	5,868	**6,112**
	270p	2,382	**2,452**

が高いほど超解像が行われたフレーム数は少ない．また，提案手法で超解像が行われたフレーム数は，単純手法と比べて，144p の映像では平均して約 609 フレーム，180p の映像では平均して約 225 フレーム，270p の映像では平均して約 100 フレーム少ない．

6.5　映像のフレームレートによる映像品質への影響

提案手法，単純手法，およびバイキュービック手法において，フレームレートが異なる 3 種類の映像を再生した場合の視聴品質を評価する．評価には，解像度が 144p，フレームレートが 24 fps，30 fps，60 fps の 3 種類の Big Buck Bunny を用いる．また，各セグメントの映像時間を 20 秒，各バッチのフレーム数は 10 枚とする．評価項目は，平均 PSNR，平均 SSIM，平均 LPIPS の 3 種類である．

3 種類のフレームレートの映像で評価を行った結果を**表 6** に示す．表 6 より，すべてのフレームレートの映像におけるすべての評価項目において，提案手法が最も高い．また，SSIM および LPIPS による評価において，提案手法と単純手法は，フレームレートが高くなると評価は低くなる．しかし，PSNR による評価では，提案手法ではフレームレートが高くなると評価が低くなる一方で，単純手法では，24 fps および 30 fps の映像に比べて 60 fps の映像における評価が高い．

6.6　バッチのフレーム数による映像の視聴品質への影響

提案手法において，バッチを構成するフレーム数の変化に応じて，再生映像における拡大手法の変化回数および映像品質を評価する．評価には，解像度が 144p，フレームレートが 24 fps の Big Buck Bunny を用いる．また，各セグメントの映像時間を 20 秒とする．評価項目は，再生映像における拡大手法の変化回数および平均 LPIPS である．

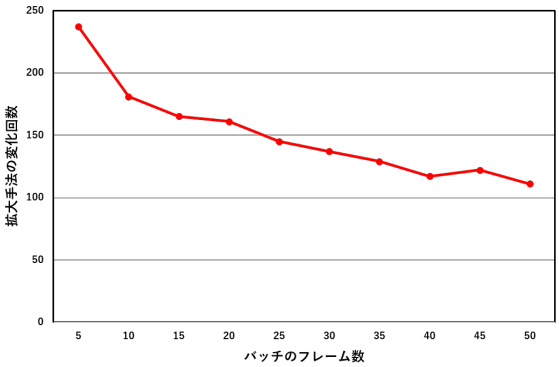

図 10　バッチのフレーム数に対する拡大手法の変化回数

はじめに，バッチのフレーム数に応じた拡大手法の変化回数を**図 10** に示す．横軸はバッチのフレーム数，縦軸は再生映像における拡大手法の変化回数である．図 10 より，バッチのフレーム数が大きくなるほど，再生映像における拡大手法の変化回数は少ない．例えば，バッチのフレーム数が 50 の場合，拡大手法の変化回数は 111 回となり，一番少ない．

次に，バッチのフレーム数に応じた平均 LPIPS の評価を**図 11** に示す．横軸はバッチのフレーム数，縦軸は全フレームの平均 LPIPS である．図 11 より，バッチのフレーム数が大きくなるほど平均 LPIPS は大きくなり，映像品質は低下する．例えば，バッチのフレーム数が 5 の場合は平均 LPIPS が 0.2687，バッチのフレーム数が 50 の場合は平均 LPIPS が 0.2701 である．

6.7　セグメントの映像時間による映像品質への影響

提案手法において，セグメントの映像時間の変化に応じた視聴品質を評価する．評価では，解像度が 144p，フレームレートが 24 fps の Big Buck Bunny を用いる．また，各バッチのフレーム数は 10 枚とする．評価項目は，平均 LPIPS を用いる．

セグメントの映像時間に応じた平均 LPIPS の評価を**図 12** に示す．横軸はセグメントの映像時間，縦軸は平均 LPIPS である．図 12 より，セグメントの映像時間が長

Vol.2021-DPS-187 No.11
Vol.2021-MBL-99 No.11
Vol.2021-ITS-85 No.11
2021/5/27

表 6　各フレームレートの視聴映像における品質評価

受信映像		提案手法			単純手法			BiC 手法		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Big Buck Bunny	24fps	**28.153**	**0.799**	**0.271**	27,452	0.795	0.279	28.101	0.790	0.320
	30fps	**28.044**	**0.795**	**0.283**	27.377	0.791	0.295	28.001	0.788	0.325
	60fps	**28.013**	**0.786**	**0.314**	27.581	0.783	0.322	27.993	0.783	0.333

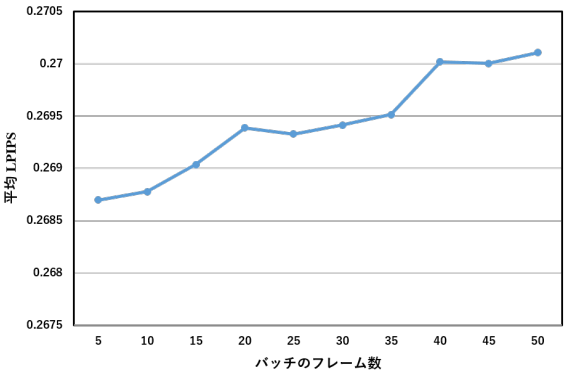

図 11　バッチのフレーム数に対する平均 LPIPS

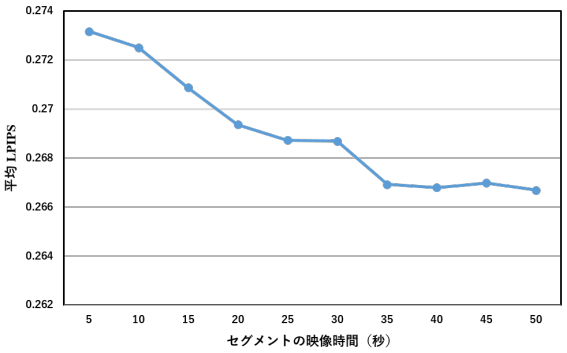

図 12　セグメントの映像時間に対する平均 LPIPS

くなるほど平均 LPIPS は小さくなり，映像品質は向上する．例えば，セグメントの映像時間が 50 秒の場合，平均 LPIPS は 0.2667 となり，映像品質は最も高い．

7. 考察

7.1 映像の種類による映像品質への影響

6.3 節の評価結果より，Tears of Steel および Big Buck Bunny では，すべての解像度の映像におけるすべての評価項目において，提案手法が最も高い．提案手法では，特徴量に基づいて超解像フレームを選択するため，他の 2 種類の手法に比べて各フレームの視覚的な平均品質は向上する．

Helzmark Homestead について，180p の映像における SSIM 以外の評価では，単純手法が最も高い．Tears of Steel および Big Buck Bunny は時間的変化が大きい映像である一方で，Helzmark Homestead は時間的変化が小さい映像であり，提案手法による超解像フレームの選択効果は小さい．以上より，提案手法は，時間的変化が少ない映像に対して有用性が低いと考えられる．

バイキュービック手法と単純手法に対して PSNR による評価を比較すると，Tears of Steel における 270p の映像，および Big Buck Bunny における 144p と 180p の映像について，BiC 手法が高い．PSNR は，対応するピクセルの画素値を単純に比較して評価する指標であり，バイキュービック法では，周辺画素の平均化によって画素値を補間する．このため，エッジやコーナーが少なく，画素値が近いピクセルが集まるフレームでは，FSRCNN による超解像に比べてバイキュービック法による拡大フレームの PSNR が高くなる場合があるためと考えられる．一方で，BiC 手法と提案手法に対して PSNR による評価を比較すると，すべての評価項目において BiC 手法が低い．特徴量が多いフレームを選択したことで，超解像による PSNR の向上が大きいフレームを選択できているためと考えられる．

7.2 映像の解像度による超解像フレーム数への影響

6.4 節の評価結果より，提案手法および単純手法において，映像の解像度が高いほどフレーム 1 枚あたりの超解像処理に必要な時間が長くなる．このため，受信する映像の解像度が高いほど，超解像が行われたフレーム数は少ない．また，提案手法において超解像が行われたフレーム数は，単純手法と比較して，144p の映像では平均して約 609 フレーム，180p の映像では平均して約 225 フレーム，270p の映像では平均して約 100 フレーム少ない．しかし，6.3 節の評価結果より，Tears of Steel および Big Buck Bunny では，すべての解像度の映像におけるすべての評価項目において，提案手法が単純手法に比べて高い．以上より，提案手法では，超解像フレーム数は少ない一方でフレームの平均品質は向上しており，超解像による視覚的な品質向上の効果が高いフレームを選択できている．

7.3 映像のフレームレートによる映像品質への影響

6.5 節の評価結果より，すべての評価項目において，フレームレートに関係なく提案手法が最も高い．また，PSNR による評価について，提案手法では，フレームレートが高くなると評価が低くなる一方で，単純手法では，24 fps や 30 fps の映像に比べて 60 fps の映像の評価が高い．エッジやコーナーが少なく，画素値が近いピクセルが集まるフレームでは，FSRCNN に比べてバイキュービック法による拡大フレームの PSNR が高くなる場合がある．このため，60 fps の映像では，FSRCNN でなくバイキュービック法で拡大されるフレームが増加し，単純手法における平均 PSNR が向上したと考えられる．

8. おわりに

本研究では，低品質な映像の受信時に特徴量が多いフレームを優先して超解像処理を行いながら動画を再生する手法を提案した．提案手法では，クライアントは一定時間分の映像をバッファリングしながら再生する．このとき，バッファリング中の映像は，再生が開始されるまでの間で，特徴量が多く視覚的な品質向上の効果が高いと予測されるフレームを優先して超解像処理を行う．

評価では，提案手法，特徴量に基づいて超解像フレームを選択しない手法，全てのフレームをバイキュービック法によって拡大する手法の 3 種類を用いて，配信映像に応じた視聴映像の視覚的な品質について，再生フレームの平均 PSNR，平均 SSIM，および平均 LPIPS で比較した．評価の結果，時間的変化が大きい映像の再生時は，解像度やフレームレートに関係なく，提案手法が他の 2 種類の手法と比べて視覚的な品質が高いことを示した．

今後の予定として，クライアントによる提案手法の有用性評価，および画像単位で領域に応じて既存の複数の超解像処理手法を利用する超解像処理手法の提案を行う．

謝辞

本研究は，文部科学省科学研究費補助金（基盤研究（C））（課題番号：18K11265）の研究助成によるものである．ここに記して謝意を表す．

参考文献

[1] Cisco Annual Internet Report (2018-2023) White Paper - Cisco (online), available from <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>（accessed 2021-03-21）.
[2] RFC8216 - HTTP Live Streaming: IETF (online), available from <https://tools.ietf.org/html/rfc8216>（accessed 2021-03-21）.
[3] Information Technology - Dynamic Adaptive Streaming over HTTP (DASH) - Part 1: Media Presentation Description and Segment Formats: ISO (online), available from <https://www.iso.org/standard/75485.html>（accessed 2021-03-21）.
[4] R. Keys: Cubic Convolution Interpolation for Digital Image Processing, IEEE Trans. Acoustic, Speech and Signal Processing, Vol.29, pp.1153-1160 (1981).
[5] C. Dong, C.C. Loy, K. He, and X. Tang: Learning a Deep Convolutional Network for Image Super-Resolution, Proceedings of the European Conference on Computer Vision (ECCV), pp.184-199 (2014).
[6] J. Kim, J. K. Lee, and K. M. Lee: Accurate Image Super-Resolution Using Very Deep Convolutional Networks, IEEE Conference on Computer Vision and Pattern Recognition, pp.1646-1654 (2016).
[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z.Wang and W. Shi: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, IEEE Conference on Computer Vision and Pattern Recognition (2017).
[8] M.S. Sajjadi, R. Vemulapalli, and M. Brown: Frame-Recurrent Video Super-Resolution, IEEE Conference on Computer Vision and Pattern Recognition, pp.6626-6634 (2018).
[9] M. Chu, Y. Xie, J. Mayer, L. Leal-Taixé, and N. Thuerey: Learning Temporal Coherence via Self-Supervision for GAN-based Video Generation, ACM Trans. Graphics, Vol.39, Issue 4, https://doi.org/10.1145/3386569.3392457.
[10] Z. Zhang and V. Sze: Fast: A Framework to Accelerate Superresolution Processing on Compressed Videos, IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.1015-1024 (2017).
[11] H. Yeo, Y. Jung, J. Kim, J. Shin and D. Han: Neural Adaptive Content-aware Internet Video Delivery, USENIX Symposium on Operating Systems Design and Implementation, pp.645-661 (2018).
[12] E. Rosten and T. Drummond: Machine Learning for High-Speed Corner Detection, European Conference on Computer Vision, pp.430-443 (2006).
[13] P.F. Alcantarilla, J. Nuevo, and A. Bartoli: Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, Proceedings of the British Machine Vision Conference (BMVC), pp.13.1-13.11 (2013).
[14] A. Vinay, A.S. Cholin, A.D. Bhat, K.N.B. Murthy, and S. Natarajan: An Efficient ORB Based Face Recognition Framework for Human-Robot Interaction, Procedia Computer Science, Vol.133, pp.913-923 (2018).
[15] Y. Li, N. Brasch, Y. Wang, N. Navab, and F. Tombari: Structure-SLAM: Low-Drift Monocular SLAM in Indoor Environments, IEEE Robotics and Automation Letters Vol.5, Issue 4, pp.6583-6590 (2020).
[16] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner: Assessing Effect Sizes of Influence Factors Towards a QoE Model for HTTP Adaptive Streaming, Proceedings of the 6th International Workshop on Quality of Multimedia Experience (QoMEX 2014), pp.111-116 (2014).
[17] C. Dong, C.C. Loy, and X. Tang: Accelerating the Super-Resolution Convolutional Neural Network, European Conference on Computer Vision, pp.391-407 (2016).
[18] The Apache HTTP Server Project (online), available from <https://httpd.apache.org/>（accessed 2021-03-21）.
[19] Tears of Steel - Mango Open Movie Project (online), available from <https://archive.org/details/Tears-of-Steel>（accessed 2021-03-21）.
[20] Big Buck Bunny (online), available from <https://download.blender.org/peach/bigbuckbunny_movies/>（accessed 2021-03-21）.
[21] Herzmark Homestead on Vimeo (online), available from <https://vimeo.com/226057477/>（accessed 2021-03-21）.
[22] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli: Image Quality Assessment: From Error Measurement to Structural Similarity, IEEE Transactions on Image Processing, Vol.13, pp.600-612 (2004).
[23] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, and O. Wang: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, Computer Vision and Pattern Recognition (2018).