情報処理学会論文誌 Vol.63 No.11 1621–1630 (Nov. 2022)

推薦論文

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

大石 貴之[1] 後藤 佑介[2,a)]

受付日 2022年2月4日, 採録日 2022年8月8日

概要：動画配信サービスでコンテンツを視聴する場合，受信映像の解像度が低下するとクライアントの視聴意欲は低下する．そこで，低品質の映像を受信した場合に受信映像を構成する各フレームの解像度を向上させ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を視聴できる．しかし，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．本論文では，低品質の映像受信時にフレーム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，再生開始までの間で，バッファに保存された映像から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ことで，視覚的な映像品質を向上させる．人間の知覚との類似度を判定する映像品質の評価では，映像の先頭フレームから順番に超解像処理を行う手法，および超解像処理を行わず画素補間で映像品質を向上させる手法と比較して，提案手法では視覚的な映像品質が向上することを示した．

キーワード：特徴量，フレーム，映像品質，映像超解像

A Processing Method for Video Super-Resolution Considering Feature Value of Frame

Takayuki Oishi[1] Yusuke Gotoh[2,a)]

Received: February 4, 2022, Accepted: August 8, 2022

Abstract: When clients watch contents in a video delivery service, their motivation to watch them decreases as the resolution of the received video decreases. To solve this problem, clients can watch high-quality video by applying the super-resolution processing methods that increases the resolution of each frame for low-quality video. However, if the client does not have sufficient computing resources, such as CPU and memory, it is difficult to perform super-resolution processing in real time for all the frames that constitute the received video. In this paper, we propose a super-resolution processing method for watching high-quality video in real time. In the proposed method, the client can improve the visual quality of the buffered video by prioritizing the frames that have more features and that are predicted to improve the visual quality. The evaluation result shows that the proposed method improves the visual video quality compared to the conventional methods.

Keywords: Features, frames, video quality, video super-resolution

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

[1] 岡山大学大学院自然科学研究科
Graduate School of Natural Science and Technology, Okayama University, Okayama 700–8530, Japan
[2] 岡山大学学術研究院自然科学学域
Faculty of Natural Science and Technology, Okayama University, Okayama 700–8530, Japan
[a)] y-gotoh@okayama-u.ac.jp

本論文の内容は 2021 年 5 月の第 187 回 DPS 研究発表会で報告され，同研究会主査により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である．

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．この再生中断をなくすため，Adaptive bitrate（ABR）[2], [3] と呼ばれるストリーミング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．ABR 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．

ABR 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変換することで高品質の映像を再生できる超解像処理技術があげられる．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，バッファに保存してから再生開始までの間で，できるだけ多くのフレームを高解像度に変換する必要がある．しかし，従来の超解像処理では，特徴量の多少を考慮せず映像の先頭フレームから順番に超解像を行うため，特徴量が少ないフレーム数が多い場合は視覚的な映像品質を大きく向上できない．

我々の研究グループでは，動画配信における映像の超解像処理技術の構築に取り組んでいる [4]．本論文では，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．

本論文は，以下のように構成される．2 章で高解像度画像の生成技術について述べ，3 章で特徴量と超解像精度の関係を説明する．4 章では提案手法について述べ，5 章で評価を行い，6 章で本論文をまとめる．

2. 高解像度画像の生成技術

2.1 画素補間

画像を拡大して高解像度化する場合，元画像を拡大した画像（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．拡大画像の画素数は元画像に比べて多くなるため，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 [5] といった手法が主に利用される．候補となる画素

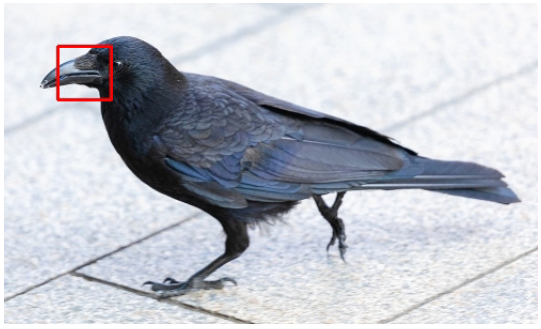


図 1 鳥の元画像
Fig. 1 Original image of bird.

図 2 画素補間による鳥の拡大画像
Fig. 2 Enlarged image of bird using interpolation: (A) Nearest neighbor interpolation, (B) Bilinear interpolation, and (C) Bicubic interpolation.

では，周辺の画素値をもとに補間することで，拡大画像の画素値を求める．

鳥を写した元画像を**図 1** に示す．また，元画像のうち赤色で示す矩形領域に対して，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といった従来手法を用いて，この領域を 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，元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，周辺画素を平均化するため，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とな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
Fig. 3 Enlarged image of bird using SRCNN.

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 節であげた画素補間による一般的な拡大手法と異なり，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，CNN を用いた超解像モデルである Super-Resolution CNN（SRCNN）[6] を用いて，この領域を 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 [7]，および敵対的生成ネットワークを用いたモデル [8] 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の精度だけでなく，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 [9], [10] 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の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．低解像度の映像を受信する状況として，サーバから低解像度の映像のみが配信されている場合，および ABR によるストリーミング配信において通信状況が悪い場合があげられる．この場合，クライアントは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で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で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 [11] 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，Motion JPEG [12] 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，超解像を行ったキーフレームをもとに生成されたフレームに対する超解像の精度は，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 [13] 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，この手法は計算資源に応じて CNN の深度を選択するのみであり，各フレームの特徴に応じて CNN を選択しない．

3. 特徴量と超解像精度

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）[14] や Accelerated KAZE（A-KAZE）[15] といったコーナ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で利用される [16], [17].

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を描画した画像を**図 4** に示す．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を描画した画像を**図 5** に示す．街の画像は，建物や車といった多くの対象物で構成されており，検出されたコーナ数は 5,407 個である．一方で，空の画像は，空と雲のみで構成されており，検出されたコーナ数は 30 個である．

3.2 元画像と復元画像との類似度評価

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

(A) 街の元画像　(B) コーナーを描画した画像

図 4　街の元画像およびコーナを描画した画像

Fig. 4　Original image with city and image with corners: (A) Original image with city and (B) Image with corners.

(A) 空の元画像　(B) コーナーを描画した画像

図 5　空の元画像およびコーナを描画した画像

Fig. 5　Original image with sky and image with corners: (A) Original image with sky and (B) Image with corners.

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 における評価項目は，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 1 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）[18]，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）[19] の 3 種類である．PSNR および SSIM では評価値が高いほど類似度が高く，LPIPS では評価値が低いほど類似度が高い．

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と比較して，3 種類すべての評価指標で高い．一方で，空の画像に対して 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と比較して，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約 0.074 となる一方で，空の画像における LPIPS による類似度の差は約 0.006 となり，小さい．

次に，バイキュービック法および SRCNN でそれぞれ拡大した街の画像のうち赤枠で示す領域を右下部で拡大した画像を**図 8** に示す．図 8 より，コーナの部分では SRCNN

(A) バイキュービック法　(B) SRCNN

図 6　縮小した街の元画像に対する手法ごとの拡大画像

Fig. 6　Enlarged image for each method for reduced original image of city: (A) Bicubic interpolation and (B) SRCNN.

(A) バイキュービック法　(B) SRCNN

図 7　縮小した空の元画像に対する手法ごとの拡大画像

Fig. 7　Enlarged image for each method for reduced original image of sky: (A) Bicubic interpolation and (B) SRCNN.

を用いた超解像による強調効果が大きい．このため，コーナ数が多い街の画像では，超解像による効果は大きい．また，コーナ数が少ない空の画像では，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 による類似度の差は小さくなる．以上より，コーナ数が多い画像では，超解像で拡大することでフレームの予測精度をより向上できる．

4. 提案手法

4.1 バッファ保存の映像に対する超解像処理

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．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，再生中断の発生回数を減少させる．リアルタイムによる動画配信では，このバッファリング時間内で動画を構成するフレームに超解像を適用する必要がある．クライアント計算機で CPU やメモリといった計算資源が十分でない場合，すべてのフレームに超解像を適用した映像再生において，バッファリング時間が短くなる．このとき，クライアントは超解像処理を行う時間を確保できず，視覚的な映像品質を十分に向上できない可能性が高くなる．

提案手法では，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時間を十分に確保できない環境を想定

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
Table 1 Similarity between restored and original images in each method for city and sky images.

	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
Fig. 8 Enlarged image of part of area (red frame) in Figure 6: (A) Bicubic interpolation and (B) SRCNN.

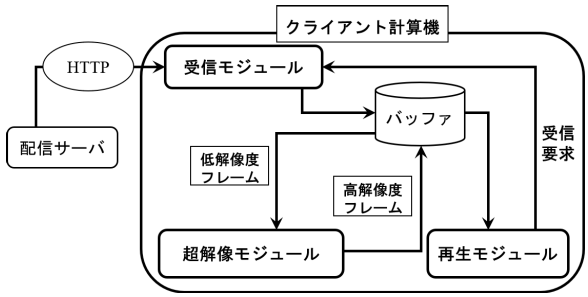


図 9 超解像映像の再生プレイヤの構成
Fig. 9 Configuration of video player for super-resolution images.

する．クライアントは，バッファに保存した映像に対して，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．

4.2 提案手法の処理手順

提案手法の処理手順は，以下のとおりである．

(1) バッファに保存した複数フレームをバッチ単位でとりまとめ
(2) バッチに含まれるすべてのフレームに対してコーナ数の合計を算出
(3) コーナ数の合計が多い順番でバッチをソート

本論文では，バッファに保存したすべてのフレームのうち映像の再生時間が早い順番で連続した一定数のフレームを 1 個のバッチで処理できるように，バッチ単位で分割する．再生するバッチが切り替わる契機における解像度の変化による視聴品質への影響について，バッチのフレーム数およびセグメントの映像時間に応じた LPIPS の変化は，5.6 節および 5.7 節でそれぞれ評価する．

ABR 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 [20]．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に対してフレームの拡大画像を生成する手法がフレーム単位で頻繁に変化すると，映像の再生中に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，フレームの拡大画像を生成する手法をバッチごとに決定することで，バッチを構成する複数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対してコーナ数の合計を算出する．3.2 節で述べたように，コーナの特徴が多いフレームは，超解像で拡大した場合に予測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，コーナを検出するアルゴリズムとして，コーナの高速検出が可能な FAST [14] を用いる．

最後に，すべてのバッチをコーナ数が多い順番にソートする．提案手法では，再生開始までの間でコーナの合計が多いバッチから順番に超解像を行うことで，映像品質をできるだけ向上する．

4.3 映像再生プレイヤ

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，配信サーバから一定時間分の映像フレーム（以下，セグメント）を受信して，バッファに保存する．バッファに保存できるセグメントの上限数を保存すると受信を停止し，再生モジュールから要求されるまで受信開始を待機する．

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，バッファに保存しており次に再生するセグメントに対して，提案手法で超解像を行う．提案手法における超解像では，CNN を用いた学習型超解像モデルである Fast SRCNN（FSRCNN）[21] を用いて，映像フレームを 4 倍に拡大する．また，バッファに保存してから再生開始までの時間が短くリアルタイムに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．

表 2 計算機の性能
Table 2 Computer performance.

Server	CPU	Intel®Pentium® 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

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，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，バッファに保存できる上限数のセグメントをできるだけ保存している状態にする．受信モジュールと配信サーバの通信プロトコルは，ABR によるストリーミング配信で主に利用される HTTP を用いる．

5. 評価

5.1 評価環境

提案手法を評価するため，4.3 節で述べた映像再生プレイヤを導入したクライアント計算機と動画配信を行うサーバは，ルータを介して同一のネットワーク上で Gigabit Ethernet により接続した．動画配信を行うサーバでは，ソフトウェアとして Apache HTTP Server [22] を用いた．評価に用いた計算機の性能を**表 2** に示す．サーバとクライアントは，映像の再生に十分な速度で通信できる．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．評価対象となる手法として，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 種類を用いる．これらの手法は，図 9 に示す超解像映像の再生プレイヤを構成する超解像モジュールに対してそれぞれ適用される．提案手法は，超解像モジュール内で各バッチを構成するすべてのフレームに対してコーナ数の合計を算出し，コーナ数が多いフレームを優先して超解像処理を行う．評価項目は，3.2 節で用いた PSNR，SSIM，LPIPS の 3 種類であり，すべてのフレームに対する平均値である．

5.2 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 [23] 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．映像を構成するシーンの切替え回数が他の映像に比べて多いため，多くの物体が複雑に配置された画像ではコーナ数が多くなる一方で，物体が少なく単純に配置された画像ではコーナ数が少なくなる．また，他の 2 種類の映像とアスペクト比を揃えるため，映像の左右部分を切り取りアスペクト比を 16:9 とした映像を用いる．

表 3 評価に用いる視聴映像の構成
Table 3 Composition of video image.

視聴映像	再生時間	解像度 (pixel)
Tears of Steel [23]	10 min.	144 x 256 (144 p)
		180 x 320 (180 p)
		270 x 480 (270 p)
Big Buck Bunny [24]	10 min.	144 x 256 (144 p)
		180 x 320 (180 p)
		270 x 480 (270 p)
Herzmark Homestead [25]	10 min.	144 x 256 (144 p)
		180 x 320 (180 p)
		270 x 480 (270 p)

Big Buck Bunny [24] はアニメーション映像であり，キャラクタの輪郭および木の模様は複雑である一方で，空やキャラクタの模様は単純である．Tears of Steel と同様に，フレームの時間的変化が大きい．また，アニメーション映像では単色で着色された物体が多くなるため，映像を構成するコーナ数は他の映像に比べて少ない．

Herzmark Homestead [25] 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．森や木々といった自然物で構成されるため，映像を構成するシーンの切替え回数は Tears of Steel に比べて少なく，多くのコーナ数を持つフレームが多い．

5.3 映像の種類による映像品質への影響

提案手法，単純手法，および 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 とする．

3 種類の映像に対して 3 種類のフレームサイズを設定し，合計 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 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，提案手法による映像品質が最も高い．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，180 p 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

表 4 視聴映像の種類に応じた品質評価
Table 4 Quality evaluation based on type of video images.

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

表 5 視聴映像のフレームレートに応じた手法ごとの品質評価
Table 5 Quality evaluation of each method based on frame rate of video image.

視聴映像		提案手法			単純手法			BiC 手法		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Big Buck Bunny	24 fps	**28.153**	**0.799**	**0.271**	27.452	0.795	0.279	28.101	0.790	0.320
	30 fps	**28.044**	**0.795**	**0.283**	27.377	0.791	0.295	28.001	0.788	0.325
	60 fps	**28.013**	**0.786**	**0.314**	27.581	0.783	0.322	27.993	0.783	0.333

い．したがって，提案手法は，時間的変化が多い映像に対して有用性が高いことが分かる．

単純手法および BiC 手法で PSNR による評価を比較すると，Tears of Steel における 270 p の映像，および Big Buck Bunny における 144 p と 180 p の映像について，BiC 手法による映像品質が単純手法に比べて高い．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．

次に，提案手法と BiC 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，提案手法による映像品質が BiC 手法に比べて高い．提案手法では，特徴量が多いフレームを選択することで，超解像による PSNR の向上率が高くなるフレームを選択できていることが分かる．

5.4 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 BiC 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 144 p，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 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 5** に示す．表 5 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による映像品質が最も高い．

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値は低くなる．しかし，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値が低くなる一方で，単純手法では，24 fps や 30 fps の映像に比べて 60 fps の映像に対する評価値が高い．エッジやコーナが少なく，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するフレーム数が増加し，単純手法による平均 PSNR が向上したことが分かる．

5.5 映像の解像度による超解像フレーム数への影響

提案手法および単純手法において，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 とする．

3 種類の映像で評価した結果を**表 6** に示す．表 6 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り，超解像が行われたフレーム数は少なくなる．

提案手法において，3 種類の映像に対して超解像が行われたフレーム数の平均は，単純手法と比較して，144 p の映像では約 609 フレーム，180 p の映像では約 225 フレーム，270 p の映像では約 100 フレームの差で少ない．しか

表 6 視聴映像に対する手法ごとの超解像フレーム数
Table 6 Number of super-resolution frames for each method for video image.

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

し，5.3 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，提案手法による映像品質が単純手法に比べて高い．したがって，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ことが分かる．

5.6 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 144 p，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．

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くなる．たとえば，バッチのフレーム数が 10 の場合における拡大手法の変化回数は 181 回となる一方で，フレーム数が 50 の場合で 111 回となり，約 38.7%減少する．提案手法では，同一の手法でバッチ内のフレームを拡大するため，バッチのフレーム数を大きくすることで，拡大手法の変化回数の増加を抑制できる．

次に，提案手法において，バッチのフレーム数の変化に応じて人間の知覚的類似性に対する影響を評価する．**図 11** に，バッチのフレーム数に応じた平均 LPIPS の評価を示す．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．たとえば，バッチのフレーム数が 5 の場合は平均 LPIPS が 0.2687 となる一方で，バッチのフレーム数が 50 の場合は平均 LPIPS が 0.2701 となり，平均 LPIPS は大きくなる．

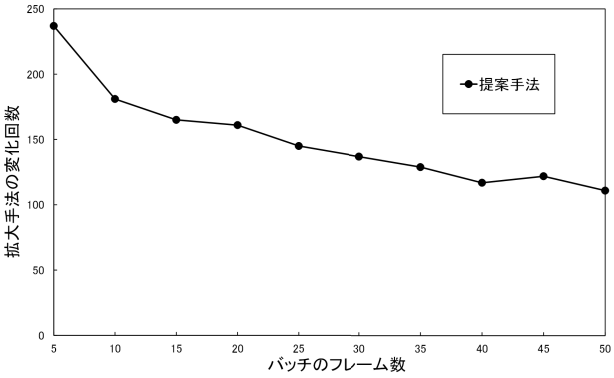


図 10 バッチのフレーム数に対する拡大手法の変化回数
Fig. 10 Number of changing enlargement methods and number of frames for each batch.

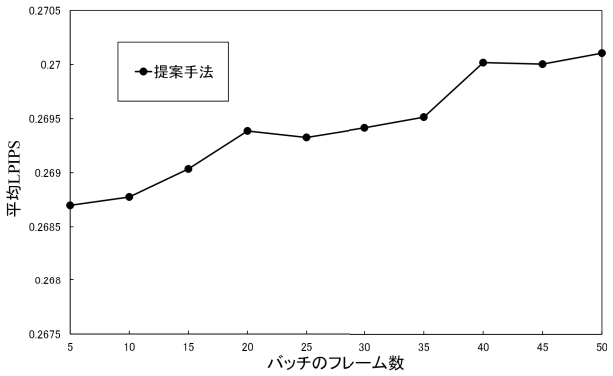


図 11 バッチのフレーム数に対する平均 LPIPS
Fig. 11 Number of frames for each batch and average LPIPS.

提案手法では，バッチのフレーム数が大きくなると各セグメントにおけるバッチ数が減少し，より少ない数のバッチから超解像を優先するバッチを選択するため，映像品質は低下する．このため，クライアントが要求する映像の視聴品質に応じて，バッチのフレーム数を決定する必要がある．

5.7 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 144 p，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 とする．評価項目は，平均 LPIPS を用いる．

提案手法において，セグメントの映像時間の変化に応じて人間の知覚的類似性に対する影響を評価する．**図 12** に，セグメントの映像時間に応じた平均 LPIPS の評価を示す．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くなるほど平均 LPIPS は小さくなり，映像品質は向上する．たとえば，セグメントの映像時間が 20 秒の場合における平均 LPIPS は 0.2694 となる一方で，セグメントの映像時間が 50 秒の場合で 0.2667 となる．提案手法では，セグメントの映像時間が長くなるほど，多くの数のバッチから超解像

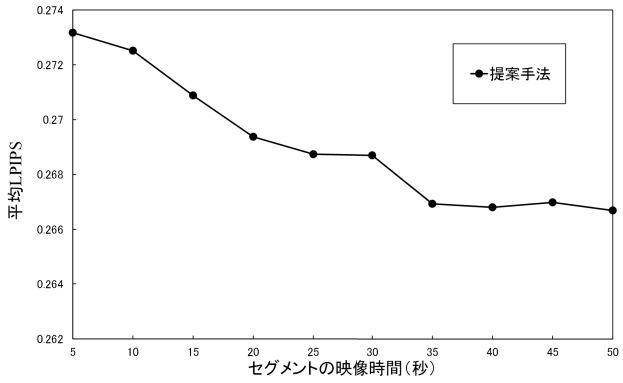


図 12 セグメントの映像時間に対する平均 LPIPS
Fig. 12 Video time of each segment and average LPIPS.

を優先するバッチを選択できる．このため，より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選択でき，映像品質は向上することが分かる．

6. おわりに

本論文では，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ことで視覚的な映像品質を向上させる手法を提案した．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に保存しながら再生するとき，再生開始までの間で，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，およびすべ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，提案手法が他の手法と比べて視覚的な品質向上の効果が高いことを示した．以上より，計算資源が十分でないクライアント計算機で超解像処理を行う場合，提案手法を適用することで，単純手法に比べて視覚的な映像品質の向上効果を高めるとともに，特徴検出にかかる負荷を減らすことで超解像処理を行うフレーム数を減少でき，視聴品質を向上できる．

今後の予定として，映像を構成する各シーンに対して複数のフレームに分割して超解像処理を行う手法の提案があげられる．今回提案した固定数のフレーム分割とは異なり，各シーンを構成するコーナ数や映像時間に応じてシーンの切替え時における解像度の変化が小さくなり，視聴品質への影響を抑えられると考えられる．また，画像単位で領域に応じて既存の超解像処理手法を用いた映像超解像処理手法の提案があげられる．

謝辞 本研究の一部は，文部科学省科学研究費補助金・基盤研究（B）（課題番号：21H03429，22H03587），および（公財）日揮・実吉奨学会の研究助成による成果である．ここに記して謝意を表す．

参考文献

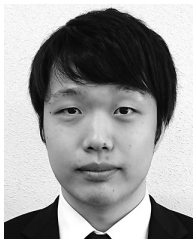
[1] Cisco Annual Internet Report (2018-2023) White Paper - Cisco (online), available from ⟨https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html⟩ (accessed 2022-02-04).
[2] RFC8216 - HTTP Live Streaming: IETF (online), available from ⟨https://www.ietf.org/rfc/rfc8216.txt⟩ (accessed 2022-02-04).
[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 2022-02-04).
[4] Gotoh, Y. and Oishi, T.: A Consideration of Delivering Method for Super-Resolution Video, *Proc. 10th International Workshop on Advances in Data Engineering and Mobile Computing* (*DEMoC-2021*), pp.268–274 (2021).
[5] Keys, R.: Cubic Convolution Interpolation for Digital Image Processing, *IEEE Trans. Acoustic, Speech and Signal Processing*, Vol.29, pp.1153–1160 (1981).
[6] Dong, C., Loy, C.C., He, K. and Tang, X.: Learning a Deep Convolutional Network for Image Super-Resolution, *Proc. European Conference on Computer Vision* (*ECCV*), pp.184–199 (2014).
[7] Kim, J., Lee, J.K. and Lee, K.M.: Accurate Image Super-Resolution Using Very Deep Convolutional Networks, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1646–1654 (2016).
[8] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. and Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.4681–4690 (2017).
[9] Sajjadi, M.S., Vemulapalli, R. and Brown, M.: Frame-Recurrent Video Super-Resolution, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.6626–6634 (2018).
[10] Chu, M., Xie, Y., Mayer, J., Leal-Taixé, L. and Thuerey, N.: Learning Temporal Coherence via Self-Supervision for GAN-based Video Generation, *ACM Trans. Graphics*, Vol.39, No.4, Article No.75, pp.75:1–75:13 (online), DOI: 10.1145/3386569.3392457 (2020).
[11] Zhang, Z. and Sze, V.: Fast: A Framework to Accelerate Superresolution Processing on Compressed Videos, *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.1015–1024 (2017).
[12] RFC2435 - RTP Payload Format for JPEG-compressed Video: IETF (online), available from ⟨https://www.ietf.org/rfc/rfc2435.txt⟩ (accessed 2022-02-04).
[13] Yeo, H., Jung, Y., Kim, J., Shin, J. and Han, D.: Neural Adaptive Content-aware Internet Video Delivery, *USENIX Symposium on Operating Systems Design and Implementation*, pp.645–661 (2018).
[14] Rosten, E. and Drummond, T.: Machine Learning for High-Speed Corner Detection, *European Conference on Computer Vision*, pp.430–443 (2006).
[15] Alcantarilla, P.F., Nuevo, J. and Bartoli, A.: Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, *Proc. British Machine Vision Conference*

(*BMVC*), pp.13.1–13.11 (2013).
[16] Vinay, A., Cholin, A.S., Bhat, A.D., Murthy, K.N.B. and Natarajan, S.: An Efficient ORB Based Face Recognition Framework for Human-Robot Interaction, *Procedia Computer Science*, Vol.133, pp.913–923 (2018).
[17] Li, Y., Brasch, N., Wang, Y., Navab, N. and Tombari, F.: Structure-SLAM: Low-Drift Monocular SLAM in Indoor Environments, *IEEE Robotics and Automation Letters*, Vol.5, No.4, pp.6583–6590 (2020).
[18] Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P.: Image Quality Assessment: From Error Measurement to Structural Similarity, *IEEE Trans. Image Processing*, Vol.13, pp.600–612 (2004).
[19] Zhang, R., Isola, P., Efros, A.A., Shechtman, E. and Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric, *Computer Vision and Pattern Recognition*, pp.586–595 (2018).
[20] Hoßfeld, T., Seufert, M., Sieber, C. and Zinner, T.: Assessing Effect Sizes of Influence Factors Towards a QoE Model for HTTP Adaptive Streaming, *Proc. 6th International Workshop on Quality of Multimedia Experience* (*QoMEX 2014*), pp.111–116 (2014).
[21] Dong, C., Loy, C.C. and Tang, X.: Accelerating the Super-Resolution Convolutional Neural Network, *European Conference on Computer Vision*, pp.391–407 (2016).
[22] The Apache HTTP Server Project (online), available from ⟨https://httpd.apache.org/⟩ (accessed 2022-02-04).
[23] Tears of Steel - Mango Open Movie Project (online), available from ⟨https://archive.org/details/Tears-of-Steel⟩ (accessed 2022-02-04).
[24] Big Buck Bunny (online), available from ⟨https://download.blender.org/peach/bigbuckbunny_movies/⟩ (accessed 2022-02-04).
[25] Herzmark Homestead on Vimeo (online), available from ⟨https://vimeo.com/226057477/⟩ (accessed 2022-02-04).

推薦文

CPU やメモリなどのクライアント側の計算資源に応じて，動画配信サービスにおける低品質の動画に対し，動画内の各フレームにおける特徴量に基づき，解像度を向上させる解像処理技術を適応するフレームを選定する方法を提案している．限られたネットワーク資源を効率よく利用することに際して，超解像処理技術は効果的な技術である一方，クライアント側の計算資源を消費するものであり，これらの資源をバランスよく利用し，配信される動画に対する満足度を向上させる実用的な取組である．今後，様々な種類に対する動画に対する評価，複数の超解像処理技術を組み合わせるなどにより，その実用性が一層高めることも期待され，本論文を推薦いたします．

（マルチメディア通信と分散処理研究会主査　田上 敦士）



大石 貴之

2019 年岡山大学工学部情報系学科卒業．2021 年同大大学院自然科学研究科博士前期課程修了．在学中は，動画配信システムの構築に従事．

後藤 佑介　（正会員）

2005 年岡山大学工学部情報工学科卒業．2007 年京都大学大学院情報学研究科システム科学専攻修士課程修了．2009 年同専攻博士後期課程修了．博士（情報学）．2009 年岡山大学大学院自然科学研究科助教を経て，2014 年同准教授．この間，豪ラトローブ大学客員研究員．放送コンピューティングおよび位置情報システムに興味を持つ．電子情報通信学会，IEEE 各会員．