Vol.2022-DPS-191 No.16
Vol.2022-MBL-103 No.16
Vol.2022-ITS-89 No.16
2022/5/26

動画配信におけるパッチの特徴量に基づく映像の超解像処理手法の提案

大橋 虎ノ介[1,a)]　大石 貴之[1]　後藤 佑介[2]

概要：
動画配信サービスに対するユーザの満足度は，サーバの配信速度やクライアントの受信速度に大きく依存する．サーバとクライアントとの間で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方法が挙げられるが，クライアントの受信動画が低解像度化すると，視聴品質は低下する．そこで，本研究では，低品質の映像受信時にフレームを複数のパッチに分割して特徴量が多いパッチ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，バッファに保存した一定時間分の映像に対して，再生開始までの間で特徴量が多いパッチを優先して超解像処理を行うことで，視覚的な映像品質が向上する．フレームの知覚的類似性を用いた映像評価では，提案手法による LPIPS の評価値は，特徴量に応じてパッチを選択しない手法と比べて約 0.03，および超解像処理を行わない手法と比べて約 0.05 の差でそれぞれ高いことを示した．

キーワード：超解像，畳み込みニューラルネットワーク，特徴量，パッチ

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で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，Adaptive Streaming と呼ばれる配信方式 [2], [3] 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 Streaming では，クライアントはサーバとの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する．一方で，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品質は低下する．

そこで，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再生することで視聴品質の低下を抑制する手法 [4] が提案されている．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再生する場合，バッファに保存した映像に対して，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から順番に超解像処理を行うことで，視覚的な映像品質は向上する．しかし，特徴量が多いフレーム内で特徴量が少ない領域に対しても同様に，フレーム単位で計算機の処理負荷が高い超解像処理を行うため，視覚的な映像品質は十分に向上できない．

本研究では，低品質の映像受信時にフレームを複数のパッチに分割して特徴量が多いパッチを優先して超解像処理を行いながら映像を再生する手法を提案する．提案手法では，バッファに保存した一定時間分の映像に対して特徴量が多いパッチを優先して超解像処理を行うことで，視覚的な映像品質が向上する．

2. 解像度変換技術

2.1 画素補間

画像を拡大する場合，原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．拡大画像は原画像に比べて多くの画素値をもつため，補間画素周辺の画素情報に基づいて，原画像では存在しない画素値を求める．画素値の補間では，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 [5] といった手法が主に利用される．

ニアレストネイバ法，バイリニア法，およびバイキュー

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

Vol.2022-DPS-191 No.16
Vol.2022-MBL-103 No.16
Vol.2022-ITS-89 No.16
2022/5/26

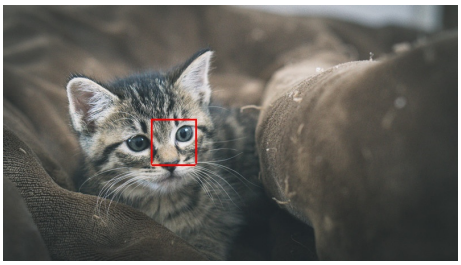

図 1 猫の原画像

図 2 各手法による猫の拡大画像

ビック法を用いて，**図 1** に示す猫を写した原画像の矩形領域をそれぞれ 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することで，原画像の画素値は変化せず，補間処理は容易となる．しかし，周辺画素の画素値を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 2×2 画素（4 画素）を基に，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となるため，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，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 4×4 画素（16 画素）を基に，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り，高周波成分を生成できないため，エッジをシャープに保つことはできない．

2.2 超解像

超解像は，動画や静止画の解像度を高めるための画像処理技術である．2.1 節で述べた一般的な画素補間による拡大とは異なり，画像の特性に基づいて画像を高解像度化する．また，超解像は，複数枚の類似画像を用いて一枚の高解像度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 種類の手法に分類される．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精度が高い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 3 SRCNN による猫の拡大画像

CNN を用いた超解像モデルである Super-Resolution Convolutional Neural Network（以下，SRCNN）[6] を用いて，図 1 において赤枠で示した画像の矩形領域を 4 倍に拡大した画像を**図 3** に示す．図 2 および図 3 より，SRCNN による画像のエッジは，他の 3 種類の手法と比較してシャープである．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に比べて高精度な超解像が可能である．また，高精度の CNN 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．

映像は連続した画像（以下，フレーム）であるため，各フレームに対して単一画像の超解像手法を適用することで，映像の超解像が可能である．しかし，映像の超解像における品質は，各フレームに対する超解像の精度とともに，フレーム間の動きの整合性が重要である．そこで，フレーム間の動きに一貫性がある映像の超解像を行う手法 [9], [10] が提案されている．

2.3 映像の超解像処理

動画配信サービスの利用時に低解像度の映像を受信した場合，超解像を適用することで受信動画を高解像度化して再生できる．低解像度の映像を受信する状況として，サーバが低解像度の映像のみを配信する場合，および Adaptive Streaming において通信状況が悪い場合が挙げられる．しかし，配信動画に対して超解像を行いながら再生する場合，受信した各フレームに対して超解像を行う時間に上限があるため，CPU やメモリといった計算資源が不足している場合は再生中断が発生する．そこで，動画を再生しながら映像の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）内のキーフレームのみに対して超解像を行うことで，他のフレームに超解像が伝播し，最終的にはすべてのフレームで超解像を行うことができる．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基に生成するフレームにおける超解像の精度は，フレームに対して超解像を直接適用した場合

Vol.2022-DPS-191 No.16
Vol.2022-MBL-103 No.16
Vol.2022-ITS-89 No.16
2022/5/26

図 4 原画像 (部屋) および変換後（ノイズ，ぼかし）の画像

に比べて低くなる．

Yeo らは，計算リソースに応じて深度を変更可能な深層 CNN モデルを用いた手法 [13] を提案している．この手法を適用することで，超解像を行いながら動画を再生できる．一方で，軽量な CNN モデルを用いた場合，動画を再生しながら超解像を行うことはできない．

3. 超解像の評価指標

3.1 画像類似性の指標

超解像によって生成された画像の品質を定量的に評価することは難しい．このため，元の高解像度画像を低解像度化した後に超解像で復元した画像と元の高解像度画像の 2 種類に対する類似度を超解像の精度とすることが一般的である [4]．本研究において，超解像映像の評価で用いる 3 種類の代表的な画像類似性の指標を説明する．

Peak Signal-to-Noise Ratio (PSNR) は，画素値の最大値と各画素値の平均誤差との比率を示す．PSNR が大きいほど画像の類似度は高いと判断できる．しかし，PSNR は 2 種類の画像に対応する画素値を単純に比較する．このため，画素に対してずれが発生すると評価値は大きく低下し，人の知覚特性と異なる評価結果となる場合がある．

Structual Similarity Index Measure (SSIM)[14] では，評価に用いる画素および周辺領域の画素がもつ画素値を基に，類似度を計算する．SSIM は，画素ごとの画素値を単純に比較せず周辺領域の画素値を利用する．このため，画素のずれに応じた評価値の低下は PSNR に比べて小さい．しかし，SSIM はスケーリングや回転といった幾何学的歪みの影響を受けやすく [15]，人の知覚特性とは異なる場合がある．

Learned Perceptual Image Patch Similarity (LPIPS)[16] は，人の知覚的類似性を学習したニューラルネットワークによる評価値である．様々な歪みを与えた画像の知覚的類似性に対して多くのユーザが類似性を評価したデータをニューラルネットワークで学習することで，人の知覚特性に基づいて類似性を評価できる．また，LPIPS は PSNR や SSIM に比べて歪み耐性が高く，LPIPS の値が小さいほど画像の知覚的類似度は高くなる．

3.2 画像類似度の比較評価

図 4 に，部屋を写した原画像，部屋にノイズを入れた画像，および部屋にぼかしを入れた画像の 3 種類をそれぞれ示す．また，図 4 のノイズ画像およびぼかし画像について，原画像との類似度の評価値を**表 1** に示す．

表 1 画像類似度の比較評価

	ノイズ画像	ぼかし画像
PSNR	17.812	**20.664**
SSIM	0.395	**0.645**
LPIPS	**0.605**	0.775

表 1 より，ぼかし画像における PSNR および SSIM の画像類似度は，ノイズ画像に比べて高い．PSNR および SSIM は，各ピクセルの誤差，および評価で用いる画素と周辺領域画素における画素値を基に，単純な計算式で評価する．一方，ぼかし画像における LPIPS の画像類似度は，ノイズ画像に比べて低い．LPIPS は，人の知覚的類似性を学習したニューラルネットワークによる評価値であり，人の知覚特性に基づいて類似性を評価する．

4. コーナー検出

4.1 コーナー検出手法

2 章で述べたように，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では高周波成分を生成できず，画像にぼやけが発生する．一方，CNN を用いた超解像ではエッジの部分がシャープになるため，高周波成分が多いコーナーの部分を含むフレームを優先して超解像を行うことで，視覚的な映像品質を向上できる．

画像処理において，画像から検出したエッジやコーナーの特徴を表す数値である特徴量を用いて高周波成分を検出するコーナー検出手法が提案されている．Features from Accelerated Segment Test (FAST) [17] および Accelerated KAZE（A-KAZE）[18] は，顔認識や Simultaneous Localization and Mapping (SLAM) といったリアルタイム処理で利用される [19], [20]．

図 5 に，部屋を映した原画像，および部屋の原画像に対して FAST を用いて検出したコーナーを描画した画像をそれぞれ示す．また，**図 6** に，空を写した原画像，および空の原画像に対して FAST を用いて検出したコーナーを描画した画像をそれぞれ示す．部屋の原画像は，ソファー，置物，および縦縞模様の壁紙で構成された複雑な画像であり，検出したコーナー数は 8,464 個である．一方，空の原画像は，空と雲のみで構成された単純な画像であり，検出したコーナー数は 4,822 個である．

4.2 特徴量を用いた原画像と復元画像の類似度評価

図 5 および図 6 の原画像を 0.25 倍に縮小し，バイキュービック法および SRCNN を用いて 4 倍に拡大した画像を**図 7**，**図 8** にそれぞれ示す．また，図 5 の原画像と図 7 の拡大画像に関する画像類似度，および図 6 の原画像と図 8

Vol.2022-DPS-191 No.16
Vol.2022-MBL-103 No.16
Vol.2022-ITS-89 No.16
2022/5/26

図 5 原画像（部屋）およびコーナー描画画像（部屋）

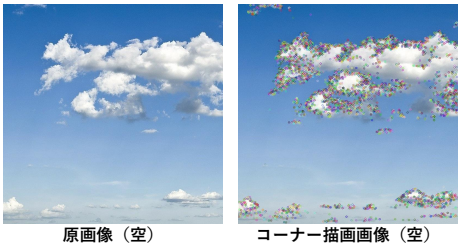

図 6 原画像（空）およびコーナー描画画像（空）

図 7 縮小した原画像（部屋）に対する各手法を用いた拡大画像

図 8 縮小した原画像（空）に対する各手法を用いた拡大画像

の拡大画像に関する画像類似度を**表 2** にそれぞれ示す．

表 2 より，部屋の画像について，SRCNN を用いた拡大画像と原画像との類似度は，バイキュービック法を用いて拡大した画像と原画像との類似度に比べて，すべての評価項目で高い．一方，空の画像について，SRCNN を用いた拡大画像と原画像との類似度は，バイキュービック法を用いて拡大した画像と原画像との類似度に比べて，PSNR と SSIM の評価値が低い一方で，LPIPS の評価値は高い．

LPIPS を用いた類似度評価について，SRCNN を用いた部屋の画像における評価値は，バイキュービック法に比べて 0.104 の差で高い．一方，SRCNN を用いた空の画像における評価値は，バイキュービック法に比べて 0.049 の差で低い．

2.1 節および 2.2 節で述べたように，バイキュービック法では高周波成分を生成できない一方で，SRCNN といった超解像では高周波成分を推定して生成できる．従って，部屋の画像といったコーナー数が多い画像では，コーナー部分で SRCNN を用いた超解像による高周波成分の推定効果が大きい．一方，空の画像といったコーナー数が少ない画像では，画素補間によるぼやけが少ないため，超解像による視覚的な品質向上の効果は小さい．

5. 提案手法

5.1 概要

本研究では，低品質の映像受信時にフレームを複数のパッチに分割し，特徴量が多いパッチ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，クライアントは再生中に中断が発生しないようにするため，一定のデータを計算機内のバッファに保存しながら動画を再生する．そこで，提案手法では，リアルタイムに超解像を行うことができない環境を想定し，バッファに保存した映像に対して，再生を開始するまでの間で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パッチを優先して超解像処理を行うことで，映像を構成するフレームに対する視覚的な品質を向上する．

5.2 提案手法の処理手順

はじめに，バッファに保存した映像を構成するすべてのフレームをバイキュービック法で拡大する．次に，バッファに保存した映像を構成するすべてのパッチに対して，各パッチに含まれるコーナー数を検出した後，特徴量に基づいて優先して超解像を行うパッチの処理順番を決定する．最後に，超解像で拡大したパッチの領域とバイキュービック法で拡大したフレームの領域を置換する．各処理の詳細について，以下の項で説明する．

5.2.1 バッファに保存した映像の拡大処理

画素補間による拡大画像を作成して解像度を高めるため，バッファに保存した映像を構成するすべてのフレームに対して，バイキュービック法で拡大処理を行う．このとき，拡大処理の終了時点で再生を開始するフレームは超解像処理を行わず，再生開始前のフレームのみに対して超解像処理を行い，解像度をさらに向上させる．提案手法では，バッファ中の映像に対して再生開始までの間で超解像を行い，再生開始に間に合わず超解像できない領域はバイキュービック法で拡大する．このため，超解像処理の前に，バッファに保存した映像を構成するすべてのフレームをバイキュービック法で拡大する．

5.2.2 パッチに含まれるコーナー数の検出

バッファに保存したフレームを正方形のパッチに分割す

Vol.2022-DPS-191 No.16
Vol.2022-MBL-103 No.16
Vol.2022-ITS-89 No.16
2022/5/26

表 2 画像補間における各手法の画像類似度の比較評価

	部屋 (バイキュービック法)	部屋 (SRCNN)	空 (バイキュービック法)	空 (SRCNN)
PSNR	23.611	**23.628**	**33.376**	33.159
SSIM	0.721	**0.728**	**0.892**	0.888
LPIPS	0.500	**0.396**	0.374	**0.325**

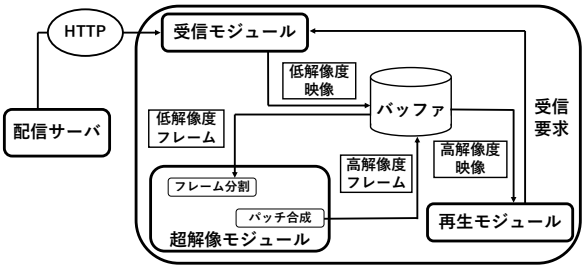

図 9 提案手法に適用する映像再生プレイヤの構成

表 3 計算機の性能

Server	CPU	Intel®Pentium(R) CPU G4400 (3.30 GHz)
	Memory	7.7 GBytes
	OS	Ubuntu 18.04.1LTS
Client	CPU	Intel®CORE(TM) i5-11700 CPU (2.50 GHz)
	Memory	16.0 GBytes
	OS	Windows 10 Home

る．本研究では，アスペクト比に基づいて，フレーム 1 枚につき $9 \times 16 = 144$ 枚のパッチを生成する．フレームのサイズは 144 p，180 p，および 270 p の 3 種類とし，パッチのサイズはそれぞれ 16×16 pixel，20×20 pixel，および 30×30 pixel とする．

2 章で述べたように，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では，高周波成分を生成できない．一方，CNN を用いた超解像ではエッジの部分がシャープになる．そこで，高周波成分であるコーナーの数が多いパッチに対して優先的に超解像処理を行う．コーナー検出のアルゴリズムは，コーナーの高速検出が可能な FAST[17] を用いる．

5.2.3 パッチの超解像処理順番の決定

4.2 節で述べたように，コーナー部分の特徴量が多い画像は，超解像で拡大した場合における精度向上の効果が高い．そこで，コーナーの数が多いパッチから順番に超解像を行うため，各パッチのコーナー数に対して順位付けを行う．

5.2.4 超解像処理後のパッチの置換

超解像は，バイキュービック法で拡大する前の低画質のパッチに対して行う．各パッチは，超解像が終了した契機で，バイキュービック法で拡大したフレームに対応する領域と置換する．すべてのパッチの超解像が終了する前に映像の再生を開始した場合，クライアントはバイキュービック法で拡大した領域と超解像を行った領域が混在した映像を再生する．

表 4 評価に用いる映像の構成

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

5.3 実装

提案手法を適用するため，**図 9** に示す映像再生プレイヤを計算機に実装した．超解像を行いながら受信映像を再生するプレイヤは，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し，2 個分のセグメントを保存すると受信を停止する．次に，再生モジュールは 1 個分のセグメントの再生終了後に受信要求を送り，受信モジュールは新たなセグメントを配信サーバから受信する．このとき，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，バッファには 2 個分のセグメントが常に保存される．なお，受信モジュールと配信サーバにおける通信プロトコルは，Adaptive Bitrate による配信で主に利用される HTTP を用いる．

超解像モジュールでは，バッファに保存されており次に再生予定のセグメントに対して，提案手法による超解像を行う．超解像を行っているセグメントの再生要求を受信すると，このセグメントの超解像を終了し，バッファに保存されている次のセグメントに対して超解像を開始する．なお，超解像では，CNN を用いた学習型超解像モデルである FSRCNN[21] を用いて，映像を構成する各パッチを 4 倍に拡大する．また，超解像を行うことができないパッチは，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，バッファに保存されている映像を再生する．1 個分のセグメントの再生が終了すると，次のセグメントの再生を開始し，新たなセグメントの受信を受信モジュールに要求する．

6. 評価

6.1 評価環境

5 章で示した映像再生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

Vol.2022-DPS-191 No.16
Vol.2022-MBL-103 No.16
Vol.2022-ITS-89 No.16
2022/5/26

表 5 各視聴映像の品質評価

視聴映像		提案手法			単純手法			バイキュービック手法		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Tears of Steel	144 p	28.966	0.834	**0.229**	28.744	0.828	0.250	**29.739**	**0.839**	0.289
	180 p	30.590	0.857	**0.202**	30.425	0.851	0.232	**31.176**	**0.859**	0.253
	270 p	32.585	0.887	**0.184**	32.439	0.883	0.203	**32.773**	**0.888**	0.209
Big Buck Bunny	144 p	26.854	0.789	**0.279**	26.369	0.781	0.307	**28.086**	**0.794**	0.342
	180 p	28.463	**0.817**	**0.251**	27.415	0.807	0.286	**29.205**	0.815	0.306
	270 p	30.310	**0.854**	**0.220**	29.682	0.848	0.243	**30.611**	0.851	0.250
Helzmark Homestead	144 p	20.333	**0.438**	**0.586**	20.290	0.436	0.593	**20.387**	0.409	0.661
	180 p	20.282	**0.435**	**0.599**	20.230	0.434	0.603	**20.285**	0.414	0.648
	270 p	**20.074**	**0.438**	**0.594**	20.024	0.435	0.596	20.050	0.426	0.617

表 6 視聴映像の超解像処理パッチ数

視聴映像		超解像処理パッチ数	
		提案手法	単純手法
Tears of Steel	144 p	1,481,451	**1,516,181**
	180 p	**1,141,930**	1,094,601
	270 p	465,212	**480,770**
Big Buck Bunny	144 p	**1,551,447**	1,521,584
	180 p	1,045,861	**1,090,126**
	270 p	461,608	**483,571**
Helzmark Homestead	144 p	1,407,466	**1,469,402**
	180 p	928,862	**1,000,838**
	270 p	440,725	**440,760**

する．動画を配信するサーバを構築するソフトウェアは，Apache HTTP Server[22] を用いた．評価に用いた計算機の性能を**表 3** に示す．サーバおよびクライアントは，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．

6.2 評価に用いる映像

評価に用いる 3 種類の映像を**表 4** に示す．すべての映像は，開始から 10 分間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．また，他の 2 種類の映像とアスペクト比を同一にするため，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[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，空およびキャラクタの模様は単純である．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は小さい．

6.3 映像の種類に応じた映像品質

提案手法，映像を構成する先頭のパッチから順番に超解像処理を行う単純手法，およびすべてのフレームをバイキュービック法で拡大するバイキュービック手法の 3 種類について，再生したすべてのフレームに対する平均品質を評価する．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 秒である．評価項目は，平均 PSNR，平均 SSIM，および平均 LPIPS の 3 種類である．

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 種類の映像による評価結果を**表 5** に示す．表 5 より，9 種類の映像に対して提案手法の LPIPS は最も高い．また，Helzmark Homestead における 270 p の場合を除き，バイキュービック手法における PSNR は最も高い．

次に，Tears of Steel の場合，バイキュービック手法における SSIM が最も高い．また，Big Buck Bunny および Helzmark Homestead の場合，Big Buck Bunny における 144 p の場合を除き，提案手法の SSIM は最も高い．

6.4 映像の解像度に応じた超解像処理パッチ数

提案手法と単純手法において，超解像処理が行われたパッチ数（以下，超解像処理パッチ数）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は 20 秒である．

3 種類の映像による評価結果を**表 6** に示す．表 6 より，提案手法および単純手法において，受信映像の解像度が高いほど超解像処理パッチ数は少ない．また，Tears of Steel における 180 p の映像，および Big Buck Bunny における 144 p の映像を除き，提案手法の超解像処理パッチ数は，単純手法に比べて少ない．

6.5 映像のフレームレートに応じた映像品質

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．評価では，解像度が 144 fps，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を 20 秒とする．評価項目は，平均 PSNR，平均 SSIM，および平均 LPIPS の 3 種類である．

3 種類のフレームレートによる映像を用いた評価結果を**表 7** に示す．表 7 より，提案手法および単純手法は，フレームレートが低いほど LPIPS の評価値が高くなる一方で，PSNR および SSIM の評価値はほとんど変化しない．また，バイキュービック手法において，PSNR，LPIPS，および SSIM の評価値は，フレームレートの違いでほとんど変化しない．

表 7　各フレームレートの視聴映像における品質評価

視聴映像		提案手法			単純手法			バイキュービック手法		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Big Buck Bunny	24fps	26.854	0.789	**0.279**	26.369	0.781	0.307	**28.086**	**0.794**	0.342
	30fps	27.642	0.792	**0.293**	26.563	0.781	0.319	**28.420**	**0.793**	0.345
	60fps	27.448	**0.788**	**0.318**	27.294	0.781	0.343	**27.747**	0.786	0.353

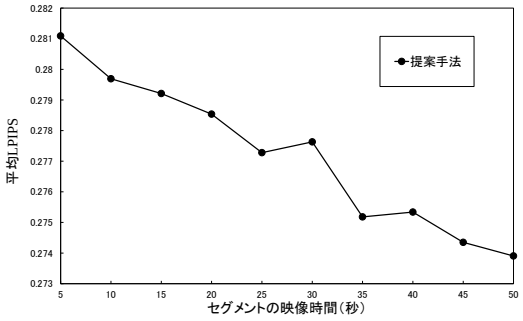

図 10　セグメントの映像時間に対する平均 LPIPS

6.6　セグメントの映像時間に応じた映像品質

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，評価指標として LPIPS を用いる．

セグメントの映像時間に応じた平均 LPIPS の評価結果を**図 10** に示す．縦軸は平均 LPIPS，横軸はセグメントの映像時間である．図 10 より，セグメントの映像時間が長いほど，超解像処理を優先して行うパッチの候補が多くなるため，平均 LPIPS の評価値は高くなり，映像品質は高くなる．

7.　考察

7.1　映像の種類に応じた映像品質

6.3 節の評価結果より，9 種類の映像に対して，提案手法における LPIPS の評価値は最も高い．提案手法は，特徴量が多く視覚的な品質向上効果が高いと予測されるパッチに対して優先的に超解像処理を行う．このため，提案手法は，単純手法およびバイキュービック手法に比べて各フレームの視覚的な品質が向上し，LPIPS の評価値は高くなる．また，Helzmark Homestead における 270 p の映像を除き，バイキュービック手法における PSNR の評価値は最も高い．バイキュービック法による超解像は，FSRCNN に比べて原画像の画素値を失う画素数が少ないため，隣接する画素間で画素の差を評価する PSNR の評価値は高くなる．

次に，提案手法とバイキュービック手法における LPIPS の評価値を比較する．3 種類の映像において，映像の解像度が高くなるほど，提案手法とバイキュービック手法における LPIPS の評価値の差は小さくなる．映像の解像度が高くなると，パッチのサイズが大きくなるため超解像の処理時間が長大化し，超解像処理を行うパッチ数が減少する．

7.2　映像の解像度に応じた超解像パッチ数

6.4 節の評価結果より，提案手法における超解像パッチ数は単純手法に比べて少なく，受信映像の解像度が高いほど少ない．提案手法は，各パッチで含まれるコーナー数を計算してすべてのパッチをソートするため，超解像を行う時間は単純手法に比べて短い．また，超解像処理にかかる時間は，入力画像の大きさに比例する．このため，受信映像の解像度が高い場合，各パッチに対する超解像の処理時間が長くなり，超解像パッチ数は減少する．

7.3　映像のフレームレートに応じた映像品質

6.5 節の評価結果より，提案手法および単純手法において，LPIPS の評価値は，フレームレートが低いほど高い．フレームレートが高い場合，セグメントに含まれるフレーム数が増え，超解像処理を行うパッチ数の割合は低くなる．フレームレートが 24 fps の場合と 60 fps の場合について，LPIPS における評価値の差の割合は，提案手法で 12.3%，単純手法で 10.5%となる．このため，超解像処理を行うパッチ数の割合の増加による視覚的な品質の向上率は，提案手法の方が大きい．

7.4　セグメントの映像時間に応じた映像品質

6.6 節の評価結果より，セグメントの映像時間が長くなると平均 LPIPS の評価値は大きくなる．例えば，セグメントの映像時間が 50 秒の場合，平均 LPIPS の評価値は 0.274 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，優先して超解像を行う候補となるパッチの数が増える．このとき，視覚的な品質向上効果がより高いと予測されるパッチを選択でき，映像品質が向上する．

8.　おわりに

本研究では，低品質の映像受信時に特徴量が多いパッチ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，クライアントがバッファに保存した映像に対して，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パッチを優先して超解像処理を行う．評価では，提案手法，映像を構成する先頭のパッチから順番に超解像処理を行う単純手法，およ

びすべてのフレームをバイキュービック法で拡大するバイキュービック手法の 3 種類を用いて，配信する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，フレーム内で特徴量が多く複雑な領域，および特徴量が少なく単純な領域がどちらも存在する場合，提案手法は他の 2 種類の手法と比べて視覚的な映像品質が高いことを示した．

今後の予定として，ユーザによる提案手法の定性的評価，および提案手法とパッチ分割を行わない手法との比較評価を行う．

謝辞 本研究は，文部科学省科学研究費補助金（基盤研究（B）課題番号：21H03429, 22H03587），日本学術振興会二国間交流事業（課題番号：JPJSBP120229932），および（公財）日揮・実吉奨学会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: Cicso Annual Internet Report (2018-2023) White Paper - Cisco (online), Cisco Systems, Inc. (online), available from ⟨https://www.cisco.com/c/ja_jp/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html⟩ (accessed 2022-02-21).

[2] Pantos, R.: HTTP Live Streaming: IETF (online), Apple, Inc. (online), available from ⟨https://tools.ietf.org/html/rfc8216⟩ (accessed 2022-02-21).

[3] for Standardization, I. O.: Information Technology - Dynamic Adaptive Streaming over HTTP (DASH) - Part 1: Media Presentation Description and Segment Formats: ISO (online), International Organization for Standardization (online), available from ⟨https://www.iso.org/standard/75485.html⟩ (accessed 2022-02-21).

[4] 大石貴之，後藤佑介：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，情報処理学会研究報告，Vol. 187, No. 11, pp. 1–8 (2021).

[5] Keys, R.: Cubic Convolution Interpolation for Digital Image Processing, IEEE TransAcoustic, *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 forImage Super-Resolution, *European Conference on Computer Vision*, 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. 105–114 (2017).

[9] Sajjadi, M. S., Vemulapalli, R. and Brown, M.: Frame-Recurrent Video SuperResolution, *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 SelfSupervision for GAN-based Video Generation, *International Conference on Learning Representations*, pp. 75:1–75:13 (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] Berc, L. M., Fenner, W. C., Frederick, R., McCanne, S. and Stewart, P.: RFC2435 - RTP Payload Format for JPEG-compressed Video: IETF (online), Lawrence Berkeley Laboratory (online), available from ⟨https://www.ietf.org/rfc/rfc2435.txt⟩ (accessed 2022-02-21).

[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] Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P.: Image Quality Assessment: from Error Measurement to Structural Similarity, *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 600–612 (2004).

[15] Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C. and Markey, M. K.: Complex Wavelet Structural Similarity: A New Image Similarity Index, *IEEE Transactions on Image Processing*, Vol. 18, No. 11, pp. 2385–2401 (2009).

[16] 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).

[17] Rosten, E. and Drummond, T.: Machine Learning for Highspeed Corner Detection, *European Conference on Computer Vision*, pp. 430–443 (2006).

[18] Alcantarilla, P. F., Nuevo, J. and Bartoli, A.: Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, *British Machine Vision Conference*, pp. 13.1–13.11 (2013).

[19] 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).

[20] 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).

[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] Foundation, A. S.: The Apache HTTP Server Project (online), Apache Software Foundation (online), available from ⟨https://httpd.apache.org/⟩ (accessed 2022-02-21).

[23] Foundation, B.: Tears of Steel — Mango Open Movie Project (online), Blender Foundation（オンライン），入手先 ⟨https://mango.blender.org/⟩ （参照 2022-02-21）.

[24] Foundation, B.: Big Buck Bunny (online), Blender Foundation (online), available from ⟨https://peach.blender.org/⟩ (accessed 2022-02-21).

[25] LLC, H. A. V.: Herzmark Homestead on Vimeo (online), Hawaii Aerial Visions LLC (online), available from ⟨https://vimeo.com/226057477/⟩ (accessed 2022-02-21).