

NeRF および特徴マップに基づくフレーム補間手法の特性評価

Characterization of frame interpolation methods based on NeRF and feature maps

速見 泰雅[†] 金 洛旭[‡] 渡辺 裕^{†‡} 中條 健^{††} 青野 友子^{††}
Taiga Hayami[†] Louxu Jin[‡] Hiroshi Watanabe^{†‡} Takeshi Chujoh^{††} Tomoko Aono^{††}
八杉 将伸^{††} 洪 秀俊^{††} 范 哲銘^{††} 猪飼知宏^{††}
Yukinobu Yasugi^{††} Sujun Hong^{††} Zheming Fan^{††} Tomohiro Ikai^{††}
[†] 早稲田大学基幹理工学部 [‡] 早稲田大学大学院基幹理工学研究科
[†]School of Fundamental [‡]Graduate School of ^{††}SHARP 株式会社
Science and Engineering, Fundamental Science and ^{††}SHARP Corporation
Waseda University Engineering, Waseda University

Abstract: 特徴マップに基づくフレーム補間はフレーム差分が大きいと補間画像の劣化が大きくなる。また NeRF は静止物体の多視点画像を学習することで自由視点画像を生成することができる。NeRF を用いることで、動画の各フレームを入力とし、フレーム間の画像を生成することができる。本稿では、動物体を含む動画フレームに対して、特徴マップに基づく補間である FILM と、NeRF に基づくフレーム補間、NeRF に加えて歪除去を目的とした NeRFLiX に基づくフレーム補間を適用し、その特性評価を行う。

1 はじめに

近年インターネットの普及や技術の向上により高品質な動画が求められている。フレーム補間は、動画のフレームレートを向上させるための技術であり、動画の品質の改善に用いられる。隣接フレームから得られる特徴マップを用いてフローを予測し、補間フレームを作成する手法が一般である。しかし、カメラ視点が大きく移動するようなフレーム差分が大きい場合は、出力される補間画像が大きく劣化する。また、この手法では画像の視点情報を扱っていないため、補間画像の視点は任意に決めることができない。

そこで本稿では、Neural Radiance Fields(NeRF) [1] を用いたフレーム補間手法について検討する。NeRF は静止した画像群から自由視点画像を生成する技術で、任意の視点からの画像を得ることができる。

比較手法として、特徴マップに基づくフレーム補間手法の Frame Interpolation for Large Motion(FILM) [2] を用いる。また、NeRF によるフレーム補間に加え、NeRFLiX [3] を用いた手法により画像の品質向上を図る。

2 フレーム補間手法

2.1 NeRF

NeRF は静止した被写体の多視点画像群を用いて、3次元空間上各座標、各方向の色を Neural Network により学習する。学習された Network に任意の視点方向を入力することでその視点から得られるであろう画像がレンダリングにより出力される。一般に NeRF は 3次元空間上の静止された情報を扱うため、動いている被写体の多視点画像には対応していない。しかし、ある程度小さな動きであれば劣化された画像として出力される。そこで、動画の各フ

レームを多視点画像として扱い、NeRF の入力画像として用いる。任意の視点画像が生成できるためフレーム間の視点からの画像を生成することで、フレーム補間が可能となる。また、FILM と異なり補間したい視点を指定することができるため、補間画像が一意に指定されない。

2.2 NeRFLiX

NeRFLiX は NeRF のレンダリング画像の品質を改善するモデルである。NeRF の学習に用いる画像が少ないと、レンダリング画像の品質が低下し、NeRF 特有の歪やノイズが生じる。一般の Denoise 処理では用いられる学習画像のノイズが NeRF のノイズと異なり、簡単に除去できない。また、学習で用いる NeRF のレンダリングのデータセットを作成するには大きな手間がかかる。

NeRFLiX は NeRF 特有の歪、ノイズを劣化プロセスにより再現した。劣化プロセスには Splatted Gaussian noise の付加、画像のピクセルの Re-positioning、ぼかしパターンを再現する Anisotropic blur の 3種類が用いられている。任意の画像データセットに対して、この劣化プロセスを紹介することで擬似的な NeRF のレンダリング画像のデータセットを作成し、Denoise 処理を適用する。

2.3 FILM

FILM は複数のスケールの画像から特徴量を用いてフローを予測するフレーム補間モデルである。フレーム間の差分や画像サイズが大きいと各フレームの対応点が取れず、補間画像が全体的に劣化する。

3 実験

FILM、NeRF、NeRFLiX を用いたフレーム補間をそれぞれ評価する。評価に用いるデータセットは Neural 3D Video Synthesis from Multi-view Video [4] において公開

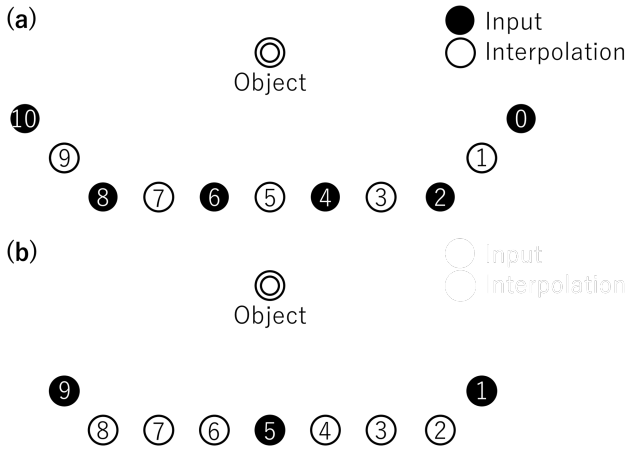


図 1: カメラ位置

された Neural 3D Video Synthesis Dataset を用いる。このデータセットは 21 台の異なる位置のカメラで被写体を撮影した動画でそれぞれの時刻は同期されている。カメラの位置情報が含まれており、NeRF で用いる。Neural 3D Video Synthesis Dataset のうち、特定のカメラ (a)、(b) を用いて、フレーム補間の評価を行う。用いたカメラはそれぞれ図 1 のように被写体を囲むように配置されている。それぞれのカメラ動画から時間の異なる 1 フレームを選択することで図 1 の左手から右手に移動するような動画フレームを作成する。予測に使用する画像 (a) と (b)、フレーム間の時間差 (1)1/30s と (2)15/30s を用意し、計 4 つの条件下で検証する。

NeRF ではカメラ画像を入力画像として学習させ、補間画像をレンダリングする。NeRF で得られたレンダリングに NeRFLiX ネットワークを介することでレンダリング画像の劣化を改善する。FILM では公開されている学習済みモデルを使用し、カメラ画像の内、隣接フレーム 2 枚づつをそれぞれ入力とする。評価指標は PSNR、SSIM、LPIPS を用いる。

4 結果

各手法の評価結果を表 1 に示す。FILM は、入力画像間が近い (a) においてフレーム間の時間差が大きくなっても劣化は小さい一方、入力画像画像が離れる (b) と大きく劣化する。また、FILM は補間フレームの正確な視点情報を保持していないため、正解画像から視点のずれた画像が出力される。これらにより NeRF に比べて PSNR を大きく落としている。NeRF に基づく補間手法は動物体の動き補間できていないが、PSNR、SSIM の評価は FILM を上回っている。

手法 (b)-(1) における補間フレームの一例を図 2 に示す。参照フレーム間が広がっている (b) では、FILM は背景を含め全体的に劣化する。一方、NeRF や NeRFLiX はノイ

表 1: 各手法の特性評価

手法		PSNR \uparrow		SSIM \uparrow		LPIPS \downarrow	
		(1)	(2)	(1)	(2)	(1)	(2)
FILM	a	20.05	20.00	0.591	0.586	0.190	0.201
NeRF	a	21.17	21.57	0.683	0.716	0.201	0.224
NeRFLiX	a	22.33	21.79	0.766	0.743	0.165	0.201
FILM	b	17.22	17.083	0.510	0.506	0.400	0.414
NeRF	b	19.18	18.91	0.553	0.532	0.306	0.327
NeRFLiX	b	19.35	19.06	0.582	0.563	0.290	0.316



図 2: 補間フレームの一例 (左上: Ground Truth、右上: FILM、左下: NeRF、右下: NeRFLiX)

ズは多くのものが背景に歪が発生しにくい。また、そのノイズは NeRF の学習に用いる画像枚数によるものの影響が大きいと考えられ、フレーム間の距離に影響されにくい。

5 まとめ

FILM、NeRF、NeRFLiX を動画のフレーム補間法として適応し、それぞれの手法について生成フレームの画像品質を評価した。フレーム間の距離やフレーム時刻を異なる条件でそれぞれ比較することで、各手法における劣化の仕方の違いを確認した。今後は、NeRF に基づくフレーム補間手法に現れる動物体部分の歪を改善していく必要がある。

参考文献

- [1] B. Mildenhall *et al.*: "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis", ECCV, 2020.
- [2] F. Reda *et al.*: "FILM: Frame Interpolation for Large Motion", ECCV, 2022.
- [3] K. Zhou *et al.*: "NeRFLiX: High-Quality Neural View Synthesis by Learning a Degradation-Driven Inter-viewpoint MiXer", CVPR 2023.
- [4] T. Li *et al.*: "Neural 3D Video Synthesis from Multi-view Video", CVPR, 2022.