

フレーム補間における順方向・逆方向ワーピングの活用

清水盛偉[†] 孫 鶴鳴[‡] 甲藤 二郎[†]早稲田大学理工学術院[†] 国立研究開発法人科学技術振興機構さきがけ[‡]

1 はじめに

フレーム補間は動画の連続フレームからその間のフレームを生成することを可能とする。この技術を用いると、元の動画よりも高いフレームレートの動画を生成することができる。近年、深層学習を用いたフレーム補間手法が多く提案されている。そのうちの一つのアプローチとしてオプティカルフローを用いた手法がある。これらの手法は、はじめに両方向オプティカルフローを求め、逆方向ワーピング処理から間のフレームを生成する。しかし、補間画像を得るには順方向ワーピングを用いる方法もある。本稿では補間フレーム I_t から入力フレーム I_0, I_1 までのフロー $F_{t \rightarrow 0}, F_{t \rightarrow 1}$ を用いたワーブ処理を逆方向ワーピングと呼ぶ。逆に入力フレーム I_0, I_1 から補間フレーム I_t までのフロー $F_{0 \rightarrow t}, F_{1 \rightarrow t}$ を用いたワーブ処理を順方向ワーピングと呼ぶ。

本稿では順方向ワーピングと逆方向ワーピングの両方を用いた補間手法を提案する。また、実験結果として、提案手法は逆方向ワーピングのみを用いた場合に比べ補間精度が向上することを示す。

2 関連研究

連続フレームからその間のフレームを生成することがフレーム補間の目的である。近年多くの深層学習をベースとした手法が提案されている。それらはいくつかの種類に分類されるが、一つのアプローチとしてオプティカルフローをもとにした手法がある。これらは両方向オプティカルフローを求め、入力フレームとフローを用いたワーブ処理により補間フレームを得る[1]。こういった手法では主に逆方向ワーピングが用いられるが、[2]は順方向ワーピングを用いることを提案した。

3 提案手法

3.1 順方向ワーピングの活用

本研究では逆方向ワーピングを活用するフレーム補間モデル Super SloMo[1]に順方向ワーピング処理を加える。図1に提案手法の概要を示す。2つの

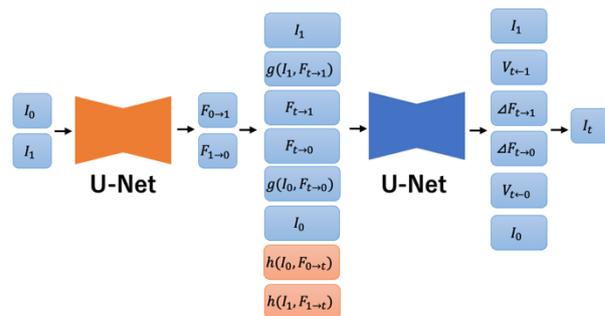


図1: 提案手法の概要

入力フレーム I_0, I_1 は初めに1つ目のU-Netに入力され、両方向オプティカルフロー $F_{0 \rightarrow 1}, F_{1 \rightarrow 0}$ を得る。次に、逆方向ワーピングを行うために必要な $F_{t \rightarrow 0}, F_{t \rightarrow 1}$ を以下の式で計算する:

$$F_{t \rightarrow 0} = -(1-t)tF_{0 \rightarrow 1} + t^2F_{1 \rightarrow 0} \quad (1)$$

$$F_{t \rightarrow 1} = (1-t)^2F_{0 \rightarrow 1} - t(1-t)F_{1 \rightarrow 0} \quad (2)$$

$I_0, I_1, F_{t \rightarrow 0}, F_{t \rightarrow 1}$ を用いた逆方向ワーピングで得られたワーブ画像を $g(I_0, F_{t \rightarrow 0}), g(I_1, F_{t \rightarrow 1})$ と表す。これらを2つ目のU-Netへの入力としていたのが元のSuper SloMoだが、ここで新たに順方向ワーピングによって得られたワーブ画像 $h(I_0, F_{0 \rightarrow t}), h(I_1, F_{1 \rightarrow t})$ を追加で入力することを提案する。順方向ワーピングには[2]で提案されたSoftmax Spalttingを用いる。また $F_{0 \rightarrow t}, F_{1 \rightarrow t}$ は以下の式で求める:

$$F_{0 \rightarrow t} = t * F_{0 \rightarrow 1} \quad (3)$$

$$F_{1 \rightarrow t} = (1-t) * F_{1 \rightarrow 0} \quad (4)$$

2つ目のU-Netでフローの差分 $\Delta F_{t \rightarrow 0}, \Delta F_{t \rightarrow 1}$ とVisibility Map $V_{t \rightarrow 0}, V_{t \rightarrow 1}$ を取得し、これらの情報から最終的な補間フレームを得る。

3.2 RAFTによるオプティカルフローの活用

3.1では順方向ワーピングに1つ目のU-Netから得られた $F_{0 \rightarrow 1}, F_{1 \rightarrow 0}$ を用いた場合の提案手法を説明した。しかし、1つ目のU-Netの出力は逆方向ワーピング処理のために最適化されたものであり、他のオプティカルフローモデルよりも精度が低い。よって順方向ワーピングにオプティカルフローモデルのRAFT[3]を利用した際も評価する。これにより精度の高い順方向ワーブ画像が得られるが、大きなオクルージョンが生じてしまう。この問題を解決す

Utilizing Forward and Backward Warping for Frame Interpolation

Joi Shimizu[†], Heming Sun[†], Jiro Katto[†]

[†] Faculty of Science and Engineering, Waseda University

[‡] PRESTO, JST, Kawaguchi-shi, Saitama, 332-0012 Japan

るために、順方向ワーブ画像に生じたオクルージョンを逆方向ワーブ画像のピクセル値で埋める処理を加える。 $h(I_0, F_{0 \rightarrow t})$ のオクルージョンは $g(I_1, F_{t \rightarrow 1})$ のピクセル値で埋め、 $h(I_1, F_{1 \rightarrow t})$ のオクルージョンは $g(I_0, F_{t \rightarrow 0})$ のピクセル値で埋める。図2に順方向ワーブ画像のオクルージョンとそれを埋めた際の例を示す。図2(a)では、特にボールや人間の手の周りにオクルージョンが生じている事がわかる。



(a) 順方向ワーブ画像(オクルージョンあり) (b) 順方向ワーブ画像(オクルージョンなし)

図2: RAFTによる順方向ワーブ画像

4 実験

元の Super SloMo と提案手法を比較する。モデルの学習に用いたデータセットは Adobe 240fps データセット、GOPRO データセット、YouTube で集めた 240fps 動画、及び iPhone で撮影した 240fps 動画である。以下に学習に用いたデータセットの詳細を示す。

表1: 学習用データセットの詳細

	Adobe240	GOPRO	YouTube	iPhone
動画数	133	33	190	83
フレーム数	124, 841	34, 874	93, 161	117, 320
解像度	720p	720p	Various resolutions	1080p

学習のエポック数は共に 250 に統一する。学習率の初期値を 0.0001 とし、100 エポックおきに学習率に 0.1 を掛ける。評価指標は PSNR とし、評価は Middlebury と DAVIS データセットで行う。

表2: Super SloMo と提案手法の精度評価

	Middlebury	DAVIS
Overlapping	27.97	-
Phase-Based [4]	31.12	-
MIND [5]	31.35	-
DVF [6]	34.34	-
Super SloMo	34.24	27.00
Ours w/o RAFT	34.43	27.04
Ours w/ RAFT	34.51	27.13

Overlapping、Phase-Based、MIND、DVF の結果には[7]で報告された数値を使う。また、Super SloMo の結果は[7]で報告されている精度とほとんど同じとなっている。

Super SloMo の1つ目のU-Netの出力情報による順方向ワーブ画像を用いた際、Middlebury、DAVIS の両方のデータセットで補間の精度が向上した。オリジナルの Super SloMo は DVF よりも精度が低いが、提案手法は DVF よりも高くなっている。また、RAFT を用いることでさらなる精度向上が確認できた。

図3には補間画像の例を示す。順方向ワーピング処理を加えることで歪みが削減できていることが確認できる。

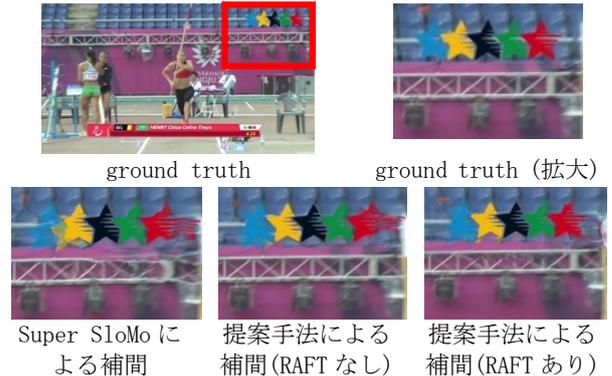


図3: 補間画像の比較

5 まとめ

本稿では逆方向ワーピングを用いたオプティカルフローベースのフレーム補間モデル Super SloMo に順方向ワーピング処理を加えることで、補間の精度が向上することを示した。また、順方向ワーピングに用いるオプティカルフローを RAFT で計算した際、さらなる精度向上が確認できた。今後は他の逆方向ワーピングを用いたフレーム補間モデルでも同様に順方向ワーピング処理を加え、精度が向上するかを確認したい。

6 謝辞

本研究は情報通信研究機構(採択番号 03801)及び JST さきがけ JPMJPR19M5 の支援を受けたものである。

参考文献

- [1] H. Jiang, D. Sun, V. Jampani, M. Yang, E. Miller, J. Kautz, "Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation", IEEE CVPR, pp. 9000-9008, 2018.
- [2] S. Niklaus, F. Liu, "Softmax Splatting for Video Frame Interpolation", IEEE CVPR, pp. 5436-5445, 2020.
- [3] Z. Teed, J. Deng, "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow", Springer ECCV, 2020.
- [4] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. S. Hornung, "Phase-Based Frame Interpolation for Video", IEEE CVPR, pp. 1410-1418, 2015.
- [5] Gucan Long, Laurent Kneip, Jose M Alvarez, Hongdong Li, Xiaohu Zhang, and Qifeng Yu. "Learning Image Matching by Simply Watching Video", Springer ECCV, 2016.
- [6] Z. Liu, R. A. Yeh, X. Tang, Y. Liu, A. Agarwala, "Video Frame Synthesis using Deep Voxel Flow", IEEE ICCV, pp. 4473-4481, 2017.
- [7] H. Lee, T. Kim T. Chung, D. Pak, Y. Ban, S. Lee, "AdaCoF: Adaptive Collaboration of Flows for Video Frame Interpolation", IEEE CVPR, pp. 5315-5324, 2020.