時間相関カメラと深層学習を用いたオプティカルフロー推定に関する研究

1235069　橋本　悠衣里　【 画像情報工学研究室 】

Study on optical flow estimation using correlation image sensor and deep learning

1235069　HASHIMOTO Yuiri　【 Image Processing and Informatics Lab. 】

1　はじめに

オプティカルフロー (OF) とは，3 次元空間での運動を画像上に射影した光学的な動きベクトルである．これまで，3 次元環境推定やジェスチャー認識などに応用され精力的に研究されている分野であり，近年では深層学習を用いた OF 推定が盛んに研究されている [1][2]．また，時間相関カメラという特殊なカメラを用いて OF 推定を行っている研究もある [3]．時間相関カメラは，各画素に入射する光強度信号と参照信号との 1 フレーム時間相関を出力するセンサであり，例えば運動により生じる輝度変化を複素数として記録することが可能である．

そこで我々はこれに着目し，1 枚の相関画像を深層学習ネットワークの入力とした OF 推定を提案した [4]．

本研究ではさらに推定精度を向上させるため，先行研究の深層学習ネットワークを改良し新たなデータセットを用いて実験を行い，学習済みモデルとの比較を行う．

2　原理

2.1　時間相関イメージセンサ

時間相関イメージセンサは，入射光と参照信号との時間相関値と強度出力を画像として出力する．入射光と参照信号との時間相関値を複素相関画像といい，参照信号として周波数 ω の複素正弦波を用いると式 (1) と表せる．

$$g(x,y) = \int_{-T/2}^{T/2} I(x,y,t)e^{-j\omega t}dt \qquad (1)$$

また，強度画像は式 (1) の $\omega = 0$ で表される．

2.2　深層学習ネットワーク

CNN を用いて OF 推定を行う手法として FlowNetS が提案されている [1]．FlowNetS は，運動前後の静止画像を入力として OF 推定を行うネットワークである．本研究では，このネットワークをベースに相関画像の実部・虚部をそれぞれ異なるチャンネルとして入力するネットワークを構築した．FlowNetS では 2 枚のカラー画像 6 チャンネルを入力としているのに対し，提案法では相関画像の実部・虚部の 2 チャンネルを入力としている．提案するネットワークの構造を図 1 に示す．提案法では，FlowNetS のチャンネル数を 1/4 倍，1 倍，2 倍，3 倍と増減させた 4 つのモデルで実験を行う．FlowNetS と提案法のモデルのチャンネル数を表 1 に示す．

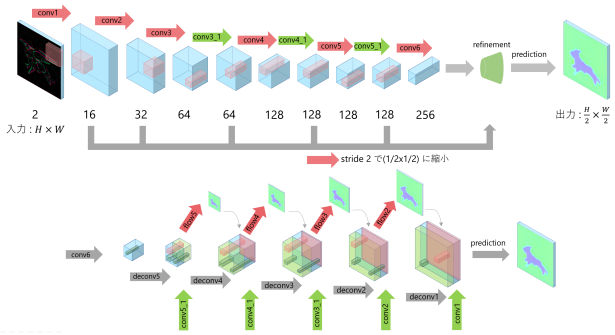


図 1: 提案法ネットワークの構図

表 1: 各ネットワークのチャンネル数

	FlowNetS チャネル数		提案法 1/4 倍		1 倍		2 倍		3 倍	
	in	out	in	out	in	out	in	out	in	out
Conv1	6	64	2	16	2	64	2	128	2	192
Conv2	64	128	16	32	64	128	128	256	192	384
Conv3	128	256	32	64	128	256	256	512	384	768
Conv3_1	256	256	64	64	256	256	512	512	768	768
Conv4	256	512	64	128	256	512	512	1024	768	1536
Conv4_1	512	512	128	128	512	512	1024	1024	1536	1536
Conv5	512	512	128	128	512	512	1024	1024	1536	1536
Conv5_1	512	512	128	128	512	512	1024	1024	1536	1536
Conv6	512	1024	128	256	512	1024	1024	2048	1536	3072
Conv6_1	1024	1024	256	256	1024	1024	2048	2048	3072	3072

FlowNetS では出力画像のサイズが入力画像の 1/4 サイズであるが，提案法では拡張部の層を追加し 1/2 サイズに変更した．さらに，学習時には式 (2) のような各層の予測フロー $\hat{\boldsymbol{f}}_i$ の誤差と重み ω_i での総和をロス関数として用い，各層のロスの計算は，式 (4) に示す予測フローと正解フロー間の全ての画素の差の 2 乗和である End Point Error(EPE) の総和を用いた．ここで，式 (2) 中の $\boldsymbol{f}, \hat{\boldsymbol{f}}$ はそれぞれ正解フローと予測フローを表す．

$$\boldsymbol{f} = \begin{bmatrix} v_x & v_y \end{bmatrix}^T, \hat{\boldsymbol{f}} = \begin{bmatrix} \hat{v}_x & \hat{v}_y \end{bmatrix}^T \qquad (2)$$

$$L_{EPE} = \sum_{i=1}^{6} \boldsymbol{\omega}_i * L_{EPE}(\boldsymbol{f}, \hat{\boldsymbol{f}}_i) \qquad (3)$$

$$L_{EPE}(\boldsymbol{f}, \hat{\boldsymbol{f}}) = \sqrt{(v_x - \hat{v}_x)^2 + (v_y - \hat{v}_y)^2} \qquad (4)$$

3　実験

本手法の有効性を確認するために，シミュレーションを行った．チャンネル数が 1/4 倍のモデルを Model1，1 倍，2 倍，3 倍のものをそれぞれ Model2，Model3，Model4 としてそれぞれで学習を行い，評価用データセットを用いて FlyingChairDatasets により学習済みである FlowNetS(pre-traind) との比較を行った．学習の際，最適化には Adam を使用し，バッチサイズを 8，学習率を 0.0001 と設定した．

3.1　データセット

相関画像は 30 枚の静止画像からマスク付き前景と後景を無作為に組み合わせ，それぞれ一様乱数により運動させ計算した．運動速度の上限は 32 画素/フレームとし，直線・回転運動を含んだ 69,753 枚の画像を生成した．また，過学習を抑えるため train で水平・垂直反転，0°，90°，180°，270° のいずれかの角度で回転，ランダム範囲での切り取りというデータ拡張を行った．

さらに，train/test で用いる画像とは別に評価用データセットとして，運動前後のカラー画像と相関画像を新たに 7830 枚生成した．図 2 に生成した運動前後のカラー画像と相関画像を示す．

図 2: 運動前後のカラー画像と相関画像

4　実験結果

test の際に最良の結果が得られた時の train/test の平均 EPE を表 2 に示す．

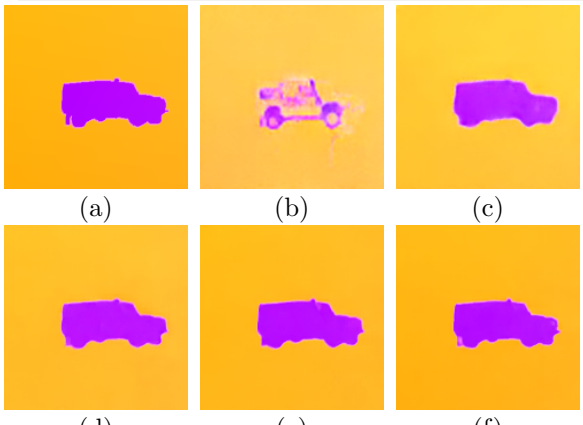
表 2: 平均 EPE

	train(55870 枚)	test(13883 枚)
Model1	0.4863	0.4015
Model2	0.2506	0.2174
Model3	0.1486	0.169
Model4	0.1839	0.1542

平均 EPE は，提案法の 4 つのモデルの中では Model4 で最も良い精度が得られた．また，評価用データを使用して得られた OF と正解ラベルを図 3 に，平均 EPE を表 3 に示す．さらに，正解ラベルとそれぞれのモデルで得られた OF との誤差の大きさをグレイスケール画像として可視化したものを図 4 に示す．これは，pre-traind の最大誤差 0.33 で正規化を行っており，誤差 0 を白，誤差 0.33 以上を黒と表示したものになっている．

表 3: 評価における平均 EPE

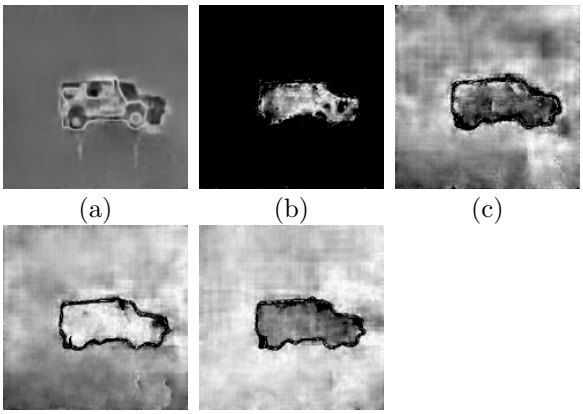
pre-traind	Model1	Model2	Model3	Model4
0.4852	0.5042	0.3646	0.3107	0.3133



(a)　(b)　(c)

(d)　(e)　(f)

図 3: OF と正解ラベル．(a) GT，(b) pre-traind，(c) Model1，(d) Model2，(e) Model3，(f) Model4



(a)　(b)　(c)

(d)　(e)

図 4: 誤差の大きさをグレイスケール画像として可視化．(a)pre-traind，(b)Model1，(c)Model2，(d)Model3，(e) Model4

表 2，3 より，チャンネル数をある程度増加させるにつれて推定精度が高くなっており，pre-traind と同じチャンネル数である Model2 以降では，提案法の方が高精度となっていることが確認できる．このことから，チャンネル数を増加させることで相関画像からより多くの特徴を得ることができていると言える．また図 4 より，pre-traind は全体的に誤差が広がっているのに対して，Model2 以降の提案法は前景と後景の境界に大きな誤差が集中し，全体的に見ると pre-traind よりも誤差が少ないことが確認できる．そのため，2 枚のカラー画像を 1 枚の相関画像に変更することで，pre-traind よりも運動に対する有益な情報が得られていると考えられる．

結果より，相関画像は運動に対する有益な情報を保持していることから，その他のネットワークも同様に入力画像を相関画像に変更することで高精度となることが期待される．

5　まとめ

本研究では 1 枚の相関画像を入力とした深層学習ネットワークを構築し，新たなデータセットを用いて pre-traind との比較を行った．その結果，pre-traind と同チャンネル数である Model2 以降全てのモデルで高精度を確認することができ，相関画像の有効性を示した．

参考文献

[1] A.Dosovitskiy, *et. al.*, ”FlowNet: Learning Optical Flow with Convolutional Networks.” In 2015 IEEE international Conference on Computer Vsion(ICCV), pp.2758-2766, 2015.

[2] IIg, E., *et. al.*, ”FlowNet 2.0:Evolution of optical flow estimation with deep networks.” In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), pp.1647-1655, 2017.

[3] S.Ando, *et. al.*, ”Correlation image sensor: two-dimensional matched detection of amplitude modulated light.” IEEE Transactions on Electron Devices, Vol.50, No.10 , pp.2059-2066, 2003.

[4] T.Kurihara, *et. al.*, ”Optical flow estimation using a correlation image sensor based on FlowNet based neural network.” In Proc. of 15th International Conference on Computer Vision, Theory and Applications(VISAPP), Vol.4, pp.847-852, 2020.