Vol.2022-CVIM-230 No.20
2022/5/12

CNNに基づく任意枚数画像からの直接・大域成分への分解

上田 宇起[1,a)]　王 超[2,b)]　川原 僚[2,c)]　岡部 孝弘[2,d)]

概要：本稿では，プロジェクタ-カメラシステムを用いて，シーンの画像を鏡面反射や拡散反射などの直接成分と相互反射や表面下散乱などの大域成分に分解する手法を提案する．物理モデルに基づいて分解を行う従来手法には，投影パタンのボケにより分解精度が低下するという問題があり，画像撮影のための投影パタンと撮影画像の分解処理の両方に改良の余地がある．そこで提案手法では，データ駆動のアプローチで，任意枚数の画像から，投影パタンのボケに頑健な画像分解を行う．具体的には，畳み込みカーネルを用いて投影パタンを表現できることに着目して，投影パタンと分解処理の両方を，CNN の枠組みで同時に最適化する．実画像を用いた実験を行い，提案手法の有効性を示す．

キーワード：成分分離，照明環境，CNN 応用，畳み込み

1. はじめに

光源に照らされたシーンの画像は，直接成分と大域成分の 2 つの成分で構成されている．直接成分は，光源から出た光が直接照らすことで生じる成分であり，鏡面反射や拡散反射が含まれる．一方，大域成分は，光源から出た光がシーンの他の点を介して間接的に照らすことで生じる成分であり，相互反射や表面下散乱が含まれる．シーンの画像をこれらの成分に分解することは，3 次元形状復元 [1]，シーンの光学的解析 [2]，質感編集 [3] などへの応用に重要である．

従来手法では，大域成分が画像の低周波数成分であることに基づいて，白黒 2 値のチェッカーパタンなどの高周波数パタンをシーンに投影して撮影した画像から，直接成分と大域成分を求めている [1]．この手法では，理想的には 2 枚の画像から成分分解が可能である．しかしながら，実際にはプロジェクタの被写界深度の浅さから焦点ボケが生じたり，カメラの解像度の限界からパタン境界にボケが発生するため理想的な画像が得られない．そのため，理想的な画像を仮定した物理モデルに基づくアプローチでは，少数の画像からの分解精度が悪化するという問題点がある．

分解処理に関しては，スパース性と平滑化に基づく手法 [4] と機械学習を用いた手法 [5] が提案されている．投影パタンについては，物理モデルや信号処理理論に基づいた従来の白黒 2 値のチェッカーパタン [4] や多値パタン [5] が用いられている．これらの研究は，投影パタンと分解処理の一方，または，各々を独立に最適化していること，および，理想的な物理モデルや信号処理理論に基づいて投影パタンを最適化していることに限界がある．

そこで，本研究では，理想的な撮影画像を仮定した物理ベースのアプローチの限界を克服すべく，データ駆動のアプローチで，任意の数の投影パタンで照明した画像から，パタン境界ボケに頑健な直接・大域成分への分解を実現する．具体的には，1×1 の畳み込みカーネルを用いて投影パタンを表現できることに着目して，投影パタンと分解処理の両方を畳み込みニューラルネットワーク (CNN) の枠組みで同時に学習する．

2. 関連研究

2.1 直接・大域成分への分解

Nayer ら [1] は，一般に大域成分が画像の低周波数成分であることに基づいて，プロジェクタから高周波数パタンを投影してカメラで撮影した画像を用いて，直接・大域成分に分解する手法を提案している．具体的には，シーンに互いに明暗が反転したチェッカーパタン 2 枚を投影し，それぞれカメラで撮影する．2 枚の撮影画像から画素ごとに線形演算を行うことで成分分解が可能となる．しかし，プロジェクタの被写界深度の浅さによる焦点ボケや，カメラの解像度の限界によるパタン境界のボケによって理想的な高周波数パタンが投影された画像の獲得は困難である．そのため，2 枚の投影パタンから分解を行うと，格子状のアー

[1] 九州工業大学 大学院情報工学府 情報創成工学専攻
[2] 九州工業大学 大学院情報工学研究院 知能情報工学系
[a)] ueda.takaoki438@mail.kyutech.jp
[b)] c_wang@pluto.ai.kyutech.ac.jp
[c)] rkawahara@ai.kyutech.ac.jp
[d)] okabe@ai.kyutech.ac.jp

ティファクトが目立った分解結果となる．

また，Subpa-Asa ら [4] は，直接・大域成分の空間的な平滑性を仮定し，それらをフーリエ基底または PCA 基底の線形結合として表現することで，単一画像から成分分解を行う手法を提案している．しかし，分解処理のみを最適化しており，分解の重要な手掛かりとなる投影パタンの最適化を行っていない．

Duan ら [5] は，分解処理に機械学習を用いた成分分解手法を提案した．投影パタンについては，ガウスノイズによるコントラストの低下を考慮して信号処理理論に基づき非バイナリ構造で作成している．しかし，投影パタン撮影時，ガウスノイズで表現できないパタン境界ボケや焦点ボケが生じる問題が存在する．そのため，信号処理理論に基づいた投影パタンの最適化には限界がある．

提案手法では，撮影画像から投影パタンも分解処理と同様に機械学習を用いて最適化を行う．これにより，実シーンにおける最適な投影パタンを獲得する．

2.2 センサと画像処理の同時最適化

一般に，ディープニューラルネットワークは，画像処理タスクのエンドツーエンドの最適化のためのツールとして使用される．しかし，ディープニューラルネットワークを画像処理に応用したほとんどの手法では，Duan ら [5] のように撮影済みの画像を入力として，画像処理を最適化しているに過ぎない．

近年，光学イメージングモデルをパラメータ化して撮像層とし，それらの層をアプリケーション層 (画像認識，画像生成，再構成などを行う) に接続し，最終的に逆伝播を利用してデータセットを学習し，センサと画像処理の同時最適化を行う手法が提案されている．Chakrabarti[6] はカラー画像の獲得のためにセンサのカラーフィルタと画像再構成手法の同時最適化を行った．Wu ら [7] は深層学習を利用した深度推定のために，レンズに付けられる位相マスクと画像再構成の同時最適化を行った．

提案手法では，上述のようなセンサではなく，成分分解の鍵となる照明環境に着目し，投影パタンと画像の直接・大域成分への分解処理の同時最適化を行う．

3. 提案手法

3.1 投影パタンの構造

従来手法では，直接・大域成分のへ分解を行う際，高周波数パタンであるチェッカーパタンをシーンに投影し，カメラで撮影した画像を用いる．

提案手法では，通常，図 1 に示すような縦 $A\times$ 横 B のブロックからなる基本パタンの繰り返しにより構成される投影パタンの輝度値を学習により最適化する．

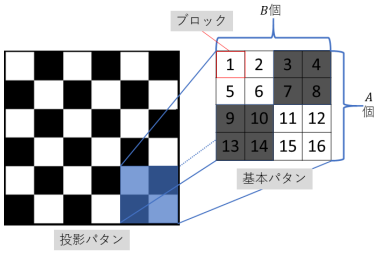

図 1 投影パタンと基本パタンの関係

3.2 畳み込みカーネルによる投影パタンの表現

一般に，図 2 の上段に示すように，任意の照明パタンで照らされたシーンの画像は，重ね合わせの原理により，単一照明に照らされたシーンの画像の線形結合で表現される．単一照明下画像の結合係数を 0 から 1 の任意の値に設定することで，任意の照明パタン下の画像を表現できる．また，図 2 の下段に示すように，CNN における 1×1 の畳み込み処理は，入力画像に対して画像ごとに重みを付与し，画素ごとに合計する処理を行う．これら 2 つの処理は等価である．

提案手法では，基本パタンを 1×1 の畳み込みカーネルの重みで表せることに着目し，CNN の枠組みで学習を行う．具体的には，基本パタンの学習を 1×1 の畳み込みカーネルの重みの学習とみなす．これにより，最適な基本パタン下の画像は，入力である単一ブロック光源下の画像に対して，最適な基本パタンを表現する 1×1 畳み込みカーネルを用いた畳み込み演算により得られる特徴マップとして表現できる．

このことから，図 3 に示すように，成分分解を行う分解ネットワークの前に 1×1 の畳み込み層を接続し，最適な投影パタンの学習を行う．この際，1×1 の畳み込み層の重みに非負制約を課す．これにより，学習された 1×1 の畳み込みカーネルの重みから最適な投影パタンを 1 枚から取得可能となり，カーネルのフィルタ数を変えることで任意の数の投影パタンが取得できる．

3.3 投影パタンと画像分解の同時最適化ネットワーク

本研究では，図 4 に示すように，画像を直接・大域成分に分解する画像分解ネットワークとして，エンコーダ・デコーダ構造にスキップ接続を加えた手法である U-Net[8] に基づいた構造を用いる．このとき，デコーダ部分を 2 つ用意し，エンコーダ部分は共通のものを使用する．この 2 つのデコーダによりそれぞれ直接成分と大域成分への分解を行う．具体的には，投影パタンと画像分解の同時最適化ネットワークに対して，学習時に，$C(=A\times B)$ 枚の単一ブロック光源下画像から $M\times N$ 画素で切り出した $M\times N\times C$ の画素値を入力サイズ，真の直接・大域成分を入力に対応する $M\times N\times1$ のサイズの画素で切り出し正解としてネットワークに与える．また，学習データと同

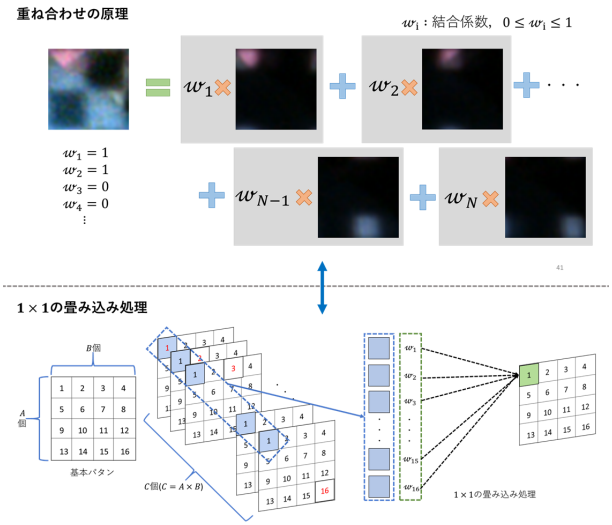

図 2　畳み込みカーネルによる任意の照明パタンの表現

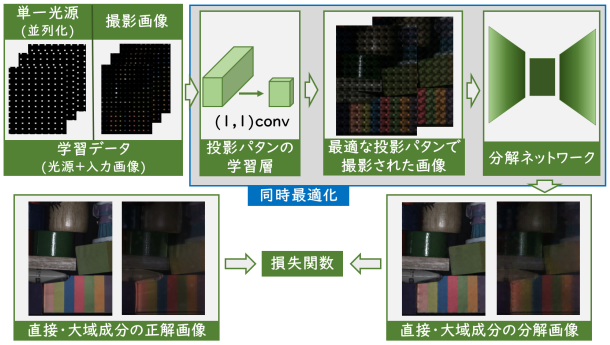

図 3　投影パタンと分解処理の同時最適化ネットワークの学習の流れ

様に交差検証用データも同じサイズでネットワークに与える．最終的にネットワークから直接成分と大域成分の値をそれぞれ $M \times N \times 1$ のサイズで抽出する．テスト時は，出力された $M \times N$ サイズの画像を並べて，シーンの画像の直接・大域成分への分解結果とする．

また，投影パタンと分解処理の同時最適化ネットワークでは，次のような損失関数 $\mathcal{L}$ を採用する．

$$\mathcal{L} = \mathcal{M}(\hat{I}_d, I_d) + \mathcal{M}(\hat{I}_g, I_g) \qquad (1)$$

ここで，$\hat{I}_d$ および $\hat{I}_g$ は真の直接成分および大域成分の画像であり，I_d および I_g は同時最適化ネットワークで抽出される直接成分および大域成分の画素値である．また，$\mathcal{M}$ は平均二乗誤差関数である．これらで構成された損失関数 $\mathcal{L}$ を最小化することによって学習される．

4. 実験

4.1 実験環境

実験では，Crosstour 製の LED プロジェクタ P970 を用いて照明を行い，FLIR 社製のカメラ Blackfly S USB3 を用いて撮影した．このとき，図 5 のように，50-50 プレート型ビームスプリッターを使用して，プロジェクタ画素とカメラ画素の対応がシーンの深度に対して不変になるように，プロジェクタとカメラをビームスプリッターを介して同軸に配置するとともに，画素間の対応を事前に較正している．

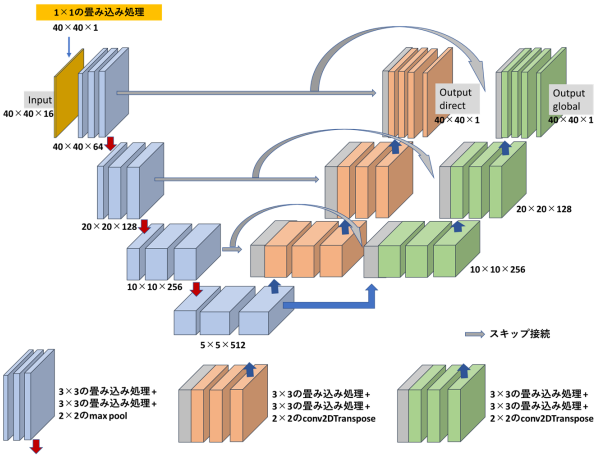

図 4　投影パタンと分解処理の同時最適化ネットワークの詳細

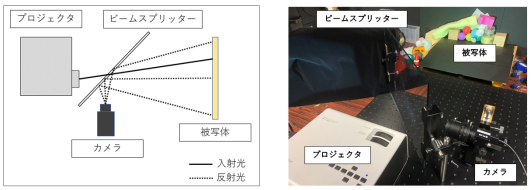

図 5　実験のセットアップ

4.2 学習データとパラメータ設定

図 5 のセットアップを用いて，9 つのシーンに対して撮影を行った．9 つのシーンのうち 6 つを学習データ，1 つを交差検証用データ，2 つをテストデータとした．被写体として，アロマキャンドル，包装紙，ティッシュ，布，ピンポン玉などの表面下散乱や相互反射の生じやすい物体を用いた．

実験では，4×4 ブロックからなる基本パタンを学習することを考えたため，シーンの撮影時，1 ブロックのみが白である投影パタン $16(= 4 \times 4)$ 枚を順に投影して撮影した．このとき，投影パタン全体に対して 1 ブロックのみ点灯するのではなく，4×4 ブロックからなる基本パタンごとに 1 ブロックのみを点灯することで，撮影を並列化して撮影回数を削減した．これにより得られた 6 つのシーンの撮影画像から，$M = 40$，$N = 40$, $C = 16$ として，40×40 画素の領域を規則的に切り取ることで作成した 224,352 枚を学習データとした．また，Nayer ら [1] の手法に基づき，チェッカーパタンをシフトさせながら撮影した 25 枚の画像から真の直接・大域成分の正解画像を作成した．

ネットワークの学習では，最適化アルゴリズムとして Adam[9] を用いて学習率を 0.0001，$\beta_1 = 0.9$，$\beta_2 = 0.999$ と設定し，ネットワークのすべての重みは He の正規分布 [10] を用いて初期化した．また，提案手法では，1×1 の畳み込みカーネルのフィルタ数を変えることで，最小 1 枚から任意の画像枚数で画像分解を行うことができる．例

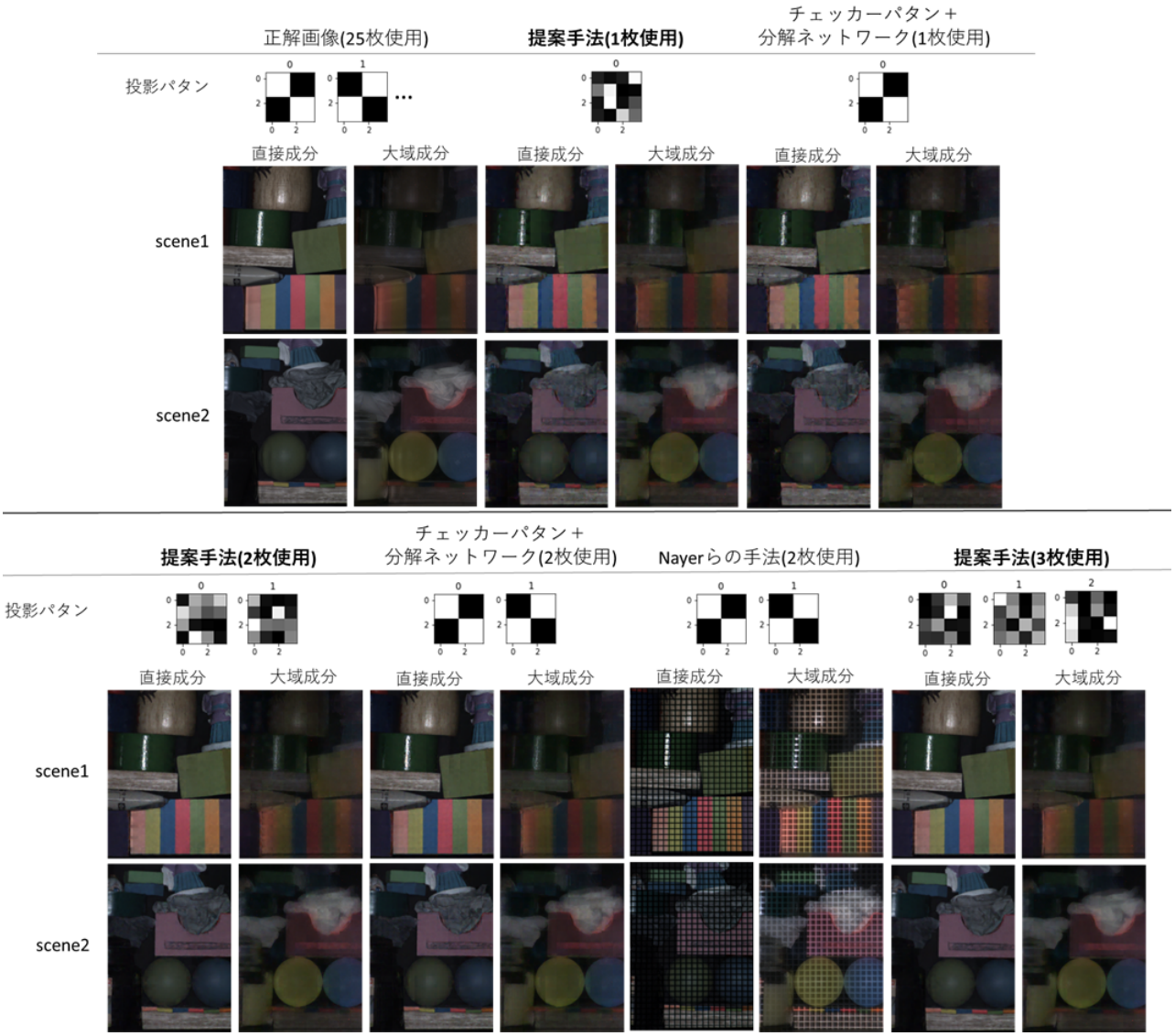

図 6　手法ごとの投影パタンと直接・大域成分の分解結果

表 1　分解画像の PSNR と SSIM

		scene1		scene2	
		直接成分	大域成分	直接成分	大域成分
提案手法 (1 枚)	PSNR	**28.64**	**32.64**	30.33	**33.44**
	SSIM	**0.915**	**0.926**	**0.881**	**0.920**
チェッカーパタン	PSNR	27.97	32.46	**30.56**	32.77
+分解 Net(1 枚)	SSIM	0.907	0.917	0.880	0.911
提案手法 (2 枚)	PSNR	**34.30**	**34.39**	**35.52**	**35.39**
	SSIM	**0.961**	**0.940**	**0.944**	**0.939**
チェッカーパタン	PSNR	34.02	34.05	35.11	34.74
+分解 Net(2 枚)	SSIM	0.959	0.929	0.942	0.931
Nayer ら (2 枚)	PSNR	17.83	17.00	19.99	19.28
	SSIM	0.561	0.530	0.571	0.647
提案手法 (3 枚)	PSNR	**35.74**	**35.18**	**37.58**	**36.50**
	SSIM	**0.972**	**0.944**	**0.962**	**0.947**

えば，フィルタ数を 1 とすることで単一画像からの成分分解が行え，2 とすると 2 枚の画像から成分分解を行う．実験では，フィルタ数を 1，2，3 に設定した 3 通りの場合でネットワークの学習を行った．また，投影パタンを学習することの有効性を示すため，チェッカーパタンを投影した画像を入力とした分解ネットワークの学習を行った．

4.3 学習済みネットワークを用いた成分分解

投影パタンと分解方法の 2 つを同時最適化した後，1×1 畳み込みカーネルより最適な投影パタンを獲得した．また，テストデータを用いて，学習済みネットワークを使用して成分分解を行った結果と使用画像の投影パタンを図 6 に示す．提案手法では少数画像からの成分分解でも格子状のアーティファクトはほとんど発生せず，定性的に良好な結果が得られたことが確認できる．

また，表 1 に PSNR と SSIM による定量評価を示す．投影パタンと分解処理を同時に最適化した場合のほうが，チェッカーパタンを用いて分解ネットワークのみを最適化した成分分解よりも定量的に良好な結果を得られたことが確認できる．これより，画像分解だけではなく，投影パタンも最適化することが重要であるということが分かる．

5. むすび

本稿では，プロジェクタ-カメラシステムを用いた直接成分と大域成分の分解手法を提案した．具体的には，投影パタンと分解処理の両方を CNN の枠組みで同時に最適化することで，任意枚数の画像から投影パタンのボケに頑健な分解を行った．今後の展望として，焦点ボケにも頑健な成分分解の学習や動的シーンへの適用に取り組みたい．

謝辞 本研究の一部は，JSPS 科研費 JP20H00612 の助成を受けた．

参考文献

[1] S. K. Nayer, G. Krishnan, M. D. Grossberg, R. Rasker, “Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination”, In Proc. ACM SIGGRAPH2006, Volume 25 Issue 3, pp.935–944, 2006.

[2] Y. Mukaigawa, Y. Yagi, and R. Raskar, “Analysis of light transport in scattering media ”, In Proc. IEEE CVPR2010, pp.153-160, 2010.

[3] M. Grossberg, H. Peri, S. Nayar, and P. Belhumeur, “Making one object look like another: controlling appearance using a projector-camera system”, In Proc. IEEE CVPR2004, pp.I-452-459, 2004.

[4] A. Subpa-Asa, Y. Fu, Y. Zheng, T. Amano, I. Sato, “Separating the Direct and Global Components of a Single Image”, Journal of Information Processing 26, pp.755–767, 2018.

[5] Z. Duan, J. Bieron, P. Peers, “Deep Separation of Direct and Global Components from a Single Photograph under Structured Lighting”, Computer Graphics Forum Vol. 39, No.7, pp.459–pp470, 2020.

[6] A. Chakrabarti, “Learning Sensor Multiplexing Design through Back-propagation”, In Proc. NIPS2016, pp.3081–3089, 2016.

[7] Y. Wu, V. Boominathan, H. Chan, A. Sanjaranarayanan, A. Veeraraghavan, “Phasecam3dlearning phase masks for passive single view depth estimation”, In Proc. IEEE ICCP2019, pp.1–12, 2019.

[8] O. Ronneberger, P. Fischer, T.Brox, “U-net: Convolutional networks for biomedical image. segmentation”, In Proc. MICCAI2015, pp234–241, 2015.

[9] D.P.Kingma, J.Ba, “Adam:A Method for Stochanstic Optimization”, In Proc. ICLR2015, 2015.

[10] K. He, X. Zhang, S. Ren, J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”, In Proc. IEEE ICCV2015, pp.1026–1034, 2015.