

# 圧縮 Epsilon Photography による撮影後の画像制御

2021 年 8 月

伊 藤 厚 史

学位論文 博士（工学）

圧縮 Epsilon Photography による撮影後の画像制御

2021年8月

慶應義塾大学大学院理工学研究科

伊 藤 厚 史

# 論文要旨

銀塩カメラからデジタルカメラへの切り替えが進み、撮影の利便性は著しく向上した。しかし、撮影者それぞれが持つ独自の好みに合致した「好ましい画像」を得るために、撮影者は未だ、フォーカス位置や絞りのサイズ、シャッタ速度や ISO 感度など、カメラの数多くのパラメータを撮影時に調整しなければならず、撮影後にそれらを変更・制御するのは従来、不可能であった。デジタルカメラ内の計算資源の向上や、光学技術の進歩に伴い、これらを自動で制御する技術は発展しているが、完全に間違えない撮影には至っておらず、ユーザの好みに応じて後から撮影画像を制御したい欲求は満たされない。ライトフィールド撮影は、フォーカス位置と撮影視点について撮影後の画像制御を可能にした。しかし、専用ハードウェアが必要であり、空間解像度を犠牲にして光線情報を取得するため、最終的な画像の解像度が低くなり、フォーカス位置や絞りの大きさの完全にフレキシブルな復元はできない。

それに対し、本論は、従来カメラを用いて、フォーカス位置や絞りの大きさ、露光時間を各々に変えて撮影された複数枚の撮影画像から、あらゆるカメラパラメータで撮影された画像を復元する技術を提案する。言い換えれば、あらかじめ設定されたパラメータによる連写画像を入力とし、従来であれば数千枚の撮像を必要とする完全な画像スタック（例：HDR の Aperture-Focus スタック）を、その 1% にも満たない 16~32 枚程度の撮像数から再構成する。

第 1 章では、本研究の背景と従来の研究を概説する。

第 2 章では、アルゴリズム構築の上で着目した、フォーカス位置・絞り値・露光レベルの全パラメータで撮影された画像スタックの統計的な冗長性を示す。

第 3 章では、数少ない撮影画像から完全な画像スタックを復元するアルゴリズムを解説する。前途の冗長性を活用した混合ガウシアンモデル表現、グリーディ・アルゴリズムにより最適な組み合わせのパラメータ組合せについての説明が含まれる。

第 4 章では、実画像における実験により本手法の効果を示す。Focus スタック, Aperture-Focus スタック, 複数露光レベルの Aperture-Focus スタックのそれぞれについて、定性・定量評価の結果を示す。

第 5 章では、画像スタック再構成が実現するアプリケーション例を示す。高精細な奥行情報を取得する Confocal ステレオ, および、撮影後のカメラ制御による画像のレタッチング, について解説する。

第 6 章では、結論として内容をまとめ、本研究の成果を要約する。

## Abstract

# Compressive Epsilon Photography for Post-Capture Control in Digital Imaging

The transition from film to digital was largely about convenience. While there has been remarkable technological breakthroughs in optical flexibility and computational capabilities of digital cameras, photography still mimics film in some unfortunate ways: the photographer is still required to set all camera parameters such as focus, aperture, exposure, and ISO at capture-time and has limited flexibility in changing these settings post-capture. Though professional photographers have mastered the art of making the correct choices during capture-time, the need to get all the camera parameters correct in the heat of the moment impedes casual photographers from acquiring breathtaking photographs.

While advances in light field photography have enabled post-capture control of focus and perspective, they suffer from several limitations including lower spatial resolution, need for hardware modifications, and restrictive choice of aperture and focus setting. In this paper, we propose “compressive epsilon photography,” a technique for achieving complete post-capture control of focus and aperture in a traditional camera by acquiring a carefully selected set of 16 to 32 images (less than 1% of complete image stack) and computationally reconstructing images corresponding to all other focus, aperture and exposure level settings.

Chapter 1 summarizes the background and prior works.

Chapter 2 describes the statistical redundancies in aperture-focal stacks (including HDR aperture-focal stacks) which we learn in the developed algorithm.

Chapter 3 describes an algorithm for reconstructing the entire image stack from a few captured images. This introduction includes our model of these stacks with a Gaussian Mixture Model by using the redundancies, and a greedy sampling strategy for selecting the best parameter settings for sampling.

Chapter 4 describes extensive results on several real data sets. Reconstructed results of focal stacks / aperture-focal stacks / multi-level exposed aperture-focal stacks are qualitatively and quantitatively analyzed.

Chapter 5 describes examples of application realized with reconstructed image stacks. This introduction includes confocal stereo results which estimates per-pixel fine depth, and post-capture re-touch effects which the user can select any focus, aperture, exposure level setting of choice.

Chapter 6 summarizes the results of this study.

第 1 章 序論.....	1
1.1 研究背景.....	1
1.2 関連研究.....	16
1.2.1 Epsilon Photography .....	16
1.2.2 圧縮センシング .....	19
1.3 本論と先行研究の関係, および, 本論の構成.....	20
第 2 章 Epsilon Photography の冗長性 .....	22
2.1 画素ごとの Intensity Profile .....	22
2.2 Focus スタックの Intensity Profile とエネルギー圧縮特性.....	24
2.3 Aperture-Focus スタックの Intensity Profile とエネルギー圧縮特性 .....	27
2.4 MLAF スタックの Intensity Profile とエネルギー圧縮特性 .....	29
第 3 章 圧縮 Epsilon Photography 手法 .....	32
3.1 提案手法の処理の流れ.....	32
3.2 幾何・輝度キャリブレーション .....	34
3.3 GMM パラメータの学習.....	36
3.4 学習済み GMM パラメータを用いたデータの再構成.....	37
3.5 サンプリング画像の選択.....	42
3.6 HDR 画像合成 .....	45
第 4 章 実験と評価.....	46
4.1 Focus スタックの再構成評価 .....	47
4.2 Aperture-Focus スタックの再構成評価 .....	54
4.3 MLAF スタックの再構成評価.....	65
第 5 章 アプリケーション .....	75
5.1 Confocal Stereo .....	75
5.2 Refocusing ユーザーインターフェース.....	77

5.3 Refocusing + HDR 画像の露出調整機能.....	81
第6章 結論.....	87
6.1 総括.....	87
6.2 課題と展望.....	87
6.3 おわりに.....	88
参考文献リスト.....	89
付録.....	95
謝辞.....	98

# 第1章 序論

序論として、本研究の背景と従来の研究を概説する。

## 1.1 研究背景

銀塩カメラからデジタルカメラへの切り替えが進み、撮影の利便性は著しく向上した。しかし、「好ましい画像」を得るために、撮影者は未だ、フォーカス位置や絞りのサイズ、シャッタ速度やISO感度など、カメラのパラメータを撮影時に調整しなければならず、撮影後に制御できる範囲は限られる。プロのカメラマンは撮影中に正しい選択をする技術を習得しているが、撮影の瞬間ですべてのカメラパラメータを正しく設定する必要があり、カジュアルな撮影者にとって、瞬間を切り出すようなドラマチックな写真を撮ることは非常に難しい。デジタルカメラ内の計算資源の向上や、光学技術の進歩に伴い、これらを自動で制御する技術は発展しているが、完全に間違えない撮影には至っているとは言えず、ユーザの好みに応じて後から撮影画像を制御したい欲求は満たされない。この点において、デジタルカメラは銀塩カメラの「置き換え」を逸脱していないと言える。撮影後の画像制御設定は、デジタル写真における最後のフロンティアのひとつとして根強く残っており、完全なポストキャプチャ制御を可能にする方法が必要とされている。

近年、こうした問題の解決を試みる取り組みが徐々に始まっている。ライトフィールド撮像[1]は、これまでは積分されていた光の角度方向に解像度を持ち、入射角度ごとに光の記録ができる撮像方法である。これにより、撮影後にフォーカス位置、絞り値、そして撮影視点、の変更を行うことを可能とした。マイクロレンズアレイを用いてこれを実現するカメラも商用化されている[2,3]。しかし、この手法は、空間解像度を犠牲にして角度解像度を得るため、最終的な画像の解像度は大幅に減少する。[4]では、異なる絞り値の画像を同時に撮影できるシステム、再構成手段が提案されている。このシステムは4種類の絞りサイズを、ミラーを使って再現しているが、ハードウェア制約は大きく、再構成に限界があった。また、高ダイナミックレンジ(HDR)なライトフィールド撮像を可能にする試み[5]も存在するが、この方式では角度解像度が低減する。総じて、撮影後の画像制御のために、通常の光学系で積分されていた情報を個別に記録するためには、こうした取得情報のトレードオフ、特殊なハードウェアが必要、といった制限が常に問題となってきた。



図 1.1 LYTRO が 2012 年に発売した Light Field カメラ[2]  
世界初の民生用 Light Field カメラ。ユーザは背面のタッチパネルを用いて、撮影後にフォーカス位置を調整することが可能になった。

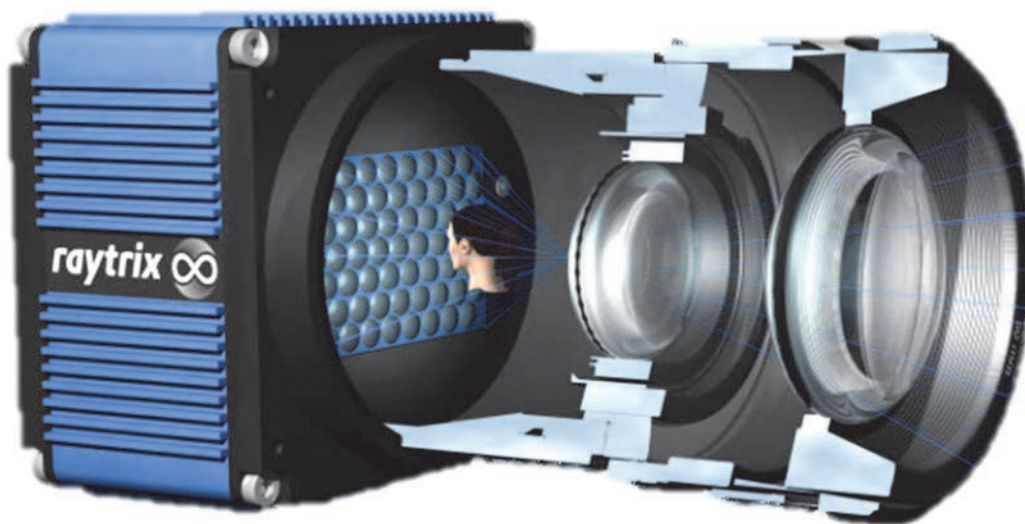


図 1.2 RAYTRIX が 2010 年に発売した Light Field カメラ”R5”[3]  
世界初の商用 Light Field カメラである。Lytro 同様にメインレンズとイメージセンサの間にマイクロレンズアレイを実装していた。



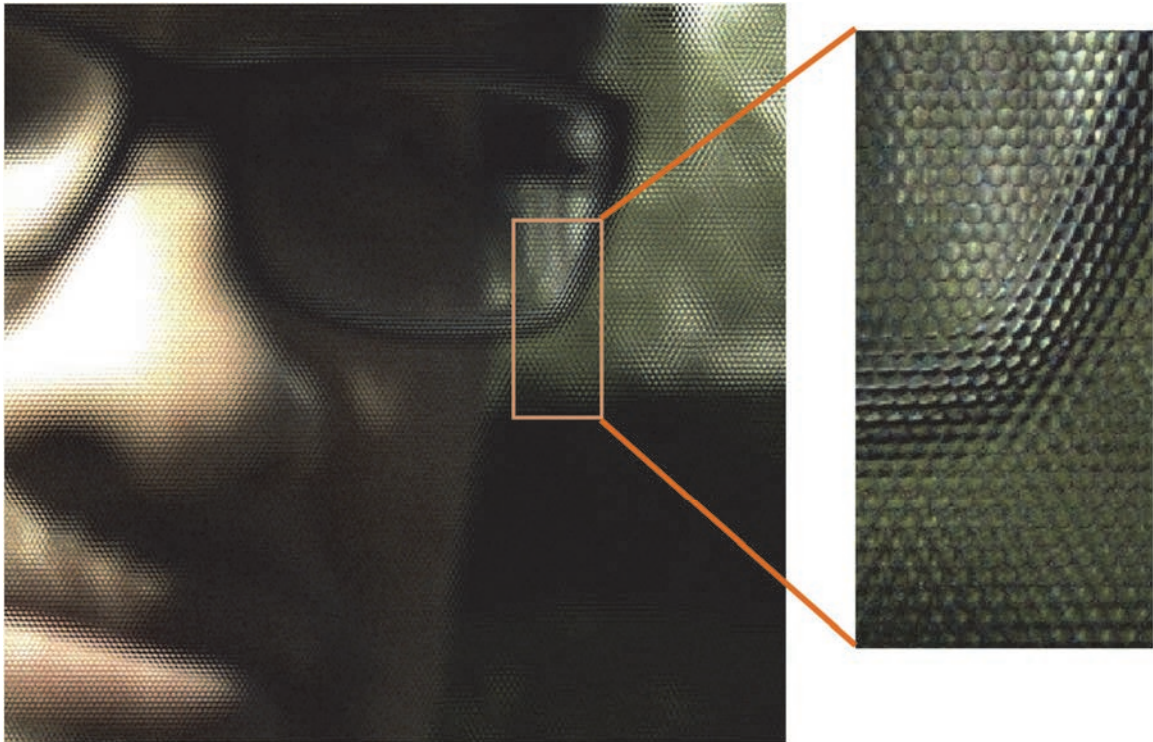


図 1.3 Light Field カメラ[1]の Raw 画像データの例(Image Credit: Todor Georgeiv)

Light Field センサで記録された再構成前の画像. マイクロレンズの効果で空間解像度が低下しているのが見て取れるが, この撮影手段により光線の角度情報を記録しており, この情報を用いて撮影後のフォーカス位置, 絞りサイズの変更を実現している. 一般に, Light Field 撮影においては, 特殊なハードウェアが必要であること, 空間解像度を犠牲にして撮影後の編集効果を実現している, といった課題があった.

一方で、Epsilon Photography[6]という概念は、この代替的な手段であり、1度の撮影においては取得に限界のある情報を、撮影パラメータ設定を変えた複数回の撮影をシーケンシャルに行うことで、撮影後の自由な制御を可能とするアプローチである。例として、フォーカスタックフォトグラフィによるデプス推定[7]やライトフィールド情報復元[8]、フォーカルスweep撮影[9]、露光ブラケット撮影による高ダイナミックレンジ (HDR) 画像生成[10]、複数の絞り値による撮影[4]、Confocal ステレオ[11]、パノラマスティッチング[12]、Lucky Imaging[13, 14]などがある。これらに共通するのは、カメラの撮像パラメータを変えて複数の撮影をすることで、画像を結合して合成イメージを生成する、または、画像とともにその他の付加的で有用な情報 (例えばデプスマップ) を得る、ことを目的としている点である。HDR 合成やフォーカルスweep撮影をはじめとして、こうした技術は一般的になってきている。

一方で Epsilon Photography の課題は、複数枚の撮影を前提としているため、静止または動きがスローな被写体に限定されること、および、データ量が大きくなること、の2点である。この課題は2次元 (またはそれ以上) に撮影パラメータを変えて撮像する場合により顕著となり、例えば[11]の場合、61 のフォーカス位置と 13 の絞り値での個別に撮影が必要となり、793 枚を数時間かけて撮影している。また、本論の 4.3 章で扱う、フォーカス位置と絞り値に加えて露光レベルの3次元に撮像パラメータを変えて撮像する画像スタック (Multi-level Aperture-Focus スタック、表 1.1 に定義を記載) の収集には、51 のフォーカス位置、22 の絞り値、3 の露光レベル、合計 3,366 回の撮像が必要となり、長時間の撮影時間が必要となり、扱うデータ量も膨大になる。

近年では、イメージセンサの応答速度が向上し、ブラケット撮影における連射機能は向上している。図 1.8 で示す CASIO 社製のカメラ“EX-10” [15]においては、9 枚画像の連射において2軸のブラケット撮影機能を有している。これにより例えば、図内で示すような異なるホワイトバランス 3 種類、異なる明るさ 3 種類の 3×3 枚の画像スタックを連射撮影し、撮影後のユーザによる選択を実現している。十数枚程度の連射撮影であれば、短い撮影時間の中で異なる撮像パラメータの画像を取得することが想定可能となってきている。

さらに近年では、ディープラーニング (DNN) 進化の恩恵を受け、限定された入力から近いパラメータの画像の復元が提案されている。通常は露光期間を変えて撮像した複数枚画像から合成される HDR 画像を1枚の撮像画像から生成する手法[16,17]や、DNN を用いて疎な視点から Light Field の再構成を行う手法[18, 19]、1枚の撮像画像から撮影後のリフォーカス機能を実現する手法[20, 21]、などがそれに当たる。

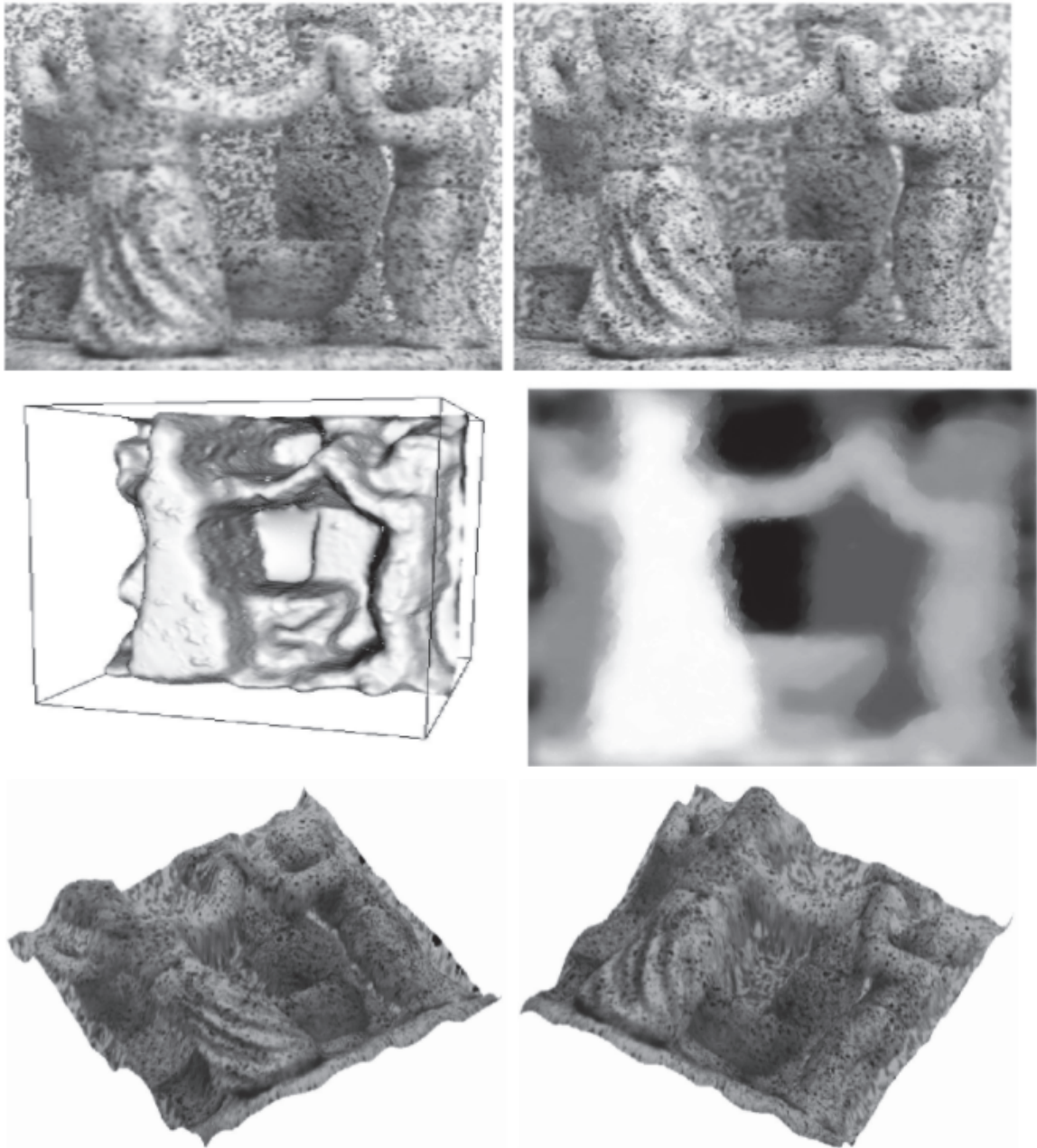


図 1.4 Depth from focus[7]の一例 (Image Credit: P. Favaro)

複数の異なるフォーカス位置で撮像した画像群(上段)から，奥行画像を生成(中段および下段)．フォーカス位置と被写体の奥行に応じて観測されるぼけ量を解析することで，デプスマップを取得する手法である．

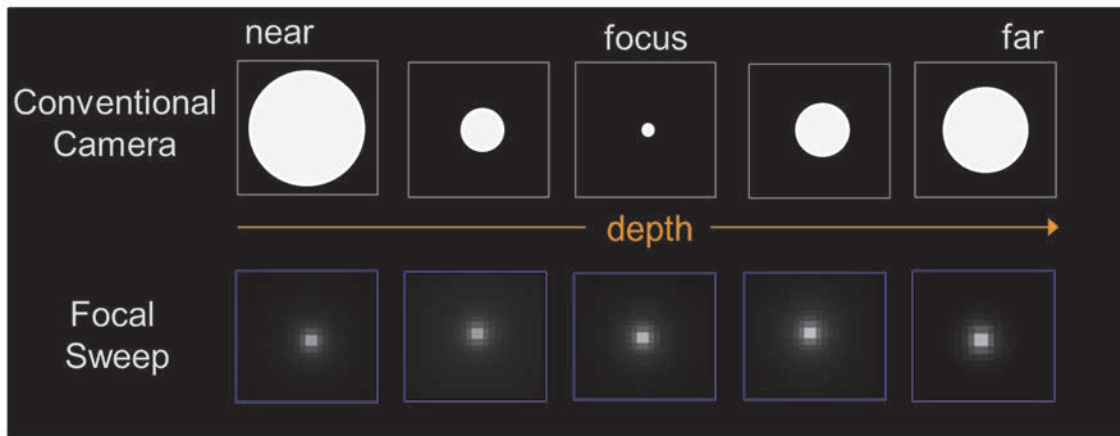
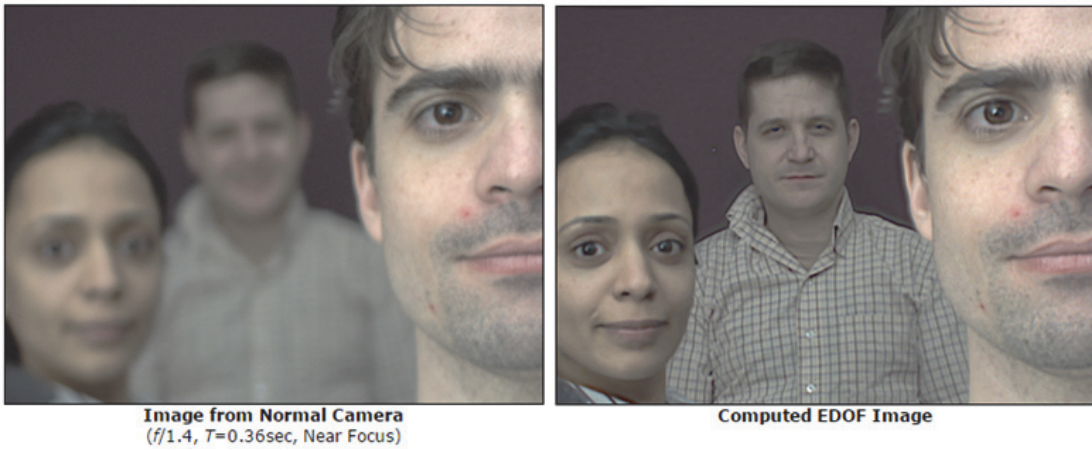
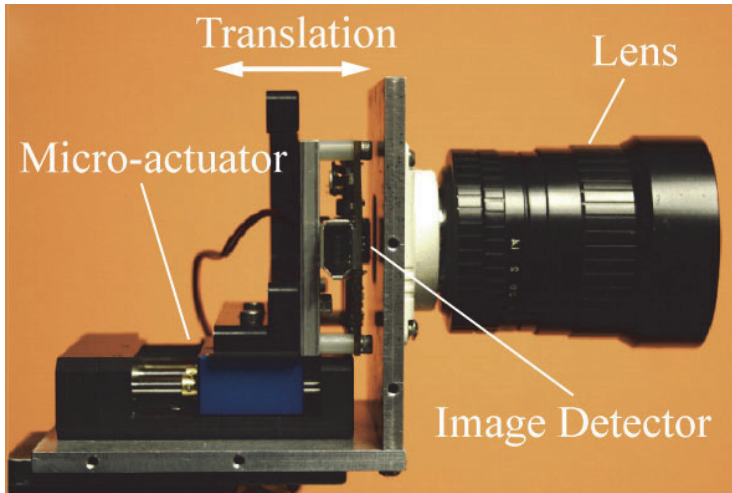


図 1.5 フォーカルスイープ撮影[9]の構成と基本原理(Image Credit: H. Nagahara)  
 フォーカス位置を前から後ろまで動かしながら撮像するカメラを開発し（上段）、  
 手前から背景までフォーカスが合った画像を再構成する（中段）、この撮像方法に  
 より手前から奥までの点広がり関数が不変となることを原理としている（下段）。

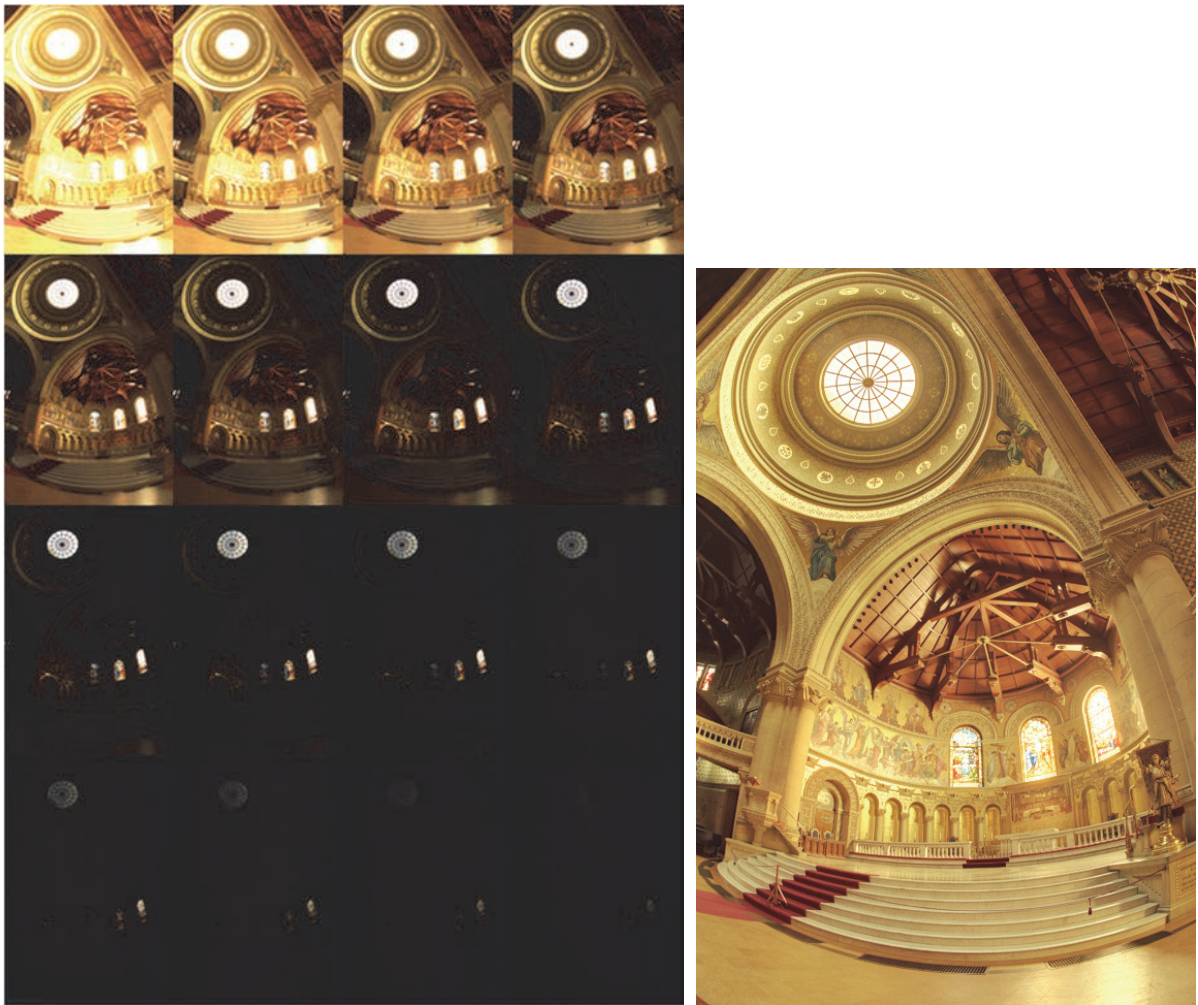


図 1.6 露光ブラケット撮影による高ダイナミックレンジ (HDR) 画像生成[10]の一例 (Image Credit: P. Debebec)

複数の異なる露光時間で撮像した画像群 (左) から, 高ダイナミックレンジの画像を合成し, 暗所から明所まで露出が合った画像を合成する (右). 通常, カメラの連射モード撮影により, 入力画像を取得し, HDR 画像を撮影後に合成する.

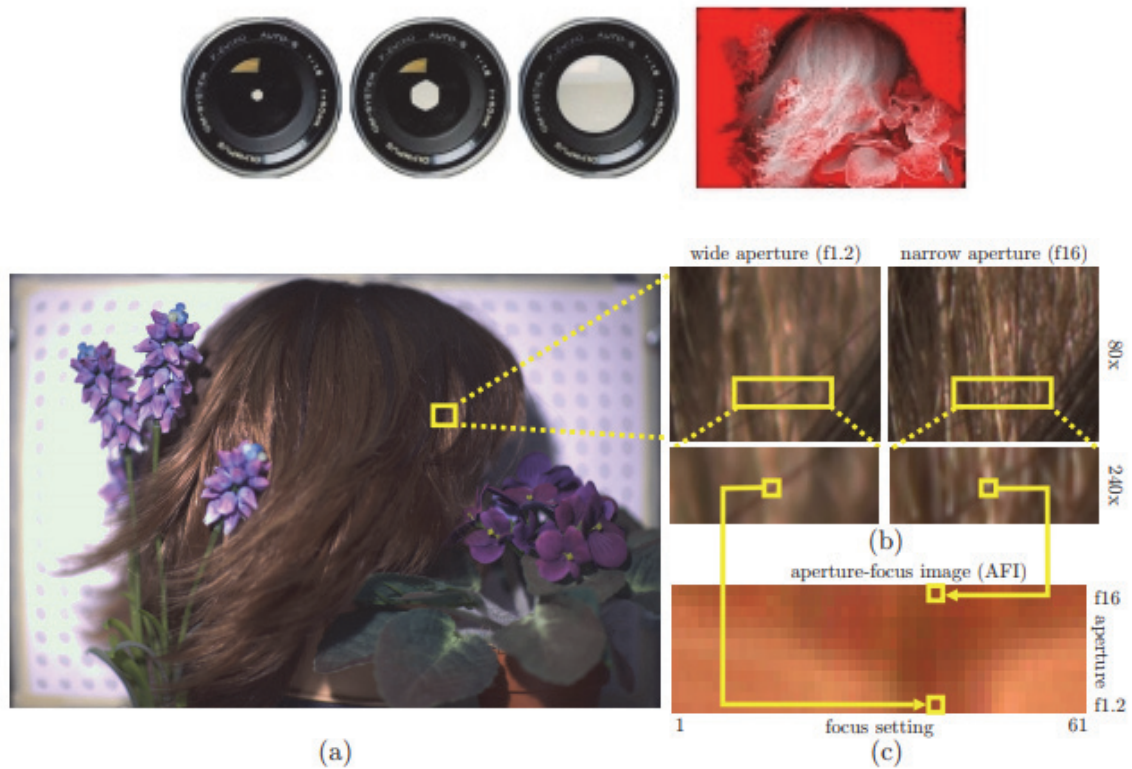


図 1.7 Confocal Stereo[11] (Image Credit: S. Hasinoff)

61 のフォーカス位置と 13 の絞り値での個別に撮影し、高精細なデプスマップ（下段）を取得する。フォーカス位置のみでなく、絞りのサイズも変えることで、より奥行きを推定する情報が得られるため（中段 c）、画素ごとの奥行き値が推定できることで、高い空間解像度のデプスマップを取得できる。

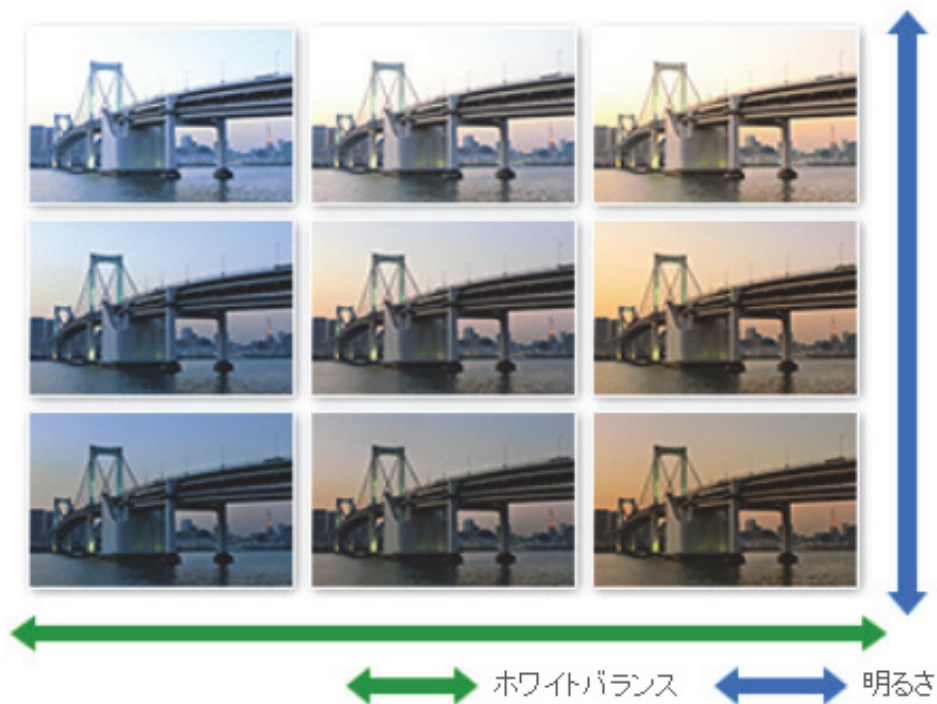


図 1.8 CASIO 社 EXILIM EX-10 [15]における 2 軸ブラケット撮影機能

([http://arch.casio.jp/dc/products/ex\\_10/](http://arch.casio.jp/dc/products/ex_10/))

9 枚画像の連射において 2 軸のブラケット撮影機能を有しており、これにより例えば、図下で示す異なるホワイトバランス 3 種類、異なる明るさ 3 種類の 3×3 枚の画像スタックを連射撮影し、撮影後のユーザによる選択を実現している。

本論の **Key Idea** は、カメラパラメータを変更したときの各画素の画素値の変化について、その冗長性を研究しモデル化することである。さらに、全てのパラメータで撮像された相当の画像スタックを、数少ない入力画像のセット（サブサンプリング）にて再構成する手法の検討である。例えば、**ISO** のみを変更した場合、輝度が飽和しない限り、画素値は概ね線形に近い挙動を示す。他のパラメータを固定にして絞り値を変えた場合、被写界深度と明るさは、連続的に変わっていく。フォーカス位置のみを変更した場合も同様で、各被写体の奥行きに応じてボケ量がスムーズに変化していく。こうしたカメラパラメータ変更に応じた個々の画素の画素値の変化は、低次元かつ低周波的であり、冗長である。この現象は、疎な信号としてモデル化が可能ではないか、と考えた。これが、本論で少数の観察された画像から、カメラの撮像パラメータをすべての値に変えて撮像された画像スタックが再構成できるのではないかと考えたモチベーションである。

図 1.9 および図 1.10 は、本論による提案、圧縮 (**Compressive**) **Epsilon Photography** の概念図である。**Epsilon Photography** を限定された撮像枚数で再構成するアプローチである。言い換えれば、計画的に決められたパラメータによる限定枚の連写撮影によって、撮影後の自由な画像制御を実現する。図 1.9 では、フォーカス位置と絞り値をそれぞれ全ての値に変えて撮像した相当の画像スタック（本論では、**Aperture-Focus** スタックと定義する）を、限定された入力画像から復元する様子を示している。図 1.9 下段で示される 4 枚の画像は、再構成された画像をピックアップしたものであり、そのパラメータは図 1.9 上段の色枠で示されている。それぞれのフォーカス位置、絞りサイズに応じたぼけの復元が示されている。一方で図 1.10 は、フォーカス位置と絞り値に加え、露光レベルも数段階に変えた、全てのパラメータ値で撮像された相当の画像スタック（本論で **Multi-level Aperture-Focus** スタックと定義する。文中では **MLAF** スタックと記載）を復元する様子を示しており、複数露光レベルの情報から高ダイナミックレンジ (**HDR**) の **Aperture-Focus** スタックが復元可能であることを示している。図 1.10 では、フォーカス位置と絞りサイズに加え、3 段階の露光レベルも含めて、全 **MLAF** スタックを復元している。図 1.10 下段では、図 1.9 と同様、上段の色枠で示された撮像パラメータ相当の、再構成された画像をピックアップして表示している。(A)および(B)では低露光、高露光の画像復元を示し、絞りサイズに応じて被写界深度が異なっている様子が見て取れる。また再構成された **MLAF** スタックは、3 段階の露光レベルの画像をそれぞれのフォーカス位置と絞りサイズに対して有するため、各フォーカス位置-絞りサイズに対して、**HDR** 画像の合成[10]が可能である。(C)および(D)ではこの例を示しており、暗所から明所まで視認しやすい明るさにて、異なる絞りサイズの画像が合成できている様子を示している。



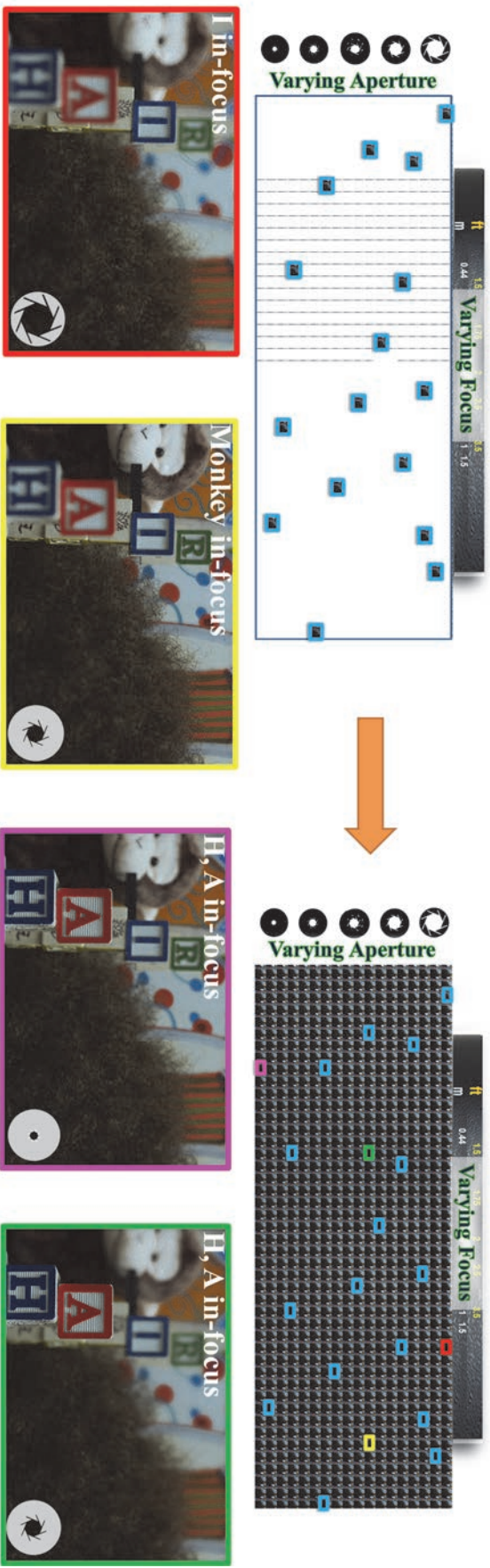


図 1.1 圧縮 Epsilon Photography の概念図

あるシーンを計画的に決められたパラメータによる数枚の連続撮影によって，撮影後の自由な画像制御を実現する．本論では，左上のように限られたパラメータで撮影された数枚の画像から，右上のように全パラメータで撮影された画像スタックを復元する．

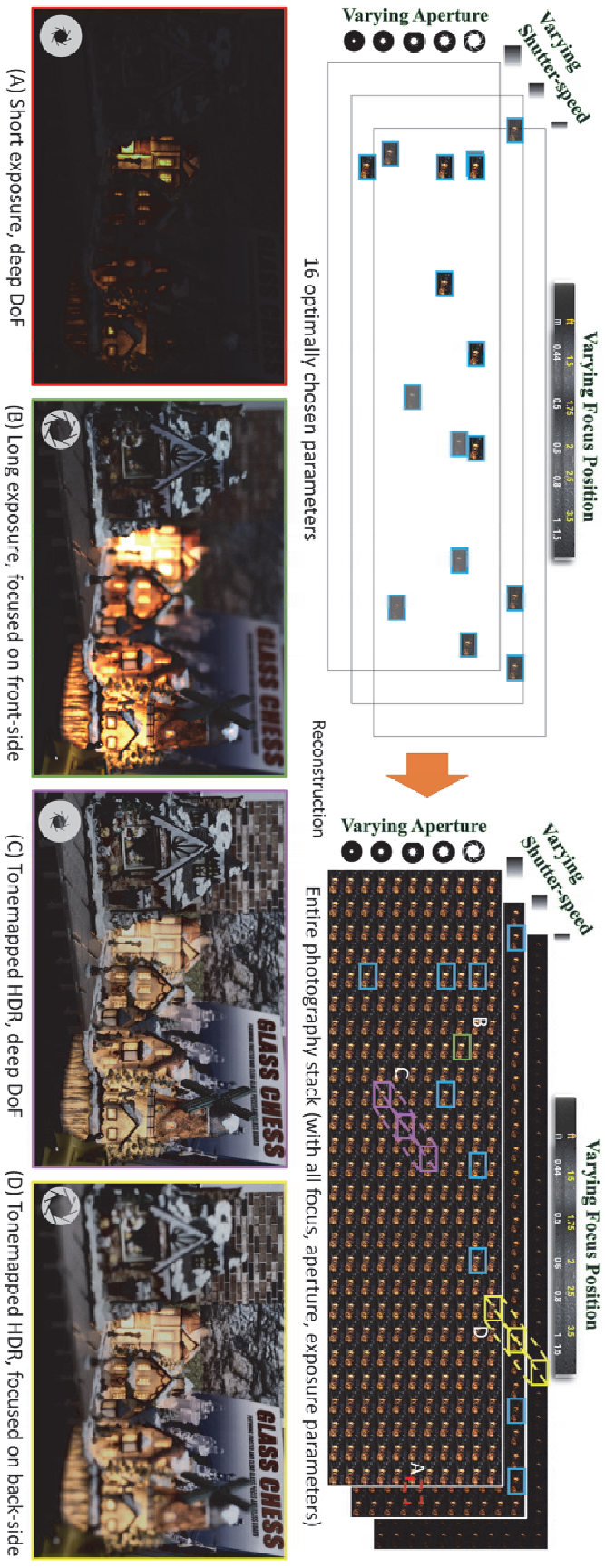


図 1.2 圧縮 Epsilon Photography による Multi-level Aperture-Focus(MLAF)スタックの復元概念図

左上のように限られたパラメータで撮影された数枚の画像から，右上のように全パラメータで撮影されたスタックを復元する．下図(A)(B)はシーン”Night Town”の各パラメータ相当の復元結果の例であり，(C)(D)では露光レベルが異なる画像スタックの復元により，HDR 画像の Refocusing を実現している．

本論の主な貢献は以下である。

●カメラパラメータ，特にフォーカス位置と絞り値，露光期間を変更したときの画素値変化の特性を研究し，低次元で疎なデータであることを確認した．Aperture-Focus スタックを用いたところでは，99%のエネルギー圧縮特性を確認した．

●画素値の特性を各画素単位で混合ガウシアンモデル（GMM）を用いて表現した．

●圧縮 Epsilon Photography の概念を提案し，撮影パラメータを計画的に選択した数少ない入力画像から全パラメータで撮影された相当の画像復元を可能とするアルゴリズムを提供した．本手法によれば，従来であれば，数千枚の撮像を必要とする，2次元や3次元の画像スタックを，その1%にも満たない枚数の撮像により復元する．

●入力画像の撮影パラメータの選択に，グリーディ・アルゴリズムを用いて，再構成の品質を最大化した．

●2次元の Aperture-Focus スタック，および，露光期間を加えた3次元の Multi-level Aperture-Focus(MLAF)スタックの復元を行う実験により，本論のアルゴリズムの有効性を証明した．



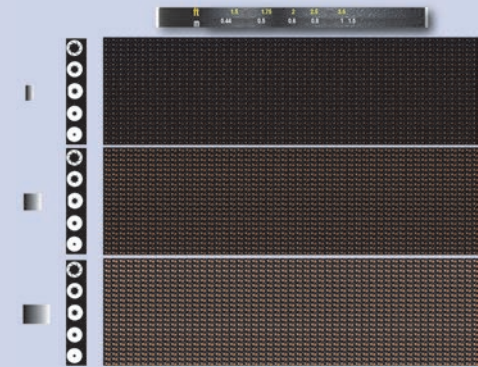
●完全な画像スタックの取得により実現されるアプリケーション（画素単位の高精細な奥行マップを取得する Confocal ステレオや，フォーカス位置・絞りサイズ，露出制御の，撮影後のフレキシブルな編集など）を，限定された枚数の撮像により実現する．

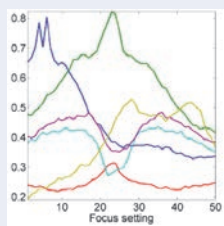
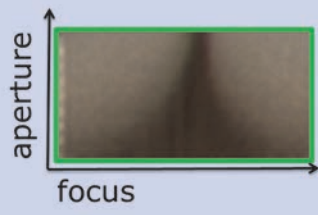
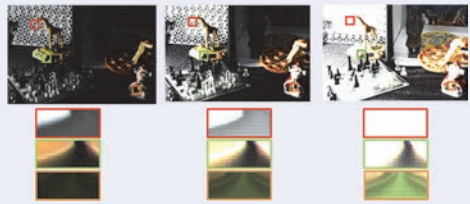
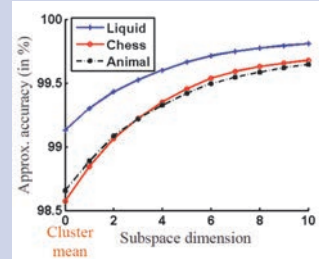
●3次元の MLAF スタックは，希少な HDR 画像の Refocus を実現する画像スタックであり，後続の研究のために収集したデータセットを公開した．

本論の提案する圧縮 Epsilon Photography を構成する重要な部分である，少数画像から完全な画像スタックを再構成するフレームワークでは，Compressive Sensing[22, 23] のアプローチを参考にした．本論では，冗長な信号である Epsilon Photography スタックを，混合ガウシアンモデル(GMM)を用いてモデル化[24]し，少ないサブサンプリングからの全スタックの復元，また，グリーディ・アルゴリズムを用いた最適なカメラパラメータのサンプリングを実現している．限定された枚数の撮像は，デジタルカメラの連射機能と相性がよく，近年のイメージセンサの高感度化や応答速度の高速化の恩恵を受け，実用化検討可能な技術である．

本論では，多くの用語が出現するため，下記の表 1.1 に新出語の定義をまとめた．本論の理解に役立てば幸いである．

表 1.1 本論で用いる用語の解説

用語	意味	図例
Epsilon Photography	1つのカメラを用いて、異なる露光レベル、異なる絞り値、異なるフォーカス位置、異なるISOなど、あらゆる撮像パラメータで撮影する手法、および、撮像された画像データから構成される画像群。	
圧縮(Compressive) Epsilon Photography	上記 Epsilon Photography を限定された撮像枚数で再構成するアプローチ（提案手法）および再構成された画像群のこと。	
Focus スタック	異なるあらゆるフォーカス位置で撮像された画像データから構成される画像群。 Epsilon Photography の1つ。	 <p data-bbox="979 1151 1246 1182">フォーカス位置 近⇄遠</p>
Aperture-Focus スタック	異なるあらゆるフォーカス位置および絞り値で撮像された画像データから構成される画像群。 Epsilon Photography の1つで2次元の Epsilon Photography。	
Multi-level Aperture-Focus スタック (MLAF スタック)	異なるあらゆるフォーカス位置、絞り値および露光レベルで撮像された画像データから構成される画像群。 Epsilon Photography の1つで、3次元の Epsilon Photography。	

<p>Intensity Profile</p>	<p>画素値の集合. 本論では特に, 画像スタックの特定の位置における画素値の集合として用いる. 下記の AFI や MLAFI を含む上位概念.</p>	
<p>Aperture-Focus Image (AFI)</p>	<p>Aperture-Focus スタックの特定の位置における画素値の集合を, 縦横軸に絞り値とフォーカス位置をとって画像として2次元表現したもの.</p>	
<p>Multi-level Aperture-Focus Image (MLAFI)</p>	<p>MLAF スタックの特定の位置における画素値の集合を, 縦横軸に絞り値とフォーカス位置をとって2次元表現し, これを露光レベル数分の画像として表現したもの.</p>	
<p>エネルギー圧縮特性</p>	<p>PCA によりモデル化した際の, 各主成分に対する分散の寄与率を累積した値. 本論では特に Intensity Profile をクラスタリングした信号に対して用いられ, 各主成分に対する分散の寄与率を全クラスについて累積した割合.</p>	

## 1.2 関連研究

### 1.2.1 Epsilon Photography

視野角(FoV), 露出, ISO, 絞り, フォーカス位置, 視点, などのさまざまな撮像カメラパラメータによって取得された複数の画像をベースとし, レンダリングや推論, デブスなどの付加情報を取得する手段は, デジタルイメージングの始まりとともに研究が行われている. カメラ設定への増分( $\epsilon$ , イプシロン)の変更を伴う一連の画像の取得に依存しているため, [6]では **Epsilon Photography** と総称しており, 本論でもこの呼称に習う. 以下, いくつかのカテゴリーに分けて, 先行研究を俯瞰する.

#### 1.2.1.1 複数枚画像を用いた De-noising/De-blurring/Super-Resolution

複数枚画像を用いた典型的なアプリケーションとして, これまで, 画像のノイズ低減が広く行われてきた[25, 26]. 超解像についても同様であり[27], 微小なカメラ動きをサブピクセルでの解像度補間に利用するのが一般的である.

また, 手振れ補正についても複数枚の情報を用いることが有効とされている[28, 29]. これは, 手振れにより発生するぼけは, 点広がり関数(Point Spread Function, PSF)の畳み込みによって数式化されるが, 複数枚画像を用いることで, 1枚の画像内では, ぼけの広がった方向に消失した高周波成分を, 別の画像では別方向にぼけが発生しているものと仮定して, 高周波成分を残留した画像から補い合う, というのが基本的な考え方である.

#### 1.2.1.2 HDR 画像合成

露光期間の異なる複数枚画像を連射撮像(ブラケット撮影)し, その一連の画像から高ダイナミックレンジ(HDR)画像を再構成する手法は, 広く用いられている一般的な手法である[10, 30]. 本論の後半においてもこの手法は登場する. 露光期間の代わりにISOの値を一連に変えて撮像し, ノイズ抑制を考慮しながらHDR画像を再構成する手法も存在する[31]. また, 絞りサイズの異なる複数枚画像からのHDR合成の試みも存在する[32]が, この手法では画像ごとに異なる被写界深度(Depth-of-Field, DoF)を考慮しなければならない.

一方, 近年では, DNNを用いて画像の飽和部分や黒つぶれ部分を復元することで, HDR画像をブラケット撮影ではなく, 1枚画像からHDR画像を合成する手法も提案されている. [18]では, 主に白つぶれ領域の復元にDNNを用いて注力することで1枚画

像からの HDR 復元を実現している。[19]では、光学現象の符号化に DNN を適用することで目的の達成を試みている。

### 1.2.1.3 Focus スタック 画像とその応用

フォーカス位置を後ろから前、あるいは前から後ろへと送りながら (sweeping) 撮像して得られた画像は、Focus スタック画像 (focal stack) と呼ばれる。この情報を用いて奥行マップが取得できることを[7, 33]では示している。また、光学系単体では実現が難しい、手前から奥までフォーカスが合った (Extended Depth of Field, EDoF) 画像の取得が可能であることが知られている[9, 34]。EDoF 画像生成により最も恩恵を受けるアプリケーションの一つは、顕微鏡 (Microscope) 撮像である[35, 36]。超拡大画像は原理的に被写界深度が浅くなり、被写体にフォーカスを合わせるのが手間である。微小に奥行が異なる被写体 (または単一被写体であってもその内の注目領域) が複数ある場合、双方同時にフォーカスを合わせることは難しい。フォーカスタック画像を用いた EDoF 画像生成により、そうした困難から解放される。本論において、まず扱うのはこの Focus スタックの冗長性である。Focus スタック画像の撮像画像を限定しても、シーンの中から後ろに渡りフォーカス位置が変わる完全なスタックが再現可能であることを示す。さらに奥行マップの生成、EDoF 画像の生成、といった代表的なアプリケーションが再構成スタックによって実現できることを示す。

### 1.2.1.4 Aperture スタック 画像とその応用

[32]では、絞りサイズを光学系が許容する最大値から最小値まで変更しながら画像を収集することで、被写界深度の撮影後のコントロールや、先に述べた HDR 画像の合成が可能であることが示された。また、[4]では特殊なハードウェアを構築し、同心円状に形成された絞りでの画像の同時撮像を提案した。撮影された複数枚の画像を処理することで、同被写体であっても被写界深度が撮影後に制御可能であることを示し、奥行き、Light Field 取得といったアプリケーションを実現した。

### 1.2.1.5 Confocal Stereo

著者の知る中で最も多い枚数の画像撮像を単一カメラで実行している Epsilon Photography である、[11]では、シーンをすべてのフォーカス位置、および、絞りサイズで撮像することで、非常に詳細な奥行マップが取得できることを示している。他の Epsilon Photography が 1 次元の撮像パラメータを変更して取得して、所望の効果を得ているのに対し、[11]では 2 次元の撮像パラメータを変更して取得する。代表的なアプリ

ケーションは、1.2.1.3 章で述べたフォーカスタックと同様に奥行マップであるが、Depth from Focus や Depth from De-Focus, および Stereo 法により推定されるデプスマップは、ほとんど空間的近傍の平滑性の過程を用いて生成される[37, 38]のに対し、Confocal Stereo は近傍の情報に頼ることなく、1画素におけるフォーカスと絞りのプロファイル (Aperture Focus Image, AFI と[11]では定義されており、本論もこれに習う) から奥行を生成できる特徴を持つ。そのため、非常に高解像度なデプスマップになり(図 1.7 に例を示す)、人間の髪の毛のような被写体についても精細なデプスマップを生成できる。

しかしながら、約 1m の奥行幅の被写体に限定した場合でも、フォーカス 61 点、絞り 13 点、合計  $61 \times 13 = 793$  枚の画像撮影が必要な手法であり、完全な静止シーン以外ではこの手法は実用的ではなかった。本論では、画素ごとのフォーカスと絞りのプロファイル(AFI)における冗長性を利用することで、限定枚数の撮像画像から全体の Aperture-Focus スタックのデータを再構成することを可能にする。これにより、非静止シーンでも Confocal Stereo 法を実用的にする。

#### 1.2.1.6 Light-field 画像取得

Light Field 撮像は、そのフレキシブルな撮像後の画像制御が可能なることから広く研究され[1, 39]、商用カメラもリリースされてきた[2,3]。イメージベースのレンダリング手法として一般的なものとなっている。Light Field 表現によれば、撮像後のフォーカス位置制御、絞り制御、視点の変更が可能である。一方で、空間解像度を犠牲にしてそうした撮像後の制御を可能としているという課題があった。これを解決するために、光学系の工夫により、空間解像度と光線解像度のトレードオフバランスを操作し、フォーカス位置制御等の効果は低下するものの、空間解像度劣化が限定的になる、Plenoptic2.0 と呼ばれる方式[40]が提案されており、また、[2]の Light Field カメラと一眼レフカメラを並べたハイブリットシステムを構築し、低下した Light Field 画像の解像度を補償する手法[41]などが提案されてきた。しかしながら、これらも各手法に応じてマイクロレンズアレイや、数十台の複数カメラが必要になるといったハードウェアの制約が必要になることが課題は依然として残っている。

Epsilon Photography を入力とする Light Field 取得方法としては、フォーカスタック画像を用いた[8]、あるいは、絞りスタック画像を用いた Light Field の再構成も提案されている[9]。また、HDR でのライトフィールド撮像を可能にする試みは、Plenoptic2.0 において、ハードウェアに工夫を加えた方法[5]として存在するが、先述の通り、この方式は角度解像度を犠牲にするトレードオフがある。本手法の提案する撮像後制御は、これら従来の手法と異なり、1台の通常カメラによる撮影で可能となり、かつ、空間解像度を犠牲としないため高精細である。近年の研究として、HDR のライトフィールド画像



データセットを公開し、視点生成や tonemap の研究を促す研究[42]が存在し、こうしたデータへの要求が高まっているが、本手法はこうしたデータセットを、通常カメラによる限定された複数枚の撮影で復元できる。

## 1.2.2 圧縮センシング

本論の提案する圧縮 Epsilon Photography を構成する重要な部分である、少数画像から完全な画像スタックを再構成するフレームワークでは、Compressive Sensing のアプローチを参考にした。関連研究について記す。

### 1.2.2.1 Compressive sensing

Compressive sensing は、観測対象データがある表現空間では「Sparse (疎)」であると仮定して、必要とする未知数の数よりも少ない観測データから、ある条件の下で対象を復元する手法である[22, 23]。この手法は、疎な信号の特徴を勘案して、L1-norm の最小化によって実現される。先行研究においては、Light Field[43, 44]、光線伝達行列(Light Transport Matrices)[45]、動画[46]やハイパースペクトラル画像データ[47]などが復元されてきた。本論は、この着眼点に習い、復元目的の信号の Sparse 性に着目・明らかにし、少ないサンプルデータから元信号を復元する課題を実現する。

### 1.2.2.2 Gaussian Mixture Model (GMM)

圧縮センシング応用のための有効なモデルとして、ガウシアン混合モデル(Gaussian Mixture Model (GMM))による表現が効果的であることが知られている[24, 48]。これを用いて、一般の画像復元[49]、動画復元[50]、Light-Field 画像の復元[51]などといった先行研究でのアプリケーション事例があげられる。GMM は、圧縮コンピューショナルカメラでの実用事例も存在している[52]。本論では、少ない撮像画像から、完全な Epsilon 画像を復元するために、GMM モデルでの再構成手段を用いる。さらに、GMM モデルを用いて、撮影する (サンプリングする) 画像のカメラパラメータの選択をグリーディ・アルゴリズムにより最適化する。

### 1.3 本論と先行研究の関係, および, 本論の構成

本論を構成する主論文は[53]および[54]の二編である. 先行研究と比較した際の位置づけは, 下記の表にまとめられる.

表 1.2 本論と先行研究の関係

		特殊HWで 実現	通常HWの連射で実現: Epsilon Photography				
種類		Light Field 撮像 [1] et. al.	露出ブラ ケット撮影 [10] et. Al.	フォーカス スタック撮 影[7] et. al.	Confocal Stereo [11]	圧縮Epsilon photo (proposed) [53]	圧縮Epsilon photo (proposed) [54]
カメラ パラ メータ	フォーカス 位置	○		○	○	○	○
	絞りサイズ	○			○	○	○
	露光レベ ル		○				○
撮像数		1枚	3~ 十数枚	10~ 数十枚	1000~ 数千枚	8~16枚	16~32枚
Note		空間解像度 が大幅に劣 化	撮影に時間が掛かる. 特に2次元以 上のカメラパラメータを扱う場合に は, 非常に遅く, 静止被写体に限定 される			2次元以上のカメラパラメー タを扱う場合でも, 限定枚数 の連射で実現可	

本論が再構成の目標とする画像スタックは, フォーカス位置と絞り値の2次元にカメラパラメータを変更して撮像した画像スタック (Aperture-Focus スタック) [53]と, フォーカス位置, 絞り値に加え, 露光レベルを変更して3次元にカメラパラメータを変更して撮像した画像スタック (Multi-level Aperture-Focus スタック, MLAF スタックとも呼称) [54]の2つである. 先行研究との関係としては, まず Light Field 撮像と比較すると, 特殊な HW を必要とせず, かつ, 最終的に得られる画像の空間解像度が劣化しない. 露出ブラケット撮影やフォーカススタック撮影といった1次元にカメラパラメータを変更した撮像では, HDR や Depth From Focus によるデプスマップ取得などの効果が期待できるが, 撮影後に操作できるパラメータは1つに限定される. 一方, 2次元にカメラパラメータを変更し, Aperture-Focus スタックを撮像する方法 (Confocal Stereo [11]) で

は、必要な撮像数が大幅に増え、1,000枚ほどの撮影が必要となり、非常に遅く、静止被写体に限られる。これに対し、提案手法では、1,000枚レベルの2次元にカメラパラメータを変更する Aperture-Focus スタックの取得に、8~16枚の画像のみを必要とし[53]、3,000枚レベルの3次元にカメラパラメータを変更する MLAF スタックの取得には、16~32枚の画像のみを必要とし[54]、いずれも通常カメラの連射撮影での取得も現実的になる。3次元にカメラパラメータを変更した画像スタックを用いた研究は著者の知る限り存在せず、新規性を主張するところである。[42]では、Light Field カメラを用いて HDR 画像を収集し、データセットを公開している類似の研究であるが、本論で扱う MLAF スタックは、より希少な解像度劣化のない HDR の Aperture-Focus スタックを合成可能である。本論の貢献は、3次元にカメラパラメータを変更して撮像した MLAF スタックを、限定枚数の撮像から再構成することであるが、それに加えて、後続の研究のために、収集した MLAF スタックのデータセットを公開している点にもある。

以上の議論を踏まえて、本論文の構成を以下に示す。

第2章では、アルゴリズム構築の上で着目した、フォーカス位置・絞り値・露光レベルの全パラメータで撮影された画像スタックの統計的な冗長性を示す。

第3章では、数少ない撮影画像から完全な画像スタックを復元するアルゴリズムを解説する。前途の冗長性を活用した混合ガウシアンモデル表現、グリーディ・アルゴリズムにより最適な組み合わせのパラメータ組合せについての説明が含まれる。

第4章では、実画像における実験により本手法の効果を示す。Focus スタック、Aperture-Focus スタック、複数露光レベルの Aperture-Focus スタックである MLAF スタック、のそれぞれについて、定性・定量評価の結果を示す。

第5章では、画像スタック再構成により実現されるアプリケーションの例を示す。高精細な奥行情報を取得する Confocal ステレオ、および、撮影後のカメラ制御による画像のレタッチング、について解説する。

第6章では、結論として内容をまとめ、本研究の成果を要約する。

# 第2章 Epsilon Photography の冗長性

Epsilon Photography[6]は、1つのカメラを異なる露光期間, 異なる絞り値, 異なるフォーカス位置, 異なるISOなど, あらゆるパラメータで撮影された画像データから構成される. この空間を埋める画像データはあまりに大きい, 空間の冗長性をモデル化することにより, 少ないサンプリングデータで再構成することが可能となる. 本章では, Epsilon Photography を異なる切り口で概観することで, その冗長性について明らかにする. この冗長性が, 本論の提案する“Compressive (圧縮) Epsilon Photography”のモチベーションである.

## 2.1 画素ごとの Intensity Profile

本論は実撮影並の画質での再構成を目指すものであるため, パラメータが異なる画像のスタックを画素ごと (Per-Pixel) でのモデル化を採用する. Per-Pixel モデルにより, 画像の解像度は保たれ, 平滑化を前提としないため, 1画素レベルのエッジなど精細な情報の復元が期待できる.

$I_{u,v}(f, \alpha, i, s)$ を画素位置 $(u, v)$ における画素値の集合 (Intensity Profile) と考える. フォーカス位置 $f$ , 絞り値 $\alpha$ , ISO感度 $i$ , 露光期間 $s$ と表すが, 画像の明るさレベル $l$ は任意定数 $C$ を用いて,  $l = C\alpha^2 * i * s$ の関係で表すことができる. ISO感度 $i$ を固定とするとき,  $I_{u,v}(f, \alpha, i, s) = I_{u,v}(f, \alpha, l(\alpha, s))$ と表されるところ, 本論では, 絞り値 $\alpha$ の変化に応じた露光期間 $s$ の制御により,  $l_j = 1, 2, \dots, L_n$ と線形かつ離散的な $L_n$ 段階に画像の明るさ (露光) レベルを調整することとし, また,  $l_{j+1} = 2l_j$ となるように明るさのステップを刻むように調整した. このとき,  $I_{u,v}(f, \alpha, l_j)$ を, 画像の露光レベル $l_j$ における Aperture-Focus Image (AFI)と呼ぶ[11]. 図 2.1 は, AFI の一例を図示したものである. AFI とは, Aperture-Focus スタックの特定の位置における画素値の集合を, 2次元表現したものと言え, 言い換えれば, 縦軸に絞り値, 横軸にフォーカス位置ごとに, 各画素位置における画素値を並べ替えた画像として表現したものである. この呼称を踏襲し, 全ての露光レベルを含む $I_{u,v}(f, \alpha, l)$ 自体を Multi-level Aperture-Focus Image (MLAFI)と表現する. 図 2.2 は, この MLAFI の一例を図示したものである. MLAFI とは, MLAF スタックの特定の位置における画素値の集合を, 複数の2次元画像として表現したものと見え, 言い

換えれば，露出レベルごとに縦軸に絞り値，横軸にフォーカス位置ごとに，各画素位置における画素値を並べ替え，露出レベルの段階数である $L_n$ 枚の画像として表現したものである．図 2.2 の例では，露出レベル数は3であるため，MLAFI は，各画素位置において3枚の AFI で構成されている，と言える．

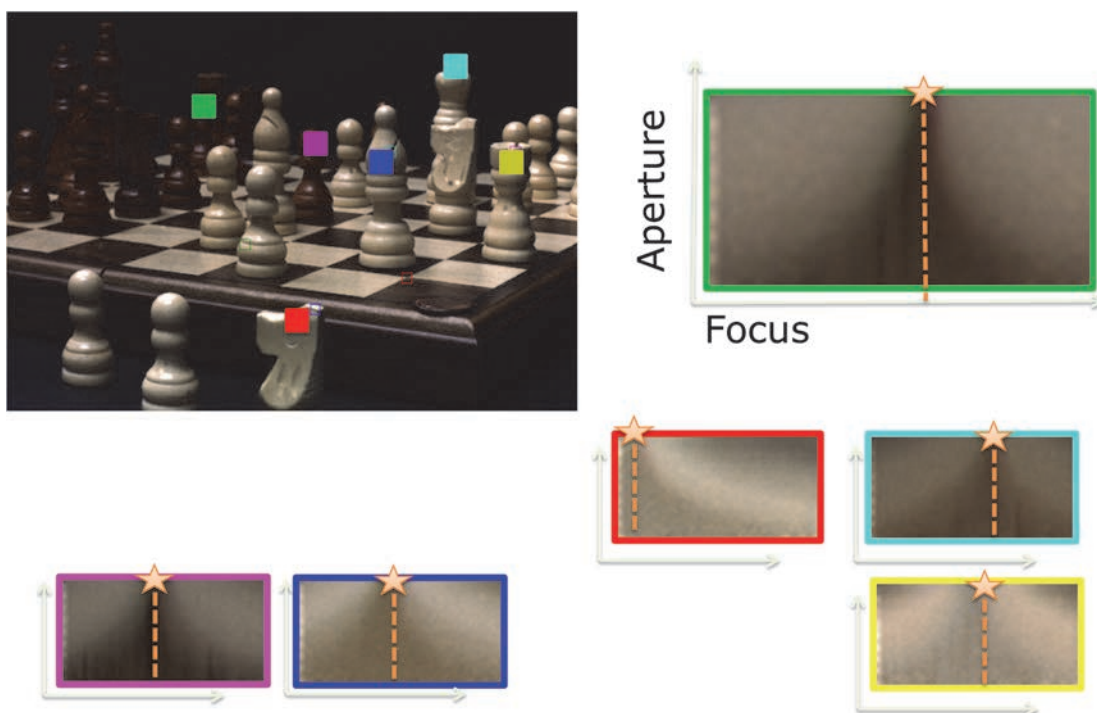


図 2.1 Aperture-Focus Image (AFI)の例:

Aperture-Focus スタックから，各色の四角が示す画素位置について，全てのフォーカス位置と絞り値の場合の画素値の集合 (Intensity Profile) を，縦軸に絞り値，横軸にフォーカス位置，に並べ替えた画像として表現している．AFI においては，あるフォーカスの値で，縦軸上の値が変化しない位置が存在し，その位置は，当該画素のカメラからの距離によって決まる．また，絞り値が小さいときに，横軸方向の輝度値が変わらないこと見て取れる．一概に，その形状が連続的であり，スムーズな変化であり，緩やかな凸形状が AFI 上で現れる位置は，被写体のカメラからの距離のみにより決まる．

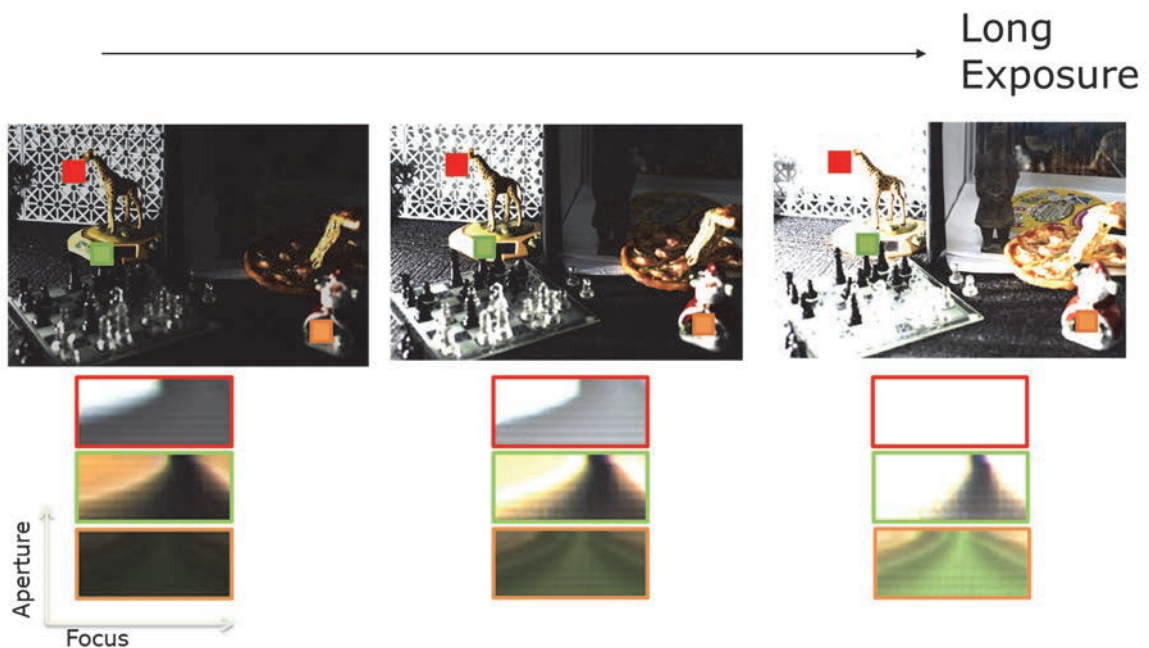


図 2.2 Multi-level Aperture-Focus Image (MLAFI)の例:

MLAF スタックから、各色の四角が示す画素位置について、全てのフォーカス位置と絞り値の場合の画素値の集合 (Intensity Profile) を、縦軸に絞り値、横軸にフォーカス位置、に並べ替えた露光レベル枚数分の画像として表現している。3つの各露光レベルにおける AFI の特徴は、Aperture-Focus スタックと同様である。一概に、それぞれの AFI は、その形状が連続的であり、変化はスムーズで、緩やかな凸形状である。一方で、各露光レベルの AFI 間の関係は、白飛びや黒つぶれしている撮像パラメータの場合の画素値を除き、明るさが露光レベルに対し、線形に増減するのみである。これをまとめた MLAFI もやはり非常に疎な情報であると言える。

## 2.2 Focus スタックの Intensity Profile とエネルギー圧縮特性

図 2.3 は、カメラパラメータの中でフォーカス位置の値のみを連続的に変えて撮像した Focus スタックについての解析を示している。ここで用いた3つのデータセットは、Canon の一眼レフカメラに焦点距離 50mm の固定レンズを用いて撮像した。絞り値は F/1.4, ISO は 100 で固定している。図 2.3(a)では、3つの Focus スタックデータセットにおいて、カラー一点で示された位置の Intensity Profile (画素値の集合) を、横軸を45点のフォーカス位置としてグラフとして示したものである。一般に、F/1.4 のような大き

な絞り値において、その形状は1つの山または谷を持つ単調な形状 (unimodal) になる。これはフォーカスが合致した点では、画素位置の信号が周辺と混ざり合うことなく観測されることで、山または谷が形成され、被写体距離からフォーカス位置が外れていくにしたがって、周辺と混ざり合った画素信号が観測されるためである。

図 2.4 では、いくつかの画素位置における Intensity Profile を、大きな絞り値 ( $F/2$ ) と、小さな絞り値 ( $F/16$ ) で比較している。大きな絞り値の場合には、山や谷の形状が現れることに対し、小さな絞り値の場合には、平坦に近い形状になっていることが見て取れる。これは、大きな絞り値の場合には、フォーカスが合っている位置では、周辺とほとんど混ざっていない信号が観測され、フォーカスが合っている位置以外では、周辺と混ざり合ったぼけによる画素値が観測されていることを示している。一方、小さい絞り値の場合には、被写界深度が深く、広い範囲のフォーカス位置で周辺とほとんど混ざっていない信号が観測されることを表している。

Focus スタックの冗長性を、k-means クラスタリングを実行して示す。各画素位置の Intensity Profile について平均と標準偏差を正規化して、これをクラスタリング処理する。クラス数を 100 としてクラス分割してみると、上位 50 クラスに 80% の画素の Intensity Profile が所属した。図 2.3(c) は、上位 9 クラスの正規化済み Intensity Profile の平均と分散を示している。この観測からわかる通り、フォーカス位置の変化に応じたぼけが作り出す Focus スタックの Intensity Profile の勾配は低周波的な変化であり、低次元でのモデル化が期待できる。

図 2.3(b) では、Intensity Profile を 100 クラスに分離した際のエネルギー圧縮特性を示している。これは、各クラスについて PCA (主成分分析) を用いてモデル化した際の、各主成分に対する分散の寄与率を全クラスについて累積した割合である。図 2.3(a) に示した 3 シーンにおいて、クラスタ中心のみでモデル化した場合、すでにすべてのシーンで 98.5% を超えるエネルギー圧縮特性が見て取れる。さらに、PCA における主成分をクラスタ中心から 10 次元まで考慮すると、3 シーン全てでエネルギー圧縮特性は、99.5% に到達する。Focus スタックが非常に "Compressive" な情報であることが見て取れ、低次元でのモデル化が可能であるという仮説を裏付けている。本論の注目するポイントである。

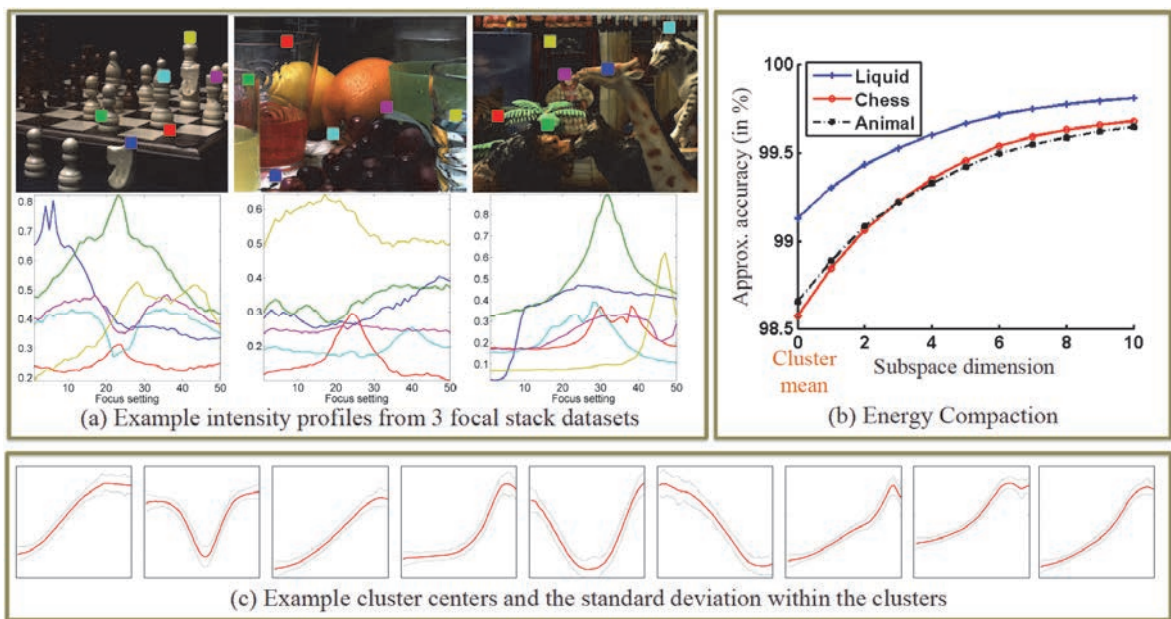


図 2.3 Focus スタックの冗長性:

(a) 3つのデータセットからピックアップした画素位置においてフォーカス位置に応じた Intensity Profile のグラフ. (b) Intensity Profile を k-means クラスタリングし 100 クラスに分離した際のエネルギー圧縮特性. (c) 正規化した上位 9 クラスの平均と分散

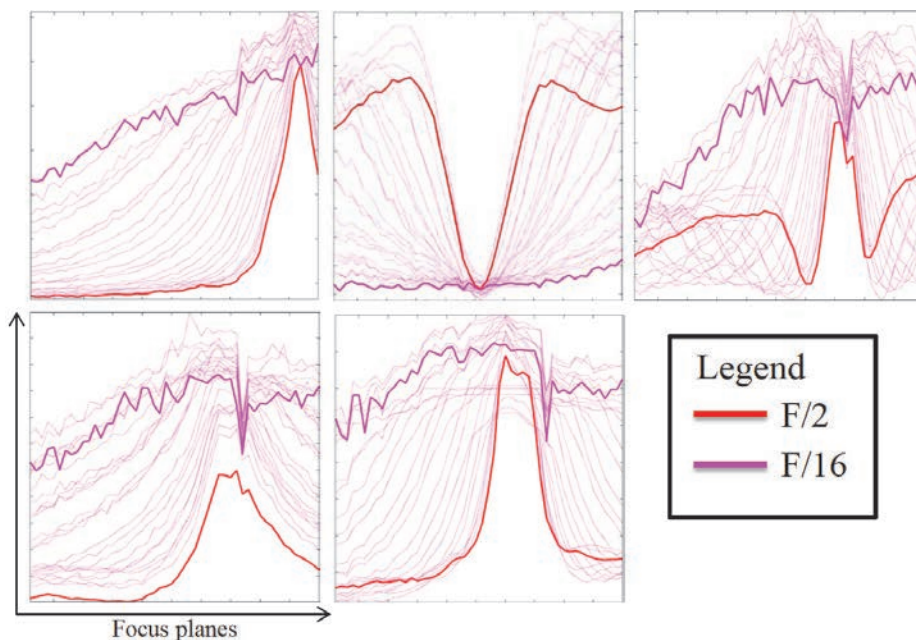


図 2.4 絞り値に応じた Focus スタックの Intensity Profile:

図 2.3 の "Animals" のシーンからいくつかの画素位置の Intensity Profile を大きな絞り値(F/2)の場合と小さな絞り値(F/16)の場合で比較したもの



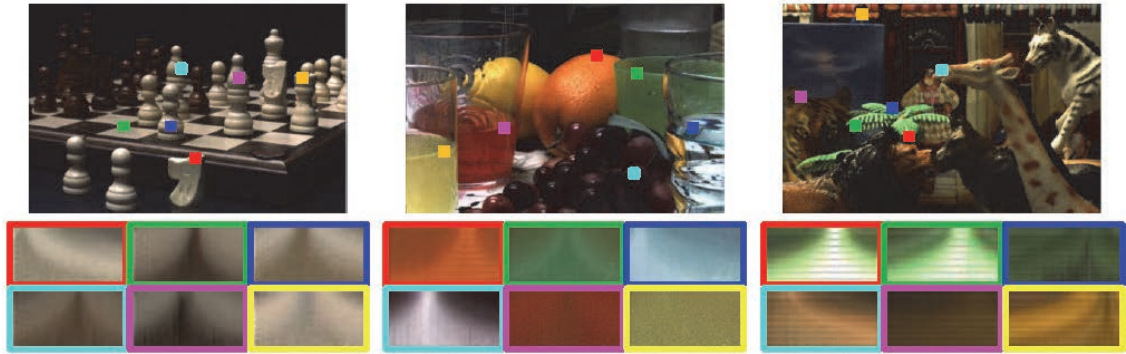
## 2.3 Aperture-Focus スタックの Intensity Profile とエネルギー圧縮

### 特性

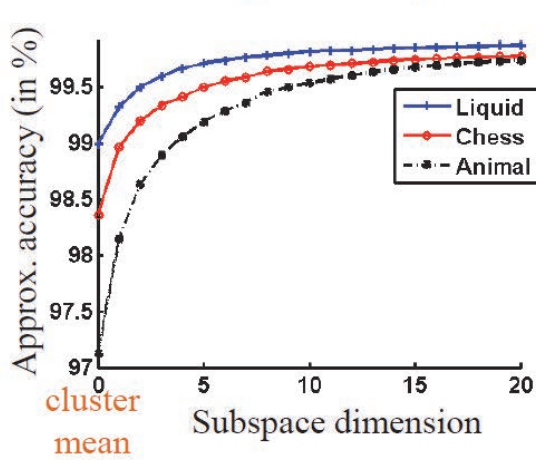
前節では、Focus スタックの Intensity Profile（画素値の集合）とそのエネルギー圧縮特性について述べた。ここでは、カメラの撮像パラメータとしてフォーカス位置に加えて、絞り値も変更して撮像した2次元の Epsilon Photography である、Aperture-Focus スタックについて、その Intensity Profile とエネルギー圧縮特性について述べる。

図 2.5(a)には、 $45 \times 18$  画素の Aperture-Focus Image (AFI)画像をいくつかの画素位置について示している。AFI とは前述の通り、Aperture-Focus スタックについて、フォーカス位置と絞り値を変更した際の各画素位置の Intensity Profile の表現方法であり、図 2.5(a)では、横軸にフォーカス位置 45 点、縦軸に絞り値 18 点について、当該画素位置の画素値を2次元画像として表現している。下記に、いくつかの AFI の特徴について記す。まず、当該画素にフォーカスが合っている位置の場合、絞り値を変更しても画素値は変化しない現象がある。この現象は、AFI においては、あるフォーカスの値で、縦軸上の値が変化しない位置が存在することで観測可能である。その位置は、当該画素のカメラからの距離によって決まる。次に、図 2.4 でも示した通り、最も小さい絞り値では、被写界深度は非常に深くなるため、フォーカス位置を変えても画素値はほとんど変わらない。この現象は、AFI においては、絞り値が小さいときに、横軸方向の輝度値が変わらないことに表れている。それぞれの AFI から一概に見て取れるのは、その形状が連続的であり、スムーズな変化であることである。緩やかな凸形状が AFI 上で現れる位置は、被写体のカメラからの距離のみにより決まる。Focus スタックと同様に、非常に疎な情報である。本論では、この冗長性をモデル化できるのではないかと考え、第3章に示す方法で実践する。

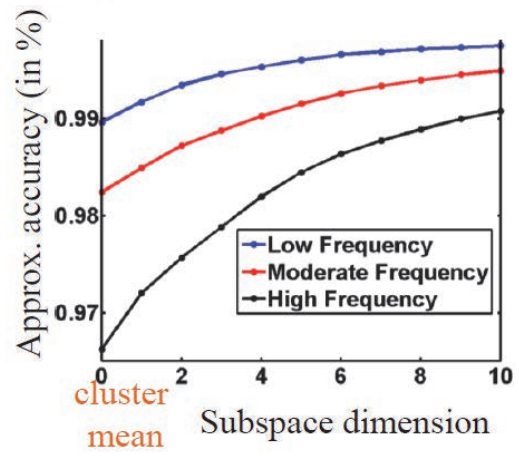
この Aperture-Focus スタックの Intensity Profile に対しても、Focus スタックと同様に、k-means クラスタリングを実行した。前途の通り、各画素位置における Intensity Profile は AFI として表現可能であるが、その平均と標準偏差を正規化して、クラスタリング処理を実施した。図 2.5(b)および(c)に、AFI のエネルギー圧縮特性を示している。エネルギー圧縮特性とは、前途の通り、各クラスを PCA によりモデル化した際の、各主成分に対する分散の寄与率を全クラスについて累積した割合である。450 のクラスに分類したとき、その中心のみを用いた場合、および、中心周辺の主成分を用いた場合である。図 2.5(b)から見て取れるように、450 のクラスタ中心のみを用いた場合において、既に全エネルギーの 97%以上となり、主成分 10 次元であれば 99%となることわかる。このエネルギー圧縮特性は、対象領域のテクスチャにほぼ依存しないことも判明した。これを示すため、最小の絞り値における勾配評価を行い、領域をテクスチャの周波数で区分し、エネルギー圧縮特性の評価を行った。図 2.5(c)に示す通り、450 のクラスタ中心



(a) Example focus aperture AFI images from 3 datasets



(b) Energy compaction across scenes



(c) Energy compaction across points of varying detail

図 2.5 Aperture-Focus スタックの冗長性:

(a) 3つのデータセット中のいくつかの画素位置における AFI (Aperture Focus Image) (b) AFI の k-means クラスタリング結果. 450 クラスのクラスタ中心のみを用いたプロットから 97% のエネルギー圧縮特性を確認. さらに 10 次元までの追加では 99% 以上であることを確認した. (c) エネルギー圧縮特性のプロットから AFI をテクスチャの周波数に分離して評価. 高周波テクスチャにおいても高い冗長性を確認した.

のみのモデルで, 低周波成分にて 99% のエネルギー圧縮特性を得たのみならず, 高周波成分においても, 96% のエネルギー圧縮特性を得た. 主成分 10 次元の場合, やはり 99% を得た. 総じて, AFI は, 非常に圧縮性の高いデータであると結論付けられる. これに

より、少ないサンプリング点から、AFI 中の全画素値が復元可能であること、を仮説として建てた。

## 2.4 MLAF スタックの Intensity Profile とエネルギー圧縮特性

次に、カメラの撮像パラメータとしてフォーカス位置、絞り値に加えて、露光レベルも変更して撮像した 3 次元の Epsilon Photography である、MLAF スタック (Multi-level Aperture-Focus スタック) について、その Intensity Profile とエネルギー圧縮特性について述べる。

図 2.6 では、MLAF スタック中のある画素位置から取得した MLAFI (Multi-level Aperture Focus Image) の特徴を示している。MLAFI とは、前述の通り、フォーカス位置、絞り値、露光レベルを変更した際の各画素位置の Intensity Profile について、露光レベル数分の 2 次元画像で表現したものであり、フォーカス位置と絞り値のパターン数である  $51 \times 22$  画素の、3 つの AFI 画像とも言い換えることができる。上段において 3 つのシーンの画像例を示し、中段では、MLAFI を上段シーンのいくつかの画素位置について示している。横軸にフォーカス位置 51 点、縦軸に絞り値 22 点、について 3 つの露光レベルで、当該位置の画素値を 2 次元画像 3 枚にて表現している。下記に、MLAFI の特性を記す。まず、3 つの露光レベルにおける AFI の特徴は、Aperture-Focus スタックにおいて説明した内容と同様である。当該画素にフォーカスが合っている位置の場合、絞り値を変更しても画素値は変化しない現象が見て取れる。あるフォーカス位置で、縦軸上の値が変化しない位置は、被写体のカメラからの距離によって決まる。一方、最も小さい絞り値では、被写界深度が深くなるため、同一の露光レベルではフォーカス位置を変えても画素値はほとんど変わらない (横軸方向の画素値はほとんど変わらない)。一概に、それぞれの AFI は、その形状が連続的であり、変化はスムーズで、緩やかな凸形状である。一方で、各露光レベルの AFI 間の関係は、白飛びや黒つぶれしている撮像パラメータの場合の画素値を除き、明るさが露光レベルに対し、線形に増減するのみである。これをまとめた MLAFI もやはり非常に疎な情報であると言え、この冗長性をモデル化できるのではないかと考えた。

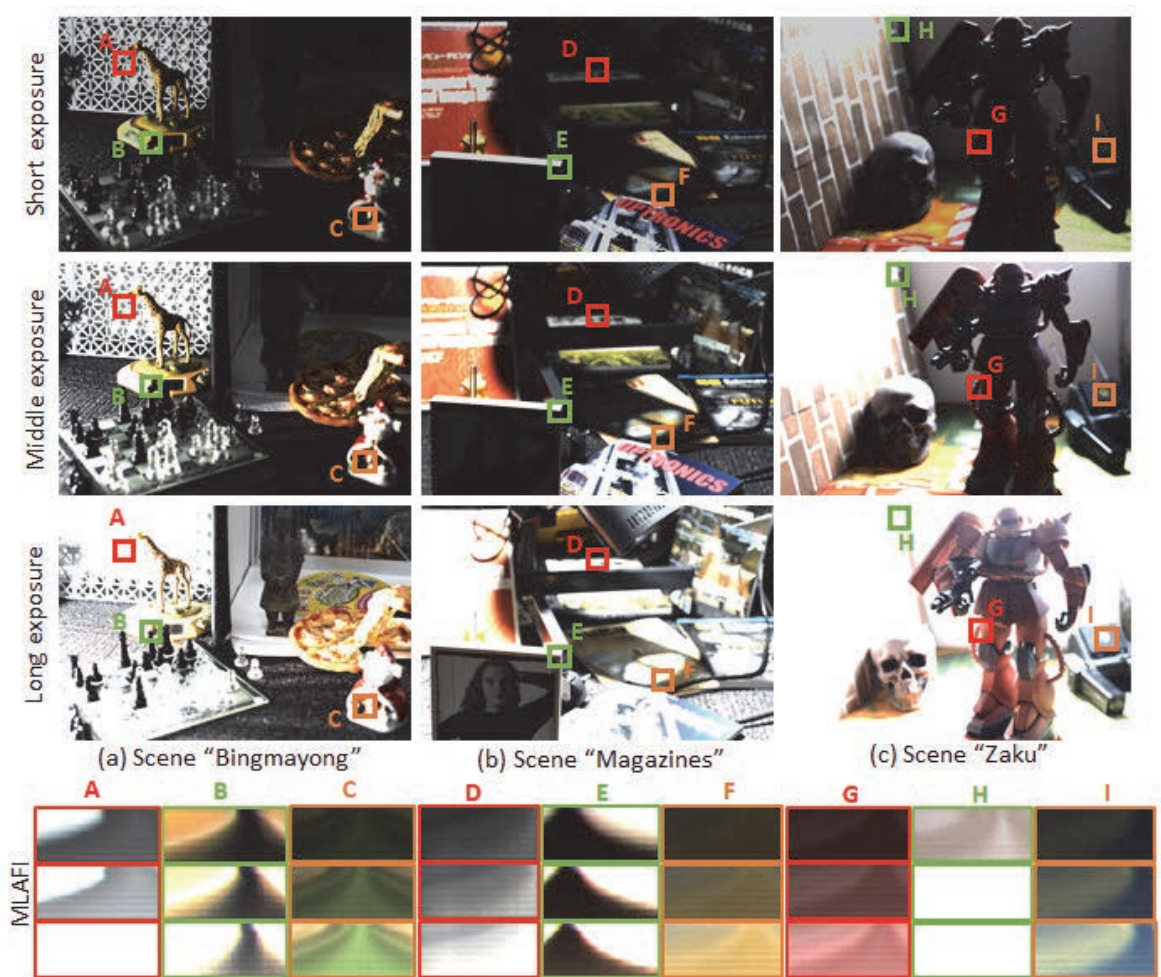


図 2.6 MLAF スタックの冗長性:

(上) 3つの画像スタック中のある特定のフォーカス位置・絞り値における3つの露光レベルの MLAFI 画像例を示している。(下) 各シーン内のいくつかの画素位置における MLAFI 画像を示している。白飛びや黒つぶれの領域を除けば、Aperture-Focus スタックにおける AFI と同様の傾向（形状が連続的で、緩やかな凸形状を有する）をそれぞれ持ち、明るさレベルが段階的になっている。

フォーカス位置，絞り値，露光レベルの全パラメータについて，撮影した MLAF 画像スタックを用意し，その各画素の Intensity Profile から構成される MLAFI の k-means クラスタリングを実行した．画素位置ごとの Intensity Profile は，やはり，平均と標準偏差を正規化して処理した．図 2.7 に，クラスタリング結果のエネルギー圧縮特性を示している．450 のクラスに分類したとき，その中心周辺の主成分を用いた場合である．450 のクラスタ中心のみを用いた場合において，既に全エネルギーの 97-98% 以上となる．このエネルギー圧縮特性は，白飛びを有する画素領域においても依存は極めて小さいことも判明した．図 2.7 右では，これを示すために，MLAFI 中に飽和輝度領域が存在する白飛びを有する画素と，有しない画素に分割し，エネルギー圧縮特性の評価を行った．白飛びがない場合，MLAFI は約 99%，主成分 10 次元までの場合は，100% 近いエネルギー圧縮特性を得た．一方で，白飛びがある画素においてもエネルギー圧縮特性の低減は 97% 程度までに留まり，その分散も 1% 未満であった．総じて，MLAFI は，非常に圧縮性の高いデータであると結論付けられる．これが，少ないサンプリング点から，MLAFI の全信号が復元可能であると見立てたモチベーションである．

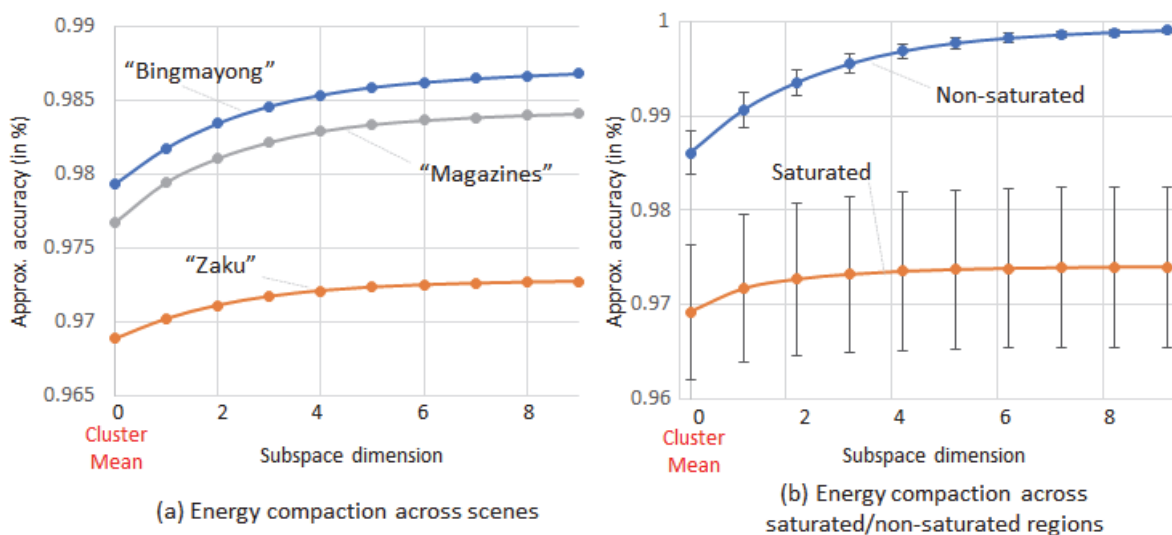


図 2.7 MLAF スタックのエネルギー圧縮特性:

MLAFI 画像の各画素を k-means クラスタリングした結果を中心周辺の主成分を用いた場合について示す．左では図 2.6 の 3 シーンについて各々評価し，右では白飛びを含む画素と含まない画素を区別して評価している．

# 第3章 圧縮 Epsilon

## Photography 手法

本論の目的は、カメラパラメータを様々に変更して撮影したときの各輝度値を、画素ごとに再構成することである。これは、本手法が実際に撮影した画像と同等の解像度での再構成可能であることを意味する。これを実現するため、画像スタック各画素の Intensity Profile である、 $I_{u,v}(f, \alpha, l)$  をモデル化し、再構成のために学習することが必要である。第2章における解析より、Epsilon Photography の各画素位置に対応する画素値の変化は、明らかにクラスタ中心に密に集まっており、混合ガウシアンモデル(GMM)で表現可能と考える。

### 3.1 提案手法の処理の流れ

図 3.1 に提案手法の処理フローを示している。まず、本手法を実施するためには、3.2 章でその詳細を示す、入力画像の幾何・輝度キャリブレーションが必要である。この図 3.1 上段に示す処理は、撮像系につき1度実施すればよい。ここで撮像系とは、カメラとレンズのセットを指す。幾何・輝度キャリブレーションの結果、実際の GMM 学習処理およびそれを用いた再構成処理において、各画像の撮像時に用いられた撮像パラメータに対して、適切なキャリブレーションパラメータが与えられる。学習・再構成処理の入力画像に対して、そのキャリブレーションパラメータを用いたキャリブレーションが実施される。

図 3.1 中段に示したのは、GMM パラメータの学習処理の流れである。ここでは、完全な画像スタック（再構成の対象となる、Focus スタック、Aperture-Focus スタック、または MLAF スタック）の撮影と入力が必要となる。入力された完全な画像スタックは、キャリブレーション処理され、GMM 学習のブロックへと入力される。ここでは、3.3 章でその詳細を説明する、GMM パラメータの学習が行われる。得られた GMM パラメータは、再構成処理時に用いられる。ここで、GMM パラメータの学習は、その学習用データセットにつき一意に決められる。言い換えれば、1つの学習データの生成につき、一度実施すればよい。さらに、GMM のパラメータが学習で決定されると、再構成時などの撮影パラメータを用いれば、最も再現性高く画像スタックが再構成されるか、その最適撮像パラメータを決定することができる。具体的な方法は 3.5 章で述べるが、本論

ではこの最適パラメータの決定を、グリーディ・アルゴリズムを用いて実施する。GMMパラメータの学習と同様に、この学習された GMM パラメータに対して最適な再構成用の撮像パラメータの決定も、用意された学習データセットにつき 1 度、学習データの生成につき 1 度実施すればよい。

図 3.1 下段に示す処理は、実際の画像スタックの再構成処理の流れである。入力は、最適撮像パラメータを用いて限定枚数分、撮像された画像データである。この入力画像データは、カメラの連射によって撮像される想定である。入力画像データは、それぞれの画像の撮像パラメータに応じたキャリブレーションパラメータでキャリブレーションされ、再構成処理ブロックに入力される。この GMM を用いた再構成処理については、3.4 章でその詳細を説明する。学習した GMM を用いて、平均二乗誤差の最小化による推定(MMSE)による近似にて、完全な画像スタックの再構成を実施する。再構成された画像スタックの内、MLAF スタックの場合には、同じフォーカス位置および絞り値をもつ画像間の HDR 合成により、HDR の Aperture-Focus スタックの合成が可能であり、ここで実施する HDR 合成処理については、3.6 章で説明する。また、画像スタックの再構成により実現される様々なアプリケーションについては、第 5 章で詳しく説明する。

ここで、1 点追記しておきたいこととして、本論では、完全な画像スタックの再構成を目的としているが、任意の撮像パラメータで撮像された相当の画像を単独で再構成することも可能であり、アプリケーションによってはそうした処理を実装できる。

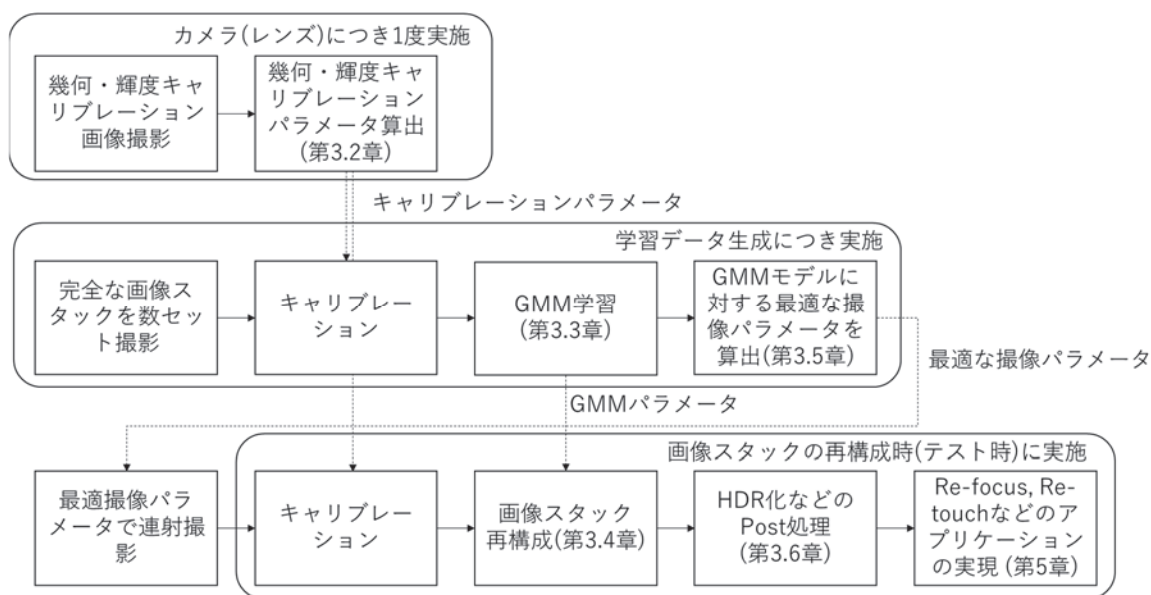


図 3.1 提案手法の流れ

## 3.2 幾何・輝度キャリブレーション

画素位置ごとの Intensity Profile をモデル化し、モデルを学習して再構成するためには、精緻な幾何および輝度キャリブレーションが必要となる。まず、幾何キャリブレーションについては、[11]が提案する手法を採用する。一般に、撮像機器のフォーカス位置を変更すると、画像中の被写体は非線形にワーピングする。ここでこのワーピングをモデル化する。[11]では、基準位置 $f_0$ からの任意のフォーカス位置 $f$ への変更によるワーピング $W_d$ は、像倍率 $m_f$ 、レンズ歪パラメータ( $k_0, k_1, k_2$ )、および、画像中心( $x_c, y_c$ )からの位置ずれ距離 $r$ 、によって正確にモデル化できることを述べている。

$$W_d(x, y) = [m_f + m_f(f - f_0)(k_0 + k_1 r + k_2 r^2)] \cdot [(x, y) - (x_c, y_c)] \quad \cdots(1)$$

これらのパラメータ推定には、黒ドットがグリッド状に並んだパターンのチャートを用いる。図 3.2 に撮像したパターンチャートの画像例を示す。以下、手順を説明する。まず、チャートをカメラの全面に配置し、最大絞り値(F/1.4)時の全フォーカス位置での画像を撮像する (図 3.2(a))。次に、撮影した画像の重心座標を画像処理で抽出する (図 3.2(b))。各黒ドットの座標を $(x, y)$ へと代入し、実測と誤差が最小となるよう(1)式の各パラメータを最適化する。推定されたパラメータを用いて、入力データセットの画像間レジストレーションを実現する。以上が幾何キャリブレーションの処理の流れである。図 3.3 では、この位置合わせの効果を、例を上げて示している。図 3.3(a)の入力画像に **Warping** 処理をし、位置合わせ後の結果画像が図 3.3 (b)である。図 3.3 (c)はその効果を示すための遠いフォーカス位置の画像であり、図 3.3 (b)と(c)の位置合わせの実現が目的である。画像中の4領域におけるズームアップが図 3.3 (d)であり、図 3.3 (a)から(b)への変換により、画素位置が整合される様子が見て取れる。当然のことながら画面中央より画面端の方がワーピング量は大きくなるが、図 3.3 の例では、概ね正確な位置へのレジストレーションが行われていることが見て取れる。図 3.3 (e)では各領域のワーピング方向と画像中心を図示しているが、これにより、フォーカス位置が近の入力画像は、遠の画像に対して、拡大されていたことがわかる。

一方、輝度キャリブレーションには、カメラ応答関数を考慮した線形化と、絞り値の大きさの変更による口径食の問題を考慮する必要がある。本論で用いる実験画像データセットはすべて Canon EOS-40D のカメラを用いて用意するため、入力輝度の線形化に関しては、Adobe 社の DNG SDK[55]のソフトウェアを利用し、Raw データにおけるカメラ応答値の線形化した。また、口径食による周辺の輝度低下を補正については、光源明るさが均一な白色画像を、絞り値を変えて撮影し、画像中の輝度の不均一性を2次関数でフィッティングして正規化を実施した。



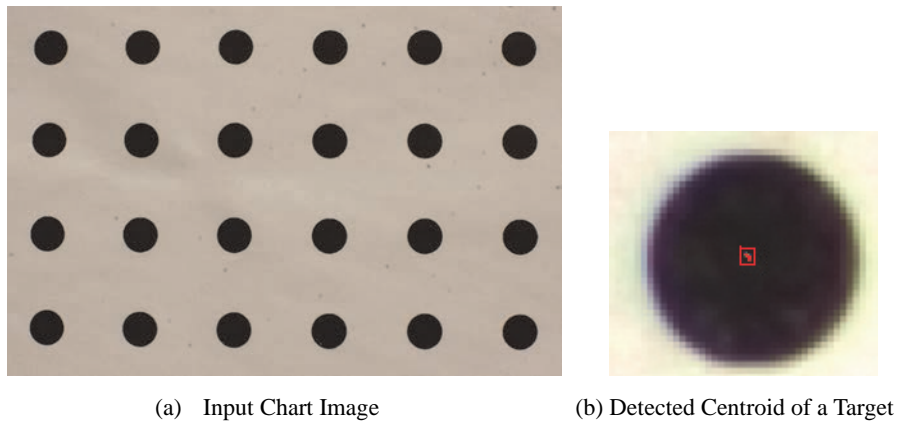


図 3.2 幾何キャリブレーションで用いたチャート:

(a)入力画像の例, (b)画像処理で抽出された黒ドットの重心位置の例. 実測値を用いて, 基準位置 $f_0$ からの任意のフォーカス位置 $f$ への変更によるワーピングモデルのパラメータを最適化する.

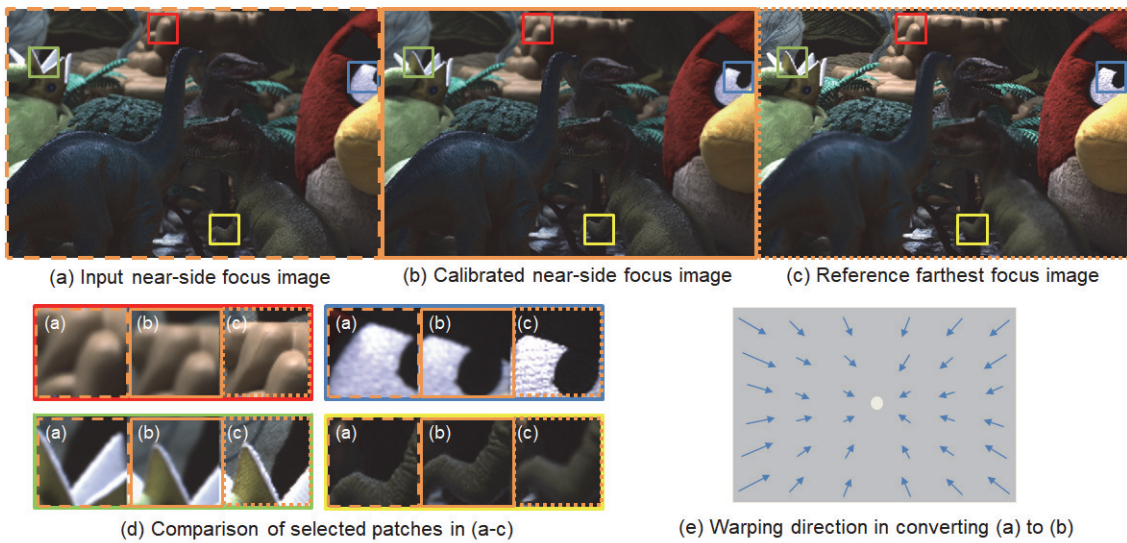


図 3.3 幾何キャリブレーションによる位置合わせ:

(a)入力画像(フォーカス位置:近), (b)キャリブレーション後の画像(フォーカス位置:近), (c)フォーカ位置:遠の画像(処理なし). (d)いくつかの領域についてズームアップした画像. キャリブレーション後の画像は, 被写体の距離に寄らず, フォーカス位置が遠い画像と位置が合っていることがわかる. (e)各領域のワーピング方向と画像中心.

### 3.3 GMM パラメータの学習

本論では、画素位置ごとの Intensity Profile をモデル化するために混合ガウシアンモデル (Gaussian Mixture Model, 以下 GMM と記載) を用いる。表現を変えれば、Aperture-Focus スタックの各画素から抽出された AFI について (または MLAF スタックの各画素から抽出された MLAFI について)、GMM でモデル化し、そのパラメータを学習する。これを進めるため、11 の Aperture-Focus スタックを撮像して用意した。全て複雑さの異なるシーンであり、後の実験に向け、6 つのデータセットで GMM パラメータの学習を行った (残りの 5 データセットはテスト用)。なお、Focus スタックにおける後の実験は、Aperture-Focus スタックのシーンと同じシーンで、固定絞り値の画像群を使用する。また、露光レベルを変えた MLAF スタックについては、別途 13 のフォーカス・絞り・露光レベルの全パラメータで撮影した画像スタックを用意した。こちらも全て複雑さの異なるシーンであり、7 セットで GMM パラメータの学習を行なった (残りの 6 セットはテスト用)。

学習用画像スタックから各画素位置の Intensity Profile を抽出し、これをそれぞれ正規化して、クラスタリング処理し、各 Intensity Profile が所属するクラスタおよび各クラスタの平均・共分散行列の計算を行った。この演算が最終的には、クラスタ  $k$  における全ての GMM パラメータである、クラスタ平均  $u^{(k)}$ 、共分散行列  $C^{(k)}$ 、を求めること同義となる。さらに、各クラスに与える初期重み  $p_k$  は、全データに対する各クラスタに所属するデータ数の割合から算出することができる。まとめると、このときモデル式はガウス関数  $N$  を用いて以下に表される。

$$f(x) = \sum_{k=1}^K p_k N(x; u^{(k)}, C^{(k)}) \quad \dots (2)$$

図 3.4 は、この処理例 (Aperture-Focus スタックの各画素から抽出された Intensity Profile (この場合は AFI と同義) から GMM パラメータを学習する場合) の概念図である。

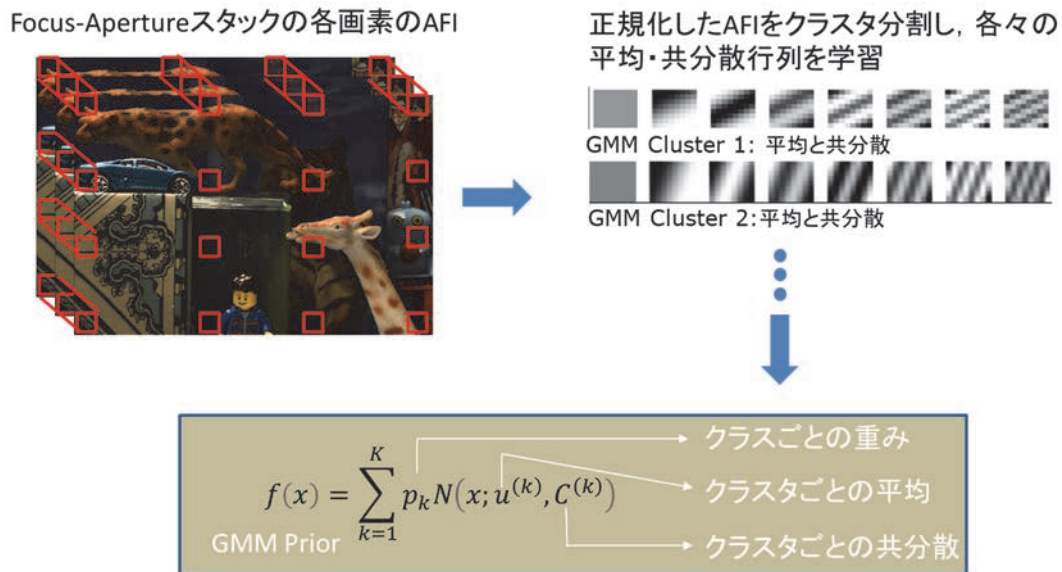


図 3.4 GMM モデル化と GMM パラメータの学習:

入力である Aperture-Focus スタックの各画素位置の Intensity Profile(この場合は AFI と同義)を正規化して, クラスタリング処理し, 各 AFI が所属するクラスタおよび各クラスタの平均・共分散行列の学習を行う. 最終的に得られた(2)式内の GMM のパラメータが取得できる.

### 3.4 学習済み GMM パラメータを用いたデータの再構成

画像再構成の問題は, 画素ごとには, サブサンプリングされた観測データから AFI や MLAFI を推定することに置き換えられる. サブサンプル観測データと完全な AFI または MLAFI の関係性は, 線形であり  $y = Hx + n$  の式で表現することが可能である. ここで,  $y$  はベクトル表現されたサブサンプル観測データ,  $x$  は完全な AFI(または MLAFI),  $H$  はサンプリングを示す変換行列であり,  $n$  は観測ノイズである.  $H$  は, 各行にひとつ 1 の値があり, それ以外は 0 の値となる行列式となり, 1 の値が実際に撮影された画像のパラメータ (サブサンプリングされた観測データ) に相当する. 図 3.5 は, 図 3.4 と同様に Aperture-Focus スタックの場合を例にとり, 各画素位置におけるサブサンプル観測データと完全な AFI の関係を概念図化したものである.

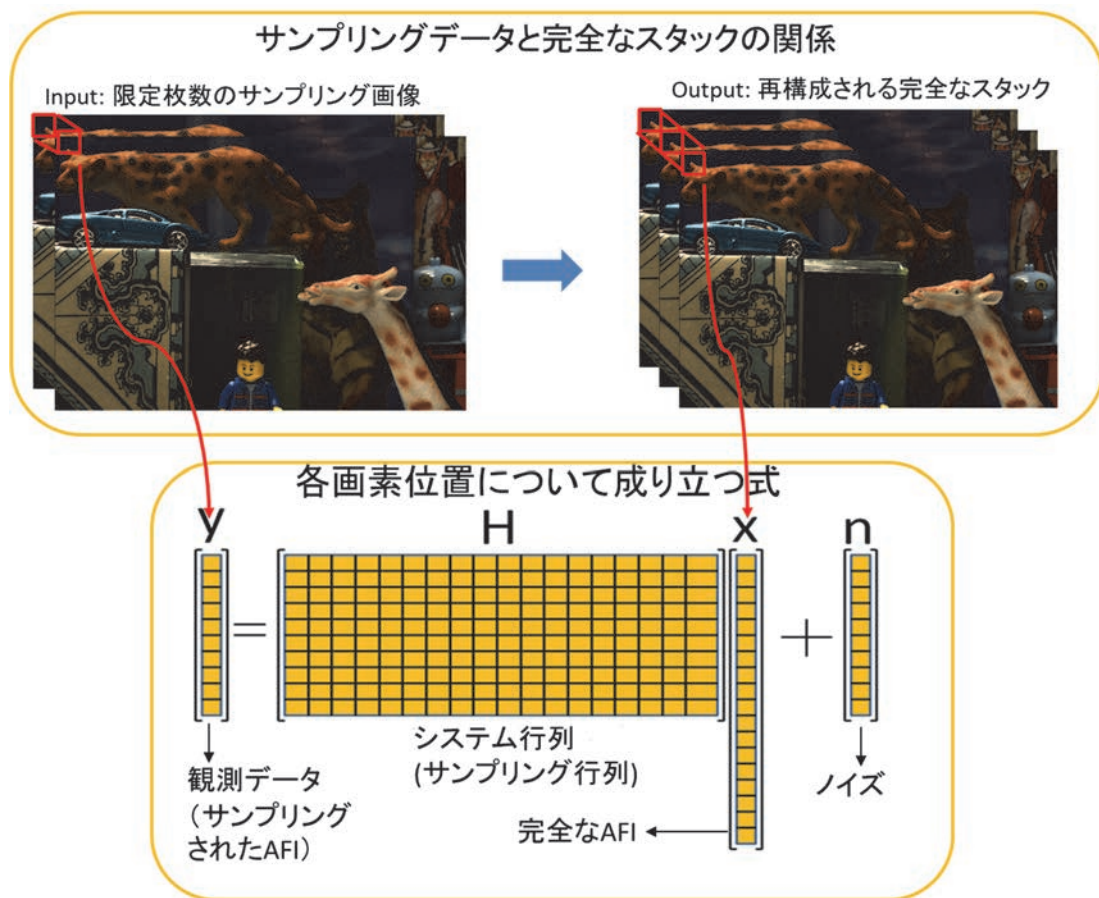


図 3.5 AFI 再構成のシステム式:

各画素位置におけるサブサンプル観測データと完全な AFI の関係を概念図化したもの。サブサンプル観測データ  $y$ ，サンプリングを示す変換行列  $H$  が既知のとき，完全な AFI である  $x$  を推定する問題と設定できる。

ノイズは，平均  $0$ ，共分散  $C_n$  のガウス分布  $N(0, C_n)$  であると仮定する。  $x$  の再構成は，平均二乗誤差の最小化による推定(MMSE)を用いて行う。この再構成の流れと原理について概念図 3.6 を示す。

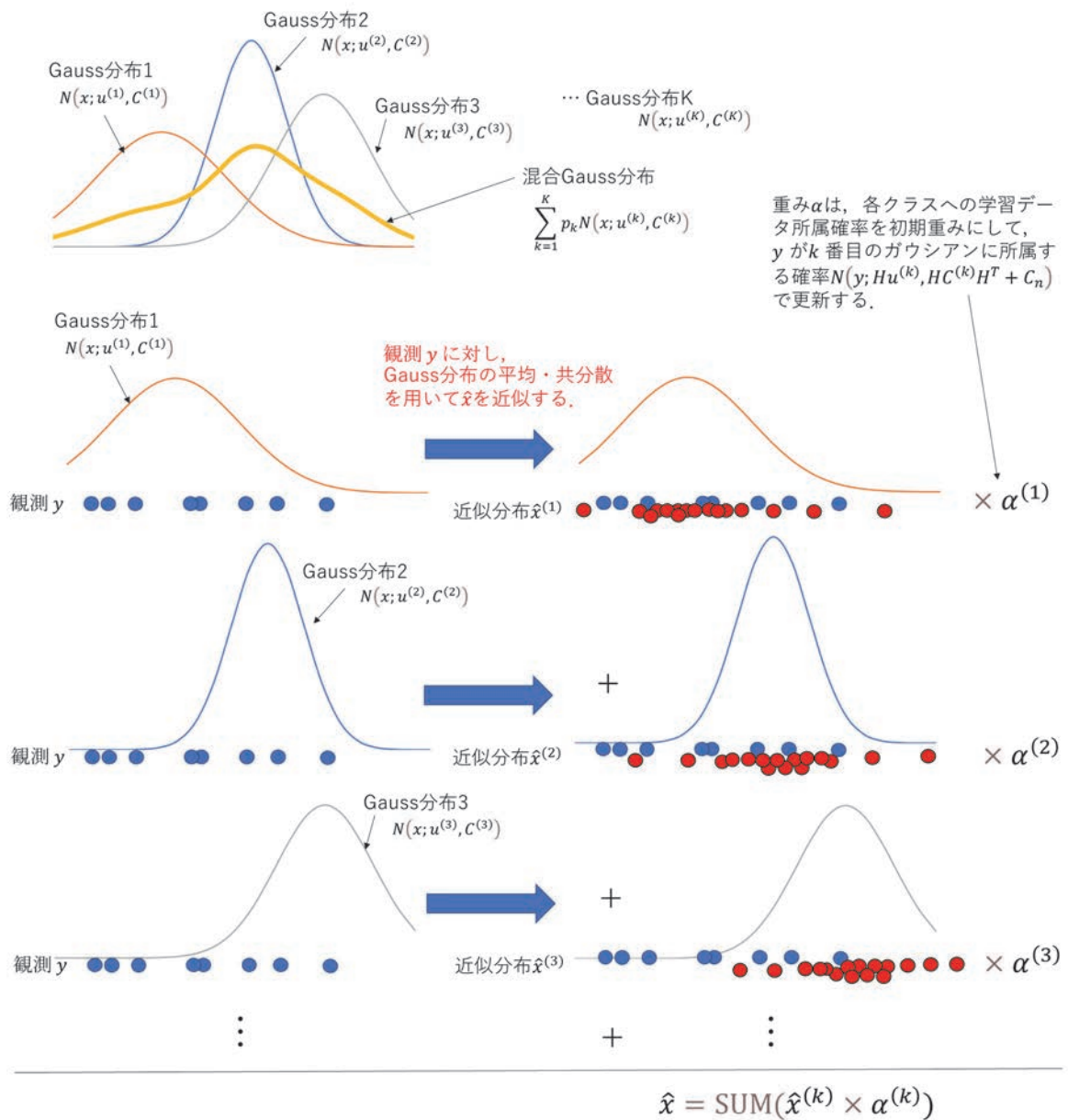


図 3.6 平均二乗誤差の最小化による推定(MMSE)を用いた元信号 $x$ の再構成:  
 上段で GMM を形成する各 Gauss 分布を示している. それぞれの Gauss 分布の平均と共分散を用いて, 観測  $y$  が与えられた際, 最小自乗法をベースとした近似により  $x$  を求める. それぞれのガウシアンを重み付け加算により, 最終的な  $x$  を決定する.

以下、式を用いて実際の計算を説明する。

観測  $\mathbf{y}$  が与えられたとき、事後分布  $f(x|\mathbf{y})$  の  $x$  について詳細には、

$$f(x|\mathbf{y}) = \sum_{k=1}^K \alpha^{(k)}(\mathbf{y}) * f^{(k)}(x|\mathbf{y})$$

と表される。

ここで重みは  $\alpha^{(k)}(\mathbf{y})$ 、 $f^{(k)}(x|\mathbf{y})$  は事後分布の  $k$  番目のガウシアンであり、

$$f^{(k)}(x|\mathbf{y}) = N(x; \mathbf{u}^{(k)}, \mathbf{C}^{(k)})$$

である。平均  $\mathbf{u}_{x|\mathbf{y}}^{(k)}(\mathbf{y})$  と、共分散  $\mathbf{C}_{x|\mathbf{y}}^{(k)}$  は、

$$\mathbf{u}_{x|\mathbf{y}}^{(k)}(\mathbf{y}) = \mathbf{u}^{(k)} + \mathbf{C}^{(k)} \mathbf{H}^T (\mathbf{H} \mathbf{C}^{(k)} \mathbf{H}^T + \mathbf{C}_n)^{-1} (\mathbf{y} - \mathbf{H} \mathbf{u}^{(k)})$$

$$\mathbf{C}_{x|\mathbf{y}}^{(k)} = \mathbf{C}^{(k)} - \mathbf{C}^{(k)} \mathbf{H}^T (\mathbf{H} \mathbf{C}^{(k)} \mathbf{H}^T + \mathbf{C}_n)^{-1} \mathbf{H} \mathbf{C}^{(k)}$$

で与えられる。また、重み  $\alpha^{(k)}(\mathbf{y})$  は、

$$\alpha^{(k)}(\mathbf{y}) = \frac{p_k f^{(k)}(\mathbf{y})}{\sum_{i=1}^K p_i f^{(i)}(\mathbf{y})}$$

であり、初期重み  $p_k$  が  $f^{(k)}(\mathbf{y})$  で更新されていることを示している。この  $f^{(k)}(\mathbf{y})$  は、 $\mathbf{y}$  が  $k$  番目のガウシアンに所属する確率であり、

$$f^{(k)}(\mathbf{y}) = N(\mathbf{y}; \mathbf{H} \mathbf{u}^{(k)}, \mathbf{H} \mathbf{C}^{(k)} \mathbf{H}^T + \mathbf{C}_n)$$

と表す。ここで、MMSE により推定される完全な画像スタックは、事後分布  $f(x|\mathbf{y})$  の平均であり、

$$\hat{\mathbf{x}}(\mathbf{y}) = \sum_{k=1}^K \alpha^{(k)}(\mathbf{y}) \mathbf{u}_{x|\mathbf{y}}^{(k)}(\mathbf{y})$$

これらを用いて最適な MMSE は下記から推定する。

$$\text{MMSE}(H) = E \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2 \quad \cdots(3)$$

本論で画像再構成アルゴリズムに GMM を選択した理由を以下の表で示す。類似研究が少なく、他の手法との比較が難しいため、定性的な説明を記す。

表 3.1 再構成手法の定性的な比較

	非線形補間 Bi-Cubic補間[60]など	スパース最適化 SPGL1[61],OMP[62]など	GMM	Deep Learning [66], [67]など
期待復元精度	×	○	◎	◎
復元処理速度	◎	×	○	◎
学習データ 低容量性	◎	◎	○	×

非線形補間は、処理が高速で非学習方式のため学習の必要はない。しかしながら、4.1章で1次元の画像スタックである、Focus スタックの復元実験において示されているように、復元精度としては大きな期待を持つことはできない。

[56, 57]では、復元の対象となる信号としては本論とは異なり、符号化露光からの動画の再構成[58, 59]を目的としたものであるが、スパース最適化 (SPGL1[61]およびOMP[62]) 手法, GMM, および, Deep Learning で画像復元を行った場合の比較評価を実施している。その中で GMM は、スパース性を用いた復元手法と比べた場合に、より高い精度が得られ、かつ、処理量が小さいことが示されている。画像の復元問題におけるスパース最適化は、学習により得られた基底と係数で表現し、係数ベクトル (行列) のスパース性を用いて最適化問題を解く方式が一般的であるが、最適化の過程で反復演算が必要となるため、高速演算に向かない。

一方、ディープラーニングを用いた手法との比較においては、ディープラーニングでの復元は高速であり、[56]においても GMM と同等レベルの復元精度を示しているが、一般に学習時に大量のデータを必要とする点において、GMM の優位性が言える。本論が復元する信号には、3つの撮像パラメータ (フォーカス位置, 絞りサイズ, 露光レベル) の 3,366 種類を各々に変えた包括的な画像データセットを含む。収集に手間がかかるのみならず、本手法は7つのデータセットを GMM の学習用に用意しているが、一般的なディープラーニングモデルの学習に対して十分かは疑問が残る。GMM での表現はエネルギー圧縮特性上、復元性能が保証されている点で、本信号への高い適用性が言える。一方で、近年のネットワークモデルや学習方法の工夫, コンピュータグラフィックスなどを用いた学習データの生成といったディープラーニング周辺技術の進化を用いることで、この課題を解決する可能性は大いにあると考える。

最後に、GMM の他にも主成分分析を直接用いた方式[63]など、学習後に反復計算なく基底をもとに信号を復元する手段も適用可能であることを追記する。本論が、この比較で GMM を選択したのは、モデルとしての表現力をベースとした期待復元精度が高いためである。本論で GMM を用いて証明される概念をもとに、別の再構成手段を試みた後続の検討が生まれることを期待する。

### 3.5 サンプリング画像の選択

本論における再構成の目標は、できるだけ少ない、フォーカス位置、絞り値と露光レベルのパターンの画像から、完全なスタックを高品質に再構成することであり、サンプリングする画像パラメータの選択は重要な点である。本論では、(3)式 MMSE を最小化することを目的とした、グリーディ・アルゴリズムを採用する[64]。まず最小化する MMSE を、

$$\text{MMSE}(H) = \sum_{k=1}^K p_k \text{Tr} \left( C_{x|y}^{(k)} \right) \quad \dots(4)$$

と記述する。ここで、 $C_{x|y}^{(k)}$  は、最適化後の各クラスタのガウシアン分布の共分散行列である。最適なカメラパラメータの探索は、入力画像枚数分のカメラパラメータの組合せを見つける問題と同義である。パラメータ空間において、フォーカス位置の次元  $N_f$ 、絞り値の次元  $N_\alpha$  (Focus スタックの場合は、 $N_\alpha = 1$ )、露光レベルの次元  $N_l$  (Focus スタックおよび Aperture-Focus スタックの場合は、 $N_l = 1$ )、を掛け合わせた次元数  $N_{all} = N_f \times N_\alpha \times N_l$  の全探索により(4)式の最適化する  $m$  個の組合せ探索を行う場合、 $\binom{N_{all}}{m}$  回の探索評価が必要であり、現実的ではない。

一方、本論の採用するグリーディ・アルゴリズムによる探索はより少ない回数での評価により、(4)式の MMSE の最適化を実現する。まず、最適な2つのパラメータ組合せを探索する。これはすなわち  $m = 2$  の問題であり、探索評価回数は  $\binom{N_{all}}{2}$  である。パラメータペアが得られると、事後の共分散である  $C_{x|y}$  を、現在までに更新されたパラメータを考慮して更新する。カメラパラメータの選択は、サンプリング行列  $H$  の各行  $h_i$  に対応する。 $i$  番目の反復後、事後のガウシアン共分散は、

$$C_{x|y,i}^{(k)} = C_{x|y,i-1}^{(k)} - C_{x|y,i-1}^{(k)} \hat{h}_i^T \left( \hat{h}_i C_{x|y,i-1}^{(k)} \hat{h}_i^T + C_n \right)^{-1} \hat{h}_i C_{x|y,i-1}^{(k)}$$

となる。ここで、初期の共分散  $C_{x|y,0}^{(k)}$  は、GMM の共分散  $C_x^{(k)}$  に対応する。このように、事後の共分散を更新することで、最適な  $m$  個のサンプリングパラメータの組合せを最適化するが、2回目以降の反復において、探索評価回数は  $N_{all}$  回である。本手法による最終的な探索回数は、 $\binom{N_{all}}{2} + (m - 2) \times N_{all}$  回となり、全探索の場合の  $\binom{N_{all}}{m}$  回と比較して、大幅な評価回数の削減を実現できる。例えば、本論で用いる Aperture-Focus スタックはフォーカス位置 45 種類・絞り値 18 種類のデータであり、16 個のパラメータの組み合わせを得ようとする、全探索の場合には、約  $1.41 \times 10^{33}$  回の探索が必要となるが、グリーディ・アルゴリズムを用いる場合は、338,985 回の探索で完結し、その差は歴然である。



このグリーディ・アルゴリズムを用いた手法では、最終的なサンプリングの数  $m$  について、 $m > 2$  となる場合には、(4)式を最小化していくパラメータ順に優先度がつけられる。例えば、 $m=8$  の場合、パラメータの探索は 8 個の組み合わせが選ばれた段階で打ち止めればよく、それ以上の  $m$  の場合には、9 個目から引き続き(4)式を最小化するパラメータを 1 つずつ選択追加していくことができる。また、このパラメータの組み合わせは、学習データセットにつき決定される。言い換えれば、学習データから GMM モデルが完成した段階で、最適なサンプリングパラメータの組み合わせは一意に決まり、再構成するデータには依存しない。

図 3.7 では、サンプリング画像の選択手段について、Aperture-Focus スタックの再構成品質を PSNR で比較したものである。この実験は、学習データとして入力した AFI を、評価用にそのまま用いたものである。(a)提案手法であるグリーディ・アルゴリズム、(b)フォーカス位置を均等に変えた(絞り値一定の)サンプリング、(c)絞り値を均等に変えた(フォーカス位置一定の)サンプリング、(d)ランダムでフォーカス位置・絞り値のペアを選択したサンプリング、(e)等間隔でフォーカス位置・絞り値のペアを選択したサンプリング、の 5 つを比較している。図 3.5 下では、それぞれの方法でサンプリングした撮像パラメータが直観的に分かるよう、16 枚の画像をサンプルした場合について図示している。

再構成性能として、グリーディ・アルゴリズムを用いたサンプリング手法は、PSNR 評価で 10dB 程度の改善が、他のサンプリング手法に対して確認され、提案手法の最適サンプリングの有効性が確認できた。他の手法間の再構成性能の比較では、一定にフォーカス位置または絞り値を固定するよりも、2 次的に撮像パラメータを選択するランダムサンプリングや、等間隔 (uniform) サンプリングの方が、高い復元結果を得られることは、言えそうである。しかしながら、その差は微小であり、結論を得るのが難しい。一方でこれらと比較した場合のグリーディ・サンプリングの性能が明確に示されたと言える。

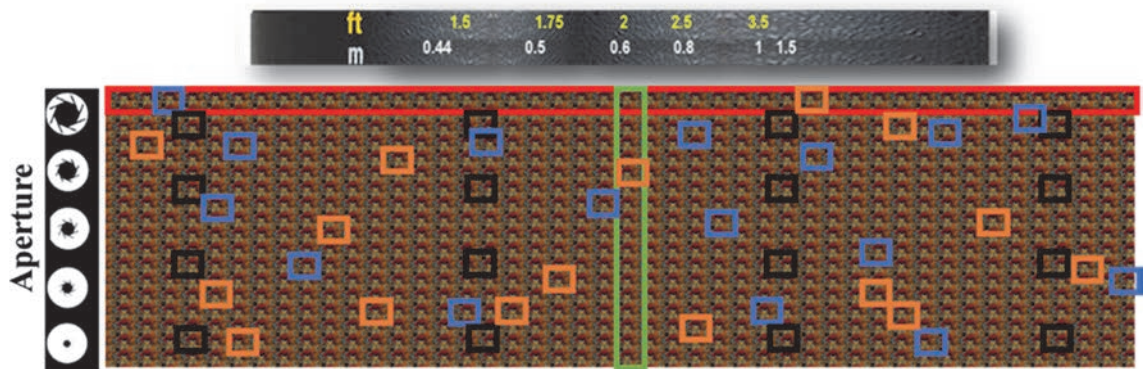
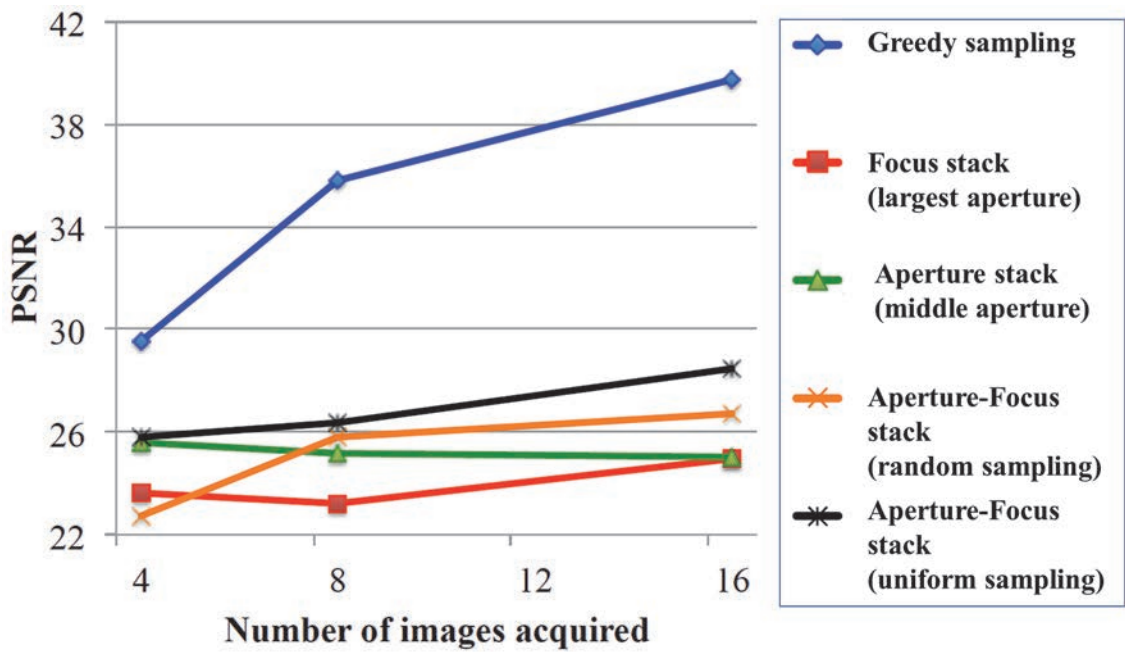


図 3.7 グリーディ・アルゴリズムによる画像選択の有効性:  
 サンプルング枚数と決定方法の PSNR 比較から、提案手法が一般的な方法に比べ、10dB 程度の改善があることを確認した。下の図では、比較したサンプルング手法について、選択されたパラメータを示しており、16 枚の画像をサンプルングした場合について、上図のグラフの色と対応させてパラメータを図示している。

### 3.6 HDR 画像合成

MLAF 画像スタックの再構成を行う場合、目的のデータは複数露光レベルの Aperture-Focus スタックとみなせるため、それぞれのフォーカス位置・絞り値において複数露光レベルの画像を用いた高ダイナミックレンジ (HDR) 画像の合成が可能となる。ここでは、本論で用いた HDR 画像合成の手法について説明する。

3.2 章においてすでに入力画像の輝度の線形化は行われており、露光レベルも既知である。よって露光レベルから得られる明るさの比率を用いて、簡易に中間の輝度値を算出することが可能である[10]。実際には Matlab の `makehdr` 関数[65]を用いてこの処理を行った (図 3.7)。第 4 章で提示される一連の結果画像において、合成された HDR 画像を表示する際には、`tonemap` 処理[66]を行って表示可能な画像に変換している。その場合の `tonemap` 処理について、特に明記がない場合には Matlab の `tonemap` 関数[67]を default 値として利用している (以降、標準 `tonemap` 処理と呼ぶ)。図 3.8 はこの標準 `tonemap` 処理の例を示している。標準 `tonemap` 処理では、明暗値のクリッピングとガンマ補正処理を実施しているが、領域ごと (default は画像を  $4 \times 4$  分割) でのヒストグラムの均等化をベースとしたアルゴリズムを用いて、その処理パラメータは自動で設定されており、暗所部も明るく表現されている。標準 `tonemap` 処理以外を適用している場合には、`tonemap` 処理のパラメータである、明暗値のクリッピング値やガンマ値を明記している。合成後の HDR 画像に対する `tonemap` 処理は使用者の好みに応じて明暗値のクリッピング値やガンマ値といった処理パラメータの選択、適用が可能であるが、5.3 章にて述べるように、撮影後にフォーカス位置や絞り値の調整と合わせて、この `tonemap` 処理を調整できることは、ユーザに大きな価値を提供できる[68]。例えば、飽和領域のぼけの印象は通常領域と大きく異なり、応じた撮影後の調整が可能となる。

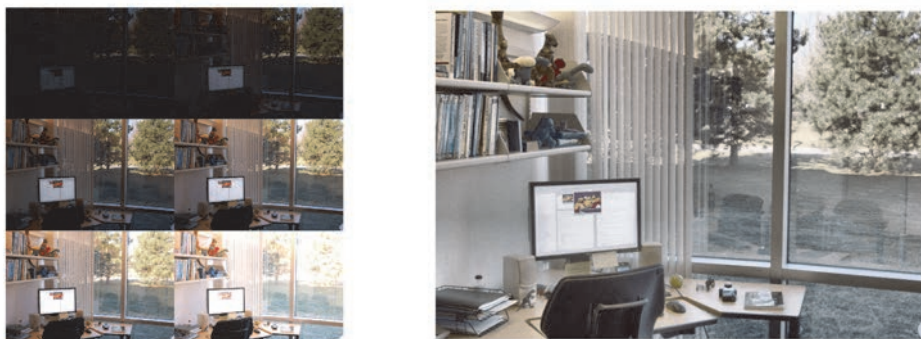


図 3.8 HDR 画像合成の例 (Image Credit: Matlab website [63]):  
左に露光期間の異なる入力画像群、右に合成後の HDR 画像 (標準 `tonempe` 処理)。

## 第4章 実験と評価

本論が提案する画像スタックの再構成アルゴリズムを実証するため、データセットを撮影し、用意した。まず、11シーンの Aperture-Focus スタックを撮像し、5シーンをテスト用、6シーンを学習用に分割した。この Aperture-Focus スタックは、18の絞り値、45のフォーカス位置の合計 810 枚の画像をそれぞれのシーンについて撮像した。18の絞り値は F/2.2 から F/16 の範囲で、45のフォーカス位置は 0.45m から 1m の範囲で変更した。撮像した Raw 画像の画素数は、1988×1296 であるが、計算とデータコストの点から、3.1 章でその詳細を説明した、幾何および輝度キャリブレーションを実施した後に、600×400 に間引きサンプリングして実験に用いた。本論のアルゴリズムは完全に画素ごとに行われる独立な処理であるため、計算量は、画素数に応じて線形に軽減されるにもかかわらず、間引きサンプリングにより再構成の品質が損なわれることは原理的にない。

さらに、13シーンの MLAF スタック (Multi-level Aperture-Focus スタック、フォーカス・絞り・露光レベルの全パラメータで撮影した画像スタック) のデータセットを撮像し、6シーンをテスト用、7シーンを学習用に分割した。MLAF スタックは、22の絞り値、51のフォーカス位置、3の露光レベル、合計 3366 枚の画像をそれぞれのシーンについて撮像した。22の絞り値は F/1.4 から F/16 の範囲で、51のフォーカス位置は約 0.8m ~2m の範囲で変更する。撮像した Raw 画像の画素数は、上と同様に 1988×1296 であるが、幾何および輝度キャリブレーションの後に、切り出しと間引きサンプリングにより、960×640 にして実験に用いた。

全てのデータセットは、Canon EOS-40D のカメラ[69]、Canon EF 50mm F1.4 固定焦点レンズ[70]を用いて、カメラは PC により自動で撮影が進むよう制御してデータセット作成を行った。本論における全ての実験では、テスト用と学習用のデータセットは完全に独立であり、オーバーラップはない。色空間は RGB であり、それぞれのカラーチャンネルごとに処理する。本論の提案する手法は、基本的に色空間は問わず適用可能であるが、元画像が飽和輝度を有する場合、カメラ内のイメージセンサの色空間である RGB からの変換を実施すると、複数チャンネルに相互影響を及ぼすため、イメージセンサの色空間のまま扱うことが望ましいと考え、RGB 色空間を適用し、カラーチャンネルごとと個別に処理することとした。

## 4.1 Focus スタックの再構成評価

まず、数枚のフォーカス位置を変えて撮像した画像からの、完全な Focus スタックを復元についての評価から開始する。1次元の Epsilon Photography であるが、本手法の有効性の確認のため、Focus スタックの評価から開始することとした。用意した評価用 Focus スタックは、前述の Aperture-Focus スタックと同じシーンであるが、全てのシーンはより大きな絞り F/1.4 にて撮影、収集した。これは、図 2.4 でも示した通り、小さい絞りと比べて大きな絞りの方が、被写界深度が浅くフォーカス位置の変化に応じたボケ量の変化が大きくなるためである。

図 4.1 は、Focus スタックの再構成評価についての概念図である。45 枚のフォーカス位置が異なる Focus スタックを収集し、これを ground-truth データセットとして用いる。中から 8 枚の画像のみを用いて、45 枚の Focus スタックの再構成を提案手法により実施し、その再構成品質の評価を行う。

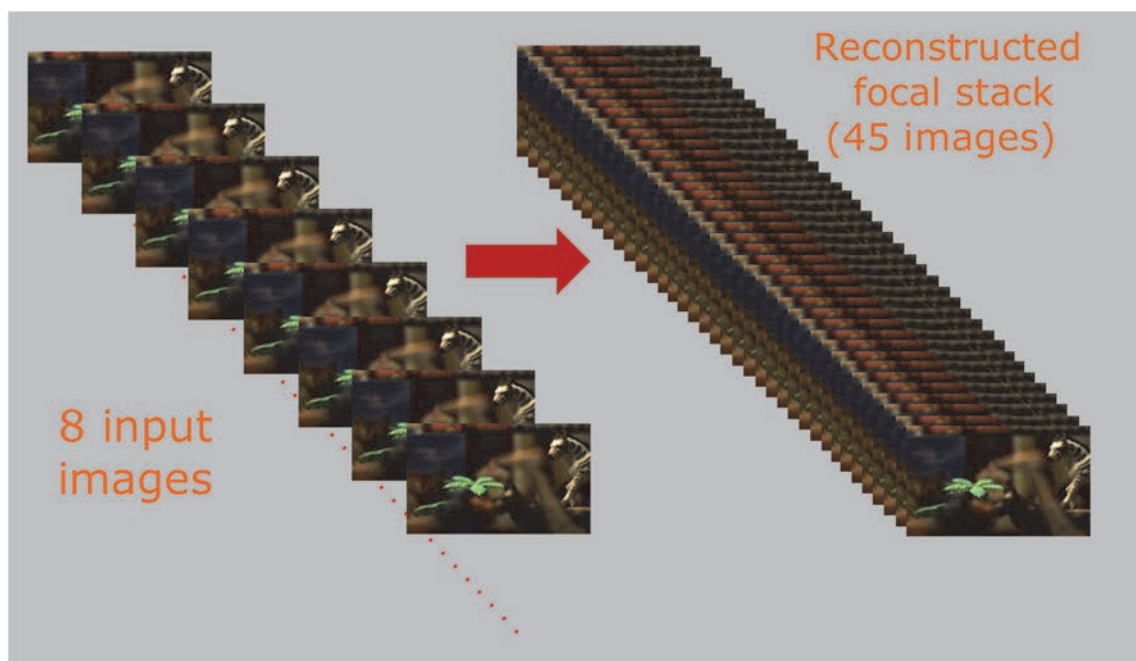


図 4.1 Focus スタックの再構成評価の概念図:

8 枚の画像を用いて、入力画像に存在しない中間位置でのフォーカス画像を含む、完全な 45 枚の画像で構成される Focus スタックを再構成する。

評価の1つのポイントは、フォーカスが合致した鮮明な領域が撮像されていないとしても、当該フォーカス位置での鮮明な画像が復元できているかどうか、に置くことができる。図 4.2 および 4.3 は、シーン”Animals”とシーン”Liquid”について、この観点に注目して評価を実施したものである。どちらの再構成処理においても、入力画像枚数は 8 枚で、画像スタック全体で 45 点の Focus スタックの復元を試みる。各図の(a)は再構成画像の 1 フレームを取り出した例であり、再構成されたスタック全体の SNR は、図 4.2 の”Animals”で 25.8 dB、図 4.3 の”Liquid”においては 32.0 dB であった。両図の(b)では、(a)内に示すピックアップした画素位置の Intensity Profile（画素地の集合）について、フォーカス位置を変えた際の遷移を、ground-truth と比較している。横軸はフォーカス位置のインデックスであり、星印のフォーカス位置で撮像した 8 枚の画像を入力としている。この Intensity Profile は、どの点においても極めて ground truth に近い復元を示している。また各図の(c)と(d)では、(a)内に示すピックアップした画像領域の再構成品質を、ground truth と並べて比較している。図 4.1(c)において図示しているが、2 つの入力画像において、手前側にフォーカスが合っている 1 枚目の入力画像では、”Leaf 1”にフォーカスが合致している。一方で、奥側にフォーカスが合っている 2 枚目の入力画像では、”Leaf 3” にフォーカスが合致している。その中間の奥行きに存在している”Leaf 2”に焦点が合っている画像は、入力画像には存在していない。再構成された中間フォーカス位置の画像から、”Leaf 1”のぼけ量は線形に増加していき、逆に”Leaf 3”のぼけ量は線形に減少していく様がよくわかる。さらには、シャープな画像が入力画像のどちらにもない”Leaf 2”では、中間画像がシャープに再構成され、フォーカスが合った画像が復元されていることも見て取れる。各図の(c)および(d)は見かけ上、ground truth と比較してもどちらかわからないほど非常によく似ているといえる。

ここでもう一つ、提案手法による Focus スタックの再構成が、広範囲に複雑な反射特性を持つ材質にも適用できていることが言える。図 4.3(c)におけるガラスのコップのように、透明な材質を撮像したときにも、そのフォーカス位置の変化が自然に見て取れる。また、図 4.3(d)の”grape 1”と”grape 2”の領域における、強い鏡面反射をもつ被写体周辺でのフォーカス位置遷移でも確認できる。さらには、図 4.3(c)の”orange”の周辺で見られる Sub-surface scattering が起きている領域においても、自然にフォーカス位置変化を再構成していることが視認できる。

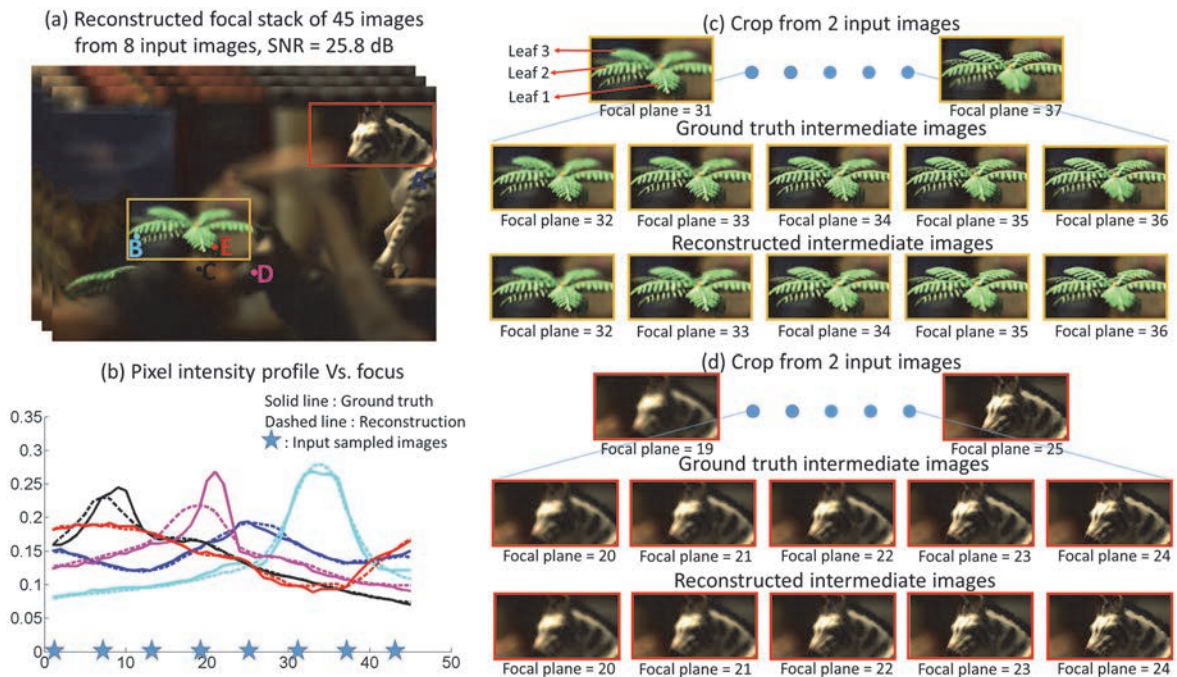


図 4.2 シーン”Animals”における Focus スタックの再構成例:

8 枚の離散的にフォーカス位置が異なる入力画像から 45 枚の Focus スタックを復元している. (a)45 枚の再構成画像から 1 枚の例. スタック全体の再構成の SNR は 25.8 dB. (b) 再構成画素を(a)からピックアップした点についての Intensity Profile の ground-truth との比較. (c)(d)ズームアップした領域における 2 枚の入力画像とその中間フォーカス位置にあたる再構成画像の品質を ground-truth と比較したもの. 2 つの入力画像において, 手前側にフォーカスが合っている 1 枚目の入力画像では, ”Leaf 1”にフォーカスが合致している. 一方で, 奥側にフォーカスが合っている 2 枚目の入力画像では, ”Leaf 3” にフォーカスが合致している. その中間の奥行きに存在している”Leaf 2”に焦点が合っている画像は, 入力画像には存在していない. 再構成された中間フォーカス位置の画像から, ”Leaf 1”のぼけ量は線形に増加していき, 逆に”Leaf 3”のぼけ量は線形に減少していく様がよくわかる. さらには, シャープな画像が入力画像のどちらにもない”Leaf 2”では, 中間画像がシャープに再構成され, フォーカスが合った画像が復元されていることも見て取れる.

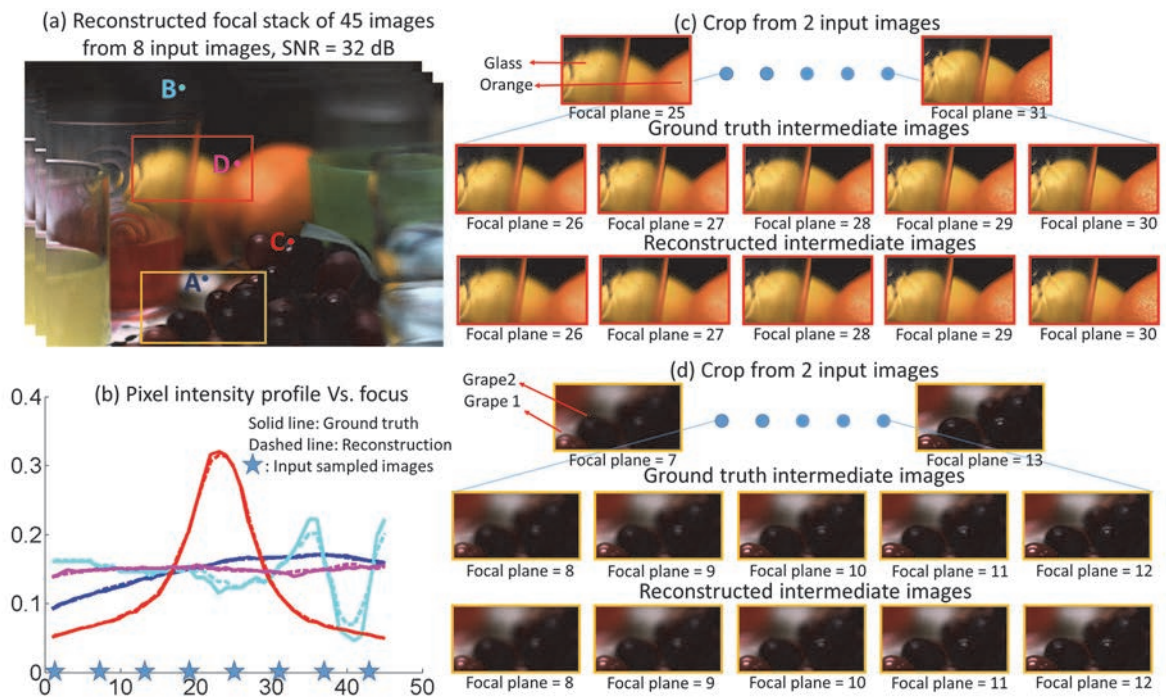


図 4.3 シーン” Liquid” における Focus スタックの再構成例:

8枚の離散的にフォーカス位置が異なる入力画像から45枚のFocusスタックを復元している。(a)45枚の再構成画像から1枚の例。スタック全体の再構成のSNRは25.8dB。(b)再構成画素を(a)からピックアップした点についてのIntensity Profileのground-truthとの比較。(c)(d)ズームアップした領域における2枚の入力画像とその中間フォーカス位置にあたる再構成画像の品質をground-truthと比較したもの。(c)におけるガラスのコップのように、透明な材質を撮影したときにも、そのフォーカス位置の変化が自然に見て取れる。また、(d)の”grape 1”と”grape 2”の領域における、強い鏡面反射をもつ被写体周辺でのフォーカス位置遷移でも確認できる。さらには、(c)の”orange”の周辺で見られるSub-surface scatteringが起きている領域においても、自然にフォーカス位置変化を再構成していることが視認できる。



Focus スタック画像を用いたアプリケーションとして、Depth from Focus[7, 33]の原理によるデプスマップの生成と、高品位な全画素に焦点が合った画像（全焦点画像 / All Focus Image, 本論では、Extended Depth of Field (EDoF)画像の呼称[71]で統一する）の生成がよく知られている。本論ではこのアプリケーションを、再構成された Focus スタックを用いて実現できることの確認と評価を行う。

本論以外の Focus スタック復元との結果を比較するため、キュービック・スプライン補間 (cubic spline interpolation) の方法にて再構成した Focus スタックを生成した。ここで実施する、Depth from Focus によるデプスマップの生成処理と、EDoF 画像の生成処理は、商用販売しているソフトウェアである HeliconSoft[72]を使用した。図 4.4 は、シーン”Chess”について、ground-truth の Focus スタックとキュービック・スプライン補間で生成した Focus スタック、および、本論の提案する GMM ベースの手法で再構成された Focus スタック、をそれぞれ入力した場合に対し、HeliconSoft が出力するデプスマップと EDoF 画像について、それぞれ画像中の一部分の拡大を含め、品質を比較したものである。生成したデプスマップは、奥行きが手前ほど暗い、奥ほど明るい輝度値で表現している。

本提案手法で再構成された Focus スタックにより生成されたデプスマップは、概ね ground-truth の Focus スタックから生成されたデプスマップと近い見た目を持っている。一部、物体の内部のテクスチャがない平坦な領域においては、奥行きのエラーが散見されるが、キュービック・スプライン補間で生成した Focus スタックから出力されたデプスマップと比較すれば、大幅に改善された結果となっている。また、EDoF 画像における比較においては、本提案手法で再構成された Focus スタックにより生成された EDoF 画像と、ground-truth の Focus スタックから生成された EDoF 画像の間に、見た目上の差はほとんどない。区別がつかないほどである。一方で、キュービック・スプライン補間で生成した Focus スタックからの EDoF 画像では、目立ったアーティファクトが存在している。提案の GMM ベース手法が高い再構成性能を持つこと、完全な Focus スタックによって実現されるアプリケーションを、少ない入力枚数で実現できることが確認された。

図 4.5 では、シーン”Liquid”について、ground-truth の Focus スタックとキュービック・スプライン補間で生成した Focus スタック、および、本論の提案する GMM ベースの手法で再構成された Focus スタック、をそれぞれ入力した場合に対し、HeliconSoft が出力するデプスマップと EDoF 画像について、拡大した図で品質を比較している。本提案手法で再構成された Focus スタックにより生成されたデプスマップおよび EDoF 画像は、キュービック・スプライン補間で生成したそれらと比較して、概ね ground-truth の Focus スタックから生成されたそれらと近い品質を持っていることが、こちらのシーンからも見て取れる。

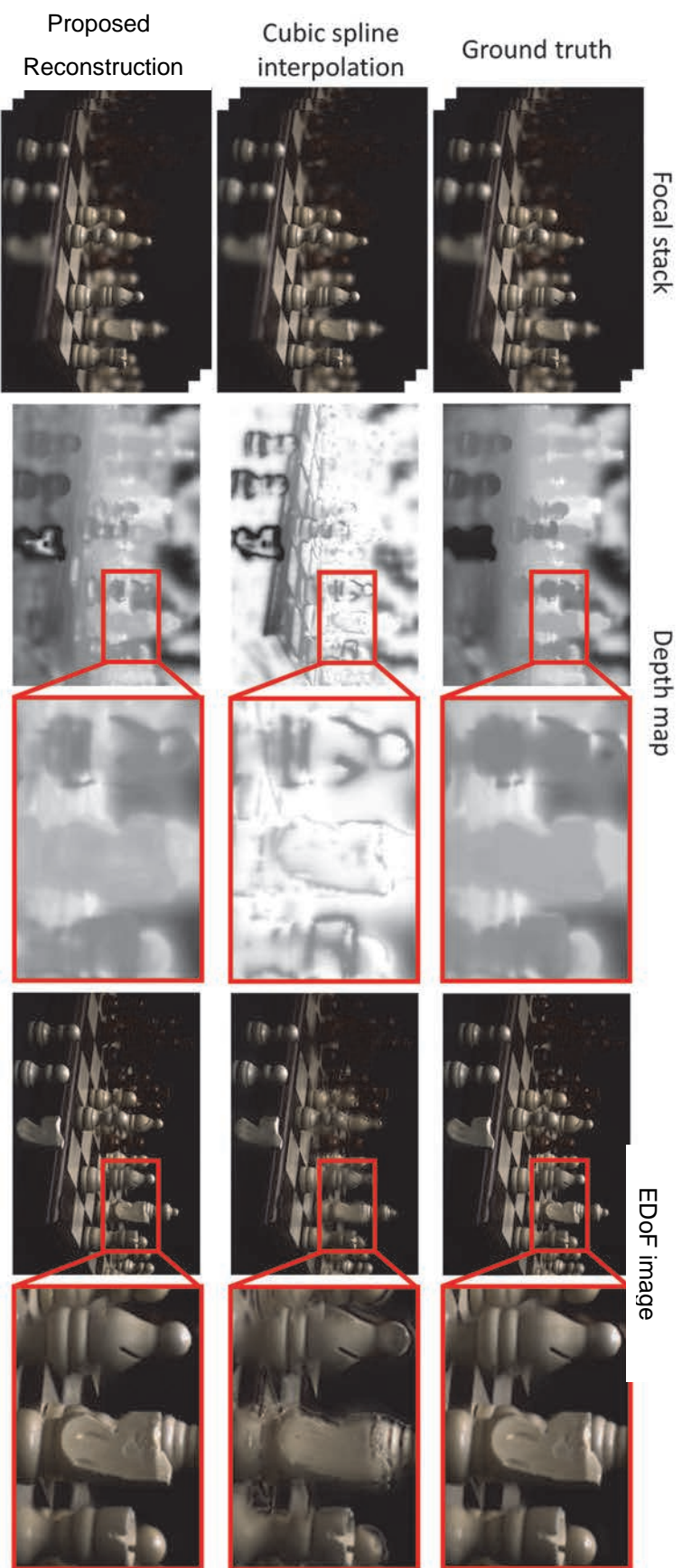


図 4.4 シーン” Chess” における再構成 Focus スタックを用いて生成したデプスマップおよび EDoF 画像:

8枚の離散的にフォーカス位置が異なる入力画像から45枚のFocusスタックを再構成し、HeliconSoft[72]に入力、デプスマップとEDoF画像(全焦点画像)を生成した。ground-truthのFocusスタック、および、キュービック・スプライン補間で生成したFocusスタックを入力した場合と比較している。

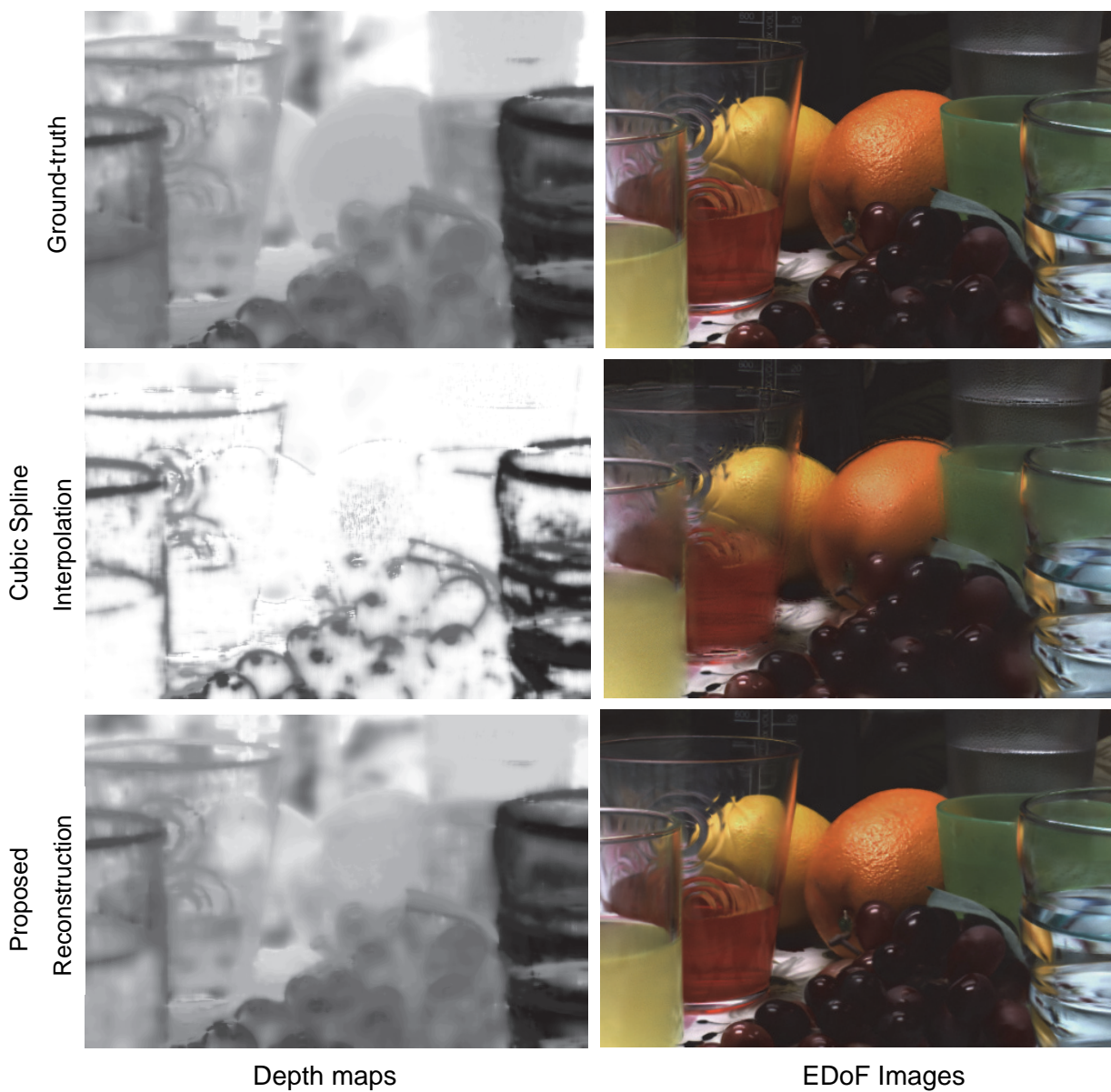


図 4.5 シーン” Liquid” における再構成 Focus スタックを用いて生成したデプスマップおよび EDoF 画像: 8 枚の離散的にフォーカス位置が異なる入力画像から 45 枚の Focus スタックを再構成し, HeliconSoft[72]に inputs, デプスマップと EDoF 画像 (全焦点画像) を生成した. ground-truth の Focus スタック, および, キュービック・スプライン補間で生成した Focus スタックを入力した場合と比較している.

## 4.2 Aperture-Focus スタックの再構成評価

次に、2次元の Epsilon Photography である、Aperture-Focus スタックの再構成評価を行う。数枚のフォーカス位置と絞り値を変えて撮像した画像からの、完全な Aperture-Focus スタックを復元についてその品質評価を行っていく。図 4.6 に示すのは、Aperture-Focus スタックの再構成評価についての概念図である。18 の絞り値, 45 のフォーカス位置の合計 810 枚の画像を `ground-truth` データセットとして収集し, 中から 8~16 枚の画像のみを用いて, 提案手法により完全な画像スタックの再構成を実現する。

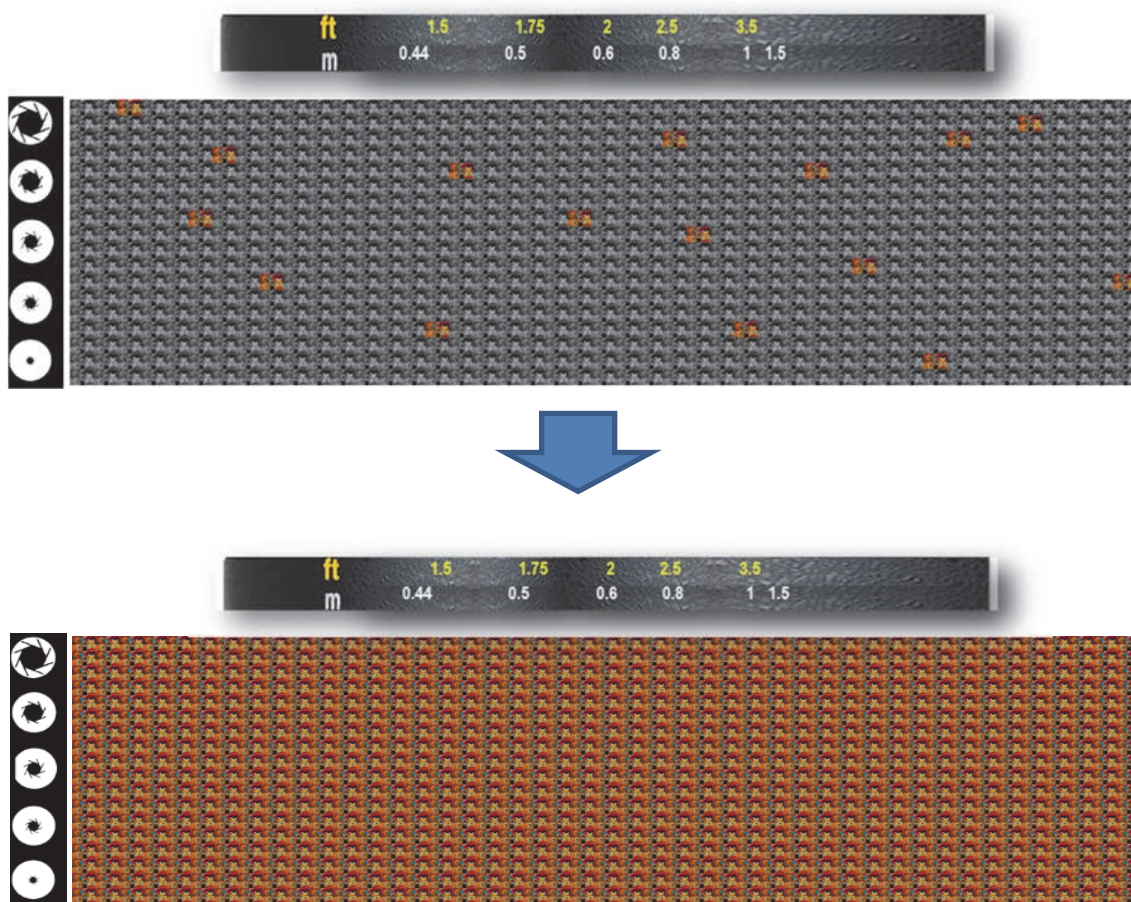


図 4.6 Aperture-Focus スタックの再構成の概念図:

8~16 枚の画像を用いて, 入力画像に存在しないフォーカス位置, 絞り値で撮像された画像を含む, 完全な 810 枚の画像で構成される Aperture-Focus スタックを再構成する。

Aperture-Focus スタックの評価として、まず入力枚数に応じた再構成品質の評価を行った。図 4.7 は、3つのデータセットに対し、入力枚数を 4, 8, 16, 32 枚のそれぞれに設定した際の、再構成品質を PSNR で評価したものである。図 4.7 より、シーン “Chess” のデータセットは、入力枚数を 4 から 32 枚に増加させてもそれほど改善はないことがわかる。一方、より複雑なシーンであるシーン “Animals” とシーン “Glassball” のデータセットにおいては、特にテクスチャの増加と相互反射の画素が存在していることを要因に、4 から 16 枚への入力増加による改善は大きく、32 枚への増加においても PSNR は若干ながら続伸している。この実験結果から、後の Aperture-Focus スタックの再構成の挙動を評価するためには、16 枚の入力がバランスがよいと考え、以降の Aperture-Focus スタックに関する実験では入力を 16 とした。

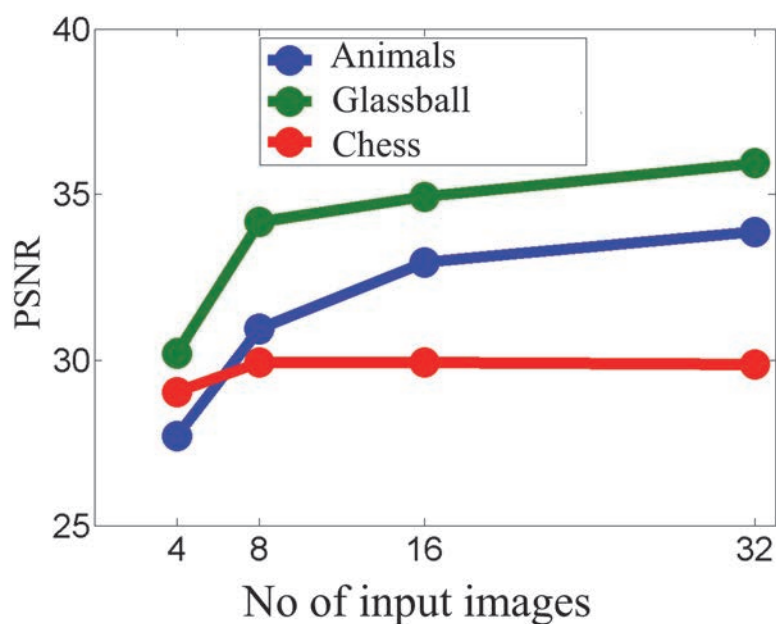


図 4.7 入力枚数による再構成画像の PSNR 比較

3つのデータセットのうち、高周波領域が多く複雑な被写体（シーン “Animals” やシーン “Glassball”）では、入力枚数が 4 から 16 枚の増加により PSNR は向上する。一方、平坦領域の多いデータセット（シーン “Chess”）では、入力枚数が 8 枚から増加させても PSNR はほとんど変わらない。

図 4.8 においては、Aperture-Focus スタックの特定の位置から画素値を集めて構成した AFI 画像について、再構成画像から得られたものと ground-truth から得られたものとを、視覚的に比較評価している。図 4.8 右の AFI 画像は、図 4.8 左の画素中の各画素位置に対して、それぞれ 45 点のフォーカス位置を横軸に、18 点の絞り値を縦軸にとり、カラーでの 2 次元画像として図示したものである。Aperture-Focus スタック中から、いくつかピックアップした点に関する AFI における評価から、その品質が高いことが確認できる。その結果、5 章において示すが、confocal stereo アルゴリズムによって得られるデプスマップは、その入力である Aperture-Focus スタックについて、直接 ground-truth を入力とした場合と、再構成データセットを入力した場合において、非常に近い結果が得られる。星点はその奥行き値を示している。

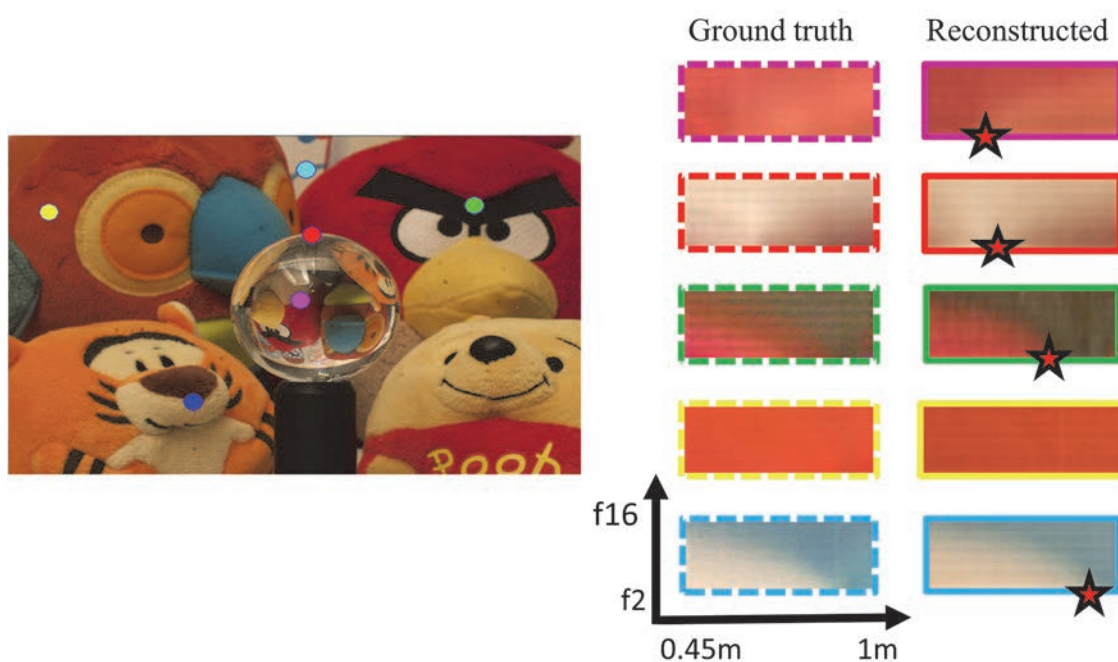


図 4.8 再構成された AFI 画像

左図で示す領域について、AFI 画像の ground-truth と再構成結果を比較したものが右図である。テクスチャ領域も奥行き境界のエッジ領域も変わらず再構成できていることが確認できる。

図 4.9 は、再構成された Aperture-Focus スタックから、異なるフォーカス位置・絞り値に相当する画像について、特定の領域をズームアップして示している。また、ground-truth として実際に撮影された Aperture-Focus スタックから、対応する撮像パラメータの画像、ズームアップ領域について並べて再構成品質を比較している。図 4.9 左上にはグリーディ・アルゴリズムで選択された入力画像の撮像パラメータと、図 4.9 右に表示している再構成画像相当の撮像パラメータをカラーの四角でマッピングしている。(図 4.9 右のズームアップ画像の枠の色が、図 4.9 左上の四角の色に対応している。) 入力画像の撮像に選択されているパラメータは全部で 16 点である。図 4.9 左上の横軸が絞り値 (左⇄右で大⇄小) であり、縦軸がフォーカス位置のインデックス (下⇄上で近⇄遠) である。ズームアップした画像相当の撮像パラメータは、選択された入力画像の撮像パラメータの全てから出来るだけ遠いもの、かつ、画質変化の大きい、絞りサイズが出来るだけ大きいもの、を選択している。図 4.9 左下は、2つのシーンの構図全体から図 4.9 右でズームアップした領域の位置を示したものである。距離の近い領域から遠い領域までまんべんなく選択している。

図 4.9 右で示された、シーン”Chess”およびシーン”Animals”の一部分のズームアップ領域について、再構成品質を ground-truth と視覚的に比較評価する。水色や赤色の枠で示された、大きな絞り値 (F/2.2) 相当の画像では、浅い被写界深度により、フォーカスが合っていない部分では大きなぼけが発生しているが、再構成画像と ground-truth の元画像は非常によく似ている。フォーカスが合っている部分においても鮮明な画質に遜色はない。一方の中間の絞り値 (紫色の F/5 相当、橙色の F/6.3 相当) や小さな絞り値 (紫色の F/8 相当) では、空間解像度が高く鮮明な画質が再現されている。この結果は、本手法が再構成ターゲットである Aperture-Focus スタックの全枚数 (810 枚) と比べて、大幅に少ない 16 枚の入力であっても、スタック全画像を再現性高く再構成できることを示している。

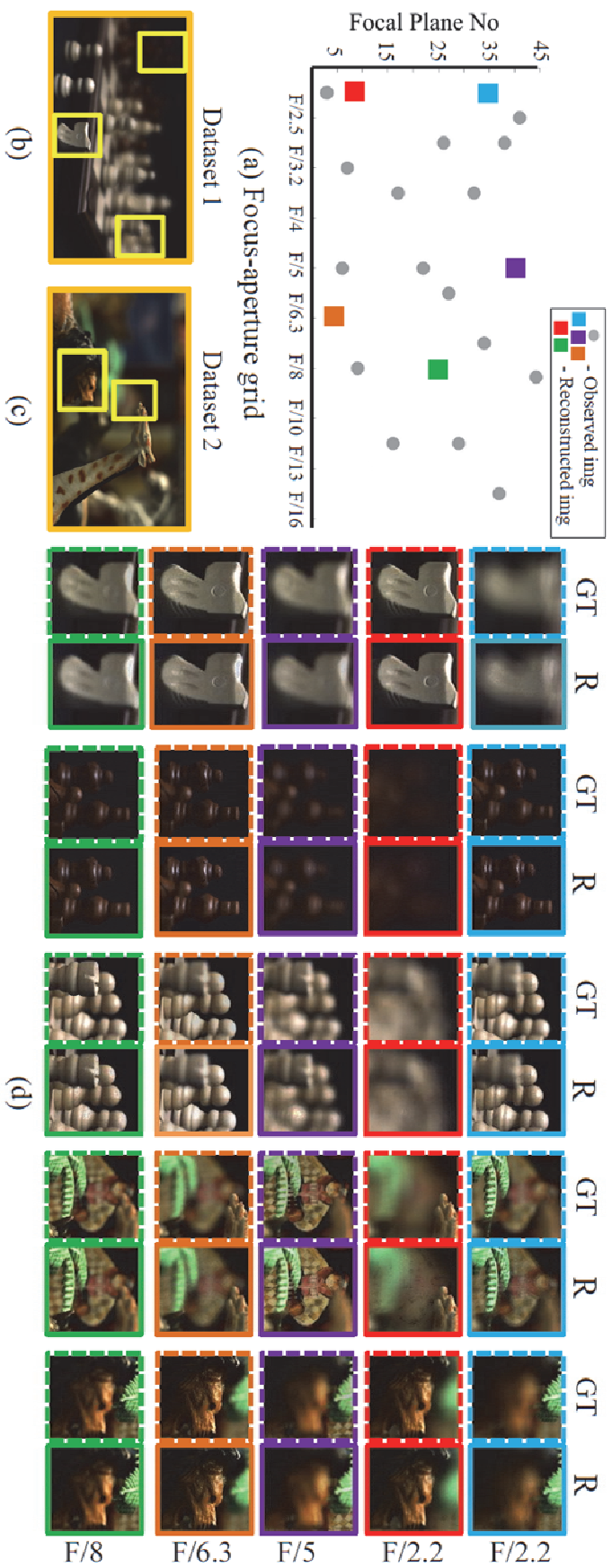


図 4.9 Aperture-Focus スタック再構成

いくつかの画像パッチで再構成結果を比較する。(a) 円で示す点が 16 の入力画像のパラメータ組合せ, (b, c) 再構成したデータセット中の画像パッチの位置, (d) 各画像パッチの再構成結果(R)と ground-truth(GT). なるべくサンプリング点から遠い再構成結果を表示している. 再構成品質が高いことを示された.



Focus スタックの場合と同様に、16枚のサンプリング撮像された画像にはフォーカスが合致したシャープなポイントが存在しなくても、このフォーカス位置でのシャープな画像が復元できているかどうか、は再構成のポイントである。図 4.10 は、シーン”Glassball”とシーン”Chess”において、パラメータが中間の画像再構成結果を示すものである。2段目の再構成画像に対して、1段目と3段目は実撮像された最も近いパラメータの画像であり、フォーカスは外れている。再構成されたフォーカスが合った画像では、テクスチャの細部がよく再現されていることが見て取れる。

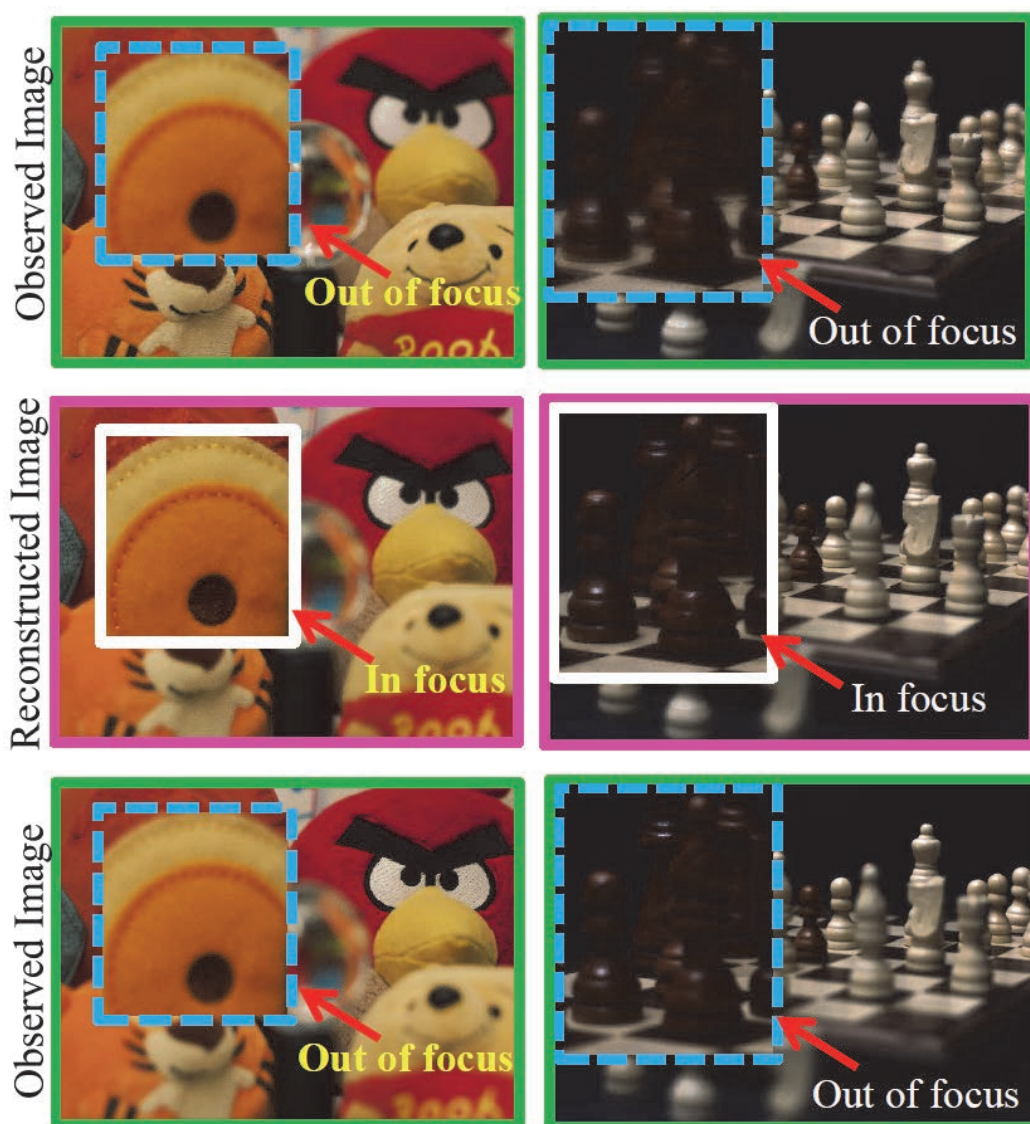


図 4.10 撮像されていないパラメータ相当の画像復元結果

図 4.11 においては、シーン “Glassball” およびシーン “Chess” のデータセットについて、異なるフォーカス位置および絞り値における再構成画像をピックアップして示している。図 4.11 (a)では、絞り値を F/2.0 (本実験で再構成する Aperture-Focus スタックの最大絞り値) に固定し、フォーカス位置を手前から奥へと変更していった場合の再構成画像を左から右へ示している。シーン “Glassball” 中の被写体である、ガラス球～Pooh～Tigger～Angry Bird の順にフォーカス面が推移していく様子が見てとれる。ぼけに関しても非常に自然に生成されており、不自然な部分は目立っていない。もう一方のシーン “Chess” についても、手前の駒から後ろの駒へとフォーカス位置が遷移していく様子が見て取れる。手前の駒に関しては、fp42 のフォーカス位置では大きなぼけ量になっているが、不自然な印象はない。絞り値が大きく、ぼけ量が大きな領域においても、本論の提案する再構成手法の高い品質が見て取れる。

また図 4.11 (b)では、図 4.11 (a)とは反対に、フォーカス位置を固定した上で、絞り値を徐々に変更した場合の再構成画像を示している。絞りの大きさが F/2.2 から F/11 へと小さくなっていくにつれ、被写界深度が広がっていき、手前から奥までフォーカスが合った、鮮明な画像となっていく様子が見て取れる。これらから、本手法が、実際に撮影していないフォーカス位置の画像を、大きな絞り値であっても、信頼性高く再構成できること、を示している。

図 4.12 は、屋外シーンで撮影した 16 枚の画像から再構成した Aperture-Focus スタックの一例である。再構成された Aperture-Focus スタックの全画像枚数は 810 枚であるが、フォーカス位置 6 点、絞り値 4 点の 24 枚を間引きして示している。ここで撮影された屋外シーン “Tennis” の画像は入力 16 枚のみである。(完全な Aperture-Focus スタックとして評価用に準備した 11 シーンのデータセットとは別のデータである。) 屋外では、光源環境や風の影響による被写体の条件の変化が起こりやすい。16 枚の撮影には、PC でカメラを制御する都合上、現状 10 分程度を要しており、一定の光源状況の変化などは避けられていない。入力画像が完全に静止した状況でないながらも、本論の提案する画像スタック再構成は実に再現性高く効果を得ている。図中のテニスラケットとボールに注目すると、フォーカス位置に応じたぼけや被写界深度といった Aperture-Focus スタックの特徴が忠実に再現されていることが見て取れる。本データには ground-truth となる画像スタックが存在しないため、定量・定性的な比較による再構成品質の評価は出来ないが、光源の揺らぎ等の変動要因が存在しても、本論の提案するアルゴリズムが一定の頑健性を持つことが確認できる。

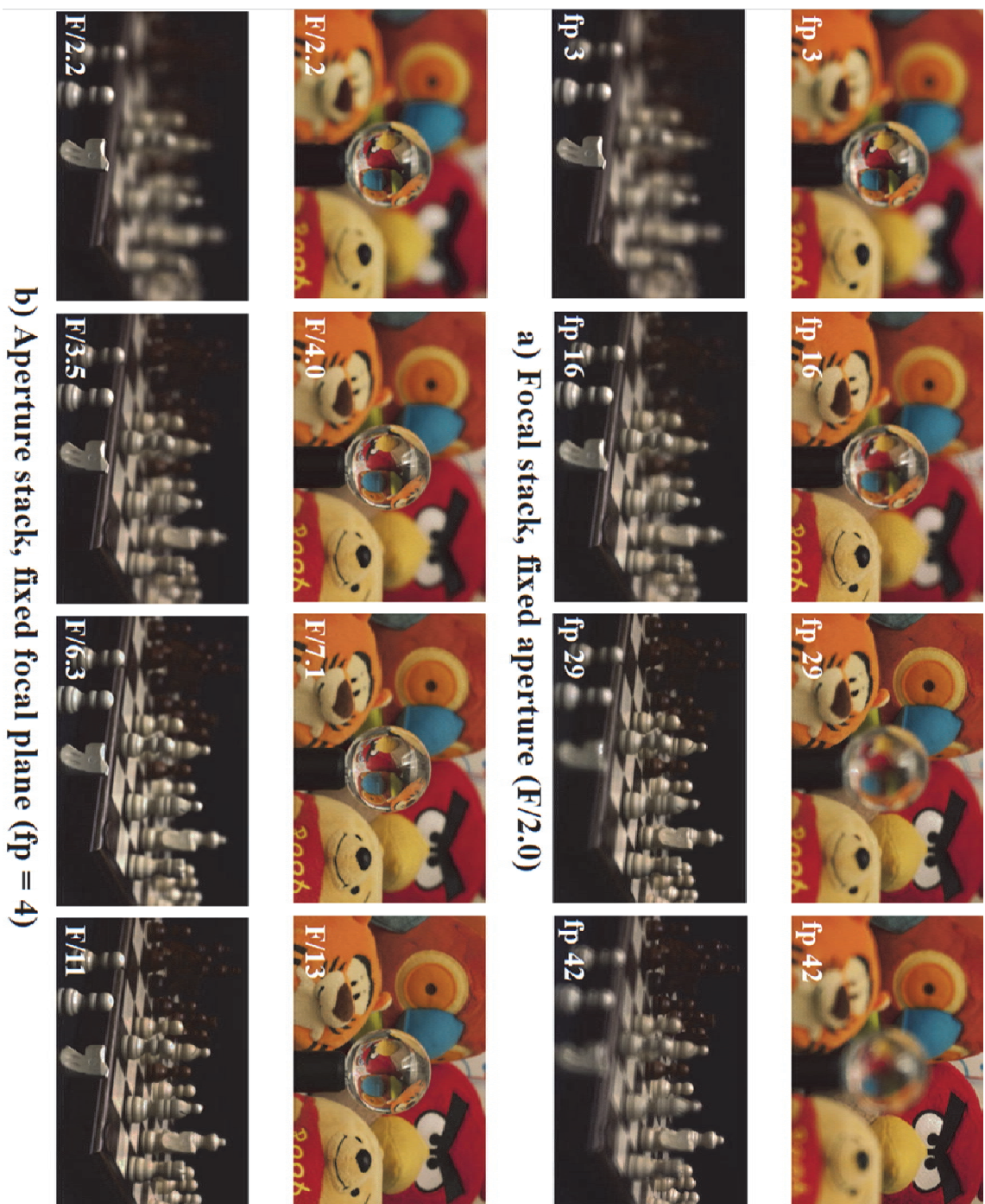


図 4.11 Aperture-Focus スタック再構成

シーン “Glassball” およびシーン “Chess” のデータセットにおいて，(a)固定の絞り値  $F/2.2$  における「Focus スタック」，ボケ量およびボケ領域がフォーカス平面  $3(fp3)$  から平面  $42(fp42)$  へと移動していることが見てとれる．(b)固定のフォーカス位置における「Aperture スタック」，左から右へ絞り値サイズが変わることで DoF が広がっているのがわかる．

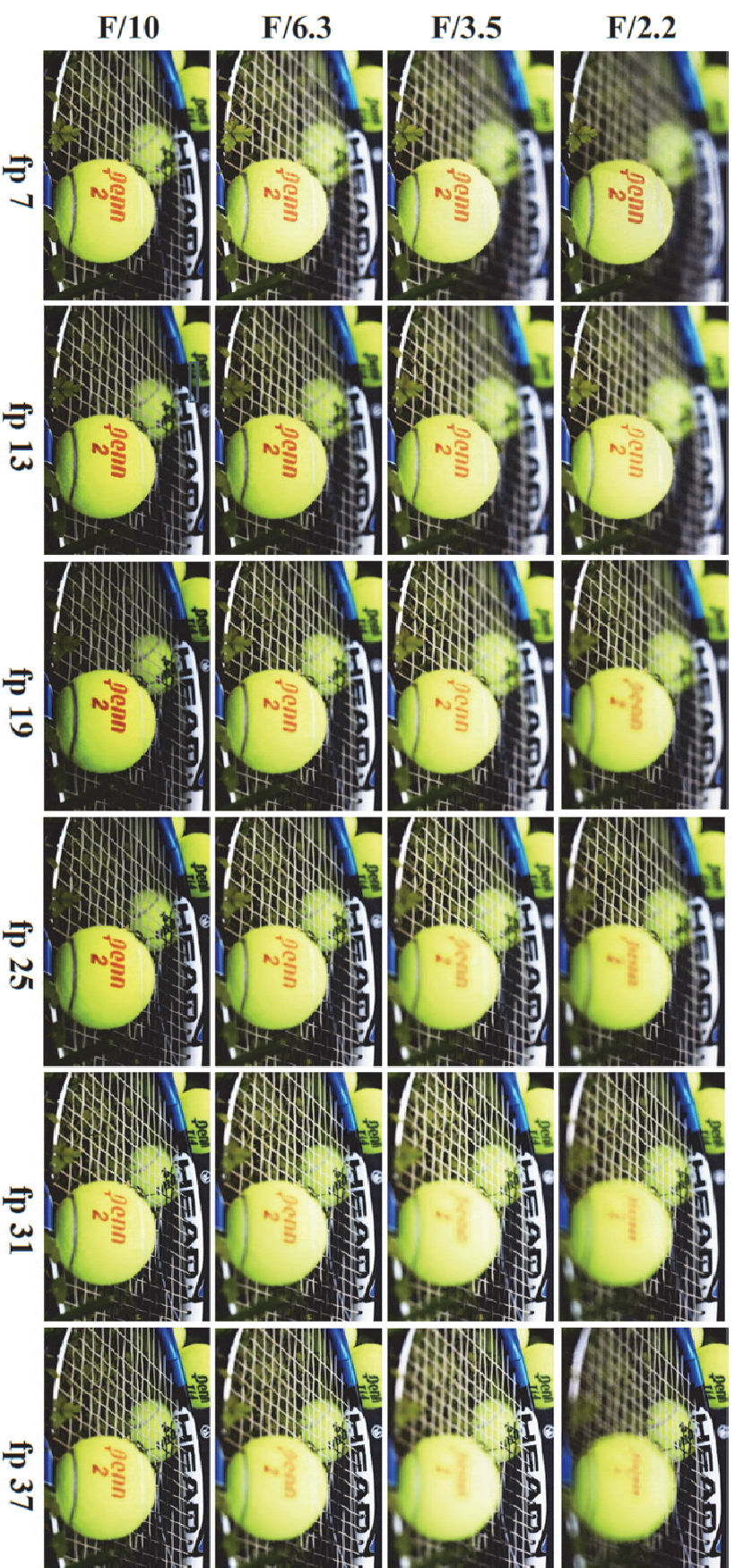
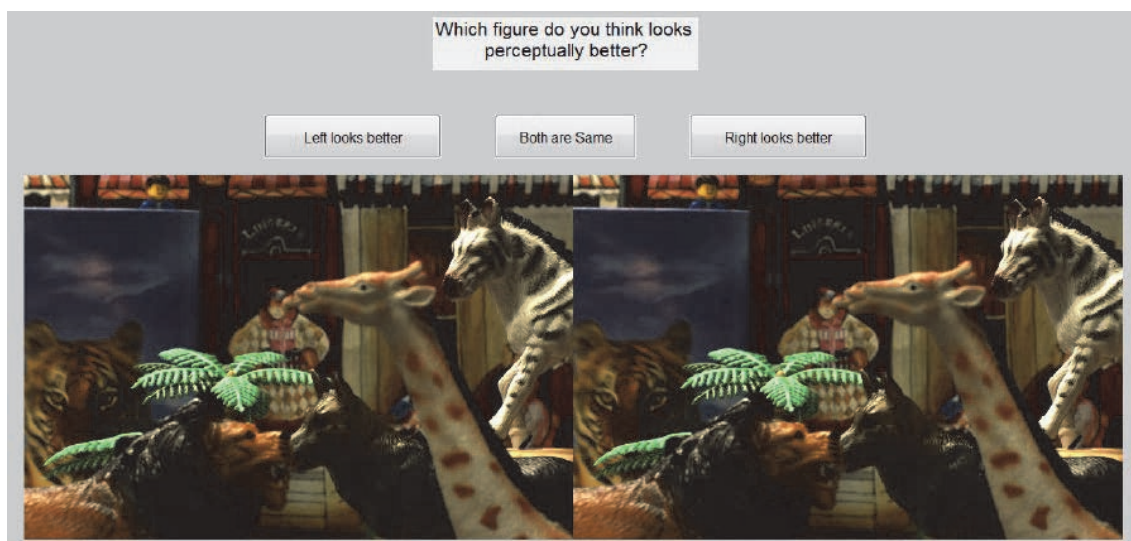


図 4.12 屋外シーン “Tennis”における Aperture-Focus スタック再構成

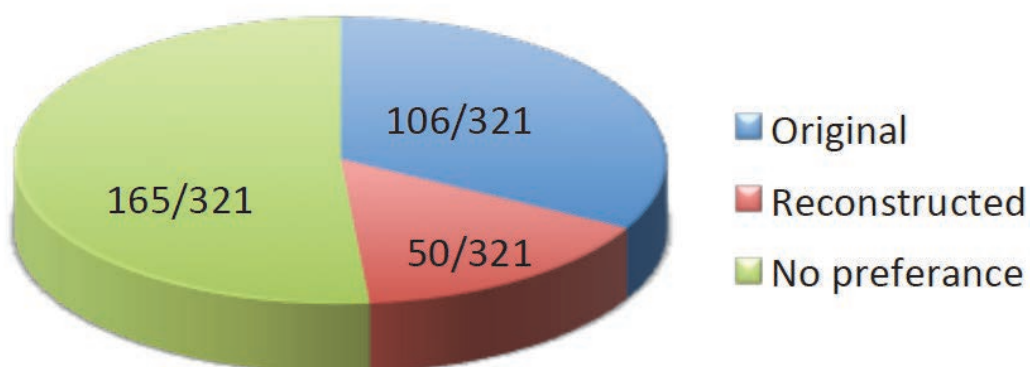
16枚の入力画像から再構成された810枚の Aperture-Focus スタックから、フォーカス位置6点、絞り値4点の24枚の再構成画像を示している。縦軸に沿って画像を観察すると、絞り値がF/2.2からF/10へと小さくなっていくにつれて被写界深度が深くなくなる様子が見て取れる。また、特に最大絞り値(F/2.2)において、横軸に沿って観察するとフォーカス面が近くから遠くへと動いていく様子が見て取れる。

画像スタックの再構成品質において、PSNR などによる定量評価は、視覚的品質を常に保証するものではないため、本論では官能評価実験[73]も合わせて実施し、主張のさらなる検証を行った。本官能評価実験は、再構成された画像が、元の画像と区別可能であるかどうかを見極めることが目的である。図 4.13 に被験者への提示例と官能評価結果を示す。手順としては、13 名の被験者に対し、1 枚は実際に撮影されたオリジナル画像、もう 1 枚を再構成された画像、の 2 枚ペアで、左右ランダムに切り替えて表示する(図 4.13(a))。シーンとしては、5 つのテスト用シーンに対して、6 セットずつ抽出し、撮像パラメータであるフォーカス位置と絞りのサイズについては、ランダムに選択した。全ての再構成画像は、16 枚の入力画像を用いて生成されたものを用いた。被験者には、「右画像が好ましい」「左画像が好ましい」「差はわからない」の 3 種類いずれかを回答してもらう。各被験者への回答制限時間は設けず、異なる領域をズームインして詳細に見ることも可能とした。30 セットの比較を行ってもらい、平均で 5 分ほどを要した。最終的に、 $13 \times 30 = 390$  セットの比較結果が得られたが、そのうち 69 セットは、表示ペアを完全に同じ画像にして表示した。これは実験の公平性を検証するための措置である。この 69 セットにおける結果は、約 87% に当たる 60/69 セットの比較において、被験者は「差がわからない」と答えた。これは、本実験の妥当性を示す 1 つの論拠といえる。

図 4.13(b)は、得られた結果の統計量である。まず、最大の約 51% (165/321 セット) の比較について、被験者は「差がわからない」と回答した。さらに驚くことに、約 15% (50/321 セット) が再構成画像の方が「好ましい」と回答した。全体の約 3 分の 1 では、被験者は元画像を好み、全体の 6 分の 1 で被験者は本論の提案手法による再構成画像を好むという結果である。言い換えれば、かなりの割合(215/321 セット, 66%程度)の被験者が、再構成画像は元画像に対し、同等または好ましい、と回答したということである。ランダムにフォーカス位置と絞りのサイズを選択しているため、比較評価された画像セットには、大きなぼけを有する画像から、手前から奥までフォーカスが合った画像までまんべんなく含まれていた。この結果は、本手法の再構成品質の高さを示す 1 つのエビデンスであると言える。



(a) Snapshot of our evaluation tool



(b) User preference statistics

図 4.13 官能評価実験

(a) 被験者に提示した画像と選択システム. 一方はオリジナル, 他方は再構成画像である. 13 人の被験者にランダムで現れる画像を 30 回ずつ比較させた. 回答は「左が好ましい」「違いはわからない」「右が好ましい」から選択させた. (b) 51%の比較において, 被験者は「違いはわからない」と回答した. 再構成結果の品質が, 元画像と遜色ないことがいえる.

### 4.3 MLAF スタックの再構成評価

フォーカス位置と絞り値を変えて撮像した Aperture-Focus スタックに加え、3つ目の撮像パラメータとして、露光レベルを変えて取得した MLAF (Multi-level Aperture-Focus) スタックの再構成処理を評価する。図 4.14 に示すのは、MLAF スタックの再構成評価についての概念図である。22 の絞り値、51 のフォーカス位置、3 段階の露光レベルの合計 3,366 枚の画像を ground-truth データセットとして収集し、中から 16~32 枚の画像のみを用いて、提案手法により完全な画像スタックの再構成を実現する。

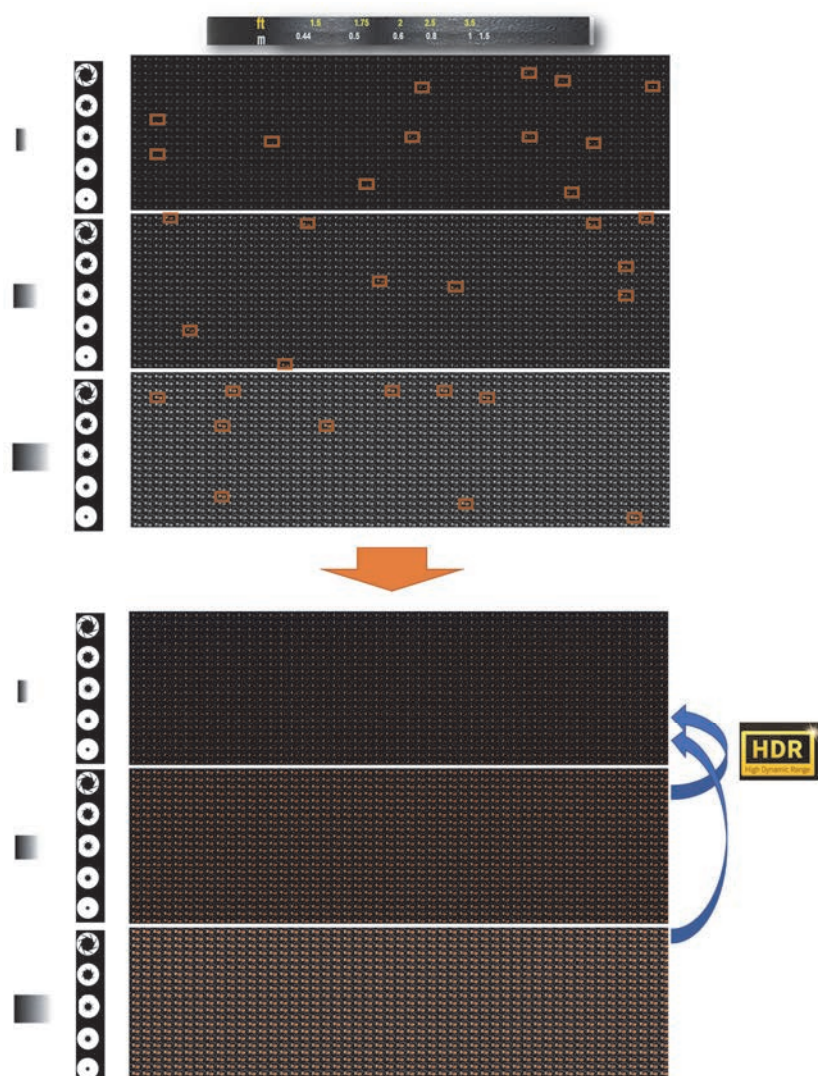


図 4.14 MLAF スタックの再構成の概念図:

16~32 枚の画像を用いて、入力画像に存在しないカメラパラメータで撮像された画像を含む、完全な 3,366 枚の画像で構成される MLAF スタックを再構成する。

再構成された MLAF スタックは、複数の露光レベルで撮像された Aperture-Focus スタックと解釈することが可能である。図 4.14 に示す通り、再構成された画像スタックから、それぞれのフォーカス位置、絞り値において、複数の露光レベルの画像を合成することで HDR 画像の取得が可能である。これは言い換えれば、HDR の Aperture-Focus スタックの再構成が可能である、と言える。

章の冒頭で述べた通り、13 種類のフォーカス位置・絞り値・露光レベルの全パラメータで撮影した画像スタックデータセットを作成し、6 シーンの評価用と、7 シーンの学習用、に分割して用いた。MLAF スタックは、22 の絞り値、51 のフォーカス位置、3 の露光レベル、合計 3,366 枚の画像をそれぞれのシーンについて撮像した。22 の絞り値は F/1.4 から F/16 の範囲で、51 のフォーカス位置は約 0.8m~2m の範囲で変更した。切り出しと間引きサンプリングにより、実験に用いる 960×640 の各画像を準備した。

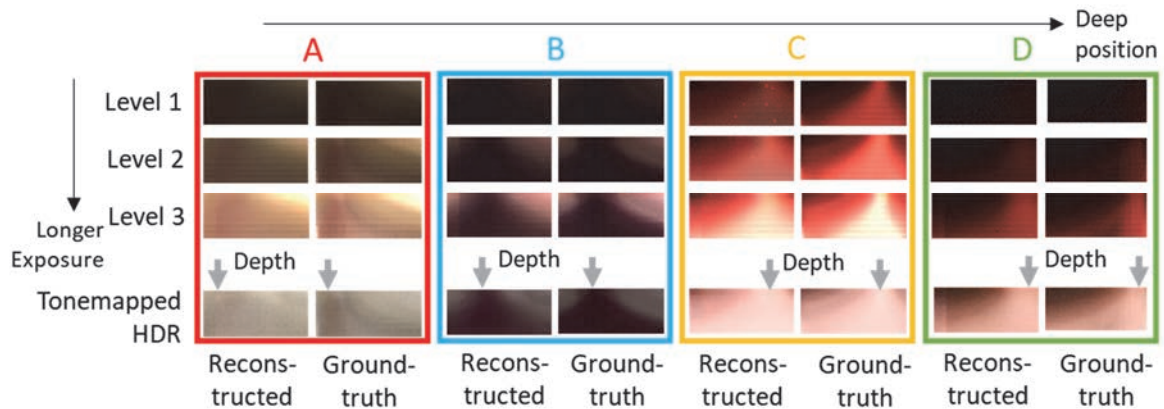
図 4.15 では、MLAF スタックから特定位置の 1 画素を取り出して作成した MLAFI (Multi-level Aperture Focus Image) について、再構成された MLAF スタックから得た MLAFI と、全ての撮像パラメータで実際に撮像された完全な MLAF スタックから得た MLAFI を ground-truth として比較し、視覚的に評価している。図 4.15 に示すシーンは“Christmas”である。これについて、撮像パラメータをグリーディ・アルゴリズムで選ばれた、入力となる 32 枚の画像から、3,366 枚の MLAF スタックを生成している。いくつかピックアップした画素の、再構成画像からの MLAFI と ground-truth (元画像) からの MLAFI との比較を、図 4.15 上でカラーの四角で示した A~D の 4 画素について行った。A~D に進んでいくに従って、近傍から遠方の画素を選んでいる。視覚的な評価から、その品質が高いことが確認できる。また、3 段階の露光レベルが違う、同一フォーカス位置と絞り値の画像から、HDR 合成を実施して、それぞれの画像から得られた AFI 画像（この場合は、HDR の Aperture-Focus スタックが生成されており、1 画素を取り出して作る画像も AFI (Aperture Focus Image) になる）についても、再構成画像と ground-truth と比較している。HDR 画像は、標準 tonemap 処理を行った後に、表示している。各露光レベルの AFI および HDR 合成後の AFI は、非常に近い形状となっており、合焦距離を山頂としたなだらかな形状が確認できる。言い換えれば、Confocal ステレオアルゴリズム[11]によるデプス位置が明確になっていると言える。





Example of reconstructed image

Corresponding ground-truth image



Evaluation of reconstructed AFI quality by comparing to ground-truth

図 4.15 再構成された MLAFI 画像:

シーン” Christmas” における再構成画像の例 (左上), それに対応する実際に撮像された元画像 (右上), MLAF スタックの中から, 上の図の A~D の位置の画素から構成した MLAFI 画像について, 再構成された MLAF スタックから得られた MLAFI と ground-truth の MLAF スタックから得られた MLAFI 画像の比較 (下). 下図において, 各フォーカス位置・絞り値における HDR 画像についての AFI 画像を, 再構成から得られたものと ground-truth から得られたものとで比較して示している. 非常に近い形状となっており, 合焦距離を山頂としたなだらかな形状が確認できる.

図 4.16 では、図 4.15 の再構成画像を示したシーン“Christmas”について、異なるフォーカス位置・絞り値・露光レベルに相当する再構成画像をズームアップして示している。図 4.16 左上に Lev1~3 として灰色のマークで示されたのはグリーディ・アルゴリズムで選択され、入力として用いられた撮像パラメータである。入力画像の撮像に選択されているパラメータは全部で 32 点である。一方、カラーの星点で示されたのが図 4.16 下でズームアップされた再構成画像相当のパラメータである。(図 4.16 下のズームアップ画像の枠の色が、図 4.16 左上の星の色に対応している。) HDR 合成画像については、ground-truth として実際に撮影された MLAF スタックから、対応する撮像パラメータの画像、ズームアップ領域について並べて再構成品質を比較している。図 4.16 左上の横軸がフォーカス位置のインデックス (左⇄右で近⇄遠) であり、縦軸が絞り値 (下⇄上で大⇄小) である。ズームアップした画像相当の撮像パラメータは、選択された入力画像の撮像パラメータの全てから出来るだけ遠いもの、かつ、画質変化の大きい、絞りサイズが出来るだけ大きいもの、を選択している。図 4.16 右上は、シーンの構図全体から図 4.16 下でズームアップした領域の位置を示したものである。距離の近い領域から遠い領域まで 3 段階に選択している。

A, B 点のパラメータでは、絞りは大きく被写界深度の浅い画像でフォーカス位置が前後に異なる。一方, D, E 点では被写界深度は深く、広いフォーカス範囲でボケが少なく鮮明な画質になっている。HDR の再構成画像と ground-truth の元画像との比較から、フォーカスが合致している点の画質も、逆にフォーカスがはずれ大きくぼけている点の画質も視覚的に、解像度の劣化もなく、高品質な再現がされていることが見て取れる。被写体のキャンディの領域では、飽和した画素を有する領域周辺で、飽和形状や色彩に若干 ground-truth からの違いが見て取れる。これは飽和領域における再構成画素に若干の劣化があり、tonemap 処理によって、ボケの形が異なるように見えていると考えられるが、なお自然な画質であることが確認できる。この結果は、本手法が再構成ターゲットである MLAF スタックの全枚数 (3,366 枚) と比べて、1%以下の枚数で大幅に少ない 32 枚の入力であっても、スタック全画像を再現性高く再構成できることを示している。

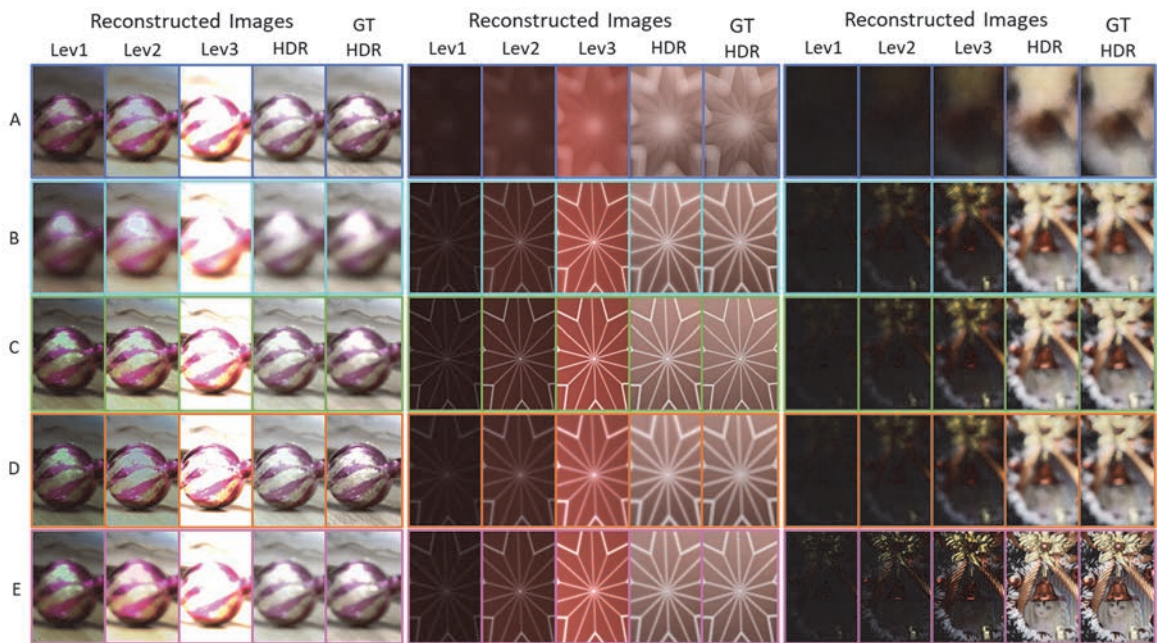
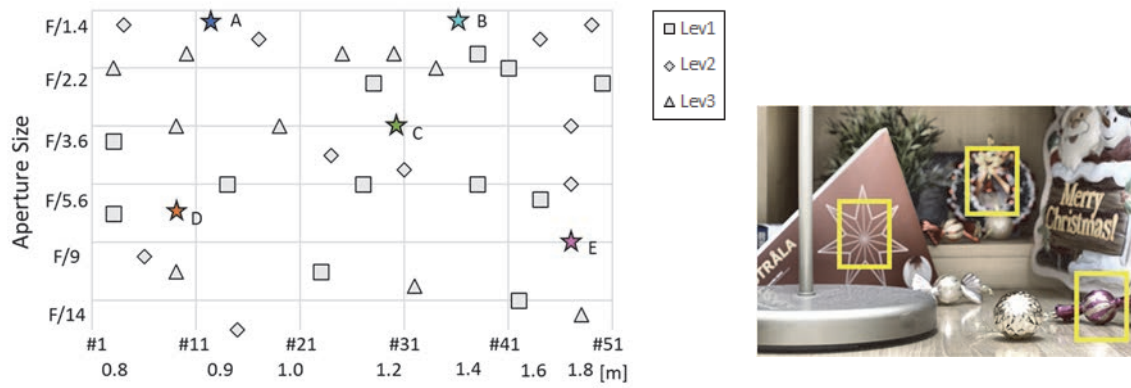


図 4.16 MLAF スタック再構成結果の比較:

シーン” Christmas” 内のいくつかの画像パッチで比較する. (左上) 灰色のマークが画像のパラメータ組合せ. 星点がここで示す再構成画像相当のパラメータ. (右上) シーン中の画像パッチの位置, (下) 各画像パッチの再構成結果と, HDR 合成画像における ground-truth(GT)との比較. サンプルング点から遠い再構成結果も品質が高いことが示された.

図 4.17 では、入力枚数に応じた MLAF スタックの再構成品質の定量評価の結果を示している。3つのデータセットに対し、入力枚数を 8~64 枚のそれぞれに設定し、再構成品質を PSNR および SSIM[74]で評価したものである。PSNR は全般に印象ほど高くない。低露光画像は大きく黒つぶれ領域が存在し、元画像のノイズ量も大きいことが理由と考えられる。一方、PSNR に対しより視覚的印象に近い画質評価指標と言われている SSIM[75]においては、非常に高い数値が得られ、高精度復元の印象を裏付けている。サンプル数は 8 枚から増えるにつれて SSIM は上昇し、3 シーンとも 24 枚または 32 枚周辺において上げ止まる傾向が見てとれた。

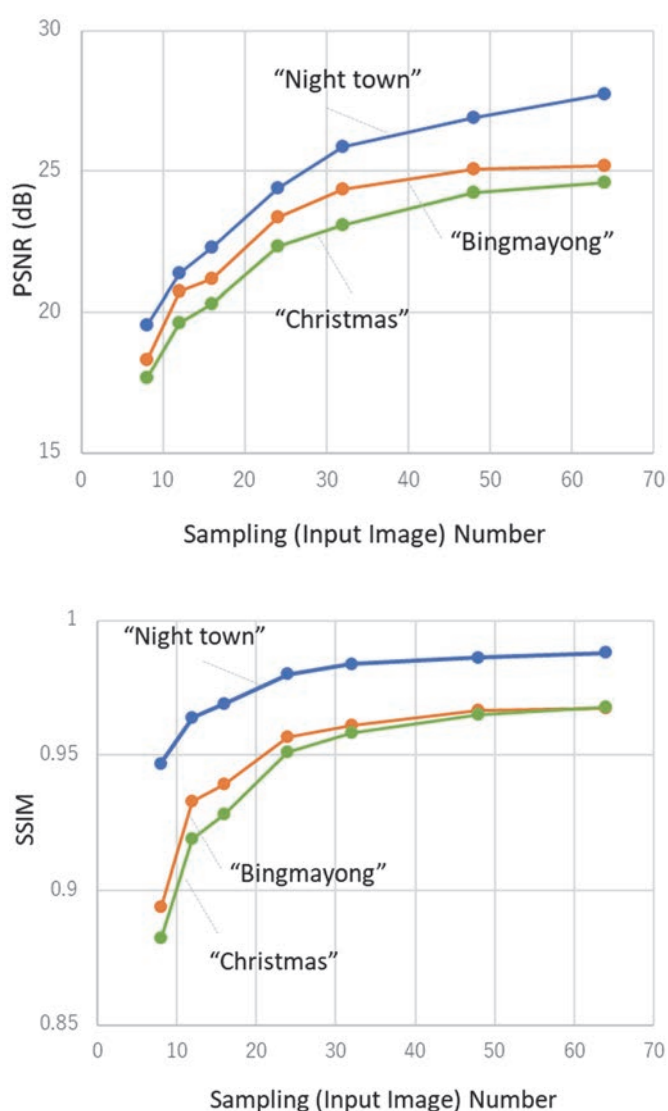


図 4.17 入力サンプリング画像枚数と画質評価値:  
3つのシーンについて PSNR(上)と SSIM(下)にてプロットした結果

図 4.18 は、シーン” Bingmayong”において、復元精度をより詳細に解析したものである。SSIM はその数値を領域マップとして表現できるため、入力サンプル枚数 8 枚と 32 枚の画像で領域的に比較している。絞りが大きく手前にフォーカスを合わせて、全体がぼけた画像を選択しているため、SSIM 平均値が低い。特に 8 枚からの復元は高周波領域における SSIM が低くなっている。図 4.18 下は、ピックアップした高周波領域を含む領域の再構成結果を入力サンプル数ごとに比較している。少ない入力枚数ではぼけが足りず、高周波が残留している印象を受けるが、入力枚数の増加とともに ground-truth のぼけに近づく。このように高周波領域がぼける場合、再構成品質が若干劣化することがわかる。図 4.17 で示した SSIM のシーンによる違いは、こうした高周波領域の面積によるものであると推察できる。

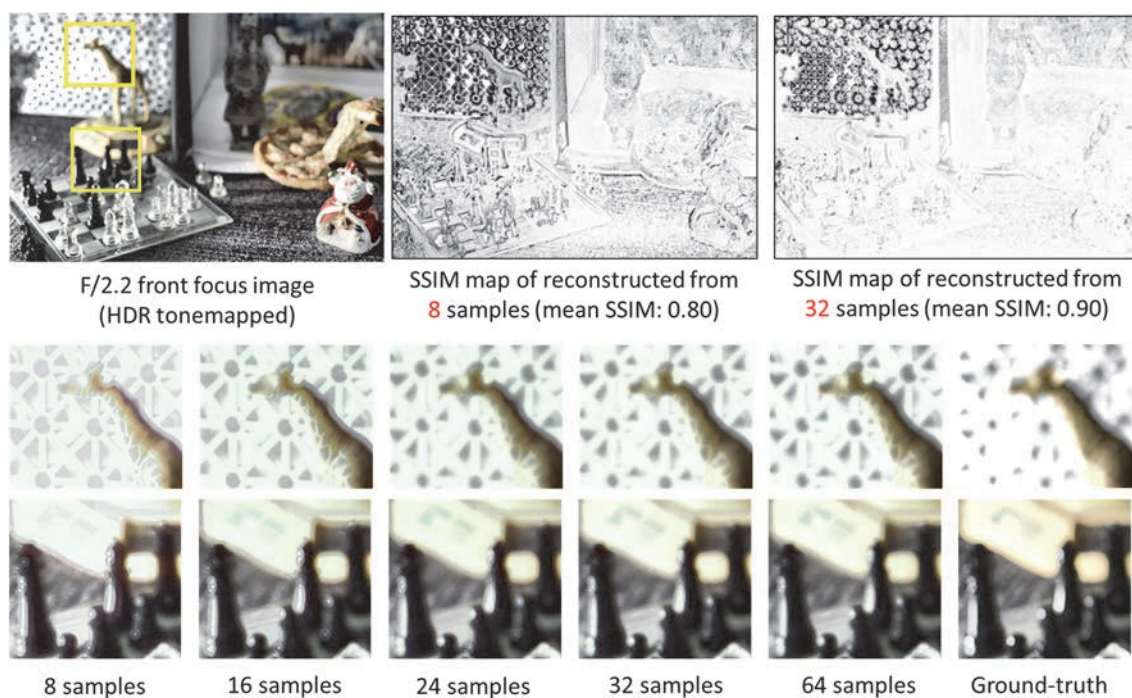


図 4.18 SSIM 値の領域評価:

SSIM マップは  $0 \leftrightarrow 1$  の値で画質の低  $\leftrightarrow$  高を表現している。高周波領域のぼけ復元について、少ない入力サンプル数の場合に画質劣化が見られることが SSIM マップ、再構成画像の双方から確認できる。

図 4.19 は、再構成された画像から HDR を合成し、標準 tonemap 処理して、Aperture-Focus スタックとして並べた例である。大きい絞りサイズの再構成画像において、フォーカス位置から遠い被写体のぼけは自然に再現されており、フォーカス位置が遠くへ動いていくとともに、その量が大きく変化していく様を確認できる。一方で、小さい絞りサイズでは、フォーカス位置が変わっても画像中のぼけの傾向は変わらず全体にくっきりした画像となっている。縦軸では、絞りの大きさが F/1.4 から F/14 へと小さくなっていくにつれ、被写界深度が広がっていく様子が見てとれる。本論の手法によれば、撮影者は撮影後に、こうしたスタックから適した画像を選択することができる。

Aperture-Focus スタックの評価の場合と同様、MLAF スタックの再構成の評価においても、SSIM などによる定量評価は視覚的品質を常に保証するものではないと考え、官能評価実験も合わせて行った。18名の被験者に対し、1枚は実際に撮影された元画像、もう1枚を再構成された画像、の2枚ペアで、左右ランダムに切り替えて side-by-side で表示する(図 4.20 上)。提示するシーンは撮像パラメータが偏らないようまんべんなく選定した。被験者には、「右画像が好ましい」「左画像が好ましい」「差はわからない」の3種類いずれかを回答してもらう。各被験者への回答制限時間は設けず、30セットの比較を行ってもらい、平均で7分ほどを要した。被験者には左右で同じ画像が表示されることがあることを事前に知らせておき、「差はわからない」との回答が妥当な場合が存在することを示唆した。最終的に、 $18 \times 30 = 540$ の比較結果が得られたが、左右に元画像と再構成画像を提示した場合の比較結果は486である。図 4.20 下は得られた結果の統計であるが、再構成画像の方を「好ましい」と25%のケースで回答があり、「差がわからない」と回答した32%と合わせ、半数以上の回答で元画像と同等以上の画質と回答している。本手法の再構成画質の高さを示していると言える。なお、この官能評価実験で用いたデータは、本論の付録で記す通り、MLAF データセット本体とともに、公開している。32枚から3,366枚を再構成した画像群と、当該カメラパラメータで撮影した元画像とを、どちらかがわからないよう side-by-side で並べたデータである。再構成の品質の高さがよくわかるサンプルであるため、是非ともご覧頂きたい。

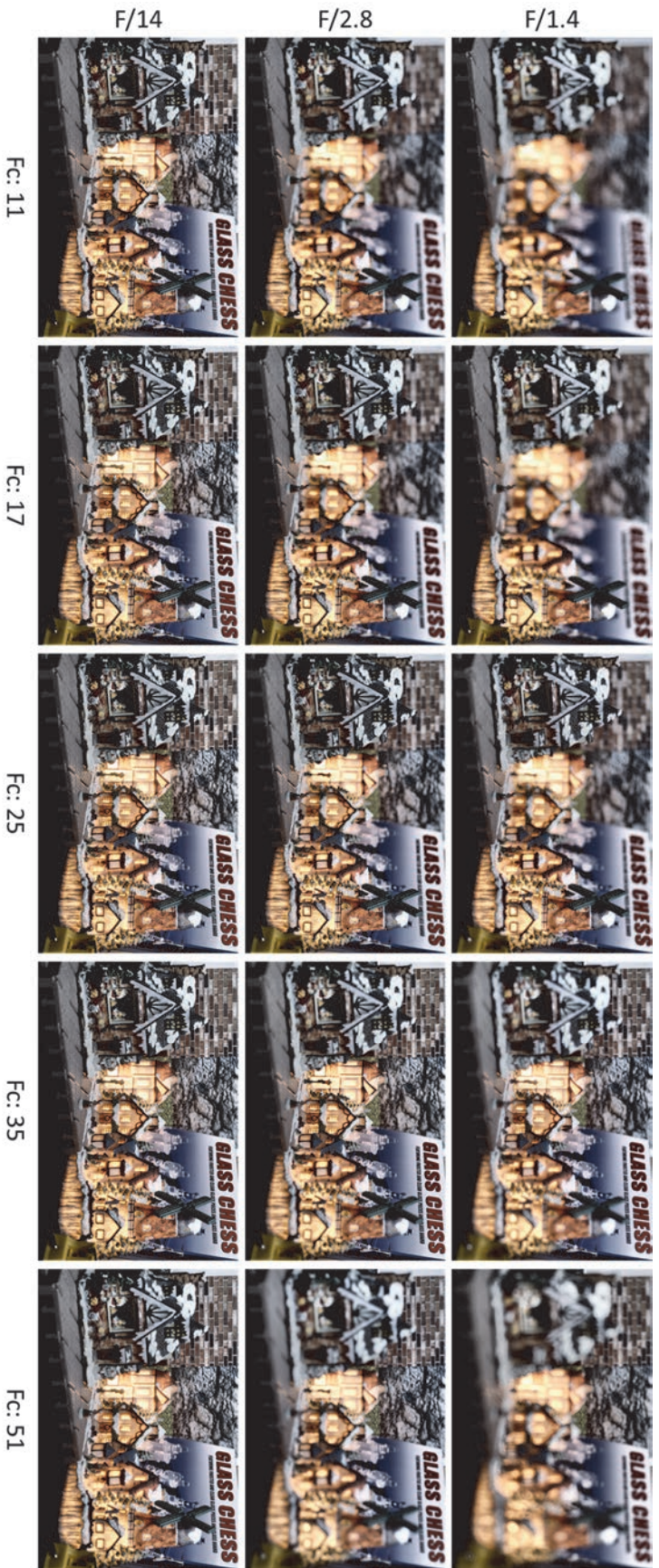
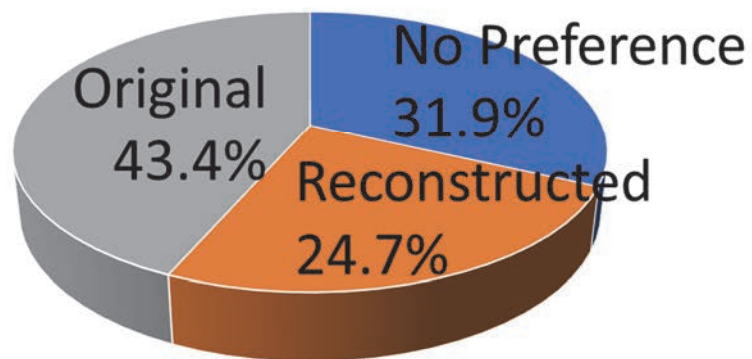


図 4.19 全ての再構成画像から合成した HDR Aperture-Focus スタックの例:

シーン”Night Town”において、横軸は固定絞り値における Focus スタック(左から右にて近から遠)、縦軸は固定フォーカス位置における Aperture スタック(上から下にて大から小)とである。各フォーカス位置・絞り値において異なる露光レベル画像から HDR 画像を合成し、標準 tonemap 処理して表示している。



Example of test image set : right is reconstructed image (subject does not know)



User preference statistics

図 4.20 官能評価実験とその結果:

提示した side-by-side 画像の一例 (上). 本例では左が撮像画像, 右が再構成画像だが, 被験者には教示せず, どちらが好ましい画像かを回答してもらう. 18 人の回答結果の統計値 (下).



# 第5章 アプリケーション

本章では、画像スタックの再構成のアプリケーションとして、高精細な奥行マップを取得する手法 **Confocal Stereo** と、フォーカス、絞り、HDR 表示パラメータを同時かつフレキシブルに、撮影後に選択・制御できることによるレタッチング効果を、いくつかの例とともに示す。

## 5.1 Confocal Stereo

**Confocal** ステレオは、**Aperture-Focus** スタックから得られる各画素位置の **AFI** 画像の特性を用いて、画素ごとのデプス値を高精度に推定することができる強力なデプス推定アルゴリズムである[11, 76]。この手法によれば、画素単位で空間的な平滑化はなく、非常に高解像度なデプスマップを生成可能である。**Aperture-Focus** スタックの全画像の撮像を必要とするため、入力データが極端に大きく、数百枚～数千枚の入力が必要である。

本論の提案による、圧縮 **Epsilon Photography** により、8～16枚の入力で同様の効果が得られることが期待できる。図 5.1 上では、シーン“Hair”における、16枚の画像入力による **Confocal** ステレオデプス推定結果を、**ground-truth** となる 810枚の完全な **Aperture-Focus** スタックを用いた **Confocal** ステレオデプス推定結果と比較して示している。このデプスマップ表現では、輝度値が高いほど遠方の奥行きを持つ画素として表現されている。提案手法は、完全な **Aperture-Focus** スタックから得られたデプスと近い結果が得られていることが目視的に見て取れる。一方、図 5.1 下では、3つのデータセットの結果において、デプス推定誤差のヒストグラムを記している。ヒストグラムの中央値が **ground-truth** から得られた **Confocal** ステレオデプスが、再構成画像スタックから得られた **Confocal** ステレオデプスと比べて、誤差が 0 であった頻度を示しており、その左右に、**Focus** 位置のデジット分の結果デプスとの誤差を、プラスマイナス別に頻度マップ化している。評価した 4シーンともに、再構成のための入力画像が 4枚から 8枚、16枚と増加するにつれて、**ground-truth** から得られた **Confocal** ステレオデプスとの差分が小さくなっていくことが見て取れる。例えば、シーン“Glassball”では、4枚画像を用いた場合のエラーが大きく、8枚画像の入力によって再構成された画像スタックから得られた **Confocal** ステレオデプスから、誤差量が急激に小さくなっていることが見て取れる。これは図 4.7 の再構成画像の **PSNR** の比較において、当該シーンの“Glassball”では、4枚から 8枚へ入力画像を増加した際の **PSNR** の改善値が（16枚、32枚と増加させた際と比べて）大きいことが、反映されていると考える。言い換えれば、画像再構成の精度が、**Confocal** ステレオによるデプスマップの精度に影響を与えていることが考察

される。結果から総じて、810枚の入力によるデプスマップに対して、本手法の8枚以上の入力によるデプスマップの誤差が極めて少ないことを、期待通りに示していると言える。本論の圧縮 *Epsilon Photography* のアプローチにより、わずか8枚の入力画像から、810枚の入力によるデプス推定と比較し、競争力が示された。810枚の画像を必要とした元のアルゴリズムが、大幅な改善が得られたと考える。

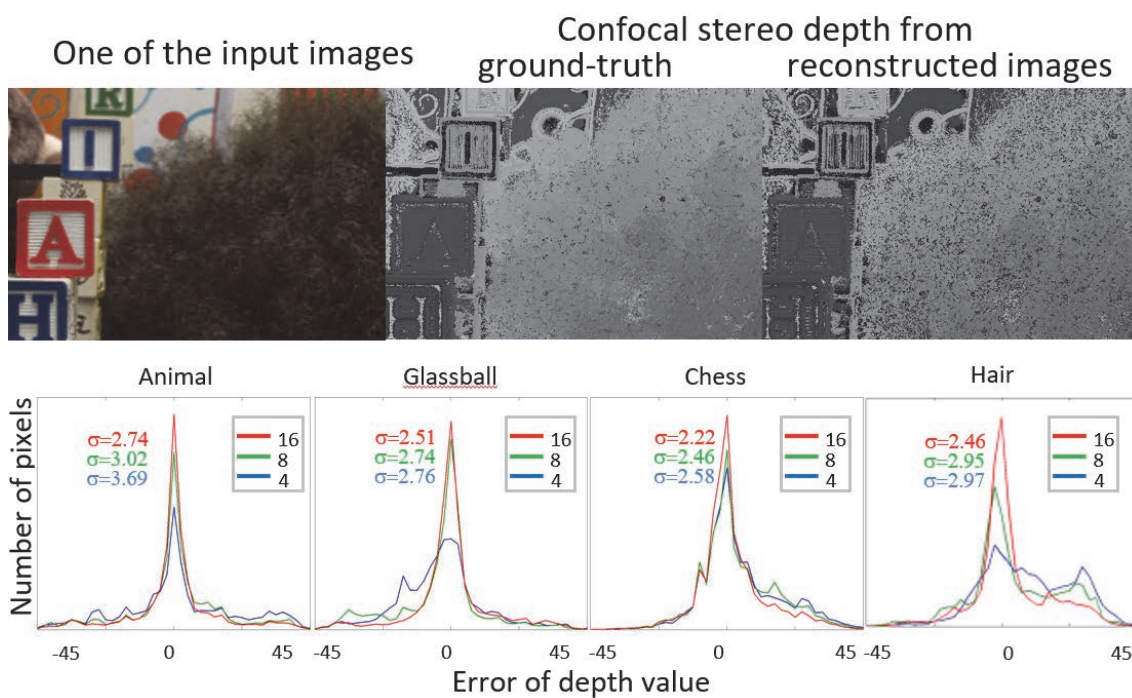


図 5.1 圧縮 Confocal ステレオ:

Aperture-Focus スタックを用いた高解像度のデプス推定手段である Confocal ステレオは、高精度であるが入力データが膨大である。本論において、4~16枚の入力により、このアルゴリズムを実現した。左上が入力画像の内から1枚、中上が ground-truth の Aperture-Focus スタック(810枚)を用いたデプスマップ、右上が再構成スタック(入力は16枚)を用いたデプスマップである。下段は、ground-truth の Aperture-Focus スタックから得たデプスとのエラー値のヒストグラムである。8枚の入力から得られた結果であっても、高品質であることがわかる。

## 5.2 Refocusing ユーザーインターフェース

Light Field 撮像における最も広く知られているアプリケーションは Refocusing であり, Lytro[2]などが提供した, タッチパネルに連動して撮像画像のフォーカス位置を撮影後に編集できるアプリケーションは一般によく知られている[77, 78]. そこで本研究においても, 再構成後の Aperture-Focus スタックを用いた Refocusing のユーザーインターフェースを, Matlab を用いて実装した. 図 5.2, 5.3, 5.4 に3つのシーンでの操作例を示す. 各図の(a)~(c)に示す通り, タッチした画素にフォーカスが合うよう, ユーザーインターフェースを設計した. さらに, 各図の(d)~(f)に示す通り, 右下のスライダーで絞り値の変更が可能である. 絞りのサイズは, 同心円の図として右下に明示している. タッチした画素のフォーカス位置の算出は 5.1 章で説明した Confocal ステレオで算出したデプスマップを用いた. 本論が提案する手法は画素単位の処理であり, 原理的に解像度が低下しない. 高解像度での Refocusing 体験の提供が可能となる.

各図の(a)~(c)に示す, 絞りサイズが大きい状況でのフォーカス位置の操作により, 奥行きに応じた自然なぼけが生成され, 立体感のある画像が得られている. Refocusing の効果はそれだけではなく, カメラのオート撮像機能が選択するカメラパラメータをユーザの好みに応じて生成できる効果も強調できる. 例えば, 図 5.4 の構図では, 一般的なオートフォーカス機能では, 中央の面積の大きな被写体にフォーカス位置が選択されるため, (a)図で示すようなキリンやライオンの位置にフォーカス位置が合うのが一般的だと考えられる. 一方, (c)図で示すような, 背景で建物から出てくる人にフォーカス位置を合わせたいというユーザの要求があるとすれば, オートフォーカスの選択したパラメータでは満たされない. 好みや文脈に応じてユーザの意図するカメラパラメータは異なり, オート機能が選択したパラメータではなくても, 失敗した撮像ではない. 撮像後の Refocusing 機能により, こうした効果が提供できる. また, 各図の(d)~(f)に示す画像にも, パラメータの設定に好みや文脈を反映することができるであろう. 各図の(f)で示すような, 手前から後ろまでフォーカスが合った, 絞りサイズの小さい画像は, 観察者がどこを見ても鮮明に感じられる. 一方で, 絞りサイズの大きい画像, 例えば図 5.4 の(d)では, おそらく観察者は, フォーカスが合っているシマウマか葉っぱに注目するであろう. 絞りサイズの大きい, 奥行き感のある画像では, 観測者に注目被写体を誘導する効果があると言え, Refocusing 機能により, 撮影後に観察者の注目被写体を操作することがとなる.

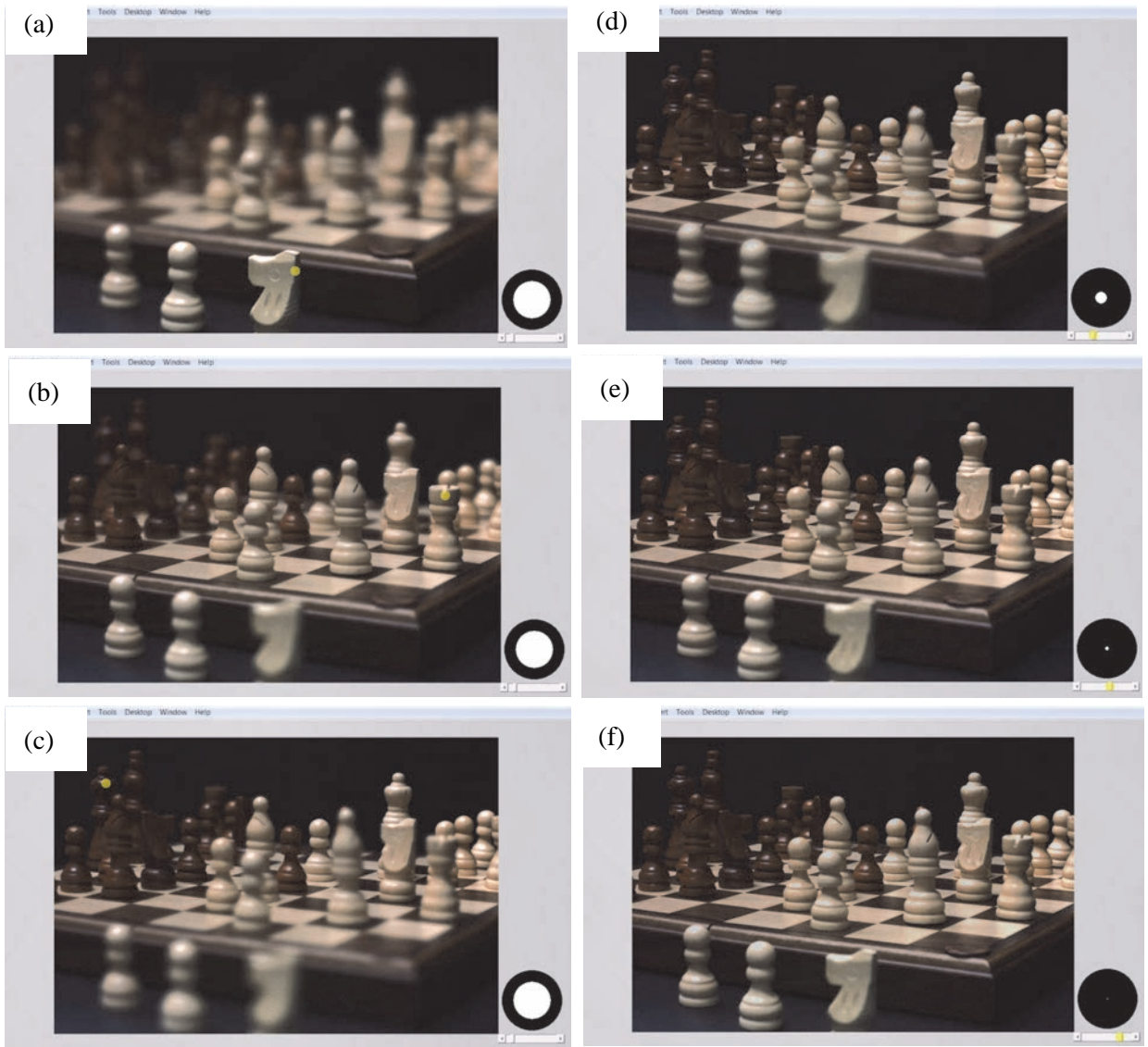


図 5.2 Refocus ユーザーインターフェースの例: シーン “Chess”



図 5.3 Refocus ユーザーインターフェースの例: シーン “Glassball”

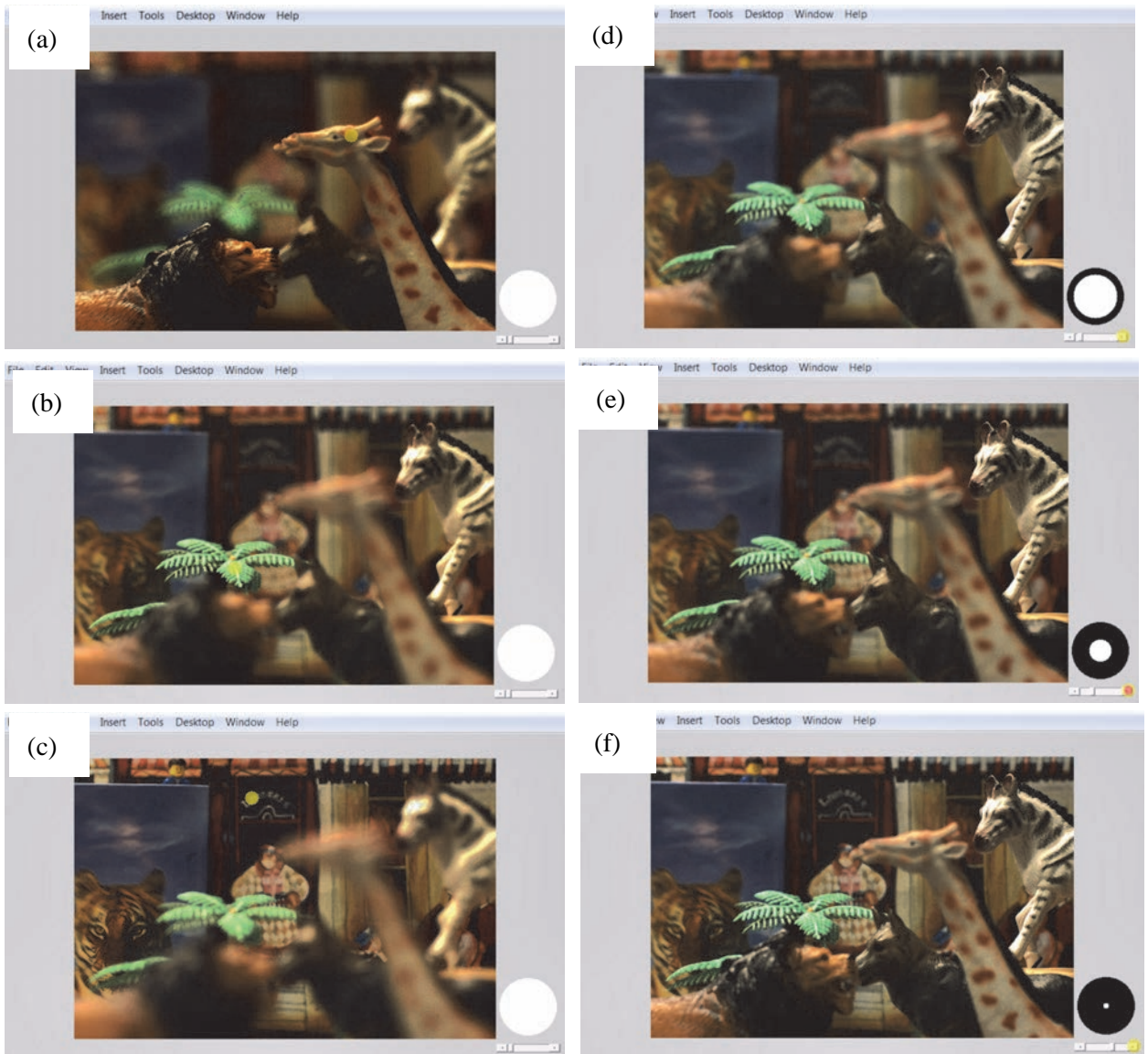


図 5.4 Refocus ユーザーインターフェースの例: シーン “Animals”

### 5.3 Refocusing + HDR 画像の露出調整機能

撮影後に合成された HDR 画像に対して、tonemap 処理のみならず、フォーカス位置や絞り値の調整といった Refocusing の適用を同時にできることは、単純な組み合わせ以上に、使用者に大きな画作りの自由度を提供できる[68, 79, 80, 81].

図 5.5 では、再構成されたシーン”Horse”の MLAF スタックを用いて、フォーカス位置一定で、絞り値を変えた画像を、いくつかの異なる tonemap 処理を実施して作成し、提示している。フォーカス位置は常に金色の馬に合わせており、縦軸で絞り値を変更しながら、横軸では、HDR 合成後に画面一様の tonemap 関数を各々に変更している。横軸では、default の線形値 ( $\gamma=1.0$ ) である図 5.5 a) に対し、b), c) では HDR 情報を利用して、背景暗所の被写体も見えるよう画作りしている一方、d), e) では背景を黒つぶれさせて前景の鏡面反射している被写体を浮き出させる画作りをしている。さらに e) は、背景の明所も飽和させずに元形状が視認できるままぼかしているのに対し、d) ではぼけと同時に明所は飽和させており、白飛びして元形状を見えにくくしている。これらも、ユーザの好みに応じて撮影後に調整が可能である。見てきたように、絞りサイズと露出制御の組み合わせで、背景明部のぼけと飽和の印象も様々に変えられることも見て取れる。各フォーカス・絞りで HDR が合成できるため、撮影後にこのようなぼけと飽和の組み合わせ制御が可能となる。繰り返しであるが、このような効果を自在に得るためには本来、3,366 枚の撮像が必要であるが、本論の提案手法によれば、その 1% に満たない、16 ~ 32 枚の撮像によりこうした効果を得ることができる。

図 5.6, 5.7 では、MLAF スタックを用いた、撮影後のフレキシブルな撮影後制御の効果を、例を挙げて示している。図 5.6 a) は、逆光効果を逆利用した画作りであり、大きな絞り値で前景をぼかしながら、露光レベルも背景に合わせている。b) では小さい絞り値で前景から背景までフォーカスを合わせつつ、露光レベルも全体をみやすく tonemap 処理しており、c) では b) の画像を大きな絞り値に変えているため、フォーカスを合わせている被写体の右目から遠い被写体をぼかしている。さらに d) では c) から tonemap 関数を変更し、背景をぼかすのみでなく白飛びさせて、被写体を浮き出させている。図 5.7 においても同様に、得られた HDR 画像を撮影後にフォーカス・絞り・tonemap の同時調整を行っている。a) は線形の  $\gamma$  値であったのに対し、b) は  $\gamma=0.33$  とし、かつ、絞り値を最小にすることで後ろから前まで鮮明で視認性の高い明るさとしており、c) では手前にフォーカスと露出を合わせ、d) ではシーン奥にフォーカスと露出を合わせることで手前は飽和しつつぼかしている。本手法が、実際に撮影していないフォーカス位置の画像を、大きな絞り値であっても信頼性高く HDR で再構成できるため、こうした撮影後の選択・制御が可能となる。

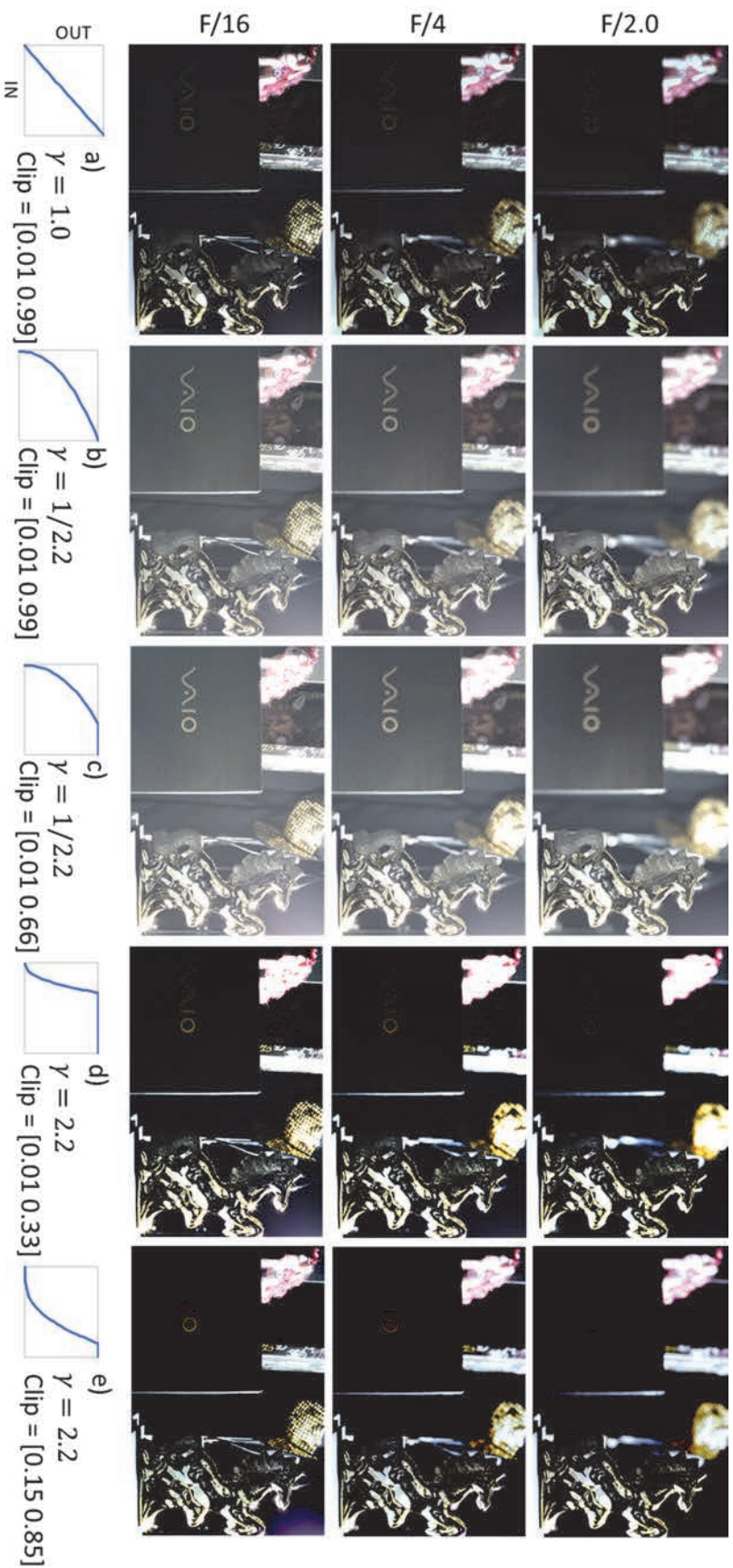


図 5.5 撮影後の絞り値と露出の同時制御の例:

シーン”Horse”において，横軸には HDR 合成後に異なる **tonemap** 処理をして並べており，縦軸は固定フォーカス位置における Aperture スタック(上台下にて大台小)となっている．**tonemap** 関数は画面一様であり，ガンマ係数と明暗のクリップ値を横軸の各々に示している．ユーザの好みに応じ，撮影後にぼけと飽和の組み合わせ制御ができる．





a) Focus on Background, Aperture F/1.4,  $\gamma = 1.6$ , Clip = [0.01 0.99]



b) Focus on Light Eye, Aperture F/16,  $\gamma = 0.33$ , Clip = [0.01 0.99]



c) Focus on Light Eye, Aperture F/1.4,  $\gamma = 0.33$ , Clip = [0.01 0.99]



d) Focus on Light Eye, Aperture F/1.4,  $\gamma = 0.45$ , Clip = [0.01 0.33]

図 5.6 再構成画像スタックを用いた撮影後のカメラ制御の例:  
シーン” Backlight” において, 撮影後にフォーカス位置・絞り値・HDR tonemap を  
様々に調整した結果. ぼけの量や位置, 白飛び・黒つぶれ領域をフレキシブルに変更・制御できる.



a) Focus on middle, Aperture F/2.8,  $\gamma = 1.0$ , Clip = [0.01 0.99]



b) Focus on middle, Aperture F/16,  $\gamma = 0.33$ , Clip = [0.01 0.99]



c) Focus on front, Aperture F/1.4,  $\gamma = 0.63$ , Clip = [0.01 0.99]



d) Focus on background, Aperture F/1.4,  $\gamma = 0.5$ , Clip = [0.01 0.5]

図 5.7 再構成画像スタックを用いた撮影後のカメラ制御の例:  
シーン” Christmas” において, 撮影後にフォーカス位置・絞り値・HDR tonemap を  
様々に調整した結果. ぼけの量や位置, 白飛び・黒つぶれ領域をフレキシブルに変  
更・制御できる.

# 第6章 結論

本章では本論文を総括し、今後の課題と展望について述べる。

## 6.1 総括

本論では、計画された撮影パラメータ（フォーカス・絞り・露光レベル）で撮影された8~32枚程度の画像から、入力100倍以上、数千枚レベルの異なるパラメータで撮影された相当の画像スタックを再構築した。アプローチにおける重要なポイントは、フォーカス・絞り・露光レベルを変更したときの画素値変化が低次元で疎なデータであることに着目し、それを定量的に確認した上で、GMM（混合ガウシアンモデル）を用いてモデル化した点である。入力画像の選択は、グリーディ・アルゴリズムにより学習時に最適化する手段を提供した。非常に高品質な画像の再構成を実現し、実画像データセットを用いた定性・定量評価により効果を確認した。また、提案手法による完全なAperture-Focusスタックの復元により、従来は数千枚の画像により実現されたConfocalステレオのようなアプリケーションを、限られた入力枚数で実現できることを示した。さらに、限られた枚数の入力により復元されるデータには、希少な高解像度なHDRのAperture-Focusスタックが含まれ、フォーカス位置・絞りサイズ、露出制御を、撮影後にフレキシブルな編集を実現する。データセットの公開共有も本論の貢献である。

## 6.2 課題と展望

課題：

主なものは、動被写体への対応である。本論で示している実験結果は、静止または非常に遅い動きの被写体に限定される。しかしながら、本手法で用いる入力画像は限定されており、カメラの連写撮影機能を用いて、オプティカルフローなどを用いた位置合わせ処理を考慮することで、将来的には動被写体への適用も期待できる。[82]や[83]では、HDR撮影向けに露光期間が異なる画像間の位置合わせや、画像内で異なるぼけを含む画像間の位置合わせ手法を提案しており、こうした手法を組み合わせることは有望である。現状用いているPer-Pixelアプローチにおいては、光学的な想定を逸脱した現象、例えば、フレアなど回折による諸々の影響[84]を含んだ復元に限界があるため、空間的近傍を含めたモデル化も、有効である。

また、本論で限定枚数での再構成を実現したとはいえ、実用化に向けてはいくつかの

課題がある。撮像枚数をより減らすこと、学習データ生成に手間がかかること、レンズの個体バラつきに対する検証が不十分であること、が挙げられる。これらに対するアプローチは下記の展望で述べる。

展望：

近年の Learning Camera の知見を取り込むことで、本論の GMM をベースとした提案手法よりも少ない入力枚数で同等画質を復元できる可能性に期待ができる。近年の研究においては、Deep Learning の恩恵を受け、1枚画像からの Refocus 機能の実現[85]、HDR 画像の復元[18, 19]、Light Field の復元本論[20, 21]、または特殊な光学系を用いた撮像データからレンズそのものの再現[86]、といった問題を、学習ベースで解く試みがなされている。本論のデータセット公開が、Epsilon Photography の再構成問題を Deep Learning で解く研究を数多く誘発する、その一助とれば幸いである。また、学習データ生成やレンズの個体バラつきに対する対応も重要な Future work である。本論では実撮影で多くの画像を収集し、レンズ個体の特性に応じたキャリブレーションを実施したが、別個体への同一キャリブレーションパラメータの適用によるパフォーマンスは未評価である。近年、機械学習の分野では、光学要素を考慮した Synthetic データを学習データとして生成、活用する試みが数多くなされ、成果を上げている[87, 88]。これは、機械学習アルゴリズムそのものが学習データセットのドメイン間での統計量の差を吸収できるようになった[89, 90]こと、および、Rendering 技術[91]などの発展により、Synthetic データそのものの質が向上したことの両側面によるものと分析できる。こうした技術と融合することで、本論の提案技術の頑健性を向上し、実用化を目指していく。

### 6.3 おわりに

本論では、銀塩カメラからデジタルカメラへの切り替えが進んで撮影の利便性は向上しても未だ残る、フォーカス位置や絞りのサイズ、シャッター速度や ISO 感度など、カメラのパラメータを撮影時に調整しなければならず、撮影後に制御できる範囲は限られる、という課題の解決に取り組んだ。完全に失敗しない撮影、ユーザの好みに応じた完全な撮影後の画像制御に向け、現状では課題は残るものの、撮像センサの感度向上や高速化の恩恵を得ながら、理想の撮像に近づけていく、その一歩となることを望むものである。課題と展望に記した通り、まだまだ多くの改善アプローチが存在する。これらをひとつひとつ解決することで、デジタルイメージングの新たな時代を切り拓くべく、研究開発に尽力していきたい。

## 参考文献リスト

- [1] M. Levoy and P. Hanrahan, "Light field rendering," Proc. ACM SIGGRAPH, pp.31–42, 1996.
- [2] LYTRO. "The lytro camera," <https://www.lytro.com/>.
- [3] RAYTRIX. "3d light field camera technology," <http://www.raytrix.de/>.
- [4] P. Green, W. Sun, W. Matusik, and F. Durand, "Multi-aperture photography," ACM Trans. Graphics 26, 3, 68, 2007.
- [5] T. Georgiev, A. Lumsdaine and S. Goma, "High Dynamic Range Image Capture with Plenoptic 2.0 Camera," Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest, SWA7P, 2009.
- [6] R. Raskar, "Computational photography: Epsilon to coded photography," Emerging Trends in Visual Computing, Springer, 238–253, 2009.
- [7] S. K. Nayar and Y. Nakagawa. "Shape from focus," IEEE Trans. on Pattern Analysis and Machine Intelligence, 16(8):824-831, 1994.
- [8] A. Levin and F. Durand. "Linear view synthesis using a dimensionality gap light field prior," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 1831–1838, 2010.
- [9] S. Kuthirummal, H. Nagahara, C. Zhou and S.K. Nayar, "Flexible depth of field photography," IEEE Trans. Pattern Analysis and Machine Intelligence 33(1), 58–71, 2011.
- [10] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," Proc. ACM SIGGRAPH, 369–378, 1997.
- [11] S. Hasinoff, and K. Kutulakos, "Confocal stereo," Int. Journal of Computer Vision 81, 1, 82–104, 2009.
- [12] M. Brown and D.G. Lowe, "Automatic panoramic image stitching using invariant features," Int. Journal of Computer Vision 74, 1, 59–73, 2007.
- [13] N. Joshi and M. F. Cohen, "Seeing Mt. Rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal," In IEEE Intl. Conf. Computational Photography, 1–8, 2010.
- [14] N. M. Law, C. D. Mackay and J. E. Baldwin, "Lucky Imaging: High angular resolution imaging in the visible from the ground," arXiv preprint astro-ph/0507299, 2005.
- [15] CASIO, "Exilim EX-10," [http://arch.casio.jp/dc/products/ex\\_10/](http://arch.casio.jp/dc/products/ex_10/)
- [16] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," ACM Transactions on Graphics (TOG), 36, 1-15, 2017.
- [17] C.A. Metzler, H. Ikoma, Y. Peng and G. Wetzstein, "Deep Optics for Single-Shot High-Dynamic-Range Imaging," Proc. IEEE Conf. Computer Vision and Pattern

Recognition, 1375–1385, 2020.

- [18] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai and Y. Liu, “Light Field Reconstruction Using Deep Convolutional Network on EPI,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2017.
- [19] N. K. Kalantari, T. C. Wang and R. Ramamoorthi, “Learning-based view synthesis for light field cameras,” ACM Transactions on Graphics (TOG), 35(6):193, 2016.
- [20] P. Sakurikar, I. Mehta, V. N. Balasubramanian, P. J. Narayanan, “RefocusGAN: Scene Refocusing Using a Single Image.” Proc. IEEE European Conference on Computer Vision (ECCV), 2018.
- [21] S. B. Dayan, D Mendlovic and R. Giryes, "Deep Sparse Light Field Refocusing," In Proc. of The British Machine Vision Conference (BMVC), 2020.
- [22] R. Baraniuk, “Compressive sensing,” IEEE Signal Processing Magazine 24, 4, 118–121, 2007.
- [23] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” IEEE Signal Processing Magazine 25, 2, 21–30, 2008.
- [24] A. Bourrier, R. Gribonval, and P. Perez, “Compressive gaussian mixture estimation,” Proc. IEEE Acoustics, Speech and Signal Processing, 2013.
- [25] T. Buades, Y. Lou, J. M. Morel and Z. Tang, “A note on multi-image denoising,” Proc. A Intl. Workshop on Local and Non-Local Approximation in Image Processing, 1-15, 2009.
- [26] L. Zhang, S. Vaddadi, H. Jin, S. K. Nayar, "Multiple view image denoising," In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2009.
- [27] D. Capel and A. Zisserman, “Super-resolution from multiple views using learnt image models,” IEEE Conf. on Computer Vision and Pattern Recognition, 2001.
- [28] F. Sroubek and P. Milanfar, “Robust multichannel blind deconvolution via fast alternating minimization,” IEEE Trans. on Image Processing, 21(4):1687-1700, 2012.
- [29] L. Yuan, J. Sun, L. Quan and H. Y. Shum, “Image deblurring with blurred/noisy image pairs,” In ACM Trans. of Graphics, Vol 26, 1, 2007.
- [30] S. Mann and R. Picard, “Being undigital with digital cameras,” MIT Media Lab Perceptual, 1994.
- [31] S. W. Hasinoff, F. Durand and W. T. Freeman, “Noise-optimal capture for high dynamic range photography,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.
- [32] S. W. Hasinoff and K. N. Kutulakos, “A layer-based restoration framework for variable-aperture photography,” Proc. IEEE Intl. Conf. Computer Vision, 2007.
- [33] P. Favaro, A. Mennucci and S. Soatto, “Observing shape from defocused images,” Intl. Journal of Computer Vision (IJCV), 52(1), 25-43, 2003.



- [34] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin and M. Cohen, “Interactive digital photomontage,” In *ACM Trans. of Graphics*, vol 23, 294-302, 2004.
- [35] J. G. McNally, T. Karpova, J. Cooper and J. A. Conchello, “Three-dimensional imaging by deconvolution microscopy,” *Methods* 19(3), 373–385, 1999.
- [36] J. B. Sibarita, “Deconvolution microscopy,” In *Microscopy Techniques*. 201–243, 2005.
- [37] E. Krotkov, “Focusing,” *Intl. Journal of Computer Vision (IJCV)*, vol. 1, 3, 223-237, 1988.
- [38] P. Grossmann, “Depth from focus,” *Pattern Recognition Letters* vol. 5, 1, 63–69, 1987.
- [39] S. Gortler, R. Grzeszczuk, R. Szeliski and M. Cohen, “The lumigraph,” In *SIGGRAPH*, 43–54, 1996.
- [40] T. Georgiev and A. Lumsdaine, "Superresolution with plenoptic 2.0 cameras", *Signal Recovery and Synthesis. Meeting of the GSA*, 2009.
- [41] V. Boominathan, K. Mitra and A. Veeraraghavan, “Improving Resolution and Depth-of-Field of Light Field Cameras Using a Hybrid Imaging System,” *IEEE Intl. Conf. on Computational Photography (ICCP)*, 2014.
- [42] M. S. K. Gul, T. Wolf, M. Bätz, M. Ziegler and J. Keinert, “A High-Resolution High Dynamic Range Light-Field Dataset with an Application to View Synthesis and Tone-Mapping,” *Proc. IEEE Int. Conf. on Multimedia & Expo Workshops*, 2020.
- [43] K. Marwah, G. Wetzstein, Y. Bando and R. Raskar, “Compressive light field photography using overcomplete dictionaries and optimized projections,” *ACM Trans. Graphics* 32, 4, 46, 2013.
- [44] S. Tambe, A. Veeraraghavan and A. Agrawal, “Towards motion-aware light field video for dynamic scenes,” *IEEE Intl. Conf. Computer Vision (ICCV)*, 2013.
- [45] P. Peers, D. K. Mahajan, B. Lamond, A. Ghosh, W. Matusik, R. Ramamoorthi and P. Debevec, “Compressive light transport sensing,” *ACM Trans. Graphics* 28,1, 3, 2009.
- [46] J. Y. Park and M. B. Wakin, “A multiscale framework for compressive sensing of video,” In *Picture Coding Symposium*, 1–4, 2009.
- [47] A. Wagadarikar, R. John, R. Willett and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging,” *Applied optics* 47, 10, B44–B51, 2008.
- [48] D. Baron, S. Sarvotham and R. G. Baraniuk, “Bayesian compressive sensing via belief propagation,” *IEEE Trans. Signal Processing* vol. 58, 1, 269–280, 2010.
- [49] G. Yu, G. Sapiro and S. Mallat, “Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity,” *IEEE Trans. Image Processing* 21, 5, 2012.

- [50] J. Yang, X. Yuan, X. Liao, P. Llull, G. Sapiro, D. J. Brady and L. Carin, "Gaussian mixture model for video compressive sensing," Proc. of Intl. Conf. on Image Processing, 2013.
- [51] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocusing using a gmm light field patch prior," Proc. of IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW), 2012.
- [52] K. Mitra, O. Cossairt and A. Veeraraghavan, "Can we beat hadamard multiplexing? data driven design and analysis for computational imaging systems," In IEEE Intl. Conf. on Computational Photography, 2014.
- [53] A. Ito, S. Tambe, K. Mitra, A. C. Sankaranarayanan and A. Veeraraghavan, "Compressive Epsilon Photography for Post-Capture Control in Digital Imaging," ACM Transactions on Graphics 33(4):1-12, 2014.
- [54] 伊藤 厚史, 青木 義満, "圧縮 Epsilon Photography に基づく任意の HDR Refocus 画像再構成," 電子情報通信学会和文論文誌 D, Vol.J104-D,No.10, 2021.
- [55] Adobe. "Adobe DNG Software Development Kit (SDK)," [https://www.adobe.com/support/downloads/dng/dng\\_sdk.html](https://www.adobe.com/support/downloads/dng/dng_sdk.html)
- [56] 吉田道隆, 長原一, 鳥居秋彦, 奥富正敏, 谷口倫一郎, "Deep learning による圧縮ビデオセンシングの再構成," 情報処理学会報告 2017-CVIM-208 (19), 2017.
- [57] M. Yoshida, A. Torii, M. Okutomi, K. Endo, Y. Sugiyama, R. Taniguchi and H. Nagahara, "Joint optimization for compressive video sensing and reconstruction under hardware constraints," In Proc. of the European Conf. on Computer Vision (ECCV), 2018.
- [58] Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga and S.K. Nayar, "Video from a Single Coded Exposure Photograph using a Learned Over-Complete Dictionary," In Proc. of IEEE International Conference on Computer Vision (ICCV), 2011.
- [59] H. Nagahara, Toshiki Sonoda, Kenta Endo, Y. Sugiyama and R. Taniguchi, "High-speed imaging using CMOS image sensor with quasi pixel-wise exposure," Proc. of IEEE International Conference on Computational Photography (ICCP), 2016.
- [60] A. Gupta, P. Bhat, M. Dontcheva, O. Deussen, B. Curless and M. Cohen, "Enhancing and experiencing spacetime resolution with videos and stills," Proc. of Intl. Conf. on Computational Photography (ICCP), 2009.
- [61] E. Van Den Berg and M. P. Friedlander, "Probing the Pareto Frontier for basis pursuit solutions", SIAM J. Scientific Comput., vol. 31, no. 2, pp. 890-912, 2008.
- [62] Y. Pati, R. Rezaifar and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition", Proc. Conf. Rec. 27th Asilomar Conf. Signals Syst. Comput., pp. 40-44, 1993.
- [63] Y. Yagi, K. Takahashi, T. Fujii, T. Sonoda, H. Nagahara: "Designing Coded Aperture

Camera Based on PCA and NMF for Light Field Acquisition", IEICE Transactions on Information and Systems, Vol. E101-D, No. 9, pp. 2190--2200, 2018.

[64] J. T. Flam, S. Chatterjee, K. Kansanen and T. Ekman, "On MMSE Estimation: A linear model under gaussian mixture statistics," IEEE Trans. Signal Processing 60, 7, 3840–3845, 2012.

[65] Matlab. "makehdr," <https://www.mathworks.com/help/images/ref/makehdr.html>

[66] F. Drago, W. L. Martens, K. Myszkowski, and N. Chiba "Design of a tone mapping operator for high-dynamic range images based upon psychophysical evaluation and preference mapping", Proc. SPIE 5007, Human Vision and Electronic Imaging VIII, 2003.

[67] Matlab. "tonemap," <https://www.mathworks.com/help/images/ref/tonemap.html>

[68] J. Kim, R. Horstmeyer, I. J. Kim and R. Raskar, "Highlighted depth-of-field photography," ACM Transactions on Graphics, Vol. 30 (3), 2011.

[69] Canon, "Canon EOS-40D," <https://global.canon/ja/c-museum/product/dslr795.html>

[70] Canon, "EF50mm F1.4 USM," <https://cweb.canon.jp/ef/lineup/standard/ef50-f14/>

[71] R. Yokoya and S. K. Nayar, "Extended Depth of Field Catadioptric Imaging Using Focal Sweep," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019.

[72] Helicon Soft, "Helicon focus," <http://www.heliconsoft.com/heliconfocus.html>

[73] R. K. Mantiuk, A. Tomaszewska and R. Mantiuk, "Comparison of Four Subjective Methods for Image Quality Assessment," Computer Graphics Forum, Vol.31(8), pp 2478–2491, 2012.

[74] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Processing, 13(4):600-612, 2004.

[75] A. Hore and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," Proc. of the Intl. Conf. on Pattern Recognition (ICPR), page 2366–2369, 2010.

[76] Q. Chen, Y. Zhang, X. Cao, Y. Zhang and H. Xiong, "Depth map estimation with 4D light fields using confocal stereo," Prof. of 2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016.

[77] J. Bolan, E. Hall, C. Clifford and B. Thurow, "Light-Field Imaging Toolkit," Software X, Vol. 5, pages 101 – 106, 2016.

[78] W. Fu, X. Tong, C. Shan, S. Zhu and B. Chen, "Implementing light field image refocusing algorithm," Proc. of IEEE International Conference on Acoustics, Speech and Signal (ICASS), 2013.

[79] D. Walton, R. Anjos, S. Friston, D. Swann, K. Aksit, A. Steed and T. Ritschel, "Beyond blur," ACM Transactions on Graphics, 40:4, pages 1-14, 2021.

- [80] J. Abbott and B. Morse, “Interactive Depth-Aware Effects for Stereo Image Editing,” *Proc. of Intl. Conf. on 3D Vision*, pp. 263–270, 2013.
- [81] L. Murmann, A. Davis, J. Kautz and F. Durand, “Computational bounce flash for indoor portraits,” *ACM Transactions on Graphics*, 35:6, pages 1-9, 2016.
- [82] M. Gupta, D. Iso and S.K. Nayar, “Fibonacci Exposure Bracketing for High Dynamic Range Imaging,” *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2013.
- [83] H. Zhang and L. Carin “Multi-shot Imaging: Joint Alignment, Deblurring, and Resolution-Enhancement,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [84] R. Kingslake, “Optics in Photography,” SPIE Publications, 1992.
- [85] X. Yang, L. Huang, Y. Luo, Y. Wu, H. Wang, Y. Rivenson and A. Ozcan, "Deep learning-based virtual refocusing of images using an engineered point-spread function," *CoRR abs/2012.11892*, 2020.
- [86] S. S. Khan, V. R. Adarsh, V. Boominathan, J. Tan, A. Veeraraghavan, and K. Mitra, “Towards photorealistic reconstruction of highly multiplexed lensless images,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV, 2019)*, pp. 7860–7869, 2019.
- [87] M.G. Retzlaff, J. Hanika, J. Beyerer and C. Dachsbacher, “Physically based computer graphics for realistic image formation to simulate optical measurement systems,” *Journal of Sensors and Sensor Systems*, 6 (1), 171 –184, 2017.
- [88] D. Schraml, “Physically based synthetic image generation for machine learning: a review of pertinent literature,” *Proc. of Photonics and Education in Measurement Science*, 2019.
- [89] M. Wang and W. Deng, “Deep visual domain adaptation: A survey,” In *Neurocomputing* 312, pp. 135–153, 2018.
- [90] S. Sankaranarayanan, Y. Balaji, A. Jain, S. N. Lim, R. Chellappa, “Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [91] M. Pharr, G. Humphreys and W. Jakob, “Physically based rendering: From theory to implementation,” third Cambridge, MA, 2017.

## 付録

本論で用いた MLAF データセット，および，官能評価テストのデータを下記に公開している。

[https://drive.google.com/drive/folders/1FX2GGv\\_beCj\\_8m5Thyr2YkI-QRobi7CO?usp=sharing](https://drive.google.com/drive/folders/1FX2GGv_beCj_8m5Thyr2YkI-QRobi7CO?usp=sharing)

公開したデータは，13 シーンの MLAF スタック（Multi-level Aperture-Focus スタック・フォーカス・絞り・露光レベルの全パラメータで撮影した画像スタック）のデータセットである。本論では 6 シーンをテスト用，7 シーンを学習用として使用した。付録図 1 および付録図 2 に，本論でのテスト・学習用の区分がわかるように，サンプル画像を示している。それぞれのデータは，22 の絞り値，51 の焦点位置，3 の露光レベル，合計 3,366 枚の画像をそれぞれのシーンについて撮像した，非常に希少なデータである。本論のデータセット公開が，Epsilon Photography の再構成問題を Deep Learning 等で解く研究を数多く誘発する，その一助とれば幸いである。

官能評価テストサンプルは，32 枚から 3,366 枚を再構成した画像群と，当該カメラパラメータで撮影した元画像とを，どちらかがわからないよう side-by-side で並べている。再構成の品質の高さがよくわかるサンプルであるため，是非ともご覧頂きたい。



” Backlight”



” Bingmayong”



”Christmas”



” Horse”



” Night town”

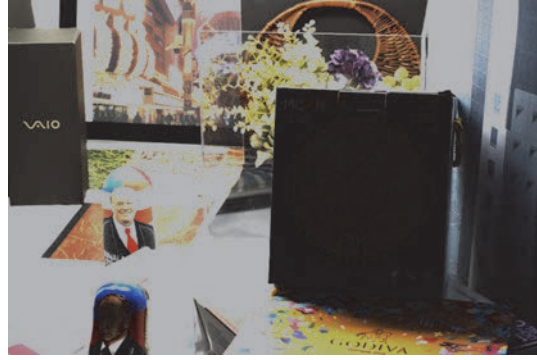


” Zaku”

付録図 1 公開画像データセット: 本論でテストに用いた 6 シーンのサンプル画像



” Caterpillar”



” Godiva”



” Magazine”



” Moon”



” Sensu”



” Star”



” Statue”

付録図 2 公開画像データセット: 本論で学習に用いた 7 シーンのサンプル画像

## 謝辞

本研究を行うに当たり，ご指導頂きました青木義満教授に深く感謝いたします。そして，池原雅章教授，斎藤英雄教授，田中敏幸教授には，副査として本論文を審査して頂きました。感謝いたします。

また，全ての共同研究者に深く感謝いたします。特に，Rice University の Ashok Veeraraghavan 教授には，本研究の基本的な方向性をアドバイスいただき，感謝に堪えません。Carnegie Mellon 大学の Aswin Sankaranarayanan 教授，IIT Madras の Kaushik Mitra 教授，Rice 大学在学時の共著者 Salil Tambe 氏，にも多大なるご支援を賜りました。感謝申し上げます。研究生活におけるサポートや主観評価実験の被験者としてもお世話になった，慶応義塾大学のラボメイトに感謝申し上げます。

最後に，所属本籍であるソニー株式会社と，上司・同僚の皆様には，私が本研究に時間を費やすことにご理解を頂き，かつご支援いただきました。深く感謝を申し上げます。