

広輝度ダイナミックレンジ画像の
生成法に関する研究

2020年3月

木下 裕磨

首都大学東京

広輝度ダイナミックレンジ画像の生成法に関する研究

2020年3月

木下 裕磨

首都大学東京

目次

第 1 章	序論	3
1.1	背景	3
1.2	目的	5
1.3	構成	5
第 2 章	広輝度ダイナミックレンジ画像撮影の課題	9
2.1	表記法	9
2.2	デジタル画像撮影の手順と多重露出画像	10
2.3	輝度のダイナミックレンジ	12
2.4	広輝度ダイナミックレンジ画像	14
2.5	WDR 画像生成法	16
2.6	WDR 画像生成における課題と本研究の位置づけ	19
2.7	まとめ	21
第 3 章	多重露出画像の補正による L-WDR 画像生成	23
3.1	はじめに	23
3.2	シーンの領域分割に基づく多重露出画像補正法	24
3.3	実験および考察	33
3.4	まとめ	51
第 4 章	シーン領域分割に基づく単一 LDR 画像からの L-WDR 画像推定	53
4.1	はじめに	53
4.2	単一 LDR 画像からの多重露出画像推定法	54
4.3	実験および考察	55
4.4	まとめ	63

第 5 章	Reinhard のグローバルオペレータに基づく高速逆トーンマッピングオペレータ	65
5.1	はじめに	65
5.2	Reinhard らのトーンマッピングフレームワーク	66
5.3	Reinhard のグローバルオペレータに基づく逆トーンマッピング法	68
5.4	実験および考察	72
5.5	まとめ	78
第 6 章	逆トーンマッピングのための深層学習ネットワーク “iTM-Net”	79
6.1	はじめに	79
6.2	逆トーンマッピングネットワーク “iTM-Net”	80
6.3	実験および考察	85
6.4	まとめ	87
第 7 章	総論	91
7.1	今後の課題	92
参考文献		95

第 1 章

序論

1.1 背景

撮像センサが扱うことができる輝度のダイナミックレンジは，現実シーンにおけるダイナミックレンジよりもはるかに狭い [1]．そのため，現在のデジタルカメラで撮影された画像では，人間が知覚している輝度のダイナミックレンジを忠実に表現するには至っていない．本論文は，一般の画像，すなわち，低ダイナミックレンジ (Low dynamic range, LDR) 画像の持つこの課題の解決を目的とし，より広い輝度のダイナミックレンジを記録した画像である広輝度ダイナミックレンジ (Wide luminance dynamic range, WDR) 画像を生成する手法を提案している．

WDR 画像は，高ダイナミックレンジ (High dynamic range, HDR) 画像とも呼ばれ，現実シーンにおける広い輝度のレンジの情報を持つ画像である．ここで，WDR 画像には，現実シーンの放射輝度と正比例する画素値を持つ U-WDR 画像や，従来の LDR ディスプレイへの表示を目的とする L-WDR 画像など，目的に応じたいくつかの種類がある．WDR 画像は，写真撮影，コンピュータグラフィックス，監視カメラ，自動運転，医用画像など多くの分野への利用が期待されている．しかしながら，撮像センサにおけるダイナミックレンジの制限により，従来のカメラを用いた WDR 画像の直接的な撮影は困難な状況にある．このような理由から，WDR 画像の生成法として，(1) 専用に設計された特殊なカメラを用いた撮影 [2-5]，(2) 同一シーンを異なる露出条件で撮影した複数枚の LDR 画像 (多重露出画像) の合成 [6-8]，(3) 単一 LDR 画像を用いた WDR 画像の推定という 3 つの方法が研究されている [9-12]．(1) の方法では高価なカメラを必要とするため，本論文では，従来のカメラを用いて実現可能な (2) の多重露出画像に基づく方法，および (3) の単一 LDR 画像に基づく方法に着目する．この研究を通して得られた知見は，複数

センサを用い多重露出画像を一度に撮影する (1) の方法の性能向上にも貢献できる。

(2) の多重露出画像に基づく方法、および (3) の単一 LDR 画像に基づく方法を用いた WDR 画像の生成には、以下に示す 3 つの状況において課題がある。

まず、不明瞭な多重露出画像からの高品質な L-WDR 画像生成である。多重露出画像を用いて高品質な L-WDR 画像を生成するためには、2 つの条件を満たす必要がある。第一の条件は撮影時に複数回シャッターを切ることによる被写体の位置ずれがないことであり、第二の条件は多重露出画像がシーンのダイナミックレンジを明瞭に記録していることである。このような背景から、第一の条件を満たさない場合に適用可能な手法として、多重露出画像間の被写体のずれを補正する手法が数多く研究されている [13–15]。しかしながら、第二の条件を満たさない場合、すなわち不明瞭な多重露出画像が入力として与えられた場合を想定した研究はこれまでにほとんど行われていない。特に、L-WDR 画像生成に適した多重露出画像撮影時の露出値や枚数を決定する方法が未だ明らかとなっていない。さらに、もし適切な露出値を決定できた場合でも、画像撮影時における時間的制約から十分な枚数の多重露出画像を撮影することは一般に困難である。したがって、不明瞭な多重露出画像を入力として与えられた場合を想定した L-WDR 画像生成法の開発が期待されている。

次に、単一 LDR 画像からの L-WDR 画像推定である。高品質な L-WDR 画像の生成は多重露出画像の合成により実現できる一方で、これまでに撮影された多くの LDR 画像には、対応する多重露出画像が存在しない。そのため、これら LDR 画像に対し、多重露出画像に基づく L-WDR 画像生成法を適用することはできない。このような理由から、単一 LDR 画像に画像強調法を適用し、明瞭な L-WDR 画像の推定を目指す研究が数多く行われている [11, 12, 16–21]。しかしながら、従来の画像強調法を用いて得られる L-WDR 画像は、過強調や強調不足により、シーン全体を明瞭に表すことができない。

最後に、単一 LDR 画像からの U-WDR 画像推定である。カメラ特性などの事前情報が利用できない場合、単一 LDR 画像からの U-WDR 画像推定は一般に不良設定問題である。そのため、U-WDR 画像推定法には、カメラ特性を仮定しそれに対応する逆トーンマッピングオペレータを用いる方法 [9, 10, 22–26]、および、単一 LDR 画像と U-WDR 画像間の関係を深層学習を用いてモデル化する方法 [27–29] の 2 つがある。逆トーンマッピングオペレータを用いる U-WDR 画像の推定法は、深層学習を用いる方法と比較して高速かつ軽量に実行可能であることから、一般の LDR 画像からの U-WDR 画像推定に加え、U-WDR 画像の情報源符号化への応用が期待されている [30]。一方、既存の逆トーンマッピングオペレータは、その実行時間の多くを内部パラメータの決定に割いていることから、パラメータの決定にかかる時間を削減することによるさらなる高速化の余地があ

る。対して、深層学習を用いる方法は、単一 LDR 画像からの高精度な U-WDR 画像推定を実現することが期待されている。しかし、教師画像として U-WDR 画像を直接利用した深層学習モデルの学習は、LDR 画像と U-WDR 画像における画素値の分布の違いなどの理由から、困難であることが先行研究により指摘されている [27]。

1.2 目的

本論文では、以上の状況を鑑み、不明瞭な多重露出画像が入力として与えられた場合を想定した L-WDR 画像生成、単一 LDR 画像からの L-WDR 画像推定法の性能向上、および U-WDR 画像推定法の高速化と性能向上を目的とする。これら目的の達成により、適切な多重露出画像の撮影が難しい状況や、そもそも多重露出画像が取得できない状況においても、より高い品質の L-WDR 画像・U-WDR 画像を生成できる。さらに、U-WDR 画像推定法の高速化により、U-WDR 画像の高効率な情報源符号化法の実現が期待できる。

各章で提案する手法は、多重露出画像の輝度補正による L-WDR 画像生成、シーン領域分割に基づく単一 LDR 画像からの L-WDR 画像生成、高速に実行可能な逆トーンマッピングオペレータ、および深層学習を用いた高性能逆トーンマッピングを実現する。各章における実験では、主観評価に加え、客観評価尺度を用いた従来法との比較を行った。これら実験の結果から、提案法が、WDR 画像生成における課題の解決に貢献することを示す。

1.3 構成

図 1.1 に、本論文中の各章の関係を示す。本論文は、以下の 7 章で構成されている。

第 1 章では、本研究分野の背景と課題を述べ、本研究の目的や本論文の構成を述べる。

第 2 章では、デジタル撮影と輝度のダイナミックレンジ、および WDR 画像技術について説明し、WDR 画像生成法の概要とその課題について述べる。

第 3 章では、シーンのダイナミックレンジを明瞭に記録していない多重露出画像および明瞭な多重露出画像の両方を入力として想定し、高品質な L-WDR 画像を生成する手法を提案する。提案法は、多重露出画像の輝度補正によって明瞭な多重露出画像を生成できる。それら多重露出画像を従来の L-WDR 画像生成法によって合成することで、高品質な L-WDR 画像が得られる。明瞭な多重露出画像の生成は、輝度に関してシーンを領域分割し、分割によって得られる各領域をよく表現する画像をそれぞれ生成することにより行われる。既存の L-WDR 画像生成法を用いた実験により、提案法の利用が、不明瞭な多重露出画像から生成される L-WDR 画像の品質を向上させることが示される。

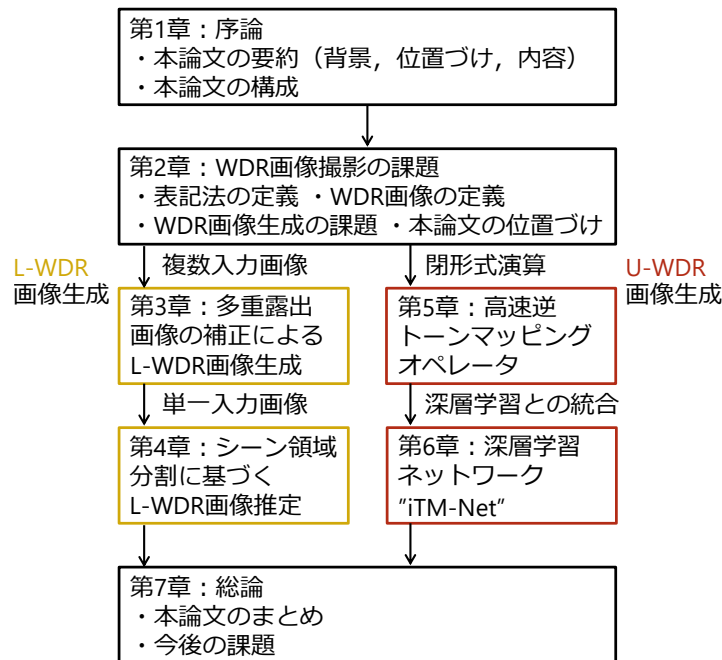


図 1.1 各章の関係

第 4 章では，第 3 章で提案する手法を拡張し，単一 LDR 画像から L-WDR 画像を推定する手法を提案する．本章では，第 3 章で提案するシーン領域分割を単一 LDR 画像に対して適用できるように拡張し，単一 LDR 画像から多重露出画像を擬似的に生成可能とする．これら擬似的に生成された多重露出画像の合成により，シーン全体を明瞭に表す L-WDR 画像が得られる．単一画像に基づく WDR 画像推定法との比較により，主観的および客観的品質の観点から提案法の有効性を確認する．

第 5 章では，高速に実行可能な逆トーンマッピングオペレータを提案する．提案法は，Reinhard らにより提案されたトーンマッピングオペレータの逆関数に基づき，単一 LDR 画像から U-WDR 画像を推定する．提案法において用いられるパラメータは閉形式で高速に計算可能でき，結果として，逆トーンマッピングオペレータの高速化が実現される．従来の逆トーンマッピングオペレータとの比較実験により，提案法は，従来法と同等の品質を持つ U-WDR 画像を，より高速に計算できることが示される．

第 6 章では，第 5 章で提案する逆トーンマッピングオペレータと深層学習を組み合わせた，逆トーンマッピングのための深層学習ネットワークを提案する．第 5 章の逆トーンマッピングオペレータは，入力 LDR 画像が Reinhard のトーンマッピングオペレータにより生成された場合に極めて高い性能を持つ．提案法は，その条件を満たすよう，入力 LDR 画像を深層ニューラルネットワークにより変換した後，得られる画像に対して第 5

第 1 章 序論

章で提案する逆トーンマッピングを実行する。これにより，従来法を上回る性能を持つ逆トーンマッピング法を実現できることが示される。

第 7 章では，本論文の総括を行い，各章で提案した手法で得られる利点や効果についてまとめを示している。

第2章

広輝度ダイナミックレンジ画像撮影の課題

本章では、本論文で用いる表記法について説明したのち、デジタル画像撮影の処理手順、輝度のダイナミックレンジ、広輝度ダイナミックレンジ画像について説明する。最後に、デジタル画像の広輝度ダイナミックレンジ化における課題を述べ、本研究の位置づけを示す。

2.1 表記法

本論文を通して以下の表記を用いる。

- a などの細字のイタリック体はスカラーを表す。
- \mathbf{a} などの太字のイタリック体小文字はベクトルを表す。ここで、特に明記しない限り、すべてのベクトルは列ベクトルであるものとする。
- (a_1, a_2, \dots, a_N) という表記は N 次元の行ベクトルを表す。
- A などの太字のローマン体大文字は行列を表す。
- ベクトルや行列の転置は、上付きの \top を用いて、 \mathbf{a}^\top, A^\top のように表記する。
- \mathbb{A} などの黒板太字のローマン体大文字は集合を表す。特に、実数全体の集合、非負の実数の集合、および正の実数の集合はそれぞれ $\mathbb{R}, \mathbb{R}_+, \mathbb{R}_{++}$ と表す。
- $\{a_1, a_2, \dots, a_N\}$ と表記した場合には、 N 個の要素 a_1, \dots, a_N を持つ集合を表す。もし、要素について曖昧性がない場合には、同様の集合をより簡潔に $\{a_n\}$ として表す。

- $p(a)$ は、確率変数 a の確率分布関数を表す。
- U および V は、画像の幅および高さを表すために用いる。
- ある画像に含まれるすべての画素の集合は、 \mathbb{P} として表記する。具体的には、 $\mathbb{P} = \{(u, v)^\top | u \in \{1, 2, \dots, U\} \wedge v \in \{1, 2, \dots, V\}\}$ として与える。ただし、二項演算子 \wedge は論理積を表す。
- 画像中のある画素 \mathbf{p} は、 $\mathbf{p} = (u, v)^\top \in \mathbb{P}$ として与える。
- カラー画像は、ベクトル値関数 \mathbf{x} を用いて $\mathbf{x} : \mathbb{P} \rightarrow \mathbb{R}^3$ として表記する。ここで、この出力 $\mathbf{x}(\mathbf{p}) = (x_r(\mathbf{p}), x_g(\mathbf{p}), x_b(\mathbf{p}))$ の要素は、RGB 色空間で表される画素値の R, G , および B 成分をそれぞれを意味するものとする。
- ある画像の輝度は、関数 $l : \mathbb{P} \rightarrow \mathbb{R}$ を用いて記す。ここで、この出力は、CIE XYZ 色空間における Y 成分に相当するものとする [31]。
- 画素の集合 $\mathbb{P}' \subseteq \mathbb{P}$ 上の輝度 l の幾何平均 $G(l|\mathbb{P}')$ を、以下のように定義する。

$$G(l|\mathbb{P}') = \exp \left(\frac{1}{|\mathbb{P}'|} \sum_{\mathbf{p} \in \mathbb{P}'} \log (\max (l(\mathbf{p}), \epsilon)) \right) \quad (2.1)$$

ここで、 $\epsilon \in \mathbb{R}_{++}$ は、 $l(\mathbf{p}) = 0$ における特異点を回避するための微小な値である。

2.2 デジタル画像撮影の手順と多重露出画像

図 2.1 に、デジタルカメラを用いた代表的な画像撮影の流れを示す [32]。本節では簡単のため、グレースケール画像を例とした流れを説明する。カラー画像の場合には、後述する関数 f の適用を R, G , および B 成分について独立に行う。

2.2.1 デジタル画像撮影の手順

センサに照射される光は、シャッターが開いている時間 $\Delta t \in \mathbb{R}_+$ の間累積される。この累積される光の、単位面積当たりのエネルギーを露出 X と呼ぶ。静的なシーンであれば、露出 X は、放射照度 $E \in \mathbb{R}_+$ とシャッタースピード Δt を用いて、以下のように表される。

$$X(\mathbf{p}) = E(\mathbf{p})\Delta t \quad (2.2)$$

ここで、 $\mathbf{p} = (u, v)^\top$ は点 (u, v) における画素を表す。撮影される画像 x の画素値 $x(\mathbf{p}) \in [0, 1]$ は、以下のように与えられる。

$$x(\mathbf{p}) = f(X(\mathbf{p})) \quad (2.3)$$

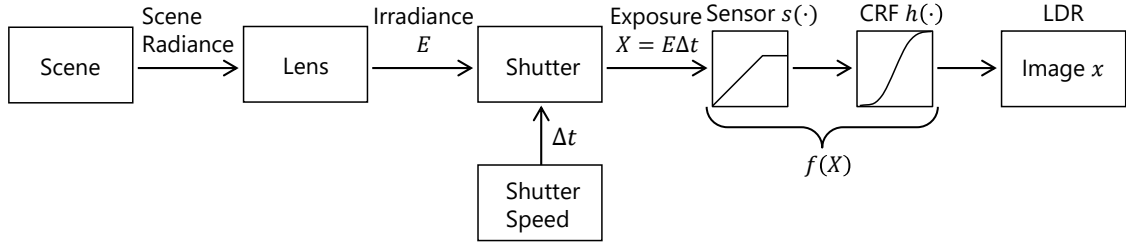


図 2.1 デジタルカメラによる画像撮影の手順

ここで、 $f: \mathbb{R}_+ \rightarrow [0, 1]$ は、センサの飽和を表す関数 $s: \mathbb{R}_+ \rightarrow [0, 1]$ とカメラが持つ非線形な応答を表す関数 (CRF) $h: [0, 1] \rightarrow [0, 1]$ の合成関数である。センサの飽和は、次式のようにモデル化できる。

$$s(t) = \begin{cases} t & (0 \leq t < 1) \\ 1 & (1 \leq t) \end{cases} \quad (2.4)$$

一方、CRF h は、シーンの輝度ダイナミックレンジの圧縮など、個々のカメラが出力画像 $x(\mathbf{p})$ の品質を向上させるために行う処理を表す [33]。

2.2.2 露出値と多重露出画像

撮影される画像の明るさは、露出の大きさによって決定される。露出の大きさを変化させながら、同一シーンを撮影することによって得られる複数枚の画像を、多重露出画像という。露出の大きさはシャッタースピードやレンズ絞り (F 値)、センサゲイン (ISO 感度) の 3 つのカメラパラメータにより調整可能であるが、レンズ絞りを変更する方法では、放射照度 E の変化に伴って画像のぼけ具合が変化してしまう。また、センサゲインを変更する方法では、高ゲインにするほどノイズが増加してしまう問題がある。これらの理由から、多重露出画像の撮影においてはシャッタースピードのみを変化させ露出を制御することが一般的である。本論文でも同様に、シャッタースピードを除くカメラパラメータは固定であるとして取り扱う。

露出の大きさは、露出値 (Exposure Value, EV) と呼ばれる数値を用いて表される。ここで、個々のカメラにより決定されるシーンに対して適切な露出値を $0[\text{EV}]$ とし、その時のシャッタースピードを Δt_0 とする。このとき、露出値 $v_i[\text{EV}]$ で画像を撮影する際のシャッタースピード Δt_{v_i} は、

$$\Delta t_{v_i} = 2^{v_i} \Delta t_0 \quad (2.5)$$

として与えられる。式 (2.2) から (2.5) を用いると、露出値 $v_i[\text{EV}]$ で撮影される画像 x_i

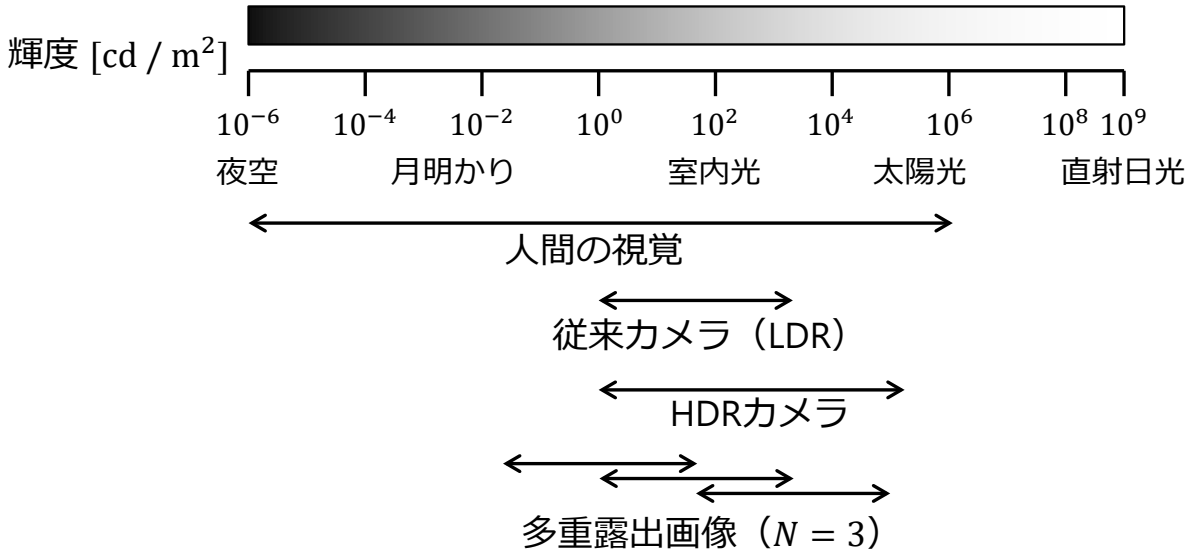


図 2.2 画像撮影における輝度のダイナミックレンジの比較

は以下のように表される.

$$x_i(p) = f(E(p)\Delta t_{v_i}) = f(2^{v_i} E(p)\Delta t_0) \quad (2.6)$$

多重露出画像は, N 個のシャッタースピード $\{\Delta t_{v_1}, \Delta t_{v_2}, \dots, \Delta t_{v_N}\}$ を用いて撮影された N 枚の画像 $\{x_1, x_2, \dots, x_N\}$ として与えられる.

2.3 輝度のダイナミックレンジ

カメラを用いて撮影可能な輝度のレンジやディスプレイで出力可能な輝度のレンジは, そのレンジにおける最小値と最大値の比であるダイナミックレンジを用いて表される.

2.3.1 画像撮影におけるダイナミックレンジ

図 2.2 には, 画像撮影における輝度のダイナミックレンジの比較を示す. 現実シーンにおける輝度のダイナミックレンジは非常に広く, 夜空などの暗所では $10^{-6}[\text{cd}/\text{m}^2]$, 太陽光では $10^9[\text{cd}/\text{m}^2]$ と, 10^{15} ものダイナミックレンジを持つ. 人間の眼は, 主に明るいシーンで働く錐状体と主に暗いシーンで働く杆状体という, 2 種類の細胞により光を知覚している. 錐状体が飽和なしに感知できる光の強さは最大で約 $10^6[\text{cd}/\text{m}^2]$ であり, 杆状体は最低で約 $10^{-6}[\text{cd}/\text{m}^2]$ の光を感知できる. よって, 人間の視覚は 10^{12} 程度のダイナミックレンジを持つ.

現在一般に用いられているカメラ (LDR カメラ) のダイナミックレンジは, スマートフォン (iPhone XS Max) で $10^{2.6}$ 程度, フルサイズセンサを持つプロ向け 1 眼レフカメラ (Canon EOS 1D X Mark II) で $10^{3.2}$ 程度であり, 現実シーンのダイナミックレンジと比較して非常に狭い [34]. したがって, 従来のカメラを用いて撮影された単一画像は, 現実シーンが持つ情報の多くを失っているといえる. 例えば, トンネル内を走行している自動車からトンネル外を撮影しようとした際には, 日光によってシーンの輝度ダイナミックレンジが非常に広くなり, トンネル外の情報が欠落してしまう場合がある. このことは, ドライブレコーダや監視カメラを用いて映像情報を記録する際などにおいて重要な問題となる.

このような背景から, 従来カメラより広いダイナミックレンジを記録可能な高ダイナミックレンジ (High dynamic range, HDR) カメラの開発が進められている. 例えば, Tocci らの開発した HDR カメラは, 1 つのカメラ内に複数のセンサを配置することによって, 約 $10^{5.1}$ ものダイナミックレンジを記録可能にした [3]. また, 従来の LDR カメラを用いた場合でも, 多重露出画像を撮影することで, 広い輝度のダイナミックレンジを記録可能である (図 2.2 参照).

2.3.2 画像表示におけるダイナミックレンジ

図 2.3 には, 画像表示における輝度のダイナミックレンジの比較を示す. 従来の LDR ディスプレイが出力できる輝度のレンジは, CRT ディスプレイで $0.1\text{--}100[\text{cd}/\text{m}^2]$ 程度, LCD ディスプレイで $0.1\text{--}400[\text{cd}/\text{m}^2]$ 程度であり, 人間が知覚可能な輝度のレンジと比べて極めて狭い [1]. そのため, カメラの HDR 化に伴って, より広い輝度のレンジを出力可能な HDR ディスプレイの開発が進められている. 例えば, EIZO 社が開発した HDR ディスプレイである ColorEdge PROMINENCE CG3145-BS は, 最大 $1000[\text{cd}/\text{m}^2]$ の輝度を出力可能である [35]. また, HDR 放送のための規格としては, Hybrid log-gamma (HLG) 方式および Perceptual quantization (PQ) 方式という 2 つの方式が ITU-R BT.2100 において採用されている [36]. このうち PQ 方式では, $0.005\text{--}10000[\text{cd}/\text{m}^2]$ という広い輝度レンジの出力を想定している.

以上より, HDR ディスプレイの開発および標準化は現在進行中であり, これらディスプレイは今後ますます普及しているものと考えられる. したがって, HDR ディスプレイの性能を生かした高品質な映像表現のためにも, 画像・映像コンテンツの広輝度ダイナミックレンジ化が期待されている.

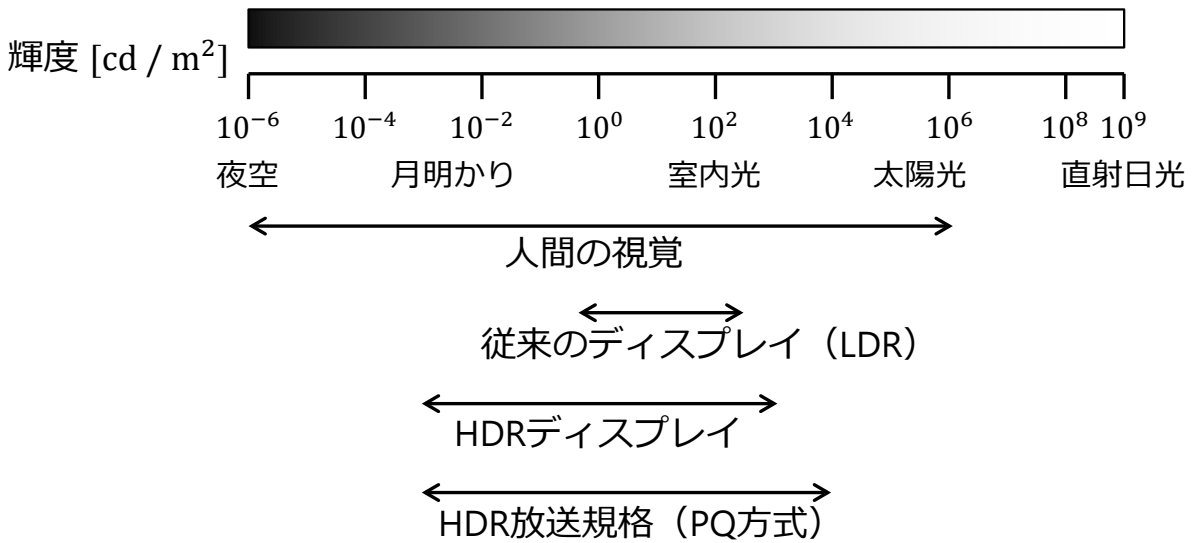


図 2.3 画像表示における輝度のダイナミックレンジの比較

2.4 広輝度ダイナミックレンジ画像

本論文では、従来のデジタルカメラで撮影可能な輝度のダイナミックレンジより広いダイナミックレンジの情報を記録した画像を、広輝度ダイナミックレンジ (Wide luminance dynamic range, WDR) 画像と呼ぶ。広いダイナミックレンジを持つ画像を指す用語として、既に“HDR 画像”という用語が一般に用いられているが、HDR 画像は、広色域かつ広輝度ダイナミックレンジの画像という意味で用いられることもある。そのため、本論文では、輝度ダイナミックレンジのみに着目していることを明確にする目的で、WDR 画像という用語を用いる。また、HDR 画像は、広色域かつ広輝度ダイナミックレンジの画像という意味で用いる。それに対して、従来のカメラで撮影される狭色域かつ低輝度ダイナミックレンジの画像を、LDR 画像という。図 2.4 には、LDR 画像、WDR 画像、HDR 画像の 3 つの用語の関係を示す。

WDR 画像は、出力ダイナミックレンジにより U-WDR 画像、H-WDR 画像、および L-WDR 画像の 3 種類に分けることができる (表 2.1 参照)。

U-WDR 画像は、ディスプレイへの出力を想定せず、現実シーンの放射輝度と正比例する画素値を持つ WDR 画像である。コンピュータビジョン分野では、物体の双方向反射率分布関数などのシーン情報を、放射輝度に基づき推定する目的で U-WDR 画像が利用される [37]。また、U-WDR 画像を用いることで、ある CRF を持つカメラによる画像撮影をコンピュータ上でシミュレーションすることが可能である。U-WDR 画像は、一般

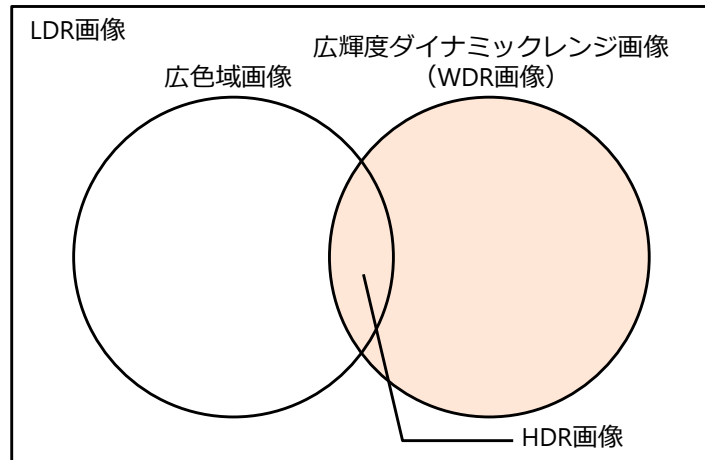


図 2.4 LDR 画像, WDR 画像, HDR 画像の違い

表 2.1 ダイナミックレンジに関する画像の分類

分類		入力ダイナミックレンジ	出力ダイナミックレンジ
LDR		Low	Low
WDR	L-WDR	High	Low
	H-WDR	High	High
	U-WDR	High	Ultra High

に放射輝度マップ, 放射照度マップ, または HDR 画像と呼ばれるが, 後述する H-WDR 画像および L-WDR 画像との違いを明確にするため, 本論文では U-WDR 画像という.

H-WDR 画像は, HDR ディスプレイでの表示を目的とした WDR 画像である. 上記の U-WDR 画像は極めて広いダイナミックレンジの情報を持つ一方で, その情報を直接表示可能なディスプレイはない. H-WDR 画像は, U-WDR 画像を HDR ディスプレイで表示するために, ダイナミックレンジを圧縮するトーンマッピング処理を利用して変換した画像である. したがって, 高品質な H-WDR 画像の生成は, 高品質な U-WDR 画像の生成を通して達成される. この理由から, 本論文では, H-WDR 画像を直接生成する方法を考慮しない.

L-WDR 画像は, LDR ディスプレイでの表示を目的とした WDR 画像である. HDR ディスプレイは開発が進行中であるが, 非常に高価な上, 性能面の課題や H-WDR コンテントの不足により普及には至っていない. 一方, 現在普及している LDR ディスプレイでは, U-WDR 画像および H-WDR 画像が持つ広いダイナミックレンジを表現することが

できない。このため、撮影時の輝度ダイナミックレンジの観点から高品質である L-WDR 画像を生成することは未だ重要な課題である。特に、WDR 画像の持つ広いダイナミックレンジの輝度情報を LDR ディスプレイ上に表示するためには、L-WDR 画像がシーン全体を明瞭に記録していることが必要とされる。ただし、本論文では、“明瞭”という用語を“被写体を視認するために適した明るさを持つこと”という意味で用い、ぼけやぶれなどの歪みの有無とは独立のものとして扱う。

次節では、これら WDR 画像の生成法についてより詳細に説明する。

2.5 WDR 画像生成法

第 2.3.1 項で述べた通り、従来の LDR カメラによる WDR 画像の直接的な撮影は、撮像センサにおけるダイナミックレンジの制限により、困難な状況にある。このような理由から、WDR 画像の生成法として、以下に示す 3 つの方法が研究されている。

1. 専用に設計された特殊なカメラ、すなわち HDR カメラを用いた撮影
2. 同一シーンを異なる露出条件で撮影した複数枚の LDR 画像 (多重露出画像) の合成
3. 単一 LDR 画像を用いた WDR 画像の推定

以降、これらの方法についての詳細を述べる。

2.5.1 HDR カメラを用いた WDR 画像の撮影

HDR カメラを用いた WDR 画像の撮影には主に、広いダイナミックレンジを記録可能な撮像センサを利用する方式、1 つのカメラに複数センサを搭載する方式、および、単一センサにおける画素単位で露出を制御する方式がある。

広いダイナミックレンジを記録可能な撮像センサを用いたカメラとして、Sony F65/F55 等がある。これらのカメラは、従来のカメラと比較してより広い、 $10^{4.2}$ 程度のダイナミックレンジを記録可能である。一方、単一の撮像センサが一度に記録可能なダイナミックレンジを広げることは、物理的制約によって難しい状況にある。この理由により、残る 2 つの方式を採用した HDR カメラの研究が盛んに行われている。

1 つのカメラに複数センサを搭載する方式では、ビームスプリッタ等を利用して各センサに照射される光量を制御することで、多重露出画像を時間ずれなしに撮影することを可能としている [2,3]。また、単一センサにおける画素単位で露出を制御する方式では、センサ上の各画素に異なるシャッタースピードを割り当てることで、時間ずれのない多重露出画像の撮影を実現する [4,5]。後者の方式では、空間的にシャッタースピードを変化させ

多重露出画像を撮影するため，得られる多重露出画像の解像度は，センサが持つ解像度より低いものとなる．これらの方式では，撮影された多重露出画像の合成として，WDR 画像を生成する．

2.5.2 多重露出画像の合成による WDR 画像の生成

多重露出画像の合成に基づく方法は，WDR 画像生成法として最も代表的なものである．この方法は，生成したい WDR 画像の種類により，異なる処理を必要とする．

U-WDR 画像の生成

多重露出画像に基づく U-WDR 画像の生成は，多重露出画像からシーンの放射照度 \mathbf{E} を計算することにより行われる [6, 7, 13, 14, 33, 38–40]．ここで， $\mathbf{E}(\mathbf{p}) = (E_r(\mathbf{p}), E_g(\mathbf{p}), E_b(\mathbf{p}))^\top$ である．LDR 画像はセンサの飽和と非線形 CRF の影響を受けて歪んでいるため，放射照度 \mathbf{E} の計算は，画素値の線形化および飽和領域の復元という 2 つの処理に分けられる．

多重露出画像に基づく U-WDR 画像の生成処理では，画素値の線形化のため，多重露出画像からカメラ特性 f の推定を行う．次に，推定処理によって得られた関数 \hat{f} の逆関数 \hat{f}^{-1} を各多重露出画像 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ に適用し，放射照度 $\{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \dots, \hat{\mathbf{E}}_N\}$ を計算する．最も代表的な U-WDR 画像生成法である Debevec らの手法 [7] は，関数 f が滑らかな単調増加関数かつその逆関数が存在するという仮定の下，関数 f と放射照度 \mathbf{E} の推定を同時に行う．この推定では， N 枚の画像からなる多重露出画像 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ とその撮影に用いられたシャッタースピード $\{\Delta t_{v_1}, \Delta t_{v_2}, \dots, \Delta t_{v_N}\}$ を用いて，次式を \hat{f} および $\hat{\mathbf{E}}_i$ に関して最小化する．

$$\sum_{\mathbf{p} \in \mathbb{P}} \sum_{i=1}^N \left(\ln \hat{f}_c^{-1}(x_{c,i}(\mathbf{p})) - \ln \hat{\mathbf{E}}_{c,i}(\mathbf{p}) - \ln \Delta t_i \right)^2 + \lambda \sum_{t \in \mathbb{I} \setminus \{\max \mathbb{I}, \min \mathbb{I}\}} \left(\frac{\partial^2}{\partial t^2} \ln \hat{f}_c^{-1}(t) \right)^2 \quad (2.7)$$

ここで， \mathbb{I} は，多重露出画像に含まれるすべての画素値の集合 $\mathbb{I} = \{x_i(\mathbf{p}) | i \in \{1, \dots, N\} \wedge \mathbf{p} \in \mathbb{P}\}$ であり， $c \in \{r, g, b\}$ である．また，LDR 画像は離散値で表現されていることから，二階微分は差分により近似する．式 (2.7) を最小化する $\hat{f}, \hat{\mathbf{E}}_i$ が，関数 f および放射照度 \mathbf{E} それぞれの推定値である．

各多重露出画像から計算される放射照度 $\{\hat{\mathbf{E}}_n\}$ は，センサの飽和による影響を受けている．一方，多重露出画像はそれぞれ異なるシャッタースピードを用いて撮影されているため，画素値が飽和している領域は各画像で異なる．この理由から，最終的な放射照度の推

定値, すなわち U-WDR 画像 $\hat{\mathbf{E}}$ は, $\hat{\mathbf{E}}_i$ の画素単位の重み付き平均として与えられる.

$$\hat{\mathbf{E}}_c(\mathbf{p}) = \frac{\sum_{i=1}^N w(x_{c,i}(\mathbf{p})) \hat{\mathbf{E}}_{c,i}(\mathbf{p})}{\sum_{i=1}^N w(x_{c,i}(\mathbf{p}))} \quad (2.8)$$

Debevec らの手法では, 次式で与えられるハット関数を重み w として利用する.

$$w(t) = \begin{cases} 2t & (t \leq 1/2) \\ -2(t-1) & (1/2 < t) \end{cases} \quad (2.9)$$

L-WDR 画像の生成

L-WDR 画像の生成では, 放射照度 \mathbf{E} の推定が必要ない. この理由から, L-WDR 画像の生成は, 関数 f の推定なしに多重露出画像を直接合成することで行われる [8,15,41–46].

例えば, Mertens らの L-WDR 画像生成法 [8] では, L-WDR 画像 $\mathbf{y}(\mathbf{p}) = (y_r(\mathbf{p}), y_g(\mathbf{p}), y_b(\mathbf{p}))$ を, 多重露出画像の重み付き平均として次式のように与える.

$$\mathcal{L}(\mathbf{y}) = \sum_{i=1}^N \mathcal{G}(w_i) \mathcal{L}(\mathbf{x}_i) \quad (2.10)$$

ここで, 重み w_i は, コントラスト, 彩度, Well-exposedness の観点から各多重露出画像 \mathbf{x}_i を評価して計算される. また, $\mathcal{L}(\mathbf{x})$ および $\mathcal{G}(\mathbf{x})$ は, 画像 (\mathbf{x}) のラプラシアンピラミッドおよびガウシアンピラミッドをそれぞれ示す.

また, L-WDR 画像は, より広いダイナミックレンジを記録する U-WDR 画像から, トーンマッピング処理により生成することも可能である. トーンマッピング処理については, 第 5.2 節にて詳しく述べる.

2.5.3 単一 LDR 画像を用いた WDR 画像の推定

単一 LDR 画像を用いた WDR 画像の推定法は, 多重露出画像を必要としないことから, 既存の LDR 画像の WDR 化を主な目的として研究されている. この方法も, 生成したい WDR 画像の種類により, 異なる処理を必要とする.

U-WDR 画像の推定

単一 LDR 画像を用いた U-WDR 画像の推定処理は, 逆トーンマッピングと呼ばれる. 前節で述べたとおり, U-WDR 画像の生成は, 画素値の線形化, および飽和領域の復元の 2 つの部分問題に分けられる. しかしながら, カメラの特性 f などの事前情報が利用できない場合には, これらは共に不良設定問題となる.

この理由から、これまでに研究されてきた逆トーンマッピング法の多くは、LDR 画像が特定の関数 f により生成されていると仮定し、その関数 f に対応する逆トーンマッピングオペレータを用いて画素値の線形化を行う [9, 10, 22–26]. 例えば, Huo らは, 逆トーンマッピングオペレータとして, S 字カーブを用いている [23]. また, さらに高精度な画素値の線形化のために, 畳み込みニューラルネットワーク (Convolutional neural network, CNN) を用いる方法も提案されている [29]. また, CNN は飽和領域の復元にも効果的であり, CNN に基づく飽和領域復元と逆トーンマッピングオペレータを用いる線形化を組み合わせた方法がいくつか提案されている [27, 28].

L-WDR 画像の推定

単一 LDR 画像を用いた L-WDR 画像の推定は, LDR 画像が持つコントラストの強調によって, シーン全体を明瞭に表す画像を生成することで行われる [11, 12, 16–21]. これら方法の中で最も代表的なものは, Histogram equalization (HE) に基づく手法である [11, 12, 16]. HE は, 輝度の分布のエントロピーを最大化することによって画像のコントラスト強調を行う. また, Retinex 理論 [47] に基づく手法も近年注目を集めている [17, 18]. Retinex 理論に基づく手法では, 画像を Reflectance と Illumination に分解し, その後 Illumination をガンマカーブなどにより処理することで強調を行う.

2.6 WDR 画像生成における課題と本研究の位置づけ

本節では, WDR 画像生成における課題と, その中での本研究の位置づけについて述べる.

2.6.1 WDR 画像生成における課題

多重露出画像に基づく方法および単一 LDR 画像に基づく方法における課題は, それぞれ以下の通りである.

多重露出画像に基づく WDR 画像生成法における課題

多重露出画像を用いて高品質な L-WDR 画像を生成するためには, 以下の条件を満たす必要がある.

- 多重露出画像撮影時に, 複数回シャッターを切ることに起因する被写体の位置ずれがないこと

- 多重露出画像がシーンのダイナミックレンジを明瞭に記録していること

一方で、三脚等に固定されていないカメラでの動く被写体の撮影は、ごく一般的なシチュエーションである。このような背景から、第一の条件を満たさない場合に適用可能な手法として、多重露出画像間の被写体のずれを補正する手法が数多く研究されている [13–15]。

しかしながら、第二の条件を満たさない場合、すなわち不明瞭な多重露出画像が入力として与えられた場合を想定した研究はこれまでにほとんど行われていない。L-WDR 画像生成のための多重露出画像を撮影する際の、最適な露出値や枚数を決定する方法は、未だ明らかとなっていない。さらに、もし適切な露出値や枚数を決定できた場合でも、画像撮影時における時間的制約から、高い露出値での画像撮影や十分な枚数の画像を撮影することは一般に困難である。したがって、第二の条件を満たさない場合を想定した、高品質な L-WDR 画像生成法の開発が期待されている。

単一 LDR 画像に基づく WDR 画像生成法における課題

単一 LDR 画像からの U-WDR 画像生成において、飽和領域の復元については Eilertsen ら、および Endo らにより一定の成果が得られている [27, 28]。一方、画素値の線形化については未だ高性能な方法が開発されていない。Marnerides らは、Min-max 正規化を用いて規格化された U-WDR 画像を CNN の学習に利用し、学習された CNN によって画素値の線形化を行う手法を提案している [29]。しかしながら、教師画像として U-WDR 画像を直接用いた CNN の学習は、LDR 画像と U-WDR 画像における画素値の分布の違いなどの理由から、困難であることが先行研究により指摘されている [27]。したがって、高精度な画素値の線形化を実行できる逆トーンマッピング法の開発が、高品質な U-WDR 画像の推定のために必要である。

また、逆トーンマッピングオペレータを用いる逆トーンマッピング法は、CNN に基づく方法より低い性能を持つものの、低い計算コストで実行できることから U-WDR 動画像の生成への応用が期待されている。さらに、逆トーンマッピングオペレータを用いる方法は、U-WDR 画像の情報源符号化に応用でき、JPEG XT として標準化されている [30]。これらの理由から、逆トーンマッピングオペレータを用いる逆トーンマッピング法の性能向上と高速化は未だに重要な課題である。

単一 LDR 画像からの L-WDR 画像推定法である HE に基づく手法、および Retinex 理論に基づく手法は、どちらにも画像中の明るい領域を過強調してしまう問題、あるいは暗い部分を十分に強調できない問題がある。したがって、画像全体を明瞭に表す L-WDR 画像推定法の開発が期待されている。

第 2 章 広輝度ダイナミックレンジ画像撮影の課題

表 2.2 提案する手法の位置づけ. ◎は従来法より優れていることを表し, ○は従来法と同等であることを示す.

提案法	入力画像	出力画像	実行速度	性能	従来法
第 3 章 (既発表論文 [48])	多重露出画像 (不明瞭)	L-WDR 画像		◎	[8, 15, 45, 46] 等
第 4 章 (既発表論文 [49, 50])	単一 LDR 画像	L-WDR 画像		◎	[11, 12, 17, 18] 等
第 5 章 (既発表論文 [26, 51])	単一 LDR 画像	U-WDR 画像	◎	○	[9, 22–24] 等
第 6 章 (既発表論文 [52])	単一 LDR 画像	U-WDR 画像		◎	[27–29] 等

2.6.2 本研究の位置づけ

本論文では, 以上の状況を鑑み, 不明瞭な多重露出画像が入力として与えられた場合を想定した L-WDR 画像生成, 単一 LDR 画像からの L-WDR 画像推定法の性能向上, および U-WDR 画像推定法の高速度化と性能向上を目的とする (表 2.2 参照).

第 3 章で提案する手法は, 入力多重露出画像の輝度を補正することにより, 不明瞭な多重露出画像が与えられた場合でさえも, 明瞭な多重露出画像を生成できる. それら多重露出画像の合成として, 高品質な L-WDR 画像生成が生成される. また, 第 3 章で提案する多重露出画像補正法は, 入力多重露出画像より多くの多重露出画像を生成することを可能とする. この特徴を利用し, 続く第 4 章では, 単一 LDR 画像から擬似的に多重露出画像を生成・合成することで, 単一 LDR 画像からの明瞭な L-WDR 画像の推定を実現する.

さらに, 第 5 章では, Reinhard のグローバルオペレータの逆関数に基づく高速逆トーンマッピングオペレータを提案し, 単一 LDR 画像からの高速な逆トーンマッピングを実現する. 加えて, 第 6 章では, この逆トーンマッピングオペレータを CNN と組み合わせた, 従来法を上回る性能を持つ逆トーンマッピングネットワーク “iTM-Net” を提案する.

2.7 まとめ

本章では, デジタル画像撮影の処理手順, 輝度のダイナミックレンジ, WDR 画像について述べた. 従来のカメラ・ディスプレイは, 現実シーンと比較して極めて狭いダイナミックレンジを持っており, HDR カメラ・ディスプレイの開発が進行中であることを示した.

また、従来のカメラより広いダイナミックレンジを記録した WDR 画像について述べた。WDR 画像は、目的とする出力デバイスのダイナミックレンジにより、U-WDR 画像、H-WDR 画像、および L-WDR 画像に分類できることを示した。

さらに、WDR 画像の生成するための 3 つの方法、すなわち HDR カメラを用いた撮影、多重露出画像からの WDR 画像の合成、および単一 LDR 画像を用いた WDR 画像の推定について述べた。U-WDR 画像の生成処理は、センサを原因とする飽和領域の復元と、非線形な関数 f の影響を取り除く画素値の線形化という処理に分けられることを述べた。単一 LDR 画像が入力として与えられた場合には、そのどちらも不良設定問題であることを示した。一方、L-WDR 画像の生成は、画素値の線形化が不要であり、多重露出画像の直接合成、あるいは単一 LDR 画像の強調により行われることを示した。

また、WDR 画像生成法における課題について説明し、その中での本研究の位置づけを述べた。

次章より、WDR 画像生成における課題を解決するための、4 つの新しい WDR 画像生成法をそれぞれ提案する。

第 3 章

多重露出画像の補正による L-WDR 画像生成

3.1 はじめに

第 2 章で述べたように，撮像センサのダイナミックレンジより広い輝度のレンジを記録した画像を生成するため，多重露出画像に基づく L-WDR 画像生成法が広く研究されている [8, 15, 41–46]．しかしながら，これら手法によって高品質な L-WDR 画像を生成するためには，第 2.6.1 項で示した 2 つの条件を満たす必要がある [53, 54]．ここで，第一の条件は被写体の位置が多重露出画像間でずれていないことであり，第二の条件は入力多重露出画像がシーンのダイナミックレンジを明瞭に記録していることである．

第一の条件を満たさない多重露出画像から高品質な L-WDR 画像を生成するために，被写体のずれに対して頑健な L-WDR 画像生成法が研究されている [15]．これらの方法では，入力多重露出画像中のある LDR 画像を基準として被写体の位置を補正し，それら補正された多重露出画像を合成する．これにより，明瞭な多重露出画像が与えられた場合には，シーンが動的な被写体を含む場合でさえ，これら手法により高品質な L-WDR 画像を生成できる．しかしながら，入力多重露出画像の明瞭さは，被写体の位置補正の精度および生成される L-WDR 画像の品質の両方に影響を及ぼす．したがって，撮影時の時間的制約により高速なシャッタースピードを用いて撮影される不明瞭な多重露出画像は，高品質な L-WDR 画像の生成を妨げる．以上のことから，第二の条件を満たさない不明瞭な多重露出画像を想定した L-WDR 画像生成法の開発が期待されている．

本章ではまず，多重露出画像の輝度を撮影後に補正することが，多重露出画像の品質向上に有効であることを指摘する．さらに，この事実に基づき，シーンの領域分割に基づく

多重露出画像補正法を提案する。提案法は、不明瞭な多重露出画像が入力として与えられた場合にさえ、高品質な L-WDR 画像の生成に適した明瞭な多重露出画像を生成可能とする。加えて、入力多重露出画像を L-WDR 画像生成に適したものへ自動で補正するために、2つのシーン領域分割法を提案する。1つ目の領域分割法では、入力多重露出画像のうち中間の明るさを持つ LDR 画像の輝度に従ってシーンの分割を行う。この方法は閉形式で表されるため、他方の分割法と比較して低い計算コストをもつ。2つ目の領域分割法は、すべての入力画像の輝度の分布を考慮することで、高品質なシーンの分割を可能とする。さらに、この方法は、 k 平均法などの従来の領域分割法にはない特徴を持つ。すなわち、2つ目の領域分割法は、補正によって得られる多重露出画像の枚数を、高品質な L-WDR 画像の生成に適したものになるよう自動的に決定できる。これは、多くの多重露出画像を合成に用いることが、計算コストを増大させるものの、生成される L-WDR 画像の品質向上のために有効であるという事実に基づく。

実験では、主観的および定量的な比較により、既存の様々な合成法と提案法を組み合わせることの有効性を評価した。主観評価の結果、提案法の利用により、シーン全体を明瞭に表現した L-WDR 画像を生成できることが示される。加えて、提案法の利用の下で多重露出画像の合成を行うことは、4つの客観評価尺度、すなわち、MEF-SSIM, Discrete entropy, Tone mapped image quality index, および Statistical naturalness の観点から最先端の多重露出画像生成法より優れた性能を持つことが明らかとなる。

3.2 シーンの領域分割に基づく多重露出画像補正法

本節では、提案するシーン領域分割に基づく多重露出画像補正法 (Scene segmentation-based luminance adjustment, SSLA) について説明する。

3.2.1 多重露出画像の輝度補正の効果

まず、不明瞭な多重露出画像の輝度を補正することが、生成される L-WDR 画像の品質に影響することを示す。図 3.1 および 3.2 は、補正された多重露出画像の例と、それらの合成により生成された L-WDR 画像をそれぞれ示す。これらの図から、多重露出画像の輝度の補正によって、生成される L-WDR 画像の品質を向上できることがわかる。言い換えれば、撮影時に適した露出値で多重露出画像が撮影できない場合でも、それら不明瞭な多重露出画像の品質を後から向上できるといえる。また、最終的な L-WDR 画像の品質は、この補正の度合いに依存する。

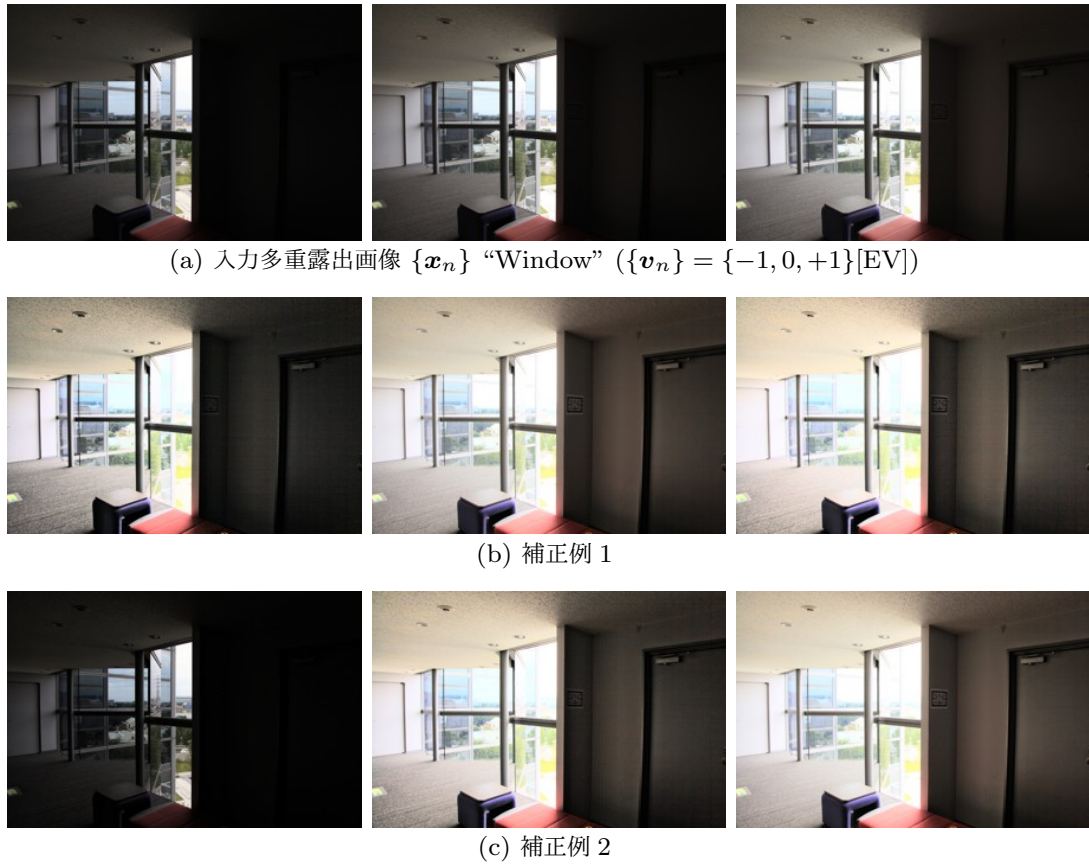


図 3.1 輝度を補正した多重露出画像の例. 輝度補正により, 多重露出画像の品質が向上する.



図 3.2 合成された L-WDR 画像の例. (a) 図 3.1(a) に示す画像を合成したもの. (b) 図 3.1(b) に示す画像を合成したもの. (c) 図 3.1(c) に示す画像を合成したもの. 合成された L-WDR 画像の品質は, 補正の度合いにより変化する.

3.2.2 提案法の概要

提案する SSLA は, 局所コントラスト強調, シーン領域分割に基づく輝度のスケールリング (Scene segmentation-based luminance scaling, SSLS), およびトーンマッピングとい

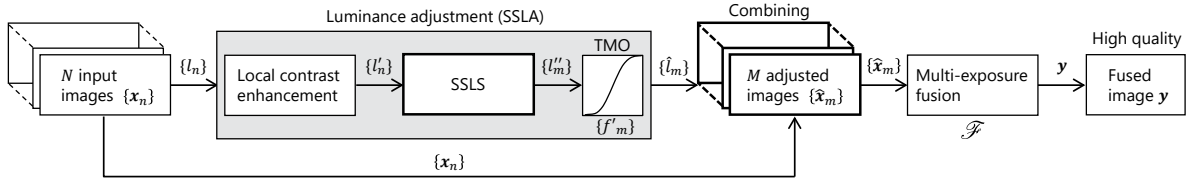


図 3.3 L-WDR 画像生成における提案法の使用法

う 3 つの処理からなる (図 3.3 参照). 本項では, 提案法の概要について説明し, SSLS の詳細については次節で述べる.

多重露出画像に基づく L-WDR 画像生成における提案法の使用法を図 3.3 に示す. 多重露出画像の品質向上のため, まず, n 番目の入力画像 x_n の輝度 l_n に対して局所コントラスト強調を行う. 次に, SSLS によって, 画素の集合 \mathbb{P} を M 個の領域 $\mathbb{P}_1, \mathbb{P}_2, \dots, \mathbb{P}_M$ に輝度値に基づき分割し, その後強調された輝度の集合 $\{l'_n\}$ をスケーリングすることで, 各領域 \mathbb{P}_m をそれぞれ明瞭に表す輝度 l''_m を得る. ここで, スケーリングにより得られる輝度の数 (補正により得られる多重露出画像の枚数) M は一般に, 入力多重露出画像の枚数 N とは異なる. 加えて, スケーリングされた輝度 l''_m にトーンマッピング処理を適用することによって, 画素値の切り捨てを回避する. 補正された多重露出画像 $\{\hat{x}_m\}$ は, トーンマッピングされた輝度 $\{\hat{l}_m\}$ と入力多重露出画像 $\{x_n\}$ を組み合わせることによって得られる. 最終的な L-WDR 画像 y は, L-WDR 画像生成法 \mathcal{F} を用いて, 補正によって得られた $\{\hat{x}_m\}$ を合成することで生成される.

図 3.3 において, 提案する SSLS は合成処理 \mathcal{F} の前に配置されなければならないことに注意する. さもなければ, 提案法は合成された L-WDR 画像から多重露出画像を生成し, 追加の合成処理が必要となる.

局所コントラスト強調

もし, 入力多重露出画像がシーンを明瞭に記録していない場合, それらのコントラストは明瞭な多重露出画像よりも低い. この理由から, Dodging and burning アルゴリズムを局所コントラスト強調のために利用する [23]. このアルゴリズムによって強調された輝度 l'_n は, 以下の式によって与えられる.

$$l'_n(\mathbf{p}) = \frac{l_n(\mathbf{p})^2}{\bar{l}_n(\mathbf{p})} \quad (3.1)$$

ここで, $\bar{l}_n(\mathbf{p})$ は, 輝度 $l_n(\mathbf{p})$ における画素 \mathbf{p} まわりの局所平均に相当する. 局所平均 $\bar{l}_n(\mathbf{p})$ は, ローパスフィルタを $l_n(\mathbf{p})$ に適用することで計算される. ここでは, バイラテ

ラルフィルタをこの目的で使用する.

バイラテラルフィルタを用いた $\bar{l}_n(\mathbf{p})$ の計算は次式の通りである.

$$\bar{l}_n(\mathbf{p}) = \frac{\sum_{\mathbf{p}' \in \mathbb{P}} l_n(\mathbf{p}') g_{\sigma_1}(\|\mathbf{p}' - \mathbf{p}\|) g_{\sigma_2}(l_n(\mathbf{p}') - l_n(\mathbf{p}))}{\sum_{\mathbf{p}' \in \mathbb{P}} g_{\sigma_1}(\|\mathbf{p}' - \mathbf{p}\|) g_{\sigma_2}(l_n(\mathbf{p}') - l_n(\mathbf{p}))} \quad (3.2)$$

ここで, $g_\sigma(t)$ は次式で与えられるガウス関数である.

$$g_\sigma(t) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{t^2}{2\sigma^2}\right) \text{ for } t \in \mathbb{R} \quad (3.3)$$

パラメータ $\sigma_1 = 16$ および $\sigma_2 = 3/255$ は, 文献 [23] に従い設定した. また, 実装には, 文献 [55] で示されているバイラテラルフィルタの実時間アルゴリズムを利用した.

トーンマッピング

スケーリングされた輝度 $l''_m(\mathbf{p})$ は, しばしば L-WDR 画像が仮定している画素値のレンジ $[0, 1]$ を超える値を持つ. したがって, これら値の切り捨てによる画素値の損失が発生する. この問題を防ぐため, トーンマッピング処理を利用して補正された輝度値が区間 $[0, 1]$ 内に収まるようにする.

補正された多重露出画像の輝度 $\hat{l}_m(\mathbf{p})$ は, トーンマッピング関数 f'_m を $l''_m(\mathbf{p})$ に適用することで

$$\hat{l}_m(\mathbf{p}) = f'_m(l''_m(\mathbf{p})) \quad (3.4)$$

として計算される. ここでは, Reinhard のグローバルオペレータをトーンマッピング関数 f'_m として利用する. [56]

Reinhard のグローバルオペレータは次式により与えられる.

$$f'_m(t) = \frac{t}{1+t} \left(1 + \frac{t}{L_m^2}\right) \text{ for } t \in \mathbb{R}_+ \quad (3.5)$$

ここで, パラメータ $L_m \in \mathbb{R}_{++}$ は, $f'_m(t) = 1$ となる t を定める. Reinhard のグローバルオペレータ f'_m は単調増加関数であることから, 各 m について $L_m = \max l''_m(\mathbf{p})$ と設定することにより, すべての \mathbf{p} について $\hat{l}_m(\mathbf{p}) \leq 1$ を満たす. したがって, 輝度値の切り捨てを防ぐことができる. 一方, トーンマッピングに起因する輝度値の変化が, 輝度値の切り捨てより好ましくない場合には, $L_m = 1$ を選択することができる. このように, この 2 つのパラメータ設定はそれぞれ長所と短所がある. 以降では, パラメータ $L_m = \max l''_m(\mathbf{p})$ を用いる.

補正された多重露出画像の合成

補正された多重露出画像 $\{\hat{\mathbf{x}}_m\}$ は, 多重露出画像に基づく任意の L-WDR 画像生成法 [8, 46] の入力として利用できる. 最終的に得られる L-WDR 画像 \mathbf{y} は, M 枚の画像 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ を単一 L-WDR 画像へ合成する関数 $\mathcal{F}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$ を用いて

$$\mathbf{y} = \mathcal{F}(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_M) \quad (3.6)$$

として得られる.

3.2.3 シーン領域分割に基づく輝度のスケーリング

提案する SSLS は, 図 3.3 に示すように, 局所コントラスト強調によって得られる輝度の集合 $\{l'_n\}$ に対して適用される. SSLS の目的は, シーン中のある特定の明るさを持つ領域 $\mathbb{P}_m \subset \mathbb{P}$ を明瞭に表す輝度 l''_m を生成することである. この目的のため, SSLS では, シーン領域分割および輝度のスケーリングという 2 つの処理を行う (図 3.4 参照). 前者の処理では, 全画素の集合 \mathbb{P} を M 個の部分集合 $\{\mathbb{P}_1, \mathbb{P}_2, \dots, \mathbb{P}_M\}$ へ分割する. 後者の処理では, 輝度 l''_m が領域 \mathbb{P}_m を明瞭に表すよう, 集合 $\{l''_m\}$ を生成する. ここで, このスケーリング処理は, \mathbb{P}_m のみならず集合 \mathbb{P} 上のすべての輝度値 $l'_n(\mathbf{p})$ について適用することに注意する. これにより, スケーリング前後で画素値の大小関係を保つことができる.

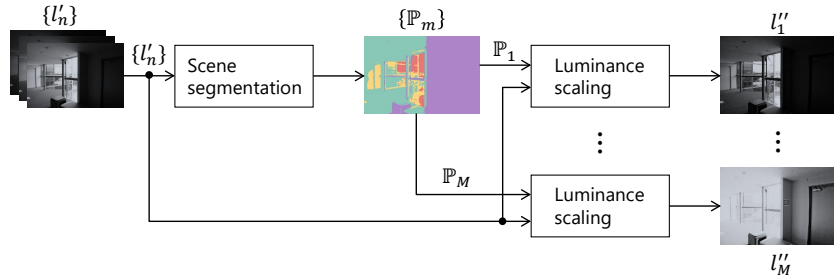
シーン領域分割

ここでは, 入力多重露出画像のシーンをそれぞれがある特定の明るさを持つ領域 $\mathbb{P}_1, \dots, \mathbb{P}_M$ へ分割する方法を議論する. ただし, 分割された領域は $\mathbb{P}_1 \cup \mathbb{P}_2 \cup \dots \cup \mathbb{P}_M = \mathbb{P}$ を満たすものとする. この処理は, 輝度に基づく画像領域分割法の一つだと考えられる. 一方, 多重露出画像に基づく L-WDR 画像生成のためのシーン領域分割は, 意味的領域分割などの一般的な領域分割問題と 2 つの観点から異なる [57, 58].

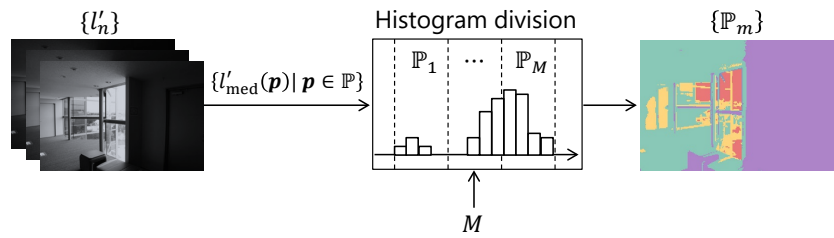
- 複数の入力画像が与えられること
- 輪郭等の画像構造には注目する必要がないこと

シーン領域分割のため, ここでは 2 つの Approach を提案する. 計算量が少ない Approach 1 では, 入力多重露出画像中で中間の明るさを持つ LDR 画像の輝度に従い, シーンを分割する. 高品位な出力が得られる Approach 2 では, すべての入力画像の輝度の分布を考慮してシーンを分割する.

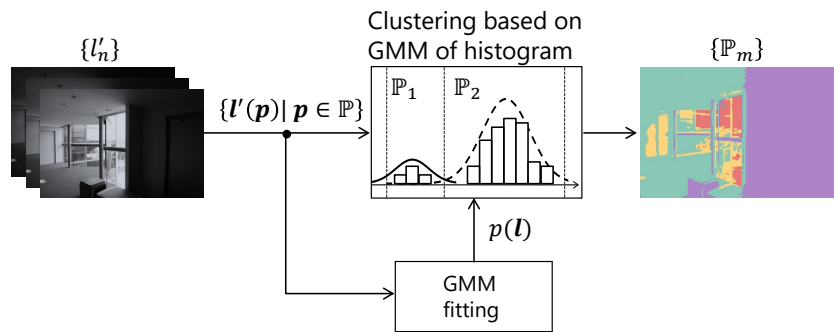
Approach 1: l'_{med} を, 集合 $\{l'_n\}$ において中間の明るさを持つ輝度とおく. l'_{med} にお



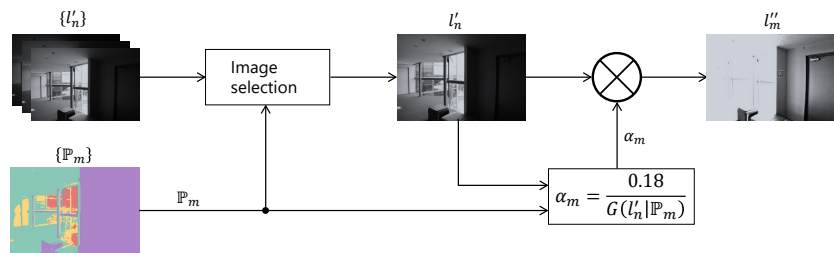
(a) SSSLs の処理フロー



(b) シーン領域分割 (Approach 1)



(c) シーン領域分割 (Approach 2)



(d) 輝度のスケールリング

図 3.4 提案するシーン領域分割に基づく輝度のスケールリング (SSLS) (a) SSSLs の概要. (b) と (c) 図 3.4(a) におけるシーン領域分割の 2 つの Approach. (d) 図 3.4(a) における輝度のスケールリング.

ける露出過多 (白飛び) あるいは露出不足 (黒潰れ) の領域は, $\{l'_n\}$ 中の他の輝度のものより小さい. すなわち, l'_{med} は, $\{l'_n\}$ 中で最も良い品質を持つものと考えられる. したがって, Approach 1 では, l'_{med} をシーン領域分割のために利用する.

Approach 1 において, $\mathbb{P}_1, \dots, \mathbb{P}_M$ は, l'_{med} の輝度のレンジを M 等分することで, 次式として与えられる.

$$\mathbb{P}_m = \{\mathbf{p} | \theta_m \leq l'_{\text{med}}(\mathbf{p}) \leq \theta_{m+1}\} \quad (3.7)$$

ここで, θ_m は

$$\theta_m = \frac{N - m + 1}{N} (\max l'_{\text{med}}(\mathbf{p}) - \min l'_{\text{med}}(\mathbf{p})) + \min l'_{\text{med}}(\mathbf{p}) \quad (3.8)$$

である.

Approach 1 はとても単純な手法であるが, 後で示されるように, 多くの入力多重露出画像について有効である.

Approach 2: Approach 1 は入力多重露出画像中の 1 枚の画像のみをシーン領域分割に使う一方で, Approach 2 は, すべての入力多重露出画像の輝度の分布を考慮する. これにより, Approach 1 がうまく働かない場合でさえ, Approach 2 によって適切にシーン領域分割を実行することが可能となる. Approach 2 では, 多重露出画像の輝度の分布をモデル化するために, 混合ガウス分布を利用する. その後, 混合ガウスモデル (GMM) に基づくクラスタリング法を利用して, 画素のクラスタリングを行う [59].

すべての入力画像の輝度値を考慮したモデルを作成するために, 画素 \mathbf{p} における多重露出画像の輝度値を N 次元ベクトル $\mathbf{l}'(\mathbf{p}) = (l'_1(\mathbf{p}), l'_2(\mathbf{p}), \dots, l'_N(\mathbf{p}))^\top$ として考える. $\mathbf{l}'(\mathbf{p})$ の分布は, GMM を用いて

$$p(\mathbf{l}'(\mathbf{p})) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{l}'(\mathbf{p}) | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (3.9)$$

としてモデル化される. ここで, K は混合数であり, π_k は混合係数である. また, $\mathcal{N}(\mathbf{l}'(\mathbf{p}) | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ は, 平均 $\boldsymbol{\mu}_k$, 共分散行列 $\boldsymbol{\Sigma}_k$ を持つ N 次元ガウス分布である.

GMM を与えられた $\mathbf{l}'(\mathbf{p})$ へ適合させるために, 変分ベイズ推論を利用する [59]. 伝統的な最尤推定と比較した変分ベイズ推論の長所は, 大きな K を与えた場合でさえ, モデルの過適合を回避できることである. この理由から, 大きな K を推論に利用した場合には, 不必要な混合要素は自動的に取り除かれる. 本研究では, 分割数 M の最大値として $K = 10$ を利用する.

ここで, 1-of- K 表現で表される (すなわち, K 個の要素のうち 1 つの要素のみ 1 であり, 他の要素がすべて 0 である) K 次元の 2 値乱数 $\mathbf{z} = (z_1, z_2, \dots, z_K)^\top$ を導入する. この定義より, $z_k (k \in \{1, \dots, K\})$ は, $z_k \in \{0, 1\}$ および $\sum_k z_k = 1$ を満たす. \mathbf{z} の上の結合分布は, 混合係数 π_k を用いて次式のように書ける.

$$p(z_k = 1) = \pi_k \quad (3.10)$$

$p(z_k = 1)$ が確率の公理を満たすために, $\{\pi_k\}$ は,

$$0 \leq \pi_k \leq 1 \quad (3.11)$$

および

$$\sum_{k=1}^K \pi_k = 1 \quad (3.12)$$

を満たさなければならない. 画素 \mathbf{p} を含む領域 \mathbb{P}_m は, 以下の条件付き確率として与えられる $\gamma(z_k | \mathbf{l}'(\mathbf{p}))$ により決定される.

$$\begin{aligned} \gamma(z_k | \mathbf{l}'(\mathbf{p})) &= p(z_k = 1 | \mathbf{l}'(\mathbf{p})) \\ &= \frac{\pi_k \mathcal{N}(\mathbf{l}'(\mathbf{p}) | \boldsymbol{\mu}_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{l}'(\mathbf{p}) | \boldsymbol{\mu}_j, \Sigma_j)} \end{aligned} \quad (3.13)$$

画素 $\mathbf{p} \in \mathbb{P}$ が与えられ, m が

$$m = \arg \max_k \gamma(z_k | \mathbf{l}'(\mathbf{p})) \quad (3.14)$$

を満たすとき, \mathbf{p} は \mathbb{P} の部分集合 \mathbb{P}_m の要素として割り当てられる.

輝度のスケーリング

スケーリングされた輝度 l''_m は, 次式として与える

$$l''_m(\mathbf{p}) = \alpha_m l'_n(\mathbf{p}) \quad (3.15)$$

ここで, パラメータ $\alpha_m > 0$ は補正の度合いを表す. この輝度のスケーリングは, カメラの特性を表す関数 f が線形であると仮定して, 露出を擬似的に変更することに相当する (第 2.2 節参照).

\mathbb{P} の部分集合 \mathbb{P}_m が与えられたとき, 領域 \mathbb{P}_m の全体的な明るさは, \mathbb{P}_m 上の輝度の幾何平均として計算できる. この理由から, 補正された多重露出画像の輝度 $l''_m(\mathbf{p})$ を, \mathbb{P}_m 上の輝度の幾何平均が 0.18 (中間のグレー) となるように求める [56].

式 (2.1) を用いて, パラメータ α_m を次式として計算する.

$$\alpha_m = \frac{0.18}{G(l'_n | \mathbb{P}_m)} \quad (3.16)$$

ここで, α_m の値が大きい場合には, 強い補正がかかり, それに伴い画像中のノイズも増幅されてしまう. このことから, できるだけ小さい α_m を選択する目的で, n を

$$n = \psi(m) = \arg \min_j (0.18 - G(l'_j | \mathbb{P}_m))^2 \quad (3.17)$$

として選択する．式 (3.15) にパラメータ α_m と n を代入することで計算される輝度 l''_m は，3.2.2 で述べたトーンマッピング処理の入力として用いられる．

3.2.4 多重露出画像の補正

補正された多重露出画像 $\{\hat{\mathbf{x}}_m\}$ は，SSLA によって補正された輝度の集合 $\{\hat{l}_m\}$ と入力多重露出画像 $\{\mathbf{x}_n\}$ を組み合わせることで得られる．ここで，各 \hat{l}_m と入力画像 \mathbf{x}_n を関連付けるために，式 (3.17) を利用する．補正された多重露出画像 $\hat{\mathbf{x}}_m$ は， \hat{l}_m ， $\psi(m)$ 番目の入力画像 $\mathbf{x}_{\psi(m)}$ ，およびその輝度 $l_{\psi(m)}$ を用いて，

$$\hat{\mathbf{x}}_m(\mathbf{p}) = \frac{\hat{l}_m(\mathbf{p})}{l_{\psi(m)}(\mathbf{p})} \mathbf{x}_{\psi(m)}(\mathbf{p}) \quad (3.18)$$

として与えられる．

3.2.5 実行手順

提案法により， N 枚の入力多重露出画像 $\{\mathbf{x}_n\}$ から L-WDR 画像 \mathbf{y} を生成する手順を以下に示す (図 3.3 参照)．

- i 各入力 LDR 画像 \mathbf{x}_n から輝度 l_n を計算する．
- ii 式 (3.1) から (3.3) を用いて $\{l_n\}$ の局所コントラストを強調し，強調された輝度 $\{l'_n\}$ を得る．
- iii \mathbb{P} を M 個の領域 $\{\mathbb{P}_m\}$ へ分割する．
 Approach 1: 式 (3.7) および (3.8) を利用
 Approach 2: 式 (3.9) から (3.14) を利用
- iv 式 (3.15) から (3.17) を用いて $\{l''_m\}$ を計算する．
- v 式 (3.4) と (3.5) に従い， $\{l''_m\}$ を $\{\hat{l}_m\}$ へトーンマッピングする．
- vi 式 (3.18) により， $\{\hat{\mathbf{x}}_m\}$ を計算する．
- vii 多重露出画像に基づく L-WDR 画像生成法 \mathcal{F} を用いて，式 (3.6) のように L-WDR 画像 \mathbf{y} を生成する．

ここで，Approach 1 と Approach 2 の違いは手順 iii のみであることに注意する．また，領域 \mathbb{P}_m の数 M は $1 \leq M \leq K$ を満たす．提案法における調整可能なパラメータを表 3.1 に示す．

表 3.1 提案法における調整可能なパラメータ

パラメータ	効果
σ_1, σ_2	局所コントラスト強調の度合いを調整する。ただし、大きな値は、ノイズやリングングを増幅させる場合がある。
L_m	トーンマッピングオペレータにおける白レベルを決定する。 $L_m \leq 1$: 低い輝度値を白レベルとして、階調を引き伸ばす。 $L_m = 1$: すべての輝度を保つ。 $1 \leq L_m$: 高い輝度値を白レベルとして、階調を圧縮する。
M, K	補正後の多重露出画像の枚数を決定する。大きい M の使用は、生成される画像の品質をより向上させる一方、計算コストを増大させる。 Approach 2 において、 M は、 $1 \leq M \leq K$ を満たす値として自動的に決定される。

3.3 実験および考察

提案する SSLA の有効性を、生成される L-WDR 画像 y の品質の観点から評価した。

3.3.1 実験条件

本実験では、カメラによる撮影で直接得られた 20 組の多重露出画像と、トーンマッピングにより生成された 550 組の多重露出画像を入力多重露出画像 $\{x_n\}$ として使用した。カメラにより撮影された多重露出画像中で、4 組は Canon EOS 5D Mark II を用いて撮影し、8 組はオンライン上で利用可能なデータベース [60] から選択した。これら 12 組の多重露出画像はそれぞれ、正の露出値, 0[EV], 負の露出値で撮影された 3 枚の LDR 画像からなる (図 3.5 および 3.6 参照)。残る 8 組の多重露出画像は、データベース [13] から選択した被写体のずれを含む多重露出画像である。トーンマッピングに基づく 550 組の多重露出画像のデータセットは、オンライン上で利用可能なデータベース [61, 62] から収集した 50 枚の U-WDR 画像をトーンマッピングすることによって作成した。各 U-WDR

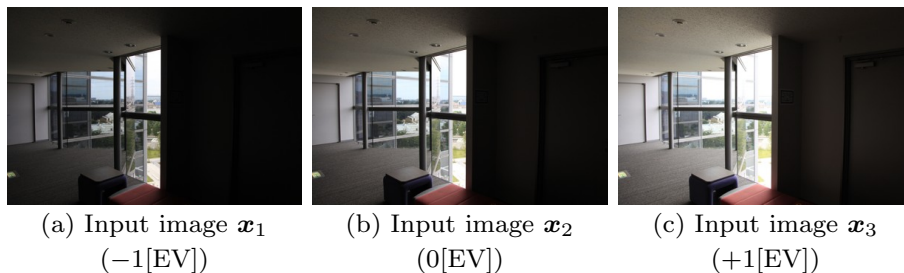


図 3.5 入力多重露出画像 “Window”. シーンの右側の領域がすべての画像で不明瞭である.

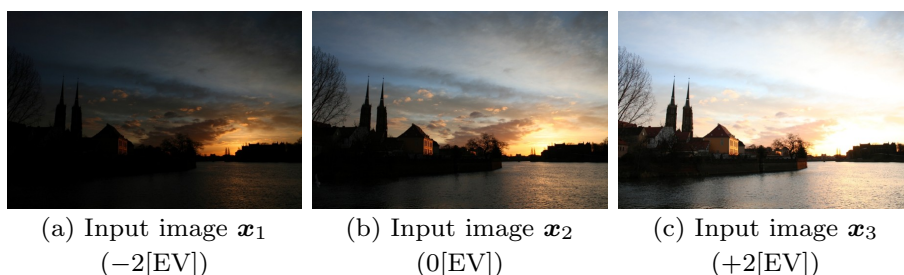


図 3.6 入力多重露出画像 “Ostrow Tumski”. タワーと右側の陸地がすべての画像で不明瞭である.

画像からは、関数 f として線形関数を用いて、それぞれ 11 組の多重露出画像を生成した。ここで、この 11 組のうち 10 組は不明瞭な多重露出画像を想定し、多重露出画像の枚数およびそれら各画像の露出値を、区間 $[2, 5]$ および区間 $[-7, 0]$ 上の乱数を用いてそれぞれ決定した。残る 1 組の多重露出画像は、露出値 $\{-7, -6, \dots, 6, 7\}$ を持つ 15 枚の LDR 画像からなる。

提案法の有効性の評価は以下の手順で実施した。

- i $\{\mathbf{x}_n\}$ から、提案法を用いて $\{\hat{\mathbf{x}}_m\}$ を生成する。
- ii \mathcal{F} を用いて $\{\hat{\mathbf{x}}_m\}$ を合成し、 \mathbf{y} を得る。
- iii 3つの客観評価尺度、すなわち、MEF-SSIM, Discrete entropy, Tone mapped image quality index (TMQI) を用いて、後述の通り \mathbf{y} の品質を評価する。

ここで、合成法 \mathcal{F} として、Mertens らの方法 [8], Sakai らの方法 [45], Nejati らの方法 [46], Li らの方法 [63], および Ma らの方法 [15] の 5 つを利用した。Approach 1 においては、 M を入力多重露出画像の枚数 N と等しくなるように設定した。Approach 2 では、 M は、 $1 \leq M \leq K = 10$ の条件の下、各シーンのために構築された GMM によって自動的に決定される。GMM の構築は、パラメータの収束状況に関わらず、変分ベイズ

推論のイテレーション数が 100 に到達した時点で打ち切った。加えて、処理の高速化のため、Approach 2 による領域分割は $\{l'_n\}$ を縮小したものに対して適用した。 l'_n を縮小した後の幅 U' および高さ V' は、 $\max(U', V') = 256$ となるように設定した。

3.3.2 主観評価

図 3.7 および 3.8 は、図 3.5 に示す入力多重露出画像 $\{x_n\}$ に対して提案する SSLA を使用して、あるいは、使用せずに合成された L-WDR 画像を示す。図 3.7(a) から (c)、図 3.8(a)、および図 3.8(b) より、輝度補正を用いないすべての合成法は、不明瞭な入力多重露出画像が与えられた際に、画像の暗い領域を明瞭に表すために十分な効果がないことがわかる。一方、提案する SSLA を用いて合成された画像は、明るい領域の品質を保ったまま暗い領域を明瞭に表している [図 3.7(d) から (i)、および図 3.8(c) から (f) 参照]。しかしながら、提案する SSLA を Li らの手法に組み合わせた場合には、シーンの明るい領域と暗い領域における輝度の逆転が発生した。この輝度の逆転は、最新の手法である Nejadi ら、および Ma らの手法それぞれに対し、提案法を組み合わせた場合には発生しなかった。図 3.6 に示す多重露出画像についての実験結果を図 3.9 に示す。図 3.9 は、図 3.7 と同様の傾向を示している。したがって、提案する SSLA は、入力多重露出画像が不明瞭な場合でさえ、それらの合成によって得られる L-WDR 画像の品質を向上させることが可能であるといえる。

Approach 1 と Approach 2 を比較すると、Approach 2 は、図 3.7 および 3.9 に示すように、暗い領域をよりはっきりと表す画像を生成することができる。この手法間の差異は、分割された領域 $\{P_m\}$ の違いに由来する。図 3.10 および 3.11 に、Approach 1 および 2 によって分割された領域 $\{P_m\}$ と、各 $\{P_m\}$ に基づいて補正された多重露出画像を示す。図 3.10(a) より、計算コストの低い Approach 1 は、画面右側のドアなどの暗い領域と、左側の床などの中間の明るさを持つ領域を分割することができなかったことがわかる。この理由から、図 3.10(c) における最も明るい画像は、シーンの最も暗い領域を明瞭に表すために十分な明るさを持たない。それに対し、Approach 2 は、大きな K が与えられた場合にも、シーンをある特定の輝度レンジを持つそれぞれの領域へ適切に分割できている。図 3.11 は、図 3.10 とほとんど同じ結果を示した。したがって、Approach 2 は、閉形式で計算される Approach 1 より、高い精度でシーンを分割できることが確認された。

また、図 3.10 と 3.11 は、Approach 1 および 2 の両方が、画像の構造を考慮していないにもかかわらず、それら構造を保持したまま補正を実行できることを示している。この結果は、提案する SSLA のためには、エッジなどの画像構造に関する情報を必要としない



図 3.7 提案法を Mertens ら, Sakai ら, Nejati らの手法と組み合わせた場合の比較 (“Window”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. 従来の多重露出画像の補正を行わない合成法は, 図 3.5 に示す不明瞭な多重露出画像から明瞭な画像を生成できない一方で, 提案法は明瞭な画像の生成を可能とする.

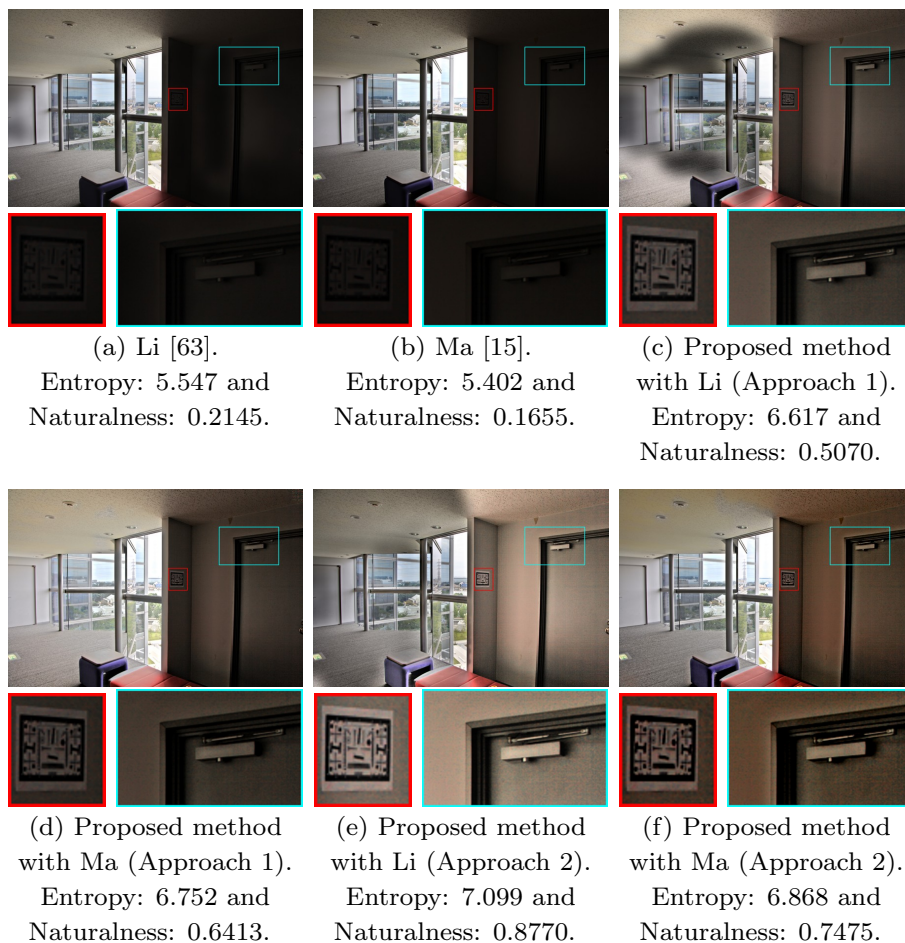


図 3.8 提案法を Li らおよび Ma らの手法と組み合わせた場合の比較 (“Window”). 枠で囲われた領域の拡大図は、各合成画像の下に示す. 従来の多重露出画像の補正を行わない合成法は、図 3.5 に示す不明瞭な多重露出画像から明瞭な画像を生成できない一方で、提案法は明瞭な画像の生成を可能とする.

ことを表している. 図 3.12 に、提案法における局所コントラスト強調の効果を示す. この強調処理は、ぼけの除去や画像構造を明瞭にすることに効果があるが、しばしば画像中のノイズやリングングを増幅させる. このような場合には、局所コントラスト強調なしの SSLA を用いることが、ノイズの増幅を防ぐために効果的である.

被写体の位置のずれを含む入力多重露出画像から生成された L-WDR 画像を、図 3.13 に示す. 入力多重露出画像に被写体の位置ずれが含まれる場合でも、合成法 \mathcal{F} として動的なシーンに頑健な手法を用いることで、提案法はゴーストアーティファクトのなしに高品質な画像を生成できる. 一方、提案法なしで合成された画像は、わずかにゴーストアーティファクトを含む. この結果から、提案法は、明瞭な L-WDR 画像を生成することを可

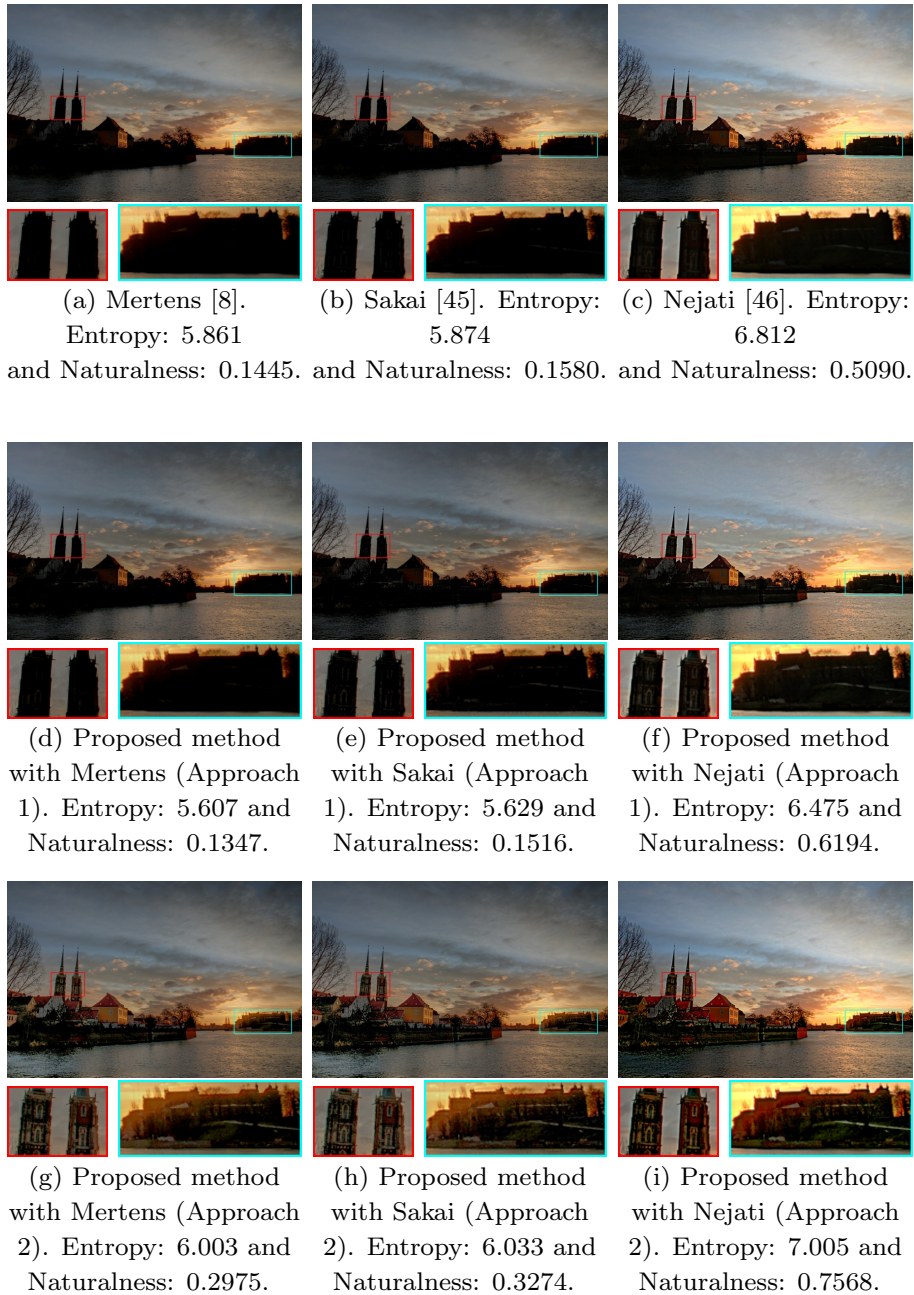


図 3.9 提案法を Mertens ら, Sakai ら, Nejati らの手法と組み合わせた場合の比較 (“Ostrow Tumski”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. 従来の多重露出画像の補正を行わない合成法は, 図 3.6 に示す不明瞭な多重露出画像から明瞭な画像を生成できない一方で, 提案法は明瞭な画像の生成を可能とする.

能にするのみならず, ゴーストアーティファクトの除去も可能にするといえる.

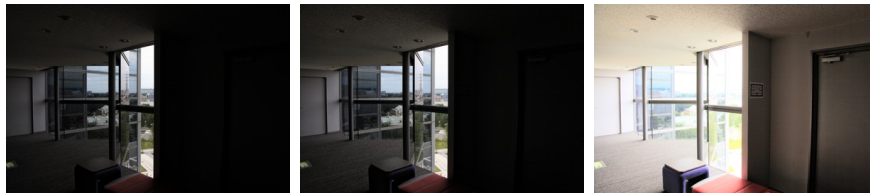
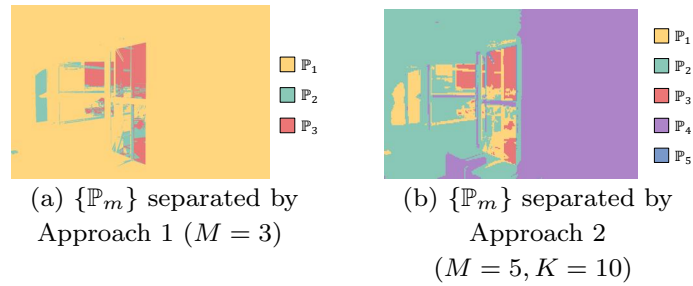
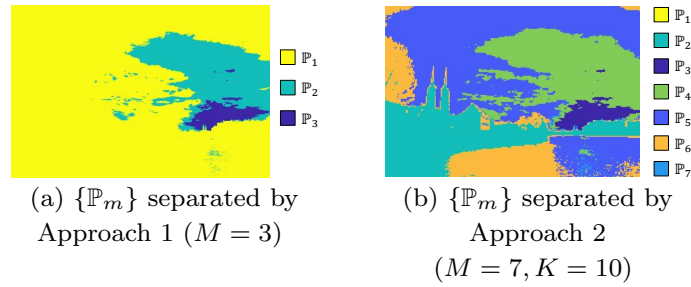
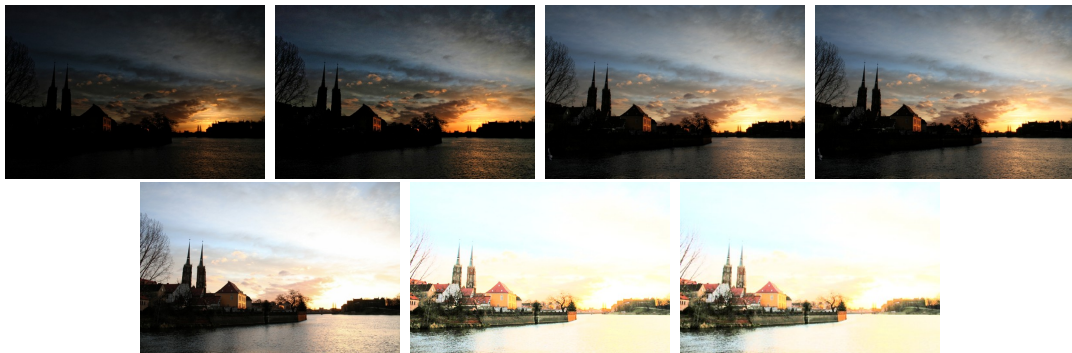


図 3.10 提案する SSLA によって補正された多重露出画像 (“Window”). Approach 1 は, ドアと床などの暗い領域と中間の明るさを持つ領域を分割できなかった. 一方, Approach 2 は, 大きな K が与えられた場合にも, ある輝度のレンジを持つ領域へシーンを適切に分割することができた.



(c) Images $\{\hat{x}_m\}$ adjusted by using $\{\mathbb{P}_m\}$ in Fig. 3.11(a) (Approach 1)



(d) Images $\{\hat{x}_m\}$ adjusted by using $\{\mathbb{P}_m\}$ in Fig. 3.11(b) (Approach 2)

図 3.11 提案する SSLA によって補正された多重露出画像 (“Ostrow Tumski”). Approach 1 は、空と水辺を分割することでできなかったが、Approach 2 は、大きな K が与えられた場合でさえ、シーンのある輝度のレンジを持つ領域へ適切に分割できた。



(a) Approach 2 with contrast enhancement.
Entropy: 6.848 and Naturalness: 0.7031.

(b) Approach 2 without contrast enhancement.
Entropy: 6.810 and Naturalness: 0.5943.

図 3.12 提案法における局所コントラスト強調の効果 (“Window”). 枠で囲われた領域の拡大図は、各合成画像の下に示す。ここで、これら画像の生成には、Nejati らの手法 [46] と提案する Approach 2 を用いた。局所コントラスト強調は画像構造を明瞭にすることができる一方で、しばしばノイズやリングングを増幅させる。



(a) Ma [15]. Entropy: 6.917
and Naturalness: 0.8262.

(b) Proposed method with Ma
(Approach 1). Entropy: 6.653
and Naturalness: 0.6731.



(c) Proposed method with Ma
(Approach 2). Entropy: 7.157
and Naturalness: 0.9701.

(d) Proposed method with Ma
(Approach 2 without local
contrast enhancement).
Entropy: 7.000
and Naturalness: 0.7076.

図 3.13 提案法を Ma らの手法と組み合わせた場合の比較 (“Lady eating”). 枠で囲われた領域の拡大図は、各合成画像の下に示す。提案する SSLA は、動的なシーンに対しても有効である。

3.3.3 定量評価

合成された L-WDR 画像の品質を定量的に比較するためには、画像品質の客観評価尺度が必要とされる。ピーク信号対雑音比 (Peak signal to noise ration, PSNR) や Structural similarity index (SSIM) は、劣化のない高品質な画像を参照画像として用いるため、本実験のためには適していない。近年、多重露出画像に基づく L-WDR 画像生成法のための Full-reference metrics がいくつか提案されている [64, 65]。しかしながら、それら尺度も高品質な多重露出画像を必要とするため、高品質な多重露出画像が存在しない場合における提案法の性能評価には適していない。したがって、本実験では、MEF-SSIM と呼ばれる尺度 [64] を、高品質な多重露出画像が利用可能な場合にのみ用いた。他の場合には、輝度値の Discrete entropy および Tone mapped image quality index (TMQI) [66] を客観評価尺度として利用した。

MEF-SSIM は、Multi-scale SSIM フレームワークと、画像パッチの Consistency を評価する尺度に基づき、画像局所構造の保存性と大域的な輝度の一貫性をバランス良く評価できる。Discrete entropy は画像に含まれる情報量を測る尺度であり、その計算には、生成された画像 \mathbf{y} の輝度値を用いた。輝度の Discrete entropy は、画像の大域コントラストを示す。TMQI は、U-WDR 画像からトーンマッピングされた画像の品質を評価する尺度であり、Structural fidelity および Statistical naturalness の 2 つの尺度から構成される。Statistical naturalness [66] は、

$$S = \frac{1}{O} \mathcal{N}(\bar{l} | 115.94, 27.99^2) \mathcal{B}(\sigma_l / 64.29 | 4.4, 10.1) \quad (3.19)$$

として定義される。ここで、 O は正規化パラメータであり、 $\mathcal{B}(\cdot | \alpha, \beta)$ は、パラメータ α および β を持つベータ分布を表す。また、 \bar{l} は $[0, 255]$ のスケールで表現される画像の平均輝度を表し、 σ_l はそのスケールでの輝度の局所分散の平均を意味する。これらガウス分布とベータ分布はそれぞれ、画像の大域的な明るさの自然さと局所コントラストの自然さを表す。Structural fidelity は U-WDR 画像を参照画像として必要とするが、Statistical naturalness は参照画像なしで計算できる。トーンマッピング処理と画像撮影処理は類似しているため、TMQI は、カメラで撮影された画像の評価にも有用である。この実験では、トーンマッピングにより生成された多重露出画像を入力として用いた場合のみ、MEF-SSIM と TMQI を評価のために利用した。Discrete entropy および Statistical naturalness は、570 組すべての場合で評価に利用した。

表 3.2 および 3.3 は、静的なシーンを撮影した 12 組の多重露出画像を入力として用いた

場合に，生成された画像を Discrete entropy と Statistical naturalness によって評価した結果を示す．それぞれのスコア (Discrete entropy $\in [0, 8]$ および Statistical naturalness $\in [0, 1]$) は，大きい値ほどより高い品質であることを示す．表 3.2 より，代表的な合成法は “Corridor 1” および “Lobby” について低い Entropy スコアを示した一方，提案する SSLA は，12 組すべての入力多重露出画像について高い Entropy スコアを示したことが確認できる．加えて，SSLA は，代表的な合成法よりも高い平均スコアを持つことがわかる．表 3.3 は，代表的な合成法と比較して，提案する SSLA がほとんどの場合において，Statistical naturalness の観点から高品質な画像を生成したことを示す．

表 3.2 静的なシーンに対する Discrete entropy スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する.

Scene	Mertens [8]		Sakai [45]		Nejati [46]		Li [63]		Ma [15]				
	w/o	Prop. 1	Prop. 1	Prop. 2	w/o	Prop. 1	Prop. 2	w/o	Prop. 1	Prop. 2			
Arno	5.987	5.931	5.833	5.844	7.016	7.016	6.793	6.312	6.095	6.072	7.054	7.100	6.813
Cave	4.938	5.920	6.277	4.946	5.942	6.290	7.044	6.115	6.840	6.999	6.082	6.782	7.043
Chinese garden	6.142	6.185	6.241	6.155	6.205	6.258	6.844	7.244	7.271	7.229	6.998	7.143	6.955
Corridor 1	2.858	5.840	5.738	2.856	5.859	5.752	6.976	3.350	7.242	7.262	2.773	7.088	7.016
Corridor 2	5.706	5.922	5.613	5.714	5.933	5.618	6.625	6.682	5.548	5.670	6.533	6.930	6.586
Estate rsa	6.577	6.604	6.585	6.560	6.584	6.573	7.014	7.001	7.015	6.914	6.802	6.921	7.014
Kluki	7.120	6.982	6.942	7.136	7.003	6.964	7.151	7.669	7.540	7.548	7.535	7.266	7.207
Laurenziana	6.672	6.478	6.599	6.678	6.490	6.609	6.898	7.585	7.239	7.417	7.267	7.091	7.008
Lobby	4.172	5.648	5.737	4.177	5.672	5.743	6.836	5.274	6.773	7.177	4.893	7.060	6.783
Mountains	6.920	6.480	6.571	6.912	6.492	6.586	6.846	6.491	6.233	6.339	6.796	6.781	6.838
Ostrow tumski	5.861	5.607	6.003	5.874	5.629	6.033	7.005	6.607	6.253	6.794	6.857	6.622	6.958
Window	5.047	5.715	6.060	5.052	5.735	6.072	6.848	5.547	6.617	7.099	5.402	6.752	6.868
Average	5.667	6.109	6.183	5.671	6.123	6.195	6.907	6.323	6.722	6.877	6.249	6.961	6.924

表 3.3 静的なシーンに対する Statistical naturalness スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する.

Scene	Mertens [8]		Sakai [45]		Nejati [46]		Li [63]		Ma [15]						
	w/o	Prop. 1	Prop. 2	w/o	Prop. 1	Prop. 2	w/o	Prop. 1	Prop. 2	w/o	Prop. 1	Prop. 2			
Arno	0.1289	0.1382	0.1233	0.1506	0.1633	0.1473	0.5164	0.5372	0.5442	0.4079	0.2852	0.3543	0.5076	0.6845	0.5851
Cave	0.0758	0.2458	0.2808	0.0758	0.2236	0.2512	0.4223	0.3413	0.0628	0.2665	0.2230	0.0347	0.4122	0.3261	0.0758
Chinese garden	0.420	0.4353	0.4612	0.4059	0.4222	0.4441	0.4609	0.4843	0.4152	0.4801	0.5316	0.4477	0.4771	0.4468	0.3863
Corridor 1	0.0004	0.1454	0.1374	0.0004	0.1998	0.1713	0.0006	0.9111	0.8792	0.0012	0.9017	0.9129	0.0006	0.8908	0.8825
Corridor 2	0.0628	0.1025	0.0744	0.0763	0.1338	0.0911	0.3368	0.5052	0.4459	0.3807	0.0957	0.1293	0.3212	0.5209	0.4632
Estate rsa	0.7919	0.8047	0.7992	0.7624	0.7755	0.7840	0.8884	0.9504	0.9872	0.9773	0.9894	0.9941	0.9006	0.9536	0.9903
Kluki	0.9665	0.9580	0.9105	0.9596	0.9432	0.8889	0.9316	0.9844	0.9749	0.5148	0.7616	0.7226	0.8659	0.9849	0.9940
Laurenziana	0.8046	0.6541	0.7387	0.8058	0.6595	0.7366	0.9330	0.9722	0.9558	0.6125	0.9408	0.8303	0.8902	0.9662	0.9319
Lobby	0.0167	0.1947	0.2093	0.0190	0.2307	0.2274	0.1019	0.9260	0.7637	0.1350	0.6566	0.6936	0.1025	0.8888	0.7285
Mountains	0.3953	0.5582	0.6231	0.3562	0.5795	0.6353	0.4702	0.7383	0.7634	0.4852	0.5860	0.6898	0.4373	0.7459	0.8171
Ostrow tumski	0.1445	0.1347	0.2975	0.1580	0.1516	0.3274	0.5090	0.6194	0.7568	0.4874	0.3090	0.8185	0.5145	0.7585	0.8162
Window	0.0569	0.0912	0.2135	0.0615	0.1142	0.2650	0.1617	0.6187	0.7031	0.2145	0.5070	0.8770	0.1655	0.6413	0.7475
Average	0.3220	0.3719	0.4057	0.3193	0.3831	0.4141	0.4777	0.7157	0.6877	0.4136	0.5657	0.6254	0.4663	0.734	0.7015

表 3.4 動的なシーンに対する Discrete entropy スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す.

Scene	w/o	Prop. 1	Prop. 2	Prop. 2 w/o CE
Baby at window	7.032	6.351	7.160	6.770
Baby on grass	7.072	6.711	7.271	6.966
Christmas rider	7.098	6.366	7.048	6.945
Feeding time	6.498	6.853	6.643	6.621
High chair	7.272	6.861	7.163	7.149
Lady eating	6.917	6.653	7.157	7.000
Piano man	6.633	6.632	6.791	6.752
Santas little helper	7.186	6.429	7.246	7.285
Average	6.964	6.607	7.060	6.936

表 3.5 動的なシーンに対する Statistical naturalness スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す.

Scene	w/o	Prop. 1	Prop. 2	Prop. 2 w/o CE
Baby at window	0.9912	0.5344	0.9830	0.9438
Baby on grass	0.9779	0.6881	0.8532	0.8629
Christmas rider	0.8797	0.5972	0.7192	0.9593
Feeding time	0.1699	0.7162	0.6393	0.2859
High chair	0.9549	0.6721	0.7856	0.8372
Lady eating	0.8262	0.6731	0.9701	0.7076
Piano man	0.6369	0.6016	0.2942	0.7637
Santas little helper	0.6302	0.3866	0.3271	0.9089
Average	0.7584	0.6087	0.6965	0.7837

表 3.4, 3.5 は, 動的なシーンを撮影した 8 組の入力多重露出画像に対する Discrete entropy スコア, および Statistical naturalness スコアをそれぞれ示す. ここで, 合成法 \mathcal{M} には, Ma らの手法を用いた. これらの入力多重露出画像は十分な数の画像からなるため, 提案法の有無による大きなスコアの違いは確認できなかった.

図 3.14 は, トーンマッピングにより生成された不明瞭な画像からなる 500 組の多重露出画像に対して, Discrete entropy, TMQI, Statistical naturalness, および MEF-SSIM を用いて品質評価を行った結果を箱ひげ図として示す. 各箱は, 第一四分位数 Q_1 から第

第 3 章 多重露出画像の補正による L-WDR 画像生成

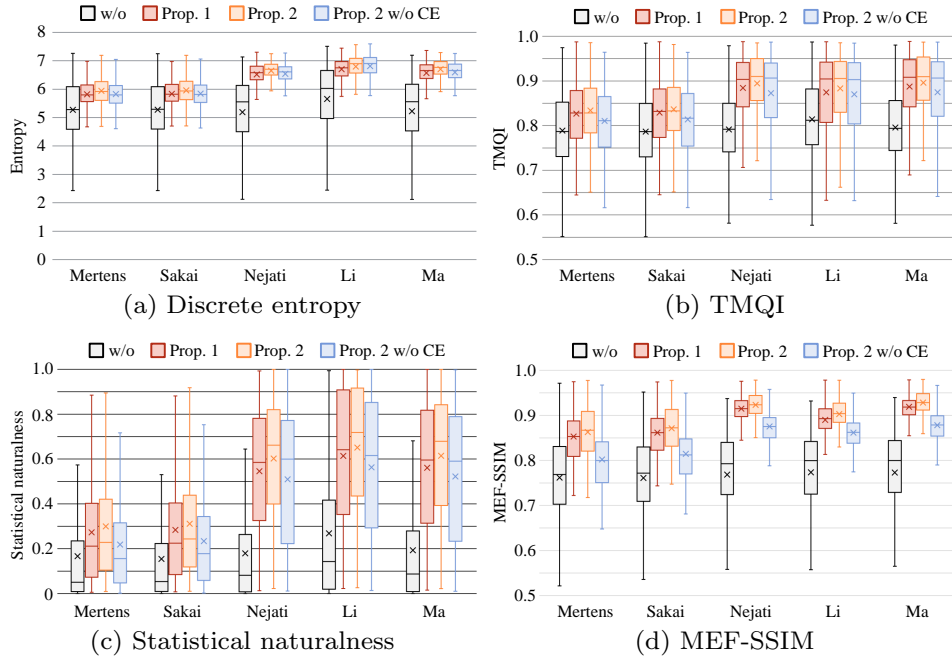


図 3.14 トーンマッピングにより生成された 500 組の多重露出画像に対する定量評価結果. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は, 中央値および平均値をそれぞれ表す.

三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれるスコアの最大値および最小値を示す. 箱の中にある横線は中央値, すなわち第二四分位数 Q_2 を表示しており, 十字は平均値を示している. Statistical naturalness と同様に, TMQI と MEF-SSIM のスコア $\in [0, 1]$ は, 大きいほど画像がより高い品質をもつことを示す. この図より, 提案法の 2 つの Approach どちらも, 補正せずに直接合成した場合に比べ, より高い品質を持つ画像を生成できることが, すべての合成法について確認できる. 特に, 同じ合成法 \mathcal{M} の使用の下では, Approach 2 がすべての尺度について最も高いスコアを示した. したがって, 提案する SSLA は, 異なる露出レベルを持つ十分な数の多重露出画像が与えられない場合, すなわち, 不明瞭な入力多重露出画像が与えられた場合において L-WDR 画像の生成を効果的に実行できる.

図 3.15 は, 十分な数の多重露出画像が入力として与えられた場合における提案法の性

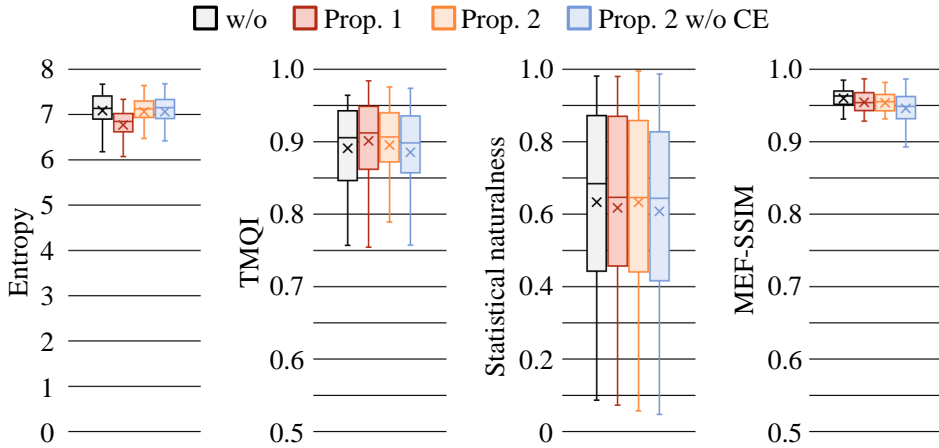


図 3.15 トーンマッピングにより生成された 50 組の条件の良い多重露出画像 (十分な数の露出を含む多重露出画像) に対する定量評価結果. 多重露出画像の合成には Ma らの手法を使用した. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は, 中央値および平均値をそれぞれ表す.

能を示している. 提案法が生成した画像に対するスコアは, Ma らの手法よりわずかに小さいものの, ほとんど同等の値を示した. この理由から, 条件の良い多重露出画像が入力として与えられた場合でも, 提案法は悪影響を与えないことが確認できる.

これらの理由から, 提案する SSLA は, 多重露出画像に基づく L-WDR 画像生成に有効であると確認できた. 提案法は, シーン全体を明瞭に表した高品質な L-WDR 画像を生成可能にする. Approach 1 および 2 を比較すると, Approach 2 はより明瞭な画像を生成できる一方, Approach 1 はより高速に実行できる.

3.3.4 計算量の比較

提案法の計算量を評価するため, 提案法の実行にかかる時間を計測した. この計測には, オンライン上で利用可能なデータベース [60] から収集した 8 組の多重露出画像を用いた. これら多重露出画像における画素数の最小値は 425, 430 であり, 最大値は 918, 400 である. また, 画素数の平均値は 533, 606.3 である.

計測は, 4.2GHz プロセッサと 64GB のメモリを持つコンピュータ上で行った (表 3.6

表 3.6 実行時間計測に用いたコンピュータスペック

Processor	Intel Core i7-7700K 4.20 GHz
Memory	64 GB
OS	Ubuntu 16.04 LTS
Software	MATLAB R2017b

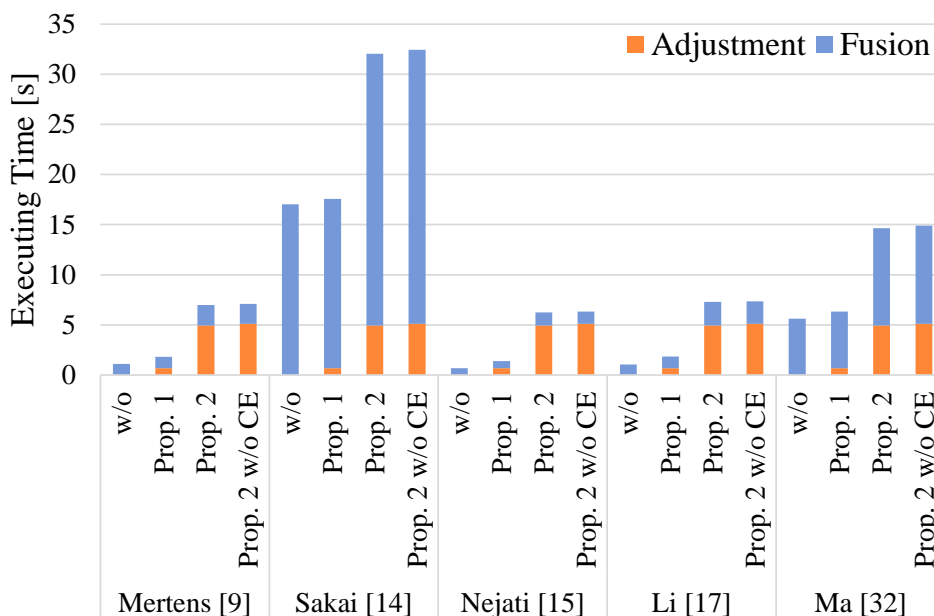


図 3.16 8 組の多重露出画像に対する平均実行時間. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す. Approach 1 は Approach 2 より高速であることが確認できる.

参照). 実行時間の計測のためには MATLAB 上で実装されている *timeit()* 関数を用い, 全手法は単一スレッドで実行された.

図 3.16 は, 8 組の多重露出画像についての平均実行時間を示している. この図より, Approach 1 を用いた補正によるオーバーヘッドは 1 秒ほどである一方, Approach 2 を用いた補正にはさらに多くの時間が必要であることが確認できる. さらに, Approach 1 との比較により, Approach 2 の使用の下では合成にかかる時間も増加していることがわかる. これは, Approach 2 で補正された多重露出画像の枚数は一般に, 入力多重露出画像の枚数よりも多くなることに起因する. したがって, Approach 2 は Approach 1 より高品質な画像生成を可能とするが, Approach 1 は Approach 2 より低い計算コストで実行できるといえる.

3.4 まとめ

本章では，多重露出画像に基づく L-WDR 画像生成のための，シーン領域分割に基づく多重露出画像補正法を提案した．本章ではまず，入力多重露出画像の輝度の補正により，最終的に生成される L-WDR 画像の品質が向上できることを示した．既存の L-WDR 画像生成法では輝度の補正を行わないのに対し，輝度の補正を行う提案法の利用は，高品質な L-WDR 画像の生成を可能にする．提案法では，入力多重露出画像を L-WDR 画像生成に適したものへ自動的に補正する目的で，シーン領域分割を実行する．このシーン領域分割のために，Approach 1 および 2 という 2 つの方法を提案した．Approach 1 は，閉形式で記述されるアルゴリズムにより，入力多重露出画像のシーンをいくつかの領域へ分割する．一方，Approach 2 では，輝度分布の GMM モデルを用いてシーンを分割する．

実験では，提案する SSLA を，5 つの L-WDR 画像生成法，すなわち Mertens らの方法，Sakai らの方法，Nejati らの方法，Li らの方法，および Ma らの方法を用いて評価した．この実験結果より，提案する SSLA による入力多重露出画像の補正は，Entropy および TMQI の観点からすべての生成法について有効であると示された．加えて，主観評価実験の結果，提案する SSLA はシーンの暗い領域と明るい領域を明瞭に表す画像の生成を可能とすると示された．とりわけ，Approach 2 は Approach 1 より明瞭な画像を生成できる．提案する SSLA は，被写体の位置のずれを含む多重露出画像に対しても有効であり，提案法と Ma らの手法を組み合わせることによってゴーストアーティファクトを除去できることが確認された．

第 4 章

シーン領域分割に基づく単一 LDR 画像からの L-WDR 画像推定

4.1 はじめに

本章では、高品質な L-WDR 画像の推定を目的として、シーン領域分割に基づく単一 LDR 画像からの多重露出画像推定法を提案する。前章で述べた多重露出画像補正法は、入力として単一 LDR 画像が与えられた場合を考慮していなかった。しかし、これまでに撮影された多くの LDR 画像コンテンツには対応する多重露出画像が存在しないことから、単一 LDR 画像を入力として想定した手法の開発は重要な課題である。

これまでに、単一 LDR 画像の強調により画像の品質向上を狙う研究は数多く行われている [11, 12, 16–21]。これら画像強調法の中で代表的なものとしては、輝度分布のエントロピーの最大化によって強調を行う Histogram equalization (HE) [11, 12, 16] と、画像を Reflectance と Illumination に分解した後 Illumination を処理することで強調を行う Retinex 理論に基づく方法 [17, 18, 47] がある。しかしながら、どちらの手法にも、画像中の明るい領域を過強調してしまう問題と暗い部分を十分に強調できない問題がある。そこで、本章では、多重露出画像に基づく L-WDR 画像生成法を応用し、単一 LDR 画像から明るい領域と暗い領域の双方を明瞭に表現した L-WDR 画像を推定する手法を提案する。

提案法では、まず、前章で提案した Approach 2 に基づくシーン領域分割を、入力の単一 LDR 画像に対して実行する。次に、領域分割の結果を用いて、単一 LDR 画像から擬似的に多重露出画像を生成する。これらの多重露出画像は、既存の多重露出画像に基づく L-WDR 画像生成法を用いて合成され、最終的な L-WDR 画像が推定される。

Discrete entropy と Statistical naturalness を尺度として用いた品質評価実験により、

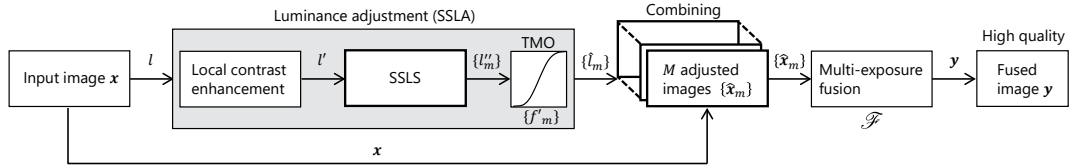


図 4.1 単一 LDR 画像に基づく多重露出画像推定法

従来の単一 LDR 画像の強調に基づく手法と比較して，提案法が高い品質の画像を生成できることが示される．加えて，提案法は，シーンの暗い画像と明るい領域双方を明瞭に表現した L-WDR 画像を生成できることが確認される．

4.2 単一 LDR 画像からの多重露出画像推定法

図 4.1 は，提案する多重露出画像推定法のご概念図である．前章で提案した多重露出画像補正法との違いは，入力として与えられる LDR 画像が 1 枚のみであることである．この理由から，局所コントラスト強調は，単一 LDR 画像から計算される輝度のみに対して適用される．また，SSLS における Approach 2 中の GMM の構築も同様に，単一 LDR 画像の輝度分布のみを用いて行う．シーンの領域分割結果に基づき， M 個の領域 $\mathbb{P}_1, \dots, \mathbb{P}_M \subset \mathbb{P}$ それぞれをよく表す画像を生成した後は，前章と同様の処理を行い L-WDR 画像が生成される．以下では，局所コントラスト強調および SSLS における前章で提案した手法との差分を説明する．

4.2.1 局所コントラスト強調における差分

提案法における局所コントラスト強調は，式 (3.1) を次式の通り置き換えて実行される．

$$l'(\mathbf{p}) = \frac{l(\mathbf{p})^2}{\bar{l}(\mathbf{p})} \quad (4.1)$$

ここで， l は，入力の単一 LDR 画像 x の輝度である．

4.2.2 SSLS における差分

提案法では, SSLS のアルゴリズムとして Approach 2 のみを用いる. ここで異なる式は式 (3.9), (3.13), (3.14), および (3.16) であり, それぞれ以下の通り置き換えられる.

$$p(l'(\mathbf{p})) = \sum_{k=1}^K \pi_k \mathcal{N}(l'(\mathbf{p}) | \mu_k, \sigma_k^2) \quad (4.2)$$

$$\begin{aligned} \gamma(z_k | l'(\mathbf{p})) &= p(z_k = 1 | l'(\mathbf{p})) \\ &= \frac{\pi_k \mathcal{N}(l'(\mathbf{p}) | \mu_k, \sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(l'(\mathbf{p}) | \mu_j, \sigma_j)} \end{aligned} \quad (4.3)$$

$$m = \arg \max_k \gamma(z_k | l'(\mathbf{p})) \quad (4.4)$$

$$\alpha_m = \frac{0.18}{G(l' | \mathbb{P}_m)} \quad (4.5)$$

ここで, $\mathcal{N}(l'(\mathbf{p}) | \mu_k, \sigma_k^2)$ は, 平均 μ_k および分散 σ_k^2 を持つ 1 次元ガウス分布である.

以上の局所コントラスト強調および SSLS によって得られた多重露出画像は, 第 3.2.2 項で説明したトーンマッピング処理の入力として用いられる.

4.3 実験および考察

本節では, 推定される L-WDR 画像 \mathbf{y} の品質の観点から, 提案法の有効性を評価するために実施した実験について説明する.

4.3.1 実験条件

本実験では, Canon EOS 5D Mark II を用いて撮影された 22 枚の LDR 画像と, データベース [60] から収集した 16 枚の LDR 画像を入力画像 \mathbf{x} として使用した. 不明瞭な入力画像を想定するため, これら画像には, 0 か負の露出値を用いて撮影されたもののみを利用した. 提案法を評価は以下の手順で行った.

1. 提案法により, \mathbf{x} から \mathbf{y} を生成する.
2. \mathbf{y} を Statistical naturalness により評価する.
3. \mathbf{y} を Discrete entropy により評価する.

表 4.1 提案法により分割された領域 $\{\mathbb{P}_m\}$ の数 M の例 ($K = 10$)

Image name	M
Chinese garden	7 (Maximum)
Estate rsa	4
Trashbox	3 (Minimum)
Window	5

本実験では、比較手法として 8 つの手法、すなわち Histogram equalization (HE), Contrast limited adaptive histogram equalization (CLAHE) [11], Adaptive gamma correction with weighting distribution (AGCWD) [12], Contrast-accumulated histogram equalization (CACHE) [16], Low light image enhancement based on two-step noise suppression (LLIE) [19], Low-light image enhancement via illumination map estimation (LIME) [17], Simultaneous reflectance and illumination estimation (SRIE) [18], Bio-inspired multi-exposure fusion framework for low-light image enhancement (BIMEF) [67], を用いた。また、提案法においては、Nejati らの手法を合成法 \mathcal{F} として利用した。

4.3.2 シーン領域分割の結果

表 4.1 は、4 枚の入力 LDR 画像 (図 4.2 から 4.6 を参照) について、提案法におけるシーン領域分割によって得られた領域の数を示す。ここで、実験で用いた 38 枚の画像中、分割された領域の数は最大 7、最小 3 であった。この表から、前章で得られた結果と同様に、大きな K が与えられた場合でさえ、提案法で構築される GMM は過適合していないことがわかる。

図 4.2 と 4.3 には、提案法により分割された領域数が最大の画像および最小の画像を示している。これらの結果から、入力 LDR 画像の輝度が狭い範囲に分布しているときには、分割される領域の少なくなることがわかる (図 4.3 参照)。加えて、図 4.2(a) と 4.2(b) (および図 4.3(a) と 4.3(b)) の比較から、入力として単一画像が与えられた場合にも、提案法におけるシーン領域分割によって、それぞれがある特定の明るさを持つ領域へ画像を分割できることが確認できる。結果として、図 4.2 (d) から (j) (および図 4.3 (d) から (f)) に示されるように、提案法によって高品質な多重露出画像を推定できる。さらに、これら多重露出画像の合成によって、高品質な L-WDR 画像が得られる (図 4.2(c) および 4.3(c))

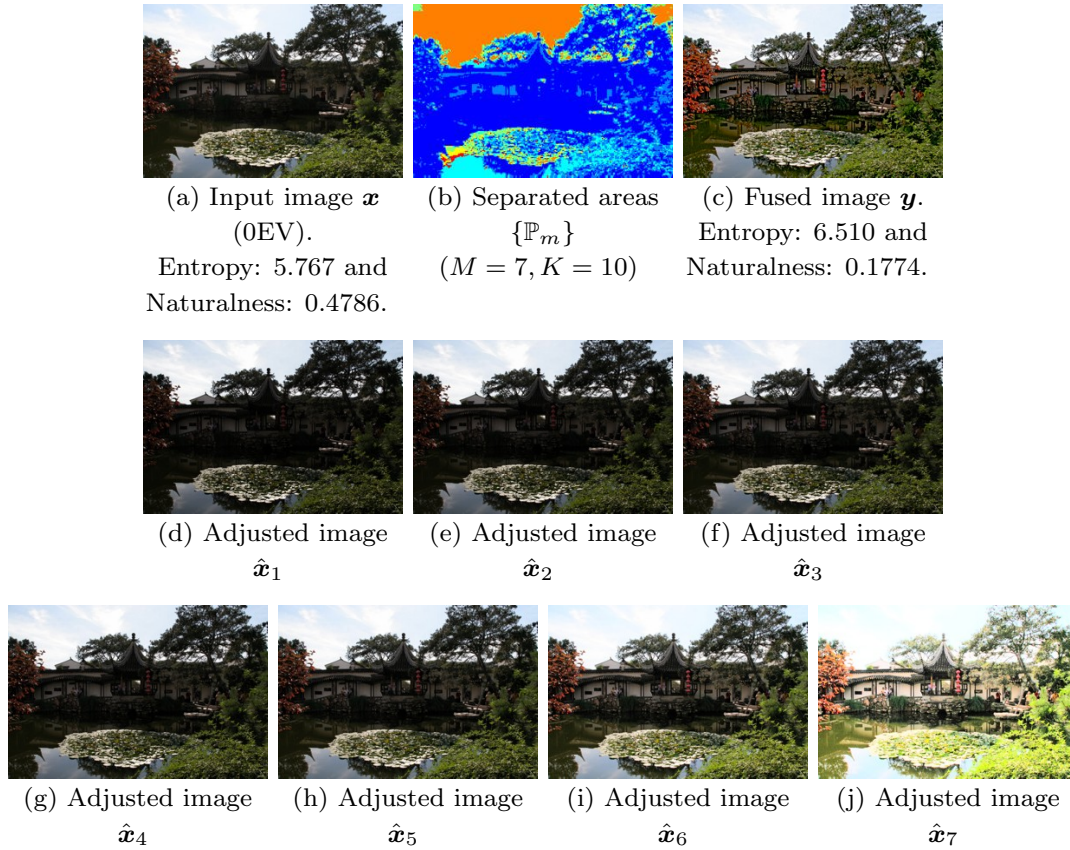


図 4.2 提案法の実行結果 (“Chinese garden”). (a): 入力 LDR 画像, (b): 提案するシーンの領域分割の結果, (c): 最終的に得られた L-WDR 画像, (d)-(j): 提案法によって推定された多重露出画像. (b) において, 各色は分割された各領域に対応する.

参照).

図 4.4 は, 固定された M および α_m の使用の下で推定された L-WDR 画像 \boldsymbol{y} を示す. ここで, (a): 入力画像, (b)-(d): M と α_m 双方を固定した場合の結果, (e): 提案法により得られた結果, (f)-(h): 固定の M と式 (4.5) により推定した α_m を用いた場合の結果である. 固定の M で画像領域を分割するために, (f)-(h) では k 平均法を利用した. 図 4.4(b)-(d) より, M と α_m の両方が固定されている場合には, 大きな M を用いることが明瞭な画像を生成するために有効であることがわかる. 一方, 小さい M を用いた場合に得られる画像は入力画像と比較してほぼ変化がなく, ほとんど L-WDR 画像推定法の効果が得られていないことがわかる. これは, α_m が固定である場合には, 入力画像の輝度が推定される多重露出画像の輝度に直接影響するためである.

M のみが固定されている場合には, 図 4.4(b) と 4.4(f) に示すように, M と α_m が共

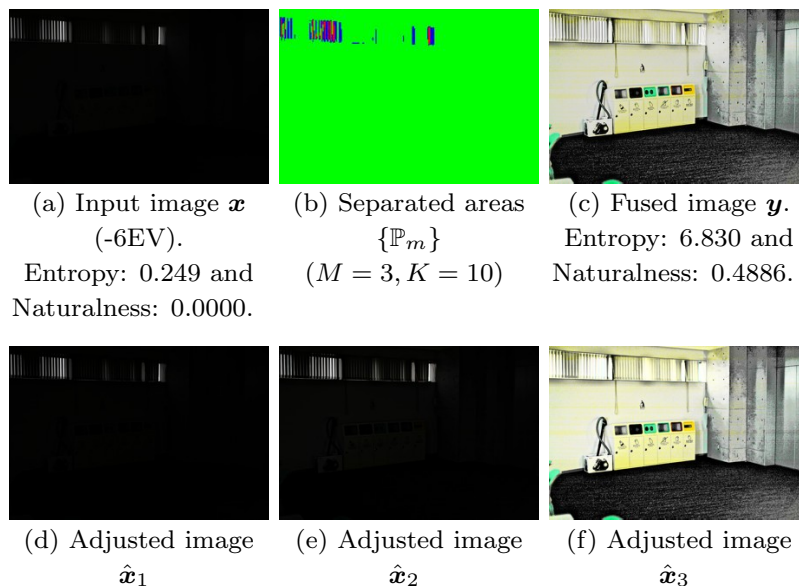


図 4.3 提案法の実行結果 (“Trashbox”). (a): 入力画像, (b): 提案するシーンの領域分割の結果, (c): 最終的に得られた L-WDR 画像, (d)-(f): 提案法により推定された多重露出画像. (b)において, 各色は分割された各領域に対応する.

に固定されている場合より細部を明瞭に表現した画像を生成できる. したがって, α_m を各領域に対して適応的に決定することは効果的である. $M = 3$ の場合と $M = 5$ の場合を比較すると, $M = 5$ を用いて生成された画像は, $M = 3$ を用いて生成された画像より明瞭であることが, 図 4.4(f) および 4.4(g) から確認できる. 加えて, $M = 5$ を用いて生成された画像は, $M = 7$ を用いて生成された画像とほとんど等しい. (図 4.4(g) および 4.4(h) 参照). したがって, α_m を適応的に決定する場合でも, 適切な M を選ぶことが高品質な L-WDR 画像の生成には必要である.

提案法は, M および α_m の両方を自動的に決定する. 提案法を用いて生成された画像は, 図 4.4(e) に示されるように, 十分明瞭である. したがって, 提案する単一 LDR 画像からの多重露出画像の推定は, L-WDR 画像の生成に有効である.

4.3.3 L-WDR 画像推定結果

入力画像 “Window” から推定された L-WDR 画像を図 4.5 に示す. この図から, 提案法は, シーンの暗い領域を強く強調していることが確認できる. 従来の画像強調に基づく手法 CLAHE, AGCWD, CACHE, LLIE, および BIMEF も強調に一定の効果がある. しかしながら, その強調効果は, 暗部を可視化するためには十分でない. 加えて, 提

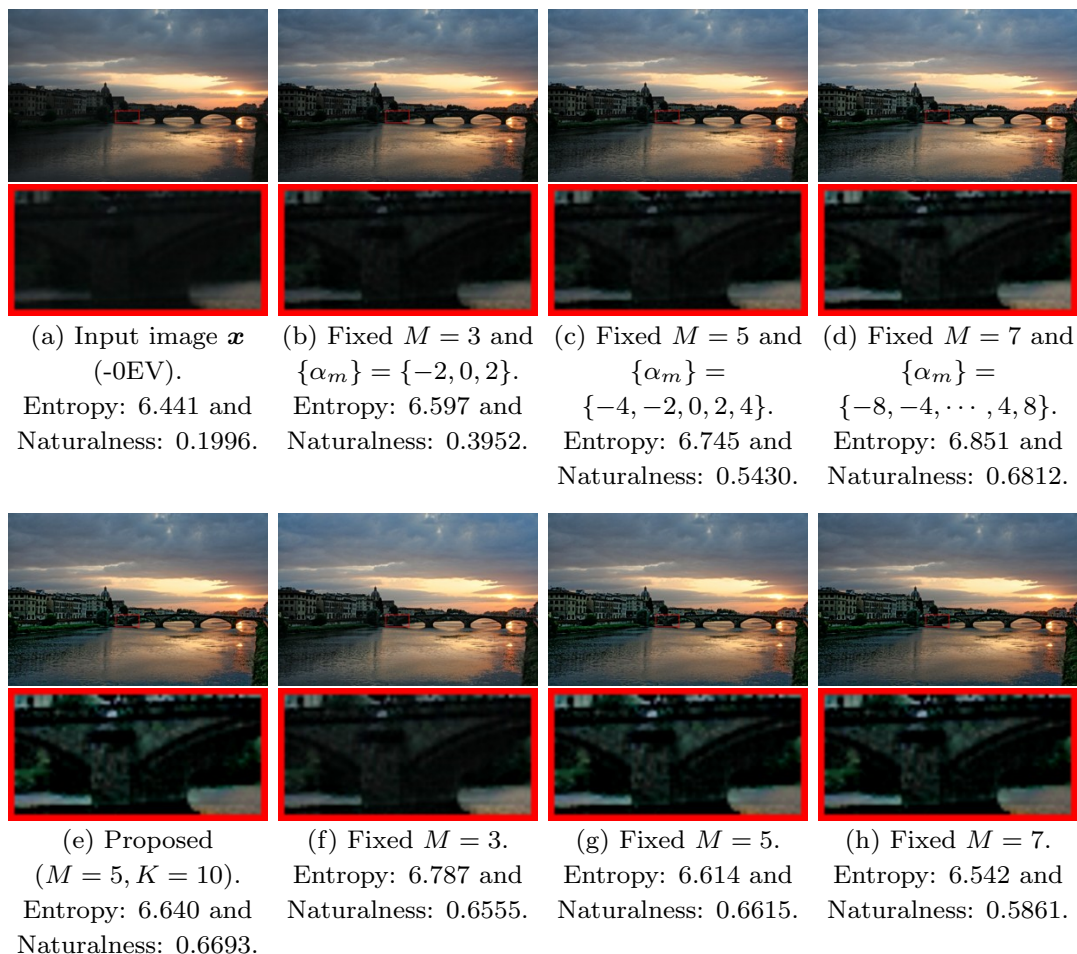


図 4.4 固定の M および α_m を使用した場合の結果画像 (“Arno”). (a): 入力画像, (b)-(d): M と α_m 双方を固定した場合の結果, (e): 提案法による結果, (f)-(h): M のみを固定した場合の結果. 枠で囲われた領域の拡大図は, 各画像の下に示される.

案法は, 明るい領域の詳細を失うことなく, すなわち過強調なしに画像品質を向上できている. 一方, HE と LIME は過強調を起こしている. 図 4.5 と同様の傾向は, 図 4.6 でも確認できる. これらの結果より, 提案法は, 明るい領域の詳細情報を保ったまま暗い領域を明瞭にすることができるといえる.

図 4.7 および 4.8 は, 38 枚の入力 LDR 画像から生成された画像の Discrete entropy と Statistical naturalness によるスコアを箱ひげ図として示したものである. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値, すなわち第二四分位数 Q_2 を表示している. 各スコア (Discrete entropy $\in [0, 8]$,



図 4.5 提案法と従来の画像強調法との比較 (“Window”). 枠で囲われた領域の拡大図は、各画像の下に示される。提案法は、過強調および強調不足なしに明瞭な画像を生成できる。

Statistical naturalness $\in [0, 1]$ は、値が大きいほど高い品質であることを示している。

図 4.7 から、9 つの手法のうち最も高い中央値を示した手法は LIME だった一方、提案法で生成された画像は、入力画像のスコアに関わらず、極めて狭い範囲に分布する高いスコアを持つことがわかる。対して、従来法によるスコアは提案法と比較して広い範囲に分



図 4.6 提案法と従来の画像強調法との比較 (“Estate rsa”). 枠で囲われた領域の拡大図は、各画像の下に示される。提案法は、過強調および強調不足なしに明瞭な画像を生成できる。

布していることがわかる。したがって、提案法は、従来法と比較して Discrete entropy の観点から高い品質の画像を生成できるといえる。提案法は、入力画像が高い Entropy スコアを持つ場合、すなわち強調の必要が無い場合に、Entropy スコアを減少させてしまうことがある。これは、提案法が多重露出画像を推定する際に、各領域の平均輝度を中間のグレーへ合わせることに起因する。

図 4.8 は、提案法と HE により生成された画像が、他の手法により生成された多くの画像より高い Statistical naturalness スコアを持つことを表す。この結果は、図 4.5 および 4.6 にあるように、提案法と HE が強く画像を強調することを示している。第三四分位数

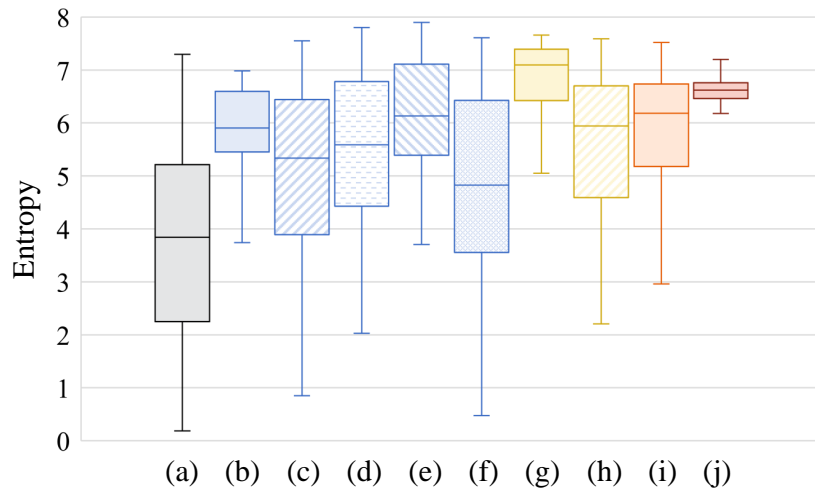


図 4.7 Discrete entropy による評価結果. (a) Input image, (b) HE, (c) CLAHE, (d) AGCWD, (e) CACHE, (f) LLIE, (g) LIME, (h) SRIE, (i) BIMEF, および (j) Proposed. 各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表す.

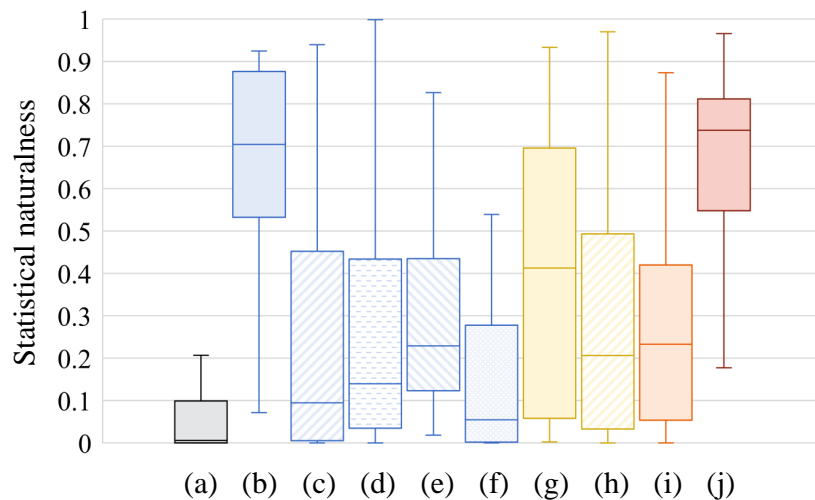


図 4.8 Statistical naturalness による評価結果. (a) Input image, (b) HE, (c) CLAHE, (d) AGCWD, (e) CACHE, (f) LLIE, (g) LIME, (h) SRIE, (i) BIMEF, および (j) Proposed. 各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表す.

を比較すると、HE は提案法を含む 9 つの手法の中で、最も高いスコアを持つ。一方、9

つの手法のうち、提案法は最も高い中央値および最も高い最大値を持つ。この理由から、提案法と HE は、Statistical naturalness の観点から同等の性能を持つといえる。ここで、HE は明るい領域の過強調を発生させるが、提案法は過強調を起こさないことから (図 4.5 および 4.6 参照)、提案法は HE より高品質な画像を生成できることが確認できる。

以上より、提案法が単一 LDR 画像から多重露出画像の推定を可能とし、それらの合成により高品質な L-WDR 画像が生成されることが確認できた。とりわけ、提案法によって生成された L-WDR 画像は、シーンの明るい領域と暗い領域の双方を明瞭に表現できることが確かめられた。

4.4 まとめ

本章では、高品質な L-WDR 画像推定を目的とし、シーン領域分割に基づく単一 LDR 画像からの多重露出画像推定法を提案した。多重露出画像の推定のため、前章で提案したシーンの領域分割に基づく多重露出画像補正法を単一 LDR 画像に適用できるよう拡張した。提案法により推定された多重露出画像の合成により、単一 LDR 画像から、シーンの明るい領域と暗い領域両方を明瞭に表す高品質な L-WDR 画像生成が実現できる。実験では、Discrete entropy および Statistical naturalness の観点から、提案法が従来の単一画像強調法を上回る性能を持つことが明らかとなった。さらに、主観評価の結果は、提案法が、単一 LDR 画像からシーンの明るい領域と暗い領域両方を明瞭に表す L-WDR 画像生成に効果的であることを示した。

第 5 章

Reinhard のグローバルオペレータに基づく高速逆トーンマッピングオペレータ

5.1 はじめに

本章では，単一 LDR 画像からの U-WDR 画像推定を目的として，Reinhard のグローバルオペレータに基づく高速逆トーンマッピングオペレータを提案する．逆トーンマッピング法とは，U-WDR 画像から L-WDR 画像を生成するトーンマッピングとは逆に，単一 LDR 画像から U-WDR 画像を推定する処理の総称である．

単一 LDR 画像のみを用いて U-WDR 画像を推定する場合には，多重露出画像における被写体の位置のずれに起因するゴーストアーティファクトが原理的に発生しない．そのため，これまでに様々な逆トーンマッピング法が開発されている [9, 10, 22–24]．Huo らは，S 字カーブを逆トーンマッピングオペレータとして用いてダイナミックレンジを拡張し，その後，暗い領域と明るい領域の局所コントラストを Dodging and burning アルゴリズムにより強調する方法を提案した [23]．Wang らは，単一 LDR 画像から擬似的に多重露出画像を生成し，それらの合成により U-WDR 画像を推定する逆トーンマッピング法を提案した [24]．

これらの逆トーンマッピングオペレータを用いる方法は，近年盛んに研究されている深層学習に基づく方法と比較して，低い性能を持つものの低い計算コストで実行できるというメリットがある．この理由から，逆トーンマッピングオペレータを用いる方法は，U-WDR 動画の生成への応用および U-WDR 画像の情報源符号化への応用 [30] が期待

されている。一方、既存の逆トーンマッピングオペレータは、その実行時間の多くを内部パラメータの決定のために割いている。したがって、パラメータの決定にかかる時間を削減することによって、さらなる高速化が実現できる。

本章で提案する高速逆トーンマッピングオペレータは、Reinhard のグローバルオペレータ [56, 68] の逆関数を逆トーンマッピングオペレータとして利用する。Reinhard のグローバルオペレータは、トーンマッピングにより生成される画像の明るさを決定する a と、U-WDR 画像の輝度 l_E の幾何平均 $G(l_E|\mathbb{P})$ という 2 つのパラメータを必要とする。したがって、この逆関数を計算する際にはそれらパラメータが必要とされる [25]。本章では、いくつかの条件の下で、 $G(l_E|\mathbb{P})$ が a から推定できることを示す。その逆に、 a は $G(l_E|\mathbb{P})$ から推定可能である。これらパラメータの推定は閉形式で記述されるアルゴリズムによって行われるため、高速に実行可能である。

従来の逆トーンマッピングオペレータと出力画像品質および実行時間を比較した実験により、提案する逆トーンマッピングオペレータは、従来法と同等の性能を維持しながら高速に実行可能であることが示される。

5.2 Reinhard らのトーンマッピングフレームワーク

本節では、準備として、Reinhard らにより提案されたトーンマッピングフレームワークである Photographic tone reproduction について説明する [56]。ここで、トーンマッピングは、U-WDR 画像 E から L-WDR 画像 x を生成する処理である。

Photographic tone reproduction を用いた L-WDR 画像 x の生成手順は以下の通りである。

1. U-WDR 画像 E の輝度 l_E を計算する。
2. l_E の幾何平均 $G(l_E|\mathbb{P})$ を式 (2.1) により計算する。
3. 2 つのパラメータ a と $G(l_E|\mathbb{P})$ を用いて、次式により l_X を計算する。

$$l_X(\mathbf{p}) = \frac{a}{G(l_E|\mathbb{P})} l_E(\mathbf{p}) \quad (5.1)$$

ここで、 $a \in [0, 1]$ は、“key value” と呼ばれる出力画像の明るさを定めるパラメータである [56]。

4. L-WDR 画像の輝度 l_x を次式として与える。

$$l_x(\mathbf{p}) = f'(l_X(\mathbf{p})) \quad (5.2)$$

ここで、関数 f' はトーンマッピングオペレータである。Reinhard のグローバルオペレータを関数 f' として用いる場合、

$$f'(t) = \frac{t}{1+t} \left(1 + \frac{t}{L^2} \right) \text{ for } t \in \mathbb{R}_+ \quad (5.3)$$

となる [56]. 式 (5.3) において、 $f'(t) = 1$ となる t を定めるパラメータ L を $L \rightarrow \infty$ とすることにより、より簡単なグローバルオペレータ

$$f'(t) = \frac{t}{1+t} \quad (5.4)$$

が得られる。

5. 次式により L-WDR 画像 \mathbf{x} を得る。

$$\mathbf{x}(\mathbf{p}) = \frac{l_{\mathbf{x}}(\mathbf{p})}{l_{\mathbf{E}}(\mathbf{p})} \mathbf{E}(\mathbf{p}) \quad (5.5)$$

5.2.1 シナリオ

提案する逆トーンマッピング法を含む、逆トーンマッピングオペレータを用いる逆トーンマッピング法には、その応用の違いから 2 つの異なるシナリオがある。

Scenario 1: 一般の LDR 画像の逆トーンマッピング

第一のシナリオは、従来の LDR カメラによって撮影された一般の LDR 画像から、逆トーンマッピングにより U-WDR 画像を推定することである。この場合には、現実シーンと LDR 画像間の関係についての情報 (例えば関数 f など) が得られない。この理由から、従来法は、LDR 画像がある関数 f によって生成されていると仮定して、それに対応する逆トーンマッピングオペレータを用いて逆トーンマッピングを行う。ここで、高精度の逆トーンマッピングの実現のため、従来法では、逆トーンマッピングオペレータにおけるパラメータを計算コストの高い処理によって推定する。

一方、提案法は、閉形式によるパラメータの推定を実現し、従来法と同等の精度の逆トーンマッピングを高速に実行可能である。提案法では、式 (5.4) に示す Reinhard のグローバルオペレータを用いた Photographic tone reproduction により LDR 画像が生成されているものとして、逆トーンマッピングを実行する。式 (5.1) と (5.4) から、Reinhard のグローバルオペレータの逆関数は次式により与えられる。

$$l_{\mathbf{E}}(\mathbf{p}) = \frac{G(l_{\mathbf{E}}|\mathbb{P})}{a} \cdot l_{\mathbf{X}}(\mathbf{p}) = \frac{G(l_{\mathbf{E}}|\mathbb{P})}{a} \cdot \frac{l_{\mathbf{x}}(\mathbf{p})}{1 - l_{\mathbf{x}}(\mathbf{p})} \quad (5.6)$$

この式の計算には 2 つのパラメータ (a および $G(l_E|\mathbb{P})$) が必要とされるが、提案法は、その両方を閉形式により高速に推定する。

Scenario 2: U-WDR 画像の再構成

第二のシナリオは、U-WDR 画像をトーンマッピングすることで生成された L-WDR 画像から、元の U-WDR 画像を再構成することである。このシナリオで逆トーンマッピングが利用される状況は、主に二層符号化法を用いる U-WDR 画像の情報源符号化がある [30]。この場合には、U-WDR 画像と L-WDR 画像間の関係についての情報 (トーンマッピング法など) を既知として扱うことができ、U-WDR 画像の高精度な再構成を高速に実行することが、計算コストおよび符号化効率の面から優れた情報源符号化法を実現するために重要である。

提案法では、Scenario 1 と同様に、式 (5.4) に示す Reinhard のグローバルオペレータを用いた Photographic tone reproduction により L-WDR 画像が生成されることを想定する。Scenario 2 においては、式 (5.6) の計算に必要な 2 つのパラメータ (a および $G(l_E|\mathbb{P})$) を記録しておくことが可能である。ここで、各 U-WDR 画像は異なる $G(l_E|\mathbb{P})$ を持つことから、式 (5.6) を正確なパラメータで実行するためには、複数の U-WDR 画像や U-WDR 動画像を L-WDR へ変換する際に、すべての $G(l_E|\mathbb{P})$ を記録しなければならない。一方、 a は、一般に複数の画像に対して共通の値が用いられる。

本章で提案する逆トーンマッピング法は、パラメータ a および $G(l_E|\mathbb{P})$ を共に記録することを必要とせず、1 つのパラメータのみから他方のパラメータを推定可能とする。この推定処理は閉形式の計算によって行われるため、提案法は、原 U-WDR 画像の高速かつ高精度な再構成を実現する。

5.3 Reinhard のグローバルオペレータに基づく逆トーンマッピング法

提案法は、Reinhard のグローバルオペレータを用いる Photographic tone reproduction の逆変換により、逆トーンマッピングを実行する。

5.3.1 Reinhard のグローバルオペレータに基づく逆トーンマッピングオペレータ

提案する逆トーンマッピングオペレータは,

$$\hat{l}_{\mathbf{E}}(\mathbf{p}) = \frac{\mathcal{G}}{\mathcal{A}} \cdot \hat{l}_{\mathbf{X}}(\mathbf{p}) \quad (5.7)$$

$$\hat{l}_{\mathbf{X}}(\mathbf{p}) = f'^{-1}(l_{\mathbf{x}}(\mathbf{p})) = \frac{l_{\mathbf{x}}(\mathbf{p})}{1 - l_{\mathbf{x}}(\mathbf{p})} \quad (5.8)$$

として与えられる. 式 (5.7) を式 (5.6) と比べると, 式 (5.6) におけるパラメータ a および $G(l_{\mathbf{E}}|\mathbb{P})$ が, \mathcal{A} および \mathcal{G} でそれぞれ置き換えられていることがわかる. ここで, パラメータ \mathcal{A} および \mathcal{G} は, 以下のようにパラメータ a あるいは $G(l_{\mathbf{E}}|\mathbb{P})$ から推定できる.

(i) a が与えられた場合には, \mathcal{A} および \mathcal{G} を以下の通り計算する.

$$\begin{cases} \mathcal{A} = a \\ \mathcal{G} = \exp\left(\frac{|\mathbb{P}|}{|\mathbb{P}_B|} \log G(\hat{l}_{\mathbf{X}}|\mathbb{P}) - \frac{|\overline{\mathbb{P}_B}|}{|\mathbb{P}_B|} \log a\right) \end{cases} \quad (5.9)$$

ここで, $\mathbb{P}_B = \{\mathbf{p} \in \mathbb{P} | l_{\mathbf{E}}(\mathbf{p}) = 0\}$ であり, $\overline{\mathbb{P}_B} = \{\mathbf{p} \in \mathbb{P} | l_{\mathbf{E}}(\mathbf{p}) \neq 0\}$ である. したがって, 計算において有限語長の影響がない場合には, 1 つのパラメータ a のみを用いて, Reinhard のグローバルオペレータにより生成された LDR 画像から HDR 画像を完全に再構成可能である.

式 (5.9) において $|\mathbb{P}_B| \neq 0$ を仮定しているため, \mathbf{E} が $|\mathbb{P}_B| = 0$ を満たす場合に式 (5.9) を計算するには, $l_{\mathbf{E}}(\mathbf{p})$ の最小値を 0 で置き換える必要がある.

(ii) $G(l_{\mathbf{E}}|\mathbb{P})$ が与えられた場合には, \mathcal{A} および \mathcal{G} を以下の通り計算する.

$$\begin{cases} \mathcal{A} = \exp\left(\frac{|\mathbb{P}|}{|\mathbb{P}_B|} \log G(\hat{l}_{\mathbf{X}}|\mathbb{P}) - \frac{|\mathbb{P}_B|}{|\mathbb{P}_B|} \log G(l_{\mathbf{E}}|\mathbb{P})\right) \\ \mathcal{G} = G(l_{\mathbf{E}}|\mathbb{P}) \end{cases} \quad (5.10)$$

式 (5.10) は, $|\mathbb{P}_B| = 0$ を満たす際でも計算可能である.

5.3.2 実行手順

提案法を用いて, U-WDR 画像 $\hat{\mathbf{E}}$ を単一画像 \mathbf{x} から推定する手順を以下に示す (図 5.1 参照).

1. L-WDR 画像 \mathbf{x} の輝度 $l_{\mathbf{x}}(\mathbf{p})$ を計算する.
2. $\mathbb{P}_B = \{\mathbf{p} \in \mathbb{P} | l_{\mathbf{x}}(\mathbf{p}) = 0\}$ とおく.

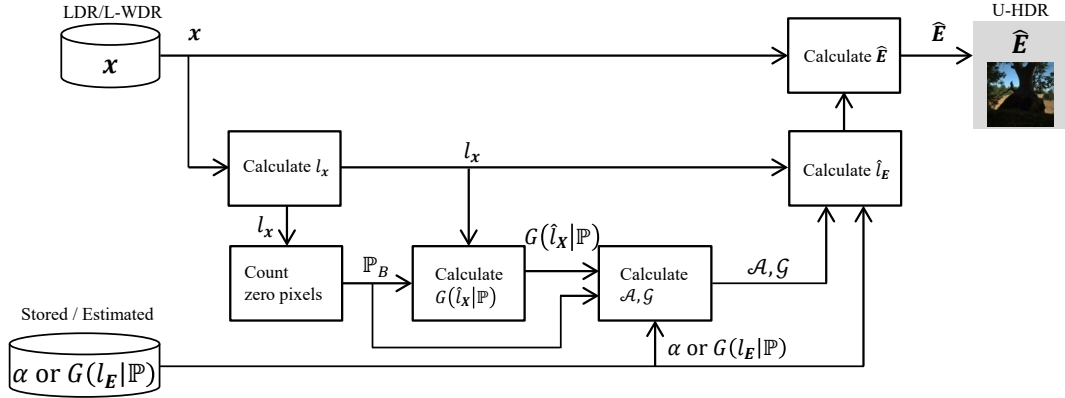


図 5.1 提案する逆トーンマッピング法

3. $G(\hat{l}_X|P)$ を計算する.
4. パラメータ A, G を計算する.
5. U-WDR 画像の輝度 $\hat{l}_E(p)$ を, A, G を用いて式 (5.7) および (5.8) により計算する
6. U-WDR 画像 \hat{E} を次式により得る.

$$\hat{E}(p) = \frac{\hat{l}_E(p)}{l_x(p)} \cdot x(p) \quad (5.11)$$

手順 4 におけるパラメータ A, G の計算は, 以下で説明するように 2 つのシナリオそれぞれで異なる.

Scenario 1

Scenario 1 では, Photographic tone reproduction が入力画像 x の生成に使用されていないため, パラメータ a および $G(l_E|P)$ はそもそも存在しない. そのため, Scenario 1 においては a を

$$a = G(\hat{l}_X|P) \quad (5.12)$$

として, 式 (5.9) に従いパラメータ A, G を計算する. 式 (5.12) の妥当性は第 5.3.3 項で示す. $G(\hat{l}_X|P)$ は x から計算できるため, 提案法は容易に実行できる.

Scenario 2

Scenario 2 では, Photographic tone reproduction を用いて入力画像 x が生成されているため, パラメータ a あるいは $G(l_E|P)$ を用いて, 次の通りパラメータ A, G を計算できる.

1. a が与えられた場合には，式 (5.9) に従い \mathcal{A}, \mathcal{G} を求める．ここで，第 5.3.1 項に示したとおり $|\mathbb{P}_B| \neq 0$ を仮定する．
2. $G(l_{\mathbf{E}}|\mathbb{P})$ が与えられた場合には，式 (5.10) に従い \mathcal{A}, \mathcal{G} を求める．

5.3.3 提案する逆トーンマッピング法の導出

本項では，式 (5.9) および (5.10) で示した，パラメータ \mathcal{A}, \mathcal{G} を計算する公式を導出する．

まず， $\hat{l}_{\mathbf{X}}$ の幾何平均 $G(\hat{l}_{\mathbf{X}}|\mathbb{P})$ を考える．ここで，式 (2.1) における ϵ は， $p \in \overline{\mathbb{P}_B}$ について， $\epsilon < \hat{l}_{\mathbf{X}}(\mathbf{p})$ および $\epsilon < l_{\mathbf{E}}(\mathbf{p})$ を満たすものとする．この幾何平均 $G(\hat{l}_{\mathbf{X}}|\mathbb{P})$ は，

$$\begin{aligned}
 G(\hat{l}_{\mathbf{X}}|\mathbb{P}) &= \exp \left(\frac{1}{|\mathbb{P}|} \left(\sum_{p \notin \mathbb{P}_B} \log \frac{a}{G(l_{\mathbf{E}}|\mathbb{P})} l_{\mathbf{E}}(\mathbf{p}) + \sum_{p \in \mathbb{P}_B} \log \epsilon \right) \right) \\
 &= \exp \left(\frac{1}{|\mathbb{P}|} \sum_{p \notin \mathbb{P}_B} \log a - \frac{1}{|\mathbb{P}|} \sum_{p \notin \mathbb{P}_B} \log G(l_{\mathbf{E}}|\mathbb{P}) \right) \\
 &\quad \cdot \exp \left(\frac{1}{|\mathbb{P}|} \sum_{p \notin \mathbb{P}_B} \log l_{\mathbf{E}}(\mathbf{p}) + \frac{1}{|\mathbb{P}|} \sum_{p \in \mathbb{P}_B} \log \epsilon \right) \\
 &= \exp \left(\frac{|\overline{\mathbb{P}_B}|}{|\mathbb{P}|} \log a + \frac{|\mathbb{P}_B|}{|\mathbb{P}|} \log G(l_{\mathbf{E}}|\mathbb{P}) \right) \tag{5.13}
 \end{aligned}$$

として計算できる．式 (5.13) から，式 (5.9) および (5.10) にそれぞれ相当する関係式が，次式として導かれる．

$$a = \exp \left(\frac{|\mathbb{P}|}{|\overline{\mathbb{P}_B}|} \log G(\hat{l}_{\mathbf{X}}|\mathbb{P}) - \frac{|\mathbb{P}_B|}{|\overline{\mathbb{P}_B}|} \log G(l_{\mathbf{E}}|\mathbb{P}) \right) \tag{5.14}$$

$$G(l_{\mathbf{E}}|\mathbb{P}) = \exp \left(\frac{|\mathbb{P}|}{|\mathbb{P}_B|} \log G(\hat{l}_{\mathbf{X}}|\mathbb{P}) - \frac{|\overline{\mathbb{P}_B}|}{|\mathbb{P}_B|} \log a \right) \tag{5.15}$$

a および $G(l_{\mathbf{E}}|\mathbb{P})$ を \mathcal{A} および \mathcal{G} でそれぞれ置き換えることによって式 (5.9) と (5.10) が導出される． $G(\hat{l}_{\mathbf{X}}|\mathbb{P})$ はトーンマッピングにより生成された L-WDR 画像から計算できるため， a か $G(l_{\mathbf{E}}|\mathbb{P})$ の一方が既知の場合にはパラメータ a が式 (5.14) により計算でき， $G(l_{\mathbf{E}}|\mathbb{P})$ が式 (5.15) により計算できる．

加えて， $|\mathbb{P}| \simeq |\overline{\mathbb{P}_B}|$ の場合，式 (5.14) より， a および $G(\hat{l}_{\mathbf{X}}|\mathbb{P})$ 間の関係式が

$$a \simeq G(\hat{l}_{\mathbf{X}}|\mathbb{P}) \tag{5.16}$$

として導かれる。この結論は、パラメータ a と $G(l_E|\mathbb{P})$ の両方が未知の場合、すなわち Scenario 1 における提案法の実行を可能とする。

5.4 実験および考察

本節では、推定される U-WDR 画像 E の品質、および実行時間の観点から、提案する逆トーンマッピング法を評価した実験について説明する。

5.4.1 提案法の評価

提案法により生成される U-WDR 画像 E の品質評価には、客観評価尺度が必要とされる。U-WDR 画像は、L-WDR 画像あるいは LDR 画像と異なり放射照度（もしくはその推定値）を直接画素値として記録していることから、PSNR や SSIM、あるいは TMQI による評価ができない。この理由から、U-WDR 画像の品質評価に関する研究が既に行われている [69]。本実験では、文献 [69] によって U-WDR 画像の品質評価に適していることが示された 2 つの評価尺度、HDR-VDP-2.2 [70] および PU encoding [71] + MS-SSIM [72] を評価に用いた。

5.4.2 実験条件

実験には、オンライン上で利用可能なデータベース [73, 74] から選択した 60 枚の U-WDR 画像を利用した (図 5.2 参照)。提案する逆トーンマッピング法の評価は、逆トーンマッピングオペレータを用いる 5 つの逆トーンマッピング法、すなわち、正確なパラメータを用いた式 (5.6) によるトーンマッピング、パラメータなしで式 (5.6) を適用するトーンマッピング [51], PMET [24], Kuo らの方法 [22] および Huo らの方法 [23] と比較により行われた。提案法を含む式 (5.6) を用いた逆トーンマッピング法を、以降、Inverse photographic tone reproduction (IPTR) と呼ぶ。本実験は、3.4GHz プロセッサとメインメモリ 16GB を持つコンピュータ上で実行した (表 5.1 参照)。

Scenario 1

HDR-VDP-2.2 および PU encoding+MS-SSIM を用いた評価には参照画像として U-WDR 画像が必要であるが、一般の LDR 画像から逆トーンマッピングを実行する場合には、参照画像として利用できる U-WDR 画像 E がない。そこで、Scenario 1 における提案法の有効性の評価は、以下の手順で行った。

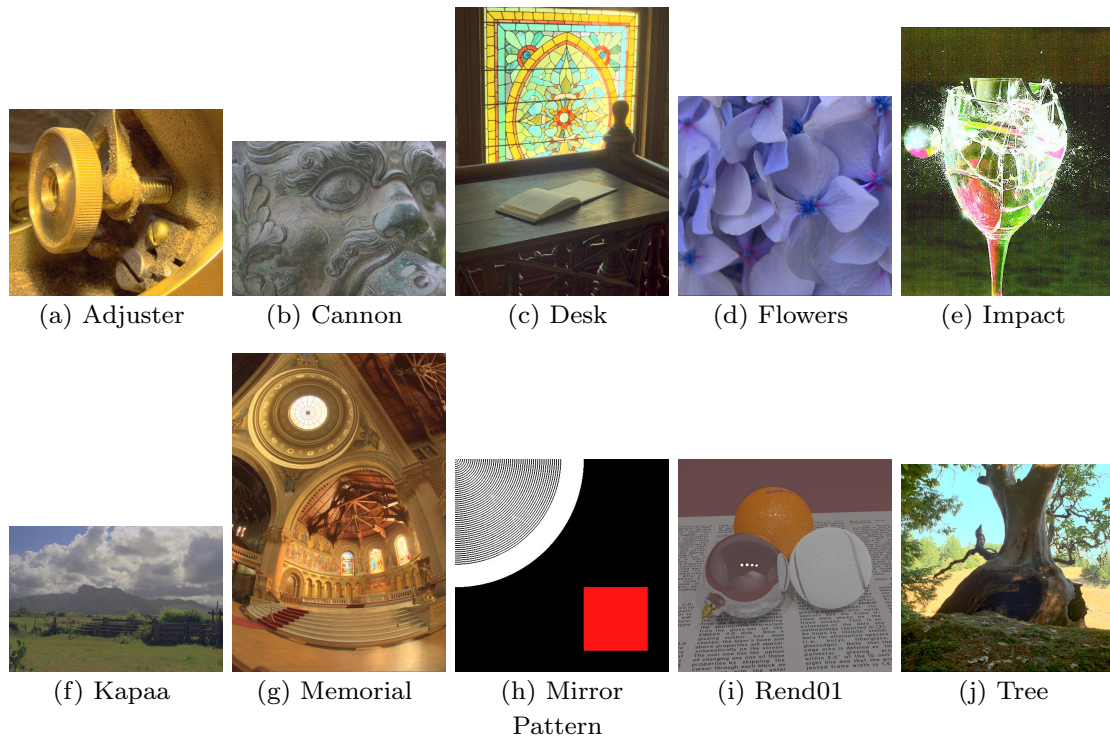


図 5.2 入力 LDR 画像の例

表 5.1 本実験で用いたコンピュータのスペック

Processor	Intel Core i7-3770 3.40GHz
Memory	16GB
OS	ubuntu 14.04 LTS
Software	MATLAB R2014b

1. 文献 [1] から選択した 10 個のトーンマッピング法を用いて, 入力 LDR 画像 \boldsymbol{x} を U-WDR 画像 \boldsymbol{E} から擬似的に生成する.
2. \boldsymbol{x} を入力画像として逆トーンマッピングを実行し, U-WDR 画像 $\hat{\boldsymbol{E}}$ を推定する.
3. \boldsymbol{E} を基準とする $\hat{\boldsymbol{E}}$ の品質を, 客観評価尺度 HDR-VDP-2.2 [70] および PU encoding [71]+MS-SSIM [72] により評価する.

手順 1 で使用した 10 個のトーンマッピング法 (TMO) は, 1. Reinhard のグローバル TMO (式 (5.4)), 2. パラメータ L を持つ Reinhard のグローバル TMO (式 (5.3)), 3. Reinhard のローカル TMO, 4. Logarithmic TMO, 5. Tumblin の TMO, 6. Schlick の TMO, 7. Chiu の TMO, 8. Ashikhmin の TMO, 9. Fattal の TMO, 10. Gamma

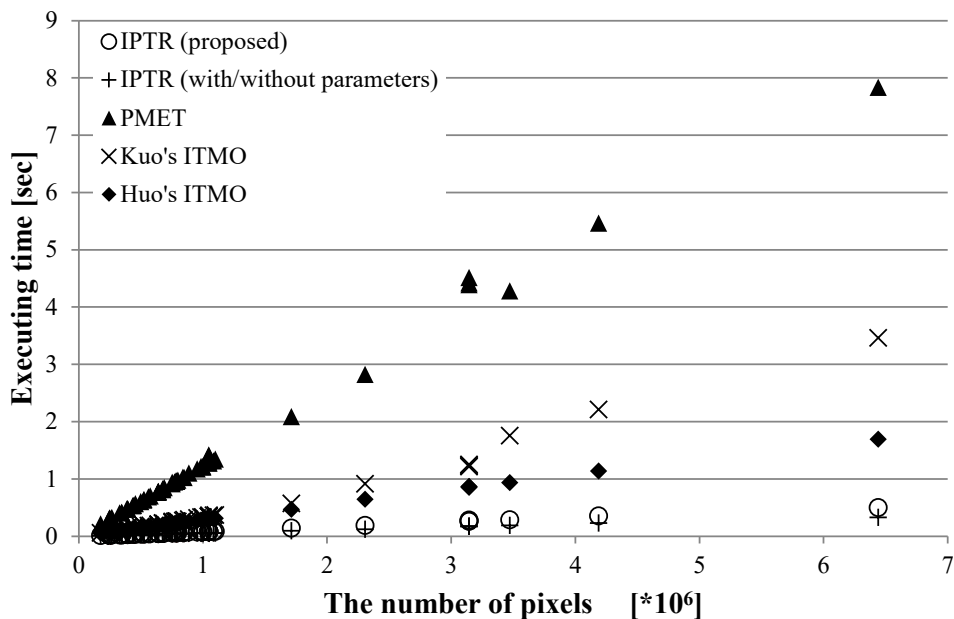


図 5.3 各逆トーンマッピング法の実行時間

TMO, である.

Scenario 2

Scenario 2 においては, Reinhard のグローバルオペレータ (式 (5.4) 参照) を用いる Photographic tone reproduction により, U-WDR 画像 E から L-WDR 画像 x を入力として用いた. この U-WDR 画像 E は, 品質評価における参照画像としても利用した. 他の条件は Scenario 1 の場合と同様である.

5.4.3 実験結果

実行時間

図 5.3 は, 60 枚の入力 LDR 画像それぞれに対し, 各逆トーンマッピング法を 100 回適用した際の平均実行時間である. この図より, 提案する逆トーンマッピング法は, 2つの正確なパラメータを利用する IPTR よりわずかに遅いものの, 他の従来法と比較して高速に実行できることがわかる. また, PMET は最も大きい計算コストを持つことが確認できる.

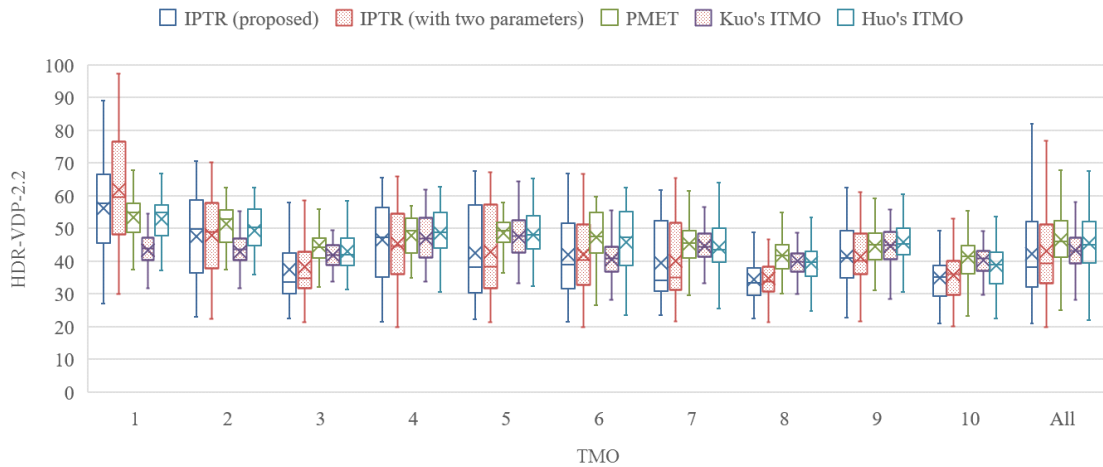


図 5.4 Scenario 1 における評価結果 (HDR-VDP-2.2). 各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表示しており、十字は平均値を示している. ここで、TMO は、1. Reinhard のグローバル TMO (式 (5.4)), 2. パラメータ L を持つ Reinhard のグローバル TMO (式 (5.3)), 3. Reinhard のローカル TMO, 4. Logarithmic TMO, 5. Tumblin の TMO, 6. Schlick の TMO, 7. Chiu の TMO, 8. Ashikhmin の TMO, 9. Fattal の TMO, 10. Gamma TMO である.

Scenario 1

図 5.4 および 5.5 は、10 個のトーンマッピングオペレータを用いて生成した入力画像 \mathbf{x} から推定された U-WDR 画像に対するスコア ($\text{HDR-VDP-2.2} \in [0, 100]$ および $\text{PU encoding} + \text{MS-SSIM} \in [0, 1]$) を箱ひげ図として示している. ここで、各スコアは値が大きいほど原 U-WDR 画像 \mathbf{E} と $\hat{\mathbf{E}}$ が近いことを示す. 各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値、すなわち第二四分位数 Q_2 を表示しており、十字は平均値を示している.

図 5.4 は、提案法が、Reinhard のグローバルオペレータにより生成された L-WDR 画像から、高い精度で U-WDR 画像を推定できることを示している. (図 5.4 中の“TMO 1”参照). 一方、他のトーンマッピング法を入力画像の生成に用いた場合には、提案法は従来法より高い性能を示さなかった. 結果として、図 5.4 中の“ALL”に示されている通り、全入力 LDR 画像について各逆トーンマッピング法を HDR-VDP-2.2 により評価し

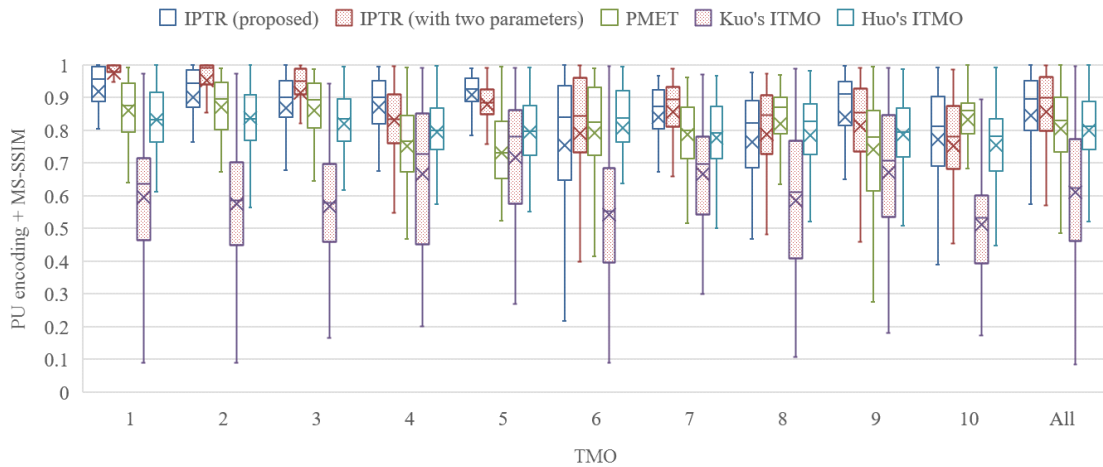


図 5.5 Scenario 1 における評価結果 (PU encoding + MS-SSIM). 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表示しており, 十字は平均値を示している. ここで, TMO は, 1. Reinhard のグローバル TMO (式 (5.4)), 2. パラメータ L を持つ Reinhard のグローバル TMO (式 (5.3)), 3. Reinhard のローカル TMO, 4. Logarithmic TMO, 5. Tumblin の TMO, 6. Schlick の TMO, 7. Chiu の TMO, 8. Ashikhmin の TMO, 9. Fattal の TMO, 10. Gamma TMO である.

た結果はほぼ同等となっている. 図 5.4 では PMET が最も高い HDR-VDP-2.2 スコアを示しているが, 図 5.5 からは中央値と平均値の双方について, 提案法が PMET, Kuo らの方法, Huo らの方法を上回る MS-SSIM スコアを示していることがわかる. MS-SSIM は 2 枚の画像間の構造的類似性を測る尺度であるため [69], 図 5.5 より, 提案法により推定された U-WDR 画像は, 他の手法と比較して, 原 U-WDR 画像との高い構造的類似性を持つことがわかる.

Scenario 2

表 5.2 および 5.3 には, U-WDR 画像 \hat{E} の品質を評価したスコアを示す.

表 5.2 および 5.3 より, 正確な 2 つのパラメータを用いる IPTR と同様の品質で, 提案法は U-WDR 画像を推定できていることが確認できる. これら 2 つの手法の微小な差は, L-WDR 画像の画素値が有限語長で表現されていることに起因する. 加えて, 提案法におけるパラメータ A あるいは G を推定することの有効性は, 提案法がパラメータを用いない IPTR より優れた結果を示していることから確かめられる. これらの結果より, 提案

表 5.2 Scenario 2 における評価結果 (HDR-VDP-2.2)

	IPTR (proposed, with a)	IPTR (proposed, with $G(l_{\mathbf{E}} \mathbb{P}))$)	IPTR (with two parameters)	IPTR [51] (without parameters)	PMET [24]	Kuo's ITMO [22]	Huo's ITMO [23]
Inverse TMO							
Adjuster	58.25	58.32	58.17	57.81	57.41	42.65	53.13
Cannon	95.00	95.53	95.53	70.69	55.17	49.92	56.58
Desk	49.45	49.76	49.76	50.50	44.66	43.88	43.89
Flowers	83.22	83.22	83.22	63.92	54.70	58.20	55.39
Impact	39.47	39.45	39.45	35.34	41.01	27.22	41.94
Kapaa	90.55	90.88	90.88	64.21	54.47	44.54	55.46
Memorial	49.60	49.71	49.61	51.61	45.33	39.60	44.65
Mirror Pattern	47.71	47.74	47.74	37.27	47.88	44.29	48.24
Rend01	80.39	80.39	80.39	56.98	72.03	42.37	77.34
Tree	54.21	55.18	55.18	55.65	51.92	34.31	51.57

表 5.3 Scenario 2 における評価結果 (PU encoding + MS-SSIM)

	IPTR (proposed, with a)	IPTR (proposed, with $G(l_{\mathbf{E}} \mathbb{P}))$)	IPTR (with two parameters)	IPTR [51] (without parameters)	PMET [24]	Kuo's ITMO [22]	Huo's ITMO [23]
Inverse TMO							
Adjuster	0.991	0.991	0.991	0.931	0.944	0.534	0.909
Cannon	1.000	1.000	1.000	0.990	0.773	0.744	0.747
Desk	0.933	0.938	0.938	0.940	0.783	0.661	0.738
Flowers	1.000	1.000	1.000	0.947	0.706	0.871	0.675
Impact	0.845	0.855	0.855	0.496	0.727	0.091	0.749
Kapaa	1.000	1.000	1.000	0.950	0.702	0.524	0.670
Memorial	0.981	0.982	0.981	0.961	0.879	0.581	0.856
Mirror Pattern	0.674	0.677	0.677	0.679	0.794	0.666	0.670
Rend01	1.000	1.000	1.000	0.803	0.989	0.487	0.999
Tree	0.973	0.978	0.978	0.944	0.892	0.295	0.868

法は、トーンマッピングされた LDR 画像から原 U-WDR 画像を再構成するために有効であるといえる。

これらの結果から、提案する逆トーンマッピングオペレータは、従来法と比較して低い計算コストで実行可能であると確かめられた。一般の LDR 画像が入力として与えられた場合には、提案法は、従来法と同等の品質を持つ U-WDR 画像を生成できた。加えて、提案法は、Reinhard のグローバルオペレータによって生成された L-WDR 画像から、高い精度で原 U-WDR 画像を生成できることが確認できた。したがって、提案する逆トーンマッピングオペレータは、2 つの Scenario 両方に有効であるといえる。

5.5 まとめ

本章では, Reinhard のグローバルオペレータの逆変換に基づく高速逆トーンマッピングオペレータを提案した提案する逆トーンマッピングオペレータは, 一般の LDR 画像からの U-WDR 画像の推定, および U-WDR 画像のトーンマッピングにより生成された L-WDR 画像から原 U-WDR 画像の再構成という 2 つの Scenario に対して適用可能である. また, 本章では, いくつかの条件下において, $G(l_{\mathbf{E}}|\mathbb{P})$ が a を用いて計算可能であることを示した. 同様に, a は $G(l_{\mathbf{E}}|\mathbb{P})$ から推定可能である. この推定は, 閉形式で記述されるアルゴリズムにより高速に実行できる.

実験によって, 提案法は, 従来の逆トーンマッピングオペレータと比較して低い計算コストで実行可能であることが示された. 加えて, 提案法は, 2 つの Scenario 双方で有効であることが示された. 構造的類似性の観点から高い品質の U-WDR 画像を推定できる.

次章では, 逆トーンマッピング法のさらなる高性能化のため, 本章で提案した逆トーンマッピングオペレータを深層学習に基づき発展させる.

第6章

逆トーンマッピングのための深層学習ネットワーク “iTMM-Net”

6.1 はじめに

本章では，前章で提案した Reinhard のグローバルオペレータに基づく逆トーンマッピング法を深層学習を用いて発展させ，さらに高性能な逆トーンマッピング法を提案する．前章までに述べてきたとおり，高性能な逆トーンマッピング法の実現には，入力 LDR 画像の撮影時に用いられたカメラの非線形な応答 f に合わせて，適切な逆トーンマッピングオペレータを入力画像に適用する必要がある．しかしながら，固定の逆トーンマッピングオペレータを用いる逆トーンマッピング法では，関数 f に合わせた画素値の線形化が実現できないため，高品質な U-WDR 画像の推定が難しい．

このような理由から，畳み込みニューラルネットワーク (Convolutional neural network, CNN) に基づく逆トーンマッピング法がいくつか提案されている [27–29]．これらの手法は，逆トーンマッピング法の性能を著しく向上させた．文献 [28] および [27] では，CNN を飽和領域の復元のために用いている一方，画素値の線形化にはこれまでと同様に CNN を用いていない．Marnerides らは，Min-max 正規化により U-WDR 画像を正規化し，それら画像を用いて学習された CNN による画素値の線形化法を提案している [29]．しかしながら，これらの正規化された U-WDR 画像の画素値は極めて狭い範囲に分布するため，それら画像を用いて学習された CNN では，十分な精度を持つ画素値の線形化を実現できない．

そこで，本章では，前章で提案した逆トーンマッピング法と CNN を組み合わせた逆トーンマッピングネットワークである “iTMM-Net” を提案する．本章では，Marnerides

らと同様に相対輝度を持つ U-WDR 画像の推定を目的とし, iTm-Net を画素値の線形化に用いる [29]. この実現のため, LDR 画像と U-WDR 画像間の非線形な関係を考慮した損失関数の提案も行う. この損失関数は, 可逆なトーンマッピングオペレータを用いて教師 U-WDR 画像を L-WDR 画像に変換し, その後, この L-WDR 画像と推定された画像との距離を計算する. 提案する損失関数は, U-WDR 画像を正則化するのみならず, U-WDR 画像の画素値を LDR 画像のように広く分布させることを可能にする. この損失関数を用いる iTm-Net は, 任意の解像度を持つ画像に適用可能とするため, 近年研究が盛んに行われている敵対的生成ネットワーク (Generative adversarial network, GAN) を利用していない.

最新の手法を含む従来の逆トーンマッピング法との比較実験により, HDR-VDP-2.2 および PU encoding + MS-SSIM の観点から, 提案法は従来法を上回る性能を持つことが明らかとなる. さらに, 提案する損失関数は, LDR 画像と U-WDR 画像間の非線形性を考慮していない通常の損失関数と比較して, CNN の性能を向上させることが示される.

6.2 逆トーンマッピングネットワーク “iTm-Net”

図 6.1 は, iTm-Net の学習手順, および学習された iTm-Net を用いた U-WDR 画像の推定手順を示している. 学習において, すべての入力 LDR 画像 \mathbf{x} は, 仮想カメラ \tilde{f} を用いることによって教師 U-WDR 画像 \mathbf{E} から生成される [27]. iTm-Net の出力画像 $\hat{\mathbf{y}}$ (以下, 予測画像) と, 教師 U-WDR 画像 \mathbf{E} 間の損失は, 非線形なトーンマッピングオペレータ f' を \mathbf{E} へ適用し計算される.

学習後は, 一般の様々な LDR 画像を, 提案する iTm-Net の入力として用いることができる. ここで, iTm-Net により得られる予測画像 $\hat{\mathbf{y}}$ は, 対応する U-WDR 画像をトーンマッピングすることで得られる L-WDR 画像に相当する. 予測画像 $\hat{\mathbf{y}}$ の線形化は, 逆トーンマッピングオペレータ f'^{-1} の適用により実行される. より詳細な学習条件は第 6.2.4 項で述べる.

6.2.1 損失関数

CNN の学習において, CNN 内のパラメータは, 損失関数を用いて評価される予測画像と教師画像間の誤差に基づき最適化される.

文献 [29] では, ℓ_1 距離に基づく損失関数 \mathcal{L}_1 とコサイン類似度に基づく損失関数 \mathcal{L}_{\cos}

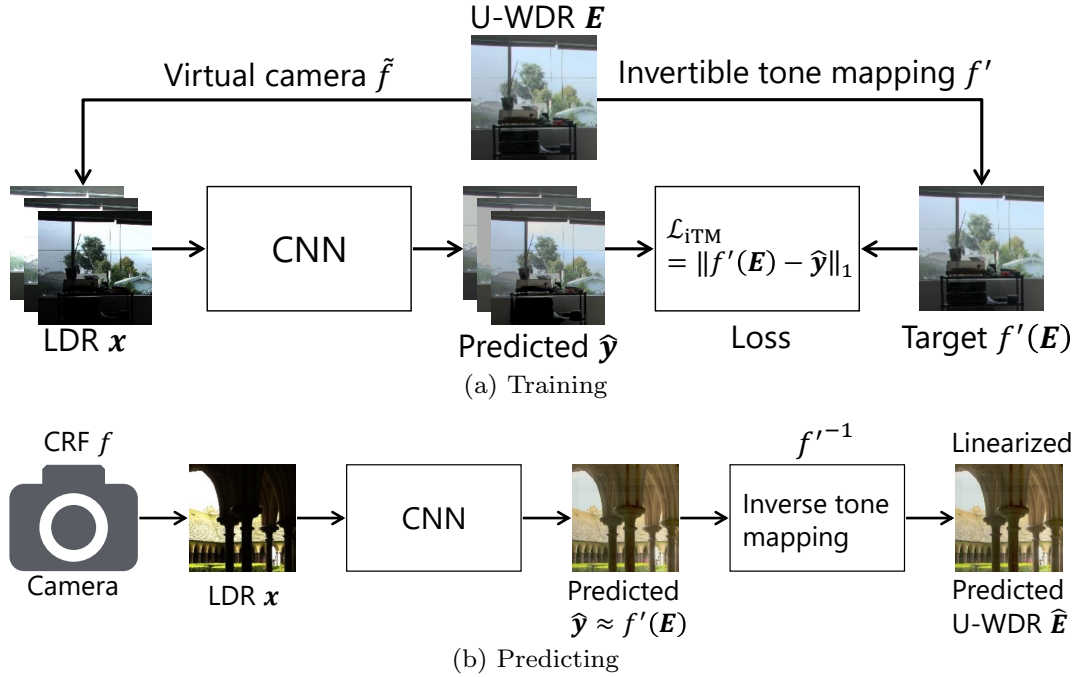


図 6.1 提案する逆トーンマッピング法

を組み合わせることで予測誤差を評価している。 \mathcal{L}_1 および \mathcal{L}_{\cos} は、以下のように与えられる。

$$\mathcal{L}_1(\hat{\mathbf{y}}, \mathbf{E}) = \frac{1}{|\mathbb{P}|} \sum_{\mathbf{p} \in \mathbb{P}} \|\mathbf{E}(\mathbf{p}) - \hat{\mathbf{y}}(\mathbf{p})\|_1 \quad (6.1)$$

$$\mathcal{L}_{\cos}(\hat{\mathbf{y}}, \mathbf{E}) = 1 - \frac{1}{|\mathbb{P}|} \sum_{\mathbf{p} \in \mathbb{P}} \frac{\mathbf{E}(\mathbf{p})^\top \hat{\mathbf{y}}(\mathbf{p})}{\|\mathbf{E}(\mathbf{p})\|_2 \|\hat{\mathbf{y}}(\mathbf{p})\|_2} \quad (6.2)$$

ここで、 $\mathbf{E}(\mathbf{p})$ および $\hat{\mathbf{y}}(\mathbf{p})$ は、画素 \mathbf{p} における U-WDR 画像 \mathbf{E} の画素値ベクトル、および予測画像 $\hat{\mathbf{y}}$ の画素値ベクトルをそれぞれ表す。また、 $|\mathbb{P}|$ は総画素数である。式 (6.1) と (6.2) を用いて、ExpandNet [29] を学習するための損失関数は、

$$\mathcal{L}_{\text{Expand}}(\hat{\mathbf{y}}, \mathbf{E}) = \mathcal{L}_1(\hat{\mathbf{y}}, m(\mathbf{E})) + \lambda \mathcal{L}_{\cos}(\hat{\mathbf{y}}, m(\mathbf{E})) \quad (6.3)$$

として与えられる。ここで、 $\lambda \in \mathbb{R}_+$ は、コサイン類似度の貢献度合いを調節するパラメータであり、 $m(\mathbf{E})$ は、画像 \mathbf{E} の輝度のレンジを $[0, 1]$ の区間へ線形に圧縮する、Min-max 正規化を意味する。しかしながら、U-WDR 画像を学習する際における Min-max 正規化の利用は適していない。これは、LDR 画像の場合とは異なり、U-WDR 画像の輝度値のレンジが極めて広い上に、輝度値が非一様に分布しているためである [27]。

この理由から、本章で提案する損失関数では、Min-max 正規化 $m(\cdot)$ の代わりに可逆のトーンマッピングオペレータ $f'(\cdot)$ を用いる。例えば、 $f'(\cdot)$ を適用した l_1 距離は、次式

として与えられる.

$$\mathcal{L}_{\text{iTM}}(\hat{\mathbf{y}}, \mathbf{E}) = \mathcal{L}_1(\hat{\mathbf{y}}, f'(\mathbf{E})) \quad (6.4)$$

本章では, 式 (5.4) に示した Reinhard のグローバルオペレータを用いる Photographic tone reproduction を, $f'(\cdot)$ とする [56] (第 5.2 節参照). このオペレータは逆関数を持つだけでなく, U-WDR 画像を高品質な L-WDR 画像にマッピングすることが可能である. また, 式 (5.1) による輝度のスケールリングは, 各 U-WDR 画像の輝度の幾何平均を a に合わせることにより, U-WDR 画像を規格化する. さらに, 式 (5.4) は, U-WDR 画像の輝度値を, LDR 画像のように, 区間 $[0, 1]$ の範囲に広く分布させる. 次章で示すように, U-WDR 画像の推定は, f' に対応する逆トーンマッピングオペレータ f'^{-1} を用いることで実行される.

6.2.2 U-WDR 画像の推定

提案する iTM-Net は式 (6.4) を損失関数として学習されるため, iTM-Net により予測される画像 $\hat{\mathbf{y}}$ は, U-WDR 画像 \mathbf{E} からトーンマッピングにより生成された L-WDR 画像 $f'(\mathbf{E})$ に相当する. したがって, U-WDR 画像の推定は, 逆トーンマッピングオペレータ f'^{-1} を予測画像 $\hat{\mathbf{y}}$ に適用することで, 次式のように実行される.

$$\hat{\mathbf{E}} = f'^{-1}(\hat{\mathbf{y}}) \quad (6.5)$$

式 (5.4) を用いる Photographic tone reproduction の利用の下では, この逆トーンマッピング f'^{-1} は, 式 (5.7), (5.8), および (5.11) を計算することによって実行される. ここで, 本章では相対輝度の推定を目的とするため, 式 (5.7) におけるパラメータ \mathcal{A}, \mathcal{G} は共に 1 とおく.

6.2.3 iTM-Net のアーキテクチャ

図 6.2 は, iTM-Net のアーキテクチャ全体を示す. iTM-Net は, ローカルエンコーダ, グローバルエンコーダ, およびデコーダから構成される. ローカルエンコーダへの入力, 解像度 $U \times V$ を持つ 24 ビットカラー画像である. 一方, グローバルエンコーダへの入力, ローカルエンコーダへの入力画像を一定の解像度 128×128 へ変換した画像である. 図 6.2 に示されているように, iTM-Net は 5 種類のブロックからなる.

3×3 Conv. + BN + ReLU このブロックでは, スライド 1, パディング 1, カーネルサイズ 3×3 の畳み込み層と Batch normalization (BN) 層 [75], Rectified linear unit 活性化関数 (ReLU) [76] を順に適用する. ローカルエンコーダとデコーダで

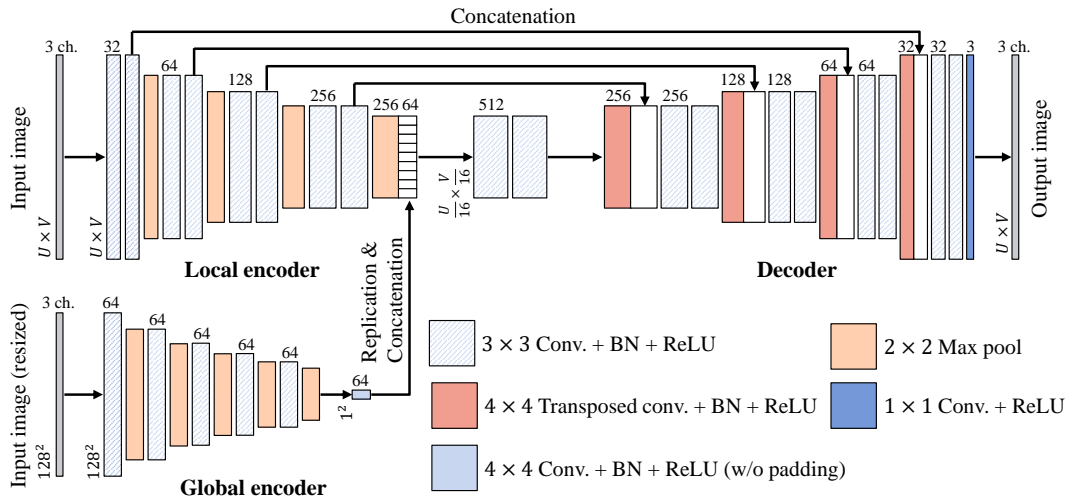


図 6.2 ネットワークアーキテクチャ. iTM-Net は、ローカルエンコーダ、グローバルエンコーダ、デコーダから構成される。各四角形はネットワーク中の各層によりそれぞれ出力される多チャンネルの特徴マップを表す。特徴マップのチャンネル数は、各四角形の上部に示している。特徴マップの解像度は、四角形の左に示している。

は、隣り合う 3×3 Conv. + BN + ReLU ブロックのカーネル数 C は共通である。このカーネル数 C は、入力側から出力側への順に $C = 32, 64, 128, 256, 512, 256, 128, 64, 32$ として与える。グローバルエンコーダにおいては、 3×3 Conv. + BN + ReLU ブロックはすべて、 $C = 64$ 個のカーネルを持つ。

2×2 Max pool このブロックでは、ストライド 2、カーネルサイズ 2×2 の最大値プーリング層によって、特徴マップをダウンサンプリングする。

4×4 Transposed Conv. + BN + ReLU このブロックでは、ストライド $1/2$ 、パディング 1、カーネルサイズ 4×4 の転置畳み込み層と BN, ReLU を順に適用する。カーネル数 C は、入力側から出力側への順に $C = 256, 128, 64, 32$ として与える。

1×1 Conv. + ReLU このブロックでは、ストライド 1、パディング 1、カーネルサイズ 1×1 の畳み込み層と ReLU を順に適用する。カーネル数は、 $C = 3$ である。

4×4 Conv. + BN + ReLU (w/o padding) このブロックでは、パディングなしのカーネルサイズ 4×4 の畳み込み層を適用する。カーネル数は、 $C = 64$ である。

iTM-Net におけるローカルエンコーダとデコーダは、U-Net のアーキテクチャとほぼ一致する [77]。また、U-Net と同様に、ローカルエンコーダとデコーダの間には Skip connection を持つ。

iTM-Net と U-Net の違いは、iTM-Net が、大域的な特徴を抽出するためのグローバルエンコーダを持つことである。最近の研究において、Marnerides らは、U-WDR 画像

を推定する場合には、U-Net の使用がブロックアーティファクトを発生させること指摘した [29]. 著者が実施した予備実験の結果、このブロックアーティファクトが発生する原因は、大域的な特徴を扱えないネットワークアーキテクチャであることが判明した. したがって、iTM-Net では、ローカルエンコーダとグローバルエンコーダにより抽出される特徴を結合することによって、ブロックアーティファクトの発生を抑制する.

このネットワークアーキテクチャに加え、提案する損失関数 \mathcal{L}_{iTM} の使用により、逆トーンマッピングの性能が向上する.

6.2.4 学習手順

iTM-Net の学習のためには、様々な条件で撮影された LDR 画像 \mathbf{x} と対応する U-WDR 画像 \mathbf{E} の組が必要である. 十分な量の学習データを準備するため、仮想カメラを用いて、U-WDR 画像 \mathbf{E} から入力 LDR 画像 \mathbf{x} を生成した [27]. 学習には、オンライン上で利用可能なデータベース [61, 62, 73, 74, 78, 79] から収集した 336 枚の U-WDR 画像を利用した.

iTM-Net の学習手順は以下の通りである.

- i 336 枚の U-WDR 画像の中からランダムに 16 枚の画像を選択する.
- ii 各 U-WDR 画像から、入力画像と教師画像の組 $(\mathbf{x}, \tilde{\mathbf{E}})$ をそれぞれ生成し、計 16 組の画像を得る. 各組は、以下の手順に従い生成する.
 - (a) U-WDR 画像 \mathbf{E} から $N \times N$ 画素の領域 $\tilde{\mathbf{E}}$ を切り出す. ここで、 N は、区間 $[0.2, 0.6]$ 上の一様乱数と、各画像 \mathbf{E} の短辺の長さ $\min(U, V)$ の積として与える. また、領域を切り出す位置は一様乱数に従い決定する.
 - (b) $\tilde{\mathbf{E}}$ を 256×256 画素にリサイズする.
 - (c) $\tilde{\mathbf{E}}$ をそれぞれ 50% の確率で左右・上下方向に反転させる.
 - (d) 式 (2.2) に従い、露出 \mathbf{X} を $\tilde{\mathbf{E}}$ から計算する. ここで、シャッタースピード Δt は、文献 [56] に従い $\Delta t_v = 0.18 \cdot 2^v / G(l_{\tilde{\mathbf{E}}}|\mathbb{P})$ として与えた (式 (2.1) 参照). また、 v は、区間 $[-2, 2]$ 上の一様乱数とした. またセンサの飽和をシミュレーションするため、式 (2.4) に従い、 \mathbf{X} の 1 を超える画素値は切り捨てた.
 - (e) 入力 LDR 画像 \mathbf{x} を、仮想カメラ \tilde{f} により以下として生成する.

$$\mathbf{x}(\mathbf{p}) = \tilde{f}(X(\mathbf{p})) = \frac{\tilde{h}(l_{\mathbf{x}}(\mathbf{p}))}{l_{\mathbf{x}}(\mathbf{p})} \mathbf{X} \quad (6.6)$$

$$\tilde{h}(t) = (1 + \eta) \frac{t^\zeta}{t^\zeta + \eta} \quad (6.7)$$

乱数 η および ζ は、平均 0.6、分散 0.1 を持つガウス分布および平均 0.9、分散 0.1 を持つガウス分布にそれぞれ従う。

- iii 16 枚の入力 LDR 画像 \mathbf{x} それぞれから、iTM-Net により 16 枚の予測画像 $\hat{\mathbf{y}}$ を得る。
- iv 予測画像 $\hat{\mathbf{y}}$ と教師画像 $\tilde{\mathbf{E}}$ 間の誤差を、式 (6.4) と Reinhard グローバルオペレータ f' を用いて計算する。ここで、 $a = 0.18$ を式 (5.1) において用いた。
- v 誤差逆伝搬により、iTM-Net のパラメータ、すなわち重み ω とバイアス b を更新する。

次節で説明する実験のためには、上記手順を 42 回繰り返したものを 1 エポックとして、iTM-Net を 1000 エポック学習した。各エポックにおける手順 i では、同一 U-WDR 画像が複数回選択されることがないようにした。

iTM-Net のパラメータの初期化には He らの手法を用いた [80]。また、誤差逆伝搬における最適化アルゴリズムとして、Adam 最適化法 [81] を利用した。ここで、Adam におけるパラメータは $\alpha = 0.002, \beta_1 = 0.9, \beta_2 = 0.999$ とした。iTM-Net は、Tensorflow をバックエンドとする Keras フレームワークにより実装した。また、学習は、単一の NVIDIA GeForce 1080Ti GPU を用いて実行した。

6.3 実験および考察

本節では、2 つの客観評価尺度と主観的な評価を用いた、提案法の性能評価実験とその結果について説明する。

6.3.1 実験条件

iTM-Net により推定される U-WDR 画像 $\hat{\mathbf{E}}$ の品質を、2 つの客観評価尺度 HDR-VDP-2.2 [70]、および PU encoding [71] + MS-SSIM [72]、により評価した。ここで、この評価には、原 U-WDR 画像 \mathbf{E} を参照画像として利用した。

これら 2 つの評価尺度は、シーンの絶対輝度を画素値として持つ 2 枚の U-WDR 画像間の差異を評価するために設計されている。そのような絶対輝度を持つ U-WDR 画像は、データセット [62] で公開されている約 100 枚の画像しか無い。そこで、44 枚の U-WDR 画像をデータセット [62] から無作為に選択し、評価実験に利用した。これら 44 枚の画像は、iTM-Net の学習には使用していない。入力 LDR 画像 \mathbf{x} は、第 6.2.4 項における手順 ii(d) および ii(e) に従い生成した。加えて、逆トーンマッピング法は相対輝度を持つ U-WDR 画像しか推定できないため、推定された U-WDR 画像 $\hat{\mathbf{E}}$ を、その輝度のレンジ

が原 U-WDR 画像 \mathbf{E} と一致するようにスケーリングした。

比較手法には、最新の手法を含む 3 つの従来法、すなわち、前章で提案した逆トーンマッピング法 (Direct ITMO) [26], Pseudo-multi-exposure-based tone fusion (PMET) [24], ExpandNet [29] を利用した。ここで、ExpandNet は CNN に基づく手法であり、他の手法は深層学習を用いない方法である。ExpandNet には、開発者らによって公開されている学習済みのモデルを用いた。さらに、提案する損失関数の有効性を明確にするため、iTm-Net は異なる 3 つの損失関数を用いて学習された。それら損失関数は、提案する損失関数 \mathcal{L}_{iTM} (iTm-Net with \mathcal{L}_{iTM}), トーンマッピングなしの標準的な l_1 損失 $\mathcal{L}_1(\hat{\mathbf{y}}, m(\mathbf{E}))$ (iTm-Net with \mathcal{L}_1), および ExpandNet で用いられている損失関数 $\mathcal{L}_{\text{Expand}}(\hat{\mathbf{y}}, \mathbf{E})$ (iTm-Net with $\mathcal{L}_{\text{Expand}}$) である。

6.3.2 実験結果

図 6.3 および 6.4 は、実験に用いた 44 枚すべての画像について、HDR-VDP および PU encoding + MS-SSIM を用いて評価した結果を箱ひげ図として示す。各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す。箱の中にある横線は中央値、すなわち第二四分位数 Q_2 を表示しており、十字は平均値を示している。これら評価尺度は共に、値が大きいほど推定された U-WDR 画像が原 U-WDR 画像に類似していることを示す。

図 6.3 および 6.4 から確認できるように、双方の評価尺度について、すべての iTm-Net は、ExpandNet を含む 3 つの従来法より高い中央値および平均値を示した。この結果から、3 つの iTm-Net によりそれぞれ推定される U-WDR 画像は、従来の方法で推定される画像より原 U-WDR 画像に類似しているといえる。推定されたすべての U-WDR 画像は原 U-WDR 画像の輝度のレンジに合わせてスケーリングされていることから、この結果は、提案法が画像の線形化を高精度に実行できることを示す。したがって、提案するネットワークアーキテクチャの使用が、ExpandNet のアーキテクチャを使用した場合より高品質な U-WDR 画像の推定を可能とすることが確認できた。

iTM-Net with \mathcal{L}_1 および iTM-Net with $\mathcal{L}_{\text{Expand}}$ と比較して、iTm-Net with \mathcal{L}_{iTM} は、両方の評価尺度について高いスコアを示した。したがって、提案する損失関数は、逆トーンマッピングのための CNN を学習するために有効である。

図 6.5 および 6.6 は、提案法を含む 6 つの比較手法により推定された U-WDR 画像の例である。ここで、LDR デバイスでは U-WDR 画像を表示できないため、推定された

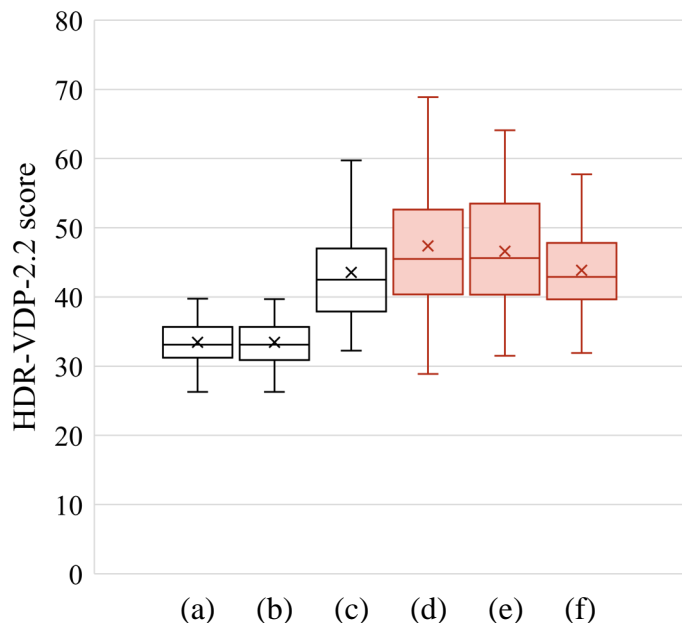


図 6.3 HDR-VDP-2.2 スコア. (a) ITMO [26], (b) PMET [24], (c) ExpandNet [29], (d) iTM-Net with \mathcal{L}_{iTM} (Proposed), (e) iTM-Net with \mathcal{L}_1 , (f) iTM-Net with \mathcal{L}_{Expand} . 各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は、中央値および平均値をそれぞれ表す.

U-WDR 画像をトーンマッピングして表示している. また, 原 U-WDR 画像の輝度のレンジに合わせたスケージングは, ここでは実行していない. 図 6.5 と 6.6 から, 提案法によって推定された U-WDR 画像は, 他の手法と比較して, 原 U-WDR 画像 \tilde{E} とより類似する高品質な画像であることがわかる.

これらの理由から, 高品質な U-WDR 画像を単一 LDR 画像から推定するために, 提案法が有効であると示された. 特に, 提案する損失関数の使用により, 逆トーンマッピングのための CNN の性能が向上すると確認された.

6.4 まとめ

本章では, 高品質な U-WDR 画像を単一 LDR 画像から推定するため, “iTM-Net” という逆トーンマッピングネットワークを提案した. iTM-Net の学習のため, U-WDR 画像と LDR 画像間の非線形な関係を考慮した損失関数もここで提案した. 提案する損失関数においては, 可逆なトーンマッピングオペレータを用いて教師 U-WDR 画像を L-WDR

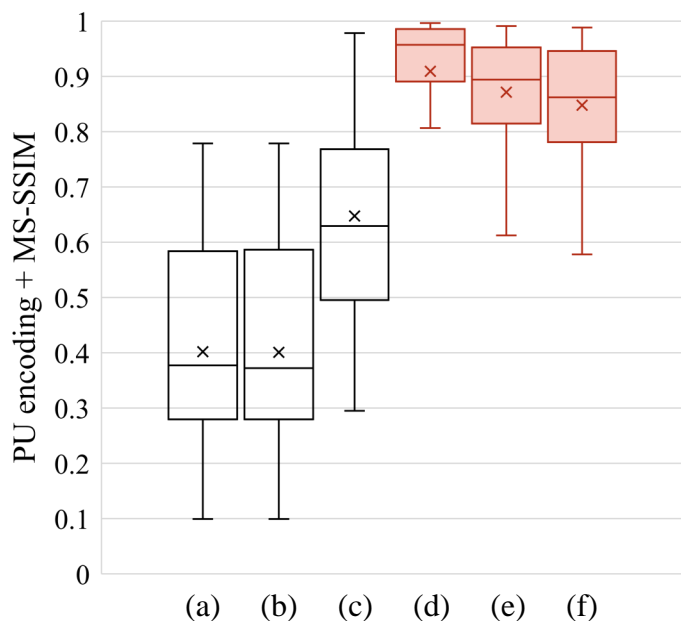


図 6.4 PU-encoding + MS-SSIM スコア (a) ITMO [26], (b) PMET [24], (c) ExpandNet [29], (d) iTM-Net with \mathcal{L}_{iTM} (Proposed), (e) iTM-Net with \mathcal{L}_1 , (f) iTM-Net with \mathcal{L}_{Expand} . 各箱は、第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し、ひげは、 $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す。箱の中にある横線および十字は、中央値および平均値をそれぞれ表す。

画像へ変換する。提案する損失関数の使用は、U-WDR 画像を正規化するのみならず、U-WDR 画像の画素値を LDR 画像のように分布させることができる。結果として、逆トーンマッピングのための CNN の性能を向上させることができる。実験により、提案する損失関数で学習された iTM-Net によって推定された U-WDR 画像は、最先端の手法を含む従来法より、HDR-VDP-2.2 および PU encoding + MS-SSIM の観点から高品質であることが明らかとなった。加えて、提案する損失関数を使用して学習される CNN は、U-WDR 画像と LDR 画像間の非線形な関係を考慮しない損失関数を用いた場合と比較して、より高い性能を持つと示された。

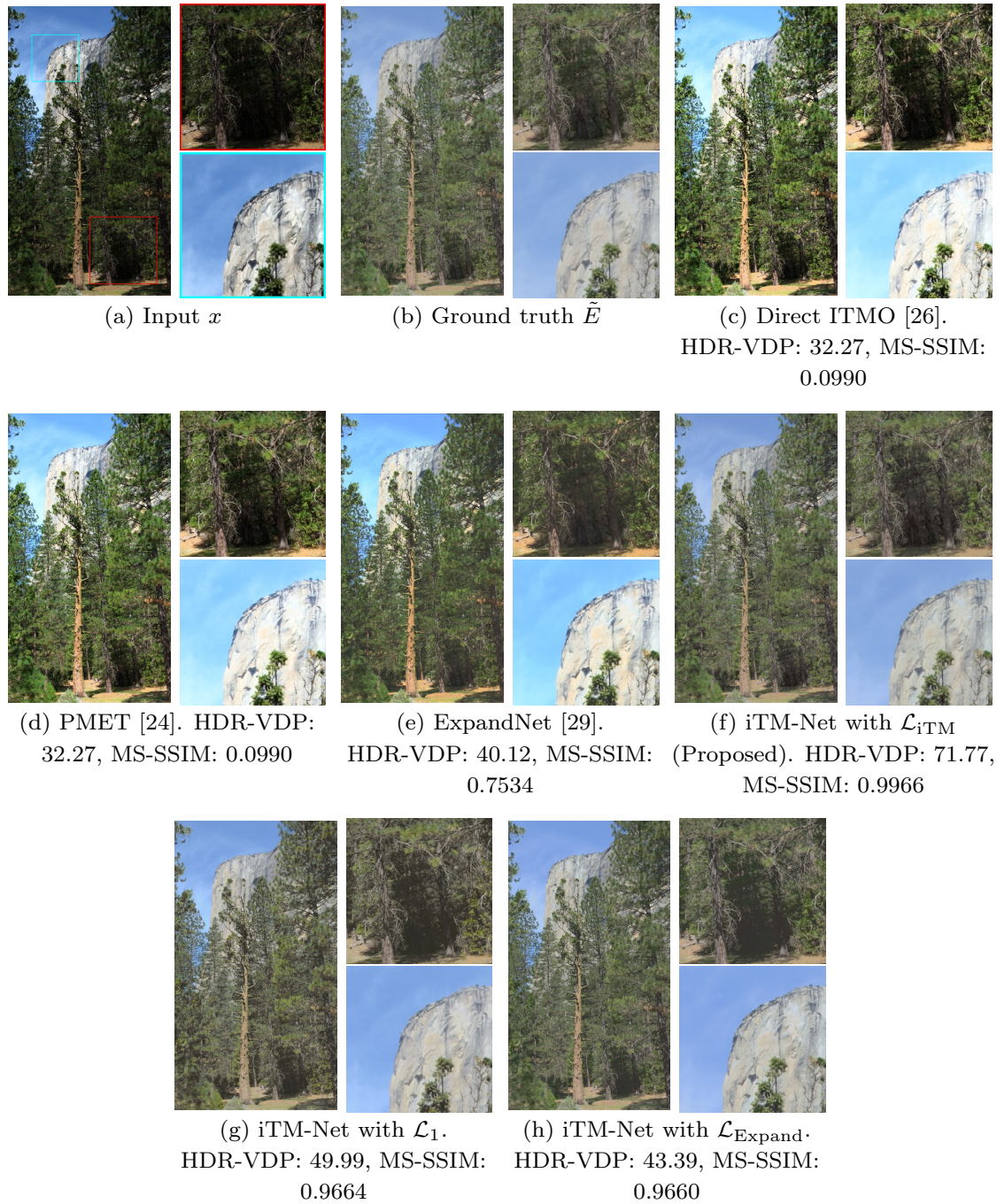


図 6.5 結果画像 “ElCapitan”. 枠で囲われた領域の拡大図は、各結果画像の右に示す. U-WDR 画像 (b)–(h) は、可視化のため、原 U-WDR 画像の輝度のレンジに合わせたスケールなしにトーンマッピングされている. 提案する iTM-Net (f) は、原 U-WDR 画像 (b) に最も類似する画像を生成した.

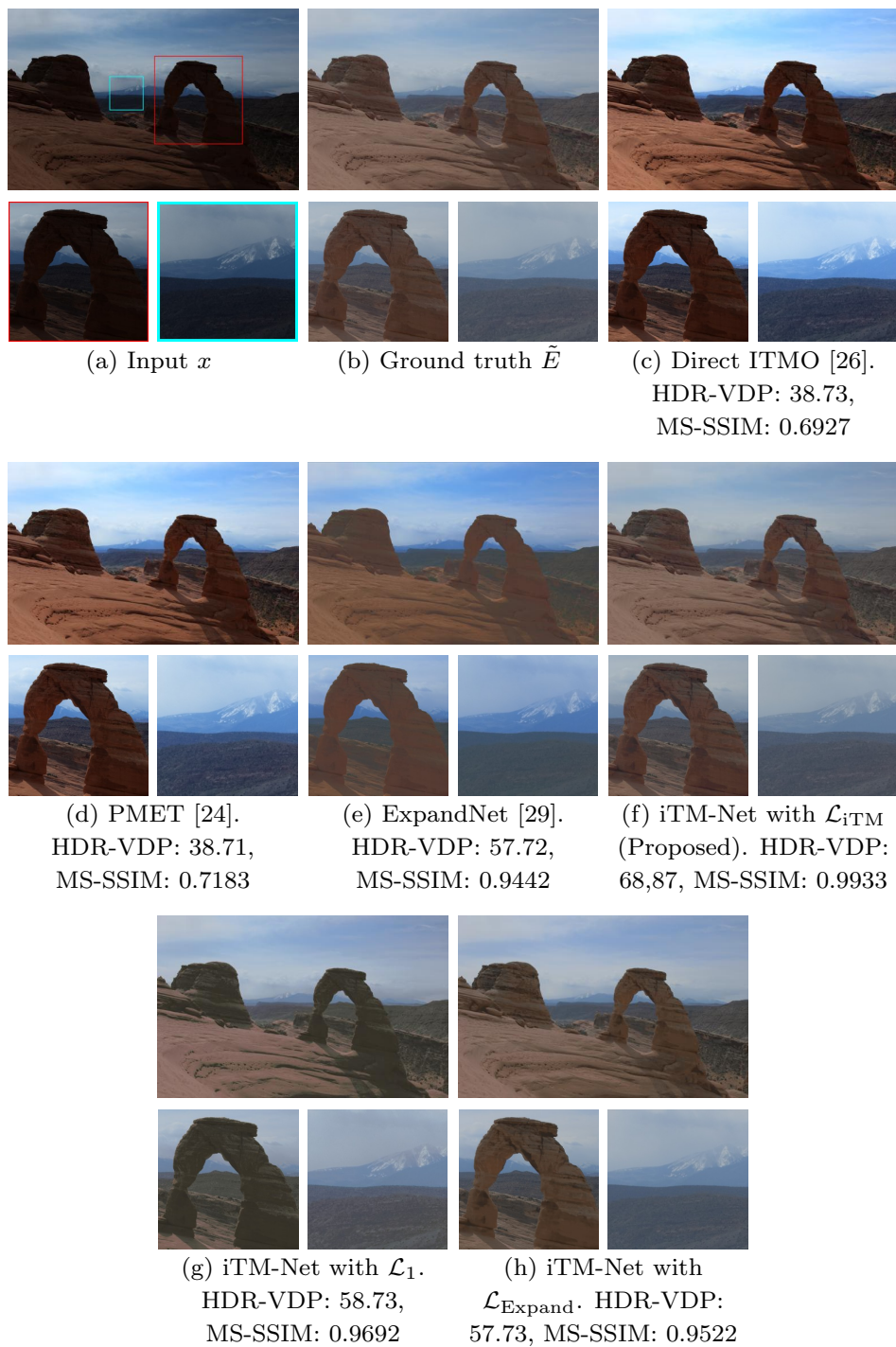


図 6.6 結果画像 “DelicateArch”. 枠で囲われた領域の拡大図は、各結果画像の下に示す。U-WDR 画像 (b)–(h) は、可視化のため、原 U-WDR 画像の輝度のレンジに合わせたスケールなしにトーンマッピングされている。提案する iTM-Net (f) は、原 U-WDR 画像 (b) に最も類似する画像を生成した。

第7章

総論

本論文では，不明瞭な多重露出画像が入力として与えられた場合を想定した L-WDR 画像生成，単一 LDR 画像からの L-WDR 画像推定法の性能向上，および U-WDR 画像推定法の高速化と性能向上を目的として，4つの手法を提案した．提案した手法を用いることで，適切な多重露出画像の撮影が難しい状況においても，撮影される不明瞭な多重露出画像を補正することで明瞭な多重露出画像を生成でき，それらの合成として高品質な L-WDR 画像を生成できる．さらに，そもそも多重露出画像が取得できない状況においても，提案した多重露出画像推定法，あるいは逆トーンマッピング法を用いて L-WDR 画像または U-WDR 画像を生成できる．

各章で述べた内容や，提案した手法の利点をまとめると，以下のようになる．第1章と第2章では，本研究の背景や目的，デジタル撮影と輝度のダイナミックレンジ，および WDR 画像技術について説明し，WDR 画像生成法における課題について述べた．

第3章では，シーンの領域分割に基づく多重露出画像補正法を提案した．提案法では，多重露出画像の輝度を補正することによって，不明瞭な多重露出画像からさえ明瞭な多重露出画像の生成を可能とした．高品質な L-WDR 画像は，補正によって得られるこれら多重露出画像を，従来の L-WDR 画像生成法を用いて合成することにより生成される．提案法による多重露出画像の補正は，局所コントラスト強調，シーン領域分割に基づく輝度補正，トーンマッピングという処理により行う．特に，シーン領域分割に基づく輝度補正は，輝度に関してシーンを領域分割し，分割によって得られる各領域をよく表現する画像をそれぞれ生成することにより行われる．ここで，このシーン領域分割のために，第3章では2つの Approach を提案した．既存の WDR 画像生成法を用いた実験により，提案法の利用が，不明瞭な多重露出画像から生成される L-WDR 画像の品質を向上させることが示された．

第 4 章では、シーン領域分割に基づき、単一 LDR 画像から多重露出画像を推定する方法を提案した。第 3 章で提案したシーン領域分割法を単一 LDR 画像に対して拡張することで、単一 LDR 画像からの擬似的な多重露出画像の生成を可能とした。これら擬似的に生成された多重露出画像の合成により、高品質な L-WDR 画像が生成される。単一 LDR 画像の強調に基づく L-WDR 画像推定法との比較により、主観的および客観的品質の観点から提案法の有効性を確認した。

第 5 章では、単一 LDR 画像から U-WDR 画像を推定する高速逆トーンマッピングオペレータを提案した。提案法では、Reinhard のグローバルオペレータの逆関数に基づき、単一 LDR 画像のダイナミックレンジを拡張する。さらに、Reinhard のグローバルオペレータの逆関数を計算するために必要な 2 つのパラメータ、 A, G が、トーンマッピングにおけるパラメータ $a, G(l_E|\mathbb{P})$ のどちらか一方を用いて、閉形式で計算できることを示した。このことが、提案法による高速な逆トーンマッピングを実現した。加えて、提案法は、Reinhard のグローバルオペレータによって生成された L-WDR 画像から、元の U-WDR 画像を高精度に復元できるという特徴を持つ。評価により、提案法は、従来法と同等の品質を持つ U-WDR 画像を高速に推定できることが示された。

第 6 章では、第 5 章で提案した逆トーンマッピングオペレータと深層学習を組み合わせ、逆トーンマッピングのための深層ニューラルネットワーク “iTM-Net” を提案した。Reinhard のグローバルオペレータで生成された L-WDR 画像が入力として与えられた場合に、第 5 章の逆トーンマッピングオペレータは極めて高い性能を持つ。そのため、一般の入力 LDR 画像からその条件を満たすような画像を CNN により予測し、第 5 章で提案した逆トーンマッピングを実行する。加えて、損失関数内でトーンマッピング処理を用いることが、逆トーンマッピングのための CNN の効果的な学習を可能とした。

7.1 今後の課題

本論文における今後の課題としては、以下の 4 つが考えられる。

第一の課題は、入力 LDR 画像のノイズを考慮した多重露出画像補正法の検討である。第 3 章で提案した多重露出画像補正法は、入力 LDR 画像のノイズが十分無視できる場合を想定していた。一方、不明瞭な多重露出画像には一般に強いノイズが含まれるため、そのようなノイズを考慮することによって、提案する補正法によって得られる多重露出画像の、さらなる品質向上が期待される。

第二の課題は、多重露出画像補正法の HDR カメラへの応用の検討である。第 2 章で述べた通り、HDR カメラは、複数のセンサを用いて一度に多重露出画像を撮影することで、

広い輝度のダイナミックレンジを記録することを可能としている。つまり、従来の LDR カメラを用いた多重露出画像の撮影と、HDR カメラを用いた多重露出画像の撮影の違いは、撮影時における時間ずれの有無である。また、HDR カメラで撮影される多重露出画像の枚数は、ハードウェアによって制限される。すなわち、HDR カメラを用いた場合でも、十分な枚数の多重露出画像を撮影することは難しい。したがって、提案法による不明瞭な多重露出画像の補正が、HDR カメラを用いて撮影される WDR 画像の品質向上に寄与すると期待できる。

第三の課題は、飽和領域の復元と輝度値の線形化の双方を同時に実現する、逆トーンマッピング法の開発である。飽和領域の復元を可能とする逆トーンマッピング法は既に提案されている。また、本論文で提案した iTM-Net は、輝度値の高精度な線形化を実現した。しかしながら、それら両方を同時に実現した手法は存在しない。したがって、そのような逆トーンマッピング法の開発により、さらに高品質な U-WDR 画像の推定が可能となることが期待できる。

第四の課題は、多重露出画像推定法および逆トーンマッピング法のデータ拡張への応用である。現在さまざまな分野への応用が研究されている深層学習は、そのモデルの構築のために、大量のデータを必要とする。より少量のデータのみを用いてモデルを学習するために、データの水増しを行うデータ拡張技術が広く用いられている。第 4 章で提案した多重露出画像推定法は単一 LDR 画像から複数の画像を生成できる。また、逆トーンマッピングにより U-WDR 画像を推定することは、仮想カメラを用いた大量の LDR 画像の生成を可能とする。したがって、これら手法をデータ拡張法として利用できる可能性がある。

参考文献

- [1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*. Natick, MA, USA: AK Peters (CRC Press), Feb. 2011.
- [2] M. Aggarwal and N. Ahuja, “Split Aperture Imaging for High Dynamic Range,” *Int. J. Comput. Vis.*, vol. 58, no. 1, pp. 7–17, Jun. 2004.
- [3] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, “A versatile HDR video production system,” *ACM Trans. Graph.*, vol. 30, no. 4, pp. 41:1—41:10, Jul. 2011.
- [4] S. Nayar and T. Mitsunaga, “High dynamic range imaging: spatially varying pixel exposures,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Hilton Head Island, SC: IEEE Comput. Soc, Jun. 2000, pp. 472–479.
- [5] V. G. An and C. Lee, “Single-shot high dynamic range imaging via deep convolutional neural network,” in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.* Kuala Lumpur: IEEE, Dec. 2017, pp. 1768–1772.
- [6] B. C. Madden, “Extended intensity range imaging,” *Univ. Pennsylvania Dep. Comput. Inf. Sci. Tech. Rep. No. MS-CIS-93-96.*, 1993.
- [7] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in *Proc. ACM SIGGRAPH*. Los Angeles, CA, USA: ACM Press, Aug. 1997, pp. 369–378.
- [8] T. Mertens, J. Kautz, and F. Van Reeth, “Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography,” *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, Mar. 2009.
- [9] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, “Inverse tone mapping,” in *Proc. Int. Conf. Comput. Graph. Interact. Tech. Australas. Southeast Asia*. Kuala Lumpur, Malaysia: ACM Press, 2006, pp. 349–356.

- [10] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, “Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs,” *ACM Trans. Graph.*, vol. 26, no. 3, pp. 39:1–39:6, Jul. 2007.
- [11] K. Zuiderveld, “Contrast Limited Adaptive Histogram Equalization,” in *Graph. gems IV*, P. S. Heckbert, Ed. San Diego, CA: Elsevier, 1994, pp. 474–485.
- [12] S.-C. Huang, F.-C. Cheng, and Y.-S. Chiu, “Efficient Contrast Enhancement Using Adaptive Gamma Correction With Weighting Distribution,” *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1032–1041, Mar. 2013.
- [13] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based hdr reconstruction of dynamic scenes,” *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203:1—203:11, Nov. 2012.
- [14] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, and I. S. Kweon, “Robust High Dynamic Range Imaging by Rank Minimization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1219–1232, Jun. 2015.
- [15] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, and L. Zhang, “Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach,” *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2519–2532, May 2017.
- [16] X. Wu, X. Liu, K. Hiramatsu, and K. Kashino, “Contrast-accumulated histogram equalization for image enhancement,” in *Proc. IEEE Int. Conf. Image Process.* Beijing: IEEE, Sep. 2017, pp. 3190–3194.
- [17] X. Guo, Y. Li, and H. Ling, “LIME: Low-Light Image Enhancement via Illumination Map Estimation,” *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.
- [18] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, “A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 2782–2790.
- [19] H. Su and C. Jung, “Low light image enhancement based on two-step noise suppression,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* New Orleans, LA: IEEE, Mar. 2017, pp. 1977–1981.
- [20] X. Ren, M. Li, W.-H. Cheng, and J. Liu, “Joint Enhancement and Denoising Method via Sequential Decomposition,” in *Proc. IEEE Int. Symp. Circuits Syst.* Florence: IEEE, May 2018, pp. 1–5.

- [21] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to See in the Dark,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* Salt Lake City, UT: IEEE, Jun. 2018, pp. 3291–3300.
- [22] P.-H. Kuo, C.-S. Tang, and S.-Y. Chien, “Content-adaptive inverse tone mapping,” in *Proc. Vis. Commun. Image Process.* San Diego, CA: IEEE, Nov. 2012, pp. 1–6.
- [23] H. Youngquing, Y. Fan, and V. Brost, “Dodging and burning inspired inverse tone mapping algorithm,” *J. Comput. Inf. Syst.*, vol. 9, no. 9, pp. 3461–3468, May 2013.
- [24] T.-H. Wang, C.-W. Chiu, W.-C. Wu, J.-W. Wang, C.-Y. Lin, C.-T. Chiu, and J.-J. Liou, “Pseudo-Multiple-Exposure-Based Tone Fusion With Local Region Adjustment,” *IEEE Trans. Multimed.*, vol. 17, no. 4, pp. 470–484, Apr. 2015.
- [25] Y. Kinoshita, S. Shiota, and H. Kiya, “Fast inverse tone mapping with Reinhard’s global operator,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* New Orleans, LA: IEEE, Mar. 2017, pp. 1972–1976.
- [26] ———, “Fast Inverse Tone Mapping Based on Reinhard’s Global Operator with Estimated Parameters,” *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E100.A, no. 11, pp. 2248–2255, Nov. 2017.
- [27] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, “HDR image reconstruction from a single exposure using deep CNNs,” *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–15, Nov. 2017.
- [28] Y. Endo, Y. Kanamori, and J. Mitani, “Deep reverse tone mapping,” *ACM Trans. Graph.*, vol. 36, no. 6, pp. 177:1–177:10, Nov. 2017.
- [29] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, “ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content,” in *Comput. Graph. Forum*, vol. 37, no. 2. Wiley Online Library, May 2018, pp. 37–49.
- [30] ISO/IEC, “ISO/IEC 18477 Information technology - Scalable compression and coding of continuous-tone still images,” 2015.
- [31] CIE, *Commission internationale de l’Eclairage proceedings*. Cambridge, UK: Cambridge University Press, 1932.
- [32] F. Dufaux, P. L. Callet, R. Mantiuk, and M. Mrak, *High Dynamic Range Video, From Acquisition, to Display and Applications*. Cambridge, MA: Academic

Press, 2016.

- [33] M. Grossberg and S. Nayar, “Determining the camera response from images: What is knowable?” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1455–1467, Nov. 2003.
- [34] “Photons to Photos.” [Online]. Available: <http://www.photonstophotos.net/Charts/PDR.htm>
- [35] EIZO, “ColorEdge PROMINENCE CG3145-BS.” [Online]. Available: <https://www.eizo.co.jp/products/ce/cg3145/index.html>
- [36] ITU, “Recommendation ITU-R BT.2100-2: Image parameter values for high dynamic range television for use in production and international programme exchange,” 2018.
- [37] S. Marschner, “Image-Based BRDF Measurement,” PhD thesis, Stanford University, 1998.
- [38] S. Mann and R. W. Picard, “On being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures,” in *Proc. IS&T*, May 1995, pp. 422–428.
- [39] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. A. Lensch, “Optimal HDR reconstruction with linear digital cameras,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* San Francisco, CA: IEEE, Jun. 2010, pp. 215–222.
- [40] A. Badki, N. Khademi Kalantari, and P. Sen, “Robust Radiometric Calibration for Dynamic Scenes in the Wild,” in *Proc. IEEE Int. Conf. Comput. Photogr.* Houston, TX: IEEE, Apr. 2015, pp. 1–10.
- [41] A. A. Goshtasby, “Fusion of multi-exposure images,” *Image Vis. Comput.*, vol. 23, no. 6, pp. 611–618, Jun. 2005.
- [42] A. Saleem, A. Beghdadi, and B. Boashash, “Image fusion-based contrast enhancement,” *EURASIP J. Image Video Process.*, vol. 2012, no. 10, pp. 1–17, Dec. 2012.
- [43] J. Wang, G. Xu, and H. Lou, “Exposure fusion based on sparse coding in pyramid transform domain,” in *Proc. Int. Conf. Internet Multimed. Comput. Serv.*, ser. ICIMCS ’15. Zhangjiajie city, Hunan, China: ACM Press, Aug. 2015, pp. 1–4.
- [44] Z. Li, J. Zheng, Z. Zhu, and S. Wu, “Selectively Detail-Enhanced Fusion of Differently Exposed Images With Moving Objects,” *IEEE Trans. Image Process.*,

- vol. 23, no. 10, pp. 4372–4382, Oct. 2014.
- [45] T. Sakai, D. Kimura, T. Yoshida, and M. Iwahashi, “Hybrid method for multi-exposure image fusion based on weighted mean and sparse representation,” in *Proc. Eur. Signal Process. Conf.* Nice: IEEE, Aug. 2015, pp. 809–813.
- [46] M. Nejati, M. Karimi, S. R. Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, “Fast exposure fusion using exposedness function,” in *Proc. IEEE Int. Conf. Image Process.* Beijing: IEEE, Sep. 2017, pp. 2234–2238.
- [47] E. H. Land, “The retinex theory of color vision,” *Sci. Am.*, vol. 237, no. 6, pp. 108–129, 1977.
- [48] Y. Kinoshita and H. Kiya, “Scene Segmentation-Based Luminance Adjustment for Multi-Exposure Image Fusion,” *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4101–4116, Aug. 2019.
- [49] —, “Automatic exposure compensation using an image segmentation method for single-image-based multi-exposure fusion,” *APSIPA Trans. Signal Inf. Process.*, vol. 7, p. e22, Dec. 2018.
- [50] Y. Kinoshita, S. Shiota, and H. Kiya, “A Pseudo Multi-Exposure Fusion Method Using Single Image,” *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E101.A, no. 11, pp. 1806–1814, Nov. 2018.
- [51] Y. Kinoshita, S. Shiota, M. Iwahashi, and H. Kiya, “An Remapping Operation without Tone Mapping Parameters for HDR Images,” *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E99.A, no. 11, pp. 1955–1961, Nov. 2016.
- [52] Y. Kinoshita and H. Kiya, “iTM-Net: Deep Inverse Tone Mapping Using Novel Loss Function Considering Tone Mapping Operator,” *IEEE Access*, vol. 7, pp. 73 555–73 563, 2019.
- [53] Y. Kinoshita, S. Shiota, H. Kiya, and T. Yoshida, “Multi-Exposure Image Fusion Based on Exposure Compensation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* Calgary, AB: IEEE, Apr. 2018, pp. 1388–1392.
- [54] Y. Kinoshita, S. Shiota, and H. Kiya, “Automatic Exposure Compensation for Multi-Exposure Image Fusion,” in *Proc. IEEE Int. Conf. Image Process.* Athens: IEEE, Oct. 2018, pp. 883–887.
- [55] J. Chen, S. Paris, and F. Durand, “Real-time edge-aware image processing with the bilateral grid,” in *ACM Trans. Graph.*, vol. 26, no. 3. ACM, Jul. 2007, p. 103.

- [56] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, “Photographic tone reproduction for digital images,” *ACM Trans. Graph.*, vol. 21, no. 3, pp. 267–276, Jul. 2002.
- [57] A. Kanezaki, “Unsupervised Image Segmentation by Backpropagation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* Calgary, AB: IEEE, Apr. 2018, pp. 1543–1547.
- [58] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
- [59] C. M. Bishop, *Pattern Recognition and Machine Learning*. NY: Springer-Verlag New York, 2006.
- [60] “HDR photography gallery.” [Online]. Available: <https://www.easyhdr.com/examples/>
- [61] P. Zolliker, Z. Barańczuk, D. Küpper, I. Sprow, and T. Stamm, “Creating HDR video content for visual quality assessment using stop-motion,” in *Proc. Eur. Signal Process. Conf.* Marrakech: IEEE, Sep. 2013, pp. 1–5.
- [62] “The HDR Photographic Survey.” [Online]. Available: <http://rit-mcsl.org/fairchild/HDRPS/HDRthumbs.html>
- [63] Shutao Li, Xudong Kang, and Jianwen Hu, “Image Fusion With Guided Filtering,” *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.
- [64] K. Ma, Kai Zeng, and Zhou Wang, “Perceptual Quality Assessment for Multi-Exposure Image Fusion,” *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
- [65] H. Rahman, R. Soundararajan, and R. V. Babu, “Evaluating Multiexposure Fusion Using Image Information,” *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1671–1675, Nov. 2017.
- [66] H. Yeganeh and Z. Wang, “Objective Quality Assessment of Tone-Mapped Images,” *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.
- [67] Z. Ying, G. Li, and W. Gao, “A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement,” *arXiv Prepr. arXiv1711.00591*, Nov. 2017. [Online]. Available: <http://arxiv.org/abs/1711.00591>
- [68] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and

- K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.
- [69] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, “Benchmarking of objective quality metrics for HDR image quality assessment,” *EURASIP J. Image Video Process.*, vol. 2015, no. 39, pp. 1–18, Dec. 2015.
- [70] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, “HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images,” *J. Electron. Imaging*, vol. 24, no. 1, p. 010501, Jan. 2015.
- [71] T. O. Aydın, R. Mantiuk, and H.-P. Seidel, “Extending quality metrics to full luminance range images,” in *Proc. SPIE Hum. Vis. Electron. Imaging XIII*, B. E. Rogowitz and T. N. Pappas, Eds. International Society for Optics and Photonics, Feb. 2008, pp. 68 060B:1–68 060B:10.
- [72] Z. Wang, E. Simoncelli, and A. Bovik, “Multiscale structural similarity for image quality assessment,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Pacific Grove, CA, USA: IEEE, Nov. 2003, pp. 1398–1402.
- [73] “GitHub - openexr.” [Online]. Available: <https://github.com/openexr/>
- [74] “High Dynamic Range Image Examples.” [Online]. Available: <http://www.anywhere.com/gward/hdrenc/pages/originals.html>
- [75] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *arXiv Prepr. arXiv1502.03167*, pp. 1–11, Feb. 2015. [Online]. Available: <http://arxiv.org/abs/1502.03167>
- [76] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in *Proc. Int. Conf. Artif. Intell. Stat.*, Ft. Lauderdale, FL, USA, Apr. 2011, pp. 315–323.
- [77] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Med. Image Comput. Comput. Interv.*, ser. LNCS, vol. 9351. Springer, Nov. 2015, pp. 234–241.
- [78] “Max Planck institut informatik.” [Online]. Available: <http://resources.mpi-inf.mpg.de/hdr/gallery.html>
- [79] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi, “Visual attention in LDR and HDR images,” in *Proc. 9th Int. Work. Video Process. Qual. Metrics Consum. Electron.*, Chandler, Arizona, Feb. 2015, pp. 1–6.
- [80] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing

Human-Level Performance on ImageNet Classification,” in *Proc. IEEE Int. Conf. Comput. Vis.* Santiago, Chile: IEEE, Dec. 2015, pp. 1026–1034.

- [81] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv Prepr. arXiv1412.6980*, pp. 1–15, Dec. 2014. [Online]. Available: <http://arxiv.org/abs/1412.6980>

目次

1.1	各章の関係	6
2.1	デジタルカメラによる画像撮影の手順	11
2.2	画像撮影における輝度のダイナミックレンジの比較	12
2.3	画像表示における輝度のダイナミックレンジの比較	14
2.4	LDR 画像, WDR 画像, HDR 画像の違い	15
3.1	輝度を補正した多重露出画像の例. 輝度補正により, 多重露出画像の品質が向上する.	25
3.2	合成された L-WDR 画像の例. (a) 図 3.1(a) に示す画像を合成したもの. (b) 図 3.1(b) に示す画像を合成したもの. (c) 図 3.1(c) に示す画像を合成したもの. 合成された L-WDR 画像の品質は, 補正の度合いにより変化する.	25
3.3	L-WDR 画像生成における提案法の使用法	26
3.4	提案するシーン領域分割に基づく輝度のスケールリング (SSLS) (a) SSLS の概要. (b) と (c) 図 3.4(a) におけるシーン領域分割の 2 つの Approach. (d) 図 3.4(a) における輝度のスケールリング.	29
3.5	入力多重露出画像 “Window”. シーンの右側の領域がすべての画像で不明瞭である.	34
3.6	入力多重露出画像 “Ostrow Tumski”. タワーと右側の陸地がすべての画像で不明瞭である.	34

3.7	提案法を Mertens ら, Sakai ら, Nejati らの手法と組み合わせた場合の比較 (“Window”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. 従来の多重露出画像の補正を行わない合成法は, 図 3.5 に示す不明瞭な多重露出画像から明瞭な画像を生成できない一方で, 提案法は明瞭な画像の生成を可能とする.	36
3.8	提案法を Li らおよび Ma らの手法と組み合わせた場合の比較 (“Window”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. 従来の多重露出画像の補正を行わない合成法は, 図 3.5 に示す不明瞭な多重露出画像から明瞭な画像を生成できない一方で, 提案法は明瞭な画像の生成を可能とする.	37
3.9	提案法を Mertens ら, Sakai ら, Nejati らの手法と組み合わせた場合の比較 (“Ostrow Tumski”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. 従来の多重露出画像の補正を行わない合成法は, 図 3.6 に示す不明瞭な多重露出画像から明瞭な画像を生成できない一方で, 提案法は明瞭な画像の生成を可能とする.	38
3.10	提案する SSLA によって補正された多重露出画像 (“Window”). Approach 1 は, ドアと床などの暗い領域と中間の明るさを持つ領域を分割できなかった. 一方, Approach 2 は, 大きな K が与えられた場合にも, ある輝度のレンジを持つ領域へシーンを適切に分割することができた.	39
3.11	提案する SSLA によって補正された多重露出画像 (“Ostrow Tumski”). Approach 1 は, 空と水辺を分割することでできなかったが, Approach 2 は, 大きな K が与えられた場合でさえ, シーンをある輝度のレンジを持つ領域へ適切に分割できた.	40
3.12	提案法における局所コントラスト強調の効果 (“Window”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. ここで, これら画像の生成には, Nejati らの手法 [46] と提案する Approach 2 を用いた. 局所コントラスト強調は画像構造を明瞭にすることができる一方で, しばしばノイズやリングングを増幅させる.	41
3.13	提案法を Ma らの手法と組み合わせた場合の比較 (“Lady eating”). 枠で囲われた領域の拡大図は, 各合成画像の下に示す. 提案する SSLA は, 動的なシーンに対しても有効である.	42

3.14	トーンマッピングにより生成された 500 組の多重露出画像に対する定量評価結果. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は, 中央値および平均値をそれぞれ表す.	48
3.15	トーンマッピングにより生成された 50 組の条件の良い多重露出画像 (十分な数の露出を含む多重露出画像) に対する定量評価結果. 多重露出画像の合成には Ma らの手法を使用した. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は, 中央値および平均値をそれぞれ表す.	49
3.16	8 組の多重露出画像に対する平均実行時間. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す. Approach 1 は Approach 2 より高速であることが確認できる.	50
4.1	単一 LDR 画像に基づく多重露出画像推定法	54
4.2	提案法の実行結果 (“Chinese garden”). (a): 入力 LDR 画像, (b): 提案するシーンの領域分割の結果, (c): 最終的に得られた L-WDR 画像, (d)-(j): 提案法によって推定された多重露出画像. (b) において, 各色は分割された各領域に対応する.	57

4.3	提案法の実行結果 (“Trashbox”). (a): 入力画像, (b): 提案するシーンの領域分割の結果, (c): 最終的に得られた L-WDR 画像, (d)-(f): 提案法により推定された多重露出画像. (b) において, 各色は分割された各領域に対応する.	58
4.4	固定の M および α_m を使用した場合の結果画像 (“Arno”). (a): 入力画像, (b)-(d): M と α_m 双方を固定した場合の結果, (e): 提案法による結果, (f)-(h): M のみを固定した場合の結果. 枠で囲われた領域の拡大図は, 各画像の下に示される.	59
4.5	提案法と従来の画像強調法との比較 (“Window”). 枠で囲われた領域の拡大図は, 各画像の下に示される. 提案法は, 過強調および強調不足なしに明瞭な画像を生成できる.	60
4.6	提案法と従来の画像強調法との比較 (“Estate rsa”). 枠で囲われた領域の拡大図は, 各画像の下に示される. 提案法は, 過強調および強調不足なしに明瞭な画像を生成できる.	61
4.7	Discrete entropy による評価結果. (a) Input image, (b) HE, (c) CLAHE, (d) AGCWD, (e) CACHE, (f) LLIE, (g) LIME, (h) SRIE, (i) BIMEF, および (j) Proposed. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表す.	62
4.8	Statistical naturalness による評価結果. (a) Input image, (b) HE, (c) CLAHE, (d) AGCWD, (e) CACHE, (f) LLIE, (g) LIME, (h) SRIE, (i) BIMEF, および (j) Proposed. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表す.	62
5.1	提案する逆トーンマッピング法	70
5.2	入力 LDR 画像の例	73
5.3	各逆トーンマッピング法の実行時間	74

5.4	Scenario 1 における評価結果 (HDR-VDP-2.2). 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表示しており, 十字は平均値を示している. ここで, TMO は, 1. Reinhard のグローバル TMO (式 (5.4)), 2. パラメータ L を持つ Reinhard のグローバル TMO (式 (5.3)), 3. Reinhard のローカル TMO, 4. Logarithmic TMO, 5. Tumblin の TMO, 6. Schlick の TMO, 7. Chiu の TMO, 8. Ashikhmin の TMO, 9. Fattal の TMO, 10. Gamma TMO である.	75
5.5	Scenario 1 における評価結果 (PU encoding + MS-SSIM). 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線は中央値を表示しており, 十字は平均値を示している. ここで, TMO は, 1. Reinhard のグローバル TMO (式 (5.4)), 2. パラメータ L を持つ Reinhard のグローバル TMO (式 (5.3)), 3. Reinhard のローカル TMO, 4. Logarithmic TMO, 5. Tumblin の TMO, 6. Schlick の TMO, 7. Chiu の TMO, 8. Ashikhmin の TMO, 9. Fattal の TMO, 10. Gamma TMO である.	76
6.1	提案する逆トーンマッピング法	81
6.2	ネットワークアーキテクチャ. iTM-Net は, ローカルエンコーダ, グローバルエンコーダ, デコーダから構成される. 各四角形はネットワーク中の各層によりそれぞれ出力される多チャンネルの特徴マップを表す. 特徴マップのチャンネル数は, 各四角形の上部に示している. 特徴マップの解像度は, 四角形の左に示している.	83
6.3	HDR-VDP-2.2 スコア. (a) ITMO [26], (b) PMET [24], (c) ExpandNet [29], (d) iTM-Net with \mathcal{L}_{iTM} (Proposed), (e) iTM-Net with \mathcal{L}_1 , (f) iTM-Net with $\mathcal{L}_{\text{Expand}}$. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は, 中央値および平均値をそれぞれ表す.	87

6.4	PU-encoding + MS-SSIM スコア (a) ITMO [26], (b) PMET [24], (c) ExpandNet [29], (d) iTM-Net with \mathcal{L}_{iTM} (Proposed), (e) iTM-Net with \mathcal{L}_1 , (f) iTM-Net with $\mathcal{L}_{\text{Expand}}$. 各箱は, 第一四分位数 Q_1 から第三四分位数 Q_3 までの範囲を表し, ひげは, $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ の範囲に含まれる最大値および最小値を示す. 箱の中にある横線および十字は, 中央値および平均値をそれぞれ表す.	88
6.5	結果画像 “ElCapitan”. 枠で囲われた領域の拡大図は, 各結果画像の右に示す. U-WDR 画像 (b)–(h) は, 可視化のため, 原 U-WDR 画像の輝度のレンジに合わせたスケーリングなしにトーンマッピングされている. 提案する iTM-Net (f) は, 原 U-WDR 画像 (b) に最も類似する画像を生成した.	89
6.6	結果画像 “DelicateArch”. 枠で囲われた領域の拡大図は, 各結果画像の下に示す. U-WDR 画像 (b)–(h) は, 可視化のため, 原 U-WDR 画像の輝度のレンジに合わせたスケーリングなしにトーンマッピングされている. 提案する iTM-Net (f) は, 原 U-WDR 画像 (b) に最も類似する画像を生成した.	90

表目次

2.1	ダイナミックレンジに関する画像の分類	15
2.2	提案する手法の位置づけ. ◎は従来法より優れていることを表し, ○は従来法と同等であることを示す.	21
3.1	提案法における調整可能なパラメータ	33
3.2	静的なシーンに対する Discrete entropy スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する.	45
3.3	静的なシーンに対する Statistical naturalness スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する.	46
3.4	動的なシーンに対する Discrete entropy スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す.	47

3.5	動的なシーンに対する Statistical naturalness スコア. 太字は, 各合成法内で比較してより高いスコアを示す. “w/o” は, 画像 y が輝度補正なしで生成されたことを意味する. 一方, “Prop. 1” および “Prop. 2” は, Approach 1 を用いる提案法および Approach 2 を用いる提案法をそれぞれ意味する. 加えて, “Prop. 2 w/o CE” は, Approach 2 を用いる提案法を局所コントラスト強調なしで適用した場合を示す.	47
3.6	実行時間計測に用いたコンピュータスペック	50
4.1	提案法により分割された領域 $\{P_m\}$ の数 M の例 ($K = 10$)	56
5.1	本実験で用いたコンピュータのスペック	73
5.2	Scenario 2 における評価結果 (HDR-VDP-2.2)	77
5.3	Scenario 2 における評価結果 (PU encoding + MS-SSIM)	77

謝辞

本研究は、著者が首都大学東京大学院システムデザイン研究科情報科学域において、多くの方々の御指導、御協力の元に進めたものであります。

はじめに、指導教官である貴家仁志教授には、本研究の全般にわたり、進行、執筆、発表に関する熱心な御指導、御助言を賜りました。ここに心より厚く御礼申し上げます。また、小野順貴教授、村松正吾教授、および藤吉正明准教授には、本論文の審査を通して貴重な御助言と御指導を賜り、深く感謝の意を表します。塩田さやか助教には、本研究のみならず、各種機器の使用法、会議の手続きなどをはじめとする各方面において大変貴重な御助言、御指導を頂きました。ここに深く感謝致します。著者が在学中にお世話になった先輩、公私にわたり良き相談相手となってくれた同輩、後輩に感謝致します。最後に、これまでの学生生活を理解し、温かい御支援を頂いた家族に心から感謝致します。

これらの支援がなければ本研究を完成させることはできませんでした。心から感謝致します。