

高ダイナミックレンジ画像の学習に向けた撮影法の開発

東海大学情報理工学部情報科学科 木下裕磨

1 はじめに

2025 年現在, 大規模な画像データセットを用いて学習された深層生成モデルが大きな注目を集めている. 例えば, 50 億枚以上の画像を用いて学習された Stable Diffusion は, 実際の写真と見間違ふほどの高精細な画像を生成できることで, 画像生成 AI として話題となった. このような深層生成モデルは, 画像を生成するのみならず, 画像の超解像やノイズの除去, 白黒写真のカラー化, 古い写真の復元など, 実現困難だった種々の画像処理へ応用できることから, 世界中の研究者により数多くの研究が進められている. しかしながら, 深層生成モデルの性能は, 学習に用いるデータセットにより制限されることに注意が必要である.

Stable Diffusion の学習に用いられた LAION-5B をはじめ, 深層生成モデルの学習に用いられる大規模画像データセットは, インターネットで公開されている画像を収集することで構築されている. これら画像の多くは, スマートフォンに搭載されたカメラやデジタル一眼カメラ, コンパクトデジタルカメラなど, 広く普及しているデジタルカメラを用いて撮影されたものであると考えられる. しかし, 撮像センサが扱うことの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. そのため, 一般的なデジタルカメラで撮影された画像では, 人間が知覚できる輝度のダイナミックレンジさえも忠実に表現するには至っていない. したがって, これら画像を用いて学習される現在の深層生成モデルでは, 現実シーンの輝度の分布を獲得できておらず, 一般的なデジタルカメラで撮影できる以上の輝度情報は表現できない.

このような背景から, 本研究は, 現実シーンの広い輝度情報を記録した高ダイナミックレンジ (High dynamic range, HDR) 画像を深層生成モデルに学習させることでさらなる性能向上を目指している. この目標を達成するため, 本研究では, (1) 大量の HDR 画像を収集するための基盤となる画像の高精度位置合わせアルゴリズム, (2) 深層生成モデルによる生成画像の主観品質の Blackbox 最適化アルゴリズム, および, (3) HDR 画像の可視化のための輝度領域分割に基づくトーンマッピング法を開発した.

2 画像の高精度位置合わせアルゴリズム

x および y として表現される二次元離散信号, 特に画像のアライメント (位置合わせ) 問題に着目する. 画素位置 $\mathbf{p} = (p_1, p_2)^\top$ における画素値は, $x[\mathbf{p}]$ および $y[\mathbf{p}]$ と表される. ここで, 添字 $\top$ はベクトルまたは行列の転置を意味する. 画像 x と y 間の空間的なズレが, 平行移動 $\Delta\mathbf{p} \in \mathbb{R}^2$, 回転 $\mathbf{R}(\theta)$, およびスケーリング $s \in \mathbb{R}^+$ の組み合わせで記述可能, すなわち, 画像 y が x の相似変換である場合, 両者の関係は以下のように定式化できる.

$$y[\mathbf{p}] = x[s\mathbf{R}(\theta)\mathbf{p} + \Delta\mathbf{p}], \quad (1)$$

ここで,

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (2)$$

本手法の目的は, 画像 x と y の間に存在する平行移動 $\Delta\mathbf{p}$, 回転角 θ , スケール係数 s を同時に推定し, 式 (1) で定義される相似変換を補正することである. この目的は, フーリエメリン変換 (Fourier-Mellin Transform; FMT) と補助関数法に基づくサブピクセル単位の相互相関最大化を組み合わせることで達成される.

2.1 アルゴリズムの流れ

アルゴリズムの流れは以下とおりである．

1. それぞれの入力画像に対し二次元離散フーリエ変換（DFT）を施し，振幅スペクトルを得る．
2. 振幅スペクトルを対数極座標（log-polar）での表現に変換する．
3. 対数極座標領域において，振幅スペクトル間の相互相関を補助関数法で最大化し，スケール s と回転 θ を推定する．
4. 得られた s と θ で画像を補正した後，画像間の相互相関を再度最大化し，平行移動 $\Delta \mathbf{p}$ を推定する．
5. 推定されたパラメータ s ， θ ，および $\Delta \mathbf{p}$ を用いて画像を補正し，整列された画像を得る．

相互相関の最大化には次節で述べる補助関数法によるアルゴリズムを用いることで，サブピクセル単位での高精度な位置合わせを実現する．

2.2 補助関数法による相互相関最大化

x および y の二次元離散フーリエ変換を $\hat{x}$ ， $\hat{y}$ とし，いずれもサイズ $N \times M$ とする．ここで， $\hat{x}$ および $\hat{y}$ は厳密な帯域制限を有していると仮定する．すなわち，少なくとも一つの要素が π である角周波数 ω について， $\hat{x}(\omega) = \hat{y}(\omega) = 0$ とする．

ここで，二次元クロススペクトル $\hat{\Phi}_2^{(xy)}(\omega_{kl}) = \hat{x}^*(\omega_{kl})\hat{y}(\omega_{kl})$ を用いると， x と y の一般化二次元相互相関関数は次式で与えられる．

$$\tilde{\Phi}_2^{(xy)}[\mathbf{p}] = \frac{1}{NM} \sum_{k \in K} \sum_{l \in L} w_{kl} \hat{\Phi}_2^{(xy)}(\omega_{kl}) \exp(j\langle \omega_{kl}, \mathbf{p} \rangle), \quad (3)$$

ここで， j は虚数単位， $w_{kl} \in \mathbb{R}^+$ は任意の重み， $\omega_{kl} = (\omega_k, \omega_l)^\top = (\frac{2\pi k}{N}, \frac{2\pi l}{M})^\top$ である．集合 K および L はそれぞれ $K = \{-N/2 + 1, -N/2 + 2, \dots, N/2\}$ ， $L = \{-M/2 + 1, -M/2 + 2, \dots, M/2\}$ と定義される． $w_{kl} = 1$ の場合，この関数は標準的な相互相関となる．一方， $w_{kl} = |\hat{\Phi}_2^{(xy)}(\omega_{kl})|^{-1}$ とした場合，この手法は位相限定相関（phase-only correlation）[1] と等価となり，音響信号処理分野では GCC-PHAT [2] とも呼ばれる．

我々は，式 (3) 中の $\mathbf{p}$ を実数値ベクトルと見なした連続関数 $\Phi_2^{(xy)}$ を考え，次式を最大化するアルゴリズムを提案した．

$$\tilde{\Delta \mathbf{p}} = \arg \max_{\mathbf{p} \in \mathbb{R}^2} \Phi_2^{(xy)}(\mathbf{p}) \quad (4)$$

この最大化に補助関数法（auxiliary function method）を利用することで，サブピクセル精度の変位推定と高速な局所最適解への収束を同時に実現している．具体的には，式 (4) を直接最大化する代わりに，下界となる補助関数 $Q(\mathbf{p}, \theta)$ を反復的に最大化する．

$$\theta^{(i)} = f(\mathbf{p}^{(i)}), \quad \mathbf{p}^{(i+1)} = \arg \max_{\mathbf{p} \in \mathbb{R}^2} Q(\mathbf{p}, \theta^{(i)}). \quad (5)$$

ここで， i は反復回数を表す．

3 シミュレーション実験

推定精度と計算コストの観点から提案法の性能を評価するため， $s = 1$ ， $\theta = 0$ の条件下でシミュレーション実験を実施した．



(a) Original

(b) Shifted

図 1: 位置ずれのある画像の組の例 (Pepper)

3.0.1 実験条件

本実験では, *airplane*, *barbara*, *cameraman*, *mandrill* に代表される 20 枚の標準画像¹を用い, 位置ずれのある画像の組を生成した. 生成した画像の組の例を図 1 に示す. 原画像の大きさは, 256×256 画素である. 位置ずれのある画像を生成するために, 原画像を縦横方向にシフトした. シフト量 $\Delta \mathbf{p}$ は, $[-10, 10] \times [-10, 10]$ 上の一様分布に従う乱数により定めた. 非整数のシフトを行う際の補間手法には, 双 3 次補間を用いた. 原画像とそれをシフトした画像の間に生じる, 他方の画素値が存在しない領域は切り取った. 最後に, 平均 0, 標準偏差 0.1 のガウス雑音を, 原画像とシフト画像のそれぞれに独立に加えた.

提案する補助関数に基づく手法 (Aux. と略記) を, 勾配上昇法 (gradient ascent, GA), 2 次元黄金分割探索 (golden section search, GSS) [3], および, 文献 [4] における 2 次曲面当てはめの 3 つの手法と比較した. GA は, 最大点 $\tilde{\mathbf{p}}$ を $\tilde{\mathbf{p}}^{(i+1)} = \tilde{\mathbf{p}}^{(i)} + \eta \nabla \check{\Phi}_2^{(xy)}(\tilde{\mathbf{p}}^{(i)})$ に従い反復的に更新する手法であり, ステップ幅 η は Armijo 条件 [5] に従うバックトラック直線探索によって決定した. GSS は, 単峰性を持つ目的関数を対象に, その勾配を用いずに最大点を探索する反復アルゴリズムである. $\check{\Phi}_2^{(xy)}$ は多峰性の関数であるが, 最大点の近傍では単峰性を持つと考えられるため, 良い初期推定値 $\mathbf{p}^{(0)}$ が与えられる場合には有効である. GA, GSS, および提案法においては, 目的関数 (式 (3) 参照) における重みを $w_{kl} = 1$ とし, 初期推定値 $\mathbf{p}^{(0)}$ は離散の相互相関 $\check{\Phi}_2^{(xy)}[\mathbf{p}]$ の最大点として, 50 回の反復更新を行った.

曲面当てはめは, 離散の目的関数に 2 次曲面を当てはめてその最大点を求めることで, もともとの目的関数の近似解を求める方法である. 文献 [4] における曲面当てはめによる位置ずれ推定では, 相互相関関数ではなく l_2 距離を目的関数としていたが, 本実験では他の手法に合わせ重みを $w_{kl} = 1$ とする離散の相互相関 $\check{\Phi}_2^{(xy)}[\mathbf{p}]$ を目的関数とした.

これら手法は, すべて Python および PyTorch を用いて実装した. 各手法により得られる推定値 $\tilde{\mathbf{p}}$ と真のシフト量 $\mathbf{p}^*$ 間の推定誤差は, 二乗平均平方根誤差 (root-mean-square error, RMSE) により評価した.

本実験は, 動作周波数 3.8 GHz の CPU および 64GB メモリを有するコンピュータ上で, GPU による演算なしに実施した. (表 1 参照). 実行時間の測定にあたっては, Python 標準ライブラリに含まれる `time.perf_counter()` 関数を用いた.

¹<https://sipi.usc.edu/database/> および http://www.ess.ic.kanagawa-it.ac.jp/app_images_j.html からダウンロードできる.

表 1: 実行時間計測に用いたコンピュータのスペック

Processor	AMD Ryzen 7 5800X (3.8 GHz, 8 cores)
Memory	64 GB
OS	Ubuntu 20.04 LTS
Language	Python 3.9.11
Library	PyTorch 2.0.1

表 2: 真のシフト量とその推定値の間の RMSE

Image	GA	GSS [3]	Curve fitting [4]	Aux. (ours)
Airplane	0.0158	0.0158	0.0489	0.0158
Barbara	0.0453	0.0453	0.1900	0.0453
Cameraman	0.0139	0.0140	0.0157	0.0140
Pepper	0.1489	0.1489	0.2241	0.1489
Woman	0.0448	0.0454	0.0820	0.0454
Average (20 images)	0.0679	0.0679	0.1202	0.0680
Std. (20 images)	0.0391	0.0391	0.0612	0.0391

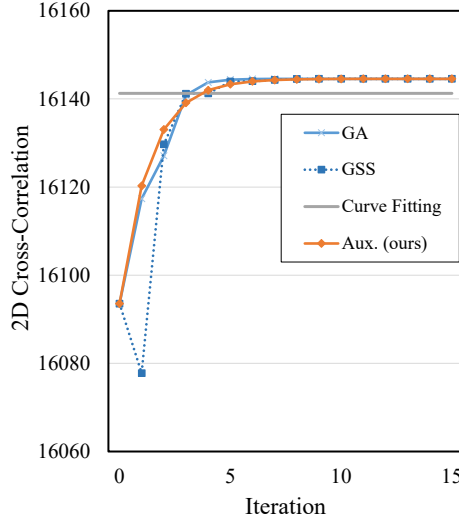


図 2: *Pepper* に対する更新毎の目的関数の推移. 曲面当てはめは反復を行わないため, 目的関数値を直線で示している.

3.0.2 実験結果

表 2 に, 評価データ内の 5 組のデータに対する真のシフト量とその推定値の RMSE と, その全評価データに対する平均および標準偏差を示す. この結果から, GA, GSS, および提案法は, 曲面当てはめと比較して高い精度でシフト量を推定可能であることがわかる.

それぞれの手法の更新ごとの目的関数値の推移を図 2 に示す. ここで, 入力画像の組は図 1 に示した *Pepper* である. 図より, GA, GSS の目的関数値は, 6 回の反復で局所最大値におおよそ収束しているのに対し, 提案法は 8 回の反復で局所最大値に収束している. 各手法の更新ごとに得られる点 $\mathbf{p}^{(i)}$ の軌跡を, 図 3 から 5 にそれぞれ示す. ここで, 2 次元相互相関関数 $\check{\Phi}_2^{(xy)}(\mathbf{p})$ の最大点と真のシフト量 $\Delta\mathbf{p}$ は, 画像の切り取り等の影響でわずかに異なることに注意されたい.

20 組のデータに対する 4 つの手法の平均実行時間を表 3 に示す. 反復法 (GA, GSS, および Aux.) に対しては, 50 回の反復にかかる経過時間を測定し, 各反復ごとの実行時間を計算した. 表より, 提案法は GA よりも約 6.8 倍, GSS よりも約 4 倍高速であり, 提案法が約 8 回の反復で収束する場合 (図 2 参照), 反復を行わない曲面当てはめと同等の実行時間であることがわかる.

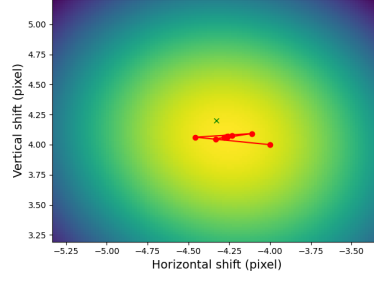


図 3: *Pepper* に対する推定値の軌跡 (GA). ここで、赤い点は推定値 $p^{(i)}$ を表し、十字は真のシフト量 p^* を表す.

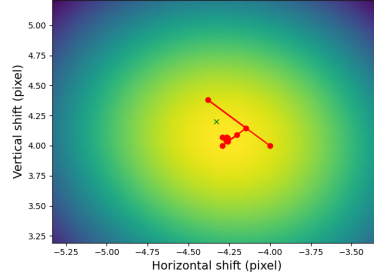


図 4: *Pepper* に対する推定値の軌跡 (GSS). ここで、赤い点は推定値 $p^{(i)}$ を表し、十字は真のシフト量 p^* を表す.

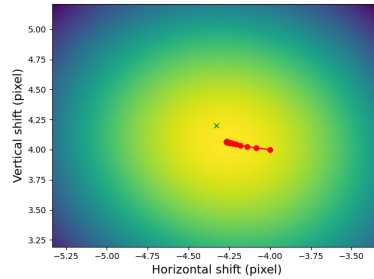


図 5: *Pepper* に対する推定値の軌跡 (Aux.). ここで、赤い点は推定値 $p^{(i)}$ を表し、十字は真のシフト量 p^* を表す.

表 3: 反復法における更新毎の実行時間 (ms)

Statistic	GA	GSS [3]	Curve fitting [4]	Aux. (ours)
Average	4.793	2.801	4.428	0.705
Std.	1.103	0.097	0.144	0.017

4 拡散モデルによる生成画像の主観品質の Blackbox 最適化

画像 x に対するユーザの嗜好を滑らかな関数 $f(x)$ として表現する. 拡散モデル $G(\cdot)$ を用いて x を生成する. 本手法の目的は, 生成された画像 $x = G(z)$ に対するユーザの主観的な品質を最大化することであり, これは以下のように定式化できる.

$$\hat{x} = \underset{x}{\operatorname{argmax}} f(x) = \underset{z}{\operatorname{argmax}} f(G(z)), \quad (6)$$

ここで, 摂動 z は標準正規分布に従う D 次元の乱数ベクトルである. この最適化問題は, 勾配法によって以下のように実行できる.

$$z_{n+1} = z_n + \eta \nabla_z (f \circ G)(z_n) = z_n + \eta \nabla_z h(z_n), \quad (7)$$

ここで, η は学習率, $h(z) = (f \circ G)(z)$, $n \in \{1, \dots, N\}$ である. しかし, $h(z)$ の勾配 $\nabla_z h(z)$ は, f が未知の人間の知覚を表しているため, 直接的には計算できない.



図 6: 初期潜在変数 $z_1 + \Delta z^{(r)}$ から生成された画像



図 7: 最終更新後潜在変数 $z_{21} + \Delta z^{(r)}$ から生成された画像

そこで、本研究では、検索勾配に基づいて $\nabla_z h(z)$ を近似する `fujii2020humangan`, `wierstra2014natural`. この近似は以下の式で与えられる.

$$\nabla_z h(z_n) \approx \frac{1}{2\sigma^2 R} \sum_{r=1}^R \Delta h(z_n, \Delta z^{(r)}) \cdot \Delta z^{(r)}, \quad (8)$$

ここで、 R は摂動の個数を表し、 $\Delta z^{(1)}, \dots, \Delta z^{(R)}$ は平均 0, 分散 $\sigma^2 I$ の正規分布に従う D 次元ベクトルである. 式 (8) の計算においては、 $h(z_n + \Delta z^{(r)})$ および $h(z_n - \Delta z^{(r)})$ の値そのものではなく、差分

$$\Delta h(z_n, \Delta z^{(r)}) = h(z_n + \Delta z^{(r)}) - h(z_n - \Delta z^{(r)})$$

を求める必要がある. したがって、ユーザは h の値を評価するのではなく、2つの画像を比較して評価を行う. すなわち、ユーザが $G(z_n + \Delta z^{(r)})$ の方が $G(z_n - \Delta z^{(r)})$ よりも優れていると判断した場合、 $\Delta h(z_n, \Delta z^{(r)}) = 1$ とし、逆の場合は $\Delta h(z_n, \Delta z^{(r)}) = -1$ とする.

4.1 実験

提案手法により拡散モデルを制御し、顔画像を生成できるかを確認するため、実験を行った. 本実験では、被験者が「男性顔」であるか否かをフィードバックし、その情報に基づき拡散モデルが男性顔画像を生成した. すなわち、被験者は、 $G(z_n + \Delta z^{(r)})$ が $G(z_n - \Delta z^{(r)})$ よりもより男性的である場合に $\Delta h(z_n, \Delta z^{(r)}) = 1$ と設定した.

提案手法におけるパラメータは、 $\eta = 1.0 \times 10^{-4}$, $\sigma^2 = 2.5 \times 10^{-7}$, $N = 20$, $R = 3$ とした.

図 6 および図 7 は、それぞれ初期潜在変数 z_1 および最終更新後の潜在変数 z_{21} から生成された画像の例を示す. これらの図を比較することで、図 7 の方が図 6 よりも男性画像の数が多いことが分かる. この結果より、ユーザのフィードバックを用いて拡散モデルを制御できることが確認された.

5 輝度領域分割に基づくトーンマッピング法

提案法の目的は、画像領域間の輝度の大小関係と局所的なコントラストを保持しながら、HDR 画像を LDR 画像へとトーンマッピングすることである. これを達成するために、提案法ではまず、従来の露出補正法 [6], [7] と同様にシーンを複数の小領域に分割する. 次に、表示輝度のダイナミックレンジ内で領域間のコントラストを最大化するように、多重露出画像の露出を決定する. HDR 画像を露出補正することにより生成した多重露出画像を合成することで、結果画像が得られる. 露出補正の効果を結果画像に十分反映させるために、合成重みを露出補正アルゴリズムで得られた露出値に基づき決定する.

5.1 露出補正

提案する露出補正法では、 ℓ_{gray} を基準とした相対露出 (単位: EV) を考える. 次式を用いてシーン輝度 ℓ_w を 0 EV に正規化する.

$$\ell_s(\mathbf{p}) = \frac{\ell_{\text{gray}}}{G(\ell_w|\mathbb{P})} \ell_w(\mathbf{p}) \quad (9)$$

$\mathbb{P}$ を M 個の領域 $\{\mathbb{P}_m\}$ に分割するために, 従来の手法 [6], [7] と同様, ℓ_s の対数分布 $p(\log \ell_s)$ を混合数 M のガウス混合モデル (Gaussian mixture model, GMM) を用いて次式の通りモデル化する.

$$p(\log \ell_s) = \sum_{m=1}^M \pi_m \mathcal{N}(\log \ell_s \mid \mu_m, \sigma_m) \quad (10)$$

ここで, π_m は混合係数である. $\mathcal{N}(\cdot \mid \mu_m, \sigma_m)$ は平均値 μ_m , 標準偏差 σ_m の正規分布であり, $\mu_m < \mu_{m+1}$ を満たすものとする. 画素 $\mathbf{p} \in \mathbb{P}$ の属する領域 $\mathbb{P}_m$ は, GMM を構成する各混合成分の負担率に基づき,

$$m = \arg \max_k \pi_k \mathcal{N}(\log \ell_s(\mathbf{p}) \mid \mu_k, \sigma_k) \quad (11)$$

として与えられる. $\mathbb{P}_1$ と平均輝度の観点から最も暗い領域を表し, $\mathbb{P}_M$ は最も明るい領域を示す.

提案法における露出補正は, 各領域の平均輝度に基づき行われる. 領域 $\mathbb{P}_m$ に対する露出補正係数 $\{\Delta t_m\}$ は, その領域に対する目標平均 μ'_m を定めたのち, 目標平均と元の平均 μ_m を用いて

$$\Delta t_m = \frac{\exp(\mu'_m)}{\exp(\mu_m)} \quad (12)$$

として与えられる. 目標平均 $\{\mu'_m\}$ の定めるアルゴリズムを次に述べる.

まず, ℓ_s の全体的な明さを合成後にも保つために, 領域 $\{\mathbb{P}_m\}$ のうち基準となる領域 $\mathbb{P}_{m_{\text{ref}}}$ を, ℓ_{gray} に最も大きく寄与する混合成分として選択する. この選択は, 次式で表される.

$$m_{\text{ref}} = \arg \max_k \pi_k \mathcal{N}(\log \ell_{\text{gray}} \mid \mu_k, \sigma_k) \quad (13)$$

基準領域 $\mathbb{P}_{m_{\text{ref}}}$ に対する目標平均 $\mu_{m_{\text{ref}}}$ には, 元々の平均 $\mu_{m_{\text{ref}}}$ をそのまま用いる.

次に, 表示輝度のダイナミックレンジを考慮して定める最小露出 v_{min} EV と最大露出 v_{max} EV に基づき, 最も暗い領域 $\mathbb{P}_1$ と最も明るい領域 $\mathbb{P}_M$ の目標平均を

$$\mu'_1 = \log 2^{v_{\text{min}}} \ell_{\text{gray}}, \quad \mu'_M = \log 2^{v_{\text{max}}} \ell_{\text{gray}} \quad (14)$$

として与える.

残る領域の目標平均は, 領域間の平均輝度の大小関係を保ちつつ, 領域間コントラストの和 $\sum_{m=1}^{M-1} \mu'_{m+1} - \mu'_m$ を最大化するように定める. これは, 以下の通り目標平均 $\{\mu'_m\}$ を等間隔に整列させることで達成される.

$$\mu'_m = \begin{cases} \frac{\mu_{m_{\text{ref}}} - \mu'_1}{m_{\text{ref}} - 1} (m - 1) + \mu'_1 & 1 < m < m_{\text{ref}} \\ \mu_{m_{\text{ref}}} & m = m_{\text{ref}} \\ \frac{\mu'_M - \mu_{m_{\text{ref}}}}{M - m_{\text{ref}}} (m - m_{\text{ref}}) + \mu_{m_{\text{ref}}} & m_{\text{ref}} < m < M \end{cases} \quad (15)$$

設定した $\{\Delta t_m\}$ を用い, 露出補正とトーンマッピングを ℓ_s に適用して, m 番目の露出画像 $\mathbf{x}_m$ の輝度を

$$\ell_m(\mathbf{p}) = f(\ell_s(\mathbf{p}) \Delta t_m) \quad (16)$$

として得る. ここで, トーンカーブ f として, Reinhard の大域 TMO を $\ell_{\text{white}} = 2^{v_{\text{white}}} \ell_{\text{gray}}$ と共に利用する.

5.2 多重露出画像合成のための重み

本節では、Mertens ら合成法 [8] を合成関数 $\mathcal{F}$ として用いることを想定し、露出補正の効果を合成画像に反映させるための新しい合成重みを提案する。 m 番目の露出画像 $\mathbf{x}_m$ に対する重み w_m を計算するため、輝度 ℓ_m と目標平均 μ'_m 間の表示輝度領域における差 d_m を次の通り求める。

$$d_m(\mathbf{p}) = g(\ell_m(\mathbf{p})) - g(\exp(\mu'_m)) \quad (17)$$

ここで、 $g(\ell) = (\gamma \circ f)(\ell)$ は、ガンマ補正関数 γ とトーンカーブ f の合成関数であり、式 (20) における合成が、ガンマ補正適用後の sRGB 領域で行われることを考慮している。

差 $d_m(\mathbf{p})$ を用いて、重み w_m を次のように計算する。

$$w_m(\mathbf{p}) = \frac{\exp(-(d_m(\mathbf{p}))^2)}{\sum_{k=1}^M \exp(-(d_k(\mathbf{p}))^2)} \quad (18)$$

式 (18) は、画素 $\mathbf{p}$ における多重露出画像の輝度 $\{\ell_m(\mathbf{p})\}$ のうち、 $\mathbf{p}$ が属する領域 $\mathbb{P}_m$ の目標平均 μ'_m と近い輝度に大きな重みをつけることを意味する。

5.3 トーンマッピングの手順

提案する露出補正と合成重みを用いて入力 HDR 画像 $\mathbf{E}$ から $\mathbf{y}$ へトーンマッピングを行う手順は、以下のようになどめられる。

I. 露出補正:

- (i) 入力 HDR 画像 $\mathbf{E}$ のシーン輝度 ℓ_w を計算する。
- (ii) 式 (9) により、 ℓ_w を 0 EV に正規化する。
- (iii) 式 (10) と (11) により $\mathbb{P}$ を M 個の領域 $\{\mathbb{P}_m\}$ に分割する。
- (iv) 式 (13) から (15) を用いて、領域 $\mathbb{P}_m$ に対する目標平均 μ'_m を決定する。
- (v) 式 (12) を用いて、領域 $\mathbb{P}_m$ の露出補正係数 Δt_m を計算する。
- (vi) 式 (16) に従い、入力画像 $\mathbf{E}$ をトーンマッピングする。
- (vii) sRGB 色空間における多重露出画像 $\{\mathbf{x}_m\}$ を

$$\mathbf{x}_m = \gamma \left(\frac{\ell_m(\mathbf{p})}{\ell_w(\mathbf{p})} \mathbf{E}(\mathbf{p}) \right) \quad (19)$$

として生成する。ここで、 γ は、RGB 値それぞれにガンマ補正を適用する関数である。

II. 多重露出合成:

- (i) 式 (17) と (18) を用いて、 $\{\ell_m\}$ と $\{\mu'_m\}$ から合成重み $\{w_m\}$ を計算する。
- (ii) 次式を用いて、多重露出画像 $\{\mathbf{x}_m\}$ のラプラシアンピラミッドを合成する。

$$\mathbf{L}_l[\mathbf{y}](\mathbf{p}) = \sum_{m=1}^M G_l[w_m](\mathbf{p}) \mathbf{L}_l[\mathbf{x}_m](\mathbf{p}) \quad (20)$$

ここで、 $\mathbf{L}_l[\mathbf{x}_m](\mathbf{p})$ は画素 $\mathbf{p}$ における m 番目の画像 $\mathbf{x}_m$ のラプラシアンピラミッドの l 番目のレベルであり、 $G_l[w_m](\mathbf{p})$ は画素 $\mathbf{p}$ における m 番目の露出画像の合成重み w_m のガウシアンピラミッドの l 番目のレベルである。

- (iii) ラプラシアンピラミッド $\mathbf{L}_l[\mathbf{y}]$ を画像領域に変換し、トーンマッピングされた画像 $\mathbf{y}$ を得る。

5.4 シミュレーション実験

提案する露出補正と合成重みを用いたトーンマッピング法を、トーンマッピングされた画像 $\mathbf{y}$ の品質の観点から評価した。以下、このトーンマッピング法を提案法と呼ぶ。

表 4: 4 つの代表的なシーンに対する TMQI スコアと, 20 個すべてのシーンに対する平均スコアおよび標準偏差

Scene	Global [10]	Local [10]	Conv. [6], [7]	Prop.
507	0.8737	0.8185	0.8467	0.9298
LabTypewriter	0.8323	0.7881	0.7869	0.8836
LasVegasStore	0.9204	0.8707	0.8300	0.9494
WillyDesk	0.8564	0.8077	0.8677	0.9326
Average	0.8923	0.8388	0.8628	0.9253
Std.	0.0398	0.0386	0.0464	0.0395

5.4.1 実験条件

本実験では, 文献 [9] から選択した 20 個の HDR 画像を入力画像 $\mathbf{x}$ として使用した. 画像 $\mathbf{x}$ を提案法によってトーンマッピングし, 得られる画像 $\mathbf{y}$ の品質を評価した.

本稿では, Reinhard の大域 TMO (Global) と局所 TMO (Local) [10], 従来の露出補正を用いた MEF (Conventional) [6], [7], および提案法 (Proposed) の 4 つの手法を比較した. MEF に基づく Conventional と Proposed では, 合成関数 $\mathcal{F}$ として Mertens の手法 [8] を使用した. また, 提案法のパラメータは $v_{\min} = -3$, $v_{\max} = 1.5$, $v_{\text{white}} = 2.5$ と設定した.

トーンマッピングされた画像の品質を評価するためには, 客観的な品質評価が必要となる. ピーク信号対雑音比 (peak signal-to-noise ratio, PSNR) や構造的類似度 (structural similarity index measure, SSIM) などの典型的な画質評価尺度には, トーンマッピングされた画像のうち理想的なものを参照画像として用いる必要があり, 本実験には適していない. そこで, 本実験では, トーンマッピングされた画像の品質評価のために設計された tone mapped image quality index (TMQI) [11] を使用した.

5.4.2 結果

入力画像 “507” および “LasVegasStore” からトーンマッピングされた画像を, それぞれ図 8 と 9 に示す. これらの図から, Conventional では全体的に灰色がかった低コントラストの画像が得られていることがわかる. 一方, Proposed は, コントラストを保持しながら HDR 画像を効果的にトーンマッピングできている.

表 4 は, 実験に用いた 20 種類の画像のうち, 4 つの代表的なシーンに対する TMQI スコアと, 20 個すべてのシーンに対する平均スコアおよび標準偏差を示している. 表より, Proposed が, 平均的に最も高い TMQI スコアを示したことが確認できる. したがって, Proposed によってトーンマッピングされた画像は, 他の 3 つの手法により得られた画像よりも高品質であることいえる.

6 おわりに

本研究では, 現実シーンの広い輝度情報を記録した高ダイナミックレンジ (High dynamic range, HDR) 画像を深層生成モデルに学習させることで, さらなる性能向上を達成することを目的に, (1) 大量の HDR 画像を収集するための基盤となる画像の高精度位置合わせアルゴリズム, (2) 深層生成モデルによる生成画像の主観品質の Blackbox 最適化アルゴリズム, および, (3) HDR 画像の可視化のための輝度領域分割に基づくトーンマッピング法を開発した. 本研究の成果は, 以下の通り発表した.

- Reiki Ota and Yuma Kinoshita, “Controlling Diffusion-based Image Generation Using Search Gradients with User Feedback,” RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, pp. 221–224, Pulau Pinang, Malaysia, Feb. 2025.
- 木下 裕磨, 貴家 仁志, “多重露出画像合成によるトーンマッピングのための輝度領域分割に基づく露出補正法,” 電子情報通信学会 信号処理シンポジウム, 2024 年 12 月.
- Yuma Kinoshita and Hitoshi Kiya, “Scene-Segmentation-Based Exposure Compensation for Tone Mapping of High Dynamic Range Scenes,” APSIPA Annual Summit and Conference, Macau, China, Dec. 2024.



(a) Global [10]



(b) Local [10]



(c) Conventional [6], [7]



(d) Proposed

図 8: 結果画像 (507)

最後に、貴財団からの支援に対し、心からの感謝の意を表する。

参考文献

- [1] C. D. Kuglin, “The Phase Correlation Image Alignment Method,” in *Proc. Int. Conf. Cybern. Soc.*, 1975, pp. 163–165.
- [2] C. Knapp and G. Carter, “The Generalized Correlation Method for Estimation of Time Delay,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.



(a) Global [10]



(b) Local [10]



(c) Conventional [6], [7]



(d) Proposed

図 9: 結果画像 (LasVegasStore)

- [3] C. He *et al.*, “Object Tracking Using the Gabor Wavelet Transform and the Golden Section Algorithm,” *IEEE Trans. Multimed.*, vol. 4, no. 4, pp. 528–538, 2002.
- [4] S. W. Hasinoff *et al.*, “Burst Photography for High Dynamic Range and Low-Light Imaging on Mobile Cameras,” *ACM Trans. Graph.*, vol. 35, no. 6, pp. 1–12, 2016.
- [5] L. Armijo, “Minimization of Functions Having Lipschitz Continuous First Partial Derivatives,” *Pacific J. Math.*, vol. 16, no. 1, pp. 1–3, 1966.
- [6] Y. Kinoshita and H. Kiya, “Automatic exposure compensation using an image segmentation method for single-image-based multi-exposure fusion,” *APSIPA Trans. Signal Inf. Process.*, vol. 7, no. e22, 2018.

- [7] Y. Kinoshita and H. Kiya, “Scene segmentation-based luminance adjustment for multi-exposure image fusion,” *IEEE Trans. on Image Process.*, vol. 28, no. 8, pp. 4101–4116, 2019.
- [8] T. Mertens *et al.*, “Exposure fusion: A simple and practical alternative to high dynamic range photography,” *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.
- [9] P. Korshunov *et al.*, “Subjective quality assessment database of HDR images compressed with JPEG XT,” in *Proc. QoMEX*, 2015, pp. 1–6.
- [10] E. Reinhard *et al.*, “Photographic tone reproduction for digital images,” *ACM TOG*, vol. 21, no. 3, pp. 267–276, 2002.
- [11] H. Yeganeh and Z. Wang, “Objective quality assessment of tone-mapped images,” *IEEE Trans. on Image Process.*, vol. 22, no. 2, pp. 657–667, 2013.