

リアルタイム放射輝度フィールドレンダリングのための 3D ガウス スプレーイング

ベルンハルト・ケルブル、インリア、コートダジュール大学、フランス
ゲオルギオス・コパナス、インリア、コートダジュール大学、フランス
THOMAS LEIMKÜHLER、マックス・プランク情報研究所、ドイツ
ジョージ・ドレッタキス、インリア、コートダジュール大学、フランス



図1. 我々の手法は、従来の手法と同等の品質で輝度場をリアルタイムにレンダリングすることを実現した[Barron et al. 2022]。従来の最速の手法と競合する最適化時間しか必要としない[Fridovich-Keil and Yu et al. 2022; Müller et al. 2022]。この鍵となるのはパフォーマンスは、リアルタイムの微分化可能なレンダラーと組み合わせた新しい3Dガウスシーン表現であり、シーンと画像の両方に大幅な高速化をもたらします。最適化と新規ビュー合成。InstantNGP [Müller et al. 2022]と同等の学習時間で、同等の品質を達成している点に注目してください。これは彼らが到達した最高品質であり、51分間のトレーニングによって、Mip-NeRF360 [Barron et al. 2022]よりもわずかに優れた最先端の品質を達成しました。

放射場法は最近、斬新な視点の合成に革命をもたらした。複数の写真や動画で撮影されたシーン。しかし、高い視覚的な品質には、依然として学習とレンダリングにコストのかかるニューラルネットワークが必要であり、最近の高速化手法では必然的に速度と品質がトレードオフされている。境界のない完全なシーン（孤立したオブジェクトではなく）と1080p解像度レンダリングでは、現在の方法ではリアルタイムの表示レートを実現できません。最先端の技術を実現するための3つの重要な要素をご紹介します。競争力のあるトレーニング時間を維持しながら視覚的な品質を維持し、そして重要なのは1080p解像度で高品質なリアルタイム（30fps）の新規ビュー合成を可能にします。まず、カメラキャリブレーション中に生成されたスパースポイントから開始し、シーンを最適化するために連続した体積放射輝度フィールドの望ましい特性を維持する3Dガウス分布でシーンを表現しながら、空きスペースでの不要な計算を避ける。次に、3Dガウス分布のインターリーブ最適化/密度制御、特に異方性共分散の最適化により、シーン; 3番目に、可視性を考慮した高速レンダリングアルゴリズムを開発し、異方性スプラッシングをサポートし、トレーニングの高速化とリアルタイムレンダリングを可能にします。最先端のビジュアル品質とリアルタイムレンダリングを実証します。いくつかの確立されたデータセットでのレンダリング。

CCSコンセプト: コンピューティング手法; レンダリング; ポイントベースモデル; ラスタライズ; 機械学習アプローチ。

両著者ともこの論文に同等の貢献をした。

著者のアドレス: Bernhard Kerbl, bernhard.kerbl@inria.fr, Inria, Université Côte d'Azur, フランス。 Georgios Kopanas, georgios.kopanas@inria.fr, インリア, コートダジュール, フランス。 Thomas Leimkühler, thomas.leimkuehler@mpi-inf.mpg.de, Max-Planck-Institut für Informatik, ドイツ。ジョージ・ドレッタキス, george.drettakis@inria.fr, インリア, コートダジュール大学, フランス。

出版権はACMにライセンスされています。ACMはこの寄稿が国の職員、請負業者、または関係者によって著作権または共同著作権者が作成されたものではありません。したがって、政府は、非独占的かつ無償で出版または公開する権利を有します。この記事複製すること、または他者に複製を許可することは、政府の目的に限ります。© 2023 著作権は所有者/著者が保有しています。出版権はACMにライセンス供与されています。0730-0301/2023/8-ART1 15.00ドル https://doi.org/10.1145/3592433

追加のキーワードとフレーズ: 新しいビュー合成、放射フィールド、3Dガウス分布、リアルタイムレンダリング

ACMリファレンスフォーマット:

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis. 2023年。「リアルタイム放射輝度場レンダリングのための3Dガウススプラッシング」ACM Trans. Graph. 42, 4, Article 1 (2023年8月), 14ページ, https://doi.org/10.1145/3592433

1 はじめに

メッシュとポイントは最も一般的な3Dシーン表現である。これらは明示的であり、高速なGPU/CUDAベースのラスタライズ。対照的に、最近のニューラルラディアンсフィールド（NeRF）法は連続的なシーン表現に基づいており、典型的にはボリュームレイマーチングを用いた多層パーセプトロン（MLP）撮影したシーンを新しい視点で合成する。同様に、最も効率的なのはこれまでのRadiance Fieldソリューションは連続表現に基づいて構築されています。例えばボクセルに格納された値を補間することによって[Fridovich-KeilとYu et al. 2022]またはハッシュ[Müller et al. 2022]グリッドまたはポイント[Xu et al. 2022]。これらの方法の連続性は最適化に役立ちますが、レンダリングに必要な確率的サンプリングはコストがかかり、ノイズの原因となります。私たちは、最高のものを組み合わせた新しいアプローチを導入します。両方の世界の3Dガウス表現により最適化が可能。最先端の（SOTA）映像品質と競争力のあるトレーニング。タイトルベースのスプラッシングソリューションは、これまで複数のデバイスで1080p解像度のSOTA品質でのリアルタイムレンダリングを保証します。公開されたデータセット [Barron et al. 2022; Hedman et al. 2018; Knapitsch et al. 2017] (図1参照)。

私たちの目標は、複数の写真から最適化された表現を作成する。典型的な従来の最も効率的な方法の何倍も速い。実際のシーン。最近の手法では高速なトレーニングが実現されている[Fridovich-Keil

1:2・ペルンハルト・ケルブル、ゲオルギオス・コパナス、トーマス・ライムクーラー、ジョージ・ドレアキス

Yu et al. 2022; Müller et al. 2022]は、現在のSOTA NeRF手法、すなわちMip-NeRF360 [Barron et al. 2022]で得られる画質を達成するのに苦労しており、この手法では最大48時間の学習時間が必要です。高速だが品質の低い放射輝度場法は、シーンに応じてインタラクティブなレンダリング時間（10〜15フレーム/秒）を実現できますが、高解像度でのリアルタイムレンダリングには至りません。

私たちのソリューションは3つの主要コンポーネントに基づいています。まず、柔軟で表現力豊かなシーン表現として3Dガウス分布を導入します。従来のNeRF類似手法と同じ入力、すなわちStructure-from-Motion (SfM) [Snavely et al. 2006]でキャリブレーションされたカメラから開始し、SfMプロセスの一環として無料で生成されるスパースポイントクラウドを使用して3Dガウス分布セットを初期化します。Multi-View Stereo (MVS)データを必要とする多くのポイントベースソリューション[Aliev et al. 2020; Kopanas et al. 2021; Rückert et al. 2022]とは対照的に、SfMポイントのみを入力として高品質な結果を実現します。NeRF合成データセットでは、ランダム初期化でも高品質な結果が得られる点に留意してください。3Dガウス分布は微分可能なボリューム表現であるため優れた選択肢であることを示していますが、2Dに投影し、NeRFと同等の画像形成モデルを使用して標準のUブレンディングを適用することで、非常に効率的にラスタライズすることもできます。私たちの方法の2番目の要素は、3Dガウス分布のプロパティ（3D位置、不透明度 α 、異方性共分散、球面調和関数（SH）係数）の最適化であり、適応型密度制御ステップとインターリーブされ、最適化中に3Dガウス分布を追加したり、場合によっては削除したりします。最適化手順により、シーンの適度にコンパクトで非構造化かつ正確な表現が生成されます（テストされたすべてのシーンで100万〜500万のガウス分布）。私たちの方法の3番目で最後の要素は、高速GPUソートアルゴリズムを使用し、最近の研究[Lassner and Zollhofer 2021]に従ったタイルベースのラスタライズに触発されたリアルタイムレンダリングソリューションです。しかし、3Dガウス表現のおかげで、ソートとUブレンディングによって可視性の順序を尊重する異方性スブラッティングを実行でき、必要な数のソート済みスブラッティングのトラバースを追跡することで高速かつ正確な後方パスが可能になります。

要約すると、私たちは次のような貢献をします。

- 高品質で非構造化の放射フィールド表現として、異方性3Dガウス分布を導入しました。
- キャプチャされたシーンの高品質表現を作成する適応型密度制御とインターリーブされた3Dガウス分布特性の最適化手法。
- 可視性を考慮したGPU向けの高速で微分可能なレンダリング手法により、異方性スブラッティングと高速なバックプロパゲーションが可能になり、高品質で斬新なビュー合成を実現します。

以前に公開されたデータセットを用いた結果、マルチビューキャプチャから3Dガウス分布を最適化し、従来の最高品質の暗黙的放射輝度場アプローチと同等以上の品質を実現できることがわかりました。また、最速の手法と同等の学習速度と品質を実現し、新規ビュー合成のための高品質なリアルタイムレンダリングを初めて実現しました。

2 関連研究

まず、従来の再構成の概要を簡単に説明し、次にポイントベースのレンダリングと輝度フィールドの作業について議論し、それらの

類似性：放射輝度分野は広大なため、本稿では直接関連する研究のみに焦点を当てます。この分野の完全な網羅性については、最近の優れたサーベイ [Tewari et al. 2022; Xie et al. 2022]をご覧ください。

2.1 従来のシーン再構成とレンダリング最初の斬新なビュー合成アプローチ

これは、最初は高密度にサンプリングされた光線場に基づいており [Gortler et al. 1996; Levoy and Hanrahan 1996]、次に非構造化キャプチャを可能にしました [Buehler et al. 2001]。Structure-from-Motion (SfM) [Snavely et al. 2006] の出現により、写真のコレクションを使用して斬新なビューを合成できるまったく新しい領域が可能になりました。SfMは、カメラのキャリブレーション中にスパースポイントクラウドを推定するもので、当初は3D空間の単純な視覚化に使用されていました。その後のマルチビューステレオ (MVS) により、何年にもわたって優れた完全な3D再構成アルゴリズムが生成され [Goesele et al. 2007]、いくつかのビュー合成アルゴリズムの開発が可能になりました [Chaurasia et al. 2013; Eisemann et al. 2013]。これらの手法はすべて、入力画像を新しいビューカメラに再投影・合成し、ジオメトリを用いて再投影をガイドする。これらの手法は多くの場合優れた結果を生み出すが、MVSが存在しないジオメトリを生成する場合、再構成されていない領域や「過剰再構成」を完全に復元できないのが一般的である。最近のニューラルレンダリングアルゴリズム [Tewari et al. 2022]は、このようなアーティファクトを大幅に削減し、すべての入力画像をGPUに保存するという膨大なコストを回避し、ほとんどの面でこれらの手法を凌駕している。

2.2 ニューラルレンダリングと放射輝度場ディープラーニング技術は、初期には新規ビュー合成に採用されていました [Flynn et al. 2016; Zhou et al. 2016]。CNNはブレンディングウェイトの推定 [Hedman et al. 2018]やテキストチャ空間解法 [Riegler and Koltun 2020; Thies et al. 2019]に使用されました。これらの手法の多くでは、MVSベースのジオメトリの使用が大きな欠点となっています。さらに、CNNを最終レンダリングに使用すると、時間的なならつきが頻繁に発生します。

新規ビュー合成のためのボリューム表現は Soft3D [Penner and Zhang 2017] によって開始され、その後、連続的に微分可能な密度場に基づいて形状を表現するボリュームレイマーチングと組み合わせたディープラーニング手法が提案されました [Henzler et al. 2019; Sitzmann et al. 2019]。ボリュームレイマーチングを使用したレンダリングは、ボリュームを照会するために必要なサンプル数が多いため、コストが非常に高くなります。Neural Radiance Fields (NeRF) [Mildenhall et al. 2020] は、品質を向上させるために重要度サンプリングと位置エンコードを導入しましたが、速度に悪影響を与える大規模な多層パーセプトロンを使用しました。NeRFの成功により、品質と速度に対処する後続の方法が爆発的に増加し、多くの場合、正則化戦略が導入されています。新規ビュー合成における画質の現状最先端技術はMip-NeRF360 [Barron et al. 2022]です。レンダリング品質は極めて優れていますが、学習とレンダリング時間は依然として非常に長くなっています。私たちは、高速な学習とリアルタイムレンダリングを提供しながら、この品質と同等、あるいは場合によってはそれを上回る品質を実現しています。

最近の手法は、主に3つの設計上の選択を活用して、より高速なトレーニングやレンダリングに重点を置いています。空間データ構造を使用して（ニューラル）特徴を保存し、その後ボリュームレイマーチング中に補間する、異なるエンコード、

そしてMLPの容量。こうした手法には、空間離散化の様々なバリエーション[Chen et al. 2022b,a; Fridovich-Keil and Yu et al. 2022; Garbin et al. 2021 ; Hedman et al. 2021; Reiser et al. 2021; Takikawa et al. 2021; Wu et al. 2022; Yu et al. 2021],コードブック[Takikawa et al. 2022],ハッシュテーブルなどのエンコーディング[Müller et al. 2022]などがあり、より小規模なMLPの使用やニューラルネットワークの完全な使用を可能としている[Fridovich-Keil and Yu et al. 2022; Sun et al. 2022]。

これらの方法の中で最も注目すべきは、ハッシュ グリッドと占有グリッドを使用して計算を加速し、より小さな MLP を使用して密度と外観を表現するInstantNGP [Müller et al. 2022],およびスパース ボクセルグリッドを使用して連続密度フィールドを補間し、ニューラル ネットワークを完全に省略できるPlenoxels [Fridovich-Keil and Yu et al. 2022] です。どちらも球面調和関数に依存しています。前者は方向効果を直接表現し、後者は色ネットワークへの入力をエンコードします。どちらも優れた結果を提供しますが、これらの方法は、シーン/キャプチャ タイプによっては、空のスペースを効果的に表現するのが困難場合があります。さらに、画像品質は、加速に使用される構造化グリッドの選択によって大きく制限され、特定のレイマーチング ステップで多くのサンプルを照会する必要があるため、レンダリング速度が低下します。私たちが使用する非構造化で明示的なGPU 対応 3D ガウス分布は、ニューラル コンポーネントなしで、より高速なレンダリング速度と優れた品質を実現します。

2.3 ポイントベースレンダリングと放射輝度フィールドポイントベース

の方法は、切断された構造化されていないジオメトリサンプル (つまり、ポイントクラウド)を効率的にレンダリングします[Gross and Pster 2011]。

最も単純な形式のポイントサンプルレンダリング [Grossman and Dally 1998] は、固定サイズ of 非構造化ポイントセットをラスタライズします。このラスタライズには、グラフィックス API のネイティブでサポートされているポイントタイプ[Sainz and Pajarola 2004] やGPU 上の並列ソフトウェアラスタライズ [Laine and Karras 2011; Schütz et al. 2022] を利用できます。ポイントサンプルレンダリングは、基礎となるデータに忠実ですが、穴ができた、エイリアシングが発生した、完全に不連続なレンダリングになります。高品質のポイントベースレンダリングに関する先駆的な研究では、円または楕円のディスク、楕円体、サーフェルなど、ピクセルよりも大きな範囲のポイントプリミティブを「スプラッティング」することでこれらの問題に対処しています [Botsch et al. 2005; Pster et al. 2000; Ren et al. 2002; Zwicker et al. 2001b]。

近年、微分可能なポイントベースレンダリング技術に注目が集まっています[Wiles et al. 2020; Yifan et al. 2019]。ポイントはニューラル特徴量で拡張され、CNNを用いてレンダリングされます[Aliev et al. 2020; Rückert et al. 2022]。これにより、高速、あるいはリアルタイムのビュー合成が可能になります。しかし、これらの手法は初期ジオメトリをMVSに依存しているため、そのアーティファクト、特に特徴のない／光沢のある領域や薄い構造といった難しいケースにおいて過剰または不足した再構成が顕著になります。

ポイントベースのUブレンディングとNeRFスタイルのボリュームレンダリングは、本質的に同じ画像形成モデルを共有しています。具体的には、色は光線に沿ったボリュームレンダリングによって生成されます。

ここで、密度f、透過率)のサンプルは、間隔X8で光線に沿って分布する。これは次のように書き直すことができる。

= \frac{1}{8} \int_0^8 c(t) dt \tag{2}

と

U8 = (1 - \exp(-f8X8)) \int_0^8 c(t) dt \tag{3}

典型的なニューラルポイントベースのアプローチ (例えば[Kopanas et al. 2022, 2021])では、ピクセルに重なり合うN個の順序付けられたポイントをブレンドすることによってピクセルの色を計算します。

= \frac{1}{N} \sum_{i=1}^N c_i \tag{4}

ここでc8は各点の色であり、U8は共分散 [Yifan et al. 2019] を持つ2Dガウス分布に学習された点ごとの不透明度を乗じて評価することによって与えられます。

式2と式3から、画像形成モデルは同じであることが明確に分かります。しかし、レンダリングアルゴリズムは大きく異なります。NeRFは、空/占有空間を暗黙的に表す連続表現です。式2のサンプルを見つけるには、ノイズと計算コストを伴う高コストのランダムサンプリングが必要です。対照的に、点は構造化されていない離散表現であり、NeRFと同様にジオメトリの作成、破壊、および移動を可能にするのに十分な柔軟性があります。これは、以前の研究[Kopanas et al. 2021]で示されているように、不透明度と位置を最適化することで実現され、完全なボリューム表現の欠点を回避します。

Pulsar [Lassner and Zollhofer 2021] は、高速な球面ラスタライズを実現しており、これが私たちのタイルベースのソートレンダラーの着想の元となっています。しかし、上記の分析を踏まえ、ボリューム表現の利点を生かすために、ソートされたスプラット上では (近似的に)従来のUブレンディングを維持したいと考えています。私たちのラスタライズは、彼らの順序に依存しない方法とは対照的に、可視順序を尊重します。さらに、ピクセル内のすべてのスプラットの勾配を逆伝播し、異方性スプラットをラスタライズします。これらの要素はすべて、結果の高画質化に貢献しています (セクション7.3を参照)。さらに、上記の以前の方法では、レンダリングにCNNも使用しているため、時間的な不安定性が生じます。それでも、Pulsar [Lassner and Zollhofer 2021] とADOP [Rückert et al. 2022]のレンダリング速度は、高速レンダリングソリューションを開発する動機となりました。

鏡面反射効果に焦点を当てたニューラルポイントカタカウスティクス[Kopanas et al. 2022]の点ベースのレンダリングトラックは、MLPを用いることでこの時間的な不安定性を克服していますが、入力としてMVSジオメトリを必要とします。このカテゴリにおける最新の手法[Zhang et al. 2022]はMVSを必要とせず、方向付けにSHを使用します。しかし、1つのオブジェクトのシーンしか扱えず、初期化にマスクが必要です。小さな解像度では高速ですが、

点数が少ないため、典型的なデータセットのシーンにどのように拡張できるかは不明です[Barron et al. 2022; Hedman et al. 2018; Knapitsch et al. 2017]。より柔軟なシーン表現のために3Dガウス分布を使用することで、MVSジオメトリの必要性を回避し、投影されたガウス分布のためのタイルベースのレンダリングアルゴリズムによりリアルタイムレンダリングを実現しています。

= \frac{1}{8} \int_0^8 c(t) dt \tag{5}

1:4 • ベルンハルト・ケルブル、ゲオルギオス・コパナス、トーマス・ライムクラー、ジョージ・ドレアキス

最近のアプローチ[Xu et al. 2022]では、ラジアル基底関数アプローチを用いて点群を用いて輝度場を表現しています。最適化においては点群の刈り込みと密度化の手法が用いられていますが、ボリウムレイマーチングを用いており、リアルタイム表示は実現できません。

人間のパフォーマンスキャプチャの分野では、3Dガウス分布がキャプチャされた人間の体を表現するために用いられてきました[Rhodin et al. 2015; Stoll et al. 2011]。さらに最近では、視覚タスクにおけるボリウムレイマーチングと組み合わせて使用されています[Wang et al. 2023]。ニューラルボリウムプリミティブも同様の文脈で提案されています[Lombardi et al. 2024]。

これらの手法は、シーン表現として3Dガウス分布を選択するきっかけとなりましたが、これらは単一の孤立した物体（人体や顔）の再構成とレンダリングという特定のケースに焦点を当てており、結果として深度複雑度が低いシーンが生成されます。これに対し、我々は異方性共分散の最適化、インターリーブ最適化/密度制御、そしてレンダリングのための効率的な深度ソートを採用することで、背景を含む屋内外の複雑なシーン全体を、そして深度複雑度が高いシーンも扱うことができます。

3 概要

我々の手法への入力は、静的シーンの画像セットと、副次効果としてスパースな点群を生成する SfM [Schönberger and Frahm 2016] によって校正された対応するカメラである。これらの点から、位置（平均）、共分散行列、不透明度 U で定義される 3D ガウス分布（セクション 4）のセットを作成し、非常に柔軟な最適化体制を可能にする。これにより、3D シーンの適度にコンパクトな表現が得られる。これは、高度に異方性のボリウム スプラットを使用して新しい構造をコンパクトに表現できるためである。放射場の方向性外観成分（色）は、標準的な方法 [Fridovich-Keil and Yu et al. 2022; Müller et al. 2023]に従って、球面調和関数（SH）を介して表現される。

2022]。我々のアルゴリズムは、3Dガウスパラメータ（位置、共分散、 U およびSH係数）の最適化手順のシーケンスと、ガウス密度の適応制御のための操作を交互に実行して、放射場表現（第5節）を作成します。

本手法の効率性の鍵は、タイルベースのラスターライザ（6節）にあります。このラスターライザは、高速ソートにより可視順序を守りながら、異方性スプラットのUブレンディングを可能にします。この高速ラスターライザには、勾配を入力できるガウス分布の数に制限なく、累積 U 値を追跡する高速逆方向パスも含まれています。本手法の概要を図2に示します。

4 微分可能な3Dガウススプラッティング

私たちの目標は、法線のない疎な点群（SfM）から始めて、高品質で斬新なビュー合成を可能にするシーン表現を最適化することです。そのためには、微分可能なボリウム表現の特性を継承しつつ、非構造化かつ明示的なプリミティブによって非常に高速なレンダリングを可能にする必要があります。私たちは、微分可能で2Dスプラットに容易に投影できるため、レンダリングにおける高速なUブレンディングを可能にする3Dガウス分布を選択しました。

我々の表現は、2次元点群を用いる従来の手法[Kopanas et al. 2021; Yifan et al. 2019]と類似しており、各点は法線を持つ小さな平面円であると仮定している。SfM点群の極度の疎性を考えると、法線を推定することは非常に困難である。同様に、

このような推定値からノイズの多い法線を最適化するのには非常に困難です。代わりに、法線を必要としない3次元ガウス分布の集合としてジオメトリをモデル化します。ガウス分布は、点（平均）を中心とするワールド空間で定義された完全な3次元共分散行列によって定義されます[Zwicker et al. 2001a]。

$$G = \frac{1}{(2\pi)^{3/2} |C|^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^T C^{-1} (x - \mu)\right) \quad (4)$$

このガウス分布は、ブレンディング処理で U に乗算されます。ただし、レンダリングするには 3D ガウス分布を 2D に投影する必要があります。Zwickerら[2001a]は、この画像空間への投影方法を示しています。視点変換が与えられた場合、カメラ座標における共分散行列 0 は以下のように与えられます。

$$0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

ここで、は射影変換の近似ヤコビアンです。Zwickerら[2001a]はまた、 0 の3行目と3列目をスキップすると、前研究[Kopanasら2021]と同様に、法線を持つ平面点から開始した場合と同じ構造と特性を持つ2次元分散行列が得られることを示しています。

明らかなアプローチは、共分散行列を直接最適化して、放射場を表現する 3D ガウス分布を取得することです。

しかし、共分散行列は、正の半デファinite行列である場合にのみ物理的に意味を持ちます。すべてのパラメータの最適化には、そのような有効な行列を生成するために容易に制約できない勾配降下法を使用します。また、更新ステップと勾配によって、無効な共分散行列が容易に生成される可能性があります。

その結果、我々は最適化のために、より直感的でありながら同等の表現力を持つ表現を選択しました。3次元ガウス分布の共分散行列は、楕円体の形状を記述するのに似ています。スケーリング行列[と回転行列]が与えられれば、対応するを見つけることができます。

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

両方の要素を独立して最適化できるように、スケーリング用の3Dベクトル B と回転を表す四元数 $@$ を別々に保存します。これらはそれぞれの行列に簡単に交換して結合することができ、 $@$ を正規化して有効な単位四元数を得ることができます。

トレーニング中の自動差別化による大きなオーバーヘッドを回避するために、すべてのパラメータの勾配を明示的に導出します。正確な微分計算の詳細については付録 A に記載されています。

この異方性共分散の表現は最適化に適しており、3Dガウス分布を撮影シーン内の様々な形状の形状に合わせて最適化することで、非常にコンパクトな表現を実現できます。図3はこのようなケースを示しています。

5 適応密度による最適化 3Dガウス分布の制御

私たちのアプローチの核となるのは最適化ステップです。このステップでは、自由視点合成のためにシーンを正確に表現する3Dガウス分布の稠密な集合を作成します。位置 0 、 U 、共分散に加えて、各ガウス分布の色 2 を表すSH係数も最適化することで、視点に依存するシーンの外観を正確に捉えます。これらのパラメータの最適化は、シーンをより適切に表現するためにガウス分布の密度を制御するステップと交互に行われます。

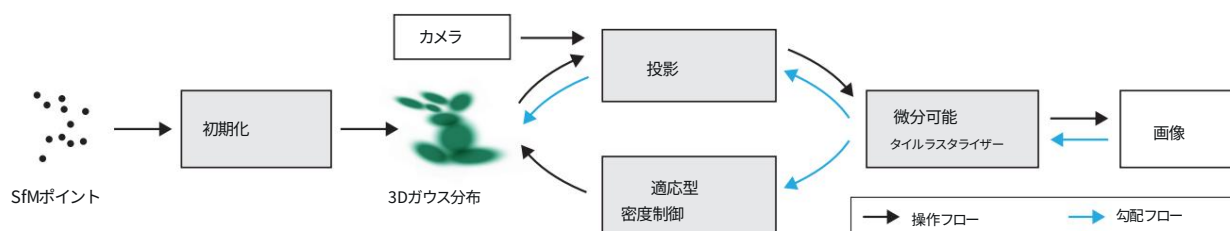


図2. 最適化はスパースなSfM点群から始まり、3Dガウス分布の集合を作成します。次に、このガウス分布の集合の密度を最適化し、適応的に制御します。最適化では、高速なタイルベースレンダラーを使用することで、SOTAの高速放射輝度場法と比較して競争力のあるトレーニング時間を実現しています。

一度トレーニングすると、レンダラーはさまざまなシーンをリアルタイムでナビゲーションできるようになります。



図3. 最適化後の3Dガウス分布を60%縮小して可視化した図（右端）。複雑な形状をコンパクトに表現する3Dガウス分布の異方性形状が明確に示されています。左は実際のレンダリング画像です。

5.1 最適化最適化は、レン

ダリングと、その結果画像とキャプチャされたデータセット内のトレーニングビューとの比較を繰り返すことで実現されます。3Dから2Dへの投影の曖昧さにより、必然的にジオメトリが誤って配置される可能性があります。そのため、最適化ではジオメトリを作成するだけでなく、誤って配置されたジオメトリを破壊または移動する必要があります。3Dガウス分布の共分散パラメータの品質は、表現のコンパクトさにとって非常に重要です。なぜなら、少数の大きな異方性ガウス分布で、広大な均質領域を捉えることができるからです。

最適化には確率的勾配降下法を用い、標準的なGPUアクセラレーションフレームワークを最大限に活用し、一部の演算にはカスタムCUDAカーネルを追加する機能も備えています。これは、最近のベストプラクティス[Fridovich-Keil and Yu et al. 2022; Sun et al. 2022]に倣ったものです。特に、高速ラスタライゼーション（第6節参照）は、最適化における主要な計算ボトルネックとなるため、最適化の効率性に極めて重要です。

Uを[0, 1]の範囲に制限して滑らかな勾配を得るために、シグモイド活性化関数を使用します。また、同様の理由から、共分散のスケールには指数活性化関数を使用します。

初期共分散行列を、最も近い3点までの距離の平均に等しい軸を持つ等方性ガウス行列として推定します。

Plenoxels [Fridovich-Keil and Yu et al. 2022] と同様の標準的な指数関数的減衰スケジューリング手法を用いますが、これはポジションのみを対象としています。損失関数はL1項とD-SSIM項を組み合わせたものです。

$$L = (1 - \lambda)L1 + \lambda D-SSIM \quad (7)$$

すべてのテストで $\lambda = 0.2$ を使用しています。学習の詳細については、セクション7.1のスケジュールおよびその他の要素。

5.2 ガウス分布の適応制御SfMから得られる初期の

疎点集合から開始し、ガウス分布の数と単位体積あたりの密度を適応的に制御する手法を適用します。これにより、初期の疎なガウス分布集合から、シーンをより適切に表現し、適切なパラメータを持つ、より密な集合へと移行することができます。最適化ウォームアップ（7.1節参照）の後、100回の反復ごとに密度を高め、本質的に透明なガウス分布、つまりUが閾値 n_U 未満のガウス分布を削除します。

ガウス分布の適応制御では、空白領域を埋める必要があります。これは、幾何学的特徴が欠落している領域（「再構成不足」）に焦点を当てるだけでなく、シーン内の広い領域をガウス分布が覆っている領域（多くの場合、「再構成過剰」に相当）にも焦点を当てます。

どちらもビュー空間の位置勾配が大きいことがわかります。

直感的に、これはおそらく、まだ適切に再構築されていない領域に対応しており、最適化ではこれを修正するためにガウス分布を移動しようとするためと考えられます。

どちらの場合も緻密化の候補として適しているため、視点空間の位置勾配の平均値がしきい値 g_{pos} を超えるガウス分布を緻密化します。テストではしきい値 g_{pos} を0.0002に設定しました。

次に、このプロセスの詳細を図4に示します。

再構成が不十分な領域にある小さなガウス分布については、新たに作成する必要がある形状をカバーする必要があります。そのためには、ガウス分布を複製し、同じサイズのコピーを作成し、それを位置勾配の方向に移動するのが最適です。

一方、分散の大きい領域にある大きなガウス分布は、より小さなガウス分布に分割する必要があります。このようなガウス分布を2つの新しいガウス分布に置き換え、そのスケールを実験的に決定した係数 $q = 1.6$ で除算します。また、元の3次元ガウス分布をサンプリング用のPDFとして用いて、ガウス分布の位置を初期化します。

最初のケースでは、システム全体の体積とガウス分布の数の両方を増やす必要性を検出し、それに応じて処理します。一方、2番目のケースでは、全体の体積は維持しつつガウス分布の数を増やします。他のボリューム表現と同様に、入力カメラに近いカメラでは最適化が行き詰まる可能性があります。このケースでは、ガウス分布の密度が不当に増加する可能性があります。ガウス分布の数の増加を抑える効果的な方法は、 $\# = 3000$ ごとにU値をゼロに近づけることです。

1ガウス分布の密度は、もちろんNeRFの密度 f と混同してはならない。文字。

1:6・ペルンハルト・ケルブル、ゲオルギオス・コパナス、トーマス・ライムクーラー、ジョージ・ドレアキス

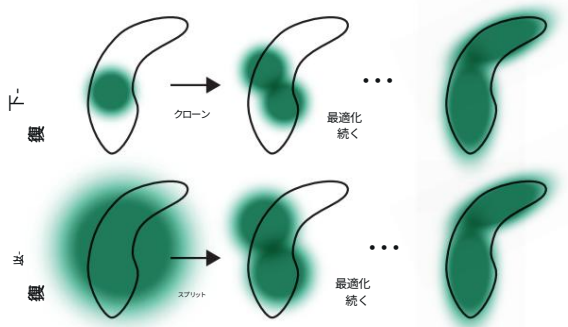


図4. 適応型ガウス分布の緻密化手法。上段（再構成中）：小規模ジオメトリ（黒枠）が十分に覆われていない場合、対応するガウス分布を複製する。上段（再構成過剰）：小規模ジオメトリが1つの大きなスプラットで表現されている場合、それを2つに分割する。

反復処理。次に、最適化により、必要に応じてガウス分布のUが増加しますが、前述のように、カリング手法により、UがnU未満のガウス分布が削除されるようになります。ガウス分布は縮小または拡大して他のガウス分布とかなり重なり合う可能性があります。ワールド空間で非常に大きいガウス分布と、ビュー空間で大きなフットプリントを持つガウス分布は定期的に削除されます。この戦略により、ガウス分布の総数を全体的に適切に制御できます。私たちのモデル内のガウス分布は、常にユークリッド空間のプリミティブのままです。他の方法 [Barron et al. 2022; Fridovich-Keil and Yu et al. 2022] とは異なり、遠くにあるガウス分布や大きなガウス分布に対して空間の圧縮、ワーピング、または投影戦略は必要ありません。

6 ガウス分布の高速微分可能ラスタライザー

私たちの目標は、全体的なレンダリングとソートを高速化して、近似Uブレンディング（異方性スプラットを含む）を可能にし、以前の研究[Lassner and Zollhofer 2021]に存在する勾配を受け取ることができるスプラットの数に対する厳しい制限を回避することです。

これらの目標を達成するために、我々は最近のソフトウェアラスタライゼーション手法 [Lassner and Zollhofer 2021] に着想を得た、ガウススプラット用のタイルベースのラスタライザを設計し、画像全体のプリミティブを一度に事前ソートすることで、従来のUブレンディングソリューション [Kopanas et al. 2022, 2021] の障害となっていたピクセルごとのソートにかかるコストを回避します。この高速ラスタライザは、任意の数のブレンドされたガウス分布に対して、追加のメモリ消費量を抑えながら効率的なバックプロパゲーションを可能にし、ピクセルあたりのオーバーヘッドは一定です。ラスタライゼーションパイプラインは完全に微分可能であり、2Dへの投影（第4節）を与えることで、従来の2Dスプラッティング手法 [Kopanas et al. 2021] と同様に異方性スプラットをラスタライズできます。

我々の手法は、まず画面を16×16のタイルに分割し、次に視錐台と各タイルに対して3次元ガウス分布を選択する。具体的には、視錐台と交差する99%信頼区間を持つガウス分布のみを残す。さらに、ガードバンドを用いて、極端な位置にあるガウス分布（つまり、平均値が近平面に近く、視錐台からはるか外側にあるガウス分布）を単純に除外する。これは、それらの投影された2次元共分散を計算すると不安定になるからである。次に、各ガウス分布を以下の式に従ってインスタンス化する。

ガウス分布は、重なり合うタイルの数に基づいて生成され、各インスタンスにビュー空間の深度とタイルIDを組み合わせたキーが割り当てられます。次に、これらのキーに基づいて、単一の高速GPU基数ソート [Merrill and Grimshaw 2010] を使用してガウス分布をソートします。ポイントのピクセルごとの順序付けは追加されず、この最初のソートに基づいてブレンディングが実行されることに注意してください。結果として、Uブレンディングは構成によっては近似値になる可能性があります。ただし、スプラットが個々のピクセルのサイズに近づくにつれて、これらの近似値は無視できるようになります。この選択により、収束したシーンで目に見えるアーティファクトを生成することなく、トレーニングとレンダリングのパフォーマンスが大幅に向上することがわかりました。

ガウス分布をソートした後、特定のタイルにスプラットする最初と最後の深度ソート済みエントリを識別して、各タイルのリストを生成します。ラスタライズのために、タイルごとに1つのスレッドブロックを起動します。各ブロックは最初にガウス分布のパケットを共有メモリに協調的にロードし、次に特定のピクセルについて、リストを前から後ろへ走査することによって色とU値を累積します。これにより、データのロード/共有と処理の両方において並列性の利点が最大化されます。ピクセル内のUの目標飽和度に達すると、対応するスレッドが停止します。一定の間隔でタイル内のスレッドが照会され、すべてのピクセルが飽和すると（つまり、Uが1になると）タイル全体の処理が終了します。ソートの詳細とラスタライズ手法全体の概要については、付録Cを参照してください。

ラスタライズ中は、Uの彩度が唯一の停止基準です。以前の研究とは対照的に、勾配更新を受け取るブレンドプリミティブの数を制限しません。この特性を強制することで、シーン固有のハイパーパラメータ調整に頼ることなく、任意の変化する深度複雑度を持つシーンを処理して正確に学習できるようになります。したがって、後方パスでは、前方パスでピクセルごとにブレンドされたポイントの完全なシーケンスを回復する必要があります。1つの解決策は、ピクセルごとにブレンドされたポイントの任意の長さのリストをグローバルメモリに格納することです [Kopanas et al. 2021]。暗黙の動的メモリ管理のオーバーヘッドを回避するために、代わりにタイルごとのリストを再度走査することを選択します。これにより、前方パスからガウス分布とタイル範囲のソートされた配列を再利用できます。勾配計算を容易にするために、それらを後ろから前へ走査するようになりました。

トラバーサルはタイル内のピクセルに影響を与えた最後の点から開始され、共有メモリへの点のロードは再び協力的に行われます。さらに、各ピクセルは、その深度が順方向パス中にその色に寄与した最後の点の深度以下の場合にのみ、点の（コストのかかる）オーバーラップテストと処理を開始します。セクション4で説明した勾配の計算には、元のブレンディングプロセス中の各ステップで累積された不透明度値が必要です。逆方向パスで徐々に縮小する不透明度の明示的なリストをトラバースするのではなく、順方向パスの終了時に合計累積不透明度のみを保存することにより、これらの中間不透明度を回復できます。具体的には、各点は順方向プロセスで最終的な累積不透明度Uを保存し、後ろから前へのトラバーサルでこれを各点のUで割って、勾配計算に必要な係数を取得します。

7 実施、結果および評価

次に、実装の詳細、現在の結果、および以前の研究やアブレーション研究と比較したアルゴリズムの評価について説明します。