ユーザ教示によるStructure-from-Motion再構成エラーの修正

金澤 爽太郎[1,2,a)]　周 矜瑶[1,2,b)]　菊池 悠太[2,c)]　小林 颯介[2,d)]　李 淳雨[2,e)]
マートリッチ ファブリス[2,f)]　五十嵐 健夫[2,g)]　樋口 啓太[2,h)]

概要：本研究では，ユーザ教示を利用して Structure-from-Motion (SfM) のプロセスにおける再構成エラーを修正する手法を提案する．SfM はカメラ画像の集合を入力として，カメラ姿勢と三次元点群を特徴点マッチングによって推定する手法である．しかし，SfM は繰り返し構造や似た構造のある環境の再構成を苦手としており，多くの特徴点マッチングの誤りを誘発するため，正確にカメラ姿勢を推定するのが難しい．自動的に誤マッチを発見し削除する機械学習手法は存在するが，精度は学習データに依存することから完璧に修正するのはいまだに困難である．人間の介入によって誤マッチの修正を行う手法も考えることができ，こういった手法は時間をかければ高精度での誤マッチの修正を期待することができるものの，人間が目で誤マッチを同定して，それらを手動で修正するのは時間のかかる作業である．我々の提案手法では，ユーザによる教示を利用して，効率的に誤マッチを削除するアプローチを導入することで，SfM エラーの修正における精度向上と効率的な介入を両立する．本手法は，ユーザがおおまかな撮影時のカメラ位置と撮影範囲を教示することで，各カメラ間の撮影範囲の重複を検証し，重複がない画像ペアのマッチが存在していた場合には，そのペアを誤マッチとみなして削除する．その後，再度 SfM を実施することで，再構成の精度を向上させる．複数のテストケースおよびユーザスタディにおける評価により，本手法が効率的に誤マッチを削除し，SfM による正確な再構成を可能にすることを確認した．

1. はじめに

Structure-from-Motion (SfM) は，複数の二次元画像からカメラの姿勢（位置と方向) およびシーンの三次元点群を推定する，三次元再構成における重要な技術である．SfM は画像ベースの三次元再構成における最初のワークフローであり，Neural Radiance Fields (NeRF) [1] や 3D Gaussian Splatting [2] をはじめとした高度な再構成技術の基盤となっている．

しかし，SfM にはいくつかの課題がある．主な問題の一つとして，撮影シーン内に類似の構造が存在する場合に，誤った特徴点マッチングが生じることが挙げられる．誤った特徴点マッチングが生じると，カメラ間の相対的な位置が誤って推定され，再構成に失敗する可能性がある．具体的には，カメラが撮影時の姿勢とは異なる姿勢のものとして推定されることによって，シーンの再構成の精度低下を引き起こす．

SfM の再構成エラーを克服するために，いくつかの方法が考えられる．最も直接的な方法は，追加で画像を撮影することである．SfM は画像間のマッチングが不十分な場合に失敗する可能性があるため，新しい視点から撮影された画像を追加することで，マッチ数を増やし，再構成精度を向上させることができる．しかし，この方法は追加の撮影が可能な場合に限られ，頻繁に変化する環境や再訪が困難な環境を再構成する場合には適用できない．

追加の撮影が現実的でない場合，特徴点マッチングの手法や閾値など，SfM のパラメータ設定を調整して再構成を改善することができる．COLMAP [3] などの SfM ソフトウェアでは，特徴点マッチングの方法や画像マッチングの探索範囲など，さまざまなパラメータの調整が可能である．しかし，どのパラメータの変更が効果的であるかの判断は非直感的で，パラメータの調整だけでは再構成エラーを修正できない場合も多い．

最近の研究では，再構成するシーンに存在する類似した

[1] 東京大学
[2] Preferred Networks
[a)] sotaro.kanazawa176@gmail.com
[b)] dayaogen@gmail.com
[c)] kikuchi@preferred.jp
[d)] sosk@preferred.jp
[e)] chunyuli@preferred.jp
[f)] fmatulic@preferred.jp
[g)] takeo@acm.org
[h)] khiguchi@preferred.jp

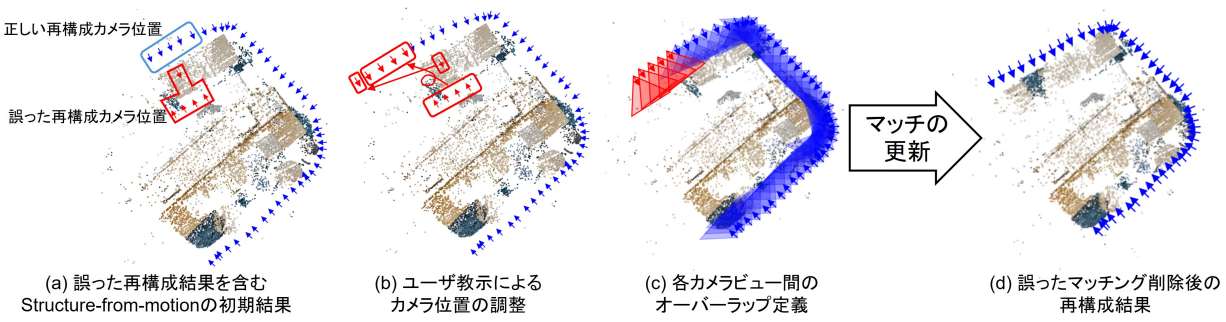

図 1 ユーザ教示による Structure-from-Motion における再構成エラー修正の流れ.

構造が特徴点の曖昧性を引き起こし，誤マッチにつながることが指摘されている [4]．例えば，前後で似た外観を持つ物体を異なる方向から撮影した場合，SfM はそれらを単一の方向から撮影されたものと誤解する可能性がある [5]．このような誤マッチに対処するために，機械学習を用いた手法が提案されている．例えば，画像ペアの分類によって誤マッチを自動的に検出・削除し，その後再構成を行うというものがある [6] ．しかし，機械学習手法は訓練データに依存しているので汎用性に限界がある．

本稿では，最小限の人間の介入で誤マッチを削除し，再構成エラーを効率的に修正するインタラクティブな手法を提案する．コンセプトとしては，ユーザ教示により誤マッチを削除し，改善されたマッチング情報に基づいて再構成を行うというものである．これに関して，直接的な方法としては，ユーザが再構成結果から姿勢推定が不正確なカメラを特定し，そのカメラ画像とマッチしたカメラ画像を目で見て比較して，それらのカメラに関する誤マッチを手動で削除するという方法が考えられる．しかし，この方法は多大な労力と時間を要するため，現実的な運用には課題がある．

我々のユーザ教示を用いた提案アプローチでは，ユーザがおおよその正しいカメラ姿勢を教示し，その情報に基づいてシステムが誤マッチを削除する．この方法は，ユーザが再構成対象の画像シーケンスと撮影環境に関する知識を持っている，もしくは再構成結果と画像シーケンスから正しい再構成結果を推測できることを活用している．ユーザは再構成されたシーンを俯瞰視点から閲覧し，姿勢推定が不正確なカメラを特定して，それらの姿勢を調整する．その後，ユーザは二次元の視錐台として定義されるカメラの撮影範囲を設定する．提案手法は撮影範囲同士が重複しない 2 つのカメラ画像間のマッチを削除し，再度再構成を行う．ユーザはカメラの姿勢を正確に設定する必要はなく，これらの姿勢は誤マッチ削除のためのガイドとして使用され，最終的な姿勢は再構成時のバンドル調整によって決定される．このプロセスは，ユーザが満足のいく結果を得るまで繰り返すことができる．

本手法を評価するために，誤マッチを誘発する状況を再現する 48 枚の画像からなる小規模な三次元再構成データセットを構築した．このデータセットに標準的な SfM 手法を適用すると，誤マッチと再構成の失敗が生じることを確認した．提案手法の有効性を検証するために，パラメータ調整や機械学習ベースの誤マッチ削除技術など，他の再構成エラー修正手法との比較を行った．さらに，ユーザが手動で誤マッチを特定・削除するベースライン手法と比較して，本手法による誤マッチ削除の有効性を評価するユーザスタディを実施した．その結果，参加者は本手法を用いることで，ベースライン手法と比較してより正確に誤マッチを削除し，より少ない労力で品質の高い再構成結果を得られることが示された．最後に，本研究における発見に基づいて，本手法の有効性と制約について議論する．

まとめると，我々の貢献は以下の 3 点である:

- ユーザの指定するカメラ姿勢と視錐台 (撮影範囲) を利用して SfM における誤マッチを削除する，ユーザ教示を用いた手法を提案する．
- 我々が作成した三次元再構成エラーが発生するデータセットを用いて，この提案手法の有効性を実証し，他の修正手法との比較評価を行う．
- ユーザースタディを通じて，我々の手法が誤マッチを目視で特定して削除する手法と比べて，より少ない労力で高精度な再構成の修正を可能にすることを示す．

2. 関連研究

2.1 Structure-from-Motion とその応用

Structure-from-Motion (SfM) は，異なる視点から撮影された二次元画像のデータセットから三次元構造を再構成する，コンピュータビジョンおよびフォトグラメトリの分野において確立された手法である．SfM はカメラの姿勢とシーンの三次元点群を同時に推定する．

初期のアルゴリズムである Tomasi と Kanade による画像間の対応から三次元構造とカメラの動きを復元する手法 [7] から始まり，現在では，COLMAP [3] のようなソフトウェアツールが利用可能で，多様な三次元再構成タスク

を効果的に実行することができる．COLMAP は，特徴点を抽出し，複数のビュー間でマッチングした後，インクリメンタルな SfM に基づいてカメラ姿勢と三次元点を最適化する．最近の研究では，Pan らが高い精度と頑健性を持つグローバルな SfM のアプローチを採用した新しい汎用システム GLOMAP を提案している [8].

最新の三次元再構成とレンダリングの技術の多くは，前処理において COLMAP に大きく依存している．例えば，NeRF (Neural Radiance Fields) [1] や，その派生手法である Instant NGP [9]，NeuS [10] などは，レンダリングパイプラインを初期化するために, COLMAP で得られる正確なカメラ姿勢を必要とする．同様に，最近の研究 [11], [12] を含む 3D Gaussian Splatting [2] とその派生手法でも，カメラパラメータの推定と三次元粗点群の再構成に COLMAP を使用している．このような最新の応用例では，特にカメラ姿勢の推定と三次元点群の生成に関して COLMAP の再構成精度を向上させることが，その後の NeRF や 3D Gaussian Splatting を用いた再構成の品質に直接寄与しており，SfM，特に COLMAP のようなツールは三次元再構成において重要な役割を果たしている．

2.2 SfM 再構成エラーの修正

Structure-from-Motion (SfM)，特に COLMAP [3] で実装されている画像マッチング手法は，視覚的に類似した，または繰り返しの続くパターンを含むシーンを扱う場合，本来マッチングが存在しない画像ペアであっても視覚的な曖昧性からそれらを誤ってマッチングさせてしまい，再構成に失敗してしまうことが多い．

繰り返しまたは類似構造を持つシーンを扱う際の COLMAP の性能を向上させるために, いくつかの方法が提案されている [4], [5], [6], [13]．例えば，Doppelgangers [6] は，視覚的に類似しているが異なる画像を区別するための機械学習ベースの自動アプローチを提案している．これらの方法は，マッチングの曖昧さを解消し，SfM パイプラインにおけるエラーを効果的に減らし，再構成結果を改善することを目的に設計されている．しかし，これら自動的な SfM エラー修正手法は，ベースとなる問題設定や訓練データに大きく依存しており，エラーを修正できないケースが発生する．その場合には，ユーザが介入してエラーを修正することになる．RealityCapture ソフトウェア[*1] の Control Points という機能は三次元再構成におけるユーザの介入手段の一つである．しかし，これらは再構成精度の向上やモデルの統合には効果的であるが，不正確に推定されたカメラ姿勢をユーザが修正するためには効率的でない．これらの課題に対処するために，本稿ではカメラ位置の推定が難しい環境で撮影されたデータにおいて COLMAP の再構成精度を向上させることを目的としたユーザ教示を用いた手法を提案する．

[*1] https://www.capturingreality.com

3. 提案手法: ユーザ教示を用いた誤マッチの削除

3.1 問題定義と対象設定

本研究で扱う問題は，ユーザが画像から三次元再構成を行うために Structure-from-Motion (SfM) を利用する際，画像間の誤ったマッチングがカメラ姿勢および三次元点群の再構成エラーを引き起こすという状況である．COLMAP [3] のような SfM ソフトウェアでは，再構成の初期段階で特徴点の抽出と画像間のマッチングが行われ，その結果はマッチングデータベースに保存される[*2]．その後，このデータベースに基づいて，カメラ間の空間的関係や対応する特徴点の三次元位置が再構成される．しかし，特徴点マッチングの際に誤マッチが存在すると，再構成エラーが引き起こされる．この問題は，撮影シーン内に視覚的に類似したパターンの構造が複数存在する場合に特に顕著である．本稿では，上述の問題に対処するために，ユーザ教示を用いたプロセスによって効果的に誤マッチを削除することで再構成エラーの修正を試みる．

我々は，特に再撮影が困難なシナリオを考慮する．追加の画像を撮影することで再構成の精度と品質は向上すると考えられるが，常に再撮影が可能とは限らない．特に，変化が伴う環境 (例: イベント会場，建設現場) やアクセスが困難な場所 (セキュリティが厳重な場所，病院など) では，同じ条件で再度データを収集することが困難である．また，本システムのユーザについて以下の仮定を置く．1) ユーザは三次元再構成および SfM のために画像データがどのように撮影されるべきかについての基礎知識を持っている．2) 再構成されるシーンについてある程度の理解があり，シーンが撮影された際にカメラが辿った経路を知っている．まとめると，自身でシーンを撮影した，あるいはデータセットの画像や追加情報を使用してシーンとその撮影条件を把握できるユーザを想定する．一方で，撮影プロセスとは無関係なユーザが，撮影された画像のみから撮影カメラ経路について学習してから提案手法を運用するという活用を否定するものではない．

3.2 手法の概要

本手法は，ユーザが撮影されたシーンに関する知識を利用して，エラーを発見し，カメラ姿勢とその視錐台を調整するというプロセスに基づく．上記のプロセス終了後，シ

[*2] より正確には，まずシステムは画像内の特徴点間のマッチングを見つける．その後，2 つの画像間で十分な特徴点のマッチングが検出されると，その画像ペアはマッチングデータベースにマッチングしたペアとして登録される．2 つの画像のペアリングが削除されると，その画像ペア間のすべての特徴点マッチングが削除される．

ステムはそれらの修正を反映するように三次元再構成結果を更新する (図 2).

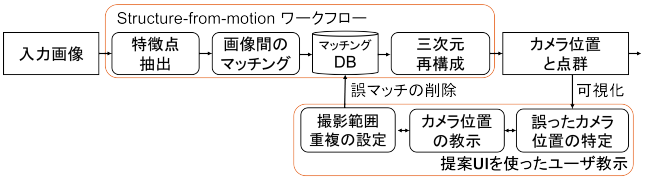

図 2 ユーザー教示を用いた Structure-from-Motion (SfM) のワークフロー．最初に画像を入力として，次に特徴点の抽出とマッチングが行われ，マッチングデータベースに結果が保存される．最初の再構成の後，カメラ位置と三次元点群が推定される．そして，ユーザーが再構成結果を見て誤ったカメラ姿勢を見つけて，手動でカメラ姿勢を調整する．調整後の新しいカメラ姿勢に基づき，システムは誤マッチを削除し，再構成結果を更新する．以上のプロセスは，すべてのカメラが正しく配置され，三次元シーンが適切に再構成されるまで繰り返される．

具体的には，ユーザは誤って配置されたカメラを特定し，その姿勢を実際の姿勢におおよそ一致するように調整する．また，これらのカメラの視錐台も調整する．2 つのカメラの視錐台が重ならない場合，システムはその 2 つのカメラ画像間のマッチを誤マッチと判断し，マッチングデータベースからその画像ペアを削除する．結果として，更新されたマッチングデータベースに基づいた後続の SfM では，カメラ姿勢及び三次元点群の再構成の精度が向上する．このプロセスを，ユーザが結果に満足するまで繰り返す．

3.3 ユーザ教示を用いた誤マッチの削除

3.3.1 誤って再構成されたカメラの特定

ユーザは，カメラ画像のシーケンスと再構成結果 (カメラ姿勢と点群) を確認し，三次元シーン内で姿勢が誤っているカメラを特定する．この際に，三次元点群の再構成エラーが，カメラ姿勢のエラーを発見する手がかりとして利用できる．さらに，カメラの軌跡の (不) 連続性も，カメラ姿勢のエラーを見つける手がかりになる．このようなエラーの特定を支援するため，システムは再構成された三次元点群とカメラ姿勢の俯瞰ビューおよびカメラ画像のシーケンスを同時にユーザに提示する．

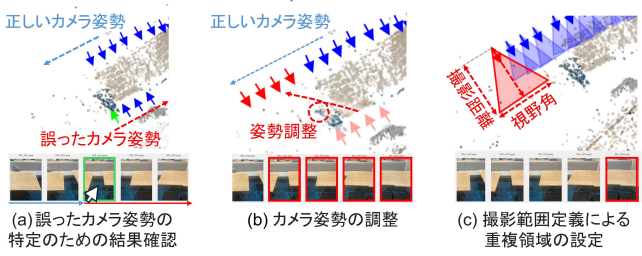

図 3 ユーザーガイドによる誤マッチの削除.

3.3.2 カメラ姿勢の手動調整

ユーザは，誤って推定されたカメラ姿勢を手動で調整し，撮影時の実際の位置に近づける．ユーザはカメラの撮影経路についての知識を活用して，カメラを実際の位置に近い場所に配置することができる．しかし，二次元ユーザインターフェースを操作しながら 三次元カメラ姿勢を正確に指定するのは難しいため，これらのユーザが指定した姿勢は単にシステムが再構成を改善するためのガイドとして利用される．誤マッチが削除された後，システムは後続する再構成の段階でカメラ姿勢を最適化するため，ユーザには正確なカメラ位置の調整は求めない．

3.3.3 誤マッチ削除をガイドする視錐台の調整

カメラ姿勢を調整した後，ユーザはカメラの「マッチング可能な」視錐台 (撮影範囲) を調整することにより，誤マッチ削除の範囲を調整できる．この「マッチング可能な」視錐台は，各カメラの実際の視錐台を模倣し，もともとは三次元であるが，簡略化のためにこれらの視錐台をカメラ位置を頂点とする二次元の二等辺三角形として表現している (図 3)．この三角形の頂角はカメラの視野角 (FoV) に対応し，カメラが撮影するエリアの幅を決定する．また，三角形の高さは撮影距離を示し，カメラがその方向で撮影できる最大距離を表す．底辺の長さ (FoV) と高さ (撮影距離) を操作することで，ユーザは各カメラの撮影範囲の重なりを正確に制御することが可能である．本手法は，視錐台が重ならないカメラ間ではマッチングが発生しないと仮定している．これは，共有している特徴点が両方のカメラビューで見える必要があるからである．ユーザは再構成結果が満足のいくものになるまでこれらのパラメータを繰り返し調整する．

視錐台を二次元で定義することにより，ユーザがカメラビューを指定するタスクが簡略化される．特に多くのカメラを扱う場合，複雑な三次元空間を解釈する際の人間の空間認識能力には限界があり，多くのカメラの重なりを三次元空間で正確に評価し定義するのは認知的に負荷が高く，ミスを招きやすい．視錐台を二次元の領域 (二等辺三角形) として表すことで，ユーザは各カメラが見えているおおよその範囲を直感的かつ効率的に指定することができる．この簡略化により認知的負荷が軽減され，カメラビュー間の重なりの特定が容易になると想定した．このアプローチは視錐台の三次元的な性質を無視しているが，ユーザーが空間的な複雑さを考慮せずに修正できるよう，使いやすさと正確さのバランスをとっている．一方で，この簡略化は，カメラの動きが主に水平であるケースには適しているが，カメラの動きがより三次元的である (例: ドローンでの垂直方向の撮影を行っているとき) ケースは扱えない場合がある．

3.4 誤マッチ削除のアルゴリズム

ユーザがカメラ姿勢を手動で調整した後，システムはデータベースを更新し，画像間の誤マッチを削除する．本手法では，全てのカメラペア間の潜在的なマッチを考慮するために，最初の画像マッチングに exhaustive マッチング[*3]を採用する．誤マッチ削除のステップでは，視錐台が重ならないマッチを削除する．アルゴリズム 1 に誤マッチ削除のプロセスを示す．誤マッチを削除した後，更新されたデータベースに基づいて再構成が行われ，対応する結果がユーザに提示される．

Algorithm 1 False Match Removal

Require: C_{moved}: Set of moved camera IDs
Require: C_{all}: Set of all camera IDs
Require: current_matches: Set of current matches
Ensure: Updated current_matches after removing false matches

```
Initialize false_matches ← ∅
for each c in C_moved do
    for each c' such that (c, c') is in current_matches do
        Compute View(c) and View(c')
        if View(c) ∩ View(c') == ∅ then
            false_matches ← false_matches ∪ {(c, c')}
        end if
    end for
end for
current_matches ← current_matches \ false_matches
```

4. プロトタイプシステム

本稿の提案手法を実証・検証するために，プロトタイプシステムを実装した．図 4 は，プロトタイプシステムのスクリーンショットを示している．(a) の矢印と (b) のカメラ画像は対応しており，(b) のシーケンスパネルを順番に捜査して，(a) 上で姿勢の推定が誤っているカメラを見つける．その後，そのカメラを選択して，キーボード (矢印キーで平行移動，「q」「r」キーで回転) で矢印を移動させる．(d) はマッチングの手動削除機能をサポートしており，ユーザ評価におけるベースライン手法として使用する (5.2 節で後述)．(e) では選択したカメラの視錐台をスライダで調整できる．(f) では誤マッチの削除と再構成の実行をボタンを用いて行うことができる．

プロトタイプシステムは，COLMAP からの SfM 再構成結果と入力画像を受け取るウェブベースのアプリケーションとして実装した．最初に，COLMAP で再構成された点群に対して地面検出を行い，カメラ姿勢情報と組み合わせて再構成の俯瞰図を生成する．ユーザが誤った位置に配置されたカメラの姿勢を調整すると，識別された誤マッチが削除対象としてマークされ，COLMAP がマッチングデータベースを更新する．誤マッチの削除と再構成の再実行のプロセスには，Pycolmap[*4] ライブラリを利用している．さらに，中間結果は段階的にバックアップされており，ユーザはコマンドラインインターフェースを通じて任意の以前のバージョンに戻ることができる．

*3 カメラ画像の全組合わせをマッチ候補として検証する手法．

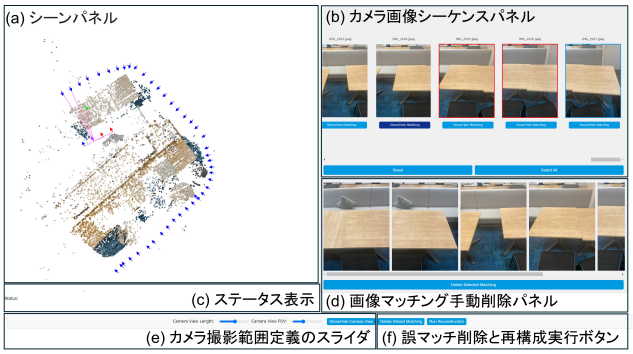

図 4 プロトタイプシステムの概要．本システムは以下のいくつかのコンポーネントで構成されている: (a) シーンパネル: カメラ姿勢を含む三次元再構成空間を俯瞰図で表示するパネル．(b) カメラ画像シーケンスパネル: ユーザが画像を確認し，各画像のマッチングを表示し，カメラ姿勢の調整のために画像を選択できるパネル．(c) ステータス表示．(d) 画像マッチング手動削除パネル: ユーザーが誤マッチを直接削除できる場所 (アノテーションおよびベースライン手法としてこの機能を使用)．(e) カメラ撮影範囲定義のスライダ群．(f) 視錐台が重ならないカメラ間のマッチを削除するボタンおよび再構成実行ボタン．

5. エラー修正手法の比較

5.1 データ収集

本稿の提案手法の有効性を評価するために，オフィス内を移動しながら iPhone 12 Pro で撮影した 48 枚の画像からなる小規模なデータセットを構築した．このシーンには同一のテクスチャを持つ複数のデスクが存在し，単純な再構成では再構成エラーが発生した．具体的には，このデータセットに対して COLMAP をシンプルなカメラモデル (Simple Pinhole Model) と exhaustive マッチングの設定で実行したところ，48 台中 21 台のカメラの推定姿勢が実際の姿勢と大きく異なっていた (図 6 (a) を参照)．

図 5 同一のテクスチャを持つ複数のデスクがあるオフィス内を移動しながら合計 48 枚の画像を iPhone 12 Pro で撮影した．左図の赤い線は撮影経路，オレンジ色は撮影方向を示す．

5.2 誤マッチの正解のアノテーション

誤マッチ削除タスクのための正解データを作成するため

*4 https://github.com/colmap/colmap/tree/main/pycolmap

に，誤マッチを手動でラベリングした．具体的には，まず，誤った姿勢で再構成されたカメラを再構成後の位置や方向から判断する．その後，そのカメラと撮影領域が十分に重ならないカメラとのマッチングを，マッチングしている全てのカメラ画像を人間の目で検査することで削除した．このタスクでは，プロトタイプシステムの手動削除機能を利用し，1 つずつマッチを削除した．ラベルはデータセットを撮影した著者の一人によって作成され，完了までに 30 分以上を要した．この手順を経て誤マッチを削除し，再構成を実行した後，正解として使用できる高品質な再構成結果が得られることを確認した．

図 6 (b) は，誤マッチ削除後の再構成結果を示している．誤マッチを時間をかけて削除することで，カメラ姿勢が正しく推定され，三次元点群の品質も向上した．しかし，人間の目での判断を通した誤マッチの手動削除は時間のかかる作業であり，実際のユースケースでこの手法を用いるのは現実的ない．図 6 (c) に示す結果は，7 章で議論するユーザスタディにおいて，参加者 (P3) が 5 分間で手動削除を行ったものである．

5.3 提案手法による誤差修正

提案手法を用いてユーザがテストデータセットのエラーを修正するプロセスを詳細に説明する．まず，ユーザは画像シーケンスと再構成結果を確認し，誤って姿勢が推定されたカメラを特定する．このデータセットの再構成結果を調べたところ，48 枚中 21 枚の画像が再構成された三次元シーン内で誤った姿勢で配置されていることが分かった．これら 21 枚の画像はさらに 6 つの連続して近傍に配置された画像群（サブシーケンス）に分類された．ユーザは各サブシーケンスから画像を選択し，対応するカメラ姿勢を調整する．その後，ユーザはカメラの視錐台を調整し，視錐台が重ならないカメラ間のマッチを削除した後に再構成を行う．図 6 (d) は，ユーザによる 5 分間の操作の結果を示しており，全てのカメラ位置が正しく再構成されている．この結果は，7 章で説明するユーザスタディの参加者の一人 (P5) によるものである．

5.4 他の手法による誤差修正

提案手法を他の 3 つの技術と比較する．最初の 1 つの手法は自動的なアプローチであり，後の 2 つはユーザの介入を必要とする．

5.4.1 機械学習に基づく曖昧性の解消

最初の手法は，事前学習された分類器を使用した自動的な誤マッチ削除アプローチの Doppelgangers である [6]．この手法は，画像ペアが正しいマッチかどうかを判断するために二値分類器を使用する．具体的には，COLMAP の再構成が失敗した後，分類器を全てのマッチした画像ペアに適用する．分類器は正しいマッチである確率を推定し，一定の閾値以下の確率を持つ全てのペアが削除され，再度再構成が試みられる．この方法の利点は，分類器の学習プロセス以外では，ユーザの介入を必要としないことである．しかし，失敗した場合は他の修正方法を使用する，または分類器を再訓練する必要がある．

データセットの COLMAP のマッチング結果に対して，著者が提供する公開されたコードと事前学習モデルを実行したところ，図 6 (e) に示すように，21 枚中 7 枚の誤ったカメラ位置が修正された．

5.4.2 パラメータ調整

パラメータ調整は，最も基本的なユーザ介入であると考えられる．COLMAP には，カメラモデル，特徴点抽出，画像ペアのマッチング，再構成など，様々な側面で多数のパラメータが存在する．これらのパラメータを適切に設定することで，再構成の成功率を向上させられる可能性がある．しかし，多数のパラメータの試行には時間がかかる上に，ユーザが再構成結果の品質を毎回評価しなければならないことを考えると，パラメータ最適化の自動化も難しい．

本実験では，5 つの異なるカメラモデルと 3 つのマッチング方法を組み合わせて 15 回の試行を行った．30 分間にわたるパラメータ調整を行った結果，OpenCV カメラモデルと sequential matcher の組み合わせの結果が最も誤ったカメラ位置が少なく，21 枚中 11 枚の誤ったカメラ姿勢が修正された (図 6 (f) 参照)．

5.4.3 RealityCapture の Control Points

最後に，商用ソフトウェアである RealityCapture[1] に実装された Control Points という機能を評価する．この手法は，3 枚以上のカメラ画像に写った複数のコントロールポイントを定義することで，画像間の位置関係に関するヒントを提供し，再構成精度を向上させることを目的としている．このアプローチは，一般的に，既に成功している再構成結果をさらに改善したり，複数の再構成結果を統合したりするのには効果的であるが，再構成プロセスが失敗し，不正確な点群が生じた場合には効率を発揮しづらい．我々のテストでは，著者の一人がデータセットの 6 つのコントロールポイントを 92 枚の画像にわたって 定義するのに 30 分を費やしたが，再構成に改善は見られなかった (図 6 (g) 参照)．

6. 他データセットへの適用結果

Yan ら [5] の 3 つのデータセット (cup, oats, ToH) に本手法を適用した．図 7 は，各データセットに対するユーザ教示を用いた再構成の結果を示している．

cup データセットは，カップの表面にある対称的なパターンにより曖昧性を示す 64 枚の画像からなるデータセットである．oats データセットは，隣り合って配置された 2 つの同一のオートミールによって曖昧性を示す 23 枚の画像を含んでいるデータセットである．ToH データセッ

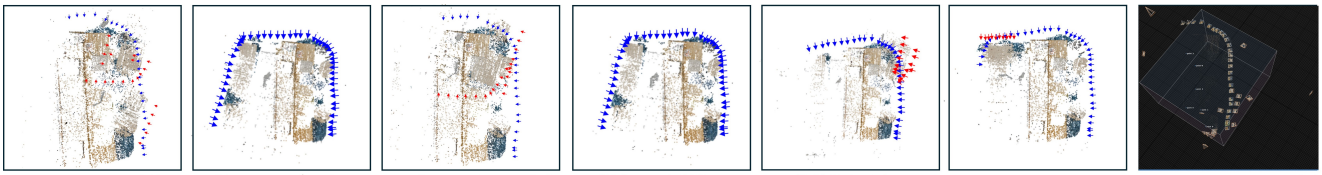

図 6 さまざまな条件下での SfM 結果の比較．(a) デフォルトのパラメータを使用した COLMAP の結果では，誤ったマッチングが原因でカメラ経路推定に大きなエラーが生じている．(b) 誤マッチを手動でアノテーションした後の COLMAP の結果であり，カメラ経路推定の改善が見られる．(c) ベースライン UI で 5 分間のユーザー介入を行った結果で，部分的な改善しか見られない．(d) 提案 UI で 5 分間の操作を行った結果であり，正しい再構築結果が得られている．(e) 機械学習手法による誤マッチ削除の結果 [6]．一定の修正効果はあるが，依然として誤ったマッチングが発生する傾向にある．(f) パラメータ調整を 15 回行った結果であり，微調整によりカメラ経路推定の改善が見られるが，再構成エラーを含む．(g) RealityCapture[1] の Control Points 機能を用いて手動にて 6 つのコントロールポイントを合計 92 枚の画像に渡って定義を通した後の再構成結果 (介入時間は 30 分).

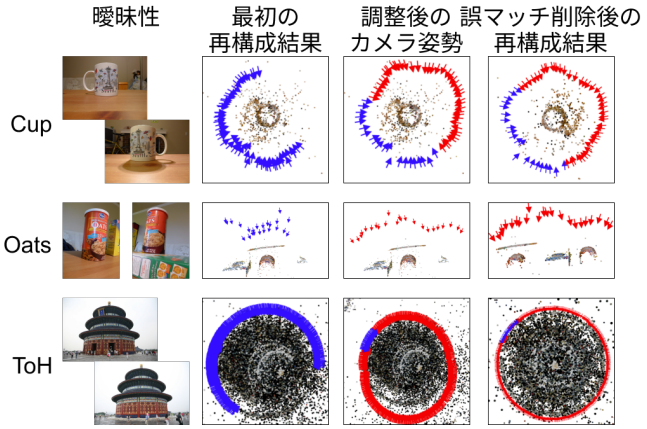

図 7 3 つのデータセット [5] における SfM の曖昧さ解消のためのユーザ介入の結果．上段が cup，中段が oats，下段が ToH データセット．青い矢印は推定されたカメラ姿勢を示し (左)，赤い矢印はユーザによってガイドされたカメラ姿勢 (中央) と最終的なカメラ姿勢 (右) をそれぞれ示している．ユーザの介入により誤ったマッチングが効果的に修正され，データセット全体でカメラ姿勢の推定精度や三次元点群再構築の質が向上している．

トは，天壇 (Temple of Heaven) を撮影した 338 枚の画像から構成されている．天壇は，一周にわたって表面に類似している模様を持つ歴史的建造物であり，これも類似パターンによる曖昧性を含んでいる．これらの曖昧性により COLMAP での再構成にはエラーが含まれるが，我々の提案手法を適用することでエラーを解消することができる．

本手法を適用した結果，全てのデータセットにおいて COLMAP の初期結果と比較して，修正されたカメラ姿勢とより高品質な三次元点群を得た．cup データセットでは，操作に 2 分程度を要し，その中で 64 枚中 43 枚のカメラ姿勢を調整した．oats データセットでは全 23 枚のカメラ姿勢の調整に 2 分程度の時間を要した．より大規模な ToH データセットでは，338 枚中 314 枚のカメラ姿勢を調整し，約 30 分を要した．

まとめると，比較的大規模なものを含む様々な曖昧性を含むデータセットにおいて，提案手法を用いたユーザによる介入が SfM の結果を向上させることを示した．

7. ユーザスタディ

提案手法の有効性を評価するためにユーザスタディを実施した．タスクは，誤マッチにより失敗した三次元再構成結果を改善することである．データセットには 5.2 で説明したものを使用した．

我々の所属機関から，平均年齢 28.6 歳 (標準偏差 6.3) の 8 名 (女性 2 名) の参加者を募集した．全ての参加者は三次元再構成に関する基本的な知識を持ち，そのうち 3 名は三次元再構成技術の開発経験があった．本研究は，所属機関の倫理審査委員会の承認を得た．

実験では，参加者がカメラ姿勢を調整して誤マッチの削除をガイドする提案手法 (3 章) と，参加者が誤マッチを目視で削除するベースライン手法 (5.2 節) を比較した．我々は，提案手法を用いることで参加者が誤マッチをより効率的に削除でき，SfM の結果を改善できるという仮説を立てた．タスクのデータセットの撮影環境に参加者が確実に慣れるため，実験はデータセット撮影に使用されたオフィスの環境で実施した．

7.1 タスク

各参加者は，提案手法とベースライン手法の両方を用いて，それぞれ 5 分間，データセットから誤マッチを削除するタスクを行った．この短いタスクの制限時間は，パイロットスタディに基づいて決定されたもので，ベースライン手法を長時間使用すると認知的負荷が大幅に増加するこ

とが判明したためである．この 5 分間の操作後，参加者は再構成を実行し，その結果を評価した．提案手法は繰り返しの伴うヒューマンインザループ手法としても利用できるが，今回は参加者の全体の操作の回数を 1 回に制限した．これは，データセットの再構成に 2〜3 分を要するので，5 分以内に再試行することが困難だったためである．提案手法とベースライン手法の使用順序は，順序による影響を排除するために参加者間で釣り合いを取った．

誤マッチ削除のインタラクション手法の比較に焦点を当てるため，誤って再構成されたカメラの特定を行う作業はタスクは含めなかった．参加者には，最初からどのカメラ姿勢が不正確であるかの情報が提供された．さらに，両条件で同じデータセットを使用した．これは，参加者が誤った姿勢のカメラとデータセットの撮影条件に関する情報を与えられていたため，データセットに関する潜在的な学習効果が結果にほとんど影響を与えないと考えたためである．

7.2 手順

実験は以下の手順で実施した．まず，各参加者からインフォームド・コンセントを取得した．実験の目的，手順，所要時間 (約 30〜50 分)，データの取り扱い，参加者の権利について説明した．次に，タスクの背景，SfM の基本概念と誤マッチが再構成品質に与える影響を説明した．参加者には，再構成結果を改善するためにこれらの誤マッチを削除することがタスクであると伝えた．その後，データセットの撮影方法を詳しく説明した．撮影環境，撮影時のカメラ位置，画像の特性などを含め，参加者がタスクをより深く理解できるようにした．続いて，提案手法とベースライン手法のシステムの使用方法を教えた．各システムの機能，誤マッチを削除する具体的な手順，重要な操作上の注意点を詳細に説明した．その後，参加者は練習用データセットを用いて最初の手法の練習を行った．その後，参加者は最初の手法のメインセッションに進んだ．5 分間で誤マッチを削除し，その後再構成プロセスを実行した．再構成中，参加者は短い休憩を取った．休憩後，参加者は同様の手順で第二の手法の練習を開始した．システムに慣れたことを確認した後，第二の手法のメインセッションを開始した．参加者は再び 5 分間で誤マッチを削除し，その後再構成プロセスを実行した．両セッションを完了した後，参加者は主観的な評価とフィードバックの提供のためにアンケートに回答した．

7.3 評価指標

ユーザによる教示が再構成結果に与える有効性を評価するために，複数の指標を使用した．具体的には，再構成されたカメラ姿勢の精度を評価するために，平行移動の平均二乗誤差 (MSE) と回転の平均絶対誤差 (MAE) を計算した．誤マッチ削除の精度を測定するために，再現率と適合率を算出した．さらに，User Experience Questionnaire Short Version (UEQ-S) を用いて，参加者の主観的な体験を評価した．

7.3.1 平行移動と回転の誤差

ユーザの介入によって得られたカメラ姿勢と，5.2 節で説明した正解データとの誤差を計算した．SfM の結果には現実世界のスケールと向きの情報が含まれていないため，正解データとユーザの再構成結果を共通の座標系に揃えた．具体的には，ユーザの介入によって変更されない 27 枚の正しく再構成されたカメラ (全 48 枚中) を用いて，ユーザの再構成結果を正解データの座標空間に変換するためのスケールと回転を推定した．座標系を揃えた後，最初の再構成で姿勢が誤っていた 21 台のカメラについて，平行移動の MSE と回転の MAE を計算した．平行移動の MSE は，正解と再構成結果の間のカメラ位置の差から計算した．回転の MAE は，各カメラペアについて回転行列を用いて角度差 (度) を計算することで求めた．

7.3.2 誤マッチ削除の再現率と適合率

誤マッチ削除のために，以下の定義を用いて再現率と適合率を計算した．

- 再現率: 参加者が正しく削除した誤マッチの数を，正解の全ての誤マッチ数で割ったもの．誤マッチ削除プロセスの網羅性を評価する．
- 適合率: 参加者が正しく削除した誤マッチの数を，参加者が削除した全マッチ数で割ったもの．誤マッチ削除の精密性を評価する．

また，再現率と適合率から F1 スコアも計算した．

7.3.3 User Experience Questionnaire 短縮版

参加者の体験を評価するために，User Experience Questionnaire Short Version (UEQ-S) を使用した．参加者は各手法について 8 つの質問に回答し，それに基づいて，実用的な品質，ヘドニック品質，overall スコアを算出した．評価は，UEQ-S で公開されているベンチマークに基づいて行った [14].

7.4 結果

表 1 は，客観的指標の結果を示している．P2 の適合率の結果を除いて，各参加者において全ての指標で本手法がベースライン手法と比較して大幅に高い精度と品質を示している．しかしながら，本手法には P5 の失敗例があり，これは 8.2 節で議論する．

全参加者における平均カメラ姿勢誤差を比較すると，本手法は平行移動の平均二乗誤差 (MSE) が 1.4227 であり，95% 信頼区間は 0.5067 から 2.3388 であった．一方，ベースライン手法では平行移動の平均二乗誤差が 6.4116 で，95% 信頼区間は 5.7085 から 7.1148 であった．信頼区間は重複しておらず，提案手法による精度の大幅な改善を示している．同様に，回転の平均絶対誤差 (MAE) につい

ても，提案手法の平均は 6.6481 (95% 信頼区間は 2.7830 から 10.5132) であったのに対し，ベースライン手法では 115.98 (95% 信頼区間は 86.70 から 145.26) であり，この結果も性能の大幅な向上を示している．

誤マッチ削除に関しても，提案手法は優れた性能を示した．提案手法の平均再現率は 0.88 であり，95% 信頼区間は 0.78 から 0.98 であった．一方，ベースライン手法の平均再現率は 0.50 (95% 信頼区間は 0.42 から 0.59) であった．提案手法の平均 F1 スコアは 0.93 (95% 信頼区間は 0.86 から 0.97) であり，ベースライン手法の 0.66 (95% 信頼区間は 0.57 から 0.73) と比較して高かった．適合率については，提案手法は平均 0.98 (95% 信頼区間は 0.96 から 1.00) であり，ベースライン手法の 0.93 (95% 信頼区間は 0.90 から 0.97) を上回った．適合率の 95% 信頼区間に若干の重なりがあるものの，全体的な結果として，提案手法がベースラインよりも誤マッチ削除の精度で優れていることを示している．

同様に，UEQ-S の結果も提案 UI はすべての指標において”positive evaluation”と評価され，ベースラインと比較して我々の手法がより優れた結果となった (図 8)．UEQ-S のベンチマークでは，我々の手法は”Above average”と評価され，ベースライン手法は”Bad”と評価された．結果，我々の手法がベースラインと比較して優れたユーザ体験を提供することを確認できた．

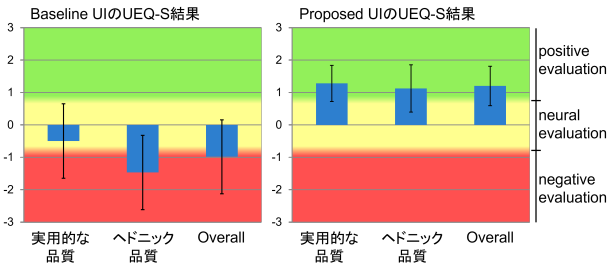

図 8 User Experience Questionnaire 短縮版 (UEQ-S) の結果．

8. 議論

8.1 本手法の利点

実験結果から，ユーザ教示を用いた手法の優位性が明らかになった．参加者は，本ツールを使用して再構成精度を大幅に向上させることができた．さらに，高い精度で誤マッチを削除できることを確認し，ユーザが限られた時間内で効率的に再構成結果を改善できることを示した．UEQ-S のスコアも，本手法が誤マッチ検出において優れたインタラクション体験を提供することの証左となった．

また，本手法は様々なシーンに適用可能であることが証明された．特に，画像数が 300 枚以上の複雑なケースでも，ユーザは誤ったサブシーケンスを持つカメラを容易に選択して再配置できた．さらに，ユーザが正確な撮影位置を知らなくても，カメラの相対位置を正確に設定することで再構成を改善することができた．

8.2 本手法の制約

実験では，参加者 P5 で操作後にモデルが分断される事例が発生した (図 9 (a) 参照)．これは，マッチングの削除の過程で，誤って正しいマッチも一部削除されたためだと考えられる．カメラの姿勢とその視錐台を定義することで削除すべきマッチングペアを指定することは可能だが，ユーザは再構成が完了するまで結果を見ることができない．しかし，提案手法はヒューマンインザループ型の設計となっているので，最初の試行が成功しなかった場合でも，元に戻して再試行できるという柔軟性を持つ．

もう一つの問題点として，我々のインターフェースは二次元の俯瞰ビューを使用しているため，複雑な三次元構造を十分に可視化するのが困難な場合がある．例えば，図 9 (b) に示すように，複数のカメラが垂直に配置されている場合，ユーザは画像のみを用いて編集している層を判断する必要がある．複雑な三次元シーンに対処するには，UI に三次元ビューを組み込むことで，カメラ位置のより正確で直感的な操作を可能にするということが考えられる．

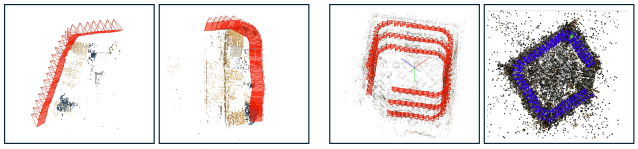

図 9 (a) 失敗例: P5 の実験結果では，再構築が 2 つの別々のモデルに分かれた．(b) 挑戦的な例: [5] の屋内データセットでは，環境内の垂直階層があるため，提案 UI の俯瞰図を使用した効果的な介入が難しくなっている．

本手法はユーザの撮影シーンに関する知見や観察に基づき介入が行われるが，最低限どのような情報があれば介入が可能であるかは明らかになっていないため，さらなるシーンに対しての修正可否の検証が必要と考える．また，2 次元マップ上に撮影時や撮影計画時のルートと，再構成結果のカメラ位置と撮影順を重畳することで，ユーザの知識を補うための UI の改善も今後検討する必要がある．

8.3 他の三次元再構成誤差修正技術との組み合わせ

本稿では，本手法が全ての再構成エラー修正が必要なケースにおいて最適な選択であると主張しているわけではない．実際に扱った例において，50 枚程度では 5 分程度でアノテーションが完了するにもかかわらず，300 枚程度のデータセットのアノテーションに 30 分以上の時間がかかることから、数百枚程度のオーダーの画像数が扱える範囲の限界であると考えられる．しかし，数千枚から数万枚の画像を含むデータセットを扱う場合でも，本インターフェー

表 1 ユーザスタディの各参加者に対する平行移動の平均二乗誤差 (MSE)，回転の平均絶対誤差 (MAE)，再現率，適合率，および F1 スコアの測定結果．P2 の適合率を除き，すべての指標で我々の手法が優れた結果を示している．

参加者	平行移動 MSE ↓		回転 MAE (度) ↓		再現率 ↑		適合率 ↑		F1 スコア ↑	
ID	ベースライン	提案手法	ベースライン	提案手法	ベースライン	提案手法	ベースライン	提案手法	ベースライン	提案手法
P1	6.41608	**1.56911**	133.267	**10.496**	0.31	**0.62**	0.88	**1.00**	0.46	**0.77**
P2	7.02564	**0.00061**	97.559	**0.668**	0.41	**1.00**	**0.96**	0.94	0.57	**0.97**
P3	5.36414	**2.09507**	76.639	**8.523**	0.57	**0.92**	0.90	**0.99**	0.70	**0.95**
P4	6.04193	**2.09787**	61.630	**8.762**	0.56	**0.83**	0.99	**1.00**	0.72	**0.91**
P5	5.33908	—	168.106	—	0.51	**0.92**	0.95	**0.98**	0.66	**0.95**
P6	6.87383	**2.10698**	121.426	**8.862**	0.61	**0.92**	0.97	**0.98**	0.75	**0.95**
P7	7.81761	**2.08914**	136.051	**8.678**	0.56	**0.82**	0.95	**1.00**	0.71	**0.90**
P8	6.41488	**0.00048**	133.221	**0.547**	0.50	**1.00**	0.88	**0.94**	0.64	**0.97**
平均	6.41165	**1.42275**	115.987	**6.648**	0.50	**0.88**	0.93	**0.98**	0.66	**0.93**

スは大規模データ内の誤った特定の部分的なシーケンスを修正するのには有用である．これらの小規模な修正が行われた後，結果は自動的なアプローチによる再構成された 3 次元点群の結合 [15] や RealityCapture の ControlPoints などの技術を使用して，より大きなデータセットに統合できる．また，本手法は他の誤マッチ修正技術を補完することができる．例えば，自動アルゴリズム [6] で識別されない誤マッチは，我々のアプローチを用いて効果的に削除できる可能性がある．我々は，既存の自動修正手法が不十分な場合に，本手法によるユーザの介入が特に有効であると考えている．

9. 結論

本稿では，ユーザが撮影シーンに関する知識を活用して，Structure-from-Motion (SfM) の再構成における誤マッチを修正するユーザ教示を用いた手法を提案した．このアプローチにより，ユーザは誤って再構成されたカメラを効率的に特定して再配置し，その視錐台を調整することで，システムが誤マッチを削除し，再構成結果を改善できる．我々が開発したユーザインターフェース (UI) はこのプロセスを支援し，ユーザが必要な介入を効果的に行えるようにした．実験結果から，本手法は再構成精度とユーザ体験の両面で，カメラ画像の誤マッチの目視による削除を上回ることが示された．

しかし，正しいマッチが誤って削除され，モデルが分断される事例など，いくつかの課題があった．これらの問題への対処は今後の研究で取り組む予定である．UI を強化して二次元と三次元のビューを両方含めることで，特に垂直方向の変化が大きいシーンにおいて，複雑な三次元カメラ配置をより明確に可視化できるようになると考えられる．また，インタラクティブなアプローチと自動アプローチの強みを組み合わせることで，SfM 再構成ワークフローの堅牢性と適用可能性をさらに高めることを目指す．

参考文献

[1] Mildenhall, B., Srinivasan, P. P. et al.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, *ECCV* (2020).
[2] Kerbl, B., Kopanas, G. et al.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering, *SIGGRAPH* (2023).
[3] Schonberger, J. L. and Frahm, J.-M.: Structure-from-motion revisited, *ICCV* (2016).
[4] Heinly, J., Dunn, E. and Frahm, J.-M.: Correcting for duplicate scene structure in sparse 3D reconstruction, *ECCV* (2014).
[5] Yan, Q., Yang, L., Zhang, L. and Xiao, C.: Distinguishing the indistinguishable: Exploring structural ambiguities via geodesic context, *CVPR* (2017).
[6] Cai, R., Tung, J., Wang, Q., Averbuch-Elor, H., Hariharan, B. and Snavely, N.: Doppelgangers: Learning to Disambiguate Images of Similar Structures, *ICCV* (2023).
[7] Tomasi, C. and Kanade, T.: Shape and motion from image streams under orthography: a factorization method, *International journal of computer vision*, Vol. 9, pp. 137–154 (1992).
[8] Pan, L., Barath, D., Pollefeys, M. and Schönberger, J. L.: Global Structure-from-Motion Revisited, *ECCV* (2024).
[9] Müller, T., Evans, A., Schied, C. and Keller, A.: Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, *SIGGRAPH* (2022).
[10] Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T. and Wang, W.: NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction, *NeurIPS* (2021).
[11] Huang, B., Yu, Z., Chen, A., Geiger, A. and Gao, S.: 2D Gaussian Splatting for Geometrically Accurate Radiance Fields, *SIGGRAPH* (2024).
[12] Fei, B., Xu, J., Zhang, R., Zhou, Q., Yang, W. and He, Y.: 3d gaussian splatting as new era: A survey, *IEEE Transactions on Visualization and Computer Graphics* (2024).
[13] Cui, Z. and Tan, P.: Global structure-from-motion by similarity averaging, *ICCV* (2015).
[14] Schrepp, M., Hinderks, A. et al.: Design and evaluation of a short version of the user experience questionnaire (UEQ-S), International Journal of Interactive Multimedia and Artificial Intelligence … (2017).
[15] Choi, S., Zhou, Q.-Y. and Koltun, V.: Robust reconstruction of indoor scenes, *CVPR* (2015).