大規模三次元復元モデルの信頼度を用いた動的点群位置合わせ

金田 綾乃[*1]　Jiazheng Chai[*1]　加藤 卓哉[*1]

Abstract – 近年，**MASt3R** のような事前学習済み大規模画像事前学習モデルは，**RGB** 画像対から高密度 **3D** 点群と信頼度マップを同時に推定できる点から注目されている．一方，古典的な **Non-Rigid ICP (NR-ICP)** は非剛体形状の位置合わせに広く用いられるが，初期位置のずれと外れ値に敏感で，局所最適解に陥りやすい．本研究では，アニメーション付き **3D** シーンを複数視点・複数時刻からレンダリングした時系列の合成データを用いて，**MASt3R** が出力する点群を **NR-ICP** の入力として採用し，さらに信頼度を損失重みに利用する有効性を検証する．

1　はじめに

3 次元点群の位置合わせはロボティクスやコンピュータビジョンにおける基盤的課題の一つであり，物体やシーンの形状復元，モーション解析等の応用事例において重要な役割を果たす．特に剛体変換を仮定した位置合わせで最も広く用いられているのが Iterative Closest Point (ICP) [3] であり, 2 つの点群間における最近傍対応と変換推定を交互に繰り返すことで, 点群間の剛体整合を逐次的に最適化する．しかし，ICP は対象が剛体形状を前提とするため，点群間に伸縮や曲げなどの形状差が存在する時系列データや異なるポーズ間においては収束性と精度が著しく低下するため，対象点群の非剛体変換を許容する Non-Rigid ICP (NR-ICP) が用いられる．

NR-ICP では各点が自由に変形可能な変換モデルを用いつつ、全体が滑らかに変形するよう正則化項を課すことで位置合わせを行う．

Amberg ら [1] は，隣接点間での局所剛性を保つ As-Rigid-As-Possible（ARAP）正則化を用い, 局所な剛体変形と，平行移動・回転・曲げを含む大域的な非剛体変形を同時に推定することで，初期位置や欠損の影響を受けにくいロバストな形状対応を実現した．さらに,Bouaziz ら [4] は, 外れ点の誤差を自動的に小さく評価する誤差関数を導入し, 剛体位置合わせを外れ点に頑健化した．

しかし，幾何学的類似性に基づいてのみ対応探索する従来手法では, 形状変形や視点差が大きい場合には対応付けの不確かさが増し，誤対応が生じやすい．

このような対応精度の限界に対し，近年は，画像やセンサから得られる点群に対して対応の信頼度（不確かさ）を付与し，外れ点除去や損失関数の重みとして活用される [2, 8]．信頼度に基づいた対応点選択によって, ロバスト性が向上する．

さらに近年, 深層学習を用いて画像から直接 3 次元情報を推定する手法が注目されている．DUSt3R は未知のカメラ姿勢や内部パラメータを持つ複数画像から,Transformer により各ピクセル対応の 3 次元点（PointMap）と信頼度を同一座標系で推定する画期的なモデルであり, 極端な視点変化に対しても高いロバスト性を示した [24]．しかし DUSt3R 入力画像ペア間の幾何的対応を明示的に扱わないため，その復元精度には限界があった．そこで後継の MASt3R[13] では, 従来の幾何学的ヘッドに加えて画像特徴マッチング用の新たなヘッドを導入し, 密な特徴量マッチング損失を組み込むことで対応精度を大幅に向上させている．MASt3R は 3 次元的な視点で画像間マッチングを行うため,LoFTR[22] や SuperGlue[17] といった 2 次元ベースの手法を上回る性能を示し, ゼロショットでの単眼距離推定など複数の 3 次元タスクにも適用可能な汎用性を備えている．

以上の背景を踏まえ，本研究は MASt3R から得られる高品質な初期点群と, 各点の信頼度を NR-ICP の対応点誤差の重みとして組み込み，その有効性について検証する．

2　関連研究

本節では，非剛体点群整合に関する代表的枠組みとロバスト化手法を概観し，学習モデルが出力する信頼度の活用という観点から本研究の位置づけを述べる．

2.1　非剛体 3D レジストレーションと変形モデル

非剛体レジストレーションは，形状が伸縮や局所変形を含む状況で対応付けと変形場推定を同時に扱う必要があり，剛体 ICP より難しい課題である．近年の体系的整理として，Deng らのサーベイが手法群（変形表現，最適化，ロバスト性，学習ベース等）を包括的にまとめている [9]．

[*1]株式会社エクサウィザーズ

古典的枠組み NR-ICP 系では，最近傍対応に基づくデータ項と変形の滑らかさ･局所剛性を担保する正則化項を反復最適化する枠組みが広く用いられる．ARAP（As-Rigid-As-Possible）に基づく局所剛性の維持 [20] や，変形グラフ（embedded deformation）による空間変形表現 [21] は，代表的な設計であり，本研究もこの系譜に従う．Li ら [14] は，対応点と変形場に加えて対応の信頼度（重み）も同時に推定し，部分重なり領域を扱う枠組みを提示している．

2.2　外れ値・部分重なりへのロバスト化と高速化

外れ値や部分重なりは，最近傍対応に基づく ICP/NR-ICP を局所解へ誘導しやすい．確率的定式化として，点集合の整合を GMM 当てはめとして扱い，コヒーレント（滑らか）な変位場を仮定する Coherent Point Drift (CPD) が知られる [15]．剛体 ICP の文脈でも，部分重なり・外れ値に対してはトリミングに基づく TrICP が提案されており [7]，局所構造の不確かさを確率モデルとして取り込む一般化 ICP（GICP）[18] も知られる．さらに ICP の推定結果の不確かさを共分散として評価する研究もあり [6, 5]，重み付けやロバスト化と親和性が高い．Bouaziz らは，外れ値の影響を抑えるロバストな目的関数設計により ICP を頑健化した [4]．さらに Jung らは，NR-ICP の線形解法を「厳密に解き切らない」単ステップ更新と前処理により加速する枠組みを提案している [12]．学習ベースの対応推定においても，対応集合の整合性や信頼度を用いて外れ値を除去する研究が進んでいる．FCGF は学習特徴に基づく対応付けを行い [8]，PointDSC は対応集合の整合性に基づく外れ値除去により頑健な整合を実現する [2]．

また，近年はロバスト誤差関数と最適化の設計により，外れ値・部分重なりを扱いつつ高速化を図る研究が進んでいる．Yao らは，滑らかなロバスト推定量と MM/準ニュートン型解法により高速な非剛体整合を実現した [25, 26]．

2.3　画像からの点群推定と信頼度（不確かさ）の活用

近年，画像から密な点群（PointMap）を推定し，同時に信頼度（不確かさ）を出力する事前学習モデルが登場している．DUSt3R は 2 枚画像から密な PointMap を推定し [24]，後継の MASt3R は 3D に基づく画像マッチングを強化することで対応精度を改善した [13]．本研究は MASt3R の信頼度を NR-ICP のデータ項に統合した場合の効果を検証する．さらに動的・非剛体シーンの再構成では，時刻間で幾何が一貫しない領域が対応付けを不安定にすることが古くから指摘されている．Newcombe ら [16] や Innmann ら [11] は，RGB-D 系列からの非剛体再構成・追跡を行い，時刻間対応の難しさとロバストな制約設計の重要性を示した．また Slavcheva ら [19] は対応探索に依存しない定式化を提案しており, Gao ら [10] も動的再構成を扱う．

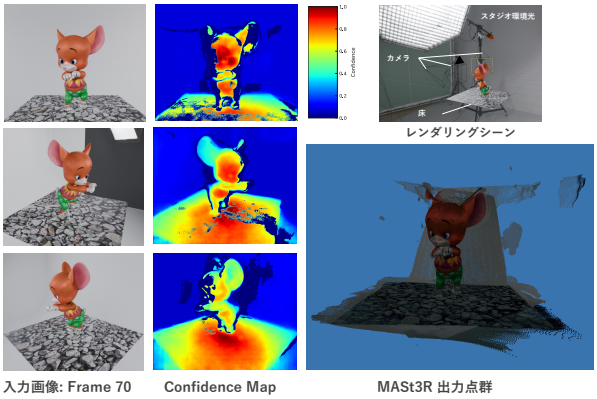

図 1　**MASt3R パイプラインの可視化例**
左から入力マルチビュー画像（Frame #70），その入力画像に対応する MASt3R が推定した信頼度マップ，推定した点群に RGB テクスチャを転写した可視化結果. 低信頼度は背景や光沢領域に集中しており，高信頼度点が主に被写体表面を構成していることが分かる．

また，VGGT[23] は，1 枚〜多数枚（数百枚規模）の画像から，カメラ内外部・深度・PointMap・3D 点トラックを単発推論するフィードフォワード型の統合モデルである．VGGT は深度や PointMap に対する不確かさ（不確実性）も推定できるため，点群整合における重み付けへの応用可能性もある．一方で，本研究は「MASt3R 点群＋信頼度」を NR-ICP に統合した場合の効果検証に焦点を当てるため，次節で述べる理由により MASt3R を採用する．

3　提案手法

本研究では，MASt3R から得られる点群と信頼度マップを活用し，NR-ICP の精度と頑健性を向上させる手法を提案する．

3.1　点群と信頼度マップの出力内容

MASt3R の幾何学ヘッドから出力される点群 (PointMap) は, 入力画像上の各ピクセルに対応する 3 次元点の集合である．画像 I_1 に対してはサイズが H　W の点群 $\mathbf{X}_1 = f\mathbf{x}_{1(u;v)}\ 2\ \mathbb{R}^3\ j\ u = 1;::;H;\ v = 1;::;Wg$ が得られ，同様に画像 I_2 から $\mathbf{X}_2$ が出力される．これらの 3 次元点は, モデル内部で選ばれた基準座標系（通常は画像 I_1 のカメラ座標系）に表現されている．これにより $\mathbf{X}_1, \mathbf{X}_2$ を統合した 3 次元点群を得る．同時に出力される信頼度マップ $\mathbf{C}_1; \mathbf{C}_2$ は各点 $X_{1(u;v)}; X_{2(u;v)}$ に対して推定された信頼度を表すスカラー値である．

信頼度 $C_{i(u;v)}\ 2\ [0;1]$ は, モデルがその対応関係をどの程度確信しているかを示す．例えばテクスチャが

乏しい領域や外れ値の可能性が高い点では低く，特徴が明瞭で対応関係が確実な点では高く出力される傾向がある．

MASt3R の学習時においては，この信頼度に対し回帰損失が課されており，モデル自体が対応点の不確かさを自己教師ありで学習する．そのため，信頼度マップは対応点の品質評価として機能し，後段の最適化において不確実な点の影響を制御する指標となるといえる (図 1)．

3.2 NR-ICP の定式化

NR-ICP では一般に，ソース点群 $\mathbf{X} = f\mathbf{x}_i g$ とターゲット点群 $\mathbf{Y} = f\mathbf{y}_i g$ 間の対応付けに基づいて以下のような損失関数 E を定義する．本研究では，損失関数として重み付き双方向 Chamfer 距離と ARAP（As-Rigid-As-Possible）正則化の線形結合を用いる．

$$L = w_{\mathrm{Cham}} L_{\mathrm{Cham}} + w_{\mathrm{ARAP}} L_{\mathrm{ARAP}}: \tag{1}$$

3.2.1 重み付き Chamfer 距離

ソース点群 $X = f\mathbf{x}_i g_{i=1}^{N}$ とターゲット点群 $Y = f\mathbf{y}_j g_{j=1}^{M}$ に対し，各点の信頼度を $\hat{w}_i$ 2 $[0;1]$ としたとき，

$$L_{\mathrm{Cham}:} = \frac{\sum_{i=1}^{N} \hat{w}_i \, d^2\big(T(\mathbf{x}_i); Y\big)}{\sum_{i=1}^{N} \hat{w}_i} + \frac{1}{M} \sum_{j=1}^{M} d^2\big(\mathbf{y}_j; T(X)\big) \tag{2}$$

と定義する．ここで $d^2(\mathbf{p}; Q)$ は最近傍点までの平方ユークリッド距離であり，$T(\,)$ は Embedded Deformation 変形場である．

3.2.2 ARAP 正則化

変形グラフの節点集合とエッジ集合に対し，局所的な剛体変形を維持するための正則化項 L_{ARAP} を導入する（詳細は [1] に準ずる）．

3.3 信頼度に基づく重み付け

本研究では，MASt3R が出力する信頼度を NR-ICP の重み $\hat{w}_i$ として活用するために，以下の 3 つのバリエーションを提案・比較する．なお，比較対象として信頼度を用いない従来の NR-ICP（$\hat{w}_i = 1$）を **w/o conf** とする．

(i) **信頼度重みのみ (conf)**
MASt3R の生信頼度を正規化した値をそのまま重みとして用いる．

$$\hat{w}_i = \frac{C(\mathbf{u}_i)}{\max_k C(\mathbf{u}_k) + ''}$$

(ii) **Mask ＋信頼度重み（mask+conf）**
信頼度が低い点は背景やオクルージョン領域である可能性が高いため，閾値　未満の点をマスク（$\hat{w}_i = 0$）し，残りの点に対して信頼度重みを適用する． 図 2 に示すように，低信頼点は主に背景側に分布しており，マスクにより外れ点の寄与を抑制できる．

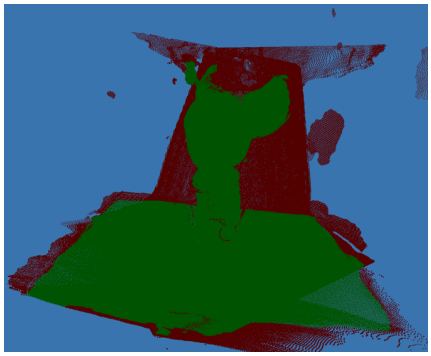

図 2 **信頼度マスクの可視化例**（$\tau = 0.3$）緑は conf ≥ 0.3 の高信頼点，赤は conf < 0.3 の低信頼点を示す．

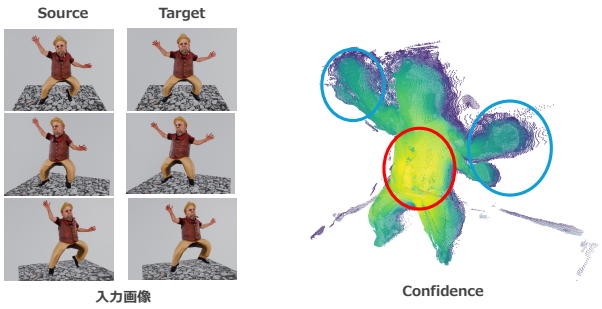

図 3 **混合入力における信頼度（mixed conf）の可視化．** 左は入力に用いたソースとターゲットの混合入力画像群．右図において，静的な胴体は高信頼（明），大きく動いた四肢は低信頼（暗）となっており，動体領域の検出が可能であることを示している．

(iii) **Mask ＋ Mixed Conf（mask+mixed conf）**
フレーム時刻の異なるターゲット画像群とソース画像群を **同一入力として混合**して MASt3R に与える手法である．フレーム間で幾何が一貫しにくい領域（四肢，衣服の揺れ等）では対応が不安定になり，出力される信頼度が相対的に低下する傾向がある．図 3 に，混合入力（Source+Target）を与えた際の信頼度分布を示す．胴体など時刻間で形状が一貫しやすい領域は高信頼となる一方，腕・脚など大きく変形・移動する領域では信頼度が低下していることが確認できる．この特性を利用し，混合入力から得られる信頼度 $\hat{w}_i^{\mathrm{mix}}$ の反転値

$$\hat{w}_i^{\mathrm{inv}} = 1 \quad \hat{w}_i^{\mathrm{mix}}$$

を動体（大変形）スコアとして重みに用いる．これにより，静的な領域よりも動的な領域の整合