

動画像からの 高精度・高密度な 3 次元点群の復元に関する検討

山尾 創輔¹ 酒井 修二¹ 伊藤 康一¹ 青木 孝文¹

概要: 本稿では、動画像から高精度・高密度な 3 次元点群を復元する手法を提案する。動画像を用いることで、限られた枚数の多視点画像ではオクルージョンとなる領域を復元することが可能である。一方で、動画像のすべてのフレームでマッチングをすると膨大な計算時間がかかるため、3 次元復元に適したフレームを選択する必要がある。提案手法は、まず、Structure from Motion を用いて対象の粗い 3 次元メッシュを求める。次に、メッシュとカメラの幾何関係からステレオマッチングに適したフレームを選択する。そして、選択されたフレームから対応する局所領域を抽出し、位相限定相関法を用いて高精度かつ密な 3 次元点群を得る。提案手法を用いることで、計算量を抑えつつ、高精度・高密度な 3 次元復元が可能である。実験により、復元誤差 1mm 以下の精度で密な 3 次元点群を復元できることを示す。

1. はじめに

3 次元復元は、物体の 3 次元形状を計算機上で復元する技術である [1]。3 次元復元技術の中でも、多視点画像からの 3 次元復元は、対象物体をカメラで撮影するという非常に簡便な作業で 3 次元情報を得られるため、広範囲な応用に適用可能である。近年では、Furukawa らの手法や Vu らの手法のように、様々な多視点画像からの 3 次元復元手法が提案されており、カメラ画像のみを用いて高品質な 3 次元復元が行えるようになった [2], [3], [4], [5]。これらの 3 次元復元手法では、比較的少ない枚数の多視点画像を入力として想定している。そのため、より多くの枚数の多視点画像から 3 次元復元を行うことで、3 次元復元精度の向上や複雑な形状の復元が期待できるが、多視点画像の撮影に手間がかかったり、画像マッチングの計算時間が膨大になるといった問題がある。これに対して、本稿では、動画像を用いることで、多視点画像からの 3 次元復元の高精度化を目的とする。動画像を用いる利点としては、(i) 連続した画像列を容易に撮影できること、(ii) 大量の多視点画像を容易に撮影できることが挙げられる。連続した画像列では、直前の画像との画像変形が小さいため、トラッキングによる容易で安定な画像マッチングが可能である。また、多視点画像の枚数が増えることで、オクルージョンの発生を抑え、3 次元復元精度の向上が期待できる。

動画像のように、多くの枚数の多視点画像から高精度に 3 次元復元を行うためには、画像マッチングを行うカメラペアを適切に選択することが重要である。特に、動画像を入力として想定する場合、非常に多くのカメラが存在するため、画像マッチングを行うカメラペアの組み合わせも数多く存在する。基線長の短いカメラペアを用いて 3 次元復元を行う場合、画像マッチングの誤差が 3 次元復元結果に与える影響が大きくなるため、3 次元復元精度が低下する。一方で、基線長の長いカメラペアを用いて 3 次元復元を行う場合、画像変形が大きくなったり、オクルージョンが発生しやすくなったりするため、画像マッチングの誤差が増加する。Furukawa らの手法 [4] に代表される多くの手法では、画像マッチングスコアの低いカメラペアの結果を 3 次元復元に利用しないことで、オクルージョンや補正できない画像変形の影響を抑えている。しかしながら、基線長の短いカメラペアのマッチング結果も基線長の長いカメラペアの結果と同等に扱うため、基線長の短いカメラペアが多く存在する場合、3 次元復元誤差が大きくなる。また、Tola らの手法 [6] では、カメラペアの視差角によって適切な基線長のカメラペアを選択しているが、画像変形やオクルージョンに関しては考慮していない。3 次元復元精度、画像変形、および、オクルージョンなどを考慮してカメラペアを選択する場合、適切なカメラペア選択は物体の 3 次元形状に依存する。しかしながら、一般に、多視点画像からの 3 次元復元では、復元する対象物体の 3 次元形状が未知であるという問題がある。

そこで、本稿では、画像マッチングおよび 3 次元復元に

¹ 東北大学 大学院情報科学研究科
Graduate School of Information Sciences, Tohoku University,
Sendai-shi, Miyagi, 980-8579, Japan

適したカメラを選択し、動画像からの高精度・高密度な 3 次元点群を復元するための新しいフレームワークを提案する。提案手法では、Structure from Motion (SfM) によるカメラ運動推定と同時に物体のおおまかな形状を復元し、そのおおまかな形状をもとに、詳細な 3 次元復元に適したカメラペアを選択する。画像マッチングを行うカメラペアを対象物体の形状に応じて選択することで、複雑な最適化を伴わない単純な手法においても高精度な 3 次元復元が可能である。さらに、提案手法では、SfM におけるフレーム間の画像マッチングと、詳細な 3 次元復元におけるカメラペア間の画像マッチングに、位相限定相関法 (Phase-Only Correlation: POC) に基づく高精度画像マッチング手法を用いる [7]。POC は、画像の位相情報に着目した画像マッチング手法であり、多視点画像からの 3 次元復元において、正規化相互相関 (Normalized Cross-Correlation: NCC) に基づく画像マッチングに比べ、高精度な画像マッチングが可能であることが報告されている [8]。SfM におけるフレーム間の画像マッチングでは、入力される多視点画像が動画像であるため、連続するフレーム間では画像変形が小さく、領域ベースのマッチング手法である POC を用いて高精度に対応付けることが可能である。詳細な 3 次元復元では、SfM で生成した大まかな形状に基づいて、選択したカメラペア間の局所的な画像変形を補正することで、POC を用いた高精度・高密度な 3 次元点群の復元が可能となる。また、本稿では、実際の動画像を用いた精度評価実験を通して、提案手法を用いて復元誤差 1mm 以下の精度で高密度な 3 次元点群を復元できることを示す。

2. 位相限定相関法に基づく画像マッチング

本章では、提案手法で重要となる基礎技術である POC について説明する。POC は、画像の位相情報のみに着目した画像マッチング手法である [7]。POC に基づく画像マッチングは、画像の振幅成分や高周波数成分を除去するため、画像の明るさ変化やノイズに対してロバストである。また、相関関数の解析的ピークモデルを用いた高精度な関数フィッティングを適用することで、サブピクセル精度の画像マッチングが可能である。

画像サイズ $N_1 \times N_2$ ピクセルの 2 次元画像信号 $f(n_1, n_2)$ および $g(n_1, n_2)$ の 2 次元離散フーリエ変換 (Discrete Fourier Transform: DFT) を $F(k_1, k_2)$, $G(k_1, k_2)$ とするとき、このとき、正規化相互パワースペクトル $R(k_1, k_2)$ を次式で定義する。

$$R(k_1, k_2) = \frac{F(k_1, k_2)\overline{G(k_1, k_2)}}{|F(k_1, k_2)\overline{G(k_1, k_2)}|} \quad (1)$$

ただし、 $\overline{G(k_1, k_2)}$ は $G(k_1, k_2)$ の複素共役である。また、 $k_1 = -M_1, \dots, M_1$, $n_2 = -M_2, \dots, M_2$ は離散周波数インデックスであり、 $N_1 = 2M_1 + 1$, $N_2 = 2M_2 + 1$ である。

2 次元 POC 関数 $r(n_1, n_2)$ は、 $R(k_1, k_2)$ の 2 次元逆離散フーリエ変換 (Inverse Discrete Fourier Transform: IDFT) であり、次式で定義する。

$$r(n_1, n_2) = \frac{1}{N_1 N_2} R(k_1, k_2) \sum_{k_1 k_2} W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \quad (2)$$

ここで、 $\sum_{k_1 k_2}$ は、 $\sum_{k_1=-M_1}^{M_1} \sum_{k_2=-M_2}^{M_2}$ を表し、 W_{N_1} , W_{N_2} は回転因子である。2 次元画像信号 $f(n_1, n_2)$ と $g(n_1, n_2)$ が互いに微小量 (δ_1, δ_2) だけ平行移動した関係にあると仮定すると、 $f(n_1, n_2)$ と $g(n_1, n_2)$ の POC 関数 $r(n_1, n_2)$ は次式で与えられる。

$$r(n_1, n_2) \simeq \frac{\alpha \sin\{\pi(n_1 + \delta_1)\} \sin\{\pi(n_2 + \delta_2)\}}{N_1 N_2 \sin\{\frac{\pi}{N_1}(n_1 + \delta_1)\} \sin\{\frac{\pi}{N_2}(n_2 + \delta_2)\}} \quad (3)$$

上式は、2 次元画像信号が微小量 (δ_1, δ_2) だけ平行移動した場合の POC 関数のピーク形状を表している。このとき、POC 関数の相関ピークの高さ α は 2 つの画像信号の類似度の指標に、相関ピークの位置座標 (δ_1, δ_2) は 2 つの信号の平行移動量に相当する。POC を用いた画像マッチングでは、実際に計算される POC 関数の離散データに対して、式 (3) で表される相関ピークの形状モデルをフィッティングし、パラメータ α と (δ_1, δ_2) を推定することで、2 つの画像信号間の類似度とサブピクセル精度の平行移動量を算出する。さらに、(i) 窓関数の適用、(ii) スペクトル重み付けを適用することで、高精度な平行移動量推定が可能である。また、画像ピラミッドを用いた階層的探索により、計算量を抑えつつ安定に高精度な画像マッチングが可能である。

3. 動画像からの高精度・高密度な 3 次元点群の復元

本章では、3 次元復元に適切なカメラペアを選択し、動画像からの高精度・高密度な 3 次元点群を復元するためのフレームワークを提案する。図 1 に示すように、提案手法は、(i) 動画像のトラッキング、(ii) 疎な 3 次元復元、(iii) 三角形メッシュの生成、(iv) 密な 3 次元復元という 4 つの処理から構成される。動画像のトラッキングでは、動画像の先頭フレームから順番に特徴点の検出と追跡を行う。次に、疎な 3 次元復元では、SfM に基づくカメラ運動の推定と物体の疎な 3 次元点群の復元を行う。そして、三角形メッシュの生成では、疎な 3 次元点群から物体の大まかな 3 次元メッシュモデルを生成する。最後に、密な 3 次元復元では、基線長およびオクルージョンを考慮して選択したカメラペアの間で POC に基づく画像マッチングを行い、物体全体を高精度・高密度に 3 次元復元する。以下では、具体的な処理について説明する。

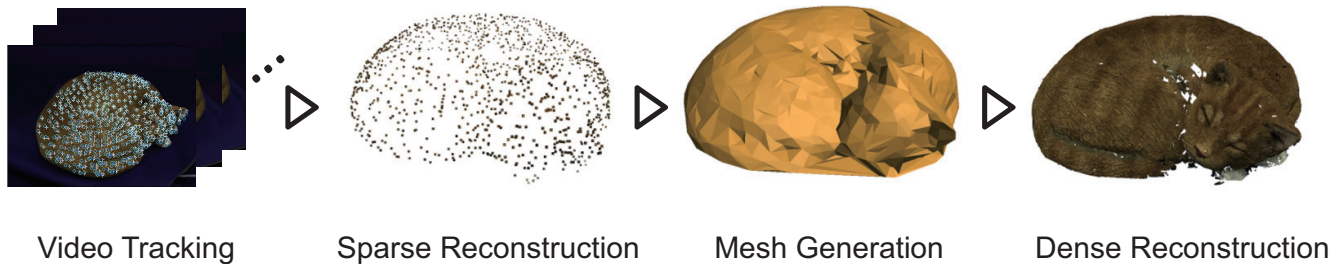


図1 提案手法による動画からの3次元復元: (i) 動画のトラッキング, (ii) 疎な3次元復元, (iii) 三角形メッシュの生成, (iv) 密な3次元復元

3.1 動画のトラッキング

動画のトラッキングでは、動画の先頭フレームから順番に特徴点の検出と追跡を行うことで、連続するフレームにおける特徴点の対応を求める。このとき、後段の処理で均等な大きさの三角形メッシュを生成するために、物体全体に特徴点が偏りなく分布するようにトラッキングを行う。また、連続するフレーム間の画像マッチングでは、フレーム間の画像変形が微小であることや、任意の点に対する対応付けが可能であることから、POCに基づく高精度な画像マッチングを用いる [9]。 K フレームの動画の i 番目のフレーム I_i ($1 \leq i \leq K$) における特徴点の検出と追跡の処理について説明する。まず、現在のフレーム I_i において、コーナ点検出により画像の特徴点を検出する。コーナ点検出には、Shi らの手法 [10] を用いることで、Harris らの手法 [11] や Rosten らの手法 [12] に比べて、より多くの特徴点を偏りなく検出することができる。このとき、特徴点の周辺 $\pm D$ ピクセルに他の特徴点が含まれないように特徴点を検出することで、点の密度を制御する。また、 $i > 1$ であれば、直前のフレーム I_{i-1} から追跡中の特徴点の周辺 $\pm D$ ピクセルを特徴点の検出領域から除くことで、計算量を大幅に削減できる。次に、POCに基づく画像マッチングを用いて、現在のフレーム I_i に含まれる特徴点に対する次のフレーム I_{i+1} の対応点を求める。ここで、 I_i に含まれる特徴点は、 I_i で新たに検出された特徴点と、直前のフレーム I_{i-1} から追跡中の特徴点の集合である。また、画像マッチングで得られる POC の相関ピーク値が閾値を下回る点については、誤対応点として除去する。以上の処理を、フレームを追加しながら繰り返すことで、動画に含まれる特徴点の連続的な対応関係を求める。

3.2 疎な3次元復元

動画をトラッキングした結果から、SfMによりカメラ運動を推定し、同時に物体全体の疎な3次元復元を行う [1], [13]。全体の処理としては、逐次的にフレームを追加しながら、線形解法と非線形最適化を繰り返す。

線形解法では、復元済みの3次元点の座標と、追跡した特徴点の座標の関係から、Direct Linear Transform (DLT)

[1], [13] に基づきカメラの外部パラメータを求める。ただし、最初の2フレームについては、5点アルゴリズム [14] を用いてカメラの外部パラメータを求める。このとき、RANSAC (RANdom SAmple Consensus) を用いることで、誤対応に対してロバストにパラメータを推定する。そして、求めたカメラの外部パラメータから、三角測量の原理に基づいて特徴点の3次元復元を行う。

非線形の最適化では、バンドル調整に基づく再投影誤差の最小化により、推定したカメラの外部パラメータおよび復元した3次元点の座標を最適化する [1], [13]。本稿では、バンドル調整の対象範囲に応じて、大域バンドル調整と局所バンドル調整の2種類を考える。大域バンドル調整は、復元されたカメラおよび3次元点群の全てを最適化の対象とする。具体的には、推定された K 個のカメラの投影行列の集合を $\mathbf{P} = \{\mathbf{p}_i\}$ ($1 \leq i \leq K$)、復元された L 個の3次元点の座標の集合を $\mathbf{Q} = \{\mathbf{q}_j\}$ ($1 \leq j \leq L$) とするとき、次式で定義されるコスト関数 $E_g(\mathbf{P}, \mathbf{Q})$ の最小化を行う。

$$E_g(\mathbf{P}, \mathbf{Q}) = \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^L \|\mathbf{m}_{i,j} - \mathbf{m}_{\text{rep}}(\mathbf{p}_i, \mathbf{q}_j)\|^2 \quad (4)$$

ここで、 $\mathbf{m}_{i,j}$ は、 \mathbf{p}_i に対応するカメラで観測された \mathbf{q}_j の画像上の座標である。また、 $\mathbf{m}_{\text{rep}}(\mathbf{p}_i, \mathbf{q}_j)$ は、 \mathbf{p}_i を用いてカメラ画像上に \mathbf{q}_j を投影したときの投影点の座標である。大域バンドル調整は、カメラパラメータや3次元復元点数が多くなるにつれて計算量が急激に増加するため、一定のフレーム間隔を空けて繰り返し行う。

局所バンドル調整は、復元直後のカメラと、そのカメラで観測される3次元点を最適化の対象とする。具体的には、推定されたカメラの投影行列 \mathbf{p}_i と、そのカメラで観測される L' 個の3次元点の座標の集合を $\mathbf{Q}' = \{\mathbf{q}'_j\}$ ($1 \leq j \leq L'$) とするとき、次式で定義されるコスト関数 $E_l(\mathbf{p}_i, \mathbf{Q}')$ の最小化を行う。

$$E_l(\mathbf{p}_i, \mathbf{Q}') = \frac{1}{2} \sum_{j=1}^{L'} \|\mathbf{m}_j - \mathbf{m}_{\text{rep}}(\mathbf{p}_i, \mathbf{q}'_j)\|^2 \quad (5)$$

局所バンドル調整は、大域バンドル調整に比べて計算量が少ないため、DLT や 5点アルゴリズムによるカメラ運動

の推定を行なう度に適用する。

式 (4), (5) で定義されるコスト関数は非線形関数であるので、非線形最小二乗アルゴリズムの1つである Levenberg-Marquardt (LM) 法を用いてコスト関数の最小化を行う。このとき、スパース行列を利用したバンドル調整を用いることで、大幅な高速化が可能である [15]。

3.3 三角形メッシュの生成

SfM で得られた疎な 3 次元点群から、Labatut らの手法 [16] を用いて三角形メッシュを生成する。Labatut らの手法では、まず、3 次元ドロネー四面体分割により、疎な 3 次元点群を頂点とする四面体を生成する。そして、四面体をノード、四面体の隣接関係をエッジとするグラフを構築する。構築したグラフに対し、グラフカットに基づくエネルギー最小化を行うことで、各四面体が対象物体の内外どちらにあるかを示す 2 値のラベリングを行う。グラフカットで最小化するエネルギー関数 E_m は次式で定義される。

$$E_m(\mathbf{S}) = E_{\text{vis}}(\mathbf{S}) + \lambda_{\text{photo}} E_{\text{photo}}(\mathbf{S}) + \lambda_{\text{area}} E_{\text{area}}(\mathbf{S}) \quad (6)$$

ここで、 \mathbf{S} は四面体を構成する三角形の集合であり、 λ_{photo} , λ_{area} は正の重みである。 $E_{\text{vis}}(\mathbf{S})$, $E_{\text{photo}}(\mathbf{S})$, $E_{\text{area}}(\mathbf{S})$ は、それぞれ \mathbf{S} の可視性、画像との一貫性および滑らかさの制約を表す項である。各項の具体的な計算方法については文献 [16] を参照されたい。三角形メッシュは、異なるラベルを持つ四面体の隣接面を抽出することで得られる。

3.4 密な 3 次元復元

生成した 3 次元メッシュモデルについて、各三角形メッシュごとに、基線長やオクルージョンを考慮してステレオ画像マッチングおよび 3 次元復元に適したカメラを選択し、選択したカメラの間で POC に基づく画像マッチングと局所的な 3 次元復元を行う。まず、注目する三角形メッシュの位置・法線方向、カメラの位置・姿勢、メッシュとカメラ間に生じるオクルージョンといった幾何的な関係から、画像マッチングの基準となる基準カメラと、基準カメラと画像マッチングする対象カメラを 2 つ選択する。そして、それぞれの対象カメラについて、基準カメラとの間で POC に基づく画像マッチングを行い、2 つの対象カメラの間の対応点を 3 次元復元する。以下では、基準カメラおよび対象カメラの選択方法と、対象カメラ間における POC に基づくステレオ画像マッチングおよび 3 次元復元について説明する。

3.4.1 カメラ選択

選択したカメラ間の基線長が短い場合、安定した画像マッチングが可能であるが、画像マッチングの誤差が 3 次元復元精度に大きく影響する。一方、選択したカメラ間の基線長が長すぎる場合、画像間の幾何的な変形が大きくなるため、正確な画像マッチングが困難になる。そこで提案

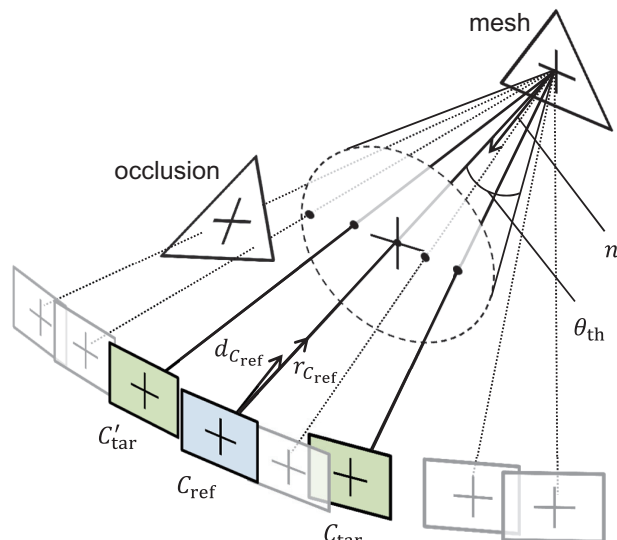


図 2 高精度・高密度な 3 次元復元に適したカメラの選択

手法では、図 2 に示すように、基線長やオクルージョンを考慮してステレオ画像マッチングおよび 3 次元復元に適したカメラを選択する。まず、カメラの画像面に投影されるメッシュの射影変形をできるだけ小さくするため、メッシュと正対するカメラを基準カメラ C_{ref} とする。次に、正確な画像マッチングが可能な範囲で、できるだけ長い基線長となるカメラを 1 つ目の対象カメラ C_{tar} とする。また、 C_{ref} を中心に、 C_{tar} と対称の位置にあるカメラを 2 つ目の対象カメラ C'_{tar} として選択することで、 C_{tar} と C'_{tar} の間で、より長い基線長での高精度な 3 次元復元が可能である。基準カメラ C_{ref} および 2 つの対象カメラ C_{tar} , C'_{tar} を選択する際の具体的な処理手順を以下に示す。

ステップ 1: K フレームの動画像から得られたカメラ群 $\mathbf{C} = \{C_1, C_2, \dots, C_K\}$ の中から、オクルージョンが生じず、メッシュの表側が見えるような位置・姿勢のカメラ群 \mathbf{C}_{vis} を以下のように抽出する。

$$\mathbf{C}_{\text{vis}} = \{C \in \mathbf{C} \mid \mathbf{n} \cdot \mathbf{d}_C < 0 \wedge \mathbf{n} \cdot \mathbf{r}_C < 0 \wedge \text{vis}(C)\} \quad (7)$$

ここで、 \mathbf{d}_C , \mathbf{n} , \mathbf{r}_C はそれぞれ、カメラ C の奥行き方向の単位ベクトル、メッシュの単位法線ベクトル、カメラ C からメッシュの重心へ向かう単位ベクトルである。 $\text{vis}(C)$ は、注目するメッシュとカメラ C の間にオクルージョンとなるメッシュが存在しない場合に真、そうでない場合に偽となる論理式である。なお、オクルージョンの判定については、Havel らの手法 [17] を用いて高速に求めることが可能である。

ステップ 2: 以下のように、メッシュと正対するカメラを基準カメラ C_{ref} として選択する。

$$C_{\text{ref}} = \arg \max_{C \in \mathbf{C}_{\text{vis}}} (\mathbf{n} \cdot \mathbf{d}_C \times \mathbf{n} \cdot \mathbf{r}_C) \quad (8)$$

ステップ 3: \mathbf{C}_{vis} のうち、基準カメラ C_{ref} の画像と正確な画像マッチングが可能なカメラ群 $\mathbf{C}_{\text{match}}$ を以下の

ように抽出する.

$$C_{\text{match}} = \{C \in C_{\text{vis}} \mid \mathbf{r}_C \cdot \mathbf{r}_{C_{\text{ref}}} > \cos \theta_{\text{th}}\} \quad (9)$$

ここで, θ_{th} は正確な画像マッチングが可能な最大の視差角である.

ステップ 4: 以下のように, 基準カメラ C_{ref} との視差角が最大となるカメラを 1 つ目の対象カメラ C_{tar} として選択する.

$$C_{\text{tar}} = \arg \min_{C \in C_{\text{match}}} (\mathbf{r}_C \cdot \mathbf{r}_{C_{\text{ref}}}) \quad (10)$$

また, 以下のように, C_{tar} との視差角が最大となるカメラを 2 つ目の対象カメラ C'_{tar} として選択する.

$$C'_{\text{tar}} = \arg \min_{C \in C_{\text{match}}} (\mathbf{r}_C \cdot \mathbf{r}_{C_{\text{tar}}}) \quad (11)$$

3.4.2 POC に基づくステレオマッチングと 3 次元復元

2 つのステレオペア $C_{\text{ref}}-C_{\text{tar}}$ および $C_{\text{ref}}-C'_{\text{tar}}$ に対し, POC に基づく画像マッチングを行う. POC に基づく画像マッチングは, 画像間の変形に平行移動のみを仮定しているため, 画像間の幾何的な変形が大きい場合, 正確なマッチングが困難である. これに対して提案手法では, Ishii らの手法 [18] を用いて, 射影変換により画像の局所的な変形を補正することで, 画像変形が大きい場合においても高精度なマッチングを行う.

C_{ref} から C_{tar} (または C'_{tar}) への射影変換行列 \mathbf{H} は次式で表される.

$$\mathbf{H} = \mathbf{A} \left(\mathbf{R} + \frac{\mathbf{t}\mathbf{n}^T}{d} \right) \mathbf{A}^{-1} \quad (12)$$

ただし, \mathbf{n} はメッシュの法線ベクトル, d は C_{ref} からメッシュへの距離を表す. \mathbf{R} および \mathbf{t} は, それぞれ C_{ref} から C_{ref} (または C'_{tar}) への回転・並進運動を表す. \mathbf{A} は, カメラの内部行列を表す. \mathbf{n}^T は \mathbf{n} の転置を表す.

次に, C_{ref} の画像と, 幾何変形を補正した C_{tar} の画像との間で POC に基づく画像マッチングを行う. 画像マッチングの基準点は基準画像上に設置することが一般的であるが, この場合, 注目するメッシュと基準カメラの間の距離や相対的な傾きによって, 3 次元復元の密度にばらつきが生じる. そこで, 3 次元空間に存在するメッシュ上に 3 次元点を一定の密度で設置し, これを基準カメラの画像に投影した点を基準点とする. これにより, メッシュと基準カメラの間の距離や相対的な傾きによらず, 物体全体に渡って一定の密度で 3 次元復元を行うことが可能である.

最後に, 2 つの対象カメラ $C_{\text{tar}}-C'_{\text{tar}}$ の間で, 三角測量の原理に基づいて対応点の 3 次元復元を行う. C_{tar} の画像座標 $\mathbf{m} = (u, v)$ について, C'_{tar} の画像座標 $\mathbf{m}' = (u', v')$ が対応するとき, 3 次元点の座標 \mathbf{M} を次式で計算することができる.

$$\mathbf{M} = \mathbf{B}^+ \mathbf{b} \quad (13)$$

ただし,

$$\mathbf{B} = \begin{pmatrix} up_{31} - p_{11} & up_{32} - p_{12} & up_{33} - p_{13} \\ vp_{31} - p_{21} & vp_{32} - p_{22} & vp_{33} - p_{23} \\ u'p'_{31} - p'_{11} & u'p'_{32} - p'_{12} & u'p'_{33} - p'_{13} \\ v'p'_{31} - p'_{21} & v'p'_{32} - p'_{22} & v'p'_{33} - p'_{23} \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} p_{14} - up_{34} \\ p_{24} - vp_{34} \\ p'_{14} - u'p'_{34} \\ p'_{24} - v'p'_{34} \end{pmatrix} \quad (14)$$

である. \mathbf{B}^+ は \mathbf{B} の擬似逆行列である. p_{ij} および p'_{ij} は, C_{tar} および C'_{tar} のそれぞれの投影行列における i 行 j 列の要素である.

POC に基づく画像マッチングにおける POC 関数の相関ピークが閾値以下となる点と, 基準点に対する再投影誤差が閾値以上となる点は, 誤対応点とする. 誤対応点が存在した場合は, 対象カメラを選択し直して, 誤対応と判定された基準点のみについて画像マッチングと 3 次元復元をやり直す. 具体的には, 対象カメラ C_{tar} とのマッチングで誤対応点が存在した場合, カメラ選択で抽出したカメラ群 C_{match} のうちでまだ選択されたことのないカメラについて, C_{tar} との視差角が最も小さいカメラを新たな C_{tar} として選択し直す.

以上の処理を全てのメッシュについて行うことにより, 対象物体全体を高精度に 3 次元復元する.

4. 精度評価実験

本章では, 図 3 に示すような, 実際に撮影した動画像を用いて実験を行い, 従来手法と提案手法の 3 次元復元精度を比較する. 従来手法では, Scale Invariant Feature Transform (SIFT)[19] に基づく SfM と, Furukawa らの手法 (Patch-based Multi View Stereo: PMVS) [4] を用いた密な 3 次元点群の復元を行なう.

4.1 実験方法

まず, 復元対象を設置した回転テーブルを自動で回転させながら, 単眼のカメラで動画像を撮影する. 復元対象は, 猫の置物, 犬の置物, デコレーションタイルの 3 種類である. 使用したカメラは Point Gray 社製の Flea 3 (FL3-U3-13E4C-C) であり, 使用したレンズは, SPACECOM 社製の JHF12M-MP である. 撮影した画像は, $1,280 \times 1,024$ ピクセルのカラー画像である. 動画像のフレーム数は, 猫の置物, 犬の置物, デコレーションタイルのそれぞれで 378, 351, 356 フレームである. カメラから対象物体までの距離は約 1m である. カメラの内部パラメータ (焦点距離および画像中心) は, 事前に撮影したチェッカーパターンを用いて, Zhang らのキャリブレーション手法 [20] に

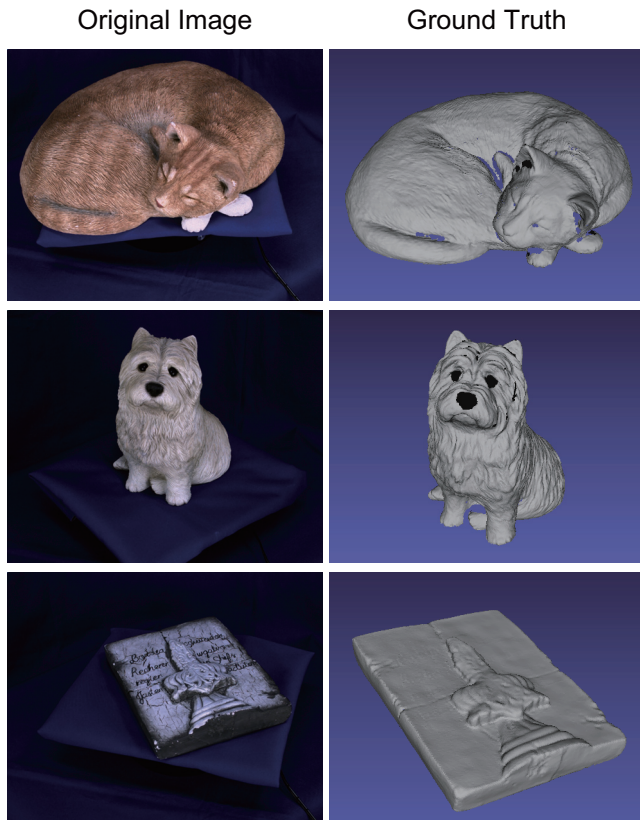


図 3 3次元復元対象と真値の3次元メッシュモデル (上から1行目: 猫の置物, 2行目: 犬の置物, 3行目: デコレーションタイル)

よって求めておく。次に、撮影した動画像から、従来手法および提案手法を用いて3次元復元を行う。従来手法および提案手法の3次元復元誤差を評価するため、コニカミノルタ社の3次元スキャナ VIVID910 を用いて真値の3次元メッシュモデル (図 3) を計測する。

撮影した動画像から、従来手法と提案手法を用いて3次元点群を復元する。動画像はブルーバックで撮影されており、青色領域を抽出することで背景領域のマスクを生成し、背景領域に対する処理を省略する。そして、復元した3次元点群と真値の3次元メッシュモデルとの間でICP (Iterative Closest Point) による位置合わせを行い、3次元復元誤差を評価する。このとき、復元した3次元点群にはスケールの不定性が存在するため、ICPによる位置合わせでは、回転・並進だけでなく、スケールも推定する。

従来手法の実装では、SfMによる疎な3次元復元にBundler [21] を、密な3次元復元にPMVS2 [4] を用いる。入力する画像は、ブルーバック領域を黒色で塗りつぶしてマスクしておく。BundlerにおけるSfMでは、カメラの内部パラメータを固定することで、カメラの外部パラメータのみを推定する。PMVS2では、NCCに基づく画像マッチングのウィンドウサイズを7×7ピクセルに、NCCの相関値の閾値を0.7に、画像を分割するセルのサイズを1×1に設定する。その他のパラメータは、[4]で設定され

表 1 3次元復元誤差のRMS [mm] (復元誤差3mm以下の点のみを対象とする)

	従来手法	提案手法
猫の置物	0.82	0.50
犬の置物	0.95	0.70
タイル	1.02	0.61

表 2 3次元復元点数

	従来手法	提案手法
猫の置物	350,403	921,767
犬の置物	109,287	1,059,561
タイル	110,988	763,681

ている値と同じ物を用いる。

提案手法におけるパラメータの設定について述べる。特徴点検出の際の点の間隔 D は、猫の置物、犬の置物、デコレーションタイルのそれぞれで20, 20, 15ピクセルとする。POCに基づく画像マッチングのウィンドウサイズを32×32ピクセルとする。動画像のトラッキングおよび密な3次元復元におけるPOC関数のピークの閾値を0.65とする。対象カメラを選択する際の視差角 θ_{th} を15°とする。三角形メッシュ上に配置する基準点の間隔を0.001とする。このときの単位長は5点アルゴリズムに用いたカメラの基線長で決まる。密な3次元復元における再投影誤差の閾値を1.5ピクセルとする。

4.2 実験結果

従来手法および提案手法の3次元復元結果と誤差マップを図4に示す。また、3次元復元誤差のRMS (Root Mean Square) と復元点数を、それぞれ表1と表2に示す。ただし、表1は、復元誤差が3mm以下の点のみを対象に計算した。結果から、提案手法は、従来手法と比較して3次元復元誤差が小さく、どの復元対象においても復元誤差1mm以下の精度で3次元復元されることが確認できる。特に、提案手法を用いることで、猫の置物の中央部分やデコレーションタイルの装飾のように、物体表面の起伏が大きく変化するような領域やオクルージョンとなりやすい領域においても、高精度に3次元復元が可能であることが確認できる。これは、物体の大まかな形状をもとにオクルージョンや基線長を考慮して復元に適したフレーム選択し、POCに基づく高精度な画像マッチングを行うためである。また、提案手法では、十分に広い基線長をもつカメラペアを用いて3次元復元を行うため、高精度な3次元点群を復元することができる。

また、靴や紙袋のように複雑な形状を持つ物体やテクスチャの少ない物体を対象に、提案手法を用いて3次元復元を行った結果を図5に示す。これらの物体については、物体表面が柔らかいため、レーザスキャナを用いて測定した3次元メッシュモデルとの比較評価を行うことができない。

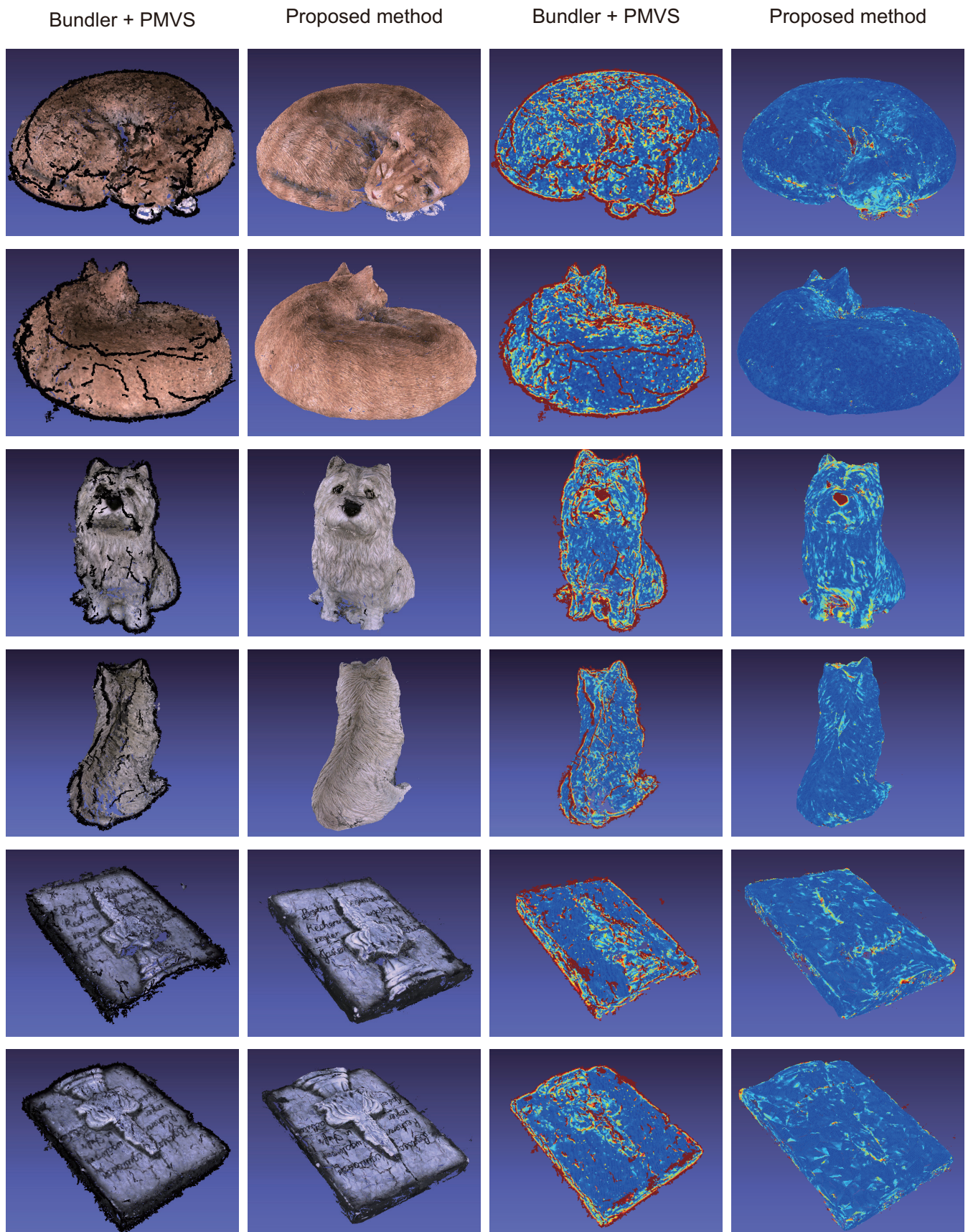


図 4 3次元復元結果(1-2列目)と誤差マップ(3-4列目): 青(0 mm)~赤(3 mm)



図 5 提案手法を用いた靴と紙袋の 3 次元復元結果 (上から 1 行目: 3 次元復元対象, 2-3 行目: 3 次元復元点群)

提案手法を用いることで, このような複雑な形状をもつ物体やテクスチャの少ない物体の 3 次元復元も可能であることが確認できる.

以上より, 提案手法を用いることで, 動画からの高精度・高密度な 3 次元点群の復元が可能であることを示した.

5. まとめ

本稿では, 基線長やオクルージョンを考慮して, 画像マッチングや 3 次元復元に適切なカメラを選択し, 選択したフレームの間で POC に基づく画像マッチングを行うことで, 高精度・高密度な 3 次元点群を復元する手法を提案した. 精度評価実験を通して, 提案手法を用いることで, 3 次元復元誤差 1mm 以下の精度で密な 3 次元点群を復元できることを示した. 今後は, 多視点からの 3 次元復元手法を適用することで, 提案手法の高精度化を行う.

参考文献

[1] Szeliski, R.: *Computer Vision: Algorithms and Applications*, Springer-Verlag New York Inc. (2010).
[2] Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R.: A comparison and evaluation of multi-views stereo reconstruction algorithms, *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 519–528 (2006).
[3] Strecha, C., von Hansen, W., Gool, L. V., Fua, P. and Thoennessen, U.: On benchmarking camera calibration

and multi-view stereo for high resolution imagery, *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 1–8 (2008).
[4] Furukawa, Y. and Ponce, J.: Accurate, dense, and robust multiview stereopsis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362–1376 (2010).
[5] Vu, H.-H., Labatut, P., Pons, J.-P. and Keriven, R.: High Accuracy and Visibility-Consistent Dense Multi-view Stereo, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 34, No. 5, pp. 889–901 (2012).
[6] Tola, E., Strecha, C. and Fua, P.: Efficient large-scale multi-view stereo for ultra high-resolution image sets, *Machine Vision and Applications*, Vol. 23, No. 5, pp. 903–920 (2012).
[7] Takita, K., Muquitt, M. A., Aoki, T. and Higuchi, T.: A sub-pixel correspondence search for computer vision applications, *IEICE Trans. Fundamentals*, Vol. E87-A, No. 8, pp. 1913–1923 (2004).
[8] Sakai, S., Ito, K., Aoki, T., Masuda, T. and Unten, H.: An Efficient Image Matching Method for Multi-View Stereo, *Proc. Asian Conf. Computer Vision*, pp. 1–8 (2012).
[9] 石井純平, 酒井修二, 伊藤康一, 青木孝文, 安藤敏之, 柳拓良: 車載魚眼カメラを用いた走行環境の 3 次元復元の検討, 第 15 回 画像の認識・理解シンポジウム, Vol. 15, No. IS-35 (2012).
[10] Shi, J. and Tomasi, C.: Good features to track, *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 593–600 (1994).
[11] Harris, C. and Stephens, M.: A combined corner and edge detector, *Proc. Alvey Vision Conf.*, Vol. 15, p. 50 (1988).
[12] Rosten, E. and Drummond, T.: Machine learning for high-speed corner detection, *Proc. European Conf. Computer Vision*, pp. 430–443 (2006).
[13] Hartley, R. and Zisserman, A.: *Multiple View Geometry*, Cambridge University Press (2004).
[14] Nistér, D.: An Efficient Solution to the Five-Point Relative Pose Problem, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 6, pp. 756–770 (2004).
[15] Lourakis, M. I. A. and Argyros, A. A.: SBA: A software package for generic sparse bundle adjustment, *ACM Trans. Math. Software*, Vol. 36, No. 1, pp. 1–30 (2009).
[16] Labatut, P., Pons, J.-P. and Keriven, R.: Robust and efficient surface reconstruction from range data, *Computer Graphics Forum*, Vol. 28, No. 8, Wiley Online Library, pp. 2275–2290 (2009).
[17] Havel, J. and Herout, A.: Yet faster ray-triangle intersection (using SSE4), *Trans. Visualization and Computer Graphics*, Vol. 16, No. 3, pp. 434–438 (2010).
[18] Ishii, J., Sakai, S., Ito, K. and Aoki, T.: Wide-baseline stereo matching using ASIFT and POC, *Proc. IEEE Int'l Conf. Image Processing*, pp. 2977–2980 (2012).
[19] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *Int'l J. Computer Vision*, Vol. 60, No. 2, pp. 91–110 (2004).
[20] Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations, *Proc. Int'l Conf. Computer Vision*, Vol. 1, pp. 666–673 (1999).
[21] Snavely, N.: Bundler: Structure from Motion (SfM) for Unordered Image Collections, <http://phototour.cs.washington.edu/bundler/>.