

NAIST-IS-MT1351064

修士論文

RGB-D カメラを用いた非剛体 3次元復元のための
3次元形状テンプレート生成

武原 光

2015年3月12日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士(工学) 授与の要件として提出した修士論文である。

武原 光

審査委員：

横矢	直和	教授	(主指導教員)
加藤	博一	教授	(副指導教員)
佐藤	智和	准教授	(副指導教員)
中島	悠太	助教	(副指導教員)
河合	紀彦	助教	(副指導教員)

RGB-D カメラを用いた非剛体 3次元復元のための 3次元形状テンプレート生成*

武原 光

内容梗概

近年、バーチャル試着システムやバーチャルペットなどの移動・変形する非剛体実物体を仮想的に提示するアプリケーションが開発されており、一般のユーザがこれらのアプリケーションを利用する機会が増えつつある。このようなアプリケーションの多くでは、人手による実物体の3次元モデル化や動きの付与が必要であり、これらの作業には高度な技術が要求されるため、一般のユーザでも利用可能で簡便なモデリング手法が求められている。非剛体物体の全周の形状と動きを自動的に復元する手法として、対象物の3次元形状テンプレートをRGB-Dカメラから得られた3次元点群に一致するように変形させる手法が提案されている。しかし、3次元形状テンプレートは変形していない状態の対象物を撮影して生成する必要があるため、対象となる人や動物に大きな負担がかかる。本研究では非剛体物体の3次元形状および動きの復元に向け、移動・変形中の非剛体物体を単一のRGB-Dカメラで全周撮影したRGB-D画像列から3次元形状テンプレートを生成する手法を提案する。撮影されたRGB-D画像から得られる3次元点群は、3次元形状テンプレートを局所的なアフィン変換で変形したもので表現されると仮定して、提案手法では、非剛体ICPアルゴリズムに基づく手法により3次元形状テンプレート生成する。このアルゴリズムでは、非剛体運動のパラメータの初期値に基づいた3次元点群間の対応点を利用してパラメータを更新することから、位置合わせの可否はこのパラメータの初期値に依存する。提案手法では、RGB-D

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1351064, 2015年3月12日.

カメラから3次元点群と同時に撮影されるRGB画像列に着目し，安定した非剛体位置合わせを実現するために，RGB画像列から得られる追跡点を利用することで，非剛体ICPアルゴリズムにおける対応点の初期値依存性を緩和する．さらに，RGB画像列から得られる画像間の特徴点对応を利用することで蓄積誤差を抑制する．また，物体の運動に対する仮定についても剛体運動から非剛体運動へと段階的にパラメータの数を上げることで，位置合わせの結果が望ましくない局所解へ陥ることを回避する．実験では，移動・変形中の人体を撮影したRGB-D画像から3次元テンプレートを生成し，提案手法の有効性を示す．

キーワード

非剛体3次元復元，テンプレート生成，非剛体位置合わせ，特徴点对応

3D Shape Template Generation for Non-rigid 3D Reconstruction Using an RGB-D Camera*

Hikari Takehara

Abstract

Recently, various applications that virtually display a moving and deforming non-rigid real object, such as a virtual fitting system and virtual pets, have been developed and are available to ordinary users. Most of these applications use a hand-crafted three-dimensional (3D) model, whose motion is manually programmed to mimic the real object. However, building such 3D models requires sophisticated skills and is practically impossible for ordinary users. Therefore, an automatic method that builds 3D models easily for the ordinary users is strongly demanded. One approach for this is to fit a 3D shape template to a sequence of 3D point cloud of the object, captured with an RGB-D camera. However, this approach requires to generate 3D shape template from the stationary object, so that this requirement gives a large burden to the target like a human and an animal. In this study, aiming at reconstructing 3D shape and motion of a non-rigid object from an RGB-D image sequence capturing the moving and deforming object, we propose a 3D shape template generation method from a single RGB-D image sequence capturing the entire directions of the moving and deforming non-rigid object. The proposed method generates the 3D shape template based on the non-rigid ICP algorithm, assuming that the transformations between 3D point clouds given by the RGB-D image and the 3D shape template are described by local affine

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1351064, March 12, 2015.

transformations. The non-rigid ICP algorithm highly depends on initial values of non-rigid transformation parameters because it is generally formulated as non-convex optimization problem and uses the initial values to obtain corresponding points in two point clouds. For stable non-rigid registration, the proposed method leverages the point trajectories in the RGB image sequence, which accompanies the point clouds. It also uses point correspondences between non-consecutive pairs of RGB images to reduce the registration error. Furthermore, the number of transformation parameters are increased during our optimization process so that the solution does not stuck to local minima. We experimentally demonstrate the capability of the proposed method by generating the 3D templates of various moving and deforming human bodies.

Keywords:

Non-rigid 3D reconstruction, template generation, non-rigid registration, feature point correspondences

目次

1. はじめに	1
2. 関連研究と本研究の位置づけ	4
2.1 3次元形状の復元手法の概要	4
2.2 剛体物体に対する全周の3次元形状の復元手法	7
2.3 非剛体物体に対する全周の3次元形状および動きの復元手法	9
2.4 本研究の位置づけと方針	13
3. 点追跡および特徴点对応を用いた3次元形状テンプレート生成	15
3.1 提案手法の概要	15
3.2 追跡点および対応点の検出	18
3.3 3次元形状テンプレート生成のためのエネルギー関数の定義	19
3.3.1 テンプレート点および3次元追跡点の位置合わせ誤差項 E_F	20
3.3.2 SIFTによる3次元対応点の位置合わせ誤差項 E_C	21
3.3.3 正規化項 E_R および E_S	22
3.4 エネルギー関数の最適化	23
4. 実験	25
4.1 実験環境とデータセット	25
4.2 3次元形状テンプレートの生成結果と考察	30
5. まとめ	43
謝辞	44
参考文献	45

目 次

1	移動・変形する非剛体実物体を仮想的に提示するアプリケーションの例	1
2	ステレオ法の原理	5
3	Visual Hull および Space Carving 法 [25] による Photo-consistent な形状	6
4	アクティブステレオ法の概要	6
5	ToF カメラの原理 [28]	7
6	RGB 画像を用いた手法により復元された剛体物体の 3 次元形状の例 [9]	8
7	デプス画像を用いた手法により復元された剛体物体の 3 次元形状の例 [31]	8
8	複数の RGB-D カメラを用いた手法により復元された非剛体物体の 3 次元形状の例 [40]	10
9	単一の RGB-D カメラを用いた手法における 3 次元形状テンプレートと復元された非剛体物体の 3 次元形状の例 [6]	12
10	対象物の小さな変形を許容する 3 次元形状復元 [44]	12
11	提案手法の概要	16
12	オプティカルフローに基づく 2 次元追跡点	17
13	SIFT による 2 次元対応点	17
14	フレーム f と各フレームの対応点の数および最大対応点数フレーム g	19
15	テンプレート点および 3 次元追跡点の位置合わせ誤差項 E_F	20
16	SIFT 対応点に基づくテンプレート点のゆるい対応付け	21
17	近傍の点で定義されたアフィン変換の類似性	23
18	3 つのステップの最適化による位置合わせのイメージ	23
19	DS1 の RGB 画像およびデプス画像の例	27
20	DS2 の RGB 画像およびデプス画像の例	28
21	DS3 の RGB 画像およびデプス画像の例	29
22	3 次元点追跡 (黒) および 3 次元対応点 (赤) の例	31

23	ステップ (ii) で推定したテンプレート点	33
24	ステップ (iii) で推定したテンプレート点	34
25	ステップ (iii) およびステップ (ii) で推定したテンプレート点の比較	35
26	ステップ (iii) で推定したテンプレート点から生成した 3 次元形状 テンプレート	36
27	初期値が剛体変換と恒等変換の場合でのステップ (ii) におけるテン プレート点の推定結果	37
28	パラメータの更新回数に対するエネルギー関数 E' (式 (9)) の推移 .	38
29	ステップ (ii) の有無によるステップ (iii) のテンプレート点の推定結果	39
30	RGB-D カメラ座標系へ変換したテンプレート点 (黒) と 3 次元追跡 点 (赤)	41
31	変換後のテンプレート点と 3 次元追跡点間の平均距離および最大 距離の推移	42

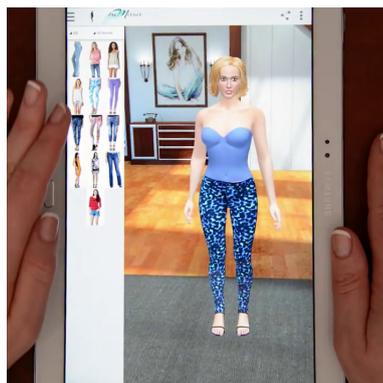
表 目 次

1	DS1 および DS2 の仕様	25
2	実験で用いたパラメータ	26

1. はじめに

近年，バーチャル試着システム [1](図 1(a)) やバーチャルペット [2](図 1(b)) などの移動・変形する非剛体実物体を仮想的に提示するアプリケーションが開発されており，一般のユーザがこれらのアプリケーションを利用する機会が増えつつある．このようなアプリケーションの多くは，人手で作成した実物体の3次元モデルや，実物体の移動・変形を模倣するようにモデルに対して付与した動きに基づいて生成したCGを提示する．しかし，3次元モデルの作成や動きの付与には高度な技術が必要であるため，一般ユーザでも利用可能で簡便なモデリング手法が求められている．

これに対して，単一のRGB-Dカメラで実物体を移動撮影して得られた時系列の画像から3次元形状や動きを復元する手法が提案されている．単純な手法として，RGB-D画像から得られる3次元点群を非剛体ICPアルゴリズム [3,4] を用いて逐次的に位置合わせすることで，対象物の全周の3次元形状を復元する手法が考えられる．このアルゴリズムは，2つの3次元点群に対して各3次元点間の距離や法線方向に基づいて各3次元点を対応付けし，対応点間の距離が近づくように移動・変形を表すパラメータを推定する．この対応付けとパラメータ推定を交互に繰り返すことにより3次元点群の位置合わせを行う．しかし，非剛体ICPアルゴリズムは，剛体ICPアルゴリズム [5] と比べて推定すべきパラメータの数が



(a) バーチャル試着システム [1]



(b) バーチャルペット [2]

図 1: 移動・変形する非剛体実物体を仮想的に提示するアプリケーションの例

多く、対応付けの曖昧性が大きいことや、一般に非凸最適化問題として定式化されることから、位置合わせの結果がパラメータの初期値に大きく依存し、安定した位置合わせが困難であるという問題がある。また、逐次的な位置合わせによりパラメータの推定誤差が蓄積するという問題がある。

蓄積誤差を抑制するために、Zollhöferら [6] や Liら [7] は、対象物の全周形状を網羅した3次元形状テンプレートをを用いて非剛体物体の全周の3次元形状および動きを復元する手法を提案している。これらの手法では、はじめに対象物が変形しない状態で、剛体に対する3次元形状復元手法を用いて3次元形状テンプレートを生成する。続いて、生成した3次元形状テンプレートを、移動・変形中の対象物を撮影した時系列の3次元点群に一致するように非剛体ICPアルゴリズム [3,4] を用いて変形することで、各時刻での対象物の動きを復元する。これらの手法では、全周の3次元形状テンプレートを生成するために、対象物が変形しない状態で全周撮影を実施する必要があるため、対象となる人や動物にとって負担が大きいという問題がある。そのため、撮影中における物体の変形を許容した3次元形状テンプレートの生成手法が望まれている。

このような課題に対して、本研究では非剛体物体の3次元形状および動きの復元に向け、移動・変形中の非剛体物体を単一のRGB-Dカメラで全周撮影したRGB-D画像列から3次元形状テンプレートを生成する手法を提案する。提案手法では、撮影したRGB-D画像から得られる3次元点群は、3次元形状テンプレート上の点(テンプレート点)を局所的なアフィン変換で変形したもので表現されると仮定して、非剛体ICPアルゴリズムに基づいてテンプレート点を求める。安定した非剛体位置合わせを実現するために、提案手法ではRGB画像から得られる時間的に連続する画像間での追跡点を利用することで、対応付けの曖昧さを緩和し、推定するパラメータの数を段階的に上げることで最適化問題の非凸性を緩和する。また、蓄積誤差を抑制するために、RGB画像から得られる時間的に連続しない画像間での特徴点对応を利用する。実験では、移動・変形中の人を撮影したRGB-D画像から3次元テンプレートを生成し、提案手法の有効性を示す。

以下、2章では、関連研究および本研究の位置づけについて述べる。3章では、本論文の提案手法である、点追跡および特徴点对応を用いた3次元形状テンプレ

ト生成について述べる。4章では，実物体を対象とした実験を行い，提案手法の有効性を示す。最後に，5章でまとめと今後の展望について述べる。

2. 関連研究と本研究の位置づけ

本章では、まず3次元形状の復元手法を概観し、次に全周の3次元形状を復元する従来研究を紹介する。全周の3次元形状の復元手法として、対象物の形状が変化しない剛体物体を対象とする手法と、形状の変化を許容する非剛体物体を対象とする手法について詳述する。最後に関連研究に対する本研究の位置付けを述べる。

2.1 3次元形状の復元手法の概要

実物体の3次元形状の復元手法には、対象物を撮影した画像から形状を推定する受動的な手法と、対象物に光や電波などを照射してその反射から形状を計測する能動的な手法がある。受動的な手法では、現在広く普及しているデジタルカメラやビデオカメラを利用して対象物を撮影することにより手軽に3次元形状を復元できる。能動的な手法は、Microsoft Kinectなどの安価な計測機器の登場により、一般のユーザにも利用されつつある。

3次元形状の受動的な復元手法として、ステレオ法 [8,9] や視体積交差法 [10–12] などが提案されている。ステレオ法は、複数の位置から撮影された複数の画像から形状を推定し、3次元形状を復元する手法である。基本的な手法である2眼ステレオ法 (図2) は、位置姿勢が既知の2台のカメラを用いて対象物を撮影し、2つの画像間の対応を求めることで、三角測量の原理に基づいて3次元位置を推定し、3次元形状を復元する手法である。利用する画像が3枚以上の場合は特にマルチビューステレオ法 [8,9] と呼ばれる。カメラの位置姿勢については、特徴点に基づく手法 [13,14] や画素値に基づく手法 [15,16] などを利用することで推定できる。特徴点に基づく手法 [13,14] は、各RGB画像に対する特徴点の抽出および、他の画像との対応点探索を行い、対応付けられた特徴点に対して一つ定義した3次元点のRGB画像への投影点とそれぞれの特徴点間の距離(再投影誤差)を最小にする3次元点およびカメラの位置姿勢を推定する。特徴点の抽出および対応付けには、局所特徴量に基づく対応点探索手法 [17–19] や、オプティカルフローに基づく対応点探索手法 [20–22] などが利用できる。画素値に基づく手法 [15,16] は、

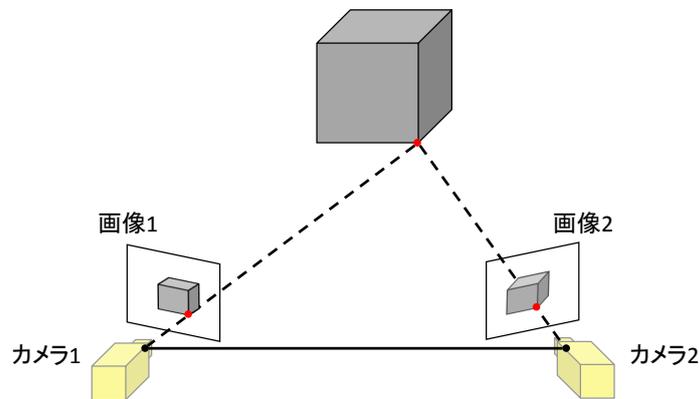


図 2: ステレオ法の原理

RGB 画像中の画素において Photo-consistency が最大となるデプス画像およびカメラの位置姿勢を推定する。

視体積交差法 [10–12] は、2次元シルエット画像を3次元空間に逆投影して得られる視体積の内側に対象物が含まれる、というシルエット制約に基づいて、2次元のシルエット画像から対象物の3次元形状を復元する手法である。この手法では、複数の位置から撮影した画像から、カメラの位置姿勢の推定、および画像中における対象物のシルエット抽出を行い、それぞれの視体積の共通部分 (Visual Hull) を求めることにより、対象物の3次元モデルを復元する。シルエットの抽出には、背景モデルに基づく背景差分法 [23] や、GrabCut [24] などのインタラクティブな手法が利用できる。視体積交差法では、シルエット画像上に現れない物体の凹み部分などの3次元形状を復元できない。これに対して、視体積交差法により対象物体の概形を生成したのち、それぞれの画像で観測される対象物表面の色の類似性を表す Photo-consistency をもとに、ステレオ法により凹み部分を含む高精度な3次元形状を復元する、Space Carving 法 [25](図 3) が提案されている。これらの受動的な復元手法で3次元形状を復元するためには、少なくとも2つ以上の異なる位置から撮影した画像が必要となる。

3次元形状の能動的な復元手法として、プロジェクタなどを用いたアクティブステレオ法 [26,27] や、Time-of-Flight (ToF) カメラを用いたレーザ測距法 [28,29] が提案されている。アクティブステレオ法 [26,27] では、プロジェクタなどの光を

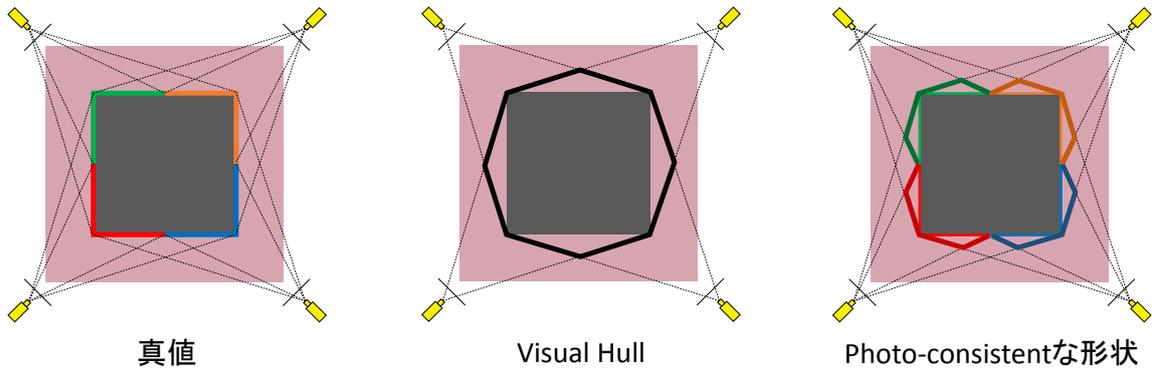


図 3: Visual Hull および Space Carving 法 [25] による Photo-consistent な形状

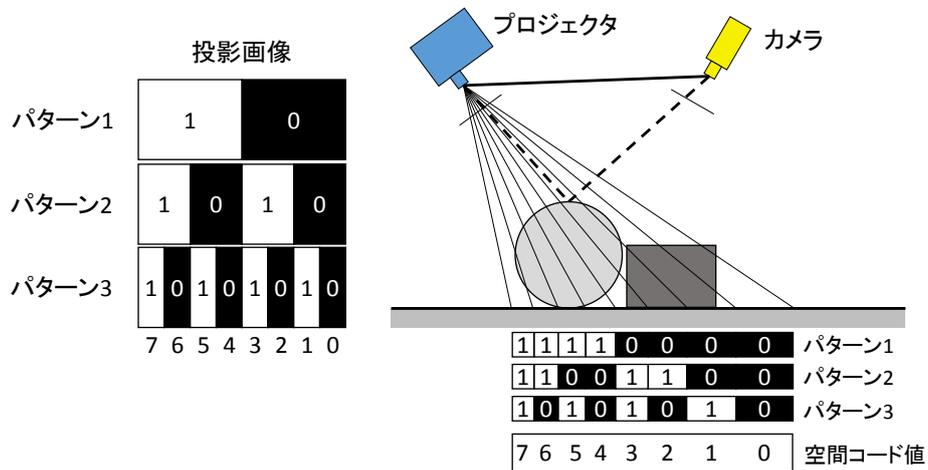


図 4: アクティブステレオ法の概要

照射する装置とカメラ間の位置姿勢が既知であるものとして、対象物に特定のパターンで光を照射し、その光の照射位置をカメラで計測することで、三角測量の原理に基づいて3次元位置を推定し、対象物の3次元形状を復元する(図4)。この手法は、前述した2眼ステレオ法における2台のカメラのうち、1台のカメラを照射装置に置き換えたものと考えることができ、2眼ステレオ法に比べて対応点探索の曖昧さが少ないため、安定してデプス画像を推定できる。

ToFカメラを用いたレーザー測距法[28,29]では、図5に示すように、照射したレーザー光が対象物に反射して撮像素子に到達するまでの時間を計測することで、

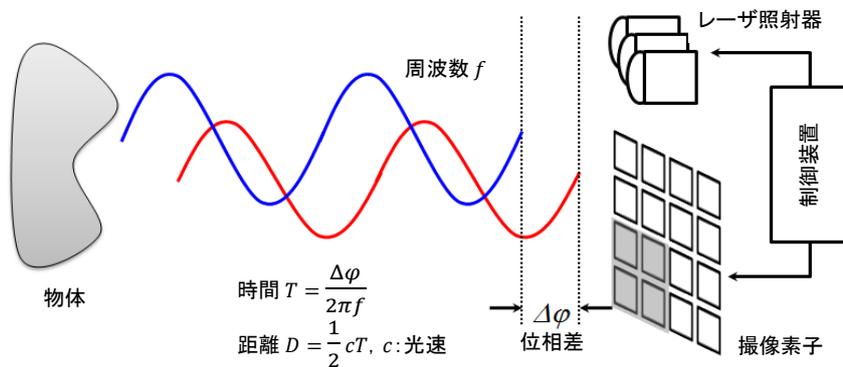


図 5: ToF カメラの原理 [28]

撮像素子から対象物までの距離を測定する。このレーザー光を 2 次元走査することで測定した撮像素子から対象物までの距離画像からデプス画像を求め、デプス画像を基に対象物の 3 次元形状を復元する。これらの能動的な復元手法では、単一の位置での計測で対象物の 3 次元形状を復元できる。

受動的な手法と能動的な手法のどちらにおいても、遮蔽などによって観測できない部位の形状を復元できないため、対象物の全周の形状を復元するためには、複数の位置での計測を統合する必要がある。全周の形状を復元する手法には、対象物の形状が変化しない剛体物体を対象とする手法と、形状の変化を許す非剛体物体を対象とする手法がある。2.2 節では、剛体物体を対象とした全周の 3 次元形状の復元手法について、2.3 節では、非剛体物体を対象とした全周の 3 次元形状の復元手法について述べる。

2.2 剛体物体に対する全周の 3 次元形状の復元手法

剛体物体を対象とした全周の 3 次元形状の復元手法は、RGB 画像を用いた手法 (受動的な手法) [9, 30] とデプス画像を用いた手法 (能動的な手法) [31] に大別できる。RGB 画像を用いた手法およびデプス画像を用いた手法で復元された 3 次元形状の例を図 6 および図 7 に示す。

RGB 画像を用いた復元手法では、はじめに複数位置から撮影した RGB 画像からカメラ位置姿勢を推定する。次に、推定したカメラ位置姿勢をもとに複数の



図 6: RGB 画像を用いた手法により復元された剛体物体の 3 次元形状の例 [9]



図 7: デプス画像を用いた手法により復元された剛体物体の 3 次元形状の例 [31]

RGB 画像から，密な 3 次元点群を推定する．最後に，得られた密な 3 次元点群から 3 次元形状を復元する．RGB 画像からカメラの位置姿勢を推定するためには，2.1 節で述べた特徴点に基づく手法 [13, 14] や画素値に基づく手法 [15, 16] などが利用される．

密な 3 次元点群の推定において，Bleyer ら [30] は，PatchMatch [32, 33] に基づいて 2 つの RGB 画像中の各画素の対応を求め，その対応とカメラ位置姿勢からデプス画像を推定し，密な 3 次元点群を求める手法を提案した．また，Furukawa ら [9] は，各 RGB 画像中から抽出した密な特徴点に対してパッチを定義し，各特徴点に対して他の RGB 画像でのそのパッチの見えが類似する特徴点対応をエッジポラ線付近から探索することで，密な 3 次元点を推定する手法を提案した．

密な 3 次元点群から 3 次元形状を復元する手法として，Curless ら [34] はカメラ

の位置に近い方を負，遠い方を正とした符号付き距離場に基づいて，3次元点群からボクセルモデルとして3次元形状を復元する手法を提案した．また，Kazhdanら [35] は，ポアソン方程式に基づいて法線付きの3次元点群からメッシュモデルとして3次元形状を復元する手法を提案した．

デプス画像を用いた手法では，はじめに複数位置で撮影したデプス画像から求めた3次元点群間の位置合わせを行う．その結果をもとに，位置合わせした3次元点群から3次元形状を復元する．

3次元点群の位置合わせを行う手法には，ICP アルゴリズム [5] がある．ICP アルゴリズム [5] では，2つの3次元点群に対して各3次元点間の距離や法線方向に基づいて各3次元点を対応付けし，対応点間の距離が近づくように回転・移動を表すパラメータを推定する．この対応付けとパラメータ推定を交互に繰り返すことにより3次元点群の位置合わせを行う．このアルゴリズムでは，対応付けの曖昧さによって位置合わせの結果がパラメータの初期値に依存するため，位置合わせの安定性が低い．これに対して，Endresら [36] はRGB-Dカメラを用いてデプス画像と同時にRGB画像を取得し，RGB画像間の特徴点对応 [17–19] を用いて明示的に対応付けを行うことで，対応付けの曖昧さを緩和する手法を提案した．

位置合わせした3次元点群からの3次元形状の復元には，RGB画像を用いた手法と同様の方法 [34, 35] が利用できる．なお，剛体物体の場合には，対象物は時間経過によって変形しないため，複数のカメラを用いた撮影と単一のカメラでの移動撮影のどちらでも同様の手法が適用できる．

2.3 非剛体物体に対する全周の3次元形状および動きの復元手法

非剛体物体の場合において，複数のカメラを用いた手法では対象物の全周を同時に撮影できるが，対象物は時間経過によって変形するため，単一のカメラでの移動撮影では各時刻における対象物の一部しか撮影できない．したがって，非剛体物体に対する全周の3次元形状および動きの復元手法は，複数のカメラを用いた手法 [37–41] と単一のカメラを用いた手法 [6, 7, 42, 43] に大別できる．

複数のカメラを用いた手法は，各時刻における対象物の全周を観測できるため，剛体に対する3次元形状の復元手法に基づいて各時刻での全周の3次元形状の

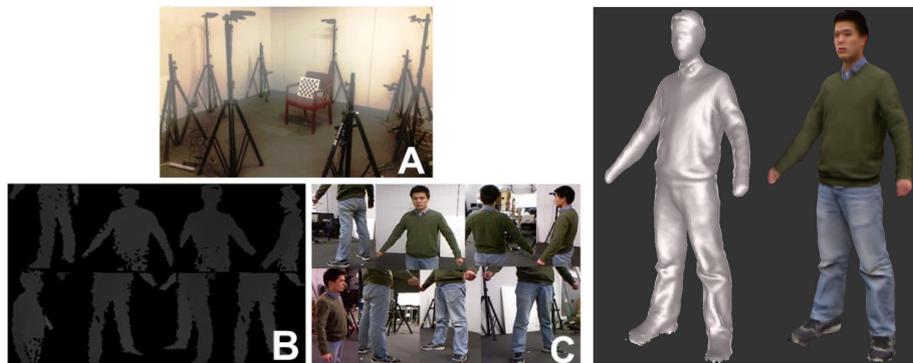


図 8: 複数の RGB-D カメラを用いた手法により復元された非剛体物体の 3 次元形状の例 [40]

復元できる。このような手法として，Starck ら [37]，De Aguiar ら [38]，Vlasic ら [39] は 8 台の RGB カメラを，Dou ら [40,41] は 8 台の RGB-D カメラを用いて様々な方向から対象物を同時に撮影することで，各時刻での 3 次元形状を復元する手法を提案している (図 8)。Starck らの手法 [37] では，各時刻での 8 つの RGB 画像からステレオ法と視体積交差法を用いて各時刻での 3 次元形状を復元する。De Aguiar ら [38] および Vlasic ら [39] の手法では，事前に作成した対象物の 3 次元形状テンプレートを各時刻での 8 つの RGB 画像へ投影し，画像中における対象物のシルエットとの重複率が最大となるように 3 次元形状テンプレートを変形することで，各時刻での 3 次元形状を復元する。Dou らの手法 [40,41] では，8 つのデプス画像から 3 次元点群を求め，1 時刻前までの 3 次元点群を融合した 3 次元形状と現時刻での 3 次元点群との非剛体位置合わせを行い，その結果をもとに現時刻での 3 次元点群を 1 時刻前の 3 次元形状に融合することで，現時刻の 3 次元形状を復元する。

これらの手法では，対象物の全周を同時に撮影できるため，全周の真の 3 次元形状を復元できるが，複数のカメラを設置するためのコストが高いことや複数のカメラ間のキャリブレーションが必要なことから，一般のユーザが使用するには適していない。

単一のカメラを用いた手法は，各時刻において対象物の一部しか観測できないため，各時刻での未観測部分を別の時刻での観測をもとに推定し，各時刻での全

周の3次元形状を復元する。単一のカメラを用いた手法において、受動的な手法では一般に同じ時刻において少なくとも2つ以上の位置での観測が必要となるため、単一のRGBカメラを用いた非剛体物体の3次元形状は困難である。一方、能動的な手法では単一の位置での計測で対象物の3次元形状を復元できるため、単一のデプスカメラを用いて非剛体物体の3次元形状を復元できる。

未観測部分を推定する手法として、事前知識として関節モデルを用いた手法 [42, 43] が提案されている。これらの手法では、対象物は複数の剛体物体からなる特定の関節モデルで表現されるものとし、関節モデルに従って3次元点群の位置合わせを行い、対象物の全周の3次元形状および動きを復元する。関節モデルは比較的少ないパラメータで表現できるため、関節モデルを用いた手法では安定した位置合わせが可能であるが、あらかじめ対象物の動きを表現できる関節モデルを対象物ごとに与える必要があるため、一般性に欠ける。

事前知識を用いない単純な方法として、デプス画像から求めた3次元点群間を非剛体ICPアルゴリズム [3, 4] を用いて逐次的に位置合わせすることで、非剛体物体の全周の3次元形状を復元する手法が考えられる。このアルゴリズムは、2つの3次元点群に対して各3次元点間の距離や法線方向に基づいて各3次元点を対応付けし、対応点間の距離が近づくように移動・変形を表すパラメータを推定する。この対応付けとパラメータ推定を交互に繰り返すことにより3次元点群の位置合わせを行う。しかし、非剛体ICPアルゴリズムは、剛体ICPアルゴリズム [5] と比べて推定すべきパラメータの数が多く、対応付けの曖昧性が大きいことや、一般に非凸最適化問題として定義されることから、位置合わせの結果がパラメータの初期値に大きく依存し、安定した位置合わせが困難であるという問題がある。また、逐次的な位置合わせによりパラメータの推定誤差が蓄積するという問題がある。

蓄積誤差を抑制するために、Zollhöferら [6] やLiら [7] は、対象物の全周形状を網羅した3次元形状テンプレートを用いてデプス画像やRGB-D画像から非剛体物体の全周の3次元形状および動きを復元する手法を提案した(図9)。これらの手法では、はじめに対象物が変形しない状態で、剛体に対する3次元形状復元手法を用いて3次元形状テンプレートを生成する。続いて、生成した3次元形状

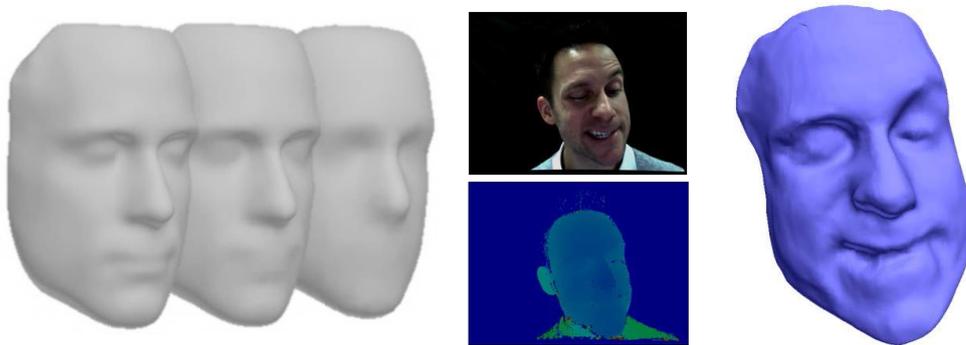


図 9: 単一の RGB-D カメラを用いた手法における 3 次元形状テンプレートと復元された非剛体物体の 3 次元形状の例 [6]

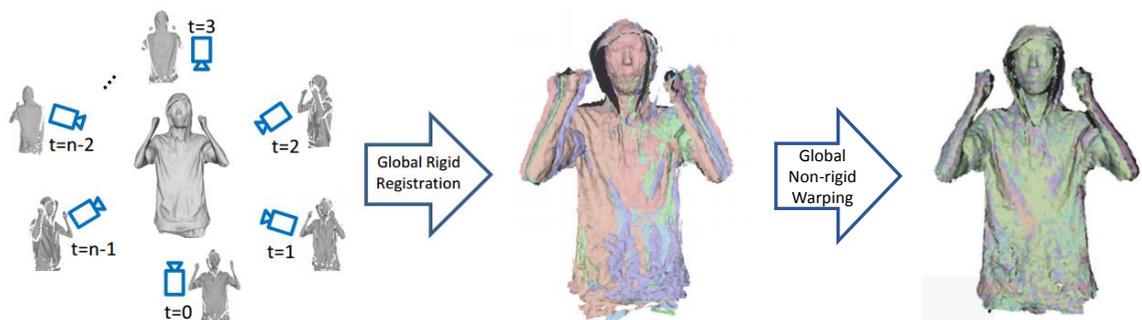


図 10: 対象物の小さな変形を許容する 3 次元形状復元 [44]

テンプレートを，移動・変形中の対象物を撮影した時系列の 3 次元点群に一致するように非剛体 ICP アルゴリズム [3, 4] を用いて変形することで，各時刻での対象物の動きを復元する．3 次元形状テンプレートは対象物を撮影した画像から生成できるため，これらの手法では事前知識を必要とせず，様々な非剛体物体に対して 3 次元形状および動きを復元できる．しかし，全周の 3 次元形状テンプレートを生成するためには，対象物が変形しない状態で全周撮影する必要があるため，対象となる人や動物にとって負担が大きいという問題がある．

また，Zeng ら [44] は，対象物の小さな変形を許容する 3 次元形状テンプレートの生成手法を提案した (図 10)．この手法では，はじめに剛体運動を仮定して 3

次元点群の位置合わせを行い、対象物の変形に起因する位置合わせ誤差を非剛体 ICP アルゴリズム [3,4] を用いて補正することで、3次元形状テンプレートを生成する。しかし、非剛体 ICP アルゴリズムは位置合わせの結果がパラメータの初期値に大きく依存するため、対象物が大きく変形する場合には適用が困難である。

2.4 本研究の位置づけと方針

本研究では非剛体物体の3次元形状および動きの復元に向け、移動・変形中の非剛体物体を単一の RGB-D カメラで全周撮影した RGB-D 画像列から3次元形状テンプレートを生成することを目的とする。前述した通り、非剛体物体の3次元形状および動きの復元には、複数のカメラを用いた手法 [37-41] と単一のカメラを用いた手法 [6, 7, 42-44] があり、複数のカメラを用いた手法では設置コストやカメラ間のキャリブレーションの面から、一般のユーザが使用するには適していないため、本研究では単一のカメラを用いた手法を対象とする。また、単一の RGB カメラを用いた従来研究では非剛体物体の3次元形状の復元が難しいため、本研究では単一の RGB-D カメラを用いて非剛体物体の3次元形状を復元する。単一の RGB-D カメラを用いた手法の中でも、実物体を計測して生成可能な3次元形状テンプレートを対象とする。従来研究 [6, 7, 44] の3次元形状テンプレート生成では、対象物が変形しない状態または小さく変形する状態で全周撮影する必要があるため、3次元形状テンプレート生成と非剛体物体の3次元形状復元にはそれぞれ別々の計測が必要となる。本研究では、移動・変形中の対象物を全周撮影した画像から3次元テンプレートを生成する手法を提案することで、同一の計測から3次元形状テンプレート生成と非剛体物体の3次元形状復元を行うことを目指す。

提案手法では、撮影した RGB-D 画像から得られる3次元点群は、3次元形状テンプレート上の点(テンプレート点)を局所的なアフィン変換で変形したもので表現されると仮定し、非剛体 ICP アルゴリズム [3,4] に基づいてテンプレート点を求める。また、RGB-D カメラを用いてデプス画像を撮影する際に、3次元点群間の対応付けに対して重要な手がかりとなる RGB 画像も合わせて取得可能であることに着目し、RGB 画像から得られる時間的に連続する画像間での追跡点を

利用することで、対応付けの曖昧さを緩和し、安定した非剛体位置合わせを実現するため。また、推定するパラメータの数を段階的に上げることで最適化問題の非凸性を緩和する。更に、蓄積誤差を抑制するために、RGB 画像から得られる時間的に連続しない画像間での特徴点对応を利用する。

3. 点追跡および特徴点对応を用いた3次元形状テンプレート生成

本章では、まず点追跡および特徴点对応を用いた3次元形状テンプレートの生成手法について概説したのち、各要素について詳述する。

3.1 提案手法の概要

提案手法の概要について図 11 に示す。提案手法では、まず移動・変形中の対象物を RGB-D カメラで全周撮影し、時系列の RGB-D 画像列を取得する。次に、非剛体 ICP アルゴリズムの問題点であった位置合わせの初期値依存性を低減するために、RGB 画像に対する点追跡に基づいて時間的に連続した一連の3次元追跡点を求める。具体的には、RGB 画像列に対してオプティカルフローに基づく点追跡手法 [22] を適用することにより得られる2次元追跡点 (図 12) を、RGB 画像とデプス画像の対応関係を利用してデプス画像へ投影し、これに基づいて2次元追跡点に対応する一連の3次元追跡点 $\mathcal{X}_i = \{\mathbf{x}_i^f | f = f_i^s, \dots, f_i^e\}$ を求める。ここで、 \mathbf{x}_i^f はフレーム f で観測された i 番目の3次元追跡点、 f_i^s および f_i^e はそれぞれ追跡点の開始フレームと終了フレームを表す。

また、オプティカルフローに基づく点追跡手法では、遮蔽により追跡点が失われた際に同じ点の再同定ができないため、点追跡手法で得られた3次元追跡点のみを利用した場合、位置合わせ誤差が蓄積される。提案手法ではさらに特徴点検出および局所特徴量に基づく特徴点の対応付けによって得られた2次元対応点 (図 13) に基づいて、時間的に連続しない3次元対応点群 \mathbf{y}_j^f を求め、位置合わせに利用することにより蓄積誤差を低減する。ここで、 \mathbf{y}_j^f はフレーム f で観測された j 番目の3次元対応点を表す。提案手法では、特徴点検出および局所特徴量に基づく対応付け手法として、SIFT [17] を用いる。

最後に、点追跡手法で得られた3次元追跡点、および SIFT で得られた3次元対応点から3次元形状テンプレートを生成する。提案手法では、点追跡手法で得られた連続するフレーム間での3次元追跡点を統一的に扱うために、一連の3次元追跡点 \mathcal{X}_i のそれぞれに対して、任意に設定した座標系上の単一の3次元

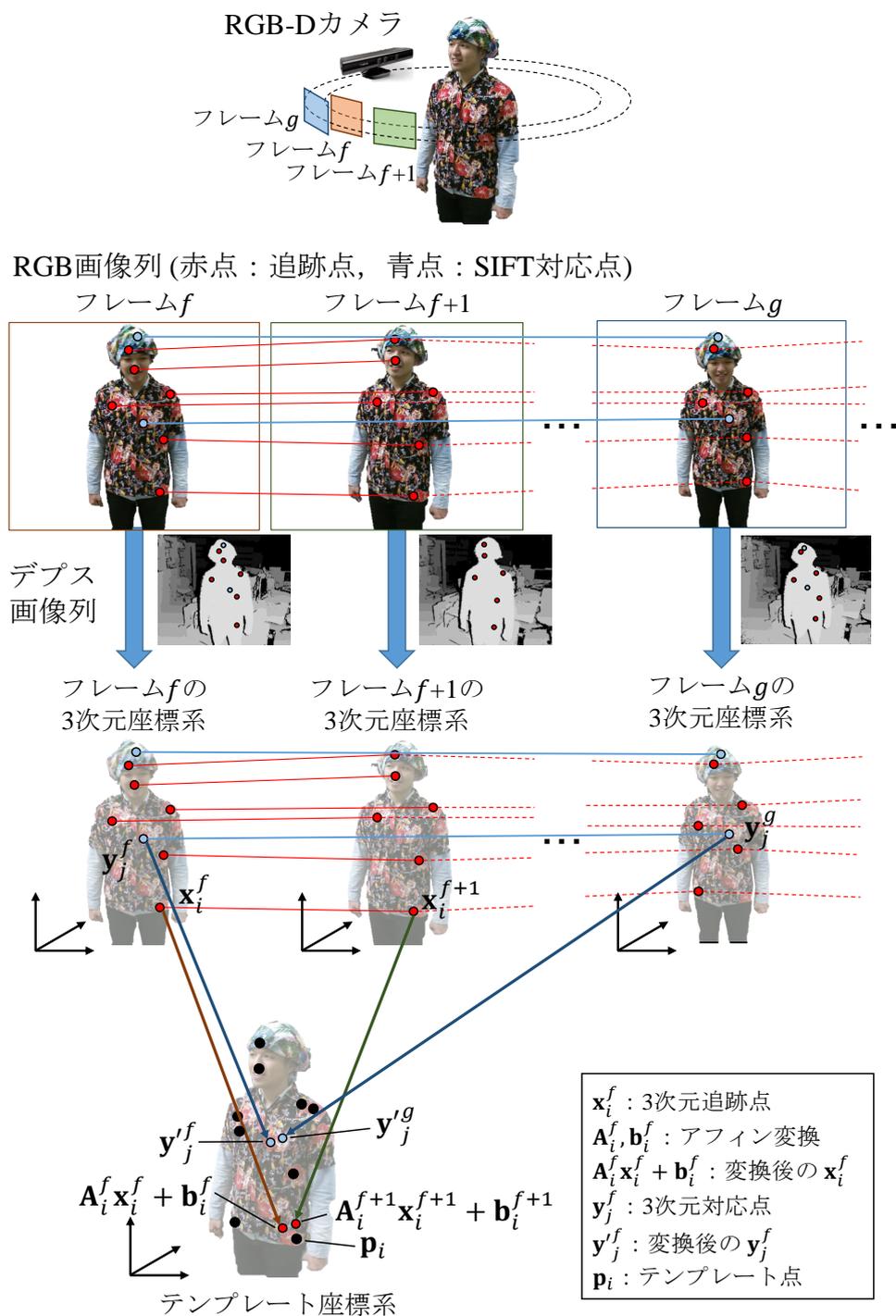


図 11: 提案手法の概要

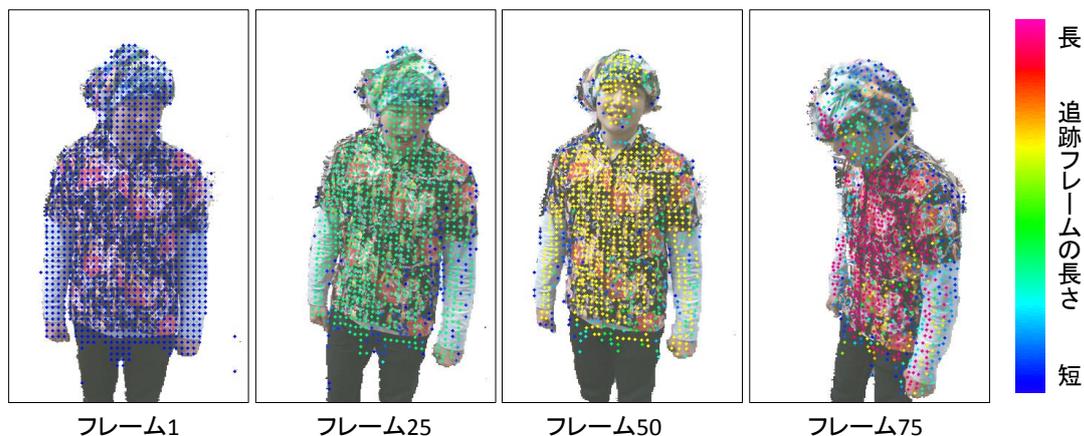


図 12: オプティカルフローに基づく 2次元追跡点



図 13: SIFT による 2次元対応点

点 \mathbf{p}_i を定義する．本稿では，この 3次元点 \mathbf{p}_i をテンプレート点と呼び，テンプレート点が定義される座標系をテンプレート座標系と呼ぶ．ここで，テンプレート点は RGB-D 画像列から得られた 3次元追跡点を任意のパラメータにより変換した点であるものとする．提案手法では，非剛体運動を記述する変換として，各 3次元追跡点 \mathbf{x}_i^f とテンプレート点 \mathbf{p}_i との間の変換を，局所的なアフィン変換 ($\mathbf{p}_i = \mathbf{A}_i^f \mathbf{x}_i^f + \mathbf{b}_i^f$) であるものと仮定する．この仮定を踏まえて， I 個のテンプレート点からなる集合 $\mathcal{P} = \{\mathbf{p}_i | i = 1, \dots, I\}$ ，および \mathbf{p}_i と \mathbf{x}_i^f の間のアフィン変換の集合 (線形成分 $\mathcal{A} = \{\mathbf{A}_i^f | i = 1, \dots, I, f = f_i^s, \dots, f_i^e\}$ と並進成分 $\mathcal{B} = \{\mathbf{b}_i^f | i = 1, \dots, I, f = f_i^s, \dots, f_i^e\}$) の関係を表す項，三次元対応点 \mathbf{y}_j^f に基づいてテンプレート点同士を対応付ける項，およびアフィン変換に関する正規化項

からなるエネルギー関数を，非線形最適化手法により最小化することにより推定したテンプレート点をもとに3次元形状テンプレートを生成する．提案手法では，3次元形状テンプレートの生成結果が望ましくない局所解へ陥ることを回避するために，推定するパラメータ数を段階的に上げてエネルギー関数を最小化する．

以上の方法を用いることで，提案手法では時間的に連続するRGB画像間の追跡点を用いた対応付けにより，安定した非剛体位置合わせを実現し，時間的に連続しないRGB画像間での特徴点对応により，蓄積誤差を抑制した位置合わせを実現する．また，多段階の最適化により，望ましくない局所解へ陥ることを回避した3次元形状テンプレートの生成を実現する．

3.2 追跡点および対応点の検出

本研究では，RGB画像列からフレーム間における点の対応付けを行うために，オプティカルフローに基づく点追跡手法 [22] を採用する．この手法は，長い時間に渡って正しく追跡でき，得られる点が比較的密であるため，対象物の表面に対して密に分布するテンプレート点を得られる．しかし，点追跡手法は時間的連続性に基づいて点を追跡するため，遮蔽などで失われた追跡点が再び出現した際にそれらの点を再同定できない．そこで，提案手法ではユーザが対象の非剛体物体の全周を1周以上撮影するものとして，局所特徴量に基づく対応点検出手法を併用し，時間的に連続しないフレーム間での3次元対応点に基づいて再出現したテンプレート点に対して対応付けを行う．

具体的には，まずオプティカルフローに基づく点追跡手法 [22] をRGB画像列に対して適用し，RGB画像列中の2次元追跡点を得る．得られた2次元追跡点をRGB画像と対応したデプス画像へ投影し，これに基づいてRGB画像列中の2次元追跡点に対応する3次元追跡点 \mathcal{X}_i を求める．次に，SIFT [17] を用いて局所特徴量に基づいて対応点を検出する．図14に示すように，RGB画像列の各フレームを注目フレーム f とし， f とその他の全てのフレームで対応点を検出し，その対応点数を数える．続いて，各注目フレームと対応するフレームを対応点数に基づいて選択する．ここでは，少数の信頼度の高い対応点を得るために，対応点数が閾値 T_C を超える時間的に連続したフレームの集合 \mathcal{S}_n を求め，それらの集合中

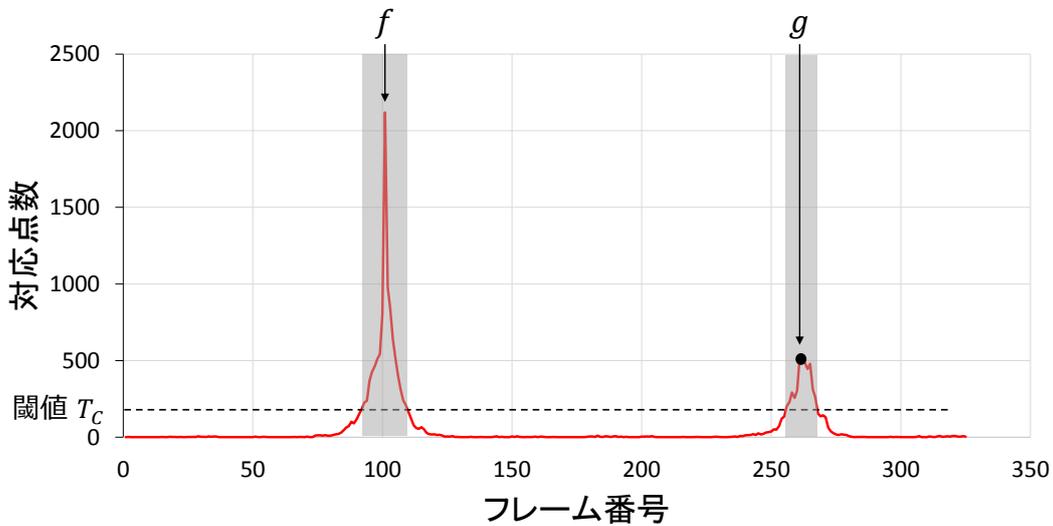


図 14: フレーム f と各フレームの対応点の数および最大対応点数フレーム g

において対応点数が最大となるフレーム g を，フレーム f との SIFT 対応フレーム対 (f, g) とする．ただし，2つのフレーム集合 S_n, S_m 間のフレーム数が T_W 以下の場合には S_n と S_m を連続したフレームの集合とみなす．図 14 では，連続したフレーム集合の中で対応点数が最大となるフレーム 261 が，注目フレーム 101 との SIFT 対応フレーム対をなす．

このような処理を全ての注目フレームに対して行い，SIFT 対応フレーム対の集合を \mathcal{G} を得る．得られた集合 \mathcal{G} の全ての SIFT 対応フレーム対 (f, g) において，RGB 画像上の対応点对に対応する 3次元点对対応点対 $(\mathbf{y}_j^f, \mathbf{y}_j^g)$ を，デプス画像を用いて求める．

3.3 3次元形状テンプレート生成のためのエネルギー関数の定義

提案手法では，テンプレート点と 3次元追跡点の位置合わせ誤差を表す項 E_F ，SIFT による 3次元対応点の位置合わせ誤差を表す項 E_C ，およびアフィン変換パラメータの正則化項 E_R, E_S により，次式のように定義されたエネルギー関数を最小化するテンプレート点の集合 \mathcal{P} およびアフィン変換の集合 (線形成分 \mathcal{A} と並

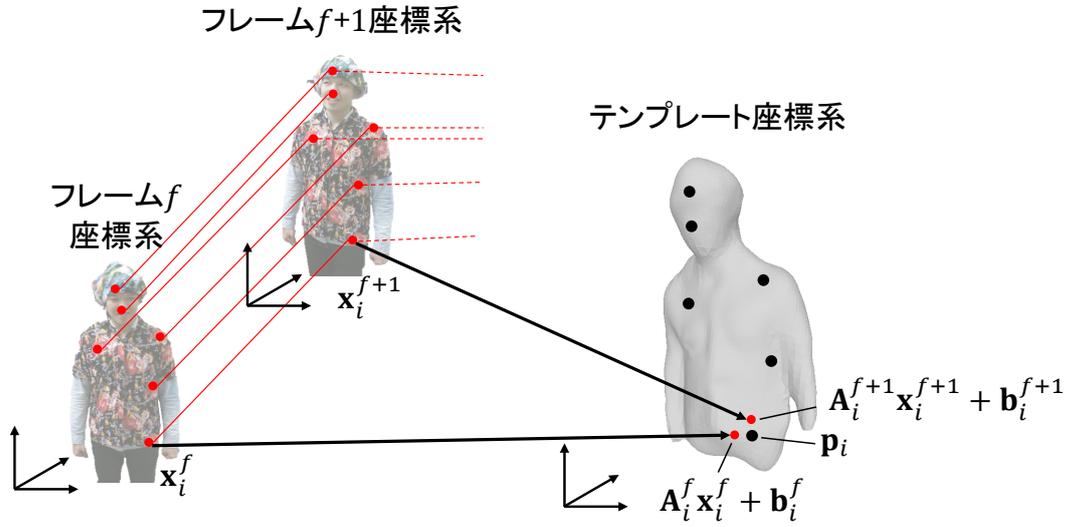


図 15: テンプレート点および3次元追跡点の位置合わせ誤差項 E_F

進成分 B) を推定する.

$$E(\mathcal{P}, \mathcal{A}, \mathcal{B}) = \alpha_F E_F(\mathcal{P}, \mathcal{A}, \mathcal{B}) + \alpha_C E_C(\mathcal{P}) + \alpha_R E_R(\mathcal{A}) + \alpha_S E_S(\mathcal{A}, \mathcal{B}) \quad (1)$$

ここで, $\alpha_F, \alpha_C, \alpha_R, \alpha_S$ は各項の重みパラメータである. 各項について以下に詳述する.

3.3.1 テンプレート点および3次元追跡点の位置合わせ誤差項 E_F

提案手法では, 3次元追跡点 \mathbf{x}_i^f と対応するテンプレート点 \mathbf{p}_i の変換が局所的アフィン変換 ($\mathbf{p}_i = \mathbf{A}_i^f \mathbf{x}_i^f + \mathbf{b}_i^f$) で記述されるものとし, テンプレート点と変換後の3次元追跡点との間の距離 (図 15) に関する項を次式により定義する.

$$E_F(\mathcal{P}, \mathcal{A}, \mathcal{B}) = \sum_{i=1}^I \sum_{f=f_i^s}^{f_i^e} \|\mathbf{p}_i - (\mathbf{A}_i^f \mathbf{x}_i^f + \mathbf{b}_i^f)\|_2^2 \quad (2)$$

ここで, E_F が小さいほどテンプレート点とアフィン変換後の追跡点が一致することを表す.

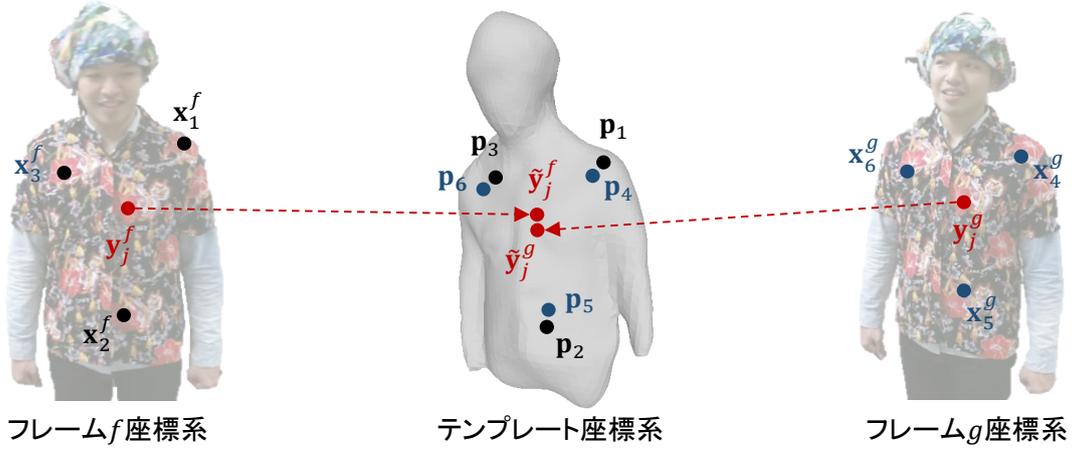


図 16: SIFT 対応点に基づくテンプレート点のゆるい対応付け

3.3.2 SIFT による 3 次元対応点の位置合わせ誤差項 E_C

追跡点が失われた際に追跡点を再同定できないため、位置合わせ誤差の蓄積により同一の追跡点に対応したテンプレート点同士の間には誤差が生じる。この誤差を低減するために、提案手法では 3.2 節で求めた SIFT 対応フレーム対 (f, g) における j 番目の SIFT 対応点 y_j^f および y_j^g をテンプレート座標系への変換後の点をそれぞれ \tilde{y}_j^f および \tilde{y}_j^g とする。これらの点は対象物上で同一の点を表すことから、これらに基づいてテンプレート点間のゆるい対応付けを行う (図 16)。提案手法では、SIFT 対応点にアフィン変換によるテンプレート座標系への変換を定義せず、対応点 y_j^f からテンプレート座標系へ変換した点 \tilde{y}_j^f は、 m 個の近傍追跡点 x_i^f と対応するテンプレート点 p_i の重み付き平均で表されるものとして、次式で与える。

$$\tilde{y}_j^f = \sum_{i \in \mathcal{M}(f,j)} w_{ji} p_i \quad (3)$$

ここで、 $\mathcal{M}(f, j)$ はフレーム f における追跡点のうち SIFT 対応点 y_j^f の近傍 m 点の添字集合、 w_{ji} は SIFT 対応点 y_j^f と近傍点 x_i^f との距離によって決定される重みである。提案手法では、Li らの手法 [7] を参考に重み w_{ji} を次のように定義

する.

$$w_{ji} = \frac{(1 - \|\mathbf{y}_j^f - \mathbf{x}_i^f\|_2^2 / \gamma^2)^3}{\sum_{k \in \mathcal{M}(f,j)} (1 - \|\mathbf{y}_j^f - \mathbf{x}_k^f\|_2^2 / \gamma^2)^3} \quad (4)$$

ここで, γ は SIFT 対応点 \mathbf{y}_j^f と $m+1$ 番目に近い追跡点との距離を表す. 以上より, テンプレート座標系上での SIFT 対応点の距離に関する項を次式により定義する.

$$\begin{aligned} E_C(\mathcal{P}) &= \sum_{(f,g) \in \mathcal{G}} \sum_j \|\tilde{\mathbf{y}}_j^f - \tilde{\mathbf{y}}_j^g\|_2^2 \\ &= \sum_{(f,g) \in \mathcal{G}} \sum_j \left\| \sum_{k \in \mathcal{M}(f,j)} w_{jk} \mathbf{p}_k - \sum_{l \in \mathcal{M}(g,j)} w_{jl} \mathbf{p}_l \right\|_2^2 \end{aligned} \quad (5)$$

ここで, E_C が小さいほど変換後の対応点間の距離が小さくなる. ただし, \mathcal{G} は 3.2 節で求めた SIFT 対応フレーム対の集合であり, j に関する和はフレーム対 (f,g) に対して得られる全ての SIFT 対応点で計算される.

3.3.3 正則化項 E_R および E_S

本手法では Li らの手法 [7] と同様に, 正則化項として以下の仮定を用いる.

- 対象物上の多くの点の動きは剛体変換に近い変換で記述される. つまり, アフィン変換の線形変換成分 \mathbf{A}_i^f は正規直交行列に近いものである.
- 近傍点の動きの多くは同一に近い変換によって記述される. つまり, ある点 \mathbf{x}_i^f をその点に対して定義されたアフィン変換 $(\mathbf{A}_i^f, \mathbf{b}_i^f)$, およびその点の近傍点 \mathbf{x}_j^f に対して定義されたアフィン変換 $(\mathbf{A}_j^f, \mathbf{b}_j^f)$ で変換した点はテンプレート座標系において近い位置に存在する (図 17).

これらを定式化すると次式となる.

$$E_R(\mathcal{A}) = \sum_{i=1}^I \sum_{f=f_i^s}^{f_i^e} \|(\mathbf{A}_i^f)^T \mathbf{A}_i^f - \mathbf{I}\|_F^2 \quad (6)$$

$$E_S(\mathcal{A}, \mathcal{B}) = \sum_{i=1}^I \sum_{f=f_i^s}^{f_i^e} \sum_{j \in \mathcal{N}(f,i)} \|\mathbf{A}_i^f \mathbf{x}_i^f + \mathbf{b}_i^f - (\mathbf{A}_j^f \mathbf{x}_i^f + \mathbf{b}_j^f)\|_2^2 \quad (7)$$

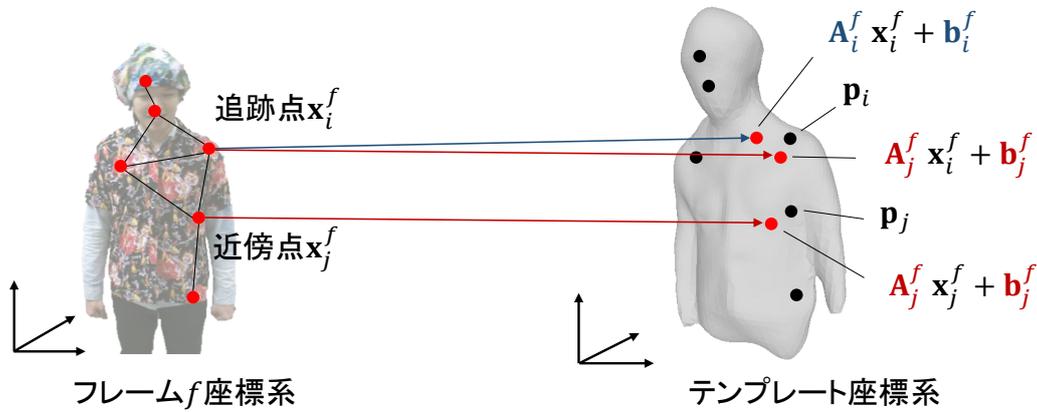


図 17: 近傍の点で定義されたアフィン変換の類似性

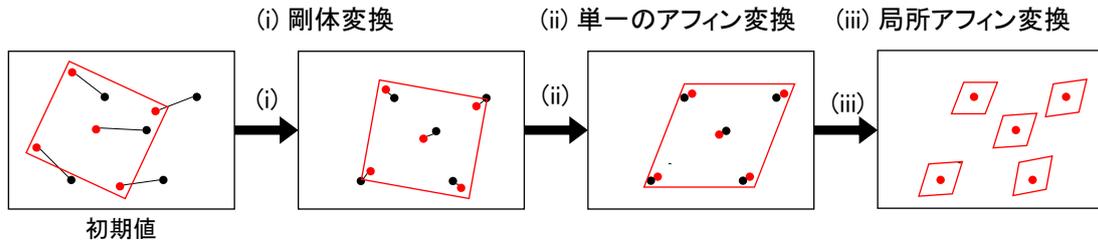


図 18: 3つのステップの最適化による位置合わせのイメージ

ここで、 E_R が小さいほど線形変換 A_i^f は正規直交行列に近く、 E_S が小さいほど点 $A_i^f x_i^f + b_i^f$ と $A_j^f x_i^f + b_j^f$ との距離が近い。ただし、 $\mathcal{N}(f, i)$ はフレーム f における i 番目の追跡点 x_i^f の n 個の近傍点の添字集合、 $\|\bullet\|_F$ はフロベニウスノルムである。

3.4 エネルギー関数の最適化

式 (1) において、 E_R は A_i^f の各要素に関する 4 次式であるため、式 (1) の最小化は非凸最適化問題であり、一般にその解は推定するパラメータ初期値に依存する。そこで提案手法では、次に示す 3 つのステップの最適化を行う (図 18)。

- (i) 追跡点に基づく対応点を用いた剛体変換による追跡点群の逐次的位置合わせ

(ii) 得られた剛体変換を初期値とし，同一フレーム内の点群に対する単一のアフィン変換を仮定した式 (1) の最適化

(iii) 各フレームの単一のアフィン変換を初期値として，各追跡点に対して定義された局所アフィン変換による変形を考慮した式 (1) の最適化

ステップ (i) では，対象物の剛体運動を仮定し，初期フレームの点群を基準として連続するフレーム間の追跡点から逐次的に剛体位置合わせを行い，剛体変換パラメータ (回転行列および並進ベクトル) を特異値分解に基づく手法 [45] により求める．これらの剛体変換パラメータから，初期フレームを基準とした剛体運動パラメータを漸化的に求める．ここでは，フレーム f における回転行列を \mathbf{R}^f ，並進ベクトルを \mathbf{t}^f とする (ただし，初期フレームでは $\mathbf{R}^1 = \mathbf{I}$ ， $\mathbf{t}^1 = \mathbf{0}$)．

ステップ (ii) では，ステップ (i) で得られた剛体運動パラメータ (\mathbf{R}^f ， \mathbf{t}^f) を初期値として，式 (1) の最小化によりテンプレート点 \mathbf{p}_i の初期値を得る．ここで，同一フレーム内の点が単一のアフィン変換にしたがって運動するものと仮定して，式 (1) に対して下記の制約を課す．

$$\forall i, \quad \mathbf{A}_i^f = \mathbf{A}^f, \mathbf{b}_i^f = \mathbf{b}^f \quad (8)$$

この制約の下で， E_S は常に 0 となることに注意して，式 (1) は次式に変形できる．

$$E'(\mathcal{P}, \mathcal{A}, \mathcal{B}) = \alpha'_F E'_F + \alpha'_C E_C + \alpha'_R E'_R \quad (9)$$

ただし，

$$E'_F = \sum_{i=1}^I \sum_{f=f_i^s}^{f_i^e} \|\mathbf{p}_i - (\mathbf{A}^f \mathbf{x}_i^f + \mathbf{b}^f)\|_2^2 \quad (10)$$

$$E'_R = \sum_f \|(\mathbf{A}^f)^T \mathbf{A}^f - \mathbf{I}\|_F^2 \quad (11)$$

ここで， α'_F ， α'_R ， α'_C は重みパラメータを表す．提案手法では，最急降下法を用いて式 (9) を最小化する．

ステップ (iii) の最適化では，ステップ (ii) で得られたテンプレート点および各フレームの単一のアフィン変換を初期値として，式 (1) を最小化するテンプレート点 \mathbf{p}_i および局所アフィン変換 (\mathbf{A}_i^f ， \mathbf{b}_i^f) を推定する．ここでは，ステップ (ii) と同様に最急降下法を用いて式 (1) を最小化する．

4. 実験

本実験では、移動・変形中の非剛体実物体を全周撮影した RGB-D 画像列から、提案手法によってテンプレート点を推定し、3次元形状テンプレートが生成可能であることを確認する。また、ステップ(ii)およびステップ(iii)での最適化に対して、テンプレート点の推定結果の考察、推定するパラメータの初期値の違いによる結果の比較、およびテンプレート点と追跡点間の位置合わせ精度の評価を行うことにより、提案手法の有効性を示す。

4.1 実験環境とデータセット

本実験では、RGB-D カメラ (Microsoft Kinect v2) を用いて、上体と頭を動かしている人物、および腕を動かしている人物、およびほとんど変形していない人物の周りを、2周撮影して得られた RGB-D 画像列からなるデータセット (DS1, DS2, DS3) に対して3次元形状テンプレートを生成する実験を行った。DS1, DS2 および DS3 について、解像度とフレーム数を表1に、RGB 画像およびデプス画像の例を図19, 図20 および図21に示す。また、本実験で利用したパラメータを表2にまとめる。

得られた RGB-D 画像には背景領域が含まれるため、本実験では人物の含まれ得るデプスの範囲を手動で設定することで、デプス画像の人物領域のみを抽出し、その領域に基づいて RGB 画像についても人物領域のみを抽出した。

表 1: DS1, DS2 および DS3 の仕様

	DS1	DS2	DS3
RGB-D 画像の解像度	960 × 540		
フレーム数	350	325	330

表 2: 実験で用いたパラメータ

	パラメータ	値
対応点検出	対応点数の閾値 T_C	80
	フレーム間の閾値 T_W	15
ステップ (i)	回転行列 \mathbf{R}^f の初期値	単位行列 \mathbf{I}
	並進ベクトル \mathbf{t}^f の初期値	零ベクトル $\mathbf{0}$
ステップ (ii)	テンプレート点 \mathbf{p}_i の初期値	零ベクトル $\mathbf{0}$
	重み α'_f	1.0
	重み α'_c	1.0
	重み α'_r	100.0
	近傍追跡点の添字集合 $\mathcal{M}(f, j)$ の個数 m	4
ステップ (iii)	重み α_f	0.1
	重み α_c	1.0
	重み α_r	1.0
	重み α_s	1.0
	近傍追跡点の添字集合 $\mathcal{N}(f, i)$ の個数 n	4

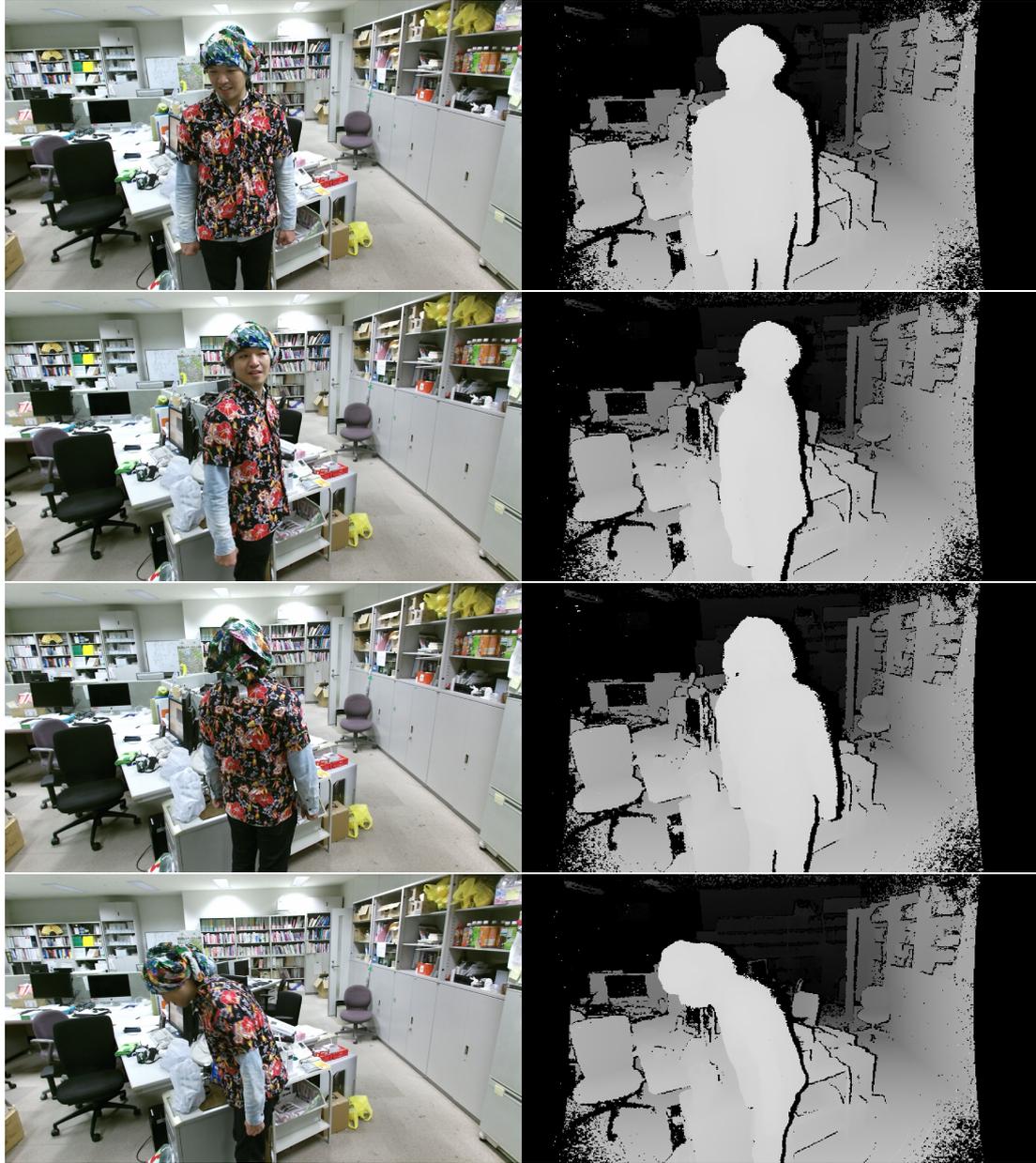


図 19: DS1 の RGB 画像およびデプス画像の例



図 20: DS2 の RGB 画像およびデプス画像の例



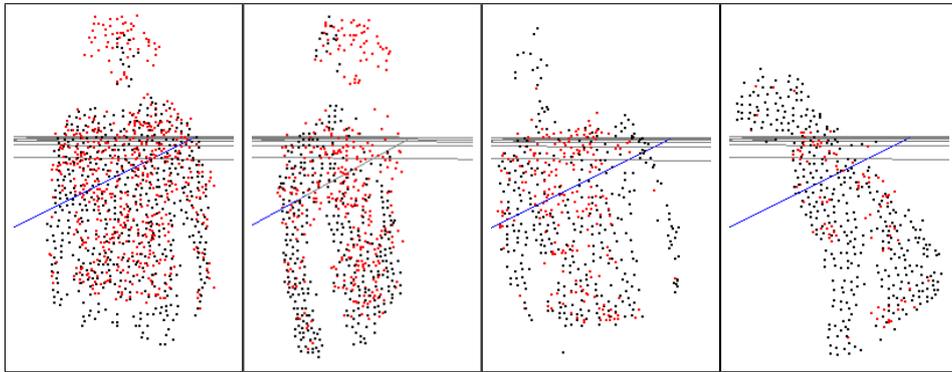
図 21: DS3 の RGB 画像およびデプス画像の例

4.2 3次元形状テンプレートの生成結果と考察

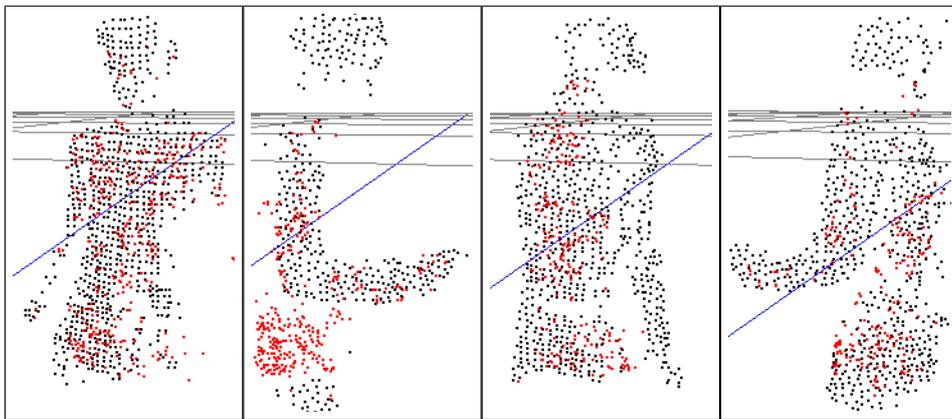
オプティカルフローに基づく点追跡 [22] および SIFT [17] による特徴点マッチングを RGB 画像に対して適用し, RGB 画像の人物領域内に含まれる追跡点, および SIFT 対応点を抽出した. それぞれのデータセットについて, 追跡点および SIFT 対応点から RGB 画像とデプス画像との対応関係を利用して求めた 3次元追跡点および 3次元対応点の例を図 22 に示す.

DS1, DS2 および DS3 のそれぞれから求めた 3次元追跡点および 3次元対応点に対して提案手法を適用することにより, テンプレート点を推定した. 以下の項目について実験結果を示し, 結果に対する考察を述べる.

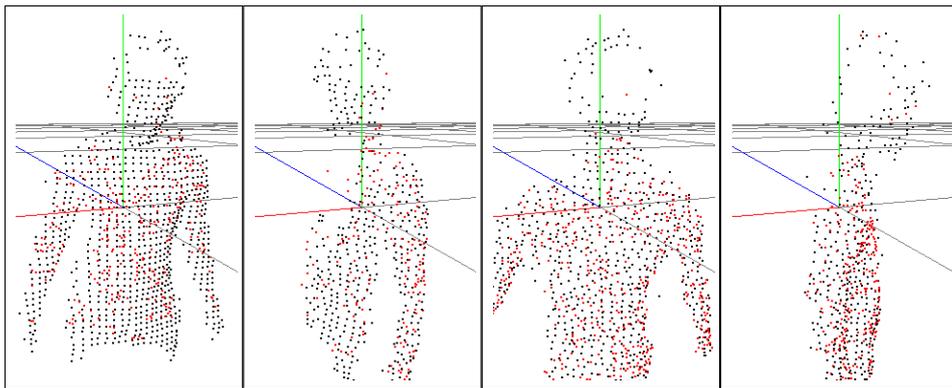
- A. ステップ (ii) のテンプレート点の推定結果
- B. ステップ (iii) のテンプレート点の推定結果
- C. 推定したテンプレート点を用いた 3次元形状テンプレートの生成結果
- D. ステップ (i) の有無によるステップ (ii) のテンプレート点の推定結果の比較
- E. ステップ (ii) の有無によるステップ (iii) のテンプレート点の推定結果の比較
- F. ステップ (ii) およびステップ (iii) における推定精度評価



(a) DS1



(b) DS2



(c) DS3

図 22: 3次元点追跡 (黒) および 3次元対応点 (赤) の例

A. ステップ (ii) のテンプレート点の推定結果

DS1, DS2 および DS3 において, ステップ (ii) でパラメータの初期値としてステップ (i) の剛体変換を用いた場合の, テンプレート点の推定結果を図 23 に示す. ステップ (ii) では同一フレーム内の変形を単一のアフィン変換として仮定しているため, DS1 での頭部の局所的な変形, および DS2 での両腕の局所的な変形を表現できず, テンプレート点が広がっている. また, DS3 はほぼ変形していないため, テンプレート点が広がらずに推定された.

B. ステップ (iii) のテンプレート点の推定結果

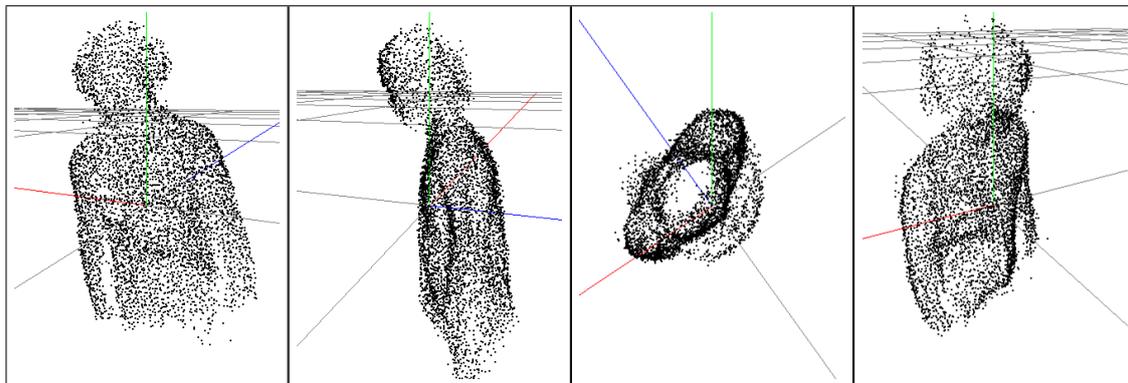
ステップ (iii) において, ステップ (ii) で得られたテンプレート点および各フレームの単一のアフィン変換を初期値として, 式 (1) の最小化により推定したテンプレート点を図 24 に示す. また, 図 25 にステップ (iii) およびステップ (ii) で推定したテンプレート点の比較結果を示す.

DS1 では, SIFT 対応点によるテンプレート点のゆるい対応付けの効果が正しく働き, ステップ (ii) で得られたテンプレート点に比べて頭の形状が小さくなった (図 25(a), 赤丸). 一方, ステップ (ii) のテンプレート点と比べて首の部分での形状のひずみが生じた (図 25(a), 青枠). この理由として, 1) SIFT 対応点によるテンプレート点のゆるい対応付けにおいて, 肩または頭の対応点の近傍に他方の追跡点が含まれており, さらにステップ (iii) において局所的なアフィン変換を許容したために, テンプレート点が E_C により対応点に引きずられた, もしくは 2) 対応点の誤対応があった, の二つが考えられる.

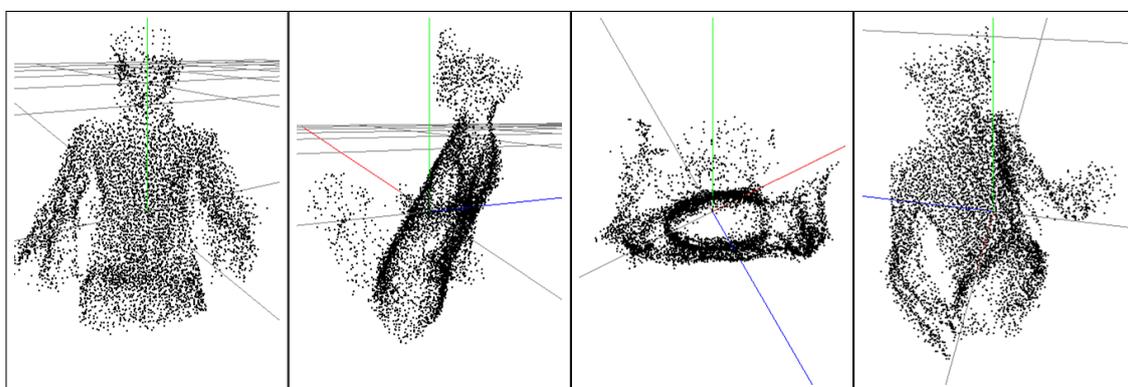
DS2 では, ステップ (ii) のテンプレート点と比較して右腕の部分で 2 つに分かれていたテンプレート点が近づいている (図 25(b), 赤枠). しかし, 腕に観測される対応点は比較的数量が少なく, 他の部分に比べてゆるい対応付けが一部にしか働かないため, ステップ (iii) において大きく移動せずに残るテンプレート点が多い (図 25(b), 青丸).

DS3 では, ほぼ同じような結果となった.

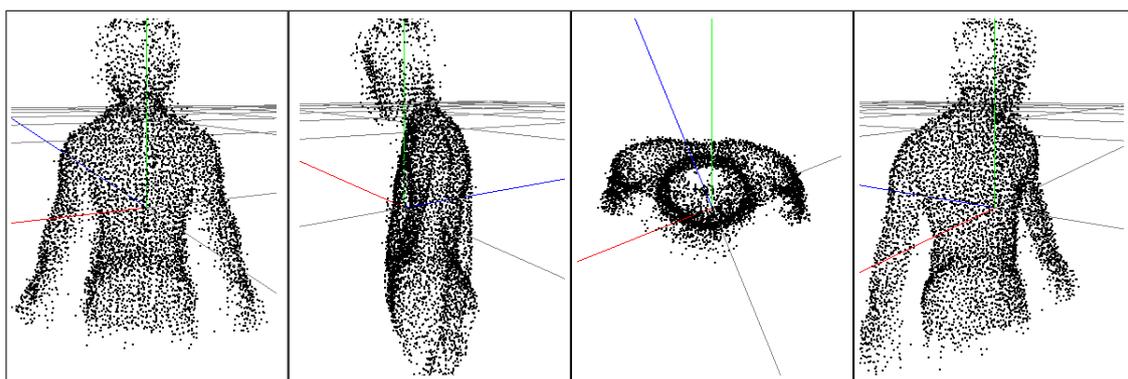
これら結果から, 生成される 3 次元形状テンプレートの質を向上させるためには, 誤対応の除去や偏りのない対応付けが必要である. また, テンプレート点のゆるい対応付けにおいて, 例えばデプス画像上での形状を考慮したジオデシック



(a) DS1

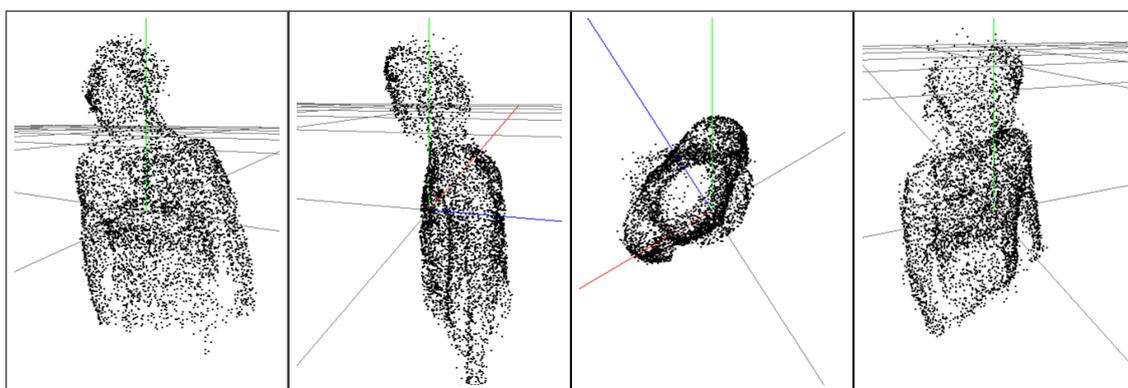


(b) DS2

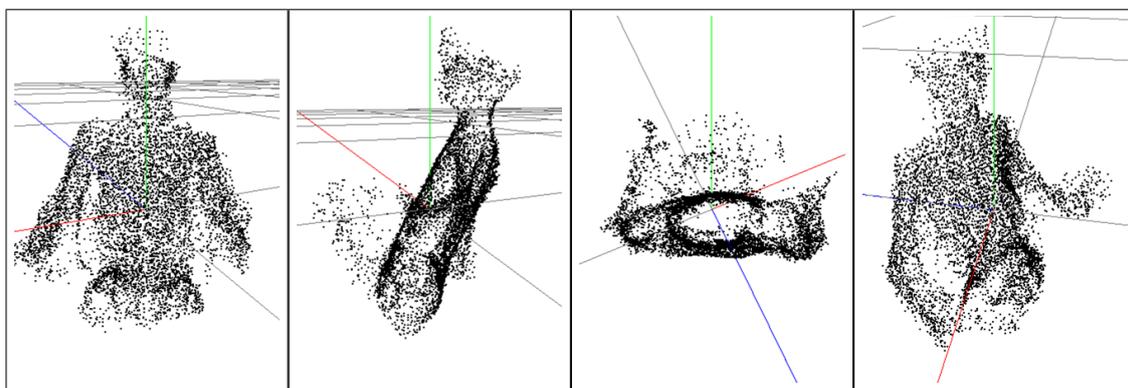


(c) DS3

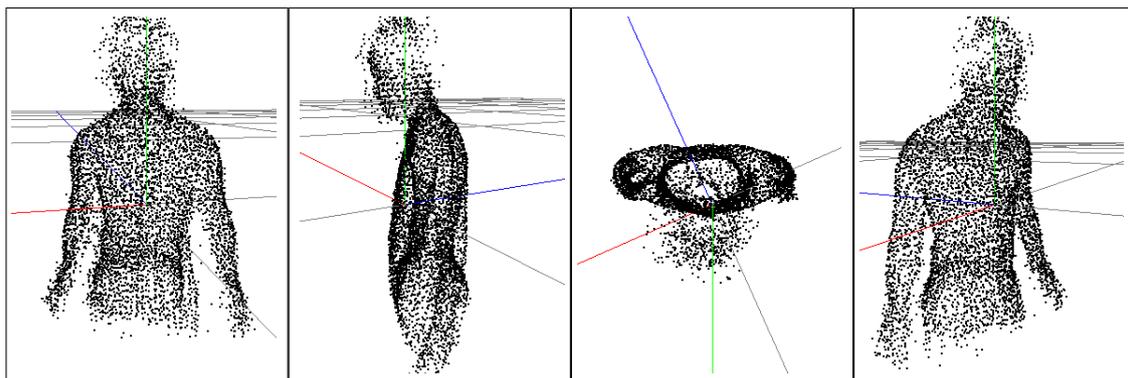
図 23: ステップ (ii) で推定したテンプレート点



(a) DS1

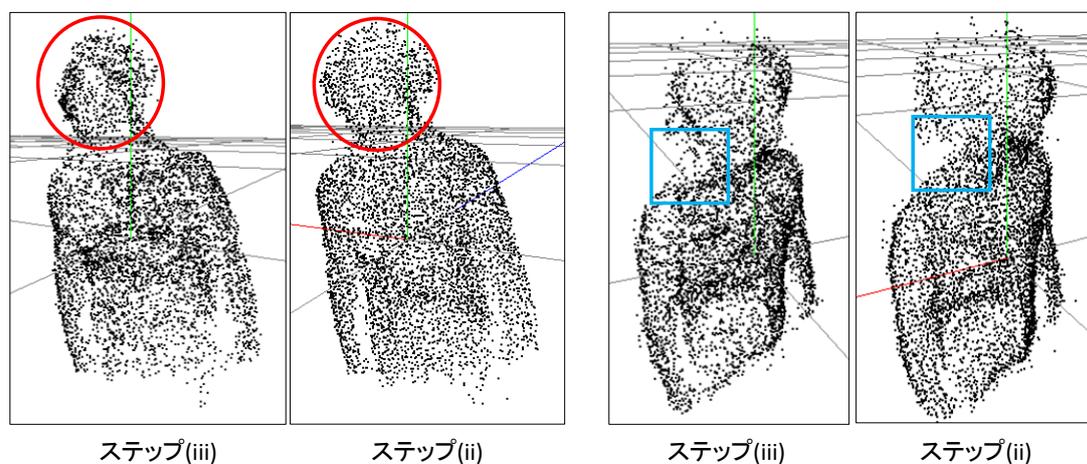


(b) DS2

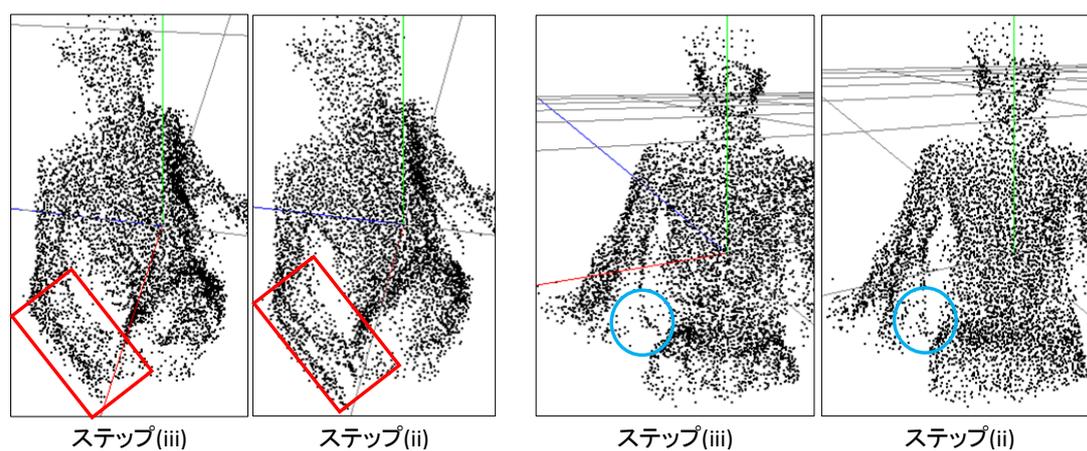


(c) DS3

図 24: ステップ (iii) で推定したテンプレート点



(a) DS1



(b) DS2

図 25: ステップ (iii) およびステップ (ii) で推定したテンプレート点の比較

距離による重みの利用することで、肩と頭などのユークリッド空間上で近傍のテンプレート点同士の影響を抑えることなども考えられる。

C. 推定したテンプレート点を用いた 3 次元形状テンプレートの生成結果

ステップ (iii) で推定したテンプレート点を用いて、ポアソン方程式に基づくメッシュ化手法 [35] を用いて 3 次元形状テンプレートを生成した結果を図 26 に示す。

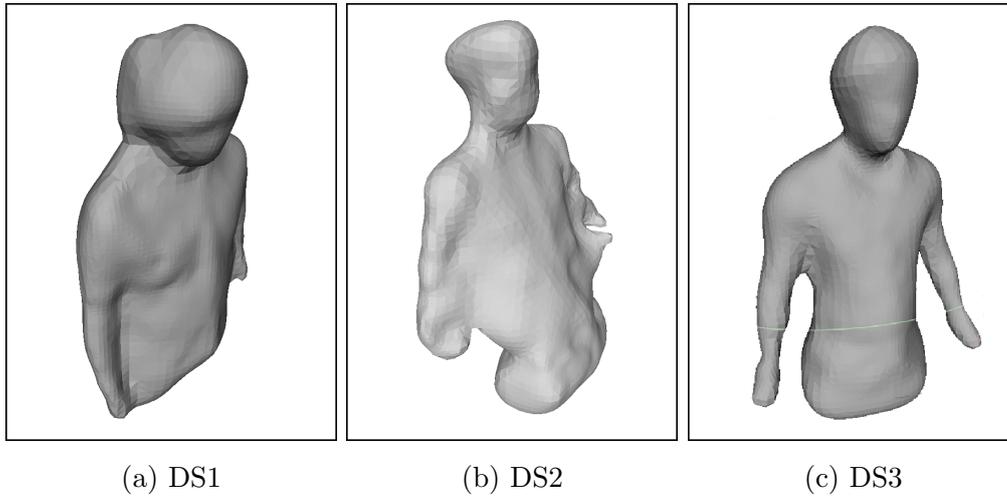
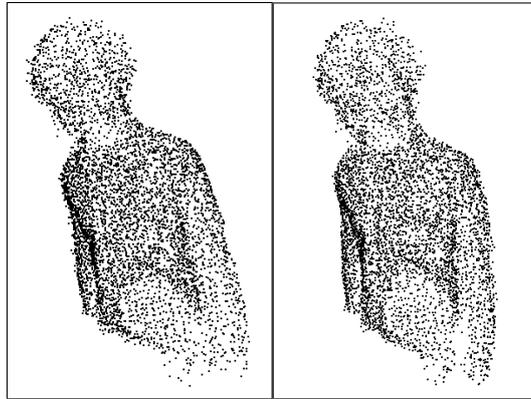


図 26: ステップ (iii) で推定したテンプレート点から生成した 3 次元形状テンプレート

図 26 において, DS1 および DS3 では対象物のおおまかな形状を表現した 3 次元形状テンプレートが生成されたが, DS2 では 3 次元形状テンプレートの腕の部分が生成されなかった. これは, 腕の部分における特徴点对応の少さに起因して, 広がったまま残ったテンプレート点が多く存在するためであると考えられる. この問題を解決するためには, 偏りの少ない対応付けが必要となる.

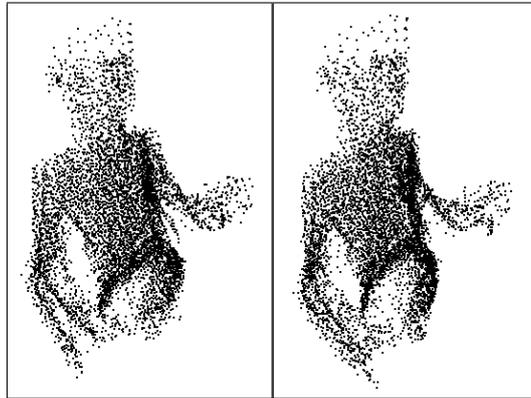
D. ステップ (i) の有無によるステップ (ii) のテンプレート点の推定結果

ステップ (ii) においてアフィン変換の初期値としてステップ (i) で得られた剛体変換パラメータ ($\mathbf{A}^f = \mathbf{R}^f$, $\mathbf{b}^f = \mathbf{t}^f$) を用いた場合と, 恒等変換 ($\mathbf{A}^f = \mathbf{I}$, $\mathbf{b}^f = \mathbf{0}$) を利用した場合に得られたテンプレート点を図 27 に示す. 同図から, これらの結果にはほとんど差がないことがわかる. しかし, パラメータの更新回数に対するエネルギー関数 E' (式 (9)) の推移 (図 28) を見ると, 恒等変換よりも剛体変換を初期値として利用した方が, 少ない更新回数でエネルギーが収束した. これは, 剛体変換の場合では大まかな姿勢が初期値として与えられているため, パラメータを少しだけ変化させるだけでエネルギーが収束するのに対し, 恒等変換の場合では対象物の裏側から撮影された点群などに対して, 初期姿勢からパラメータを大きく変化させる必要があり, エネルギーが収束するまでに時間がかかるためであると考えられる. また, ステップ (ii) で推定するパラメータ数が, RGB-D 画像列



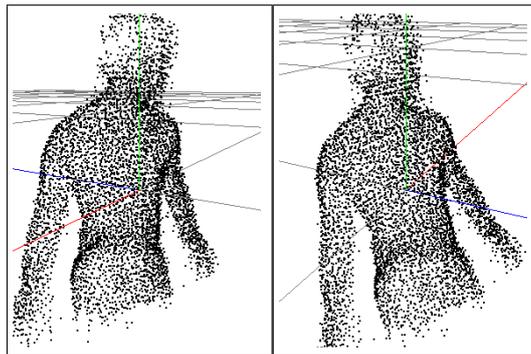
剛体変換 (DS1)

恒等変換 (DS1)



剛体変換 (DS2)

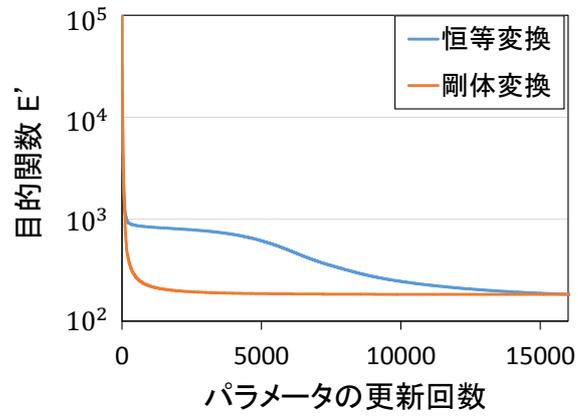
恒等変換 (DS2)



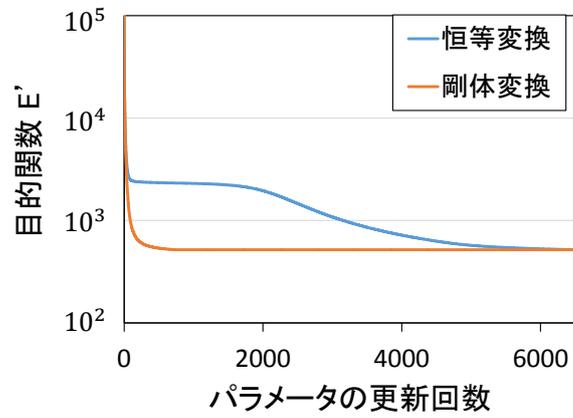
剛体変換 (DS3)

恒等変換 (DS3)

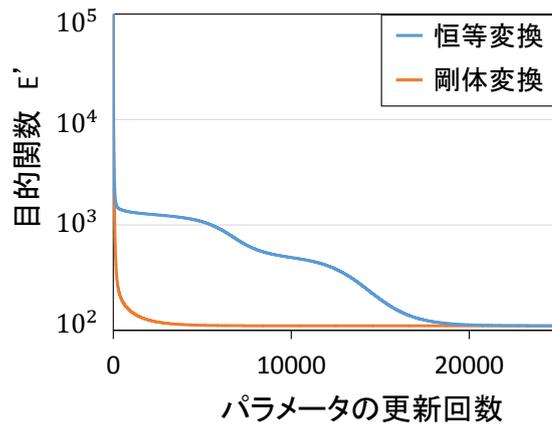
図 27: 初期値が剛体変換と恒等変換の場合でのステップ (ii) におけるテンプレート点の推定結果



(a) DS1



(b) DS2



(c) DS3

図 28: パラメータの更新回数に対するエネルギー関数 E' (式 (9)) の推移

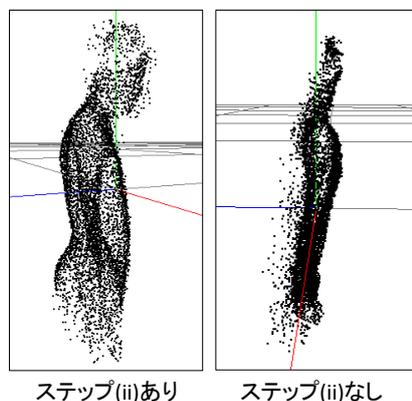
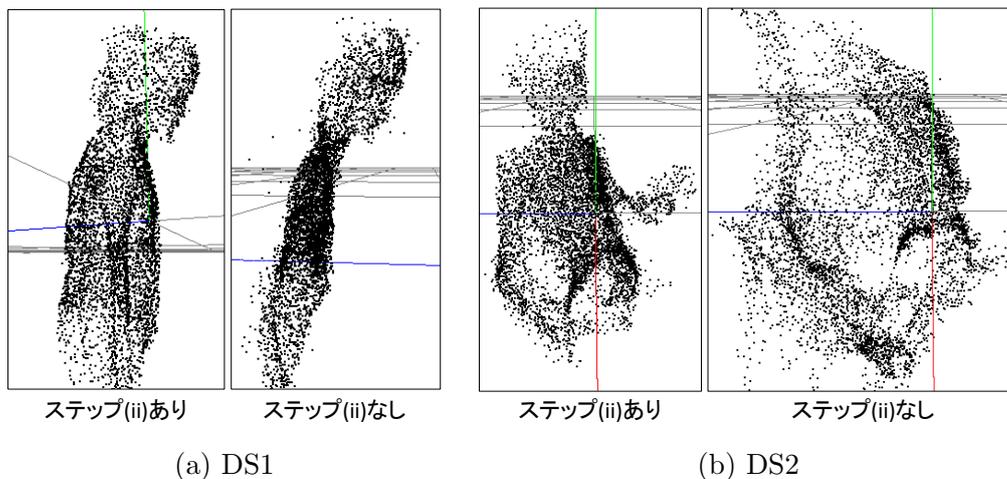


図 29: ステップ (ii) の有無によるステップ (iii) のテンプレート点の推定結果

から取得した 3 次元追跡点の数に対して少ないため、ステップ (ii) での推定結果が局所解に陥らなかったと考えられる。

E. ステップ (ii) の有無によるステップ (iii) のテンプレート点の推定結果の比較

ステップ (iii) においてテンプレート点およびアフィン変換の初期値としてステップ (ii) で得られた結果を用いた場合と、零ベクトルおよび恒等変換 ($\mathbf{p}_i = \mathbf{0}$, $\mathbf{A}_i^f = \mathbf{I}$, $\mathbf{b}_i^f = \mathbf{0}$) を用いた場合での推定したテンプレート点の違いを図 29 に示す。

図 29 のステップ (ii) の最適化を行わない場合のテンプレート点の推定結果において、DS1, DS3 では形状が潰れており、DS2 ではテンプレート点が広がっている

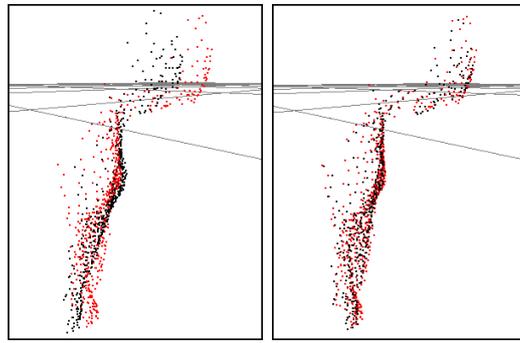
る。この結果から、ステップ (ii) の最適化を行わない場合には、望ましくない局所解に陥ることを確認した。

F. ステップ (ii) およびステップ (iii) における推定精度評価

テンプレート点と追跡点間の位置合わせ精度を評価するために、ステップ (ii) およびステップ (iii) で推定したアフィン変換の逆変換を用いてテンプレート点 \mathbf{p}_i を RGB-D カメラ座標系へ変換した点 $(\mathbf{A}_i^f)^{-1}(\mathbf{p}_i - \mathbf{b}_i^f)$ と 3次元追跡点 \mathbf{x}_i^f との間の距離 $d = \|\mathbf{x}_i^f - (\mathbf{A}_i^f)^{-1}(\mathbf{p}_i - \mathbf{b}_i^f)\|_2$ を全ての 3次元追跡点に対して求めた (図 30)。これらの平均値と最大値の推移を図 31 に示す。

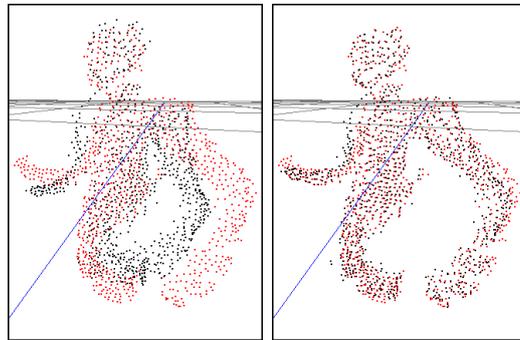
図 30 より、ステップ (ii) では同一フレーム内の変形を単一のアフィン変換として仮定しているため、RGB-D カメラ座標系へ変換したテンプレート点と 3次元追跡点が大きく離れている部分が存在していることがわかる。また、ステップ (iii) では局所的なアフィン変換を仮定しているため、変換したテンプレート点と 3次元追跡点がほとんど一致していることがわかる。

また、図 31 より、ステップ (ii) およびステップ (iii) において、距離 d がともに小さくなっており、平均距離は 0.01 [m] 程度となった。一方、ステップ (iii) における最大距離は減少しなかった。これは、距離が最大となる追跡点において、推定されたアフィン変換の線形成分が軸の反転を含む変換となったためであると考えられる。この問題は、アフィン変換の線形成分の行列式が正になるように制約を設けることにより解決できると考えられる。



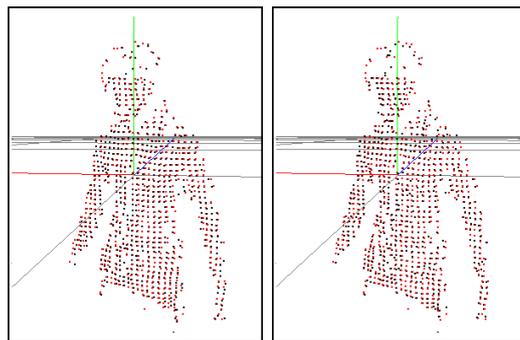
ステップ (ii) ステップ (iii)

(a) DS1



ステップ (ii) ステップ (iii)

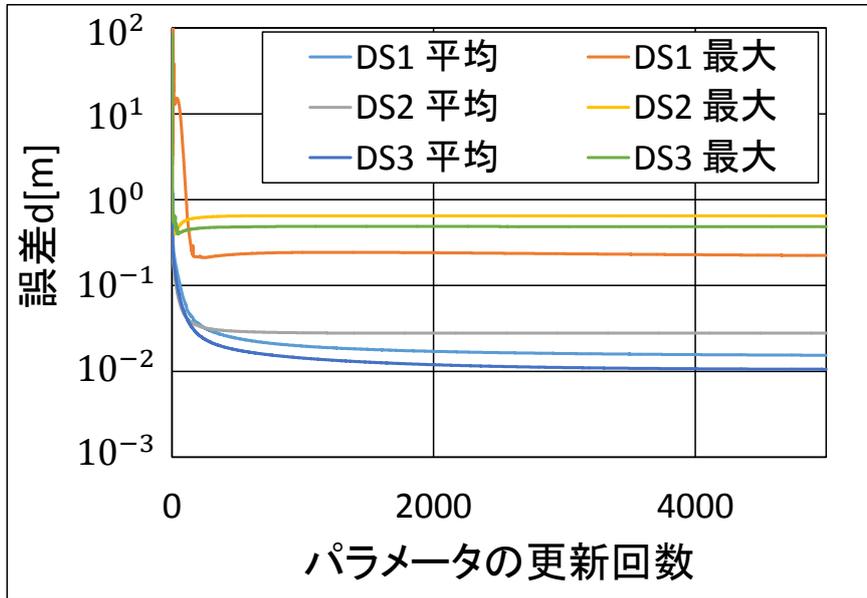
(b) DS2



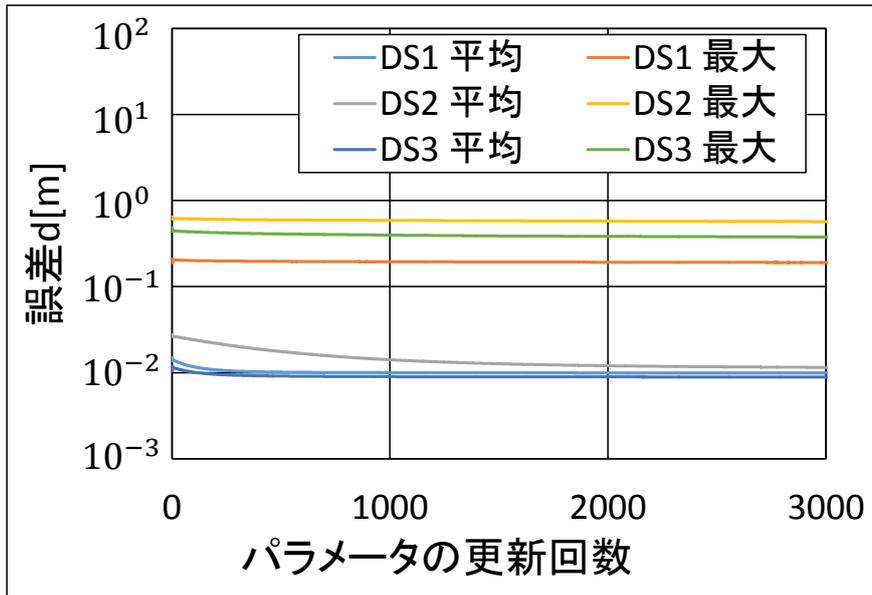
ステップ (ii) ステップ (iii)

(c) DS3

図 30: RGB-D カメラ座標系へ変換したテンプレート点 (黒) と 3次元追跡点 (赤)



(a) ステップ (ii)



(b) ステップ (iii)

図 31: 変換後のテンプレート点と3次元追跡点間の平均距離および最大距離の推移

5. まとめ

本研究では非剛体物体の3次元形状および動きの復元に向け、移動・変形中の非剛体物体を単一のRGB-Dカメラで全周撮影したRGB-D画像列から3次元形状テンプレートを生成する手法を提案した。提案手法では、非剛体ICPアルゴリズムに基づいて3次元形状テンプレート上の点を推定することで3次元形状テンプレートを生成する。安定した非剛体位置合わせを実現するために、提案手法ではRGB画像から得られる時間的に連続する画像間での追跡点を利用することで、非剛体ICPアルゴリズムにおける対応付けの曖昧さを緩和する。また、時間的に連続しない画像間での特徴点对応を利用することで位置合わせ誤差の蓄積を抑制する。さらに、推定するパラメータの数を段階的に上げて最適化することで望ましくない局所解に陥ることを回避する。本論文では、移動・変形中の非剛体実物体を全周撮影したRGB-D画像列から、提案手法によってテンプレート点を推定し、3次元形状テンプレートが生成可能であることを確認することを目的として、テンプレート点の推定結果の考察、推定するパラメータの初期値の違いによる結果の比較、およびテンプレート点と追跡点間の位置合わせ精度の評価を行い、提案手法の有用性を確認した。この結果から、特徴点マッチングによる対応点の誤対応による3次元形状テンプレートへの影響などに関する知見を得た。今後は、1) ロバスト推定を取り入れたエネルギー関数の設計などによって、誤対応に起因する問題の低減などについて検討する。また、提案手法ではRGB画像から追跡点や特徴点对応を抽出しているため、テクスチャの少ない対象物に対しては追跡点や特徴点对応の抽出が困難なため、本手法の適用が難しい。そこで、2) RGB画像だけでなくデプス画像も考慮した追跡点や特徴点对応の抽出を行うことで、テクスチャの少ない対象物に対しても頑健に3次元形状テンプレートを生成可能にする。さらに、3) 生成した3次元形状テンプレートを用いて対象物の動きを復元することで、提案手法の有効性を評価する。

謝辞

本研究を進めるにあたり，その全過程において細やかな御指導，御鞭撻を頂いた視覚情報メディア研究室 横矢 直和 教授に心より感謝いたします。そして，本研究の遂行にあたり，有益な御助言，御鞭撻を頂いたインタラクティブメディア設計学研究室 加藤 博一 教授に厚く御礼申し上げます。さらに，本研究の全過程を通して，始終温かい御指導をして頂いた視覚情報メディア研究室 佐藤 智和 准教授に深く感謝申し上げます。また，研究に関して的確な御助言を頂いた視覚情報メディア研究室 中島 悠太 助教，河合 紀彦 助教に厚く御礼申し上げます。また，研究室での生活を支えて頂いた視覚情報メディア研究室 石谷 由美 女史に心より感謝いたします。本研究で使用したデータセットの撮影に協力して頂いた視覚情報メディア研究室 黄 頌友 氏および 林 佑亮氏に深く感謝いたします。

参考文献

- [1] TriMirror Virtual Fitting Room. <http://www.trimirror.com/en/about/>.
- [2] FooPets — Real Virtual Pets Online. <http://www.foopets.com/>.
- [3] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. In *Proc. Eurographics Symposium on Geometry Processing (SGP)*, pp. 1421–1430, 2008.
- [4] B. Amberg, S. Romdhani, and T. Vetter. Optimal step nonrigid ICP algorithms for surface registration. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, 8 pages, 2007.
- [5] P. J. Besl and N. D. McKay. Method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 14, No. 2, pp. 239–256, 1992.
- [6] M. Zollhöfer, M. Nießner, S. Izadi, C. Rehmann, C Zach, M. Fisher, C. Wu, A. Fitzgibbon, C. Loop, C. Theobalt, and M. Stamminger. Real-time non-rigid reconstruction using an RGB-D camera. *ACM Trans. on Graphics (TOG)*, Vol. 33, No. 4, pp. 156:1–156:12, 2014.
- [7] H. Li, B. Adams, L. J. Guibas, and M. Pauly. Robust single-view geometry and motion reconstruction. *ACM Trans. on Graphics (TOG)*, Vol. 28, No. 5, pp. 175:1–175:10, 2009.
- [8] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1434–1441, 2010.
- [9] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 32, No. 8, pp. 1362–1376, 2010.

- [10] H. Baker. Three-dimensional modelling. In *Proc. Int'l Joint Conf. on Artificial Intelligence (IJCAI)*, Vol. 2, pp. 649–655, 1977.
- [11] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 16, No. 2, pp. 150–162, 1994.
- [12] G. K.M. Cheung, T. Kanade, J.-Y. Bouguet, and M. Holler. A real time system for robust 3D voxel reconstruction of human motions. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 714–720, 2000.
- [13] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *Int'l Journal of Computer Vision*, Vol. 59, No. 3, pp. 207–232, 2004.
- [14] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. IEEE and ACM Int'l Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 225–234, 2007.
- [15] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison. DTAM: Dense tracking and mapping in real-time. In *Proc. IEEE Int'l Conf. on Computer Vision (ICCV)*, pp. 2320–2327, 2011.
- [16] J. Engel, T. Schöps, and D. Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *Proc. European Conf. on Computer Vision (ECCV)*, pp. 834–849. 2014.
- [17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int'l Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, 2004.
- [18] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Proc. European Conf. on Computer Vision (ECCV)*, pp. 404–417, 2006.

- [19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: an efficient alternative to SIFT or SURF. In *Proc. IEEE Int'l Conf. on Computer Vision (ICCV)*, pp. 2564–2571, 2011.
- [20] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int'l Joint Conf. on Artificial Intelligence (IJCAI)*, Vol. 81, pp. 674–679, 1981.
- [21] P. Sand and S. Teller. Particle video: Long-range motion estimation using point trajectories. *Int'l Journal of Computer Vision*, Vol. 80, No. 1, pp. 72–91, 2008.
- [22] N. Sundaram, T. Brox, and K. Keutzer. Dense point trajectories by GPU-accelerated large displacement optical flow. In *Proc. European Conf. on Computer Vision (ECCV)*, pp. 438–451, 2010.
- [23] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground–background segmentation using codebook model. *Real-Time Imaging*, Vol. 11, No. 3, pp. 172–185, 2005.
- [24] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics (TOG)*, Vol. 23, No. 3, pp. 309–314, 2004.
- [25] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *Int'l Journal of Computer Vision*, Vol. 38, No. 3, pp. 199–218, 2000.
- [26] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, Vol. 37, No. 4, pp. 827–849, 2004.
- [27] S. Zhang and P. Huang. High-resolution, real-time 3D shape acquisition. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW)*, pp. 28–37, 2004.

- [28] M. Hansard, S. Lee, O. Choi, and R. P. Horaud. *Time-of-Flight Cameras: Principles, Methods and Applications*. Springer Science & Business Media, 2012.
- [29] L. Li. Time-of-flight camera - an introduction. Technical report, Texas Instruments - Technical White Paper, 2014.
- [30] V. Pradeep, C. Rhemann, S. Izadi, C. Zach, M. Bleyer, and S. Bathiche. MonoFusion: Real-time 3D reconstruction of small scenes with a single web camera. In *Proc. IEEE Int'l Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 83–88, 2013.
- [31] R. A. Newcombe, A. J. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *Proc. IEEE Int'l Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 127–136, 2011.
- [32] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman. PatchMatch: a randomized correspondence algorithm for structural image editing. *ACM Trans. on Graphics (TOG)*, Vol. 28, No. 3, pp. 24:1–24:10, 2009.
- [33] M. Bleyer, C. Rhemann, and C. Rother. PatchMatch Stereo - stereo matching with slanted support windows. In *Proc. British Machine Vision Conf. (BMVC)*, Vol. 11, pp. 14.1–14.11, 2011.
- [34] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proc. ACM Conf. on SIGGRAPH*, pp. 303–312, 1996.
- [35] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. pp. 61–70, 2006.
- [36] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An evaluation of the RGB-D SLAM system. In *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, pp. 1691–1696, 2012.

- [37] J. Starck and A. Hilton. Surface capture for performance-based animation. *IEEE Computer Graphics and Applications*, Vol. 27, No. 3, pp. 21–31, 2007.
- [38] E. De Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H. Seidel, and S. Thrun. Performance capture from sparse multi-view video. *ACM Trans. on Graphics (TOG)*, Vol. 27, No. 3, pp. 98:1–98:10, 2008.
- [39] D. Vlastic, I. Baran, W. Matusik, and J. Popović. Articulated mesh animation from multi-view silhouettes. *ACM Trans. on Graphics (TOG)*, Vol. 27, No. 3, pp. 97:1–97:9, 2008.
- [40] M. Dou, H. Fuchs, and Frahm J. Scanning and tracking dynamic objects with commodity depth cameras. In *Proc. IEEE Int’l Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 99–106, 2013.
- [41] M. Dou and H. Fuchs. Temporally enhanced 3D capture of room-sized dynamic scenes with commodity depth cameras. In *Proc. IEEE Virtual Reality (VR)*, pp. 39–44, 2014.
- [42] I. Oikonomidis, N. Kyriazis, and A. A. Argyros. Efficient model-based 3D tracking of hand articulations using Kinect. In *Proc. British Machine Vision Conf. (BMVC)*, pp. 101.1–101.11, 2011.
- [43] T. Schmidt, R. Newcombe, and D. Fox. DART: Dense articulated real-time tracking. In *Proc. Robotics: Science and Systems*, 9 pages, 2014.
- [44] M. Zeng, J. Zheng, X. Cheng, and X. Liu. Templateless quasi-rigid shape modeling with implicit loop-closure. In *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 145–152, 2013.
- [45] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 9, No. 5, pp. 698–700, 1987.