

NAIST-IS-MT1351124

修士論文

固定RGB-Dカメラを用いた
手持ち物体のモデリングシステム

黄 頌友

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。

黄 頌友

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)

固定 RGB-D カメラを用いた 手持ち物体のモデリングシステム*

黄 頌友

内容梗概

近年、拡張現実や仮想現実などの技術を用いて実物体を仮想的に提示するアプリケーションが一般のユーザに利用され始めている。実物体の仮想提示には、その3次元モデルを用いてCGを描画する手法が広く用いられるが、実物体の3次元モデル生成には一般的に高度な技術を要する。これに対して、実物体をRGB-Dカメラで移動撮影して得られた点群に基づいて、その実物体を自動で3次元モデル化する手法が提案されている。しかし、この手法では実物体を固定撮影することを前提としているため、対象物体の接地面などが撮影できず、3次元モデルに欠損が生じるという問題がある。これに対して、特に手で把持可能な小物体を対象とし、固定RGB-Dカメラを用いて把持した実物体を回転移動させながら撮影することで、欠損のない3次元モデルを生成するシステムが提案されている。ただし、これらのシステムでは把持部分と実物体を区別するために、精密に測定されたロボットアームによる把持や、把持部分を検出するための特殊な環境内での撮影を前提としており、一般のユーザによる手軽な3次元モデル生成には適さない。

そこで本研究では、手で把持可能な実物体を対象とし、一般のユーザが把持した実物体を回転移動しながら、固定RGB-Dカメラで撮影することで、手軽に利用可能な3次元モデリングシステムを提案する。提案システムでは、ユーザの手と対象物体の位置関係が変化することに着目し、RGB-Dカメラから得られる点群から剛体運動で記述できない点を除去する。具体的には、各RGB-D画像の点

*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-MT1351124, 2015年3月12日.

群から抽出された3次元特徴点を利用して、まずRANSACにより実物体の剛体運動を推定する。続いて、ICPアルゴリズムで剛体運動パラメータの推定結果を高精度化した上で、剛体運動により記述できない点を除去する。最後に、推定された剛体運動に基づいて各画像の点群を統合し、3次元モデルを生成する。また、提案システムではRGB-Dカメラで撮影されたRGB画像をテクスチャとして利用することでテクスチャ付き3次元モデルの自動生成を実現する。実験では、(1)生成された3次元モデルを用いた描画結果の品質、(2)本システム利用時におけるユーザの手に対応する点群の除去率、の2点に着目して提案手法の有効性を検証する。

キーワード

3次元復元, RGB-Dカメラ, 手持ち物体, RANSAC

3D Reconstruction of a Handheld Object with a Fixed RGB-D Camera*

Sung-yu Huang

Abstract

Applications of virtual reality and augmented reality, such as virtual furniture arrangement, are getting popular among average users in these years, displaying a real-world object as a virtual image. A well-known method for displaying a real-world object as a virtual object uses a 3D model of the object. Since 3D modeling is often a time-consuming and cumbersome task, some methods are proposed to build such a 3D model automatically from point clouds obtained using an RGB-D camera as a 3D scanner, to enable users to build 3D models without excessive burden. However, these methods may suffer from missing surfaces in regions where the RGB-D camera cannot observe, such as supporting surfaces. To prevent such missing surfaces in the 3D model, particularly for a hand-held object, some methods are proposed that allow users to scan the complete object by turning it around in front of the RGB-D camera, so that the object can be captured from any direction. However, these methods require a precisely measured robot arm or a special environment for capturing to distinguish hands from the object.

In this study, focusing on modeling a hand-held rigid object, we propose a modeling system for average users. The system allows users to scan a complete object by turning the object around in front of the RGB-D camera by their own hands as the existing works. Note that it does not require any other additional

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1351124, March 12, 2015.

tools nor platforms. Observing that the geometrical relationship between the target object and hands changes during capturing, we distinguish the points on hands from ones on the target object by finding points that cannot be described by rigid-body motion. More specifically, our system extracts 3D feature points in the point clouds, and use them to estimate the dominant rigid-body motion with RANSAC. After refining the result of rigid-body motion estimation by the ICP algorithm, we remove points that do not fit to the refined rigid-body motion. Finally, we merge point clouds from all RGB-D frames and build the 3D model of the object. In order to render the model in high quality, the proposed system applies RGB image to the model as texture. In our experiments, we focus on the following two points: (1) how the quality of the obtained 3D model is satisfactory and (2) how many points of hands are removed by the system.

Keywords:

3D reconstruction, RGB-D camera, hand-held object, RANSAC

目次

1. はじめに	1
2. 関連研究と本研究の位置付け	3
2.1 3次元モデル生成	3
2.1.1 RGBカメラ画像を用いる手法	3
2.1.2 RGB-Dカメラ画像を用いる手法	5
2.2 小物体の3次元モデル生成インタフェース	6
2.3 本研究の位置付けと概要	7
3. 把持部を考慮した3次元モデル生成システム	9
3.1 提案手法の概要	9
3.2 撮影用ユーザインタフェース	11
3.3 特徴点検出とマッチング	12
3.4 RANSACを用いた位置合わせ	14
3.5 把持部を考慮した位置合わせの高精度化	15
3.6 把持部を除去した対象物のモデル生成	17
4. 実験	19
4.1 提案システムを用いたモデル生成実験	19
4.2 持ち手の除去率評価実験	25
5. まとめ	33
謝辞	34
参考文献	35

目 次

1	実物体を3次元モデル化の応用例	2
2	Agarwalらによる大規模な特徴点3次元位置推定結果 [14]	4
3	視体積獲得のイメージ図 [15]	5
4	RGB-D カメラ	5
5	事前計測したロボットアームによる対象物特定 [7]	7
6	Weiseらのインタフェース [8]	7
7	提案手法の入力例	8
8	提案手法の流れ	10
9	撮影用のユーザインタフェース	11
10	ISS 特徴点の検出例	12
11	対応付けによる位置合わせ結果例	13
12	RANSACによる点群位置合わせの流れ	16
13	対象物の形状ボリューム推定例	18
14	入力の一例	19
15	対象物 A と入力画像例	20
16	対象物 B と入力画像例	20
17	対象物 C と入力画像例	21
18	対象物 D と入力画像例	21
19	対象物 E と入力画像例	22
20	対象物 F と入力画像例	22
21	対象物 G と入力画像例	23
22	対象物 H と入力画像例	23
23	対象物 A の復元結果	26
24	対象物 B の復元結果	26
25	対象物 C の復元結果	27
26	対象物 D の復元結果	27
27	対象物 E, F, G, H の復元結果	28
28	対象物 A に対するラベルの例	29

29	対象物 A における ROC 曲線	30
30	対象物 B における ROC 曲線	30
31	対象物 C における ROC 曲線	31
32	対象物 D における ROC 曲線	31
33	対象物 A, B, C, D における ROC 曲線	32
34	撮影角度による色変化の例	33

表 目 次

1	Microsoft Kinect V2 のカメラ規格	24
2	本研究に用いる閾値	24

1. はじめに

実物体の3次元モデル化は、文化財のアーカイブ [1] や拡張現実を用いた実物体の配置シミュレーション [2] など、様々な目的で利用されている (図 1). しかし、実物体を3次元モデル化するには、人手での長時間におよぶ複雑な作業、または3D スキャナなど高価な機械が必要なため、簡便に実物体を3次元モデル化できる手法が求められている.

これに対し、剛体の対象物を前提として、RGB カメラやデプスカメラにより、対象物の色や形状を計測し、自動的に対象物の3次元モデルを生成する手法が提案されている [3, 4]. また、カラー画像とデプス画像を同時に計測可能な Microsoft Kinect や Asus Xtion PRO といった RGB-D カメラを用いて、対象物の形状とテクスチャを復元する手法も提案されている [5, 6]. しかし、これらの手法では対象物の自己遮蔽や計測位置の制限による対象物の未計測部分が復元できずに3次元モデルの欠損となる.

この問題に対し、特に手で把持可能な小物体を対象とし、把持した実物体を移動・回転させながら固定カメラを用いて撮影することで、全方位から対象物を撮影する3次元モデル生成システムが提案されている [7, 8]. Krainin ら [7] は、ロボットアームによる把持を用いてモデルを生成した. また、Weise ら [8] は把持部を検出するために、特殊な環境内で単色の手袋を着けて対象物を撮影する、インタラクティブな3次元モデル生成システムを提案した. これらの手法は対象物と把持部を区別するために、ロボットアームという高価な機材、もしくは把持部を検出するための特殊な環境が必要で、一般ユーザが利用しにくいという問題がある.

本研究では、一般ユーザによる使用を想定し、素手で把持した対象物を固定 RGB-D カメラを用いて撮影する3次元モデル生成システムを提案する. 提案システムでは、素手で対象物を移動・回転させる過程で手が対象物に対して非剛体運動することから、撮影された物体の上の任意の点が剛体運動で記述できるかを判定することにより対象物と把持部を自動的に区別する. 具体的には、入力点群ストリームの各点群から3次元特徴点 ISS [9] を検出し、CSHOT 特徴量 [10] を用いて点群間の特徴点を対応付ける. 続いて、対応点に対して剛体運動を仮定した



(a) 発掘された鋳のアーカイブ [1]



(b) 家具配置シミュレーション [2]

図 1: 実物体を 3 次元モデル化の応用例

RANSAC[11]により各点群を位置合わせする。このときに、前述のように把持部は非剛体運動となることから、RANSACにおける外れ値を把持部として大まかに除去する。更に、ICPアルゴリズムを利用して位置合わせ結果を高精度化した後、符号付き距離場[12]に基づき残存する把持部の点群を除去し、対象物の形状とテクスチャを復元する。本研究では3次元モデル化実験と把持部除去実験を行い、単色の背景と同色の手袋やロボットアームなどを利用せずに対象物のみの3次元モデル化が可能であることを示す。

本論文では、2章で3次元モデル生成に関する従来研究を概観し、本研究の位置づけについて述べる。次に3章で提案システムの概要、位置合わせ手法、モデル生成方法について記述する。4章では、提案システムを用いて各種対象物に対する復元実験、および生成した3次元モデルの把持部除去率に対する評価実験のそれぞれの概要と結果を述べる。最後に、5章でまとめおよび今後の展望を述べる。

2. 関連研究と本研究の位置付け

本章では、まず剛体の実物体を3次元モデル化する従来研究について概観し、次に対象物を把持して撮影する3次元モデル生成インターフェースを紹介する。最後に関連研究に対する本研究の位置づけを述べる。

2.1 3次元モデル生成

実物体の3次元モデルを生成するには、対象物の形状を計測する必要がある。従来はレーザスキャナなど形状計測用の機器が使われてきた[13]。近年では、広く普及しているRGBカメラにより形状を計測する手法が多く提案されており[3, 14, 15, 16, 17]、さらにRGB-Dカメラを利用する手法が広く注目されている[5, 6, 7, 8, 18]。本節では、RGBカメラとRGB-Dカメラを用いる3次元モデル生成手法について紹介する。

2.1.1 RGBカメラ画像を用いる手法

RGBカメラ画像を用いる代表的な手法として、ステレオ法を用いる手法と視体積交差法が提案されている。ステレオ法は、異なる位置から対象物を撮影した複数枚の画像から、対象物の3次元モデルを生成する手法である。この手法は、三角測量の原理に基づき、各画像の撮影位置から画像中の点のデプス値を求めて対象物の形状を推定する。Campbellら[19]は複数の入力画像に対し各画像のデプス画像を推定し、デプス画像を統合することでモデル生成する手法を提案した。Taylor[20]は特徴点对応を用いて各特徴点の3次元位置とカメラ位置姿勢を推定し、さらに各特徴点に基づき適切なメッシュとテクスチャを生成する手法を提案した。Pradeepら[3]は、GPUの使用により単一カメラを用いたリアルタイム3次元モデル生成が可能であることを示している。一方、ステレオ法を利用する際には、各画像の撮影位置を推定する必要があるが、これにはSFM (Structure from Motion) と呼ばれる特徴点の動きを用いる手法が利用される。Wuら[21, 22]はGUIシステムにより簡単にカメラの位置姿勢を推定できるソフトウェア VisualSFM を公



図 2: Agarwal らによる大規模な特徴点 3 次元位置推定結果 [14]

開している。Agarwal ら [14] は市街地の画像を大量に収集し、インターネットから収集した画像による大規模なカメラ位置姿勢推定が可能であることを示した。ステレオ法では、形状の推定に対応点を用いるため、テクスチャのないシーンの形状を推定できない。このため、物体の 3 次元モデル生成を目的とする場合、復元できる対象はテクスチャが豊富な物体に限定される。

視体積交差法は、対象物の 2 次元シルエット画像から、対象物のモデルを生成する手法である [15]。この手法では、シルエットを実空間に投影した視錐 (Visual Corn) の中に対象物が含まれるシルエット制約を利用し、多視点から視錐の共通領域である視体積 (Visual Hull) を求めることで対象物の形状を推定する。Cheung ら [17] は各カメラ中心から対象物の輪郭に対する接線を考え、接線上にある各シルエットの内部に存在する線分を対象物の輪郭線とし、移動撮影した画像ストリームを入力としたモデル生成手法を提案した。Matusik ら [23] は視体積を利用して計算コストを軽減し、仮想視点における実物体の見た目をリアルタイムにレンダリング可能なアルゴリズムを提案した。また、不殿ら [24] は手持ちカメラによる画像ストリームを入力とし、AR マーカーを用いた撮影インタフェースを提案している。この手法は、撮影時点での復元結果をユーザに提示することに加え、未計測部分を撮影する指示を与えることにより、良好な 3 次元モデルが得られる。視体積交差法には、シルエットには物体の凹な形状が反映されないことに起因して、対象物の凹部分を再現できないという欠点がある。

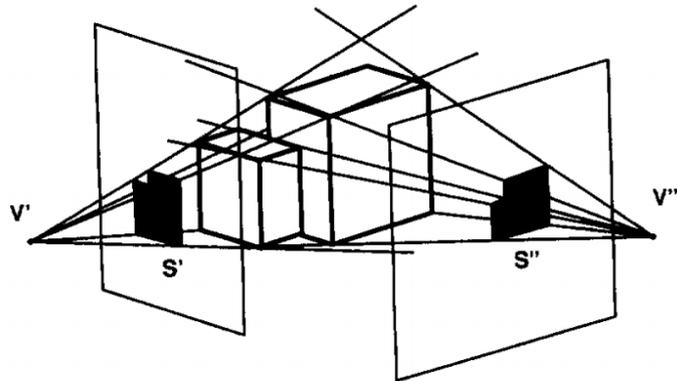


図 3: 視体積獲得のイメージ図 [15]



(a) Microsoft Kinect V2



(b) Asus Xtion PRO LIVE

図 4: RGB-D カメラ

2.1.2 RGB-D カメラ画像を用いる手法

RGB-D カメラは RGB 画像と同時にデプス画像獲得できるカメラである．近年では Microsoft Kinect (図 4a) と Asus Xtion PRO (図 4b) などの RGB-D カメラが一般ユーザでも入手可能である．RGB-D カメラから得られるデプス画像は点群データとして利用可能であり，複数の点群を統合することで対象物の形状を推定できる．また，RGB-D カメラを用いる手法はテクスチャの有無に影響されず凹形状も再現可能という利点があり，近年 RGB-D カメラを用いた 3 次元モデル生成手法が多く提案されている [5, 6, 7, 8, 18]．Newcombe ら [5] は入力点群を逐次的に位置合わせすることにより，リアルタイムに剛体シーンの 3 次元モデル生

成が可能であることを示した。また，Endres ら [6] はカラー画像から 2 次元特徴点を検出し，特徴点の位置合わせ結果を点群位置合わせの初期値として用い，カメラの自己位置推定と環境の 3 次元モデル生成を同時に行う RGB-D SLAM を提案した。Zhou ら [18] は RGB-D SLAM の生成したモデルから注目領域を検出することにより，局所の形状を保持しつつ広い範囲における 3 次元モデル生成手法を提案した。これらの手法は，部屋や机といった比較的広範囲のモデル生成を目的としており，移動撮影して得られた RGB-D ストリームを入力に用いる。

2.2 小物体の 3 次元モデル生成インタフェース

2.1.2 項では RGB-D カメラで計測したデータを用いて 3 次元モデルを生成する方法を紹介した。しかし，対象物の自己遮蔽や計測位置の制限により，カメラの移動撮影での未計測部分が発生する可能性がある。そこで，対象物を把持可能な小物体であることを前提とし，対象物をカメラ前方で把持し，回転・移動させながら撮影することにより，これらの未計測部分を減少させる手法が提案されている [7, 8]。ただし，これらの手法には，いかに対象物と把持する部分を区別するかという課題がある。

この課題に対し，Krainin ら [7] は，ロボットアームを用いて対象物を把持する手法を提案した。この手法では，事前に計測したロボットアームの形状と各関節の角度といった情報を用いて，撮影したシーンからロボットアームと対象物を区別し，対象物のみを 3 次元モデルを生成する。また，ロボットアームの動きを適切にプログラムすることにより，対象物を全自動にスキャンすることが可能である。一方，Weise ら [8] はユーザ自身の手で対象物を把持する手法を提案した。この手法では，単色の撮影スペースを準備し，ユーザが同色の手袋を着用することで対象物を特定する。さらに，撮影時点で位置合わせの結果をプレビューとして表示することにより，即時に未計測部分の提示と追加撮影が可能である。

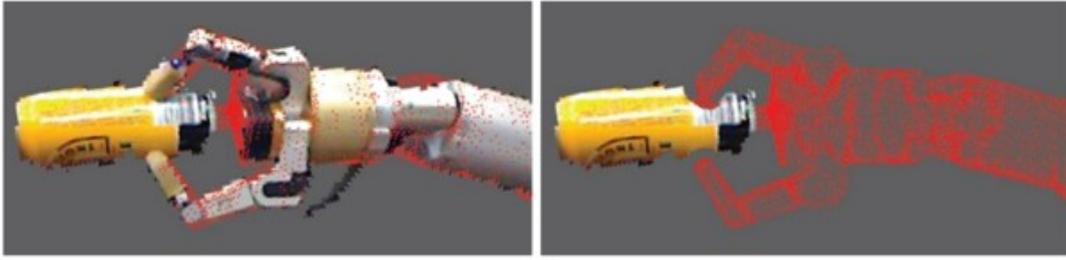


図 5: 事前計測したロボットアームによる対象物特定 [7]



図 6: Weise らのインタフェース [8]

2.3 本研究の位置付けと概要

本研究は、把持可能な剛体小物体を対象とした一般ユーザが手軽に利用できる 3次元モデル生成システムを開発することを目的とする。また、テクスチャが欠如する対象物を想定し、RGB-D カメラを用いて対象物を計測する。本研究の目的に対し、Krainin ら [7] らと Weise ら [8] の手法は、対象物と把持部を区別するために、ロボットアームといった特殊なデバイスや対象物を検出するための撮影環境を用いており、一般ユーザが手軽に使用できないといった問題がある。

従って本研究では、特殊なデバイスや撮影環境を利用せず、素手による把持の RGB-D ストリーム入力に対して対象物と把持部を区別する (図 7)。対象物を把持して撮影する場合、対象物は剛体であることから、次の性質を満たす。

1. 対象物の運動は剛体運動で記述できる。
2. 対象物を回転・移動させる過程で把持部は変形・移動する。



図 7: 提案手法の入力例

これらの性質から，入力点群において剛体運動で記述できない点が把持部に属すると考えられるため，これを利用して対象物と把持部を区別する．

本研究では，上記の性質に基づく把持部と対象物の区別のために，特徴点对応と RANSAC に基づくシステムを提案する．提案手法は，テクスチャの有無に依存しない 3 次元特徴点，および特徴量を利用して得られる点群間の対応点を利用し，RANSAC によって得られた剛体運動で記述できない点を把持部とすることで，特殊な環境を仮定することなく，対象物と把持部を区別する．

3. 把持部を考慮した3次元モデル生成システム

本章では、一般のユーザのための対象物を把持する3次元モデル生成システムの詳細を述べる。本研究では、3次元モデル生成に関する知識・技術を持たない一般のユーザでも、素手で把持した剛体の対象物を簡単にモデル化可能なシステムを構築することを目的とする。提案システムでは、RGB-Dカメラを用いて対象物を撮影したRGB-Dストリームを入力とする。また、Kraininら[7]やWeiseら[8]のシステムと同様に対象物を把持撮影することで対象物の全周を撮影し、欠損のない3次元モデルを生成する一般ユーザが簡単に利用可能なシステムを実現する。

以下ではまず、提案システムの設計と処理の流れを概説した後、それぞれの処理について詳述する。

3.1 提案手法の概要

提案システムでは、ユーザが撮影したRGB-Dストリームに含まれるRGB-Dフレームから3次元点群を抽出し、対象物の動きが剛体運動で記述できるものとして、点群を逐次的に位置合わせることにより3次元モデルを生成する。一般に、剛体運動を仮定した点群の位置合わせでは、ICPアルゴリズム[25]が広く利用されるが、ICPアルゴリズムは剛体変換のパラメータの初期値に大きく依存し、特に提案システムでは手などの把持部、もしくは写り込んだユーザの体の一部などにより対象物上の点群を位置合わせできない可能性がある。そこで提案システムでは、Henryら[26]の手法と同様に、3次元点群中の対応点に対して剛体運動を仮定したRANSACを適用することで対象物の位置合わせを行う。また、この位置合わせの結果に基づいて3次元点群中の対象物以外の点を除去し、その上でICPアルゴリズムにより位置合わせを高精度化する。

以上の手順において、対象物を把持した手などは、非剛体運動となるため、RANSACにより外れ値と判定される。しかし、手の領域が大部分を占める場合、手の動き方によっては手に対して剛体運動を当てはめる可能性がある。そこで、提案システムでは、撮影用のユーザインタフェースを提供することで、腕や体な

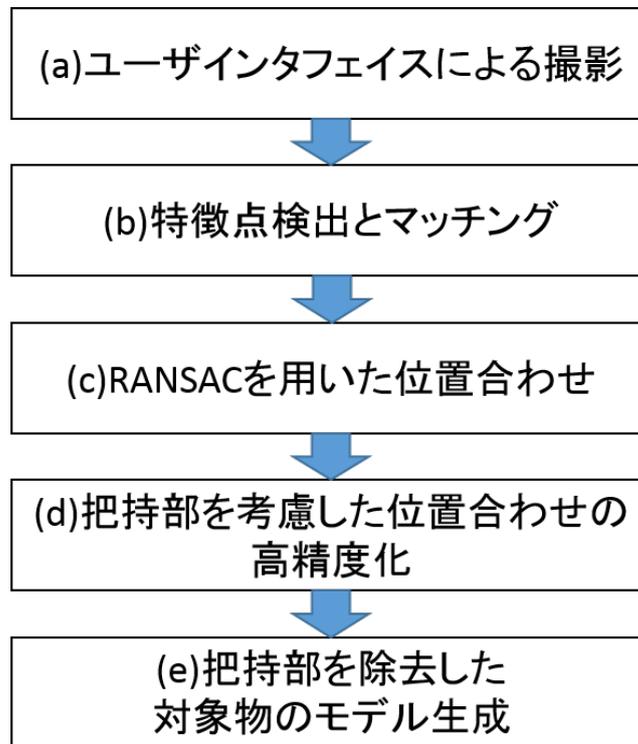


図 8: 提案手法の流れ

ど対象物以外の点群を事前に削減する。

図 8 に提案システムの概要を示す。提案システムでは、まず (a) 撮影用のユーザインタフェースによって RGB-D ストリームを取得する。(b) RGB-D ストリームから抽出した 3 次元点群から 3 次元点に対する特徴点および記述子を抽出し、3 次元点間の対応点を求める。この対応点に対して、(c) RANSAC を逐次的に適用することにより対象物を位置合わせした上で、(d) 対象物以外の点群を除去し、ICP アルゴリズムにより位置合わせを高精度化する。最後に (e) 符号付き距離場により 3 次元モデルを生成する。以下では、提案システムの各要素について詳述する。

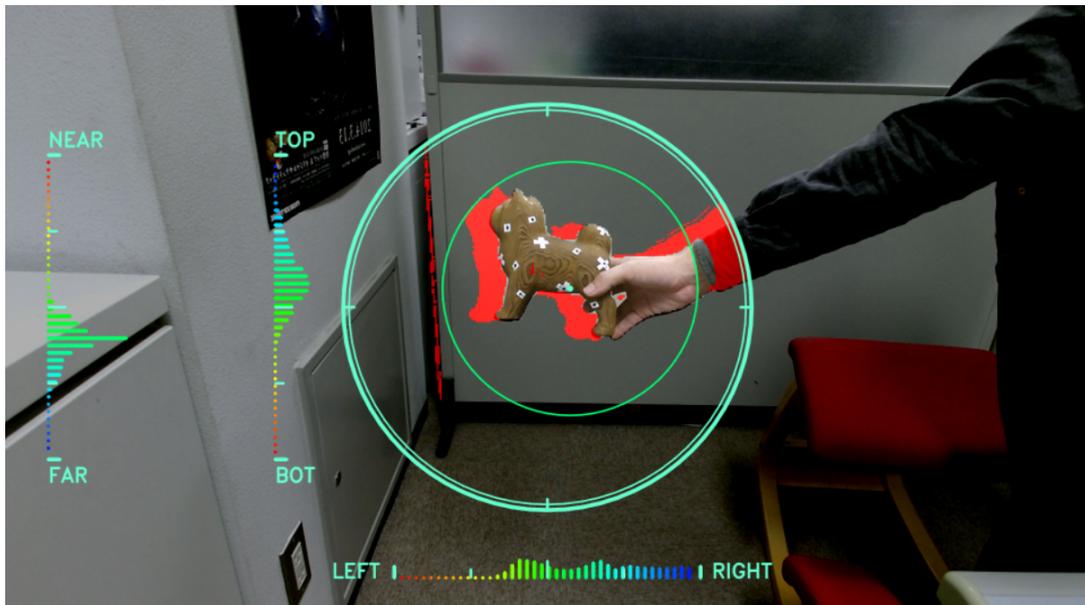


図 9: 撮影用のユーザインタフェース

3.2 撮影用ユーザインタフェース

前述のとおり，提案システムでは RANSAC において腕や体の一部などに対して剛体運動を当てはめる際の誤りを低減するために，予めこれらの部位を取り除く必要がある．そこで，点群の取得範囲を限定した上で，その範囲を視覚的に提示する撮影用のユーザインタフェースを提供する．これにより，前腕部などに対応する 3次元点の数を低減し，背景部分に対応する点群を除去すると共に，ユーザが対象物を RGB-D カメラの奥行き取得可能範囲内で動かせるようサポートする．

図 9 に提案システムが提供する撮影用ユーザインタフェースを示す．このインタフェースは，横方向，縦方向，および奥行き方向についてそれぞれ 3次元点群のヒストグラムをユーザに提示することで，撮影範囲内における対象物の位置を把握できるようにする．また，撮影範囲内部と外部にある 3次元点を，異なる輝度を用いて表示することで対象物が撮影範囲内にあることを提示する．提案システムは，このインタフェースから得られた N フレームの RGB-D ストリームから，撮影範囲内の 3次元点群 $Y_n = \{\mathbf{y}_{ni} | i = 1, \dots, |Y_n|\}$ ($n = 1, \dots, N$) および RGB 画像を抽出し，これらに対して以降の処理を適用する．

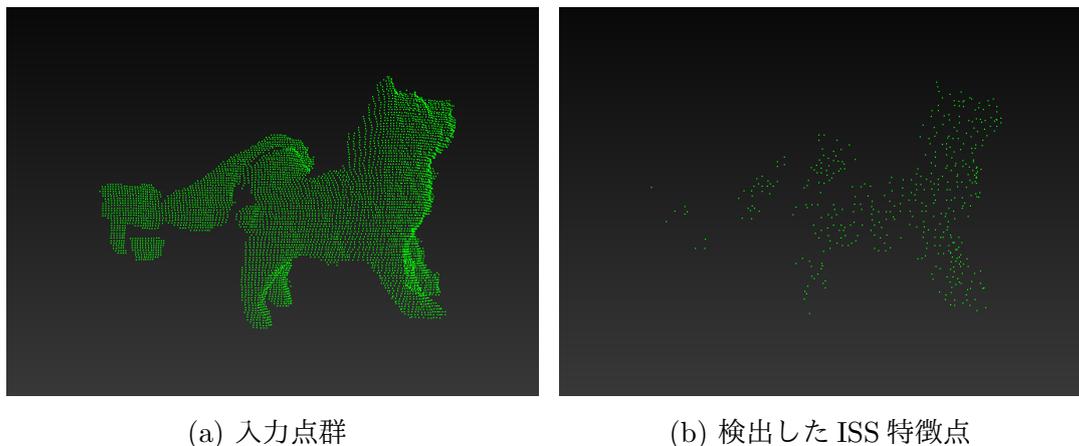


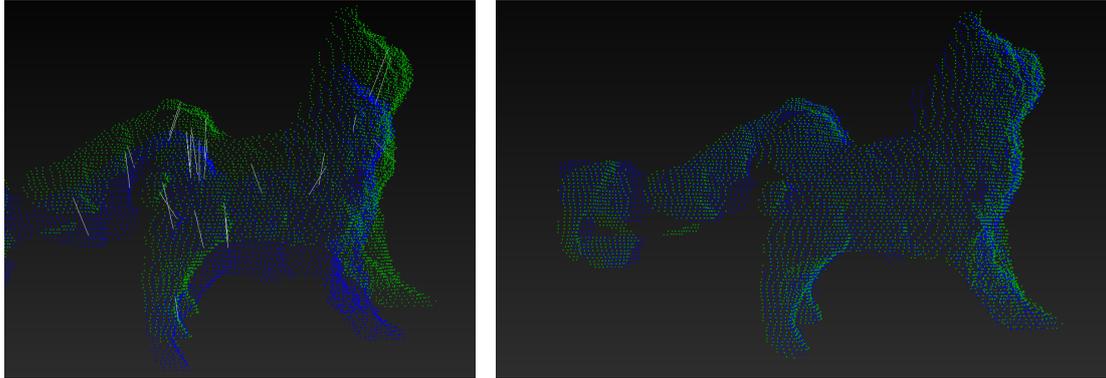
図 10: ISS 特徴点の検出例

3.3 特徴点検出とマッチング

提案システムでは，RANSAC による位置合わせのために，3次元点群間の対応点を利用する．従来システム [6, 8, 26] では，点群の位置合わせに RGB 画像で検出された特徴点と，その特徴点に関する特徴量が用いられている．しかし，RGB 画像上の特徴点や特徴量を用いる場合，例えば対象物が単一の色から構成される場合などにおいて，対応点が得られない可能性があり，3次元形状の取得がテクスチャに依存しないという RGB-D データの利点を損なう．そこで，本研究では3次元の特徴点，および特徴量を用いて位置合わせを行う．

3次元形状に用いられる特徴検出法として，点群データに用いる ISS (Intrinsic Shape Signatures) [9] や，デプス画像に用いる NARF (Normal Aligned Radial Feature) [27] などが提案されている．また RGB 画像に用いられる特徴点検出手法を3次元形状に拡張した手法も提案されている [28]．Filipe ら [29] はこれら特徴検出法の再現性について検証し，ISS [9] のスコアが最も高いことを示した．提案システムでは，RGB-D カメラを利用して点群を取得することから，これらすべての手法が利用可能であるが，Filipe らの結果に従って特徴検出手法として ISS を採用する (図 10)．

3次元点群中の任意の点のマッチングに用いられる特徴量として，現在までに PFH (Point Feature Histogram) [30]，FPFH (Fast Point Feature Histogram) [31]，



(a) CSHOT による対応付け (白線)

(b) 対応による位置合わせ結果

図 11: 対応付けによる位置合わせ結果例

SHOT (Signature of Histograms of Orientations) [32], CSHOT (Color SHOT) [10] などが提案されている。これらの特徴量は注目点の周辺にある法線分布状況をヒストグラムとして記述し、3次元点群による形状特徴をユニークかつ再現性を持って記述することができる。これらのうち、CSHOT 特徴量は法線方向のみではなく、周辺の色変化も記述した特徴量を抽出する。本研究では、対象物に形状的特徴が欠如している場合を考慮し、CSHOT 特徴量 [10] を用いて特徴点のマッチングを行う。

提案システムでは、対応点に対する RANSAC によって 3次元点を位置合わせする。このとき、対応点の誤りにより位置合わせの誤りが生じる可能性がある。そこで、以下の 2段階の処理により対応点の誤りを低減する。

RGB-D ストリームから得られる n 番目と m 番目の 3次元点群から検出された ISS 特徴点の集合をそれぞれ $X'_n = \{\mathbf{x}'_{ni} | i = 1, \dots, |X'_n|\}$ と $X'_m = \{\mathbf{x}'_{mj} | j = 1, \dots, |X'_m|\}$ とする。提案システムでは第 1 段階として、 \mathbf{x}'_{ni} に対応する CSHOT 特徴量の 2 近傍を X'_m に含まれる全特徴点に対応する CSHOT 特徴量から探索する。得られた第 1 近傍と第 2 近傍との CSHOT 特徴量空間におけるユークリッド距離の比が閾値 θ_{corr} 以上であれば、 \mathbf{x}'_{ni} とその第 1 近傍 \mathbf{x}'_{mj} を対応点候補とする。

ここで、一般に \mathbf{x}'_{ni} の近傍点が \mathbf{x}'_{mj} の場合においても、 \mathbf{x}'_{mj} の近傍点は \mathbf{x}'_{ni} とは異なる可能性がある。この時、 \mathbf{x}'_{mj} の近傍点が \mathbf{x}'_{ni} となれば、 \mathbf{x}'_{ni} と \mathbf{x}'_{mj} の近傍に他の点が存在しないと言える。第 2 段階では、第 1 段階で得られた対応点候

補 \mathbf{x}'_{ni} と \mathbf{x}'_{mj} について, \mathbf{x}'_{mj} の CSHOT 特徴量空間における近傍点を X'_n から探索し, \mathbf{x}'_{ni} が第 1 近傍として得られた場合にのみ \mathbf{x}'_{ni} と \mathbf{x}'_{mj} を対応点とする (図 11). 提案システムでは, 逐次的に 3 次元点を位置合わせするため, X'_n と $m < n$ となる全ての X'_m について上記の対応点探索を行う.

3.4 RANSAC を用いた位置合わせ

提案システムでは, n 番目の 3 次元点群 Y_n をそれまでに位置合わせされた点群集合 P_{n-1} に対して, 剛体運動を仮定した RANSAC[11] を用いて位置合わせすることで RGB-D ストリームから得られるすべての点群を逐次的に統合する. ここで, P_{n-1} は次式のように 3 次元点群とそれぞれに対応した剛体変換 (回転行列と並進ベクトル) の集合とする.

$$P_{n-1} = \{(Y_m, \mathbf{R}_m, \mathbf{t}_m) | m = 1, \dots, n-1\} \quad (1)$$

ただし, 最初の点群 Y_1 の座標系を世界座標系とし ($\mathbf{R}_1 = \mathbf{I}, \mathbf{t}_1 = \mathbf{0}$), 他の点群についてはこの座標系への変換を求める. このとき, 3 次元点群 Y_n を Y_{n-1} に対して位置合わせる単純な方法では, 誤差の蓄積により点群にズレが生じる可能性がある. そこで, 提案システムは P_{n-1} に含まれるすべての点群に対して RANSAC により Y_n を位置合わせし, 最もインライア数が多いものを採用することにより, 誤差の蓄積を低減する. 位置合わせの流れを図 12 に示す.

まず提案システムは, Y_n を P_{n-1} に含まれる点群 Y_m に対して位置合わせした際のインライア数を算出する. 前節で得られた Y_n と Y_m の対応点の集合をそれぞれ Y'_n と Y'_m とする. ここで, \mathbf{y}'_{ni} と \mathbf{y}'_{mi} が対応するものとする. 剛体運動を仮定した RANSAC では, 対応点からランダムに 3 点を選択し, 次式の最小自乗問題を解くことでこれら 3 点を一致させる剛体変換 (\mathbf{R}, \mathbf{t}) を求める.

$$\arg \min_{\mathbf{R}, \mathbf{t}} \sum_{i=1}^3 \|(\mathbf{R}_m \mathbf{y}'_{mi} + \mathbf{t}_m) - (\mathbf{R} \mathbf{y}'_{ni} + \mathbf{t})\|^2 \quad (2)$$

この最小自乗問題は特異値分解によって解くことができる [11].

続いて，得られた (\mathbf{R}, \mathbf{t}) によって元の点群 Y_n を変換する．変換後の点群を $\bar{Y}_n = \{\bar{\mathbf{y}}_{ni} | i = 1, \dots, |Y_n|\}$ とすると， $\bar{\mathbf{y}}_{ni}$ は次式により与えられる．

$$\bar{\mathbf{y}}_{ni} = \mathbf{R}\mathbf{y}_{ni} + \mathbf{t} \quad (3)$$

\bar{Y}_n に含まれるすべての点について， Y_m を $(\mathbf{R}_m, \mathbf{t}_m)$ で変換した点群 \bar{Y}_m から最近傍点を探索する． \bar{Y}_n に含まれるある3次元点 $\bar{\mathbf{y}}_{ni}$ の最近傍点を $\bar{\mathbf{y}}_{mj}$ とすると， θ_{inlier} を閾値として，次式を満たすときに $\bar{\mathbf{y}}_{ni}$ をインライアとする．

$$\min_{\bar{\mathbf{y}}_{mj}} \|\bar{\mathbf{y}}_{ni} - \bar{\mathbf{y}}_{mj}\| < \theta_{\text{inlier}} \quad (4)$$

ただし，上記の最小化は P_{n-1} に含まれるすべての Y_m から変換した \bar{Y}_m の $\bar{\mathbf{y}}_{mj}$ について求める．以上によって得られるインライア数を RANSAC における評価尺度として用いる． \bar{Y}_n に含まれるすべての3次元点に対して得られるインライア数を評価尺度として用いることにより，3.3 節の手順により得られる対応点に誤対応が多く含まれる場合においても，位置合わせの誤りを低減する．

提案システムは， Y_n と Y_m に対して RANSAC により，上記の評価尺度を最大にする剛体変換を求める処理を P_{n-1} に含まれるすべての Y_m に対して繰り返し，インライア数をもっとも大きくなる点群に対する変換 (\mathbf{R}, \mathbf{t}) を求める．さらに，この (\mathbf{R}, \mathbf{t}) により決定されたすべてのインライア $\bar{\mathbf{y}}_{ni}$ を用いて剛体変換を再推定して $(\mathbf{R}_n, \mathbf{t}_n)$ を決定する (図 11)．

3.5 把持部を考慮した位置合わせの高精度化

提案システムでは，点群の位置合わせを高精度化するために ICP アルゴリズム [25] を利用する．ICP アルゴリズムは，最近傍点による対応付けと，対応点からの剛体運動変換推定を繰り返すことで剛体変換を求める手法である．しかし，ICP アルゴリズムは，結果が入力点群の初期位置に大きく依存し，さらにノイズに対しても脆弱である．そこで，提案システムは前節で得られた点群の位置合わせ結果を利用し，対象物の形状をボリュームとして推定する．さらに，このボリュームに基づいて把持部など非対象物の点群を除去した後に ICP アルゴリズムで位置合わせを高精度化する．

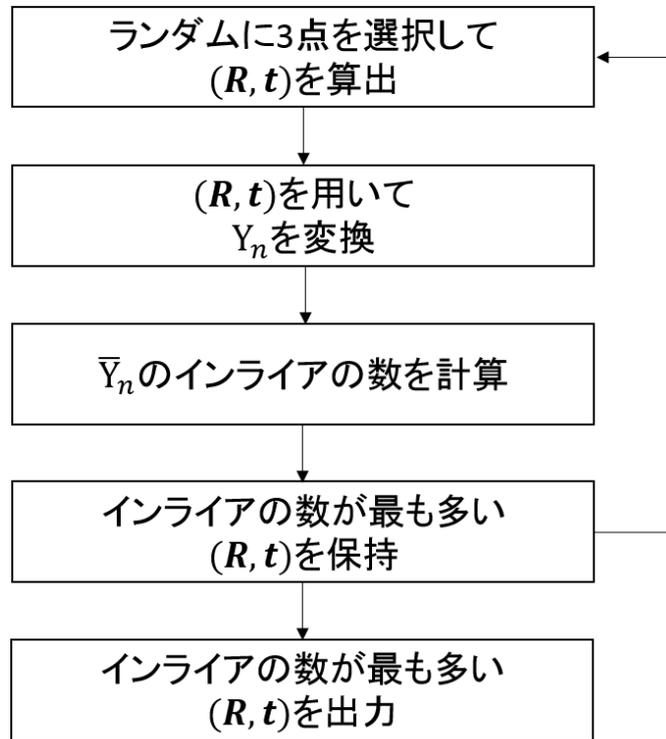


図 12: RANSAC による点群位置合わせの流れ

対象物の形状を推定するために、符号付き距離場 (Signed Distance Field, SDF) [12] を利用する。SDF は、複数のデプス画像から対象物の形状を推定する手法である。ここでは、世界座標系の適切に位置にボリューム $V = \{\mathbf{v}_k | k = 1, \dots, |V|\}$ を設定し、それぞれのボクセルに対応する SDF を f_k とする。また、RGB-D カメラの内部パラメータを既知とし、RGB-D カメラから得られた点群 Y_n に対応するデプス画像を D_n とする。ボクセル \mathbf{v}_k をデプス画像 D_n に投影したときの座標 (u_{nk}, v_{nk}) 、および RGB-D 座標系における \mathbf{v}_k のデプス値 d_{nk} は、RGB-D カメラの内部パラメータと、前節で得られた $(\mathbf{R}_n, \mathbf{t}_n)$ を用いて決定される。これらを用いて f_k は次式により得られる。

$$f_k = \sum_{n=1}^N d_{nk} - D_n(u_{nk}, v_{nk}) \quad (5)$$

ただし、 $D_n(u_{nk}, v_{nk})$ はデプス画像 D_n の (u_{nk}, v_{nk}) における値を表す。

SDFにおけるデプス値の差 $d_{nk} - D_n(u_{nk}, v_{nk})$ は、対象物の内部にあるボクセルに対して常に負の値となる一方、把持部は対象物を回転・移動させる過程で変形・移動することから、把持部に対応するボクセルは正と負の値が得られる。従って、十分なデータが計測され、把持部が頻繁に持ち換えられる場合、 f_k が負となるボクセルが対象物の形状を表すと考えられる。つまり、対象物のボリューム V_{obj} (図 13) は次式により与えられる。

$$V_{\text{obj}} = \{\mathbf{v}_k | f_k < 0, k = 1, \dots, |V|\} \quad (6)$$

次に、求めた対象物の形状ボリューム V_{obj} を利用して、入力点群から対象物に含まれない点群を除去する。位置合わせされた点群集合 $P_N = \{(Y_n, \mathbf{R}_n, \mathbf{t}_n) | n = 1, \dots, N\}$ の各点 \mathbf{y}_{ni} が対象物上の点であるかを判断する場合、点 \mathbf{y}_{ni} を $(\mathbf{R}_n, \mathbf{t}_n)$ により世界座標系に変換した点 $\bar{\mathbf{y}}_{ni}$ と対象物の形状ボリュームの距離 $d(V_{\text{obj}}, \bar{\mathbf{y}}_{ni})$ を求め、この距離が閾値 θ_{obj} 以下である場合、点 \mathbf{y}_{ni} を対象物の点群に属すると判断する。ここで、点 $\bar{\mathbf{y}}_{ni}$ と対象物の形状ボリュームの距離 $d(V_{\text{obj}}, \bar{\mathbf{y}}_{ni})$ は、次式のように定義する。

$$d(V_{\text{obj}}, \bar{\mathbf{y}}_{ni}) = \min_{\mathbf{v}_k \in V_{\text{obj}}} \|\bar{\mathbf{y}}_{ni} - \mathbf{v}_k\| \quad (7)$$

以上の手順に従い得られた点群のみを用いて ICP アルゴリズムで位置合わせを高精度化する。ここで最終的に得られた点群集合を $P'_N = \{(Y'_n, \mathbf{R}'_n, \mathbf{t}'_n) | n = 1, \dots, N\}$ とする。

3.6 把持部を除去した対象物のモデル生成

前節の処理により更新された剛体変換を用いて、SDFにより改めてボリュームを求める。また、各ボクセルに対して RGB 画像に基づいてその色を決定する。

ボクセルの色を決定する際は、対象ボクセルについて各 RGB 画像から色を収集し、それらの平均値をそのボクセルの色とする。具体的には、ボクセル \mathbf{v}_k の色 \mathbf{c}_k を決定する際は、各 RGB 画像においてボクセル \mathbf{v}_k が観測されているかを判定する。まず、剛体変換 $(\mathbf{R}'_n, \mathbf{t}'_n)$ と既知の内部パラメータを利用して、ボクセル \mathbf{v}_k をデプス画像 D_n に投影する。前節で述べた手法と同様に、投影された座

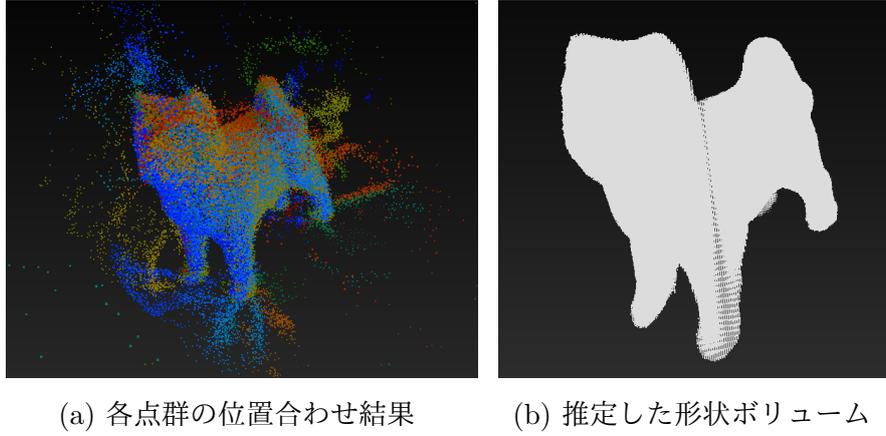


図 13: 対象物の形状ボリューム推定例

標を (u'_{nk}, v'_{nk}) とし, \mathbf{v}_k のカメラ座標系におけるデプス値を d'_{nk} とする. また, (u'_{nk}, v'_{nk}) における D_n の値を $D_n(u'_{nk}, v'_{nk})$ とし, 次式を満たすときにボクセル \mathbf{v}_k が n 番目の RGB 画像に観測されたものと判定する.

$$|d'_{nk} - D_n(u'_{nk}, v'_{nk})| < \theta_{\text{visible}} \quad (8)$$

ただし θ_{visible} は事前に決定された閾値である. 続いて, 上式を満たす n の集合を W_k とし, 対応する RGB 画像 \mathbf{I}_n を用いて, 次式によって \mathbf{c}_k を求める.

$$\mathbf{c}_k = \frac{1}{|W_k|} \sum_{n \in W_k} \mathbf{I}_n(\mathbf{v}_k) \quad (9)$$

ただし, $\mathbf{I}_n(\mathbf{v}_k)$ は RGB 画像 \mathbf{I}_n に投影された \mathbf{v}_k の座標における色を表す.

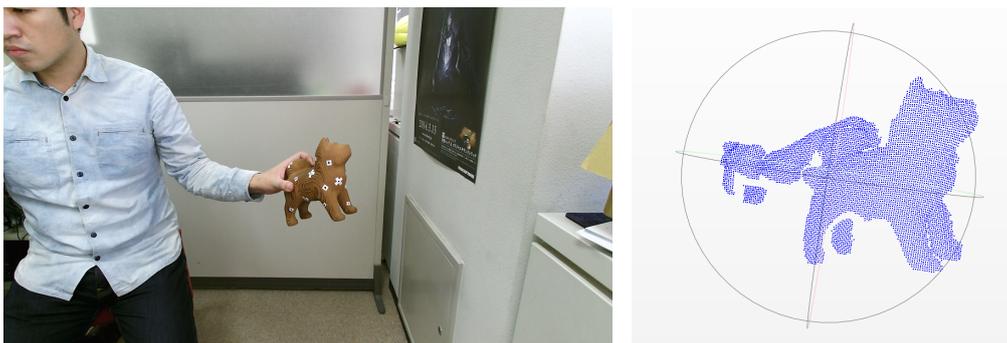
4. 実験

提案システムの有効性を検証するために，提案システムを利用したモデル生成実験と生成モデルを用いた把持部除去率評価実験を行った．以下では各実験の内容と結果を詳述する．

4.1 提案システムを用いたモデル生成実験

本実験は，提案システムを用いて実物体のモデル生成が可能であることを確認することを目的とする．実験では，提案システムのインタフェースを用いて，ユーザが対象物を把持し，RGB-D カメラの前方で回転・移動することにより対象物の全方位から対象物のデプス画像と RGB 画像を撮影する．さらに，撮影したデータを用いて対象物の 3 次元モデルを生成する．

対象物の撮影には Microsoft Kinect V2 を利用した．表 1 にこの RGB-D カメラの仕様を示す．また，これに関する各種内部パラメータは Microsoft Kinect SDK から取得した．撮影インタフェースで設定する計測の有効範囲は Kinect V2 の RGB カメラの光軸から 0.2 m 以内，深度 0.5 m から 1 m までの円筒状空間とした．図 14 に得られた RGB 画像の一例と，有効範囲から抽出した点群を示す．3 章で言及した各閾値は表 2 に示すように設定した．実験に用いた 8 種類の対象物 A-H を図 15-22 中 (a) に，入力画像の例を図中 (b) に示す．



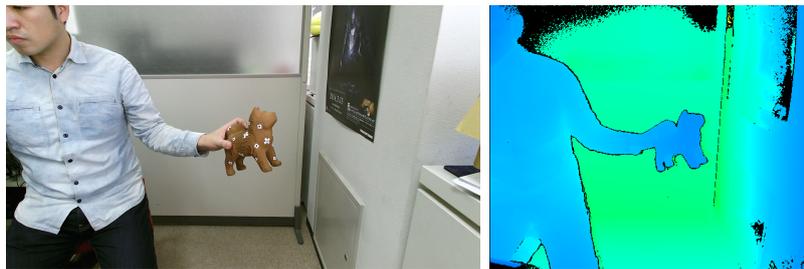
(a) 撮影した RGB 画像

(b) 有効範囲内の点群

図 14: 入力の一例



(a)

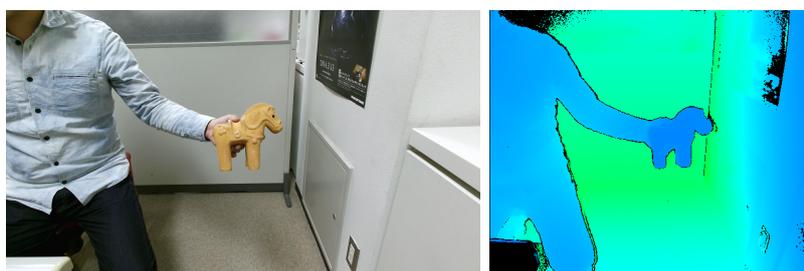


(b)

図 15: 対象物 A (a) と入力画像例 (全 150 フレーム)(b)



(a)



(b)

図 16: 対象物 B (a) と入力画像例 (全 170 フレーム)(b)



(a)

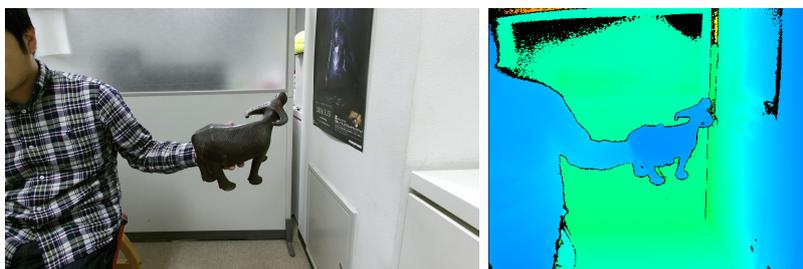


(b)

図 17: 対象物 C (a) と入力画像例 (全 150 フレーム)(b)



(a)



(b)

図 18: 対象物 D (a) と入力画像例 (全 120 フレーム)(b)



(a)

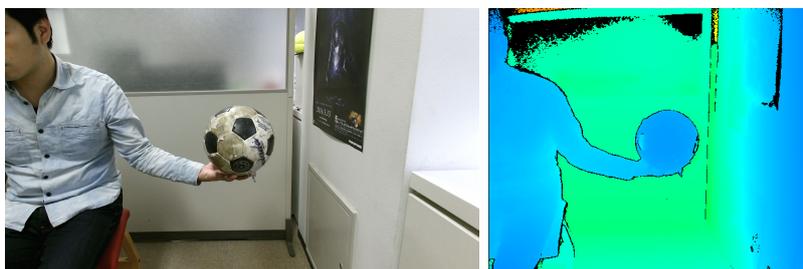


(b)

図 19: 対象物 E (a) と入力画像例 (全 170 フレーム)(b)

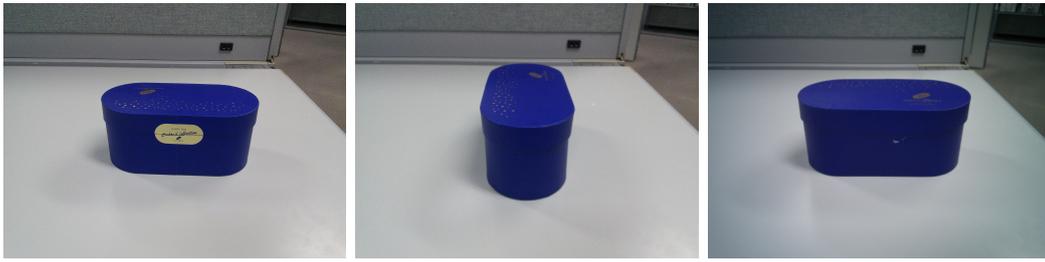


(a)



(b)

図 20: 対象物 F (a) と入力画像例 (全 140 フレーム)(b)



(a)

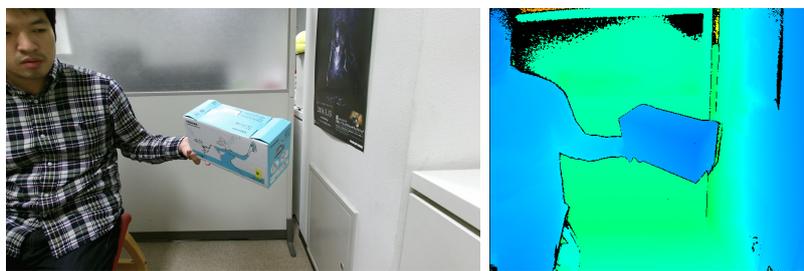


(b)

図 21: 対象物 G (a) と入力画像例 (全 130 フレーム)(b)



(a)



(b)

図 22: 対象物 H (a) と入力画像例 (全 170 フレーム)(b)

表 1: Microsoft Kinect V2 のカメラ規格

項目	規格
RGB カメラ解像度	1920×1080
RGB カメラフレームレート	30 fps
デプスカメラ解像度	512×424
デプスカメラフレームレート	30 fps
最小奥行距離	0.5 m
最大奥行距離	4.5 m

表 2: 本研究に用いる閾値

項目	設定値
θ_{corr}	1.1
θ_{inlier}	10 mm
θ_{obj}	10 mm
θ_{visible}	5 mm

モデル生成の結果，対象物 A-D についてはモデルが生成された．図 23-26 では各対象物を一辺 1 mm の単位ボクセルで生成したモデルを示す．一方，対象物 E-H については位置合わせが破綻したため，形状を再現するモデルが得られなかった (図 27)．位置合わせが破綻した原因は以下の二点にまとめられる．

1. 対象物の形状を正しく計測不能である．
2. 重複の形状とテクスチャに対する特徴点の誤対応が発生する．

1 の場合では，対象物 E の表面に光沢があり，その表面の相互反射によりデプスカメラが正確に形状を計測できなかったため，位置合わせが破綻した．これは，デプスカメラの計測原理による問題で，異なる計測原理を利用したデプスカメラ

に変えることで解決できると考える。2の場合では、対象物 F-H が対称であるため、重複の形状とテクスチャが多数存在する。これが特徴点の誤対応を発生させ、誤った位置合わせする状況が見られる。これは、提案システムでは対応付けするときに重複の形状とテクスチャを考慮していないのが原因だと考える。

また、モデル生成に成功した対象物について、図 26 のように、対象物以外の色が割り当てられる問題が見られた。これは、ボクセルの RGB 画像上への投影位置に誤差があったことが原因であると考えられる。同モデルの口に該当する箇所の形状に凹みがあることも、一部のフレームにおける位置合わせの誤差が大きいということを支持するものであると考えられる。

4.2 持ち手の除去率評価実験

本実験は、提案システムを用いて把持部などの対象物に属さない点群が除去可能であるかを検証することを目的とする。ここではまず、点群 Y_n に対応する RGB 画像上で対象物に属する画素に対して手動でラベルを割り当て (図 28)、このラベルに基づいて Y_n のうち対象物に属さない 3 次元点 \mathbf{z}_{ni} を特定した。次に、得られた 3 次元点 \mathbf{z}_{ni} を提案手法によって推定された剛体変換 $(\mathbf{R}_n, \mathbf{t}_n)$ により世界座標系に変換した点群を $\bar{Z}_n = \{\bar{\mathbf{z}}_{ni} | i = 1, \dots, |Z_n|\}$ を用いて、これらが除去可能であるかの評価基準として以下で定義される真陽性率 (True Positive Rate; TPR) と偽陽性率 (False Positive Rate; FPR) を算出した。

$$\text{TPR} = \frac{\bar{Z}_n \text{ のうち、近傍に } V_{\text{obj}} \text{ に属するボクセルが存在しない点の数}}{\bar{Z}_n \text{ に含まれる点の数}} \quad (10)$$

$$\text{FPR} = \frac{\bar{Y}_n - \bar{Z}_n \text{ のうち、近傍に } V_{\text{obj}} \text{ に属するボクセルが存在しない点の数}}{\bar{Y}_n - \bar{Z}_n \text{ に含まれる点の数}} \quad (11)$$

ただし、 \bar{Y}_n は Y_n を剛体変換 $(\mathbf{R}_n, \mathbf{t}_n)$ により世界座標系に変換した点群であり、 V_{obj} は提案システムによって得られた対象物のボクセルモデルである。また、ボクセルが存在するとは、 $\bar{\mathbf{z}}_{ni}$ もしくは $\bar{\mathbf{y}}_{ni}$ の最近傍のボクセルまでの距離がある閾値以下であることを表す。

ここでは、入力された RGB 画像のうち、8 枚ごとにラベルを割り当て、最近傍のボクセルまでの距離における閾値を 0 mm から 200 mm の間で 0.1 mm ごと

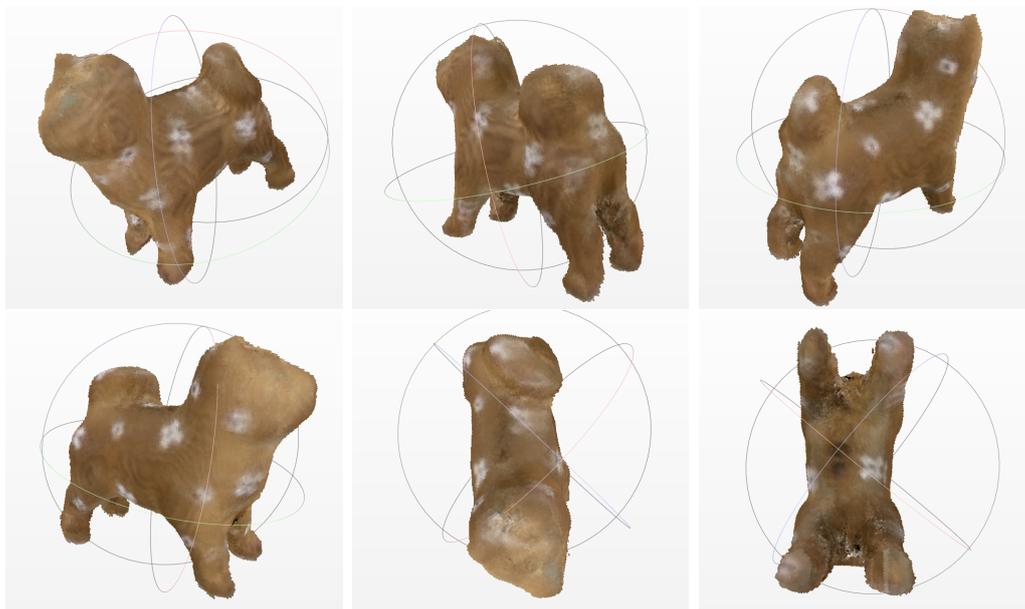


図 23: 対象物 A の復元結果

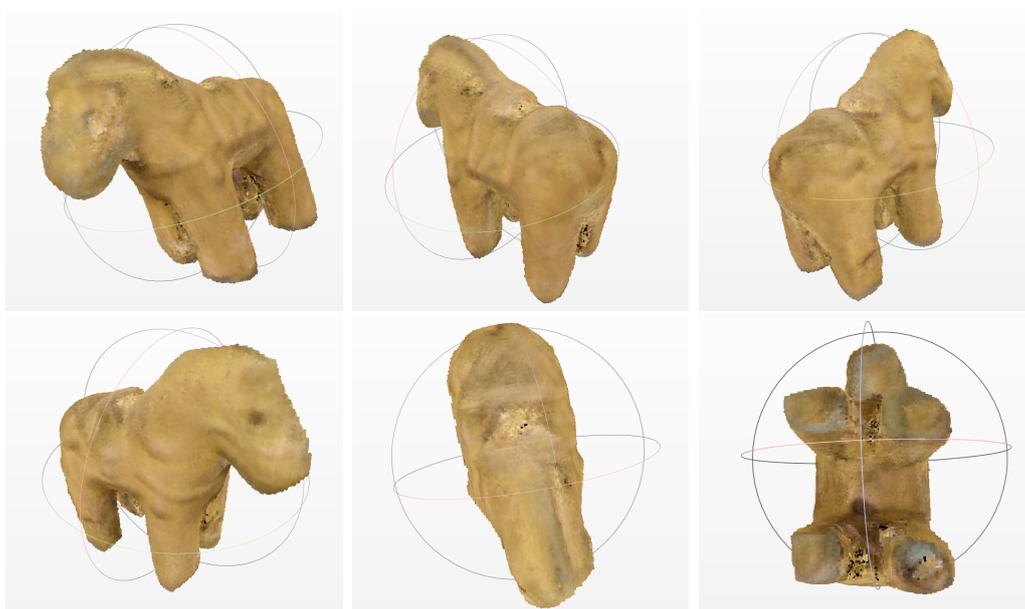


図 24: 対象物 B の復元結果

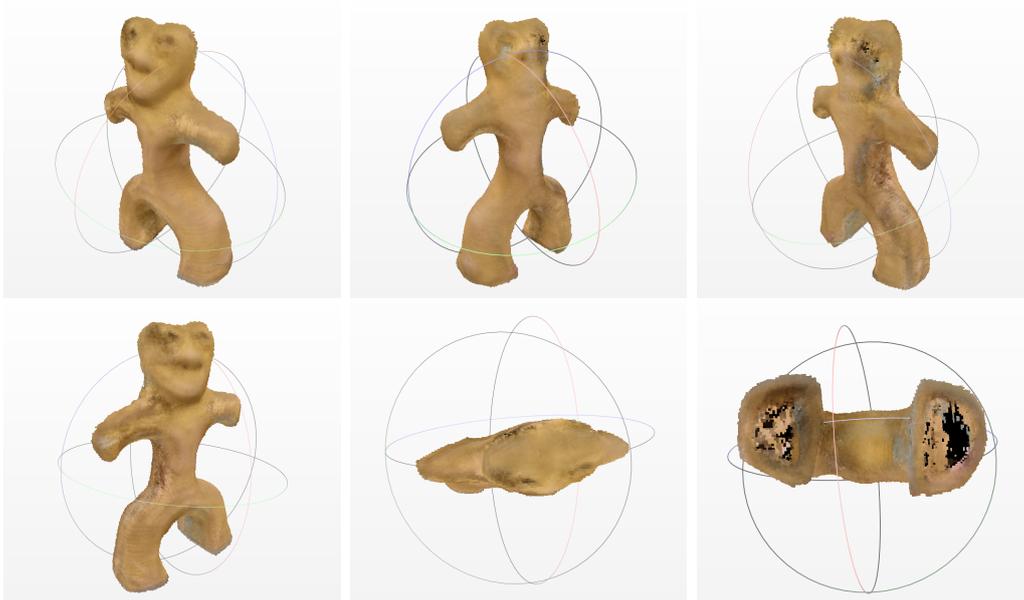


図 25: 対象物 C の復元結果

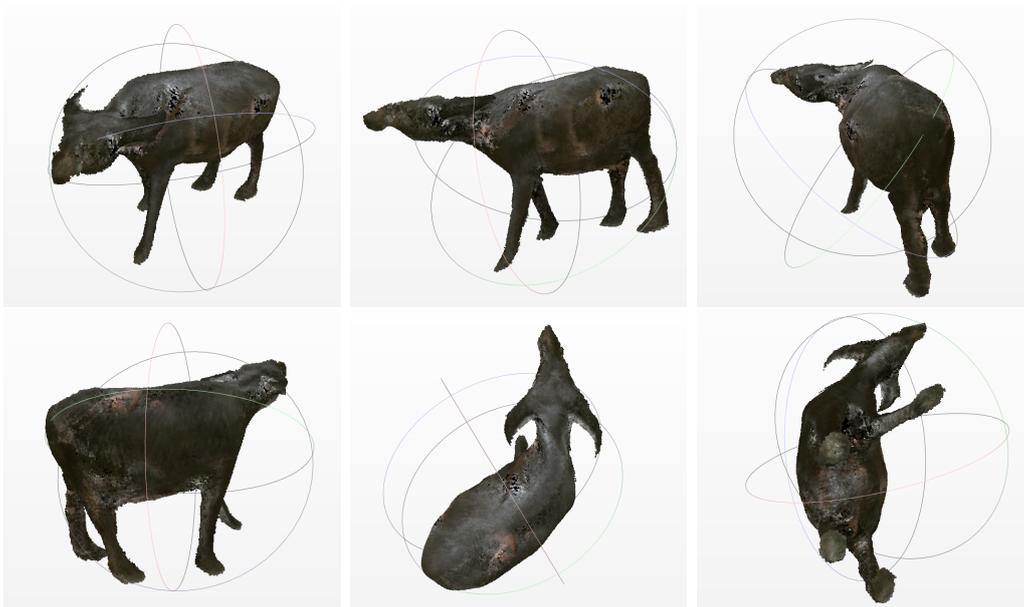


図 26: 対象物 D の復元結果

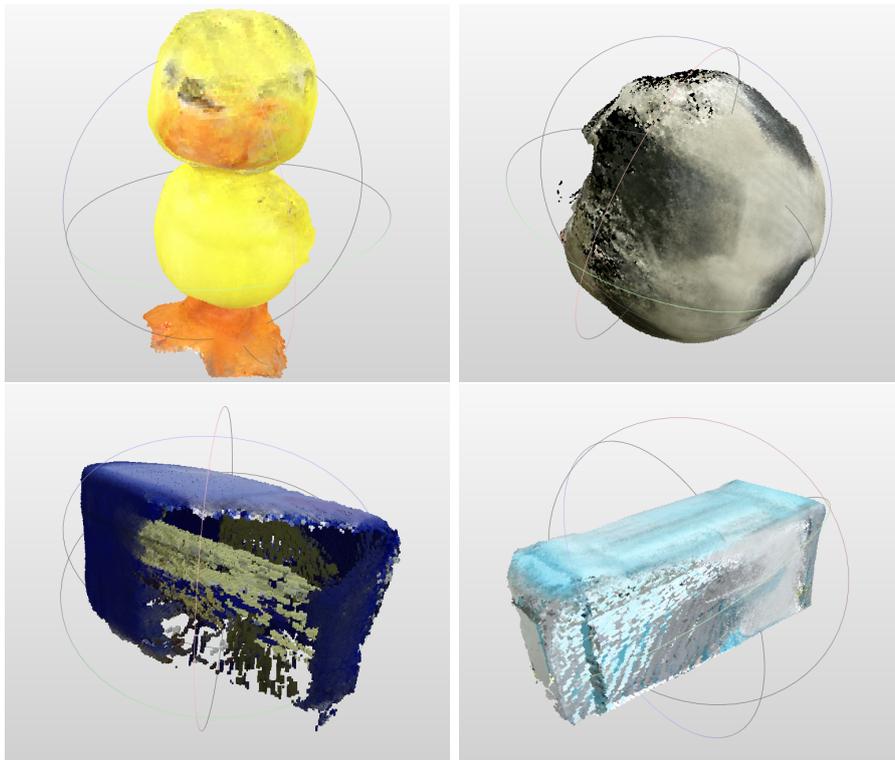


図 27: 対象物 E, F, G, H の復元結果



図 28: 対象物 A に対するラベルの例

に上式により FPR と TPR を算出し、ROC 曲線を生成した。またここでは、特徴点マッチングによる位置合わせ結果から求めた対象物のボクセルモデルと、その後 ICP を利用して高精度化した位置合わせ結果から求めた対象物のボクセルモデルの双方について結果を算出し、比較した。

4.1 節においてモデル生成に成功した対象物に対し、前述の ROC 曲線を生成した結果を図 29-33 に示す。図 29-33 では、特徴点マッチングのみによるボクセルモデルに対する ROC 曲線 (青線, feature と表記) と、その後 ICP を利用して求めたボクセルモデルに対する ROC 曲線 (赤線, feature+ICP と表記) を示している。閾値を 1.0mm に設定する場合、特徴点マッチングのみと ICP で高精度化した位置合わせ結果に対し、対象物 A, B, C, D における (TPR, FPR) の数値はそれぞれ (0.9137, 0.1866) と (0.9084, 0.1821) であった。この結果から、多くの場合において正しく点群を除去できたことが分かる。また、4.1 節でモデルの生成に使用したパラメータにおける (TPR, FPR) の数値は (0.7568, 0.0258) であった。

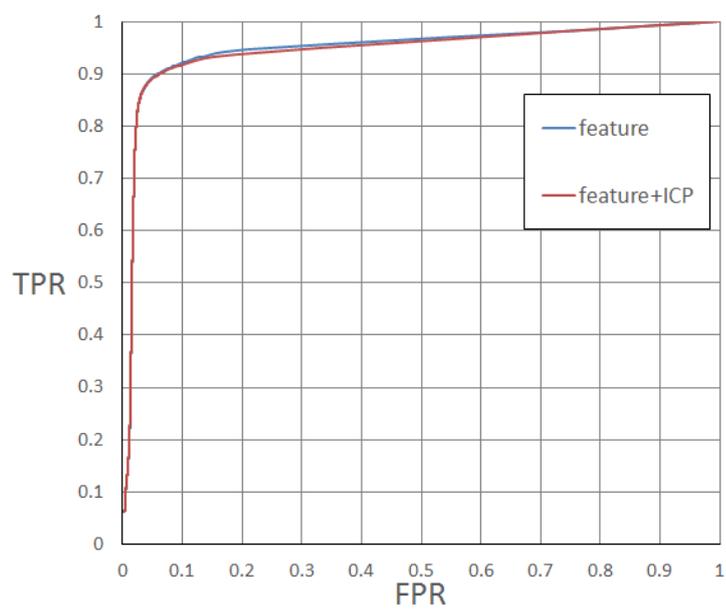


図 29: 対象物 A における ROC 曲線

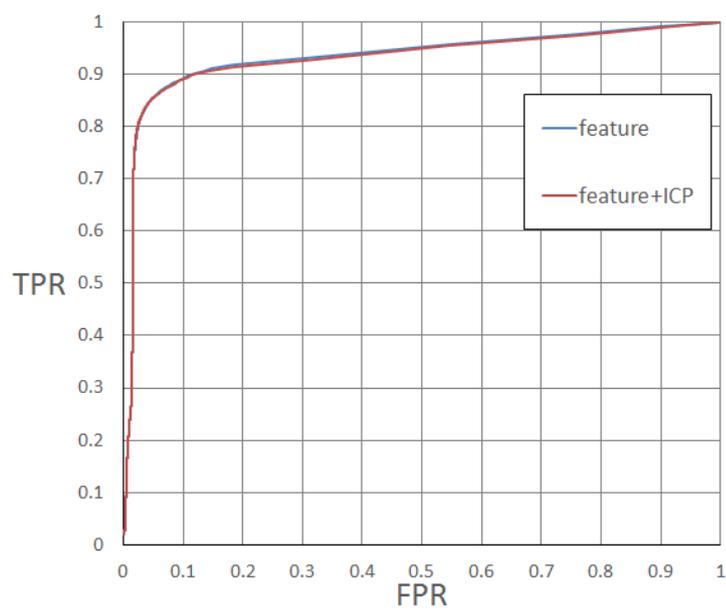


図 30: 対象物 B における ROC 曲線

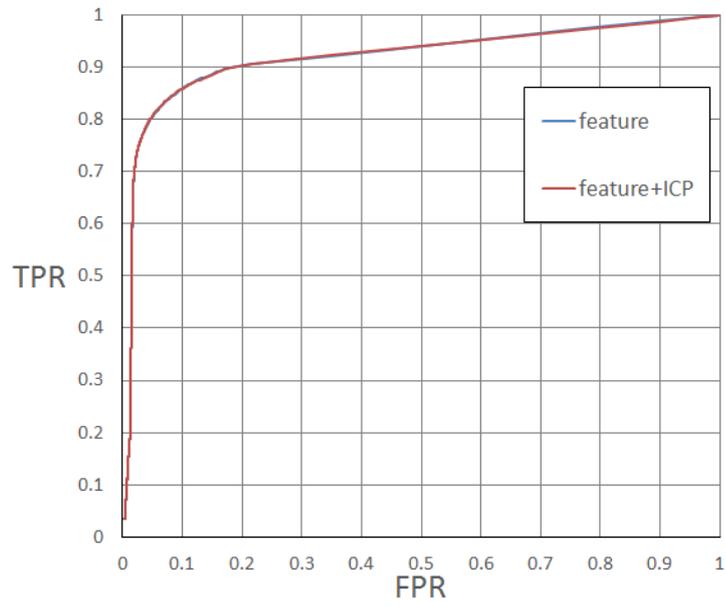


図 31: 対象物 C における ROC 曲線

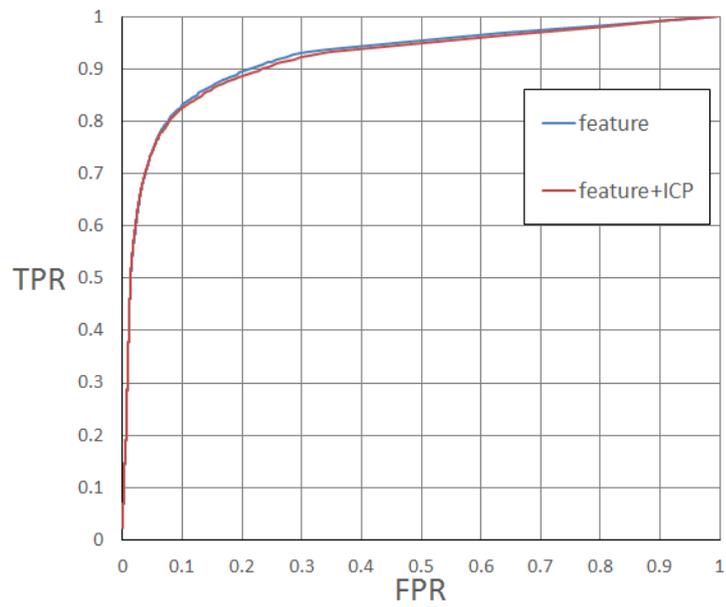


図 32: 対象物 D における ROC 曲線

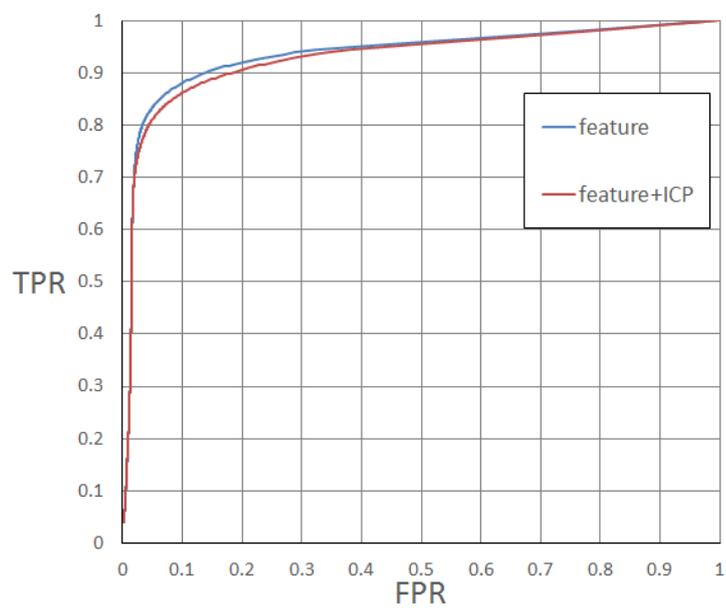


図 33: 対象物 A, B, C, D における ROC 曲線



図 34: 撮影角度による色変化の例

5. まとめ

本論文では素手で対象物を把持し、回転・移動させながら撮影した RGB-D ストリームを入力とした 3次元モデル生成システムを提案した。提案システムでは、対象物を回転・移動させる過程で把持部は変形・移動するという性質に着目し、3次元特徴点による対応点を利用して各点群の剛体変換を求め、剛体変換により位置合わせされない点を把持部として除去し、対象物の 3次元モデルを生成する。

モデル生成実験では、把持部を除去した 3次元モデルを生成可能であることを明らかにした。また、把持部の除去実験では、提案システムが多くの場合において正しく把持部を除去できるという結果を示した。これにより、提案システムは一般のユーザが手で把持可能な剛体小物体の 3次元モデルを生成する際の負担を低減できると考える。ただし、モデル生成実験では表面に光沢のある物体、または形状とテクスチャが重複する物体に対して正確なモデルを生成できなかった。前者はデプスカメラの計測原理による問題であり、原理の異なるデプスカメラに変更することにより改善されると考える。後者については重複の形状とテクスチャを考慮した対応付け手法の適用が必要だと考える。

また、提案システムでは生成モデルの色に RGB 画像の平均値を用いるが、撮影方式により、各フレームの対応点における色にばらつきが生じる (図 34) ため、同一対象物のモデルの色が大きく変わる可能性がある。これは、物体の本来の色成分と照明による明度変化を分析できる Intrinsic Image Decomposition [33] などの手法を利用し、本来の色成分のみを用いてモデルに適用することにより解決できると考える。

謝辞

本研究を進めるに当たり，終始暖かくご指導，ご鞭撻頂いた視覚情報メディア研究室 横矢直和 教授に心より感謝申し上げます。また，本学での研究におけるあらゆる面で，多大なご助言，ご協力を賜りましたことを深く感謝致します。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，本研究を進めるに当たり，終始細やかなご指導，ご助言頂いた視覚情報メディア研究室 佐藤智和 准教授に厚く御礼申し上げます。本研究へのご助言，ご協力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 助教に深く感謝致します。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。あらゆる面において，多大なるご助言を頂いた視覚情報メディア研究室 青砥隆仁氏に深く感謝いたします。本研究の遂行にご助言をくださり，実験等にご協力して頂きました，視覚情報メディア研究室方々に深く感謝致します。そして，留學生活において多大なるご協力を頂いた NAIST の方々に深く感謝致します。最後に，暖かく見守って下さいました両親に心より感謝の意を表します。

参考文献

- [1] Hippo tusk harpoon | african fossils. <http://africanfossils.org/tools/hippo-tusk-harpoon>. 最終閲覧 2015.01.12.
- [2] IKEA カタログ 2015. http://onlinecatalogueasia.ikea.com/JP/ja/IKEA_Catalogue/. 最終閲覧 2015.01.12.
- [3] 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 *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 83–88, 2013.
- [4] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. *ACM Transactions on Graphics.*, pp. 438–446, 2002.
- [5] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 127–136, 2011.
- [6] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An evaluation of the RGB-D SLAM system. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2012.
- [7] M. Krainin, P. Henry, X. Ren, and D. Fox. Manipulator and object tracking for in-hand 3D object modeling. *International Journal of Robotics Research*, Vol. 30, No. 11, pp. 1311–1327, 2011.
- [8] T. Weise, T. Wismer, B. Leibe, and L. J. V. Gool. Online loop closure for real-time interactive 3D scanning. *Computer Vision and Image Understanding*, Vol. 115, No. 5, pp. 635–648, 2011.

- [9] Y. Zhong. Intrinsic shape signatures: A shape descriptor for 3D object recognition. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 689–696, 2009.
- [10] F. Tombari, S. Salti, and L. Di Stefano. A combined texture-shape descriptor for enhanced 3D feature matching. In *Proceedings of the IEEE International Conference on Image Processing*, pp. 809–812, Sept 2011.
- [11] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2004.
- [12] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *Proceedings of the ACM International Conference on Computer Graphics and Interactive Techniques*, pp. 303–312, 1996.
- [13] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 519–528, 2006.
- [14] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building rome in a day. In *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, 2009.
- [15] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 150–162, 1994.
- [16] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3121–3128, 2011.
- [17] K. M. Cheung, S. Baker, and T. Kanade. Visual hull alignment and refinement across time: A 3D reconstruction algorithm combining shape-from-

- silhouette with stereo. In *Proceedings of the IEEE Computer Society Computer Vision and Pattern Recognition*, p. 8, 2003.
- [18] Q. Zhou and V. Koltun. Dense scene reconstruction with points of interest. *ACM Transactions on Graphics*, Vol. 32, No. 4, pp. 1–8, 2013.
- [19] N. D. Campbell, G. Vogiatzis, C. Hernández, and R. Cipolla. Using multiple hypotheses to improve depth-maps for multi-view stereo. In *Proceedings of the European Conference on Computer Vision*, pp. 766–779, 2008.
- [20] C. J. Taylor. Surface reconstruction from feature based stereo. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 184–190, 2003.
- [21] C. Wu. Towards linear-time incremental structure from motion. In *Proceedings of the International Conference on 3D Vision*, pp. 127–134, 2013.
- [22] C. Wu, S. Agarwal, B. Curless, and S. M. Seitz. Multicore bundle adjustment. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3057–3064, 2011.
- [23] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. In *Proceedings of the ACM International Conference on Computer Graphics and Interactive Techniques*, pp. 369–374, 2000.
- [24] 不殿健治, 佐藤智和, 横矢直和. ハンドヘルドビデオカメラを用いた撮影支援インタフェースを有するインタラクティブ三次元モデリングシステム. *日本バーチャルリアリティ学会論文誌*, pp. 599–608, 2005.
- [25] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. In *Proceedings of the British Machine Vision Conference*, pp. 662–670, 2001.
- [26] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGBD mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Pro-*

ceedings of the RGB-D: Advanced Reasoning with Depth Cameras Workshop, 2010.

- [27] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard. NARF: 3D range image features for object recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics*, 2010.
- [28] D. G. Lowe. Local feature view clustering for 3D object recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 682–688, 2001.
- [29] S. Filipe and L.A. Alexandre. A comparative evaluation of 3D keypoint detectors. In *proceedings of the 9th Conference on Telecommunications*, pp. 145–148, 2013.
- [30] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz. Persistent point feature histograms for 3D point clouds. In *Proceedings of the International Conference on Intelligent Autonomous Systems*, 2008.
- [31] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3D registration. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 1848–1853, 2009.
- [32] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *Proceedings of the European Conference on Computer Vision*, pp. 356–369, 2010.
- [33] Q. Chen and V. Koltun. A simple model for intrinsic image decomposition with depth cues. In *Proceedings of the IEEE International Conference on Computer Vision*, p. 8, 2013.