

博士論文

高精度ウィンドウマッチングに
基づく多視点 3 次元復元
に関する研究

情報基礎科学専攻

酒井 修二

Multi-View 3D Reconstruction Based on High-Accuracy Window Matching

by

Shuji SAKAI

Graduate School of Information Sciences, Tohoku University, 2015

Copyright © 2015 Shuji SAKAI,

All Rights Reserved.

目次

第 1 章	緒言	1
第 2 章	多視点 3 次元復元に関する基礎的考察	9
2.1	まえがき	9
2.2	多視点 3 次元復元	10
2.2.1	多視点画像の取得	10
2.2.2	カメラパラメータ推定	13
2.2.3	3 次元復元	17
2.3	多視点ステレオ	19
2.3.1	多視点ステレオアルゴリズムの分類	20
2.3.2	多視点ステレオにおけるウィンドウマッチング	23
2.4	むすび	26
第 3 章	多視点ステレオのための高精度ウィンドウマッチング	27
3.1	まえがき	27
3.2	位相限定相関法	28
3.2.1	位相限定相関法の基本原理	28
3.2.2	位相限定相関法の高精度化	34
3.3	多視点ステレオのための位相限定相関法に基づくウィンドウマッチング	37
3.3.1	ステレオビジョンにおける位相限定相関法の適用	39

3.3.2	正規化視差による位相限定相関関数の統合	42
3.3.3	位相限定相関法に基づくウィンドウマッチングによる奥行き推定	47
3.4	マッチングウィンドウの幾何補正による高精度化	51
3.4.1	3次元平面によるステレオ画像間の変形	51
3.4.2	マッチングウィンドウの幾何補正による画像変形の軽減	53
3.5	性能評価	56
3.5.1	実験用データセット	57
3.5.2	比較手法	62
3.5.3	奥行き推定精度の評価	65
3.5.4	計算コストの評価	79
3.6	むすび	81
第4章	高精度ウィンドウマッチングに基づく多視点ステレオアルゴリズム	83
4.1	まえがき	83
4.2	位相限定相関法に基づく多視点ステレオアルゴリズム	84
4.2.1	位相限定相関法と階層的探索を用いたデプスマップ生成	85
4.2.2	重み付きメディアンフィルタを用いたデプスマップの高精度化	91
4.2.3	グラフカットに基づく誤対応点除去	94
4.2.4	閾値処理に基づくアーチファクト除去	97
4.2.5	デプスマップからの3次元点群生成	101
4.3	精度評価実験	101
4.3.1	実験環境	101
4.3.2	3次元復元精度の評価	106
4.4	文化財の3次元復元への応用	121
4.4.1	3次元復元処理の流れ	125
4.4.2	瑞巖寺欄間の撮影条件	126
4.4.3	瑞巖寺欄間の3次元復元結果	127

4.5	むすび	127
第 5 章	結言	131
参考文献		135
謝辞		145

第 1 章

緒言

近年，計算機の技術的發展に伴い，データ容量の多い 3 次元データが様々な分野で利用されるようになった [1], [2]．例えば，産業の分野では，エレクトロニクス製品や自動車などの 3 次元情報による製品検査に，医療の分野では，外科や歯科における人体の形状の解析・診断に，芸術・考古学の分野では，文化遺産の保護を目的とした美術品や遺跡といった文化財のデジタルアーカイブに用いられている．それにより，3 次元データを取得するための 3 次元復元技術に対する需要が高まっている．物体の高精度な 3 次元形状を低いコストで取得することができれば，3 次元データを用いた処理は，さらに広範囲の分野において，さらに多用途に向けた発展が期待できる．

3 次元復元は，物体の 3 次元形状や構造を計算機上で復元する技術である．一般に，3 次元復元は，能動型と受動型の 2 つに分類することができる [1], [3], [4]．能動型の 3 次元復元は，対象物体にレーザやパターン光を投影し，その挙動をカメラにより観察することで物体の 3 次元形状を取得する [1], [3]．能動型の 3 次元復元の代表的なものとして，プロジェクタ・カメラによりステレオ計測を行う能動型ステレオや，光の発信から受信までの時間から距離計測を行う Time-of-Flight が挙げられる．一般に，能動型の 3 次元復元は，精度や安定性に優れ，現在実用化されている 3 次元復元技術の主流となっている．しかし，撮影装置が大型で重くなりやすく，撮影環境が限定されるといった問題がある．特に，復元対象が大規模な場合，大型の撮影装置により繰り返し撮影する必要があるため，撮影作業にかかる負担は大きい．また，能

動型の3次元復元では、撮影可能な対象物体の大きさや3次元復元精度が撮影装置に大きく依存するため、用途ごとに専用の撮影装置を開発する必要がある。これらの理由から、能動型の3次元復元は用途が限定される。

一方、受動型の3次元復元は、カメラの位置や焦点距離などの撮影条件を変化させながら複数枚の画像を撮影し、撮影されたカメラ画像の違いから物体の3次元形状を計算する [3], [4]。受動型の3次元復元の代表的なものとして、カメラの位置を変えながら画像を撮影する多視点3次元復元と、カメラの焦点距離を変えながら画像を撮影する Shape from Focus (もしくは、Shape from Defocus) が挙げられる [3]–[9]。中でも、多視点3次元復元は、撮影の自由度、および、精度が高い3次元復元技術である。多視点3次元復元は、カメラによる画像取得という非常に簡便な作業により入力データが取得できるため、撮影にかかる負担が少なく、撮影環境も限定されにくい。また、近距離用のレンズや遠距離用のレンズ、工業用カメラや汎用デジタルカメラなど、さまざまな撮影装置を利用可能であり、用途ごとに専用の撮影装置を開発する必要がない。そのため、多視点3次元復元は、産業における小さな工業部品から、文化財のデジタルアーカイブにおける建築物や都市全体といった大規模な対象まで、さまざまな対象を復元可能である。しかしながら、現在、精度やロバスト性の観点から、多視点3次元復元の実用化に対する課題は多い。もし、多視点3次元復元の精度・ロバスト性が向上すれば、3次元復元技術のより広範囲な実用化が期待できる。そこで、本論文では、多視点3次元復元の精度・ロバスト性を向上させ、より広範囲の用途に利用可能な3次元復元の実現を目標とする。

多視点3次元復元は、多視点画像の取得、カメラパラメータ推定、3次元復元の大きく3つのステップに分けられる [8], [9]。多視点画像の取得は、カメラを用いて異なる位置から複数枚の画像を撮影するステップであり、多眼カメラによる撮影と単眼カメラによる撮影がある。カメラパラメータ推定は、カメラの焦点距離や位置などの撮影条件を表すカメラパラメータを推定するステップであり、事前に校正する手法と、撮影された多視点画像から直接推定する手法がある。多視点3次元復元におけるカメラパラメータ推定は、原理的に能動型3次元復元の校正手法と同様であり、十分な精度でカメラパラメータ推定が行える手法が確立されている。そのため、多視点3次元復元の精度・ロバスト性の向上には、3次元復元のステップを改善する必要がある。多視点3次元復元における3次元復元のステップは、多視点画像とカメラパ

ラメータから物体の 3 次元形状を復元するステップであり, Structure from Motion (SfM), 視体積交差法 (もしくは, Visual-hull, Shape from Silhouette と呼ばれる), 多視点ステレオ (Multi-View Stereo: MVS) がある [3]. SfM は, カメラパラメータ推定と同時に 3 次元復元を行う手法であり, 通常, コーナー点などの特徴点のみの 3 次元復元を行う. そのため, 物体のおおまかな形状しか復元できず, その用途は限られる. 視体積交差法は, 多視点画像における対象物体のシルエットに基づく手法である. この手法は, 背景と前景の正確な分離が必須であり, 背景を抽出しやすいようにブルースクリーンなどの単純な背景で撮影する必要がある. また, 高精度な 3 次元復元のために必要な画像枚数が膨大になったり, 原理的に凹面の復元ができないといった問題がある. 多視点ステレオは, 多視点画像における対象物体の表面テクスチャに基づく手法である. SfM および視体積交差法に対して, 対象物体の詳細な形状を復元可能, 複雑な背景にも適用可能, 少ない枚数の画像からでも高精度な 3 次元復元が可能といった利点がある. 本論文では, 3 次元復元アルゴリズムの中でも精度・ロバスト性が高い, 多視点ステレオに着目し, 多視点ステレオの高精度化を目指す.

多視点ステレオアルゴリズムは, 様々な技術の複合として構成される [3]. 例えば, カメラ選択, 3 次元点群生成, 誤対応点除去, 3 次元メッシュ生成, 3 次元メッシュの最適化などである. Seitz らは, 多視点ステレオアルゴリズムを, (i) 3 次元ボリュームの最適化に基づくアルゴリズム [10]–[12], (ii) 3 次元メッシュの最適化に基づくアルゴリズム [13]–[16], (iii) 特徴領域拡張に基づくアルゴリズム [17]–[19], (iv) デプスマップ統合に基づくアルゴリズム [20]–[27] の 4 つに分類した [5]. いずれのアルゴリズムにおいても, 3 次元点群生成, 3 次元メッシュ生成, ボクセル最適化, メッシュ最適化などの, 3 次元座標を決定する処理に, 多視点画像間のウィンドウマッチング (画像全体ではなく画像中の局所領域同士のマッチング) が適用され, その性能が処理全体の性能に大きく影響する. そのため, 多視点ステレオにおいて, ウィンドウマッチングは, アルゴリズム全体の精度・ロバスト性を決定する最も重要な構成要素のひとつである. 多視点ステレオにおけるウィンドウマッチングの精度が向上すれば, より高精度な多視点ステレオが実現できる. このような背景のもと, 本論文では, 多視点ステレオにおけるウィンドウマッチング手法に着目する.

従来の多視点ステレオアルゴリズムでは, 多視点画像間のウィンドウマッチング手法とし

て、Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), 正規化相互相関 (Normalized Cross-Correlation: NCC) [28], DAISY に基づくマッチング [29] などが用いられている。多視点ステレオアルゴリズムにおいて、これらのマッチング手法は、3 次元形状のパラメータ (例えば、3 次元座標や法線ベクトルなど) の尤度を算出するために利用される。例えば、代表的な多視点ステレオアルゴリズムであるプレーンスイーピング [3], [5], [20] では、3 次元点の奥行きを変化させながらウィンドウマッチングを繰り返し、最もマッチングスコアが高くなる奥行きを選択する。このとき、高精度な 3 次元復元のためには、ステレオ画像上のピクセルとピクセルの間、いわゆるサブピクセルにおけるマッチングスコアを算出する必要がある。従来のウィンドウマッチング手法の中で、DAISY に基づくマッチングは、原理的にピクセルレベルのマッチングしかできず、精度よりも処理時間を重視した手法である。一方、SAD, SSD, NCC に基づくウィンドウマッチングでは、サブピクセルレベルのマッチングに、画像補間に基づく手法か関数フィッティングに基づく手法が用いられる。画像補間に基づく手法では、マッチングスコア算出の入力データであるマッチングウィンドウについて、サブピクセルの画素値を補間により求めることで、サブピクセルレベルのマッチングを行う [3], [28]。しかしながら、多視点ステレオでよく用いられる線形補間やキュービック補間について、これらの補間によって生成されたウィンドウ間のマッチングスコアでは、真の 3 次元点の奥行きで最大になるとは限らず、誤差が生じる。一方、関数フィッティングに基づく手法では、ピクセルレベルのマッチングスコアの変動に対し、それぞれのマッチング手法に適したピークモデル関数をフィッティングすることでサブピクセルレベルのマッチングを行う [3], [30]。しかしながら、SAD, SSD, NCC のいずれも、そのピーク形状が入力画像に依存するため、任意の入力画像に対するピークモデル関数を求めることはできない。そのため、実際には、実験的に求められたピークモデル関数を利用している。このように、従来のウィンドウマッチング手法では、サブピクセルレベルのマッチングに対して、たとえノイズなどの外乱が入らない場合でも、マッチングウィンドウの画像信号によっては誤差が生じる可能性がある。

これに対して、本論文では、多視点ステレオのための位相限定相関法 (Phase-Only Correlation: POC) に基づくウィンドウマッチングを提案する [31]–[33]。位相限定相関法は、画像の位相成分のみに着目した画像マッチング手法である [34]–[37]。この位相限定相関法では、理想

的な平行移動モデルを仮定した際に，入力画像に依存しないピークモデルを求めることができる．本論文では，位相限定相関法を多視点画像間のウィンドウマッチングに適用し，そのピークモデルを用いた関数フィッティングにより，3次元点の奥行きを高精度に推定する．従来の位相限定相関法に基づくステレオビジョンでは，2眼ステレオのみを対象にしていた [36], [37]．そのため，単純に位相限定相関法を多視点画像間のウィンドウマッチングに適用するだけでは，その視差の違いから，複数のステレオペアから計算される位相限定相関関数を同一の座標系で扱うことができず，3視点以上の多視点画像間のウィンドウマッチングに対して位相限定相関法を適用することができない．これに対し，本論文では，正規化視差の概念に基づく多視点画像間の相関関数の統合を提案し，異なるステレオペアから計算される位相限定相関関数を同一の座標系で扱うことで，位相限定相関法を多視点画像間のウィンドウマッチングに適用する．複数のステレオペアの位相限定相関関数を用いることで，従来の2眼ステレオの場合に比べ，ノイズやオクルージョンの影響を抑え，より高精度な奥行き推定が可能である．また，本論文では，多視点ステレオのための位相限定相関法に基づくウィンドウマッチングに対して，マッチングウィンドウの幾何補正による高精度化を適用する [38], [39]．物体表面の法線を考慮した幾何補正を行うことで，基線長の長いステレオカメラのようにステレオ画像間の変形が大きい場合においても，高精度に3次元点の奥行きを推定できる．

さらに，本論文では，多視点ステレオのための位相限定相関法に基づくウィンドウマッチングをもとに，高精度多視点ステレオアルゴリズムを提案する．具体的には，(i) 位相限定相関法と階層的探索を用いたデプスマップ生成，(ii) 重み付きメディアンフィルタを用いたデプスマップの高精度化，(iii) グラフカットに基づく誤対応点除去，(iv) 閾値処理に基づくアーチファクト除去を提案する．3次元点の座標計算に，位相限定相関法に基づくウィンドウマッチングを用いることで，多視点ステレオアルゴリズム全体においても，3次元復元精度の向上が可能である．また，提案手法の応用として，デジタルアーカイブを目的とした文化財の3次元復元を行う．実際の文化財を多視点画像から3次元復元することで，提案手法の有効性を示す．

精度評価実験では，多視点ステレオにおけるウィンドウマッチングのみの精度比較と，多視点ステレオアルゴリズム全体の精度比較を行う．まず，従来の多視点ステレオで用いられる

ウィンドウマッチング手法を分類し、ウィンドウマッチングによる奥行き推定精度を比較することで、ウィンドウマッチング手法が 3 次元復元結果に与える影響を明らかにする。次に、多視点ステレオアルゴリズム全体の評価を通して、提案手法により、従来の多視点ステレオアルゴリズムと比べ高精度な 3 次元復元が行えることを示す。

本論文は、以下の内容を取りまとめたものであり、以下に示す 5 章より構成される。

第 1 章は、本論文の背景、目的、および、その概要を述べた緒言である。

第 2 章では、本研究の中心である多視点 3 次元復元について概説する。はじめに、多視点画像の取得、カメラパラメータ推定、3 次元復元といった一般的な多視点 3 次元復元の流れについて説明する。次に、3 次元復元処理の中でも本研究で着目する多視点ステレオについて述べる。そして、多視点ステレオアルゴリズムにおけるウィンドウマッチングの役割について述べ、従来の多視点ステレオアルゴリズムで用いられるウィンドウマッチング手法の特徴をまとめる。

第 3 章では、多視点ステレオのための高精度ウィンドウマッチング手法について述べる。まず、提案手法で重要となる基礎技術として、位相限定相関法の基本原理を概説する。そして、位相限定相関法のステレオビジョンへの適用、および、正規化視差に基づく位相限定相関関数の統合について述べ、多視点ステレオのための位相限定相関法に基づくウィンドウマッチング手法を提案する。さらに、画像変形の大きいステレオ画像に対処するために、多視点ステレオのための位相限定相関法に基づくウィンドウマッチングに対して、マッチングウィンドウの幾何補正による高精度化を適用する。性能評価実験では、多視点ステレオアルゴリズムにおけるウィンドウマッチング手法のみについて精度および計算コストを比較し、提案手法を用いることで、従来のウィンドウマッチング手法と比べ 3 次元点の奥行きを高精度に推定できることを示す。

第 4 章では、高精度ウィンドウマッチングに基づく多視点ステレオアルゴリズムについて述べる。第 3 章で提案した多視点ステレオのための位相限定相関法に基づくウィンドウマッチング手法に対して、階層的探索を用いたデプスマップ生成、重み付きメディアンフィルタを用いたデプスマップの高精度化、グラフカットに基づく誤対応点除去、閾値処理に基づくアーチファクト除去を適用することで、位相限定相関法に基づく多視点ステレオアルゴリズムを提案

する．そして，公開データセットを用いた実験により，従来の多視点ステレオアルゴリズムと比較し，提案手法を用いて高精度な 3 次元復元が行えることを示す．また，提案手法の応用として，デジタルアーカイブを目的とした文化財の 3 次元復元について述べ，本文で提案する 3 次元復元の有効性を論じる．

第 5 章は，結言である．

以上，本論文の企図するところを概説した．

第 2 章

多視点 3 次元復元に関する基礎的 考察

2.1 まえがき

多視点 3 次元復元は、異なる視点から撮影された複数枚のカメラ画像を用いて物体の表面形状を復元する技術であり、コンピュータビジョンの分野において最も注目されている研究のひとつである [3], [5], [6], [8], [9]。多視点 3 次元復元は、レーザやパターン光を投影する能動型の 3 次元復元に比べ、撮影装置が小型・軽量であり、撮影環境が限定されにくい。また、汎用のカメラを利用できるため、専用の撮影装置が必要なく、導入が簡単であるといった利点もある。近年、デジタル画像技術の発展や、デジタルカメラの高性能化、および、さまざまな 3 次元復元アルゴリズムの進歩に伴い、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。これにより、多視点 3 次元復元は、コンピュータビジョンの研究分野だけでなく、産業、医療、芸術などの、さまざまな分野において、能動型の 3 次元復元に代わる技術として注目されるようになった。

本章では、多視点 3 次元復元に関する基礎的考察を行う。まず、多視点 3 次元復元の全体の処理について、(i) 多視点画像の取得、(ii) カメラパラメータ推定、(iii) 3 次元復元のそれぞれの観点から説明する。次に、本論文で着目する多視点ステレオ (Multi-View Stereo: MVS) について、多視点ステレオの分類とそれぞれの特徴、および、従来の多視点ステレオにおける

ウィンドウマッチングについて述べる。

2.2 多視点 3 次元復元

多視点 3 次元復元の流れは、(i) 多視点画像の取得、(ii) カメラパラメータ推定、(iii) 3 次元復元の 3 つのステップに分けられる [9]。多視点画像の取得は、多眼カメラによる撮影と、単眼カメラによる撮影がある。カメラパラメータの推定は、既知パターンによる事前校正と、撮影された多視点画像から直接推定する手法がある。3 次元復元は、カメラパラメータ推定と同時に疎な 3 次元復元を行う Structure from Motion (SfM)、多視点画像のシルエットに基づく手法である視体積交差法、多視点画像のテクスチャに基づく手法である MVS がある。本節では、それぞれのステップについて、その原理と特徴を概説する。

2.2.1 多視点画像の取得

多視点画像の取得方法は、多眼カメラによる撮影と、単眼カメラによる撮影の大きく 2 つに分けられる。単眼カメラによる撮影は、さらに、単眼カメラとロボットアーム・回転テーブルによる移動撮影、単眼カメラによる自由な移動撮影、インターネットからの画像取得の 3 つに分けられる。また、どの画像取得手法を用いるかによって、利用可能なカメラパラメータ推定手法が異なる。撮影される多視点画像に求められる要件を以下にまとめる。

- 対象物体の復元領域が少なくとも 2 枚以上の画像に写っている。
- 多視点画像間で対象物体の形状が変化しない。
- 多視点画像間で光源の状態が大きく変化しない。
- 復元領域全体にピントが合っている。
- 画像のノイズやぼけ、モーションブラーが少ない。
- 復元領域にハレーションや黒つぶれがない。

以下では、4 つの画像取得手法のそれぞれについて概説する。

(I) 多眼カメラによる撮影

多眼カメラによる撮影では、3 次元復元対象を中心に複数のカメラを固定し、多視点画像を撮影する。カメラを固定することで、カメラパラメータ推定と 3 次元復元のそれぞれに使用する画像を異なるタイミングで撮影した場合でも、カメラパラメータが変化せず、既知パターンによる事前校正が利用可能である。また、カメラを同期して撮影することで、移動物体の 3 次元復元や、動画像の撮影による対象形状の時間変化の観察が可能である。さらに、多視点画像がすべて同時に撮影されるため、多視点画像間で光源の状態を統一できるという利点もある。一方で、撮影装置が大型になるため、撮影スタジオの大きさなどの理由から、3 次元復元対象は比較的小さな物体に限られ、建築物などの大きな物体を撮影することができない。また、多くのカメラを必要とするため、撮影装置にかかる費用が多くなるといった問題がある。

(II) 単眼カメラとロボットアーム・回転テーブルによる移動撮影

単眼カメラとロボットアーム・回転テーブルによる移動撮影では、カメラがロボットアームの先端に固定され、ロボットアームおよび回転テーブルを動かすことにより、多視点画像を撮影する [5], [8], [9], [40], [41]。また、ロボットアームの利用が困難な場合、カメラは三脚などに固定される。対象物体とカメラ・ロボットアーム・回転テーブルの位置関係の例を図 2.1 に示す。この方法は、ロボットアームや回転テーブルの動きを記録し、常に同じカメラ位置から多視点画像を撮影することで、多眼カメラを用いる場合と同様に、既知パターンによる事前校正を利用できる。また、ロボットアームや回転テーブルの大きさから、撮影対象は比較的小さな物体に限られる。一方で、この手法では、多視点画像を同時に取得することができないため、移動物体の撮影は不可能である。また、回転テーブルを使用する場合、回転テーブルが動くたびに光源と物体の位置関係が変化するため、多視点画像間の光源の状態を統一することが難しい。事前に校正を行う場合、校正を行ったときと同じ条件で 3 次元復元対象の撮影を行うために、ロボットアームや回転テーブルに高い繰り返し精度が要求される。MVS 評価用の代表的なデータセットである Middlebury のデータセット [5], [42]、および、Jensen らのデータセット [41], [43] は、この手法により撮影されたものである。

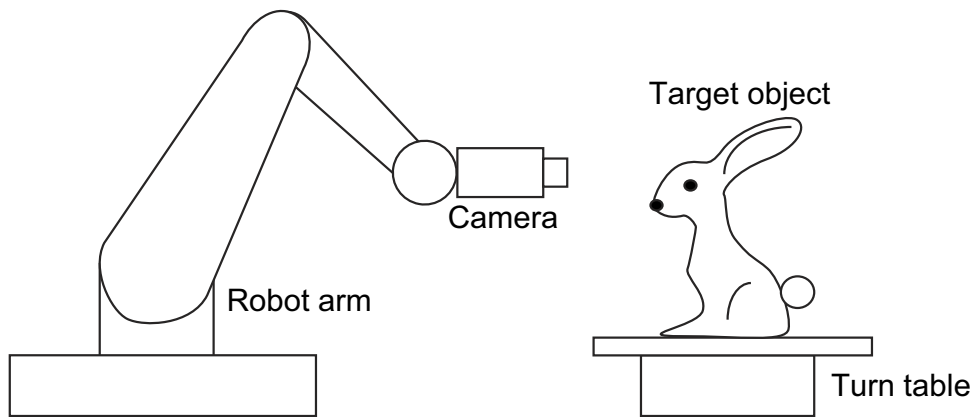


図 2.1 単眼カメラとロボットアーム・回転テーブルによる移動撮影

(III) 単眼カメラによる自由な移動撮影

多眼カメラやロボットアームなどを用いる撮影方法に対して、より撮影の制限が少ない撮影方法として、単眼カメラによる自由な移動撮影がある [6], [8], [9], [44]–[48]。カメラの移動撮影では、撮影者が自由に移動しながら同一の対象を撮影することで、多視点画像を取得する。この方法は、多眼カメラやロボットアームなどを用いる手法に比べて、撮影の制限が少なく、テーブルの上に乗るくらいの小さな物体から、建築物のような大きな物体まで撮影可能である。光源と物体の位置関係を変えずにカメラだけを移動させることで、多視点画像間の光源の状態を統一することが可能である。一方で、カメラの移動撮影では、事前に校正を行うことが困難であり、撮影された多視点画像から直接カメラパラメータを推定する必要がある。MVS 評価用の代表的なデータセットである Strecha らのデータセット [6], [49] はこの手法により撮影されたものである。

現在、多視点 3 次元復元において、その撮影の自由度から、単眼カメラによる自由な移動撮影が特に注目されている。この背景として、Scale-Invariant Feature Transform (SIFT) [50] や Speeded-Up Robust Features (SURF) [51] などの特徴ベースの対応付け手法により安定して多視点画像間の対応付けが可能になったこと、および、バンドル調整 [3], [52]–[55] により多視点画像間の対応付け結果のみから高精度なカメラパラメータ推定が可能になったことが挙げられる。

(IV) インターネットからの画像取得

インターネットからの画像取得では，オンラインの写真共有サイトなどを利用し，キーワード検索により復元対象の多視点画像を取得する [56]–[62]．オンライン写真共有サイトの代表的なものとして，Flickr [63] が挙げられる．画像撮影の作業が不要であり，キーワードを設定するだけで多視点画像を取得できるため，画像取得に必要な作業量が非常に少ないこと，および，インターネットを経由することで遠隔地の復元対象についても画像取得が可能であることが利点として挙げられる．一方で，撮影環境を制御することができないため，多視点画像間の光源の状態を統一することができない，撮影に用いるカメラを統一することができない，事前に校正を行うことができないなどの問題がある．特に，インターネット上の画像は，異なる人が異なる時間に異なる天候のもとで異なるカメラ・レンズで撮影したものであり，多視点 3 次元復元において最も難しい入力画像となる [57], [59], [62]．さらに，キーワード検索だけでは，復元対象以外の画像も候補として選択されたため，キーワード検索により収集した画像の中から，多視点 3 次元復元に利用する画像を適切に選択する必要がある．また，キーワード検索により復元対象を指定するため，復元対象は一意に決定できる物体に限定される．そのため，有名な文化財・芸術品・建築物などの 3 次元復元に適した手法といえる．

2.2.2 カメラパラメータ推定

カメラパラメータ推定では，各視点の画像におけるカメラの内部パラメータおよび外部パラメータを求める [3], [4], [9], [55]．内部パラメータは，レンズの焦点距離や画像中心，撮像素子のサイズと画像サイズの関係などを含み，カメラ座標系の 3 次元座標とデジタル画像座標系の 2 次元座標の関係を表す．外部パラメータは，その画像をどの位置からどの方向で撮影したかという情報を含み，世界座標系の 3 次元座標におけるカメラの位置関係を表す．本小節では，多視点 3 次元復元において一般的に利用されるピンホールカメラモデルを想定し，その内部・外部パラメータを推定する作業について述べる．多視点 3 次元復元に利用されるカメラパラメータ推定は，既知パターンによる事前校正と，撮影された多視点画像から直接推定する手法の 2 つに大きく分けられる．以下では，それぞれのカメラパラメータ推定手法について概説する．

(I) 既知パターンによる事前校正

既知パターンによる事前校正では、3 次元形状が既知の物体を複数回撮影し、既知パターンの 3 次元座標と画像上の 2 次元座標との対応関係からカメラパラメータを推定する (図 2.2(a)) [3], [4], [55], [64]。このとき、既知パターンとして平板や直方体を用い、表面のテクスチャとして格子パターンやチェッカーパターン、整列された円などを用いることで、コーナー検出や円検出などにより平板や直方体上の 3 次元座標と画像上の 2 次元座標の対応関係を得ることができる。中でも、図 2.2(b) に示されるような、チェッカーパターンの印刷された平板を用いる Zhang らの手法が広く利用されている [64], [65]。このように得られた 3 次元座標と 2 次元座標の関係から、想定するカメラモデルにおいて最も整合するパラメータを推定する。つまり、カメラパラメータが正確であれば、すべての特徴点について、既知の 3 次元座標を画像上に投影した点と、対応する 2 次元座標との誤差が小さくなるはずである (図 2.2(a))。平板の既知パターンについて、1 枚の画像のみから、その特徴点の 3 次元座標と 2 次元座標の整合性を取るカメラパラメータは一意に決まらない。そこで、既知パターンの傾きや位置を変えた複数枚の画像について、整合性を取ることで、カメラパラメータを決定する。内部パラメータのみの推定、および、内部・外部パラメータ両方の推定には最低 3 枚の画像が、既知の内部パラメータからの外部パラメータのみの推定には最低 1 枚の画像が必要である。

既知パターンによる校正は、非常に安定性が高く、また、外部パラメータにスケールの不定性が生じないことが利点として挙げられる。一方で、校正に用いる既知パターンを高精度に作成する必要がある。カメラの解像度が低い場合、既知パターンの作成に求められる精度は現実的であるが、カメラの解像度が高い場合、既知パターンの作成に要求される精度が非常に高く、作成が困難である。例えば、 $5,000 \times 4,000$ ピクセルのカメラで 100 mm 四方の既知パターンを視角 45 度で撮影する場合、1 ピクセル以下の精度でカメラパラメータを推定するためには、既知パターンを 0.025 mm 以下の精度で作成しなくてはならない [9]。

多視点画像の取得に、多眼カメラを用いる場合には、すべてのカメラの内部・外部パラメータを同時に推定する。一方、ロボットアーム・回転テーブルを用いる場合には、すべてのカメラの内部・外部パラメータを同時に推定する方法だけでなく、先にカメラの内部パラメータだけを求め、次に外部パラメータを推定する方法も用いられる。このとき、ロボットアーム・回

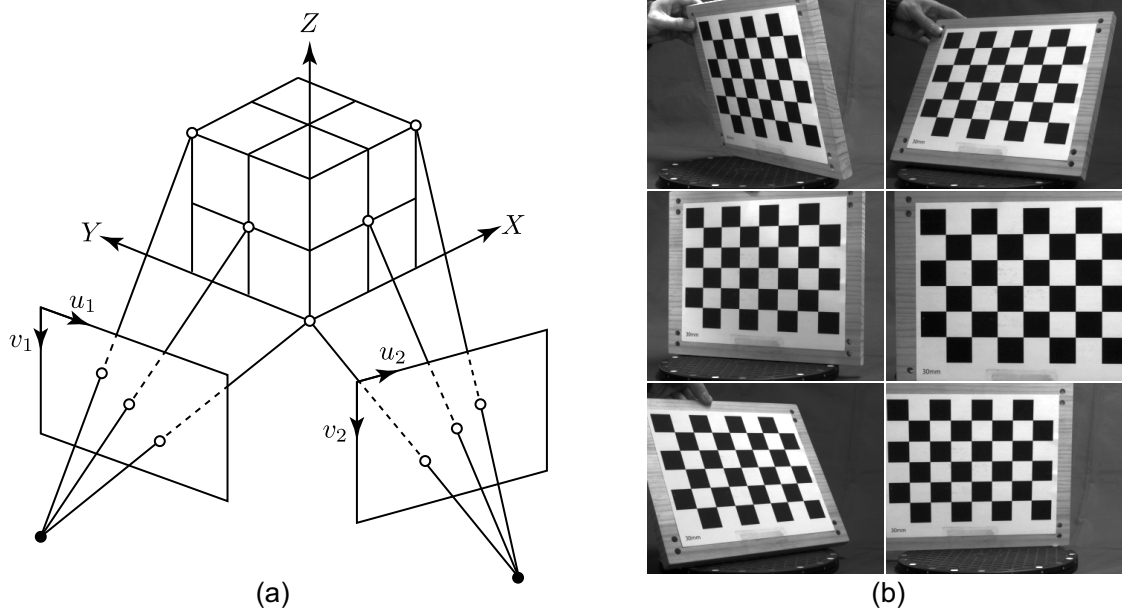


図 2.2 既知パターンによる校正: (a) 3 次元座標と 2 次元座標の整合性, (b) Zhang からの手法によるチェッカーパターンの例

転テーブルの制御情報から外部パラメータを直接求めたり, 最適化処理の初期値に用いることも可能である. 多視点画像の取得に, 単眼カメラによる自由な移動撮影や, インターネットからの画像取得を用いる場合, 3 次元復元に利用する多視点画像と全く同じ状態で既知パターンの撮影を行うことができないため, 既知パターンによる校正は利用できない.

(II) 多視点画像からのカメラパラメータ推定

多視点画像からのカメラパラメータ推定では, 3 次元復元に用いる多視点画像から直接カメラパラメータを推定する [3], [7], [9], [55], [56], [58]. この処理は, 後述する Structure from Motion (SfM) と同一の処理として扱われる場合もあるが, 本論文では, SfM におけるカメラパラメータ推定と 3 次元復元の処理を分けて説明する. 多視点画像からのカメラパラメータ推定では, まず, 多視点画像間の対応関係を得る. 多視点画像間の対応付けには, ステレオ画像間の画像変形にロバストな対応付け手法が必要なため, SIFT や SURF などの特徴ベースの対応付け手法 [50], [51] が用いられる. このように得られた多視点画像間の対応関係から,

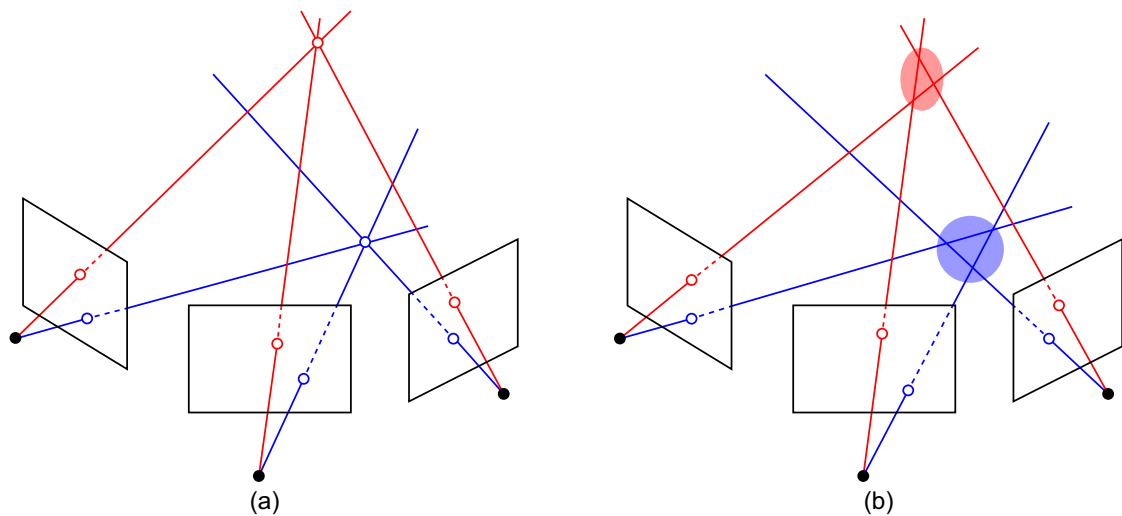


図 2.3 多視点画像からのカメラパラメータ推定: (a) カメラパラメータが正確であれば, すべての対応点ペアの視線が 1 点で交わる, (b) 不正確なカメラパラメータでは, 対応点ペアの視線が 1 点で交わらない

想定するカメラモデルにおいて最も整合するカメラパラメータを推定する。つまり, カメラパラメータが正確であれば, すべての対応点ペアについて, その視線が 3 次元座標上の 1 点で交わるはずである (図 2.3)。この問題の代表的な解法として, 8 点アルゴリズム [55], 5 点アルゴリズム [66], P3P 問題の解法 [67], バンドル調整 [52]–[54] などが挙げられる。2 視点の対応関係からエピポーラ方程式の F 行列を求める場合, 8 点アルゴリズムが用いられる。内部パラメータが既知の 2 視点の対応関係から E 行列を求める場合, 5 点アルゴリズムが用いられる。3 視点以上の対応関係について, 既に他の 2 視点の対応から 3 次元座標が計算されている場合, P3P 問題の解法が適用される。また, いずれかの手法により推定したカメラパラメータを初期値として, 3 次元復元した点を再度画像上に投影した際の誤差 (再投影誤差) を最小化するように, カメラパラメータを最適化する処理をバンドル調整と呼ぶ。

多視点画像からのカメラパラメータ推定は, 多視点画像の取得方法に依存せず, あらゆる多視点画像に適用可能である。一方で, 多視点画像間の対応付けに誤対応が含まれると, カメラパラメータの推定精度が著しく低下する。そのため, RANdom SAmple Consensus

(RANSAC) [68] などを用いて誤対応を正確に除去することが重要となる。また、多視点画像から直接カメラパラメータを推定する場合、原理的に外部パラメータにスケールの不定性があり、3次元復元結果の大きさを一意に決定することができない、復元対象が完全な平面だと正確なカメラパラメータが推定できないといった問題がある。

単眼カメラによる自由な移動撮影の場合、事前に校正することができないため、多視点画像から直接カメラパラメータを推定する必要がある。内部・外部パラメータの両方を推定することもできるが、撮影中に焦点距離などを固定することで、内部パラメータのみ事前に校正し、外部パラメータのみを多視点画像から推定することもできる。一方、インターネットからの画像取得では、内部パラメータについても、事前に校正することができない。そのため、常に、内部・外部パラメータの両方を多視点画像から推定する必要がある。このとき、安定性向上のため、画像のヘッダ情報などから近似的に計算した内部パラメータを、バンドル調整の初期値として利用できる。

2.2.3 3次元復元

3次元復元のステップでは、前述の2つのステップで取得した多視点画像とカメラパラメータを入力として、対象物体の3次元形状(3次元点群, 3次元メッシュモデル, 3次元ボリュームなど)を復元する。3次元復元には、SfM, 視体積交差法, MVSの3つのアルゴリズムがある[3]。以下では、それぞれの3次元復元アルゴリズムについて概説する。

(I) Structure from Motion (SfM)

SfMは、2.2.2節で述べた多視点画像からのカメラパラメータ推定の際に利用した特徴点のみを3次元復元する[7],[9]。通常、多視点画像からのカメラパラメータ推定では、カメラパラメータと同時に特徴点の3次元座標を推定するため、カメラパラメータ推定の処理と3次元復元の処理を分けることはできない。SfMを3次元復元の処理に含めるかどうかは文献により異なる。また、特徴ベースの対応付け手法が画像上の局所的なテクスチャに基づくことから、広義では、SfMはMVSの一種であるといえる。SfMは、カメラパラメータ推定に利用した特徴点のみを復元するため、対象物体のおおまかな形状しか復元できない。また、復元し

た特徴点をもとに，3 次元メッシュモデルを生成する場合についても，MVS と比べて粗い 3 次元メッシュモデルしか復元できない．そのため，拡張現実 (Augmented Reality: AR) や，人物のジェスチャ認識など，物体のおおまかな形状で十分な用途には利用できるが，工業製品の 3 次元形状による検査や，文化財のデジタルアーカイブなどの物体の詳細な形状が求められる用途には適さない．

(II) 視体積交差法

視体積交差法は，多視点画像間のシルエットに基づく 3 次元復元アルゴリズムで，Visual-hull, Shape from Silhouette と呼ばれる [3], [69], [70]．すべての視点の画像について，画像上のピクセルを対象物体が投影されたピクセル（前景）と対象物体が投影されていないピクセル（背景）に分離し，復元される 3 次元モデル（ボクセルやメッシュモデル）がすべての画像上の前景のみに投影されるように，3 次元モデルを推定する．ボクセルに基づくアルゴリズムでは，3 次元空間を整列した立方体（ボクセル）で分割し，各ボクセルをすべての画像上に投影し，1 つ以上の画像で背景となるボクセルを除去することで，3 次元復元を行う．一方，3 次元メッシュモデルに基づくアルゴリズムでは，現在のメッシュモデルに対して 1 枚ずつ画像を追加していき，追加された画像上の前景と背景の境界線をメッシュモデルに逆投影し，現在のメッシュモデルと境界線が交差する領域が新しいメッシュモデルの物体表面になるように，メッシュモデルを更新していく．いずれの手法においても，視体積交差法では，前景と背景の正確な分離が 3 次元復元精度を決定するため，ブルースクリーンなどを用いて単純な背景で画像を撮影する必要がある．そのため，ある程度背景を制御できる屋内の撮影には適しているが，背景の制御が難しい屋外の撮影やインターネットからの画像収集は適さない．また，高精度な 3 次元復元のために必要な画像枚数が膨大になる，原理的に凹面が復元できないといった欠点がある．

(III) 多視点ステレオ

MVS は，多視点画像間の物体表面テクスチャに基づく手法である [3], [8], [9]．詳しい原理については後述し，本節ではその特性について記述する．多視点画像において，対象の物体表面テクスチャは，物体の 3 次元形状と各視点のカメラパラメータによって，画像上の写り方が

変化する．そのため，カメラパラメータが既知の場合，正しい 3 次元形状を仮定すれば，多視点画像間に投影された物体表面テクスチャで整合性が取れる．従来の MVS では，ウィンドウマッチングにより多視点画像間のテクスチャの整合性を表すスコアを算出し，マッチングスコアが高くなるように 3 次元形状を推定する．MVS は，最低 2 枚の画像から利用でき，視体積交差法と比べて，少ない枚数の画像から高精度な 3 次元復元が可能である．さらに，復元対象の表面テクスチャが背景に依存しないため，複雑な背景においても 3 次元復元精度が低下しない．特に，屋外環境では，背景の制御が難しいため，視体積交差法よりも MVS が適している．また，凹面についても凸面と同様の原理で復元できるため，対象物体の形状が限定されない．このように，MVS は，SfM および視体積交差法と比べ，物体の詳細な形状を復元できることや，撮影環境・対象物体の形状に制限が少ないことが利点として挙げられる．以上のような特長から，MVS は，多視点 3 次元復元の中でも特に実用性が高く，高精度な 3 次元復元手法である．一方で，屋外における昼と夜の画像のように，光源の状態が大きく変化すると，画像上の物体表面テクスチャが大きく変化するため，対処できない．また，単色の物体のようにテクスチャがない物体については，正確な復元が難しい．そのため，インターネットからの画像取得のように，撮影環境が全く制御できない場合には，取得した画像の中から MVS に適した画像を選択して利用する必要がある [57],[62] ．

2.3 多視点ステレオ

MVS は，多視点画像間のテクスチャに基づく 3 次元復元であり，多視点 3 次元復元に用いられる 3 次元復元アルゴリズムの中でも，精度・ロバスト性が高く，多くの物体に適用可能である．現在までに，数多くの MVS アルゴリズムが提案されているが [3],[5],[9]，いずれのアルゴリズムにおいても，3 次元点群生成，3 次元メッシュ生成などの，3 次元座標を決定する処理に多視点画像間のウィンドウマッチングが適用され，その性能が処理全体の性能に大きく影響する．本節では，まず，MVS アルゴリズムの分類について述べ，次に，MVS アルゴリズムにおける重要な構成要素であるウィンドウマッチングについて述べる．

2.3.1 多視点ステレオアルゴリズムの分類

Seitz らは, MVS アルゴリズムを, (i) 3 次元ボリュームの最適化に基づくアルゴリズム, (ii) 3 次元メッシュの最適化に基づくアルゴリズム, (iii) 特徴領域拡張に基づくアルゴリズム, (iv) デプスマップ統合に基づくアルゴリズムの 4 つに分類した [5]. 本小節では, それぞれのアルゴリズムについてまとめる.

(I) 3 次元ボリュームの最適化に基づくアルゴリズム

3 次元ボリュームの最適化に基づくアルゴリズムは, 多視点画像と 3 次元ボリュームからエネルギー関数を定義し, そのエネルギー関数が最小になるような 3 次元ボリュームを推定することで, 3 次元復元を行う [10]–[12]. 一般に, 自由な 3 次元ボリュームの表現を定義することは難しく, また, 複雑な 3 次元ボリュームの表現はエネルギー関数を複雑にする. そこで, 多くの場合, 3 次元ボリュームをボクセルによって表現し, あるボクセルが 3 次元復元対象内部かどうかを最適化により求める. このようなボクセルの内外判定は, グラフカットによる 2 値のラベリング問題として解かれる. このとき, グラフのエッジに相当するエネルギー関数の計算には, 多視点画像間のウィンドウマッチングが適用され, ウィンドウマッチングによりあるボクセルが物体表面である場合の多視点画像間のテクスチャの整合性を数値化する.

3 次元ボリュームの最適化に基づくアルゴリズムは, 3 次元復元精度と処理時間がボクセルの分解能に大きく依存する. 高精度な 3 次元復元を行うためには, ボクセルを細かく設定する必要があり, 処理時間が増加する. また, 物体全体を内包するボクセル空間を定義するために, 3 次元復元対象の大まかな形状を入力する必要がある. これらの特徴から, 3 次元ボリュームの最適化に基づくアルゴリズムは, 視体積交差法により得られた 3 次元ボリュームや, 他の MVS アルゴリズムにより得られた 3 次元復元結果を初期値として用いることが多い. また, 大規模な 3 次元復元を行う場合, 想定するボクセル空間の大きさが膨大になるため, 適用が難しい.

(II) 3 次元メッシュの最適化に基づくアルゴリズム

3 次元メッシュの最適化に基づくアルゴリズムは, 多視点画像と 3 次元メッシュモデルから

エネルギー関数を定義し、そのエネルギー関数が最小になるような 3 次元メッシュモデルを推定することで、3 次元復元を行う [13]–[16]。一般に、自由な 3 次元メッシュモデルの表現を定義することは難しく、また、複雑な 3 次元メッシュモデルの表現は最小化するエネルギー関数を複雑にする。そこで、多くの場合、3 次元メッシュモデルを多角形メッシュによって表現し、多角形の頂点を移動させたり、多角形を分割したりすることで最適化を行う。このような最適化には、最急降下法や Levenberg-Marquardt 法 (LM 法) などの非線形最適化手法 [71] が用いられる。このとき、最小化するエネルギー関数の計算には、多視点画像間のウィンドウマッチングが適用され、ウィンドウマッチングにより、ある仮定した 3 次元メッシュのパラメータに対する多視点画像間のテクスチャの整合性を数値化する。

3 次元メッシュの最適化に基づくアルゴリズムは、MVS アルゴリズムの中でも特に高精度な 3 次元復元を実現している [15], [16]。一方で、初期値によって問題が収束しなかったり、最適化の際の処理時間が膨大になるといった問題がある。また、局所的な最適解に収束するのを防ぐために、初期値のメッシュモデルを高精度に求める必要がある。これらの特徴から、3 次元メッシュの最適化に基づくアルゴリズムは、視体積交差法により得られた 3 次元ボリュームや、他の MVS アルゴリズムにより得られた 3 次元復元結果を初期値として用いることが多く、3 次元復元結果の高精度化に適している。

(III) 特徴領域拡張に基づくアルゴリズム

特徴領域拡張に基づくアルゴリズムは、(i) コーナー点などの特徴点について 3 次元復元を行い、(ii) 復元結果を周囲の点に繰り返し伝搬することで、物体全体の 3 次元復元を行う [17]–[19]。特徴点の 3 次元復元には、SIFT または SURF のような特徴ベースの対応付け手法や、正規化相互相関 (Normalized Cross-Correlation: NCC) などの領域ベースのウィンドウマッチングを用いたパッチベースの対応付け手法が用いられる。中でも、Furukawa らの MVS アルゴリズム [19] は、さまざまな環境下で適用でき、現在、最も広く利用される MVS アルゴリズムのひとつである。Furukawa らのアルゴリズムでは、Difference-of-Gaussian (DoG) と Harris のコーナー検出により特徴点を検出し、特徴点の 3 次元復元を行う。その後、誤対応点除去と 3 次元復元結果の伝搬を繰り返し行うことで、物体全体の 3 次元メッシュモデルを

復元する。このとき、特徴点の 3 次元復元や伝搬の処理において、ある仮定した 3 次元点のパラメータ (3 次元座標や法線方向) の尤度を、ウィンドウマッチングにより算出する。

一般に、特徴領域拡張に基づくアルゴリズムは、3 次元ボリュームの最適化やメッシュの最適化に基づくアルゴリズムに比べて、処理時間が短く、また、視体積交差法の 3 次元復元結果などの初期値が必要ないといった利点がある。さらに、局所領域ごとに 3 次元復元の処理が独立するため、都市全体の復元などの大規模な復元にも適用しやすい。一方で、3 次元復元を特徴点から伝搬させるため、復元対象によっては、特徴点の数が少なく物体全体が復元できなかったり、特徴点の少ない領域で局所的に 3 次元復元精度が低下するといった問題がある。特に、物体表面のテクスチャが少なく、局所的に単色に近い領域の復元は難しい。また、特徴点の 3 次元復元結果には誤対応点が含まれるため、安定した 3 次元復元を行うためには、誤対応点除去を正確に行うことが重要である。

(IV) デプスマップ統合に基づくアルゴリズム

デプスマップ統合に基づく手法は、(i) 多視点画像から複数のデプスマップを生成し、(ii) 複数のデプスマップを統合することで、3 次元復元を行う [20]–[27]。デプスマップとは、画像上のすべてのピクセルについて、その視点における 3 次元点の奥行きを保存したマップである。つまり、デプスマップの生成は、各視点について、すべての画像座標における 3 次元点の奥行きを推定する問題となる。このとき、各 3 次元点の奥行き推定には、ウィンドウマッチングが適用され、ウィンドウマッチングにより、ある仮定された 3 次元点のパラメータに対する尤度を算出する。デプスマップの統合は、視点ごとに生成されたデプスマップについて、それぞれのデプスマップの整合性を取るように、3 次元点群や 3 次元メッシュモデルを生成する。

デプスマップ統合に基づくアルゴリズムは、複雑な最適化を必要としないため、3 次元ボリュームの最適化やメッシュの最適化に基づくアルゴリズムと比べて、局所的な最適解に収束するといった問題がなく、視体積交差法の 3 次元復元結果などの初期値が必要ないといった利点がある。また、デプスマップ生成が、参照視点ごとに独立するため、多視点画像の枚数が多くなっても、1 つの参照視点ごとの計算コストは変化しない。そのため、大規模な 3 次元復元にも適用しやすい。さらに、特徴領域拡張に基づくアルゴリズムと比べて、対象物体のテク

スチャによって 3 次元復元結果が特徴点周辺に限定されるといった問題もない。このように、デプスマップ統合に基づくアルゴリズムは、復元対象が限定されにくく、さまざまな環境下で適用できることから、現在の MVS アルゴリズムの主流となっている [9]。一方で、特徴領域拡張に基づくアルゴリズムと同様に、各視点におけるデプスマップには誤対応点が含まれるため、安定した 3 次元復元を行うためには、誤対応点除去を正確に行うことが重要である。

2.3.2 多視点ステレオにおけるウィンドウマッチング

前述した MVS アルゴリズムの 4 つの分類において、3 次元点群生成、3 次元メッシュ生成、ボクセル最適化、メッシュ最適化などの、3 次元座標を決定する処理には多視点画像間のウィンドウマッチングが適用される。そのため、MVS におけるウィンドウマッチングは、アルゴリズム全体の性能を決定する重要な要素である。従来の MVS アルゴリズムでは、多視点画像間のウィンドウマッチング手法として、Sum of Absolute Differences (SAD)、Sum of Squared Differences (SSD)、NCC [3],[28]、DAISY に基づくマッチング [29] が用いられる。これらのウィンドウマッチング手法により算出されたマッチングスコア (MVS では photo-consistency と呼ばれる) は、3 次元点や 3 次元メッシュモデルのパラメータ (3 次元座標や法線方向) の尤度として利用される。例えば、特徴領域拡張やデプスマップ統合に基づく手法でよく用いられる 3 次元点の奥行き推定の様子を図 2.4 に示す。図 2.4 では、参照視点 V_R 上のデジタル画像座標 m について、その視線上の 3 次元点 M の奥行き Z を変化させながら、参照視点 V_R と近傍視点 C_0, C_1, C_2 間でマッチングスコアを算出する。このとき、正しい奥行きが仮定されると、その投影点周辺の画像上のテクスチャが等しくなるため、マッチングスコアが高くなる。そのため、3 次元点 M の奥行き Z を推定する問題は、最もマッチングスコアが高くなる Z を探索する問題となる。このマッチングスコアが最高になる奥行きを探索する問題について、もっとも単純な解法として、総当たりによる探索が挙げられ、プレーンスイーピングと呼ばれる [3],[5],[20]。プレーンスイーピングでは、奥行き Z を刻み幅 ΔZ で変化させながらウィンドウマッチングを繰り返し、最もマッチングスコアが高くなる Z を真の奥行きとして選択する。このとき、高精度な 3 次元復元を行うためには、3 次元点の奥行き変化が、ステレオ画像上でサブピクセル (1 ピクセル以下) の変動になるため、

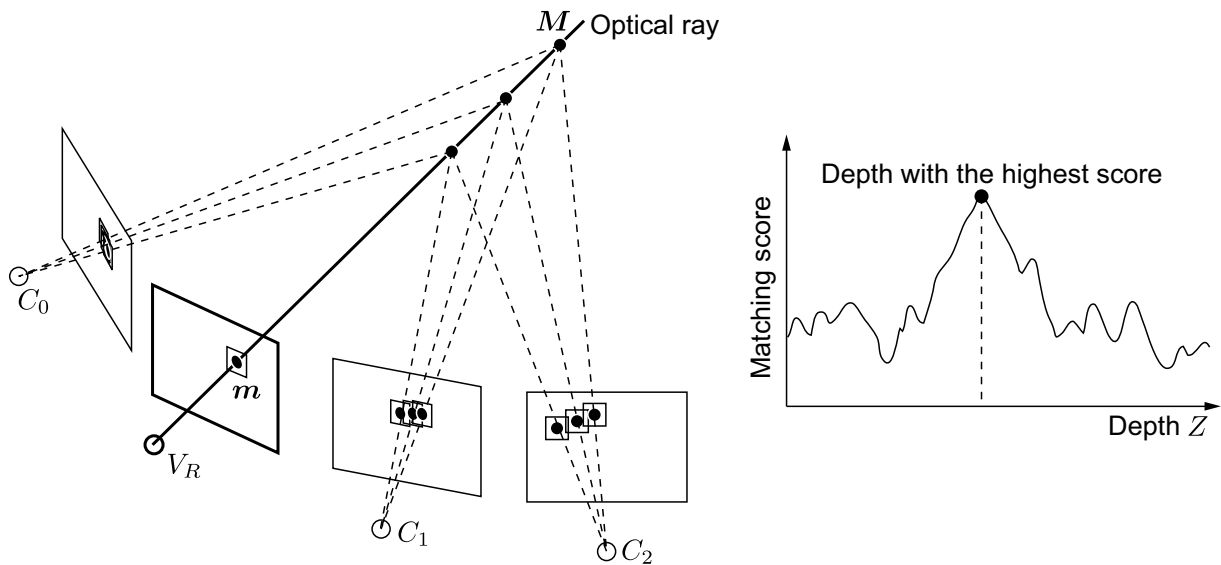


図 2.4 MVS におけるウィンドウマッチング (3 次元点の奥行き変化とマッチングスコアの関係)

サブピクセルレベルのマッチングが必要である。MVS におけるウィンドウマッチングにおいて、どのようにマッチングスコアを算出するか、および、サブピクセルレベルのマッチングをどのように実現するかで、3 次元復元精度は大きく変化する。

現在、MVS アルゴリズムの中でも精度が高いアルゴリズムの多くは、NCC に基づくウィンドウマッチングを適用している [15], [16], [19], [20], [23], [24], [27], [72]。Goesele らは、NCC に基づくウィンドウマッチングをプレーンスイーピングの枠組みで使用し、複数のステレオペアから計算されたマッチングスコアを統合することで高精度なデプスマップを生成した [20]。Campbell らは、Goesele らの手法に対して、近傍ピクセルの情報を利用することで、誤対応点の発生を抑制した [24]。また、Zabulis ら、Bradley ら、Furukawa らは、3 次元点の奥行きだけでなく、3 次元点の法線ベクトルも考慮することで、カメラに正対した平面以外にも対処し、精度・ロバスト性を向上させている [19], [23], [72]。

一方、SAD, SSD に基づくウィンドウマッチングは、精度よりも処理時間に重点を置いたアルゴリズムで利用されている [22], [73], [74]。Cornells らは、マッチングスコアに SSD を使い、アルゴリズムを GPU に実装することでリアルタイムの多視点 3 次元復元を実現した

[73]. さらに, Gallup らは, SAD に基づくウィンドウマッチングにおいて, カメラに正対したウィンドウ以外を考慮することで, 3 次元復元精度を向上させている [22]. また, Newcombe らは, そのリアルタイム性を生かし, 動画像を用いたオンラインのカメラ位置推定と密な 3 次元復元を実現した [74]. しかしながら, 近年では, 計算機の高性能化とともに, NCC に基づく手法においても高速に処理できるようになり, SAD, SSD の NCC に対する利点は小さくなりつつある [75], [76].

SAD, SSD, NCC に基づくウィンドウマッチングでは, サブピクセルレベルのマッチングに, 画像補間に基づく手法か, 関数フィッティングに基づく手法が用いられる. 画像補間に基づく手法では, マッチングスコア算出の入力データであるマッチングウィンドウについて, サブピクセルの画素値を補間により求めることで, サブピクセルレベルのマッチングを行う [3], [28]. 3 次元形状のパラメータ探索に, 総当たりによる探索や, 非線形最適化による探索を適用する場合, 画像補間に基づく手法が用いられる. しかしながら, 多視点ステレオでよく用いられる線形補間やキュービック補間について, これらの補間によって生成されたウィンドウ間のマッチングスコアでは, 真の 3 次元点の奥行きで最大になるとは限らず, 誤差が生じる. 一方, 関数フィッティングに基づく手法では, ピクセルレベルのマッチングスコアの変動に対し, それぞれのマッチング手法に適したピークモデル関数をフィッティングすることでサブピクセルレベルのマッチングを行う [3], [30]. しかしながら, SAD, SSD, NCC のいずれも, そのピーク形状が入力画像に依存するため, 任意の入力画像に対するピークモデル関数を求めることはできない. そのため, パラボラ関数や等角直線関数など, 実験的に求められたピークモデル関数を利用している.

一方, Li ら, および, Tola らは, DAISY に基づくマッチングを用いた MVS アルゴリズムを提案している [25], [26]. DAISY に基づくマッチングは, 基線長の長いステレオペアのような画像変形の大きいステレオ画像間でもロバストにマッチングが行えるが, 原理的にサブピクセルレベルのマッチングができない. Li らは, 精度が低いデプスマップについて, デプスマップ統合の際にバンドル調整を適用することで, 3 次元復元精度を向上させている [25]. また, Tola らは, デプスマップの統合の際に, 基線長の長いステレオ画像の対応付け結果を優先することで, 3 次元復元精度を向上させている [26]. ただし, SAD, SSD, NCC に基づく

ウィンドウマッチングにおいても、マッチングウィンドウ間の幾何補正を行うことで、基線長の長いステレオペアに対処することが可能であり、ウィンドウマッチングによる 3 次元点の奥行き推定の観点から、DAISY による奥行き推定精度はその他のマッチング手法と比べて低い。DAISY に基づくマッチングは、その計算コストの少なさから、大規模な多視点画像セットに適した手法である。

2.4 むすび

本章では、多視点 3 次元復元に関する基礎的考察を行った。まず、多視点 3 次元復元の全体の処理について、(i) 多視点画像の取得、(ii) カメラパラメータ推定、(iii) 3 次元復元のそれぞれの観点から説明した。次に、本論文で着目する MVS について、MVS アルゴリズムの分類とそれぞれの特徴、および、従来の MVS におけるウィンドウマッチングについて述べた。MVS におけるウィンドウマッチングは、主に、3 次元形状を計算する処理に適用され、MVS の 3 次元復元精度を決定する重要な要素である。特に、高精度な 3 次元復元では、ウィンドウマッチングの際に、サブピクセルのマッチングをどのように実現するかが重要となることを述べた。

第 3 章

多視点ステレオのための高精度ウィンドウマッチング

3.1 まえがき

多視点ステレオ (Multi-View Stereo: MVS) におけるウィンドウマッチングは、アルゴリズム全体の精度・ロバスト性を決定する重要な構成要素である。ウィンドウマッチングを精度が向上することで、MVS による 3 次元復元精度の向上が期待できる。従来の MVS におけるウィンドウマッチングでは、サブピクセルレベルのマッチングに対して、実験的に一定の成果を達成しているが、たとえノイズなどの外乱が入らない場合でも、マッチングウィンドウの画像信号によっては誤差が生じる可能性がある。これに対して、本章では、MVS のための位相限定相関法 (Phase-Only Correlation: POC) に基づくウィンドウマッチングを提案する [31]–[33]。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 [34]–[37]。この POC では、理想的な平行移動モデルを仮定した際に、入力画像に依存しないピークモデル関数を求めることができる。提案手法では、POC のピークモデル関数を 3 次元点の奥行き推定に利用することで、高精度な 3 次元復元を行う。しかしながら、単純に多視点画像間のマッチングに POC を適用するだけでは、その視差の違いから、複数のステレオペアから計算される位相限定相関関数 (POC 関数) を統合できない。そこで、本章では、正規化視差の概念に基づく多視点画像間の POC 関数の統合を提案し、多視点画像間のウィンドウマッチングに

POC を適用する。また、本章では、MVS のための POC に基づくウィンドウマッチングに対して、マッチングウィンドウの幾何補正による高精度化を適用する [38], [39]。物体表面の法線を考慮した幾何補正を行うことで、基線長の長いステレオペアのように画像間の変形が大きい場合においても、高精度に 3 次元点の奥行きを推定できる。

以下では、MVS のための POC に基づくウィンドウマッチングについて述べる。まず、提案手法で重要となる基礎技術である POC について述べる。次に、MVS のための POC に基づくウィンドウマッチングを提案し、POC を用いて 3 次元点の奥行きを推定する手法を述べる。さらに、提案するウィンドウマッチング手法の高精度化として、マッチングウィンドウの幾何補正について述べる。そして、さまざまなウィンドウマッチング手法と、奥行き推定精度、および、計算コストを比較することで、提案手法の性能評価を行う。

3.2 位相限定相関法

本節では、提案手法で重要となる基礎技術である POC について述べる。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 [34], [77]–[80]。2 つの画像信号から 1 つの相関値、もしくは、相違度を算出する Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), 正規化相互相関 (Normalized Cross-Correlation: NCC) のような画像マッチングと異なり、POC は 2 つの画像信号から 1 つの相関関数を計算する。この相関関数の相関ピーク座標を求めることで、2 つの画像信号間の平行移動量を推定することが可能である。以下では、まず POC を用いた平行移動量推定の基本原理について説明し、次に、POC を用いた平行移動量推定の高精度化について述べる。

3.2.1 位相限定相関法の基本原理

一般に、POC は、1 次元、2 次元、3 次元の信号のように、任意の次元の信号に対して定義できる [79], [80]。提案手法では、ステレオ画像間の局所的な平行移動量推定に POC を利用する。ステレオ画像では、ある点の奥行き変化に対するステレオ画像上の平行移動がエピポーラ線上の 1 次元の移動に限定されるため [3], [4]、本論文では 1 次元 POC を用いる。以下では、

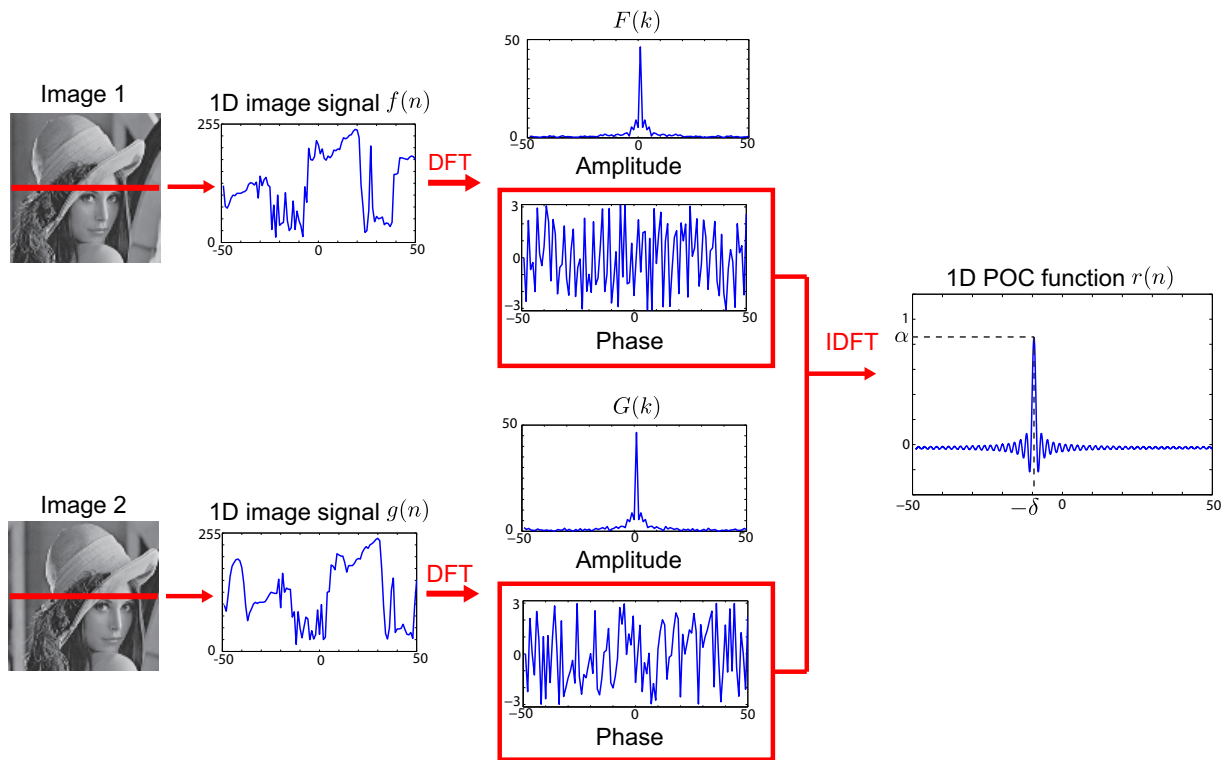


図 3.1 1次元位相限定相関法 (1次元 POC)

2つの1次元画像信号間の平行移動量を1次元POCを用いて推定する手法について述べる。

はじめに、2つの1次元画像信号を $f(n)$ および $g(n)$ とする。ここで、定式化の便宜上、1次元画像信号の離散空間インデックスを $n = -M, \dots, M$ とする。ここで、 M は正の整数であり、1次元画像信号の長さ N は $N = 2M + 1$ で与えられる。なお、本節では、離散空間インデックスを正負対称にとり、1次元画像信号の長さを奇数にしているが、これはPOCにおいて必須ではない。すなわち、通常よく用いられるように非負の離散空間インデックスを用い、1次元画像信号の長さを任意の正の整数になるように一般化することができる。

1次元画像信号 $f(n)$ および $g(n)$ から、1次元POC関数 $r(n)$ を計算する際の処理の流れを図3.1に示す。1次元画像信号 $f(n)$ および $g(n)$ の1次元離散フーリエ変換 (Discrete

Fourier Transform: DFT) をそれぞれ $F(k)$ および $G(k)$ として次式で与える .

$$F(k) = \sum_{n=-M}^M f(n)W_N^{kn} = A_F(k)e^{j\theta_F(k)} \quad (3.1)$$

$$G(k) = \sum_{n=-M}^M g(n)W_N^{kn} = A_G(k)e^{j\theta_G(k)} \quad (3.2)$$

ここで, $W_N = e^{-j\frac{2\pi}{N}}$ は回転因子を, $A_F(k)$ および $A_G(k)$ はそれぞれ 1 次元画像信号 $f(n)$ および $g(n)$ の振幅成分を, $\theta_F(k)$ および $\theta_G(k)$ はそれぞれ $f(n)$ および $g(n)$ の位相成分を表す . また, 離散空間インデックスと同様に, 離散周波数インデックスを $k = -M, \dots, M$ とする . このとき, $f(n)$ と $g(n)$ の正規化相互パワースペクトル $R(k)$ を次式で与える .

$$R(k) = \frac{F(k)\overline{G(k)}}{|F(k)\overline{G(k)}|} = e^{j(\theta_F(k)-\theta_G(k))} = e^{j\theta(k)} \quad (3.3)$$

ここで, $\overline{G(k)}$ は $G(k)$ の複素共役を表す . また, 位相差スペクトル $\theta(k)$ は, $f(n)$ の位相成分と $g(n)$ の位相成分の差であり, $\theta(k) = \theta_F(k) - \theta_G(k)$ である . 1 次元 POC 関数 $r(n)$ は, 正規化相互パワースペクトル $R(k)$ の 1 次元逆離散フーリエ変換 (Inverse Discrete Fourier Transform: IDFT) として, 次式で定義される [34] .

$$r(n) = \frac{1}{N} \sum_{k=-M}^M R(k)W_N^{-kn} \quad (3.4)$$

以下では, 1 次元画像信号 $f(n)$ と, $f(n)$ が微小量 δ だけ平行移動した信号 $g(n)$ における POC について考える . まず, $f(n)$ を, 連続空間で定義された 1 次元信号 $f_c(x)$ を標本化間隔 T で標本化した信号として, 次式で定義する .

$$f(n) = f_c(x)|_{x=nT} \quad (3.5)$$

ここで, x は実数である . また, $f(n)$ を微小量 δ だけ平行移動した信号 $g(n)$ を, $f_c(x)$ を δ だけ平行移動した信号 $f_c(x - \delta)$ を標本化間隔 T で標本化した信号として, 次式で定義する .

$$g(n) = f_c(x - \delta)|_{x=nT} \quad (3.6)$$

ここで, δ は信号長 N と比べ十分に小さい ($\delta \ll N$) 実数であり, デジタル画像座標におけるサブピクセルレベルの平行移動量に相当するものとする . このとき, $f(n)$ および $g(n)$

の 1 次元 DFT $F(k)$ と $G(k)$ の間には，次式の近似が成り立つ*1．

$$\begin{aligned}
 G(k) &\simeq F(k)W_N^{k\delta} \\
 &= F(k)e^{-j\frac{2\pi}{N}k\delta} \\
 &= A_F(k)e^{j(\theta_F(k)-\frac{2\pi}{N}k\delta)}
 \end{aligned} \tag{3.7}$$

ここで，上式を本論文における理想的な平行移動モデルとする．式 (3.7) の近似を考慮すると， $f(n)$ と $g(n)$ の正規化相互パワースペクトル $R(k)$ は，

$$\begin{aligned}
 R(k) &= e^{j(\theta_F(k)-\theta_G(k))} \\
 &= e^{j(\theta_F(k)-\theta_F(k)+\frac{2\pi}{N}k\delta)} \\
 &= e^{j\frac{2\pi}{N}k\delta}
 \end{aligned} \tag{3.8}$$

となる．よって，式 (3.4) から， $f(n)$ と $g(n)$ の POC 関数 $r(n)$ は次式で与えられる．

$$\begin{aligned}
 r(n) &= \frac{1}{N} \sum_{k=-M}^M R(k)W_N^{-kn} \\
 &\simeq \frac{1}{N} \sum_{k=-M}^M e^{j\frac{2\pi}{N}k\delta} W_N^{-kn} \\
 &= \frac{1}{N} \sum_{k=-M}^M e^{j\frac{2\pi}{N}k(n+\delta)} \\
 &= \frac{1}{N} \frac{e^{-j\pi(n+\delta)} - e^{j\pi(n+\delta)}}{e^{-j\frac{\pi}{N}(n+\delta)} - e^{j\frac{\pi}{N}(n+\delta)}} \\
 &= \frac{1}{N} \frac{\sin(\pi(n+\delta))}{\sin\left(\frac{\pi}{N}(n+\delta)\right)}
 \end{aligned} \tag{3.9}$$

上式は，1 次元画像信号 $f(n)$ と $g(n)$ が理想的に微小量 δ だけ平行移動した場合における 1 次元 POC 関数の一般形である．式 (3.9) における $-\delta$ は，POC 関数のピーク座標に相当する．そのため，POC を用いた平行移動量推定では，1 次元画像信号 $f(n)$ と $g(n)$ から計算された POC 関数 $r(n)$ のピーク座標 $-\delta$ を求めることで， $f(n)$ と $g(n)$ の間の平行移動量を推定することができる．平行移動量 δ が整数の場合，POC 関数 $r(n)$ は Kronecker のデルタ関

*1 式 (3.7) が近似であることは，連続空間信号と離散空間信号に対するフーリエ変換の性質の違いに起因する．連続空間のフーリエ変換では，式 (3.7) が等式として成立する．

数になり，POC 関数 $r(n)$ が最大となる離散空間インデックス n から平行移動量 δ が計算できる．

$$\delta = - \operatorname{argmax}_n r(n) \quad (3.10)$$

一方，平行移動量 δ が非整数の場合，POC 関数 $r(n)$ のピーク座標は標本化間隔の間に存在するため，平行移動量 δ を式 (3.10) で推定できない．そこで，計算された POC 関数の離散値に対し，ピーク座標 $-\delta$ をパラメータとして式 (3.9) をフィッティングすることで，平行移動量 δ を推定する．図 3.2 は，POC を用いた平行移動量推定の例として，理想的に $\delta = 2.2$ だけ平行移動した 1 次元信号 $f(n)$ および $g(n)$ による計算結果である．図 3.2(d) に示すように，フィッティングされた式 (3.9) のピークモデルでは，標本化間隔の間にあるピーク座標が計算された $r(n)$ の離散値から推定されている．

式 (3.9) は，1 次元画像信号 $g(n)$ が 1 次元画像信号 $f(n)$ に対して理想的に δ だけ平行移動した場合の POC 関数である．しかしながら，実用上 $f(n)$ と $g(n)$ 間の変形が理想的な平行移動のみであることは少なく， $f(n)$ と $g(n)$ の間にはさまざまな画像変形やノイズが加わる． $g(n)$ にノイズを加えた場合のシミュレーション結果を図 3.3 に示す．図 3.3 に示す通り，1 次元画像信号 $f(n)$ と $g(n)$ の間の画像変形が平行移動のみの場合，その位相差スペクトル $\theta(k)$ は線形関数で表される^{*2}．一方で， $g(n)$ にノイズが加わると，位相差スペクトル $\theta(k)$ にもノイズが加わり，POC 関数 $r(n)$ のピーク値（ピークの高さ）が低くなる．このように，POC 関数のピーク値は，1 次元画像信号 $f(n)$ と $g(n)$ 間の変形が理想的な平行移動のみに近似できるかの指標として利用できる． $f(n)$ と $g(n)$ 間の変形が平行移動により近似できる場合，つまり， $f(n)$ と $g(n)$ の位相差スペクトル $\theta(k)$ の線形関数との誤差が小さいほど，POC 関数 $r(n)$ のピーク値は高くなる．そこで，実際のカメラ画像に適用する場合を考慮し，1 次元画像信号 $f(n)$ と $g(n)$ 間の変形が理想的な平行移動でない場合にも対処するため，式 (3.9)

^{*2} 図 3.3 において，線形関数である $\theta(k)$ が， $\theta(k) = -\pi$ および $\theta(k) = \pi$ となるインデックスで 2π だけ折り返しているのは，位相差スペクトル $\theta(k)$ が $[-\pi, \pi]$ の区間で循環しているためである．つまり，位相差スペクトルでは， π と $-\pi$ の区別がないことに注意されたい．

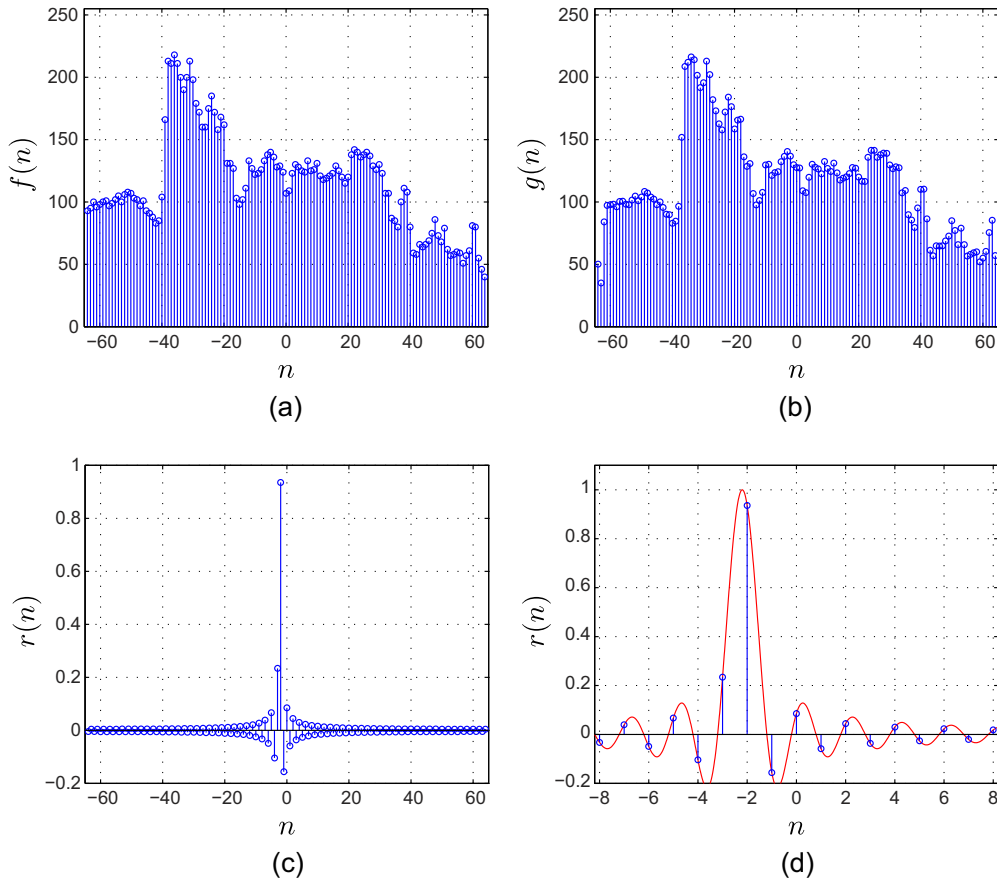


図 3.2 1 次元 POC を用いた平行移動量推定: (a) 1 次元信号 $f(n)$, (b) $f(n)$ に対して微小量 δ だけ平行移動した 1 次元信号 $g(n)$, (c) $f(n)$ と $g(n)$ の 1 次元 POC 関数 $r(n)$, (d) $r(n)$ の最大値付近での関数フィッティング結果

に POC 関数のピーク値を表す変数 α を加える .

$$r(n) = \frac{\alpha \sin(\pi(n + \delta))}{N \sin\left(\frac{\pi}{N}(n + \delta)\right)} \quad (3.11)$$

この式 (3.11) を 1 次元 POC 関数のピークモデルとする [34],[36] . また , サブピクセルの (非整数の) 平行移動量を推定する際の関数フィッティングにおけるパラメータはピーク値 α とピーク座標 $-\delta$ とする . 本論文では , 関数フィッティング手法として , 非線形最小 2 乗法の 1 つである Levenberg-Marquardt 法 (LM 法) [71] を用いる . LM 法により , 計算された POC 関数の離散値 $r(n)$ と POC 関数のピークモデルとの誤差が最小になるように α と δ を

求める．また，ノイズの影響が POC 関数のサイドローブに現れやすいことから，POC 関数 $r(n)$ の離散値が最大となる離散空間インデックス付近の離散値のみを関数フィッティングに使用することで，平行移動量推定に対するノイズの影響が抑制できる．多くの POC を用いた平行移動量推定の応用では，POC 関数の離散値が最大となる離散空間インデックス付近の 3 から 7 点のみを利用する．図 3.3 における POC 関数 $r(n)$ のグラフ上に描画した赤の実線は，式 (3.11) のフィッティング結果の一例である．

3.2.2 位相限定相関法の高精度化

以下では，実際のカメラ画像に適用する際に重要となる，1 次元 POC を用いた平行移動量推定の高精度化手法 [34]–[37] について述べる．

(I) 窓関数の適用

POC 関数を計算する際に利用する DFT では，取り扱う画像信号が周期的に循環することを仮定している．一方，実際にカメラ画像の平行移動量推定では，画像端で信号が循環していないため，マッチングを行う 1 次元画像信号 $f(n)$ と $g(n)$ 間に平行移動が存在すると， $f(n)$ と $g(n)$ で画像端の不連続な信号の現れ方が異なるという問題がある．この不連続性の影響を低減するために，マッチングを行う 1 次元画像信号 $f(n)$ および $g(n)$ に対して窓関数 $w(n)$ を適用する．本論文では，次式で定義される 1 次元ハニング窓を乗じる．

$$w(n) = \frac{1 + \cos\left(\frac{\pi n}{M}\right)}{2} \quad (3.12)$$

このとき，ハニング窓を乗じる前に，1 次元画像信号 $f(n)$ および $g(n)$ からそれぞれの平均値 \bar{f} と \bar{g} を引くことで，周波数領域の直流成分付近におけるハニング窓の影響を低減することができる．特に，信号長 N が小さいときには，ハニング窓の影響は無視できず，平均値を減算してからハニング窓を適用することで，平均値を減算しない場合に比べて精度が向上することを実験的に確認している．ステレオ画像間の局所的なマッチングウィンドウに対して POC を適用する場合，信号長が比較的短い場合が想定されるため，本論文では，常に，画像信号から平均値を減算してから窓関数を適用する．また，POC を用いた平行移動量推定では，信号長 N に対して，原理的に $\pm N/2$ の範囲の平行移動量しか推定できない．ハニング窓を適用する

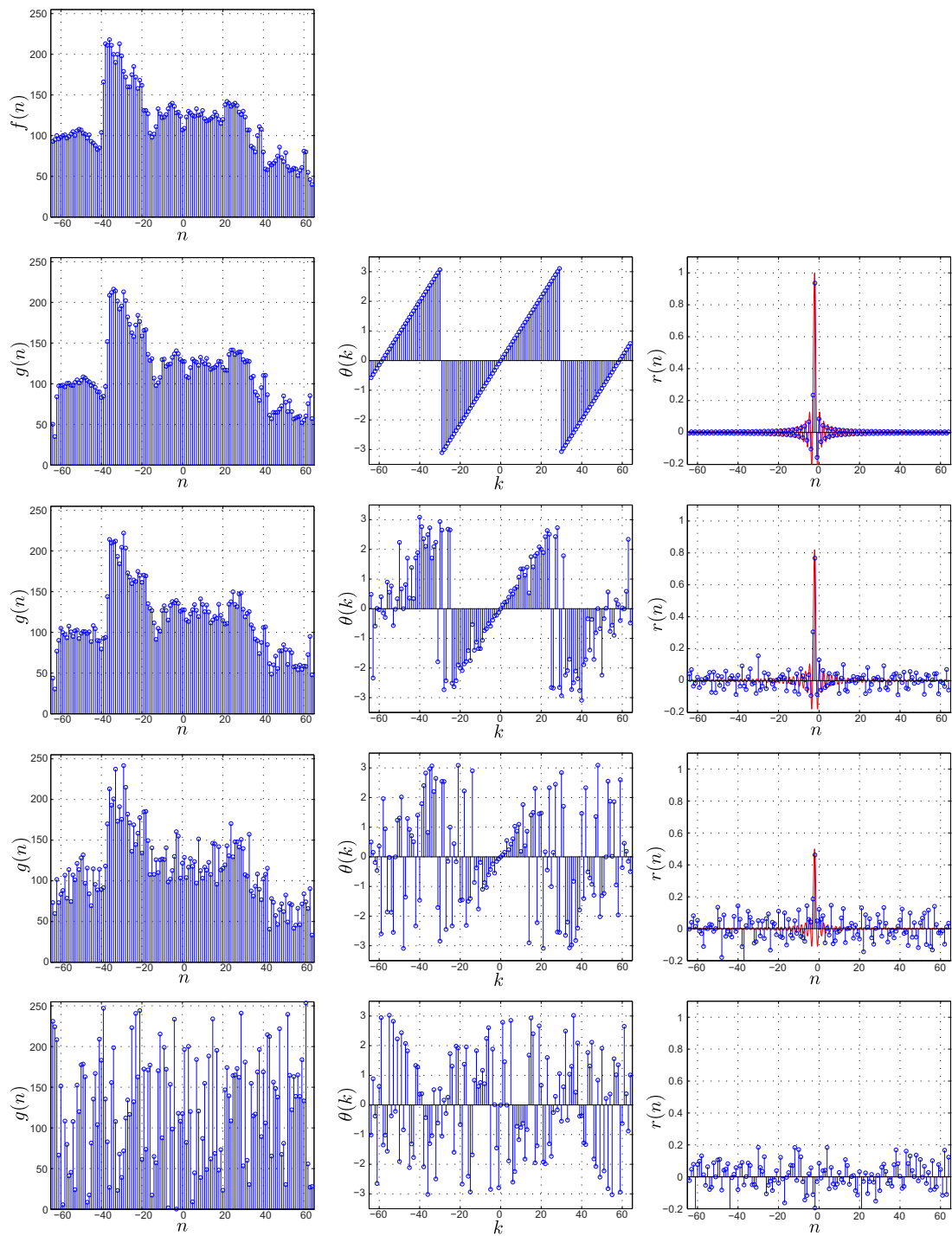


図 3.3 ノイズに対する POC 関数のふるまい (1 段目: 1 次元信号 $f(n)$, 2 段目: $f(n)$ を理想的に平行移動した 1 次元信号 $g(n)$ とその位相差スペクトル $\theta(k)$ および POC 関数 $r(n)$, 3 段目: $g(n)$ に微小な乱数を加えた場合の POC 関数 $r(n)$, 4 段目: $g(n)$ により大きな乱数を加えた場合の POC 関数 $r(n)$, 5 段目: $g(n)$ を乱数で生成した場合の POC 関数 $r(n)$)

と、有効な信号長が短くなるため、推定可能な平行移動量の範囲は $\pm N/4$ となることに注意されたい。

(II) スペクトルの重み付け

自然画像の場合、そのエネルギーは低周波数成分に集中し、高周波数成分のエネルギーは相対的に小さいことが知られている。そのため、カメラのノイズやエイリアシングなどの外乱が加わると、低周波数成分に比べ、高周波数成分の SN 比が低下する。一方で、POC における正規化相互パワースペクトル $R(k)$ の計算では、それぞれの信号の DFT $F(k)$ および $G(k)$ を各周波数成分の振幅の大小によらず、常に振幅を 1 に正規化して合成することを意味しており、相対的に SN 比が低い高周波数成分における外乱が強調されることになる。そこで、信頼性の低い高周波数成分の影響を抑制するために、正規化相互パワースペクトル $R(k)$ に対して、低域通過型のスペクトル重み付け関数 $H(k)$ を適用する。本論文では、次式で表されるガウス型スペクトル重み付け関数を乗じる。

$$H(k) = e^{-2\pi^2\sigma^2\frac{k^2}{N^2}} \quad (3.13)$$

ここで、 σ はスペクトル重み付け関数の半値幅を決定するための定数であり^{*3}、本論文では $\sigma = \sqrt{0.5}$ を使用する。このとき、正規化相互パワースペクトルを重み付けすることで、POC 関数のピークモデルも変化する。式 (3.13) のガウス型スペクトル重み付け関数を使用した際の 1 次元 POC 関数のピークモデルは、次式のようになる [34]。

$$\begin{aligned} r(n) &= \frac{1}{N} \sum_{k=-M}^M H(k)R(k)W_N^{-kn} \\ &\simeq \frac{\alpha}{\sqrt{2\pi\sigma}} e^{-\frac{(n+\delta)^2}{2\sigma^2}} \end{aligned} \quad (3.14)$$

ここで、 α は、POC 関数のピーク値を表す変数である^{*4}。式 (3.14) は、式 (3.13) を乗算した正規化相互パワースペクトルの正確な IDFT ではない。ガウス型スペクトル重み付け関数

^{*3} 一般的に、ガウス関数の数式において、分散を決定するための変数 σ は、指数関数の指数部の分母に記述される。式 (3.13) を σ が指数部の分子に乗算されるように定義した理由は、式 (3.14) のピークモデルを一般的なガウス関数の形にするためである。

^{*4} 式 (3.14) における α は、0 から 1 の値をとる。一方で、 $\alpha = 1$ の場合においても、ピーク座標における POC 関数の値は $\frac{1}{\sqrt{2\pi\sigma}}$ となることに注意されたい。

の信号端が十分に小さい場合 ($H(M) \ll 1$ の場合), 式 (3.14) の近似が成り立つ. 本論文で使用する $\sigma = \sqrt{0.5}$ の場合, 式 (3.14) が十分に近似として使用できることを実験的に確認している. また, スペクトル重み付けを用いる場合, 平行移動量推定のための関数フィッティングに用いるピークモデルは, 式 (3.14) となる. フィッティングパラメータは, ピーク値 α とピーク座標 $-\delta$ である.

(III) 複数の 1 次元画像信号の利用

1 次元 POC をステレオ画像間のウィンドウマッチングに適用する場合, 注目点の座標を中心にマッチングウィンドウを設定し, 設定されたウィンドウ間で 1 次元 POC 関数を計算する. 3.2.1 節では, 1 組の 1 次元画像信号のみを用いて平行移動量を推定する手法について説明した. しかしながら, ウィンドウマッチングのように信号長 N が比較的小さい場合, 実際のカメラ画像において 1 組の 1 次元画像信号のみを用いると, ノイズなどの外乱の影響により誤差が増加することを実験的に確認している. そこで, マッチングウィンドウを設定する際に, 注目点の近傍から複数の 1 次元画像信号の組を抽出し, それらの 1 次元画像信号の POC 関数を平均化して使用する [36], [37]. 図 3.4 は, 実際のステレオ画像における, 1 組の 1 次元画像信号における POC 関数と, 注目点近傍の 15 組の 1 次元信号から計算された POC 関数の平均の一例である. ここで, 信号長は $N = 32$ である. 図 3.4 において, 1 組の 1 次元画像信号のみを用いて求めた 1 次元 POC 関数の Peak-to-Noise Ratio (PNR) は低い, 15 個の 1 次元 POC 関数を平均化することで PNR が向上している. この 1 次元 POC 関数の平均をピーク値 α およびピーク座標 $-\delta$ の推定に用いることで, ノイズの影響を抑え, 高精度に平行移動量を推定することが可能である.

3.3 多視点ステレオのための位相限定相関法に基づくウィンドウマッチング

本節では, MVS のための POC に基づくウィンドウマッチングを提案する [31]–[33]. 従来の MVS のためのウィンドウマッチングでは, 2 つの画像信号から 1 つのマッチングスコアを算出するため, ある仮定した 3 次元点の奥行きが尤もらしいかどうかの判断にしか利用で

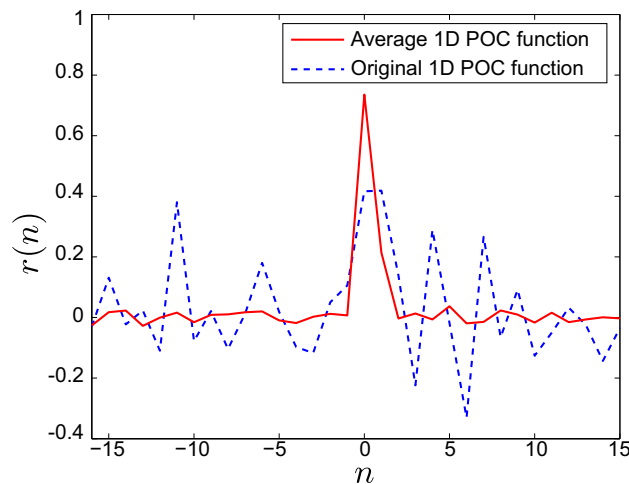


図 3.4 1 次元 POC 関数の平均化による Peak-to-Noise Ratio (PNR) の向上

きない [3], [9]. これに対し, MVS のための POC に基づくウィンドウマッチングでは, 1 回のウィンドウマッチングで, ある仮定した 3 次元点の奥行きが最も尤もらしい奥行きからどれくらい離れているかを推定することが可能である. このとき, 提案手法では, POC のピークモデルを用いた関数フィティングにより, 3 次元点の奥行きを高精度に推定する. また, 従来の POC に基づくステレオビジョンでは, 2 眼ステレオのみを対象にしていたため, 複数のステレオペアから計算される POC 関数を同一の座標系で扱うことができず, 3 視点以上の多視点画像間のウィンドウマッチングに対して POC を適用することができなかった [36], [37]. これに対し, 本論文で提案する MVS のための POC に基づくウィンドウマッチングでは, 正規化視差の概念に基づき, 異なるステレオペアから計算される POC 関数を同一の座標系で扱い, 統合することで, 多視点画像間のウィンドウマッチングに POC を適用する. 複数のステレオペアの POC 関数を用いることで, 従来の 2 眼ステレオの場合に比べ, ノイズやオクルージョンの影響を抑え, より高精度な奥行き推定が可能である.

以下では, まず, 単純化された条件について, ステレオビジョンにおける POC の適用方法について述べる. 次に, POC に基づくウィンドウマッチングを一般化するために, 正規化視差の概念について説明し, 正規化視差による POC の統合について述べる. そして, MVS のための POC に基づくウィンドウマッチングを提案し, 提案手法を用いた奥行き推定について述べる.

3.3.1 ステレオビジョンにおける位相限定相関法の適用

はじめに、非常に単純化した例について、ステレオビジョンにおけるカメラの 3 次元幾何と、POC のステレオ画像間のウィンドウマッチングへの適用について説明する。ステレオ画像を参照視点 V_R と近傍視点 C とし、ステレオペア V_R-C を平行ステレオペアとする。このとき、平行ステレオペア V_R-C について、参照視点 V_R 上のある 2 次元座標 $m = [u, v]^T$ に対応する視線上の 3 次元点 $M = [X, Y, Z]^T$ を求める問題を考える。ここで、3 次元点 M の座標系は参照視点 V_R のカメラ座標系とする。平行ステレオペア V_R-C はエピポラ線が水平座標軸と平行であり、その基線長を B とする。また、参照視点 V_R と近傍視点 C それぞれの内部パラメータ A および A' を以下のように設定する。

$$A = A' = \begin{bmatrix} \beta & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.15)$$

ここで、 β はカメラの焦点距離を、 (u_0, v_0) は画像中心を表す^{*5}。このとき、 M の参照視点 V_R 上の投影点 m および、近傍視点 C 上の投影点 $m' = [u', v']^T$ は以下の式で表される。

$$s\tilde{m} = AM \quad (3.16)$$

$$s\tilde{m}' = A(M - [B, 0, 0]^T) \quad (3.17)$$

ここで、 s はスケールを表す変数であり、 \tilde{m} は m の同次座標表現である。また、 m と m' の間に以下の関係が成り立つ。

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} u - \frac{\beta B}{Z} \\ v \end{bmatrix} = \begin{bmatrix} u - d \\ v \end{bmatrix} \quad (3.18)$$

上式で、 $d = u - u' = \frac{\beta B}{Z}$ は一般に視差と呼ばれる [4]。それぞれの幾何関係を図 3.5 に示す。このとき、対称物体の形状が 3 次元点 M において、平行ステレオペアの画像平面と平行な平面であると仮定する。画像化の際にノイズなどが生じない理想的な撮影では、参照視点 V_R 上で m を中心に設定されたマッチングウィンドウ（エピポラ線上の局所的な 1 次元画像信

^{*5} 一般的に、焦点距離は f で表されるが [4]、本論文では、マッチングを行う 1 次元画像信号 f との混同を避けるため、 β を用いる。

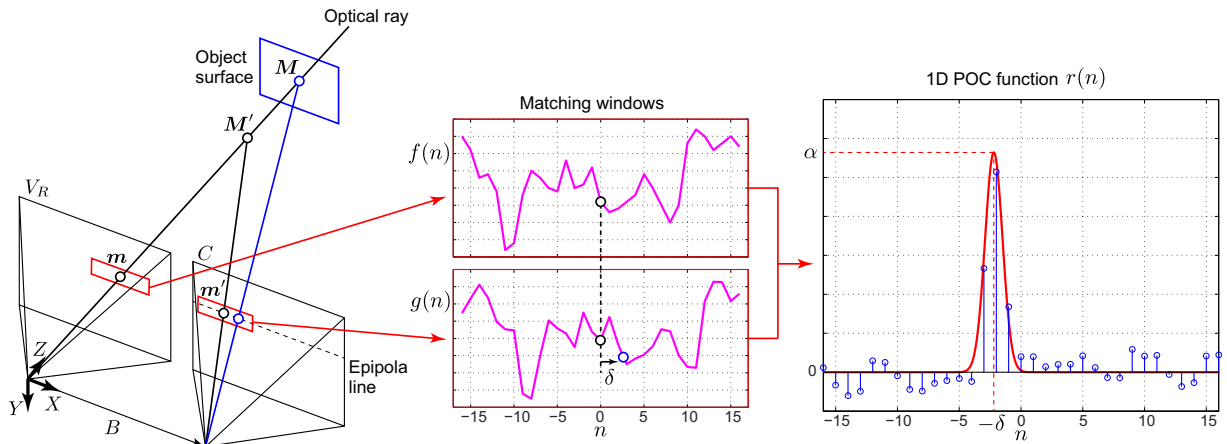


図 3.5 平行ステレオにおける 1 次元 POC に基づくウィンドウマッチング

号) の $f(n)$ と、近傍視点 C 上で m' を中心に設定されたマッチングウィンドウ $g(n)$ は、3次元点 M における対象物体表面のテクスチャを連続空間で定義した信号 $f_c(x)$ を同一の標本化間隔 $T = Z/\beta$ で標本化した同一の信号として表される。

$$f(n) = g(n) = f_c(x)|_{x=nT} \quad (3.19)$$

次に、参照視点 V_R 上の視線上で、真の 3 次元点 M から $\Delta M = [\Delta X', \Delta Y', \Delta Z']^T$ だけ平行移動した 3 次元候補点 $M' = [X', Y', Z']^T$ が与えられた場合を考える。ここで、 $M' = M - \Delta M$ である。このとき、 M' の参照視点 V_R 上の投影点 m と、近傍視点 C 上の投影点 m' との間には以下の関係が成り立つ。

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} u - \frac{\beta B}{Z'} \\ v \end{bmatrix} \quad (3.20)$$

上式において、 M' に対応する視差 d' を $d' = d + \delta = \frac{\beta B}{Z'}$ とする。このとき、参照視点 V_R 上で m を中心に設定されたマッチングウィンドウ $f(n)$ は式 (3.19) と変わらないが、近傍視点 C 上で m' を中心に設定されたマッチングウィンドウ $g(n)$ は、 δ だけ平行移動して標本化される。

$$g(n) = f_c(x - \delta)|_{x=nT} \quad (3.21)$$

上式のように、3 次元候補点 M' が真の 3 次元点 M からずれている場合、その投影点 m および m' を中心に設定されたマッチングウィンドウ $f(n)$ と $g(n)$ 間には平行移動が生じる。

このマッチングウィンドウ間の平行移動量 δ は、3 次元候補点 M' と真の 3 次元点 M との位置ずれ量 ΔM と大きく関係する。よって、ある 3 次元候補点 M' が与えられた場合、その投影点を中心にマッチングウィンドウを設定し、ウィンドウ間の平行移動量を求めることで、真の 3 次元点 M を推定することが可能である。

$$M = \begin{bmatrix} (u - u_0)B/(d' - \delta) \\ (v - v_0)B/(d' - \delta) \\ \beta B/(d' - \delta) \end{bmatrix} \quad (3.22)$$

ステレオビジョンに POC を適用する場合、このマッチングウィンドウ間の平行移動量推定に POC を利用する。3 次元候補点 M' が真の 3 次元点 M に十分近い場合、つまり、 M がマッチングウィンドウ内に投影される場合には、POC を用いて平行移動量を推定することができる。3.2.2 節の窓関数の適用を行う場合、有効な情報をもつ信号幅が信号長よりも短くなるため、実用上は、 M が $\pm N/4$ の範囲に投影される必要がある。3 次元候補点 M' と真の 3 次元点 M のずれ量が大きい場合、つまり、 M がマッチングウィンドウ内に投影されない場合には、物体表面テクスチャが繰り返しパターンでない限り、 $f(n)$ と $g(n)$ 間の変形が平行移動のみ近似できなくなるため、POC 関数のピーク値 α が著しく低下する。このように、POC 関数のピーク値を用いることで、初めに仮定した 3 次元候補点 M' が、真の 3 次元点 M の座標を計算するのに妥当か判断することができる。つまり、3 次元候補点 M' を参照視点 V_R における視線上で変化させながら POC に基づくウィンドウマッチングを繰り返し、最も POC 関数のピーク値 α が高くなるときに算出された 3 次元点 M を真の 3 次元点として選択することで、視線上の 3 次元座標を推定できる。また、実際のステレオ画像に適用する場合、高精度な 3 次元復元のためには、3.2.2 節で述べたように複数の 1 次元画像信号を利用することが重要である。利用する 1 次元画像信号のライン数を L とすると、マッチングウィンドウが投影点を中心に $N \times L$ ピクセルの領域に設定され、そのマッチングウィンドウについて L 個の POC 関数が平均化され、3 次元座標の算出に利用される。

3.3.2 正規化視差による位相限定相関関数の統合

ここまでは、1組のステレオペアについて、ステレオ画像間のウィンドウマッチングにより3次元点の奥行きを算出する方法について述べた。一方で、1組のステレオペアでは、画像のノイズにより正しい奥行きでもPOC関数のピーク値が低下したり、繰り返しパターンなどにより間違った奥行きでも高いピーク値が計算されたりすると、誤対応点が発生する。これに対して、多くのMVSアルゴリズムでは、ウィンドウマッチングの際に、複数のステレオペアで得られたマッチングスコアを統合することで、3次元復元精度およびロバスト性を向上させている。複数のステレオペアのマッチング結果を統合することで、ノイズや繰り返しパターンなどの影響を低減し、誤対応の発生を抑制できる。また、ある1組のステレオペアにおいてオクルージョンとなる領域についても、異なる基線長方向のステレオペアではオクルージョンでなくなるため、1つの参照視点から復元可能な領域が多くなるといった利点もある。本論文で提案するMVSのためのPOCに基づくウィンドウマッチングにおいても、複数のステレオペアで計算されたマッチング結果を統合することで、その精度・ロバスト性の向上が期待される。しかしながら、異なるステレオペアで計算された1次元POC関数では、ステレオペアにおける視差の違いから、そのピーク座標が異なるという問題がある。このような、ピーク座標の異なるPOC関数を統合することは困難である。そこで、本論文では、複数のステレオペアで計算されたPOC関数を統合するために、正規化視差の概念を導入する[31]–[33]。従来の多視点画像間における相関関数の統合手法として、Okutomiらの手法[81]があるが、Okutomiらの手法では、すべてのカメラが直線上にあることを仮定しており、一般的な多視点画像に適用できない。本論文で提案する正規化視差に基づく相関関数の統合は、Okutomiらの手法をより一般化したものであり、カメラが同一の直線上にない場合においても相関関数を統合することが可能である。

カメラパラメータが既知の多視点画像 $V = \{V_0, \dots, V_{K_{\text{all}}-1}\}$ のうち、参照視点 $V_R \in V$ とその近傍視点 $C = \{C_0, \dots, C_{K-1}\} \subseteq V - \{V_R\}$ を入力として考える。ここで、 K_{all} および K は、それぞれ多視点画像の数と近傍視点の数である。 V_R および C から K 組のステレオペア V_R-C_i ($i = 0, \dots, K-1$) を得る。それぞれのステレオペアについて、ステレオ平行

化を行い，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ を生成する．この K 組の平行ステレオペアについて，各ステレオペアで計算された 1 次元 POC 関数を正規化視差に基づき統合し，3 次元点の奥行き推定を行う．

まず，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ における 3 次元点 $M = [X, Y, Z]^T$ のステレオ画像上の視差を考える．ここで， $C_i \in C$ である．また，3 次元点の世界座標は参照視点 V_R のカメラ座標と一致するように設定する．参照視点 V_R と近傍視点 C_i でステレオ平行化を行う際に，参照視点におけるカメラ座標の回転行列を R_i とし，次式で表す．

$$R_i = \begin{bmatrix} R_{i11} & R_{i12} & R_{i13} \\ R_{i21} & R_{i22} & R_{i23} \\ R_{i31} & R_{i32} & R_{i33} \end{bmatrix} \quad (3.23)$$

このとき，世界座標における 3 次元点 M と平行化後のカメラ座標における 3 次元点 $M_i^{\text{rect}} = [X_i, Y_i, Z_i]^T$ の関係は次式で表される．

$$M = R_i M_i^{\text{rect}} \quad (3.24)$$

また，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ における 3 次元点 M_i^{rect} とその視差 d_i の関係は次式で表される [4] ．

$$M_i^{\text{rect}} = \begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \begin{bmatrix} (u_i - u_{0i})B_i/d_i \\ (v_i - v_{0i})B_i/d_i \\ \beta_i B_i/d_i \end{bmatrix} \quad (3.25)$$

ここで， (u_i, v_i) は平行化後の参照視点 $V_{R,i}^{\text{rect}}$ における 3 次元点 M_i^{rect} の対応点のデジタル画像座標を， (u_{0i}, v_{0i}) は平行化後の参照視点 $V_{R,i}^{\text{rect}}$ における画像中心を， β_i は平行化後の焦点距離を， B_i は平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ の基線長を表す．式 (3.24) と式 (3.25) より，3 次元点 M と平行化後の視差 d_i との関係は次式で表される．

$$M = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R_i \begin{bmatrix} (u_i - u_{0i})B_i/d_i \\ (v_i - v_{0i})B_i/d_i \\ \beta_i B_i/d_i \end{bmatrix} \quad (3.26)$$

同様に，参照視点 V_R と近傍視点 $C_j \in C - \{C_i\}$ の間で次式が成り立つ．

$$M = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{R}_j \begin{bmatrix} (u_j - u_{0j})B_j/d_j \\ (v_j - v_{0j})B_j/d_j \\ \beta_j B_j/d_j \end{bmatrix} \quad (3.27)$$

式 (3.26) と式 (3.27) より，

$$\begin{aligned} X &= \frac{B_i}{d_i}(R_{i11}(u_i - u_{0i}) + R_{i12}(v_i - v_{0i}) + R_{i13}\beta_i) \\ &= \frac{B_j}{d_j}(R_{j11}(u_j - u_{0j}) + R_{j12}(v_j - v_{0j}) + R_{j13}\beta_j) \end{aligned} \quad (3.28)$$

$$\begin{aligned} Y &= \frac{B_i}{d_i}(R_{i21}(u_i - u_{0i}) + R_{i22}(v_i - v_{0i}) + R_{i23}\beta_i) \\ &= \frac{B_j}{d_j}(R_{j21}(u_j - u_{0j}) + R_{j22}(v_j - v_{0j}) + R_{j23}\beta_j) \end{aligned} \quad (3.29)$$

$$\begin{aligned} Z &= \frac{B_i}{d_i}(R_{i31}(u_i - u_{0i}) + R_{i32}(v_i - v_{0i}) + R_{i33}\beta_i) \\ &= \frac{B_j}{d_j}(R_{j31}(u_j - u_{0j}) + R_{j32}(v_j - v_{0j}) + R_{j33}\beta_j) \end{aligned} \quad (3.30)$$

となり，3次元点 M に対する，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ における視差 d_i と，平行ステレオペア $V_{R,j}^{\text{rect}}-C_j^{\text{rect}}$ における視差 d_j との間には次の関係が成り立つ．

$$\begin{aligned} d_i &= \frac{R_{i11}(u_i - u_{0i}) + R_{i12}(v_i - v_{0i}) + R_{i13}\beta_i}{R_{j11}(u_j - u_{0j}) + R_{j12}(v_j - v_{0j}) + R_{j13}\beta_j} \frac{B_i}{B_j} d_j \\ &= \frac{R_{i21}(u_i - u_{0i}) + R_{i22}(v_i - v_{0i}) + R_{i23}\beta_i}{R_{j21}(u_j - u_{0j}) + R_{j22}(v_j - v_{0j}) + R_{j23}\beta_j} \frac{B_i}{B_j} d_j \\ &= \frac{R_{i31}(u_i - u_{0i}) + R_{i32}(v_i - v_{0i}) + R_{i33}\beta_i}{R_{j31}(u_j - u_{0j}) + R_{j32}(v_j - v_{0j}) + R_{j33}\beta_j} \frac{B_i}{B_j} d_j \end{aligned} \quad (3.31)$$

式 (3.31) で示したように，異なる平行ステレオペアにおける視差 d_i と d_j はある倍率によって関係付けられ，その倍率はカメラパラメータと参照視点における 3次元点の対応点のデジタル画像座標に依存する．以上より，各ステレオペアにおける視差の倍率を考慮することで，正規化視差 d を定義する．平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ ($i = 0, \dots, K - 1$) が与えられたとき，各平行ステレオペアにおける視差 d_i と正規化視差 d の間には以下の関係が成り立つ．

$$d_i = s_i d \quad (3.32)$$

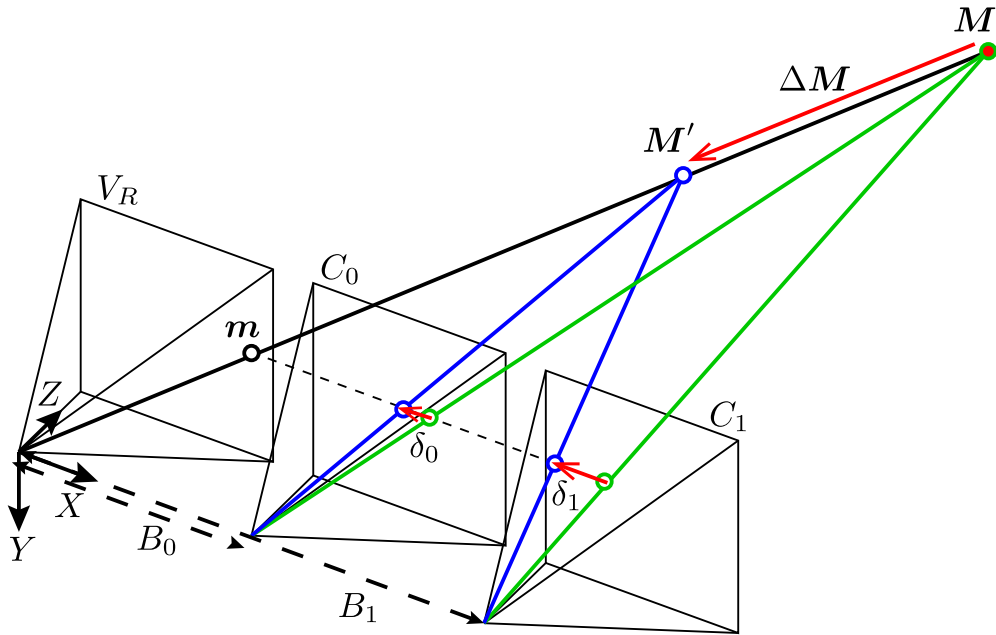


図 3.6 3次元点の変移と視差の変移の関係

ここで、 s_i は視差の倍率である．平行化前の参照視点における注目点 (u, v) について，平行化後の各平行ステレオペアにおける視差倍率 s_i を式 (3.31) から次式のように定義する [32] ．

$$s_i = \frac{(R_{i31}(u_i - u_{0i}) + R_{i32}(v_i - v_{0i}) + R_{i33}\beta_i)B_i}{\frac{1}{K} \sum_{l=0}^{K-1} (R_{l31}(u_l - u_{0l}) + R_{l32}(v_l - v_{0l}) + R_{l33}\beta_l)B_l} \quad (3.33)$$

このとき，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ における 3次元点 M_i^{rect} と正規化視差 d の関係は，次式で与えられる．

$$M_i^{\text{rect}} = \begin{bmatrix} (u_i - u_{0i})B_i/(s_i d) \\ (v_i - v_{0i})B_i/(s_i d) \\ \beta_i B_i/(s_i d) \end{bmatrix} \quad (3.34)$$

次に，図 3.6 のように，3次元点 M に対して，参照視点 V_R の視線を微小量 $\Delta M = (\Delta X, \Delta Y, \Delta Z)$ だけ平行移動させた点 $M' = M - \Delta M$ について，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ ($i \in \{0, \dots, K-1\}$) における視差の変動について考える．式 (3.26) より，3次元点 M' は，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ における視差 d'_i を用いて

$$M' = \mathbf{R}_i \begin{bmatrix} (u_i - u_{0i})B_i/d'_i \\ (v_i - v_{0i})B_i/d'_i \\ \beta_i B_i/d'_i \end{bmatrix} \quad (3.35)$$

と表される．ここで， M を真の 3 次元点とすると，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ における視差に変移 δ_i が生じるので， M に対する視差 d_i と M' に対する視差 d'_i の関係は

$$d'_i = d_i + \delta_i \quad (3.36)$$

となる．このとき，3.3.1 節で述べたように，局所的な画像変形を平行移動のみと仮定すると， M' の対応点を中心に $V_{R,i}^{\text{rect}}$ から切り出したウィンドウ f_i と C_i^{rect} から切り出したウィンドウ g_i の間には δ_i の平行移動が生じていると考えられる．この平行移動量 δ_i は， f_i と g_i の 1 次元 POC 関数 r_i を計算し，そのピーク座標を推定することで求められる．しかしながら，異なる平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ および $V_{R,j}^{\text{rect}}-C_j^{\text{rect}}$ ($j \in \{0, \dots, K-1\} - \{i\}$) において，平行移動量 δ_i と δ_j は必ずしも一致せず，それぞれの平行ステレオペアから切り出されたウィンドウで計算される 1 次元 POC 関数 r_i と r_j のピーク座標は異なる．例えば，図 3.6 では，ステレオペア $V_R - C_0$ の基線長 B_0 よりも $V_R - C_1$ の基線長 B_1 の方が長く，同じ奥行きの変化 ΔZ に対して $V_R - C_1$ 上の対応点の変動 δ_1 の方が $V_R - C_0$ 上の対応点の変動 δ_0 よりも大きくなる．そのため， $V_R - C_0$ から計算される POC 関数のピーク座標 $-\delta_0$ と $V_R - C_1$ から計算される POC 関数のピーク座標 $-\delta_1$ は異なっており，このようなピーク座標が異なる POC 関数を統合することは困難である．

そこで，MVS のための POC に基づくウィンドウマッチングでは，正規化視差における視差倍率 s_i に基づき各平行ステレオペアで切り出すウィンドウを拡大縮小することで，異なる平行ステレオペアで計算される 1 次元 POC 関数のピーク座標を一致させる． M における正規化視差を d ， M' における正規化視差を $d' = d + \delta$ とすると式 (3.34) より M' は

$$\mathbf{M}' = \mathbf{R}_i \begin{bmatrix} (u_i - u_{0i})B_i/(s_i(d + \delta)) \\ (v_i - v_{0i})B_i/(s_i(d + \delta)) \\ \beta_i B_i/(s_i(d + \delta)) \end{bmatrix} \quad (3.37)$$

と表される．よって，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ において対応点を中心に切り出したウィンドウ f_i と g_i 間の平行移動量は $s_i\delta$ となる．ここで，基準となるウィンドウサイズ N に対して， f_i, g_i を切り出す際のウィンドウサイズを s_iN とする．この s_iN の大きさの 1 次元画像信号 f_i, g_i を N の大きさにそろえるため， $1/s_i$ 倍に拡大縮小した信号をそれぞれ \hat{f}_i, \hat{g}_i とする．このとき， \hat{f}_i, \hat{g}_i 間の平行移動量は δ となり， \hat{f}_i, \hat{g}_i から計算される 1 次元 POC 関

数 \hat{r}_i のピーク座標は $-\delta$ となる．平行ステレオペア $V_{R,j}^{\text{rect}}-C_j^{\text{rect}}$ ($j \in \{0, \dots, K-1\} - \{i\}$) について同様に考えると，切り出すウィンドウのサイズ $s_j N$ は $s_i N$ と異なるが，1次元 POC 関数 \hat{r}_j のピーク座標は $-\delta$ となり，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ で計算される 1次元 POC 関数 \hat{r}_i のピーク座標と一致する．実際に計算される POC 関数の例を図 3.7 に示す．図 3.7(a) は，視差の倍率を考慮せず，すべての平行ステレオペアにおいて同じウィンドウサイズ N でウィンドウを切り出した場合の POC 関数である．この場合，ウィンドウ間の平行移動量 δ_i はそれぞれのステレオペアごとに異なり，計算される POC 関数のピーク座標は一致しない．これに対して，図 3.7(b) は，各ステレオペアにおいて，正規化視差における視差倍率 s_i に基づき，切り出すウィンドウサイズを $s_i N$ とした場合の POC 関数である．図 3.7(b) より，視差の倍率を考慮することで，それぞれのステレオペアにおける POC 関数のピーク座標 $-\delta$ が一致すること確認できる．以上より，異なる平行ステレオペアで計算される 1次元 POC 関数を統合することが可能となる．本論文では，複数の 1次元 POC 関数の統合として，各平行ステレオペアで計算された 1次元 POC 関数 \hat{r}_i ($i = 0, \dots, K-1$) を平均化し，平均 POC 関数 \hat{r}_{ave} とする．この平均 POC 関数 \hat{r}_{ave} を 3次元点の奥行き推定に用いる．

ここで，オクルージョンにより 3次元点 M が近傍視点 $C_i \in C$ の画像に写っていない場合や，物体境界において複数の視差を持つ領域でウィンドウを抽出した場合は，そのウィンドウから計算される 1次元 POC 関数 \hat{r}_i に非常に大きい誤差が生じることが予想される．一方で，マッチングウィンドウ間の画像変形が平行移動のみに近似できない場合は，そのウィンドウから計算される 1次元 POC 関数のピーク値 α_i が低下する．そこで，複数のステレオペアにおける 1次元 POC 関数の平均化の際に，POC 関数のピーク値 α_i が閾値 th_{corr} を超えるもののみを用いて計算することで，オクルージョンや物体境界の影響などを抑えることが可能である．

3.3.3 位相限定相関法に基づくウィンドウマッチングによる奥行き推定

本小節では，3.3.1 節，および，3.3.2 節に基づき，MVS のための POC に基づくウィンドウマッチングを提案する [32]．以下では，参照視点 $V_R \in V$ とその近傍視点 $C = \{C_0, \dots, C_{K-1}\} \subseteq V - \{V_R\}$ における，参照視点 V_R 上の任意の点 $m = [u, v]^T$ について，

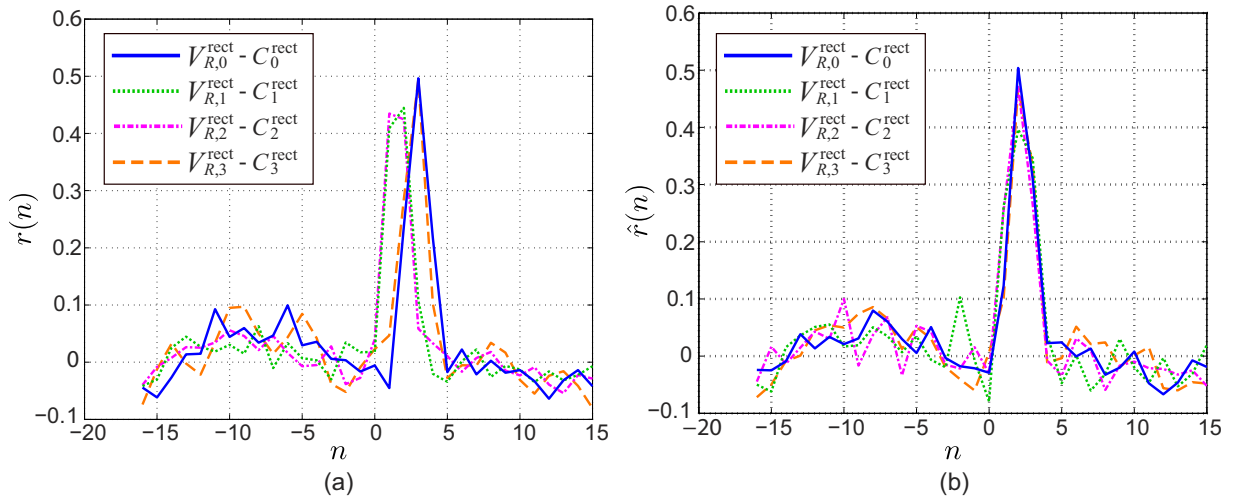


図 3.7 正規化視差に基づく POC 関数の統合: (a) ウィンドウサイズ N のマッチングウィンドウから計算された 4 組の 1 次元 POC 関数 r_i , (b) ウィンドウサイズ $s_i N$ のマッチングウィンドウから計算された 4 組の 1 次元 POC 関数 \hat{r}_i

その点の奥行き初期値 Z_{init} が与えられたときに、提案手法を用いて、 m に対応する視線上の 3 次元点 M と真の奥行き Z を求める方法について説明する。処理の流れを図 3.8 に示す。

Step1: 参照視点 V_R の内部パラメータ A_{V_R} と、デジタル画像座標 m 、奥行き初期値 Z_{init} から、3 次元候補点 M' を決定する。

$$M' = Z_{\text{init}} A_{V_R}^{-1} \tilde{m} \tag{3.38}$$

Step2: 参照視点 V_R と近傍視点 C から平行ステレオペア $V_{R,i}^{\text{rect}} - C_i^{\text{rect}}$ ($i = 0, \dots, K - 1$) を生成し、3 次元候補点 M' の各平行ステレオペアにおける対応点の座標を求める。

$V_{R,i}^{\text{rect}} - C_i^{\text{rect}}$ について、 M' の $V_{R,i}^{\text{rect}}$ 上の対応点を $m_i = [u_i, v_i]^T$ 、 C_i^{rect} 上の対応点を

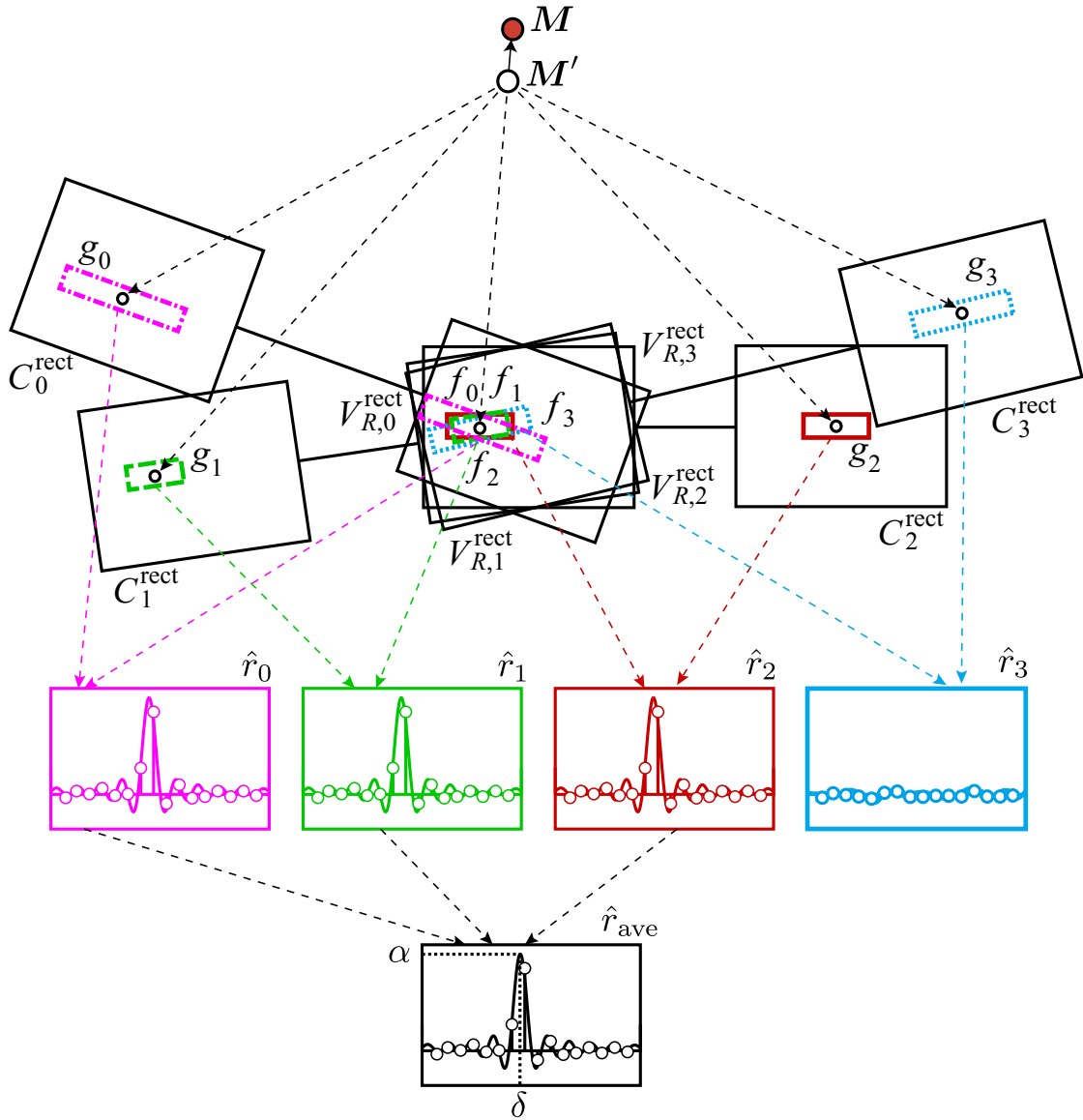


図 3.8 MVS のための POC に基づくウィンドウマッチングによる奥行き推定

$m'_i = [u'_i, v'_i]^T$ とすると, m_i および m'_i は次式で決定される *6 .

$$s\tilde{m}_i = A_i R_i^{-1} M' \tag{3.39}$$

$$s\tilde{m}'_i = A'_i (R_i^{-1} M' - t_{cam,i}) \tag{3.40}$$

ここで, A_i は平行化後の参照視点 $V_{R,i}^{rect}$ の内部パラメータを, A'_i は平行化後の近傍

*6 参照視点 V_R では, ステレオペアとなる近傍視点 $C_i \in C$ のカメラパラメータに応じて, 平行化後の参照視点 $V_{R,i}^{rect}$ の座標系が異なる. そのため, 本来 1 枚の画像である参照視点 V_R についても, 平行化後の M' の対応点座標 m_i は, ステレオペアごとに異なることに注意されたい.

視点 C_i^{rect} の内部パラメータを, R_i は平行化におけるカメラ座標系の回転を, $t_{\text{cam},i}$ は平行化後のカメラの並進ベクトルを表す. $t_{\text{cam},i}$ は, 平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ の基線長 B_i について, $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ をエピポラ線が水平座標軸と平行になるように平行化した場合 $t_{\text{cam},i} = [B_i, 0, 0]^T$, $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ をエピポラ線が垂直座標軸と平行になるように平行化した場合 $t_{\text{cam},i} = [0, B_i, 0]^T$ で表される. また, 3次元候補点 M' の各ステレオペアにおける視差を d'_i とする. d'_i は, $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ が水平ステレオの場合 $d'_i = (u_i - u_{0i}) - (u'_i - u'_{0i})$, $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ が垂直ステレオの場合 $d'_i = (v_i - v_{0i}) - (v'_i - v'_{0i})$ で表される. ここで, (u_{0i}, v_{0i}) は平行化後の参照視点 $V_{R,i}^{\text{rect}}$ の画像中心を, (u'_{0i}, v'_{0i}) は平行化後の近傍視点 C_i^{rect} の画像中心を表す.

Step3: $i = 0, \dots, K - 1$ について, m における視差倍率 s_i を式 (3.33) から決定する. このとき, 各ステレオペアにおける視差 d_i と正規化視差 d の関係は, 視差倍率 s_i を用いて $d_i = s_i d$ と表される. 平行化後の参照視点 $V_{R,i}^{\text{rect}}$ から座標 m_i を中心に $s_i N \times L$ ピクセルで切りだされたマッチングウィンドウを f_i , 近傍視点 C_i^{rect} から座標 m'_i を中心に $s_i N \times L$ ピクセルで切りだされたマッチングウィンドウを g_i とする. ここで, N は基準となるマッチングウィンドウの信号長を, L はライン数を示す. f_i および g_i を $N \times L$ の大きさにそろえるため, エピポラ線方向に $1/s_i$ 倍に拡大縮小し, \hat{f}_i, \hat{g}_i とする. \hat{f}_i および \hat{g}_i から 1次元 POC 関数 \hat{r}_i を計算し, そのピーク値を α_i とする.

Step4: \hat{r}_i のうち POC 関数のピーク値 α_i が th_{corr} 以上のものを平均化し, \hat{r}_{ave} とする. \hat{r}_{ave} に対し, 関数フィッティングを行い, POC 関数のピーク値 α とピーク座標 $-\delta$ を得る. δ より, 次式を用いて真の 3次元点 $M = [X, Y, Z]^T$ を算出する.

$$M = R_i \begin{bmatrix} (u_i - u_{0i})B_i / (s_i(d' - \delta)) \\ (v_i - v_{0i})B_i / (s_i(d' - \delta)) \\ \beta_i B_i / (s_i(d' - \delta)) \end{bmatrix} \quad (3.41)$$

以上の Step1 から Step4 により, m に対応する視線上の 3次元点 M と真の奥行き Z を得る. また, 奥行き初期値 Z_{init} (もしくは, 3次元候補点 M') の奥行き推定における妥当性を示す値として, POC 関数のピーク値 α を 3次元点 M の信頼度とする.

3.4 マッチングウィンドウの幾何補正による高精度化

3.3 節では，マッチングウィンドウ間の画像変形を平行移動のみ近似するため，対象物体の表面形状が局所的に参照視点 V_R の画像平面と平行な平面であると仮定した．実際には，復元する対象物体が必ずしも画像平面と平行でなく，むしろ，この仮定が成り立たないことが多い．そのため，単純な長方形のマッチングウィンドウでは，ステレオ画像間の変形が大きい場合に，マッチングウィンドウ間の変形を平行移動のみに近似できなくなり，POC に基づくウィンドウマッチングの奥行き推定精度が低下する．特に，基線長が長いステレオペアや，対象物体がカメラに対して大きく傾いている場合には，ステレオ画像間の画像変形が大きくなるため，マッチングウィンドウ間の変形を無視できない．そこで，MVS のための POC に基づくウィンドウマッチングに対して，マッチングウィンドウの幾何補正による高精度化を行う [38], [39], [82]–[84]．本論文では，対象物体の局所的な形状を平面に近似し，その平面のパラメータ（3次元座標と法線ベクトル）からマッチングウィンドウ間の変形を算出し，その画像変形を軽減するようにマッチングウィンドウの形状を変化させる．対象物体の局所的な形状を考慮したマッチングウィンドウ間の幾何補正を行うことで，ステレオ画像間の変形が大きい場合においても，高精度に奥行きを推定することができる．

本節では，MVS のための POC に基づくウィンドウマッチングにおける，マッチングウィンドウの幾何補正による高精度化について述べる．まず，3次元平面によるステレオ画像間の画像変形について説明する．次に，3次元平面を仮定した際の，マッチングウィンドウ間の幾何補正手法を提案し，MVS のための POC に基づくウィンドウマッチングの高精度化を行う．

3.4.1 3次元平面によるステレオ画像間の変形

ある3次元平面を平行ステレオペア $V_R^{\text{rect}}-C^{\text{rect}}$ で撮影した場合の，参照視点 V_R^{rect} と近傍視点 C^{rect} 間の画像変形について説明する．ここで，3次元平面を表すパラメータを，平面上の3次元点 $M' = [X', Y', Z']^T$ および法線ベクトル $n = [n_X, n_Y, n_Z]^T$ とする．また，参照視点 V_R^{rect} の内部パラメータ A ，および，近傍視点 C^{rect} の内部パラメータ A' をそれぞれ

次式で表す．

$$\mathbf{A} = \begin{bmatrix} \beta & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.42)$$

$$\mathbf{A}' = \begin{bmatrix} \beta & 0 & u'_0 \\ 0 & \beta & v'_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.43)$$

はじめに，平行ステレオペア $V_R^{\text{rect}}-C^{\text{rect}}$ において，エピポーラ線が水平座標軸に平行である場合について考える．このとき， $V_R^{\text{rect}}-C^{\text{rect}}$ 間のカメラ座標系の回転 \mathbf{R}_{cam} および並進ベクトル \mathbf{t}_{cam} は次式で与えられる．

$$\mathbf{R}_{\text{cam}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.44)$$

$$\mathbf{t}_{\text{cam}} = \begin{bmatrix} B \\ 0 \\ 0 \end{bmatrix} \quad (3.45)$$

ここで， B は平行ステレオペア $V_R^{\text{rect}}-C^{\text{rect}}$ の基線長である．また，水平ステレオの場合， $v_0 = v'_0$ が成り立つ．このとき， \mathbf{M}' および \mathbf{n} で表される 3 次元平面を，各視点に投影すると，参照視点 V_R^{rect} 上の点 m と近傍視点 C^{rect} 上の点 m' の間に次式の関係が成り立つ．

$$s\tilde{m}' = H\tilde{m} \quad (3.46)$$

ここで， H は射影変換行列を表し，次式で定義される [4]．

$$\mathbf{H} = \mathbf{A}' \left(\mathbf{R}_{\text{cam}} + \frac{\mathbf{t}_{\text{cam}} \mathbf{n}^T}{\mathbf{M}' \cdot \mathbf{n}} \right) \mathbf{A}^{-1} \quad (3.47)$$

$$= \begin{bmatrix} 1 + \frac{Bn_X}{\mathbf{M}' \cdot \mathbf{n}} & \frac{Bn_Y}{\mathbf{M}' \cdot \mathbf{n}} & d_u \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.48)$$

ここで， d_u は，

$$d_u = \frac{-u_0 B n_X - v_0 B n_Y + \beta B n_Z}{\mathbf{M}' \cdot \mathbf{n}} + u'_0 - u_0 \quad (3.49)$$

である．式 (3.48) が示すように，3 次元平面の平行ステレオペア間における画像変形はアフィン変換行列で表される．厳密には，ステレオ画像間の変形を表す行列 H は，水平座標方向の拡大縮小とスキュー，および，平行移動の組み合わせで構成される．

次に，平行ステレオペア $V_R^{\text{rect}}-C^{\text{rect}}$ において，エピポーラ線が垂直座標軸に平行である場合について考える．このとき， $V_R^{\text{rect}}-C^{\text{rect}}$ 間のカメラ座標系の回転 R_{cam} は水平ステレオの場合と変わらないが，並進ベクトル t_{cam} は次式で与えられる．

$$t_{\text{cam}} = \begin{bmatrix} 0 \\ B \\ 0 \end{bmatrix} \quad (3.50)$$

また，垂直ステレオの場合， $u_0 = u'_0$ が成り立つ．このとき，参照視点 V_R^{rect} と近傍視点 C^{rect} 間の画像変形を表す射影変換行列 H は次式で表される．

$$H = \begin{bmatrix} 1 & 0 & 0 \\ \frac{Bn_X}{M' \cdot n} & 1 + \frac{Bn_Y}{M' \cdot n} & d_v \\ 0 & 0 & 1 \end{bmatrix} \quad (3.51)$$

ここで， d_v は，

$$d_v = \frac{-u_0 B n_X - v_0 B n_Y + \beta B n_Z}{M' \cdot n} + v'_0 - v_0 \quad (3.52)$$

である．式 (3.51) に示す通り，垂直ステレオの場合，平行ステレオペア $V_R^{\text{rect}}-C^{\text{rect}}$ 間の画像変形は，垂直座標方向の拡大縮小とスキュー，および，平行移動となる．

3.4.2 マッチングウィンドウの幾何補正による画像変形の軽減

3.4.1 節で述べたように，復元対象の局所的な 3 次元平面を仮定すると，ステレオペア間の画像変形は，エピポーラ線方向の拡大縮小とスキュー，平行移動で表される．本節では，これらの画像変形を軽減するようなマッチングウィンドウを設定することで，MVS のための POC に基づくウィンドウマッチングによる奥行き推定の高精度化を行う [38], [39]．このとき，平行移動量 d_u （もしくは d_v ）はマッチングウィンドウの中心座標のずれであり，MVS のための POC に基づくウィンドウマッチングでは 3 次元候補点 M' によってマッチングウィンドウの中心座標が決定されるため， d_u を考慮する必要はない．そのため，マッチング

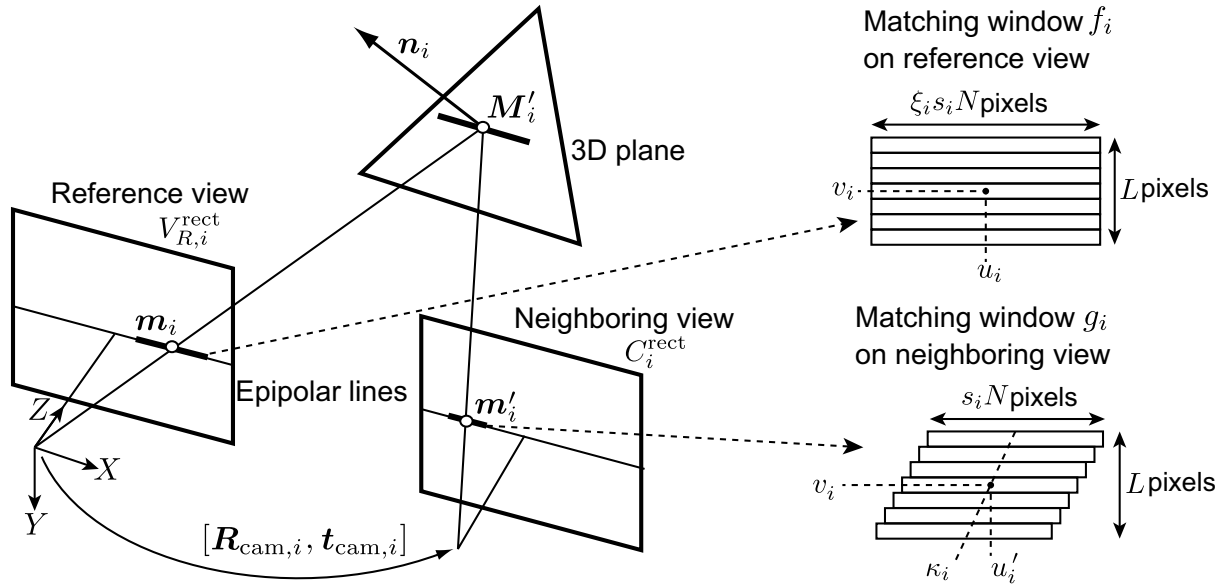


図 3.9 対称物体の局所的な 3 次元形状を考慮したマッチングウィンドウの幾何補正

ウィンドウ間の拡大縮小とスキューのみを軽減するように，マッチングウィンドウを設定すれば良い．以下では，参照視点 $V_R \in V$ とその近傍視点 $C = \{C_0, \dots, C_{K-1}\} \subseteq V - \{V_R\}$ について，3 次元候補点 M' ，および，法線ベクトル n が与えられたときに，マッチングウィンドウ間の幾何補正による画像変形の軽減を行いながら，MVS のための POC に基づくウィンドウマッチングを用いて真の 3 次元点 M を求める方法について説明する．また，ある平行ステレオペア $V_{R,i}^{\text{rect}} - C_i^{\text{rect}}$ について， $V_{R,i}^{\text{rect}} - C_i^{\text{rect}}$ が水平ステレオである場合を例に，マッチングウィンドウの幾何補正の様子を図 3.9 に示す．

まず，3.3 節で述べたように，MVS のための POC に基づくウィンドウマッチングでは，参照視点 V_R と近傍視点 C から平行ステレオペア $V_{R,i}^{\text{rect}} - C_i^{\text{rect}}$ ($i = 0, \dots, K - 1$) を生成する．このとき，平行化後の参照視点 $V_{R,i}^{\text{rect}}$ における，3 次元候補点 M'_i と法線ベクトル n_i は次式で与えられる．

$$M'_i = R_i^{-1} M' \tag{3.53}$$

$$n_i = R_i^{-1} n \tag{3.54}$$

ここで， R_i は，ステレオ平行化の際の参照視点におけるカメラ座標系の回転を表す．

次に，3 次元候補点 M'_i と法線ベクトル n_i により表現される 3 次元平面によるステレオ

画像間の画像変形を軽減するようにマッチングウィンドウを設定する．提案手法では，図 3.9 に示す通り，平行ステレオペア $V_{R,i}^{\text{rect}}-C_i^{\text{rect}}$ について，参照視点 $V_{R,i}^{\text{rect}}$ のマッチングウィンドウ f_i を ξ_i で拡大縮小し，近傍視点 C_i^{rect} におけるマッチングウィンドウ g_i を κ_i でスキューすることで，マッチングウィンドウ間の画像変形を軽減する．ここで，図 3.9 において， $\mathbf{m}_i = [u_i, v_i]^T$ は M' の $V_{R,i}^{\text{rect}}$ 上の対応点を， $\mathbf{m}'_i = [u'_i, v'_i]^T$ は M' の C_i^{rect} 上の対応点を， N は基準となるウィンドウサイズを， s_i は視差の正規化のための拡大縮小率を， L はマッチングウィンドウのライン数を表す．式 (3.48) より，水平ステレオの場合の拡大縮小率 ξ_i とスキューの傾き κ_i は次式で表される．

$$\xi_i = \left(1 + \frac{B_i n_{X,i}}{M'_i \cdot \mathbf{n}_i} \right)^{-1} \quad (3.55)$$

$$\kappa_i = \frac{B_i n_{Y,i}}{M'_i \cdot \mathbf{n}_i} \quad (3.56)$$

式 (3.51) より，垂直ステレオの場合の拡大縮小率 ξ_i とスキューの傾き κ_i は次式で表される．

$$\xi_i = \left(1 + \frac{B_i n_{Y,i}}{M'_i \cdot \mathbf{n}_i} \right)^{-1} \quad (3.57)$$

$$\kappa_i = \frac{B_i n_{X,i}}{M'_i \cdot \mathbf{n}_i} \quad (3.58)$$

$V_{R,i}^{\text{rect}}$ におけるウィンドウサイズを $\xi_i s_i N \times L$ ピクセル， C_i^{rect} におけるウィンドウサイズを $s_i N \times L$ ピクセルとすることで，ステレオ画像間の局所的な拡大縮小を軽減する．ここで，参照視点のウィンドウを ξ_i 倍するのではなく，近傍視点のウィンドウを $1/\xi_i$ 倍した場合，POC 関数 \hat{r}_i のピーク座標が $1/\xi_i$ 倍されるため，異なるステレオペア $V_{R,j}^{\text{rect}}-C_j^{\text{rect}}$ ($j \in \{0, \dots, K-1\} - \{i\}$) から計算される POC 関数 \hat{r}_j と \hat{r}_i のピーク座標が一致しないことに注意されたい．また，近傍視点におけるマッチングウィンドウを κ_i でスキューさせ，エピポーラ線と直行する座標軸のウィンドウ中心からのずれに応じて，ラインを切り出す中心座標を変化させることで，ステレオ画像間の局所的なスキューを軽減する．マッチングウィンドウ f_i および g_i を， $N \times L$ ピクセルの長方形に補正したマッチングウィンドウを \hat{f}_i, \hat{g}_i とする．このとき， \hat{f}_i および \hat{g}_i 間では，式 (3.48) (もしくは式 (3.51)) で表される 3 次元平面に基づく拡大縮小とスキューが軽減され，3 次元候補点 M' と真の 3 次元点 M の奥行きの変移に起因する平行移動 δ のみが残る．

最後に, \hat{f}_i および \hat{g}_i から 1 次元 POC 関数 \hat{r}_i を計算し, 平均化された POC 関数 \hat{r}_{ave} から真の 3 次元点 M を計算する. 3 次元点 M の計算方法は, 3.3.3 節で述べた通りである. また, このとき, POC 関数のピーク値 α は, はじめに仮定した 3 次元候補点 M' および法線ベクトル n の妥当性を示し, 推定された 3 次元点 M の信頼度として利用できる.

3.5 性能評価

本節では, 実験により, 提案手法を評価する. 本実験では, MVS アルゴリズムにおけるウィンドウマッチングのみの違いによる, 3 次元点の奥行き推定精度や計算コストの違いを調べる. MVS 評価用のデータセットとして, Middlebury のデータセット [5], [42], Strecha らのデータセット [6], [49], Jensen らのデータセット [41], [43] が有名である. 一方で, これらのデータセットは, MVS アルゴリズム全体の評価を目的としており, ウィンドウマッチングによる奥行き推定のみの比較を行うことが難しい. 奥行き推定のみの比較のためには, 各視点における真値の奥行きが必要である. 上記データセットのうち, Strecha らのデータセット Fountain-P11 は, 真値の 3 次元メッシュモデルが提供されているため, 真値の奥行き情報を算出できる. しかし, それ以外のデータセットでは, 真値の 3 次元メッシュモデルが提供されておらず, 本実験には使用できない. そこで, 本実験では, Fountain-P11 の他に, MVS 評価用の新しい多視点画像データセットを作成し, 実験に使用する. また, 実験で比較する従来手法として, 従来の MVS アルゴリズムで用いられるウィンドウマッチング手法のうち, NCC に基づくウィンドウマッチングに着目し, サブピクセル推定・幾何補正の有無などで分類し, 使用する. 本節では, 作成したデータセット, および, データセット Fountain-P11 を用いて, これらの従来手法との比較実験を行い, 提案手法の有効性を示す.

以下では, まず, 実験に用いる多視点画像データセットについて説明し, 比較する従来手法について述べる. 次に, 各ウィンドウマッチングによる奥行き推定精度を評価する. そして, 各ウィンドウマッチング手法の計算コストを評価する.

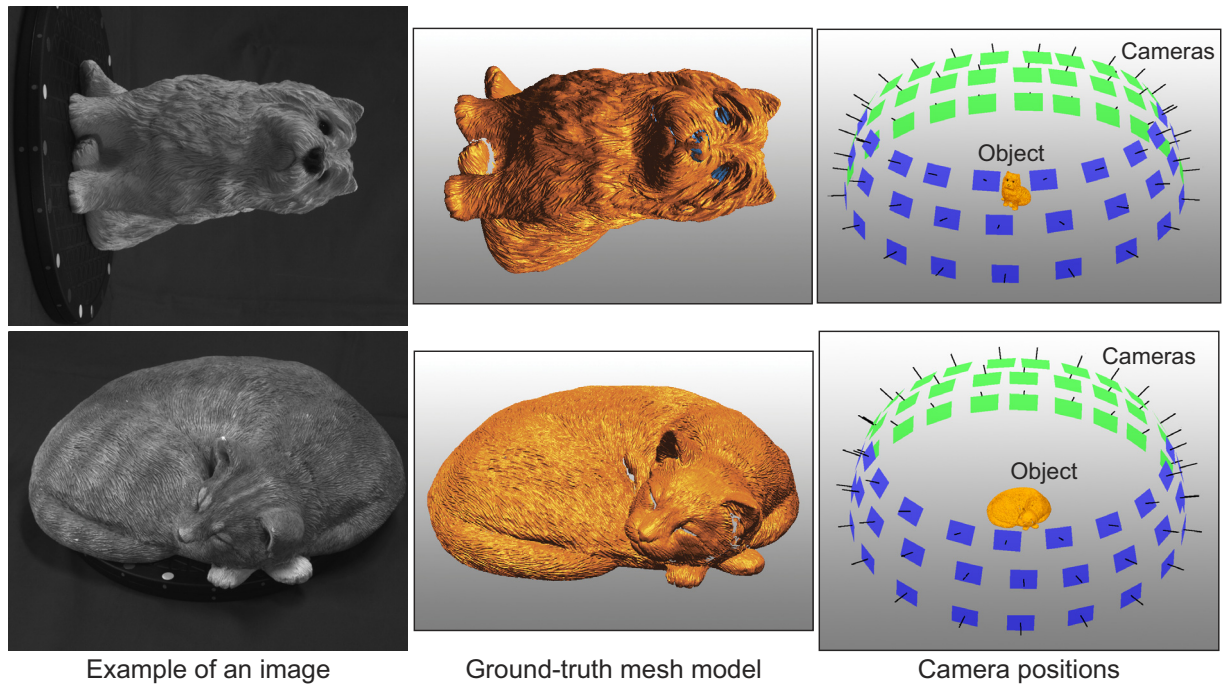


図 3.10 多視点画像データセット (上段: データセット dog , 下段: データセット cat)

3.5.1 実験用データセット

本論文では，ウィンドウマッチング手法の違いによる 3 次元点の奥行き精度を比較するため，真値の 3 次元モデルを伴う多視点画像データセットを作成し，実験に使用する．図 3.10 に作成したデータセットの画像例，真値のメッシュモデル，視点位置を示す．以下，データセットの作成方法について詳しく説明する．

対象物体は陶器製の犬の置物と猫の置物である．物体表面は，光沢が少ない拡散反射面であり，細かいテクスチャで塗られている．これらの対象物体を回転テーブルに載せ，カメラの高さを 3 パターン，回転テーブルの角度を 20 パターンで変えながら 60 枚の画像を撮影する．カメラは，PointGrey 社製のデジタルカメラ Flea3 (FL3-U3-13Y3M-C) を，レンズは，ミュートロン社製のメガピクセル低ディストーションレンズ (焦点距離 10 mm , FV1022) を用いる．撮影される画像は， $1,280 \times 1,024$ ピクセルの白黒画像である．対象物

体とカメラとの距離は、およそ 700 mm である。多視点画像 V について、カメラの高さ方向を表すインデックス i と、カメラの角度方向を表すインデックス j を用いて、各視点を $V_{i,j}$ ($i \in \{0, 1, 2\}, j \in \{0, 1, \dots, 19\}$) と表す。カメラの高さを表すインデックス i が小さいほど、低い位置から対象を撮影した視点であるとする。また、回転テーブルは $j = 0$ から $j = 19$ で約 1 周するため、視点 $V_{i,0}$ と $V_{i,19}$ は隣接する視点となる。真値のメッシュモデルは、Steinbichler 社製の 3 次元デジタイザ COMET5 を使用して計測した。COMET5 は、奥行き計測精度が 0.025 mm となるように設定した。

撮影した多視点画像と真値のメッシュモデルを用いて、各視点のカメラパラメータを推定する。本章で行う精度評価実験では、多視点画像の各視点における、対象物体の真値の奥行きが必要である。そのため、多視点画像と真値のメッシュモデルとで整合性が取れるように、カメラパラメータを推定する必要がある。本論文で作成するデータセットでは、Scale-Invariant Feature Transform (SIFT) に基づく対応付けと Structure from Motion (SfM) を組み合わせ、各視点のカメラパラメータを得る。このとき、真値のメッシュモデルを用いて再投影誤差を計算することで、カメラ画像とメッシュモデルの整合性を図る。視点 $V_{i,j}$ の内部パラメータを $A_{i,j}$ 、外部パラメータを $R_{i,j}, t_{i,j}$ (回転行列と並進ベクトル) としたとき、各視点のカメラパラメータを推定する手順を以下に記す。

Step1: カメラの内部パラメータを事前に Zhang らの手法により求める [64], [65]。カメラのキャリブレーション用の画像を撮影する際と、3 次元復元用の多視点画像撮影の際で、焦点距離や絞りなどを固定し、すべての視点の内部パラメータを同一の内部パラメータ A であるとする。

$$A_{i,j} = A \quad (i \in \{0, 1, 2\}, j \in \{0, 1, \dots, 19\}) \quad (3.59)$$

Step2: SIFT に基づく対応付け手法 [50] により、すべての視点ペアにおける対応点ペアを得る。

Step3: まず、SIFT に基づく対応付け結果のみを用いて、5 点アルゴリズム [66]、および、Kneip らの手法 [67] により、各視点の外部パラメータ $R_{i,j}, t_{i,j}$ を求める。求めた $R_{i,j}, t_{i,j}$ を初期値として、バンドル調整により、各視点の外部パラメータの最適化を

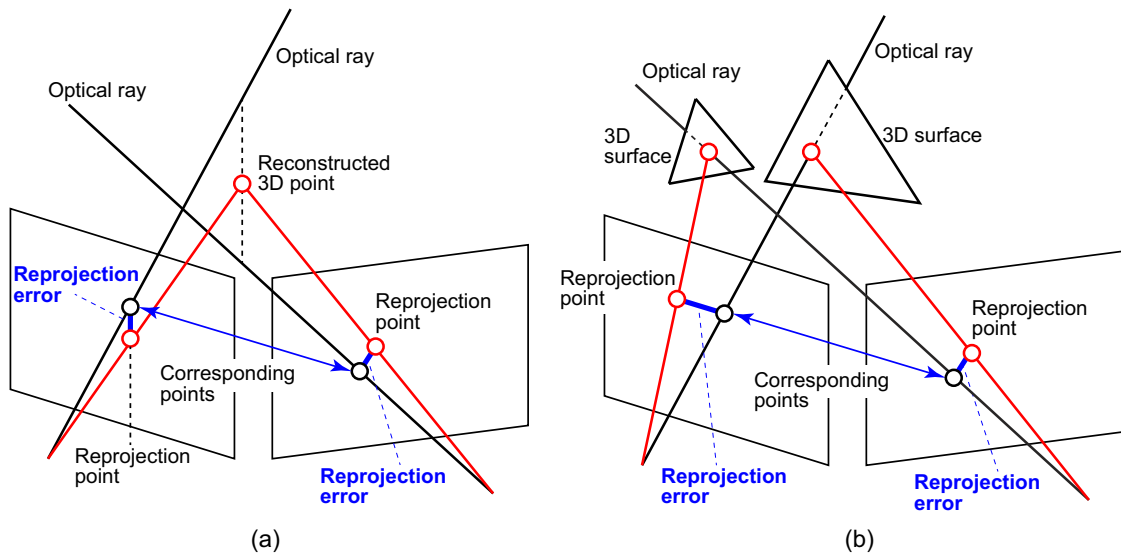


図 3.11 再投影誤差: (a) 対応点から計算した 3 次元復元点の再投影誤差, (b) 真値の 3 次元メッシュモデルから計算した再投影誤差

行う [52]–[54]。バンドル調整では、各視点の外部パラメータ $R_{i,j}, t_{i,j}$ を最適化のパラメータとして、全対応点ペアの再投影誤差 (図 3.11(a)) の 2 乗和を最小化する。また、同時に、疎な 3 次元復元結果を得る。

Step4: ICP アルゴリズム [53], [85] を用いて、疎な 3 次元復元結果と真値のメッシュモデルを位置合わせする。Step3 で得た 3 次元点では、スケールの不定性があるため、物体の絶対的な大きさがわからない。そのため、ICP アルゴリズムでは、疎な 3 次元復元結果と真値のメッシュモデル間の回転と平行移動のみだけでなく、スケールも推定する。推定した回転・平行移動・スケールを用いて、各視点の外部パラメータ $R_{i,j}, t_{i,j}$ を更新する。

Step5: 再度、バンドル調整により、各視点の外部パラメータの最適化を行う。このとき、真値のメッシュモデルを用いて再投影誤差を計算することで、多視点画像とメッシュモデルの整合性を取る (図 3.11(b))。最適化後の外部パラメータを、各視点の外部パラメータ $R_{i,j}, t_{i,j}$ とする。

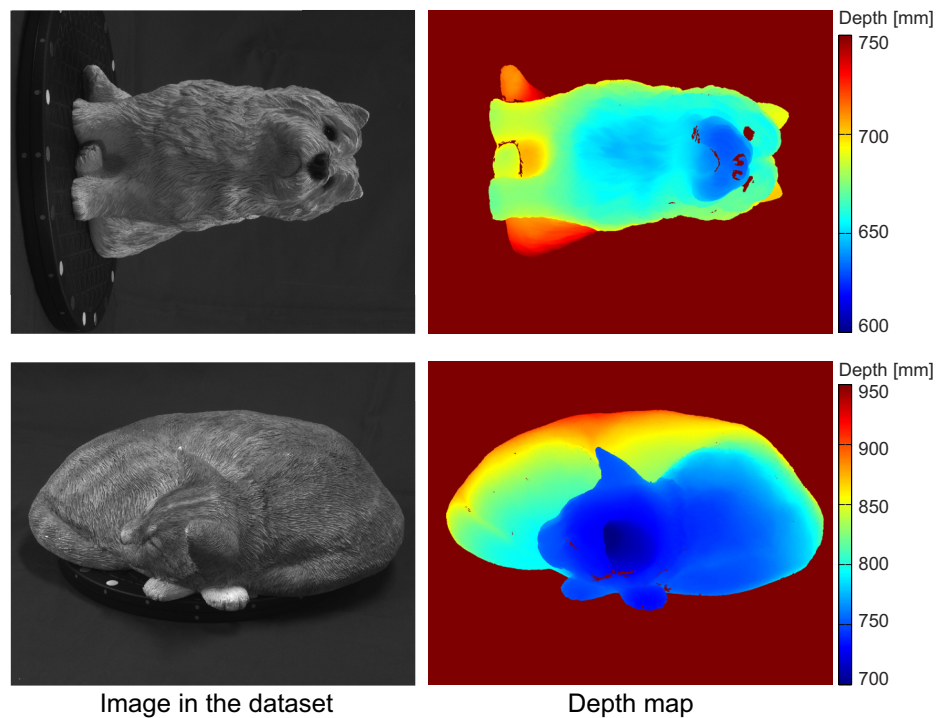


図 3.12 データセット中の画像とその視点に対応するデプスマップ (上段: データセット dog, 下段: データセット cat)

以上の Step1 から Step5 により, データセット dog および cat のカメラパラメータを推定した. このように, 作成したデータセット dog および cat では, 真値のメッシュモデルから, 各視点の真値のデプスマップ (奥行き情報) を作成することが可能である. 推定したカメラパラメータを用いて真値のメッシュモデルから生成したデプスマップの例を図 3.12 に示す. 図 3.12 より, 各視点について, その視点からの真値のメッシュモデルまでの奥行きが正しく求められており, 多視点画像と真値のメッシュモデルの整合性が取れていることが確認できる. また, このとき, SIFT に基づく対応点ペアについて, メッシュモデルを用いた再投影誤差の Root Mean Square (RMS) は, dog および cat どちらのデータセットについても 0.4 ピクセル以下であり, 画像上でサブピクセルの精度で位置合わせができていることが確認できる.

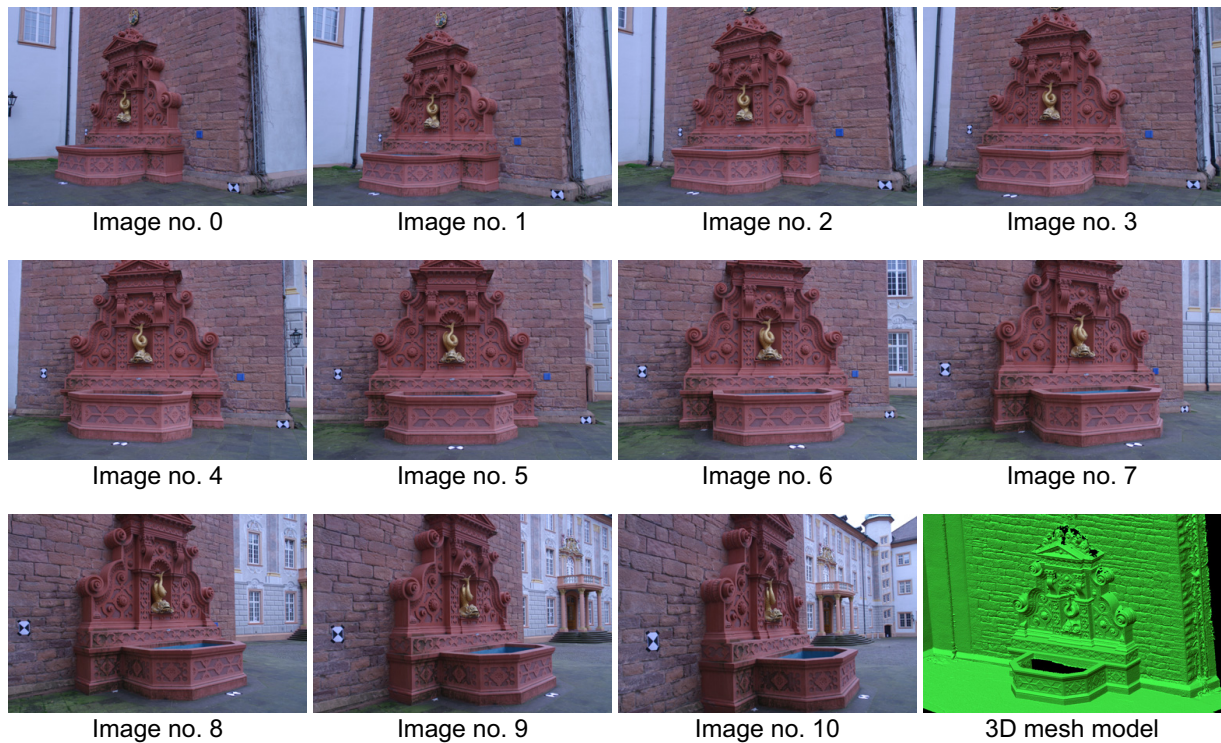


図 3.13 データセット Fountain-P11 における多視点画像と真値のメッシュモデル

また、本章では、データセット dog および cat と同様に、真値の 3 次元データを伴うデータセットとして、Strecha らによる公開データセット Fountain-P11 [6], [49] も実験に用いる。Fountain-P11 には、11 枚の多視点画像とそのカメラパラメータ、および、真値のメッシュモデルが提供されている^{*7}。多視点画像 V について、視点番号を表すインデックス i を用いて、各視点を V_i ($i \in 0, 1, \dots, 10$) と表す。データセット Fountain-P11 の各画像と真値のメッシュモデルを図 3.13 に示す。各視点のカメラパラメータは、多視点画像と真値のメッシュモデルで整合性が取れるように求めたものであるため、各視点について、真値のメッシュモデルから真値のデプスマップを作成することができる。また、データセット Fountain-P11 に含まれる原画像は、 $3,072 \times 2,048$ ピクセル (full) のカラー画像であるが、その他に、 $1,536 \times 1,024$ ピクセル (half)、および、 768×512 ピクセル (quarter) に縮小した画像について比較を行う。

^{*7} [49] で提供されているカメラパラメータについて、その外部パラメータに誤差が含まれていることを実験的に確認した。そこで、本論文では、データセット Fountain-P11 についても、データセット dog および cat と同様の方法で外部パラメータを最適化し、実験に利用する。

3.5.2 比較手法

本章の性能評価実験では, MVS アルゴリズム全体でなく, MVS におけるウィンドウマッチングのみの精度や計算コストを比較する. 比較するウィンドウマッチング手法について, マッチングスコア算出, サブピクセル推定, ウィンドウの幾何補正の有無について分類し, 表 3.1 にまとめる. NCC に基づくウィンドウマッチングでは, サブピクセル推定を総当たりで行う手法 (BF), 関数フィッティングで行う手法 (FF), 非線形最適化で行う手法 (LM) がある. この中で BF と LM は, 画像補間に基づく手法である. また, POC に基づくウィンドウマッチングでは, サブピクセル推定が関数フィッティングに基づく手法である. ウィンドウマッチングの幾何補正は, NCC を用いる場合には射影変換に, POC を用いる場合にはアフィン変換に従う*⁸. 本節の実験では, これらの 8 つの手法を, 単純なプレーンスライピングに適用し, それぞれのウィンドウマッチングによる 3 次元点の奥行き推定精度と計算コストを比較する. 以下では, 各手法の実装について具体的に述べる.

NCC+BF (+Homography):

多視点画像間のウィンドウマッチングに NCC を用いる. 参照視点 V_R の視線上のある 3 次元点の奥行きについて, 参照視点 V_R 上の投影点を中心に設定されたマッチングウィンドウと, 近傍視点 C 上の投影点を中心に設定されたマッチングウィンドウで NCC を計算し, その奥行きの尤度とする. このとき, 複数のステレオペアで計算された NCC の値を平均化することで, 精度を向上させる [20]. NCC の平均化の際には, 計算された NCC の値が閾値 th_{corr} を超えたもののみを平均化することで, オクルージョンなどに対処する. BF では, 3 次元点の奥行き Z を刻み幅 ΔZ で変化させながらマッチングスコアを計算し, 最もマッチングスコアが高くなる奥行きを真の奥行きとして選択する. このとき, サブピクセル位置におけるマッチングウィンドウは画像補間

*⁸ NCC に基づく手法と POC に基づく手法の幾何補正における, 射影変換とアフィン変換の違いは, ステレオ平行化の有無に起因する. NCC に基づく手法においてもステレオ平行化を行う場合, そのマッチングウィンドウの幾何補正は, アフィン変換に従う.

表 3.1 多視点ステレオにおけるウィンドウマッチングの分類

Matching	Sub-pixel estimation	Geometric correction	Reference
NCC	Brute force (BF)	—	[20], [24]
		Homography (H)	[23], [27], [72]
	Function fitting (FF)	—	
		Homography (H)	
	Nonlinear optimization (LM)	—	
		Homography (H)	[19]
POC	Function fitting (FF)	—	Proposed method
		Affine (A)	

により生成する．本実験では，多くの MVS アルゴリズムで用いられる線形補間を適用する．ウィンドウマッチングによる奥行き探索のパラメータは以下の通りである．マッチングウィンドウサイズは 17×17 ピクセルとする．このウィンドウサイズは，マッチングウィンドウ内の有効な信号長が提案手法と同等になるように設定した．奥行き変化の刻み幅 ΔZ は，近傍視点群の中で最も基線長が長いステレオ画像上において $1/10$ ピクセルとなるように設定する．NCC を平均化する際の閾値 th_{corr} は 0.5 とする．マッチングウィンドウの幾何補正を行う場合，マッチングウィンドウ間の画像変形は射影変換に従う．3次元点の奥行きの他に，法線ベクトルの向きを仮定し，3次元座標と法線ベクトルから射影変換行列を推定する．本実験では，参照視点 V_R の画像平面に正対する法線と， X 軸と Y 軸方向に 45 度ずつ傾けた合計 9 つの法線について試行し，最もマッチングスコアが高くなる法線を選択する．

NCC+FF (+Homography):

多視点画像間のウィンドウマッチングに NCC を用いる．サブピクセル精度の奥行き Z は，ピクセルレベルのマッチング結果に対してピークモデル関数をフィッティングすることで求める．ピクセルレベルのマッチング結果は，NCC+BF において $\Delta Z = 1$

ピクセルのときの結果を用い、マッチングスコアが最大となる奥行きとその前後のマッチングスコアに対して関数フィッティングを行う。本実験では、フィッティングするピークモデル関数としてパラボラ関数を用いる [28]。マッチングウィンドウサイズは 17×17 ピクセルとする。

マッチングウィンドウの幾何補正を行う場合、マッチングウィンドウ間の画像変形は射影変換に従う。本実験では、NCC+BF ($\Delta Z = 1$) においてマッチングスコアが最大となる法線を使用し、射影変換行列を求める。

NCC+LM (+Homography):

多視点画像間のウィンドウマッチングに NCC を用いる。サブピクセル精度の奥行き Z は、マッチングスコアが最大となる奥行き付近で、奥行き Z を最適化パラメータとして非線形最適化によりマッチングスコアの最大化を行うことで推定する。本実験では、非線形最適化手法として LM 法 [71] を用いる。また、最適化の際の奥行き Z の初期値は、NCC+BF における $\Delta Z = 1$ ピクセルのときの結果について、マッチングスコアが最大となる奥行きとする。マッチングウィンドウサイズは 17×17 ピクセルとする。また、サブピクセルレベルの奥行きに対応するマッチングウィンドウは画像補間により生成する。本実験では、BF と同様に、線形補間を用いる。

マッチングウィンドウの幾何補正を行う場合、マッチングウィンドウ間の画像変形は射影変換に従う。本実験では、NCC+BF ($\Delta Z = 1$) においてマッチングスコアが最大となる法線を使用し、射影変換行列を求める。

POC (+Affine):

多視点画像間のウィンドウマッチングに POC に基づく手法を用いる。ある 3 次元候補点 M' の奥行き Z_{init} を与えたとき、奥行き Z_{init} から真の 3 次元点 M の奥行き Z を推定する手順は、3.3.3 節で述べた通りである。本節の実験では、3 次元候補点の奥行き Z_{init} を刻み幅 ΔZ で変化させながら POC に基づくウィンドウマッチングを繰り返し、POC 関数のピーク値 α が最も高くなるときに推定された Z を真の奥行きとして選択する。3.2.2 節で述べた通り、POC では、信号長 N に対して $\pm N/4$ の平行移動量を推定することが可能である。そこで、本実験では、奥行き変化の刻み幅 ΔZ

を，近傍視点群の中で最も基線長が長いステレオ画像上で $N/4$ ピクセルの刻み幅になるように設定する．POC に基づくウィンドウマッチングのパラメータは以下の通りである．ウィンドウサイズは $N \times L = 32 \times 17$ ピクセルとする．ハニング窓の影響を考慮すると，このウィンドウサイズにおける有効な信号長は，NCC に基づくマッチングの有効なウィンドウサイズと同等である．複数のステレオペアから計算された POC 関数を平均化する際の閾値 th_{corr} は 0.5 とする．

マッチングウィンドウの幾何補正を行う場合，マッチングウィンドウ間の画像変形はアフィン変換に従う．3次元候補点の奥行き Z_{init} の他に，法線ベクトル n を仮定し，真の3次元点の奥行き Z を推定する手順は，3.4.2 節で述べた通りである． Z_{init} と n を変化させながら POC に基づくウィンドウマッチングを繰り返し，POC 関数のピーク値 α が最も高くなるときの n ，および，推定された Z を選択する．法線ベクトル n について，本実験では，参照視点 V_R の画像平面に正対する法線と， X 軸と Y 軸方向に 45 度ずつ傾けた合計 9 つの法線について試行する．

3.5.3 奥行き推定精度の評価

本小節では，MVS におけるウィンドウマッチング手法による 3次元点の奥行き推定精度を比較する．データセット dog および cat では，60 枚の多視点画像の中から 21 枚の画像を選択し参照視点 V_R とする．

$$V_R \in \{V_{i,j} | \{i,j\} \in \{\{0,0\}, \{0,3\}, \{0,6\}, \{0,9\}, \{0,12\}, \{0,15\}, \{0,18\}, \\ \{1,1\}, \{1,4\}, \{1,7\}, \{1,10\}, \{1,13\}, \{1,16\}, \{1,19\}, \\ \{2,2\}, \{2,5\}, \{2,8\}, \{2,11\}, \{2,14\}, \{2,17\}, \{2,19\}\}\} \quad (3.60)$$

近傍視点群 $C = V - \{V_R\}$ は, $V_R = V_{i,j}$ について 3-4 枚の隣接視点を選択する .

$$C = \begin{cases} \{V_{h,l} | \{h,l\} \in \{\{i+1,j\}, \{i, \text{mod}(j+1, 20)\}, \{i, \text{mod}(j-1, 20)\}\}\} \\ \quad \text{(if } i = 0\text{)} \\ \{V_{h,l} | \{h,l\} \in \{\{i+1,j\}, \{i-1,j\}, \{i, \text{mod}(j+1, 20)\}, \{i, \text{mod}(j-1, 20)\}\}\} \\ \quad \text{(if } i = 1\text{)} \\ \{V_{h,l} | \{h,l\} \in \{\{i-1,j\}, \{i, \text{mod}(j+1, 20)\}, \{i, \text{mod}(j-1, 20)\}\}\} \\ \quad \text{(if } i = 2\text{)} \end{cases} \quad (3.61)$$

ここで, $\text{mod}(A, B)$ は, B を法とする A の最小非負剰余を表す . 一方, データセット Fountain-P11 では, 11 枚の多視点画像の中から 9 枚の画像を選択し参照視点 V_R とする .

$$V_R \in \{V_i | i \in \{1, 2, \dots, 9\}\} \quad (3.62)$$

近傍視点群 $C = V - \{V_R\}$ は, $V_R = V_i$ について 2 枚の隣接視点を選択する .

$$C = \{V_l | l \in \{i-1, i+1\}\} \quad (3.63)$$

いずれのデータセットにおいても, データセットに含まれるすべての画像が, 参照視点, もしくは, 近傍視点の少なくともどちらかにおいて 1 度は利用されるように, 参照視点・近傍視点群を選択した . それぞれのデータセットについて, 実験に用いた参照視点 V_R と近傍視点群 C の例を図 3.14 に示す .

それぞれのデータセットについて, すべての V_R におけるすべてのデジタル画像座標について 3 次元点の奥行きを求め, 真値のメッシュモデルから計算される奥行きと比較する . 各手法が算出した奥行き $Z_{\text{calculated}}$ と真値の奥行き $Z_{\text{ground-truth}}$ の誤差は, 次式で定義される正規化誤差 e を用いて評価する .

$$e = \frac{Z_{\text{calculated}} - Z_{\text{ground-truth}}}{Z_{\text{ground-truth}}} \quad (3.64)$$

データセット Fountain-P11 については, それぞれの解像度の画像において奥行きの探索を行う点数を同じにするため, full の場合は 4 ピクセル間隔の格子状に配置した座標について, half の場合は 2 ピクセル間隔の格子状に配置した座標について, quarter の場合はすべてのピクセルについて奥行き探索を行う . また, 背景のように, V_R において真値のメッシュモデルが存在しない領域については, 奥行き推定の結果を無視し, 誤差の評価を行わない .

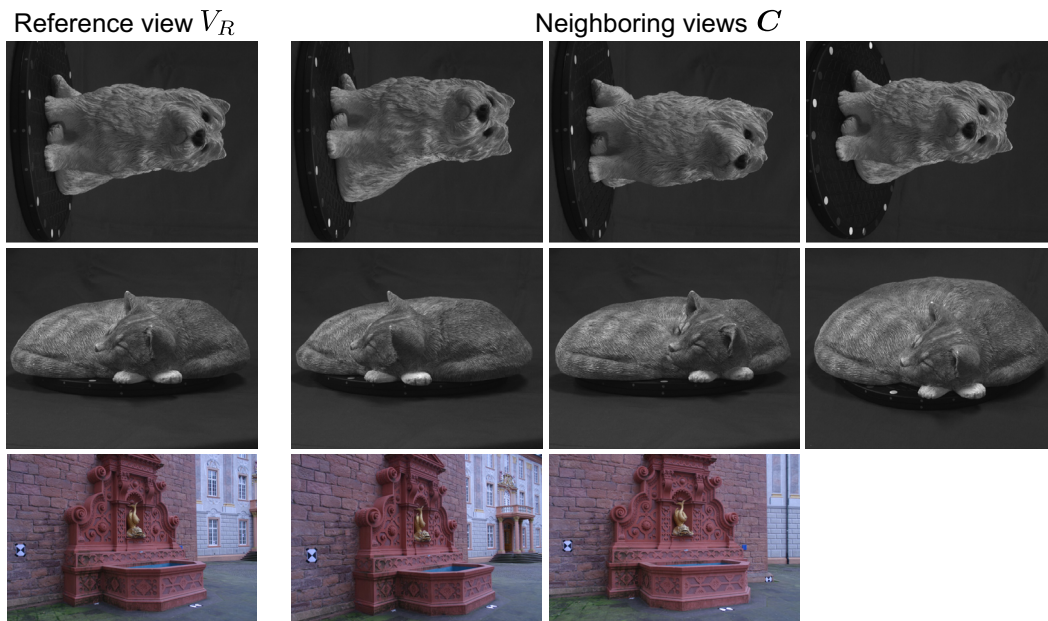


図 3.14 参照視点 V_R と近傍視点群 C の例 (1 段目: データセット dog, 2 段目: データセット cat, 3 段目: データセット Fountain-P11)

(I) データセット dog, cat の実験結果

図 3.15 に, それぞれのデータセットにおける誤差ヒストグラムを示す. ここで, ヒストグラムは 0.01% のビンで作成した. 図 3.15 上段は, 誤差とその誤差を持つ点の数を表すヒストグラムであり, 図 3.15 下段は, 3 次元復元点のうちその誤差を持つ点の割合を表すヒストグラムである. 図 3.15 のヒストグラムより, 各手法で, ヒストグラムが最大となる誤差はほぼ変わらず, 分布の広がり異なることが確認できる. そこで, 本実験では, 3 次元復元精度を表す定量的な指標として, 正規化誤差 e の絶対値の中央値を用いる. それぞれのデータセットにおける, 総復元点数と誤差の絶対値の中央値を表 3.2 にまとめる. 各手法の結果について誤対応点除去は適用していないが, すべてのステレオペアにおけるマッチングスコアが閾値を下回った場合, その点が 3 次元復元されないため, 各手法の総復元点数は異なる. そのため, 点数のヒストグラムと割合のヒストグラムは異なった傾向を示す. また, 図 3.16, および, 図 3.17 に, データセット dog の 3 次元復元結果として, 画素値で色をつけた 3 次元点群と, 誤差の絶対値で色をつけた 3 次元点群を示す. 同様に, 図 3.18, および, 図 3.19 に,

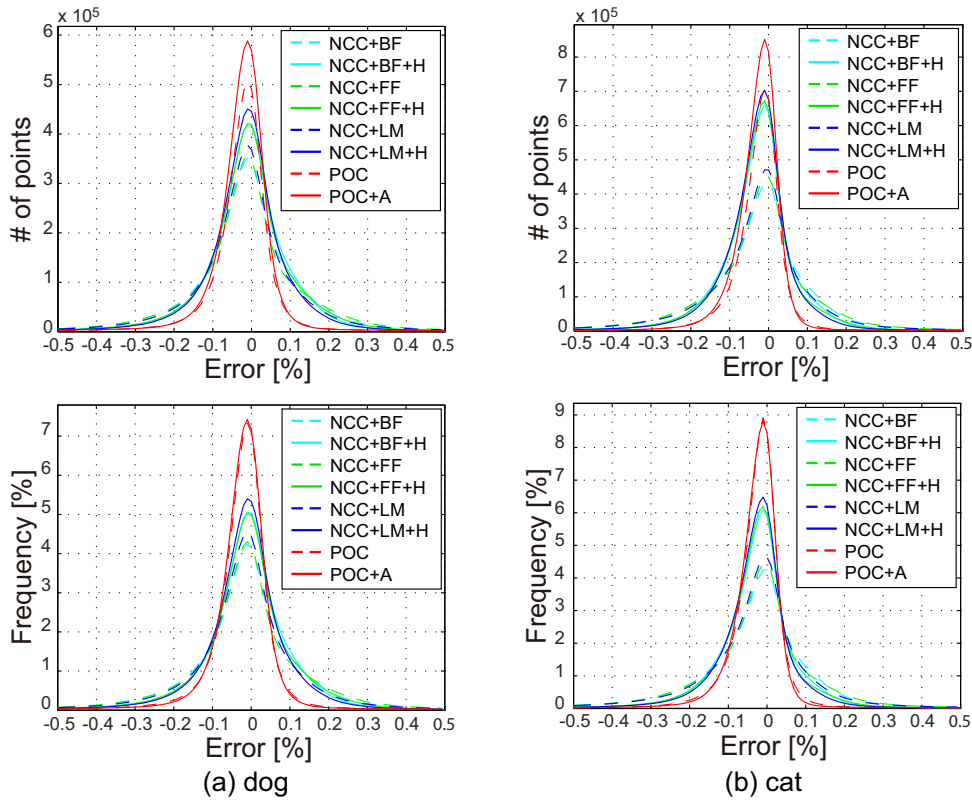


図 3.15 データセット dog, cat における誤差ヒストグラム (上段: 点数のヒストグラム, 下段: 点の割合のヒストグラム)

データセット cat の 3 次元復元結果として, 画素値で色をつけた 3 次元点群と, 誤差の絶対値で色をつけた 3 次元点群を示す. ここで, 各 3 次元点群は, 図の視認性のため, 誤差の絶対値が非常に大きい誤対応点を手動で除去して表示した.

まず, NCC に基づくウィンドウマッチングと POC に基づくウィンドウマッチングの比較を行う. NCC に基づく手法では, 幾何補正の有無に関わらず, 誤差ヒストグラム, 3 次元復元点数, 誤差の絶対値の中央値, 3 次元復元点群のいずれにおいても, サブピクセル推定手法による大きな違いはなく, 同様の傾向を示している. 誤差の絶対値の中央値から, BF, FF, LM の中では LM の精度が高いが, その差は小さい. 以下では, サブピクセル推定手法 (BF, FF, LM) についての区別はせず, NCC に基づくウィンドウマッチングとして扱う. 図 3.15 の誤差ヒストグラムより, 幾何補正の有無のそれぞれについて, POC に基づく手法が, NCC

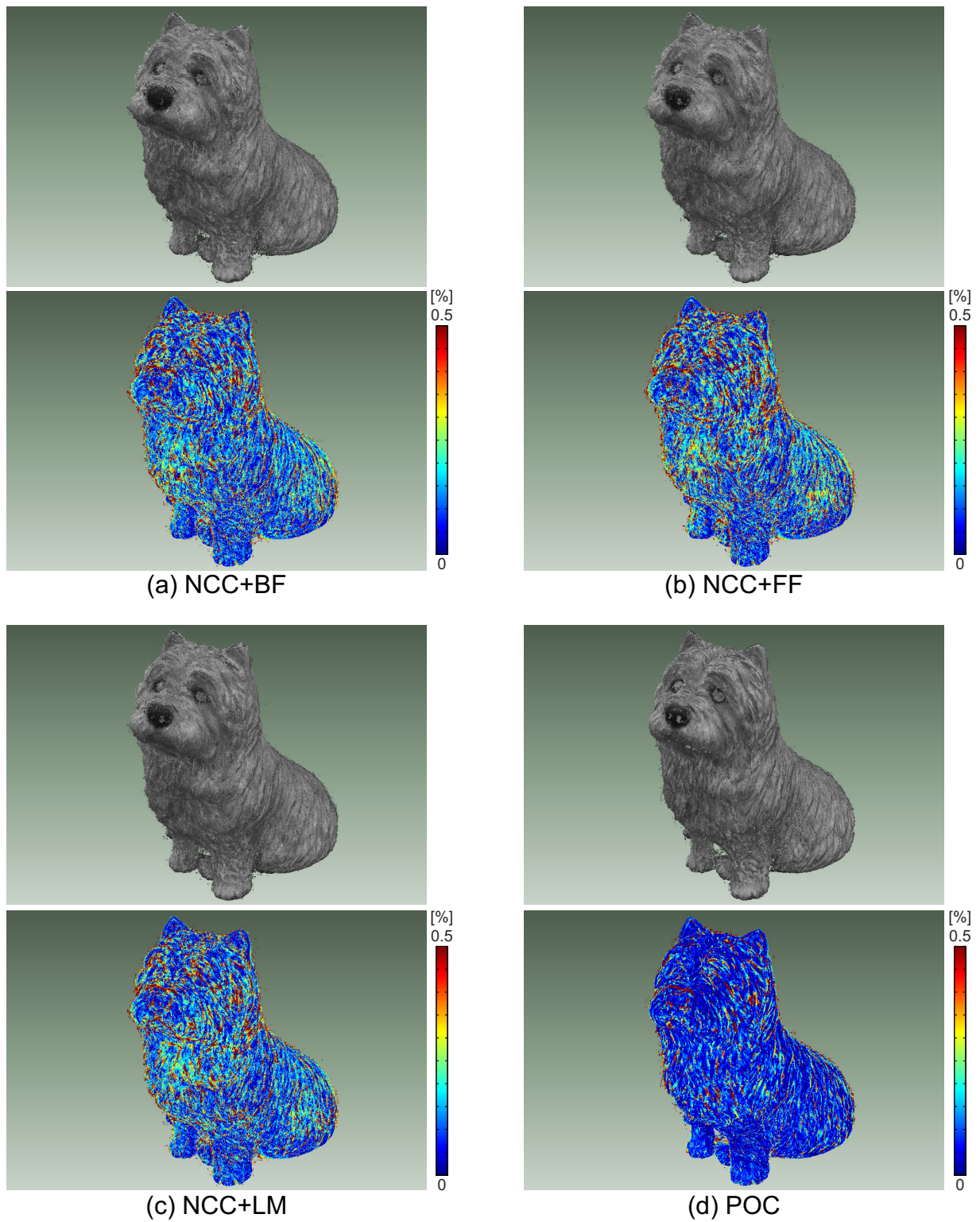


図 3.16 データセット dog における，幾何補正を行わない場合の 3 次元復元点群と誤差（それぞれの手法について，上段：3 次元復元点群，下段：3 次元復元誤差の絶対値）

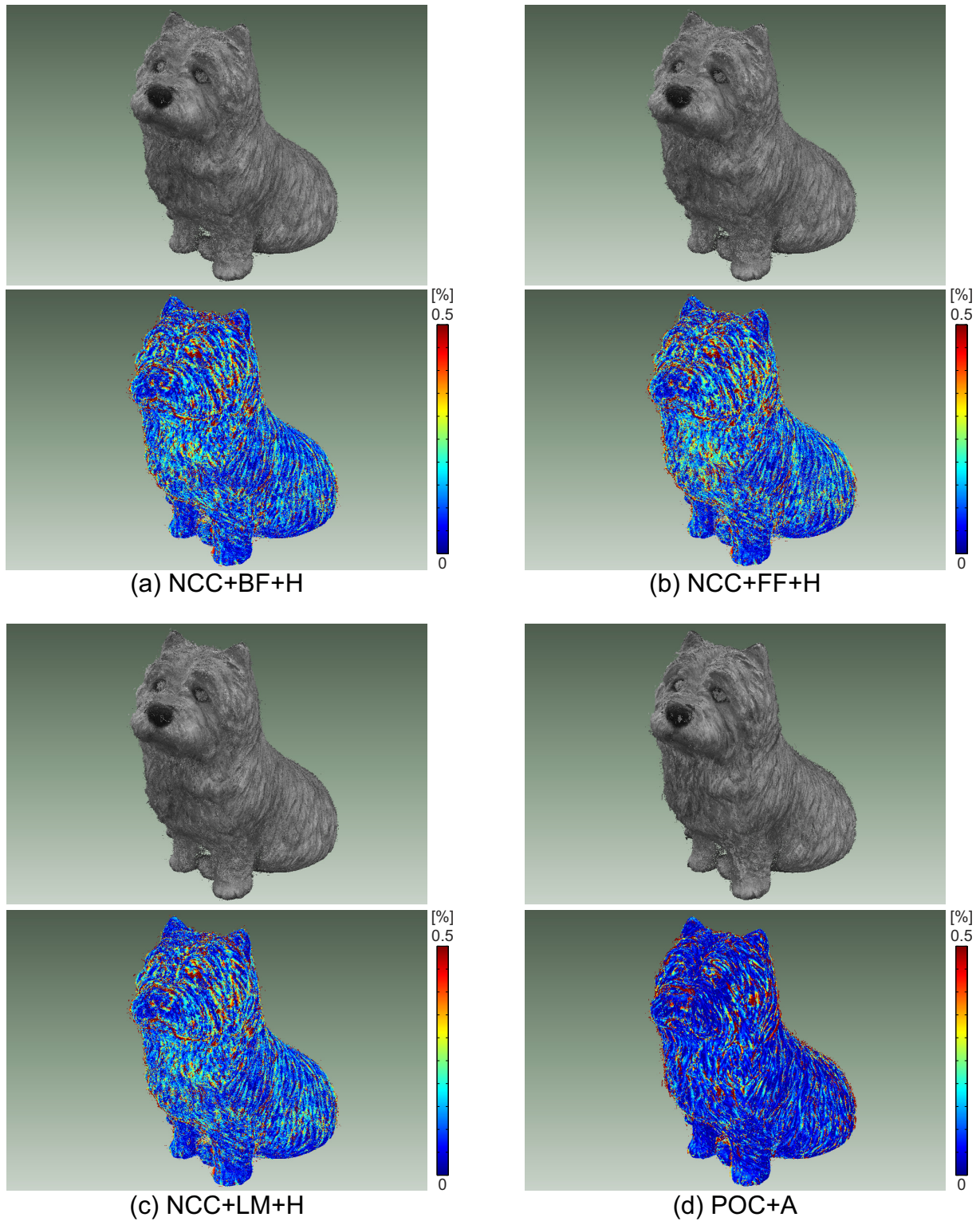


図 3.17 データセット dog における，幾何補正を行う場合の 3 次元復元点群と誤差（それぞれの手法について，上段: 3 次元復元点群，下段: 3 次元復元誤差の絶対値）

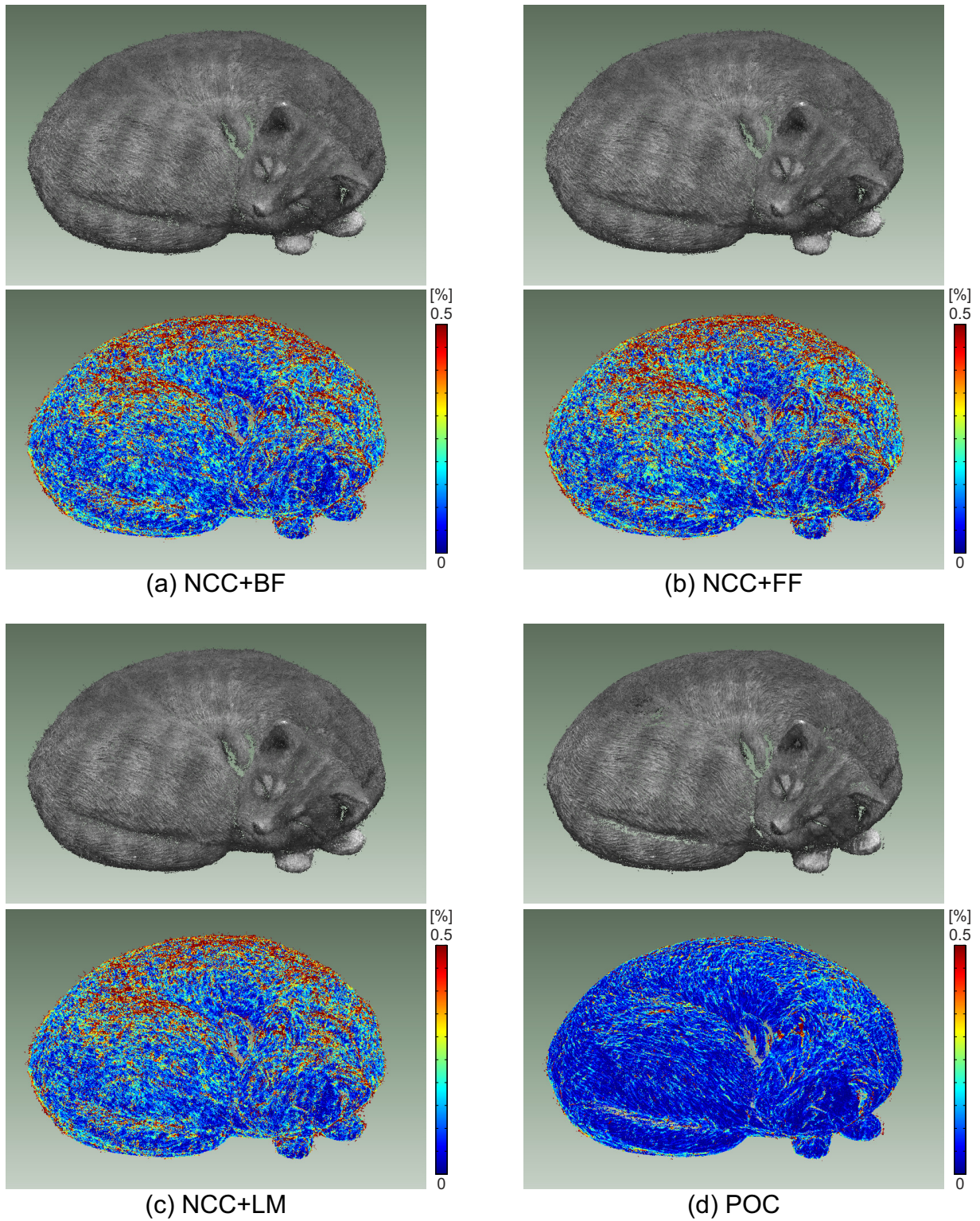


図 3.18 データセット cat における，幾何補正を行わない場合の 3 次元復元点群と誤差（それぞれの手法について，上段：3 次元復元点群，下段：3 次元復元誤差の絶対値）

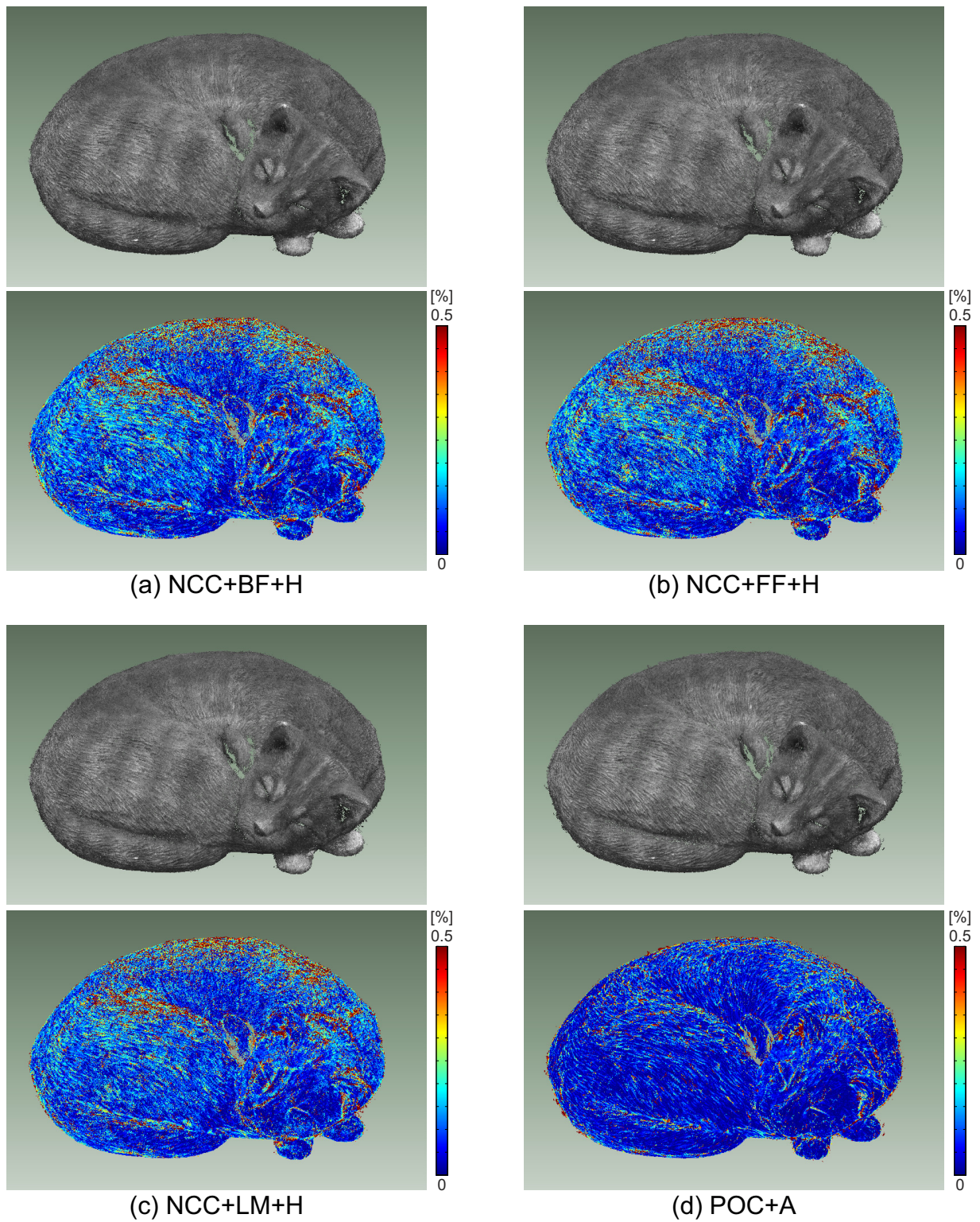


図 3.19 データセット cat における，幾何補正を行う場合の 3 次元復元点群と誤差（それぞれの手法について，上段: 3 次元復元点群，下段: 3 次元復元誤差の絶対値）

表 3.2 データセット dog, cat における 3 次元復元点数と誤差の絶対値の中央値

	# of points [10^6]		median of absolute values of error [%]	
	dog	cat	dog	cat
NCC+BF	7.51	9.30	0.0688	0.0699
NCC+FF	7.46	9.25	0.0684	0.0673
NCC+LM	7.46	9.25	0.0626	0.0629
POC	6.15	7.48	0.0351	0.0310
NCC+BF+H	7.91	10.44	0.0564	0.0496
NCC+FF+H	7.88	10.43	0.0569	0.0478
NCC+LM+H	7.88	10.43	0.0519	0.0454
POC+A	7.21	9.06	0.0348	0.0307

に基づく手法と比べ、誤差の絶対値が小さい点数が多く、誤差の絶対値が大きい点数が少ないことから、奥行き推定精度が高い。表 3.2 の誤差の絶対値の中央値についても、同様の結果が現れており、NCC に基づく手法に比べ、POC に基づく手法の誤差は約 $2/3$ から $1/2$ となっている。また、図 3.16 から図 3.19 の 3 次元復元点群についても、NCC に基づく手法の点群に比べ、POC に基づく手法の点群の方が、誤差の絶対値が大きい赤い点が少なく、誤差の絶対値が小さい青い点が多いことが確認できる。これらの誤差の差は、ステレオ画像上においてサブピクセル相当の誤差の差である。つまり、NCC に基づく手法と POC に基づく手法の奥行き推定精度の差は、主に、サブピクセルレベルのマッチングをどのように実現するかに起因する。NCC+BF (+H) および NCC+LM (+H) では、サブピクセルの画素値を線形関数で近似する。NCC+FF (+H) では、サブピクセルのマッチングスコアをパラボラ関数で表現する。一方で、POC (+A) では、POC 関数の相関ピークモデルの関数フィッティングにより、離散データの標本化間隔の間にあるピーク座標を求め、3 次元点の奥行きを推定する。解析的に求めたピークモデルによりサブピクセルレベルのマッチングを実現し、3 次元点の奥行き推定に利用することで、POC に基づくウィンドウマッチングは、NCC に基づく手法よりも高

精度に奥行きを推定できる。

次に、マッチングウィンドウ間の幾何補正の有無について比較する。図 3.15 の点数のヒストグラム、および、表 3.2 の 3 次元復元点数より、NCC、POC のどちらについても、マッチングウィンドウの幾何補正を行うことで、3 次元復元点数が増加している。NCC に基づく手法では、図 3.15 の割合のヒストグラムおよび、表 3.2 の誤差の絶対値の中央値より、幾何補正を行うことで奥行き推定精度も向上していることが確認できる。一方、POC に基づく手法では、幾何補正により 3 次元復元点数は増加しているが、奥行き推定精度に大きな差はない。POC に基づくウィンドウマッチングでは、マッチングウィンドウ間の画像変形が平行移動のみに近似できない領域では、マッチングの精度が低下するため、奥行き推定精度も低下する。一方で、マッチングウィンドウ間の画像変形が平行移動のみに近似できない場合、POC 関数のピーク値が著しく低下するため、幾何補正を行わないと、物体の画像変形が大きい領域において、3 次元点が復元されない。そのため、奥行き推定精度が低い点が復元されず、結果として、幾何補正の有無による差が、奥行き推定精度ではなく 3 次元点数に現れたと考えられる。図 3.16 と図 3.17 の POC に基づく手法の結果を比較すると、幾何補正を行うことで、犬のお腹部分の 3 次元点が復元されている。同様に、図 3.18 と図 3.19 の POC に基づく手法の結果を比較すると、猫のしっぽ周辺の 3 次元点群に差が生じている。このように、どの参照視点においても、画像変形が大きくなる領域では、マッチングウィンドウ間の幾何補正が必要である。一方で、奥行き推定精度のみに着目すると、POC に基づく手法では、幾何補正を行わない場合においても、復元できた点について、高精度なマッチングが行える可能性を示している。

また、提案手法 (POC+Affine) について、近傍視点数の変化に対する奥行き推定精度を比較する。ここで、各近傍視点は、参照視点 1 枚につき、基線長の短い順に選択する。表 3.3 に、近傍視点数を変化させた場合の、POC+Affine の 3 次元復元点数と誤差の絶対値の中央値を示す。表 3.3 より、近傍視点数が増えるに従い、3 次元点数が増加し、誤差が小さくなっていることが確認できる。これは、多くの近傍視点を利用することで、参照視点におけるオクルージョン領域が減少し、奥行き推定可能なピクセルが増えたため、および、複数のステレオペアの POC 関数を統合することでノイズなどの外乱の影響が低減し、奥行き推定精度が向上

表 3.3 POC+Affine における，近傍視点数に対する 3 次元復元点数と誤差の絶対値の中央値

# of neighboring views	# of points [10^6]		median of absolute values of error [%]	
	dog	cat	dog	cat
1	6.43	8.70	0.0447	0.0377
2	7.01	9.51	0.0386	0.0330
3	7.47	9.76	0.0336	0.0322

したためである．このように，提案手法では，多視点画像間のマッチングを行うことで，従来の POC のような 2 視点のみを用いた 3 次元復元に比べ，3 次元点の奥行きを高精度に推定することが可能である．

(II) データセット Fountain-P11 の実験結果

図 3.20 に，full ($3,072 \times 2,048$ ピクセル)，half ($1,536 \times 1,024$ ピクセル)，quarter (768×512 ピクセル)のそれぞれの画像サイズにおける誤差ヒストグラムを示す．ここで，ヒストグラムは 0.01% のビンで作成した．図 3.20 上段は，点数のヒストグラムであり，図 3.20 下段は，割合のヒストグラムである．各画像サイズにおける，総復元点数と誤差の絶対値の中央値を表 3.4 にまとめる．また，図 3.21，および，図 3.22 に，画像サイズ half のときの，画素値で色をつけた 3 次元点群と，誤差の絶対値で色をつけた 3 次元点群を示す．ここで，各 3 次元点群は，図の視認性のため，誤差の絶対値が非常に大きい誤対応点を手動で除去して表示した．

図 3.20 より，画像サイズ quarter および half の結果については，データセット dog，cat と同様の傾向の結果が得られている．NCC に基づく手法に比べ POC に基づく手法が，点数および割合のどちらについても奥行き推定精度が高いことが確認できる．また，NCC，POC に基づく手法のどちらについても，幾何補正を行うことで 3 次元復元点数，および，奥行き推定精度が向上している．これらの傾向は，表 3.4，および，図 3.21，図 3.22 の 3 次元点群からも確認できる．一方で，画像サイズ full の結果については，点数のヒストグラムについて，

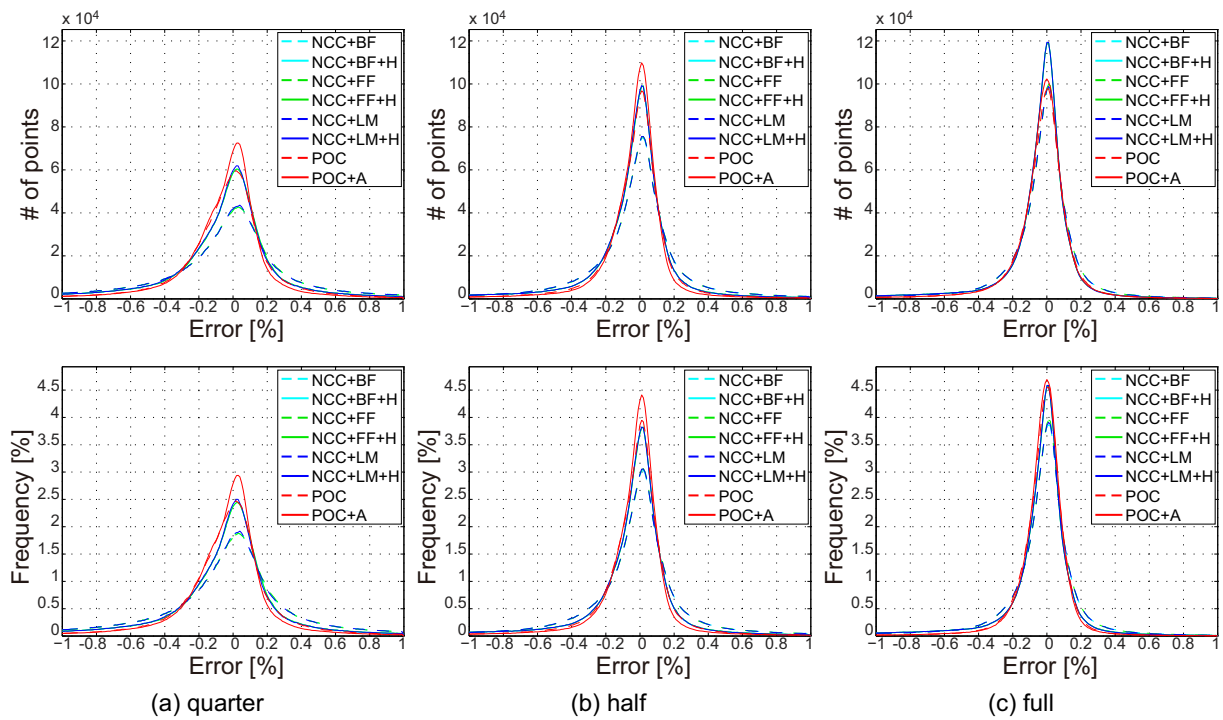


図 3.20 データセット Fountain-P11 における誤差ヒストグラム (上段: 点数のヒストグラム, 下段: 点の割合のヒストグラム)

NCC に基づく手法よりも POC に基づく手法の点数が少ない。この原因として、画像サイズが大きい場合、ウィンドウサイズが相対的に小さくなるため、テクスチャが少ない領域が増えたこと、カラー画像のため、原画像においてカラーフィルタのベイヤーパターンによる影響が無視できないことが考えられる。このような領域では、物体表面テクスチャの信号に対し、ノイズやカラーフィルタの違いなどの外乱が大きくなるため、POC 関数のピーク値が大幅に低下する。その結果、3次元点が復元されなかったと考えられる。一方、図 3.20 の割合のヒストグラム、および、表 3.4 の誤差の絶対値の中央値より、復元できた点については、NCC に基づく手法に比べ POC に基づく手法の奥行き推定精度が高いことが確認できる。

また、表 3.4 より、ウィンドウマッチング手法に依らず、画像サイズが大きくなるほど奥行き推定精度が向上することが確認できる。これは、画像サイズが大きくなるほど、ステレオ画像上の 1 ピクセル当りのマッチング誤差に対して、3次元点の奥行きの誤差が小さくなるため

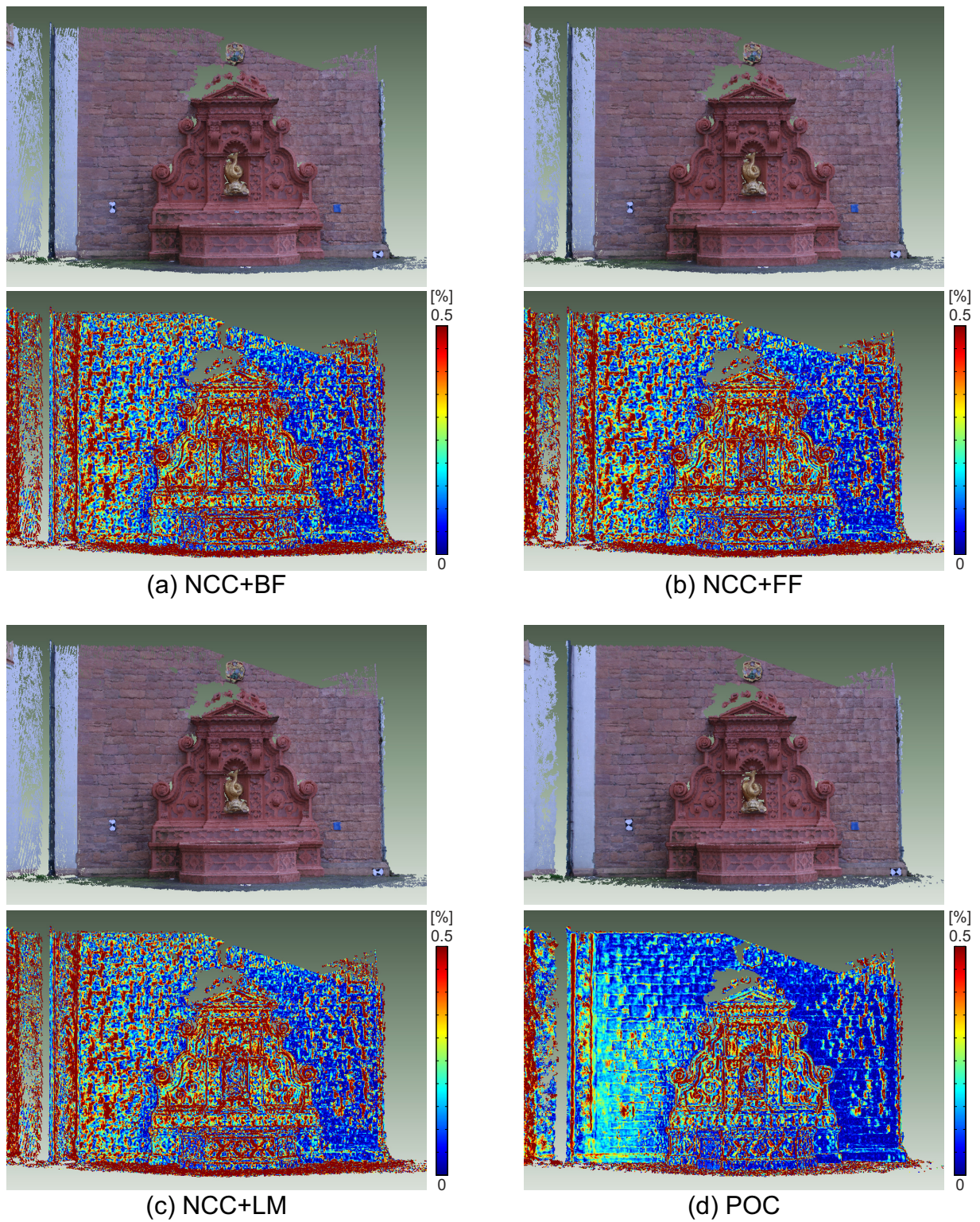


図 3.21 データセット Fountain-P11 (half) における，幾何補正を行わない場合の 3次元復元点群と誤差（それぞれの手法について，上段: 3次元復元点群，下段: 3次元復元誤差の絶対値）

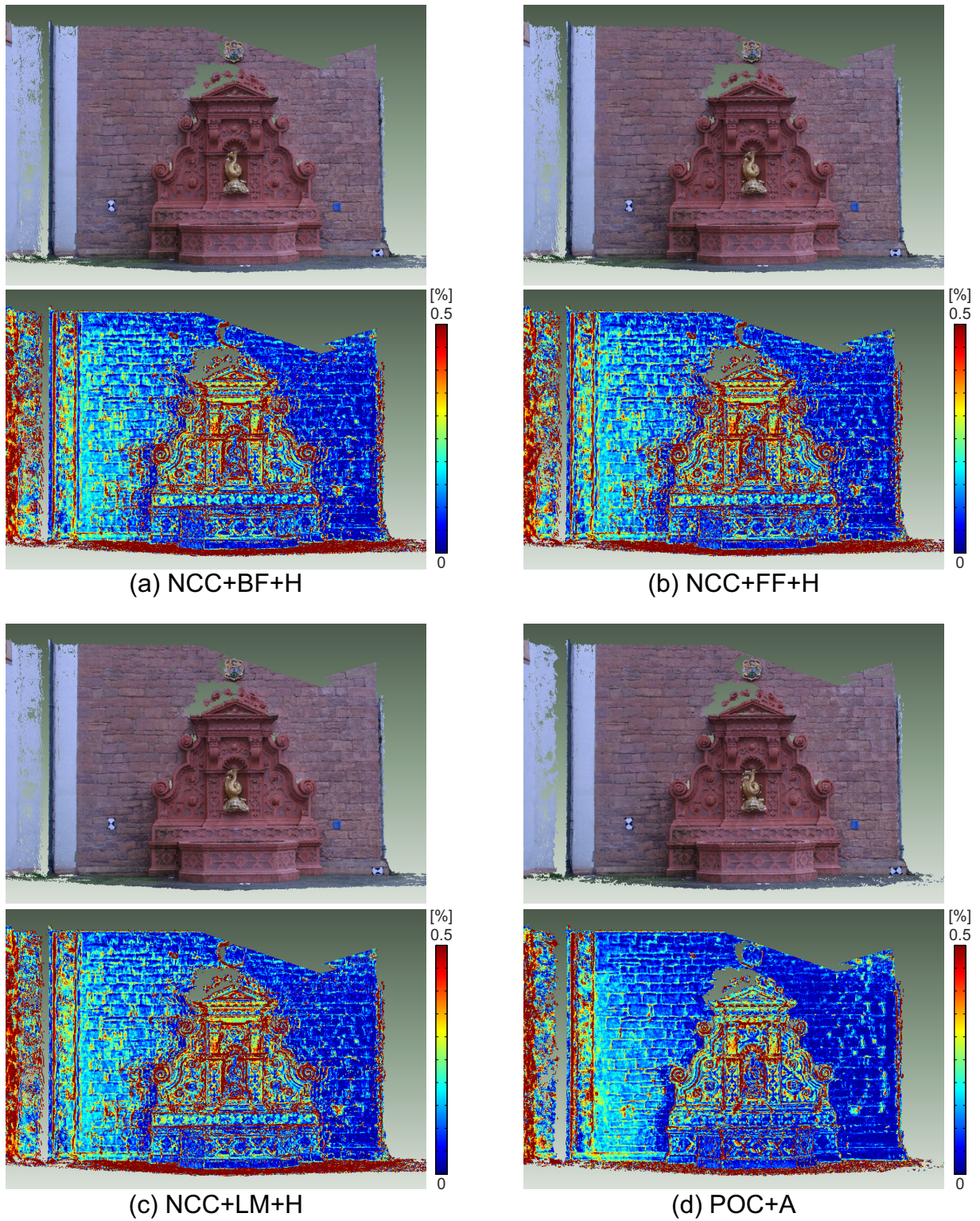


図 3.22 データセット Fountain-P11 (half) における，幾何補正を行う場合の 3 次元復元点群と誤差（それぞれの手法について，上段: 3 次元復元点群，下段: 3 次元復元誤差の絶対値）

表 3.4 データセット Fountain-P11 における 3 次元復元点数と誤差の絶対値の中央値

	# of points [10^6]			median of absolute values of error [%]		
	quarter	half	full	quarter	half	full
NCC+BF	2.29	2.55	2.53	0.1700	0.1122	0.0804
NCC+FF	2.30	2.55	2.52	0.1702	0.1115	0.0795
NCC+LM	2.30	2.56	2.53	0.1694	0.1117	0.0799
POC	2.44	2.53	2.40	0.1209	0.0817	0.0636
NCC+BF+H	2.50	2.68	2.61	0.1301	0.0859	0.0671
NCC+FF+H	2.50	2.68	2.60	0.1301	0.0859	0.0671
NCC+LM+H	2.50	2.68	2.60	0.1290	0.0856	0.0669
POC+A	2.49	2.57	2.45	0.1065	0.0715	0.0604

である．画像サイズに依らずステレオ画像上で同等の精度でマッチングできれば，画像サイズを大きくすることで 3 次元点の奥行き推定精度は向上する．一方で，3 次元復元点数は，画像サイズ half が最も多く，画像サイズ full の復元点数の方が小さい．これは，POC に基づく手法だけでなく，NCC に基づく手法についても，テクスチャの少ない領域におけるノイズや，カラーフィルタのベイヤーパターンによる影響が無視できないためである．

3.5.4 計算コストの評価

本小節では，MVS におけるウィンドウマッチングの計算コストについて比較する．参照視点 V_R 上のある 1 点のデジタル画像座標について，一定の探索範囲内の奥行きを求めるのに必要な計算コストで評価する．MVS のための POC に基づくウィンドウマッチングは，1 回のマッチングで，ステレオ画像上でウィンドウサイズの $\pm 1/4$ に相当する範囲内の奥行きを推定できる．本小節では，NCC に基づくウィンドウマッチングにおいても，同等の範囲の奥行き推定に必要な計算コストを求める．本論文では，処理に必要な加算・乗算・除算・平方根

表 3.5 1 点の 3 次元点の奥行きを決定するのに必要な計算コスト

	Additions	Multiplications	Divisions	Square roots	Total cost
NCC+BF	751,400	312,460	5,780	5,780	3,885,860
NCC+FF	75,143	31,250	579	578	388,621
NCC+LM	145,860	60,654	1,122	1,122	754,314
POC	40,000	34,496	2,176	1,088	312,064
NCC+BF+H	6,762,600	2,812,140	52,020	52,020	34,972,740
NCC+FF+H	676,263	281,218	5,203	5,202	3,497,309
NCC+LM+H	740,350	307,865	5,695	5,695	3,828,715
POC+A	360,000	310,464	19,584	9,792	2,808,576

演算の数で計算コストを評価する．表 3.5 に，それぞれの手法における，1 点の奥行き探索に必要な演算の回数と計算コストを示す．ここで，計算コストは，加算・乗算・除算・平方根の演算回数の重み付き和であり，重みをそれぞれ 3 : 5 : 6 : 6 とした．なお，この重みは，一般に広く利用されるプロセッサのアーキテクチャである Intel®Core™Microarchitecture における各演算の該当命令のレイテンシに基づいて決定した [86]．ここで，NCC+LM (+H) の計算コストは，入力画像に依存するため，データセット Fountain-P11 (full) における平均計算コストとする．

まず，NCC に基づく手法と POC に基づく手法の計算コストを比較する．表 3.5 より，POC に基づく手法は，NCC に基づく手法に比べて計算コストが小さいことが確認できる．NCC に基づく手法では，そのサブピクセル推定によって，計算コストが大きく変化する．特に，NCC+BF では，非常に細かい刻み幅でマッチングを繰り返すため，計算コストが膨大になる．これに対して，NCC+FF では，ピクセルレベルの奥行き変化のみでマッチングスコア算出をするため，計算コストが少ない．POC に基づくウィンドウマッチングでは，NCC+FF と同様に，POC 関数の離散値（ピクセルレベルのマッチングスコアに相当する）に対してピークモデルをフィッティングするため，少ない計算コストで奥行き推定が可能である．

次に、幾何補正の有無について比較する。本実験では、NCC に基づく手法と POC に基づく手法のどちらについても、マッチングウィンドウの幾何補正を行う場合、単純に 9 つの法線ベクトルについて試行した。そのため、計算コストは、NCC+LM を除いて、NCC に基づく手法と POC に基づく手法のどちらでも、幾何補正を行わない場合に比べ 9 倍に増加している。幾何補正を伴う NCC に基づくウィンドウマッチングを用いた従来の MVS アルゴリズム [19], [27] では、このような単純な法線ベクトルの試行ではなく、周囲の画像座標の情報を伝搬するなどして計算コストを抑えている。このように、マッチングウィンドウ間の幾何補正を行う場合には、計算コストの増加を考慮して適用することが重要である。

3.6 むすび

本章では、MVS のための高精度ウィンドウマッチング手法について述べた。まず、POC の基本原理、および、POC を用いた平行移動量推定の高精度化手法を述べた。次に、MVS のための POC に基づくウィンドウマッチングを提案した。3 次元点の奥行き変化をステレオ画像間の平行移動量に近似し、その平行移動量を POC を用いてサブピクセル精度で推定することで、高精度な奥行き推定が可能である。また、提案手法では、複数のステレオペアから計算される POC 関数を正規化視差に基づき統合することで、精度・ロバスト性を向上させた。さらに、画像変形の大きいステレオペアに対応するため、マッチングウィンドウの幾何補正を適用した。仮定した 3 次元平面とカメラパラメータから算出される拡大縮小率・スキューの傾きを用いて画像変形を軽減することで、ステレオペア間の画像変形が大きい場合においても、マッチングウィンドウ間の変形を平行移動のみに近似可能である。最後に、実験により、さまざまなウィンドウマッチング手法と奥行き推定精度、および、計算コストを比較することで、提案手法の性能評価を行った。従来の NCC に基づくウィンドウマッチングと比べ、提案手法を用いることで、より高精度に 3 次元点の奥行きを推定できることを示した。

第 4 章

高精度ウィンドウマッチングに基づく 多視点ステレオアルゴリズム

4.1 まえがき

多視点ステレオ (Multi-View Stereo: MVS) アルゴリズムは、カメラ選択、3次元点群生成、誤対応点除去、3次元メッシュ生成、3次元メッシュの最適化など、様々な技術の複合として構成される [3]。第 3 章では、MVS のための位相限定相関法 (Phase-Only Correlation: POC) に基づくウィンドウマッチングを提案し、従来のウィンドウマッチング手法と比べ、高精度に 3次元点の奥行きを推定できることを示した。本章では、第 3 章で提案した MVS のための POC に基づくウィンドウマッチングをもとに、POC に基づく MVS アルゴリズムを提案する。MVS のための POC に基づくウィンドウマッチングにより高精度に 3次元座標を計算することで、MVS アルゴリズムの処理全体における 3次元復元結果の精度向上が望める。しかしながら、ウィンドウマッチングを用いた 3次元点の奥行き推定だけでは、誤対応点やアーチファクト (MVS アルゴリズムで物体表面に復元される、本来、対象物体に存在しない構造物) の発生を完全に防ぐことができず、高精度な 3次元復元が困難である。これは、あるひとつの視点の、あるひとつの画像座標におけるマッチング結果だけでは、テクスチャが少ない領域や物体境界・オクルージョン境界において、真の奥行きのマッチングスコアが低くなったり、誤対応となる奥行きのマッチングスコアが高くなったりした場合、推定された 3次

元点の奥行きが正しいのか誤対応なのか判断できないためである。そこで、提案手法では、近傍ピクセルのマッチング結果や、近傍視点のマッチング結果との整合性を図る。まず、近傍ピクセルのマッチング結果を用いるために、階層的探索を用いたデプスマップ生成を提案し、重み付きメディアンフィルタを用いた高精度化を適用する。次に、近傍視点のマッチング結果の整合性を取ることで、誤対応点除去、および、アーチファクト除去を行う。これにより、各視点の各ピクセルの奥行きを独立に推定し、単純に世界座標系に統合する場合に比べ、誤対応点・アーチファクトの発生を抑え、高精度な3次元復元が可能になる。

以下では、POCに基づくMVSアルゴリズムについて述べる。まず、MVSのためのPOCに基づくウィンドウマッチングを用いたプレーンスイーピングに対して、(I) 階層的探索を用いたデプスマップ生成、(II) 重み付きメディアンフィルタを用いたデプスマップの高精度化、(III) グラフカットに基づく誤対応点除去、(IV) 閾値処理に基づくアーチファクト除去を適用し、POCに基づくMVSアルゴリズムを提案する。次に、公開データセットを用いた実験により、従来のMVSアルゴリズムと比較し、提案手法を用いて高精度な3次元復元が行えることを示す。そして、提案手法の応用として、デジタルアーカイブを目的とした文化財の3次元復元について述べ、提案手法の有効性を示す。

4.2 位相限定相関法に基づく多視点ステレオアルゴリズム

本章では、第3章で提案したMVSのためのPOCに基づくウィンドウマッチングをもとに、POCに基づくMVSアルゴリズムを提案する。提案手法は、Seitzらの4つ分類[5]における、デプスマップ統合に基づくアルゴリズムに従う。まず、各視点におけるデプスマップを生成し、次に、それらのデプスマップを統合することで3次元復元を行う。デプスマップの生成では、MVSのためのPOCに基づくウィンドウマッチングによる奥行き推定を、プレーンスイーピングと階層的探索の枠組みの中で利用する。さらに、階層的探索における各階層のデプスマップに対して、重み付きメディアンフィルタを適用することで、デプスマップの高精度化を行う。デプスマップ統合では、グラフカットに基づく誤対応除去、および、閾値処理に基づくアーチファクト除去により、複数のデプスマップ間の整合性を取りつつ、3次元点群を

生成する．以下では，(I) MVS のための POC に基づくウィンドウマッチングと階層的探索を用いたデプスマップ生成，(II) 重み付きメディアンフィルタを用いたデプスマップの高精度化，(III) グラフカットに基づく誤対応点除去，(IV) 閾値処理に基づくアーチファクト除去，(V) デプスマップからの 3 次元点群生成の各処理について説明する．

4.2.1 位相限定相関法と階層的探索を用いたデプスマップ生成

本章で提案する POC に基づく MVS アルゴリズムは，デプスマップ統合に基づく MVS アルゴリズムであり，本小節では，そのデプスマップ生成について説明する．提案手法におけるデプスマップ生成は，プレーンスイーピング [3], [9] と階層的探索 [31], [34] を組み合わせて使用する．プレーンスイーピングは，総当たり探索によるデプスマップ生成手法であり，参照視点の各ピクセルについて，奥行き Z や法線ベクトル n を変化させながらウィンドウマッチングを繰り返し，最もマッチングスコアの高くなる奥行き Z を，デプスマップ上のそのピクセルの値とする [20]．ウィンドウマッチングに MVS のための POC に基づくウィンドウマッチングを使用する場合，奥行きの初期値 Z_{init} と法線ベクトル n を変化させながら奥行き推定を行い，最も POC 関数のピーク値 α が高くなる時に推定された奥行き Z を，デプスマップ上のそのピクセルの値とする [32], [38]．このような単純なプレーンスイーピングでは，局所的にテクスチャが少ない領域において，画像のノイズにより正しい奥行きでも POC 関数のピーク値が低下したり，局所的な繰り返しパターンにより間違った奥行きでも高いピーク値が計算されたりすると，誤対応点が発生する．MVS のための POC に基づくウィンドウマッチングでは，複数のステレオペアの POC 関数を統合することで，ノイズや繰り返しパターンなどの影響を低減しているが，誤対応点の発生を完全に防ぐことはできない．一方で，このように誤対応点が発生しやすい領域でも，マッチングウィンドウのサイズを大きくすることで，マッチングのロバスト性が向上する．これは，より広い範囲の信号を考慮することで，ウィンドウ内にテクスチャが多い領域が含まれたり，ウィンドウ内のテクスチャが繰り返しパターンでなくなるためである．しかしながら，サイズの大きいマッチングウィンドウを使用すると，物体の細かい形状が復元できないといった問題が発生する．そこで，提案手法では，MVS のための POC に基づくウィンドウマッチングを用いたプレーンスイーピングに対して，階層的探索を

適用する [31] . 階層的探索では, 原画像を $1/2$ ずつ縮小した階層画像を作成し, 上位層の奥行き推定結果を下位層に伝搬しながらデプスマップを生成する. 各階層におけるマッチングウィンドウサイズを固定することで, 上位層では画像全体に対して相対的に大きいウィンドウサイズで, 下位層では相対的に小さいウィンドウサイズでマッチングを行う. さらに, 提案手法におけるデプスマップ生成では, より詳細なデプスマップを生成するため, 最下位層において, ウィンドウサイズを小さくしながらマッチングを繰り返す. これにより, 誤対応点の発生を抑制しつつ, 物体の細かい形状の 3 次元復元が可能になる.

以下では, カメラパラメータが既知の多視点画像 $V = \{V_0, \dots, V_{K_{\text{all}}-1}\}$ について, 1 枚の参照視点 $V_R \in V$ におけるデプスマップ $I_Z(u, v)$ を求める問題を考える. 入力は, 参照視点 V_R とその近傍視点 $C = \{C_0, \dots, C_{K-1}\} \subseteq V - \{V_R\}$ とする. ここで, K_{all} および K は, それぞれ多視点画像の数と近傍視点の数である. 出力は, 参照視点 V_R におけるデプスマップ $I_Z(u, v)$ および相関値マップ $I_\alpha(u, v)$ とする. 提案手法におけるデプスマップ生成は, (i) 階層画像の作成, (ii) 最上位層の処理, (iii) 中間層の処理, (iv) 最下位層の処理で構成される. 処理の流れを図 4.1 に示す. (i) から (iv) の各処理について詳しく説明する.

(i) 階層画像の作成

画像を $1/2$ ずつ縮小しながら階層画像を作成する. 階層数 H , 階層番号 $h = 0, \dots, H-1$ について, h 階層における参照視点を V_R^h , 近傍視点群を C^h とする. このとき, V_R^h の画像サイズは, V_R^{h-1} の画像サイズの $1/2$ となる. 同様に, 階層ごとのデプスマップを $I_Z^h(u, v)$, 階層ごとの相関値マップを $I_\alpha^h(u, v)$ とする. また, 処理の中間データとして必要になる法線ベクトルマップについて, 階層ごとの法線ベクトルマップを $I_\psi^h(u, v)$, $I_\phi^h(u, v)$ とする. ここで, $I_\psi^h(u, v)$ は参照視点のカメラ座標系における法線ベクトルの X 成分を, $I_\phi^h(u, v)$ は法線ベクトルの Y 成分を表す. 各階層のデプスマップ $I_Z^h(u, v)$, 相関値マップ $I_\alpha^h(u, v)$, 法線ベクトルマップ $I_\psi^h(u, v)$, $I_\phi^h(u, v)$ のサイズは, その階層の参照視点 V_R^h の画像サイズと同じであり, h が大きくなるにつれ $1/2$ ずつ小さくなる.

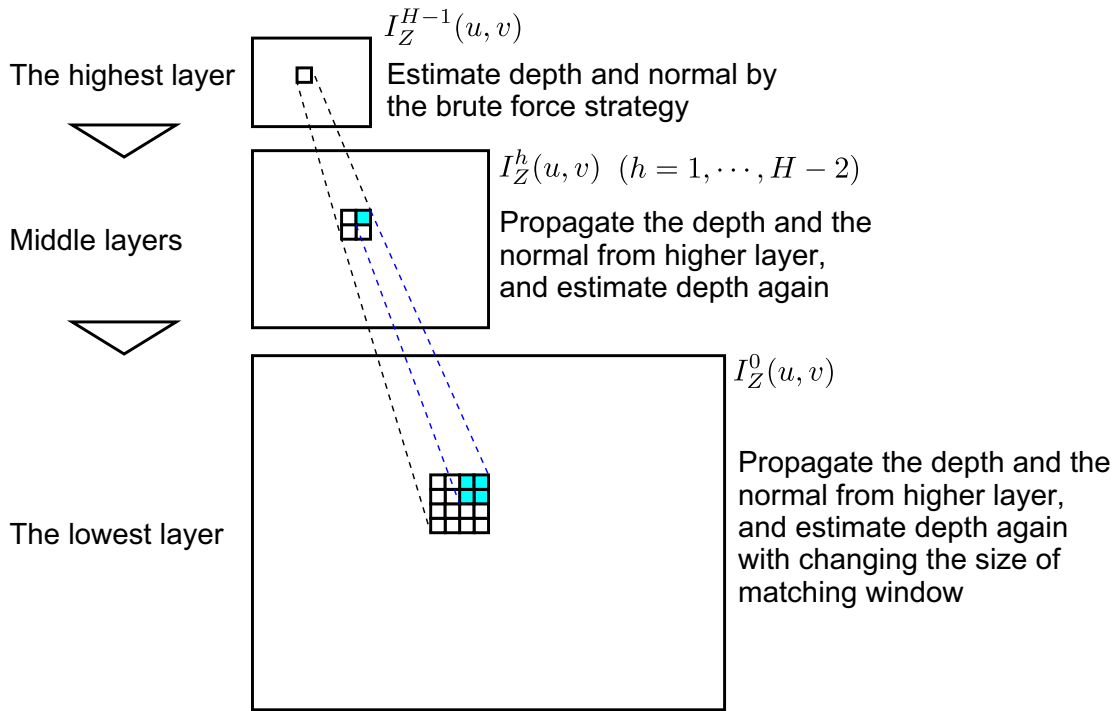


図 4.1 階層的探索を用いたデプスマップ生成

(ii) 最上位層の処理

最上位層では，MVS のための POC に基づくウィンドウマッチングを用いたプレーン スweepingにより各ピクセルの奥行きと法線ベクトルを推定する．階層 $H - 1$ の各ピクセル (u, v) について，参照視点 V_R^{H-1} の 3 次元候補点の奥行き Z_{init} と法線ベクトル $n = [n_X, n_Y, n_Z]^T$ を変化させながら，MVS のための POC に基づくウィンドウマッチングにより真の奥行き Z を繰り返し推定する．MVS のための POC に基づくウィンドウマッチングのウィンドウサイズは， $N \times L$ ピクセルとする．このとき，奥行きの初期値 Z_{init} は，近傍視点群の中で最も基線長が長いステレオ画像上で $N/4$ ピクセルに相当する刻み幅で変化させる．また，法線ベクトル n は，参照視点 V_R の画像平面に正対する法線と， X 軸と Y 軸方向に 45 度ずつ傾けた合計 9 つの法線について試行する．参照視点 V_R^{H-1} および近傍視点群 C^{H-1} について，奥行きの初期値 Z_{init} と法線ベクトル n を仮定したとき，真の 3 次元点の奥行き Z を推定する手順は，3.4.2 節で述べた通りである．奥行きの初期値 Z_{init} と法線ベクトル n の変化に対

して、最も POC 関数のピーク値 α が高くなるときに推定された奥行き Z 、POC 関数のピーク値 α 、および、仮定した法線ベクトル \mathbf{n} から、ピクセル (u, v) におけるデプスマップ $I_Z^{H-1}(u, v)$ 、相関値マップ $I_\alpha^{H-1}(u, v)$ 、法線ベクトルマップ $I_\psi^{H-1}(u, v)$ 、 $I_\phi^{H-1}(u, v)$ の値を決定する。

$$I_Z^{H-1}(u, v) \leftarrow Z \quad (4.1)$$

$$I_\alpha^{H-1}(u, v) \leftarrow \alpha \quad (4.2)$$

$$I_\psi^{H-1}(u, v) \leftarrow n_X \quad (4.3)$$

$$I_\phi^{H-1}(u, v) \leftarrow n_Y \quad (4.4)$$

ここで、 \leftarrow は代入を表す。階層 $H-1$ のすべてのピクセルについて、上記手順により、デプスマップ $I_Z^{H-1}(u, v)$ 、相関値マップ $I_\alpha^{H-1}(u, v)$ 、法線ベクトルマップ $I_\psi^{H-1}(u, v)$ 、 $I_\phi^{H-1}(u, v)$ の値を決定する。ただし、MVS のための POC に基づくウィンドウマッチングにおいて、すべてのステレオペアの POC 関数のピーク値が閾値を下回った場合、3次元点の奥行きが算出されないため、そのピクセルのそれぞれのマップの値を 0 とする。また、階層番号 h について、 $h \leftarrow H-1$ とする。

(iii) 中間層の処理

階層番号を 1 つ減らし、 $h \leftarrow h-1$ とする。中間層の処理では、1 つ上位の階層の結果を初期値として、MVS のための POC に基づくウィンドウマッチングにより奥行き推定を行う。階層 h の各ピクセル (u, v) について、奥行き初期値 Z_{init} および法線ベクトル $\mathbf{n} = [n_X, n_Y, n_Z]^T$ を、

$$Z_{\text{init}} \leftarrow I_Z^{h+1} \left(\left[\frac{u}{2} \right], \left[\frac{v}{2} \right] \right) \quad (4.5)$$

$$\mathbf{n} \leftarrow \frac{1}{\sqrt{\psi^2 + \phi^2 + 1}} \begin{bmatrix} \psi \\ \phi \\ 1 \end{bmatrix} \quad (4.6)$$

とする。ここで、

$$\psi = I_\psi^{h+1} \left(\left[\frac{u}{2} \right], \left[\frac{v}{2} \right] \right) \quad (4.7)$$

$$\phi = I_\phi^{h+1} \left(\left[\frac{u}{2} \right], \left[\frac{v}{2} \right] \right) \quad (4.8)$$

である．以上のように決定した奥行きの初期値 Z_{init} および法線ベクトル \mathbf{n} について，参照視点 V_R^h および近傍視点群 C^h に対して MVS のための POC に基づくウィンドウマッチングを行い，真の 3 次元点の奥行き Z を推定する．推定された奥行き Z ，POC 関数のピーク値 α ，および，仮定した法線ベクトル \mathbf{n} から，ピクセル (u, v) におけるデプスマップ $I_Z^h(u, v)$ ，相関値マップ $I_\alpha^h(u, v)$ ，法線ベクトルマップ $I_\psi^h(u, v)$ ， $I_\phi^h(u, v)$ の値を決定する．

$$I_Z^h(u, v) \leftarrow Z \quad (4.9)$$

$$I_\alpha^h(u, v) \leftarrow \alpha \quad (4.10)$$

$$I_\psi^h(u, v) \leftarrow n_X \quad (4.11)$$

$$I_\phi^h(u, v) \leftarrow n_Y \quad (4.12)$$

階層 h のすべてのピクセルについて，上記手順により，デプスマップ $I_Z^h(u, v)$ ，相関値マップ $I_\alpha^h(u, v)$ ，法線ベクトルマップ $I_\psi^h(u, v)$ ， $I_\phi^h(u, v)$ の値を決定する．ただし，MVS のための POC に基づくウィンドウマッチングにおいて，すべてのステレオペアの POC 関数のピーク値が閾値を下回った場合，そのピクセルのそれぞれのマップの値は 0 とする． $h > 0$ である限り，この処理を繰り返す．

(iv) 最下位層の処理

最下位層の処理では，第 1 階層の結果を初期として，MVS のための POC に基づくウィンドウマッチングにより奥行き推定を行う．さらに，ウィンドウサイズを小さくしながら，マッチングを繰り返し，より詳細なデプスマップを生成する．

まず，第 1 階層の結果から，最下位層の各ピクセル (u, v) について，奥行き初期値 Z_{init} および法線ベクトル $\mathbf{n} = [n_X, n_Y, n_Z]^T$ を，

$$Z_{\text{init}} \leftarrow I_Z^1 \left(\left[\frac{u}{2} \right], \left[\frac{v}{2} \right] \right) \quad (4.13)$$

$$\mathbf{n} \leftarrow \frac{1}{\sqrt{\psi^2 + \phi^2 + 1}} \begin{bmatrix} \psi \\ \phi \\ 1 \end{bmatrix} \quad (4.14)$$

とする．ここで，

$$\psi = I_{\psi}^1 \left(\left[\frac{u}{2} \right], \left[\frac{v}{2} \right] \right) \quad (4.15)$$

$$\phi = I_{\phi}^1 \left(\left[\frac{u}{2} \right], \left[\frac{v}{2} \right] \right) \quad (4.16)$$

である．以上のように決定した奥行きの初期値 Z_{init} および法線ベクトル \mathbf{n} について，参照視点 V_R^0 および近傍視点群 C^0 に対して MVS のための POC に基づくウィンドウマッチングを行い，真の 3 次元点の奥行き Z を推定する．ここで，ウィンドウサイズは $N \times L$ ピクセルとする．推定された奥行き Z ，POC 関数のピーク値 α ，および，仮定した法線ベクトル \mathbf{n} から，ピクセル (u, v) におけるデプスマップ $I_Z^0(u, v)$ ，相関値マップ $I_{\alpha}^0(u, v)$ ，法線ベクトルマップ $I_{\psi}^0(u, v)$ ， $I_{\phi}^0(u, v)$ の値を決定する．

$$I_Z^0(u, v) \leftarrow Z \quad (4.17)$$

$$I_{\alpha}^0(u, v) \leftarrow \alpha \quad (4.18)$$

$$I_{\psi}^0(u, v) \leftarrow n_X \quad (4.19)$$

$$I_{\phi}^0(u, v) \leftarrow n_Y \quad (4.20)$$

次に，ウィンドウサイズを小さくしながら，マッチングを繰り返し，より詳細なデプスマップを生成する．ウィンドウサイズ $N \times L$ を $1/2$ ずつ小さくする．

$$N \leftarrow N/2 \quad (4.21)$$

$$L \leftarrow L/2 \quad (4.22)$$

奥行き初期値 Z_{init} および法線ベクトル $\mathbf{n} = [n_X, n_Y, n_Z]^T$ を，

$$Z_{\text{init}} \leftarrow I_Z^0(u, v) \quad (4.23)$$

$$\mathbf{n} \leftarrow \frac{0}{\sqrt{\psi^2 + \phi^2 + 1}} \begin{bmatrix} \psi \\ \phi \\ 1 \end{bmatrix} \quad (4.24)$$

とする．ここで，

$$\psi = I_{\psi}^0(u, v) \quad (4.25)$$

$$\phi = I_{\phi}^0(u, v) \quad (4.26)$$

である．奥行きの初期値 Z_{init} および法線ベクトル n について，参照視点 V_R^0 および近傍視点群 C^0 に対して MVS のための POC に基づくウィンドウマッチングを行い，真の 3 次元点の奥行き Z を推定し， $I_Z^0(u, v)$ ， $I_\alpha^0(u, v)$ ， $I_\psi^0(u, v)$ ， $I_\phi^0(u, v)$ の値を更新する．

$$I_Z^0(u, v) \leftarrow Z \quad (4.27)$$

$$I_\alpha^0(u, v) \leftarrow \alpha \quad (4.28)$$

$$I_\psi^0(u, v) \leftarrow n_X \quad (4.29)$$

$$I_\phi^0(u, v) \leftarrow n_Y \quad (4.30)$$

ただし，MVS のための POC に基づくウィンドウマッチングにおいて，すべてのステレオペアの POC 関数のピーク値が閾値を下回った場合，そのピクセルのそれぞれのマップの値は 0 とする． $N \geq N_{\text{min}}$ である限り，この処理を繰り返す．ここで， N_{min} は，最小のウィンドウサイズを表すパラメータである．

最下位層のすべてのピクセルについて，上記手順により，デプスマップ $I_Z^0(u, v)$ ，相関値マップ $I_\alpha^0(u, v)$ ，法線ベクトルマップ $I_\psi^0(u, v)$ ， $I_\phi^0(u, v)$ の値を決定する．最終的なデプスマップ $I_Z(u, v)$ および相関値マップ $I_\alpha(u, v)$ を，最下位層におけるデプスマップ $I_Z^0(u, v)$ ，および，相関値マップ $I_\alpha^0(u, v)$ とする．

$$I_Z(u, v) \leftarrow I_Z^0(u, v) \quad (4.31)$$

$$I_\alpha(u, v) \leftarrow I_\alpha^0(u, v) \quad (4.32)$$

提案手法では，すべての参照視点 $V_R \in V$ について，上記処理によりデプスマップ $I_Z(u, v)$ ，および，相関値マップ $I_\alpha(u, v)$ を生成する．

4.2.2 重み付きメディアンフィルタを用いたデプスマップの高精度化

4.2.1 節で述べた階層的探索では，上位層の 1 つのピクセルのマッチング結果が，下位層の複数のピクセルに影響を与えるため，上位層における誤対応が最終的なデプスマップの多くのピクセルの誤対応につながる．また，上位の階層ほど，相対的に画像サイズに対するウィンド

ウサイズが大きくなるため、物体境界・オクルージョン境界の影響範囲が大きくなる。この問題に対し、提案手法では、重み付きメディアンフィルタによるデプスマップの高精度化を行う。重み付きメディアンフィルタは、2 眼ステレオにおける視差マップ生成の高精度化に利用されており、孤立した誤対応や、物体境界・オクルージョン境界における誤対応の修正が可能である [87]–[89]。提案手法では、各階層におけるデプスマップ $I_Z^h(u, v)$ 、相関値マップ $I_\alpha^h(u, v)$ 、法線ベクトルマップ $I_\psi^h(u, v)$ 、 $I_\phi^h(u, v)$ が生成されるたびに、重み付きメディアンフィルタを $I_Z^h(u, v)$ 、 $I_\alpha^h(u, v)$ 、 $I_\psi^h(u, v)$ 、 $I_\phi^h(u, v)$ に適用する。これにより、上位層において突発的に発生した誤対応や、物体境界・オクルージョン境界付近で発生するアーチファクトの影響が下位層に伝搬することを防ぐ。ただし、最終的なデプスマップである最下位層のデプスマップに対しては、MVS のための POC に基づくウィンドウマッチングによる奥行き推定結果そのものを用いるために、重み付きメディアンフィルタを適用しない。また、提案手法では、重み付きメディアンフィルタにおける重み関数として、原画像におけるバイラテラル重みと、相関値マップの積を用いる。

以下では、階層 h におけるデプスマップ $I_Z^h(u, v)$ を例に、提案手法で使用する重み付きメディアンフィルタについて説明する。重み付きメディアンフィルタは、フィルタリング後の注目ピクセルの値を、注目ピクセル周辺のピクセルの値の重み付き中央値で決定するフィルタである。ソートされた数列 $(a_k)_{k=0,1,\dots,N-1}$ と各要素の重み $(w_k)_{k=0,1,\dots,N}$ が与えられた際に、数列 (a_k) の重み付き中央値 a_{med} は、次式で与えられる。

$$a_{\text{med}} = a_n \quad (4.33)$$

$$n = \underset{l}{\operatorname{argmin}} \left| \sum_{i=0}^l w_i - \sum_{i=l+1}^{N-1} w_i \right| \quad (4.34)$$

重み付きメディアンフィルタでは、ある注目ピクセルについて、注目ピクセル周辺のマップ値（ウィンドウ） $I_{\text{med}}(i, j)$ とその重み $w_{\text{med}}(i, j)$ をソートし、数列 (a_k) および (w_k) として、重み付き中央値を求める。ここで、 (i, j) は、フィルタのウィンドウ内の座標であり、 $i = -M_w, \dots, M_w$ 、 $j = -M_w, \dots, M_w$ である。また、フィルタのウィンドウサイズは、 $N_w \times N_w$ ($N_w = 2M_w + 1$) ピクセルである。デプスマップ $I_Z^h(u, v)$ における座標系とウィンドウ $I_{\text{med}}(i, j)$ における座標系の関係を図 4.2 に示す。このとき、注目ピクセル (u', v') と、

階層 h におけるデプスマップ $I_Z^h(u, v)$ について，ウィンドウ内のマップ値 $I_{\text{med}}(i, j)$ は，

$$I_{\text{med}}(i, j) \leftarrow I_Z^h(u' + i, v' + j) \quad (4.35)$$

となる．また，重み $w_{\text{med}}(i, j)$ は，

$$w(i, j) \leftarrow I_\alpha^h(u' + i, v' + j) \exp\left(-\frac{i^2 + j^2}{2\sigma_{w1}^2}\right) \exp\left(-\frac{(I_V^h(u', v') - I_V^h(u' + i, v' + j))^2}{2\sigma_{w2}^2}\right) \quad (4.36)$$

となる．ここで， (u', v') はデプスマップ $I_Z^h(u, v)$ の座標系におけるフィルタのウィンドウ中心を， $I_V^h(u, v)$ は階層 h における参照視点のカメラ画像の輝度値を表し， σ_{w1} および σ_{w2} はパラメータである．式 (4.36) において，考慮されている仮定は，以下に示す通りである．

- 注目ピクセルとデジタル画像座標が近いピクセルは奥行き値も近い．
- 注目ピクセルとカメラ画像上の輝度値が近いピクセルは奥行き値も近い．
- POC 関数のピーク値が低いピクセルは奥行き値の信頼性が低い．

また，式 (4.35) および式 (4.36) について， $I_Z^h(u, v)$ および $I_V^h(u, v)$ の画像外の値は 0 とする．このように決定されたウィンドウ内のマップ値 $I_{\text{med}}(i, j)$ とその重み $w_{\text{med}}(i, j)$ について， $I_{\text{med}}(i, j)$ をソートした数列 (a_k) と， $I_{\text{med}}(i, j)$ と同様の並び替えを $w_{\text{med}}(i, j)$ に対して行った重み数列 (w_k) から，式 (4.34) により (a_k) の重み付き中央値 a_{med} を計算し，フィルタリング後のデプスマップ $I_Z^h(u', v')$ の値とする．

$$I_Z^h(u', v') \leftarrow a_{\text{med}} \quad (4.37)$$

デプスマップ $I_Z^h(u, v)$ 上のすべてのピクセルについて注目ピクセル (u', v') を変えながら，重み付き中央値を計算し，代入することで，デプスマップ $I_Z^h(u, v)$ を更新する．

デプスマップ $I_Z^h(u, v)$ と同様に，相関値マップ $I_\alpha^h(u, v)$ ，法線ベクトルマップ $I_\psi^h(u, v)$ ， $I_\phi^h(u, v)$ についても，上記の重み付きメディアンフィルタにより，マップを更新する．このとき，フィルタのウィンドウ内のマップ値 $I_{\text{med}}(i, j)$ は，式 (4.35) と異なり，それぞれのマップ値を代入する必要があるが，重み $w_{\text{med}}(i, j)$ の計算には式 (4.36) を用いる．

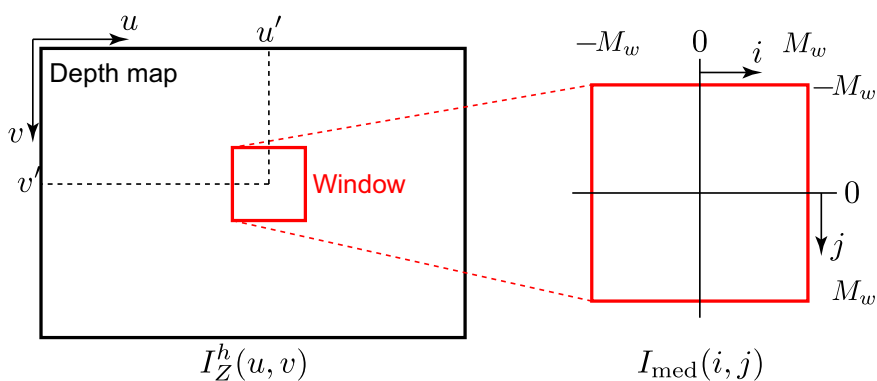


図 4.2 デプスマップ $I_Z^h(u, v)$ における座標系とウィンドウ $I_{\text{med}}(i, j)$ における座標系の関係

4.2.3 グラフカットに基づく誤対応点除去

4.2.1 節, および, 4.2.2 節で述べたアルゴリズムによって生成されたデプスマップを統合することで, 3次元点群を生成する. しかしながら, 各視点におけるデプスマップから算出される3次元点群を世界座標系に変換するだけでは, 最終的な3次元復元結果に多くの誤対応点が残される. これは, デプスマップ生成のステップでは, 各視点のデプスマップに含まれる誤対応点を完全に排除することが難しいためである. 一方で, ひとつの視点のデプスマップにおいて誤対応点となる領域でも, 同一の領域を復元する別の近傍視点では, 誤対応点とならない可能性がある. そこで, ある参照視点のデプスマップについて, カメラ位置に近い近傍視点のデプスマップとの整合性を取ることで, 参照視点における誤対応点を除去する. 提案手法におけるデプスマップ生成は, 階層的探索を適用しているため, 誤対応点がデプスマップ上でクラスタ状に発生しやすい. 本小節では, このようなクラスタ状に発生した誤対応点について, 近傍視点との整合性を取るために, グラフカットに基づく誤対応点除去を提案する. グラフカットは, S ノードと T ノードを含む重み付き有向グラフについて, S-T カット (カット上を S から T に向かうエッジの重みが最小になるカット) を求めることで 2 値のラベリングを行う手法である [90]. 有向グラフの重みに, エネルギー関数の各項を割り当てることで, エネ

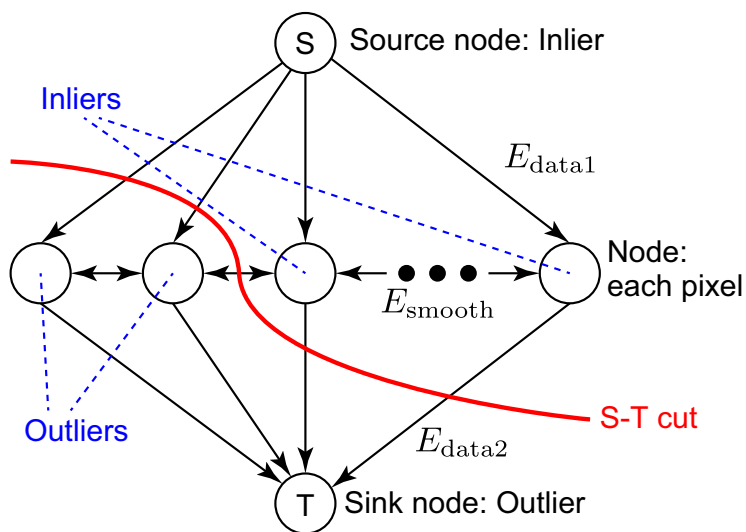


図 4.3 誤対応点除去に用いるグラフ

ルギー最小化問題を解くことができる．画像処理に適用する場合，近傍ピクセルとの連結関係を考慮した 2 値のラベリングに使用される．

以下では，ある参照視点 $V_R \in V$ について，デプスマップ $I_Z(u, v)$ 上の各ピクセル $m = (u, v)$ が正対応点（正しい 3 次元復元点）か誤対応点かの 2 値のラベリング問題をグラフカットで解く．まず，カットするグラフは図 4.3 のように定義する．各ピクセル m をノードとし，S ノードからすべてのピクセルのノードに向かうエッジ，すべてのピクセルのノードから T ノードへ向かうエッジを追加する．また，隣接する 4 近傍のピクセル（ピクセル (u, v) に対して， $(u - 1, v)$ ， $(u, v - 1)$ ， $(u + 1, v)$ ， $(u, v + 1)$ の 4 つのピクセル）間について対応するノード間に双方向のエッジを追加する．このとき，S ノードを正対応（Inlier）のラベル，T ノードを誤対応（Outlier）のラベルとする．また，すべてのピクセルのノードを表す集合を N ，すべてのピクセル間のエッジを表す集合を E とする．次に，各ノードに重みを追加する．S ノードから各ピクセルのノードに向かうエッジの重みを $E_{data1}(m)$ ，各ピクセルのノードから T ノードに向かうエッジの重みを $E_{data2}(m)$ ，隣接ピクセル $m_1 - m_2$ の関係を表すエッジ

の重みを $E_{\text{smooth}}(\mathbf{m}_1, \mathbf{m}_2)$ とする．このとき，このグラフカットは，

$$\begin{aligned}
 E_{\text{all}}(\mathbf{N}, \mathbf{E}) = & \sum_{\mathbf{m} \in \mathbf{N}} E_{\text{data1}}(\mathbf{m}) + \sum_{\mathbf{m} \in \mathbf{N}} E_{\text{data2}}(\mathbf{m}) \\
 & + \sum_{\{\mathbf{m}_1, \mathbf{m}_2\} \in \mathbf{E}} E_{\text{smooth}}(\mathbf{m}_1, \mathbf{m}_2)
 \end{aligned} \quad (4.38)$$

を最小化する問題となる． $E_{\text{data1}}(\mathbf{m})$ は，そのピクセルの正対応点らしさを表し，ピクセル \mathbf{m} が正対応の時に $E_{\text{data1}}(\mathbf{m})$ が大きいほど，S ノードから \mathbf{m} のノードへ向かうエッジが S-T カットに含まれにくくなる．提案手法では，POC 関数のピーク値が高いピクセルは正対応である可能性が高いと仮定し， $E_{\text{data1}}(\mathbf{m})$ を次式で定義する．

$$E_{\text{data1}}(\mathbf{m}) = \begin{cases} \lambda_{E1} I_{\alpha}(\mathbf{m}) & \text{if } \mathbf{m} \text{ is inlier} \\ 0 & \text{otherwise} \end{cases} \quad (4.39)$$

$E_{\text{data2}}(\mathbf{m})$ は，そのピクセルの誤対応点らしさを表し，ピクセル \mathbf{m} が誤対応の時に $E_{\text{data2}}(\mathbf{m})$ が大きいほど， \mathbf{m} のノードから T ノードへ向かうエッジが S-T カットに含まれにくくなる．提案手法では，近傍視点への再投影点における奥行きが大きく異なるピクセルは誤対応である可能性が高いと仮定し， $E_{\text{data2}}(\mathbf{m})$ を次式で定義する．

$$E_{\text{data2}}(\mathbf{m}) = \begin{cases} \lambda_{E2} \sum_{C_k \in C} \left(1 - \exp \left(-\frac{(Z' - I_{Z, C_k}(\mathbf{m}'))^2}{2\sigma_{E1}^2} \right) \right) & \text{if } \mathbf{m} \text{ is outlier} \\ 0 & \text{otherwise} \end{cases} \quad (4.40)$$

ここで， C は V_R のデプスマップ作成時の近傍視点群である．また， Z' は，参照視点 V_R の 3 次元点を近傍視点 C_k の座標系に変換した際の 3 次元点の奥行きを， $I_{Z, C_k}(\mathbf{m}')$ は，近傍視点 C_k のデプスマップにおける \mathbf{m}' ピクセルの奥行きを表し， Z' ， \mathbf{m}' は次式の関係を満たす．

$$\begin{aligned}
 \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} &= I_Z(\mathbf{m}) \mathbf{R}_{C_k} \mathbf{A}^{-1} \tilde{\mathbf{m}} + \mathbf{t}_{C_k} \\
 s\tilde{\mathbf{m}}' &= \mathbf{A}_{C_k} \mathbf{M}'
 \end{aligned} \quad (4.41)$$

ここで， \mathbf{R}_{C_k} は参照視点 V_R のカメラ座標から近傍視点 C_k のカメラ座標への回転を， \mathbf{t}_{C_k} は参照視点 V_R のカメラ座標から近傍視点 C_k のカメラ座標への平行移動量を表す．また， \mathbf{A} および \mathbf{A}_{C_k} は，それぞれ参照視点 V_R および近傍視点 C_k の内部パラメータを表す．この

$E_{\text{data}2}(m)$ は、デプスマップ $I_Z(u, v)$ の近傍視点におけるデプスマップとの整合性をとるための項である。

$E_{\text{smooth}}(m_1, m_2)$ は、隣接ピクセルとのラベルの一致しやすさを表し、 $E_{\text{smooth}}(m_1, m_2)$ が大きいほど、 m_1 - m_2 間のエッジが S-T カットに含まれにくくなる。提案手法では、隣接ピクセルと奥行き之差が大きくなるとして正対応・誤対応のラベルが変化しやすいと仮定し、 $E_{\text{smooth}}(m_1, m_2)$ を次式で定義する。

$$E_{\text{smooth}}(m_1, m_2) = \lambda_{E3} \exp\left(-\frac{(I_Z(m_1) - I_Z(m_2))^2}{2\sigma_{E2}^2}\right) \quad (4.42)$$

式 (4.39)、式 (4.40) および式 (4.42) において $\lambda_{E1}, \lambda_{E2}, \lambda_{E3}$ は各コストの優先度を表し、 σ_{E1}, σ_{E2} はパラメータである。以上により定義されたグラフについて、グラフカットにより S-T カットを求め、S ノードと同じラベルのピクセルを正対応点、T ノードと同じラベルのピクセルを誤対応点とする。このように決定された誤対応点について、デプスマップ $I_Z(u, v)$ および相関値マップ $I_\alpha(u, v)$ の誤対応点のピクセルの値を 0 とする。上記の処理を、多視点画像 V に含まれるすべての視点のデプスマップに対して行う。

4.2.4 閾値処理に基づくアーチファクト除去

アーチファクトとは、MVS アルゴリズムにより復元される現実には存在しない物体表面上の構造物であり、参照視点において物体境界・オクルージョン境界となる領域で復元される傾向がある。MVS では、マッチングウィンドウという小領域ごとに 3 次元点の奥行きを推定するため、物体境界で背景となるピクセルについても、ウィンドウ内に含まれる前景の奥行きを推定し、アーチファクトとして復元される。4.2.3 節で対象とした誤対応点と違い、本小節で対象とするアーチファクトは、カメラ位置の近い近傍視点のデプスマップ上で差が現れにくく、4.2.3 節の誤対応点除去では除去されない。そのため、カメラ位置の大きく異なる視点のデプスマップとの整合性を取る必要がある。一方で、カメラ位置が大きく異なる視点では、対象物体における復元領域も変わりやすいため、デプスマップ同士に差がある場合に、それがアーチファクトによるものなのか、異なる領域を 3 次元復元したものなのか、判断するのが難しい。そこで、提案手法では、アーチファクトを含む複数枚のデプスマップについて、デプ

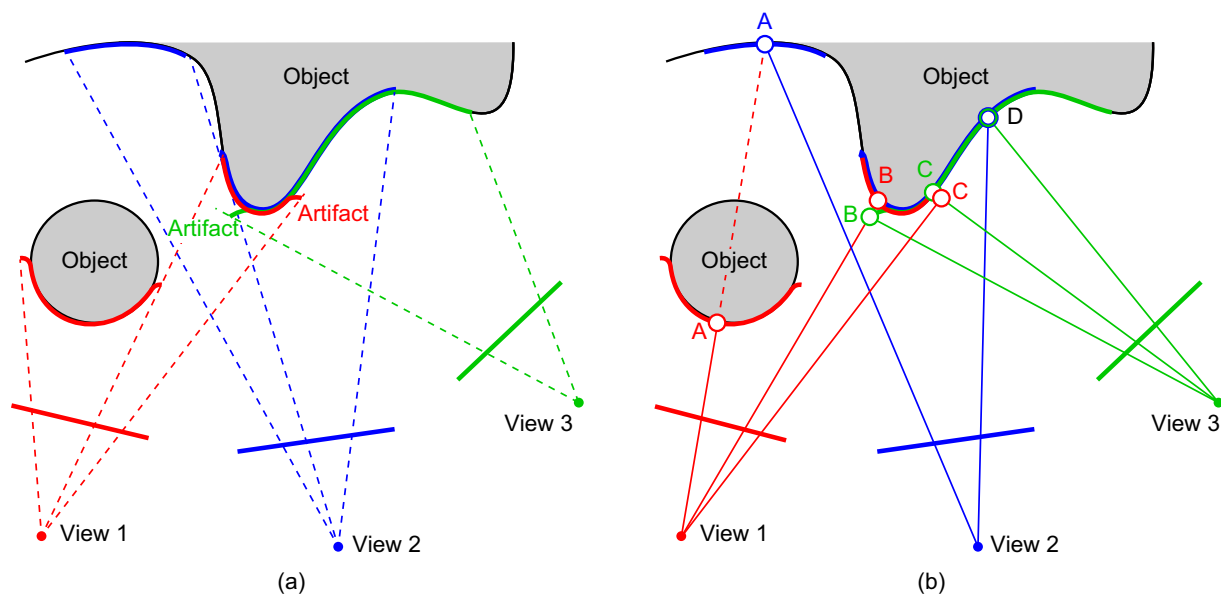


図 4.4 デプスマップの整合性: (a) 各視点から 3 次元復元される物体表面とアーチファクト (b) デプスマップ間の整合性と 3 次元復元点の関係 (A: 物体の異なる領域を復元した正対応点同士でデプスマップに差が生じる場合, B, C: アーチファクトによりデプスマップに差が生じる場合, D: 同一の領域を復元した正対応点でデプスマップの整合性が取れる場合)

スマップ同士の整合性が取れなくなる状態を場合分けし, アーチファクトを閾値処理 [27] によって除去する.

図 4.4 にアーチファクトが発生した場合の複数のデプスマップの関係の例を示す. 図 4.4(a) では, 視点 1 と視点 3 のデプスマップにおいて, 物体境界・オクルージョン境界でアーチファクトが発生している. このとき, 視点 1 から物体境界・オクルージョン境界となる領域でも, 別の視点からは境界にならない. また, アーチファクトは物体の外側に発生しやすく, ある視点から復元されたアーチファクトは, 別の視点のデプスマップにおける奥行きよりも手前側に復元される. 図 4.4(b) は, デプスマップ間の整合性と 3 次元点の関係を表す. A から D の各位置について, デプスマップ間の整合性と 3 次元点の関係を以下にまとめる.

- A の位置では, 視点 1 と 視点 2 のデプスマップで整合性が取れないが, どちらの視点

から復元される 3 次元点も正対応点である．これは，カメラ位置が大きく異なる視点から，対象物体の異なる領域を復元した状態である．このとき，視点 1 と視点 2 のデプスマップの差は比較的大きくなる．

- B, C の位置では，視点 1 と視点 3 のデプスマップ間で整合性が取れず，B の位置では視点 3 から復元される 3 次元点が，C の位置では視点 1 から復元される 3 次元点がアーチファクトとなる．このとき，視点 1 と視点 3 のデプスマップの差は比較的小さく，常にアーチファクトが正対応点よりもカメラ側に復元される．
- D の位置では，視点 2 と視点 3 のデプスマップの差が少なく，整合性が取れている．このことより，D の位置では，視点 2 と視点 3 で対象物体の同一の領域を復元したと判断できる．

以上の特性から，閾値処理により，参照視点 V_R のデプスマップ $I_Z(u, v)$ について，次式の条件を満たすピクセル m をアーチファクトとする．

$$th_1 < \frac{I_{Z,C_k}(m') - Z'}{I_{Z,C_k}(m')} < th_2 \quad (4.43)$$

ここで， C_k は，次式の条件を満たす視点である．

$$\deg(V_R, C_k) > th_3 \quad (4.44)$$

また， Z' は，参照視点 V_R の 3 次元点を近傍視点 C_k の座標系に変換した際の 3 次元点の奥行きを， $I_{Z,C_k}(m')$ は，近傍視点 C_k のデプスマップにおける m' ピクセルの奥行きを， $\deg(V_R, C_k)$ は，視点 V_R と C_k の視差角を表し， Z' ， m' は式 (4.41) を満たす． th_1 ， th_2 ， th_3 はパラメータである．図 4.4 の A, B の位置における，各変数の関係を図 4.5 に示す． C_k の候補が複数存在する場合，すべての候補について式 (4.43) を評価し，1 つ以上の候補で条件を満たす場合はアーチファクトとする．式 (4.43) および式 (4.44) の条件式の意味は以下の通りである．

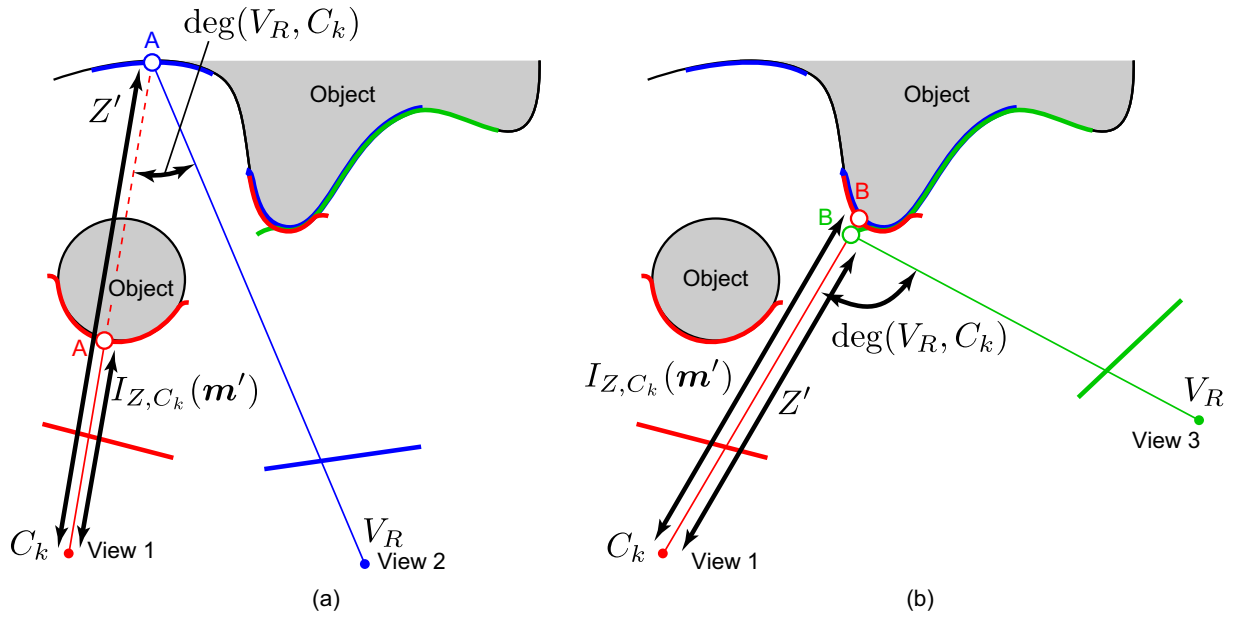


図 4.5 閾値処理によるアーチファクト除去における各変数の関係: (a) A の位置において視点 2 のデプスマップから復元された点は正対応点として判断され, (b) B の位置において視点 3 のデプスマップから復元された点はアーチファクトとして判断される .

- 奥行き之差が th_1 以下の点は, 同一の領域を復元した点であるとして除去しない .
- 奥行き之差が th_2 以上の点は, 対象の別の部分を復元した点として除去しない ,
- 視野角が th_3 以下の視点のデプスマップは, アーチファクトの判定に適さないため, 使用しない .

デプスマップ $I_Z(u, v)$ 上のすべてのピクセルについて, 式 (4.43) および式 (4.44) を評価し, 条件を満たすピクセルについてデプスマップ $I_Z(u, v)$ および相関値マップ $I_\alpha(u, v)$ を 0 とすることで, アーチファクトを除去する . 上記の処理を, 多視点画像 V におけるすべての視点 V のデプスマップに対して行う .

4.2.5 デプスマップからの3次元点群生成

4.2.3 節の誤対応点除去, および, 4.2.4 節のアーチファクト除去を適用した後のデプスマップ群とカメラパラメータから, 対象物体の3次元点群を生成する. 多視点画像 V について, 視点 $V_R \in V$ のデプスマップを $I_Z(u, v)$, 内部パラメータを A , 外部パラメータを R, t とする. このとき, デプスマップ上のピクセル m の3次元点 M の世界座標は次式で計算される.

$$M = R^{-1}(I_Z(m)A^{-1}\tilde{m} - t) \quad (4.45)$$

ただし, $I_Z(u, v)$ および $I_\alpha(u, v)$ の値が0のピクセルについては, 3次元点の奥行きが推定されていない, もしくは, 誤対応点・アーチファクトとして除去されたとして, 3次元点を復元しない. 視点 V_R のデプスマップ $I_Z(u, v)$ におけるすべてのピクセル m について, 式 (4.45) により3次元点を生成する. この処理を, 多視点画像 V に含まれるすべての視点について行うことで, 対象物体全体の3次元点群を生成する. 各視点から生成された3次元点群の座標系は世界座標系に統一されているため, 後処理としての点群の位置合わせを必要としない.

4.3 精度評価実験

本章では, Jensen らの公開データセットを用いた実験により, 従来のMVSアルゴリズムと3次元復元精度を比較し, 提案手法の有効性を確認する. まず, 実験環境として, Jensen らの公開データセット, および, 比較する従来手法について説明する. 次に, 提案手法と従来手法のそれぞれで生成した3次元点群について, その誤差を求め, 3次元復元精度を評価する.

4.3.1 実験環境

本実験では, MVS 評価用のデータセットとして, Jensen らのデータセット [41], [43] 用いる. Jensen らのデータセットは, 3次元点群の評価を行うデータセットであり, MVS 評価用のデータセットの中でも, データセットの規模が大きく, 128 個の対象物体に対して, それぞれの多視点画像と真値の3次元点群が利用できる. また, 各対象物体について, いくつかの

MVS アルゴリズムによる 3 次元復元点群も提供されている。対象物体は、建物の模型、商品パッケージ、野菜、建築素材、動物の置物の 5 つの分類について、さまざまな特性の物体がある。データセットには、対象物体ごとに、49 視点、もしくは、64 視点の多視点画像が含まれる。これらの多視点画像は、ロボットアームの先端に固定されたデジタルカメラにより撮影されたものである。各視点の画像は、 $1,600 \times 1,200$ ピクセルのカラー画像である。ロボットアームの動きを記録し、対象物体を変えながら、同一の 49 視点、もしくは、64 視点の多視点画像を撮影する。49 視点の多視点画像は、対象物体を設置する撮影ステージに対して、半径約 500 mm の球状の視点位置から撮影されたものである。64 視点の多視点画像は、49 視点の視点位置に加え、半径約 650 mm の球状の視点位置から撮影された画像を含む。さらに、各視点について、8 種類の光源設定の画像が選択できる。これら 8 種類の光源設定は、撮影ステージ上部に取り付けられた 18 個の LED 照明の一部を点けたり消したりすることで制御され、対象物体と照明の位置関係は視点ごとに変化しない^{*1}。各視点のカメラパラメータは、既知パターンによる校正により推定されたもので、Jensen らの提供する内部・外部パラメータが利用できる。データセットに含まれる多視点画像のカメラパラメータは、対象物体ごとに変化しない。本実験では、128 個の対象物体の中から、40 個の物体を選択し、実験に用いる。実験に使用する対象物体と、Jensen らのデータセットにおける物体番号を図 4.6、図 4.7 に示す。これらの 40 個の対象物体は、すべての従来手法の結果が提供されている物体の中から、Jensen らの 5 つの分類で偏りが生じないように選択した。

比較する従来手法として、Campbell らの手法 [24]、Furukawa らの手法 [19]、Tola らの手法 [26] を用いる。いずれの手法も、Middlebury のランキングサイト [42] や、Strecha のランキングサイト [49] にて高い性能を示す MVS アルゴリズムである。従来手法の精度評価には、Jensen らのデータセット内に含まれる 3 次元復元点群を利用する。Campbell らの手法は、アルゴリズム内のウィンドウマッチングに 3.5.2 節の分類における NCC+BF を用いる。Furukawa らの手法は、アルゴリズム内のウィンドウマッチングに 3.5.2 節の分類における NCC+LM+Homography を用いる。Tola らの手法は、DAISY に基づくアルゴリズムで

^{*1} ロボットアームが視点ごとに変動しているため、ロボットアームの影により、視点ごとの光源の状態は、完全に統一されていない。



図 4.6 実験に使用する対象物体 (1 から 20 個目)

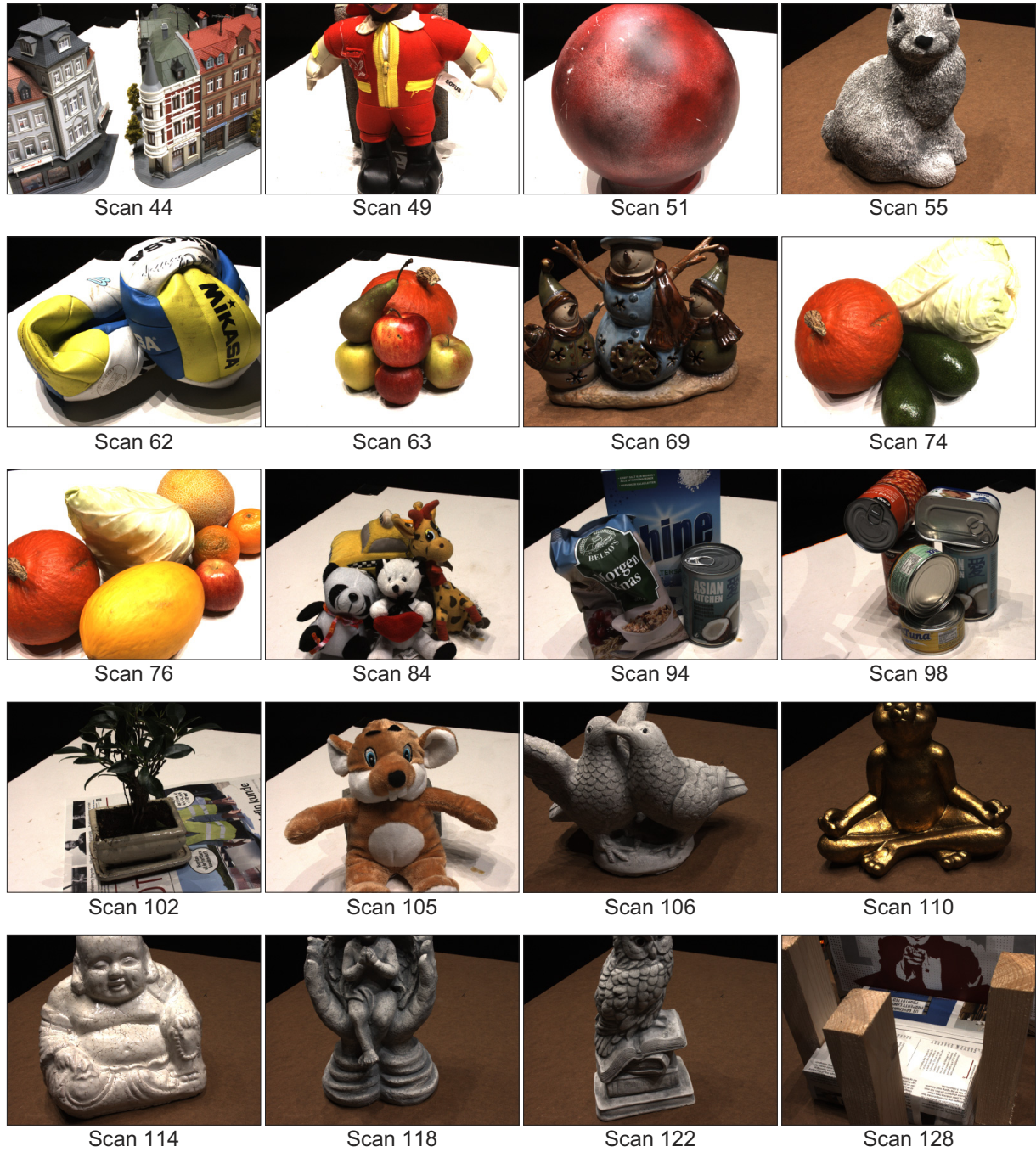


図 4.7 実験に使用する対象物体 (21 から 40 個目)

ある．

Jensen らのデータセットには，対象物体によって，49 枚の多視点画像と 64 枚の多視点画像のデータセットがあるが，提案手法については 64 枚のデータセットについても，はじめの 49 枚のみを用いる．8 種類の光源設定については，対象物体ごとに，ハレーションや黒つぶれの領域が最も少なくなる光源設定を手動で選択する．対象物体の 1 つのデータセット内について，光源設定は，全視点の画像で統一する．このように選択した 49 枚の画像に対して，提案手法により，すべての視点のデプスマップを生成し，3 次元点群を復元する．提案手法は，次に示す 3 つのバリエーションについて実行する．

- ・ Sakai1: MVS のための POC に基づくウィンドウマッチングによる奥行き推定に対して，階層的探索を用いたデプスマップ生成 (4.2.1 節) と重み付きメディアンフィルタを用いたデプスマップの高精度化 (4.2.2 節) のみを適用する．
- ・ Sakai2: Sakai1 に加えて，グラフカットに基づく誤対応点除去を (4.2.3 節) を使用する．
- ・ Sakai3: Sakai2 に加えて，閾値処理に基づくアーチファクト除去を (4.2.4 節) を使用する．

このとき，各対象物体の各視点の画像について，ハレーションや黒つぶれ領域を除去するために，マスク画像を生成し，マスク外の領域についてはデプスマップの値を求めないようにする．本実験において，マスク画像は，原画像を 0 から 255 の輝度値を持つグレースケール画像に変換し，輝度値が 5 以下，または，250 以上のピクセルを除外するように作成した．以上のような条件について，すべての対象物体について 3 次元復元を行い，その精度を従来手法と比較する．

提案手法のパラメータは以下の通りである．まず，Sakai1, Sakai2, Sakai3 について，デプスマップ生成の際に，参照視点 V_R ごとに，最も基線長の短いものから 4 つの視点を近傍視点群 C とする．実際に，デプスマップ生成に用いた参照視点 V_R および近傍視点群 C の例を図 4.8 に示す．階層数は， $H = 3$ である．ウィンドウサイズは，階層番号 $h > 0$ のときに $N \times L = 32 \times 17$ ピクセル，最下位層 ($h = 0$) のときに最小で $N_{\min} \times L_{\min} = 8 \times 5$ ピクセルとする．複数のステレオペアの POC 関数を平均化する際の閾値は $th = 0.5$ とする．重

み付きメディアンフィルタについて、フィルタのウィンドウサイズは 11×11 ピクセルとし、パラメータは $\sigma_{w1} = 2$ ピクセル、 $\sigma_{w2} = 15$ とする。ただし、 σ_{w2} は、カメラ画像の輝度値が 0 から 255 の場合の値である。次に、Sakai2, Sakai3 について、グラフカットに基づく誤対応点除去のパラメータは、 $\sigma_{E1} = \sigma_{E2} = 15$ mm、 $\lambda_{E1} : \lambda_{E2} : \lambda_{E3} = 1 : 1 : 10$ とする。最後に、Sakai3 について、閾値処理に基づくアーチファクト除去のパラメータは、 $th_1 = 0.3\%$ 、 $th_2 = 4\%$ 、 $th_3 = 30$ 度とする。

4.3.2 3次元復元精度の評価

本実験では、3次元復元精度の評価に、Jensen らのデータセットに含まれる精度評価用プログラムを利用する。Jensen らの精度評価用プログラムでは、MVS アルゴリズムで復元した 3次元点群 P_{data} と真値の 3次元点群 P_{truth} を入力とし、2つの距離集合 D_{accuracy} と D_{complete} を出力する。 D_{accuracy} は、3次元復元点群 P_{data} の各 3次元点から最近傍の真値の 3次元点までの距離集合であり、文献 [41] における“accuracy”に相当する。 D_{complete} は、真値の 3次元点群 P_{truth} の各 3次元点から最近傍の 3次元復元点までの距離集合であり、文献 [41] における“complete”に相当する。 D_{accuracy} および D_{complete} の関係を図 4.9 に示す。本実験では、これら 2つの距離集合の中央値 $\text{median}(D_{\text{accuracy}})$ 、および、 $\text{median}(D_{\text{complete}})$ を 3次元復元精度の指標として使用する。 $\text{median}(D_{\text{accuracy}})$ は、復元した 3次元点群がどれくらい高精度であるかを表す指標である。ただし、復元できなかった領域があっても影響を受けないため、3次元復元点数が少ないと小さくなる傾向にある。 $\text{median}(D_{\text{complete}})$ は、対象物体のどれくらい多くの領域を高精度に復元できたかを表す指標である。ただし、誤対応点が増えても影響を受けないため、3次元復元点数が多いと小さくなる傾向にある。図 4.9(d) では、復元された 3次元点が正確であるが、対象物体の詳細な形状を復元できていない。一方で、図 4.9(e) では、復元された 3次元点に不正確な点が含まれるが、対象物体の詳細な形状を復元できている。このように、 $\text{median}(D_{\text{accuracy}})$ 、および、 $\text{median}(D_{\text{complete}})$ のどちらの指標も復元された 3次元点群の精度を表すものであり、両方の指標を考慮する必要がある。そこで、本論文では、“accuracy”と“complete”の複合的な指標として、 $\text{median}(D_{\text{accuracy}} \cup D_{\text{complete}})$ も併せて使用する。

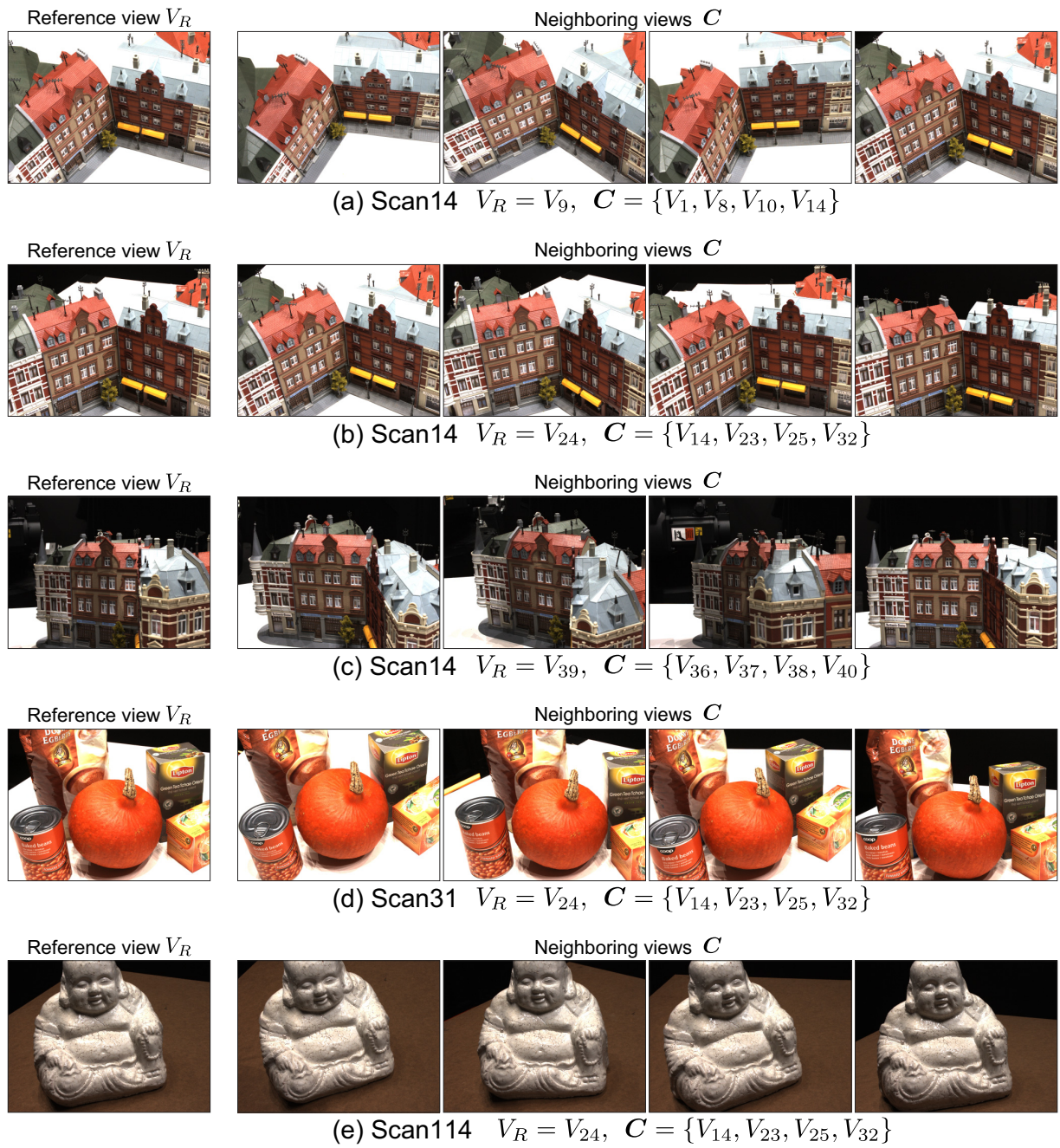


図 4.8 Jensen らのデータセットによる実験で使用する参照視点 V_R と近傍視点群 C の例

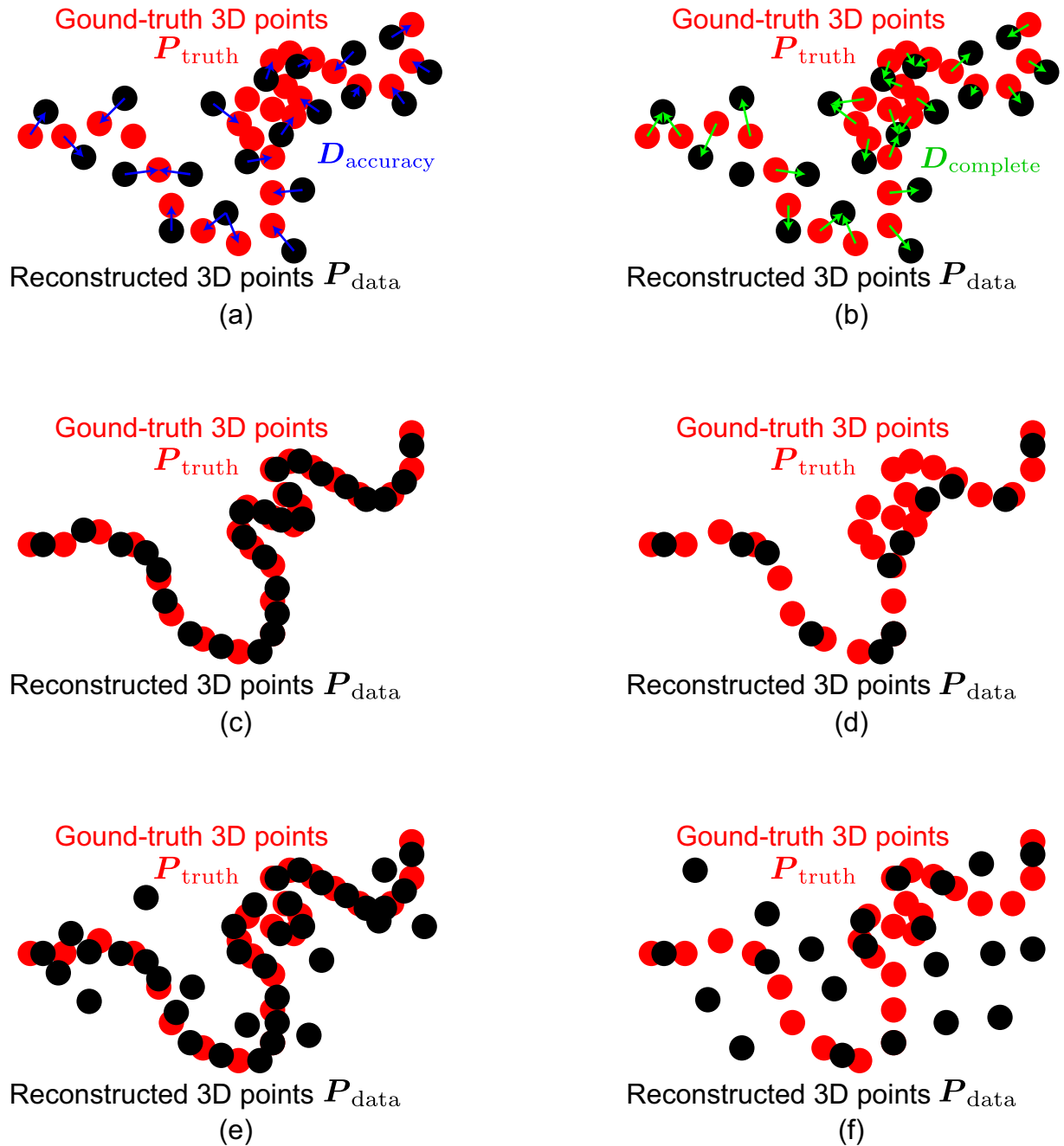


図 4.9 3次元点復元精度の評価指標: (a) 3次元復元点から真値までの距離 $D_{accuracy}$, (b) 真値から3次元復元点までの距離 $D_{complete}$, (c) $median(D_{accuracy})$ と $median(D_{complete})$ の両方が小さい場合, (d) $median(D_{accuracy})$ は小さいが $median(D_{complete})$ が大きい場合, (e) $median(D_{complete})$ は小さいが $median(D_{accuracy})$ が大きい場合, (f) $median(D_{accuracy})$ と $median(D_{complete})$ の両方が大きい場合

それぞれの対象物体について，図 4.10 に $\text{median}(D_{\text{accuracy}})$ の結果を，図 4.11 に $\text{median}(D_{\text{complete}})$ の結果を，図 4.12 に $\text{median}(D_{\text{accuracy}} \cup D_{\text{complete}})$ の結果を示す．ただし，文献 [41] と同様に，それぞれの中央値の計算において，20 mm 以上の誤差を持つ点を誤対応点として除去し，中央値の計算に影響を与えないようにした．まず，従来手法と提案手法の 3 次元復元精度を比較する．実験結果より，提案手法の性能は，Campbell らの手法と同様の傾向を示す．Furukawa らの手法，Tola らの手法と比べると，全体的に $\text{median}(D_{\text{accuracy}})$ が大きく， $\text{median}(D_{\text{complete}})$ が小さい．Sakai1 では，ほぼすべての物体において，どの従来手法よりも $\text{median}(D_{\text{complete}})$ が小さい結果となった．一方， $\text{median}(D_{\text{accuracy}})$ については，Tola らの手法より誤差が大きいものの，すべての物体で Campbell らの手法よりも，いくつかの物体で Furukawa らの手法よりも誤差が小さい．また， $\text{median}(D_{\text{accuracy}})$ と $\text{median}(D_{\text{complete}})$ の両方を考慮した指標である $\text{median}(D_{\text{accuracy}} \cup D_{\text{complete}})$ については，すべての物体において，Sakai3 の結果が，どの従来手法の結果よりも誤差が小さい．

実験に用いた対象物体の中で，いくつかの特徴的な復元結果を図 4.13，図 4.14，図 4.15，図 4.16，図 4.17，図 4.18 に示す．ただし，Tola らの 3 次元復元結果について，Jensen らの提供する 3 次元復元点群に色情報が含まれておらず，本節では，図示のために，カメラ画像と 3 次元点群の関係から適当に色情報を与えた．Scan 34 は，すべての手法について，3 次元復元誤差が最も小さい対象物体である．物体表面のテクスチャが多く，形状も単純な構造である．Scan 11，Scan 31 は，物体表面のテクスチャが少なく，一般に 3 次元復元が難しいとされる物体を含む．特に，Scan 11 は，物体全体でテクスチャが少なく，すべての手法で誤差が大きい．Scan 24，Scan 49，Scan 63 は，提案手法と従来手法の特徴的な差が見られる対象である．3 次元復元結果から確認できる各手法の特徴とその比較は以下の通りである．

- Campbell らの手法は，どの対象物体についても， D_{accuracy} の色付き点群で，全体的に誤差の大きい赤い点が多く，他の手法に比べ不正確な 3 次元点を多く復元していることが確認できる．各対象物体の画素値の色付き点群からも，多くの誤対応点やアーチファクトを含んでいることが確認できる．一方で，Furukawa らの手法，Tola らの手法と比べ，対象物体の多くの領域を復元しており， D_{complete} の誤差が小さい．

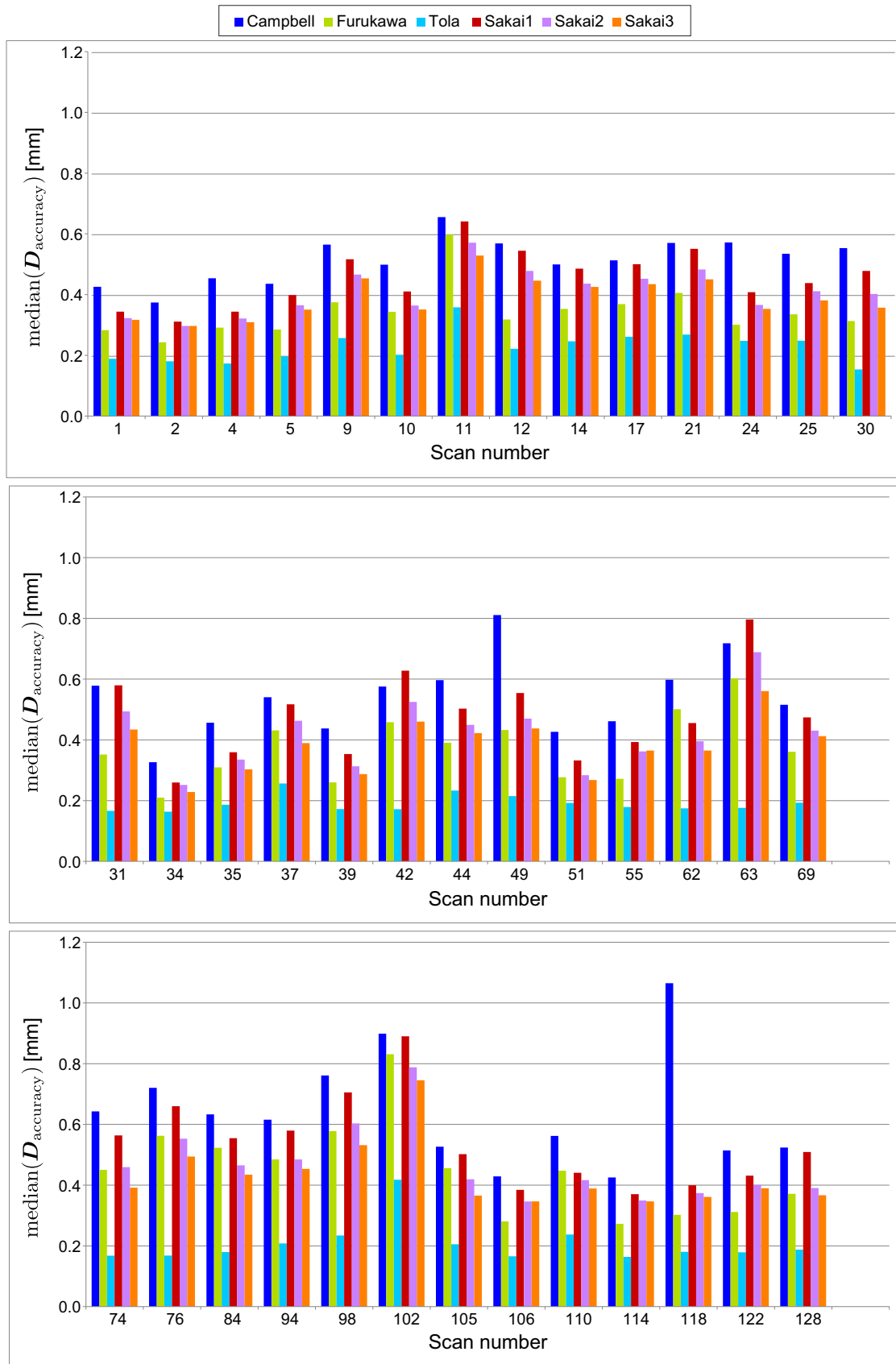


図 4.10 それぞれの対象物体における $\text{median}(D_{\text{accuracy}})$

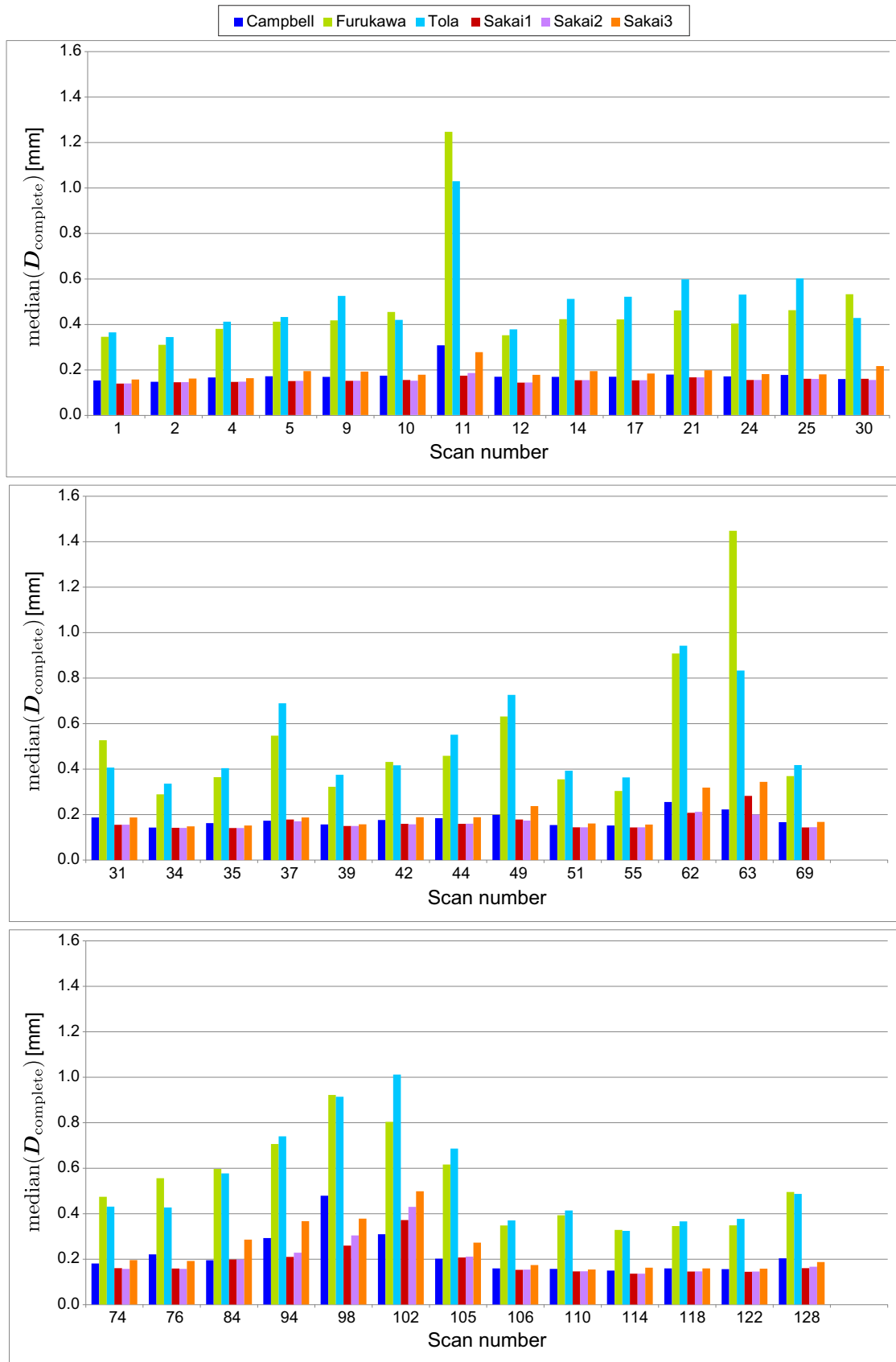


図 4.11 それぞれの対象物体における $\text{median}(D_{complete})$

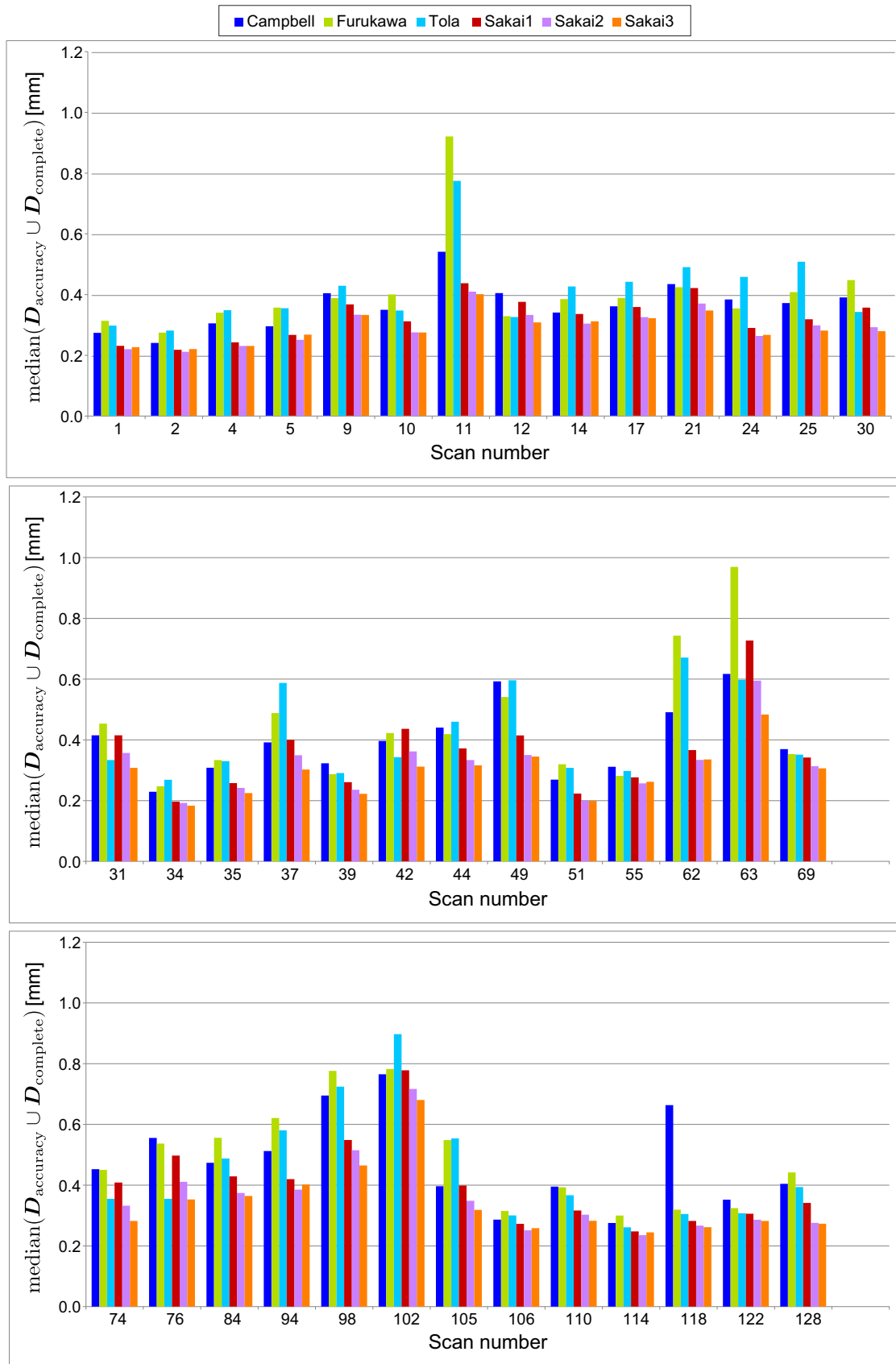


図 4.12 それぞれの対象物体における $\text{median}(D_{\text{accuracy}} \cup D_{\text{complete}})$

- Furukawa らの手法は，テクスチャが多い物体において， $D_{accuracy}$ ， $D_{complete}$ の誤差が小さく，3次元復元点の精度が高い．一方で，テクスチャが少ない Scan 11 や Scan 31 において，復元されていない領域が多く， $D_{complete}$ の誤差が大きい．これは，Furukawa らの手法が，画像上の特徴点を起点として復元を行うため，テクスチャが少ない領域で特徴点が検出されず，3次元点の伝搬による復元もできなかったためである．また，単純な背景においてアーチファクトを復元しやすく，Scan 49，Scan 63 では物体表面に，本来対象物体に存在しない色の3次元点を復元している．
- Tola らの手法は，どの対象物体についても， $D_{accuracy}$ の誤差が小さい一方で，復元されていない領域が多く， $D_{complete}$ の誤差が大きい．これは，Tola らの手法が，信頼値の低い点を除外することで高精度な3次元復元を行う手法であり，テクスチャが少なく信頼値が低い領域は復元できないためである．また，物体の詳細な形状を復元できず，平滑化されたような3次元点群が復元される傾向にある．このような場合，図 4.9(d) に示したように， $D_{complete}$ の誤差が大きくなる．特に，Scan 24 の結果では，窓枠のように形状が細かい領域で $D_{complete}$ の誤差が大きくなっていることが確認できる．これは，基線長の長いステレオペアをマッチングに用いる Tola らの手法において，基線長の長いステレオ画像では，細かい形状を有する領域で画像変形が非常に大きく，かつ，複雑になり，信頼値が大幅に低下するためである．
- Sakai3 は，3つの従来手法よりも多くの領域を復元しており， $D_{accuracy}$ ， $D_{complete}$ の誤差が小さいことを確認できる．図 4.10 において，Sakai3 は，Furukawa らの手法，Tola らの手法に比べて誤差が大きく，Campbell らの手法と同様の傾向を示している．しかし，3次元復元結果から，Sakai3 において $D_{accuracy}$ の誤差が大きい点は，真値の3次元点が存在しない領域に多く存在していることが確認できる．例えば，Scan 24 の屋根の部分や，Scan 49 の靴やくちばし，Scan 63 の果物と床の間などである．その他の領域に関しては，Furukawa らの手法，Tola らの手法と同様に誤差の小さい青い点が多く，Campbell らの手法と比べて $D_{accuracy}$ の誤差が小さいことが確認できる．一方で，どの対象物体においても，物体の多くの領域を復元できており，Furukawa らの手法，Tola らの手法と比べ， $D_{complete}$ の誤差が小さい．特に，Scan 11 や Scan 31 で

は、テクスチャの少ない領域においても高精度な 3 次元点を復元できている。

次に、Sakai1, Sakai2, Sakai3 について、3 次元復元精度を比較する。図 4.10, 図 4.11, 図 4.12 より、Sakai1 と Sakai2 の結果を比較すると、 $\text{median}(D_{\text{complete}})$ にはほとんど差がないが、すべての物体について $\text{median}(D_{\text{accuracy}})$ が小さくなっている。つまり、グラフカットに基づく誤対応点除去により、誤対応点のみを正しく除去できている。一方、Sakai2 と Sakai3 の結果を比較すると、 $\text{median}(D_{\text{accuracy}})$ は小さくなっているが、 $\text{median}(D_{\text{complete}})$ が大きくなっている。つまり、閾値処理に基づく誤対応点除去では、アーチファクトだけでなく、正しく復元できていた 3 次元点も除去している。ただし、 $\text{median}(D_{\text{accuracy}} \cup D_{\text{complete}})$ については、多くの物体で Sakai2 よりも Sakai3 の方が誤差が小さくなっており、3 次元復元点群の精度向上につながっている。

Sakai1, Sakai2, Sakai3 の 3 次元復元結果について、実験に用いた対象物体の中で、いくつかの特徴的な復元結果を図 4.19, 図 4.20, 図 4.21 に示す。色付き 3 次元点群より、Sakai1 の 3 次元復元結果には多くの誤対応点が含まれていることが確認できる。これらの誤対応点は、デプスマップ上では小さい領域であるが、49 視点すべてのデプスマップを単純に統合することで、各デプスマップにおける誤対応点がすべて復元されている。そのため、最終的な 3 次元復元点において目視で確認できるほど多くの誤対応点が復元されている。これに対して、Sakai2 では、Sakai1 の復元結果にみられる誤対応点の多くが除去されており、 D_{accuracy} における誤差の大きい赤い点が少なくなっていることが確認できる。また、Sakai1 において正しく復元されていた正対応点については、除去されていないため、 D_{complete} には差が見られない。一方で、Sakai2 の 3 次元復元結果においても、物体表面付近に小さい誤差を持ったアーチファクトが残っていることが確認できる。例えば、図 4.19 に示した Scan 4 の結果における鳥のとさかや、図 4.21 に示した Scan 63 の結果におけるりんごの表面では、対象物体に存在しない色の点が復元されている。また、図 4.20 に示した Scan 25 の結果では、アーチファクトにより物体表面の輪郭が不正確になり、3 次元復元点群をレンダリングした画像の輪郭にノイズがみられる。これに対して、Sakai3 の 3 次元復元結果では、物体表面のアーチファクトが除去され、3 次元復元点群の見た目が真値に近づいている。 D_{accuracy} について

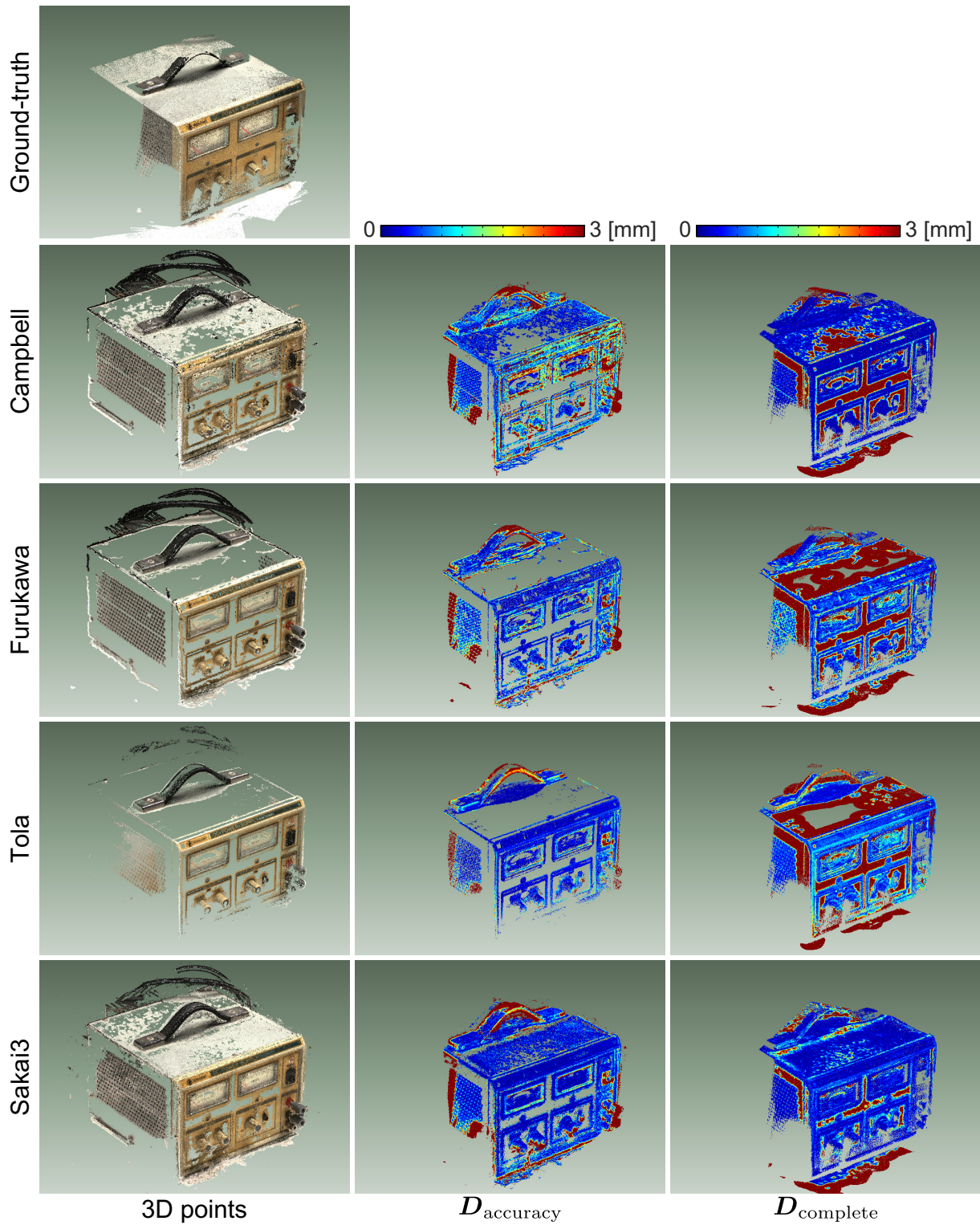


図 4.13 Scan 11 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

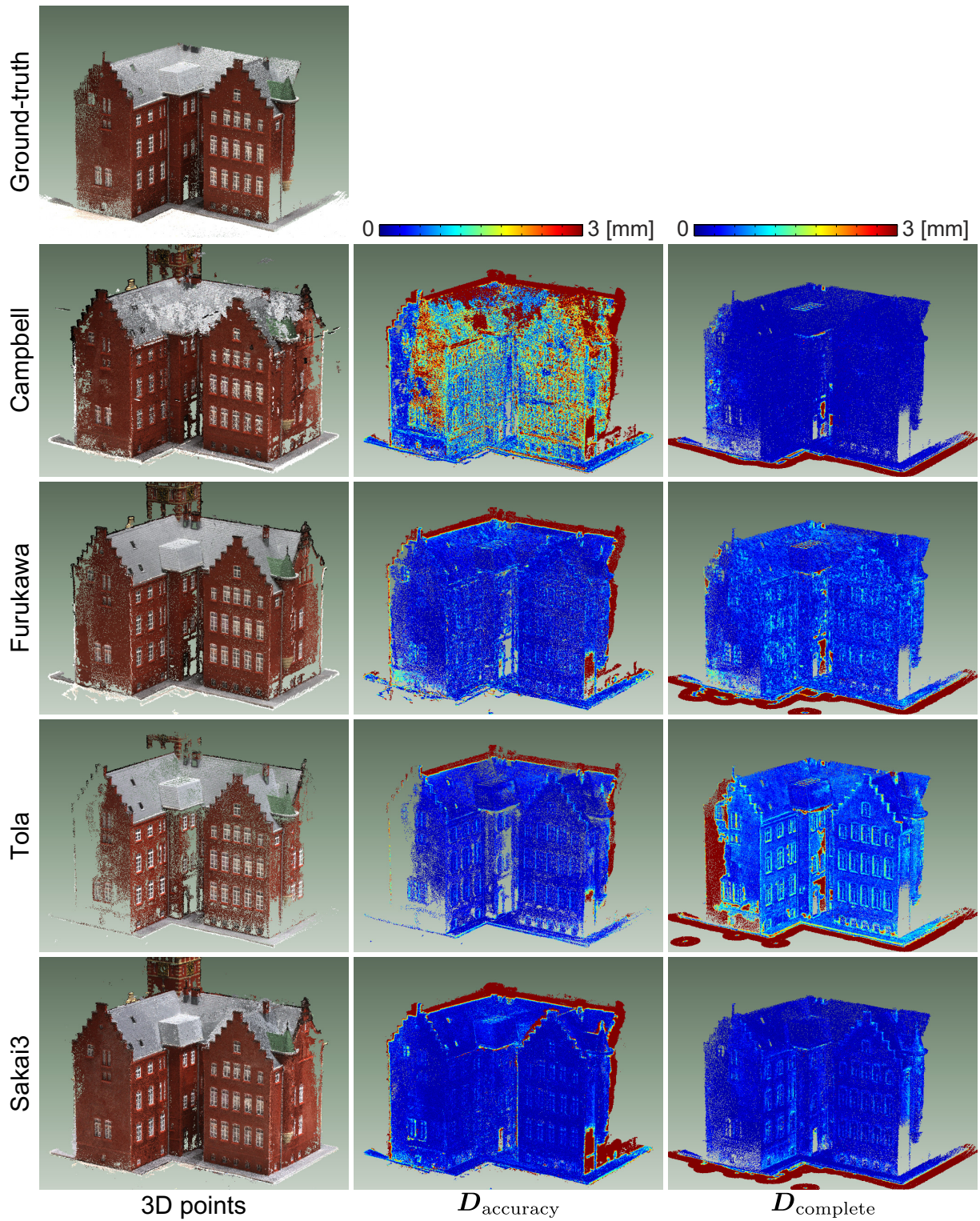


図 4.14 Scan 24 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

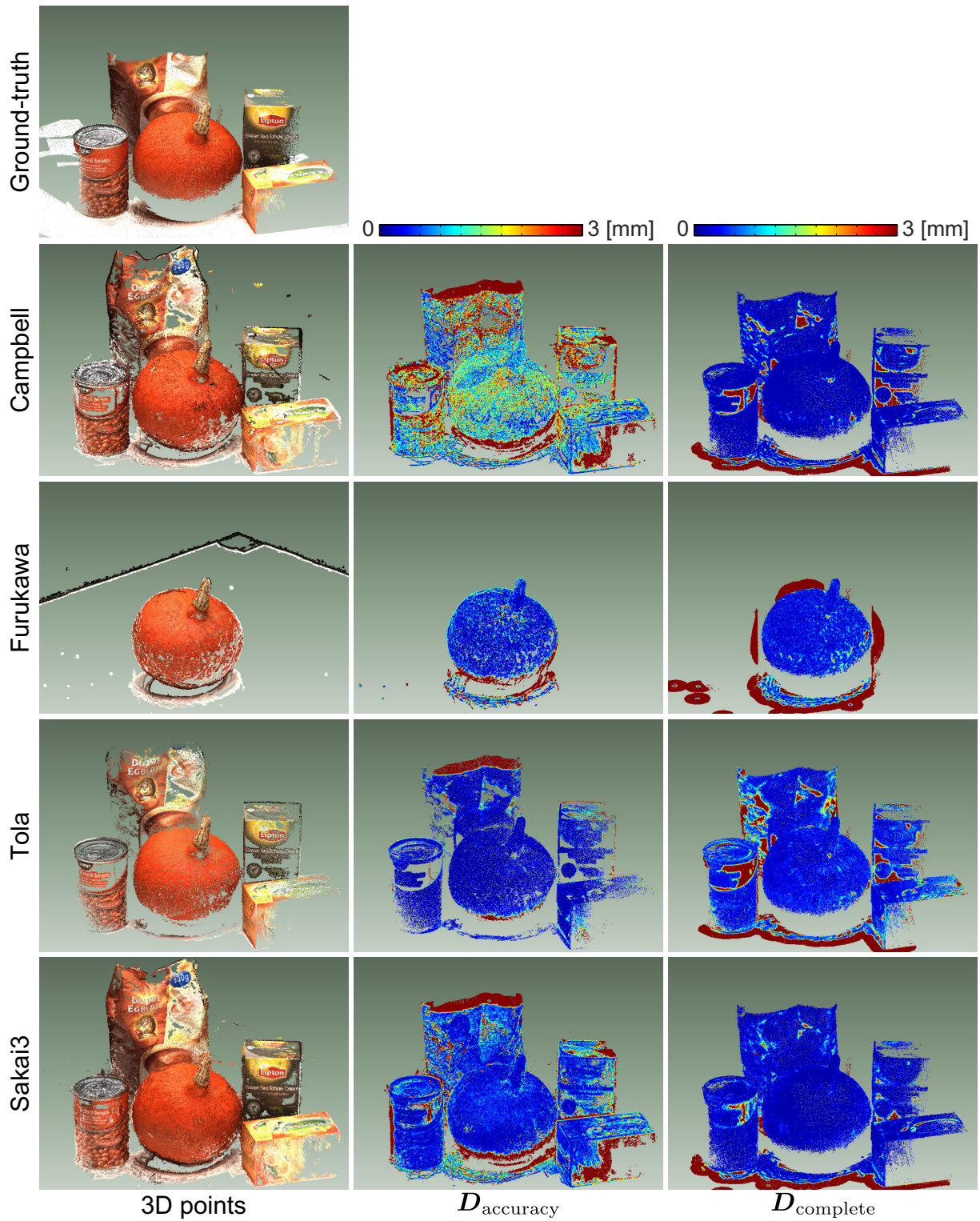


図 4.15 Scan 31 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

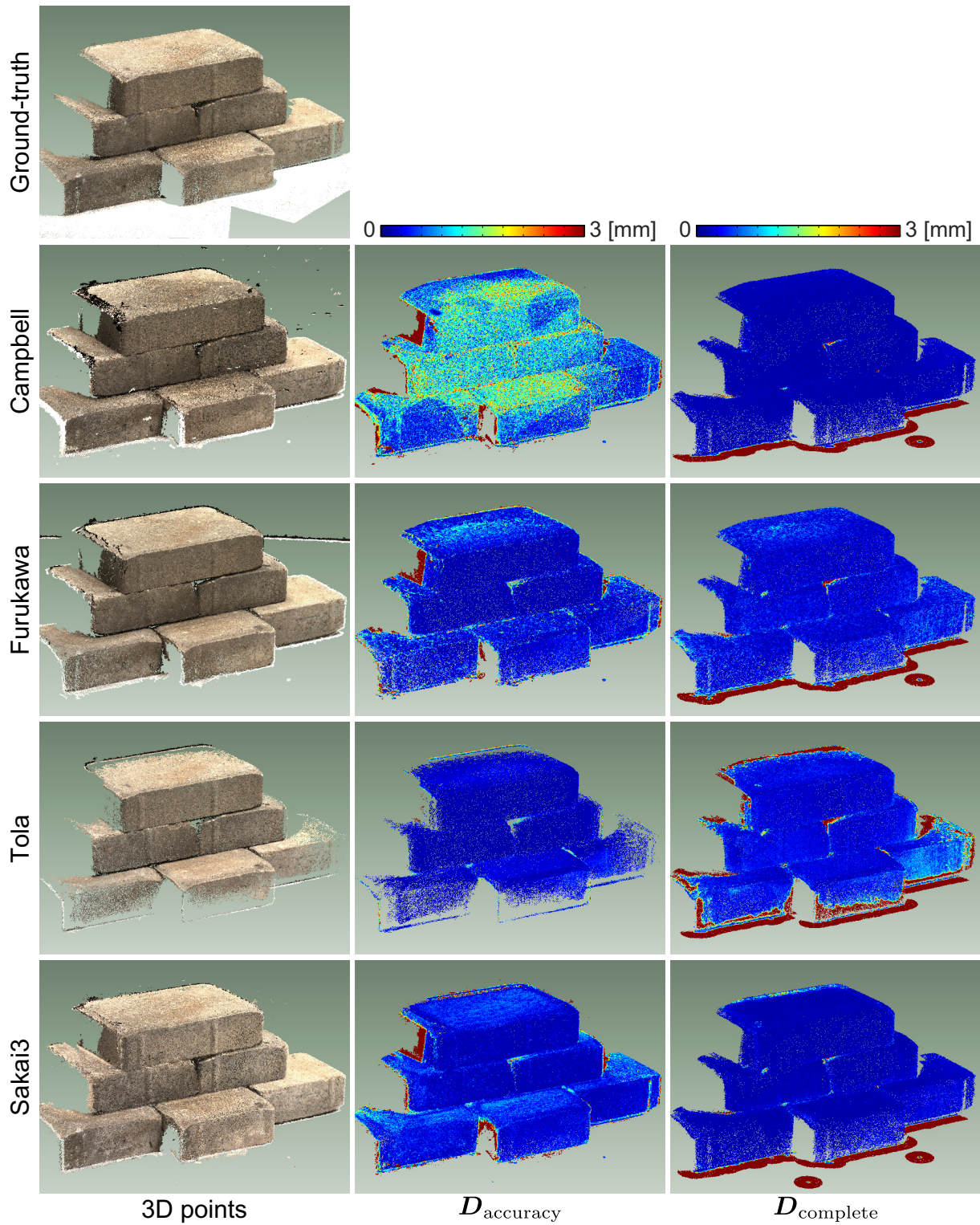


図 4.16 Scan 34 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

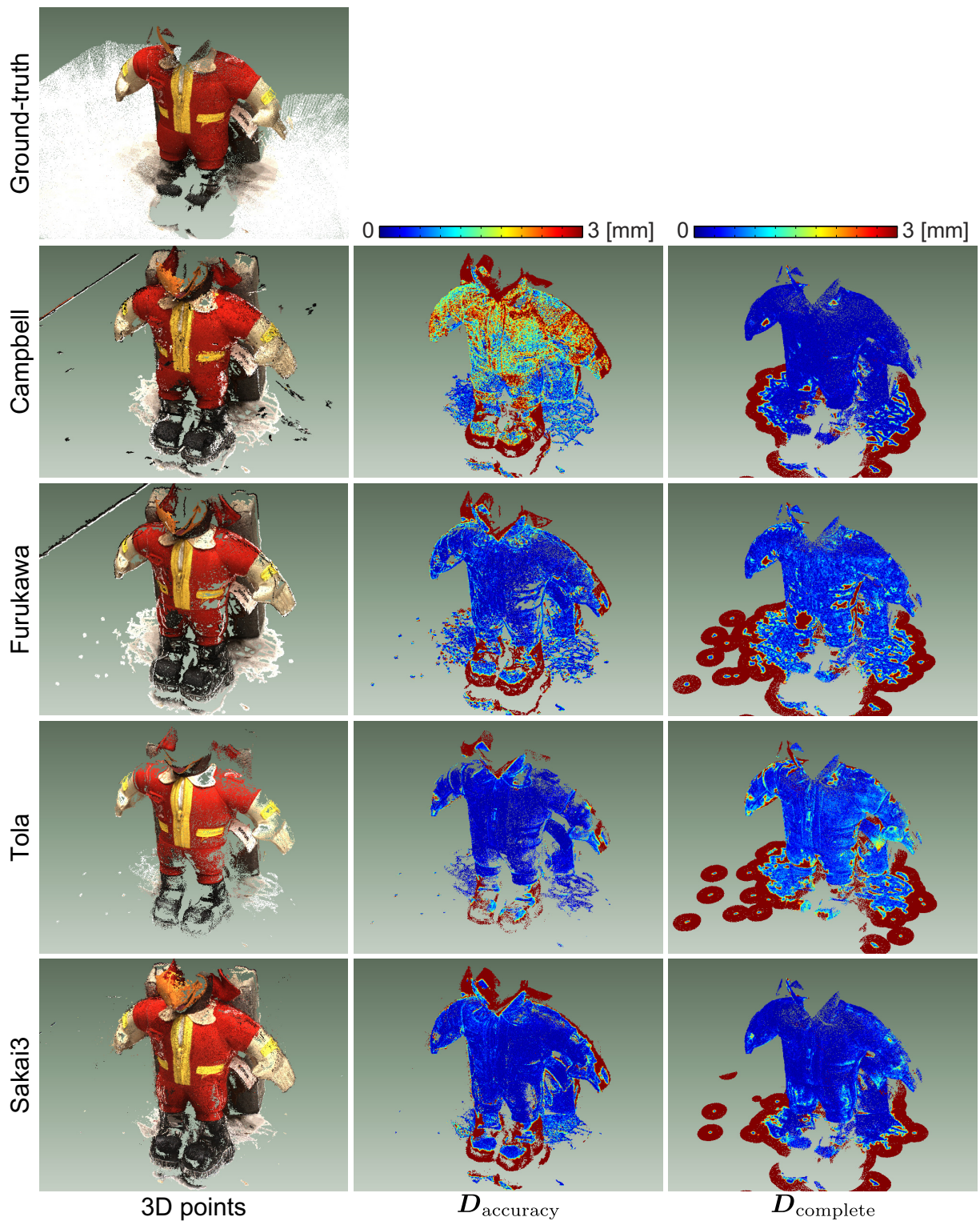


図 4.17 Scan 49 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

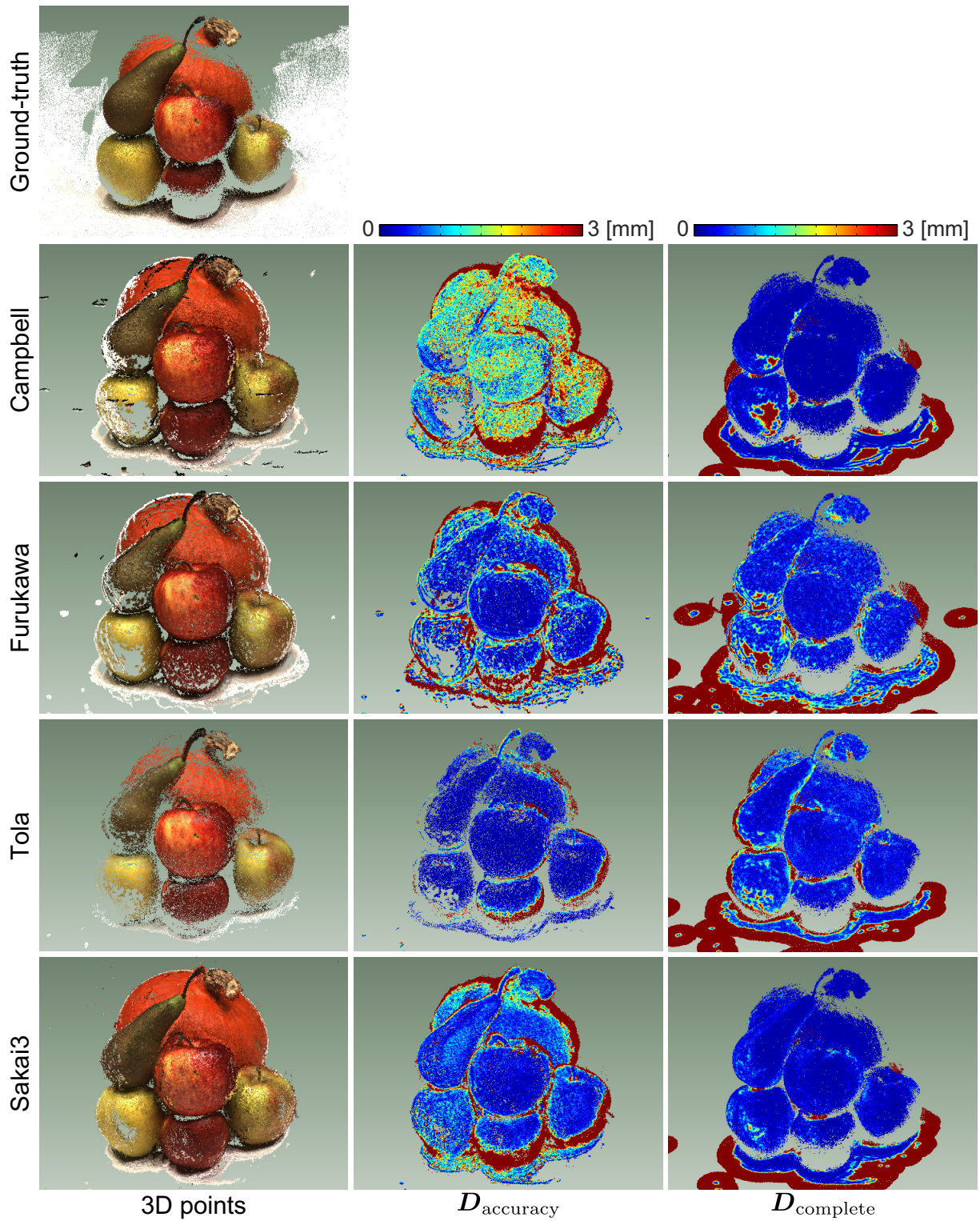


図 4.18 Scan 63 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

も、誤差の大きい赤い点が減少し、青い点の割合が高くなっていることが確認できる。また、図 4.11 では $\text{median}(D_{\text{complete}})$ が Sakai2 に比べ Sakai3 の方が誤差が大きいが、図 4.19、図 4.20、図 4.21 の 3 次元復元結果では、 D_{complete} の変化は局所的である。つまり、物体全体の D_{complete} の誤差が増加したのではなく、物体の一部で D_{complete} の誤差が増加したため、 $\text{median}(D_{\text{complete}})$ が大きくなっている。例えば、Scan 25 におけるバルコニーの柵や、Scan 63 における左のりんごの表面では、アーチファクト除去により正しく復元されていた 3 次元点まで除去され、 D_{complete} の誤差が大きくなっているが、物体全体の D_{complete} の色は大きく変化していない。

以上、本節では、実験を通して提案手法を評価した。Jensen らの公開データセットを用いた実験により、Campbell らの手法、Furukawa らの手法、Tola らの手法と 3 次元復元精度を比較し、 $\text{median}(D_{\text{accuracy}} \cup D_{\text{complete}})$ の指標において、従来手法よりも誤差が小さいことを確認した。また、グラフカットに基づく誤対応点除去、閾値処理に基づくアーチファクト除去の有無を比較し、それぞれの処理の有効性を確認した。

4.4 文化財の 3 次元復元への応用

本節では、本論文で提案する POC に基づく MVS アルゴリズムの応用として、デジタルアーカイブを目的とした文化財の 3 次元復元 [91] について述べる。文化財のデジタルアーカイブは、文化財の記録・修復・研究のため、文化財の画像や 3 次元形状などを計算機上に保存することを目的とする [2]。現在、文化財の 3 次元形状を保存するためには、能動型の 3 次元復元技術が適用され、3 次元スキャナ・3 次元デジタイザと呼ばれる 3 次元計測器が利用されている。しかしながら、それらの 3 次元計測器の多くは、小型の手持ち式のものでも数 kg 以上の重さがあり、かつ、有線であるため、3 次元形状の取得の際に文化財の破損の危険が伴う。また、計測可能な距離や範囲、および、3 次元計測器の大きさから、利用できる環境が限定される。このような背景から、文化財のデジタルアーカイブでは、従来の 3 次元計測器よりも小型で軽量の撮影装置による 3 次元形状の保存技術が望まれている。

そこで、本論文で提案する POC に基づく MVS アルゴリズムにより、カメラ画像のみから

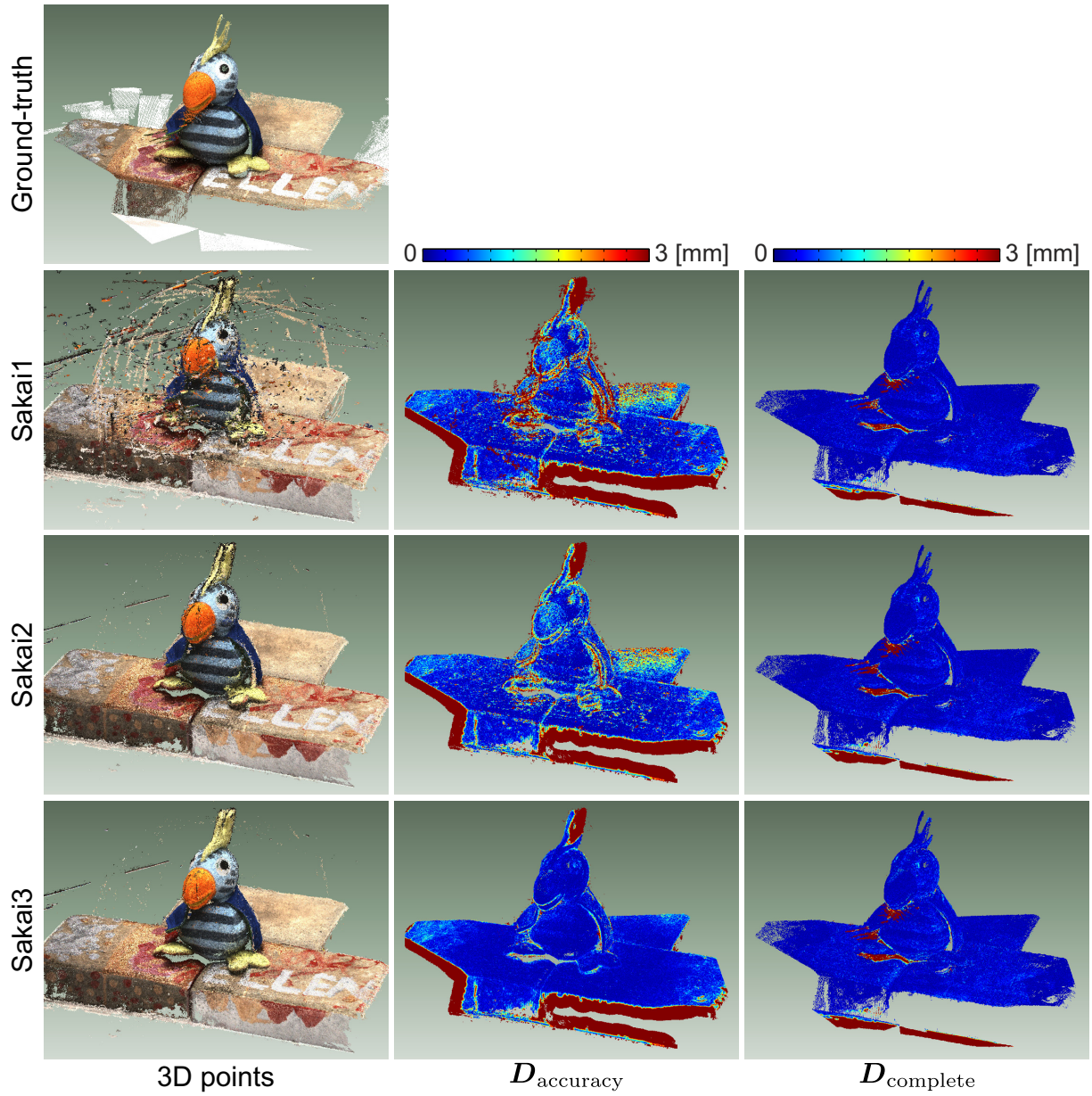


図 4.19 Scan 4 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 3 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

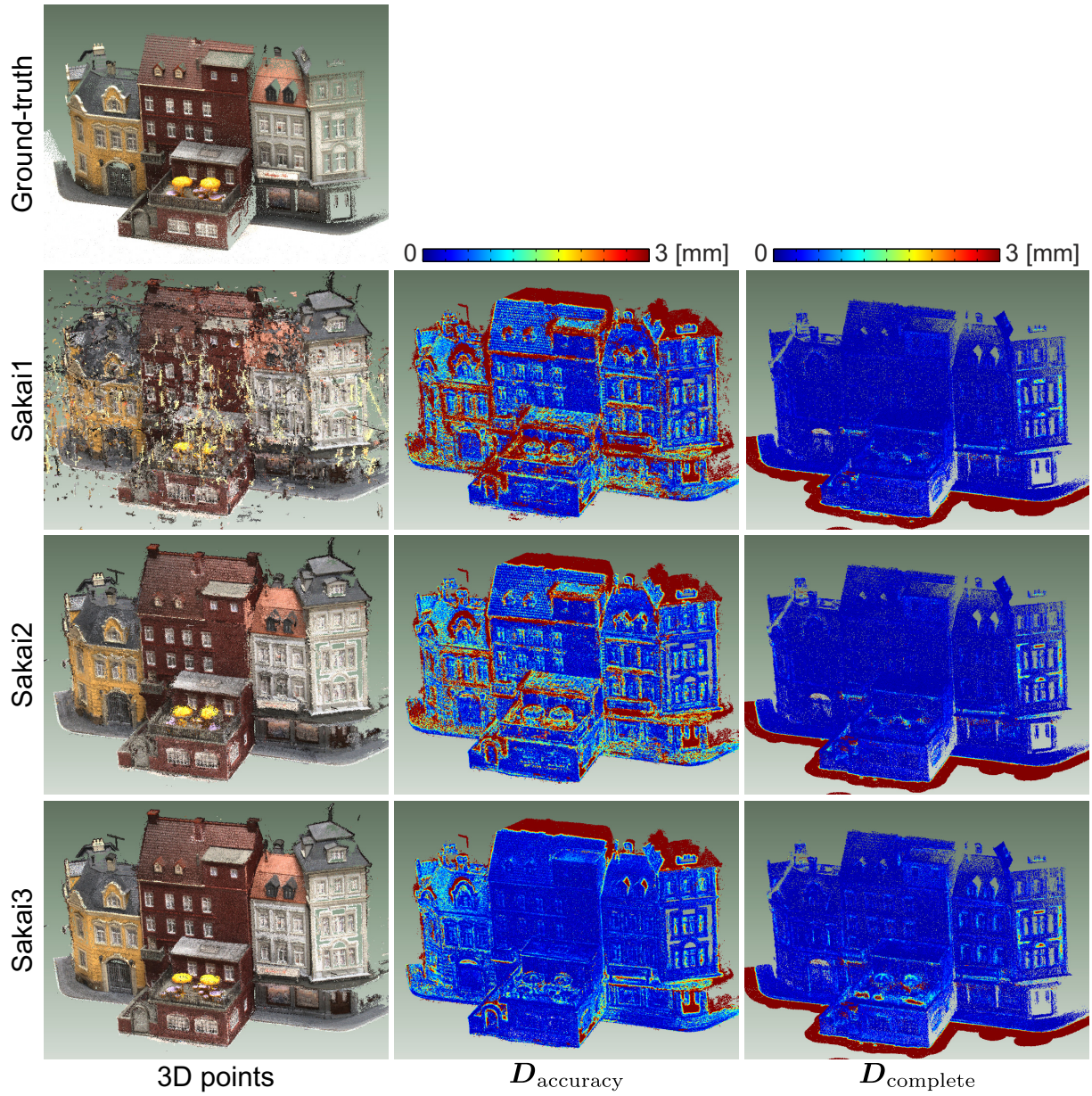


図 4.20 Scan 25 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 2 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

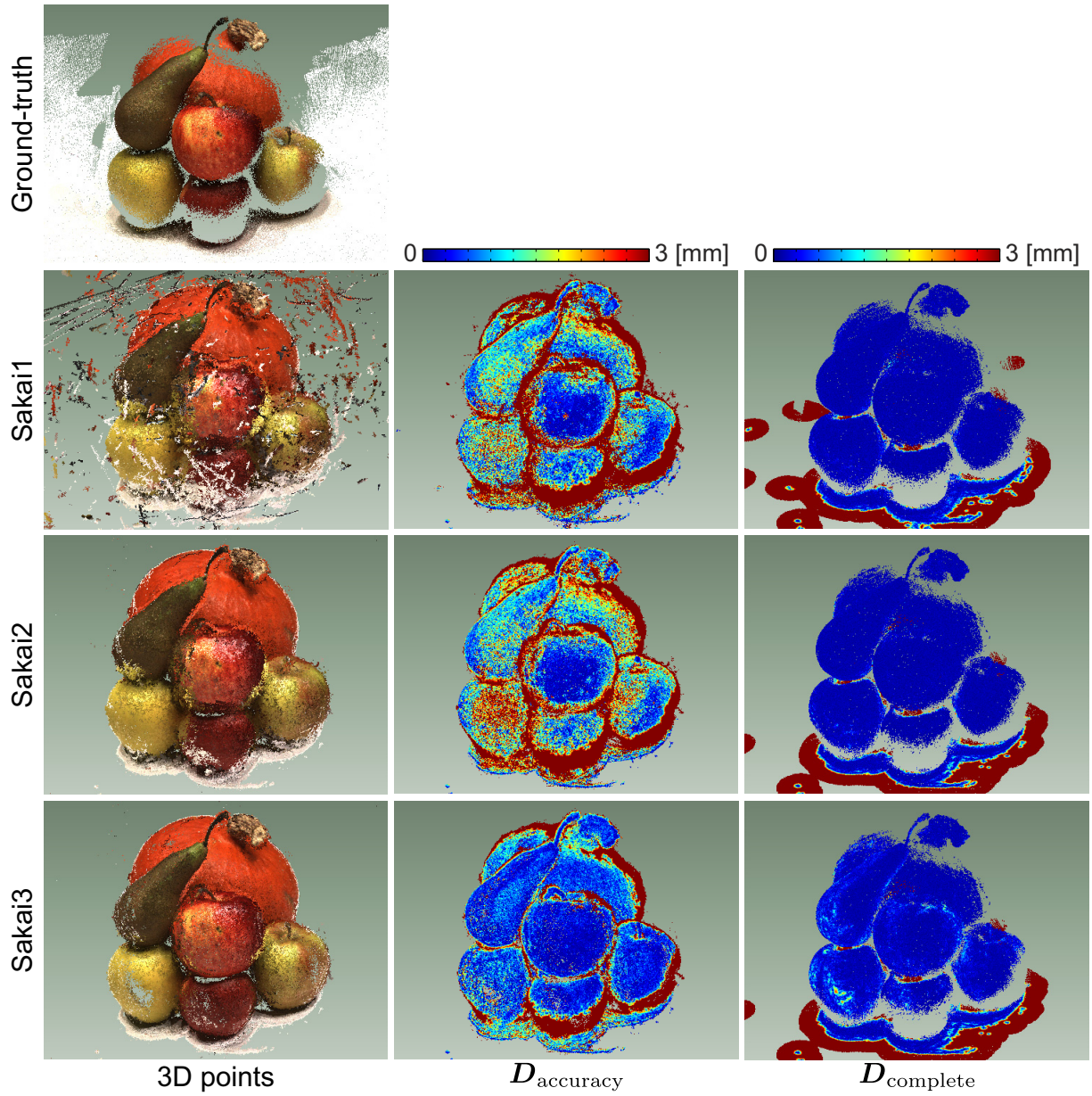


図 4.21 Scan 63 の 3 次元復元結果 (1 列目: 画素値で色をつけた 3 次元点群, 2 列目: $D_{accuracy}$ で色をつけた 3 次元復元点群, 3 列目: $D_{complete}$ で色をつけた真値の 3 次元点群)

の文化財の3次元復元を行う。3次元復元に必要な入力データがカメラ画像のみであるため、従来の3次元計測器に比べ、撮影装置が小型・軽量であり、撮影の負担も少ない。3.5節、および、4.3節で示したように、POCに基づくMVSアルゴリズムは、カメラ画像のみから1mm以下の精度で対象物体を復元可能である。そのため、適切なカメラ画像を撮影すれば、十分な精度で文化財のデジタルアーカイブのための3次元復元に利用できる。以下では、実際の文化財である瑞巖寺の欄間の3次元復元を行い、提案手法により高精度かつ簡便に文化財の3次元形状を取得できることを示す。

4.4.1 3次元復元処理の流れ

POCに基づくMVSアルゴリズムを文化財の3次元復元に適用する場合、一般的な多視点3次元復元の流れに従い、(i) 多視点画像の取得、(ii) カメラパラメータ推定、(iii) 3次元復元の3つのステップで対象物体の3次元形状を取得する。

まず、多視点画像の取得には、単眼カメラを用いた自由な移動撮影を用いる。文化財を対象とする場合、対象物体に様々な大きさの物体が想定され、かつ、対象物体を移動させることが難しい場合がある。そのため、多眼カメラによる撮影や、ロボットアーム・回転テーブルの使用は難しい。また、対象の文化財の画像が必ずしもインターネット上に公開されているとも限らないため、インターネットからの画像取得も困難である。一方で、単眼カメラを用いた自由な移動撮影では、撮影装置が小型・軽量であり、狭い空間での撮影にも適している。さらに、撮影されるカメラ画像をある程度制御可能であり、画像の明るさや焦点距離を制御しつつ、適切な距離から対象物体を撮影できる。

次に、カメラパラメータ推定には、多視点画像からのカメラパラメータ推定を用いる。多視点画像の取得に、単眼カメラを用いた自由な移動撮影を採用したため、撮影される多視点画像について、既知パターンを用いて事前校正を行うことが難しい。そのため、対象物体を撮影した多視点画像から直接カメラパラメータを推定する必要がある。本節では、カメラパラメータ推定に、Bundler [56], [58], [92] を利用する。Bundler では、Scale-Invariant Feature Transform (SIFT) [50] に基づく対応付け結果に対して、5点アルゴリズム [66]、P3P問題の解法 [67]、バンドル調整 [52]–[54] を用いてカメラパラメータを推定する。



(a)



(b)

図 4.22 瑞巖寺の欄間の撮影 [91]: (a) 復元対象の欄間, (b) 撮影の様子

最後に, 多視点画像とカメラパラメータから, 4.2 節で提案した POC に基づく MVS アルゴリズムを用いて 3 次元復元を行う. さらに, POC に基づく MVS アルゴリズムの出力した 3 次元点群に対して, Bradley らの手法 [23], および, Poisson Surface Reconstruction (PSR) [93] を用いて 3 次元メッシュモデルを生成する.

4.4.2 瑞巖寺欄間の撮影条件

図 4.22 に, 復元対象の欄間のカメラ画像と撮影の様子を示す. 欄間の大きさは, 幅 × 高さ × 厚さが約 $2,400 \times 815 \times 80$ mm である. この欄間を, 1,500 mm から 2,000 mm 離れた

距離から、欄間の中央にカメラを向けて撮影し、多視点画像を取得する。撮影に使用したカメラは、ニコン社製のデジタルカメラ D3 である。画像サイズは、 $4,000 \times 3,000$ ピクセルである。また、撮影の際にスタンド式の蛍光灯を設置し、700 ルクス程度の明るさを確保した。多視点画像は、欄間の表面の 131 枚、裏面の 83 枚を撮影した。撮影に要した時間は、合計 31 分である [91]。

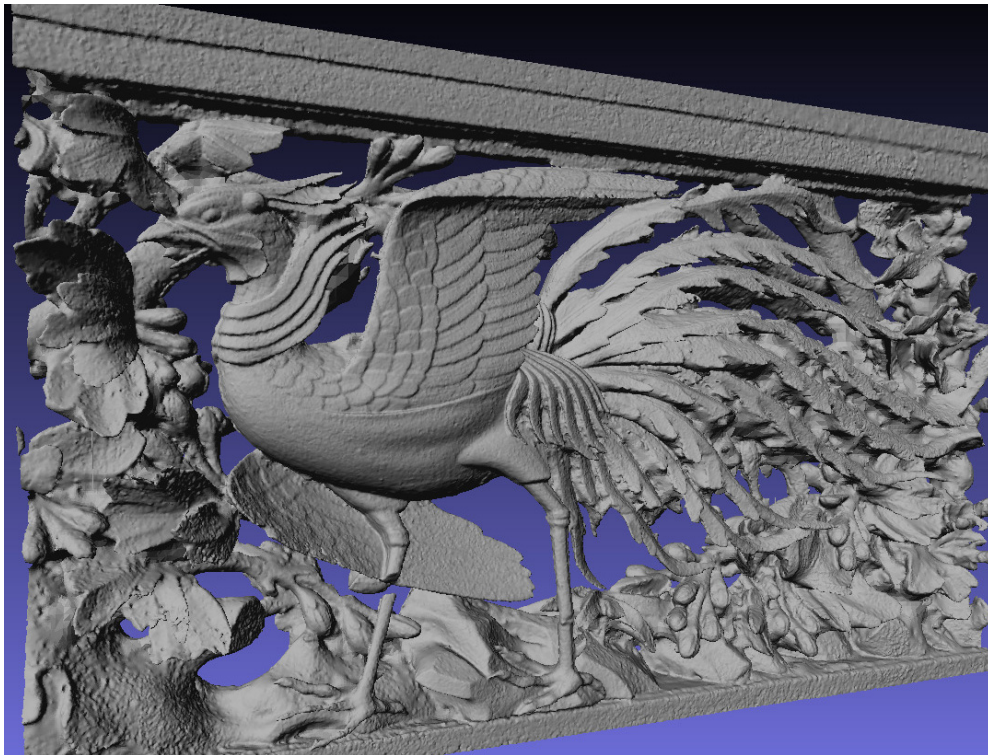
4.4.3 瑞巖寺欄間の 3 次元復元結果

撮影した 214 枚の画像のうち、表面で 25 枚、裏面で 54 枚の画像を手動で選択し、3 次元復元を行った。3 次元復元の計算機処理に要した時間は、約 7 時間である。図 4.23 に、欄間の 3 次元復元結果を示す。3 次元復元結果より、欄間の大部分については、対象の細かい形状まで復元できていることが確認できる。ただし、鳳凰の尾の部分は、オクルージョンが発生しやすく、復元結果が不正確になっている。

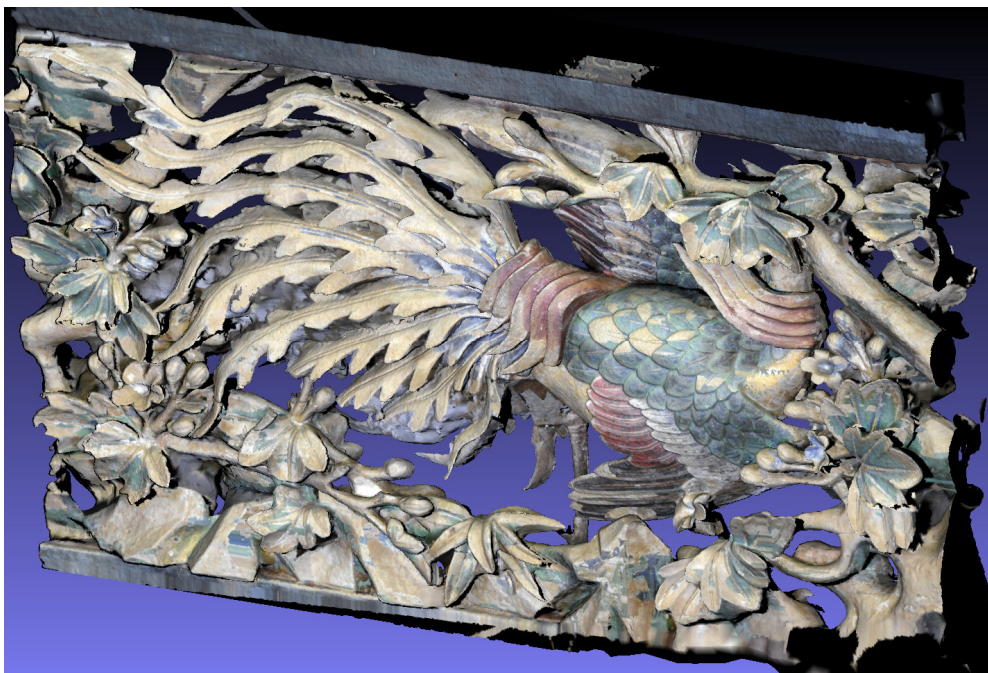
本実験では、3 次元復元精度を比較するため、GOM 社製の 3 次元デジタイザ ATOS (カタログ上の 3 次元復元誤差: 0.05 mm) にて、同一の物体を復元した。提案手法による 3 次元復元結果について、ATOS の復元結果との差は 0.3 mm 以下であった [91]。一方で、手持ち式の小型 3 次元スキャナの 3 次元復元精度は、カタログ上のスペックで 0.5 mm 前後である。以上より、提案手法は、手持ち式の 3 次元スキャナと同等以上の精度で 3 次元復元が可能であり、文化財のデジタルアーカイブにおいて、十分に利用できるものである。

4.5 むすび

本章では、高精度ウィンドウマッチングに基づく MVS アルゴリズムとして、POC に基づく MVS アルゴリズムについて述べた。まず、MVS のための POC に基づくウィンドウマッチングを用いたプレーンスイーピングに対して、(i) 階層的探索を用いたデプスマップ生成、(ii) 重み付きメディアンフィルタを用いたデプスマップの高精度化、(iii) グラフカットに基づく誤対応点除去、(iv) 閾値処理に基づくアーチファクト除去を適用し、POC に基づく MVS アルゴリズムを提案した。次に、公開データセットを用いた実験により、従来の MVS アルゴ



(a)



(b)

図 4.23 瑞巖寺の欄間の 3 次元復元結果 [91]: (a) テクスチャなし, (b) テクスチャあり

リズムと比較し，提案手法を用いて高精度な 3 次元復元が行えることを示した．さらに，提案手法の応用として，デジタルアーカイブを目的とした文化財の 3 次元復元について述べ，提案手法の有効性を示した．

第 5 章

結言

以上，第 2 章から第 4 章まで，高精度ウィンドウマッチングに基づく多視点 3 次元復元について述べた。

第 2 章では，多視点 3 次元復元に関する基礎的考察を行った。まず，多視点 3 次元復元の一連の処理として，多視点画像の取得，カメラパラメータ推定，3 次元復元のそれぞれについて概説した。次に，3 次元復元処理の中でも，3 次元復元精度，および，ロバスト性の観点から，多視点ステレオに着目し，多視点ステレオアルゴリズムの分類と特性について述べた。そして，多視点ステレオアルゴリズムの構成要素の中でも，ウィンドウマッチング手法が，アルゴリズム全体の性能を決定する重要な要素であることを述べた。最後に，従来の多視点ステレオアルゴリズムにおけるウィンドウマッチングの役割と，従来のウィンドウマッチング手法の問題点について述べた。

第 3 章では，多視点ステレオのための高精度ウィンドウマッチング手法について述べた。まず，提案手法で重要となる基礎技術として，位相限定相関法について述べた。次に，位相限定相関法のステレオビジョンへの適用方法，および，正規化視差に基づく位相限定相関関数の統合について述べ，多視点ステレオのための位相限定相関法に基づくウィンドウマッチング手法を提案した。さらに，多視点ステレオのための位相限定相関法に基づくウィンドウマッチングに対して，マッチングウィンドウの幾何補正を適用し，画像変形の大きいステレオ画像に対処した。性能評価実験では，従来の多視点ステレオアルゴリズムにおけるウィンドウマッチング

手法を分類し、その精度と計算コストについて、提案手法と比較した。実験を通して、提案手法により、従来のウィンドウマッチング手法と比べ、高精度に3次元点の奥行きを推定できることを示した。

第4章では、高精度ウィンドウマッチングに基づく多視点ステレオアルゴリズムについて述べた。第3章で提案した多視点ステレオのための位相限定相関法に基づくウィンドウマッチング手法に対して、階層的探索を用いたデプスマップ生成、重み付きメディアンフィルタを用いたデプスマップの高精度化、グラフカットに基づく誤対応点除去、閾値処理に基づくアーチファクト除去を適用することで、位相限定相関法に基づく多視点ステレオアルゴリズムを提案した。精度評価実験では、Jensenらの公開データセットを用いて、さまざまな特性の対象物体について、従来の多視点ステレオアルゴリズムと提案手法で3次元復元精度を比較した。実験を通して、提案手法により、従来の多視点ステレオアルゴリズムに比べ、高精度な3次元復元が行えることを示した。さらに、提案手法の応用として、デジタルアーカイブを目的とした文化財の3次元復元について述べた。実際の文化財を撮影したカメラ画像から、提案手法を用いて3次元復元を行い、提案手法の有効性を示した。

以上、本論文では、高精度ウィンドウマッチングに基づく多視点3次元復元に関する研究として、多視点ステレオのための位相限定相関法に基づくウィンドウマッチング、および、位相限定相関法に基づく多視点ステレオアルゴリズムを提案し、その有効性を確認した。多視点ステレオのための位相限定相関法に基づくウィンドウマッチングでは、位相限定相関法のピークモデルを用いた関数フィッティングにより、従来のウィンドウマッチングに比べ、多視点画像間のサブピクセル相当の奥行きを高精度に推定することが可能である。また、位相限定相関法に基づく多視点ステレオアルゴリズムでは、位相限定相関法に基づくウィンドウマッチングを用いた高精度な奥行き推定に加え、階層的探索や重み付けメディアンフィルタにより誤対応点の発生を抑制し、グラフカットや閾値処理により不正確な3次元復元点のみを除外することで、高精度な3次元点群生成が可能である。提案手法を多視点3次元復元の枠組みで利用することで、カメラ画像のみから、能動型の3次元復元技術に匹敵する高精度な3次元形状を復元できる。文化財のデジタルアーカイブでは、小型・軽量の装置による3次元形状の取得が望まれており、カメラ画像のみを用いた高精度3次元復元が特に有効である。本論文では、実

際の文化財の 3 次元復元を行い，提案手法が文化財の保存を目的とした 3 次元復元に十分利用できる技術であることを示した．

今後，本論文における高精度ウィンドウマッチングに基づく多視点 3 次元復元技術をさらに発展させるために取り組むべき課題として，画像補間を伴わないウィンドウマッチング手法の提案と，処理の高速化が挙げられる．本論文で提案した多視点ステレオのための位相限定相関法に基づくウィンドウマッチングでは，ステレオ平行化やウィンドウ切り出しの際に画像補間を必要とする．一方で，サブピクセル相当の奥行き分解能を考慮した時，3 次元復元精度に対する画像補間の影響は無視できない．ウィンドウマッチングの際のサブピクセルの画素値の推定を位相限定相関関数の計算の中で解決することで，さらなる高精度化につながると考えられる．また，現状，提案手法による 3 次元復元では，復元処理に数時間要するため，その用途が処理時間を重視しないものに限られる．一方，提案手法における処理時間の大部分は，階層的探索におけるデプスマップ生成の最上位層の処理時間である．最上位層の処理に，精度が低くても，高速なデプスマップ生成を適用することで，処理時間の大幅な改善が期待できる．以上のような課題を解決し，より高精度で，より高速な多視点 3 次元復元が実現すれば，多視点 3 次元復元技術がさまざまな用途で利用されるようになる．多視点 3 次元復元の用途としては，文化財のデジタルアーカイブや，都市の 3 次元マップ生成，医療における人体計測などが挙げられる．特に，本論文でも着目した文化財のデジタルアーカイブでは，多視点 3 次元復元技術に対する需要が高い．カメラ画像のみから高精度に対象の 3 次元形状を復元できれば，能動型の 3 次元復元において，装置の大きさから撮影が困難であった物体に対して，3 次元復元が可能になる．また，汎用のデジタルカメラが利用できるため，世界中のどこの美術品・建築物であっても，専用の撮影装置を運ぶ必要がなく，現地の作業者がカメラ画像の撮影を行うことができる．これにより，文化財のデジタルアーカイブにおいて，3 次元復元技術が盛んに利用されるようになると思われる．

参考文献

- [1] 吉澤徹, 最新光三次元計測. 朝倉書店, 2006.
- [2] 池内克史, 大石岳史, 3次元デジタルアーカイブ. 東京大学出版会, 2010.
- [3] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer-Verlag New York Inc., 2010.
- [4] 徐剛, 辻三郎, 3次元ビジョン. 共立出版, 1998.
- [5] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-views stereo reconstruction algorithms,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 519–528, June 2006.
- [6] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen, “On benchmarking camera calibration and multi-view stereo for high resolution imagery,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, June 2008.
- [7] 鳥居秋彦, 岡谷貴之, 延原章平, “多視点 3次元復元の研究動向,” 研究報告コンピュータビジョンとイメージメディア, Vol. 176, No. 1, pp. 1–22, Mar. 2011.
- [8] Y. Furukawa, “Multi-view 3D reconstruction techniques,” 研究報告コンピュータビジョンとイメージメディア, Vol. 176, No. 12, pp. 1–9, Mar. 2011.
- [9] 八木康史, 斎藤 英雄 (編), コンピュータビジョン最先端ガイド 5. アドコム・メディア株式会社, 2012.
- [10] S. Tran and L. Davis, “3D surface reconstruction using graph cuts with surface constraints,” *Proc. European Conf. Computer Vision*, Vol. 2, pp. 219–231, May 2006.
- [11] G. Vogiatzis, P. H. Torr, and R. Cipolla, “Multi-view stereo via volumetric graph-cuts

- and occlusion robust photo-consistency,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 29, No. 12, pp. 2241–2246, Dec. 2007.
- [12] S. N. Sinha, P. Mordohai, and M. Pollefeys, “Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh,” *Proc. Int’l Conf. Computer Vision*, pp. 1–8, Oct. 2007.
- [13] A. Zaharescu, E. Boyer, and R. Horaud, “Transformesh: A topology-adaptive mesh-based approach to surface evolution,” *Proc. Asian Conf. Computer Vision*, pp. 166–175, Nov. 2007.
- [14] A. Delaunoy, E. Praos, P. Gargallo, J. P. Pons, and P. Sturm, “Minimizing the multi-view stereo reprojection error for triangular surface meshes,” *Proc. British Machine Vision Conf.*, Sept. 2008.
- [15] V. H. Hiep, P. Labatut, J. P. Pons, and R. Keriven, “Towards high-resolution large-scale multi-view stereo,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1430–1437, June 2009.
- [16] V. H. Hiep, R. Keriven, P. Labatut, and J. P. Pons, “High accuracy and visibility-consistent dense multiview stereo,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 34, No. 5, pp. 889–901, May 2012.
- [17] M. Lhuillier and L. Quan, “A quasi-dense approach to surface reconstruction from uncalibrated images,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 3, pp. 418–433, Mar. 2005.
- [18] M. Habbeck and L. Kobbelt, “A surface-growing approach to multi-view stereo reconstruction,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.
- [19] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multiview stereopsis,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1362–1376, Aug. 2010.
- [20] M. Goesele, B. Curless, and S. M. Seitz, “Multi-view stereo revisited,” *Proc. IEEE*

- Conf. Computer Vision and Pattern Recognition*, pp. 2402–2409, June 2006.
- [21] C. Strecha, R. Fransens, and L. V. Gool, “Combined depth and outlier estimation in multi-view stereo,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2394–2401, June 2006.
- [22] D. Gallup, J. M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, “Real-time plane-sweeping stereo with multiple sweeping directions,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.
- [23] D. Bradley, T. Boubekeur, and W. Heidrich, “Accurate multi-view reconstruction using robust binocular stereo and surface meshing,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, June 2008.
- [24] N. D. F. Campbell, G. Vogiatzis, C. Hernandez, and R. Cipolla, “Using multiple hypotheses to improve depth-maps for multi-view stereo,” *Proc. European Conf. Computer Vision*, pp. 766–779, Oct. 2008.
- [25] J. Li, E. Li, Y. Chen, L. Xu, and Y. Zhang, “Bundle depth-map merging for multi-view stereo,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2569–2776, June 2010.
- [26] E. Tola, C. Strecha, and P. Fua, “Efficient large-scale multi-view stereo for ultra high-resolution image sets,” *Machine Vision and Applications*, Vol. 23, No. 5, pp. 908–920, Sept. 2012.
- [27] S. Shen, “Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes,” *IEEE Trans. Image Processing*, Vol. 22, No. 5, pp. 1901–1914, May 2013.
- [28] 奥富正敏 (編), デジタル画像処理. CG-ARTS 協会, 2004.
- [29] E. Tola, V. Lepetit, and P. Fua, “DAISY: An efficient dense descriptor applied to wide-baseline stereo,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, pp. 815–830, May 2010.
- [30] M. Shimizu and M. Okutomi, “Sub-pixel estimation error cancellation on area-based

- matching,” *Int’l J. Computer Vision*, Vol. 63, No. 3, pp. 207–224, July 2005.
- [31] 酒井修二, 伊藤康一, 青木孝文, 運天弘樹, “位相限定相関法に基づく多視点画像からのデプスマップ生成の検討,” 第 14 回画像の認識・理解シンポジウム, Vol. 14, No. IS4-10, pp. 1347–1354, July 2011.
- [32] S. Sakai, K. Ito, T. Aoki, T. Masuda, and H. Unten, “An efficient image matching method for multi-view stereo,” *Proc. Asian Conf. Computer Vision*, pp. 1–8, Nov. 2012.
- [33] 酒井修二, 伊藤康一, 青木孝文, 増田智仁, 運天弘樹, “多視点ステレオのための位相限定相関法に基づく画像マッチング手法とその性能評価,” 第 75 回 情報処理学会 全国大会, Vol. 75, No. 1D-1, pp. 2–23–2–24, Mar. 2013.
- [34] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, “High-accuracy subpixel image registration based on phase-only correlation,” *IEICE Trans. Fundamentals*, Vol. E86-A, No. 8, pp. 1925–1934, Aug. 2003.
- [35] K. Takita, M. A. Muquit, T. Aoki, and T. Higuchi, “A sub-pixel correspondence search for computer vision applications,” *IEICE Trans. Fundamentals*, Vol. E87-A, No. 8, pp. 1913–1923, Aug. 2004.
- [36] T. Shibahara, T. Aoki, H. Nakajima, and K. Kobayashi, “A sub-pixel stereo correspondence technique based on 1D phase-only correlation,” *Proc. Int’l Conf. Image Processing*, pp. V–221–V–224, Sept. 2007.
- [37] 柴原琢磨, 沼徳仁, 長嶋聖, 青木孝文, 中島寛, 小林孝次, “1 次元位相限定相関法に基づくステレオ画像の高精度サブピクセル対応付け手法,” 電子情報通信学会論文誌 D, Vol. J91-D, No. 9, pp. 2343–2356, Sept. 2008.
- [38] 酒井修二, 伊藤康一, 青木孝文, 増田智仁, 運天弘樹, “多視点ステレオのための位相限定相関法に基づく画像マッチングの高精度化,” 研究報告コンピュータビジョンとイメージメディア, Vol. 186, No. 22, pp. 1–8, Mar. 2013.
- [39] 酒井修二, 伊藤康一, 青木孝文, 渡邊隆史, 運天弘樹, “多視点ステレオのための高精度ウィンドウマッチング手法とその精度評価,” 第 17 回 画像の認識・理解シンポジウム,

- Vol. 17, No. SS1-47, pp. 1–2, July 2014.
- [40] C. H. Esteban and F. Schmitt, “Silhouette and stereo fusion for 3D object modeling,” *Computer Vision and Image Understanding*, Vol. 96, No. 3, pp. 367–392, Dec. 2004.
- [41] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Anæs, “Large scale multi-view stereopsis evaluation,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 406–413, June 2014.
- [42] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “Multi-View Stereo.” <http://vision.middlebury.edu/mview/>.
- [43] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Anæs, “MVS Data Set 2014 — DTU Robot Image Data Sets.” http://roboimagedata.compute.dtu.dk/?page_id=36.
- [44] J. Stühmer, S. Gumhold, and D. Cremers, “Real-time dense geometry from a hand-held camera,” *Pattern Recognition*, Vol. 6376, pp. 11–20, 2010.
- [45] P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, and M. Pollefeys, “Live metric 3D reconstruction on mobile phones,” *Proc. Int’l Conf. Computer Vision*, pp. 65–72, Dec. 2013.
- [46] 三浦衛, 酒井修二, 石井純平, 山尾創輔, 伊藤康一, 青木孝文, “カメラの移動撮影に基づく2視点からの3次元形状計測とその性能評価,” *映像情報メディア学会誌*, Vol. 68, No. 4, pp. J135–J143, Apr. 2014.
- [47] 山尾創輔, 三浦衛, 酒井修二, 伊藤康一, 青木孝文, “カメラの移動撮影に基づく簡便なオンライン3次元計測システムの検討,” *研究報告コンピュータビジョンとイメージメディア*, Vol. 192, No. 2, pp. 1–8, May 2014.
- [48] 山尾創輔, 三浦衛, 酒井修二, 伊藤康一, 青木孝文, “カメラの移動撮影に基づくオンライン3次元計測システム,” *第20回画像センシングシンポジウム*, Vol. 20, No. DS2-03, June 2014.
- [49] C. Strecha, “Multi-View Evaluation.” <http://cvlab.epfl.ch/data/>.
- [50] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int’l J. Computer Vision*, Vol. 60, No. 2, pp. 91–110, Nov. 2004.

- [51] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: Speeded up robust features,” *Proc. European Conf. Computer Vision*, pp. 404–417, May 2006.
- [52] 岡谷貴之, “バンドルアジャストメント,” 研究報告コンピュータビジョンとイメージメディア, Vol. 167, No. 37, pp. 1–16, June 2009.
- [53] 八木康史 斎藤 英雄 (編), コンピュータビジョン最先端ガイド 3. アドコム・メディア株式会社, 2010.
- [54] M. I. A. Lourakis and A. A. Argyros, “SBA: A software package for generic sparse bundle adjustment,” *ACM Trans. Math. Software*, Vol. 36, No. 1, pp. 1–30, Mar. 2009.
- [55] R. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge University Press, 2004.
- [56] N. Snavely, S. M. Seitz, and R. Szeliski, “Photo tourism: Exploring image collections in 3D,” *ACM Trans. Graphics (Proc. ACM SIGGRAPH)*, Vol. 25, No. 3, pp. 853–846, July 2006.
- [57] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz, “Multi-view stereo for community photo collections,” *Proc. Int’l Conf. Computer Vision*, pp. 1–8, Oct. 2007.
- [58] N. Snavely, S. M. Seitz, and R. Szeliski, “Modeling the world from internet photo collections,” *Int’l J. Computer Vision*, Vol. 80, No. 2, pp. 189–210, Nov. 2008.
- [59] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, “Building Rome in a day,” *Proc. Int’l Conf. Computer Vision*, pp. 72–79, Oct. 2009.
- [60] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz, and R. Szeliski, “Reconstructing Rome,” *IEEE Computer*, Vol. 43, No. 6, pp. 40–47, June 2010.
- [61] S. Agarwal, Y. Furukawa, N. Snavely, B. Curless, S. M. Seitz, and R. Szeliski, “Building Rome in a day,” *Commun. ACM*, Vol. 54, No. 10, pp. 105–112, Oct. 2011.
- [62] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski, “Towards internet-scale multi-view stereo,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1434–

- 1441, June 2010.
- [63] “Flickr.” <https://www.flickr.com/>.
- [64] Z. Zhang, “Flexible camera calibration by viewing a plane from unknown orientations,” *Proc. Int’l Conf. Computer Vision*, pp. 666–673, Sept. 1999.
- [65] J. Y. Bouguet, “Camera calibration toolbox for matlab.” http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [66] D. Nistér, “An efficient solution to the five-point relative pose problem,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 6, pp. 756–770, June 2004.
- [67] L. Kneip, D. Scaramuzza, and R. Siegwart, “A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation,” *Proc. Int’l Conf. Computer Vision and Pattern Recognition*, pp. 2969–2976, June 2011.
- [68] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, Vol. 24, No. 6, pp. 381–395, June 1981.
- [69] B. G. Baumgart, *Geometric modeling for computer vision*. Stanford University, 1974.
- [70] A. Laurentini, “The visual hull concept for silhouette-based image understanding,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 150–162, Feb. 1994.
- [71] 金谷健一, *これなら分かる最適化数学*. 共立出版, 2005.
- [72] X. Zabulis and K. Daniilidis, “Multi-camera reconstruction based on surface normal estimation and best viewpoint selection,” *Proc. Int’l Symp. 3D Data Processing, Visualization and Transmission*, pp. 733–740, Sept. 2004.
- [73] N. Cornells and L. V. Gool, “Real-time connectivity constrained depth map computation using programmable graphics hardware,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1099–1104, June 2005.
- [74] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, “DTAM: Dense tracking and

- mapping in real-time,” *Proc. Int’l Conf. Computer Vision*, pp. 2320–2327, Nov. 2011.
- [75] C. Zach, M. Sormann, and K. Karner, “High-performance multi-view reconstruction,” *Proc. Int’l Symp. 3D Data Processing, Visualization and Transmission*, pp. 113–120, June 2006.
- [76] A. Wendel, M. Maurer, G. Graber, T. Pock, and H. Bischof, “Dense reconstruction on-the-fly,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1450–1457, June 2012.
- [77] C. D. Kuglin and D. C. Hines, “The phase correlation image alignment method,” *Proc. Int’l Conf. Cybernetics and Society*, pp. 163–165, 1975.
- [78] D. J. Fleet, “Phase-based disparity measurement,” *CVGIP: Image Understanding*, Vol. 53, No. 2, pp. 198–210, 1991.
- [79] 青木孝文, 伊藤康一, 柴原琢磨, 長嶋聖, “位相限定相関法に基づく高精度マシンビジョン—ピクセル分解能の壁を越える画像センシング技術を目指して—,” *IEICE Fundamentals Review*, Vol. 1, No. 1, pp. 30–40, Jan. 2007.
- [80] 青木孝文, 伊藤康一, 本間尚文, “位相情報に基づく画像マッチング技術とその応用展開 —3D ビジョンからバイオメトリクスまで—,” *電子情報通信学会誌*, Vol. 90, No. 8, pp. 680–685, Aug. 2007.
- [81] M. Okutomi and T. Kanade, “A multiple-baseline stereo,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 353–363, Apr. 1993.
- [82] S. Sakai, K. Ito, T. Aoki, and H. Unten, “Accurate and dense wide-baseline stereo matching using SW-POC,” *Proc. Asian Conf. Pattern Recognition*, pp. 335–339, Nov. 2011.
- [83] 酒井修二, 伊藤康一, 青木孝文, 運天弘樹, “基線長変化にロバストなステレオ画像間の高精度対応付け手法,” *電子情報通信学会論文誌 D*, Vol. J95-D, No. 8, pp. 1609–1622, Aug. 2012.
- [84] 酒井修二, 伊藤康一, 青木孝文, “基線長変化にロバストな高精度ステレオ画像対応付け手法,” *画像ラボ*, Vol. 25, No. 6, pp. 47–53, June 2014.

- [85] P. J Besl and N. D. McKay, “A method for registration of 3-D shapes,” *EEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239–256, Feb. 1992.
- [86] “Intel 64 and IA-32 architectures optimization reference manual.” <http://www.intel.com/content/dam/doc/manual/64-ia-32-architectures-optimization-manual.pdf>.
- [87] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3017–3024, June 2011.
- [88] A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, “Fast cost-volume filtering for visual correspondence and beyond,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 35, No. 2, pp. 504–511, Feb. 2013.
- [89] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu, “Constant time weighted median filtering for stereo matching and beyond,” *Proc. Int’l Conf. Computer Vision*, pp. 49–56, Dec. 2013.
- [90] 八木康史 斎藤 英雄 (編), コンピュータビジョン最先端ガイド 1. アドコム・メディア株式会社, 2008.
- [91] 渡邊隆史, 運天弘樹, 酒井修二, 伊藤康一, 青木孝文, 北野信彦, 酒巻仁一, 伊奈仁, 新野一浩, “多視点ステレオ技術を用いた瑞巖寺欄間の三次元形状計測,” 日本文化財科学会 第31回大会, July 2014.
- [92] N. Snavely, “Bundler: Structure from Motion (SfM) for Unorderd Image Collections.” <http://www.cs.cornell.edu/~snavely/bundler/>.
- [93] M. Kazhdan, M. Bolitho, and H. Hoppe, “Poisson surface reconstruction,” *Proc. Symp. Geometry Processing*, pp. 61–70, June 2006.

謝辞

本論文は、著者が東北大学 大学院情報科学研究科 情報基礎科学専攻 計算機構論分野（青木・本間研究室）において行った研究を取りまとめたものであります。本研究を推し進めるにあたり、多くの方々からご協力とご助言を頂きました。

恩師青木孝文教授には、研究に対する懇切なるご指導と終始変わらぬ励ましを頂きました。研究だけでなく、何事に対しても情熱的に取り組む先生のご姿勢から、非常に多くのことを学ばせて頂きました。学部4年次から博士後期課程修了まで、6年間に渡り先生のもとで研究を行えたことは、私にとって大変な幸運であり、今後の人生の大きな糧になると確信しております。ここに改めて、深く感謝の意を表します。

本論文をまとめるにあたり、川又政征教授ならびに篠原歩教授より、それぞれのご専門の立場から有意義なご意見を賜りました。ここに深く感謝いたします。

本間尚文准教授には、本研究を進めるにあたり、様々なご討論とご助言を頂きました。ここに改めて深く感謝いたします。

凸版印刷株式会社の運天弘樹博士、増田智仁博士、渡邊隆史博士、猪瀬健二氏には、共同研究を通して、多視点3次元復元技術および文化財のデジタルアーカイブに関するご討論とご助言を頂きました。特に、多視点3次元復元技術の実用の観点からは、大変有益なご意見を頂きました。また、文化財のデジタルアーカイブに関して、実際の文化財のデータをご提供して頂きました。ここに改めて深く感謝いたします。

宮城県産業技術総合センターの太田晋一氏には、多視点画像データセットの作成にあたり、データの計測にご協力頂きました。ここに改めて深く感謝いたします。

日本電信電話株式会社の土田勝博士には、マルチバンドカメラを用いた色再現という非常に

興味深いトピックについてご教授頂きました。また，マルチバンドカメラの多視点 3 次元復元への利用についてご討論させて頂きました。ここに改めて深く感謝いたします。

東社シーテック株式会社の真鍋俊之氏には，アルゴリズムの実装やプロトタイプ作製などで大変お世話になりました。ここに改めて深くお礼申し上げます。

伊藤康一助教には，学部 4 年次より研究全般に関する直接的なご指導と活発なご討論を頂くとともに，本論文の執筆においても様々なご助言を頂きました。ここに改めて深く感謝いたします。

青木・本間研究室の先輩である宮澤一之博士（現 三菱電機株式会社）には，位相限定相関法の基礎原理，および，位相限定相関法に基づくステレオビジョンについて，有益なご議論とご指導を頂きました。高橋徹博士（現 日本電気株式会社）には，多視点 3 次元復元に関するご指導を頂くとともに，本研究に対して多大なるご助言を頂きました。ここに改めて深く感謝いたします。

青木・本間研究室の青山章一郎氏，三浦衛氏には，研究全般に関する活発なご討論を頂くとともに，本論文の執筆においても様々なご助言を頂きました。また，研究室の同期として，様々な面でご助力を頂きました。ここに改めて深く感謝いたします。

青木・本間研究室の後輩である石井純平氏（現 東北電力株式会社），青木・本間研究室の山尾創輔氏には，多視点 3 次元復元に関するご討論を頂くとともに，本研究を進めるにあたり，多大なるご助力を頂きました。佐藤拓杜氏，丸木大樹氏，平館睦基氏には，研究全般を通してさまざまなご意見やご協力を頂きました。ここに改めて深くお礼申し上げます。

日頃の研究室生活において様々な面でご協力頂いた研究室諸氏に心よりお礼申し上げます。

このように，本研究を進めるにあたり，多くの方々からご協力とご助言を頂きました。私一人の力では，到底このような形で研究を取りまとめ，論文として著すことはできませんでした。諸氏に重ねて厚くお礼申し上げます。

2015 年 1 月 20 日