

局所特徴量を用いた物体の姿勢推定の 高精度化に関する研究

小堀 訓成

目次

第 1 章	序論	1
1.1	背景	2
1.1.1	生活支援ロボット開発の位置づけ	2
1.1.2	Mobile Manipulator の現状と課題	3
1.1.3	姿勢推定技術に注目する理由	7
1.2	本研究の位置づけ	8
1.3	姿勢推定技術の現状と本研究のアプローチ	9
1.3.1	姿勢推定技術の現状と問題点	9
1.3.2	本研究のアプローチ	12
1.3.3	目標精度	13
1.4	本論文の構成	14
第 2 章	姿勢推定に用いる局所特徴量記述子の事前性能評価	15
2.1	局所特徴量記述子の研究動向	15
2.2	SIFT (Scale-Invariant Feature Transform)	16
2.2.1	スケールスペースにおける極値探索	17
2.2.2	キーポイントの検出	18
2.2.3	キーポイントの向きの算出	20
2.2.4	特徴量記述子の表現	21
2.3	特徴量マッチング	21
2.4	SIFT の性能評価試験	23
2.5	まとめ	25
第 3 章	局所特徴量とジャイロセンサの相補的な活用による姿勢推定	27

3.1	ロボットの姿勢推定に関する先行研究	28
3.2	ジャイロセンサとカメラの相補性	29
3.3	提案手法を構成する要素技術：拡張 Kalman フィルタ	29
3.3.1	観測方程式と最尤推定	30
3.3.2	状態方程式と誤差伝搬	32
3.4	提案手法：ジャイロセンサと SIFT を利用した姿勢推定	33
3.4.1	システム構成	33
3.4.2	工夫点 1：カメラから求めた姿勢の信頼性の判定	35
3.4.3	工夫点 2：Kalman フィルタの状態量設計と更新方法	38
3.4.4	工夫点 3：ジャイロセンサを利用した画像処理	39
3.4.5	提案手法の特徴	41
3.5	実験と考察	41
3.5.1	実験条件・手順	41
3.5.2	実験結果	43
3.5.3	カメラから姿勢が求まらない場合の検証	47
3.5.4	ジャイロセンサを利用した画像処理の検証	49
3.5.5	制御への適用に関する考察	50
3.6	まとめ	51
第 4 章	局所特徴量を活用した“検出しやすい”姿勢推定用マーカ	53
4.1	提案するマーカの特徴	53
4.2	位置・姿勢推定マーカに関する先行研究	54
4.3	スケールスペース理論に基づく検出しやすい符号化パターン	56
4.3.1	検出処理	57
4.3.2	認識処理	57
4.4	提案手法	57
4.4.1	マーカ検出：キーポイント検出によるマーカの絞り込み	57
4.4.2	マーカ認識：M 系列を用いたデータパターン	61
4.5	実験と考察	63
4.5.1	実験条件・手順	63
4.5.2	ボケに対する検出・認識性能の比較	64

4.5.3	雑然とした背景環境下での多様な配置に対する検出・認識性能の比較	67
4.5.4	距離に対する検出・認識性能の分析	69
4.5.5	角度に対する検出・認識性能の分析	69
4.6	まとめ	69
第 5 章	局所特徴量を活用した“検出・認識しやすい”姿勢推定用マーカ	73
5.1	姿勢推定マーカの課題	74
5.2	先行研究：ランダムドットマーカ	75
5.2.1	特徴量表現	75
5.2.2	登録アルゴリズム	77
5.2.3	検索アルゴリズム	77
5.3	提案手法	78
5.3.1	工夫点 1：Spectral SIFT とフィルタリングを用いたドット検出	80
5.3.2	工夫点 2：ドットスケールによるデータ表現とデータ欠損時の対策	82
5.4	実験と考察	83
5.4.1	ドット検出方法の比較	83
5.4.2	様々なマーカ配置におけるデータデコードの頑健性の分析	86
5.4.3	距離に対する検出・認識性能の分析	87
5.5	まとめ	89
第 6 章	結論	91
6.1	本研究の成果	91
6.2	今後の課題と展望	93
6.2.1	今後の課題	93
6.2.2	将来技術としての応用	94
6.2.3	さらなる開発の方向と長期的な展望	94
	謝辞	97
	本人主筆の研究業績	109
	本人主筆のその他の研究業績	111

第 1 章

序論

本論文は、生活支援ロボットの自律的な行動に必要な“姿勢推定”に関する研究成果をまとめたものである。

少子高齢化、生産年齢人口の減少が進展するなか、ロボット技術は、製造業の生産現場、医療・介護現場、農業・建設・インフラの作業現場などの幅広い分野で、人手不足の解消、過重な労働からの解放、生産性の向上などの社会課題を解決する可能性を有している。従来型の製造現場のロボットは、予め決められた物体に対して、カメラ配置も最適化された条件下で、把持操作等の予め決められた動作を繰り返すものである。そのため、生産性は向上するものの、用途が限定されている。

一方、近年は、移動型アームロボット（以下、Mobile Manipulator）と呼ばれる、より汎用的なタスクを遂行するためのロボットが開発されている。これは、ロボット自身が自由に移動でき、ロボットアームで様々な物体を把持できるようなハードウェア構成を有している。このようなロボットをより幅広い分野で活用するには、ロボットが環境中の自らの“位置・姿勢”を認識し、さらに、環境中の物体を認識して、その“位置・姿勢”までを把握するすることが不可欠である。ロボットのハードウェアは普及しつつあるものの、それに見合うだけのソフトウェア開発が追いついていないのが現状である。Mobile Manipulator を日常生活の家事支援で活用することは、少子高齢化の背景からも特に期待されている。また、技術的な観点からもその応用先は広く、製造ラインをより自由度をもって構築することができ、Industrie 4.0 [1] で推進されている多品種大量生産への貢献が期待される。

そこで本論文では、Mobile Manipulator に課す汎用的なタスクとして日常生活の家事

支援を想定し，そのための基礎研究として，姿勢推定を軸に (i) ロボットの姿勢推定，及び，(ii) ロボットが把持動作を行う物体の姿勢推定に関する研究成果をまとめる．

本章では，まず 1.1 節で Mobile Manipulator における“位置・姿勢推定”の必要性，さらに“姿勢推定”に注力した背景について述べる．次に，本研究の位置づけを 1.2 節で述べる．更に，“姿勢推定”の現状技術とその課題，課題を解決するためのアプローチを 1.3 節で述べた後，本論文全体の構成を 1.4 節で述べる．

1.1 背景

生活支援ロボットの開発に関する背景を 1.1.1 項で，Mobile Manipulator の現状と課題を 1.1.2 項で述べる．そして，姿勢推定技術に注目する理由を 1.1.3 項に記す．

1.1.1 生活支援ロボット開発の位置づけ

経済産業省は「日本再興戦略」改訂 2014 [2] にて“ロボットによる新たな産業革命”を掲げ，ロボット技術の活用により生産性の向上を実現し，企業の収益力向上，賃金の上昇を図ることなどを基本方針として定めた．さらに，その実現に向けて，ロボット革命実現会議を複数回にわたって開催し，その結果，“ロボット新戦略”として以下の 3 つの目標を定めている．

1. センサ，人工知能などの技術進歩により，従来はロボットと位置づけられてこなかったモノまでもロボット化すること．例えば，自動車，家電，携帯電話や住居までもがロボットの 1 つとなること．
2. 製造現場から日常生活の様々な場面でロボットが活用されること．
3. 社会課題の解決やものづくり・サービスの国際競争力の強化を通じて，新たな付加価値を生み出し，利便性と富をもたらす社会を実現すること．

“ロボット新戦略”では上記の実現にあたって「自律化」，「情報端末化」，「ネットワーク化」の 3 つが重要な技術テーマであることにも言及している．

ロボットによる日常生活の家事支援は，上記の目標 2 に該当する．さらに，生活支援ロボットの自律的な行動とは，物を取って人に渡す，床の掃除を行う，洗濯を行うなどの家事全般に対して必要なタスクを総称する．

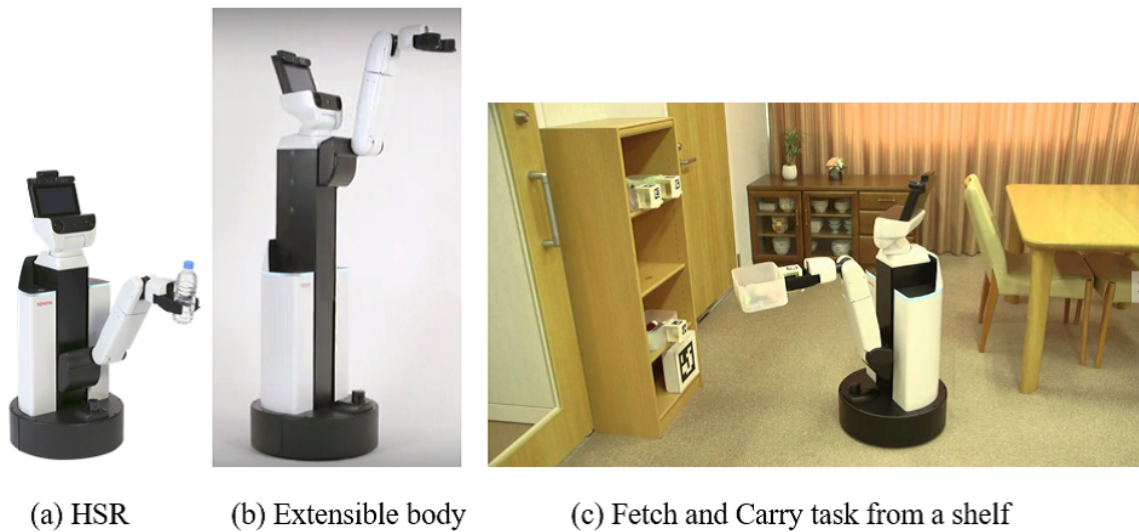


図 1.1: トヨタ自動車（株）の生活支援ロボット：HSR

1.1.2 Mobile Manipulator の現状と課題

本項では、生活支援ロボットを実現するための Mobile Manipulator について、具体的な製品とそれらに関する技術的な課題について説明する。

これまで多くの Mobile Manipulator が開発されてきた。代表的な開発事例として、トヨタ自動車（株）では、図 1.1 に示すような生活支援ロボット HSR（Human Support Robot）[3][4] を開発している。これは、手足の不自由な障がい者や高齢者などの家庭内における自立生活支援を目標としている。ROS というロボット向け OS を搭載しており、基本機能はオープンソースとして活用し、その上で開発ができる。そのため、ROS のコミュニティで開発が進められている各種ソフトウェア（例えば、自己位置推定など）は、そのまま使うことができるという利点がある。さらに現在は、HSR の開発コミュニティに参画する国内外の研究機関にロボットとそのソフトウェアが配布され、要素技術開発の更なる加速が図られている。RGB-D カメラ、ステレオカメラ、広角カメラ、マイクロフォンアレイ、ディスプレイが頭部に搭載され、脚部は全方位台車になっている。また床面側の下部には、慣性センサとレーザレンジセンサが搭載されている。ボディは伸張し、目的に応じて長さを変化させることができる。アーム先端のハンドには、グリップとバキューム機構、さらにカメラが付いている。ペットボトルのような大きな剛体から鉛筆やカードなどの小さな剛体、更に衣服やタオルなどの柔軟物体までを把持することが可能で

あり，所定の箇所に運び，渡すことまでを基本機能として提供している．この一連の動作は，Fetch and Carry タスク（物取りタスク）と呼ばれ，生活支援における基本機能と位置づけられている．

また同様のロボットは，図 1.2 に示す Fetch [5]（Fetch Robotics 社），図 1.3 に示す PR2 [6]（Willow Garage 社）からも提供されており，いずれも ROS [7] を搭載し，ハードウェア構成も類似している．現状，これらのロボットは主に研究用途向けであるが，物取りタスクや洗濯物の折り畳みなど，家事支援での自律的な動作が研究開発されている．

一方，研究用途ではなく，商用販売されているロボットも存在する．例えば，図 1.4 (a) に示す Baxter [8]（Rethink Robotics 製）や図 1.4 (b) に示す NEXTAGE [9]（川田



(a) Fetch



(b) Fetch and Carry task

図 1.2: Fetch Robotics 社のロボット : Fetch [5]



(a) PR2 robot



(b) Folding task for a towel

図 1.3: Willow Garage 社のロボット : PR2 [6]



図 1.4: 製造ラインで活躍する人間型ロボット (出典: (a) <http://www.nihonbinary.co.jp/>, (b) <http://www.imart.co.jp/>)

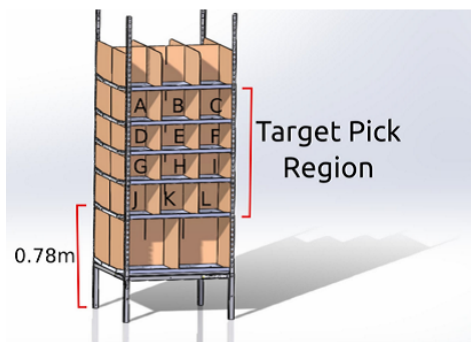


図 1.5: 物流システムで活躍する移動台車ロボット: Kiva Systems [11]

工業製)に代表される人間型ロボットが工場の生産ラインで広く活用されるようになりつつある。近年、工場のIoT化と無人化が急速に進んでおり、ロボットが人間の代わりとして、簡単な把持作業や部品配列作業を実施する。このBaxterやNEXTAGEは、ティーチング機能を有しており、事前に人がロボットを操作し、所望の動作を行うことで、動作を学習させることができる。省スペースであり、様々な作業ができる自由度をもつのが特徴である。またBaxterやNEXTAGEは、脚部として自律移動可能な台車を搭載していないが、移動台車型ロボットと組み合わせることで、Mobile Manipulatorになることも可能である。図1.5に、Kiva Systems (現在、Amazon Robotics社[11])のロボットを示す。Mobile Manipulatorの移動台車の部分だけをロボットとして製品化したものであ



(a) Example of Fetch and Carry task in Amazon Picking Challenge
(Left: Baxter, Right: Mitsubishi Electric Corp.)



(b) Shelf used in the challenge



(c) Items used in the challenge

図 1.6: Amazon Picking Challenge (出典 : <https://www.amazonrobotics.com/>)

る。Amazon 社の物流倉庫で使用され、商品棚を自動搬送する。既に、計 15,000 台が配送センタで稼働している [11]。

これらのロボットは、Mobile Manipulator に分類されるロボットであり、特に、Baxter, NEXTAG, Kiva Systems は用途を特化して商品化したものといえる。

ここで、自律的に Fetch and Carry タスクを行う上での必要技術（言い換えると、Mobile Manipulator の課題）をまとめる。必要な技術は、大きく以下の 2 つに大別できる。

1. グローバル座標系：自己位置・姿勢推定

事前に用意した環境のグローバル座標系におけるロボット自身の自己位置・姿勢を同定すること。

2. ローカル座標系：把持対象物の相対位置・姿勢推定

ロボットに搭載されたカメラから把持対象物体までの相対的な位置・姿勢を同定すること。

このような課題設定は、2015 年の ICRA (International Conference on Robotics and Automation) に併設して企画された Amazon Picking Challenge [10] コンテストでも取り上げられた。Amazon Picking Challenge では、図 1.6 に示す棚と把持対象物体が事前に公開され、ロボットは棚まで自律的に移動し、棚から物体を掴み、それを所定の箇所に配置することを競う。物体ごとに難易度に応じた得点が設定されている。さらに、物体を途中で落としたり、傷つけた場合は減点される。棚の中の物体の配置は、事前に知らされていない。

グリップの把持パターンは作り込みで良く、またアームの軌道生成は従来技術で対処できるため、上述した必要技術 (Mobile Manipulator の課題) さえ精度良く達成できれば、Amazon Picking Challenge の課題は解決できると期待する。2015 年の ICCV (International Conference on Computer Vision) では、“International Workshop on Recovering 6D Object Pose” という Workshop も立ち上がり、この課題に対して研究者の関心が高まっている。

1.1.3 姿勢推定技術に注目する理由

本項では、前項で明らかにした Fetch and Carry タスクを行う際の必要技術に対して、“姿勢推定” に研究の方向性を絞った理由を説明する。Fetch and Carry タスクを行う際の必要技術は、“グローバル座標系：自己位置・姿勢推定” と “ローカル座標系：把持対象物の相対位置・姿勢推定” に大別されるが、それらのうち、各々における位置推定は、以下に述べる手法で実用的なレベルに到達している。

1. グローバル座標系：レーザ SLAM による自己位置推定

HSR や Fetch, PR2 といったロボットはレーザ SLAM [12] という技術を用いて自己位置推定を行っている。これは、周辺の 2 次元環境を予めレーザ計測して 2 次元の環境マップとして表現し使用する。レーザレンジセンサによる計測値と環境マップを比較し、一致度の高い箇所を確率的探索 (パーティクルフィルタ) によって算出し、自己位置を同定するものである。手法としては、環境地図と自己位置同定を同時に推定することができるが、実用的な使用方法ではオフラインで静的に環境地図を準備する方法が一般的である。この手法は、床面に平行に取り付けしたレーザレンジセンサを用いるため、推定される位置は 2 次元座標系に限られる。またセン

サの測距精度が高いため、1 cm 単位の自己位置精度を出すことができる。レーザ SLAM は、全方位に対する標本数 (270 deg. 程度の領域に 1 deg. 刻みの解像度が可能) も多い。そのため、人などの動的障害物がある場合やごみ箱や椅子が環境地図から移動した場合にも頑健である。また Kiva Systems については、床面にバーコードを張ることで、環境中からロボットの位置同定をしている。空間的に広い環境ではバーコードを使用し、空間的に狭い環境ではレーザ SLAM を使うようなハイブリッド形式も可能で、実用的なレベルにある。

2. ローカル座標系 : Faster RCNN による把持対象物の相対位置推定

深層学習手法の 1 つである Faster RCNN [17] は、入力画像から物体の領域箇所の特定とその物体のカテゴリ認識を同時に行うことができる方法である。距離画像センサからの距離情報を加味することで、把持対象物体までの相対的な 3 次元位置を精度良く推定でき、かつ、1 Hz 以上の速度での動作が可能である。ただし、現状では、物体の姿勢までを考慮したうえでの位置推定はできていない。例えば、ペットボトルであれば平面に鉛直に配置し、それを横から見た形状は認識できても、真上から見た形状は正しく認識できない。学習データにも依存するが、物体の姿勢をある程度限定すれば、物体の位置推定は実用レベルにある。

一方、姿勢推定については、グローバル座標系／ローカル座標系ともに、精度よく推定ができる手法がまだ存在しない。そのため、ロボットの位置や把持対象物体の位置が分かっても、それらの姿勢を正確に推定できないため、物体を的確に把持できない問題がある。そこで本研究では、自律的に Fetch and Carry タスクを行う際の必要技術の基礎検討として、“姿勢推定” に注目して取り組んだ。

1.2 本研究の位置づけ

本研究の目的は、ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の“姿勢推定”の高精度化を画像処理技術により解決することである。画像処理技術では、カメラ画像間の幾何学的な対応がとれる局所特徴量を用いた手法が姿勢推定に有効である。局所特徴量を用いて姿勢推定の高精度化を実現するには、局所特徴量の (1) ロボッ

トの姿勢推定への活用方法（第 3 章），（2）把持対象物体の姿勢推定への活用方法（第 4 章，第 5 章），の 2 つを明らかにする必要がある．ここで述べる姿勢の高精度化とは，上述の（1）では姿勢そのものの推定値の高精度化である．一方，上述の（2）では，検出されたマーカの位置・姿勢が目標値誤差以下の条件を満たした場合の“マーカの検出率・認識率の向上”を指す．

次節では，姿勢推定技術の現状の課題を述べ，その上で，局所特徴量を用いた理由について詳細に説明する．また，局所特徴量の活用方法（上述の（1），（2））について，基本的な方針を説明する．

1.3 姿勢推定技術の現状と本研究のアプローチ

本節では，前節で説明した（1）ロボットの姿勢推定及び，（2）把持対象物体の姿勢推定に関する既存技術の現状と問題点，本論文で採用するアプローチ，目標精度について述べる．

1.3.1 姿勢推定技術の現状と問題点

図 1.7 に姿勢推定技術の現状を記す．縦軸は，（1）ロボットの姿勢推定，（2）把持対象物体の姿勢推定で分類し，横軸は，実用性フェーズか，基礎的な研究フェーズかで分類した．まず，ロボットの姿勢推定について説明する．

1. 実用性フェーズ

ジャイロセンサを用いた手法は，安定して姿勢を推定することができるものの，一般にはセンサ誤差の問題があり，長時間にわたって高精度に推定することが難しいという問題がある．

2. 研究フェーズ

局所特徴量を用いた方法は，シーンの中からテクスチャなど特徴的な模様に対して特徴点を抽出し，時刻間の画像で特徴点との対応づけることで高精度に算出することができる．しかし，シーン中の被写体にテクスチャがない場合は，特徴点が得られないため，姿勢をフレーム毎に求めることができないという問題がある．

レーザではなく，RGB-D センサを用いた SLAM 手法も存在し，DVO-SLAM [21]，LSD-SLAM [22] といった方法がこれまで提案されている．RGB-D センサ

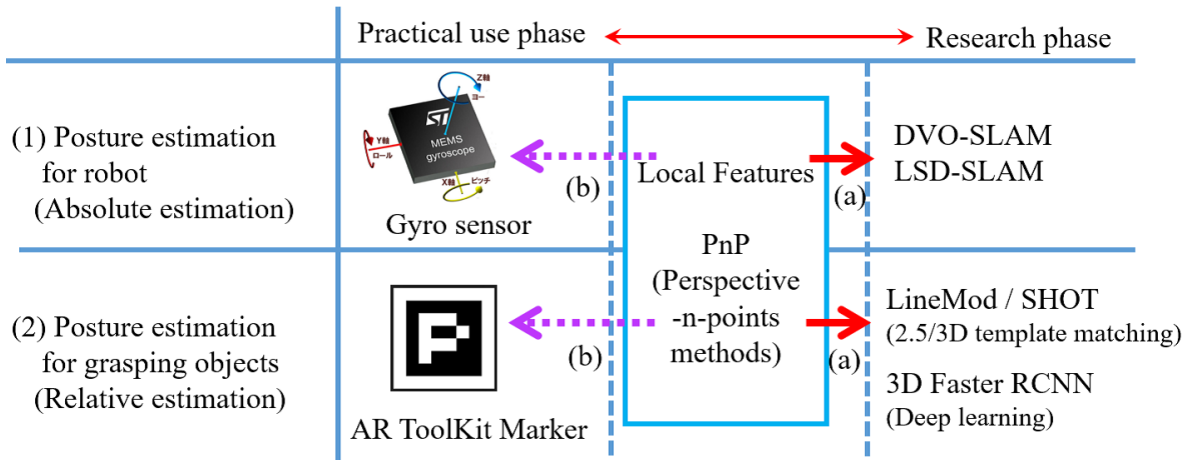


図 1.7: 現状の姿勢推定技術

を用いた SLAM は、位置と姿勢を同時に求めることができるが、レーザ SLAM と比較して精度的に低く、俊敏にロボットが移動した場合は失敗しやすいなどの問題がある。また、ジャイロセンサ同様に慣性航法の演算になるため、時間方向に対しての積算誤差の問題もある。さらに DVO-SLAM や LSD-SLAM の中でも、局所特徴量を用いることで特徴点の対応づけを行い、姿勢と併進の変位量を精度良く求めるといった組み合わせ方法もある（図 1.7 の矢印 (a)）。

次に、把持対象物体の姿勢推定について説明する。

1. 実用性フェーズ

ARToolKit マーカ [16] を用いることで、対象物の姿勢を容易に推定することができる。しかし、マーカはカメラに対して比較的に見えやすい向きと距離に置く必要があり、物体が様々な姿勢で置かれた場合などは検出が難しいという問題がある。

2. 研究フェーズ

局所特徴量を用いた姿勢推定方法は、事前に物体のテクスチャから特徴点を抽出し、それとカメラ視点の情報を対でデータベースに登録しておく。入力画像から得られる特徴点と登録されたデータベース内の特徴点の間で照合し、登録されたカメラ視点の情報を加味して姿勢算出を行う。テクスチャがある物体にしか適用できないという問題がある。

さらに、LineMod [18] や SHOT [24] といった 2.5 次元、3 次元でのテンプレー

トマッチングがある。これらは、物体の 3 次元 CAD モデルを事前に準備し、その様々な姿勢を 2.5 次元 RGB-D 画像（または、3 次元特徴量）テンプレートとして記録し、テンプレートとの照合により姿勢推定を行うものである。SHOT は、局所特徴量を 3 次元点群に拡張したものである（図 1.7 の矢印 (a)）。各物体に対して様々な姿勢のテンプレートと照合するため、計算コストも高い。さらに、特徴的な形状でないと高い検出精度が得られないといった実用面での問題がある。またマーカ同様に、カメラに対して比較的に見えやすい向きと距離に置く必要があり、物体の様々な配置で検出するのは難しいという問題も存在する。

また近年は、深層学習（Deep Learning）の 1 つで画像処理特有のニューラルネットワークである Convolutional Neural Network（CNN：畳み込みニューラルネットワーク）[25] が登場し、一般物体認識の精度が飛躍的に向上することが Large Scale Visual Recognition Challenge（ILSVRC）で示されている。そのうえで、物体の位置推定までを行う Faster RCNN [17] や、それを 3 次元に拡張して、物体の姿勢までを推定する手法 [26] などもある。後者は、検出率 60～80% 前後であり、未知物体の姿勢までを実用的に求めることは依然難しい。

局所特徴量は、上述した問題はあるものの、Computer Vision の発展に大きく貢献した技術である。SIFT（Scale-Invariant Feature Transform）[27] をはじめとした局所特徴量記述子（Local Feature Descriptor）によって、画像間の幾何学的な対応づけが正確にできるようになった点は大きな進歩である。SIFT だけでなく、SURF [28]、BRIEF [29]、ORB [30]、BRISK [31]、CARD [32]、AKAZE [33]、Spectral-SIFT [34] など、現在も改良が重ねられている。他研究への応用例も多く、図 1.7 の矢印 (a) などが該当する。しかし、図 1.7 の矢印 (b) のように局所特徴量をジャイロセンサやマーカに活用した先行研究例はない。

また、ロボットと把持対象物体の姿勢推定の双方ともに、実用性フェーズの課題が、研究フェーズにおいても同様に存在することにも注目する。実用性フェーズの課題の方が、優先度が高い課題と考える。

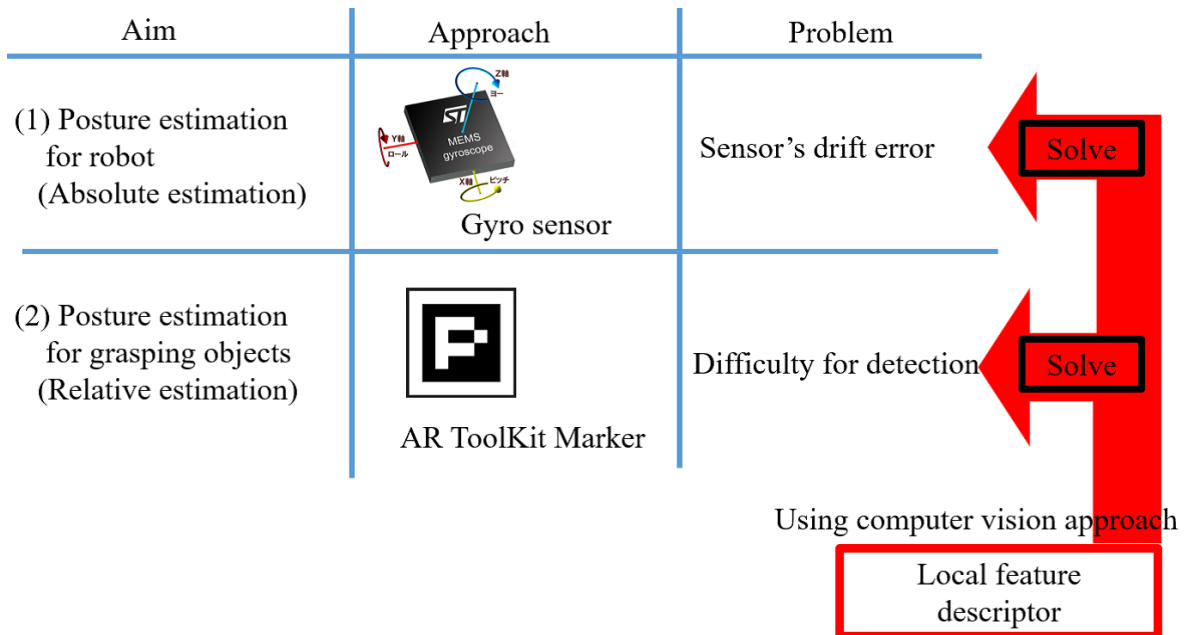


図 1.8: 本研究のアプローチ

1.3.2 本研究のアプローチ

前項で述べた姿勢推定技術の現状と問題点を踏まえ、本研究のアプローチを述べる。図 1.7 (1) のロボットの姿勢推定ではジャイロセンサによる手法が、図 1.7 (2) の把持対象物体の姿勢推定ではマーカによる手法が、現状では最も実用的な方法である。図 1.7 において、実用性フェーズと研究フェーズの中間に位置するのが、局所特徴量を用いた方法である。従来、SLAM や SHOT などの図 1.7 の研究フェーズに位置する方法に局所特徴量を活用した事例はある。しかし、局所特徴量をジャイロセンサやマーカに活用した先行研究例はない。また、時間方向に対しての積算誤差（ジャイロセンサ）、物体の様々な配置で検出するのが難しさ（マーカ）は、同様の課題が図 1.7 の研究フェーズに位置する方法にも存在する。SLAM は積算系の演算であり、LineMod, SHOT, 3D Faster RCNN では、物体の配置による物体の見え方によって難しさが変わる。

そこで本研究では、実用性フェーズの課題を優先し、これらの手法に改良を加えることを考え、図 1.8 に示すアプローチを採る。従来、局所特徴量をジャイロセンサやマーカに活用した先行研究例はないため、局所特徴量をうまく活用することで従来手法の問題点解

決を図る方法を選択する。その構成は、

$$\begin{cases} \text{ジャイロセンサ} & + & \text{局所特徴量} \\ \text{ARToolKit マーカ} & + & \text{局所特徴量} \end{cases}$$

である。具体的には、ジャイロセンサに対しては、センサからの積算誤差を局所特徴量を用いて除去することを試みる。また ARToolKit マーカに対しては、マーカの様々な配置・距離における検出・認識の困難さが課題としてあり、マーカの検出方法に局所特徴量のキーポイント検出の考え方を導入した“検出・認識しやすい”マーカを開発する。いずれも従来の実用的な手法に対して、局所特徴量を活用することで姿勢の高精度化を図る手法である。

1.3.3 目標精度

ここでは、(1)、(2) の各々について、その目標精度について述べる。

図 1.8 (1) のロボットの姿勢推定を考える際、使い得る最高級のジャイロセンサは光ファイバジャイロ（以降、FOG (Fiber Optic Gyro)）である。1 時間使用時のドリフト誤差は約 3 deg. 程度であり、航空機などに使用されている大がかりかつ高価なセンサである。一方、一般にロボットに搭載されるジャイロセンサは MEMS 型の振動型ジャイロセンサであり、小型かつ安価なセンサである。これらは、FOG とは計測方式・精度・コストとも大きく異なる。経験的に、FOG と同等の精度があれば、所定の箇所への到達位置精度に対して十分であると考え、ロボットの把持タスクでは 1 回の動作あたり、部屋の移動に数分（～2 分程度）、アーム制御に最大 1 分の計 3 分程度の時間を要する。そのため目標精度として、2 分間（移動時間）において FOG と同水準を実現できることが望ましい。

一方、**図 1.8 (2)** の把持対象物体の姿勢推定では、位置のずれ量を 5 mm 以下、角度誤差を 5 deg. 以内を目標とする。“International Workshop on Recovering 6D Object Pose” のチャレンジコンテストでは、位置ずれ 5 cm, 角度誤差 5 deg. 以内が一般である。しかし、筆者が実際に認識ピッキングロボットを用いた経験では、角度誤差については把持タスクの手先位置精度に対して十分であるが、位置のずれ量については対象物体が小さくなった際に問題になる。そこで、本研究では、位置のずれ量の目標値を 5 mm 程度に設定する。

1.4 本論文の構成

以下に、本論文の構成を示す。

第 2 章では、局所特徴量記述子 (Local Feature Descriptor) について説明する。数多く提案されている局所特徴量記述子のうちで、SIFT の位置付けを解説し、本研究で SIFT を利用した理由を説明する。また SIFT を利用した特徴量マッチングについて説明する。さらに予備実験として、視点の違いにおける特徴量の照合性能の評価結果を示し、SIFT 単体の性能を確認する。

第 3 章では、“ジャイロセンサ + SIFT” によるロボットの姿勢推定手法を提案する。ジャイロセンサの問題点、SIFT を利用するうえでの問題点について説明し、高精度化のために両者を組み合わせる方法について説明する。そして、実験結果、考察をまとめる。

第 4 章では、“マーカ + SIFT” による把持対象物体の姿勢推定手法を提案する。従来のマーカについて、その手法と問題点について説明する。そして、マーカの様々な配置に対して、SIFT を活用することでマーカを検出しやすくできることを説明する。そして、実験結果、考察をまとめる。

さらに、第 5 章では、“マーカ + SIFT” による把持対象物体の姿勢推定手法において、検出しやすいだけでなく、認識しやすくなる改良をおこなった。任意の距離・配置においてもマーカを検出・認識できるように、ドットのみ (ドットの配置とスケール) でマーカを表現する方法と Spectral SIFT [34] を活用する方法を説明する。そして、実験結果、考察をまとめる。

最後に、第 6 章では、全体を通したまとめと今後の展望を述べる。

第 2 章

姿勢推定に用いる局所特徴量記述子の事前性能評価

本研究の目的は，ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の姿勢推定の高精度化を画像処理技術により解決することである．画像処理技術では，カメラ画像間の幾何学的な対応がとれる局所特徴量を用いた手法が姿勢推定に有効である．本章では，姿勢推定の高精度化に対して，局所特徴量を活用することの有効性を事前検証する．具体的には，局所特徴量の向きに対する単体性能を確認する．

局所特徴量を用いた姿勢推定手法は，画像間の照合を実現する強力な手法 [23] であり，物体認識，3 次元復元，画像検索など，様々に応用されている．局所特徴量の算出は，キーポイントの検出と特徴量記述の 2 段階の処理から構成される．そのどちらもが，幾何学的な変化（拡大・縮小・回転・アフィン変換）や照明変化に対して頑健な設計が施されている．本章では，局所特徴量記述子の研究動向と本研究で用いた SIFT 特徴量 [27] について説明し，視点の角度変位に対する頑健性を事前評価する．

2.1 局所特徴量記述子の研究動向

2004 年に提案された SIFT 特徴量（以降，SIFT）を皮切りに，様々な局所特徴量記述子が開発されている．特徴抽出の高速化を行った SURF [28]，特徴量間の誤対応を低減するために特徴量の記述性能を改善した PCA-SIFT [35]，アフィン不変な対応点探索かつ回転に頑健な ASIFT [36]，連続的なスケールでキーポイント検出が可能な Spectral-SIFT [34]，ボケに頑健かつ特徴量抽出の高速化を実現した AKAZE [33]，などがある．AKAZE

や Spectral-SIFT では、スケールスペース解析について更なる追及がなされ、解析的にキーポイントを検出できるようになった点が大きな発展である。

上述した局所特徴量記述子は、特徴量表現としては実数値を採る。一方、メモリ消費の削減と対応点探索の高速化のために、2 値表現の特徴量も開発されている。画素間の輝度差から 2 値特徴量を生成する BRIEF [29]、それをスケールと回転の不変性に発展させた BRISK [31] や ORB [30] などである。特徴量表現では、2 値表現化が進み、Hamming 距離や Binary Hashing の考え方を採用したものが流行である。

様々な局所特徴量記述子がある中で、対応点マッチングの性能評価では、AKAZE、もしくは SIFT が最も安定しているとされる [33]。また、キーポイント検出では、Spectral SIFT が最も頑健な検出手法である。そこで本研究では、SIFT をジャイロセンサと組み合わせることで高精度な姿勢推定を実現する手法を提案する（第 3 章）。更に、SIFT 及び Spectral SIFT をマーカと組み合わせることで検出しやすいマーカパターンを提案する（第 4 章、第 5 章）。

2.2 SIFT (Scale-Invariant Feature Transform)

姿勢推定に利用する SIFT (Scale Invariant Feature Transform) について述べる。SIFT は、特徴点の検出と特徴量の抽出を行う手法 [37] である。Lindeberg は、スケールスペースを用いることで画像の構造を解析し、blob の検出と自動スケール選択を行う手法 [38] を提案した。スケールスペースを用いた解析とは、帯域幅が可変の Gaussian フィルタで信号波形をぼかすことにより、様々な尺度で信号の基本構造の解析を行うことを指す。Lowe は局所領域の特徴量記述という考えを拡張し、スケールスペースを用いることで、画像のスケール変化や回転に不変な特徴量を記述する SIFT を提案した [27, 39]。SIFT は、回転・スケール変化等に不変な特徴量を表現するため、画像の照合などに用いられている。SIFT の処理は、特徴点（以下、キーポイントと呼ぶ）の検出 (detection) と特徴量の記述 (description) の 2 段階からなり、各処理は以下の流れとなる。

1. キーポイントの検出 (detection)

- a. Scale space extrema detection (2.2.1 項)
- b. Keypoint localization (2.2.2 項)

2. 特徴量の記述 (description)

- $$\begin{cases} \text{a. Orientation assignment (2.2.3 項)} \\ \text{b. Keypoint descriptor (2.2.4 項)} \end{cases}$$

「1a. Scale space extrema detection」では、スケールとキーポイントを検出し、「1b. Keypoint localization」では、1a. で検出されたキーポイントから特徴点として適さない点を削除し、その後サブピクセル推定を行う。「2a. Orientation assignment」では、回転に不変な特徴を得るためにキーポイントの向きを求める。「2b. Keypoint descriptor」では、2a. で求めた向きに基づいてキーポイントの特徴量を記述する。以降に各処理の詳細を述べる。

2.2.1 スケールスペースにおける極値探索

1 段階目のキーポイント検出では、以下に述べる DoG (Difference-of-Gaussian) 処理を用いてスケールスペースにおける極値探索を行い、キーポイントの位置とスケールを決定する。キーポイント候補点は、スケールが異なる Gaussian 関数 $G(x, y, \sigma)$ と入力画像 $I(u, v)$ を畳み込んだ平滑化画像 $L(u, v, \sigma)$ の差分 (DoG 画像) から求める。

$$L(u, v, \sigma) = G(x, y, \sigma) * I(u, v) \quad (2.1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2.2)$$

上記の 2 式から、DoG 画像は以下の式より算出する。

$$D(u, v, \sigma) = L(u, v, k\sigma) - L(u, v, \sigma) \quad (2.3)$$

ここで、 u と x 、 v と y は、各々 2 次元画像の x 方向、 y 方向の画素位置を指す。標準偏差 σ を変数として k 倍し、ぼかす度合を変化させていく。異なる σ 間での差分画像を DoG 画像と呼び、SIFT ではこの DoG 画像の極値を探索する。この様子を図 2.1 に示す。これは、縦・横の 2 次元の画像情報に対して大きさ (スケール) の次元も加え、スケールの次元に対してパラメータを割り当てたときに極値があれば、それを特徴点の候補とするという考え方である。このスケールの極値は、同じ画像パターンにおいて画像の大きさが 2 倍になれば、スケールの極値も 2 倍になる性質を持っている。このため、キーポイントは画像の拡大・縮小 (スケール変化) に対して同じ場所に現れる。

極値検出の方法は、DoG 画像の注目画素と周辺画素間で値の比較を行う。これは、注目画素のスケールで 8 点、1 つ上のスケールで 9 点、1 つ下のスケールで 9 点の計 26 点

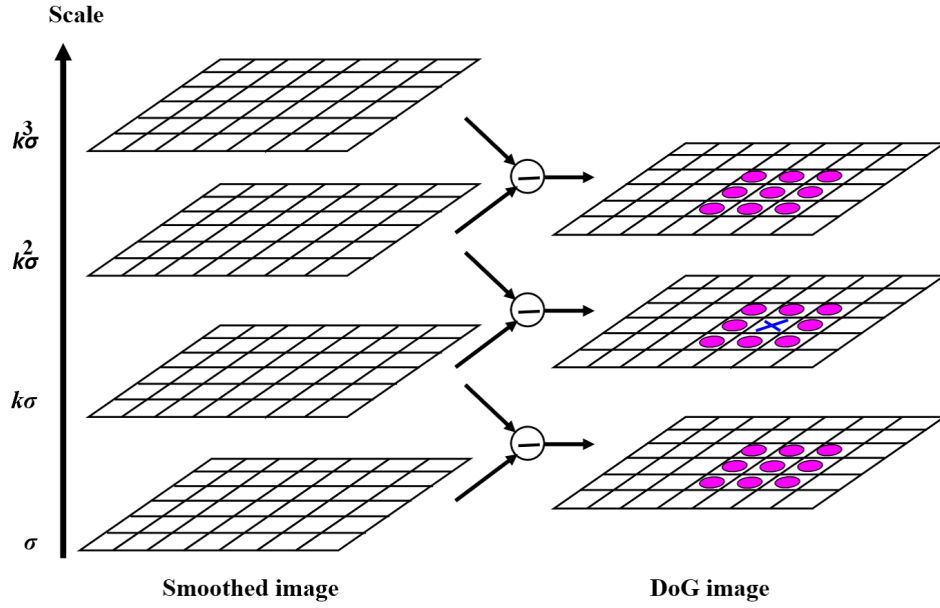


図 2.1: DoG 処理と極値検出

と比較し，注目画素が極大または極小の場合をキーポイント候補にする．図 2.1 にその様子も示す．

2.2.2 キーポイントの検出

キーポイント候補からキーポイントの選出を行う．まず，検出されたキーポイントから特徴点として適さない点（エッジの上にある点）を削除する．次に，キーポイントの位置精度を高めるためサブピクセル推定を行う．

エッジ上に存在するキーポイント候補点の削除方法について述べる．Harris コーナ検出法 [40] を利用する．キーポイント候補点における 2 次元 Hessian 行列 H を計算し，主曲率を求める． D_{xx} は，キーポイント候補位置における x 方向の DoG 出力値の 2 次微分を示す．

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \quad (2.4)$$

ここで，Hessian 行列から求められる第 1 固有値 α と第 2 固有値 β ($\alpha > \beta$) とする．この α と β の大小関係によって，キーポイント出現箇所の性質を判定することができる．これを図 2.2 に示す．つまり， $\gamma = \alpha / \beta$ の値が大きければエッジ上の点と判定し，候補から除く．SIFT では，エッジでない輝度の平坦な部分のみをキーポイントにする．

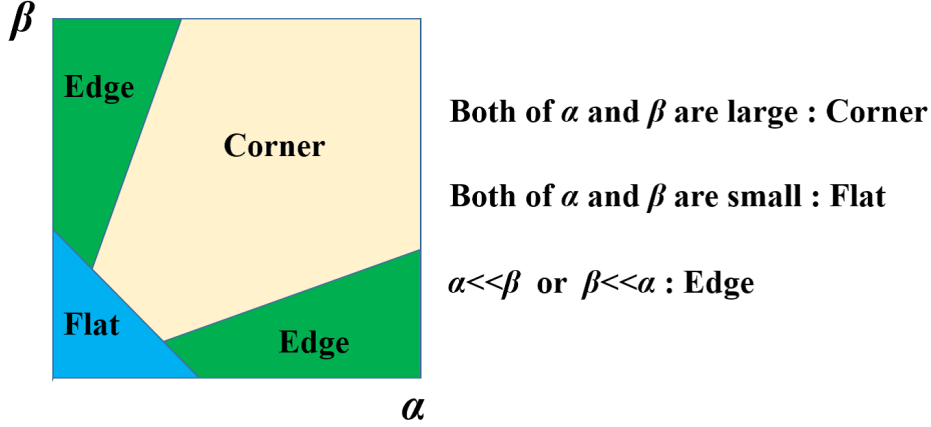


図 2.2: Hesse 行列の固有値が示すキーポイントの性質

ただし，第 4 章，第 5 章で示す提案方法では，エッジのみを選別する．その理由については，各章を参照されたい．

次に，キーポイントの位置精度を高めるためのサブピクセル推定について述べる．ここでは，パラボラフィッティングと呼ばれる手法を用いる．これは，2 次関数を当てはめ，対称軸位置をサブピクセル推定値とするものである．ある点 $\mathbf{x} = (x, y, \sigma)^T$ における DoG 関数 $D(\mathbf{x})$ について 2 次の Taylor 展開を行う．

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (2.5)$$

\mathbf{x} の導関数を 0 として解くと，以下の式が得られる．

$$\frac{\partial^2 D}{\partial \mathbf{x}^2} \hat{\mathbf{x}} = - \frac{\partial D}{\partial \mathbf{x}} \quad (2.6)$$

この式は次のように表現できる．

$$\begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial x\sigma} \\ \frac{\partial^2 D}{\partial yx} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial y\sigma} \\ \frac{\partial^2 D}{\partial \sigma x} & \frac{\partial^2 D}{\partial \sigma y} & \frac{\partial^2 D}{\partial \sigma^2} \end{bmatrix} \begin{bmatrix} x \\ y \\ \sigma \end{bmatrix} = - \begin{bmatrix} \frac{\partial D}{\partial x} \\ \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial \sigma} \end{bmatrix} \quad (2.7)$$

これから以下のように変形し，キーポイント候補点のサブピクセル位置とスケールを求めることが可能である．実際の計算では，演算の反復数の上限値を決めて，サブピクセル位

置の修正を行う．本論文では，反復数を5回とした．

$$\begin{bmatrix} x \\ y \\ \sigma \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 D}{\partial x^2} & \frac{\partial^2 D}{\partial xy} & \frac{\partial^2 D}{\partial x\sigma} \\ \frac{\partial^2 D}{\partial yx} & \frac{\partial^2 D}{\partial y^2} & \frac{\partial^2 D}{\partial y\sigma} \\ \frac{\partial^2 D}{\partial \sigma x} & \frac{\partial^2 D}{\partial \sigma y} & \frac{\partial^2 D}{\partial \sigma^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial D}{\partial x} \\ \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial \sigma} \end{bmatrix} \quad (2.8)$$

2.2.3 キーポイントの向きの算出

検出したキーポイントに対して，特徴量記述を行う．まず，検出したキーポイントの向きを求める．具体的には，平滑化画像 $L(u, v, \sigma)$ の勾配強度 $m(u, v)$ と勾配方向 $\theta(u, v)$ を次式より算出する．

$$m(u, v) = \sqrt{f_u(u, v)^2 + f_v(u, v)^2} \quad (2.9)$$

$$\theta(u, v) = \tan^{-1} \frac{f_v(u, v)}{f_u(u, v)} \quad (2.10)$$

$$f_u(u, v) = L(u + 1, v) - L(u - 1, v) \quad (2.11)$$

$$f_v(u, v) = L(u, v + 1) - L(u, v - 1) \quad (2.12)$$

次式より勾配強度 $m(x, y)$ と勾配方向 $\theta(x, y)$ から勾配ヒストグラム h を作成する．

$$h = \sum_x \sum_y w(x, y) \delta[\theta', \theta(x, y)] \quad (2.13)$$

$$w(x, y) = G(x, y, \sigma) m(x, y) \quad (2.14)$$

$$\delta[\theta', \theta(x, y)] = \begin{cases} 1 & (\theta' = \theta(x, y)) \\ 0 & (\theta' \neq \theta(x, y)) \end{cases} \quad (2.15)$$

勾配ヒストグラム h は，横軸が36方向に量子化された向き，縦軸が各向きに対する強度を示す．勾配強度 $m(x, y)$ に Gaussian 窓 G をかけることで，キーポイント中心ほど，勾配強度の強さがそのまま反映される．また δ は，Kronecker の δ である． θ' が勾配方向 $\theta(x, y)$ と同じ向きであったときのみ1を出力する．このように作成された勾配ヒストグラムの中で，最も強度が高いピーク値の向きをキーポイントの向きとして採用する．

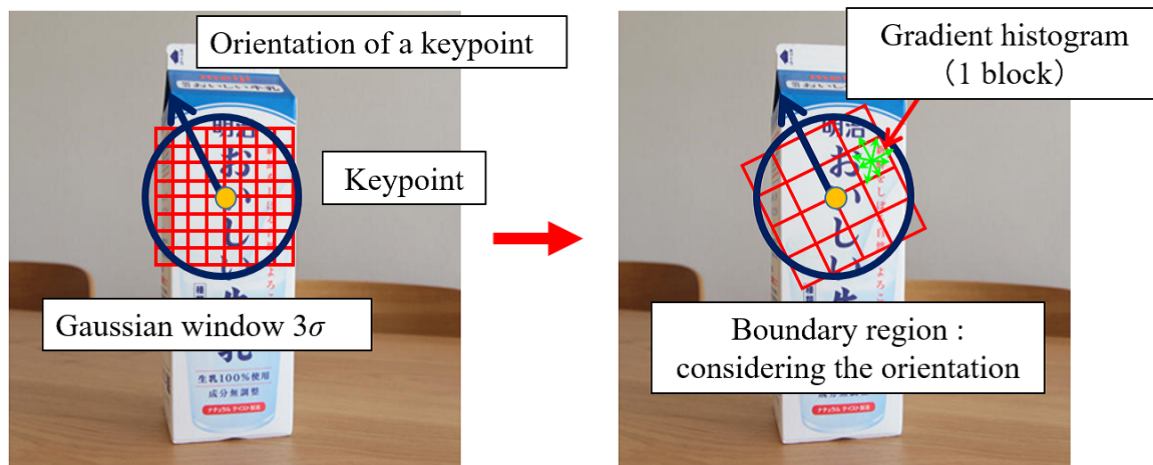


図 2.3: キーポイントの向きとスケールを考慮した特徴量記述

2.2.4 特徴量記述子の表現

特徴量は、128 次元ベクトルで表現される。これは、キーポイントの向きが決まった後、その周辺領域の勾配情報を再度ヒストグラムとして表現したものになる。周辺領域を考慮して、キーポイントの向きを 4×4 ブロックに分割し、各ブロックを 8 方向の勾配ヒストグラムで表現する。図 2.3 にその様子を示す。キーポイントの向きを考慮した勾配ヒストグラムのため、回転の変化に対して不変な特徴量になる。また周辺領域の大きさは先ほど求めたスケール σ に基づいて決定する。そのためスケールの変化に対しても不変な特徴量になる。

2.3 特徴量マッチング

SIFT 特徴量の性質は、再現性が高く、拡大・縮小や回転に対して影響を受けにくい。同じ物体であれば、同じ特徴量ベクトルを複数有するため、その対応づけ（マッチング）ができれば画像検索などが可能になる。マッチングの方法は多く開発されており、KD 木 [41] や KD 木を拡張した FLANN [42] といった手法がある。ここでは、本論文で用いる KD 木について説明する。

特徴量マッチングでは、高次元の特徴量を高速に対応づけることが重要である。総当たり的な方法で対応点探索を行った場合の計算量は、 n_i を i フレームの画像にある特徴点の

数とすると,

$$\sum_i (n_i \sum_{j \neq i} n_j) \quad (2.16)$$

と非常に大きくなり, 特徴点の数が多い場合には非現実的である. KD 木は 2 分探索を K 次元空間へ一般化した方法であり, 分散が最大となる次元を選択し, その次元における中間値でデータを分割する. 各葉ノードが保持するデータ数がしきい値以下になるまでこれを繰り返す. N 個の特徴点に対して, KD 木の深さは, $\log N$ であるから, KD 木による近傍点探索の計算量は, 以下のように改善される.

$$\sum_i (n_i \sum_{j \neq i} \log n_j) \quad (2.17)$$

KD 木の作成方法を説明する. 最も分散が大きい軸で, 中央値を基準にデータを分割していく. 図 2.4 に 2 次元の例を示す. まず, 点 f で 2 分割し, 分割された領域ごとに再度分散が大きい軸とその中央値を求める. この場合は, 点 e と点 g が該当する. この操作を反復し, 最も細分化された領域に該当するデータの数 (葉ノードの数) がしきい値以下になった場合に分割を終了する. この例では, データ属性が 5 つになる. 言い換えれば, 5 個の領域にクラスタリングされたことになる. 同じ領域にクラスタリングされたデータは, 近傍データとして識別される.

SIFT 特徴量の場合は, これを 128 次元空間 (特徴量の次元) で実施する. また, 時系列の画像フレーム間で特徴点マッチングする場合は, まず時刻 $t-1$ における画像から得

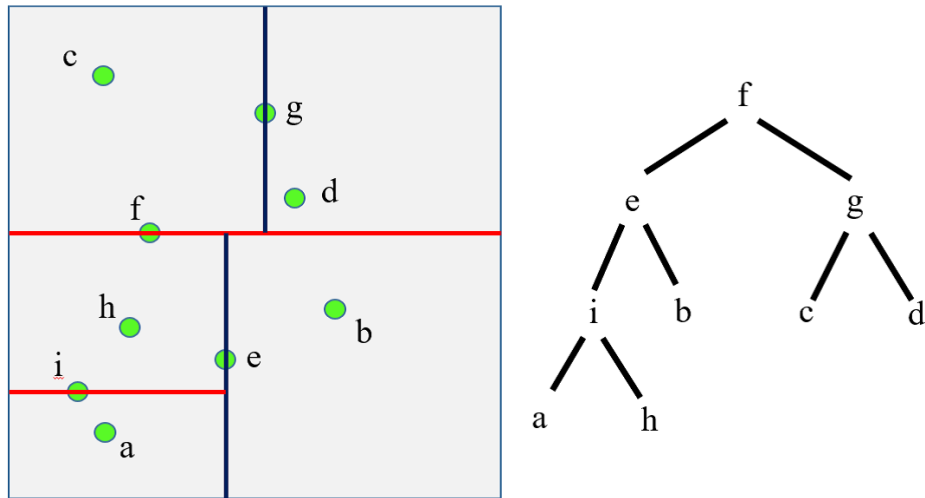


図 2.4: KD Tree の作成例

られた特徴点群で KD 木を構成する．時刻 $t-1$ に得られた各特徴点が属するクラスを判断できる．同様に，時刻 t における画像から得られた特徴点群を，時刻 $t-1$ において作成した KD 木を参照し，属するクラスを確認する．同じクラスに該当する時刻 $t-1$ の特徴点が最近傍点候補となる．この候補の中で，特徴量誤差が最小のものを最近傍点とする．

2.4 SIFT の性能評価試験

実際に SIFT による特徴点のマッチング性能を事前評価した．SIFT 特徴量と KD 木を用いた結果を図 2.5 に示す．カメラ前方 1 m に物体を配置し，カメラを水平方向（ヨー方向）に回転角度を変えて，画像を計測した．正面方向を 0 deg. とし，右方向を正，左方向を負とした．結果から，キーポイントそのものはテクスチャがある箇所に反応していることが分かる．また対応点数は，5 deg. の角度変化で 87 点，30 deg. の角度変化で 46 点の対応点が得られた．最低でも 8 点の対応が得られれば，幾何学的な変化量は求まるた

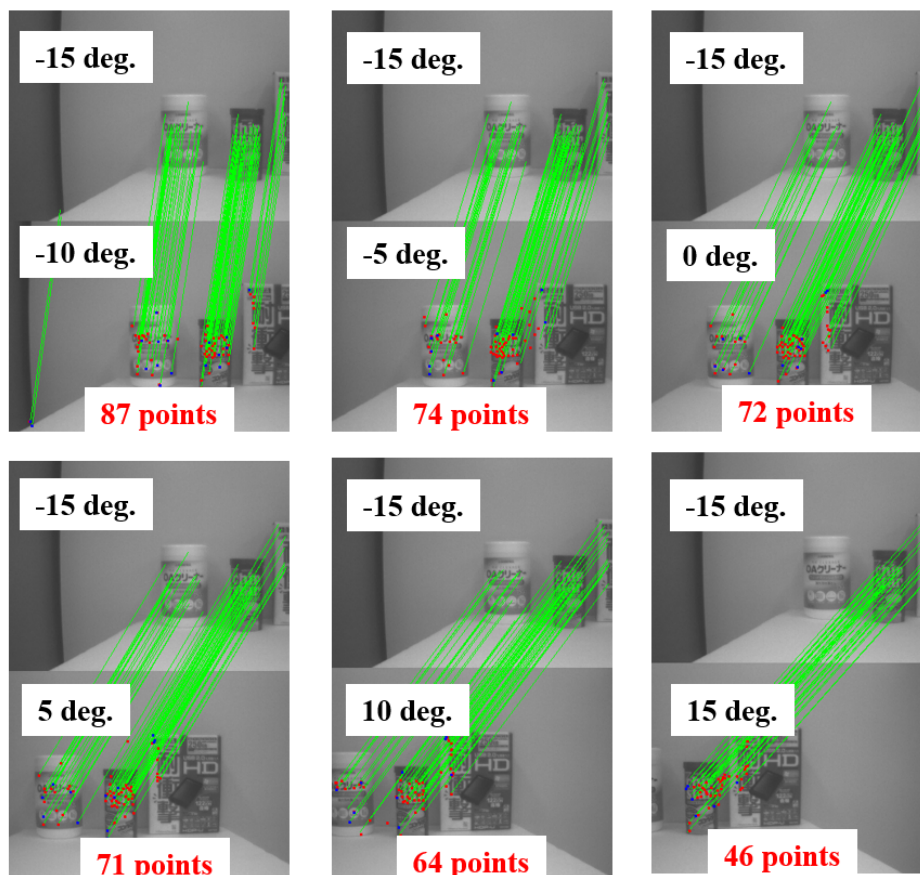


図 2.5: 角度変位に対する SIFT 特徴量のマッチング性能

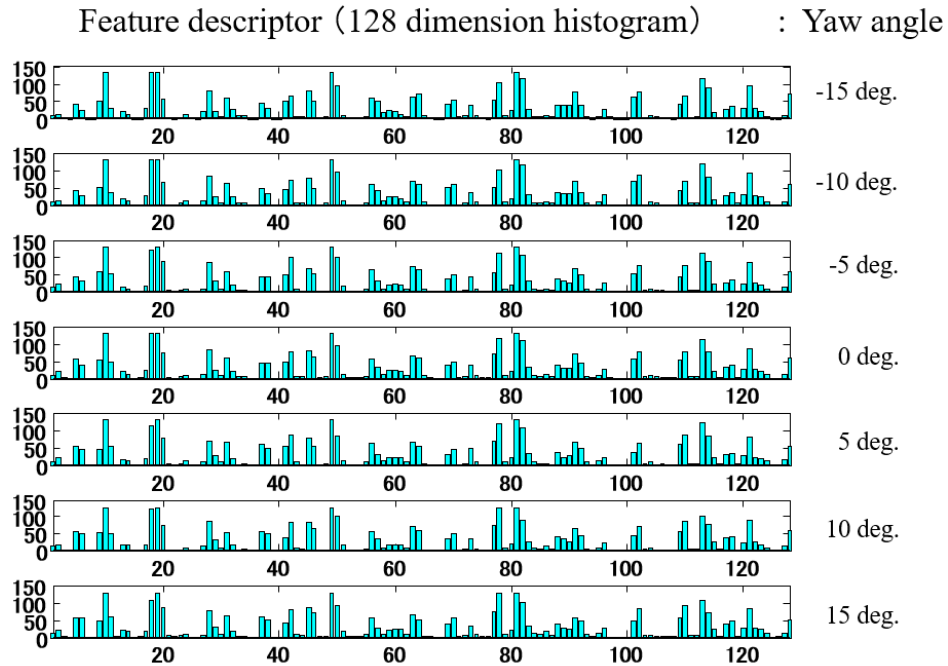


図 2.6: 角度変位に対する特徴量ヒストグラムの変化

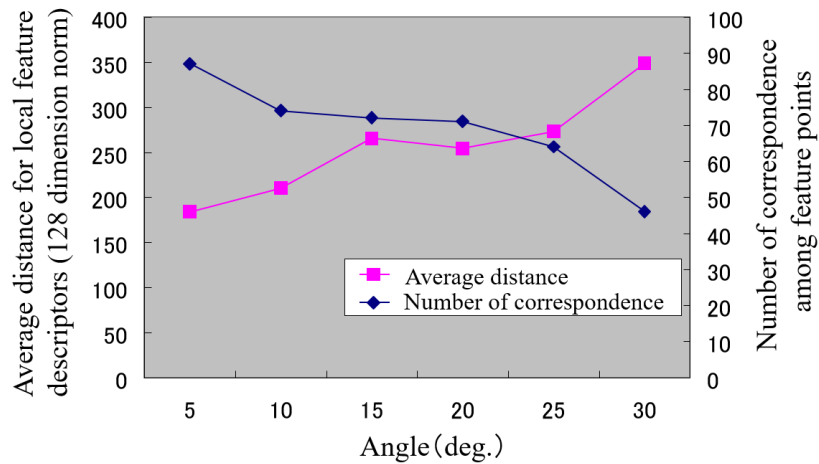


図 2.7: 角度変位に対する特徴量距離と対応点数

め、十分な精度である。

更に、対応が得られた特徴量そのものを表示して比較した。その様子を図 2.6 に示す。-15 deg. から 15 deg. の範囲で、視点が異なっているが、全て類似した波形になっている。またマッチングが得られた点間における 128 次元特徴量の距離を図 2.7 に示す。角度差 30 deg. 近辺から距離が増加し、対応点が減少していることが分かる。現実には 1 フ

レームの間に 30 deg. もカメラの向きが変わるほど，ロボットは速い動作をしないため，SIFT を活用できると判断した．

2.5 まとめ

画像処理技術を用いた姿勢推定の高精度化において，局所特徴量を活用することの有効性を事前検証した．まず局所特徴量の研究動向と本研究で用いた SIFT 特徴量およびその照合方法について説明した．次に，視点の角度変位に対する“SIFT 特徴量の頑健性”を確認した．約 30 deg. 程度の視点の変位にまで特徴点照合ができるため，姿勢の高精度化に応用できると結論した．

第 3 章

局所特徴量とジャイロセンサの相補的な活用による姿勢推定

本研究は、ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の姿勢推定の高精度化を画像処理技術により解決する。前章では、画像処理技術として、局所特徴量の活用が有効であることを事前検証した。これに対して、本章では、局所特徴量を活用して、ロボットの姿勢推定を高精度化する方法について提案する。ジャイロセンサの問題点、局所特徴量を利用するうえでの問題点を明らかにし、各々の組み合わせ方を工夫することで、ロボットの姿勢推定の高精度化を図る。

具体的には、移動ロボットの高精度な姿勢推定手法として、MEMS 型の振動型ジャイロセンサと単眼カメラを組み合わせたシステムを提案する。ジャイロセンサを用いて姿勢推定を行う場合、温度などの影響によりドリフト誤差が発生するため、時間経過に対して姿勢推定精度が低下する。一方、カメラを用いた姿勢推定手法では、ドリフト誤差が発生しない反面、特徴点を得られない場合があるため継続的な姿勢推定は難しい。提案手法では、ジャイロセンサの出力とカメラ画像から算出した特徴点（SIFT 特徴量）より求めた姿勢出力を拡張 Kalman フィルタの枠組みで組み合わせ、上記センサの各々がもつ問題の解決を図る。特に、カメラから姿勢が求まらない場合の対応、カメラから求める姿勢の精度向上方法を提案し、実用面を考慮した組み合わせ方法を示す。提案手法により姿勢の推定精度が向上することを実験により示す。

3.1 ロボットの姿勢推定に関する先行研究

自動車やロボットなどの移動体において、自らの姿勢を知ることは重要である。自動車では、横滑り安定装置などの運動制御に姿勢の情報が利用されている [43]。また、ロボットにおいてもヒューノイドロボットの倒立制御 [44] や Segway [45] を代表とする 2 輪搭乗型移動ロボットの倒立制御などに利用されている。姿勢の推定にはジャイロセンサ（角速度センサ）を使うのが一般的である。特に、近年 MEMS（Micro Electro Mechanical Systems）型の振動型ジャイロセンサ [46] が登場し、小型かつ安価になった。その一方で、基板温度などの影響によって誤差が発生する問題は依然残っている。姿勢は角速度の積分計算により算出されるため、時間経過による累積誤差（ドリフト誤差）が大きな問題となっている。

一方、カメラでは SIFT 特徴量 [27] のように回転・縮小・拡大・照明の変化に頑健な特徴量が開発され、カメラ画像間で特徴点の対応付けが正確にできるようになった。これにより、カメラからも姿勢を正確に求められる。しかし実環境では、カメラから特徴点の情報を得られない場合も多く、継続的に姿勢を求めることができない問題がある。

また、ジャイロセンサとカメラを組み合わせた手法も提案されている [47–56]。これらの手法では、まずカメラを用いて、空間中の目印となる特徴点とカメラの幾何関係により姿勢を算出する。そして、カメラから求めた姿勢とジャイロセンサから求めた姿勢の間で Kalman フィルタを構築することにより、姿勢を推定する。しかし、これらの研究のほとんどが人工的な環境での利用を前提としている。文献 [47, 48, 51, 54] は人工的なランドマークを環境中に配置しており、文献 [49, 55] はラインエッジが観察される環境での利用に限られる。特に、自動車やロボットの走行環境を考えれば、人工的な環境を用意するのは難しい。文献 [50, 52, 53, 56] は実環境での実験を行っているが、画像から特徴点が抽出できない場合の議論がなされていない。実環境においては画像から特徴点を安定して抽出できない場合も多く、カメラから姿勢が求まらない場合について議論が必要である。拡張現実（Augmented Reality）では特徴点が検出されない場合、ジャイロセンサの計測値で姿勢を補間している [54, 55, 57]。この間、高精度化は図れないといった問題がある。

表 3.1: ジャイロセンサとカメラの相補的關係

	Camera	Gyro sensor
Strong	If feature-points matching succeeds, high accuracy	Continuous output sampling rate: 10 kHz
Weak	No feature-points or mis-matching sampling rate: 10~30 Hz	Drift error occurs, Low accuracy

3.2 ジャイロセンサとカメラの相補性

ジャイロセンサとカメラの個々の利点と欠点を整理し、表 3.1 に示す。ここで、ジャイロセンサの出力は角速度に時間を乗じた姿勢の変位である。同様に、カメラの出力は時刻が異なる画像間における特徴点の対応から求めた姿勢の変位である。以降、特に明示しない場合には、姿勢の表現は姿勢および姿勢の変位を包括的に指すものとする。

ジャイロセンサから求まる姿勢は精度があまり高くないが、必ず値は出力される。また応答性も良く、数 10 kHz の動作まで計測できる。一方、カメラから求まる姿勢は特徴点の対応づけが正確であれば高精度であるが、特徴点が抽出できない場合は求まらない。また特徴点の対応づけが失敗した場合、求まる姿勢は真値から大きく外れる。更に画像そのもののデータ量が多いことから計算コストが高いため、1 Hz 程度の速さでしか処理できず、応答性が悪い。整理すると、以下に示すように双方の欠点を利点で補えることが分かる。

相補関係 1. 高精度化： カメラの利点⇒ジャイロセンサの欠点

相補関係 2. 定常出力： ジャイロセンサの利点⇒カメラの欠点

ジャイロセンサとカメラが相補的な役割を担うように、相補関係 1. と相補関係 2. の作用を同時に満たすことができれば、先行研究の課題を解決できると考えた。

3.3 提案手法を構成する要素技術：拡張 Kalman フィルタ

まず提案手法で利用した Kalman フィルタ [58–60] について説明する。Kalman フィルタは、最尤推定法を時間方向に拡張させたものである。最尤推定法は“与えられた測定

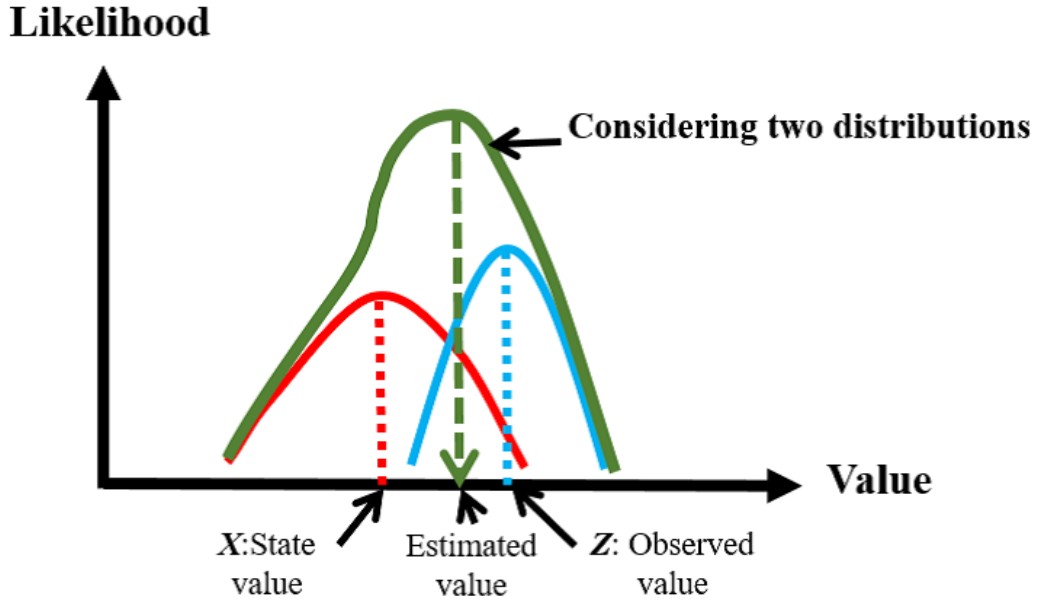


図 3.1: Kalman フィルタの考え方

値 Z に対して、最もありそうな X を選ぶこと”を行う。それは、条件付き確率 $P(X|Z)$ が最大となるような X を選ぶことであり（ここで、 X ：状態量， Z ：測定量）， $P(X|Z)$ を事前確率に変換することで X を算出する方法である。

図 3.1 に Kalman フィルタの考え方を示す。先に説明した $P(X|Z)$ が最大となるような X とは、状態量 X のばらつき（分散）と測定値 Z のばらつき（分散）を考慮した確率分布の極値であることが分かる。また、Kalman フィルタは、観測方程式（最尤推定法）と状態方程式（時間方向拡張）の 2 つから構成される。以下、これら 2 つについて説明する。

3.3.1 観測方程式と最尤推定

M 個の測定値 Z から X （例えば、3 軸のセンサ値：ヨー、ロール、ピッチ）を推定する問題を考える。事前に、 Z と X の関係式（観測方程式）は以下で与えられているとする。

$$Z = HX + v \quad (3.1)$$

ここで、 v は測定誤差で、その平均と分散も与えられているとする。

$$E(v) = 0 \quad (3.2)$$

$$E(vv^T) = R \quad (3.3)$$

また、状態量 \mathbf{X} の平均と分散も与えられていると仮定する.

$$E(\mathbf{X}) = \bar{\mathbf{X}} \quad (3.4)$$

$$E(\mathbf{X}\mathbf{X}^T) = \mathbf{M} \quad (3.5)$$

ここで、与えられた測定値 \mathbf{Z} に対して、最もありそうな \mathbf{X} を選ぶ. これは、 $P(\mathbf{X}|\mathbf{Z})$ を最大となるような \mathbf{X} を選ぶことと同値である. そこで、 $P(\mathbf{X}|\mathbf{Z})$ を事前確率に変換して考える. これは、以下の Bayes の公式を使用すると、

$$P(\mathbf{X}|\mathbf{Z}) = \frac{P(\mathbf{Z}|\mathbf{X})P(\mathbf{X})}{P(\mathbf{Z})} \quad (3.6)$$

この右辺を最大とする \mathbf{X} を選べは良い. ここで、

$$\begin{cases} P(\mathbf{X}) : \text{平均} : \bar{\mathbf{X}}, & \text{分散} : \mathbf{M} \text{ の正規分布} \\ P(\mathbf{Z}) : \text{平均} : \mathbf{H}\bar{\mathbf{X}}, & \text{分散} : \mathbf{H}\mathbf{M}\mathbf{H}^T + \mathbf{R} \text{ の正規分布 (※ 1)} \\ P(\mathbf{Z}|\mathbf{X}) : \text{平均} : \mathbf{H}\bar{\mathbf{X}}, & \text{分散} : \mathbf{R} \text{ の正規分布} \end{cases}$$

を式 (3.6) に代入して、整理すると以下になる.

$$\begin{aligned} P(\mathbf{X}|\mathbf{Z}) &= \frac{|\mathbf{H}\mathbf{M}\mathbf{H}^T + \mathbf{R}|^{1/2}}{2\pi^{n/2}|\mathbf{M}|^{1/2}|\mathbf{R}|^{1/2}} \exp[-0.5(\mathbf{Z} - \mathbf{H}\mathbf{X})^T \mathbf{R}^{-1}(\mathbf{Z} - \mathbf{H}\mathbf{X}) \\ &\quad + (\mathbf{X} - \bar{\mathbf{X}})^T \mathbf{M}^{-1}(\mathbf{X} - \bar{\mathbf{X}}) - (\mathbf{Z} - \mathbf{H}\bar{\mathbf{X}})^T (\mathbf{H}\mathbf{M}\mathbf{H}^T) \\ &\quad + \mathbf{R}^{-1}(\mathbf{Z} - \mathbf{H})\bar{\mathbf{X}}] \end{aligned} \quad (3.7)$$

また、先の (※ 1) の計算は以下に従う.

$$\begin{aligned} E[(\mathbf{Z} - \bar{\mathbf{Z}})(\mathbf{Z} - \bar{\mathbf{Z}})^T] &= E[\mathbf{H}(\mathbf{X} - \bar{\mathbf{X}}) + \mathbf{v}\mathbf{H}(\mathbf{X} - \bar{\mathbf{X}}) + \mathbf{v}^T] \\ &= \mathbf{H}E[(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T]\mathbf{H}^T + E[\mathbf{v}\mathbf{v}^T] \\ &= \mathbf{H}\mathbf{M}\mathbf{H}^T + \mathbf{R} \end{aligned} \quad (3.8)$$

よって、式 (3.6) を最大化する \mathbf{X} は、以下の評価関数 $f(\mathbf{X})$ を最小化する \mathbf{X} を求めることと同値である.

$$f(\mathbf{X}) = \frac{(\mathbf{Z} - \mathbf{H}\mathbf{X})^T \mathbf{R}^{-1}(\mathbf{Z} - \mathbf{H}\mathbf{X})}{2} + (\mathbf{X} - \bar{\mathbf{X}})^T \mathbf{M}^{-1}(\mathbf{X} - \bar{\mathbf{X}}) \quad (3.9)$$

評価関数 $f(\mathbf{X})$ は 2 次関数であるため、以下を解けば良い.

$$\frac{\delta f(\mathbf{X})}{\delta \mathbf{X}^T} = 0 \quad (3.10)$$

式 (3.10) は、次のように整理でき、これを \mathbf{X} について解く．

$$\mathbf{M}^{-1}(\mathbf{X} - \bar{\mathbf{X}}) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{Z} - \mathbf{H}\mathbf{X}) = 0 \quad (3.11)$$

ここで $\hat{\mathbf{X}}$ を $P(\mathbf{X}|\mathbf{Z})$ を最大化する \mathbf{X} とすると、以下の形で導出できる．

$$\begin{aligned} \hat{\mathbf{X}} &= \bar{\mathbf{X}} + \mathbf{P}\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{Z} - \mathbf{H}\bar{\mathbf{X}}) \\ \mathbf{P}^{-1} &= \mathbf{M}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \end{aligned} \quad (3.12)$$

\mathbf{P} は最尤推定値 $\hat{\mathbf{X}}$ の周りの \mathbf{X} の共分散行列になっており、 \mathbf{R} は非負定値なので、推定誤差は必ず減少することがいえる．

3.3.2 状態方程式と誤差伝搬

Kalman フィルタとは、先に示した最尤推定法を時間方向の動的プロセスに拡張したものである．具体的には、 \mathbf{X} が時間方向に次式で遷移することを与える（状態方程式）．

$$\mathbf{X}_1 = \Phi_0 \mathbf{X}_0 + \Gamma_0 \boldsymbol{\omega}_0 \quad (3.13)$$

ここで、 \mathbf{X}_1 、 \mathbf{X}_0 は \mathbf{X} の時刻 $t = t_1$ 、 t_0 の時の値である ($t_1 > t_0$)．また $\boldsymbol{\omega}_0$ は外乱ベクトルであり、その平均を $\bar{\boldsymbol{\omega}}_0$ 、さらにその分散 \mathbf{Q}_0 も与えられるとする．今、 \mathbf{X}_0 の平均 $\hat{\mathbf{X}}_0$ 、共分散 \mathbf{P}_0 であるとする、時刻 $t = t_1$ の時の平均と分散は、次に従う．

$$\bar{\mathbf{X}}_1 = \Phi_0 \hat{\mathbf{X}}_0 + \Gamma_0 \bar{\boldsymbol{\omega}}_0 \quad (3.14)$$

$$\mathbf{M}_1 = \Phi_0 \mathbf{P}_0 \Phi_0^T + \Gamma_0 \mathbf{Q}_0 \Gamma_0^T \quad (3.15)$$

上記は状態方程式に基づく平均値とその誤差分散の拡散を意味する．そして、時刻 t_1 に遷移したのちに測定をしたとすると最尤推定を適用でき、以下のように修正できる．

$$\begin{aligned} \hat{\mathbf{X}}_1 &= \bar{\mathbf{X}}_1 + \mathbf{P}_1 \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{Z}_1 - \mathbf{H}\bar{\mathbf{X}}_1) \\ \mathbf{P}_1^{-1} &= \mathbf{M}_1^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \end{aligned} \quad (3.16)$$

Kalman フィルタは以上の状態方程式と観測方程式の計算を再帰的に反復することで、フィルタとしての効果を発揮する．この様子を図 3.2 に示す．

前述した状態方程式、観測方程式は、共に線形モデルを前提にしている．線形モデルで定式化できない場合は、関係式を現在の状態量の周辺で Taylor 展開し、線形近似を行う．これを拡張 Kalman フィルタと呼ぶ．

3.4 提案手法：ジャイロセンサと SIFT を利用した姿勢推定

3.4.1 システム構成

全体のシステム構成を図 3.3 に、以下に手順を示す。

Step 1: ジャイロセンサでは、センサ値に Low-pass フィルタ (LPF) を施して白色雑音を除いた後、回転行列 \mathbf{R}_{gyro} を算出する。

Step 2: カメラでは、SIFT 特徴量を利用して異なる時刻における画像間の特徴点を対応付け、モーションステレオの原理により回転行列 \mathbf{R}_{img} を算出する。 \mathbf{R}_{img} の導出では、画像処理を工夫する。具体的にはジャイロセンサから求めた \mathbf{R}_{gyro} を用いて、特徴点の誤対応を除去する。更に、画像から基礎行列を算出する際の逐次計算の初期値として \mathbf{R}_{gyro} を利用し、精度の向上を図る。

Step 3: \mathbf{R}_{img} に対して信頼性判定を行い、十分な信頼性があると判定された場合のみ、 \mathbf{R}_{img} を観測量として Kalman フィルタを適用する。Kalman フィルタは、カメラ画像から求めた \mathbf{R}_{img} とジャイロセンサから求めた \mathbf{R}_{gyro} が等しいという拘束条件の下、ジャイロセンサの誤差 \mathbf{X} (バイアス誤差 \mathbf{b} , オフセット誤差 $\delta\boldsymbol{\omega}$) を状態量として推定する。

Step 4: フィードバック機構を設け、推定した誤差量を常時ジャイロセンサ値から取り除く。 \mathbf{R}_{img} に十分な信頼性がないと判定された場合は、Kalman フィルタを適用せず、前時刻までに推定した誤差量をジャイロセンサ値から除く。

Step 5: ロール、ピッチ、ヨー 3 軸の姿勢および角速度をシステムの出力として算出す

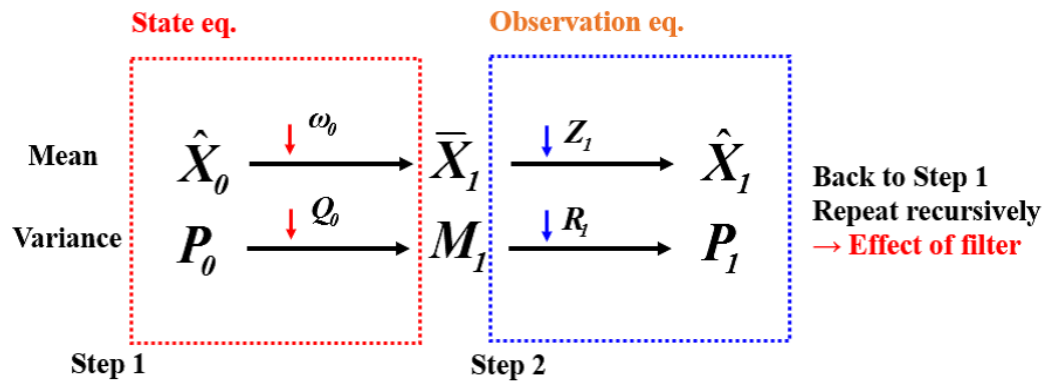


図 3.2: Kalman フィルタ 全体の流れ

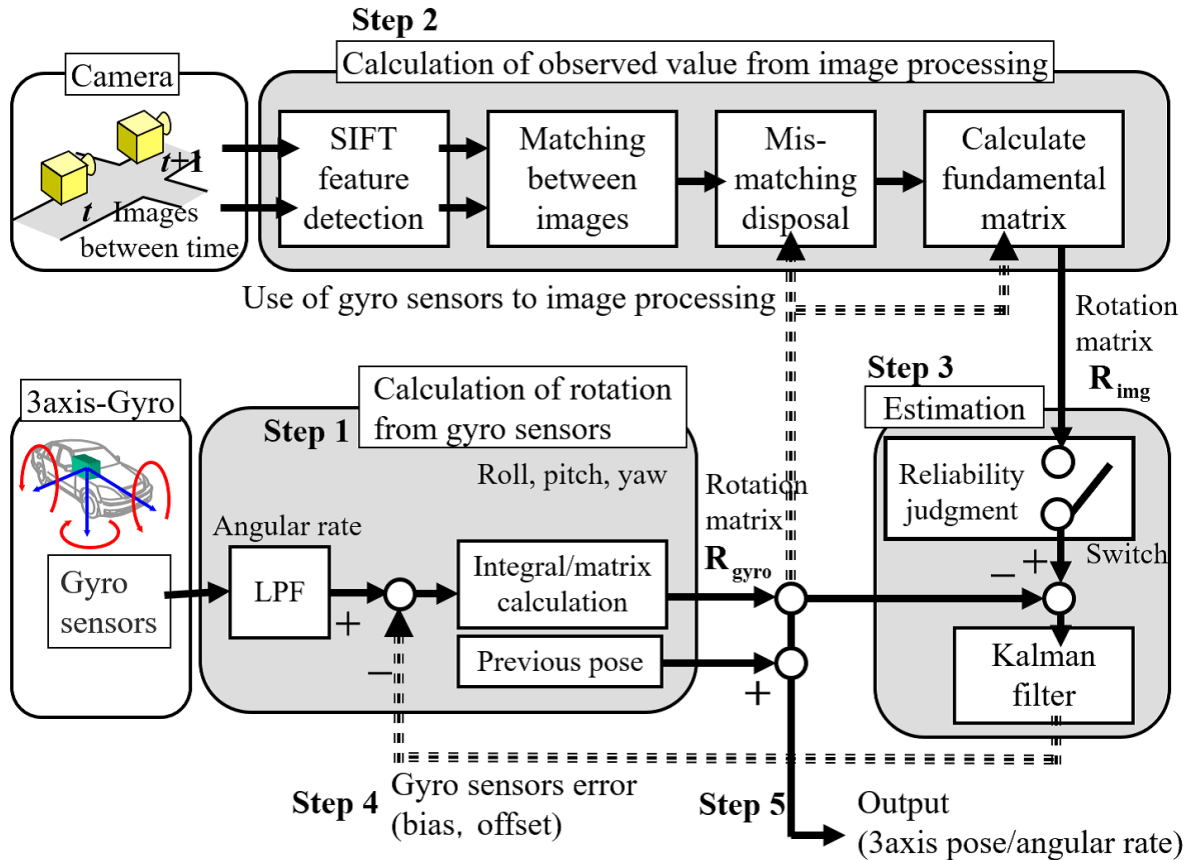


図 3.3: 提案手法のシステム構成

る。

提案手法では、3.2 節で述べた相補関係が機能する。相補関係 1. の作用は、Step 3 で信頼性判定を行い、カメラから精度良く姿勢を算出した状態のみ Kalman フィルタを適用することで実現する。一方、相補関係 2. の作用は、Step 4 で同様に信頼性判定を実施し、姿勢が十分な精度で求まらない場合は Kalman フィルタを適用せずに、前時刻で推定された状態量をジャイロセンサの誤差としてそのまま利用することで実現する。ジャイロセンサの誤差として LPF を施した後の低周波成分のみに着目することで、毎時刻カメラから姿勢が求まらなくても、断片的に求まれば、精度が維持できる仕組みである。

以降の各節で、カメラから求まる姿勢の信頼性判定、Kalman フィルタ状態量の設計、ジャイロセンサを利用した画像処理に分けて詳しく述べる。

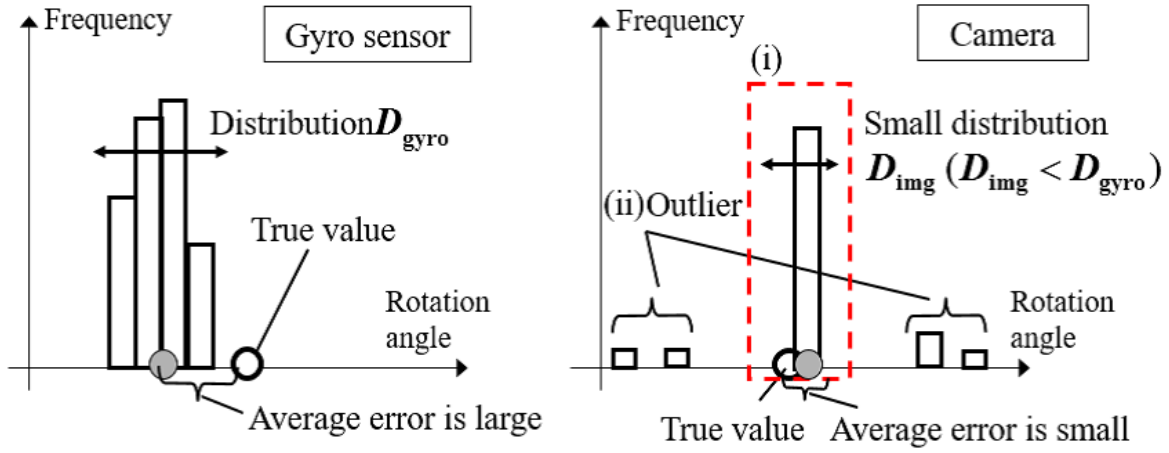


図 3.4: ジャイロセンサとカメラから求まる姿勢の出力頻度分布（概念図）

3.4.2 工夫点 1：カメラから求めた姿勢の信頼性の判定

図 3.4 にジャイロセンサとカメラから求まる姿勢について、出力されるデータ分布の概念図をヒストグラムで示す。ジャイロセンサは常に値が出力されるが、誤差がある。また極端に大きな外れ値もない。一方、カメラから求める姿勢に注目すると、大きく 2 つの場合に分けられる。(1) 特徴点の対応づけが正確で高精度に姿勢が求まる場合、(2) 特徴点が得られず姿勢が算出されない、もしくは特徴点の誤対応によって計算誤差が大きい（外れ値）場合、である。

Kalman フィルタは状態量及び観測量の 2 つの平均と分散に基づいて推定値を決定する。分散が小さければ、推定値はその分布が示す平均の方向に引っ張られる。ここで、平均は毎時刻得られる計測値であり、分散は事前に与える設計値である。場合 1. では、カメラ側の分散を小さくすることで、Kalman フィルタによる精度の改善が期待できる。一方、場合 2. に対応するためには、分散として非常に大きな値をとる必要があり、精度の改善は難しい。そのため、カメラから得られる姿勢の信頼性判定を実施し、場合 1. のみ Kalman フィルタを適用する。場合 2. に対しては、Kalman フィルタの状態量の設計、更新方法を工夫する。詳細は 3.4.3 項で述べる。

カメラから求めた姿勢の信頼性判定について説明する。姿勢は異なる時刻における画像間で特徴点の対応づけを行うことで求める。2 カメラ間の関係は以下の式で表現される。

$$\mathbf{x}^T \mathbf{E} \mathbf{x}' = \mathbf{m}^T \mathbf{F} \mathbf{m}' = 0 \quad (3.17)$$

$$\mathbf{E} = \mathbf{h}_{\text{img}} \times \mathbf{R}_{\text{img}} \quad (3.18)$$

ここで、 \mathbf{x} と \mathbf{x}' 、および \mathbf{m} と \mathbf{m}' は時刻 t , $t+1$ でのカメラ座標系における任意の点の同次座標である。ここで基本行列 \mathbf{E} と基礎行列 \mathbf{F} は、 $\mathbf{F} = \mathbf{A}^{-T} \mathbf{E} (\mathbf{A}')^{-1}$ の関係にある。ただし、 \mathbf{A} と \mathbf{A}' はカメラの内部行列である。時刻間のカメラの変位は \mathbf{E} または \mathbf{F} で表現され、 \mathbf{h}_{img} は時刻間におけるカメラの並進ベクトル、同様に \mathbf{R}_{img} は回転行列である。 $\mathbf{a} \times \mathbf{B}$ は \mathbf{B} の各列ベクトルと \mathbf{a} の外積を並べた行列を算出する演算とする。実際に姿勢を算出する際には2画像間で対応づけられた複数の特徴点の組を利用する。しかし、誤対応や特徴点の検出位置に誤差があるため、全ての特徴点が式 (3.17) を満たすとは限らない。カメラから求めた姿勢、すなわち \mathbf{F} は以下の式によって信頼性を評価できる。

$$C = \frac{1}{n} \sum_{i=1}^n \left[\frac{(\mathbf{m}_i^T \mathbf{F} \mathbf{m}'_i)^2}{l_1^2 + l_2^2} + \frac{(\mathbf{m}'_i^T \mathbf{F} \mathbf{m}_i)^2}{l'_1{}^2 + l'_2{}^2} \right] \quad (3.19)$$

$$l_i = \mathbf{F} \mathbf{m}_i \quad l'_i = \mathbf{F}^T \mathbf{m}_i$$

i は任意の特徴点を指し、 n は特徴点の数を示す。 C は、特徴点とその特徴点に対応するエピポーラ平面までの Euclidean 距離の2乗和の平均値である [66]。式 (3.19) を用いて、カメラから求めた姿勢の信頼性判定をする。具体的には、RANSAC [65] により繰り返し特徴点を選び、計算された値が $C < C_t$ を満たす場合、信頼性が十分であると判定する。ここで、 C_t はカメラから求まる姿勢の誤差分散が ± 0.002 deg. 以下になるように設計し、 $C_t = 0.042$ とした。詳細は 3.4.2.1 に記す。

3.4.2.1 信頼性判定：しきい値の設計方法

誤差伝播法則を元に C の値と \mathbf{R}_{img} の誤差分散の対応付けを行う。Kalman フィルタは観測量、状態量の分散値設計が精度に影響する。分散値は事前に与える設計値である。本手法では、 C の値を介して観測量 \mathbf{R}_{img} の分散値を評価できるため信頼性判定が可能である。つまり、事前設計した観測量の分散値に比べ、 C の値を介して推定される観測量の分散値が大きい場合は Kalman フィルタを実施しない。以下に、 C と \mathbf{R}_{img} の誤差分散の対応付け、および C_t の導出について述べる。

式 (3.17) は $\mathbf{x}_i = [\alpha_1^i, \alpha_2^i, \alpha_3^i]^T$, $\mathbf{x}'_i = [\alpha_1^i, \alpha_2^i, \alpha_3^i]^T$ とすると、以下に変形できる。

$$\mathbf{U}_i \mathbf{E} = 0 \quad (3.20)$$

ここで、 $\mathbf{U}_i = [\xi_1^i, \xi_2^i, \dots, \xi_9^i] = [\alpha_1^i \alpha_1^i, \alpha_1^i \alpha_2^i, \dots, \alpha_3^i \alpha_3^i]$, $\mathbf{G} = [g_{11}, g_{12}, \dots, g_{33}]^T$, g_{ij} は \mathbf{E}

の i 行 j 列の要素を表し、式 (3.22) において \mathbf{G} は $\overline{\mathbf{G}}$ (計算値) として正しいと仮定し、特徴点の座標に誤差がある場合を考えると以下の式が導かれる。

$$\Delta \mathbf{U}_i \overline{\mathbf{G}} = \Delta \eta_i \quad (3.21)$$

Δ は誤差を意味する。誤差量が右辺 $\Delta \eta$ として表れる。特徴点の数 i を計算の便宜上 9 とすると式 (3.23) は式 (3.24) で表現できる。さらに分散の関係は式 (3.25) である。

$$\Delta \mathbf{P} \overline{\mathbf{G}} = \Delta \eta \quad (3.22)$$

$$\text{cov}(\Delta \eta) = \overline{\mathbf{G}}^T \text{cov}(\Delta \mathbf{P}) \overline{\mathbf{G}} \quad (3.23)$$

ここで、 $\Delta \mathbf{P} = [\Delta \mathbf{U}_1, \Delta \mathbf{U}_2, \dots, \Delta \mathbf{U}_9]^T$ である。 $\Delta \mathbf{P}$ の共分散行列を $\text{cov}(\Delta \mathbf{P})$ と書く。次に、先に求めた誤差量 $\Delta \eta$ と \mathbf{G} の関係を考える。 $\mathbf{P} = \overline{\mathbf{P}}$ (計測値) とすると、 $\Delta \mathbf{G} = \overline{\mathbf{P}}^{-1} \Delta \eta$ であるため、分散の関係は以下である。

$$\text{cov}(\Delta \mathbf{G}) = \overline{\mathbf{P}}^{-1} \text{cov}(\Delta \eta) \overline{\mathbf{P}}^{-T} \quad (3.24)$$

式 (3.25), 式 (3.26) をまとめると

$$\text{cov}(\Delta \mathbf{G}) = \overline{\mathbf{P}}^{-1} \overline{\mathbf{G}}^T \text{cov}(\Delta \mathbf{P}) \overline{\mathbf{G}} \overline{\mathbf{P}}^{-T} \quad (3.25)$$

式 (3.27) より特徴点の 3 次元座標のばらつきが基本行列の精度へ与える影響の程度が分かる。次に、 $\text{cov}(\Delta \mathbf{P})$ を求める。画像平面上の特徴点座標を $\mathbf{m}_i = [u^i, v^i]^T$, その分散を σ_u^2, σ_v^2 で表す。 \mathbf{x}_i はカメラの幾何関係より以下の式で表わされる。

$$[\alpha_1^i, \alpha_2^i, \alpha_3^i]^T = \begin{bmatrix} 1 \\ (1/f)(u_i - \text{usize}/2)\text{psize} \\ (1/f)(v_i - \text{vsize}/2)\text{psize} \end{bmatrix} \quad (3.26)$$

ここで、 f は焦点距離、 usize , vsize は画角の大きさ、 psize は画素の大きさである。 ξ_k^i の分散を $V(\xi_k^i)$ で表現すると

$$V(\xi_k^i) = \left(\frac{\delta \xi_k^i}{\delta u} \right)^2 \sigma_u^2 + \left(\frac{\delta \xi_k^i}{\delta v} \right)^2 \sigma_v^2 \quad (3.27)$$

誤差伝播法則より式 (3.28) を u, v で偏微分した形で表現できる。式 (3.27) の $\text{cov}(\Delta \mathbf{P})$ は $V(\xi_k^i)$ を要素にとり以下で表現する。

$$\text{cov}(\Delta \mathbf{P}) = \begin{bmatrix} V(\xi_1^1) & V(\xi_2^1) & \cdots & V(\xi_9^1) \\ V(\xi_1^2) & V(\xi_2^2) & \cdots & \vdots \\ \vdots & \ddots & & \vdots \\ V(\xi_1^9) & V(\xi_2^9) & \cdots & V(\xi_9^9) \end{bmatrix} \quad (3.28)$$

\mathbf{R}_{img} の i 行 j 列の各要素を r_{ij} , 式 (3.18) を $\mathbf{R}_{\text{img}} = \Phi(g_{ij})$ で表現する. さらに $\text{cov}(\Delta \mathbf{G})$ の対角成分を $(\sigma_{G_1}^2, \sigma_{G_2}^2 \cdots \sigma_{G_9}^2)$ とする. 求めたい \mathbf{R}_{img} の分散は式 (3.31) になり, 式 (3.27)~式 (3.30) より σ_u^2, σ_v^2 で記述できる.

$$V(r_{ij}) = \sum_{i,j} \left(\frac{\delta \Phi(g_{ij})}{\delta g_{ij}} \right)^2 \sigma_{G_{3i+j}}^2 \quad (3.29)$$

式 (3.19) の C は分散値 σ_u^2, σ_v^2 と対応する. 式 (3.31) によって観測量の分散を把握できる. ここで, Monte Carlo 法より特徴点の位置をランダムに発生させ, カメラから求まる姿勢 (姿勢変位) の誤差分散が ± 0.002 deg. 以下になるよう事前に C_t を決定した.

3.4.3 工夫点 2 : Kalman フィルタの状態量設計と更新方法

カメラから得られる姿勢の信頼性が十分でないと判定された場合, Kalman フィルタを適用せずに前時刻で推定された状態量を用いる. そこで, 推定する状態量が重要となる.

まずジャイロセンサの特性を把握する. 図 3.5 に静止時において温度変化を与えた際のジャイロセンサの出力結果を示す. 静止状態であるため常時ゼロが真値である. しかし, ± 0.23 deg./s の高周波の白色雑音が常に存在するうえ, 温度の上昇と共に緩やかに増加する低周波の誤差が見られる. 白色雑音は時間平均をとれば 0 であるため, LPF で除くことができる. 一方, 低周波の誤差は温度などの外的要因による時間変化であり, 取り除くのは難しい. そこで, この誤差を Kalman フィルタで取り除くことを試みる.

「カメラから求めた回転行列=ジャイロセンサから求めた回転行列 (姿勢変位)」の拘束条件から姿勢の変位を推定し, 積分することで姿勢を求める. 観測方程式は以下の通りである.

$$\mathbf{R}_{\text{img}} = \mathbf{R}_{\text{gyro}}(\mathbf{X}) \quad (3.30)$$

状態量 \mathbf{X} はジャイロセンサの誤差とする. これは先に示した低周波の誤差である. この低周波の誤差を推定できれば, 信頼性判定により Kalman フィルタを適用しない場合においても, 前時刻までに推定された誤差量で代用できる. これは, 低周波の誤差は, 短時間において変化が少ないからである.

状態量 \mathbf{X} の設計は, 以下の式に示すようにジャイロセンサの値を真値と誤差に分けて行う.

$$\omega_i = (1 + b_i)\hat{\omega}_i + \delta\omega_i \quad (3.31)$$

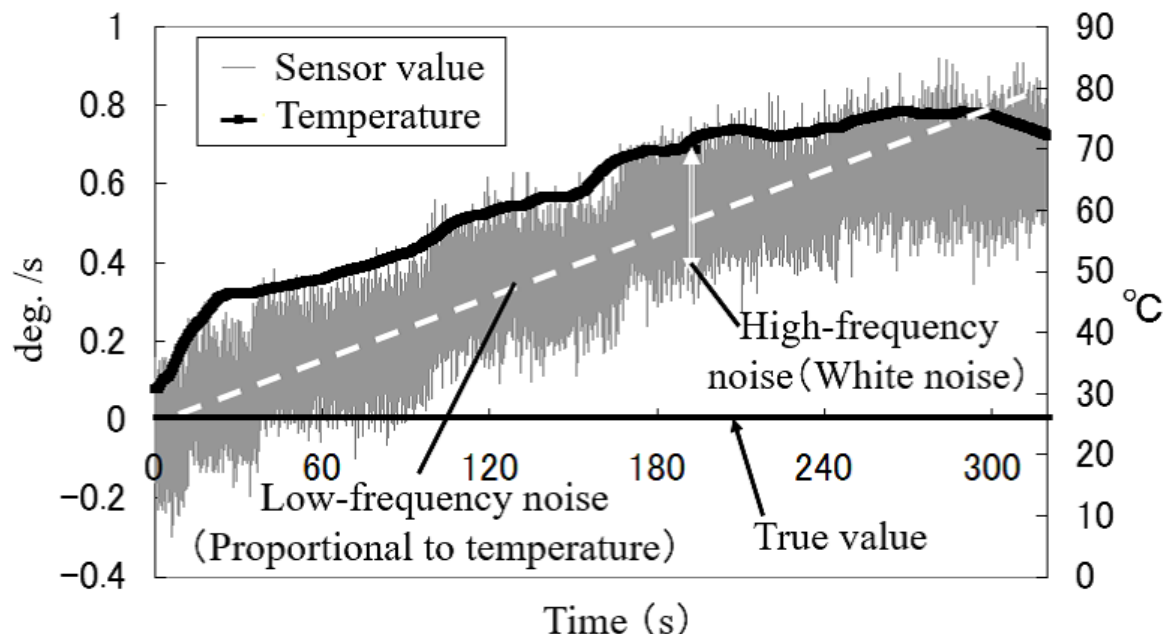


図 3.5: 温度とジャイロセンサの精度

ここで ω_i は角速度の計測値, $\hat{\omega}_i$ は真値, b_i はバイアス誤差, $\delta\omega_i$ はオフセット誤差とする. i はロール, ピッチ, ヨーの 3 軸を意味し, $i \in \{r, p, y\}$ とする. 状態量 $\mathbf{X} = [b_r, b_p, b_y, \delta\omega_r, \delta\omega_p, \delta\omega_y]^T$ と定義する. オフセット誤差 $\delta\omega_i$ は電源の揺らぎなどで生じ, バイアス誤差 b_i は検出した電荷を増幅する場合などで発生しやすい [67, 68]. 半導体は温度依存性があるため, いずれも常時推定する必要がある. \mathbf{X} を拡張 Kalman フィルタの枠組みで推定する [63, 69, 70].

3.4.4 工夫点 3：ジャイロセンサを利用した画像処理

特徴点の誤対応例を図 3.6 (点 A~D) に示す. 本研究では, 特徴点として SIFT 特徴量を利用し, 特徴点のマッチングには KD 木探索 [41] を基にした Best-Bin-First (BBF) アルゴリズム [71] による最近傍探索処理を利用する. マッチングは特徴量空間中での距離を用いて行われるため, 少なからず何らかの誤対応が発生する. そこで, ジャイロセンサから得られる姿勢の情報を利用して誤対応を除去する. まず, 回転行列 \mathbf{R}_{gyro} と並進ベクトル \mathbf{h}_{gyro} を求める. \mathbf{R}_{gyro} はジャイロセンサの計測値から算出される. 一方, \mathbf{h}_{gyro} は座標系を図 3.7(b) のようにとり, 光軸と進行方向が一致すると仮定して $(1, 0, 0)^T$ とする. 時刻 $t+1$ の特徴量を \mathbf{R}_{gyro} と \mathbf{h}_{gyro} より時刻 t の画像に射影する. 射影された点

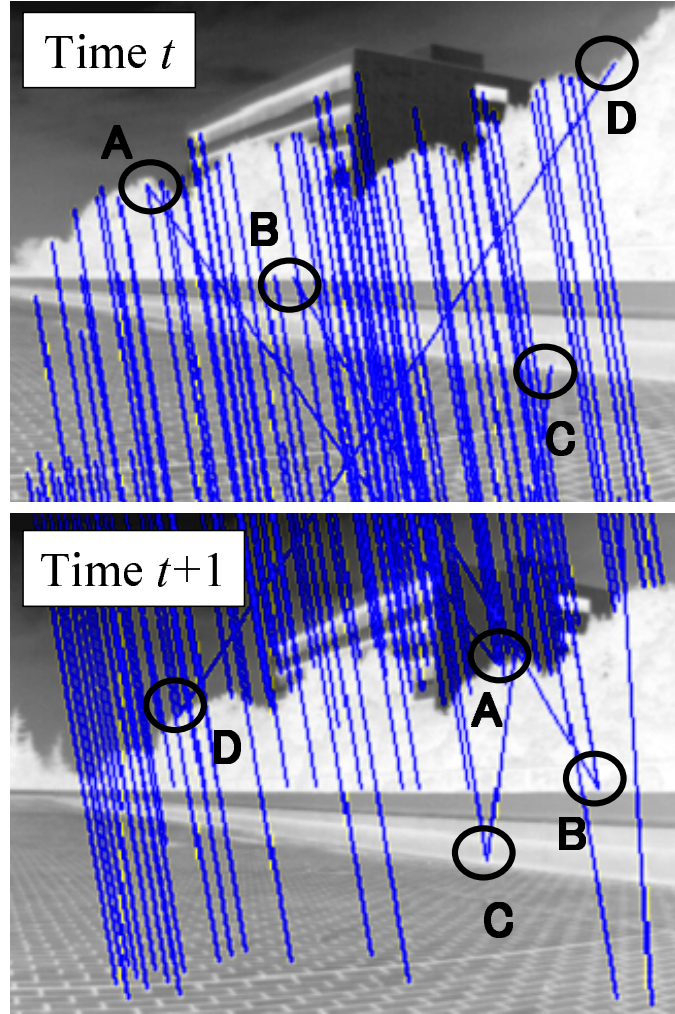


図 3.6: ミスマッチングの例

の N 画素四方内に時刻 t の画像の特徴点があれば対応点として利用する．それ以外は誤対応として対応点群から削除する． $N = (1/\text{psize})\{\max(\omega_i T) + 0.025\}f$ とした． f は焦点距離， psize は画素の大きさ， $\max(\omega_i T)$ は 3 軸の姿勢変化量の最大値 [rad] である．

ジャイロセンサの情報は特徴点の対応づけだけでなく，基礎行列 \mathbf{F} の導出にも利用できる．これにより，カメラ画像から求める姿勢の精度向上が可能である． \mathbf{F} の導出には，式 (3.19) の評価関数を元に逐次計算法である Levenberg-Marquardt (LM) 法 [66] を用いる． \mathbf{F} の初期値として，式 (3.18) とジャイロセンサから求めた \mathbf{R}_{gyro} と \mathbf{h}_{gyro} から算出された値を設定する．LM 法は勾配法を基にしているため，局所最適解に陥る可能性がある．しかし，ジャイロセンサの情報より \mathbf{R}_{gyro} と \mathbf{h}_{gyro} を用いて解の近傍から探索することで，安定した精度を実現できる．

3.4.5 提案手法の特徴

MEMS 型の振動型ジャイロセンサと単眼カメラを組み合わせた，提案する姿勢推定方法は，高精度かつ安定した姿勢推定を実現する．提案方法は，カメラから姿勢が求まらない場合も考慮している．具体的には，カメラから求める姿勢の信頼性判定と Kalman フィルタの状態量の設計とその更新方法を工夫した．信頼性判定では，事前にジャイロセンサを画像処理に利用し，以下の処理を適用した．

- (a) 特徴点の対応付けにおける誤対応の除去を行う．
- (b) ジャイロセンサの姿勢を初期値として解の近傍から計算する．
- (c) カメラから求めた姿勢を評価する．
- (d) Kalman フィルタの状態量の設計，更新方法については，軸ごとにオフセット誤差，バイアス誤差を状態量とする．
- (e) ジャイロセンサの時間変化に緩やかな誤差特性に着目し，カメラからの姿勢が求まらない場合は前時刻までの推定値を引き継ぐ．

(a) については，拡張現実の分野にて特徴点の探索にジャイロセンサを用いる手法がある [53, 61, 62]．(d) の状態量の設計は他のセンサ系において同様の設計手法がある [63, 64]．しかし，(a) から (c) の工夫点を統合し，Kalman フィルタを選択的に実施する機構と (e) の更新方法に着目している事例は過去にない．さらに (c) では RANSAC [65] を用いる点は一般的であるが，Kalman フィルタで用いる観測量の分散値に着目した信頼性判定方法は新しい．提案手法は，実環境のようなカメラから断片的にしか姿勢が求まらない場合でも精度向上を実現する．

3.5 実験と考察

3.5.1 実験条件・手順

車両 (MobileRobot 製 Pioneer3DX) の上に MEMS 型ジャイロセンサ (BOSCH 製 SMG074) とカメラを搭載し，1,500 mm/s 以下の移動速度で走行，計測した．実験車両にはこれらのセンサとは別に，評価用に光ファイバジャイロ (FOG) とサーボ加速度計 (日本航空電子工業製 JM3403/JA-5VC4) を搭載した．図 3.7 に実験環境を示す．FOG

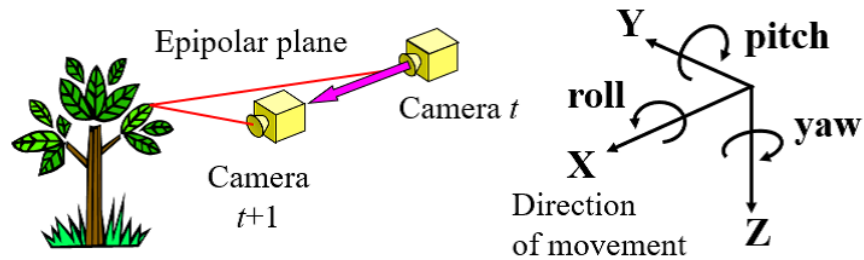
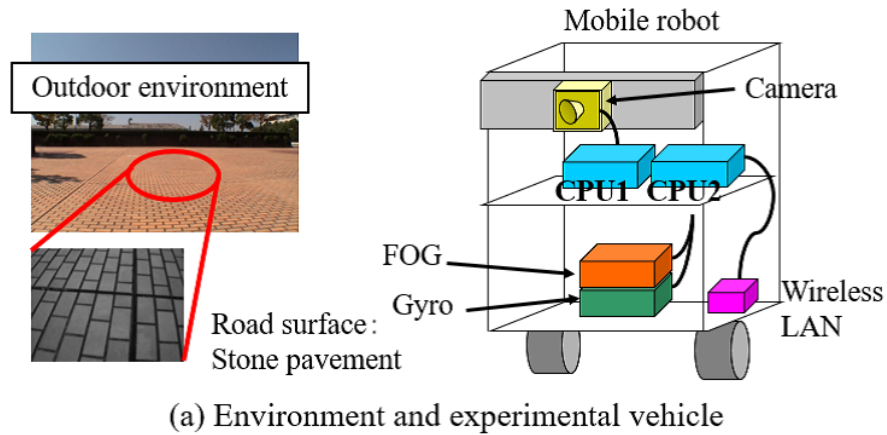
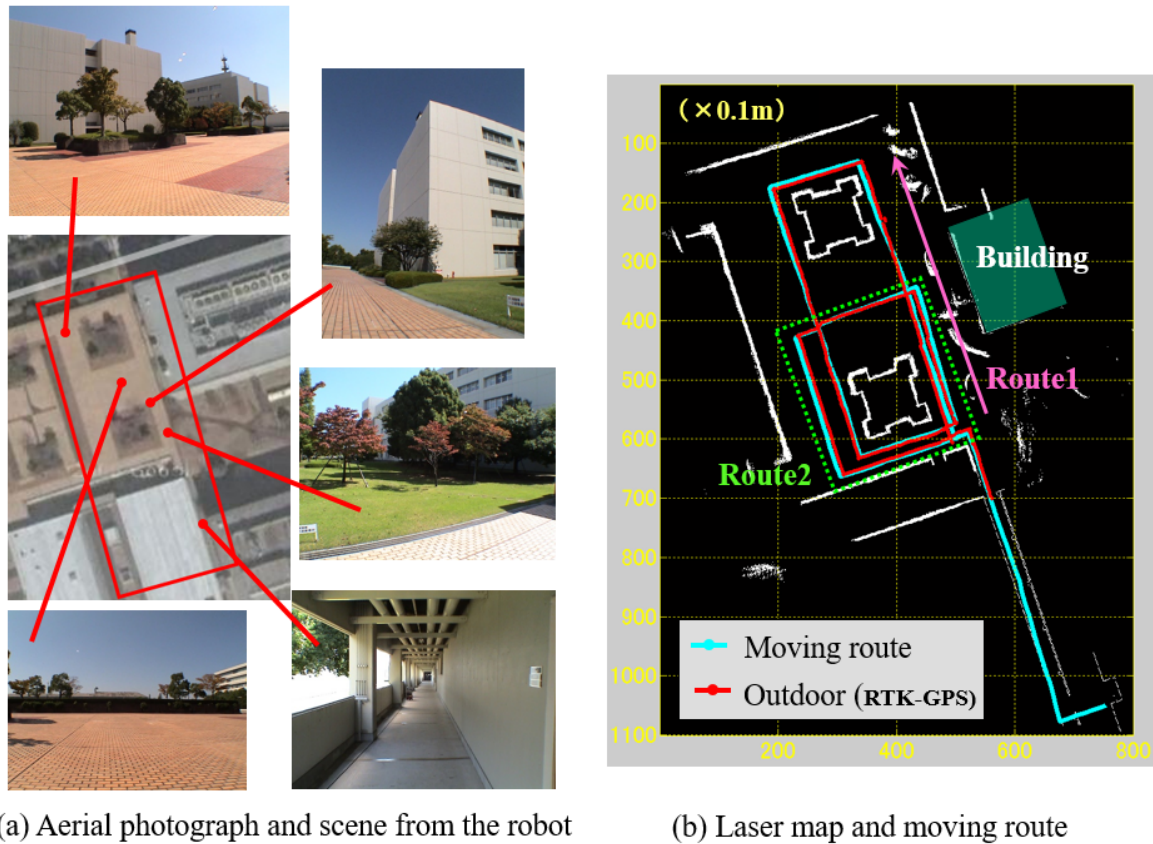


図 3.7: 実験環境

は高価であるが、その誤差は1時間で約3 deg.であり、非常に高精度である。よって本実験では姿勢の真値として利用することにした。計測周波数は、ジャイロセンサが100 Hz、カメラ画像が10 Hzである。また実験はデータを計測後、オフラインで行った。計算周波数はカメラ画像に合わせ10 Hzとし、ジャイロセンサの計算周波数は平均処理によりダウンサンプリングしてカメラの計測周波数に合わせた。走行環境は屋外であり、走行路面は石畳である。またカメラの姿勢が路面に対して水平となり、光軸が車両の進行方向と一致するようにし、カメラを他の用途にも利用できる配置にした。各軸のとり方は図 3.7 に従う。また、図 3.8 に走行経路と周辺環境を示す。屋内、屋外の双方の環境を延べ1.5 km 程度（約300 m × 5 セット）走行した中から、経路1の直線路（経路長：約60 m、走行時間：900 s）と経路2の旋回路（経路長：約100 m、走行時間：1,700 s）の2つのデータセットを用いて評価した。



(a) Aerial photograph and scene from the robot

(b) Laser map and moving route

図 3.8: 走行ルート

3.5.2 実験結果

1. 経路 1：直線路走行

図 3.8 の経路 1 に示す直線路を走行した結果を示す。表 3.2 に、得られた姿勢推定精度をロール、ピッチ、ヨー各軸に対する平均誤差で示す。比較手法 1：ジャイロセンサのみで計測，比較手法 2：信頼性判定でカメラからの姿勢が不十分な場合にジャイロセンサで代用して計測，そして提案手法の 3 つを比較した。提案手法はジャイロセンサ単体と比較して，いずれの軸も 900 フレーム間での時間平均にして 3~6 deg. の精度改善を果たした。またカメラからの姿勢が求まらない場合，ジャイロセンサに切り替えて姿勢計測するよりも提案手法の方が精度改善があった。

図 3.9 に，ロール・ピッチ・ヨー方向の姿勢をグラフで示す。ヨー軸では，ジャイロセンサは徐々に真値からずれていき，900 フレーム後には，8 deg. 程度の誤差

表 3.2: 経路 1 (直線路) における姿勢推定精度 (単位: deg.)

Method	Roll	Pitch	Yaw
Compared method 1	5.53	7.94	3.94
Compared method 2	3.43	3.85	1.55
Proposed method	2.73	1.99	0.84

表 3.3: 経路 2 (旋回路) における姿勢推定精度 (単位: deg.)

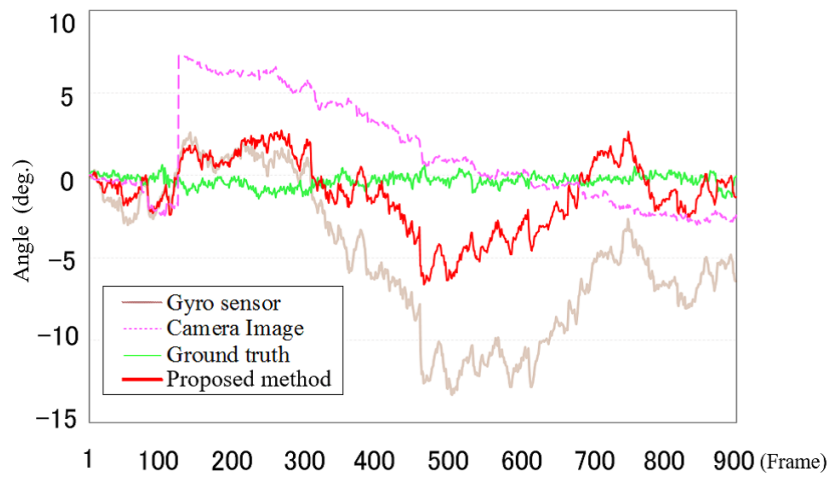
Method	Roll	Pitch	Yaw
Compared method 1	3.85	7.83	35.63
Compared method 2	3.10	3.55	6.72
Proposed method	1.56	4.27	3.82

となった. またカメラで姿勢を求めた場合も途中で誤差が発生した. これは遮蔽によって画像から特徴点が得られなかった, もしくは基礎行列を十分な精度で求めることができなかったためと考えられる. 実環境においては, カメラによる姿勢算出が困難である場合があることを確認できた. 一方, 提案手法は常時ほぼ真値に近い値を示した. またロールとピッチ軸でも, 提案手法の効果を確認できた.

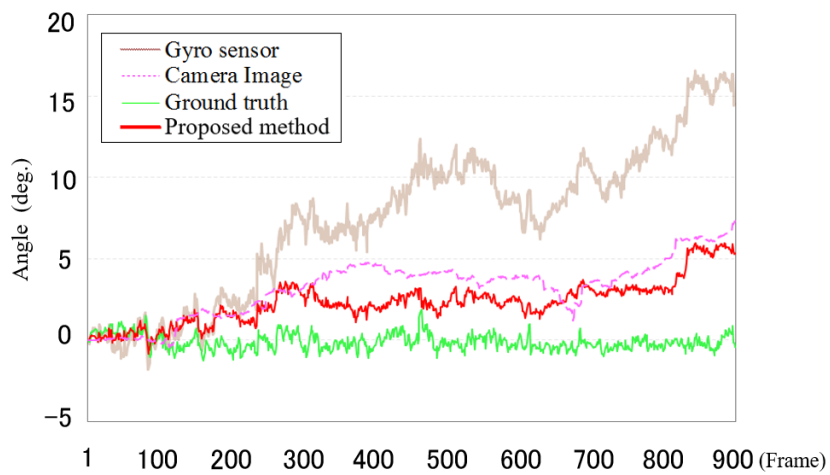
2. 経路 2 : 旋回路走行

図 3.8 の経路 2 に示す旋回路を走行した結果を示す. 表 3.3 に得られた姿勢推定精度をロール, ピッチ, ヨー各軸に対する平均誤差で示す. また, 図 3.10 に各軸方向の姿勢を示す. いずれの軸も提案手法の効果を確認できる. 旋回路においては, ジャイロセンサ単体で計測した場合, 曲がる際に大きな誤差がヨー軸方向に発生した. 一方, 提案手法では曲がる際も精度良く推定できていることが分かる. 各軸で比較すると, 直線路同様に, ヨー軸の推定精度が最も良い. 光軸と移動する方向との角度変化量が大きい軸が最も効果があると考えられる. 特に, ロールやピッチ軸はロボット走行時の石畳みの高周波の振動を拾うため, 10 Hz 程度のカメラ画像の計測周波数では, FOG 同等の精度でロボットの走行動作を再現するのが難しいと考える.

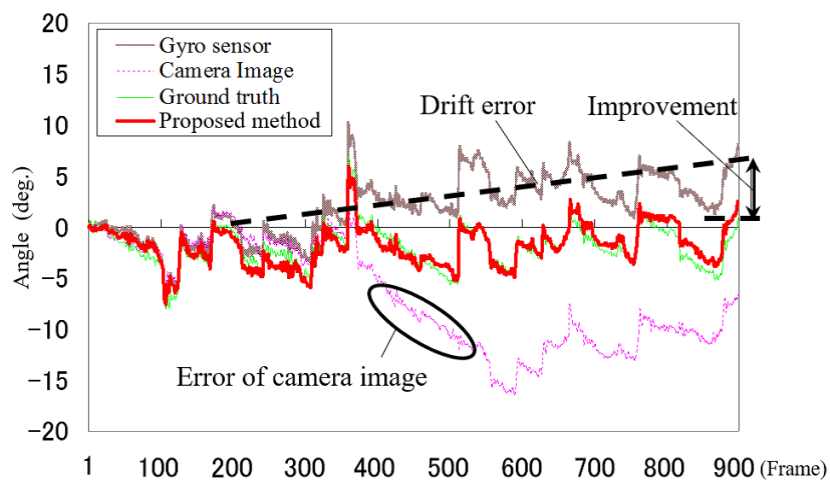
提案手法はカメラから姿勢が求まらない場合, 信頼性判定によりカメラからの情報を参



(a) Roll

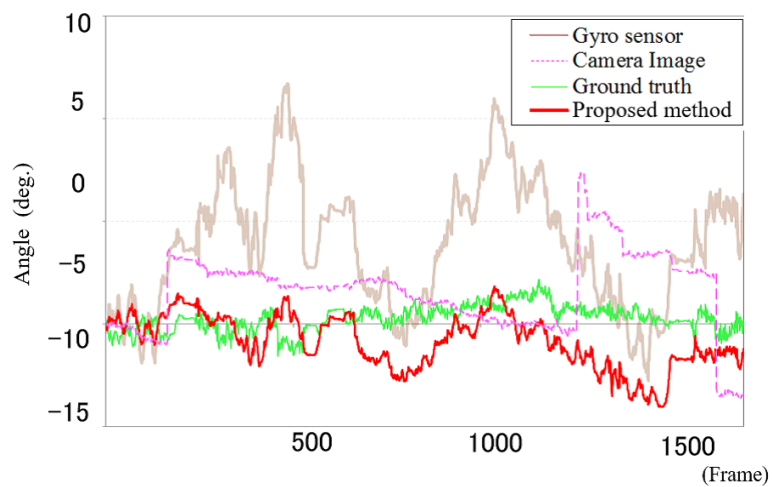


(b) Pitch

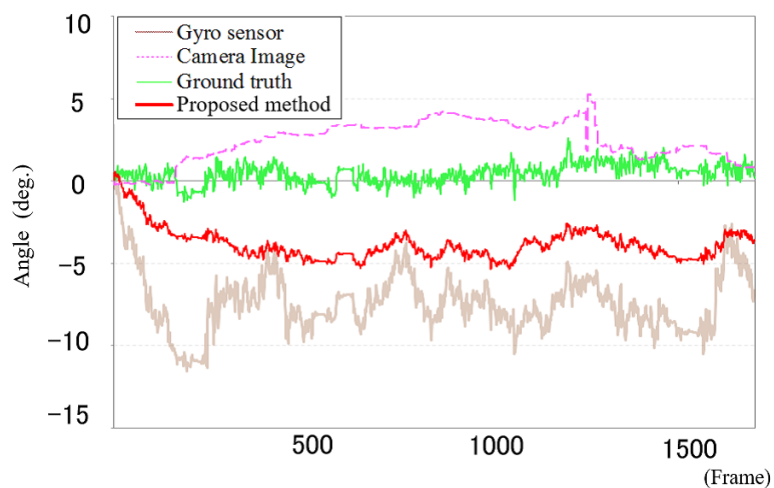


(c) Yaw

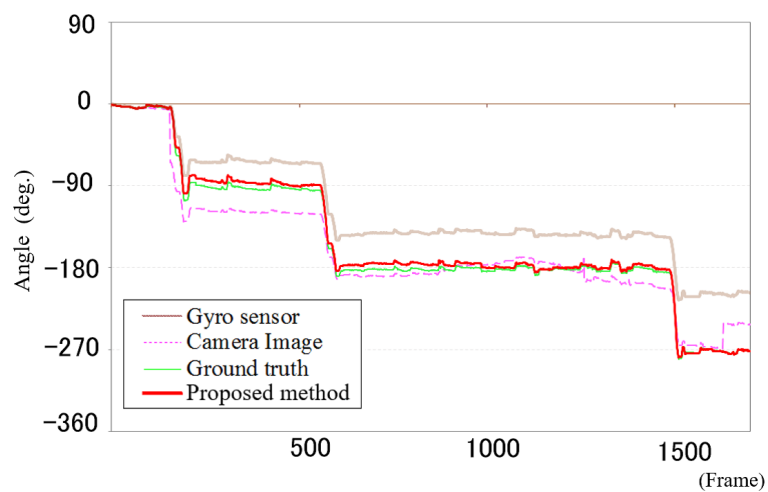
図 3.9: 経路 1 (直線路) における各軸の姿勢



(a) Roll



(b) Pitch



(c) Yaw

図 3.10: 経路 2 (旋回路) における各軸の姿勢

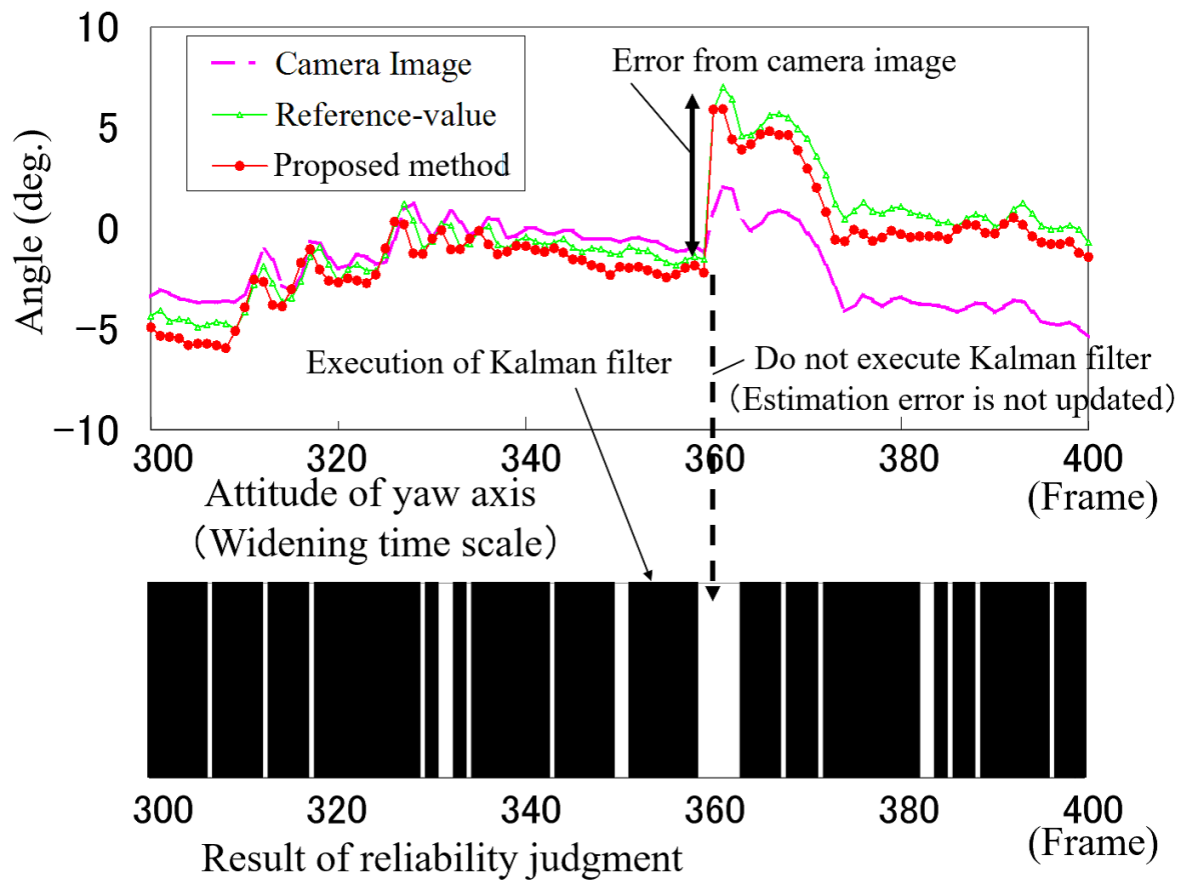


図 3.11: 経路 1 におけるカメラで計算された姿勢の信頼性判定結果

照せず Kalman フィルタを適用しない。そして、その場合には前時刻までに推定されたジャイロセンサの誤差量とジャイロセンサ値から姿勢を導出する。ジャイロセンサの誤差量として低周波の誤差に着目することにより、カメラから姿勢が求まらない場合にも、高い推定精度を保つことができる。このような方法により、実環境においても高精度かつ継続的な出力が可能なが実験によって確認できた。提案手法で意図した通りに、ジャイロセンサとカメラの相補的な利用を実現できたといえる。

3.5.3 カメラから姿勢が求まらない場合の検証

カメラから姿勢が求まらない場合について、経路 1 を例に、姿勢の信頼性判定の効果と Kalman フィルタの状態量について検証する。

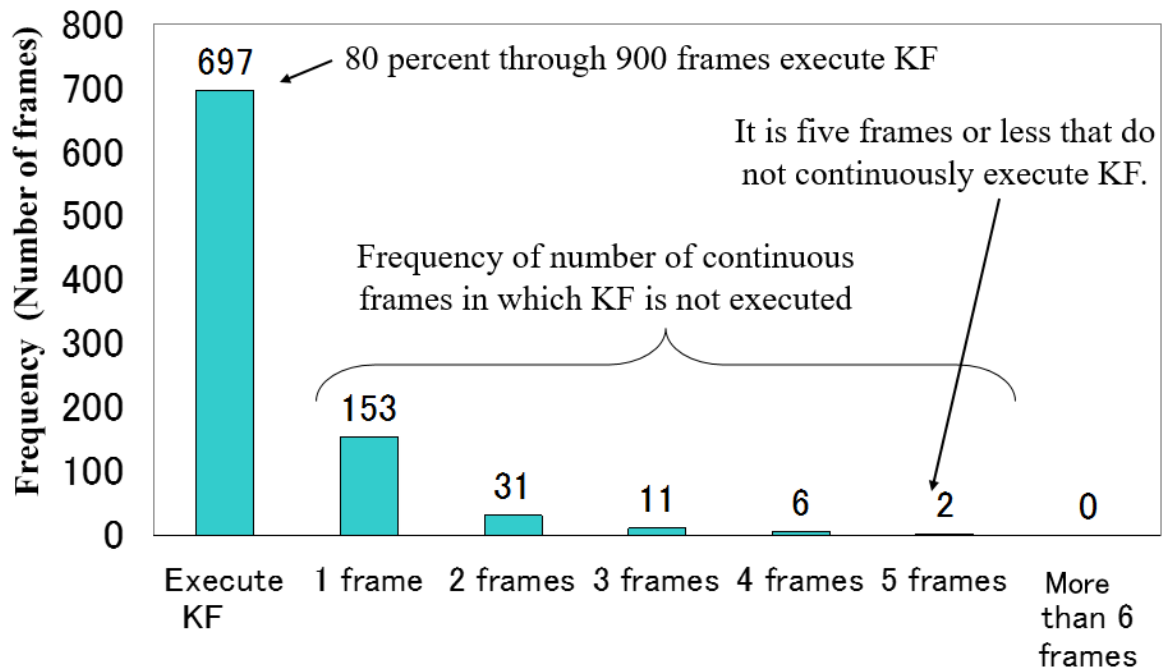


図 3.12: 経路 1 における推定誤差を引き継いだ連続フレーム数とその頻度の関係

3.5.3.1 カメラから求めた姿勢の信頼性判定結果

図 3.11 の上のグラフは図 3.9 のヨー軸の一部を時間軸に対して拡大表示したグラフである。図 3.11 の下のグラフは同じ時間軸に対する Kalman フィルタ適用の有無を示す。カメラから得られた姿勢の信頼性が低いと判定され、Kalman フィルタが適用されなかった場合を空白で示す。一方、カメラから姿勢が求まり Kalman フィルタが適用された場合を塗りつぶして示す。真値と比較すると、カメラの姿勢の誤差が大きい場合、Kalman フィルタが適用されなかったことが分かる。このことから少なくとも経路 1 では、信頼性判定が正しく行われたことを確認した。また、図 3.11 の空白の箇所では、カメラから取得される特徴点の数が少ない場合や通行人によって動的に遮蔽された場合であった。提案手法は、このような走行環境や走行時の状況に左右されにくい構成であることが示唆された。

3.5.3.2 Kalman フィルタの状態量に関する検証

カメラから姿勢が求まらない場合は、ジャイロセンサの推定誤差を時間方向に引き継ぐ。このため、精度を保ったままどの程度連続して推定誤差の引き継ぎが可能であるかを

表 3.4: 経路 1 における誤対応の発生数 (合計 197,505 点)

Image processing + Gyro	126
Image processing only	2,740

確認する必要がある。これに関して、経路 1 において推定誤差を引き継いだ連続フレーム数とその頻度の関係を図 3.12 に示す。900 フレーム中、Kalman フィルタを適用したフレーム数は 697 であった。残り 203 フレームは Kalman フィルタを適用せず、推定誤差の引き継ぎを行った。そのうち、153 フレームは 1 フレーム間のみ、50 フレームは 2 フレーム以上連続して引き継ぎを実施した。本実験では、最大 5 フレームまでジャイロセンサの推定誤差を引き継いでおり、短時間であれば精度を維持できることを確認した。実環境では断片的にしかカメラから姿勢が求まらない場合があるため、ジャイロセンサの低周波成分の誤差に着目した Kalman フィルタの設計は有効であることが示唆された。

3.5.4 ジャイロセンサを利用した画像処理の検証

ジャイロセンサを利用した特徴点のマッチング及び、ジャイロセンサを用いた基礎行列算出について、経路 1 を例に、その効果を検証する。

3.5.4.1 ジャイロセンサを用いたマッチング結果

経路 1 900 フレームにおける誤対応発生数を表 3.4 にまとめる。ここでは、FOG とサーボ加速度計で得られた回転行列 \mathbf{R}_{ref} と並進ベクトル \mathbf{h}_{ref} を元に時刻 $t+1$ の特徴点を時刻 t の画像に射影、射影された点の 3 画素四方内に時刻 t の画像の特徴点がない場合を誤対応とした。表中の画像処理は BBF アルゴリズムを、画像処理 + ジャイロセンサは BBF アルゴリズムにジャイロセンサによるフィルタリング処理を加えた場合を意味する。ジャイロセンサによりフィルタリング処理をすることで誤対応の発生数を減少させることができた。1 フレームあたり平均 3.0 から 0.14 まで誤対応の発生数が下がり、精度改善に寄与できた。しかし、 N 画素四方の窓幅内に誤対応が発生することがあったため、依然 126 個分の誤対応が残った。

表 3.5: 経路 1 におけるカメラから求めた姿勢変化量とジャイロセンサの精度 (単位: deg./s)

	Accuracy	Resolution
Image processing + Gyro	± 0.0012	± 0.21
Image processing only	± 0.0197	± 1.3
Gyro sensor	± 1.0 bias 0.25 offset	± 0.00571
Fiber Optic Gyro (FOG)	± 0.00083	± 0.019

3.5.4.2 ジャイロセンサを用いた基礎行列算出結果

経路 1 において, カメラから求めた姿勢推定精度に関する検証結果を表 3.5 に示す. カメラから求まる 1 フレーム間の姿勢の変化量 (ピッチ軸) を元に算出したものである. 表の画像処理は LM 法を, 画像処理+ジャイロセンサは LM 法の初期値にジャイロセンサから求めた値を与える方法を指す. 精度は経路 1 の 900 フレーム間の平均誤差を, 分解能は精度を満たす姿勢変位の最小値である. また, 使用したジャイロセンサと FOG の性能も併せて表 3.5 に示す. LM 法の計算回数は 5,000 回で打ち切った. 基礎行列の計算にジャイロセンサを利用することで, 画像処理だけで行うよりも精度, 分解能が向上することを確認した.

LM 法は勾配法の枠組みによる逐次計算であり, 与える初期値によって局所最適解に陥る問題がある [72]. 提案手法はジャイロセンサの情報より解の近傍から探索を開始できるため, 安定した収束を実現できる. また, 基礎行列の計算には最小 2 乗法を固有値計算で解く方法 [66, 73] があるが, 本実験では計算誤差が発生し, 満足に求まらなかった. これは, 図 3.7 に示すように, カメラの光軸と進行方向がほぼ一致する状態となるため, エピポラ平面を形成するための 3 角形が細長くなりやすいためである.

3.5.5 制御への適用に関する考察

本章では, 実用面での課題に焦点を当てた手法を提案した. 制御への適用を考えると, 姿勢そのものよりも, 姿勢変位 (角速度) が重要となる場合が多い. 提案手法は角速度の出力も可能である. カメラを用いることで精度改善ができるジャイロセンサとして考えれ

表 3.6: 経路 1 における画像処理の計算時間 (単位 : ms)

	SIFT	KDTree	BBF	LSM	KF
Computation time	645.9	9.0	96.7	36.7	1.8

ば, コスト的にも有用であり制御への利用も期待できる. また制御への適用を検討する際, 実時間処理が必要になる. 表 3.6 に経路 1 における各処理の計算時間を Intel Pentium M 1.2 GHz 環境下で計測した結果を示す. 表中の SIFT は SIFT 特徴量の導出, KDTree は KD 木の作成, BBF は最近傍探索処理, LSM は基礎行列計算, KF は Kalman フィルタの計算時間である. 現状の構成では実時間処理が難しいが, この問題は CPU の分散処理, GPU (Graphics Processing Unit) の利用, FPGA (Field Programmable Gate Array) によるハードウェア化などで解決可能と考える. このような構成でも光ファイバジャイロなどに比べ, 安価に設計ができる.

3.6 まとめ

本研究では, ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の“姿勢推定”の高精度化を画像処理技術により解決する. 本章では, 局所特徴量を活用することで, ロボットの姿勢推定を高精度化する方法を考案した. ジャイロセンサとカメラ, 双方の利点で欠点を補完することにより, 高精度で頑健にロボットの姿勢を推定する手法を提案し, その効果を確認した.

既存研究の課題であった, カメラから姿勢が求まらない場合 (遮蔽の場合も含む) の方策に対しては, カメラから得られる姿勢の信頼性が高いときのみジャイロセンサとの Kalman フィルタを適用することにより姿勢精度の向上を実現した. 提案手法は, 特徴点が得られずカメラから姿勢が求まらない場合でも, ジャイロセンサの誤差の低周波成分を推定することにより短時間であれば精度維持が可能である. 信頼性判定では, カメラから得られる姿勢のばらつき (観測量の分散値) を元に判定する. 姿勢精度の期待値に合わせた判定が可能である. カメラからの姿勢算出の際は, 特徴点の選択, 基礎行列計算にジャイロセンサを利用することで誤対応の低減と精度の向上を実現した.

第4章

局所特徴量を活用した“検出しやすい”姿勢推定用マーカ

本研究は、ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の姿勢推定の高精度化を画像処理技術により解決する。これに対して、本章では、画像処理技術として局所特徴量を活用して、把持対象物体の姿勢推定を高精度化する方法について提案する。マーカを用いた把持対象物体の姿勢推定では、マーカの様々な配置・距離に対する検出と認識の難しさがある。そこで、マーカの検出方法に局所特徴量のキーポイント検出の考え方を導入した“検出しやすいマーカ”を開発し、把持対象物体の姿勢推定の高精度化（マーカの検出率・認識率の向上）を図る。

4.1 提案するマーカの特徴

作業対象である複数の物体が任意の距離や姿勢で配置されたとしても、それらを検出・認識し、かつその姿勢を的確に推定できることが重要である。このような問題設定に対して、作業対象に直接マーカを貼付することで物体の検出・認識と姿勢推定を行う手法が実用化されている。この際、作業対象が小さな物体や曲率を持った物体の場合は、取手が付いた箱にマーカを貼付し、箱の中に作業対象物体を入れる方法が採られる [3, 4]。貼付したマーカと物体の把持操作箇所の位置関係を事前に記録しておくことで、ロボットは把持操作を行うことができる。現在用いられているマーカの多くは、被写体であるマーカとカメラの距離・位置関係をある程度限定することで実用に耐え得る性能を得ている。そのため、マーカそのものを検出しやすい条件下でないと正しく動作しない。このようなマーカ

を用いる際には、(1) ボケが発生する、(2) 雑然とした背景環境下で様々な距離や姿勢でマーカが配置される、の2つが大きな問題となる。ロボットが様々な作業を行うためには、(2) は解決すべき重要な問題であり、マーカの距離や配置に対する高い自由度が求められる。また(2) を解決するためには、(1) のボケがある環境を扱うことも重要である。これは、マーカが被写界深度 (depth of field) の範囲外に配置される場合にボケが発生するためである。

上記のような背景から、本章では把持作業を想定した作業ロボットのマーカに対して、次の2つの実用課題の解決を目的とした新しいマーカを提案する。

- (i) ボケに依存せず検出しやすい
- (ii) 背景と配置に依存せず検出しやすい

これらを実現するために、容易にマーカを見つけるための“検出”処理と、それに必要なマーカのデザインパターンを提案する。従来、“検出”に焦点を当てた手法は提案されておらず、また上記はマーカをロボットに用いる際の特有の課題である。なお本章では、マーカを見つけることを“検出”、検出した候補の中から正確にマーカを同定することを“認識”として区別する。

提案手法によるマーカは、空間中の姿勢、位置、大きさ (スケール) のいずれも事前に分かっていないような条件下でも容易に見つけることが可能であり、かつ姿勢やスケール変化に対しても頑健である。このマーカは、従来のマーカのように姿勢や大きさが異なる複数のテンプレートとの照合も不要である。具体的には、SIFT のキーポイント検出 [27] で有名なスケールスペース理論 [74] を活かした符号化マーカを提案する。スケールスペースの極値を算出し易いマーカ形状 [75] とマーカのみを選別する判別フィルタを用いることで効率的にマーカの検出を行う。次に、マーカの同定に、M 系列 [76] を用いることで、マーカのパターンがボケた場合でも認識できるようにする。

4.2 位置・姿勢推定マーカに関する先行研究

最初に、マーカを用いるアプローチを採る理由について説明し、その後マーカの先行研究を紹介する。

マーカを利用せず、物体のモデルを事前に登録または学習したうえで、その検出・認識

と 3 次元の姿勢推定を行う手法が研究されている [18–20]. これらは、物体の 3 次元 CAD モデルを事前に準備し、その様々な姿勢を 2 次元 RGB-D 画像（または RGB 画像）テンプレートとして記録し、テンプレートとの照合により検出・認識と姿勢推定を行うものである。各物体に対して様々な姿勢のテンプレートと照合するため計算コストが高く、また特徴的な形状でないと高い検出精度を得られないといった実用面での課題が多い。また既知の物体でなく、一般物体の 3 次元姿勢推定を行うものもある [26]. しかし、精度の点では、多様な物体を扱うほどの汎用性を備えた実用水準にはない。物体の置かれ方、遮蔽、照明条件、検出（領域の切り出し）、計算コスト、物体の種類に対する拡張性など、まだ多くの課題がある。

一方、マーカは実用性には長けている。そこで、従来のマーカ技術を整理する。現在最も広く用いられているマーカである QR コード [77] は、エッジやコーナといったパターンで構成されたマーカである。目印である隅 3 箇所のコーナパターンを検出するために、カメラとマーカの距離が近く、比較的正面から撮影する必要がある。また、遠方からの撮影においては、パターンがボケて観測されるため、符号化開口 [78] や超解像 [79] を利用して検出性能を改善する試みもなされている。しかし、マーカそのものを検出しやすい条件での利用を想定して設計されており、作業ロボットのように遠方のマーカを検出したいような応用には不向きである。

ARToolKit [16] で用いられるマーカは、環境中のカメラ位置・姿勢の推定に広く利用されている。マーカは黒の太枠と内部の 2 値パターンで表現される。黒の太枠の隅 4 箇所の点を求め、さらに 4 辺の長さを比較することで検出を行う。カメラとマーカの間の相対的な位置・姿勢も、黒の太枠の隅 4 箇所から推定する。このマーカは、任意の配置に対して位置・姿勢を推定するのに有効である。しかし、菱形など類似形状に対して、誤検出が多く発生するという問題がある。

一方、色情報を用いたマーカも多数提案されている。ChameleonCode [80], ColorZip [81], ColorOCM [82], ColorBit [83], C-Band [84], HCCB [85] などがその代表例である。これらのマーカは色の付いたパターンの集合として表現され、パターンの色や隣接する色の変化に基づいて特徴（情報）表現を行っている。図書館における本の管理などで実用化されており、本の背表紙に貼り付けたマーカを一定の距離から計測することにより、複数のマーカの同時検出と認識を可能としている。これらのマーカの検出は、QR コードと同様に特徴的なパターンを捕捉する方法 [80, 81, 84, 85] と色の配列でマーカそのものを特

定する方法 [82, 83] に分類される．前者は QR コード同様の課題が存在する．後者は，背景に同様の色が存在する場合やマーカが遠方で小さく撮影され，色情報がボケて混合，欠損する場合に検出が困難である．マーカの歪みや配置に対しても高い自由度がある反面，色を正確に取得できない状況では検出することができない．

また，遠方から検出できるマーカとして，Nested Marker [86], Bokode [87] がある．Nested Marker はフラクタルパターンを使用し，カメラとマーカの間が遠距離であっても近距離の幾何学的なパターンの見え方が同じになるように設計されている．Bokode はマーカが光学素材から構成され，投光器を必要とする．投光した光を受光したマーカは，ピンホールカメラの原理によってそのマーカよりも大きなパターンを空中に投影する．両者とも距離と認識（データの復号性能）の問題に焦点を当てており，カメラとマーカとの幾何関係は正面のみを対象としている．しかし，マーカの配置が正面でない場合においては，QR コード同様の課題が存在する．一方，ボケに対して頑健なマーカとして，Mono-Spectrum Marker [88] がある．これは，パターンの要素が単一かつ低周波な色で表現される．認識方法のみ議論されており，検出方法は ARToolKit マーカと同じであり，同様の課題がある．

以上で述べたように，従来のマーカでは，検出しやすい条件下で使用されることを想定しており，(i) 雑然とした背景環境下でマーカが様々な距離や姿勢で配置される，(ii) ボケが発生する，といった条件下での検出については十分考慮されていない．そこで，本章では，マーカを見つけ出す“検出”の観点からマーカ検出に改良を行う．

4.3 スケールスペース理論に基づく検出しやすい符号化パターン

画像から自動的に情報を得ようとするとき，見るべきスケールを事前に知ることはできない．この問題に対処するため，画像をスケールによってパラメータづけられた族として表現し，すべてのスケールを同時に扱うスケールスペース理論 [74] が提案されている．提案するマーカの特徴は，スケールスペースの極値検出を活かしたマーカ検出であり，マーカの大きさ（スケール）の推定やマーカの配置に依存せず容易にマーカの候補を見つけることが可能である．

4.3.1 検出処理

様々に配置されたマーカを正確に見つけることは難しい．一方 SIFT 特徴量もスケールスペースの極値検出（キーポイント検出）を利用し，拡大・縮小に対する不変性を特長とする．この性質を活かしてマーカを設計すれば，マーカまでの距離に依存せず容易に検出することが可能になると期待できる．また SIFT 特徴量はキーポイント周辺の勾配情報の記述によって回転と並進に対する不変性を有しているが，これはキーポイントを前提にしており，キーポイントそのものも回転と並進に対して不変性を有していると考えられる．つまり，スケールスペースの極値検出を応用すれば，マーカの向きにも依存せず容易に検出することが可能と考える．図 4.1(a-1) にマーカ画像の検出パターン，(b-1) に (a-1) を輝度反転した画像を示す．さらに，図 4.1(a-2) と (b-2) は，(a-1) と (b-1) の輝度を各々 3 次元表現したものである．スケールスペースの極値検出は，図 4.1(a-2) に示す画像中の輝度の張りに囲まれた輝度の窪み（極小値）と図 4.1(b-2) に示す画像中の輝度の窪みに囲まれた輝度の張り（極大値）に強く反応する性質がある．そこで，この性質を利用したマーカを設計する．具体的には，中心部が円形のリング状のマーカで，リング部が輝度の窪み，中心部の円形が輝度の張りに相当した検出パターンを考案する．このマーカは，スケールスペースの極値が算出され易い模様であり，算出されるスケール値に基づいてマーカの大きさも推定できる．

4.3.2 認識処理

遠方からの撮影ではマーカのパターンが劣化する．そこで，“自己相関が高く，相互相関が低い”ことで良く知られる M 系列 [76] を用いた符号化をする．これにより，計測した符号パターンと設計した符号パターンの間で急峻な相関を示すように設計することが可能であり，マーカ中の符号パターンを正確に認識（同定）できる．

4.4 提案手法

4.4.1 マーカ検出：キーポイント検出によるマーカの絞り込み

マーカの検出手順を図 4.2 に示す．まず SIFT のキーポイント検出を用いて，マーカの候補点探索を行う．次に，得られた候補点に対して，判別フィルタを用いてマーカの候

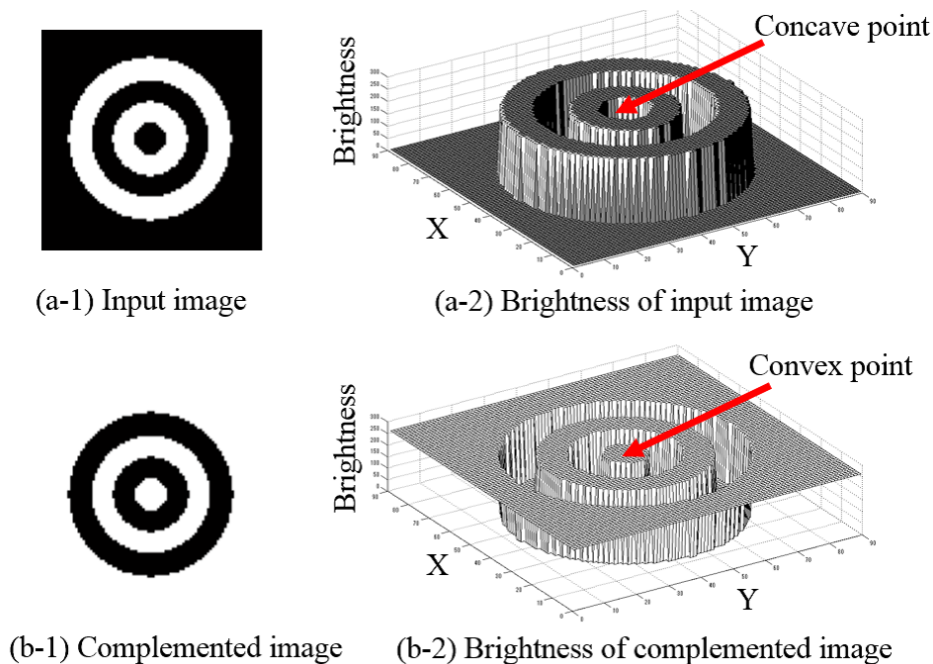


図 4.1: スケールスペースの極値検出に反応する輝度パターン

補を絞り込む。最後にマーカの判定処理を行う。

4.4.1.1 マーカの構造

図 4.3(i) にマーカの例を示す。SIFT のキーポイント検出でマーカの中心位置をとることができるようにするためにリング構造とする。さらに、RGB 色表現の G 成分を用いて検出パターンである 2 重リング構造を作成し、R 成分と B 成分でデータパターンを表現する。G 成分による 2 重リング構造は、マーカの位置検出のほか、4.4.1.4 で述べるパターンの相関値計算においても使用する。

4.4.1.2 キーポイント検出によるマーカ候補の絞り込み

入力画像から G 成分のみを抽出し、さらに輝度値の反転処理を行う（リング部が窪み、マーカ中心部が張りに相当）。そして、輝度の張り（極大値）のみをスケール空間（DoG 画像群 [27]）から探す。そして、キーポイントの位置とスケールを算出する。算出方法は第 2 章を参照されたい。当初、極大値と極小値の双方を候補として設計していたがどちらか 1 つで問題ない、極大値のみ候補としたため反転処理を実施する。図 4.4 に SIFT のキーポイント検出によるマーカ候補点を示す。誤った候補は検出されるにしても、いずれ

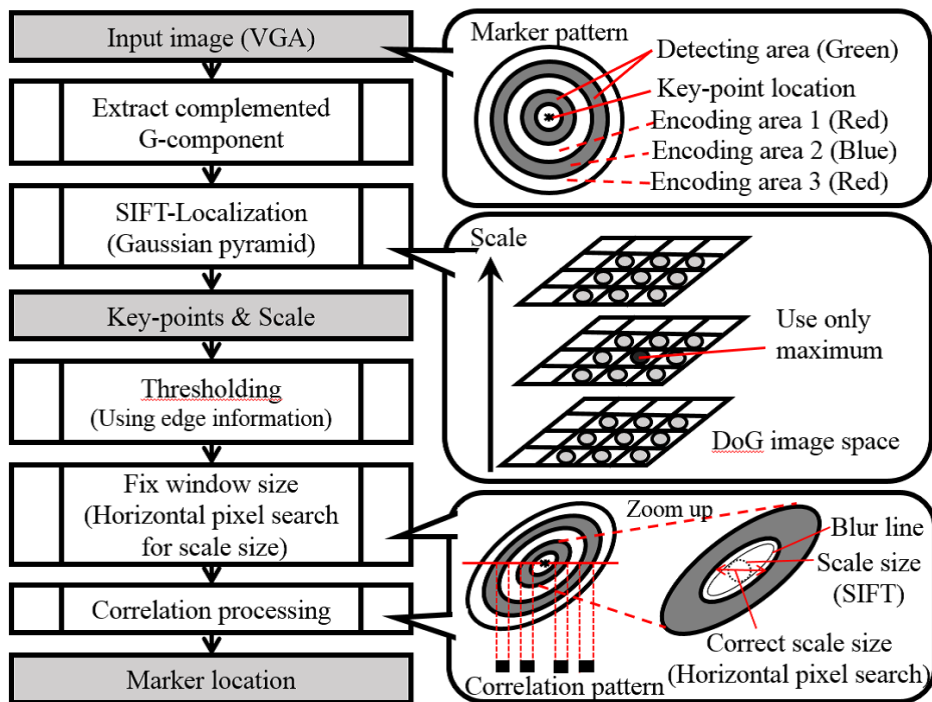


図 4.2: マーカの検出手順

も真のマーカ中心には必ずキーポイントが出現しているのが分かる。

4.4.1.3 判別フィルタによるマーカ候補の絞り込み

キーポイントの周辺の勾配情報より XY 方向のエッジの傾向度合に応じて、マーカの候補点をさらに絞り込む。XY 方向共にエッジ成分が強い箇所を残し、他の候補は除く。条件は、 $|\text{Det}(\mathbf{H})| > 1.0 \times 10^{-5}$ とし、判別式は式 (4.1) により求める。ここで、行列 \mathbf{H} は、式 (4.2) に示すような DoG 画像の Hessian 行列である。

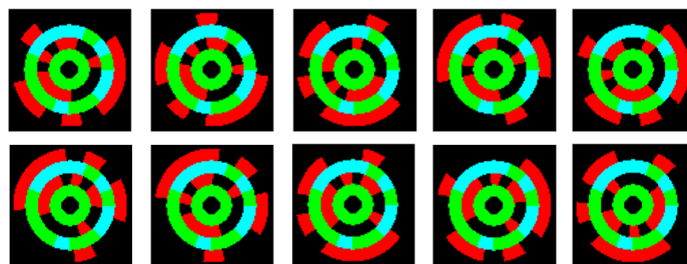
$$\text{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 \quad (4.1)$$

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (4.2)$$

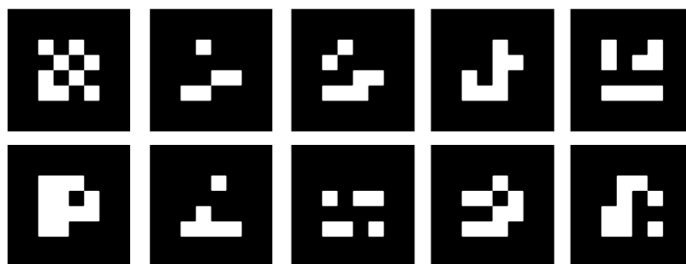
判別フィルタによる絞り込みの効果を図 4.5 に示す。凸凹の小さな起伏がある壁面に多くのキーポイントが検出されるが、判別フィルタを適用した後は、マーカのみ検出されたことが分かる。

4.4.1.4 パターン相関によるマーカの判定

キーポイントの位置とそのスケールも使用して、マーカ候補から真のマーカを判定する。



(i) Proposed markers (10 categories)



(ii) ARToolKit markers (10 categories)

図 4.3: マーカのデザインパターンの比較

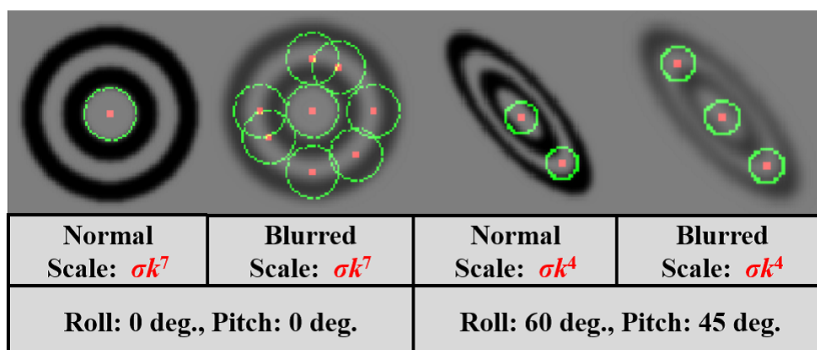


図 4.4: キーポイントの位置とその精度：赤い点はキーポイントを、緑の円はスケールサイズを示す． σ はスケールの初期値であり， k はスケールのステップの刻み幅である．

まず以下の手順によりスケールの再計算を行う．これは、マーカの配置がカメラに対して傾いている場合に、推定したスケールと実際のマーカの大きさの間で誤差が生じるためである．4.4.1.2 で推定されたスケールを η とした場合、キーポイントの位置 (x, y) を中心に、縦横 1.5η 四方の領域で輝度平均値 Φ を算出する（ここで、係数の 1.5 は経験的に決定した設計値）．そして、キーポイントの位置から水平方向に連続する、 Φ より輝度値が大きいピクセルの数を左右各々、 P_r , P_l とする．そして、 $(P_r + P_l)/2$ を最終的なス

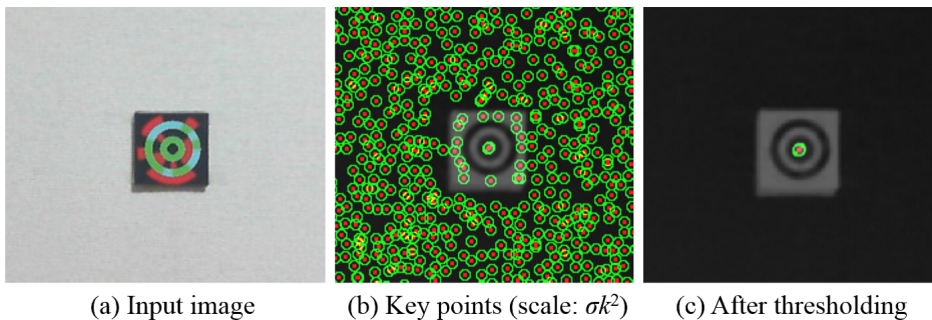


図 4.5: 判別フィルタによるキーポイントの絞り込み

ケール値 $\hat{\eta}$ とする．ここで，位置 (x, y) の輝度を $I(x, y)$ で表現する．

$$\hat{\eta} = (P_r + P_l)/2 \quad (4.3)$$

$$P_r = \sum_{i=1}^{0.75\eta} f(x+i, y) \quad P_l = \sum_{i=1}^{0.75\eta} f(x-i, y)$$

$$f(x, y) = \begin{cases} 1 & (I(x, y) > \Phi) \\ 0 & (\text{otherwise}) \end{cases}$$

マーカであることの判定は，2重リングで構成される G 成分の 2 値化パターンがマーカ設計比と同じか否かで判断する．2 値化パターンは，推定されたキーポイント位置を中心に水平成分 1 ラインのみ， $\hat{\eta}$ をもとに求めたマーカの大きさ分の画素で比較する．この際に相関値が $\text{Corr} \geq 0.8$ を満たすものをマーカと判定する．相関式は以下に従った．

$$\text{Corr} = \sum_{p=1}^n S_p O_p \quad (4.4)$$

ここで， S_p は計測パターン， O_p はマーカの設計パターンであり，いずれも 2 値化されて ± 1 のいずれかで表現されているものとする．また n [pixel] は推定したスケール $\hat{\eta}$ に基づいて算出したマーカの大きさである．リング形状のため，射影変換があった形状でもリング中心部を通る水平ラインのパターン比は不変であり，キーポイントとスケールさえ正しければパターン相関は一致する．

4.4.2 マーカ認識：M 系列を用いたデータパターン

マーカの認識手順を図 4.6 に示す．登録されたマーカのうち，どのマーカであるかを認識する．ターンテーブルの角度検出に M 系列符号を用いた先行事例 [89] があり，それ

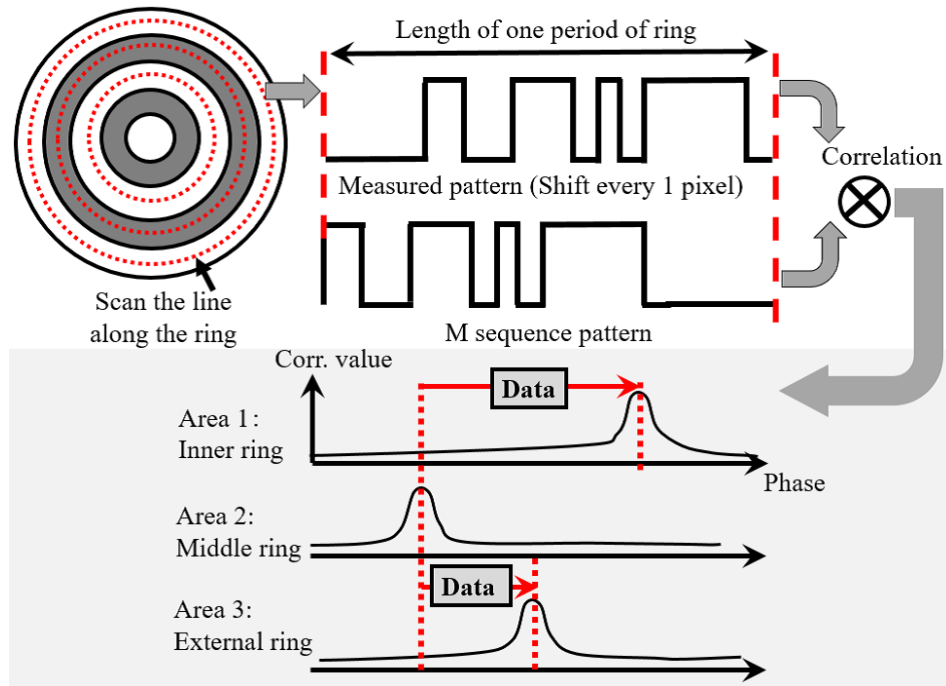


図 4.6: マーカの認識手順

をマーカに応用した。

4.4.2.1 データエリアのスキanning

データは3つのリング状の部分（図 4.2 の Encoding area を参照）に符号化されている。この部分を抽出するために、4.4.1.4 で推定したスケール値 $\hat{\eta}$ を用いて楕円形状フィッティングを行う。

4.4.2.2 M 系列を用いたデータパターン生成

リング1周の画素列が M 系列1周期に対応する。M 系列は4次の原始多項式 $a^4 + a + 1$ を使用し、 $2^4 - 1$ 個の $\{1, -1\}$ の要素を持った列ベクトルである。ここでは、16 個目の要素に -1 を追加した p を符号化・復号に用いる。なお、ここで1は輝度値の255に、 -1 は輝度値の0に対応する。

4.4.2.3 M 系列の位相差によるデータ表現・復号処理

図 4.6 に示す3つのリング（内側、中央、外側）の各々に、M 系列である p の位相シフトしたパターンを配置し、それらの位相差でデータを表す。リング1周は7 bit（128 分

表 4.1: ボケ環境下での精度 (F 値)

Marker	Detection	Decoding
ARToolKit	0.925	0.780
Proposed	1.000	0.978

解能 : 1 量子化単位あたり約 2.8 度) とする. 復号は, リングごとに, 1 周分のパターンと同じ長さに合わせた p との間で, 1 ピクセルずつシフトして畳み込み演算を行い, 最も相関値が高い地点のシフト量を記録する. 3 つのリング間でピーク地点のシフト量を比較し位相差として算出する. 中央リングと内側リング間の位相差で上位 7 bit, 中央リングと外側リング間の位相差で下位 7 bit, 計 14 bit のデータ表現が可能である.

4.5 実験と考察

4.5.1 実験条件・手順

提案するマーカの有効性を確認するために, マーカの検出及び認識の性能を ARToolKit マーカと比較して評価した. 具体的には, ボケと雑然とした背景環境下での多様な配置 (距離・姿勢) に対する頑健性を比較した. さらに, マーカの幾何学的配置に対する提案マーカの頑健性を評価するために, 距離と向きの 2 点の影響について分析した. 本実験では, 提案マーカと ARToolKit マーカ共に, 各々 14 bit 分 (16,384 個) のマーカを登録した上で, 実験には図 4.3 に示した 10 種類のマーカを用いた. マーカの大きさはいずれも 9 cm 四方である. これは, 生活支援ロボット [3, 4] で棚からの物取りタスクに使用しているマーカサイズに合わせた. 撮影にはスマートフォン (LG エレクトロニクス社製, Nexus5) 内蔵のカメラを使用し, 640×480 画素の解像度で撮影した. ただし, ボケに対する性能評価では, 焦点面の位置調整が可能な PointGrey 製のカメラ Cameleon を使用した. 同様に解像度は, 640×480 画素で撮影した. また提案するマーカ, ARToolKit マーカ共に検出・認識速度は Intel 社製 Core i5 を搭載した PC で 1 fps で可能であり, マーカを認識してから把持操作を開始することが十分可能である.

本実験では, マーカの検出位置が実際の位置に対して 5 mm 以内の場合に検出成功とし, マーカに埋め込まれたデータを正しく復号できた場合に認識成功とした. 検出実験,

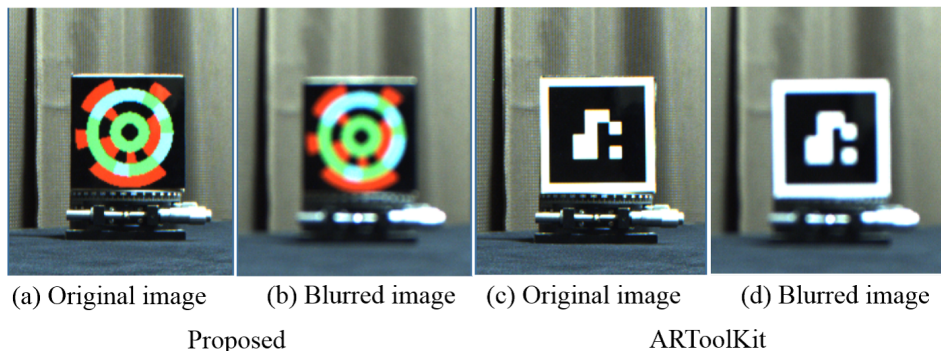


図 4.7: Blurred pattern used in the experiment.

認識実験ともに、マーカの数に対する Precision（適合率）と Recall（再現率）の調和平均である F 値を用いた。

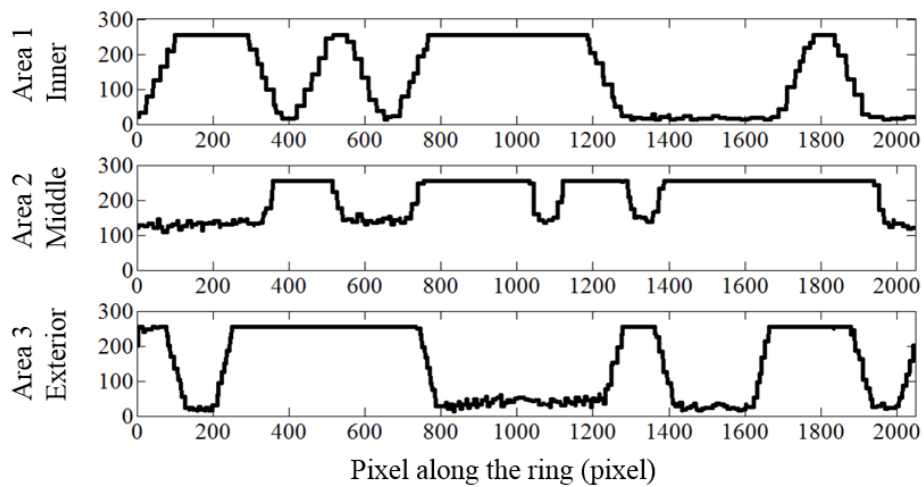
スケールスペースの極値検出のパラメータとして、スケールの刻み幅は $k = 2^{1/3}$ ，初期スケールは $\sigma = 1.2$ とし，1 オクターブを 3 ステップで刻み，3 オクターブまで探索した（計 9 ステップ）。

4.5.2 ボケに対する検出・認識性能の比較

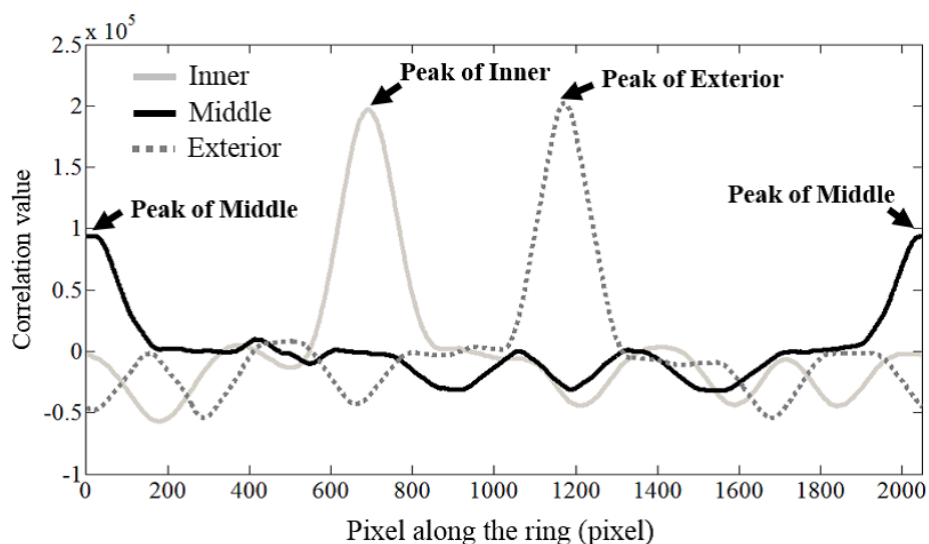
カメラの焦点面の位置調整をして，ボケの程度を変化させた条件下で実験した．カメラとマーカの配置を固定し，カメラから 1 m の距離にマーカを配置した．10 種類の各マーカを 100 回計測して計 1,000 個のデータで評価した．ボケ度合の様子を図 4.7 に示す．検出・認識結果を表 4.1 にまとめた．この表より，検出・認識いずれにおいても，提案マーカは ARToolKit マーカに比べて優れた結果を示した．この結果から，提案マーカはボケの影響を受けにくいといえる．この点は次に述べる，(i) 検出処理，(ii) 認識処理 の観点からも確認できる．

4.5.2.1 ボケ発生時の検出性能

図 4.4 に，シミュレーションによりボケ画像を作成しキーポイントの検出位置を確認した結果を示す．マーカ候補点であるキーポイントは，ボケ画像においても正しく検出できていることが分かる．マーカの配置が斜めであっても同様であることが分かる．このように，提案マーカではボケによる検出性能の劣化は小さいと考えられる．これはスケール



(a) M-sequence signal in a blurred pattern.



(b) Decoded signal in a blurred pattern.

図 4.8: ボケ画像に対するマーカのデータ信号

スペースでの極大値を検索しているためであり、点広がり関数 (point spread function) が Gaussian 関数で近似できるようなボケについては、どれだけボケても極大値を示す箇所は極大値であり続けるため影響を受けにくい。

一方、ARToolKit は輪郭線抽出を行うことで矩形を検出し、その縦横 1 辺の長さを計測して正方形に近いものをマーカとして検出する。ARToolKit マーカの検出性能が低い

表 4.2: 雑然とした環境下での多様な配置に対する精度 (F 値)

Marker	Detection	Decoding
ARToolKit	0.957	0.527
Proposed	0.990	0.921

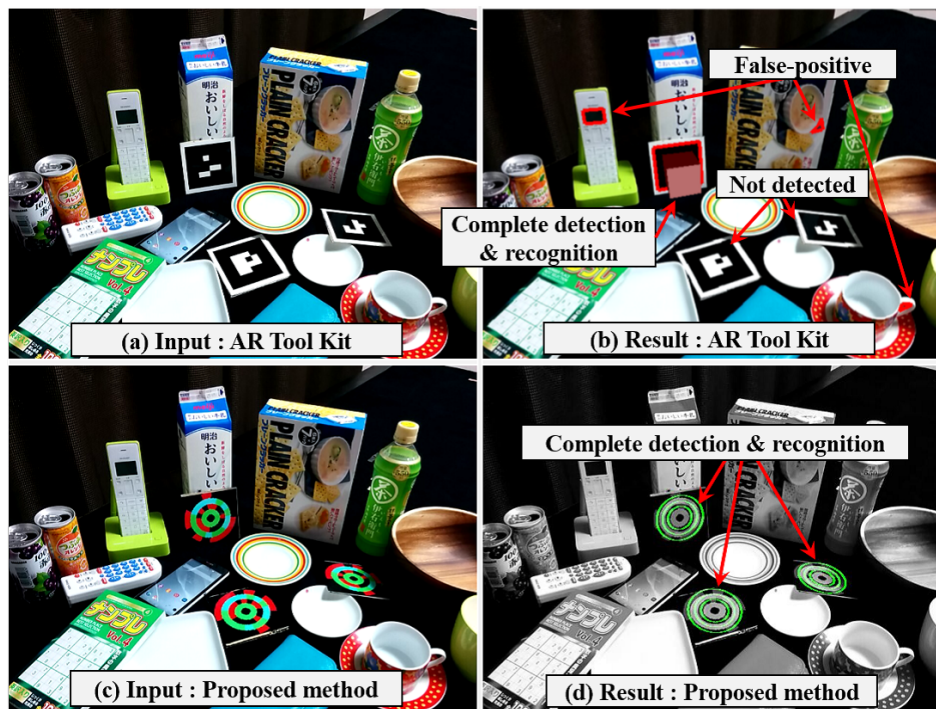


図 4.9: 雑然とした環境下でのマーカ検出結果

のは、ボケの影響で輪郭線抽出後にマーカ形状を矩形として判断することができないためである。さらに ARToolKit は、パターンの輝度に対する相関値を用いてマーカを探索するため、ボケにより輝度値に変化が生じるとその影響を強く受けてしまうため性能が低下したと考えられる。

4.5.2.2 ボケ発生時の認識性能

図 4.7 に示した実際のボケ画像に対するマーカのデータ信号を図 4.8(a) に、その畳み込み計算結果を図 4.8(b) に示す。マーカは図 4.3(i) の 2 段目右側のマーカを使用し、ボケを含む画像である。図 4.8(a) を見ると、ボケの影響で信号は矩形波を平滑化したような形状に変化している。しかし、図 4.8(b) では急峻なピークが 3 つ (外輪, 中輪, 内輪)

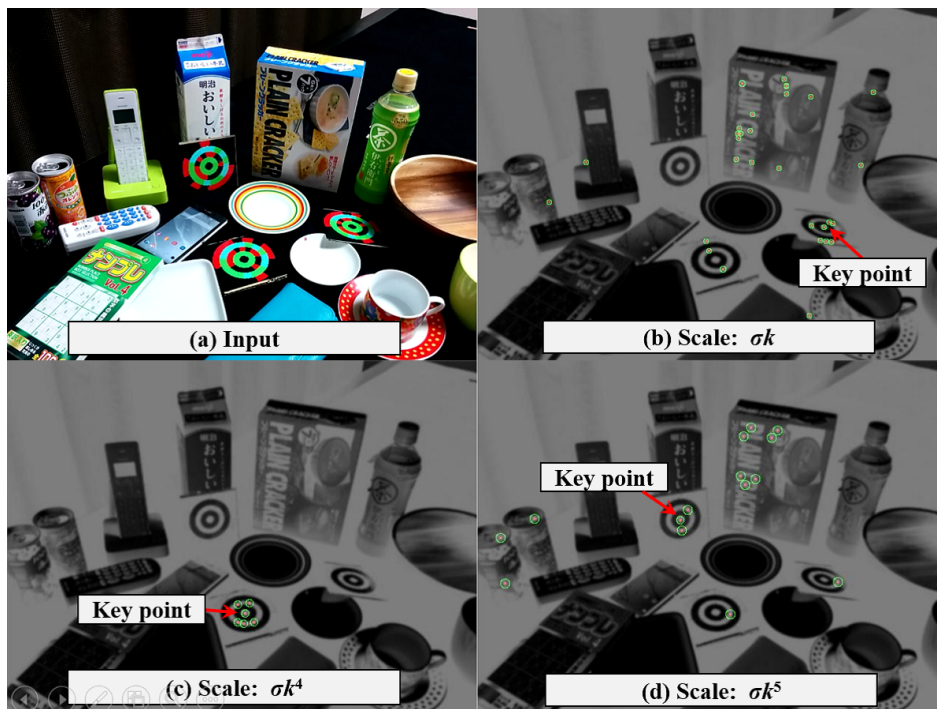


図 4.10: キーポイント検出によるマーカ候補結果

現れていることが分かる．このピークの位相差から正確に復号が可能である．また，データの符号化に M 系列を使用することにより，輝度の影響も受けにくくなっていることが分かる．以上より，提案マーカは，認識性能においてもボケに対して頑健であるといえる．

4.5.3 雑然とした背景環境下での多様な配置に対する検出・認識性能の比較

図 4.9 に示すような雑然とした背景雑音下で，物体の影がマーカに重なるような環境を 5 種類用意して実験した．10 種類のマーカそれぞれを毎回様々な距離・姿勢で配置し，固定カメラから 100 回計測することで 1 つの環境につき計 1,000 個のデータを用意した．また，提案マーカと ARToolKit でマーカとカメラは同じ場所に配置して実験した．その結果を表 4.2 にまとめる．この表より，検出・認識いずれにおいても，提案マーカが大きく優れた結果を示した．

また図 4.9(b) は ARToolKit マーカの結果，図 4.9(d) は提案マーカの結果の 1 例である．検出できた ARToolKit マーカには赤線の枠を，認識できたものには 3D オブジェクトを表示させた．一方，提案マーカは，検出・認識共に成功したものにデータ部のリン

表 4.3: 距離に対する評価 (F 値)

Distance	Detection	Decoding
1 m	0.985	0.984
2 m	1.000	0.876

表 4.4: 角度に対する評価 1 m 地点にて (F 値)

Yaw angle	Detection	Decoding
0 deg	0.985	0.984
30 deg	1.000	0.970
60 deg	1.000	0.942

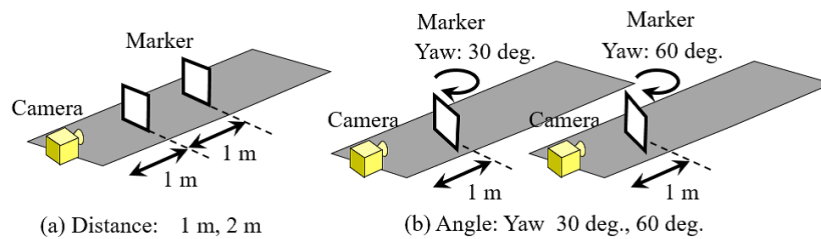


図 4.11: 距離と角度に対する実験条件

グ形状部へ線を重畳して描画した．この図から分かるように，ARToolKit マーカでは誤検出が多く現れていることが分かる．また手前 2 つの ARToolKit マーカを検出できていない．これは周辺物体の影の影響でパターンを正確に検出できないためである．これは、表 4.2 の結果にも顕著に表れている．これらのことから，ARToolKit マーカは輝度値の変化に敏感であるといえる．一方，提案マーカは雑多な配置でも的確にマーカを検出することができ，さらにマーカに影が重なっても正確に認識することが可能であった．以下，背景雑音下での検出性能の観点で考察する．

4.5.3.1 背景雑音下の検出性能

図 4.10 に，4.4.1.2 の判別フィルタを用いてマーカ候補を選出した結果を示す．各マーカは配置も異なり，スケールも異なるが，マーカの中心を的確に検出できることが分かる．判別フィルタ適用後のマーカ候補の数は，スケールスペース空間の DoG 画像 1 枚あ

たり，数個から最大 10 個程度に抑えられた．このような背景雑音下でキーポイントの数が増えないのは，判別フィルタを適用しているためである．提案マーカはマーカ候補を的確に絞り込むことができている．さらにマーカが配置された姿勢や距離が多様であっても，同一の枠組みの中で検出できるのが優れている点である．つまり，マーカの姿勢や大きさに合わせたテンプレートの照合処理は不要である．

4.5.4 距離に対する検出・認識性能の分析

提案マーカの距離に対する頑健性を確認するために，カメラから 1 m，2 m の距離で性能評価を行った．その様子を図 4.11(a) に示す．マーカはカメラに対して正面を向くよう配置した．黒色の無反射紙を下に配置し，マーカ以外の被写体が写らない環境で実験した．マーカの画像上の大きさは，距離 1 m の時 54 画素，2 m の時 27 画素 四方であった．検出・認識結果を表 4.3 に示す．いずれも高い検出・認識性能を示した．

4.5.5 角度に対する検出・認識性能の分析

提案マーカの角度に対する頑健性を確認するために，カメラから 1 m の距離で，水平方向に 0 deg.，30 deg.，60 deg. 傾けて配置したマーカで性能評価を行った．その様子を図 4.11(b) に示す．検出・認識結果を表 4.4 に示す．角度が大きく傾いても検出性能は変わらなかった．これはキーポイントの検出が向きの変化に対しても頑健であるためである（第 2 章参照）．図 4.4 にも示したが，傾いて配置されたマーカであってもキーポイントの位置は正確に検出できる．また結果より，60 deg. 傾いて配置した場合でも精度良く検出できた．このことから，マーカが傾いて配置されても，その中心を検出することが可能であることを確認した．

4.6 まとめ

本研究では，ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の“姿勢推定”の高精度化を画像処理技術により解決する．本章では，局所特徴量を把持対象物体の姿勢推定に活用する方法を提案した．従来のマーカを用いた把持対象物体の姿勢推定では，マーカの様々な配置・距離における検出と認識の困難さが課題としてあった．本章では，検出方法に焦点を当て，(i) ボケに依存せず検出しやすい，(ii) 背景と配置に

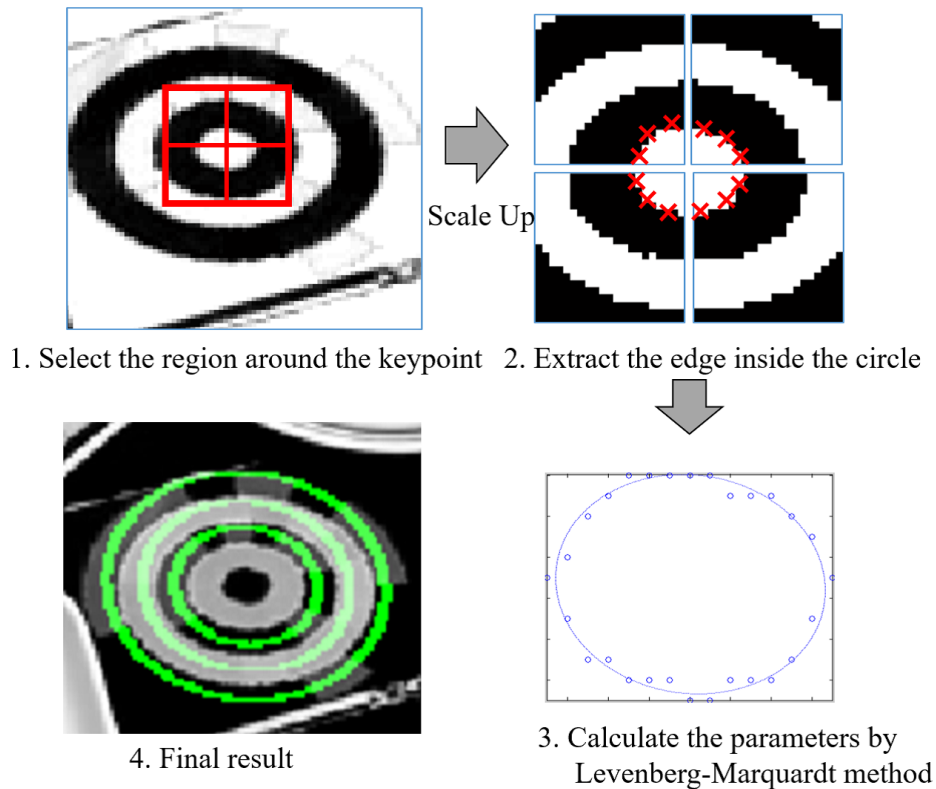


図 4.12: 楕円形状フィッティングの手順 (成功例)

依存せず検出ししやすい符号化マーカを提案し、上記の課題を解決した。そして、ロボットや拡張現実で多く使われている ARToolKit マーカに対して、マーカの検出率、認識率のうえで優位性を確認し、把持対象物体の姿勢の高精度化を実現した。

具体的には、スケールスペースの極大値として算出しやすいパターンを設計し、SIFT のキーポイント検出によってマーカ検出を実現した。更に、マーカ候補を絞り込むための判別フィルタと配置によるスケール誤差の補正方法も提案した。これらの工夫により、実環境での実験の結果、配置、ボケに頑健で、雑然とした背景環境下でも極めて高い精度でマーカの位置を検出できることを示した。

このマーカは、従来のマーカのように姿勢や大きさが異なる複数のテンプレートとの照合は不要である。そのため、様々な配置に対応するために計算コストが増加していく問題がない。またスケールスペースのスケール値に基づいてマーカの大きさを推定できる点も特長であり、その点を M 系列符号のデータデコード領域の推定に利用した。

また提案マーカは、表 4.3 では 2 m より 1 m の方が、表 4.4 では 30 deg., 60 deg. より 0 deg. の方が検出精度が若干低かった。これは、SIFT で扱うスケールが離散値であ

Levenberg-Marquardt 法で推定するパラメータ $A \sim E$:

$$\begin{pmatrix} A \\ B \\ C \\ D \\ E \end{pmatrix} = \begin{pmatrix} \sum X_i^2 Y_i^2 & \sum X_i Y_i^3 & \sum X_i^2 Y_i & \sum X_i Y_i^2 & \sum X_i Y_i \\ \sum X_i Y_i^3 & \sum Y_i^4 & \sum X_i Y_i^2 & \sum Y_i^3 & \sum Y_i^2 \\ \sum X_i^2 Y_i & \sum X_i Y_i^2 & \sum X_i^2 & \sum X_i Y_i & \sum X_i \\ \sum X_i Y_i^2 & \sum Y_i^3 & \sum X_i Y_i & \sum Y_i^2 & \sum Y_i \\ \sum X_i Y_i & \sum Y_i^2 & \sum X_i & \sum Y_i & \sum 1 \end{pmatrix}^{-1} \begin{pmatrix} -\sum X_i^3 Y_i \\ -\sum X_i^2 Y_i^2 \\ -\sum X_i^3 \\ -\sum X_i^2 Y_i \\ -\sum X_i^2 \end{pmatrix}$$

$$\text{中心の座標: } X_0 = \frac{AD - 2BC}{4B - A^2} \quad Y_0 = \frac{AC - 2D}{4B - A^2}$$

$$\text{楕円の傾き: } \theta = \frac{\tan^{-1}\left(\frac{A}{1-B}\right)}{2}$$

$$\text{X軸の長さ: } a = \sqrt{\left(X_0 \cos \theta + Y_0 \sin \theta\right)^2 - E \cos^2 \theta - \left\{\left(X_0 \sin \theta - Y_0 \cos \theta\right)^2 - E \sin^2 \theta\right\} \frac{\sin^2 \theta - B \cos^2 \theta}{\cos^2 \theta - B \sin^2 \theta}}$$

$$\text{Y軸の長さ: } b = \sqrt{\left(X_0 \sin \theta - Y_0 \cos \theta\right)^2 - E \sin^2 \theta - \left\{\left(X_0 \cos \theta + Y_0 \sin \theta\right)^2 - E \cos^2 \theta\right\} \frac{\cos^2 \theta - B \sin^2 \theta}{\sin^2 \theta - B \cos^2 \theta}}$$

図 4.13: Levenberg-Marquardt 法による楕円形状推定

り、スケールの刻み幅や初期スケール値によって感度が高いマーカの大きさが変わるためである。そのため、スケールを連続値として扱う Spectral SIFT [34] などの利用も検討する必要がある、第 5 章で検討する。

解決すべき残課題としては、より遠方に配置されて解像度が低いマーカの認識精度の向上がある。検出実験においては、マーカの解像度が低い場合でもその中心を精度良く検出できた。しかし、マーカが傾いた場合には、楕円形状のフィッティング、スケールの再計算の精度が不十分なため、認識に失敗した例が見られた。図 4.12 に楕円形状のフィッティングの手順を示す。キーポイント周辺のエッジ点を抽出して、楕円を構成するパラメータを推定する（図 4.13 参照）。マーカが遠方にあり、斜めに傾いた場合は、エッジ点を正しく抽出できなくなることが容易に想像される。本章では、楕円形状の周上に M 系列のデータを直接エンコードする手法を採った。この方法はボケに頑健な方法であったが、マーカが傾いた配置に対しては脆弱である。そこで、遠方からの認識に対して頑健なパターンを更に検討する必要がある。具体的にはドットのみでマーカを構成できれば問題が解決できると考え、これも第 5 章で検討する。

第 5 章

局所特徴量を活用した“検出・認識しやすい”姿勢推定用マーカ

本研究は、ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の姿勢推定の高精度化を画像処理技術により解決する。これに対して、第 4 章では、把持対象物体の姿勢推定に対して、局所特徴量を活用したマーカを提案した。マーカを用いた把持対象物体の姿勢推定では、マーカの様々な配置・距離に対する検出と認識の難しさがある。第 4 章では、局所特徴量のキーポイント検出の考え方を導入した“検出しやすいマーカ”を提案した。これに対して、本章では、検出しやすい特徴に加え、“認識しやすいマーカ”を提案する。

このマーカにおいて、以下の問題があった。

- (i) SIFT で扱うスケールが離散値であるため、マーカとカメラ間の距離によっては正しくマーカを検出・認識できない問題
- (ii) 遠方からマーカを正しく認識できない問題

上記の問題点 (i) は局所特徴量をマーカに活用する上での SIFT 特有の問題である。これは、マーカの検出に SIFT でなく、Spectral SIFT [34] を用いることで解決を図る。一方、問題点 (ii) は、前述したマーカ特有の課題である。マーカ検出とデータ表現の双方を区別せずに同じドットのみで実現することで対策する。マーカ検出領域とデータ領域を分けなため、マーカの大きさをより小さくすることができ、遠方からの認識にも頑健になる。

具体的には、提案するマーカはランダムなドットパターンから構成され、ドットの検出に Spectral SIFT [34] を用いる。そして、第4章で提案した“ドットのエッジ強度による判別フィルタ”と“ドットのスケール推定による判別フィルタ”を用いることで、的確にマーカのドットのみを絞り込む。さらに、ランダムドットから構成される幾何特徴として、ドットから構成される3角形の面積比（アフィン不変な量）を用い、その面積比を Locally Likely Arrangement Hashing（以降、LLAH）[91] という方法で高速に検索する。つまり、ドットの並び方に基づいてマーカを検出する。一方で、データ表現はドットの大小により符号化し、マーカ検出に用いたのと同じドットを用いることで、マーカ検出領域とデータ領域を共通化する。

5.1 姿勢推定マーカの課題

従来の ARToolKit マーカはマーカの姿勢を推定することできるが、雑然とした環境では4角形や菱形などの類似形状に誤検出してしまう。さらに、焦点ボケや外形の枠が少しでも影などに干渉されると検出できない問題がある（問題点 (a)）。これは、第4章で議論したマーカの検出精度が、背景やマーカの配置に依存してしまう問題であり、既に第4章で解決済である。

一方、第4章で提案したマーカは、雑然とした環境下での見つけやすさは実現できている。また、ARToolKit マーカ [16] のように必要な画像パターンを事前に準備する必要もない。しかし、データを円層状にエンコードするため、形状（円もしくは楕円）を正しくとらえることができない場合には、エンコードが難しくなってしまう。特に、マーカの被写体が遠く、画像での大きさが小さい場合などは、正確に楕円形状フィティングができず、データを正確に読めない（問題点 (b)）。これは前述した、(ii) 遠方からマーカを正しく認識できない問題に該当する。

またランダムドットマーカ [92] がこれまで提案されている。このマーカは、ランダムに配置したドットより、特徴パターンを表現する。特微量は、ドットから構成される3角形の面積比（アフィン不変な量）であり、その検索には LLAH [91] を用いている。この方法は、ドット検出の工程にエッジ検出（輝度による2値化処理）を行う。周囲にマーカ以外の物体が存在しない環境では正しく動作するが、雑然とした環境ではマーカ以外の箇所でもエッジが多く検出され、マーカを正しく検出することが難しい（問題点 (c)）。これは、

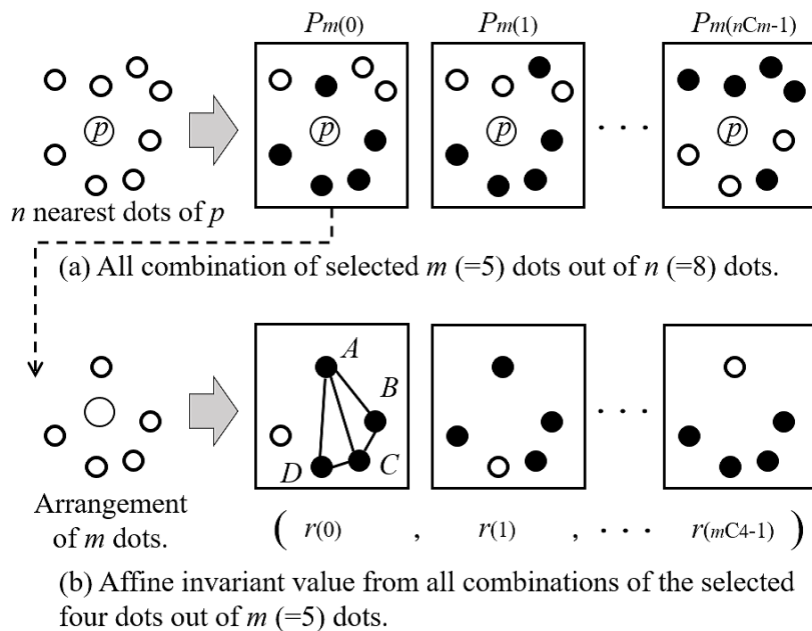


図 5.1: 特徴量の安定性と特徴量表現

上述した問題点 (a) と同様で、マーカの検出精度が、背景やマーカの配置に依存してしまう問題であり、既に第 4 章で解決済である。

上述の問題点 (c) は、第 4 章で提案した SIFT を活用した方法によって、ドット検出することで解決できると推察する。また、上述の問題点 (b) は、ランダムドットマーカのように、データをドットの配置で表現することで解決できると考える。

本章で提案する“Spectral Random Dot マーカ”は、ランダムドットマーカを改良し、ドットの検出に Spectral SIFT を活用し、更にデータの表現はドットの配置とそのスケールによって表現する。上述した問題点 (a), (b), (c) の全てを解決する方法である。

5.2 先行研究：ランダムドットマーカ

5.2.1 特徴量表現

先行研究であるランダムドットマーカについて説明する。ランダムドットマーカは、ランダムドットから構成される幾何特徴として、ドットから構成される 3 角形の面積比（アフィン不変な量）を用いる。そして、その 3 角形の面積比を LLAH [91] と呼ばれるハッシュに基づく方法で高速に検索することで、マーカを検出する。マーカに対して遮蔽が発生した状況でも、マーカを認識できる利点がある。LLAH は、文書検索などで実用化され

```

For each  $p \in \{\text{feature points}\}$  do
   $P_n$  : Class of  $n$  neighboring dots from dot  $p$ 
  For each  $P_m \in \{\text{All combinations of } m \text{ dots out of } P_n\}$  do

    (1) Select  $p_0$  from  $P_m$  randomly, make  $L_m = (p_0, \dots, p_{m-1})$  : A
    sequence of feature dots from the elements of  $P_m$ , whose
    order is clockwise from  $p_0$ .

    (2) Make a subsequence (  $L_{4(0)}, \dots, L_{4(i)}, \dots, L_{4(nC_m-1)}$  ) from
     $L_m$  by selecting four dots

    For  $i = 0$  to  $mC_4-1$  do
      (3) Calculate the affine invariant value from  $L_{4(i)}$  :  $r(i)$ 
    end For

    (4) Calculate the index of hash table and register “marker ID,
    Dot ID,  $r(0), \dots, r(mC_4-1)$ ”

  end For
end For

```

図 5.2: 登録処理アルゴリズム

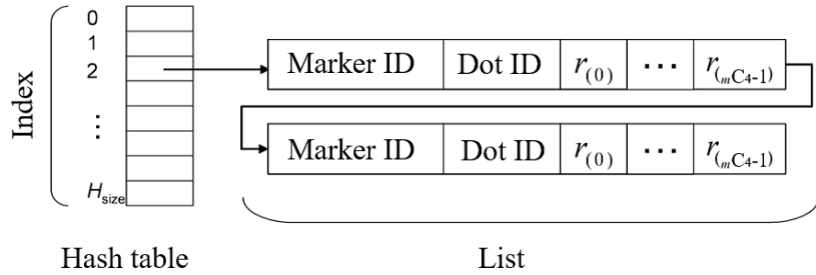


図 5.3: ハッシュ表とリスト

ており，10,000 ページの文書から該当する文書を 100 ms 程度で検索可能である．

(A) エッジ処理，(B) 2 値化処理，(C) 膨張／収縮処理，(D) ラベリング処理を経て，ドット（特徴点）を検出する．その後，このドットを識別する．具体的には，注目点と周囲の点群との幾何学的な配置から特徴点を同定する．手がかりは，同一平面上の 4 点 ABCD から以下の式で計算されるアフィン不変量を用いる．

$$r_{(i)} = \frac{T(A, C, D)}{T(A, B, C)} \quad (5.1)$$

ここで， $T(A, B, C)$ は，頂点 A，B，C をもつ 3 角形の面積である．近傍点の配置から特徴点 p の特徴量を計算する場合， p の近傍 4 点から計算されるアフィン不変量を特徴量

とする．しかし，射影変換に歪みが発生した場合は近傍点に異なるものが得られることがある．そのため，近傍 4 点から計算されるアフィン不変量は安定ではない．そこで，この問題を解決するために，より広い範囲での局所領域を用いる．近傍 n 点から全ての m 点 ($m \leq n$) の組み合わせ $P_{m(0)}, P_{m(1)}, \dots, P_{m({}_nC_{m-1})}$ を調査する． n 点のうち m 点の値が同じであれば，同じ m 点を計測できたと仮定して，特徴量の安定性を確保する．この様子を図 5.1 (a) に示す．本論文の特徴量表現は，さらに， m 点 ($m > 4$) から全ての 4 点の組み合わせを求め，そこから計算されるアフィン不変量を離散化した値の列 ($r_{(0)}, r_{(1)}, \dots, r_{({}_mC_4-1)}$) で表現される．この様子を図 5.1 (b) に示す．

5.2.2 登録アルゴリズム

登録処理のアルゴリズムを図 5.2 に示す．ここで，Marker ID とは，マーカを識別するための番号，Dot ID とは，特徴点を識別するための番号である．Marker ID と Dot ID は共に特徴量と紐づいており（図 5.3 参照），LLAH より求めることができる．ハッシュ表により，特徴量のインデックスを管理してメモリに格納する．ハッシュ表のインデックスは，以下の計算式より求める．

$$H = \sum_{i=0}^{{}_mC_4-1} (r_{(i)} k^i) \bmod H_{\text{size}} \quad (5.2)$$

ここで， $r_{(i)}$ はアフィン不変量の離散値， k は離散値数で 31 とした． H_{size} はハッシュ表の大きさで， $2^{32} - 1$ である．またアフィン不変量の離散値は，次式と表 5.1 に従う．

$$r_{(i)} = \begin{cases} 1 & (\tau < \phi_{(1)}) \\ k & (\phi_{(k)} \leq \tau < \phi_{(k+1)}) \\ 31 & (\tau \geq \phi_{(31)}) \end{cases} \quad (5.3)$$

5.2.3 検索アルゴリズム

検索処理のアルゴリズムを図 5.4 に示す．LLAH では，Marker ID への投票によって検索を行う．登録処理と同様の計算式で，ハッシュ表のインデックスを算出する．インデックスより図 5.3 に示すリストを得る．リストの各項目 ($r_{(0)}, r_{(1)}, \dots, r_{({}_mC_4-1)}$) が一致している場合は，Marker ID への投票を行う．最終的に，最大の得票数を得たマーカを検索結果とする．

表 5.1: アフィン不変量の離散値

k	$\phi_{(k)}$	k	$\phi_{(k)}$	k	$\phi_{(k)}$	k	$\phi_{(k)}$
1	0.0336963	9	0.366761	17	0.795556	25	1.75568
2	0.0819746	10	0.413245	18	0.873983	26	2.02188
3	0.1315070	11	0.461068	19	0.950395	27	2.39235
4	0.1762730	12	0.496194	20	1.012950	28	3.00036
5	0.2178020	13	0.533282	21	1.094030	29	4.05752
6	0.2557620	14	0.588809	22	1.210200	30	6.21834
7	0.2960170	15	0.652065	23	1.356790	31	13.1458
8	0.3302200	16	0.720762	24	1.532510		

```

For each  $p \in \{\text{feature dots}\}$  do
   $P_n$  : Class of  $n$  neighboring dots from dot  $p$ 
  For each  $P_m \in \{\text{All combinations of } m \text{ dots out of } P_n\}$  do
    For each  $p_0 \in P_m$  do
      (1) Select  $p_0$  from  $P_m$  randomly, make  $L_m = (p_0, \dots, p_{m-1})$  : A
      sequence of feature dots from the elements of  $P_m$ , whose
      order is clockwise from  $p_0$ .
      (2) Make a subsequence (  $L_{4(0)}, \dots, L_{4(i)}, \dots, L_{4(nC_m-1)}$  ) from  $L_m$ 
      by selecting four dots
      For  $i = 0$  to  $mC_4-1$  do
        (3) Calculate the affine invariant value from  $L_{4(i)}$  :  $r(i)$ 
        end For
        (4) Access to the hash table by calculating the table index
        (5) Check (  $r(0), \dots, r_{(mC_4-1)}$  ), if OK then Vote to marker ID
      end For
    end For
  end For
end For

```

図 5.4: 検索処理アルゴリズム

5.3 提案手法

提案手法では、前述のランダムドットマーカに対して、以下2つの改良を施す。

1点目は、ドットの検出に Spectral SIFT を使用し、マーカのドットのみを検出するように雑音除去する点である。これにより、効率よくマーカのドットを検出できる。従来の

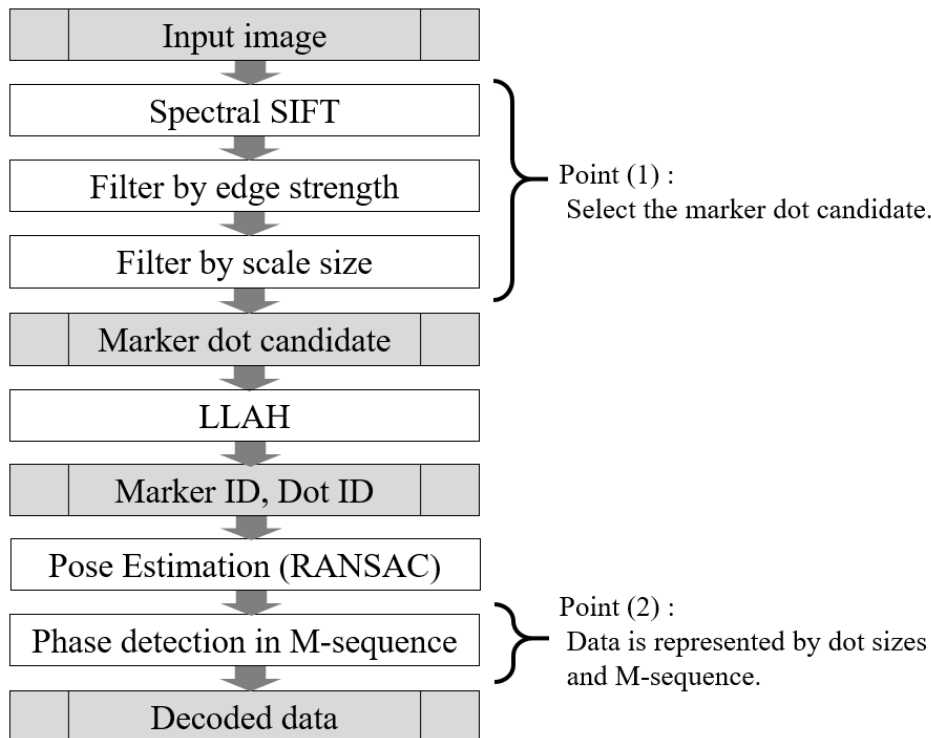


図 5.5: 提案手法の処理手順

ランダムドットマーカでは、エッジ検出，膨張・収縮を行う方法を用いているが，雑然とした環境でのマーカの検出・認識は難しい．雑然とした環境では，多くのドットが候補として検出され，その後の特徴点の識別に多くの時間が費やされるためである．また雑音となるドットが多いため，特徴点の識別精度も悪くなる．

2 点目は，ドットの大きさの大小を用いてデータを表現する点である．また，マーカを見つける検出処理とデータ表現をエンコード・デコードする認識処理の双方同じドットを用いてる．そのため両者を区別するのに比べて，マーカの大きさを小さくできる．これに対して，ARToolKit マーカは，検出のための枠とデータ表現のための領域が分かれているため，マーカの大きさを小さくすることが難しい．また，提案手法では，ドットの大きさの大小パターンの表現に M 系列の位相差を用いる．これにより特徴点検出ができない点があっても，頑健にデータ表現を行うことができる．従来のランダムドットマーカでは，256 種類のデータ表現をする場合，256 通りのランダムドットを事前に準備する必要がある．これは ARToolKit マーカと同様で，表現したいデータ数分パターンを用意しないといけない．提案手法は，データ表現を行うドット数を 8 点とすると，32 通りのドットパターンで，256 通りのパターンを表現できる（ドット数を 16 点とすれば，16 通りの

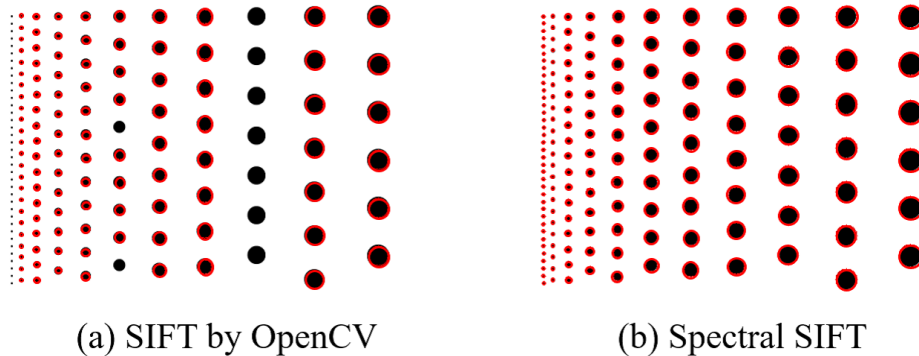


図 5.6: ドット検出の比較 (SIFT, Spectral SIFT) : 検出結果を赤丸で示す.

ドットパターンで、やはり 256 通りのパターン表現が可能).

提案手法の全体構成を図 5.5 に示す.

5.3.1 エッセンス 1 : Spectral SIFT とフィルタリングを用いたドット検出

SIFT により、輝度の張りまたは窪みがある箇所にキーポイントが出現する. 第 4 章では, SIFT のキーポイント検出を利用し, 2 重円のパターンを効率良く検出する手法を提案した. またパターンが斜めに配置されても, ある程度の範囲であれば検出できることも示した. SIFT では, 離散化されたスケール空間で DoG (Difference of Gaussian) を作成し, その極値からキーポイントを求める. 離散化されたスケールでは, 被写体の大きさとスケールによっては, ドットの検出ができない.

一方, Spectral SIFT は, 任意のスケールを扱える. Gaussian スケール空間および, SIFT などを用いられる Scale Normalized LoG 空間に対してスペクトル分解を施し, 固有値問題を積分方程式の問題に帰着させ, 多項式で近似することで連続的な固有解を求める. Spectral SIFT は, 連続値のスケールを扱えるため, 被写体の大きさ (画像上のドットの大きさ) によらず, ドットを検出できる. SIFT 及び Spectral SIFT でドットを検出した様子を図 5.6 に示す. Spectral SIFT では, 全てのドットを検出できている. つまり, Spectral SIFT を用いることで, 被写体をカメラから任意の距離に置くことができるようになる. 提案手法はこの点に着目し, ランダムドットマーカのドットを Spectral SIFT で検出する.

Spectral SIFT を用いて検出すると, マーカのドット以外にも多くの点を検出してしまふ. そこで, 以下の 2 つの処理を経て, 候補点を絞り込む. 1 つ目は, 第 4 章で述べた

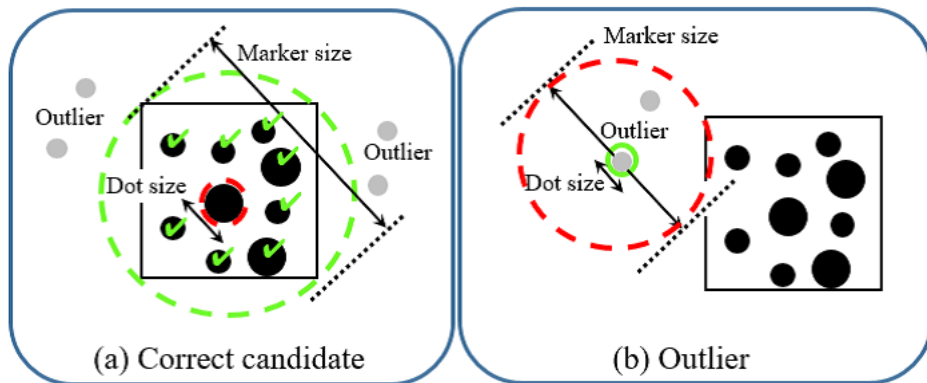


図 5.7: ドットのスケール推定によるフィルタリング処理

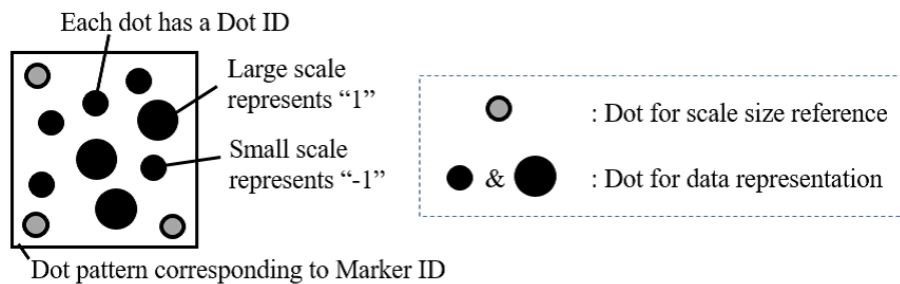


図 5.8: ドットスケールの大小によるデータ表現

エッジの強度度合による判別フィルタ，2 つ目はドットのスケール推定による判別フィルタである．前者は第 4 章と同じため，ここでは，後者のドットのスケール推定による判別フィルタについてのみ説明する．

図 5.7 に示すように，Spectral SIFT を用いてドットの大きさを推定することができる．マーカのドットの大きさは，大きい場合と小さい場合の 2 種類ある．データ表現として各々 1，-1 を割り当てた．大きい方のドットとマーカの大きさとの比は，事前に定義する設計値である．

各ドット候補に対して，推定したドットのスケールに基づいて，大きい方のドットであると仮定してマーカの大きさを推定する．その範囲内に，ある範囲内のスケール値を持つドットの数が多い値 $T (= 8)$ 以上あった場合は，マーカのドット候補として採用する．それ以外のドットは雑音として除外する．

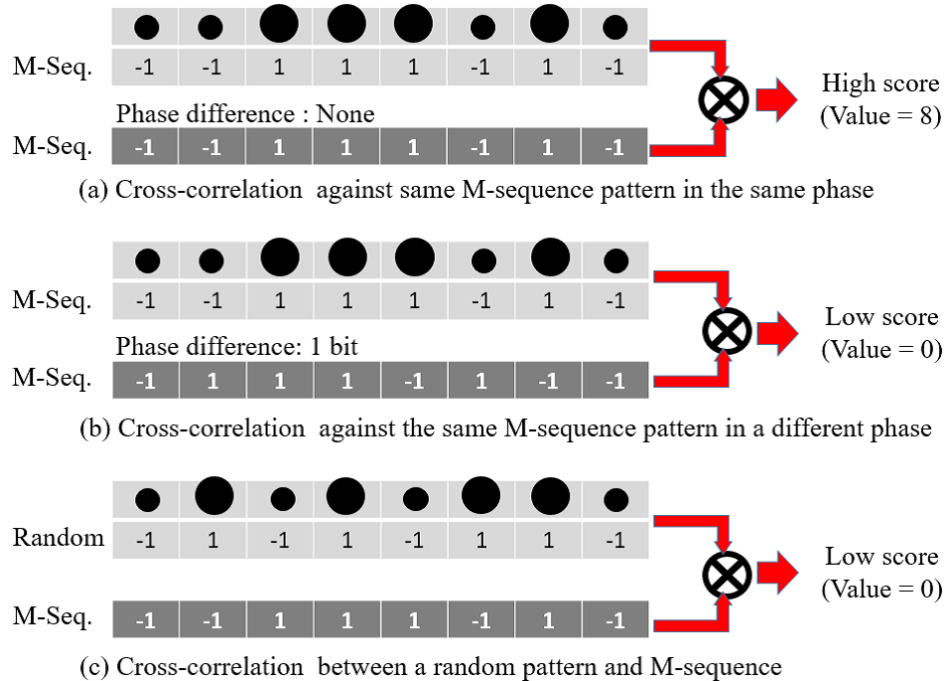


図 5.9: ドットスケールの M 系列によるデータ表現

5.3.2 エ夫点 2：ドットスケールによるデータ表現とデータ欠損時の対策

マーカのドット候補を特定した後、ドットの大きさの判定を実施する．図 5.9 に示すマーカパターンの隅 3 点（右下，左下，左上のコーナドット）を基準点として，それらとの比較する．

更に，Dot ID の順番に従って，ドットスケールの大きさを M 系列の位相として表現し，Marker ID と位相の組み合わせを最終的なデータ表現とする．M 系列を使用した点は，LLAH において，Dot ID の識別が失敗した場合にも復元できるようにするためである．M 系列は，自己相関が高く，相互相関が低い特性を持つ．同じ M 系列で位相差が異なれば，相互相関は低い．一方，位相が同期していれば，数点の Dot ID の識別に失敗しても，自己相関は高くなる．そこで，M 系列の位相差をデータ表現として利用すれば，LLAH による Dot ID の誤識別に対しても頑健になる．この様子を図 5.9 に示す．

図 5.9 (a) では，M 系列の位相が同期しているため，相関が高い．一方，図 5.9 (b) では M 系列間の位相がずれているため，相関が低い．また，図 5.9 (c) のように M 系列とランダムパターンの間では，相関が低い．例えば，図 5.9 (a) の例で，Dot ID が 2

点正しく求まらなかったとしても、相関値は“4”で高い。

ここで使用する M 系列は、7 ビット M 系列の末尾に便宜上 -1 を加えた 8 ビット系列を用いる。ドットのスケールパターンは、8 ビット M 系列 $[-1, -1, 1, 1, 1, -1, 1, -1]$ を用いた。これは、8 点のドットを用いてデータ表現したいためで、8 ビット M 系列であれば、どのような系列でも良い。8 ビット M 系列では、位相が 1 ビットシフトしたものは、**図 5.9** (b) であり、同様にビットシフトを行うと 8 通りの位相シフトパターンが作成できる。よって、データの表現は、M 系列の位相差 (8 通り) とドットパターンである Marker ID (32 通り) の組み合わせの、256 通りを表現することにした。

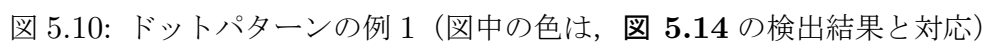
5.4 実験と考察

解像度 $1,280 \times 720$ 画素の USB カメラを用いて実験した。ただし、計算処理に用いる画像解像度は VGA に変換してから使用した。使用したマーカパターンの大きさは 2.5 cm 四方で、構成ドット数は 11 点、そのうちデータ表現に 8 点、コーナ点を 3 点とした。使用したマーカパターンを**図 5.10** と**図 5.11** に示す。

5.4.1 ドット検出方法の比較

ドットの検出において、従来手法のランダムドットマーカで使用されているエッジ処理による手法と、提案する Spectral SIFT を用いる方法とを比較する。また提案手法は、Spectral SIFT の後段処理として、雑音除去を施している。

図 5.12, **図 5.13** に結果を示す。**図 5.12** (b) より、従来手法では、マーカのドット以外のエッジ上の点が多く検出されていることが分かる。一方、**図 5.13** (b) より提案手法では、的確にマーカのドットを検出できていることが分かる。しかし、壁面上の小さな凹凸や他のテクスチャにも反応してしまっている。これを、エッジ強度による判別フィルタで選別したのが、**図 5.13** (c) である。細かい窪みに反応したキーポイントを除去できたことが分かる。さらに、**図 5.13** (d) では、ドットスケールの推定による判別フィルタを用いて、マーカのドット以外の点を除去し、所望のマーカのドットのみが得られた。ここで、LLAH への入力として、極力マーカのドットだけに絞り、計算負荷を抑えることが大切である。本実験では、Intel Core i5 CPU 上で 1 Hz 以上の速さで動作した。



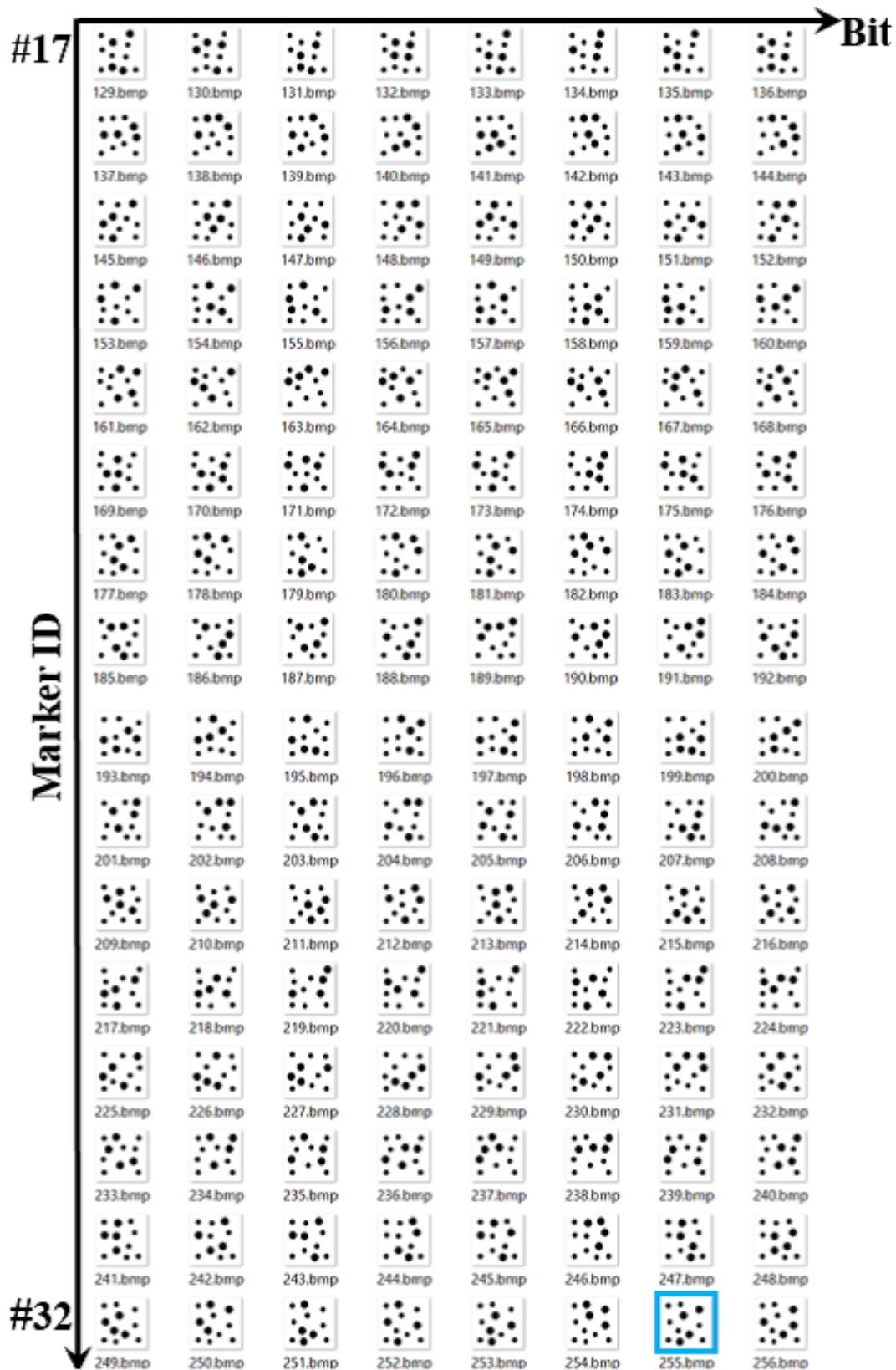


図 5.11: ドットパターンの例 2 (図中の色は, 図 5.14 の検出結果と対応)

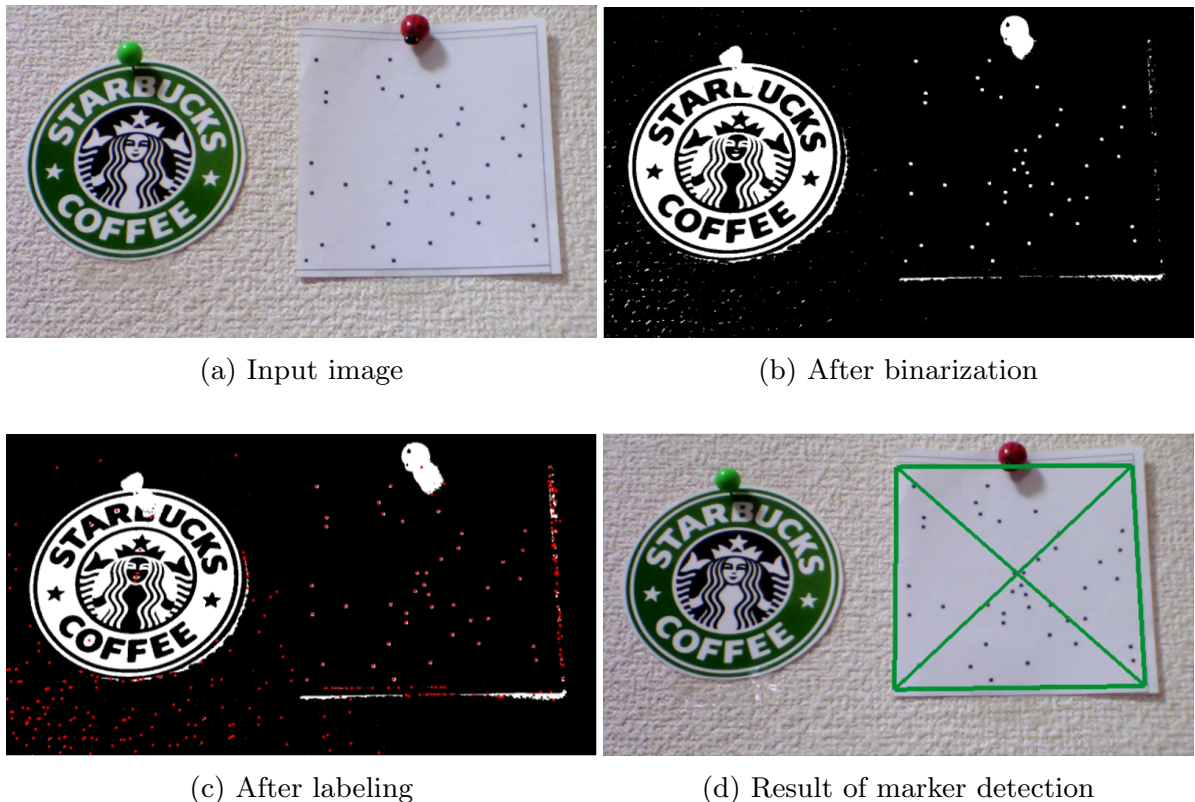


図 5.12: Random Dots Markers によるドットの検出

5.4.2 様々なマーカ配置におけるデータデコードの頑健性の分析

データデコードの結果を図 5.14 に示す. 図 5.10 と図 5.11 の中から 5 つマーカを選択して実験に用いた. 各マーカはカメラから 70 cm 前方に様々な向きで配置した. 図 5.14 (d) にデコードが成功した結果を示した. マーカの配置が様々な向きであっても, 正しくマーカの姿勢を推定することができた. これは, ロボットの把持タスクを行う上で重要である. さらに, 缶やコップといった曲面からなる物体に貼りつけたマーカに対しても, デコードが成功した. これは, アフィン不変量が離散値であること, さらに, マーカの大きさが小さいことの 2 つから, 多少の湾曲があっても特徴量の値が変わらなかったためである. 図 5.14 (b) は, LLAH に入力する点群である, 先の実験でも示したように, 背景雑音は全く検出されないことが確認できる. Spectral SIFT によるドット検出や雑音除去が機能していることが分かる. 更に, 図 5.14 (c) は LLAH により, 点を同定した結果である. 点の色が Marker ID と対応しており, いくつかの点では誤検出

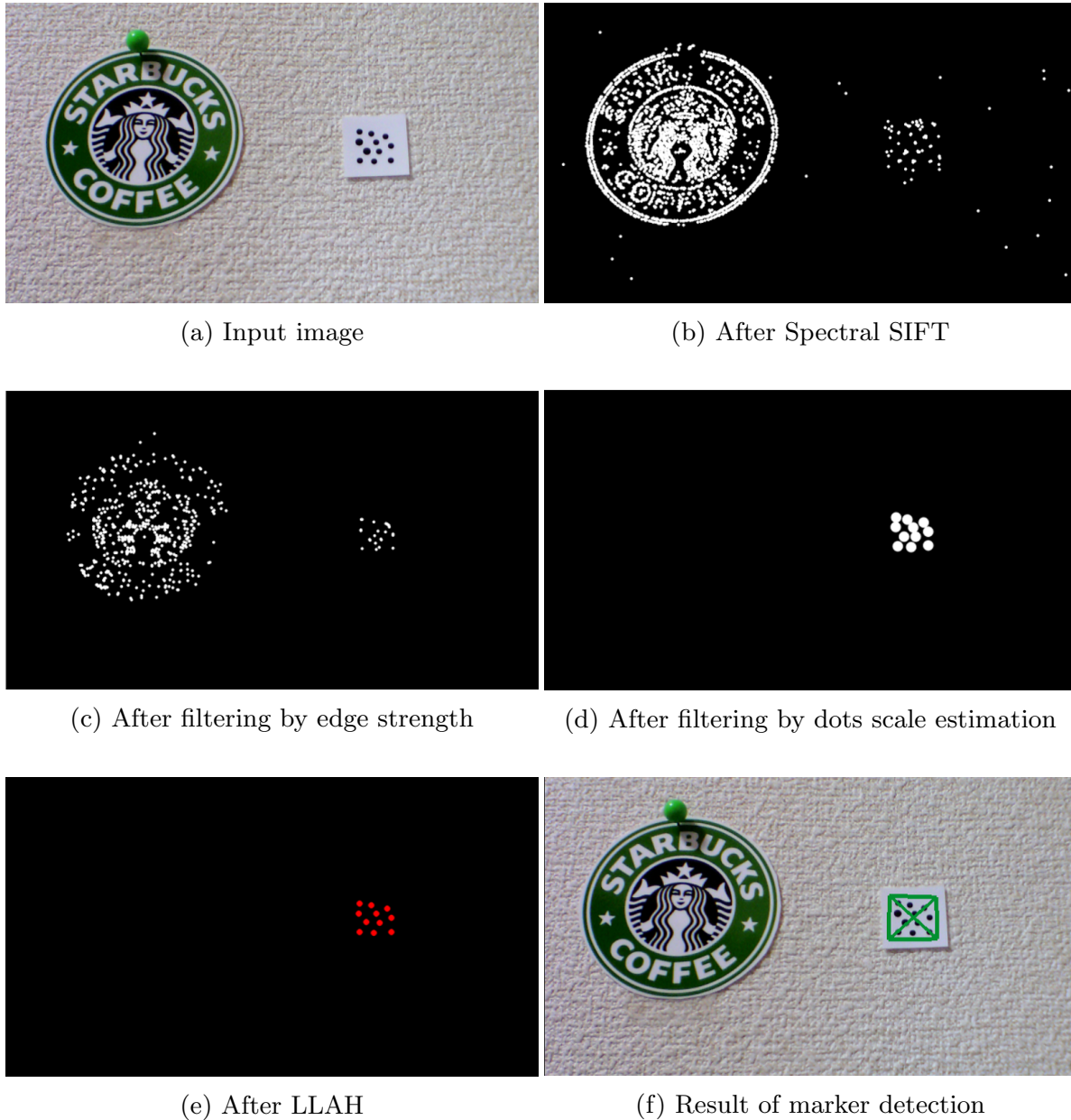


図 5.13: Spectral Random Dot マーカによるドットの検出

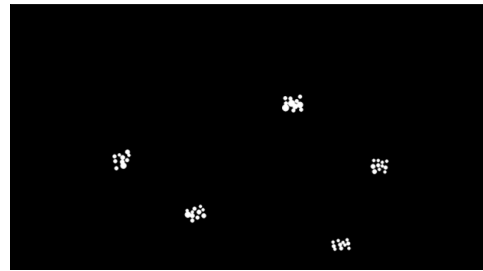
がある。しかし、RANSAC を行うことで Marker ID が誤検出されたドットは除去され、最終的には正しい姿勢を求めることができた。数点の誤検出があっても、M 系列を用いた位相差によるデータ表現は正しくデコードできることが分かる。

5.4.3 距離に対する検出・認識性能の分析

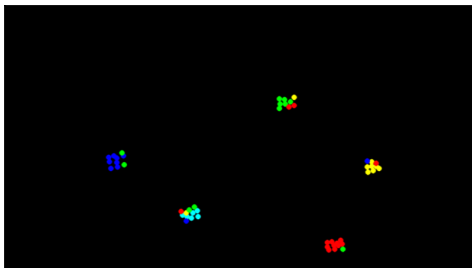
図 5.15 にマーカまでの距離を変えた場合のマーカの検出・認識結果を示す。提案す



(a) Input image



(b) After Spectral SIFT and denoising filtering



(c) After LLAH



(d) Result of the identifications of markers and their decoding

図 5.14: 様々な配置でのマーカ検出の例

るマーカは、100 cm まで検出可能であった。各距離ごと、マーカとカメラを固定して 1,000 frame 計測し Recall と Precision を算出した。その結果を表 5.2 に示す。Recall は距離と共に減少するが、それでも 85.2 % 以上の高い値を維持した。5.3.1 項で述べたように、任意の距離に対してドット検出を正しくできることを確認した。

第4章で提案したマーカとの比較をする。表 4.3 の結果では、200 cm 地点での F 値が 87.6 % だった。また、マーカの大きさが 27 画像四方であった。一方、100 cm 地点での Spectral Random Dot マーカは、コーナ一点の間隔が 20 画素と小さいにも関わらず、F 値は 88.0 % と上回ることができた。

また ARToolKit マーカで、マーカの大きさが 2.5 cm 四方の場合に、同様の実験を行ったところ、100 cm 地点で Recall が 86.1 %, Precision が 63.7 %, F 値が 73.22 % だった。この場合も、提案するマーカの方が、遠方からの認識に対してより良いことが分かる。更に、ARToolkit マーカでは検出パターンとデータ領域が分かれているのに対して、提案するマーカではパターンの検出とデータ表現の領域が共通である。そのため、同じマーカサイズあたりでは、提案するマーカの方が、データ領域を広くとることができる。

表 5.2: 距離に対する認識性能 (単位: %)

Distance (cm)	Recall	Precision	F 値
30	94.2	100	97.0
50	91.5	100	95.6
80	87.3	100	93.2
100	85.2	91.2	88.0

5.5 まとめ

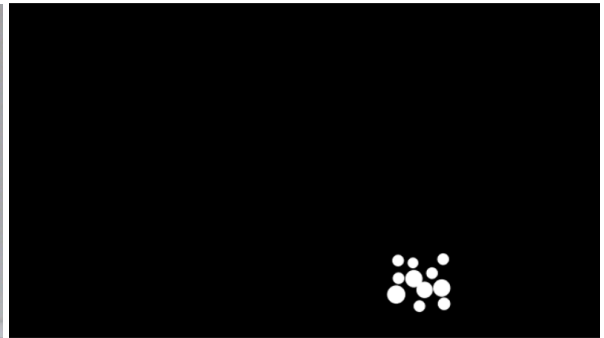
本研究では、ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の“姿勢推定”の高精度化を画像処理技術により解決する。マーカを用いた把持対象物体の姿勢推定では、マーカの様々な配置・距離における検出と認識の困難さが課題としてあった。本章では、第4章で提案した局所特徴量を活用したマーカを更に発展させ、検出だけでなく、認識もしやすいマーカに改良した。

本章で提案した Spectral Random Dot マーカは、従来のランダムドットマーカを改良し、Spectral SIFT を用いてマーカのドットを検出、ドットの並び方でマーカを検出し、更に、データの表現についてはドットの大きさで判断した。ドットの大きさは、Spectral SIFT から精度高く推定ができるため、第4章のマーカのようなデータ領域の推定（楕円形状フィッティング）は不要である。

具体的な方法は、Spectral SIFT のキーポイント検出から得たドットの幾何学的な並び方で、LLAH によりマーカを判別する。ドットの検出においては、“エッジ強度による判別フィルタ”と“ドットのスケール推定による判別フィルタ”を用いることで、効率的にドットを選別し、計算負荷を抑えることができることを確認した。また Spectral SIFT を活用することで、任意の距離・配置においても、マーカを見つけられることを示した。更に、データを、ドットスケールの大小により表現することで、検出領域とデータ領域を共通にすることができた。これにより、遠方からの認識が可能になった。



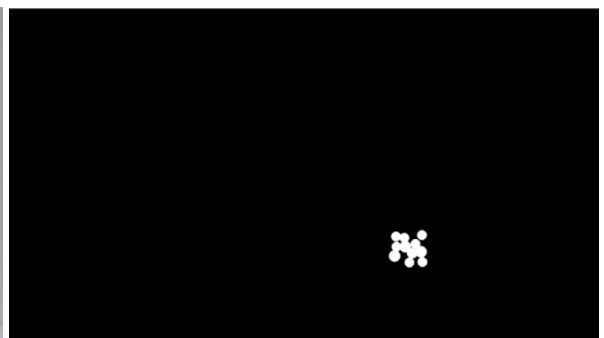
(a) Result for 30 cm from the camera



(b) Extracted dots in (a)



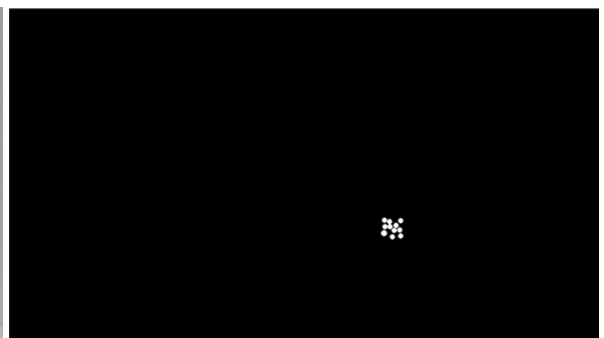
(c) Result for 50 cm from the camera



(d) Extracted dots in (c)



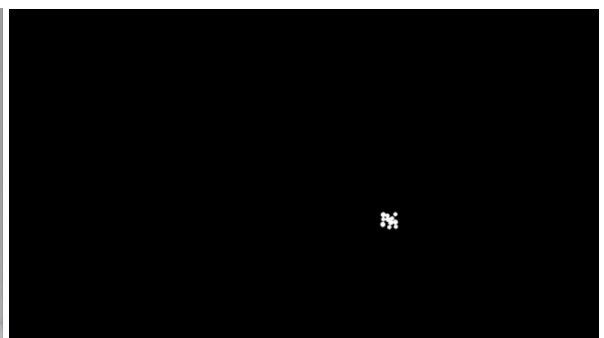
(e) Result for 80 cm from the camera



(f) Extracted dots in (e)



(g) Result for 100 cm from the camera



(h) Extracted dots in (g)

図 5.15: 距離に対する検出性能

第 6 章

結論

6.1 本研究の成果

本研究は、ロボットの姿勢推定とロボットが把持動作を行う物体の姿勢推定の双方の姿勢推定の高精度化を画像処理技術により解決することである。Mobile Manipulator における Fetch and Carry タスクを将来の応用先とし、そのための基礎検討として (1) ロボットの姿勢推定, (2) 把持対象物体の姿勢推定の 2 つの課題に取り組んだ。既存の構成であるジャイロセンサやマーカを使うことを前提としたうえで、それらに加えて局所特徴量記述子を導入することで、各々の課題を解決した。ジャイロセンサについては“ドリフト誤差”であり、マーカについては“様々な配置・距離における検出の困難さ”が課題であった。

以下に、各章で得られた結果を記す。

第 2 章では、本研究で用いた局所特徴量記述子である SIFT について説明した。さらに、カメラ視点の変動に対する頑健性を事前評価した。特徴点照合において、約 30 deg. 程度までの視点移動には十分に対応できることが分かった。想定しているロボットは、1 フレーム間 (10 fps とした場合, 0.1 s 間) で 30 deg. も大きく変化することがないため、十分な性能であると判断した。

第 3 章では、SIFT による画像フレーム間の対応づけから姿勢変化量を求め、ジャイロセンサの積算誤差を拡張 Kalman フィルタで除去することでロボットの高精度な姿勢推定ができることを示した。また、ジャイロセンサと SIFT を相補的に活用することを提案した。具体的には、画像処理により姿勢が正しく得られない場合や誤対応をする場合には、ジャイロセンサから基礎行列の初期値を求めるとともに、ジャイロセンサの出力から

誤対応を除去した。さらに、画像処理で求める姿勢推定に判定基準を設定することで、時間方向に継続して姿勢推定を高精度化できることを確認した。特に、変化量が大きいヨー方向の姿勢は FOG 相当の精度を示した。移動する方向とカメラの光軸が同じ向きであるため、基礎行列を精度良く求めることが難しい場合があった。また、ロール・ピッチ方向には、走行時の石畳みの振動のような高周波の振動があるため、10 Hz 程度のカメラ計測ではその動作をとらえることが周波数的に難しい場合もあった。前者については、移動方向に対して鉛直な向きにカメラを配置すればよく、複数のカメラを用いることで対策が可能と考えている。また後者については、信号処理をハードウェア化（例えば、FPGA で演算処理）することで、カメラの計測周波数を改善することが可能と考える。

第 4 章では、SIFT のキーポイント検出を利用して、把持対象物体に貼付したマーカを見つけやすくするように改良を実施した。ボケや雑然とした背景環境下でも、従来の ARToolKit マーカと比較して、高精度で検出できることを示した。また、円形の輝度の凸凹であれば、マーカの傾きが 60 deg. あってもキーポイント検出ができることを確認した。さらに、提案したマーカは、スケール（大きさ）の推定も同時にできることを示した。

従来の ARToolKit マーカのように、カメラからの見えやすい位置にマーカを配置する必要がなくなり、把持対象物体とカメラ間の自由度がより向上した。雑然とした環境下でもマーカ認識ができ、ARToolKit マーカが 52.7 % の F 値に対して、提案するマーカは、92.1 % と性能を大幅に向上させることに成功した。

データの符号化は、円層状（楕円）に従って、画素毎の輝度値を M 系列のパターンで表現した。ボケに対して頑健に機能することを確認できた。難点として、マーカの配置が傾いた場合に楕円が潰れてしまい、データ領域を正しく読み取ることができない場合があることが判明した。そのため、次の第 5 章でデータの符号化を検討し、ドットのみで表現できるように改良をした。

第 5 章では、第 4 章のマーカに対して、更に遠方からも認識しやすいように改良を施した。提案方法は、Spectral SIFT のキーポイント検出から得たドットの幾何学的な並び方で、LLAH を用いてマーカを判別した。Spectral SIFT を活用することで、SIFT では扱うことができなかった任意の距離からマーカを見つけられることを示した。ドットのスケール推定を活用した背景雑音除去を提案し、マーカのドットのみを的確に抽出することに成功した。さらに、ドットの並び方でマーカを検出し、ドットのスケールの大小でデータを表現をすることで、検出パターンとデータパターンを共通化することに成功した。こ

れによりマーカの大きさを小さくすることができ、同じ大きさのマーカであれば遠方から認識しやすくなった。

マーカの大きさが 2.5 cm 四方で、1 m 程度の距離から、F 値 88 %（誤差、5 mm, 5 deg. 以下の条件で評価）で認識できる。提案するマーカは、マーカの大きさを小さくすることができ、容易に物体に貼付できることが特徴である。

以上、提案した 2 つの姿勢推定手法は、実用的な従来法（ジャイロセンサ・マーカ）に局所特徴量をうまく活用したものである。提案手法は、DVO-SLAM や RGBD-SLAM などの 3D-SLAM に局所特徴量を用いた方法と比較して、以下のような利点があると考えている：(1) 環境中に人などの動的な障害物がいても良い、(2) 環境中の構造物にも大きく依存しない。3D-SLAM の場合、単調な構造物の連続（例：一面が壁）や幾何学的な特徴が掴みにくい環境（例：距離を測れないほどひらけた空間）では姿勢精度を正しく推定できない問題がある。しかし、提案するロボットの姿勢推定手法は、石畳の床のような類似したテクスチャでも問題なく動作した。更に、ロボットが走行中に通行人による遮蔽が発生したが、ジャイロセンサがあるため問題なく動作した。同様に、把持対象物体の姿勢推定手法では、環境の構造物に影響されず、マーカのみを抽出することを得意とする。このような環境への依存は 3D-SLAM と比較して少ない点も提案手法の特徴である。

6.2 今後の課題と展望

6.2.1 今後の課題

まずロボットの姿勢推定においては、近年 GPU の性能が飛躍的に向上し、SIFT も高速に実時間で計算できるようになった。さらに、局所特徴量記述子の勾配方向を CNN で学習する研究 [93] も提案され、より高精度に特徴量記述子間の対応づけをできるようになりつつある。本研究では、カメラ画像からの信号とジャイロセンサの信号の同期処理について非常に苦労した。信号の同期処理まで考慮したハードウェアを構築し、SIFT もしくは、局所特徴量記述子を学習した CNN を実時間処理できるように実装することで、ロボットのみならず自動車やドローンなど他の用途における姿勢推定にも活用できると考えている。特に、姿勢推定と制御をオンラインで同期実行した制御系も含めた評価は今後の課題である。

また、第 5 章で提案したマーカは、把持対象物体の姿勢推定において、ARToolKit マー

カよりも様々な配置からの検出やその姿勢推定には優れている。マーカを検出することで、その位置と姿勢が求まり、更にマーカにエンコードする情報をマーカの位置・姿勢に対する把持パターンとすることでロボットの把持タスクに特化したマーカができる。把持パターンを設計し、ロボットの実機を通した実験を行うことで提案するマーカを評価したい。そして、ロボット用の作業マーカとしての標準化とその普及方法も検討していきたい。また、キーポイント検出方法もまだ開発され続けており、アフィン変換に不変なキーポイント検出方法 [90] については、その効果を評価する価値があると考えている。

6.2.2 将来技術としての応用

第3章と第5章の手法を活用し、ロボットの把持タスクに応用していきたいと考えている。本論文では、“ジャイロセンサ+局所特徴量”、“マーカ+局所特徴量”の各々の構成を実験・評価したが、“ジャイロセンサ+マーカ+局所特徴量”といったよりタイトなカップリング構成も可能であると考ええる。

また、個別の技術をスピニアウトして応用することも可能である。例えば、把持対象物体の姿勢推定において開発したマーカ技術は、Virtual Reality への応用がまず考えられる。マーカが遮蔽されたり、マーカが様々な配置であってもコンテンツを表示させたい状況で使用できる。また、ランドマークとしてマーカを複数使用することでドローンなど移動体の姿勢推定にも応用できると考えている。この際、第3章で提案した手法を組み合わせることでより高精度な姿勢推定が可能である。

また、SIFT でドットを検出する方法は、様々な配置された点字の認識などにも応用できると考える。また認識したい領域に対しての位置合わせ（例：QR コードの検出パターン）などにも利用できると考える。

6.2.3 さらなる開発の方向と長期的な展望

計測した画像や点群を、平面、円柱、円錐、球などの基本形状に置き換える“基本形状への当てはめ”という先行研究がある。2次元の画像から輪郭を与えるとその部分を基本形状に置き換え3次元モデルを作成する 3-Sweep [94] や点群から3次元物体を全探索する Sliding Shape [95]、RANSAC を繰り返すことで基本形状を推定するものもある [96]。[96] は、CGAL [97] と呼ばれるオープンライブラリで使用可能である。

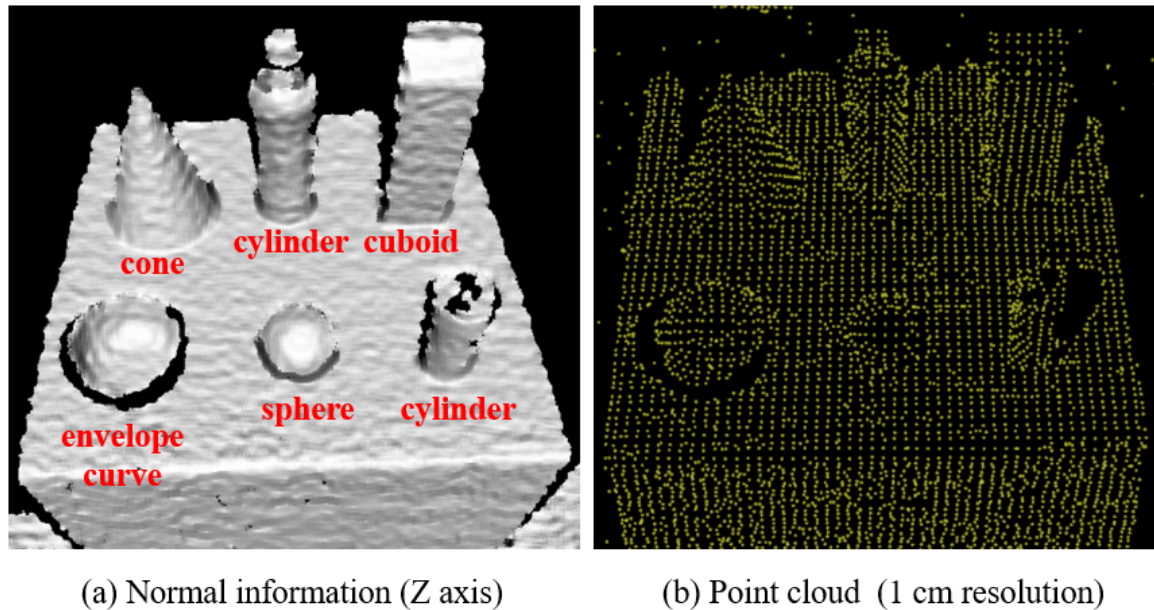


図 6.1: プリミティブ形状に対する 3 次元点群の様子

第 5 章で提案したマーカは、3 次元空間中の点（点群）により構成することも可能であり、3 次元点群の基本形状への当てはめにも活用できると考えている。平面、円柱、円錐、球などの基本形状の点群を予めドットパターンとして用意しておけば、同じ LLAH の枠組みで、同様のドットパターンに対する識別ができる。基本形状に置き換えれば、ロボットは物体の把持すべき箇所をマーカなしで判断できる。

試しに様々な物体形状を配置し、距離画像センサで撮影した。図 6.1 に、その Z 方向の法線成分とその点群を示す。図 6.1 では、各基本形状ごとに点の間隔などが異なることが分かる。アフィン不変の性質は変わらないため、式 (5.1) に示した 3 角形の面積比は、3 次元空間でも利用可能である。図 6.1 のような場合は、RANSAC により平面を検出し、Euclidean Cluster Extraction 法 [100] で各物体を検出することができるはずである。しかし、物体同士が接近し遮蔽がある場合は容易でないため、3 次元点群を基本形状毎に分類することは有用であると考えている。

基本形状の検出を行う際に用いる点群については、検討の余地がある。全点群に対して適用すると計算コストが大きくなるので、SIFT の考えを 3 次元形状に導入した方法 [98, 99] を用いて、3 次元形状の凸凹を的確に捉えた特徴ある点群を選別して、基本形状の当てはめに応用できる可能性がある。

本論文で扱った“カメラ画像、ジャイロセンサ、マーカ”によるセンシング技術は将来

的には統合され，民生製品としてより普及していくと考えられる．その基礎技術として本論文が活用されることを期待する．特に，今後のロボットは，生産現場から家庭内の生活支援ができるように変化していくと考えている．1 家に 1 台ロボットが普及するためには，タスクの設定を限定したうえで，ロボットが人間以上にそのタスクをこなせる必要がある．そのための 1 つの技術として，本論文が少しでも貢献できれば本望である．

謝辞

本博士論文は、筆者が社会人博士として名古屋大学大学院情報科学研究科博士後期課程に在籍した2008年10月から2011年9月、3年間休学を経て、2014年10月から2015年10月までの研究成果をまとめたものである。在職する傍ら、業務とは全く関係のない領域で、研究の取り組むことをしてきました。研究する時間の捻出に苦難することもありましたが、研究する楽しさと考え抜く楽しさ、自らの知的探求心を満たすことができたと思っております。本論文の執筆や研究において、名古屋大学の先生方には本当にお世話になったと思っております。

村瀬洋先生に心より感謝致します。先生からは、研究に熱心な議論、指導を頂く中で、研究者としての心の在り方を教わったと思っております。元々、早稲田大学に修士学生でいた頃、ニューロサイエンスがきっかけで、パターン認識を勉強しました。そして、面白いと感じ、自身の方向性を考えていた時代がありました。村瀬洋先生をはじめ、渡辺澄夫先生、中野馨先生、甘利俊一先生の本を読みあさり、その分野の第1線で活躍している研究者の方々に憧れを持っていました。実際に、先生にお会いし、指導を頂けたことが自身にとって光栄です。また先生は、研究の楽しさやワクワク感を常に持ち続けいる方で、加えて、大きな研究成果があっても大変謙虚な姿勢でいられることが印象的でした。一流の研究者としてのあるべき手本（鑑み）を学ぶことが出来たと思っております。心から深く感謝申し上げます。

井手一郎先生に心より感謝致します。先生からは、研究に熱心な議論、指導を頂く中で、研究者としての基本を教わったと思っております。研究を他人に理解してもらうための、論理構成や説明の仕方、論文の正しい書き方を教わったと思っております。おそらく先生の指導がなかったら、研究者の基本である論文をちゃんと書けなかったとも思います。これからも、エンジニア兼研究者として、論文を書いていこうと思っております。これからの研究で良い成果があった時、質の高い論文、世界に通用する論文が書けることは、先生のおかげ

だと思ってます。心から深く感謝申し上げます。

出口大輔先生に心より感謝致します。先生からは、研究に熱心な議論、指導を頂く中で、技術的な部分を多く教わったと思ってます。トレンドの流れが速い画像処理の分野で、最新の論文から色々な手法などを教えて頂けたと思ってます。議論していて楽しいと素で感じる事が出来ました。この博士課程を通して振り返っても、村瀬先生、井手先生、出口先生、高橋先生とのアイデアを議論する時間が自身にとって新鮮で、楽しかったと思います。心から深く感謝申し上げます。

2015 年からは、自身の業務も画像処理を専門とする研究になりました。今年実施している共同研究を始め、先生皆様には、末永くご指導を頂きたいと思っております。何卒、宜しくお願い申し上げます。

最後に、支えてくれて妻の弥生に感謝致します。結婚、闘病、出産など色々なことがありました。思ったようにいかなかったても良い、今日より明日が少しでも良くなれば。あきらめない心を持つ大切さを感じ、それを共に実践できた月日だったと思ってます。研究を楽しむ気持ちを持って、研究を続けることが出来たのは、弥生のおかげだと思ってます。心からありがとう。

参考文献

- [1] German Institute for Standardization: “Industrie 4.0 German Standardization Roadmap”,
<http://www.din.de/blob/65354/f5252239daa596d8c4d1f24b40e4486d/roadmap-i4-0-e-data.pdf> [2016/08/01 Access]
- [2] 経済産業省: “日本再興戦略”,
<http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/honbun2JP.pdf> [2016/08/01 Access]
- [3] 山本貴史, 齋藤史倫, 橋本国松, 池田幸一: “生活支援ロボット HSR の開発”, 第 30 回日本ロボット学会学術講演会予稿集, 3C2-1 (2012)
- [4] 齋藤史倫, 橋本国松, 池田幸一, 山本貴史, 畠中規, 飯島浩: “生活支援ロボット HSR の試作と実証評価”, 第 30 回日本ロボット学会学術講演会予稿集, 3C2-2 (2012)
- [5] Fetch Robotics: “Fetch”, <http://fetchrobotics.com/> [2016/08/01 Access]
- [6] T. Chen, M. Ciocarlie, S. Cousins, P. Grice, K. Hawkins, K. Hisao, C. Kemp, C. King, D. Lazewatsky, A. Leeper, H. Nguyen, A. Paepcke, C. Pantofaru, W. Smart, and L. Takayama: “Robots for humanity: A case study in assistive mobile manipulation”, IEEE Robotics and Automation Mag., Vol.20, No.1, pp.30–39 (2013)
- [7] Open Source Robotics Foundation: “ROS”, <http://www.ros.org/> [2016/08/01 Access]
- [8] Rethink Robotics: “Baxter”, <http://www.rethinkrobotics.com/baxter/> [2015/09/01 Access]
- [9] 川田工業 (株): “NEXTAGE”, <http://nextage.kawada.jp/> [2015/09/01 Access]
- [10] Amazon Robotics : “Amazon Picking Challenge”,

- <http://amazonpickingchallenge.org/> [2016/08/01 Access]
- [11] Kiva Systems: “Kiva”, <https://www.amazonrobotics.com/> [2016/08/01 Access]
- [12] H. Durrant-Whyte and T. Bailey: “Simultaneous localisation and mapping (SLAM): Part I The essential algorithms”, *IEEE Robotics and Automation Mag.*, Vol.2, No.1, pp.469–476 (2006)
- [13] O. Ghita and P. Whelan: “A bin picking system based on depth from defocus”, *Int. J. of Machine Vision and Applications*, Vol.13, No.4, pp.234–244 (2003)
- [14] Itseez: “OpenCV”, <http://opencv.org/> [2016/08/01 Access]
- [15] Open Perception Foundation: “PCL”, <http://pointclouds.org/> [2016/08/01 Access]
- [16] H. Kato and M. Billinghurst: “Marker tracking and HMD calibration for a video-based augmented reality conferencing system”, *Proc. 2nd IEEE/ACM Int. Workshop on Augmented Reality*, pp.85–94 (1999)
- [17] S. Ren, K. He and R. Girshick, and J. Sun: “Faster R-CNN: Towards real-time object detection with region proposal networks”, *Proc. 2015 Int. Conf. on Neural Information Processing Systems*, pp.85–94 (2015)
- [18] S. Hinterstoisser, C. Cagniart, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit: “Gradient response maps for real-time detection of texture-less objects”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.34, No.5, pp.876–888 (2012)
- [19] E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton, and C. Rother: “Learning 6D object pose estimation using 3D object coordinates”, *Computer Vision, ECCV 2014, 13th European Conf. Procs., Part II, Lecture Notes in Computer Science*, Vol.8690, pp.536–551, Springer (2014)
- [20] A. Doumanoglou, R. Kouskouridas, S. Malassiotis, and T. Kim: “Recovering 6D object pose and predicting next-best-view in the crowd”, *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition*, pp.3583–3592 (2016)
- [21] C. Kerl, J. Sturm, and D. Cremers: “Dense visual SLAM for RGB-D cameras”, *Proc. 2013 Int. Conf. on Intelligent Robot Systems*, pp.2100–2106 (2013)
- [22] J. Engel, T. Schops, and D. Cremer: “Large-scale direct monocular SLAM”,

- Proc. 2014 European Conf. on Computer Vision, pp.834–849 (2014)
- [23] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard: “An evaluation of the RGB-D SLAM system”, Proc. 2012 Int. Conf. on Robotics and Automation, pp.1691–1696 (2012)
- [24] F. Tombari, S. Salti, and L.D. Stefano: “SHOT: Unique signatures of histograms for surface and texture description”, Proc. 2010 European Conf. on Computer Vision, pp.356–369 (2010)
- [25] A. Krizhevsky, I. Sutskever, and G. E Hinton: “Imagenet classification with deep convolutional neural networks”, Proc. 2012 Int. Conf. on Advances in Neural Information Processing Systems, pp.1097–1105 (2012)
- [26] S. Song and J. Xiao: “Deep sliding shapes for amodal 3D object detection in RGB-D images”, Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition, pp.809–816 (2016)
- [27] D. G. Lowe: “Distinctive image features from scale-invariant keypoints”, Int. J. of Computer Vision, Vol.60, No.2, pp.91–110 (2004)
- [28] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool: “Speeded-up robust features (SURF)”, Int. J. of Computer Vision and Image Understanding, Vol.110, No.3, pp.346–359 (2008)
- [29] M. Calonder, V. Lepetit, C. Strecha, and P. Fua: “BRIEF: Binary robust independent elementary features”, Proc. 2010 European Conf. on Computer Vision, pp.778–792 (2010)
- [30] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski: “ORB: An efficient alternative to SIFT or SURF”, Proc. 2011 IEEE Int. Conf. on Computer Vision, pp.2564–2571 (2011)
- [31] S. Leutenegger, C. Margarita, and R. Siegwart: “BRISK: Binary robust invariant scalable keypoints”, Proc. 2011 IEEE Int. Conf. on Computer Vision, pp.2548–2555 (2011)
- [32] M. Ambai and Y. Yoshida: “CARD: Compact and real-time descriptors”, Proc. 2011 IEEE Int. Conf. on Computer Vision, pp.97–104 (2011)
- [33] P. Alcantarilla, J. Nuevo, and A. Bartoli: “Fast explicit diffusion for accelerated

- features in nonlinear scale spaces”, Proc. 24th British Machine Vision Conf. (BMVC), pp.1–11 (2013)
- [34] 上瀧剛, 内村圭一: “スペクトル理論のパターンマッチングへの応用とその性能評価”, 電子情報通信学会論文誌 (D), Vol.J96-D, No.8, pp.1664–1674 (2013)
- [35] Y. Ke and R. Sukthankar: “PCA-SIFT: A more distinctive representation for local image descriptors”, Proc. 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.506–513 (2004)
- [36] J. Morel and G. Yu: “ASIFT: A new framework for fully affine invariant image comparison”, Society for Industrial and Applied Mathematics J. of Imaging Sciences, Vol.2, No.2, pp.438–469 (2009)
- [37] 藤吉弘亘: “Gradient ベースの特徴抽出 — SIFT と HOG — ”, 情報処理学会 研究報告, 2007-CVIM-160 (35) (2007)
- [38] T. Lindeberg: “Scale-space theory: A basic tool for analysing structures at different scales”, J. of Applied Statistics, Vol.21, No.2, pp.224–270 (1994)
- [39] D. G. Lowe: “Object recognition from local scale-invariant features”, Proc. 1999 IEEE Int. Conf. on Computer Vision, pp.1150–1157 (1999)
- [40] C. Harris and M. Stephens: “A combined corner and edge detector”, Proc. 4th Alvey Vision Conf., pp.147–151 (1988)
- [41] R. セジウィック: “領域探索”, アルゴリズム C 探索・文字列・計算幾何, (株) 近代科学社, pp.205–220 (1996)
- [42] M. Muja and D. G. Lowe: “Fast approximate nearest neighbors with automatic algorithm configuration”, Proc. 2009 Int. Conf. on Computer Vision Theory and Applications, pp. 331–340 (2009)
- [43] 鈴木将人, 深谷克己, 浅田宏起: “車両運動統合制御 (VDIM) の現状と展望”, トヨタ・テクニカル・レビュー, Vol.55, No.1, pp.48–53 (2006)
- [44] R. Tajima, D. Honda, and K. Suga: “Fast running experiments involving a humanoid robot”, Proc. 2009 IEEE Int. Conf. on Robotics and Automation, pp.1418–1423 (2009)
- [45] Ninebot: “Segway 製品ホームページ”, <http://www.segway.com/> [2016/08/01 Access]

- [46] R. Neul, U. Gomez, K. Kehr, W. Bauer, J. Classen, C. Doring, E. Esch, S. Gotz, J. Hauer, B. Kuhlmann, C. Lang, M. Veith, and R. Willig: “Micromachined angular rate sensors for automotive applications”, *IEEE Sensors*, Vol.7, No.2, pp.302–309 (2007)
- [47] M. Song, J. Park, Y. Park, and Y. S. Park: “Fusion filter for orientation estimation of biped robot”, *Motion and Vibration Control*, pp.285–294, Springer Netherlands (2009)
- [48] S. You and U. Neumann: “Fusion of vision and gyro tracking for robust augmented reality registration”, *Proc. 2001 IEEE Virtual Reality Conf.*, pp.71–78 (2001)
- [49] H. Rehbinder and B. K. Ghosh: “Pose estimation using line-based dynamic vision and inertial sensors”, *IEEE Trans. Automatic Control*, Vol.48, No.2, pp.186–199 (2003)
- [50] M. Veth and J. Raquet: “Fusing low-cost image and inertial sensors for passive navigation”, *J. Institute of Navigation*, Vol.54, No.1, pp.11–20 (2007)
- [51] D. Hong, H. Lee, H. Cho, Y. Park, and J. H. Kim: “Visual gyroscope integration of visual information with gyroscope for attitude measurement of mobile platform”, *Proc. 2008 Int. Conf. on Control, Automation and Systems*, pp.503–507 (2008)
- [52] C. Teuliere, L. Martin, C. Leroux, and E. Pissaloux: “Vision and inertial sensor fusion for 3D self-localization in unknown environment”, *Proc. 9th ESA Workshop on Advanced Space Technologies for Robotics and Automation*, Session 1.4.1.2 (2006)
- [53] G. Klein and T. Drummond: “Robust visual tracking for non-instrumented augmented reality”, *Proc. 2003 Int. Symp. on Mixed and Augmented Reality*, pp.113–122 (2003)
- [54] E. Foxlin and L. Naimark: “VIS-Tracker: A wearable vision-inertial self-tracker”, *Proc. 2003 IEEE Virtual Reality Conf.*, pp.199–206 (2003)
- [55] B. Jiang, U. Neumann, and S. You: “A robust hybrid tracking system for outdoor augmented reality”, *Proc. 2004 IEEE Virtual Reality Conf.*, pp.3–10

- (2004)
- [56] G. Bleser, and D. Stricker, “Advanced tracking through efficient image processing and visual-inertial sensor fusion”, J. Computer Graphics, Vol.33, No.1, pp.59–72 (2009)
 - [57] K. Satoh, S. Uchiyama, and H. Yamamoto, “A head tracking method using bird’s-eye view camera and gyroscope”, Proc. 2004 Int. Symp. on Mixed and Augmented Reality, pp.202–211 (2004)
 - [58] 加藤寛一郎： “最適制御入門，レギュレータとカルマン・フィルタ”，東京大学出版会（1987）
 - [59] 金井喜美雄，坪内孝司： “ビークル”，（社）計測自動制御学会，pp.142–173（2003）
 - [60] 片山 徹： “新版 応用カルマンフィルタ”，朝倉書店，第 5 章（2000）
 - [61] S. You, U. Neumann, and R. Azuma, “Hybrid inertial and vision tracking for augmented reality registration”, Proc. 1999 IEEE Virtual Reality Conf., pp.260–267 (1999)
 - [62] M. Kanbara, H. Fujii, H. Takemura, and N. Yokoya: “A stereo vision-based augmented reality system with an inertial sensor”, Proc. 2000 IEEE and ACM Int. Symp. on Augmented Reality 2000, pp.97–100 (2000)
 - [63] E. Shin and N. El-Sheimy: “Accuracy improvement of low cost INS/GPS for land applications”, Proc. 2002 National Technical Meeting of the Institute of Navigation, pp.146–157 (2002)
 - [64] E. Foxlin: “Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter”, Proc. 1996 Virtual Reality Annual Int. Symp., pp.184–195 (1996)
 - [65] M. A. Fischler and R. C. Bolles: “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”, Comm. ACM, Vol.24, No.6, pp.381–395, (1981)
 - [66] 徐剛，辻三郎： “エピポーラ幾何”，3次元ビジョン，共立出版（株），pp.61–77（2005）
 - [67] 前田龍太郎，小林健，池原毅，単学伝： “MEMS のはなし”，日刊工業新聞社（2008）
 - [68] 坂本修，新井昭文，山崎喜一郎，熊谷秀夫，塩沢龍雄，三村道彦，古田美直： “ジャ

- イロ活用技術入門”，多摩川精機（編），（株）工業調査会（2002）
- [69] 小堀訓成，出口大輔，高橋友和，井手一郎，村瀬洋：“カメラとジャイロセンサを組み合わせた移動ロボットの高精度な姿勢推定”，画像の認識・理解シンポジウム（MIRU）2009 論文集，pp.1757–1763（2009）
- [70] 小堀訓成，出口大輔，高橋友和，井手一郎，村瀬洋：“カメラとジャイロセンサを組み合わせた高精度な姿勢推定”，動的画像処理実利用化ワークショップ（DIA）2010 講演論文集，pp.1757–1763（2010）
- [71] J. Beis and D. Lowe: “Shape indexing using approximate nearest-neighbour search in high-dimensional spaces”, Proc. 1997 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.1000–1006（1997）
- [72] 山下信雄，福島雅夫：“Levenberg-Marquardt 法の局所収束性について（最適化の数理解科学）”，京都大学数理解析研究所講究録，No.1174，pp.161–168（2000）
- [73] 金谷健一：“3次元剛体運動の画像解析”，画像理解3次元認識の数理解，森北出版（株），pp.79–114（2005）
- [74] T. Lindeberg: “Scale-space theory: A basic tool for analysing structures at different scales”, J. of Applied Statics, Vol.21, No.2, pp.224–270（1994）
- [75] 小堀訓成，出口大輔，井手一郎，村瀬洋：“検出しやすくボケにロバストな符号化マーカの提案”，電子情報通信学会技術研究報告，IE2014-59（2014）
- [76] 松尾憲一：“スペクトラム拡散技術のすべて”，東京電機大学出版局（2005）
- [77] （株）デンソーウェーブ：“QRcode.com”，<http://www.denso-wave.com/qrcode/> [2015/09/01 Access]
- [78] 武田祐一，日浦慎作，佐藤宏介：“符号化開口を用いた Depth from Defocus とステレオ法の融合”，電子情報通信学会論文誌（D），Vol.J96-D，No.8，pp.1688–1700（2013）
- [79] S. Farsiu, M. Robinson, M. Elad, and P. Milanfar: “Fast and robust multiframe super resolution”, IEEE Trans. on Image Processing, Vol.13, No.10, pp.1327–1344（2004）
- [80] （株）シフト：“カメレオンコード”，<http://www.shift-2005.co.jp/> [2015/09/01 Access]
- [81] （株）カラーコードラボラトリーズ：“ColorZip”，<http://www.colorzip.co.jp/>

- [2015/09/01 Access]
- [82] 本田匠, 金子俊一: “色特徴を用いたロバストパターン照合技術と実応用”, OplusE, Vol.35, No.12, pp.1375–1380 (2013)
- [83] ビーコア (株): “カラービット”, <http://www.colorbit.jp/> [2015/09/01 Access]
- [84] K. Miyaoku, A. Tang, and S. Fel: “C-Band: A flexible ring tag system for camera-based user interface”, Virtual Reality, ICVR2007, 2nd Int. Conf. Procs., Lecture Notes in Computer Science, Vol.4563, pp.320–328, Springer (2007)
- [85] D. Parikh and G. Jancke: “Localization and segmentation of a 2D high capacity color barcode”, Proc. 2008 IEEE Workshop on Applications of Computer Vision, pp.1–6 (2008)
- [86] K. Tateno, I. Kitahara, and Y. Ohta: “A nested marker for augmented reality”, Virtual Reality, ICVR2007, 2nd Int. Conf. Procs., Lecture Notes in Computer Science, Vol.4563, pp.259–262, Springer (2007)
- [87] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar: “Bokode: Imperceptible visual tags for camera based interaction from a distance”, ACM Trans. on Graphics, Vol.28, No.3, Article no.98 (2009)
- [88] M. Toyoura, H. Aruga, M. Turk, and X. Mao: “Detecting markers in blurred and defocused images”, Proc. 2013 Int. Conf. on Cyberworlds, pp.183–190 (2013)
- [89] 上田智章: “実験研究 符号パターンをカメラで撮影して位置・角度を測る”, Interface, CQ 出版社, pp.14–21 (2013)
- [90] T. Hasegawa, M. Ambai, K. Ishikawa, G. koutaki, Y. Yamauchi, T. Yamashita, and H. Fujiyoshi: “Multiple-hypothesis affine region estimation with anisotropic LoG filters”, Proc. 2015 IEEE Int. Conf. on Computer Vision, pp.585–593 (2015)
- [91] T. Nakai, K. Kise, and M. Iwamura: “Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval”, 7th Int. Workshop DAS, Lecture Notes in Computer Science, Vol.3872, pp.541–552 (2006)
- [92] H. Uchiyama and H. Saito: “Random dot markers”, Proc. 2011 IEEE Virtual Reality Conf., pp.35–38 (2011)
- [93] K. Yi, Y. Verdie, P. Fua, and V. Lepetit: “Learning to assign orientations to feature points”, Proc. 2016 IEEE Conf. on Computer Vision and Pattern

- Recognition, pp.107–116 (2016)
- [94] C. Tao, Z. Zhe, S. Ariel, H. Shi-Min, and C. Daniel: “3-sweep: Extracting editable objects from a single photo”, *ACM Trans. on Graphics*, Vol.32, No.6, pp.195–204 (2013)
- [95] S. Shuran and X. Jianxiong: “Sliding shapes for 3D object detection in depth images”, *Proc. 2014 European Conf. on Computer Vision*, pp.634–651 (2014)
- [96] R. Schnabel, R. Wahl, and R. Klein: “Efficient RANSAC for point-cloud shape detection”, *Proc. 2007 Computer Graphics Forum*, Vol.26, No.2, pp.214–226 (2007)
- [97] Utrecht University: “The Computational Geometry Algorithms Library”, <http://www.cgal.org/> [2016/08/01 Access]
- [98] F. Tombari, S. Salti, and L. Stefano: “Performance evaluation of 3D keypoint detectors”, *Int. J. of Computer Vision*, Vol.102, No.1, pp.198–220 (2013)
- [99] A. Zaharescu, E. Boyer, and R. Horaud: “Keypoints and local descriptors of scalar functions on 2D manifolds”, *Int. J. of Computer Vision*, Vol.100, No.1, pp.78–98 (2012)
- [100] P. Marion, R. Kwitt, B. Davis, and M. Gschwandtner: “PCL and ParaView — Connecting the dots”, *Proc. 2012 IEEE Point Cloud Processing*, pp.80–85 (2012)

本人主筆の研究業績

学術雑誌論文：2件

1. 小堀訓成, 出口大輔, 井手一郎, 村瀬洋: “作業用ロボットのための符号化マーカの提案 ～多様な配置やボケ環境下でも検出し易い符号化マーカ～”, 電気学会論文誌, Vol.136-C, No.9, pp.1367–1375 (2016-9)
2. 小堀訓成, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: “MEMS ジャイロセンサと単眼カメラを利用した高精度で頑健な姿勢推定”, 計測自動制御学会論文集, Vol.47, No.10, pp.442–449 (2011-10)

査読付き国際会議論文：2件

1. Norimasa Kobori, Daisuke Deguchi, Ichiro Ide, and Hiroshi Murase: “Proposal of a spectral random dots marker using local feature for posture estimation”, IEEE Virtual Reality Conf., (2017-3) ※投稿中
2. Norimasa Kobori, Daisuke Deguchi, Tomokazu Takahashi, Ichiro Ide, and Hiroshi Murase: “Accurate and robust attitude estimation of a moving vehicle using MEMS gyroscopes and a monocular camera”, Proc. 2010 Workshop on Picture Coding and Image Processing, WP3-11 (2010-7)

国内会議口頭発表：4件

1. 小堀訓成, 出口大輔, 井手一郎, 村瀬洋: “姿勢推定のための局所特徴量を用いた符号化マーカの提案”, 電子情報通信学会画像工学会, IE2016-65 (福岡大; 2016-10)
2. 小堀訓成, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: “検出しやすくボケにロバストな符号化マーカの提案”, 電子情報通信学会画像工学会, IE2014-59 (大阪大; 2014-12)
3. 小堀訓成, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: “カメラとジャイロセンサを組み合わせた高精度な姿勢推定”, DIA2010 動的画像処理実利用化ワークショップ, O2-2 (甲府市ぴゅあ総合; 2010-3)
4. 小堀訓成, 出口大輔, 高橋友和, 井手一郎, 村瀬洋: “カメラとジャイロセンサを組み合わせた移動ロボットの高精度な姿勢推定”, 第12回 画像の認識・理解シンポジウム, IS3-56 (くにびきメッセ; 2009-7)

本人主筆のその他の研究業績

学術雑誌論文：1 件

1. 小堀訓成, 鈴木健嗣, ピトヨ ハルトノ, 橋本周司: “尤度情報に基づく温度分布を用いた強化学習法”, 人工知能学会論文誌, Vol.20, No.4D, pp.297–305 (2005-4)

査読付き国際会議論文：1 件

1. Norimasa Kobori, Kenji Suzuki, Hartono Pitoyo, and Shuji Hashimoto: “Learning to control a joint driven double inverted pendulum using Nested Actor Critic Algorithm”, Proc. 9th Int. Conf. on Neural Information Processing, pp.2610–2614 (2002-12)

国内会議口頭発表：1 件

1. 小堀訓成, 橋本周司: “尤度情報を利用した強化学習法”, 情報処理学会第 66 回全国大会, 2X-6 (慶応大 ; 2004-3)

受賞：1 件

1. 情報処理学会第 66 回全国大会, 学生奨励賞 (2004 年 3 月)