

NAIST-IS-MT1551115

課題研究

Visual-SLAM における Pose Graph Optimization 及びバンドル調整が推定精度に与える影響に関する調査

山本 雄貴

2017 年 3 月 16 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

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

山本 雄貴

審査委員：

加藤 博一 教授	（主指導教員）
向川 康博 教授	（副指導教員）
Christian Sandor 准教授	（副指導教員）
武富 貴史 助教	（副指導教員）

Visual-SLAM における Pose Graph Optimization 及びバンドル調整が推定精度に与える影響に関する調査*

山本 雄貴

内容梗概

カメラから取得された画像から自己位置姿勢推定と環境地図作成を同時に行う Visual-SLAM(Simultaneous Localization and Mapping) は、コンピュータビジョン、ロボティクス、拡張現実感などの分野において研究が行われている。中でも単眼カメラを用いた Visual-SLAM は、これまでに様々な手法が提案されている。特に、ORB-SLAM[1] は、Bag-of-Words による Loop Detection と 7 自由度の Pose Graph Optimization を Visual-SLAM に組み込むことによって、大規模環境においても高精度な推定を実現している。しかし、Pose Graph Optimization および後段で実行されるバンドル調整が Visual-SLAM の最終的な推定精度に与える影響については、検証が十分に行われていない。そこで、本研究では、Pose Graph Optimization およびバンドル調整においてキーフレーム全体を推定対象とした場合と一部のキーフレームのみを推定対象とした場合の推定誤差について検証を行った。検証では、シミュレーション環境および実環境における精度評価実験を行った。その結果、カメラの移動距離が増加することによる距離誤差の増加及びカメラがループ軌道を描かない場合において、Pose Graph Optimization 及びバンドル調整は SLAM の精度に対し良い影響を与えることが確認された。また、Pose Graph Optimization 及びバンドル調整の回数の増加が蓄積された距離誤差及び姿勢誤差に対して良い影響を与えることが確認された。

キーワード

Visual-SLAM, ORB-SLAM, Pose Graph Optimization, バンドル調整

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 課題研究, NAIST-IS-MT1551115, 2017 年 3 月 16 日.

Investigation of the Influence of Pose Graph Optimization and Bundle Adjustment on Estimation Accuracy in Visual-SLAM*

Yuki Yamamoto

Abstract

Visual-SLAM(Simultaneous Localization and Mapping) which simultaneously performs self position / posture estimation and environmental map creation from images obtained from cameras is being widely studied in the fields of computer vision, robotics and augmented reality. In these fields, various Visual-SLAM methods that employ single monocular camera are highly popular. Particularly, ORB-SLAM provides high estimation accuracy even in large-scale environments through Loop Detection by Bag-of-Words and Pose Graph Optimization with seven degrees of freedom. However, the effects of Pose Graph Optimization and the Bundle Adjustment on final estimation accuracy lacks sufficient amount of verification. Therefore, in this study, we investigated the estimation error when Pose Graph Optimization and Bundle Adjustment set the entire key frame as the estimation target and the estimation error when only some key frames are targeted for estimation. In our verification, we conducted accuracy evaluations in both simulated and real environments. As a result, we confirmed that Pose Graph Optimization and Bundle Adjustment have positive influence on accuracy of SLAM; when distance error increases due to increase in camera movement where the camera's trajectory is loop-free. We also confirmed that Pose Graph Optimization and repetitive Bundle Adjustments have positive effect on reducing the accumulated error.

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1551115, March 16, 2017.

Keywords:

Visual-SLAM, ORB-SLAM, Pose Graph Optimization, bundle adjustment

目次

図目次	vi
表目次	vii
第 1 章 はじめに	1
第 2 章 ORB-SLAM の概要	2
2.1 Tracking スレッド	2
2.1.1 特徴点の抽出	3
2.1.2 カメラ位置姿勢の推定	3
2.1.3 キーフレームの決定	4
2.2 Local Mapping スレッド	5
2.2.1 マップの初期化	5
2.2.2 マップへのキーフレームの挿入	6
2.2.3 新たなマップポイントの作成	6
2.2.4 ローカルバンドル調整	7
2.3 Loop Closer スレッド	8
2.3.1 ループの検出	9
2.3.2 ループの結合	10
2.3.3 Pose Graph Optimization	12
2.3.4 バンドル調整	14
2.4 最適化の組み合わせ	14
第 3 章 評価実験	17
3.1 ループクローズに対する変更	17
3.2 検証方法	18
3.3 シミュレータ環境	20
3.4 実験結果	22
3.4.1 カメラシーケンス 1	24

3.4.2	カメラシーケンス 2	29
3.4.3	カメラシーケンス 3	34
3.4.4	実環境	39
3.5	考察	42
第 4 章	まとめ	45
	謝辞	46
	参考文献	47

図目次

2.1	ORB-SLAM の概要	2
2.2	Loop Closer スレッドを構成する主な 2 つのタスク	9
2.3	ループクローズが行われていない状態	10
2.4	k_i が補正された状態	11
2.5	ループを構成するキーフレームが線形に補正された状態	12
2.6	Raul Mur-Artal らの示す Covisibility Graph(引用 : [1])	13
2.7	Raul Mur-Artal らの示す Essential Graph(引用 : [1])	14
2.8	2 つのループを持つカメラの動き	15
3.1	補正された K_i の位置姿勢 $T_i w$	18
3.2	カメラ位置姿勢の真値	20
3.3	カメラシーケンス 1 のカメラの動き	22
3.4	カメラシーケンス 2 のカメラの動き	22
3.5	カメラシーケンス 3 のカメラの動き	22
3.6	シーケンス 1 での各キーフレーム毎の距離誤差 (両固定)	27
3.7	シーケンス 1 での各キーフレーム毎の距離誤差 (バンドル固定)	27
3.8	シーケンス 1 での各キーフレーム毎の距離誤差 (Pose Graph 固定)	28
3.9	シーケンス 1 での各キーフレーム毎の距離誤差 (両非固定)	28
3.10	シーケンス 2 での全キーフレームの距離誤差 (両固定)	32
3.11	シーケンス 2 での全キーフレームの距離誤差 (バンドル固定)	32
3.12	シーケンス 2 での全キーフレームの距離誤差 (Pose Graph 固定)	33
3.13	シーケンス 2 での全キーフレームの距離誤差 (両非固定)	33
3.14	シーケンス 3 での全キーフレームの距離誤差 (両固定)	37
3.15	シーケンス 3 での全キーフレームの距離誤差 (バンドル固定)	37
3.16	シーケンス 3 での全キーフレームの距離誤差 (Pose Graph 固定)	38
3.17	シーケンス 3 での全キーフレームの距離誤差 (両非固定)	38
3.18	推定されたカメラパス	39
3.19	測定環境	40

表目次

3.1	カメラシーケンス 1 での最終誤差	23
3.2	カメラシーケンス 2 での最終誤差	23
3.3	カメラシーケンス 3 での最終誤差	24
3.4	実環境での最終誤差	24
3.5	カメラシーケンス 1 での距離誤差	25
3.6	カメラシーケンス 1 での「オリジナル」の距離誤差	25
3.7	カメラシーケンス 1 での姿勢誤差	26
3.8	カメラシーケンス 1 での「オリジナル」の姿勢誤差	26
3.9	カメラシーケンス 2 での距離誤差	30
3.10	カメラシーケンス 2 での「オリジナル」の距離誤差	30
3.11	カメラシーケンス 2 での姿勢誤差	31
3.12	カメラシーケンス 2 での「オリジナル」の姿勢誤差	31
3.13	カメラシーケンス 3 での距離誤差	35
3.14	カメラシーケンス 3 での「オリジナル」の距離誤差	35
3.15	カメラシーケンス 3 での姿勢誤差	36
3.16	カメラシーケンス 3 での「オリジナル」の姿勢誤差	36
3.17	実環境での距離誤差	41
3.18	実環境での姿勢誤差	42

第 1 章 はじめに

単眼カメラを用いて自己位置姿勢推定および環境地図の作成を行う Visual-SLAM(Simultaneous Localization and Mapping) は，手軽に拡張現実感環境を構築することが出来ることから，コンピュータビジョン，ロボティクス分野のみでなく拡張現実感分野においても広く研究がなされている．中でも，2007 年に発表された PTAM[2] は自己位置姿勢推定と環境地図作成の処理を異なるスレッドで実行することによって，推定処理において多くの特徴点を利用することを可能とし高精度かつ安定的な推定を実現した．しかし，広域な環境においては，蓄積した誤差を解消することが難しいという問題があった．これに対して，2013 年に Bag of Words を用いた位置認識と，Pose Graph Optimization による蓄積誤差の解消を可能とした ORB-SLAM が発表された．ORB-SLAM では，推定の途中で再度同じ位置に戻ってきた際に，Pose Graph Optimization およびバンドル調整を実行することによって，大規模な環境においても蓄積した誤差を解消することができる．しかし，Pose Graph Optimization およびバンドル調整において，最適化対象のキーフレームの取り扱いに関する詳細な検討はこれまでに行われていない．そこで，本研究課題では，Pose Graph Optimization およびバンドル調整において，最適化対象となるキーフレームを限定することが最終的な推定精度に与える影響について検証を行った．

第 2 章 ORB-SLAM の概要

ORB-SLAM は図 2.1 の様に「Tracking」「Local Mapping」「Loop Closer」の 3 つのスレッドで構成されており，これらが並行で動作することで SLAM を行っている．そこで，本章では第 2.1 節で Tracking スレッドの処理を，第 2.2 節で Local Mapping スレッドの処理を，第 2.3 節で Loop Closer スレッドの処理を説明し，その後，最適化戦略について説明する．

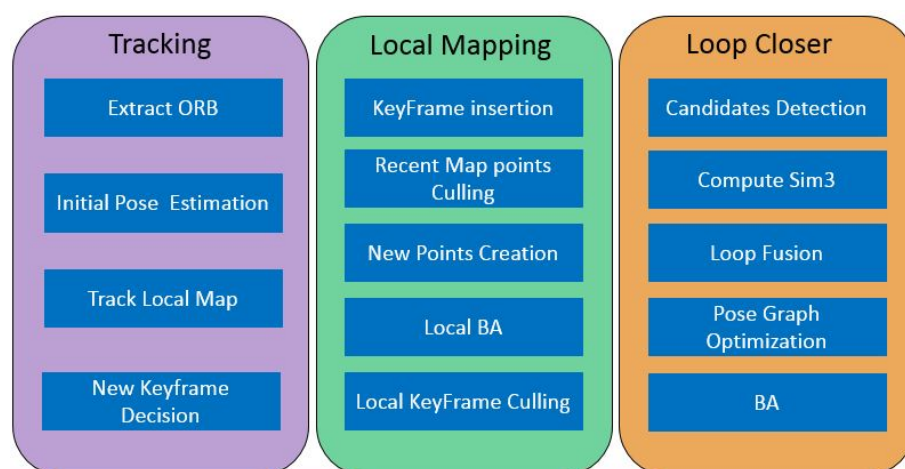


図 2.1 ORB-SLAM の概要

2.1 Tracking スレッド

本節では Tracking スレッド (図 2.1 : Tracking) の処理を説明する．Tracking スレッドではフレームごとのカメラの位置姿勢の推定，キーフレームの決定を行う．Tracking スレッドでの処理は以下のような手順で行われる．

1. 特徴点の抽出
2. カメラ位置姿勢の推定
3. キーフレーム (小節 2.1.2 で説明する) の決定

2.1.1 特徴点の抽出

単眼 SLAM において、カメラ位置は一般的に異なるフレーム間の特徴点の対応から求められる。そのため、Tracking スレッドではまずカメラ位置を推定するために、フレーム中から特徴点を抽出しなければならない。この時、ORB-SLAM では処理の高速化の為にトラッキング、マッピング、relocalization、ループクローズ等の処理に全て同じ特徴が使用されている為、抽出速度の早い特徴が求められている。また、精度の高い位置姿勢推定の為にスケール不変性と回転不変性も必要である。これらの条件を満たす特徴として、ORB-SLAM では ORB 特徴 [3] を使用している。ORB 特徴は FAST[4] と BRIEF[5] にスケール不変性と回転不変性を組み込んだアルゴリズムであり、一般的な画像特徴である SIFT[6] や SURF[7], A-KAZE[8] と比べて、1 画像に対する特徴の抽出速度が高速である。

2.1.2 カメラ位置姿勢の推定

ORB 特徴が求められると、次に現フレームのカメラ位置姿勢を求める。まず初めに、現フレームの前のフレームから、等速直線運動モデルを用いて大体のカメラ位置姿勢を推定する。大体のカメラ位置姿勢が推定されると、そのフレームをローカルマップに投影する。ローカルマップには現在のフレームとマップポイントを共有するキーフレーム K_1 と、 K_1 とマップポイントを 15 点以上共有するキーフレーム K_2 、そして現在のキーフレームとほとんどのマップポイントを共有する参照キーフレーム $K_{ref} \in K_1$ が含まれている。この時、マップポイントとは ORB 特徴のワールド座標系での座標値であり、以下の情報を保持している。

- ワールド座標系での 3 次元座標値 $\mathbf{X}_{w,i}$
- 視線方向 n_i : その点を観測しているキーフレームの光学中心と点を結ぶ光線
- ORB 特徴へのファイルディスクリプタ $\mathbf{D}_i$
- その点が観測される最大の距離 d_{max} と最小距離 d_{min}

また、キーフレームとは現在世界の視覚情報を保持するフレームのことであり、キーフレームには以下の情報が格納されている。

- 世界座標系からカメラ座標系への変換行列である，位置姿勢 T_{iw}
- 焦点距離と主要点を含む，カメラの内部行列
- フレーム内で抽出されたすべての ORB 特徴
- キーポイント

ローカルマップに現フレームを投影すると， K_1 ， K_2 で観測されたマップポイントは現フレームで以下のように検索される．

- 現フレームのマップポイントが投影されている点 x を計算し，画像外の場合は破棄する．
- 現在の視線 v とマップポイントの視線 n の角度を計算し， $V \cdot n < (60^\circ)$ の場合は破棄する．
- マップポイントからカメラ中心までの距離 d を計算し， $d \notin [d_{max}, d_{min}]$ の場合は破棄する．
- フレーム内のスケールを d/d_{min} で計算する．
- フレームの対応付けが行われていない ORB 特徴を，予測されたスケールを用い， x の周辺でディスクリプタ D と比較することで，最良の対応付を行う．

以上により，現フレームで観測されたマップポイントを用いて，カメラ位置姿勢を最適化することにより，現フレームの位置姿勢を求める．

2.1.3 キーフレームの決定

ORB 特徴の抽出に成功し，カメラ位置姿勢が推定されると，その位置姿勢を推定したフレームがキーフレームであるかどうかの判定を行う．ORB-SLAM では，一度キーフレームと決定されると不変となる PTAM と違い，マップポイントが少なくとも 3 つのキーフレームの同じスケール，またはより小さなスケールで 90 % 以上共通している場合，削除することが出来る．フレームは以下の条件をみたすことでキーフレームとして決定される．

- relocalization から 20 フレームが経過している
- 少なくとも 50 点以上の対応がある

- 現在のフレームと最も多くのマップポイントを共通するキーフレーム K_{ref} と比べて 90 %未満のマップポイントを追跡する
- キーフレームが最後にローカルマップに挿入されてから少なくとも 10 フレーム経過している, または 20 フレーム経過しローカルバンドル調整 (小節 2.2.4 で説明する) の終了シグナルをローカルマップに送信している

フレームがキーフレームと決定されると Local Mapping スレッドに移行する.

2.2 Local Mapping スレッド

本節では Local Mapping スレッド (図 2.1 : Local Mapping) の処理を説明する. Local Mapping スレッドでは Tracking スレッドでキーフレームと決定されたフレーム K_i をマップに挿入し, マップの最適化と更新を行う. Tracking スレッドでの処理は以下のような手順で行われる.

1. マップの初期化
2. マップへのキーフレームの挿入
3. 新たなマップポイントの作成
4. ローカルバンドル調整

2.2.1 マップの初期化

十分な ORB 特徴を抽出すると, 初期カメラ位置姿勢の推定とマップポイントの計算を行う. まず, 初期カメラ位置姿勢の推定は以下の手順で行われる.

1. 現在のフレーム F_c と参照フレーム F_r 内の ORB 特徴の一致 $x_c \leftrightarrow x_r$ を検索する. 十分な一致が見つからない場合は F_r をリセットする.
2. ホモグラフィ H_{cr} と基礎行列 F_{cr} を並列に求める. これらの値は [9] で説明されている様に, 8 点法と DLT アルゴリズムを使用して計算される.
3. シーンが「平面」「平面に近い」「視差が低い」場合は H_{cr} を, シーンが「平面でない」「十分な視差がある」場合は F_{cr} を使用して位置姿勢を推定しな

ければ，間違った位置姿勢が推定される可能性がある．その為，以下のロバストヒューリスティックを計算することにより， H_{cr} と F_{cr} を切り替えている．

$$R_H = \frac{S_H}{S_H + S_F} \quad (2.2.1)$$

$R_H > 0.45$ の場合は H_{cr} を用い，それ以外の場合は F_{cr} を用いる．

こうして求められた初期のカメラ位置姿勢を追跡することで，各フレームのカメラ位置姿勢を求めている．マップポイントは，上記の手順の際に対応付けられた異なるキーフレームの ORB 特徴を三角測量することで求めることが出来る．

2.2.2 マップへのキーフレームの挿入

マップの初期化に成功すると，Local Mapping スレッドでは，初めに，Tracking スレッドによってキーフレームと決定されたフレーム K_i をマップに挿入する．ORB-SLAM ではマップの表現に Covisibility Graph（重み付き無向グラフであり，キーフレームをノード，マップポイントを重みとし，異なるキーフレーム間に共通のマップポイントが θ 個存在した場合キーフレーム間にエッジを張ることが出来る．Strasdat ら [10] はこの共通の θ を 15～30 点としている．）を使用している為， K_i はノードとして挿入され， K_i と共通のマップポイントが θ 個存在するキーフレームとの間にエッジが張られる．また，この時 D. Galvez-Lopez ら [11] の方法と同じように，キーフレームを Bags of Words に変換し，データベースに保存しておく．

2.2.3 新たなマップポイントの作成

新しいキーフレームがマップに挿入されると，新しくマップポイントが計算される．マップポイントは異なるキーフレーム間で ORB 特徴を三角測量することにより作成される．PTAM では新しく挿入されたキーフレームに最も近いキーフレームのみを使用してマップポイントを計算していたが，ORB-SLAM では Covisibility Graph 上で新しいキーフレームと大部分のマップポイントを共有している N 個の

キーフレームを使用してマップポイントを計算している．この時，Vocabulary tree の第2ノードを参照することで無駄な計算を省き，計算コストを抑えている．マップポイントの作成に成功した場合，そのマップポイントが正しいかどうかの判定が行われる．正しいマップポイントと判断されるには以下の2つの条件を満たす必要がある．

- そのマップポイントが投影されていると予測されているフレームの内，25%以上のフレームで実際に追跡している．
- 複数のキーフレームがマップポイント作成後に挿入された場合，少なくとも3つのキーフレームでそのマップポイントが表示されている．

これらの条件を満たした場合にのみ，正しいマップポイントと判断されマップに反映される．

2.2.4 ローカルバンドル調整

新しいキーフレームが挿入され，正しいマップポイントが計算されると，ローカルバンドル調整が行われる．バンドル調整とは，推定されたカメラ位置とそのフレームのマップポイントに誤差が含まれていた場合に生まれる，画像上の特徴点の誤差（再投影誤差）が最小になるように最適化を行うことである．ここでキーフレーム i からマップポイント j を観測した際の誤差項 $e_{i,j}$ は次式のように表すことが出来る．

$$e_{i,j} = x_{i,j} - \pi_i(\mathbf{T}_{iw}, \mathbf{X}_{w,j}) \quad (2.2.2)$$

$x_{i,j} \in \mathbb{R}^2$ はキーフレーム i 上でのマップポイント j の画像座標， $\mathbf{T}_{iw} \in SE(3)$ はキーフレーム i の位置姿勢， $\mathbf{X}_{w,i} \in \mathbb{R}^3$ はマップポイントの3次元位置を表している (w はワールド座標系基準)．また， π_i は投影関数であり，次式で表される．

$$\pi_i(\mathbf{T}_{iw}, \mathbf{X}_{w,i}) = \begin{bmatrix} f_{i,u} \frac{x_{i,j}}{z_{i,j}} + c_{i,u} \\ f_{i,v} \frac{y_{i,j}}{z_{i,j}} + c_{i,v} \end{bmatrix} \quad (2.2.3)$$

$$\begin{bmatrix} x_{i,j} & y_{i,j} & z_{i,j} \end{bmatrix}^T = \mathbf{R}_{i,w} \mathbf{X}_{w,j} + \mathbf{t}_{i,w} \quad (2.2.4)$$

$\mathbf{R}_{iw} \in SO(3)$ と $\mathbf{t}_{iw} \in \mathbb{R}^3$ はそれぞれ $\mathbf{T}_{iw}$ の回転と平行移動を表しており、 $(f_{i,u}, f_{i,v})$ と $(c_{i,u}, c_{i,v})$ はそれぞれカメラ i の焦点距離と主点を表している。この時、最適化を行うべきコスト関数は次式で表すことが出来る。

$$C = \sum_{i,j} \rho_h(e_{i,j}^T \mathbf{\Omega}_{i,j}^{-1} e_{i,j}) \quad (2.2.5)$$

ρ_h は Huber robust コスト関数 [12], $\mathbf{\Omega}_{i,j}$ は共分散行列を表している。バンドル調整ではマップに保存されている全てのキーフレームとマップポイントを最適化の対象として扱うが、ローカルバンドル調整では K_i とマップポイントを θ 個共有しているキーフレーム及び K_i 自体のみ最適化され、その他のキーフレームは固定されたままである。

2.3 Loop Closer スレッド

本節では Loop Closer スレッド（図 2.1： Loop Closer）の処理を説明する。Loop Closer スレッドでは Local Mapping スレッドで処理が終了したキーフレーム K_i を受け取り、ループの検出とループの補正を行うことで、グローバルなマップの最適化を行う。Loop Closer スレッドは図 2.2 の様に、「ループの検出」と「ループの補正」という大きな 2 つのタスクに別れており、それぞれのタスクの主な動作は以下のような手順で行われる。

1. ループの検出
2. ループの結合
3. Pose Graph Optimization
4. バンドル調整

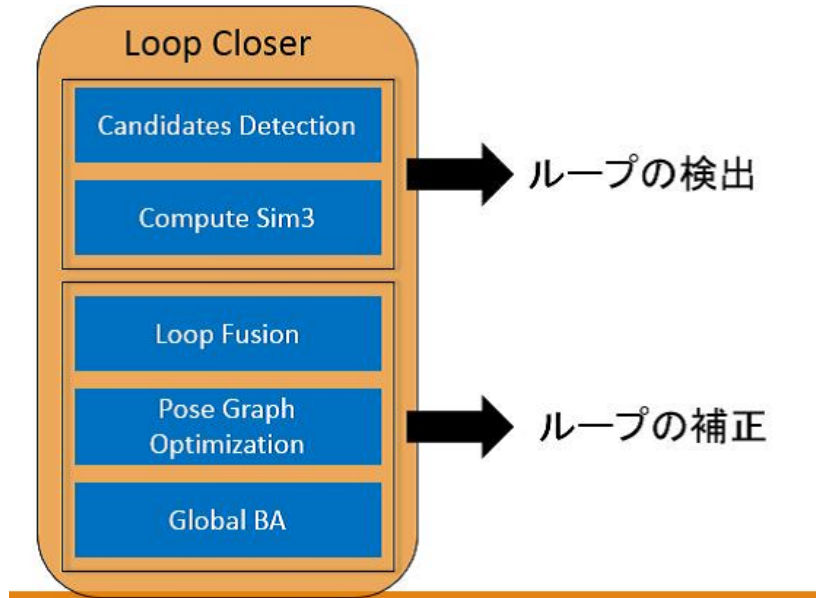


図 2.2 Loop Closer スレッドを構成する主な 2 つのタスク

2.3.1 ループの検出

ループの検出には Bag of Words を使用する．まず初めに K_i の Visual Words と Covisibility Graph 内で K_i にエッジが張られているキーフレームの類似度を計算し，最小スコアを S_{min} として保持する．次にデータベースを参照し，同じように類似度を計算する．この時スコアが S_{min} 以下のものと， K_i にエッジが張られているキーフレームを結果から破棄する．そして，それ以外のものをループ候補とする．ループのロバスト性を高めるために，ループ候補は 3 つ以上検出される必要がある．ループ候補が検出されると，ループ候補となる複数のキーフレームの中から 1 つのループ候補を決めなければならない．ORB-SLAM ではまず，現在のキーフレーム K_i と各ループ候補のキーフレームの特徴点对応を求める．次にこれらの対応を元に Horn[13] らの方法を用いて相似変換行列 $S_i l$ を計算する．各ループ候補で相似変換行列を計算出来たら，これらに RANSAC アルゴリズムを適用する．これにより，最も類似度の高いキーフレーム k_l がループ候補の最有力として決定される．ループ候補として 1 つのキーフレームが決定されると，次にループの補正処

理を行う。

2.3.2 ループの結合

K_l と K_i の間にループが検出され、キーフレームが図 2.3 の様に保存されていたとする（青い四角形がキーフレーム、黒線がカメラの移動を表している）。この時、ループ補正の第 1 段階としてキーフレームを線形に補正する。まず、相似変換行列 $S_i l$ を使用して K_i のワールド座標系での位置姿勢 T_{iw} を図 2.4 の様に補正する。次に、補正前のキーフレーム同士の相対関係を使用して、図 2.5 の様にループを構成するキーフレームに補正を適用する（点線は修正されたカメラ移動を示している）。この時、補正前のキーフレームで観測されたマップポイントは、補正されたキーフレームに投影される。これによりループを構成するキーフレームが補正される。最後に、ループ補正の最終段階として、線形に補正したキーフレームに対し Pose Graph Optimization を行うことでキーフレームを非線型に補正する。

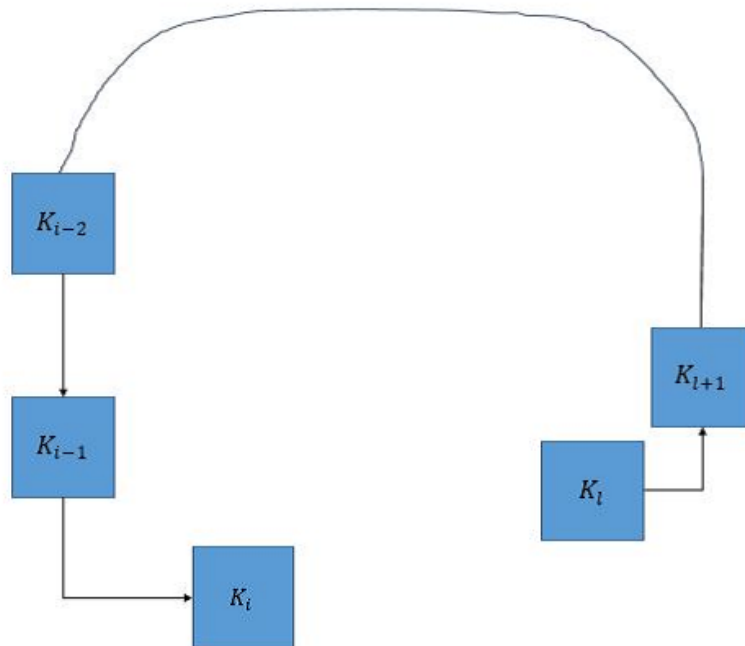


図 2.3 ループクローズが行われていない状態

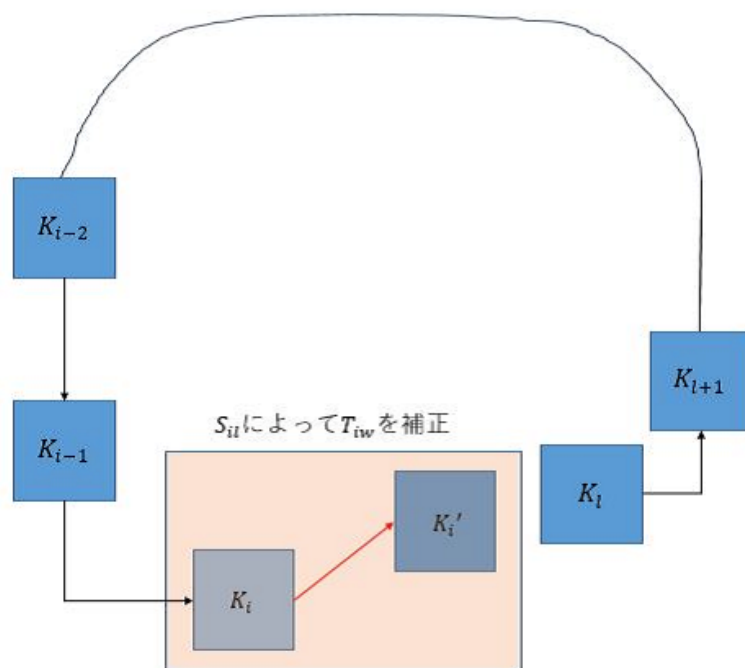


図 2.4 k_i が補正された状態

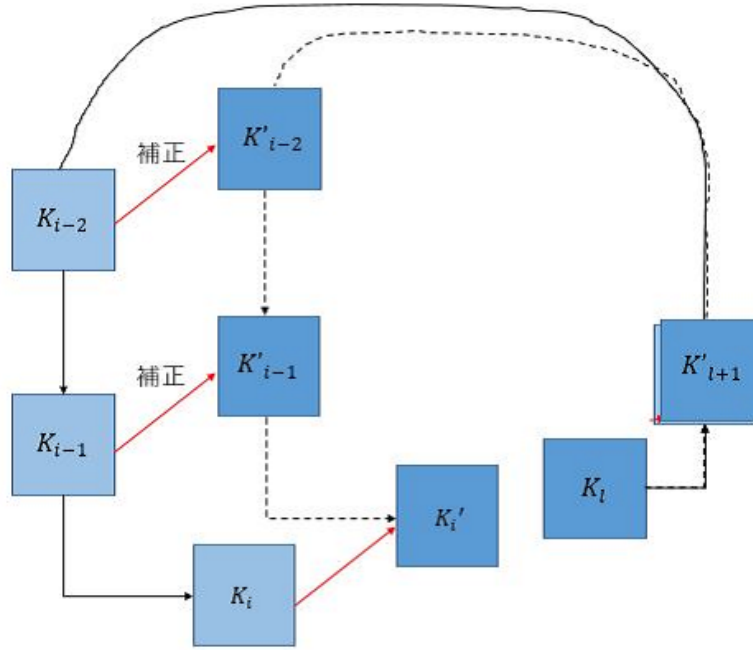


図 2.5 ループを構成するキーフレームが線形に補正された状態

2.3.3 Pose Graph Optimization

ループ補正の最終段階として、先程線形に補正したキーフレームに対し、更に Pose Graph Optimization を行う事でキーフレームを非線型に補正する．ORB-SLAM では Essential Graph を使用したキーフレームの位置姿勢の最適化を行っている．Essential Graph は高密度である可能性がある Covisibility Graph の内、グラフを構成する必要最小限のエッジと全てのノードを含んだグラフであり、スパニングツリーを基にして作成される．Raul Mur-Artal らの示す Covisibility Graph 図 2.6 と Essential Graph 図 2.7 を比べると、かなりのエッジ (図 2.6, 図 2.7: 緑線) が減少していることが確認できる．

バイナリエッジのポーズグラフが与えられると、エッジの誤差 $e_{i,j}$ は、

$$e_{i,j} = \log_{sim(3)}(\mathbf{S}_{ij}\mathbf{S}_{jw}\mathbf{S}_{iw}^{-1}) \quad (2.3.1)$$

によって与えられる． $\mathbf{S}_{i,j}$ はキーフレーム間の相似変換であり、スケールを 1 と設定している．ループクローズのエッジの場合， $\mathbf{S}_{i,j}$ は Horn ら [13] の方法により求められ， $\log_{sim(3)}$ [14] により接空間に変換されるため， $e_{i,j}$ は $\mathbb{R}^7$ となる．グラフ最適化の目標は，コスト関数を最小化する事で $\text{Sim}(3)$ の位置姿勢を最適化することである．コスト関数 C は次のようになる．

$$C = \sum_{i,j} (e_{i,j}^T \mathbf{\Lambda}_{i,j} e_{i,j}) \quad (2.3.2)$$

ここで， $\mathbf{\Lambda}_{i,j}$ は [14] と同じように，エッジの情報行列を表している．

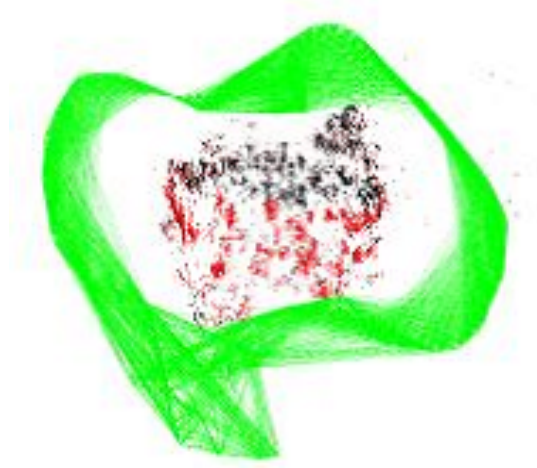


図 2.6 Raul Mur-Artal らの示す Covisibility Graph(引用 : [1])



図 2.7 Raul Mur-Artal らの示す Essential Graph(引用 : [1])

2.3.4 バンドル調整

ポーズグラフ最適化が終了すると、バンドル調整が行われる。この処理はローカルバンドル調整と同じ処理を行っているが、ローカルバンドル調整と違い Covisibility Graph 内に保存されている全てのマップポイントとキーフレームの位置姿勢に対して処理が実行される。

2.4 最適化の組み合わせ

ORB-SLAM では、LoopCloser スレッド内の Pose Graph Optimization 及びバンドル調整により一度位置姿勢が補正されたキーフレームに対し、再度 Pose Graph Optimization 及びバンドル調整を適用した場合、キーフレームの位置姿勢が再び補正される。具体的には、カメラが図 2.8 の地点 A から $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A \rightarrow B$ （紫線→緑線→赤線）の順に移動したとする。この時、地点 $A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$ の順に移動した時点でループが存在する為、紫線及び緑線上を移動する際に保存されたキーフレームの位置姿勢が、Pose Graph Optimization とバンドル調整により補正される。次に地点 B に移動した際に、地点 $B \rightarrow C \rightarrow D \rightarrow A \rightarrow B$ の

カメラ移動に対し、同様にループの存在が探索されたとすると、ORB-SLAM では緑線及び赤線を移動する際に保存されたキーフレームの位置姿勢が、Pose Graph Optimization 及びバンドル調整により補正される。この時、緑線上のキーフレームは、Pose Graph Optimization 及びバンドル調整により 2 回補正が行われていることになる。

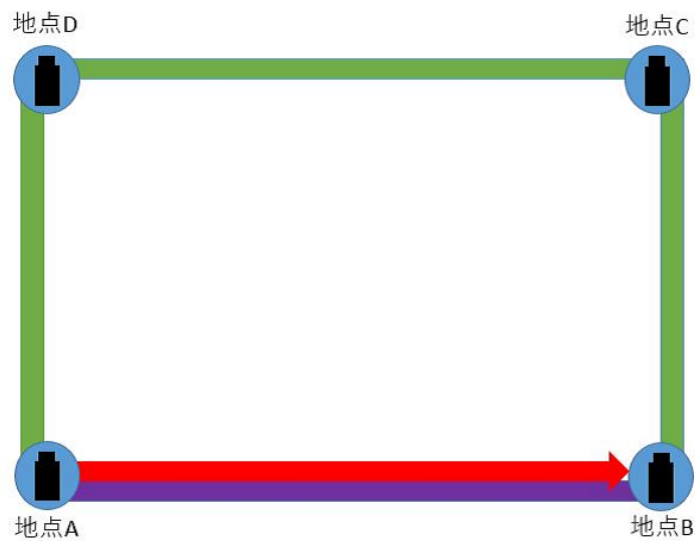


図 2.8 2 つのループを持つカメラの動き

その為、Pose Graph Optimization 及びバンドル調整が SLAM の精度に与える影響を評価するためには、「Pose Graph Optimization 及びバンドル調整により補正されたキーフレームに対しては再度補正を行わない（紫線上のキーフレームに対してのみ、補正処理を行う）」等、以下の 4 種類の組み合わせにおいて最適化を実行する必要があると考えられる。

1. Pose Graph Optimization 及びバンドル調整により 1 度補正されたキーフレームに対しては再度補正を行わない（固定する）
2. バンドル調整処理時のみ固定する（Pose Graph Optimization 及びバンドル調整で 1 度補正されたキーフレームの位置姿勢に対して、再度 Pose Graph Optimization が実行された場合は再度補正処理を行うが、再度バンドル調

- 整が行われた場合は補正処理を行わない)
3. Pose Graph Optimization 処理時のみ固定する (Pose Graph Optimization 及びバンドル調整で 1 度補正されたキーフレームの位置姿勢に対して, 再度 Pose Graph Optimization が実行された場合は補正処理を行わず, 再度バンドル調整が行われた場合は再度補正処理を行う)
 4. Pose Graph Optimization 及びバンドル調整により 1 度補正されたキーフレームに対して, 再度補正処理を行う (固定しない)

また, これら 4 組の Pose Graph Optimization 及びバンドル調整の組み合わせを用いて, SLAM の精度に与える影響を調査する方法として, 本研究では Pose Graph Optimization 及びバンドル調整前後のカメラの位置姿勢の真値と推定値を比較している. しかし, ORB-SLAM のループクローズでは, カメラシーケンス内にループが存在した場合のみ Pose Graph Optimization 及びバンドル調整を行っている為, ループが存在しないカメラシーケンスではカメラ位置姿勢を補正することが出来ない. そこで, 本研究では小節 2.3.1 及び小節 2.3.2 で説明した, 類似度を用いたループの検出部と相似変換行列を用いた K_i の補正部を, AR マーカーを使用して求めるように変更を行うことで, ループが存在しないカメラシーケンスにおいても Pose Graph Optimization 及びバンドル調整が行われるように変更している. 具体的な検証方法およびループクローズに対する変更部分については次章で説明する.

第 3 章 評価実験

Pose Graph Optimization 及びバンドル調整が推定精度に与える影響を調査するために、最適化前後のカメラの位置姿勢とその真値を比較する実験を行った。この時、ループの存在しないカメラ移動に対しても対応できるように、ループクローズの一部を変更した。本章では、まずループクローズにどのような変更を加えたのかを説明し、次に検証方法について述べ、最後に実験結果とその考察について述べる。

3.1 ループクローズに対する変更

キーフレームが図 3.1 の様に保存されているとすると、ORB-SLAM では節 2.3 で説明したような手順で最適化が行われる。この時、ORB-SLAM ではループの検出に類似度を使用している。しかし、この方法ではループを含まないカメラシーケンスではループクローズを行うことが出来ない。そこで本実験では、ORB-SLAM に ARToolkit[15] を組み込み、AR マーカーを使用してループクローズと同等のことを行った。具体的には、まず、ループの検出部 (小節 2.3.1) に対し変更を行った。ORB-SLAM では類似度の最も高いキーフレームをループ候補のキーフレームとしているが、変更後は、2 回同じ AR マーカーが認識された時、初めにマーカーを認識したキーフレームをループ候補 (K_l)、2 回目にマーカーを認識したキーフレームをループ終わりのキーフレーム (K_i) とする様に変更した。また、ループの結合 (小節 2.3.2) についても変更を行っている。ORB-SLAM では類似度 S_{il} を使用し、図 2.4 の様に K_i を補正しているが、変更後は、同じマーカーを認識しているキーフレーム間の相対位置関係を利用して K_i の補正を行っている。具体的には、 K_l の位置姿勢 (初期カメラ位置との相対位置) を T_1 、 K_i の位置姿勢を T_2 とし、 K_l で認識されたマーカーとカメラの座標変換行列を T_3 、 K_i で認識されたマーカーとカメラの座標変換行列を T_4 とすると、 K_l から見た K_i の位置関係 T_{il} は以下の式で表すことが出来る。

$$T_{il} = T_4 * T_3^{-1} \quad (3.1.1)$$

上記の式により求められた T_{il} と K_l の位置姿勢 T_1 から、補正された K_i の位置姿勢 T'_2 を以下のように求めることが出来る．

$$T'_2 = T_{il} * T_1 \quad (3.1.2)$$

この位置姿勢 T'_2 を，類似度を使用して補正された， K_i の位置姿勢 T_{iw} の代わりに，図 3.1 の様に補正している．以下，ループクローズに対して変更を行った最適化を「Pose Graph Optimization」と呼称している．

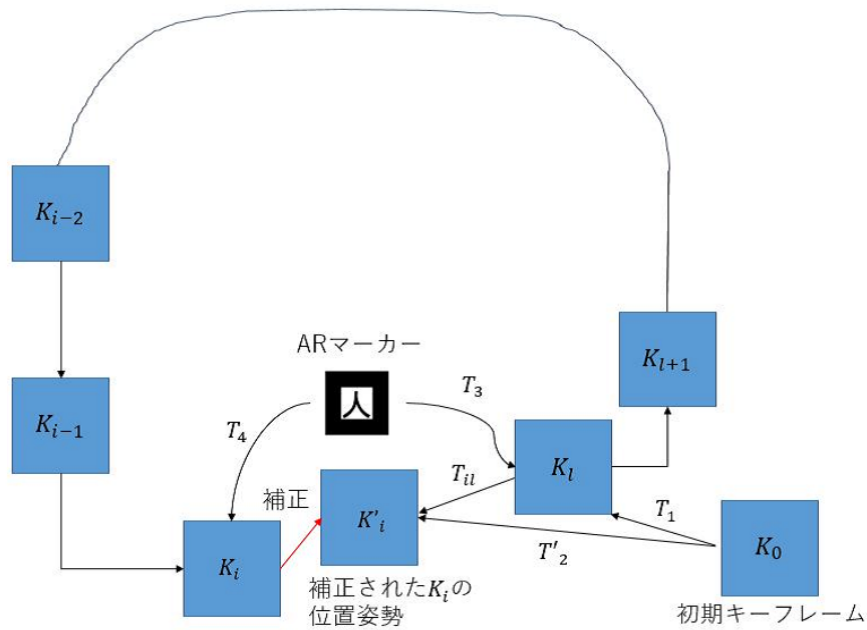


図 3.1 補正された K_i の位置姿勢 T_{iw}

3.2 検証方法

シミュレータ環境を構築し，カメラ位置姿勢の推定結果と真値を比較することで推定精度の評価を行った．評価では以下の5つの条件で比較した．また，実環境下でカメラを動かし，シミュレーターと同様の条件で評価を行った．

1. Pose Graph Optimization 及びバンドル調整により補正されたキーフレームに対しては再度補正を行わない（固定する）
2. バンドル調整処理時のみ固定する
3. Pose Graph Optimization 処理時のみ固定する
4. Pose Graph Optimization 及びバンドル調整により 1 度補正されたキーフレームに対して、再度補正処理を行う（固定しない）
5. 変更前の Pose Graph Optimization 及びバンドル調整を行う。

距離誤差（位置誤差）はカメラ位置間のユークリッド距離として算出し、姿勢誤差は真値の逆行列と推定値を掛けた値を、ロドリゲスの回転公式を用いて変換し、そのノルムを計算することで算出した。具体的には、位置姿勢の真値を T_t 、推定値を T_s とすると、真値の逆行列と推定値を掛けた値 T_{ts} は以下の式で表される。

$$T_{ts} = T_s * T_t^{-1} \quad (3.2.1)$$

この T_{ts} をロドリゲスの回転公式を用いて行列 R_{ts} に変換する。

$$R_{ts} = \begin{bmatrix} x_r & y_r & z_r \end{bmatrix}^T \quad (3.2.2)$$

そして、 R_{ts} のノルム L_{ts} を以下のように求め、これを度数法に直し、姿勢誤差とした。

$$L_{ts} = \sqrt{x_r^2 + y_r^2 + z_r^2} \quad (3.2.3)$$

また、距離誤差及び姿勢誤差は、誤差測定時にマップに保存されている全てのキーフレームにおいて上記誤差算出法を使用し、その平均値から求めている。

更に、現実環境下でもシミュレーター環境下と同様に評価実験を行った。この時、現実環境下でのカメラの位置姿勢の真値をマーカー間の位置姿勢から求めている。具体的には、図 3.2 の様にマーカー 1 からマーカー 4 までの 4 枚のマーカーが設置されており、それぞれのマーカー間の位置関係 T_{12} , T_{23} , T_{34} が分かっているとす。そして初期キーフレーム K_1 でマーカー 1 を、キーフレーム K_m でマーカー 4 を認識しているとす、それぞれのキーフレームが認識しているマーカーの座標系からカメラ座標系への変換行列を T_1 , T_i とする。この時、キーフレーム K_m のカメラの位置姿勢の真値 T_{1m} は以下の式で求めることが出来る。

$$T_{1m} = T_i * T_{34} * T_{23} * T_{12} * T_1^{-1} \quad (3.2.4)$$

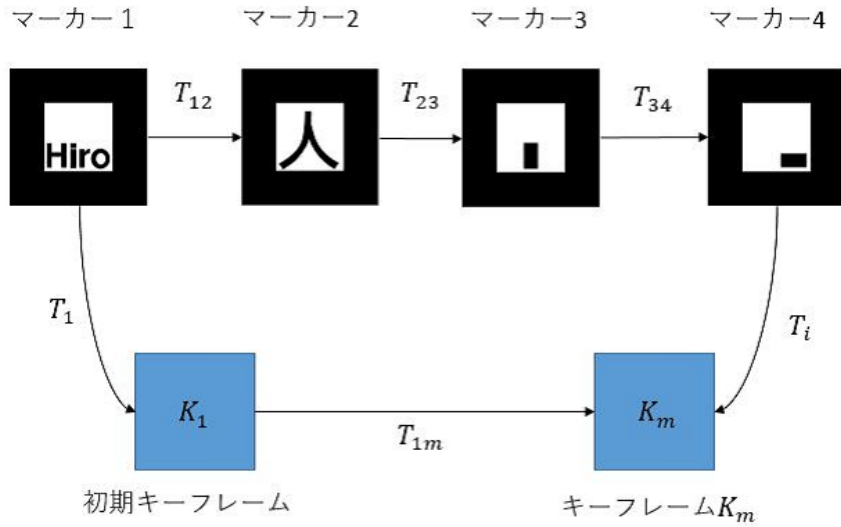


図 3.2 カメラ位置姿勢の真値

3.3 シミュレータ環境

OpenGL を使用してシミュレータ環境を構築し、「カメラの動き」「設置されたマーカーの数」を変更した 1 から 3 までの 3 つのカメラシーケンスを作成した．以下にそれぞれのカメラシーケンスの動きを簡潔に表す．

- カメラシーケンス 1 : 2 枚のマーカーを設置し、カメラを 1 方向を向くように固定した状態で、反時計回りに円を描くように 1 週半カメラを動かした．(図 3.5)
- カメラシーケンス 2 : シーケンス 1 とカメラの動き方は変えず、カメラの移動距離を増加させている．(図 3.4)
- カメラシーケンス 3 : シーケンス 2 とカメラの動き方、移動距離を変えずマーカーを 3 枚に増やしている．(図??)

更に、それぞれのカメラシーケンスの違いについて、まず初めに「カメラの動き」の変更について説明する．カメラの位置姿勢推定に対する蓄積誤差は、カメラの移

動距離によって変化する。その為、カメラが移動し始めてから止まるまでのカメラの移動距離を変更している。また、「設置されたマーカ-の数」の変更について説明する。ポーズグラフ最適化は AR マーカ-を 2 度以上認識する毎に行われる。その為、マーカ-の設置枚数を変更することで最適化の回数による推定精度への影響を確認している。

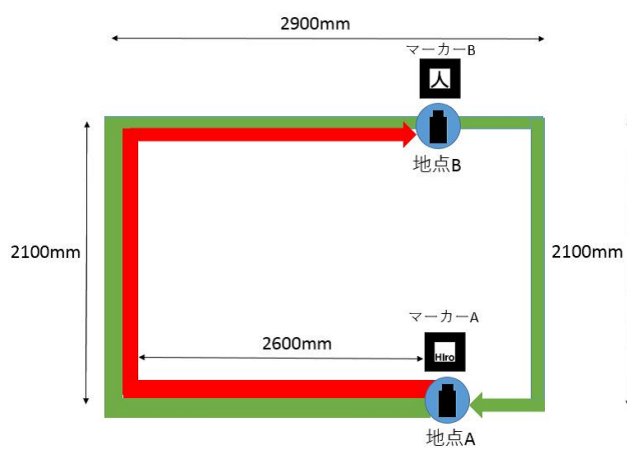


図 3.3 カメラシーケンス 1 のカメラの動き

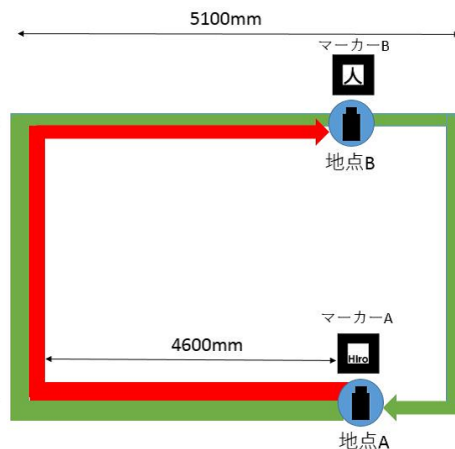


図 3.4 カメラシーケンス 2 のカメラの動き

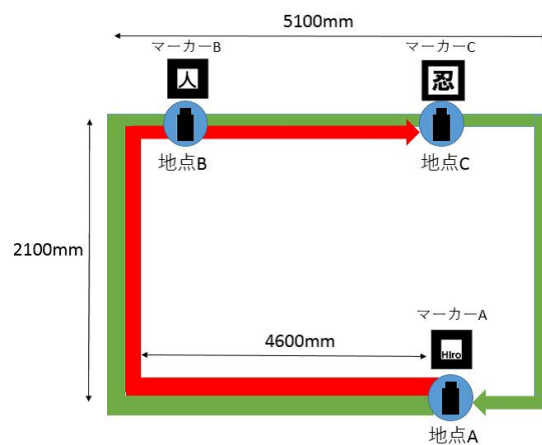


図 3.5 カメラシーケンス 3 のカメラの動き

3.4 実験結果

ループを発見すると、その時マップに保存されている各キーフレームそれぞれの誤差の平均を最適化前の誤差，Pose Graph Optimization が行われた直後にマッ

プに保存されている各キーフレームの誤差の平均を Pose Graph Optimization 後の誤差, Pose Graph Optimization が行われ, 更にバンドル調整が行われた直後にマップに保存されている各キーフレームの誤差の平均をバンドル調整後の誤差として, ループを発見する毎に測定している. まず小節 3.2 で説明した 3 シーケンス及び実環境それぞれでの最終距離誤差及び最終姿勢誤差を表 3.1, 表 3.2, 表 3.3, 表 3.4 に示す. この時, 最終距離誤差及び最終姿勢誤差とは, 各シーケンスで最後に行われたバンドル調整終了時にマップに保存されていたキーフレームの距離誤差及び姿勢誤差の平均値である. 例えば, あるシーケンスでループが 3 度発見されたとすると, 3 度めのループ発見時に行われたバンドル調整が終了した時点でマップに保存されている, 各キーフレームの誤差の平均値が最終誤差となる.

表 3.1 カメラシーケンス 1 での最終誤差

	カメラシーケンス 1	
条件	最終距離誤差 [mm]	最終姿勢誤差 [度]
両固定	34.70732	0.64879
バンドル調整固定	30.42985	0.70030
Pose Graph 固定	37.13808	0.60230
両非固定	35.56207	0.51579

表 3.2 カメラシーケンス 2 での最終誤差

	カメラシーケンス 2	
条件	最終距離誤差 [mm]	最終姿勢誤差 [度]
両固定	85.20977	0.90511
バンドル調整固定	94.56001	1.09052
Pose Graph 固定	101.72670	0.93415
両非固定	76.63296	1.05183

表 3.3 カメラシーケンス 3 での最終誤差

	カメラシーケンス 3	
条件	最終距離誤差 [mm]	最終姿勢誤差 [度]
両固定	67.64799	0.74555
バンドル調整固定	89.55339	1.27352
Pose Graph 固定	76.58587	0.82363
両非固定	87.91056	0.96844

表 3.4 実環境での最終誤差

	実環境	
条件	最終距離誤差 [mm]	最終姿勢誤差 [度]
両固定	57.11464	1.35205
バンドル調整固定	105.29035	1.71785
Pose Graph 固定	85.22844	0.82363
両非固定	82.68325	1.46719

3.4.1 カメラシーケンス 1

カメラシーケンス 1 では、1 方向に向きを固定したカメラをパスが長方形を描くように反時計回りに動かしている。具体的には、図 3.5 の地点 A において、マーカー A を認識した状態からカメラを動かし、左に 2600mm、前に 2100mm、右に 2900mm、後ろに 2100mm、左に 2600mm、前に 2100mm、右に 2600mm 移動させた後、地点 B を 2 回目に通過した時点で停止させている。この時、地点 A 及び地点 B を 2 回目に通過した際にループが発見され、Pose Graph Optimization 及びバンドル調整が行われる。その為、カメラシーケンス 1 では合計 2 度の Pose Graph Optimization 及びバンドル調整が行われることになる。この時、「変更前の Pose Graph Optimization 及びバンドル調整を行う」を除く 4 条件下で得られた距

離誤差及び姿勢誤差は表 3.5, 表 3.7 の様になった. また, 「変更前の Pose Graph Optimization 及びバンドル調整を行う」条件下で得られた距離誤差及び姿勢誤差は, 表 3.6, 表 3.8 の様になった. また, 最後の最適化 (シーケンス 1 では 2 度目のバンドル調整) が終了した時, マップに保存されている各キーフレーム毎の距離誤差を, 各条件毎にそれぞれ図 3.6, 図 3.7, 図 3.8, 図 3.9 に示す.

表 3.5 カメラシーケンス 1 での距離誤差

		距離誤差 [mm]	
条件	キーフレーム取得時	A 認識	B 認識
両固定	最適化処理前	54.06005	39.95970
	Pose Graph Optimization 後	37.64018	36.09212
	バンドル調整後	31.86715	34.70732
バンドル調整固定	最適化処理前	65.16226	30.43273
	Pose Graph Optimization 後	35.88875	32.05061
	バンドル調整後	25.35048	30.42985
Pose Graph 固定	最適化処理前	55.49368	35.93005
	Pose Graph Optimization 後	46.21104	35.20297
	バンドル調整後	37.14838	37.13808
両非固定	最適化処理前	54.49131	35.54189
	Pose Graph Optimization 後	36.48911	36.18095
	バンドル調整後	30.01437	35.56207

表 3.6 カメラシーケンス 1 での「オリジナル」の距離誤差

キーフレーム取得時	距離誤差 [mm]
最適化処理前	63.54356
ループクローズ後	36.67494
バンドル調整後	35.26353

表 3.7 カメラシーケンス 1 での姿勢誤差

		姿勢誤差 [度]	
条件	キーフレーム取得時	B 認識	A 認識
両固定	最適化処理前	1.03664	0.92138
	Pose Graph Optimization 後	0.62844	0.71394
	バンドル調整後	0.53809	0.64879
バンドル調整固定	最適化処理前	1.19161	0.68671
	Pose Graph Optimization 後	0.63162	0.75033
	バンドル調整後	0.49048	0.70030
Pose Graph 固定	最適化処理前	0.99175	0.63854
	Pose Graph Optimization 後	0.74933	0.65646
	バンドル調整後	0.573851	0.602298
両非固定	最適化処理前	0.79285	0.57991
	Pose Graph Optimization 後	0.56881	0.58844
	バンドル調整後	0.45793	0.51579

表 3.8 カメラシーケンス 1 での「オリジナル」の姿勢誤差

キーフレーム取得時	姿勢誤差 [度]
最適化処理前	0.77371
ループクローズ後	0.50428
バンドル調整後	0.40474



図 3.6 シーケンス 1 での各キーフレーム毎の距離誤差（両固定）

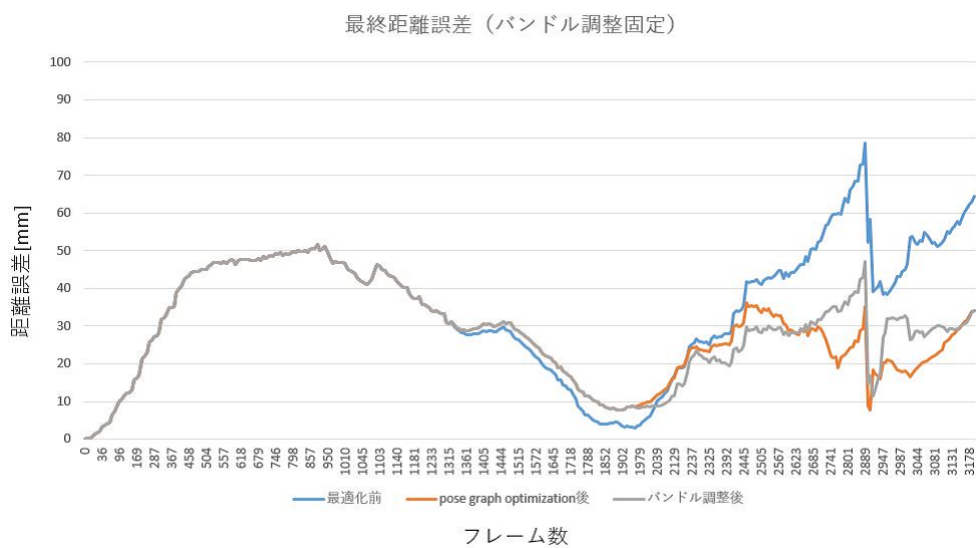


図 3.7 シーケンス 1 での各キーフレーム毎の距離誤差（バンドル固定）

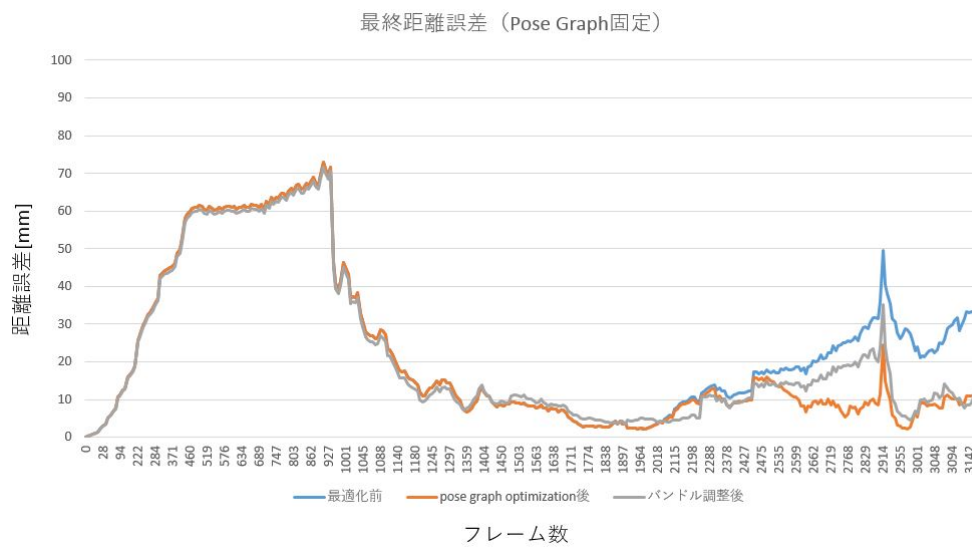


図 3.8 シーケンス 1 での各キーフレーム毎の距離誤差 (Pose Graph 固定)

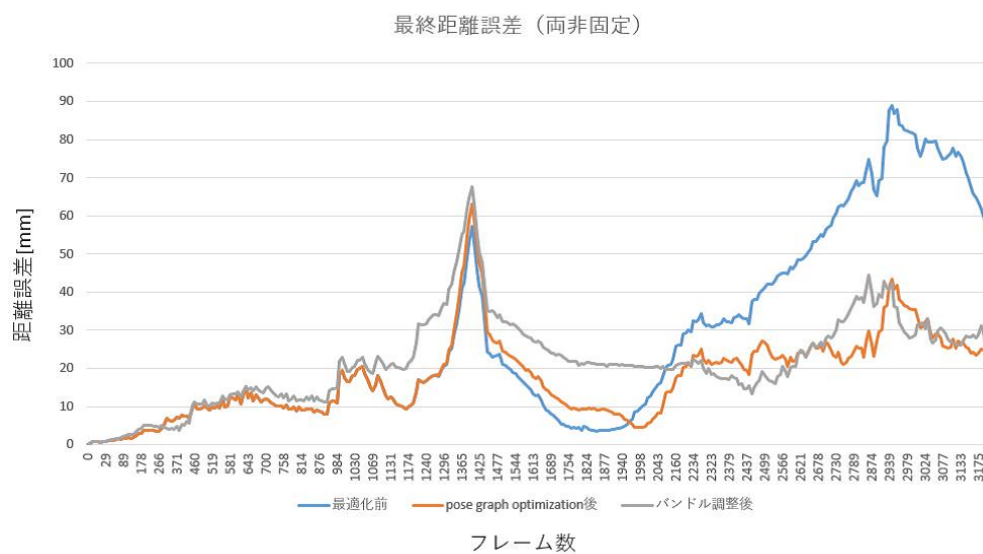


図 3.9 シーケンス 1 での各キーフレーム毎の距離誤差 (両非固定)

3.4.2 カメラシーケンス 2

カメラシーケンス 2 では，カメラシーケンス 1 と同様に，1 方向に向きを固定したカメラをパスが長方形を描くように反時計回りに動かしている．但し，カメラの移動距離の変化が精度に与える影響を確認するために，シーケンス 1 よりカメラの移動距離を増加させている．具体的には，図 3.4 の地点 A において，マーカー A を認識した状態からカメラを動かし，左に 4600mm，前に 2100mm，右に 5100mm，後ろに 2100mm，左に 4600mm，前に 2100mm，右に 4600mm 移動させた後，地点 B を 2 回目に通過した時点で停止させている．この時，地点 A 及び地点 B を 2 回目に通過した際にループが発見され，Pose Graph Optimization 及びバンドル調整が行われる．その為，カメラシーケンス 2 ではそれぞれ 2 度ずつ Pose Graph Optimization 及びバンドル調整が行われることになる．この時，「変更前の Pose Graph Optimization 及びバンドル調整を行う」を除く 4 条件下で得られた距離誤差及び姿勢誤差は表 3.9，表 3.10 の様になった．また，「変更前の Pose Graph Optimization 及びバンドル調整を行う」の条件下で得られた距離誤差及び姿勢誤差は，表 3.11，表 3.12 の様になった．また，最後の最適化（シーケンス 2 では 2 度目のバンドル調整）が終了した時，マップに保存されている各キーフレーム毎の距離誤差を，各条件毎にそれぞれ図 3.10，図 3.11，図 3.12，図 3.13 に示す．

表 3.9 カメラシーケンス 2 での距離誤差

		距離誤差 [mm]	
条件	キーフレーム取得時	A 認識	B 認識
両固定	最適化処理前	102.25680	106.36441
	Pose Graph Optimization 後	60.04953	87.92717
	バンドル調整後	61.17793	85.20977
バンドル調整固定	最適化処理前	102.54960	114.74162
	Pose Graph Optimization 後	77.14964	95.11846
	バンドル調整後	73.87953	94.56001
Pose Graph 固定	最適化処理前	100.87390	126.88110
	Pose Graph Optimization 後	74.93199	106.07132
	バンドル調整後	75.35537	101.72672
両非固定	最適化処理前	96.27743	113.64040
	Pose Graph Optimization 後	49.26476	83.58267
	バンドル調整後	51.81492	76.63296

表 3.10 カメラシーケンス 2 での「オリジナル」の距離誤差

キーフレーム取得時	距離誤差 [mm]
最適化処理前	70.89723
ループクローズ後	65.80391
バンドル調整後	67.07543

表 3.11 カメラシーケンス 2 での姿勢誤差

		姿勢誤差 [度]	
条件	キーフレーム取得時	A 認識	B 認識
両固定	最適化処理前	2.58446	0.98079
	Pose Graph Optimization 後	0.86124	0.96423
	バンドル調整後	0.81392	0.90511
バンドル調整固定	最適化処理前	1.92517	1.12129
	Pose Graph Optimization 後	1.05524	1.12406
	バンドル調整後	0.91934	1.09052
Pose Graph 固定	最適化処理前	2.07927	1.13410
	Pose Graph Optimization 後	0.92882	1.05956
	バンドル調整後	0.86308	0.93415
両非固定	最適化処理前	2.33298	1.03948
	Pose Graph Optimization 後	0.96298	1.23061
	バンドル調整後	0.91509	1.05183

表 3.12 カメラシーケンス 2 での「オリジナル」の姿勢誤差

キーフレーム取得時	姿勢誤差 [度]
最適化処理前	1.32479
ループクローズ後	1.34208
バンドル調整後	1.25737

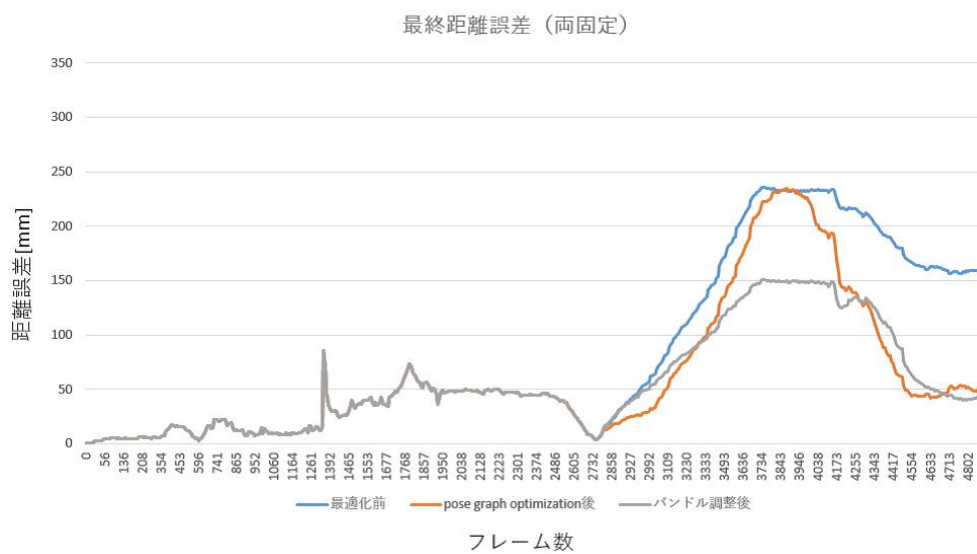


図 3.10 シーケンス 2 での全キーフレームの距離誤差（両固定）

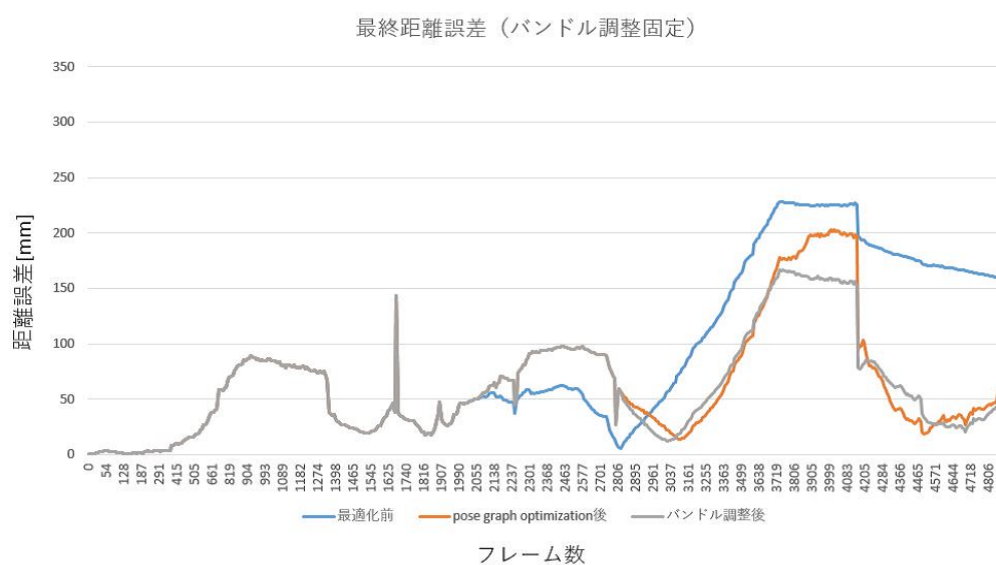


図 3.11 シーケンス 2 での全キーフレームの距離誤差（バンドル固定）

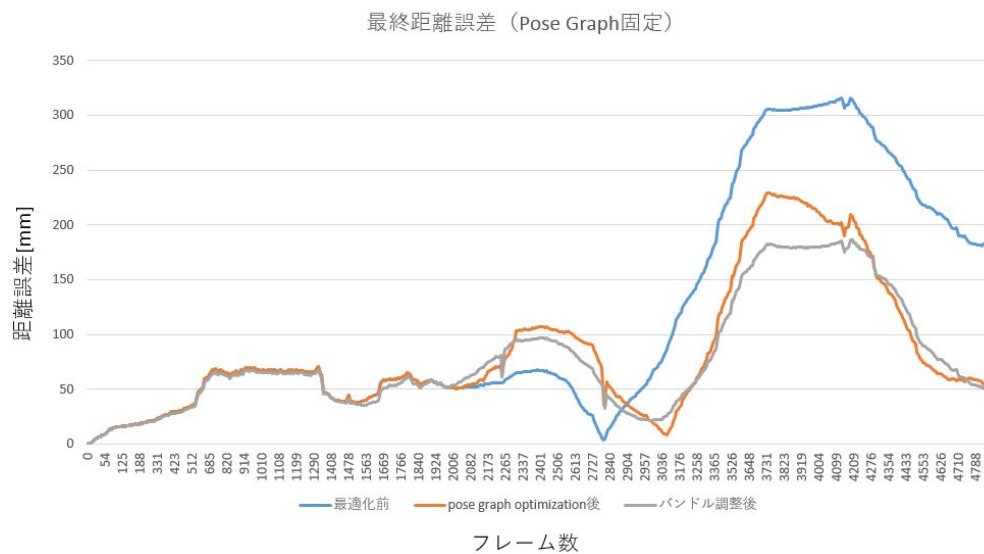


図 3.12 シーケンス 2 での全キーフレームの距離誤差 (Pose Graph 固定)

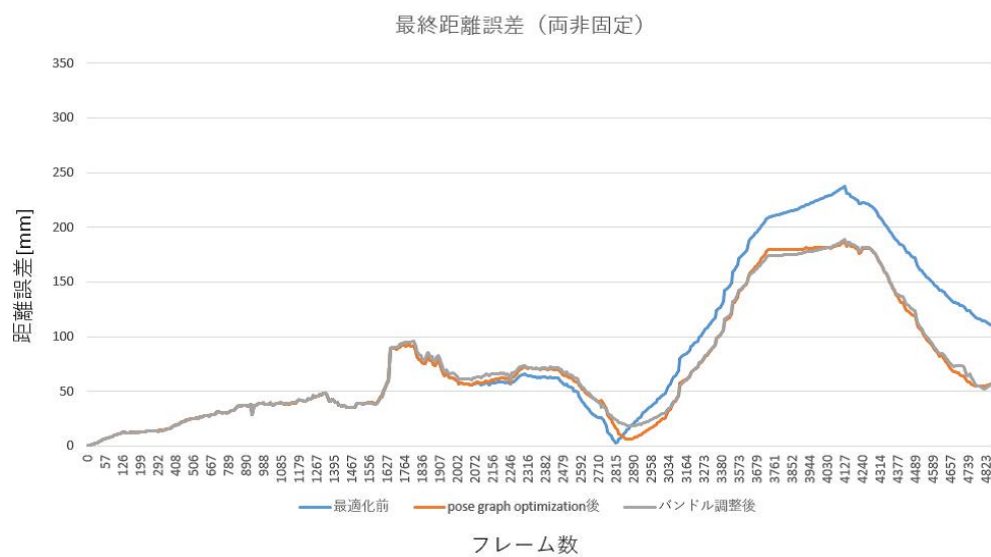


図 3.13 シーケンス 2 での全キーフレームの距離誤差 (両非固定)

3.4.3 カメラシーケンス 3

カメラシーケンス 3 では、カメラシーケンス 1, 2 と同様に、1 方向に向きを固定したカメラをパスが長方形を描くように反時計回りに動かしている。また、カメラの移動距離はカメラシーケンス 2 と同じ距離だけ移動させている。但し、Pose Graph Optimization 及びバンドル調整の回数の変化が精度に与える影響を確認するために、マーカーの数を 1 枚増加することにより、Pose Graph Optimization 及びバンドル調整の回数を増加させている。具体的には、図 3.4 の地点 A において、マーカー A を認識した状態からカメラを動かし、左に 4600mm, 前に 2100mm, 右に 5100mm, 後ろに 2100mm, 左に 4600mm, 前に 2100mm, 右に 4600mm 移動させた後、地点 B を 2 回目に通過した時点で停止させている。この時、地点 A, 地点 B 及び地点 C をそれぞれ 2 回目に通過した際にループが発見され、Pose Graph Optimization 及びバンドル調整が行われる。その為、カメラシーケンス 3 では 3 度ずつ Pose Graph Optimization 及びバンドル調整が行われることになる。この時、「変更前の Pose Graph Optimization 及びバンドル調整を行う」を除く 4 条件下で得られた距離誤差及び姿勢誤差は表 3.9, 表 3.10 の様になった。また、「変更前の Pose Graph Optimization 及びバンドル調整を行う」の条件下で得られた距離誤差及び姿勢誤差は、表 3.11, 表 3.12 の様になった。また、最後の最適化（シーケンス 3 では 3 度目のバンドル調整）が終了した時、マップに保存されている各キーフレーム毎の距離誤差を、各条件毎にそれぞれ図 3.14, 図 3.15, 図 3.16, 図 3.17 に示す。

表 3.13 カメラシーケンス 3 での距離誤差

		距離誤差 [mm]		
条件	キーフレーム取得時	A 認識	B 認識	C 認識
両固定	最適化処理前	98.21227	91.17370	75.39526
	Pose Graph Optimization 後	59.14475	70.98284	68.15693
	バンドル調整後	57.68488	66.14965	67.64799
バンドル調整固定	最適化処理前	93.84474	87.27951	89.61283
	Pose Graph Optimization 後	55.50191	85.86154	89.62795
	バンドル調整後	57.39843	84.39850	89.55339
Pose Graph 固定	最適化処理前	111.67884	110.41187	84.71156
	Pose Graph Optimization 後	69.25443	94.20285	78.05428
	バンドル調整後	74.19199	77.39601	76.58587
両非固定	最適化処理前	105.12748	91.65058	85.99174
	Pose Graph Optimization 後	56.43236	89.19603	88.91829
	バンドル調整後	61.04476	84.73544	87.91056

表 3.14 カメラシーケンス 3 での「オリジナル」の距離誤差

キーフレーム取得時	距離誤差 [mm]
最適化処理前	110.72025
ループクローズ後	66.17143
バンドル調整後	72.16627

表 3.15 カメラシーケンス 3 での姿勢誤差

		姿勢誤差 [度]		
条件	キーフレーム取得時	A 認識	B 認識	C 認識
両固定	最適化処理前	2.22120	0.69008	0.87077
	Pose Graph Optimization 後	0.79890	0.75083	0.75348
	バンドル調整後	0.63379	0.73779	0.74555
バンドル調整固定	最適化処理前	1.54552	0.88805	1.22503
	Pose Graph Optimization 後	0.78844	1.11160	1.28287
	バンドル調整後	0.67811	1.17067	1.27352
Pose Graph 固定	最適化処理前	1.74808	0.94929	0.84622
	Pose Graph Optimization 後	0.75629	0.89353	0.80653
	バンドル調整後	0.75608	0.82895	0.82364
両非固定	最適化処理前	1.71539	0.58678	0.83823
	Pose Graph Optimization 後	0.56501	0.99956	0.96059
	バンドル調整後	0.51206	0.83075	0.83075

表 3.16 カメラシーケンス 3 での「オリジナル」の姿勢誤差

キーフレーム取得時	姿勢誤差 [度]
最適化処理前	1.88235
ループクローズ後	1.51219
バンドル調整後	1.19440

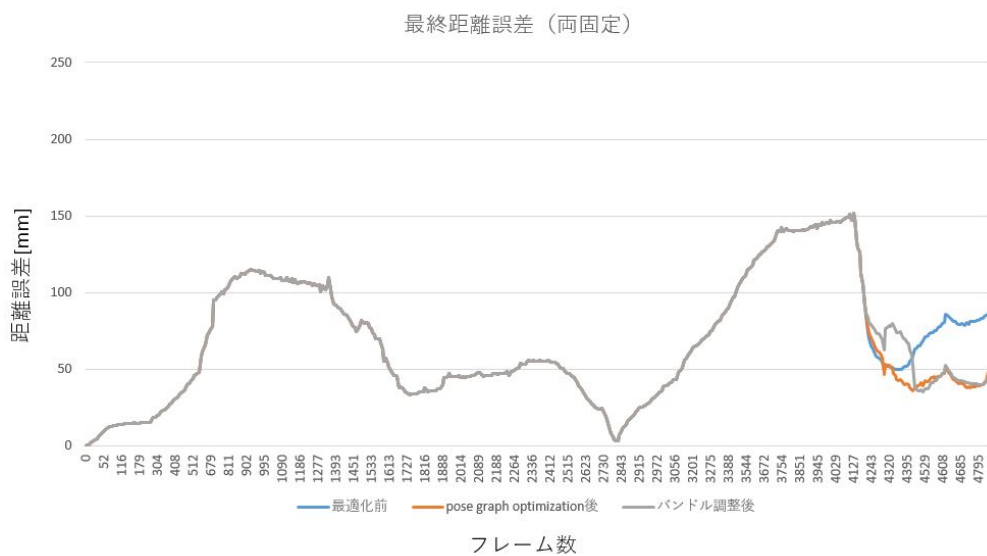


図 3.14 シーケンス 3 での全キーフレームの距離誤差（両固定）

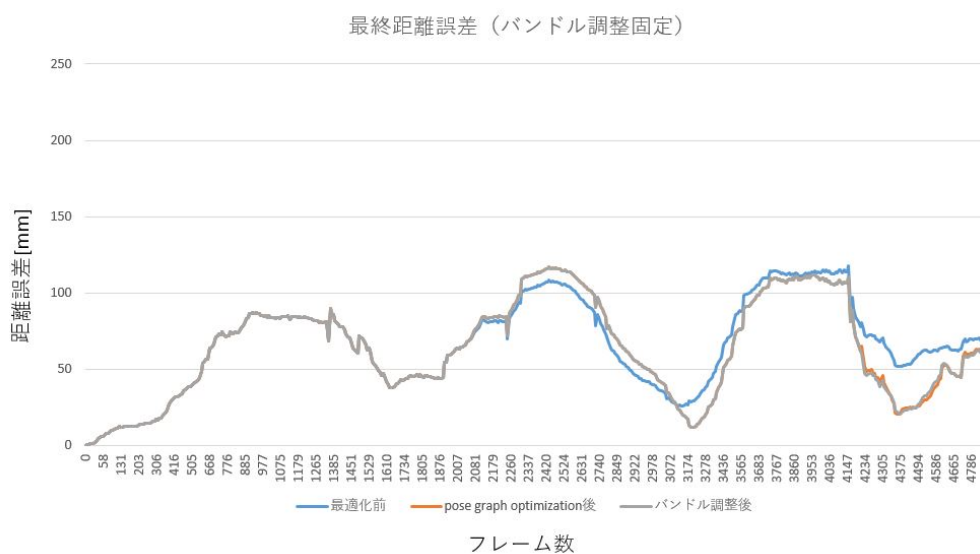


図 3.15 シーケンス 3 での全キーフレームの距離誤差（バンドル固定）

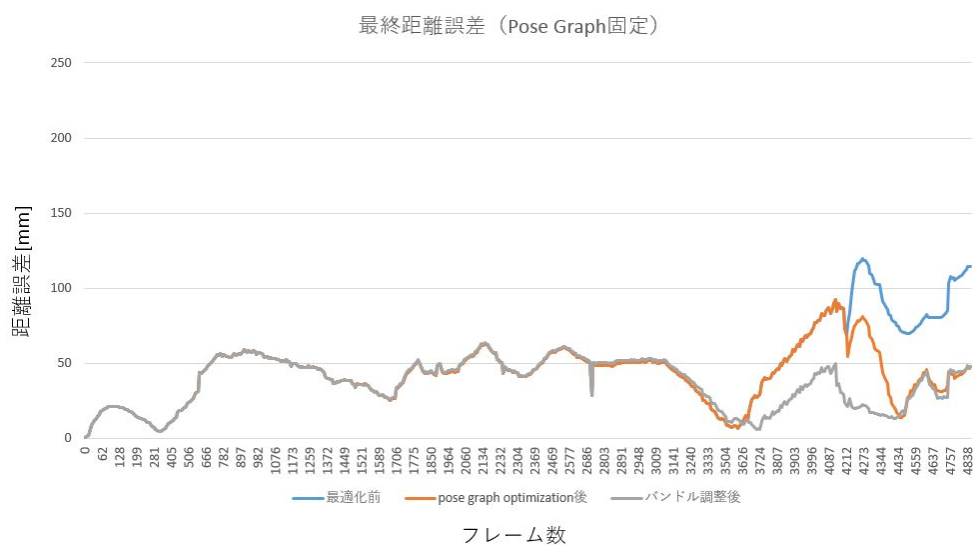


図 3.16 シーケンス 3 での全キーフレームの距離誤差 (Pose Graph 固定)

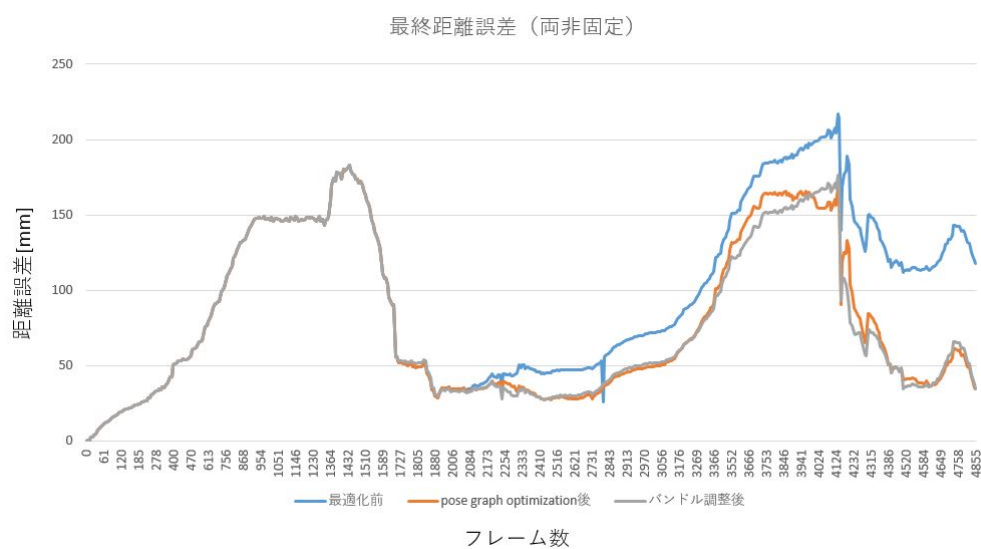


図 3.17 シーケンス 3 での全キーフレームの距離誤差 (両非固定)

3.4.4 実環境

実環境測定では，シーケンス 1, 2, 3 と違い，カメラの向きを固定せず，カメラが直線上を往復するようにカメラを動かした．この時ループ検出用のマーカー 2 枚 (マーカー A 及びマーカー B) と真値作成用のマーカー 10 枚，計 12 枚のマーカーを設置して実験を行った．また，カメラはループ検出用のマーカーを 2 度ずつ認識し，計 2 回の Pose Graph Optimization 及びバンドル調整が行われた．また，真値を作成するために，常にマーカーを 2 枚以上認識している状態でカメラを動かすことでマーカーの位置関係を全て取得した．推定されたカメラパスを図 3.19 に環境の写真を図??に示す．そして，「変更前の Optimization 及びバンドル調整を行う」を除く 4 条件下で得られた距離誤差及び姿勢誤差は表 3.17，表 3.18 の様になった．また，「変更前の Pose Graph Optimization 及びバンドル調整を行う」の条件ではループが検出されなかった為，最適化は行われなかった．

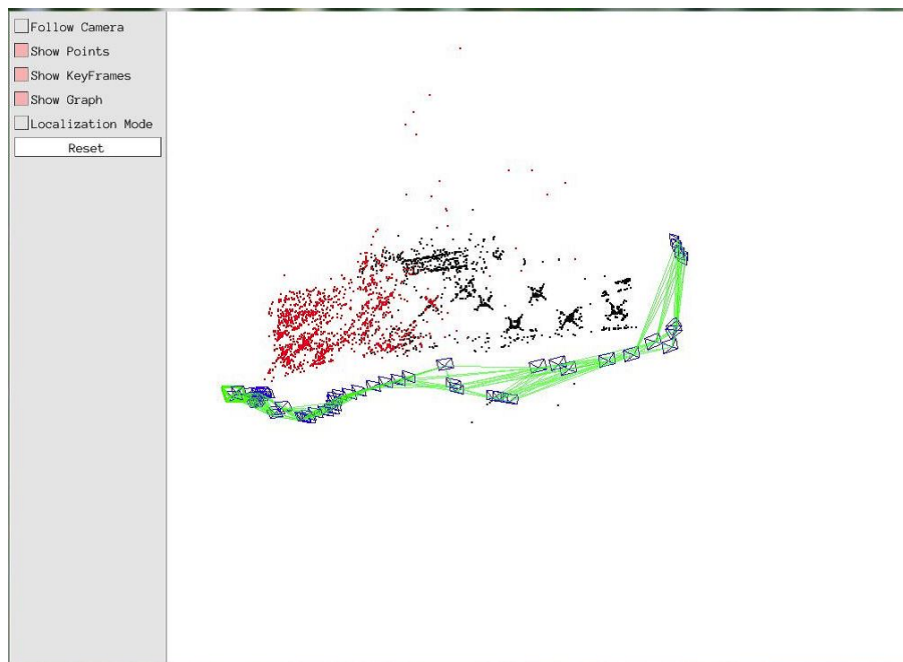


図 3.18 推定されたカメラパス



图 3.19 測定環境

表 3.17 実環境での距離誤差

		距離誤差 [mm]	
条件	キーフレーム取得時	A 認識	B 認識
両固定	最適化処理前	76.86206	58.10341
	Pose Graph Optimization 後	57.88054	56.83604
	バンドル調整後	61.32011	57.11463
バンドル調整固定	最適化処理前	129.17094	102.21813
	Pose Graph Optimization 後	104.14460	87.2256
	バンドル調整後	105.29035	88.88736
Pose Graph 固定	最適化処理前	120.71457	87.89497
	Pose Graph Optimization 後	93.64003	84.89293
	バンドル調整後	92.93773	85.22844
両非固定	最適化処理前	102.94631	79.84152
	Pose Graph Optimization 後	80.38859	64.02665
	バンドル調整後	82.68325	68.83704

表 3.18 実環境での姿勢誤差

		姿勢誤差 [度]	
条件	キーフレーム取得時	A 認識	B 認識
両固定	最適化処理前	1.38592	1.29856
	Pose Graph Optimization 後	1.06426	1.30847
	バンドル調整後	1.37766	1.35205
バンドル調整固定	最適化処理前	1.34035	1.34612
	Pose Graph Optimization 後	1.04363	1.12923
	バンドル調整後	1.42612	1.22366
Pose Graph 固定	最適化処理前	1.50158	1.36390
	Pose Graph Optimization 後	1.07733	1.34530
	バンドル調整後	1.50053	1.71785
両非固定	最適化処理前	1.41194	1.25666
	Pose Graph Optimization 後	1.15437	1.195091
	バンドル調整後	1.26717	1.46719

3.5 考察

まず、カメラの移動距離が誤差に与える影響を調査するために、カメラシーケンス 1 とカメラシーケンス 2 の結果を比較する。カメラシーケンス 1 と 2 ではカメラの動き方は同様であり、カメラの移動距離を変更している。この 2 つのシーケンスを比較すると、距離誤差において、地点 A で行われたポーズグラフ最適化及びバンドル調整の結果は、カメラシーケンス 1 及び 2 において、共に 20mm 以上の誤差の減少が見られた。特にカメラシーケンス 1 では「バンドル調整固定」の条件で最適化前に測定された 65.16226mm の誤差が 25.35048mm まで減少しており、カメラシーケンス 2 では「両非固定」の条件で最適化前に測定された 96.27743mm の誤差が 51.81492mm まで減少していた。また、地点 B で行われたポーズグラフ最適化及びバンドル調整最適化の結果、カメラシーケンス 1 では誤差の減少はあまり

見られなかった。しかし、カメラシーケンス 2 では 20mm 以上の誤差の減少が見られた。これらの結果から、カメラの移動距離の増加による距離誤差の蓄積に対して、ポーズグラフ最適化及びバンドル調整は精度に対し、良い影響を与えていると推測される。また、姿勢誤差に関しては、カメラシーケンス 1 及び 2 の両シーケンスにおいて大きな誤差の減少は確認出来なかった。その為、カメラが向く方向を固定している場合、ポーズグラフ最適化及びバンドル調整は姿勢誤差にあまり影響を与えていないと推測される。

次に、最適化の回数が誤差に与える影響を調査するために、カメラシーケンス 2 とカメラシーケンス 3 の結果を比較する。カメラシーケンス 2 と 3 ではカメラの移動距離、動き方をは同様であるが、カメラシーケンス 3 ではカメラシーケンス 2 と比べ、マーカーを 1 枚増やすことで最適化の回数を増加させている。この 2 つのシーケンスを比較すると、距離誤差において、2 回ポーズグラフ最適化及びバンドル調整が行われたカメラシーケンス 2 と比べ、3 回ポーズグラフ最適化及びバンドル調整が行われたカメラシーケンス 3 では、最終的な距離誤差が全ての条件において減少していることが確認できた。その為、蓄積された距離誤差に対して、最適化回数を増加させることは SLAM の精度に良い影響を与えていると推測できる。また姿勢誤差に対しても同様に、カメラシーケンス 2 よりカメラシーケンス 3 の方が最終姿勢誤差が減少していた。その為、姿勢誤差に対しても距離誤差と同様に、SLAM の精度に良い影響を与えていると推測される。

更に、最終的にマップに保存されている各キーフレーム毎の誤差をシーケンス毎に比較する為に、各キーフレーム毎の距離誤差を比較する。各キーフレーム毎の距離誤差では、マーカーを認識しているキーフレームが最も誤差が小さくなり、そのキーフレームを中心として距離誤差が増加すると考えられる。しかし、結果は全てのシーケンスにおいて、最も誤差が少ないキーフレームはマーカーを認識しているキーフレームではなかった。これは、マーカーを認識しているキーフレームには蓄積誤差の他に「マーカー認識誤差」が含まれている。その為、マーカーを認識しているキーフレームが最小誤差とはならなかったと推測される。

最後に、実環境での測定結果について考察を行う。実環境での測定において、変更前のループクローズではループが検出されず、最適化が行われなかった。しかし、マーカーを使用した最適化（条件 1 ～ 4）の結果、最大で距離誤差が 41mm

程度減少していた．その為，ループが検出されない環境においても，Pose Graph Optimization 及びバンドル調整は SLAM の精度に良い影響を与えていると推測される．

第 4 章 まとめ

本論文では，Pose Graph Optimization 及びバンドル調整が SLAM の精度に与える影響を調査するために，Pose Graph Optimization 及びバンドル調整前後のカメラ位置姿勢の推定値と真値を比較することで，それぞれの最適化が SLAM の精度に与える影響を調査する実験を行った．この時，ORB-SLAM にはシミュレータ環境を構築して作成したカメラシーケンスまたは実環境で実際にカメラを動かして得たシーケンスを入力とした．実験の結果，カメラの移動距離が増加することによる距離誤差に対して，Pose Graph Optimization 及びバンドル調整は良い影響を与えていることが確認された．また，最適化回数を増加させることは，蓄積された距離誤差及び姿勢誤差に対して良い影響を与えることが確認された．そして，実環境において，カメラがループ軌道を描かない場合においても Pose Graph Optimization 及びバンドル調整は SLAM の精度に対し良い影響を与えることが確認された．

謝辞

本研究を行うにあたり，主指導教員として，研究のみならず研究生活等の面でも入学より2年間，ご指導ご鞭撻を賜りましたインタラクティブメディア設計学研究室の加藤 博一教授に心より厚くお礼申し上げます．学内での発表の場において，研究に対する貴重なご意見を頂いた，副指導教員である光メディアインターフェース研究室，向川 康博教授に心より感謝いたします．研究会等の場において，様々なご意見を頂きました副指導教員の Christian Sandor 准教授．何気ない質問に対しても真摯に対応してくださった山本 豪志郎助教授及び，Alexander Plopski 助教授．並びに事務関係などの様々な手続きの際に御助力頂いた上野真紀子秘書に，心よりお礼申し上げます．また，本研究において，基礎から応用の理論まで手厚くサポートして頂いたにも関わらず，何度もご迷惑をかけ，その度に叱咤激励して下さい，手厚いご指導を頂いた武富 貴史助教授に心より深く感謝申し上げます．そして研究のみならず，日常生活の中で様々なアドバイスやご支援を頂いた，インタラクティブメディア設計学研究室の先輩，同輩，後輩の皆様に心よりお礼申し上げます．最後に，大学院で学ぶ機会を与えてくださり，研究生活を長年に渡り支えてくれた家族に心よりお礼申し上げます．

参考文献

- [1] Raúl Mur-Artal, J.M.M. Montiel and Juan D. Tardós, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” *IEEE Transactions on Robotics*, pp. 1147-1163, 2015.
- [2] G. Klein and D. Murray, “Parallel tracking and mapping for small AR workspaces,” *IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234 , 2007.
- [3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: an efficient alternative to SIFT or SURF,” in *IEEE International Conference on Computer Vision*, Barcelona, pp. 2564– 2571, 2011.
- [4] Rosten, E.and Drummond, “T. Machine learning for hirh-speed corner detection,” *Lecture Notes in Computer Science*, pp. 430-443 , 2006.
- [5] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “BRIEF: Binary Robust Independent Elementary Features,” in *European Conference on Computer Vision*, pp.778–792, 2010.
- [6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, pp. 91–110, 2004.
- [7] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded Up Robust Features,” in *European Conference on Computer Vision*, pp. 404–417, 2006.
- [8] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” in *British Machine Vision Conference*, 2013.
- [9] R. Hartley and A. Zisserman, “Multiple View Geometry in ComputerVision, 2nd ed,” *Cambridge University Press*, 2004.
- [10] H. Strasdat, A. J. Davison, J. M. M. Montiel, and K. Konolige, “Double window optimisation for constant time visual SLAM,” in *IEEE International Conference on Computer Vision* , pp. 2352–2359, 2011.
- [11] D. G ´ alvez-L ´ opez and J. D. Tard ´ os, “Bags of binary words for fast

- place recognition in image sequences,” IEEE Transactions on Robotics, pp. 1188–1197, 2012.
- [12] H. Strasdat, “Local Accuracy and Global Consistency for Efficient Visual SLAM,” Ph.D. dissertation, Imperial College, 2012.
 - [13] B. K. P. Horn, “Closed-form solution of absolute orientation using unit quaternions,” Journal of the Optical Society of America , pp. 629–642, 1987.
 - [14] H. Strasdat, “Local Accuracy and Global Consistency for Efficient Visual SLAM,” Ph.D. dissertation, Imperial College, 2012.
 - [15] <http://www.msoft.co.jp/ar/info/artoolkit-release-ver5-3-1/>