CH-002

ジオタグ付き画像を活用した大規模三次元復元の改善
Correction of Large-Scale SLAM Reconstruction Using Geo-Tagged Images

石見 和也† 山崎 俊彦† 相澤 清晴†
Kazuya Iwami Toshihiko Yamasaki Kiyoharu Aizawa

1 はじめに

画像処理の分野において，映像からの高精度な三次元復元は早期から取り組まれてきた重要な課題である．Simultaneous Localization and Mapping（SLAM）はこの課題における主なアプローチであり，6-DoF のカメラ姿勢と周囲の環境の三次元マップを同時に復元することにより，高精度な三次元復元を実現してきた [1, 2]．

しかし，ステレオカメラや LIDAR と異なり，単眼カメラは距離を直接測定することが不可能なため，単眼カメラを用いた SLAM には徐々にスケールの誤差が蓄積してしまう問題（スケールドリフト問題）があることが報告されている [3]．この問題に対処するため，ループクロージャー [3] や既存の地理情報への位置合わせ [4, 2] といったアプローチが提案されてきた．ループクロージャーとは，カメラが再び同じ場所を観測した際に，蓄積した誤差を解消する手法である．この手法は SLAM において最も一般的な補正方法であるが，同じ場所を観測する動画でないと適応できない．既存の地理情報への位置合わせには，点群や建物の三次元モデルといった GIS（地理情報システム）の情報を利用した手法が挙げられる．しかし，この手法には予め高精度な三次元復元を行う必要がある，情報の存在する限られた環境下でしか利用できない，といった制約が存在する．このような制約を緩和するため，本稿で我々は他の GIS 情報を利用することを試みた．

近年，ジオタグ付き画像のデータセットが世界的に拡張されてきていることから，我々は Google Street View から取得したジオタグ付き画像を利用して，三次元復元の結果を改善する手法を提案する．ジオタグ付き画像を用いて映像の三次元復元結果の座標と世界座標との対応を取得する方法は Agawarl らにより提案された [5]．しかし，入力映像と Google Street View のジオタグ付き画像は，照明環境や視点，周囲の建築物などの変化により点対応の取りづらい問題であることが知られているため，疎な対応のみを用いてスケールドリフトを含む大きな蓄積誤差を改善する必要がある．一般的な位置情報を統合したバンドル調整 [6] では局所解に収束しがちな問題を解決するため，我々は線形変換・スケールドリフトを考慮したポーズグラフ最適化・バンドル調整という，疎から精密な 3 種類の変形手法を提案する．

また，長距離走行動画のデータセットを用いた実験により，3 種類の変形それぞれ，及びフレームワーク全体の有効性を検証する．

2 関連研究

ここでは，GPS や車のオドメトリ等の他センサーを使用せずに三次元復元の結果を改善する手法について言及する．

ループクロージャー: ループクロージャーとは，カメラが再び同じ場所を観測した（カメラの軌道がループを描いた）際に，同じ場所を撮影した 2 つのカメラ姿勢間の蓄積誤差を改善する手法である．Lu 及び Milios らは最初にこの問題をポーズグラフの最適化として定式化した [7] が，当初はまだ非常に計算コストが高い問題であった．近年では，Olson ら [8] や Grisetti ら [9] によりポーズグラフ最適化の手法が高速化され，Strasdad ら [3] により単眼カメラの SLAM におけるスケールドリフト問題に考慮したポーズグラフ最適化が提案された．近年提案されている SLAM [1, 2] はこの技術を取り込み高精度な自己位置推定を実現しているが，この手法はループの存在しない映像では有効に働かないことを注意されたい．

既存の地理情報への位置合わせ: 既存の点群を利用した新たな入力映像の三次元復元は，SLAM の一部として研究されてきた [10, 2]．しかし，密な特徴点のトラッキングが要求されるため，照明環境や周囲の環境の変化の影響を受けやすい特徴がある．また，Lothe ら [11] や Tamaazoustira ら [4] によって，既存の建物の三次元モデルを三次元復元結果の補正に利用する手法が提案された．しかし，これらの手法は単純な形状の建築物に囲まれた場所でしか適応できない問題がある．

3 提案手法

3.1 概要

はじめに，本手法の概要図を図 1 に示す．本手法は大きく分けて
（1）ジオタグ画像を利用し，SLAM 及び世界座標系間の三次元対応を取得（章 3.2）
（2）その対応を用いて三次元復元マップを変形
という 2 つの過程に大別される．また，（2）の変形部分は初期線形変換，Sim（3）での制約付きポーズグラフ最適化，バンドル調整という 3 種類の粗から精密な変形で構成される．それぞれの詳細は章 3.3, 章 3.4，及び章 3.5 に示す．

本研究において，世界座標系は三次元座標系 (x, y, z) で表され，その xz 平面はメートル単位の直交平面座標系であるユニバーサル横メルカトル (UTM) 座標系に対応する．また y 軸は地平面からの高度（メートル単位）に対応する．UTM 座標系上の点は緯度経度に変換することが可能である．

†東京大学, The University of Tokyo

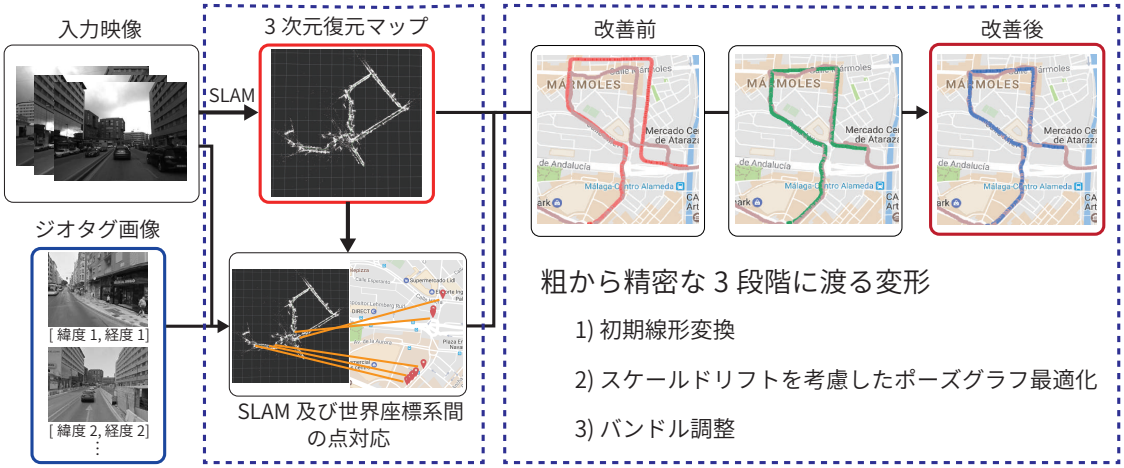


図 1: 提案手法の概要

3.2 ジオタグ画像を介した SLAM 及び世界座標系間の対応取得

ここでは，提案手法の前半部分である，ジオタグ画像を用いて SLAM 及び世界座標系間の三次元点の対応 $C_{\text{SLAM-world}}$ を取得する部分を説明する．我々は Agarwal らの提案した手法 [5] を，より大規模な環境でも適応できるよう拡張することでこれを実現する．章 3.2.1 では，使用した三次元復元手法の詳細を，章 3.2.2 では，ジオタグ画像を収集する方法の詳細を説明する．その上で，章 3.2.3 及び章 3.2.4 で $C_{\text{SLAM-world}}$ を取得する方法を説明する．

3.2.1 三次元復元

三次元復元には，単眼 SLAM の中で最も優れた手法の一つである ORB-SLAM [2] を使用する．ただし，本フレームワークは，他の特徴点を利用した SLAM にも適応可能な汎用的な手法であることに注意されたい．

ORB-SLAM では，三次元復元において重要なフレーム（映像を構成する画像）がキーフレームとして選択されるが，本手法の以降の処理ではキーフレームのみを扱うことで，効率良い処理を行う．三次元復元の際に取得された三次元特徴点と，それらに対応するキーフレーム上の観測点の対応を $C_{\text{fp-kf}}$ と定義する．

3.2.2 ジオタグ画像の収集

Google Street View [12] は，路上を対象とした検索可能な GIS であり，世界中を対象とした最も大規模なジオタグ画像データセットの中の一つである．全てのジオタグ画像は高解像度な RGB パノラマ画像と高精度な緯度経度のペアとして与えられる [13]．我々はそれぞれのパノラマ画像を，入力映像と同じ画角で水平等 8 方向に切り出すことで，ジオタグ画像群として実験に使用する．ジオタグ画像には世界座標が割り当てられているため，ジオタグ画像のカメラ位置が SLAM 座標系内で推定可能であれば，SLAM 及び世界座標間の対応を取得することができる．

3.2.3 画像検索及び特徴点マッチング

ここでは，キーフレームに対応するジオタグ画像を検索し，その上でキーフレームとジオタグ画像間の対応する特徴点の組を取得する．まず，全てのキーフレームに対して，類似度の高い k 枚のジオタグ画像を検索する．検索システムには SIFT 特徴量を用いた bag-of-words アプローチを利用した [5]．キーフレームとジオタグ画像の組に対して，ORB 特徴点 [14] の検出及び特徴点マッチングを行う．映像中のフレームと Google Street View の画像間の特徴点マッチングは誤対応を多く含む傾向があるため [15]，Virtual Line Descriptor（kVLD）[16] を用いて誤対応のマッチングを除去する．更に，キーフレームとジオタグ画像の組の中で，特徴点の対応が 5 つ未満の組を除去する．

3.2.4 ジオタグ画像のカメラ姿勢推定

$C_{\text{SLAM-world}}$ を取得するために，まず SLAM 地図中の三次元特徴点とそれらのジオタグ画像における観測点の対応 $C_{\text{fp-geo}}$ を取得する．$C_{\text{fp-geo}}$ は章 3.2.3 で取得した特徴点の対応と章 3.2.1 で取得した対応 $C_{\text{fp-kf}}$ を組み合わせることで取得する．$C_{\text{fp-geo}}$ を取得した後，$C_{\text{fp-geo}}$ の三次元特徴点をジオタグ画像に再投影した際の誤差を最小化することで，SLAM 座標系におけるジオタグ画像のカメラ姿勢を推定する．最終的に，SLAM 座標系におけるジオタグ画像のカメラ姿勢と，ジオタグ画像に紐付けられた世界座標の組を取得することで，$C_{\text{SLAM-world}}$ を得る．

3.3 初期線形変換（ILT）

ここでは，SLAM 及び世界座標系間の対応 $C_{\text{SLAM-world}}$ を用いて，三次元復元マップを世界座標系に粗に変形する手法を示す．SLAM 及び世界座標系における対応点の座標は大きく異なるため，この変形による粗な初期解を使用しなければ以降の高精度な変形手法は上手く働かない．

まず，全てのカメラ位置がおおよそ同一平面上に存在すると仮定し，その平面が xz 平面に一致するよう三次元マップを回転させる．その際に用いるカメラの乗

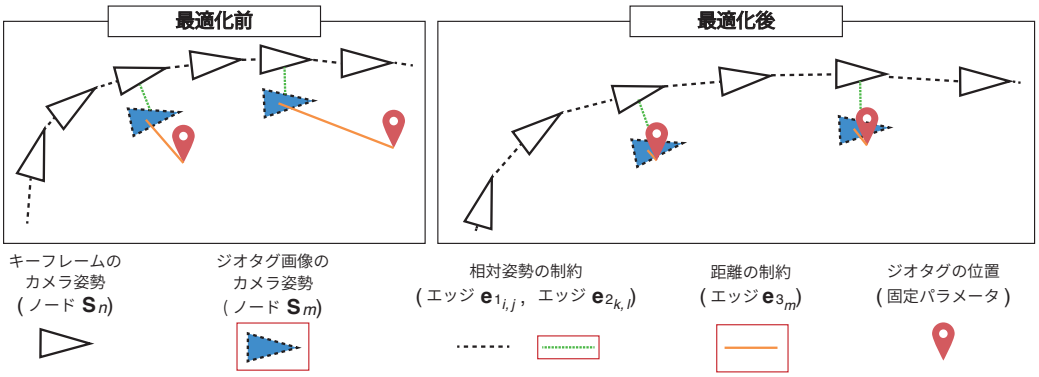


図 2: 提案するポーズグラフ最適化の概要図．新しくポーズグラフに追加したノードとエッジは赤枠で囲われている．

る平面は全カメラ位置の主成分分析により推定する．

次に，$C_{\text{SLAM-world}}$ における SLAM 座標上の点 $\mathbf{p}$ を，世界座標上の点 $\mathbf{p}'$ に変換する変換行列（式 1）を推定する．

$$\mathbf{p}' = \begin{bmatrix} s*\cos(\theta) & 0 & -s*\sin(\theta) & a \\ 0 & s & 0 & 1 \\ s*\sin(\theta) & 0 & s*\cos(\theta) & b \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{p} \quad (1)$$

変換行列中の 4 パラメータ $[a, b, s, \theta]$ は非線形最小二乗問題を RANSAC とレーベンバーグ・マルカート法を用いて解くことにより推定される．推定された変換行列により，キーフレーム，及びジオタグのカメラ姿勢や，三次元復元マップの三次元特徴点の位置を変換する．ここでの二種類の変換はどちらも三次元相似変換の一種であり，これらの変換によりスケールドリフトは改善されないことに注意されたい．

3.4 Sim(3) での制約付きポーズグラフ最適化（PGO）

ここではスケールドリフトを考慮した非線形変形を行うための新しいポーズグラフ最適化を提案する．それにより，元の三次元復元マップの構造を維持しつつスケールドリフトのみ改善する処理と，2 座標系間の対応点 $C_{\text{SLAM-world}}$ を近づける処理を同時に行うことが可能となる．図 2 に提案するポーズグラフ最適化の概要を示す．

表記． 三次元剛体変換 $\mathbf{G} \in \text{SE}(3)$ と三次元相似変換 $\mathbf{S} \in \text{Sim}(3)$ は式 2 のように定義される．ただし，$\mathbf{R} \in \text{SO}(3)$，$\mathbf{t} \in \mathbb{R}^3$，$s \in \mathbb{R}^+$ とする．SO(3)，SE(3)，Sim(3) はいずれもリー群に属しており，$\mathfrak{so}(3)$，$\mathfrak{se}(3)$，$\mathfrak{sim}(3)$ はそれぞれに対応するリー代数である．リー群は指数写像により対応するリー代数に変換され，また逆変換である対数写像も定義される．本稿では，リー代数を係数のベクトル表記によって記す．例を挙げると，$\mathfrak{sim}(3)$ は 7 次元のベクトル $(\omega_1, \omega_2, \omega_3, \sigma, \nu_1, \nu_2, \nu_3)^{\mathrm{T}}$ で表記され，その指数写像 $\exp_{\text{Sim}(3)}$ は式 3 のように定義される．ただし，$\mathbf{W}$ はロドリゲスの公式に似た形の項である．Sim(3) についての詳細は [3] を参照されたい．

$$\mathbf{G} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad (2)$$

$$\exp_{\text{Sim}(3)} \begin{pmatrix} \boldsymbol{\omega} \\ \sigma \\ \boldsymbol{\nu} \end{pmatrix} = \begin{bmatrix} e^{\sigma} \exp_{\text{SO}(3)}(\boldsymbol{\omega}) & \mathbf{W}\boldsymbol{\nu} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad (3)$$

提案するポーズグラフ最適化． 一般的に，6-DoF のカメラ姿勢やカメラ姿勢間の相対変換は SE(3) の要素として表現される．一方で，本最適化ではカメラ姿勢やその相対変換を Sim(3) の要素として扱う．SE(3) から Sim(3) への変換は，回転行列の R と並進ベクトルの $\mathbf{t}$ を変化させず，スケール成分の s を 1 とすることで行う．カメラ姿勢やその相対変換を Sim(3) の要素として扱うアイデアは，単眼 SLAM のスケールドリフトに対処するため Strasdat らにより提案された [3]．我々はこのアイデアを拡張し，ジオタグ画像のカメラ位置をジオタグの持つ世界座標に近づけることで，三次元復元マップのスケールドリフトを改善する手法を提案する．提案するポーズグラフ最適化は，以下の 2 種類のノードと 3 種類のエッジから構成される．

- ノード $\mathbf{S}_n \in \text{Sim}(3)$, $n \in \{1, 2...N\}$: n 番目のキーフレームのカメラ姿勢
- ノード $\mathbf{S}_m \in \text{Sim}(3)$, $m \in \{1, 2...M\}$: m 番目のジオタグ画像のカメラ姿勢
- エッジ $\mathbf{e}_{1_{i,j}}$, $(i, j) \in C_1$: i, j 番目のキーフレームのカメラ姿勢間の相対変換による制約 (式 4)
- エッジ $\mathbf{e}_{2_{k,l}}$, $(k, l) \in C_2$: k, l 番目のジオタグ画像のカメラ姿勢間の相対変換による制約 (式 5)
- エッジ $\mathbf{e}_{3_m}$, $m \in \{1, 2...M\}$: ジオタグ画像のカメラ姿勢 $\mathbf{S}_m$ と対応するジオタグの世界座標 $\mathbf{y}_m$ との距離 (式 7)

$$\mathbf{e}_{1_{i,j}} = \log_{\text{Sim}(3)}(\Delta\mathbf{S}_{i,j} \cdot \mathbf{S}_i \cdot \mathbf{S}_j) \in \mathbb{R}^7 \quad (4)$$

$$\mathbf{e}_{2_{k,l}} = \log_{\text{Sim}(3)}(\Delta\mathbf{S}_{k,l} \cdot \mathbf{S}_k \cdot \mathbf{S}_l) \in \mathbb{R}^7 \quad (5)$$

$$\begin{aligned}\mathbf{e}_{3_m} &= -\frac{1}{s_m}\mathbf{R}_m^{\mathrm{T}}\mathbf{t}_m - \mathbf{y}_m \\ &= -e^{\sigma_m}\exp_{\mathrm{SO}(3)}(\boldsymbol{\omega}_m)^{\mathrm{T}}\mathbf{W}_m\boldsymbol{\nu}_m - \mathbf{y}_m \\ &\in \mathbb{R}^3\end{aligned} \tag{6}$$

ただし，N はキーフレームの総数，M はキーフレームと対応を持つジオタグ画像の総数である．また，C_1 は三次元復元において，同一の三次元特徴点を観測しているキーフレームの組であり，C_2 はキーフレームと対応するジオタグ画像の組である．$\Delta\mathbf{S}_{i,j}$ は，最適化前の $\mathbf{S}_i$ と $\mathbf{S}_j$ の間の相対変換を Sim(3) に変換したものであり，この値は最適化の間固定される．

我々は Strasdat らのポーズグラフに新たにノード $\mathbf{S}_m$，エッジ $\mathbf{e}_{2_{k,l}}$，エッジ $\mathbf{e}_{3_m}$ を追加した．$\mathbf{e}_{1_{i,j}}$ 及び $\mathbf{e}_{2_{k,l}}$ の最小化は，ゆるやかなスケールの変化を除いてカメラ姿勢間の相対変換の変化を抑えるよう働く．また，$\mathbf{e}_{3_m}$ の最小化は，ジオタグ画像のカメラ位置をジオタグの持つ世界座標に近づけるよう働く．提案するポーズグラフ最適化のコスト関数は以下の通りである．

$$\begin{aligned}E(\mathbf{S}_1, \mathbf{S}_2, \ldots \mathbf{S}_{N+M}) = \lambda_1 &\sum_{(i,j)\in C_1} \mathbf{e}_{1_{i,j}}^{\mathrm{T}}\mathbf{e}_{1_{i,j}} \\ + \lambda_2 &\sum_{(k,l)\in C_2} \mathbf{e}_{2_{k,l}}^{\mathrm{T}}\mathbf{e}_{2_{k,l}} \\ + \lambda_3 &\sum_{m\in\{1,2\ldots M\}} \mathbf{e}_{3_m}^{\mathrm{T}}\mathbf{e}_{3_m}\end{aligned} \tag{7}$$

このコスト関数をレーベンバーグ・マルカート法で最小化することにより，キーフレーム及びジオタグ画像のカメラ姿勢 $\mathbf{S}_n$，$\mathbf{S}_n$ を推定する．更に，この最適化による変換を三次元復元マップの三次元特徴点の位置にも反映させる [3]．

3.5 バンドル調整（BA）

最後に，ジオタグ画像との制約を含めたバンドル調整により三次元復元マップの変形を行う．ジオタグ画像との制約を組み合わせるため，一般的な $C_{\text{fp-kf}}$ の再投影誤差だけでなく，$C_{\text{fp-geo}}$ の再投影誤差も合わせて最小化する．ただしバンドル調整の間，ジオタグ画像のカメラ姿勢は固定する．この変形は十分に良い初期解が与えられれば，更に高精度に三次元復元マップを変形しうる．

4 実験

本章では実データを用いて提案手法により改善される精度を定量評価する．また，提案手法中の 3 段階に渡る変形手法について，それぞれの精度への影響及び有用性を検証する．

4.1 データセット及び実装詳細

本実験では The Málaga Stereo and Laser Urban Data Set （Málaga データセット） [17] というスペインの都心部を長距離に渡って撮影した走行映像データセットを用いた．Málaga データセットの映像は解像度が 1024 × 768，フレームレートが 20 fps となっている．我々は，その映像から 2 種類の映像（video 1, 2）を切り出し評価に用いた．2 種類の映像はどちらもループを描かず，軌跡長は 1 km 以上である．全てのフレームには一秒ごとに取得された GPS の位置情報が関連付けされているが，10 m 以上誤差を含む場合もしばしば観測された．

表 1: 提案手法による改善結果．

	video 1		video 2	
Method	Ave [m]	SD	Ave [m]	SD
Baseline	56.1	45.1	36.8	57.6
Ours	**5.7**	2.4	**6.7**	0.1

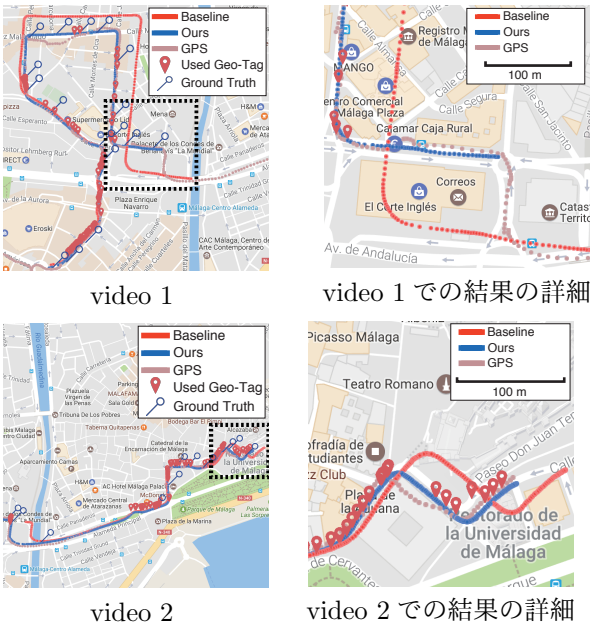


図 3: 提案手法による改善結果（Google Map 上に表示）．

4.2 評価指標

定量的に比較するため，後述の Ground Truth（GT）の位置と，GT の付いたキーフレームのカメラ位置の距離（メートル単位）の平均（Ave）及び標準偏差（SD）を評価指標として用いる．今回付与されていた GPS の位置情報は正確さに欠けたため，Google Street View の三次元地図や走行映像を参考に，手動で幾つかのキーフレームに GT として絶対位置を割り当てた．

4.3 提案手法全体の評価

我々は，提案手法の有用性を検証するため，2 つの走行映像を用いて手法の適応前（Baseline）と適応後（Ours）のカメラ位置の誤差を比較した．適応前のカメラ位置には，三次元復元の結果を相似変換により世界座標上に割り当てたカメラ位置を使用した．表 1 には数値評価の結果を，図 3 には変換前後のキーフレームの軌跡を Google Street View 上に可視化したものを示す．図 3 から明らかなように，Baseline にはスケール

表 2: 3 種類の変形手法の様々な組み合わせにおける改善精度の比較（video 1 での結果）

	ILT	PGO	BA	Ave [m]	SD
#1		✓		*	*
#2	✓			56.1	45.1
#3	✓		✓	18.5	1.9
#4	✓	✓		9.0	4.7
Ours	✓	✓	✓	**5.7**	2.4

誤差が蓄積し，全体では 30m 以上もの誤差が生じていることが分かる．これは，映像のカメラ位置の軌跡が 1km 以上もの長距離であるにも関わらず，ループが存在しないためスケールドリフトを解消できなかったためである．一方で我々の手法は適切にスケールドリフトを改善し，三次元復元の結果を十分に改善した．

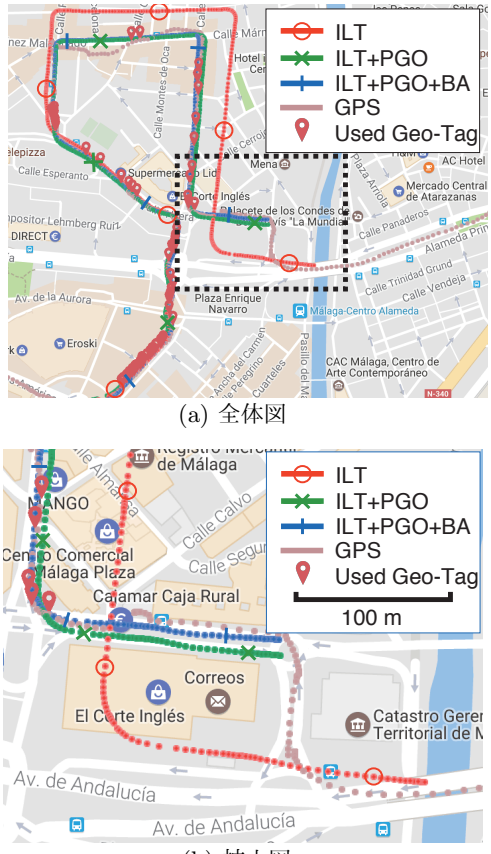
4.4 3 段階の変形手法の影響分析

ここでは，提案手法中の 3 段階に渡る変形手法について，それぞれの改善精度への影響，及び必要性を検証する．そのために，3 段階の変形手法の様々な組み合わせを video1 に適応し，カメラ位置の誤差を比較した．表 2 と図 4 に結果を示す．*は意味のある値が得られなかったことを示す．表 2 より，全ての変形手法を適応した場合に最も高精度な位置推定が実現されていることが分かる．図 4 は結果の一部を Google Map 上に表示したものである．線形変換（ILT）のみの場合は 100 m 程の誤差が生じているが，ポーズグラフ最適化（PGO）によりスケールドリフトが解消され，その結果最終的にバンドル調整（BA）が上手く働いているのが確認できる．

5 考察

本稿では，三次元復元結果を改善するため，ジオタグ付き画像を活用するフレームワークを提案した．今回用いた映像のように，1 km を超える長距離撮影映像で，且つ映像の軌跡がループを描かない場合は，三次元復元の結果に実スケールで数 10 m 程の誤差が生じる場合がある（表 1 の Baseline を参照）．これは，一般的な SLAM 等の映像からの三次元復元では，映像の軌跡がループを描いた（同一の場所を再度観測した）場合にのみしか誤差の蓄積を改善できないためである．しかし，本手法では，スケールドリフトを改善する処理を統合することで，絶対位置推定に対して三次元復元の構造を適切に利用することが可能となった．しかし，現状では 6 m 程の誤差が残っているためその原因を考察する．

1 つ目の原因として，GT の位置が正確でないことが挙げられる．本稿では GPS の精度が十分でないため，より高精度な GT を目視により手動で付けたが，1 m 程の誤差が含まれていると考えられる．2 つ目の原因として，Google Street View から取得したジオタグ画像に紐付いている緯度経度に誤差が含まれていることが挙げられる．我々は使用した映像を精査する中で，最大 3 m 程の誤差が生じている場所があることを確認した．3 つ目の原因として，映像とジオタグ画像で共有する特徴点が遠距離に存在し且つ少量であるという問題が挙げられる．これは，特徴点マッチングの精度が十分でない点や，カメラが車道に並行な方向を向いている点，カメラの画角が狭い点などに起因する．その結果，章 3.2.4 でジオタグのカメラ姿勢を推定する際，カメラから特徴点群へ向かう方向の誤差が大きく生じてしまう問題が生じる．推定結果と GT の位置関係を目視で確認すると，誤差は主に車道に並行な向きに生じていることを確認した．



(a) 全体図

(b) 拡大図

図 4: 3 種類の変形手法の一部の組み合わせにおける改善精度の比較（video1 の結果を Google Map 上に表示）．

以上の状況を踏まえると，まずは高精密な手法評価のため，GT 及び評価手法の改善・再検討が不可欠と考えられる．また，Google Street View の精度に関しては関与できないため，適応可能範囲が限定的にはなるが，自らより高精度なジオタグ付き画像データセットを生成する方向性も考えられる．三つ目の原因に取り組むためには，特徴点マッチングの手法の改良が有用であると考えられる．

6 まとめ

本稿では，三次元復元結果を改善するために，Google Street View 等から取得したジオタグ付き画像を活用するフレームワークを初めて提案した．まず，ジオタグ画像を介して三次元復元の座標系及び世界座標系間の

疎な点対応を取得する．そして，その対応を元に 3 段階の変形を適用することで，三次元復元における蓄積誤差を解消し高精度な三次元地図を得る．特に，単眼カメラからの三次元復元において特徴的な蓄積誤差であるスケールドリフトに対処するため，我々は 2 段階目の変形手法にあたるスケールドリフトを考慮したポーズグラフ最適化を提案した．

長距離を撮影した実映像を使用した実験を通して，3 段階からなる変形手法それぞれの必要性，及び本フレームワークが十分に三次元復元結果を改善可能であることを検証した．

参考文献

[1] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-scale direct monocular slam,” ECCV, pp.834–849, 2014.

[2] R. Mur-Artal, J.M.M. Montiel, and J.D. Tardos, “ORB-SLAM: a versatile and accurate monocular slam system,” IEEE Transactions on Robotics, vol.31, no.5, pp.1147–1163, 2015.

[3] H. Strasdat, J. Montiel, and A.J. Davison, “Scale drift-aware large scale monocular slam,” Robotics: Science and Systems IV, pp.73–80, 2010.

[4] M. Tamaazousti, V. Gay-Bellile, S.N. Collette, S. Bourgeois, and M. Dhome, “Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment,” CVPR, pp.3073–3080, 2011.

[5] P. Agarwal, W. Burgard, and L. Spinello, “Metric localization using google street view,” IROS, pp.3111–3118, 2015.

[6] M. Lhuillier, “Incremental fusion of structure-from-motion and gps using constrained bundle adjustments,” TPAMI, vol.34, no.12, pp.2489–2495, 2012.

[7] F. Lu and E. Milios, “Globally consistent range scan alignment for environment mapping,” Autonomous robots, vol.4, no.4, pp.333–349, 1997.

[8] E. Olson, J. Leonard, and S. Teller, “Fast iterative alignment of pose graphs with poor initial estimates,” Robotics and Automation, pp.2262–2269, 2006.

[9] G. Grisetti, R. Kümmerle, C. Stachniss, U. Frese, and C. Hertzberg, “Hierarchical optimization on manifolds for online 2d and 3d mapping,” ICRA, pp.273–278, 2010.

[10] G. Klein and D. Murray, “Parallel tracking and mapping for small ar workspaces,” ISMAR, pp.225–234, 2007.

[11] P. Lothe, S. Bourgeois, F. Dekeyser, E. Royer, and M. Dhome, “Towards geographical referencing of monocular slam reconstruction using 3d city models: Application to real-time accurate vision-based localization,” CVPR, pp.2882–2889, 2009.

[12] “Google Street View”. `https://www.google.com/streetview/`.

[13] B. Klingner, D. Martin, and J. Roseborough, “Street view motion-from-structure-from-motion,” ICCV, pp.953–960, 2013.

[14] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: An efficient alternative to sift or surf,” ICCV, pp.2564–2571, 2011.

[15] A.L. Majdik, Y. Albers-Schoenberg, and D. Scaramuzza, “Mav urban localization from google street view data,” IROS, pp.3979–3986, 2013.

[16] Z. Liu and R. Marlet, “Virtual line descriptor and semi-local matching method for reliable feature correspondence,” BMVC, pp.16–1, 2012.

[17] J.-L. Blanco-Claraco, F.-Á. Moreno-Dueñas, and J. González-Jiménez, “The málaga urban dataset: High-rate stereo and lidar in a realistic urban scenario,” The International Journal of Robotics Research, vol.33, no.2, pp.207–214, 2014.