

視野を共有しない車載カメラ間のvSLAMを用いた外部パラメータ校正

西口 和希^{1,a)} 内山 英昭¹ 早川 和孝² 足立 淳² トマ ディエゴ¹ 島田 敬士¹ 谷口 倫一郎¹

概要: 全周囲を観測可能な複数の車載カメラのように、視野を共有しないカメラ間の外部パラメータ校正は、専用の校正治具または校正環境を用いて行われている。一般に、このような校正はカメラを車体に固定後や車検時のみに行われるが、経年劣化や車体振動などによってカメラ間の位置変化が発生するため、高頻度で校正を行うことが望ましい。そこで本稿では、前後左右に設置された車載カメラ間の外部パラメータ校正問題に対し、各カメラが共通するシーンを撮影可能な切り返しを伴う後退入庫の動作に着目し、専用の校正環境を必要としない未知環境における外部パラメータ校正手法を提案する。提案手法では、撮影のタイミングが同期された複数カメラの画像列を入力とし、初めに、各カメラで独立にキーフレームに基づく単眼カメラの Visual SLAM(vSLAM)を行うことで、マップを生成する。次に、各マップのキーフレームに対し、画像検索に基づいて各マップ間で共通するシーンを撮影したキーフレームを検出することで、独立した全てのマップを1つに統合する。最後に、カメラ間の外部パラメータ校正のために、各カメラから複数の同期した画像を選択し、カメラ間の位置姿勢関係が固定であることを制約としたバンドル調整を行うことで、高精度にカメラ間の外部パラメータを算出する。実験では、2台のカメラ間の校正精度を定量的に評価した。

1. はじめに

車の周辺環境の認識や可視化を通じて運転支援を行うために、カメラを搭載する車が普及しつつある。例えば、一般の乗用車においても、駐車時に上部視点から見た俯瞰画像を生成して車の位置を可視化するシステムが実用的に用いられている。このような可視化を実現するためには、車体に設置された複数カメラ間の位置姿勢関係を表す外部パラメータが既知である必要がある。一般には、車検等の機会に専用の校正機器を用いてカメラ間の外部パラメータの校正が行われる。しかし、校正後に車体の振動や経年劣化などにより校正された位置からずれを生じる場合、誤認識や誤った可視化となる問題を生じる。安全面の観点から、外部パラメータに変化が生じていることを検出するためには、頻繁に校正が行われることが好ましい。しかし、従来の校正手法は、専用の機器や環境を必要とすることから、車の利用者が頻繁に校正を行うのは困難である [1], [2], [3], [4], [5]。より安全性の高い運転支援を行うためには、専用の機器や環境を必要とすることなく外部パラメータの校正を行うことのできる枠組みが必要となる。

未知環境下におけるカメラの外部パラメータの算出手法として、Visual SLAM(vSLAM)が挙げられる。vSLAMとは、カメラで撮影された画像列から未知環境の構造を環境地図(マップ)として復元し、同時にカメラの位置姿勢を算出する手法である [6]。特に、入力画像列の一部をキーフレームとして選択することでマップの高精度化のためのバンドル調整を行う手法が用いられている [7]。vSLAMを用いた場合、各カメラで構築したマップ内にシーンの重複が存在すれば、重複部分の構造を用いてマップを統合することで、各カメラの位置姿勢を同一座標系で算出可能となる。すなわち、各カメラのvSLAMで生成されるキーフレーム間の位置姿勢関係を同一座標系で表現することで、車載カメラのような視野を共有しないカメラにおいても、外部パラメータ校正を行うことができる。

vSLAMを用いた視野を共有しないカメラの外部パラメータ校正手法として、文献 [8]が挙げられる。この文献では、ロボットを想定した円形の台にカメラを固定し、台の中心を軸として回転させながらvSLAMを実行する。次に、各カメラで作成されるマップをキーフレーム間のマッチングに基づいて一致させ、マップ間の相似変換行列を計算することで、カメラ間の外部パラメータを算出する。しかし、この手法では、カメラの配置や動かし方に制約があるため、車載カメラの外部パラメータ校正に適用すること

¹ 九州大学

² アイシン精機株式会社

^{a)} nishiguchi@limu.ait.kyushu-u.ac.jp

は難しい。また、文献 [9] では、各カメラで一定間隔の画像をキーフレームを同期させて vSLAM させているが、カメラから空間までの距離を考慮せずにキーフレームを選択しているため、屋外の実環境では vSLAM がスケールドリフトする。このため、精度が大きく低下する。

本稿では、車の前後左右に設置された車載カメラを対象とし、屋外の未知環境における車載カメラ間の外部パラメータ校正手法を提案する。特に、図 1 に示すように、共通するシーンを撮影可能な切り返しを伴う後退入庫の動作を利用する。入力として、撮影のタイミングが同期された複数のカメラの画像列を取得する。初めに、各カメラで独立にキーフレームに基づく単眼カメラの vSLAM を行うことで、マップを生成する。ここで、各カメラから撮影されるシーンまでの距離が不定であるため、各カメラのキーフレームは非同期である。次に、各マップのキーフレームに対し、特徴点を用いた画像検索に基づいて共通するシーンの検出および特徴点マッチングを行うことで、独立した各マップを 1 つに統合する。最後に、カメラ間の外部パラメータ校正のために、各カメラから複数の同期した画像を選択し、カメラ間の位置姿勢関係が固定であることを制約としたバンドル調整を行うことで、高精度にカメラ間の外部パラメータを算出する。実験では、2 台のカメラ間の校正精度を定量的に評価した。

2. 提案手法

2.1 概要

vSLAM を利用して視野を共有しないカメラの外部校正を行うためには各カメラで作成されるマップに構造の重複が不可欠である。すなわち、各カメラで同一シーンが撮影されていることが前提となる。そこで我々は、車に無理のない動作として短い時間で同一シーンを撮影することが可能な駐車動作に着目した。具体的には、各カメラにシーンの共有が生じるような切り返しを伴う後退入庫の動作に着目し、車の駐車時の画像列を入力としてカメラ間の外部パラメータの校正を行う。

提案手法は、図 2 に示すように、次の 3 つの処理に分けられる。

vSLAM 前後左右の各カメラから、駐車動作中に撮影された画像列を入力とし、各カメラで独立に vSLAM を実行することで、キーフレームとマップを生成する。

マップの統合 vSLAM によって得られた各カメラのキーフレームを入力とし、画像検索に基づいて共通するシーンを撮影したキーフレームを探索し、さらに特徴点マッチングを行うことで、各マップで共通する特徴点の対応付けを行う。最後に、特徴点の対応付けから各マップ間の相似変換行列を算出することで、全カメラのマップを同一座標系に統合したマップを出力する。

外部パラメータ校正 統合されたマップで、異なるカメラ

の位置姿勢を算出することで、カメラ間の外部パラメータの校正結果として出力する。

2.2 vSLAM

vSLAM では、画像上に映る特徴点の三角測量によって環境の 3 次元構造、すなわちマップが再構成される。この際、三角測量のために使用される画像とそのカメラの位置姿勢を総称してキーフレームと呼ぶ。高精度なマップを生成するためには、ベースラインと呼ばれるキーフレームの間隔を適切に設定する必要がある。例えば、カメラから対象物体の距離が遠い場合に、キーフレームの間隔を短くしてしまうと、復元される点の誤差の蓄積が大きくなりスケールドリフトを引き起こす要因となる。従来研究では、キーフレーム選択の際にカメラからシーンまでの距離が考量されておらず、一定間隔で同期させたキーフレームを利用していた [8], [9]。しかし、前後左右に設置された車載カメラでは、各カメラからカメラに映るシーンまでの距離が異なる場合が多い。このため、従来手法の vSLAM の枠組みでは、外部パラメータの校正精度が大きく低下する。

そこで、キーフレームを用いた vSLAM の一つである ORB-SLAM [7] にて用いられているキーフレーム選択手法を利用する。具体的には、特徴点のトラッキングができていない点数の変化率に対してしきい値処理を行う。これは、カメラからシーンが近い場合には特徴点の移動が大きくなり、遠い場合には小さくなるためである。図 3 に示すように、車載カメラではカメラによってシーンまでの距離が異なるため、上述のキーフレーム選択基準を利用することで、シーンの深度に応じて異なる間隔でキーフレームが挿入されていることが確認できる。

2.3 マップの統合

従来手法では、校正機器の回転動作によって異なるカメラ間での共通するシーンの観測が保証されていたのに対し、提案手法では車の駐車動作によりこれを保証している。そこで、初めに、各カメラで共通シーンを撮影したキーフレームを探索するために、特徴点の Bag of Visual Words(BoVW) を用いた画像検索を利用する [10]。各マップのキーフレームに対し、BoVW を用いて特徴量を算出し、各マップのキーフレームのマッチングを行う。次に、特徴量の近い上位 N 個のマッチングされたキーフレームを共通シーンを候補として検出する。さらに、特徴点マッチングを用いて、2 つのキーフレームが共通シーンを撮影しているかどうかを判別するとともに、マップの統合を行う。前後左右に設置された車載カメラで撮影される後退入庫時の画像列では、撮影される共通のシーンにおいて視点変化が生じやすく、従来研究と比較して、特徴量が大きく変化しやすいため、しきい値を用いた特徴点マッチングが困難になる。このため、本研究では、特徴点マッチングの

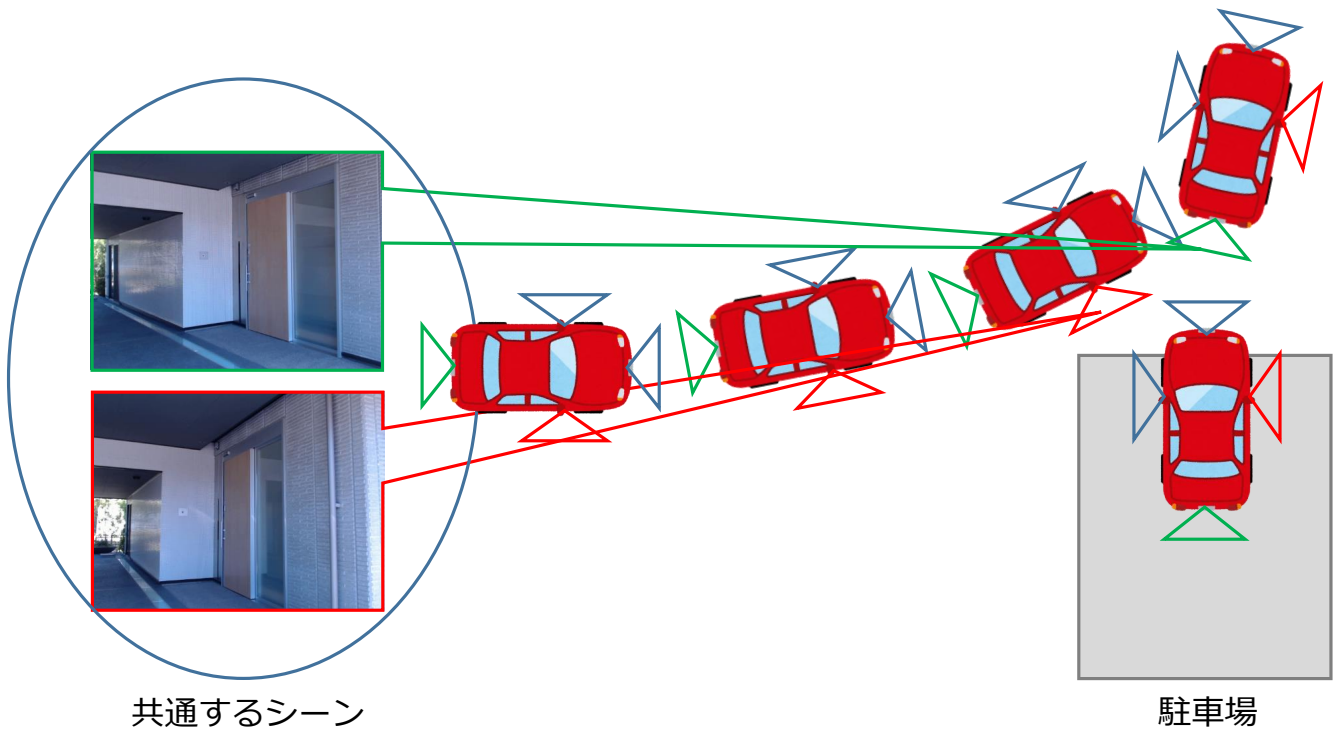


図 1: 切り返しを伴う後退入庫

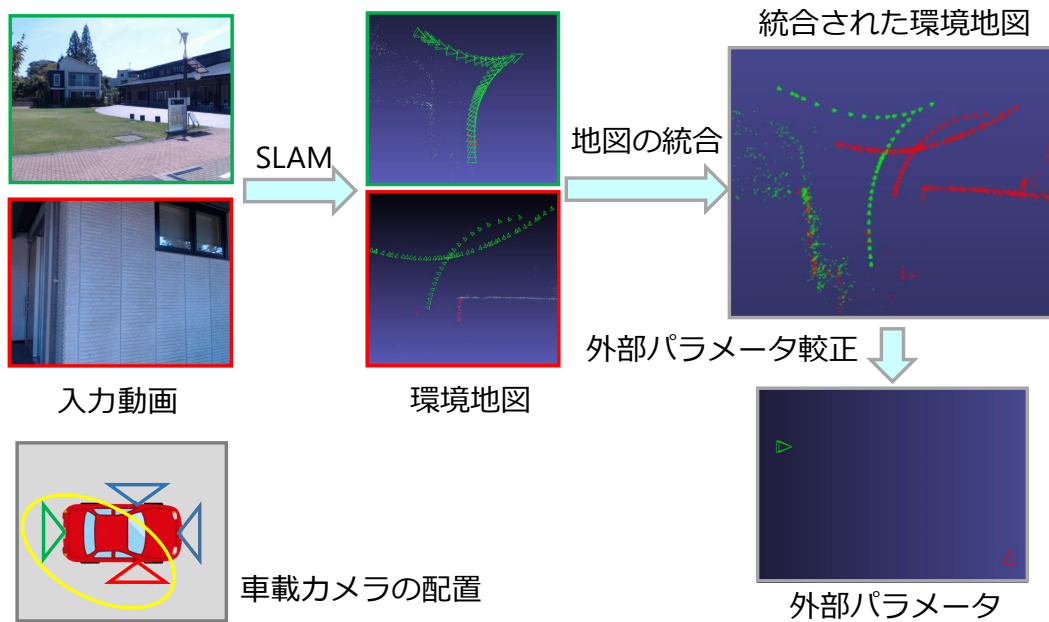
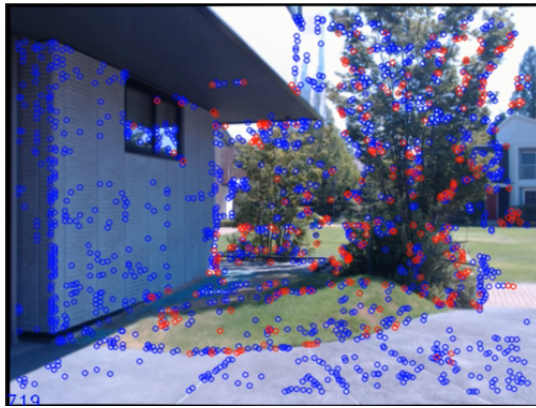


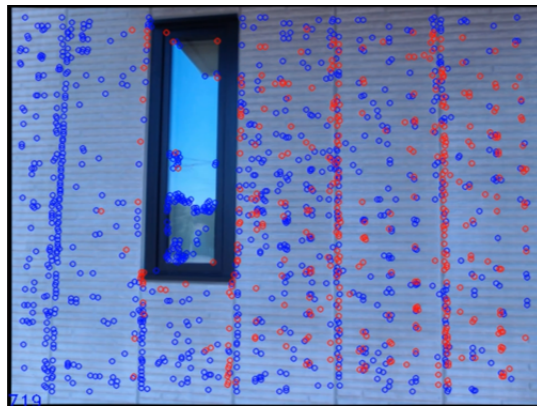
図 2: 提案手法の概要

際に同一特徴と認める特徴量のしきい値を下げることで、より多くの対応点を初めに算出する。しかし、しきい値を下げることは特徴点の誤対応の増加につながるため、誤対応除去の枠組みとして、GMS を利用する [11]。GMS では、マッチングの近傍領域の対応点数によってそのマッチングを評価し、誤対応を除去している。この手法により、マッチング時の特徴量のしきい値処理への依存度を低減できるため、視点変化に対する頑健性を向上させた。

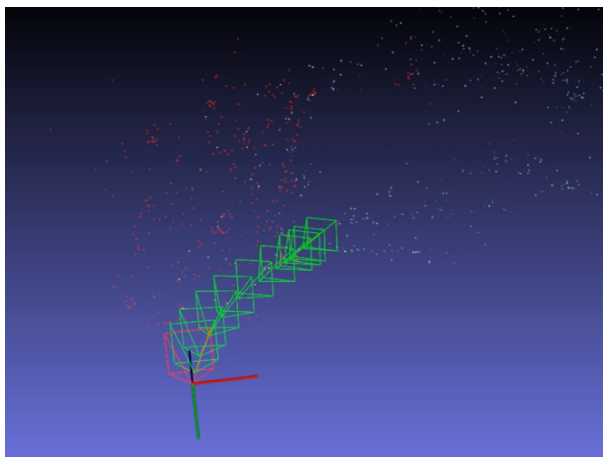
最後に、特徴点マッチングにより得られた対応点を用いて相似変換行列の計算を行う。各カメラの vSLAM は独立に実行されているため、単眼カメラの vSLAM の特徴として、スケールの不定性がある。これは、キーフレームを用いた単眼 vSLAM では、最初の 2 枚の画像の距離を 1 とするスケールとなるため、各マップの単位が異なっているという意味である。このため、一般的に用いられる同一の単位間の変換を表す剛体変換行列を用いることはできな



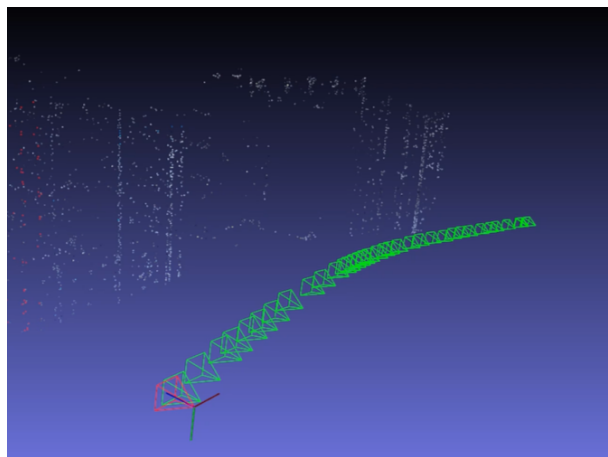
(a) 後ろカメラ



(b) 右カメラ



(c) 後ろカメラのキーフレーム



(d) 右カメラのキーフレーム

図 3: 車載カメラから撮影したシーンにおける vSLAM の例

い。そこで、スケールの違いも含む相似変換行列 (Sim3) を算出する。マップ間で対応付けられたキーフレームの特徴点に対し、各マップの 3 次元座標を対応付ける。次に、umeyama 法を利用して、マップ間の Sim3 を算出する [12]。umeyama 法を用いる際に注意すべき点として、誤対応が含まれる場合に精度が大きく低下することがある。そのため、誤対応除去を行う枠組みとして RANSAC を利用した。RANSAC を用いた Sim3 の算出では、Sim3 が計算可能となる最小点数である 3 点を入力された対応点をランダムに選択して Sim3 を算出し、選択された対応点以外を変換した際の誤差を評価する。Sim3 を算出することで、スケールが統一された 1 つのマップを作成できる。

RANSAC を用いた Sim3 の算出後の対応点数は十数点程度と少ない場合が多い。これは前述したように、後退入庫時の画像列では共通するシーンを撮影した場合にも視点変化を生じているため、特徴量に基づく特徴点マッチングのみからは、得られる対応点数が少ないためである。そこで、Sim3 によるマップを統合後に対応点を増加させる処理を行うことで、より高精度な統合を実現する。ここで、マップ 1 のキーフレームを K1、画像検索及び特徴点マッ

ングによって K1 に対応付けられたマップ 2 のキーフレームを K2 とする。初めに、マップ 2 の中で、K2 の近傍にあるキーフレームを選択する。これは K2 に映るマップ点があるマップ 2 内の他のキーフレームである。次に、そのキーフレームと K1 の間で、特徴点マッチングを行う。これにより K2 に映っていない特徴点が K1 に映っている場合に対応付けることができ、マップ間の対応点を増加させることができる。この対応付けは、Sim3 によるマップ統合後に行われるため、全てのマップが 1 つの座標系に統一されている。したがって、ORB-SLAM における投影に基づく特徴点マッチングを利用した [7]。

各マップを統合し、さらに対応点を増加させたのち、バンドル調整を用いてマップとキーフレームの最適化を行う。これにより、Sim3 による相似変換のみでは解決が困難な各マップ内に存在する軽微なスケールドリフトや誤差の蓄積による位置のずれの解消が見込まれる。図 4 に、マップ統合の結果の例を示す。図 4(a) の相似変換のみによるマップの統合結果に対して、(b) のバンドル調整後の結果では 2 つのカメラの軌跡が平行な軌道となっていることが確認できる。

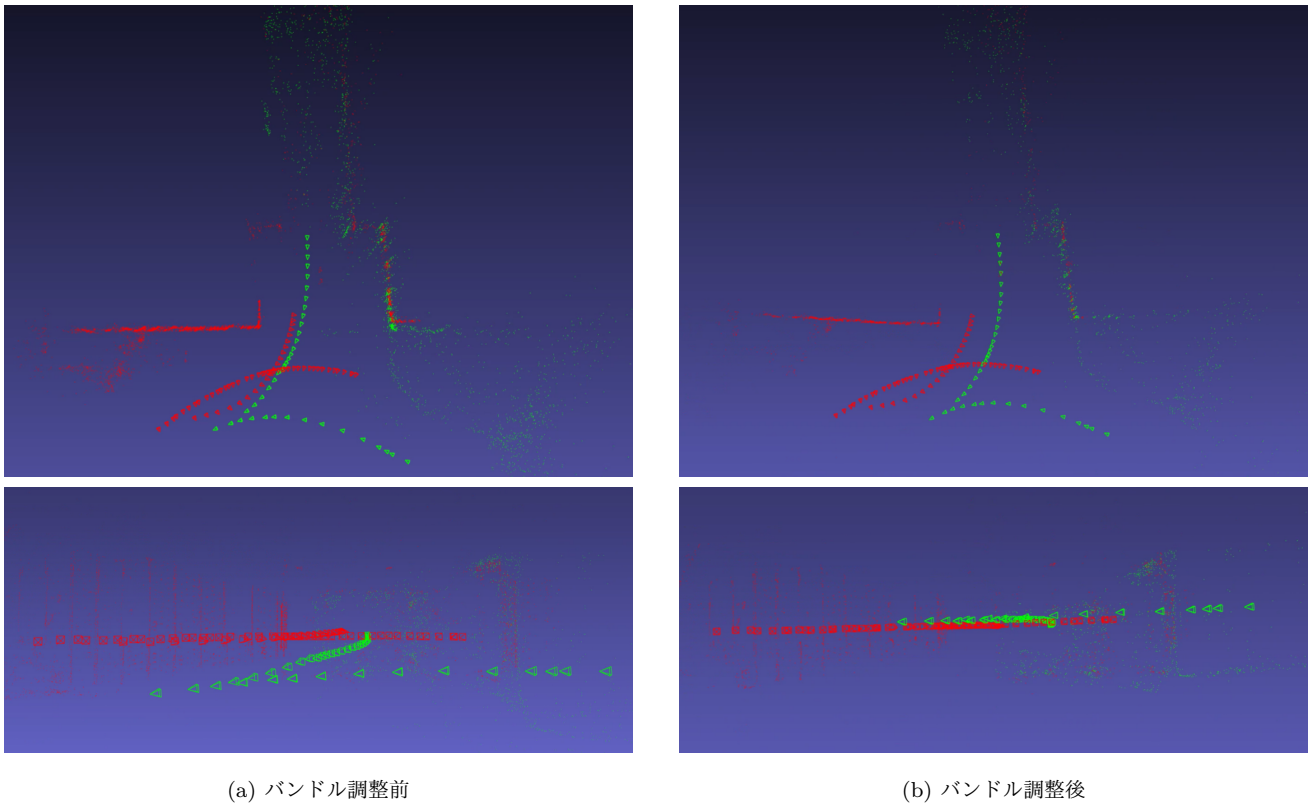


図 4: 統合された環境地図 ((a),(b) いずれも後側のカメラで撮影された点群・カメラの軌跡を緑, 同様に右側を赤で示す)

2.4 外部パラメータの較正

マップ統合を行うことで、各カメラのマップ点及びキーフレームが、1つの座標系で表現された。これにより、統合されたマップ上で、各カメラ間の位置姿勢を算出することで、外部パラメータの較正が可能となる。従来手法では各カメラでのキーフレームを同期させていたため、キーフレームの位置関係を較正結果として利用した。それに対し、提案手法では各カメラでキーフレーム選択を独立させているため、従来手法で行われていたキーフレームの位置姿勢関係の算出では較正ができない。

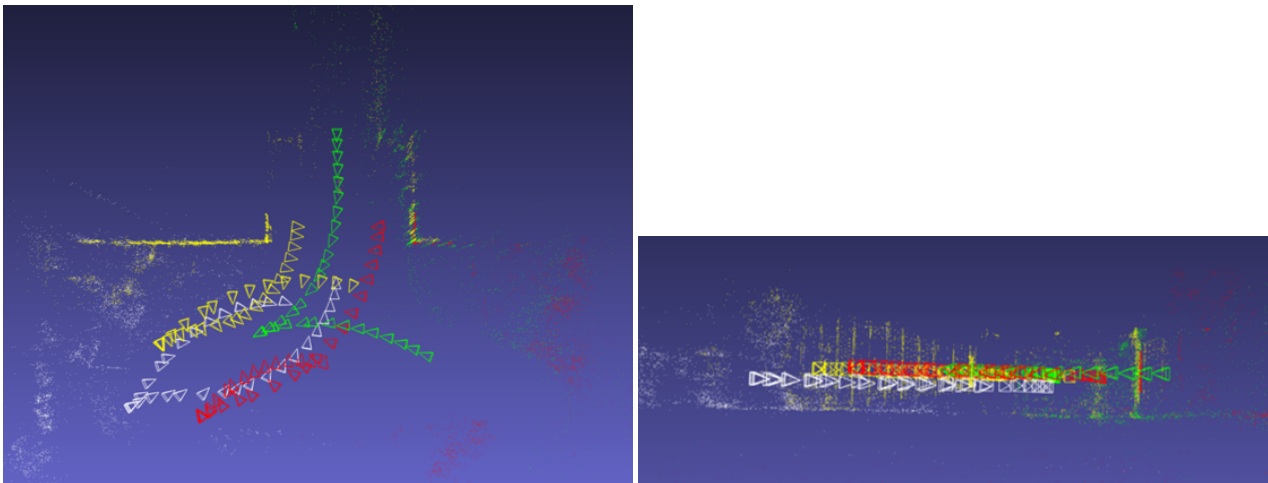
そこで、外部パラメータの較正を行うために、キーフレームとは別に新たに各カメラ間で同期した画像を選択する。この画像を同期フレームと呼ぶ。同期フレームはvSLAMを実行中に固定長の時間間隔で選択され、その時点の特徴点とカメラの位置姿勢が保存される。vSLAMの実行中やマップ統合の際に行われるバンドル調整によってキーフレームの位置が変化するのに対し、保存された同期フレームはそれらの処理に含まれていない。これは、同期フレームを種々の最適化に含めた場合、通常のキーフレームとの区別が計算処理上不可能になってしまうため、適切な間隔で選択されたベースラインの関係が破綻する恐れがあるためである。したがって、同期フレームを保存した時点でのカメラの位置姿勢を比較するだけでは外部パラメータの較正を行うことができない。そこで、マップ統合後のマップ

に対して、改めてカメラの位置姿勢を算出する操作が必要となる。これにより、外部パラメータの較正が可能となる。

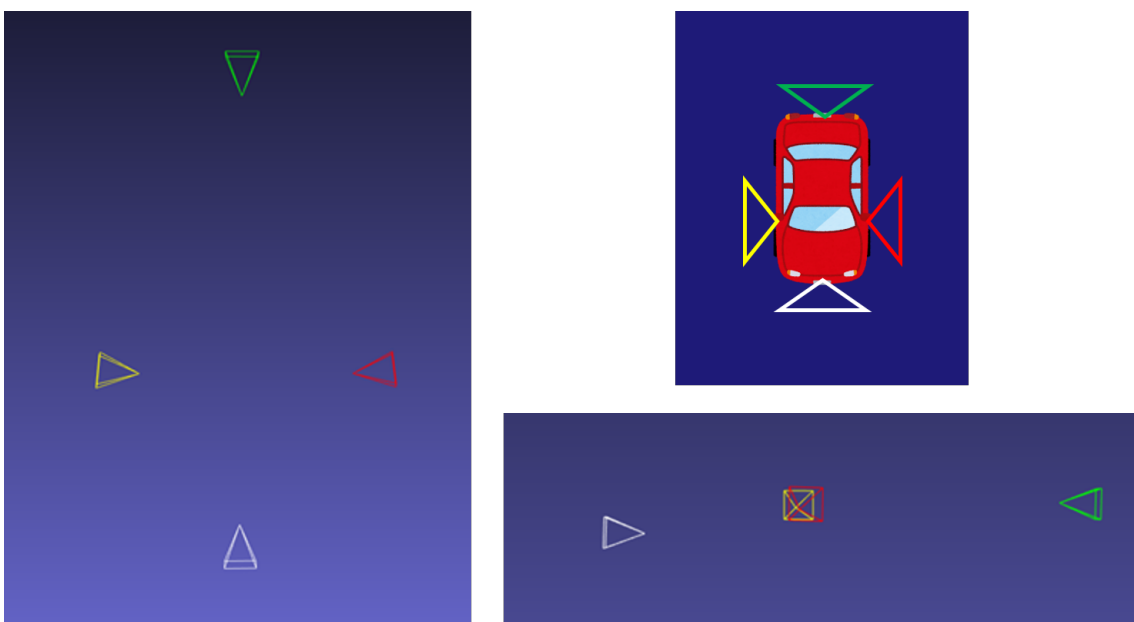
マップ統合の際に行ったバンドル調整では、各カメラ間の位置姿勢関係が固定である制約が組み込まれていない。これは、各カメラのキーフレームが非同期で選択されていたためである。ここで導入した同期フレームは、各カメラで同一時刻に撮影した画像であるため、これらのカメラの位置姿勢は常に一定となる。そこで、カメラ間の位置姿勢関係を共通する変数としたバンドル調整を行う。具体的には、各カメラのうち1つを基準として他のカメラポーズを基準カメラからの相対ポーズとして表現することによって、基準カメラ以外のカメラポーズを、同期した基準カメラのカメラポーズと相対ポーズによるパラメータ表現に変換する。このパラメータにおける各カメラポーズで再投影誤差を最小化することによって、通常のバンドル調整同様最適化が実現される。図5に、前後左右4台のカメラで撮影したマップの統合結果と外部パラメータの較正結果を示す。

3. 実験

定量的に外部パラメータの較正精度を検証する。ここでは、図6に示すように、台車に固定した視野を共有しない2台のカメラを利用する。外部パラメータの真値を取得するために、隣接するカメラとの間に視野の共有が生じるように追加のカメラを設置し、各カメラ間の外部パラメータを



(a) 統合された環境地図



(b) 外部パラメータ

図 5: 4 台のカメラを使用した場合の結果

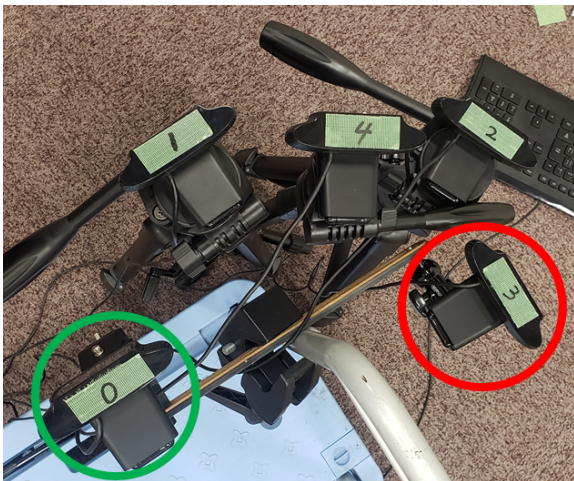
チェスボードを用いて算出することで、2 台のカメラ間の外部パラメータを間接的に算出した。表 1 に真値を示す。

較正した 2 台のカメラを用いて、屋内の実験環境において、計 8 シーンを撮影し、精度評価を行った。外部パラメータの精度評価には、文献 [13] の指標を利用した。外部パラメータは回転と並進、という単位の異なる 2 つの成分から構成されているため、それぞれの成分を独立に評価する場合が多い。文献 [13] では、真値の逆変換と推定された外部パラメータの積を評価行列として基準姿勢からのずれとして算出する。ただし、評価に際して真値と推定された値のスケールが一致している必要があるが、提案手法では単眼 vSLAM を利用しているため、スケールは不定である。今回は、推定された値の並進部分のスケールを真値に一致させることでスケール合わせを行った。

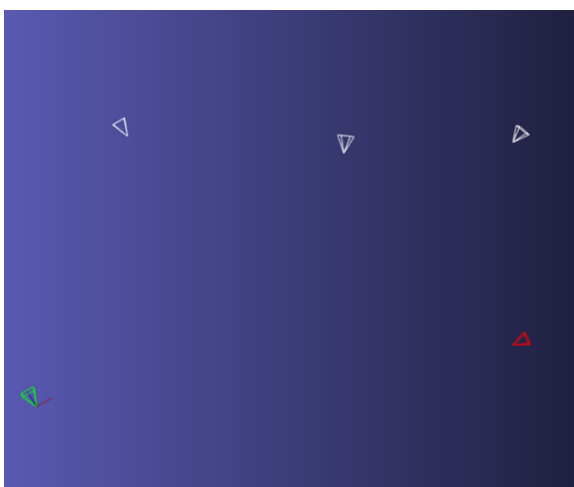
表 1: 真値の較正結果

	x	y	z
並進 [mm]	277.16	38.01	-127.48
回転 [degree]	96.76	77.64	105.77

撮影を行った計 8 シーンに対し、2 シーンは共通シーンのテクスチャ不足により、マップ統合において共通構造を検出することができなかった。また、他の 2 シーンでは、vSLAM 実行中にトラッキングが失敗した。これらは、vSLAM が特徴点に基づく処理であるため、周辺環境のテクスチャが十分に必要であることを示している。外部パラメータを算出できた 4 シーンに対し、較正結果と真値との誤差を表 2 に示す。



(a) 実験用のカメラのセットアップ



(b) 校正結果

図 6: 2 台のカメラを用いた校正実験

表 2: 真値と比較した際の推定誤差

	x	y	z	deg x	deg y	deg z
scene0	16.81	4.32	-8.62	0.099	0.54	0.046
scene2	-24.43	16.22	1.15	0.011	1.081	0.56
scene5	6.14	-16.87	8.14	0.050	1.096	0.0036
scene6	-4.24	-7.78	-1.60	0.064	1.086	0.47

4. おわりに

本稿では、視野を共有しない車載カメラ間の外部パラメータ校正を屋外の実環境下で行う方法を提案した。校正済みの実験器具を用いた評価を行い、これによって定量的に誤差を評価した。今後は実際の車に設置したカメラでの評価を行う。

参考文献

- [1] Strauß, T., Ziegler, J. and Beck, J.: Calibrating multiple cameras with non-overlapping views using coded checkerboard targets, *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 2623–2628 (2014).
- [2] Yin, L., Wang, X., Ni, Y., Zhou, K. and Zhang, J.: Extrinsic parameters calibration method of cameras with non-overlapping fields of view in airborne remote sensing, *Remote Sensing*, Vol. 10, No. 8, p. 1298 (2018).
- [3] Zhu, C., Zhou, Z., Xing, Z., Dong, Y., Ma, Y. and Yu, J.: Robust plane-based calibration of multiple non-overlapping cameras, *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, pp. 658–666 (2016).
- [4] Ataer-Cansizoglu, E., Taguchi, Y., Ramalingam, S. and Miki, Y.: Calibration of non-overlapping cameras using an external SLAM system, *2014 2nd International Conference on 3D Vision*, Vol. 1, IEEE, pp. 509–516 (2014).
- [5] Heng, L., Lee, G. H. and Pollefeys, M.: Self-calibration and visual slam with a multi-camera system on a micro aerial vehicle, *Autonomous robots*, Vol. 39, No. 3, pp. 259–277 (2015).
- [6] Taketomi, T., Uchiyama, H. and Ikeda, S.: Visual SLAM algorithms: a survey from 2010 to 2016, *IPSJ Transactions on Computer Vision and Applications*, Vol. 9, No. 1, p. 16 (2017).
- [7] Mur-Artal, R., Montiel, J. M. M. and Tardos, J. D.: ORB-SLAM: a versatile and accurate monocular SLAM system, *IEEE transactions on robotics*, Vol. 31, No. 5, pp. 1147–1163 (2015).
- [8] Carrera, G., Angeli, A. and Davison, A. J.: SLAM-based automatic extrinsic calibration of a multi-camera rig, *2011 IEEE International Conference on Robotics and Automation*, IEEE, pp. 2652–2659 (2011).
- [9] Heng, L., Li, B. and Pollefeys, M.: Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry, *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 1793–1800 (2013).
- [10] Nister, D. and Stewenius, H.: Scalable recognition with a vocabulary tree, *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2, Ieee, pp. 2161–2168 (2006).
- [11] Bian, J., Lin, W.-Y., Matsushita, Y., Yeung, S.-K., Nguyen, T.-D. and Cheng, M.-M.: Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4181–4190 (2017).
- [12] Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns, *IEEE Transactions on Pattern Analysis & Machine Intelligence*, No. 4, pp. 376–380 (1991).
- [13] Sturm, J., Engelhard, N., Endres, F., Burgard, W. and Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems, *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 573–580 (2012).