

マルチモーダルステレオカメラにおける 単眼 SfM のスケール推定

角倉 慎弥^{1,a)} 櫻田 健^{2,b)} 河口 信夫^{1,c)} 中村 良介^{2,d)}

概要: 本研究では、マルチモーダルステレオカメラにおける単眼 SfM のスケール推定手法を提案する。コンピュータビジョン・ロボティクス分野では、観測機器の簡略化を主な目的として単眼 SfM のスケール推定について研究が行われている。単眼 SfM のスケールを推定するには、例えば IMU や GPS などのカメラ以外のセンサ情報、あるいはカメラのセットアップなどに関する事前知識を用いることが一般的である。本研究では、RGB・FIR カメラなどの異なるモダリティを持つ 2 台のカメラからなるステレオカメラシステム (マルチモーダルステレオカメラ) における、単眼 RGB 画像による SfM のスケール推定手法を提案する。異なるモダリティを持つ画像の間では特徴点对応を正しく推定することが難しいため、それらの対応点からスケールを推定することは困難である。そこで本研究では、FIR 画像間のエッジポラ拘束の残差の最小化によってスケールを推定する。FIR 画像は一般的に解像度が低くテクスチャが乏しいため、少数の対応点しか得られない。そのため提案手法では、複数の FIR 画像ペアの対応点を用いて全体最適化を行うことでスケールを推定する。さらに、推定したスケールを初期値としてバンドル調整を行うことで、推定スケールの精度を向上する。本研究では、まずエッジポラ拘束に基づいたスケールの推定手法を定式化した。続いて、シミュレーションおよび実環境下での実験により提案手法のスケール推定精度と安定性を検証し、応用例として RGB-FIR カメラシステムを用いた 3 次元温度マップの構築を行った。

キーワード: スケール推定, マルチモーダル画像, 遠赤外線, 単眼 SfM

1. はじめに

本研究では、異なる波長域・モダリティの画像を同時に撮影できるステレオカメラシステム (マルチモーダルステレオカメラ) における単眼 SfM のスケール推定手法を提案する (図 1)。コンピュータビジョン・ロボティクス分野では、スマートフォンやデジタルカメラなどのデバイスの発展に伴い、カメラシステムの簡略化を主な目的として単眼 SfM のスケール推定に関する研究が盛んに行われている [1], [2]。単眼 SfM のスケールを推定する方法としては、Global Positioning System (GPS) などの衛星測位システム (navigation satellite system: NSS) や慣性計測装置 (inertial measurement unit: IMU) などから得られる情報と画像から得られる情報を統合する手法 [3] や、カメラのセットアップなどの事前知識を用いる手法 [4], [5] が挙げられる。本研究では、RGB カメラと遠赤外線 (far-

infrared: FIR) カメラのように、SIFT [6], SURF [7] あるいは ORB [8] といった局所特徴量で画像間の対応点を直接得ることが困難なステレオカメラシステムにおける単眼 SfM のスケール推定手法を提案する。

FIR カメラの製造技術の向上に伴い FIR カメラの用途が拡大しており、例えば建築物・建造物の診断や暗所での動物・歩行者の検出、火山活動の観測などに FIR カメラが広く用いられている。近年では、ドローンなどの無人航空機 (unmanned aerial vehicle: UAV) に搭載した RGB・FIR カメラのステレオカメラシステムによる建築物・建造物等の撮影および診断が行われており、そのための製品も販売されている。本研究では、そのような RGB・FIR カメラペアのように、異なる波長域・モダリティの画像を同時に撮影できるステレオカメラシステムをマルチモーダルステレオカメラと呼ぶ。マルチモーダルステレオカメラを用いた撮影では異なる波長域の画像を同時に取得できるが、大量の RGB・FIR 画像ペアを参照する診断には依然として多大な労力が必要となる。そのような調査・診断を効率化する方策として、Structure from Motion (SfM) [9], [10] と Multi-view Stereo (MVS) [11], [12], [13] をベースとした 3 次元温度マップの構築が挙げられる。具体的には、RGB 画像による SfM と MVS を用いてシーンの 3 次元モデルを構築し、そのモデルに FIR 画像から得られる温度情報を投

¹ 名古屋大学大学院 工学研究科
Graduate School of Engineering, Nagoya University

² 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

a) sumikura@ucl.nuee.nagoya-u.ac.jp

b) k.sakurada@aist.go.jp

c) kawaguti@nagoya-u.jp

d) r.nakamura@aist.go.jp

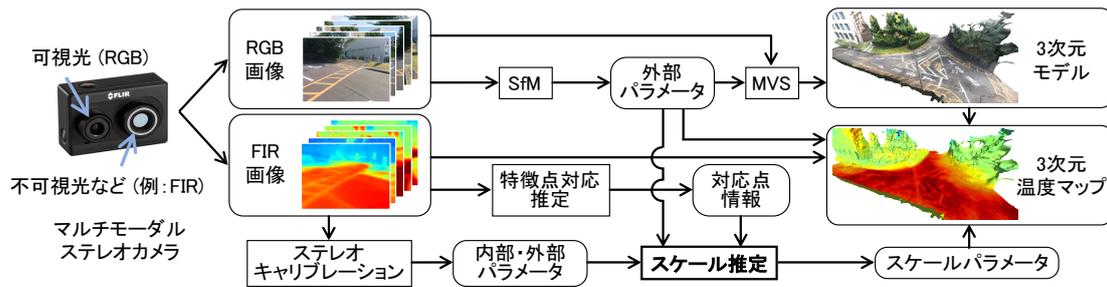


図 1 提案手法のフローチャートと応用例 (3次元温度マップの構築)。

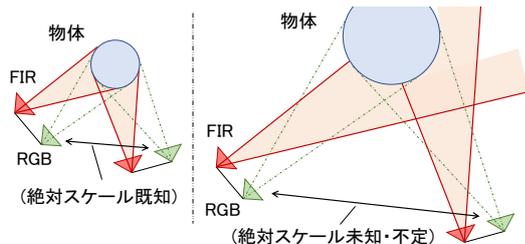


図 2 RGB-FIR カメラシステムにおける温度投影の模式図. 緑色の線と赤色の線は、それぞれ RGB・FIR 画像中に写っている物体の投影線を表している. RGB カメラ姿勢の絶対スケールが未知あるいは不定のときは、FIR 画像の温度情報を物体に再投影する際に投影線が物体からずれてしまう。

影して 3次元温度マップを構築する. 単眼 RGB 画像による SfM・MVS では推定された 3次元モデルの絶対スケールが不定であるため、FIR 画像から得られる温度情報を 3次元モデル上に投影するには 3次元モデルの絶対スケールの推定が必要になる (図 2). しかし、前述の通り RGB 画像と FIR 画像の間での直接の対応点推定は難しく、それらの対応点を元にしたスケールの推定は困難である. さらに FIR 画像間においても、得られる特徴点対応の数は比較的少ない、これは主に、FIR 画像が低解像度であることや、遠赤外線画像のテクスチャが可視光画像のそれに比べて乏しいことが原因である. Deep neural networks (DNN) などの機械学習ベースの手法 [14], [15], [16] を用いれば異なるモダリティの画像間での対応点推定が可能だが、そのような手法を適用する場合はシーン毎・カメラ毎のトレーニングデータの準備が求められ、多大な労力が必要となりコストが大きい.

そこで本研究では、マルチモーダルステレオカメラシステムから得られる情報のみを使って単眼 RGB 画像による SfM の絶対スケールを推定するために、FIR 画像間の少数の対応点と RGB・FIR カメラ間の相対姿勢を用いる (図 1). 具体的には、FIR 画像間のエピソード拘束の残差の最小化によって絶対スケールパラメータを推定する. 本研究の達成点は以下の通りである.

- マルチモーダルステレオカメラにおける単眼 SfM のスケール推定アルゴリズムの定式化
 - 仮想環境および実環境下における提案手法の有効性の検証
 - 提案手法の応用例の例示 (3次元温度マップの構築)
- 本報告は以下で構成される. 第 2 章で関連研究をま

め、第 3 章で問題の定式化とスケール推定アルゴリズムの詳細を説明する. 続いて、第 4 章で提案手法の安定性をシミュレーションによって検証し、第 5 章で実環境下における精度評価を行う. 最後に、第 6 章で本報告をまとめる.

2. 関連研究

2.1 単眼 SfM のスケール推定

既存の単眼 SfM のスケール推定手法として、カメラ以外のセンサ情報を用いる手法や、カメラのセットアップ等の事前知識を用いる手法が挙げられる.

IMU を用いる手法では、IMU から得られる加速度・角速度と画像群から得られる姿勢情報を拡張カルマンフィルタ (extended Kalman filter: EKF) を用いて統合し、スケールパラメータを推定する [3]. また、カメラ以外のセンサ情報として GPS などの衛星測位システムから得られる位置情報を用いることもできる. 具体的には、単眼画像の SfM から得られるカメラ位置と衛星測位によって得られるカメラ位置の相似変換を求めることでスケールパラメータを推定する.

一方で、カメラのセットアップに関する事前知識を用いるスケール推定手法も存在する. Scaramuzza らは、カメラが搭載されている車両の非ホロノミック拘束を利用した手法を提案している [4]. Kitt らの研究では、地面の検出およびカメラの地面からの高さ情報を利用している [5]. また、単眼画像による SfM や simultaneous localization and mapping (SLAM) では、スケールドリフトの補正のためにグラフ最適化とループ検出によるループクロージングが用いられる [17], [18], [19].

本研究では、単眼 SfM の絶対スケールの推定のためにマルチモーダルステレオカメラの情報のみを使用し、カメラ以外のセンサ情報等は用いない. これは主に、提案手法を様々な用途に応用可能にするための方策である. 例えば、UAV による建築物・建造物の診断では、搭載された IMU は振動等によるノイズを受けることが考えられる. また GPS の信号も、構造物の下部や屋内では受信が困難である. そのため提案手法では、マルチモーダルステレオカメラから得られる情報のみを利用する.

2.2 3次元温度マップの構築

一般的に FIR 画像のテクスチャは RGB 画像より乏しい. そのため、FIR 画像を使った診断を容易にするために、FIR

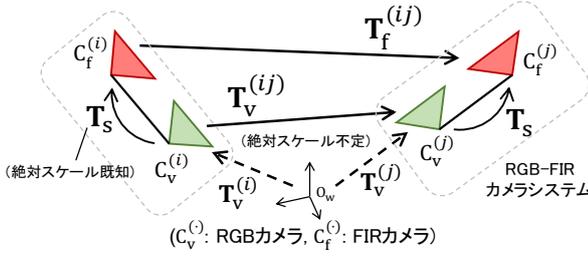


図 3 2 視点 i, j における各カメラ姿勢の定義. $\mathbf{T}_v^{(i)}$ と $\mathbf{T}_f^{(i)}$ はそれぞれ RGB カメラ $C_v^{(i)}$ と FIR カメラ $C_f^{(i)}$ の絶対姿勢を表している. また, \mathbf{T}_s は同視点の $C_v^{(i)}$ と $C_f^{(i)}$ の間の相対姿勢を表している. $\mathbf{T}_v^{(ij)}$ は, 同じタイプのカメラ ($C_v^{(i)}, C_v^{(j)}$) 間の相対姿勢を表している.

カメラとそれ以外のセンサを組み合わせることで 3 次元温度マップを構築する研究が広くなされている. Oreifej らは, light detection and ranging (LiDAR) センサを用いて屋内の 3 次元温度マップを構築する手法を提案している [20]. また, デプスカメラを用いて Kinect Fusion アルゴリズム [21] によりシーンの 3 次元モデルを構築し, そのモデルに温度情報を投影することも行われている [22], [23].

シーンの 3 次元モデルを構築する手法として, RGB 画像の SfM・MVS を用いることもできる. この手法では, 前述の通り 3 次元モデルの絶対スケールの推定が必要となる. Ham らは, RGB 画像と FIR 画像の間で直接対応点を推定して絶対スケールを推定している [24]. この手法は, シーンの温度テクスチャが豊富な場所にしか適用できない. Truong らの手法では, RGB 画像と FIR 画像でそれぞれ SfM・MVS を行い, それらの点群の位置合わせを行うことでスケールパラメータを推定している [25]. この手法も, FIR 画像のみでの安定した SfM・MVS が必要となるため適用できるシーンに限られる.

LiDAR センサの測距距離はデプスカメラより広いが, センサの重量が大きいことや価格が高いことが欠点となる. デプスカメラはシーンの 3 次元形状を絶対スケールで測定できるが, 測距範囲が限られるため広い範囲の計測・診断には適さない. 前述の通り, 本研究は建築物・建造物等の広範囲な 3 次元温度マップの構築を目的としているため, RGB・FIR カメラペアのようなマルチモーダルカメラシステムを対象とするスケール推定手法を提案する.

3. スケール推定

3.1 問題の定式化

本章では, 提案手法の詳細を説明する. マルチモーダルステレオカメラの例として, RGB カメラと FIR カメラからなる RGB-FIR カメラシステムを用いる. 図 3 は, 2 視点の RGB-FIR カメラシステムからなる系における絶対姿勢および相対姿勢の定義を表している.

いま, RGB 画像の集合 $\{I_v^{(1)}, I_v^{(2)}, \dots, I_v^{(n)}\}$ および FIR 画像の集合 $\{I_f^{(1)}, I_f^{(2)}, \dots, I_f^{(n)}\}$ が与えられているとする. それぞれの k 番目の画像, すなわち $I_v^{(k)}$ と $I_f^{(k)}$ は, RGB-FIR カメラシステムを用いて同時に撮影されているものとする. ここで, RGB-FIR カメラシステムの RGB・FIR カ

メラ間の相対姿勢を

$$\mathbf{T}_s = \begin{bmatrix} \mathbf{R}_s & \mathbf{t}_s \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad (1)$$

とする. \mathbf{R}_s と \mathbf{t}_s は, それぞれ RGB・FIR カメラ間の回転行列と並進ベクトルを表している. これらの行列およびベクトルは, キャリブレーションにより事前に推定できる. k 番目の RGB 画像 $I_v^{(k)}$ および FIR 画像 $I_f^{(k)}$ は, それぞれ k 番目の RGB カメラ $C_v^{(k)}$, FIR カメラ $C_f^{(k)}$ から撮影されているものとする. つまり, $C_v^{(k)}$ と $C_f^{(k)}$ は RGB-FIR カメラシステムのステレオカメラペアをなす. さらに, RGB カメラ $C_v^{(k)}$ と FIR カメラ $C_f^{(k)}$ の姿勢を, それぞれ $\mathbf{T}_v^{(k)}$ と $\mathbf{T}_f^{(k)}$ で表す. $\{\mathbf{T}_v^{(k)} | k = 1, 2, \dots, n\}$ は, RGB 画像群 $\{I_v^{(k)} | k = 1, 2, \dots, n\}$ を用いる単眼 SfM によって推定できる. ただし, 絶対スケールは未知である.

$\mathbf{T}_v^{(i)}$ と $\mathbf{T}_v^{(j)}$ を用いると, $C_v^{(i)}$ と $C_v^{(j)}$ の間の相対姿勢は $\mathbf{T}_v^{(j)}\mathbf{T}_v^{(i)-1}$ で計算できる. 単眼 RGB 画像による SfM のスケール不定性を解消するためにスケールパラメータ s を導入し, $C_v^{(i)}$ と $C_v^{(j)}$ の相対姿勢 $\mathbf{T}_v^{(ij)}$ を求めると,

$$\mathbf{T}_v^{(ij)} = \begin{bmatrix} \mathbf{R}_v^{(ij)} & s \cdot \mathbf{t}_v^{(ij)} \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad (2)$$

のように表せる. ここで, $\mathbf{R}_v^{(ij)}$ は $\mathbf{T}_v^{(j)}\mathbf{T}_v^{(i)-1}$ の回転行列ブロック, $\mathbf{t}_v^{(ij)}$ は $\mathbf{T}_v^{(j)}\mathbf{T}_v^{(i)-1}$ の並進ベクトルブロックを表している. すなわち, 並進ベクトル $s \cdot \mathbf{t}_v^{(ij)}$ の長さが実際の RGB カメラ ($C_v^{(i)}, C_v^{(j)}$) 間の距離と合うように s を定めることが絶対スケールの推定問題である.

3.2 スケールパラメータ s の導出

$\mathbf{T}_v^{(j)}$ と \mathbf{T}_s を用いて視点 i, j の FIR カメラ ($C_f^{(i)}, C_f^{(j)}$) 間の相対姿勢 $\mathbf{T}_f^{(ij)}$ を計算すると,

$$\begin{aligned} \mathbf{T}_f^{(ij)} &= \mathbf{T}_s \mathbf{T}_v^{(ij)} \mathbf{T}_s^{-1} \\ &= \begin{bmatrix} \mathbf{R}_s \mathbf{R}_v^{(ij)} \mathbf{R}_s^{-1} & s \cdot \mathbf{R}_s \mathbf{t}_v^{(ij)} + (\mathbf{I} - \mathbf{R}_s \mathbf{R}_v^{(ij)} \mathbf{R}_s^{-1}) \mathbf{t}_s \\ \mathbf{0}^\top & 1 \end{bmatrix} \end{aligned} \quad (3)$$

$$= \begin{bmatrix} \mathbf{A}^{(ij)} & s \cdot \mathbf{b}^{(ij)} + \mathbf{c}^{(ij)} \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad (4)$$

となる. ここで, $\mathbf{A}^{(ij)} = [\mathbf{a}_1^{(ij)} \ \mathbf{a}_2^{(ij)} \ \mathbf{a}_3^{(ij)}] = \mathbf{R}_s \mathbf{R}_v^{(ij)} \mathbf{R}_s^{-1}$, $\mathbf{b}^{(ij)} = \mathbf{R}_s \mathbf{t}_v^{(ij)}$, $\mathbf{c}^{(ij)} = (\mathbf{I} - \mathbf{R}_s \mathbf{R}_v^{(ij)} \mathbf{R}_s^{-1}) \mathbf{t}_s$ である. FIR カメラ ($C_f^{(i)}, C_f^{(j)}$) 間の essential matrix $\mathbf{E}^{(ij)}$ は, FIR カメラ間の相対姿勢 $\mathbf{T}_f^{(ij)}$ から求めることができ,

$$\begin{aligned} \mathbf{E}^{(ij)} &= [s \mathbf{b}^{(ij)} + \mathbf{c}^{(ij)}]_\times \mathbf{A}^{(ij)} \\ &= s \cdot [\mathbf{b}^{(ij)} \times \mathbf{a}_1^{(ij)} \quad \mathbf{b}^{(ij)} \times \mathbf{a}_2^{(ij)} \quad \mathbf{b}^{(ij)} \times \mathbf{a}_3^{(ij)}] \\ &\quad + [\mathbf{c}^{(ij)} \times \mathbf{a}_1^{(ij)} \quad \mathbf{c}^{(ij)} \times \mathbf{a}_2^{(ij)} \quad \mathbf{c}^{(ij)} \times \mathbf{a}_3^{(ij)}] \end{aligned} \quad (5)$$

となる. $\mathbf{p}_k^{(i)} = [x_k^{(i)}, y_k^{(i)}, 1]^\top$ と $\mathbf{p}_k^{(j)} = [x_k^{(j)}, y_k^{(j)}, 1]^\top$ を FIR 画像 ($I_f^{(i)}, I_f^{(j)}$) 間における k 番目の対応点とすると,

$(I_f^{(i)}, I_f^{(j)})$ 間のエピソード拘束条件は

$$\mathbf{p}_k^{(j)\top} \mathbf{E}^{(ij)} \mathbf{p}_k^{(i)} = 0 \quad (8)$$

と書ける. なお, $\mathbf{p}_k^{(i)}, \mathbf{p}_k^{(j)}$ は正規化画像座標系 [26] で表されているものとする. 正規化画像座標 $\mathbf{p}_k^{(i)}$ は,

$$\mathbf{p}_k^{(i)} = \mathbf{K}_f^{-1} \begin{bmatrix} u_k^{(i)} \\ v_k^{(i)} \\ 1 \end{bmatrix}^\top \quad \text{with } \mathbf{K}_f = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

で定義される. \mathbf{K}_f は, FIR カメラの内部パラメータである. 加えて, 正規化画像座標は

$$\mathbf{p}_k^{(i)} = \mathbf{X}_l^{(i)} / Z_l^{(i)} \quad (10)$$

でも得られる. ここで, $\mathbf{X}_l^{(i)} = [X_l^{(i)}, Y_l^{(i)}, Z_l^{(i)}]^\top$ は l 番目の 3 次元点の座標であり, i 番目の FIR カメラ $C_f^{(i)}$ 基準の座標系で表されている. なお, $\mathbf{X}_l^{(i)}$ は FIR 画像 $I_f^{(i)}$ 上の特徴点 $\mathbf{p}_k^{(i)}$ と対応しているものとする.

エピソード拘束条件を表した式 (8) を変形すると,

$$\mathbf{u}_k^{(ij)} (s \cdot \mathbf{f}^{(ij)} + \mathbf{g}^{(ij)}) = 0 \quad (11)$$

という形になる. 行ベクトル $\mathbf{u}_k^{(ij)}$, 列ベクトル $\mathbf{f}^{(ij)}, \mathbf{g}^{(ij)}$ は全て 9 次元であり, 成分を書き下すと

$$\mathbf{u}_k^{(ij)} = [x_k^{(i)} x_k^{(j)}, x_k^{(i)} y_k^{(j)}, x_k^{(i)}, y_k^{(i)} x_k^{(j)}, y_k^{(i)} y_k^{(j)}, y_k^{(i)}, x_k^{(j)}, y_k^{(j)}, 1] \quad (12)$$

$$\mathbf{f}^{(ij)} = \begin{bmatrix} [\mathbf{b}^{(ij)} \times \mathbf{a}_1^{(ij)}]_1, [\mathbf{b}^{(ij)} \times \mathbf{a}_1^{(ij)}]_2, \dots, [\mathbf{b}^{(ij)} \times \mathbf{a}_3^{(ij)}]_2, [\mathbf{b}^{(ij)} \times \mathbf{a}_3^{(ij)}]_3 \end{bmatrix}^\top \quad (13)$$

$$\mathbf{g}^{(ij)} = \begin{bmatrix} [\mathbf{c}^{(ij)} \times \mathbf{a}_1^{(ij)}]_1, [\mathbf{c}^{(ij)} \times \mathbf{a}_1^{(ij)}]_2, \dots, [\mathbf{c}^{(ij)} \times \mathbf{a}_3^{(ij)}]_2, [\mathbf{c}^{(ij)} \times \mathbf{a}_3^{(ij)}]_3 \end{bmatrix}^\top \quad (14)$$

となる. 対応する特徴点に誤差が全く含まれなければ式 (11) は完全に成立する. しかし実際の特徴点には幾分かの誤差が含まれており, さらにスケールパラメータ s も未知であるため式 (11) が常に成り立つとは限らない. そこで, エピソード拘束による残差 $e_k^{(ij)}$ を

$$e_k^{(ij)} = \mathbf{u}_k^{(ij)} (s \cdot \mathbf{f}^{(ij)} + \mathbf{g}^{(ij)}) \quad (15)$$

で定義する. 同様に, 残差ベクトル $\mathbf{e}^{(ij)}$ を

$$\mathbf{e}^{(ij)} = \mathbf{U}^{(ij)} (s \mathbf{f}^{(ij)} + \mathbf{g}^{(ij)}) \quad \text{with } \mathbf{U}^{(ij)} = \begin{bmatrix} \mathbf{u}_1^{(ij)\top} & \mathbf{u}_2^{(ij)\top} & \dots & \mathbf{u}_m^{(ij)\top} \end{bmatrix}^\top \quad (16)$$

で定義する. m は FIR 画像 $(I_f^{(i)}, I_f^{(j)})$ 間の特徴点对応数である. 最小二乗法により, スケールパラメータ s は

$$s = \arg \min_{s \in \mathbb{R}} \frac{1}{2} \sum_{i,j} \|\mathbf{e}^{(ij)}\|^2 \quad (17)$$

で得られる. すなわちスケール推定問題は, 目的関数

$$J(s) = \frac{1}{2} \sum_{i,j} \left\| \mathbf{U}^{(ij)} (s \cdot \mathbf{f}^{(ij)} + \mathbf{g}^{(ij)}) \right\|^2 \quad (18)$$

が最小となる s を定める問題に帰着する. $dJ(s)/ds = 0$ となる s を求めると,

$$s = - \frac{\sum_{i,j} (\mathbf{f}^{(ij)\top} \mathbf{U}^{(ij)\top} \mathbf{U}^{(ij)} \mathbf{g}^{(ij)})}{\sum_{i,j} (\mathbf{f}^{(ij)\top} \mathbf{U}^{(ij)\top} \mathbf{U}^{(ij)} \mathbf{f}^{(ij)})} \quad (19)$$

となる.

3.3 スケールパラメータ s の異なる導入方法

式 (2) では, s を RGB カメラ $(C_f^{(i)}, C_f^{(j)})$ 間の相対並進ベクトル $\mathbf{t}_v^{(ij)}$ に乗ずることでスケールパラメータを導入している. 代わりに, スケールパラメータ s を RGB · FIR カメラ間の相対並進ベクトル \mathbf{t}_s に乗ずることでスケールパラメータを導入できる. なぜなら, \mathbf{t}_s に s を乗ずることは $\mathbf{t}_v^{(ij)}$ に $1/s$ を乗ずることと幾何的に等価だからである. その場合, スケールパラメータの推定式の導出は

$$\mathbf{T}_v^{(ij)} = \begin{bmatrix} \mathbf{R}_v^{(ij)} & \mathbf{t}_v^{(ij)} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \quad \mathbf{T}_s = \begin{bmatrix} \mathbf{R}_s & s \cdot \mathbf{t}_s \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad (20)$$

に基づいて行われる. 式 (20) を元にスケール推定式を導出する場合, 式 (5) 中の $\mathbf{A}^{(ij)}, \mathbf{b}^{(ij)}, \mathbf{c}^{(ij)}$ はそれぞれ, $\mathbf{A}^{(ij)} = [\mathbf{a}_1^{(ij)} \mathbf{a}_2^{(ij)} \mathbf{a}_3^{(ij)}] = \mathbf{R}_s \mathbf{R}_v^{(ij)} \mathbf{R}_s^{-1}$, $\mathbf{b}^{(ij)} = (\mathbf{I} - \mathbf{R}_s \mathbf{R}_v^{(ij)} \mathbf{R}_s^{-1}) \mathbf{t}_s$, $\mathbf{c}^{(ij)} = \mathbf{R}_s \mathbf{t}_v^{(ij)}$ となる. 以降の導出過程は変わらない.

以降, 式 (1) と式 (2) に基づいたスケール推定アルゴリズムをアルゴリズム (1), 式 (20) に基づいたスケール推定アルゴリズムをアルゴリズム (2) と表記する.

3.4 スケールパラメータを考慮したバンドル調整

アルゴリズム (1) または (2) における式 (19) を用いたスケールの初期推定の後, バンドル調整 [27] を行う. RGB カメラ姿勢自体は, スケール推定以前に絶対スケールを除き高精度で得られている. そのため, RGB カメラ姿勢の回転行列・並進ベクトルを定数, スケールパラメータ s を変数としたバンドル調整を行う.

FIR 画像 $I_f^{(i)}$ 上の再投影誤差 $\delta_{k,l}^{(i)}$ は, グローバル座標系で表した l 番目の 3 次元点 $\mathbf{X}_l = [X_l, Y_l, Z_l]^\top$ を用いると,

$$\delta_{k,l}^{(i)} = \mathbf{x}_k^{(i)} - \pi^{(i)}(s, \mathbf{X}_l) \quad (21)$$

と書ける. ここで, $\mathbf{x}_k^{(i)}$ は FIR 画像 $I_f^{(i)}$ 上の 2 次元特徴点であり, 3 次元点 \mathbf{X}_l と対応している. $\pi^{(i)}(\cdot)$ は i 番目の FIR カメラにおける投影関数であり,

$$\pi^{(i)}(s, \mathbf{X}_l) = \begin{bmatrix} f_x \frac{X_l^{(i)}}{Z_l^{(i)}} + c_x, f_y \frac{Y_l^{(i)}}{Z_l^{(i)}} + c_y \end{bmatrix}^\top \quad (22)$$

と書ける. $\mathbf{X}_l^{(i)} = [X_l^{(i)}, Y_l^{(i)}, Z_l^{(i)}]^\top$ は, アルゴリズムによって導出手順が異なり, アルゴリズム (1) では

$$\mathbf{X}_l^{(i)} = \mathbf{R}_s \mathbf{R}_v^{(i)} \mathbf{X}_l + s \cdot \mathbf{R}_s \mathbf{t}_v^{(i)} + \mathbf{t}_s \quad (23)$$

となり、一方でアルゴリズム (2) では

$$\mathbf{X}_l^{(i)} = \mathbf{R}_s \mathbf{R}_v^{(i)} \mathbf{X}_l + \mathbf{R}_s \mathbf{t}_v^{(i)} + s \cdot \mathbf{t}_s \quad (24)$$

で得られる。

以上を用いると、再投影誤差によるコスト関数 $\mathcal{L}(\cdot)$ は

$$\mathcal{L}(s, \{\mathbf{X}_l\}, \mathbf{K}_f; \{\mathbf{T}_v^{(i)}\}, \mathbf{T}_s) = \sum_{i,k,l} \rho \left(\left\| \delta_{k,l}^{(i)} \right\|^2 / \sigma_r^2 \right) \quad (25)$$

と書ける。 $\rho(\cdot)$ は Huber 損失関数であり、 σ_r は再投影誤差の標準偏差である。最適化されたスケールパラメータ s は以下の式で得られる。

$$s = \arg \min_{s \in \mathbb{R}, \{\mathbf{X}_l\}, \mathbf{K}_f} \mathcal{L}(s, \{\mathbf{X}_l\}, \mathbf{K}_f; \{\mathbf{T}_v^{(i)}\}, \mathbf{T}_s) \quad (26)$$

式 (26) は非凸最適化問題であり、Levenberg–Marquardt 法などの反復解法で解くことが必要になる。その初期値はアルゴリズム (1) または (2) の式 (19) を用いて定められる。

4. 仮想環境下での安定性評価

第 3 章において、スケールパラメータ s を導入する位置が異なる 2 種類のスケール推定アルゴリズムを示した。この章では、2 種類のアルゴリズムそれぞれについて特徴点の誤差および RGB–FIR カメラシステムのベースライン長に対する推定安定性を検証する。具体的には、まず RGB–FIR カメラシステムによる仮想的な撮影データを用意する。続いて、それらの特徴点座標にノイズを与えた状態でアルゴリズム (1) および (2) それぞれを用いてスケールを推定し、安定性を検証する。なお、予備実験によりノイズが無い環境下ではアルゴリズム (1) と (2) の両方でスケールパラメータが正しく推定できることを確認している。

4.1 実験手順

- (1) 3次元点 $\mathbf{X}_i \in \mathbb{R}^3$ ($i = 1, 2, \dots, n_p$) を一辺が D の立方体空間中にランダム散布する。
- (2) 3次元空間中に n_c 個の RGB–FIR カメラシステムをランダムに配置する。具体的には、カメラシステムの RGB · FIR カメラ間の相対姿勢 \mathbf{T}_s が与えられた状態で、RGB カメラ姿勢 $\mathbf{T}_v^{(k)}$ ($k = 1, 2, \dots, n_c$) をランダムに設定し、FIR カメラ姿勢 $\mathbf{T}_f^{(k)}$ ($k = 1, 2, \dots, n_c$) を $\mathbf{T}_f^{(k)} = \mathbf{T}_s \mathbf{T}_v^{(k)}$ で設定する。
- (3) k 番目の FIR カメラ ($k = 1, 2, \dots, n_c$) について、3次元点 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{n_p}$ を $\mathbf{T}_f^{(k)}$ を用いて再投影し、各 3次元点に対応する正規化画像座標を求める。
- (4) 各正規化画像座標に平均 0 · 標準偏差 $\sigma_n \geq 0$ のガウシアンノイズを加える。
- (5) アルゴリズム (1) および (2) でスケールパラメータ s を推定する。この際、式 (15) の残差に基づいた外れ値除去を行う。なお、RGB カメラ位置はスケールがされていないため推定スケールの真値は 1.0 である。

今回の実験では、 $n_p = 1000$ 、 $D = 2000$ および $\sigma_n = 0.001$ に設定した。また、RGB · FIR カメラの相対姿勢 \mathbf{T}_s は

$$\mathbf{T}_s = \begin{bmatrix} \mathbf{R}_s & \mathbf{t}_s \\ \mathbf{0}^\top & 1 \end{bmatrix} \quad \text{with } \mathbf{R}_s = \mathbf{I} \text{ and } \mathbf{t}_s = \begin{bmatrix} d & 0 & 0 \end{bmatrix}^\top \quad (27)$$

と設定した。 $d > 0$ は RGB · FIR カメラ間のベースライン長である。

4.2 特徴点検出誤差の影響

複数のベースライン長において仮想環境下での撮影データの準備・ノイズ負荷・スケールパラメータ s の推定を 100 回行い、アルゴリズム (1) では $1/s$ の、アルゴリズム (2) では s の 100 回平均および標準偏差を求めた。図 4 はベースライン長 d と推定スケールの平均・標準偏差の関係を示している。

図 4b を見ると、ベースライン長が比較的長い領域 ($0.1 < d$) では、推定スケールの平均値が 1.0 であるとともに標準偏差が 0.0 に漸近しており、スケールパラメータ s が安定して推定されていることが分かる。また、ベースライン長が比較的短い領域 ($d < 0.1$) では、ベースライン長が短くなるほど標準偏差が増加しているが、推定スケールの平均値は 1.0 のままで保たれている。一方、図 4a を見ると、ベースライン長が比較的短い領域 ($d < 0.1$) において推定スケールの平均値の精度が図 4b より低下している。加えて、図 4a の標準偏差は図 4b のそれより大きくなっており、スケールの推定が安定していない。

以上をまとめると、アルゴリズム (2) によって推定されたスケールパラメータはアルゴリズム (1) によって推定されたものより精度が高く、アルゴリズム (2) のほうが特徴点に与えたノイズに対してよりロバストに推定が行えることが分かる。加えて、RGB–FIR カメラシステムのベースライン長は長い方が安定したスケール推定を行える。

5. 実環境下での精度評価

5.1 評価手法

この章では、実環境で撮影したデータに提案手法を適用して絶対スケールが正しく推定できることを確認する。検証のために、RGB カメラ間の実際の距離が既知であるデータセットを用意する必要がある。そのため本実験では、RGB–FIR カメラシステムを図 5 に示す架台に固定し、架台のステージを 100 [mm] の間隔で垂直あるいは水平方向に動かしながら RGB · FIR 画像を撮影してカメラ間距離が既知な撮影データを用意した。

いま、 $d^{(ij)}$ を RGB カメラ ($C_v^{(i)}, C_v^{(j)}$) 間の実際の距離とし、 $L^{(ij)}$ を単眼 SfM の結果から求めた ($C_v^{(i)}, C_v^{(j)}$) 間の距離とする。前述の通り単眼 RGB 画像の SfM にはスケール不定性があるため、一般的には $d^{(ij)} \neq L^{(ij)}$ である。このとき、RGB カメラ間の実際の距離の推定値 $\hat{d}^{(ij)}$ は、

$$\hat{d}^{(ij)} = s \cdot L^{(ij)} \quad (\text{アルゴリズム (1) の場合}) \quad (28)$$

$$\hat{d}^{(ij)} = L^{(ij)} / s \quad (\text{アルゴリズム (2) の場合}) \quad (29)$$

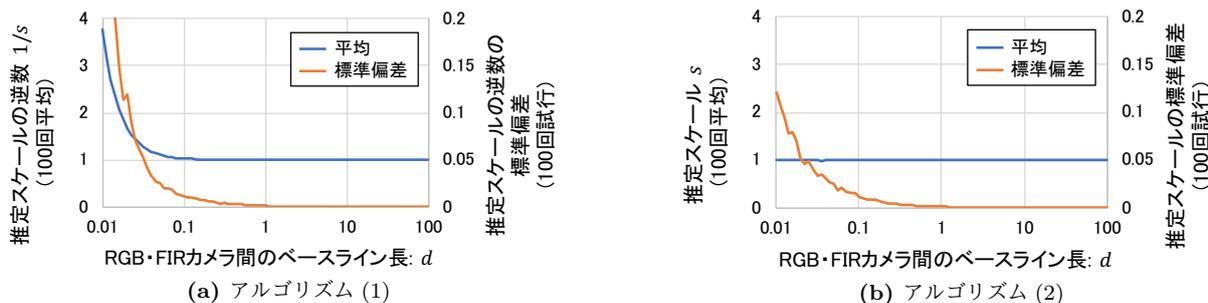


図 4 特徴点にノイズを与えた仮想環境下での、アルゴリズム (1) および (2) による推定スケールの平均 (左縦軸) および標準偏差 (右縦軸)。平均および標準偏差は 100 回の試行により求めた。図 4a および図 4b において、横軸は RGB-FIR カメラシステムのベースライン長 d を表しており、 $[10^{-2}, 10^2]$ の範囲で変化させている。なお推定スケールの真値は 1.0 である。推定スケールの精度および安定性が、特に $d < 0.1$ の範囲で大きく異なる。



図 5 RGB-FIR カメラシステムおよび実験で用いた架台。架台はグリッド状に並んだ視点で RGB・FIR 画像を撮影するのに用いた。架台のステージにカメラシステムを固定し、ステージを垂直・水平に動かしながら撮影を行った。

で得られる。 s はアルゴリズム (1) または (2) で推定したスケールパラメータである。さらに、推定値 $\hat{d}^{(ij)}$ の相対誤差 $\epsilon^{(ij)}$ は以下で得られる。

$$\epsilon^{(ij)} = \frac{\hat{d}^{(ij)} - d^{(ij)}}{d^{(ij)}} \times 100[\%] \quad (30)$$

本実験で用いた RGB-FIR カメラシステムを図 5 に示す。カメラシステムの RGB カメラは LUMIX DMC-G8 (Panasonic Corp.) または FLIR Duo R (FLIR Systems, Inc.) に搭載された RGB カメラを用いる。FIR カメラは FLIR Duo R に搭載された FIR カメラを用いる。

実験の手順は以下の通りである。

- (1) グリッド視点の RGB・FIR 画像のペアを図 5 に示したカメラシステムを用いて撮影する。また、単眼 SfM とスケール推定を安定させるために、自由視点で撮影した RGB・FIR 画像ペアを追加する。
- (2) 撮影した RGB 画像を用いて単眼 SfM の処理を行う。
- (3) FIR 画像間の対応点をの SIFT [6] 特徴量を用いて求める。なお、5 点アルゴリズム [28], [29] による RANdom SAmple Consensus (RANSAC) を用いて誤対応を除いた。
- (4) アルゴリズム (1) および (2) を用いてスケールパラメータ s を推定する。
- (5) 以下で定義される平均相対誤差 $\bar{\epsilon}$ を計算する。 N はグリッド視点で撮影された RGB 画像の枚数である。

$$\bar{\epsilon} = \frac{2}{N(N-1)} \sum_{i < j} \epsilon^{(ij)} \quad (31)$$

FIR 画像の特徴点の検出および特徴量の計算を行う際

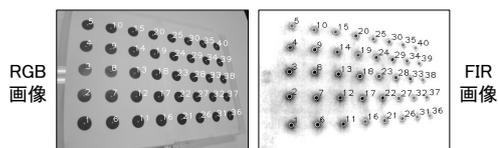


図 6 キャリブレーションボードの RGB・FIR 画像。



図 7 評価用環境および SfM で推定した RGB カメラ姿勢。

には、FIR 画像をグレースケール画像に変換する必要がある。FLIR Duo R は各ピクセルが物体の放射温度情報を持った FIR 画像を出力する。それらをグレースケール画像に変換するために、FIR 画像ごとに全ピクセル値の平均 μ および標準偏差 σ_p を求め、元の FIR 画像のピクセル値の $[\mu - 2\sigma_p, \mu + 2\sigma_p]$ の範囲を $[0, 2^8 - 1]$ にマッピングする。

RGB・FIR カメラ間のベースライン長がスケール推定に与える影響を考察するため、評価用に用意した環境の RGB・FIR 画像ペアを 4 種類のベースライン (273, 192, 113, 26 [mm]) のカメラシステムで撮影した。ベースライン長が 273, 192, 113 [mm] のカメラシステムでは、RGB カメラとして LUMIX DMC-G8 を用いた。ベースライン長が 26 [mm] のカメラシステムでは、RGB カメラとして FLIR Duo R に搭載された RGB カメラを用いた。RANSAC によるランダム性を考慮し、4 種類のベースライン長それぞれについてスケールの推定および $\bar{\epsilon}$ の計算を 100 回行い、それらの平均および標準偏差を求めた。

5.2 RGB-FIR カメラシステムのキャリブレーション

提案手法を実環境のデータに適用する場合、RGB-FIR カメラシステムの事前キャリブレーションが必要となる。我々は、チェスボードパターンなどの平面パターンを用いるキャリブレーション手法 [30] を用いて RGB-FIR カメラシステムのキャリブレーションを行った。キャリブレーションボードの平面パターンは、RGB カメラはもちろん

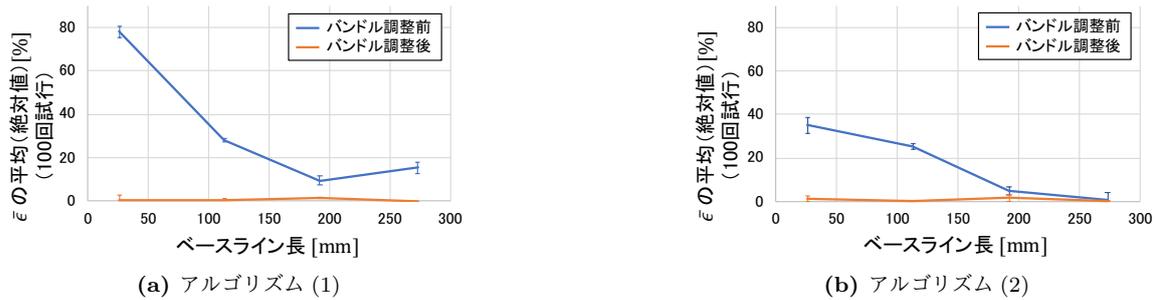


図 8 4 種類のベースライン (26, 113, 192, 273 [mm]) それぞれにおける $\bar{\epsilon}$ の 100 回平均。エラーバーは 100 回試行による $\bar{\epsilon}$ の $\pm 1\sigma$ の範囲を表している。バンドル調整前では、アルゴリズム (1) および (2) どちらにおいてもベースライン長が長くなるほど誤差の絶対値が小さくなっていることが分かる。加えて、アルゴリズム (1) による推定値の誤差がアルゴリズム (2) による推定値の誤差より大きいことが読み取れる。一方、バンドル調整後ではアルゴリズム (1) および (2) どちらにおいても $\bar{\epsilon}$ の平均値はほぼ 0 になっている。

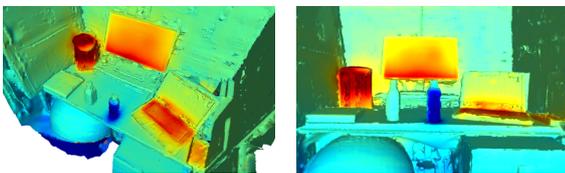


図 9 ベースライン長が 273 [mm] の RGB-FIR カメラシステムによる評価用環境 (図 7) の 3 次元温度マップ。

FIR カメラからも認識できる必要がある。そこで、グリッド上に並べられた黒色円形パターンの各中心に小型の発熱体を挿入したキャリブレーションボードを製作した (図 6)、キャリブレーション対象の RGB-FIR カメラシステムを用いて複数視点で RGB・FIR 画像を撮影し円形パターンの検出を行った後に、カメラシステムの内部・外部パラメータを Zhang の手法 [30] を用いて推定した。

5.3 評価結果

評価用のデータを撮影した環境を図 7 に示す。グリッド状に並んだ視点の数は、8 (縦) \times 10 (横) の計 80 視点である。また、単眼 SfM およびスケール推定を安定させるために 50 ペアの RGB・FIR 画像を追加した。RANSAC のランダム性を考慮し、提案手法によるスケール推定を 100 回行った上での $\bar{\epsilon}$ の 100 回平均および標準偏差を図 8 に示す。図 8a および図 8b は、それぞれアルゴリズム (1) と (2) を用いた際の相対誤差 $\bar{\epsilon}$ の絶対値を表している。

図 8a および図 8b どちらにおいても、カメラシステムのベースライン長が長くなるほどバンドル調整前の $\bar{\epsilon}$ の絶対値が小さくなっている。加えて、4 種のベースライン長いずれにおいても図 8a の誤差の絶対値は図 8b の誤差の絶対値より大きいことが読み取れる。これらの結果は第 4 章のシミュレーションの結果と同じ傾向を示している。つまり、バンドル調整を行わない場合は RGB-FIR カメラシステムのベースライン長が長いほど高精度なスケール推定が行えることが分かる。また、2 種類のアルゴリズムによる推定値の精度および安定性の違いが仮想環境下と実環境下の両方の実験で示された。

一方、バンドル調整後では図 8a と図 8b どちらにおいて

も $\bar{\epsilon}$ の絶対値がほぼ 0 になっている。特に、ベースライン長が 26 [mm] のときのアルゴリズム (1) による推定値の相対誤差 $\bar{\epsilon}$ は、バンドル調整前では -78.0 [%] であるのに対し、バンドル調整後では 0.543 [%] であった。また、ベースライン長が 273 [mm] のときは、アルゴリズム (1) および (2) による推定値の相対誤差 $\bar{\epsilon}$ がそれぞれ -0.146 , -0.148 [%] (いずれもバンドル調整後) であり、絶対スケールを高精度に推定できている。標準偏差についても、バンドル調整前に比べてバンドル調整後はバラつきが小さくなっている。以上をまとめると、アルゴリズム (1) および (2) による推定値はどちらもバンドル調整の初期値として十分な精度であるとともに、バンドル調整によって絶対スケールパラメータの推定精度を高められていることが分かる。

最後に、ベースライン長が 273 [mm] の RGB-FIR カメラシステムで取得したデータと推定したスケールパラメータを用いて構築した 3 次元温度マップを図 9 に示す。温度が高い部分 (電気ポット等)、低い部分 (ペットボトル) への温度の投影が正しく行われていることが確認できる。

6. 結論

本研究では、RGB・FIR カメラペアのような異なる波長域・モダリティのカメラで構成したステレオカメラシステムにおける単眼 SfM のスケール推定手法を提案した。RGB・FIR 画像のような異なるモダリティの画像間では直接の対応点推定が困難であるため、我々はカメラシステムの外部パラメータおよび FIR 画像間の少数の対応点を用いてスケールを推定する手法を提案した。提案手法は FIR 画像間のエピソード拘束条件に基づいており、本研究報告ではスケールパラメータの導入箇所が異なる 2 種類のスケール推定アルゴリズムを定式化した。また、仮想環境および実環境下での実験を通して 2 種類のアルゴリズムによる推定値の精度および安定性を検証した。その結果、スケールパラメータの導入箇所によって精度・安定性に違いが生じることが分かった。さらに、提案手法で推定したスケールパラメータを初期値としてスケールを考慮したバンドル調整を行うと、スケールパラメータの推定精度がさら

に向上することを確認した。また、提案手法の応用例として RGB-FIR カメラシステムを用いた 3 次元温度マップの構築を行った。

謝辞 この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務と公益財団法人堀科学芸術振興財団第 26 回研究助成の結果得られたものである。

参考文献

- [1] Davison, A. J., Reid, I. D., Molton, N. D. and Stasse, O.: MonoSLAM: Real-Time Single Camera SLAM, *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 29, No. 6, pp. 1052–1067 (2007).
- [2] Klein, G. and Murray, D.: Parallel tracking and mapping for small AR workspaces, *International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 225–234 (2007).
- [3] Nützi, G., Weiss, S., Scaramuzza, D. and Siegwart, R.: Fusion of IMU and Vision for Absolute Scale Estimation in Monocular SLAM, *Journal of Intelligent & Robotic Systems*, Vol. 61, No. 1, pp. 287–299 (2011).
- [4] Scaramuzza, D., Fraundorfer, F., Pollefeys, M. and Siegwart, R.: Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints, *International Conference on Computer Vision (ICCV)*, pp. 1413–1419 (2009).
- [5] Kitt, B. M., Rehder, J., Chambers, A. D., Schonbein, M., Lategahn, H. and Singh, S.: Monocular Visual Odometry using a Planar Road Model to Solve Scale Ambiguity, *European Conference on Mobile Robots* (2011).
- [6] Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision (IJCV)*, Vol. 60, No. 2, pp. 91–110 (2004).
- [7] Bay, H., Tuytelaars, T. and Van Gool, L.: SURF: Speeded Up Robust Features, *European Conference on Computer Vision (ECCV)*, pp. 404–417 (2006).
- [8] Rublee, E., Rabaud, V., Konolige, K. and Bradski, G.: ORB: an efficient alternative to SIFT or SURF, *International Conference on Computer Vision (ICCV)*, pp. 2564–2571 (2011).
- [9] Agarwal, S., Snavely, N., Simon, I., Seitz, S. M. and Szeliski, R.: Building Rome in a day, *International Conference on Computer Vision (ICCV)*, pp. 72–79 (2009).
- [10] Schönberger, J. L. and Frahm, J.-M.: Structure-from-Motion Revisited, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113 (2016).
- [11] Furukawa, Y. and Ponce, J.: Accurate, Dense, and Robust Multi-View Stereopsis, *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 32, No. 8, pp. 1362–1376 (2010).
- [12] Jancosek, M. and Pajdla, T.: Multi-view Reconstruction Preserving Weakly-supported Surfaces, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3121–3128 (2011).
- [13] Schönberger, J. L., Zheng, E., Pollefeys, M. and Frahm, J.-M.: Pixelwise View Selection for Unstructured Multi-View Stereo, *European Conference on Computer Vision (ECCV)*, pp. 501–518 (2016).
- [14] DeTone, D., Malisiewicz, T. and Rabinovich, A.: Toward Geometric Deep SLAM, *arXiv preprint arXiv:1707.07410* (2017).
- [15] Han, X., Leung, T., Jia, Y., Sukthankar, R. and Berg, A. C.: MatchNet: Unifying Feature and Metric Learning for Patch-Based Matching, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3279–3286 (2015).
- [16] Zagoruyko, S. and Komodakis, N.: Learning to Compare Image Patches via Convolutional Neural Networks, *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4353–4361 (2015).
- [17] Grisetti, G., Kummerle, R., Stachniss, C. and Burgard, W.: A Tutorial on Graph-Based SLAM, *Intelligent Transportation Systems Magazine*, Vol. 2, No. 4, pp. 31–43 (2010).
- [18] Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K. and Burgard, W.: g2o: A General Framework for Graph Optimization, *International Conference on Robotics and Automation (ICRA)*, pp. 3607–3613 (2011).
- [19] Strasdat, H., Montiel, J. and Davison, A. J.: Scale drift-aware large scale monocular SLAM, *Robotics: Science and Systems (RSS)* (2010).
- [20] Oreifej, O., Cramer, J. and Zakhor, A.: Automatic generation of 3d thermal maps of building interiors, *ASHRAE transactions*, Vol. 120, p. C1 (2014).
- [21] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S. and Fitzgibbon, A.: KinectFusion: Real-time dense surface mapping and tracking, *International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 127–136 (2011).
- [22] Mller, A. O. and Kroll, A.: Generating High Fidelity 3-D Thermograms With a Handheld Real-Time Thermal Imaging System, *IEEE Sensors Journal*, Vol. 17, No. 3, pp. 774–783 (2017).
- [23] Vidas, S., Moghadam, P. and Bosse, M.: 3D thermal mapping of building interiors using an RGB-D and thermal camera, *International Conference on Robotics and Automation (ICRA)*, pp. 2311–2318 (2013).
- [24] Ham, Y. and Golparvar-Fard, M.: An automated vision-based method for rapid 3D energy performance modeling of existing buildings using thermal and digital imagery, *Advanced Engineering Informatics*, Vol. 27, No. 3, pp. 395–409 (2013).
- [25] Phuc Truong, T., Yamaguchi, M., Mori, S., Nozick, V. and Saito, H.: Registration of RGB and Thermal Point Clouds Generated by Structure From Motion, *International Conference on Computer Vision Workshop (ICCVW)* (2017).
- [26] Hartley, R. I. and Zisserman, A.: *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition (2004).
- [27] Triggs, B., McLauchlan, P. F., Hartley, R. I. and Fitzgibbon, A. W.: Bundle Adjustment — A Modern Synthesis, *Vision Algorithms: Theory and Practice*, pp. 298–372 (1999).
- [28] Nistér, D.: An efficient solution to the five-point relative pose problem, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 26, No. 6, pp. 756–770 (2004).
- [29] Stewénius, H., Engels, C. and Nistér, D.: Recent developments on direct relative orientation, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 60, pp. 284–294 (2006).
- [30] Zhang, Z.: A Flexible New Technique for Camera Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 22, pp. 1330–1334 (2000).