

# 単眼カメラ搭載移動ロボットを用いた SIFT 特徴量に基づく特徴追跡と Structure from Motion による SLAM

静岡大学 井上 渉, 川西 亮輔, 山下 淳, 金子 透

## SLAM by SIFT Feature Tracking and Structure from Motion Using a Mobile Robot Equipped with a Single Camera

Wataru Inoue, Ryosuke Kawanishi, Atsushi Yamashita and Toru Kaneko, Shizuoka University

In this paper, we propose a SLAM (Simultaneous Localization and Mapping) technique by a SIFT (Scale Invariant Feature Transform) feature tracking and a structure from motion method using a mobile robot equipped with a single camera. The method extracts and tracks feature points in an image sequence to get corresponding points in image pairs taken during the robot movement. Feature points are extracted and tracked by using a SIFT method. The relative relations of camera positions and orientations are estimated by the positions of the corresponding points. With these relations and image coordinates of the feature points, 3D coordinates of these points are calculated by triangulation. Experimental results show the effectiveness of the proposed method.

### 1 緒言

近年のロボット産業の発展に伴い、警備ロボットや介護ロボットなどの、自律移動ロボットの導入が期待されている。一般に、自律移動ロボットが行動するためには、周囲の地図情報を用いてロボットの位置を知る必要がある。しかし、自律移動ロボットが未知環境で行動する場合にはロボット自身が周囲環境を計測し、地図を生成する必要がある。このように自律移動ロボットが未知環境で行動するためには、地図生成と自己位置推定を同時に行う必要がある。地図生成と自己位置推定を同時に行う手法は総称して SLAM (Simultaneous Localization and Mapping) と呼ばれている。

SLAM では高速かつ正確な計測ができるとしてレーザ距離センサを用いることが多いが、カメラを用いた場合、距離情報だけでなく色や模様などの視覚情報も同時に取得できるため、本研究ではカメラを用いて計測を行う。

一般にステレオ計測では、カメラ間の距離（基線長）が長い方が計測精度が良い。そこで 1 台のカメラをロボットに搭載して動かし、モーションステレオを行う。これにより基線長を長く取ることができ、計測精度の向上が可能である。本手法ではカメラの移動前後の相対的な位置・姿勢を同時に推定し、被写体の計測を行う Structure from Motion (SfM) と呼ばれる計測手法 [1] により地図生成を行う。

SfM を行う場合、動画像中の物体の点や角など、周囲と比べて特徴的な点（特徴点）を抽出し、カメラの移動前後の 2 画像間で対応している点（対応点）を取得する必要がある。本手法では、SIFT (Scale Invariant Feature Transform) 特徴量 [2] を用いて特徴点の抽出、対応付けを行う。SIFT は各画素に回転・スケール変化・照明変化による画像の変化に不変な特徴量を記述するため、精度良く特徴点を追跡することが可能である。そこで本研究では、SIFT 特徴量に基づく特徴追跡による SLAM の手法を提案する。

### 2 計測手法

#### 2.1 SIFT

SIFT 特徴量は以下の手順で記述される [2]。

1. DoG (Difference of Gaussian) によりスケールと特徴点の候補点を検出する。
2. 検出された候補点の中から DoG 出力の小さい点とエッジ上の点を削除し、サブピクセル位置推定を行う。
3. 回転に不変な特徴を得るために各特徴点のオリエンテーションを求める (図 1)。ここでオリエンテーションとは特徴点周辺の輝度勾配のヒストグラムがピークになる方向である。
4. 特徴量記述領域を 1 で求めたスケールをもとに決定し、3 で求めたオリエンテーションに合わせて回転させる。そして、領域を  $4 \times 4$  の 16 ブロックに分割し、ブロックごとに輝度勾配を 8 方向に量子化し、128 次元の特徴ベクトルを記述する (図 2)。

SIFT 特徴量は、画像座標、スケール、オリエンテーション、特徴ベクトルなどの情報を持っている。記述された SIFT 特徴量を用いて各画像間で対応付けを行い、位置・姿勢推定、対応点の 3 次元計測を行う。

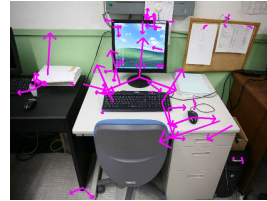


図 1: オリエンテーション

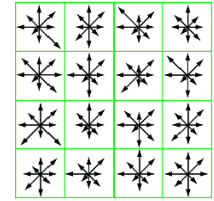


図 2: 特徴ベクトル

#### 2.2 外れ値除去

特徴点追跡によって得られた対応点はすべてが正しく対応しているとは限らない。対応が間違っている点が計測に含まれると位置・姿勢の推定に悪影響を及ぼす。そこで、位置・姿勢の推定に悪影響を及ぼす点を外れ値として除去する。これには、RANSAC [3] を用いる。

#### 2.3 位置・姿勢の推定

カメラから特徴点の 3 次元空間上の点に向かう光線のベクトルを、光線ベクトル  $\mathbf{r}$  とする。移動前後の画像の光線ベクトルをそれぞれ  $\mathbf{r}_i = [x_i, y_i, z_i]^T$ ,  $\mathbf{r}'_i = [x'_i, y'_i, z'_i]^T$  とし、カメラ間の位置・姿勢の情報からなる基本行列  $\mathbf{E}$  を求める。 $\mathbf{E}$  は式 (1) で表され、式 (1) を変形すると式 (2) が得られる。

$$\mathbf{r}'_i{}^T \mathbf{E} \mathbf{r}_i = 0 \quad (1)$$

$$\mathbf{u}_i^T \mathbf{e}_i = 0 \quad (2)$$

ただし、

$$\mathbf{u}_i = [x_i x'_i, y_i x'_i, z_i x'_i, x_i y'_i, y_i y'_i, z_i y'_i, x_i z'_i, y_i z'_i, z_i z'_i]^T$$

$$\mathbf{e} = [e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33}]^T$$

( $e_{ab}$  は  $\mathbf{E}$  の  $a$  行  $b$  列成分)

$\mathbf{E}$  は 8 点以上の対応する光線ベクトル対に対応する連立方程式、すなわち式 (3) を解いて求める。

$$\min_{\mathbf{e}} \|\mathbf{U}\mathbf{e}\|^2 \quad (3)$$

ただし、 $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]^T$  である。ここで  $n$  は特徴点数とする。 $\mathbf{e}$  は  $\mathbf{U}^T \mathbf{U}$  の最小の固有値に対応する固有ベクトルとして求められ、 $\mathbf{E}$  が求まる。 $\mathbf{E}$  より回転行列  $\mathbf{R}$  と並進移動ベクトル  $\mathbf{t}$  を求める。

#### 2.4 対応点の 3 次元座標算出

算出された回転行列  $\mathbf{R}$  と並進移動ベクトル  $\mathbf{t}$  を用いて、対応点の 3 次元計測を行う。移動前後のカメラ位置から 3 次元空間中の特徴点へ向かう光線ベクトルが交わる点  $\mathbf{p}_i$  を計測結果として算出する。

## 2.5 計測結果の限定

一般に、カメラの移動方向の延長線付近、あるいはカメラから遠い位置にある対応点ほど計測精度が悪い。そこで計測精度の良い対応点を使用するため、計測結果  $p_i$  を 2 画像上の特徴点座標  $[u_i, v_i]^T$ ,  $[u'_i, v'_i]^T$  で偏微分し、それぞれの和をとって誤差を評価する  $g_i$  を算出する (式 (4))。そしてベクトル  $g_i$  のノルムが式 (5) を満たす計測結果を以後の処理に用いる [1]。ただし、 $h$  は閾値である。

$$g_i = \left| \frac{\partial p_i}{\partial u_i} \right| + \left| \frac{\partial p_i}{\partial v_i} \right| + \left| \frac{\partial p_i}{\partial u'_i} \right| + \left| \frac{\partial p_i}{\partial v'_i} \right| \quad (4)$$

$$\|g_i\| < h \quad (5)$$

## 2.6 バンドル調整

2.3 節で推定した位置・姿勢は特徴点に含まれる誤差を考慮していないため、必ずしも良い推定結果とは限らない。そこで各特徴点の計測誤差を考慮して位置・姿勢を推定する。まず 2.3 節で推定した位置・姿勢を初期値とする。そして、特徴点の 3 次元座標値を画像上に再投影したときの座標と、元の座標の差である再投影誤差の二乗和を最小化するバンドル調整法 [4] と呼ばれる手法を行う。再投影誤差和を計算する際、特徴点ごとに計測誤差を式 (4) で表される  $g_i$  で評価する。その  $g_i$  を用いて定義される重みをつけることで、より精度の良い推定をすることができる [1]。

## 2.7 計測結果の修正・統合

各観測点間の計測結果を統合し、地図を生成する。その際、計測値には誤差が含まれている。そこで、本手法では計測値に誤差が含まれる場合にも物体の状態を精度良く推定できる拡張カルマンフィルタを用いて計測結果の統合を行う。

## 3 実験

### 3.1 精度評価実験

まず、提案手法の精度を評価するための実験を行った。直角な壁を対象に 1 回の計測を行い、カメラは手動で並進移動させ基線長は 10cm、画像の解像度は  $320 \times 240$  pixels で行った。カメラの移動前の位置と壁の角までの距離は約 55cm である。

提案手法による計測結果を上から見た図を図 3 に示す。青色の矢印はカメラの移動方向、緑色の点はカメラの移動前の位置、青色の点はカメラの移動後の位置、赤色の点は計測点である。

精度評価の結果を表 1, 2 に示す。表 1 は壁の A 面と B 面の最小 2 乗平面を求め、そこからの標準偏差を、表 2 は A 面と B 面の最小 2 乗平面の法線のなす角度を表したものである。A 面、B 面共に標準偏差が 2.5mm 前後、面のなす角が誤差 0.9deg で計測でき、提案手法で精度良く計測ができることが確認できた。

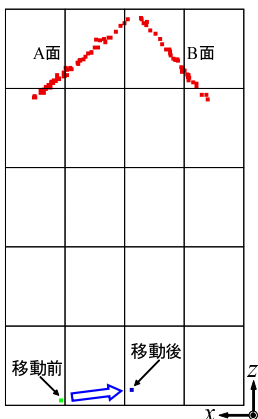


図 3: 計測結果

表 1: 標準偏差

	標準偏差
A 面	2.52mm
B 面	2.12mm

表 2: 法線のなす角

	法線のなす角
A 面-B 面	90.9deg
真値	90deg

### 3.2 移動ロボットを用いた総合実験

次に複数回計測を行い、室内環境の地図の生成を行った。実験装置にはカメラを搭載した移動ロボットを使用した (図 4)。実験は図 5 のような室内環境でロボットを直進移動させ行い、対象物 (本) と壁の計測を行った。計測は 1 フレームごとに行い全部で 18 回行った。ロボットから壁までの距離は約 150cm、画像の解像度は  $320 \times 240$  pixels、ロボットの移動速度は約 20mm/s である。CPU には Core2Duo 2.4GHz、メモリは 2GB を用いた。

提案手法による計測結果を上から見た図を図 6(a), (b) に、鳥瞰図を図 6(c) に示す。図 6(a) は 9 回目まで、図 6(b) は 18 回

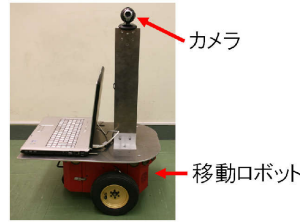
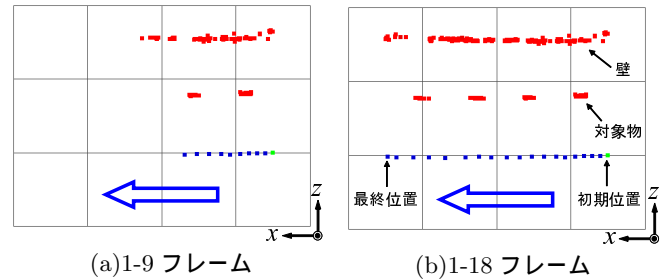


図 4: 移動ロボット



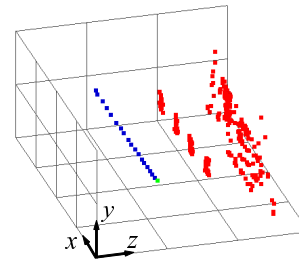
図 5: 実験環境

目までの計測結果を示している。青色の矢印はロボットの移動方向、緑色の点はロボットの初期位置、青色の点はロボットの移動軌跡、赤色の点は計測点である。なお、全フレームを通して 1 回しか観測されなかった計測点は、計測結果が不安定であるため除去した。1 回の計測の処理速度は平均 7.28sec、2 回以上計測された点は全部で 324 点であった。図 6 を見るとロボットの移動軌跡、対象物と壁の位置関係が表現されていることがわかる。



(a)1-9 フレーム

(b)1-18 フレーム



(c) 鳥瞰図

図 6: 計測結果

## 4 結言

本研究では、SIFT 特徴量に基づく特徴点追跡による SLAM の手法を提案した。SIFT 特徴による精度の良い追跡により、精度の良い計測が実現でき、本手法の有効性が示された。

今後の課題として、SIFT 特徴を利用した物体の認識、勾配情報を主成分分析 (PCA) を用いて次元圧縮することにより、高速なマッチングを行なえる PCA-SIFT [5] の導入などが考えられる。

## 参考文献

- [1] R. Kawanishi, A. Yamashita and T. Kaneko: "Construction of 3D Environment Model from an Omni-Directional Image Sequence", Proceedings of the 3rd Asia International Symposium on Mechatronics, TP1-3(2), pp.1-6, 2008.
- [2] D. G. Lowe: "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol.60 No.2 pp.91-110, 2004.
- [3] M. A. Fischler and R. C. Bolles: "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Communications of the ACM, Vol.24, pp.381-395, 1981.
- [4] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: "Bundle Adjustment - A Modern Synthesis", Vision Algorithms: Theory & Practice, Springer-Verlag LNCS Vol.1883, pp.298-372, 2000.
- [5] Y. Ke and R. Sukthankar: "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.511-517, 2004.