

# 単眼・両眼推定器を複合した自己位置と環境の同時推定 Simultaneous Localization and Mapping Using Monocular and Binocular Estimators

坂口 雄介<sup>†</sup>    Ngo Thanh Trung<sup>†</sup>    長原 一<sup>†</sup>    谷内田 正彦<sup>†</sup>  
Yusuke Sakaguchi    Ngo Thanh Trung    Hajime Nagahara    Masahiko Yachida

## 概要

自律ロボットが何らかの作業を行うには外界情報の獲得と、ロボット自身の位置を把握する必要がある。そのため、従来より未知環境におけるロボットの自己位置推定と環境マッピングを同時に行う Simultaneous Localization and Mapping (SLAM) の研究が盛んに行われている。視覚センサを用いた SLAM には両眼カメラや単眼カメラを用いた手法が提案されているが、それぞれ、一長一短がある。本論文では、単眼と両眼推定それぞれの長所を組み合わせることで、精度の高い自己位置推定と密な環境マップの生成を実現する手法を提案する。

## 1. はじめに

近年、オフィスや一般家庭における家事支援ロボットのような自律ロボットがテレビで紹介され、またそのニーズも高まっている。このようなロボットが一般的な環境において自律移動するには、周囲の環境形状とロボット自身の位置を把握する必要がある。このような、ロボットの自己位置と環境マップを同時に推定する問題は、Simultaneous Localization and Mapping (SLAM) [1][2] と呼ばれ、様々な手法が提案されている。SLAM の研究において、視覚センサは安価で受動的なセンシングであることから、これまで盛んに利用されてきた。視覚センサを用いた SLAM の手法には、大きく分けて単眼カメラと両眼カメラの2つのアプローチがある。単眼カメラを用いた手法は、ロボットの移動に伴う画像列の視差から奥行きを推定する Structure From Motion (SFM) [3][4] により、環境の構造やロボットの動きを推定するが、画像だけでは、環境マップやロボットの軌跡の実際のスケールが分からないといった問題がある。一方、両眼カメラを用いた手法 [5][6] では、両眼画像から得られる視差から、推定を行うため、ランドマークが両眼で観測される必要がある。そのため、推定されるランドマークは、オクルージョンやミスマッチングなどにより減少するため、推定される環境マップが粗になる傾向にある。

以上のような従来研究に対して本手法では、2つの全方位視覚センサを持つ移動ロボットにより得られる情報をもとに、両眼での推定と単眼での推定を複合して、自己位置推定と環境マッピングを行う手法を提案する。両眼と単眼の推定を複合することにより、両眼で対応がとれた特徴点を用いて、そのランドマークとロボットの自己位置を高精度に推定し、片方のカメラにのみ写っている特徴点も単眼でのモーションステレオ視によって推定し、密な環境マップの生成を実現する。提案手法の有効性を評価するために、実際にロボットシステムを構築し、実環境における検証を行った。

## 2. 自己位置推定と環境マッピングの原理

### 2.1 ロボットと環境設定

本手法が想定する場所は、オフィスなどの屋内環境である。このような場所では、柱などの直線的なエッジを比較的容易に多数抽出することが可能である。このエッジをランドマークとして利用することで自律ロボットの誘導を行う。移動ロボットは、進行方向と平行に取り付けられた2つの全方位視覚センサとロボットの直進速度と回転速度を測ることができるオドメトリセンサを有している。世界座標系とその座標系におけるロボットの座標設定を図1に示す。ロボットの座標はカメラ1の座標と一致し、 $(x_R, y_R, \phi)$  である。ここで  $\phi$  は2つのカメラを結んだ直線と世界座標の  $x$  軸となす角である。2つのカメラ間距離 (基線長) を  $d$  で表す。カメラ1及びカメラ2におけるランドマーク観測方位を基線となす角として、それぞれ  $\alpha_1, \alpha_2$  とする。また、ロボットの直進速度を  $V$ 、回転速度を  $\omega$  とし、ロボットの初期位置は世界座標系の原点とする。

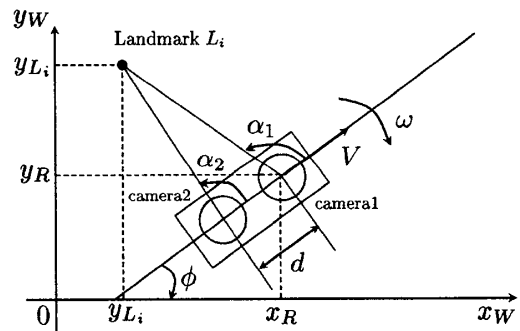


図1: ロボットの座標設定

### 2.2 提案アルゴリズム

図2に示すように、方位情報 ( $\alpha_1, \alpha_2$ ) と速度情報 ( $V, \omega$ ) を入力し、両眼で対応のとれる点を推定する。観測されたランドマークの方位情報から、両眼カメラ間で対応がとれたものを、両眼推定器に入力し、両眼での対応点を推定する。対応が取れなかったランドマークの特徴点に関しては、単眼推定器に入力し、両眼で推定されたロボットの位置をもとに、そのランドマークの位置を推定する。以上の処理を各時刻において繰り返し行うことで、各時刻のロボットの位置姿勢と環境の形状を推定する。

### 2.3 両眼対応の推定

両眼視差を用いた距離復元において、両眼画像に表われる同じランドマークの特徴点の対応を推定する必要がある。図3で表しているように、環境中のランドマーク (A, B) より得られる画像特徴点は、両眼画像それぞれ2

<sup>†</sup>大阪大学大学院基礎工学研究科

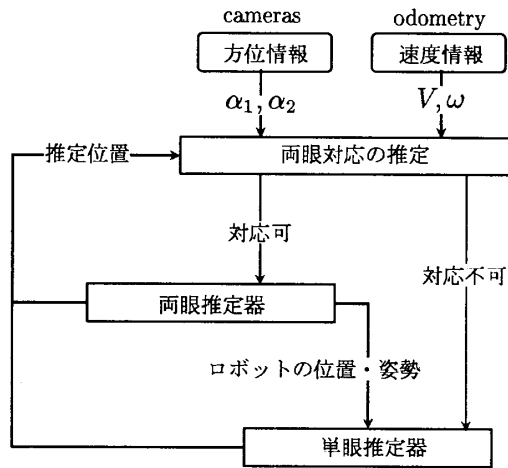


図 2: 提案手法の全体の流れ

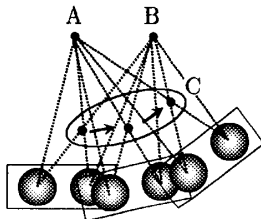


図 3: 誤った対応付けの動き

点得られ、その対応の組み合わせから得られる推定結果は3通り (A,B,C) となる。ここで、正しく対応がとれているランドマーク (A,B) の推定はほぼ静止し、誤った対応がとられているランドマーク (C) の推定はロボットの動きに伴って変位する。本研究では、複数の組み合わせを推定し、そのランドマーク推定位置の標準偏差を評価することで、両眼カメラの対応点マッチングをおこなう。図4に示す  $\beta_i[k-j]$  ( $j=0, \dots, n-1$ ) は、時刻  $k-j$  における推定ランドマーク  $L_i$  が世界座標の  $x$  軸からなす角を表し、以下のように表すことができる。

$$\beta_i[k-j] = \arctan\left(\frac{Y_{L_i}[k-j] - Y_R[k]}{X_{L_i}[k-j] - X_R[k]}\right) \quad (1)$$

観測された  $n$  枚の時系列画像列から、各時刻のランドマーク推定位置  $L_i$  から  $\beta_i$  を算出し、その平均を  $\beta_{oi}$  とする。これらを用いて、推定ランドマーク  $L_i$  の時刻  $k$  における標準偏差を次式のように表すことができる。

$$\text{angdev}(L_i, k) = \sqrt{\frac{1}{n} \sum_{j=0}^{n-1} (\beta_i[k-j] - \beta_{oi})^2} \quad (2)$$

この偏差の大きさを評価することで、両眼対応点を決定できる。

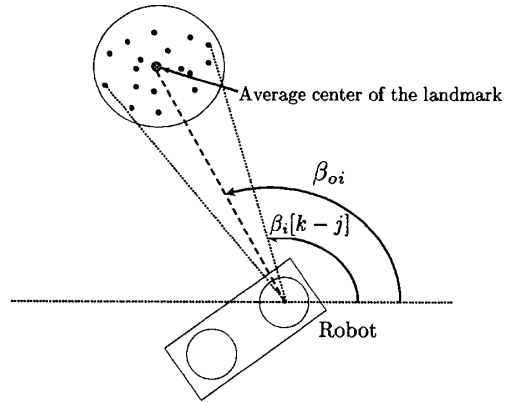


図 4: ランドマーク推定位置の偏差

## 2.4 推定アルゴリズム

本手法において、ランドマークの観測方程式は、非線形式で表されるため、これらの式をテーラー展開により線形化し、カルマンフィルタを適用する。カルマンフィルタの“時間更新”では速度情報 ( $V, \omega$ ) を用いて次の時刻でのロボットの状態を予測し、“観測更新”で視覚センサからの方位情報 ( $\alpha_1, \alpha_2$ ) で更正する。両眼対応の可否によりランドマークを分類し、以下の両眼推定器・単眼推定器のそれぞれで推定を行う。

### 2.4.1 両眼推定器

#### a. 時間更新アルゴリズム

時刻  $k$  でのロボットの位置・姿勢を  $(x_R[k], y_R[k], \phi[k])$ 、 $i$  番目のランドマークの座標を  $X_{L_i}[k] = (x_{L_i}[k], y_{L_i}[k])$  で表す。速度ベクトル  $(V[k], \omega[k])$  は時刻  $k$  における直進速度  $V$  と回転速度  $\omega$  を表しており、サンプリングタイム  $\delta t$  毎にオドメトリセンサで測定する。  $(n_V[k], n_\omega[k])$  は時刻  $k$  における速度ベクトルの測定ノイズである。時間更新による状態方程式は以下ようになる。

$$\begin{aligned} \hat{x}_R^-[k+1] &= \hat{x}_R[k] + (V[k] + n_V[k])\delta t \cos(\hat{\phi}[k]) \\ \hat{y}_R^-[k+1] &= \hat{y}_R[k] + (V[k] + n_V[k])\delta t \sin(\hat{\phi}[k]) \\ \hat{\phi}^-[k+1] &= \hat{\phi}[k] + (\omega[k] + n_\omega[k])\delta t \\ \hat{X}_{L_i}^-[k+1] &= \hat{X}_{L_i}[k] \end{aligned} \quad (3)$$

#### b. 観測更新アルゴリズム

時刻  $k$  におけるカメラ1とカメラ2でのランドマーク  $L_i$  の観測値を  $(\alpha_{1i}[k], \alpha_{2i}[k])$  と表し、観測ベクトルを  $Y[k] = (\alpha_{11}[k], \alpha_{21}[k], \dots, \alpha_{1N}[k], \alpha_{2N}[k])^T$  とする。ここで  $N$  は時刻  $k$  におけるランドマークの数である。このときの、両眼における観測方程式は図(1)より、次式のようにになる。

$$\begin{aligned}\hat{x}_{L_i}[k] &= \hat{x}_R[k] + \frac{d \sin(\alpha_{2i}[k]) \cos(\alpha_{1i}[k] - \phi[k])}{\sin(\alpha_{1i}[k] - \alpha_{2i}[k])} \\ \hat{y}_{L_i}[k] &= \hat{y}_R[k] + \frac{d \sin(\alpha_{2i}[k]) \sin(\alpha_{1i}[k] - \phi[k])}{\sin(\alpha_{1i}[k] - \alpha_{2i}[k])}\end{aligned}\quad (4)$$

### 2.4.2 単眼推定器

#### a. 時間更新アルゴリズム

単眼推定器ではランドマーク位置のみを推定する。時刻  $k$  における単眼推定器で推定する  $j$  番目のランドマーク座標を  $L_j = (x_{L_j}[k], y_{L_j}[k])$  とおくと、状態方程式は次式の様に記述できる。

$$\hat{X}_{L_j}^-[k] = \hat{X}_{L_j}[k-1] \quad (5)$$

#### b. 観測更新アルゴリズム

両眼推定器においてロボットの位置姿勢を推定するため、単眼推定器では時刻  $k, k-1$  におけるカメラの位置と姿勢が既知であると仮定できる。すなわち、両眼推定器より得られる時刻  $k, k-1$  におけるロボットの位置姿勢情報  $(x_R[k], y_R[k], \phi[k])$  と  $(x_R[k-1], y_R[k-1], \phi[k-1])$ 、および、カメラ 1 における各時刻のランドマーク  $L_j$  の観測角  $\alpha_{1j}[k], \alpha_{1j}[k-1]$  より、モーションステレオによる観測方程式は次の式のように表される。

$$\begin{aligned}x_{L_j}[k] &= x_R[k] + \frac{d' \sin(\alpha_{1j}[k-1]) \cos(\alpha_{1j}[k] - \phi[k])}{\sin(\alpha_{1j}[k] - \alpha_{1j}[k-1] - \phi[k] + \phi[k-1])} \\ y_{L_j}[k] &= y_R[k] + \frac{d' \sin(\alpha_{1j}[k-1]) \sin(\alpha_{1j}[k] - \phi[k])}{\sin(\alpha_{1j}[k] - \alpha_{1j}[k-1] - \phi[k] + \phi[k-1])}\end{aligned}\quad (6)$$

ここで、 $d'$  は 2 時刻間でロボットが移動した距離で、以下のように表すことができる。

$$d' = \sqrt{(x_R[k] - x_R[k-1])^2 + (y_R[k] - y_R[k-1])^2} \quad (7)$$

カメラ 2 における単眼推定器についても、カメラ 1 と同様に定義できる。

## 3. 実験

移動ロボットには ActivMedia Robotics 社製の Pioneer 2-DXe を用い、基線長が 226mm である全方位視覚センサを取り付けた。PC のシリアルポートと RS232C ケーブルで接続し、ロボットの制御を行う。画像の取り込みには、Euresys 社製のビデオキャプチャボード Picolo Alert を使用した。本手法では、ロボットの移動環境は柱など多くの垂直エッジが存在し、平坦な床面からなる建物の人工環境を前提とする。この様な環境中の垂直エッジをランドマークの特徴点した。全方位画像に対して垂直エッジ検出を行い、その方位を入力情報として利用し

た。提案手法との比較を行うために、単眼カメラ画像に SFM を用いた SLAM 手法についても従来手法として同様に実験を行った。SFM はカメラからの画像データのみで推定を行うため、実際の距離がわからない。そこで、得られた結果のロボット軌跡からスケールファクターを求めスケールリングすることで、比較対象とした。

### 3.1 実験手法

図 5 に、本実験環境を略図として示す。垂直エッジを検出しやすいように、意図的に角を黒くした黄色い直方体の箱とロボットを配置し、 $t=0$  のロボットの位置を世界座標系の原点とした。また、図中にランドマークとして検出されると思われる箇所に点を打った。ロボットに  $x_w$  の正の方向に直進運動させるコマンドを PC から送り、前進させた。本実験では、5fps で画像を取り込み・保存した。1 回の試行において、2 つのカメラでそれぞれ約 350 枚のフレームの撮像を行い、合計 3 回実験を行った。ここまでの処理で得られたデータを入力データとして提案手法、従来手法それぞれで処理を行った。

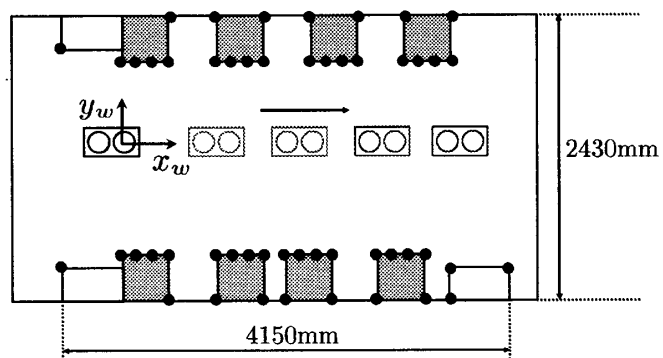


図 5: 実験環境

### 3.2 結果と考察

3 回の試行における提案手法および従来手法から得られるロボット軌跡、ランドマークの位置推定の平均誤差を表 1 に示す。ロボット軌跡は、各時刻での推定結果と真値との距離の平均をとり平均誤差とし、ランドマークは、推定結果より推定されたランドマークの ID を調べ、真のランドマークとそれに対応する推定結果との距離の平均をとり平均誤差とした。この結果から、ロボット軌跡、ランドマークともに提案手法のほうが推定の精度が高いことがわかる。

提案手法および従来手法によるロボットとランドマークの位置推定の結果の 1 つを図 6 に示す。推定結果をわかりやすくするために黄色い箱に意図的に垂直エッジをつけた位置と垂直エッジとして検出される可能性がある位置を True Landmarks とした。図 6 より、従来手法に比べ、提案手法は推定したランドマークの数が多く、生成された環境マップが密になっていることが視覚的に確かめられる。しかし、提案手法において、同じランドマークの推定を 2 回行ってしまっている箇所がある。これは、両眼推定器と単眼推定器それぞれで推定してしまったためと思われる。本手法では両眼推定器と単眼推定器でラ

ランドマークの状態変数を共有していないため、始め両眼推定器で推定していたランドマークが、後に単眼推定器で推定してしまう場合がある。また、ロボットの進行方向や後方に対して方位が小さい隣り合ったランドマークは、画像データでは非常に近くに存在するので、追跡に失敗し、同一のランドマークを2回推定したためと考えられる。このような単眼推定器と両眼推定器で重複して推定したランドマークや、追跡失敗のため重複して推定したランドマークを考慮に入れ、人工的に設けたランドマーク46個の内、推定できたランドマーク数を表2に示す。両眼推定器だけではSFMよりも少ないが、両眼・単眼を複合することで、推定できたランドマーク数は単眼推定のSFMを上回っており、提案手法はより密な環境マップを生成できたといえる。

表 2: 推定ランドマーク数

	提案手法			SFM
	両眼	単眼	複合	
1	14	30	32	26
2	15	34	35	29
3	13	33	33	27

4. おわりに

本論文では、ロボットの自己位置推定と環境マッピングを同時に行うSLAMの手法において、両眼による推定と単眼による推定を複合することによって、両眼による高精度なロボットの自己位置推定と、単眼による密な環境マップの生成を実現する手法を提案した。実際に、実環境における実験を行い、提案手法では従来手法に比べ高精度に推定でき、さらに単眼での推定を複合することで推定できるランドマークの数が増加し、環境マップがより密になることを確認した。さらに、本研究ではリアルタイム処理を目指しているため、1ステップ当たりの処理時間を計測した。2枚の入力画像から特徴点検出・追跡には約55ms、1ステップ当たりのSLAMの処理は約120msかかることがわかり、将来的には5fpsでのリアルタイム処理が可能であると思われる。

参考文献

- [1] H.P. Moravec and A. Elfes, "High resolution maps from wide angle sonar". *IEEE, International Conference on Robotics and Automation*, pp. 116- 121, 1985.
- [2] A.J. Davison, "Real-time Simultaneous Localization and Mapping with a single camera" *Pros. IEEE Intl. Conf. on Computer vision*, pp. 1403-1410, 2003.
- [3] T.J. Broida, S. Chandrashekar, R. Chellappa, "Recursive 3-D Motion estimation from a monocular image sequence", *IEEE, Trans. on Aerospace and Electronic systems*, Vol. 26, No. 4, 1990.
- [4] A.J. Azarbayejani and A. Pentland, "Recursive estimation of motion structure and focal length", *IEEE, Trans. on Pattern Analysis and Machine Intelligent*, Vol. 17, No. 6, 1995.
- [5] M.A. Garcia and A. Solanas, "3D Simultaneous localization and Modeling from Stereo Vision", *Proc. IEEE, Intl. Conf. on Robotics and Automation*, New Orleans, LA, 2004.
- [6] C. Drocourt, L. Delahoche, B. Marhic and A. Clerentin, "Simultaneous Localization and Map Construction Method using omnidirectional stereoscopic information", *Proc. IEEE, Intl. Conf. on Robotics and Automation*, Vol. 1, pp. 894- 899, Washington. DC, 2002.

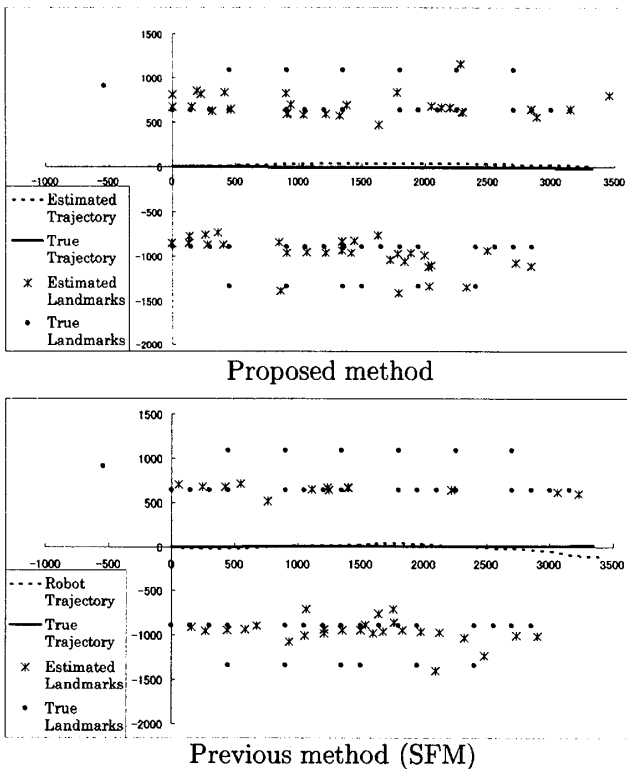


図 6: 推定結果

表 1: 平均誤差

	提案手法	平均誤差 (mm)	
		ロボット	ランドマーク
1	提案手法	39.42	136.40
	SFM	120.50	216.52
2	提案手法	64.80	144.28
	SFM	200.41	188.84
3	提案手法	31.80	104.89
	SFM	99.40	182.16