

全方位画像列を用いた 3 次元環境モデル生成

3D Environment Modeling from an Omni-directional Image Sequence

○ 川西亮輔 (静岡大) 正 山下 淳 (静岡大) 金子 透 (静岡大)

Ryosuke KAWANISHI, Shizuoka University, f0730034@ipc.shizuoka.ac.jp
Atsushi YAMASHITA, Shizuoka University
Toru KANEKO, Shizuoka University

A 3D environment model is important for tasks of autonomous mobile robots. In this paper, we propose a method for 3D environment modeling based on structure from motion using an omni-directional camera installed on a mobile robot. The method extracts and tracks feature points in an omni-directional image sequence to get corresponding points in image pairs taken during the movement of a robot. 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. The individual measurements are integrated by scale matching. By using 3D Delaunay triangulation, triangular meshes is constructed. The method makes the triangular meshes compatible with the physical shape from texture information of images. By texture-mapping, a 3D environment model is generated. Experimental results showed the effectiveness of the proposed method.

Key Words: Omni-directional Image, 3D Environment Model, Structure from Motion

1. 序論

移動ロボットが経路生成や自己位置同定を行うためには地図情報が必要である。しかし、ロボットが未知環境で自律的に行動する場合にはロボット自身が周囲環境を計測し、地図を作成する必要がある。このとき、未知環境下における自律移動ロボットの活動のための地図情報としては 3 次元の幾何学的な情報を持つ環境モデルが有効である。また、テクスチャを含むことで、より有用な情報を持つことができる。

地図生成にはカメラによって視野内の物体の計測を行うアプローチがある。これには魚眼レンズを用いたカメラによる計測[1]や周囲 360 度の視野を持つ全方位カメラによる計測[2]がある。ロボットに搭載することを考慮すると、ロボットの前後左右を一度に撮影できる全方位カメラが有効である。

周囲環境の計測には、対象の計測とロボットの位置・姿勢推定を同時に行うことができる Structure from Motion (SFM) と呼ばれる手法を用いる。本手法では、動画像中から物体の角などの特徴的な点 (特徴点) を追跡し、ロボットの移動前後の 2 画像間で対応する特徴点 (対応点) を取得する。対応点情報を元に、ロボットの位置・姿勢推定と対象の計測を行う。

環境モデルの生成には、得られた計測点群から三角網を構築し、三角網に周囲環境のテクスチャマッピングを行い、色情報を持つ環境モデルを生成する。

SFM による計測の精度向上のためには、より精度良く位置・姿勢を推定することが重要である。また三角網を構築することによるモデリングでは、計測対象の形状に合った三角網を構築する必要がある。そこで本研究では、より精度良く位置・姿勢を推定するための手法の提案および計測対象の形状に対して三角網を最適化し実環境に即した環境モデルを生成することを目的とする。

2. 処理の概要

実験装置は移動ロボットに、ビデオカメラに双曲面ミラー

を取り付けた全方位カメラを搭載したものを使用する (図 1)。移動ロボットを環境中で走行させ、走行中に撮影した動画像を用いて計測を行う。

本研究の処理手順を図 2 に示す。まず動画像を取得する。最初の画像中から特徴点を抽出し、以後のフレームで追跡を行う。追跡を行った画像間で対応点が取得できる。対応点情報を用いて線形推定により位置・姿勢を推定し、推定された位置・姿勢を用いて対応点の 3 次元座標を算出する。その後、非線形推定によって位置・姿勢をより精度良く推定する。こうして最終的に得られた位置・姿勢を用いて対応点の 3 次元座標を再度算出する。最後に複数回の計測結果を合成する。



Fig.1 Mobile Robot and Omni-directional Image

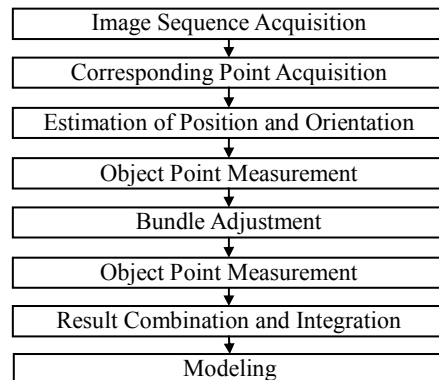


Fig.2 Process

3. 環境計測

3.1 対応点抽出

ロボットの移動前後の2画像間における対応点を取得するため、移動前の画像から特徴点を抽出、移動後の画像まで追跡する。本手法では、Lucas Kanade Trackerと画像をピラミッド構造化して探索する手法を組み合わせた特徴点追跡手法[3]を用いる。

特徴を定量的に評価したものを特徴量と呼ぶ。特徴量が大きいほど追跡が容易な点で画像間の対応がとりやすい。逆に特徴量小さい点ほど対応が取りにくいいため追跡に失敗する可能性が高い。そこで、本手法では特徴量が大きい順に特徴点を抽出する。ここで抽出する特徴点の総数を抽出数と呼ぶ。

3.2 基本行列Eの算出

カメラから3次元空間中の点に向かう光線を光線ベクトル $\mathbf{r}=[x,y,z]^T$ とする。本研究で用いる全方位カメラはカメラ前方に双曲面ミラーを装着したものであり、この光線ベクトル \mathbf{r} は、双曲面のミラー側の焦点から3次元空間中の点へ向かう光線が双曲面上で反射する点へのベクトルで表せる(図3)。

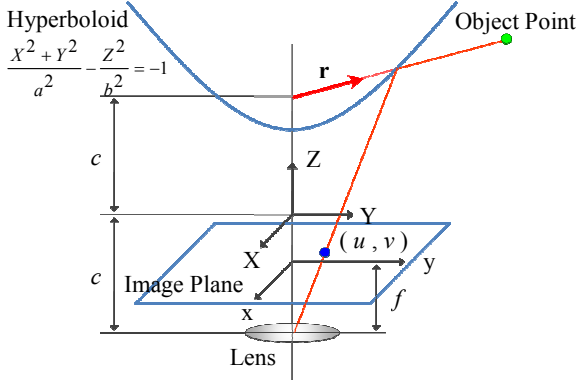


Fig.3 Calculation of Reflection Point

特徴点の画像座標 $[u,v]^T$ から式(1)(2)により \mathbf{r} を求める。ここで、 a, b, c は双曲面のパラメータであり、 f は焦点距離である。

$$\mathbf{r} = \begin{bmatrix} su \\ sv \\ sf - 2c \end{bmatrix} \quad (1)$$

$$s = \frac{a^2 \left(f \sqrt{a^2 + b^2} + b \sqrt{u^2 + v^2 + f^2} \right)}{a^2 f^2 - b^2 (u^2 + v^2)} \quad (2)$$

そして、後の演算で特徴点ごとの偏差を生じさせないように $\|\mathbf{r}\|$ の大きさを単位ベクトルに変換し、正規化する。

計測に用いる2画像間の対応点の光線ベクトル $\mathbf{r}_i=[x_i, y_i, z_i]^T$, $\mathbf{r}'_i=[x'_i, y'_i, z'_i]^T$ を用いて幾何関係によりカメラ間の位置・姿勢情報からなる行列を求める。この行列は基本行列 \mathbf{E} と呼ばれ、式(3)で表される。式(3)を変形すると式(4)が得られる。

$$\mathbf{r}'_i{}^T \mathbf{E} \mathbf{r}_i = 0 \quad (3)$$

$$\mathbf{u}_i{}^T \mathbf{e} = 0 \quad (4)$$

ただし、

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

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

(e_{ab} は \mathbf{E} の a 行 b 列成分)

基本行列 \mathbf{E} は8点以上の対応する光線ベクトル対に対する連立方程式、すなわち式(5)を解いて求める。ここで、 n は特徴点の数とする。

$$\min_{\mathbf{e}} \|\mathbf{Ue}\|^2 \quad (5)$$

ただし、 $\mathbf{U}=[\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]^T$ である。 \mathbf{e} は $\mathbf{U}^T \mathbf{U}$ の最小固有値に対応する固有ベクトルとして与えられ、基本行列 \mathbf{E} が求まる。

3.3 外れ値除去

特徴点追跡によって得られた対応点は全てが正しく対応しているとは限らない。対応を誤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 (RANdom SAMple Consensus) [4]を用いる。

まず基本行列 \mathbf{E} が決定する最小の対応点数である8個の特徴点をランダムで抽出し \mathbf{E}_{rand} を求める。そして光線ベクトル $\mathbf{r}_i, \mathbf{r}'_i$ に対して式(6)を満たす特徴点の数を k とする。

$$|\mathbf{r}'_i{}^T \mathbf{E}_{\text{rand}} \mathbf{r}_i| < q \quad (6)$$

ただし q は閾値である。この $\mathbf{E}_{\text{rand}}, k$ を求める処理をあらかじめ設定した回数繰り返す。そして k が最大となったときの式(6)を満たさない特徴点を外れ値とする。その後、外れ値を除いた特徴点を用いて式(5)より基本行列 \mathbf{E} を算出する。

3.4 抽出数の自動決定

一般に8組の特徴点があれば位置・姿勢の推定が行えるが、実際は各種の誤差の影響があり、抽出数が少ないと正確な位置・姿勢の推定は困難である。環境の詳細なモデル化のためには抽出数が多いほど良いが、抽出数が多すぎると特徴量の小さい点が抽出され、位置・姿勢推定に悪影響を及ぼす(3.1節)。そのため、RANSACによって外れ値とされる特徴点の数(外れ値数)が増加する。外れ値数が多いほど位置・姿勢の推定精度が悪く考えられる。したがって、本手法では位置・姿勢の推定精度が十分に期待できる最大の外れ値数となるとき抽出数を最適な抽出数と考え、これを自動的に決定する。ここで以下の式(7)(8)を定義する。

$$\zeta(k) = \frac{\left| w \sum_{i=k}^{k+w} i \cdot O_i - \sum_{i=k}^{k+w} i \sum_{i=k}^{k+w} O_i \right|}{w \sum_{i=k}^{k+w} i^2 - \left(\sum_{i=k}^{k+w} i \right)^2} \quad (7)$$

$$\zeta(k) < G \quad (8)$$

ここで O_i は抽出数 i のときの外れ値数、 G は閾値、 $\zeta(k)$ は抽出数が k から $k+w$ まで変化するときの外れ値数の増加量を最小二乗法で求めたものである。 w を適切に設定することで安定して外れ値数の増加量を求めることができる。

追跡が容易で画像間の対応がとりやすい特徴点の数は環境ごとに異なるが、いずれにしてもその数は限られている。したがって、適切な抽出数を超えると外れ値数が急激に増加するという現象がいずれの観測点にも共通して起こることを利用し、式(8)を満たす最大の k を抽出数として算出する。

3.5 位置・姿勢の初期推定

基本行列 \mathbf{E} は回転行列 \mathbf{R} と並進移動ベクトル $\mathbf{t}=[t_x, t_y, t_z]^T$ により式(9)で表される。

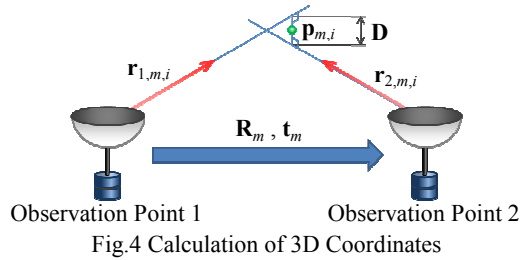
$$\mathbf{E} = \mathbf{R}\mathbf{T} \quad (9)$$

ただし、 $\mathbf{T} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$ である。

\mathbf{E} に特異値分解を行い、 \mathbf{R} と \mathbf{T} を求める。

3.6 対応点の3次元計測算出

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。移動前と移動後のそれぞれの観測点のカメラ位置から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を計測結果とする(図4)。



3.7 計測結果の限定

計測精度は一般にカメラの移動方向の延長線付近ほど、あるいはカメラから遠いものほど悪い。そこで計測精度の良い対応点を使用するため、計測結果 $\mathbf{p}_{m,i}$ を2画像上の特徴点座標 $[u_{m,i}, v_{m,i}]^T$, $[u'_{m,i}, v'_{m,i}]^T$ で偏微分したのから式(10)で誤差を評価するベクトル \mathbf{g} を求める。そしてベクトル \mathbf{g} のノルムが式(11)を満たす計測結果を以後の処理に用いる。ただし、 h は閾値である。

$$\mathbf{g} = \begin{bmatrix} \frac{\partial \mathbf{p}_{m,i}}{\partial u_{m,i}} + \frac{\partial \mathbf{p}_{m,i}}{\partial v_{m,i}} + \frac{\partial \mathbf{p}_{m,i}}{\partial u'_{m,i}} + \frac{\partial \mathbf{p}_{m,i}}{\partial v'_{m,i}} \\ \mathbf{g}_y \\ \mathbf{g}_z \end{bmatrix} \quad (10)$$

$$\|\mathbf{g}\| < h \quad (11)$$

3.8 バンドル調整

3.5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であるとは限らない。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。これには3.4節で推定した位置・姿勢を初期値とし、特徴点の再投影誤差の二乗和を最小化するバンドル調整法[5]と呼ばれ

る手法を用いる。再投影誤差和を計算する際、特徴点ごとに計測誤差を評価し適切に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。

ここで再投影誤差とは、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、元の画像座標との差である。これが小さいほど位置・姿勢の推定精度が良い。特徴点の再投影されたときの座標は式(12)~(14)より求められる。

$$\begin{pmatrix} u \\ v \\ -c+f \end{pmatrix} = d' \begin{pmatrix} dx \\ dy \\ dz+2c \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -c \end{pmatrix} \quad (12)$$

$$d = \frac{cz + b\sqrt{x^2 + y^2 + z^2}}{\left(\frac{b}{a}\right)^2 (x^2 + y^2) - z^2} \quad (13)$$

$$d' = \frac{f}{dz + 2c} \quad (14)$$

画像の再投影誤差和を以下のように定義する。

$$E_f = \sum_p r_{fp} \|\mathbf{x}_{fp} - \mathbf{x}'_{fp}\|^2 \quad (15)$$

ここで、 E_f は第 f フレームでの再投影誤差和、 r_{fp} は第 f フレームの p 番目の特徴点に対する重み係数、 \mathbf{x}_{fp} は特徴点の元の画像座標、 \mathbf{x}'_{fp} は再投影したときの画像座標である。重み係数は特徴点ごとに誤差を評価することにより決定する。誤差の評価には3.7節の式(10)で表されるベクトル \mathbf{g} のノルムを用い、以下の式で定義する。

$$r_{fp} = \frac{1}{\|\mathbf{g}_{fp}\|} \quad (16)$$

ここで、 \mathbf{g}_{fp} は第 f フレームの p 番目の対応点の誤差ベクトルである。これにより対応点の計測誤差を考慮し、適切に重みをつけることができる。

3.9 計測結果の合成・統合

各観測点間で得られた計測結果を、各観測点の位置・姿勢推定結果を用いて重ね合わせ、合成する。

同じ特徴点を示す計測結果でも、含まれる誤差の違いから観測点間ごとに計測結果は異なる。したがって、各観測点間の計測結果を重ね合わせるとずれが生じ、同じ特徴点を示す計測結果が複数存在することになる。そこで本手法では空間をボクセルで区切り、各特徴点の計測誤差を考慮した評価値(式(16))で重み付けをした投票を行うことで、特徴点の3次元位置を推定し計測結果を統合する。

4. モデリング

4.1 三角網の生成

得られた計測点群から3次元のドロネー分割により三角網を構築する。しかし、ドロネー分割によって得られる多面体構造は計測対象の形状を考慮していないため、物理的な辺に

矛盾する三角網が生成されることがある。

そこで本手法では、矛盾を検出するテンプレートを導入して三角網を物理的な形状に適合するように最適化する手法[6]を適用する(図5)。これにより実際の環境と矛盾しないモデルを生成することができる。

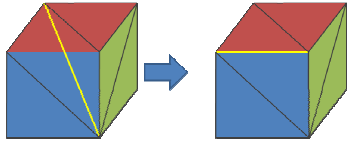


Fig.5 Optimizing Triangular Mesh

4.2 テクスチャの貼り付け

各観測点で撮影した画像の中でテクスチャの解像度が最大となる画像を選び、三角網の各面に貼り付けることで、モデルにより詳細な色情報をもたせることができる。

5. 実験

5.1 実験環境

計測は屋内環境(L字廊下, 図6)で行った。ロボットの移動速度は約10cm/sで、移動中に全方位カメラで撮影した5fpsの画像列を入力とした。入力画像のサイズは1920×1080pixelsである。

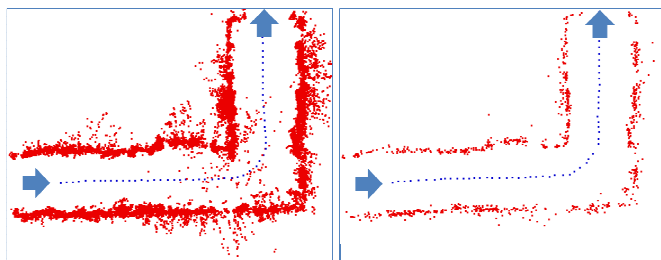


Fig.6 Environment Image

5.2 実験結果

提案手法により得られた計測結果を以下に示す(図6)。青い点がロボットの軌跡、赤い点が計測点である。提案手法(3.4, 3.8節)を適用しない場合(a)に比べ、提案手法を適用した場合(b)の結果は、計測結果のばらつきが少なく、廊下形状を正確に表しており、精度の良い計測ができていていることが分かる。

- Measurement Point
- Observation Point



(a) Result without Our Method (section3.4, 3.8) (b) Result with Our Method

Fig.6 Integrated Measurement Data

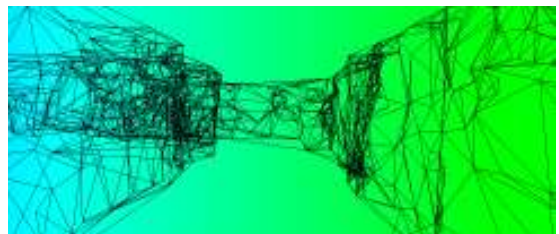
提案手法により得られた三角網および3次元環境モデルを以下に示す(図7)。生成された環境モデルに貼り付けられた

テクスチャの歪みが少ないことから、実環境と矛盾しない三角網が構築できているといえる。しかし、オクルージョンなどの影響から計測点が疎な部分が存在し、モデル化が困難となる問題がある。これは今後改善する必要がある。

6. 結論

本論文では、より精度良く位置・姿勢推定をするための手法の提案をした。また、計測対象の形状に対して三角網を最適化する手法を適用した。その結果、計測精度および環境モデルの精度が向上し、本手法の有効性が示された。

今後の課題としては、複数の観測点の位置・姿勢を同時に推定する手法の構築、計測点が疎な部分に新たに計測点を追加する手法の構築などが考えられる。



(a) Wire Frame



(b) 3D Model

Fig.7 Result of 3D Modeling

文献

- [1] 宮川勲, 若林佳織, 荒川賢一: "魚眼投影モデルに基づく画像系列からの平面運動と3次元形状の復元", 電子情報通信学会論文誌 D-II, Vol.J87-D-II, No.5, pp.1120-1132, 2004.
- [2] Ronald Bunschoten and Ben Krose: "Robust Scene Reconstruction from an Omnidirectional Vision System", IEEE Transactions on Robotics and Automation, Vol.19, No.2, pp.351-357, 2003.
- [3] J. Y. Bouguet: "Pyramidal Implementation of the Lucas Kanade Tracker Description of the Algorithm", Intel Corporation, 2000.
- [4] 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.
- [5] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: "Bundle Adjustment - A Modern Synthesis", Vision Algorithms: Theory & Practice, Springer-Verlag LNCS 1883, 2000.
- [6] 中辻敦忠, 菅谷保之, 金谷健一: "ビデオ画像からの形状復元のための三角網の最適化", 情報処理学会研究報告 CVIM-149-2, pp.7-14, 2005.