

全方位カメラを用いた 3 次元環境モデリング

川西亮輔[†] 山下淳[†] 金子透[†]

[†] 静岡大学工学部 〒432-8561 静岡県浜松市中区城北 3-5-1
E-mail: †{f0730034,tayamas,tmtkane}@ipc.shizuoka.ac.jp

あらまし 本論文では全方位画像列を用いて周囲環境のモデルを構築する手法を提案する。カメラの移動前後で取得した 2 画像間における対応点情報を用い、撮像された物体の計測とカメラの位置・姿勢推定を同時に行う Structure from Motion (SFM) により周囲環境の計測を行う。計測点群から構築した三角網にテクスチャマッピングを施すことで環境モデルを生成する。本研究では計測精度の向上のためにカメラの位置・姿勢の効果的な推定手法を提案する。また、より実環境に即した環境モデルを生成するために三角網の最適化手法を適用する。実験結果より、本手法の有効性が示された。

キーワード 全方位カメラ, Structure from Motion, 三角網, 環境モデル

Construction of 3D Environment Model by an Omni-Directional Camera

Ryosuke KAWANISHI[†], Atsushi YAMASHITA[†], and Toru KANEKO[†]

[†] Faculty of Engineering, Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka 432-8561, Japan
E-mail: †{f0730034,tayamas,tmtkane}@ipc.shizuoka.ac.jp

Abstract Map information is important for path planning and self-localization when mobile robots accomplish autonomous tasks. In unknown environments, however, mobile robots should generate an environment map by themselves. We propose a method for 3D environment modeling by a mobile robot. A 3D environment model can be generated from the result of 3D measurement using image data. To realize a 3D measurement of objects more efficiently, the robot uses an image sequence acquired by an omni-directional camera which has a wide field of view. The measurement method is based on structure from motion (SFM). A triangular mesh is constructed from measurement data. Experimental results showed the effectiveness of the proposed method.

Key words Omni-Directional Camera, Structure from Motion, Triangular Mesh, Environment Model

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。

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 [1]。しかし、一般的な視野のカメラを用いてロボットの周囲 360° の計測を行うには撮影回数を多くする必要があり [2]、1 枚ごとに撮像できる視野が狭いため画像間の対応付けの処理が煩雑になるという問題が生じる。この問題に対して、広い視野を持つ撮像センサが考案されている。これには魚眼レンズを用いたカメラによる計測 [3] や周囲 360° の視野を

持つ全方位カメラによる計測 [4] [5] 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。全方位カメラは広い視野を有するという特徴により周囲環境の計測や認識に有効であることが示されている [6]。

全方位カメラを用いた計測手法として、複数の全方位カメラによりステレオ計測を行う手法 [7] [8] がある。一般にステレオ計測はカメラ間の距離（基線長）が長いほど精度が良いが、複数の全方位カメラをロボットに搭載することを考えた場合、基線長を長くすることは困難である。そこで、1 台のカメラを搭載したロボットを移動させることにより基線長を調節できる計測手法 [9] であるモーションステレオが有効だと考えられる。

また、より高精度な計測を行うために、全方位カメラと GPS [10] や超音波 [11]、レーザ [12]などを同時に使用し、センサフュージョンにより計測を行う手法がある。しかしセンサ

フュージョンは、計測対象・状況が限定される、キャリブレーションの煩雑さが増す、装置構成が大きくなるなどの問題がある。

そこで本手法では、カメラ1台で周囲環境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手法である Structure from Motion (SFM) を用いる。まず動画像中で物体の角などの視覚的に特徴のある点(特徴点)を抽出・追跡を行い、ロボットの移動前後の2画像間に対応する特徴点(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周囲環境の計測を行う。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うことで、色情報を持つ環境モデルを生成することができる。

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を提案し、計測対象の形状に対して三角網を最適化することで実環境に即した環境モデルを生成する。

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。

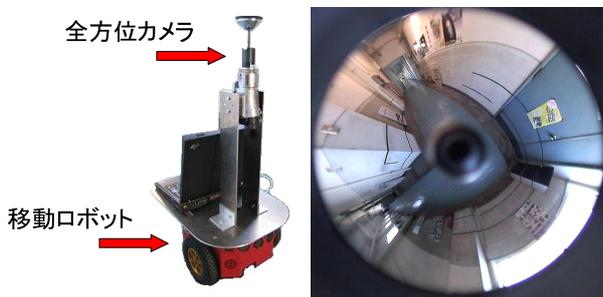


図1 移動ロボットと全方位画像

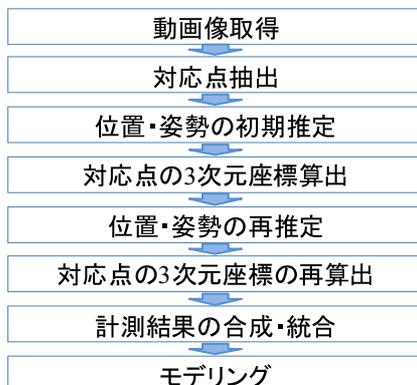


図2 処理の流れ

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に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。

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[13]を用いる(図3)。

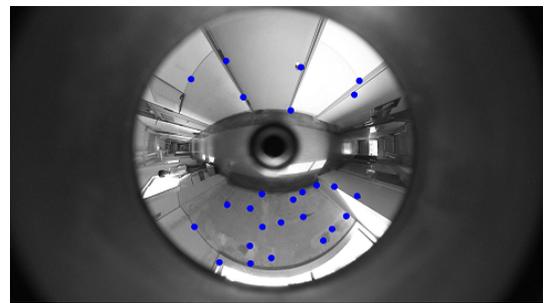


図3 特徴点抽出

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。逆に特徴量が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4)。

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

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 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．

$$\tilde{\mathbf{r}}_2^T \mathbf{E} \tilde{\mathbf{r}}_1 = 0 \quad (3)$$

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

ただし，

$$\mathbf{u} = [x_1 x_2, y_1 x_2, z_1 x_2, x_1 y_2, y_1 y_2, z_1 y_2, x_1 z_2, y_1 z_2, z_1 z_2]^T$$

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

(e_{ab} は E の a 行 b 列成分)

基本行列 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．

$$\min_e \|\mathbf{U} \mathbf{e}\|^2 \quad (5)$$

ただし， $\mathbf{U} = [u_1, u_2, \dots, u_n]^T$ 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．

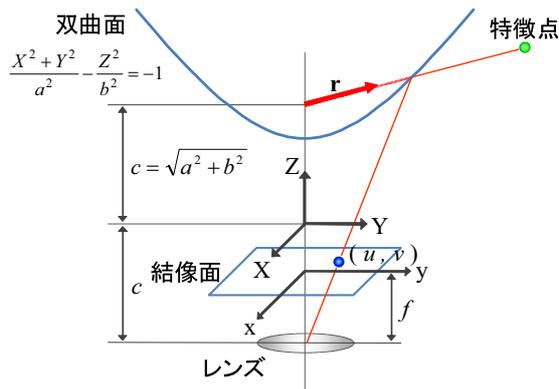


図 4 全方位カメラの座標系

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)[14] を用いる．

まず基本行列 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 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．

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

ただし q は閾値である．この E_{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 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出た場合，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いと考えられる．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．

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

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

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起こる．これを利用し，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 E は回転行列 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， $\mathbf{T} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$ である．

E に特異値分解を行い， R と T を求める．

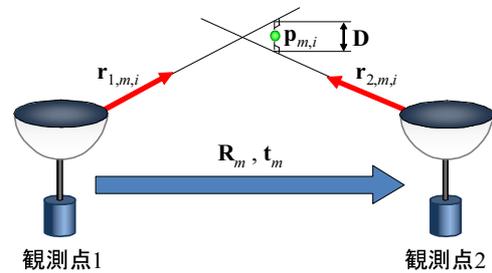


図 5 対応点の 3 次元計測

3.6 対応点の 3 次元座標算出

算出された回転行列 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 5)．

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。

$$\mathbf{g} = \left| \frac{\partial \mathbf{p}_{m,i}}{\partial u_{1,m,i}} \right| + \left| \frac{\partial \mathbf{p}_{m,i}}{\partial u_{2,m,i}} \right| + \left| \frac{\partial \mathbf{p}_{m,i}}{\partial v_{1,m,i}} \right| + \left| \frac{\partial \mathbf{p}_{m,i}}{\partial v_{2,m,i}} \right| \quad (10)$$

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

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 [15] 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。

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

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

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

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 [16] を適用する (図 6)。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。

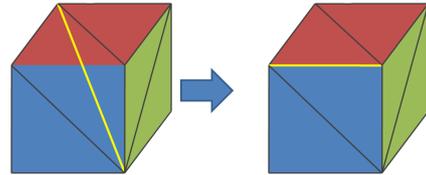


図 6 三角網の最適化

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

屋内環境 (図 7(a) L 字廊下, (b) 部屋) でそれぞれ計測を行った。L 字廊下における計測では全方位カメラを搭載した移動ロボットを約 10cm/s の速度で走行させながら撮影を行い、フレームレートは 5fps とした。部屋における計測では人が全方位カメラを手を持って移動しながら撮影を行い、フレームレートは 10fps とした。ともに入力画像サイズは 1920 × 1080pixels である。



(a) L 字廊下

(b) 部屋

図 7 実験環境

廊下の計測結果を上から見た図 8 に示す。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 (3.3, 3.4, 3.7, 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ることが分かる。

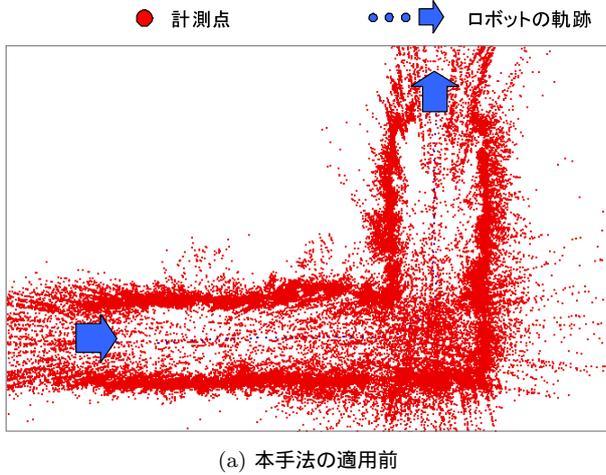


図 8 廊下の計測結果 (上視点)

提案手法により得られた廊下の 3 次元環境モデルの鳥瞰図および各視点図を図 9, 図 10 に示す。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。

部屋の計測結果を上から見た図 11 に示す。提案手法 (3.3, 3.4, 3.7, 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、人物の進行方向の延長線付近の精度の悪い計測結果が除去されており、部屋の壁や部屋内に存在する物体の形状を計測することができる。

提案手法により得られた部屋の 3 次元環境モデルの各視点図を図 12 に示す。視点の変化にも対応しており、部屋内の物体の形状がモデル化できている。

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、より正確にカメラの位置・姿勢を推定するための手法として、複数の観測点の位置・姿勢を同時に推定する手法の構築が考えられる。また、より実環境に忠実なモ

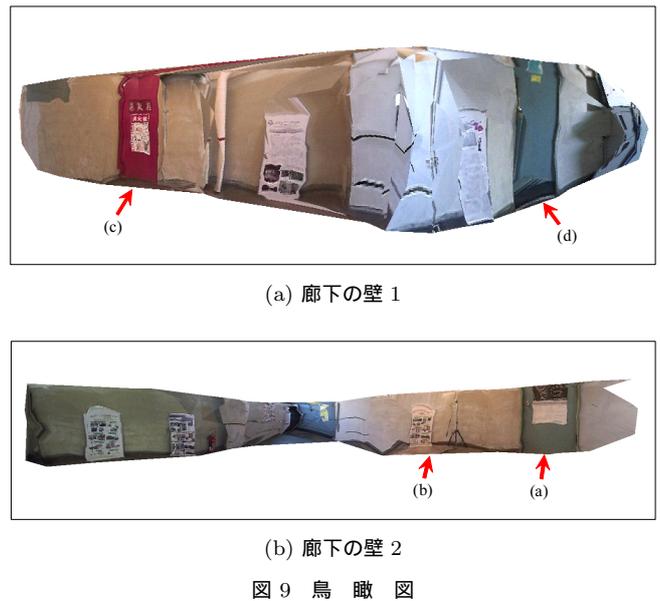
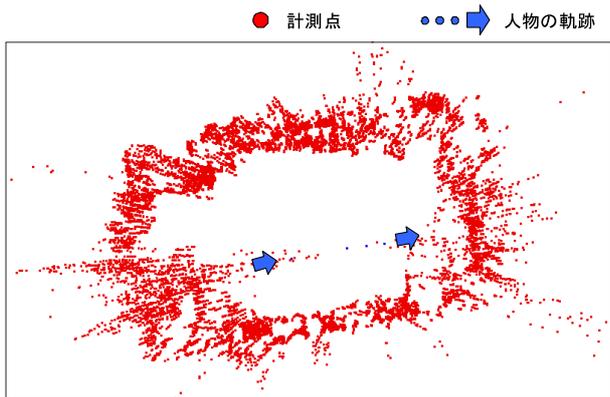
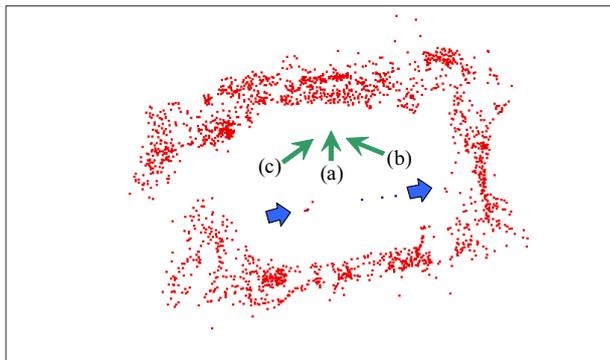


図 10 廊下のモデリング結果 (左: 実際の環境, 右: 環境モデル)



(a) 本手法の適用前



(b) 本手法の適用後

図 11 部屋の計測結果（上視点）



視点 (a)



視点 (b)



視点 (c)

図 12 部屋のモデリング結果（左：実際の環境，右：環境モデル）

デルを生成するために，計測点数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．

謝 辞

本研究の一部は，文部科学省 科学研究費補助金 若手研究 (B) 20700184，および財団法人 立石科学技術振興財団の補助を受けた．

文 献

- [1] A. J. Davison: " Real-Time Simultaneous Localisation and Mapping with a Single Camera ", Proceedings of 9th IEEE International Conference on Computer Vision, Vol. 2, pp. 1403-1410, 2003.
- [2] H. Ishiguro, M. Yamamoto and S. Tsuji: " Omni-Directional Stereo ", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 257-262, 1992.
- [3] 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, Vol. J87-D-II, No. 5, pp. 1120-1132, 2004.
- [4] R. Bunschoten and B. Krose: " Robust Scene Reconstruction from an Omnidirectional Vision System ", IEEE Transactions on Robotics and Automation, Vol. 19, No. 2, pp. 351-357, 2003.
- [5] C. Geyer and K. Daniilidis: " Omnidirectional Video ", The Visual Computer, Vol. 19, No. 6, pp. 405-416, 2003.
- [6] J. Gluckman and S. K. Nayar: " Ego-motion and Omnidirectional Cameras ", Proceedings of 6th International Conference on Computer Vision, pp. 999-1005, 1998.
- [7] J. Takiguchi, M. Yoshida, A. Takeya, J. Eino and T. Hashizume: "High Precision Range Estimation from an Omnidirectional Stereo System", Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.263-268, 2002.
- [8] 石川 智也, 山澤 一誠, 横矢 直和: " 複数の全方位動画からの自由視点画像生成 ", 画像の認識・理解シンポジウム (MIRU2005) 講演論文集, pp. 1396-1403, 2005.
- [9] M. Tomono: " 3-D Localization and Mapping Using a Single Camera Based on Structure-from-Motion with Automatic Baseline Selection ", Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 3353-3358, 2005.
- [10] J. Meguro, Y. Amano, T. Hashizume and J. Takiguchi: " Omni-Directional Motion Stereo Vision Based on Accurate GPS/INS Navigation System ", Proceedings of 2nd Workshop on Integration of Vision and Inertial Sensors, 2005.
- [11] S. Wei, Y. Yagi and M. Yachida: " Building Local Floor Map by Use of Ultrasonic and Omni-Directional Vision Sensor ", Proceedings of the 1998 IEEE International Conference on Robotics and Automation, pp. 2548-2553, 1998.
- [12] 浅井 俊弘, 神原 誠之, 横矢 直和: " 全方位距離画像と全方位カラー画像の統合による屋外環境の三次元モデル化 ", 画像電子学会誌, Vol. 34, No. 5, pp. 529-538, 2005.
- [13] J. Y. Bouguet: " Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm ", OpenCV, Intel Corporation, 2000.
- [14] 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, No. 6, pp. 381-395, 1981.
- [15] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: " Bundle Adjustment -A Modern Synthesis ", Vision Algorithms: Theory & Practice, Springer-Verlag LNCS 1883, 2000.
- [16] 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.