

多画像からの3次元復元のための 復元精度を考慮した画像選択

酒匂祐太郎^{1,a)} 菅谷保之^{1,b)}

概要: 多画像からシーンの3次元復元を行う際の復元精度を考慮した画像選択手法を提案する。近年, Photo Tourism を基にした3次元復元手法が多く提案されている。これらの多くは, 多数の入力画像から1枚ずつ復元に使用する画像を追加しながらシーン全体の3次元復元を実現するものである。本研究では, 復元に用いる画像間の視差を用いた画像選択の指標を提案し, 従来手法よりも効率的に精度の高い3次元復元が実現できることを示す。シミュレーションデータと実画像を用いた実験により, 提案手法の有効性を確認した。

Image selection method for accurate 3-D reconstruction from a large collection of images

YUTARO SAKO^{1,a)} YASUYUKI SUGAYA^{1,b)}

Abstract: In order to reconstruct the 3-D shape of an object from a large collection of images, we propose a new image selection method for reconstructing the 3-D shape accurately. Recently, many 3-D reconstruction methods based on the Photo tourism are proposed. They first reconstruct a 3-D shape from two images selected from input images, and then incrementally reconstruct the whole 3-D shape by adding an image one by one. In general, the accuracy of the reconstructed 3-D shape becomes high by using images whose disparity between them are large. In this consideration, we compute disparities among all image pairs and reconstruct the 3-D shape one by one by using the image whose disparity is the largest. Using simulation data and real images, we confirmed that our proposed method was effective for 3-D reconstruction from large collections of images.

1. はじめに

近年, Photo Tourism[5] に代表されるような, あるシーンをさまざまな視点から撮影した多数枚の画像からそのシーンの3次元復元をする研究が盛んに行われている。SnavelyらはPhoto Tourismの技術をオープンソースのソフトウェアBundler[5], [6], [7]として公開している。Bundlerでは, まず2枚の画像から5点アルゴリズム[4]によって計算した基礎行列を用いて対応点の3次元復元を行う。その復元結果を利用して, Direct Linear Transformation法(以下ではDLT法と呼ぶ)[2]により新たな画像のカメラ位置を推定し, その画像に含まれる対応点の3次元復元を行う。また, その結果を初期値としてバンドル調整[8]を行うことで, カメラ位置および3次元復元結果の精度を向上させる。これをくり返すことで, 全体の3次元復元を行う。

この手法ではDLT法によって計算する値は精度よりも安定性を重視しており, バンドル調整を新たな画像を追加するたびにを行うことで高精度な復元を可能にしている。しかし, バンドル調整の回数が多いため計算コストが高いという問題がある。

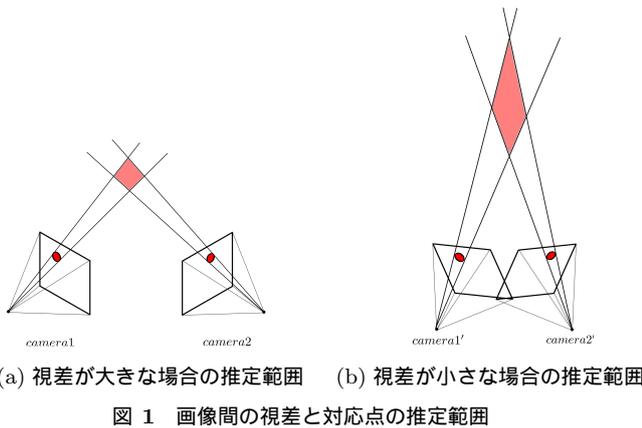
本研究では, Snavelyらの多画像からの3次元復元の枠組みの中で, DLT法による3次元復元精度が向上する画像の選択手法を提案する。Bundlerのようなシステムでは, 追加画像のカメラ運動の推定精度はそれまでの復元結果に依存し, 3次元復元の精度はカメラ運動の推定結果に依存するため, 画像を追加する順番は最終的な復元結果に大きく影響を与える。特に, 画像枚数が増えるほど画像選択が重要になる。また特徴点対応に基づく3次元復元では, 復元に用いる画像間の視差が大きいほうが復元精度がよくなる特徴がある。

本研究ではまず, 復元済みの画像群との視差を考慮して追加する画像を選択することで, 復元精度が向上することをシミュレーション実験により示す。そして, Bundlerと比べてバンドル調整の回数が少なくても, 同等の結果が得ら

¹ 豊橋技術科学大学情報・知能工学系
Department of Computer Science and Engineering, Toyohashi University of Technology

a) sako@iim.cs.tut.ac.jp

b) sugaya@iim.cs.tut.ac.jp



れることを実画像実験により示す。

2. 追加画像の選択手法

特徴点对応に基づく 3 次元復元では、一般に画像間の視差が大きいが対応点の復元精度が良くなる特徴がある。図 1 は 2 画像から対応点の 3 次元位置を三角測量で求める際の 3 次元位置の推定範囲を示している。図 1(a) はカメラ間の視差が大きな場合、図 1(b) はカメラ間の視差が小さな場合を示している。視差が小さな場合は奥行き方向に推定範囲が広くなり、視差が大きいが推定精度が高くなる。そこで、本研究では復元済みの画像群との視差が大きくなる画像を優先的に選択する手法を提案する。また、画像間の視差だけでなく、対応点の数も考慮して画像選択を行う。

2.1 視差の計算

3 次元シーンは一般には射影変換では表現できないが、2 つのカメラ間の位置姿勢の変化が小さければ対応点間には射影変換がある程度成立すると考えられる。そこで、本研究では射影変換を満たす対応点の割合に注目して画像間の視差を測る。

具体的には、RANSAC[1] により対応点から射影変換を計算し、そのインライアの割合を計算する。射影変換のインライアの割合から視差の大きさを表すスコアを計算する。対応点を次のように 3 次元の同次座標系で表現する。

$$\mathbf{x}_\alpha = \begin{pmatrix} x/f_0 \\ y/f_0 \\ 1 \end{pmatrix}, \mathbf{x}'_\alpha = \begin{pmatrix} x'/f_0 \\ y'/f_0 \\ 1 \end{pmatrix}, \alpha = 1, \dots, N \quad (1)$$

ここで、 f_0 は画像サイズ程度に設定するスケール定数である。この 2 点間に射影変換が成立する場合、次式を満たす。

$$\mathbf{x}'_\alpha = Z[\mathbf{H}\mathbf{x}_\alpha] \quad (2)$$

ここで、 \mathbf{H} は射影変換行列、 $Z[\cdot]$ はベクトルの第 3 成分を 1 とする正規化作用素である。そこで、画像サイズの 0.4% をしきい値として、次式を満たす対応点をインライアと判

断する。

$$\|\mathbf{x}'_\alpha - Z[\mathbf{H}\mathbf{x}_\alpha]\| \leq \frac{0.004 \max(W, H)}{f_0} \quad (3)$$

W, H はそれぞれ画像の幅と高さを表し、 $\max(a, b)$ は a, b のうち大きい値を返す関数である。そして、計算した射影変換に対して対応点のインライアの割合を r とし、視差の大きさの指標 $score_d$ を式 (4) で求める。

$$score_d = \begin{cases} 0.1 & (r = 0) \\ \frac{1}{r} & (otherwise) \end{cases} \quad (4)$$

2.2 追加画像の選択

復元済みの画像群との視差が大きく、対応点が多い画像を選択する方法について述べる。

まず復元済みの画像群との視差が大きい画像を選択するために、全ての復元済みの画像と未復元の画像間の視差の大きさの指標 $score_d$ を式 (4) により計算し、その平均値を $Score_d$ とする。 $Score_d$ が大きいほど復元済みの画像群との視差が大きいことを表す。そして、既知の 3 次元点との対応点数を $Score_m$ とする。 $Score_d$ と $Score_m$ が両方大きくなるような画像を選択することで復元済みの画像群との視差が大きく、対応点が多い画像を選択する。追加画像を選択する手順を以下にまとめる。

- (1) $Score_d$ を各画像について求める。
- (2) 各画像と既知の 3 次元点との対応点数を $Score_m$ として求める。
- (3) $Score_d, Score_m$ のそれぞれの最大値、最小値を max_d, max_m および min_d, min_m として求める。
- (4) $Score = \frac{Score_d - min_d}{max_d - min_d} + \frac{Score_m - min_m}{max_m - min_m}$ とする。
これは、対応点数と視差をそれぞれ 0 から 1 に正規化して足しあわせたものである。

3. シミュレーション実験

画像間の視差の大きさを考慮した画像を優先して追加することによる 3 次元復元の精度向上を確認するため、全ての画像間で曲面格子の全特徴点が得られており、各画像間で特徴点の対応が正しいような状況で復元済みの画像との視差が大きいものを選んだ場合とそうでない場合を比較した。また、対応点数の変化に対する 3 次元復元精度の変化も調べた。

3.1 シミュレーションデータ

3.1.1 曲面格子

シミュレーションデータとして、以下に示す曲面格子を透視投影したものをを用いる。式 (5) は中心が X 軸に平行で、原点 $(0, 0, 0)$ を通り、格子間隔が D の 6×12 の $1/4$ 円筒格子を Z 軸回りに ϕ 度回転したものである。曲面格子を図 2 に示す。実験では、 $D = 1400, \phi = 30$ とする。

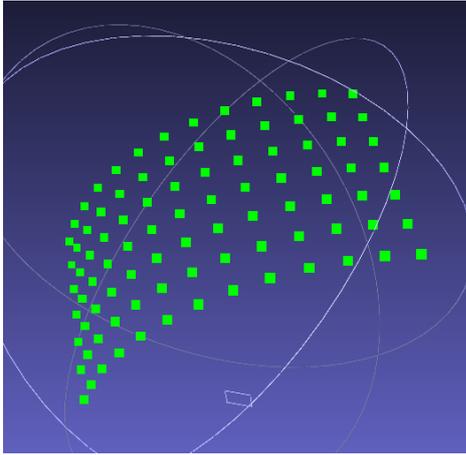


図 2 曲面格子

$$r = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} lD \\ \frac{24D}{\pi} \sin \frac{m\pi}{24} \\ \frac{24D}{\pi} (1 - \cos \frac{m\pi}{24}) \end{pmatrix}, \quad (5)$$

$$l = 0, \pm 1, \pm 2, \pm 3$$

$$m = 0, \pm 1, \pm 2, \pm 3, \pm 4, \pm 5, \pm 6$$

3.1.2 透視投影とカメラ配置

カメラの回転を R 、並進を t とするとある空間上の 3 次元点 X を透視投影した点 x は、式 (6) となる。

$$x = \text{diag}(f, f, 1)Z[R^T(X - t)] \quad (6)$$

ここで、 f はカメラの焦点距離、 $\text{diag}(a, b, c)$ は a, b, c を対角要素に持つ対角行列を表す。これに対して、カメラの並進 t を $Z < 0$ かつ原点からの距離が [1000 : 20000] の範囲の 1 様乱数で決定する。また回転行列 R は任意軸周りの $[-90^\circ : +90^\circ]$ の範囲の回転により生成する。そして、焦点距離 $f = 1000$ ピクセル、 2000×2000 ピクセルの画像面を仮定して、シミュレーションデータを作成する。実験には、画像内に少なくとも一つ以上の点が投影された画像のみを使用する。

3.2 精度の評価

実験では、初期復元として 3 画像から 3 次元復元を行い、その 3 次元復元結果を元に追加復元を行っていく。全ての復元済み画像と未復元画像間で式 (4) の $score_d$ を求め、その平均値を $Score_d$ とする。追加画像の選択方法は、 $Score_d$ が最も大きなものを選択する方法、 $Score_d$ が最も小さなものを選択する方法の精度比較を行った。計算したカメラの回転 R 、並進 t の精度を評価するために、真値との RMS 誤差を式 (7) で計算した。ここで、真の回転、並進をそれぞれ、 \hat{R}, \hat{t} とした。また、画像枚数は N 枚である。

$$E_R = \sqrt{\frac{1}{N} \sum_{\alpha=1}^N \|R_\alpha - \hat{R}_\alpha\|^2} \quad (7)$$

$$E_t = \sqrt{\frac{1}{N} \sum_{\alpha=1}^N \left(\cos^{-1} \frac{(t_\alpha, \hat{t}_\alpha)}{\|t_\alpha\| \|\hat{t}_\alpha\|} \right)^2}$$

3.3 視差の大きさを考慮した場合の精度比較

まず、シミュレーションデータに平均 0、分散 σ の正規分布に従う誤差を加えて実験を行った。画像の枚数を 30 枚として、 σ を 1 から 5 まで、1 刻みで変化させて、各 σ に対して試行を行った。 E_R および E_t の値を図 3 に示す。横軸は対応点に与えた誤差の分散値 σ を表す。図 3 より、 $Score_d$ が大きな画像を選択する方法は視差の小さな画像を選択するよりも誤差が増えても高精度かつ安定な結果が得られることがわかる。

次に、シミュレーションデータに平均 0、分散 1 の正規分布に従う誤差を加えて、画像の枚数を 30 枚から 150 枚まで、30 枚ずつ変化させて実験を行った。 E_R および E_t の値を図 4 に示す。横軸は画像の枚数を表す。図 4 より、 $Score_d$ が大きな画像を選択する方法は小さな画像を選択するよりも画像数が増えていっても、高精度かつ安定な結果が得られることがわかる。

3.4 対応点数と復元精度

シミュレーションデータをランダムに 4 つ (第 1, 2, 3, 4 画像) 作成し、そのシミュレーションデータに平均 0、分散 1 の正規分布に従う誤差を加えて、第 1, 2, 3 画像から初期復元を行う。次に、その初期復元の結果から DLT 法によって第 4 画像のカメラパラメータを推定する。そして、第 4 画像の対応点数を 10 点から 91 点まで変化させた時の精度の変化を調べた。各対応点数につき、100 回の試行の平均を誤差として、回転の誤差 E_R と、並進の誤差 E_t を求めた結果を図 5 に示す。ここで、横軸対応点数、縦軸はそれぞれ E_R, E_t を表す。この結果より、対応数が多い方が DLT 法によるカメラパラメータの推定精度が良いことが確認できた。

4. 実画像実験

提案手法で追加画像選択を行う場合と、Bundler のように既知の 3 次元点との対応数のみから画像選択を行った場合を比較するため、実画像実験を行った。実験に使用した 17 枚の画像を図 6 に示す。

4.1 実行結果

提案手法の画像選択を行って復元して行った結果を図 7(a) に示す。また、Bundler の選択手法により画像選択を行って復元して行った結果を図 7(b) に示す。そして、これらを初期値として、1 回のみバンドル調整を行った結果を

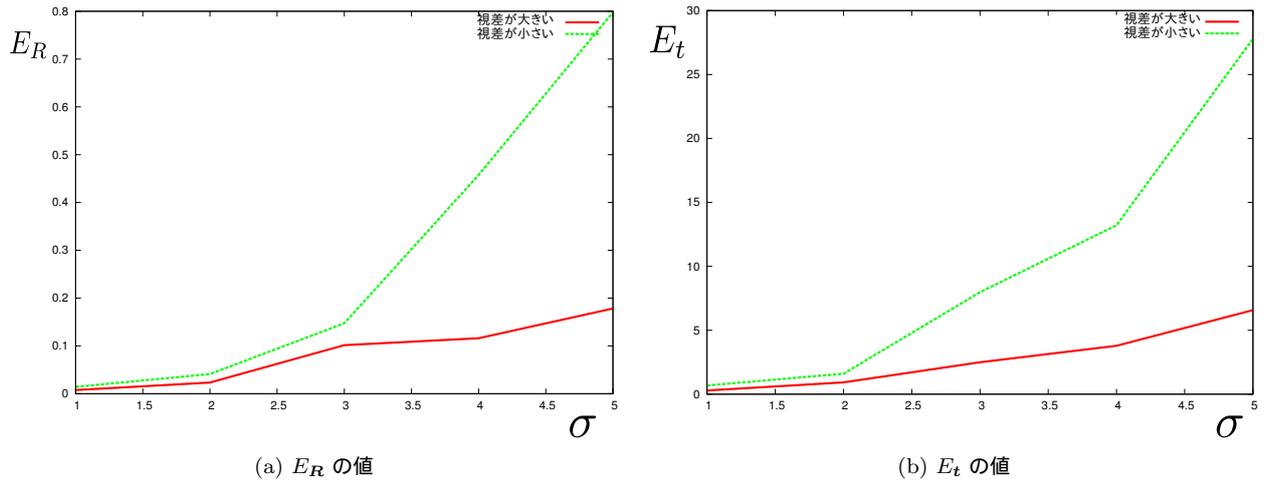


図 3 誤差の分散を変化させた時のカメラパラメータの推定精度

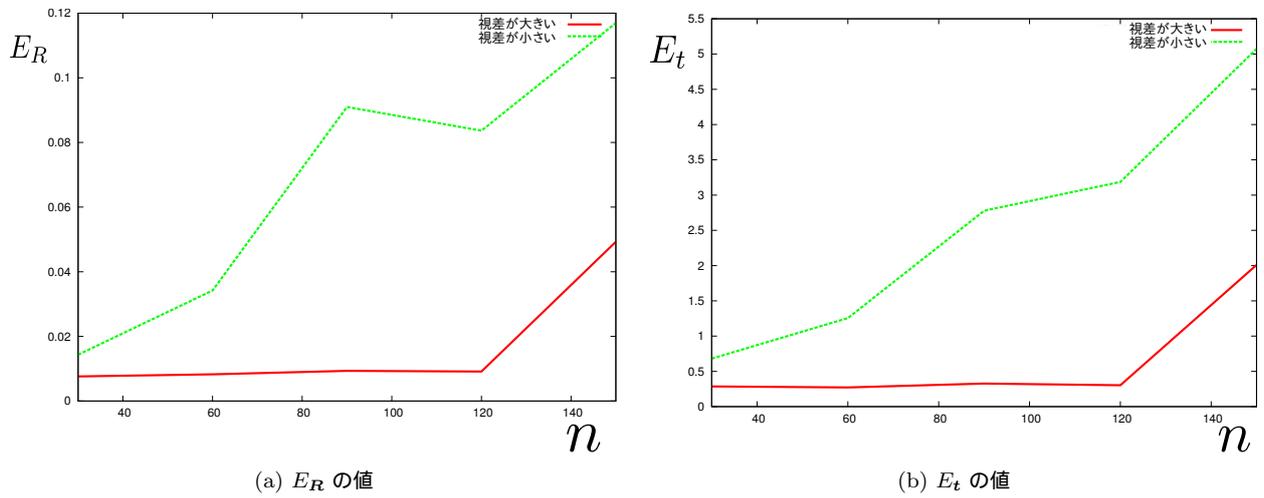


図 4 画像枚数を変化させた時のカメラパラメータの推定精度

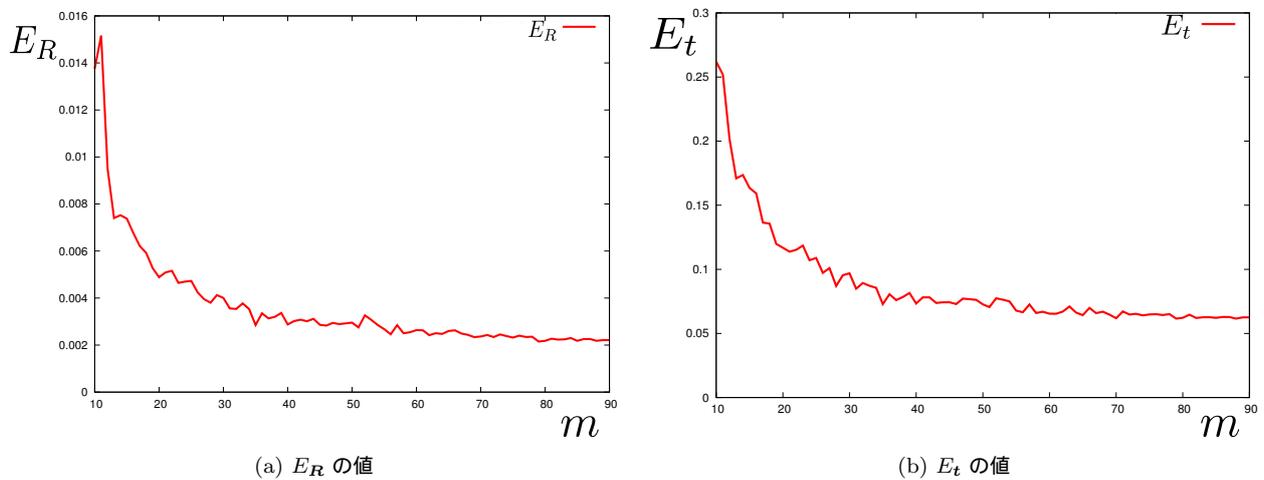


図 5 対応点数を変化させた時のカメラパラメータの推定精度

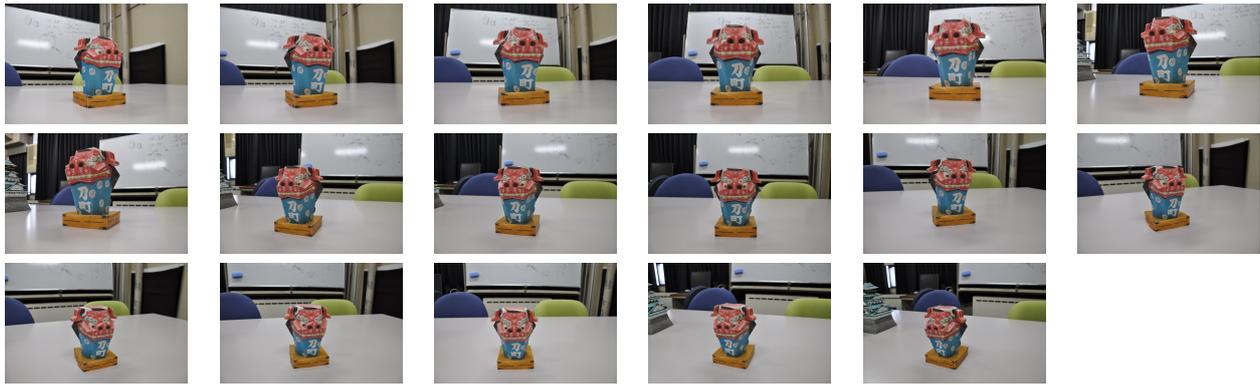
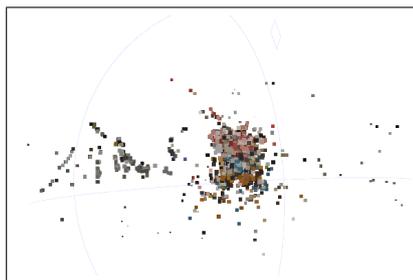
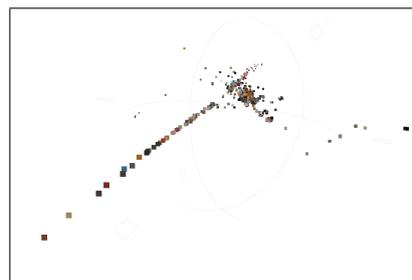


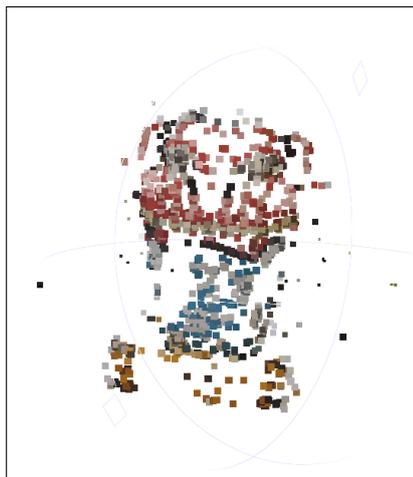
図 6 実験に使用した 17 枚の実画像



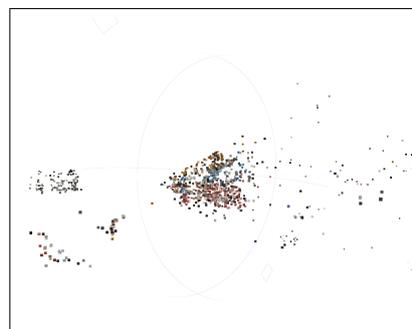
(a) 提案手法による復元結果



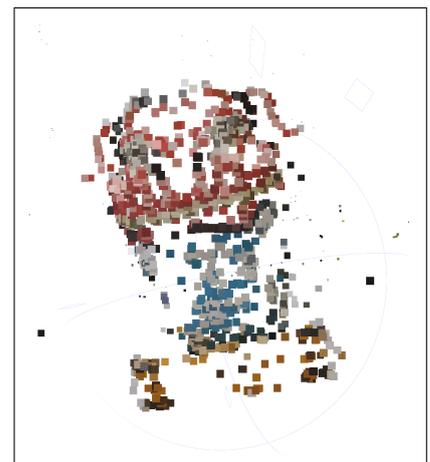
(b) Bundler の選択手法による復元結果



(c) (a) にバンドル調整をした結果
(バンドル調整 1 回)



(d) (b) にバンドル調整をした結果
(バンドル調整 1 回)



(e) Bundler による復元結果
(バンドル調整 34 回)

図 7 それぞれの手法による 3 次元復元結果

それぞれ図 7(c), (d) に示す。また, Bundler を用いて 3 次元復元を行った結果を図 7(e) に示す。Bundler の手法では 1 画像追加ごとに追加カメラのバンドル調整と復元済みカメラ全体のバンドル調整の 2 回, 全体で 34 回行っている。バンドル調整には Bundler で使われているものと同じ手法 [3] を用いた。

Bundler の選択手法で画像を選択して 3 次元復元した結果は復元点群がまとまらなかったが, 提案手法で画像を選択し 3 次元復元した結果は物体に沿うようにまとまった。また, Bundler の選択手法で 3 次元復元した結果を初期値としてバンドル調整を行うと, 初期値の精度が悪いため良好な結果を得ることができなかった。一方, 提案手法で 3

次元復元した結果を初期値としてバンドル調整を行うと, Bundler で 3 次元復元した結果とほぼ同等の結果を得た。

またこの結果を定量的に評価するために, 図 7(c), (d), (e) の 3 次元復元結果に対して復元点の再投影誤差を調べた。それぞれの結果に対する 1 点あたりの再投影誤差は, 0.432, 46.23, 0.298 となり, Bundler による 3 次元復元結果が一番再投影誤差が小さくなった。提案手法の結果にバンドル調整を施した場合, 1 回しかバンドル調整を行っていないにもかかわらず同様にバンドル調整を 1 回適用した比較手法に比べて再投影誤差が小さくなり, Bundler の結果に並ぶ高精度な結果を得ることができた。

4.2 考察

バンドル調整は、全てのカメラパラメータと3次元復元点を同時に最適に調節するため、非常に膨大な計算をしており、画像枚数が増えればかなりの時間がかかる。そのため大規模な復元になれば、Bundlerのように画像更新ごとにバンドル調整を行うと計算時間が膨大になる。これに対して提案手法のように追加復元する画像の選択方法を工夫することで、画像の追加ごとにバンドル調整を行わなくても最後に1度だけバンドル調整を適用するだけで精度のよい3次元復元が実現できることが確認できた。そのため、復元に使用する画像が非常に大量になる大規模なシーンの3次元復元を行う場合には、提案手法が有効であると考えられる。

5. まとめ

本研究では、多画像からシーンの3次元復元を行う際の復元精度を考慮した画像選択手法を提案した。具体的には復元に用いる画像間の視差を用いた画像選択の指標を提案し、従来手法よりも効率的に精度の高い3次元復元が実現できることを示した。まず、シミュレーション実験により提案手法の仮説が正しいことを確認し、実画像を用いて実シーンの3次元復元を行い、少ないバンドル調整の回数で高精度な結果を得ることができることを確認した。

謝辞: 本研究の一部は日本学術振興会科学研究費(若手研究 B 23700202)の助成による。

参考文献

- [1] M. A. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM*, 24-6, pp. 381-395, 1981.
- [2] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.
- [3] Lourakis M. and Argyros A., The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg Marquardt algorithm, *Technical Report 340*, Inst. of Computer Science-FORTHCVPR, 2004.
- [4] Nister D., An efficient solution to the five-point relative pose problem, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6) pp. 756-770, 2004.
- [5] Noah Snavely, Steven M. Seitz, and Richard Szeliski, Photo Tourism: Exploring image collections in 3D, *ACM Transactions on Graphics (SIGGRAPH Proceedings)*, 25(3) pp. 835-846, 2006.
- [6] Noah Snavely, Bundler: Structure from Motion (SfM) for Unordered Image Collections, <http://www.cs.cornell.edu/snavely/bundler/>.
- [7] Noah Snavely, Steven M. Seitz, and Richard Szeliski, Modeling the World from Internet Photo Collections, *International Journal of Computer Vision*, 80-2, pp. 189-210, 2007.
- [8] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. Fitzgibbon, Bundle adjustment-A modern synthesis, *Vision Algorithms: Theory and Practice*, pp. 298-375, 2000.