

画像列からの3次元形状復元における 非線形最小2乗法の高速度化と安定化に関する研究

右田 剛史††, 浅田 尚紀†

† 広島市立大学 情報科学部 知能情報システム工学科

‡ 現在 岡山大学 工学部 情報工学科

あらまし 運動からの形状復元問題は、非線形最小2乗法による再投影誤差最小化として定式化できる。一意の解が得られる線形の定式化も知られているが、再投影誤差最小化の方が適用範囲が広く精度も高い等の利点がある。しかし、非線形問題であるため計算量の削減(高速化)と局所解の回避(安定化)が重要な課題である。本研究では、これらの課題を扱う。また、屋外建物の全周形状復元問題を対象とし、画像数数百枚、特徴点数数百点の規模の実験を行い、提案手法の有効性を示す。

Fast and Stable Method for 3D Shape Recovery from an Image Sequence using Non-Linear Least Squares

T. Migita†† and N. Asada†

† Department of Intelligent Systems, Hiroshima City Univ.

‡ Presently with Faculty of Engineering, Okayama Univ.

Abstract Simultaneous recovery of shape and motion from an image sequence is directly formulated as minimization of reprojection errors via the nonlinear least squares. Although several linear formulations which ensure unique solution are also known, the nonlinear approach can be applied to much broader configurations, and gives more accurate results. We investigate fast optimization algorithm and several methods to avoid local minima for nonlinear reconstruction problem. We test our methods on recovering entire 3D shape from a building image sequence with several hundred images and several hundred feature points.

1 はじめに

運動からの形状復元問題 [1]~[8] は、物体を様々な方向から撮影した未校正画像列を用いて、対象物体の形状と各カメラの位置・姿勢を求めるものであり、CG や VR のモデル生成や3次元デジタルアーカイブなどへの応用が期待され、盛んに研究されている。

画像から3次元構造を推定する手法は、各画像中で特徴点を検出し画像間に対応付ける問題 [9, 10] と、これらの対応情報から特徴点とカメラの幾何学的な関係を推定する問題 [1]~[8] に分けて考えることができる。後者の問題に対する代表的な手法として、2枚の透視投影画像を用いる8点アルゴリズム [3] や、線形投影カメラで得られた画像列を用いる因子分解法 [4] が知られている。これらの問題設定の下では線形に定式化され厳密な議論を行う事ができるが、実環境の建物を全方位から撮影した画像列等に直接適用する事はできない。

一方、形状復元の最も直接的な定式化は、推定したモデルを画像に投影した結果と観測を比較したときの差異(再投影誤差の残差2乗和)を最小化する非線形最小2乗法によるもの [1] である。これは、最尤推定に基づく明解な定式化であり、前述の線形解法よりも広範な問題設定に適用でき、全ての入力情報を一度に用いた高精度な解を求められる等の利点

がある。しかし、非線形問題であるため、解法に関しては、全ての問題設定に対する一様な議論は困難である。また、大規模な入力を想定する場合、計算効率の議論も必要である。

本研究では、この非線形解法による形状復元に関連する以下の各課題について議論する。

● 反復解法の計算量低減(高速化)

一般に非線形問題の解は解析的に表現できないため、近似解の逐次更新による反復解法に頼ることになる。形状復元問題では Levenberg-Marquardt 法が多用されている [1, 5] が、これは画像数や特徴点数の3次の量に比例する計算量を要する。一方、共役勾配法を用いると特徴点の延べ出現数(特徴点数と画像数の積以下)に比例する計算量に抑えることができるため、大規模問題ではこちらの方が有利である。更に、これら2手法の長所を組み合わせることによって性能を向上させた「ブロック対角行列による前処理付共役勾配法(BDCG法)[11]」について述べる。未知数が1000個程度のデータを用いた実験では従来法の約1/5~1/20の計算時間で解を求めることができ、大規模な問題では更なる効果が期待できる。このような高速化は、この最適化を要素技術として用いる対話的な形状復元システムや、リアルタイムの画像処理システムなどへの応用において極めて重要な基礎技術である。

更に、この計算法は形状復元に依存する仮定を用いていないため、類似の数学的構造を持つ「未定要素のある特異値分解 [12]」や「Intensity Manifold [12]」等の問題にも適用可能であると期待できる。

- 適切な初期値の設定と局所解の回避 (安定化)

8点アルゴリズムや因子分解法などの基本手法が直接適用できない問題設定を扱う場合、問題を分割して部分毎に基本手法を適用し、それらの結果を統合した後、非線形最小2乗法によって再投影誤差を最小化する解に補正する (全体での最適化, bundle-adjustment) という方法を採用する機会が多い [8]。

非線形問題において、局所解を避け正解を得るために、近似解を初期値として用いる方法は有効であるが、初期値に要求される精度を明らかにすることは重要な課題である。これによって、初期値の準備に要する複雑な手順を簡略化することが期待できる。

本研究では、建物の全周復元を対象とした場合に、撮影方法の理想モデル (注視回転モデル) を用いた極めて単純な初期値から再投影誤差最小化の探索が正解に収束することを実験的に示す。この結果は、bundle-adjustment の初期値推定に用いられる極めて複雑な手順を大幅に簡略化できる可能性があることを示すものである。

- 逐次復元法

前述の理想モデルが有効ではない問題設定の場合には、部分データによる復元から始め、徐々に扱うデータを増やすことで、全情報に対する最適解を求める。このような考え方は文献 [4, 6, 7] 等にも見られるが、本研究では以下の点が異なる。データの増加に伴って復元に誤りが生じる可能性があることを考慮し、データを単調に増やすのではなく、試行錯誤探索によってデータを追加/削減する。また、逐次復元の各段階で再投影誤差を最小化する最適解を求めることにより、誤差の蓄積を避ける。

本研究では、この探索を対話的な形状復元システムに組み込んだ。全自動で正しい解を得られる画像列もあるが、人手の操作を伴う方が適用範囲は広い。

- 異種情報の併用

一般に画像列のみを用いた形状復元は困難であるため、地図、ジャイロ (角度センサ)、GPS (位置センサ)、レーザ測距儀などの画像以外のセンサを併用した復元の安定化も試みられている [15, 16, 17]。このような観点から、撮影角度を既知とした時の形状復元の線形化・安定化を議論する。

本論文の構成は以下の通りである。2章では形状復元問題の定式化と、高速計算アルゴリズムについて述べる。3章では全周囲形状復元の理想モデルとして注視回転モデルを導入し、これに基づく簡易な初期値設定による全周復元の収束可能性について実

験的に検討する [13]。4章では、より難しい形状復元問題を解くための逐次復元の探索アルゴリズム [14] を示す。5章では画像情報のみでは解くことが難しい問題を安定に解くため、カメラの角度を併用した線形化法 [18] を示し、角度測定の誤差と、復元形状の品質の関係を調べる。6章では結論を述べる。

2 形状復元の定式化と高速解法

2.1 運動からの形状復元の定式化

対象物体の形状を P 個の特徴点で表す。各特徴点は物体座標系に対して静止しており、その座標を $s_p = (x_p, y_p, z_p)^T$, ($1 \leq p \leq P$) と書く。物体を撮影した F 枚の画像列を考え、第 f 画像 ($1 \leq f \leq F$) での物体座標系とカメラ座標系の関係を、回転行列 R_f と並進ベクトル $t_f = (t_{fx}, t_{fy}, t_{fz})^T$ で表す。これらのパラメータを用いると、第 p 特徴点の第 f 画像における座標 $u_{fp} = (u_{fp}, v_{fp})^T$ は次式によって表される。

$$u_{fp} = P[R_f s_p + t_f] \quad (1)$$

ここで、 P は次式で定義される透視投影の作用素である。

$$P \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{l}{z} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2)$$

運動からの形状復元は、 u_{fp} の集合が与えられた時に、式 (1) の関係に基づいて、右辺の形状パラメータ s_p と運動パラメータ R_f , t_f を求める問題である。観測に誤差が含まれることを考慮すると、この式は厳密には満たされないため、次のように残差2乗和を最小化する問題として定式化される。

$$\arg \min_{s_p, R_f, t_f} \sum_{(f,p)} |P[R_f s_p + t_f] - u_{fp}|^2 \quad (3)$$

以下、本節では、最適化の準備として、運動パラメータの表現法を扱う。

非線形最適化計算において回転行列を表現する場合、4元数を用いると都合がよい [1]。4元数 $q = (q_0, q_1, q_2, q_3)$ は、図1に示す 3×3 行列に対応付けられる。この行列は回転行列の定数倍 (スケール回転行列) であり、ノルム $(q_0^2 + q_1^2 + q_2^2 + q_3^2)$ が1の4元数は回転行列を表す。このように、三角関数を必要とせず2次式で回転行列を表すため、微分が容易、計算が高速である等の利点がある上、4つのパラメータの対称性が高く特異点や不均一性がない [8] 等の利点もある。ただし、本来自由度が3しかない回転

$$Q(\mathbf{q}) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix}$$

図 1: 4 元数とスケール回転行列

行列を 4 つのパラメータで表す冗長な表現となっているため、この冗長性を除く工夫が必要である。

形状復元では回転並進に続いて透視投影が作用するため、式 (1) の回転行列をスケール回転行列に置き換えた次式を用いることができる。

$$\mathbf{u}_{fp} = \mathcal{P} [Q(\mathbf{q}_f) \mathbf{s}_p + \mathbf{t}_f] \quad (4)$$

透視投影 (式 (2)) では、非 0 の実数 k に対して $\mathcal{P}[\mathbf{s}] = \mathcal{P}[k\mathbf{s}]$ が成立するため、上の式は次式のように書き換えることができる。

$$\mathbf{u}_{fp} = \mathcal{P} \left[\frac{Q(\mathbf{q}_f)}{|\mathbf{q}_f|^2} \mathbf{s}_p + \frac{\mathbf{t}_f}{|\mathbf{q}_f|^2} \right] \quad (5)$$

つまり、 $Q(\mathbf{q}_f)/|\mathbf{q}_f|^2$ と $\mathbf{t}_f/|\mathbf{q}_f|^2$ が本来の回転行列と並進ベクトルである。

本研究では、最適化の過程では式 (4) の投影モデルを用いる。この式では、4 元数の正規化は透視投影の除算に吸収されているため、微分が容易になり、計算量も低減され、負の数の平方根や 0 による除算等の問題を避けられるため、都合が良い。一方、本来 6 自由度の位置・姿勢を 7 つのパラメータ (\mathbf{q}, \mathbf{t}) で表しているため、非 0 の実数 k に対して ($k\mathbf{q}, k^2\mathbf{t}$) は全て同一の位置・姿勢を表していることになるが、非線形最適化計算ではこの曲線上のどの点に収束してもよいため計算の不安定性を生じるものではない¹。経験的にも、この冗長性が収束に悪影響を及ぼさないことを確認している。

次節の議論の都合上、式 (3) の代わりに、目的関数を次の形で表すことにする。

$$E(\mathbf{w}) = \frac{1}{2} \mathbf{r}(\mathbf{w})^T \mathbf{r}(\mathbf{w}) \quad (6)$$

ここで、 \mathbf{w} は形状と運動のパラメータを全て含む $N = 3P + 7F$ 次元ベクトルである。

$$\mathbf{w} = (x_1, y_1, z_1, \dots, x_P, y_P, z_P, \\ t_{1x}, q_{10}, q_{11}, q_{12}, q_{13}, t_{1z}, t_{1y}, \dots, \\ t_{Fx}, q_{F0}, q_{F1}, q_{F2}, q_{F3}, t_{Fz}, t_{Fy})^T$$

¹文献 [8] では、原点から非常に遠い特徴点は 3 パラメータ (x, y, z) で表現するよりも、射影的に 4 パラメータ (x, y, z, w) \rightarrow ($x/w, y/w, z/w$) で表現する方がよいとしている。これは、3 パラメータ表現では極めて大きな数に収束するには多数の反復を要するが、4 パラメータ表現では w を 0 に近づけるだけで無限遠近くの点を表現でき、この方が最適化の反復回数ははるかに少なく済むためである。本論文の場合も、これと同様である。

また、 \mathbf{r} は推定座標と観測座標の残差 $\mathcal{P}[Q(\mathbf{q}_f) \mathbf{s}_p + \mathbf{t}_f] - \mathbf{u}_{fp}$ を全て並べた $2M$ 次元ベクトルである。

M は特徴点の述べ観測数であり、全画像で全特徴点が観測できる場合には FP となる。比 $M/(FP)$ を観測率と呼び、次章では形状復元問題の難易度の指標として用いる²。また、達成した残差の最小値を E^* として、 $\sqrt{2E^*/M}$ は RMS 残差であり、復元精度の 1 つの指標となる。

2.2 非線形最適化の高速解法

式 (6) は非線形最適化問題であり、解を解析的に表現することはできないため、勾配に基づいて推定値を逐次更新する反復法で解く。解の更新は次式で表される E の 2 次近似に基づいてなされるが、この更新方法が非線形最適化の効率を大きく左右する。

$$E(\mathbf{w}_k - \mathbf{d}) \approx E(\mathbf{w}_k) - \mathbf{d}^T \mathbf{g}_k + \frac{1}{2} \mathbf{d}^T \mathbf{H}_k \mathbf{d} \quad (7)$$

ここで、 \mathbf{g}_k は点 \mathbf{w}_k における E の勾配 ∇E 、 \mathbf{H}_k は点 \mathbf{w}_k におけるヘッセ行列 ($\partial^2 E / \partial w_i \partial w_j$) であり、次式のヤコビ行列 \mathbf{J}_k を用いて $\mathbf{H}_k \approx \mathbf{J}_k^T \mathbf{J}_k$ と近似できる。

$$\mathbf{J}_k = \begin{bmatrix} \frac{\partial r_1}{\partial w_1} & \frac{\partial r_1}{\partial w_2} & \dots & \frac{\partial r_1}{\partial w_N} \\ \frac{\partial r_2}{\partial w_1} & \frac{\partial r_2}{\partial w_2} & \dots & \frac{\partial r_2}{\partial w_N} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial r_{2M}}{\partial w_1} & \frac{\partial r_{2M}}{\partial w_2} & \dots & \frac{\partial r_{2M}}{\partial w_N} \end{bmatrix} \quad (8)$$

2.2.1 Levenberg-Marquardt 法 (LM 法)

式 (7) の右辺を最小化する \mathbf{d} は $\mathbf{H}_k^{-1} \mathbf{g}_k$ である。しかし、右辺は左辺の近似であるため、この値は必ずしも左辺を最小化するわけではない。このため、 $\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{H}_k^{-1} \mathbf{g}_k$ と更新する Newton 法は、発散する危険がある。これを避けるために、 $\mathbf{w}_{k+1} = \mathbf{w}_k - (\mathbf{H}_k + \mu_k \mathbf{I})^{-1} \mathbf{g}_k$ と更新する Levenberg-Marquardt 法 (以下、LM 法) が、形状復元問題においては多用されている [1, 5]。なお、 μ_k は解の遠くでは大きく、解の近くでは小さく選ばれる。この行列 $\hat{\mathbf{H}}_k :=$

²文献 [4] で用いられている「観測行列の fill-ratio」と等価なものである。

$J_k^T J_k + \mu_k I$ は次のような構造の N 次対称行列となる。

$$\dot{H}_k = \begin{array}{|c|c|} \hline \begin{array}{c} S \\ \vdots \\ V \end{array} & U \\ \hline V & T \\ \hline \end{array} \quad (9)$$

空白部分の要素は0である。 S と T の部分はブロック対角行列であり、ブロックの大きさはそれぞれ 3×3 と 7×7 である。この行列の特殊構造を利用して、 $O(FP \cdot \min(3P, 7F))$ の計算量で Cholesky 分解を行う方法が知られている [1, 5, 8] が、この計算量では大規模問題には適していない。文献 [5] では、時系列画像を仮定した近似によって、特徴点数に比例する計算量に抑えているが、この近似が一般に有効であるとは限らない。

2.2.2 共役勾配法 (CG 法)

目的関数が2次関数の場合、かつ H_k が大規模疎行列の場合には、(前処理付)共役勾配法(以下、CG法)が有効な手段であることが知られている [19]。以下に解を更新法を示す。

$$\begin{aligned} \mathbf{w}_{k+1} &= \mathbf{w}_k - \alpha_k \mathbf{d}_k \\ \mathbf{d}_k &= C^{-1} \mathbf{g}_k + \beta_k \mathbf{d}_{k-1} \end{aligned} \quad (10)$$

ここで、

$$\begin{aligned} \alpha_k &= \arg \min_{\alpha} E(\mathbf{w}_k - \alpha \mathbf{d}_k) \\ \beta_k &= \mathbf{g}_k^T C^{-1} \mathbf{g}_k / |\mathbf{g}_{k-1}|^2 \end{aligned}$$

ただし、最初の探索方向は $\mathbf{d}_0 = C^{-1} \mathbf{g}_0$ である。また、 C は前処理行列である(前処理を行わない場合は $C = I$)。この手法は、元来、 N 次元の連立方程式(凸2次関数の最小化)の解法であり、計算量は、係数行列が密の場合は $O(N^3)$ 、疎の場合は $O(N^2)$ 、前処理行列をうまく選ぶと $O(N)$ に近づく。また、非線形最適化にも応用されている³。この手法を用いると、形状復元は特徴点の延べ出現数 M に比例する計算量に抑えることができるため、大規模問題ではLM法よりも有利である⁴。更に、以下ではヘッセ行列をブロック対角部分で近似したものを前処理行列として用いて性能を向上させた「ブロック対角行列による前処理付共役勾配法(Block Diagonal matrix preconditioned CG method, 以下BDCG法)[11]」について述べる。

³ただし、非線形最適化に対して前処理を用いた例は確認できない。

⁴文献 [2] ではLM法はCG法よりも格段に効率が良くいと述べているが、これは問題規模が小さく連立方程式求解の手間が大きな問題にならないためと考えられる。

2.2.3 BDCG 法

線形連立方程式に対する前処理付共役勾配法に倣って、非線形最適化の共役勾配法にも前処理行列を導入することを考える。前処理行列としては、ヘッセ行列の近似であり、連立方程式の求解が容易であるものが望ましい。このような行列を構成するために、式(9)の \dot{H} の U と V の部分を0で置き換える方法が有効であると考えられる(以後、この行列を \ddot{H} と書く)。行列 \ddot{H} では、 U や V の各要素は2次元ベクトルの内積であるのに対して、対角部分 S 、 T はそれぞれ $6P$ 次元、 $14F$ 次元ベクトルの内積で表されるため、 U, V 部分に比べて S, T のブロック対角部分の要素の絶対値が非常に大きくなる [5]。また、 F と P がともに大きいほど対角成分の値も大きくなり、 U, V を無視することによる相対誤差は小さくなる。ブロック対角部分のみで近似した \ddot{H} は、 $O(N)$ の計算量で Cholesky 分解が可能であり、必要な記憶容量も $N^2 \rightarrow 7N$ と大幅に削減することができる。 U, V を省略することによりLM法の $\dot{H}^{-1} \mathbf{g}$ と比べて探索方向に誤差が含まれることになるが、この影響を共役方向生成 ($\beta_k \mathbf{d}_{k-1}$ の項)によって補うのが本手法の特徴である。

また、式(10)中の α_k の決定は、 α の1次元関数 $E(\alpha)$ を最適化する直線探索であるが、厳密に $E(\alpha)$ を最小化する必要はない。CG法では2次関数に対する最適性を満たせば十分であるため、本研究では2次近似による方法で直線探索を行うこととした。詳細は付録に示すが、この方法は黄金分割法 [19] より少ない回数で探索を終了する。

なお、 \ddot{H}^{-1} は毎回更新するわけではない。探索が進行し、探索点が \dot{H} を求めた点から離れるに従って探索方向 \mathbf{d}_k が最適探索方向からずれ、解の収束性が低下する。そこで、通常のCG法における定期的な再出発 [19] を拡張し、再出発と同時に \ddot{H} を更新することとした。解の収束性を高めるためには、反復計算における探索方向の精度をある程度維持する必要があるため、通常のCG法の再出発に比べて周期を短く設定し、後述の実験では16反復毎に \ddot{H} の再計算を行った。この再出発による解収束の効果と計算コストをLM法と比較すると次のように考えることができる。 \ddot{H} の再計算の計算量が $O(M)$ であるのに対してLM法の1反復の計算量が $O(FP \cdot \min(3P, 7F))$ であることを考えると、 \ddot{H} の計算から次の再計算までの間の解探索によってLM法の1反復と同等の解収束の効果が得られれば、計算コストの点で十分有利である。

次に、BDCG法はCG法と同様にLM法に比べて多くの反復回数を要する点について考察する。線

形連立方程式に対する前処理付き共役勾配法の反復回数は、未知数の数よりも係数行列の固有値の密集具合によることが知られている [19]。同様に、非線形最適化問題においても反復回数を決定する要素はヘッセ行列の固有値の分布となる。文献 [20] の解析を考慮すると、反復回数は特徴点の配置と運動の性質によって決まるものと考えられ、実験的にも反復回数が特徴点や画像数にほとんど影響されないことを確認している。したがって、反復回数は少ないが計算量が $O(FP \cdot \min(3P, 7F))$ の LM 法より、反復回数が問題の規模にはほぼ無関係で計算量が $O(M)$ の BDCG 法の方が、画像数 F や特徴点数 P が大きな問題になるほど計算速度の点で有利になる。

なお、BDCG 法は解の探索方向を求める際にヘッセ行列の近似を行っているが、式 (6) そのものの最小化については何ら近似を行っていないため、理論的にも実験的にも LM 法や CG 法と同じ精度の解が得られることを確認している。

2.3 実験

提案手法の有効性を確認するためシミュレーションデータおよび実画像を用いた実験を行った。次章で述べる「注視回転による初期値設定法」を用いて大局的な最適解が得られる問題に適用し、LM 法、CG 法、BDCG 法の計算速度を比較した。実験に用いた計算機は、Pentium 4, 2.0GHz, OS は Vine Linux 2.5 である。

なお、性能比較のために実装した LM 法は行列の疎構造を利用した Cholesky 分解を用いたものであり、従来の高速化 LM 法 [1, 5] と同等の計算量である。

2.3.1 シミュレーション実験

対象物体として、球面上に $P = 60, 120, 240$ 個の特徴点をほぼ均等に配置した 3 種類の物体を用いた。図 2 は 240 点の球の例であり三角形パッチの頂点が特徴点を表す。カメラの前で並進しながら 360 度回転する物体を撮影した 3 画像列 (画像数 $F = 25, 50, 100$ 枚) を作成し、全部で 9 例のデータセットを用いて形状復元を行った。未知数の個数 ($N = 3P + 7F$) は表 1 に示すように最小 355, 最大 1420 である。このデータでは、半径 100 の球が距離 1000 程度離れて投影されているので射影歪みは大きくないが、どの画像も全特徴点の半数が球の裏側で見えないという隠蔽を生じている。なお、画像上での物体の像の大きさは 150 画素程度であり、観測された特徴点位置には標準偏差 $\sqrt{2}$ 画素の正規乱数の誤差を与えた。

注視回転初期値を用いて形状復元した結果は、い

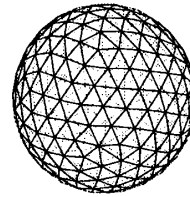


図 2: シミュレーション実験に用いた球

表 1: 未知数の個数

Frames \ Points	60	120	240
25	355	535	895
50	530	710	1070
100	880	1060	1420

ずれの手法も RMS 残差が小数点以下 3 桁まで一致し、形状の 3 次元誤差は全ての例で球の半径の $1/110$ 以下であった。このことから、3 手法とも同じ精度の解が得られることが分かる。

形状復元に要した 3 手法の計算時間を表 2 に示す。いずれのデータセットにおいても LM, CG, BDCG の順に計算時間が減少し、3 手法の中では BDCG 法が最も高速である。図 3 は F, P の増加に対する計算時間の変化を表すグラフである。 F, P がそれぞれ 2 倍, 4 倍に増えた時、BDCG 法、CG 法の計算時間がほぼ線形に増加するのに対して、LM 法は急速に増大する傾向がある。これは、CG 法と BDCG 法の計算量が M に比例するのに対して、LM 法では $FP \cdot \min(3P, 7F)$ に比例するためである。なお、CG 法と BDCG 法の計算時間にはほぼ定数倍の差があり、この実験では 3 倍程度であった。

一方、3 手法の反復回数を表 3 に示す。LM 法はいずれのデータセットに対しても反復回数は 5 回と最も少ない。これに対して、CG 法は 120 回程度、BDCG 法は 30 回程度の反復計算を行っている。しかし、上述のように反復計算 1 回あたりの計算量は LM 法が多く、CG 法と BDCG 法は少ないため、全体の計算時間では CG 法や BDCG 法の方が有利である。しかも、LM 法と BDCG 法は F, P の増加に対して反復回数がほとんど変化しないため、問題の規模が大きくなる程 BDCG 法が有利となる。

2.3.2 実画像を用いた実験

実画像からの形状復元に 3 手法を適用した結果を示す。対象物体として車の模型 ($65 \times 70 \times 155$ mm) を用い、カメラ-物体間の距離と焦点距離を変化させ

表 2: 計算時間の比較 (単位は 10^{-3} 秒)

Frames\Points	60	120	240	
25	LM	67.7	122.5	278.4
	CG	27.3	49.3	146.6
	BDCG	9.0	17.2	34.8
50	LM	132.1	491.2	877.1
	CG	55.3	101.6	242.3
	BDCG	18.6	35.1	69.1
100	LM	270.9	860.8	7533.9
	CG	138.2	208.8	499.4
	BDCG	38.8	72.0	141.5

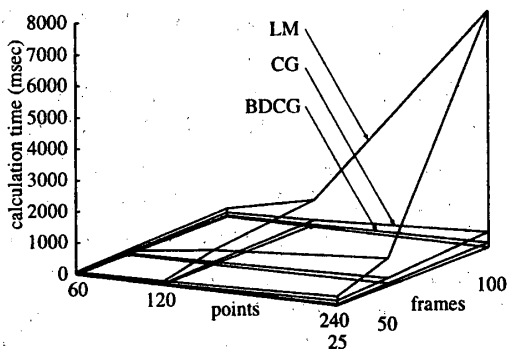


図 3: LM 法, CG 法, BDCG 法の計算時間の比較

表 3: 反復回数の比較

Frames\Points	60	120	240	
25	LM	5	5	5
	CG	105	88	124
	BDCG	28	28	29
50	LM	5	5	5
	CG	108	101	113
	BDCG	29	29	29
100	LM	5	5	5
	CG	139	107	127
	BDCG	32	31	31

て射影歪みの異なる 3 種類の画像列を撮影した。表 4 に各画像列の性質を表す各種パラメータと第 1 フレームの画像を示す。画像サイズは 640×480 であり、各画像列から Tomasi-Kanade の手法 [9] を用いて物体上の特徴点を検出・追跡した。

注視回転初期値を用いて形状復元したときの 3 手法の計算時間、反復回数、RMS 残差の比較を表 5 に示す。3 手法とも残差が一致していることから、復元形状に精度の差がないことが分かる。3 手法の計算時間と反復回数の違いは、シミュレーション実験と同じ傾向を示し、計算時間は LM 法 > CG 法 > BDCG 法の順に短く、反復回数は CG 法 > BDCG 法 > LM 法の順に少ない。いずれの画像列も BDCG 法の計算が最も速く、最大で CG 法の 3.9 倍、LM 法の 8.8 倍の高速化が実現できている。

なお、この実験では特徴点の隠蔽が生じない画像列を用いたため、因子分解法を適用することができ。そこで、因子分解法で求めた形状 (近似解) を初期値として非線形最適化の 3 手法を適用したところ、BDCG 法の計算時間が最も短く、最大で CG 法の 9.5 倍、LM 法の 18.4 倍の高速化が実現された。

3 注視回転初期値による全周復元

3.1 注視回転による初期値

式 (3) の最小化は非線形問題であるため、局所解を避け正しい解を得るためには適切な初期値を設定する必要がある。

ここでは、実環境の建物を周囲 360 度から撮影した画像列を想定し、全周囲形状を一括復元することを考える。なお、この問題設定は、復元すべき情報に対して利用できる情報が少ないため、局所解が現れやすい困難な問題設定である。即ち、実環境の建物の周辺には木立や他の建造物が存在するため、遠方から撮影できず、建物に近接した位置から撮影した画像列を使うことが多い。近距離からの撮影では、各画像に現れる物体の部分が減少し、一周撮影に必要な画像数が増加する。このように、部分的な情報を多数つなげて全体を復元する必要がある。反対に、物体の片面のみを復元する場合等、全特徴点が全画像に現れる場合には局所解は現れ難い。

全周撮影の直観的な撮影方法は図 4(a) に示す円周撮影の形となる。これは、各カメラ位置で 1 枚の画像を撮影しながら円状の軌跡を描いて移動するものである。この時、カメラ軌跡はほぼ平面上に収まる。一方、(b) の疑似広角撮影は、円周撮影と同様の移動をしながら、各カメラ位置でカメラ方向を変化させて複数の撮影を行うものであり、同一の特徴点をより広い角度から撮影することができるため、

表 4: 画像列 1, 2, 3 の性質

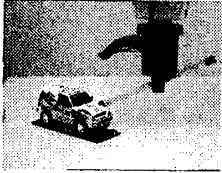
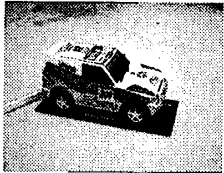

	画像列 1	画像列 2	画像列 3
特徴点数 × 画像数	90×81	187×19	269×16
未知数の個数	837	694	919
対象物までの距離 (mm)	1100	300	150
カメラの焦点距離 (pixel)	1800	870	870
物体の回転量 (度)	33	24	20
第 1 フレームの画像			

表 5: 注視回転初期値による 3 手法の性能比較 (計算時間の単位は 10^{-3} 秒)

	画像列 1			画像列 2			画像列 3		
	計算時間	反復回数	残差	計算時間	反復回数	残差	計算時間	反復回数	残差
LM	915	9	1.132	209	7	1.190	394	11	1.336
CG	378	157	1.132	193	159	1.190	318	226	1.336
BDCG	104	45	1.132	53	45	1.190	81	58	1.336

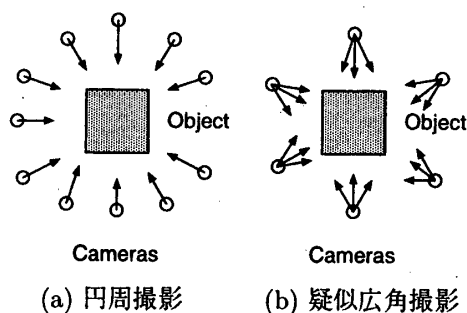


図 4: 全周撮影の方法

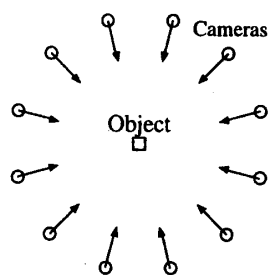


図 5: 注視回転

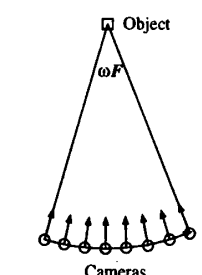


図 6: 45 度の注視回転

円周撮影よりも有効となる場合がある。

これらの撮影方法の理想モデルとして、図 5 の注視回転モデルを導入する。これは、物体を中心にカ

メラが等速円運動をしながら軌道中心方向を観測するものである。これに基づいて初期値を構成し、全周形状の一括復元を行う。

復元の推定パラメータは、カメラの位置、姿勢、物体形状の 3 種類から成る。このうち、形状情報は一般に未知であり、初期値として何らかの仮定を導入することは現実的でない。そこで、形状の初期値としては全特徴点が原点に位置するという縮退した状態を用いる。カメラの位置、姿勢情報については、注視回転を仮定して設定する。この初期値 (以下、注視回転による初期値と呼ぶ) を数式で表すと次式のようにになる。

$$\begin{cases} R_f = R_y^{\omega f} & t_f = (0, 0, 1)^T \text{ for all } f \\ s_p = (0, 0, 0) & \text{ for all } p \end{cases} \quad (11)$$

ここで、 R_y^ω は y 軸 (地面に対して垂直な軸) 周りの回転行列であり、回転量は ω である。撮影状況に応じて与えるべきパラメータは、全回転量 ωF (または、画像間の平均回転移動速度 ω) である。全周復元では、撮影時のカメラ軌跡の方向 (ωF の符号) を考慮して、 $\omega F = 360$ 度または $\omega F = -360$ 度のいずれかを選ぶ。

このような初期値設定⁵により、どの範囲の問題

⁵文献 [1] でも、観測データから簡単な方法で初期値を構成し、形状復元が行えることを示しているが、本手法は観測データさえ用いない更に簡単なものとなっている。

を解くことができるかが重要であるが、非線形性のため、解析的な検討は困難である。経験的に得られた傾向を以下に列挙する。本手法の有効性については、実験で評価する。

- 対象物体の形状にはほとんど依存せず、1点に縮退した初期値から正解を得ることができる。
- 全回転量の大きさにもそれほど影響を受けない。実際の回転量が270度であっても360度を初期値として正解がえられる。
- カメラ軌跡の方向は重要で、撮影時と異なる設定をすると、局所解に陥る可能性が高い。
- 疑似広角撮影では、各撮影点で仰角や方位角が大幅に変動するが、これに対しても、式(11)を初期値として正解が得られる。また、疑似広角撮影の方が円周撮影よりも安定して解が得られる。
- 対象物体に近接して撮影するほど局所解に陥りやすく、復元が失敗する可能性が高くなる。

なお、全周より狭い範囲の部分形状復元に対して、撮影状況にあわせて ωF を定めることにより、同じ枠組で復元が行える。この際、全周復元の場合と同様に、カメラ軌跡の方向は重要であるが、全回転量の大きさにはそれほど影響を受けない傾向がある。

3.2 解の修正

注視回転による初期値は多くの場合に正解を得ることができるが、問題設定によっては局所解に陥ることもある。このような場合にも、式(3)の枠組の中で、特徴点の削除と再追加を繰り返すことにより、解を修正して、局所解から抜け出すことができる場合がある。そのために以下のような手順をとる。

- 局所解の原因となった特徴点を除去する。
- 残った特徴点のみを用いて、形状と運動を再推定する。
- 除去した特徴点を再追加する。

この手順のために、式(3)の目的関数において総和の範囲を限定した次式を導入する。

$$E(\mathbf{x}, L) = \sum_{(f,p) \in S(L)} |P[R_f \mathbf{s}_p + \mathbf{t}_f] - \mathbf{u}_{fp}|^2 \quad (12)$$

ここで、 $S(L)$ は考慮すべき添字の集合で、 L は制御パラメータである。特徴点 p が L 枚以上の画像に現れる場合のみ $S(L)$ に (f,p) を含めるものとする。

理論上は、特徴点は2画像以上で観測されていれば($L=2$)復元可能であるが、少数の画像でしか現れない特徴点が誤った位置に推定されることによって、カメラ位置や姿勢が誤って推定され、局所解となる場合が多い。手順(i)では、 L を増加させるこ

とによって、局所解の原因となっている特徴点(群)を一旦除去する。手順(ii)では、残った点を用いて、(12)を最小化することにより、カメラ位置、姿勢を推定しなおす。多数の画像で観測されている特徴点のみを用いると、カメラ位置、姿勢や形状を正しく推定できる傾向がある。手順(iii)では、手順(ii)で修正された推定値に基づいて、除去した特徴点群を含めた全ての特徴点を用いて式(12)を最小化する。これによって、局所解から脱出する。一回で解を修正しきれない場合、 L の増加、減少を複数回繰り返すことで修正を行う。

一般に、局所解では、 $E(\mathbf{x}, L)$ の値が大きいのに対して、局所解の原因となっている特徴点が除去できると、 $E(\mathbf{x}, L)$ の値が急激に減少する。この性質を使って、 $E(\mathbf{x}, L)$ の閾値処理によって、パラメータ L を決定する。

この手順によって必ずしも局所解から抜け出せるわけではないが、パラメータ L を増加させても十分な数の特徴点が残る画像列に対しては、ある程度の効果が期待できる。

3.3 実験

本節では、全周復元問題に対して、提案した初期値設定がどの程度有効であるかを確かめるため、数値的に生成した疑似建造物を用いた模擬実験、並びに、実画像での実験結果を示す。

3.3.1 シミュレーション画像による復元

凹凸のある建物を想定し、球を変形させた次式の曲面を対象物体とする。ただし、 $x-z$ 平面を地面と考える。

$$\begin{bmatrix} [1 - \nu + \mu \cos(\rho(\varphi - \pi/4))] \cdot \cos \theta \cos \varphi \\ [1 - \tau] \cdot \sin \theta \\ [1 - \nu + \mu \cos(\rho(\varphi - \pi/4))] \cdot \cos \theta \sin \varphi \end{bmatrix} \\ |\theta| \leq \pi/2, \quad |\varphi| \leq \pi$$

この曲面は球の θ - φ 表現を変形したもので、 ν, τ は縦・横・奥行の比を決めるパラメータ、 ρ は凹凸の数、 μ は凹凸の深さを表す。 $\nu = \tau = \mu = 0$ のときこの曲面は球となる。球面上で点密度がほぼ一定となるように特徴点を設定し、上の式に従って変形したものを対象物体とする。更に、近接点を結んだ3角形群で曲面を構成し、特徴点の隠蔽判定に用いる。

真円からある程度離れたカメラ軌跡を生成するために、視点の決定にも上の曲面の式を用いる。対象物体とは異なるパラメータを用いて、対象物体よりも大きな曲面を作り、 $\theta = 0$ で切った断面の曲線を上下/左右に平行移動させたものに沿って視点を設

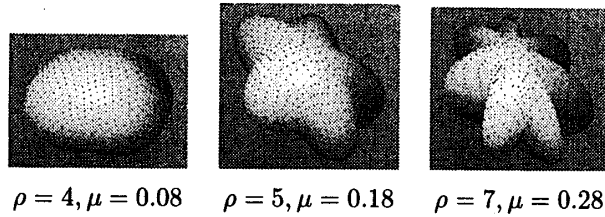
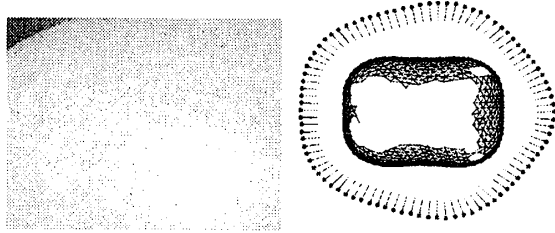
図 7: 対象物体 ($\nu = \tau = 0.2$)

図 8: 4角形を撮影した例 (最も近い距離)

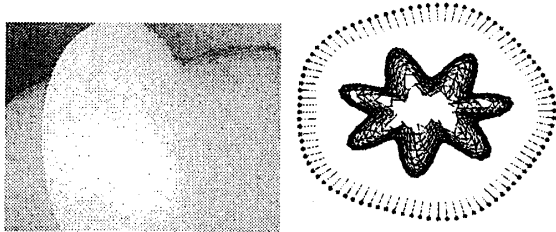


図 9: 7角形を撮影した例 (最も遠い距離)

定する。光軸方向は、この曲線の接線によって決めるものとする。このため、光軸方向は必ずしも物体中心を向いていない。

形状、画像数等を以下のように変化させ、合計 135 種類のデータについて復元の成功率を調べた。

形状: 図 7 に示す 4 角形、5 角形、7 角形を用いた。角数の増大と共に物体の凹凸も増加させたため、4 角形の場合ほとんど凹凸はないが、7 角形の場合、物体は深く凹んでおり隠蔽が起こりやすい。隠蔽が多いと特徴点対応は少なくなり、復元困難となる。

特徴点数: 各形状につき 300, 600, 1200 と変化させた。特徴点数が少ないほど隣接画像での対応数が減り、復元困難となる。

画像数: 物体周囲を一回転する 5 角形 ($\rho = 5, \mu = 0.03, \nu = 0.1$) の軌跡に沿って 30, 60, 90ヶ所から、22.5 度見上げる角度で画像を作成した。各画像の大きさは 600x450pixel で、焦点距離は画像の横幅と同じ (横画角が約 53°) とした。画像数が少ないほど隣接画像間の対応が少なく復元困難となる。

カメラ-物体間距離: 距離が遠い場合復元は容易で、カメラが物体に近づくと隣接画像間での対応が

減り復元困難となる。ここでは、復元成功率が大きく変化するような 5 段階の距離を用いた。図 8, 9 に画像および物体とカメラ軌跡の上面図の例を示す。5 段階のうち最も遠い距離の場合でも、各画像には物体の一部しか入らないことがわかる。上面図において周辺部の線付の点はカメラ位置と光軸を表す。その内側の点の集合は物体の概形を表す特徴点である。なお、画像列の 2 画像以上に現れない特徴点は表示されていないため、特徴点は主に物体辺縁部に集中している。

上記の設定の違いによって、与えられた 2 次元情報からの形状復元の困難さが異なるが、これを総合的に表す指標として、特徴点の観測率を用いる。

復元には、全回転量が 360 度の注視回転を仮定した初期値を用いた。観測に誤差は加えず、復元結果の RMS 残差が 0.1pixel 未満の場合を復元成功と判定した。

図 10(a) は円周撮影法において、観測率と復元成功率を示す。白領域は成功、灰色領域は修正によって成功、黒領域は失敗を表す。平均出現率と復元成功率には相関が認められ、与えられた問題が解けるかどうかを表す指標として平均出現率が適切であることが確認できる⁶。観測率が 15% 以上の全ての場合で注視回転の初期値から正解が得られている。10% 程度でも、解の修正により正解が得られる場合があった。平均出現率の観点から本手法は、全周復元の比較的広い範囲の問題に対して、設定した初期値から非線形最適化により正解を得られることが示された。

同様に図 10(b) に疑似広角撮影法における復元成功率を示す。画像数は 6×10 (10ヶ所から各 6 枚), 6×20 , 6×30 の 3 通り、他のパラメータは前記実験と同じである。同一位置から複数の画像を撮影しているが、この情報は問題を解く際の制約には利用しておらず、初期値も円周撮影と同じものを用いた。復元結果の上面図の一例を図 11 に示す。平均出現率が 11% 以上の全ての場合に注視回転の初期値から正解が得られた。また、平均出現率が 5% 程度でも、解の修正により正解が得られる場合があった。円周撮影よりも疑似広角撮影の方が、設定した初期値からより離れた問題設定となっているが、結果的には、平均出現率が低く、カメラ-物体間距離が近い場合でも、円周撮影の場合より成功率が高い。これは、同じ画像対応数でも、疑似広角撮影の方がカメラ位置の空間的分布が広がるために生じる利点と考えられる。なお、ここでは用いなかったが、疑似広角撮影では、同一位置から撮影した複数の画像の

⁶文献 [4] では、観測率が 16% の問題を扱っている。

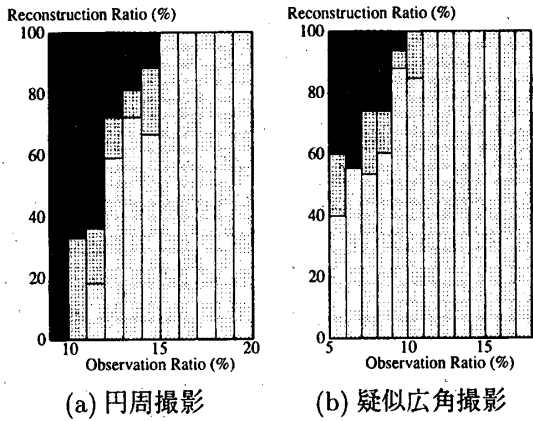


図 10: 平均出現率と復元成功率

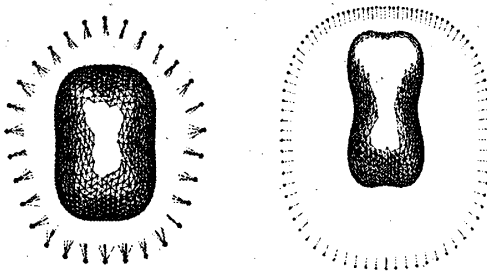


図 11: 疑似広角撮影法による復元結果
図 12: 速度や距離に偏りがある場合の復元結果

カメラ位置に対して、カメラ位置が同一であることを利用して定式化することにより、推定すべきパラメータが減少し、更に容易かつ高精度に解を求めることが期待できる。このように、全周復元に対しては、円周撮影よりも疑似広角撮影の方が、より適切であると考えられる。

カメラの移動速度や距離に偏りのある場合(図12)に対しても良好な結果が得られることを確認しており、設定した初期値は汎用性の高いものと言える。

3.3.2 実画像を用いた復元

ここでは、実画像列からの復元結果を2例示す。

図13に、原爆ドームを撮影した画像29枚中の2枚と復元結果の上面図を示す。これは、設計図と比べて良好な復元であることを確認している。この実験例では、建物から離れた位置から撮影を行っており、観測率は17%と高いため、注視回転の初期値から局所解に陥ることなく正解が得られ、解の修正は不要であった。

図14の例では、建物に近い位置から撮影しており、建物のごく一部しか撮影されていない画像を含む。そのため、観測率は6%と低い。また、近距離

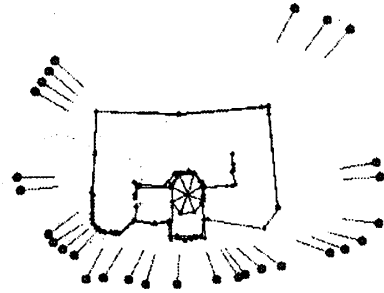


図 13: 原画像の例と復元結果原爆ドーム ($P = 122, F = 29, r = 17\%$)

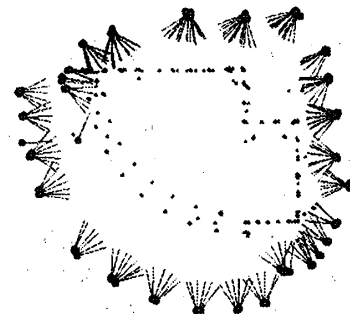
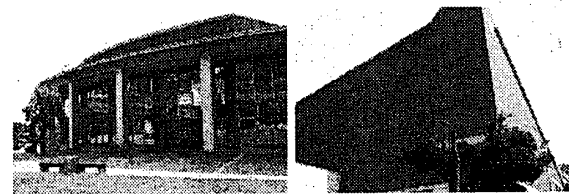


図 14: 原画像の一部と復元結果講堂 ($P = 189, F = 280, r = 6\%$)

から広範囲を撮影するため、光軸を上下左右に大きく動かしながら撮影しており、疑似広角撮影となっている。この例では注視回転の初期値からでは部分的に正しくない結果が得られた。その後、解の修正を行ったところ、図14の復元結果が得られた。図右上方の建物の垂直な角が良好に復元されているなど、建物全体が良好に復元できていることが確認できる。

その他、100枚以上からなる全周実画像列5例においても良好な復元結果が得られた。

4 逐次復元法

前節で述べた一括復元後の解の修正は、一括復元の解の大部分が正しい場合に、一部に生じた誤りを修正するものであった。しかし、注視回転とは大きく異なる問題設定や、カメラが物体に近づき観測率の極めて低い場合には、一括復元の結果中で信頼できる部分が減少し、修正さえ不可能となる場合が現れる。このような問題設定では、全情報を一度に用いることを諦める必要がある。

多くの場合、全情報に対する一括復元は不可能であっても、入力情報の部分集合に対しては一括復元可能である。その部分復元(または、初期復元)を元に部分集合を逐次拡大し、それに伴い部分復元を全体での最適解に更新する方法が考えられる。

具体的には、処理対象画像 f を $s \leq f < e$ の範囲に限定し、この範囲で L 画像以上に出現する特徴点 p を用いる。この条件を満たす添字 (f, p) の集合を $S(s, e, L)$ と定義する。最適化の目的関数は、この集合の全要素に亘る総和とする。

$$E(\mathbf{x}; s, e, L) := \sum_{S(s, e, L)} |\mathcal{P}[R_f \mathbf{s}_p + \mathbf{t}_f] - \mathbf{u}_{fp}|^2 \quad (13)$$

ここで \mathbf{x} は、求めるべき全パラメータを含むベクトルである。画像数を F とすると、本来最小化すべき式(3)の目的関数は $E(\mathbf{x}; 0, F, 2)$ と書ける。なお、前節の修正操作は、集合を定義する3つのパラメータの内 L のみを用いたものであり、この枠組に含まれている。

逐次復元の手順は次のようになる。

(i) 対象画像列から、簡単に復元できる初期部分情報 $S(s_0, e_0, L_0)$ を選び、初期復元 $\mathbf{x}(s_0, e_0, L_0)$ を得る。

この時の復元が局所解に陥らないようにするには、次のような条件を満たすように3つのパラメータ s_0, e_0, L_0 を選ぶと良い。

- カメラが物体から比較的遠く、各画像の視野の重なりが多く共通特徴点が多い
- 部分画像列の最初と最後のカメラ位置が十分に離れている

(ii) 部分情報 $S(s_i, e_i, L_i)$ を逐次拡大し、その添字集合に対応する最適解 $\mathbf{x}(s_i, e_i, L_i)$ を求め、最終的に正解 $\mathbf{x}(0, F, 2)$ を得る。解の更新 $\mathbf{x}(s_{i-1}, e_{i-1}, L_{i-1}) \rightarrow \mathbf{x}(s_i, e_i, L_i)$ の方法には、複数の選択肢(詳細は次節で述べる)が存在し、選択肢の選び方によっては局所解に陥り正解に辿り着くことができない。従って、正解に辿り着くような選択肢の経路を、試行錯誤によって探索する必要がある。

適切な操作を選択するためには、局所解の判定が必要である。この判定は、現在の推定値 \mathbf{x} による式

(13)の残差によって、ある程度行うことができる。特徴点位置精度が1画素以内のデータに対しては、正しい解のRMS残差は概ね1画素程度の量となるので、それより大きめの閾値を設定すれば良い。残差が十分に小さい場合、推定値は局所解ではないと考え、新しい画像を追加する。残差が大きい場合、局所解と考え、修正の選択肢を順に試して局所解からの脱出を図る。全ての方法に失敗したら、新規画像を追加する前の状態にバックトラックする。このため、画像列の追加や、解の更新の履歴はスタックで管理しておく。

4.1 解の更新の選択肢

ここでは、部分画像列に基づく現在の推定を $\mathbf{x}_i := \mathbf{x}(s_i, e_i, L_i)$ と書く(以後、添字 i は省略し、添字 $i+1$ は $+$ と略記する)。

本節における解の更新とは、 \mathbf{x}, s, e, L に基づいて $P_+ := (s_+, e_+, L_+)$ を決め、非線形最適化の初期値 $\tilde{\mathbf{x}}(s_+, e_+, L_+)$ を構成し、最適化 $\tilde{\mathbf{x}}_+ \rightarrow \mathbf{x}_+$ を行うことである。

解の更新を $P \rightarrow P_+$ の変化によって分類すると、以下の5種類が考えられる。ただし、更新後に定義域を越えるものは除く。

- $P_+ = (s, e+1, L)$: 新規1画像の追加。
- $P_+ = (s-1, e, L)$: 新規1画像の追加。
- $P_+ = (s, e, L)$: 情報集合は変えず、局所解からの脱出を行う。
- (d), (e) $P_+ = (s, e, L \pm 1)$ これらは、前節で述べた解の修正操作である。

(a), (b) の場合、画像の追加に伴い、推定すべきパラメータも増加する。具体的には、追加した画像を撮影したカメラ(以下、新規カメラと呼ぶ)の位置・姿勢、及び、出現画像数が L 枚以上となり新たに復元可能となった特徴点(以下、新規特徴点と呼ぶ)の3次元位置が追加される。これらの未知パラメータは \mathbf{x} で推定されていないので、 $\tilde{\mathbf{x}}_+$ を構成する際に新たに推定する必要がある。これは、以下の方法で行うことができる。

- 新規カメラの位置・姿勢

追加した画像に既存の特徴点が6つ以上写っているなら、新規カメラの位置は外部パラメータ較正の手法により推定できる。これは、次の最小化により実現できる。

$$\min_{\mathbf{t}_f, R_f} \sum_p |\mathcal{P}[R_f \mathbf{s}_p + \mathbf{t}_f] - \mathbf{u}_{fp}|^2 \quad (14)$$

R_f は回転行列であるため、上式は解析的には解けないが、線形近似により初期推定を行い、反復改良することは可能である。

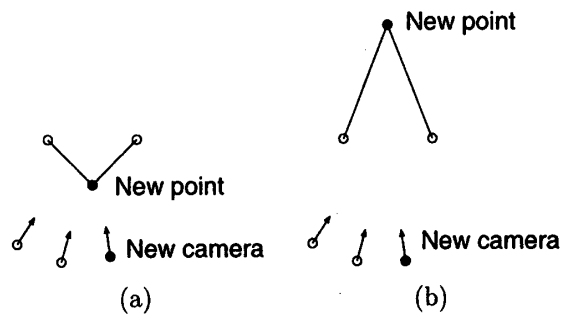


図 15: 物体形状の鏡映反転

なお、この総和に用いる特徴点 p を L 画像以上に現れるものに限定すると、 \hat{x}_+ は L の関数となる。このとき、 L の値によって正解の収束圏に入る場合と入らない場合に分かれる可能性がある⁷。 $L=2$ として 2 画像のみに現れる特徴点も利用すると推定精度が悪くなることが多いので、通常はもう少し大きく $L=4$ 程度にする方が良い結果を得やすい。これは、特徴点の観測数が多いほど最小 2 乗法の平均化効果で、推定の信頼性が高まるためである。

● 新規特徴点の位置

新規特徴点が既存の 2 画像以上で観測されていれば、それらの画像を撮影したカメラの位置・姿勢は既知なので、特徴点の位置は容易に推定できる。

特徴点は、カメラ位置 $-R_f^T t_f$ と、投影面上の特徴点位置を結ぶ直線上にある。この直線の方法ベクトルは $d_f := R_f^T (u_{fp} | l)^T$ である。これらの直線の交点、もしくは、これらの直線からの距離の 2 乗和を最小化する点は、次式のように線形最小 2 乗法で表される。

$$\arg \min_{s_p} \sum_f |d_f^\perp (s_p + R_f^T t_f)|^2 \quad (15)$$

ここで、 $x^\perp := I - xx^T/x^T x$ は正射影の行列である。

これらの方法により推定された情報は、既存カメラの位置・姿勢や、既存特徴点の位置を真値であると仮定した計算により推定されているため、時には信頼性の乏しい推定値となる。これを初期値として用いた最適化 $\hat{x}_+ \Rightarrow x_+$ の結果、様々な局所解に陥る可能性がある。代表的な局所解は次のようなものである。

● 形状の鏡映反転 (図 15)

本来、新規特徴点が図 15(a) の位置に復元されるべきところを、鏡映反転して図 15(b) の様に復元される。このとき、新規特徴点は無限遠近くまで「飛ばされる」ことが多い。

⁷ L をどう選んでも局所解しか得られない場合、入力特徴点に欠陥があるので、特徴点を追加・修正すべきである。

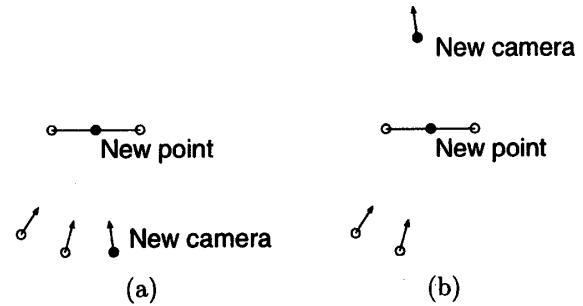


図 16: カメラ位置の鏡映反転

● カメラが鏡映対称位置に出現 (図 16)

本来、図 16(a) の位置に復元されるべき新規カメラが、物体に対して反対側の図 16(b) の位置に復元される。線形投影や 2 画像の場合には、誤ったカメラ位置を正しい位置に変換することは比較的容易であるが、透視投影の複数の画像を用いている本手法では、カメラ位置を修正する解析的な変換は明らかではない。なお、この問題は、復元対象が平面に近い場合に起きやすいため、これを回避するには、様々な距離の特徴点を与えることが有効である。

最適化の結果として、これらの局所解が出現した場合でも、以下の幾つかの方法で正解に補正することが期待できる。これが (c),(d),(e) の $P_+ = (s, e, L + \{0, 1, -1\})$ の場合である。

(c) カメラ位置の再推定

あるカメラの現在の推定値を破棄して、較正によりカメラ位置を再推定する。ここで式 (14) の初期値として現在の推定に依存しない線形近似解法を用いることによって、式 (13) の最適化のみでは回避できない局所解からの脱出が期待できる。

(d),(e) 特徴点位置の再推定

最終目的は $L=2$ として式 (13) の残差 2 乗和を最小化することであるが、形状復元の過程では 2 画像にしか現れない特徴点は局所解の原因と成り易い。そこで、一度 L を大きくして、推定精度が悪い特徴点の推定位置を破棄し、再び L を 2 に近づけながら x を更新する事で、特徴点の推定位置は改善され、式 (13) のみでは回避できない局所解からの脱出が期待できる。これは、前節でも述べた。

なお、(d)(e) 各一回の対として用いるのではなく、それぞれ k 回適用し、 $L \rightarrow \dots \rightarrow L+k \rightarrow \dots \rightarrow L$ となるように用いることが多い。

4.2 実験

本手法の適用例として、建物群に囲まれた中庭での環境復元を挙げる。画像は図 17 を含む 65 枚、特徴点数 215、観測率 13% である。図 18 に復元過程を示す。(a) は、右上の建物の前方でカメラが平行



図 17: 中庭の画像列 (65 枚中 2 枚)

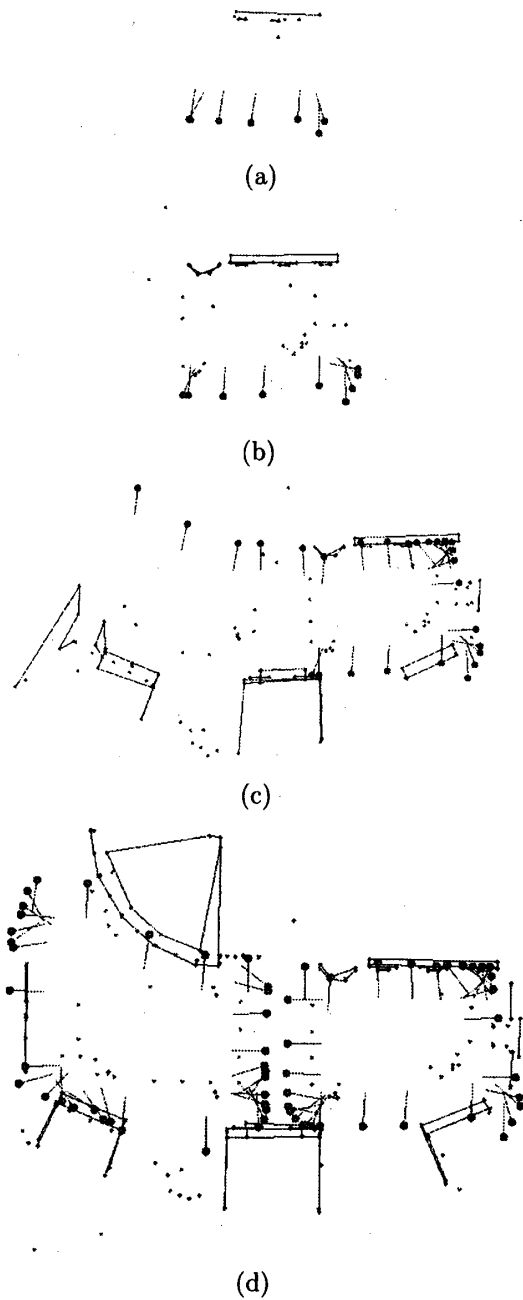


図 18: 逐次復元例

移動している部分画像列からの初期復元結果である。(b) や (c) 付近で局所解が生じるため多くの探索を要するが、ここでは探索後の正解のみを示した。(d) は最終結果である。特徴点群 (建物群) の内部で、カメラが左右 2 つの四角形状の軌跡を描いている様子が復元されている。これは自動処理可能であり、計算時間は十数秒である。

5 カメラ姿勢情報を用いた線形形状復元

ここでは、画像間に対応づけられた特徴点の 2 次元座標に加えて、カメラの姿勢を既知とすることによって、各特徴点と各カメラの 3 次元位置の推定を線形化する手法 [18] を示す。姿勢情報は、三軸角度センサ等を用いて得ることができる。この定式化は、線形投影を仮定せず、回転や並進運動に関する制約も、特徴点の隠蔽に関する制約も必要としない上、固有値問題によって解が明確に表されるという特徴を持つ。

目的関数 (3) を線形化するには、カメラの姿勢 R_f を既知とする他に、次式のような重みを用いる必要がある。

$$\sum_{(f,p)} \lambda_{fp}^2 |\mathcal{P}[R_f s_p + t_f] - u_{fp}|^2$$

$$= \sum_{(f,p)} \left\| \begin{bmatrix} 1 & 0 & -u_{fp} \\ 0 & 1 & -v_{fp} \end{bmatrix} [R_f | I] \begin{bmatrix} s_p \\ t_f \end{bmatrix} \right\|^2 \quad (16)$$

ただし、 $u_{fp} = (u_{fp}, v_{fp})^T$ は焦点距離 1、画像中心を (0,0) とする正規化カメラ座標系で表された画像座標である。重み λ_{fp} は射影的奥行きであるが、透視投影の中にある λ_{fp} による除算と打ち消し合い、変形後には現れない。

この目的関数は、未知の特徴点座標 s_p と並進 t_f に関する 2 次同次式となるので、次のように書ける。

$$x^T A x$$

ここで $x = [s_1^T \ s_2^T \ \dots \ s_P^T \ t_1^T \ t_2^T \ \dots \ t_F^T]^T$ 、 A は次のような構造を持つ対称行列である。

$$A = \begin{array}{c|ccc} S_1 & U_{11} & \dots & U_{1F} \\ & \vdots & & \vdots \\ & & & \\ & S_P & U_{P1} & \dots & U_{PF} \\ \hline & & T_1 & & \\ & \text{対称} & & \ddots & \\ & & & & T_F \end{array}$$

この行列の各要素は次のようになる。

$$\begin{aligned} S_p &= \sum_f R_f^T P_{fp} R_f \\ U_{fp} &= R_f^T P_{fp} \\ T_f &= \sum_p P_{fp} \end{aligned}$$

ここで、

$$P_{fp} = \begin{bmatrix} 1 & 0 & -u_{fp} \\ 0 & 1 & -v_{fp} \\ -u_{fp} & -v_{fp} & u_{fp}^2 + v_{fp}^2 \end{bmatrix}$$

である。ただし、当該特徴点はその画像で観測されていない場合は $P_{fp} = 0$ とする。

5.1 解法

入力とする u_{fp} や R_f が誤差を含む場合、真の解は $\mathbf{x}^T A \mathbf{x} = 0$ を満たさないが、 $\mathbf{x}^T A \mathbf{x}$ を最小化する \mathbf{x} を最小 2 乗の意味での最適解と考えることができる。通常、このような問題の最適解は A の最小固有値に対応する固有ベクトルとして得られるが、本問題では座標系並進の自由度があり、これに対応して A は最適解以外に 3 つの 0 固有値を持つ。

また、行列 A の典型的な大きさは 1000×1000 を越えるため、計算効率を考慮する必要がある。よって、逆乗乘法に基づく次のような方法を用いる。適当な初期ベクトル \mathbf{x}_0 を乱数で生成し、次式の計算を収束するまで繰り返す。

$$\mathbf{x}_{k+1} = \mathcal{N}[A^-(I - UU^T)\mathbf{x}_k]$$

ここで、 $\mathcal{N}[\cdot]$ は正規化演算子、 A^- は一般逆行列である。 U は、並進の自由度に対応する 3 つの固有ベクトルを列ベクトルとして持つ ($AU = 0$) 行列で、次のように定義できる。

$$U = \tau[I \mid \cdots \mid I \mid -R_1^T \mid \cdots \mid -R_F^T]^T$$

ここで、 $\tau (= 1/\sqrt{F+P})$ は各列ベクトルを単位の大きさにするための定数である。 I は 3×3 単位行列で、これは P 個並ぶ。

右辺の行列 $A^-(I - UU^T)$ を明示的に計算することは時間の無駄で、実際の計算は $\mathcal{N}[L(L^T(\mathbf{x}_k - U(U^T \mathbf{x}_k)))]$ の順序で行う。ここで、 $A^- = LL^T$ は、 A の Cholesky 分解において 0 の対角要素に注意する事によって得られる。逆乗乘法の大半の計算時間は、最初に一度だけ行うこの分解に費やされる。

問題の規模や用いる計算機にも依るが、後の実験では PentiumIII 1GHz の Linux PC 上の C/C++ 言語でこの算法を実装し、1 秒前後で解を得ている。

5.2 誤差の影響

ここでは、回転行列 R_f に含まれるノイズの影響を解析する。 R_f は真値 \bar{R}_f と微小なノイズの積に分解できる。これに対応して、式 (17) の行列 A は、真値 \bar{A} とノイズによる微小変動 εB の和に分解され、次のように書ける。

$$\begin{aligned} R_f &\approx \bar{R}_f(I + \omega_f[\mathbf{n}_f]_x) \\ A &= \bar{A} + \varepsilon B \end{aligned}$$

ここで、 $I + \omega_f[\mathbf{n}_f]_x$ は微小回転 (その回転軸は \mathbf{n}_f 、回転角は ω_f) の 1 次近似であり、 \mathbf{n}_f は単位ベクトルである。微小な ω_f の代表値 (標準偏差や絶対値最大要素などが考えられる) を ε とすると、 \bar{A} と B のノルムは同等と考えて良い。 \bar{A} の固有値を λ_i 、対応する固有ベクトルを \mathbf{e}_i とすると、 A の最小固有値 λ_1 と対応するベクトル \mathbf{e}_1 の誤差は、摂動定理により、次のように近似できる。

$$\begin{aligned} \Delta \lambda_1 &= \varepsilon (\mathbf{e}_1^T B \mathbf{e}_1) + O(\varepsilon^2) \\ \Delta \mathbf{e}_1 &= \varepsilon \sum_{i \neq 1} \frac{\mathbf{e}_i^T B \mathbf{e}_1}{\lambda_i} \mathbf{e}_i + O(\varepsilon^2) \end{aligned}$$

ここから、形状誤差はノイズレベル ε に比例することが判る。また、最小以外の固有値 λ_i が微小であれば、対応する固有ベクトル方向の誤差が大きくなる。

このように、最小から 2 番目以降の固有値・ベクトルは、形状推定の数値的不安定性を示す。解法として乗乘法を用いた場合、2 番目以降の固有値が最小固有値に近いときには収束が遅いが、この状況は復元形状が数値的に不安定であることも示している。また、文献 [20] では、式 (3) の Hesse 行列の固有値を解析し、各特徴点が多く方向から撮影されているほど第 2 固有値は大きくなり、形状復元の曖昧さが減ると述べているが、本手法の行列 A にも同様の傾向があると考えられる。

5.3 実験

ここでは、提案手法の有効範囲を示すためにシミュレーション実験を行う。与えられたカメラ姿勢情報に誤差がある場合を考慮し、誤差による解の挙動を調べ、角度取得に要求される精度を推定する。

単純な幾何形状を単純な軌跡に沿って投影した数値データを用いたのでは、実画像で生じる性質を忠実に再現できない。そこで、現実的なシミュレーションを行うために実画像列に基づいて入力データを合成した。その合成手順を以下に示す。

1. 実シーンを撮影した画像列に対して、特徴点を入力し前章までに述べた方法で物体の 3 次元形状とカメラ位置・姿勢を復元する。

なお、原画像は 1024x768 画素で、焦点距離は約 1000 画素である。

2. 復元されたカメラ姿勢にノイズを加えたものを用いて、復元された特徴点を再投影し、画像中の特徴点座標を得る。このとき、原画像で隠蔽されている特徴点は再投影しない。
3. このように再投影された特徴点 2 次元座標と 1. で得られたカメラ姿勢を、ここでの入力として用いる。

上記手順 1. で得られた推定値は、撮影対象の真の形状とは必ずしも一致しないが、手順 2. で合成された特徴点軌跡に対しては、意図的に誤差を加えない限り真値であり、以後の復元形状評価の基準となる真の形状が判っていることになる。これは、実画像に基づく合成データであるため、実物体の複雑な形状や、実世界での現実的な隠蔽効果が継承されており、実データを用いた場合に近い固有値分布を持つ A を再現できると考えられる。

各カメラ姿勢には次のようにして誤差を加える。

1. 標準正規乱数を要素とする 3 次元ベクトル n を作り、これを正規化することによって単位球面上で一様に分布する方向ベクトル $N[n]$ を得る。
2. $\omega(\text{rad})$ は、平均 0 で、標準偏差 $\sigma(\text{度})$ の回転角とする。なお、以下で σ の値を表示する際には度を単位とする。
3. 微小回転行列 $Q = \exp(\omega[N[n]]_x)$ を作り、与えられた R_f の右から掛ける。

このような誤差を加えた場合の復元形状には大きく 2 種類の劣化が見られる。一つは形状全体の伸縮等の変形であり、この影響は、復元形状と真値を最適に重ねる 3 次元のアフィン変換を求め、その行列の単位行列からのずれによって評価できる。但し、アウトライアの影響を避けるため、残差が閾値を越える点は位置合わせの対象外とする。第二の劣化は各点に固有のもので、異なる乱数系列で誤差を加えたときに、各特徴点は真の特徴点位置を中心として「振動」する。観測画像数が少ない特徴点において、この振動が特に激しい。これらの劣化の方向や程度は前章の摂動定理によって予言されるものである。

5.3.1 実験結果

図 19 の 2 枚を含む 144 枚からなる画像列を用いた結果を示す。右の画像のように建物のごく一部しか含んでいない画像も多く、復元が困難な画像列で

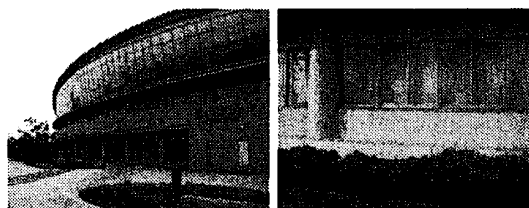


図 19: スタジアムの画像列 144 枚中 2 枚

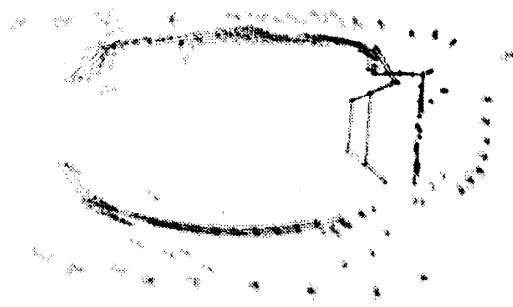


図 20: 復元結果 ($\sigma = 2$)

ある。特徴点数は 239 で、平均 11 枚の画像で対応づけられている。

図 20, 21 は、異なる誤差を加えて得られた複数の結果を重ねて表示したものである。復元の度に結果が異なる部分は不鮮明となっている。図 20 に示す $\sigma = 2$ の場合は比較的良好な結果が得られている。一方、図 21 に示す $\sigma = 5$ の場合の復元結果には、かなりの劣化が見られる。形状の伸縮も深刻で 3 次元アフィン変換では除去されないが、式 (3) の初期値に用いると、容易に最適解を得ることができた。

他の実験でも、標準偏差が 5 度程度なら、非線形最適化の初期値としては十分であると言える。

6 まとめ

本研究では、非線形最小 2 乗法に基づく画像列からの 3 次元形状復元問題を対象に、以下の問題を

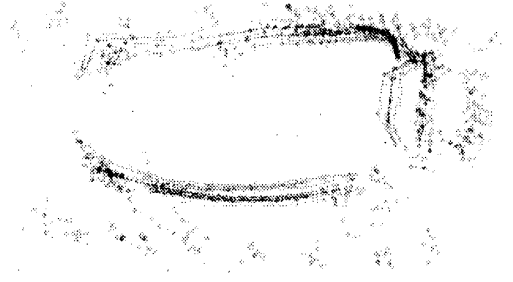


図 21: 復元結果 ($\sigma = 5$)

扱った。

- 非線形最適化計算の高速アルゴリズムとして「ブロック対角行列による前処理つき共役勾配法 (BDCG 法)」を提案しその評価を行った。
- 全周形状復元問題に対して、注視回転という簡単なモデルを用いて撮影位置と姿勢の初期値を与え、全情報を一度に用いて解く方法を提案した。さらに、この方法だけでは正解が得られない場合を考慮して、解の修正により局所解から抜け出し正解を得る方法について述べた。
- 一括復元が困難な場合を対象として、部分情報から始め徐々にデータを増やす逐次復元法を示し、実画像に適用した。
- 撮影時のカメラの角度を用いた形状復元の線形化法を示し、角度測定誤差と形状復元精度の関係を調べた。

今後の課題として、特徴点追跡との統合や、より広範な問題設定への対応が挙げられる。

参考文献

- [1] R. Szeliski and S. B. Kang: "Recovering 3D Shape and Motion from Image Streams using Non-Linear Least Squares," CVPR, pp.752-753, (1993).
- [2] E. V. E. Kumar, A. Tirumalai and C. Jain: "A Non-linear Optimization Algorithm for the Estimation of Structure and Motion Parameters," CVPR, pp.136-153, (1989).
- [3] R. I. Hartley: "In Defence of the 8-point Algorithm," PAMI 19-6, pp.580-593, (1997).
- [4] C. Tomasi, T. Kanade: "Shape and Motion from Image Streams under Orthography: a Factorization Method," IJCV 9(2), pp.137-154, (1992).
- [5] P. F. McLauchlan: "A Batch/Recursive Algorithm for 3D Scene Reconstruction," CVPR, pp.738-743, (2000).
- [6] T. Sato, M. Kanbara, H. Takemura and N. Yokoya: "3-D Reconstruction from a Monocular Image Sequence by Tracking Markers and Natural Features," 14th Vision Interface, pp.157-164, (2001).
- [7] M. Pollefeys, F. Verbiest, and L. V. Gool: "Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery," 7th ECCV, pp.II-837-851, (2002).
- [8] B. Triggs, P. F. McLauchlan, R. Hartley and A. W. Fitzgibbon: Bundle adjustment — A modern synthesis, *Lecture Notes in Computer Science*, No.1883, pp.298-375, (2000).
- [9] C. Tomasi and T. Kanade: "Detection and Tracking of Point Features," CMU-CS-91-132, (1991).
- [10] 金澤靖, 金谷健一: "段階的マッチングによる画像モザイク生成," 信学論, Vol. J86-D-II, No. 6, (2003-6).
- [11] 右田剛史, 天野晃, 浅田尚紀: "3次元形状・運動復元のための高速非線形最適化計算法," 情報処理学会論文誌, vol.44, no.11, pp.2864-2872, (2003).
- [12] D. Jacobs: "Linear fitting with missing data: Application to structure-from-motion and to characterizing intensity images," CVPR, pp.206-212, (1997).
- [13] 棕木雅之, 右田剛史, 天野晃, 青山正人, 浅田尚紀: "再投影誤差最小化による建物画像列からの全周形状の一括復元," 情報研報 CVIM-142, pp.81-88, (2004).
- [14] T. Migita, A. Amano and N. Asada: "Entire Building Shape Recovery from Near Distance Images," MVA2002, pp.277-280, (2002).
- [15] D. P. Robertson and R. Cipolla: "Building Architectural Models from Many Views Using Map Constraints," 7th ECCV, pp.II-155-169, (2002).
- [16] 松川昌洋, 神原誠之, 山澤一誠, 竹村治雄, 横矢直和: "GPSと三軸角度センサを用いた多視点ステレオによる屋外環境構造の再構成" 信学技報, PRMU99-266, pp.37-42, (2000).
- [17] C. Früh and A. Zakhor: "3D Model Generation for Cities Using Aerial Photographs and Ground Level Laser Scans," CVPR, pp.31-38, (2001).
- [18] A. Amano, T. Migita and N. Asada: "A Linear Shape from Motion Algorithm using Rotation Information of the Cameras," 16th Vision Interface, pp.299-304, (2003).
- [19] 今野浩, 山下浩: "非線形計画法," 日科技連, (1978).
- [20] R. Szeliski and S. B. Kang: "Shape Ambiguities in Structure from Motion," PAMI 19-5, pp.506-512, (1997).

A 直線探索

式 (10) 中の $\arg \min E(\mathbf{w}_k - \alpha \mathbf{d}_k)$ は、目的関数 $E(\alpha)$ を 2 次関数 $G(\alpha)$ で近似し、 $G'(\alpha) = 0$ となる α を求める。近似の条件は、 $E(\alpha)$ と $G(\alpha)$ が 3 点で一致すること ($E(\alpha) = G(\alpha)$, $\alpha \in \{0, \delta, 2\delta\}$) であり、 δ は条件 ($G(\delta) < G(0)$ かつ $G(\delta) < G(2\delta)$) を満たすように選ぶ。まず、適当な初期値 ($\delta = 1$ や前回の δ の値等) を選び、上記の条件を満たすまで、 $G(0), G(\delta), G(2\delta)$ の大小関係に基づき δ を 2 倍あるいは $1/2$ 倍に変更する。この手順は、上記の条件が緩やかであるため、黄金分割法よりも少ない反復回数で $G(\alpha)$ が定まり最適な α が得られる。