

初期モデルとしてアフィン投影を用いた 多視点透視投影画像からの運動と形状の復元

植芝 俊夫 富田 文明

電子技術総合研究所
つくば市梅園 1-1-4

E-mail: {ueshiba | tomita}@etl.go.jp

カメラが透視投影に従うという条件の下で、複数の画像で得られた対応点からカメラ運動とシーン形状を復元する因子分解法を提案する。アフィン投影カメラモデルから出発して計測行列のランクが4になるような射影的奥行きを逐次的に推定し、得られた計測行列を因子分解して射影空間におけるシーンの3次元情報を復元する。これによって、従来の透視投影画像に対する因子分解法で必要だった fundamental matrix の計算等のノイズに敏感な操作を排除し、すべての視点から得られる情報を均等且つ同時に扱った安定な復元が可能になった。さらに、従来のアフィン投影モデルにおける計量拘束を拡張して透視投影条件下での計量拘束を導き、カメラの内部パラメータが与えられればユークリッド空間における復元が可能であることを示す。

A Factorization Method for Multiple Perspective Views Using Affine Projection as the Initial Camera Model

Toshio Ueshiba, Fumiaki Tomita

Electrotechnical Laboratory
1-1-4 Umezono, Tsukuba-shi, Ibaraki 305, Japan

E-mail: {ueshiba | tomita}@etl.go.jp

In this paper, we propose a factorization-based method for shape and motion recovery from point correspondences in image sequences under perspective camera model. We define a measure representing "rank-4-ness" of the measurement matrix. The structural parameters called *projective depths* are iteratively estimated so that this measure is minimized starting from affine camera model where all the projective depths equal to unity. Camera motion and shape in 3D projective space are then recovered by factoring the measurement matrix computed from the obtained projective depths. Thus the stable reconstruction is attained by utilizing the information from the all views simultaneously and uniformly. Moreover, we derive metric constraints for perspective camera and show that the Euclidean reconstruction is possible if all the internal parameters of the cameras are given. The validity of the proposed method is confirmed by experiments with real data.

1 はじめに

異なる視点から得た複数の画像を用いてカメラ間の相対的位置関係とシーンの3次元情報を同時に復元する問題は *structure from motion* と呼ばれ、コンピュータビジョンの中心的な課題の一つとして多くのアルゴリズムが提案されてきた。その中で、Tomasi-Kanadeによって提案された因子分解法 [1] は簡便かつ安定性の高い優れた手法である。これは、カメラがアフィン投影に従うという仮定のもとでは、画像上で観測される特徴点の2次元座標を並べてできる計測行列が、カメラ運動とシーン中での特徴点の3次元位置を表す2つの行列の積に分解できるという性質を利用したものである。Tomasiらはカメラモデルとして *orthographic projection* を採用したが、その後、*weak perspective* カメラモデルや *paraperspective* カメラモデルへの拡張 [2] も行われている。

最近、因子分解法を透視投影カメラモデルの場合に拡張する手法がいくつか提案されている [3, 4, 5, 6, 7]。透視投影画像に因子分解法を適用する際に困難をもたらすのは、射影的奥行き (*projective depth*) と呼ばれるシーンの構造パラメータが未知であるため、計測行列も未知となり因子分解を直ちに適用できないことによる。したがって、何らかの方法で射影的奥行きを求めることが重要なポイントとなる。

Christy-Horand は、*paraperspective* カメラモデルから出発し、計測行列の線形因子分解を繰り返して逐次的に透視投影モデルに近づける形状復元法を示した [3]。この方法は因子分解の各ステップでいわゆる計量拘束 (*metric constraint*) を解いてユークリッド空間におけるカメラ運動とシーンの3次元構造を求め、それを用いて奥行きパラメータを真値に近づくように更新している。しかし、計量拘束を適用するためにはカメラの内部パラメータが必要であり、カメラがキャリブレーションされていない場合には有効でない。

一方、Sturm-Triggs は、キャリブレーションされていない透視投影画像列に対して繰り返しを必要としない因子分解法を提案した [4]。この手法では、まず2枚の画像のペアを何組か取り出して *fundamental matrix* を計算することにより射影的奥行きを求め、次にそれをもとに構成した計測行列を因子分解してカメラ運動とシーン形状を復元するという方針をとっている。しかし、個々の画像ペアから *fundamental matrix* を独立に求め、それらから計算された射影的奥行きを整合性を保つように統合するという手法では、*fundamental matrix* の推定精度が計測行列や最終的な復元解の精度に直接影響すると考えられる。一般的に言って、2枚の画像

の特徴点の対応から *fundamental matrix* を求める操作は特徴点位置の計測誤差に対して極めて脆弱であることが知られており、そのような過程を経て得られる計測行列が十分信頼性の高いものであるとは限らない。*fundamental matrix* の計算を安定化する試みとして、原点移動とスケールングによって入力データを正規化するテクニック [8] も提案されているが、著者らの経験によればこれでも十分でない場合がある。因子分解法の本来的利点は、多数の視点から得られた情報を均等且つ同時に扱うことにより、特定の視点での情報に誤差が含まれていても全体として安定な結果を得られることにあり、何組かの画像のペアを取り出してその *fundamental matrix* を求めてから因子分解を行うという方針は、この利点を生かしたのものとは言い難い。

本稿では、まず計測行列のランクが4になるような射影的奥行きを求め、次に因子分解によってカメラ運動とシーン形状を復元する手法を提案する。観測誤差が存在する実際の条件下では計測行列が厳密にランク4になるとは限らないが、“ランク4らしさ”を示す評価関数を定義し、これを最小化する非線形最適化として奥行きパラメータの推定問題を定式化する。また、非線形最適化においては真値に十分近い初期値を与えることが重要であるが、我々はアフィン投影カメラモデルに基づけば射影的奥行きがすべて1になることを利用し、これを初期値とする。これによって、すべての視点における情報を均等且つ同時に扱いながら、内部パラメータ未知のカメラで撮像された透視投影画像列から射影空間におけるカメラ運動とシーン形状を復元することができる。さらに、従来のアフィン投影カメラモデルにおける計量拘束を透視投影カメラモデルへ拡張することによって、カメラの内部パラメータが既知ならばカメラ運動とシーン形状をユークリッド空間において復元することができることを示す。本手法の有効性を実画像を用いて検証した実験結果も示す。

2 問題の記述とカメラモデル

2.1 問題の記述

シーン中の P 個の特徴点を3次元ベクトル \mathbf{x}_j ($j = 1, 2, \dots, P$) で表す。 F 個の異なる視点 $i = 1, 2, \dots, F$ から \mathbf{x}_j を観測して各画像にその像が得られたとし、これを2次元ベクトル \mathbf{m}_{ij} で表す。すると、カメラを透視投影に従うものとしてモデル化すれば、その投影式は

$$\lambda_{ij} \begin{bmatrix} \mathbf{m}_{ij} \\ 1 \end{bmatrix} = \mathbf{K}_i \begin{bmatrix} \mathbf{R}_i^\top & -\mathbf{R}_i^\top \mathbf{t}_i \end{bmatrix} \begin{bmatrix} \mathbf{x}_j \\ 1 \end{bmatrix} \quad (1)$$

となる。ここで、 \mathbf{t}_i , \mathbf{R}_i は i 番目のカメラの位置と姿勢を表す3次元ベクトルと3次元直交行列であり、 \mathbf{K}_i は内部パラメータを表す 3×3 上半三角行列である [9]。

$\mathbf{K}_i \mathbf{R}_i^T$ の最初の2行および第3行をそれぞれ 2×3 行列 \mathbf{P}_i^T と3次元ベクトル \mathbf{p}_i^T で表せば、上式は

$$\lambda_{ij} \begin{bmatrix} \mathbf{m}_{ij} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_i^T & -\mathbf{P}_i^T \mathbf{t}_i \\ \mathbf{p}_i^T & -\mathbf{p}_i^T \mathbf{t}_i \end{bmatrix} \begin{bmatrix} \mathbf{x}_j \\ 1 \end{bmatrix} \quad (2)$$

と表される。ただし、 $\lambda_{ij} \equiv \mathbf{p}_i^T (\mathbf{x}_j - \mathbf{t}_i)$ である。

我々の目的は、特徴点の像の位置 \mathbf{m}_{ij} からカメラの運動 $(\mathbf{R}_i, \mathbf{t}_i)$ およびシーンの3次元形状 (\mathbf{x}_j) を求めることである。

2.2 透視投影モデルとアフィン投影モデルの関係

P 個の特徴点の中から任意の1点を取り、これを $\bar{\mathbf{x}}$ とする。また、画像 i における $\bar{\mathbf{x}}$ の像を $\bar{\mathbf{m}}_i$ とする。すると

$$\Delta \mathbf{m}_{ij} = \mathbf{m}_{ij} - \bar{\mathbf{m}}_i, \quad \Delta \mathbf{x}_j = \mathbf{x}_j - \bar{\mathbf{x}} \quad (3)$$

なる平行移動により、画像座標系およびワールド座標系の原点をそれぞれ $\bar{\mathbf{m}}_i, \bar{\mathbf{x}}$ に一致させることができる。この座標変換により、(2)は

$$\begin{aligned} \lambda_{ij} \begin{bmatrix} \Delta \mathbf{m}_{ij} \\ 1 \end{bmatrix} &= \begin{bmatrix} \mathbf{I}_2 & -\bar{\mathbf{m}}_i \\ \mathbf{0}_2^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{P}_i^T & -\mathbf{P}_i^T \mathbf{t}_i \\ \mathbf{p}_i^T & -\mathbf{p}_i^T \mathbf{t}_i \end{bmatrix} \begin{bmatrix} \mathbf{I}_3 & \bar{\mathbf{x}} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_j \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T & (\mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T)(\bar{\mathbf{x}} - \mathbf{t}_i) \\ \mathbf{p}_i^T & \mathbf{p}_i^T (\bar{\mathbf{x}} - \mathbf{t}_i) \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_j \\ 1 \end{bmatrix} \end{aligned}$$

となる。ここで、 $\bar{\mathbf{m}}_i$ が $\bar{\mathbf{x}}$ の像であるから、 $\bar{\mathbf{m}}_i = \mathbf{P}_i^T (\bar{\mathbf{x}} - \mathbf{t}_i) / \mathbf{p}_i^T (\bar{\mathbf{x}} - \mathbf{t}_i)$ 、すなわち $(\mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T)(\bar{\mathbf{x}} - \mathbf{t}_i) = \mathbf{0}_2$ であることを考慮し、さらに $\bar{\lambda}_i \equiv \mathbf{p}_i^T (\bar{\mathbf{x}} - \mathbf{t}_i)$ 、 $\Delta \lambda_{ij} \equiv \mathbf{p}_i^T \Delta \mathbf{x}_j$ とおけば、上式は

$$\begin{aligned} \left(1 + \frac{\Delta \lambda_{ij}}{\bar{\lambda}_i}\right) \begin{bmatrix} \Delta \mathbf{m}_{ij} \\ 1 \end{bmatrix} &= \begin{bmatrix} (\mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T) / \bar{\lambda}_i & \mathbf{0}_2 \\ \mathbf{p}_i^T / \bar{\lambda}_i & 1 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_j \\ 1 \end{bmatrix} \quad (4) \end{aligned}$$

と書ける。これを透視投影カメラモデルと呼ぶ。

カメラから基準点までの奥行き $\bar{\lambda}_i$ に比べてシーンの奥行き方向の変動 $\Delta \lambda_{ij}$ が充分小さければ、上式左辺の $\Delta \lambda_{ij} / \bar{\lambda}_i$ を無視できる。すなわち、(4)の近似として

$$\Delta \mathbf{m}_{ij} = \frac{1}{\bar{\lambda}_i} (\mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T) \Delta \mathbf{x}_j \quad (5)$$

を得る。これをカメラのアフィン投影モデルと呼ぶ。特にカメラの内部パラメータが既知の場合は、このモデルは paraperspective camera と呼ばれ、基準点 $\bar{\mathbf{x}}$ のまわりで(2)の \mathbf{m}_{ij} をTaylor展開して1次項までをとったものに相当することが知られている [10]。

3 射影的復元:アフィン投影から透視投影へ

本章では、カメラの内部パラメータ((1)の \mathbf{K}_i)が未知という条件のもとで、射影空間におけるカメラ運動とシーン形状を復元するための因子分解法について述べる。

3.1 計測行列と射影的奥行き

$\mu_{ij} \equiv 1 + \Delta \lambda_{ij} / \bar{\lambda}_i$ 、 $\mathbf{m}_{ij} \equiv [\Delta \mathbf{m}_{ij}, 1]^T$ 、 $\mathbf{x}_j \equiv [\Delta \mathbf{x}_j, 1]^T$ とおき、また

$$\mathbf{Q}_i^T \equiv \begin{bmatrix} (\mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T) / \bar{\lambda}_i & \mathbf{0}_2 \\ \mathbf{p}_i^T / \bar{\lambda}_i & 1 \end{bmatrix} \quad (6)$$

と定義すれば、すべての画像および特徴点について(4)を行列形式で書き下して

$$\begin{aligned} &\overbrace{\begin{bmatrix} \mu_{11} \mathbf{m}_{11} & \cdots & \mu_{1P} \mathbf{m}_{1P} \\ \vdots & \ddots & \vdots \\ \mu_{F1} \mathbf{m}_{F1} & \cdots & \mu_{FP} \mathbf{m}_{FP} \end{bmatrix}}^{\mathbf{W}} \\ &= \underbrace{\begin{bmatrix} \mathbf{Q}_1^T \\ \vdots \\ \mathbf{Q}_F^T \end{bmatrix}}_{\mathbf{Q}^T} \underbrace{\begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_P \end{bmatrix}}_{\mathbf{X}} \quad (7) \end{aligned}$$

を得る。この $3F \times P$ 行列 \mathbf{W} は計測行列 (measurement matrix) と呼ばれ、(7)はこれがカメラ運動を表す $3F \times 4$ 行列 \mathbf{Q}^T およびシーンの形状を表す $4 \times P$ 行列 \mathbf{X} に分解できることを示している。

一方、 μ_{ij} を画像 i における特徴点 j の射影的奥行き (projective depth) という。透視投影モデルの下ではこれらは未知パラメータであり、従って計測行列も未知である。しかし、 $\mu_{ij} = 1 + \Delta \lambda_{ij} / \bar{\lambda}_i$ であるから、§2.2で述べたように、アフィン投影モデルの下ではこれらはすべて1となる。

3.2 計測行列の因子分解とその射影的不定性

今、射影的奥行き μ_{ij} が与えられて計測行列 \mathbf{W} が計算できるとする。このとき、 \mathbf{W} を(7)のようにカメラ運動 \mathbf{Q}^T とシーンの3次元構造 \mathbf{X} の積に分解することを因子分解と呼ぶ。これは、次のように特異値分解(SVD)を用いて容易に実行できる。

計測行列 \mathbf{W} を

$$\mathbf{W} = \mathbf{V} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N) \mathbf{U}^T \quad (8)$$

と特異値分解する。ただし、 $N \equiv \min(3F, P)$ であり、 \mathbf{U} および \mathbf{V} は、それぞれ

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}_N, \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}_N \quad (9)$$

を満たす $P \times N, 3F \times N$ 行列である。また、 $\sigma_1, \dots, \sigma_n$ は \mathbf{W} の特異値 (非負) であり、その値の大きい順に並んでいるものとする。仮に特徴点 \mathbf{m}_{ij} に計測誤差が含まれないとすれば、(7) からわかるように計測行列のランクは高々 4 であるから、5 番目以降の特異値はすべて 0 になる。よって、 \mathbf{U} および \mathbf{V} の最初の 4 列を取り出したものをそれぞれ $\mathbf{U}^*, \mathbf{V}^*$ とすれば、

$$\mathbf{Q}^T = \mathbf{V}^*, \quad \mathbf{X} = \text{diag}(\sigma_1, \dots, \sigma_4) \mathbf{U}^{*T} \quad (10)$$

が \mathbf{W} の分解の 1 つを与える。

ただし、この分解は一意でない。実際、任意の 4×4 正則行列 \mathbf{H} に対して $\mathbf{W} = \mathbf{Q}^T \mathbf{X} = (\mathbf{Q}^T \mathbf{H}^{-1})(\mathbf{H} \mathbf{X})$ であるから、

$$\mathbf{Q}'^T \equiv \mathbf{Q}^T \mathbf{H}^{-1}, \quad \mathbf{X}' \equiv \mathbf{H} \mathbf{X} \quad (11)$$

も \mathbf{W} の分解を与える。

さらに、射影的奥行き μ_{ij} にも次の不定性がある [4, 7]。すなわち、ある μ_{ij} ($i = 1, \dots, F; j = 1, \dots, P$) について \mathbf{W} が (7) のようにカメラ運動とシーンの 3 次元形状に分解できるとすると、ある i を固定して μ_{ij} ($j = 1, \dots, P$) をすべて α_i 倍しても \mathbf{Q}_i が α_i 倍されるだけで \mathbf{W} が分解可能という性質は保存される。同様に、ある j を固定して μ_{ij} ($i = 1, \dots, F$) をすべて β_j 倍しても \mathbf{x}_j が β_j 倍されるだけで分解可能性に変わりはない。すなわち、 μ_{ij} が分解可能な \mathbf{W} を与えるならば、任意の $\alpha_i \neq 0$ ($i = 1, \dots, F$) および $\beta_j \neq 0$ ($j = 1, \dots, P$) に対して

$$\mu'_{ij} = \alpha_i \mu_{ij} \beta_j \quad (i = 1, \dots, F; j = 1, \dots, P) \quad (12)$$

も分解可能な \mathbf{W} を与える正しい射影的奥行きである。

以上をまとめて、ある射影的奥行き μ'_{ij} に関する計測行列 $\mathbf{W}(\mu')$ を因子分解して得られるカメラ運動 \mathbf{Q}'_i および形状 \mathbf{x}'_j とそれらの真値 \mathbf{Q}_i , \mathbf{x}_j とは、未知の正則行列 \mathbf{H} および未知の定数 α_i, β_j ($i = 1, \dots, F; j = 1, \dots, P$) によって

$$\mathbf{Q}'_i{}^T = \alpha_i \mathbf{Q}_i{}^T \mathbf{H}^{-1}, \quad \mathbf{x}'_j = \beta_j \mathbf{H} \mathbf{x}_j \quad (13)$$

と関係付けられる。これは推定値と真値が未知の 3 次元射影変換で結ばれることを示しており、シーンを 3 次元射影空間と考えたときにその座標枠の取り方に任意性が残ることを表すものと解釈できる [11, 12]。この不定性を残してカメラ運動とシーン形状を復元することを射影的復元と呼ぶ。

3.3 共役勾配法による射影的奥行きの推定

前節までの議論により、射影的奥行き μ_{ij} が得られれば、計測行列を因子分解することによってカメラ運動とシーン形状の射影的復元が可能であることがわかった。では、如何にして μ_{ij} を推定できるだろうか？

仮に μ_{ij} が得られたとすれば、 \mathbf{m}_{ij} に誤差がなければ計測行列 \mathbf{W} のランクは 4 であり、その 5 番目以降の特異値はすべて 0 になる。一方、誤差がある場合でもそれが微小ならば \mathbf{W} は rank 4 に近く、5 番目以降の特異値も 0 に近いと予想される。そこで、 \mathbf{W} の "rank 4 らしさ" を示す指標として

$$J \equiv \sum_{n=5}^N \sigma_n^2 \quad (14)$$

なる評価関数を定義し、これを最小化するような μ_{ij} として射影的奥行きを推定する。

(14) を最小化することは非線型最適化問題であり、反復演算が必要になる。ここでは、汎用的な最小化手法の 1 つである共役勾配法 [13] を用いる。そのためには、以下の 2 点を解決しなければならない。

初期値 非線型最適化では、最適解に十分近い初期値を与えることが重要である。我々の場合は、アフィン投影をカメラの初期モデルとし、 $\mu_{ij} = 1$ を射影的奥行きの初期値とすることができる。

評価関数の勾配 共役勾配法を適用するためには、任意の点で評価関数の勾配 (各パラメータに関する 1 階微分) が計算できなければならない。これは、以下のようにして求めることができる。

\mathbf{W} を (8) のように特異値分解したとき、 \mathbf{U} と \mathbf{V} の第 n 列をそれぞれ $\mathbf{u}_n, \mathbf{v}_n$ とする。すると、 $\sigma_n = \mathbf{v}_n^T \mathbf{W} \mathbf{u}_n$ である。これを微分し、 $\mathbf{W} \mathbf{u}_n = \sigma_n \mathbf{v}_n$, $\mathbf{W}^T \mathbf{v}_n = \sigma_n \mathbf{u}_n$ および $\mathbf{u}_n^T \mathbf{u}_n = \mathbf{v}_n^T \mathbf{v}_n = 1$ であることを用いれば、

$$\begin{aligned} \frac{\partial \sigma_n}{\partial \mu_{ij}} &= \frac{\partial \mathbf{v}_n^T}{\partial \mu_{ij}} \mathbf{W} \mathbf{u}_n + \mathbf{v}_n^T \frac{\partial \mathbf{W}}{\partial \mu_{ij}} \mathbf{u}_n + \mathbf{v}_n^T \mathbf{W} \frac{\partial \mathbf{u}_n}{\partial \mu_{ij}} \\ &= \sigma_n \frac{\partial \mathbf{v}_n^T}{\partial \mu_{ij}} \mathbf{v}_n + \mathbf{v}_n^T \frac{\partial \mathbf{W}}{\partial \mu_{ij}} \mathbf{u}_n + \sigma_n \mathbf{u}_n^T \frac{\partial \mathbf{u}_n}{\partial \mu_{ij}} \\ &= \mathbf{v}_n^T \frac{\partial \mathbf{W}}{\partial \mu_{ij}} \mathbf{u}_n \end{aligned}$$

を得る。したがって、 $\mathbf{u}_n = [u_{1n}, u_{2n}, \dots, u_{pn}]^T$, $\mathbf{v}_n = [\mathbf{v}_{1n}^T, \mathbf{v}_{2n}^T, \dots, \mathbf{v}_{fn}^T]^T$ (\mathbf{v}_{in} は 3 次元ベクトル) と書けば、

$$\begin{aligned} \frac{1}{2} \frac{\partial J}{\partial \mu_{ij}} &= \sum_{n=5}^N \sigma_n \frac{\partial \sigma_n}{\partial \mu_{ij}} \\ &= \sum_{n=5}^N \sigma_n \mathbf{v}_n^T \frac{\partial \mathbf{W}}{\partial \mu_{ij}} \mathbf{u}_n \\ &= \sum_{n=5}^N \sigma_n \mathbf{v}_{in}^T \mathbf{m}_{ij} u_{jn} \quad (15) \end{aligned}$$

となって、評価関数 J の勾配が計算できる。

なお、§3.2で述べたように、 μ_{ij} には(12)で表される冗長性があるので、一般性を失うことなく1つの画像(たとえば $i=1$)および1つの特徴点(たとえば $j=1$)を任意に選んで $\mu_{ij} = \mu_{i1} = 1 (i=1, \dots, F; j=1, \dots, P)$ と固定する。すなわち、独立な奥行きパラメータは $(F-1) \times (P-1)$ 個である。

また、実際の計算では \mathbf{W} の各要素の大きさが極端に異なることによる数値的不安定を避けるため、 $\Delta \mathbf{m}_{ij}$ を正規化[8]しておくことが有効である。ただし、既に各画像において基準点の像 $\bar{\mathbf{m}}_i$ が原点になるように平行移動してあるので、[8]と異なりスケーリングだけを調整する。具体的には、各画像 i について $\|\Delta \mathbf{m}_{ij}\| (j=1, \dots, P)$ の平均値が $\sqrt{2}$ になるように定数を乗じる。

3.4 透視投影画像のための因子分解アルゴリズム

以上より、透視投影画像の因子分解は以下のようなアルゴリズムにまとめられる。

1. P 個の特徴点を F 枚の画像に渡って追跡し、その画素座標値 $\mathbf{m}_{ij} (i=1, \dots, F; j=1, \dots, P)$ を取り出す。
2. 特徴点の1つを任意に選び、各画像 i についてこの点の像が原点となるように(3)によって座標系を平行移動する。さらに、ノルムの平均値が $\sqrt{2}$ になるよう $\Delta \mathbf{m}_{ij} (j=1, \dots, P)$ にスケーリング定数を乗じる。
3. 射影的奥行き初期値を $\mu_{ij} = 1 (i=1, \dots, F; j=1, \dots, P)$ とする。
4. (7)によって計測行列 \mathbf{W} を作り、その特異値分解を求める。
5. (15)によって評価関数 J の勾配を求め、併せてその共役方向を求める[13]。
6. 共役方向への μ_{ij} の改善量 $\Delta \mu_{ij} (i=2, \dots, F; j=2, \dots, P)$ を求める[13]。 $\Delta \mu_{ij}$ による J の変化が十分小さければ7.に進む。そうでなければ $\mu_{ij} \leftarrow \mu_{ij} + \Delta \mu_{ij}$ と更新して4.に戻る。
7. (10)によって計測行列 \mathbf{W} をカメラ運動 \mathbf{Q}^T と形状 \mathbf{X} に因子分解し、さらに2.で乗じた定数で各 \mathbf{Q}_i^T の最初の2行を除いてスケーリングを回復する。

4 ユークリッド的復元: カメラ内部パラメータが既知の場合

§3では、画像上での特徴点の位置のみが与えられたときに射影変換の不定性を残してカメラ運動とシーン形状を復元する手法を示した。本章では、さらにすべてのカメラの内部パラメータが既知である場合に、この射影的復元解からユークリッド空間での3次元構造を求める方法について述べる。

4.1 透視投影カメラモデルの計量拘束

カメラの内部パラメータが既知ならば、(1)において $\mathbf{K}_i^{-1}[\mathbf{m}_{ij}^T, 1]^T$ を改めて $[\mathbf{m}_{ij}^T, 1]^T$ とおき直すことにより、一般性を失うことなく $[\mathbf{P}_i, \mathbf{p}_i] = \mathbf{R}_i$ とすることができる。したがって、今後は $[\mathbf{P}_i, \mathbf{p}_i]$ が直交行列であるとして差し支えない。

§3で得られた射影空間におけるカメラ運動およびシーン形状をそれぞれ $\mathbf{Q}_i^T (i=1, \dots, F)$ 、 $\underline{\mathbf{x}}_j (j=1, \dots, P)$ とする。一方、これらをユークリッド空間で記述したものをそれぞれ $\mathbf{Q}_i^T, \underline{\mathbf{x}}_j$ とすると、両者は(13)で表される3次元射影変換で結ばれる。我々の目的は、この射影変換を表す 4×4 正則行列 \mathbf{H} を求めて、射影空間における記述 $(\mathbf{Q}_i^T, \underline{\mathbf{x}}_j)$ をユークリッド空間におけるそれ $(\mathbf{Q}_i^T, \underline{\mathbf{x}}_j)$ に補正することである。

P 個の特徴点 $\underline{\mathbf{x}}_j (j=1, \dots, P)$ のうち基準点 $\underline{\mathbf{x}}$ に相当するものを $\underline{\mathbf{x}}$ とする。ユークリッド形状 $\underline{\mathbf{x}}_j$ を記述する座標系は、§2.2でこの基準点を原点にとっているから、射影変換 \mathbf{H} は $[\mathbf{0}_3^T, 1]^T$ を $\underline{\mathbf{x}}$ に写す。すなわち、ある非零のスカラー $\bar{\beta}$ が存在して、

$$\underline{\mathbf{x}} = \bar{\beta} \mathbf{H} \begin{bmatrix} \mathbf{0}_3 \\ 1 \end{bmatrix}$$

である。すなわち \mathbf{H} の第4列は $\underline{\mathbf{x}}$ の定数倍となるが、 \mathbf{H} は射影変換行列であり定数倍の任意性があるから、その第4列は $\underline{\mathbf{x}}$ に一致するものとして差し支えない。したがって、 4×3 行列 \mathbf{A} を用いて \mathbf{H} は次の形に表される。

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & \underline{\mathbf{x}} \end{bmatrix} \quad (16)$$

(16)と(6)を(13)の第1式に代入すれば、

$$\mathbf{Q}_i^T \begin{bmatrix} \mathbf{A} & \underline{\mathbf{x}} \end{bmatrix} = \frac{\alpha_i}{\lambda_i} \begin{bmatrix} \mathbf{P}_i^T - \bar{\mathbf{m}}_i \mathbf{p}_i^T & \mathbf{0}_2 \\ \mathbf{p}_i^T & \bar{\lambda}_i \end{bmatrix} \quad (17)$$

を得る。§2.2で、すべての画像についてその座標系原点を基準点 $\underline{\mathbf{x}}$ の像にとっているから、 $\mathbf{Q}_i^T \underline{\mathbf{x}}$ の第1成分および第2成分は0になる。したがって、(17)の(1,4)成分および(2,4)成分は自動的に満たされている。また、(3,4)成分についても、(17)の最初の3列を満たす \mathbf{A} 、 α_i/λ_i を求めた後に α_i を $\mathbf{Q}_i^T \underline{\mathbf{x}}$ の第3成分に一致するように決めればよい。結局、(17)の最初の3列のみを考慮すればよく、これらから $[\mathbf{P}_i, \mathbf{p}_i]$ をその直交性を利用して消去すれば

$$\mathbf{Q}_i^T \mathbf{A} \mathbf{A}^T \mathbf{Q}_i = \frac{\alpha_i^2}{\lambda_i^2} \begin{bmatrix} \mathbf{I}_2 + \bar{\mathbf{m}}_i \bar{\mathbf{m}}_i^T & -\bar{\mathbf{m}}_i \\ -\bar{\mathbf{m}}_i^T & 1 \end{bmatrix} \quad (18)$$

を得る。

(16)と(18)は射影変換 \mathbf{H} が満たすべき制約を表す。アフィン投影の場合に倣って、これらを計量拘束(metric constraint)と呼ぶことにする。

4.2 射影変換の回復

$Q_i^T \equiv [a_i, b_i, c_i]$, $\bar{m}_i \equiv [\bar{m}_i, \bar{n}_i]^T$ とおき、

$$B \equiv AA^T \quad (19)$$

と定義すれば、(18)は

$$\begin{cases} a_i^T B a_i &= (1 + \bar{m}_i^2) \alpha_i^2 / \lambda_i^2 \\ b_i^T B b_i &= (1 + \bar{n}_i^2) \alpha_i^2 / \lambda_i^2 \\ a_i^T B b_i &= \bar{m}_i \bar{n}_i \alpha_i^2 / \lambda_i^2 \\ a_i^T B c_i &= -\bar{m}_i \alpha_i^2 / \lambda_i^2 \\ b_i^T B c_i &= -\bar{n}_i \alpha_i^2 / \lambda_i^2 \\ c_i^T B c_i &= \alpha_i^2 / \lambda_i^2 \end{cases}$$

となる。これらから α_i^2 / λ_i^2 を消去すれば

$$\begin{cases} a_i^T B a_i - (1 + \bar{m}_i^2) c_i^T B c_i &= 0 \\ b_i^T B b_i - (1 + \bar{n}_i^2) c_i^T B c_i &= 0 \\ a_i^T B b_i - \bar{m}_i \bar{n}_i c_i^T B c_i &= 0 \\ a_i^T B c_i + \bar{m}_i c_i^T B c_i &= 0 \\ b_i^T B c_i + \bar{n}_i c_i^T B c_i &= 0 \end{cases} \quad (20)$$

を得る。これは、 4×4 対称行列 B の独立な 10 個の成分に関する線形同次方程式であり、2 枚以上の画像が得られれば線形演算で (定数倍の不定性を残した) 最小自乗解が求められる。

次に、 B から (19) を満たす A を取り出す。 A は 4×3 行列であるから、(半) 正定値行列 B は full rank にはなり得ず、その固有値の 1 つは 0 である。そこで、 4×4 直交行列 T を用いて B を

$$B = T \text{diag}(\nu_1, \nu_2, \nu_3, 0) T^T$$

とスペクトル分解し (ν_1, ν_2, ν_3 および 0 は B の非負固有値)、 T の最初の 3 列を取り出した 4×3 行列を T^* として

$$A = T^* \text{diag}(\sqrt{\nu_1}, \sqrt{\nu_2}, \sqrt{\nu_3}) \quad (21)$$

とすれば、 A が求められる。

なお、 B に定数倍の不定性が残ることから、同様に A にも定数倍の不定性が残る。これは (16) で A と \underline{x} の比が決まらないことを意味するが、このことは問題にならない。なぜなら、(17) よりわかるように、 A を k 倍することはすべての奥行きパラメータ λ_i ($i=1, \dots, F$) を一斉に $1/k$ 倍することを意味し、これは Q_i^T, \underline{x}_i を記述しているユークリッド空間において等方的にスケールを変化させていることに過ぎないからである。画像情報のみを扱っている限りはシーンのスケールを任意に決めることができるので、 A と \underline{x} の比も任意である。

以上のようにして求められた行列 H によって、射影空間で記述されたカメラ運動とシーンの形状をユークリッド空間での記述に変換することができる。

5 実験結果

本手法の有効性を確認するために、実画像を用いた実験を行った。

5.1 積木のシーン

図 1 は 3 つの視点から見た積木のシーンである。まず各画像からエッジ点を抽出し、連続しているものを連結する。次に角検出器を適用して屈曲点を抽出し、これを特徴点とした。特徴点位置を sub pixel の精度で検出することは行っていない。画像間の特徴点の対応は手作業で与え、これらから本手法によってカメラ運動とシーン形状を復元した。特徴点数は 35 点であった。

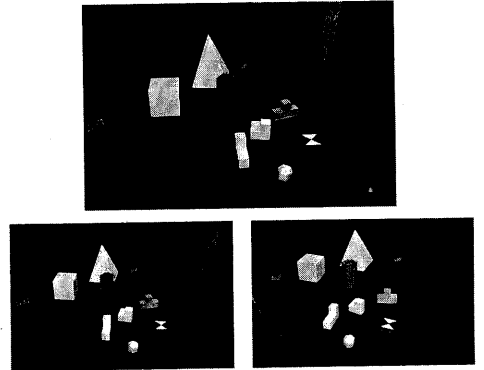


図 1: 積木のシーン

まず、カメラの内部パラメータを与えずに射影的復元を行った。各カメラ i についてその運動を表す Q_i^T が得られるので、任意の 2 画像間でエピポーラ線を計算できる。誤差が無ければ各特徴点はそれに対応するエピポーラ線上に乗るはずである。そこで、各特徴点から対応するエピポーラ線までの距離によって復元の誤差を評価した。表 1 に 2 枚の画像の組み合わせ各々について平均誤差と最大誤差を示す。平均誤差は 1 pixel 以内、最大誤差も 5 pixel 以内に収まっている。さらに、実際に計算されたエピポーラ線の一部を図 2 に示す。3 つの視点が同一直線上にないので、各特徴点において他の 2 枚の画像との間の 2 本のエピポーラ線は一致せずに交差していることがわかる。

combination of images	mean error (pixel)	max. error (pixel)
1st-2nd	0.718	2.22
1st-3rd	0.904	3.07
2nd-3rd	0.709	4.34
total	0.777	4.34

表 1: 特徴点とエピポーラ線との距離

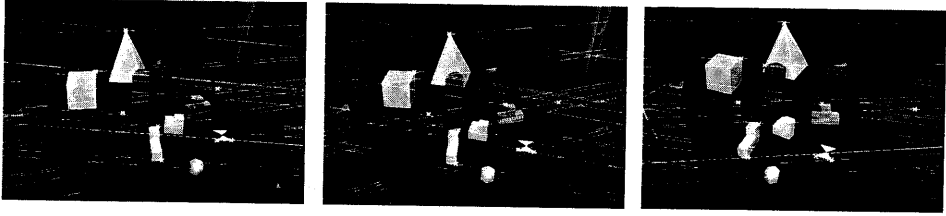


図 2: 積木のシーンから求められたエピポーラ線

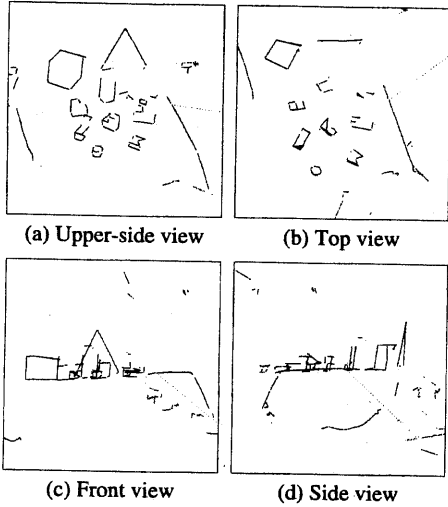


図 3: 積木のシーンの 3 次元情報復元

次に、カメラの内部パラメータを与えてユークリッド的復元を行った結果を図 3 に示す。ただし、特徴点のみ 3 次元情報を復元しても数が少なく形状が把握しにくいので、ユークリッド的復元によって得たカメラの位置と姿勢を我々が開発した 3 眼ステレオビジョンシステム [14, 15] に幾何情報として与え、改めて原画像からステレオによって密な 3 次元情報を抽出した。

5.2 ビルのあるシーン

図 4 は、横方向に移動しながら 5 つの視点からビルを撮像したシーンである。対応付けを与えた特徴点数は 20 であった。このシーンは動きが少なく、隣接する 2 画像間で fundamental matrix を正しく推定するのは困難であった。第 2-3 画像間で fundamental matrix を推定*して計算されたエピポーラ線を図 5 に示す。横方

*正規化を伴う 8 点アルゴリズム [8] で初期値を求め、それを非線形最適化で改善した。

向に移動しているにもかかわらず epipole が画像内にあり、明らかに誤った推定である。したがって、このシーンに対して、fundamental matrix の推定を伴う従来の方法は適用困難と言える。

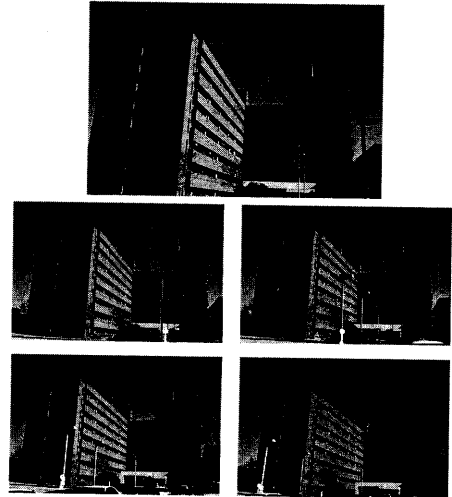


図 4: ビルのシーン

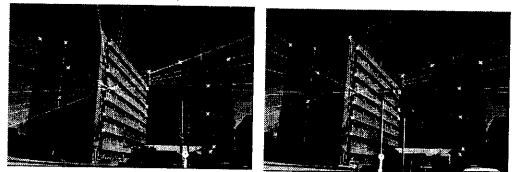


図 5: fundamental matrix 推定の失敗

本手法で射影的復元を行い、同じ画像間のエピポーラ線を計算した結果を図 6 に示す。ビルに対してほぼ真横に移動しているため、エピポーラ線はビル正面の

水平方向線に一致するはずであるが、実際そのように正しく推定されている。

さらに、カメラの内部パラメータを与えてユークリッドの復元を行った結果を図7に示す。積木のシーンと同様、得られたカメラの位置と姿勢を3眼ステレオビジョンシステムに与え、第2,4,5画像から密な3次元復元を行った。直角であるべき箇所がそのように推定されており、ユークリッド空間における3次元情報が正しく復元されていることがわかる。

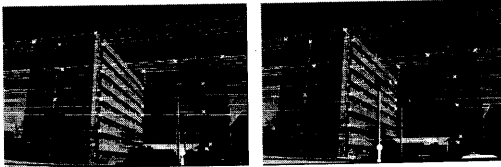


図6: 本手法によるエピポラ線の推定

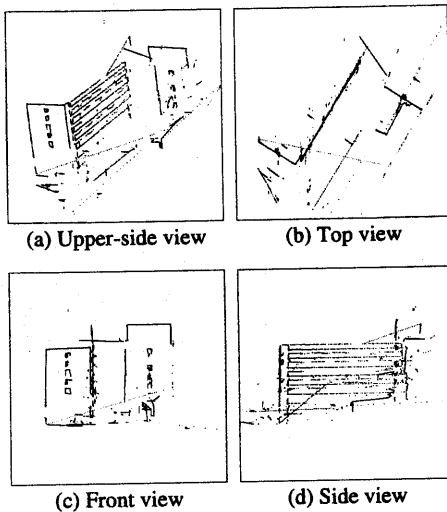


図7: ビルのシーンの3次元情報復元

6 おわりに

複数の透視投影画像の対応点からカメラ運動とシーン形状を復元するための因子分解法を提案した。本手法の特徴は、すべての視点から得られる情報を均等且つ同時に扱い、fundamental matrixの計算等のノイズに敏感な操作を排除していることにある。これによって、射影空間におけるカメラ運動とシーン形状の安定な復元が可能になった。さらに、従来のアフィン投影モデ

ルにおける計量拘束を拡張し、カメラの内部パラメータが与えられればユークリッド空間における復元が可能であることを示した。

現在、射影的奥行きを反復推定には共役勾配法を用いているが、十数回の反復を要しその収束はあまり早くない。評価関数の2階微分(Hessian)が計算できればNewton法やLevenberg-Marquardt法などのより高速な手法が採用できるので、今後はこの点について検討を進めたい。

参考文献

- [1] C. Tomasi and T. Kanade. Shape and Motion from Image Streams under Orthography: a Factorization Method. *International Journal of Computer Vision*, Vol. 9, No. 2, pp. 137-154, 1992.
- [2] 金出武雄, コラッドポールマン, 森田俊彦. 因子分解法による物体形状とカメラ運動の復元. 電子情報通信学会論文誌, Vol. J76-D-II, No. 8, pp. 1497-1505, 1993.
- [3] S. Christy and R. Horaud. Euclidean Reconstruction: from Paraperspective to Perspective. In *Proc. 4th European Conference on Computer Vision*, Vol. 2, pp. 129-140, 1996.
- [4] P. Sturm and B. Triggs. A Factorization Based Algorithm for Multi-Image Projective Structure and Motion. In *Proc. 4th European Conference on Computer Vision*, Vol. 2, pp. 709-720, 1996.
- [5] B. Triggs. Factorization Methods for Projective Structure and Motion. In *Proc. Computer Vision and Pattern Recognition*, pp. 845-851, 1996.
- [6] 出口光一郎, B. Triggs. 多視点透視投影画像の因子分解. 信学技報, No. PRMU96-139, pp. 81-88, January 1997.
- [7] 出口光一郎. 透視投影条件の下での因子分解法による多視点画像からの形状復元. 情処研報, No. CVIM106-6, pp. 35-42, July 1997.
- [8] R. I. Hartley. In Defense of the 8-point Algorithm. In *Proc. 5th International Conference on Computer Vision*, pp. 1064-1070, 1995.
- [9] O. D. Faugeras. *Three-Dimensional Computer Vision*. MIT Press, 1993.
- [10] A. Sugimoto. Object Recognition by Combining Paraperspective Images. *International Journal of Computer Vision*, Vol. 19, No. 2, pp. 181-201, 1996.
- [11] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. 2nd European Conference on Computer Vision*, pp. 563-578, 1992.
- [12] R. Hartley, R. Gupta, and T. Chang. Stereo from Uncalibrated Cameras. In *Proc. Computer Vision and Pattern Recognition*, pp. 761-764, 1992.
- [13] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1988.
- [14] 河井良浩, 植芝俊夫, 石山豊, 角保志, 富田文明. セグメントベースステレオにおける連結性と対応評価. 信学技報, No. PRMU96-135, pp. 49-56, January 1997.
- [15] 植芝俊夫, 河井良浩, 石山豊, 角保志, 富田文明. セグメントベースステレオにおける対応バスの探索. 信学技報, No. PRMU96-137, pp. 65-72, January 1997.