

— 招待論文 —

因子分解法による物体形状とカメラ運動の復元

正員 金出 武雄[†] 非会員 コンラッド ポールマン[†]
正員 森田 俊彦^{††}

A Factorization Method for Shape and Motion Recovery

Takeo KANADE[†], Member, Conrad J. POELMAN[†], Nonmember
and Toshihiko MORITA^{††}, Member

あらまし 時系列画像からの特徴点追跡結果を用いて、3次元的なカメラの相対運動と対象物の形状を求める因子分解法は、線型な定式化を行い、数値計算的に安定な行列の特異値分解を用いたため、安定に運動と形状を復元することができる。本論文では、Tomasiと金出による正射影モデルにおける因子分解法の考え方を発展させ、より中心射影に近い scaled orthographic projection と paraperspective projection への拡張を行うと共に、特徴点追跡の不確かさを考慮に入れて復元の精度を高める重み付け因子分解法を示す。

キーワード 因子分解、追跡、運動、形状、特異値分解、画像

1. まえ

時系列画像をと同時に求め得される重要な手法は、通常連続的手法で(オプティカル^{(2),(3)}、速度場⁽⁴⁾、座標値⁽⁵⁾を求める中間案⁽⁶⁾が枚の画像間で用いて、最小2乗⁽⁷⁾最近では、多観測的⁽⁸⁾な平均化^{(9)~(11)}が行われている。

しかし、従来不安定であると

線型となり、それを反復法で解くためには精度の良い初期値が必要である。これら二つの問題^{(3),(9),(10),(12)}により、実際の画像に対して安定に動作する手法はまだ確立されていない。

時系列画像からの動きと形状の復元問題は、もともと画像の変化を取り扱うものであるから、ノイズに対する慎重な配慮が必要である。特に、計測値の微小差の取扱いには注意を要する。例えば、微小な時間間隔で得た2枚の画像はほぼ同じであるから、その差分はノイズに敏感である。また、動きが微小の場合には、回転と並進がほぼ同様の画像変化をもたらすため、その違いがノイズに埋もれやすい。従って、解を安定化するためには、できるだけ線型な定式化のもとで、ノイズの増幅を巧妙に避け、情報を平均化する工夫が必要である。このような指摘はこれまでいく度も繰り返され、ノイズの分析も行われているが、それを補う実用的手法はまだ確立されていなかった。

これに対して Tomasi と金出⁽¹³⁾は、物体の角などの特徴点を順次追跡⁽¹⁴⁾した上で、行列の因子分解(factorization)を利用してカメラ運動と物体形状を復元する手法を提案した。この手法は運動や形状に対する知識や条件を必要としない。また、正射影モデルに基づいて線型な定式化を行い、数値計算的に安定な行列の特異値分解を用いたため、解は極めて安定である。

[†] カーネギーメロン大学計算機科学科、米国
School of Computer Science, Carnegie Mellon University, Pittsburgh, PA15213, USA
^{††} (株)富士通研究所、川崎市
FUJITSU LABORATORIES LTD., Kawasaki-shi, 211 Japan

ここで式(5)を以下のように書き換えておく。

$$x_{sp} = m_f^T s_p + cx_f, \quad y_{sp} = n_f^T s_p + cy_f \quad (6)$$

但し、

$$cx_f = -i_f^T t_f, \quad cy_f = -j_f^T t_f, \\ m_f = i_f, \quad n_f = j_f \quad (7)$$

とした。ワールド座標原点を対象物の重心にとれば、

$$\sum_{p=1}^P s_p = 0 \quad (8)$$

であるから、式(6)の並進の項 cx_f は以下のように W の各行 x_{sp} の平均に等しい。

$$\frac{1}{P} \sum_{p=1}^P x_{sp} = \frac{1}{P} \sum_{p=1}^P (m_f^T s_p + cx_f) \\ = \frac{1}{P} m_f^T \sum_{p=1}^P s_p + cx_f = cx_f \quad (9)$$

cy_f についても同様である。従って、計測行列 W' の要素 x'_{sp}, y'_{sp} は、

$$x'_{sp} = x_{sp} - cx_f = m_f^T s_p \\ y'_{sp} = y_{sp} - cy_f = n_f^T s_p \quad (10)$$

と、二つのベクトルの内積のみで表され、これを行列で書くと、 W' は、

$$W' = MS = \begin{bmatrix} m_f^T \\ \vdots \\ m_f^T \\ n_f^T \\ \vdots \\ n_f^T \end{bmatrix} \begin{bmatrix} s_1 & \cdots & s_P \end{bmatrix} \quad (11)$$

のように二つの行列積に分解できる。 W' は、 $2F \times 3$ の行列 M と $3 \times P$ の行列 S の積から作られたものであるから、階数は 3 以下である。逆に階数が 3 以下の行列は必ず式(11)のように分解できる。そして、 M にはカメラの回転運動情報だけが、また S には物体の形状情報だけが含まれているので、各々からカメラの姿勢 i_f, j_f と物体形状 s_p を即座に求めることができる。

2.3 特異値分解

時系列画像を処理して得た計測行列 W' を式(11)のように因子分解するには、線形代数における特異値分解⁽¹⁵⁾の手法を用いる。特異値分解定理によれば、任意の $2F \times P$ 行列 W' は、以下の三つの行列の積に分解できる。

$$W' = U \Sigma V^T \quad (12)$$

ここで、 U は $2F \times P$ の直交行列、 Σ は $P \times P$ の対角行列、 V は $P \times P$ の直交行列である。 Σ の対角要素は、

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_P \geq 0 \quad (13)$$

と降順に並んでいる。このとき、 W' の階数が 3 であれ

ば σ_4 以降は 0 となる。従って W' は、 U の最初の 3 列である $2F \times 3$ 行列 U' 、 3×3 の対角行列 Σ' 、 V^T の最初の 3 行である $3 \times P$ 行列 V'^T の積に等しく、

$$W' = U' \Sigma' V'^T \quad (14)$$

である。そこで、

$$\hat{M} = U', \quad \hat{S} = \Sigma' V'^T \quad (15)$$

とおけば、一つの分解結果

$$W' = \hat{M} \hat{S} \quad (16)$$

が得られる。

一般には計測行列 W' にノイズが含まれるため、階数は必ずしも 3 とはならない。しかしその場合でも、 σ_4 以降の特異値は極めて小さく、 W' の階数はほぼ 3 である。従って、同様に σ_4 以降を 0 として分解を行えば、最小 2 乗的意味での最良の近似解を得る。

2.4 拘束条件

特異値分解を用いて W' を式(16)のように二つの行列積に分解できた。しかし実はこの分解は一意ではない。事実、任意の可逆な 3×3 行列 A を用い、

$$M' = \hat{M} A, \quad S' = A^{-1} \hat{S} \quad (17)$$

とすれば、 $M' S'$ も W' の分解結果である。なぜなら、

$$M' S' = (\hat{M} A)(A^{-1} \hat{S}) = \hat{M}(AA^{-1})\hat{S} = \hat{M} \hat{S} = W' \quad (18)$$

となるからである。従って、式(16)で求めた \hat{M} と \hat{S} は、一般には式(11)の M 、 S とは異なり、両者を以下のように関係づける適当な 3×3 行列 A が存在する。

$$M = \hat{M} A \quad (19)$$

$$S = A^{-1} \hat{S} \quad (20)$$

ところが、行列 M の各行はカメラの姿勢を表す単位ベクトル i_f^T, j_f^T の並びであるから、

$$m_f^T m_f = n_f^T n_f = 1 \\ m_f^T n_f = 0 \quad (21)$$

を満たす必要がある。従って、式(21)を拘束条件として式(19)から行列 A を求めることができ、その結果、一意な分解を得ることができる。

2.5 実験結果

因子分解法が安定に動作することを確認するため、実画像を用いた実験を行った。カメラの動きは未知でもよいが、ここでは、精度評価のため高精度のステージで制御した。対象物は建物の模型（高さ約 15 cm）、カメラの焦点距離は 200 mm、対象物までの奥行きは約 3 m である。

カメラの位置姿勢を変えながら 150 枚の画像を取り込み、388 個の特徴点を抽出し追跡した。図 3 に第 1 フレームの画像と抽出した特徴点を示す。

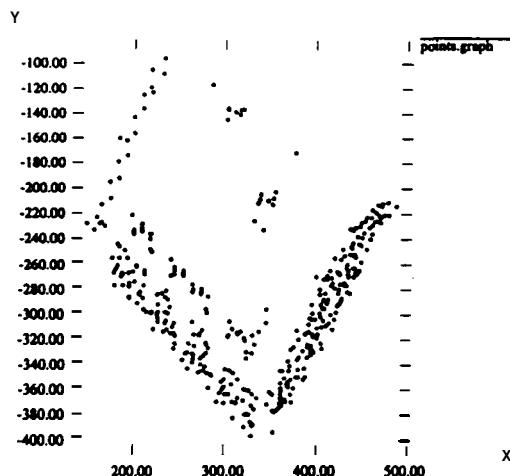


図7 建物の形状復元結果を斜め上から見た様子
Fig. 7 A view of the recovered shape from approximately above the building.

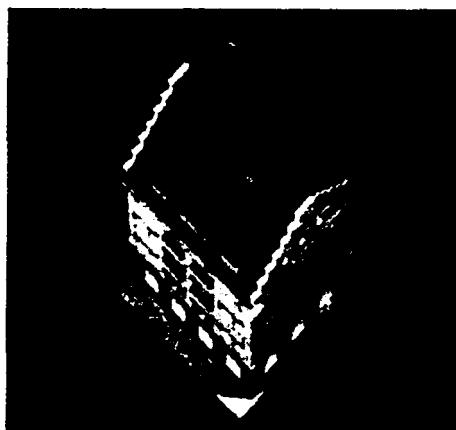


図8 建物を実際に斜め上から見た様子
Fig. 8 A real picture from above the building.

についても安定に動作することを確認した。

3. 投影モデルの拡張

Tomasi と金出は正射影を仮定した因子分解法を示した。しかし、正射影は実際のカメラのモデルとして使われる中心射影の近似として十分とは言えない。例えば、中心射影では、対象が光軸に沿って並進すると投影の大きさが変化する（スケール効果）し、また、画像面に平行に動いた場合には対象物がやや回転したように見える（位置効果）のに対し、正射影では上記二つの効果を表現できない。

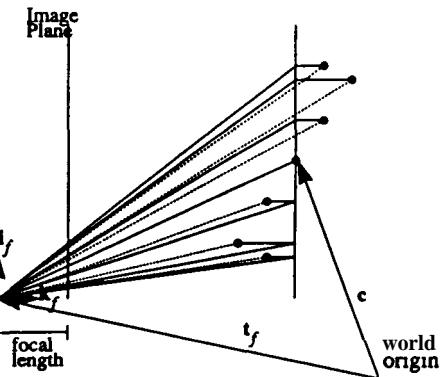


図9 2次元の scaled orthographic projection ● デル
Fig. 9 Scaled orthographic projection in two dimensions. Dotted lines indicate true perspective projection.

ここでは、中心射影により近い近似モデルとして知られる scaled orthographic projection 並びに paraperspective projection を用いた因子分解法⁽¹⁶⁾を示す。前者はスケール効果を考慮に入れ、後者はスケール効果と位置効果の両方を考慮に入れたものである。どちらの場合にも、計測行列 W' のランクはやはり 3 以下であり、復元の基本的枠組みは正射影の場合と全く同様であることを示すことができる。

3.1 Scaled Orthographic Projection

weak perspective としても知られる scaled orthographic projection は図 9 に示すように二つのステップからなる。（1）まず、対象物の重心 c を通り画像面に平行な仮想平面を設け、各点をその平面上に正射影する。（2）次に、その点を画像面上に中心射影する。このとき、投影位置は以下の式で与えられる。

$$x_{sp} = \frac{1}{z_f} i_f^T (s_p - t_f), \quad y_{sp} = \frac{1}{z_f} j_f^T (s_p - t_f) \quad (22)$$

但し、 z_f はカメラ中心から重心までの奥行き

$$z_f = k_f^T (c - t_f) \quad (23)$$

であり、焦点距離は簡単のため 1 とした。式(22)はちょうど式(5)を z_f でスケーリングしたものとなる。ここで、正射影の場合と同じく物体重心に座標原点をとつて $c=0$ なるようにし、式(22)を次のように書き換える。

$$x_{sp} = m_f^T s_p + cx_f, \quad y_{sp} = n_f^T s_p + cy_f \quad (24)$$

但し、

$$z_f = -k_f^T t_f,$$

$$cx_f = -\frac{k_f^T t_f}{z_f}, \quad cy_f = -\frac{j_f^T t_f}{z_f} \quad (25)$$

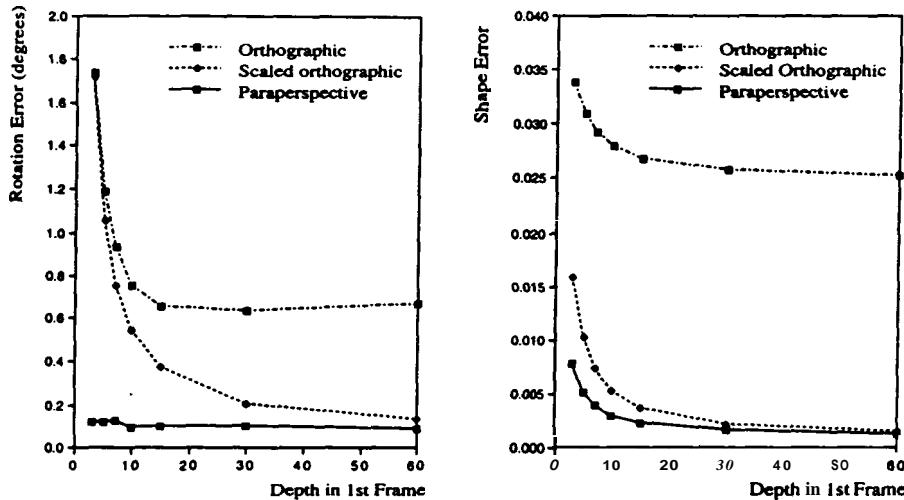


図 11 各投影モデルを用いた因子分解法の比較
Fig. 11 Comparison of the factorization methods under the three projection models.

図 11 に、カメラの回転運動と物体形状の復元誤差を、奥行き(第 1 フレームにおける対象物までの距離を物体の大きさで正規化した値)を横軸として比較した結果を示す。どのモデルを用いた場合でも、奥行きが大きくなるほど中心射影への近似の度合が高くなり、精度が向上する傾向にある。但し正射影モデルは他のモデルに比べて誤差が大きい。paraperspective モデルの精度が最も高く、特にカメラ運動については奥行きが小さくても誤差は極めて小さい。scaled orthographic では、奥行きがある程度大きくなれば paraperspective と同様の精度となる。このように、scaled orthographic あるいは paraperspective モデルに基づく因子分解法を利用すれば、より高精度の復元結果を得ることができる。

4. 重み付け因子分解法

因子分解法では、多数の特徴点の追跡結果を用いて運動と形状を復元する。このとき、特徴点近傍の濃度パターンが不明りょうな場合や、パターンが時間的に変動する特徴点は、追跡結果が不確かであり、誤差も大きいと考えられる。従って、全特徴点の追跡結果を均等に扱うよりは、個々の不確かさに応じた重み付けを行って処理した方が、復元の精度が高い。例えば、局所的濃度パターンの差の 2 乗和を最小化するように動きの対応を求める Tomasi と金出の追跡法⁽¹⁴⁾では、2 乗和の最小値を不確かさとして用いることができる。以下では、特徴点追跡の不確かさに応じた重み付けを行い、復元精度を高める手法を示す。

4.1 列操作による重み付け

各特徴点 p の追跡結果の不確かさを λ_p とし、計測行列 W' の第 p 列を対応する不確かさ λ_p で割る操作を考える。すなわち、 λ_p による対角行列

$$\Lambda = \text{diag}(1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_P) \quad (32)$$

を W' の右から掛けて、

$$W_1 = W' \Lambda \quad (33)$$

なる重みづけられた計測行列 W_1 をつくる。このような列操作は行列の階数を変えないので、 W_1 の階数はやはり 3 以下である。 W_1 を $2F \times 3$ の行列 \tilde{M}_1 と $3 \times P$ の行列 \tilde{S}_1 の積

$$W_1 = \tilde{M}_1 \tilde{S}_1 \quad (34)$$

に分解し、 $\tilde{M} = \tilde{M}_1$ 、 $\tilde{S} = \tilde{S}_1 \Lambda^{-1}$ とおけば、

$$W' = W_1 \Lambda^{-1} = \tilde{M}_1 \tilde{S}_1 \Lambda^{-1} = \tilde{M} \tilde{S} \quad (35)$$

であるから、これまで同様、 W' の分解を得ることができる。各列のノルムをいったん $1/\lambda_p$ に小さくして特異値分解を行うことで、その列の寄与が少なくなり重み付けの効果が生じる。なお、この重み付けはどの投影モデルに対しても適用可能である。

4.2 重み付けの効果

重み付けの効果を確認するため、3.と同様のシミュレーション画像を用いた実験を行った。但しここでは、計測行列 W' の各列ごとに標準偏差の異なるノイズを加え、その標準偏差を不確かさ λ_p と考えて重み付けを行う。フレーム数・特徴点数は 3.と同じだが、カメラの回転

feasibility of motion and structure from noisy time-varying image velocity information", Int. J. Comput. Vision, **5**, 3, pp. 239-269 (1990).

- (10) Broda T. J., Chandrashekhar S. and Chellappa R. : "Recursive 3-D motion estimation from a monocular image sequence", IEEE Trans. AES, **26**, 4, pp. 639-655 (1990).
- (11) Matthies L., Szeliski R. and Kanade T. : "Kalman filter-based algorithms for estimating depth from image sequences", Int. J. Comput. Visiton, **3**, 3, pp. 209-238 (1989).
- (12) Jerian C. P. and Jain R. : "Structure from motion—a critical analysis of methods", IEEE Trans. SMC, **21**, 3, pp. 572-587 (1991).
- (13) Tomasi C. and Kanade T. : "Shape and motion from image streams : a factorization method". Technical Report CMU-CS-91-172, CMU (1991).
- (14) Tomasi C. and Kanade T. : "Shape and motion from image streams : a factorization method—Part 3, Detection and tracking of point features", Technical Report CMU-CS-91-132, CMU (1991).
- (15) Press W. H., Flannery B. P., Teukolsky S. A. and Vetterling W. T. : "Numerical recipes in C : the art of scientific computing", Cambridge University Press (1988).
- (16) Poelman C. J. and Kanade T. : "A paraperspective factorization method for shape and motion recovery". Technical Report CMU-CS-92-208, CMU (1992).
- (17) Ohta Y., Maenobu K. and Sakai T. : "Obtaining surface orientation from texels under perspective projection", 7th IJCAI, pp. 746-751 (1981).

(平成5年5月12日受付)

森田 俊彦



昭59東大・工・産業機械卒。同年(株)富士通研究所に入社。以来、ロボットビジョン、宇宙ロボットの研究に従事。現在、宇宙メカトロニクス研究部に所属。平4年7月より、米国カーネギーメロン大学に客員研究员として滞在中。日本ロボット学会、IEEE会員。

金出 武雄



昭48京大・工・大学院博士課程了。同年同大情報工学科助手。昭51同大助教授。昭55米国カーネギーメロン大学計算機科学科高等研究員。昭60同大教授。平4同大ロボティクス研究所 Director, 工博。日本ロボット学会会員。IEEE Senior Member, AAAI Fellow.

コンラッド ポールマン



1990年米国マサチューセッツ工科大・計算機科学ならびに航空宇宙卒。現在カーネギーメロン大学にて Ph. D. 取得中。計算機科学専攻。コンピュータビジョン、画像理解等の研究に従事。