コンピュータビジョン 99－9
（１９９６．３．２１）

CVCV-WG 特別報告:コンピュータビジョンにおける技術評論と将来展望 (VIII)
運動からの３次元復元 —複数運動の扱いを中心に—

志沢 雅彦

新技術事業団 「さきがけ研究２１」[1]

あらまし

動画像からの３次元復元に関する研究は、従来からＣＶにおける中心的テーマの一つであった。動画像からの３次元復元において、複数物体が行う複数運動の扱いは困難な問題である。従来は、複数物体の扱いは、コンピュータビジョンにおける標準的手法で実現されてきた。本報告では、動画像から複数運動を推定する方法の中で、比較的最近提案された注目すべき新手法に焦点を当ててレビューする。

CVCV-WG reports: Technical Review and Views in Computer Vision (VIII)
—Three-dimensional Structures from Motion: A Focus on Multiple Motions—

Masahiko Shizawa

PRESTO, JRDC, Research Development Corporation of Japan

Abstract

Reconstruction of three-dimensional structure from motion in images has been one of the central themes of computer vision. In this area of research, the multiple-object motions in an image are sources of difficulty of the methods for 3D reconstruction of realistic scenes. The multiple-object motions have been handled by using standard techniques in computer vision. This report focuses on the recently proposed interesting development of theories and algorithms for the multiple-object structure-from-motion problems.

[1]連絡先：東京大学工学部計数工学科 (113 文京区本郷 7-3-1), e-mail: shizawa@simplex.t.u-tokyo.ac.jp

1　はじめに

コンピュータビジョンにおいて複数物体、複数運動、複数属性など「複数対象」を扱わなければならない状況は多い。この問題は「セグメンテーション問題」とも深く関連し、従来から困難視されてきた。コンピュータビジョンの問題を数理解析的に扱う際には、できるだけ単純化された状況を想定するのが普通であるから、「複数」性は、あらかじめ解析から除いて、いわば「単数」化することが多かった。しかし、現実の画像は、そのほとんどが複数物体、複数属性などの複数要因の絡み合いを含む。このことが、コンピュータビジョンの数理解析的アプローチと実際応用との大きなギャップの一因であったと考えられる。つまり、数理解析的な手法で得られるアルゴリズムは、多くの場合に現実の画像へそのまま応用することが困難である。

従来は、複数情報を数理解析的に厳密に扱うための道具がコンピュータビジョンには存在しなかった。Marrの言うところの視覚情報処理の理解のための３つの水準で言えば、常にアルゴリズムとインプリメンテーションのレベルで解決が試みられてきたといえる。複数の異なる属性を持つ領域に画像を分割する問題は、総じてセグメンテーション問題と呼ばれてきた。セグメンテーション問題には、よく知られている様に、「鶏が先か？卵が先か？」というジレンマがある。つまり、それぞれ単一の物体または単一の要因からなる部分集合にデータを分離するプロセスは、各物体あるいは属性の推定結果を必要とする。しかし、各属性の推定が可能であるためには、データは分離できていなければならない。

しかし、近年、数理解析的およびアルゴリズム的側面からもこの「複数」性を直接扱おうとする研究が散見される。これは、より現実に近い状況でのコンピュータビジョンの応用を意識した方向であるとも言えるし、透明視[2]の様な人間の高度で精緻な視覚機構を説明するための計算理論であるともいうことができる。いままでに提案された方法によって、数理解析的アルゴリズムを実際応用に適用することがただちに著しく容易になると結論づけることは現時点では無理である。しかし、原理的限界や原理的困難を明らかにすることも、コンピュータビジョンの研究としての使命である。従来手法と組み合わせれば、制約が少なく、かつ、より信頼性が高い実現を可能にするという意味で、コンピュータビジョンに全体として進歩をもたらすと考えられる。

本稿では、運動からの３次元復元の分野において、複数運動に関連して最近行われた研究を中心に紹介する。

[2]例えば、人間は複数枚の重なりあう表面を知覚することができる。このときもっとも奥の面以外は、透明なシートに見える。この視覚現象を透明視という。

2　複数運動を扱うアルゴリズムの分類

まず、複数運動を扱うために用いられるアルゴリズムの分類を以下に行う。これは、手法の面から分類したものである。これらの手法がそれ単独で用いられることはあまりない。実際のアルゴリズムは、異なる手法を複数組み合わせて構成されたものが多い。

2.1　投票法、ハフ変換

未知パラメータ空間における投票法（あるいはハフ変換）は、もっとも素朴な方法である。しかし、運動からの３次元復元においては、未知パラメータ数が大きいため、素朴な実現方法では投票空間が高次元配列となり、空間計算量が著しく大きくなってしまう。そこで、適応的に投票空間の量子化密度を変化させる方法など、種々の工夫がこらされている。複数運動を扱った研究としてもっとも古く、頻繁に参照されている研究は、Fennema[9] のオプティカルフローのセグメンテーションアルゴリズムである。これは、物体の３次元運動を扱うものではないが、２次元速度空間における投票法を用いて複数運動物体の分離を行った最初の論文である。３次元運動を扱ったものには、Adiv[1] の論文がある。これは、オプティカルフローが与えられたという前提のもとで、異なる３次元物体の占める領域に画像を分割している。パラメトリックなモデルフィッティングと投票法を組み合わせたハイブリッドのアルゴリズムである。

2.2　クラスタリング

この方法の基本的手順は次の通りである。画像から３次元復元に必要な少数のデータを取り出し、３次元復元アルゴリズムを解く。これを、ランダムあるいは系統的に画像から取り出したデータに対して多数回実行する。得られる運動パラメータを運動パラメータ空間でクラスタリングする。クラスタリングには、統計的パターン認識の方法などが応用できる [10] [12] [15]。しかし、透視投影を仮定した運動からの３次元復元においては、パラメータ推定に必要とされるデータの最小数が比較的大きい。したがって、複数運動に対して、取り出したデータから正しい単一運動を推定できる確率は非常に低くなる。そのため、大量のノイズの中でクラスタリングを行う必要がある。しかも、運動からの３次元復元は非線形逆問題であるため、同一物体の運動パラメータの計算結果が、パラメータ空間で局在して分布する保証がない。したがって、手法的にこなれたクラスタリングのアルゴリズムを用いたとしても、安定性を保証することが難しい。３枚

の画像に直交投影された特徴点対応からの３次元復元において、方程式数が運動パラメータ数よりも多いことを利用して、クラスタリングのための剛体性検査を容易にできると主張されている [6] [20] [21] [22]。

2.3　線過程を用いたマルコフ場モデル

Geman & Geman らによって提案された画像のマルコフ場モデルを応用し、画像面上で物体境界の不連続点を線過程としてモデル化すると、アニーリング法またはその変形による超並列アルゴリズムとして実現できる。この方法を動画像からの３次元復元に応用する場合には、オプティカルフローの不連続によって仕切られた画像面上の領域として物体の記述が得られるという仮定が必要である。Murray & Buxton [13] は、Adiv[1] と同様の問題設定に対して、マルコフ場モデルを用いた最適化問題として定式化した。

2.4　単一運動だけに同調するパラメータ推定法

外れ値の影響を受けにくいロバスト統計などの手法を応用し、単一の運動だけに適合し、その他の運動に対応するデータを外れ値と見なしたパラメータ推定を行う [14] [18]。この方法では、どの物体の運動が結果として出力されるのかは、初期値に強く依存する。そのため、複数のロバスト推定法を合成し、多モード情報を扱うことができるM推定法が用いられることもある [8]。

2.5　多レイヤー表現

動画像を複数枚のレイヤーで表現する方法である。それぞれのレイヤーは、不透明性マップ、オプティカルフローまたは３次元運動パラメータ、テクスチャマップなどの属性情報を持つ。各レイヤーは、異なる運動または異なる奥行きを持つ物体表面に対応する。この表現を動画像から求めるために、M推定法や、クラスタリング手法が用いられる [2] [3] [8]。

2.6　画像フレームのずらし差分

一個の物体の運動パラメータが知られていれば、隣接する画像フレーム間でずらし差分を行い、画像から、その物体のパターンを消去できる。この事実を用いると、運動パラメータの推定と画像フレームのずらし差分をくり返し実行して、複数物体の運動を検出できる。連続する３フレームの画像から２個の運動を計算する方法が提案されている [4]。

2.7　「因子分解法」の応用

「因子分解法 (Factorization Method)」は、任意枚数の画像フレーム間に対応付けられた特徴点座標を要素とする行列の特異値分解から、剛体の３次元運動パラメータと３次元構造パラメータを最適に計算する方法である。最近この方法を変形して、複数物体の構造を復元するアルゴリズムが提案されている [5][7]。特に、文献 [7] の方法の興味深い点は、特異値分解は一度だけで、あとは、行列要素の並べ替えによるブロック対角行列への変形によってセグメンテーションが行える点である。このブロック対角行列への変形については、最適化による方法とグラフ理論のアルゴリズムを用いた方法が適用できるとされている。しかし、計算量の見積もりが定かではない。しかも、物体数の決定法などは、まだ研究の余地があるようである。しかし、パラメータ推定のアルゴリズムである特異値分解が１度だけの実行で足りる事実は、従来のアルゴリズムにはない特長である。これについては、第3.1節で概要を説明する。

2.8　線形重ね合わせの原理から導かれるテンソル積結合の応用

複数物体の特徴点対応データをデータ空間の分布と考えると、複数物体のデータは、データ空間で加法的に重畳されたものと考えることができる。つまり線形重ね合わせの原理が成立する。すると、演算子形式の重ね合わせの原理から、複数剛体のための基本拘束方程式を求めることができる。得られる方程式は、単一剛体の拘束方程式に異なる運動パラメータを与えたものをテンソル積で結合し、一つの方程式に合成したものである。この結合された方程式は、複数剛体から生じる特徴点の運動を直接拘束する拘束方程式である。この拘束方程式を用いると特徴点を各物体に分離することなく、複数物体の運動パラメータを求める解析解が得られる。運動パラメータが求まれば、それを用いて特徴点を分離することは容易である。この研究は、データを分離することなく、複数運動に対する解析解を直接得られることを示した点が新しい。従来は、「鶏と卵」の関係は避けられないと考えられていたセグメンテーションとパラメータ推定の手法に新しい視点をもたらした。つまり、物体数が既知ならばセグメンテーションを行わなくても複数組の運動パラメータを求めることが可能である。従来は、手法のレベルで複数の対象が扱われてきた。線形重ね合わせの原理は、それに対して、拘束方程式のレベルで複数対象を直接扱うことができるので、複数運動に関する計算理論としての解析が可能になった [16] [17] [19]。これに関しては3.2節で概要を説明する。

3　文献紹介

3.1　因子分解法による運動からの複数物体の3次元復元

因子分解法は、多フレームの特徴点対応の座標データから、3次元運動の運動パラメータを計算する方法である。最近、この因子分解法を発展させて、複数物体の複数運動を扱うことができる方法が発表された [7]。この「多体因子分解法」の要点を以下に説明する。なお、複数物体の扱いに直接関係の無い部分の詳しい説明は省いた。

3.1.1　「因子分解法」

物体上の特徴点の3次元座標を $\mathbf{p}_\alpha = [X_\alpha, Y_\alpha, Z_\alpha]^T$ $(\alpha = 1, 2, \cdots, N)$ とする。このベクトルの同次座標表示を $\tilde{\mathbf{p}}_\alpha = \left[\mathbf{p}_\alpha^T, 1\right]^T = [X_\alpha, Y_\alpha, Z_\alpha, 1]^T$ とすると、3次元剛体運動 $\mathbf{p} \mapsto \mathbf{p}'$ によって、この同次座標ベクトル $\tilde{\mathbf{p}}_\alpha$ は、次の変換を受ける。

$$\begin{bmatrix} X'_\alpha \\ Y'_\alpha \\ Z'_\alpha \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_\alpha \\ Y_\alpha \\ Z_\alpha \\ 1 \end{bmatrix} \tag{1}$$

ただし、$\mathbf{R} = (R_{ij})(i = 1, 2, 3; j = 1, 2, 3)$ は回転行列であり、$\mathbf{t} = [t_1, t_2, t_3]^T$ は並進ベクトルである。

Z 軸をレンズの光軸として、平行投影を仮定すると、特徴の画像面上での投影位置座標は、(X'_α, Y'_α) である。したがって、3次元復元に用いることができるのは、次の2個の方程式である。

$$\begin{aligned} X'_\alpha &= R_{11}X_\alpha + R_{12}Y_\alpha + R_{13}Z_\alpha + t_1 \\ Y'_\alpha &= R_{21}X_\alpha + R_{22}Y_\alpha + R_{23}Z_\alpha + t_2 \end{aligned} \tag{2}$$

F 枚の画像フレーム間で第 α 特徴点の対応が、

$$(x_{1\alpha}, y_{1\alpha}), (x_{2\alpha}, y_{2\alpha}), \cdots, (x_{F\alpha}, y_{F\alpha}) \tag{3}$$

と与えられていると仮定する。このとき、$2F \times N$ の「測定値行列」 $\mathbf{W}$ を次の形に書く。

$$\mathbf{W} = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \cdots & \cdots & \cdots \\ x_{F1} & \cdots & x_{FN} \\ y_{11} & \cdots & y_{1N} \\ \cdots & \cdots & \cdots \\ y_{F1} & \cdots & y_{FN} \end{bmatrix} \tag{4}$$

「運動行列」$\mathbf{M}$ は、次式で定義する。

$$\mathbf{M} = \begin{bmatrix} R_{11}^1 & R_{12}^1 & R_{13}^1 & t_1^1 \\ \cdots & \cdots & \cdots & \cdots \\ R_{11}^F & R_{12}^F & R_{13}^F & t_1^F \\ R_{21}^1 & R_{22}^1 & R_{23}^1 & t_2^1 \\ \cdots & \cdots & \cdots & \cdots \\ R_{21}^F & R_{22}^F & R_{23}^F & t_2^F \end{bmatrix} \tag{5}$$

各フレームにおける剛体の運動は、原点からの座標変換で表現することにした。右肩の数字は画像のフレーム番号を表す。

「形状行列」$\mathbf{S}$ は、次式で定義する。

$$\mathbf{S} = \begin{bmatrix} X_1 & \cdots & X_N \\ Y_1 & \cdots & Y_N \\ Z_1 & \cdots & Z_N \\ 1 & \cdots & 1 \end{bmatrix} \tag{6}$$

すると、次の基本方程式が成り立つ。

$$\mathbf{W} = \mathbf{MS} \tag{7}$$

行列 $\mathbf{M}$ と行列 $\mathbf{S}$ の階数は、ともに一般には4であるから、行列 W の階数も一般には4である。この事実と、行列 W の特異値分解を用いると、行列 $\mathbf{M}$ と行列 $\mathbf{S}$ を以下に様に求めることができる。

まず、行列 W を特異値分解し、特異値の大きい方から4個だけを取り出す。その部分結果を

$$\mathbf{W} = \mathbf{U\Sigma V}^T \tag{8}$$

と書く。ただし、$\mathbf{\Sigma}$ は、大きい方から4個の特異値を成分とする対角行列 $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3, \sigma_4)$ である。また、$\mathbf{U} \in R^{2F\times 4}$, $\mathbf{V} \in R^{N\times 4}$ である。$\hat{\mathbf{M}} \stackrel{\mathrm{def}}{=} \mathbf{U\Sigma}^{1/2}$, $\hat{\mathbf{S}} \stackrel{\mathrm{def}}{=} \mathbf{\Sigma}^{1/2}\mathbf{V}^T$ とすると、ある非特異行列 $\mathbf{A} \in R^{4\times 4}$ を用いて $\mathbf{M} = \hat{\mathbf{M}}\mathbf{A}$、$\mathbf{S} = \mathbf{A}^{-1}\hat{\mathbf{S}}$ と書ける。行列 $\mathbf{M}$ の成分が回転行列の成分と並進ベクトルの成分で構成されていて、原点を物体の重心にとるという条件と仮定を用いると、行列 $\mathbf{A}$ を決定することができる。この部分の説明は省略する。

3.1.2　多体因子分解法

以下では、複数物体の場合を説明する。一般の形状をした2物体が異なる3次元運動をする2体問題を考える。この場合には、退化の無い一般の場合に、行列 $\mathbf{W}$ の階数は8である。物体1に N_1 点、物体2に N_2 点の特徴点が存在する場合の測定値行列 $\mathbf{W}^*$ を次の形に書く。

$$\mathbf{W}^* = [\mathbf{W}_1|\mathbf{W}_2] \tag{9}$$

ただし、$\mathbf{W}_1, \mathbf{W}_2$ は、それぞれ物体1、物体2の測定値行列とする。$\mathbf{W}_1$ と $\mathbf{W}_2$ は、それぞれ、次の形に特異値分解されると仮定する。

$$\begin{aligned} \mathbf{W}_1 &= \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T = \mathbf{M}_1\mathbf{S}_1 \\ \mathbf{W}_2 &= \mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^T = \mathbf{M}_2\mathbf{S}_2 \end{aligned} \tag{10}$$

$\mathbf{W}^*$ に対応する運動行列、形状行列を次式で定義する。

$$\mathbf{M}^* = [\mathbf{M}_1|\mathbf{M}_2] \tag{11}$$

$$\mathbf{S}^* = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{bmatrix} \tag{12}$$

特異値分解の結果から、行列$\hat{\mathbf{M}}_1 = \mathbf{U}_1\mathbf{\Sigma}_1^{1/2}$, $\hat{\mathbf{S}}_1 = \mathbf{\Sigma}_1^{1/2}\mathbf{V}_1$, $\hat{\mathbf{M}}_2 = \mathbf{U}_2\mathbf{\Sigma}_2^{1/2}$, $\hat{\mathbf{S}}_2 = \mathbf{\Sigma}_2^{1/2}\mathbf{V}_2$ を計算する。非特異行列 $\mathbf{A}_1, \mathbf{A}_2$ を用いて、求めたい運動行列と形状行列は、$\mathbf{M}_1 = \hat{\mathbf{M}}_1\mathbf{A}_1$, $\mathbf{S}_1 = \mathbf{A}_1^{-1}\hat{\mathbf{S}}_1$, $\mathbf{M}_2 = \hat{\mathbf{M}}_2\mathbf{A}_2$, $\mathbf{S}_2 = \mathbf{A}_2^{-1}\hat{\mathbf{S}}_2$ と表現される。このとき、行列 $\mathbf{A}$ に対応する行列は、

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}_1 & 0 \\ 0 & \mathbf{A}_2 \end{bmatrix} \tag{13}$$

である。特徴点の物体への対応が未知である一般の場合には、行列 $\mathbf{A}^*$は、式 (13) の様にブロック対角形ではないので、単一運動の場合の行列 $\mathbf{A}$ に関する拘束を課すことができない。したがって、何等かの方法で測定値のセグメンテーションが必要である。

セグメンテーションには「形状相互作用行列」$\mathbf{Q} = \mathbf{V}\mathbf{V}^T$ を用いる。この行列は、対称行列である。2物体の場合の形状相互作用行列 $\mathbf{Q}^*$ は、次式の形式になる。

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{S}_1^T(\mathbf{S}_1\mathbf{S}_1^T)^{-1}\mathbf{S}_1 & 0 \\ 0 & \mathbf{S}_2^T(\mathbf{S}_2\mathbf{S}_2^T)^{-1}\mathbf{S}_2 \end{bmatrix} \tag{14}$$

この行列は、形状パラメータだけに依存し、運動パラメータに依存しないという著しい特徴がある。式 (9) の様に測定値が物体１、２に分離されていない、つまり、セグメンテーションが未知の一般の場合でも、形状相互作用行列の成分の値は集合としては不変である。したがって、逆に、式 (14) のブロック対角構造にその成分を並びかえることができれば、それがセグメンテーションを行ったことになる。並べ替えでは、行と列を同時に交換する操作を繰り返す。論文では、ブロック対角化を行うために、最適化手法とグラフのアルゴリズムを用いているが詳細は明らかでない。物体数が未知の場合にも、特異値分解は1回の実行で済む。しかし、セグメンテーションが未知の段階で、特異値分解の階数だけから正しく物体数を決定することが容易かどうかは明らかでない。$\mathbf{Q}$ の行と列の交換に対応して、行列 $\mathbf{M}^*, \mathbf{S}^*$ の行と列の入れ換えを行い、運動パラメータと形状パラメータが式 (11),(12) のごとく各物体に分離される。

3.2 線形重ね合わせの原理から導かれるテンソル積結合を用いた複数物体の3次元構造復元

複数物体について、データの線形重ね合わせが可能な形に超関数を用いてデータを記述し、そのデータに作用する演算子を用いた拘束方程式の記述法を提案した。この方法によって、複数の異なるパラメータを持つ拘束方程式をクロネッカーのテンソル積で結合した方程式が得られる。この方法は、フレーム間の特徴点対応からの3次元構造復元と、オプティカルフローからの3次元構造復元のいずれにも適用可能である。ここでは、透視投影下での、２フレーム間の特徴点対応から、一般の３次元構造をもつ物体の場合と、平面物体の場合の３次元構造復元の計算理論を２物体の運動に拡張した例を説明する。この理論では、他の方法とは異なり、まったくセグメンテーションを必要としないアルゴリズムが導かれる。また、複数組の運動パラメータを陽に含む拘束方程式が導かれるため、標準的な線形あるいは非線形パラメータ推定の方法をそのまま用いることができる。なお、本章での変数などの表記は、文献 [11] の表記法を一部参考にしたもので、原論文の表記法とは若干異なる。

3.2.1 一般の３次元構造をもった２運動物体の３次元復元

一般の３次元構造を２枚の透視投影像の特徴点対応情報から計算するアルゴリズムでは、基本行列を用いた方法がよく知られている。これを複数剛体の運動推定に拡張した [16] [17]。カメラのレンズ中心に原点をとり、第１画像フレームの特徴点の３次元座標を $\mathbf{p}_\alpha = (x_\alpha, y_\alpha, f)(\alpha = 1, 2, \cdots, N)$、それに対応する第２画像フレームの特徴点座標を $\mathbf{p}'_\alpha = (x'_\alpha, y'_\alpha, f)(\alpha = 1, 2, \cdots, N)$ とする。fは、焦点距離である。これらの単位ベクトルを、$\mathbf{m}_\alpha = \left(m_{\alpha(1)}, m_{\alpha(2)}, m_{\alpha(3)}\right) = \mathbf{p}/\|\mathbf{p}\|$, $\mathbf{m}'_\alpha = \left(m'_{\alpha(1)}, m'_{\alpha(2)}, m'_{\alpha(3)}\right) = \mathbf{p}'/\|\mathbf{p}'\|$、第１画像フレームにおける３次元物体上の第$\alpha$特徴点と原点との距離を r_α, 第２画像フレームにおける３次元物体上の第α特徴点と原点との距離を r'_αとする。r_α と r'_α は、物体構造パラメータと呼ばれる。このとき、剛体の運動方程式は、

$$r'_\alpha\mathbf{m}'_\alpha = \mathbf{R}(r_\alpha\mathbf{m}_\alpha) + \mathbf{t} \tag{15}$$

である。この方程式において、３本のベクトル $\mathbf{m}'_\alpha, \mathbf{R}\mathbf{m}_\alpha, \mathbf{t}$ が線形従属の関係にある事実を用いると次のエピポーラ方程式が得られる。

$$\left(\mathbf{m}'_\alpha, \mathbf{G}\mathbf{m}_\alpha\right) = 0 \tag{16}$$

$\mathbf{G} = \mathbf{t} \times \mathbf{R}$ は、3×3 行列で「基本行列」と呼ばれる。この基本行列の値は、8 組の特徴点対応情報から、方程式 (16) を用いて定数倍の不定性を除いて推定できる（これは「8点アルゴリズム」と呼ばれる)。$\mathbf{G}$ から運動パラメータ $\mathbf{t}$ と $\mathbf{R}$ を求める公式は数種が知られている。運動パラメータが求まれば、物体構造パラメータを求める公式を用いて３次元物体構造の復元が可能である。

さて、2 物体 A, B の運動の場合には、その基本行列を $\mathbf{G}^{\mathrm{A}}$ と $\mathbf{G}^{\mathrm{B}}$ とすると

$$\left(\mathbf{m}'_\alpha, \mathbf{G}^A\mathbf{m}_\alpha\right)\left(\mathbf{m}'_\alpha, \mathbf{G}^B\mathbf{m}_\alpha\right) = 0 \tag{17}$$

が基本方程式である。この積は、クロネッカーのテンソル積のもっとも単純な場合である。この方程式は、線形

パラメータを用いた次の形に変形できる。

$$\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\sum_{l=1}^{3} G^{AB}_{(ik)(jl)} m'_{\alpha(i)} m_{\alpha(j)} m'_{\alpha(k)} m_{\alpha(l)} = 0 \tag{18}$$

ここで、テンソル $G^{AB}_{(ik)(jl)}$ は、36個の成分をもち、添字 i と kの間、添字 jと lの間でそれぞれ対称性をもつ。このテンソルと基本行列 $\mathbf{G}^{A}$, $\mathbf{G}^{B}$ との関係は次式である。

$$G^{AB}_{(ik)(jl)} = \frac{1}{4}\left(G^{A}_{ij}G^{B}_{kl} + G^{A}_{kj}G^{B}_{il} + G^{A}_{il}G^{B}_{kj} + G^{A}_{kl}G^{B}_{ij}\right) \tag{19}$$

方程式 (18) と正規化条件 $\left\|G^{AB}_{(ik)(jl)}\right\| = 1$ を用いて、36個の線形パラメータ ($G^{AB}_{(ik)(jl)}$) を計算する。35組の特徴点対応があれば、この計算は可能である。36点以上ある場合には、最小2乗法を用いる。こうして得られた線形パラメータから、方程式 (19) を解いて、基本行列 $\mathbf{G}^{A}$ と $\mathbf{G}^{B}$ を求めることができる。但し、この方程式系は、未知数の数よりも方程式数が多い過剰決定方程式系であり、かつ、非線形代数方程式系であるから、その解法には特別の工夫が必要である。文献では、固有値問題を有限回解いて解を得る手順を導いた。また、特徴点数が不足する場合や2個の物体の運動が一致する場合など方程式が退化する場合の意味を実験的に推測した。

3.2.2 平面構造をもった2運動物体の3次元復元

3次元構造の特殊な場合として、平面が3次元運動する場合が詳しく研究されてきた。文献 [19] は、透視投影下での平面の運動を扱う理論とアルゴリズムを2平面が異なる3次元運動をする場合に拡張した。本節では、それを説明する。$\mathbf{x}$ を3次元座標として、平面の方程式を次式に書く。

$$(\mathbf{n}, \mathbf{x}) = 1 \tag{20}$$

ただし、ベクトル $\mathbf{n}$ は、平面の法線ベクトルだが、その大きさは、零以外の任意の値をとれる。3次元剛体運動の方程式 $\mathbf{x}' = \mathbf{R}\mathbf{x} + \mathbf{t}$ から次の方程式を得る。

$$\mathbf{x}' = (\mathbf{R} + \mathbf{t}\mathbf{n}^{T})\mathbf{x} \tag{21}$$

$\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{n}^{T}$とおき、この 3×3 行列を「変換行列」と呼ぶことにする。

方程式 (21) と特徴点に関する前節の表現法を用いると、次の方程式が成立する。

$$r'_{\alpha}\mathbf{m}'_{\alpha} = \mathbf{H} r_{\alpha}\mathbf{m}_{\alpha} \tag{22}$$

この方程式は、3次元ベクトル $\mathbf{m}'_{\alpha}$ と3次元ベクトル $\mathbf{H}\mathbf{m}_{\alpha}$ が平行であることを表している。したがって、それらのベクトル積は零ベクトルである。

$$\mathbf{m}'_{\alpha} \times (\mathbf{H}\mathbf{m}_{\alpha}) = \mathbf{0} \tag{23}$$

これは、3本の方程式を成分として持つが、そのうち独立な成分方程式数は2である。したがって、一般には、4個の特徴点対応があれば、変換行列 $\mathbf{H}$ は定数倍の不定性を持って復元できる。定数倍の不定性をもった $\mathbf{H}$ から3次元運動パラメータ $\mathbf{R}, \mathbf{t}$ と平面の法線ベクトル $\mathbf{n}$ を復元する公式は知られている。

方程式 (23) は、テンソル積結合を用いて異なる3次元運動をする2平面の場合に拡張できる。

$$\left\{\mathbf{m}'_{\alpha} \times \left(\mathbf{H}^{A}\mathbf{m}_{\alpha}\right)\right\} \otimes \left\{\mathbf{m}'_{\alpha} \times \left(\mathbf{H}^{B}\mathbf{m}_{\alpha}\right)\right\} = 0 \tag{24}$$

この方程式は、変換行列 $\mathbf{H}^{A}$ と $\mathbf{H}^{B}$ のテンソル積 $H^{AB}_{ijkl} = H^{A}_{ij}H^{B}_{kl}$ の同次線形形式である。このとき、添字 jと l を入れ替えても方程式が変化しないため H^{AB}_{ijkl} の対称成分 $H^{AB}_{ik(jl)}$ しか求められない。また、添字 i と kの間には対称性は無いが、その対称部分と反対称部分の間で、方程式が完全に独立になるため、1組の同次連立線形方程式系としては同時に解くことができない。常に方程式の階数が1だけ落ちしてしまう。そこで、添字 i と kに関する対称成分と反対称成分に対応する2組の同次連立方程式系を独立に解くアルゴリズムを提案した。

対称成分

$$H^{AB}_{(ik)(jl)} = \frac{1}{4}\left(H^{AB}_{ijkl} + H^{AB}_{kjil} + H^{AB}_{ilkj} + H^{AB}_{klij}\right) \tag{25}$$

は、36個の成分をもち、一般に12組の特徴点対応があれば定数倍の不定性を持って計算できる。但し、少なくとも4組の特徴点対応がそれぞれの平面上になければならない。

反対称成分

$$H^{AB}_{[ik](jl)} = \frac{1}{4}\left(H^{AB}_{ijkl} - H^{AB}_{kjil} + H^{AB}_{ilkj} - H^{AB}_{klij}\right) \tag{26}$$

は、18個の成分をもち、一般に17組の特徴点対応があれば定数倍の不定性を持って計算できる。

対称成分を表すテンソル $H^{AB}_{(ik)(jl)}$は、一般の3次元構造の復元の場合に現われたテンソル $G^{AB}_{(ik)(jl)}$と形式的には同一であるから、この場合と同じアルゴリズムを用いて $H^{AB}_{(ik)(jl)}$から変換行列 $\mathbf{H}^{A}$と $\mathbf{H}^{B}$ を求めることができる。

しかし、文献 [19] では、対称成分と反対称成分を両方用いたより単純なアルゴリズムを提案した。さらに、各平面上の特徴点数が不足する場合と、運動パラメータが一致する場合に方程式の階数がどのように低下するかを実験的および理論的に検討した。

4 まとめ

複数の3次元物体の複数運動を動画像から推定する方法を紹介した。特に、最近提案された新しいアプローチに

重点をおいた。新しい方法は、いずれも「鶏と卵」の関係にあると言われてきたセグメンテーションとパラメータ推定の間の関係に新しい光を当てる結果であると言える。今後は、計算量とノイズに対する振る舞いなどの解析を通じて、手法としての特性と限界を見極めることが必要である。退化判定とモデル選択の手法を応用して、物体数の系統的な判定法を与えることが課題として残されている。

また、これらの研究は、コンピュータビジョンの分野で見出されたものであるが、そこで提案された数理的手法は、コンピュータビジョンの枠をこえて、ある種の非線形逆問題に適用できる一般性を持っていることを最後に指摘しておきたい。コンピュータビジョンの応用を探るだけでなく、コンピュータビジョンの研究で開発されたいろいろな数理技法を他の分野に応用可能な形に一般化してゆくことも重要と考える。

謝辞

初稿に関してコメント頂いたＣＶＣＶのメンバーの方々に感謝します。

参考文献

[1] Adiv, G: "Determining three-dimensional motion and structure from optical flow generated by several moving objects", *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-7, 4, pp.384-401 (1985).

[2] Wang, J.Y.A. and Adelson, E.H.: "Layered representation for motion analysis", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, pp.361-366 (June, 1993).

[3] Adelson, E.H.: "Layered representations for Vision and Video", *Proc. IEEE Workshop on Representation of Visual Scenes*, Mass., MA, pp.3-9 (June, 1995).

[4] Bergen, J.R., Burt, P., Hingorani, R. and Peleg, S.: "A three-frame algorithm for estimating two-component image motion", *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-14, 9, pp.886-896 (1992).

[5] Boult, T.E. and Brown, L.G.: "Factorization-based segmentation of motions", *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, pp.179-186 (October, 1991).

[6] Chen, H.H. and Huang, T.S.: "Maximal matching of 3-D points for multiple-object motion estimation ", *Pattern Recognition*, vol.21, 2, pp.75-90 (1988).

[7] Costeira, J. and Kanade, T.: "A multi-body factorization method for motion analysis", *Proc. IEEE Fifth International Conference on Computer Vision*, Boston, MA, pp.1071-1076 (1995).

[8] Darrel, T. and Pentland, A.: "Robust estimation of a multi-layered motion representation", *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, pp.173-178 (October, 1991).

[9] Fennema, C.L. and Thompson, W.: "Velocity determination in scenes containing several moving objects", *Computer Graphics and Image Processing*, vol.9, pp.301-315 (1979).

[10] Gear, C.W.: "Feature grouping in moving objects", *Proc. Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, Texas (Novermber, 1994).

[11] Kanatani, K.: "Geometric Computation for Machine Vision", Oxford University Press, Oxford (1993).

[12] Liou, S.-P. and Jain, R.C.: "An approach to three-dimensional image segmentation", *CVGIP: Image Understanding*, vol.53, 3, pp.237-252 (1991).

[13] Murray, D.W. and Buxton, B.F.: "Scene segmentation from visual motion using global optimization", *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-9, 2, pp.220-228 (1987).

[14] Peleg, S. and Rom, H.: "Motion based segmentation", *Proc. International Conference on Pattern Recognition*, vol.1, Atlantic City, NJ, pp.109-113 (June, 1990).

[15] Schunck, B.G.: "Image flow segmentation and estimation by constraint line clustering", *IEEE Trans. Pattern Analysis and Machine Intelligence*, PAMI-11, 10, pp.1010-1027 (1989).

[16] Shizawa, M.: "Transparent 3D motions and structures from point correspondences in two frames: a quasi-optimal, closed-form, linear algorithm and degeneracy analysis", *Proc. First Asian Conference on Computer Vision*, Osaka, Japan, pp.329-334 (Nov.,1993).

[17] 志沢雅彦："運動立体視におけるトランスペアレンシー：２透視投影像における点対応からの準最適線形

アルゴリズム", 電子情報通信学会論文誌 D-II, **J77-D-II**, 2, pp.286-300 (1994.2).

[18] Soatto, S. and Perona, P.: "Three dimensional transparent structure segmentation and multiple 3D motion estimation from monocular perspective image sequences", *IEEE Workshop on Motion of Non-Rigid and Articulated Objects,* pp.228-235 (1994).

[19] Sugimoto, A. and Shizawa, M.: "Two-plane structures and Motions from Point Correspondences in Two Images", ATR Technical Report TR-H-155, ATR Human Information Processing Research Laboratories (1995).

[20] Tsukune, H. and Aggarwal, J.K.: "Analyzing Orthographic Projection of Multiple 3D Velocity Vector Fields in Optical Flow", *Computer Vision, Graphics and Image Processing,* vol.42, pp.157-191 (1988).

[21] Ullman, S.: "The interpretation of structure from motion", *Proc. Royal Society of London,* vol. B203, pp.405-426 (1979).

[22] Ullman, S.: **The Interpretation of Visual Motion**, MIT Press, Cambridge MA (1979).