

## 連載講座



## コンピュータビジョンのための幾何学

## 3. 動画像の扱い†

出口 光一郎††

## 1. はじめに

動画像には、対象が運動している場合もあるし、カメラが運動しながら画像を得ている場合もある。どちらの場合も、動画像は、対象の形状について豊富な情報を与えてくれる。今回は、動画像の解析の基になる幾何学について考察しよう。

一番最初の回で述べたように、たとえば対象がある運動をしたとき画像はどのように変化するか、また、カメラがある動きをしたとき画像はどのように変化するかを記述しようとする、とたんに複雑な式となってしまう、扱いが難しくなる。ただし、その結果は複雑であるが、それを得るまでのステップを順を追ってきちんと押さえていくことが肝心である。そうすることで、複雑な式も見通しのよい解釈ができる。

まず、動画像から対象の3次元形状が復元できる。この基本的な原理は、やはり前回述べたステレオと同じである<sup>1)2)</sup>。図-1にそのしかけを示す。

図-1(a)はステレオによる立体形状の認識である。もし対象が静止しているとしたら、両眼で一度に画像を得る必要はなく、(b)のように片眼を動かせばよい。すると、(c)のように眼は静止していて、相対的に対象が動いても同じことである。ただし、正確には、このときは(d)のように、対象の大きさ、距離、速度についてのあいまいさが残ってしまう(スケール(目盛り)が決められない)。この1つの自由度が残ってしまうことは、以後の様々な場面でその意味付けができるし、一方で、少し注意がいることでもある。

ここで重要なことは、上記の対象についてのスケールの自由度が1つ残ってしまうが、その条件付きで、カメラ、または、対象の動きが画像の動きだけから復元しようということである。また同時に対象の形状も、同じ自由度を残して復元できる。

さて、動画像の解析の原理は図-1であるとしたが、この図ではステレオの場合と同様に、基本的には2枚の画像の間での対応付けに基礎を置いている。当然、その意味で解析の原理はステレオと同一である。ただし、動画像として画像の系列が与えられていると、対象の運動の追跡が可能になり、2枚の画像だけが与えられた場合より対応点の抽出は容易である。

さらに、このときのもう1つの特徴としては、画像内で各点がどのように動いているかを、各点の速度ベクトル場として与えられる場合があることである。この場合は、ある時点で、画像内の各点が、画像上でどの方向にどのくらいの大きさの速度を持って移動しているかが与えられる。上記の第一の場合の2枚の画像を得た時間間隔を十分小さくして、いわば瞬間の運動をとらえていることに相当する。画像上でその濃淡値が流体のように流れると見なしている。その流速の場をオプティカルフロー (Optical flow) という。

この2つのどちらに注目するかのアプローチでの大きな違いは、まず第一に、画像から動きの情報を検出するとき何をもとにするか(画像からの動きの抽出のしかた)と、もう1つは、動きを表現するパラメータの与え方(特に、回転の表現のしかた)の2つである。画像から動きが抽出できれば、上で述べた対象の3次元形状の復元の基本的しかけは同一である。

まず、それらを対比しながら、続いてそれぞれの特有の幾何学について述べていこう。

† Image Motion Analysis by Koichiro DEGUCHI (University of Tokyo, Faculty of Engineering, Department of Mathematical Engineering and Information Physics).

†† 東京大学工学部計数工学科

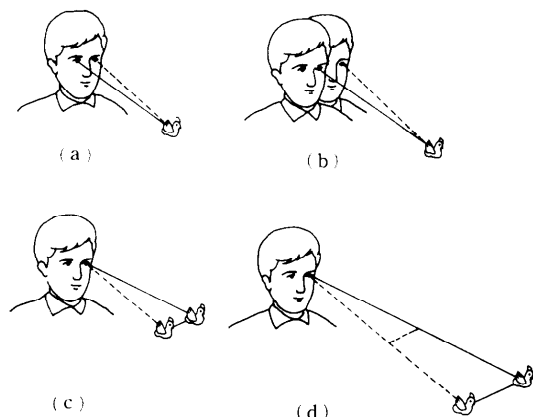


図-1 対象の動きから形状が分かるしかけ。(a)両眼によるステレオ、(b)一眼でも眼を動かせば同じこと、(c)対象の方が動いても同じこと、(d)ただし、2倍の距離で2倍の大きさのものが2倍の速さで動いたときと区別がつかない。

## 2. 空間の動きと画像の動き

### 2.1 対象の動きの表現

動画像には、運動している対象を静止したカメラでとらえた場合と、静止している対象をカメラを動かしてとらえた場合と、さらに、両方が動いている場合もありうる。このとき、対象はその形を変えないものとする(剛体 (Solid body) という)。雲や生物や人の顔など、刻々形を変える対象の動画像の研究も盛んになりつつあるが、まだ、一般的に扱える手法が定着するには至っていない。

まず、基本的な原理を述べよう。

今回は、図-2 に示すように、前方を移動する対象の物体を固定したカメラでとらえる場合を扱う。カメラでは、物体の移動前、移動後のそれぞれの画像が形成される。

ここで、両画像での対応する点は検出できるものとする。ある一点に注目して、この場合の画像化を幾何学的に示すと、図-3 のようになる。すなわち、カメラ中心を原点としたカメラ座標系で、3次元空間中の点  $\mathbf{x}=(x, y, z)^T$  が  $\mathbf{x}'=(x', y', z')^T$  に移動したとする。この移動を、原点のまわりの回転を表す行列  $R$  と平行移動ベクトル  $t$  で表すと、

$$\mathbf{x}' = R\mathbf{x} + t \quad (1)$$

前回のステレオでの扱いは、対象が静止していてカメラが移動した場合にあたる。この両者は相

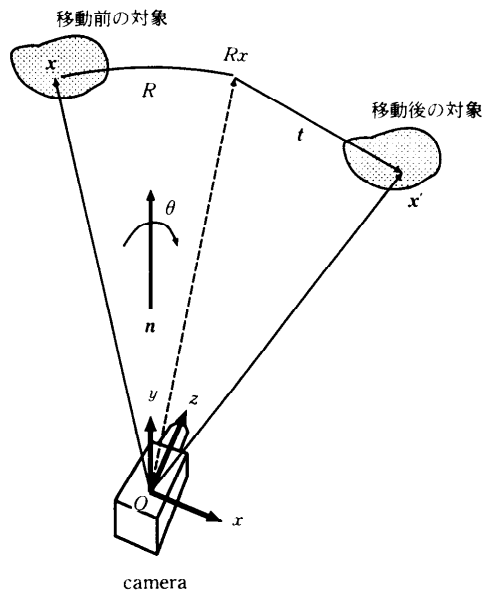


図-2 対象物体の運動とそのパラメータ

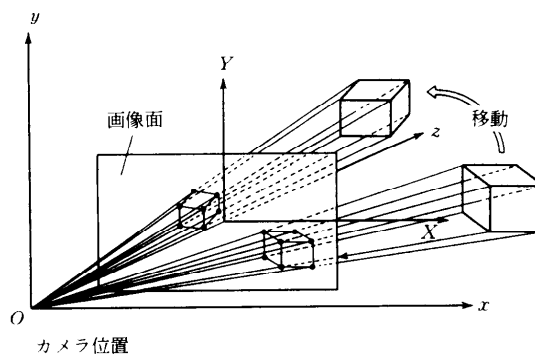


図-3 対象の運動とその画像

対的なもので、よく整理して考えれば明らかであるが、混同されている場合がよくあるので注意が必要である\*。

第1回で述べたように、この回転行列  $R$  というのは、具体的にはいろいろな表し方がある。いずれにしても3つの独立なパラメータで表現できる。すなわち、各成分は独立ではない。これは、 $R$  の成分は9個あっても、回転を表す行列は正規直交であり、 $RR^T = I$  という性質から6個の制約条件が課せられ、結局自由度は3であること

\* ステレオの場合のように、あるカメラ座標での3次元空間中の点  $\mathbf{x}=(x, y, z)$  が、カメラが移動して、新しいカメラ座標で  $\mathbf{x}'=(x', y', z')$  に移動したとする。この両座標の関係が(1)であるとき、座標系の回転は  $R$ 、平行移動は  $-R^{-1}t$  である。このときに、カメラが静止して対象が運動したと見なせば、対象の回転は  $R$ 、平行移動は  $-t$  である。

からも説明される。

この回転行列の具体的な表記法は、第1回で示したもののほかに、たとえば、 $x$ 軸の回りに角度 $\alpha$ だけ対象を回転し、その新しい位置の対象を $y$ 軸の回りに $\beta$ 、さらに、 $z$ 軸の回りに $\gamma$ だけ回転するものとする（この順序は重要である）、

$$R = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix}$$

$$= \begin{bmatrix} \cos \beta \cos \gamma & -\cos \alpha \sin \gamma + \sin \alpha \sin \beta \cos \gamma \\ \cos \beta \cos \gamma & \cos \alpha \cos \gamma + \sin \alpha \sin \beta \sin \gamma \\ -\sin \beta & \sin \alpha \cos \beta \\ \sin \alpha \sin \gamma + \cos \alpha \sin \beta \cos \gamma \\ \sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma \\ \cos \alpha \cos \beta \end{bmatrix} \quad (3)$$

もう1つの表し方は、方向余弦が $\mathbf{n}=(n_1, n_2, n_3)$  ( $n_1^2+n_2^2+n_3^2=1$ )である軸の回りに角度 $\theta$ だけ回転するとして、

$$R = \begin{bmatrix} n_1^2 + (1-n_1^2)\cos \theta & n_1 n_2(1-\cos \theta) + n_3 \sin \theta & n_1 n_3(1-\cos \theta) - n_2 \sin \theta \\ n_1 n_2(1-\cos \theta) - n_3 \sin \theta & n_2^2 + (1-n_2^2)\cos \theta & n_2 n_3(1-\cos \theta) + n_1 \sin \theta \\ n_1 n_3(1-\cos \theta) + n_2 \sin \theta & n_2 n_3(1-\cos \theta) - n_1 \sin \theta & n_3^2 + (1-n_3^2)\cos \theta \end{bmatrix} \quad (4)$$

(3)の表記は、その成り立ちのように、回転を各座標軸の回りに次々に回転していくときには分かりやすいが、逆に、 $R$ が与えられて、それからその由来である各軸回りの回転角 $\alpha, \beta, \gamma$ を導き出すことは容易ではない。むしろこの表記は、各回転角 $\alpha, \beta, \gamma$ が微小であるとき、すなわち、 $\cos \alpha=1, \sin \alpha=\alpha$ などに見なせる場合、

$$R \approx \begin{bmatrix} 1 & -\gamma & \beta \\ \gamma & 1 & -\alpha \\ -\beta & \alpha & 1 \end{bmatrix} \quad (5)$$

となり、後のオプティカルフローの表現にとって重要になる。

一方、(4)の表記で用いた回転のパラメータについては、 $R$ から、具体的な値を導くことが容

易である。

まず、回転の軸方向を表すベクトル $\mathbf{n}$ については、この方向の点は $R$ によって回転しても動かない。したがって、

$$R\mathbf{n}=\mathbf{n} \quad (6)$$

これは、 $\mathbf{n}$ が行列 $R$ の固有値1に対する固有ベクトルであることを表し、標準的な計算手法で $\mathbf{n}$ は求まる。また、 $n_1^2+n_2^2+n_3^2=1$ であることから、行列 $R$ の対角要素の和は、

$$R_{11}+R_{22}+R_{33}=1+2\cos \theta \quad (7)$$

である。

以上の議論で、対象の運動を表す変数は、 $R$ の自由度は今示したように3で、平行移動ベクトル $\mathbf{t}$ が3個。したがって、空間を移動する自由度は6である。また、先に述べたようにスケールの分の1自由度が残るので、運動のパラメータの自由度は、結局、5である。実際、 $\mathbf{t}$ については、その方向だけが決まり、空間での絶対的な移動量の大きさは決定できないということである。そこで、以下、 $\mathbf{t}$ を単位ベクトル、すなわち、

$$\|\mathbf{t}\|=(t_x^2+t_y^2+t_z^2)^{1/2}=1 \quad (8)$$

とする。

## 2.2 画像上の対象の動き

さて、動画像から対象の3次元形状を復元する問題は、前回のステレオと基本的に同じ枠組みに落ちることを述べた。ただし、ステレオでは、特性の異なる2つのカメラを前提としたのに対して、今度は、1つのカメラでの時間の異なる2つの画像を対象とする。ステレオの言葉で言えば、まったく同じ内部パラメータを持つ2つのカメラを用いていることとほぼ等価である。

しかし、今度は対象形状の復元だけでなく、動き(回転行列 $R$ と、平行移動ベクトル $\mathbf{t}$ )の復元にも興味がある。実際、動きが復元できれば、形状の復元はステレオと同じ問題となる。ただし、動きも含めて復元するには、カメラの内部パラメータを外部パラメータ(カメラの位置、姿勢)から切り離して正確に校正する必要がある。

すなわち、第1回で述べたように、空間の点の座標(齊次座標)とその画像座標(齊次座標)とは透視投影行列によって結び付けられ、その透視投影行列は、内部パラメータを表す行列と外部パラメータを表す行列の積であった。そして、ステレオの場合の対象形状の復元では、この透視投影

行列が分かれば十分であったが、画像座標に基づく動きの復元には、この2つを分離する必要がある。カメラの内部パラメータを正確に求め、画像と対象の関係をカメラ座標で表現する必要がある。

さて、図-2に示したように、対象上の点がカメラ座標で(1)と表される運動をするとき、3つのベクトル  $t, Rx, x'$  が一平面上に乗る。これが、以後の運動パラメータを求めるための基礎となる。

この3つのベクトルが一平面上にあるときは、 $\wedge$  が外積を表すとして、

$$(t \wedge Rx) \perp x', \quad (9)$$

であり(3つのベクトルはどの順序でもよい)、したがって、

$$((t \wedge Rx), x') = 0 \quad (10)$$

とも表される。

画像面が  $z=1$  であるとしたとき、対応点それぞれの画像上の座標は、

$$\begin{cases} X = \frac{x}{z} \\ Y = \frac{y}{z} \end{cases}, \quad \begin{cases} X' = \frac{x'}{z'} \\ Y' = \frac{y'}{z'} \end{cases} \quad (11)$$

そこで、(10)の両辺を  $zz'$  で割れば、 $X=(X, Y, 1)'$ 、 $X'=(X', Y', 1)'$  として、

$$((t \wedge RX), X') = 0 \quad (12)$$

を得る。

前回、同じ対象を違う視点から見た画像上の対応点の対の画像座標  $m, m'$  の間には、

$$m' F m = 0 \quad (13)$$

という関係があることを述べた。ここで、カメラの内部パラメータを表す行列を  $A$  とすれば、

$$F = (A^{-1})' T R A^{-1} \quad (14)$$

であり、この中の行列  $T$  は、 $t = (t_1, t_2, t_3)'$  に対して、

$$T = \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix} \quad (15)$$

である。

画像座標(画像平面上にとられた2次元の座標)  $m$  に対して、 $X = A^{-1} m$  は、まさに、この点のカメラ座標による表記で(第1回)、(11)で与えられる。

すなわち、

$$E = TR \quad (16)$$

とおくと、(13)は、対応点の画像のそれぞれのカメラ座標での表記  $X, X'$  に対して、

$$X'^T E X = 0 \quad (17)$$

を得る。ここで、行列  $E$  は、(16)のように(歪み対象行列)  $\times$  (回転行列)と分解でき、 $\text{rank}(E)=2$  であることに注意が必要である。

一方、行列  $T$  は、 $t \wedge X = TX$  を与えるものであった。したがって、(17)と(12)は、まったく同等の式である。

### 2.3 動きによる形状計測の線形解法

カメラの前で対象が運動するとき、その前後での画像の間には、(17)、または、(12)の関係が成り立つことを示した。すると、問題は、「2つの画像上の  $n$  対の点の組  $(X_i, X'_i)$ 、 $(i=1, 2, \dots, n)$  から、そのすべての点に対して(17) (または、(12))を満たすような運動パラメータ  $R$  と  $t$  を決定せよ。」

そして、このときの制約条件として、「行列  $R$  は回転行列であるので、正規直交、すなわち  $RR' = I$  である。」

が付加される。

この問題に対して、Longuet-Higgins<sup>9)</sup>によって、また、Tsai, Huangらによって、簡単に明快な解法(より完全な形では文献4))がほぼ同時に示された。それらは、本質的には同一であるので、以下文献4)に沿ってその解法のあらましを述べ、それにまつわる問題を述べる。

式(17)を  $E$  の要素について方程式と見なせば、8組以上の対応点の座標の組  $X_i = (X_i, Y_i)$ 、 $X'_i = (X'_i, Y'_i)$  ( $i=1, \dots, n$ ) が与えられれば、

$$B = \begin{bmatrix} X_1 X'_1 & X_1 Y'_1 & X_1 & Y_1 X'_1 & Y_1 Y'_1 & Y_1 & X'_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n X'_n & X_n Y'_n & X_n & Y_n X'_n & Y_n Y'_n & Y_n & X'_n \\ Y'_1 & 1 \\ \vdots & \vdots \\ Y'_n & 1 \end{bmatrix} \quad (18)$$

とおいて、

$$Be = 0 \quad (19)$$

ただし、 $e$  は、 $E$  の要素を並べたベクトルで、

$$e = (E_{11}, E_{12}, E_{13}, E_{21}, E_{22}, E_{23}, E_{31}, E_{32}, E_{33})' \quad (20)$$

この方程式は、前回、基礎行列を2枚の画像の対

応点の組から求めたのとまったく同様にして解くことができる。

すなわち、画像の読み取りノイズなどにより(19)を正確に満たす解が存在しない場合にも対応するため、

$$\|Be\| \rightarrow \text{最小} \quad (21)$$

となる  $e$  をもって、(19)の解とする。このような解は、(19)を満たす  $e$  が存在するときも含め、 $B^T B$  の最小固有値に対応する固有ベクトルとして与えられる(本来は、固有値0に対する固有ベクトル)。この計算は、行列  $B$  の特異値分解を用いることで、容易にできる。ただし、この固有ベクトルの大きさはこれだけでは決まらないが、(8)の  $\|t\|=1$  に対応させるためには、

$$\begin{aligned} \|e\|^2 &= \sum_{i,j} E_{ij}^2 = \text{trace}(EE^T) \\ &= \text{trace}(TT^T) = 2\|t\|^2 \end{aligned} \quad (22)$$

より、 $\|e\| = \sqrt{2}$  とする。

こうして求めた  $E$  を、続いて、 $R$  と  $t$  とに分解する必要がある。しかし、この  $E$  は、(16)のように(歪み対象行列)×(回転行列)と分解できるとは限らない。行列  $E$  にとって、一番重要な  $\text{rank}(E)=2$  という性質について何も配慮していなかったからである。そこで、文献5)では、 $\hat{E}=TR$  と分解できる、 $E$  に一番近い行列、すなわち、行列のノルム<sup>\*\*</sup>  $\|E-\hat{E}\|$  を最小とする行列  $\hat{E}$  を求めることで分解を行っている。

それには、 $E$  をさらに特異値分解し、

$$E = SDQ^T \quad (23)$$

ただし、 $S$ 、 $Q$  は正規直交行列、 $D$  は対角行列で、その3つの対角成分が  $d_1 \geq d_2 \geq d_3 \geq 0$  となるように、 $S$ 、 $Q$  は選ばれる。このときの  $S = [s_1 \ s_2 \ s_3]$ 、 $Q = [q_1 \ q_2 \ q_3]$  とすれば、求める  $R$  と  $t$  は、次のように2組与えられる。

$$R = \mp s_2 q_1 \pm s_1 q_2 + s_3 q_3 \quad (\text{複合同順})$$

このうち、対象への奥行きが正となる方の解に絞る。

この解法(8点アルゴリズムと呼ばれる)は、単純で標準の線形計算ライブラリを用いて、高速に実行できる。また、 $n \geq 8$  のとき、与えられた  $n$  個の対象点が動作前後の2つの相対的なカメラ位置とともに、ある特別な配置をしていなければ

(*surface condition* と呼ばれる)、解が唯一に与えられることも分かっている<sup>6)</sup>。

#### 2.4 ノイズと制約条件下での解法

この線形アルゴリズムをそのまま用いたのでは、小さなノイズに対しても、思いがけないほどの大きな誤差を与えることがあると指摘されている。

まず、数値計算上の不安定さを取り除いておく必要がある。画像座標の組として、(100, 100) というオーダの値を用いると、(18)での  $B$  は、 $10^4$  のオーダの値と1とを含み、さらには、 $B^T B$  の成分の間には、 $10^8$  倍の開きが生じる。画像座標をうまくスケールリングすることで、誤差の影響を小さくできる場合があることが文献7)で述べられている。

さらに、本質的な問題が、Phillip<sup>8)</sup>によって指摘されている。

このアルゴリズムは、画像の読み取りノイズがなければ、(歪み対象行列)×(回転行列)と分解できる中間行列  $E$  が得られるという関係が基本となっており、ノイズの存在下では、先に述べたように第1段階で最適化された  $E$  はそのように分解できるとは限らない。第2段階の最適化では、その  $E$  に「フロベニウスノルム」の意味で一番近い、(歪み対象行列)×(回転行列)に分解できる中間行列  $\hat{E}$  を求め、 $\hat{E}=TR$  としている。しかし、全体を通して見たとき、この2段階の最適化は意味がない。

その結果、いくら点の数  $n$  を大きくしても、推定される運動パラメータの平均とその真の値との間に一定の偏差が残ることが知られている。ただし、その偏差はある程度見積もれるので、そこで、推定量の偏りを取り除きながらこの線形解法にそって運動パラメータを求めていく手法が提案されている<sup>9),10)</sup>。

ここで、基本になっているのは、すべての対応点の画像座標について、(17)または(12)が満たされているはずだという考え方である。これに対するまず第一の直観的な解法は、

$$J = \sum_{i=1}^n ((RX_i) \wedge t, X_i)^2 \quad (25)$$

を、 $R$  が回転行列で  $\|t\|=1$  という制約のもとで最小化するものである。ただし、これらの制約は評価式を高次の非線形式としてしまい、実際には

\*\* ここでの  $\|\cdot\|$  は、内部の行列の各成分の2乗和の平方根を表す。行列のフロベニウス・ノルムと呼ばれる。

反復法によらねばならず、適切な初期値を必要とする。文献11)では、 $R$  および  $t$  への制約のもとで(25)式を直接に最小化する問題へもう一度立ち返った上で、評価関数を変形して計算の手間を減らす研究が報告されている。

また、このときたとえば「四元数」を用いることで、定式化が簡潔になる可能性が指摘されていて<sup>12)</sup>、いくつかの解法が提案されている。これらによれば、ノイズの存在下では、線形アルゴリズムに比べて、直接(25)式を最小化することによってはるかに良い結果が得られてはいる。しかし、依然、それでも十分な精度は得られていないこともあわせて報告されている。

### 2.5 何を最小化すべきか

最初に述べたように、ノイズのない理想的な場合は、解として得られた回転行列  $R$  によって、3つのベクトル  $t$ ,  $Rx$  および  $x'$  は一平面上に乗る。画像の読み取りにノイズが混入すると、これらのベクトルが一平面上にあるようにする  $R$ ,  $t$  が得られない。前節での評価(25)を最小化すると、すべての対応点に対してそれぞれでこの3つのベクトルを「なるべく一平面上にある」ようにしようというつもりであった。しかしながら、このような意図に対して、評価「 $\sum((t \wedge Rx), X')$ 」が適切であるかは明らかではない。

$((t \wedge Rx), X')$  は、3つのベクトル  $t$ ,  $Rx$ ,  $X'$  のベクトル3重積と呼ばれ、この絶対値はこれらの3つのベクトルを隣り合う稜とする平行6面体の体積を表す(図-4)。3つのベクトルが一平面上にあれば、この平行6面体の体積は0である。しかし、この体積を最も小さくするのは3つのベクトルが一番平面的(?)になったときとは、単純には言えない。このことが、先の偏りを生む要因になっている。

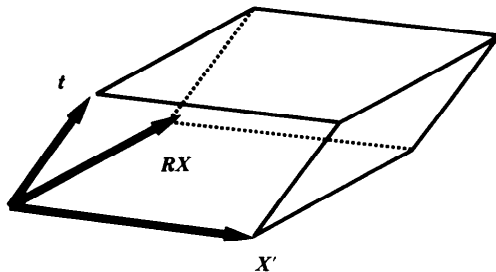


図-4 ベクトル  $t$ ,  $Rx$ ,  $X'$  で張られる平行6面体。この体積が  $|((t \wedge Rx), X')|$  である。

偏りがない推定量にはいろいろあるが、その代表的なものが最尤推定量である。このような画像からの運動の推定に最尤推定を導入したのは Adiv<sup>13)</sup>である。運動パラメータの最尤推定量は、観測された画像上の動きと、推定したパラメータによって復元した画像の理論的な動きとの差を最小にするものとして与えられるが、その計算には対象までの真の距離を必要とする上、非線型評価関数の最小化問題に帰着されるので、反復法や全数探索が必要になる。

また、まったく違った立場として、結果的に誤差を大きくするような点を計算から除外するという手法によって、運動パラメータの推定の精度を上げるアプローチも報告されている<sup>14)</sup>。ステレオの場合と同様に、1点でも誤差の比較的大きな点が混入すると、その影響を克服するには膨大な数の精度の良い点を必要とする。そこで、そのような点を  $M$  推定と呼ばれるロバスト推定の手法で判定して計算から除外する。

以上のように、この問題は完全にはまだ解決されていないと言えるが、はるかに簡単な計算で運動パラメータを得る線形アルゴリズムの貢献は大きい。現在、様々な改良が試みられている。

## 3. オプティカルフロー

### 3.1 オプティカルフローからの形状復元

最初に述べたように、対象が空間で移動するか、または、カメラが移動すると、画像は画像面上で流れを生じる。前章までの移動前後の2枚の画像のその撮影の時間間隔を微小にしていって、そのときの各対応点間を結んだベクトルを時間間隔で割ると、画像上の各点の速度場が図-5のようにできる。この速度場をオプティカルフローと呼んでいる。

オプティカルフローからも、基本的には同じ原理で、対象の運動と3次元形状が復元できる。ただし、前回のステレオで述べたように、画像間の視線の差が小さいとその形状復元の精度は落ちる。一方で、画像の動きが小さいこと、多数の連続した画像系列を利用できることなどから、対応点の決定は容易になり、画像上の多くの点で速度ベクトルを得ることができ、3次元形状の復元には、この両者でうまくバランスをとることが必要である。

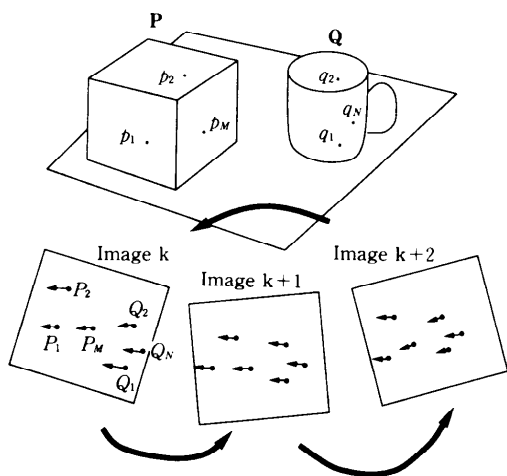


図-5 対象の動きとオプティカルフロー

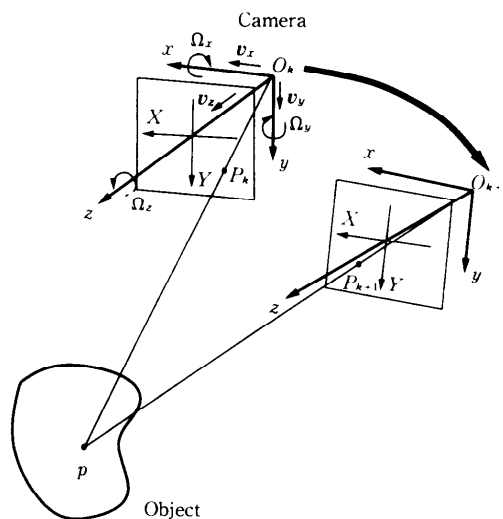


図-6 オプティカルフローの表現のための運動パラメータ

画像からのオプティカルフローの検出法については、多くの研究がある。これは、3次元形状の復元だけでなく、動画像のデータ圧縮でも、大きな役割を持つからである。様々な手法のサーベイと比較検討が文献15)に述べられている。また、オプティカルフローに基づく3次元形状の復元と、それにまつわる問題が文献16)で、多角的に具体的な実験結果とともに論じられている。

以下に、前章までの原理の延長としての、オプティカルフローからの対象の運動と3次元形状の復元について簡単にまとめておく。

### 3.2 運動パラメータとオプティカルフロー

座標系を図-6のように定義する<sup>23)</sup>。

前に述べたように、実際には対象が移動する場合でも、カメラが移動するとしても実質的には同じである。ここでは、次回に述べるロボットにカメラを搭載して移動する場合を想定して、カメラの動きによってオプティカルフローが生じる場合を示す。カメラは、カメラ座標の  $x, y, z$  軸回りの回転をしつつ、平行移動をする。静止したカメラに対し対象が移動するときは、最初の脚注<sup>☆</sup>に従って、運動パラメータの表記を入れ替えてやればよい。

ここで、 $V$  を、カメラの平行移動成分  $t$  の速

度で  $V = \left( \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right)^T$  とし、また、 $\Omega = (\Omega_x, \Omega_y, \Omega_z)^T$  を、カメラのカメラ座標  $x, y, z$  軸回りの回転の角速度とする。(3)の表記を用いれば、 $\Omega_x = \frac{d\alpha}{dt}$ ,  $\Omega_y = \frac{d\beta}{dt}$ ,  $\Omega_z = \frac{d\gamma}{dt}$  である。

$k$  枚目の画像から  $dt$  時間後の  $k+1$  枚目の画像の間に、カメラの移動によって、点  $p$  がカメラ座標で表した空間では  $x = (x, y, z) \rightarrow x + dx = (x + dx, y + dy, z + dz)$  と移動したとする。すると、カメラ座標に対する点  $p$  の平行移動の速度成分は、 $-V$  であり、また、回転角速度の成分は、回転行列  $R$  の微分について、 $\frac{dR}{dt} = -R \Omega$  となることから、

$$\frac{dx}{dt} = -V - \Omega \wedge x \tag{26}$$

ここで、右辺の第2項は、 $\frac{dR}{dt}$  が反対称行列になり、ベクトルの外積演算に帰着されることから、このように表記できる。

一方、この間に点  $p$  の画像が  $P_k$  から  $P_{k+1}$  へ移動したとすると、画像座標  $X$  とカメラ座標で表した空間の座標  $x$  とは(11)の  $X = \frac{1}{z} x$  と結び付いていることから、この両辺を時間で微分して、

$$\dot{X} = \frac{1}{z} \left( \dot{x} - x \frac{\dot{z}}{z} \right) \tag{27}$$

<sup>23)</sup> この図では、座標軸をいわゆる右手系にとっている。これまで、座標軸の方向はあまり注意してこなかった。これは、それぞれの場面で座標系が一貫していれば、右手系であるか左手系であるかは問題なかった。しかし、回転の角速度を導入し、それを角速度ベクトルとして表現すると、今度は注意が必要である。以下の式は右手系での表現である。

ここで、 $\hat{z}$  は  $z$  軸 (光軸) 方向の単位ベクトルで、 $\dot{x} \cdot \hat{z}$  は  $\dot{x}$  の  $z$  成分である。

これに、(26) を代入して整理すると、画像面上の動きは、

$$\dot{X} = -\frac{V}{z} + X \frac{V \cdot \hat{z}}{z} - \Omega \wedge X + X(\Omega \wedge X \cdot \hat{z}) \quad (28)$$

各成分に分解すれば、

$$\dot{X} = \left( -\frac{V_x}{z} - \Omega_y + \Omega_z Y \right) + X \left( \frac{V_z}{z} + \Omega_x Y - \Omega_y X \right) \quad (29)$$

$$\dot{Y} = \left( -\frac{V_y}{z} - \Omega_z X + \Omega_x \right) + Y \left( \frac{V_z}{z} + \Omega_x Y - \Omega_y X \right) \quad (30)$$

となる。これが、オプティカルフローと形状を結び付ける方程式である。

この方程式は、次の興味ある内容を語っている。まず、3次元形状の復元は各点の奥行きを求めることであり、結局上式では  $z$  を求めることに相当するが、この奥行き  $z$  は、その運動の平行移動の速度成分との比の形でいつも現れている。つまり、オプティカルフローから、 $z$  を独立して求めることはできない。この場合も、(奥行き)/(スピード) についてのスケールの曖昧さが顔を出しているわけである。

さて、画像上の  $n$  点についての速度ベクトルが与えられると、(29)(30) の2式より  $2n$  個の式と、空間の座標と画像座標の関係を表す(11)の  $2n$  個の式との、合計  $4n$  個の式が与えられる。一方、未知数は、 $n$  点の空間座標の  $3n$  と、上記のスケール分を除いた運動パラメータ5個の  $3n+5$  個ある。したがって、オプティカルフローから、対象の運動と形状の両方を求めるためには、5点以上の点での速度がいる。

また、奥行き  $z$  とオプティカルフロー  $\dot{X}$ ,  $\dot{Y}$  が与えられると、(27) は  $V$ ,  $\Omega$  に関して線形になる。そこで、3点以上について、奥行き  $z$  とオプティカルフローが与えられると、運動が完全に求められる。逆に、 $V$ ,  $\Omega$  が既知ならば、各点についての  $\dot{x}$ ,  $\dot{y}$  から、その点の奥行き  $z$  を決定することができる。

#### 4. おわりに

最初の回でも述べたが、対象の運動とその画像上での動きを記述するためには、以上のようにかなり込み入った数式を必要とした。ただ、順を追って、丁寧に解いていくことで、その仕組み自体はそれほど難しいことではない。

さて、ここまでは、対象とその画像といっても、実は空間の点しか扱ってこなかった。対象が幾何学的な形状のものであれば、その形状を決めるパラメータと、やはり対象の投影で作られる幾何学形状のパラメータの対応を考える方が、応用の場面では実用的であろう。次回(最終回)は、このような立場で、動画像の扱いをもう一度整理して、ロボットビジョンなどの応用を考えてみる。

#### 参考文献

- 1) Aggarwal, J. K. and Nandhakumar, N.: On the Computation of Motion from Sequences of Images, *Proceedings of IEEE*, Vol. 76, No. 8, pp. 917-935 (1988).
- 2) 金谷健一: 3次元運動のコンピューテーション, *光学*, Vol. 21-10, pp. 650-656 (1992).
- 3) Longuet-Higgins, H. C.: A Computer Algorithm for Reconstructing a Scene from Two Projections, *Nature*, Vol. 293, No. 8, pp. 133-135 (1981).
- 4) Tsai, R. Y. and Huang, T. S.: Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Faces, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-6, pp. 13-27 (1984).
- 5) Weng, J., Huang, T. S. and Ahuja, N.: Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-11, pp. 451-475 (1989).
- 6) Zhuang, X., Huang, T. S. and Haralick, R. M.: Two-view Motion Analysis: A Unified Algorithm, *Journal of Optical Society of America*, Vol. A 3, pp. 1492-1500 (1986).
- 7) Hartley, R.: In Defence of the 8-point Algorithm, *Proc. 5th ICCV*, pp. 1064-1070 (1995).
- 8) Phillip, J.: Estimation of Three-Dimensional Motion of Rigid Objects from Noisy Observation, *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI-13, pp. 61-66 (1991).
- 9) Tagawa, N. et al.: Un-biased Linear Algorithm for Recovering Three-Dimensional Motion from Optical Flow, *IEICE Trans. Info. & Syst.*, Vol. E 76-D, No. 10 (1993).



- 10) 金谷：画像理解のための統計学：画像の幾何学的解釈の信頼性評価, 情報処理, Vol. 37, No. 1, pp. 52-60 (Jan. 1996).
- 11) Spetsakis, M. E. and Aloimonos, J.: Optimal Visual Motion Estimation: A Note, IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-14, pp. 959-964 (1992).
- 12) Horn, B. K. P.: Closed-Form Solutions of Absolute Orientations Using Unit Quaternions, J. Opt. Soc. America, A, Vol. 4, No. 4, pp. 629-642 (1987).
- 13) Adiv, G.: Determining Three-Dimensional Motion and Structure from Optical Flow Generated by Several Moving Objects, IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-7, pp. 384-401 (1985).
- 14) Haralick, R. M. et al.: Pose Estimation from Corresponding Point Data, IEEE Trans. System, Man and Cyber., Vol. SMC-19, pp. 1426-1446 (1989).
- 15) Barron, J. et al.: Performance of Optical Flow Techniques, International Journal of Computer Vision, Vol. 12, No. 1, pp. 43-77 (1994).
- 16) Murray, D. and Buxton, B.: Experiments in the Machine Interpretation of Visual Motion, The MIT Press (1990).  
(平成8年3月8日受付)



出口光一郎(正会員)

1976年, 東京大学大学院修士課程修了(計数工学)。同年より東京大学工学部助手, 講師を経て, 1984年, 山形大学工学部情報工学科助教授, 1988年, 東京大学工学部計数工学科助教授, 現在に至る。この間, 1991年~1992年, 米国ワシントン大学客員準教授。コンピュータビジョン, 画像計測, 並列コンピュータの研究に従事。計測自動制御学会, 電子情報通信学会, 形の科学会, IEEE等各会員。