2. 画像理解のための計算理論

2-1 計算理論は如何に役立つか

金 谷 健 一[†]

1. ま え が き

近年、ビデオカメラから入力した画像を計算機で解析することによる製造工程や品質検査の自動化がますます進んでいる。しかし、実用化されているのは、製品や部品の平面部分に描かれた文字、記号や特定のパターンの読取り、あるいは材料の表面や断面の検査などの、主として2次元の解析が主である。一方、カメラ画像は近くのもの大きく、遠くのもの小さく写る「透視変換」としてモデル化できるの、画像を解析すれば、奥行きを含めた物体の3次元的位置や3次元形状も計算できる。画像による3次元認識が可能になれば、産業における潜在的応用は極めて大きいと期待される。今日「画像認識・理解」に対する関心が急速に高まっているのはこのためであると思われる。

しかし、研究としては盛んでも、具体的な産業応用はそれほど進んてはいないようである。確かに、環境の計測や障害物の回避のための画像解析システムを搭載した移動ロボットや自動走行車は盛んに試作され、テモも行われているが、本格的な実用化がなされているとはいいがたい。その原因は、煮詰めればコストと精度・信頼性の問題であろう。コストの問題は技術革新のテンポのみならず、政治、経済、社会的な要因にも左右されやすく、将来的な見通しが困難である（例えば新素子の開発による計算機のコスト/計算能力の大幅な低下もあり得るし、労働条件の変化、政府の政策変化、法律改正等に起因する人件費の高騰によって自動化へのコストの許容水準が高まる可能性もある）それに対比し、精度・信頼性の問題は純粋に工学的な研究によって解決すべき問題である。

ところが、従来の研究・開発においては、精度・信

頼性が中心的課題であっても、そのアプローチは主観的、経験的であることがほとんどである。代表的なのは、設計者がその主観から精度・信頼性を向上させられると思われる手法（同一量を別々の方式で計算して平均をとる、同一量の荷たす複数の方程式に最小二乗法を適用する、カルマンフィルタを盲目的に適用する、等）を導入し、システムを試作し、シミュレーションデータあるいは実データを入力して、その精度・信頼性を経験的に評価し、不十分であればさらに別の手法を導入して試験する、というものである。しかし、これでは試験データに対して基準が満足されても一般性がない。一般性を保証するためには、画像のノイズや各種の処理の精度・信頼性を理論的に評価する誤差の数理モデルの導入によって、システムの出力の精度・信頼性を導出したり、その最適性を証明したりすることのできる理論体系を確立することが必要である。

しかし、画像認識・理解システムの開発においては、そのような数理的なアプローチに対する反発が極めて根強く存在する。その主な主張は次のようなものである。

- まず、画像処理を伴うシステムでは誤差の要因が複雑であり、理論的にモデル化しようとすれば種々の理想化が避けられない。例えば、理論化の都合上から誤差は独立な正規分布に従うと仮定せざるを得ないことも多い。その場合、実際の誤差が独立でなかったり、正規分布からずれていれは、理論的な予想は実際の挙動とは必ずしも一致しない。
- その結果として、経験を積んだ設計者の工夫による方が、実際に試験してみると、理論解析に基づいて設計したものよりも性能が良いことが多い。また、分布の独立性、正規性のような詳細な仮定を用いない方が、モデル化できない例外的な擾乱にも強いはずである。
- そもそも画像認識・理解システムは、人間の視覚

[†] 群馬大学 工学部 情報工学科

“Computational Theory for Image Understanding, The Role of Theoretical Analysis” by Kenichi Kanatani (Department of Computer Science, Gunma University, Kiryu, Gunma)

認識機能を模倣しようとすることから出発しているものであり、人間は複雑な環境下でも理論解析や最適化計算などせずに、うまく処理している。すべての要因を忠実に解析できない以上、機械化にあたっては、この人間の処理に学ぶのがよい。

画像認識・理解システムを含む知的処理システムの研究では、「人工知能」、「ファジィ」、「ニューロ」などと呼ばれるような非数理的アプローチが主流である。確かにそのようなアプローチも重要であろうが、例えば理想化を伴ってでも、数理的に解決できる問題は数理的に解決しようという態度がなければ、いつまでも研究段階に始終し、論文やデモばかりが量産され、肝心の産業応用に進まないのではなかろうか。数理的アプローチによると、理想化による理論的予想と実際の挙動とが一致していれば理想化に用いた仮定が成立していると結論されるし、大きくずれるようであれば理論中に仮定していない何らかの要因が存在することを意味し、その要因の理論的解明が次の研究課題となる。このように、理想化のプロセスは次々と発展する可能性を秘めている。それに対して、主観的な工夫によって満足できる試験結果が得られたとしても、あるいは得られないとしても、その原因についての理解はそれ以上進まないことが多い。

筆者が最近取り組んでいる研究は、画像の誤差の数理モデルを導入し、処理の精度を理論的に評価する理論体系を確立しようとするものである。この理論は画像データから計算される諸量の不確定性を共分散行列で評価して、2次元画像上および3次元空間中での幾何学的計算の最適化を追及する、いわば幾何学と統計学を統合する理論体系である。

2. 最適推定の理論

「画像認識・理解」の基本となるのは、画像あるいはシーンに関する何らかの“知識”をもとにして、画像やそれから計算されるデータに解釈を与えることである。定性的な知識は“ルール”と呼ばれることが多く；定量的な知識は“拘束”と呼ばれることが多い。定量的な拘束は（一般には複数の）方程式として表される。

$$F^{(k)}(\mathbf{a}, \mathbf{u}) = 0, \quad k=1, \dots, L \quad (1)$$

これは幾何学的、光学的、物理的な諸法則を表している。幾何学的な拘束とは、例えば物体の複数の特徴点が空間中あるいは画像上で“共線”である（同一直線上にある）とか、複数の特徴点や線分が空間中で“共面”である（同一平面上にある）とか、複数の線分が“共点”である（延長すれば同一の交点を持つ）とか、ある線分が別の線分と平行であるとか直交してい

るというような知識のことである。光学的な拘束とは、物体の表面が拡散反射面であるとか鏡面反射面であるとか、あるいはその合成であるとか、光源は何個どちらの方向にあって、そのスペクトル特性はどうであるというような知識のことである。また、物理的拘束とは、物体は力学的につり合いの位置にあるとか、剛体運動をしているとか、並進速度あるいは回転角速度が一定であるとか、ニュートンの運動方程式に従っているというような知識のことである。

方程式(1)の変数は2組に分けられる。一組 $\mathbf{a}$ は画像（一般にはセンサ）から計測できる量であり、もう一組 $\mathbf{u}$ は推定したい未知パラメータである。この拘束を満たすべき多数のデータ $\{\mathbf{a}_\alpha\}$ が観測されたとする。誤差のない正確なデータであれば、未知数の個数に等しい個数のデータを(1)式に代入すると一般には $\mathbf{u}$ が求まるはずである。しかし誤差があれば、一般には(1)式が満たされない。このとき、どのようにパラメータ $\mathbf{u}$ を推定するのがよいであろうか。

これを理論的に解析するために、次のような誤差モデルを導入する。各データ $\mathbf{a}$ は真の値 $\bar{\mathbf{a}}$ に誤差 $\Delta\mathbf{a}$ が加わって

$$\mathbf{a}_\alpha = \bar{\mathbf{a}}_\alpha + \Delta\mathbf{a}_\alpha \quad (2)$$

であるとする。そして、誤差 $\Delta\mathbf{a}_\alpha$ は各 α で独立で、期待値 $\mathbf{0}$ の正規分布に従う確率変数と仮定する。その共分散行列を

$$E[\Delta\mathbf{a}_\alpha \Delta\mathbf{a}_\alpha^T] = V[\mathbf{a}_\alpha] \quad (3)$$

と置く。ただし $E[\cdot]$ は期待値を表し、記号 T は転置を表す。各 $\mathbf{a}_\alpha$ の成分は独立ではなく、何らかの制約、例えば $\|\mathbf{a}_\alpha\|=1$ 等の条件が課されていてもよい。以下、ベクトルの内積を $(\mathbf{a}, \mathbf{b})$ で、ノルムを $\|\mathbf{a}\|$ と表す。

誤差 $\Delta\mathbf{a}_\alpha$ は確率変数であるから予測不可能であるが、共分散行列 $V[\mathbf{a}_\alpha]$ は撮像系や画像処理の精度、センサの特性から定まる既知量である。パラメータ $\mathbf{u}$ の最適な推定は次の評価関数の最小化として定式化できる（これはある意味での“最尤推定”に基づいているが、統計学で通常用いられているものとは異なる。詳細省略）。

$$J[\mathbf{u}] = \sum_{\alpha=1}^N \sum_{l=1}^L W_\alpha^{(kl)}(\mathbf{u}) F^{(k)}(\mathbf{a}_\alpha, \mathbf{u}) F^{(l)}(\mathbf{a}_\alpha, \mathbf{u}) \rightarrow \min \quad (4)$$

$$W_\alpha^{(kl)} = ((\nabla_{\mathbf{a}} F_\alpha^{(k)}, V[\mathbf{a}_\alpha] \nabla_{\mathbf{a}} F_\alpha^{(l)})_{\bar{\mathbf{r}}} \quad (5)$$

ただし $\nabla_{\mathbf{a}} F^{(k)} = (\partial F^{(k)} / \partial a_i)$ であり、 $F_\alpha^{(k)} = F^{(k)}(\mathbf{a}_\alpha, \mathbf{u})$ 、 $\nabla_{\mathbf{a}} F_\alpha^{(k)} = \nabla_{\mathbf{a}} F^{(k)}(\mathbf{a}_\alpha, \mathbf{u})$ と略記した。また (kl) 要素が A_{kl} の行列を (A_{kl}) と書き、その逆行列およびその (kl) 要素を共に $(A_{kl})^{-1}$ と書く。記号 $(A_{kl})_{\bar{\mathbf{r}}}$ はランク

r の“一般逆行列”を表す．行列 $(W_a^{(kl)})$ のランク r は (1) 式の拘束のうちの独立なものの個数に等しい (証明省略)．

これによって，点列に対する直線の最適当てはめ，複数の直線の共通の交点の最適な推定，コニック等の曲線の最適当てはめ，運動する物体の 3 次元運動の最適な推定，オプティカルフローからの最適な 3 次元復元などが可能となる (詳細省略)．

3. 最適推定値の性質

パラメータ $\mathbf{u}$ の最適な推定値 $\hat{\mathbf{u}}$ を求めるには (4) 式の評価関数 $J[\mathbf{u}]$ の最小値を探索すればよいが， n 次元ベクトル $\mathbf{u}$ に何らかの制約条件 (例えば $\|\mathbf{u}\|=1$) が課されていて，自由度が $n' (\leq n)$ に限定されているもよい．したがって， $J[\mathbf{u}]$ の定義域は $\mathbf{u}$ の全空間ではなく，その n' 次元部分多様体となる．いま何らかの数値探索アルゴリズム (最急降下法，共役勾配法，準ニュートン法，その他) によって，その部分多様体上で最適推定値 $\hat{\mathbf{u}}$ が計算されたとする．この最適推定値は第 1 近似として次の性質をもつ (証明省略)．

1. 最適推定値 $\hat{\mathbf{u}}$ は統計的に不偏であり，その期待値 $E[\hat{\mathbf{u}}]$ は真の値 $\bar{\mathbf{u}}$ に一致する．
2. “モーメント行列”を

$$\mathbf{M} = \left(\sum_{a=1}^N \sum_{l=1}^L \bar{W}_a^{(kl)} (\mathbf{P}_U \nabla_{\mathbf{u}} \bar{F}_a^{(k)}) (\mathbf{P}_U \nabla_{\mathbf{u}} \bar{F}_a^{(l)})^T \right) \quad (6)$$

と定義すると，最適推定値 $\hat{\mathbf{u}}$ の共分散行列 $V[\hat{\mathbf{u}}]$ はその一般逆行列で与えられる．

$$V[\hat{\mathbf{u}}] = \mathbf{M}^{-1} \quad (7)$$

3. (1) 式の拘束のうちの独立なものの個数を r ，パラメータ $\mathbf{u}$ の自由度を n' とすると，残差

$$J[\hat{\mathbf{u}}] = \sum_{a=1}^N \sum_{k,l=1}^L \bar{W}_a^{(kl)} \hat{F}_a^{(k)} \hat{F}_a^{(l)} \quad (8)$$

は自由度 $rN - n'$ の χ^2 分布に従う．

ただし，これが厳密に成立するのは共分散行列 $V[\mathbf{a}_a]$ をデータ $\mathbf{a}_a$ の期待値 $\bar{\mathbf{a}}_a$ で評価した場合である．しかし $\bar{\mathbf{a}}_a$ は未知であるから，数値計算では $V[\mathbf{a}_a]$ を観測データ $\mathbf{a}_a$ において評価せざるを得ない．したがって，実際に数値計算できるのは理論的な最適推定値 $\hat{\mathbf{u}}$ の近似値である．また $\mathbf{u}$ の“真の値” $\bar{\mathbf{u}}$ とは，各データ $\mathbf{a}_a$ の期待値 $\bar{\mathbf{a}}_a$ に対して拘束 $F^{(k)}(\bar{\mathbf{a}}_a, \mathbf{u}) = 0$ を成立させる $\mathbf{u}$ の値のことであり，そのような値 $\bar{\mathbf{u}}$ の存在は仮定している．なお，(6)，(8) 式でバーとハットはそれぞれ変数 $\mathbf{u}$ に真の値 $\bar{\mathbf{u}}$ および推定値 $\hat{\mathbf{u}}$ を代入することを意味している．また $\nabla_{\mathbf{u}} F^{(k)} = (\partial F^{(k)} / \partial u_i)$ であり， $\mathbf{P}_U$ は $\mathbf{n}$ の制約を表

す多様体の接空間への直交射影子である．

ところでパラメータ推定の立場からは，

(a) (1) 式の制約条件が成立して，

(b) 観測されるデータはその制約を満たす量の，誤差を伴った実現値である．

という仮定に基づいている．しかし，上の結果を利用すれば逆に，そのような仮定が成立するかどうかを確かめることができる．すなわち，(a)，(b) を“仮説”とし，仮説に基づいて推定した $\hat{\mathbf{u}}$ に対する $J[\mathbf{u}]$ 残差を計算し，自由度 $rN - n'$ の χ^2 分布を用いて指定した信頼水準で仮説が棄却できるかどうかの検定を行うことができる．これによって，特徴点が共線であるか，特徴点や線分が共面であるか，線分が共点であるか，ある線分が別の線分と平行であるとか直交しているか，などが体系的に同一基準で検定できる (詳細省略)．

4. 最適補正の理論

前節の手続きによってパラメータが推定されたとき，各データは誤差のために拘束を厳密には満たしていない．しかし，3 次元環境のモデル化のために各データが拘束を厳密に満たすように補正しなければならないことも多い．いま N 個の幾何学的対象が存在し，それぞれパラメータ $\mathbf{u}_1, \dots, \mathbf{u}_N$ で記述されるとする．各 $\mathbf{u}_i$ は統計的には独立であるとし，その共分散行列を $V[\mathbf{u}_i]$ とする．ただし各 $\mathbf{u}_i$ の成分は独立ではなく，何らかの制約 (例えば $\|\mathbf{u}_i\|=1$ 等) が課されていてよいとする．このとき N 個の幾何学的対象 $\mathbf{u}_1, \dots, \mathbf{u}_N$ を補正して $\mathbf{u}'_1, \dots, \mathbf{u}'_N$ とし，拘束

$$F^{(k)}(\mathbf{u}'_1, \dots, \mathbf{u}'_N) = 0, \quad k=1, \dots, L \quad (9)$$

を満足させたいとする．このとき最適な線形補正は次のように与えられる (これは“マハラノビス計量”による最適化に基づいている．詳細省略)．

$$\mathbf{u}'_i = \mathbf{u}_i - V[\mathbf{u}_i] \sum_{k,l=1}^L W^{(kl)} F^{(k)} \nabla_{\mathbf{u}_i} F^{(l)} \quad (10)$$

$$W^{(kl)} = \left(\sum_{i=1}^N (\nabla_{\mathbf{u}_i} F^{(k)}, V[\mathbf{u}_i] \nabla_{\mathbf{u}_i} F^{(l)}) \right)^{-1}_r \quad (11)$$

$$V[\mathbf{u}_i]' = \mathbf{P}_{U_i} V[\mathbf{u}_i] \mathbf{P}_{U_i} \quad (12)$$

ただし $\mathbf{P}_{U_i}$ は各 $\mathbf{u}_i$ の制約を表す部分多様体の接空間への直交射影子であり， r は (9) 式の拘束の独立なものの個数である．補正した $\{\mathbf{u}'_i\}$ が拘束条件 (9) を完全には満足していなければ $\mathbf{u}_i$ ， $V[\mathbf{u}_i]$ を $\mathbf{u}'_i$ ， $V[\mathbf{u}'_i]$ と置き換えて，この補正を反復する．

これを用いれば，物体の特徴点が共線であると判定され，それらを通る直線が推定されたとき，各点をその直線上へ統計的に最適に投影できる．また特徴点や

線分が共面あること、あるいは線分が共点であることが判定され、共通の平面や共通の交点が推定されたとき、個々の点や線分が共面あるいは共点であるように統計的に最適に補正できる。あるいは空間中のある線分が別の線分と平行であるとか直交するべきであるという知識があるとき、そのような線分が厳密に平行となる、あるいは直交するように統計的に最適に補正できる（詳細省略）。

5. 拡張カルマンフィルタ

(4)式の最小化ではデータ $\{a_\alpha\}$ が一括して利用できると仮定している。しかし α が時刻を表し、 $\alpha=1, 2, 3, \dots$ に対して順にデータが得られるとしたら、リアルタイム処理として望ましいのは、時刻 α での最適推定値 $\hat{u}_\alpha$ とその共分散行列 $V[\hat{u}_\alpha]$ に対して、データ $a_{\alpha+1}$ が得られた時点で次の推定値 $\hat{u}_{\alpha+1}$ とその共分散行列 $V[\hat{u}_{\alpha+1}]$ を計算することである。またこれは時系列でない場合も、(4)式の最小化のひとつの数値解法を与えているとみなせる。なぜなら、推定値の更新の量が小さければ更新過程は線形計算で近似できるので、これによってひとつの反復解法が得られるからである。

これを計算すると、これは“内部状態が変化しない”「線形動的（静的？）システム」の内部状態と観測過程に独立な正規分布に従う誤差が加わる系に対する「カルマンフィルタ」と等価でもある。この理由で、この逐次更新過程は「拡張カルマンフィルタ」と呼ばれる。具体的な更新規則は次のようになる。

$$\hat{u}_{\alpha+1} = \hat{u}_\alpha - V[\hat{u}_\alpha] \sum_{k,l=1}^L \tilde{W}_\alpha^{(kl)} F_\alpha^{(k)} \nabla_u F_\alpha^{(l)} \quad (13)$$

$$V[\hat{u}_{\alpha+1}] = V[\hat{u}_\alpha] - \sum_{k,l=1}^L \tilde{W}_\alpha^{(kl)} (V[\hat{u}_\alpha] \nabla_u F_\alpha^{(k)}) (V[\hat{u}_\alpha] \nabla_u F_\alpha^{(l)})^\top \quad (14)$$

$$\tilde{W}_\alpha^{(kl)} = ((\nabla_u F_\alpha^{(k)}, V[\hat{u}_\alpha] [\nabla_u F_\alpha^{(l)}] + (\nabla_u F_\alpha^{(k)}, V[\hat{u}_\alpha] \nabla_u F_\alpha^{(l)})^{-1} \quad (15)$$

ただし $F_\alpha^{(k)} = F^{(k)}(a_{\alpha+1}, \hat{u}_\alpha)$, $\nabla_u F_\alpha^{(k)} = \nabla_u F^{(k)}(a_{\alpha+1}, \hat{u}_\alpha)$ と略記した。しかし、上の更新過程はこれを導く途中での線形近似が粗く、解は統計的に不偏ではないことが知られている。近似をより詳しくして偏差を除去するには、式中の $\nabla_u F_\alpha^{(k)}$ を次式で定義する「有効勾配」に置き換えればよい。

$$\nabla_u^* F_\alpha^{(k)} = \nabla_u F_\alpha^{(k)} + \sum_{m,n=1}^L (W_\alpha^{(km)})^{-1} F_\alpha^{(n)} \nabla_u W_\alpha^{(mn)} \quad (16)$$

このような補正を加えたものは「偏差補正拡張カルマンフィルタ」と呼ばれている。ここでは簡単のために

パラメータ u に制約のない場合を示したが、拘束が存在する場合は式中で、 u の制約を表す部分多様体の接空間への直交射影 $P_{\hat{u}_\alpha}$ を導入し、逆行列を一般逆行列に置き換えたりする変形が必要となる（詳細省略）。

このような拡張カルマンフィルタによれば、極めて長い時系列画像からの運動推定や、あるいは時系列でなくても多数のデータの読出しを伴う推定を効率的に行うことができる。後者の例としては、長い画素列に対する曲線当てはめやオプティカルフローの3次元解釈がある。オプティカルフローが全画素（例えば 512×512 画素）で定義され、その x, y 成分がそれぞれ原画像と同じ画素数を持つメモリーフレーム2枚に分けて記憶されているとすると、1つずつフローの成分を読出して演算して行けば、読み終わったときに解が得られる。

6. くりこみ法

画像認識・理解に現れる拘束のほとんどは、次のような線形方程式で表すことができる。

$$(a^{(k)}, u) = 0, \quad k=1, \dots, L \quad (17)$$

もともとの形が線形でなくても、適当な変数変換によってこのような形に帰着できることが多く、交点の計算、直線や曲線当てはめ、点や直線や平面の補正、3次元運動解析、オプティカルフローの3次元解釈などの問題がこの範疇に入る。(17)式から明らかなように、ベクトル u の定数倍の不定性が残る。そこで $\|u\|=1$ と正規化する（他のどんな正規化を用いても等価である）。また、誤差の絶対量を表す定数を ϵ とする。このとき最適化(4)式は次の形に書ける。 $V^{(kl)}[a_\alpha]$ は $\epsilon=1$ と仮定した $a_\alpha^{(k)}$ と $a_\alpha^{(l)}$ との共分散行列である。

$$J[u] = \sum_{k=1}^L \sum_{l=1}^L W_\alpha^{(kl)}(u) (a_\alpha^{(k)}, u) (a_\alpha^{(l)}, u) \rightarrow \min \quad (18)$$

$$W_\alpha^{(kl)} = ((a_\alpha^{(k)}, V^{(kl)}[a_\alpha] a_\alpha^{(l)})^{-1} \quad (19)$$

汎用のライブラリプログラムによる数値探索を行うには、探索の初期値を与えなければならないが、真の解に十分近い初期値を与えなければ、探索が誤った極小値に落ち込む恐れがある。しかし、上の形であれば探索を行わずに直接的に計算する次のような方法が考えられる。これを「くりこみ法」と呼ぶ。ただし、パラメータ u には $\|u\|=1$ 以外には制約はなく、最適化の定義域は $(n-1)$ 次元単位球面 S^{n-1} であるとする。

1. $c=0$, $W_\alpha^{(kl)} = \delta_{kl}$ と置く (δ_{ij} はクロネッカのデルタ)。

2. モーメント行列 M と行列 N を次のように計算する.

$$M = \left(\sum_{a=1}^N \sum_{l=1}^L \bar{W}_a^{(kl)} \mathbf{a}_a^{(k)} \mathbf{a}_a^{(l)\top} \right) \quad (20)$$

$$N = \sum_{a=1}^N \sum_{l=1}^L W_a^{(kl)} V^{(kl)}[\mathbf{a}_a] \quad (21)$$

3. 行列

$$\hat{M} = M - cN \quad (22)$$

の最小固有値 λ に対する単位固有ベクトル $\mathbf{u}$ を計算する.

4. $|\lambda| \approx 0$ であれば $c, \hat{M}, \mathbf{u}$ を返して終了する.
5. 定数 c と重み行列 ($W_a^{(kl)}$) を次のように更新する.

$$c \leftarrow c + \frac{\lambda}{(\mathbf{u}, N\mathbf{u})} \quad (23)$$

$$W_a^{(kl)} \leftarrow ((\mathbf{u}, V^{(kl)}[\mathbf{u}_a])^{-1}) \quad (24)$$

6. ステップ 2 へ戻る.

この方法は収束が速く、次の特徴がある.

- 返された $\mathbf{u}$ は不偏な推定値 $\hat{\mathbf{u}}$ であり、その期待値は真の値 $\bar{\mathbf{u}}$ に一致する.
- 推定値 $\hat{\mathbf{u}}$ の共分散行列は $V[\hat{\mathbf{u}}] = \varepsilon^2 \hat{M}^{-1}$ で与えられる.
- rNc/ε^2 が自由度 $rN - n + 1$ の χ^2 分布に従う.
- ε^2 の不偏推定値が $\hat{\varepsilon}^2 = c/(1 - (n-1)/rN)$ で与えられる.

このように、単に推定値が効率的に計算できるだけでなく、その共分散行列まで同時に計算でき、しかも“誤差の大きさ ε が未知でもよい”. 共分散行列には定数倍の不定性があっても解 $\hat{\mathbf{u}}$ には影響せず、その絶対量を表す定数 ε が推定できるからである. 画像やセンサの誤差モデルを立てるとき、誤差の“幾何学的構造”をモデル化することは比較的容易であるが、その絶対量を推定することは困難であることが多い. 例えば、画像から検出した特徴点の位置には等方的で場所によらないガウス分布に従う誤差が含まれていると仮定するのは、多くの場合に現実的であるが、その分布の広がり特徴づける平方二乗平均の大きさを正確に推定することは容易ではない. 上に述べたくりこみ法では誤差の幾何学的モデルさえ立てればよく、その大きさは事後的に推定できる. これは例えていえば、点列に直線を当てはめるとき、各点の位置に含まれている誤差を当てはめた直線とデータ点とのずれから事後的に評価することに相当している.

7. 線形化手法

変数変換によって(17)式の形に帰着する例として最も単純なのは、拘束が

$$(\phi^{(k)}(\mathbf{a}), \mathbf{u}) = 0 \quad (25)$$

の形をしている場合である. ただし $\phi^{(k)}(\mathbf{a}) = (\phi_i^{(k)}(\mathbf{a}))$ であり、各 $\phi_i^{(k)}(\mathbf{a})$ は $\mathbf{a}$ の非線形関数である. このときは $\mathbf{a}_a$ の代わりに $\hat{\mathbf{a}}_a^{(k)} = \phi^{(k)}(\mathbf{a}_a)$ をデータとみなせばよい. データ $\mathbf{a}_a$ の共分散行列 $V[\mathbf{a}_a]$ が与えられているとき、変換したデータ $\hat{\mathbf{a}}_a^{(k)}$ の共分散行列 $V^{(kl)}[\hat{\mathbf{a}}_a]$ を計算するのは容易である (変換のヤコビ行列で合同変換すればよい). このとき、誤差の4次の項まで考慮した「2次のくりこみ法」が定式化できる (詳細省略).

重要なのは、拘束が次の形をしている場合である.

$$(\mathbf{a}^{(k)}, \phi(\mathbf{u})) = 0 \quad (26)$$

ただし $\phi(\mathbf{u}) = (\phi_i(\mathbf{u}))$ であり、各 $\phi_i(\mathbf{u})$ は $\mathbf{u}$ の非線形関数である. このときは $\mathbf{u}$ の代わりに $\hat{\mathbf{u}} = \phi(\mathbf{u})$ を未知パラメータとすればよい. 新しいパラメータ $\hat{\mathbf{u}}$ を「基本パラメータ」と呼ぶ. これは「線形化手法」と呼ばれている. これによって、例えばくりこみ法によって解 $\hat{\mathbf{u}}$ が簡単に計算できる. ただし、以下に述べる問題が生じる.

ベクトル $\mathbf{u}, \hat{\mathbf{u}}$ の次元をそれぞれ $n, \bar{n}$ とするとき、 $\bar{n} < n$ であれば、計算した $\hat{\mathbf{u}}$ から元の $\mathbf{u}$ が定まらない. この場合は、その拘束を適用すること自体が不適切である.

一方 $\bar{n} > n$ であれば、連立方程式 $\phi_i(\mathbf{u}) = \hat{u}_i, i=1, \dots, \bar{n}$ を解いて $\mathbf{u}$ が定まるはずであるが、誤差があると解が存在するとは限らない. 変数 $\hat{\mathbf{u}}$ の作る $\bar{n}$ 次元空間において、方程式 $\hat{u}_i = \phi_i(\mathbf{u})$ は n 次元部分多様体 Σ を定義している. そのような部分多様体 Σ 上の点は「分解可能」とであるという. 誤差がなければ線形化手法によって計算した $\hat{\mathbf{u}}$ は Σ 上に求まるはずであるが、誤差のために一般には Σ から離れたところに求まり、分解可能ではない. このときは次のようにするのが簡便である.

部分多様体 Σ は陰関数による方程式

$$\Psi_k(\hat{\mathbf{u}}) = 0, \quad k=1, \dots, \bar{n} - n \quad (27)$$

として表すことができる. これを「分解可能条件」と呼ぶ. 計算した $\hat{\mathbf{u}}$ の共分散行列は (例えば、くりこみ法によって) 簡単に計算できるから、4節に示した手法によって $\hat{\mathbf{u}}$ を分解可能条件が満たされる $\hat{\mathbf{u}}'$ に統計的に最適に補正すればよい. 補正した $\hat{\mathbf{u}}'$ は分解可能であるから、対応する $\mathbf{u}$ を (例えば数値計算によって) 求めることができる.

このような線形化手法が活用できる代表的な例は、動画画像からの3次元復元とオプティカルフローの3次元解釈である. 真に求めたいのは、並進と回転、あるいは瞬間的な並進速度と回転速度からなる「運動パラメータ」であるが、それらの満たす拘束 (「エビ極線拘

束条件」と呼ばれる)は運動パラメータに関して非線形である。しかし、運動パラメータを変換した「基本行列」あるいは「基本パラメータ」という量に関しては線形であり、くりこみ法で容易に計算できる。それを計算した後に、それを運動パラメータに統計的に最適に分解すればよい(詳細省略)。

8. む す び

本解説では、画像認識・理解における数理的方法論の役割を論じ、定量的推定の背後にある数学的な体系を一般的に記述した。画像認識・理解の個々の問題に精通している人には、これが極めて広い範囲の問題に適用できることに気がつくと思うが、そうでない人にとっては抽象的過ぎたかも知れない。しかし個々の問題を取り上げると、その問題の特殊性に目を奪われ、一般的な体系によって解決できることに気がつきにくい。現状では問題ごとに個別に、例えば3次元形状への曲面当てはめを研究している人はそのみに没頭し、オプティカルフローの3次元解釈を研究している人はまたその問題のみに没頭し、互いに同じ手法によって問題が解決するという事実はまだ気づいていないようである。今後は本解説に示した一般論によって画像認識・理解における数理的方法が体系化されること

が期待される。

(1993年9月14日受付)

【参 考 文 献】

- 1) 金谷健一: “画像理解/3次元認識の数理”, 森北出版 (1990)
- 2) 金谷健一: “3次元運動のコンピュータシミュレーション”, 光学, **21**, 10, pp. 650-656 (Oct. 1992)
- 3) K. Kanatani: “Geometric Computation for Machine Vision”, Oxford Univ. Press, Oxford (1993)
- 4) K. Kanatani: “3-D Interpretation of Optical Flow by Renormalization”, Int. J. Comput. Vision, **11**, 3, pp. 267-282 (1993)
- 5) 金谷健一: “画像の3次元解釈の統計的信頼性”, 情報学論, **34**, 10 (Oct. 1993)
- 6) J. Porrill: “Optimal Combination and Constraints for Geometrical Sensor Data”, Int. J. Robotics Research, **7**, 6, pp. 66-77 (1988)
- 7) J. Porrill: “Fitting Ellipses and Predicting Confidence Envelopes Using a Bias Corrected Kalman Filter”, Image & Vision Comput., **8**, 1, pp. 37-41 (1990)
- 8) N. Tagawa, T. Toriu and T. Endoh: “Un-biased Linear Algorithm for Recovering Three-dimensional Motion from Optical Flow”, IEICE Trans, E76-D, **10**, pp. 1263-1275 (1993)

かなたに けんいち
金谷 健一 昭和54年、東京大学大学院計数工学専攻(数理工学)博士課程修了。昭和63年、群馬大学工学部情報工学科教授。工学博士。

2-2 上手に拘束条件を設定するには

—多重視覚情報の抽出を例にして—

志 沢 雅 彦[†]

1. ま え が き

画像は単なる輝度値の配列ではなく、物理法則に基づいて生成された外界の情景が幾何学的変換によって投影されたものである。その生成には、多くの要因が同時に複雑に絡み合っている。これは、例えば図1のような自然環境を想像すればわかる。しかし、従来のコンピュータビジョンのアルゴリズムは、このよ

うな複雑な状況を正しく扱うことができなかった。しかし人間は、このような状況下でも比較的困難を感じずに物理世界の属性を理解することができる。その典型的な例が透明視(トランスペアレンシー)(例えば、本小特集4.参照)であろう。本稿では、計算機に人間と同等の視覚機能を与えるという立場から、一般に同一種の複数の画像生成要因が画像の同一点に同時に影響を与える場合に、これらの生成要因を推定することを透明視と定義する。

一方、ビジョン一般における最も困難な問題のひとつとして、セグメンテーションがあげられる。これは、画像を個々の物体の占める領域に領域分割する問題であって、ビジョンの応用を考える場合、常に最初

[†] 株式会社 ATR 人間情報通信研究所

“A Nice Way to Create Constraints for Vision—Extracting Multiple Visual Attributes—” by Masahiko Shizawa (ATR Human Information Processing Research Laboratories, Kyoto)