コンピュータビジョン：単に数学的問題なのか

若　　者　　ミヒャエル　ヒルド*

先日，休憩時間にうちの研究室の助教授と，コンピュータビジョンの分野において増え続ける“mathematization”（数学化）という，面白い現象について話しあった．コンピュータビジョンのうちのいくつかの特定な面でclosed-form analytic solutionsを試みるなど，かなり複雑な数学的分析をしている論文が，近頃いくつも発表されている．この現象は“shape-from-motion”やCTスキャンからの三次元再構成問題において非常に顕著なのだがエッジ抽出（例えばCanny operator等）やテクスチャー領域分割（例えばUnserの一連の論文等）といった低いレベルの分野でも同じ傾向がみられる．とりわけ，先頃出版されたLenz著のGroup Theoretical Methodsについての本などは，画像理解問題に関する論文というより，むしろ数学のコースの教科書のようにみえる．限られた範囲の問題に，より一層詳細な数学的分析を加えようとするこの傾向の意義について，果たしてこれは画像を「理解」するのに必要な手段を考案することを目的とするこの分野の発展に本当に役立つのだろうか，また，こういう努力の中に高度に確実，かつ一般的な（幅広い範囲での応用に適するという意味で）人工ビジョン・システムを最終的に作り出せる可能性が本当にあるのだろうか，と助教授と私は考えた．このような，はっきりとは定義づけられないが基本的な問題に対し，根拠十分な判断をくだすのは大変難しいのだが，いろいろと慎重熟考してみる価値はあるだろう．

画像理解の究極の目的は，物質的な対象物やそれらの互いの関係を認識することである．物質のイメージは，光の伝播，反射，回析，吸収，散乱，収差等といった物理的現象を経て形成される．画像理解のこの面は，非常に重要であり，かなりよく理解されている．二次元イメージ・データから三次元的形状を決定する，中心的問題のことを考えれば容易に証明できるように，考慮すべき問題点は別の所に起因するのである．よくわかっているプロセスによって，画像が形成される時，奥行きの情報を取り去ることにより，三次元的物体のデータ・セットの次元は，画像の上で二次元に減少する．奥行きの喪失は，画像から三次元的世界への逆写像を拘束不十分（underconstrained）にする原因となり，拘束不十分の問題に正確で一意な解決がないということは，よく知られているのである．この逆写像の問題は数学的にはill-posedである．一方生物のビジョン・システム，特に人間の視覚系は，このill-posedな問題を，少なくとも定性的に解くことができ，また，量的解決のかなり正確な評価さえ，ある程度出せるのである．

拘束不十分な問題を数学的に取り組みやすくする一つの方法は，拘束を加えること，すなわち，ある種の場面や形状，あるいは，照明条件等を仮定するのである．拘束が本当に有効であれば，拘束不十分の問題は，御しやすくなるのだろうが，有効でなければ，間違った解決を「もとめる」ことになり，解決が正しいか否かを確かめる方法すらないかもしれない．例えば，形状再現問題でよく使われている拘束は，regularization methodsに沿うように，すべての表面がそれぞれなめらかであるとする仮定である．しかし，それぞれの面がなめらかではない自然の形状は数多く，たとえ正常な状態でそれぞれの面がなめらかであっても，腐食，ペンキのしみ，油等によってざらざらとした形状を呈することがある．だから，単に追加の拘束

*ヒルド　ミヒャエル(Michael HILD)，工学部電子制御機械工学科，白井研究室，産業機械

を加えるだけでは，一般的に，コンピュータ・ビジョンにおける，拘束不十分の数学的問題に対する，強固で正しい解決を求めるための役には立たない．しかしながら，完全に拘束を無視するというのも不可能である．例えば，形状の空間的変化の周波数が，常に，場面における照明の空間的変化の周波数より高いということは，基本的なことである．この仮定により，我々はスライド画の投影を正確に解釈でき，それは，この仮定なしでは不可能なのである．そういった基本的な性質の仮定は，不可避であり，この仮定が無効な場面では，間違った解釈をひきだしてしまう．

拘束不十分の数学的問題を解くより良い方法は，付加的な情報を求めることである．例えば，形状構成問題で，一対のステレオ画像を用いるのは，有望なことだ．というのは，双方の画像における三次元的物体の，ある点の画像上の座標がわかれば，方程式は解けるからである．双方の画像における物体の，ある点の座標の計測には，双方の画像において，その点を求めることが必要であり，これは，対応づけ問題として知られている．双方の画像において，その点を含んでいる局所的領域は，二次元のエピポーラー線上で正確に照合できるものと仮定する．この照合プロセスは，かならずしも単純で正確なものではなく，その点が，オクルージョンのために，どちらか片方の画像でしか見えない時は，照合させることが可能にさえなるということが，経験からわかる．そういう場合には，ミス・マッチさえおこりうるのである．オクルージョンの問題を正しく取り扱うために，画像のどこにオクルージョンがおこるのか，照合を試みる前に知る必要がある．しかしこれには，双方の一眼画像における，部分の認識が必要となり，部分を認識する前には，形状を再構成しなければならず，今までに述べてきたように，それを二次元から三次元への逆写像によって直接試みるのは，扱いにくいことなのである．実際，我々はステレオを応用して逆写像を解決したいと思っていたのだから，もし我々が形状再構成問題の正確な解決を期待しているのなら，我々がここで取り組んでいるのは，典型的な，卵が先かニワトリが先か，といった問題だということになる．画像列が分析される動画像分析でも，基本的に同じ理論が適用される．その場合，複数の対応づけの問題を解くことが必要となるのだが，一つの対応づけ問題すら正確に解けないなら，どうしてこれがうまくいくなどと予期できよう．私の見解からすれば，一眼視野からの部分認識に代わるべき良い手段はなく，もしそれが正確に可能でないならば，定性的に行なわなければならない．そうすれば，定性的な解決が見つかった時に，それをステレオ，あるいは動画像列を用いて，精密に，そして正確にすることができるし，ステレオあるいは動画像列の分析は，一眼分析の定性的結果によって導かれるはずである．

ここでのキーワードは，「定性的」であり，量的の逆であることを強調している．最近の談話で，メリーランド大学のRosenfeld教授は，定性的記述を「どこでもよい部分的記述」と定義していた．これは，たしかに非常に一般的な定義である．私自身の意見では，「定性的」の意味は，アフィン変換のもとで不変である幾何学的特徴だけが意味を持つ，位相学の概念に非常に近く，すなわち，頂点，端，面，そしてそれら相互の配置と順列は重要だが，それらの部分の絶対的サイズや，互いの絶対的距離は，重要でないということである．すべての絶対的寸法を完全に無視することは，おそらく不可能だろうが，寸法をできるだけ正確にしようとしたり，あらゆる場所に寸法をいれようとしたりは，するべきでない―この意味で，定性的記述は，部分的なものとなるが，そこには，できるかぎり多くの位相学的特性や属性が含まれているべきである．

投影のもとでゆがみが限られている一群の局所的特徴から，物体は部分的に認識されなければならない．例えば，形状のテクスチャーは，この目的にとって有力な候補である．そのような局所的特徴に基づいた「認識」は，最終的な結果とはみなされない．むしろ，それには，知識に基づいた推測，あるいは，仮説の性質があり，結晶を結ぶ際の核のように，そのまわりにさらに多くの知識が集まり，部分的モデルと矛

盾がないかチェックできるのである．このプロセスはボトム・アップから作動する．次の段階では，発見された仮説を使って，仮定した知識ベースの，関連するセクションに進み，そこから，より細かい点が導き出され，それらが，低レベルの認識結果と共に，低レベルの画像データを再評価するのに使われる．これは即ち，認識がボトム・アップというよりトップ・ダウンの性質のモードに，あるいは，モデル・ベースドの認識に飛び込んだということである．このモードにあるということは，ボトム・アップのプロセスに戻ることを，永久に除外しなければならないというわけではなく，特定の画像領域では，それまでより良い核を供給し，その核が再びトップ・ダウンのプロセスを誘発するように，ボトム・アップのプロセスを再開しなければならないかもしれない．このフリップ・フロップ行動は，モデルと低レベルのデータ間に矛盾がない状態になるまで続くだろう．それまで，我々が画像に示された場面について得る知識は，完全なものではなく，場面の空間的スケッチを描かなければならない三次元的画面認識には，十分でない．これまでの結果を考慮にいれた上で，これからは，我々が予期できる箇所におきる，場面の奥行きによるゆがみの形跡を探してみよう．例えば，曲がった形状の物体に起こるかも知れない陰影の変化，テクスチャーの勾配，寸法の変化や，遠近法効果のような，深さのための計測のゆがみなどである．予期した箇所で，そのようなゆがみを見つけることは，画像の中に，三次元的物体の定性的記述が始まったことを示している．

前のパラグラフで述べたような定性的方法は，最初に場面に形状の記述をしたうえで，それらを物体の認識に使おうとするのではない．それよりむしろ，テクスチャーのような，低レベルの情報を使って，場面に含まれているいくつかの物体を，部分的に認識しようとし，物体の認識に十分に信頼性があると考えられる時（モデルと画像データ間に矛盾がない時）のみ，形状の定性的特徴が決定されるのである．その時点で，何らかの理由でもし必要とあれば，どこか大切な形状の量的分析が，ステレオ・プロセス等によって，次になされるかもしれない．このようなやり方には二層の長所がある．即ち，1）形状の正確な記述が，不可能でないとしても，求めるのが困難なため，避けられ，2）比較的低いレベルで物体を認識しようとすることにより，組み合わせの爆発が避けられるーというのは，実際の場面の高レベル描写は，無理のないプロセス時間で処理するには，たいてい複雑すぎるためである．この案の短所は，まだ十分に理解されていないという点だ．これは，制御構造に非常に頼っており，制御構造は，センサー読み取りデータ，抽出された特徴のデータ，そして，モデル・データによって作動する．このような環境に存在する制御構造は，何らかの方法で知識ベースのそれぞれの部分に連結されている，多くの異なった状態の変換規則を含み，分散しているはずである．特に，低レベル認識に用いられる規則は，アブダクティブ推論法の性質をもっており，その一方，トップ・ダウン・プロセスを制御する規則は，演繹的性質を多く備えている．そのような制御構造の詳細がどのようになっているかは，はっきりしていない．今一つの問題は，低レベルの認識，あるいは仮説を可能にする特徴を見いだすことである．先に，テクスチャーを例に出したが，多くの立証されているテクスチャー分類法は，陰影や変化する照明を含んだ三次元場面に応用するには，あまり頼りにならない．時には，与えられた画像解像度ではテクスチャーが弱すぎて役にたたず，何か他の低レベルの特徴でアブダクティブ認識を進めたく思うことがあるが，それは，どの特徴であろうか．他の心配は，場面の「定性的表現」に関するもので，それは研究課題として，あまり進んでいない．

今までに述べたこれらの方法が万一実現可能であるとわかれば，コンピュータ・ビジョンにおける数学の役割は，主に画像形成，ノイズ，幾何学的モデリング，特徴の抽出といった物理的レベルでのプロセスに限られ，それ以外の面では，知識表現，推論方式，様々な制御構造が深く研究されてきている人工知能が大きく関わってくるのである．