3. システム構築への具体的アプローチ

3-1 画像理解の研究は産業に役立つのか

池 内 克 史[†]

1. ま え が き

画像を扱う手法を研究する一分野に画像理解がある。人間の視覚機能の代替システム（あるいはロボットの目）を作ることを目的とする研究である。例えば、両眼立体視や距離画像計測装置のようなセンサを利用して、指定された物体を探し出したり、ロボットの位置を推定するようなシステムを作る研究である。

画像理解の研究成果を生産システムに使用すればよいという議論は古くからある。画像理解の研究が人間の視覚機能の代替を目的とするのであるから、画像理解システムは人間の代わりを果たすことができるはずである。したがって、画像理解システムを生産現場に持ち込むことにより、作業者数を減らすことができ生産価格の低減が図れる。計算機は人間のインスペクタのように疲労しないので、生産された製品の品質の均一化も期待できる。

しかしながら、一部の2次元処理を除いて画像理解の手法が生産システムに応用された例は少ないようである。これには各種の要因が考えられる。主なものとしては、①センサ計算機本体を含めた画像処理ハードウェアのコストが高価であったこと、②画像理解ソフトウェアの開発に携わるプログラマの数が少なく、かつ開発期間が長期にわたる、このため、ソフトウェアの開発コストが非常に大きかったこと、③画像理解の研究上、人間の視覚処理機能の代替ということがあまりに強調されすぎ、一般的ではあるが非効率なアルゴリズムが偏重され、このため、提案された多くのプログラムの実処理時間が長くなってしまう、実際の応用には向かない傾向があったことなどが主な理由だと

考えられる。

一方において、近年の計算機の処理能力の増加は目覚ましいものがある。100 MIPSを超えるワークステーションが珍しくなくなってきた。しかも IC テクノロジーの進歩により、これらのワークステーションやセンサ本体のコストが急激に低下してきた。

これらの外的要因に加えて、画像理解の研究においても、タスクオリエンティッドといった考え方が提唱されてきた。ビジョンモジュールの研究に際して、あまりに一般的に（その結果、非効率）に考えるのではなく、問題に即してその問題の特性を考慮してコンサイスなアルゴリズムを考えようとするものである¹⁾。この考え方の下に、各応用に応じたアルゴリズムの設計が可能になり、効率の良いプログラムが開発可能になってきた。この結果、高速の処理が期待できるようになってきた。さらに、タスクオリエンティッドの考え方に基づき、画像理解システムの設計法のシステム化が図られ、各種プログラムのシステムの生成法さらには自動生成の研究も進んで来た。このためプログラム作成の手間の軽減が図れ、開発期間、開発コストが減少してきた。本稿では、この新しいアプローチによるビジョン研究に焦点を当てて解説してみたい。

2. 物体認識プログラムの自動生成

物体認識とは、画像の中から特定の物体を探し出したり、物体の位置姿勢を決定したりするものである。この物体認識は、視覚システムの応用分野に大きな割合を占める。例えば、ベルトコンベア上を移動する物体を選別するシステムや、山積みされた物体を1つずつつかみ上げるシステムなどが代表的な応用例である。

これらの応用ではプログラムの高速性が要求される。例えば先の選別システムやビンピッキングシステムは、システムの開発目的が人間の作業者の代替を目指すものであるから、秒単位の処理時間しか許されな

[†] カーネギーメロン大学 コンピュータ科学部
 "Approaches for System Buildup; Is Image Understanding Research Useful to Industrial Vision" by Katsushi Ikeuchi (School of Computer Science, Carnegie Mellon University, Pennsylvania, USA)

い。こういった高速化のためには、物体認識を一般的な手法で解くのではなく、対象物体に即した特徴を有効に利用するアルゴリズムを開発する必要がある。一方において、こういった対象物体に即した物体に固有のプログラムをそのつど開発するのでは、非常に高価な開発費がかかってしまう。このため、認識プログラムを自動あるいは半自動で生成する技術が、タスクオリエンティドビジョンのためには必須である。

1970年代の初期、スタンフォード研究所(Stanford Research Institute(SRI))で2次元形状の認識プログラムを実際の画像より自動形成するSRIビジョンマシンが開発された²⁾。このシステムは、オフラインモード(学習モード)とオンラインモード(実行モード)の2つのモードをもつ。オフラインモードでは、まず閾値処理により各認識対象物体の輪郭線を抽出する。この輪郭線情報より認識に有用な面積、慣性モーメントといった約20ほどの幾何特徴量を計算する。この候補特徴のなかから各対象物体を識別するのに有用な特徴を選び出す。選び出された有用な特徴だけを用いて判別用の2分木を生成する。オンラインモードでは、システムは先に述べた2分木を用いて高速に物体を識別する。このシステムはきわめて短時間に物体認

識プログラムが自動生成され、非常に簡便であった。しかしながら、このシステムは2次元形状しか扱えないという制限が大きかった。

CMUでは、幾何モデルより3次元物体の認識プログラムを自動生成する研究を行っている^{3)~5)}。図1にその概念図を示す。1990年ごろビンピッキング用の物体認識プログラムを生成するシステムが完成した³⁾。ビンピッキング問題は以下の3つの好条件をそなえている。

- ・前提1: オクルージョンなし—ビンピッキングでは、ロボットが一度につかめる物体は1つしかないで、画像内のすべての物体を認識する必要はなく、山の頂上近くにある物体を1つ認識すればよい。このため、他の物体によるオクルージョンを考えなくてよい。
- ・前提2: 距離変化なし—山の概略の位置は既知なので、距離による物体の形状変化を考える必要がない。
- ・前提3: 物体識別なし—山はたいていの場合、同一物体からなっているので物体の識別は必要なく、位置姿勢の決定のみが問題となる。

こういったタスクの特性を利用して効率的な認識プ

図1 Vision Algorithm Compiler

ログラムを自動生成するシステム (Vision Algorithm Compiler (VAC) for bin picking tasks)が設計された。

3次元物体は観測者の位置ならびに観測方向に応じて無数の2次元形状を見せる。これらは見かけの形状がトポロジ的に同一な有限のグループに分類できる。この形状グループのことをアスペクトと呼ぶ。このアスペクトを利用して認識の処理を大きく2段階に分類できる。すなわち、まず未知の姿勢をアスペクトに分類し、次にアスペクト内で物体の正確な位置を決定する。この方針に従う認識プログラムをVACが生成する。生成過程は、アスペクトの生成、アスペクト分類部の生成、姿勢決定部の生成の3つに分かれる。

VACはアスペクト生成を実験的に行う³⁾。前提2より物体までの概略の距離は既知なので、姿勢による自由度のみを考慮する。物体が座標原点にあるとすると観測者は距離一定の球上に存在する。この球を小三角形を用いて分割する。各三角形の中心がサンプル点となる(初期のシステムでは80点)。前提1よりオクルージョンがないと仮定できるので、各サンプル点でCGを用いて80枚のオクルージョンのないサンプル画像を得る。各80枚の画像を領域分割し、トポロジを調べ、アスペクトを決定する。このとき、同時に面積や慣性モーメントといった各種の幾何特徴量も計算し、各アスペクトでの特徴量の変化を求める。

次に、VACはアスペクト分類部を生成する³⁾。各アスペクトでの特徴量の変化を見比べ、分類のための幾何特徴ならびに閾値を求め、SRIビジョンシステムと同様な2分判別木をルールベースシステムにより生成する。2分木の各終端ノードは1つのアスペクトに対応する。実行時にはこの2分木に従って特徴を比較していくことで、未知の領域がアスペクトに分類できる。

姿勢決定部の生成は、アスペクト内では見える面の組合せが既知であることと、前提1のオクルージョンがないことを利用して行われる⁴⁾。まず、各アスペクトで見える面のうち最大のものを選び出す。この面上に座標系を決定する方法を先の2分木の終端ノードに登録する。

このシステムは、他の物体によりオクルージョンのある物体の物体認識プログラムが生成できない。この欠点を補うため、現在開発中のシステムでは、アスペクト分類部の2分木をやめ、ネットワーク形式のものにしている⁵⁾。アスペクト生成の基本は同一であるが、サンプル回数を増やし、各特徴量の分布関数を得る。得られた画像を領域分割し、幾何特徴から各領域の各面への対応づけの確からしさを計算する。互いに隣りあう領域の対応づけから確からしさを増減し、最

図2 認識結果

終的なアスペクト分類を行う。最終的な位置姿勢決定は、各アスペクトごとに用意された3次元モデルを直接画像の上で動かす。図2に、このシステムにより生成されたプログラムによる認識結果を示す。

このシステムでは、以前のシステムに比較し、オクルージョンを許したため、事前に計算できる部分が減少し、扱える場合は拡大したものの、生成されたプログラムの効率は低下している。

3. センサプランニングの自動生成

与えられた問題に対して、適当なカメラの視点を考えたり、焦点距離やズームを変えたり、さらには照明条件を改良したりして効果的な画像を取り込むことは、画像理解システムの効率を向上させ、処理精度を向上させる上で非常に重要な要素となる。これまでは、多くの場合これらは、画像処理エキスパートの勘に頼ってアドホックに決められて来た。これを、タスクに応じて要求される画像の質が得られるように、自動的にセンサのパラメータを決定しようとするのが、センサプランニングと呼ばれる研究分野である。

研究の方向としては、カメラに関するものと照明に関するものに大分類できる。カメラに関するものとしては、①ある特徴を他の物体にオクルードされることなく観測できること、②ある特徴を必要な精度内で観測できることなどを目的とし、最適観測点、最適カメラパラメータ、などを決定する。照明に関するものと

図 3 オクルージョン回避問題

しては、①ある特徴がコントラスト良くとらえられること、②ある特徴の精度がある誤差範囲で求まること、などを目的とし最適照明点を決定する。

カメラに関するもので、オクルージョンを回避する問題（あるいは visibility 領域を決定する問題）については、電総研の坂根の研究がある⁶⁾（図 3 参照）。このアルゴリズムはロボットの作業監視を主目的として開発された。対象とする物体なりロボットの手が視野内で一定の大きさに撮影されるためには、物体からカメラまでの距離が一定である必要がある。したがって、可能な視点を全空間で探索する代わりに注視対象物体を中心とした仮想球面上を探索している。この球面を小三角形で分割する（図 3(a)左）。各三角形の中心と球の中心を結ぶ直線上で、この直線が物体表面と交差

するか周辺物体と交差するかでオクルージョンの有無を調べる。球面上でのオクルージョンのない三角形をつなぎあわせてオクルージョンのない領域を求める（図 3(a)右）。オクルージョンのない領域の中心方向を観測位置として出力するものである。図 3(b)にこの手法で計算されたハンドアイ作業中の観測可能領域の例を示す。コロンビア大学の Tarabanis も同様の問題を多面体間の干渉を調べる計算機幾何学的手法で解くアルゴリズムを提案している⁷⁾。

これらの手法では、対象を観測できるカメラ位置が決定できる。対象が見えるか否か以外にも考慮すべき要因は多い。SRI の Cowan は画像の精度、画像の視野、焦点距離、オクルージョンの 4 つを同時に考えてカメラの位置を求めるシステムを提案している⁸⁾。図 4 は Cowan のアルゴリズムを用いて対象物体のインスペクションに最適なカメラの位置を計算した例である。図 4 におけるタスクは、物体の底の部分にある穴の形状をチェックすることで、必要とされる精度は 1 mm が画像上で 1.5 pixel 以上に写ることである。レンズの焦点距離が既知とすると図 4(c)に示したような精度による拘束空間が求まる。近い物体は大きく写り、遠くの物体は小さく写る。したがって、カメラは一定距離内に存在する必要がある。図で拘束域が単純な球形になっていないのは、斜めから 2 つのエッジを眺めると間隔が狭くなり、正面から眺めたとき最大距離となるためである。レンズの焦点距離が与えられると、厳密には、ある一定の距離にある特徴しか画像として取り込めない。通常はこの距離の前後に、焦点深度と呼ばれる量に応じてピント良く画像が取り込める領域が存在する。この焦点深度、焦点距離によるカメラ位置の拘束を示したのが図 4(d)である。図 4(e)は先の 3 つの穴がオクルージョンなく見える領域である。これら 3 つの拘束を足しあわせて、可能なカメラの位置が図 4(f)のように求められている。

画像処理のためには、照明は欠かすことができない。与えられた物体上のエッジに対して、そのエッジがコントラスト良く写ることは画像処理にとって非常に重要である。これらの問題も SRI の Cowan⁹⁾ や電総研の坂根¹⁰⁾ らにより研究されている。たとえば、Cowan は乱反射モデルの下に各画素のエッジの両側での反射率の差に基づき、Sobel オペレータを掛けたとして、一定以上のコントラストが得られる光源の位置を求めている。

4. 視覚による組み立てプログラムの自動生成

少し変わった例として、本節では現在我々が CMU

(a) Bracket photograph.

(b) Bracket model.

(c) Resolution constraint region.

(d) Focus constraint region.

(e) Field-of-View constraint region.

(f) Viewpoints satisfying constraints (c), (d), and (e).

図 4 視点決定問題

図 5 APO システムのための面接触状態の分類

において行っている、視覚による組み立てプログラムの自動生成について紹介する。このシステム (Assembly Plan from Observation System, 略して APO システム) は、視覚をうまく利用することにより、人間の作業動作からロボットの組み立てプログラムを生成する¹¹⁾¹²⁾。

人間が行う組立作業を視覚システムで観察し、同じ作業を実行するロボットシステムでは、人間の行った動作が作業の中でどのような目的を持っているかなどの、動作の意味を抽出する必要がある。そのためには、行動理解の枠組みとなる作業モデルの構築が不可欠である。APO システムでは、組立作業を対象物どうしの接触を実現する動作ととらえ、目標となる面接触状態を運動に対する拘束に着目して分類し、作業対象物体の状態遷移とその状態間の遷移を起こす動作とを組合せたものとして組立作業の作業モデルを構築した。図5にガウス球面を用いて可能な運動方向と接触面の法線の関係を示す。接触面はその法線方向に対応したガウス球面上の点で表す。運動方向もその方向に対応したガウス球面上の点で表される。解析により、10の典型的な面接触状態が得られた。次に、各関係での可能な運動を考えることにより状態の遷移を得た。各可能な状態遷移に必要な動作テンプレートを割り振る。各テンプレートは、作業動作前の接触状態と作業動作後の接触状態、その状態間の遷移を実現する作業動作、動作実行に必要なパラメータから構成される。以上より作業モデルが作成された。図6(次ページ)に作業モデルならびに実時間処理における作業モデルの使用の様子を示す。

- (1) 人間がカメラおよびレンジファインダの前で1つの物体を初期状態から終了状態へと動かす。
- (2) 明るさの変化からシステムは何かが起こったことを知り、レンジファインダを起動する。
- (3) 人間の作業終了後レンジファインダが起動さ

れ、距離画像を取り込む。現在のシステムでは、この方式により、以前 (before)、途中 (during)、以後 (after) の3つの距離画像が得られる。

- (4) 以前と以後の距離画像の差分を取ることで、移動した物体の領域を切り出す。この移動領域の領域分割を行う。
 - (5) あらかじめ用意されている物体の幾何モデルとの照合により操作された物体の種類、位置・姿勢を決定する。これらの結果は幾何モデル内に現在のワールドモデルとして記述される。ここでは、移動した領域のみに注目しているため、認識が容易になり、高速化も図れる。
 - (6) ワールドモデルを利用して操作された物体と他の物体との面接触を抽出し、初期面接触状態と終了面接触状態を決定する。この例では、S-to-Dと決定された。ここでも、接触面は視覚からは観測できないにもかかわらず、幾何モデルを援用することで、見えない接触面が求められている。
 - (7) 面接触状態遷移に対応する作業モデルが取り出される。この例ではS-to-D作業モデルが取り出された。
 - (8) 途中と以後の距離画像から手の領域を抽出する。物体の位置が既知のため物体の近傍の領域のみを探す。この領域に楕円体をフィットさせる。この楕円体と物体の幾何モデルより把持点の情報を得る。
 - (9) S-to-D作業モデル内の組立動作テンプレートに必要な動作パラメータを幾何モデルから抽出する。
 - (10) マニピュレータにより同じ動作を実行する。
- これらのステップにより、例えば写真1のような組立作業を人間に習って実行することが可能となった。

写真1 ロボット作業

図 6 作業モデルを利用した行動認識例

5. む す び

本稿では、タスクオリエンティドビジョンの枠組みのなかで押し進められている各種のプログラムの自動生成法について紹介した。これらの手法の出現により、効率の良いプログラムが短期間に生成できるようになってきた。このため、低い開発コストで高速の画像理解システムが得られるようになってきた。これらの技法の進歩により、画像理解の研究結果が直接、産業用のビジョンに役にたつ日も近い。

(1993年10月25日受付)

〔参 考 文 献〕

- 1) 池内克史：“タスクオリエンティドビジョン”，信学誌，**73**，4，pp. 360-365 (1991)
- 2) G. J. Gleason and G. J. Agin：“A Modular Vision System for Sensor-Controlled Manipulation and Inspection”，Proc. of the 9th Intern. Symposium on Industrial Robots，pp. 57-70 (1979)
- 3) K. Ikeuchi and T. Kanade：“Automatic Generation of Object Recognition Program”，Proceedings of IEEE，**76**，8，pp. 1016-1035 (1988)
- 4) K. Ikeuchi and K. S. Hong：“Determining Linear Shape Change: Toward Automatic Generation of Object Recognition Programs”，CVGIP，**53**，2 (1991)
- 5) M. D. Wheeler and K. Ikeuchi：“Sensor Modeling, Markov Random Fields, and Robust Localization for Recognizing Partially Occluded Objects”，Proc. DARPA Image Understanding Workshop (1993)
- 6) 坂根，石井，柿倉：“ハンドアイ行動シミュレータ：HEAVEN システムにもとづく視覚センサのオクルージョン回避”，日本ロボット学誌，**4**，4，pp. 41-54 (1986)
- 7) K. Tarabanis, R. Y. Tsai, and P. K. Allen：“Satisfying the Resolution Constraint in the MVP Machine Vision Planning System”，Proc. DARPA Image Understanding Workshop (1990)
- 8) C. K. Cowan and P. D. Kovesi：“Automatic Sensor Placement for Vision Task Requirements”，IEEE Trans. PAMI (May 1988)
- 9) C. K. Cowan and A. Bergman：“Determining the Camera and Light Source Location for a Visual Task”，Proc. IEEE Conf. on Robotics and Automation (May 1990)
- 10) S. Sakane and T. Sato：“Automatic Planning of Light Source and Camera Placement for an Active Photometric Stereo System”，Proc. IEEE Conf. on Robotics and Automation，pp. 1080-1087 (1991)
- 11) K. Ikeuchi, T. Suehiro, P. Tanguy, M. Wheeler：“Assembly Plan from Observation”，Annual Research Review 1990, The Robotics Institute
- 12) K. Ikeuchi, M. Kawade, T. Suehiro：“Assembly Task Recognition with Planar, Curved and Mechanical Contacts”，Proc. IEEE Conf. on Robotics and Automation，2，pp. 688-694 (1993)

いけうち かつし
池内 克史 昭和48年、京都大学工学部機械工学科卒業。昭和53年、東京大学大学院情報工学専攻博士課程修了。マサチューセッツ工科大学人工知能研、電総研を経て、現在、カーネギーメロン大学計算機科学部教授(研究系)。人間の視覚機能、明るさ解析、物体認識、ハンドアイロボットの研究に従事。工学博士。