

解説

画像理解アーキテクチャ

大田 友一

筑波大学電子情報工学系 〒305 つくば市天王台 1-1-1

(1990年8月15日受理)

Architectures for Image Understanding

Yuichi OHTA

Institute of Information Sciences and Electronics, University of Tsukuba,
1-1-1, Tennodai, Tsukuba 305

1. はじめに

画像を手がかりとして、元の世界（通常は3次元）の記述を構築しようとする画像理解の研究は、1975年に開始された米国 DARPA (Defence Advanced Research Projects Agency) の画像理解プロジェクトの主要テーマとして推進された。

1970年代後半に Marr¹⁾ を中心として提唱された、2次元の画像情報から3次元の情景情報を復元する過程を計算理論的に明らかにしようとする動きが、1980年代前半に「画像理解の理論」と呼ばれる概念として確立され、以後、現在に至るまで shape-from-X をはじめとして多くのアルゴリズムが提案されてきた。ちなみに、それ以前には、画像に関する理論としては、2次元の画素配列を処理する際の幾何学的な扱いや計算量の検討など、「画像処理の理論」という概念しかなかったわけである。

いうまでもなく、画像理解研究の目標は、画像から得られる情報を通して、計算機が外界の状況を把握することを可能とする視覚システムの実現である。上記のアルゴリズム群は、基本的には、情景の一局面を捕らえている画像の情報を、情景の情報に変換するための、いわば「規則集」にすぎない。これらを部品として組み合わせ、画像理解のためのシステムを構築することが必要である。このような、画像理解システム構築のためのハードウェアやソフトウェアのアーキテクチャのことを、ここでは、「画像理解アーキテクチャ」と呼ぶ。

画像理解におけるシステム化研究の必要性は、1980年

代半ばから、ことあるごとに叫ばれてきたが²⁾、なかなか具体的な方向を見いだせない状態が続いた。DARPA では、1984年から自律走行車 (ALV) プロジェクト³⁾を開始し、これが、画像理解のシステム化研究に弾みをつけるものと期待されたが、現実に自動車を走らせるためには、今すぐに使える手法が必要、という観点から、視覚機能としては、レーザー光を用いたレンジファインダによる距離画像の利用が中心となった。画像の利用は、道路端の検出と追跡などに限られ、一部、画像理解的な考え方の応用はあるにしても、画像理解アルゴリズムのシステム化を主目標に掲げた研究はあまり進展しなかった。

最近、種々の並列計算機が実用機として利用可能となり、画像理解の研究グループにも徐々に普及するのに伴って、画像理解のシステムアーキテクチャにも、ようやく一つの方向が定まりつつあるように思える。すなわち、「情報の統合」、「アルゴリズムの統合」、および、それを実現的に実行可能にする「並列処理」がキーワードである。

2. 画像理解ベンチマークと並列計算機

2.1 並列計算機

画像処理を目的とした並列計算機アーキテクチャとして、従来、最も一般的であったのは、図1(a)に示したように、CPU とメモリを含むセルを2次元メッシュ状に結合したものであり、通常、各 CPU は中央の制御に従って SIMD 型の処理を行う。個々のセルを画素や隣接画素の集合に対応させることにより高い並列性が得ら

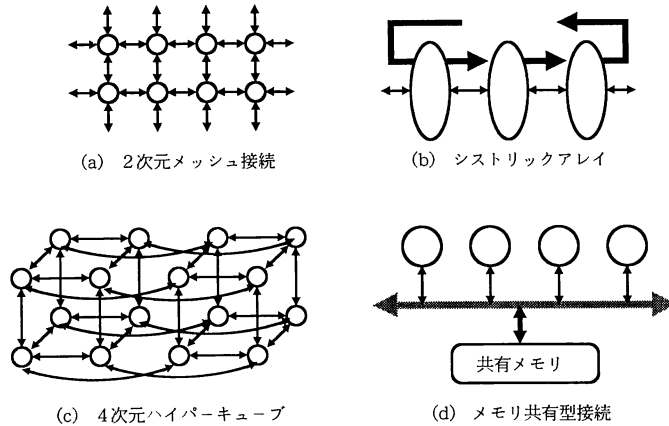


図1 並列計算機の構成

れる特長がある。しかし、構造上の制約から、その特性を発揮できるのは、比較的小さい範囲の近傍画素に基づく処理を、全画面に一律に施す場合に限られてしまう。この点では、ラインバッファを用いて局所並列演算を行うという、現在、画像処理専用マシンに広く採用されている方式と性能面で競合する部分が多いが、後者のほうが、コスト対効果比の面では優れている。

両者の中間的存在とも考えられるのが、(b)に示したようなストリッパレイであり、個々にプログラム可能なストリッパセルが、広いバンド幅を持つリンクで1次元状に結合されている。カーネギーメロン大学が開発した Warp マシンでは、パイプライン型の処理を行うほかに、セル間リンクのバンド幅の広さを利用して、全画像データを各セルにばらまき、画面分割型の処理を行うこともできる。

一方、画像理解では、画像の部分ごとに、異なった仮説に基づく演算を適用するという処理形態が必要であり、全画面に一律な操作を行うことを前提としたアーキテクチャでは対処が困難である。また、画像上で位置的に離れた部分を関連づけて処理する必要もあるが、近傍部分としかリンクを持たないアーキテクチャでは、この種の処理に対して効率が悪いという問題がある。

個々にプログラム可能なセルを、(c)に示したような d 次元の超立方体の頂点に配置し、隣接頂点のセルとの間にリンクを持たせたアーキテクチャを、ハイパーキューブマシンと呼んでいる。 d 次元のハイパーキューブマシンは、 2^d 個のセルを持ち、各セルは d 本のリンクを持つ。また、セルとセルの最大距離は、 $d-1$ であり、セル数の対数オーダーでしか増加しない。

図1(a)に示した2次元メッシュ接続の欠点を補うた

め、2次元メッシュによる接続に加えて、各セル間にハイパーキューブのリンクを設置することもできる。コネクションマシン (CM-1)⁴⁾は、この構成を持つ SIMD型のハイパーキューブマシンである。新バージョンの CM-2 では、16セルに1個の割合で浮動小数点演算器が付加されている。

上記の方式では、いずれも、各セルごとにプロセッサとメモリを持っているのに対して、(d)に示すように、共有の大きなメモリを持ち、複数のプロセッサをメモリを介して接続した、共有メモリ型のアーキテクチャも実用化されている。共有メモリ上に、大規模な共有データ構造や、画像データを置くことができ、タスク並列性を活かした処理に適している。実用機の例としては、Seqent の S81 などがある。この方式の欠点は、メモリバスのバンド幅がボトルネックになり並列度が上がりにくい点であるが、これを回避するため、プロセッサとメモリを多段階のスイッチングネットワークで結合した例もある。

2.2 画像理解ベンチマーク

従来、画像処理のベンチマークとしては、コンポリューションや、座標変換など、個別の処理を行わせて性能を測ろうとしていた。しかし、画像理解では、多様な処理を多段階に行う必要があり、個々の段階の処理の高速性ととも、各段階での処理の受渡し時のオーバーヘッドが問題となる。したがって、ベンチマークとしては、個別の処理の集合ではなく、画像理解における典型的な処理の流れに従って処理を結合したものが要求される。

DARPA では、現在利用可能な種々の並列計算機の画像理解への適合性を調査するため、以下のようなシナリオに従って、一連の画像処理を行うベンチマークを

1988年に設定した⁵⁾。

認識の対象とするのは、「長方形の紙片で造ったモビールを、真上から見た情景」を想定した合成画像であり、それに対応する距離画像も与えられているものとする。対象物体のモビールのモデルは与えられているものとし、画像中からモビールの位置や姿勢を抽出するのがタスクである。モデルと情景とは、平行移動と2次元的回転を含み、個々の紙片の位置関係も多少の変動を含む。画像サイズは、512×512画素で、濃淡画像は1画素8ビットの整数、距離画像は1画素32ビットの浮動小数で表現されている。

さらに、実際にプログラムを作成するプログラムの技量が、ベンチマークテストの結果に及ぼす影響を小さくするため、雑音除去のための前処理から、長方形紙片の検出と記述の作成、モデル中の紙片との照合、照合紙片の拡張・検証に至る処理手順を、40項目近くにわたって詳細に規定している。

表1は、各アーキテクチャの種々の処理への適合性を観るために、各部分の処理時間が全体に占める割合(%)を示したものである⁶⁾。参考のために、処理全体に対する相対速度比も示してある。Sun 3/160は通常アーキテクチャの代表、Sequent S81はメモリ共有型並列計算機(実験では8個のプロセッサを使用)、Warpはストリックマシン(10セル)である。また、IUAはマサチューセッツ大学で計画している図2に示すような3階層の並列処理系からなる画像理解用並列計算機⁷⁾で、表の結果はシミュレータによるものである。なお、コネクションマシンによる処理は、低レベル部分にしか適用されなかったため、表からは除外してある。

モデルとの照合の可能性を1個ずつ試行錯誤的に調べ

表1 処理ステップごとの処理時間の割合(%)と全体の相対速度比

処理内容	アーキテクチャ			
	SUN 3	IUA	S81	Warp
オーバヘッド	0.6	17.9	2.4	26.1
連結成分ラベル付け	3.5	0.1	5.0	6.1
長方形領域抽出	0.8	20.7	0.6	8.4
メディアンフィルタ	30.9	0.7	12.3	16.3
ソーベルフィルタ	17.0	3.5	6.0	0.7
長方形紙片の照合	3.1	15.6	1.2	0.6
モデル照合の拡張	40.9	38.6	64.2	37.8
結果の表示	3.1	2.9	8.3	4.0
処理全体の相対速度比	1	10,000	3	12

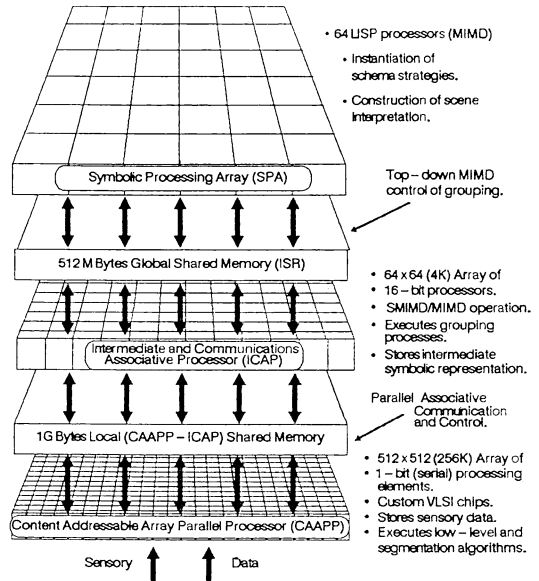


図2 マサチューセッツ大学の IUA⁷⁾

ていく必要のある「モデル照合の拡張」処理が、いずれのアーキテクチャの場合も、最大のボトルネックになっていることがわかる。

3. 画像理解における情報の統合

3.1 情報統合の必要性

画像情報を手がかりとして情景の情報を得ようとする画像理解過程では、画像から得られる情報の量と質が非常に貧弱であることが、問題の解決を困難にする大きな要因となっている。これは、

- 1) 画像に記録される情報が、種々の情景情報の複合したものであり、得たい情報がそのなかに埋もれている、
 - 2) 情景の3次元空間から、画像の2次元空間へ写像する過程で、奥行き情報が失われている、
 - 3) 背面に位置したり、他の物体による隠れのため、画像には全く記録されない情景情報がある、
 - 4) 情景中の雑音、撮像過程の雑音により、画像から抽出した情報にも、誤りや雑音が含まれる、
- などによる。

これに対処するための一つの方法が、複数の経路で得られた情報を組み合わせることにより、情景情報の量と質を向上させることであり、以下のような試みが行われている。

3.2 センサー融合と仮想センサー

センサー融合 (sensor fusion) とは、互いに異なった

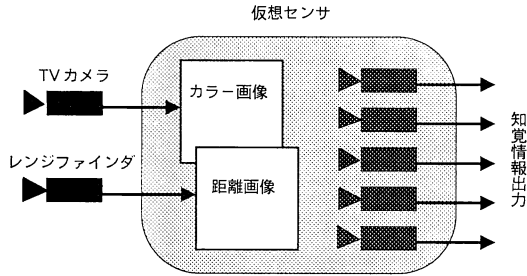


図3 仮想センサーの概念

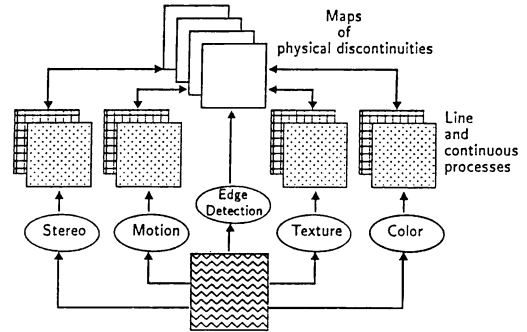
特徴を持つ複数のセンサーの出力を統合し、得られる画像特徴や情景特徴の量と質の向上を図ろうとする概念である。たとえば、カラー TV カメラとレーザーレンジファインダーを組合せ、色付き距離画像を得るなどの試みがある。融合により得られる効果という観点からは、大きくは競合的融合と相補的融合に分けられる。どちらかといえば、相補的な融合は情報の量、競合的な融合は情報の質を向上させる。色付き距離画像を得るのは、相補的融合の例である。

異種のセンサーではなく、同一のセンサーを異なった条件で適用して得た情報を統合することもできる。たとえば、ロボットが移動するための環境認識において、一つの地点からの観測では、十分な3次元情景情報が得られない。そこで、ロボットの移動に伴って一連の観測を行い、それを統合することにより環境モデルを順次更新する。

一方、図3に示すように、TV カメラなどの物理センサーと、それにより得られるデータを処理して、画像特徴や情景特徴に変換するアルゴリズムを組み合わせた系全体を、一種のセンサーと考え、仮想センサー (virtual sensor) と呼ぶ。物理センサーと同様、仮想センサーは、その動作に関して明確な仕様を持ち、上位の処理レベルからの要求に従って、所定の知覚情報を供給する。shape-from-X などの個々の画像理解アルゴリズムは、画像を入力とし、情景特徴を出力する仮想センサーと考えられる。仮想センサーの出力を統合することも、広い意味でのセンサー融合であろう。

3.3 MIT ビジョンマシン

MIT では、仮想センサーの統合システムとして、図4に示すようなシステムを構築している⁹⁾。アームにより可動なステレオカメラにより撮影された画像を入力とし、情景中の物理的エッジのマップを出力とする。入力画像は、エッジ抽出モジュール、奥行きマップを出力する STEREO モジュール、オプティカルフローを出力する MOTION モジュール、簡単なテクスチャ度のマッ

図4 MIT のビジョンマシン⁹⁾

プを出力する TEXTURE モジュール、物体表面の反射率を出力する COLOR モジュールに供給される。各モジュールはコネクションマシン上で実行される。各モジュールの出力に対して、ノイズ除去や内挿が行われるとともに、他のモジュールの出力と統合して、奥行き、面方向、正反射成分、表面反射率、影などの物理的エッジのマップを生成する。マップ内部の画素間の調整や、マップ間のデータの統合は、MRF (Markov random field) の枠組み⁸⁾ を利用している。

4. 画像理解アルゴリズムの統合

4.1 協調処理の必要性

画像理解を困難にするいま一つの要因は、対象としている情景に関する的確な状況把握が難しいことである。一般に、画像理解のアルゴリズムは、前述のような貧弱な情報をカバーするため、処理対象とする情景に関してかなり大きな制約を設け、その条件下でのみ所定の性能を発揮するように構成されている。したがって、画像理解システムが多様な情景に対応するためには、状況に応じて適切なアルゴリズムを自動的に選択することが必要となる。

処理状況に応じ、適用するアルゴリズムを柔軟に切り換えて制御する問題は、対象と処理アルゴリズムに関する知識、および、推論を含む高度の判断機構が相互に作用する問題としてとらえられる。しかし、画像理解では、画像から対象に関する正確な情報を抽出すること自体が容易ではなく、それに基づいてアルゴリズム切り換えの的確な判断を行うことは現実には非常に困難である。

あるアルゴリズムが、ある状況に適しているか否かは、そのアルゴリズムを実際に適用した結果に基づいて判断するのが最良であることは明白であり、それぞれ専門を異にする複数のアルゴリズムを互いに対等の立場で

並行して実行し、良好な結果を出力するアルゴリズムを動的に選択する方式が画像理解システムの枠組みとして有望である。また、前述のように、各アルゴリズムの出力結果を統合するだけでなく、処理過程において、互いの情報を融通しあい協調的に動作する枠組みをとることにより、いっそう、柔軟性の高い処理を実現できよう¹⁰⁾。

4.2 並行処理によるアルゴリズムの統合例

4.2.1 ステレオ視アルゴリズムの統合

適用範囲の広い柔軟なステレオ対応探索処理の実現のために、図5に示すように、異なる画像特徴を用いて対応探索を行う三つのステレオ対応探索アルゴリズム、すなわち、動的計画法を用いたエッジに基づく区間対応法、相関を用いた特徴点に基づく粗密対応探索法、連結エッジセグメントの対応探索法を統合する¹¹⁾。

個々のアルゴリズムを専門に実行するモジュールを一つずつ用意し、各モジュールが別々にかつ並行して処理を進める。各アルゴリズムの結果はモジュールごとに個別に格納し、通信によって互いの処理結果を参照する。このシステムは、メモリ共有型並列計算機 Sequent S81 上に構築され、実際に並列処理を行っている¹²⁾。

各モジュールは、他のモジュールの支援なしに解ける部分から処理を進めるので、画像のある局所部分に注目すれば、アルゴリズムの適用方法が、その局所部分の周囲の処理状況に適應して、自動的に決定されることになる。

三つのモジュールにおける探索結果を統合するためには、視差画像の生成を専門に実行するモジュールを設け、各モジュールが提供する視差情報と信頼度をもとに、統合視差マップを生成する。

4.2.2 並行トップダウン処理

顔画像から目や口などの位置を抽出する処理は、顔部品の配置がある程度決まっていることから、あらかじめ

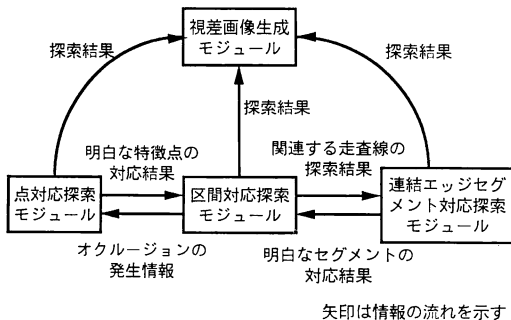


図5 ステレオ視アルゴリズムの協調的統合

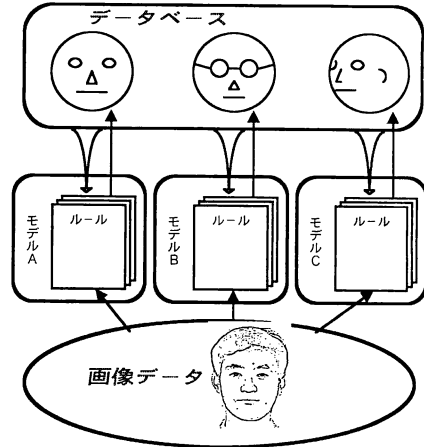


図6 並行トップダウン処理による顔解析

設定した顔のモデルにしたがって、トップダウン的に解析を進めることが有効なことはよく知られている。

しかし、入力顔画像が、向きに関して自由度を持つ、あるいは眼鏡や髭といった付属物をもつ可能性があるなどの多様性がある場合には、複数のモデルを設定し、適用すべきモデルを選択しつつ処理を進める必要がある。このとき、モデルの選択を行う判断に必要な情報を正確に画像から得るのは容易ではない。また選択したモデルが間違っただけであった場合、処理をどの段階までバックトラックすべきかといった試行錯誤的な制御を有効に実現することは現実には困難な場合が多い。

これらの問題点を解決するには、図6に示すように、見え方の異なる入力に対応した複数個のモデルを構成し、各モデルを並行に動作させる並行トップダウン処理方式が有効である¹³⁾。

各モデルは、互いに重複する部分も多いので、階層構造を持ったオブジェクトの集まりとして構成し、共有可能な部分結果は互いに利用しあうことにより、重複した処理を行うことを避ける。各モデルと入力画像との適合度を評価し、最良の処理結果を得たモデルを最終結果として選択する。

並行トップダウン処理方式は、最終結果に直接には関係しないような「無駄な」処理を許容することによって、画像理解システムの構築を容易にするものといえるが、VLSI 技術の著しい発展によるハードウェアコストの低下を見込めば、今後の画像理解システムアーキテクチャの一つの方向を示唆するものと考えられよう。

5. む す び

本稿では、「情報の統合」, 「アルゴリズムの統合」という観点から, 最近の画像理解システムのソフトウェアアーキテクチャについて述べ, それらの処理をサポートすべき並列計算機について概観した. 画像理解には多様な演算が必要であり, 単一のハードウェアアーキテクチャで対応することは困難と思われる. マサチューセッツ大学の IUA のように, 複数の並列計算機アーキテクチャを組合せる試みは興味深い.

文 献

- 1) D. Marr 著: ビジョン (1982), 乾, 安藤訳 (産業図書, 1987).
- 2) 大田友一, 金出武雄: “米国における画像理解研究”, 情報処理, **24** (1983) 1446-1452.
- 3) 金出武雄: “米国における自律移動ロボット研究の動向”, 日本ロボット学会誌, **5** (1987) 44-51.
- 4) W. D. Hills: *The Connection Machine* (The MIT Press, 1985).
- 5) C. C. Weems, *et al.*: “An integrated image understanding benchmark: Recognition of a 2.5D mobile,” Proc. of 1988 DARPA Image Understanding Workshop (1988) pp. 111-126.
- 6) C. C. Weems, *et al.*: “A report on the results of the DARPA integrated image understanding benchmark exercise,” Proc. of 1989 DARPA Image Understanding Workshop (1989) pp. 165-192.
- 7) C. C. Weems, *et al.*: *Int. J. Comput. Vision*, **2** (1989) 251-282.
- 8) S. Geman and D. Geman: “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans.*, **PAMI-6** (1984) 721-741.
- 9) T. Poggio, *et al.*: “The MIT vision machine,” Proc. of 1988 DARPA Image Understanding Workshop (1988) pp. 177-198.
- 10) 大田友一, ほか: “並列型画像理解—協調処理と並行トップダウン処理—”, 通信学会秋季全国大会論文集, SD-9-1 (1989).
- 11) M. Watanabe and Y. Ohta: “Cooperative integration of multiple stereo algorithms,” Proc. of 3rd ICCV (1990).
- 12) 渡辺正規, 大田友一, 池田克夫: “画像理解のための並列処理機構”, 電子情報通信学会論文誌, **J73-D-II** (1990) 820-829.
- 13) 角 保志, 大田友一: “並行トップダウン処理方式による顔画像の解析”, 情報処理学会研究報告, CV 61-3 (1989).