

UMASSにおけるVISION研究

Vision Research at UMASS

美濃 導彦
Michihiko MINOH

京都大学工学部
Faculty of Engineering, Kyoto University

あらまし 1987年3月から約1年間,UMASSへ留学する機会をえたので,そこで行われていた研究を紹介する.

Abstract Since I got a chance to stay at UMASS for one year, we introduce the research there.

1. はじめに

1987年3月から1988年3月までの約1年間,米国マサチューセッツ州立大学(略称UMASS)へ留学する機会をえた.UMASSは,ボストンの西約90マイル(162km)のコネチカット川(Conneticut river)のバレイに位置しているアムハースト(Amherst)という町にある.このバレイは,パイオニアバレイと呼ばれていることからわかるように,アメリカの中では歴史があり,古いものが好きなアメリカ人には観光地として人気がある.

UMASSは,1863年農業学校として設立され,以後色々な学部を開設して今では総合大学に発展している.医学部がボストンにあるので,UMASSがボストンにあると誤解しがちである.クラーク博士の出身校で,北海道大学と関係があり,北大から毎年2人の交換留学生が来ている.

UMASSのランキングは,町の人に言わせるとアメリカ全体の中で中の上に最近なっらしい.田舎にあるので,学生は全体的には,のんびりしている.しかし,ポリマーサイエンスの分野では世界一との評判があり,全世界から研究者が集まってきた.日本人はほとんどがポリマーの関係者であった.その中で,最近,COINS(Computer and Information Science)が注目されている.

UMASSのコンピュータビジョン研究室は,Prof. Riseman, Prof. Hanson と数人の研究者及

び約20名の大学院生,2名のプログラマ,数人のビクタにより構成されている.ここでは,この12年間,風景画像を認識・理解する画像理解システムVISIONSの開発に多くの努力がなされてきた.彼らは,領域分割のプログラムから出発し,初期のプロットタイプを6年ほど前に完成した[1].その後,個々のコンポーネントの改良と研究に必要なプログラム開発環境の構築,及びシステムの拡張とその応用に研究の重点を移してきた.現在は,ハイレベル処理の改良,ローレベル処理の並列プロセッサへのインプリメント,及び動画像処理の研究とそれをロボットへ応用する研究が中心となっている.

2. VISIONSの概要[2,3,4]

VISIONSは,屋外風景画像を認識・理解することを目標に開発された汎用的画像理解システムである.その構成を図1に示す.図からもわかるようにこのシステムは3つの階層(ローレベル・インタメディアイトレベル及びハイレベル)からなる.

ローレベル処理では,画像を領域分割したり,線抽出したりして画像からトークン(領域・線などの総称)を生成する.これらのトークンに対して,種々の特徴を計算し,ISR(Intermediate level Symbolic Representation)という名のデー

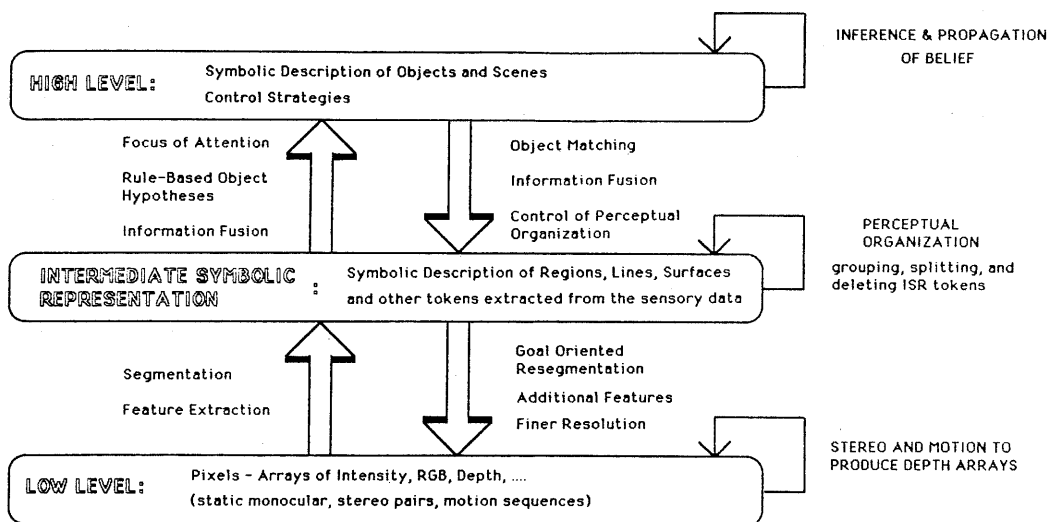


図1 VISIONSのシステム構成

データベースに蓄える。

インターメディアイトレベルの処理は、これらのトークンをハイレベル処理が扱い易いようにグループ化することで、ローレベル処理から得られる情報をハイレベル処理に引き渡すためのインタフェース的な処理を行う。

ハイレベル処理はインターメディアイト処理の結果を利用して、ISRデータベースにあるトークンにラベル付けを行う処理である。認識対象となる個々のオブジェクトに対して、スキーマと呼ばれる解釈ルーチンが並列に動作し、ブラックボードを介して通信する環境（スキーマシステム）を開発して実験している。

以下、まず計算機の構成について述べた後、それぞれの階層の概要を説明する。

2.1 計算機システムの構成

VISIONSのグループは、VAX750を4台、クラスタとして利用している。即ち、全てのディスクをデックネットにより共有し、端末はターミナルサーバを介して接続されているので、最初にどのVAXにログインするかを入力する。1台のVAXがネットワークサーバとなり、TCPIPも動いている。これをどうして、CSNET・ARPANETと結合されている。また、研究室には、6台のLISPマシン（TIのEXPLORER、それぞれ研究室でPhDをとった学生の名

前をホスト名にしている）があり、これらは、NFSによりVAXとディスクを共有する。

画像の入出力関係の装置は、ITVカメラが1台、ディスプレイ装置が4台、ディスプレイを写真にとる装置が1台だけである。この中には、グールドのIP8500も含まれている。実験に使う画像の多くは他の大学で入力されたもので、ネットワークを介して貰っている。彼らが持っているイメージのデータベースには、距離の画像もあるし、レンジデータもあるし、動画像もたくさんある。従って、この程度の入出力装置で十分なようである。

VAXのOSはVMSで、多くの人がこれで満足している。UNIXを使わないのは、メーカーがきちんとサポートしているOSの方が安心なこと、UNIX以前に多くのプログラムを蓄積していることなどが理由である。しかし、UNIXには関心があるので、UNIXの多くのユーティリティを彼らがVMS上に移植している。エディタは、VMS上のemacsを利用し、プログラム言語はLISP（Common LISP）とCで、後で述べるようにその中心はLISPである。文書の作成は、VMS上にインプリメントされているTEXを利用する。従って、彼らには、端末としてのパソコンは必要でなく、研究室にはVT100などのターミナルとワークステーションだけで十分である。

このようにハードウェアシステムの構成には何の特徴もない。それどころか、この構成では後

で述べるように、かなりの不都合な点が生じている。彼らにとっては、研究はアイデアであり、それが実証出来れば多少時間がかかっても平気で、システムの性能の改善、よりよいシステム作りはスタッフの仕事と割り切っている。

2.2 ローレベル処理

ローレベル処理の大きな役割は、画像からトークンを生成することで、領域分割や線抽出などの画像処理を行う。画像処理のアルゴリズムを開発するための環境としてローレベルビジョンシステム (Low Level Vision System: LLVS) [7, 8, 9] がある。

LLVSは、VAXのVMS上でLISPを利用して画像処理を行えるように開発されたシステムである。LLVS内では、全ての画像はPLANEとして統一的に扱われる。LLVSの主な機能は、①PLANEの入出力、及び表示、②イメージオペレータの開発支援 (LISPとCの結合の部分がLISPのマクロで書かれているので、その辺りの詳細を知らなくてもCと結合が出来る) ③画像をアクセスするC関数の提供、④イメージオペレータの形式の統一とそれらの蓄積、⑤ヘルプ記述の提供とマニュアルの完備、等である。TVカメラから入力した画像やMTにある画像などは、簡単にLLVSの形式に変換できる。また、ゲールドIP8500形式の画像も簡単に取り込める。

現在、領域分割としては、ローカルヒストグラムを用いる手法を利用している [6, 10]。画像を16x16画素の正方形の部分画像に規則的に分割し、各部分画像に対してヒストグラムの谷を検出する方法で領域分割を施す。この時、隣接の部分画像の情報をうまく利用して分割による疑似輪郭の除去と分割線にまたがる領域の検出を行っている。

線抽出としては、おもに直線を抽出する手法を2種類利用している。ラプラシアンゼロクロス点のエッジの方向と強さを基にヒューリスティックスを用いて結合して直線を得る方法 [11] とグラディエントの傾きの方向が揃っているところで一気に直線を抽出する方法 [5] の2つである。

現在は、もう画像処理のアルゴリズムの研究をしている学生はいない。このレベルの問題はトップダウン処理を取り込むことで、PhDの学生が

システムを作成して、実験している。彼は、LLVS以前のシステムを利用しているので、多くの努力がシステム間の整合をとることに費やされた。

ハイレベル処理では、VAXの処理速度の問題もあり、各画像に対して1つの理想的な処理結果 (パラメータをその画像に合わせてチューニングしたもの) を用いて、研究を行なっている。

ロボティックスのグループがローレベル処理をIP8500を用いて実時間でやりたいようだが、いまのところUMASSで作成しているハードウェアを待ちながら研究している。

2.3 インターメディアイトレベルシンボリックレプレゼンテーション (Intermediate-level Symbolic Representation - ISR)

ISRは、ラベル画像又は線抽出結果の画像から作成されるリレーショナルデータベースである。個々の領域・線をそれぞれ1つのトークンとし、全ての特微量はこのトークンに対して計算する。領域の場合、各トークンはその領域が占める範囲もその特徴として持っている。この情報を利用して、いろいろな特徴を動的に計算するメカニズムが提供されている。これらの機能はLISPのパッケージとしてLLVSのもとで実現されている。現在、種々の特徴を計算するプログラムが用意されている。

さらに、これらの特徴をキーとしてこのデータベースをアクセスする機能も備わっている。例えば、ある特微量の値がある範囲内にあるトークンを得ると言うような機能である。これらの機能は、ハイレベルシステムにより利用され、現在注目している領域がどのオブジェクトに属するかを決定する。このようにISRはハイレベルとローレベルのインターフェースとして機能する。

インターメディアイトレベルの処理は、トークンをグループ化する事 (Perceptual Organization) である。まとまりのよい領域を抽出したり、平行線、直角、共線、円弧などを抽出して新しいトークンとしてISRデータベースに登録する [12, 13]。また、統計的知識に基づいて、例えば領域トークンの”道路らしさ”を評価して、一つの特微量としてISRに付加する処理も考えられる [14, 15]。一般的には、これらの処理は、ジェネラルな (画像に依存しな

い) 知識を利用するボトムアップ処理である。この結果、1枚の画像から千個程度生成されたトークンは、数千個に増加する。

2.4 ハイレベル処理

スキーマとは、物体の認識に必要な宣言的知識と手続き的知識を同時に表現する枠組みでフレームと同じ概念である。ハイレベルシステムでは、各物体を解釈する多くのスキーマがそれぞれ通信しながら全体として矛盾のない解釈を生成する。各スキーマは、複数の知識源を実行するストラテジを持っている。手続き的知識はストラテジの中に効率的に実現される。物体に関する知識は、マップ関数の形で各ストラテジの中に蓄積されている。

スキーマの作成を支援する環境がスキーマシステムまたはスキーマシェルと呼ばれる[16]。このシステムはLISPマシン上に実現され、スキーマの生成・削除及びスキーマ間の通信をサポートする。スキーマ間の通信は黒板を用いて行う。黒板に対して各スキーマは自由にセクションを設定し、メッセージを書き込む。書かれたメッセージは、各スキーマが自由に読み出すことが出来るが、必要なものだけを選択するために関数が設定でき、条件に合うものだけを選択する機能があり、効率を高めている。さらに、各スキーマは、その内部に局所的な黒板を持っている。構成は大局的黒板と同じであり、同一のスキーマに含まれるストラテジの相互通信に使われる。

スキーマシステムは、並列処理を原則としているので、現在はTIのLISPマシン上で並列処理をシミュレーションして実現されている。このためか、オーバーヘッドも大きく、現在のところ、1枚の画像のISRが与えられて、それを解釈するのに数時間かかっている。また、現在はISRをVAXのLLVSで作成し、それをネットワークを介してLISPマシンにロードしてハイレベル処理を行っている。これは、TIのLISPマシンが純粋にLISPしか実行できないためである。ISRのアクセス関数もこれに対応してLLVS用とLISPマシン用の2種類がある。このため、画像の認識処理を最初から通して実現することは、現在のシステムではほとんど不可能である。

3. VISIONSの関連研究

1枚の画像の解釈は、それ自体が興味深い研究であるが、実用的には3次元データや動画画像が扱えることが望ましい。現在は、VISIONSのシステムを3次元を扱えるように拡張したり、動画画像処理に応用したりすることに重点が移ってきている。同時にシステムをハードウェア化するプロジェクトも進んでいる。以下、それぞれについて簡単に説明する。

(1) 3次元データの獲得とその利用

3次元のデータ、主に視点からの相対距離と物体表面の法線ベクトルは、いろいろな手法で得られる。ここでは、動画画像から距離を求める手法、ステレオ画像による手法、輪郭線から法線ベクトルを計算する方法が研究されている。これらは、独立に研究されており、現在のところ、VISIONSシステムとの関係はない。距離を計算する目的が物体を認識することであるならば、早い段階から関係を持たせる方が面白い。3次元データをどう利用するかについては、現在は何も研究されていない。3次元モデルを視点解析によりビューモデルに変換し、いわゆるインタープリテーションツリーをつくって画像の特徴とマッチングさせる研究も行われている。

(2) 動画画像処理

動画画像処理の研究は、ロボットのナビゲーションと関連して進められている。センサーの動きが平行移動と仮定して、複数のフレームよりFOE (Focus of Expansion)を計算し、距離を得る手法は、CMUのプロジェクトで利用されている。また、センサーの動きと同時に複数の独立に動いている物体がある環境で得られたオプティカルフローを解釈する研究もある[17]。最近では、実用化を目指した高速化の研究も進められている。例えば、画像から選択した特徴点(線・特異点など)をフレーム間に対応づけるときに、フローベクトルを利用して、注目している点の存在位置を予測し、ある程度のエリアを設定してその中だけを探索することにより、効率を上げる研究がある。動画画像処理においても、早い段階からVISIONSシステムと結合することが望ましく、今後の研究は動画画像処理が中心になりそうである。

(3) アーキテクチャの研究

VISIONSを高速化するためには、新しいアー

キテクチャの開発が必要である。特に, VISIONS は, これまでに述べたように3つの階層からなり, 各階層における処理はそれぞれ大変異なっている。従って, アーキテクチャもこれに合わせて構成する必要がある (図2参照)。

ローレベルは, 画素に対応する512X512の1bitALUからなり, Content Addressable Array Parallel Processor (CAAPP) と呼ばれる。この階層は, SIMDで動作し, カスタムLSIとして設計されている。現在は, 初期バージョンのLSIが出来たところで, デバッグ中である。

インターメディアイトレベルは, 64X64のSIMDアレイでTIのシグナルプロセッサを用いて設計が進められている。このレベルでは, トークンのグルーピング処理をその主な目的としている。1つのプロセッサは, ローレベルの8X8のプロセッサとダイレクトに通信が出来る。現在, 詳細設計が終わったところで, ボードが出来るまでにまだ時間がかかりそうである。Intermediate and Communications

Associative Processor (ICAP) と呼ばれている。シミュレータがLISPマシン上にあって, パラレルアルゴリズムの研究も同時に進められている。しかし, まだ設計が変更になることが多く, 学生さんが嘆いている。

ハイレベルは, 64個のLISP関数を実行する32bitのプロセッサ(68020等を使う予定)からなり, 主にスキーマシステムを効率的に実行する。現在は, 何も手が着けられていない状態でどんな設計になるかも未定である。

プロトタイプとして, 1-8-64X64をまず動かそうという計画である。民間企業と共同で行っているが, お金の問題もあり, まだまだ時間がかかりそうである。プロトタイプでどの程度の性能が出せるかが問題で, 世の中のハードウェアの性能改善の速度を考えると, 構想全体のシステムが実際に作成されるかどうかは疑問である。

4. VISIONSの問題点

VISIONSのシステムは, 長い年月をかけて作成されてきた大変よく考えられたシステムである。改良に改良を重ねてきた結果, 現在のような形に落ち着いている。ここでは, このシステムや研究についての私見を述べる。

◎ システム全体にもう少しトップダウン的

University of Massachusetts Image Understanding Architecture

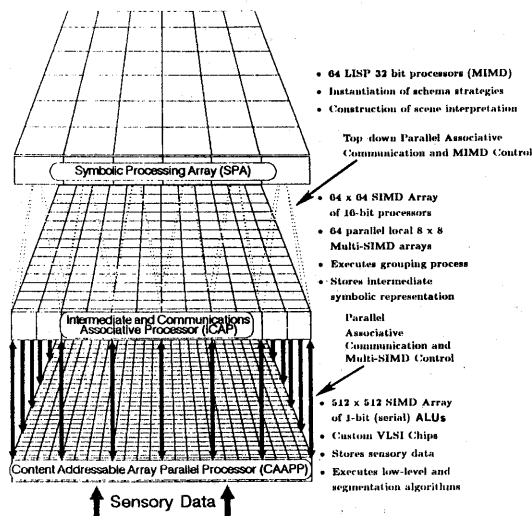


図2 UMASS画像理解アーキテクチャ

な処理を強化するべきである。トークンの解釈の段階では, このような階層構成をとっているシステムでは, トップダウン処理は比較的实现し易いが, 問題はトークンの生成の段階にいかにつトップダウン処理を組み込むかである。トークンは, 画像から何等かの処理により生成されるが, 基本的にはそれほど信頼性の高いものではない。特に領域トークンの場合はトークンのグループ化だけでなく, 分割や修正も必要である。この時に, 重要な点はハイレベルシステムからどのような情報がトップダウン処理のために得られるかと言う点である。

◎ 3次元データや動画像をVISIONSシステムの枠組みで扱うことは困難で系統的に何等かの拡張が必要である。

◎ 画像を入力してから結果を得るまでを通して実験できる環境が必要である。この種のシステムは, 出来るだけ多くのデータに対して実験することが大切である。そのためには, 新しい画像を入力してから結果を得るまでが自動的に実行される環境が必要である。ローレベルシステムが質の悪い結果を出したときに, スキーマシステムが

とのような振舞いをするか、興味深い。また、このような実験では、初期の領域分割のパラメータ設定をどのようにすればよいかが大きな問題になる可能性がある。

◎ スキーマシステムは、現在、形、色などの2次元情報のみを用いて画像を解釈している。これを、何等かの形で3次元の情報が扱えるようにしたとき、スキーマが現在のように比較的シンプルに書けるかどうかの問題である。

◎ スキーマは将来的には自動プログラミングできるようにすべきである。スキーマは、各物体毎に作成しなければならない。これを人間がしていたのでは、到底大きなシステムは作れない。最低限必要なものを人間が与えて、それをもとに自動生成することが今後必要であろう。

◎ 同じ研究を10年以上も研究室全体で続けていることはすばらしい。しかし、その弊害もある。例えば、システムの的に問題が出てきても歴史的な重み（ソフトウェアが多く蓄えられている）があり、なかなか改善されないことや、ドキュメントが完備されていないために、多くのノウハウが存在することである。UMASSでは、この種の弊害を極力なくすようスタッフが協力している。

◎ 日本に比べて、数学的な理論を重視する態度を学生に徹底的に身に付けさせようと言う努力がなされている。UMASSは、特に数学関係の著名な人がいるのでその雰囲気は強い。

[謝辞] 今回の出張に関しては、坂井利之教授（現龍谷大学）をはじめ、金出武雄教授（CMU）、白井良明教授（大阪大学）にたいへんお世話になりました。心より感謝致します。また、出張を快く許可して下さいました京都大学工学部情報工学教室の矢島教授をはじめ職員の方々に御礼を申し上げます。

[参考文献]

[1] Terry E. Weymouth; Using Object Description In a Schema Network For Machine Vision, COINS Technical Report 86-24

[2] Allen R. Hanson, Edward M. Riseman; A Methodology for the Development of General Knowledge-Based Vision System, COINS Technical Report 86-27

[3] Allen R. Hanson, Edward M. Riseman; The VISIONS Image Understanding System -1986, COINS Technical Report 86-62

[4] Allen R. Hanson, Edward M. Riseman; Summary of Progress in Image Understanding Research at the University of Massachusetts, COINS Technical Report 87-20

[5] J. Brian Burns, Allen R. Hanson, Edward M. Riseman; Extracting Straight Lines, COINS Technical Report 84-29

[6] Ralf R. Kohler; Integrating Non-Semantic Knowledge into Image Segmentation Processes, COINS Technical Report 84-04

[7] James H. Burrill; Low Level Vision System, COINS Technical Report 87-14

[8] Robert Heller; Show-Plane and Friends Users Manual, COINS Technical Report 87-15

[9] James H. Burrill, Robert Heller; Using LLVS under VMS, COINS Technical Report 87-41

[10] J. Ross Beveridge, Joey Griffith, Ralf Kohler, Allen R. Hanson, Edward M. Riseman; Segmenting Images Using Localized Histograms and region Merging, COINS Technical Report 87-20

[11] Michael Boldt, Recharad Weiss; Token-Based Extraction of Straight Lines, COINS Technical Report 87-104

[12] John Dolan, George Reynolds, Les Kitchen; Piecewise Circular Description of Image Curves Using Constancy of Grey-level Curvature, COINS Technical Report 86-33

[13] George Reynolds, J. Ross. Beveridge; Searching for Geometric Structure in Images of Natural Scenes, COINS Technical Report 87-03

[14] George Reynolds, Deborah Strahman, Nancy Lehrer, Les Kitchen; Plausible Reasoning and the Theory of Evidence, COINS Technical Report 86-11

[15] Nancy Lehrer, George Reynolds, Joey Griffith; A Method for Initial Hypothesis Formation in Image Understanding, COINS Technical Report 87-04

[16] Bruce A. Draper, Robert T. Collins John Brolio, Joey Griffith, Allen R. Hanson, Edward M. Riseman; Tools and Experiments in the Knowledge Directed Interpretation of Road Scenes, COINS Technical Report 87-05

[17] Gilad Adiv; Interpreting Optical Flow, COINS Technical Report 85-35