

第5回人工知能国際会議におけるビジョン部門

白井 良明

(電子技術総合研究所)

1. はじめに

IJCAI-77は8月22日~24日、MITで行われた。一般講演は表1のような11部門に分け、約4セッション並行して発表された。すでに自然言語は長尾、知識の表現と定理の証明は古川から本研究会で紹介された。

今回は応募論文の数が多かったため、約半数を短論文とし、発表時間も半分の15分となった。表1からわかるように、ビジョンはとくに短論文の割合が多かった。

ビジョンは9セッションにわたって発表されたが、筆者はその一部に出席する機会を得た。本会議の前に8月16日~19日、MITでコンピュータビジョンに関する日米セミナーを行ったが、そこで得た知識なども参考にし、ビジョン部門の概要と、適当な論文を紹介する。

表1. 部門別論文数

部門	長論文	短論文
自然言語	2/	2/
知識の表現	13	12
知識の獲得	8	9
プログラム援助	6	6
問題解決と探索	10	10
定理の証明	8	6
ビジョン	16	25
ロボット	6	8
特殊システム	10	7
応用	6	4
言語とシステム	4	4

2. 両眼立体視

米国ではスタンフォード大学の Binford の所で両眼立体視を研究しているという事になっているが、下記の論文ではじめて明らかになった。

D. B. Gennery: A Stereo Vision System for an Autonomous Vehicle.

これはTVカメラを乗せて動きまわる車のための両眼立体視で、高分解能相関法を述べている。2つの画像の一方の点を推定し、その点を中心とした小さな正方形領域に対応する他方の領域を求める問題である。ここでは、全く同じ明るさの2つの領域が理想的には存在し、実際にはそれに独立なガウスノイズが加わっているモデルを考えている。指定された点に対応する領域の中心を (x_m, y_m) とし、両領域の明るさの差を、領域内の各点の明るさの差を要素とするベクトル ϵ で表わす。 (x_c, y_c) を (x_m, y_m) の候補点とし、 $P(x_m, y_m = x_c, y_c | \epsilon)$ という確率を求める。これはベイズの定理から得られる。すべての候補点について確率を求め、

$$x_0 = \frac{\sum P(x_m, y_m = x_c, y_c | \epsilon) x_c}{\sum P(x_m, y_m = x_c, y_c | \epsilon)} \quad y_0 = \frac{\sum P(x_m, y_m = x_c, y_c | \epsilon) y_c}{\sum P(x_m, y_m = x_c, y_c | \epsilon)}$$

で対応点を決定する。従来は領域間の明るさの相関を求め、そのピークを対応点としていた。相関と上の確率は単調な関係にある。したがって従来の方法では確率のピークを求めたが、新しい方法では確率分布の重心を求めることになる。

この手法は数学的にはきちんとしているが、両画像の見る位置の差による歪みを無視していることに大きな問題がある。

3. 探索の Locusモデル

カーネギーメロン大学では、音声理解システムプロジェクトを行っていたが、そこで用いた探索の手法を、新しく始めた画像理解システムに適用しようとしている。この手法はマルチバンド写真の領域分けに適している。すなわち各画素がどの領域に属するかを決める場合、画素の持つマルチバンドの値からそれが各領域となる尤度が計算でき、いっほう大局的に領域が満たすべき位置関係が与えられている場合である。例之び図1.のような領域A、B、Cの間の位置関係が与えられている、画素は2×4の場合を考える。

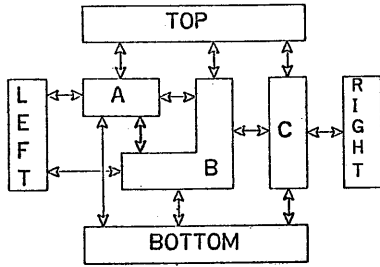


図1. 領域の位置関係

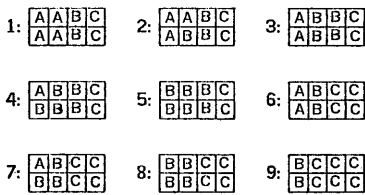


図2. 可能な解釈

可能な解釈は図2の9通りに限られる。これを各画素をノードとする隣接関係のグラフで表わせば図3となる。いっほう各画素の位置によらない尤度が与えられているので、これを用いて適切な解釈を行おうとするのである。ここではテレビの順次走査と同じ順序で画素の解釈の確からしさを計算する。もしそれが0.5以下であれば、探索を中止する。このようにして図4のような探索が行われる。解釈を行うためには、この探索を逆にたどりながら最も確かな解釈を採用していく。この結果図2の4番目が採用される。この手法は探索において常に一定の確からしさ以上の候補を保持しながら他を捨てていくことにより、探索の数が増大するのを防ごうとしている。音声理解の場合にもいえることであるが、可能な解釈の数が限られている場合にのみ有効である。また音声の場合には、探索の順序が時間によって一々に決められたが二次元画像では果して順次走走方式がよいか議論の中心となった。

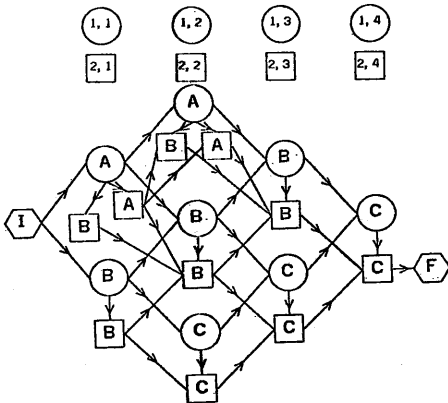


図3. 可能な解釈のグラフ表示

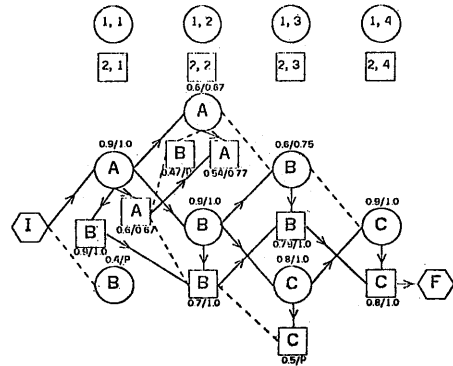


図4. 探索の経路(数字は尤度)

4. 曲面体の認識

3方向から見た曲面体の線画を作、て認識する研究を述べる。

R. Shapira and H. Freeman: Reconstuction of Curved-Surface Bodies from a Set of Imperfect Projections.

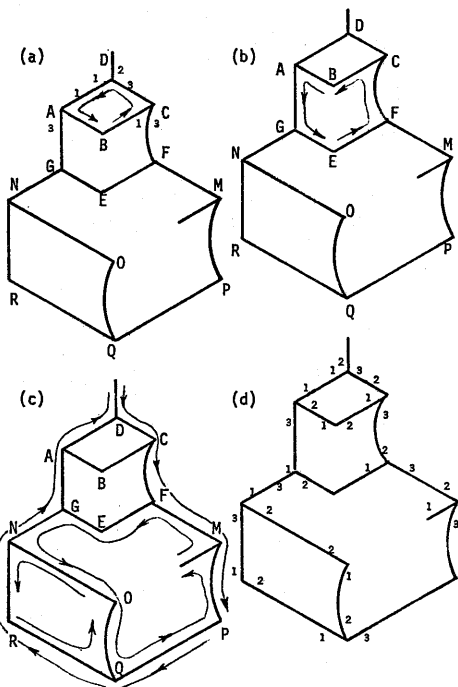


図5. 線列の順序

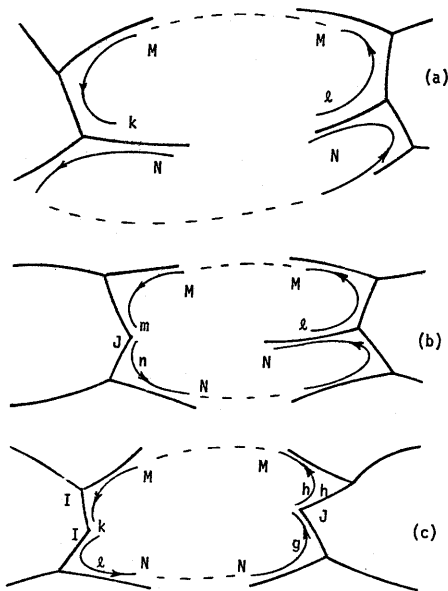


図6. 欠けている線の挿入

線画にはすべての頂点が誤りなく存在し、欠けている線はあってもよいが余分な線がないものとする。物体の頂点には必ず3つの面が交わっていて、見る位置は一般位置(少しの変位で線画の頂点の形が大きく変わらない)としている。一つの面を囲む線に反時計まわりに矢印をつければ、一つの線はその両側で反対方向の矢印がつけられる。図5.(a)のDのように3つの線が交わる頂点では、他の部分とは関係なく矢印をつけることができる。ところが頂点Bは、線の数が2つか3つか確定していないので矢印が決まらない。しかしBCDABという順序は定まっている。つまり閉じた面が形成されている。そこでBではABCという順序を決定することができる。Bは3つの面が交わる頂点であるから、ABCという順を持つ面があれば、CBAという順序を持つ面は存在しない。したがって図5.(b)のように2つの面が存在することがわかる。つぎに頂点Eに着目すれば、Eの上側にはすでに2つの面があるので、FEGという順序は面を形成することがわかる。このような法則を用いれば、図5の線画はあり得ないことを証明できる。まずRでQRNという順序を仮定すれば、図5(c)の矢印が決まる。これは頂点Mから出る不完全な線を2回通ることになり矛盾が生じる。つぎにNRQという順序を仮定すれば、頂点における矢印は図5.(d)で表わされる(数字は図5(a)を参照)。頂点Pではどちら側が面であると仮定しても矛盾をきたす。

このような推論を用いながら不完全な3つの線画の頂点や線の対応づけを行おうとするのである。一つの線画の頂点は他の線画の一定の線上のどこかの点と対応する。その線上にある頂点が対応頂点の候補となる。はじめに2つの絵ごとに対応頂点の候補を見つけしておく。まず3つの絵のすべてで対応のとれる頂点を対応づける。つぎに2つの絵を用いて線の対応をとる。これはすでに対応が確定

した頂点を手がかりにして、それにつながっている線の対応を一定のヒューリスティックな法則により決定する。最後に物体の面の対応を求める。ここでも前述の法則を用いて探索の範囲をせばめている。また図6のように欠けている線を補うこともできる(2つの面の存在が確定し、その間に線がなければ補う)。

以上の手法は拘束条件をうまく利用し、かなり一般性もあるが、実用的には曲面物体の頂点がすべて求められているという前提を満たすことが困難であろう。

5. システム

汎用ビジョンシステムを目標としている2つの研究を紹介する。はじめはマサチューセッツ大学で研究しているVISIONSシステムで、以下の3つの論文がそれを扱っている。著者はプロジェクトリーダーのE. Risemanと学生達およびハンアシャー大学のA. Hansonのグループである。

J. Prager et al. : I. Segmentation Process in the VISION System.

T. Williams et al. : II. Model-Building in the VISION System.

K. Konolige et al. : III. Between Region and Objects..... Surface and Volume.

システムは図7の概念図のように低レベル処理と高レベル処理が分けていて、前者では並列処理を何回もくり返して特徴抽出を行い、領域、境界線、境界線の端点(RSE構造)をインターフェイスにして後者と交信する。Iでは弛緩法によって屋外風景の境界を出し、それとは別に領域法によって得た領域分けの結果の共通な境界線を求めることによって、ノイズに強い線画を作っている。この手法は文献E. R. Riseman and M. A. Abib

: Computational Techniques in the Visual Segmentation, Computer Graphics and Image Processing, 6, P221 (1977).

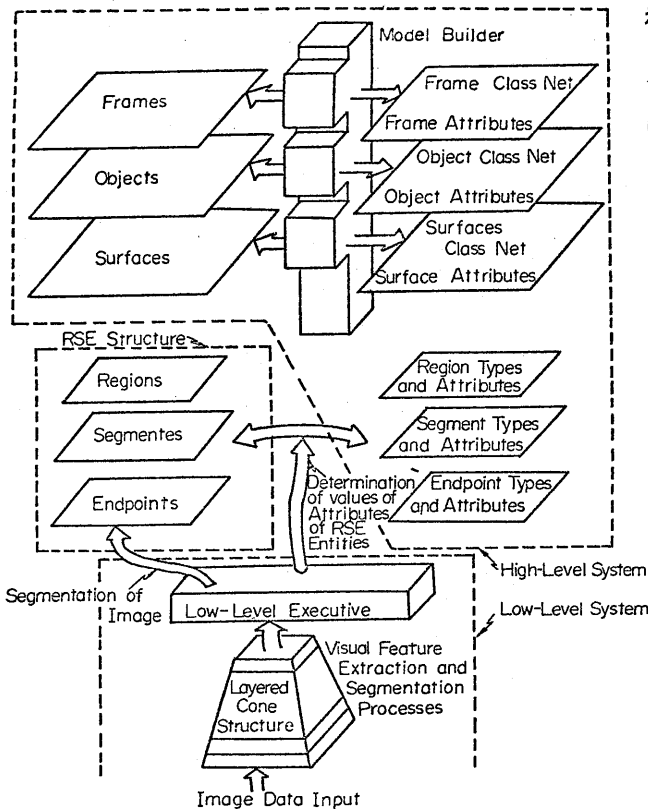


図7. VISIONの概念図

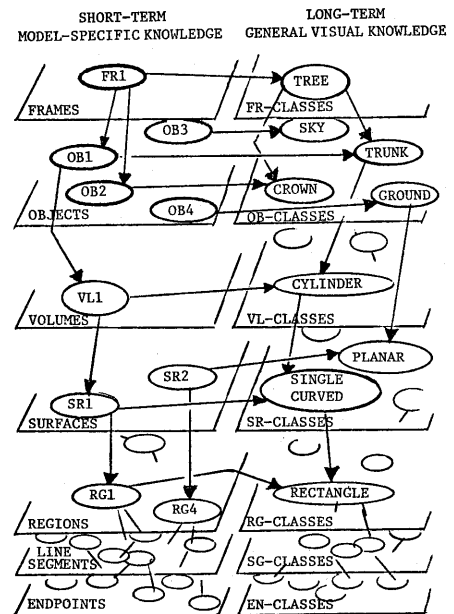
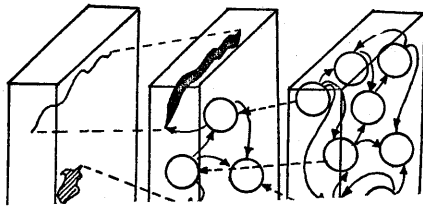


図8. モデル作成の例

に詳しく述べてある。ⅡとⅢは高レベル処理を述べている。図7の上部右がシステムの持つ知識で、図8の長期記憶に相当する。それは階層構造をしている。画像が与えられると各階層の物に対応する部分を見つけにくい。これをモデル作成と呼んでいる。たとえば図8の右側に 木は幹と樹冠から成り、幹は円筒で、円筒は単一曲面で、それは四角な領域となるという知識がある。四角な領域PGIが、右の四角に対応すると仮定してVL1, OB1, FR1.という対応を作る。それを確かめるためには樹冠の存在を知ればよく、今度は右のCROWN から下へたどりながら対応づけを行なっていく。この手法で 木、家、道などの重なって見える10個程の領域を解釈した。

次にロチェスター大学で行っている研究を紹介する。

D.H. Ballard, C. M. Brown and J. A. Feldman: An Approach to Knowledge-Directed Image Analysis
このシステムはセマンティックネットワークで表わしたモデルを用いて画像を理解していく。モデルは記号で表わされた要素間の関係であり、画像はこれとは異なる構造であるため両者の中間に図9のスケッチマップを設ける。画像理解とは、

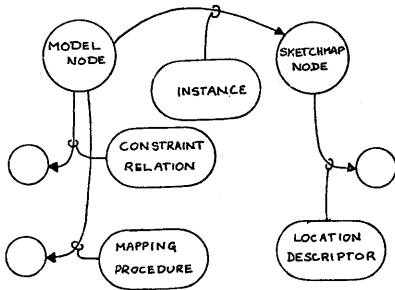


スケッチマップを作ることである。その構造はモデルと同様なネットワークで、各ノードはモデルと対応がとれたものである。そのノードは画像の境界線とか領域との対応が取れたものでもある。

モデルのノード間のリンクはお互いの関係を表わしているが それには次のような種類がある。

- ・拘束 ----- 関係の成立する確率、関係の値の期待値、関係の成立する条件。
- ・位置 ----- どこにそれがあるかを表わすプログラムを指す。
- ・写像手続 ----- それを見つけるプログラムを指す。

図9. 画像理解の概念図



したがって一つのノードに対応するものを画像から探す場合には、図10のような知識によってスケッチマップと画像の写像を求めることになる。

画像から何かを探す場合には、人が制御プログラムを書く必要がある。このプログラムは上述の知識を仮定したものであるため、細かい部分は隠されている。ALGOLの形式で書かれたプログラムの例が示してあるが、かなりわかり易い。

図10. モデルとスケッチマップ

このシステムはARPAの画像理解システムのプロジェクトとして研究されている。実際に行、た仕事は、港の船を見つけたり、胸のX線写真からあばら骨を見つけているという簡単なことにすぎない。ユーザがプログラムを書かなければならないという欠点はあるが、かなり自由に動作を指定できる利点がある。これは画像処理専門家向けのシステムを目標しているものと思われる。

6. 画像理解システムのプロジェクト

1975年から始

められたARPA
の上記プロジェ
クトに關係する
ものがいくつが
発表された。ス
タンフォード大
学からは、2章
で述べた両眼立
体視の他に一つ
ある。テレビカ
メラを載んで動
きまわる車が、
自分の位置を知
り、障害物を避
けるための研究
である。一度見
たシーンのめぼ
しい物が、少し
動いた後でどこ
に見えるかを高
速に求めるため
の手段が論じて

表2. ARPA の画像理解システムプロジェクト

#	Contractor		Project Leader	Theme
	University	Industry		
1	Stanford		Binford	Navigation
2	USC	Hughes Aircraft	Andrews	Smart Sensor Feature Extraction
3	CMU	CDC	Reddy	Change Detection Digital Architecture
4	Maryland	Westinghouse	Rosenfeld	CCD-based Processor Low Level Processing
5	Purdue	Honeywell	Fu Haung	Low Level Image Modeling
6	MIT		Winston	Registration by Shading
7	Rochester		Feldman	Knowledge-Directed Image Analysis
8	SRI		Hart	Semantic Driven Segmentation
9	Utah		?	

ある。(H.P. Moravec : Toward Automatic Visual Obstacle Avoidance.)

カーネギーメロン大学からは、3章で述べた論文 (S. M. Rubin and R. Reddy : The Locus Model of Search and Its Use in Image Interpretation) と、異なる時に写した2つの航空写真から、両者を記号で表現して変化を知ろうとする K. Price and R. Reddy : Change Detection and Analysis in Multispectral Images がある。

SRIからは、地形の特徴点の3次元モデルを用いて、航空写真との対応を求め H. G. Barrow, J. M. Tenenbaum, R. C. Bolles and H. C. Wolf : Parametric Correspondance and Chamfer Matching : Two New Techniques for Image Matching と、地図情報を用いて航空写真を解釈する H. G. Barrow, R. C. Bolles, T. G. Garvey, J. H. Cremers, J. M. Tenenbaum and H. C. Wolf : Experiments in Map-Guided Photo Interpretation。後者は港、道路、線路などの位置に基づいて、そこにある船や自動車、貨車を見つける説である。前述のロチェスター大学とともにいかにも軍と関係のあるプロジェクトであるという感じがする。

7. おわりに

米国の研究ばかりを述べたが、今回は英国以外に西独、カナダの論文が多かった。日本からの論文は第一級と思われるがここでは紹介しなかった。