

1992 年コンピュータビジョンとパターン認識会議参加報告

浅田 稔*, 佐藤 嘉伸*, 大和 淳司**, 山本 裕之***, Jiang Yu Zheng****

* 大阪大学, **NTT HI 研, *** キヤノン情報システム研, ****ATR 通信

* 〒 565 吹田市山田丘, asada@robotics.ccm.osaka-u.ac.jp

あらまし

1992 年 6 月, 米国イリノイ大学で開催されたコンピュータビジョンとパターン認識会議に出席したので, その概要を報告する.

和文キーワード コンピュータビジョンとパターン認識会議

A Report on CVPR'92

Minoru ASADA*, Yoshinobu SATO*, Junji Yamato**, Hiroyuki
YAMAMOTO***, Jiang Yu Zheng****

*Osaka University, **NTT HI Lab. ***Canon Information Systems
Research Center, ****ATR Communication Systems

Abstract

This report describes an overview of the 1992 Computer Vision and Pattern Recognition Conference,
which was held at Champaign, Illinois, USA during June 15th and 18th.

英文 key words Computer Vision, Pattern Recognition, Champaign, Illinois

1 はじめに

1992年6月15日から18日にかけてイリノイ州シャンペインで開催されたIEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR'92)に参加したので、その概要を報告する。今年の会議では、表1に示す様に18の本セッション・3つのポスターセッションに分かれ、89件の本セッション発表・70件のポスターセッション発表が行われた他、3件の招待講演・2件のパネル討論・2つのワークショップが催され、毎日約400名の参加があった。今年の会議の特徴の1つは、ビジョンの問題を動作(能動的なセンサの調整・回転・移動等及び処理戦略の制御)と結びつけて解決を試みる研究報告がしめる割合が多いということである。Purposive Visionのワークショップ、2つのActive Visionのセッションで、合計18件の研究が発表されている。さらに関連した研究発表を含めると、全体の約13ビジョンの研究に関するものと考えられ、昨年までと比べると急増している。

以下では、各報告者からの原稿を浅田の独断で追加・修正などを行なってまとめたものであるが、必ずしも全てをカバーしていない。これは、Current Topicを中心にするという意図として受け止めて頂きたい。

表1 Contents of Conference

A1	Active Vision I	5
A2	Calibration & Pose Estimation	5
A3	Active Vision II	4
B1	Shape Description & Recovery	5
B2	Invariants	5
B3	Handwriting & Neural Networks	4
C1	Motion Sequences & Optical Flow	6
C2	Shape from Texture/Focus	5
C3	Range Images	4
D1	Perceptual Organization & ..	6
D2	Object Recognition	5
D3	Applications	5
E1	Navigation & Recognition	5
E2	Stereo	5
E3	Morphology	5
F1	Shape from Shading/Photometric Stereo	5
F2	Statistical Model for Low-Level Vision	5
F3	Segmentation	5
P1	Poster Session I	70
招	Future Direction for Computer Vision	1
待	Biological Image Representation	
講	and Visuo-Motor Control	1
演	Development and Applications of a Low-cost Space-variant Active Vision System	1
パ	Future Direction of Computer	1
ネ	Vision Research	1
ル	A Report on DARPA Image Understanding Environment Project	1
	Workshop on Purposive Vision	8
	Symposium on Physics-Based Vision	1

2 Workshop on Purposive Vision

本ワークショップでは、Purposive Visionに関する研究成果8件が発表された。Purposive Visionの定義は定かでないが、発表された内容から推測すると、「タスク・オリエンティッドな能動的なセンサの回転・移動・内部パラメータの制御や処理戦略の制御を伴うビジョンの研究」と考えることができるであろう。ワークショップではこの分野での著名な研究者が、1件45分をかけて発表を行った。参加者は約80名であった。

コロムビア大のKenderは2つのビジョン・システム、すなわち人間等のバイオリジカルなビジョン・システムと現状のコンピュータ・ビジョン・システムを比較し、その相違、コンピュータ・ビジョン・システムに欠けている機能を整理したうえで、タスク指向のコンピュータ・ビジョンの研究の重要性を検討した。バイオリジカル・ビジョン・システムは、Foveation, Nonuniform Density, Color, Binocularity, Converge, Response Time, Iris, Focus & Zoom, Shutting Down等の機能を利用して環境に適応している。従って、コンピュータ・ビジョンの研究においても、1)ある環境下において目的を達成するためのタスク指向のビジョン、2)様々な環境に適応していくためのGenetic Algorithm的なアプローチの研究が今後必要となるという発表者の考えが説明された。

ジェノバ大(イタリア)のSandiniは、物体の特徴認識・マニピュレーション(行動)戦略の制御にアクティブなVisuo-Motor機能の利用に関する幾つかのアイデアを発表した。ビジョンの目的は行動である。これまでの研究では、

1. Look, 2. Plan, 3. Close the eye, 4. Move

の4段階の動作でシステムが構成されており、LookとMoveが時間的に分離されている。このようなシステムでは、視覚処理に対する要求が必要以上に大きくなる。Active・Purposive Visionでは、1)視覚処理はマニピュレーションの制御ループの一部であり、2)視覚処理は其中で学習と行動中のモニタリングを行う必要がある。この発表では、ロボットが物体を押しながら目的地へ進むことを題材に、オブティカル・フローを利用したシステム事例が示された。発表者は、Non-uniformなCCDセンサの開発者として著名である。

従来のマシン・ビジョン・システムでは、入力画像の認識処理(Discovery Mode)に処理の重点がおかれていた。ミシガン大のRamesh Jainは、タスク指向の予測が認識処理を制御し、効率的に処理が行えるシステムを発表した。システムは、Perception, Cognition, Actionの3つの機能の協調動作で構成される。これらの処理機能は環境知識を仲介として情報交換を行うため、環境知識から「的確な時点での的確な情報」を検索(Knowledge Caching)出来る必要がある。ここでは、システムのタスクに応じたContext-based Cachingという方法が発表された。ここで述べられた3つの機能の協調によるビジョン・システムの構築は、従来からこの分野で解決されなければならないと考えられている問題である。発表からだけでは、発表者のシステムが実際のビジョンの問題に適応できるかどうか、疑問が残った。

イリノイ大の Ahuja は、2つのカメラの視線方向・フォーカス・絞り・ズームを動的に制御してシーン内を走査し、ステレオ情報を基に正確なシーンの3次元マップを入力するシステムを発表した。システムは、1) シーンのラフな3次元情報をステレオ画像をもとに入力、2) この情報を利用して次にカメラが観測する点を選び (Target Selection), その点にフォーカスを合わせる (Homing), 3) このフォーカスをもとにその点までのより正確な距離を計算し、2つのカメラの中心をその点に合わせる (Fixation), 4) その点の周りにウィンドウ (Fovea) を設定し2つのカメラを利用してステレオ情報からウィンドウ内のさらに正確な3次元情報を復元、5) これまでの3次元情報と融合し (Merging), 次の観測点を決定する、6) 2) からの処理を繰り返す。これらの動作により、シーン全体の正確な3次元情報が復元できる。Target Selection の基準には、距離・面の滑らかさ・オクルージョン・複雑さが考えられるが、このシステムでは、次の観測点までの3次元距離の推定値とカメラの回転する角度の重み付け加算値が最小になる点を利用している。このシステムは疎密制御で3次元情報の復元を行っている点でユニークである他、ステレオ情報の利用を画像中心の小さなウィンドウ (中心か) に限定している点でユニークである。これは、2つのカメラが独立に回転することで実現できているが、これにより光学的なレンズの歪み等による誤差を最小に抑えることができる。

シカゴ大の Swain は、ビジョンとアクションを1つのアーキテクチャ内で考察していく必要があるというアイデアの説明を行った。アクティブ・ビジョンの枠組み内では、リアル・タイム性の要求から Focus of Attention・最小限の記述・タスク指向のアルゴリズムが必要となる。これらは、即ち、アクティブ・ビジョンに、1) カメラの移動、2) 不完全な処理結果の融合、3) Incremental なアルゴリズムが必要であることを示している。これらの要求を満たすために、発表者らは Active Sensing, Behavior Control, Reactive Planning (おおまかなプランを状況に応じたプリミティブな行動に分割する機能) から構成されるシステムを構成中であり、そのアイデアを発表した。発表者のアイデアは、先の Sandini や Jain のアイデアと共通している。

3 Active Vision I

このセッションの前半の3件は、アクティブ・ビジョンの枠組みでセンサ (距離センサやカメラ) を次にどの方向へ、どこへ移動させると、対象物体の3次元形状復元に適した情報が得られるか、そのアルゴリズムに関する検討結果が報告された。4件目の発表では、動物体を移動中のカメラ・ヘッド (2つのカメラとその回転機構) を利用してどのようにして追従するか、その事例が発表された。最後の発表では、能動的に光源位置を移動することにより物体上の点までの距離を計測するアルゴリズムが紹介された。参加者は、約200名であった。

マギル大 (カナダ) の Whaite らは、距離データから求めた3次元のモデルの予測誤差に基づいて、次の視点 (レンジ・ファインダ) 位置と方向を求めるアイデアを発表した。こ

こでは、利用した予測誤差の求め方が有効であることを、シミュレーション・データを用いて確認している。この手法を実シーン (仮定した物体モデル以外の形状物体を含んだ) に適用するためには、さらに改良が必要であると考えられる。

メリーランド大の Herve らは、物体形状を求めるためには、通常の Shape-from-shading では非線系な方程式を解く必要があるのに対し、カメラが移動することにより線系方程式を解くのみでよいという理論的な研究を発表した (以前に発表者らにより報告されているが)。さらに、その解を安定にする方向へカメラを移動するというアイデアは、アクティブ・ビジョンの特徴を活用している。

ウィスコン大の Kutulakos らは、Occluding Contour を利用した Shape from Shading Method で視線方向が面の曲率の主軸方向と一致した時、Contour から形状復元が可能であることから、もっとも形状復元しやすい視点を、面の曲率と Contour の曲率の関係をを用いて求めた。視点設定の簡単なアルゴリズムを提案し、実験結果を示した。

ロチェスター大の Brown らは、両眼による実時間追跡システムを発表した。対象物体を画像中心に保っているの、追従操作に必要な画像処理は、すべて画像中心の小さなウィンドウ内 (Fovea) で行える。このウィンドウを設けないと、追従動作に支障が起ることが報告されていることが興味深い。現在のシステムでは、約8Hzの物体の振動には追従できる。なお、カメラ・ヘッドのハードウェアに関しては、4月に報告を行った SPIE シンポジウムの報告書を参照されたい。

ハーバード大の Clark は、Photometric Stereo 法に能動性 (光源の移動) を持たせて、物体上の点までの距離 (従来の Photometric Stereo では物体の形状のみ) を線系の方程式で解けるということを示している。実データを利用した実験結果はまだ不完全ではあるが、興味深い。Depth from light source motion といえる手法である。

4 Active Vision II

このセッションの最初の発表では、照明・視点位置による物体の見えの変化による問題を、カメラ・照明を移動することで解消し、物体認識を行うシステムのプロトタイプが紹介された。2件目の発表では、タスク特有の知識をベイズ・ネット で表現し、その知識をもとに処理ウィンドウの移動・処理モジュールの選択を制御する枠組みが紹介された。3件目の発表は移動物体の抽出を微小運動するカメラから抽出する手法、4件目は人間が行う物体の組み立て動作をビジョン・システムとロボットで再現するシステムが紹介された。

トロント大の Tsotsos らは、照明・視点位置による見掛けの形状の変化を、視点位置を Standard View へ常に移動することで回避して、物体認識を行なう手法を提案した。未知の物体の画像を基に Standard View へカメラを移動するアルゴリズムがアド・ホックであり、適用範囲が非常に限られると考えられる。

ロチェスター大の Brown らは、視点 (Fovea) の移動もネット で表現している点で、従来のベイズ・ネット をビジョンの問題に適用した手法と異なる方法を提案した。

メリーランド大の Sharma and Aloimonos は、カメラを

微小に移動させることで、従来の手法のように画像データからの定量的な数値を用いずに、定性的な推定から移動物体の抽出を行う方法を提案した。実シーンに適応した結果が乏しく、適用性が疑問である。

CMUのIkeuchi and Suehiroは、人間の組み立て作業を観察(距離画像を利用)することにより、ロボットをプログラムするアプローチを提案している。システムは、人間の作業前と作業後の距離画像を比較し、部品間の接面関係の変化を抽出する。これらの変化は、Procedure Treeで表現されており、それを基にロボットのMove, insert等の動作を決定できる。この発表は他のActive Visionのセッションの発表と異なり、センサの能動的な移動を伴わない。Active Visionの範疇より、むしろPurposive Visionの範疇であろう。

5 その他のアクティブビジョン関係の発表

Shape from Texture/Focusのセッションにおいて、レンズのフォーカスを変化させながら連続的に入力した画像(群)を基に、シーンの奥行情報を復元する手法が2件発表された。

ポスターセッションにおいても1件、フォーカスの変化による距離情報の入力発表されていたが、これは画像中の1点(中心)の距離を計算するのみで新規性は感じられなかった。

その他ポスターセッションにおいて、人間の目の動きを模擬するための手法が2件、アクティブな移動がおこなえるシステムにおけるカメラ・パラメータの校正の手法が1件発表された。

以上の発表から、アクティブ・ビジョンに関する研究の現状は、ill-definedの問題をアクティブにセンサを制御することによりwell-definedの問題に換えて解決しようとする理論的な研究・視点の移動のアルゴリズム・システムのコンセプトの研究が中心であり、実際のシステムの構築に関する研究が少ないう印象を受けた。これは、1つにアクティブにセンサをコントロールするハードウェア・ソフトウェアが一般に利用できる機会が少ないことが原因であると感じる。これらのシステムの構築は、機械・制御・画像処理・光学・素子技術など様々な分野の技術を要する複合研究課題である。しかし、レンジ・ファインダの入手が容易になるに従い距離画像処理の研究が普及してきているのと同様、これらのハードウェア(カメラ・ヘッド等)が一般に利用可能になるに従い、この分野の研究もシステム構築を含んだ研究に移っていくと考えられる(以上、山本、浅田)。

6 Shape Description and Recovery

このセッションは会議の第1日目の最初のセッションであり、かつ、今話題のActive Visionの裏セッションであったため聴衆がやや少なかった。Medioniは彼の発表の冒頭で、「形状記述は非常に大事なテーマであり、本来ならもっと注目されるべきなのに。」と述べていた(同感である)。5件のうち3件は、一般化円錐(Generalized Cone:GC)の復元に関

するものであった。相変わらずStraight Homogeneous GCという数学的性質が良いクラスのみを対象とする研究が多いなか、Sato et.al.は、piecewiseなGC当てはめにより、Non-straight Non-homogeneous GCを距離画像から復元することを試みている。残りの2件は2次元形状の「階層的」軸表現に関するものであった。Rom & Medioniは、曲率解析、局所対称、大局的対称などを組み合わせることにより、人間の直感に近い階層的パーツ分割法を提案した。Ogniewicz & Ilgはボロノイ線図に基づく階層的骨格化手法を提案した。今後は、これらの形状記述を用いた「パーツに基づく物体認識」手法を本格的に検討する必要があるだろう。

7 Perceptual Organization and Curve Description

6件の発表のうち4件が知覚的組織化(群化処理)に関するものであった。この分野は、問題の定義自体が難しい。ベイズネットを用いたもの、共通曲率を持つような線素の群化を多重スケールで行うものなどが発表されたが、個々の手法の性能評価は難しい。知覚的組織化の問題に関して、性能評価がきっちり行えるような枠組みの検討が必要であると感じる。残りは多重解像度閉曲線解析に関するもので、記号記述への変換法や収縮を伴わない平滑化法があった。

8 Range Images

特に目立った傾向は見られない。Mirwa & Boyerは、「AIC(赤池の情報量基準)」を通常の最小2乗法だけでなく、ロバスト推定法において適用できるよう改良を加えた。そして、距離画像に関数モデルを当てはめる際のモデル選択に利用した。AICは、データに最もうまく当てはまるモデルを選択する規準として統計学ではよく知られている。Monga et.alは、CT装置等により得られる3D濃淡画像から、エッジ検出、薄面化、表面記述という処理過程を経ることなく、直接的に表面曲率を計算する手法を提案している。その他に、距離画像の多視点統合による3Dモデル構成、および、時系列距離画像を用いた自動走行車の航路決定に関する発表があった。

9 Object Recognition

最近では、認識システムの構築よりも、理論的な解析を指向する傾向が強い。これは、「物体認識」が一分野として確立されつつある証拠であろう。Grimsonを中心にMITでは、Formalアプローチを進めており、各種認識手法の計算の複雑さや、測定誤差に対するロバスト性の解析を行っている。今回、Grimson et.al.は、2D画像から3D物体を認識する際に、画像特徴の測定誤差が誤認識率に与える影響を解析した。開発した曲面物体の認識手法に同様の誤差解析をBasriも自身も行った。大規模モデルベースの取扱もこの分野の重要テーマである。Flynnは、モデルベースにおいて、同一物体内および異物体間の共通部分を予め登録することにより、無駄な仮説生成をなくし、さらに特徴の顕著さの測度を導入するこ

とによる認識の効率化を提案している。Eggert et.al. の Scale Space Aspect Graph は、去年の CVPR の前の Aspect Graph に関するパネル討論で多重解像度化を考えるべきだという指摘があったところなので、非常にタイムリーな印象を受けた。Aspect Graph を Scale を反映する 1 パラメータ族として、そのパラメータに沿って必ず Graph の複雑さが単調に減少するように表現したい。Scale パラメータとして視角 (visual angle) を用いることを提案しており、議論は、3 次元多面体の 2 次元投影ではなく、次元を 1 つ縮退させた 2 次元多角形の 1 次元投影で行っている。視角 σ に対して、2 次元多角形のある辺が可視である、すなわち、その辺をある撮像点から見たときの視角が σ より大きくなるような撮像点の領域は、円周角の定理に基づき、その辺を弦とするある円弧内となる。 σ を大きくするにつれその領域は小さくなる。多角形全体に対して、これをプロットし、円弧の各交差領域が Aspect Graph のノードに対応する。 σ を連続的に変化させることにより Scale Space Aspect Graph が生成される。残念ながら、 σ の増加に伴って Graph の複雑さが単調に減少するという保証はない。また、3 次元多面体への拡張も今後の課題である。

10 Poster-1

各ポスターセッションでの議論は非常に活発に行われていた。Medioni が「Best Paper は、おそらく Poster Session の中にあるだろう。なぜなら、(既存の価値観にとらわれた)査読者は、真に新しい研究に気づかない場合が多いからである。」と述べていた。24 件のうち、興味を引いたものとして、Levy Vehel et.al による Multifractal 理論、Sato et.al. による 3 次元形状の階層的パーツ分割手法、および、Zabrodsky et.al による対称性の連続的測度に関する発表があった。

11 Symposium: Physics-based Vision

色(color), 放射測定 (radiometry), カメラモデル (camera modeling) の直列 3 セッションで、発表、およびパネル討論が行われた。新しい研究成果の発表ではなく、この分野でこれまで顕著な成果をあげた研究者 14 名が、"keynote-style" で発表を行う形をとった。Organizer の中心は、CMU の S. Shafer で、発表の質も高く、Purposive Vision の Workshop とは対照的に、パネル討論は活発であった。Physics-based Vision の定義は、Horn の研究に端を発する、画像生成過程の綿密な解析を基盤とするアプローチと考えることができ、偏向 (polarization) や照度差ステレオ (photometric stereo) などセンサの工夫により複雑な推論なしに表面特性を直接的に得ようという工学的アプローチと、通常のカラー画像や濃淡画像から画像生成過程の解析に基づき表面の不変特性を推定するという視覚の計算モデル的アプローチに大別できる。色のセッションでのパネル討論における印象的な質問として、「あなたの方法を確かめるのに何種類の画像で実験を行ったか?」というものがあった。残念ながら、どの答えも高々数枚であった。確かに、Shape from Shading 等においても、より現実に即した表面反射モデルを用いるようになってきて

おり、計算モデル的アプローチは、Toy World から Realistic World に着実に進歩している。しかし、Real World までには、まだかなりの道程があるという印象をもった。とはいっても、まだまだ基盤のしっかりしない Purposive Vision とは異なり、一分野として確立しているという印象は強い。また、これを機に Physics-based Vision の成果が 3 冊の論文集にまとめられた。CV 関係だけでなく、光学関係の専門誌も含む広い範囲から集められたので、価値あるものになっていると Shafer は述べていた (以上、佐藤)。

12 Invited talk II

Development and Applications of a low-cost, space-variant active vision system, E.Schwaratz. NYU

超小型カメラを使用したアクティブビジョンシステムの開発と応用例の紹介である。画像は DSP により log-polar 変換され、中心部の解像度が高い。カメラはモータでなくコイルでドライブされ、高速化を実現した。応用として、走っている (模型の) 自動車のナンバープレート読みとりのデモビデオが上映された。今後移動ロボットに搭載する予定である。log-polar 変換については、イタリアの Sandini らが開発した RVS(Retina vision system) のカタログが会場で配布されていた。こちらは、CCD の画素配列そのものを log-polar としたもので、4600 ECU で発売予定とされている。

13 Application

5 件の発表があった。1 件目は顔認識、2 件目は動作認識、3, 4 件目はパーツの Visual Inspection、5 件目は心臓壁のモデリングであった。人間にまつわるもの、生産自動化関連、医用関連と、現状の vision の応用研究に即したものになっている。B.S.Manjunath et.al. は Gabor-wavelet で curvature の変化を特徴として抽出し、グラフマッチングで辞書との比較を行ない、顔認識をおこなった。J.Yamato et.al は、人物像のメッシュ特徴をベクトル量子化してえたシンボル列 HMM で学習、認識することで、動作認識をおこなった。いづれも対象に関する知識を陽に利用しない点で限界はあるが、適用対象によっては有効であろう。J.Mundy et.al は、I2F:Image Interpretation Foundation と呼ばれるシステムのためのコンセプトと実験の報告をおこなった。オブジェクト指向と制約ベースモデリングが基本になっている。テンプレートの作成は人間がインタラクティブに行なうが、人間に使いやすくするにはどうしたらよいか、というのも重要なテーマであろう。100talk でもあげられたポイントの一つである。B. R. Modayur et.al は、CAD モデルをベースにした寸法検査システムに関する発表をおこなった。従来あやふやだった、幾何公差の扱いを厳密におこない、計測アルゴリズムと合格/不合格決定の方法について検討している。精密測定において画像を単独で使用することは精度の上でむづかしい点がこののではないと思われるが非接触で測定できるなどの特性を生かした応用が可能だろう。

14 Motion Sequence and Optical Flow

6件の発表があった。オブティカルフローの計算法自体に関する発表はなく、それぞれ、フローを利用した Range Estimation, モデル獲得, 物体の強調などの応用的なもの、そして、非剛体の対応点決定, FOEの計算法, 最後に各種のオブティカルフロー計算法のパフォーマンス評価であった。M. Irani et.al は multi-resolution の gradientをつかった motion の計算により、複数物体の追跡をおこない、画像中の物体の強調, 分離, 解像度向上などをおこなった。ビジョンの分野で培われた手法は画像の加工, ビデオハンドリングなどの分野へも様々な応用が考えられる。マルチメディア流行りの昨今この種の適用が増えるかも知れない。J.L.Barron, D.J.Fleet et.al は各種の optical flow 計算法の実画像を対象とした、定量的な評価をおこなった。評価した計算法は、differential ベースのものとして、Horn and Schunk, Lucas and Kanade, Uras の方法, 領域マッチングベースの Anandan, さらに、Gabor filter のエネルギーベースの Heeger, 位相ベースの Fleet and Jepson (評価者自身) であった。対象画像は、合成画像と、樹木の自然画像の移動, 拡大と森の中の移動であった。もとの精度が良いのは Fleet and Jepson という(手前味噌気味の)結果であったが、計算量が非常に多いため、総合的には Lucas and Kanade が実用的ということであった。手法の評価が重要だという点は、Reddy の invited talk でも強調された点である。また、パネルセッションでも統一的な評価のための画像データベースを作るべきだという意見も出ていた。しかしそこでも、またこの発表でも、対象とした画像に特化することを懸念する人がいた(以上、大和)。

15 Invited Talks I

Future directions for computer vision, Professor Raj Reddy, Robotics Institute (所長), CMU

この講演者は、70年代ビジョン発足の時の大先輩であり、歴史の観点から現在あるいは将来何ができるか、何をやるべきかの意見を述べた。1. Multimedia, Image data base を利用する(最近画像蓄積の技術が進歩した。そのため、従来の少ない画像から視覚の問題解決を果たすという概念は適応しなくなり、より多くの画像を使うことも考えられる。)。2. Powerful Sensor と Uncertainty, 3. Basic Approach と Physical based vision (視覚現象を起こした物理的原因から調べる)、さらに、次のことをやらなければならないと以下のように述べた。

1. 応用を重視する(アメリカに対する話、日本は別)。従来の 100 年の自動化の概念から離れて、90 しまう。残った 10

2. Benchmark — 共通の基準を設け、共通の画像を用意し、研究者全員が使えるようにする。これらの画像に対して、研究者各自で開発したアルゴリズムを走らせ、検証や評価を行なう。成功したものを全員に配って使う。そのため、研究者たちは何でもゼロから始める必要はなくなり、コンピュータビジョンの進歩を促進させることができる。

3. Learning of Observation. 観察から理解する。システムは、人々が同じタスクをどのようにやっているかを観察することによって自分の知覚能力を上げる。このような学習は、経験を積み重ねによって達成できる。その過程では、誤動作から、質問をすることから、質疑応答が繰り返される。

4. Learning via simulation — Virtual Reality も大いに関係する。

5. Train Robot — Neural Network は刺激など単純なものにはよいかも知れないが、ビジョンのような複雑の対象に対しては不向きではないか。

16 Invited Talks III

Development and applications of a low-cost space-variant active vision system, Professor Eric Schwartz, Courant Institute of Mathematical Sciences and Medical School, New York University

人間の網膜の視覚セル分布と同様なセンシングユニットをカメラの中に配置する。その結果、視野の中心部が高分解能、周りに低分解能分布になり、全体のセルの数が少なくなつて、視覚処理が素早くできる。この講演は、このようなシステムの実際の構築過程を紹介した。開発されたカメラは小さいモータに駆動される。研究用タイプと商用タイプの二つの種類がつくられたが、商用のタイプは日本のものより小さく(レンズは親指ぐらい)、そのため、小さいモータでも高速にカメラを動かせる。このカメラシステムを用いて走っている車を追跡し、そのナンバを読み取る実験のビデオも上映した。カメラの視野のセル分布は、従来の正方形行列と異なる為、計算機内部のデータ構造も独自に設計した。勿論、アルゴリズムも全部開発しなければならない。視野中高分解能の領域が小さい為、常に視線を動かして、興味のあるところに向いていく。つまり、分解能の不足は動的観測によって補足される。逆に、このような少ない画素では、高速な画像処理が可能となる。そのため、カメラをより速く移動させることが、このカメラシステムでは容易に実現できた。それはまさに人間の視覚機構と同等であると言える。しかし、本世紀末高分解能かつ広い視野での早い処理も実現できるかもしれないと反論する人もいる。私の感覚では、それは保証できない。人間の知覚をコンピュータの力を借りている分野では、このようなシステムは非常に理想的ではないかと思われる。なぜなら、ここは人間相手を満足させるため、将来実時間の処理が要求される。又、各瞬間の画像理解も全領域に渡っているものではない。そのため、このような能動的カメラは近い将来実際に使われるかも知れない。

17 Beckman Institute 見学

ATR と比べてやや小さく、教会のような建物であり、内部は華やかでまるでホテルのような Beckman Institute 中で、地元大学の研究が数多く展示されていた。内容はそれほど深いものではなかったが、一目見てすぐ有意な情報が得られるよいものであった。設備は移動ロボット、アーム、能動的ス

テレオカメラヘッドなど一通り揃っている。Huang 教授のグループは画像からの運動解析、心臓の変形モデルの計測と表示などを行なっている。顔など理解や表示の研究を提唱しているが、成果と言えるものはまだない。Ahuja 教授のところは移動ロボット制御及び環境理解、テクスチャーからの3次元獲得、能動的センシングなどの研究を行なっている(以上, Zheng)。

18 パネル討論：Future Directions of Computer Vision Research

去年ハワイで同会議のNSFのワークショップの報告(資料150ページにも及ぶ)で、現地(ハワイ)のプログラムや、日本から一人参加した白井先生などのはなしでは、初日に3つのパネル討論を実施し、その後、6つのグループに分かれて、細かな討議をやった模様である。ここでは、主に6つのグループからの総括報告がなされた。以下、ハワイでの3つのパネル内容を始めに簡単にまとめる。

最初のパネルは「New Research Directions」と題するもので、南カリフォルニア大のBowyerが座長をつとめた。論点としては、「Generic Vision」と「Task-oriented Vision」に、関するもので「Generic」を「一般幾何モデル」と読むか、「物体の機能」に基づくものにとるかで、定義が異なることの確認(Bowyerのグループでは「機能」モデル関連の研究を行なっているためか? ポスターでも1件発表されていた)が、なされたこと、また、ともに明確な定義とよい実例が必要と言うあたりまえの結論が出されていた。

次のセッションは「Technological and Conceptual Obstacles」と題するもので、ロチェスター大のNelsonが座長をつとめたようである。ここでは悪質な論文を駆逐しなければいけないという意見がだされ、それに対する具体的な方法も検討された。NelsonがAloimonosのもとで博士を取得していることから、Active, Purposive, Task-oriented, Qualitative Visionなど、はやりのテーマが論議的になったようである。

最後は「How Do We Build Vision Systems」で、SRIのLeclercが座長をつとめた。実システムとしての要件として、画像生成過程と再構成問題の理解、タスクに応じたビジョンモジュールの作成とその相互関係の把握、特定システムの構築による一般化視覚への理解促進などが挙げられた。

これらのパネルに続く6つのセッションでは、ワークショップ参加者全員がいずれかのグループに入って討論を重ねた。これらについて、数人が代表してCVPR'92で簡単な総括を行なった。誰がまとめたか定かではないが(筆者の記憶が曖昧)、以下に簡単に述べる。

「Early Processing」では、これまでの初期計算を含め、以下が課題であると結んでいる。それらは、画像表現(尺度や方向に依存した分解、さまざまな種類の特徴を抽出するためのそれらの用法)、シーン境界の抽出、グルーピングと知覚的組織化、初期視覚モジュールの統合、異種センサ情報の統合、センサ自体のモデリングである。発表は確か当地イリノイ大のAhuja(?)。

「Modeling and Representation」では、画像表現、面

表現、物体表現、そして、画像モデリング、構造的モデリングと形状モデリング、静力学的・動力学的モデリング、不確実性モデリングなどが研究項目として挙げられた(当たり前か?)。特に4つのモデリングを統一視することが、さまざまな応用に対応でき基礎を構築すると結んでいる。

「Representation and Recognition」のセッションでは4つの項目が論議され、いずれもMundyカーブと称された「反比例」カーブがもてはやされた。4つの項目は、Grouping and Segmentation, Matching and Indexing, Model Acquisition, そして、Theory Validationである。図に最後のトピックのカーブのラベルを示す。これらとHPCC(High Performance Computer and Communications)との関連について論議された。発表者はUSCのMedioni(?)。

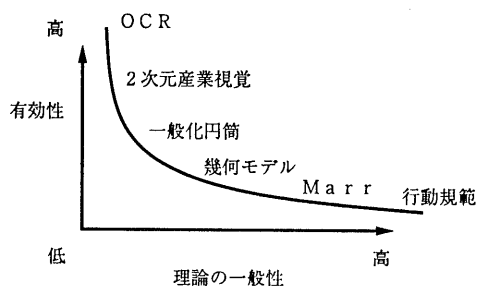


Figure 1: 理論の有効性の現状

「Robotics and Navigation」では、研究項目として、不整地環境の表現と戦略、動的環境での航行、正確なモデルや「感覚-立案-行動」サイクルに代わるもの、センサおよびマニピュレーションの統合、視覚に基づくマニピュレーションなどが挙げられ、実用化システムとして「人」との協調などが論議された。発表はAloimonos(?)。

「Industrial Application」では、撮像技術、エンドユーザーインターフェース、タスク指定、高速高効率、高速検証と正確な位置決め、システム統合などが研究項目として挙げられたと同時に、ビジョン教育についても産業界との結びつきが必要であるとのコメントが出された。Shapiroが発表(?)。

「Databases, Software Libraries, and Benchmarks」では、CV研究のインフラストラクチャの構築と画像データベースの両面から論議され、CV研究成果のデータやプログラムの共有化を目指した巨大データベース構築による科学技術の基盤促進、データベース研究の促進などが唱われた。

以上、NSF主催のCVの将来研究に関するワークショップの総括パネルであったが、どの結論もさして新しい印象を受けなかった。

19 Navigation and Recognition

セッションタイトルから分かるように、Navigation関連が3(?)件と認識関係が2件であった。UMassのグループからの2件あり、一つはShallo Structureと彼らが読んでいる点や直線などのトークンマッチングをAffine変換の4つのパラ

メータで拘束することにより、特徴の追跡が可能であることを示した。Shallow Structure は、観測者から物体までの距離に比べて、物体自身が持つ奥行きが小さい時にそれを Depth Constant といった特徴物で、一般性があると思われるものの、実験では本当に Shallow のものしかあつたことがなかった。もう 1 件は、最初に近似透視射影 (2 次元の回転と並進プラス尺度) で、ランドマークのあたりをつけ、次に真の透視射影によって、ロボットの正確な運動を算出するもので、近似による高速性を唱っている。三つめはハワイからフロリダのマイアミ大学に移った Negahdaripour の発表で、ステレオ時空間勾配情報の直接利用による (いわゆる直接法)、運動と構造の復元で、平均視差を初期値として、繰り返し法で精度良く求める手法の提案と実験結果の提示があったが、従来の彼らの研究の続きで新しさは感じられない。

残り 2 件はいずれも MIT の AI ラボからで、3 次元物体の 2D ビューマッチングに関連するものである。3 次元物体の特徴点の 2 次元投影点を 2 次元の別空間内の 2 直線で表現できることを示し、これを用いてマッチングを行なっている。もう 1 件は、多重分解能、ハフ変換、探索に基づく認識などのアイデアを混合した RAST と呼ばれる探索アルゴリズムを用いて、変換空間を動的に分割することにより、高速な認識を目指すもので、アイデアが紹介されたが、実際の結果は、充分には示されていなかったようである。

20 Shape from Shading/Photometric Stereo

Shape from Shading では、2 件が特異点 (最大明度点) を出発点として、形状復元 (高さ復元) を行なう手法を発表した。一つは、光源が観測者と同じ方向であるとの仮定をおいて (後に一般化)、高さの変化と明度の関係から、形状を逐次求めるもの、また、もう一つも特異点 (局所凹、局所凸、または鞍点) を出発点として、もっとも光源から遠ざかる点に面情報を伝播していく手法 (minimum down hill principle) を利用して、数回の繰り返しで、安定した解が得られたことを発表した。これまでの Shape from Shading が面の法線方向を復元することを目指していたのに対し、高さの変化と捉えることに (CVPR'91 でも SRI から同様の発表があった) 置き換えたものが多くなっている。

Photometric stereo では、CMU から 4 方向の光源を利用して面の属性、特に Specular Lobe と形状を復元する手法が提案された。もう 1 件は、単眼の Shape from Shading を 2 方向の光源を利用して統合したもので、ともに、アイデア的には新しいものとはいえなかった。

Wolff は、一様乱反射について、これまでの余弦則による単調減少関数ではない現象が見られることを実験を通して示した。これは、Physics-Based Vision の典型で、彼自身もあとの Physics-Based Vision のワークショップで講演を行っていた。このセッションではないが、関連する論文として、ポスターで Asada et al. によって、弱一様乱反射仮定による形状復元手法が提案された。これは、通常の余弦則ではなく単調減少関数のみを想定することにより、輪郭線情報との統合により、光源方向、面の属性、形状、姿勢を復元する

ものである。

21 その他

ステレオ関係では、セグメンテーションの問題と関連づけたものが 2 件、階層的照合が 2 件、オクルージョン領域をベイス則を利用して推定するものが 1 件あったが、いずれも過去の手法の拡張・修正版の印象を受けた。「統計モデルを用いた初期処理 (Statistical Models for Low-Level Vision)」では、従来、アドホックに行なっていた処理を統計学を用いて整理するとともに、改善する目的でなされているように見えるが、必ずしも従来手法より優れているという印象は受けなかった。Morphology 関連では、トピックとしてあまり持てはやされていないので、聴衆も少なかったようである。論文内容より、事ある毎に、「論文採択に問題がある」と文句をいっていたワシントン大の Haralick が印象的であった。かれは、CVPR の会議そのものが CV に偏重し過ぎて、特に motion がやたら多過ぎると主張し、会議自体のあり方に問題を投げかけていた。

全体としては、新しい研究がどんどんでてくるというより、既存手法の洗い直しが多いという印象をもった。これは、ビジョンの研究分野そのものが熟成している証拠かもしれない。既存の枠組をはずれた、新しい応用を目指したものは、少なかったが、それらが新鮮に見えたのは、ビジョンの目的や応用をもう一度考え直す時期にきているかもしれないという印象を受けた (以上、浅田)。