

激論：これからのCV

司会：久野 義徳（大阪大学）

パネリスト：

大田 友一（筑波大学）、金谷 健一（群馬大学）

上田 博唯（日立電子）、松山 隆司（京都大学）

1980年代に花開いた Marr の Computational Vision を踏まえ、最近では、新たなコンピュータビジョンの研究方向を探る試みが様々な角度から行われている。本パネル討論では、「これからのコンピュータビジョン研究の進むべき方向」について、4人の研究者の方々に意見表明およびそれらを基にした討論を行ってもらい、今後のコンピュータビジョンの在り方を明らかにする。

Debate: Prospects of Computer Vision

Coordinator: Y. Kuno (Osaka Univ.)

Panelist:

Y. Ohta (Tsukuba Univ.), K. Kanatani (Gunma Univ.),

H. Ueda (Hitachi Electronics), T. Matsuyama (Kyoto Univ.)

Following the computational vision proposed by D. Marr, which brought fruitful results in the computer vision research in the 1980s, varieties of new ideas have been proposed to find future directions of computer vision. This panel discussion invites four computer vision researchers to clarify promising prospects of computer vision.

イメージプロセッシングおよびコンピュータビジョン研究の流れを大雑把にまとめると、

1960年代：（統計的）パターン認識理論の確立と文字認識への応用

1970年代：デジタル画像処理技術の発展と各種画像処理装置や応用システムの開発

1980年代：Marrの提案した視覚の計算理論（computational vision）に基づく「見え方の科学（shape from X（X: shading, texture, contour, motion））」に関する計算アルゴリズムの考案

となり、理論研究（2値画像）→技術開発（濃淡、カラー画像）→理論研究（3次元形状、距離、運動情報）というサイクルをスパイラル状に着実に進んできたといえる。

最近では、ビジョンの問題を数理的観点からより精密に分析し、それを解くための計算アルゴリズムを考案するといった数理的アプローチのほか、アクティブビジョン、センサフュージョン、タスク指向型ビジョン、定性的ビジョン、分散協調型画像理解、といった新たなビジョン研究の方向性を見いだそうとする試みがいろいろ提案されている。

一方、応用面から見ると、最近では、従来のロボットや工業応用を想定した、機械部品のような人工物の認識ではなく、高度なヒューマン・インターフェイスの実現という観点から人間の身体や顔といった自然物を認識対象とした研究が盛んに行われており、複雑に変形する手足の動作の追跡・認識や応答性のよいインターフェイスを実現するために動画（ビデオ映像）の実時間処理を行うのが一般的になっている。

さらに、近年のマルチメディアブームに加えて、次世代テレビジョンへのデジタル方式導入の流れなどによって、通信や符号化、データベースのように従来コンピュータビジョンとは別の分野と考えられてきた画像関連分野においても、一般のシーンを認識・理解する技術の確立が要求されている。

こうした状況を踏まえ、本パネル討論では、「これからのコンピュータビジョン研究の進むべき方向」について、以下の4人の研究者の方々に意見表明およびそれらを基にした討論を行ってもらい、今後のコンピュータビジョンの在り方を明らかにする。（松山記）

司会： 久野 義徳（大阪大学）

パネリスト： 大田 友一（筑波大学）

金谷 健一（群馬大学）

上田 博唯（日立電子）

松山 隆司（京都大学）

3次元画像メディアとコンピュータビジョン

大田 友一

筑波大学 電子情報工学系

<http://www.image.esys.tsukuba.ac.jp>

画像メディアの歴史は長い

絵画 - 100%の人力依存
写真 - 写真術の発明 = 入力 of 自動化
高臨場感映像 - 動き, 色, 大視野,
高精細, 立体感

視覚機能はすばらしい

視覚の発達は進化の選択

視覚を扱うコンピュータ科学

コンピュータビジョン
センサーとしての技術
優秀な視覚を模擬する
視覚より劣るので見栄えがしない
不良設定問題
CVの能力は幼児期を出ていない
コンピュータグラフィクス
インタフェースとしての技術
優秀な視覚を活用する
視覚に訴えるので見栄えがする
良設定問題
CGは写真術発明以前の絵画時代

CVは3次元写真術 になり得るか

CV技術 + CG技術で 新しい3次元画像メディアを 創製できるか

3次元画像表示の ニーズと実状は

両眼立体視再現技術
周波数分割, 時分割, 空間分割
運動立体視再現技術
受動的方法, 能動的方法

能動的運動視差提示技術の 実状は

3Dモデルベース
(実写不可, 視域自由)
2D画像ベース
(実写可能, 視域限定)
オンラインカメラベース
(実写のみ, 1対1限定)

3D写真術に基づく方法では 実写可能, 視域自由, 1対N

3D映画術が発明できれば...

画像メディアは CV技術の新大陸か

使えない新原理から 使える新技術へ

キーワードは, 冗長性の活用と人間との協調

人間の視覚を代行するCVから 人間の視覚と協調するCVへ

参考文献

田村, 池田編:
知能情報メディア - マルチメディアは
進化する, 総研出版(1995)

科学技術としてのコンピュータビジョン

金谷健一

群馬大学工学部

「コンピュータビジョン(CV)」とは画像からそこに写っている3次元実世界に関する情報を引き出す技術のことである。入力画像が2次元の文字や記号やボタンであるときに「ボタン認識」であり、数値や記号の論理演算としてさまざまな識別、分類のプログラミング技法が開発された。しかし入力画像が「3次元実世界」となると実世界の物体と画像との対応関係にはさまざまな物理的、光学的、電気的、統計的な要因が関係し、ボタン認識と比べて必要な知識が格段に増加する。したがってCVには次の要素が必要である。

CV = (画像) + (プログラミング) + (数理解析)

ここでいう「画像」とは入力装置(ビデオカメラ)や出力装置(ディスプレイ)を始め、変換、伝送、記憶、その他すべての画像信号や画像データの処理を含めたものを意味する。「プログラミング」とは数値や記号の処理アルゴリズムだけでなく、参照するデータベースやシステムアーキテクチャまで含んでいる。「数理解析」とは2次元・3次元の幾何学的関係、物理法則、光学的関係、濃淡・色、誤差、精度・信頼性を初め、現象のモデル化とその解析であり、正にCVの核心である。

「画像」と「数理解析」を除いて「プログラミング」の部分の独立させたものが「人工知能」ともいえる。初期のCVはこの人工知能的プログラミング技術を画像と結合させたものであり、設計者の直観と思い付きからさまざまなルールやストラテジーを試行錯誤的に導入し、システムを試作してはその性能を実験的に確認していた。

その後、実世界の複雑な現象を基本要因に分解し、それぞれを厳密に解析するという科学的方法論が取り入れられるようになり、ようやくCVは思い付きの発明工夫から客観的な科学技術となろうとしている。

今後望まれることは「要素技術の体系化」である。ともすると目的ごとに都合がよさそうに思える技法を工夫してシステムを組み立てることが多いが、目的の異なるシステムでも要素技術として共通するもの

が多い。それらを取り出し、前提を明確にし、効率、性能、最適などを理論的に解明することが必要である。

これは従来から行なわれてはいたが、過去のCVの研究では各種の処理を効率的に実行するアルゴリズムやデータ構造の工夫が大きな部分を占めていた。これは利用できる計算機の能力が限られていて、資源を最大限に活用する必要があったからであるが、ともすると効率に目を奪われて、その処理が実世界と画像との関係においてどういう「客観的な意味」を持つのかを十分に考慮しないことが多かった。

例えば「エッジ検出」を「画像の濃淡値から“人間が目で見えて物体境界だと思ふ曲線”を抽出する方法」と主観的にとらえると、脱出不可能な迷路に落ち込む。主観を排し、どういう濃淡ボタンに対してどういう曲線を抽出するのかを客観的に定式化しなければならない。そもその「実世界のあるもの(たとえば人物)を検出する手法」などは存在しない。どんな検出手法でも、その手法で検出できない場合がある。追及すべきものは「これこれの条件のもとではこれこれの性質を満たすものを抽出する」というように客観的に定義された手法である。

従来は存在しないまぼろしの手法を追い求めてさまざまな処理を考案しては実験を行ない、主観評価により「多くの場合」に有効であるなどと主張することがよくあった。しかし、たまたま試した実験例に対して80%有効な手法より、30%しか有効でなくても有効である条件とその性質が明示されている手法のほうが科学技術として意味がある。その条件が満たされる場合に使用すればよいし、逆にその条件が満たされるようにコントロールを加えることもできる。

例えば「領域分割」をニューロや分散協調処理を組み合わせた「統合処理」で実現して「多くの場合」に有効であっても、それがどういう条件のもとでどういう性質をもつ領域を出力しているのかを客観的に明示できないものは科学技術と呼ぶことはできない。そのような、実世界と画像との関係に対する客観的な定式化なしに試行錯誤するという「主観主義」を排することが、今後の科学技術としてのCVの発展のためには必要である。

† 376 桐生市天神町 1-5-1 群馬大学工学部情報工学科
Tel: (0277)30-1841, Fax: (0277)30-1801
E-mail: kanatani@mail.cs.gunma-u.ac.jp

これからのCV

上田博唯

日立電子(株) 開発研究所

1 はじめに

CV研究には大きく理論研究派とシステム思考派とでもいう二つの立場がある。前者の立場では画像を解析してシーンを自動記述し尽くすことが目的となるのに対し、後者ではコンピュータを如何に使いこなして優れた実用システムを実現するかが主眼点となる。これらを対比しつつ過去を振り返り、これからのCV研究のあり方について考察してみたい。

2 四半世紀を振り返って

筆者は企業の研究所に身を置き、二十数年画像認識の分野で研究開発を行ってきた(従って自分自身は自然とシステム思考寄りとなった)。

1970年頃のCVはまだ若い研究分野であり、理論も実践もひとりの研究者が兼ね備えるのがむしろ普通であった。当時の筆者のテーマは無人化を目的とする工業応用であり、理論研究的な面を追及すると実用的な処理速度が得られず妥協することも多くあった。それは若干の劣等感をともなうものであったけれども、人間に取って代わって組み立てや目視検査をより高速により正確に実行できるシステムの実現は充実感のある仕事であった。ところが数年のうちに無人化できることはやりつくし、後には極度に困難な研究課題ばかりが残っているような状況となった。CVの理論、技術、ハードウェア、全ての能力が人間のそれに比べてはるかに劣っているのだから当然のことである。この頃から研究者は、純粋に理論研究に向かうか、システム思考に向かうかの選択を迫られるようになったように思う。

1980年代になり、理論研究ではMarrなどによる新しい枠組みが提案され、再び種々の進展があった。実用へのこだわりを一時的に放棄したことによって、理論の洗練が捗ったのかも知れない。一方のシステム思考派は、如何にしてコンパクトなシステムで、より高度な問題を実時間で処理できるようにするかに多大な苦勞をしていた。筆者の場合は、やっと出回り始めた16ビットマイクロプロセッサでどの

程度の画像処理ができるかの研究を手始めに、画像処理用DSPとこれを用いた文書画像処理システムの開発、そしてこのDSPを64個用いたマルチプロセッサシステムの開発に従事していた。

1990年代になると、メモリやハードディスクの高密度化が進み、コンピュータの能力も大きく向上したために、画像処理における制約がかなり緩和された。これによって、理論研究派、システム思考派ともに大きく新しい展開を迎えることとなった。それとともに両者の立場の差も、くっきりと見えるようになってきた。前者がコンピュータに与えた知識のみでの解析に力点を置くのに対し、後者ではシステムに人間を積極的に含めるようにした研究アプローチが多くなったように思われる。

3 これからのCV研究

こうやって見てきた理論研究派とシステム思考派であるが、ときにはお互いをけなしあう場面もよく見受けられた。シミュレーションでのみ正しく動作する理論、トイ・プログラムに対し、理論的境界を説明できないシステム、大本営的発表等々。しかし最近になって、CVの研究対象がマルチメディアやヒューマンインターフェースに広がってきたことにより、状況が変わりつつある。明確な理論が無ければ複雑・高度化したシステムが設計できないシステム思考派と、システムの実運用からのフィードバックがないと研究が進められない理論研究派という構図が出てきたのである。

認知科学的アプローチ、すなわち人間を情報処理システムとみなし、その情報処理過程に合わせてシステムを設計することなど、学際的な試みも増加している。考えてみれば人工知能の一分野であるCVは人間の知能を工学的に実現することを目的としているから、人間のこともっと知る必要があるのは当然である。

これらの動きの中から、また新しいCV研究の飛躍が生じるであろうと期待している。

視野を広げよう、現実世界を見つめよう

松山 隆司

京都大学大学院 工学研究科 電子通信工学専攻

1 広い視野に立った CV 研究

「CV とはどんな研究をする分野なのか？」

私の考えでは、本来 CV は、情報科学の立場から視覚認識の問題を考える研究分野であり(事実 CV フォーラム'82 ではそうだった)、1980年代から主流になったいわゆる Computational Vision (数理的手法による2次元画像からの3次元幾何学、光学、時間的情報の復元)はその1つのアプローチにしか過ぎない。Computational Vision の成功により、Computer Vision = Computational Vision という見方が定着しているのは事実であるが、そうした1つの見方にだけこだわり続けることは、研究活動を蝸壺に追い込み、返ってその発展を阻害することになりかねない¹。

そこで、まず狭まった視野を本来のものに戻し、今後は広く視覚認識の問題を考えることにしたい。すなわち、図1に示した、知能システムの構成という視点から視覚認識の問題を考えて行こうというわけである。最近活発に研究されているアクティブビジョンや分散協調型画像理解は、こうした視点から視覚認識の問題に迫るアプローチであるといえる [1]。

次にもう1つ別の方向への視野の拡張として、図2に示した、視覚情報メディア処理という視点から視覚認識問題を考えるものがある。本研究会が新たに目指そうとしているイメージメディアの処理は、この視点に立ったものである。

2 現実世界で有効に機能する CV アルゴリズム・システム

「CV は役に立たない？」

一般に情報科学・工学はモデリングの学問であるといえる。モデリングにおいて最も重要なのは、モデルが現実世界の問題をどの程度うまく表現しているかということであり、常にモデルと現実世界との整合性を調べる実験・検証が必要である。また、モデル自身の持つ性質を研究する場合も絶えず現実世界を踏まえた議論が必要である。CV が上のような批判を受ける一因は、現実世界を横に置き、モデル自身に関する議論に、はまりこんでしまったためであろう。

現実世界で有効に機能する CV アルゴリズムやシステムを構築するための基本方針として、最近我々は次のようなものを考えている。

1. まず画像を撮ることから考えよう。：通常のカメライは人が見るための画像を撮ることを目的として設計されており、多くの「くせ」や限界がある。そこでまず第一に、CV のための画像の撮り方を考えなければならない [2]。

¹だからといって、Computational Vision の研究を続ける必要がないといっているわけではない。

2. タフでなければ動まらない。：高精度化や高速化も重要であるが、頑健性の実現を第一に考えたアルゴリズム、システム開発を考えるべきである [3]。
3. スペードのエースは存在しない。チームプレイで対応しよう。：1つのモデル・アルゴリズムによって、多様な現実世界の状況すべてに対処することはできない。多様な状況に柔軟に適應する機能を実現するには、適材適所でモデル・アルゴリズムを切り換えるためのシステム構成を考える必要がある。分散協調型画像理解の研究 [4] はこうした考えによる。

- [1] 松山:AI マップビジョン研究から見た統合アーキテクチャ, 人工知能学会誌, Vol.10, No.6, pp.888-894, 1995
[2] 松山:多重画像の統合-高精度画像計測と多機能画像生成-, 電子情報通信学会誌, Vol.79, No.5, 1996
[3] 松山:多角的視覚情報の統合, 電子情報通信学会誌, Vol.74, No.4, pp.349-353, 1991
[4] 松山:分散協調処理による画像理解, 計測と制御, Vol.31, No.11, pp.1149-1154, 1992

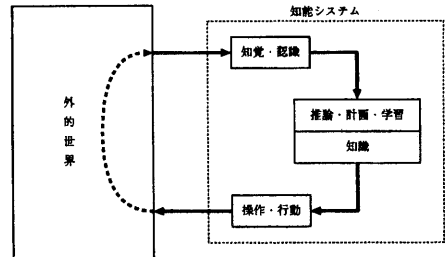


図1: 知能システムの構成

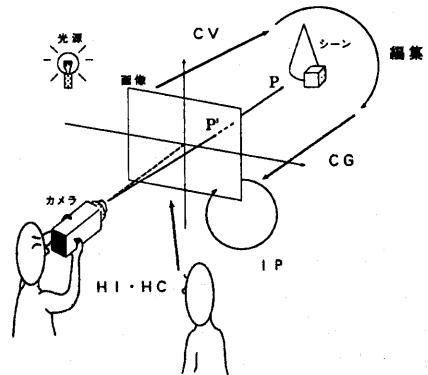


図2: 視覚情報メディア処理の世界