

コンピュータとヒューマンビジョンとの接点

長尾 智晴

東京工業大学 工学部 像情報工学研究施設

〒226 横浜市緑区長津田町 4259

1. はじめに

計算機を用いた画像の認識・理解を目的としたコンピュータビジョン (Computer Vision; 以後 CV と略記) に関する研究が, 30 年以上前から国内外で盛んに行なわれている^{1,2)}. 特に近年, 高性能ワークステーション, CCDカメラやイメージスキャナなど, 計算機および画像関連の周辺機器の高性能化・低価格化に伴って, 情報系の研究者が画像を扱うことが容易になり, CV 関連の研究者の数が増えてきている. 印刷文字や郵便番号の認識, 製品製造プロセスでの不良品検出など, CV 技術が実際に産業の一部として実用化されている例も少なくない. しかしながら, CV の最終目標を人間の視覚 (Human Vision; 以後 HV と略記) 機構と同等 (以上) の機能をもつ機械による視覚の実現であると考えれば, 現状の到達度は, まだまだ低いと言わざるを得ない. CV 研究の発展には, 人間の視覚および脳の理解が不可欠であり, HV 研究による成果から得られた知見の参照が非常に重要である. また逆に, CV 技術の HV への何らかのフィードバックも考えられる. そこで本稿では, CV の現状と問題点を示し, HV との比較を行なうことによって, CV と HV との接点を見出すとともに, 今後の CV と HV との新しい相補的な関係へ向けての展望について述べる.

2. コンピュータビジョンの現状と課題

2.1 CV 研究の特徴

前記のように, CV 研究の最終目標は機械による視覚機構の実現であるが, はじめから視覚

機構の構築を目指すことはあまりに困難であり, 工学的アプローチの限界を超えるものである. このため CV では, 処理対象や状況を限定するとともに, 利用できる先見的知識はできるだけ有効に利用しようとする立場をとる場合が多い. また現状では, ごく一部の例外を除いて, 一般的なプログラム蓄積型計算機を用いており, そのプログラムは研究者によって試行錯誤的に開発されている. これらのことから, CV 研究の一般的な特徴として, 次の2点が挙げられる.

(1) 処理対象が限定されており, 他の対象には適用不能な場合が多い.

(2) アルゴリズムの蓄積はあるが, それらが理論的・体系的にまとめられていない.

このような特徴をもつ CV について, 2.2 および 2.3 で対象別に研究の現状を示す.

2.2 文字・図形認識 (2次元CV)

元来2次元のデータを対象とした CV として, 文字・文書認識, 図面及び平面図形などの認識分野がある. この分野での CV の現状と課題を表1にまとめて示す.

表1中の各認識対象について次に述べる.

(1) 印刷文字

印刷文字を認識対象にする場合, 入力文字のフォントが特定できれば, 認識は比較的容易であり, 計算機による高速な処理が実現できる. 基本的には, 入力文字と, 計算機内にあらかじめ用意した各文字の標準パターンとの間の重ね合せによるテンプレートマッチングを行なう. しかしながら, フォントが特定できない場合や, 入力文字に多くのノイズが含まれる場合は

認識率が低下する。

(2) 手書き文字

手書き文字では個人差が大きいため、各文字の標準パターンを定義することができない。このため、入力文字に対する何らかの幾何学的な特徴抽出を行ない、特徴空間でのクラスタリングによって対象を認識する必要があり、認識率は印刷文字に比較するとかなり低下する。また、走り書きでは、各文字の切り出しも問題になる。一方、筆記過程の情報を利用するオンライン認識では、文字の書き順や部首による認識が可能であるため、楷書体であれば高い認識率が期待でき、実際に製品化されているものもある。一般に、計算機による文字認識の研究は終了したかのように理解されていることが多いが、実際には解決されていない課題が山積しているのが現状である。

(3) 文書

文章、図表、画像の混在した文書画像に対しては、はじめに原画像を各構成要素に分割する領域分割を行なう。文章の領域については、行を切り出した後、各文字を分離して認識を行なう。例えば名刺や書式の定まった書類のような対象については、かなり精度良く認識することが可能であるが、領域分割自身が困難な対象も多い。また、封筒に書かれた住所など、手書きの部分は各文字の切り出しが困難である場合が

表1 文字・図形認識での CV の現状と課題

認識対象	現状	課題
印刷文字	フォント既知・ノイズ小ならほぼ100%の認識率	マルチフォントの認識とノイズの影響の低減
手書き文字	楷書なら、オフライン、オンラインともにかなり認識できる	走り書きは文字の切り出し・認識とも困難
文書	書式が大体既知なら文書の読み取りは可	文章、図、表、画像などの領域分割
図面	書式が厳密な回路図等は実用段階だが地図や既存図面は困難	他の図形と重畳した文字や図形の切り出しが困難
平面図形	輪郭線情報などを用いた特定図形の切り出しが可能	図形の自由度が大きく、他の図形と重畳があると困難

多い。

(4) 図面

書式が厳密な図面に対しては、直線・曲線のベクトル化³⁾、文字・図形の切り出しが比較的容易であり、高い認識性能が期待できる。一方、地図などのように、図形間の重畳が多い対象や、書式が不明な図などは認識が困難である⁴⁾。

(5) 平面図形

平面図形の認識処理は、例えば工場内での部品選択、航空写真の解析などで必要となる技術である。ここでは認識対象のモデルの記述と対象画像に対する画像処理が問題になる。これらの処理は一見容易に思われるが、実際には困難な課題である⁵⁾。

以上のように、対象、書式や画像取得環境を限定した場合は、当初設定した処理目的を満たす手法やアルゴリズムが多数開発されている。しなしながら、対象や状況の変化、ノイズなどの影響により、認識率は一般に著しく低下する傾向がある。

2.3 シーン理解 (3次元 CV)

処理対象の画像が、屋内や屋外のシーンである場合、対象が2次元図形の場合よりも認識が一般に困難になる。シーンの理解の問題は、設定状況や認識対象、処理目的が表2に示すように多種多様であり、汎用的な手法は存在しない。

一般的な前処理として、計算機に入力された原画像に対して、ノイズ除去や平滑化、エッジ強調などのフィルタリングを行なった後、意味のある領域ごとに画像中の領域を分割するセグ

表2 シーン理解での状況や目的の例

設定状況	認識対象	処理目的
・ 計算機内の仮想環境 ・ 部屋の中 ・ 廊下 ・ 工場内 ・ 屋外 ・ 道路上 ・ 交差点 など	・ 積み木などの単純な多面体 ・ 関数による形状 ・ 任意曲面の物体 ・ 具体的な対象 (人間、顔、荷物、自動車等) ・ 不特定の対象 など	・ 対象の識別 ・ 対象の抽出 ・ 対象の形状測定 ・ 対象の動き認識 ・ 3次元環境認識 ・ ロボットの行動制御 ・ 視点移動制御 など

メンテーションを行なうことが多い。しかしながら、それ以降の処理は千差万別である。ここでは、表2中の処理目的別に現状と課題について述べる。

(1) 対象の識別

認識対象の3次元物体が孤立図形として与えられた場合に、それが既知の物体のどれであるか、あるいは未知の物体であるかを判定する処理である。この処理は、画像処理による対象物体の切り出し、物体からのエッジ抽出、面抽出などの特徴抽出の精度に大きな影響を受ける。実際のシーンをCCDカメラなどで撮影して得た実画像は、計算機内で生成したノイズを含まないテスト用画像とは比較にならないほど処理が困難である。また、積み木のような単純な物体とは異なり、形状特徴が明確に定義できない、言い替えれば、対象物体のモデルとしてどのような情報を用いればよいかわからないような対象、例えば人間の顔や任意の曲面から構成される物体などの認識は非常に困難である⁶⁾。

(2) 対象の抽出

(1)とも関連するが、背景が複雑な場合や他の物体と混在している場合は、対象を画像から切り出す必要があり、この処理自体が一般に困難である⁷⁾。セグメンテーションによって、物体ごとの明確な領域分割が行なわれた場合は、対象の抽出は比較的容易であるが、一般に領域の境界は不明確であり、対象を精度よく切り出すことができない場合が多い。

(3) 対象の形状測定

対象の3次元データを取得することを目的とした形状測定を行なう場合は、左右2台のステレオカメラを用いたり、対象を回転台に乗せてさまざまな向きから撮影したり、スリット光やレーザー光などを対象に投光するなどの専用装置を用いる。このような専用装置を用いる場合は、精度よく形状計測を行なうことが可能である。しかしながら、そのような装置を用いない場合は、対象の形状に対する先見的な知識を有効に利用するなどの工夫が必要であるとともに、一般に精度の高い計測は期待できない。

(4) 対象の動き認識

動画像中の連続するフレーム間の差分情報から、画像中の動きの場を表わすいわゆるオプティカルフローや、物体の移動方向・速度・加速度などを求める動画像処理手法が提案されている。対象が剛体でない場合の各部分の動きの解析や、移動物体が多数ある場合のそれぞれの分離⁸⁾など、解決されていない課題も多い。

(5) 3次元環境認識

あらかじめ空間配置の既知な複数台のカメラ、あるいは移動可能なカメラを用いて撮影した一連の画像から、カメラの周囲の環境の3次元構造を認識する処理である。これまでに、実際に自律ロボットに搭載したカメラによって得た画像列から、部屋内の3次元地図を作成する試みなどがなされている。しかしながら、環境が複雑な場合は一般に処理が困難である。

(6) ロボットの行動制御

自律ロボットが3次元環境を認識しながら自分自身の移動方向や速度を決定する問題であり、カメラから得た画像の処理の他に、行動を制御するための戦略を設計する必要がある。一般にはif～then～形式のプロダクションルールがよく用いられるが、それらのルールの優先順位の設定や行動の評価など、人工知能分野の技術が必要とされる。

(7) 視点移動制御

人間は常に視点を移動させながら有効な認識処理を行なっているが、それを機械に行なわせようとするものであり、局所的な高解像度視野と、大局的な低解像度視野を併用するとともに、それらの動きを制御しようとする試みが行なわれている。

2.4 今後の研究方向

2.2, 2.3で述べた現状と課題をもとに、CVをより学問的に成熟した分野へと高めるため、あるいは汎用的な方法として体系化するための最近の考え方について述べる。

(1) 逆問題としての定式化

Marr⁹⁾が提唱したように、画像の認識は2次元の画像からの3次元空間の構築であり、数

学的には3次元から2次元への写像の逆問題であるとみなす考え方がある。そして Shape from disparity, Shape from shading, Shape from texture, shape from motion など、一般に Shape from X 法と呼ばれる各種の方法論が提唱されている。これらはいずれも数学的には美しい問題の定式化であり、計算機シミュレーション上は有効であるが、実画像では成立しない場合が多いため、他の手法との併用が必要である。

(2) 能動的認識

自律ロボットを意識した方法論であり、与えられた画像を受動的に認識しようとするのではなく、情報が足りない場合に別の角度から見た画像を積極的に得て、認識に用いようとするもので、タスクオリエンテッドビジョン¹⁰⁾と呼ばれる。この場合は、どの角度から見れば効率が良いかというカメラコントロールの問題を扱う必要がある。利用できる複数のセンサの統合を行なうマルチセンサフュージョンとも関連が深い。

(3) 最適化問題としての定式化

ステレオ画像での左右対応点の決定での弛緩法の利用や、エネルギー最小化を用いた物体の輪郭線の抽出¹¹⁾などの例に見られるように、CVでの問題を一種の最適化問題としてとらえようとする試みがある。また、近年情報科学の分野で注目されている、生物の進化プロセスから着想された最適化法である遺伝的アルゴリズム (Genetic Algorithm; GA)¹²⁾に代表される進化的計算法 (Evolutionary Computation) を用いて、CVをとらえようとする試みがある¹³⁾。これらは、アルゴリズムの蓄積にとどまっていたこれまでのCVを、最適化という考え方で体系化できる可能性を秘めていると考えられる。

(4) 脳研究との関連

先に述べたように、CVは突き詰めて考えれば人工脳の設計問題となり、非常に学際的な研究領域である。近年、進化的計算法を用いて、神経細胞間が任意に結合された神経回路網を自動的に生成するための研究¹⁴⁾が行なわれており、CVとしての機能をもつ神経回路網の設計

に対する応用が期待されている。いずれにしても、CV研究は、神経回路網研究や人工知能研究、さらに視覚心理物理、脳に関する医学的・生理学的研究と非常に密接な関係がある。今後は、これら関連分野の知見の有効利用、および協力体制の強化が必要である。

3. CVとHVの接点

ここでは、2.で述べたCVと、HVとの関係について考える。

3.1 ハードウェア的観点

はじめにハードウェア的観点による両者の比較を表3に示す。ただし、ここではCVを自律ロボットの視覚とし、HVを視覚系を含む脳および人間としてとらえている。

近年、LSIの高集積化と駆動系の小型化に伴い、超小型の自律ロボットに何らかの作業を行なわせるマイクロマシンに関する研究が、ロボティクスでの一分野を形成している¹⁵⁾。また、人間の視覚機構を模した画像撮像装置、超小型CCDステレオカメラによる視点移動装置などが開発され、視覚以外の各種センサの高性能化も著しい。このため、情報処理部であるコンピュータによるソフトウェアが優れたものになれば、CVの機能をHVにさらに近づけることができ、自律ロボットが担うことができる作業が大幅に拡張され、産業および社会に大いに貢

表3 ハードウェア的観点によるCVとHVの比較

	コンピュータ ビジョンCV [自律ロボット]	ヒューマン ビジョンHV [人間]
組織	無機質 (可塑性なし)	有機体 (可塑性あり)
画像入力部	CCDカメラ等 による撮像系 (低解像度)	眼球および 視細胞・視神経系 (高解像度)
視点制御系	モータ等駆動系	筋肉など
情報処理部	コンピュータ	視覚神経系・脳
視覚以外の センサ	距離センサ、集音 器、速度センサ等	聴覚、触覚、 味覚等
作業	マニピュレータ	腕・手
移動装置	車輪・歩行装置	足

献することが期待できる。

3.2 ソフトウェア的観点

次に、ソフトウェア的観点による CV と HV の関係について表 4 に示す。

表 4 からわかるように、CV では HV にとって負担が大きい、単純かつ大量の作業を高速に処理することができる利点がある反面、柔軟な判断や、学習・適応などの高度な情報処理については、現状では HV に比較してかなり劣っているとと言える。神経回路網に代表される並列分散処理の理論が確立され、進化的計算法などの強力な最適化法によって、任意の機能をもつ神経回路網を自動的に設計することが容易になれば、従来は困難であった高度な情報処理機能をもつ CV を実現することが可能になると考えられる。

3.3 将来展望

これまでに CV の現状と問題点、HV との比較について述べた。そして現状の CV 技術によって実現されている機能は、HV の機能のほんの一部に過ぎず、我々が日常生活において無意識的かつ容易に行なっている HV の高度な視覚情報処理には遠く及ばないことを示した。現状では、CV 研究と HV 研究の間に明確な境界

表 4 ソフトウェア的観点による CV と HV の比較

	コンピュータビジョン CV [コンピュータソフトウェア]	ヒューマンビジョン HV [視神経系を含む脳内情報処理]
処理方式	逐次方式	並列分散方式
制御方法	コンピュータプログラム	神経回路網
処理速度	CPU自身の演算速度は高速。全体の性能がCPUに大きく依存。	神経細胞間の信号伝達速度は低い。回路網である脳の処理速度は超高速
単純作業	高速	低速
高度な視覚情報処理 (学習・適応・推論等)	劣っている	優れている

があると言わざるを得ない。これは、前者の目的が視覚機構の工学的応用であり、後者の目的が視覚機構の解明であることにもよるが、視覚神経系を含む脳があまりに複雑な機構であり、一方で得られた知見を、他方に直接的に応用することが困難であるため、他方の研究に興味を持たない研究者が、前者・後者を問わず、非常に多いからであると考えられる。しかしながら、視覚機構の解明と、その研究成果の社会への還元のためには、両分野の研究者の接近と相互理解が不可欠である。これまでも、局所視野と大局視野の相補的利用に着目して開発された神経回路網であるネオコグニトロン¹⁰など、視覚心理物理で得られた知見の工学的応用の成功例がいくつかある。将来はさらに密接な相補的關係の確立が期待される。また、HV に関する研究成果の CV への応用だけでなく、その逆も考えられる。例えば、心理物理実験を通じた考察によって得られたモデルを、神経回路網などの工学的手段を用いて実際に設計することによって、何らかの機能の存在や未知の現象を予測することができたとすれば、それは非常に建設的で興味深いことではないだろうか。

4. まとめ

本稿では、コンピュータビジョン (CV) の現状と問題点、ヒューマンビジョン (HV) との比較、ならびに両者間の関係の将来展望について述べた。今後、両分野の研究者の交流が今まで以上に活発化し、学際領域である視覚機構の解明とその工学的応用がますます発展することを期待したい。

文 献

- 1) 安居院 猛, 長尾智晴: 画像の処理と認識. 昭晃堂, 1992.
- 2) 小川英光: パターン認識・理解の新たな展開. 電子情報通信学会, 1994.
- 3) 長尾智晴, 安居院 猛, 長橋 宏, 呉 永煥: 適応ブロック分割法によるラスター・ベクタ変換. 画像電子学会誌, 21, 310-316, 1992.
- 4) 長尾智晴, 安居院猛, 中嶋正之: 塊状図形分離と

- キーキャラクタ探索による2万5千分の1地形図からの文字列抽出. *テレビジョン学会誌*, 44, 1044-1052, 1990.
- 5) T. Nagao, T. Agui and H. Nagahashi: An extraction method of partially occluded objects utilizing partial shape vectors. *Transactions of IEICE*, E73, 410-417, 1990.
 - 6) 安居院 猛, 中嶋正之, 長尾智晴: 静止濃淡情景画像からの顔領域の抽出. *電子情報通信学会論文誌*, J74-D11, 1625-1627, 1991.
 - 7) T. Nagao, T. Agui and H. Nagahashi: 2D and 3D object recognition using a genetic algorithm. *Machine Graphics and Vision*, 3, 275-286, 1994.
 - 8) 安居院 猛, 長尾智晴, 斉藤智明, 中嶋正之: ブロック特徴を用いた歩行人物の抽出と追跡. *テレビジョン学会誌*, 45, 1213-1220, 1991.
 - 9) D. Marr: *Vision* ビジョン. 産業図書, 1987.
 - 10) 池内克史: タスクオリエンテッドビジョン. *電子情報通信学会誌*, 74, 360-365, 1991.
 - 11) M. Kass, A. Witkin and D. Terzopoulos: SNAKES. *International Journal of Computer Vision*, 1, 321-331, 1988.
 - 12) 安居院 猛, 長尾智晴: ジェネティックアルゴリズム. 昭晃堂, 1993.
 - 13) 長尾智晴, 安居院 猛, 長橋 宏: 遺伝的手法を用いた2値図形のパターンマッチング. *電子情報通信学会論文誌*, J76-D11, 3, 557-565, 1993.
 - 14) T. Nagao, T. Agui and H. Nagahashi: Structural evolution of neural networks having arbitrary connections by a genetic method. *Transactions of IEICE*, E76-D, 689-697, 1993.
 - 15) T. Fukuda and T. Ueyama: Cellular Robotics and Micro Robotic Systems. *World Scientific Series in Robotics and Automated Systems*, 10, World Scientific, 1994.
 - 16) K. Fukusima, S. Miyake and T. Ito: Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on SMC*, SMC-13, 826-834, 1983.