

## 展 望

## 人と共存し発展するロボットのためのビジョン

Vision Studies for Robots Supporting Humans

石 黒

浩\*

\*大阪大学基盤工学研究科知能・システム創成専攻

Hiroshi Ishiguro\* \*Osaka University

## 1. ロボットビジョンの起源

## 1.1 ロボットビジョンとアクティブビジョン

ロボットビジョンという言葉が最初に使われたのは、筆者の経験の中では、Horn による著書 Robot Vision [1] である。幾何学的、計算論的なコンピュータビジョンの教科書として、当時は多くの読者に読まれた。しかしながら、今に言うロボットビジョンとは多少意味が異なる。コンピュータビジョンにおいて、しばらくは Ballard のコンピュータビジョン [2] が有名だったが、その内容は、どちらかというところ人工知能を意識した内容であった。Horn の場合はそれとは異なる意味で Robot Vision という言葉を使っていた。ロボットのための視覚機能として、新たな考え方が提案されたのは、Aloimonos の Active Vision [3] や Ballard の Animate Vision [4] に始まる。無論、それまでも、Machine Vision 等の呼び名で、工場で働くロボットの視覚機能の研究はされていたが、より、人間やロボットに近い視覚機能の実現を目指すという意味では、Active Vision や Animate Vision が最初であろう。

Active Vision や Animate Vision の基本的な考え方は、得られた画像を解析するだけでなく、解析しやすいように、カメラを能動的に移動させるというものである。Active Vision では、カメラの運動によって、観測者中心の座標系から物体中心の座標系への変換が容易になること等が示されている。また、Animate Vision では人間のような注意制御を用いることの重要性も指摘された。

そして、特に日本においてロボット研究が盛んになるにつれ、ロボットの視覚として、いわゆるロボットビジョンとして、特にロボット関係者には Active Vision や Animate Vision の研究の方向性が支持されるようになった。その根底には、視覚認識の研究には、認識モデルを獲得するための身体が必要であるという考え方がある。経験することができる身体を持ち、経験に応じて視覚を制御することによ

り、より高度な認識が実現できるのである。

## 1.2 コンピュータビジョンとロボットビジョン

しかしながら、このようなロボットビジョンは、本当に視覚の問題を解きやすくしたのであるか？ 例えば、Tomasi [5] らによって提案された、shape and motion without depth の研究とその発展では、従来よりも性能のよい視覚認識結果を示しただけでなく、直交射影という仮説を適用するために、物体を部分的な固まりとしてとらえるというアイデアも提案されており、それは注意制御にも似ている。また、顔認識機能も、ロボットという概念をまったく用いることなく、今では実用レベルにまで発展してきた。特に最近のデジタルカメラに搭載されている画像処理機能には目を見張るものがあり、人間の顔を検出する機能や、高感度の CCD を用いて、常に安定した画像が撮影できる機能が実現されている。

すなわち、従来のコンピュータビジョンにおいても、その機能は徐々にロボット研究で必要とされる人間の機能に近づきつつある。

一方で、ロボットを意識したビジョン研究は、独自の発展を遂げただろうか？ 1990 年代において、移動ロボットやマニピュレータの研究に盛んに視覚センサが使われるようになってから、特に、ロボットビジョンという言葉が耳にするようになった。しかしながら、コンピュータビジョンからロボットビジョンに代わって、従来のコンピュータビジョンで解けなかった問題が解けるようになったかというところ、必ずしもそうは言い切れない。むしろ、ロボットビジョンとして取り組むことの意義は、従来のコンピュータビジョンの研究において、実時間性やロバスト性の向上を意識することにあった。

コンピュータビジョンは、応用を目指したパターン認識を基本とする研究と、幾何学を基本とする計算論的な研究に分けられる。Horn は従来の人工知能的なコンピュータビジョンと区別するために、後者の意味で、ロボットビジョンという言葉を使ったようであるが、その他の研究では、むしろ逆で、ロボットビジョンは、ロボットの視覚機能として、より実用性の高いコンピュータビジョンの研究を目指してきた。

原稿受付 2009 年 4 月 3 日

キーワード: Robot Vision, Active Vision, Distributed Vision, Panoramic Vision

\*〒 565-8531 豊中市待兼山町 1-3

\*Toyonaka-shi, Osaka

本稿では、ロボットビジョンの初期の研究から最近の研究までを傍観しながら、今後ロボットビジョン研究の展開と重要となるであろう研究テーマについて議論する。

## 2. リアルタイムビジョンとパノラミックビジョン

### 2.1 リアルタイムビジョン (real-time vision)

1990年ごろから盛んになったロボットビジョンの象徴的な研究は、Active VisionとPanoramic Visionであろう。Active VisionのコンセプトはAnimate VisionやActive Visionとして、アメリカから発信されたが、ほぼ同時期に、日本でも、ロボットの視覚機能として注目すべき研究が展開されていた。東京大学のグループを中心にしたリアルタイムビジョンの研究である。

國吉ら[6]は、人間の行為をロボットにまねさせるために、人間の行為を実時間で観測する実時間の視覚システムを開発した。背景差分を元に、手を検出し、物体との接触情報を実時間で認識するシステムで、物体をつかむ、話す、物体を積み上げる等の行為を認識することができた。コンピュータビジョンとしての情報処理そのものは、それほど複雑ではなく、むしろ単純であったが、実時間で認識できるシステムの特徴をうまく生かすことができている。

このリアルタイムビジョンの研究で重要なのは、高速な処理によって連続的に画像を処理することにより、一回一回の視覚情報処理は単純であっても、時間軸にそって状態変化をとらえることにより、ロバストな認識が可能になるという点である。例えば、物体をつかむという行為を1枚の画像を元に認識するなら、画像の中から、手を探し出し、手の形や物体の形を正確に認識しなければならず、それでもまだ、それは物体をつかもうとしているのか、物体を手放そうとしているのかという曖昧性が残る。一方、手を見つけた後、実時間でその手を追跡すれば、手を見失うこともなく、また、物体との接触関係から、行為を比較的正确に認識できる。

すなわち、時間とともに変化する状況を、必要最小限に捕らえていくことによって、認識するというアプローチである。これはアクティブビジョンにおける注意制御に通じる考えであり、まさにその注意制御を実現したものである。

### 2.2 パノラミックビジョン (panoramic vision)

リアルタイムビジョンの研究と同時期に、ロボットの視覚として研究が盛んになったのは、筆者もかかわったパノラミックビジョンや全方位視覚 (omnidirectional vision) と呼ばれる研究である。この研究は大阪大学のグループを中心に盛んになり、その後、ワークショップにも発展し、今なおそのワークショップは続いている。

アクティブビジョンやリアルタイムビジョンが、狭い視野を高速に処理することを目指したのに対して、パノラミックビジョンの研究は、広い視野情報を処理することによ

ってもたらされるメリットを模索した。

人間は特定の物を注視する行動だけでなく、環境全体を見渡して自分の位置を調べるという行動もとる。パノラミックビジョンはまさにそのような環境を広く見渡す行為をロボットに取り入れるために研究された[7][8]。

リアルタイムビジョンが、視野を限定して時間方向に情報を探索したのに対して、パノラミックビジョンは、視野を空間的に広げるアプローチを取った。そして、リアルタイムビジョンがロボティクスにおいて物体を操作するマニピュレーションの研究に用いられたのに対して、パノラミックビジョンは、環境を認識する機能を元に、ナビゲーションに用いられた。

パノラミックビジョンの特殊なケースである全方位視覚のメリットは、移動に伴って画面上で観測されるFOE (Focus Of Expansion, 動きの無限遠点) が一つだけでなく、二つとも観測できることである[9]。この性質を利用すれば、環境に対してロボットを正確に誘導することができる[8]。さらに、全方位視覚にはそれが水平方向に周期信号であるという強い性質がある。この性質により、全方位画像はフーリエ級数展開により、方位とは独立な成分 (強度成分) と方位に依存する成分 (位相成分) に分解できるとともに、それらは、視覚のみをもとにしたロボットの誘導を実現する[10]。

### 2.3 ハードウェアの開発

ロボットビジョンはロボットのための視覚認識技術である。故に、独自のハードウェアや光学装置が開発されている。アクティブビジョンでは、人間の網膜構造に似せて、Log-polar mappingの視覚センサ[11]が、開発されたり、中心付近が周辺よりも拡大されている、魚眼レンズに似たレンズが開発されてきた。これらは、オプティカルフローの計算や物体追跡においてメリットを持つ。

一方、パノラミックビジョンでは、全方位の視覚情報を一台のカメラで得るための全方位カメラが数多く開発されている[12]。このような全方位カメラを用いれば、カメラを回転させることなく、瞬時に全方位の視覚情報を得ることができる。

これらの光学系ハードウェアは、画像処理の効率化に大きな効果をもたらしている。ロボットビジョンでは、このような光学系の設計を伴いながら、実時間性やロバスト性を実現してきた。

## 3. 近年のロボットビジョン

### 3.1 分散視覚 (distributed vision)

視覚認識の技術の発展に伴い、ロボットそのものも開発が進んできた。特にロボットの分野では、工場や特殊な環境で働くための技術であるナビゲーション (navigation) やマニピュレーション (manipulation) の研究に加えて、日



図1 レーザレンジファインダとカメラを用いた人間追跡のためのセンサネットワーク（ATR 知能ロボティクス研究所開発）

常生活で人とかわりながら働くロボットのための技術であるインタラクション（interaction）の研究が盛んになりつつある。

日常生活で人とかわりながら働くロボットにとって、困難なのは、人間と同じ視覚能力が期待される点である。人間を相手にしない技術開発では、環境をロボットやコンピュータのために作り込むことができた。しかしながら、日常生活において、人間とかわるためには、オープンな環境で人間並みの認識能力を発揮できる視覚認識の技術が必要となる。

そのような背景から研究が始まったのが分散視覚の研究である。分散視覚では環境に必要なだけカメラを配置する。ロボットに搭載された少数のカメラで、環境の状態を認識することは困難である。しかしながら、環境に十分な数のカメラを配置すれば、複雑な処理をしなくとも必要情報を瞬時に得ることができる[13]。

もっとも、いくらコンピュータビジョンの技術が進んだといっても、カメラだけで環境の様子を正確に把握するのはまだ限界がある。故に、複数のセンサを組み合わせたシステムが開発されている。筆者はこれを知覚情報基盤（perceptual information infrastructure）と呼ぶ。この知覚情報基盤で最近注目されているセンサシステムの一つが、ATR 知能ロボティクス研究所が開発したセンサネットワークである[14]（図1参照）。このセンサネットワークでは、複数のレーザレンジファインダで人間の位置を検出した後に、複数のカメラで人間の顔の位置を検出している。レーザレンジファインダと組み合わせることにより、カメラでは致命的な問題である照明条件の問題を克服している。

このほか、人間追跡のシステムは、RFID タグ等、他のセンサと組み合わせたものが様々開発されている。無論、カメラだけで追跡するシステムも数多く研究されているが、利用できる場所や環境は限られている。いずれにしろ重要なのは、人間追跡の機能が実環境で十分に利用できるようになってきたことである。これによって、ロボットの活動の場は、特定の環境から日常生活の場へと広がりがつつある。

### 3.2 顔認識

近年注目すべきコンピュータビジョン技術の一つは、顔

認識技術である。顔認識は、すでにデジタルカメラの基本的な機能の一つにまでなってきた。

1990 年ごろから、顔認識の研究が盛んになり、平均顔画像などのフィルタを用いる方法やニューラルネットを用いる用法が提案された。この顔認識の技術を確立させるための鍵は、コンピュータの高速化と小型化である。それまでにコンピュータビジョンで用いられていた特徴に比べて、顔の検出はより多くのモデルを必要とする。安定な顔認識を実現するには、できるだけ多様なモデルで総合的に判断することが必要なのであるが、コンピュータの高速化はそれを可能にしてきた。

この顔認識の技術は、現在開発されている日常活動型ロボットのほとんどすべてに実装されており、人間とかわるロボットにおいて必要不可欠の技術となっている。

### 3.3 SLAM

一方、移動ロボットの視覚もずいぶん進展を見せてきた。パノラマビジョンの研究で一時期盛んになった視覚移動ロボットの研究は、その後、SLAM（Simultaneous Localization and Mapping）の研究として発展し、今なお盛んである。SLAM が盛んになったのは、確率的な推定方法によって、大規模な地図を作成できることが示されてからである[15]。ただし、ここでも、カメラよりもレーザレンジファインダが多く利用されている。

## 4. 今後のロボットビジョン

### 4.1 実用に近づきつつあるコンピュータビジョン

コンピュータビジョン研究そのものは、工場で用いられているマシンビジョンの研究以来、久しぶりに実用化が期待される研究に進展してきた。特に顔認識やカメラネットワークの研究はすでに実用化されており、研究においても、かつてのように研究者が個別に独自のアイデアを練っていた時代から、特定の技術に焦点を絞り、ベンチマークをもとに、その性能を競う時代になった。音声認識研究と同じ様相を呈している。

音声認識の研究が進み、ほとんどのパーソナルコンピュータが標準的にその機能を搭載しているが、顔認識についても、パーソナルコンピュータの標準機能になる日は近いと期待される。

### 4.2 携帯電話と自動車と家庭電化製品

コンピュータビジョン技術の実用がさらに期待されるのは、携帯電話と自動車の分野であろう。携帯電話はもはや単なる電話機ではなく、カメラや Web ブラウザも備えるようになってきた。携帯電話は人間の体に張り付いて、いわばロボットのように環境を移動する。このカメラがさらに進化すれば、日常環境で活動するロボットの視覚機能と共通する点は数多く出てくると想像される。

期待されるもう一つの応用が、自動車である。自動車その

ものも知能化が進み、ロボットと区別がなくなりつつある。自動的に白線を検出して運転をアシストする機能はすでに実用化されている。今の ITS (intelligent transportation system) 研究を見ると、ドライバーの認識技術が盛んに研究されており、ドライバーの状態をカメラで認識しながら、運転をアシストする機能も近い将来に実現されるであろう。

また、センサネットワークの技術もさらに発展すると期待される。街中の監視カメラは不振な人物や困っている人を自動的に検出し、また、家庭内では人間の活動に応じて、テレビやエアコンなどの電化製品を制御するようになる日も近い。

#### 4.3 ロボットビジョンの展望

かつてのロボットビジョンの研究は、実用化が見えていなかったが故に、研究者がそれぞれにロボットを作り、そのための視覚機能の実現を目指していた。しかしながら、ロボットにおいてもロボットビジョンにおいても、実用化が見てくると、研究の焦点も変わってくる。

ロボットに関しては、無論、人間型ロボットがそこかしこで働くような時代がすぐに来るわけではないが、確実に世の中はロボット化が進んでおり、最新型の家電をはじめとして、自動的に動作するものは数多くある。故に、今後ロボットビジョンの研究もこういった実世界におけるロボットのためのビジョン研究がより盛んになると考えられる。

一方で、実用化の先には、より高度な視覚機能の実現が望まれるようになる。その一つの方向性は、認知科学や脳科学との融合であると想像する。人間の行動や表情などが、コンピュータによって理解できるようになれば、さらに、人間に適応したサービスを実現するために、より深い人間に関する知識をもとに、システムを開発するようになる。しばらくは実用化の研究が盛んに行われるであろうが、その先には、人間の認知モデルに基づいたより高次の認識を実現するロボットビジョンの研究が展開されるであろう。そしてそのときには、ロボットビジョン研究は人間の視覚認識機能を実現するという、コンピュータビジョンが目指してきた目的を再び共有することになる。おそらくは、新たな研究枠組みが提案されることになる。

#### 参 考 文 献

- [1] B.K. Horn: Robot vision. Mcgraw-Hill College, 1986.
- [2] D. Ballard and C. Brown: Computer vision. Prentice-Hall, 1982.
- [3] J. Aloimonos, et al.: "Active vision," Proc. Image Understanding Workshop, pp.552-573, 1987.
- [4] D.H. Ballard: "Reference frames for animate vision," Proc. Int. Joint Conf. Artificial Intelligence, pp.1635-1641, 1989.
- [5] C. Tomasi and T. Kanade: "Shape and motion without depth," Proc. Int. Conf. Computer Vision, pp.91-95, 1990.
- [6] 國吉康夫, 井上博允, 稲葉雅幸: "人間が実演してみせる作業の実時間視覚認識とそのロボット教示への応用", 日本ロボット学会誌, vol.9, no.3, pp.295-303, 1991.
- [7] J.Y. Zheng and S. Tsuji: "Panoramic representation of scenes for route understanding," Proc. Int. Conf. Pattern Recognition, pp.161-167, 1990.
- [8] H. Ishiguro, M. Yamamoto and S. Tsuji: "Omnidirectional stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.14, no.2, pp.257-262, 1992.
- [9] R.C. Nelson and J. Aloimonos: "Finding motion parameters from spherical flow fields," Proc. Workshop on Computer Vision, pp.145-150, 1987.
- [10] E. Menegatti, T. Maeda and H. Ishiguro: "Image-based memory for robot navigation using properties of the omnidirectional images," Robotics and Autonomous Systems, vol.47, issue 4, pp.251-267, 2004.
- [11] M. Tistarelli and G. Sandini: "Robot navigation using an anthropomorphic visual sensor," Proc. Int. Conf. Robotics and Automation, pp.374-383, 1990.
- [12] R. Benosman and S.B. Kang: Panoramic vision. Springer-Verlag New York, Inc., 2001.
- [13] 石黒浩: "分散視覚システム", 日本ロボット学会誌, vol.16, no.6, pp.778-781, 1998.
- [14] T. Kanda, D.F. Glas, M. Shiomi, H. Ishiguro and N. Hagita: "Who will be the customer? a social robot that anticipates people's behavior from their trajectories," Proc. UbiComp, pp.380-389, 2008.
- [15] S. Thrun, D. Fox, W. Burgard and F. Dellaert: Artificial Intelligence, vol.128, pp.99-141, 2001.



石黒 浩 (Hiroshi Ishiguro)

1991 年大阪大学大学院基礎工学研究科博士課程修了。工学博士。その後、山梨大学、大阪大学、京都大学、カリフォルニア大学、京都大学、和歌山大学を経て、2002 年より現職および ATR 知能ロボティクス研究所客員室長。(日本ロボット学会正会員)