

Computer Vision 研究の将来展望 - Robotics における視覚研究としての Computer Vision -

石黒 浩 驚見 和彦 天野 晃 浅田 稔
久野 義徳 渡辺 駿 八木 康史

Young Image Understanding Experts Organization

本稿では、Computer Vision 研究の将来展望に関する討論を報告する。著者間で一致した見解は次のようである。認識とは、無限の情報を含む外界から選択された情報の組合せに対して記号(タスクに対して妥当な行動を導く記号)を割り当てる過程であり、認識結果はタスク遂行結果に対する妥当性によって評価されるべきである。この考えに基づき、我々は、実環境でロバストに働くモジュール(環境、機構、タスクに依存した)を獲得し、組み合わせることにより、実世界を実時間で認識する視覚システムの実現を目指す。

Promising Directions in Computer Vision - Computer Vision for Robotics -

Hiroshi ISHIGURO Kazuhiko SUMI Akira AMANO Minoru ASADA
Yoshinori KUNO Mutsumi KUNO Yasushi YAGI

Young Image Understanding Experts Organization

This paper reports discussions on promising directions in Computer Vision. Our idea is as follows: Recognition is a process to assign symbols guiding proper actions for given tasks to the data(information) selectively extracted in an open world, and the results should be evaluated through the task execution. Based on this idea, we aim to develop a vision system recognizing real worlds in real time by integrating robust behavioral modules.

1 はじめに

Computer Vision 研究の目的は、Computer に視覚認識の機能を持たせることである。すなわち、視覚センサを通して得られた画像を解析し、画像に映し出された物体が何であるか、また、どこにどのように位置するか調べ、コンピュータやロボットに与えられたタスクに利用可能な情報を提供する。

過去 20 年以上にわたる Computer Vision の研究において、上記の問題を解決すべく様々なアイデアが提案されてきた。しかし、どのアイデアも実世界で起こりうる様々な変動に対処できるものではなく、人間のように柔軟で頑強な認識機能を Computer で実現できたとは言いがたい。近年に至り、多くの研究者が、Computer Vision のゴールに対して明確なビジョンを見失いつつあるよう思われる。

本稿では、このような Computer Vision 研究のブレークスルーとなるべく研究方針について討論した結果を報告する。ただし、本稿はあくまでも議事録的性格の強いものであり、コンセプト論文としての体裁を整えたものではないことをお断りしておく。

Computer Vision 研究のブレークスルーを模索するにあたり、まず 2 章で、Computer Vision における認識とは何かをもう一度考え、Computer Vision 本来の目的から見直す。我々の疑問は、身体を持たない機械が認識できるかということであり、逆に、身体を持つロボットの視覚としての Computer Vision のみが、Computer Vision 本来の目的を達成できると考える。3 章では、身体を持たないが故に解けない Computer Vision の幾つかの問題について考え、Computer Vision 研究における Robot の必要性を主張する。そして、4 章において、我々のイメージするシステム(ロボット)について述べ、5 章で、現状の研究から、イメージするシステムに向けてのアプローチを議論する。

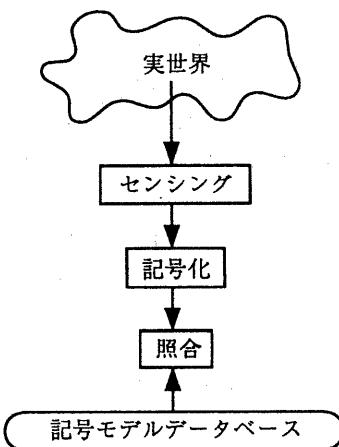


図 1: 従来の Computer Vision における認識

2 Computer Vision における認識について

2.1 従来の Computer Vision における認識

従来の Computer Vision 研究では、認識はおおむね以下のプロセスを指していると考えられる。

1. 外界を投影した画像から記号表現を作り出す。
2. 人間が正当と理解しているモデルと対比することにより、得られた記号表現の正当性を評価する。

Fig.1に従来の Computer Vision における認識過程を示す。この認識過程では、認識(図中では照合)結果が、センシングや、記号化にフィードバックされない。

これに対して我々は、以下のような直感的な疑問を持つ。

1. 情報の部分性についての考察が不十分である。すなわち、無限の情報を含む外界から適切な

情報を選択し、それが対象としているタスクにとって最適であることを検証するという、最も困難な問題を扱っていない。

2. 物理モデルとの対比が認識ではない。タスクに対する妥当性が認識結果となるべきである。

2.2 認識に対する我々の解釈

前節の情報の部分性とタスクに関わる2つの疑問に対し、我々は以下のように考える。

情報の部分性への対処

世界を完全に記述することは不可能であり、可能なのは、その中の一部の情報を得ることだけである。すなわち、情報収集の範囲を自分に関係する領域に限定し、獲得すべき情報の種類も自分を基準にして決定するしかない。

ここで、“自分”という言葉を使ったが、この自分という存在なしには情報を限定することができない。人間があらかじめ環境条件を整え、注意を向けるべきものを指示できるような場合なら、知的に見えるふるまいをするものは作れるだろう。しかし、そのような限定を解除したオープンワールドで行動できる知性は、注意を限定するための中心としての自分を持たなければならない。

この自分という存在と、それが存在する空間を対応づけるもっとも適当な手段は身体を持つことであると考える。すなわち、物理世界において情報を収集、処理するシステムには、身体を持つロボットが必要となる。

タスクに対する妥当性評価に基づく認識

従来の研究では、画像から作り出された記号表現と、物理モデルとの対応を認識していたが、完全な記号表現、完全な物理モデルを作り出すことは、情報の部分性の問題から見て不可能であり、そのため、これまでに作られた画像認識システムは、実世界で起こりうる様々な変動（照明変動やセンサ系のゆらぎ等）に弱いものであった。これに対

し我々は、認識を物理モデルとの対応ではなく、与えられたタスクに対し妥当な解答を示すための情報収集の過程そのものと考える。すなわち、身体を介した個々のタスクに対する妥当性の評価に基づく認識を基礎として、システムを構築すべきと考える。

私は、身体性を持つロボットにおける認識を、以下のようにとらえる。

- 認識とは、無限の情報を含む外界から、身体を介して選択された情報の組合せに対して記号（タスクに対して妥当な行動を導く記号）を割り当てる過程であり、認識結果は、身体を介して遂行されるタスクの遂行結果の妥当性によって評価されるべきである。

3 ロボットの必要性

Computer Vision の研究が、オープンワールドの問題を解決しようとしたとき、情報の部分性や、認識結果の妥当性評価において、身体を持つロボットが必要になる。従来の Computer Vision で扱ってきた幾つかの問題を通して、このロボットの必要性を考えてみよう。

3.1 Model Based Vision

Model Based Vision は、Computer Vision における代表的な研究である。従来の研究では、センサから得られる情報を人間が正当と信ずる記号化の結果得られるモデルの記述と、予め記憶されているモデルを対応させるというものであった。Model Based Vision の延長線上で、実世界の様々な物体の認識を試みると、以下の問題が生じる。

1. 全てのモデルを人間の手によりコンピュータに与えることは不可能である。
2. 人間の用意したモデルは必ずしもコンピュータにとって妥当なモデルではない。

全ての認識に利用できる完全なモデルを用意することは不可能である。これは、フレーム問題[4]や、オープンワールドの問題[5]に共通するものであり、人工知能研究全般で大きな問題になっている。故に、モデルを獲得する能力が必要であり、前章で述べたように、モデルの獲得には、身体を持つことが重要な意味を持つ。

反論として、例えば、ネットワーク上の他のコンピュータからモデルを獲得し、多くのモデルを共有するというアイデアがあるかもしれない。これには、あるコンピュータが所有しているモデルを、目的の違う他のコンピュータが利用できる一般的なモデル表現への変換が存在する、という前提が必要となる。しかし、全てのモデルをあらわす唯一無比な一般的表現などありえない。仮にあつたとしても、変換コストが高ければ、実際の局面では役にたたない恐れがある。また、それが何らかの形で損傷を受けた場合、システム全体が破綻をきたすおそれがあり、環境へのロバスト性を問うと、これは大きな問題である。

人間のセンサとコンピュータのセンサ(例えば、CCD カメラ)との間には大きな相違があり、人間が与えるモデルが、コンピュータにとって妥当なモデルとなっている保証はなにもなく、むしろ妥当なものとなっていないと考えられる場合が多い。これまでの多くの研究が、実世界で起こりうる様々な変動に弱くもろいものであることが、それを示しているのではないだろうか。

コンピュータにとって妥当なモデルとは、センサ情報と物理世界における適切な行動を結び付けるものであり、モデルを利用するものが、物理世界と自らの関係を把握しながら獲得すべきである。

Held and Hein[3] の論文にあるように、子猫の実験で、自ら運動を生成している場合は、知覚形成(奥行き感など)が可能であるが、他から駆動されている箱の中にいる猫は、例え、首振りが可能でも知覚が形成されない。これは、触覚など環境と自らの身体との関係を直接知覚する内界センサによる、環境内における行為の確認なしには、外

界センサからもたらされる情報が意味づけされないことを示している。この例は工学的な根拠にはならないが、身体を持つことの重要性を示唆するものである。

反論として、情報の抽象化が進めば、センサの構造の違いは克服できるという意見があるかもしれない。しかし、センサが異れば、抽象化の手法も変わるはずであり、やはり、人間のセンサとロボットのセンサとの隔たりは埋め尽くせない。現在のシステムの中には、予め記憶するモデルに、モデルの抽象度に応じた構造化を行っているものもあるが、その構造化は十分な根拠に基づくものではない。

3.2 視覚移動ロボットの環境モデル獲得

環境モデルを獲得する視覚移動ロボットも、Computer Vision の一分野として長年研究されてきたものである。この研究においても、以下の問題点がある。

1. 人間がロボットに与えるセンサ情報の記号化手法は、必ずしもロボットにとって最適なものではない。
2. ロボットは、自ら再利用可能なモデルを獲得していない。

先に述べたように、人間のセンサとロボットのセンサとの間には大きな相違があり、人間が与えるセンサ情報の記号化手法は、ロボットのセンサや機構に対して、十分可能なものとなっているかは定かではない。

再利用可能なモデルの獲得においても、これもまた先に述べたように、センサとアクチュエータの両者の利用が必要である。センサとアクチュエータを通して獲得されたモデルでなければ、逆にそのモデルから、適切な行動を獲得することは難しい。

従来の視覚移動ロボット研究では、身体を持つロボットを使っていても、身体を用いた能動的な観測に関しては考察されていなかった。また、Active Vision[1] は、このような問題を取り扱おうとした

研究枠組みであるが、未だ上記の問題に解答を与える研究成果は報告されていない。

3.3 顔の表情認識

近年、Computer Vision の新たな認識対象として、人物の顔画像の認識が盛んである。この研究でも、先に述べた Model Based Vision と同様の問題を、そのまま抱えながら研究が進められている。表情認識における問題として以下のものがある。

1. 状況のフィードバックなしに表情は理解できない。

これまでの研究では、顔の構成要素の絶対的な動きのモデルをもとに、表情認識が試みてきた。そこでは、その表情が発現した状況がどのようなものであろうと、それとは独立に表情認識が行われる。しかし、人間の表情には個人差が大きく、このような従来のモデルではロバストな認識は不可能である。

本来表情は、2者間で発生した状況とそのときの表情の関係から、意味付けされるべきであり、さらに、意図的に新たな状況を発生させることにより、相手の表情変化を引き起こし、その意味付けの正当性を確認する必要がある。すなわち、この表情認識においても、先の問題と同様に、状況を能動的に変化させる行為を介した認識でなければならぬ。

もっとも、表情認識においては、必ずしも、物理的な身体を持つロボットが必要ではない。2者間の状況を変化させる何らかの手段があれば、コンピュータによる表情認識は可能であろう。しかし、人間との間に様々な状況を発生させるために、身体が有効に利用できることは確かである。

3.4 行動認識

行動認識は、複数のエージェントが存在する環境において、通信機器を使わずに情報を伝達する。他者の行動を取り込むことにより、環境から行動を学習する、といったエージェントの機能を実現

するための重要な課題である。このような、人間やロボットの行動認識に関しては、その重要性は意識されているものの、研究例は少ない。この研究では以下のことに注意すべきである。

1. 行動認識のためのモデルはロボット自らが獲得しなければならない。

行動認識は、表情認識以上に身体を持つことの意味が大きい。他者の行動は、まず他者の身体を自分の身体にマッピングし、そのマッピングを元に自らが行動し、その行動の効果を知ることで、理解することできる。すなわち、行動認識においては、相手の行動がマッピングできる身体が必要不可欠である。他から行動認識のためのモデルが与えられては、相手の行動が物理世界にいかなる効果をもたらすかは知る由もなく、2章で述べた我々のめざす認識は実現できない。

3.5 物理的身体の必要性

本章の最後に、我々はあくまで、物理的身体(ロボット)を用いた実世界における実験を重要視していることを述べておきたい。

まず、シミュレーションだけでロボット研究ができるとは考えられない。

我々が達成したいのは、物理世界で行動する知能ロボットである。仮想的な環境で作られたシステムは、仮想的な世界の複雑さには対処でき、知的に振る舞えても、それが物理世界で行動できるとは限らない。仮想的な世界に取り込むことできる複雑さは、人間がモデル化できるだけの複雑さであり、現実世界とは本質的に異なる。

もっとも、あるプロトタイプシステムの振舞いを特徴付けるために、人間がモデル化できる環境において、評価や実験を行うことは意味がある。

しかし気をつけなければならないのは、このようなシミュレーションが意味を持つのは、問題の本質を適切にとらえている場合に限られることがある。問題の本質を適切にとらえることは非常に難しく、十分な注意が必要であり、シミュレーション

ンは研究開発における一つのツールとして有用であるが、最終的には、物理的身体を持つロボットによって、全ての研究は検証される必要がある。

主張したいのは、シミュレーションは意味を持たないということではなく、実世界の複雑さを直接扱う枠組みとしては、仮想的な環境に閉じてはいけないということである。

4 目指すシステムに対するイメージ

先に述べた新たな方策により実現されるであろう、システムに対するイメージについて述べる。目指すシステムに対するイメージはおおむね以下のようなものである。このイメージは Behavior Based System[2]への共感から得られたものである。

- 幾つかの基本的な行動モジュール群から構成される。
- 行動モジュール群の相互作用により、情報の部分性がうまく補われ、タスクが達成される。

Fig.2は、我々のシステムのイメージを示す。従来の Computer Vision 研究では、認識機能を実現するにあたり、システムアーキテクチャは重視されていなかったが、我々は、そのシステムアーキテクチャが認識機能実現に深く関わると考える。図に示すシステムは、基本的に、センシング/行動モジュール群から構成される。モジュール群は、実世界と、記号化/計画モジュールによって作られる仮想世界の両世界に働きかける。ただし、この図は、現時点におけるおよそのイメージを示したものにすぎないことを断っておく。

このシステムのイメージにおいて、どのような行動モジュールを用意するか、いかに行動モジュールを獲得するかにおいて幾つかの直感を持っている。

どのような行動モジュールを用意するか

我々の目指すシステムでは以下のようなモジュールの存在を想定している。特に、4つ目のモジュー

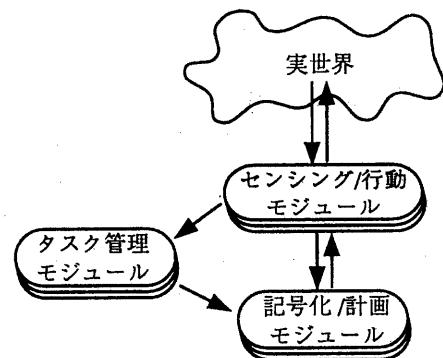


図 2: 我々のイメージする認識システム

ルは、これまでの Behavior Based System との違いを特徴づける意味で重要である。

1. 従来の Behavior Based System に見られるような、反射的な行動を実現するモジュール
2. 記号推論を行うモジュール、行動の計画を立てる。
3. 感情または、それよりも大きな枠組みでイメージ（ここで、目指すシステムに対するイメージを作ろうとしているが、まさにそのイメージ）を司るモジュール。
4. 意志/意思を持つモジュール、この存在により、不足した情報が補われる。

いかに行動モジュールを獲得するか

行動モジュールの獲得方法は、研究者により様々であるが、主に以下の 2つが可能性のあるものとして挙げられる。

- 認知科学の知見を利用して、従来のビジョンの構成方法に欠けていた考え方を明らかにし、より良い工学的手法を見つける。例えば、人間が発達過程で行う模倣の機能をロボットに取り入れることにより、人間と同様の巧妙な

注意制御とモジュールの獲得（学習）を実現する。

- 現状の Robotics の延長として、行動モジュールの獲得を考える。例えば、ロボット記述言語を強化し、行動モジュールの記述を容易にする。

5 現状の Computer Vision からイメージしたシステム実現に向けて

イメージしたシステム実現に向けて、現状の Computer Vision 研究がどのような方向に進むべきか、さらに具体的に考えてみる。

これまで述べたように、我々は、以下のような方針で、システムの実現を目指す。

- 実環境でロバストに働くモジュール（環境、機構、タスクに依存した）を獲得し、組み合わせることにより、実世界を実時間で認識する視覚システムを実現する。

また、このシステムを実現する課程で、タスク、環境、機構の関係を明らかにし、さらには、認識とは何かについて工学的見解を得ることができると考える。何度も繰り返すが、タスク、環境、機構をふまえることにより、我々が考えるところの、真の認識に近づける。

以下では、システム実現における 4 つの段階にわけ、議論を進める。

5.1 実世界に則した環境/機構に依存するモジュールの実現

ロバストな情報収集のための、行動を伴う視覚探索におけるアルゴリズムを開発し、多くの事例を集めることにより、その一般的性質を検討する。信号が有する情報をどこまで引き出せるかという点から、人間系に無関係に、定量的な性能評価が行なえるはずである。このモジュールの一般的な

性質を基に以下の 2 つのアプローチで研究に取り組む。

1. 学習により視覚探索アルゴリズムを獲得する。
学習アルゴリズムの収束性から、モジュールの最小単位を定義する。また、人間や他のロボット行為を模倣することで、学習を早める。
2. アダプティブなモジュール記述言語を開発し、それを基にモジュールを記述する。このモジュール記述言語は、収束性が保証される範囲で学習アルゴリズムを有効に利用するためのものである。

また、ここではセンサ開発も重要な課題の一つである。従来よりも、ロバストな情報が得られるセンサが必要である。例として、全方位視覚センサや多重解像度視覚センサ等が考えうる。

5.2 環境や機構には依存しない、タスク実行モジュールの実現

このレベルでは、モジュール間の関係をより明確にし、次のモジュールの組み合わせにつなげる。

一般的なタスクモジュールを獲得するために、タスクの最小単位（いわゆる Behavior）の工学的な定義を行い、それをもとに全てのタスクの構造を説明する必要がある。ここでも、2 つのアプローチが考えられる。

1. タスク単位での学習を行う。ここで学習アルゴリズムには、般化能力の優れたものが要求される。
2. 前節のモジュール記述言語を発展させたタスク記述言語を開発し、それによりタスク実行モジュールを実現する。すなわち、学習アルゴリズムではなく、プログラマの注意深い洞察により、タスクの普遍性を探り、タスク記述言語を実現する。

5.3 注意制御の実現

システム全体のタスク（上記のタスクは、システムタスクのサブタスク）を遂行するために、サブタスクの起動を最適に制御する方法、すなわち、注意制御の実現方法を検討する。注意制御の実現方法としては、以下の2つが考えられる。

1. 学習のみによりサブタスク起動の優先順位を獲得する。また、模倣により学習を加速する。
2. 数多くの注意制御の例から、プログラマの注意深い洞察により、注意制御方法を獲得する。

また、ここでは注意制御を実装するコンピュータシステムについても考えておく必要がある。複数のモジュールを並列実行可能にする、並列コンピュータが必要であろう。

6 おわりに

最後に、研究の評価そのものがどのように変わるべきかについて述べておきたい。これまでの研究は、解くべき明らかな問題が存在していたため、その評価基準は比較的明確であった。しかし、ロボットの可能性を摸索する上記のような研究では、明らかな評価基準を持たないままに、研究成果を世に出さざるを得なくなる。

すなわち、タスク指向型のシステム開発に関する研究を評価する、新たな尺度が必要である。これまでの研究評価に加えて、その難しさに応じたタスクの分類を行い、そのシステムの達成度を評価できる尺度を用意する必要がある。

YIUEOについて

Young Image Understanding Experts Organization (略称:YIUEO) は、関西地区の若手 Computer Vision 研究者間での情報交換を目的として発足した研究会組織である。今回の報告は、この YIUEO メンバーの中で、今後の Computer Vision は、Robotics における視覚研究として位置づけら

れるべきと考える7人が集まり討論した結果である。各自の所属(E-mailアドレス)は以下の通り。

石黒浩、京都大学 工学部 情報工学教室
ishiguro@kuis.kyoto-u.ac.jp
鷲見和彦、三菱電機 株式会社
sumi@fas.sdl.melco.co.jp
天野晃、京都大学 工学部 情報工学教室
amano@kuis.kyoto-u.ac.jp
浅田稔、大阪大学 工学部 電子制御機械工学科
asada@robotics.ccm.eng.osaka-u.ac.jp
久野義徳、大阪大学 工学部 電子制御機械工学科
kuno@cv.ccm.eng.osaka-u.ac.jp
渡辺睦、東芝関西研究所
mutty@krl.toshiba.co.jp
八木康史、大阪大学 基礎工学部 システム工学科
y-yagi@sys.es.osaka-u.ac.jp

参考文献

- [1] D. H. Ballard, Reference Frames for Animate Vision, Proc. Int. Joint Conf. Artificial Intelligence, pp. 1635-1641, 1989.
- [2] R. A. Brooks, Intelligence without Representation, Artificial Intelligence, Vol. 47, pp. 139-159, 1991.
- [3] R. Held and A. Hein, Movement-Produced Stimulation in the Development of Visually Guided Behaviors, J. Comparative and Physiological Psychology, Vol. 56, No. 5, pp. 872-876, 1963.
- [4] 松原仁、橋田浩一、情報の部分性とフレーム問題の解決不能性、人工知能学会誌, Vol. 4, No. 6, pp. 695-703, 1989.
- [5] 井上博充、感覚と行動の統合による機械知能の発現機構の研究、第12回日本ロボット学会学術講演予稿集, pp. 151-152, 1994.