

# 「ロボットの世紀」

秋田純一(8,12,14,17,18,19,22 回)

## ロボットとわたし

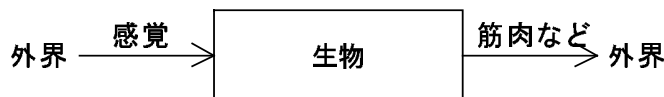
まずはじめに白状しておきます。ロボットの世紀、ということなのですが、実は秋田の元々の専門はロボット工学ではありません。なのにどうしてこんなことを書いているかという、次のような二つの理由があるからです。

まず一つの理由は、以前湧源クラブ基金をいただいて有志でRoboCupというロボット工学と人工知能のチャレンジに参加したことがあることです。これはいまでも引き続き継続していますが、ロボット工学の専門性から、という形よりは、むしろ趣味や興味本位が大きな部分を占めています。

もう一つの理由は、秋田の元々の専門にあります。秋田は、学生のころから電子工学科で「集積回路」の研究をしてきました。学生のころの研究は、ひとことでいえば「コンピュータをいかに速くするか」という研究でした。それはそれで面白かったのですが、卒業するあたりのころから、コンピュータ自身(ハードウェア)の発展の今後に疑問と不安を持つようになりました。それと時期を同じくして前述のRoboCupに触れ、それに参加していくうちに、「ロボットの目」というものは、まだまだやるべきことがたくさんあるようだ、と感じるようになりました。そんなわけで、集積回路の研究を続けながら、その応用としてロボットを念頭においていることが多くなりました。そんなわけで、ロボット工学、というよりは、ロボットの目の集積回路の話になりますが、おつきあいください。

## 生物・ロボット・感覚・センサ

私たち人間をはじめとして、生物が生きていく上で、いろいろな感覚によって外界の情報を得ています。例えば人間であれば、五感といって主に視覚、聴覚、嗅覚、触覚、味覚をもっていますし、他の動物でもそのような感覚をもっていて、外界から情報を得て活動しています。つまり生物を単純化したシステムとして考えれば、次の図のようなシステムとして考えることができます。



つまり外界から各種の感覚によって情報を得て、それに基づいて何らかの判断を行って、その結果に基づいて行動をしている、というわけです。

特に動物においては、いろいろな感覚の中でも特に視覚、つまり「目」から得る情報に、非常に大きく依存していることが知られています。それは例えば私たちが目を閉じて外を歩くことを考えてみれば(あるいは実践してみれば)よくわかります。

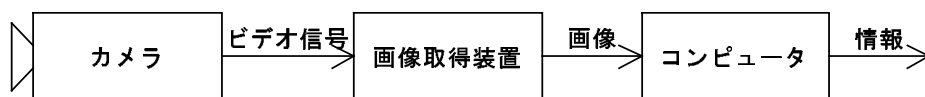
さて本題のロボットですが、ロボットも究極に単純化すれば、先ほどの図のようなシステムとして考えることができます。ロボットが外界の情報を得るために用いる装置をセンサ(sensor)と呼びますが、基本的に外界の情報を得て活動していることには変わりはありません。

動物の活動では、視覚情報が非常に重要であることは先ほど述べましたが、そうするとロボットの活動でも、視覚情報が非常に重要であると考えられます。これは言い方を変えれば、動物でもロボットでも、活動する上では、目で見た情報が最も情報量が多く、有用である、というように考えられます。

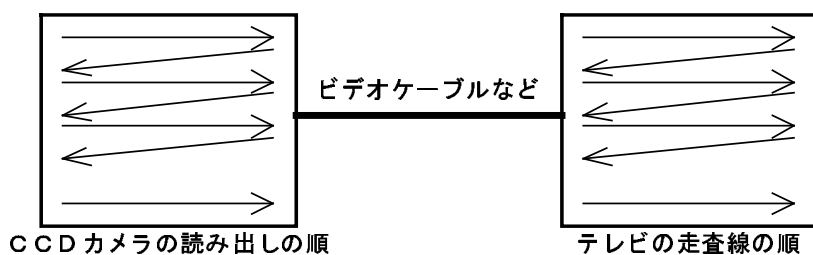
## ロボットの視覚

このような考えから、ロボットの行動においても視覚情報を活用しよう、という考えや試みは非常に古くからありました。もちろん現在でも、ロボットの感覚として最も有用な(でも実は最も扱いにくい)ものは視覚情報です。通常ロボットの目として使われるのはカメラです。細かくいえば**CCD**カメラと呼ばれるものが圧倒的によく用いられます。これを使って、画像取得装置(フレームグラバと呼ばれるものなど)を使って目で見た情報を写真のような「画像」として取り込み、それをパソコンなど

のコンピュータに転送して、さまざまな「画像処理」と呼ばれる処理を行って、必要な情報を得るのが一般的です。つまり次の図のような感じです。



ここで実はミソ(というか問題)なのは、カメラとしてビデオカメラを用いることです。ビデオカメラというのは、もともとはテレビにつないで映像を映すためのものです。つまり「人間が見るための画像」を作り出すためのものであって、「ロボットが見るための画像」を作り出すためのものではない、ということです。テレビで映像をつくりだす仕組みはラスタースキャンと呼ばれるもので、映像を左上から右下に順に走査線を走らせて、光の点の集合として映像をつくるわけですが、**CCD**カメラの構造は、まさにこの順に映像中の点を読み出して取り出しています。読み出された映像中の点の情報(色や明るさなど)は、一つ一つの点の情報が時間に沿って順番に送られる、時系列信号と呼ばれる形態で転送され、テレビに映されます。



つまり本来は人間が見るための画像を作るためのビデオカメラを、ロボットの目にも流用している、ということなわけです。これはビデオカメラの研究が進んでいて精度・色再現性などがよいなどの技術的な理由や、普及しているため安価で入手がしやすい、という経済的な理由が主ですが、いずれにしても、本来は違う目的のためのものをロボットの視覚に流用している、というのが現状なわけです。

ロボットの視覚系とも呼ぶべきものを、先ほどのような構成で使うもう一つのミソ(こちらは主にメリット)なのは、画像の「理解」にコンピュータを用いることです。つまりプログラムを書いて、いろいろな画像処理のプログラムを使って画像の特徴量(物体の形状や位置、色など)を求めたりするわけです。この方法では、プログラムの作り方によっていろいろな情報を得ることができ、またその情報を求める方法の変更も容易である、という利点があります。

このように入手が容易で柔軟性に富むカメラを用いたロボットの視覚系ですが、根本的に一つの大きな問題点があります。それは、この方法ではコンピュータで一つ一つの光の点の情報を処理するため、時間がかかるということです。一般に画像信号に含まれる点の数は膨大なことが多いため、技術の進歩でいくらコンピュータが速くなっても、基本的には一つ一つの点の情報を順に処理する(逐次処理と呼ばれます)方法では時間がかかってしまいます。

NTSC信号と呼ばれる普通のビデオカメラの映像信号程度の解像度を、NTSC信号の規格である毎秒30枚ずつ処理するぐらいであれば、最近の速くなったコンピュータであれば十分可能です。しかし毎秒30枚というのは1枚あたり33ミリ秒ですが、これは33ミリ秒ごとにコマ送りされる映像をみているようなものです。人間が目で見るとは、残像などの効果もあってそれほどコマ送りになっている、という気はしませんが、ロボットにとっては、それで外界の情報をすべて得るわけですから深刻です。例えば毎秒30m(時速100kmぐらい)で動くボールは、33ミリ秒の間に1mも動いてしまいます。カメラの視野を横幅30m(野球のホームベースと2塁ベースぐらいの距離)としてもその1/30ですから、画面の1/30ぐらいごとにとびとびに動いているように見えます。つまり横幅60cmのテレビであれば、2cmずつとびとびに映っているボールを見て、「動いているボール」として理解しなければいけないこととなります。これは実際にやってみると、けっこう大変な作業です。つまりCCDカメラをつかったロボットにとって、こんな速い(といってもバッティングセンターの一番

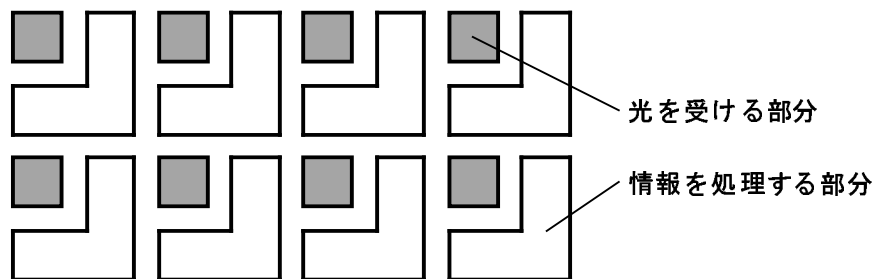
遅いボールぐらいの速さ)ボールを、「動いているボール」として理解するのは、非常に困難です。

かといって、毎秒30枚というビデオカメラの規格よりも多くの例えば毎秒1000枚程度の画像をとればボールはほとんど止まって見えるので、「動いているボール」として理解するのは容易になるでしょうが、毎秒1000枚もの画像をCCDカメラから読み出すには、仮に100万画素のビデオカメラだとしても1画素あたり10億分の1秒、つまり1GHzで読み出しを行わなければなりません。これはとても現実的な数字ではありませんし、しかもこんなにたくさんの画像情報をコンピュータで順番に処理するのは、いかに最近のコンピュータが速くなったといっても難しい話です。また、画像のより細かいところまで見ようとして、画像の解像度を高くすると、やはり処理すべき情報量が膨大になり、時間がかかってしまうことになります。(ちなみに毎秒1000枚ぐらいであれば、ロボットの視覚系としては十分です)

このように、カメラ+コンピュータという構成のロボットの視覚系では、根本的に処理速度の問題をかかえています。これは、カメラで撮って、それを転送して、コンピュータで順に処理をする、という構成をとっている限りは、なかなか解決できないやっかいな問題です。コンピュータを何台も使うことも可能ですが、根本的な解決にはなりません。いずれにしても、1秒あたりの画像の枚数と、画像の解像度は、用途と技術的制約と予算的制約から妥協して決めるのが一般的です。

## ビジョンチップ、またはシリコン網膜

このように根本的に速度と解像度の両立が困難なカメラ+コンピュータという構成のロボットの視覚系ですが、これを解決するものとして研究されているものの一つに「ビジョンチップ (Vision Chip)」あるいは「シリコン網膜 (Silicon Retina)」と呼ばれるものがあります。これは10年程前に生まれた比較的新しい概念で、以後広く研究が行われています。



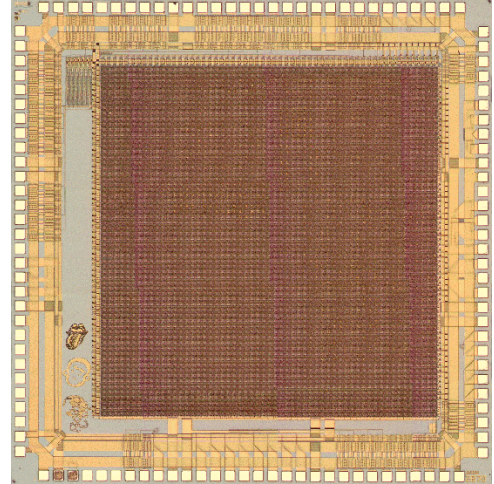
これは、もともとカメラで光を電気信号に変えて、それをビデオケーブルで転送して、コンピュータに順番に処理をしているんだったら、光を受ける部分と情報を処理する部分をいっしょにしてしまえ、という発想です。つまり「光の点」ごとにコンピュータを並べてしまうようなものですが、これが実現可能となった背景には、集積回路の技術が進歩があります。つまり昔はコンピュータも大きかったのです。そんなものは作ることができずはなかったのですが、集積回路の技術の進歩によってコンピュータを小さく作ることができるようになり、ついには「光の点」(だいたい10 $\mu$ m四方程度の領域)に、単純な機能のものではありますがコンピュータを載せられるようになったわけです。このようなものが有用であるのは、画像信号が画素ごとに独立性が非常に高く、独立に処理することができる、ということも理由です。また動物の受光器官である網膜にも、このように光受容体(coneやrodと呼ばれる視細胞)と簡単な信号処理システム(水平細胞など)が統合されて存在する、という事実もあります。

ただしこのビジョンチップにも難点があって、どうしてもコンピュータを載せている分だけ解像度が低くなってしまって細かいところまでわからない、ということがあります。そこでほとんどの研究は、あまり複雑でない情報処理を行うことを目指しています。例えば画像のノイズを除去したり、エッジ部分だけを強調したりといった、「初期視覚問題(Early Vision Problem)」と呼ばれる処理を行うものがほとんどです。つまり、やはり人間が見るための映像を作ることを目指しているわけですが、ロボットの視覚系としては、やはり視覚情報の「意味」がほしいところです。

## ぱっと見センサ

最初に書いたような背景で、最近秋田が研究しているのは、画像の「意味」を理解できるような画像センサ(あるいはカメラ)を作ることができるようなビジョンチップです。さすがに複雑な画像の「意味」を理解するのは難しいので、まずは、画像の簡単な意味を理解できるようなもの、私たちがものを見てなんとなく得る、いわば「ぱっと見」とでも呼ぶべき情報だけでも、非常に高速に得ることができるもの、いわば「ぱっと見センサ」を作ろうとしています。

まず「ぱっと見」として最初に取り組んだのは、視野内にある物体の位置を知るためのものです。私たちがものを見る時、まず、「あれ?このあたりになんかあるな」と思い、続いてその付近を「なんだろう?」とよく見て考えて認識する、という手順をふんでいると考えられます。そこでその最初の段階である、「このあたりになんかあるな」という情報だけを高速に知るための画像センサです。細かい回路構成についてはここでは述べませんが、右の図は、最近試作した、視野内にある物体の位置を高速に知ることができる画像センサです。1辺の大きさは4.5mmで、約3000個の画素があります。普通の画像センサ(カメラ)と違うのは、普通のカメラでは出てくる信号は「絵」ですが、この「ぱっと見センサ」では、出てくる信号は「座標」になります。つまり視野内の物体の座標が、すぐに得られるわけです。(ちなみにRoboCup的には視野内のオレンジ色の物体はボールなので、ボールの座標がすぐに得られるわけです)



他にとして考えているのは、例えば視野内の特定方向の線分の情報を得るセンサです。これは視野内の線分(サッカーフィールドのサイドラインなど)を知るのには好都合です。

他に漠然と考えているのは、不規則な画素配置の画像センサです。CCDカメラの画素は正方形の格子状に並んでいますが、これは先に述べたようにテレビに映す信号をつくるためです。しかし視野内の物体の表現、という意味では、この格子状の画素の配置は必ずしも好都合ではなく、例えば斜めの直線はガタガタの点の集合と見えてしまいます。これに対して網膜上の視細胞はこのように整然とはならんでおらず、ほとんど不規則に並んでいます。これは信号理論的にはホワイトノイズと考えられますが、これはすべての周波数(空間周波数)を含む配置と考えられるため、逆にいえばどのような形状でも自然に表現できる可能性をもっているはずです。問題は不規則に並んだ画素をどのように扱うか、ということですが、動物の網膜の場合は、一つ一つの視細胞から視神経が脳につながっているわけで、その配置が不規則でも構わないわけですが、コンピュータで扱うのは困難です。この点を解決できるうまい方法があれば、ひょっとしたら今までの概念とはまったく異なるロボットの視覚系ができる、かも、しれません。

## 最後の目標

このようにロボットの視覚系というものをいろいろ考えてくると、前から気にはなっていて、興味はあったんだけど、なかなか近づかずに今まで生きてきたんだけど、最近はどうしても避けて通れないんじゃないかという気がしてならないものがあります。

それは「脳」です。

さきほどの「ぱっと見」もそうですし、特定方向線分の検知というのは、実際にそういうものに反応する神経細胞があることも知られています。あとはこれをうまく統合してやれば、本当の視覚情報の「意味」を知ることができる、ロボットの視覚系ができるはずなのですが、はたしてどうなることやら。今後の精進です。

(あきたじゅんいち: 公立ほこだて未来大学 システム情報科学部 情報アーキテクチャ学科)