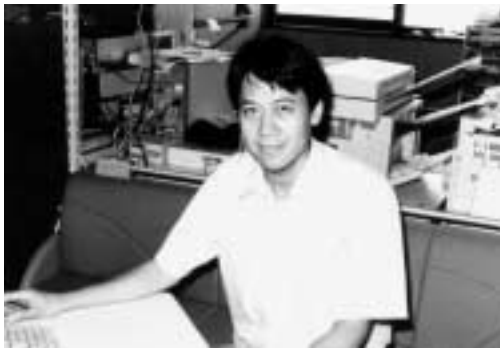




# 動く物体を追跡するために

## 熊澤 逸夫 研究室 ~ 計算工学専攻



熊澤 逸夫 助教授

真新しい西8号館に位置する熊澤研究室。ここでは主に、物体追跡技術の研究が行われている。これは画像中の任意の物体をコンピュータに探索させ、その位置や形状についての情報を記憶させることで動く物体を追跡していく技術である。様々な分野に応用できるとして世界中でも多くの研究機関でこの技術が研究されている。現在は追跡すべき対象に色情報を付加し、その色を追うことで対象を追跡する方法が主流であるが、熊澤研究室はニューラルネットワークという仕組みを使ったユニークな方法で物体追跡を研究している。



# ニューラルネットワークって？

「ニューラルネットワーク」という言葉は多くの読者には耳慣れないものであろう。これは、脳内の神経細胞（ニューロン）がつくるネットワークをモデル化したものである。

脳内ではニューロンとよばれるもの（図1(a)）がいくつか互いに結合し、ネットワークを形成している。入力端子である樹状突起を通して複数のニューロンから電気パルスが送られてくると、細胞内部の電位が上昇する。そしてその電位の総和がある値を超えると、次のニューロンへ軸索という出力端子を通して電気パルスを送る。これを繰り返すことによって、各ニューロン間に情報が伝わっていく。

こうした脳内のニューロンをモデル化したものが図1(b)だ。左の3つの矢印は他のニューロンからの入力、右の矢印が出力である。いくつかのニューロンから信号が送られてきた時に、その入力の総和が条件を満たしていれば1という情報を、満たしていなければ0という情報を次のニューロンへ送る。このようにして、複数のニューロンを結合させ、情報を伝えていくことによりネットワークを形成していく。

そして、ニューラルネットワークの最大の特徴であるのが「学習」だ。学習とは、ニューロンが受け取った入力に対して必要とされる信号を出力するよう、自らを変化させていく過程のことである。この学習能力があるために、不完全な情報が

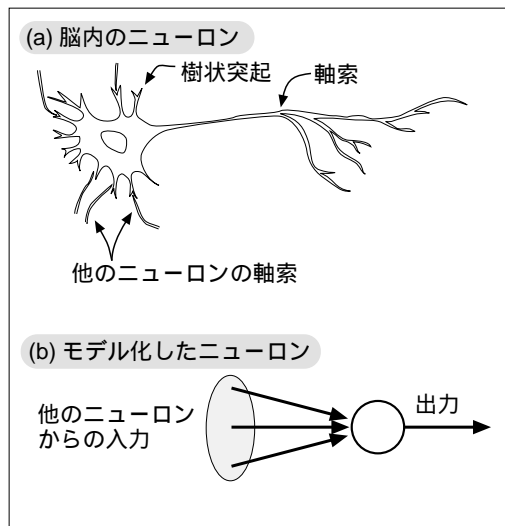


図1 脳内のニューロンとニューロンモデル

与えられたり、ネットワークの一部が壊れたりしていてもある程度ならば正しい処理を行うことができる。このため、従来のコンピュータには難し

い処理を比較的容易に実行できるとしてあらゆるところで注目されている。



## 追跡対象の輪郭を近似！

実際にコンピュータに物体を追跡させるには、画像中の対象物体の形を記憶しなければならない。現在、動画を表すときには1秒間に約30枚の静止画を使うのが一般的である。しかし、これを全て記憶させようとするとは莫大な量のメモリが必要となるので、記憶にかかる時間も長くなってしまふ。そうなると、物体の動く速さに比べて画像を記憶する速さが遅くなるので、リアルタイムで追跡するのは大変難しい。

この問題を解決するために、熊澤研究室では物体の輪郭を直線や2次曲線で近似して、記憶させている。なぜなら、コンピュータに追跡させるような物体はほぼ全てがいくつかの直線と2次曲線を組み合わせることで近似できるからだ。例えば球状のボールなら円で表すことができるし、手のような複雑な形も指の部分は直線、指先の丸い部分は楕円を組み合わせることで近似できる。

直線や2次曲線は、 $x, y$ を変数、 $a-f$ を定数として、簡潔な数式で表すことができる。

直線  $ax + by + c = 0$   
 2次曲線  $ax^2 + by^2 + cxy + dx + ey + f = 0$

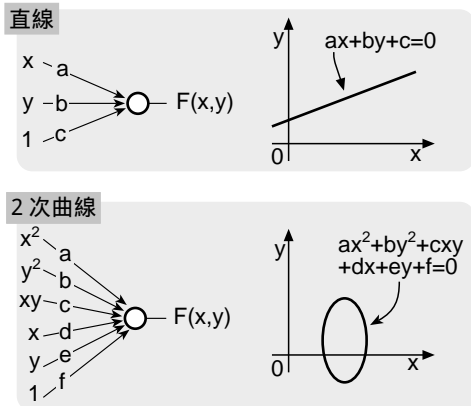


図2 直線・2次曲線の表し方

これをニューラルネットワークで表示したものが図2だ。 $x, y$ などの入力をする、次のニューロンに $a-f$ という係数の情報を信号に付加して送ることで、右のような直線・二次曲線が表せるようになっている。

また、以下のようにすることで、複数の直線や2次曲線を使って輪郭を近似できる。図3を見てほしい。左の3つのニューロンのうち、アは2次曲線の内側に $x, y$ の入力があるときに次のニューロンに1という信号を送り、そうでないときは0という信号を送る。同様に、イは直線1の下に、ウは直線2の上に $x, y$ の入力があるときに次のニューロンに1という信号を送り、そうでないときは0という信号を送る。こうすると、図のような2次曲線と直線で囲まれた形を近似できる。これを組み合わせることでより複雑な、手のような形も近似することもできる。

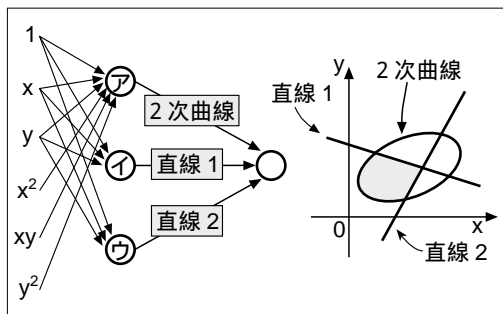


図3 囲まれた図形の近似方法



## 物体追跡の具体的方法

それでは実際にどうやって物体追跡をさせるのか、順を追って説明していこう。

まず、追跡すべき対象の含まれた画像を与え

る。するとニューラルネットワークはこれを初期値として記憶する。そして、先に述べたような方法で「自動的に」その形を近似する。つまり、係

数の値やどこをどのような関数で近似するのか、といったことを人間がいちいち指示する必要がないのだ。これはニューラルネットワークが、学習能力を持つためである。

次に2番目の画像を与える。この画像中の追跡対象は、もちろん最初に与えられたものと同じ形状ではない。対象が近づいてきていれば大きく見えるだろうし、回転しているかもしれない。こうして変形した物体を、どのようにニューラルネットワークで表すのだろうか。

ここで、輪郭を数式で表していることが生きてくる。数式で表すことにより、変換が容易に行えるからだ。ニューラルネットワークでは、 $x', y'$  (初期値) から  $x, y$  (2番目の画像中の形状) への変換を自動的に行える (  $\theta$  で回転、 $A$  で拡大・縮小、 $a, b$  で平行移動 )。

$$\begin{pmatrix} x \\ y \end{pmatrix} = A \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$$

これに対応するのが図4の左側の部分だ。そして、図の右側の部分で輪郭の近似が行われる。これは先の図3を拡張したものだ。

しかし、ここで1つの疑問が生じる。上に挙げたような変換で、形状の変化した画像を確実に表現することはできるのか、と。しかし先述したように、1秒の動画を表すのに30枚の静止画を使うのだから、ある静止画から次の静止画に移るまでにかかる時間は約30分の1秒である。つまり、よ

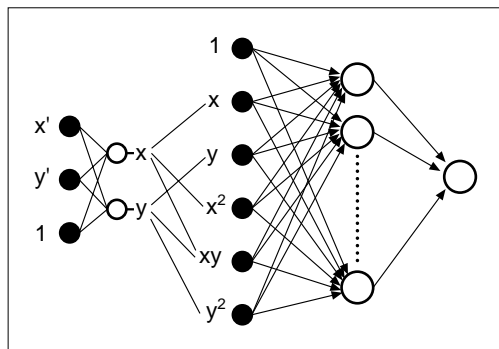


図4 変換の仕組み

ほど動作が速くない限り、その2枚の画像間では対象の位置や形状の変化はごく微小であるので、変換を行うことでほとんどの場合近似できる。

ここで重要なのが、新しくネットワークを構成し直すのではなく、もとのネットワークの記憶から変換を施すことによって近似するということだ。これは前者に比べ計算が短時間で済むので、追跡の速度が飛躍的に向上する。現在、1枚の画像から対象の輪郭を近似するのに2秒から5秒ほどしかかからないそうだ。

以上のことを繰り返して、実際に追跡したのが図5だ。このような複雑な形でも、複数の直線と2次曲線を組み合わせることで、比較的きれいに近似できている。また、対象の運動に伴う様々な変形にも対応して追跡していることがわかる。

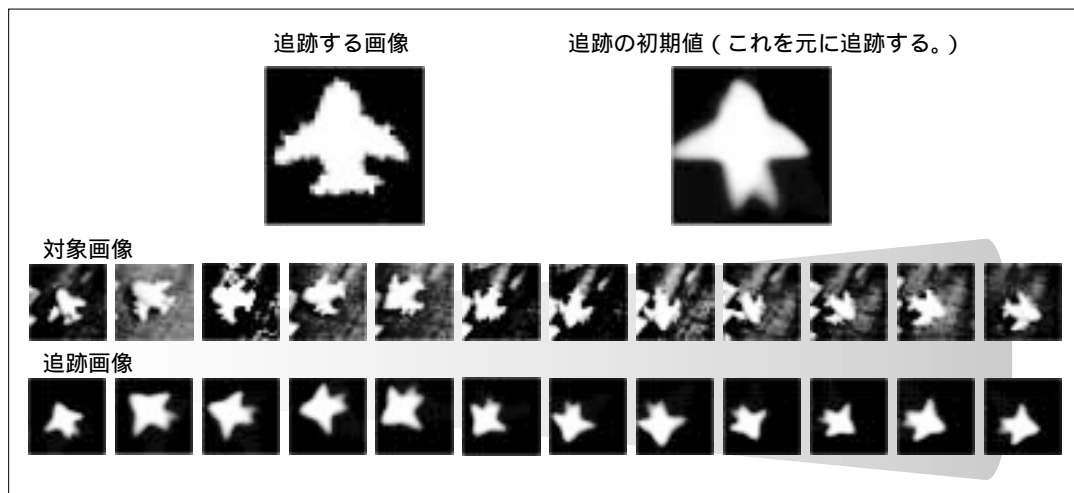


図5 動く物体の近似と追跡



## 物体追跡技術の応用

物体追跡技術にニューラルネットワークを使うことによって様々な利点が生まれることは既に述べた。しかし、現実には追跡すべき物体はほとんどが3次元の情報を持つにも関わらず、今は2次元の追跡しかできていないことや、追跡に大変時間がかかることなどが、実用化に際して解決すべき問題点だ。

ボールのようにどこから見ても形のあまり変わらないものならば2次元でも十分追跡できる。しかし手やはさみなどの物体は回転する方向によって全く違った形状になるため、コンピュータが追跡対象を認識できなくなってしまうかもしれない。これは物体の輪郭を近似する方法で追跡するために生じてしまう問題とも言える。

そこでこの問題を解決するために、2次元の追跡から3次元へと応用すべく研究を行っている。だが新たな問題として、3次元では視点によって対象の背後にまわって見えなくなる部分が出てくることが挙げられる。そのため複数の視点を用意しなければならない上に、そうした背後の情報をも記憶させるために膨大な情報量と計算量が必要になってしまう。しかし、熊澤研究室では既に物体の形をニューラルネットワークで表すことに成功している。そして、来年度中にも3次元での物体追跡を実現させる予定である。

また、追跡にかかる時間についての問題だが、現在普及しているようなコンピュータの処理速度が向上しても、リアルタイムで物体追跡を行うのは難しいと言われている。その理由として、現在のコンピュータは1つの処理を1つのコンピュータに割り振るので、速度が速くても処理時間に限界があることが指摘されている。そこで、処理を複数のコンピュータに分散させることで処理時間

を短くできる、物体追跡専用のハードウェアを先生は提案している。これが実現すれば、動く物体をリアルタイムで追跡することも夢ではない。

では具体的に、動く物体を瞬時に追跡できるようになれば、どんなことに応用できるだろう。

まず挙げられるのが、自動車の自動走行だ。各自動車メーカーが競ってこれを研究・開発しているが、実現するには、前の車に合わせて自動的に速度を変える技術が不可欠である。これに大きく貢献するのが物体追跡技術だ。例えば前の車の速度が上がれば遠ざかっているのが小さくなっていくように見えるし、遅くなれば近づいてくるのが大きくなっていくように見える。これを利用して、小さく見えればアクセルを踏み、大きく見えればブレーキを踏むというような一連の行動を記憶させておけば、自動的に速度を調節できる。

また、ロボットへの適用も期待されている。現段階でも、あらかじめ情報を与えたり行動を記憶させたりしておけば、一見人間と同じような振る舞いをさせることもできる。しかし、ロボット自身に判断・行動させようとするときに不可欠なのが視覚機能だ。その中でも、3次元物体を認識・追跡する能力は特に重要である。

例えば、ロボットにテーブルの上の積木を取らせるとしよう。と言っても、これには物体追跡が必要ないと思えるかもしれない。しかしロボットが動くことによって、積木の位置や形状が変わってしまうため、3次元での物体追跡が行えなければ、実現は難しい。これ以外に歩行やスポーツをさせる時にも同じことが言える。

将来、ジェスチャーを理解できるようなロボットができれば……そういう夢を持って研究していきたい、と先生はおっしゃっていた。

---

本文中では物体追跡について述べたが、ニューラルネットワークを使って物体の輪郭を近似するという方法は画像圧縮にも応用できる。

画像中に人と背景が含まれている場合に、コンピュータがそれらの輪郭を認識できれば、人はほとんど圧縮せずに背景の圧縮率を上げることで、

全体の圧縮率を上げることができる。

これまで、全体を均一に圧縮する技術が主流だったが、このように重要度によって圧縮率を変えられるような、理想的な圧縮技術が実用化される日もそう遠くないかもしれない。

(木佐貫 奈央)