ロボットの眼は人間の視覚にどこまで近づけるか？

――現在のコンピュータシステムと脳型ハードウェア――

九州工業大学 大学院生命体工学研究科 脳情報専攻

森江 隆

URL: http://www.brain.kyutech.ac.jp/ ˜morie

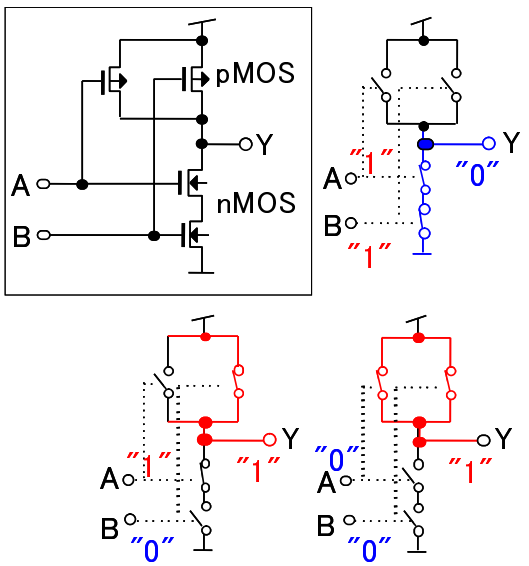
e-mail: morie@brain.kyutech.ac.jp

1　はじめに

昨今，人型・ペット型ロボットなどの開発が進み，テレビでその様子を見たり，博覧会で実物を見たりした人が多いことと思います。最先端のロボットには人の顔を認識し，あたかも人間と会話をしているようにおしゃべりをするものもあります。このようなロボットの物や情景を見る力（視覚）はどの程度のレベルにあるのか，人の視覚能力に匹敵する物はできるのか，この講演では，そういう観点から，人の視覚の不思議さ，すばらしさをわかって頂くとともに，それに近づこうとする工学的な試みについて，お話ししたいと思います。ロボットの視覚といっても，実際に行っているのはシリコンでできた集積回路（LSI），つまりコンピュータです。この講演では，著者の研究室で行っている集積回路を使った人工視覚の試みの一部を紹介します。

2　コンピュータと脳の視覚

私たち人間は雑踏の中からでも友人の顔を瞬時に見分けることができます。また，目隠しをされてどこかに連れて行かれても，目隠しをはずされた瞬間に周囲の情景を瞬時に判断・理解することができます。しかし，このような人間にとってたやすいことが，現在のコンピュータでは全くできません。



論理機能	NAND (NOT AND)
論理関数	$Y = \overline{A \cdot B}$
論理記号	A, B, Y
真理値表	A B Y: 0 0 1; 0 1 1; 1 0 1; 1 1 0

図 1: デジタルコンピュータの基本素子（CMOS NAND ゲート）（入力が 1 のとき，nMOS はオン，pMOS はオフ。入力が 0 のときはその逆になる。）

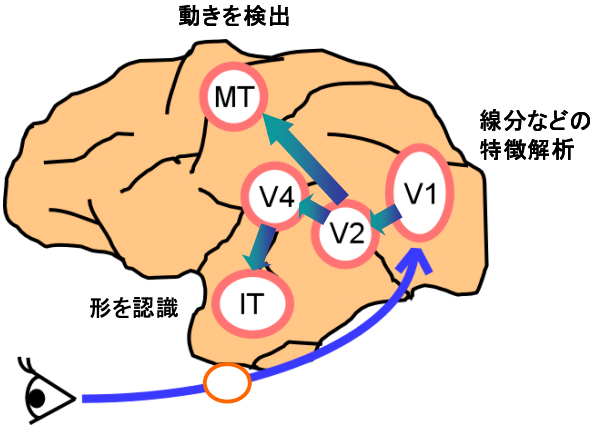


図 2: 脳の視覚システム

今のコンピュータは，トランジスタのオン・オフの組み合わせだけで実現された，2 進数とブール代数を実行する「デジタル」方式です。最も簡単な論理回路である 2 入力 CMOS NAND ゲートと呼ばれるものを図 1 に示します[1]。図に示した「真理値表」は二つの入力 A, B が 0 か 1 をとる場合のそれぞれで出力 Y が 0,1 のいずれになるかを示したものです。二つの入力がともに 1 になったときだけ出力が 0 になる，つまり AND の否定（NOT）ということで NAND と呼ばれます。これを組み合わせればすべての論理と四則計算が実現でき，あらゆるデジタル的な計算ができるということで，現在のコンピュータは「万能計算機」とみなされます。これまで，デジタルコンピュータの代表であるパソコンでは，このデジタル計算を行うただ一つのユニット（CPU）を高速化することで進歩してきました。現在では動作周波数が GHz のレベルに達しており，10^9 回/秒の処理を行えるようになっています[2]。

では，今のコンピュータ（ロボット）ではどのようにして視覚を実現し，情景を理解するのでしょうか？ この研究分野はコンピュータ・ビジョンと呼ばれ，長い研究の歴史があります。ただ，驚くべきことに，現在に至るまで，脳の仕組みを真似た視覚ハードウェアを作るということはほとんど行われていません。コンピュータ・ビジョンでのほとんどすべての努力は，現状のコンピュータ上でのソフトウェアでの実現を前提に，計算法をさまざまに工夫することに費やされてきました。現在まで，わずかに眼の網膜レベルのハードウェア化が試みられているにすぎません。

一方，脳は，膨大な数（$\sim 10^{11}$）の神経細胞（ニューロン）が，ミリ秒オーダの処理スピードしかないけれども膨大な結合（ニューロン当たり $\sim 10^3$）を介して動作している超並列システムです。人間の優れた視覚能力も，この超並列フィードバックシステムというハードウェア上で実現されています。私たちは物や情景を瞬時に何の苦労も無しに認識しているのでその大変さがわかりませんが，実はそのために様々な画像処理と膨大な計算を行っています。図 2 は眼から入った信号が脳の後ろ（後頭葉）の初期視覚野（V1）に入り，内部で処理されていく様子を簡単に示しています。おもしろいことに，脳では，V1 辺りで入力画像を線分などの特徴にバラバラに分解した後，側頭葉に向かう「形を認識する」経路と，頭頂葉に向かう「動きを検出する」経路に分かれて処理が進みます。その後，それらの処理結果が統合されて，意識に上る認識に至るわけです。

このように，現代の大脳生理学の研究から脳の各所の機能分担などはわかってきていますが，視覚という機能が全体としてどのように実現されているかという「モデル化」の部分はまだわかっていないことが極めて多いのです。さらに，たとえ視覚のモデルができたとしても，ロボットの眼が直ちに人間のようになるかというとそうではありません。人間とロボット（コンピュータ）ではそのハードウェアの構成が全く異なるからです。本稿の以下の節では，ロボットの眼を人の視覚に近づけるための研究の一

[1] 二つの相補的動作をするトランジスタ pMOS と nMOS を組み合わせた回路が CMOS です。

[2] しかし，昨年，パソコンの CPU で大きなシェアを有しているインテルという会社は，とうとうこの 1 ユニットでの高速化という路線を修正して，複数の CPU を載せたチップを作るようになりました。高速化に限界が来て，コスト的に見合わなくなってきたからです。

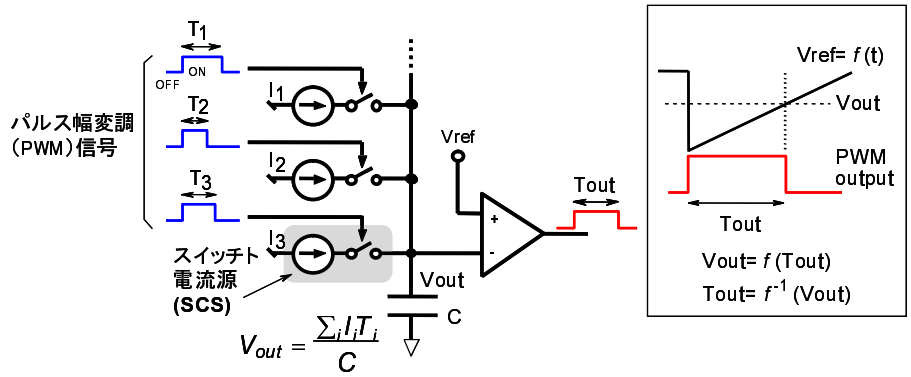


図 3: スイッチト電流源での積和演算の原理：アナログ・デジタル融合回路

例として，視覚機能のモデル化を中心に，現在のデジタルコンピュータの構成とは異なる，脳の機能・構成を真似た「脳型ハードウェア」の試みについて紹介します。

3　パルス信号を用いた回路方式：アナログ・デジタル融合回路

上に述べたように，コンピュータで主流のデジタル方式は，スイッチのオン・オフに相当するビットという単位で情報を表現します。この方式では，ビット数を多くすればいくらでも計算精度を上げることができますが，回路規模と消費電力がどんどん大きくなります。脳を真似た計算をする場合，計算精度はそれほど必要ではありません。実際，神経細胞が行っている計算精度は数ビット程度といわれています[3]。その代わり，多くの回路が並列的に同時に動作することが重要です。デジタル方式は回路の素子数が多くなり，LSI チップ上の占有面積が大きくなるために，たくさんの回路をチップ上に並べることが難しく，超並列動作には向いていません。

そこで私たちはこの目的のために，パルス信号を用いた回路方式を提案しています。私たちが主に用いているパルス信号は，電圧方向では 2 値ですが，時間軸方向（パルス幅）に連続的な値（アナログ値）を有する「パルス幅変調（PWM）信号」と呼ばれるものです。図 3 はこの PWM 信号で積和演算（掛け算と足し算）をする回路の原理を示しています。つまり，PWM 信号で電流源をスイッチするとパルス幅の時間だけ電流が流れるので，電流値 I_i とパルス幅 T_i の積が電荷としてコンデンサ C に流れ込みます。もし，このスイッチと電流源の組（スイッチト電流源: SCS）を共通のコンデンサに接続しておけば，キルヒホッフの第一法則により電流の加算が自動的に行われるため，積和演算 $\sum_i I_i T_i$ が簡単に実行できるわけです。コンデンサに溜まった電荷は電圧 V_{out} として表現されますから，これを時間的に変化する電圧波形 V_{ref} と比較することで，パルス幅 T_{out} を有する PWM 信号を作り出すことができます。私たちはこのように，半デジタル的な信号（PWM 信号）でアナログ的な演算を行う回路方式を「アナログ・デジタル（AD）融合方式」と呼んでいます。AD 融合方式は積和演算を少数の素子から成る簡単な回路で実行できるので，チップ上にたくさんの演算回路を搭載することができ，超並列でかつ低消費電力で動作するシステムを構成することができます。

4　視覚系の脳の入口，第一視覚野（V1）のモデル：ガボールフィルタ

視覚系の脳の入り口である第一視覚野では，画像のエッジの方向および濃淡周期の検出を行うガボール変換（フィルタリング）が行われています。ガボール変換は図 4（左）に示すように，cos 波と sin 波をある範囲内に局在化させた波形（カーネル）で畳み込み[4]を行う演算で，数学的には複素数関数で定義

[3]例えば，5 ビット精度とすると $2^5 = 32$，つまり 32 レベルの値が表現できます。

[4]図 9 を参照。

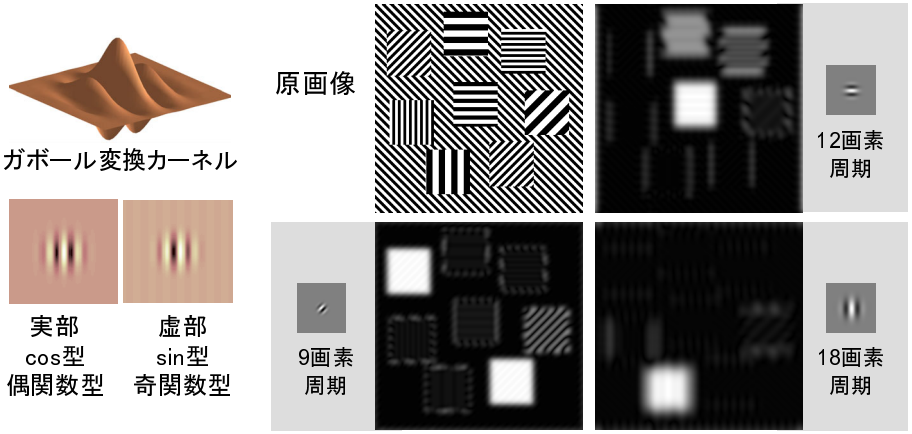


図 4: ガボールフィルタカーネルと縞模様（濃淡周期）の検出

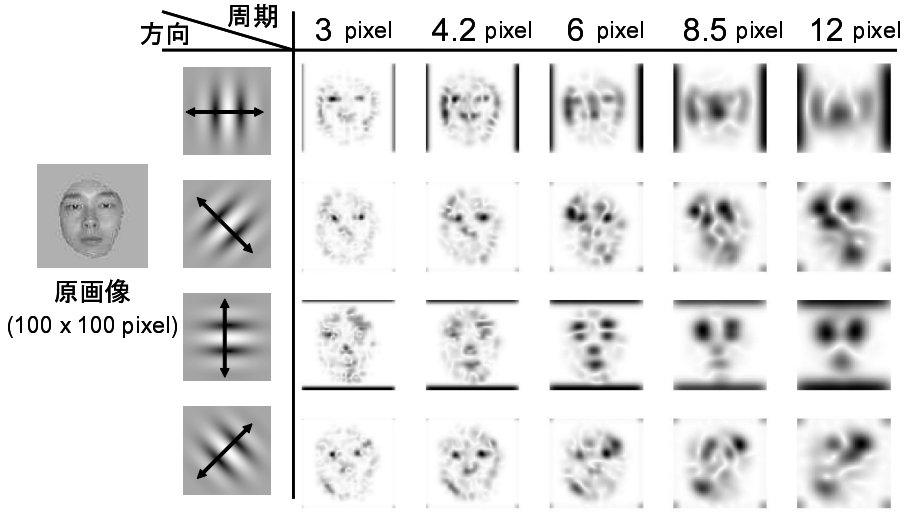


図 5: ガボールフィルタによる顔の解析結果

されます。例えば図 4（右）では，様々な方向と濃淡周期のパターンをガボールフィルタで検出している（図で白くなった部分）ことがわかります。また，人の顔に適用した例を図 5 に示しますが，このような特徴量を用いて，顔を認識する方法が実用化されています。

この変換は照明の変化などの影響を受けにくいので，物体認識のための優れた特徴抽出法として知られていますが，膨大な演算量を必要とするため，これまで実用的な認識システムに採用された例はそれほど多くありませんでした。そこで，このタイプの変換をハードウェア的に実行するために，抵抗ネットワークで計算する回路モデルとアルゴリズムを開発しました。このモデルは画像の画素に相当して回路を用意し，それらが並列で動作する，画素並列方式です。1 画素当たりの回路を図 6 に示します。また，AD 融合方式に基づいて，画素並列動作を行うガボールフィルタ LSI を開発しました。画素回路の LSI パターンレイアウト図とチップ写真を図 7 に示します。このチップは約 1cm 角の大きさに，64x64 画素分の処理回路を搭載しています。

このチップをパソコンと組み合わせて，画像のガボールフィルタリングを実行するシステムを開発しました。図 8 にシステムの概観とストライプパターンを検出している処理画面を示します。

5　形を認識する視覚系のモデル：畳み込みニューラルネットワーク

形を認識する経路での視覚系は，非常に抽象化すると，初期視覚野での線分特徴検出結果から様々な特徴を組み合わせ，複雑な図形を認識していく過程として理解できます。そのモデルとして，畳み込み

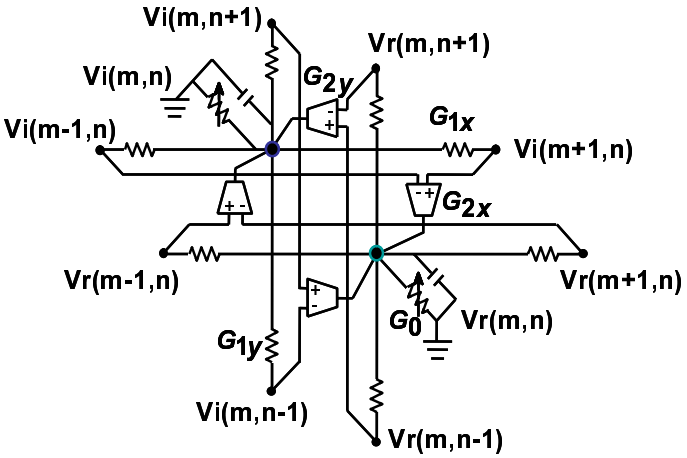


図 6: ガボールフィルタを実現する 2 層抵抗ネットワーク（1 画素当たりの回路）

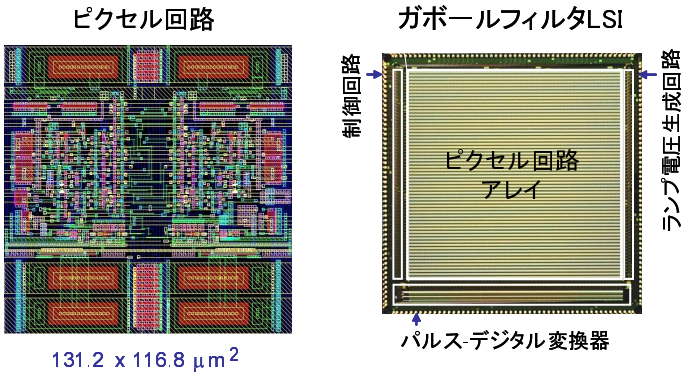


図 7: ガボールフィルタ LSI 画素回路とチップ写真

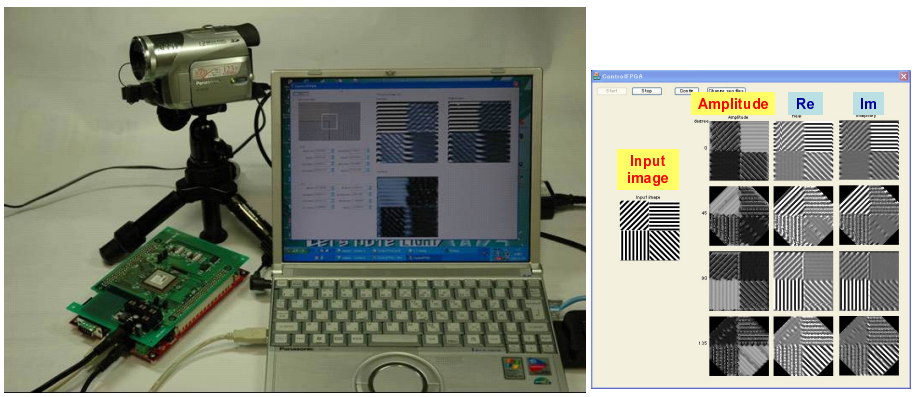


図 8: ガボールフィルタシステムの外観（パソコンの横にあるのが，ガボールフィルタ LSI を搭載したボード）と処理画面（4 方向のストライプパターンが検出されていることがわかる）

ニューラルネットワーク（Convolutional Neural Networks）が知られています。畳み込み（フィルタリング）とは図 9 に示すように，ある画素の近傍画素の値に所定の重み付けをした量を足し合わせる演算です。この重み（荷重）の組をカーネルと呼びます。上に述べたガボールフィルタも畳み込み演算の一種です。

畳み込みニューラルネットワークは，図 10 に示すように，前段の特徴量（特徴クラス）を所定の範囲内（受容野と呼ばれる）で畳み込み計算をする層（FD）と，細かい位置ずれを問題にしないようにぼかし効果を加える層（FP）を交互に配置した多層構造のネットワークです。このモデルでは，膨大な数の積和演算が必要なので，それを高速に実行する LSI チップを AD 融合方式を用いて開発しました。チッ

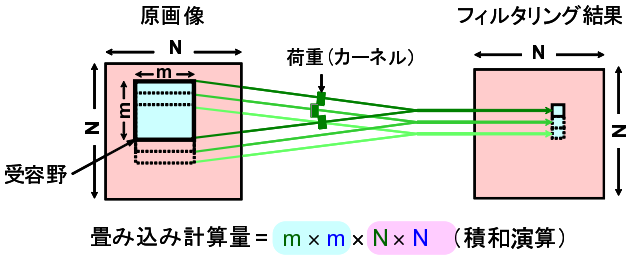


図 9: 畳み込み計算

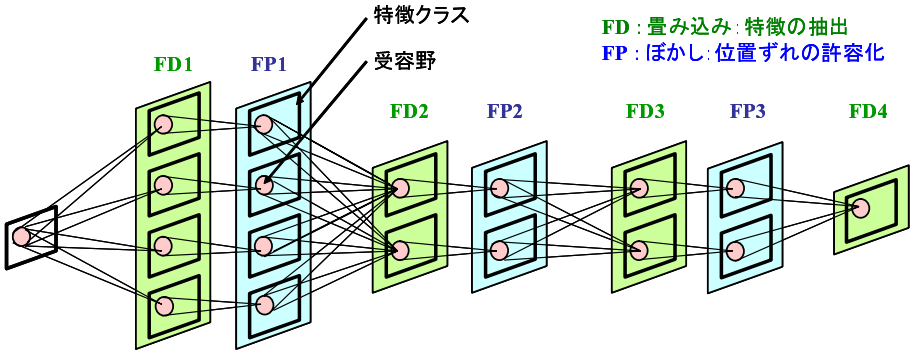


図 10: 畳み込みニューラルネットワークの構造

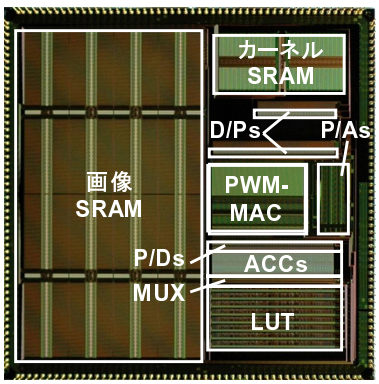


図 11: 畳み込みニューラルネットワーク LSI

プ写真を図 11 に示します。約 1cm 角のチップに，画像やカーネルデータを保存するメモリ（SRAM）や，デジタル値と PWM 信号を相互に変換する変換回路（D/P, P/D）PWM 信号で積和演算を行う回路（PWM-MAC），積和演算結果をさらに足し合わせる回路（ACC），非線形変換を行う回路（LUT）などが搭載されています。この試作した LSI でシステムを構成し，顔の検出に成功しました。図 12 にその様子を示します。

6　大まかな領域を検出する視覚のモデル：抵抗ヒューズネットワーク

理想的な画像と異なり，現実の画像には必ずノイズが含まれています。ノイズを除去するには「ぼかし」を加えればよいのですが，全体をぼかすとエッジの情報もぼけてしまいます。細かいノイズを除去して，かつエッジ情報をきちんと取らないと物と物の境界を認識できません。この機能はまた，画像の細かい部分を無視して，大まかな情報を得ようとするときにも用いることができます。この相反する条件を満たす処理を実行する画像処理モデルに抵抗ヒューズネットワークがあります。このモデルもまた，脳の視覚系で行われている計算の一部を模擬しています。

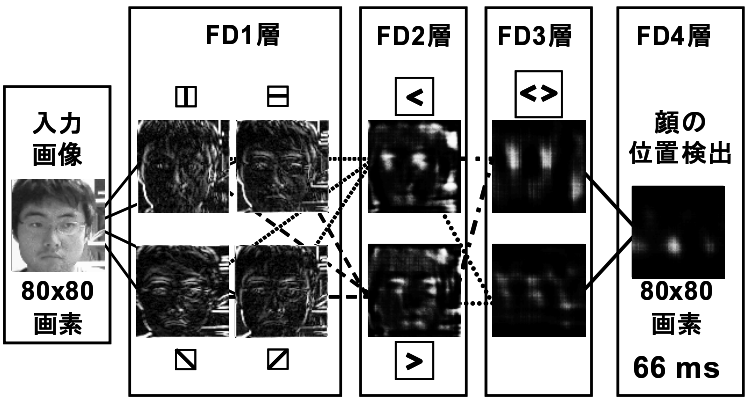


図 12: 畳み込みニューラルネットワーク LSI による顔検出

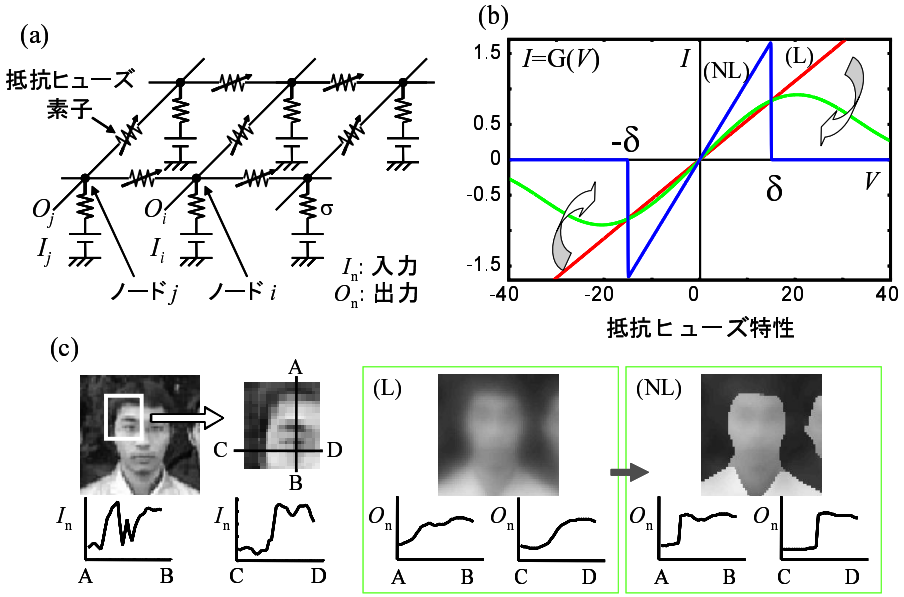


図 13: 抵抗ヒューズネットワークの動作原理

このモデルもガボールフィルタと同様に，図 13(a) に示すように，抵抗のネットワークで実現できます。各ノードが各画素に対応し，各ノードの間は抵抗ヒューズと呼ばれる素子で結合されます。この素子は両端の電圧差が小さいときは抵抗として働きます（電流が電圧に比例する）が，電圧差が大きくなると，結合が切れて電流が流れなくなります（ヒューズが切れた状態になる）。ノード間が抵抗で結合されていると，ある画素の値がその周囲に拡散してしまうので，ぼかしの効果を実現できます。ただし，この特性のままで，いきなり入力画像を処理すると，結合が切れる電圧差（しきい値 δ）を越えるか，越えないかで，ぼかしが入るか入らないかが一意に決まってしまいますので，ノイズを除去しながらエッジを保持するという機能は実現できません。そこで，この素子の特性を最初は通常の抵抗のようにしておき，徐々に抵抗ヒューズの特性に変化させます（図 13(b)）。これにより，最初は画像全体がぼかされ，その後徐々にエッジのように画素値の差の大きいところではヒューズ状態で結合が切れてぼかし効果がなくなり，一方，細かいノイズや領域はぼかしが入って消えてしまうという結果になります（図 13(c)）。こうして，ノイズを除去しながらエッジを残すという処理が実現できます。

私たちはこのモデルを LSI 上で実行させ，通常の動画像で上記の処理を可能にしました。図 14 に私たちが構築したシステムの処理画面を示します。顔の細かい部分が無視され，顔領域だけが取り出されていることがわかります。

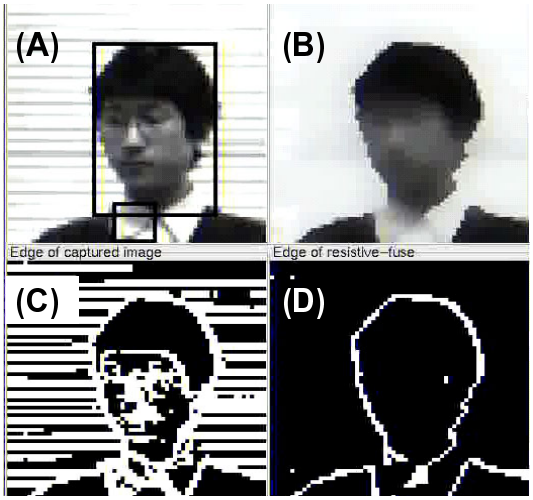


図 14: 抵抗ヒューズネットワークによる処理の例（ブラインドの前に立っている人の画像 (A) をそのままエッジ抽出すると (C) のようにブラインドの横縞や顔の中のさまざまな輪郭が出てしまう。抵抗ヒューズネットワークで処理した結果 (B) をエッジ抽出すると (D) のように顔領域のエッジだけが抽出できる。）

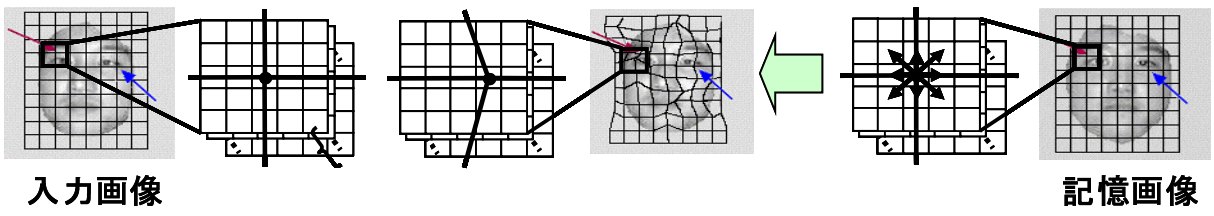


図 15: エラスティック・グラフ・マッチングの原理

7 顔や物体などの歪みに柔軟な認識モデル：エラスティック・グラフ・マッチング

人の顔を認識しようとするときに問題となるのが，表情や照明の当たり具合などで，同一人物でも 2 次元の顔画像は全く同じでない，という点です。そのため，同じ人の顔であっても，2 枚重ねて照合（マッチング）をしても一致することはありません。そこで，顔認識のためにさまざまな手法が提案されています。

脳を真似ていない人工的なモデルでは，まず計算する量が少ないことが重要です。それは通常の逐次処理的なコンピュータで実行することを仮定しているからです。簡単な認識法として，例えば，顔画像を荒っぽいモザイクにしてマッチングをとる方法があります。そこそこの認識精度が得られますが，モザイク化することにより，細かい違いが消されてしまいますので，認識の精度を上げるのは難しいと考えられます。

一方，脳での視覚系の抽象的なモデルとして，ガボール特徴を比較点を変えていきながらマッチングをとる「エラスティック・グラフ・マッチング」と呼ばれる手法が提案されています。図 15 のように，まず，記憶している顔画像と今認識しようとしている入力画像とで，同じ格子を当てはめ，各格子点同士でのガボール特徴の一致度を調べます[5]。次に，一方の格子を少し歪めて，一致度を計算します。もし，歪めた方が一致度が高ければ，その歪み方を採用して，さらに歪めてみます。こうして，格子の変形を繰り返し，最も一致度が高く，かつ格子の歪み具合が小さい条件を探します。この条件で，ガボール特徴の一致度と格子の歪みの小ささの兼ね合いをみる量を計算し，記憶している画像すべてについて最も値の大きな記憶画像を答えとします。

[5]ガボール特徴は先に述べたガボール変換で得られた特徴ですが，その画素近辺での様々な濃淡周期・方向を調べているので，照明の当たり具合にあまり依存しない特徴といえます。

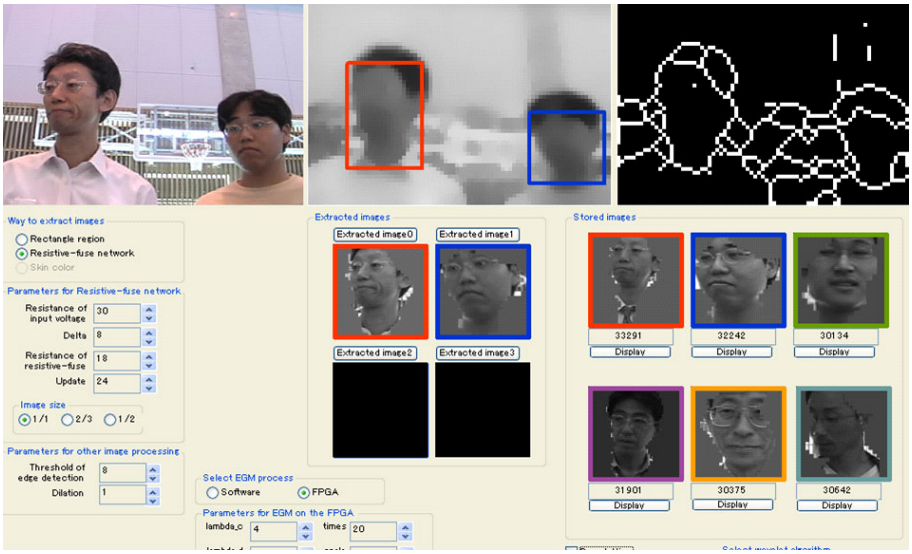


図 16: 顔認識システムの表示画面

容易に想像されるように，この方式は計算する量が極めて大きくなります。それでもこの方法は認識精度が高いので，計算の仕方を工夫して，ソフトウェア処理で実用化されています。

私たちはこの「エラスティック・グラフ・マッチング」を LSI 上で実行するシステムを開発しました。先に述べた抵抗ヒューズネットワークによる「大まかな領域抽出」により顔領域を取り出し，それが誰の顔かを認識することができます。そのシステムの処理画面を図 16 に示します。

顔認識と一口に言っても様々な条件があります。一般に，その人が自分を認識してもらおうとしている場合，例えば入室管理しているドアの前で自分の顔を認識して欲しい場合，銀行のキャッシュディスペンサからお金を引き出したいために，顔を認識して欲しい場合，などでは，人が認識しやすいように振る舞います（あえて変な顔はしない）し，周囲に紛らわしい背景はありません。顔だけをカメラの前にきちんと出すでしょう。このような状況では人工的な方式でも高い認識精度が得られます。しかし，様々な状況の中で人が認識して欲しいと思ってない場合，例えば雑踏の中で友人を見つけるような場合は，脳に学んだ，計算量が多いけれども，確かな認識が可能な方式が必要になってくると考えられます。

8　主観的輪郭を作り出すモデル

図 17（左）に示すような図形を見てください。なにもないところに正三角形が見えると思います。この図形はカニッツァの三角形と呼ばれ，脳の視覚系が物の輪郭を補うことによる効果を示しています。これを「主観的輪郭」と呼びます。これを再現するモデルを考え，現在集積回路化を図っています。

このモデルでは，図 17（右）に示すように，まず黒いパックマン[6]の三角形の頂点を形成する部分の方位をガボールフィルタを用いて検出します。次に各画素に状態量を定義し，ガボールフィルタで検出した位置からその方位に状態量を拡散をします。向かい合うパックマンからの拡散が重なり合うと，次にそれを細線で近似し，最後にその細線を合成して，カニッツァの三角形を再現します。ここで処理のポイントとなる拡散は隣り合う画素間で行われ，画素並列で動作するので，集積回路上で動作させるのに適しています。同様な処理は脳の中でも行われていると推測されています。

[6]昔のゲームに出てきたキャラクタ

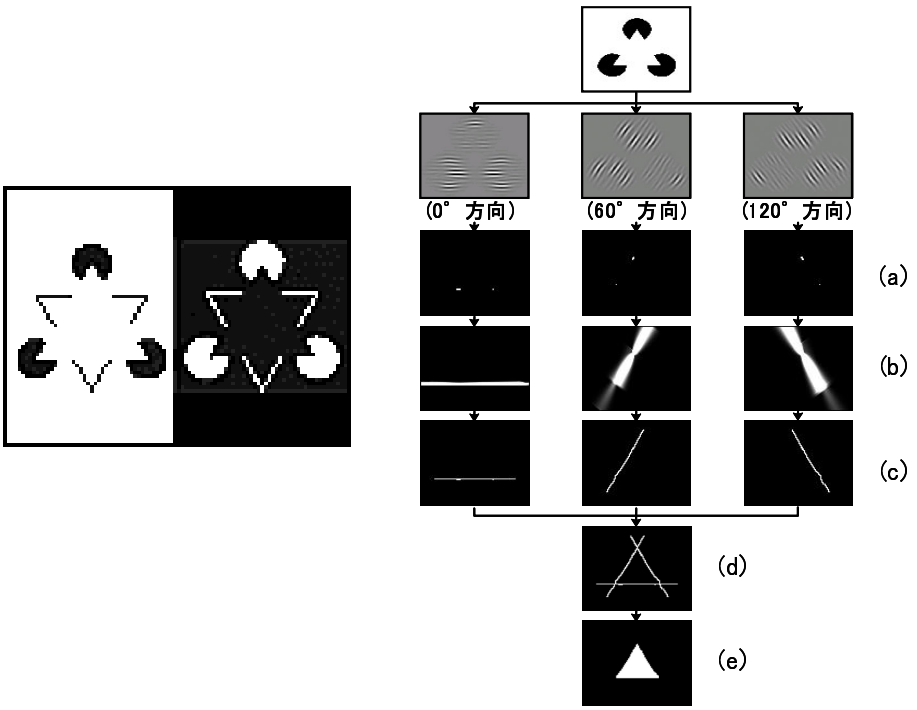


図 17: 主観的輪郭の生成

9 おわりに

以上，脳の視覚系をモデルとした処理とその集積回路での動作例を紹介しました。ここに紹介したのは視覚処理の「形の認識」に関するごく一部であり，脳では，他に動きの検出，色の検出など複雑な処理を同時にかつ瞬時に行っています。既存のコンピュータで各処理を瞬時に行わせるにはそれぞれで最新の高性能パソコンを必要とします。本稿では，それを LSI チップで行った結果を示しました。しかし，その性能はごく限られたレベルであり，人間が行っている処理の足元にも及びません。例えば，ガボールフィルタ LSI では 4 方向で数周期の計算ができるだけですが，人間は数度刻み方位で非常に広い範囲の空間周波数（濃淡周期）を検出しています。

このように考えると，現在のコンピュータ技術（集積回路技術）が今後さらに進んでも，人間の視覚能力に到達できるとは到底思えません。しかし，困難であるからこそ，チャレンジする価値があります。工夫すれば，限定された状況ながら優れた視覚をもつロボットを作り出せる可能性はあります。また，実用的価値だけでなく，脳を模倣する努力の中から脳を本当に理解するきっかけが生まれる可能性もあります。今後，この方面に若い方々が興味をもって，活躍して頂けることを期待します。

参考文献

参考文献は特に示しませんが，著者のホームページ（http://www.brain.kyutech.ac.jp/ ˜morie）で関連文献を紹介しています。