

# 唇検出による非接触キーボードおよびポインティングデバイス

太田広志・白滝順・富川武彦

電気電子工学科

None-contacting keyboards and pointing devices of lip detection

Hiroshi OTA, Jun SHIRATAKI, Takehiko TOMIKAWA

## Abstract

This system supports that physically handicapped persons operate computers. The position and the shape of a lip operates a computer indirectly. A lip is detected on the picture from the video camera using Snakes (active contour model). Therefore, operations which are key input etc., are possible without contacting with input devices. There is also no necessity for attachment and detachment of devices, and we think that this system reduces users' burdens. However, we have not reached to stable detection of the lip in various lighting conditions in the present system. It is expected by adding the further improvement that it is usable also as assistance of a general-purpose pointing device.

**Keywords:** active contour model, human-machine interface, none-contacting

## 1 はじめに

昨今、インターネットの普及などにより、コンピュータをコミュニケーションの道具として用いることが多くなってきている。平成12年版の通信白書によると日本国内の15歳から69歳までのインターネット利用者は約3000万人であり、平成17年には8000万人にも及ぶとされ、さらなるインターネットの普及が進むと予想される。

一方、体が不自由な人のためのPCに対する入力装置は、高価なものであったり、また、取り付け、取り外しやキャリブレーションなどに介助者の助けが必要であるという現状にある<sup>1)2)</sup>。また、コンピュータを利用することができる人とできない人、もしくは利用することが難しい人の情報格差が問題となっている。この情報格差を解消し生活の質を向上するために、主に手足が不自由な人を対象として、手足や頭部などに接触せずにキーボードとマウスの代替とするシステムを構築した。本システムでは、非接触であることから、取り付け、取り外し、キャリブレーションの作業をほとんどせずに操作ができることが期待できる。本システムでは、ビデオカメラで取り込んだ頭部周辺の画像をPCに取り込み、Snakes(動的輪郭モデル)<sup>3)</sup>の原理を用いて唇の位置、形状を認識する。そして、その位

置、形状からPC画面上のポインタを操作し、種々のソフトウェアや作成したソフトウェアキーボードを操作する(図1参照)。

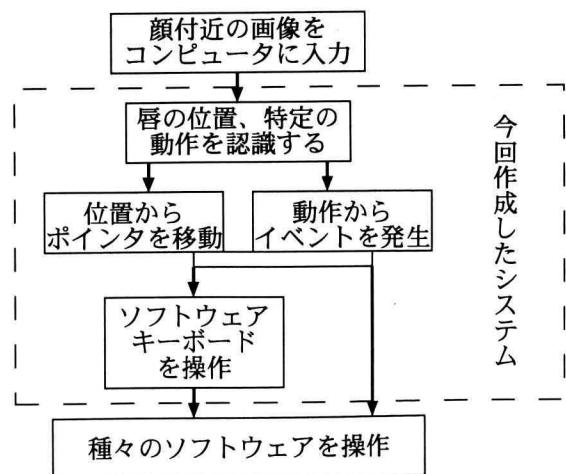


図1 システムの流れ

## 2 唇の認識

### 2.1 Snakes(動的輪郭モデル)

Snakes とは、エネルギー最小化原理に基づく動的モデルである。このモデルでは snake と呼ばれるサンプル点の集合を、物体の輪郭などに近づけるために、サンプル点を結んだ曲線の持つエネルギー  $E_{int}$ 、画像のもつエネルギー  $E_{image,snake}$  に外から加えるエネルギー  $E_{con}$  の和を小さくするようにサンプル点を移動する。

ここでのエネルギーは、例えば、 $E_{int}$  はサンプル点を結んだ曲線の滑らかさをあらわし、滑らかなときに小さくなるエネルギー、 $E_{image}$  は線画などの場合は明るさをあらわし、線上にサンプル点に乗ったときに小さくなるエネルギー、 $E_{cos}$  はユーザの意志(サンプル点を動かしたい方向など)をあらわすエネルギーであるとする事ができる。

$s$  をサンプル点とすると式 (1) のように表すことができる。

$$E = \int (E_{int}(s) + E_{image}(s) + E_{con}(s)) ds \quad (1)$$

Snakes では、画像全てを計算の対象とせず初期状態から次第に移動し、エネルギーの極小値を求めるため、全体的な輪郭抽出には適さない。しかし、前段との相関が強い動画像などの場合は、前段の位置情報から、snake の初期位置を決定することができるため、極小値を求めることで効率的に輪郭の抽出ができることになる。

### 2.2 信号の選択

ビデオカメラからの入力信号 (RGB) のすべてを、唇の位置、形状認識における計算 (上で示した原理に基づく計算) に用いることは、計算コストを増大させる原因となる。

そこで、ビデオカメラからの入力信号を唇の位置、形状認識における計算コストをおさえるために、計算に用いる前に良質の信号に変換する必要がある。

本システムにおいて唇の位置、状態を認識する際、良質のデータとは、唇の色をよくあらわすデータであるということが出来る。

そこで、唇の色を表現するのに適していると考えた、3つの信号 (RGB から生成できる信号) の比較をし、どの信号が唇の輪郭でもっとも大きく変化するかを調べた。その3つの信号は、ビデオカメラから入力した RGB 信号の赤の明るさをあらわす値  $R$ 、色相をあら

わす値  $H^1$ 、入力信号の赤の明るさをあらわす値  $R$  と緑の明るさをあらわす値  $G$  の比率  $R/G$  である。

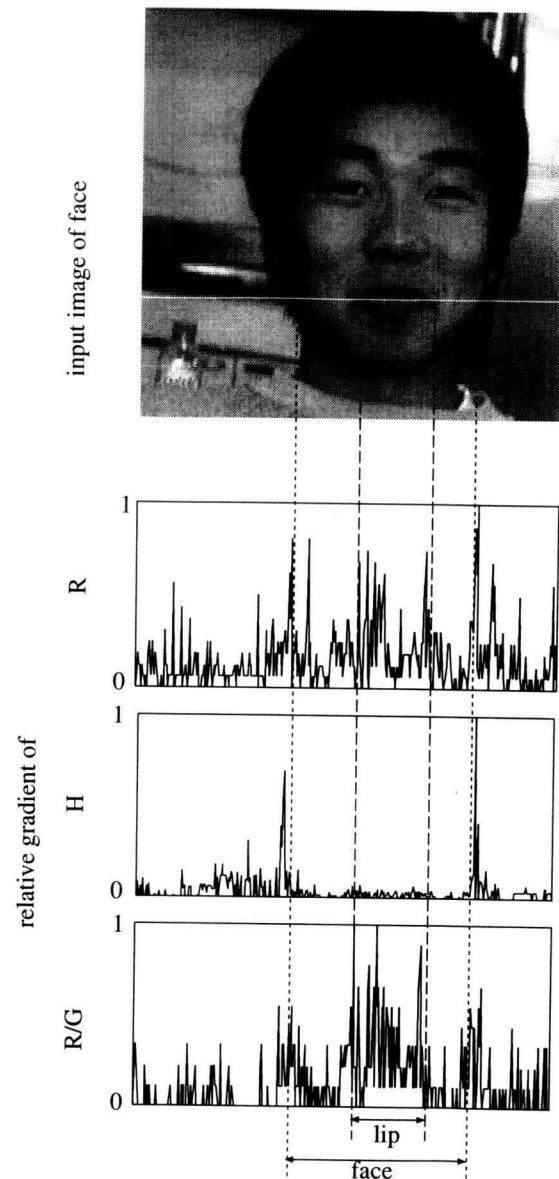


図2 R,H,R/G の相対変化度

図2は、入力画像の唇をとる線上における  $R, H, R/G$  の値の変化 (差分) の最大を1、最小を0

<sup>1</sup>色相をあらわす  $H$  は、

$$H = \begin{cases} \frac{(G-B)}{(V-v)} & (R=V) \\ 2 + \frac{(B-R)}{(V-v)} & (G=V) \\ 4 + \frac{(R-G)}{(V-v)} & (B=V) \end{cases}$$

であらわすことができる。ここで、 $RGB$  はそれぞれ赤、緑、青の明るさをあらわし、 $V$  は  $RGB$  の3信号の中の最大値、 $v$  は最小値である。

として、それぞれの変化を示したものである。

この図2から、このときの入力画像においては、Rの変化の大きい場所は、顔の輪郭および唇の輪郭、Hでは顔の輪郭、R/Gでは、唇の輪郭であるということがわかる。また、他の複数の入力画像でも同様の結果であった。

よって、本システムではRGBの入力画像を前処理として、R/G信号に変換し、後段の唇の位置、形状認識に利用することとした。

### 2.3 静止画像での輪郭抽出

本システムでは唇の位置と形状によって、ソフトウェアなどを操作する。そこで、唇の位置と形状を認識するために、前述のSnakesの原理を用いて、唇の輪郭を抽出する。

本システムでは唇の輪郭の抽出のみが検出できればよいので、前述の唇の色の特徴や唇の形状の特徴を用いてエネルギーを定義することができる。(図3参照)

- $E_a$ : snakeの滑らかさを表すエネルギー  
注目サンプル点の1つ前のサンプル点から注目サンプル点へのベクトルと注目サンプル点から1つ後のサンプル点へのベクトルの内積の $\cos\theta$ の成分に-1をかけたもの
- $E_g$ : 画像の勾配を表すエネルギー  
入力されたR/Gの信号の注目サンプル点と隣接画素の差分を自乗したものに-1をかけたもの
- $E_d$ : snakeの動きの方向を表すエネルギー  
注目サンプル点とサンプル点の集合の重心との距離

全体のエネルギーを $E_a, E_g, E_d$ を用いて表わすと、

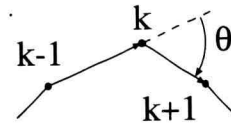
$$E = \sum_{k=1}^n (w_a E_a(k) + w_g E_g(k) + w_d E_d(k))$$

となる。ここで、 $k$ はサンプル点の番号、 $n$ はサンプル点の個数、 $w_a, w_g, w_d$ はそれぞれ $E_a, E_g, E_d$ の荷重である。

しかし、本システムでは計算の簡略化のために注目サンプル点の評価をするごとに、サンプル点を隣接8サンプル点の中からエネルギーの小さい位置に移動している。また、サンプル点を移動するごとにエネルギーの値を更新して次の計算に用いている。そのため、実際には

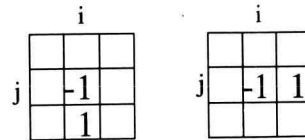
$$E(k) = w_a E_a(k) + w_g E_g(k) + w_d E_d(k)$$

となり、この計算をサンプル点すべてに行い、計算の実行回数を終了条件とする。



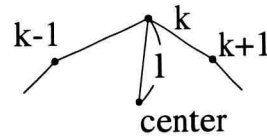
$$E_a = -\cos\theta$$

(a)



$$E_g = -(\text{grad}(\text{image}[i,j]))^2$$

(b)



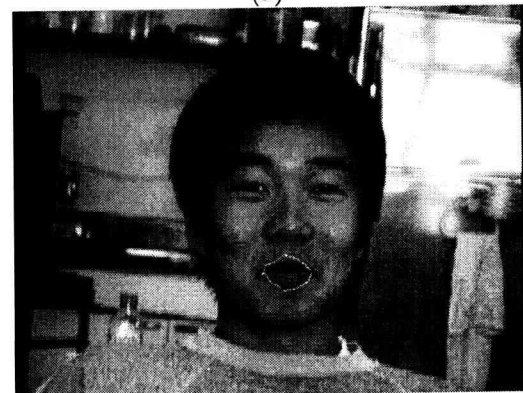
$$E_d = 1$$

(c)

図3 定義したエネルギー



(a)



(b)

図4 静止画像での輪郭抽出の例

図4は実際に唇の輪郭を抽出した例である。初期状態(a)に配置されたsnakeは、エネルギーが小さくなるように、変形し、計算回数が終了条件と合致したところで計算を終了し、(b)で示す唇の輪郭を抽出する。

本システムでは荷重を少しずつ変化させ、経験的にsnakeを唇の輪郭の近くに収束させることができる荷重を見つけ出すことにより決定している。しかし、微小な荷重の変化が輪郭の抽出に対して大きく影響を与えるため、ニューラルネットワーク(NN)や遺伝的アルゴリズム(GA)などを用いることで、適切な値を自動的に設定することができれば、輪郭の抽出が容易になると考えられる。

## 2.4 動画像での輪郭抽出

動画像は、静止画像(フレーム)の連続であると考えることができる。唇の位置は1つ前のフレームの唇の位置とほぼ同じであるため、その唇の位置情報を用いることにより、容易にサンプル点の初期位置を決定できる。また、画像全体においての唇の認識ではなく、局所的なものであるので計算量がすくなく、容易に唇の位置を検出できる。

ここで、1つ前のフレームを利用することは、最初の唇の位置の探索をどうするかという問題を生じる。しかしながら、本システムでは入力画像のほぼ中心に唇が存在するという条件で探索を開始することによって解消できる。これは、利用者が唇を入力画像のなかで任意の位置に合わせることができるため、最初に自分の唇の位置をディスプレイに表示される顔周辺の画像を見ながら初期位置に唇を合わせることができることによる。

## 2.5 検討

これらの唇の位置、形状の認識アルゴリズムにより、ほぼ、唇を追跡し、形状の認識をすることができた。しかし、口を大きく開けるような場合は、唇がsnakeの初期状態の外側にでてしまうため、唇の検出ができなくなってしまうことがあり、今後、形状の変化に柔軟対応できれば、口の動きを認識することができるようになると思われる。

また、唇からsnakeが大きくはずれた場合、唇の位置、形状の認識ができなくなってしまう。そのため、唇の認識ができているかできていないかを判断できるようにし、一定時間認識できていないときは、画面の中央部に戻り、認識を始めるなどの機能を実現することが必要である。

## 3 入力装置のエミュレーション

### 3.1 ポインタの移動

ポインタを移動する操作として、マウスと同様な機構を利用するとマウスを持ち上げる操作を実現しなければならないということがわかる。従って、マウス同様な機構ではなく、ジョイスティックの機構と同様な機構とする方が使いやすいと判断した。ポインタを移動させるために、画像の中心から認識した唇の重心座標の距離( $l$ )と方向を求め、それをポインタの速度と対応させる方法をとった。

ここで、ポインタの速度と唇の位置を対応させるために、線形の式ではなく、高次の式を用いて対応させた。高次の式によって対応させることにより、ポインタをアイコンに合わせるなどの細かい操作、および画面広域の移動などの大きな操作の両方の操作に対応させることができるようになる。

しかし、唇の位置を静止させ、ポインタを目標物の上に留めることは非常に困難である。また、実際に唇の位置が静止しても唇の位置を認識する際にまわりの環境の雑音などで、認識する唇の位置がずれることが容易に考えられる。そこで、自転車のブレーキのように「遊び」の領域を設けることとした。

この「遊び」の領域は画像エリアの中心においた半径 $r$ の円とし、その領域の端の円弧からの垂直距離を移動距離に反映するもので、唇が領域の内部にある場合はポインタは移動しないようにする。

図5は、唇の位置が遊びの領域の内部ではポインタが静止し、領域外部にでるとポインタを高次の式によって決定された速度で移動することを示したものである。

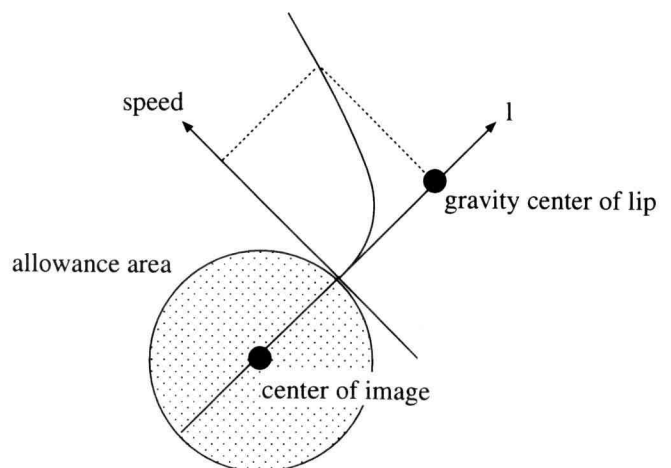


図5 唇の位置とポインタの速度および遊び領域の関係

### 3.2 唇形状の入力

本システムでは唇の形状を認識することでクリックなどを入力するため、唇の形状を唇をつぐむ状態 (図 6(a))、つぼめる状態 (同 (b))、また、通常の状態 (同 (c)) の 3 状態によって認識する。

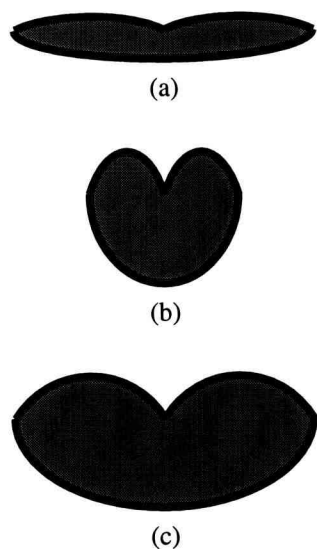


図 6 識別する唇の形状

形状の認識をするために、輪郭を抽出した snake の縦横比 (*ratio*) を用いた。この縦横比は  $x$ 、 $y$  それぞれの最大値、最小値から、幅 (*width*)、高さ (*height*) を求め、その比を式 (2) であらわしたものである。

$$ratio = \frac{height}{width + height} \quad (2)$$

本システムでは経験的に  $ratio < 0.3$  のときに口をつぐんだとき、 $ratio > 0.5$  のとき口をつぼんだときとした。

以上の 3 状態だけでは、一般的にソフトウェアに用いられるダブル「クリック」、「ドラッグ」などを代替することができない。そこでクリックモードを切替えることで代替する方法をとった。このモードの内容を表 1 に示す。ここで、「なし」のモードは、本システムでのソフトウェアキーボードのようにクリックが不必要なときのためであり、「なし」のモードから通常モードに復帰するには、通常モードを選択するトグルボタンの上で 4 秒静止するようにした。

これらのモードで実行することにより、一般的にマウスで操作することができると考えられる。

図 7 に唇の位置、形状を認識し、ポインティングデバイスの代替となるソフトウェアの実行例を示す。

図 7 では、左の表示部で実際に輪郭抽出に用いられる画像 ( $R/G$  の値) と抽出された唇の輪郭が表示さ

表 1 モードによるクリックの種類の変化

モード	つぐんだ状態	つぼめた状態
通常	左クリック	右クリック
なし	なし	なし
中クリック	中クリック	通常モードに
ドラッグ	左ドラッグ	ドロップ
ダブルクリック	左ダブルクリック	通常モードに

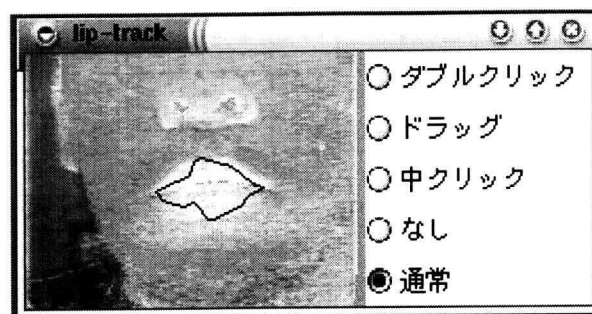


図 7 ポインティングデバイスの代替ソフトウェアの実行例

れるようになっており、右の操作部でモードを変更することができる。また、左の表示部の画像は、鏡像となっており、唇を右に移動すると右に動くようになっている。

### 3.3 キーボード

本システムでは、図 8 に示すソフトウェアキーボードを作成し、文字の入力を必要とするソフトウェアを操作できるようにした。

このソフトウェアキーボードは、文字入力部分の下にあるモード選択ボタンでモードを切替えることで平仮名、カタカナ、英数字、日常的な漢字 (少量) が入力できるようになっている。文字をソフトウェアに入力する手順としては以下の通りである。

- (1) 入力場所選択キーで文字を入力したいソフトウェアの場所を選択する。
- (2) 入力したい入力モードを選択する。(初期状態では平仮名)
- (3) キーを押し文字入力部分に文字を入力する。
- (4) 送信ボタンを押して文字を送信する。

このソフトウェアキーボードでは、クリックを多用するため、上述のポインティングデバイスの代替ソフ



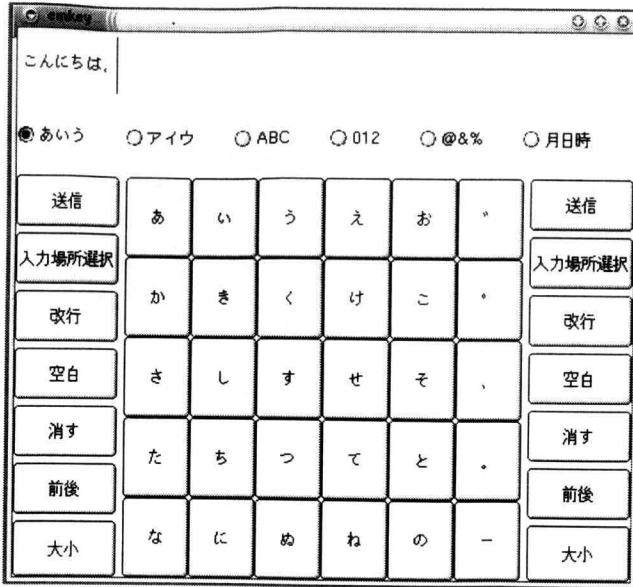


図 8 ソフトウェアキーボードの実行例

ソフトウェアでは、唇の形状を頻繁に変える必要がある。頻繁な唇の形状変化は、利用者にとって負担となるため、キーボード上で4秒静止した場合は、クリックしたものと処理をおこない負担の軽減を図っている。

本システムでの文字入力速度は、「こんにちははおようこんばんは」と繰り返し入力するタスクにおいて、健常者が入力した場合180秒で12文字程度が入力可能であり、1文字あたり15秒程度の速度であった。この速度は、視線入力式コミュニケーション装置<sup>2)</sup>の66文字画面で、1文字あたり5~6秒程度であることと比べて、約1/3の速度となっている。

### 3.4 検討

現時点では、2種類の形状のみを認識しているため、特には前処理や学習を必要としていない。しかし、数種類のクリックを実現するために、クリックのモードを切替えて実現しているが、NNなどを用いて唇の形状が柔軟に認識できるようになれば、唇の形状とクリックの種類を対応させることができ、直観的に操作ができると考えられる。また、唇と肌の間の色の変化が大きくない場合は、口紅を用いるなどの処置が必要となる可能性がある。

ソフトウェアキーボードにおいては、唇の形状が柔軟に認識できれば、子音のキーと母音の唇の形状を用いて、入力ができる、より入力が容易にできると考えられ、文字入力速度の向上が期待できる。

## 4 おわりに

本システムでは、唇の位置、状態の変化を認識することができるようになった。その結果、本システムを利用することで、首を動かすことができれば、手足を使わずにソフトウェアの操作ができ、情報の受信、発信が可能となる。よって、手足が不自由である人のコミュニケーションの手助けをすることができ、生活の質を向上させることができると考えられる。

一方、現時点では、長時間利用するときの利用者への負担や操作性に対する評価が不十分であるため、今後それらの評価を行う必要がある。また、常にシステムを実行していただけるため、介助者の手助けがなくてもPCが操作でき、介助者の負担を減らすことができると期待している。

## 参考文献

- 1) 岡本, 山田, 高木, “制御感を重視した重度肢体不自由の人のための入力装置”, 信学論, Vol.J80-D-II, No.7, pp.1870-1877, 1997.
- 2) 伊藤, 数藤, 伊福部, “重度肢体不自由者向けの視線入力式コミュニケーション装置”, 信学論, Vol. J83-D-I, No.5, pp.495-503, 2000.
- 3) Michel Kass, Andrew Witkin, Demetri Terzopoulos, “Snakes:Active Coutour Models”, International Journal of Computer Vision pp321-331, 1988
- 4) M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision, Books/Cole Publishing Company, 1999.
- 5) 飯尾 淳, Linux による画像処理プログラミング, オーム社, 2000
- 6) 井上 八木 林 中須 三谷 奥井, C 言語で学ぶ実践画像処理, オーム社, 1999.
- 7) 宮本 患博, 計測用カメラを用いたソフトウェアキーボードの製作, 富川研究室 97 年度卒業論文, 1998
- 8) 永野 裕久, 西野 由哲, 非接触キーボードの検討, 富川研究室 96 年度卒業論文, 1997