

# 画像処理を用いた操作入カインタフェース

秋藤 俊介<sup>1</sup>, 徳田 圭世<sup>2</sup>, 広瀬 正<sup>1</sup>

<sup>1</sup>日立製作所システム開発研究所, <sup>2</sup>日立西部ソフトウェア

利用者の動作を認識し、操作入カインタフェースに利用することを目的として画像処理を用いた手形状認識手法を開発している。開発手法は、まずテレビカメラから入力した画像を色の割合の差を用いて手部分だけ抽出し、これと慣性2次モーメントが等価な楕円を使って手首位置を求める。次に手首周りの画素の距離分布を求め、分布曲線の極大点のなかでしきい値を超えた点の個数を計測し、指本数とする。しきい値は手の形状モデルに従って設定した。また、入力画素を2次元空間で等間隔に離散的に選択することにより処理時間の短縮を図った。

## Direct Manipulation Interface using Image Processing

Shunsuke AKIFUJI <sup>1</sup>, Tamayo TOKUDA <sup>2</sup>, Tadashi HIROSE <sup>1</sup>

<sup>1</sup>Systems Development Laboratory, HITACHI, Ltd., <sup>2</sup>HITACHI Seibu Software, Ltd.

This paper describes a new hand shape recognition method for man-machine interface using image processing. The number of the fingers is gotten as follow; (1) extracts a user's hand from a background picture used a hue of the skin color, (2) estimates the wrist position from an ellipse which is the equivalent to the hand in two dimensional inertia moment, (3) calculates the distribution curve of distance around the wrist. The number of the fingers is that of relative maximum of the curves over the thresholds, which set based on hand shape model. The method also uses only the pixels at intervals to reduce processing time.

## 1. 結 言

我々は、TVカメラから入力した利用者の動作を認識し、各種操作を行う非装着型動作認識インタフェースシステムの開発を進めている。人間の動作は、手、腕、頭や足を使った全身の動作など多彩である。本研究では基礎的な機能として背景から手部分を抽出し、伸長した指本数を実時間で認識する手法を開発、評価を行った。本研究と同様に画像処理を利用したユーザインタフェースとして以下のものが発表されている。

- (1) 実画像から得られる情報と人間の運動に関する知識から人物の動きを再構築する手法 [1]
- (2) バイプライン型画像処理装置を用いた手形状認識手法 [2,3]
- (3) 色別されたピンポン玉をつけた手袋の3次元形状を計測するシステム [4]

これらは、実時間処理のため画像処理専用ハードウェアを用いている。実用化のためには、現状のワークステーションで利用される30~100Mips程度のCPU (Central Processing Unit) で処理できることが望ましい。本研究では、入力した画像の中から少数の画素を選択し、それらの分布特性を解析することにより通常ワークステーションでの実現を目指す。以下、2章では手形状認識手法、3章では評価実験について述べる。

## 2. 手形状認識手法

### 2.1 手首からの距離分布による指本数計測

一般に画像処理では、計測の精度を高めるために画像を構成する要素である画素を細かくすることが望ましいが、細かくすれば画素の個数が増加し、処理時間がかかる。本研究では、処理時間を短縮するため、画素全てではなく、ある間隔で選択した画素だけを処理の対象とし、少ない画素から得た情報で形状を推定するために、画素の分布を解析する手法を用いた。

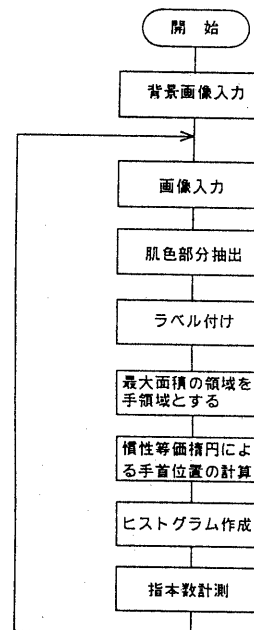


図2.1 手形状認識のフローチャート

提案手法の概略を図2.1に示す。始めに手がカメラの撮影範囲に存在しない状態の背景画像を撮影し、以降毎回、入力した画像と背景の対応する画素の差を求め、この差がある値の範囲内にあるものを肌色部分とする。ここで、手の影などによる背景の変化に影響されないようにするために、単純に2画像間の差を求めるのではなく、肌色を強調するように工夫した。具体的には、式(1)に示すように肌色に多く含まれる赤成分と含まれない青成分の2色について、それぞれ背景と入力画像で差を求め、両者の和を画素間の差とする。

$$dP_{x,y} = |Ri_{x,y} - Rb_{x,y}| + |Bi_{x,y} - Bb_{x,y}| \quad \dots (1)$$

ただし、 $dP_{x,y}$  : 画像間の(x,y)画素の差、

$Ri_{x,y}, Bi_{x,y}$  : 入力画像の(x,y)画素の

赤、青成分の強度、

$Rb_{x,y}, Bb_{x,y}$  : 背景画像の(x,y)画素の  
赤、青成分の強度

肌色を強調ようにした場合では、色の成分が肌色に近いものも存在する。一般的な画像処理で多用されるラベル付け処理[5]を施して連続する画素をひとまとまりの領域とし、各領域の中で

最も面積が大きいものを手領域とする。

手の部分と指先の方向を求めるため、利用者は必ず手首まで袖がある衣服を着用していると仮定し、手領域を慣性2次モーメントが等しい楕円で近似する(図2.2)。この楕円は式(2)から(6)で求める。楕円の長軸方向が指先方向であり、楕円の長軸上の2頂点の内一方が手首位置である。(7)を満たす2個の $\theta$ のうち(8)を最小とするものを楕円の傾きとする。

次に手首を中心として、等角度間隔 $\alpha$ に分割し、それぞれの扇形の領域で手首から手の周囲までの距離を求め、ヒストグラム(分布曲線)を作成する。これを2.2で述べるしきい値を用いて解析し、指本数を計測する。

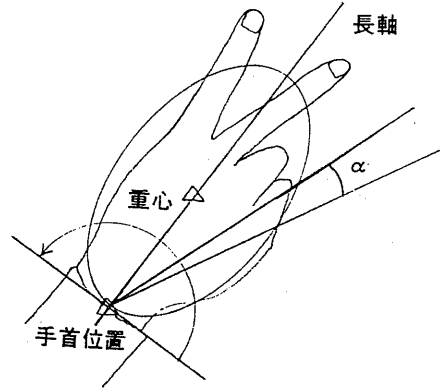


図2.2 慣性等価楕円による近似

$$P_x = \frac{D_x}{N} \sum_{k=0}^{N-1} X_k \quad \dots (2)$$

$$P_y = \frac{D_y}{N} \sum_{k=0}^{N-1} Y_k \quad \dots (3)$$

$$D1 = \sqrt{\frac{2m(2,0)}{\sqrt{m(0,2)m(2,0)}\sqrt{\pi}}} \quad \dots (4)$$

$$D2 = \sqrt{\frac{2m(0,2)}{\sqrt{m(0,2)m(2,0)}\sqrt{\pi}}} \quad \dots (5)$$

$$\text{長径} = \max(D1, D2) \quad \dots (6)$$

$$\text{短径} = \min(D1, D2) \quad \dots (7)$$

$$m(i,j) = D_x^i D_y^j \sum_{k=0}^{N-1} (X_k - P_x)^i (Y_k - P_y)^j \quad \dots (8)$$

$$\tan^2 \theta + \frac{m(2,0) - m(0,2)}{m(1,1)} \tan \theta - 1 = 0 \quad \dots (9)$$

$$f(\theta) = m(2,0)\sin^2 \theta - 2m(1,1)\sin \theta \cos \theta + m(0,2)\cos^2 \theta \quad \dots (10)$$

ただし、 $P_x, P_y$  : 重心のX, Y座標

$X_k, Y_k$  : 手領域の画素のX, Y座標

$D_x, D_y$  : X, Y方向のサンプリング間隔

$N$  : 手領域の画素の個数

## 2.2 しきい値の設定方法

処理の結果得たヒストグラムを解析すると、以下のような特徴があることが分かった。

- (a) 伸長している指は極大値として現われ、小さい場合でも手首から長径以上離れた範囲に存在する。
- (b) 握っている指も、ビデオカメラと手の角度によっては極大値として出現する。
- (c) 伸ばしている指の本数が1本の場合、それが長い指と短い指では慣性等価楕円の長径と短径の比が異なる。ここで、長い指は、人差指、中指、薬指であり、短い指は、親指と小指である。

これらの特徴から、以下のようにしきい値を設定し、指本数を計測する(図2.3)。手首から等価楕円の長径以上離れた極大点を求め、それらを伸長している指の候補とする。これらの候補の中で、しきい値1以上の極大点を伸長した長い指とし、個数を数える。しきい値1は、等価楕円が円に近い場合は、それよりわずかに外側になるよう設定し、等価楕円が細長い場合は、楕円から少しだけ内側になるようにする。

次に、この段階で検出されない短い指の数を計測するため、さらに他のしきい値で指本数を計測

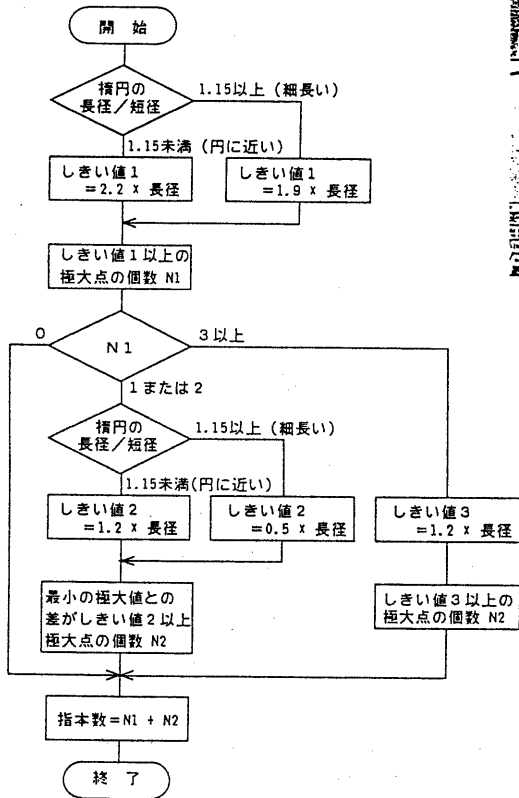


図2.3 しきい値設定手順のフローチャート

し、しきい値1を用いて検出した指本数に加える。しきい値1で、検出された指の本数が3以上である場合は、握っている部分が小さく、極大値として出現する可能性は少ない。逆に、検出した指が1または2本の場合は、握っている部分の影響が大きくなる可能性がある。そこで、3本以上の場合は、単純にしきい値1より小さなしきい値2を設け、それを超えた極大点を求める。しきい値1で検出した指が1または2本の場合は、手首からの距離を用いて計測することは困難なので、折れ曲がった指に対応する最小の極大値を求め、これとの差がしきい値3以上であるものの個数を指本数とする。

例えば図2.4に示すような指を2本だけ伸ばしている場合は、長径以上の極大点a、b、c、

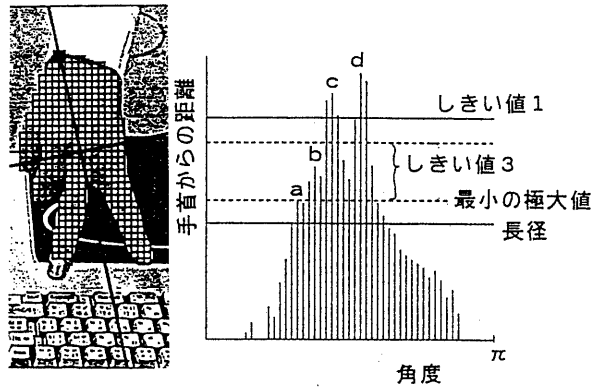


図2.4 手の形状とヒストグラム  
(指を2本だけ伸ばしている状態)

dが伸ばした指の候補となる。まず、しきい値1以上の極大点であるcとdを伸ばした指とする。次に指本数が2なので最小の極大値を求め、これとの差がしきい値3である極大点を求める。この例ではaが最小の極大点であるが、bとaとの差はしきい値3を超えないのでどの極大点も該当しない。従って指本数はcとdの2本となる。

### 3. 実験

#### 3.1 実験の目的と方法

2. 1で述べたように、指本数の計測精度を高めるためには処理対象とする画素を細かく、多くすることが望ましい。しかし、画素の個数増加とともに処理時間も増大する。本実験では、選択する画素間の距離、サンプリング間隔を様々に変化させ、指本数が計測可能な最大の間隔を求め、この間隔で実時間処理が可能であることを示す。

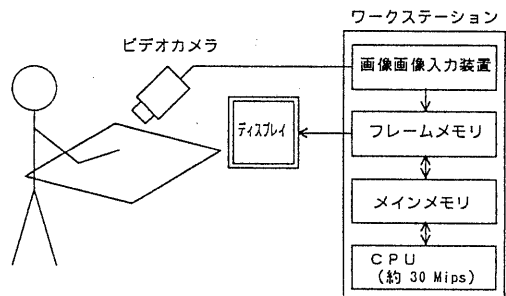


図3.1 実験システムの構成

開発したシステムの構成を図3.1に示す。利用者の画像はビデオカメラで撮影し、画像入力装置を通してワークステーション(約30Misp)に入力する。実験は以下の条件で行った。

- (1) 特別な照明は設けず、室内の天井に設置した蛍光灯を用いた。
- (2) デスクトップでの使用を想定し、机の上にキーボードがある背景に手を重ねた。
- (3) 入力した画像の大きさは、横646、縦486である。

撮影は、手の面積がカメラの撮影範囲に比べて比較的大きなもの/小さなものの2種類、手が手の角度が撮影範囲と垂直方向と一致するもの/斜めなものとの組合せ、計4種類について行った。まず、背景画像用として手が撮影範囲に入っていない状態で数秒間撮影する。その後、手を広げた状態で撮影範囲に差し入れ、順に指を折り、最後に握った状態にした(表3.1)。

表3.1 実験に用いた画像とサンプリング間隔

画像の大きさ	手の角度	サンプリング間隔
大	まっすぐ	24, 16, 8
	斜め	24, 16, 8
小	まっすぐ	16, 8, 4, 2
	斜め	16, 8, 4, 2

正誤は、撮影したビデオ画像の中から1フレームを抜き出し、処理結果として出力される指本数と画像を比較して判定した。同時に処理対象とした画素の個数を数えた。処理速度は、ビデオ動画すべてについて処理を行い、その間の平均処理回数を数えることによって求めた。両者ともにヒストグラムが得られる範囲でサンプリング間隔を変えて処理を行った。

### 3.2 手領域の面積と正誤の関係

処理結果の例を図3.2、図3.3に示す。図3.2は表3.1で手の大きさが「大」で手の角度が「まっすぐ」、指を5本伸ばしている画像をサンプリング間隔16で処理した場合の出力画像である。図3.3は、手の大きさが「小」で手の角度が「斜め」、指を1本伸ばしている画像をサンプリング間隔8で処理した場合の出力画像である。図中、縦横の線分の交差点が用いた画素である。

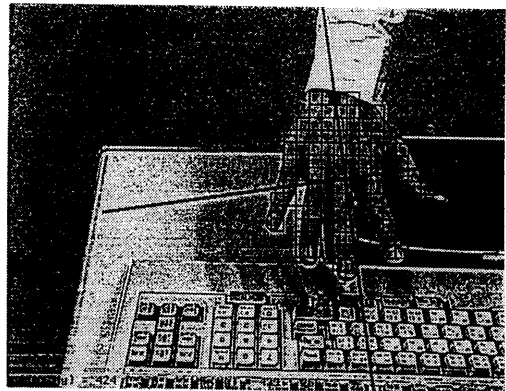


図3.2 出力画面の例(1)

(手の大きさ:大, 手の角度:まっすぐ, 間隔:16)

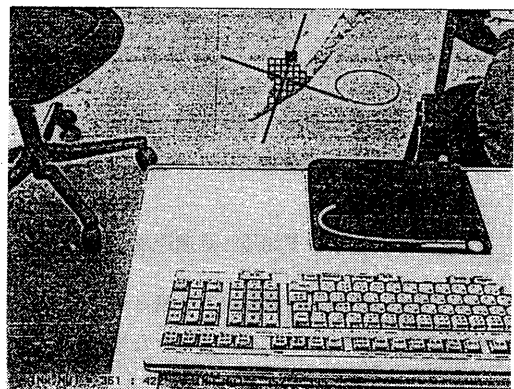


図3.3 出力画面の例(2)

(手の大きさ:小, 手の角度:斜め, 間隔:8)

計測した指本数と処理対象とした画素の個数との関係を図3.4に示す。横軸は処理対象とした手領域の画素の個数であり、縦軸は伸長した指の本数である。図中、□は2章で述べた処理による結果が正しい場合、×は誤っている場合である。処理対象とした画素の個数が多いほど正解を得られる割合が増加していることが分かる。

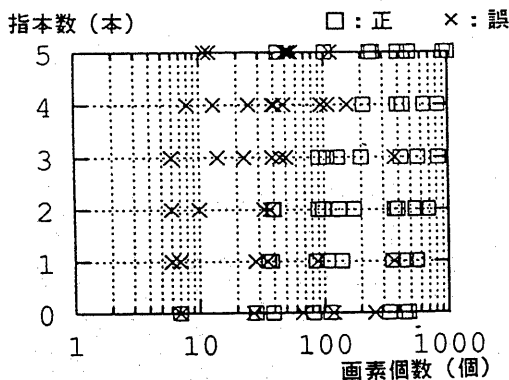


図3.4 手領域の画素個数と正誤

手を上げたときと握ったときの面積の割合はほぼ2倍程度である。従って同じサンプリング間隔を用いたとき対象とする画素の個数の下限と上限の比は約2倍程度となる。この仮定のもとで適当な画素数を求める。図3.4で対象画素の個数が50から100の範囲では、正答率45%(11個中、正解5)であり、100から200の範囲では、67%(12個中、正解8)、200から400では77%(13個中、正解10個)であった。従って70%程度の正答率を得るためには、対象となる画素が100から200の範囲となるようにサンプリング間隔を調整すれば良い。

### 3.3 サンプリング間隔と処理速度

サンプリング間隔と処理速度を表3.2に示す。正答率70%を得るためには、対象となる画素を100から200にする必要がある。これは、手の大きさが「大」では、ほぼサンプリング間隔が16の場合であり、平均処理回数は約10回/秒となる。手画像の大きさが「小」では、サンプリング間隔4

となり、平均処理回数は約2回/秒となる。手が小さい場合、撮影範囲に比較して手の領域は限定されているにもかかわらず、領域を求めるために画像全体で細かく画像間の差を求めており、この処理時間が大きくなったためと考える。10回/秒を得るためには、入力画像をすべて処理しないで、画像中に手の含まれる処理ウインドを設け、この範囲だけで処理するようにする必要がある。

表3.2 サンプリング間隔と処理時間

画像の大きさ	サンプリング間隔	処理速度(回/秒)
大	2 4	16.32
	1 6	10.68
	8	3.76
小	1 6	13.32
	8	6.28
	4	1.75
	2	計測不能

## 4. まとめ

- (1) ビデオカメラから入力した手の画像の指本数を計測する手法を開発した。まず、ビデオカメラから入力した画像を色情報を用いて手部分だけ抽出し、これと慣性2次モーメントが等価な楕円から手首位置を求める。次に手首から手の輪郭までの距離分布を求め、段階的に設定する複数のしきい値を用いて分布曲線の極大点の個数を計測し、指本数を計測する。入力画素を等間隔で離散的に選択することにより処理時間の短縮を試みた。
- (2) 適切なサンプリング間隔を得るため実験を行った。正答率と処理対象画素の個数との関係から、70%程度の正答率を得るためには、対象となる画素が100から200の範囲となるようにサンプリング間隔を調整すれば良いことが分かった。
- (3) サンプリング間隔と処理速度の関係から、正答率70%を得るためには、手画像の大きさが

大の場合でサンプリング間隔が16である必要がある。この場合、平均処理回数は約10回／秒となった。手画像の大きさが撮影範囲に比較して小さい場合、ヒストグラムを得るためにはサンプリング間隔を小さくする必要があるが、手部分を抽出するための処理時間が大きくなる。実時間処理のためには、入力画像をすべて処理しないで、画像中に手の含まれる処理ウインドを設け、この範囲だけで処理するようにする必要がある。

#### 参考文献

- 1) 石井浩史, 他 : ステレオ画像を用いた人物の3次元動作の認識と再構成, 「画像の認識・理解シンポジウム(MIRU' 92)」, 情報処理学会, 電子情報通信学会, pp.135-142 (1992).
- 2) 石淵耕一, 他 : バイプライン型画像処理装置を用いた実時間手形状認識, Human Interface, Vol.7, pp.275-280 (1992).
- 3) 渡辺和文, 他 : バイプライン型動画画像処理プロセッサの共同作業への適用性, テレビジョン学会技術報告, IPCV92-19 (1992).
- 4) 岡本恭一, 他 : 定性的運動認識を用いたヒューマンインタフェースシステム, 電子情報通信学会論文誌 D-II, J76-D-II, No.8, pp.1813-1821 (1993).
- 5) 安居院, 他 : 画像の処理と認識, pp.67-68, 昭晃堂(1992).