

例示的動作入力支援のための画像処理による 指本数計測手法の提案

3 J-4

秋藤俊介¹ 徳田圭世² 広瀬正¹ 矢島敬士¹

¹ 日立製作所システム開発研究所 ² 日立西部ソフトウェア

1. はじめに

人と人の意図の伝達手段には文字や音声など言語的なものと表情や動作など非言語的なものがある。ある分析によると通常の会話では非言語的な手段による伝達が約70%を占めているといわれている[1]。計算機と人間のインタフェースでは、コンピュータグラフィックス(CG)に代表されるように計算機側からの非言語的な手段は整いつつあるが、計算機に入力する手段としてはマウスやトラックボールなど位置を伝える装置があるに過ぎない。計算機を新しい応用分野へ展開する上で、非言語的な入力手段による意図の伝達が重要である。そこで、非言語的な伝達のなかで人間の動作、特に手や腕の意識的動作である例示的動作を入力し、計算機の各種操作インタフェースに利用することを目的として画像処理による動作認識インタフェースを開発している。例示的動作は手や腕を使って示されることが多く、これらの形状と動きを細かく検出する必要がある。本発表では、この最初の試みとして手の伸ばした指の本数を計測する手法を提案する。

提案手法では、ヒューマンインタフェースに用いる観点から検出精度(正答率)よりも応答時間を重視し、正確さを求めるために入力画素すべてを処理するのではなく、2次元空間で離散的に画素を選択することにより処理時間の短縮を試みた。以下、2章では手形状認識手法の概略、3章では実時間処理の可能性を示すために行った実験について述べる。

2. 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。一方、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、入力した画素全体ではなく、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手法を提案する。

Proposal of Finger Number Counting Method
using Image Processing for Inputting Hand
Illustrators

Shunsuke AKIFUJI¹, Tamayo TOKUDA²,
Tadashi HIROSE¹, Hiroshi YAJIMA¹

¹HITACHI, Ltd., ²Hitachi Seibu Software, Ltd.

提案手法の概略を図1に示す[2]。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。

次に肌色部分で連続する画素に対してラベル付け処理を施して、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。さらに、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。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。

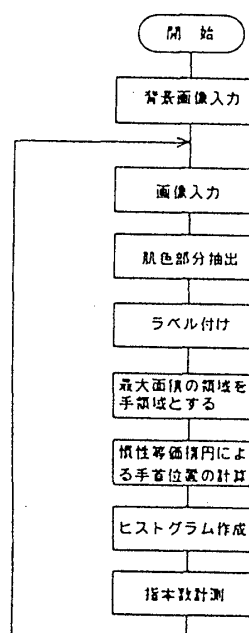


図1 指本数計測のフローチャート

手首を中心として、等角度間隔に分割し、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。楕円の長径を基準に複数のしきい値を設定し、これを越えたヒストグラム上の極大値を数え、指本数とする。

提案手法では、手首からの距離分布の極大値の個数を得られれば良いので、手の周囲の連続した全ての画素を処理対象としなくても、指の間隔以下のサンプリング間隔で選択した画素だけを扱えば指の本

数を計測できる。また、ヒストグラムを求めるための基準点(手首位置)は、楕円で近似した形状から得ているが、この形状も離散的に選択した画素から得た形状とほぼ等しいと考えられ、求める手首位置は大きく違わない。従って、一定間隔で画素を選択し、これらだけを処理しても計測精度の低下は少ないと考えた。一方、処理対象となる画素数は数分の一から数十分の一になるので処理時間は大いに短縮できる。

3. 実験

本実験では、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。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。入力した画像の大きさは横646、縦486である。

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。実験に用いた画像データは、手の面積が撮影範囲の約1/6である、「大」の大きさ「大」と約1/50である「小」の2種類である。

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2に示す。今回の実験で用いた画像に関しては70%程度の正答率を得るためには、対象となる画素の個数が100~200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。

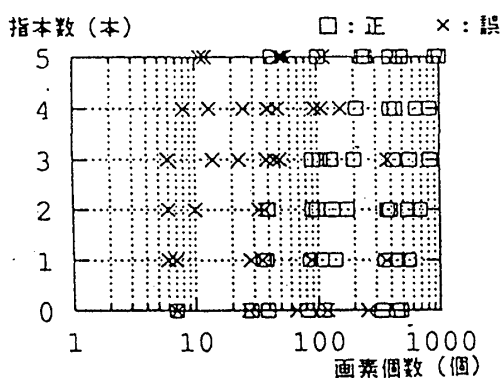


図2 手領域の画素個数と正誤

サンプリング間隔と処理速度を表1に示す。ここで、処理の対象とする画素を100~200にするためには、手画像が「大」の場合では、サンプリング間隔が16とすれば良く、この場合、平均処理回数は約10回/秒となる。手画像が「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、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。処理時間を短縮するためには、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1 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	24	16.32
	16	10.68
	8	3.76
小	16	13.32
	8	6.28
	4	1.75

4. まとめ

- (1) 非言語的な伝達のなかで人間の動作、特に手や腕の意識的動作である例示的動作を入力し、計算機の各種操作インタフェースに利用することを目的として画像処理による動作認識インタフェースを開発している。例示的動作は手や腕を使って示されることが多く、これらの形状と動きを細かく検出する必要がある。この最初の試みとして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提案した。
- (2) 提案手法では、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、手首位置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。
- (3) 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。実験の結果、実時間処理の見通しを得た。

参考文献

- 1) ヴァーガス(石丸 訳)：非言語コミュニケーション，新潮選書(1987)。
- 2) 秋藤，他：画像処理を用いた操作入力インタフェース，情報処理学会研究報告ヒューマンインタフェース 51-3(1993)。