

動き情報の検出とヒューマンインタフェースへの応用

待井 康弘

クンラボン ユーニバン

酒井 善則

東京工業大学工学部

あらまし

本研究ではユーザの身振り・手振りを用いた通信機器のコマンド入力为目标とした、より感覚的なヒューマンインタフェースについて検討する。その一環として、人の上半身の動画像を対象とした手振りの認識を行う。手振りの持つ情報として、位置関係、移動の方向・速さなどの抽出が必要となってくる。これらの手振り情報の抽出を動き領域における色情報に注目して行う。すなわち、手振りにおける手領域を差分画像およびエッジをもとにした領域分割画像より抽出した色情報により判別し、抽出した手領域とシルエット画像より、上半身を剛体の連結体で表現するスティックモデル化を行う。このモデルの動きをパラメータで表現し、あらかじめシステムに用意された辞書を参照することによってユーザが手振りによって示す”意味”の解釈を行う方式を提案する。

MOTION ESTIMATION OF THE HANDS AND IT'S APPLICATION TO HUMANINTERFACE

Yasuhiro Machii

Kulpong Yunibhand

Yoshinori Sakai

Faculty of Engineering, Tokyo Institute of Technology

Abstract

In this paper, we discuss the sensuous human interface for the purpose of using hand gestures as commands. The input image is moving image of the upper part of user's body. To recognize the hand gesture, position, direction and speed of movement of the hands have to be known. In the proposed method, hand color is extracted using differential image and hand is segmented using color information. Segmented hand region is then used for recognition. The upper part of body is described by stick model based on silhouette image and region of arms. We represent movement of the stick model by parameters. Hand gesture is translated by matching these parameters with the command in dictionary.

1 はじめに

人間対機械のコミュニケーションの手段としては、現在のところキーボード操作に代表されるような入力手段が主流となっている。これは文章やプログラムの入力等に対しては合理的な手段と言えるであろうが、人間側にかかる負担という点においては必ずしも優れているとは言えないであろう。上記のような論理的な手段に対して、人間対人間のコミュニケーションに用いられる、手話に代表されるような、身振り・手振りによる意図の伝達手段は感覚的手段とすることができる。このようなヒューマンインタフェースという観点から機械に対するコミュニケーションも、より自然な、そしてより感覚的なものが望ましいと考えられる。そこで我々は身振り・手振りをを用いたインタフェースを対象とするシステムを考える。

身振り・手振りの使用形態は

- 位置指示器：指示動作により、空間上の位置を指定する。
- 程度表現器：回転・伸縮等の程度を表現する。
- イメージ表現器：物の形・概念などを表現する。

の3つに大きく分類される^[1]。この位置指示、程度表現を認識するデバイスとして、データグローブ^[2]等の特殊装置を使用して認識を行う方式もあるが、ユーザの自由度、違和感等を考慮した場合、画像情報のような非接触なセンサを用いることが望ましい。

このような、2次元画像により身振り・手振りを認識する手法として、対象を手などに限定し、エッジ情報をもとに重心を求め、各指の方向を識別する手法^[3]、動画像の場合では、フレームごとの差分をもとに対象物の抽出を行う手法やあらかじめ用意した背景画像と動画像間の差分を用いる手法^[4]、輪郭情報と運動情報、空間的位置情報等よりセグメンテーションを行い、2次元構造を記述する手法^[5]等が考案されている。しかしながら、それぞれの状況下によって用いる手法も様々であり、確立された方法はまだないと言ってよい。

身振り・手振りを認識する際の問題点として、カメラ一台の単眼視による入力はステレオ視による入力等に比べて処理が容易であるが、手の領域の分離が困難である。また、手にマーク等をつけてマークの色で認識する場合などにおいても、あらかじめマークの色情報を与える必要がある。そして、抽出した手の情報をどのように解釈するか。等が

あげられる。本論文では、これらの解決策として、手振りにおいて移動量が最も多い領域が手の部分であると考え、この動き部分の抽出を行う。この抽出された動き部分の領域における色情報に基づき、入力動画像の各フレームの腕部分の領域抽出を行う。そして、これらの腕領域を中心に、モデル化^[6]を行い、モデルの動きによって手振りを認識するスティックフローモデルを提案する。

システムの構成を図1に示す。動画像から対象物の動き検出と色情報の抽出^[7]について2章で詳しく述べ、次に人体のシルエット画像、色情報をもとにしたモデル化の手法および、パラメータの記述について3章で述べる。最後に実験結果を示す。

2 手の動き検出

2.1 動き情報による領域抽出

人間が手振り動作を行うとき、動きの大きい部分は手の部分であり、有する情報量の多い部分もまた手の部分であると考えられる。また、人間は対象を認識する際に視覚情報として、エッジ、差分等の要素の総括的判別のもとに認識を行っているが、その中で色によって対象物を認識・判断する場合も少なくない。そこで、動きの大きい手の部分の色に注目し、抽出した手の部分の色情報を基に動画像中の各フレームにおいて手の領域を抽出する方式を提案する。しかし、長袖を着ている場合、腕まくりをしている場合など様々な場合が考えられ、動き部分における色情報といっても何種類もの色情報が抽出される場合もありうる。このような場合には、手の部分を色別に領域に分割し、動き量の大ききで識別することによって適応できるものと思われる。今回は半袖を着用している場合を対象として、認識・

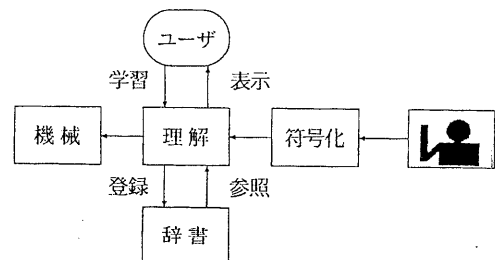


図1: システムの構成

モデル化およびパラメータ表現までの検討結果を中心に報告する。

2.2 差分画像の生成

動画内の2枚のフレーム間において、画素値の差分の絶対値を取ると、2枚のフレーム間で画素値が変化している部分、すなわち画像の中で動きのある部分が抽出される。これにより手振り動作における手の部分の色情報の抽出を行う。

あらかじめ背景は白色としているため、差分画像の中で、動きの大きい部分ほど顕著に現れてくるはずである。しかし、逆にエッジのところなど、他の部分のわずかな動きも差分として出てきてしまったり、照明条件やカメラの特性などにより、ノイズがかなり含まれてしまう。そこでノイズ除去の対策として、閾値をもうけ、差分画像からある大きさ以下の移動領域を除去する。

実際の差分画像の作成には、以下の式(1)を用いる。

$$d_{ij}(x,y) = \begin{cases} 1 & \text{if } |f_i(x,y) - f_j(x,y)| > Th \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

ただし、入力画像1フレームは、720×480画素、RGB24ビットであり、 $f_i(x,y)$ は第*i*フレームの(*x*,*y*)座標における輝度値を表すものとし、*Th*は閾値を $d_{ij}(x,y)$ は作成される差分画像の値を示す。

2.3 動き部分における色情報の抽出

まず、対象とする動画から抽出した差分画像より、動き部分画像を作成する。3枚の入力輝度画像を f_i, f_j, f_k とする。動き部分は差分画像 d_{ij}, d_{jk} の共通部分、すなわち論理積部分をもって表わす。式を以下に示す。

$$m_j(x,y) = d_{ij}(x,y) \cdot d_{jk}(x,y) \quad (2)$$

$m_j(x,y)$ は f_j における動き部分画像を表わす。

しかし、頭の部分などのブレなどが存在するため、エッジ部分等においてノイズが残ってしまう。そこで、ノイズを取り除く作業を行う。このノイズ除去の手法としては、対象画像のエッジをもとに領域分割を行い、動き部分の画素の多い領域を求める。動きの部分を領域として捕らえることにより領域の動きとして認識することができるため、エッジ近辺における微妙な動きから生じたノイズの除去を行うことができ、手振りにおける”手”の部分の抽出がより安定して行うことが可能となる。

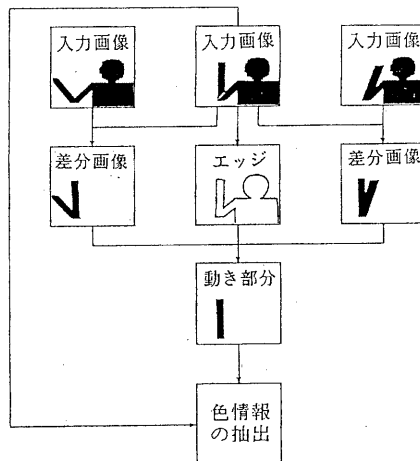


図2: 色情報の抽出

次に、作成した動き領域より色情報の抽出を行う。動き部分と認識された領域において入力画像よりRGB成分の

平均値 R_m, G_m, B_m

標準偏差 $\sigma_R, \sigma_G, \sigma_B$

を求める。これを手の領域の色情報として、各フレームの手部分の抽出を行い、モデル化する際にこれを基に行うものとする。ここまでの流れを図2に示す。

3 スティックフローモデル

人間の体を剛体の連結体とみなし、各剛体部をスティック、剛体の連結部をノードで近似し、動作をこれらのモデルの動きベクトルで表現し、パラメータ記述を行う。しかし、スティックフローモデルで表現する際、ノードの位置を原画像から検出し、各フレームにおいてノードの流れを追跡する必要がある。そこでノードを動きのあるノードとないノードに分別し、動きのあるノードは動き部分領域の色情報をもとに検出し、動きのないノードに関しては、初期フレームのシルエット画像より検出するものとする。上半身を対象としているため、ノードの割り当てられる部位として、頭部、首部、肩部、肘部、手部の5つに限定を行うこととする。モデル化の例を図3に示す。

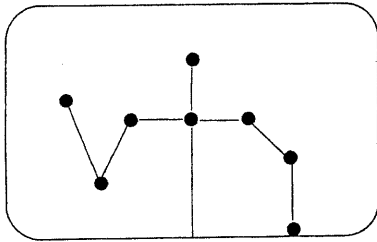


図 3: モデル化の例

3.1 シルエット画像によるノード検出

今回は手振り動作に限定しているため、頭部、首部、肩部のノードを動きのない部分として初期フレームのシルエット画像より検出するものとする。

アルゴリズムの概略を以下に示す。

1. シルエット画像の作成

エッジ画像より背景、ノイズを除去し、対象を切り出し、2値化を行ったものをシルエット画像とする。

2. 体の中心線の抽出

まず作成したシルエット画像の胴の幅を検出する。これは入力を上半身画像と限定しているため、胴の部分は画像の下部の途切れている部分の幅で表わし、その中心をもって体の中心線とする。

3. 頭部、肩部のノード検出

頭部のノードは中心線上の頭頂部に割り当てるものとする。

肩部のノードの検出は、シルエット画像において x-方向と y-方向における画素数のヒストグラムを作成し、あらかじめ推定しておいた領域より、シルエット画像より肩部のノードを決定する。

4. 首部のノード検出

首部のノードは体の中心線と左右肩部ノードを結んでできる線との交点をもって表わす。

3.2 色情報によるノード検出

動きの大きい部分においては、色情報を用いたノード検出を行う。2.3章で求めた動き部分領域である手の部分の色情報を用いて、各フレームにおける手の領域を以下のように抽出する。

に抽出する。

```

if
     $R_m - 2\sigma_R < f_R(x, y) < R_m + 2\sigma_R$ 
    and  $G_m - 2\sigma_G < f_G(x, y) < G_m + 2\sigma_G$ 
    and  $B_m - 2\sigma_B < f_B(x, y) < B_m + 2\sigma_B$ 
then  $f(x, y)$  は手領域候補の画素である。
else  $f(x, y)$  はその他である。
    
```

ここで $f_R(x, y)$, $f_G(x, y)$, $f_B(x, y)$ は $f(x, y)$ における RGB 成分の値である。以上のようにして求められた手領域候補の画素より、各フレームにおける手部分を判別し、肘部、手部のノードを検出する。以下にアルゴリズムを述べる。

1. 領域分割

対象としているフレームに対して、エッジをもとにした領域分割⁸⁾を行う。

2. 手領域候補の抽出

分割した i 番目の領域内の画素数を N_i 、その領域内における手領域候補の画素数を S_i とすると、

$$S_i / N_i > Th$$

となる領域を手の領域候補としてあげる。

3. ノイズ除去および領域の統合

候補としてあげた領域の内、細かな線などで分断されている領域に対し、膨張収縮を施し、領域の統合を行う。また、面積がある閾値以下である領域についてはノイズと判断し除去する。

4. 手部分の判別

手候補としてあげられた領域に対して重心を求め、頭、両肩のノードより成る三角形の範囲に重心が含まれる領域は顔の領域と判断し、他の面積の大きい領域に対して1つの場合に対しては片手、2つの場合に対しては両手として判断する。

5. 手の軸及び向き判別

手の重心を通る任意の直線を引く。この直線と手の領域の境界との交点で、しかも重心の両側にある2点を手、肘のノードの候補とする。この直線をすべての方向に回転させ、候補となるノード間の距離が最も長くなる方向を手の軸と判断する。また、手と肘

の判別だが手の幅が極小値をとるところを手首と判断し、手首に近い方を手、遠い方を肘のノードとして割り当てる。

3.3 パラメータ記述

手振り動作を表現する情報としては大きく分けて

- 手の位置情報
- 動きの方向及び大きさ

の2つがあげられる。これらの情報をスティックフローモデルの動きによって記述する。手振り・身振り動作は個人差があり、曖昧性を有する言語であると言える。このためシステムで使用する辞書は個人別に用意し、認識結果を辞書で参照し、動作の意味を判断する。今回は手振り動作を対象としているため、手の位置情報は体の中心線を座標軸とした相対的位置関係を用い、また、動きの大きさなどは動作の速さなどに対応できる記述方法が望まれる。一つの動作でも、各部の動きの総合として成り立っている。このことを考慮し、また、上半身を対象としているため頭部、左右手部の動きの総括として動作を表現する方式を提案する。そのため、図4のようなTree構造を用いて記述する。この記述は2フレーム間の動きに対するものであり、一連の動作ではフレーム数分の記述を用いて各ノードの位置関係、移動の方向、大きさを求める。図4にある coordinate, motion

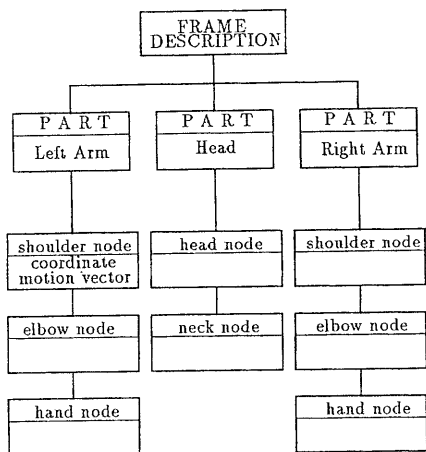


図4: パラメータのフレーム記述

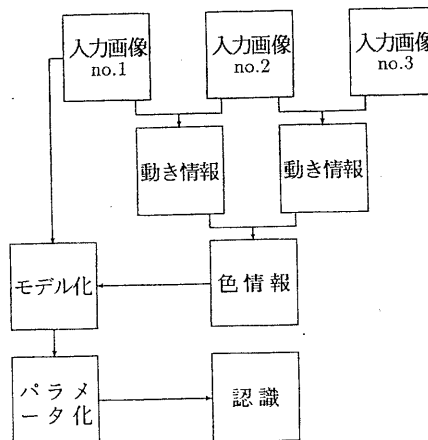


図5: 認識までの流れ

vector は、各ノードにおける座標、動きベクトルの情報を表わす。このようなフレーム間記述を用いることによって、フレーム間記述の数を加減することにより動きの速さに依存しない認識が可能となると思われる。この認識までの流れを図5に示す。

4 実験結果

本システムにおける手の認識の結果の例を図6-9に示す。入力ビデオカメラによってフレームバッファに取り込まれた手振りの画像とする。図6は入力画像の中の1フレームを表わす。前後のフレームとの差分を求め、論理積をとったものが図7の動き部分画像で、これからノイズを取り去り、動きの大きい領域における色情報をもとに、動き部分と同じ色の領域を求める。次にシルエット画像より頭、肩のノードを抽出した画像が図8である。今回は手の部分を肌色と限定をしているため、図9に示すように手と顔の部分が抽出されていることがわかる。これらをもとに抽出された部分の重心とノードの位置関係より手部分の判別を行い、手部分のモデル化を行う。

5 むすび

本論文では、身振り・手振りによるヒューマンインタフェースのための、手振り認識の一方式の提案・検討を行ったものである。提案方式は、手振りの動き情報より手部分の色情報を抽出し、これをもとに各フレームにおける手部分の認識を行い、モデル化を行うことによってモデルの動きにより、手振り動作の認識を行うものである。このモデルは身体を剛体の連結体とし、各ノードの動きベクトルによってパラメータ化するため、スティックフローモデルと名付けた。また、連続するフレームのパラメータ記述を用いることによって動作の速さによらない柔軟な認識が可能である。

しかし、スティックフローモデルでは、2次元的な動きしか表現できないという制限もあり、また手部分の色に制約を付すなどの条件があるため、3次元情報を用いたモデル化などを行い、より制約の少ない認識法が必要であろう。また、システムに柔軟性をもたせるために手振りの曖昧性を吸収できるような、学習機能を含めた認識体系の考案などが今後の課題であろう。

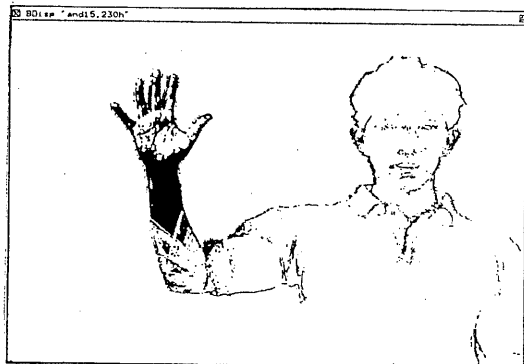


図 7: 動き部分画像



図 6: 入力画像

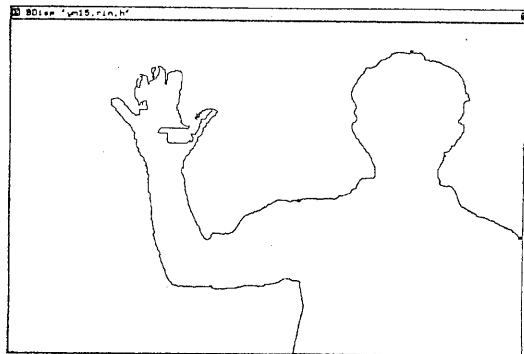


図 8: 頭部・肩部のノード

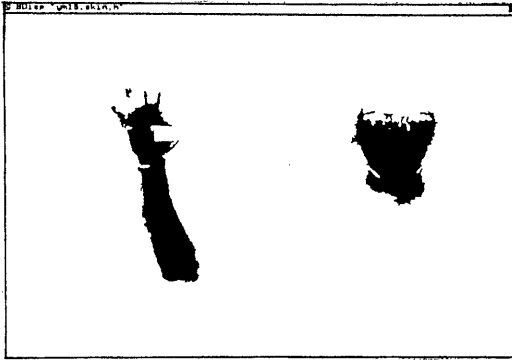


図 9: 顔および腕部分

参考文献

- [1] 福本雅朗, 間瀬健二, 末永康仁: "Finger-Pointer:画像処理を用いた仮想指示棒", 信学技報, HC91-12(1991)
- [2] 高橋友一, 岸野文郎: "手振り認識方法とその応用", 信学論, J73-D-II, 12, pp.1985-1992(1990)
- [3] 大西剛, 竹村治雄, 岸野文郎: "手形状記述の一提案", 信学春全大 A-250(1991)
- [4] 催享振, 安居院猛, 中嶋正之, 横前高広: "動画画像処理による通行車両とそのナンバープレート領域の抽出", 信学論, J73-D-II, 3, pp.392-398(1990)
- [5] 森川博之, 原島博: "逐次的領域分割に基づく動画画像の構造的記述", 1991年画像符号化シンポジウム(PCSJ91), 6-7, pp.137-140
- [6] 木本伊彦, 安田靖彦: "知的動画画像符号化のための動物体のフレーム表現の一方式", 信学論, J72-D-II, 9, pp.1356-1363(1989)
- [7] クラボン ユーニバン, 木下宏揚, 酒井善則: "学習機能をもった視覚言語処理システムに関する基礎検討", 信学技報, HC90-21(1990)
- [8] 横矢, 浅野, 大久保, 田中: "濃淡画像からの対象物の切り出し", 情報処理, Vol.19, 8, pp. 730-737(1978)