差分画像を利用した人体の動作認識

亀田 能成　　美濃 導彦　　池田 克夫

京都大学工学部

〒 606-01 京都大学工学部情報工学科
e-mail:kameda@kuis.kyoto-u.ac.jp

本稿では、差分画像を利用した動画像からの人体の動作認識方法について述べる。差分画像は動画像から抽出できる簡便な特徴量である。本手法では人体を関節物体としてモデル化し、その動作を関節物体の角度の変化として捉える。すなわち、モデルマッチング手法を基本とし、差分画像とモデルの画像平面への投影像との関係から、モデルの各部分の状態を運動・静止・隠蔽の 3 状態に分類する。その状態に応じて関節角度の制御を行い、人体の動作を認識する。応用としては非接触式のマンマシンインターフェースなどが挙げられる。

動画像, 差分画像, 動作認識, モデルマッチング, 人体

A Human Motion Recognition Method Using Difference Images

KAMEDA Yoshinari　MINOH Michihiko　IKEDA Katsuo

Faculty of Engineering, Kyoto University

Department of Information Science, Faculty of Engineering,
Kyoto University, 606-01, Japan
e-mail:kameda@kuis.kyoto-u.ac.jp

We propose a human motion recognition method using a difference image sequence. A difference image is useful because it is easy to extract it from a image sequence. Our method is based on a model matching method. The human body model consists of solid parts. We classify the parts into three modes; moving, stationary, and occlusion. A part is labeled with one of three modes by considering the relationship between the difference image and the projected region of that part on the image plane. We calculate joint angles of a part according to its mode at each frame.

image sequence, difference image, motion recognition, model matching, human body

1　はじめに

実世界の三次元物体の中でも、関節物体のように変形しうるものに対する姿勢推定・動作推定は画像認識の分野において重要な研究である。特に、人体は関節の数も多くかつその回転範囲が広いので認識が難しいとされるが、マンマシンインターフェースの観点から社会的には認識要求が高い。

人間などを対象にその動作全体の認識を行う研究は行われている [3, 5, 6] が、動作の分類が主体であり、対象となる人体の空間的な動作を直接計測しているわけではない。関節物体そのものの姿勢推定を行う研究については、VR へのインターフェースを目指す研究 [4] がある。これは単眼視カメラを前提に三次元的な指の動きを検出するものであるが、実時間処理を第一義にしているため、特徴点の抽出そのものはカラーマーカーを利用し、画像取得の際に制約を課している。

そこで、本稿では、動画像から差分画像を生成し、それを利用して人体の動作認識を空間的に行う方法について提案する。差分画像を利用すれば、画像内の前景と背景とを区別する処理は不要である。また、本方法は、人体モデルの各体節ごとにモデルマッチング処理を行う。その際、体節ごとに運動・静止・隠蔽の 3 状態を定め、その状態に応じて、差分画像を利用する。画像情報が利用できない隠蔽状態については、人体の動作に慣性が働くとの制約を付加して対応する。

なお、動画像中には対象となる人体以外に運動する物体はないものと仮定する。また、カメラパラメータは既知とする。

以下、2章では人体のモデル化について述べ、3章では人体の動作に関して本研究で設けた慣性の制約について述べる。4章では動作における人体モデルの各部分の状態について考察し、差分画像との関係を述べる。そのあと、本稿で提案する方法を 5章で説明し、6章で本方法を用いた実験を示す。最後に、7章で結論を述べる。

2　人体のモデル化

本稿では、人体を関節物体モデルで表現する。ここで、関節物体とは、関節角度によって、かつそれによってのみその姿勢が一意に決定される物体をいう。人体モデルにおいては、以下のような制約を設ける。

- 一つの関節は最大三軸までの関節回転軸の組み合わせで実現される。
- 人体モデルを構成する体節は関節によって互いに接続される。
- 各体節は剛体である。
- 各体節は木構造状に連結されている。

人体モデルは、木構造のデータ構造で表現する。このデータ構造を姿勢決定木と呼ぶ。姿勢決定木の各ノードは人体モデルの一つの体節に相当し、その体節の幾何形状と、親ノードとの間の関節に関する情報を持つ。関節角度 $a_{ik}(t)$ は時刻 t におけるノード i の軸 k まわりの関節角度を表す。

なお、姿勢決定木中の根ノード r において、関節接続点は三次元世界座標系内での根ノードの位置を表す。また、その関節角度 $\mathbf{a}_r(t) = \{a_{rk}(t) | 0 \leq k < n_r\}$ は三次元世界座標系に対する回転を表す。

3　慣性を考慮した動作記述

本研究では、人体の動作に慣性が働くと仮定する。正確には、人体モデルを構成する各ノードにおいて、外乱要因がなければ回転の角速度は保存されると仮定する。これは自然な仮定である。実際は常に角速度が保存された関節の動きをしているわけではないので、角速度保存則に加えて、角速度の時間的変化を見込まなければならない。

いま、ある時刻 t において、人体モデルの一つのノードの運動を考える。このノードが表す体節の時間 Δt 内における運動は、等角速度運動と外乱要素の合成と考えられ、時間 Δt 後に予測される関節角度の範囲は以下の式で拘束できる [1]。

$$|a_{ik}(t + \Delta t) - (a_{ik}(t) + v_{ik}(t)\Delta t)| \leq \Delta R \quad (1)$$

ただし、$a_{ik}(t)$ は時刻 t における関節 i の軸 k

の角度、$v_{ik}(t)$ はその角速度を意味する。ΔR は、運動の種類によって異なる値をとるが、人体の場合、その上限はほぼ一定であると観察される。

4 各ノードの状態と差分画像

4.1 ノードの状態

人体の運動をカメラから観察することを考える。このとき、人体内部の体節を見かけ上、運動・静止・隠蔽の 3 種類の状態に分類する。ただし、注目する体節自身の関節が回転運動していなくとも、その体節が他の体節の関節運動によって移動する場合は、運動状態であるとする。また、隠蔽状態とは、注目する体節が他の体節によって隠蔽されている状態であり、実際に運動しているか静止しているかを問わない。

これに対応して、人体モデルの各ノードを排他的に 3 状態に分類する。各状態において関節角度を求める際、次のようなことを考慮しなくてはならない。

1. 運動状態のノード
   カメラから観察される。関節角度の推定に差分画像情報を利用できる。
2. 静止状態のノード
   カメラから観察される。関節角度の推定に画像情報を利用できるが、対応する画像平面上でフレーム間差分は生じない。
3. 隠蔽状態のノード
   カメラから観察されない。関節角度の推定に画像情報を利用できないので、3章の角速度保存則の制約を用いる。

ここで、時刻 t における運動状態のノードの集合を $\mathcal{M}_t$、静止状態のノードの集合を $\mathcal{S}_t$、隠蔽状態のノードの集合を $\mathcal{O}_t$ とする。

4.2 差分画像

動画像から抽出しうる特徴量として、連続する二枚の画像からの画素間の差分を用いる。差分値は、簡単に計算でき、動作認識に用いる場合、前景と背景を区別せずともよいという長所をもつ。そこで、本研究ではこの差分値について考察し、その利用法を提案する。

いま、時間間隔 Δt で撮影された時刻 t_{n-1} と t_n の画像を $\mathbf{I}_{t_{n-1}}$, $\mathbf{I}_{t_n}$ とする $(t_{n-1}+\Delta t = t_n)$。画像 $\mathbf{I}_{t_n}$ 中の画素 $\mathbf{x} \in \mathbf{I}_{t_n}$ について、その差分値 $s(\mathbf{I}_{t_n}, \mathbf{x})$ を式 (2) で定義する。この差分値に応じて、画素 $\mathbf{x} \in \mathbf{I}_{t_n}$ は二種類に分類される。ただし、次式において $p(\mathbf{I}_t, \mathbf{x})$ は画像 $\mathbf{I}_t$ の位置 $\mathbf{x}$ における画素値を表す。

$$s(\mathbf{I}_{t_n}, \mathbf{x}) = |p(\mathbf{I}_{t_n}, \mathbf{x}) - p(\mathbf{I}_{t_{n-1}}, \mathbf{x})| \qquad (2)$$

- 静止領域 $\mathbf{S}_{t_n} = \{\mathbf{x} | s(\mathbf{I}_{t_n}, \mathbf{x}) = 0\}$
  これは、画素値 $p(\mathbf{I}_{t_n}, \mathbf{x})$ が一定であることを意味している。これを満たす条件は二つある。
  - 背景の静止物体か、ないしは静止状態にある人体の体節が投影されているとき。
  - 時刻 t_{n-1} ないし t_n もしくはその両方において運動状態にある人体の体節が投影され、かつ差分値が 0 であるとき。

  人体の表面の色・形状の多様性及びその運動から考えると、一般的に後者のような条件は非常に稀であり、考慮から外してもよいものと考えられる。従って、$\mathbf{S}_{t_n}$ は静止状態の体節 $i \in \mathcal{S}_{t_n}$ が投影されてできた領域と見なせる。
- 運動領域 $\mathbf{M}_{t_n} = \{\mathbf{x} | s(\mathbf{I}_{t_n}, \mathbf{x}) \neq 0\}$
  運動領域は、以下に定義するような生成動領域 $\mathbf{G}_{t_n}$、継続動領域 $\mathbf{K}_{t_n}$、あるいは消滅動領域 $\mathbf{V}_{t_n}$ の和集合である。ただし、$\mathbf{P}(\mathcal{A})$ はノード集合 $\mathcal{A}$ に含まれる全てのノードの投影領域の和集合とする。

$$\mathbf{G}_{t_n} = \mathbf{P}(\mathcal{S}_{t_{n-1}}) \cap \mathbf{P}(\mathcal{M}_{t_n}) \qquad (3)$$
$$\mathbf{K}_{t_n} = \mathbf{P}(\mathcal{M}_{t_{n-1}}) \cap \mathbf{P}(\mathcal{M}_{t_n}) \qquad (4)$$
$$\mathbf{V}_{t_n} = \mathbf{P}(\mathcal{M}_{t_{n-1}}) \cap \mathbf{P}(\mathcal{S}_{t_n}) \qquad (5)$$

時刻 t_n の差分画像の中で、時刻 t_n で顕著に変化するのは $\mathbf{G}_{t_n}, \mathbf{V}_{t_n}$ である。従って、これらの領域を抽出できれば有用であると考えられる。

5 マッチング処理方法

本章では、これまで述べてきた人体モデルのノードの 3 状態や差分画像に関する考察を組み合わせて、人体動作認識を行うモデルマッチング処理について説明する。

5.1 処理の概要

本方法は、人体モデル中のノード毎に処理を行う。一枚の差分画像についてノード毎にマッチング処理を行い、その時刻での関節角度を求める。さらに、フレーム間で関節角速度保存則を考慮して、次フレームにおける処理の前に関節角度の予測を行い、隠蔽状態のノードに対応する。全体の処理の流れを図 1 で示す。

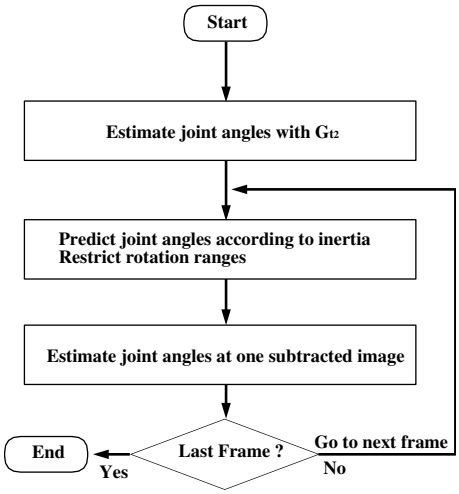


図 1　処理の流れ

なお、初期処理として、第 1 フレームのシルエット画像に対して姿勢が与えられたものとする。この姿勢については人手で合わせるか、別の姿勢推定法 [2] を利用することが考えられる。また、この時の関節角度の可動範囲は、その姿勢から $\pm\Delta R$ の範囲を指定する。角速度については、 $v_{ik}(1) = 0.0 \; \forall \; i, k$ とする。

時刻 t_n における差分画像と人体モデルの姿勢 $\{\mathbf{a}(t_n)\}$ との関係が $\mathbf{H}(\{\mathbf{a}(t_n)\}) \equiv \mathbf{K}_{t_n} \cup \mathbf{G}_{t_n}$ を満たせば、画像平面上で正しくマッチングされたことになる。ただし、$\mathbf{H}(\{\mathbf{a}(t_n)\})$ は姿勢 $\{\mathbf{a}(t_n)\}$ を画像平面に投影してできる領域である。一般に、$\mathbf{I}_{t_{n-1}}, \mathbf{I}_{t_n}$ から式 (2) に基づいて生成された差分画像 $\mathbf{D}_{t_n}$ では、$\mathbf{K}_{t_n} \cup \mathbf{G}_{t_n}$ が求められない。しかし、第 2 フレームの処理の場合だけは仮定から $\mathbf{H}(\{\mathbf{a}(t_1)\}) = \mathbf{K}_{t_2} \cup \mathbf{V}_{t_2}$ と見なせる。これから $\mathbf{G}_{t_2} = \mathbf{M}_{t_2} \cap \overline{\mathbf{H}(\{\mathbf{a}(t_1)\})}$ が得られるので、式 (6) の最小化を考えることにする。ただし、$f(\mathbf{A})$ は領域 $\mathbf{A}$ の面積を求める関数であり、$\oplus$ は二つの領域に対する排他的論理和演算子である。

$$E\left[\{\mathbf{a}(t_2)\}\right] = f\left(\left(\mathbf{G}_{t_2} \oplus \mathbf{H}(\{\mathbf{a}(t_2)\})\right) \cap \overline{\mathbf{H}(\{\mathbf{a}(t_1)\})}\right) \tag{6}$$

一方、第 n フレーム $(n \geq 3)$ の処理のときは、マッチング処理の進行につれて $\mathbf{H}(\{\mathbf{a}(t_{n-1})\})$ の精度の劣化が不可避であり、$\mathbf{H}(\{\mathbf{a}(t_{n-1})\}) = \mathbf{K}_{t_n} \cup \mathbf{V}_{t_n}$ が仮定できなくなる。その代わり、処理の進行によって関節速度の予測が可能になるので、$\mathbf{H}(\{\mathbf{a}(t_n)\}) \in \mathbf{M}_{t_n}$ を満たせばマッチングが正しく行われているものとし、式 (7) の最小化を考える。

$$E\left[\{\mathbf{a}(t_n)\}\right] = f\left(\mathbf{M}_{t_2} \oplus \mathbf{H}(\{\mathbf{a}(t_n)\})\right) \tag{7}$$

以上の手順をまとめると、図 2のようになる。

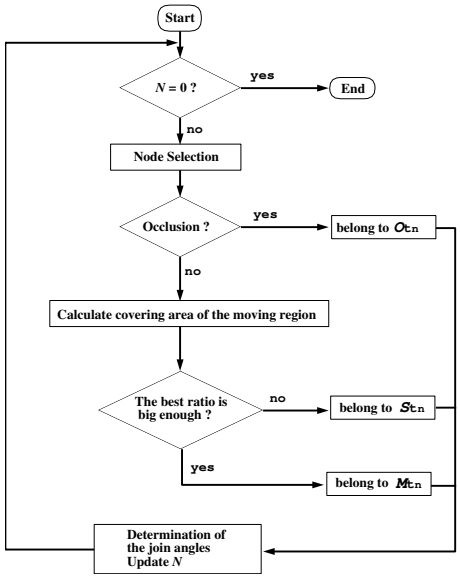


図 2　フレーム上での処理

まず、処理開始時は、未処理ノード数 N は人体モデルのノード数に等しい。この中から姿勢決定木の根ノードに近い未処理のノード p を取りだし、既に関節角度決定済のノードの影に隠れているかどうかを人体モデルで調べる。もし隠蔽されていれば、そのノードを隠蔽状態と判断し、画像情報が利用できないのでその関節角度を関節角速度の分だけ移動した予測値に設定する。

隠蔽されてないノードについては、そのノードの関節可動範囲を適当なサンプリング間隔 s でサンプリングし、それぞれについて式 (6) ないし (7) を最小化するようなマッチング評価関数 $F[\mathbf{a}_p(t_n)]$ を計算する。その中で最も値の大きいサンプルについて、運動領域識別値 μ と比較し、それより小さく、かつそのノード p から根ノードに至る姿勢決定木上の経路に運動状態のノードが存在しない場合、そのノードは静止状態にあると判断する。静止状態のノードについては、許された関節可動範囲のなかで最も前フレームでの関節角度に近い関節角度を採用し、関節角速度を零にする。

一方、それ以外の場合は、ノード p を運動状態であると判断する。運動状態のノードについては、最大値を与えたサンプル点をその関節角度とする。

このようにそれぞれの状態に応じて関節角度を決定したのち、 N の値を一つ減らし、処理の最初に戻る。

以後、 5.2節と 5.3節でマッチング評価関数 $F[\mathbf{a}_p(t_n)]$ について述べる。

5.2 第 2 フレーム

式 (6) を最小にする姿勢 $\{\mathbf{a}(t_2)\}$ を解析的に求めることは、 $\{\mathbf{a}(t_2)\}$ の変化に対する $\mathbf{H}(\{\mathbf{a}(t_2)\})$ の変化の予測が困難であるため、難しい問題である。そこで、この最小化問題をノード毎の問題に変換するため、次の式を考える。運動状態にあるノード p において、式 (8) はこれを最小化することでノード p の関節角度を求める問題となる。ここで、 $\mathbf{T}(\mathbf{a}_p(t_2))$ はノード p の画像平面上への投影領域を表す。式 (8) を図式化したのが図 3である。

$$F\left[\mathbf{a}_p(t_2)\right] = f\left(\left(\mathbf{G}_{t_2} \oplus \mathbf{T}(\mathbf{a}_p(t_2))\right) \cap \overline{\left(\mathbf{H}(\{\mathbf{a}(t_1)\}) \cup \mathbf{C}_{t_2}\right)}\right) \qquad (8)$$

5.3 第 n フレーム $(n \geq 3)$

式 (7) の問題設定の場合、前フレームの人体モデルの姿勢が正しいときには、本フレームで全く静止していても、正しく動作認識したのと同一に評価されるが、関節角速度が設定されているのでそのような局所解には陥らないと考えられる。

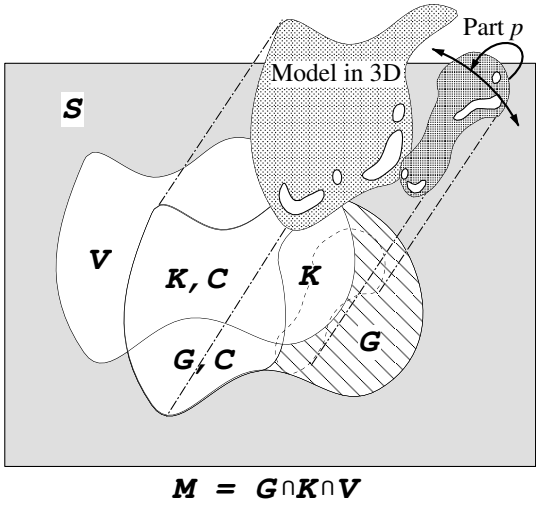


図 3　ノードのマッチング処理

これを前節同様にノードごとの最小化問題に変換することを考えると、式 (9) のようになる。

$$F\left[\mathbf{a}_p(t_n)\right] = f\left(\left(\mathbf{M}_{t_n} \oplus \mathbf{T}(\mathbf{a}_p(t_n))\right) \cap \overline{\mathbf{C}_{t_n}}\right) \qquad (9)$$

6 実験

実験として、男性の 40 フレームの歩行運動に適用した際の実験結果を図 4に示す。画像の大きさは横 600 画素、縦 360 画素であり、 $\Delta R = 6.0$ 度、サンプリング間隔 $s = 2.0$ 度、運動領域識別値 $\mu = 2,500$ 画素とした。結果から、体節どうしが重なりあう場合にも、本方法による動作認識が成功していることがわかる。

本方法では、人体が複雑な動作を行った場合にはまだ対応できない。この原因のひとつは、一旦角速度が付き過ぎるとその角速度の変更がされる前に見当違いの運動領域とマッチングされだしてしまい、復帰できなくなることがあげられる。

7 おわりに

本稿では、動画像から差分画像を生成し、それを利用して人体の動作をモデルマッチング手

法により認識する方法について提案し、実験を行った。本方法では、人体モデルの各体節ごとに運動・静止・隠蔽の３状態を定め、それによって差分画像情報を利用するかどうかを判定している。

今後は、人体の複雑な動作にも対応していく予定である。

参考文献

[1] 亀田能成, 美濃導彦, 池田克夫. モデルを用いた人体の動作推定法 — 慣性の導入 —. 信学全大, Vol. spring, No. D-658, 1995.

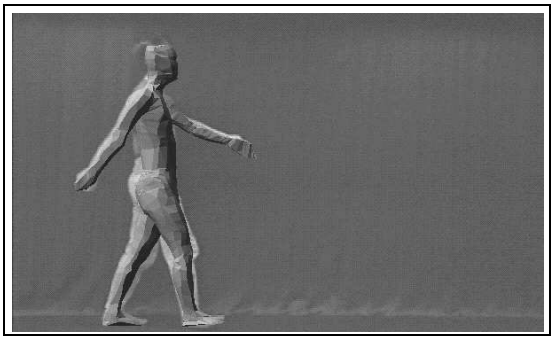
[2] 亀田能成, 美濃導彦, 池田克夫. シルエット画像からの関節物体の姿勢推定法. 信学論, Vol. D-II, No. 1, pp. 26–35, 1996.

[3] 高橋勝彦, 関進, 小島浩, 岡隆一. ジェスチャー動画像のスポッティング認識. 信学論, Vol. J77-D-II, No. 8, pp. 1552–1561, August 1994.

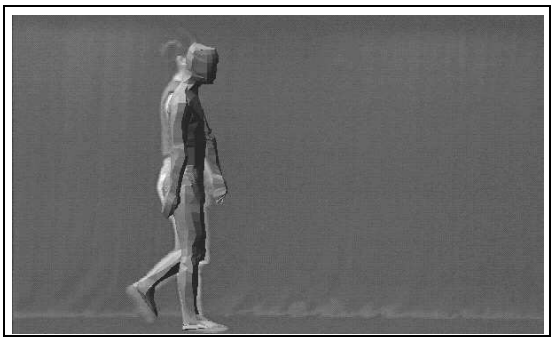
[4] 中嶋正之, 柴広有. 仮想現実世界構築のための指の動きの検出法. 信学論, Vol. J77-D-II, No. 8, pp. 1562–1570, August 1994.

[5] 牛田博英, 山口亨, 高木友博. ファジー連想記憶システムを用いた動作認識. 信学論, Vol. J77-D-II, No. 8, pp. 1571–1581, August 1994.

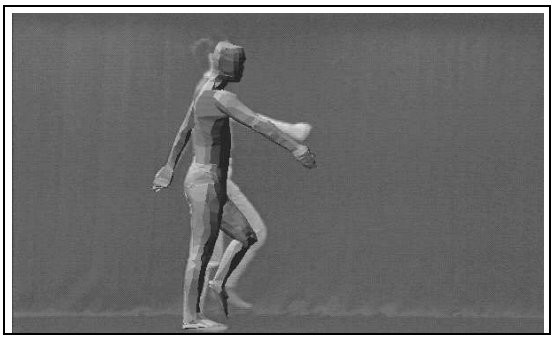
[6] 安富敏, 森英雄, 清弘智昭. 歩行のリズムに基づく歩行者検出の一手法. 信学論, Vol. J78-D-II, No. 4, pp. 608–617, April 1995.



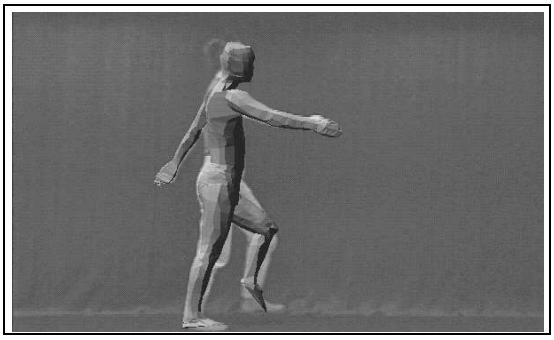
第 1 フレーム



第 13 フレーム



第 27 フレーム



第 32 フレーム

図 4　人体の動作認識結果