

2. 認識技術の現状

2-1 身振りの自動認識

山本正信[†]

1. ま え が き

人間の動作を定量的に捉えることができれば、CGアニメーションやテレビゲームにおけるキャラクタの動作の生成、ロボットマニピュレータの軌道の生成、人間と機械との円滑なコミュニケーション、テレビ会議のための画像の高速伝送、安全のためのセキュリティモニタリング等、数多くの分野での活用が期待される。そのため、動作の定量化に向けて様々な研究が活発に行われてきている。

人間の動作を捉える装置として、データスーツやデータグローブなどが利用可能である。これらの装置は、人体に特殊な器具を取り付けることが多い。そのため、人間の自由な動作を捉えることが難しい。また、セキュリティ関連では、事前に器具を取り付けること自体意味をなさない場合がある。

テレビカメラからの映像を利用すれば、人体に接触することなく動作の測定が可能である。映像記録さえ残っていれば、過去の動作を解析し再現することも可能である。本稿ではテレビ映像等の動画像を使って、身振りを測定し認識を行う手法の現状を紹介する。

2. 身振りの動画像計測

人体は木構造を持つ多関節モデルで表されることが多い。このとき、人体の姿勢は各関節角で表され、身振りは関節角速度で表される。これを身振りの運動パラメータと呼んでいる。

運動パラメータは、このモデルの投影像が人体像にフレーム毎に一致するよう決定される⁵⁾⁷⁾⁹⁾¹²⁾¹⁴⁾。パラメータは解析的に得られることもあるが³⁾、一般には探索操作を必要とする。この探索の範囲は多関節構造

からの拘束を利用すれば、かなり狭めることができる。また、Best first search²⁾やビーム探索法¹⁴⁾、GA¹⁰⁾などのような効率的な探索手法の適用も試みられている。

一方、コストのかかるモデルと人体像の照合は追跡開始フレームだけで行い、変位分を時空間勾配法¹⁶⁾や対応点追跡³⁾により推定し、変位分の累積により追跡を行う方法も提案されている。図1はインラインスケートの追跡結果である。また、図2は、そのときのひざの屈伸角速度を示す。

人体のモデルと人体像との照合には、計算コスト以外にも、次のような問題点が含まれている。(1)モデルと人体像が一致したとしても、人体の姿勢が正確に測定できているとは限らない。人体の形状を正確にモデル化することは難しく、またたとえ正確にモデル化したとしても、人体形状はたえず変化する。したがって、近似モデルとの画像上での照合であるため、3次元的に正確に照合できているという保証はない。(2)手や足が自身の胴体等によって隠される場合がある。このようなオクルージョンが発生したときに、隠された部分の動きを知ることが難しい。(3)動きが投影面上に現れにくい場合がある。例えば、カメラ光軸方向の動きは見かけの動きが小さいため、そこから3次元的な動きを復元するのは難しい。

これらの問題点を解決(あるいは回避)するために、異なる位置に置かれた複数のカメラの利用が提案されている。多視点動画像の使用により、照合の曖昧さやオクルージョン問題²⁾¹¹⁾¹³⁾、動きの縮退問題⁴⁾¹³⁾の解消が可能である。

3. 身振りの認識

身振りの認識システムは、他の認識システムと同様に、既知パターンの教示による学習課程と未知パターンとの照合課程の2つのプロセスから構成される。

身振りが運動パラメータで記述されているとき、そ

[†] 新潟大学 工学部 情報工学科

“Recognition Technologies; Gesture Recognition” by Masanobu Yamamoto (Department of Information Engineering, Niigata University, Niigata)

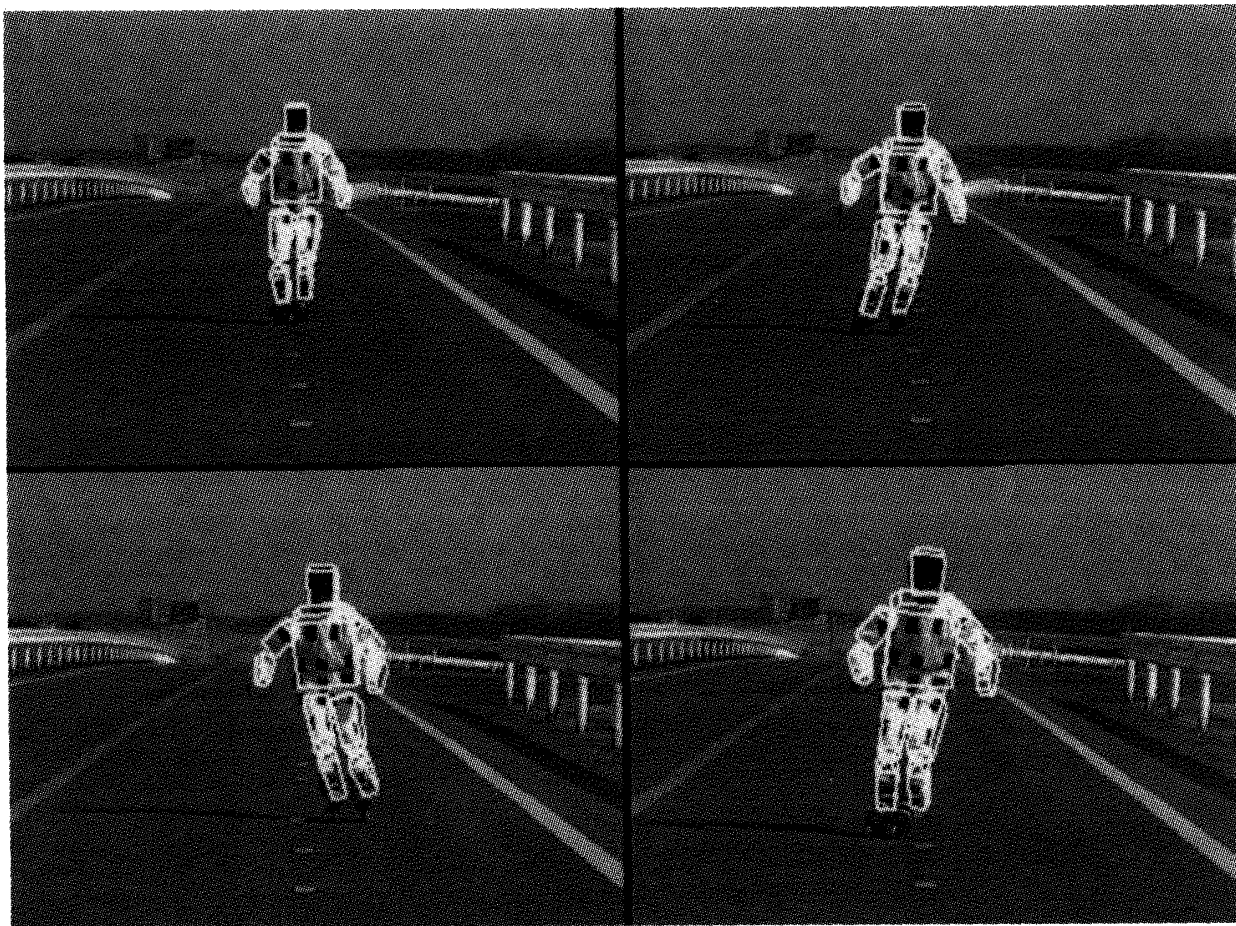


図 1 追跡結果⁶⁾ (多関節モデルが人体像に重ねて表示されている)

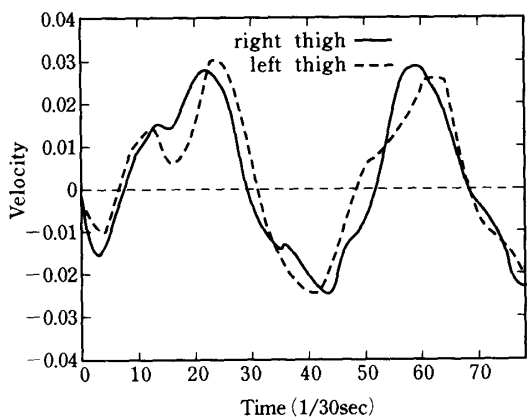


図 2 左右のひざの屈伸角速度⁶⁾

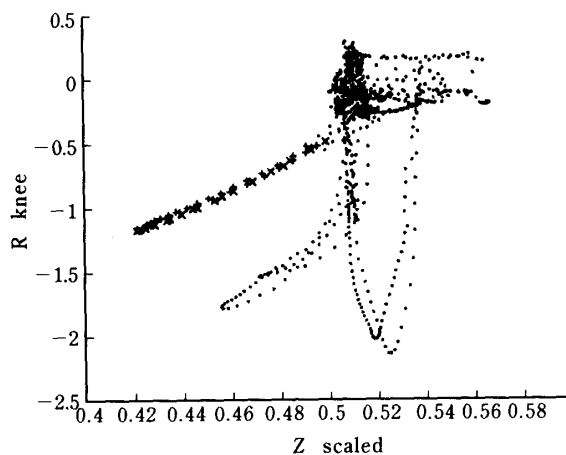


図 3 バレエのステップの位相軌道¹⁾

のパラメータ空間でのプロットは、時間の経過と共に軌跡を描く。この軌跡は位相軌道と呼ばれる。図 3 はバレエのステップを踊っている 2 人の踊り手の位相軌道が示されている。バレエのステップのひとつである Plie の位相軌道が、2 人の踊り手についてそれぞれ × 印と + 印で示されている。この図から明らかなよう

に、位相軌道を使えば、踊り手によらずステップの種類を識別することが可能である。すなわち、それぞれのステップを表す位相軌道を教示しておけば、位相軌道同士の照合によりステップ名を認識することができる¹⁾。

一方、「上手」「下手」といった動作の技量も運動パ

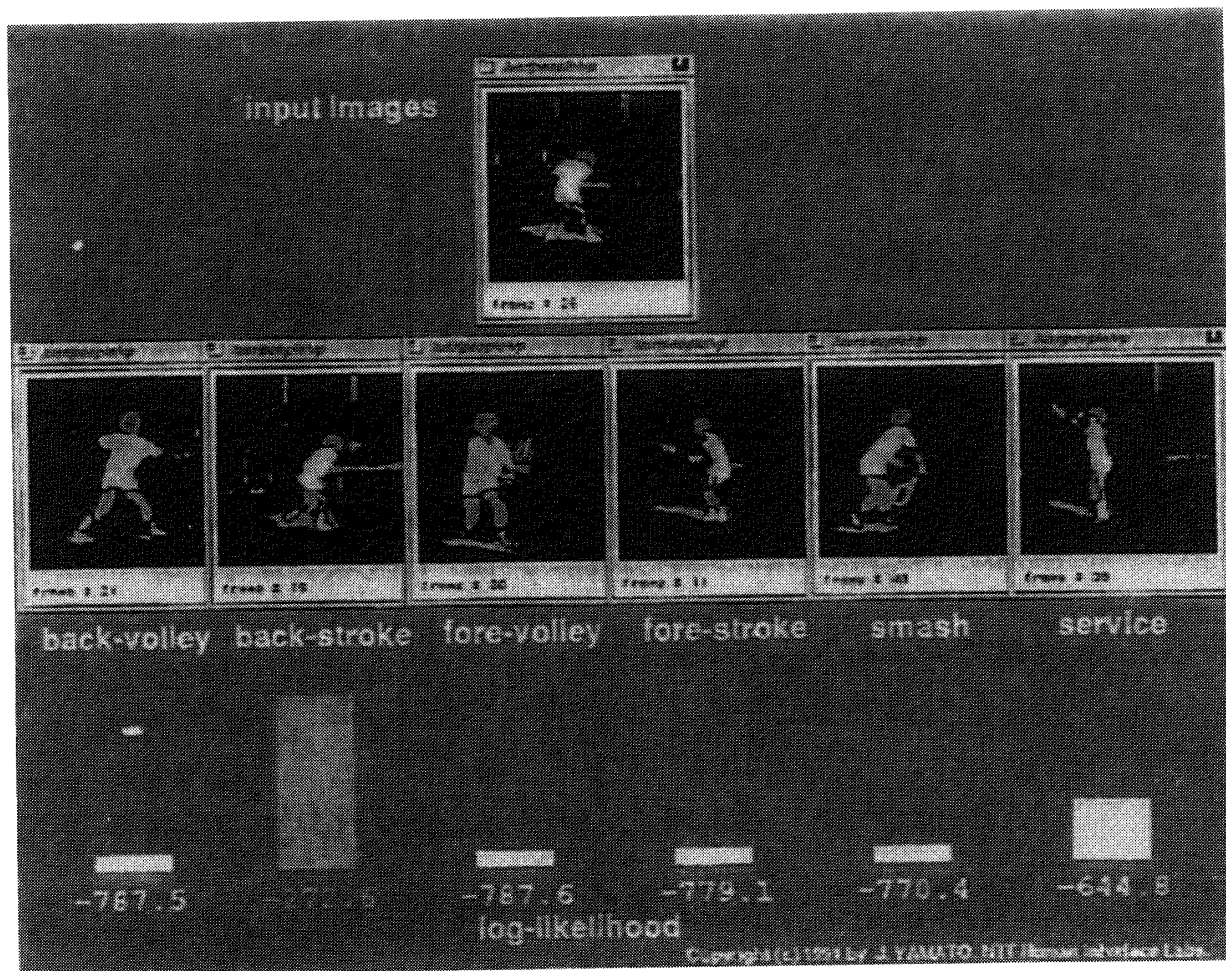


図4 テニスのストロークの認識結果¹⁷⁾

ラメータから判定することができる⁶⁾。例えば、運動パラメータを周波数分析したとき、低周波成分に対する高周波成分の比率が小さいほど滑らかな動作を表し、「上手な」動作となる。スキーなど(図1)の動作では、図2に示されるように両足が同じ動きをしているほど上手であるとされているが、これは両足の運動パラメータの相関係数の大きさによって評価することができる。

運動パラメータを使わず、動画同士を直接照合することによっても、身振りを認識することができる。その際、照明の変動などの影響を避けるため、濃淡画像ではなく、2値画像やエッジ画像を用いることが多い。また、人物像は一定の大きさに正規化しておく必要がある。

教示された身振りと未知の身振りとの照合は、動画間の差の総和や相関値によって評価される。差が小さいほど、また相関値が大きいほど、両者の類似度が高いとされる。この類似度の計算は動画の全域にわたって行われるため、膨大な計算量を必要とする。連

続DP(動的計画法)は、この計算を効率的に行うことができる¹⁵⁾。認識例として「いいえ」、「拍手」、「ばいばい」等の簡単な動作の認識が、汎用のマシンを使って実時間で達成可能となっている。

パラメトリック固有空間法を使えば、教示画像と対象画像との相関演算の計算量を少なくすることができる⁶⁾。まず、様々な個人の動画からなる画像の集合から画像の固有ベクトルを計算する。このとき、身振りの特徴は少数の固有ベクトルの線形和で表すことができる。この固有ベクトルを基底とする線形空間を固有空間とする。身振りの動画はこの固有空間内に1つの軌跡として射影される。教示動作も対象動作も固有空間内の軌跡として表したとき、動画同士の相関演算は、固有空間に射影された軌跡同士の相関演算に置き換えることができる。固有空間の次元は元の画像のサイズよりはるかに小さいので、計算量も少なくなる。

動画や運動パラメータ同士の照合が、信号レベルでの照合であるのに対し、動作パターンを記号列で表

し、記号列の照合による認識も行われている。記号レベルの照合法として隠れマルコフモデル(HMM)の利用が提案されている¹⁷⁾。

身振りを、身体の幾つかの代表的な姿勢を記号として、その記号列で表す。各時刻の姿勢は、状態遷移システムを観測した結果であるとする。状態は確率的に推移し、状態を直接観測することはできず、どの状態が観測されるかも確率的に与えられるものとする。動作の学習は、既知の動作パターンについて、この動作が生成できるように遷移モデルの生起確率を決定する。動作ごとに生起確率を決定した後、未知の動作パターンに対して、そのようなパターンを最も高い確率で生成する動作パターンをその動作名とする。図4は、この手法でテニスのストロークを認識した結果である。HMMの利点は、記号化の段階での認識ミスをリカバーできること、学習パターンと未知パターン間の時間的な伸縮に強いことである。

これまで提案された身振りの認識手法は、音声認識(連続DP¹⁵⁾やHMM¹⁷⁾や制御工学(位相軌道¹⁾、離散事象システム⁴⁾等、他の分野で実績のある手法を発展させたものが多い。このことは、時系列データを扱う他の解析法の適用可能性を示唆している。また、身振りの認識独自の手法が見出される余地も残されている。

4. む す び

動画像を使って身振りの測定と認識を行う手法について現状を紹介した。最近の画像処理関連の国際会議では、人体の動画像解析は最も人気の高いセッションとなっている。また、身振りの動画像解析専門のワークショップ(Non-Rigid and Articulated MotionやFace and Gesture Recognition)も定期的で開催されるようになってきた。身振りの動画像解析には、広大な応用領域が広がっており、今後の発展が期待される。

(1997年4月28日受付)

[参 考 文 献]

- 1) L. W. Campbell and A. F. Bobick: "Recognition of Human Body using Phase Space Constraints", Proc. of 5th ICCV, pp. 624-630 (1995)
- 2) D. M. Gavrila and L. S. Davis: "3-D Model-based Tracking of Humans in Action: A Multi-view Approach", Proc. of IEEE CVPR'96, pp. 73-80 (1996)
- 3) 岩井儀雄, 八木康史, 谷内田正彦: "単眼動画像からの手の3次元運動と位置の推定", 信学論, **J80-D-II**, 1, pp. 44-55 (1997)
- 4) I. A. Kakadiaris and D. Metaxas: "Model-based Estimation of 3D Human Motion with Occlusion based on Active Multi-viewpoint Selection", Proc. of IEEE CVPR'96, pp. 81-87 (1996)
- 5) 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, **J79-D-II**, 1, pp. 26-35 (1996)
- 6) 近藤拓也, 山際貴志, 山中光司, 山本正信: "動画像からの動作感性情報の抽出", 信学論, **J80-D-II**, 1 pp. 247-255 (1997)
- 7) 倉掛正治, ラマカント・ネバティア: "関節のある動物体の記述と追跡", 信学論, **J76-D-II**, 4, pp. 854-862 (1993)
- 8) H. Murase, R. Sakai: "Moving Object Recognition in Eigenspace Representation: Gait Analysis and Lip Reading", Pattern Recognition Letters, **17**, pp. 155-162 (1996)
- 9) 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, **J77-D-II**, 8, pp. 1562-1570 (1994)
- 10) J. Ohya and F. Kishino: "Human Posture Estimation from Multiple Images Using Genetic Algorithm", 12th ICPR, pp. 750-753 (1994)
- 11) J. M. Rehg and T. Kanade: "DigitEyes: Vision-based Human Hand Tracking", CMU-CS-93-220 (1993)
- 12) K. Rohr: "Towards Model-based Recognition of Human Movements in Image Sequences", CVGIP: Image Understanding, **59**, 1, pp. 94-115 (1994)
- 13) 佐藤明知, 川田聡, 大崎喜彦, 山本正信: "多視点動画像からの人間動作の追跡と再構成", 信学論, **J80-D-II**, 6 pp. 1581-1589 (1997)
- 14) 島田伸敬, 白井良明, 久野義徳: "確率に基づく探索と照合を用いた画像からの手指の3次元姿勢推定", 信学論, **J79-D-II**, 7, pp. 1201-1217 (1996)
- 15) 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トティング認識", 信学論, **J77-D-II**, 8, pp. 1552-1561 (1994)
- 16) 山本正信, 川田聡, 近藤拓也, 越川和忠: "ロボットモデルに基づく人間動作の3次元動画像追跡", 信学論, **J79-D-II**, 1, pp. 71-83 (1996)
- 17) 大和淳司, 大谷淳, 石井健一郎: "隠れマルコフモデルを用いた動画像からの人物の行動認識", 信学論, **J76-D-II**, 12, pp. 2556-2563 (1993)



やまもと まきのぶ
山本 正信 1973年,九州工業大学工学部制御工学卒業。1975年,東京工業大学大学院修士課程修了。同年,電総研入所。コンピュータビジョンの研究に従事。1989~1990年,カナダ国立研究協議会招聘研究員。1992年より新潟大学工学部情報工学科教授。工学博士。