

距離センサを用いた特定動作に対する動作判定システムの開発 Development of a System for Evaluating Human Motions Using Depth Sensor

榮門 恒希[†] 浮田 宗伯[†] 萩田 紀博[†]
Koki Eimon Norimichi Ukita Norihiro Hagita

1. はじめに

1.1 背景

人間の動作の正確性が要求される場面やタスクは数多く存在する。例えば、医療分野においては、リハビリ動作の正確性が症状の改善効果や安全性に大きな影響を与えられている[1]。また、スポーツ分野においては競技選手の動作が正確でなければ、パフォーマンスの低下や怪我の危険を招いてしまう[2]。現在、こうした動作の正確性の評価は人間が直接行うことが通常であるが、理学療法士やコーチといった専門家を、動作評価が必要なすべてのケースに派遣するには人的コスト等の種々な制約がある。このことから自動で人間の技能的な動作の正確性を評価できるシステムが必要とされている。

こうした背景から、膝関節症患者の脚部に装着した加速度計の計測結果から症状を推定する研究[3]をはじめとしたリハビリ支援の自動化システム[4]や、マーカやラインの入った専用服を着てカメラの前でゴルフスイングをした際の画像を解析して理想的なスイングに向けてのアドバイスするシステム[5]など、活発に研究開発が進められている。しかしながら、これらの研究では人体に何かしらの計測器具(たとえば、加速度センサや筋電位計測器など)を装着することが前提となっている。これらの器具には、人体に装着するための手間、特殊な器具入手に関するコストという負担が常につきまとうため、こうしたシステムが普及していくための大きな障害となっている。さらに本質的な問題として、そうした装着型計測機器がリアルで自然な動きの計測を阻害することにもなってしまう。そのため、以上の制約を受けない非接触で安価なデバイスによって動作を計測することが望ましい。

1.2 関連研究

非接触センサによる動作計測として、光学カメラを用いた人体姿勢推定研究が数多く行われている。これらは使用するカメラの数で二つに大別される。一つ目は、計測対象を囲むように複数の同期カメラで撮影する手法で[6]、二つ目は、一方向からの撮影画像の見えから姿勢を推定する手法[7]である。

前者の手法は、一方向からの見え情報のみに依存する後者と比較すると情報量が豊富なので推定精度は高くなる傾向があるが、システムの規模が大きく計算量も多くなる。システムの規模が大きくなると、金銭的コストが高騰し、最終的なシステムを一般家庭に対しても普及させることは難しくなるという点も現実問題として考慮すべきである。

一方、後者のタイプの研究では、一方向観測だけでは情

報量が少なく、自己遮蔽などの問題を直接的に解決できないということがネックであったが、近年 Microsoft 社製の Kinect などの距離センサの急速な普及によって、デプス画像から得られる距離情報を利用した姿勢推定の研究が非常に活発になっている。Kinect による姿勢推定法で特に有名な文献[8]では、デプス画像の各画素のパーツラベル(頭、右肩、等)を識別し、その識別結果から実時間で関節位置推定を行うことを可能にした。この手法は、推定の高速化といかなる姿勢、または大きく変動する姿勢に対しても頑健に姿勢推定できることを目的としている。これに対し、本研究で目標とする特定動作における動作判定システムを実現するためには、限られた変化の姿勢をより正確に推定できる手法のほうが望まれる。

一方、文献[6, 9]では、画像から抽出した特徴量から Motion Capture によって計測した真値の姿勢データへの回帰画像を学習しておき、未知の画像に対してその回帰画像を適用することで姿勢推定を行う。こうした回帰画像ベースの手法は観測される人の動きが学習データ中のそれと類似している場合には他手法と比べて高精度な推定が可能であるといった利点がある一方、学習データとテストデータのずれが大きくなると、推定が破綻してしまうという欠点もある。しかし、この欠点は「目標とする正しい動作に近い動作のみを評価する」という本研究のゴールに対しては、大きな問題とならない。そこで本研究では、特定の動作に特化した高精度な姿勢推定を回帰画像ベースで行う。

上記の方法で推定される結果は、フレーム毎に独立した姿勢である。動作とは姿勢の時系列集合であり、動作評価に際しては時系列に即した解析が必要である。時系列パターン認識においては、DP(Dynamic Programming) マッチングのような動作軌道の連続的なマッチングに基づく認識手法があり、本研究でもこの手法を採用する。

1.3 目的

以上のことを踏まえて本研究では、安価な非接触センサである Kinect を用いて姿勢推定を行い、その結果から自動で動作評価を行うシステムの開発を目的とする。なお、このシステムでは正確(熟練)、非正確(非熟練)の動作の違いが大きく、熟練被験者の確保が容易であるという理由から野球の投球動作(オーバースロー)を対象動作とすることとする。本研究では投球動作という特定動作を対象を絞っているため、より高精度な姿勢推定のために[6, 9]に見られる統計的モデル学習のアプローチを取り入れることとする。また、動作判定ではいつ、どこで誤った動作を行ったのかを具体的に判定できるシステムを目指す。本稿では一連の動作が正確か非正確かを判定するシステムを開発する。

2. システム概要

本システムは図 1 に示すように、姿勢推定処理と動作判定処理の 2 つのフェーズから構成され、各フェーズはシス

[†] 奈良先端科学技術大学院大学 NAIST

テム開発者が行うオフライン処理と利用者が行うオンライン処理から構成される。

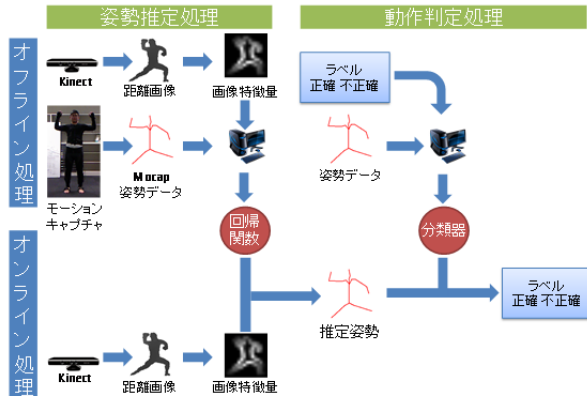


図 1 システム概要

3. 姿勢推定処理

姿勢推定処理におけるオフライン処理では、Kinect の出力であるデプス画像に HOG[10]を適用して得られる画像特徴量空間からジャイロ式 Motion Capture(IGS-190)から得られる 18 関節の 3 次元位置の姿勢空間への写像関数として学習する。学習は一般的な統計的学習手法である GPR(Gaussian Process Regression)を用いる。学習のためのデータ取得は、Motion Capture スーツを着用した熟練者と非熟練者に試技動作を行なってもらい Kinect でその様子を撮影して姿勢データとデプス画像を得る。

姿勢推定処理におけるオンライン処理では、Kinect の出力であるデプス画像に HOG を適用して得られる画像特徴量と学習によって作成された写像関数から姿勢を推定する。

4. 動作判定処理

動作判定におけるオフライン処理では姿勢データとその姿勢データに対応するクラスラベル(上手, 下手)を対応付けて学習させて記録しておく。

オンライン処理では、入力動作と学習動作の間のフレーム対応付けを行う必要がある。同じ試技でも、被験者によって動作の推移の時間幅は均等ではない。よって DP を用いて全ての学習動作の入力動作との間の同期をとっておく。次に、入力動作の各フレームにおいて、すべての学習動作における同期フレームとの類似度(54 次元姿勢ベクトル間の距離の逆数)を計算する。類似度最小の学習動作が正しい動きクラスの学習データであればこのクラスに投票し、他方のクラスであればそのクラスに投票する。最終的により多くの票を獲得したクラスが入力動作の判定結果とする。

5. 実験

5.1 姿勢推定精度

将来的なシステム運用イメージでは、学習データ獲得時のみ Motion Capture により姿勢データの Ground-truth を得るが、本実験では、姿勢推定の評価を行うために、テストデータ撮影時にも Motion Capture を使用した。Ground-truth との 3 次元関節位置の誤差を平均することで算出する。得られた推定誤差は最良のケースで 12.8cm となった。

5.2 動作 2 値判定精度

本実験では各テスト動作全体を通して、それが上手か下手かの 2 値判定処理を行った。判定精度は入力動作中の判定成功シーケンス数の割合で評価する。姿勢推定を介した動作判定では 50%前後の成功率が得られた。また、クラス分類の評価のために、Ground-truth 同士を用いて動作判定精度の評価を行った。絶対座標系の姿勢データを用いた判定では 108 の各入力動作に対し 93.5 %の成功率を示した。

5.3 考察

姿勢推定誤差は 12.8 cm と良くない結果が出た。原因としては、今回の姿勢推定は一つの画像特徴量から全身姿勢への写像を求めていたが、この手法だと局所的な距離情報の誤差に全身の姿勢が強く影響を受けてしまったことが考えられる。これを解消するため、各ピクセルで得られる特徴量と各ピクセルに対応する関節の写像を求め、全ての特徴量の多数決から関節位置を推定するといった手法が考えられる。また、シルエットが無い場所には関節は無いという幾何学的制約を加えることも有効な解決策と言える。

6. まとめ

距離センサを用いて特定動作の動作判定システムを開発している。現状では、姿勢推定精度が良くないため十分な成果が得られていない。今後は推定精度を向上させ、判定処理についてもより細かく判定できるシステムを目指す。

謝辞

本研究において基礎を作り上げ大きな助力を頂いた NAIST 修了生 花光政和さんに心より感謝申し上げます。

参考文献

- [1] K. S. Thomas, K. R. Muir, M. Doherty, A. C. Jones, S. C. O'Reilly, and E. J. Bassey, "Home based exercise program for knee pain and knee", *British Medical Journal*, Vol.325 (2002).
- [2] 諏訪 正樹, 西山 武繁, "アスリートが「身体を考える」ことの意味", 人工知能学会身体知研究会 (2009).
- [3] P. E. Taylor, G. J. M. Almeida, T. Kanade, and J. K. Hodgins, "Classifying Human Motion Quality for Knee Osteoarthritis Using Accelerometers", *IEEE Engineering in Medicine and Biology Society* (2010).
- [4] 桑江 豊, 沼田 崇之, 関根 正樹, 三好 寿顕, 辻 美和, 岡部 一郎, 原 圭太, 藤元 登一郎, 田村 俊世, "ウェアラブルモーションセンサを用いた要支援認定者の低頻度運動療法の検証", 生体医工学シンポジウム (2011).
- [5] 植田 勝彦, 大貫 正秀, 白井 良明, 島田 伸敬, "映像を用いたゴルフスイング診断システムインストラクターがいなくてもスイングチェックができるシステム" 画像ラボ, 1月号 (2007).
- [6] N. Ukita and T. Kanade, "Gaussian Process Motion Graph Models for Smooth Transitions among Multiple Actions", *Computer Vision and Image Understanding*, Vol.116, Issue 4 (2012).
- [7] C. -S. Lee and A. M. Elgammal, "Coupled Visual and Kinematic Manifold Models for Tracking", *In International Journal of Computer Vision*, Vol.87, No.1-2 (2010).
- [8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-Time Human Pose Recognition in Parts from Single Depth Images", *IEEE Computer Vision and Pattern Recognition* (2011).
- [9] R. Urtasun and T. Darrell, "Sparse Probabilistic Regression for Activity-independent Human Pose Inference", *IEEE Computer Vision and Pattern Recognition* (2008).
- [10] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *IEEE Computer Vision and Pattern Recognition*, Vol.1 (2005).