

単一のウェアラブルカメラを用いた人間の歩行動作推定

渡辺 義浩^{*1}

畑中 哲生^{*2}

小室 孝^{*3}

石川 正俊^{*4}

Human Gait Estimation Using a Wearable Camera

Yoshihiro Watanabe^{*1}, Tetsuo Hatanaka^{*2}, Takashi Komuro^{*3} and Masatoshi Ishikawa^{*4}

Abstract – We focus on the growing need for a technology that can achieve motion capture in outdoor environments. The conventional approaches have relied mainly on fixed installed cameras. With this approach, however, it is difficult to capture motion in everyday surroundings. This paper describes a new method for motion estimation using a single wearable camera. We focused on walking motion. The key point is how the system can estimate the original walking state using limited information from a wearable sensor. This paper describes three aspects: the configuration of the sensing system, gait representation, and the gait estimation method.

Keywords : motion capture, motion estimation, walking motion, wearable sensor, high-speed camera

1 はじめに

近年、日常の生活環境下で人間の位置や姿勢を取得するセンシングシステムのニーズが高まっている。このようなシステムを用いて、人間の動作を推定・予測することで、様々な支援応用が可能になると考えられる。応用展開に向けて重要な点は、そのセンシングが場所に制限されないことである。この点について、これまでのモーションキャプチャの技術は十分ではなかった。

従来のモーションキャプチャは、主にコンピュータグラフィックスのアニメーションを生成するために設計されていた。典型的なシステムでは、複数のカメラを対象の周囲に設置し、関節に装着された光学マーカを観測することで、各マーカの3次元位置を計測し、身体の姿勢を取得していた [1, 2, 3]。近年では、マーカレスでのキャプチャ技術も開発されている [4]。これらは主に、モーションキャプチャのために専用化された屋内のスタジオにおいて実施されることが前提となっていた。一方で、同様のシステム構成のもと、屋外環境においてもモーションキャプチャを行う手法が提案されている [5]。同提案では、太陽光スペクトルに存在する吸収線の波長帯域のLEDをマーカとして利用することで、屋外でもモーションキャプチャが可

能になると報告されている。

このようなシステムでは、単独の人間を観測するために、複数のカメラが各視点からマーカなどの遮蔽が生じないように固定設置されている必要がある。この条件は、計測環境を対象に応じて自由に調整できる場合は問題ないが、日常生活下でのセンシングを目標とする場合、異なる技術が必要になると考えられる。

このような要請に対して、ウェアラブルなセンサで構成されるシステム技術は有効であると考えられる。例えば、Vlasicらは、加速度センサ、ジャイロセンサ、超音波源、及びマイクロフォンを用いた新たなモーションキャプチャを開発した [6]。同システムは日常生活下において様々な動作を取得することを目的として設計されており、実験では自転車に乗る動作や、スキーの動作、卓球の動作などが捉えられることが実証されている。一方、同システムでは、多くのセンサを身体の理想的な部位に装着する必要があり、手軽ではないと考えられる。また、超音波を用いて身体部位の位置を計測しているため、混雑した環境下で同システムを複数台利用することは難しいと考えられる。

また、同様に、加速度センサやジャイロセンサなどで構成されるセンサモジュールを身体の各部位に装着するものとして、市販化されているXsens社のシステム [7] や、より小型化を目指したシステム [8] などがある。これらの技術では、センサ結果の積分によって、移動位置を特定する必要があるため、累積誤差が蓄積する問題がある。

この他、身体に装着された加速度センサのみを用いて動作を認識する研究例も報告されている [9]。ただし、同研究では基本動作が実施されたことを検出する

^{*1}東京大学 情報理工学系研究科

^{*2}コナミ株式会社

^{*3}埼玉大学 理工学研究科

^{*4}東京大学 情報理工学系研究科

^{*1}Graduate School of Information Science and Technology, University of Tokyo

^{*2}KONAMI Corporation

^{*3}Graduate School of Science and Engineering, Saitama University

^{*4}Graduate School of Information Science and Technology, University of Tokyo

ことを目的としており、連続的な姿勢を推定するものではない。

さらに、歩行動作に特化したウェアラブルな動作取得システムも開発されている [10]。同システムでは、磁気センサを用いて、腰とかかとの相対位置を計測する。歩行中に足が接地すると、足底に取り付けられたボタンが押され、その時点での腰と両足の位置を加算することで歩行軌道が生成される。上記の例に比べると、センサ数は少ないが、システムは複雑な設計となっている。また、地面に足が接地した際の移動距離を積分することで位置を計測しているため、累積誤差の影響が大きいと考えられる。

このように、ウェアラブルなセンサ構成を導入した場合、ユーザ位置の取得情報に累積誤差が蓄積することを避けることが難しい。これに対して、センサが搭載されたシステムの動きと周囲環境の地図を同時に推定することで、同問題を解決する技術が提案されている。例えば、カメラのみで周囲環境の3次元マップとカメラの位置姿勢を推定するものに Structure from Motion(SfM)[11]がある。この他にも、移動ロボットなどでの応用を想定し、複数のセンサを組み合わせて使う技術として、Simultaneous Localisation and Mapping (SLAM)がある [12, 13]。SfMは撮像された画像群の全てを使って推定を行うが、SLAMでは新たに撮像された画像が入力される度に推定結果を更新する方式となっている。このような技術では、加速度センサなどのようなセンサと異なり、周囲環境のランドマークを発見して同定することができるため、累積誤差の問題を回避することができている。同技術は、ロボットなどの位置を推定するためのものであり、ウェアラブルなカメラによって人間の身体姿勢を推定するためのものではないが、手法の基本構成は導入が可能であると考えられる。

このような考えを導入したものとして、身体の各部位に複数台のカメラを搭載して、モーションキャプチャを実施するシステムが提案された [14]。同システムでは、最低16台のカメラを身体に装着し、取得された映像から SfM 技術によって、各カメラの位置姿勢を推定し、身体動作を取得している。ウェアラブルセンサによるモーションキャプチャ技術として新たな構成に取り組むものであるが、身体中にセンサを取り付ける面で、日常生活での支援応用ではなく、より特殊な用途を応用展開として想定している技術であると考えられる。

このほか、モーションキャプチャ技術によって取得された身体姿勢を用いて、どのような動作が行われているかを識別する手法も提案されている [15, 16]。しかし、同手法は全身の関節情報を用いて、明確に分類

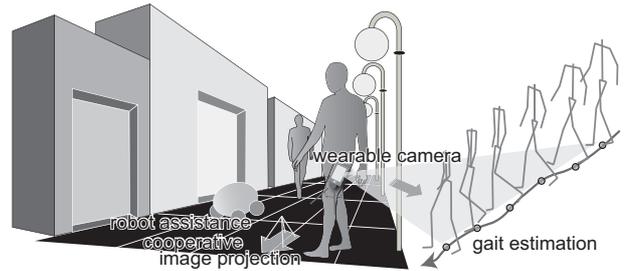


図1 単一のウェアラブルカメラを用いた歩行動作推定

Fig.1 Human gait estimation using a wearable camera.

可能な異なる動作を識別するものである。また、本論文では、動作の識別ではなく、身体の位置と姿勢の推定の性能が最大化されることを目標としている点で異なっている。

このように、少ない数のウェアラブルセンサを用いることで、身体の姿勢と大域的な位置を推定する技術はニーズが高いが、実現が困難であった。そこで、本論文では、動作を歩行に限定することで、システム構成の単純化を図る。歩行動作は近似的に周期的な挙動として記述できる可能性があるため、限られたセンサデータからでも、オリジナルの歩行動作を完全に復元することができると考えられる。本論文の技術構想を図1に示す。本技術は、図に示されるように、ユーザに装着された単一のカメラから、歩行時の連続的な姿勢とその移動を推定するものである。例えば、推定された動作を利用して、映像投影システムや、ロボット、移動体などを制御することで、協調的な支援環境を実現することが可能になると考えられる。このように、歩行動作に限られた場合でも、必要なセンサがカメラ1つの単純なシステム構成となる点や、その応用展開が幅広いことから本技術は有用性が高いと考えられる。

2 単一のウェアラブルカメラを用いた歩行動作推定

本節では、対象となる人間に装着された単一のウェアラブルカメラを用いて、歩行動作を推定するためのシステム構成について述べる。近年、カメラを始めとしたセンサ類のサイズとコストに関して、急激な削減化が進んでいる。実際、多くのモバイル端末には、1つ以上のセンサが搭載され、入力などの支援に用いられている。このように、いつでもどこでも、人間がセンサを装着した状態にいることは前提条件として設定可能であると考えられる。特に、カメラを用いた場合は、外界を捉えることができるため、加速度センサなどのような場合に生じる累積誤差を軽減することができる。と期待できる。

提案システムの構成を図1に示す。本システムでは、計測対象となるユーザは自身の片方の大腿部に単一のカメラを下向きに装着する。設置位置が既知であれば、

右脚と左脚のどちらに装着しても良いとする。このように装着されたカメラは、ユーザの動作によって生じる外界の地面の変動を捉えることになる。本論文では、この構成のみでユーザの歩行動作が推定可能であることを示す。提案する手法では、カメラによる観測情報が、人間の動作モデルに基づいて生成されることを利用し、動作姿勢とグローバルな位置の両方を推定する。

本手法はリアルタイム処理が可能な構成を目標とする。具体的には、入力画像が取り込まれるたびに、同データに対応する出力値をフレーム時間(使用するカメラのフレームレートの逆数)程度の遅延で応答することを目指す。このため、ビデオクリップ単位ではなく、フレーム単位で認識結果を取得する。このように、動作の識別(歩いているか、走っているか、曲がっているかなどの動作の種類を同定する)のではなく、対象の位置と身体姿勢を捉えることを主眼に置いており、一連の動作が完了した後に何が起こっていたのかを認識するのではなく、動作の最中に認識をすることを目的とする。本手法では、システムが保持する歩行姿勢が個人に特化することを前提としているため、予め得られるデータ量は少数であるとする。

なお、上記で仮定したカメラの視線方向は必須の前提条件ではないと考えられる。本論文では、下向きの視線にすることで、ユーザの周囲の他の人間や車などの移動物体を、カメラの像面から取り除き、画像処理を単純化することを意図している。今回は提案手法の基本的な実証を目的として取り組んだため採用しなかったが、カメラを前向きに設置することも構成案として考えられる。この場合は、周囲環境を広く捉えることができる。これによって、位置を特定するための情報量が大幅に増えることが期待されるため、ランドマークのみを発見する画像処理の問題を解決できれば、有効であると考えられる。

このような構成のもと、本タスクは動作が特定のモデルに基づいているという状況下で単一カメラの映像からオリジナルの動作を推定する問題となる。ただし、動作モデルは唯一ではなく、特定の規則に従って、時間とともに変化しうるものとする。まず、3節と4節ではこのような歩行動作のモデルについて述べる。次に、5節では提案モデルに基づく推定手法を述べる。

3 歩行動作の表現

撮像された動画像における変動から実際の動作を推定するために、予め歩行動作モデルを事前に準備する必要がある。特に、本手法では、ユーザごとに歩行動作のモデルを準備するものとする。

まず、これまでに提案されている歩行動作のモデルについて述べる。倒立振り子モデルは、2足歩行の単純なモデルとして広く使われている [17]。例えば、主な

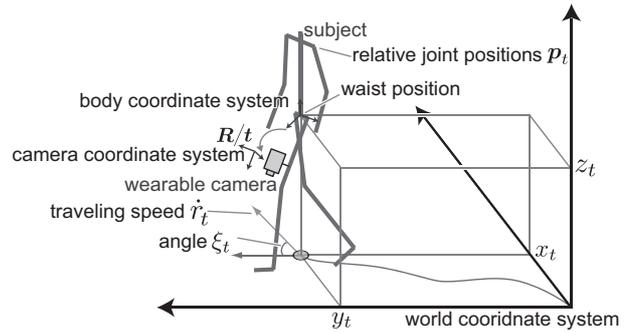


図2 パラメータの定義
 Fig.2 Parameter definitions.

応用例として、ロボットの歩行動作パターン生成がある。しかし、倒立振り子モデルは人間の歩行動作を正確に表したものではないため、今回のように不十分なセンサデータを補完するためには不十分であると考えられる。

一方、コンピュータグラフィックスの分野では、モーションキャプチャから取得されたデータを直接用いて動作のアニメーションを生成している。例えば、Unumaらは、モーションキャプチャを用いて取得された歩行動作をアニメーション化するために、フーリエ解析を導入した [18]。多くの歩行動作は周期動作として特徴づけられるため、このような周波数領域での解析は有効であると考えられる。同事例では、歩行動作が低周波数成分の情報のみで表現が十分に可能であることが示されている。この点は、歩行動作のモデルを少ないパラメータで表現できるため、本タスクの推定においても有効であると考えられる。そこで、提案手法では、同手法に基づく歩行動作のモデル化を図った。

ここで、歩行動作を表現するパラメータの定義について述べる。特に、本論文では、歩行中の各時刻の位置・姿勢を歩行状態と呼ぶ。各時刻の歩行状態 μ_t は次式で更新される。

$$\mu_t = \begin{bmatrix} x_t \\ y_t \\ z_t \\ \xi_t \\ \mathbf{p}_t \end{bmatrix} = \begin{bmatrix} x_{t-1} + \dot{r}_t(\eta_{t-1}) \cos(\xi_{t-1}) \\ y_{t-1} + \dot{r}_t(\eta_{t-1}) \sin(\xi_{t-1}) \\ z(\eta_{t-1}) \\ \xi_{t-1} + \dot{\xi}(\eta_{t-1}) \\ \mathbf{p}(\eta_{t-1}) \end{bmatrix} \quad (1)$$

上式で、 $[x_t, y_t, z_t]$ は時刻 t における腰の位置である。 \dot{r}_t は腰の進行方向の速度である。 $\dot{\xi}_t$ は同角速度である。また、歩行動作は、位相 η_t による周期動作としてパラメータ化される。各関節は腰位置を原点とする相対座標系として記述され、 $\mathbf{p}_t = \{\mathbf{p}_t^i | i = 1, \dots, N_p\}$ で表す。図2にパラメータの定義を表す。

本表現では、環境に固定された座標系、ユーザの身体に固定された座標系、及びカメラに固定された座標系の3つの座標系がある。環境と身体の間座標系変

換は、歩行状態によって記述できる。また、カメラは身体に固定されているため、身体とカメラの間の座標系の変換は事前に分かっているものとする。

この定義のもと、サンプルの歩行データを事前に収集し、パラメータ $\mathbf{d}_t = [z_t, \dot{r}_t, \dot{\xi}_t, \mathbf{p}_t]$ の挙動を解析する。歩行のサンプルは通常のもーションキャプチャシステムによって取得するものとする。本手法では、収集された周期データをフーリエ級数展開を用いて、5次までの周波数成分で表現する。例えば、腰の高さは次式で表わされる。

$$z_t = \sum_{n=0}^5 a_n^1 \sin(n\eta_t + b_n^1). \quad (2)$$

このように、歩行動作は (a_n^i, b_n^i) を変数とするモデルとしてパラメトリックに表現される。ここで、添え字 i は腰の高さを始めとする各パラメータを識別する番号である。以上の表現により、あるサンプルの歩行状態は、特徴ベクトル $\mathbf{f} = \{a_n^i, b_n^i | i = 1, \dots, N_p + 3\}$ を用いて保持される。

4 歩行動作の分類

人間の歩行動作には、複数の種類があると考えられる。本節では、歩行動作を分類する手法について述べる。本論文では、分類されたそれぞれの動作を歩行形態と呼ぶ。また、安定的な歩行において形態数は有限であるとし、システムは十分な数の歩行形態を予め取得することを前提とする。

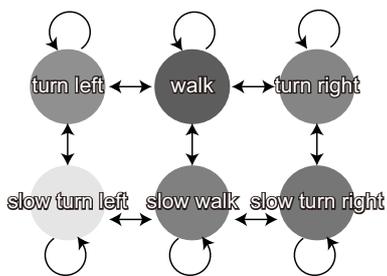


図3 動作の遷移図例

Fig. 3 Transition diagram of walking forms.

まず、多数の歩行サンプルをモーションキャプチャシステムを用いて取得する。次に、取得されたサンプルを、3節で述べた周波数解析を適用して特徴ベクトル \mathbf{f}_k に変換する。さらに、この収集されたデータ $\{\mathbf{f}_k : k = 1, \dots, N_s\}$ を次元圧縮し、分類のための空間へ射影する。この射影された各点を、クラスタリング手法によって分類する。本論文では、主成分分析を適用し、特徴ベクトルを3次元まで圧縮した。次式に具体的な変換を示す。

$$\mathbf{g}_k = \mathbf{U}(\mathbf{f}_k - \bar{\mathbf{f}}) \quad (3)$$

ここで、 \mathbf{U} は分析によって得られた正規直交基底、 $\bar{\mathbf{f}}$ は全歩行データの平均、 \mathbf{g}_k が圧縮された特徴量である。

次元削減をされた特徴量空間において、距離が近い2つの歩行データは同じ歩行形態か、似た歩行形態であると考えられることができる。そこで、これらの歩行データに対して、k-means クラスタリング [19] を適用し、複数のクラスタに分類する。分類された各クラスタを、個別の歩行形態として登録する。特に、各クラスタの重心に最も近い点を同クラスタ内の代表歩行データとする。

本手法では、歩行形態に分類後、各形態間の遷移図を生成する。本手法における遷移図の概念図を図3に示す。本手法の遷移図では、遷移が起こりうる形態間の接続関係を保持するものとする。このような遷移図を用いることで、形態が時系列に遷移する条件を予め保持することができるため、推定時の探索範囲を制限し、ロバスト化を図ることができると考えられる。なお、本論文では、収集されたサンプルから、形態間の遷移を観測し、接続関係を生成するものとする。

5 歩行状態の推定手法

5.1 全体の概要

提案手法では、ウェアラブルカメラで観測された像面上の変動から、歩行形態と状態を推定する。本推定では、歩行状態を $\tilde{\boldsymbol{\mu}}_t = [x_t, y_t, \xi_t, \eta_t, \dot{\eta}_t]$ で定義する。3節で定義された完全な歩行状態の情報 $\boldsymbol{\mu}_t$ は、この推定された状態 $\tilde{\boldsymbol{\mu}}_t$ と形態 \mathbf{f}_k から復元することができる。

本手法では、センサのサンプリングごとに歩行状態を推定する。手法の概要を図4に示す。本手法では、複数の状態候補を生成する。時刻 t での候補総数は M_t とする。各候補を用いて、歩行形態の選択、歩行状態の予測、歩行状態の尤度推定、歩行状態のリサンプリング、歩行状態の更新の5ステップを実施する。状態予測と更新には、拡張カルマンフィルタを用いる。この5ステップを適用した結果、計算された尤度が最も高い状態候補を各時刻で決定し、その時刻における歩行姿勢及び位置を取得する。本手法では、状態遷移モデルを動的に入れ替えているため、拡張カルマンフィルタの収束性が悪い可能性がある。そこで、上述のように状態候補を複数用意することでその悪化を防いでいる。推定が発散した候補は、リサンプリングの処理によって、自動的に切り捨てられ、精度を保証した候補のみが残るため、動作の安定化を図ることができると考えられる。各ステップの詳細は次節以降で述べる。

なお、本論文では、カメラの画像上で観測された特徴点の環境における世界座標が分かるものとする。な

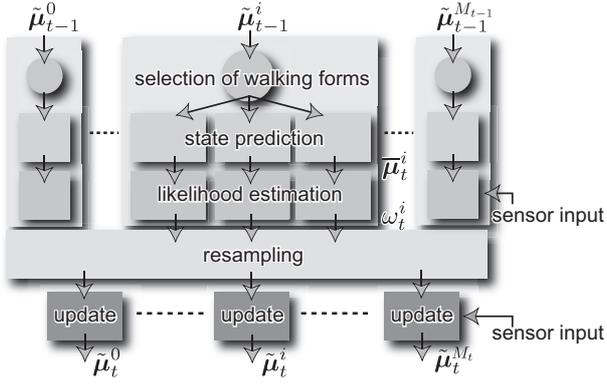


図4 歩行動作推定の概要図

Fig. 4 Overview of the walking estimation technique.

お、この仮定はSLAMの枠組みを導入することで取り除くことが可能である。具体的には、拡張カルマンフィルタの状態変数を拡張することで導入できると考えられる。また、カメラの内部パラメータは事前に取得されているものとする。

5.2 歩行形態の選択

時刻 $t-1$ に生成された i 番目の状態候補を $\tilde{\mu}_{t-1}^i$ とする。同状態の時刻 t における歩行形態は、遷移図に基づいて選択される。例えば、状態候補 $\tilde{\mu}_{t-1}^i$ が図3の形態“turn left”で更新されたとする。この場合、遷移図に基づき、同状態は形態“turn left”, “walk”, 及び“slow turn left”の歩行データを用いて検証される。

5.3 歩行状態の予測

前節で述べたステップでは、状態候補 $\tilde{\mu}_{t-1}^i$ のための時刻 t での歩行形態が選択された。この選択された歩行形態の特徴 f_k を用いて、状態予測が適用される。次式が予測処理である。

$$\tilde{\mu}_t^i = w(\tilde{\mu}_{t-1}^i) + \epsilon_t = \bar{\mu}_t^i + \epsilon_t. \quad (4)$$

ここで、 ϵ_t はプロセスノイズである。 w は非線形の動作モデルである。このステップでは、予測値 $\bar{\mu}_t^i$ のみを計算する。モデル w として、状態 $\tilde{\mu}_t^i$ のうちのパラメータ $[x_t, y_t, \xi_t]$ は、式(1)に基づいて予測される。また、位相パラメータは次式に基づいて予測される。

$$\begin{bmatrix} \eta_t \\ \dot{\eta}_t \end{bmatrix} = \begin{bmatrix} \eta_{t-1} + \dot{\eta}_{t-1} \\ \dot{\eta}_{t-1} \end{bmatrix}. \quad (5)$$

5.4 歩行状態の尤度推定とリサンプリング

i 番目の状態候補の尤度 ω_t^i は、1時刻前の $t-1$ における尤度を用いて、次式で計算される。

$$\omega_t^i = C\lambda_t^i\omega_{t-1}^i. \quad (6)$$

連続する2時刻では、同じ歩行形態で動作する確率が高いと考えられる。そこで、候補が同じ歩行形態で

予測されていた場合には、より高い尤度を与える操作を導入する。この操作は定数パラメータ C によって制御される。もし、連続する2時刻での形態が同じである場合、 C は1よりも大きな値をとり、そうでない場合は1とする。

パラメータ λ_t^i は、予測された状態値の信頼度を表わしている。この信頼度の計算方法について述べる。5.1節で述べたように、本手法では観測された特徴点の世界座標とカメラの内部パラメータを既知としているため、予測された状態によって記述される歩行姿勢を用いて、同世界座標をカメラの画像面上に投影することができる。投影された点と対応する観測された点の間の誤差を用いて、同信頼度 λ_t^i は下式で定義される。

$$\lambda_t^i = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(e(\mathbf{u}_t, \mathbf{h}(\bar{\mu}_t^i)) - \nu)^2}{2\sigma^2}\right\} \quad (7)$$

$$e(\mathbf{u}_t, \mathbf{h}(\bar{\mu}_t^i)) = \frac{\sum_{k=1}^{N_t} \|\mathbf{u}_t^k - \mathbf{h}^k(\bar{\mu}_t^i)\|}{N_t} \quad (8)$$

ここで、 \mathbf{u}_t は N_t 個の観測された画像上の特徴点位置である。具体的には、 $\mathbf{u}_t = [u_1, v_1, u_2, v_2, \dots, u_{N_t}, v_{N_t}]^t$ である。また、 $\mathbf{h}(\bar{\mu}_t^i)$ は対応する既知の空間点を画像面上に投影した位置であり、 \mathbf{u}_t と同じ規則で点位置の座標を格納している。 \mathbf{h} では、選択された歩行形態と予測された歩行状態 $\bar{\mu}_t^i$ によって、3節で述べたモデルに基づいて算出されるユーザの腰の位置と大腿部の姿勢、及び既知のカメラ設置位置から、注目する空間点の画像投影点を計算する。予測された特徴点の位置と観測点の位置の誤差は、実際の歩行が保持した各歩行モデル通りではないことに起因するものであり、ガウス分布を用いて表わすものとした。本論文の実験では、平均 ν と分散 σ^2 をそれぞれ0, 10に設定した。

リサンプリングのステップでは、尤度が高い候補のみを残し、他の候補を切り捨てることで、状態候補のフィルタリングを行う。残された候補に対して、次節で述べられる状態更新が適用される。なお、文字表記の便宜上、状態候補の番号付けはこの操作後に初期化されるものとする。

5.5 歩行状態の更新

前節の操作で取得された予測誤差は、センサの誤差やシステムが保持する歩行動作と観測された動作間の誤差に起因するものである。このような誤差を、観測されたセンサデータを用いたパラメータの更新操作によって除去する。観測モデルは、次式で表わされる。

$$\mathbf{u}_t = \mathbf{h}(\bar{\mu}_t) + \delta_t. \quad (9)$$

このとき、更新操作は拡張カルマンフィルタの枠組みに基づき、次式によって実施される。

$$\bar{\Sigma}_t = \mathbf{G}_t \Sigma_{t-1} \mathbf{G}_t^T + \mathbf{P}_t \quad (10)$$

$$\mathbf{K}_t = \bar{\Sigma}_t \mathbf{H}_t^T (\mathbf{H}_t \bar{\Sigma}_t \mathbf{H}_t^T + \mathbf{Q}_t)^{-1} \quad (11)$$

$$\tilde{\boldsymbol{\mu}}_t = \bar{\boldsymbol{\mu}}_t + \mathbf{K}_t (\mathbf{u}_t - \mathbf{h}(\bar{\boldsymbol{\mu}}_t)). \quad (12)$$

ここで、 \mathbf{G}_t は 5.3 節で述べられた予測モデル \mathbf{g} のヤコビアンである。また、 \mathbf{H}_t は 5.4 節で述べられた観測モデル \mathbf{h} のヤコビアンである。さらに、 $\mathbf{P}_t, \mathbf{Q}_t, \Sigma_t$ は、プロセスノイズと観測モデル、及びシステムの共分散行列である。

状態候補のうち、前節で計算された尤度が最も高いものを現時刻の歩行状態として選択する。選択された状態候補の歩行形態、更新された歩行状態を用いて、全身の動作姿勢を取得する。

前節までに述べた予測と更新は、拡張カルマンフィルタの枠組みを踏襲するものである。一方で、本手法では、状態予測モデル \mathbf{g} が形態の遷移とともに切り替わる。これは、通常の拡張カルマンフィルタとは異なる仕様であるが、実験の結果、大きな問題とはならないことを確認している。これは、拡張カルマンフィルタ自体のロバスト性ととも、形態の遷移時に起こるパラメータの変化が滑らかであるためであると考えられる。

最後に、提案手法の計算量について述べる。本手法の計算時間は、ほぼ本節の処理に依存する。大まかに見積もると、拡張カルマンフィルタを候補数分繰り返すだけの処理が必要となる。なお、本手法の場合、観測ベクトルが状態変数ベクトルよりも次元数が高いため、逆行列の演算に依存し、同フィルタの更新に要する計算量は、観測された特徴点数を N_t 個とすると、 $O((2N_t)^{2.4})$ 程度になると見積もることができる [12]。このことから、候補数を M_t 個とすると、全体の計算量は $O(M_t(2N_t)^{2.4})$ となる。このように、本手法は、最新の CPU を駆使することで、十分に実時間で実行可能な量であると考えられる。

6 実験

6.1 歩行動作ライブラリを用いたシミュレーション実験

本実験では、Carnegie Mellon University Graphics Lab のモーションキャプチャライブラリ [20] を用いた。使用した歩行は 5 種類である。ここでは、2 種類の “slow walk”, 2 種類の “walk”, 1 種類の “run” を使用した。

まず、3 節で述べた手法を用いて、5 種類全ての動作データを歩行特徴量 \mathbf{f} に変換した。次に、4 節で述べた手法を用いて、同特徴量を分類した。分類結果を

図 5 に示す。クラスタリングの結果、同じクラスに属した射影点は同じ記号で表している。本実験では、3 種類に分類された。3 種類は、“slow walk”, “walk”, “run” であった。このように、得られた分類結果が、予め視覚的に異なるものとして判断された形態の分類と一致していることが確認された。また、実験では、歩行遷移を図中の黒線で接続する関係として定義した。さらに、システムが保持する歩行データとして、“run” は、分類に使用したデータを登録した。“walk” と “slow walk” は、分類に使用した 2 種類から 1 種類を選択し、登録した。

本実験では、3 つの歩行形態をつなげることで、入力 of 歩行動作を生成した。具体的には、“slow walk”, “walk”, “run” の順で動作が行われるものを生成した。また、“slow walk” と “walk” には、システムに登録されなかったデータを用いた。さらに、生成した動作に基づいて、ランドマークの世界座標をカメラ画像面に再投影したものを、センサ入力として用いた。本実験では、右足の大腿部の外側にカメラを設置した。なお、本実験では 10 点以上のランドマークが視野内に安定して捉えられる環境を設計した。

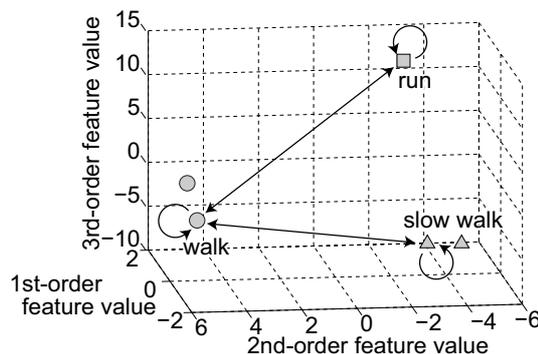


図 5 歩行動作ライブラリを用いた分類結果と遷移図

Fig. 5 Classification result and transition diagram.

推定結果を図 6 に示す。図で最上段に描画された対象が入力データ、第 2 段目の対象が推定結果である。各画像は、同じ時刻のものが縦に配置されている。また、図 7 に、関節位置の誤差平均の時間遷移を示す。図より、誤差は 10cm 以内に抑えられており、精度が高いことが確認できる。本実験では、入力 of 歩行動作と保持する動作が異なっていたり、形態が遷移するタイミングが未知であるにも関わらず、推定時の形態の遷移や状態の更新は正常に達成されており、オリジナルの歩行動作を復元することに成功している。

次に、提案手法を段階的に推定を行う手法と比較した。この段階的推定手法は、カメラの位置姿勢推定、歩行形態識別、動作姿勢の決定、座標系の位置合わせの 4 ステップを個別に実施するものである。カメラの

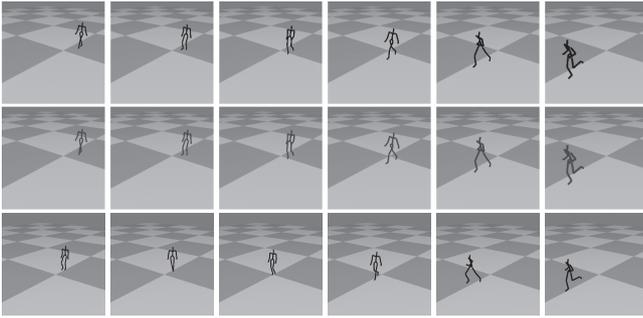


図6 入力歩行動作と推定歩行動作の比較. 上段: 入力動作, 中段: 提案手法の推定結果, 下段: 比較手法の推定結果.

Fig.6 The input human gait and the estimated gait.

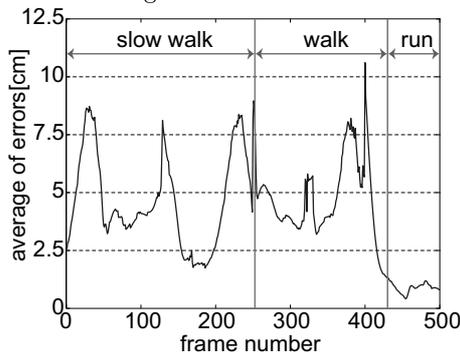


図7 推定動作から復元された関節位置の平均誤差 (提案手法)

Fig.7 Average errors in joint positions.

位置姿勢推定では、画像位置からカメラの3次元位置姿勢を推定する。歩行形態識別では、特徴量を用いて、歩行形態の識別を行う。具体的には、まず、推定されたカメラの位置姿勢を、観測した絶対座標系に依存しない正規化された特徴量に変換し、同特徴量に基づいて、保持している様々な歩行形態のうち、現時刻における動作がどれかを識別する。動作姿勢の決定では、識別された形態における状態を推定する。座標系の位置合わせでは、システムが保持している動作を記述している座標系とカメラの位置姿勢推定における座標系の間の変換を推定する。

このように、各ステップに分離することによって、カメラ位置姿勢のための手法 [21] やサポートベクターマシン (SVM)[22] などのテクニックを利用できる。ただし、提案手法が全てのステップを分割することなく、統合的に取り込むことで、観測情報であるカメラの画像の情報を最終的な推定まで利用し、同観測情報に最も整合する出力を推定するのに対して、このような段階的な推定手法では、分解された各ステップにおいて生じる誤差が蓄積する問題があると考えられる。

具体的には、比較のための段階的推定手法は次の通りに設計した。まず、カメラの位置姿勢推定に関しては、PnP問題を解いた。次に、歩行形態識別における

特徴量として、過去 F フレームのカメラの位置姿勢 $\mathbf{R}_f/\mathbf{T}_f$ を用いた。このうち、保持している動作の座標系と観測時の世界座標系に関して、重力方向の軸は一致していると仮定して、並進成分は Z 軸の成分のみ用いた。また、回転行列は、現時刻の姿勢に対する相対的な姿勢変換 $\tilde{\mathbf{R}}_{f-j} = \mathbf{R}_{f-j}\mathbf{R}_f^t$ を過去 $F-1$ フレームに対して算出し、同行列を表す3パラメータを用いた。回転行列のベクトル変換には、Rodriguesの公式を用いた。従って、特徴量の次元数は、 $(F-1)*3+F$ となる。本実験では $F=10$ とした。この特徴量を用いた歩行形態識別には、マルチクラスSVMを利用した。本実装では、同SVMのツールとして [23] を用いた。カーネル関数には Radial Basis Function を用いた。動作姿勢の決定には、識別された動作から算出される Z 軸のカメラの位置が、推定された位置と最も近いものを探索した。最後に、取得した動作姿勢に基づいた場合のカメラの位置姿勢を算出し、推定された位置姿勢と整合するように、動作姿勢全体の座標変換を推定した。並進成分は水平面の移動のみを考慮し、 X 軸と Y 軸方向のみを合わせた。また、回転成分は、 Z 軸周りの回転のみを合わせた。なお、同回転の補正によって、正確に値が一致することはないため、非線形最適化によって同パラメータを算出した。

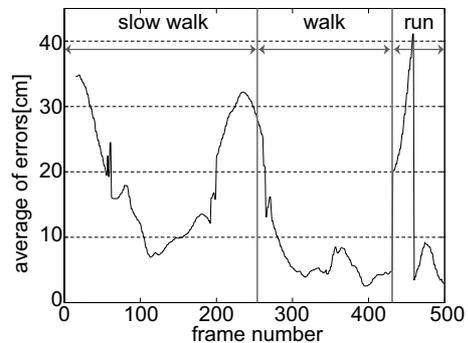


図8 推定動作から復元された関節位置の平均誤差 (比較手法)

Fig.8 Average errors in joint positions.

上記の実験と同様の動作を入力した場合の動作推定の結果を図6の最下段に示す。また、関節位置の平均誤差を図8に示す。なお、歩行形態識別における識別率は69.8%であった。これらの結果に示されるように、段階的手法では精度の高い身体姿勢の推定が困難であることが分かった。

6.2 実環境下での動作実験

本節では、提案手法の実環境下での動作実験について述べる。実験環境の写真を図9に示す。今回は、右足大腿部の外側にカメラを装着した。使用したカメラはBasler社のpiA640-210gc(センササイズ:1/3inch)である。本実験では、解像度 400×400 、フレームレート 250fps(露光時間 4ms) で撮像を行った。このとき使

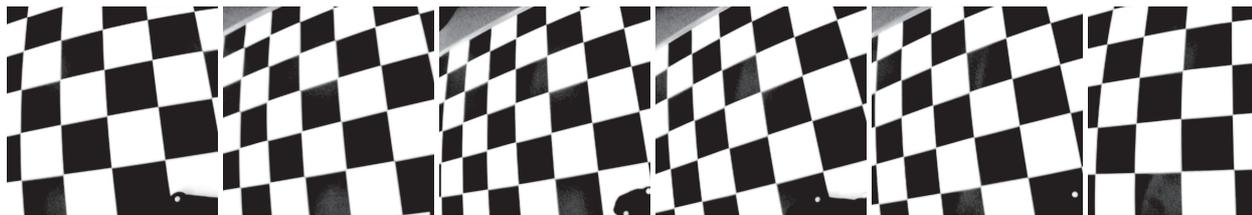


図 11 歩行中にウェアラブルカメラが撮像した画像例
 Fig. 11 Input images captured by the camera mounted on the user's leg.

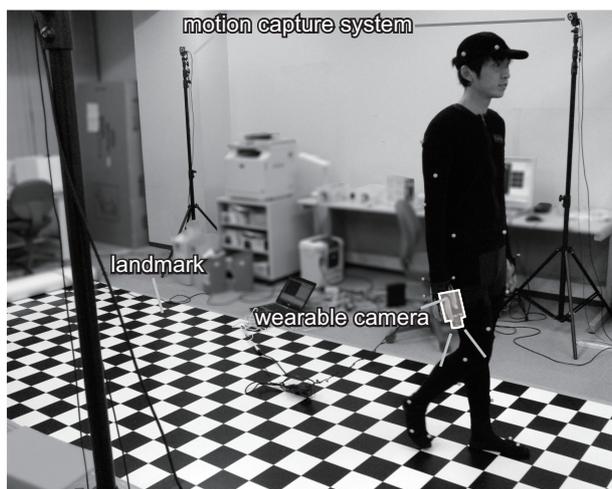


図 9 実験環境の写真
 Fig. 9 Photograph of the experimental environment.



図 10 ウェアラブルカメラの設置状況
 Fig. 10 Installation of the wearable camera.

用したレンズは焦点距離 3.5mm である。なお、カメラの視野に足が映り込まず、カメラと脚の間にはスペーサを設置した。設置状況の詳細を図 10 に示す。カメラは固定設置されており、本実験の歩行中にずれが生じないことを確認した。

また、チェッカーパターンを地面に設置し、地面のランドマークの世界座標が既知であると仮定した。チェッカーパターンの各正方形の 1 辺の長さは 15.15cm である。

なお、5.4 節で述べたカメラの観測モデル h では、レンズ歪みが考慮されていない。そこで、入力画像は、歪み補正を適用した上で利用した。例として、直進歩行をしたときにウェアラブルカメラによって撮影された画像を図 11 に示す。ここでは、全画像から 50 フレーム毎 (200ms) の画像を示している。

まず、準備としてユーザのサンプル歩行動作を収集した。収集時に用いたモーションキャプチャシステムは、NaturalPoint 社の OptiTrack である。サンプルされたデータを周波数解析し、特徴量に変換した。同特徴量を分類した結果を図 13 に示す。分類の結果、6 つの形態に分類された。分類された各動作は、“walk”, “left-turning walk”, “right-turning walk” の 3 種類の低速動作と中速動作である。それぞれの動作について、3 試行が取得されている。システムには、クラスタの中心に最も近いものを登録した。図 12 に、6 形態の動作に関して、腰の高さの時系列挙動を例として示す。

以上のセットアップの下、対象者が歩行している際の推定を行った。推定では、直進して歩いている途中で、左折し、再び直進する歩行を動作入力とした。また、低速での動作を意識して歩行することを指示した。撮像画像での特徴点の検出には Harris 検出器を用いた。本実験では、推定を行った各画像における観測点数は、平均 16 点、標準偏差 11 であった。推定の結果を図 14 に示す。図では、歩行の様子を捉えた写真と、同時刻の推定結果を比較している。

また、観測された特徴点の画像上の位置と、対応する空間点を画像上に投影した位置の平均誤差の時間推移を図 15 に示す。図に示されるように、全域においてほぼ 10 画素以下の誤差に抑えられていることがわかる。また、特徴点が検出されなかったり、保持した歩行データと大きく異なる動作をした瞬間があった場合にも、推定が発散せず、安定していることがわかる。

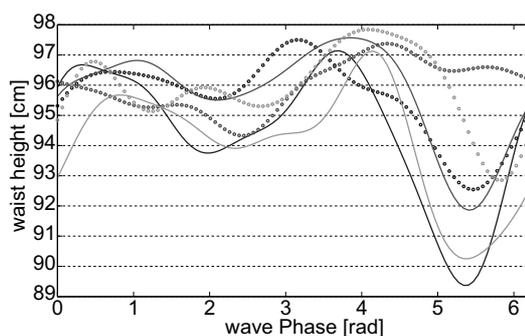


図 12 6 つの異なる形態における腰の高さの挙動
 Fig. 12 Sample waveform of six different motions.

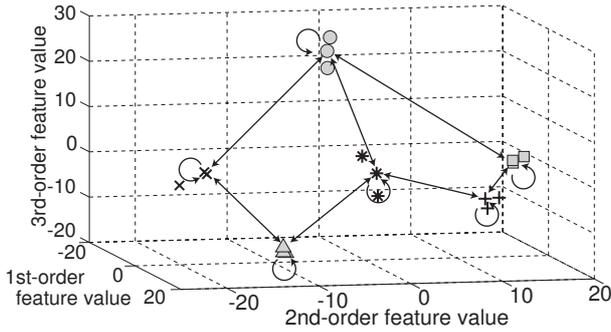


図 13 実環境下実験での分類結果と遷移図
Fig. 13 The classified result using the actual sensor data.

次に、図 16 に歩行形態の時間変化を示す。本実験では、もし対象者がシステムが保持する理想的な歩行動作と同じ動きをしたと仮定すると、観測された歩行は 3 種類の形態で記述可能であると考えられる。しかし、推定では遷移にばらつきが見られる。これは、歩行動作をある短い瞬間で捉えた場合に、似た動作が他の形態にも含まれている場合に生じるものである。推定される姿勢に異常がなければ、このような形態の遷移は致命的な問題ではない。例えば、式 (6) で表現される尤度による制御だけでなく、歩行形態の時間変化のダイナミクスモデルを導入することで、推定精度を高めることも可能であると考えられる。

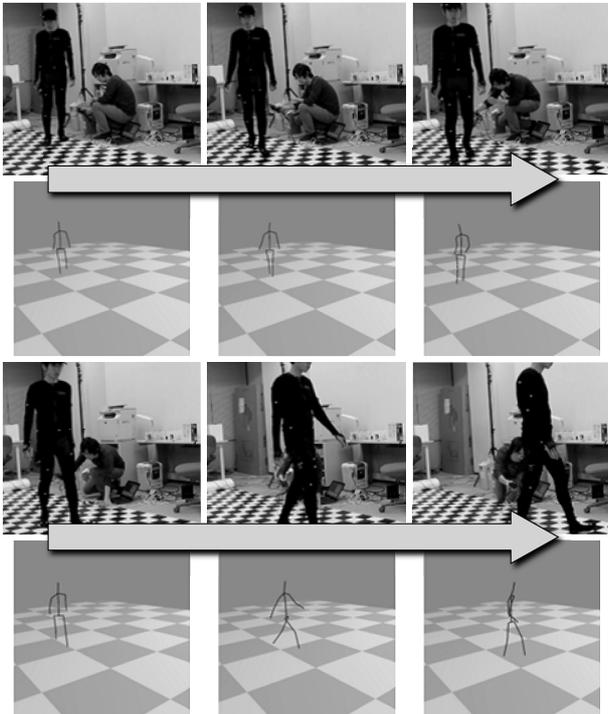


図 14 歩行動作の推定結果
Fig. 14 Estimation results of walking motion.

7 考察

提案手法に基づく歩行動作推定は、ウェアラブルカメラ 1 つのみで構成されているため、屋外環境下でも

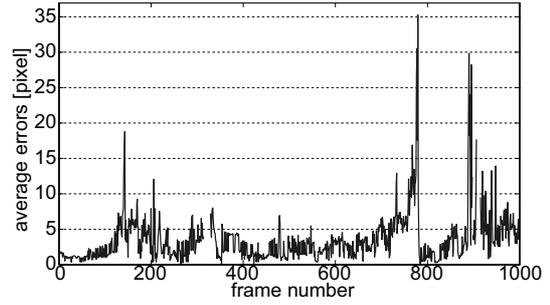


図 15 再投影誤差の時系列変化。推定された歩行姿勢に基づき、環境中の空間点をカメラの画像面上に投影し、実際の観測点との誤差を評価した。

Fig. 15 Time variation of errors evaluated by projecting the spatial points onto the image plane based on the estimated pose.

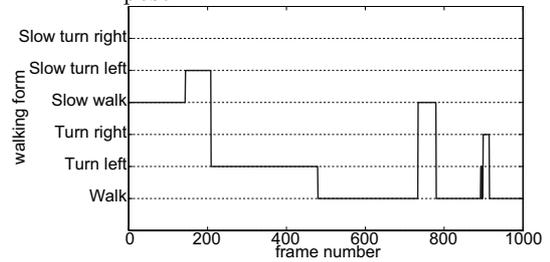


図 16 歩行形態の時系列変化
Fig. 16 Time variation of the walking forms.

場所を選ばずに実施できる点で有効であると考えられる。歩行動作に注目した手法となっているが、日常生活下において同動作のモーションキャプチャを実現できる点は応用展開においても価値が高いと考えられる。また、カメラを用いているため、従来の加速度センサやジャイロセンサなどのモジュールに比べて、移動に伴う蓄積誤差を回避できるメリットがある。

一方で、人間の動作は高速であるため、ウェアラブルなセンサとしてカメラを導入する場合は、高いフレームレートでの撮像が必要となることが分かった。この点は、文献 [14] でも指摘されており、走行などの動作を伴うときには必須の機能になると考えられる。本論文ではこの点に対して、250fps のフレームレートで実験を実施し、この問題を回避している。この程度の速度のフレームレートを備えたカメラ機能は、コンパクトデジタルカメラにも搭載されており [24]、ウェアラブルセンサとしても技術的に導入が可能であると考えられる。

本手法は、歩行モデルを組み込むことで、動作計測を不十分なセンサ情報からオリジナルの情報を復元する問題として取り扱っている。この点は、歩行中の下肢の動きを、片方の足に装着された単一のカメラだけで取り込むというシステム構成の単純化の面で利点を生み出したが、計算量が大きくなる問題がある。この点については、携帯情報端末のニーズの拡大とともに、

組込みシステムの計算能力も飛躍的に上がってきており、技術的なボトルネックにはならないと考えられる。

今回の手法では、予め環境の3次元構造が既知であることを想定していた。例えば、特定の場所については、予め計測された環境構造の情報を、その場にいるユーザのセンシングシステムに提供することで、このような前提条件を満たすことが可能になると期待されるが、場所に依存しないサービスを想定した場合には問題になると考えられる。この点については、1節でも述べたように、SLAMやSfMなどのような3次元の環境構造マップの作成とカメラ位置姿勢を同時推定する技術を組み込むことで回避できると考えられる。本論文では、独自の技術となる歩行動作推定部のみについて提案と実証を行ったが、歩行動作の状況のみを記述した変数セットに、ランドマークの位置を組み込み、包括的に推定することでこのような機能も付加的に取り込むことができると考えられる。

また、本手法では、予めシステムが各個人の歩行動作を登録することを前提としている。歩行動作は人によって特有のものであることが知られており、歩容認証として個人認証にも使われている [25, 26]。このことから、本手法のように動作の姿勢を推定する場合には、個人に特化する方式をとる必要があると考えられる。一方で、歩行動作を登録するためには、3次元動作を記録できる環境が整備されている必要がある。これについては、3次元計測技術の発展により、コンシューマ向けにモーションキャプチャを家庭で実施できるシステムもあり [27]、本技術のメインターゲットとなるユーザへの導入が困難になるレベルでは無いと考えられる。

次に、手法の構成について考察を述べる。提案手法では複数の歩行動作を切り替えて用いているが、システムが保持する歩行動作を1種類にして、状態遷移を行わずに推定の安定化を図る手法も考えられる。この場合にも、近似的な歩行動作を達成することはできる可能性はあるが、本論文では正しい歩行姿勢を推定することを目的としており、1種類に限定する方策は同目的と合致しない。

次に、実験結果について述べる。動作ライブラリを用いた実験では、誤差が10cm以内に抑えられており、高い精度を達成することが示されている。また、実環境下での実験も同様に推定に成功していることが確認された。それぞれの実験結果では、各時刻で推定された歩行形態とその状態から、全身の動作がモデルに基づいて復元されている。ここでは下肢のみが一致することを目標としていたが、歩行のように安定的な周期動作の場合、上肢に関しても推定動作と実際の動作に一致が見られることが分かった。

今回は、周期的な歩行についてのみ検証を行ったが、転ぶなどの非周期的な動作も、保持するモデルを拡張することで対応することが次の課題である。また、平地だけでなく、階段を登るなどの動作も取り込んでいく必要がある。この場合、地形とリンクした歩行動作推定が必要となるため、前述の単眼カメラによる周囲環境の3次元復元が、推定の安定化のために効果が大きいと期待される。

8 まとめ

幅広い応用分野において、日常環境下でいつでもどこでもモーションキャプチャを可能とする技術の重要性が高まっている。しかし、これまでのモーションキャプチャは、設置カメラを利用するものが主流であり、複雑な環境下では遂行が困難であった。また、ウェアラブルセンサを利用するものは、センサを冗長につける構成となっていた。そこで、本論文では、単一のウェアラブルカメラのみを用いた歩行動作の取得技術を提案した。歩行動作に特化した場合でも応用展開の可能性は高く、本手法は強力であると考えられる。

提案手法は、ユーザの動作によって生じるカメラ内での変動のみを用いて、身体姿勢と大域的な位置の両方を推定することができる。また、多様な歩行動作をコンパクトなパラメータ表現によって記述する方法を導入し、その分類や推定が有効に機能するように設計されている。さらに、遷移図による歩行形態の切り替えと拡張カルマンフィルタによる状態推定を統合し、歩行を階層的に復元する手法となっている。実験の結果、提案手法が実環境下においても正常に動作することを確認した。

参考文献

- [1] S. Rosenthal and J. Pella. The process of motion capture: Dealing with the data. *Computer Animation and Simulation*, 97:3-18, 1997.
- [2] VICON. <http://www.vicon.com/>.
- [3] NaturalPoint, Inc. OptiTrack. <http://www.naturalpoint.com/optitrack/>.
- [4] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81:231-268, 2001.
- [5] 新居英明, N. L. Htat, 稲見昌彦, A. D. Cheok. 直射日光対応型ウェアラブル光学モーションキャプチャ装置. 第16回日本バーチャルリアリティ学会大会論文集, pages 17-18, 2011.
- [6] D. Vlastic, R. Adelsberger, G. Vannucci, J. Barnwell, M. Gross, W. Matusik, and J. Popovic. Practical motion capture in everyday surroundings. *ACM Transaction on Graphics*, 26:35, 2007.
- [7] Xsens. MVN. <http://www.xsens.com/>.
- [8] C. M. N. Brigante, N. Abbate, A. Basile, A. Carmelo Faulisi, and S. Sessa. Towards miniaturization of a mems-based wearable motion capture system. *IEEE Transactions on Industrial Electronics*, 58(8):3234-3241, 2011.

- [9] J. Frank, S. Mannor, and D. Precup. Activity and gait recognition with time-delay embeddings. In *Proc. 24th AAAI Conference on Artificial Intelligence*, 2010.
- [10] A. Hamaguchi, M. Kanbara, and N. Yokoya. User localization using wearable electromagnetic tracker and orientation sensor. In *Proc. IEEE International Symposium on Wearable Computers*, pages 55–58, 2006.
- [11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [12] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents series)*. The MIT Press, 2005.
- [13] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.
- [14] T. Shiratori, H. S. Park, L. Sigal, Y. Sheikh, and J. K. Hodgins. Motion capture from body-mounted cameras. *ACM Transactions on Graphics*, 30(4):Article No.31, 2011.
- [15] 下坂 正倫, 祢次金 佑, 森 武俊, 佐藤 知正. 動作の時間的連続性を考慮した boosting アルゴリズムに基づくオンライン日常動作認識. 日本ロボット学会誌, 25(6):906–912, 2007.
- [16] 高野 渉, 今川 洋尚, クリッチ ダナ, 中村 仁彦. クリストルボール：運動の記号推論を通じた未来の行動予測. 日本ロボット学会誌, 29(8):745–751, 2011.
- [17] M. Garcia, A. Chatterjee, A. Ruina, and M. Coleman. The simplest walking model: Stability, complexity, and scaling. *ASME Journal of Biomechanical Engineering*, 120:281–288, 1998.
- [18] M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *Proc. ACM SIGGRAPH*, pages 91–96, 1995.
- [19] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *Journal of the Royal Statistical Society*, 28:100–108, 1979.
- [20] CMU Graphics Lab Motion Capture Database. <http://mocap.cs.cmu.edu/>.
- [21] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 1–10, 2007.
- [22] C. M. Bishop. *パターン認識と機械学習 - ベイズ理論による統計的予測*. Springer, 2007.
- [23] SVM^{multiclass}. http://svmlight.joachims.org/svm_multiclass.html.
- [24] CASIO Computer Corp. EXILIM. <http://casio.jp/dc/>.
- [25] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.
- [26] 榎原靖, 佐川立昌, 向川康博, 越後富夫, 八木康史. 周波数領域における方向変換モデルを用いた歩容認証. 情報処理学会論文誌. コンピュータビジョンとイメージメディア, 48(1):78–87, 2007.
- [27] Microsoft Corp. Kinect for Xbox 360. <http://www.xbox.com/>.

[著者紹介]

渡辺 義浩 (正会員)



2007年東京大学大学院情報理工学系研究科博士課程修了。2004-2007年日本学術振興会特別研究員。2007年同大学院リサーチフェロー。2008年同研究科特任助教。2011年同研究科助教，現在に至る。高速ビジョン，超並列プロセッサ，実時間画像センシング，映像メディア，デジタルアーカイブ等に関する研究に従事。博士(情報理工学)。2008年度日本ロボット学会論文賞。2008年度3次元画像コンファレンス優秀論文賞。2009年度エリクソン・ヤング・サイエンティスト・アワード。2011年度画像センシングシンポジウム優秀学術賞。2011年度日本バーチャルリアリティ学会論文賞。

畑中 哲生



2010年東京大学大学院情報理工学系研究科修士課程修了。同年，コナミ株式会社入社。

小室 孝 (正会員)



2001年東京大学大学院工学系研究科計数工学専攻博士課程修了。同年科学技術振興事業団研究員。2002年東京大学大学院情報理工学系研究科システム情報学専攻助手，2005年同講師。2011年埼玉大学大学院理工学研究科数理電子情報部門准教授。高速画像センシング，三次元ユーザーインターフェース等に関する研究に従事。博士(工学)。

石川 正俊 (正会員)



1979年東京大学大学院修士課程修了。同年通産省工業技術院製品科学研究所に入所。1989年東京大学工学部計数工学科助教。現在，東京大学情報理工学系研究科創造情報学専攻教授。2002年度東京大学総長特任補佐，2004年東京大学副学長，2005年東京大学副学長(2006年まで)。超並列・超高速ビジョン，センサフュージョン，メタパーセプション等に関する研究に従事。工学博士。1998年，2001年，2008年，2010年日本ロボット学会論文賞，1984年，2010年計測自動制御学会論文賞，2004年度同学会技術賞・友田賞，1990年応用物理学会光学論文賞，2005年同学会光・電子集積技術業績賞，2011年度日本バーチャルリアリティ学会論文賞，受賞。2011年紫綬褒章受章。