三次元人体 CG モデルを用いた多眼視マーカレス モーションキャプチャ*

老 田 隆^{*1}, 堀 純一郎^{*2}
高 橋 和 彦^{*3}, 橋 本 雅 文^{*3}

Markerless Human Motion Capture Using Multiple Images of 3D Articulated Human CG Model

Takashi OIDA, Junichiro HORI,
Kazuhiko TAKAHASHI^{*4} and Masafumi HASHIMOTO

^{*4} Department of Science and Engineering, Doshisha University,
1-3 Miyakodani, Tatara, Kyotanabe-shi, Kyoto, 610-0321 Japan

This paper proposes markerless human motion capture using a 3D articulated CG model of the human body. The method of estimating human body posture is based on 2D matching between human silhouettes extracted from camera images and model silhouettes projected onto the corresponding camera planes in virtual space. Candidates for human model silhouette are generated using a Monte Carlo filter-based algorithm, and the normalised core-weighted XOR distance is introduced to calculate the likelihood rate between silhouettes. Experimental results show the feasibility and effectiveness of the proposed method for achieving markerless human motion capture.

Key Words: Motion Capture, Human CG Model, Silhouette Image, Monte Carlo Filter, Multi-Camera System

1. 緒 言

モーションキャプチャは、人物の姿勢や表情を動的に計測する技術であり、スポーツや医療の分野において、身体動作を解析するためのデータを収集したり、映画やゲームなどのエンタテインメントの分野において、人間らしい動きを再現する三次元キャラクターを制作したりする目的で、広範に利用されている⁽¹⁾。現在実用されているモーションキャプチャは、計測の対象者にセンサやマーカ等の機器を装着する方式が殆どであり、これらは計測精度が高い反面、システムが高額になる、機器の装着に時間を要する、身体に拘束を与えるため自然な動作を制限したりする等の欠点がある。こうした背景から、近年は機器を装着することなく身体動作を計測するマーカレスモーションキャプチャへの需要が高まっており、これを実現する方法の一つとして、カメラにより取得する画像に対して画像処理のみで人体の姿勢や動きのパラメータを計測する手法の研究が盛んに行われている^{(2)~(5)}。

画像処理によるマーカレスモーションキャプチャにおいては、カメラからの観測画像より得られる特徴量に対して、あらかじめ設定した人体モデルを照合することで、モデルを定義するパラメータを推定し、そのパラメータから人体の姿勢や動きのパラメータを決定する手法がよく用いられている⁽⁶⁾。特に近年は、計算機の能力が飛躍的に向上したため、観測画像から得られる特徴量としてマルチカメラによる多視点画像や距離画像を用いて人物の三次元形状を復元したボリュームデータを使用し、これに対して人体部位の形状や骨格を立体で表現したモデルをマッチングする手法が多数提案されている^{(7)~(10)}。しかし、三次元形状の復元には、それぞれのカメラの位置や姿勢からなる外部パラメータを知る必要があり、チェッカーボード等の参照物体を用いたキャリブレーションによりカメラパラメータの推定が必要である。このカメラキャリブレーションの作業は、時間を要す上に、キャリブレーション後もカメラの完全な固定が難しい等の理由から、少しの衝撃でも外部パラメータに変化をもたらし、復元結果に異常を来す問題がある⁽¹¹⁾。

本論文では、多視点画像を用いたマーカレスモーションキャプチャの開発を目的として、モーションキャプチャの対象となる人物の三次元形状をコンピュータ

* 原稿受付 2010年4月7日。

^{*1} (株)リコー(☎104-8222 東京都中央区銀座8-13-1)[前:同志社大学大学院工学研究科]。

^{*2} (株)NTTデータ(☎135-6033 東京都江東区豊洲3-3-3)[前:同志社大学工学部]。

^{*3} 正員,同志社大学理工学部(☎610-0321 京田辺市多々羅都谷1-3)。

E-mail: katakaha@mail.doshisha.ac.jp

Fig. 1 Configuration of actual and virtual multi-camera systems for human motion capture.

ラフィックス (CG) により近似した三次元人体 CG モデルを仮想空間に構築し、実空間における人物の観測情報と仮想空間におけるモデルの観測情報についてマッチングを行うモーションキャプチャアルゴリズムを提案する。提案する手法では、実空間で人物の観測画像を取得する視点に対応したカメラ平面にモデルを投影して得られるシルエット情報を用いて二次元マッチング処理を行う。そのため、観測画像からの人物の三次元形状復元が不必要であり、カメラキャリブレーションにおける外部パラメータの問題を回避することができる。提案するマーカレスモーションキャプチャの実現可能性を検討するため、マルチカメラシステムにより取得した実画像シーケンスに対して人物の姿勢推定を行い、その有効性を示す。

2. モーションキャプチャのアルゴリズム

2.1 アルゴリズムの概要 本論文において提案するモーションキャプチャの概念図を図 1 に示す。このモーションキャプチャシステムは、実空間における対象人物の撮影系と、仮想空間における三次元人体 CG モデルの撮影系で構成される。これにより、2つの撮影系における観測画像中心のモデルマッチングを行い、実空間上の人物姿勢に一致する仮想空間上の三次元人体 CG モデルを求める。具体的には、実空間の撮影系において取得する対象人物の画像シーケンスに対して、仮想空間の撮影系において実空間と同一の視点に投影する三次元人体 CG モデルの画像をフレームごとにマッチングするもので、マッチングにはそれぞれの画像から抽出する人物あるいはモデルのシルエット領域の情報（以下、シルエット画像）を特徴量として用いる。また、シルエット画像は異なる複数視点より取得し、各視点におけるマッチング結果を総合的に扱うことで、奥行き情報が得られないことによるオクルージョンの課題に対応する。

2.2 三次元人体 CG モデル モデルマッチングに使用する人体モデルには、図 2 に示すような CG の

Fig. 2 3D articulated CG human model.

三次元人体モデルを用いる。この人体 CG モデルは、10 の関節を持ち、頭、胸、腰、そして左右の上腕、前腕、大腿、下腿の 11 の部位から構成される。各部位は別々の CG モデルとして作成され、それぞれ接続する関節において基準姿勢と可動方向からなる回転軸を定義している。各関節における回転軸は、図 2 において、CG の三次元座標（左手系）に従って、軸番号 1, 2, 3 により示している。そして、これらの部位モデルを基準姿勢から各回転軸に関して回転し、接続関節へ平行移動することで、全身の人体 CG モデルを再現している。従って人体 CG モデルは、各関節における部位モデルの接続状態により姿勢が定義され、これをパラメータとして自動生成される。尚、ここで作成する人体 CG モデルは、シルエット画像を用いたモデルマッチングを行う理由から、対象人物の表面形状を近似する必要があるため、あらかじめ対象人物の形状を計測した情報をもとに、三次元 CG モデリングソフトウェアを用いて作成する。

2.3 モンテカルロフィルタによるモデル姿勢の予測制御 対象の人物姿勢に一致する人体 CG モデルを求めるために、逐次ランダムな姿勢の人体 CG モデルを複数生成し、その中より最も姿勢の近い人体 CG モデルを推定結果として決定するモンテカルロ法的な姿勢推定を行う。しかし、無数のパターンが存在する人間の姿勢に対して、人体 CG モデルをランダムに生成するだけでは、対象の人物姿勢に一致する人体 CG モデルを獲得することは困難であり、非効率的である。そこで、人物姿勢の運動状態を定義する状態空間モデルを導入し、前フレームにおける姿勢情報を活用したモンテカルロフィルタ⁽¹²⁾(MCF)による予測制御を行って、適切な人体 CG モデルを生成する。

式 (1), (2) に導入する状態空間モデルを示す。これは、人体 CG モデルの姿勢を定義するパラメータである関節の回転軸に関する回転量を状態変数として関節

ごとに定義したものである。

$$\mathbf{x}_i(t+1) = \mathbf{F}_i \mathbf{x}_i(t) + \mathbf{G}_i \mathbf{v}_i(t) \quad (1)$$

$$\mathbf{y}_i(t) = \mathbf{H}_i \mathbf{x}_i(t) + \mathbf{w}_i(t) \quad (2)$$

ここで、 $\mathbf{x}_i$ は状態ベクトル、 $\mathbf{F}_i$ 、 $\mathbf{G}_i$ はシステム行列、 $\mathbf{v}_i$ はシステムノイズベクトル、そして、 $\mathbf{y}_i$ は観測ベクトル、 $\mathbf{H}_i$ は観測行列、 $\mathbf{w}_i$ は観測ノイズベクトルである。また、 t はサンプリング時刻であり、本提案手法の場合、取得した画像シーケンスのフレーム番号を意味する。そして、 $i(=1, 2, \dots, 10)$ は関節を識別する番号であり、1 から順に腰、首、左肩、左肘、右肩、右肘、左股、左膝、右股、右膝に対応している。また、図 2 に示すように、関節により定義される回転軸の数が異なるため、状態ベクトルと観測ベクトルは関節により異なっている。3 軸の回転軸を持つ関節 $i=1, 2, 3, 5, 7, 9$ については $\mathbf{x}_i(t) = [x_1(t), x_1(t-1), x_2(t), x_2(t-1), x_3(t), x_3(t-1)]^T$ 、 $\mathbf{y}_i(t) = [y_1(t), y_2(t), y_3(t)]^T$ であり、2 軸の回転軸を持つ関節 $i=4, 6$ については $\mathbf{x}_i(t) = [x_1(t), x_1(t-1), x_2(t), x_2(t-1)]^T$ 、 $\mathbf{y}_i(t) = [y_1(t), y_2(t)]^T$ 、1 軸のみの回転軸を持つ関節 $i=8, 10$ については $\mathbf{x}_i(t) = [x_1(t), x_1(t-1)]^T$ 、 $\mathbf{y}_i(t) = y_1(t)$ である。尚、 $\mathbf{F}_i$ 、 $\mathbf{G}_i$ 、 $\mathbf{H}_i$ については、人間の種々の動作における姿勢の時間的な変化を滑らかなものであると仮定し、状態ベクトルの時間変化が原則として一定、すなわち時間に関する 2 階差分値が 0 となるように定義している⁽¹³⁾。

次に、MCF のアルゴリズムについて説明する。MCF では、適用する状態空間において起こり得る状態ベクトルの実現値を無数に生成する。そして、これらの実現値を状態空間モデルにより遷移し、観測される情報をもとにフィルタリングすることで、実現値の分布を実際の状態ベクトルに近く、信頼できるものに更新していく。ここで、状態ベクトルは人体 CG モデルの姿勢を定義するパラメータであるから、これらの実現値もとに人体 CG モデルを生成することで、逐次適切に分布する人体モデル群が生成できる。なお、実現値は無数に生成する様子から、一般に「粒子」と呼ばれる。本論文においても以降は「粒子」と記す。

以下に、具体的な処理の流れを示す。ただし、ここで示す粒子 $\{\mathbf{p}_i^{(j)}\}_{j=1}^m = \{\mathbf{p}_i^{(1)}, \dots, \mathbf{p}_i^{(m)}\}$ は、関節 i に関する状態遷移後の粒子(予測分布粒子)であり、粒子 $\{\mathbf{q}_i^{(j)}\}_{j=1}^m = \{\mathbf{q}_i^{(1)}, \dots, \mathbf{q}_i^{(m)}\}$ は、関節 i に関するフィルタリング後の粒子(フィルタ分布粒子)で、これは次時刻の状態遷移の基になる。また、システムノイズ、観測ノイズについても同様に、粒子 $\{\mathbf{v}_i^{(j)}\}_{j=1}^m = \{\mathbf{v}_i^{(1)}, \dots, \mathbf{v}_i^{(m)}\}$ 、 $\{\mathbf{w}_i^{(j)}\}_{j=1}^m = \{\mathbf{w}_i^{(1)}, \dots, \mathbf{w}_i^{(m)}\}$ として分布を近似する。

Step 1 $i=1, \dots, 10$ について、状態の初期分布に従

う乱数ベクトルを m 個発生し、初期のフィルタ分布粒子 $\{\mathbf{q}_i^{(j)}(0)\}_{j=1}^m$ とする。

Step 2 $i=1, \dots, 10$ について、システムノイズ分布に従う乱数ベクトルを m 個発生し、システムノイズ粒子 $\{\mathbf{v}_i^{(j)}(t)\}_{j=1}^m$ を生成する。また、観測ノイズ分布に従う乱数ベクトルを m 個発生し、観測ノイズ粒子 $\{\mathbf{w}_i^{(j)}(t)\}_{j=1}^m$ を生成する。

Step 3 $i=1, \dots, 10$ について、状態空間モデルに基づき、フィルタ分布粒子 $\{\mathbf{q}_i^{(j)}(t-1)\}_{j=1}^m$ を遷移し、予測分布粒子 $\{\mathbf{p}_i^{(j)}(t)\}_{j=1}^m$ を得る。

Step 4 予測分布粒子 $\{\mathbf{p}_i^{(j)}(t)\}_{j=1}^m (i=1, \dots, 10)$ を、5 つのグループ粒子 $\{\mathbf{p}_1^{(j)}(t), \mathbf{p}_2^{(j)}(t)\}_{j=1}^m$ 、 $\{\mathbf{p}_3^{(j)}(t), \mathbf{p}_4^{(j)}(t)\}_{j=1}^m$ 、 $\{\mathbf{p}_5^{(j)}(t), \mathbf{p}_6^{(j)}(t)\}_{j=1}^m$ 、 $\{\mathbf{p}_7^{(j)}(t), \mathbf{p}_8^{(j)}(t)\}_{j=1}^m$ 、 $\{\mathbf{p}_9^{(j)}(t), \mathbf{p}_{10}^{(j)}(t)\}_{j=1}^m$ にグルーピングし、それぞれのグループ粒子をもとに、個別部位のみの人体 CG モデルを生成する。これに対し、シルエット画像を用いたモデルマッチング(2.4 節にて詳述)を行うことで個別グループごとの尤度を部位間の関連性を考慮して計算する。腕モデルと脚モデルは胴体モデルに接続する部位であるために、腕粒子 $\{\mathbf{p}_3^{(j)}(t), \mathbf{p}_4^{(j)}(t)\}_{j=1}^m$ 、 $\{\mathbf{p}_5^{(j)}(t), \mathbf{p}_6^{(j)}(t)\}_{j=1}^m$ および脚粒子 $\{\mathbf{p}_7^{(j)}(t), \mathbf{p}_8^{(j)}(t)\}_{j=1}^m$ 、 $\{\mathbf{p}_9^{(j)}(t), \mathbf{p}_{10}^{(j)}(t)\}_{j=1}^m$ の尤度計算を行うためには、予め胴体モデルの配置がされている事が必須となる。そこで、はじめに胴体粒子 $\{\mathbf{p}_1^{(j)}(t), \mathbf{p}_2^{(j)}(t)\}_{j=1}^m$ に関する尤度を計算する。次に、その尤度で上位 $a\%$ の胴体粒子それぞれに対して左腕粒子を組み合わせて胴体と左腕だけを持つ人体モデルを $m' = m \times (m \times a / 100)$ 個構成し、このモデルに対するマッチングで求まる尤度を左腕粒子の尤度とする。同様の処理を右腕粒子、左脚粒子、右脚粒子についても並列に行うことで各粒子の尤度を計算する。次に、尤度が r 位の胴体粒子に対して、左腕粒子、右腕粒子、左脚粒子、右脚粒子の各尤度上位 $a\%$ の粒子をランダムに組み合わせて新たな粒子を m 個生成する。この処理を $r=1, \dots, (m \times a / 100)$ について行い、総合的な予測分布粒子 $\{\mathbf{p}_1^{(j)}(t), \dots, \mathbf{p}_{10}^{(j)}(t)\}_{j=1}^{m'}$ を生成する。

Step 5 総合的な予測分布粒子 $\{\mathbf{p}_1^{(j)}(t), \dots, \mathbf{p}_{10}^{(j)}(t)\}_{j=1}^{m'}$ をもとに、全身の三次元人体 CG モデル姿勢を生成する。これに対し、シルエット画像を用いたモデルマッチングを行うことで最終的な尤度を計算する。最も高い尤度が得られた粒子をもとに生成される人体 CG モデルを時刻 t における対象人物の姿勢推定

結果とする。

Step 6 総合的な予測分布粒子 $\{p_1^{(j)}(t), \dots, p_{10}^{(j)}(t)\}_{j=1}^{m'}$ において、尤度に関して上位 $b\%$ の粒子を尤度に比例する確率で m 個リサンプリングし、各関節についてのフィルタ分布粒子 $\{q_i^{(j)}(t)\}_{j=1}^m (i=1, \dots, 10)$ を得る。

Step 7 Step 2 に戻る。

ここで、**Step 4** においてグループごとに尤度計算を行う理由について説明する。**Step 5** での最終的な尤度計算に用いる総合的な予測分布粒子は、**Step 3** で生成される各関節の予測分布粒子をそのまま組み合わせるのが自然である。しかし、これらの予測分布粒子が存在する状態空間は非常に高次元であるため、単純に予測分布粒子を組み合わせるモデルの姿勢を生成しても、対象人物の姿勢に近いモデルの姿勢を1つも得られない可能性がある。また、**Step 4** において述べたように、各関節パラメータの間には部位モデルの接続関係から関連性が存在しているため、単純な予測分布粒子の組み合わせでは、ある関節パラメータは人物姿勢のものと近い値を表しているが、接続する部位の関節パラメータが大きく異なるために、連動して低い尤度を与える予測分布粒子を生じる場合がある。そこで、粒子型フィルタの最適性に関する特徴を犠牲にすることになるが、連続する部位間の関連性を保持して探索空間を低次元化して探索を容易にするため、提案手法では予測分布粒子を代表的な部位ごとのグループに分けて尤度を求めている。

2.4 シルエット画像を用いたモデルマッチング

モデルマッチングでは、実空間と仮想空間それぞれの撮影系において取得する多視点シルエット画像を用いて、人物と人体CGモデルの尤度を計算する。ここで、人物の n 視点シルエット画像を $S = \{S_1, \dots, S_n\}$ 、人体CGモデルの n 視点シルエット画像を $T = \{T_1, \dots, T_n\}$ (ただし、前景画素値を1、背景画素値を0) とすると、2つの姿勢間の尤度 $\alpha(S, T)$ は次の手順により計算される。

まず、視点ごとにシルエット画像間の正規化 core-weighted XOR 距離 (以下、nwXOR 距離) を計算する。シルエットのマッチングの手法として、排他的論理和 (XOR) に基づく対応画素値の比較がある。これは、2枚のシルエットにおける値の一致していない対応画素の総数が全体に占める割合を誤差として定義する方法であり、ハミング距離の考え方と類似していることから XOR 距離と呼ばれる。しかし、人体形状のシルエットのマッチングにおいては、輪郭線付近における画素の一致・不一致と骨格線付近の画素の一致・不一致

Fig. 3 Example of weighted distance function of silhouette for matching (left to right: input image, silhouette image and weighted distance function).

を同等に扱う XOR 距離をそのまま用いるのは適切ではない。そこで、先行研究⁽¹⁴⁾に倣い、XOR 距離に対して輪郭線からの距離に比例し、かつ骨格線を強調する重みを加えた nwXOR 距離を利用する。nwXOR 距離は、2枚のシルエット画像の対応画素に関する重みつき XOR の平均を距離として定義した値である。ここで、視点 $l (= 1, \dots, n)$ におけるシルエット画像間の nwXOR 距離を $d_l(S_l, T_l)$ とすると、次の式 (3) により求まる。

$$d_l(S_l, T_l) = \frac{1}{|W|} \sum_{k \in W} \{1 - \delta_{S_l(k), T_l(k)}\} g(k) \quad (3)$$

$$g(k) = D(\bar{S}_l) + \beta \frac{D(S_l)}{D(S_l) + D(\bar{S}_{ls})} \quad (4)$$

ここで、 W はシルエットを含む画素の集合であり、 $|W|$ はシルエット画像全体の画素数を意味する。また、 k は注目画素を識別する番号である。 $g(k)$ は画素ごとに定義される重みであり、式 (4) により求まる。式 (4) における $\bar{S}_l$ は S_l の反転画像、 S_{ls} は S_l の骨格線画像であり、関数 D は二値画像の距離変換を示す。また、右辺の第1項目、第2項目はそれぞれ外側領域、内側領域に関する重みを意味し、ここでは両項の値のスケールを正規化するために係数 β を設けている。図3に入力される人物の画像とそれに対するシルエット画像、重み画像の例を示す。重み画像では、重みが大きいほど輝度が高くなるように表示している。重み画像より、nwXOR 距離に基づいたマッチングが、輪郭線から離れるほど画素値の一致を重視し、かつ骨格線上の画素値の一致を重視する手法であることがわかる。

次に、視点ごとに計算した nwXOR 距離を、ガウス関数によりそれぞれ正規化し、それらを相乗することで2つの姿勢間の尤度 $\alpha(S, T)$ を計算する。尤度 $\alpha(S, T)$ を求める式は、次の式 (5) のとおりである。

$$\alpha(S, T) = \prod_{l=1}^n \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{d_l^2(S_l, T_l^{(j)})}{2\sigma_l^2}\right) \quad (5)$$

ここで、 σ_l は視点 l において分布する nwXOR 距離の分散である。

Fig. 4 Overview of multi-camera system.

3. 姿勢推定実験

3.1 実験概要 提案手法の実現可能性を検討するために、図4に示すマルチカメラシステムにより撮影した人物の画像シーケンスに対して、提案手法により人物の姿勢を推定するモーションキャプチャシステムを構築し、実験を行う。システムの実装には、C++言語を採用し、開発環境としてMicrosoft Visual Studio (マイクロソフト社)を利用している。また、CGを扱うため、ゲームやマルチメディア処理用のAPIであるMicrosoft DirectX (マイクロソフト社)も利用した。三次元人体CGモデルの作成には、LightWave 3D (エヌジーシー社)を利用し、部位モデルごとにポリゴンデータファイルを用意する。本実験では、対象人物を成人男性1名とし、着衣状態における大まかな形状を事前に計測し、手動で三次元人体CGモデルを作成している。なお、本実験におけるモーションキャプチャは計算コストの都合により、オフライン処理システムとしてパーソナルコンピュータ (DELL XPS730, インテル® Core™ 2 Extreme プロセッサ QX9770, 12MB L2 キャッシュ, 3.2GHz, 1600MHz FSB) に実装した。

3.2 対象動作の計測 図4に示すマルチカメラシステムは、正六角形状に対象人物を囲み、交互に高低差がつくように設置された6台のCCDカメラ (IEEE1394カメラ"DragonFly2", point grey社)の同期を取り、画像の撮影およびシルエットの抽出を行うシステムである。マルチカメラシステムでは、計測の精度と種々のコストを考慮してカメラの台数や設置方法を決定する必要がある。本実験では、対象人物の画像取得に関して、全てのカメラの視野内に観測可能であることが必要であるが、三次元モデリングレベルの精密さは不要であること、後述する誤差評価で使用する市販の光学式モーションキャプチャにおける最小構成カメラ台数の要件を考慮して使用するCCDカメラを6台としている。同等の視野を持つカメラを均等間隔で配置することで、6台のカメラ程度でも、直立

Fig. 5 Feature points of the human body.

状態の姿勢に関しては、あらゆる人体の向きに対して $2(m) \times 2(m) \times 2(m)$ の空間について安定に観測可能な環境を実現している。シルエットの抽出手法には、一般的な室内環境でも使用可能な背景差分法を用いており、RGB表色系における輝度および色差を利用し、各画素毎に前景領域と背景領域を分離している⁽¹⁵⁾。本実験では、このシステムを用いて15[fps]の6視点シルエット画像シーケンスを10[s]分取得した。

また、このCCDカメラによる人物動作の撮影と同時に、市販の光学式モーションキャプチャシステムを使用して対象人物の特徴点の三次元位置計測を行う。このモーションキャプチャシステムは、対象人物の特徴点にマーカを装着し、周囲に配置した6台の赤外線カメラ (USB赤外線カメラ"OptiTrack FLEX:v100", Natural Point社)により取得した画像を、ソフトウェアARENA™ Motion Capture (Natural Point社)により解析することでリアルタイムにモーションキャプチャを行うシステムであり、1[mm]以下の誤差でマーカの三次元位置を計測することができる。本実験では、この計測結果を推定対象の真値として利用した。なお、人物の特徴点は図5に示す関節位置や先端位置からなる14点とした。

3.3 キャリブレーション 提案手法において構築する実空間の撮影系と仮想空間の撮影系は、同一の撮影条件にする必要がある。そこで、キャリブレーションとして以下に示す2つの作業を行う。

(1) 実カメラの歪み補正 一般に、カメラにより取得した画像には、レンズの収差により中心から離れるにしたがって歪みが生じる。これに対し、仮想空間に配置する仮想カメラにはレンズの概念がないため、三次元人体CGモデルを投影した画像に歪みは生じない。提案手法では、実カメラと仮想カメラの観測画像をマッチングするが、これは歪みのある画像と歪みのない画像のマッチングであり不適切である。そこで、チェッカーボードを用いた一般的なカメラキャリブレーション

ン⁽¹⁶⁾により実カメラの内部パラメータと歪み係数を推定し、取得した実画像シーケンスに歪み補正を施す。

(2) 仮想カメラのパラメータ決定 仮想カメラの設置は、実カメラの設置と異なり、投影処理を実現する射影行列を定義することを意味する。射影行列は、カメラの位置や光軸の向き等、カメラの仕様や配置に関するパラメータにより求まる。したがって、実空間の撮影系を仮想空間で再現するためには、実カメラの配置から射影行列のパラメータを推定する必要がある。ここで、本実験システムにおいて使用する Direct X の仮想カメラにおける射影行列のパラメータとは、カメラの三次元位置座標、光軸ベクトルとそれに対するカメラ本体の姿勢、画角の4種類である。カメラの三次元位置座標については、実空間に原点および三次元座標軸を設定し、実測した値を用いる。画角については、使用するカメラの仕様に従う。光軸ベクトルとそれに対するカメラ本体の姿勢については、大きさや位置が既知の参照物体を使用し、実カメラで撮影した参照物体に仮想カメラで撮影した参照物体が一致するパラメータを探索することで推定した。なお、効率的なパラメータ探索のために遺伝的アルゴリズムを使用した。

提案手法では実カメラの外部パラメータを必要としないため、マルチカメラシステム設置後にカメラの移動等が生じて実カメラは再度のキャリブレーションが不要であり、仮想カメラについてのみ再度パラメータを探索すればよい。従って、マルチカメラシステムの設置やその精度に関しては自由度が高くなっている。

3.4 実験結果と考察 推定結果の例として、被験者 A (成人男子, 172cm, やせ型) による2つの動作に対して姿勢推定を行った結果を図6に示す。図6(a)は、その場での足踏みする動作 (以下, Stepping), 図6(b)は、直立した状態で両腕を身体の前で回転させる動作 (以下, Swing arms) であり、代表的なフレーム (図中, #フレーム番号で記述) について各視点における推定対象の人物姿勢を表す入力画像と推定結果の人体 CG モデルの画像を示したものである。推定の条件は, Stepping に対しては, 胴体, 腕, 脚の各個別部位に関する粒子数を 300, 各尤度に関して上位 10 % の粒子の組み合わせ, 全身の粒子数を 9000 とし, Swing arms に対しては, 胴体, 腕, 脚の各個別部位に関する粒子数を 400, 各尤度に関して上位 10 % の粒子の組み合わせ, 全身の粒子数を 16000 とした。これらの値は予備実験を通じて試行錯誤的に決定した。図6に示す通り, どのフレームに関しても推定対象の姿勢に近い推定されており, 良好な推定結果が得られていると

言える。尚, 推定に要する時間は, 対象とした2つの動作について平均 22 分程度であった。

図6に示した動作の推定結果における推定誤差の変化を図7に示す。ここで, 推定誤差とは, 図5に示す人物の特徴点について, 光学式モーションキャプチャにより計測した特徴点 $\{f_i\}_{i=1}^{14}$ と, 推定結果である人体 CG モデルの関節状態および各部位モデルの大きさから算出した特徴点 $\{\hat{f}_i\}_{i=1}^{14}$ の各座標の三次元ユークリッド距離平均を推定誤差 e として定義したもので, 次の式 (6) より求まる。

$$e = \frac{1}{14} \sum_{i=1}^{14} \|f_i - \hat{f}_i\| \quad (6)$$

推定開始の初期フレームでは, 人体 CG モデルの生成がランダム初期値から行われるために推定誤差が比較的大きくなるが, 以降のフレームでは, MCF の効果により誤差の変動が抑制されている。また, 図6より, 提案手法による姿勢推定の推定誤差は平均 0.1[m] 程度であることがわかる。図8に被験者 B (成人男子, 174cm, やせ型) の Stepping に対する推定結果の例, 図9に推定誤差の変化を示す。これより, 異なる被験者についても姿勢推定が可能であることが確認できる。推定誤差が平均 0.13[m] 程度と被験者 A の結果に比べて大きくなっているが, これは被験者の体格が極端に異なることより被験者 A の推定に用いた人体 CG モデルを被験者 B の推定に用いたためであり, モデルの形状を調整することで推定誤差の改善は可能である。提案手法によるマーカレスモーションキャプチャは, スポーツ解析やリハビリ評価など厳密な計測精度を要求するアプリケーションに対しては, 必ずしも十分な分解能であるとはいえないが, マン・マシンインタフェースやビデオゲームインタフェースなどへの応用においては, 十分実用可能な範囲であると考えられる。

4. 結 言

本論文では, 多視点画像を用いたマーカレスモーションキャプチャの開発の一環として, 仮想空間上にモーションキャプチャの対象となる人物の三次元形状を近似する三次元人体 CG モデルを構築し, 実空間の撮影系において取得する対象人物のシルエット画像に対して, 仮想空間の撮影系において実空間と同一の視点に投影したモデルのシルエット画像をマッチングするモーションキャプチャアルゴリズムを提案した。6 台の CCD カメラを用いたマルチカメラシステムで取得した実画像シーケンスに対して提案手法による姿勢推定実験を行い, アルゴリズムの実現可能性を確認

Fig. 6 Estimation results (subject A, top: target posture, bottom: estimated posture with highest likelihood).

するとともに、光学式モーションキャプチャによる計測結果との比較から、提案手法による推定誤差が平均

0.1[m]であることを明らかにした。

Fig. 7 Averaged estimation error (subject A).

Fig. 8 Estimation results (subject B, top: target posture, bottom: estimated posture with highest likelihood).

Fig. 9 Averaged estimation error (subject B).

文 献

(1) T. Kanekubo and A. Miyazawa, "Latest technologies and trends of 3D character animation in computer games",

- Journal of the Institute of Electrical Engineers of Japan*, Vol.129, No.9, (2009), pp.616–619. (in Japanese)
- (2) L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis", *Pattern Recognition*, 36(3), (2003), pp. 585–601.
- (3) T. B. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis", *Computer Vision and Image Understanding*, 104(2), (2006), pp. 90–126.
- (4) D. A. Forsyth, O. Arikian, L. Ikemoto, J. O' Brien, and D. Ramanan, "Computational studies of human motion: Part 1, tracking and motion synthesis", *Foundations and Trends in Computer Graphics and Vision*, 1, (2006), pp. 77–254.
- (5) R. Poppe, "Vision-based human motion analysis: An overview", *Computer Vision and Image Understanding*, 108, (2007), pp. 4–18.
- (6) N. Shimada, D. Arita, and T. Tamaki, "Model Fitting of Articulated Objects", *IPSS SIG Notes (CVIM)*, Vol. 2006, No. 51, (2006), pp. 375–392. (in Japanese)
- (7) K. M. (G) Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time part II: application to human modeling and markerless motion tracking", *International Journal of Computer Vision*, Vol. 63, No. 3, (2005), pp. 225–245.
- (8) B. Michoud, E. Guillou, H. Briceno and S. Bouakaz, "Real-time marker-free motion capture from multiple cameras", in *Proceedings of 2007 IEEE 11th International Conference on Computer Vision*, (2007), pp. 1–7.
- (9) K. Ogawara, X. Li, and K. Ikeuchi, "Marker-less human motion estimation using articulated deformable model", in *Proceedings of 2007 IEEE International Conference on Robotics and Automation*, (2007), pp. 46–51.
- (10) K. Takahashi, Y. Nagasawa, and M. Hashimoto, "Markerless human motion capture from voxel reconstruction with simple human model", *JSME Journal of Advanced Mechanical Design, Systems, and Manufacturing*, 2(6), (2008), pp. 985–997.
- (11) S. Sakamoto, K. Shoji, F. Toyama, and J. Miyamichi, "Optimization of Camera Parameters in Volume Intersection", *IEEJ Transactions on Electronics, Information and Systems*, Vol. 125, No. 7, (2005), pp. 1024–1029. (in Japanese)
- (12) G. Kitagawa, Introduction to time series analysis, Iwanami Pub. Co., (2005), pp.209–222.
- (13) N. Ichimura, "Stochastic filtering of motion trajectory using a self-organizing state space model", *IPSS Transactions on Computer Vision and Image Media*, Vol. 43, No. 11, (2002), pp. 92–104. (in Japanese)
- (14) R. Okada, B. Stenger, "A single camera motion capture system for human-computer interaction", *IEICE Transactions on information and systems*, Vol. E91-D, No. 7, (2008), pp. 1855–1862.
- (15) K. Takahashi, "Remarks on background subtraction using color information", *IEEJ Transactions on Electronics, Information and Systems*, Vol. 123, No. 3, C(2003), pp. 627–628. (in Japanese)
- (16) NAIST OpenCV programming book publishing team, OpenCV Programming Book, Mainichi Communications, (2007), pp. 280–296. (in Japanese)