

身体モデルを用いた 3D モーションキャプチャ 3D Motion Capturing Using Body Model

萩戸 翔† Sho Shishido 齋藤 宗彦† Munehiko Saito 安藤 敏彦† Toshihiko Ando

1. はじめに

人間が対面して行うコミュニケーションには、言葉の他に身振り、表情、うなずきなど、非言語的な側面が含まれ、それによりコミュニケーションが円滑で快適なものになると考えられる。本研究では非言語コミュニケーションを調べるためのモーションキャプチャシステムの開発を行っている。一般的なモーションキャプチャシステムでは、専用のキャプチャスーツや特殊機材などの周辺機器が必要となり、撮影場所や動作位置が限定されるといった問題がある。本研究では安価で、持ち運びの容易なデジタルビデオカメラ（以下 DVC）を使用し、L フレームとワンドを用いた動的キャリブレーション、および身体モデルを用いた被験者のモデリングを採用する。これらの手法により、簡易かつ高速なモーションキャプチャシステムの実現が期待できる。

2. モーションキャプチャシステムの概要

本研究では、被験者の測定には 3 台の DVC を使用する。DVC を使用する場合、各カメラ映像（2 次元座標系）と実空間（3 次元座標系）を対応させるカメラキャリブレーションが必要になる。複数の視点から同じ目標を測定できれば、各視点の 2 次元座標とカメラパラメータから 3 次元座標を求めることができる。キャリブレーションの手法としては、ウェービングワンド法¹⁾という比較的時間のかからない方法を採用する。実際は 2 台の DVC からの情報があれば、3 次元座標を求めることが可能であるが、一方のカメラで目標を見失うと 3 次元座標を求めることができなくなる。そのため、本研究では最低 2 台のカメラで目標を測定できるように、冗長性を兼ねた 3 台の DVC を使用する。

また、本研究の測定対象として、主にパフォーマンス・アートの演技者を取り上げる。この際、なるべく演技者に制約を与えない環境を作ることが重要となる。従来は、被験者の関節位置にカラーマーカーを取り付け、画像処理でマーカーを抽出することで動作のモデリングを行ってきた。しかし、この手法ではマーカーが被験者の動作の妨げになるといった問題が発生する。そこで本研究では、マーカーの代わりに身体の構造を模したモデルを用意し、撮影後の動画像において、モデルを被験者の輪郭に重ね合わせることで動作のモデリングを行う。この手法ならば、被験者に測定器具を取り付ける必要がないため、被験者はより自然な動作をすることが可能となる。

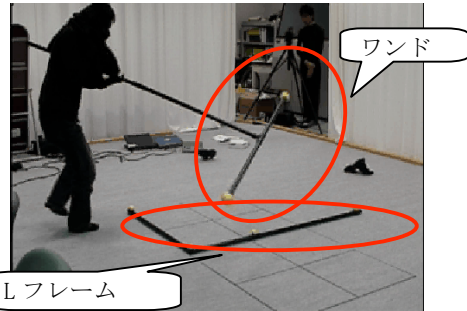


図1. キャリブレーション器具

3. 提案手法の概要

3.1 動的キャリブレーション

キャリブレーション器具として、図1に示すLフレームとワンドを使用する。LフレームはL字の長軸に2つ、短軸に1つ、長軸と短軸の交点に1つ、計4つのマーカーを取り付けた器具で、3次元座標系の原点と座標軸を定義する。ワンドは棒の両端にマーカーを取り付けた器具で、撮影空間内で動かすことによって多くのキャリブレーションデータを発生させる。Lフレームとワンドのマーカーの抽出には色フィルタを用いており、抽出の際の誤差を減らすために、作業者は黒の衣装を着用している。

実際の3次元座標からカメラ映像上の2次元座標への変換には、(1)式に示すピンホールカメラモデルを使用する。

$$\begin{aligned}
 \lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} &= P \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = KMRT \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = K \begin{bmatrix} I & 0 \\ R & [I \ -\vec{t}] \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} \alpha & c & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -t_1 \\ 0 & 1 & 0 & -t_2 \\ 0 & 0 & 1 & -t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (1)
 \end{aligned}$$

行列 K はカメラごとに固有で、カメラの内部パラメータによる補正のための変換行列を表す。R は回転、T は並進を行う行列で、R と T によって 3 次元座標系をカメラの視点に固定されたカメラ座標系に変換する。R と T の要素は Levenberg-Marquardt 法により求められる。

3.2 身体モデル

本研究では、身体を図2に示す多関節モデルで表現し、これをもとに、主要な部位の3次元座標を導出する。身体モデルは身体の主要な関節を木構造で表現したもので、部位間の矢印の向きは親から子への向きを示す。各部位 i には固有の座標系 Σ_i を設定し、原点を関節位置（腰部では重心位置）に置く。腰部を木構造の最上位に置き、 $i=1$ とする。腰部の親は簡単のためにカメラ座標系としておく。各部位の姿勢は、その親の座標系を基準に表す。時刻 t に

† 仙台電波工業高等専門学校

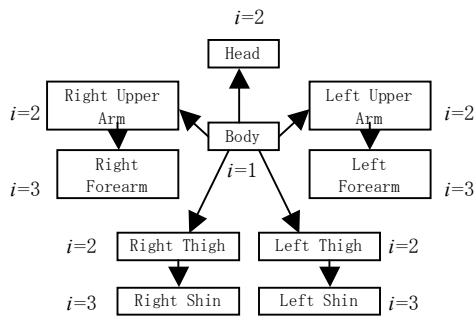


図2. 身体モデル

において、親の部位に対する x , y , z 軸周りの回転量を $\theta_{xi}(t)$, $\theta_{yi}(t)$, $\theta_{zi}(t)$, 親の部位に対する x , y , z 軸方向の移動量を $a_i(t)$, $b_i(t)$, $c_i(t)$ とすると、座標系 Σ_i から親の座標系 Σ_{i-1} への変換は同次変換行列 T_i を用いて

と表せる。ただし、 $Q_i(t)$ は回転行列で、 x , y , z 軸周りの回転行列をそれぞれ $Q_x(\theta_x)$, $Q_y(\theta_y)$, $Q_z(\theta_z)$ とすると

となる。また、 $S_i(t)$ は並進ベクトルで、

である。撮影後の動画像において、適当な開始フレームを選び、身体モデルを手動で被験者の輪郭に重ね合わせることで、 $T_i(t)$ の近似値が得られる。このとき、被験者がカメラに対して正面を向いているフレームを選ぶことで、身体の陰に隠れる部位が少なくなり、照合が容易に行える。この照合は 2 次元の画像上で行き、奥行き方向の姿勢については、各部位の 2 次元座標を 3 次元座標に変換することで決定する。

実際の測定では、開始フレームのみ手動で姿勢を調整し、以後は動画像から被験者の姿勢の変化分を推定し、モデルを自動追跡させる。モデルの追跡は、各フレームにおいて $T_i(t)$ の値を求め、モデルの姿勢を逐次更新していくことで実現できる。モデルの姿勢は $Q_i(t)$, $S_i(t)$ によって決定されるので、 $Q_i(t)$, $S_i(t)$ の要素を運動パラメータと呼ぶ。運動パラメータの導出については、山本らによって提案された身体の追跡法^{2),3)}を採用している。この手法では、測定対象は 1 人に限定されているが、本研究ではコミュニケーションデータの収集のため、必要に応じて複数の身体モデルを用意し、複数の演技者の測定を同時に行う。

3.3 オクルージョンへの対策

被験者の測定において、被験者の身体部位が別の部位を隠してしまう現象が発生する。これをオクルージョンと呼ぶ。カメラ 1 台の動画像においてオクルージョンが発生すると、陰に隠れた部位は正確に追跡することができなくなる。本研究では 3 台のカメラを使用しているため、あるカメラでオクルージョンが発生しても、残りのいずれかのカメラに対象の部位が移っていれば追跡は可能である。オクルージョンへの対策としては、なるべく被験者の全身を撮影できるよう、被験者を囲むようにカメラを配置することが考えられる。

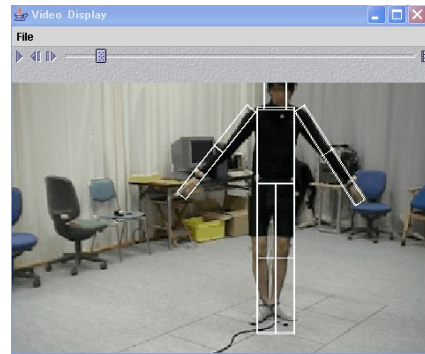


図3. 追跡の様子

4. 実験：対象部位の追跡

実験として、被験者が腕を上下に振りながら回転する動作を取り上げた。測定対象部位は腰部、頭部、上腕、下腕、上脚、下脚とし、追跡の様子を図 3 に示す。この実験の場合、被験者はカメラの中央で回転しているため、奥行き方向の移動量が少なく、初期姿勢からほとんど変化していない。そのため、奥行き姿勢は比較的精度良く追跡できる。また、腰部の回転が動作の中心になっており、腰部の追跡精度が他の部位の追跡精度に影響を与えている。一方、被験者が腕を上下に振っているため、腕が交差したり、カメラに対して側面を向いたときにオクルージョンが発生する。これに対しては、オクルージョンが発生したフレームを検出し、手動でカメラを切り替えて追跡を続行した。現段階では、オクルージョンが発生するたびに手動でカメラの切り替えを行っており、追跡の手間が煩雑になっている。結果的に、ある程度の追跡は実行できるものの、長時間の追跡では運動パラメータの導出において誤差が蓄積され、実用的な精度は得られていない。

5. まとめ

本研究では、簡易かつ高速なモーションキャプチャを行うための手法を提案した。撮影はキャリブレーションも含めて 1 時間程度で実行でき、演技者は測定器具を取り付けることなく、より自然な演技をすることが可能になった。しかし、動作の追跡において初期姿勢の照合や、オクルージョンへの対策は手動で行っており、測定の手間が増大している。また、運動パラメータの精度が不十分な場合、追跡が十分に行えないという問題が明らかになった。今後の課題として、オクルージョンの自動検出、および運動パラメータの精度向上が挙げられる。

参考文献

- 1) E. Stenborg and L. Svensson, "Complete Camera Calibration by Waving Wand", Technical Report EX040/2005, Chalmers Univ., 2005.
- 2) 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 次元動画追跡", 信学論, J79-D-II, 1, pp. 71-83, 1996
- 3) 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.