

# 和装人物の単視点姿勢推定のための遮蔽部位3次元位置推定

梶原 遼<sup>1,a)</sup> 延原 章平<sup>1</sup>

**概要:** 本論文は単眼画像を入力とした和装人物の3次元骨格運動推定を深層学習によって実現することを目的としたものであり、そのために必要となる学習データの中でも特に和服によって遮蔽された部位の3次元位置を、モーションキャプチャーデータ間の非剛体位置合わせと逆運動学推定を利用して推定する手法を提案する。

## 1. はじめに

コンピュータビジョン分野では2次元画像から3次元形状推定を行う代表的な手法としてステレオビジョンや視体積交差法 [1] など幾何的に解く手法があるが、ただ1枚の画像から3次元形状復元を行う手法として深層学習アーキテクチャによる回帰 [2][3]、あるいは回帰と分類を合わせた問題として解く手法が存在する [4]。入力を1枚のRGB画像とし、出力を3次元形状推定に関する特徴量とすることが主である。すなわち、あらかじめ学習されたものを用意しておけば、必要なのは1台のRGBカメラと1台のコンピュータだけで、光学式モーションキャプチャーの赤外線カメラのような特別な装置は必要ない。また過去に撮影された画像からも3次元形状推定が行える。3次元形状推定の対象を人間に限った場合、骨格推定 [3] と表面形状の推定 [4] の2つの手法が考えられるが、本研究では前者を想定する。

このような学習に基づく手法においては、人物のアピアランスと3次元骨格姿勢の組を学習データとして用意しなくてはならないが、特に着衣が洋服のような身体を緊密に包む窄衣型ではなく、和服のような身体に布をかけて着る懸衣型である場合は、衣服が必ずしも身体に沿わない平面で構成されるために衣服の下の遮蔽された関節位置の推定が人間の目でさえも難しく、したがって学習データの取得そのものが困難である。

このような理由から既存の骨格を推定する手法では、懸衣型の衣服は想定されていないことが多い。そこで本研究では和装人物の学習用データ作成方法として、同じ特性を示す運動軌跡間で対応付けを行う手法、およびそれらを用いて測定可能な露出関節間で求めた対応付けから逆運動学

的に遮蔽関節を推定する手法を提案する。最終的な目的は古い時代劇の映像から役者の骨格推定が可能となることである。

## 2. 関連研究

既存のデータ作成法について代表例をいくつか述べる。HaslerらはThe Captury[5]というマーカレスモーションキャプチャー用のソフトウェアを開発した。これはあらかじめアクターをスキャンしてモデルを作成し、そのモデルを変形させることによりリアルタイムに人物の3次元形状を計測する。そのため、この方法では既存の手法と違いマーカや撮影のためのスタジオなど特別な設備が必要がない。しかし、アクターモデル作成の際には身体の形が外側にはっきりとは現れない和服は想定されていない。

骨格推定を行う手法として他にMicrosoft Kinectが知られているが、これは深度画像からの回帰問題として3次元骨格の推定を行う。赤外線式モーションキャプチャーなどと比べるとはるかに精度が劣り、また和装人物が学習データに含まれていないために推定に失敗することも多い。

このような理解に基づいて、本研究では赤外線モーションキャプチャーによる高精度3次元位置測定データを用いて、和装時のデータを作成する。

## 3. 提案手法

前提条件を以下に述べる。

- アクターは殆ど同じ演技を繰り返すことのできるプロであるとする。
- 和装時のモーションキャプチャー用マーカ集合は、モーションキャプチャースーツ時のマーカ集合の部分集合であるとする。具体的には、和服はモーションキャプチャーの上に着ることとし、結果として和服から露出している頭、両手、両足に付着しているマーカが

<sup>1</sup> 京都大学大学院情報学研究科

<sup>a)</sup> kajiwara@vision.kuee.kyoto-u.ac.jp

測定可能とする。

- 撮影に用いるカメラはすべてキャリブレーション済みであるとする。

学習用データの作成方法について以下に述べる。

**STEP1** アクターにモーションキャプチャスーツを着た状態と、その上に和服を着た状態とで2回全く同じ演技をしてもらい、赤外線式モーションキャプチャでマーカの3次元位置を測定する。マーカの位置から、アクターの関節位置を推定する。

**STEP2** 和装状態でも測定できる関節に関して、2回の演技の中で同じ演技動作であるとみなせる同一関節におけるデータの対応を求める。

**STEP3** STEP2で求めた対応付けを基に、逆運動学的に解くことで和装時の遮蔽関節の3次元位置を求める。

**STEP4** 求めた3次元位置を和装状態の画像に投影し、画像上の関節位置とする。

以下、モーションキャプチャスーツを着用した状態で測定したデータを非和装時のデータ、スーツの上から和服を着た状態で測定したデータを和装時のデータと呼ぶことにする。

$M$ : 非和装時の全関節の時系列3次元データの集合

$C$ : 和装時の全関節の時系列3次元データの集合

$N_m$ :  $M$ 中のデータの総フレーム数

$N_c$ :  $C$ 中のデータの総フレーム数

$J$ : 非和装時の関節の集合

$J'$ : 和装時でも位置が求められる関節の集合、 $J' \subset J$ とする。

STEP2とSTEP3では関節の3次元位置を用いるが、STEP4ではマーカの3次元位置を用いる点に注意する。

#### 4. STEP1: マーカ位置から関節位置への変換

赤外線式モーションキャプチャで各マーカの3次元位置を求め、各マーカをあらかじめ作成しておいたアクターの3次元モデルのマーカに対応させ、そこから関節位置を決める。これはすべてコンピュータグラフィックスにおけるキャラクター作成作業と同様であり、今回は人手で行った。

#### 5. STEP2: 3次元運動軌跡間の時間的・空間的対応付け

関節集合  $J'$  に属する関節に関して、非和装状態の3次元データを同一演技動作時の和装状態の3次元データへ対応付けするために、 $M$  と  $C$  の組み合わせ最適化問題を解くことを考える。そのためのデータに関するコストになりうる要素として考えられるのは、

1. 対応点間における距離と時間差
2. アクターの演技速度差
3. アクターの運動軌道の曲率差

以上が考えられる。さらに、対応付けの妥当性を表す要素が以下のように考えられる。

4. 対応関係が前後と平行であること、つまり図1のように対応関係が離れすぎたり、図2のように入れ替わったりしていないこと（対応の連続性）
5. ある関節の対応関係が別の関節の対応関係と外れすぎっていないこと（演技動作の同一性）

以上5つの要素を考慮したコスト関数の最小化問題を考える。

#### 5.1 データに関するコストの定式化

上述の要素の1から3について述べる。3次元データ  $m \in M$  の対応先  $c \in C$  は  $m$  と同一の関節に属する3次元データから選ばれるものとする。

##### 5.1.1 距離と時間差

3次元データ  $m \in M$  と  $c \in C$  にそれぞれ対応するフレーム番号  $n_m, n_c$  を付加し、4次元データ  $m', c'$  とする。 $L_1$  距離の重み付け線形和を用いて、 $m'$  と  $c'$  の距離は

$$d_m(n_m, n_c) = w_d \sum_{k=1}^3 |m_k - c_k| + w_f |n_m - n_c| \quad (1)$$

ただし、4次元目をフレーム番号とし、 $m' = (m_1, m_2, m_3, n_m)^\top$ ,  $c' = (c_1, c_2, c_3, n_c)^\top$  とする。

##### 5.1.2 速度

注目フレームに対し、前フレームでの3次元位置と次フレームでの3次元位置の差の絶対値を平均した中心差分を速度とする。ただし、最初のフレームの速度は次のフレームとの前進差分、最後のフレームの速度は前のフレームとの後退差分とする。すなわち、 $m \in M$  に関して、フレーム番号が  $n$  の3次元データを  $m(n)$  と置くと、

$$v_m(n) = \begin{cases} \frac{m(n+1) - m(n-1)}{2} & (1 < n < N_m) \\ m(n+1) - m(n) & (n = 1) \\ m(n) - m(n-1) & (n = N_m) \end{cases} \quad (2)$$

$c \in C$  に関しても同様に求める。速度に関するコストはとする。

##### 5.1.3 曲率

ある曲線が存在し、 $s$  を曲線上の長さとする、曲線上の点は  $r(s)$  と書ける。単位接ベクトルは

$$T = \frac{dr}{ds} = \frac{dr}{dt} \frac{dt}{ds} = \frac{v}{\frac{ds}{dt}} \quad (3)$$

ただし、 $v$  は速度である。曲率は

$$\chi = \left| \frac{dT}{ds} \right| = \left| \frac{\frac{dT}{dt}}{\frac{ds}{dt}} \right| \quad (4)$$

である。以上から、 $m \in M$  に関して、

$$\frac{ds_m}{dt} = |v_m(n)| \quad (5)$$

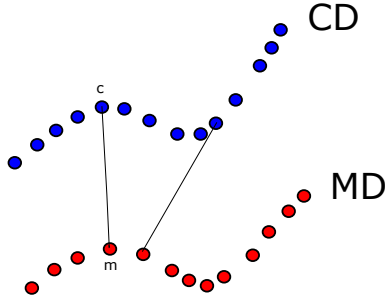


図 1 あるフレームと次のフレームの対応先が離れすぎている場合の 1 例

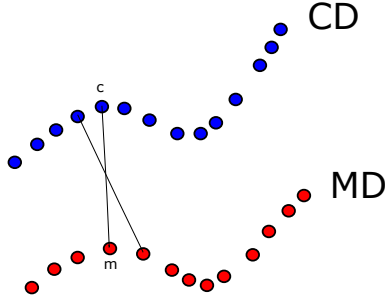


図 2 あるフレームと次のフレームの対応先が交差してしまっている場合の 1 例

$$\mathbf{T}_m(n) = \frac{\mathbf{v}_m(n)}{|\mathbf{v}_m(n)|} \quad (6)$$

$$\frac{d\mathbf{T}_m}{dt} = \begin{cases} \frac{\mathbf{T}_m(n+1) - \mathbf{T}_m(n-1)}{2} & (1 < n < N_m) \\ \mathbf{T}_m(n+1) - \mathbf{T}_m(n) & (n = 1) \\ \mathbf{T}_m(n) - \mathbf{T}_m(n-1) & (n = N_m) \end{cases} \quad (7)$$

$$\chi_m(n) = \frac{\left| \frac{d\mathbf{T}_m}{dt} \right|}{|\mathbf{v}_m(n)|} \quad (8)$$

$\mathbf{c} \in C$  に関しても同様に求める。曲率に関するコストは

$$d_{\chi_m}(n_m, n_c) = |\chi_m(n_m) - \chi_c(n_c)| \quad (9)$$

以上のコストの線形和を  $M$  側のあるフレームを、 $C$  側のあるフレームに対応付ける場合のコストとして定義する。つまりある関節において、 $n_m$  フレーム目のデータ  $\mathbf{m} \in M$  の  $n_c$  フレーム目のデータ  $\mathbf{c} \in C$  に対するコストは

$$D_m(n_m, n_c) = d_m(n_m, n_c) + w_v d_v(n_m, n_c) + w_\chi d_{\chi_m}(n_m, n_c) \quad (10)$$

と定義する。

## 5.2 対応付けの妥当性に関するコストの定式化

上述の要素の 4 と 5 にあたるコストに関して述べる。

### 5.2.1 対応の連続性

ある関節において、フレーム番号  $n_m$  のデータ  $\mathbf{m} \in M$  がフレーム番号  $n_c$  のデータ  $\mathbf{c} \in C$  と対応付けされ、フレーム番号  $n_m + 1$  のデータ  $\mathbf{m}' \in M$  がフレーム番号  $n'_c$  のデータ  $\mathbf{c}' \in C$  と対応付けされたとする。2 回の演技の速度がほぼ同じだと仮定すれば、対応の連続性とはこの 2

つの対応関係が空間的には平行になることを、時間的には  $n'_c = n_c + 1$  を満たすことに相当する。図 1 や図 2 のように対応付けが離れすぎたり、入れ替わったりしているほど大きなコストを与えるためには、以下のように定式化すればよい。

$$C_{m1}(n_m, n_c, n'_c) = |(n_c - n_m) - \{n'_c - (n_m + 1)\}| \quad (11)$$

演技速度の多少のずれにより、時間的に対応が少し離れていても許容できるが、入れ替わりは許容できない。よって、入れ替わりに対してより大きなコストを課すべきでだと考え、さらに以下のコストを付与する。

$$C_{m2}(n_c, n'_c) = \begin{cases} 0 & (n'_c - n_c \geq 0) \\ w_{penalty} \times |n'_c - n_c| & (n'_c - n_c < 0) \end{cases} \quad (12)$$

### 5.2.2 演技動作の同一性

同じ演技をするのだから、ある関節でのフレーム番号  $n_m$  のデータ  $\mathbf{m} \in M$  が対応する先  $\mathbf{c} \in C$  のフレーム番号  $n_c$  と別の関節でのフレーム番号  $n_m$  のデータ  $\mathbf{m}_o \in M$  が対応する先  $\mathbf{c}_o \in C$  のフレーム番号  $n_{c_o}$  は基本的には一致するべきである。これに対応するものとして、 $n_c$  が  $n_{c_o}$  に近いほど小さくなるようなコストを考え、単純に

$$C_{m_o}(n_c, n_{c_o}) = |n_c - n_{c_o}| \quad (13)$$

を演技動作の同一性に対するコストとする。

以上 3 つのコストの線形和を以て対応付けの妥当性に関するコストとできる。

$$C_m(n_m, n_c, n'_c, n_{c_o}) = w_{c1} C_{m1}(n_m, n_c, n'_c) + w_{c2} C_{m2}(n_c, n'_c) + w_o C_{m_o}(n_c, n_{c_o}) \quad (14)$$

## 6. MAP-MRF 問題としての定式化

以上までで述べたコストの総和を最小化する組み合わせ最適化問題は、次に述べる MAP-MRF 問題に帰着できる。以下の式で表されるエネルギー関数を最小化することを考える。

$$E(L) = \sum_{v \in V} D(L_v) + \sum_{u, v \in E} w_{u,v} C(L_u, L_v) \quad (15)$$

ここで、 $V$  はノードの集合、 $E \subset V \times V$  は、ノード間の隣接関係を表す集合である。今回の状況でいえば、 $V = M$  である。この式は 1 階のマルコフ確率場 (MRF) における事後確率最大化問題 (MAP 推定) のときによく現れる式である。データ  $d$  が教示されたとき、事後確率  $P(l|d)$  を最大化するようなラベル  $l$  を求める問題を考えると、ベイズの定理より

$$P(l|d) = \frac{P(d|l)P(l)}{P(d)} \quad (16)$$

となる。ここで  $P(d)$  はデータにのみ依存するので分子の

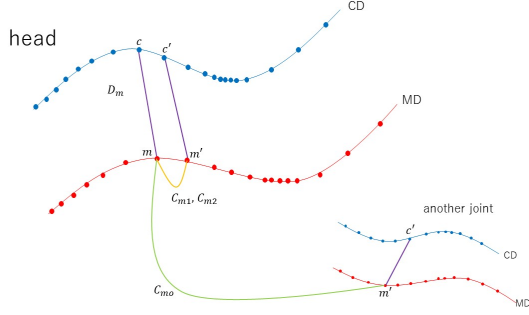


図 3 各コストと対応付けの関係

最大化を考えればよいので、分子のみについて対数を取り、 $-1$  倍すれば、

$$E(l) = -\ln P(d|l) - \ln P(l) \quad (17)$$

この式の最小化問題になる。式 (15) の右辺第 1 項、第 2 項がそれぞれ式 (17) の右辺第 1 項、第 2 項に対応している。それぞれをデータ項、平滑化項と呼ぶ。前節までの最適化問題をこの問題に置き換えると、

$$E(n_m, n_c, n'_c, n_{c_o}) = \sum_{m \in M} D_m(n_m, n_c) + \sum_{m \in M} C_m(n_m, n_c, n'_c, n_{c_o}) \quad (18)$$

$n'_c, n_{c_o}$  は一つにまとめることができる。 $m$  と結ばれるエッジのもう一方の端点と同じ関節なら  $C_{m_o} = 0$  とし、異なる関節なら  $C_{m1} = C_{m2} = 0$  とすればよい。

$$E(n_m, n_c, n'_c) = \sum_{m \in M} D_m(n_m, n_c) + \sum_{m \in M} \{w_{c1} C_{m1}(n_m, n_c, n'_c) + w_{c2} C_{m2}(n_c, n'_c) + w_o C_{m_o}(n_c, n'_c)\} \quad (19)$$

式 (19) 右辺第 1 項、第 2 項がそれぞれ式 (15) 右辺第 1 項、第 2 項に対応している。

図 3 にコストと対応付けの関係を表す。 $c$  は  $m$  の対応先データ、 $m'$  は  $m$  の次のフレームの同じ関節のデータ、もしくは  $m$  と同じフレームだが別の関節のデータで、 $c'$  は  $m'$  の対応先データである。

## 7. STEP3 : 逆運動学による遮蔽関節位置の推定

前節で和装状態でも測定できる関節  $j \in J'$  に関して、 $M, C$  間対応の取り方を述べた。続いて得られた対応関係を用いた遮蔽関節の位置推定を行う。

### 7.1 フレーム単位の剛体位置合わせ

時間的な関節軌道の連続性は考慮せず、各フレーム毎に解くものとする。ある代表的な関節  $j_{re}$  を  $J'$  から選ぶ。非和装時の関節の位置データ全体を平行移動させ、非和装時と和装時の  $j_{re}$  の位置を一致させておく。

### 7.2 フレーム単位の非剛体位置合わせのための関節角をパラメータとする非線形最適化

逆運動学的に最適化問題として解き、遮蔽関節の位置推定を行うことを考える。骨の長さを固定し、関節角をパラメータとした非線形最適化問題とする。

$\theta$  : 関節の回転パラメータを要素とする関節角ベクトル

$X_j$  : 関節  $j \in J$  の 3 次元位置、

$d$  : 対応先との距離上限値

として、

$$\min_{\theta, X_{root}} \sum_{j \in J'} |X_j^m - X_j^c|$$

$$\text{s.t. } |X_j^m - X_j^c| < d$$

ここで、関節角ベクトル  $\theta$  と 3 次元位置  $X_j$  の階層構造を使った定義を述べておく。関節  $j$  に紐付けられた位置パラメータを  $(X_{j,pos}, Y_{j,pos}, Z_{j,pos})$  とし、回転パラメータを  $(\theta_{j,x}, \theta_{j,y}, \theta_{j,z})$  と置く。回転パラメータは各座標軸を軸とした回転の角度 (オイラー変換)、位置パラメータは並進を表す。回転と並進をそれぞれ  $R_j, t_j$  とおくと、同次座標における変換は

$$M_j = \begin{pmatrix} R_j & t_j \\ 0 & 1 \end{pmatrix} \quad (20)$$

である。この変換は座標を関節  $j$  に紐付けられた座標系へ変換することを意味する。親関節座標系における関節  $j$  の座標  $X_j^p$  は

$$\begin{pmatrix} X_j^p & 1 \end{pmatrix}^\top = M_{parent} M_j \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}^\top \quad (21)$$

と求まる。これを再帰的に繰り返し関節の座標を世界座標系に変換する。

$$\begin{pmatrix} X_j \\ 1 \end{pmatrix} = M_{world} M_{root} \cdots M_{parent} M_j \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (22)$$

$\theta$  は各関節に割り当てられた回転パラメータを要素にもつベクトルである。

$$\theta = \left( \theta_x^{j_1} \quad \theta_y^{j_1} \quad \cdots \right)^\top \quad (j_1, j_2, \dots \in J) \quad (23)$$

$X_{root}$  は階層構造最上位にある関節の 3 次元位置である。この定義に従えば、最適化において骨格の長さは固定される。

## 8. 評価実験

以下、長さの単位は全て mm で統一するものとする。

### 8.1 撮影環境

モーションキャプチャ撮影は東映東京撮影所で行った。

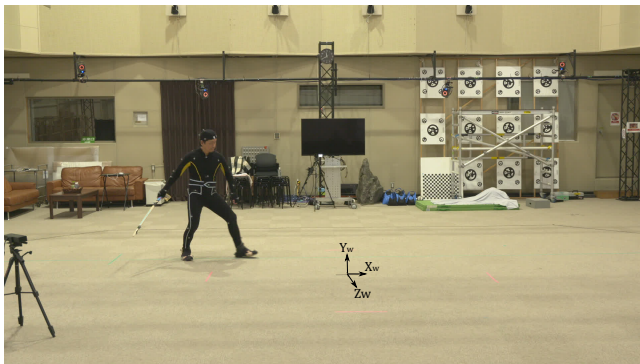


図 4 撮影の様子・世界座標系の取り方

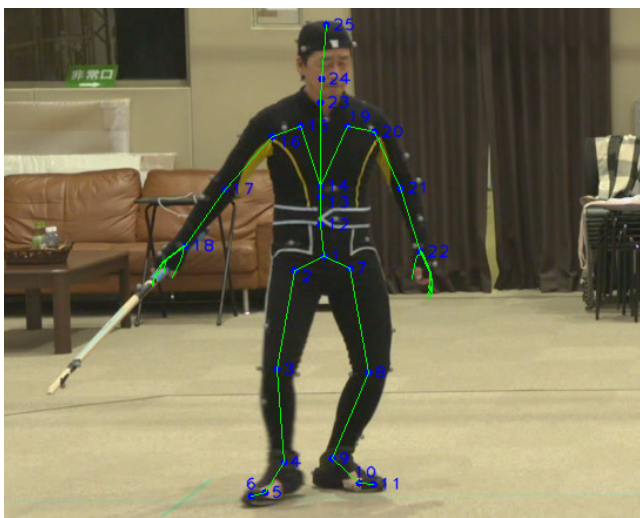


図 5 アクターの関節位置の定義

その様子を図 4 に示す。

- スタジオ床の十字線の交点を世界座標系の原点とする (図 4)。
- 撮影カメラは業務用を 2 台, 赤外線カメラは 24 台。
- 撮影カメラは 59.94fps, 赤外線カメラはちょうど 2 倍の 119.88fps に設定。

マーカの 3 次元位置の測定精度は mm 単位である。

マーカは刀に計 4 つ, アクターには計 70 個取り付けた。関節の位置は頭部, 両肩, 両肘, 両手首, 両手, 脊椎, 骨盤, 両肘, 両足首, 両足にあてた。その具体的なアクターの関節の位置を図 5 に, 番号と関節名の対応を表 1 に示す。頭部と足は 2 か所, 脊椎と骨盤は 3 か所を関節位置として分けた。

和装時の測定可能関節集合は

$J' = \{\text{head, right wrist, left wrist, right ankle, left wrist}\}$ であった。末端の関節であることが逆運動学の制約として強く働く。

第 4 節で述べたようにマーカの位置からアクターの関節位置を特定したが, その際, 画像とデータ間でフレーム対応を取れるように, 撮影カメラに合わせて関節データを 59.94fps に合わせた。アクターには時代劇の殺陣の映像を

番号	関節の種類
1	hip
2	right pelvis
3	right knee
4	right ankle
5	base of right toe
6	end of right toe
7	left pelvis
8	left knee
9	left ankle
10	base of left toe
11	end of left toe
12	lower spine
13	middle spine
14	upper spine
15	right clavicle
16	right shoulder
17	right elbow
18	right wrist
19	left clavicle
20	left shoulder
21	left elbow
22	left wrist
23	neck
24	head
25	top of head

表 1 関節名の定義

見せ, 主役と同じ演技をしていただくようにした。撮影は複数のシーケンスで行ったが, 今回用いたのは 1 シーケンス分のデータのみである。なお, 演技時間は数秒から二十秒程度なので, 総フレーム数は高々数百から千数百である。

## 8.2 MAP-MRF 問題

### 8.2.1 設定

非和装時データの全 829 フレーム中 230 から 620 まで, 和装時データの全 881 フレーム中 260 から 700 までを用いた。すなわち,  $N_m = 390, N_c = 440$  である。まず,  $C$  の全データを対象とした式 (1) を用いた Nearest Neighbor 法で, 注目データ  $m \in M$  の最近傍 200 個を求め, そのうち閾値以上のものを省いた。ただし, 近傍点は同じ関節に属するものから選んだ。これを求めるのに本研究では KDTree を用いた。それら 200 個以下のデータを  $m$  の対応先候補とした。これを  $M$  に含まれる全データに対して行った。

対応先候補に絞って, 最適な対応付けを第 5 節で述べたコストの最小化問題として求めた。最適化の手法は 2 段階構成で, 1 段階目は大域的最適解を求めるよう木の再重み付けメッセージ送受信の逐次拡張 (TRWS) [6] を, 2 段階目は局所的最適解を求めるために繰り返し条件付きモード (ICM) [7] を用いた。1 段階目は最大 10 回まで, 2 段階目は最適解が見つかるまで解いた。

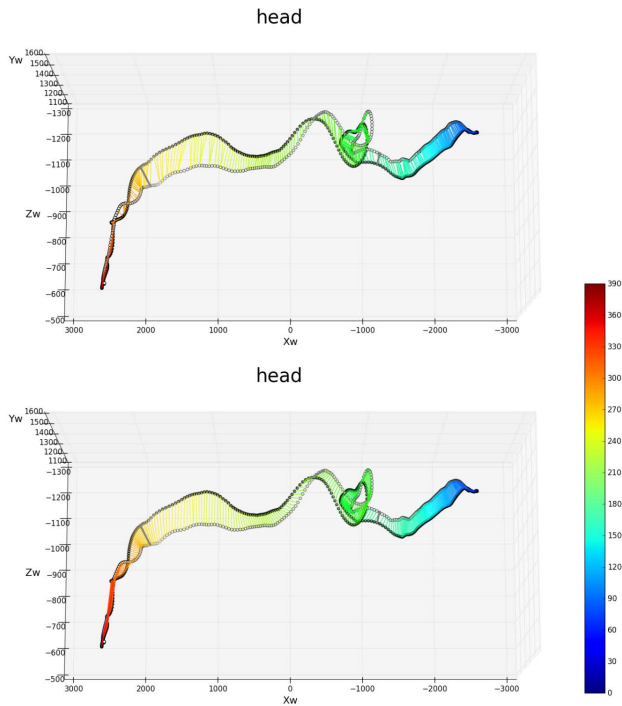


図 6 コストを変化させたときの対応付け結果

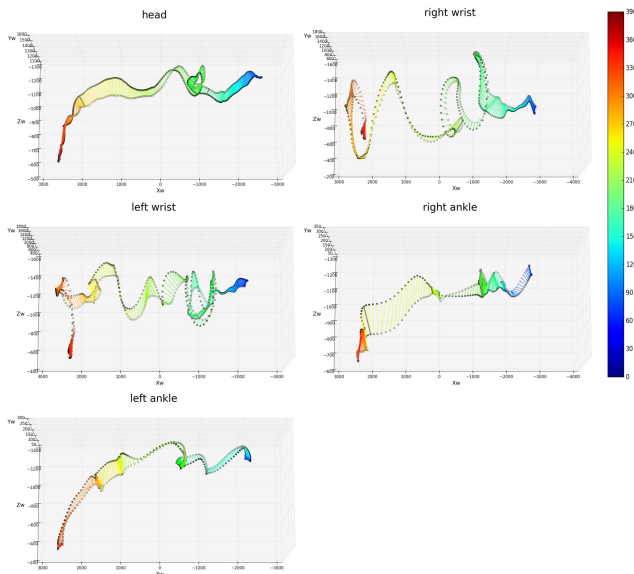


図 7 全関節の対応付け結果

コスト関数各項の係数を以下のように定めた。

$$\begin{aligned}
 w_d : 1 & & w_f : 15 & & w_v : 3 & & w_\chi : 500 \\
 w_{c1} : 600 & & w_{c2} : 200 & & w_o : 300 & & 
 \end{aligned}$$

なお運動軌跡上で対応関係が明らかな点をいくつか選び、それらについては対応先候補を1つにあらかじめ限定することで、正解とする対応を手動で何ペアか与えた。

### 8.2.2 結果

図 6, 図 7 で、黒点が非和装時、白点が和装時のデータである。灰色の線は正解対応としてあらかじめ与えたものを示す。カラーバーは  $M$  側のフレーム番号を表す。まず、

各コストの有用性について確かめた。代表して関節 head について示す。図 6 上図にコストを  $D_m$  のみにした場合を示す。ループ軌道の部分が大きく対応付けに失敗している。入れ替わりや1点に集中して割り当てられていることも多い。図 6 下図にコストを  $D_m, C_{m1}, C_{m2}$  とした場合を示す。ループ軌道の部分はうまく対応付けることができ、対応付けの入れ替わりも存在しない。他の関節に関しても入れ替わりが生じないことが確認できた。このことは  $C_{m1}$  と  $C_{m2}$ , 特に  $C_{m1}$  が入れ替わりを阻止する役割を強く担っていることを表している。 $C_{m0}$  がなくてもパラメータの重みを適切に調整すれば入れ替わりは起こらない。図 7 にさらにコスト  $C_{m0}$  を付加した場合の全関節の対応付け結果を示す。

### 8.3 非線形最適化による遮蔽関節位置の推定

第 7.2 節で与えた最適化問題を COBYLA 法で解いた。パラメータを以下のように設定した。

$$j_{re} : \text{head}$$

$$\text{上限値 } d : 5$$

また、階層構造の最上位関節を hip とした。

### 8.4 画像への投影

あらかじめ撮影しておいたチェスボード映像から内部パラメータと歪みパラメータを、非和装時の 3 次元データとそれに対応する演技画像を用いて外部パラメータを求めた。パラメータは以下になっている。

チェスボードの格子点間隔 : 38.6 (ただし,  $\pm 0.03$  程度の誤差あり)

チェスボード画像数  $K_c : 4$

演技画像数  $K_p : 5$

和装時の画像は、代表して関節 head で、非和装時のデータと対応付けされた和装時のデータに対応する画像を選んだ。

### 8.5 結果

#### 8.5.1 非和装時データの投影結果

非和装時の関節 3 次元位置の投影結果を図 8 に示す。この画像は複数のフレームの投影結果を 1 枚の画像に表示している。これにより各種カメラパラメータが正確に求められていることが確認できる。

#### 8.5.2 IK 後の 3 次元データを和装時画像に投影した結果

図 9 に、和装状態で撮影した画像へ、非和装時の全関節 3 次元データを投影したもの (緑点), それらを第 7.2 節に示す最適化で変換したデータを投影したもの (赤点), 和装時の関節  $j \in J'$  の 3 次元データを投影したものを示す。1 つ目のフレームはうまく求まっているように見える。2 つ目と 3 つ目は青点自体が関節位置からずれているので関節 head の対応付けから求めた和装画像とデータ



図 8 非和装時データの投影結果



図 9 IK 後データの投影結果

間のフレーム対応がうまくいっていない可能性が高い。また、4つ目の right shoulder のように非和装時と和装時のポーズのずれから遮蔽部位の位置推定がずれることが投影結果から分かる。

## 9. 考察

$M$  から  $C$  への対応関係を求めたが、 $C$  の方が総フレーム数が多く、 $M$  の複数のデータが  $C$  の1つのデータに割り当てられることもあり、対応付けを受けた  $C$  のデータが飛び飛びになる。それゆえに  $C$  の各データと和装時の演技映像の各フレームの画像との対応がうまく取れない。再サンプリングなどの工夫が必要と考えられる。

また、フレーム番号によるコストは非和装時と和装時の演技速度が殆ど同じであるから有効となりうる。今回用いたシーケンスは演技速度が殆ど同じであったが、演技速度が少しずれた演技シーケンスもあった。そこで、フレーム番号ではなく累積距離による代用も試みたが、第 8.2.1 節で述べた信念伝搬法 (Belief Propagation) と最急降下法による2段階の最適化が局所解に陥ってしまった。最適化手法についても更に検討をする必要がある。

## 10. 結論

本論文では、遮蔽部位を含む人体関節の3次元位置推定を目的として、和装時と非和装時の2通りのモーションキャプチャで得られた3次元運動系列間での対応付けと、その対応関係に基づいた逆運動学による関節位置推定手法

を提案し、実験によってその有効性を検証した。

提案手法により画像を入力とした和装人物の3次元骨格姿勢推定のための学習データ作成が可能となった。今後は

- 計測された運動軌跡の再サンプリングなどによる、より滑らかな対応付けの実現
- 運動軌跡上で特徴的な箇所抽出と、それらの間の事前対応付けによる、最適化問題への制約の導入
- 逆運動学による遮蔽関節位置推定における時間的連続性の考慮と、推定位置が撮影画像上で対象領域内部へと投影されることの保証

などを検討するとともに、実際に機械学習による和装人物の3次元骨格姿勢推定が可能か否かを検証することが今後の課題である。

謝辞 本研究の一部は京都大学インキュベーションファンド採択事業「デジタルヒューマン」の助成を受けて行った。

## 参考文献

- [1] Yano, T., Nobuhara, S. and Matsuyama, T.: 水中物体向け視体積交差法, 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2013-CVIM-187, No. 13, pp. 1-7 (2013).
- [2] Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F. and Navab, N.: Deeper Depth Prediction with Fully Convolutional Residual Networks, *arXiv* (2016).
- [3] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Real-time Multi-Person 2D Pose Estimation using Part Affinity Fields, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7291-7299 (2017).
- [4] Güler, R. A., Trigeorgis, G., Antonakos, E., Snape, P., Zafeiriou, S. and Kokkinos, I.: DenseReg: Fully Convolutional Dense Shape Regression In-the-Wild, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6799-6808 (2017).
- [5] Nils Hasler et al.: The capture, <http://thecapture.com/>.
- [6] Kolmogorov, V.: Convergent Tree-Reweighted Message Passing for Energy Minimization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 10, pp. 1568-1583 (online), DOI: 10.1109/TPAMI.2006.200 (2006).
- [7] Besag, J.: On the Statistical Analysis of Dirty Pictures, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 48, No. 3, pp. 259-302 (1986).