

修士論文

陸上 100m 走における単眼カメラと 高精度 3 次元人体モデルを用いた 3 次元位置姿勢推定

石川 拓志

奈良先端科学技術大学院大学

先端科学技術研究科

知能社会創成科学プログラム

主指導教員: 加藤 博一 教授

インタラクティブメディア設計学 研究室（情報科学領域）

令和 2 年 3 月 12 日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に
修士 (工学) 授与の要件として提出した修士論文である.

石川 拓志

審査委員：

加藤 博一 教授	(主指導教員, 情報科学領域)
小笠原 司 教授	(副指導教員, 情報科学領域)
神原 誠之 准教授	(副指導教員, 情報科学領域)
藤本 雄一郎 助教	(副指導教員, 情報科学領域)

陸上 100m 走における単眼カメラと 高精度 3 次元人体モデルを用いた 3 次元位置姿勢推定*

石川 拓志

内容梗概

陸上 100m 走競技において、走行姿勢を定量的に計測することは重要である。計測結果から良い走行動作、悪い走行動作を定量的に分析することによって、走行動作の相違を数値で理解することができ、改善に有効な分析やトレーニング策定を行うことができる。しかし、現在の計測手法は特殊装置を用いなければならず実際のトレーニング現場で使用することは難しい。

本研究では 1 台の RGB カメラと高精度人体 3 次元モデルを用いて走行動作を計測する手法を提案する。本手法は、撮影された画像から陸上 100m 走環境における人物位置を特定し、その後画像内の人物領域と CG 画像内の高精度人体 3 次元モデル領域を重ね合わせることで、位置姿勢を推定する。位置姿勢を推定する際、陸上 100m 走という限られた状況を利用し、多自由度の姿勢パラメータ推定問題を取り扱いやすい問題に置き換える処理を行う。実験では、目標画像を CG 画像とするシミュレーション環境において、真値との比較を行った。その結果、単一フレーム処理において、脚関節の誤差が大幅に減少する効果が確認できた。

キーワード

陸上 100m 走, RGB カメラ, 人体 3 次元モデル, 3 次元位置姿勢推定, CG

*奈良先端科学技術大学院大学 先端科学技術研究科 修士論文, 令和 2 年 3 月 12 日.

Accurate 3D pose estimation with monocular camera and accurate 3D human model in 100m sprint scene*

Hiroshi Ishikawa

Abstract

It is important to measure 3D pose quantitatively in 100m sprint scene. Analysing good pose or bad pose with quantitative data allows us to understand the difference between poses and to practice efficiently. However, traditional method to analyze 3D pose quantitatively needs special facility, so it is difficult to use in practical scene.

This thesis proposes 3D pose estimation with monocular RGB camera and accurate 3D human model in 100m sprint scene. This method specifies the position of a person in 100 m scene from the captured image, and then estimates the 3D pose by fitting human 3D model region in the CG image into the person region in the image. When estimating 3D pose, by using the limited situation of 100m scene to replace the multi-degree-of-freedom pose estimation problem with an easy-to-handle problem. In the experiment, The effect of significantly reducing errors in the leg joint was confirmed..

Keywords:

100m sprint, RGB camera, 3D human model, 3D pose estimation, CG

*Master's Thesis, Graduate School of Science and Technology, Nara Institute of Science and Technology, March 12, 2020.

目次

図目次	v
第 1 章 序論	1
1.1 研究背景	1
1.2 研究目的	3
1.3 本論文の構成	3
第 2 章 関連研究	4
2.1 RGB カメラを用いた人物の 3 次元位置姿勢推定	4
2.1.1 人体モデルを用いない推定手法	4
2.1.2 人体モデルを用いた推定手法	5
2.2 陸上短距離走への応用	6
2.3 本研究の位置付け	7
第 3 章 1 台の RGB カメラと高精度 3 次元人体モデルを用いた 3 次元位置姿勢推定	8
3.1 提案手法の概要	8
3.2 人物位置推定 (初期位置推定)	9
3.2.1 3 次元既知点を利用したカメラの位置姿勢推定	12
3.2.2 人物領域特徴点を利用した人物の位置推定	13
3.3 人物位置姿勢推定	15
3.3.1 実画像人物領域のセグメンテーション	15
3.3.2 CG 画像 (3D モデル) のレンダリング	16
3.3.3 実画像と CG 画像の人物領域類似度計算	18
3.3.4 非線形最適化による位置姿勢パラメータ推定	19
第 4 章 陸上短距離走という限定された状況を利用した最適化問題の置き換え	21
4.1 主成分分析による次元削減	21

4.2	学習データとの差分推定	22
第 5 章	実験	24
5.1	人物位置推定 (初期位置推定)	24
5.1.1	実験環境	24
5.1.2	実験結果	26
5.2	人物位置姿勢推定	27
5.2.1	実験環境	27
5.2.2	実験結果	28
	単一フレーム処理	29
	動画像処理	32
5.3	考察	36
第 6 章	おわりに	37
6.1	まとめ	37
6.2	今後の課題と展望	37
謝辞		39
参考文献		40

図目次

1.1	モーションキャプチャ装置の使用例 [8]	2
1.2	接地測定装置の使用例 [5]	2
2.1	3次元人体モデルを用いない3次元位置姿勢推定手法の例 [12]	5
2.2	3次元人体モデルを用いた3次元位置姿勢推定の例 [16]	6
2.3	走行姿勢のマッチング [29]	7
3.1	提案手法の流れ	9
3.2	座標系の関係	11
3.3	走行動作の撮影例	11
3.4	PnP 問題 [4]	12
3.5	Structure from Motion [21]	13
3.6	人物領域の切り取り部分	16
3.7	セグメンテーション結果	16
3.8	撮影時の様子	17
3.9	取得した人体 3D モデル	17
3.10	スケルトンモデル	17
3.11	画像の人物領域に対する Dice の適用	19
3.12	最適化対象の 17 関節	20
4.1	時間軸最適化の流れ	23
5.1	人物位置推定対象の画像	25
5.2	人物位置推定の実験環境	25
5.3	切り出した画像	26
5.4	セグメンテーション結果	26
5.5	人物位置姿勢推定の実験環境	28
5.6	横 655 画素・縦 472 画素の画像	28
5.7	単一フレーム処理の結果	30

5.8	単一フレーム処理の結果 (正面から見た姿勢)	30
5.9	類似度の推移	33
5.10	動画像処理の結果	33
5.11	動画像処理における 24 フレーム時 (最も高い類似度で収束) の結果	34
5.12	動画像処理における 24 フレーム時 (最も高い類似度で収束) の結 果 (正面から見た姿勢)	34

第 1 章 序論

1.1 研究背景

陸上短距離 100m 競技において、アスリートはゴールタイム向上のために日々トレーニングに取り組んでいる。競技力を改善し、ゴールタイムを向上させるには、走行姿勢を定量的に計測・分析することが重要である。走行動作を定量的に計測することによって、望ましい走行動作、望ましくない走行動作の相違を数値で理解することができ、改善に有効な分析やトレーニング策定を行うことができる。

現在、走行動作を定量的に分析する方法として、姿勢を計測するための特殊装置を用いる手法がある。使用される手法の代表例としては、モーションキャプチャ [11, 15] や接地測定装置 [20, 26] を用いた手法などがある。モーションキャプチャシステムでは、図 1.1 のように対象人物の体にマーカを貼り付けることで、位置姿勢のデータを取得する。また、接地測定装置は図 1.2 のようなセットアップを行うことで、足の設置時間や着地地点を計測することができる。しかし、このような特殊装置を使用すれば、セットアップが煩雑、実際の競技環境の再現性が低いなどの問題が発生してしまう。また、特殊な装置を使用できる環境は限られており、陸上短距離走を行う多くのアスリートは利用することが難しい。そこで、上記のような特殊な装置を使わずに走行動作を定量的に分析する方法が望まれている。



図 1.1 モーションキャプチャ装置の使用例 [8]

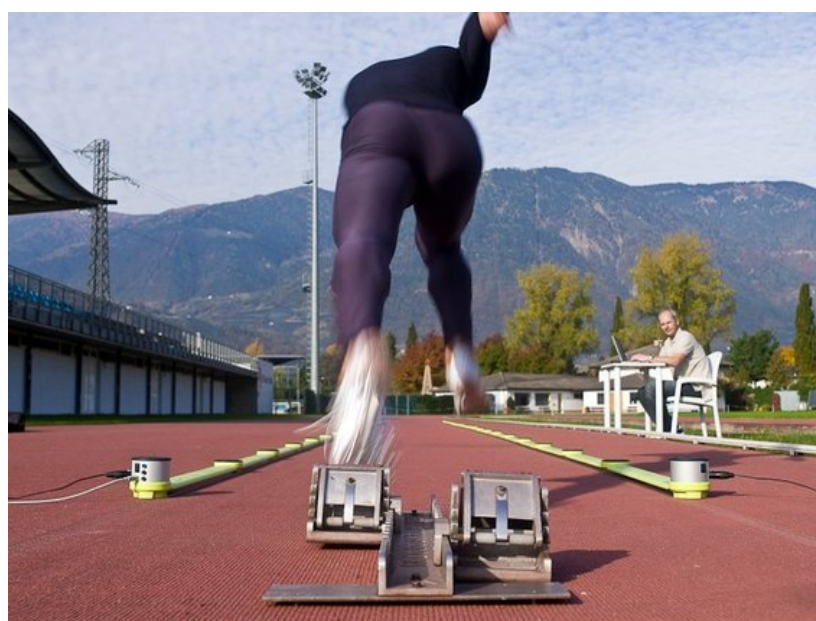


図 1.2 接地測定装置の使用例 [5]

1.2 研究目的

本研究では、特殊装置を使用することで発生する問題を解決するため、特殊装置を使用せずに走行時のアスリートの3次元位置姿勢を高精度に求める手法の実現を目的とする。そこで、RGBカメラ1台と高精度にパーソナライズされた人体3Dモデルを使用して、陸上短距離走におけるアスリートの3次元位置姿勢を推定する手法を提案する。特殊装置を使用しないという目的を達成するため、RGBカメラ1台で3次元位置姿勢を行う。また、高精度な3次元位置姿勢推定の実現という目的を達成するため、高精度にパーソナライズされた人体3次元モデルを使用する。高精度にパーソナライズされた人体3次元モデルが事前に準備されていることを前提とし、実画像の人物領域と、人体3次元モデルがレンダリングされたCG画像の人物領域を重ね合わせ、正確に重なった時の人体3Dモデルの位置姿勢パラメータを出力値とすることで、高精度な3次元位置姿勢推定を試みる。

実験では、実画像の役割を、人体3Dモデルを使用してレンダリングしたCG画像とすることで、関節角度における真値との比較を行い、3次元位置姿勢推定の精度を評価する。

1.3 本論文の構成

2章では、RGBカメラを用いた3次元位置姿勢推定における関連研究と、RGBカメラを用いた3次元位置姿勢推定が陸上短距離走に応用されている研究を紹介し、それらの問題を述べる。3章では提案手法の流れと、それぞれの処理を説明し、4章では陸上100m走という限定された状況を利用することで多自由度の位置姿勢推定問題を取り扱いやすい問題に置き換える方法を述べる。5章では実験の結果を示し、考察を行う。最後に6章では本研究のまとめと今後の課題、展望を述べる。

第 2 章 関連研究

本研究に関連する研究として、RGB カメラを用いた 3 次元人体位置姿勢推定に関する研究を紹介する。RGB カメラを用いた 3 次元人体位置姿勢推定に関する研究は、3 次元人体モデルを使わずに推定する手法と 3 次元人体モデルを用いて推定する手法に分けることができる。さらに、RGB カメラを用いた 3 次元位置姿勢推定を陸上短距離走に応用した研究を紹介する。

2.1 RGB カメラを用いた人物の 3 次元位置姿勢推定

2.1.1 人体モデルを用いない推定手法

人体モデルを使わずに 3 次元位置姿勢を推定する手法は、直接 3 次元座標位置を回帰する研究 [9] や、3 次元ヒートマップを用いる研究 [19, 24]、2 次元姿勢位置推定 [28, 10] の結果を使用して、3 次元位置姿勢を推定する研究 [12, 18, 30] などが存在する。これらの研究は、一度推論モデルを学習すれば、RGB カメラで対象人物を撮影するだけで図 2.1 のように人物のスケルトンモデルを取得することができるため、様々な場面での活用が可能である。

しかし、2 次元位置姿勢情報を使用せずに 3 次元位置姿勢を直接推定する手法に関しては、大規模な学習データセットの構築が困難という問題がある。また、2 次元位置姿勢情報を用いて 3 次元位置姿勢を推定する手法に関しては、2 次元位置姿勢情報の精度に依存するが、2 次元位置姿勢情報の学習データセット作成は人手によって行われることが多く、誤差が発生する可能性がある。このような問題から、陸上競技における 3 次元位置姿勢推定にこれらの研究を適用することは、精度の点で問題がある。

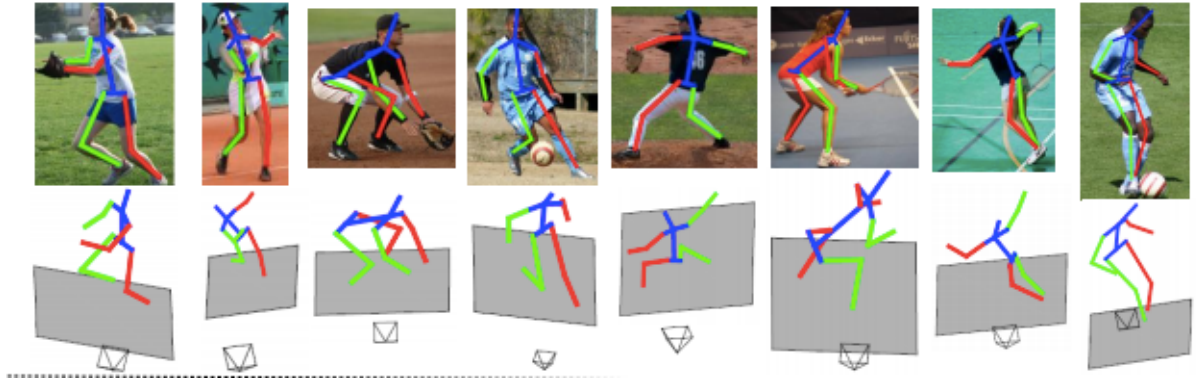


図 2.1 3次元人体モデルを用いない3次元位置姿勢推定手法の例 [12]

2.1.2 人体モデルを用いた推定手法

人体モデルを用いた3次元位置姿勢推定は、事前に用意された人体モデルのパラメータ推定を行うことで3次元位置姿勢推定を行う。人体モデルを用いた推定手法は、スケルトンモデルを使う手法と、メッシュモデルを使う手法の大きく2つに分けることができる。文献 [23] は、対象人物に対する2次元位置姿勢推定を複数台のカメラで行い、その結果を用いて人体スケルトンモデルを作成し、作成したスケルトンモデルの姿勢パラメータを推定している。メッシュ情報を持つ人体モデルを使用する研究 [16, 13] では、画像の特徴量から、図 2.2 のようなメッシュ情報を持った人体モデルのパラメータを推定することで、3次元位置姿勢を推定している。しかし、これらのメッシュ情報を持つ人体モデルは特定の個人に最適化しておらず、モデルには誤差が発生する。そのようなモデルを使用した3次元位置姿勢推定では、結果に誤差が生じる。



図 2.2 3次元人体モデルを用いた3次元位置姿勢推定の例 [16]

2.2 陸上短距離走への応用

RGB カメラを用いた人体位置姿勢推定を陸上短距離走に応用した研究 [29] では、手法 [10] を用いて2次元人体位置姿勢推定を行い、出力された4つのキーポイント(右手首・左手首・右足首・左足首)のユークリッド距離を使用して定義された類似度を用いて走行姿勢のマッチングを行なっている。走行動作の特徴である周期性を利用することで、手法 [10] の結果に対してロバストなマッチングを実現している。図 2.3 は、50m 地点の走行動作の姿勢と 90m 時点の走行動作の姿勢のマッチングを行なった結果である。手法 [10] による関節検出の精度が低い場合でも、類似している姿勢にマッチングしていることが分かる。しかし、この研究では、姿勢のマッチングのみを行なっており、姿勢の定量的な分析ができていない。

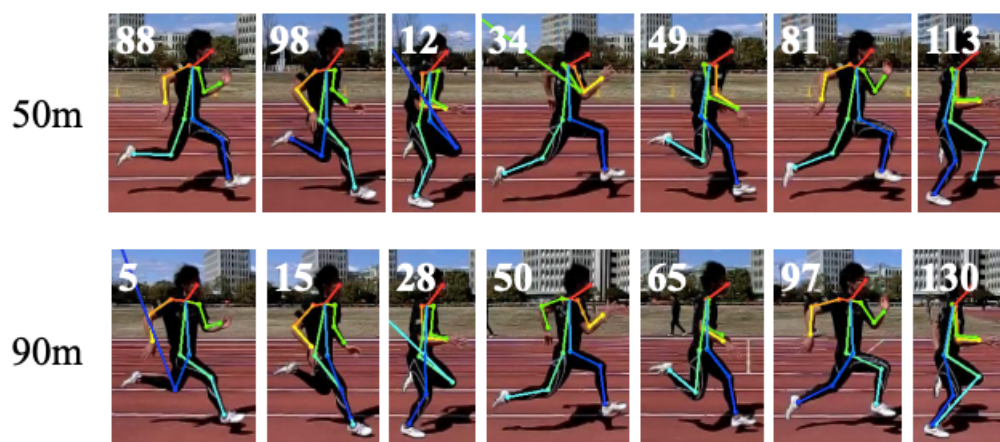


図 2.3 走行姿勢のマッチング [29]

2.3 本研究の位置付け

本研究では、陸上 100m 走において、特殊装置を使用せずに高精度な 3 次元人体位置姿勢推定を実現することを目的としている。したがって、個人に最適化された高精度人体メッシュモデルを使用し、陸上 100m 走という限定された状況を利用する 3 次元人体位置姿勢推定手法を提案する。本研究の対象は、特定の陸上 100m 走のトップアスリートである。本研究の位置付けを表 2.1 に示す。

表 2.1 本研究の位置付け

3 次元位置姿勢推定手法		汎用シーン対象	陸上シーン対象
人体モデル無し		関連研究 [9] 等	関連研究 [29]
人体モデル有り	汎用モデル	関連研究 [16] 等	なし
	最適化モデル	なし	本研究

第 3 章 1 台の RGB カメラと高精度 3 次元人体モデルを用いた 3 次元位置姿勢推定

3.1 提案手法の概要

本章では、1 台の RGB カメラと事前に用意された高精度個人最適化人体 3 次元モデルを用いた 3 次元人体位置姿勢推定手法について説明する。提案手法では、陸上 100m 走環境における対象人物の位置と姿勢を推定する。そのため、本手法は対象人物の位置を推定するための処理群と、求めた位置を初期値として対象人物の位置姿勢を推定する処理群の大きく 2 つに分かれる。提案手法の流れを図 3.1 に示す。人物位置推定では、3 次元シーンの点が既知であるという前提で、それらの点を利用してカメラの位置姿勢推定を行う。ここで、撮影カメラの内部パラメータは事前のカメラキャリブレーションで既知とする。そして、求めたカメラの位置姿勢と人物領域特徴点、直線上を走る陸上 100m 走の状況を利用して、人物の位置を求める。人物位置姿勢推定では、実画像の人物領域と、人体 3 次元モデルがレンダリングされた CG 画像の人物領域を重ね合わせ、正確に重なった人体 3 次元モデルの位置姿勢パラメータを推定する。そのために必要な処理として、実画像人物領域のセグメンテーション、CG 画像 (人体 3 次元モデル) のレンダリング、実画像と CG 画像の人物領域類似度計算、非線形最適化による位置姿勢パラメータ推定を行う。本章では、それぞれの処理に関して詳細に説明する。

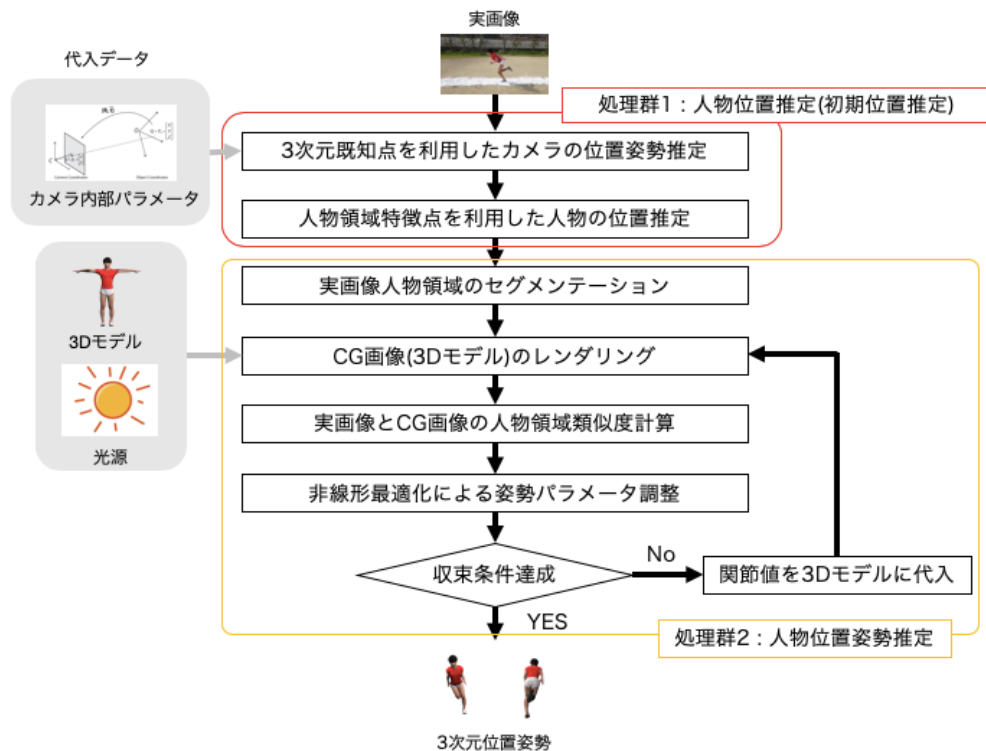


図 3.1 提案手法の流れ

3.2 人物位置推定 (初期位置推定)

撮影画像から陸上 100m 走環境における人物位置を推定する。陸上 100m 走において走行動作のスタート位置を原点とする座標系を設定し、その座標系での人物位置を陸上 100m 走環境における人物位置とすることで、人物位置の推定を行う。本手法では、走行動作のスタート位置を原点とする座標系をワールド座標系、撮影カメラが原点の座標系をカメラ座標系、対象人物の腰を原点とした座標系をオブジェクト座標系と呼ぶ。求めたいものは、ワールド座標系におけるオブジェクト座標系原点の位置である。ワールド座標系におけるオブジェクト座標系の位置を求めるため、まずワールド座標系におけるカメラの位置姿勢を 3 次元既知点を利用して推定する。ここで、3 次元既知点とは、ワールド座標系における座標値が既知の点とする。そして、求めたカメラの位置姿勢と人物領域特徴点を利用して人物位置を推定する。ここで、シーンの前提として、3 次元既知点が 6 点以上あるとし、撮影カメ

ラの内部パラメータは事前のカメラキャリブレーションにより既知とする。それぞれの処理について、下記に詳細に記述する。図 3.2 に各座標系の関係を示す。また、図 3.3 は、実際に走行動作を撮影した画像である。図 3.3 の画像サイズは縦 2160 画素、横 3840 画素であり、カメラ機器は osmo pocket[2] を利用した。図 3.3 ではサッカーボールをワールド座標系の指定位置に事前に設置することで、3 次元シーンにおける既知点を作成した。

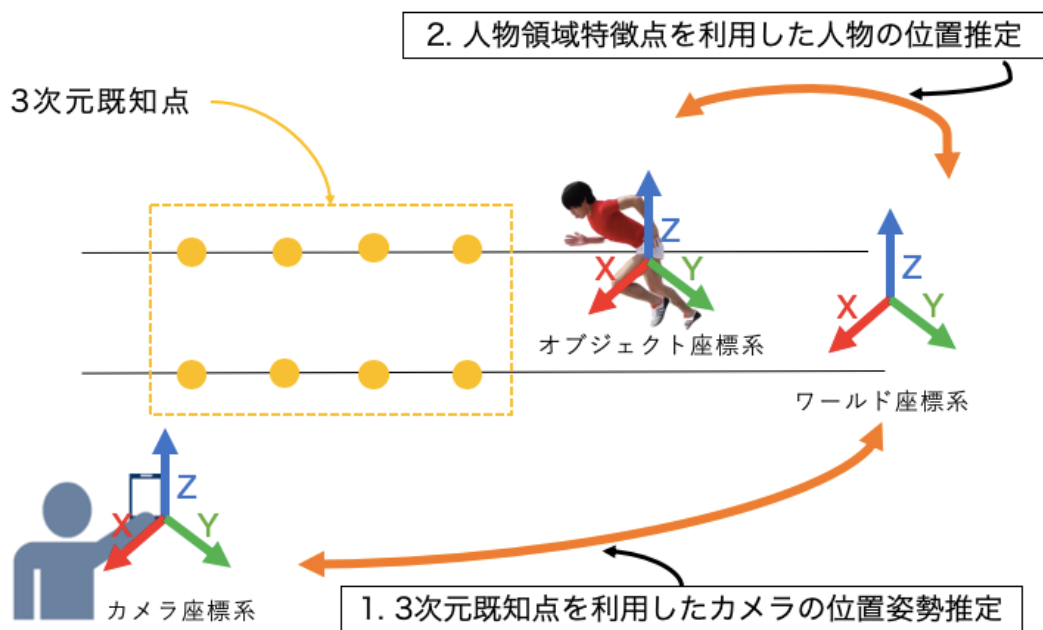


図 3.2 座標系の関係



図 3.3 走行動作の撮影例

3.2.1 3次元既知点を利用したカメラの位置姿勢推定

ワールド座標系における座標値が既知の点が6点以上あるという前提の下、ワールド座標系におけるカメラの位置姿勢を求める。ワールド座標系座標値が既知の点が画像内に6点以上写っていれば、PnP問題として定式化でき、カメラの位置姿勢は求まる。PnP問題とは、図3.4が示すように、3次元位置が既知の n 点とそれに対応する画像上の n 点からカメラの位置姿勢を求める問題である。

しかし、実際の3次元既知点が画像内に写らない場合、PnP問題として定式化できない。そこで、撮影された走行動作の動画像において、画像間の対応位置からカメラの相対的な位置姿勢を求めておき、PnP問題が定式化できる画像で求めたカメラ位置姿勢と、相対的なカメラ位置姿勢の関係から、動画像全フレームにおけるワールド座標系のカメラの位置姿勢を求める。ここで、画像間の対応位置からカメ

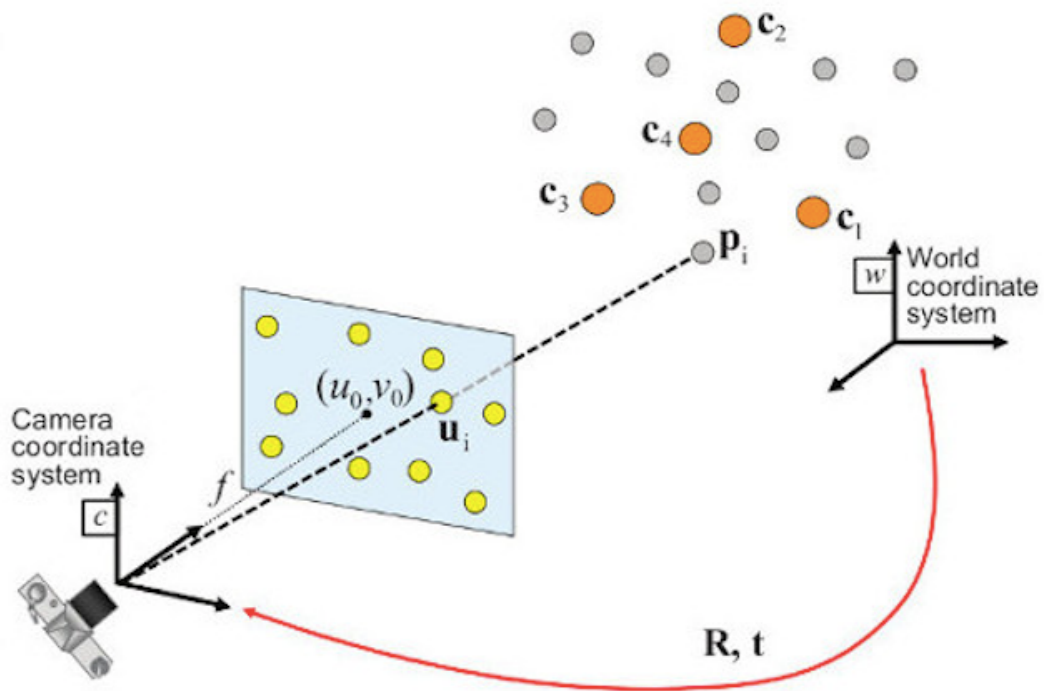


図 3.4 PnP 問題 [4]

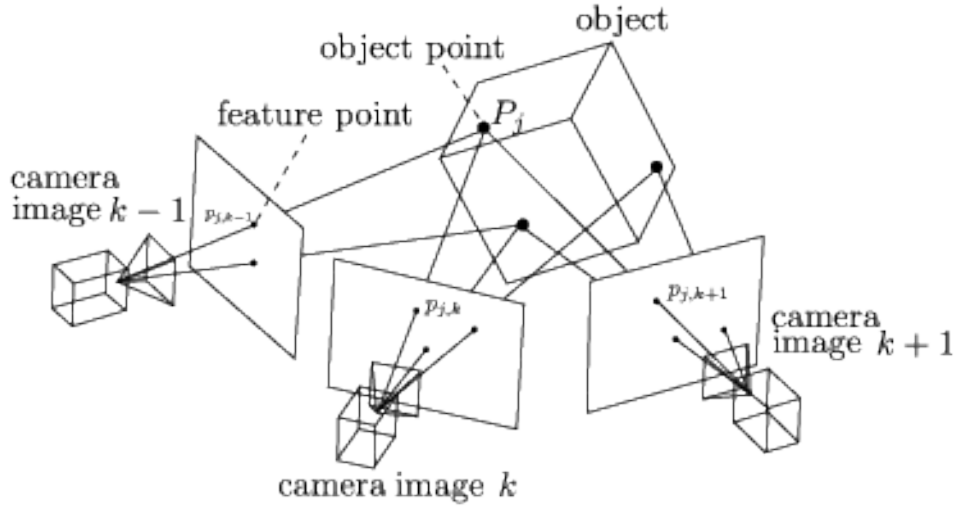


図 3.5 Structure from Motion [21]

ラの相対的な位置姿勢を求める際、Structure from Motion[21] の問題 (図 3.5) として定式化する。Structure from Motion では、実画像の大きさと 3 次元シーンにおける大きさの対応関係を決めるパラメータであるスケールは求まらない。スケールを求めるためには、画像内の特徴点の動きに対応する 3 次元シーンの動きや、画像内び複数の特徴点の距離に対応する 3 次元シーンのサイズが必要である。本手法では、3 次元シーンにおいて既知の点が 6 点以上あるという前提であるため、それら既知の点からスケールを求める。Structure from Motion を行う際は、オープンソースを利用すれば比較的簡単にパラメータを求めることができ、OpenMVG[22] などが利用可能である。

3.2.2 人物領域特徴点を利用した人物の位置推定

3.2.1 で求めたカメラ位置姿勢と、画像座標系における人物領域の特徴点座標値から、ワールド座標系における対象人物の位置を推定する。人物領域の特徴点は、セグメンテーション等の技術を利用して求めることができると仮定する。そして、対象人物はワールド座標系における横軸上を移動していると仮定すると、画像座標系における特徴点座標値は、カメラ内部パラメータとカメラ位置姿勢パラメータを

用いてワールド座標系における座標値に変換できる。

画像座標系の特徴点座標値と、ワールド座標系の特徴点座標値の関係式について 3.2.1 に示す。特徴点について、カメラ座標系における位置を (x_c, y_c, z_c) 、ワールド座標系における位置を (x_w, y_w, z_w) とすると、

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (3.2.1)$$

ここで、 f をカメラの焦点距離、 (c_x, c_y) を主点 (principal point)、ワールド座標系とカメラ座標系の関係を表す外部パラメータを r, t を用いて表す。

また、スケールパラメータ h を導入し、画像座標系におけるピクセル位置を (x_p, y_p) とすると、カメラ座標系の値は以下の式で表せる。

$$\begin{pmatrix} hx_p \\ hy_p \\ h \\ 1 \end{pmatrix} = \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} \quad (3.2.2)$$

よって、3.2.1 と 3.2.2 より、以下の式が導出できる。

$$\begin{pmatrix} hx_p \\ hy_p \\ h \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} \quad (3.2.3)$$

ここで、求めたいパラメータは (x_w, y_w, z_w) である。 h 以外のパラメータは既知であるため、式変形してまとめると、

$$\begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11}h + a_{12} \\ a_{21}h + a_{22} \\ a_{31}h + a_{32} \\ 1 \end{pmatrix} \quad (3.2.4)$$

ここで、 a で表されるパラメータ群は既知である。また、人物はワールド座標系において $y_w = 0$ の軸に沿って走行を行うと仮定すると、

$$\begin{pmatrix} x_w \\ 0 \\ z_w \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11}h + a_{12} \\ a_{21}h + a_{22} \\ a_{31}h + a_{32} \\ 1 \end{pmatrix} \quad (3.2.5)$$

となり、パラメータ h の値が決まる。 h の値が決まると、 (x_w, y_w, z_w) の値が求まり、ワールド座標系における人物領域の特徴点座標値が求まる。

3.3 人物位置姿勢推定

3.2 で求めたカメラ位置姿勢と人物位置を使用して、人物位置姿勢推定を行う。人物位置姿勢推定では、実画像の人物領域と、人体 3 次元モデルがレンダリングされた CG 画像の人物領域を重ね合わせ、正確に重なった人体 3 次元モデルの姿勢パラメータを推定する。そのために必要な処理として、実画像人物領域のセグメンテーション、CG 画像 (人体 3 次元モデル) のレンダリング、実画像と CG 画像の人物領域類似度計算、非線形最適化による姿勢パラメータ推定を行う。セグメンテーションにより実画像の人物領域を抜き出し、CG 画像の人体 3 次元モデル領域と実画像の人物領域がどの程度重なっているかを類似度計算で確認する。そして、その類似度の値を目的関数として最適化を行うことで、実画像の人物領域と、CG 画像の人体 3 次元モデル領域を重ね合わせる。ここで、最適化の初期値となる姿勢データは、事前 to 取得した学習データを参照することで求めることができると仮定する。

3.3.1 実画像人物領域のセグメンテーション

セグメンテーションとは、画像をピクセル単位で分類する技術のことで、対象物体と背景を切り分けて認識するような際に用いられる。近年は深層学習の発展で精度の高い分類が可能になっており、オープンソースや web サービスとして利用できる手法 [14][6] が増加している。本手法では、3.3.3 で後述する類似度を計算する際に、人物領域のセグメンテーションが必要である。そこで、remove.bg[6] という web サービスを用いたセグメンテーションのテスト結果を図 3.7 に示す。ここで、図 3.6 は今回撮影した画像を拡大した画像である。非常に精度の高いセグメンター



図 3.6 人物領域の切り取り部分



図 3.7 セグメンテーション結果

ション結果が確認できる。

3.3.2 CG 画像 (3D モデル) のレンダリング

3.2 の処理を行うことで、ワールド座標系におけるカメラの位置姿勢と対象人物の位置が求まる。これらのパラメータと人体 3D モデルを使用して CG 画像をレンダリングする。本論文では、人体 3D モデルは superscanstudio の設備 [7] で取得したものを使用する。図 3.8 は、実際に撮影した様子を示している。また、図 3.9 は今回使用した人体 3D モデルであり、図 3.10 は人体のスケルトンモデルである。



図 3.8 撮影時の様子



図 3.9 取得した人体 3D モデル

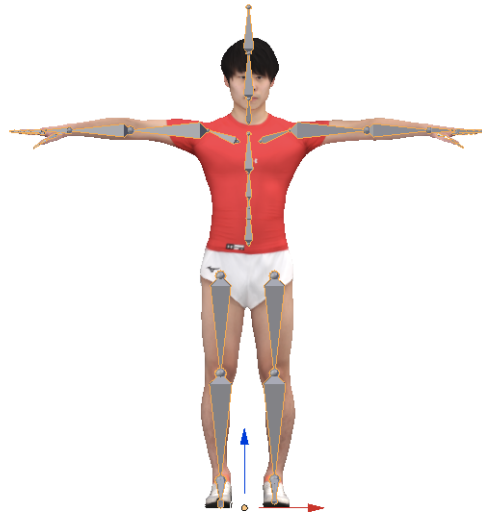


図 3.10 スケルトンモデル

3.3.3 実画像と CG 画像の人物領域類似度計算

3D モデルをレンダリングした後、CG 画像における人体 3D モデルの領域と、撮影画像の人物領域の類似度計算を行う。本手法では、テンプレートマッチングで用いられる基本的な類似度指標の一つである Zero-means Normalized Cross-Correlation(以下 ZNCC) と、領域の重なりを示す Dice を利用する。そして、ZNCC の値と Dice の値を掛けた値を類似度として定義する。ZNCC を用いることで人物領域における画素の情報を利用することができ、Dice を用いることで重なっている人物領域面積の情報を利用することができる。

- ZNCC

ZNCC は以下の式 3.3.1 で表される。

$$\text{ZNCC} = \frac{\sum_x \sum_y (I(x, y) - \mu_I)(T(x, y) - \mu_T)}{\sqrt{\sum_x \sum_y (I(x, y) - \mu_I)^2} \sqrt{\sum_x \sum_y (T(x, y) - \mu_T)^2}} \quad (3.3.1)$$

ここで、 $I(x, y)$, $T(x, y)$ は画像座標 (x, y) における画素値を表し、 μ_I , μ_T は対象領域 I , T の平均画素値を表す。ZNCC では、 I , T をそれぞれ画像領域内の濃度値分布と考え、統計量としての相互相関係数を計算する。

- Dice

Dice は以下の式 3.3.2 で表され、集合に対する類似度として使用されることが多い。

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (3.3.2)$$

ここで、 X は集合 X の要素数、 Y は集合 Y の要素数を表し、 $|X \cap Y|$ は集合 X と集合 Y に共通する要素の数を表す。したがって、Dice は 2 つの集合の平均要素数と共通要素数の割合を表す。

本手法では、画像の人物領域に対して Dice の指標を用いる。よって、人物領域の重なりが大きければ値が大きく、重なりが小さければ値が小さくなる。図 3.11 に Dice を画像の人物領域に適用する流れを示す。

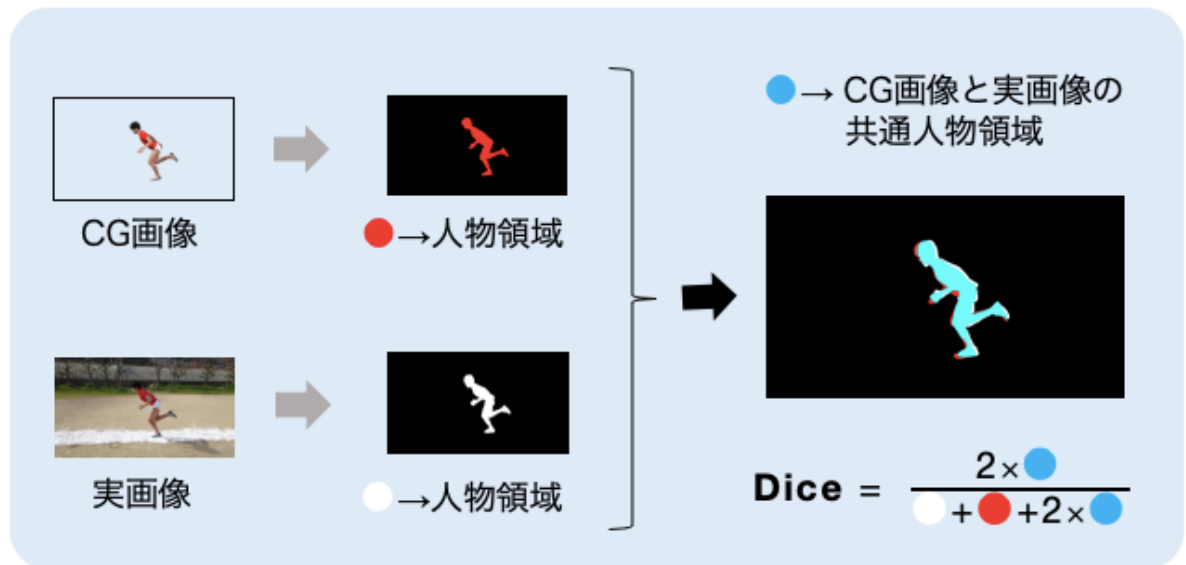


図 3.11 画像の人物領域に対する Dice の適用

3.3.4 非線形最適化による位置姿勢パラメータ推定

3.3.3 で定義した類似度を目的関数、人体 3 次元モデルのパラメータを決定変数として定式化し、数理最適化 (以下最適化) を行う。ここで、人体 3 次元モデルのパラメータとは、オブジェクト座標系原点の位置とオブジェクト座標系のワールド座標系に対する姿勢、そして 3 次元モデルのそれぞれの関節角度のことであり、関節角度の自由度は 1 関節あたり 3 つである。5 章の実験では、オブジェクト座標系原点の位置とオブジェクト座標系のワールド座標系に対する姿勢を最適化対象に含めず、17 関節を最適化対象としているため、関節空間の次元数は $17 \times 3 = 51$ 次元である。図 3.12 における青い関節が最適化対象の関節である。

姿勢パラメータ推定問題は、多自由度の最適化問題に置き換えることができるが、自由度が多くこのままでは解くことが難しい。したがって、トップアスリートの陸上 100m 走という限られた状況であることを利用し、多自由度の最適化問題を取り扱いやすい問題に置き換えて解く必要がある。具体的な方法は 4 章に詳細に記述する。

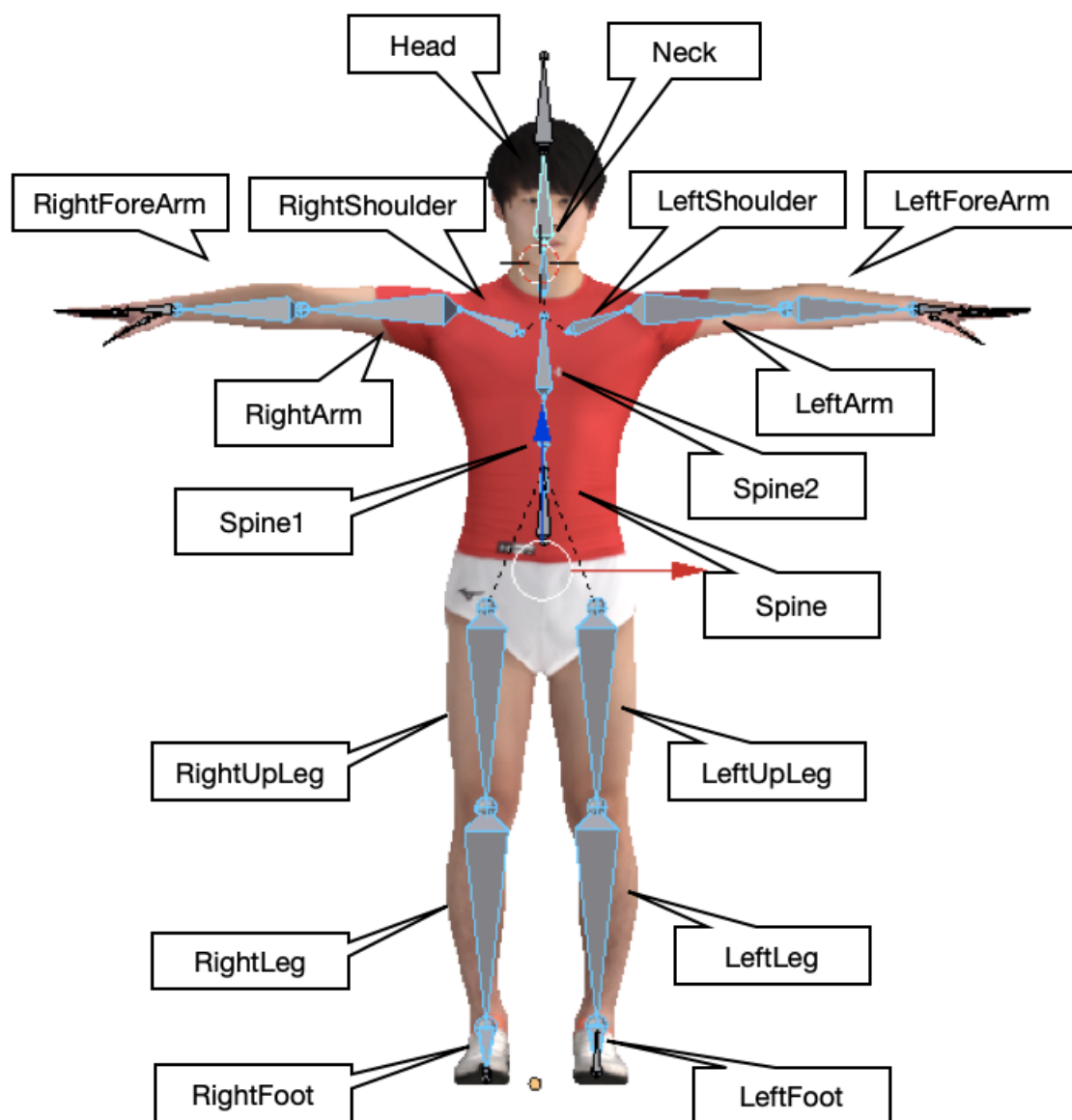


図 3.12 最適化対象の 17 関節

第 4 章 陸上短距離走という限定された状況を利用した最適化問題の置き換え

多自由度の最適化問題は局所解問題が発生するため解くことが難しい。そこで、本研究の対象は陸上 100m のトップアスリートという状況を利用して、多自由度の位置姿勢推定問題を取り扱いやすい問題に置き換えて解く。

多自由度の位置姿勢推定問題を解くことが難しい原因として、多自由度空間のため目的関数が複雑になること、初期値と最適解の距離が遠いことが考えられる。したがって、多自由度空間の次元を削減することで多自由度空間を次元が少ない空間に置き換え、かつ初期値を最適解に近づけた値に置き換える。多自由度空間の次元を削減する方法として、事前取得した走行データを使用して主成分分析を行い、次元削減された固有値空間で最適化を行う。走行動作は周期的な動作で、かつそれぞれの関節の変位は関連するため、主成分分析は有効であると考えられる。また、初期値を最適解に近づけて最適化を行う方法として、トップアスリートにおいてスプリント毎における走行姿勢の変化は極めて微小な変動でしかないという仮定をもとに事前取得した選手個人の走行データを参照し、そのデータとの差分を推定する。ここで、事前取得した走行データのことを学習データと呼ぶこととする。

4.1 主成分分析による次元削減

事前取得した走行データを用いて主成分分析を行い、次元削減された固有値空間で最適化を行う。走行データは、モーションキャプチャで取得する関節角度のことであり、事前にモーションキャプチャを使用して取得しておく。最適化に使用する主成分軸の数は、累積寄与率や実験結果から決めることとする。

ここで、最適化問題の定式化について述べる。目的関数は、3.3.3 で定義した類似度を使用し、決定変数は、固有値空間の主成分軸である。以下の式に示す。

$$\max_{\mathbf{x}} \text{ZNCC}(\mathbf{x}) \times \text{Dice}(\mathbf{x}) \quad (\mathbf{x} = x_1, x_2 \cdots, x_i) \quad (4.1.1)$$

ここで、 $\mathbf{x}$ は決定変数ベクトルを表し、ベクトルの要素は主成分軸の値である。

x_1 は第一主成分軸の値、 x_2 は第二主成分軸の値を表し、 i 次元空間で最適化を行う。

また、目的関数を最適化する手法として、勾配法 [31] を用いる。勾配法は、目的関数の偏導関数を用いて最適解を見つけるアルゴリズムである。勾配法のアルゴリズムは様々存在 [27, 25] するが、本手法では adam[17] と呼ばれるアルゴリズムを使用する。

4.2 学習データとの差分推定

事前取得した走行データを参照し、観測対象のデータと最も似ている走行データを求め、その走行データを初期値として最適化を行うことで、学習データとの差分推定を行う。観測対象のデータと最も似ている走行データを求めるため、多自由度の関節空間の値を 1 次元の時間軸で捉え、その時間軸で最適化を行う。走行動作は周期動作であるため、時間軸最適化を行うことによって、観測対象のデータが学習データ 1 周期の時間軸の中でどの位置の姿勢の値に類似しているかを求める。具体的には、以下の処理を行うことで時間軸最適化を実現する。図 4.1 に流れを示す。図中の赤点が最適化対象の値、青点が事前に取得した走行データである。下記が時間軸最適化のアルゴリズムである。処理手順の番号と図 4.1 に記述されている番号は対応している。

1. 最適化対象姿勢における多自由度の関節値にユークリッド距離が最も近い 2 つの学習データを選択し、距離の比を計算
2. 学習データにおいて、選択した学習データに対応する時間軸の値を選択し、1 で計算した距離の比を用いて最適化対象姿勢に最も似ている学習データの時間軸の値を求める
3. 2 で求めた時間軸の値に対応する多自由度の関節値を学習データから求める
4. 3 で求めた関節値を人体モデルに代入し、目的関数の計算を行い、時間軸で勾配計算、パラメータ更新
5. 1-4 の処理を収束まで繰り返す

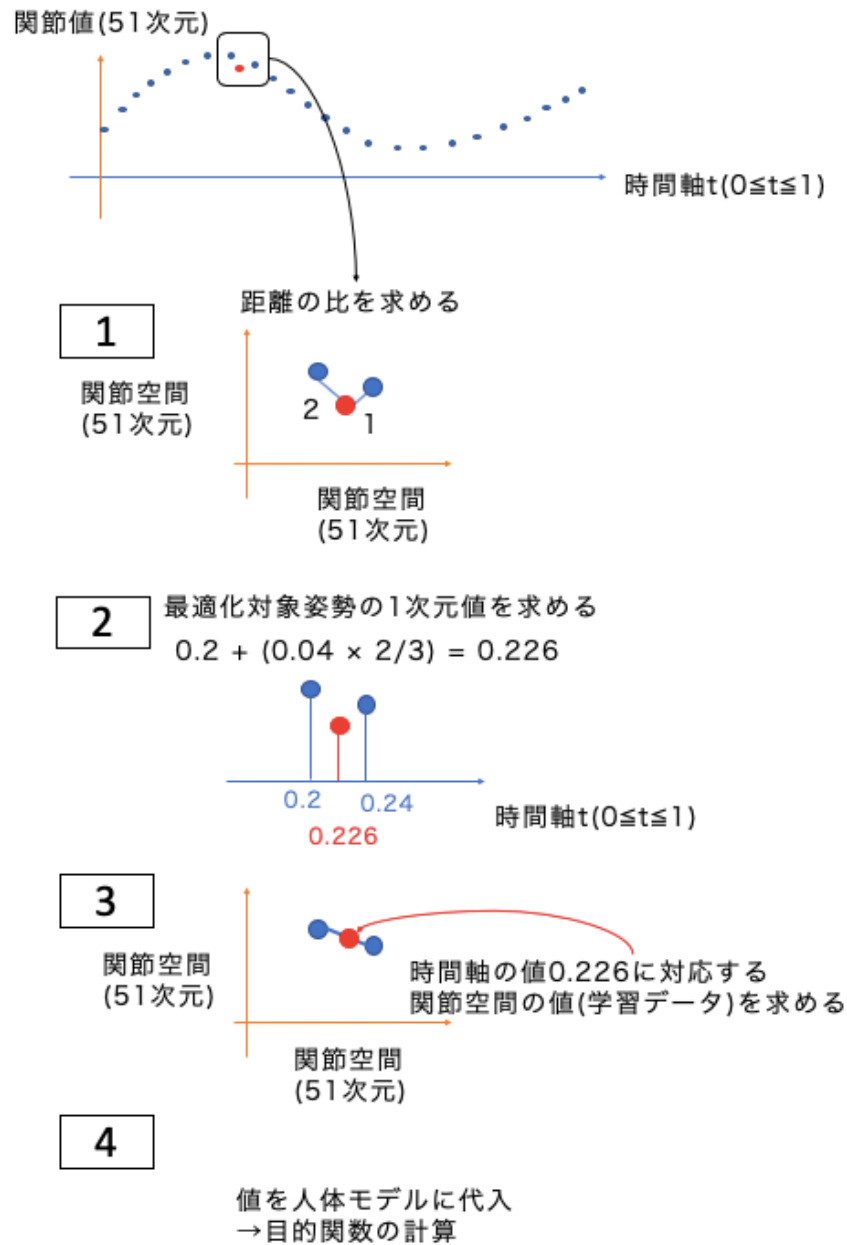


図 4.1 時間軸最適化の流れ

第 5 章 実験

提案手法の有効性を確認するため、実験を行った。人物位置推定 (初期位置推定) の実験では、実画像から、走行動作開始位置を原点とするワールド座標系における人物位置を求めることが可能か検証した。また、人物位置姿勢推定では推定する関節値の精度検証を目的として実験を行った。

5.1 人物位置推定 (初期位置推定)

5.1.1 実験環境

人物位置推定では、実画像を対象とする。図 5.1 に本実験で対象とした画像を示す。図 5.1 ではサッカーボールをワールド座標系の指定位置に事前に設置することで、3次元シーンにおける既知点を作成した。サッカーボールは走行方向に対して平行方向に 5m 間隔、垂直方向に 2m 間隔で計 12 個設置した。図 5.2 に実験環境を示す。そして、3次元既知点を利用したカメラの位置姿勢推定には既存ライブラリの OpenMVG[22] を使用する。また、画像人物領域の特徴点は、セグメンテーションを行うことで人物領域を抜き出すことで指定する。本実験では、セグメンテーションで抜き出した人物領域において、縦軸の座標値が最も大きい画素を頭の位置とすることでワールド座標系における頭の位置を求める。セグメンテーションでは `remove.bg`[6] という web サービスを利用する。

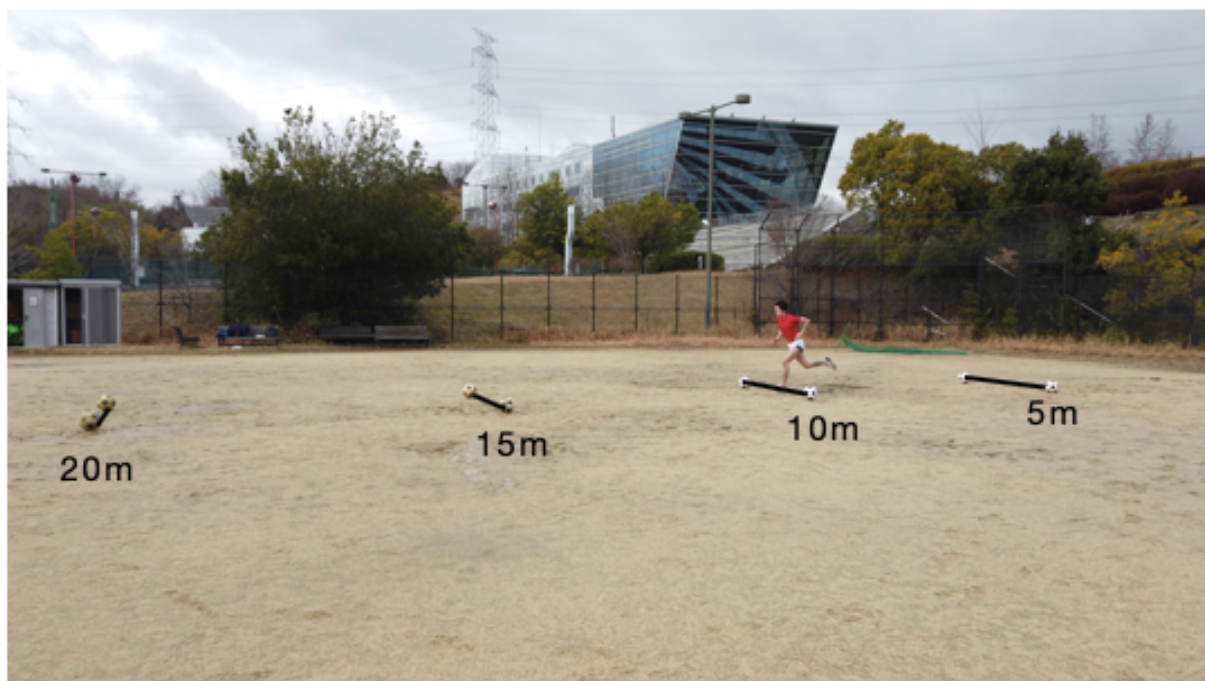


図 5.1 人物位置推定対象の画像

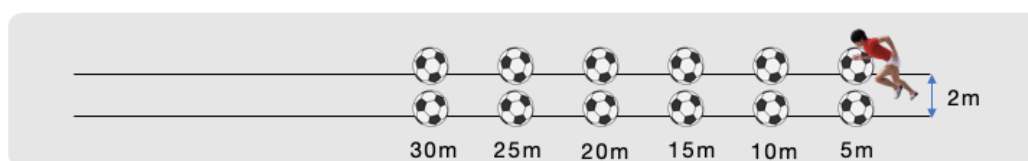


図 5.2 人物位置推定の実験環境

5.1.2 実験結果

セグメンテーションをして人物領域を抜き出した結果を図 5.4 に示す。セグメンテーションは、画像全体 (図 5.1) に対して実施するのではなく、人物が写っている部分を切り出して (図 5.3) 実施した。そして、セグメンテーションで求めた頭位置の画像上の座標値 (2460, 925) を使用し、ワールド座標系における頭位置の座標値 (x, y, z) を求めた結果、出力値は $(9.95, 0.00, 1.61)$ となった。ここで、 x は走行方向の軸であり、進行方向に正である。また、 z は人物の高さ (身長) 方向の軸であり、上方向が正である。そして、出力値の単位はメートルである。対象画像 (図 5.1) を観察すると、10m 地点に設置したサッカーボール付近に人物が位置しており、 x 方向の推定結果はおおよそ正しいことが分かる。また、走行を行なっている人物の身長は 1.69m であり、画像上の姿勢は少し前傾しているため、 z 方向の推定結果も同様におおよそ正しいことが分かる。



図 5.3 切り出した画像



図 5.4 セグメンテーション結果

5.2 人物位置姿勢推定

5.2.1 実験環境

実画像を対象とすると人物姿勢の真値を取得することができないため、今回は実画像の役割 (目標姿勢) を CG 画像とするシミュレーション実験を行う (図 5.5)。シミュレーション実験を行うことで、目標姿勢の関節角度と、最適化によって目標姿勢に近づける処理を行う姿勢 (以下最適化対象姿勢) の関節角度の比較が可能になり、精度検証を行うことができる。ここで、目標姿勢と最適化対象姿勢の位置とワールド座標系に対するオブジェクト座標系の姿勢は等しいとする。CG 画像をレンダリングする際には、blender[1] という 3DCG ソフトウェアを使用する。また、学習データは、鹿屋体育大学の陸上競技部の選手の走行データを使用する。鹿屋体育大学から提供されたデータは trc 形式のファイルで、人体 3 次元モデルに直接代入できないため、3DCG ソフトウェアの Motion Builder[3] を用い、trc ファイルを fbx ファイルに変換し、fbx ファイルで保存された人体 3D モデルの関節角度情報を走行データとする。今回使用した走行データは 1 周期 33 フレームであり、単一フレーム処理では、前フレームにおける姿勢の真値を初期値として最適化を行うことで次フレームの目標姿勢にどれだけ近づけることができるかを検証した。動画画像処理では、事前取得した走行データを参照し、観測対象の姿勢に最も似ている走行データの値を初期値として最適化を行うことで、1 周期通して処理を行った。ここで、観測対象の姿勢と事前取得した姿勢は同一人物のものである。CG 画像のサイズは、実際の実画像における人物領域のサイズを想定し、横 655 画素、縦 472 画素とした。図 5.6 に横 3840 画素、横 2160 画素で実際に撮影した画像から人物領域を切り出した画像を示す。サイズは横 655 画素、縦 472 画素である。

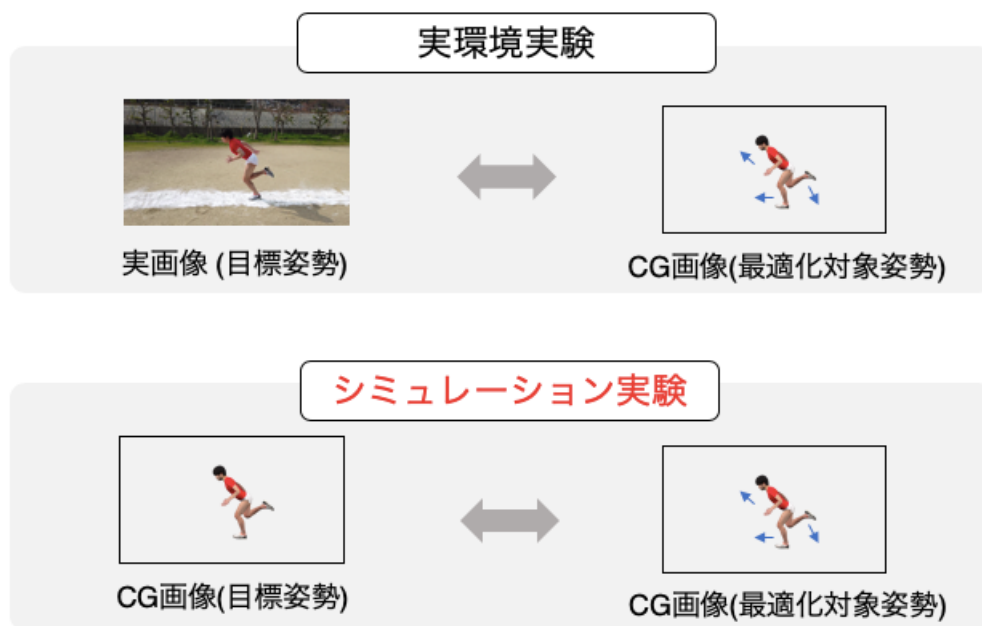


図 5.5 人物位置姿勢推定の実験環境

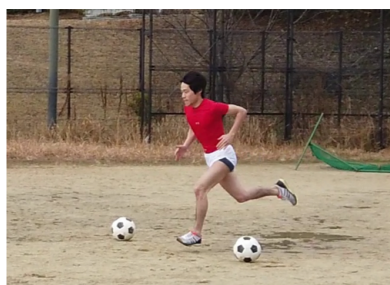


図 5.6 横 655 画素・縦 472 画素の画像

5.2.2 実験結果

前フレームにおける姿勢の真値を初期値として最適化を行う単一フレーム処理の実験結果と、1 周期通して最適化処理を行う動画像処理の実験結果を示す。今回の実験では、最適化に使用する固有値空間の主成分軸の数を 15 とした。

単一フレーム処理

単一フレーム処理では学習データとの差分推定は行なっておらず、前フレーム姿勢の真値を初期値として、次元削減を行なった固有値空間で最適化を行なった。図 5.7 は、対象フレームにおける、最適化前の姿勢 (左グレー姿勢) と最適化後の姿勢 (右グレー姿勢) を目標画像 (カラー画像) に重ね合わせた結果である。最適化後の姿勢と目標姿勢は重なり部分が大きく、画像上における姿勢の違いはほとんど見られない。図 5.8 は最適化前の姿勢 (左グレー姿勢) と最適化後の姿勢 (右グレー姿勢) を目標画像 (カラー画像) に重ね合わせた結果を正面から見た図である。図 5.7 では最適化後の姿勢において大きな違いは見られなかったが、正面から確認すると LeftArm 等の関節に少し違いが見られる。最適化前の姿勢における目標姿勢との関節値の誤差と、最適化後の姿勢における目標姿勢との関節値の誤差を表 5.1 に示す。表 5.1 では、それぞれの関節の回転誤差と類似度を記述した。回転誤差は回転ベクトル表現を使用して記述する。ここで、回転ベクトル表現とは、回転をある回転軸とその軸周りの回転角によって表す表現方法のことであり、表 5.1 に記述した回転誤差の値は基準回転軸周りの回転角である。表 5.1 の一番下の行に示す類似度は、最適化前は 0.3168 であるのに対して、最適化後は 0.9279 となり、最適化によって類似度が向上したことが分かる。それぞれの関節ごとに回転誤差の値を見ると、最適化前の姿勢において大きく誤差が見られていた LeftUpLeg、RightUpLeg の回転誤差は減少している。LeftUpLeg と RightUpLeg の値は表 5.1 に青字で記述した。しかし、表 5.1 の赤字で記述した LeftArm において、最適化後の姿勢においても 15.8 度になっており、誤差が大きい。最適化を行う際のカメラ姿勢で生成した画像上における姿勢の違い (図 5.7) がほとんど見られなくても、カメラ視線奥行き方向の不定性により、誤差が発生している (図 5.8) と考えられる。

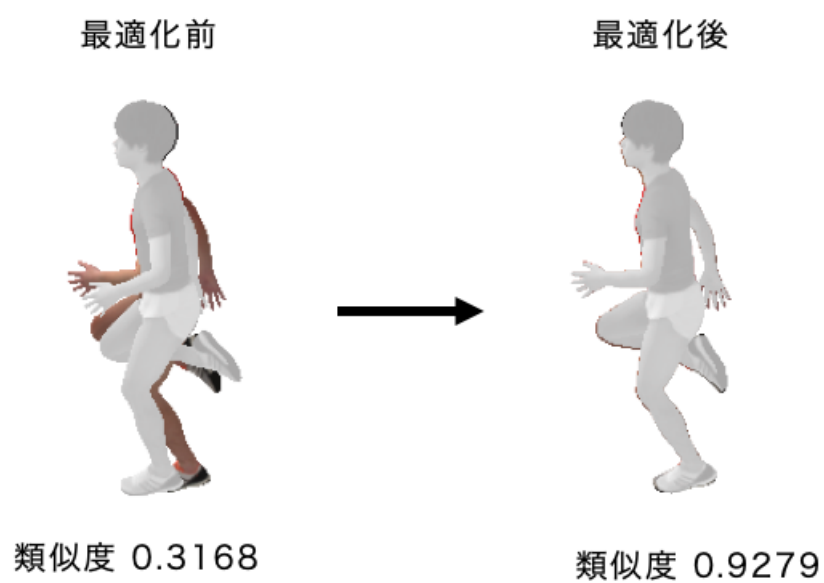


図 5.7 単一フレーム処理の結果

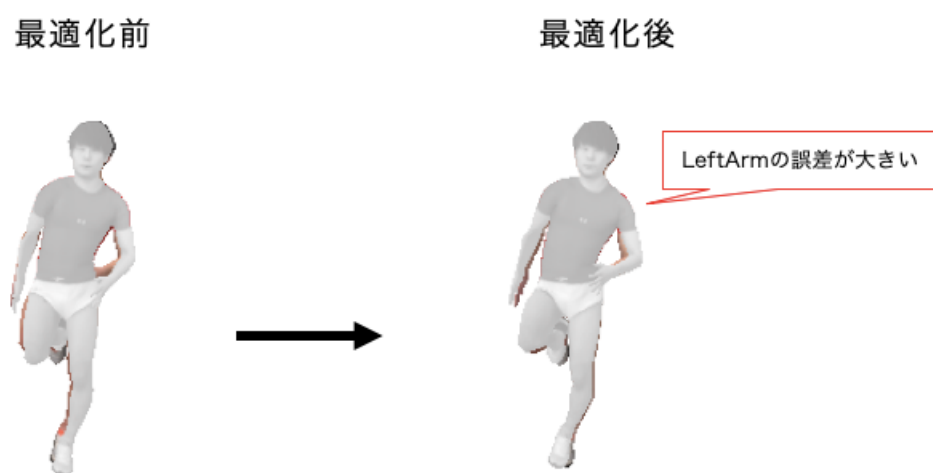


図 5.8 単一フレーム処理の結果 (正面から見た姿勢)

関節	最適化前の誤差	最適化後の誤差
Spine	4.7	2.7
Spine1	5.1	2.3
Spine2	0.0	0.0
Neck	2.5	1.4
Head	2.4	3.0
LeftShoulder	2.3	4.4
LeftArm	14.1	15.8
LeftForeArm	5.7	7.3
RightShoulder	1.9	1.3
RightArm	7.5	7.5
RightForeArm	3.2	5.0
LeftUpLeg	7.4	2.2
LeftLeg	7.0	5.1
LeftFoot	9.4	9.1
RightUpLeg	16.6	5.1
RightLeg	3.1	2.0
RightFoot	2.3	1.0
類似度	0.3168	0.9279

表 5.1 単一フレーム処理の結果

動画像処理

動画像処理では、事前取得した走行データを参照し、観測対象の姿勢に最も似ている走行データの値を初期値として最適化を行う。事前取得した走行データを参照する理由として、動画像処理では単一フレーム処理のように前フレームの真値を初期値に設定できないため、フレームが蓄積するにつれて初期値と最適解の距離が大きくなってしまい、最適化の結果がうまく収束しないためである。トップアスリートにおいてスプリント毎における走行姿勢の変化は極めて微小な変動でしかないということを前提とすると、事前取得した走行データを参照することで初期値と最適解の距離が大きく離れてしまうことを防ぐことができると考えられる。動画像処理の結果として、類似度の推移を図 5.9 に示し、5 フレーム間隔で収束した姿勢 (グレー姿勢) を目標画像 (カラー画像) に重ね合わせた結果を図 5.10 に示す。図 5.9 から、それぞれのフレームにおいて、最適化の目的関数である類似度はおよそ 0.6 から 0.85 の間で収束していることがわかる。最も高い値は 24 フレーム時の類似度の値で 0.8345 である。図 5.11 に、24 フレーム時における学習データ参照結果の姿勢 (左グレー姿勢) と、固有値空間最適化後の姿勢 (右グレー姿勢) を目標画像 (カラー画像) に重ね合わせた結果を示し、正面から見た結果を図 5.12 に示す。また、最適化前の姿勢 (学習データ参照結果の姿勢) における目標姿勢との関節値の誤差と、最適化後の姿勢における目標姿勢との関節値の誤差を表 5.2 に示す。赤字で記述した関節は、回転ベクトル表現での回転誤差を表す回転角が 2 桁を超えており、単一フレーム処理と比較して大きな誤差が発生した。事前取得した走行データと観測対象の姿勢の類似性が低いため、初期値と最適解の距離が離れてしまい、単一フレーム処理と比較して精度が低下したと考えられる。また、RightArm、RightForeArm の誤差が大きい理由として、胴体部位に RightArm、RightForeArm が隠れてしまうオクルージョン問題によって、RightArm、RightForeArm の情報が画像上に反映されず、最適解に収束されなくなってしまうと考えられる。

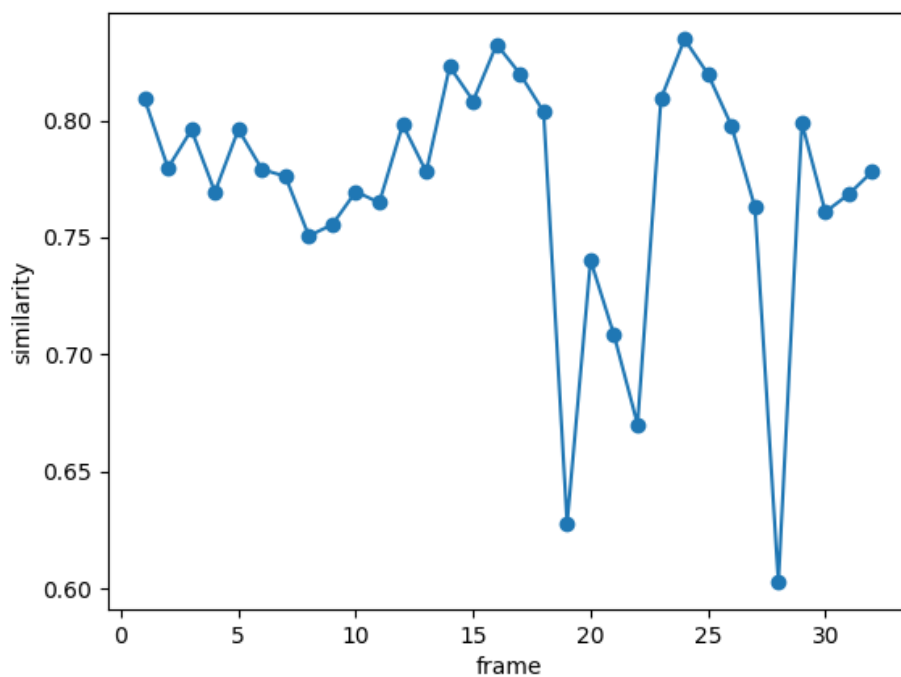


図 5.9 類似度の推移

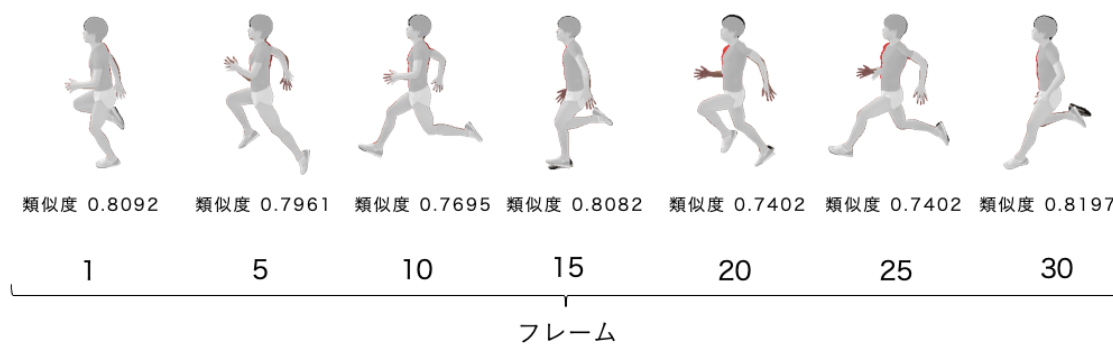
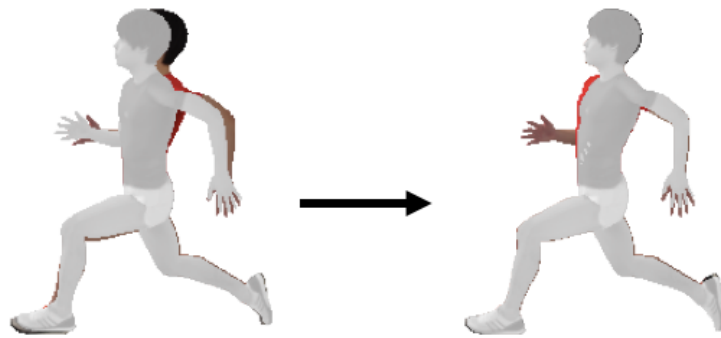


図 5.10 動画像処理の結果

最適化前(学習データ参照結果)

固有値空間最適化後



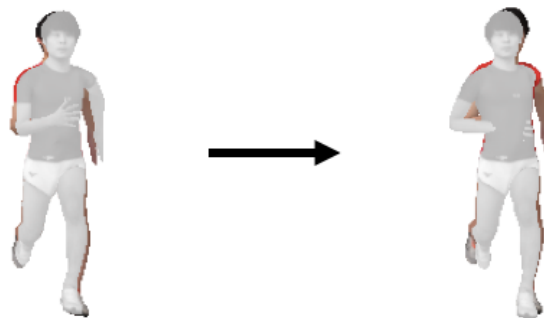
類似度 0.4471

類似度 0.8345

図 5.11 動画像処理における 24 フレーム時 (最も高い類似度で収束) の結果

最適化前(学習データ参照結果)

固有値空間最適化後



類似度 0.4471

類似度 0.8345

図 5.12 動画像処理における 24 フレーム時 (最も高い類似度で収束) の結果 (正面から見た姿勢)

関節	最適化前の誤差	最適化後の誤差
Spine	5.2	6.2
Spine1	9.9	6.7
Spine2	0.0	0.0
Neck	1.7	1.7
Head	10.7	3.7
LeftShoulder	4.1	11.0
LeftArm	2.2	22.9
LeftForeArm	15.6	5.8
RightShoulder	2.2	1.6
RightArm	17.0	38.5
RightForeArm	3.7	18.2
LeftUpLeg	3.6	3.3
LeftLeg	8.3	6.4
LeftFoot	2.4	4.6
RightUpLeg	6.3	3.9
RightLeg	3.8	6.9
RightFoot	5.1	6.1
類似度	0.4471	0.8345

表 5.2 動画像処理における 24 フレーム時 (最も高い類似度で収束) の結果

5.3 考察

本章では、人物位置推定と人物位置姿勢推定の実験を行なった。人物位置推定では、本提案手法において望ましい出力を確認できた。人物位置姿勢推定では、シミュレーション環境において、本提案手法による人物姿勢推定の精度を確認した。その結果、単一フレーム処理であれば、脚関節の関節誤差を大幅に減少させることが確認できた。単一フレーム処理に関して、結果の画像を見ると2次元上の姿勢の誤差はほとんど見られない。ただし、3次元姿勢における精度に関してカメラ視線奥行き方向に不定性やオクルージョン問題が発生し、正しい値に収束しない関節が存在することが分かった。また、学習データからの差分推定による動画像処理においては、単一フレーム処理と比較して精度が低下した。

奥行き方向の不定性に関する原因として、画像上では姿勢が等しくなっても、カメラ視線方向の奥行きに誤差が発生してしまい、3次元姿勢では誤差が生じることが考えられる。また、人体に隠れて画像に写らない RightArm のような関節に関して、オクルージョン問題が発生してしまった。提案手法では主成分分析による次元削減により、値の変位範囲を制限することで奥行き問題とオクルージョン問題の解決に取り組んだが、有効ではなかった。奥行き問題とオクルージョン問題を解決する手法として、探索する最適化空間に制約を付けて最適化を行う方法や、奥行き問題やオクルージョン問題が発生する関節を判定しその関節値に学習データをそのまま参照する手法が考えられる。しかし、奥行き問題やオクルージョン問題が発生する関節を判定する処理の場合、判定処理をどのように設計するのかが問題になるため、検討が必要である。また、動画像処理による精度低下問題の原因として、参照する学習データと、目標データ(観測対象のデータ)の類似性が低いことが考えられる。この問題を解決するためには、どの程度の類似性を持つ学習データが必要なのかを検討し、必要な学習データ(1周期)を、学習データ群(複数周期)から選択する処理を行う必要がある。必要な類似性が検討でき、必要な類似性を持つ学習データを参照することで、動画像処理における誤差低下問題は解決できると考えられる。

第 6 章 おわりに

6.1 まとめ

本論文では、陸上 100m 走競技におけるアスリートの 3 次元位置姿勢を推定するため、個人最適化された人体 3 次元モデルと RGB カメラ 1 台を用いた手法を提案した。提案手法は、人物位置推定 (初期位置推定) と人物位置姿勢推定の大きく 2 つの処理群に分けることができる。人物位置推定では、3 次元既知点を利用してカメラの位置姿勢を求め、その情報と人物領域の特徴点を利用して位置を推定する。人物位置姿勢推定では、セグメンテーションにより抽出された実画像の人物領域と CG 画像の人体 3 次元モデルの領域で類似度計算を行い、類似度を目的関数、人体 3 次元モデルの位置姿勢パラメータを決定変数として非線形最適化を行うことで 3 次元位置姿勢を推定する。

実験では、人物位置推定と人物位置姿勢推定を行った。人物位置推定では対象の実画像からワールド座標系における人物位置を求めることができるかを検証し、望ましい結果を確認できた。人物位置姿勢推定では、実画像を対象とせず、CG 画像同士で行うことで精度検証を行なった。具体的には、最適化対象姿勢が、本提案手法によって目標姿勢にどの程度近づけることができるかを確認した。その結果、単一フレーム処理であれば、脚関節の関節誤差を大幅に減少させることが確認できた。しかし、関節角度の値に不定性が発生してしまうことや、動画像処理においては精度が低下してしまうことなどの問題も発生した。

6.2 今後の課題と展望

本手法に関しての発展は、まず 3 次元位置姿勢精度の向上である。本手法では、シミュレーション環境かつ単一フレーム処理において、脚関節の関節誤差を大幅に減少させることができることを確認した。しかし、関節角度の値に不定性が発生する問題や、動画像処理における精度低下の問題を解決することができなかったため、これらの問題を解決することが必要である。不定性問題を解決する手法としては、最適化空間に制約を付ける方法や、不定性問題が発生する関節を判定し、その関節

値に対して学習データをそのまま参照する手法が考えられる。また、本手法では主成分分析を行う際に 1 周期すべてのデータを用いているが、主成分分析に用いるデータを 1 周期から特定姿勢部分のデータに限定することで、関節変位の範囲を制限でき、不定性問題が解決できる可能性がある。動画像処理における精度低下問題を解決する手法としては、学習データの類似性を上げるなどの方法が考えられる。

また、今回の人物位置姿勢推定の実験では人物位置とワールド座標系に対するオブジェクト座標系の姿勢を最適化対象に含めず、姿勢パラメータである関節の角度のみを最適化対象にして実験を行なった。今後は人物位置とワールド座標系に対するオブジェクト座標系の姿勢を最適化対象のパラメータに含めて実験を行う必要がある。

次に、実画像への本手法の適用が考えられる。現在は人物位置姿勢推定においてシミュレーション環境での実験のみであるが、実画像を対象にすることで、照明やセグメンテーションの精度による類似度への影響などの問題が発生すると考えられる。照明の条件やセグメンテーションの精度によって結果がどのように変わるのかを確認し、条件にロバストな類似度になるように類似度設計をしなければならない。

さらに、最適化時間の短縮も必要である。本手法において、1 周期 33 フレームを動画像処理する際、時間が約 3 時間程度かかった。陸上短距離走における本手法の実用化を考えると、データ取得からフィードバックまでの時間が短ければ短いほど有効であるため、アルゴリズムを再設計し、最適化時間の短縮に取り組む必要がある。

謝辞

本研究は、奈良先端科学技術大学院大学 先端科学技術研究科 インタラクティブメディア設計学研究室 加藤博一のご指導の下で行いました。多岐にわたるご指導やご助言を賜りました加藤博一教授に深く感謝いたします。

学内発表の場におきまして、適切にご指摘を頂きました本学先端科学研究科 ロボティクス研究室小笠原司教授に深く感謝いたします。

研究方針や論文執筆にあたり適切な御教示や御助言を頂きましたインタラクティブメディア設計学研究室の神原誠之准教授に深く感謝いたします。

日頃の研究生活において様々なご意見やご指導を頂きましたインタラクティブメディア設計学研究室の藤本雄一郎助教に深く感謝いたします。

研究方針において様々なご意見を頂きましたインタラクティブメディア設計学研究室の武富貴史客員准教授に深く感謝いたします。

研究活動において多岐にわたるご指導を頂きましたオタゴ大学の Alexander Plopski 博士研究員に深く感謝いたします。

本研究を進めるにあたり、データ提供にご協力頂きました鹿屋体育大学の永原隆助教に深く感謝いたします。

研究生生活を送る上で様々なサポートをして頂きました上野真紀子女史ならびに小川暁子女史に深く感謝いたします。

研究に対してのご助言や様々な形での支援を頂きました先輩、同輩ならびに後輩たちに深く感謝いたします。

最後に、長い学生生活を支えていただいた両親、家族並びに友人たちに心から感謝いたします。

参考文献

- [1] blender.org - home of the blender project - free and open 3d creation software. <https://www.blender.org/>. Accessed: 2019-12-18.
- [2] Dji osmo pocket – スタビライザー搭載ハンドヘルドカメラ – dji. <https://www.dji.com/jp/osmo-pocket>. Accessed: 2019-12-23.
- [3] Motionbuilder | リアルタイム 3d キャラクター アニメーション ソフトウェア | area japan | autodesk. <https://area.autodesk.jp/product/motionbuilder/>. Accessed: 2020-1-3.
- [4] Opencv: Real time pose estimation of a textured object. https://docs.opencv.org/master/dc/d2c/tutorial_real_time_pose.html. Accessed: 2019-01-25.
- [5] Optojumpnext - オプトジャンプネクスト | S & C スポーツ科学計測テクノロジー スポーツパフォーマンス分析. <http://sandcplanning.com/solution/category/detail/?cd=7>. Accessed: 2019-12-18.
- [6] Remove background from image – remove.bg. <https://www.remove.bg/>. Accessed: 2019-12-18.
- [7] スーパースキャンスタジオ | super scan studio | super scan studio は、コミュニケーションの新しい「場所」であり、可能性を探る拠点です。 <http://www.superscanstudio.com/>. Accessed: 2019-12-18.
- [8] モーションキャプチャ | sports performance research center. <https://splab.nifs-k.ac.jp/ja/kinou/motion-capture/>. Accessed: 2019-12-18.
- [9] Mathieu Salzmann Vincent Lepetit Bugra Tekin, Isinsu Katircioglu and Pascal Fua. Structured Prediction of 3D Human Pose with Deep Neural networks. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*. BMVA Press, 2016.
- [10] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2D pose estimation using part affinity fields. In *The IEEE Confer-*

- ence on Computer Vision and Pattern Recognition (CVPR), 2017.
- [11] Cappadona and Julianne. Kinematic and spatiotemporal analysis between sprint drills and maximal sprinting. In *PTHMS Undergraduate Publications. 1*, 2013.
 - [12] Ching-Hang Chen and Deva Ramanan. 3D human pose estimation = 2D pose estimation + matching. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
 - [13] Riza Alp Guler and Iasonas Kokkinos. Holopose: Holistic 3D human reconstruction in-the-wild. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
 - [14] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017.
 - [15] Joseph Hunter, Robert Marshall, and Peter McNair. Interaction of step length and step rate during sprint running. *Medicine and science in sports and exercise*, 2004.
 - [16] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.
 - [17] Diederik Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, 2014.
 - [18] Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little. A simple yet effective baseline for 3D human pose estimation. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017.
 - [19] Dushyant Mehta, Srinath Sridhar, Oleksandr Sotnychenko, Helge Rhodin, Mohammad Shafiei, Hans-Peter Seidel, Weipeng Xu, Dan Casas, and Christian Theobalt. Vnect: Real-time 3D human pose estimation with a single RGB camera. *ACM Transactions on Graphics*, 36(4), 2017.
 - [20] Andrea Monte, Valentina Muollo, Francesca Nardello, and Paola Zamparo. Sprint running: how changes in step frequency affect running mechanics

- and leg spring behaviour at maximal speed. *Journal of sports sciences*, 2016.
- [21] Pierre Moulon, Pascal Monasse, and Renaud Marlet. Adaptive structure from motion with a contrario model estimation. In *The Asian Computer Vision Conference (ACCV 2012)*. Springer Berlin Heidelberg, 2012.
 - [22] Pierre Moulon, Pascal Monasse, Romuald Perrot, and Renaud Marlet. Openmvg: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*. Springer, 2016.
 - [23] T. Ohashi, Y. Ikegami, K. Yamamoto, W. Takano, and Y. Nakamura. Video motion capture from the part confidence maps of multi-camera images by spatiotemporal filtering using the human skeletal model. 2018.
 - [24] Georgios Pavlakos, Xiaowei Zhou, Konstantinos G Derpanis, and Kostas Daniilidis. Coarse-to-fine volumetric prediction for single-image 3D human pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
 - [25] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 1999.
 - [26] Giuseppe Rabita, Sylvain Dorel, Jean Slawinski, Eduardo Sáez de Villarreal, Antoine Couturier, Pierre Samozino, and Jean-Benoît Morin. Sprint mechanics in world-class athletes: A new insight into the limits of human locomotion. *Scandinavian Journal of Medicine & Science in Sports*, 2015.
 - [27] Sebastian Ruder. An overview of gradient descent optimization algorithms. <http://arxiv.org/abs/1609.04747>. Accessed: 2019-01-27.
 - [28] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
 - [29] Yoshiteru Yamamoto, Keisuke Doman, Yoshiya Hotta, and Yoshito Mekada. Running video matching algorithm focusing on the periodicity and linearity of running motion. In *Congress on Sport Sciences Research and Technology Support 2018*, 2018.

- [30] Xingyi Zhou, Qixing Huang, Xiao Sun, Xiangyang Xue, and Yichen Wei. Towards 3d human pose estimation in the wild: A weakly-supervised approach. In *The IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [31] 金谷 健一. これなら分かる最適化数学 : 基礎原理から計算手法まで. 共立出版, 2005.