演技力向上を目的とした 動画データにおける骨格解析方法の提案

釜谷 尚宏[1,a)] 松井 加奈絵[1,b)]

概要：本研究では，人の演技を行っている様子を収録した動画データに対し骨格解析を行うことで，演技に対するフィードバックを演者に行うための骨格推定手法を提案する．ドラマや映画といった作品制作の過程において，演者は繰り返し同じセリフ，動きを用いた演技を求められる場合がある．しかしながら，この繰り返しにより演者の中で慣れが生まれてしまい，演技上，人物にとって初めての場面にも関わらず，その先の流れを汲み取った動画が生まれるといった，不自然さが生まれてしまう恐れがある．そのため，演者は繰り返し同じ芝居を新鮮なリアクションを伴って行う訓練が必要となるが，現状繰り返し演技が成立しているか判定するのは講師等の人間であるため，学習コストがかかってしまう．本研究では，繰り返し演技を行っているデータの動きに着目した骨格推定を用いた解析手法を提案し，将来的に自動判定につながる環境構築を目指す．本提案の有用性を確認するために，実際に演技データを収集し，そのデータを提案の手法を用いて解析した．その結果，動画データの収集方法に改善の必要があることがわかったものの，ある一定の動きの一致度を算出することができた．

A Proposal of an Evaluation Method on a Performance Improvement using Skeletal Analysis

Naohiro KAMATANI[1,a)] Kanae MATSUI[1,b)]

1. 概要

動画解析を行う環境が整いつつあることで，スポーツなど身体の動きの分析およびフィードバックが大きな成果をもたらす業界で，動画解析を用いた身体の動きの洗練化が普及しつつある．例えば，スポーツ業界ではスポーツテックのひとつとして，スポーツを行っている様子を撮影し，解析結果を用いるトレーニング方法が用いられつつあり，今後身体の動きの鍛錬を行う現場では，動画解析が大きな役割を果たすと考えられる．

本研究は，身体の動きのうち演技に着目し，演技の動画データを用いて演技力の向上を目指す．演技において，身体の動きを伴なうものは多数あるが，今回は繰り返し同じ演技を求められる状況に着目し，繰り返しの演技の動作の類似度を骨格推定を用いて数値化し，演者へのフィードバックとする．ドラマや映画といった作品制作において，繰り返しの演技は演者は求められる場合が多いが，同じ演技の繰り返しにより演者の中で慣れが生まれてしまい，不自然さが生まれてしまう恐れがある．そのため，演者は繰り返し同様の演技を，初回に行った演技と同様，もしくは類似の感覚で行うための訓練を行う必要があるが，この訓練を第三者視点で評価する方法は，演技講師といった人間による判別しかない．

そこで，本研究では動画データを用いた解析方法を提案し，人を介さずに演者自身が演技における動きの評価をフィードバックとして得ることが可能となる環境を提供することを目的とした．本論文では，まず提案の骨格推定について説明し，実際に実験協力者から得た演技の動画データを提案の手法で解析した結果について述べ，本提案の有用性について論じる．

[1] 東京電機大学
Tokyo Denki University, Adachi, Tokyo 120-8551, Japan
[a)] 17aj042@ms.dendai.ac.jp
[b)] matsui@mail.dendai.ac.jp

2. 関連研究

本章では，提案手法の根幹となる画像データを用いた骨格推定方法の先行研究について述べる．身体の動きそのものを解析するのではなく，動画から対象者の骨格を導き出し，その骨格の動きを数値化し，解析する方法である．以下，動画データを用いた骨格推定に使用される手法について述べる．

2.1 OpenPose における人間の骨格推定

映像から骨格推定を行う際に使用する手法として，OpenPose がある [1]．これは，骨格データと合致する人物映像をニューラルネットワークを用いて学習させ，その学習データを元に姿勢を推定する手法および GitHub 上で公開されているライブラリである．この手法を用いることにより，身体の動きを検知するデバイスである Kinect のように深度データを使わずに映像データのみで骨格情報が得ることができ，さらに複数の人物が同時に動作していても推定が可能となる．

OpenPose は，以下に示すように様々な分野で活用されている．スポーツの分野では自転車ロードレース競技において選手を識別するための研究や [2]，テニスの選手を姿勢変化を可視化し，その姿勢を評価する研究に用いられている [3]．スポーツ以外では，人の歩行データから歩きスマホを行っている人物を認識する方法として利用されている [4]．

2.2 演技データにおける骨格推定

OpenPose は，他の研究では得られる姿勢変化の特徴量を様々な分野に応用しているが，本研究では，その姿勢変化の検出を身振り手振りといった行動のタイミングが重要となる演技に適応する．演技においては，音声解析や表情解析といったアプローチが取られているものの，動画データを用いたフィードバック方法のアプリケーションやソフトウェアの普及には至っていない．しかしながら，スポーツと同様に身体の動きが非常に重要な領域であり，かつ練習成果の判別を演者自身が行うことが難しいことから，本研究では OpenPose を用いた演技データの解析に取り組む．

また，映像作品に求められる演技において，演者は繰り返し同じ演技を求められる時間内で行う必要がある．例えば，ドラマや映画の撮影現場において，1 シーンを撮影する際に演者は同一の演技を何度も繰り返す場合がある．その場合，画角やライティングといった撮影方法の観点から，また映像シーンの観点から，演者は決められた時間に定められた動きを繰り返すが，何度も同一の動きを行うには繰り返しの演技を行っているうちに，動作に慣れが生まれてしまうため，次の動作を制御しながら動くという，通常は起こらない動作を行う．そのため，演者は一般的に最初に決定した演技を基準とし，その基準点から外れないように演技することがヒアリングからわかった．そのため，今回は繰り返し演技における動画データに対し，提案の解析手法を用いることにより，基準となったデータと他の繰り返し演技がどの程度乖離しているのか数値化することに取り組む．

3. 演技動画データの解析手法

本章では，OpenPose を用いた繰り返し演技データの解析手法について述べる．以下，実装方法，骨格推定について詳細を述べる．

3.1 実装方法

演技における動画データを利用した骨格推定を行うために，骨格検出アルゴリズムである tf-pose-estimation における骨格検出を行うための学習済みモデルを用いて，顔面上の目や鼻および関節など，18 種類のポイントを座標としてフレーム毎に抽出し，保存することとした．

繰り返し演技の基準となる 1 回目の動画データのスコアリングから，2 回目，3 回目といった繰り返し演技データが，初回の骨格の動きがどの程度乖離しているのかを判別する．骨格推定で抽出した座標を用いて，以下の処理を実施する．

- フレーム間における座標移動量の計算
- 当該フレームごとに 1 回目とその他を比較した誤差率の計算
- 誤差率の平均化し一致度とするプログラムの実装

3.1.1 骨格推定

まず，骨格推定は上述の tf-pose-estimation を用いて行う．tf-pose-estimation は，機械学習に用いられたソフトウェアライブラリである TensorFlow を用い，OpenPose を実装したものである．これに全身が写った画像を入力すると，18 種類のポイントを座標として変換し，list 形式で出力される．OpenPose で使用されているモデルは，図 1 のように，人間の身体の各所に番号がラベリングされている．また，図 1 の番号に対応する場所の名称は，表 1 の通りである．

映像を 1 フレームずつ tf-pose-estimation に入力することで，そのフレームにおける姿勢情報を取得し，動画ごとに表 2 のような CSV 形式に整形するプログラムを通して出力する．

3.1.2 スコアの算出

次に，フレームごとの時間差分座標を用いてスコアを関節ごとに算出する．まず，算出に必要な座標が推定できているか確認する．n 回目の動画の f フレームを比較する際に必要な座標は，n 回目の動画の f フレーム，$f+1$ フレーム，1 回目の動画の f フレーム，$f+1$ フレームの 4 つであ

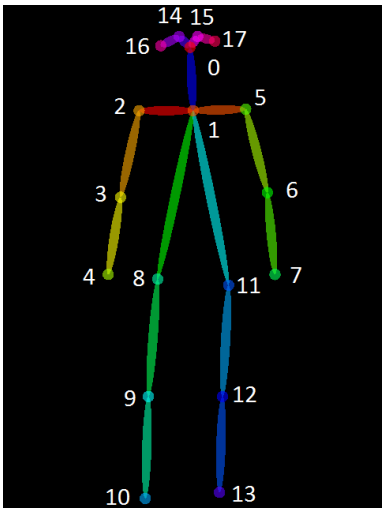

図 1　OpenPose によるモデルの関節番号

表 1　番号対応表

番号	場所
0	Nose
1	Neck
2	RShoulder
3	RElbow
4	RWrist
5	LShoulder
6	LElbow
7	LWrist
8	RHip
9	RKnee
10	RAnkle
11	LHip
12	LKnee
13	LAnkle
14	REye
15	LEye
16	REar
17	LEar

る．今回，同一関節において 4 つのフレームのうちいずれかが推定されていない場合は，計算から省くこととした．

次に，比較する n 回目の映像の f フレームの座標 (xn_f, yn_f)，$f+1$ フレーム目の座標 (xn_{f+1}, yn_{f+1})，1 回目の映像の f フレーム目の座標 $(x1_f, y1_f)$，$f+1$ フレーム目の座標 $(x1_{f+1}, y1_{f+1})$ を CSV データより取得する．ここで，それぞれの移動量 $Dn_x, Dn_y, D1_x, D1_y$ を求める．

$$Dn_x = xn_{f+1} - xn_f \tag{1}$$

$$Dn_y = yn_{f+1} - yn_f \tag{2}$$

$$D1_x = x1_{f+1} - x1_f \tag{3}$$

$$D1_y = y1_{f+1} - y1_f \tag{4}$$

これらの移動量を以下の式に代入し，n 回目と 1 回目の f フレームにおける誤差率 Ex_f, Ey_f を算出する．

表 2　出力した CSV の一部

human	frame	point	x	y
1	0	0	650	160
1	0	1	648	250
1	0	2	568	252
1	0	3	540	362
1	0	4	558	422
1	0	5	734	252
1	0	6	756	362
1	0	7	734	424
1	0	8	612	454
1	0	11	690	462
1	0	12	742	510
1	0	13	722	692
1	0	14	634	144
1	0	15	666	146
1	0	16	610	158
1	0	17	688	162
1	1	0	650	160
1	1	1	648	250
1	1	2	566	252
1	1	3	540	364
1	1	4	558	422

$$Ex_f = \frac{(Dn_x - D1_x)}{D1_x} \tag{5}$$

$$Ey_f = \frac{(Dn_y - D1_y)}{D1_y} \tag{6}$$

ここで，$D1_x, D1_y$ がそれぞれ 0 の場合はこの式では計算できない．そこで，$D1_x, D1_y$ が 0 の場合において，$D1_x = Dn_x$ または $D1_y = Dn_y$ の場合は n 回目も 1 回目も不動であるため，Ex_f と Ey_f に直接 0 を代入することとした．それ以外の $D1_x, D1_y$ が 0 の場合は，1 回目では不動であるが n 回目では動作していることになるため，Ex_f または Ey_f に直接 1 を代入した．

これら x 座標と y 座標の誤差率 Ex_f，Ey_f を平均して f フレーム目の誤差率 E_f とする．

$$E_f = \frac{Ex_f + Ey_f}{2} \tag{7}$$

この値を算出可能な全てのフレームにおいて算出し，誤差率の平均を求める．

算出可能なフレームの総数を f' とした場合，n 回目と 1 回目における誤差率 $\bar{E}$ は以下のようになる．

$$\bar{E} = \frac{\sum_{k=0}^{f'} E_k}{f'} \tag{8}$$

この誤差率の絶対値を用いて，一致度として $Score$ を算出する．

$$Score = (1 - |\bar{E}|) \times 100 \tag{9}$$

次に，この $Score$ を 18 種類のポイントごとに出力し，最後にそれらの数値を平均したものも出力する．

表 3　使用機材

番号	名称	用途	台数	機材名
1	カメラ	表情の収録	1	FDR-AX700
2	カメラ	概観映像の収録	3	GoPro HERO 6 BLACK
3-1	三脚	カメラ補助器具	1	VCT-VPR1
3-2	三脚	カメラ補助器具	3	Pole Pod EX
4-1	モニタ	映像フィードバック	1	EX-LD2071TB
4-2	モニタ	映像フィードバック	1	On-Lap 1503I

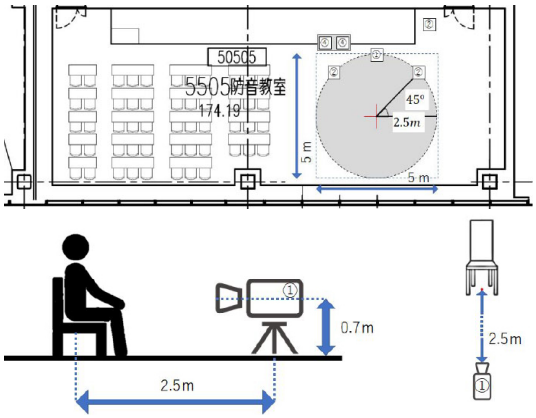

図 2　カメラの配置図

表 4　実験協力者の属性

ID	性別	年代	属性	撮影回数
A	男性	30 代	俳優（演技経験あり）	3
B	男性	20 代	学生（演技経験なし）	3
C	男性	20 代	学生（演技経験なし）	2

以上が，本研究における演技における動画データの解析手法である．

4.　演技における動画データの収集

本解析手法の評価を行うために 1 名の俳優の方の演技の動画データ，また演技の経験がない 2 名のデータを収録した．以下収録方法，および環境について述べる．

4.1　実験環境

動画データの収集のための実験は，東京電機大学東京千住キャンパス 5 号館 5 階の防音教室で実施した．撮影に使用した機材を表 3 に示す．また，カメラの配置図を図 2 に示す．カメラは計 4 台を使用し，うち番号 1 のカメラは前方 2.5m の位置に設置し，正面の表情含む全身の映像を収録した．また，番号 2 の 3 台の GoPro のうち，2 台を斜め 45 度前方の半径 2.5 m の位置に配置し全身の映像，残りの 1 台を用いて概観の撮影を行った．今回解析に用いたデータは，番号 1 の正面に設置したカメラの映像である．

4.2　実験協力者

本実験を行うにあたり，3 名の実験協力者の演技を撮影した．3 名の属性は，表 4 の通りである．協力者 A と B は 3 回，協力者 C は 2 回の撮影を行った．

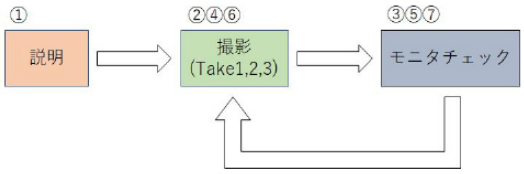

図 3　撮影手順

図 4　撮影した動画

4.3　演技の収録方法

上述の通り，ドラマや映画の収録現場では，ひとつのシーンに対し，NG や別角度での再撮影などの理由により何度も撮影を繰り返す場合がある．そのため演者はひとつのパターンの演技を繰り返し実施するうちに演技が固定化することで，不自然に見える場合がある．本研究では演者に課した題に対して演者が示す初回の反応（以下ファーストリアクション）を撮影しそれ以降撮影およびフィードバックを重ねることにより，演技の上達の過程や，ファーストリアクションとの乖離を観察する．

このようなデータを収集するため，3 名の実験協力者に「哀れみ」の感情を想起する質問をし，1 回目に自然な回答とその様子を録画し，この動画を基礎シナリオとし，2 回目，3 回目はこのシナリオを模倣した演技を実施してもらうものとした．この撮影手順の略図を図 3 に示す．この場合，実験協力者は 2 回目，3 回目の撮影前に 1 回目の動画データを 5〜10 分間閲覧する．

図 4 は撮影したものを編集した動画から抜粋したものであり，図内左から 1 回目，2 回目，3 回目の順に並べたものである．今回はこれら 1 回目のファーストリアクションを基礎データとし，この基礎データの骨格の動きから 2 回目と 3 回目のデータを比較した．

5.　解析結果

実験で得た 3 名の演技における動画データに対し，提案の解析手法を用い解析を行った．以下，詳細を述べる．

実験協力者 A の 1 回目と 2 回目の演技における動きを骨格推定を用いて解析した結果を表 5 に，1 回目と 3 回目を比較した結果を表 6 に示す．

次に実験協力者 B の 1 回目と 2 回目の演技における動きを骨格推定を用いて解析した結果を表 7 に，1 回目と 3 回

表 5　協力者 A の 1 回目と 2 回目の比較

場所	一致度 [%]
Nose	74.416
Neck	72.769
RShoulder	83.611
RElbow	75.730
RWrist	77.314
LShoulder	80.317
LElbow	72.873
LWrist	61.362
RHip	36.503
RKnee	34.101
RAnkle	22.285
LHip	28.603
LKnee	25.713
LAnkle	20.833
REye	80.025
LEye	79.358
REar	80.317
LEar	80.859
Average	60.388

表 6　協力者 A の 1 回目と 3 回目の比較

場所	一致度 [%]
Nose	74.625
Neck	74.479
RShoulder	84.946
RElbow	74.030
RWrist	71.101
LShoulder	74.604
LElbow	72.602
LWrist	54.712
RHip	33.625
RKnee	21.122
RAnkle	12.500
LHip	24.341
LKnee	25.350
LAnkle	60.000
REye	81.234
LEye	79.608
REar	80.234
LEar	81.443
Average	60.031

表 7　協力者 B の 1 回目と 2 回目の比較

場所	一致度 [%]
Nose	76.668
Neck	77.693
RShoulder	74.214
RElbow	64.109
RWrist	82.412
LShoulder	68.351
LElbow	74.023
LWrist	0.000
RHip	68.541
RKnee	84.890
RAnkle	83.133
LHip	55.158
LKnee	75.979
LAnkle	60.365
REye	73.475
LEye	77.502
REar	75.119
LEar	77.169
Average	73.459

表 8　協力者 B の 1 回目と 3 回目の比較

場所	一致度 [%]
Nose	77.772
Neck	79.416
RShoulder	75.859
RElbow	78.067
RWrist	90.088
LShoulder	76.595
LElbow	60.550
LWrist	93.363
RHip	56.968
RKnee	79.244
RAnkle	80.936
LHip	47.596
LKnee	60.182
LAnkle	62.859
REye	76.619
LEye	78.459
REar	75.172
LEar	81.354
Average	73.950

目を比較した結果を表 8 に示す．

最後にに実験協力者 C の 1 回目と 2 回目の演技における動きを骨格推定を用いて解析した結果を表 9 に示す．

2 回目との比較で平均が 60.4%，3 回目との比較で平均が 60.0%という一致度となった．場所別に見ると，概ね 70〜80%の一致度として出ているものもあり，2 回目よりも 3 回目の方が高い場所も見られた．

6. 考察

本章では，結果から得られた数値について考察を述べる．

6.1 解析する動画データに関する考察

解析結果から，3 名とも下半身の動きに対する一致度が他の場所に比べて低いことが分かる．今回は座った状態での撮影であったため，図 5 のように膝回りが正しく推定されずに評価が下がっているように見られる．また，協力者 A の映像は，くるぶしより先が見切れていたため，10(RAnkle) と 13(LAnkle) が正しく推定されず，評価が全体的に下がっている傾向があった．

これらの改善点として，立ち姿勢での撮影にしたり，カメラを引いて全身を撮影したりするなど，撮影時の動画収

表 9　協力者 C の 1 回目と 2 回目の比較

場所	一致度 [%]
Nose	82.485
Neck	78.490
RShoulder	77.217
RElbow	69.732
RWrist	87.357
LShoulder	77.085
LElbow	75.198
LWrist	83.933
RHip	43.118
RKnee	33.419
RAnkle	46.774
LHip	48.932
LKnee	34.854
LAnkle	64.831
REye	82.572
LEye	86.304
REar	81.958
LEar	86.567
Average	68.935

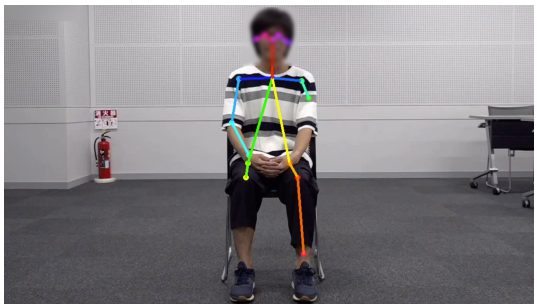

図 5　正しく推定されない例

集状況を改善する必要がある．また，上半身だけの場合は，8 から 13 の場所を異常値として無効にすることで，平均値に影響を出さないようにできると考える．また，OpenPose は 2 次元データでの推定あるが故に，解析には限度がある．正しく識別するには，映像データに加えて Kinect のようなデバイスで取得できる深度データを解析に加えることも有効であると考える．

6.2　解析手法に関する考察

今回の解析手法で行っていないものとして，台詞との同期が挙げられる．セリフを発したタイミングがずれると動作を行うタイミングもずれてしまうものであるが，その後の動作は 1 回目と同じタイミング正確に行っている可能性がある．この点を考慮し，音声の解析によるセリフの同期をすべきと考える．

7.　まとめ

本稿では，同一の演技を繰り返し行う動画データ用いて骨格推定および評価を実施した．繰り返し同様の演技を行うための訓練において，本提案手法はひとつの修練における評価になり得ると考えられる．

今後の課題として，この方法では行っていない台詞との同期や異常値の除去がある．より正確な分析を行う為に，これらを今後の課題とし，更なる精度の向上を目指す．

謝辞

本研究活動は，文部科学省による Society 5.0 実現化研究拠点支援事業によって行われたものである．

参考文献

[1] Cao, Zhe and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser : ”Realtime multi-person 2d pose estimation using part affinity fields”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7291–7299 (2017)

[2] 高木政徳，石川孝明，渡辺裕：”姿勢情報を用いた自転車選手識別手法に関する一検討”，映像情報メディア学会冬期大会 25C-5 (2017)

[3] 黒瀬龍之介，林昌希，石井壮郎，岡村麻人，青木義満：”姿勢推定を用いたテニス映像の姿勢傾向分析”，映像情報メディア学会技術報告 ME2016-130 (2016)

[4] 加藤君丸，渡辺裕：”姿勢推定による肘と肩の角度情報を用いた歩きスマホ認識”，映像情報メディア学会技術報告 13D-4 (2017)

[5] CMU Perceptual Computing Lab(Online)，入手先 ⟨https://
/github.comCMUPerceptualComputing-Lab/openposeblobmasterdocmediakeypoints_pose⟩（参照 2020-05-18）.