

二次元姿勢情報に基づく 速度，加速度，躍度を利用した 動作自動教示システム

原科 堯宗^{1,a)} 高丸 尚教¹

概要：近年，画像処理や機械学習を用いたスポーツやダンスの動作解析，教示するシステムの研究・開発が進んでいる．例えば，モーションキャプチャデータに対して動作解析を行っているものや，深度カメラ・RGB カメラによって撮影された動画データに対して人物姿勢推定を行った後に動作解析を行っているものがある．本研究では，能楽などの無形文化財の保存や技能継承のための舞踏動作の意味論レベルにおける特徴抽出を目的として，(1) 人物を RGB カメラによって撮影された動画データに対して OpenPose を用いて人物二次元姿勢を推定する．(2) 推定された関節座標の時系列データから算出された動作軌跡，速度，加速度，躍度に対して動作解析を行う．(3) これらの解析結果に cos 類似度を使用して動作の人物間の誤差を算出する．この一連の処理系を有する動作の教示システムを提案する．

1. はじめに

近年，画像処理や機械学習を用いたスポーツやダンスの動作を解析し，教示するシステムの研究・開発が進んでいる．このシステムは，モーションキャプチャによって獲得できる人物の関節座標に対して動作解析を行っているもの [1], [2], [3] や，深度カメラ・RGB カメラによって撮影された動画データに対して二次元もしくは三次元の人物姿勢推定によって推定された関節座標に対して動作解析を行っているもの [5], [6] がある．これらの教示システムによって，近年問題視されている能楽などの無形文化遺産や日本舞踊などの芸能の技能継承に役立てられている．しかし，モーションキャプチャによる動作解析では，データ採取のために特別な環境と機材が必要であり，踊り手への負担が大きい．また，RGB カメラによる動画から推定された姿勢情報を使用した研究 [8], [9] では，人物間の関節座標などの姿勢情報のみを比較している．そのため，ある程度似た動作であれば高い類似性を算出してしまう問題がある．そこで，本研究では，リアルタイム自動教示システムの開発を目標として，人物を単一の RGB カメラによって撮影された動画データに対して OpenPose[10] を用いて人物二次元姿勢推定を行う．その後，推定された関節座標の時系列データから算出された各関節毎の速度，加速

度，躍度に対して動作解析を行う．これらの解析結果に対して計算コストの低いとされる cos 類似度を使用して動作の人物間の誤差を算出している．これにより人物間の舞踏動作の類似度を提示する教示システムを提案する．先行研究では，踊りのようなリズムに合わせた動作に対する動作解析を行う際に，音楽の情報を導入した動作解析 [1] が存在する．しかし，本研究では，二次元関節座標から算出可能な値のみを使用することを重要視して動作解析を行う．また，3D シミュレーション上で撮影した人物動作に対して提案手法を実施することで，動作の速度，加速度，躍度を用いた動作解析の有用性を提示する．そして，動作の速度，加速度，躍度に対して cos 類似度を用いた比較を行うことで似た動作の細かな違いを検出することが可能であることを示す．

2. 関連研究

モーションキャプチャにより得られた関節データを用いた動作解析の研究は過去に多く存在する．その中でも中澤は [2]，そうらん節などの踊りの動作をモーションキャプチャによって取得し，モーションセグメントを検出するために音楽情報を導入した動作解析を行っている．これによりモーションセグメントを意味論レベルで正しく検出することを可能にしている．また，この研究では躍度最小モデルを用いて動作軌跡を合成することでヒューマノイドによる舞踏動作の模倣を行っている．白鳥は [1]，音楽のリズム

¹ 中部大学

^{a)} tr21012-2165@sti.chubu.ac.jp

を考慮するためにビートトラッキングを導入した舞踏動作の解析を行っており、モーションキャプチャによって獲得した動作のデータから特に「止め」などのモーションセグメントを検出し、踊りの本質を失わない動作解析を可能としている。鹿内は [3], モーションキャプチャにより日本舞踏の動作を取得し、関節座標間のベクトルの傾きに対して動作解析を行い動作の印象との関係性を評価した。Ran は [4], モーションキャプチャにより取得したダンスモーションに対してヒルベルト__ファン変換を用いて人物の関節の動きを周波数的に解析している。これらのモーションキャプチャを用いた動作データの取得では、専用の機材によるデータ取得を行うため、被験者への時間的、肉体的に負担が大きく大量のデータを取得することが困難であるという問題がある。

モーションキャプチャによる動作データ取得よりも被験者に負担が掛からない手法として、深度・RGB 動画を対象にした動作解析が挙げられる。深度動画を対象とした動作解析の手法として、柴田は [5][6], Kinect から取得した姿勢データに対して時系列間で \cos 類似度を算出している。これにより、モーションセグメントを検出している。そして、検出されたモーションセグメントに対してラベル付を行うことで動作認識を可能にしている。RGB カメラにより撮影された動画を対象にして動作解析を行っている研究も多数存在する。清水は [7], RGB カメラから取得した人物動作動画から速度情報をもとにモーションセグメントを検出した後に躍度最小モデルを用いた動作生成を行うことでシミュレーション上のキャラクタによる人物動作の模倣を実現している。

モーションキャプチャや RGB カメラから取得した動作データに対して人物間の動作の類似性を算出し動作教示を可能とするための研究は多数存在する。柴田は、検出したモーションセグメントと人物姿勢の \cos 類似度を算出することで動作認識を行っている。西澤は [8], 人形のデッサンの能力向上も目的としたデッサン人形とデッサンによる人物画の二次元姿勢情報を \cos 類似度によって比較することでデッサンの結果の良し悪しを評価可能であることを示した。栗井は [9], 個人の照合のために人物が同じ動作を複数回行い、それぞれの二次元姿勢情報から \cos 類似度を算出している。この際に関節間のベクトルだけでなく足首同士のベクトルなどの多数のベクトルを入力とすることで個人照合を可能としている。

3. 使用手法, 提案手法

3.1 提案手法の全体像

提案手法の流れを図 1 に示す。動画を入力として、OpenPose により二次元姿勢を推定する。その後、推定された二次元姿勢情報に対して時間微分を行うことで速度、加速度、躍度を算出し、動作解析を行うことで動作特徴の抽出

が可能であることを示す。また、 \cos 類似度により人物間の動作の細かな違いを検出可能であることを示す。

3.2 OpenPose による二次元姿勢推定

本研究では、二次元姿勢推定手法に OpenPose を採用している。OpenPose は、RGB カメラから取得される動画像に対して二次元姿勢推定を行い人物二次元関節座標を推定する手法である。本研究では、Blender によって出力された人物 3D モデルの動作を記録した動画を入力として、関節座標の時系列データを出力する。図 2 に OpenPose によって推定された二次元姿勢を示す。

3.3 Blender による人物動作データの取得

本研究では、人物の動作データを取得するに当たって、3D シミュレータの Blender を使用する。Blender 内の 3D モデルに MMD(MikuMikuDance) によって作成されたモーションデータを読み込ませることで、様々な動作データの撮影が可能としている。

3.4 人物動作の速度, 加速度, 躍度の算出

本研究では、OpenPose により推定された二次元関節座標に対して時間微分を行うことで、速度、加速度、躍度を算出している。速度は、動作の始めと終わりを観測するため絶対値で算出している。式 (1), (2), (3) にそれぞれ速度、加速度、躍度を求める手法を示す。

$$v(t_n) = \left| \frac{x(t_{n+1}) - x(t_n)}{t_{n+1} - t_n} \right| \quad (1)$$

$$a(t_n) = \frac{v(t_{n+1}) - v(t_n)}{t_{n+1} - t_n} \quad (2)$$

$$j(t_n) = \frac{a(t_{n+1}) - a(t_n)}{t_{n+1} - t_n} \quad (3)$$

3.5 \cos 類似度による動作データの比較

本研究では、人物間の動作の違いを比較するために \cos 類似度を使用する。 \cos 類似度は、2 種類のベクトル \vec{a}, \vec{b} を入力とし、ベクトル間の角度の類似度を表現する関数である。出力される値が 1 に近ければ類似しており、0 に近ければ類似していない事を示している。本研究では、 \cos 類似度へ入力するベクトルは、 t と $t-1$ の間の速度、加速度、躍度を入力とする。式 (4) に使用する \cos 類似度を算出する手法を示す。

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \quad (4)$$

4. 実験検証

4.1 実験環境

表 1 に本研究の実験環境を示す。

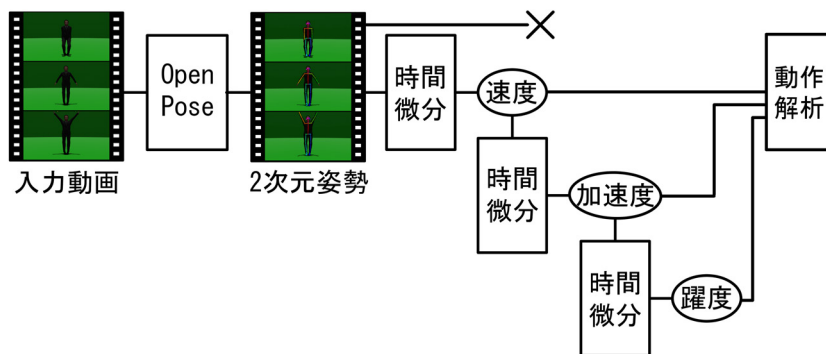


図 1 手法の全体像



図 2 OpenPose による二次元姿勢推定の例



図 3 実験で使用した人物モデル [11]

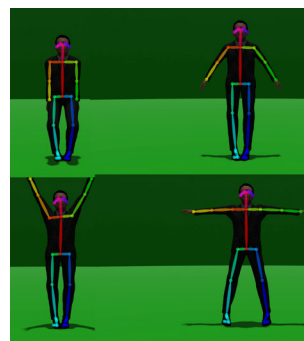


図 4 体操動作に対する二次元姿勢推定

表 1 実験環境

種類	型番
OS	ubuntu 20.04
CPU	AMD x570
RAM	8 GB
GPU	RTX3060
VRAM	10 GB
CUDA	11.1
cuda	8.2.1
OpenPose	1.7.0
Blender	2.9

表 2 実験で使用したカメラと撮影環境

各種設定	値
地面からの高さ (m)	0.85
被験者からの距離 (m)	3
縦 × 横 × 高さ (m)	90 × 25 × 25
解像度	1920 × 1080
焦点距離 (m)	2.7

4.2 3D モデルに対する二次元姿勢推定

本研究では、試験的なデータの取得のために、Blender 内の 3D モデルによるラジオ体操第 2 の動作に対して二次元姿勢推定を行う。これにより、ラジオ体操第 2 における 15 個の二次元関節座標の時系列データが出力される。図 3 に実験で使用したモデルの外観を示す。図 4 に体操動作に対する二次元姿勢推定の様子を示す。表 2 に Blender 内で使用したカメラのスペックと撮影環境を示す。図 5 に撮影環境の様子を示す。

本研究では、IntelRealsence を模したカメラスペックで撮影を行う。カメラとの人物モデルの距離を 3m とする。

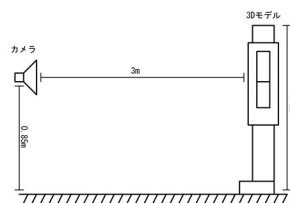


図 5 撮影環境の様子

この距離は、体を大きく動かした際に体がカメラに接触しないことや演者がストレスにならない距離を考慮した結果である。カメラから地面までの距離を 0.85m とする。この距離は、カメラに人物の全身をカメラに収める且つ、体の縦幅の中心を取るために、日本人の平均身長を 1.7m としたときの半分の値である。

4.3 速度、加速度、躍度に対する動作解析

本研究では、人物二次元姿勢推定した結果に対して時間微分を用いて速度、加速度、躍度を算出する。これら 3 種類の値の波形を 3D モデルの動作と照合し、波形から動作の特徴を抽出できることを確認する。今回使用する動作は

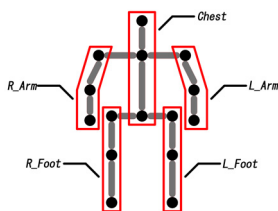


図 6 人物二次元姿勢のパーツ分割の方法

ラジオ体操第2の種目113の中で、種目1と種目11を使用する。種目1の動作は、小さな跳躍を三回繰り返す動作となっている。跳躍の最中に左右の肘を小さく横ぶりしている。種目11の動作は、大きな跳躍を三回繰り返す動作となっている。二回目の跳躍の最中に腕と脚を開いて着地した後、腕と足を閉じながら再び跳躍して着地する。腕は、左右に約90度開く。脚は約45度開く。

本実験では、関節を人体のパーツ毎に分割し、パーツ内で各フレームの平均値を算出する。算出された各パーツ毎の波形を用いて動作解析を行う。図6に人物二次元姿勢のパーツ分割の方法を示す。図7に種目1の動作の波形を、図8に種目11の動作の波形を示す。種目1、種目11は共に跳躍する動作であるため、Chest, R.arm, L.arm, R.foot, L.footのパーツは共通してY軸が周期的に増減している。この波形から跳躍の大きさに応じて波形の大きさが変化している。跳躍時の勢いの違いは、加速度、躍度の波形から確認可能である。種目1、種目11の動作は、ともにChestのパーツはY軸のみ増減しており、X軸はほとんど動いていない。このことから今回の動作には体を捻る動きが無いことを読み取ることが可能である。種目1の動作のR.arm, L.armの動作では、跳躍のたびに腕を横に揺らしているため、X軸の波形が常に小さく波打っている。しかし、R.arm, L.armのY軸の波形はChestの波形ほとんど差が無いため、両腕が上下に動いていないことを読み取ることが可能である。種目11の動作のR.arm, L.armの波形では、4205フレームから4215フレームの間で波形が他のフレームよりも大きくなっている。この動きは両腕を左右に開いた動作に対応していることがわかる。同様に、4225フレームから4240フレームにかけてX軸の増減がY軸の波形の増減にも影響している。この動きは、両腕を閉じる動作に対応していることを読み取ることが可能である。また、ChestのX軸の波形にはなかった膨らみが、Y軸の値が増加したフレーム間と同じ区間で起きている。このことから両腕を左右に大きく開いたことが読み取ることが可能である。R.foot, L.footにおいて種目1の動作では、X軸の波形の増減が無く、Y軸のみ増減している。種目11の動作では、脚を開く動作に対応し、X軸の膨らみが起きている。しかし、Y軸の波形には、R.arm, L.armのような大きな変化が現れていない。このことから、跳躍中の脚の動きは大きな動きではないことがわかる。

表 3 人物動作の cos 類似度による比較結果

	種目1と種目11の動作	2種類の種目1の動作
二次元姿勢	99.10	99.98
速度	24.18	82.62
加速度	22.78	84.51
躍度	7.79	73.84

本実験により、関節をパーツ毎に統合し、相互の関係性を考慮しながら解析することで動作の特徴を抽出できることを示した。また、加速度、躍度の波形を用いることで動作の勢いや動作の機敏さの特徴を抽出可能であることを示した。

4.4 cos 類似度による人物動作の比較

本研究では、似ている動作に対して、動作間の人物姿勢を入力とした cos 類似度と、動作間の速度、加速度、躍度の変化量を入力とした cos 類似度を比較する。本実験では、類似動作の他に、OpenPoseでは、入力動画のコーデックの違いによる関節座標の細かな誤差が発生する。この性質を利用して、一つの動作データから複数の人物の二次元姿勢のシーケンスデータを取得することと同義のデータの取得が可能であると考えられる。このデータを同じ動作であるが少し異なる動作として扱う。具体的には、AVIとMPEG-4を使用して二次元姿勢のシーケンスデータを取得する。

本実験では、類似動作と同じ動作であるが少し異なる動作についての cos 類似度を比較する。この実験により、速度、加速度、躍度に対する cos 類似度が、二次元姿勢に対する cos 類似度よりも、細かな動作の違いに対して反応しやすいことを示す。類似動作として、ラジオ体操第2の種目の中で種目1と種目11の脚を閉じた跳躍1回分のシーケンスデータを使用する。もう一方の同じ動作であるが少し異なる動作として、種目1の脚を閉じた跳躍1回分のシーケンスデータを AVI と MPEG-4 を使用して取得した。図9に種目1と種目11の人物姿勢に対する cos 類似度を示す。図10, 11に種目1と種目11の速度、加速度、躍度に対する cos 類似度を示す。図12に2種類の種目1の人物姿勢に対する cos 類似度を示す。図13, 14に2種類の種目1の速度、加速度、躍度に対する cos 類似度を示す。表3に各動作の cos 類似度による比較の結果を示す。

ラジオ体操第2に含まれる小さな跳躍動作と大きな跳躍動作を比較する。動作の細かな違いとして、跳躍する際の勢いが挙げられる。この勢いという要素は姿勢からは読み取れないため、人物姿勢を入力とした cos 類似度では一致に近い99.13点を算出された。速度、加速度、躍度の変化量を入力とした cos 類似度では、24.18点、22.78点、7.79点を算出された。小さな跳躍の AVI と MPEG-4 を比較した場合、人物姿勢を入力とした cos 類似度では99.98点を算出された。速度、加速度、躍度の変化量を入力とした cos

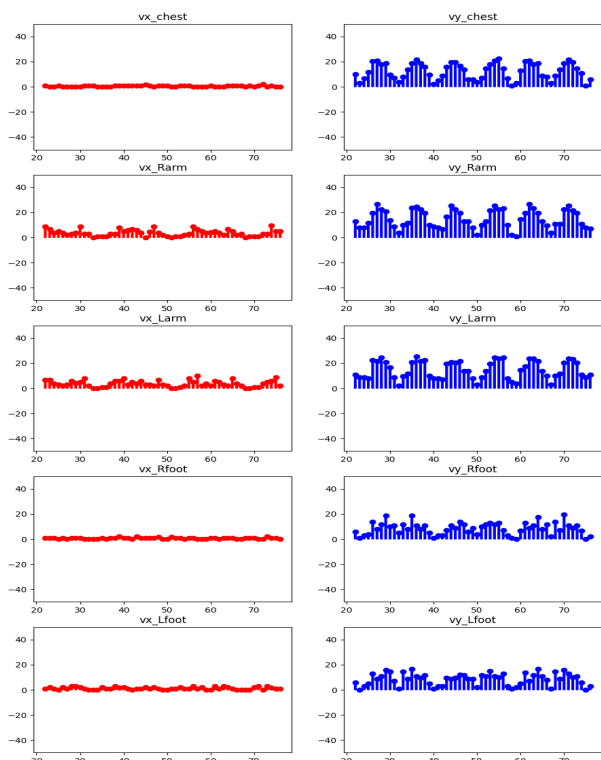


図 7 1 番の速度波形
(上から Chest, R_arm, L_arm, R_foot, L_foot
左: X 軸, 右: Y 軸)

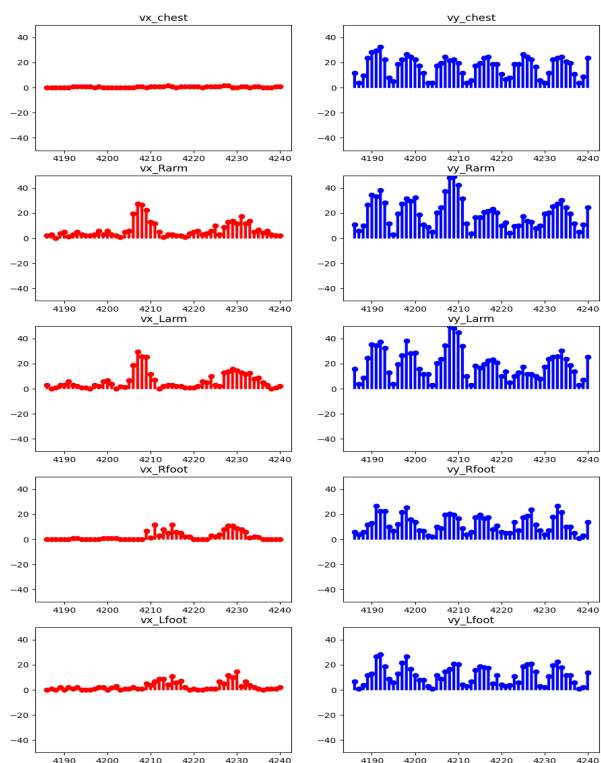


図 8 11 番の速度の波形
(上から Chest, R_arm, L_arm, R_foot, L_foot
左: X 軸, 右: Y 軸)

類似度では, 82.62 点, 84.51 点, 73.48 点を算出された.

以上の結果から, 人物姿勢を入力とした \cos 類似度では, 点数の変動は 1% 程度である. この結果に対して, 速度, 加速度, 躍度の変化量を入力とした \cos 類似度では, 勢いの違いに敏感に反応することが確認された. つまり速度, 加速度, 躍度を利用した教示システムの有用性を示すことが可能と考える.

5. まとめと今後の展望

本研究では, 人物の二次元姿勢から算出された速度・加速度・躍度を用いた動作解析によって人物間の動作の誤差を算出する動作の教示システムを提案し, 動作に対する波形の分析と動作の比較が可能であることを示した.

今後は, 三次元の動作解析に応用したい. そして, 実空間での日本舞踊の踊りや能楽の舞等に対して応用したモーションセマンティックスの推定により技能継承システムの向上を行いたい. その後, 舞い手の個人的動作特徴を抽出し個人の動作の癖などを含めた技能保存を行いたい.

参考文献

- [1] 白鳥貴亮, et al. "楽音のリズムを考慮した舞踊動作の解析." 第 21 回日本ロボット学会学術講演会 (2003).
- [2] 中澤篤志, et al. "モーションキャプチャによる全身運動解析と模倣ロボット-「じょんがら」節を HRP-1S に踊らせる." 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM) 2004.113 (2004-CVIM-146) (2004): 31-39.
- [3] 鹿内菜穂, et al. "見返りすぎない美人: 日本舞踊における振り向く動作の特徴と女性らしさの印象." 佐賀城下町『竈帳』と町絵図を用いた GIS 分析の試み: 49.
- [4] 董然, et al. "ヒルベルト-ファン変換を用いたダンスモーション解析." 電子情報通信学会論文誌 D 102.12 (2019): 843-853.
- [5] 柴田佳幸. "コサイン類似度を用いた人の姿勢の時系列データに基づく日常動作の自動分類とその認識." 大学院研究年報 理工学研究科編 46 (2016).
- [6] 柴田佳幸. "コサイン類似度を用いた人の姿勢の時系列データに基づく日常動作の自動分類とその認識." 大学院研究年報 理工学研究科編 46 (2016).
- [7] 清水ありさ, et al. "インタラクティブキャラクタにおける手先の追いかけ動作の再現." エンタテインメントコンピューティングシンポジウム 2015 論文集 2015 (2015): 551-556.
- [8] 西澤博大, et al. "姿勢推定を援用した実人物モデルの描画学習支援システム." 画像電子学会研究会講演予稿 画像電子学会第 285 回研究会講演予稿. 一般社団法人 画像電子学会, 2018.
- [9] 粟井修司, et al. "2 次元姿勢推定による歩容情報を用いた人物照合." 人工知能学会全国大会論文集 第 34 回全国大会 (2020). 一般社団法人 人工知能学会, 2020.
- [10] Cao, Zhe, et al. "OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields." IEEE transactions on pattern analysis and machine intelligence 43.1 (2019): 172-186.
- [11] 優しい世界を, "一般副総理 ASO", <https://3d.nicovideo.jp/works/td72040> (2021/12/21)

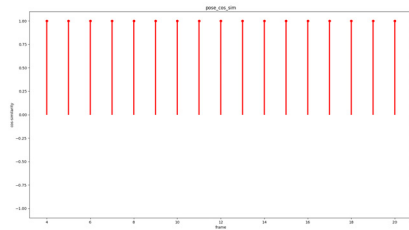


図 9 種目 1 と種目 11 の動作間の人物姿勢に対する cos 類似度

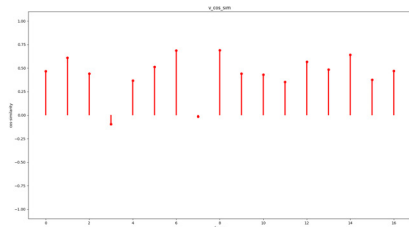


図 10 種目 1 と種目 11 の動作間の速度に対する cos 類似度

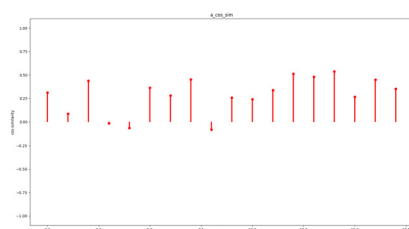


図 11 種目 1 と種目 11 の動作間の加速度に対する cos 類似度

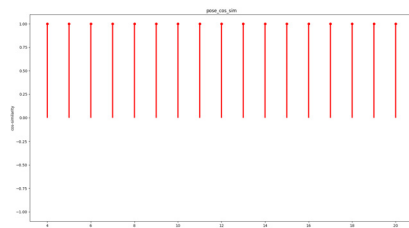


図 12 2 種類の種目 1 の動作間の人物姿勢に対する cos 類似度

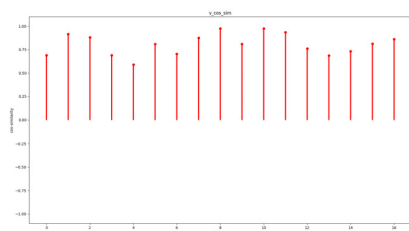


図 13 2 種類の種目 1 の動作間の速度に対する cos 類似度

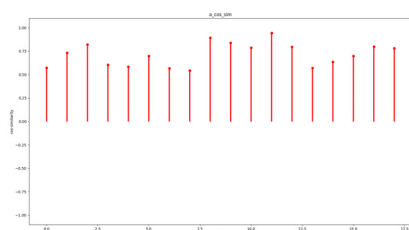


図 14 2 種類の種目 1 の動作間の加速度に対する cos 類似度