

修士論文概要書

Summary of Master's Thesis

Date of submission: (01/29/2020)

専攻名(専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	稲田 健太郎	指導 教員 Advisor	渡辺 裕 〇
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番号 Student ID number	5118F010-9 ^{CD}		
研究題目 Title	姿勢情報を用いた複数ダンサーの動作評価 Evaluation of Multiple Dancers Using Pose Information				

1. まえがき

中学校の体育授業において、2012年度からダンス科目が必修化された。また、YouTubeなどの動作サイトに自分たちのダンス動画をアップロードする人が増加している。現状では、一般人がダンスを踊り、ダンスは視聴者の主観で評価されるものであり、評価する人によって評価が異なる。複数人のダンスが一致しているかどうかを数値で表現することで、一つの指標となり公平な主観評価を支援できる。

本論文では、複数人が同じ振り付けのダンスをする動画を対象とし、動画内の全ダンサーから評価の基準となる時系列データを作成し、動作のタイミングを評価する手法を検討する。

2. 関連研究

2.1. OpenPose

Caoらが提案したOpenPose[1]を用いて2次元画像での姿勢推定を行う。OpenPoseは、単一画像から複数人の肩・肘・目など18点または25点のキーポイントの位置推定を行い、キーポイントの2次元画像上の座標情報を取得する。

2.2. Soft-DTW

Cuturiらが提案したSoft-DTW[2]は、Dynamic Time Warping(DTW)の2系列間のDTW距離計算を微分可能にした手法である。長さが任意の2時系列データ $A = \{a_1, a_2, \dots, a_M\}$, $B = \{b_1, b_2, \dots, b_N\}$ が与えられたとき、平滑化パラメータ $\gamma (> 0)$, 2データ間の距離関数 C を用いて2時系列データのコスト行列 d_γ を式(1,2)のように定義する。このときのSoft-DTWによる合計距離 $D_\gamma(A, B)$ は、 $d_\gamma(M, N)$ となる。

$$d_\gamma(m, n) = C(a_m, b_n) + \min_\gamma \left\{ \begin{array}{l} d_\gamma(m-1, n), \\ d_\gamma(m, n-1), \\ d_\gamma(m-1, n-1) \end{array} \right\} \quad (1)$$

$$\min_\gamma \{s_1, s_2, \dots, s_H\} = -\gamma \log \sum_{h=1}^H e^{-s_h/\gamma} \quad (2)$$

また、 I 個の入力時系列データ $Q_i (i \leq I)$ から式(3)の $f(P)$ を最小化することにより、平均時系列データ P を作成する。 P の初期値は、全入力時系列データの平均とする。

$$f(P) = \sum_{i=1}^I D_\gamma(P, Q_i) \quad (3)$$

3. 提案手法

3.1. 平均類似度による対応付け

2時系列多次元データ長さ M の A , 長さ N の B が入力された場合、式(4)によってベクトル系列 A' , B' を作成する。次に、平均類似度が最大になるような A' と B' の対応付けを行う。平均類似度とは、累積類似度をそれまでの累積対応付け数で割ったものである。このときの類似度関数 Sim を式(5)のように定義する。

$$A' = \{(a_2 - a_1), (a_3 - a_2), \dots, (a_M - a_{M-1})\} \\ = \{\vec{a}'_1, \vec{a}'_2, \dots, \vec{a}'_{M-1}\} \quad (4)$$

$$Sim(\vec{a}, \vec{b}) = \begin{cases} 1 & \|\vec{a}\|_2 = 0 \text{ and } \|\vec{b}\|_2 = 0 \\ 0 & \|\vec{a}\|_2 = 0 \text{ xor } \|\vec{b}\|_2 = 0 \\ \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 \|\vec{b}\|_2} & \text{otherwise} \end{cases} \quad (5)$$

3.2. 基準時系列データの作成

動画内の全ダンサーのOpenPoseによる姿勢推定結果から、両手首と両足首の4部位に対して動作評価の基準となるデータを作成する。まず、2.2節の平均データ作成手法を用いて、全ダンサーの時系列データから平均時系列データを作成する。次に、この平均時系列データと各ダンサーの時系列データを、平均類似度による対応付けを行う。全ダンサーの対応付けの合計時間差を計算する。平滑化パラメータ γ を変化させて、合計時間差が最小になるときの平均時系列データを評価の基準時系列データとする。

3.3. 動作変化点の検出

各部位の基準時系列データから、動作の方向が $\pi/4$ [rad]以上変化するフレーム、動作が開始するフレーム、動作が終了するフレームを検出する。時系列多次元データの変化量ベクトルは、動作のスピードと方向を表す。よって、任意フレームにおいて、当該データと前データの変化量ベクトルと次データと当該データの変化量ベクトルの内積が $\sqrt{2}/2$ 以下になるとき、当該フレームが動作変化点となる。また、当該データと前データの変化量ベクトルと次データと当該データの変化量ベクトルのどちらかのみが零ベクトルであるときも、動作変化点とする。

3.4. 動作評価

各部位に対して、各ダンサーと基準時系列データ

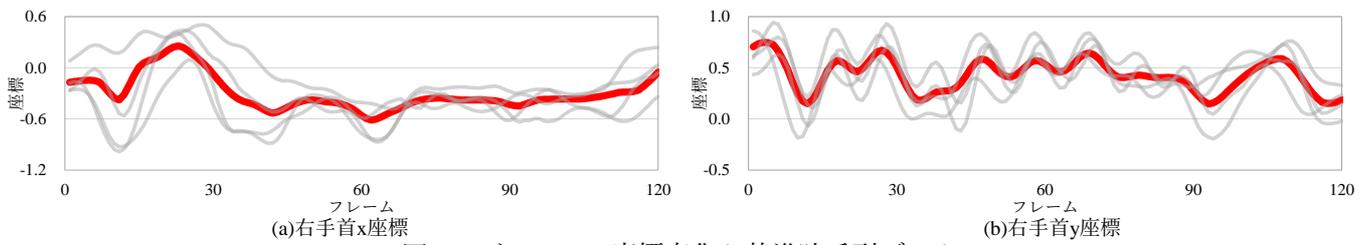


図1 ダンサーの座標変化と基準時系列データ

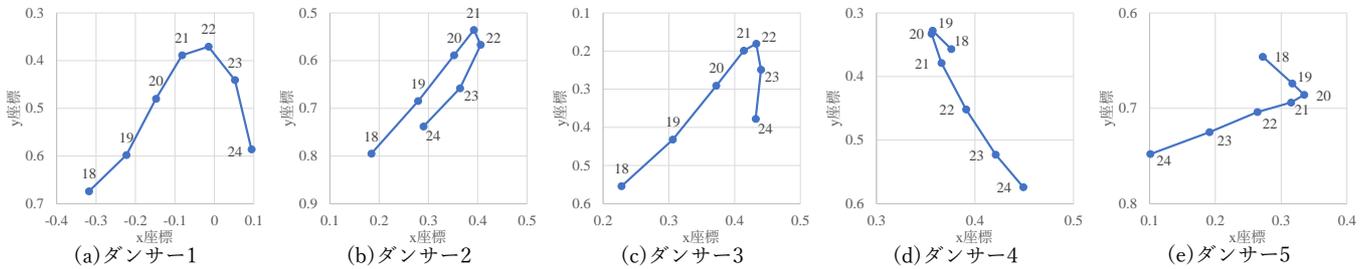


図2 右手首の座標変化 (18 フレーム～24 フレーム)

を平均類似度による対応付けを行い、3.2節で検出した動作変化点と対応付けられたフレームとのフレーム差 Δt を算出する。ダンサーの動作を評価するために、閾値 T 、 P を設定する。 $\Delta t < -T$ が成り立つときその動作が基準より早いと判断し、 $\Delta t > T$ が成り立つときその動作が基準より遅いと判断する。早い(遅い)と判断した数の動作変化点数における割合が P を越える場合、当該部位は基準より早い(遅い)と評価する。4部位のうち1部位以上でも早い(遅い)と評価した場合、そのダンサーの総合評価を「早い(遅い)ところがある」とする。ここで、早いと遅いが存在する場合、そのダンサーは「早いところがあり、かつ遅いところがある」と評価する。

4. 評価実験

4.1. 基準時系列データの作成

評価実験では、5人が同じ振り付けのダンスを踊る5秒間の動画を使用し、左から順にダンサー1から5とした。3.2節で提案した手法を用いて、基準時系列データを作成した。右手首における全ダンサーの座標変化と作成した基準時系列データを図1に示す。図中の灰色と赤色の線は、それぞれ全ダンサーの座標変化と基準時系列データを表す。図1から、作成した基準時系列データが、全ダンサーの座標変化の基準として適切であるといえる。

4.2. 動作変化点の検出

3.3節で提案した手法により検出した動作変化点と、各ダンサーの動作変化フレームを比較する。右手首の動作変化点として検出された21フレーム付近の各ダンサーの座標変化を図2に示す。図2より、全ダンサーが21フレーム付近に動作が変化していると推測でき、動作変化点を検出できるといえる。

4.3. 動作評価

3.4節で提案した手法により、閾値 T 、 P をそれぞれ2、1/3と設定し、各ダンサーを評価する。また、提案手法を評価するために、ダンス視聴者6人に各

ダンサーが周りに比べて早い、正常、遅いと感じるかどうかなンケートを実施した。提案手法による評価結果とアンケートで各ダンサーが早いまたは遅いと回答した人数を表1に示す。表の「早い、遅い」は、評価結果が「早いところがあり、かつ遅いところがある」である。表1より、提案手法による評価とアンケート結果が一致する場合と相違がある場合がある。人間が主観的に感じられた動作が早いダンサーと遅いダンサーの検出だけでなく、主観的に感じられない早いところや遅いところの検出ができたといえる。これにより、ダンス視聴者の公平な主観評価を支援できたといえる。

表1 評価結果とアンケート結果

	提案手法	アンケート	
		早い[人]	遅い[人]
ダンサー1	正常	0	2
ダンサー2	正常	1	0
ダンサー3	正常	0	2
ダンサー4	早い、遅い	5	0
ダンサー5	早い、遅い	2	0

5. むすび

動画内の全ダンサーから基準時系列データの作成手法、動作変化点の検出手法、動作評価手法を提案した。また、提案手法による評価と主観評価を比較した。提案手法により、適切な基準時系列データを作成できた。また、全ダンサーの動作変化フレームが、検出した動作変化点付近であることを確認した。また、視聴者が感じられない早いダンサーと遅いダンサーの検出ができ、公平な主観評価を支援できた。

参考文献

- [1] Z. Cao, T. Simon, S-E Wei, Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), No.121, pp.1302-1310. July 2017.
- [2] M. Cuturi, M. Blondel, "Soft-DTW: a Differentiable Loss Function for Time-Series", The International Conference on Machine Learning (ICML), Vol. 70, pp.894-903

2019 年度

早稲田大学大学院基幹理工学研究科情報理工・情報通信専攻 修士論文

姿勢情報を用いた複数ダンサーの動作評価

Evaluation of Multiple Dancers Using Pose Information

稲田 健太郎

(5118F010-9)

提出日：2020.01.29

指導教員：渡辺 裕 印

研究指導名：オーディオビジュアル情報処理研究

目次

第1章 序論.....	1
1.1 研究背景.....	1
1.2 研究目的.....	1
1.3 本論文の構成.....	2
第2章 関連技術.....	3
2.1 まえがき.....	3
2.2 OpenPose.....	3
2.3 コサイン類似度.....	4
2.4 Dynamic Time Warping.....	4
2.5 Angular Metrics for Shape Similarity.....	5
2.6 Soft-DTW.....	7
2.6.1 Soft-DTW の概要.....	7
2.6.2 Soft-DTW による平均時系列データ作成.....	9
2.7 むすび.....	10
第3章 提案手法.....	11
3.1 まえがき.....	11
3.2 提案手法概要.....	11
3.3 検出座標値の正規化.....	12
3.4 基準時系列データ作成.....	13
3.5 動作変化点の検出.....	17
3.6 評価手法.....	18
3.7 むすび.....	19
第4章 評価実験, 結果および考察.....	20
4.1 まえがき.....	20
4.2 実験概要.....	20

4.3	実験 1	23
4.4	実験 2	28
4.5	実験 3	32
4.5.1	提案手法による評価	32
4.5.2	主観評価との比較	35
4.6	むすび	36
第 5 章	結論	37
5.1	結論	37
5.2	今後の課題	37
	謝辞	38
	参考文献	39
	図一覧	40
	表一覧	41
	研究業績	42

第1章 序論

1.1 研究背景

義務教育の一部である中学校の体育授業では、それまで選択性だったダンス科目が 2012 年度から必修化された[1]。これにより、原則として中学生全員がダンスを経験することとなり、担当教員からダンスの出来栄えなどによって評価されている。文部科学省の中学校指導要領保健体育編では、ダンスのカリキュラムが「創作ダンス」、「フォークダンス」、「現代的なリズムのダンス」で構成されている。また、3種類のダンスを通して、「仲間と踊る楽しさを味わい、イメージをとらえた表現ができるようにする」と指導要領に記載されている[2]。しかし、担当教員はダンスを経験していない場合がある。文部科学省が発行する指導要領によって評価基準が設けられているが、ダンスの評価は担当教員の主観評価に依存している。

また、アイドルグループやアーティストが簡単なダンスを踊りながら歌うことが多くなっている。そのダンスが流行となり、複数人数で踊る様子をスマートフォンなどで動画に撮り「踊ってみた動画」と名付けて YouTube などの動画サイトにアップロードする人が増加している。その動画を視聴する不特定多数の人物から、動きのキレやかかわいらしさなどといった様々な尺度で動画内の人物が躍るダンスの出来栄えが評価されている。

これらのことから、若い世代を中心にダンスというものが、以前と比較して身近なものに変化しているといえる。今まではテレビ等でアイドルグループやパフォーマーが躍っている姿を見るものであったのに対して、現在では実際に自分たちが踊り他人に評価されるものもしくは他人のダンスを評価するものに変化している。

1.2 研究目的

ダンスは視聴者の主観で評価されるものであり、評価する人によって評価が異なる。複数人のダンスが一致しているかどうかを数値で表現することで、一つの指標となり公平な主観評価を支援できる。また、指導者が不在のダンスグループなどに対して、技術向上を手助けできる。

本研究では複数人のダンス動画を対象として、動画内の全ダンサーから動作のタイミング評価の基準を作成する。次に、各人物に対して動作のタイミングが評価基準とどれだけ一

致しているかを評価する。ダンス動画から OpenPose によって取得した身体座標データを利用し、ダンサーの動作を評価する手法を提案する。

1.3 本論文の構成

以下に本章以降の構成を示す。

第 1 章は本章であり、本論文の研究の背景および目的について述べている。

第 2 章では、本研究で用いる身体部位座標の取得手法である OpenPose、本研究で用いる時系列データ分析の関連技術を説明する。

第 3 章では、複数人のダンス動画から各ダンサーの評価手法を提案する。

第 4 章では、第 3 章で述べた提案手法における実験および結果について示す。

第 5 章では、本研究のまとめと今後の課題を示す。

第2章 関連技術

2.1 まえがき

本章では本研究で利用する関連技術について述べる。

2.2 OpenPose

画像内の人物の部位を推定する手法として、Cao らが提案した Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields (OpenPose) [3]がある。この手法は、画像内に映る複数の人物の部位座標を取得できる。首、肘、膝などの主要な 15, 18 もしくは 25 部位を検出可能である。さらに、画像内に手や顔が十分な大きさで存在する場合、指の関節や顔のパーツの輪郭も検出できるモデルがある。検出結果は、画像上での各部位の x 座標, y 座標および検出の確度が出力される。18 部位を検出できるモデルでは、鼻、首、両肩、両肘、両手首、両腰、両膝、両足首、両目、両耳が検出される。25 部位を検出できるモデルでは、18 部位に加えてつま先やかかとなど足の詳細な部位が検出できる。本研究では 25 部位を検出できるモデルを使用した。OpenPose の検出結果の例として、入力画像と出力画像を図 2.1 に示す。



図 2.1 OpenPose の検出結果

左：入力画像 右：出力画像

OpenPose は、検出する部位の存在する確率を示す Confidence Maps と部位同士を接続させていく Part Affinity Field で構成されている。従来の姿勢推定手法は、画像中の人物を検出した後、検出人物に対して部位を検出する。しかし OpenPose では人物検出せず、部位を検出し部位同士を接続する。身体の一部が物体や人物によって重なり画像内に存在しない場合でも、存在している部位を検出できるだけでなく、存在していない部位も推定できる。しかし、存在しない部位の推定精度は低く、正しく推定できない場合がある。OpenPose は、画像だけでなく動画に対しても有効であるが、フレーム毎に検出するため時系列を無視して検出している。

2.3 コサイン類似度

ベクトル同士の類似度計算手法として、コサイン類似度がある。コサイン類似度は、ベクトル同士の成す角度によって類似度を表現する。三角関数のコサインと同様に、-1 から 1 の範囲で値をとり、1 に近いほど類似度が高く、-1 に近いほど類似度が低い。 \vec{x} と \vec{y} のベクトルの成す角を θ としたとき、 \vec{x} と \vec{y} のコサイン類似度は式 (2.1) のように表すことができる。

$$\cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|_2 \|\vec{y}\|_2} \quad (2.1)$$

2.4 Dynamic Time Warping

2 時系列データ間の類似度を算出する手法として、Dynamic Time Warping (以下, DTW) がある。DTW は、2 時系列データの各要素の距離を総当たりで計算し、2 時系列データ間の合計距離 (DTW 距離) が最小になるように対応付けをする。ここで、2 時系列データの長さが異なる場合でも、DTW による対応付けが可能である。

2 時系列データ $A = \{a_1, a_2, \dots, a_M\}$, $B = \{b_1, b_2, \dots, b_N\}$ の DTW 距離 $D(A, B)$ を式 (2.2) のように定義する。

$$D(A, B) = d(M, N) \quad (2.2)$$

2 時系列データ間の距離関数 C とし、一方の時系列データの要素 a_m ともう一方の時系列データの要素 b_n の距離を $C(a_m, b_n)$ とするとき、 $d(m, n)$ を式 (2.3) のように再帰的に算出する。

$$d(m, n) = C(a_m, b_n) + \min \left\{ \begin{array}{l} d(m-1, n), \\ d(m, n-1), \\ d(m-1, n-1) \end{array} \right\} \quad (2.3)$$

このときの条件式は式 (2.4) である.

$$\begin{aligned} d(0,0) &= 0, & d(m,0) &= d(0,n) = \infty \\ 1 \leq m \leq M, & & 1 \leq n \leq N, & & m, n \in \mathbb{N} \end{aligned} \quad (2.4)$$

DTW の欠点として, ノイズなどに弱いことが挙げられる. ノイズが発生すると, 本来対応すべき要素間の距離と実際の距離が異なることがあり, 不正確な対応付けが生じる.

DTW の距離関数として 2 要素間のユークリッド距離が用いられることが多い. このとき時系列データの時間方向の伸縮や移動に強いが, 空間方向の伸縮や移動に弱い. つまり, 2 時系列データが同じ軌跡だが空間方向に移動している場合, 正確な対応付けができない. DTW による対応付けの例を図 2.2 に示す.

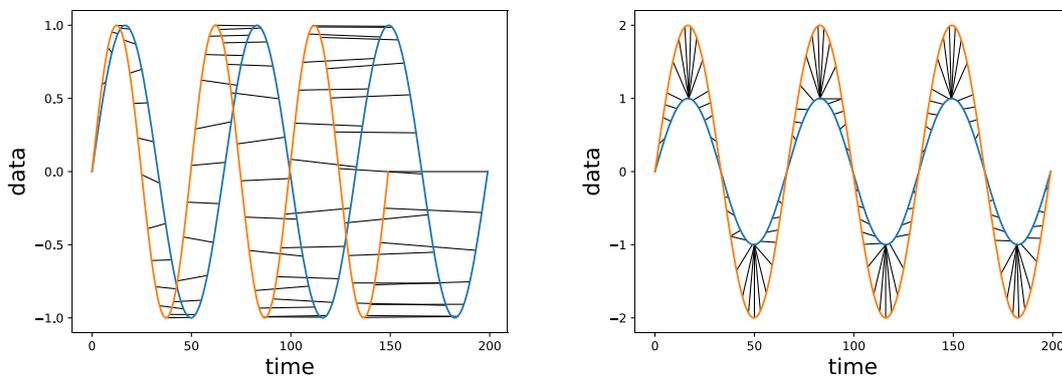


図 2.2 時間・空間方向に伸縮された時系列データに対する DTW 例

左：時間方向の伸縮 右：空間方向の伸縮

2.5 Angular Metrics for Shape Similarity

DTW を拡張した 2 時系列データ間の類似度算出手法として, 中村らが提案した Angular Metrics for Shape Similarity (以下, AMSS) [4]がある. 通常の DTW は, 2.4 節の式 (2.3) のように 2 時系列データ間の DTW 距離が最小になるような対応付けをする. これに対して

AMSS は、時系列データからベクトル系列へ変換し、2 ベクトル系列間の類似度を算出する。距離関数としてコサイン類似度を用いて、合計距離が最大になるような対応付けをする。

2 時系列データ $A = \{a_1, a_2, \dots, a_M\}$, $B = \{b_1, b_2, \dots, b_N\}$ の AMSS による対応付けの手順を以下に示す。時系列データ A から、式 (2.5) に従ってベクトル系列 $A' = \{\vec{a}'_1, \vec{a}'_2, \dots, \vec{a}'_{M-1}\}$ を作成する。時系列データからベクトル系列への変換の概要を図 2.3 に示す。

$$\begin{aligned} A' &= \{(a_2 - a_1), (a_3 - a_2), \dots, (a_M - a_{M-1})\} \\ &= \{\vec{a}'_1, \vec{a}'_2, \dots, \vec{a}'_{M-1}\} \end{aligned} \quad (2.5)$$

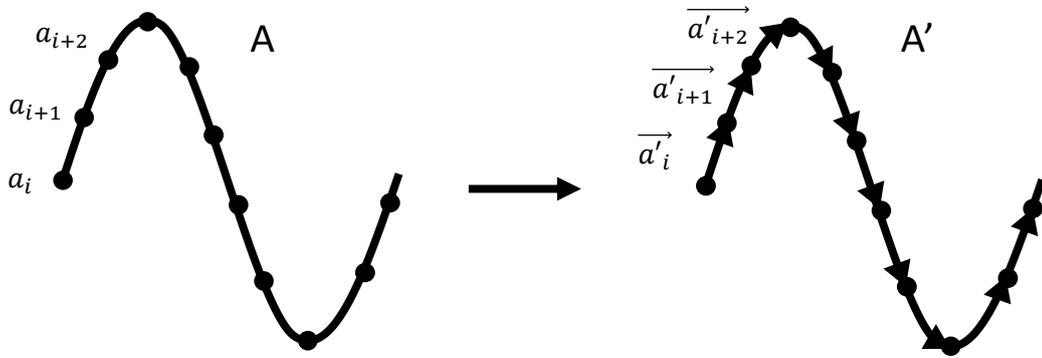


図 2.3 時系列データからベクトル系列への変換

同様に時系列データ B から、ベクトル系列 $B' = \{\vec{b}'_1, \vec{b}'_2, \dots, \vec{b}'_{N-1}\}$ を作成する。2 ベクトル系列 $A' = \{\vec{a}'_1, \vec{a}'_2, \dots, \vec{a}'_{M-1}\}$, $B' = \{\vec{b}'_1, \vec{b}'_2, \dots, \vec{b}'_{N-1}\}$ の合計類似度 $D'(A', B')$ を式 (2.6) のように定義する。

$$D'(A', B') = d'(M - 1, N - 1) \quad (2.6)$$

\vec{a}'_m と \vec{b}'_n の類似度 $S(\vec{a}'_m, \vec{b}'_n)$ として、 $d'(m, n)$ を式 (2.7) のように再帰的に算出する。

$$d'(m, n) = S(\vec{a}'_m, \vec{b}'_n) + \max \left\{ \begin{array}{l} d'(m - 1, n), \\ d'(m, n - 1), \\ d'(m - 1, n - 1) \end{array} \right\} \quad (2.7)$$

ここで類似度 $S(\vec{a}'_m, \vec{b}'_n)$ は、 \vec{a}'_m と \vec{b}'_n の成す角 θ_{mn} を用いて式 (2.8) のように定義する。

$$S(\vec{a}'_m, \vec{b}'_n) = \begin{cases} \frac{\vec{a}'_m \cdot \vec{b}'_n}{\|\vec{a}'_m\|_2 \|\vec{b}'_n\|_2} & \theta_{mn} < \frac{\pi}{2} \\ 0 & \theta_{mn} \geq \frac{\pi}{2} \end{cases} \quad (2.8)$$

このときの条件式は式 (2.9) である.

$$\begin{aligned} d'(0,0) &= 0, & d'(m,0) &= d'(0,n) = -\infty \\ 1 \leq m \leq M-1, & & 1 \leq n \leq N-1, & & m, n \in \mathbb{N} \end{aligned} \quad (2.9)$$

式 (2.8) のように 2 ベクトルの成す角 θ_{mn} が $\pi/2$ [rad] より大きいとき, AMSS では 2 ベクトル間の類似度 S を 0 に設定している. これにより, ノイズにより 2 つのベクトルの向きが真逆に近い場合でも, ベクトル同士の類似度は 0 であるため合計距離の最小化に影響を与えづらくなる. また, AMSS は時系列データの要素間の変化量によって対応付けをするため, 時系列データの空間方向の伸縮や移動に強い. つまり, 2 時系列データの軌跡の類似度による対応付けをする. 以上より AMSS は, DTW の欠点であるノイズに弱いことと時系列データの空間方向の伸縮や移動に弱いことに対して, DTW よりも正確な対応付けが可能である.

2.6 Soft-DTW

2.6.1 Soft-DTW の概要

DTW 距離を最小化する操作を微分可能にした手法として, Cuturi らが提案した Soft-DTW[5]がある. DTW は DTW 距離が最小になるような 2 時系列データの対応付けをするが, Soft-DTW は対応付けを行わずに DTW 距離を計算する. Soft-DTW は時系列データの空間方向のスケールによって, 平滑化パラメータ $\gamma (\geq 0)$ を決定する必要がある. このとき, スケールが小さいほど γ は小さくなり, 小さい値が DTW 距離に与える影響が大きくなる.

2 時系列データ $A = \{a_1, a_2, \dots, a_M\}$, $B = \{b_1, b_2, \dots, b_N\}$ の Soft-DTW による DTW 距離 $D_\gamma(A, B)$ を式 (2.10) のように定義する.

$$D_\gamma(A, B) = d_\gamma(M, N) \quad (2.10)$$

2.4 節の式 (2.3) と同様に 2 時系列データ間の距離関数 C とし、一方の時系列データの要素 a_m ともう一方の時系列データの要素 b_n の距離は $C(a_m, b_n)$ とするとき、 $d_\gamma(m, n)$ を式 (2.11) のように再帰的に算出する。Soft-DTW の DTW 距離算出方法の概要を図 2.4 に示す。

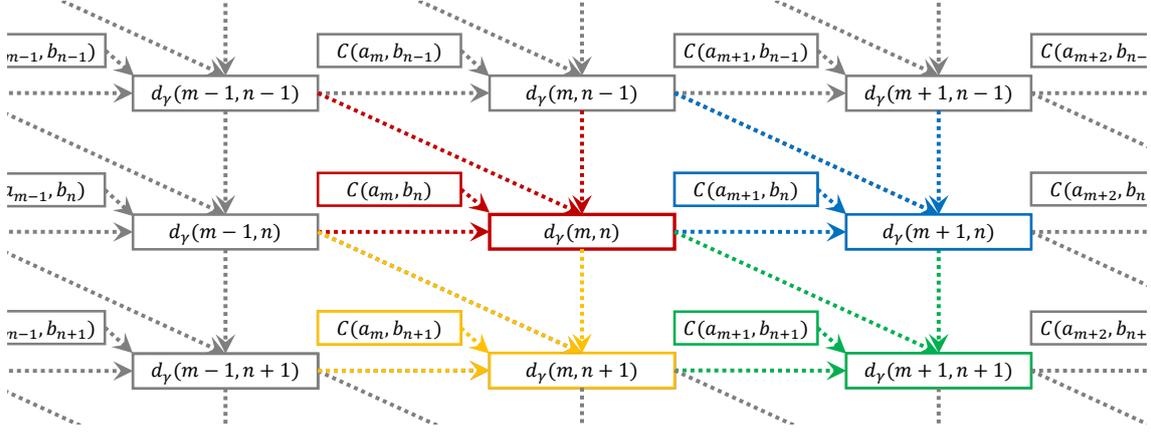


図 2.4 Soft-DTW の概要

$$d_\gamma(m, n) = C(a_m, b_n) + \min_\gamma \left\{ \begin{array}{l} d_\gamma(m-1, n), \\ d_\gamma(m, n-1), \\ d_\gamma(m-1, n-1) \end{array} \right\} \quad (2.11)$$

また、 $\gamma = 0$ を満たすとき Soft-DTW ではなく通常の DTW を行うため、 $\min_\gamma \{s_1, s_2, \dots, s_H\}$ を式 (2.12) のように定義する。

$$\min_\gamma \{s_1, s_2, \dots, s_H\} = \begin{cases} -\gamma \log \sum_{h=1}^H e^{-s_h/\gamma} & \gamma > 0 \\ \min \{s_1, s_2, \dots, s_H\} & \gamma = 0 \end{cases} \quad (2.12)$$

このときの条件式は式 (2.13) である。

$$\begin{aligned} d_\gamma(0,0) &= 0, & d_\gamma(m,0) &= d_\gamma(0,n) = \infty \\ 1 \leq m \leq M, & & 1 \leq n \leq N, & \quad m, n \in \mathbb{N} \end{aligned} \quad (2.13)$$

2.6.2 Soft-DTW による平均時系列データ作成

Cuturi らは 2.6.1 節の Soft-DTW を用いた複数の入力時系列データから平均時系列データを作成する手法を提案している。平均時系列データと各入力時系列データとの Soft-DTW による DTW 距離の合計が最小となるように、平均時系列データを作成する。つまり、式(2.14)の関数 $f(P)$ を最小化する問題を解くことと同値になる。

$$f(P) = \sum_{i=1}^I \frac{\lambda_i}{l_i} D_{\gamma}(P, Q_i) \quad (2.14)$$

ここで、 P は任意の要素数 L の平均時系列データ、 I は入力時系列データ数、 Q_i は i 番目の入力時系列データ、 l_i は i 番目の入力時系列データの要素数、 λ_i は i 番目の入力時系列データの重み ($\lambda_1 + \lambda_2 + \dots + \lambda_I = 1$) を表している。 P の初期値は、全入力時系列データの平均とする。このとき、 l_i が L と異なる場合は 3 次スプライン補間を用いて、 Q_i が要素数 L になるように補間する。

Soft-DTW を用いて、複数時系列データから平均時系列データを作成する例を図 2.5 に示す。灰色の実線は正弦波を時間方向に移動した 5 個の入力時系列データを表し、赤色の実線は作成した平均時系列データを表している。

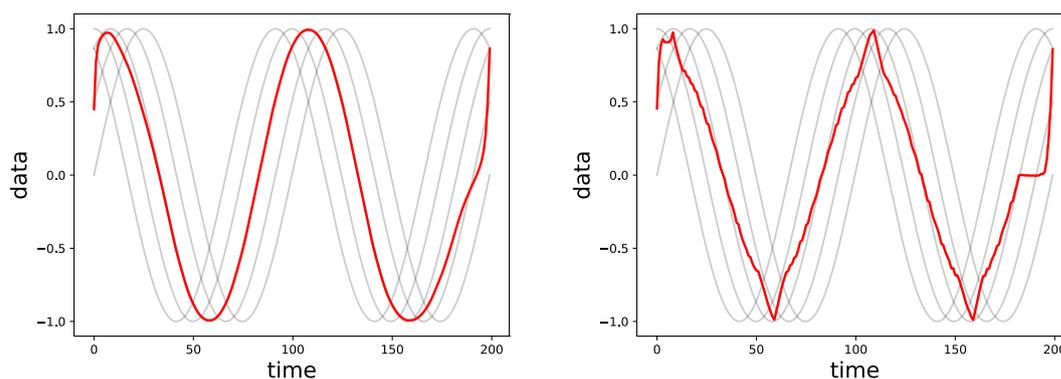


図 2.5 Soft-DTW による平均時系列データの作成

左 : $\gamma = 0.01$ 右 : $\gamma = 0.0001$

図 2.5 のように、平均時系列データの軌跡は γ によって変化する。 γ が大きくなるほど滑らかな軌跡を描く平均時系列データが作成でき、 γ が小さくなるほど粗い軌跡を描く平均時系

列データが作成できる．作成する平均時系列データの妥当性をどのように定義するかによって， γ の適切な値が変化する．

2.7 むすび

本章では，本研究に関連する技術や従来手法について示した．

第3章 提案手法

3.1 まえがき

本章では，提案手法の具体的な内容について述べる。

3.2 提案手法概要

本研究の提案手法は，ダンス動画からダンサーの動作一致度を評価する方法である．複数人が同じ振り付けのダンスをしている動画を対象とする．提案手法の概要を図 3.1 に示す．

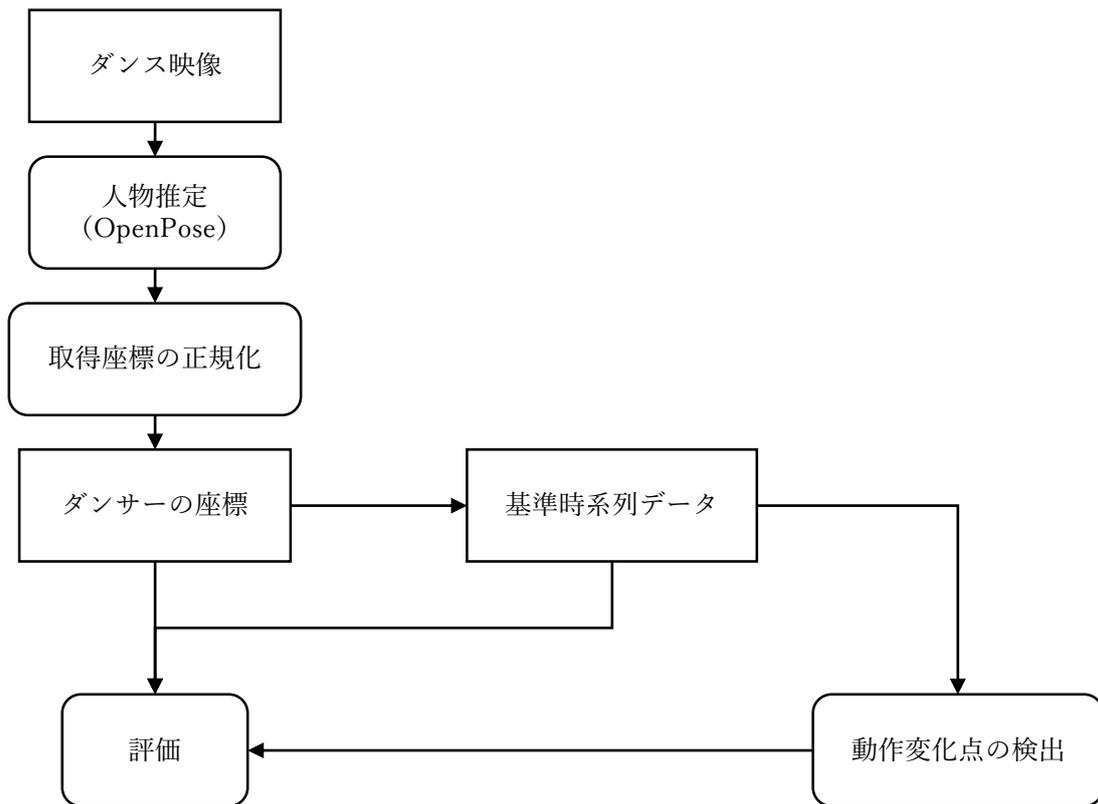


図 3.1 提案手法の概要

ダンス動画から OpenPose を用いて、ダンサーの部位座標値を取得する。ダンスは肩より先の腕部分や腰より先の脚部分の動作が細かく、ダンスの表現において重要になる。そのため、OpenPose の推定部位のうち、首、両手首、腰の中心、両足首の 6 部位のみを扱い、動作の評価には両手首、両足首の 4 部位を使用する。OpenPose の推定失敗フレームの座標値を前後の推定成功座標から推測する。ダンサーの身長や立ち位置によって取得座標系のスケールが異なるなどの理由から座標値の正規化を行う必要がある。次に、動画内の全ダンサーの座標時系列データから、評価基準となる時系列データを作成する。ここで、創作ダンスなどでは正解モデルの作成が困難であることから、動画内のダンサーから評価基準を作成する。作成した基準時系列データから動作が変化するフレームを検出する。この動作変化点において、動画内のダンサーの動作が変化するフレーム、動作が開始するフレーム、動作が終了するフレームが評価基準とどれだけずれているかを評価する。

3.3 検出座標値の正規化

OpenPose による推定は、フレームごとに行われる。推定が失敗した場合、そのフレームの推定失敗部位の座標値は 0 になる。推定失敗区間は、前後の推定成功結果から補正する。推定失敗区間内は、区間前後の推定成功結果の座標値差を等速に変化しているとして補正する。座標値のフレーム補正の概要を図 3.2 に示す。

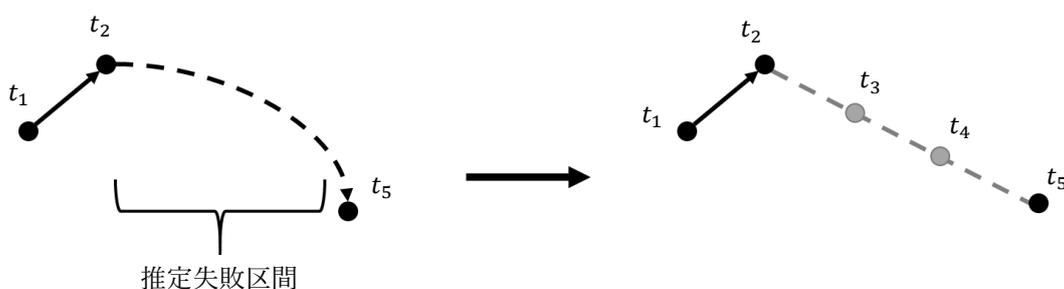


図 3.2 座標値のフレーム補正の概要

本研究で用いるダンス動画は、定点カメラで撮影された動画である。カメラ位置が固定されていても、動画内のダンサーは常に動いている。そのため、各人物の座標値も常に変化している。そこで正確な評価を行うために、各ダンサーに対して首が原点になるように両手首、両足首の 4 部位の座標値を補正する。このときの番号と部位名の対応を表 3.1 に示す。補正

関数を f ，検出ダンサー数を I ，人物 i の首の座標値を $\mathbf{Y}_{i,neck} = (x_{i,neck}, y_{i,neck})$ とするとき，式 (3.1) のように各部位 j の座標を移動する．

表 3.1 番号と部位名の対応

番号	部位名
1	右手首
2	左手首
3	右足首
4	左足首

$$f(\mathbf{Y}_i) = \mathbf{Y}_{i,j} - \mathbf{Y}_{i,neck} = (x_{i,j} - x_{i,neck}, y_{i,j} - y_{i,neck}) \quad (3.1)$$

このときの条件式は式 (3.2) である．

$$\begin{aligned} \mathbf{Y}_i &= \{\mathbf{Y}_{i,1}, \mathbf{Y}_{i,2}, \dots, \mathbf{Y}_{i,4}\} \\ 1 \leq i \leq I, \quad 1 \leq j \leq 4, \quad i, j \in \mathbb{N} \end{aligned} \quad (3.2)$$

また，ダンサーの動画内の立ち位置や身長によって大きさが変化する．本研究では，正面から撮影されたダンス動画を使用するため，立ち位置と胴体の長さが比例している．そこで，各ダンサーに対して首と両腰の中心の距離が1になるように4部位の座標をスケール変換する．変換関数を g ，検出ダンサー数を I ，首の座標値を $\mathbf{Y}_{i,neck} = (x_{i,neck}, y_{i,neck})$ ，腰の中心座標値を $\mathbf{Y}_{i,midhip} = (x_{i,midhip}, y_{i,midhip})$ とするとき，式 (3.3) のように座標を移動する．

$$g(\mathbf{Y}_i) = \frac{\mathbf{Y}_i}{|\mathbf{Y}_{i,neck} - \mathbf{Y}_{i,midhip}|} \quad (3.3)$$

3.4 基準時系列データ作成

3.3 節で示した手法により補正された動画内の全ダンサーの時系列データから，2.6.2 節の Soft-DTW による平均時系列データ作成手法を用いて評価の基準となるモデルを作成する．

両手首, 両足首の4部位独立して基準モデルを作成する. なぜならば, 部位によって動作のスピードや複雑度が変化するため, 適切な Soft-DTW の平滑化パラメータ γ が異なる. よって, 部位ごとに基準時系列データを作成する必要がある.

時系列データの変化量は動作のスピードと方向を表し, ダンスの評価では動作のスピードと方向も重要な要素である. そのため, 2.4 節で示した通常の DTW による対応付けより 2.5 節で示した AMSS による対応付けの方が, 本研究では適している. 2.5 節の手法は類似度の合計が最大になるように対応付けをするが, 対応付けの数が多くなれば類似度の合計が大きくなる. そのため, 対応付けが不適切な場合がある. そこで, 累積類似度ではなく平均類似度を用いた対応付けを行う Average Similarity DTW (以下, ASDTW) を提案する. 平均類似度とは, 累積類似度をそれまでの累積対応付け数で割ったものである. これにより, 対応付けの数に左右されることなく適切な対応付けが可能となる.

多次元データ系列 $A = \{a_1, a_2, \dots, a_M\}$, $B = \{b_1, b_2, \dots, b_N\}$ から, 式 (2.5) に従ってベクトル系列 $A' = \{\vec{a}'_1, \vec{a}'_2, \dots, \vec{a}'_{M-1}\}$, $B' = \{\vec{b}'_1, \vec{b}'_2, \dots, \vec{b}'_{N-1}\}$ を作成し, ASDTW による対応付けをする. A' と B' の合計類似度 $ASDTW(A', B')$ を式 (3.4) のように定義する.

$$ASDTW(A', B') = asd(M-1, N-1) \quad (3.4)$$

時系列データ間の類似度関数を Sim と表し, 一方の時系列データの要素 a_m ともう一方の時系列データの要素 b_n の類似度を $Sim(a_m, b_n)$ とするとき, $asd(m, n)$ と累積対応付け数 $L(m, n)$ を式 (3.5) のように再帰的に算出する.

$$\begin{aligned} asd(m, n) &= Sim(\vec{a}'_m, \vec{b}'_n) + asd(m+u, n+v) \\ L(m, n) &= L(m+u, n+v) + 1 \end{aligned} \quad (3.5)$$

なおこのとき, (u, v) を式 (3.6) に, S_1, S_2, S_3 を式 (3.7) に, 条件式を式 (3.8) に示す. ここで, (u, v) は u と v の値の組み合わせを表す.

$$(u, v) = \begin{cases} (-1, 0) & S_1 > \max\{S_2, S_3\} \\ (0, -1) & S_2 > S_3 \\ (-1, -1) & otherwise \end{cases} \quad (3.6)$$

$$\begin{aligned}
S_1 &= \frac{\text{Sim}(\vec{a}'_m, \vec{b}'_n) + \text{asd}(m-1, n)}{L(m-1, n) + 1} \\
S_2 &= \frac{\text{Sim}(\vec{a}'_m, \vec{b}'_n) + \text{asd}(m, n-1)}{L(m, n-1) + 1} \\
S_3 &= \frac{\text{Sim}(\vec{a}'_m, \vec{b}'_n) + \text{asd}(m-1, n-1)}{L(m-1, n-1) + 1}
\end{aligned} \tag{3.7}$$

$$\begin{aligned}
\text{asd}(0,0) = 0, \quad \text{asd}(m,0) = \text{asd}(0,n) = \infty \\
1 \leq m \leq M-1, \quad 1 \leq n \leq N-1
\end{aligned} \tag{3.8}$$

このときの類似度関数として、コサイン類似度を用いる。2 データの変化量ベクトルがともに零ベクトルであるとき、ともに動いていないため、類似度は1となる。2 データの変化量ベクトルの一方のみが零ベクトルであるとき、一方が動いているのに対してもう一方が動いていないという状態であるため、類似度を0とする。また、動作の方向が $\pi/2$ [rad]以上異なるとき、類似度を一律に0にする。2次元ベクトル \vec{a}, \vec{b} の類似度 $\text{Sim}(\vec{a}, \vec{b})$ を式(3.9)のように定義する。

$$\text{Sim}(\vec{a}, \vec{b}) = \begin{cases} 1 & \|\vec{a}\|_2 = 0 \text{ and } \|\vec{b}\|_2 = 0 \\ 0 & (\|\vec{a}\|_2 = 0 \text{ xor } \|\vec{b}\|_2 = 0) \text{ or } (\vec{a} \cdot \vec{b} < 0) \\ \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 \|\vec{b}\|_2} & \text{otherwise} \end{cases} \tag{3.9}$$

4 部位の x 座標と y 座標のそれぞれに対して、2.6.2 節の手法によって平均時系列データを作成する。このとき、動画内の全ダンサーの時系列データを入力時系列データとする。Soft-DTW の平滑化パラメータ γ を変化させて、各部位の x 座標と y 座標の時系列データで構成される時系列 2 次元データから適切な γ を探索する。作成した平均時系列データと各ダンサーを平均類似度による対応付けを行う。2 時系列データ間の時間方向の差が小さいとき、動作のタイミングが近いと評価できる。よって、全ダンサーの全対応付けの合計時間差が最小になる γ を適切な値とし、このときの平均時系列データを基準時系列データとする。ここ

で、 K は γ の探索範囲を表し、 $\gamma = 10^{-k}$ とし $0 \leq k \leq K, k \in \mathbb{N}$ の範囲で変化させる。検出ダンサー数を I 、動画の総フレーム数を T 、入力時系列データを Y としたとき、各部位の基準時系列2次元データ *Ref_data* の作成手法を Algorithm 1 に示す。このとき、`softdtw_barycenter`(Y, γ)を入力時系列データ Y 、平滑化パラメータ γ のときの Soft-DTW による平均時系列データ作成関数 (式 (2.14)) とする。このとき、動画内の各ダンサーの時系列データを等しく扱うため、各入力時系列データの重みである λ_i は一律になるように設定する。また、`ASDTW_path`(A, B)を2時系列データ A, B の ASDTW による対応付けを決定し、全対応付けが各行に格納された行列を作成する関数とする。ここで、2時系列データ A, B の `ASDTW_path`(A, B)によって行列 *path* を作成したとき、 m 番目の対応付けの A のフレームと B のフレームをそれぞれ `path[m][0]`および `path[m][1]`と表せるとする。このとき、 $Y[i][j]$ はダンサー i 、部位 j の x 座標と y 座標の時系列データで構成される時系列2次元データを示す。

Algorithm 1 基準時系列データ作成手法

Require: I, J, T, K, Y

Ave_data \leftarrow `zero_matrix`($2, T$), *Ref_data* \leftarrow `zero_matrix`($4, 2, T$)

for $j \leftarrow 1$ to 4 **do**

for $k \leftarrow 1$ to $K-1$ **do**

for $h \leftarrow 1$ to 2 **do**

Ave_data[h] \leftarrow `softdtw_barycenter`($Y[:, j][h], 10^{(-k)}$)

end for

sum $\leftarrow 0$, *min_sum* \leftarrow *infinity*

for $i \leftarrow 1$ to I **do**

path \leftarrow `ASDTW_path`(*Ave_data*, $Y[i][j]$)

delta \leftarrow `path`[:, 0] - `path`[:, 1]

sum \leftarrow *sum* + `sum`(*delta* * *delta*)

end for

if *sum* < *min_sum* **then**

min_sum \leftarrow *sum*

Ref_data[j][:] \leftarrow *Ave_data*[:]

end if

end for

end for

return *Ref_data*

3.5 動作変化点の検出

3.4 節で作成した 4 部位の x 座標および y 座標の時系列データで構成される基準時系列 2 次元データから、各部位の動作が変化するフレームを検出する。本研究では動作の方向が $\pi/4$ [rad]以上変化するフレーム、動作が開始するフレーム、動作が終了するフレームを動作変化点と定義する。時系列 2 次元データの変化量ベクトルは、動作のスピードと方向を表す。よって、基準時系列データの任意フレームにおいて、当該データと前データの変化量ベクトルと次データと当該データの変化量ベクトルの方向がの方向が $\pi/4$ [rad]以上異なるとき、つまり 2 ベクトルの内積が $\sqrt{2}/2$ 以下になるとき、当該フレームが動作変化点となる。また、当該データと前データの変化量ベクトルと次データと当該データの変化量ベクトルのどちらかのみが零ベクトルであるとき、動作開始または動作終了となる。そのため、このフレームが動作変化点となる。動作変化点の検出をすべての部位に対して行う。基準時系列 2 次元データを Ref_data , $dot(a,b)$ を a と b の内積, $norm(a)$ を a のノルムとし, 部位 j において検出した動作変化フレームを $mv_list[j]$ に格納するとしたときの動作変化点の検出手法を Algorithm 2 に示す。

Algorithm 2 動作変化点の検出手法

Require: Ref_data

```
 $mv\_list \leftarrow [ \text{for } j \leftarrow 1 \text{ to } 4 \text{ do } [ ] \text{ end for } ]$ 
for  $j \leftarrow 1$  to  $4$  do
   $R_j \leftarrow Ref\_data[j]$ ,  $M \leftarrow R_j.length$ 
   $R'_j \leftarrow zero\_matrix(M)$ ,
  for  $m \leftarrow 2$  to  $M$  do  $R'_j[m] \leftarrow R_j[m] - R_j[m-1]$  end for
  for  $m \leftarrow 2$  to  $M-1$  do
     $inner \leftarrow dot(R'_j[m], R'_j[m+1])$ 
    if  $norm(R'_j[m]) == 0$  and  $norm(R'_j[m+1]) == 0$  then
      continue
    else if  $norm(R'_j[m]) * norm(R'_j[m+1]) == 0$  or  $inner / (norm(R'_j[m]) * norm(R'_j[m+1])) \leq \sqrt{2}/2$  then
       $mv\_list[j].append(m)$ 
  end for
return  $mv\_list$ 
```

3.6 評価手法

同じ振り付けのダンスをしているダンスグループの評価は、グループ全体が正確に揃っているかどうかで評価される[6]。各ダンサーの動作の変化が周りのダンサーに比べて早いとダンス視聴者が感じる時、そのダンサーは早いと評価される。逆に、遅いとダンス視聴者が感じる時、そのダンサーは遅いと評価される。そして、早いと感じるときと遅いと感じるときが混在する場合、そのダンサーは、動画内に早いところがあり、かつ遅いところがあると評価される。全ダンサーの動作の方向が変化するフレーム、動作が開始するフレーム、動作が終了するフレームが一致しているとき、視聴者はダンスが揃っていると評価する。よって本研究では、各部位の動作変化点と対応付けされたダンサーのフレームが、動作変化点からどれだけずれているかによって、ダンサーの動作のタイミングを評価する。つまり、3.5節で検出した基準時系列データの動作変化点のフレームと各ダンサーの対応付けされたフレームの差を算出し、閾値 T を用いることでダンサーの動作が早い、正常、遅いか評価する。

まず、4部位において基準時系列データと各ダンサーの時系列データをASDTWによる対応付けを行う。このとき、動作変化点と対応付けされたダンサーのフレームがそのダンサーの動作変化点となり、フレーム差 Δt を算出する。フレーム差 Δt が正のとき基準よりも動作が遅いことを、負であるとき基準よりも早いことを示す。ここで、基準時系列データの動作変化点とダンサーの複数フレームと対応付けされたときは、動作変化点として最も適切なフレームを探索する。対応付けられたフレームの中で、最も動作の方向が変化したフレームをダンサーの動作変化点とする。つまり、動作変化点の候補フレームのなかで、前フレームの変化量ベクトルと次フレームの変化量ベクトルのコサイン類似度が最小になるフレームが動作変化点となる。閾値 T を設定し、平均値 Δt に $-T < \Delta t < T$ が成り立つとき、動作は正常であると判断する。また、 $\Delta t < -T$ が成り立つとき動作は早いと判断し、 $\Delta t > T$ が成り立つとき動作は遅いと判断する。動作の判断の概要を図3.3に示す。

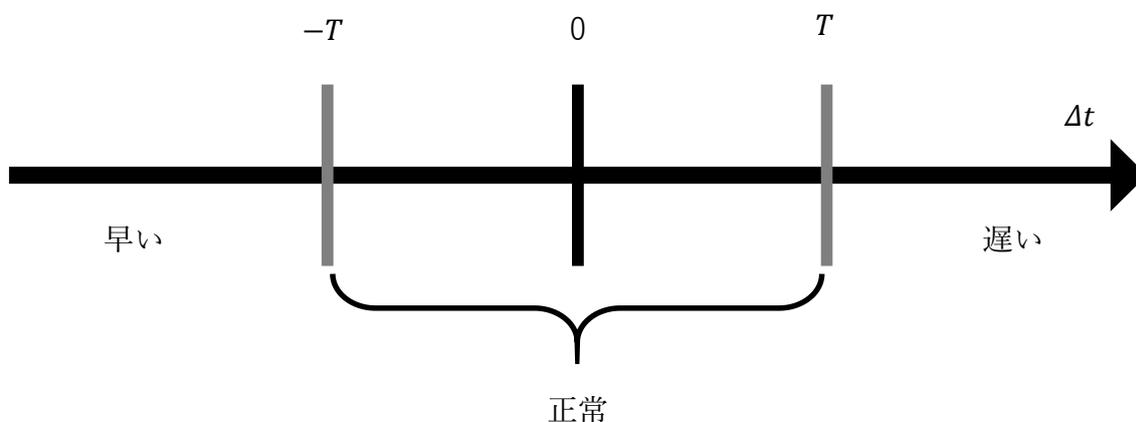


図 3.3 判断の概要

次に、部位ごとに早いと判断した数、遅いと判断した数を集計する。早いと判断した数の動作変化点数における割合が閾値 P を越える場合、当該部位は基準より早いと評価する。また、遅いと判断した数の動作変化点数における割合が閾値 P を越える場合、当該部位は基準より遅いと評価する。閾値 P を越えない場合、当該部位は基準に沿った動作をしていることになり、正常と評価する。4部位のうち1部位以上でも早い、または遅いと評価した場合、そのダンサーの総合評価を「早いところがある」、または「遅いところがある」とする。ここで、4部位の評価に早いと遅いが存在する場合、動画内で早いところと遅いところがあるとして、「早いところがあり、かつ遅いところがある」と評価する。逆に、4部位の評価がすべて正常である場合、そのダンサーは「正常」と評価する。これらにより、各ダンサーの動作がどれだけ評価基準に沿った動作をしているか評価できる。本研究では、3.4節と3.5節で述べたように、評価基準は動画内の全ダンサーの動作から作成している。そのため、本手法によって各ダンサーの動作が、周りのダンサーからずれていないかがわかる。

3.7 むすび

本章では、提案手法の具体的な内容について示した。動画内の全ダンサーから評価基準となるデータを作成し、基準と各ダンサーの評価手法について示した。

第4章 評価実験，結果および考察

4.1 まえがき

本章では，ダンス動画を用いて第 3 章で説明した提案手法の評価実験，結果および考察について述べる．本研究での実験環境を表 4.1 に示す．

表 4.1 実験環境

OS	Ubuntu 14.04 LTS
CPU	Intel® Core™ i7-2600K 3.4GHz 4Core
GPU	NVIDIA GeForce GTX 1070

4.2 実験概要

提案手法の評価実験として，動画内の全ダンサーが同じ振り付けのダンスをしている動画を対象に，3 種類の実験と主観評価との比較をする．まず，動画の OpenPose による部位推定を行い，3.3 節で示した手法により取得座標値の補正をする．実験 1 では，3.4 節の提案手法を用いて，4 部位の基準時系列データを作成する．実験 2 では，3.5 節の提案手法を用いて，基準時系列データから動作変化点を検出する．実験 3 では，3.6 節の提案手法を用いて，各ダンサーを評価する．また，ダンス視聴者にアンケートを実施し，提案手法による評価と主観評価を比較する．3 種類の実験は，3 本の 5 秒間の動画で行う．それぞれの動画のダンサー数 I ，動画のフレームレート，動画の総フレーム数 T を表 4.2 に示す．また，使用動画のダンスの様子として，動画開始 1 秒間の OpenPose 推定結果を図 4.1，図 4.2 および図 4.3 に示す．

表 4.2 使用動画の詳細

	動画 1	動画 2	動画 3
ダンサー数 I	5	7	4
フレームレート [fps]	24	30	24
フレーム数 T	120	150	120

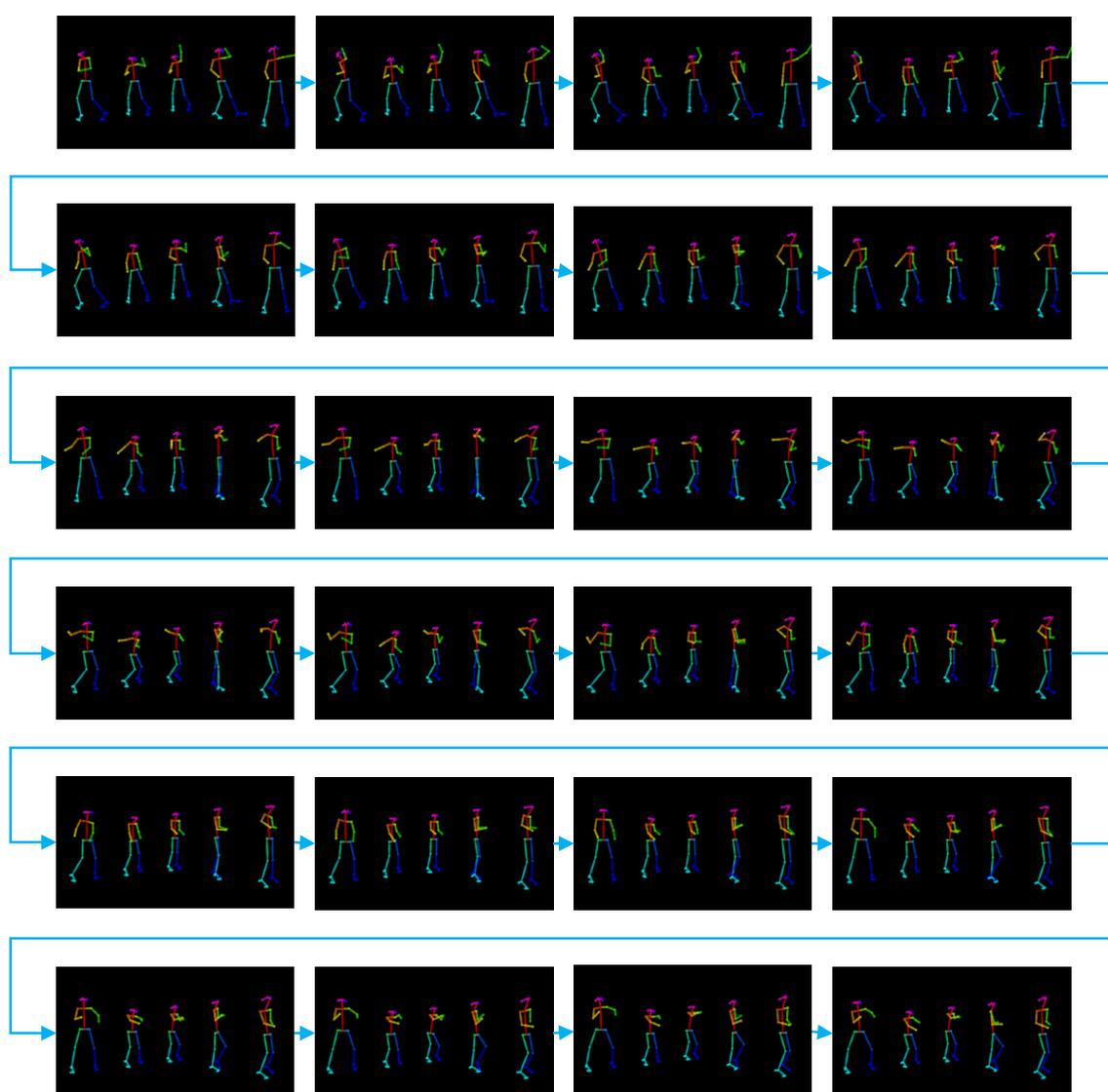


図 4.1 動画 1 の開始 1 秒間



図 4.2 動画 2 の開始 1 秒間



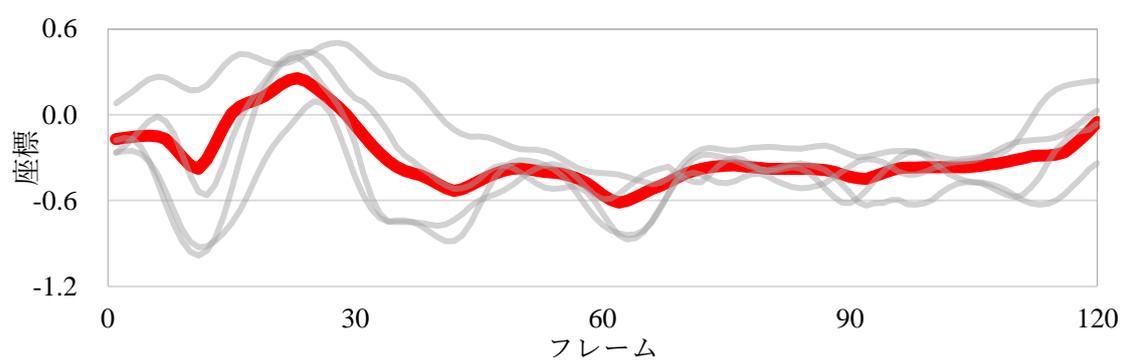
図 4.3 動画 3 の開始 1 秒間

4.3 実験 1

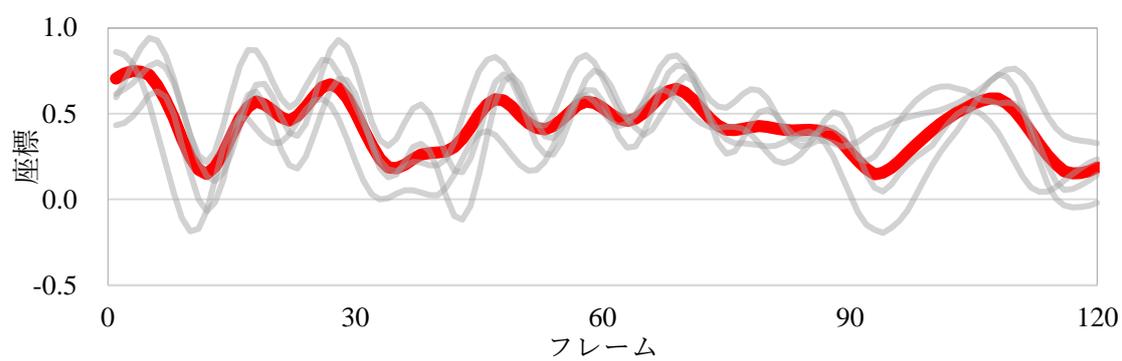
3.4 節で提案した手法により作成した基準時系列データを確認する。平滑化パラメータ γ を変化させ、適切な値を探索する。次に、作成した基準時系列データと全ダンサーの座標変化を比較し、基準データとして適正かどうか評価する。

本研究では、平滑化パラメータ γ の適正な値の探索範囲である K を、3本の動画のどちらに対しても6と設定した。両手首、両足首の4部位に対して、基準時系列データを作成した。作成した時系列データと全ダンサーの座標変化を比較する。ここでは、右手首と右足首につ

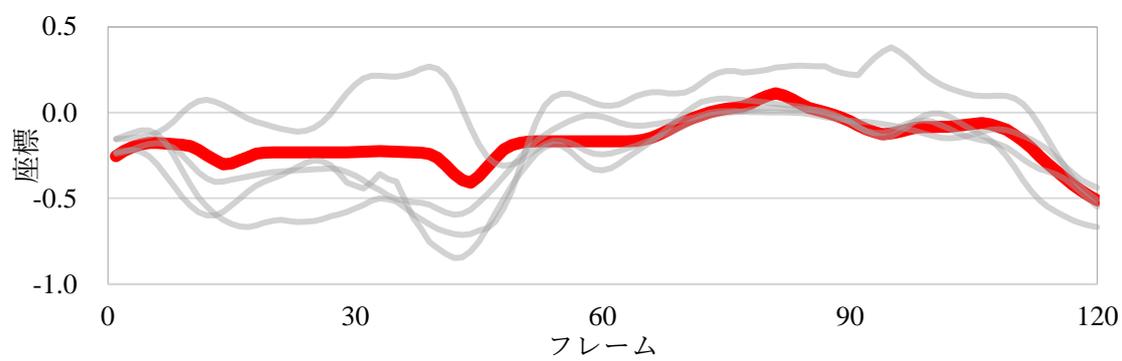
いてのみ結果を示す。動画 1, 動画 2 および動画 3 における基準時系列データとダンサーの座標変化をそれぞれ図 4.4, 図 4.5 および図 4.6 に示す。これらの図において, 赤色の線は基準時系列データ, 灰色の線は全ダンサーの時系列データを表す。



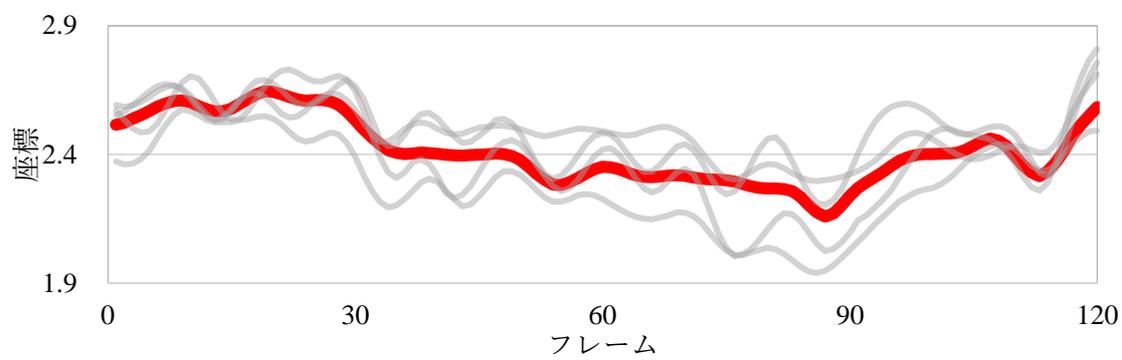
(a)右手首 x 座標



(b)右手首 y 座標



(c)右足首 x 座標



(d)右足首 y 座標

図 4.4 動画 1 の基準時系列データ

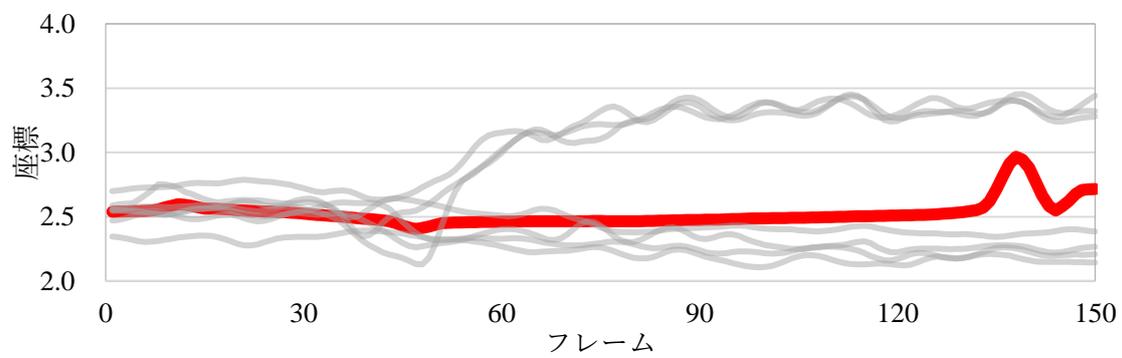
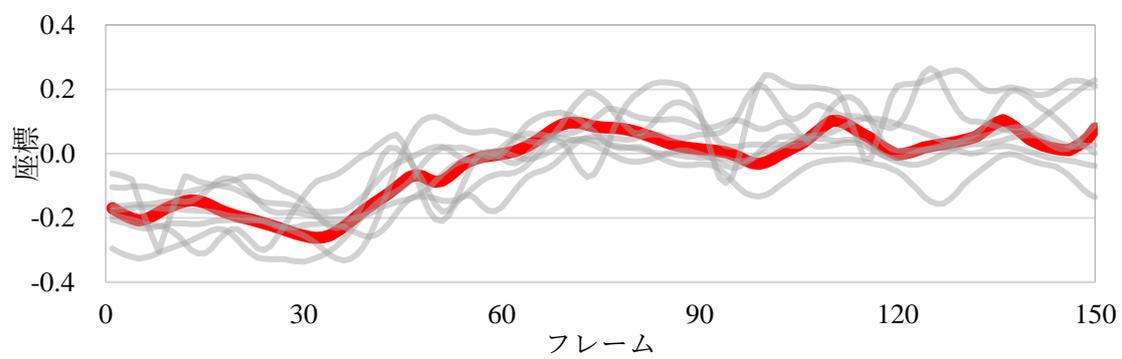
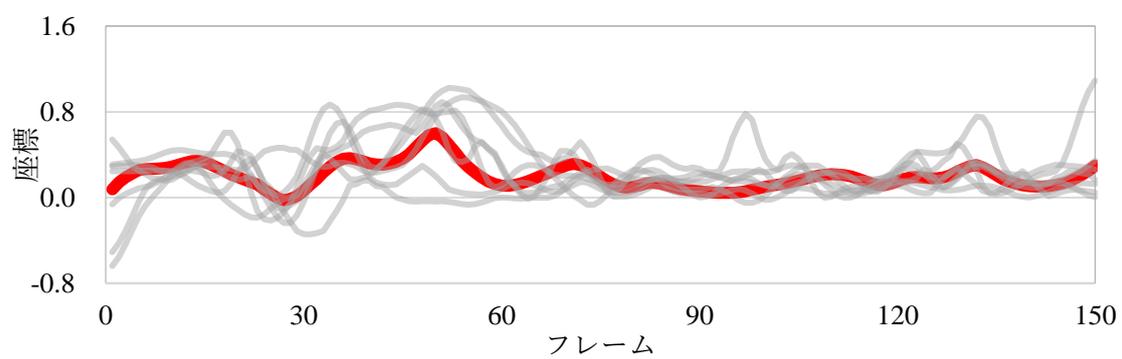
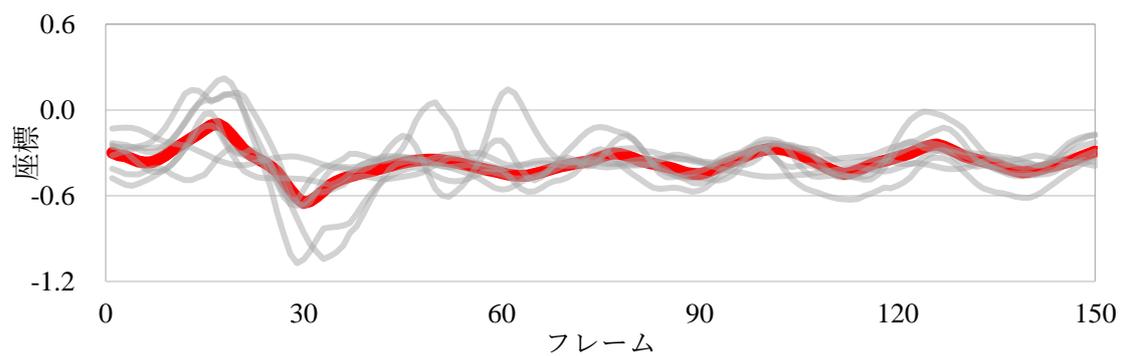
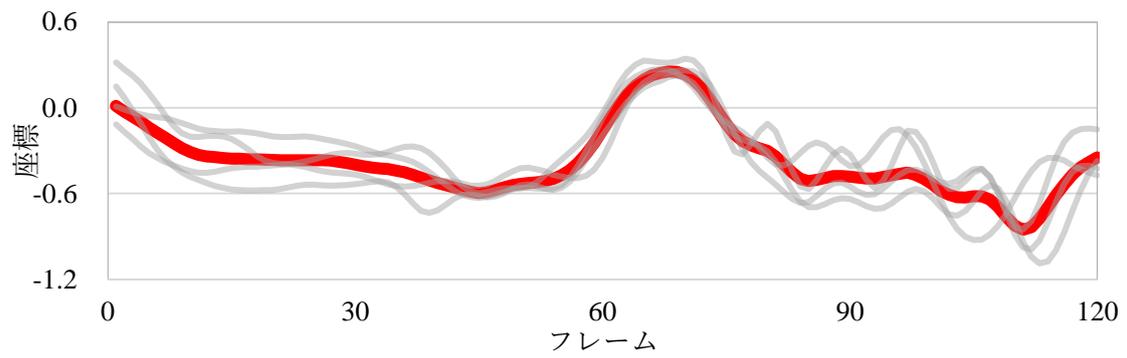
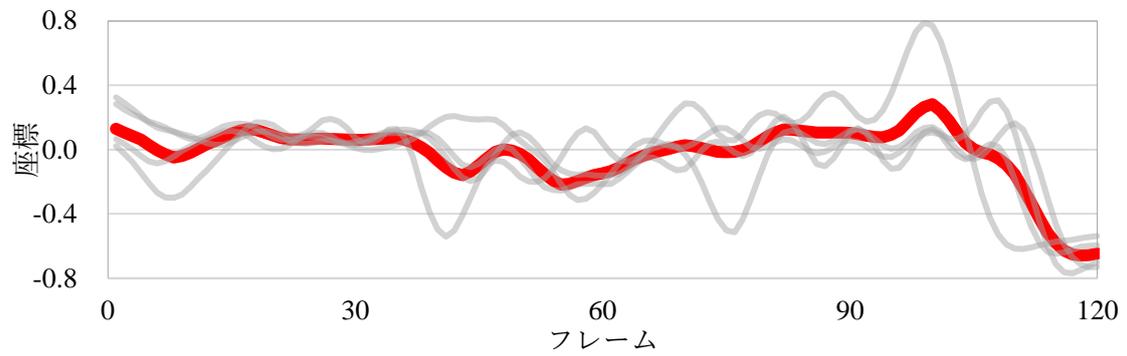


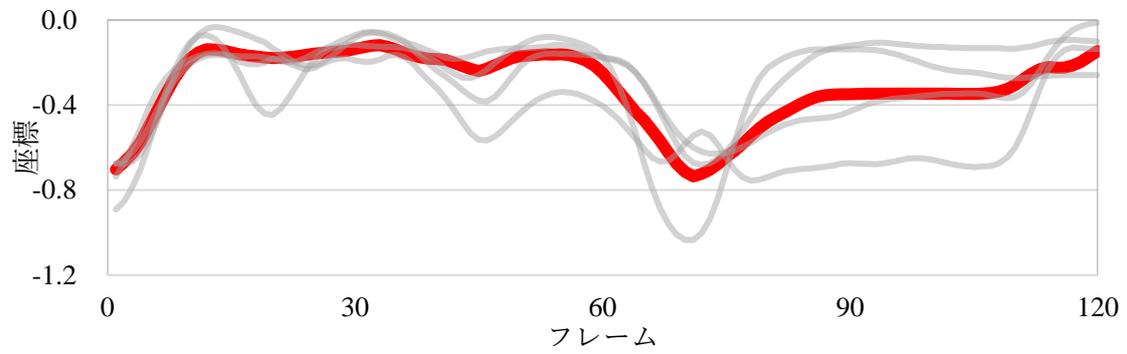
図 4.5 動画 2 の基準時系列データ



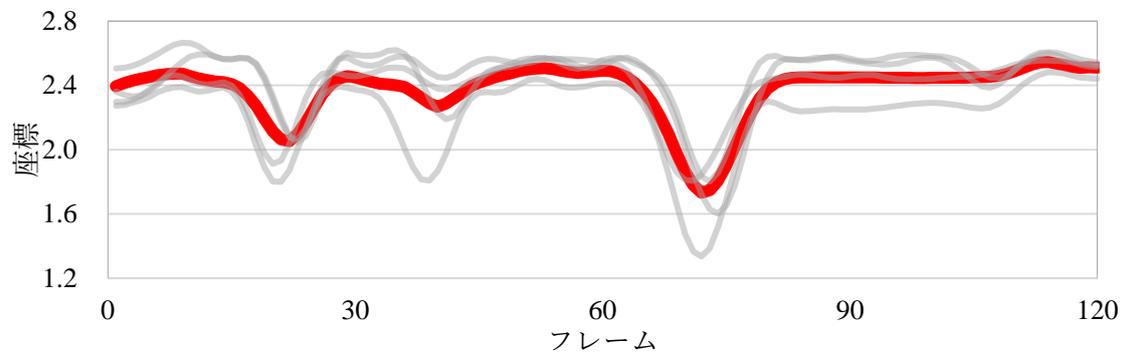
(a) 右手首 x 座標



(b) 右手首 y 座標



(c) 右足首 x 座標



(d) 右足首 y 座標

図 4.6 動画 3 の基準時系列データ

図 4.4, 図 4.5 および図 4.6 より, 基準時系列データが動画内の全ダンサーの座標変化から大きくずれていないことが確認できる. 特に, 図 4.4(b)や図 4.6(d)では, 基準時系列データが動画内の全ダンサーの座標変化と同様な変化をしていることがわかる. 図 4.4(b)や図 4.6(d)では, 全ダンサーの座標が同様な値であり, 座標変化が一様である. つまり, 全ダンサーの座標変化の極値が, 同じ時刻に集中している. そのため, ダンサーの座標変化の特徴が同様になり, 基準時系列データがそれらの特徴の多くを保持できるといえる. しかし, 図 4.4(d)の 40 フレーム付近, 70 フレーム付近や図 4.5(d)に注目を見ると, ダンサーの座標変化に対して基準時系列データがほぼ変化していないことが確認できる. 図 4.4(d)の 40 フレーム付近, 70 フレーム付近や図 4.5(d)では, ダンサーの座標が分散しており, 座標変化にまとまりがない. つまり, 全ダンサーの座標変化の極値が様々な時刻に散在している. そのため, ダンサー同士の座標変化の特徴を相殺してしまい, 基準時系列データがダンサーの座標変化の特徴を保持できていないことがわかる.

以上のことから, 提案手法を用いることで, 動画内の全ダンサーの動作にまとまりがある場合には, 適切な基準データを作成することができる. しかし, ダンサーによって踊りのタイミングが極端にバラバラになってしまうと適切な基準時系列データの作成が難しくなる.

4.4 実験 2

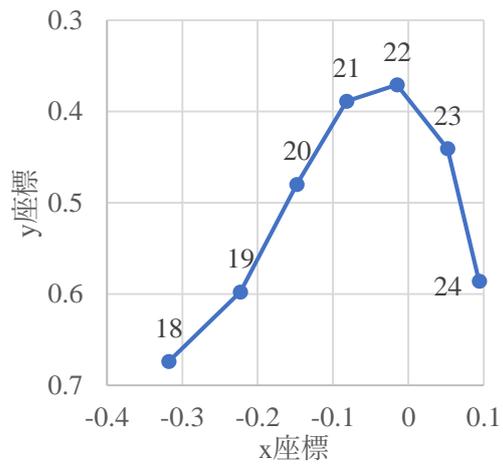
3.5 節で提案した手法により実験 1 で作成した基準時系列データから動作変化点を検出する. 基準時系列データの任意フレームにおいて, 当該データと前データの変化量ベクトルと次データと当該データの変化量ベクトルの内積が $\sqrt{2}/2$ 以下になるとき, および当該データと前データの変化量ベクトルと次データと当該データの変化量ベクトルのどちらかのみが零ベクトルであるときを, 動作変化点として検出する. これらの動作変化点と動画内の全ダンサーの動作変化フレームを比較して, 評価基準として適正か確認する. まず, 動画 1, 動画 2 および動画 3 の右手首, 左手首, 右足首, 左足首の動作変化点数を表 4.3 に示す.

表 4.3 動作変化点数

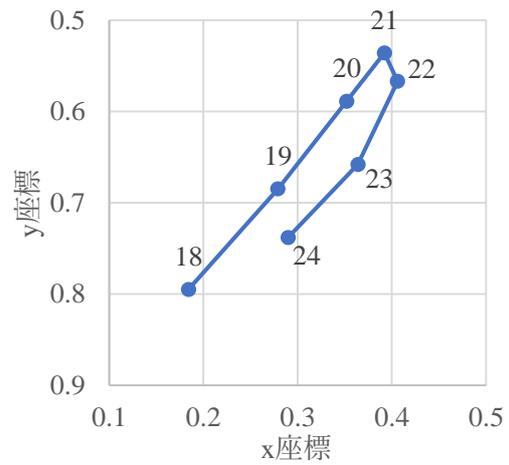
	右手首	左手首	右足首	左足首
動画 1	16	11	12	16
動画 2	16	11	13	14
動画 3	13	14	13	13

各部位の動作変化点の前後 3 フレームの各ダンサーの当該部位の座標変化を示す。ダンサーの座標変化から、そのダンサーの動作が変化しているフレームが推測できる。動作変化点と全ダンサーの動作変化推測フレームを比較する。ここでは、結果の一例を示す。

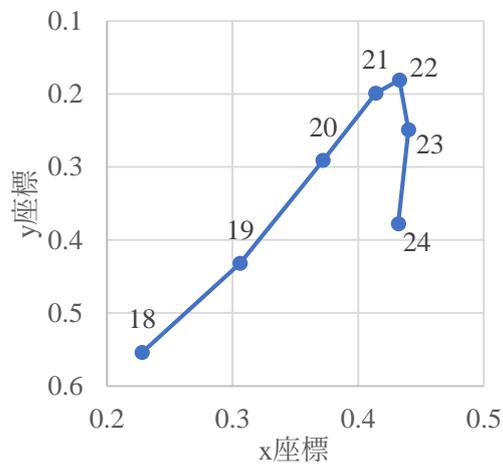
動画 1 の右手首では、21 フレームが動作変化点として検出された。この動作変化点の前後 3 フレーム、つまり 18 フレームから 24 フレームの各ダンサーの右手首の座標変化を図 4.7 に示す。また、動画 3 の右足首では、71 フレームが動作変化点として検出された。この動作変化点の前後 3 フレーム、つまり 68 フレームから 74 フレームの各ダンサーの右足首の座標変化を図 4.8 に示す。これらの図において、プロット付近の数字はフレームである。なお、画像中の座標軸と同様になるように、y 軸を反転している。



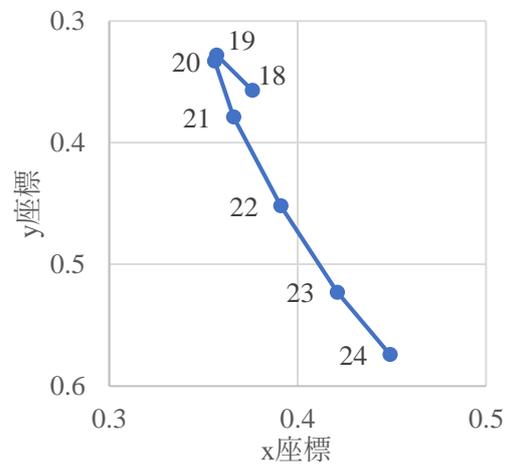
(a)ダンサー1



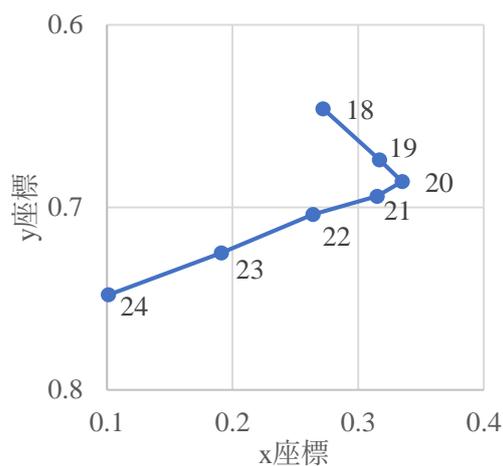
(b)ダンサー2



(c)ダンサー3



(d)ダンサー4



(e)ダンサー5

図 4.7 動画1の右手首の座標変化 (18 フレーム~24 フレーム)

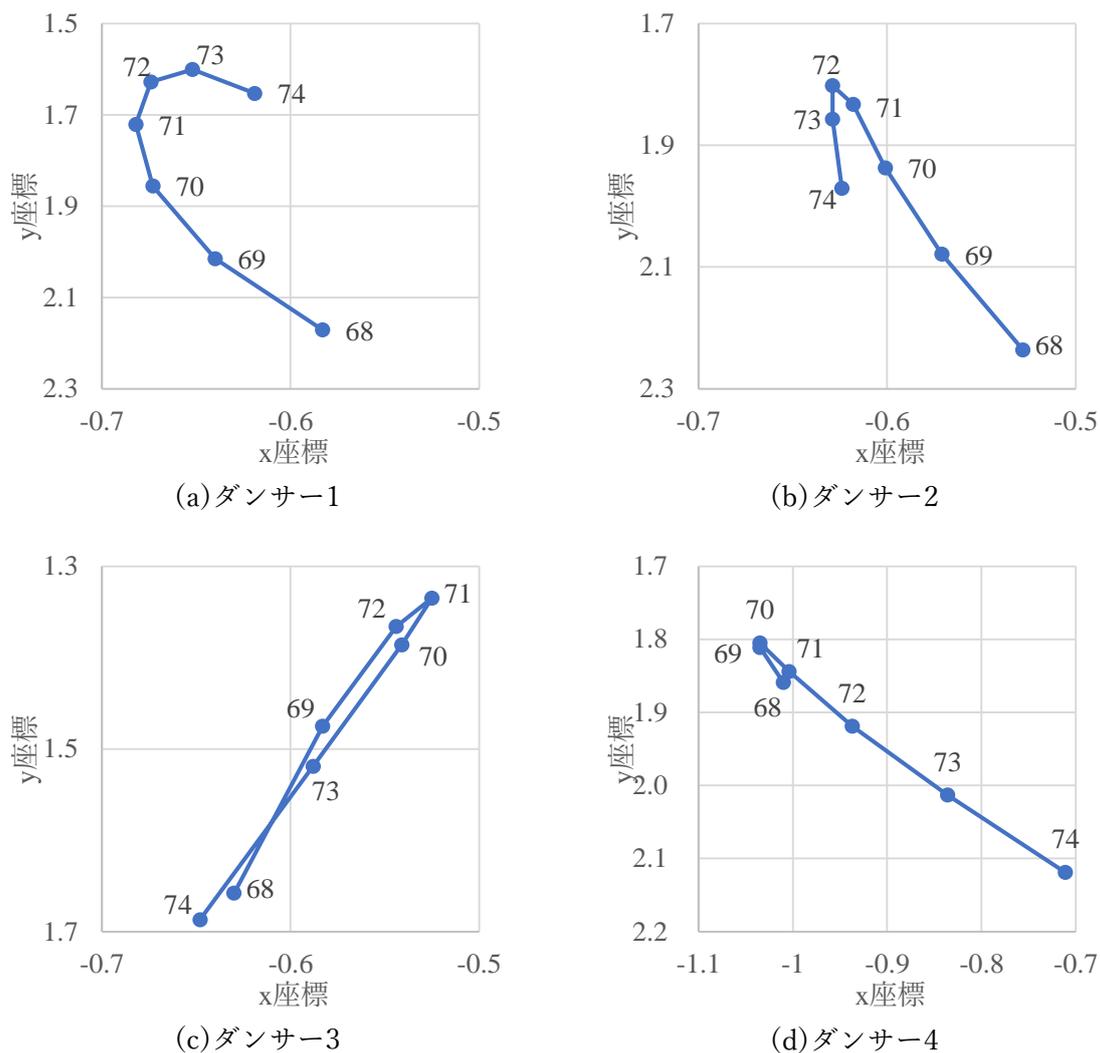


図 4.8 動画3の右足首の座標変化 (68 フレーム~74 フレーム)

図 4.7 から、ダンサー1 は 22 フレームで、ダンサー2 は 21 フレームで、ダンサー3 は 22 フレームで、ダンサー4 は 20 フレームで、ダンサー5 は 20 フレームで動作が変化していると推測できる。このとき、全ダンサーの動作変化フレームの平均は 21 である。よって、21 フレームが動作変化点として適切であると確認できる。図 4.8 から、ダンサー1 は 72 フレーム付近で、ダンサー2 は 72 フレームで、ダンサー3 は 71 フレームで、ダンサー4 は 70 フレーム付近で動作が変化していると推測できる。このとき、全ダンサーの動作変化フレームの平均は 71.25 である。よって、71 フレームが動作変化点として適切であると確認できる。以上のことから、提案手法によって検出した動作変化点が、ダンサーの動作変化フレームに近いことが確認できた。そのため、検出した動作変化点が、ダンサーの動作を評価する時刻として適切であることがいえる。

4.5 実験3

4.5.1 提案手法による評価

3.6 節で提案手法により，各動画のダンサーの動きを評価する．人間の視覚の時間分解能は 0.05 秒から 0.1 秒の間である [7]．そのため，評価基準との動作の時間差が時間分解能を越える場合，ダンス視聴者が動作のずれを認知できるといえる．よって，本実験では動画 1，動画 2 および動画 3 の閾値 T を分解能の範囲になるようにする．フレームレートが 24fps であるとき，1 フレーム間が約 0.042 秒となり，フレームレートが 30fps であるとき，1 フレーム間が約 0.033 秒となる．このことより，各動画の閾値 T を表 4.4 のように設定した．

表 4.4 使用動画の閾値 T

	動画 1	動画 2	動画 3
T [フレーム]	2	3	2

また，ダンサーの動作評価の閾値 P を $1/3$ と設定した．動画 1，動画 2 および動画 3 において，各ダンサーの各部位の動作が早い，または遅いと判断した数をそれぞれに表 4.5，表 4.6 および表 4.7 に示す．次に，表 4.3 と閾値 P からダンサーの各部位の動作が，基準より「早い」，「正常」，「遅い」の評価できる．この結果から，各ダンサーの総合評価が，「早いところがある」，「正常」，「遅いところがある」，「早いところがあり，かつ遅いところがある」の中から決定する．この各ダンサーの総合評価を表 4.8 に示す．このとき，各動画内の一番左に存在するダンサーをダンサー 1 とし，左から順にダンサー番号を付けた．

表 4.5 各ダンサーの判断数（動画1）

		右手首	左手首	右足首	左足首
ダンサー1	早い	1	1	2	0
	遅い	1	0	1	5
ダンサー2	早い	2	1	3	3
	遅い	0	0	0	0
ダンサー3	早い	2	2	1	1
	遅い	1	0	3	4
ダンサー4	早い	2	2	2	6
	遅い	0	4	4	1
ダンサー5	早い	5	5	4	1
	遅い	0	0	0	6

表 4.6 各ダンサーの判断数（動画2）

		右手首	左手首	右足首	左足首
ダンサー1	早い	3	3	4	1
	遅い	7	0	4	2
ダンサー2	早い	3	3	3	0
	遅い	2	2	5	7
ダンサー3	早い	1	2	0	2
	遅い	6	2	3	8
ダンサー4	早い	1	0	8	0
	遅い	2	4	3	3
ダンサー5	早い	2	3	3	2
	遅い	0	1	4	6
ダンサー6	早い	5	2	1	2
	遅い	2	1	6	3
ダンサー7	早い	6	3	6	10
	遅い	1	1	2	1

表 4.7 各ダンサーの判断数（動画3）

		右手首	左手首	右足首	左足首
ダンサー1	早い	3	0	1	3
	遅い	0	8	1	2
ダンサー2	早い	2	0	0	0
	遅い	3	0	3	3
ダンサー3	早い	1	1	4	3
	遅い	4	0	0	2
ダンサー4	早い	2	5	5	2
	遅い	0	0	0	1

表 4.8 ダンサーの総合評価

	動画1	動画2	動画3
ダンサー1	正常	遅いところがある	遅いところがある
ダンサー2	正常	遅いところがある	正常
ダンサー3	正常	遅いところがある	正常
ダンサー4	早いところがあり、 かつ遅いところがある	早いところがある	早いところがある
ダンサー5	早いところがあり、 かつ遅いところがある	遅いところがある	(計測データなし)
ダンサー6	(計測データなし)	遅いところがある	(計測データなし)
ダンサー7	(計測データなし)	早いところがある	(計測データなし)

提案手法によるダンサーの評価は表 4.8 のようになった。この手法による結果を評価するために、4.5.2 節でダンス視聴者による主観評価と比較を行う。

4.5.2 主観評価との比較

ダンス視聴者にアンケートを実施し、視聴者の主観による評価を確認する。各動画に対して、ダンス視聴者は、各ダンサーの動作が周りに比べて早いと感じる瞬間の有無と遅いと感じる瞬間の有無を回答する。本研究では、6 人にアンケートを実施した。動画 1、動画 2、および動画 3 の各ダンサーが周りのダンサーに比べて、早いと感じると回答した人数と遅いと回答した人数を表 4.9 に示す。

表 4.9 アンケート結果

		動画 1	動画 2	動画 3
ダンサー 1	早い	0	0	0
	遅い	2	2	0
ダンサー 2	早い	1	2	0
	遅い	0	0	6
ダンサー 3	早い	0	0	0
	遅い	2	0	0
ダンサー 4	早い	5	3	6
	遅い	0	1	0
ダンサー 5	早い	2	0	(計測データなし)
	遅い	0	1	(計測データなし)
ダンサー 6	早い	(計測データなし)	1	(計測データなし)
	遅い	(計測データなし)	0	(計測データなし)
ダンサー 7	早い	(計測データなし)	0	(計測データなし)
	遅い	(計測データなし)	3	(計測データなし)

動画 1 のダンサー 4 とダンサー 5 の提案手法による評価は、どちらも早いところがあり、かつ遅いところがあるとなった。しかし、アンケートではダンサー 4 に対して早いと感じる回答者は多かったに対して、遅いと感じると回答した人はいなかった。また、ダンサー 5 に

は早いと感じる人はアンケート回答者の全体の 1/3 しかいなかった。動画 2 では、提案手法によるダンサー評価は、全ダンサーが早いところ、または遅いところがあった。これに対して、アンケートでは各ダンサーに対して多くてもアンケート回答者の半数以下しか早い、または遅いと回答されなかった。また、全ダンサーに違いは見られなかったとアンケートに回答していた回答者がいた。これは、動画 2 のダンスが小さくて動作の違いが感じられなかったからであると考えられる。動画 3 では、ダンサー 4 に対しては提案手法による評価とアンケート結果が一致していた。しかし、提案手法ではダンサー 1 が早いところがあると評価していたが、アンケートではダンサー 2 が早いと回答されていた。提案手法による評価と主観評価が一致する場合と相違がある場合があり、人間が主観的に感じられた動作が早いダンサーと遅いダンサーの検出だけでなく、主観的に感じられない早いところや遅いところの検出ができたといえる。しかし、アンケートでは全員が遅いところがあると回答した動画 3 のダンサー 2 が、提案手法では正常と評価された。そのため、評価手法内の閾値 T 、閾値 P の設定を変える必要がある。

4.6 むすび

本章では、ダンス動画を用いて第 3 章で説明した提案手法の評価実験、結果および考察について述べた。また、ダンス視聴者に対してアンケートを実施し、提案手法による結果と比較した。

第5章 結論

5.1 結論

本研究では、複数ダンサーが同じ振り付けのダンスをする動画を対象とし、ダンサーの姿勢情報から作成した基準データをもとに、各ダンサーの動作を評価する手法を提案した。また、2時系列データの類似度を求める手法として、平均類似度を用いて対応付けをする手法を提案した。提案手法による作成した基準データについて、全ダンサーの座標変化と比べたとき、全ダンサーの座標変化が一様である場合に評価基準として適切である確認した。また、提案手法によって基準データから検出した動作変化点が、動画内の全ダンサーの動作が変化するフレームに大きく差がなく、適切であると確認できた。次に、提案手法によるダンサーの動作評価とアンケートによる主観評価の比較を行った。これにより、ダンス視聴者が感じる動作が早いダンサーと遅いダンサーの検出だけでなく、主観的に感じられない早いダンサーと遅いダンサーの検出ができた。

5.2 今後の課題

提案手法によって複数時系列データから基準データを作成する場合に、全ダンサーの動作のタイミングが不揃いの時は適切な基準データが作成できない問題点がある。そのため、基準データを作成する手法の改良が必要である。提案手法では正常と評価されたダンサーが、アンケートでは遅いと感じると回答されていた。したがって、動作が早い、正常、遅いの評価を決定する手法に改良の余地がある。

謝辞

本研究の実験環境を与えてくださり、研究の方向性等の丁寧かつ熱心なご指導を頂いた渡辺裕教授に心から感謝いたします。

本研究のきっかけを与えてくださり、様々なご提案を頂きました早稲田大学国際情報通信センターの石川孝明様に心から深謝いたします。

また、日頃から御意見やアドバイスをくださった笠井裕之教授、研究室の皆様に御礼申し上げます。

特に同僚である松田卓也氏、赤塚紘輝氏、梅田聖也氏には、日々の会話やゲームなどを興じ、様々な部分で研究の支えとなったことを感謝いたします。

最後に、私が学業に専念できる環境を与え、ここまで育ててくださった家族に深謝いたします。

参考文献

- [1] 文部科学省 (2011), “新学習指導要領に基づく中学校向け「ダンス」リーフレット,” http://www.mext.go.jp/a_menu/sports/jyujitsu/1306098.htm, (2020年1月18日現在)
- [2] 文部科学省 (2017), “中学校学習指導要領 (平成 29 年告示) 解説保健体育編,” https://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2019/03/18/1387018_008.pdf, (2020年1月18日現在)
- [3] Z. Cao, T. Simon, S-E Wei, Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), No.121, pp.1302-1310, Jul. 2017.
- [4] 中村, 滝, 野宮, 上原, “AMSS:時系列データの効率的な類似度測定手法,” 電子情報通信学会論文誌.D, 情報・システム, J91-D(11), pp.2579-2588. Nov. 2018.
- [5] M. Cuturi, M. Blondel, “Soft-DTW: a Differentiable Loss Function for Time-Series,” The International Conference on Machine Learning (ICML), Vol. 70, pp.894-903, 2017
- [6] ダンス部専門フリーマガジン&ウェブ ダンスク! (2017-12), “「日本ダンス大会」審査委員長 TAKAHIRO がココだけで明かす〜ダンス部大会必勝のポイント 4 つ!,” <https://d-s-k.jp/interview/30671/>, (2020年1月18日現在)
- [7] 鬼沢, “残像に関する研究(1): 残像研究の展望,”アルテスリベラル第 6 号, pp.9-43, Oct. 1969

図一覧

図 2.1	OpenPose の検出結果	3
図 2.2	時間・空間方向に伸縮された時系列データに対する DTW 例.....	5
図 2.3	時系列データからベクトル系列への変換	6
図 2.4	Soft-DTW の概要	8
図 2.5	Soft-DTW による平均時系列データの作成.....	9
図 3.1	提案手法の概要.....	11
図 3.2	座標値のフレーム補正の概要.....	12
図 3.3	判断の概要.....	18
図 4.1	動画 1 の開始 1 秒間.....	21
図 4.2	動画 2 の開始 1 秒間.....	22
図 4.3	動画 3 の開始 1 秒間.....	23
図 4.4	動画 1 の基準時系列データ	25
図 4.5	動画 2 の基準時系列データ	26
図 4.6	動画 3 の基準時系列データ	27
図 4.7	動画 1 の右手首の座標変化 (18 フレーム~24 フレーム)	30
図 4.8	動画 3 の右足首の座標変化 (68 フレーム~74 フレーム)	31

表一覧

表 3.1	番号と部位名の対応	13
表 4.1	実験環境	20
表 4.2	使用動画の詳細.....	21
表 4.3	動作変化点数	28
表 4.4	使用動画の閾値 T	32
表 4.5	各ダンサーの判断数（動画 1）	33
表 4.6	各ダンサーの判断数（動画 2）	33
表 4.7	各ダンサーの判断数（動画 3）	34
表 4.8	ダンサーの総合評価	34
表 4.9	アンケート結果.....	35

研究業績

- [1] 稲田, 大澤, 山川: “OpenPose”, 動画情報メディア学会誌「私の研究ツールー第 107 回ー」, Vol.73 No.5, pp.932-934, Sep. 2019
- [2] 稲田, 石川, 渡辺: “機械学習を用いた複数ダンサーの 3 次元基準モデル作成”, 電子情報通信学会総合大会, D-12-60, Mar. 2019
- [3] 稲田, 石川, 渡辺: “OpenPose を用いた複数ダンサーの動作の同期および姿勢評価”, 2018 年映像情報メディア学会冬季大会, 13D-2, Dec. 2018
- [4] 稲田, 石川, 渡辺: “OpenPose を用いた複数ダンサーの姿勢評価の基礎検討”, 映像情報メディア学会スポーツ情報処理時限研究会 (メディア工学研究会共催) ,ME2018-109 (SIP2018-03-06), pp.21-24, Dec. 2018
- [5] 稲田, 石川, 渡辺: “OpenPose を用いた複数ダンサーの動作のずれ評価の基礎検討”, 2018 年画像符号化シンポジウム・2018 年映像メディア処理シンポジウム (PCSJ/IMPS2018), P-5-7, Nov. 2018
- [6] 稲田, 石川, 渡辺: “OpenPose を用いた複数人のダンスの一致度評価”, 情報処理学会全国大会, 2U-01, Mar. 2018