

「踊ってみた動画」を対象としたダンスのキレ採点システムの開発

関西学院大学大学院 理工学研究科 人間システム工学専攻 片寄研究室 47021828 森鈴果

1 はじめに

動画共有サイトの1つである「YouTube」の人気ジャンルとして「踊ってみた動画」がある。視聴者によっては「キレッキレ」等のコメントが寄せられることもあり、単なる鑑賞に留まらず、視聴者からダンスの評価を受けられることも「踊ってみた動画」が人気となった1つの要因であると言える。このような背景から、筆者はカラオケ採点システムの精密採点機能のようにダンスの採点システムを提供することで、「踊ってみた動画」に新たな楽しみ方を提供できるのではないかと考えている。

筆者らは、これまで「恋ダンス」とダンスの「キレ」を採点対象としたダンス採点システムを開発してきた[1]。しかし、このシステムでは、ダンス全体に対して採点を行っているため、総体的な点数しか表示することができない。そこで本研究では、短期間の区間によってダンスを採点することを目標とする。

2 ダンスの評価

ダンスにおいて、計算機による評価はこれまでにいくつか試みられてきたが、キレに着目した研究は見られない。そこで、本節では計算機によるダンスの評価を行った従来研究を紹介し、ダンスのキレの定義について触れる。

2.1 ダンスの評価に関する研究

ダンスの質が評価された研究として、古市らの研究がある[2]。この研究では、骨格情報のみでヒップホップダンスの「個性」を識別できることが明らかになっている。また Niewiadomski らは、ダンスの「軽さ」と「脆さ」の評価を行っている[3]。このように、骨格情報を利用することで計算機によるダンスの評価を可能としている。

2.2 ダンスのキレ

ダンススクール TOKYO STEPS ARTS[4]によると、ダンスのキレは以下の視点に基づいて評価できるとされる。

1. 動きをびたっと止めることを意識する
2. 指先や足先まで気を配る
3. 初速を意識して動きに緩急をつける
4. 大きく躍動的な動きで踊る

「踊ってみた動画」から指先や足先の身体の情報は捉えにくいことから(2)指先や足先まで気を配ることを評価することは難しく、(4)大きく躍動的な動きで踊ることは、振付による影響を大きく受けてしまう。そのため、「素早い動作をびたっと止める」ことに着目することで、計算機によるキレ採点が可能になると考えられる。

3 アプローチ

本節では短期間でのキレを対象とした採点システムの構成方法について検討する。筆者ら[1]やダンスの感情分類を行った Wang らの研究[5]では、CNN と LSTM を組み合わせることで、総体的なキレの認識精度が向上している。そこで本節では、短期的な区間によるキレの認識精度の向上に向けて(3.1)学習モデルの設計、(3.2)キレ情報の付与、(3.3)データセットの作成を示す。

3.1 学習モデルの設計

「踊ってみた動画」には、背景や服装などダンス動作以外の情報が含まれている。そこで、姿勢推定技術によって骨格情報を取得し、動画からダンス動作のみを抽出する。筆者は、ダンス動作

の抽出に2次元姿勢推定技術を利用してきた[1]が、ダンスは3次元空間で踊られる。そのため、3次元姿勢推定技術の利用を検討する。

人間のモーションデータとしては、関節の位置座標や角度がよく用いられる。ダンスジャンル分類[6]では、この関節角度に対しフレーム間差分を利用することで、骨格情報の速度変化を取得していた。そこで、フレーム間差分した関節の位置座標・角度による骨格情報の利用を検討する。

3.2 キレ情報の付与

筆者はこれまで、動画に付されたコメントに基づいてキレ情報付与する手法をとってきた[1]。しかし、この方法では総体的なラベルの付与となってしまい、キレに関係しない動作についてもキレがあるという情報を付与している可能性がある。そのため、短期的なキレ採点においては、専門家によるキレ情報の付与が必要であると考えられる。

3.3 データセットの作成

ダンスは時系列パフォーマンスであり、キレにおいて「素早い動作をびたりと止めている」瞬間だけでなく、止めの動作までの素早さ、また止めの動作からの素早さも重要である。そこで、学習モデルに入力するデータについて、時系列を考慮する。

4 実装

本節では、アプローチに基づき、短期間の区間によるダンス動作を対象としたキレ採点システムを実装について述べる。

4.1 モデル

3次元姿勢推定技術を利用する動画は、人物が複数存在しないという条件がある。「踊ってみた動画」はアイドル等のダンスを模倣したものが多く存在することから、動画内に複数人含まれる場合がある。そのため、YOLO-V5 と Byte-Track を利用し人物検出した結果を用いて Detectron2 と VideoPose3D による3次元姿勢推定を行った。

VideoPose3D によって取得する骨格情報は3次元の位置座標である。そのため、位置座標から関節角度の計算を行う。関節角度の3次元空間による情報を維持するため、関節角度をある関節と隣接する2つの関節の3点間によるなす角の $xy \cdot yz \cdot xz$ 方向と xyz 方向の4方向から計算する。この位置座標と関節角度に対し、フレーム間差分を行った。

4.2 データセット

本実験で対象とする振付は、「踊ってみた動画」として多くの人がアップロードしている BTS の「Dynamite」とする。初めに、YouTube から「Dynamite 踊ってみた」を検索キーワードとして関連度順上位に存在し、キレのあるダンスを踊る人が含まれる動画8本、キレのないダンスを踊る人が含まれる動画6本、どちらも含まれる動画1本を取得した。その後、人物検出によって、キレのあるダンスを踊る人24人、キレのないダンスを踊る人21人を取得した。

人物検出によって取得したキレのあるダンスを踊る人に対し、10年以上のダンス経験を持つ筆者によって、2.2節で示したキレの定義である「素早い動作をびたりと止める」ことが出来ている箇所を抽出し、「キレがある」という情報を付与する。キレのないダンスを踊る人については、「キレがある」という情報を付与した振付と同じ振付を踊る箇所に対し、「キレがない」という情報を付与した。

このキレ情報を元に、時系列を考慮したデータセットを作成する。ここでは、キレ情報を付与した拍に対し、前後拍と合わせ

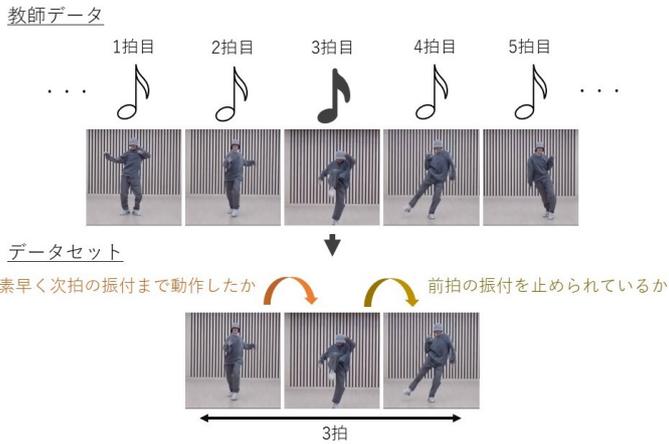


図 1: 3 拍目が「素早い動作をびたりと止められている」ときのデータセット

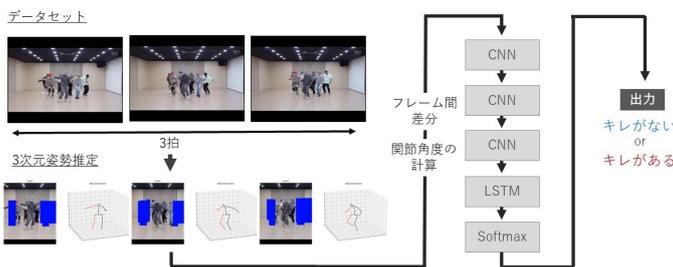


図 2: 学習の流れ

表 1: 実験結果

	2次元	3次元
位置座標	81.5 %	82.7 %
関節角度	79.0 %	81.6 %

た 3 拍のウィンドウ幅によってデータセットを作成することで、短期間の区間によるダンス動作とする (図 1). これらの作業によって、キレのあるダンスデータ 299 件、キレのないダンスデータ 343 件を取得した. このデータセットの作成から学習を行うまでの流れを図 2 に示す.

4.3 学習条件

学習条件としては、活性化関数に ReLU 関数、プーリングに MaxPooling を用いた. また、バッチ正規化を LSTM, Dropout を CNN と LSTM の出力層に用いた. 損失関数にはクロスエントロピーを用い、バッチサイズを 10 とし、400 エポックまで 0.001 の学習率で訓練した.

5 結果

10 分割交差検証 (1 分割ごとに、異なる人物をテストデータに振り分け、テスト用データの人物は、訓練用データと検証用データに含まれないようにする) を用いて、ダンスのキレ識別の性能を評価した (表 1). 最も精度が高かったのは、3 次元による関節の位置座標を骨格情報としたデータセットで精度は 82.7 % であった.

関節の位置座標、角度による骨格情報のどちらも、2 次元より 3 次元情報の精度が高かったことから、3 次元の骨格情報を用いることは、ダンスのキレを採点するにあたって有効であると言える.

2 次元、3 次元情報ともに関節の角度よりも位置座標を用いた方が高い精度となった. 関節角度の精度が十分でなかった理由として、本システムで用いる姿勢推定技術が VideoPose3D であり、出力である位置座標から関節角度を計算するという作業を介

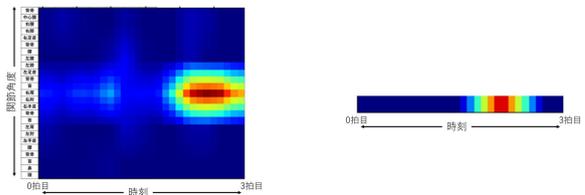


図 3: Grad-CAM による 2 層目の CNN によるヒートマップ

図 4: Grad-CAM による 3 層目の CNN によるヒートマップ

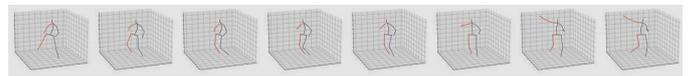


図 5: 「キレがある」と出力した場合のデータセットのダンス動作

したことが考えられる.

6 考察

学習モデルでは、CNN によってキレの特徴抽出を行っていた. この CNN がどの関節に、あるいはどの時刻に着目したのか調査するため、Grad-CAM を利用し、CNN の判断根拠について可視化する.

図 3 と図 4 は、キレのあるダンスデータを入力したとき「キレがある」と出力したデータセット (図 5) の 2 層目と 3 層目の CNN の畳み込み層に Grad-CAM を用いたヒートマップである. 2 層目の CNN では右腕の動きに、3 層目の CNN では強調すべきフレームの時刻に着目できていることがわかる. この結果から、CNN の畳み込み層では動きの大きかった関節の部位・フレームに着目しており、キレの特徴抽出に利用していると考えられる.

7 まとめ

本研究では、「踊ってみた動画」を対象とした精密なダンス採点システムの実装を行った. ダンスのキレの中でも「素早い動作をびたりと止める」ことに着目し、3 拍という短期間の区間によるダンス動作を対象としたキレ採点システムの実装を試みた. この結果、3 次元の位置座標を利用することが有効であるという結果が得られた.

今後の展望として、他の振付への適応を目指すことで、より多くの人が楽しめる「踊ってみた動画」の採点システムが実装できると考える.

参考文献

- [1] 森鈴果, 田中哉汰, 橋田光代, 片寄晴弘. 恋ダンスの達人: 深層学習を利用したダンスのキレ採点システム. Technical Report 2, 関西学院大学, 関西学院大学, 福知山公立大学, 関西学院大学, mar 2021.
- [2] 古市冨佳, 阿部和樹, 齋藤光, 中村聡史. ヒップホップダンスにおける骨格情報のみによる個人識別の検討. 情報処理学会論文誌, Vol. 62, No. 11, pp. 1792-1805, nov 2021.
- [3] Radoslaw Niewiadomski, Maurizio Mancini, Stefano Piana, Paolo Albornò, Gualtiero Volpe, and Antonio Camurri. Low-intrusive recognition of expressive movement qualities. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction, ICMI '17*, p. 230-237, New York, NY, USA, 2017. Association for Computing Machinery.
- [4] TOKYO STEPS ARTS. キレキレのダンスで観客を魅了するための 4 つのコツ&3 つのトレーニング方法.
- [5] Simin Wang, Junhuai Li, Ting Cao, Huaijun Wang, Pengjia Tu, and Yue Li. Dance emotion recognition based on laban motion analysis using convolutional neural network and long short-term memory. *IEEE Access*, Vol. 8, pp. 124928-124938, 2020.
- [6] Shuhei Tsuchida, Satoru Fukayama, Masahiro Hamasaki, and Masataka Goto. Aist dance video database: Multi-genre, multi-dancer, and multi-camera database for dance information processing. *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, pp. 501-510, 11 2019.