

ダンスジャンル自動識別による振り付け 理解支援インタフェースの検討

小林 優花^{1,a)} 田之上隼人³ 鈴木 裕真³ 西原 大輝³ 堀 隆之³ 寺田 努² 土田 修平^{1,b)}

概要：本研究では SNS 上のストリートダンス動画を対象に、複数ジャンルを時系列的に自動識別・可視化するシステムを提案する。AIST++Dataset を用いて ST-GCN++ を学習し、10 種類のジャンルを分類した結果、最も高い識別率はクリップ単位で平均 0.91、動画単位で平均 0.71 であった。さらに、再生時間に合わせてジャンル要素を動的に表示するインタフェースを開発し、ユーザの振り付け構造把握を容易にした。今後はユーザスタディを通じて、初心者が興味のあるジャンルを効率的に発見できるかを評価し、他のダンススタイルへの適用を含め学習体験の向上を目指す。

1. 研究の背景と目的

ストリートダンスの普及に伴い、TikTok や Instagram といった多様な SNS プラットフォーム上で、数多くのダンス動画が共有されている。これは、ダンスに対する関心の高まりを示唆しており、ダンスを学びたいと望む人々が増加していると推察される。一方で、ダンスについての知識量が少ない初学者がこのような投稿されている魅力的なダンスを学びたいという際に問題が生じる可能性がある。例えば、特定のダンス動画に興味を持った際、そのダンスジャンルが何を指しているのかが明確でない場合、そのダンスを学ぶための適切なレクチャー素材を見つけるのが困難となる。ハッシュタグによるジャンルの明示がある場合もあるが、すべての動画に付与されているわけではない。また、フリースタイルダンスでは複数のダンスジャンルが含まれており、組み合わせられたダンスジャンルを非専門家が特定することは難しい。さらに、プロフェッショナルなダンサーであっても、幅広いダンスジャンルに精通していない限り、一つのダンスに含まれる複数のダンスジャンルを特定することは容易ではない。

そこで本研究では、ダンス動画から抽出したモーションの時系列データを分析し、時刻ごとにダンスジャンルをラベリングすることで、ダンス動作に含まれるダンスジャンルの詳細を知ることができるシステムを構築する。これにより、ユーザは関心のあるダンスジャンルや特定の動作

が取り入れられた区間を可視的に確認しながら学習でき、より効率的なダンスの習得が期待できる。また、動画中のジャンル情報をもとに、ユーザが学びたいジャンルが含まれる動画を検索・発見することも容易になる。

2. 関連研究

ダンス動画の動作認識や検索に関する研究は、マルチモーダル情報の活用など多角的に進められている。例えば、映像と楽曲を統合した解析 [4] の研究は見られるが、これらは一つの動画を単一のジャンルとして扱うことが多い。また、民俗舞踊のように時代や地域によって動きが変化するダンスに注目し、データベース化やメタデータを用いた文化的背景の分析を行う例も報告されている [1]。一方、本研究では、単一のダンス動画内に含まれる複数ジャンルを時系列的に自動識別し、振り付け理解や学習支援につなげる。また、識別結果を可視化や検索機能と組み合わせることで、ユーザが目的のジャンル要素を素早く把握できるインタフェースの構築を目指している。既存研究ではジャンル識別の精度や映像・音楽情報の解析を単純に評価する例が多いが、本研究はその識別率を活用し、振り付け構造を可視化・理解する仕組みを提供する。

3. 実験

本稿では、提案システムの実現に向けて、ジャンル分類の推定精度を検証し、複数のダンス動作におけるラベリング結果を分析する。

3.1 前処理

検証用のデータセットとして AIST++Dataset [3] を用

¹ お茶の水女子大学

² 神戸大学

³ ソフトバンク株式会社

a) g2320511@ocha.ac.jp

b) tsuchida.shuhei@ocha.ac.jp

いた。AIST++Dataset は AIST Dance DB [5] に基づいて構築されており、3 次元・2 次元モーションの時系列データが含まれている。使用したデータは Advanced Dance に該当し、正面カメラから撮影されたダンス動画 210 本から抽出された 2 次元データを使用した。これらは 10 のダンスジャンルで構成されている。各ジャンルには、3 人のダンサーによる 21 本のダンス動画が含まれており、各ダンサーはオリジナルの振り付けを踊っている。合計で 30 人のダンサーによる 210 本のダンス動画の 2 次元キーポイントデータを使用した。210 本のダンス動画をトレーニングデータ (140 本)、テストデータ (70 本) に分割した。各ジャンルにおいては、2 人のダンサーによる 14 本の動画をトレーニングデータに使用し、残りの 1 人のダンサーによる 7 本の動画をテストデータに使用した。また、ウィンドウサイズは 100,200,300 の三つを用意し、20 または 30 ずつスライドさせながら 2 次元キーポイントデータを分割した (クリップ)。

3.2 実験設定

本研究では、時系列 2 次元キーポイントデータを扱うための学習モデルとして、Graph Neural Network の一種である ST-GCN++ [2] を使用した。ST-GCN++ は、既存の ST-GCN モデルを基に計算効率の向上と時間方向のエンコーディングを改良したモデルである。モデルのトレーニングには、ST-GCN++ の論文に準拠して、Optimizer は Stochastic Gradient Descent (SGD) を使用した。SGD の設定は、学習率 (lr) を 0.1, momentum を 0.9, weight decay を 0.0005 とした。Learning policy としては、Cosine Annealing を採用し、最小学習率 (min lr) を 0 に設定した。モデルの入力としては、各クリップのフレーム数を 100 と設定した。クリップのフレーム数が 101 以上の場合は、均等にサンプリング (UniformSample) を行い、データを抽出した。元動画の FPS が 30FPS 程度であり、前後フレーム間の情報にあまり差がないため、サンプリングによる影響は少ないと判断した。また、最大で 300 フレームに対して 100 フレームのサンプリング程度、30FPS の動画のうちの 2 フレーム飛ばし程度であるため、時系列的な情報量は十分であると判断した。トレーニングデータのバッチサイズは 512, エポック数は 80 と設定した。データ拡張としては、RandomScale と RandomRot を使用し、入力データタイプには、関節座標 (Joint) と関節座標の差分 (Joint Motion) を採用した。3 分割交差検証によりモデルの推定精度を評価する。

4. 結果

分割されたモーション (クリップ) ごとで、関節座標 (Joint) データを用いたジャンル推定精度は、ウィンドウサイズは 300 で、Stride が 20 の際の平均 0.91 が最も高かつ

表 1 分割されたモーション (クリップ) ごとでのジャンル推定精度

Window size	Stride	Acc (Joint)			
		Split1	Split2	Split3	Ave.
100	20	0.88	0.88	0.87	0.88
100	30	0.90	0.85	0.88	0.88
200	20	0.85	0.92	0.88	0.89
200	30	0.90	0.90	0.82	0.87
300	20	0.87	0.97	0.88	0.91
300	30	0.90	0.85	0.84	0.86

表 2 動画ごとでのジャンル推定精度

Window size	Stride	Acc (Joint)			
		Split1	Split2	Split3	Ave.
100	20	0.64	0.64	0.66	0.65
100	30	0.67	0.63	0.67	0.65
200	20	0.68	0.66	0.68	0.67
200	30	0.75	0.67	0.68	0.70
300	20	0.69	0.72	0.72	0.71
300	30	0.71	0.71	0.71	0.71

た (表 1)。また、動画毎で関節座標 (Joint) データを用いたジャンル推定精度は、ウィンドウサイズは 300 で、Stride が 20 または 30 の際の平均 0.71 が最も高かった (表 2)。

5. 提案手法

本研究では、ダンス動画に連動し時間経過とともに 10 種類のダンスジャンルの要素を可視化するインタラクティブなシステムを作成した。本システムでは、ユーザーが視聴したいダンス動画の名前を入力すると、入力された動画名と一致するデータを csv ファイルから抽出し、グラフとして可視化する仕組みを提供する。可視化には、JavaScript のライブラリである p5.js を使用し、動画の再生時間に連動して、グラフ上のバーが動的に変化するように設計した。さらにバーが通過した際の数値を表示する機能も実装し、各ジャンルの要素がどの程度現れているのかを直感的に把握できるインターフェースを実現した。複数のジャンル動作を含んだダンス動画を検索できるため、ダンスの振り付けを考える際に活用することができる。

6. まとめ

本研究では、ストリートダンス動画に対して、各瞬間の動作に含まれるダンスジャンルを自動的に推定し、可視化する手法を提案した。具体的には、AIST++Dataset の 2 次元キーポイントデータを用いて ST-GCN++ で時系列解析を行い、10 種類のダンスジャンルを分類した。ウィンドウサイズやストライド幅を複数パターン設定した実験を通じて、動画を細かく分割したクリップ単位での推定精度が平均 0.91、動画単位での推定精度は平均 0.71 であることを示した。また、推定したジャンル情報を動画の再生時間と連動させ、どの時刻にどのジャンルが含まれているかを

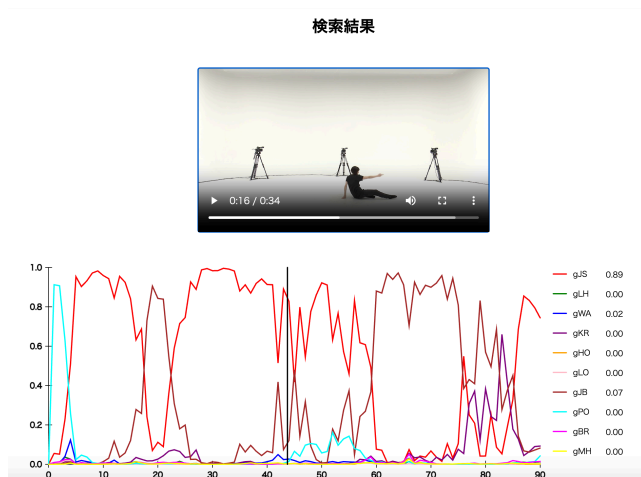


図 1 インタフェース画面

インタラクティブに可視化するシステムを開発した。

これにより、複数のジャンル要素を含んだダンスを把握しやすくなり、ユーザーは自分の学びたいダンスの詳細をより簡単に知ることが可能となる。今後の課題としては、既存の動画閲覧システムと比較して、ユーザーが目的に近いダンス動画をより効率的に発見できるかを検証するユーザスタディの実施が挙げられる。

参考文献

- [1] Aristidou, A., Shamir, A. and Chrysanthou, Y.: Digital Dance Ethnography: Organizing Large Dance Collections, *Journal on Computing and Cultural Heritage*, Vol. 12, No. 4 (2019).
- [2] Duan, H., Wang, J., Chen, K. and Lin, D.: PYSKL: Towards Good Practices for Skeleton Action Recognition, *Proceedings of the 30th ACM International Conference on Multimedia, MM '22*, pp. 7351–7354 (2022).
- [3] Li, R., Yang, S., Ross, D. A. and Kanazawa, A.: AI Choreographer: Music Conditioned 3D Dance Generation with AIST++, *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, ICCV '21*, pp. 13381–13392 (2021).
- [4] Tsuchida, S., Fukayama, S. and Goto, M.: Query-by-Dancing: A Dance Music Retrieval System Based on Body-Motion Similarity, *Proceedings of the 25th International Conference on MultiMedia Modeling, MMM '19*, pp. 251–263 (2019).
- [5] Tsuchida, S., Fukayama, S., Hamasaki, M. and Goto, M.: 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 '19*, pp. 501–510 (2019).