

YOLOv7-pose-estimation を用いたダンス動画における同期度の定量評価

Quantitative Evaluation of Synchronous Dance
Using YOLOv7-Pose-Estimation

黒崎 大希*

Daiki Kurosaki

黒木 修隆*

Nobutaka Kuroki

沼 昌宏*

Masahiro Numa

1. はじめに

近年はスポーツやエンターテインメントの練習や採点にも AI が取り入れられている。ダンスにおいても例外ではなく、個人のダンス動画を撮影することによって全身の骨格を推定し、決められたポーズが正しく行われているかを評価するシステムが存在する。また、Zhou, 矢谷ら [1] は複数人の骨格を同時に検出・比較することで、集団ダンスの同期度を評価した。しかし、集団のダンスの中の特定の人物動作を評価できる例は少ない。

2. 提案手法

単眼カメラによって撮影された集団および個人のダンスの同期度を、YOLOv7 pose estimation [2] と OC-SORT [3] を用いて定量評価する手法を提案する。

2.1 YOLOv7-pose-estimation と OC-SORT

本研究で使用した YOLOv7-pose-estimation は、図 2.1 に示す 17 個の関節点の座標情報に加えて、それぞれの関節点の信頼度や、検出された人物のバウンディングボックスを出力できる。YOLOv7-pose-estimation を利用した例を図 2.2 に示す。

さらに提案手法では OC-SORT を用いて時間方向にも関節を追い続けることで、それぞれの人物に注目した評価が可能になる。

2.2 集団全体のダンスの同期度の定量評価

本研究では 2 種類の方法を用いてダンスの同期度の定量評価を行う。1 つ目は図 2.1 に示す 17 箇所の座標を利用して同期度を評価する手法である。具体的には以下の手順によって評価を行う。 f フレーム目の m 人目の k 番目の関節点 $\mathbf{d}_{f,m,k}$ について、 M 人の座標の平均値 $\mathbf{t}_{f,k}$ は、

$$\mathbf{t}_{f,k} = \frac{1}{M} \sum_{m=1}^M \mathbf{d}_{f,m,k} \quad (1)$$

であるため、それぞれの関節点の M 人の分散 $\mathbf{v}_{f,k}$ は、

$$\mathbf{v}_{f,k} = \frac{1}{M} \sum_{m=1}^M (\mathbf{d}_{f,m,k} - \mathbf{t}_{f,k})^2 \quad (2)$$

と表すことができる。これらを全てのフレーム、すべての関節に

ついて平均し、全体のずれ \mathbf{z} を、

$$\mathbf{z} = \frac{1}{F \times 17} \sum_{f=1}^F \sum_{k=0}^{16} \mathbf{v}_{f,k}, \quad (3)$$

$$Z = \|\mathbf{z}\|_1 \quad (4)$$

のように求める。 Z が低いほど映像内の人物のダンスの同期度が高いことを表している。

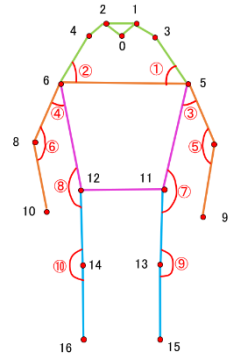
2 つ目は図 2.1 に示す 10 箇所の角度 θ の値を使用してダンスの同期度を評価する手法である。具体的には以下の手順で評価を行う。 f フレーム目の m 人目の n 番目の箇所の角度 $\theta_{f,m,n}$ の平均値 $T_{f,n}$ は、

$$T_{f,n} = \frac{1}{M} \sum_{m=1}^M \theta_{f,m,n} \quad (5)$$

となるため、それぞれの角度の分散 $V_{f,n}$ は、

$$V_{f,n} = \frac{1}{M} \sum_{m=1}^M (\theta_{f,m,n} - T_{f,n})^2 \quad (6)$$

と表される。これを全てのフレーム、すべての関節について平均し、ずれ度合い E を、



座標 d	角度 θ
0. 鼻	①左肩
1. 左目	②右肩
2. 右目	③左脇
3. 左耳	④右脇
4. 右耳	⑤左ひじ
5. 左肩	⑥右ひじ
6. 右肩	⑦左腰
7. 左ひじ	⑧右腰
8. 右ひじ	⑨左膝
9. 左手首	⑩右膝
10. 右手首	
11. 左腰	
12. 右腰	
13. 左膝	
14. 右膝	
15. 左足首	
16. 右足首	

図 2.1 検出関節点と使用角度



図 2.2 YOLOv7-pose-estimation 使用例

*神戸大学大学院工学研究科,

Graduate School of Engineering, Kobe University

$$E = \frac{1}{F \times 10} \sum_{f=1}^F \sum_{n=1}^{10} V_{f,n} \quad (7)$$

のように求める。E の値が低いほど映像内の人物のダンスの同期度が高いことを示している。

2.3 特定の人物の同期度の定量評価

特定の人物のダンスの同期度は以下の手順によって評価する。fフレーム目における m 人目の関節角度 $C_{f,m}$ を

$$C_{f,m} = (\theta_{f,m,1}, \theta_{f,m,2}, \theta_{f,m,3}, \dots, \theta_{f,m,10}) \quad (8)$$

とする。一方、fフレーム目における M 人の関節角度平均 J_f は、

$$J_f = (T_{f,1}, T_{f,2}, T_{f,3}, \dots, T_{f,10}) \quad (9)$$

と表すことができる。両者のコサイン類似度 $s_{f,m}$ は、

$$s_{f,m} = \frac{C_{f,m} \cdot J_f}{\|C_{f,m}\| \|J_f\|} \quad (10)$$

と表される。これを動画全体で平均することにより、m 人目の人物のシンクロ度 S_m を、

$$S_m = \frac{1}{F} \sum_{f=1}^F s_{f,m} \quad (11)$$

のように求める。S_m は値が大きいくほど m 人目の人物と集団全体のダンスの同期度が高いことを示している。

3. 評価実験

3.1 集団全体のダンスの同期度の評価結果

提案した評価指標 Z および E を用いて、以下の 4 つのダンス動画の評価を行った。

動画 A：一人の動画をコピーし結合することによって動画内の人物が全く同じ動作となっている動画

動画 B：プロ全員が同じ振り付けを行っている動画

動画 C：アマチュア全員が同じ振り付けを行っている動画

動画 D：プロのうち一人が異なる振り付けを行っている動画

実験結果を図 3.1 に示す。いずれの評価指標でも同じ傾向が見られた。完全なコピーを並べただけの動画 A では、ずれの評価がほぼ 0 であることから、YOLOv7-pose-estimation の検出精度の高さがうかがえる。また、その他のずれの評価は明らかに動画 B<動画 C<動画 D の順に大きくなっており、これらは主観的な評価と一致する。

3.2 特定の人物の同期度の定量評価

提案したシンクロ度 S_m を用いて以下 2 つのダンス動画の評価を行った。

動画 F：人物全員が同じ振り付けを行っている動画

動画 G：一人のみが異なる振り付けを行っている動画

実験結果を図 3.2 に示す。図 3.2(b)における斜線の棒グラフが異なる振り付けを行った人物のシンクロ度である。この値だけ明らかに他の人物の値より低いことから、提案した定量評価が主観評価をよく反映していることがわかる。以上のことから、提案手法はダンスの定量評価において有益な指標の一つになると考える。

4. まとめ

本研究では集団のダンスにおける全体および特定の人物の同期度を、YOLOv7 pose estimation と OC-SORT を用いて評価した。実験の結果、いずれの同期度も主観の評価とよく一致していることを確認した。今後はより多くの動画について人による評価と相関の高い客観的評価手法の開発に取り組みたい。

5. 参考文献

- [1] Zhongyi Zhou, Anran Xu, Koji Yatani, “SyncUp: Vision-based Practice Support for Synchronized Dancing” Proceedings of the ACM on interactive, Mobile, Wearable and Ubiquitous Technologies volume5, Issue3, Article No.143, pp.1-25,2021.
- [2] Maji, D., Nagori, S., Mathew, M., & Poddar, D. (2022). “Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2637-2646).
- [3] Cao, J., Pang, J., Weng, X., Khirodkar, R., & Kitani, K.(2023). “Observation-centric sort: Rethinking sort for robust multi-object tracking.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9686-9696).

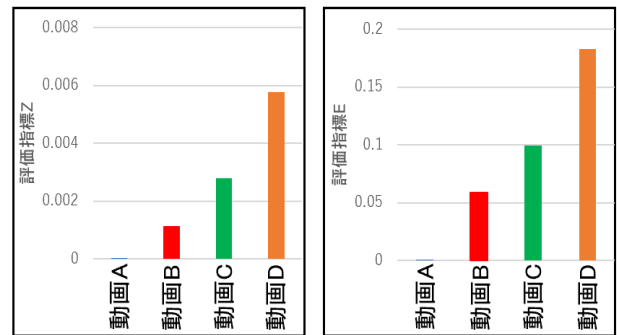


図 3.1 4 つの集団ダンスにおけるずれの評価

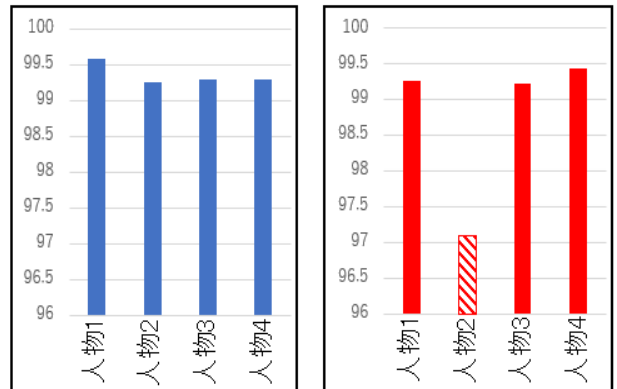


図 3.2 特定の人物のシンクロ度