

コマ送り提示を用いたダンス学習支援のための VR アプリケーション

時田 聡実^{1,a)} 石黒 祥生^{1,b)}

概要：スポーツ学習において、動画観察は一般的な学習方法である。しかし、動画では重要なフレームとそれ以外が一定の速度で提示されるため、重要な動作やポーズのタイミングを正確に捉えることが難しく、特にダンスのようなタイミングが重要な身体運動の学習においては認知負荷が高い。そこで我々は、動画からキーフレームを抽出し、コマ送りで提示する手法により、認知負荷を軽減し、ダンス学習効果を向上させると考える。動画とコマ送りを比較した事前実験を通じて、コマ送りでは次の動作の予測が難しく、またダンス学習において手首の学習が特に困難であることが示唆された。この結果を元に、コマ送りと動作軌跡の可視化を組み合わせた VR アプリケーションを開発し、従来の動画ベースの手法との学習効果を比較する評価実験を行った。結果、提案手法は従来の手法と比べ、リズム感と振付の正確性を共に向上させる傾向が見られた。初心者、経験者、プロを対象とした実験結果から、コマ送り提示の利点を確認し、身体運動の学習における潜在的な使用方法の可能性を議論する。

1. はじめに

ダンス、武道、スポーツなどの身体運動において、学習者は新しい運動技術を効率的に学ぶ必要がある。ダンスは、異なる部位で異なる動作を組み合わせた複雑な身体運動であり、ダンス学習の複雑さは過去の研究で明らかになっている [1], [2]。一般的なダンス学習方法として、動画観察による方法が用いられている。しかし、この方法ではダンス学習を次に述べるような理由から効率的にサポートできず、動画を用いたインタラクティブな研究が行われている。(a) 情報過多による注意の分散: 動画は逐次変化する多くの要素を同時に表示するため、注意が分散してしまい、手や足などの細部の要素まで確認することができない。(b) タイミングの正確性: ダンス動作はタイミングが重要であるが、動画では重要なフレームとそれ以外が一定の速度で提示されるため、重要な動作やポーズのタイミングを正確に捉えることが難しい。(c) 三次元ポーズの推測: 2D 画面を見て、正確な三次元ポーズを把握するのは難しい。

情報過多による注意の分散へのアプローチとして、Zhou らは複数人によるシンクロダンスを支援するため、シンクロしていない部分を定量的に可視化するシステム SyncUp [3] を提案している。また、タイミングの正確性へのアプローチとして、Drobny らは初心者のダンサー向けに、ビート開

始時にドラム音のフィードバックを提供する手法 Saltate を提案している [4]。三次元ポーズの推測へのアプローチとしては、ヘッドマウントディスプレイ (Head Mounted Display; HMD) を用いたシステムがいくつか提案されており、三次元アバターを多方面から観察しながら学習できるシステム [5], [6] や、動作軌跡を三次元空間に可視化するシステム [7], [8], [9] がある。

動画を用いた学習方法以外に、提示するフレーム数を減らすことで認知負荷を下げるための研究が行われている。ポーズのタイミングに合わせてボールやアバターを提示することによりダンスができる VR のダンスリズムゲームが存在するが、学習効果に関しては不明である [10], [11], [12]。ダンス学習以外では、Pei-Yu Chi らがコンピュータ上でのタスクにおいて、動画とコマ送りのミックスメディアチュートリアルシステムである MixT を提案し、複雑な直接操作の理解を助けることを示唆している [13]。

そこで本研究では、動画を用いたダンス学習における、情報過多による注意の分散の課題とタイミングの正確性の課題を解決するために、動画から特定のキーフレームを抽出し、コマ送りで提示する手法により、認知負荷を減らし、学習効果を向上させることができると考える。まず、事前実験では、ダンス動画からキーフレームを抽出し、抽出されたフレームを音楽と共にコマ送りで提示する 2D の検証システムを作成した。それを用いて、我々は参加者 12 名に、3 種類のダンスジャンルに関して動画、コマ送り、そ

¹ 東京大学

^{a)} tokidasatomi@g.ecc.u-tokyo.ac.jp

^{b)} ishiy@acm.org

して先出しコマ送りによるダンス学習タスクを設計し、学習効果を実験により検証した。その結果、コマ送りでは次の動作の予測が難しく、また全手法を通じて特に手首の学習が難しいことがわかった。次に、三次元ポーズの推測の課題へのアプローチとして、事前実験の結果を元に、コマ送り手法と手首の三次元動作軌跡の可視化を組み合わせたVRダンス学習アプリケーションであるDRFを設計した。提案システムを評価するために参加者12名による評価実験を実施した結果、DRFは動画ベースの学習手法と比べ、リズム感と振付の正確性を共に向上させる傾向が見られた。

2. 関連研究

2.1 動画ベースのダンス学習支援システム

Zhouらは、多くのダンサーが学習手法として動画を使用しているものの、従来の動画インターフェースは、複数のダンサーが同期していないポーズを効果的に特定するのをサポートしていないことを指摘した。そして、複数のダンサーのポーズの類似性と動きの時間的な整合性を定量的に可視化するフィードバックシステムであるSyncUpを開発した[3]。同様に、動画を用いたインタラクティブな学習システムが提案されてきた。例えば、土田らは、自分が模範動作を行なっている動画を生成し閲覧することにより、ダンス動作を教えるシステムを提案している[14]。また、同一のダンスを反復的に練習した際の動作の差異や変化を可視化することで、ダンスのスキル向上に寄与する可視化システム[15]や、特定の型に沿った身振り練習を支援するため、動作の誤りを補正する6種類の指示を行うシステム[16]がある。

これらの研究は、参照動画の学習中に認知負荷を軽減することに焦点を当てている。しかし、高橋らが動画と音声両方の誤りを同時に表示するとユーザの混乱を招く可能性を指摘するように[16]、1秒あたりに提示される視覚情報の量はまだ多大である。提案するコマ送り手法では、従来の動画ベースの手法が提示する多量のフレームをキーフレームのみに減らすことで、認知負荷を軽減することを目的としている。

2.2 動画を分割することによる効果

Rivièreらは、ダンサーが動画から振りを覚える際の行動を調査し、ダンサーは動画の振りを分割して覚えていることを明らかにした[17]。さらに、手作業でダンス動画を分割するためのツールを開発した。同様に、日本舞踊の動画や[18]、幅広いジャンルのダンス動画の振り[19]を機械学習を用いて個々の動きへと自動で分割する手法が提案されている。しかし、抽出した個々の動画を見ながらの学習では動画の複雑さの解決には至っていない。

Pei-Yu Chiらは、教育コンテンツの最も一般的な形式で

ある静止画のコマ送り提示と動画提示を組み合わせた混合チュートリアルを自動生成するシステムであるMixTを提案し、2つの手法の利点を組み合わせることができると指摘した[13]。DavisとAgrawalaは、動画内の物体の動作にリズムカルなパターンを取り入れたビジュアルビートの概念を開発した。彼らのシステムは、ビジュアルビートを背景音楽のビートと合わせることでダンスのような動きを合成することができる[20]。

商用のVRダンスリズムゲーム[10],[11],[12],[21]や、画面上に表示されるキャラクターの動きに合わせて踊るゲーム[22],[23]も存在する。これらのゲームでは、提示されるポーズが大幅に簡略化されており、全てのダンスのポーズが表現されているわけではない。現在の商用のゲームを模倣したダンスゲームプロトタイプを使用した研究によれば、ゲームから着想を得たインターフェース要素だけでは実際の人間のダンサーの映像に代わることはできないことが示唆されている[24]。

本研究の提案システムでは、全てのキーフレームのダンスポーズを簡略化せず、3D人型モデルとして提示する。これにより、正確なダンスポーズと音楽のリズムを同時に理解することができると思う。

2.3 VRを使用したスポーツトレーニング支援

先行研究は、スポーツトレーニングにおけるスキル向上においてVRシステムの潜在的な可能性を実証している[25]。モーション推定技術を使用することで、コンピュータはユーザの動きを追跡し、臨場感あるトレーニング体験を提供することができる。例えば、ダンストレーニングの教材[6],[26],[27]、バレエパフォーマンスの外見をチェックする方法[28]、またスキートレーニングの教材[29],[30]などに応用されている。また、HMDには、ユーザにディスプレイの位置を気にする必要なく、立体的な視覚体験を提供できるという利点があり、この利点を活かしたスポーツ学習手法が提案されている。例えば、HMDを通じて、ダンス学習者が自分の身体の動きを外部観察者として観察し、異なるアバター表現が学習者に与える影響を調べた研究がある[31],[32],[33]。

我々は、モーション推定技術を使用して3D人型モデルを作成し、動作軌跡と一緒にVR空間で提示することで、没入感のあるトレーニング環境を提供する。

3. 事前実験

インターフェースデザインを決めるにあたり、動画とコマ送りの利点と欠点を理解するため、複数の提示手法を用いたダンス学習タスクを実施した。参加者は20歳から25歳($M: 22.6, SD: 1.38$)の大学生12名(男性6名、女性6名)で、そのうち6名は部活動やダンススクールでのダン

ス経験があり（経験者と呼ぶ）、残りの6名は中学校・高校等での授業、文化祭などの校内イベントで実施された集団でのダンス程度の経験がある（未経験者と呼ぶ）。

3.1 実験課題

参加者は、以下に説明する3つの異なるダンスジャンル（Hip-hop, Lock, Jazz）のそれぞれにつき3つずつ、合計9つのダンスステップを学習した。

3.1.1 学習対象のダンス動作

学習対象のダンス動作は、Hip-hop, Lock, Jazzの3つのジャンルから構成されている。各ジャンルにつき、AIST Dance Video Database [34]の基本ダンス動画から3つの異なるダンスステップの動画をトリミングし、参照動画として使用した。これらのステップは、Hip-hopの場合はH1からH3, Lockの場合はL1からL3, Jazzの場合はJ1からJ3としてラベル付けた。全ての動画は最初に元のテンポの3/8の速さ（22.5 fps）に変更した。

3.1.2 学習手法

参加者は、2D画面に投影された以下の3つの学習手法を使用して、各ジャンルから3つのステップを学習した。

動画: 参照動画がそのまま提示される。

コマ送り: 音楽と同期して、抽出されたキーフレームが提示される。

先出しコマ送り: 音楽と同期して、抽出されたキーフレームが、動作の1拍前から動作のタイミングまでの間提示される。フレームは1拍の間で75%から徐々にサイズが大きくなり、動作のタイミングで100%の大きさとなって消える。

3.1.3 キーフレーム抽出プロセス

コマ送りと先出しコマ送りでは、参照動画から抽出されたキーフレームのみを提示する。抽出方法は、まず1拍ごとに1フレームが自動的に抽出される（図1(a)）。その後、同じポーズが維持されている場合、最初の拍のフレームのみが保持され、後続のフレームは省略される（図1(b)）。省略されたフレームでは、前のフレームが続けて表示される。このフレームの抽出プロセスは、ダンス経験が10年以上ある第1著者によって最初に行われ、その後、ダンス教師によって確認された。

3.2 実験手順

参加者はスタジオに入り、まず実験の説明を受けた。また、実験中には動作や間違いに関する具体的なアドバイスや指導は一切されないことを伝えられた。実験は、学習フェーズとテストフェーズの二つのフェーズから構成された。学習フェーズとテストフェーズは、各ステップごとに1回実施された。

学習フェーズ: 学習フェーズでは、参加者はダンスの動



図1 キーフレームの抽出プロセスの一例。最初に、(a) 各フレームはダンス動画から1拍ごとに自動的に抽出される。その後、(b) ダンス教師が重複するフレームを手動で省略する。

きを示す2D画面を見ながら、ダンスステップを4回練習した。また、参加者にはテストに備えてダンスステップを学ぶよう指示された。

テストフェーズ: テストフェーズでは、参加者は学習フェーズで学んだダンスステップを音楽を聴きながら行った。このフェーズでは画面には何も表示されない。各ステップはテストフェーズで1回だけテストされた。

個々のダンス経験が実験結果に影響を与えないようにするため、参加者が3つの学習方法を使用する順序は、可能な限り参加者間で被りが少なくなるようバランスをとった。参加者は外部ディスプレイ（133 cm × 75 cm）を見ながら練習した。実験の様子は、参加者の正面から録画された。撮影に使用したカメラはGoPro HERO11 Miniで、システムはMicrosoft PowerPointを使用して作成された。精神的な負荷を考慮して、実験中に部屋にいるのは参加者のみであった。参加者は、すべてのダンスステップを練習しテストした後、アンケートに回答した。

3.3 評価方法

テストフェーズ時に撮影された参加者が踊っている動画と参照動画間において、動作にどの程度差があるかのコストを次のように算出した。まず、参加者が踊っている動画のうちテストフェーズ時に撮影された部分をトリミングし、30 fpsでフレームに分割した。参照動画も同様にして30 fpsでフレームに分割した。次に、参加者の動画および参照動画に対してBEV [35]を適用した。BEVは、単一のRGB画像から71の人体骨格の3Dピクセル位置を算出するポーズ推定器である。ポーズ検出の精度については、本研究の範囲外であるため、正式な実験は行っていない。しかし、BEVではシンプルな背景の前で踊るダンサーのポーズの検出に成功していることが、我々の予備的な調査で確認された。したがって、本実験に使用する参照動画にはシンプルな背景の前での撮影されたものを選択し、実験中の参加者のパフォーマンスもシンプルな背景の前で撮影した。次に、ダンス動作処理の領域における先行研究 [3]と同様に、我々は14のキーポイント（鼻、首、肩、肘、手首、腰、膝、足首）を使用して13の特徴ベクトルを計算し

表 1 3つのジャンルにおける各学習手法間の効果量 (=Cohen's d).

ジャンル	動画とコマ送り	動画と先出しコマ送り	コマ送りと先出しコマ送り
Hip-hop	0.13 (非常に小さい)	0.20 (小さい)	0.36 (小さい)
Lock	0.04 (非常に小さい)	0.24 (小さい)	0.33 (小さい)
Jazz	0.00 (なし)	0.08 (非常に小さい)	0.09 (非常に小さい)

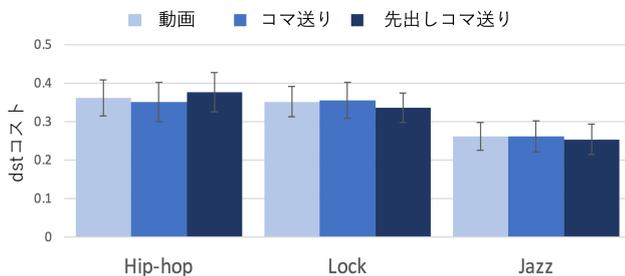


図 2 各学習手法ごとの dst コストの平均値。dst コストが低いほど、参照動画と参加者のダンス動作の類似度が高いことを示す。

た。身長やカメラからの距離が異なることによるスケールの問題を回避するため、特徴ベクトルを対応する単位ベクトルに正規化した。これらの 13 の単位ベクトルをアルゴリズムの入力として使用し、与えられたフレーム t における各身体部位 i に対する dst コスト ($dst(i, t)$) を算出する。dst コストは、以下の式に従って各ボディパーツ i に対する与えられたフレーム t における各ボディパーツの絶対差を示す。

$$dst(i, t) = |\vec{v}_P(i, t) - \vec{v}_R(i, t)| \quad (1)$$

$\vec{v}_P(i, t)$ は参加者の i 番目 ($i \in 1, 2, \dots, 13$) の単位ベクトルであり、 $\vec{v}_R(i, t)$ は参照動画中のダンサーの i 番目の単位ベクトルである。dst コストが大きければ、 i 番目の部位のポーズの差異が大きいことを意味する。

3.4 結果と考察

3.4.1 各学習手法における dst コストの平均値

各学習手法における dst コストの平均値を図 2 に示す。エラーバーは標準誤差を示す。3 種類のジャンルにおける dst コストの平均値について、各学習手法間の Cohen d 効果量 [36] を計算した。結果を表 1 に示す。Hip-hop と Lock において、動画と先出しコマ送り間、またコマ送りと先出しコマ送り間に小さな効果が見られた ($d > 0.20$)。Jazz においては、各学習手法間で非常に小さい効果が見られた ($d > 0.01$)。このことから、Hip-hop と Lock においては Jazz と比べて、各学習手法間の効果に差がある可能性がある。

3.4.2 各骨格座標における dst コスト

各学習手法における、12 名の参加者の骨格座標ごとの dst コストの平均値を図 3 に示す。各ジャンルにおいて、学習手法ごとに 13 の特徴ベクトルに対応する骨格座標の dst コストを要因として、一要因の参加者間分散分析を行った結果、全てのジャンルの全ての学習手法において、有意

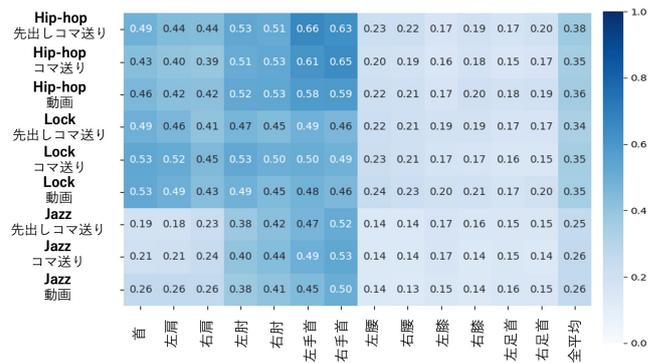


図 3 各学習手法における骨格座標ごとの dst コストの平均値。

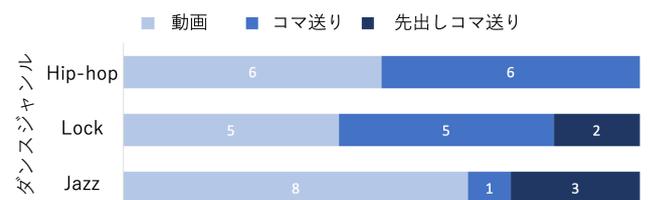


図 4 主観評価アンケート結果。

な差が見られた ($p < 0.05$)。また、首、肩、肘、手首といった上半身の dst コストは他の関節よりも大きいことが示唆され (図 3)、手首の dst コストが最も大きかったことから、提案するインターフェースには手首の動作軌跡を可視化することとした。

3.4.3 アンケート調査

図 4 は、各ジャンルにおいて、最も効果的と考えられる学習手法に関する主観評価アンケートの回答結果を示している。Hip-hop と Lock に関しては、動画を好むと回答した参加者の数とコマ送りを好むと回答した参加者の数が同数で最多であった (Hip-hop では 6 名が動画、同じく 6 名がコマ送りと回答、Lock では 5 名が動画、同じく 5 名がコマ送りと回答)。一方、Jazz の場合、参加者の大多数が動画を最も効果的な学習手法として回答した。

さらに、参加者はアンケートのコメント欄で、動画、コマ送り、先出しコマ送りの利点と欠点についてコメントした。動画の利点として、次のポーズの準備動作を見ることができ、全体の構図を明確に理解できることが挙げられた。コメントの内容を以下に示す。「動画では、準備動作を見ることができるため、次の動きを予測することができた。」「Jazz はテンポが遅いため、動画の方が理解しやすかった。」一方、「足の動きと手の動きを同時に確認するのが難しい」という複数の身体部位の動作を同時に確認するのが難しいという意見もあった。コマ送りの利点として、動きの遷移とタイミングを理解しやすいたことが挙げられた。コメントの内容を以下に示す。「動きと動きのつながりがわかりやすい。」「タイミングが学びやすい。」「慣れない Hip-hop や Lock は、コマ送りの方が動きが簡単

で学びやすかった。」一方、コマ送りの欠点としては、コマ間の動作を提示されないがために、どのように動いたらいいかわからないという点を指摘された。コメントの内容を以下に示す。「次の動作の予測がつきにくい」「足の動き、手の動きがわからなかった」「コマの間をどうしたらいいかわからなかった。」先出しコマ送りについては、「枠にくるタイミングを見極めるのが難しかった。」という意見があり、最も学習しやすい手法として選択した参加者が少なかったことも考慮すると、学習に適していない可能性が示唆された。本手法は、リズムゲームなどで、枠線にくる前から事前にタイミングを見ることができシステムを模倣して作成した。2Dでは奥行きを直接提示することができないため、理解が余計に難しくなることが示唆された。先出しコマ送り手法は2Dでの身体動作学習に向いていない可能性がある。

このことから、我々は、コマ送りシステムの欠点として、(1) フレーム間の動きを提示されないため、次の予測がつきにくいことと、(2) 異なる部位で学習難易度に差があり、特に手首の学習が最も難しいことをあげる。これを解決するべく、提案システムには奥行きを提供できるというVRの利点を活かした先出しコマ送り手法に加え、手首の動作軌跡の可視化を行う。また、先出しコマ送りのシステムの表示方法を改善するため、ポーズをはっきりと認識できる3D人型モデルを用いる。ダンスジャンルに関しては、Jazzのようなゆっくりとした動きのジャンルにおいては動画の方が適している可能性があり、Hip-hopやLockなどの振り付けが曲の拍に合わせて作られており、コマ間の動作が予測しやすいジャンルにおいてはコマ送りが適している可能性がある。これらを踏まえ、後の評価実験はHip-hopのステップに注目して行った。

4. DRF: Dance with Rhythmic Frames

提案するシステム「DRF: Dance with Rhythmic Frames」は、曲のリズムに基づくコマ送り提示と正確な三次元動作軌跡の可視化による、VRダンス学習アプリケーションである。HMDで動作するVRインターフェースとして、3DゲームエンジンであるUnity^{*1}を用いて開発され、Meta Quest 2^{*2}上で表示される。

本システムの概要を図5に示す。ユーザはVRシーン内のステージの中央に立つ。3D人型モデルは、フロアの端にあるサークルから、ポーズのタイミングの4拍前に現れ、ユーザに向かって一定の速度で移動し、ポーズのタイミングで赤線に到達し、消える。3Dモデルのポーズは鏡映されており、ユーザは単にポーズを模倣することでダンスステップを学ぶことができる。

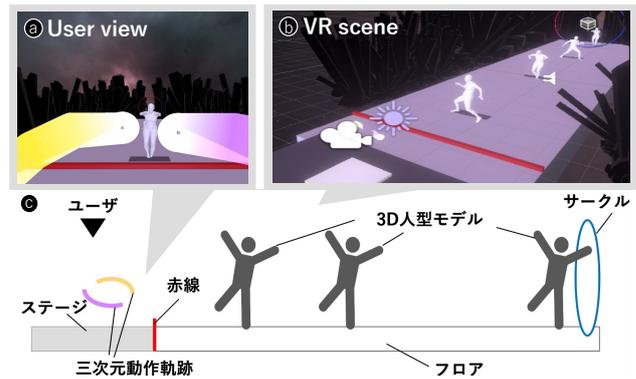


図5 DRFのVRシーンの構成図。ユーザは3D人型モデルと三次元動作軌跡を同時に観察することができる。



図6 BEV [35]を用いて、単一のRGB画像から3D人型モデルを作成する方法。入力画像(左)、入力画像上に重ねられた3D人型モデル(中央)、上から見た3D人型モデル(右)。

4.1 単眼カメラ映像によるモーション推定

3D人型モデルと、三次元動作軌跡の可視化を行うにあたり、モーション推定には、2D動画内の人物の三次元骨格を機械学習で推定するBEV [35]を用いる。BEVによって推定した三次元骨格座標をもとに、3Dメッシュ人型モデルを作成することができる(図6)。モーション推定を行うにあたり、3D人型モデルの作成には、ダンス動画から第3.1.3項の方法で抽出されたキーフレームを入力とし、三次元動作軌跡の可視化には、参照動画を60fpsで抽出したフレームを入力とした。

4.2 リズムに基づくコマ送りでの3D人型モデル表示

作成された3D人型モデルを、曲のリズムに合わせて提示するために、リズムに基づくMIDIファイルを作成して使用する。MIDIファイルには音を発するタイミングや消すタイミングの情報が含まれているため、UnityでMIDIファイルを読み込み、音を発するタイミングに合わせて学習者の前方に3Dモデルを提示することで、リズムに基づくコマ送り提示での3Dモデル表示を行う。また、次のポーズをわかりやすくするために、3Dモデルは学習者の前方から学習者の方向に一定速度で近づいてくるシステムとする。3Dモデルは特定の位置まで進み、ポーズをとるタイミングに合わせて消える。次のポーズの3Dモデルは、一つ前の3Dモデルが消えると同時に前方に確認することができる。

4.3 三次元動作の軌跡の可視化

ダンス動作の軌跡を学習者の周囲に線で可視化すること

*1 Unity, <https://unity.com>

*2 Meta Quest 2, <https://www.meta.com/jp/en/quest/products/quest-2/>

により提示する。これにより、学習者は提示された情報に合わせて身体を動かすだけで正確な三次元姿勢を理解することができると考えられる。

BEV で推定した各画像の三次元骨格座標には、71 個の関節の x 座標, y 座標, z 座標が含まれている。本システムでは、事前実験の結果から、学習が特に難しい両手首の関節座標の軌跡を三次元空間に描くものとする。得られた三次元骨格座標は Unity で使用する座標に合わせて、Python で座標変換を行う。変換した座標を Unity で読み込み、曲のテンポに合わせて各関節の位置座標を移動させることにより、両手の動作の軌跡を三次元空間内に可視化する。図 5(a) の黄色の線は左手首の軌跡を、ピンク色の線は右手首の軌跡を表している。軌跡は透明であり、ユーザは常に 3D モデルを観察することができる。

5. 評価実験

事前実験を通じて得られた知見に基づいて設計された DRF を用いて、動画ベースの手法とコマ送りでの手法による VR 上でのダンス学習効果を比較するための評価実験を行った。参加者は 20 歳から 27 歳 ($M: 22.9, SD: 4.08$) の大学生 12 名 (男性 5 名, 女性 7 名) で、そのうち 5 名は部活動やダンススクールでのダンス経験があり (経験者と呼ぶ)、残りの 7 名は中学校・高校等での授業、文化祭などの校内イベントで実施された集団でのダンス程度の経験がある (未経験者と呼ぶ)。個人のダンス経験が実験結果に影響を与えないようにするため、実験を通して経験者と未経験者のバランスを慎重にとった。過去の VR 経験の有無に関しては、全ての参加者が、「ほとんどない。」または「全くない。」と答えた。

5.1 実験課題

参加者は以下に説明する 2 種類のダンスステップを異なる学習手法で学習した。

5.1.1 学習対象のダンス動作

学習対象のダンス動作は、Hip-hop のジャンルの 2 種類のダンスステップ (ステップ A, ステップ B) である。これらのステップは、AIST Dance Video Database [34] の Middle Hip-hop のジャンルの基本ダンス動画のものを使用し、4 分音符を 1 拍として、8 拍分の長さでトリミングした。学習を容易にするために、全ての動画は最初に元のテンポの $\frac{3}{8}$ の速さ (22.5 fps) に変更した。

5.1.2 学習手法

参加者は、以下の 2 つの学習手法を使用して、2 つのステップを学習した。図 7 は以下に説明する 2 種類の学習手法の VR シーンを示している。

動画ベース: 参加者の正面の同じ位置でダンス動作を行う動的な 3D 人型モデルが提示される (図 7 (a))。このモ

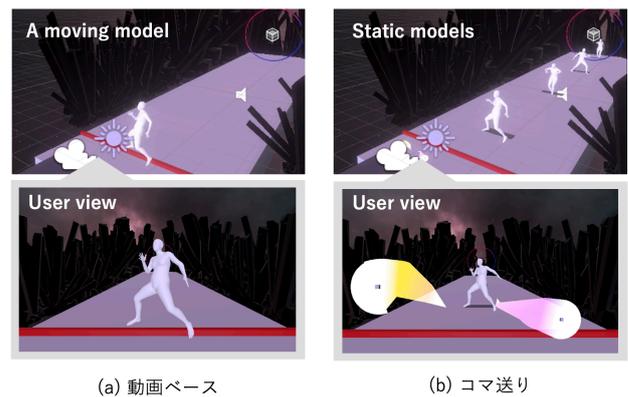


図 7 2 種類の学習手法の VR シーン。3D 人型モデルの提示方法が異なる。



図 8 評価実験に使用した 2 種類のダンスステップにおけるコマ送りで提示されるフレーム。

デルは、参加者の前で参照動画の人物が踊っているかのようにスムーズに動く。3D 人型モデルは、各参照動画から BEV を用いて作成される。

コマ送り: コマ送りシステムには、第 4 章で説明した DRF を使用する (図 7 (b))。参加者の正面で、ダンスポーズをした静的な 3D 人型モデルが、参加者の方向に一定の速度で接近する。モデルは第 3.1.3 項の方法で抽出されたフレームから作成される。図 8 はステップ A とステップ B の抽出されたフレームを示す。

5.2 実験手順

参加者はスタジオに入り、まず実験の説明を受け、2 種類のダンスステップを 2 つの異なるシステムで学ぶことを確認した。その後、HMD を装着してシステムが正常に動作することを確認した。また、実験中には動作や間違いに関する具体的なアドバイスや指導は一切されないことを伝えられた。実験は、事前実験と同様に、学習フェーズ (図 9 (a)) とテストフェーズ (図 9 (b)) の二つのフェーズから構成された。学習フェーズとテストフェーズは、各ステップごとに 2 回実施される。

学習フェーズ: 学習フェーズでは、参加者は VR システムを使用して、ダンスステップを 3 回学習した。参加者は HMD の内蔵スピーカーから聞こえてくる音楽と、HMD を通じて観察できる 3D 人型モデルをもとに学習を行った。

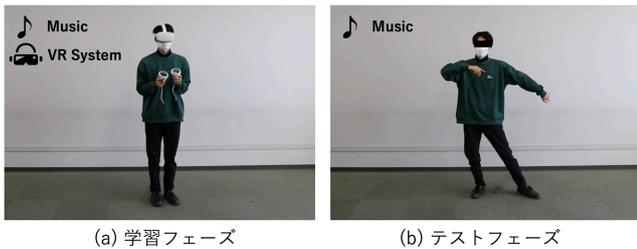


図 9 評価実験における参加者の様子.

また、VR 空間における手の位置を把握できるよう、ユーザは VR 空間内で視覚化されるコントローラーを持って学習した。また、参加者にはテストに備えてダンスステップを学ぶよう指示された。

テストフェーズ: テストフェーズでは、参加者は学習フェーズで学んだダンスステップのテストを HMD を装着せず、音楽のみを聴きながら行った。各ステップは 1 回だけテストされた。

参加者が 2 つの学習方法を使う順番と、2 つのステップを学ぶ順番は、可能な限り参加者間で被りが少なくなるようバランスをとった。参加者は、すべてのダンスステップを練習しテストした後、アンケートに回答した。HMD は Meta Quest 2 を使用し、全ての実験は広くて通気性の良い部屋で行われた。また、参加者のダンス学習の様子は、学習者の前方 5 メートルの距離から Canon EOS 6D で録画された。

5.3 評価方法

テストフェーズ時に撮影された参加者が踊っている動画と、参照動画を比較することで、定量的なシステム評価を行った。評価には、2 回目のテストフェーズ時に撮影された参加者の動画を使用した。

5.3.1 参照動画とのポーズ距離

事前実験と同様に、テストフェーズ時に撮影された参加者が踊っている動画と参照動画間において、動作にどの程度差があるかのコスト ($dst(i, t)$) を算出した。参加者の動画と参照動画は、それぞれ 30fps でフレームに分割した後、評価に使用した。

5.3.2 プロダンサーによる評価

参加者を撮影した動画について、振付とリズムを正確に踊れているかについて、ダンス経験 7 年以上でダンス教師経験のある 3 名のプロダンサーに 1 点から 7 点までの 7 点満点で評価を依頼し、3 名の平均値を評価対象とした。我々は、参加者のダンス動画とその評価を共有するためのオンラインフォームを作成し、集計はオンラインフォームを通じて行われた。

はじめに、プロダンサーはダンスステップ (ステップ A, ステップ B) の参照動画を視聴し、動画内のダンサーの動作が 7 点 (満点) であることを確認した。その後、彼らは

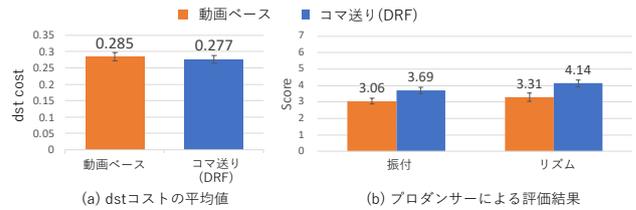


図 10 評価実験の結果

各参加者のダンス動画を視聴し、振付の正確さとリズムの正確さの 2 つの項目について、それぞれ採点を行った。各プロダンサーは、独自の評価基準に従って評価を行った。また、本実験の目的を知らず、動画内の人物がどちらのシステムで学習したかを知らない状態で評価を行った。

5.4 結果

2 種類の学習手法における参加者 12 名の dst コストの平均値を図 10 (a) に示す。エラーバーは標準誤差を示す。経験者と未経験者それぞれの dst コストの平均値について、Cohen d 効果量 [36] の大きさを計算した。経験者 5 名については、小さな効果が見られた ($d > 0.20$)。また、未経験者 7 名については、非常に小さな効果が見られた ($d > 0.01$)。dst コストについて、学習手法ごとに骨格座標に関する 13 個の特徴ベクトルを要因とした参加者間分散分析を行った結果、どちらも有意差が見られた ($p < 0.05$)。さらに、コマ送り手法において三次元動作軌跡の可視化を行った、左手首と右手首の座標について調査した。左手首と右手首それぞれの dst コストについて、学習手法間の効果量を計算した結果、左手首において中程度の効果 ($d > 0.50$) が見られ、右手首においては小さな効果 ($d > 0.2$) が見られた。

次に、プロダンサーによる評価結果について確認する。図 10 (b) は、3 名のプロダンサーによって評価された振付の正確さとリズムの正確さの平均点を示す。エラーバーは標準誤差を示す。12 名の参加者による 2 つの学習手法の評価点について、学習手法間の効果量を計算した結果、振付およびリズムの評価条件の両方で大きな効果 ($d > 0.80$) を示した。次に、経験者と未経験者それぞれについて、2 つの学習手法による効果量を調べた結果、経験者、未経験者ともに振付およびリズムの評価条件の両方で大きな効果 ($d > 0.80$) を示した。このことから、経験者と未経験者の双方において、プロダンサーはコマ送りで学習したステップを高く評価したことがわかる。

5.5 アンケート評価

参加者は (i) 動作のわかりやすさ、(ii) 覚えやすさ、(iii) 学習のしやすさの 3 つの項目に基づいて、2 種類の学習手法を評価した。それぞれが 1 から 7 までの 7 段階で評価され、1 は動画ベースの手法を強く好むことを示し、7 はコマ

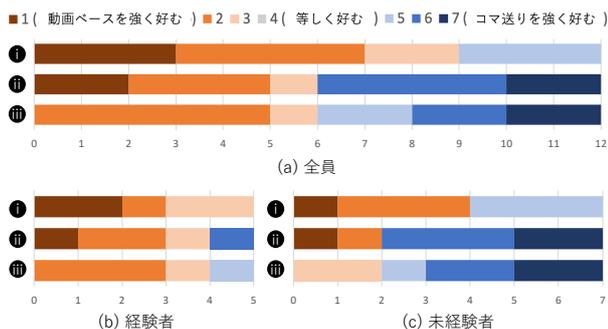


図 11 主観評価アンケート結果.

送りでの手法 (DRF) を強く好むことを示す。図 11 (a) は、全ての参加者の結果を示す。(i) について、12 名中 9 名の参加者が動画ベースの手法をより好むと回答した。よって、動画ベースの手法はコマ送りでの手法と比べて、動きを理解しやすいことが示唆される。図 11 (b) は、経験者 5 名の結果を示す。(i) について、全ての参加者が動画を好み、(ii) と (iii) について、5 名中 4 名が動画を好むと回答した。これは、動画ベースの手法がダンス経験のある学習者にとって使いやすいことを示唆している。図 11 (c) は、未経験者 7 名の結果を示す。(i) について、7 名中 4 名が動画ベースの手法を好む一方、(ii) と (iii) について、7 名中 5 名がコマ送りの手法を好むと回答した。これは、コマ送りでの手法が初心者にとって記憶しやすく、使用しやすいことを示唆している。

次に、アンケート評価とプロダンサーによる評価との関係の分析を行う。アンケートで動画ベースの手法が学習しやすいと答えた 6 名のうち、プロによる評価でも動画ベースの手法が高得点だった割合は、振付は 2 名 (33%)、リズムも 2 名 (33%) であった。また、アンケートでコマ送り手法が学習しやすいと答えた 6 名のうち、プロによる評価でもコマ送り手法が高得点だった割合は振付は 6 名 (100%)、リズムは 5 名 (83%) であった。このことから、プロによる評価は、アンケート結果によらず、コマ送り手法で学習した場合の評価点が高い傾向にあることがわかる。

また、参加者は HMD を装着してのダンス学習が快適であったかどうかを 1 (非常に快適) から 7 (非常に不快) の 7 点スケールで評価した。その平均は 3.66 で、やや快適であることがわかった。

次に、実験後に参加者とプロダンサーに行ったインタビュー結果について確認する。参加者からは、コマ送り手法が動きの理解をサポートする点について、「途中の動作が切り捨てられており、静止画像を理解するのは動画と比べて簡単だった。」「動画では、間の動きも全部覚えなければならず、一連の流れの理解が追いつかなかった。」「次のポーズまでに時間を与えてくれるため動きがわかりやすかった。」という肯定的なコメントが複数見られた。一方、「フレームだけだと途中経過がわからないため混乱した。」

「コマ送りでは間の動きを自分で想像して動く必要があったため、大変だった。」という否定的なコメントも見られ、途中の動作を正確に捉える難しさが指摘された。

また、動きの覚えやすさに関しては、「8 種類のポーズを覚えるだけで良かったため覚えやすかった。」という肯定的なコメントがある一方で、複数の経験者から、「いつも動画でしかダンス動作を見ないため、コマ送りは慣れておらず覚えづらかった。」という否定的なコメントも見られた。経験者の中には、動画に慣れており、コマ送りが不慣れに感じる人がいたことが示唆された。

その他のコメントからは、他のスポーツにも効果的である可能性が指摘された。例えば、「サッカーのキックの動きを練習するために使いたい。」「中国拳法や空手の練習など、型が決まっている種目の練習にはとても向いていると思う。」といった具体的な利用例が挙げられ、コマ送り手法が複雑な動きを含む他のスポーツの練習にも適している可能性が指摘された。

VR アプリケーションへの印象に関しては、「実際のダンス教室にいるかのような感覚を覚えた。」「ゲームとしてリリースすると面白いだろう。」「リズムゲームをプレイしている感じで、楽しかった。」などのコメントが見られ、HMD を使用した VR アプリケーションを使用することで、ユーザは実際にダンスの授業を受けているかのように感じることが示唆される。

また、プロダンサーはコマ送り手法について以下のようにコメントした。「コマ送りはリズムが正確に踊ることができるかもしれない。しかし、最終的には細かいリズムやアップダウンを見る必要があると思う。」「初心者にとって、コマ送りシステムを導入するのはいいと思う。日本では 2014 年から小学校の授業でダンスが必修化されたが、指導方法がわからない教師も多い。このシステムは、簡単にダンスを習得でき、またゲーム感覚でダンスを学べるため、効果的だと思う。」つまり、コマ送り手法は初心者に効果的であるが、中級以上のレベルでは、抽出されたフレームだけでなく、より詳細な情報が必要だと考えられる。三次元動作軌跡に関しては、「正確な手の軌跡を表示するのは良かったと思う。」「最初に両手の軌跡だけ表示し、そのあとに足の幅なども表示するとより良いシステムになるのではないか。」のように、手の軌跡は重要であるが、さらに他の体の部位の軌跡を表示することも検討すべきだと述べた。実験では、試験的に手の軌跡だけを表示したが、今後はユーザに合わせて他の部位の軌跡も表示することができるシステム検討も必要だと考えられる。

6. 考察

6.1 DRF の利点と欠点

評価実験から、コマ送り手法 (DRF) は動きの理解と覚

え方において、利点と欠点があることがわかった。経験豊富なダンサーはダンス学習に動画を使うことに慣れているため、動画ベースの手法が動きを理解するのに適している可能性がある。一方、初心者にとっては、フレーム数が少ないDRFは、ポーズの理解や記憶がしやすく、また次のポーズの前に動きを理解する時間を作ることで、動きを理解しやすくしている。特に、プロダンサーによる評価から、DRFはダンス学習において振付とリズムの正確性を有意に向上させることがわかった。

6.2 骨格間の学習難易度の差

評価実験における骨格座標間のdstコストの分析の結果、骨格座標間には有意な差があった。つまり、ダンス学習において、正しく学習しやすい部位と学習しにくい部位があることがわかる。実験結果からは、先出しコマ送り手法を用いることで、ユーザは一つのフレームを観察する時間が長くなり、習得が難しい詳細の部位にまで注意を払う時間を持つことができる可能性が考えられる。

特に左手首の学習効果を向上させた動作軌跡の可視化は、ダンスの学習に効果的であることを示唆している。しかし、インタビューの回答からは、3D人型モデルと手の軌跡の両方を同時に観察して学習することは困難であると指摘されており、一度に両手首の軌跡を確認するのは難しく、左手首を意識して学習した参加者が多かった可能性がある。このことから、3D人型モデルに注目して学習した後に、動作軌跡を付加して正確な座標を学ぶ方法など、習熟度に応じた手法も今後研究するべきであると考えられる。

6.3 ダンスの評価方法の妥当性

現在、ダンス研究におけるダンス評価方法は多岐にわたる。代表的なものとして、機械学習を用いたポーズ推定により骨格座標の差を計算する方法や、人による評価が挙げられる。評価実験では、dstコストの平均値からは小さな効果しか見られなかったものの、プロダンサーによる評価結果からは大きな効果が見られた。このことから、単純な骨格座標の差では、プロダンサーが考える「ダンスの上手さ」は測れない可能性が考えられる。我々は、評価実験において、双方を使用したため、より説得力のある結果を提示することができたと考える。

7. 結論と今後の課題

スポーツ学習において、動画観察は一般的な学習方法であるが、ダンスのような複雑な身体運動の学習においては認知負荷が高く、ダンス学習を効率的にサポートできていない。本論文は、動画からキーフレームを抽出し、コマ送りで提示する手法により、認知負荷を軽減し、ダンス学習効果を向上させることができるかどうかを調べることを目

的としている。まず、動画とコマ送りを比較した事前実験を通じて、コマ送りでは次の動作の予測が難しく、また手首の学習が特に困難であることが示唆された。この結果を元に、コマ送りと両手首の動作軌跡の可視化を組み合わせたVRアプリケーションであるDRFを開発し、従来の動画提示と学習効果を比較した評価実験を行った。結果、提案手法は従来の動画提示と比べ、リズム感と振付の正確性を共に向上させる傾向が見られた。

本稿における評価実験では、DRFがHip-hopやLockなどのジャンルに適していることが示されたが、Jazzなどのテンポが遅いダンスでは、従来の動画ベースの手法や代替手法が必要とされた。また、事前実験及び評価実験では、参加者が学びやすい基本的なステップを選択したが、将来的にはプロダンサー向けの高度なステップにも適応できるかを検討する必要がある。また、現在の実装ではコマ送り手法におけるフレーム抽出は手動で行われている。将来的には、フレーム抽出の自動化がコンテンツ制作のコストを削減できる可能性がある。さらに、DRFにはユーザへのフィードバック機能が実装されていない。ダンス結果をプロダンサーが採点したところ、本提案手法を用いた方が上手に踊れているとわかったが、ユーザからは自分が上手に踊れているかわからないというコメントがあった。実際のダンス教室でのレッスンでは、教師がフィードバックを行うように、今後の研究でフィードバック手法の開発を検討する予定である。

参考文献

- [1] Bettina Bläsing, Beatriz Calvo-Merino, Emily S. Cross, Corinne Jola, Juliane Honisch, and Catherine J. Stevens. Neurocognitive control in dance perception and performance. *Acta Psychologica*, Vol. 139, No. 2, pp. 300–308, 2012.
- [2] Nicholas EV Foster Virginia B Penhume Falisha J Karpati, Chiara Giacosa and Krista L Hyde. Dance and the brain: a review. *Annals of the New York Academy of Sciences 1337*, Vol. 1 (2015), p. 140–146, 2015.
- [3] Zhongyi Zhou, Anran Xu, and Koji Yatani. Syncup: Vision-based practice support for synchronized dancing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 5, No. 3, sep 2021.
- [4] Dieter Drobny, Malte Weiss, and Jan Borchers. Saltate! a sensor-based system to support dance beginners. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, p. 3943–3948, New York, NY, USA, 2009. Association for Computing Machinery.
- [5] Yoko Usui, Katsumi Sato, Shinichi Watabe, and Erina Yanagida. Vr teaching materials for dance practice. In *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, pp. 178–183, Toyama, Japan, 2019. CPS(IEEE Computer Society).
- [6] Simon Senecal, Niels Alexander Nijdam, Andreas Aristidou, and Nadia Magnenat-Thalmann. Salsa dance learning evaluation and motion analysis in gamified virtual reality environment. *Multimedia Tools and Applications*,

- Vol. 79, pp. 24621 – 24643, 2020.
- [7] Natsuki Hamanishi and Jun Rekimoto. Motion-specific browsing method by mapping to a circle for personal video observation with head-mounted displays. In *Augmented Humans Conference 2021*, AHs'21, p. 240–250, New York, NY, USA, 2021. Association for Computing Machinery.
 - [8] Sunada Haruya, Kiyoko Yokoyama, and Tsuyoshi Matsukawa. Study on visualization of movement information measured by motion capture system. *IEICE technical report*, Vol. 114, No. 25, pp. 113–118, 05 2014.
 - [9] Masaki Oshita. Motion volume: Visualization of human motion manifolds. In *Proceedings of the 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '19, New York, NY, USA, 2019. Association for Computing Machinery.
 - [10] Beat Games. Beat saber, 2019.
 - [11] emergeWorlds. Dance collider, 2019.
 - [12] Andromeda Entertainment. Audio trip, 2020.
 - [13] Pei-Yu Chi, Sally Ahn, Amanda Ren, Mira Dontcheva, Wilmot Li, and Björn Hartmann. Mixt: Automatic generation of step-by-step mixed media tutorials. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, p. 93–102, New York, NY, USA, 2012. Association for Computing Machinery.
 - [14] Shuhei Tsuchida, Haomin Mao, Hideaki Okamoto, Yuma Suzuki, Rintaro Kanada, Takayuki Hori, Tsutomu Terada, and Masahiko Tsukamoto. Dance practice system that shows what you would look like if you could master the dance. In *Proceedings of the 8th International Conference on Movement and Computing*, MOCO '22, New York, NY, USA, 2022. Association for Computing Machinery.
 - [15] 真美川西, 修平土田, 貴之伊藤. ダンスモーションの反復練習とその上達過程の可視化. 第 85 回全国大会講演論文集, Vol. 2023, No. 1, pp. 183–184, 02 2023.
 - [16] 高橋雅人. “振り” の練習を支援するインタラクティブシステム. インタラクシオン 2004 論文集, pp. 97–104, 2004.
 - [17] Jean-Philippe Rivière, Sarah Fdili Alaoui, Baptiste Caramiaux, and Wendy E. Mackay. Capturing movement decomposition to support learning and teaching in contemporary dance. *Proc. ACM Hum.-Comput. Interact.*, Vol. 3, No. CSCW, nov 2019.
 - [18] T. Shiratori, A. Nakazawa, and K. Ikeuchi. Detecting dance motion structure through music analysis. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.*, pp. 857–862, 2004.
 - [19] 五十嵐 健夫遠藤 輝貴. 振りの理解を助けるためのダンス動画の自動分割. 日本ソフトウェア科学会研究会資料シリーズ (Web), 2022.
 - [20] Abe Davis and Maneesh Agrawala. Visual rhythm and beat. *ACM Trans. Graph.*, Vol. 37, No. 4, jul 2018.
 - [21] Odders Lab. Ohshape, 2020.
 - [22] Ltd. Konami Amusement Co. Dance evolution, 2010.
 - [23] Ltd. Konami Amusement Co. Dancerush stardom, 2018.
 - [24] Emiko Charbonneau, Andrew Miller, and Joseph J. LaViola. Teach me to dance: Exploring player experience and performance in full body dance games. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE '11, New York, NY, USA, 2011. Association for Computing Machinery.
 - [25] Benoit Bideau, Richard Kulpa, Nicolas Vignais, Sébastien Brault, Franck Multon, and Cathy Craig. Using virtual reality to analyze sports performance. *IEEE Computer Graphics and Applications*, Vol. 30, No. 2, pp. 14–21, March 2010.
 - [26] Jacky C.P. Chan, Howard Leung, Jeff K.T. Tang, and Taku Komura. A virtual reality dance training system using motion capture technology. *IEEE Transactions on Learning Technologies*, Vol. 4, No. 2, pp. 187–195, 2011.
 - [27] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. Youmove: Enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, p. 311–320, New York, NY, USA, 2013. Association for Computing Machinery.
 - [28] Matthew Kyan, Guoyu Sun, Haiyan Li, Ling Zhong, Paisarn Muneesawang, Nan Dong, Bruce Elder, and Ling Guan. An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. *ACM Trans. Intell. Syst. Technol.*, Vol. 6, No. 2, mar 2015.
 - [29] Hidetaka Katsuyama, Erwin Wu, and Hideki Koike. Using rhythm game to train rhythmic motion in sports. In *SIGGRAPH Asia 2022 Posters*, SA '22, New York, NY, USA, 2022. Association for Computing Machinery.
 - [30] Jana Hoffard, Xuan Zhang, Erwin Wu, Takuto Nakamura, and Hideki Koike. Skisim: A comprehensive study on full body motion capture and real-time feedback in vr ski training. In *Proceedings of the Augmented Humans International Conference 2022*, AHs '22, p. 131–141, New York, NY, USA, 2022. Association for Computing Machinery.
 - [31] Saliha Akbas, Asim Evren Yantac, Terry Eskenazi, Kemal Kuscü, Sinem Semsioğlu, Onur Topal Sumer, and Asli Oztürk. Virtual dance mirror: A functional approach to avatar representation through movement in immersive vr. In *Proceedings of the 8th International Conference on Movement and Computing*, MOCO '22, New York, NY, USA, 2022. Association for Computing Machinery.
 - [32] Shuo Yan, Gangyi Ding, Zheng Guan, Ningxiao Sun, Hongsong Li, and Longfei Zhang. Outsideme: Augmenting dancer's external self-image by using a mixed reality system. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, p. 965–970, New York, NY, USA, 2015. Association for Computing Machinery.
 - [33] Georgios Tsampounaris, Katerina El Raheb, Vivi Katifori, and Yannis Ioannidis. Exploring visualizations in real-time motion capture for dance education. New York, NY, USA, 2016. Association for Computing Machinery.
 - [34] 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. In *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, Delft, Netherlands, November 2019.
 - [35] Yu Sun, Wu Liu, Qian Bao, Yili Fu, Tao Mei, and Michael J. Black. Putting people in their place: Monocular regression of 3D people in depth. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, June 2022.
 - [36] J. Cohen. Statistical power analysis for the behavioral sciences (2nd ed.). routledge. 1988.