

Title	VR を活用したスポーツ技術支援システム:動画解析によるフィードバック
Author(s)	廣野, 一也
Citation	
Issue Date	2026-03
Type	Thesis or Dissertation
Text version	author
URL	https://hdl.handle.net/10119/20502
Rights	
Description	Supervisor:金井 秀明, 先端科学技術研究科, 修士(知識科学)

修士論文

VR を活用したスポーツ技術支援システム：動画解析によるフィードバック

廣野 一也

主指導教員 金井 秀明

北陸先端科学技術大学院大学
先端科学技術研究科
先端科学技術専攻
(知識科学)

令和 8 年 3 月

Abstract

In recent years, with the proliferation of video-sharing services such as YouTube, learning sports and dance using videos has become commonplace. However, learning solely through viewing conventional 2D videos makes it difficult to check areas obscured in the demonstration or grasp detailed joint angles during the process. Furthermore, the inability of learners to receive specific feedback hinders practice.

Therefore, this study proposes a system that uses monocular camera footage as input, performs temporal and spatial alignments, and enables feedback within a VR space. This system has three main features. First, using a monocular camera to capture spatial information enables support without requiring expensive measurement equipment or sensors to be worn. Second, it adopts Temporal Cycle-Consistency Learning (TCC) as the temporal alignment method. This allows for automatic synchronization of user videos with reference videos by defining event frames only in the reference video, even when there are differences in movement speed or form shape between the reference and the user. This enables synchronization of meaningful movement phases like “take-back” and “impact.” Third, for feedback within the VR space, we implemented features beyond a simple skeleton overlay display. These include color-based highlighting based on motion deviation (root-mean-square deviation [RMSD]) and a slow-motion playback function. This enables learners to understand the areas and timing that require correction.

Furthermore, to verify the usefulness of this system, we conducted an evaluation experiment on golf swings. The experiment involved nine subjects. The results showed no statistically significant physical improvement in the swing form (i.e., no statistically significant reduction in RMSD). However, qualitative evaluation via questionnaires revealed that many subjects reported being able to “clearly recognize their own habits and differences from the model.” This indicates that the system is useful as a tool for promoting “awareness of movement” during the initial stages of motor learning.

目次

第 1 章	序論	1
1.1	本研究の目的	1
1.2	研究背景	1
第 2 章	関連研究	4
2.1	動作解析に関する研究	4
2.2	本研究の位置づけ	7
第 3 章	提案システム	9
3.1	Phase1：データ入力と時間アライメント	10
3.2	Phase2：3 次元骨格推定と空間アライメント	14
3.3	Phase3：VR 空間でのフィードバック	19
3.4	まとめ	23
第 4 章	評価実験	24
4.1	実験概要	24
4.2	定量的評価	26
4.3	定性的評価	27
第 5 章	議論	35
5.1	定量的評価に対する考察	35
5.2	定性的評価に対する考察	36
5.3	フィードバック手法に関する考察	36
5.4	骨格推定精度とアライメントの課題	37
5.5	ユーザーインターフェースとフィードバックシステムの課題	37

第 6 章	結論	39
6.1	提案手法について	39
6.2	実装システムについて	39
6.3	構築したシステムに基づく実験結果について	40
6.4	今後の課題について	40
謝辞		41
参考文献		42

目次

1.1	投球時の比較する部分 [3]	2
1.2	フィードバック（左：AR 右：ディスプレイ） [4]	2
1.3	スコアリングシステム [5]	3
2.1	処理ステップ [6]	4
2.2	お手本とユーザーの動作比較 [7]	6
2.3	ユーザーの骨格情報取得方法 [8]	7
3.1	システム概要図	9
3.2	TCC の処理 [10]	10
3.3	入力動画（左：お手本動画 右：ユーザー動画）	13
3.4	出力動画（左：お手本動画 右：ユーザー動画）	14
3.5	VideoPose3D 後の骨格データ比較	17
3.6	Kabsch Algorithm 後の骨格データ比較	18
3.7	VR 空間でのフィードバック	19
3.8	VR 空間でのフィードバック（左：重ね合わせ 右：横並び）	22
3.9	VR 空間でのフィードバック（左：誤差強調 右：肉付き骨格）	23
4.1	フィードバック手法ごとのアンケート結果（Q9～Q11）	31
4.2	フィードバック手法の順位	32

表目次

3.1	骨格のグループ化	22
4.1	被験者の属性一覧	25
4.2	各被験者における実験前後の RMSD 値および変化量	27
4.3	アンケートの集計結果 (Q1～Q8)	30

第 1 章

序論

1.1 本研究の目的

近年、YouTube やその他の動画配信サービスのような ICT を用いて、スポーツやダンスの練習をすることは一般的になっている [1]. 一般に、スポーツなどの複雑な運動スキルの習得において、学習者に自身の動作情報などを外部から提示する「拡張フィードバック」は極めて重要である. 特に、Sigrist らのレビューによれば、視覚的フィードバックは、動作の空間的な位置関係や軌跡を理解させる上で高い有効性を持つことが示されている [2]. しかし、従来の 2 次元的な動画を視聴する方法では、お手本の死角となる部分や、詳細な関節角度を正確に把握することが難しい. また、具体的なフィードバックを得ることができない点が練習の質を低下させる要因となっている.

そこで本研究では、スポーツ初学者がモーションキャプチャーなどの高価な計測機器を用いず、単眼カメラ映像 (2D 動画) のみを入力として、お手本の動作を VR 環境で理解・模倣できる学習環境の構築を目的とする. 具体的には、入力された動画から推定した 3 次元骨格情報を活用し、VR 空間内でユーザーとお手本の姿勢誤差を多様な視覚的フィードバックとして提示するシステムを開発する. さらに、開発したシステムにおいて 4 種類の異なる視覚的フィードバック手法を実装し、被験者実験を通じて、それぞれの表示形式が学習者の理解度や動作改善に与える効果を比較・検証する.

1.2 研究背景

昨今ではスポーツやダンスの練習を、VR や AR などの ICT を用いて行う手法が注目を集めている.

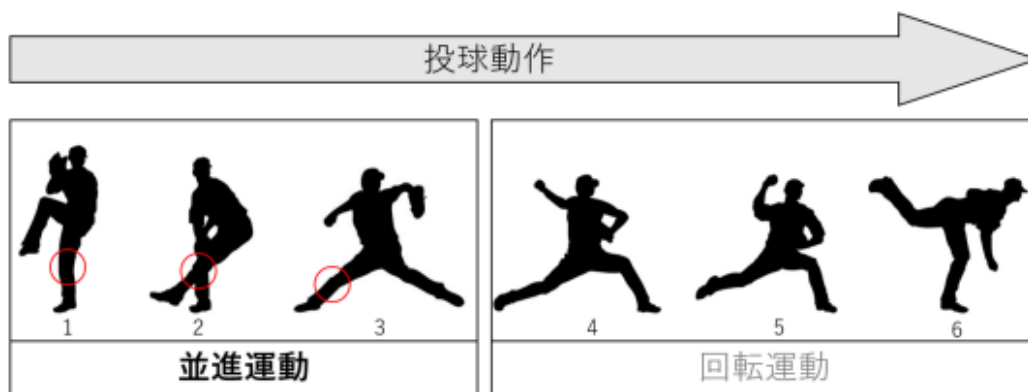


図 1.1 投球時の比較する部分 [3]

ICT を用いた先行研究として，大加戸は VR を用いた野球の目標動作とユーザー動作の差をリアルタイムで確認しながら練習できるシステムを提案している [3]．このシステムでは，MediaPipePose を用いたリアルタイム姿勢推定と VR を組み合わせ，図 1.1 に示す通り，投球動作における右膝関節の角度に着目し比較を行っている．しかしこのシステムは，特定の関節角度の比較には有効であるものの，全身の比較や，一連の投球動作全体の流れをとらえることができず，技術習得には不十分であるという側面がある．

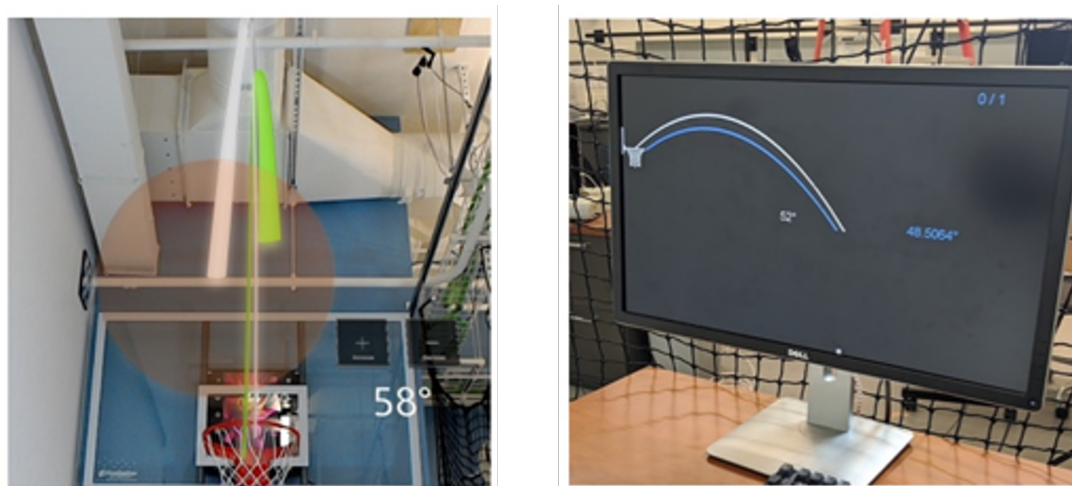


図 1.2 フィードバック（左：AR 右：ディスプレイ） [4]

また，Lin らは，バスケットボールのフリースローの練習において，リアルタイムの視覚的フィードバックを提供する AR とディスプレイを用いた 2D 可視化システムを構築した [4]．このシステムでは，図 1.2 に示す通り，シュートの軌道などを可視化しているが，

2D ディスプレイや AR グラスなどの平面的な表示デバイスを用いるため、身体の「奥行き」や「ねじれ」といった 3 次元的な位置関係を把握することが困難である。



図 1.3 スコアリングシステム [5]

他にも、専用センサや特殊カメラが必要な従来のスポーツ分析の課題を解決するため、Zhang らは RGB カメラと深層学習を用いて動作をリアルタイムで高精度に解析、評価するシステムを構築した [5]。しかしこのシステムでは、図 1.3 に示すように、ユーザーの動作に対する評価やスコアリングは行うものの、お手本となる熟達者の動作そのものは提示されないため、学習者が自身の動作との差異を視覚的に比較することはできない。

これらの課題を解決するため本研究では、VR 空間においてお手本とユーザーの全身の骨格情報を 3 次元的に可視化し、多様な視点から動作の差異を把握可能なフィードバックシステムを構築する。

第 2 章

関連研究

本章では，スポーツ動作の解析や可視化に関する先行研究について述べ，それらの課題を整理した上で，本研究の位置づけとアプローチについて述べる．

2.1 動作解析に関する研究

スポーツにおける動作習得支援に向けて，センシング技術を用いた動作の計測・解析や，その結果に基づいた定量的評価およびフィードバックを行う様々なシステムが提案されている．

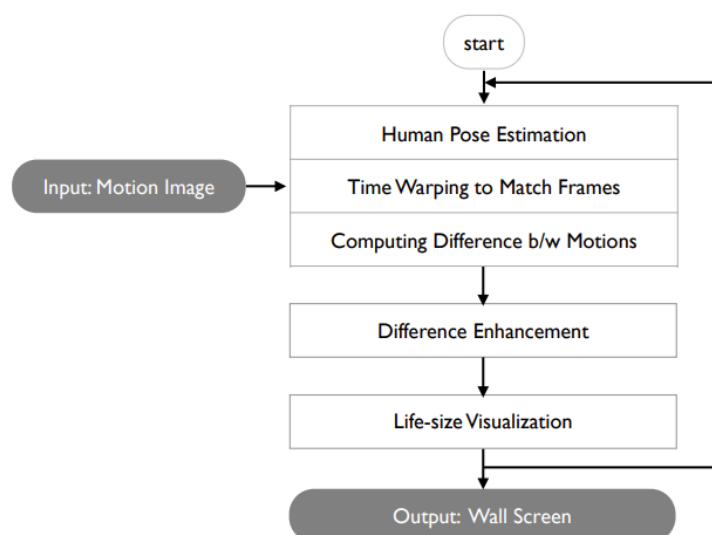


図 2.1 処理ステップ [6]

Umezu らは、野球の投球動作におけるお手本とユーザーの動作差異を可視化するシステムを提案している [6]。システムの処理ステップを図 2.1 に示す。この研究ではまず、カメラや動画から入力された映像 (Input: Motion Image) に対し、「Human Pose Estimation」のステップにおいて、OpenPose を用いた姿勢推定を行う。この際、推定される 18 点のキーポイントのうち、スポーツ動作への影響が少ない目と耳の 4 点を除いた計 14 点の座標データを抽出する。

次に、「Time Warping to Match Frames」のステップへ進む。ここでは、お手本とユーザーの動作速度の違いを補正して進行を合わせるために、抽出されたキーポイント情報を用いて改良を行った DTW 法 (動的時間伸縮法) による時間アライメントを行う。同期されたデータは「Computing Difference b/w Motions」へと渡され、フレームごとに両者の骨格情報の差分が計算される。さらに、ユーザーが自身の改善点を直感的に理解できるよう、「Difference Enhancement」のステップにて、差異が大きい部位を強調 (赤色表示) して表示する。

最終的な出力として、「Life-size Visualization」処理が行われ、プロジェクターを用いて大型壁面スクリーン (Output: Wall Screen) に等身大でフィードバック映像が投影される。

しかし、Umezu ら自身が指摘しているように、学習効果を高めるためにはスロー再生や自由な視点移動といった詳細な比較機能の実装が求められる。加えて、2 次元骨格情報に基づく比較では、動作の「奥行き」などの 3 次元的な情報を正確に把握できないという課題も残されている。

一方、Umezu らのシステムの 2 次元解析の限界や、時系列解析手法である DTW 法の課題に着目した研究もある。Geisen らは、DTW 法が時間軸を正規化する過程で、スポーツ動作において極めて重要な「固有の順序パターン」や「動作のリズム」を消失させてしまうと指摘している。そこで彼らは、これらの問題を回避しつつ空間的なズレを定量的に評価する手法として、慣性計測ユニット (IMU) を用いたモーションキャプチャースーツと Kabsch Algorithm を組み合わせたシステムを提案している [7]。

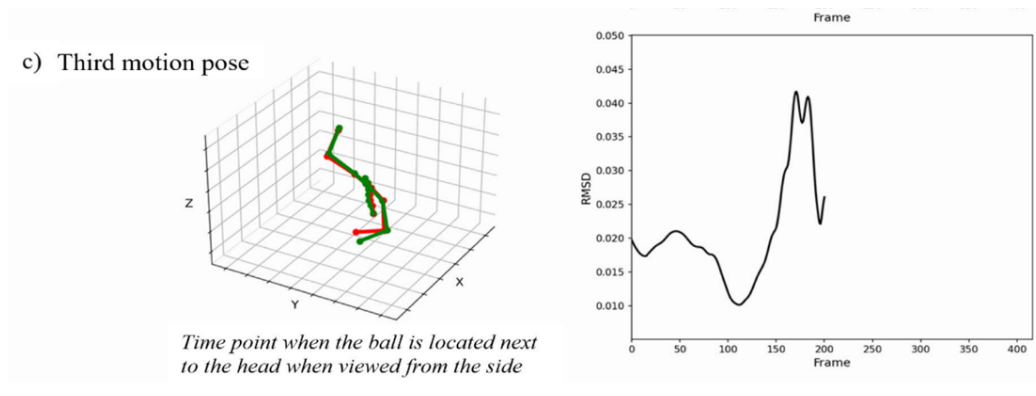


図 2.2 お手本とユーザーの動作比較 [7]

この研究では、バレーボールやハンドボールを対象に、全身の骨格情報を取得し、Kabsch Algorithm を用いてお手本とユーザーの骨格情報の向きや位置を最適に重ね合わせている。図 2.2 の左はバレーボールのスパイク動作におけるインパクトの瞬間を比較したものである。さらに、図 2.2 の右のグラフが示すように、その際の 2 乗平均平方根誤差 (RMSD) を計算することで、動作のズレを定量的に可視化することに成功している。

しかし、DTW 法のような自動的な時間正規化を行わないため、運用上の課題が残されている。Kabsch Algorithm は空間的なアライメントには有効であるが、時間的なアライメント機能を持たない。そのため、比較を行いたい動作のキーポイント（例：バレーボールのインパクトの瞬間など）を比較するには、ユーザー自身が手動でタイミングを調整し、フレームを特定する必要がある。このような手動による同期作業は、システムの利便性を損なう要因となる。

また、VR 技術を活用したトレーニングシステムも提案されている。Hiramoto らは、テニスのサーブ練習を対象に、プロ選手の単眼カメラ映像から 3 次元骨格を推定し、VR 空間内でユーザーの動作と比較可能なシステムを提案している [8]。

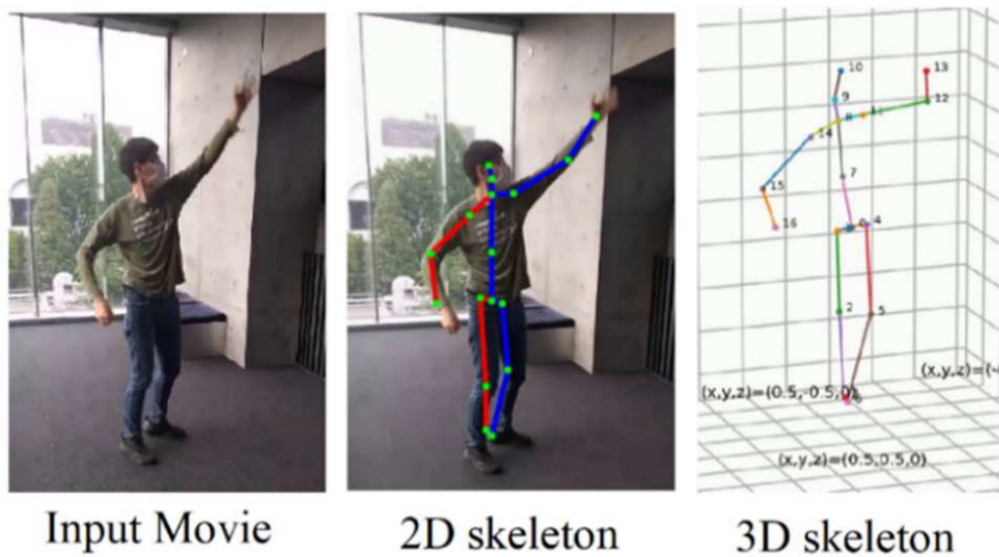


図 2.3 ユーザーの骨格情報取得方法 [8]

しかし、図 2.3 に示すように、このシステムはユーザーの動作計測にモーションキャプチャセンサーの装着を必要とするため、手軽な利用の妨げとなる。加えて、現状の機能はお手本のアバターとユーザーの動きを並列または重ねて表示し、視覚的に「確認」することに主眼が置かれており、具体的な改善箇所の指摘を行う機能の実装は課題として残されている。

さらに、より高精度な 3 次元解析や VR 空間での可視化を実現するために、大規模なセンサーシステムを用いた研究も行われている。Guo らは、複数の LiDAR とカメラを組み合わせたマルチモーダルな計測システムを提案し、バスケットボール選手の動きを精緻にアバター化して VR 空間で観戦可能にするシステムを構築した [9]。これはエンターテインメントの分野では画期的な成果であるが、広範囲なセンサーの設置を必要とする。したがって、準備の手間やコストの観点から、学習者が日々の練習で反復的に利用する「トレーニング支援ツール」として、この技術をそのまま適用することは困難である。

2.2 本研究の位置づけ

前節で述べたように、既存の動作比較システムには「2 次元情報による奥行き情報の欠如」「DTW 法による動作リズムの消失」「手動同期による利便性の低下」といった課題が存在する。加えて、3 次元解析や VR 活用を行う関連研究においては、ユーザーにウェアラブルセンサーの装着を求める点や、大規模な設備を必要とする点が、日常的な利用の障

壁となっている。さらに、Hiramoto らが今後の課題として挙げているように、単に動作を可視化・比較するだけでなく、具体的な修正箇所をユーザーに明示するフィードバック機能の実装も求められている。

そこで本研究では、これらの課題を包括的に解決するシステムを提案する。まず、時間アライメント手法として、Temporal Cycle-Consistency Learning (TCC) を採用する。TCC は、各フレームの座標位置を単純に比較するのではなく、一連の動作の流れや文脈を学習して埋め込み空間を作成する手法である。これにより、ユーザーとお手本のフォームの形状に差異がある場合でも、「テイクバック」や「インパクト」といった動作の意味的なフェーズを特定し、自動的に同期させることが可能となる。また、本研究では、Geisen らが懸念した厳密な動作のリズムそのものよりも、VR 空間内での視覚的フィードバックにおいて最も重要となる「動作フェーズの正確な一致」を優先する。次に、空間的な情報の取得に関しては、高価な計測機器を用いず、単眼カメラによる入力を採用する。最後に、フィードバック手法として、単なる重ね合わせ表示ではなく、動作のズレに基づいた色変化による強調表示機能を実装する。これにより、「どこを直すべきか」を直感的に認識させることを実現する。

第 3 章

提案システム

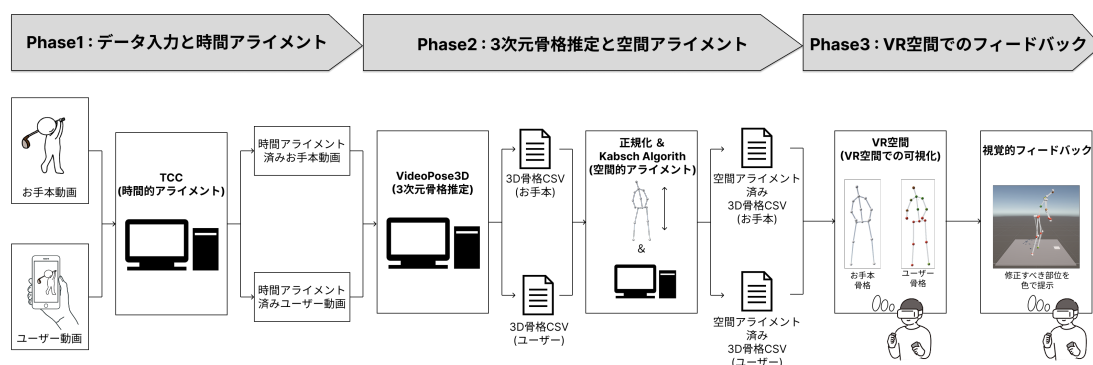


図 3.1 システム概要図

本章では、本研究で実装したフィードバックシステムの概要を述べる．本提案システムは、お手本動画とユーザー動画を入力とし、時間アライメント、空間アライメントを行い、その結果を VR 空間で表示するシステムである．その処理の流れを図 3.1 に示す．処理工程は、3 つのフェーズで構成される．

Phase1：データ入力と時間アライメント

Phase2：3 次元骨格推定と空間アライメント

Phase3：VR 空間でのフィードバック

次節では、各段階で実行される処理について、順を追って説明する．

3.1 Phase1：データ入力と時間アライメント

本フェーズでは，入力されたユーザー動画をお手本動画の時間軸に合わせるための同期処理を行う．本システムでは，単なるフレームのマッチングではなく，TCC の特徴量を用いた「時間順序を考慮した最近傍探索」と「区間ごとの時間伸縮」を組み合わせることで，自然で整合性の取れた時間アライメントを実現する．

以下に，本フェーズの中核となる技術（TCC）と，予備実験で明らかになった課題，およびそれを解決するための具体的な実装アルゴリズムについて述べる．

3.1.1 Temporal Cycle-Consistency (TCC)

まず，Phase1 で使用するアルゴリズムについて紹介する．

本提案システムでは，Dwibedi らが提案した自己教師あり学習の手法 [10] を使用する．この手法は，ラベル付けされていない動画群から，アクションの進行段階といった「時間的な意味」を持つ特徴を自動的に学習するものである．この学習を実現するための核となる概念が，時間的サイクル一貫性（Temporal Cycle-Consistency：TCC）である．TCC の基本的なアイデアは，「ある動画のフレームを別の動画の特徴空間を經由してマッピングし，再び元の動画に戻した際に，元のフレームと一致するか」という一貫性を利用することにある．

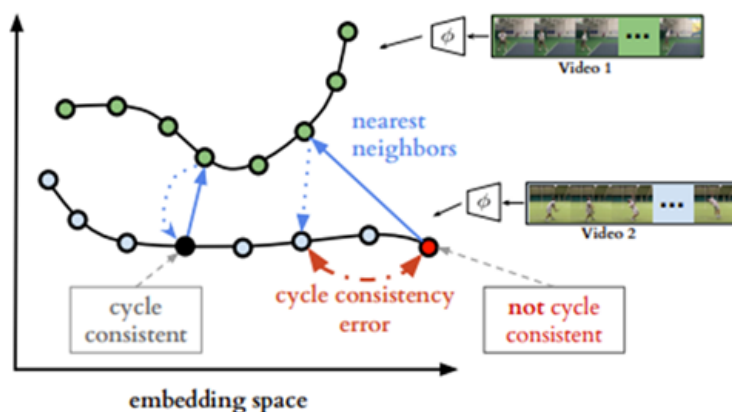


図 3.2 TCC の処理 [10]

図 3.2 は TCC のアルゴリズムである．TCC は以下の 4 つのステップを繰り返すこと

で学習を行っている。

1. Video2 の任意のフレーム（図中の黒点）を基準点として選択する
2. Video1 の特徴空間内において、その基準点に最も類似する（距離が近い）フレームを探索する
3. Video1 で特定されたフレームを起点とし、再び Video2 内を探索する
4. 戻ってきた点が、元の基準点とどれだけ近いかを評価し、その誤差を最小化するようにモデルを更新する

この TCC のアルゴリズムを用いることで、本提案システムはお手本とユーザーの動作を同期させ、比較することを可能にする。

次に、動画の時間アライメントを行うための特徴抽出器の構築について述べる。本提案システムでは、TCC モデルの学習用データセットとして、ゴルフスイング解析用データセットである GolfDB[11] を採用した。GolfDB は、YouTube 上に公開されているプロゴルファーによるスイング動画を 1,400 本収録した大規模データセットである。本研究では、この GolfDB を用いて TCC モデルの学習を行い、スイング動作の特徴を捉えた特徴量抽出器を構築した。

3.1.2 予備実験と課題

しかし、この GolfDB を用いて学習した TCC モデルを、そのままお手本動画とユーザー動画の時間アライメントに適用したところ、十分な精度が得られないという課題が生じた。スイング全体を一度に同期させようとする、個人の癖やスイング速度の違いにより、インパクトなどの重要な局面でタイミングのズレが生じることが確認された。

そこで本研究では、すべてを TCC の自動処理に任せるのではなく、お手本動画における重要なイベントを事前に定義し、それを基準にアライメントを行う手法を取り入れた。具体的には、一連のゴルフスイング動作を比較・評価するために、スイング動作を構成する主要なイベントとして「Address」「Top」「Impact」「Finish」の 4 点を定義した。Hume ら [12] は、ゴルフスイングのバイオメカニクスの解析において、これらのイベントがスイングの各フェーズ（バックスイング、ダウンスイング、フォロースルー）を区分する重要な境界点であると述べている。本手法では、この知見に基づき、お手本動画に対してこれら 4 つのイベントフレームをあらかじめ指定した上で、学習済み TCC モデルによって抽出された特徴空間上での類似度計算を行うアプローチを採用した。ただし、単純な類似度計算だけでは、似たようなフォーム（例：アドレスとインパクト直後）を取り違えるリス

クがある．そこで次節では，これらのイベントを時系列の整合性を保ちながら高精度に特定し，動画全体を同期させる具体的な実装アルゴリズムについて述べる．

3.1.3 実装アルゴリズム：時間順序を考慮した最近傍探索と区間ごとの時間伸縮

前述の課題を解決するため，本システムでは以下のアルゴリズムを実装した．具体的な処理フローは以下の3段階で構成される．

(1) 特徴量抽出と埋め込み

まず，お手本動画とユーザー動画の全フレームに対し，学習済み TCC モデルを適用し，埋め込みベクトルを抽出する．抽出されたデータは NPY 形式で保持され，各フレームが動作の意味的特徴を表す高次元ベクトルとして表現される．なお，本システムではお手本動画を基準として使用しており，その「Address」「Top」「Impact」「Finish」に対応するフレーム番号および特徴量は事前に定義されている．

(2) 順序制約付きイベント探索

次に，ユーザー動画の中から，お手本動画の各イベントに対応するフレームを特定する．ここでは，TCC の特徴空間上のユークリッド距離に基づいて探索を行うが，誤検出を防ぐため，「時系列順序制約」を適用する．単純な最近傍探索では，似た姿勢（例：アドレスとインパクト直後の姿勢）を取り違え，時間の順序が逆転するリスクがある．これを防ぐため，本実装では「あるイベント（例：Top）を探索する際，直前のイベント（例：Address）が検出されたフレームよりも後方の区間のみを探索範囲とする」という制約を加えた．これにより，スイングの時系列順序を保ったまま，各イベントに最も類似したフレームを特定することを可能にした．

(3) フェーズごとの時間伸縮

イベントフレームの特定後，各イベント間の区間ごとに動画の再生速度を調整する．スイング動作を「Address-Top」「Top-Impact」「Impact-Finish」の3つのフェーズに分割し，それぞれの区間について，ユーザー動画のフレーム数をお手本動画のフレーム数に合わせて線形補間による伸縮を行う．例えば，ユーザーの Impact から Finish がお手本より遅い（フレーム数が多い）場合は，フレームを間引くことでお手本と同じ所要時間に短縮する．逆に速い場合は補間により延長する．

本手法によりアライメント精度が向上した要因として、TCC が事前の学習によって「ゴルフスイングの意味的な特徴」を適切に獲得していたことが挙げられる。TCC の特徴空間上では、同じ動作（例：インパクトの瞬間）であれば、撮影環境や人物が異なっても類似度が高くなる性質がある。本手法では、この性質を利用してスイングの「基準点」を正確に同期させ、その間のフレームを補間することで、全体として整合性のとれた高精度なアライメントを実現した。

3.1.4 入出力と処理結果

以上の処理プロセスにおける入出力データと結果について、本実験で得られた代表的な処理結果の一例を用いて述べる。

(1) 入力データ



図 3.3 入力動画（左：お手本動画 右：ユーザー動画）

主な入力は「お手本動画（Reference Video）」と「ユーザー動画（User Video）」である。これらはスイング速度や開始タイミングが大きく異なるため、処理前の段階では同一時刻における動作は全く一致していない。図 3.3 は本実験の入力データである。お手本動画は 82 フレーム目でスイングを完了（Finish）している。しかし、ユーザー動画では本

来の Finish が 284 フレーム目にあるため、同時刻である 82 フレーム目の時点では、ユーザーはまだ「Address（構え）」の状態にあり、動作に大きな乖離が見られる。

(2) 出力データと処理結果



図 3.4 出力動画（左：お手本動画 右：ユーザー動画）

本フェーズの最終出力は、お手本動画と時間的に完全に同期するように再構成された「アライメント済みユーザー動画」である。

図 3.4 に処理結果を示す。ここでは入力時と同じ 82 フレーム目の瞬間を比較している。

処理前は Address の状態であったユーザー動画が、提案手法（順序制約付きイベント探索と時間伸縮）を適用されたことにより、お手本動画と同じ「Finish」の姿勢をとっていることが確認できる。これは、本来 284 フレーム目にあったユーザーの Finish 動作が、お手本の時間軸（82 フレーム目）に合わせて正しく同期されたことを示している。

3.2 Phase2：3 次元骨格推定と空間アライメント

Phase 1 において時間的な同期が完了したユーザー動画とお手本動画に対し、本フェーズでは 3 次元骨格情報の推定および空間的なアライメントを行う。単純に 3 次元座標を推定しただけでは、カメラの位置や被写体の体格差により座標系が一致しないため、直接的

な比較が不可能である．そこで本システムでは，VideoPose3D と，Kabsch Algorithm を用いてこの問題を解決した．

3.2.1 3次元骨格推定 (VideoPose3D)

2次元の動画から3次元の骨格情報を取得するために，Pavlo らが提案した VideoPose3D を採用した [13]．本システムにおける具体的な処理フローは以下の通りである．

(1) 2次元キーポイント検出

まず，物体検出ライブラリである Detectron2 を用いて，各フレームから人物の関節位置（17 点の COCO フォーマット）を検出する．これにより，映像が2次元座標データ (x, y) に変換される．

(2) 3次元骨格推定

検出された2次元座標系列を VideoPose3D に入力し，カメラ座標系における3次元骨格座標 (x, y, z) を出力する．これにより，動画内の人物の姿勢が3次元データとして復元される．しかし，この段階では「カメラからの距離」「身長」「身体の向き」といった外部パラメータが未調整であるため，次節の空間アライメントが必要となる．

3.2.2 空間アライメント (Kabsch Algorithm)

抽出された2つの3次元骨格データ（お手本とユーザー）を比較可能な状態にするため，空間的な位置合わせを行う．本処理には，異なる座標系を持つ2つの点群間の最適な回転行列と並進ベクトルを求めるアルゴリズムである「Kabsch Algorithm」を用いた．具体的には以下の4つの手順で空間アライメントを行う．

1. 重心の算出と移動

始めに，カメラに対する立ち位置の違いを解消するため，それぞれの点群の重心を算出する．重心を $centroid_P, centroid_Q$ とすると，以下の式で計算される．

$$centroid_P = \frac{1}{N} \sum_{i=1}^N p_i, \quad centroid_Q = \frac{1}{N} \sum_{i=1}^N q_i \quad (3.1)$$

算出された重心を各座標から引くことで，中心位置を揃えた各点の座標

$p_{i,centered}, q_{i,centered}$ を得ることができる.

$$p_{i,centered} = p_i - centroid_P, \quad q_{i,centered} = q_i - centroid_Q \quad (3.2)$$

2. スケールの正規化

ユーザーとお手本では体格に違いがあるため、そのままでは比較ができない。そこで、重心から各関節までの距離の2乗平均平方根を算出した。結果を、 $scale_P, scale_Q$ と定義する。

$$scale_P = \sqrt{\frac{1}{N} \sum_{i=1}^N \|p_{i,centered}\|^2}, \quad scale_Q = \sqrt{\frac{1}{N} \sum_{i=1}^N \|q_{i,centered}\|^2} \quad (3.3)$$

このスケール比を用いて、ユーザーの点群 $p_{centered}$ をお手本の大きさに合わせるよう正規化を行った。

$$p_{norm} = p_{centered} \cdot \frac{scale_Q}{scale_P} \quad (3.4)$$

3. 回転行列の算出

次に、ユーザーの骨格をお手本に重ね合わせるための最適な回転行列 R を求める。まず、2つの点群間の共分散行列 H を計算する。

$$H = \sum_{i=1}^N p_{i,norm} q_{i,centered}^T \quad (3.5)$$

この行列 H に対して特異値分解 (SVD) を適用し、構成要素 U, V を取得する。

$$H = U \Sigma V^T \quad (3.6)$$

これらを用いて、仮の回転行列 R を以下の式で求める。

$$R = V U^T \quad (3.7)$$

4. 鏡像の補正

SVD を用いた計算では、数学的に最適な解として「鏡像 (左右反転)」が出力される場合がある。人体の骨格において鏡像は物理的にあり得ないため、行列式を確認して補正を行う。もし回転行列の行列式 $\det(R)$ が負 (< 0) の場合、行列 V の最後の列の符号を反転させた V' を用いて、再度 R を計算する。

$$R = V' U^T \quad (\text{if } \det(R) < 0) \quad (3.8)$$

最終的に、この R を用いてユーザーの骨格座標を回転させ、お手本の座標系と一致させる。

3.2.3 入出力と処理結果

以上の処理プロセスにおける入出力データと結果について、本実験で得られた代表的な処理結果の一例を用いて述べる。

(1) 入力データ

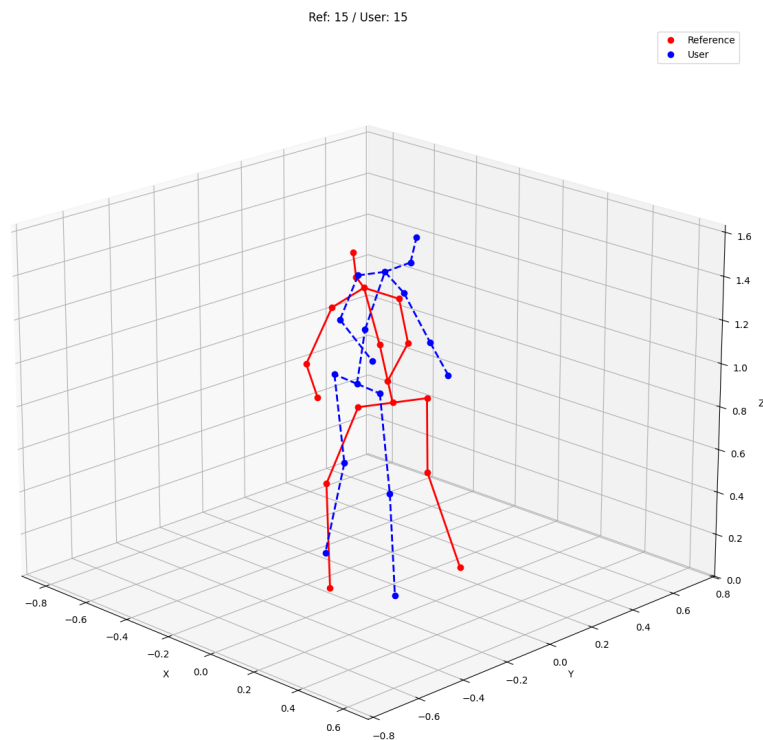


図 3.5 VideoPose3D 後の骨格データ比較

本フェーズへの入力は、Phase1 で時間同期された「ユーザー動画」である。本システムでは、まずこの動画に対して VideoPose3D を適用し、各フレームの 3 次元骨格座標を推定する。図 3.5 に、この 3 次元骨格推定後（空間アライメント前）の 15 フレーム目の 3 次元骨格データを示す。青色がユーザー、赤色がお手本の骨格を表している。この段階のデータは、VideoPose3D によって別々の動画から推定されたものであるため、カメラ位置の違いによりユーザーとお手本の間に空間的な不一致が生じていることが確認できる。したがって、この推定された生データが、次に行う空間アライメント処理（Kabsch Algorithm）への直接的な入力となる。

(2) 出力データと処理結果

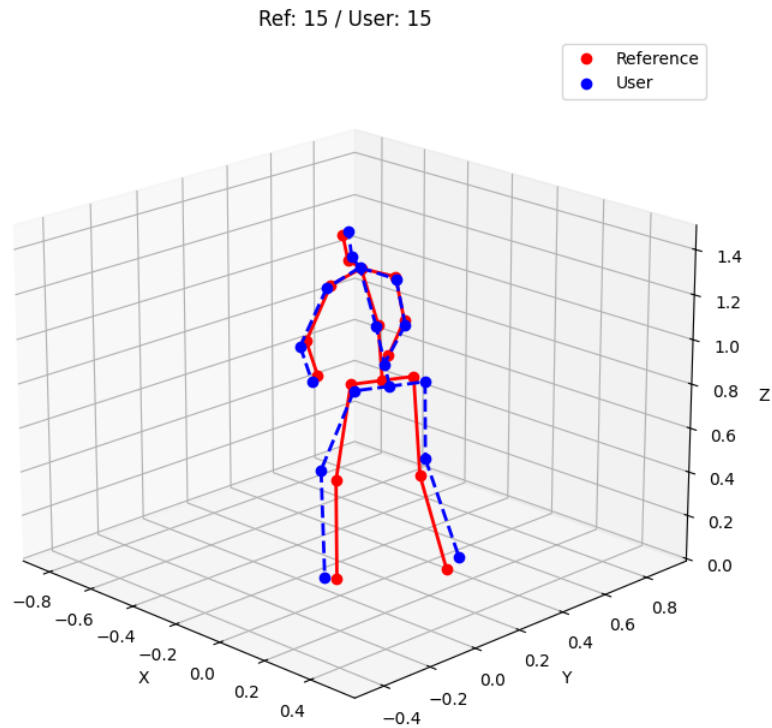


図 3.6 Kabsch Algorithm 後の骨格データ比較

本フェーズの最終出力は、Kabsch Algorithm を用いて位置・大きさ・向きを統一した「空間アライメント済み 3 次元骨格データ」である。図 3.6 に Kabsch Algorithm 適用後の 15 フレーム目の骨格比較結果を示す。提案手法（重心補正・スケール正規化・回転行列の適用）を適用したことで、ユーザー（青）とお手本（赤）の骨格が同一座標上で正しく重なり合っていることが確認できる。これにより、撮影環境や体格差の影響を排除し、純粋な「動作フォームの違い」のみを定量的に比較・評価する準備が整った。

3.3 Phase3：VR 空間でのフィードバック

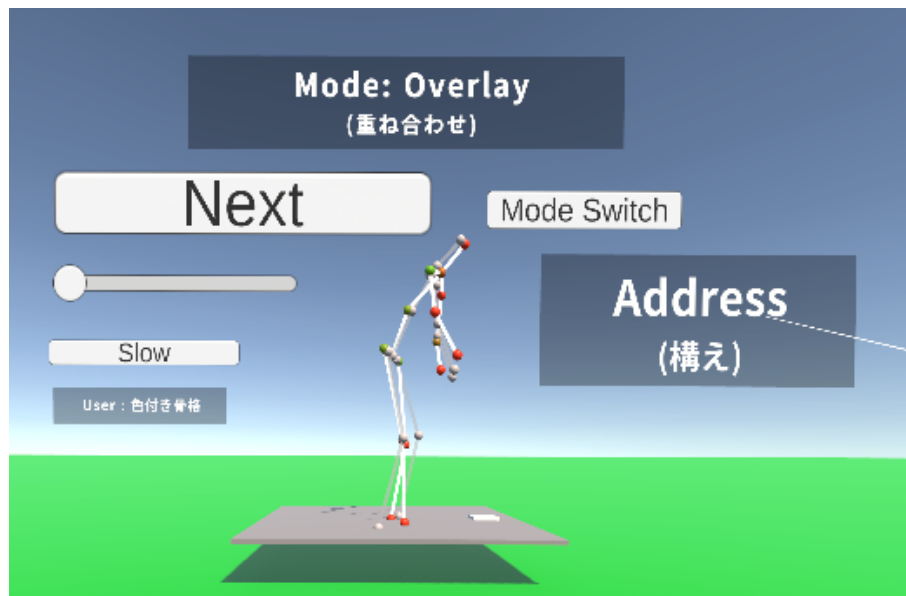


図 3.7 VR 空間でのフィードバック

Phase3 では、Phase1 および Phase2 で処理された 3 次元骨格座標データ（CSV 形式）を用い、VR 空間内でのフィードバック提示を行う。VR 機器には HTC 社の VIVE Pro 2 を使用し、フィードバック環境の構築には Unity を用いた。VR 空間におけるフィードバックの様子を図 3.7 に示す。ユーザーは利き手にコントローラを握り、トリガーやトラックパッドを操作することで、動画の再生・停止、フィードバックモードの切り替え、骨格の回転を自由に行うことができる。本システムにおいて、ユーザーが実行可能な操作は以下の通りである。

- Next ボタン
 - 動作イベント（Address, Top, Impact, Finish）間の区間再生
- ModeSwitch ボタン
 - 4 種類の視覚的フィードバックモードの切り替え
- Slow ボタン
 - 0.25 倍速によるスローモーション再生

各機能の詳細について述べる。まず「Next ボタン」は、ボタンを押すと、現在の時点か

ら次の主要な動作イベント（Address, Top, Impact, Finish）までの区間のみを再生し、イベント到達時に自動的に一時停止する機能である。動作全体を一度に連続再生するのではなく、イベント間の遷移に限定して再生することで、ユーザーは各フェーズ間の身体の動きの変化や流れを、区切りながら詳細を確認することが可能となる。

次に「ModeSwitch ボタン」は、ボタンを押すごとに、視覚的フィードバックの表示形式を「重ね合わせ」「横並び」「誤差強調」「肉付き骨格」の順で切り替えることができる機能である。ユーザーは自身の好みや確認したいポイントに応じて、最適な表示方法を選択できる。

最後に「Slow ボタン」は、動画および骨格モデルの再生速度を標準の 0.25 倍速に変更する機能である。ゴルフスイングのような高速な動作においては、通常の再生速度では確認することが困難な関節の微細な動きやタイミングのズレが生じやすい。川崎らは、高速な物体操作を伴うけん玉の VR トレーニングにおいて、対象や動作をスローに提示することで「技の動きが分かりやすくなる」「身体の使い方が覚えやすくなる」といった学習支援効果が確認されたと報告している [14]。したがって本システムでは、ユーザーが自身の動作を詳細に分析し、正しいフォームを習得・定着させるための補助機能として、このスロー再生機能を採用した。

3.3.1 視覚的フィードバックの実装と処理プロセス

本システムでは、Phase2 で作成した空間アライメント済みの 3 次元骨格データ（CSV 形式）を入力とし、Unity 上で視覚的フィードバックを生成する。以下に、入力データの定義と、各フィードバックモードにおける具体的な描画処理および誤差計算アルゴリズムについて述べる。

(1) 入力データと骨格モデルの構築

本システムにおけるフィードバック生成の基盤となる入力データは、以下の 2 種類の CSV ファイルである。

- お手本動画の骨格データ（Reference CSV）：
 - お手本となるスイング動画から抽出された、未加工の 3 次元骨格座標データを保持したファイルである。
- アライメント済みユーザー動画の骨格データ（User CSV）：
 - ユーザー動画に対し、Phase1 による時間アライメントおよび Phase2 による

空間アライメントを施したファイルである。これは、お手本動画の時間軸および座標系に合わせて正規化された状態にある。

システムは起動時にこれらの CSV ファイルを読み込み、Unity 空間内に「お手本 (Reference)」と「ユーザー (User)」それぞれの骨格モデルを生成する。骨格モデルは、関節位置を示す「球体 (Sphere)」と、関節間を接続する「ボーン (Bone)」で構成される。ボーンの描画には、「LineRenderer (線画)」と、身体の奥行きや立体感を表現するための「Cylinder (円柱)」の 2 種類を用意し、後述するフィードバックモードに応じて動的に切り替える実装とした。

(2) 誤差に基づく色変化アルゴリズム

「重ね合わせ」「横並び」「肉付き骨格」の 3 モードでは、ユーザーの動作精度を直感的に伝達するため、お手本との空間的乖離に応じた色変化フィードバックを実装した。具体的には、各フレームにおけるユーザーとお手本の対応する関節間のユークリッド距離を算出する。この算出された距離は、事前に設定した誤差を基準として正規化される。本システムでは、この最大許容誤差を 0.3m (30cm) に設定した。

この設定値は、身体の大きさとゴルフスイングの特性に基づいている。30cm という空間的乖離は、一般的な成人の頭部サイズやおよそ胴体の厚みに相当する距離である。スイング軌道においてこの規模のズレが生じている場合、フォームの修正が必要な状態であると判断できる。そのため、ユーザーが直感的に「大きくズレている」と認識すべき基準として、この値を採用した。

システムはこの正規化された値に基づき、緑色（一致時）から赤色（最大乖離時）へと遷移するカラーグラデーションを参照して、対応する関節およびボーンのマテリアル色をリアルタイムに更新する。これにより、ズレている部位が即座に赤く強調される仕組みを実現した。

(3) 部位別誤差強調 (Error Focus) の処理

「誤差強調」モードでは、全身を細かく見るのではなく、最も修正すべき部位を特定するために、身体部位ごとの誤差集計処理を行っている。全身の関節を表 3.1 に示す 6 つの部位グループ（頭部、胴体、腰部、右腕、左腕、下半身）に分類し、各グループにおける二乗平均平方根誤差 (RMSD) を算出する。システムは全グループの RMSD を比較し、最大値を持つグループを特定する。そして、その該当部位のオブジェクトのみを赤色で描画し、それ以外の部位を標準色で表示することで、ユーザーの注意を最大誤差部位に誘導

する。

表 3.1 骨格のグループ化

部位名 (Label)	構成関節 (Joints)	インデックス
頭部 (Head)	Head	10
胴体 (Torso)	Spine, Neck	7, 8
腰部 (Waist)	Hip (Center), R_Hip, L_Hip	0, 1, 4
右腕 (Right Arm)	R_Shoulder, R_Elbow, R_Wrist	11, 12, 13
左腕 (Left Arm)	L_Shoulder, L_Elbow, L_Wrist	14, 15, 16
下半身 (Legs)	R_Knee, R_Ankle, L_Knee, L_Ankle	2, 3, 5, 6

3.3.2 処理結果の提示例

前節で述べた処理によって生成される，実際の VR 空間でのフィードバック映像について述べる。

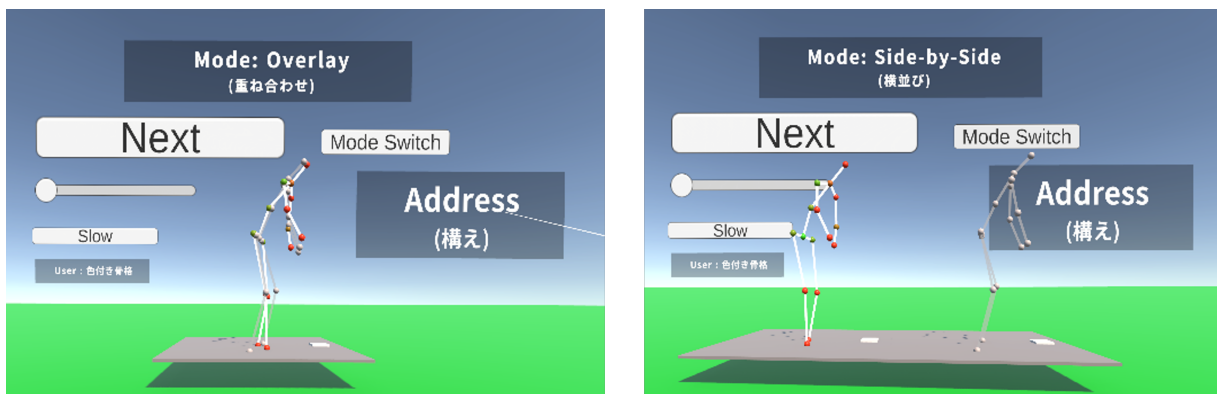


図 3.8 VR 空間でのフィードバック（左：重ね合わせ 右：横並び）

図 3.8 の左は，お手本とユーザーの骨格を同一座標上に重ね合わせて表示する「重ね合わせ (Overlay)」モードであり，右は 2 つの骨格を左右に並列して表示する「横並び (Side-By-Side)」モードである。

具体的なフィードバック手法として，重ね合わせおよび横並びモードでは，全身 16 点の関節（キーポイント）ごとに，お手本とユーザーの骨格座標の誤差に応じた色変化を実装した。誤差が小さい場合は緑色，誤差が大きくなるにつれて赤色へと変化させることで，ユーザーへ直感的な視覚的フィードバックを提供する。

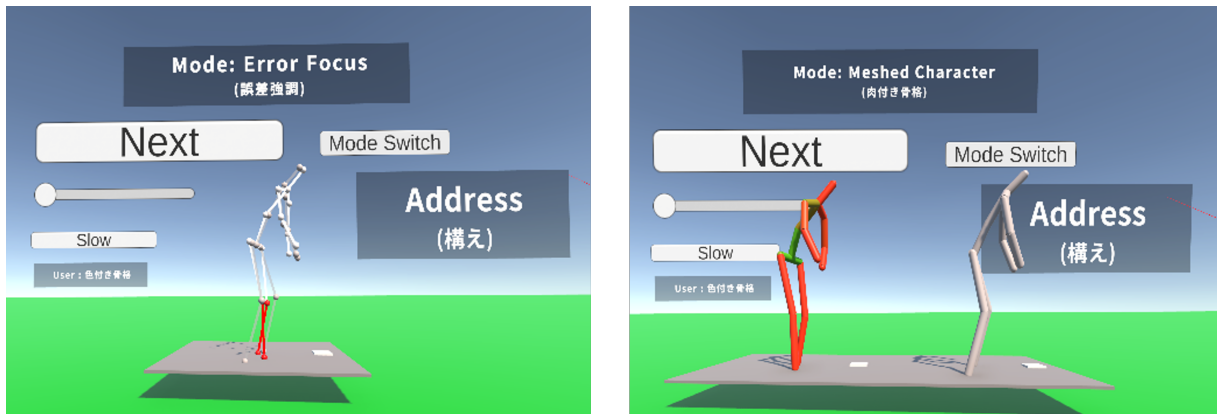


図 3.9 VR 空間でのフィードバック（左：誤差強調 右：肉付き骨格）

図 3.9 の左は「誤差強調 (Error Focus)」モード、右は「肉付き骨格 (Meshed Character)」モードである。誤差強調モードでは、お手本とユーザーの骨格を重ね合わせた上で、身体部位を表 3.1 に示す 6 つのグループに分類する。そして、各グループごとの誤差（2 乗平均平方根誤差：RMSD）を計算し、最も誤差が大きい部位のみを赤色で強調表示する。これにより、ユーザーは重点的に修正すべき部位を即座に認識できる。

一方、肉付き骨格モードは、線画表現である骨格の視認性を向上させることを目的としている。Unity のシリンダーオブジェクト（円柱）を使用して骨格モデルに肉付けを行うことで、身体の奥行きや立体感をより把握しやすい表示形式とした。また、本モードにおいても重ね合わせ・横並びモードと同様に、誤差に応じたカラーグラデーションによる色変化が適用される。これにより、ユーザーは立体的なモデルから、動作のズレを把握することが可能となる。

3.4 まとめ

本章では、本提案システムの構成と処理フローについて述べた。本提案システムでは、まず入力されたお手本動画とユーザー動画に対して、TCC を用いた時間アライメントを実行する。次に、VideoPose3D による 3 次元骨格推定を行い、得られた骨格データに対して正規化および Kabsch Algorithm を用いた空間アライメントを適用する。最終的に、時間・空間的な同期が完了した 2 つの 3 次元骨格データを Unity 上で可視化し、VR 環境を通じてユーザーへ視覚的フィードバックを提供する。

次章では、提案システムの評価実験および結果について述べる。

第 4 章

評価実験

4.1 実験概要

本実験は、提案システムの効果検証や 4 種類のフィードバックシステムを評価することを目的として行った。なお、本実験は北陸先端科学技術大学院大学 KSEC から承認（承認番号：KSEC-B20260010603）を受けたうえで実施された。

実験内容は、提案システムの定量的評価および定性的評価を行った。定量的評価では、提案システムの有効性を検証するため、被験者 9 名に対し、システム使用前と使用後のゴルフスイングにおけるお手本動作との RMSD を比較した。また定性的評価では、実験終了後に実施したアンケート（7 段階評価および自由記述）を基に評価を行った。なお、本実験では 9 名（男性 8 名，女性 1 名）が参加した。被験者属性は表 4.1 の通りである。

表 4.1 被験者の属性一覧

被験者	性別	ゴルフスイングを見たことがあるか	ゴルフスイングの経験があるか	ゴルフスイングの経験年数
user01	男性	いいえ	いいえ	0 年
user02	男性	はい	いいえ	0 年
user03	男性	はい	いいえ	0 年
user04	女性	はい	いいえ	0 年
user05	男性	はい	いいえ	0 年
user06	男性	はい	いいえ	0 年
user07	男性	はい	いいえ	0 年
user08	男性	はい	いいえ	0 年
user09	男性	はい	はい	1 年～3 年未満

4.1.1 実験タスクおよび手順

本実験は、以下の 5 つのステップで進行した。実験を通して、被験者一人当たりの所要時間は約 30 分から 40 分程度であった。

1. 事前学習とスイング撮影

まず、被験者にお手本となるスイング動画（2D）を提示した。この際、視聴時間や再生回数に制限は設けず、被験者自身がスイングのイメージを十分に掴み、満足するまで自由に視聴させた。被験者が「動作を理解した」と判断した後、その動作を模倣してスイングを行わせ、その様子を iPhone 13 のカメラ機能を用いて撮影した。このデータを「システム使用前（Before）」の評価データとした。

2. データ処理

撮影された動画に対し、提案システムによる時間・空間アライメント処理を実行した。

3. VR 空間でのフィードバック体験

データ処理完了後、被験者は VR ヘッドマウントディスプレイ（HMD）を装着し、VR 空間内で自身のスイングとお手本スイングの比較を行った。フィードバックの体験時間は 4 分から 5 分程度とした。被験者にはこの時間内で、4 種類のフィードバックモード（重ね合わせ、横並び、誤差強調、肉付き骨格）を一通り試すとともに

に、コントローラを用いた自由な視点変更や再生操作を通じて、学習するよう指示した。

4. フィードバック後のスイング撮影

VR によるフィードバックの直後に、HMD を外した状態で再度スイングを行わせ、その様子を撮影した。これを「システム使用後 (After)」の評価データとした。なお、フィードバックで得た気づきを忘れないうちに動作へ反映させるため、体験終了後、速やかに撮影を行った。

5. アンケートへの回答

すべての撮影終了後、Google フォームを用いて、システムの使用感や学習効果に関する主観評価アンケート（定性的評価）に回答させた。

次節 (4.2,4.3) で、定量的評価および定性的評価の方法、結果について述べる。

4.2 定量的評価

4.2.1 実験方法

動作の定量的評価には、お手本動作とユーザー動作の間の 2 乗平均平方根誤差 (RMSD: Root Mean Square Deviation) を用いた。RMSD は以下の式で算出される。

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{p}_i - \mathbf{q}_i\|^2} \quad (4.1)$$

ここで、 N は関節点数 ($N = 17$)、 $\mathbf{p}_i$ はアライメント後のユーザーの第 i 関節座標、 $\mathbf{q}_i$ はお手本の第 i 関節座標を表す。本システムでは VideoPose3D の出力座標系を用いており、RMSD の値はメートル (m) に相当する。例えば、 $\text{RMSD} = 0.10$ は、関節位置が平均して 10cm 離れていることを意味する。

なお、純粋な動作フォームの差異を評価するため、第 3 章で述べた、時間・空間アライメント及び正規化を行った。評価値には、スイング動作区間 (Address から Finish まで) における全フレームの RMSD の平均値を使用した。また、本実験は、システム使用前後の有意差検定にノンパラメトリック検定であるウィルコクソンの符号付順位検定を用いた。

4.2.2 実験結果

9名の被験者に対し、システム使用前（Before）と使用后（After）のRMSDを比較した。実験データを表4.2に示す。

使用前のRMSDの中央値は0.112（四分位範囲 IQR=0.012）、使用後の中央値は0.114（IQR=0.017）であった。この差の統計的有意性を確認するためウィルコクソンの符号付順位検定（両側検定）を行った結果、有意水準5%において有意な差は認められなかった（ $T = 17.0$, $p = 0.57$ ）。したがって、本実験において全被験者に一律なRMSDの減少効果は統計的に確認されなかった。

表 4.2 各被験者における実験前後の RMSD 値および変化量

被験者	Before RMSD (m)	After RMSD (m)	差分 (m)
user01	0.104	0.109	+0.005
user02	0.121	0.143	+0.022
user03	0.120	0.114	-0.006
user04	0.111	0.121	+0.010
user05	0.132	0.112	-0.020
user06	0.101	0.153	+0.052
user07	0.109	0.110	+0.001
user08	0.138	0.099	-0.039
user09	0.112	0.127	+0.015
平均 (SD)	0.116 (0.012)	0.121 (0.017)	+0.005

4.3 定性的評価

4.3.1 実験方法

実験終了後、Web アンケートツール（Google フォーム）を用いて、提案システムおよびフィードバック手法に関する定性的評価を行った。回答形式として、Q1 から Q11 までは「1：全くそう思わない」から「7：非常にそう思う」までの7段階のリッカート尺度を用いた。また、Q12では4種類のフィードバック手法に対する順位（1位から4位）を回答させ、最後に自由記述欄を設けて感想や意見を収集した。

質問項目の評価指針

本実験におけるアンケート項目は、システムの使用感を評価する SUS (System Usability Scale) や、VR でのトレーニングに関する先行研究の指標を参考にしつつ、本研究の目的である「動作習得支援の効果」を詳細に検証するために独自に設計した。具体的には、以下の 4 つの観点に基づいて構成されている。

1. 自己身体の認識とシステム没入感 (Q1, Q2) :

VR 空間上に表示された骨格モデルが、ユーザー自身として違和感なく知覚され、お手本との区別が明確になされているかを評価する。

2. VR での視覚的フィードバックの有用性 (Q3~Q6) :

提案システムの核となる「3 次元的なズレの発見」や「奥行き情報の理解」において、VR 表示や色変化機能が有効に機能したかを評価する。これは、従来の 2D 映像に対する本システムの優位性を検証するものである。

3. 学習効果への寄与 (Q7, Q8) :

フィードバック等の情報が、実際の次のアクション（身体の動かし方）のイメージ形成に繋がるのか評価する。

4. フィードバック手法特性の比較検証 (Q9~Q11) :

実装した 4 種類の手法の特性を比較するために設けた項目である。

- **Q9 (骨格の奥行き・ねじれ) :** 線画（骨格のみ）と立体（肉付き骨格）の視認性の違いを評価する。
- **Q10 (改善点の視認性) :** 特定部位の強調（誤差強調モード）が、エラー箇所での即時発見に寄与するかを評価する。
- **Q11 (全体バランスの把握) :** 2 つのモデルの配置方法（「重ね合わせ」と「横並び」）による、全体像の把握しやすさの違いを評価する。

また、具体的な質問項目は以下の通りである。

- Q1. 表示された骨格モデルを「客観的な自分」として自然に見ることができた。
- Q2. 「お手本」と「自分」の区別は直感的につきやすかった。
- Q3. 自分のスイングとお手本のスイングの「ズレ」がどこにあるか、容易に発見できた。
- Q4. 身体「奥行き（前後）方向」の位置関係やズレが理解しやすかった。

- Q5. 平面のビデオ（2D）で見るよりも，VR（3D）で見るほうが，動きの改善点がわかりやすいと感じた．
- Q6. 色が変わるフィードバックは，修正すべき箇所を見つけるのに役立った．
- Q7. フィードバックを見た後，次のスイングで『具体的にどう体を動かせばよいか』がイメージできた．
- Q8. このシステムを使うことで，練習の効率が上がると感じた．
- Q9. 骨格の「奥行き」や「ねじれ（手首や腰の回旋）」が直感的に理解しやすかった．
- Q10. 改善点が一目で分かりやすかった．
- Q11. 修正すべき箇所だけでなく，体全体のバランス（他の部位との関係）も把握しやすかった．
- Q12. フィードバックの順位付け
- Q13. 1 位の理由（記述）

4.3.2 実験結果

表 4.3 アンケートの集計結果 (Q1～Q8)

被験者	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
user01	6	3	4	4	5	6	5	6
user02	6	6	7	7	7	4	6	6
user03	6	6	6	6	7	6	6	6
user04	5	5	5	4	6	6	6	6
user05	6	3	4	3	2	6	5	6
user06	6	5	7	6	6	4	5	6
user07	7	6	6	7	7	6	6	6
user08	6	5	5	4	6	6	5	5
user09	7	6	6	7	7	6	6	7
平均	6.1	5.0	5.6	5.3	5.9	5.6	5.6	6.0
標準偏差	0.6	1.2	1.1	1.6	1.6	0.9	0.5	0.5
中央値	6	5	6	6	6	6	6	6

実験後に実施したアンケート（Q1～Q13 および自由記述）の集計結果に基づき，前述した評価指針の 4 つの観点に沿って結果を述べる．まず，システム全体の評価（Q1～Q8）を表 4.3 に示す．

1. 自己身体の認識とシステム没入感（Q1, Q2）

Q1「表示された骨格モデルを客観的な自分として見られたか」については，平均値が 6.1 ($SD = 0.6$) となり，全項目の中で最も高い評価かつ安定した結果が得られた．全ての被験者が 5 以上の回答をしており，違和感なく認識できていたことが示された．一方，Q2「お手本と自分の区別」については平均 5.0 ($SD = 1.2$) であった．多くの被験者が肯定的であったが，user01 および user05 が評価「3」をつけるなど，認識のしやすさに個人差が見られた．

2. VR での視覚的フィードバックの有用性（Q3～Q6）

Q3「ズレの発見」および Q6「色変化の有用性」は，ともに平均 5.6 を示し，標準偏差もそれぞれ 1.1, 0.9 と比較的安定した評価が得られた．Q5「2D に対する VR の優位性」は

平均 5.9 と高い値を示したが、標準偏差は 1.6 と大きくなった。これは、大部分の被験者が高い評価（6 または 7）を与えた一方で、user05 が「2」と回答した影響によるものである。同様に、Q4「奥行き方向の理解」についても平均 5.3 (SD = 1.6) となり、user05（回答：3）や user08（回答：4）のように、奥行きの把握に難しさを感じる被験者が一部存在した。

3. 学習効果への寄与（Q7, Q8）

この観点の項目は、被験者間でのばらつきが最も小さく、一貫して高い評価が得られた。Q8「練習効率の向上」は平均 6.0 (SD = 0.5) であり、Q7「次の動作イメージの形成」も平均 5.6 (SD = 0.5) であった。全被験者が本システムによる学習支援効果に対して肯定的（5 以上）な評価を示した。

4. フィードバック手法特性の比較検証（Q9～Q11）

フィードバック手法ごとの特性を比較するため、Q9（奥行きの理解）、Q10（改善点の視認性）、Q11（全体バランスの把握）について、手法別に 7 段階で評価させた結果を図 4.1 に示す。

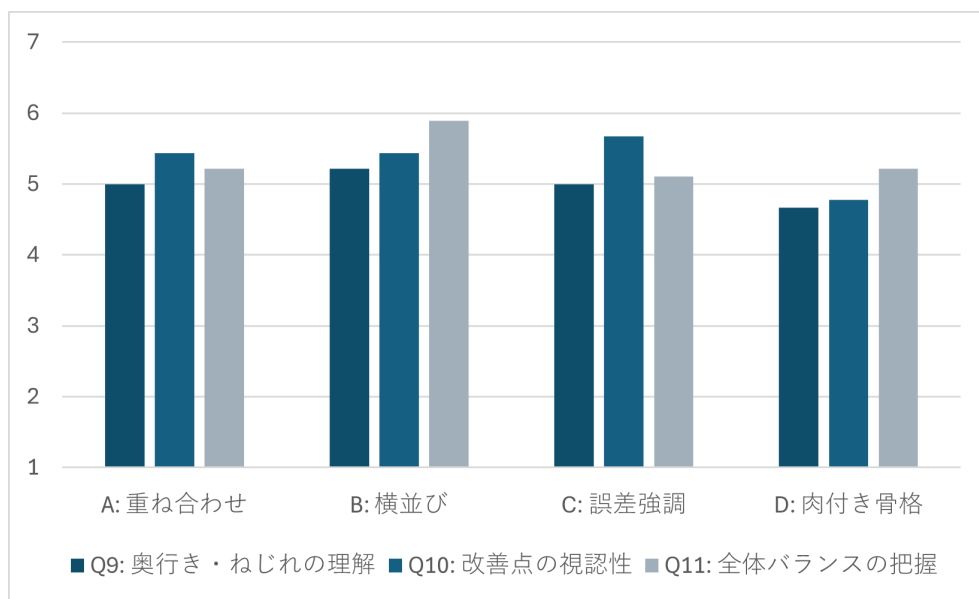


図 4.1 フィードバック手法ごとのアンケート結果（Q9～Q11）

結果として、「Q11：全体バランスの把握」においては、「B: 横並び」が最も高い評価（平均 5.89）を示した。また、「Q10：改善点の視認性」においては、「C: 誤差強調」が平均 5.67 で最も高い値となり、次いで「A: 重ね合わせ」と「B: 横並び」が同値（平均 5.44）

で続いた。一方で、「D: 肉付き骨格」に関しては、バランス把握（Q11）では平均 5.22 であったが、改善点の発見（Q10）および奥行きの理解（Q9）においては、それぞれ平均 4.78, 4.67 となり、他の 3 手法と比較して低い値を示した。

5. フィードバック手法の総合順位と選定理由（Q12, Q13）

被験者に 4 つの手法の中で最も練習に適していると感じた順に順位を付けさせ、1 位を 4 点、2 位を 3 点、3 位を 2 点、4 位を 1 点としてスコアを算出した結果を図 4.2 に示す。

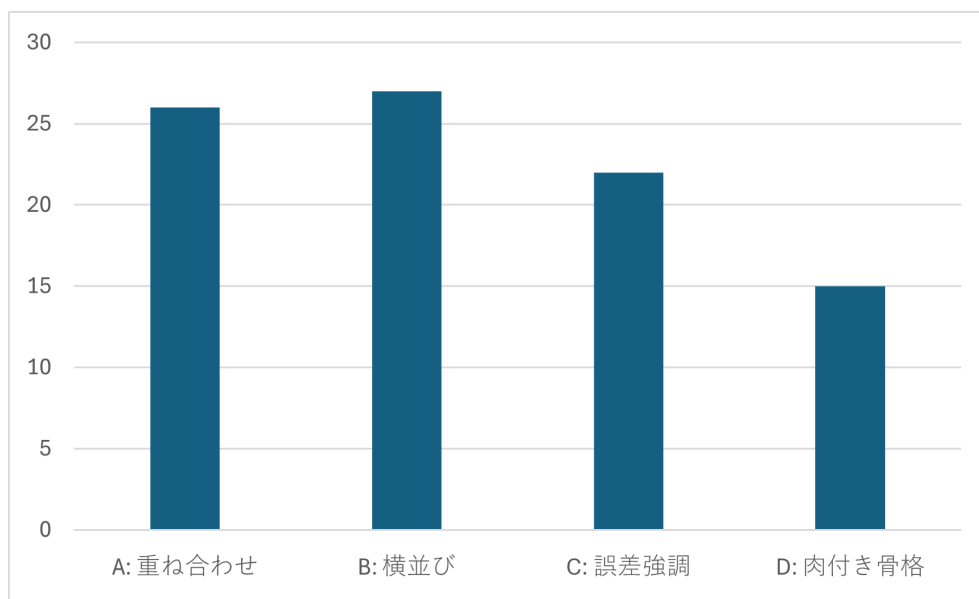


図 4.2 フィードバック手法の順位

集計の結果、「B: 横並び」が総合スコア 27 点で最も高い値を示した。次いで「A: 重ね合わせ」が 26 点となり、これら 2 つの手法のスコア差はわずか 1 点であった。一方で、「C: 誤差強調」は 22 点、「D: 肉付き骨格」は 15 点となり、上位 2 手法と比較してスコアに開きが見られた。特に「D: 肉付き骨格」に関しては、1 位に選出した被験者は 0 名であり、最も低い評価となった。

Q12 の順位付けにおいて、各手法を 1 位に選出した理由（Q13）として得られた主な回答を以下に示す。

- **A: 重ね合わせ（スコア 2 位）**

「自分のスイングと手本のスイングを動画で見比べる以上に、具体的なズレが分かりやすかった」という、直接的な比較のしやすさが評価された。

- **B: 横並び（スコア 1 位）**

「自分と手本が真横に置かれた状態であったため様々な方向から比較を行うことが出来た」や、「悪いところと全体像が見やすかった」といった意見が挙げられた。また、「重ね合わせでおかしいところを見つけてこっち（横並び）に戻ってきたらより理解しやすい」といった、併用による効果を挙げる意見も見られた。

- **C: 誤差強調（スコア 3 位）**

「お手本と自身の骨格のズレが強調されることによって、特にわかりやすかった」と、修正箇所の発見しやすさが評価の理由となった。

なお、「D: 肉付き骨格」を 1 位に選出した被験者は不在であったため、当該手法に関する選定理由は得られなかった。

6. 自由記述による意見・感想

まず、システムの実用性向上に向けた意見として、機能追加や改善に関する具体的な提案が得られた。

- 「実際に体を動きと同期してリアルタイムで指摘をもらえるとより効果的と予想する」
- 「スイングを段階的に見るのも修正点が分かりやすくよかったが、通しでみる機能があると嬉しい」
- 「フィードバックを確認している途中で手本のビデオも見ることが出来ればより細かな確認ができる」

一方で、VR 空間での 3D 表示に関しては肯定的な意見が多く寄せられた。特に、「客観的な視点による気付き」と「空間的な理解のしやすさ」の 2 点において、本システムの有効性を支持する具体的なコメントが得られた。

- **動作の客観視と気付きについて**

「自分の全体像を客観的に見ることで全然ないので良かった。言葉で説明されるより自分で見た方がどこ直せばいいか分かりやすい」や、「3D モデルとして自分の動きを第三者視点で見ることは、自分の無意識な動きに気づくきっかけとなり、癖を修正するための用途としても使える」といった感想が挙げられた。これより、三人称視点での振り返りが自己の動作認識に有効であることが示された。

- **空間的・立体的な理解について**

「3D で表示されるお手本が映像で見るよりも、遥かに体の運び方や腕の構え方がわかりやすかった」や、「VR でみることによってゴルフスイングの” 奥行き” 部分を

体系的に理解できた」との回答が得られた。これらは、平面映像では捉えにくい奥行き情報の提示において、本システムが優位性を持つことを示している。

第 5 章

議論

本研究では、お手本動画とユーザーの模倣動画に対し、時間・空間アライメントを適用し、その結果を VR 空間内でフィードバックとして提示するシステムを提案した。評価実験の結果、定量的評価 (表 4.2) においてはゴルフスイングの統計的に有意な改善効果は確認されなかったものの、定性的評価 (表 4.3) においてはシステムの有用性が示唆された。一方で、骨格推定の精度や、お手本とユーザーの動きの正確な比較には課題が残る結果となった。また、システムの操作性に関しても改善の余地があり、より直感的なユーザーインターフェースや補助機能の導入が求められる。次節 (5.1 から 5.5) では、これらの実験結果に基づき、提案システムの課題とその要因を考察し、解決策について述べる。

5.1 定量的評価に対する考察

本実験の定量的評価において、システム使用前後での RMSD に統計的な有意差は認められず、一部の被験者では RMSD が増加する結果となった。

この要因として、新しい動作フォームを身体に定着させるための練習時間が不足していた点が挙げられる。本実験における VR 空間でのフィードバック時間は 4~5 分程度であり、数回のスイングを確認する程度に留まった。ゴルフスイングのような全身を使う複雑な動作において、既存のゴルフスイングを修正し、新しい動作を定着させるためには、十分な反復練習と時間が必要不可欠である。また、被験者はシステムからのフィードバックによって「どう動くべきか」という頭での理解はできたものの、それを実際の動きとして再現し、安定させるまでの反復回数が足りなかったと考えられる。その結果、修正しようとする意識と実際の身体動作の間に乖離が生じ、数値上の改善 (RMSD の減少) には至らなかったと推察される。

5.2 定性的評価に対する考察

定量的評価の結果とは対照的に、アンケートによる定性的評価では、多くの被験者から「自分の動きの癖がわかった」「手本との違いが理解できた」といった肯定的な回答が得られた。これは、本システムが「即時的な動作矯正」には結びつかなくとも、「動作への気づき」を促すツールとして機能したことを示している。RMSD のような定量的な評価には表れない微細な意識の変化や、身体部位への注意の向け方の変化をユーザー自身が実感しており、動機づけや学習の初期段階における支援ツールとしての有用性は確認されたといえる。

特に「表示された骨格モデルを客観的な自分として見られたか (Q1)」や「2D に対する VR の優位性 (Q5)」が高く評価された背景には、ゴルフスイングという立体的な運動を観察する上で、VR 技術が適していた点が挙げられる。ゴルフスイングは、上半身を回旋させる複雑な運動であり、単一視点の 2D 映像では、奥行き方向の動きや、身体のねじれを正確に把握することが困難である。本システムは、VR 空間内で視点を自由に変更できることで、ユーザーが「自分を様々な視点から見る」ことを可能にした。これにより、平面映像では見落とされていた奥行き情報の理解に繋がったと考えられる。

一方で、「お手本と自分の区別 (Q2)」に関しては、比較的低く評価された背景として、重ね合わせ表示における視覚情報の混雑が要因として考えられる。本システムでは、ユーザーとお手本の骨格を同一空間上に重ねて表示する手法を採用したが、両者の動作が空間的に重なる局面においては、多数の骨格が入り乱れて表示される。その結果、ユーザーは瞬時に「どちらが自分の骨格か」を判別する際に戸惑いを感じた可能性がある。このことから、今後の改善策として「自分の骨格だけを点滅させる」や「お手本とのコントラストを強める」といった、視覚的な識別性を高める工夫が必要である。

5.3 フィードバック手法に関する考察

実験結果より、フィードバック手法によってユーザーの認知特性が異なることが明らかになった。まず、全体的なバランス把握 (Q11) で最も高評価だった「B: 横並び」は、手本とユーザーのモデルが横並びで表示されているため、モデル同士の重なりによる視認性の悪化が発生しない。これにより、ユーザーは姿勢全体を俯瞰的に捉えることが容易であったと考えられる。対照的に、改善点の発見 (Q10) で優位だった「C: 誤差強調」は、色による強調表示がユーザーの注意を引きつける役割を担っており、即時的なエラー認識を支

援したといえる。一方で、最も写実的な「D: 肉付き骨格」が低評価に留まった点は注目したい。フォーム修正においては、必ずしもリアルさが正解ではなく、棒人間のように情報を適切に抽象化した可視化こそが、本質的な動作理解に寄与することを示唆している。

以上のことから、単一の手法ですべての学習をカバーすることは困難である。理想的なシステムとしては、全体像を把握する「マクロな視点（横並び）」と、細部のズレを修正する「ミクロな視点（重ね合わせ・誤差強調）」を、ユーザーが目的に応じて柔軟に切り替えられる仕組みが有効であると結論付けられる。

5.4 骨格推定精度とアライメントの課題

本システムでは VideoPose3D を用いた単眼カメラからの 3D 骨格推定を採用しているが、ゴルフスイングの動作において、単眼推定の精度に関する課題が 2 点挙げられる。

第一の課題は、「自己遮蔽（オクルージョン）」による推定精度の低下である。ゴルフスイングの「バックスイングのトップ」や「フィニッシュ」といった局面では、腕が頭の後ろに回ったり、身体に巻き付いたりするため、カメラから見て手足が隠れる現象が発生する。VideoPose3D などの手法では、見えていない部位を予測によって補完するが、ゴルフのような複雑な動作ではその予測が外れやすく、関節位置が不自然に飛ぶ現象がいくつか確認された。

第二の課題は、単眼カメラ特有の「奥行きの曖昧性」である。2D 映像から 3D 座標を復元する際、奥行き方向の推定誤差は、RMSD の算出値にノイズとして混入するだけでなく、空間アライメント（Kabsch Algorithm）の計算結果をも不安定にさせる要因となった。

これらの課題に対する解決策として、複数台のカメラを用いる「マルチビュー推定」や、慣性計測ユニットを用いたモーションキャプチャスーツの導入が必要であると考えられる。

5.5 ユーザーインターフェースとフィードバックシステムの課題

実験を通じて、システムの操作性が直感的ではないという課題も浮き彫りになった。現在の UI では、フィードバックを確認するためにコントローラによる操作を行う必要があり、これが練習への集中を削ぐ要因となっていた可能性がある。また、実験中にも数人の被験者から操作面での質問が見受けられた。そのため、ユーザーがスイングの改善のみに

注力できるよう，ユーザーインターフェースの改善や，VR 空間内でリアルタイムに自分の練習中の 3D モデルが表示されるような機能の追加が求められる．

第 6 章

結論

本研究では、スポーツ初学者が高価な計測機器を用いず、単眼カメラ映像のみを入力として、熟達者の動作を効果的に模倣できる学習支援環境の構築を目的とした。本章では、提案手法、実装システム、実験結果、および今後の課題の 4 つの観点から本研究の成果を総括する。

6.1 提案手法について

本研究では、撮影環境や体格の異なるユーザーとお手本動画を、3 次元空間内で正確に比較するための手法を提案した。具体的には、姿勢推定技術 (VideoPose3D) により 2D 映像から 3 次元骨格情報を抽出し、TCC を用いた時間アライメントと、Kabsch Algorithm を用いた空間アライメント、およびスケール正規化処理を行い、お手本動画とユーザー動画の骨格情報を VR 空間での動作比較を可能にした。これにより、高価な計測機器を必要とせず、スマートフォン等の単眼カメラのみで動作比較を実現可能にし、トレーニングのコストを低減する手法を確立した。

6.2 実装システムについて

提案手法に基づき、ユーザーが自身の動作とお手本の動作を VR 空間内で比較・分析できるフィードバックシステムを実装した。本システムは、算出された身体部位ごとの差異に基づき、修正すべき箇所を色で強調表示する機能や、視点を自由に変更できる機能を備えている。特に、フィードバック手法として「重ね合わせ」「横並び」「誤差強調」「肉付き骨格」などの複数のフィードバックモードを実装した。これにより、平面ディスプレイで

は把握することが困難であった「奥行き方向のズレ」や「身体の回旋」を、3 次元的に理解させることを可能にした。

6.3 構築したシステムに基づく実験結果について

被験者 9 名を対象とした評価実験の結果、以下の知見が得られた。定量的評価においては、4～5 分程度の短時間の練習では、ゴルフスイングの物理的な改善（RMSD の統計的に有意な減少）は確認されなかった。これは、ゴルフスイングのような全身を使う複雑な動作において、既存のゴルフスイングを修正し、新しい動作を定着させるためには、十分な反復練習と時間が必要不可欠であると考えられる。

一方で、定性的評価においては、多くの被験者が「自身の癖や手本との違いを明確に認識できた」と回答し、VR 特有の複数視点からの観察が動作理解に寄与することが示された。

以上の結果から、本システムは即時的なパフォーマンス向上ツールとしては課題を残したものの、学習者の「動作への気づき」を促進し、動機づけや学習の初期段階における支援ツールとしての有用性は確認されたといえる。

6.4 今後の課題について

本研究の課題と今後の展望として、以下の 3 点が挙げられる。

第一に、「骨格推定精度の向上」である。現在は単眼カメラのみを使用しているため、ゴルフスイング特有のオクルージョンが発生する局面での推定誤差が課題となった。これに対し、複数台のカメラを用いる「マルチビュー推定」や、慣性計測ユニットを用いたモーションキャプチャスーツの導入が必要であると考えられる。

第二に、「操作性およびフィードバック機能の改善」である。自分とお手本の差を一目で理解できるような補助や、コントローラーを用いた操作の簡易化が求められる。また、フィードバックを即座に動作修正に反映させるため、自身の 3D モデルをリアルタイムに提示する機能の実装など、より直感的なシステムへの改良が必要であると考えられる。

第三に、「長期的効果の検証」である。本実験では 4～5 分程度の短時間のフィードバックであったため、学習の定着までは確認されなかった。今後は、本システムを継続的に使用した場合に、RMSD の有意な減少や、学習効果の定着などが起きるかどうか検証する必要がある。

謝辞

本研究にあたり，主指導教員として，毎週のゼミや個人面談において多大な助言とご指導をいただいた，北陸先端科学技術大学院大学 先端科学技術研究科 創造社会デザイン研究領域 金井 秀明准教授に深く感謝申し上げます。

また，ご多忙である中，実験参加にご協力いただいた北陸先端科学技術大学院大学の学生の皆様に感謝申し上げます。

参考文献

- [1] 水島 宏一. 器械運動のデジタル資料の検討. *スポーツ教育学研究*, 35(1):1–13, 2015.
- [2] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review*, 20(1):21–53, 2013.
- [3] 大加戸 翔真. 動作推定を用いた野球投球動作の習得方法の提案. 修士論文, 北陸先端科学技術大学院大学, 2024.
- [4] Tica Lin, Rishi Singh, Yalong Yang, Carolina Nobre, Johanna Beyer, Maurice A Smith, and Hanspeter Pfister. Towards an understanding of situated ar visualization for basketball free-throw training. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2021.
- [5] Baoheng Zhang, Carol Yx Chen, Louis Cw Chan, and Wilton Fok. Intelligent sports performance scoring and analysis system based on deep learning network. In *2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 17–21, 2020.
- [6] Nobuyuki Umezu and Souta Akiyama. Visualizing difference between bodily movements of athletes and users learning to play baseball. In *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pages 252–254, 2022.
- [7] Mai Geisen, Florian Seifriz, Frowin Fasold, Michal Slupczynski, and Stefanie Klatt. A novel approach to sensor-based motion analysis for sports: Piloting the kabsch algorithm in volleyball and handball. *IEEE Sensors Journal*, 24(21):35654–35663, 2024.
- [8] Yuichiro Hiramoto, Mohammed Al-Sada, and Tatsuo Nakajima. Designing a 3d human pose estimation-based vr tennis training system. In *2023 IEEE 29th*

International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), pages 267–268, 2023.

- [9] Wenxuan Guo, Zhiyu Pan, Ziheng Xi, Alapati Tuerxun, Jianjiang Feng, and Jie Zhou. Sports analysis and vr viewing system based on player tracking and pose estimation with multimodal and multiview sensors, 2024.
- [10] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. Temporal cycle-consistency learning, 2019.
- [11] William McNally, Kanav Vats, Tyler Pinto, Chris Dulhanty, John McPhee, and Alexander Wong. Golfdb: A video database for golf swing sequencing, 2019.
- [12] P A Hume, J Keogh, and D Reid. The role of biomechanics in maximising distance and accuracy of golf shots. *Sports Medicine*, 35(5):429–449, 2005.
- [13] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 3d human pose estimation in video with temporal convolutions and semi-supervised training, 2019.
- [14] 川崎 仁史, 脇坂 崇平, 笠原 俊一, 齊藤 寛人, 原口 純也, 登嶋 健太, and 稲見 昌彦. けん玉できた！ vr：5 分間程度の vr トレーニングによってけん玉の技の習得を支援するシステム. In **エンタテインメントコンピューティングシンポジウム 2020 論文集**, volume 2020, pages 26–32. 情報処理学会, 2020.