ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE AND MANAGEMENT

УДК 004.89
https://doi.org/10.23947/2687-1653-2025-25-3-221-232

Оригинальное эмпирическое исследование

Подход к реконструкции модели тела на основе ограниченного набора данных о двигательной активности рук

А.Д. Обухов ✉, **Д.В. Теселкин**
Тамбовский государственный технический университет, г. Тамбов, Российская Федерация
✉ obuhov.art@gmail.com

EDN: HLYDVW

Аннотация

Введение. Точная реконструкция модели тела человека крайне важна для визуализации цифровых аватаров в виртуальных тренажерах и реабилитационных системах. Однако использование экзоскелетных систем может привести к перекрытию и экранированию датчиков, что затрудняет работу систем отслеживания. Это подчеркивает актуальность задачи реконструкции модели тела человека на основе ограниченного набора данных о движениях рук, как в сфере реабилитации, так и в спортивной подготовке. Существующие исследования сосредоточены либо на масштабных IMU-сетях, либо на полном видеоконтроле, не рассматривая вопрос реконструкции модели тела на основе данных о движениях рук. Цель данной работы заключается в разработке и тестировании методов машинного обучения, направленных на восстановление координат модели тела с использованием ограниченных данных, например, информации о положении рук.

Материалы и методы. Для проведения исследования была сформирована виртуальная имитационная среда, в которой виртуальный аватар выполнял различные движения. Эти движения фиксировались камерами с видом от первого лица и боковой. В качестве эталонных данных сохранялись положения ключевых точек модели тела относительно точки спины. Рассматривалась задача регрессии, целью которой было восстановление положения рук пользователя в полной модели его тела в пяти различных вариациях, включающих координаты ключевых точек, извлеченные из видео и виртуальной сцены. Задача также подразумевала сравнение различных моделей регрессии, среди которых были линейные модели, деревья решений, ансамбли, а также три глубокие нейронные сети (DenseNN, CNN-GRU, Transformer). Точность оценивалась с использованием MAE и среднего Евклидова отклонения сегментов тела. Проведены экспериментальные исследования на пяти наборах данных, размер которых варьировался от 25 до 180 тысяч кадров.

Результаты исследования. Эксперименты показали, что ансамбли (LightGBM) наиболее эффективны в большинстве ситуаций. Среди нейросетевых моделей наименьшую погрешность обеспечила модель на базе CNN-GRU. Обучение моделей на последовательности из 20 кадров не дало значительного улучшения. Применение модуля инверсной кинематики на ряде сценариев позволяет снизить погрешность до 3 %, но в ряде случаев ухудшает итоговый результат.

Обсуждение. Анализ полученных результатов показал низкую точность реконструкции при использовании наборов данных от компьютерного зрения, а также отсутствие превосходства сложных моделей перед более простыми ансамблями и линейными моделями. Тем не менее, обученные модели позволяют с некоторой погрешностью восстанавливать положение ног пользователя для более достоверного отображения цифровой модели его тела.

Заключение. Полученные данные показывают сложность решения задачи реконструкции модели тела человека при использовании ограниченного объема данных, а также большую погрешность у ряда моделей машинного обучения. Сравнение моделей на различных наборах данных показало низкую применимость данных от первого лица, не содержащих информацию о расстоянии до рук. С другой стороны, использование в качестве входной информации абсолютных значений положения рук позволяет осуществить реконструкцию модели тела со значительно меньшей погрешностью.

Ключевые слова: реконструкция модели тела человека, машинное обучение, виртуальные тренажеры, ограниченные данные

Благодарности. Авторы благодарят руководителя научного проекта М.Н. Краснянского, доктора технических наук, профессора, ректора Тамбовского государственного технического университета за организацию научно-исследовательского процесса.

Финансирование. Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках проекта «Разработка иммерсивной системы взаимодействия с виртуальной реальностью для профессиональной подготовки на основе всенаправленной платформы» (124102100628-3).

Для цитирования. Обухов А.Д., Теселкин Д.В. Подход к реконструкции модели тела на основе ограниченного набора данных о двигательной активности рук. *Advanced Engineering Research (Rostov-on-Don)*. 2025;25(3):221–232. https://doi.org/10.23947/2687-1653-2025-25-3-221-232

Original Empirical Research

Reconstructing a Full-Body Model from a Limited Set of Upper-Limb Motion Data

Artem D. Obukhov ✉, **Daniil V. Teselkin**
Tambov State Technical University, Tambov, Russian Federation
✉ obuhov.art@gmail.com

Abstract

Introduction. Accurate reconstruction of the human body model is required when visualizing digital avatars in virtual simulators and rehabilitation systems. However, the use of exoskeleton systems can cause overlapping and shielding of sensors, making it difficult for tracking systems to operate. This underlines the urgency of the task of reconstructing a human body model based on a limited set of data on arm movements, both in the field of rehabilitation and in sports training. Existing studies focus on either large-scale IMU networks or full video monitoring, without considering the issue of reconstructing a body model based on arm motion data. The objective of this research is to develop and test machine learning methods aimed at reconstructing body model coordinates using limited data, such as arm position information.

Materials and Methods. To conduct the study, a virtual simulation environment was created in which a virtual avatar performed various movements. These movements were recorded by cameras with a first-person and side view. The positions of the keypoints of the body model relative to the back point were saved as reference data. The regression task considered was to reconstruct the user's arm positions in a full body model in five different variations, including keypoint coordinates extracted from a video and a virtual scene. The task also involved comparing different regression models, including linear models, decision trees, ensembles, and three deep neural networks (DenseNN, CNN-GRU, Transformer). The accuracy was estimated using MAE and the mean Euclidean deviation of body segments. Experimental studies were conducted on five datasets, whose size varied from 25 to 180 thousand frames.

Results. The experiments showed that ensembles (LightGBM) were best-performing in most situations. Among neural network models, the CNN-GRU-based model provided the lowest error. Training models on a sequence of 20 frames did not give significant improvement. Using the inverse kinematics module on a number of scenarios allowed reducing the error to 3%, but in some cases worsened the final result.

Discussion. The analysis of the results obtained showed low reconstruction accuracy when using computer vision datasets, as well as the lack of superiority of complex models over simpler ensembles and linear models. However, the trained models allowed, with some error, for the reconstruction of the position of the user's legs for a more reliable display of the digital model of his body.

Conclusion. The data obtained showed the complexity of solving the problem of reconstructing a human body model using a limited amount of data, as well as a large error in a number of machine learning models. The comparison of models on different datasets proved low applicability of first-person data that did not contain information on the distance to the arms. On the other part, using absolute values of arm positions as input information provided for the reconstruction of the body model with significantly less error.

Keywords: reconstruction of the human body model, machine learning, virtual simulators, limited data

Acknowledgements. The authors would like to thank the head of the scientific project, M.N. Krasnyansky, Dr.Sci. (Engineering), Professor, Rector of TSTU, for organizing the research process.

Funding Information. The research is done with the financial support from the Ministry of Education and Science of the Russian Federation within the framework of the project “Development of an Immersive Virtual Reality Interaction System for Professional Training Based on an Omnidirectional Platform” (124102100628-3).

For Citation. Obukhov AD, Teselkin DV. Reconstructing a Full-body Model from a Limited Set of Upper-Limb Motion Data. *Advanced Engineering Research (Rostov-on-Don)*. 2025;25(3):221–232. https://doi.org/10.23947/2687-1653-2025-25-3-221-232

Введение. Виртуальные тренажеры, интегрированные с управляемыми экзоскелетами, позволяют моделировать физические нагрузки и реабилитационные упражнения в контролируемой среде [1]. Для достижения максимального эффекта погружения необходимо точное отслеживание всей кинематики тела пользователя, чтобы сформировать виртуальный аватар, соответствующий реальным движениям человека. Однако верхние экзоскелеты могут перекрывать датчики, требующие прямого визуального контроля (например, HTC Vive Tracker), а также создавать электромагнитные помехи и ограничивать обзор внешних камер, что требует использования дополнительных маркеров [2]. В таких условиях традиционные системы отслеживания, такие как инфракрасные маркеры и множественные камеры захвата, не могут постоянно обеспечивать полный набор данных о положении всех сегментов тела [3]. Таким образом, перед исследователями стоит задача реконструкции полной модели тела на основе ограниченной информации, например, данных о положении рук или кистей.

Одним из возможных решений является использование носимых датчиков, таких как инерциальные измерительные устройства (IMU). Однако для полной реконструкции модели тела требуется достаточное количество сенсоров (не менее 11, а часто и до 18 элементов) [4]; при уменьшении числа сенсоров точность данных резко снижается. В то же время технологии компьютерного зрения активно развиваются и все чаще применяются в системах виртуальной реальности для отслеживания рук и пальцев, что делает задачу реконструкции модели тела по ограниченным данным, полученным только от рук пользователя, особенно актуальной [5].

Задача восстановления полной скелетной модели человека по ограниченному набору визуальных данных (например, движений рук) имеет значительное практическое и научное значение. В классических маркерных системах захвата движения использовались инфракрасные камеры и маркеры на суставах, а для безмаркерных решений разрабатываются и успешно внедряются алгоритмы на основе компьютерного зрения. Современные сверточные нейронные сети, такие как OpenPose, BlazePose и MediaPipe Pose, способны обнаруживать 2D положение ключевых точек тела без дополнительных меток [6]. Эти методы эффективно определяют видимые точки (руки, плечи, таз и т. д.), однако без глубинной информации от одиночной камеры расстояние до тела не восстанавливается, что затрудняет полную 3D-реконструкцию модели тела. Решение данной проблемы может быть найдено с помощью применения стереокамер и методов триангуляции [7]. Используя такие подходы, современные модели (например, MediaPipe Pose) могут отслеживать до 33 ключевых точек с погрешностью порядка 1–2 см, что позволяет в реальных условиях получать 3D-координаты основных суставов (например, кистей, локтей, коленей), комбинируя данные с нескольких камер и минимизируя ошибку проекции. Тем не менее, такие системы отслеживания часто оказываются непригодными, если камеры видят лишь руки, и нужно оценивать остальную часть скелета на основании движений кистей без прямого визуального контроля. Это крайне актуально в системах виртуальной реальности, где камеры присутствуют только на шлеме и фиксируют в рабочей зоне лишь кисти пользователя. В связи с этим необходимо рассмотреть существующие подходы к решению данной проблемы.

Основное направление работы над данной задачей включает использование методов на основе регрессии или нейронных сетей, которые способны дополнять позу, опираясь только на частичные данные [4]. Например, регрессионные модели, обученные на видеопарах с частично замаскированными телами и руками, могут восстанавливать отсутствующие части тела в сложных условиях [8, 9]. Это указывает на то, что современные модели действительно способны выводить полную позу тела по частичной визуальной информации о руках. В других исследованиях используются нейросетевые архитектуры, ориентированные на последовательность движений, такие как рекуррентные сети (LSTM/GRU) и особенно Transformer [10, 11]. Например, в работе [12] описывается AvatarPoser — модель на базе трансформера, которая прогнозирует полную 3D-позу тела (включая ноги и туловище) по положению головы и рук. Эта система извлекает глубокие признаки из поступающих сигналов движений и разделяет глобальное перемещение тела и локальные ориентации суставов. Для точного согласования позы также производится оптимизация конечностей с использованием метода обратной кинематики [12]. Более того, идея улучшения устойчивости предсказаний при отсутствии видимости реализована в модели EgoPoser, которая также опирается на механизмы Transformer для учета прерывистых данных о движении рук, обеспечивая стабильные предсказания [13]. Стоит отметить, что обучение таких моделей требует разметки полных поз, что приводит к необходимости использовать большие датасеты, такие как Human3.6M, CMU MoCap/AMASS, MPI-INF-3DHP и другие, где имеется синхронизированное видео и 3D-скелет [14, 15]. Однако существующих датасетов, сопоставляющих вид от первого лица с полной моделью тела, недостаточно, что делает задачу сбора и сопоставления таких данных весьма актуальной. Формирование такого датасета можно организовать, воссоздавая движения человека в виртуальной сцене, где можно гибко настроить положение виртуальных камер для записи видео и получить точные координаты точек тела с необходимой частотой [4].

В данном исследовании основным предметом для внедрения полученных научных результатов рассматривается система виртуального тренажера на основе верхнего управляемого экзоскелета. Современные модели VR-шлемов ориентированы на позиционирование по камерам, предполагая, что основной источник информации о движении рук будет поступать от встроенной камеры шлема, используемой для распознавания рук. Кроме того, для расширения экспериментальной базы и выявления закономерностей в движениях человека предполагается наличие внешней системы покадрового захвата, фиксирующей положение тела пользователя в целом. Целью данной работы является

разработка и тестирование методов машинного обучения для восстановления координат тела на основе частичных данных о положении рук. Завершая исследование, планируется сравнение как классических регрессионных моделей, так и нейросетевых, включая современные архитектуры, основанные на механизмах внимания, что позволит оценить преимущества каждого подхода в различных экспериментальных сценариях.

Материалы и методы. Сначала была рассмотрена процедура сбора и первичной обработки данных. Данные собирались в виртуальной среде, где имитировался процесс использования VR-шлема с камерой; все данные (вид от первого лица, вид с боковой камеры) отслеживались виртуальными камерами. Далее видео обрабатывались моделями библиотеки MediaPipe, что позволяло осуществить детекцию рук для выделения 21 ключевой точки ладони, а также извлечь с боковой виртуальной камеры данные о 33 ключевых точках модели тела. Параллельно в виртуальном пространстве фиксировались «истинные» метрические координаты всех сегментов тела (18 ключевых точек стандартной цифровой модели аватара, заданной в игровом движке Unity), включая точки положения рук. Эти реальные координаты (эталон) формировали целевой набор Y для большинства сценариев. Использование виртуальной камеры позволяло обойти ограничения с физическими сенсорами и получить эталонную информацию о позе тела. MediaPipe был выбран в качестве основного фреймворка трекинга рук благодаря модульной системе графов обработки и готовым ML-моделям (детектор ладони и полная модель тела). Ввели сокращение «CV» для тех данных, что были получены в ходе обработки компьютерным зрением и моделями MediaPipe (обозначим как cx_i, cy_i, cz_i), а под «эталоном» понимали метрические координаты точек тела (обозначим как vx_i, vy_i, vz_i), записанные в виртуальной сцене относительно спины пользователя.

Далее рассмотрели процедуру подготовки данных для различных сценариев регрессии. Для анализа моделей машинного обучения и их возможностей были сформированы пять наборов данных (экспериментов), различающихся тем, какие признаки X использовались и какие целевые переменные Y предсказывались:

1) Набор 1 «Руки (вид от первого лица) → Руки (эталон)»: $X = \{(cx_i, cy_i, cz_i)\} \in \mathbb{R}^{63}$, $i = 1–21$ — координаты ключевых точек рук при виде от первого лица (63 значения), $Y = \{(vx_i, vy_i, vz_i)\} \in \mathbb{R}^{18}$, $i = 1–6$ — метрические координаты тех же точек рук (18 значений).

2) Набор 2 «Руки (вид от первого лица) → Тело (эталон)»: $X = \{(cx_i, cy_i, cz_i)\} \in \mathbb{R}^{63}$, $i = 1–21$ — координаты рук (полученные из вида от первого лица, 63 значения), $Y = \{(vx_i, vy_i, vz_i)\} \in \mathbb{R}^{54}$, $i = 1–18$ — метрические координаты всех точек тела (54 значения). Таким образом осуществляется полная реконструкция тела по данным рук.

3) Набор 3 «Руки (вид от первого лица) → Тело (CV)»: $X = \{(cx_i, cy_i, cz_i)\} \in \mathbb{R}^{63}$, $i = 1–21$ — координаты рук (вид от первого лица на основе CV, 63 значения), $Y = \{(vx_i, vy_i, vz_i)\} \in \mathbb{R}^{99}$, $i = 1–33$ — координаты 33 точек тела из дополнительного видео сбоку (99 значений). Отличается от предыдущей задачи тем, что регрессии осуществляются исключительно по CV данным.

4) Набор 4 «Тело (CV) → Тело (эталон)»: $X = \{(cx_i, cy_i, cz_i)\} \in \mathbb{R}^{99}$, $i = 1–33$ — координаты точек тела (вид с боковой камеры, 99 значений), $Y = \{(vx_i, vy_i, vz_i)\} \in \mathbb{R}^{54}$, $i = 1…18$ — метрические координаты всех точек тела (54 значения). Задача состоит в проверке точности прямого преобразования данных от CV в метрические величины 18 ключевых точек.

5) Набор 5 «Руки (эталон) → Тело (эталон)»: $X = \{(vx_i, vy_i, vz_i)\} \in \mathbb{R}^{18}$, $i = 1–6$ — метрические координаты точек рук (18 значений), $Y = \{(vx_i, vy_i, vz_i)\} \in \mathbb{R}^{54}$, $i = 1–18$ — метрические координаты всего тела (54 значения). От набора 2 отличается тем, что используются только эталонные данные, таким образом, проверяется сам факт реконструкции движения по ограниченному набору точек.

Далее рассмотрим модели, используемые для решения указанных пяти задач регрессии. Архитектуры всех моделей в различных задачах будут схожи; отличия для каждого набора заключаются лишь в размерности входа и выхода. В общей сложности рассматриваются два класса моделей: классические регрессионные модели из библиотеки Scikit-Learn (а также модели XGBoost и LightGBM) и нейросетевые модели на базе фреймворка Keras [16, 17].

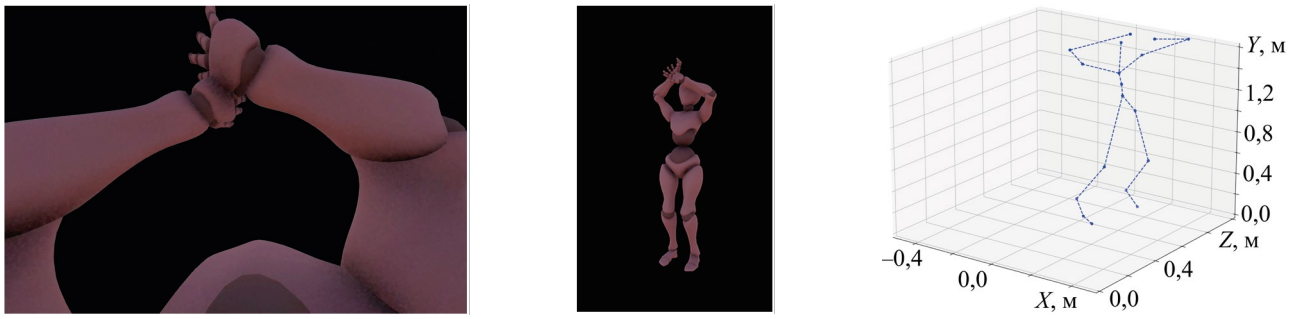
К классическим моделям относятся: линейная регрессия (LinearRegression), ElasticNet (с L1/L2-регуляризацией), ансамбли деревьев (RandomForestRegressor, HistGradientBoostingRegressor), бустинг (XGBRegressor, LightGBMRegressor) и KNN-регрессор. Поскольку целевая переменная включает несколько выходов (координаты точек), модели обернуты в MultiOutputRegressor, что позволяет одновременно предсказывать все параметры. Все древовидные модели настроены на 100 деревьев и глубину, равную 5, в то время как бустинговые модели имеют learning_rate = 0,05.

Далее рассмотрим нейросетевые архитектуры. Полносвязная сеть (обозначим как DenseNN). Входной слой соответствует размерности признаков X (здесь и далее зависит от набора данных), за ним располагаются 4 полносвязных слоя: 256, 512, 1024 и 128 нейронов с активацией ReLU и слоями разреживания Dropout (25 %). Модель завершается выходным слоем размерностью Y (также зависит от набора данных). Используется нормализация пакетов (BatchNorm) и оптимизатор Adam (lr = 1e–3) с функцией потерь MSE.

Сверточно-рекуррентная сеть (CNN-GRU). После входного слоя применяются 1D-свертка (128 фильтров, kernel = 3) и BatchNormalization. Далее следует слой GRU (128 единиц) с возвратом последовательности. Реализован механизм внимания: плотный слой с активацией tanh над выходом GRU выдает веса кадров, которые затем с помощью softmax перемножаются с выходом GRU и суммируются. Затем идут полносвязный слой из 128 нейронов с активацией ReLU и Dropout (30 %), после чего следует выходной слой. Оптимизатор Adam (lr = 1e–3) с функцией потерь MSE.

Трансформер (Transformer). В начале применяются несколько 1D-сверток (kernel = 3, dilation_rate 1, 2 и 4) для создания локального контекста, затем добавляется слой Squeeze-and-Excite для адаптивной фильтрации каналов [18]. Далее вводятся обучаемые позиционные эмбеддинги и 3 энкодерных блока трансформера, в каждом из которых реализована MultiHeadAttention (4 головы, ключ размером 64/4), последующее суммирование и нормализация, а затем двухслойная плотная сеть (размер 256, 64) с Dropout — снова суммирование и нормализация. После энкодеров производится GlobalAveragePooling1D, затем полносвязная прослойка из 128 нейронов с активацией ReLU и Dropout (25 %), после чего следует линейный выход. Оптимизатор Adam (lr = 1e–3) с функцией потерь MSE.

На основании проведенного обзора и имеющегося опыта в данной области можно предложить подход к решению рассматриваемых задач регрессии. Формируется датасет анимации типовых движений человека, который применяется к виртуальному аватару в имитационной сцене. Движения записываются с помощью нескольких виртуальных камер: одна из них расположена на уровне глаз аватара (вид от первого лица), а вторая наблюдает за ним сбоку (боковая камера), охватывая его во весь рост. Дополнительно фиксируются метрические значения 18 точек модели тела. Итоговый вид данных для каждого источника представлен на рис. 1.



а) *б)* *в)*

Рис. 1. Исходные данные: *а* — кадр с камеры от первого лица; *б* — кадр с боковой камеры; *в* — скелет, построенный по эталонным данным

Видеоданные обрабатываются соответствующими моделями (MediaPipe Pose/Hands), после чего координаты точек сохраняются в массивы. Затем в рамках предлагаемого подхода осуществляется обучение моделей машинного обучения, которые на основе одних исходных данных (например, информации о руках) формируют полную 3D-конфигурацию тела. После прогнозирования позы можно дополнительно скорректировать локтевые и коленные суставы, чтобы длины сегментов и положения конечностей лучше соответствовали сигнатурам рук. Также для оценки вклада временного контекста в точность решения задачи реконструкции предлагается провести дополнительный эксперимент по решению задачи регрессии для каждого набора не по данным одного кадра, а некоторой последовательности из N-кадров.

В рамках данной работы не будет сделан акцент на коррекции модели тела после реконструкции на основе правил обратной кинематики. Основная цель заключается в обучении и сравнении набора регрессоров (линейные, древовидные, KNN модели) и нейросетей (DenseNN, CNN-GRU и Transformer) для определения наиболее точной модели. Выбор осуществляется по метрикам средней абсолютной ошибки (MAE), суммарного отклонения (евклидово расстояние) по всем точкам модели от эталонных, а также по оценке вычислительной сложности (времени прогнозирования). Это позволит решить задачу реконструкции модели тела на основе ограниченного набора информации о движениях рук. Кроме того, в рамках экспериментального раздела будут рассмотрены и другие варианты регрессии. Расчет будет производиться по следующим формулам:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

$$\Delta = \frac{1}{MJ}\sum_{n=1}^{M}\sum_{j=1}^{J} ||y_{n,j} - \hat{y}_{n,j}||, \quad ||v|| = \sqrt{v_x^2 + v_y^2 + v_z^2},$$

где y_i — истинное значение; $\hat{y}_i$ — прогноз модели, N — число сравниваемых значений, M — число кадров; J — число суставов (ключевых точек), $y_{n,j}, \hat{y}_i \in \mathbb{R}^3$ — истинный и предсказанный 3-D вектор позиции j-го сустава в n-м кадре.

Результаты исследования. В соответствии с описанной методикой был осуществлён сбор данных по 11 типам различной сложной анимации, включая перемещения тела, прыжки и активные движения. Девять типов использовались для обучения, а два — для валидации (данные из них не участвовали в процессе обучения). Общий объём составил 239 968 записей, однако на каждом этапе проводилась фильтрация и отбор записей в случае, если один из источников не возвращал корректные значения (чаще всего это касалось получения координат рук с

помощью компьютерного зрения). Таким образом, для наборов 1–3 было отобрано 25 и 8 тысяч записей для тренировки и валидации, для наборов 4 и 5 — 183 и 56 тысяч соответственно. В процессе обучения тренировочная выборка была дополнительно разделена в соотношении 75/25. Размерность данных по каждому эксперименту была указана выше при описании соответствующих наборов. На рис. 2 представлены сравнительные результаты всех моделей по всем наборам данных по метрике MAE, на рис. 3 — по метрике суммарного отклонения, на рис. 4 — сравнение моделей по времени вычисления одного прогноза. Далее сравним полученные результаты.

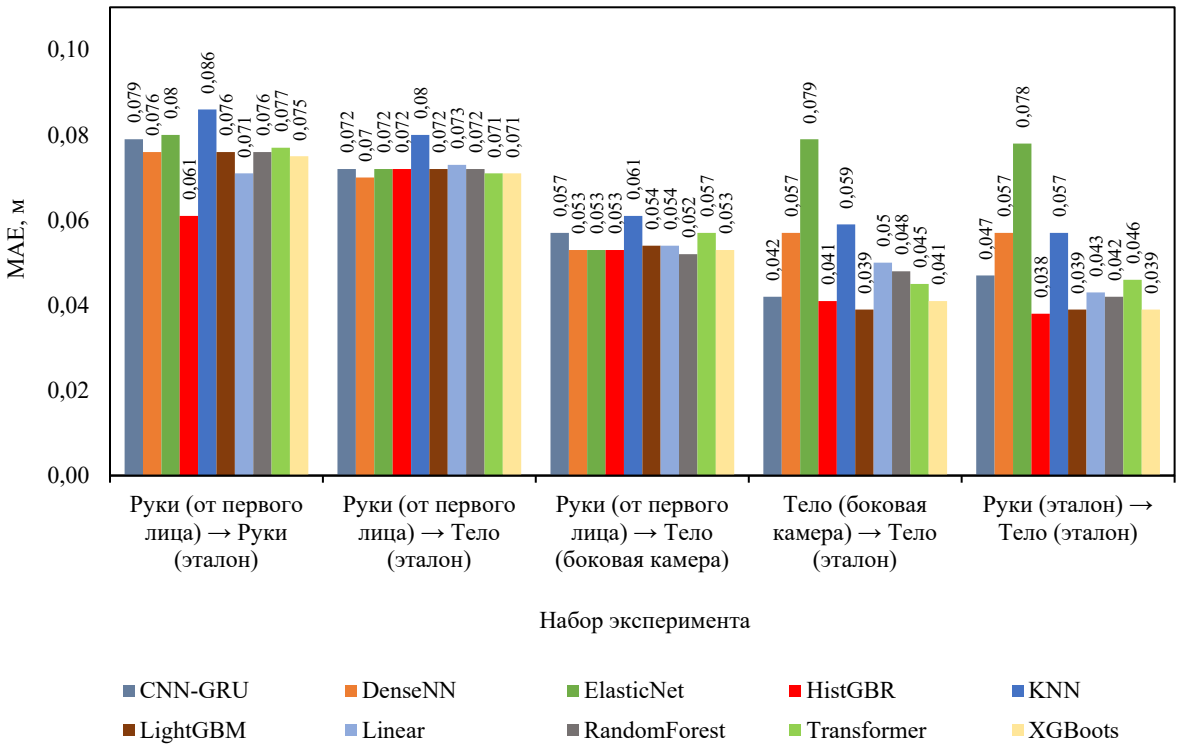


Рис. 2. Сравнение моделей по метрике MAE

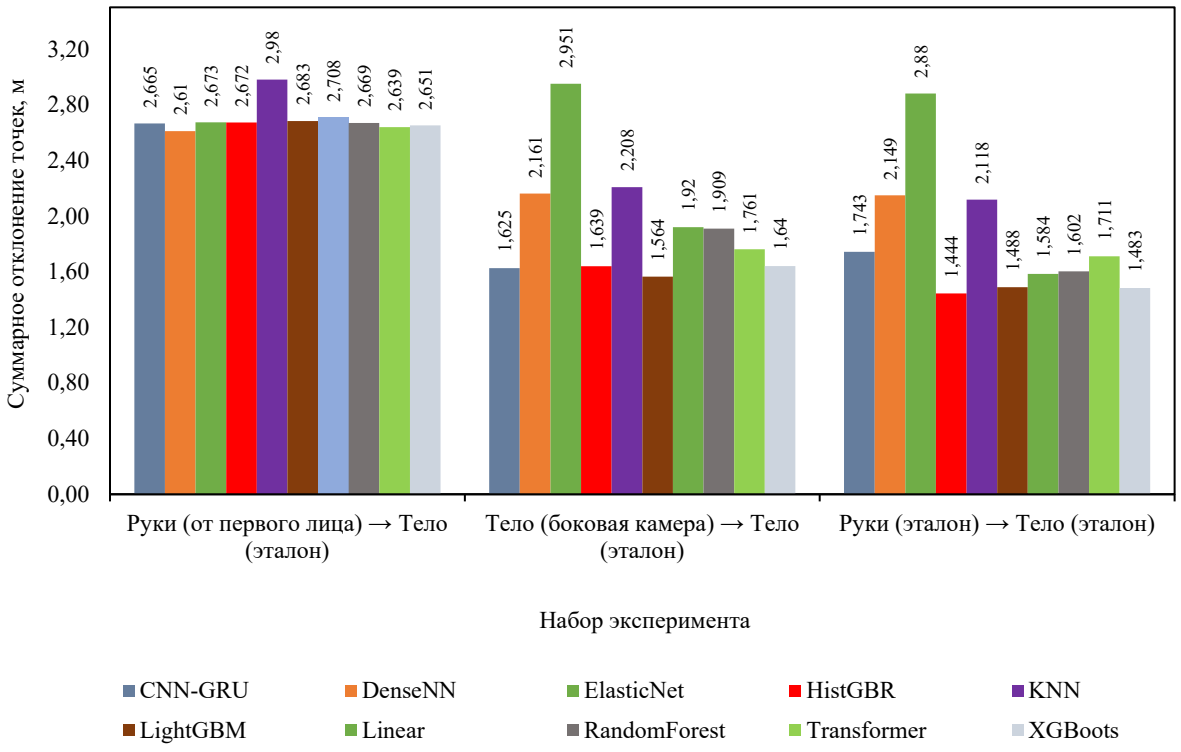


Рис. 3. Сравнение моделей по суммарному отклонению

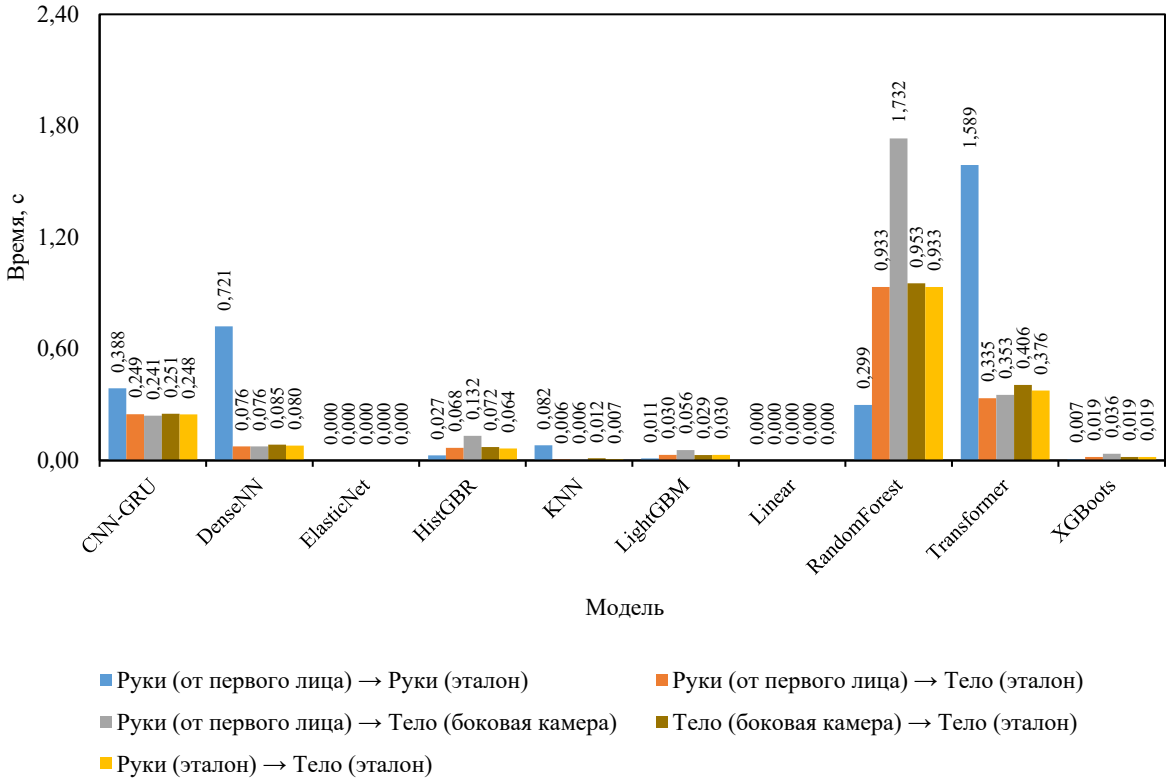


Рис. 4. Сравнение моделей по производительности

Анализ полученных данных показал неоднородность поведения моделей при смене источника входной информации и на различных метриках. В большинстве сценариев наименьшую погрешность по MAE демонстрируют градиентные ансамбли (HistGBR, LightGBM, XGBoost и RandomForest). Нейросетевые модели показывают себя хуже, особенно если учесть сложные задачи реконструкции тела на основе данных о руках (CV). Тем не менее, если оценивать все модели по MAE, то однозначного лидера выделить не удается. С другой стороны, суммарное отклонение всех точек (рис. 3) значительно проясняет ситуацию при решении трех задач регрессии. Наблюдается превосходство ансамблей, как и в предыдущем случае, но среди нейросетевых моделей наилучшей оказывается CNN-GRU. Полученные значения суммарного отклонения, находящиеся в диапазоне от 1,4 до 3,5 метров, свидетельствуют о низкой эффективности решения задачи регрессии всеми моделями, особенно на наборе «Руки (вид от первого лица) → Тело (эталон)». Оценивая производительность моделей по времени вычисления, можно отметить, что классические модели машинного обучения (линейные и ансамбли) обладают достаточной производительностью для использования в режиме реального времени. В то же время CNN-GRU, Transformer и особенно Random Forest крайне затратны по вычислениям, что делает их применимыми только в офлайн-системах (не реального времени). Для DenseNN часто наблюдается длительный расчет при первом вызове модели.

Принимая во внимание имеющийся опыт в задачах реконструкции тела, важно оценивать модели не только по указанным метрикам, но и визуально. Для этого будет осуществлена реконструкция скелетов тела по наборам 2, 4 и 5 с использованием моделей LightGBM и CNN-GRU. Данное сравнение (рис. 5) позволит оценить, как наиболее точная архитектура (LightGBM) визуально отличается от более сложной (CNN-GRU).

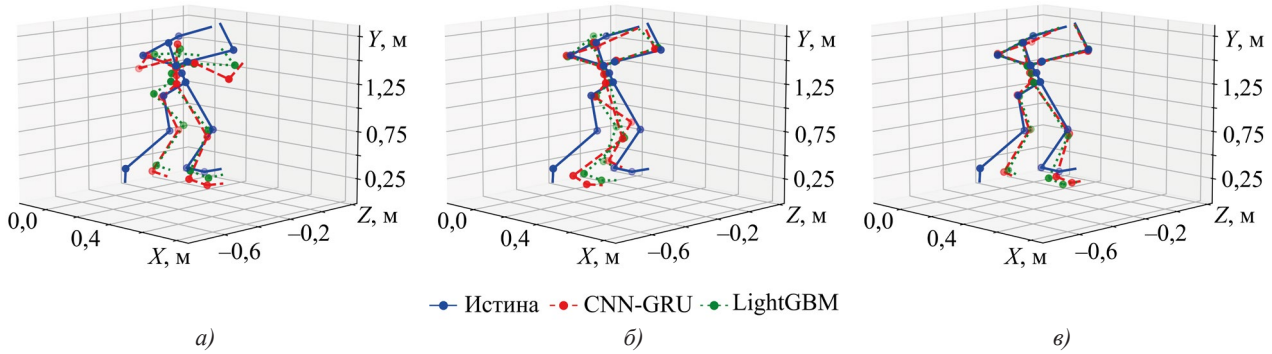


Рис. 5. Визуальное сравнение моделей CNN-GRU и LightGBM: *а* — на наборе «Руки (вид от первого лица) → Тело (эталон)»; *б* — на наборе «Тело (CV) → Тело (эталон)»; *в* — на наборе «Руки (эталон) → Тело (эталон)»

Визуальное сравнение демонстрирует, что между данными CV и реальным положением существует заметная разница, поскольку камера от первого лица не способна точно определить реальную глубину и расстояние до рук. Это приводит к приближенному расположению верхних частей тела (первый график — рис. 5). При использовании данных всего тела CV также наблюдается значительная погрешность, хотя поза в известной степени совпадает. Третий набор, основанный на данных о руках из эталона (что может быть достигнуто путём извлечения координат контроллеров виртуальной реальности или датчиков абсолютного положения), показывает, что верхняя часть тела реконструируется достаточно точно, тогда как ноги — лишь приблизительно, с большой погрешностью. Таким образом, для всех трёх наборов и обеих моделей можно говорить лишь о приближенной реконструкции, что в целом соответствует результатам метрик суммарного отклонения на рис. 3.

Далее был проведён эксперимент по обучению перечисленных моделей не на единственном кадре, а на последовательности из 20 кадров. Это позволяет выявить некоторые динамические характеристики и увеличить объём исходной информации. Поскольку определяющей метрикой, как показало визуальное сравнение, является суммарное отклонение, рассмотрим только его (рис. 6). В целом, использование последовательности кадров незначительно снизило суммарное отклонение; некоторые модели даже показали худшие результаты. С визуальной точки зрения (рис. 7) наблюдается определённое улучшение у модели LightGBM, когда качество восстановления значительно возросло, даже при реконструкции тела на основе данных рук (вид от первого лица). Это касается и двух других наборов данных. Однако для нейросетевой модели в целом значительных улучшений не выявлено.

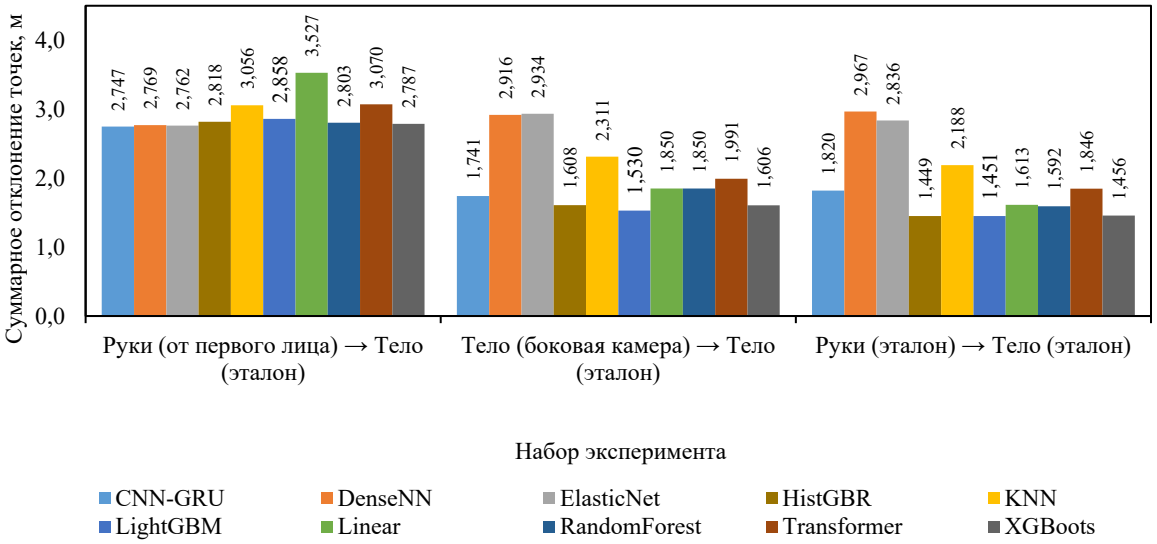


Рис. 6. Сравнение моделей по суммарному отклонению (при обучении на последовательности кадров)

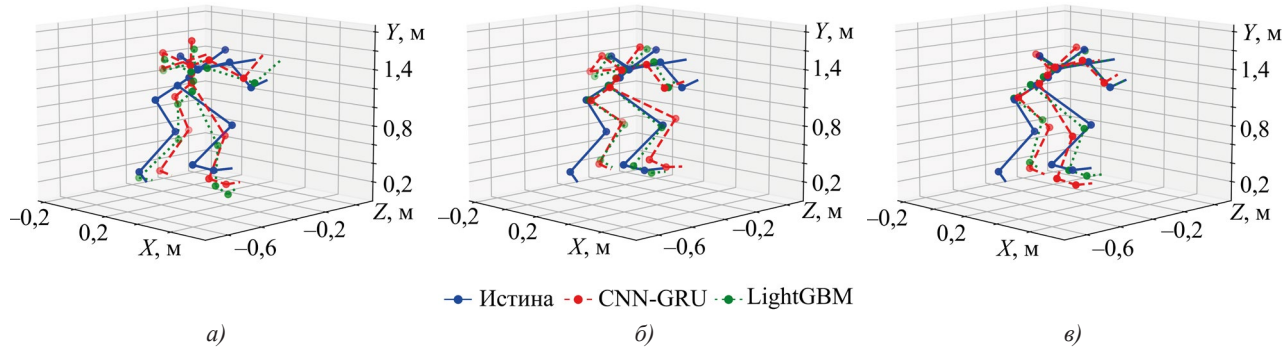


Рис. 7. Визуальное сравнение моделей CNN-GRU и LightGBM (при обучении на последовательности кадров): *а* — на наборе «Руки (вид от первого лица) → Тело (эталон)»; *б* — на наборе «Тело (CV) → Тело (эталон)»; *в* — на наборе «Руки (эталон) → Тело (эталон)»

В завершение эксперимента был проведён опыт по внедрению корректировки точек на основе модели инверсной кинематики (IK). Для этого, после прогнозирования точек тела с использованием моделей машинного обучения, применялся разработанный модуль инверсной кинематики, который сначала корректирует конечные звенья (кисти и стопы) методом FABRIK [19, 20] с учётом угловых ограничений локтей и коленей. Затем модуль перераспределяет возникшие смещения между тазом и грудным отделом, автоматически выравнивая ось позвоночника. Результаты работы данного модуля представлены на рис. 8.

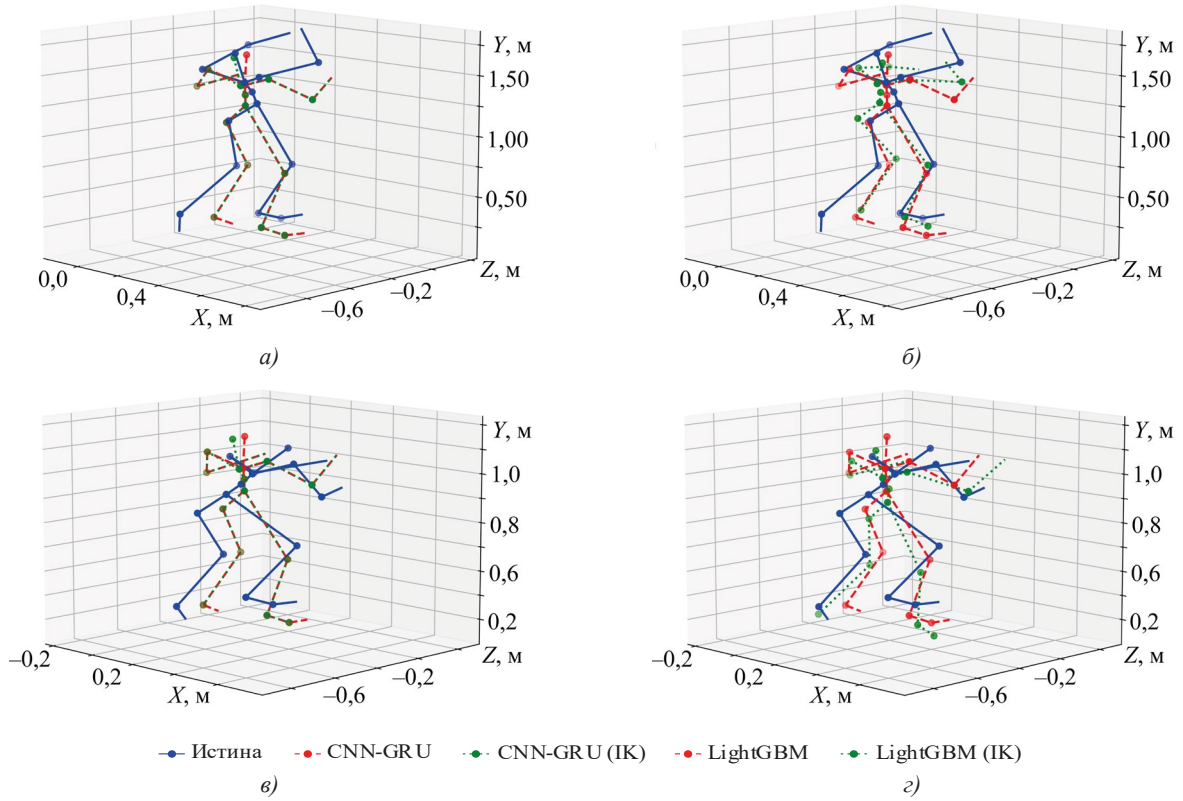


Рис. 8. Визуальное сравнение моделей без и с коррекцией модулем инверсной кинематики (с указанием суммарного отклонения до и после коррекции) на наборе «Руки (вид от первого лица) → Тело (эталон)»: *а* — CNN-GRU (до = 3,511, после = 3,436 м); б — LightGBM (до = 3,183, после = 3,112 м); в — CNN-GRU на последовательности кадров (до = 2,952, после = 2,991 м); г — LightGBM на последовательности кадров (до = 3,261, после = 3,306 м)

Полученные визуализации и численные оценки демонстрируют, что внедрение предложенной двухпроходной инверсной кинематики в целом снижает суммарное евклидово отклонение суставов от эталона для одиночных кадров, однако эффект варьируется в зависимости от типа модели и позиции тела. В первом эксперименте для модели CNN-GRU суммарное отклонение уменьшилось с 3,511 до 3,436 метров, а для LightGBM — с 3,183 до 3,112 метров, что соответствует улучшению примерно на 2–3 %. Графически это проявляется в более естественном выравнивании головы и уменьшении «перегибов» в локтях и коленях. Во втором эксперименте, основанном на 20 кадрах и другой анимации, наблюдается иная картина: для CNN-GRU ошибка возросла с 2,952 до 2,991 метров, а для LightGBM — с 3,261 до 3,306 метров. Замечено, что процедура коррекции стремится выпрямить скелет, что в данном случае только усугубляет ситуацию. Это указывает на то, что геометрические ограничения, примененные постфактум, могут улучшить статическую анатомическую правдоподобность, но в сложной анимации ухудшать текущую позу.

Обсуждение. Проведенное исследование выявило несколько закономерностей. Во-первых, реконструкция полной модели тела на основе ограниченного набора данных возможна, особенно когда исходные и выходные данные получены из одного источника, что подтверждается качественной реконструкцией модели тела по положению рук. Однако выявлены значительные проблемы в восстановлении положения ног пользователя, так как недостаточно информации о движениях рук для прогнозирования сложной анимации. В-третьих, использование положения рук из видеопотока от первого лица, полученное с помощью компьютерного зрения, для реконструкции полной модели тела приводит к высокой погрешности из-за отсутствия данных о расстоянии до рук, имея лишь их положение относительно глаз пользователя. Предварительная обработка данных, смоделированных в виртуальной среде, также показала трудности с распознаванием рук при сложной анимации, что негативно сказалось на процессе обучения.

При сравнении различных архитектур машинного обучения в рамках данной задачи стоит отметить, что более простые линейные модели демонстрируют хорошие результаты в прогнозировании положения сегментов тела, поскольку между исходными и выходными данными существуют четкие зависимости, которые можно аппроксимировать этими моделями. Сложные нейросетевые модели также решают аналогичную задачу, показывая большую гибкость в работе с комплексными входными данными, однако они не отличаются высокой производительностью, а процесс их обучения затратен. В визуальном сравнении нейросетевые модели не продемонстрировали высокой эффективности, показывая результаты, сравнимые или даже худшие.

Проведенный эксперимент показывает, что использование сильно ограниченного по информационной ценности источника данных (информация о положении рук от системы компьютерного зрения является именно таким источником) приводит к значительной погрешности в решении задачи регрессии. Во-первых, объект отслеживания часто выходит из поля зрения и не распознается моделью (это явно видно в снижении объема данных для обучения в наборах 1 и 2). Во-вторых, отсутствие корректных данных о глубине, т.е. расстоянии до рук, затрудняет их абсолютное позиционирование. В системах виртуальной реальности этот аспект нивелируется за счет триангуляции с использованием данных нескольких камер, однако в рамках проведенного моделирования нейросетевая модель для распознавания рук не отражала корректные координаты по оси Z. Потенциальным решением проблемы и темой следующего исследования могло бы стать получение данных непосредственно с шлемов виртуальной реальности, оснащенных интегрированными камерами, что расширило бы обучающую выборку за счет данных о естественных движениях и обеспечило бы более качественный захват рук в виртуальном кадре, поскольку система захвата шлема могла бы возвращать координаты в метрическом пространстве для цифровой модели, предоставляя набор типа 5 («Руки (эталон) → Тело (эталон)»).

Анализируя актуальность исследования в рамках предметной области, следует сравнить его с существующими работами. Главным отличием является ограничение на использование данных о движениях рук, так как более точным подходом считается использование не менее 5 точек тела для дальнейшей реконструкции [21]. Это подтверждают и наши предыдущие исследования [4], в которых оптимальным количеством точек для реконструкции указано не менее 5–7, полученных с использованием эталонной системы отслеживания.

Важно отметить, что во многих VR-приложениях и играх внедряется система отслеживания, основанная только на контроллерах и шлеме, с последующей реконструкцией упрощенного положения тела с использованием алгоритмов инверсной кинематики, что позволяет распространять движение рук на все тело. Как подчеркивают авторы [21], в таких системах одни и те же показания датчиков, расположенных на руках, могут соответствовать множеству разных полных поз, что указывает на необходимость дополнительной настройки инверсной кинематики, чтобы избежать артефактов, а обученные модели должны подбирать правдоподобный вариант. Поэтому сложность задачи без дополнительных источников информации о положении ног или туловища остается высокой. Проведенное исследование подчеркивает данную проблему, указывая на необходимость поиска и сбора дополнительных источников информации для достижения, как минимум, отображения «Руки (эталон) → Тело (эталон)», а в идеале — для распознавания всей траектории движения, что поможет более точно спрогнозировать положение других частей тела. Перспективным направлением здесь может быть использование не только предобученных нейронных сетей (например, MediaPipe), но и захват всей информации об окружающем мире, что позволит лучше сегментировать руки, а, возможно, туловище и ноги пользователя.

Еще одним ограничением проведенного исследования является отсутствие оценки влияния размера обучающей выборки на качество моделей. В данной работе были собраны данные 11 различных типов анимации, для валидации использовались два дополнительных типа, но учитывая объемы и вариативность движений, набор должен быть значительно больше. Тем не менее, исследование ставило целью сравнение моделей в рамках заданной задачи, что продемонстрировало неоднозначность их эффективности по сравнению с классическими линейными моделями и ансамблями. Это также указывает на необходимость дальнейшего улучшения архитектуры моделей.

Наконец, этап коррекции модели тела на основе кинематической модели, реализованный через наложение анатомических ограничений и переоценку позы, дал неоднозначные результаты — на одной позе это снизило суммарное отклонение, а на другой, наоборот, увеличило. С другой стороны, следует учитывать, что модуль инверсной кинематики должен работать с уже искаженными данными о руках и голове в случае набора 1, поэтому переход к более качественному набору данных может снизить погрешность в модуле кинематики.

Заключение. Таким образом, в результате проведенных исследований был разработан подход к прогнозированию модели тела на основе ограниченного набора точек, включающий этапы обработки данных, решения задач регрессии и применения модуля инверсной кинематики для корректировки модели тела. Проведены соответствующие экспериментальные исследования, которые показали, что модели типа LightGBM (среди ансамблей) и CNN-GRU с механизмом внимания (среди нейросетевых моделей) продемонстрировали наилучшие результаты по выбранным метрикам. Сравнение также показало низкую точность реконструкции модели тела при использовании моделей (ElasticNet, KNN, DenseNN), что свидетельствует о их слабой обобщающей способности. В ходе визуального сравнения выявлены противоречия в качестве реконструкции скелета при выполнении сложной анимации, поскольку положение рук недостаточно для определения положения ног и головы. Кроме того, применение коррекции на основе инверсной кинематики не всегда обосновано для сложных поз, так как наложение анатомических ограничений и переоценка позы могут приводить к дополнительным искажениям.

Сравнение разработанных моделей также позволяет сделать выводы о степени их применимости: модели, обученные на наборе данных от первого лица, не позволяют достоверно реконструировать модель тела, показывая высокую визуальную погрешность, что ограничивает их использование только теоретическим сравнением; в то время как модели, обученные на реальных положениях рук (набор 5), показывают более достоверные прогнозы положения тела, что может быть востребовано в виртуальных тренажерах без достаточного набора датчиков. Поскольку модели, обученные на наборе 5, работают с абсолютными положениями рук, это обеспечивает их универсальность при выборе системы отслеживания, так как данные о положении рук могут быть получены не только с помощью системы компьютерного зрения, но и контроллеров виртуальной реальности или инерциальных датчиков, отслеживающих положение рук.

Данное исследование формирует несколько направлений для дальнейшей работы в рамках задачи реконструкции модели тела. Проведенные сравнительные эксперименты моделей машинного обучения показали, что для успешного решения поставленной задачи необходим сбор большего объема информации о движениях человека, расширение датасета и реализация более эффективных моделей обучения с большей обобщающей способностью.

Список литературы / References

1. Tiboni M, Borboni A, Vérité F, Bregoli Ch, Amici C. Sensors and Actuation Technologies in Exoskeletons: A Review. *Sensors*. 2022;22(3):884. https://doi.org/10.3390/s22030884

2. Vélez-Guerrero MA, Callejas-Cuervo M, Mazzoleni S. Artificial Intelligence-Based Wearable Robotic Exoskeletons for Upper Limb Rehabilitation: A Review. *Sensors*. 2021;21(6):2146. https://doi.org/10.3390/s21062146

3. Zihe Zhao, Jiaqi Wang, Shengbo Wang, Rui Wang, Yao Lu, Yan Yuan, et al. Multimodal Sensing in Stroke Motor Rehabilitation. *Advanced Sensor Research*. 2023;2(9):2200055. https://doi.org/10.1002/adsr.202200055

4. Obukhov A, Dedov D, Volkov A, Teselkin D. Modeling of Nonlinear Dynamic Processes of Human Movement in Virtual Reality Based on Digital Shadows. *Computation*. 2023;11(5):85. https://doi.org/10.3390/computation11050085

5. Kuan Cha, Jinying Wang, Yan Li, Longbin Shen, Zhuoming Chen, Jinyi Long. A Novel Upper-Limb Tracking System in a Virtual Environment for Stroke Rehabilitation. *Journal of NeuroEngineering and Rehabilitation*. 2021;18:166. https://doi.org/10.1186/s12984-021-00957-6

6. Jen-Li Chung, Lee-Yeng Ong, Meng-Chew Leow. Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet*. 2022;14(12):380. https://doi.org/10.3390/fi14120380

7. Obukhov AD, Dedov DL, Surkova EO, Korobova IL. 3D Human Motion Capture Method Based on Computer Vision. *Advanced Engineering Research (Rostov-on-Don)*. 2023;23(3):317–328. https://doi.org/10.23947/2687-1653-2023-23-3-317-328

8. Islam MdM, Nooruddin Sh, Karray F, Muhammad G. Human Activity Recognition Using Tools of Convolutional Neural Networks: A State-of-the-Art Review, Data Sets, Challenges, and Future Prospects. *Computers in Biology and Medicine*. 2022;149:106060. https://doi.org/10.1016/j.compbiomed.2022.106060

9. Obukhov A, Dedov D, Volkov A, Rybachok M. Technology for Improving the Accuracy of Predicting the Position and Speed of Human Movement Based on Machine-Learning Models. *Technologies*. 2025;13(3):101. https://doi.org/10.3390/technologies13030101

10. Титаренко Д.Ю., Рыжкова М.Н. Возможности использования нейросетей для распознавания ошибок при выполнении физических упражнений. *Радиотехнические и телекоммуникационные системы*. 2024;(3):62–72. https://doi.org/10.24412/2221-2574-2024-3-62-72

Titarenko DYu, Ryzhkova MN. Possible Neural-Network Use for Error Recognition during Physical Exercising. *Radio Engineering and Telecommunications Systems*. 2024;(3):62–72. https://doi.org/10.24412/2221-2574-2024-3-62-72

11. Hung Le Viet, Han Le Hoang Ngoc, Khoa Tran Dinh Minh, Son Than Van Hong. A Deep Learning Framework for Gym-Gesture Recognition Using the Combination of Transformer and 3D Pose Estimation. *Cybernetics and Physics*. 2024;13(2):161–167. https://doi.org/10.35470/2226-4116-2024-13-2-161-167

12. Jiaxi Jiang, Paul Streli, Huajian Qiu, Andreas Fender, Larissa Laich, Patrick Snape, et al. AvatarPoser: Articulated Full-Body Pose Tracking from Sparse Motion Sensing. In book: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T (eds). *Computer Vision — ECCV 2022*. Cham: Springer; 2022. P. 443–460. https://doi.org/10.1007/978-3-031-20065-6_26

13. Jiaxi Jiang, Paul Streli, Manuel Meier, Christian Holz. EgoPoser: Robust Real-Time Egocentric Pose Estimation from Sparse and Intermittent Observations Everywhere. In book: Leonardis A, Ricci E, Roth S, Russakovsky O, Sattler T, Varol G (eds). *Computer Vision — ECCV 2024*. Cham: Springer; 2024. P. 277–294. https://doi.org/10.1007/978-3-031-72627-9_16

14. Baradel F, Groueix Th, Weinzaepfel Ph, Brégier R, Kalantidis Y, Roges G. Leveraging MoCap Data for Human Mesh Recovery. In: *Proc. IEEE/CVF Conference on 3D Vision (3DV)*. New York City: IEEE; 2021. P. 586–595. https://doi.org/10.1109/3DV53792.2021.00068

15. Seong Hyun Kim, Sunwon Jeong, Sungbum Park, Ju Yong Chang. Camera Motion Agnostic Method for Estimating 3D Human Poses. *Sensors*. 2022;22(20):7975. https://doi.org/10.3390/s22207975

16. Kumar S, Srivastava M, Prakash V. Advanced Hybrid Prediction Model: Optimizing LightGBM, XGBoost, Lasso Regression and Random Forest with Bayesian Optimization. *Journal of Theoretical and Applied Information Technology*. 2024;102(9):4103–4115. URL: https://jatit.org/volumes/Vol102No9/32Vol102No9.pdf (дата обращения: 01.06.2025).

17. Nidhi Dua, Shiva Nand Singh, Vijay Bhaskar Semwal. Multi-Input CNN-GRU Based Human Activity Recognition Using Wearable Sensors. *Computing*. 2021;103(7):1461–1478. https://doi.org/10.1007/s00607-021-00928-8

18. Vosco N, Shenkler A, Grobman M. Tiled Squeeze-and-Excite: Channel Attention with Local Spatial Context. In: *Proc. IEEE/CVF International Conference on Computer Vision Workshops*. New York City: IEEE; 2021. P. 345–353. https://doi.org/10.1109/ICCVW54120.2021.00043

19. Колпащиков Д.Ю., Гергет О.М., Данилов В.В. Сравнение алгоритмов FABRIK обратной кинематики для многосекционных непрерывных роботов. *Известия высших учебных заведений. Машиностроение*. 2022;753(12):34–45. https://doi.org/10.18698/0536-1044-2022-12-34-45

Kolpashchikov DYu, Gerget OM, Danilov VV. FABRIK-Based Comparison of the Inverse Kinematic Algorithms Operation Results for Multi-Section Continuum Robots. *BMSTU Journal of Mechanical Engineering*. 2022;753(12):34–45. https://doi.org/10.18698/0536-1044-2022-12-34-45

20. Lamb M, Lee S, Billing E, Högberg D, Yang J. Forward and Backward Reaching Inverse Kinematics (FABRIK) Solver for DHM: A Pilot Study. In: *Proc. 7th International Digital Human Modeling Symposium*. 2022;7(1):26. https://doi.org/10.17077/dhm.31772

21. Qiang Zeng, Gang Zheng, Qian Liu. DTP: Learning to Estimate Full-Body Pose in Real-Time from Sparse VR Sensor Measurements. *Virtual Reality*. 2024;28(2):116. https://doi.org/10.1007/s10055-024-01011-1

Об авторах:

Артём Дмитриевич Обухов, доктор технических наук, доцент кафедры «Системы автоматизированной поддержки принятия решений» Тамбовского государственного технического университета (392000, Российская Федерация, г. Тамбов, ул. Мичуринская, 112), SPIN-код, ORCID, ResearchGate, ScopusID, ResearcherID, obuhov.art@gmail.com

Даниил Вячеславович Теселкин, ассистент кафедры «Системы автоматизированной поддержки принятия решений» Тамбовского государственного технического университета (392000, Российская Федерация, г. Тамбов, ул. Мичуринская, 112), SPIN-код, ORCID, ResearchGate, ScopusID, dteselk@mail.ru

Заявленный вклад авторов:

А.Д. Обухов: разработка концепции, получение финансирования, проведение исследования, разработка методологии, научное руководство, написание черновика рукописи, написание рукописи, предоставление ресурсов, административное руководство исследовательским проектом.

Д.Л. Теселкин: курирование данных, формальный анализ, разработка программного обеспечения, валидация результатов, визуализация, написание черновика рукописи.

Конфликт интересов: **авторы заявляют об отсутствии конфликта интересов.**

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the Authors:

Artem D. Obukhov, Dr.Sci. (Eng.), Associate Professor of the Department of Automated Decision Support Systems, Tambov State Technical University (112, Michurinskaya Str., Tambov, 392000, Russian Federation), SPIN-code, ORCID, ResearchGate, ScopusID, ResearcherID, obuhov.art@gmail.com

Daniil V. Teselkin, Assistant Professor of the Department of Automated Decision Support Systems, Tambov State Technical University (112, Michurinskaya Str., Tambov, 392000, Russian Federation), SPIN-code, ORCID, ResearchGate, ScopusID, dteselk@mail.ru

Claimed Contributorship:

AD Obukhov: conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, writing – original draft preparation, writing – review & editing.

DV Teselkin: data curation, formal analysis, software, validation, visualization, writing – original draft preparation.

Conflict of Interest Statement: **the authors declare no conflict of interest.**

All authors have read and approved the final version of manuscript.

Поступила в редакцию / Received 24.06.2025
Поступила после рецензирования / Reviewed 20.07.2025
Принята к публикации / Accepted 31.07.2025