

Проблема оценки позы человека: задачи, методы, решения

И. А. Шергин¹, А. П. Рыжов²

В статье рассматривается проблема оценки позы человека, возникающая во многих приложениях систем компьютерного зрения. Рассмотрены различные постановки задач и подходов к решению проблемы, приводится обзор существующих методов решения и наборов данных. Дополнительно исследуется применение методов в задаче обучения физическим навыкам.

Ключевые слова: машинное обучение, глубокое обучение, оценка позы человека, компьютерное зрение.

1. Введение

Компьютерное зрение - это быстро развивающаяся область компьютерных наук. Человеческое тело является одним из самых непредсказуемых элементов для этой области.

Оценка позы человека (*Human pose estimation, HPE*) - это подобласть компьютерного зрения которая, позволяет идентифицировать и прогнозировать положение и ориентацию человеческого тела на изображении или видео. Последние достижения в развитии сверточных нейронных сетей значительно ускорили прогресс в этой области.

Различные процессы жизнедеятельности человека, такие как ходьба, бег, прыжки, танцы, или иные физические упражнения являются объектами изучения систем оценки позы человека - Рис. 1.

Целью данной работы является обзор задач в области Human Pose Estimation, методов их решения, и реализованных систем. Работа организована следующим образом: в разделе 1 приводится общая постановка задачи HPE; в разделе 2 обсуждаются вариации оценки позы человека;

¹Шергин Иван Андреевич — аспирант кафедры интеллектуальных информационных технологий ф-та ВМК МГУ, e-mail: i.a.shergin@gmail.com.

Shergin Ivan Andreevich — graduate student, Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Chair of Intelligent Information Technology.

²Рыжов Александр Павлович — доктор технических наук, профессор кафедры интеллектуальных информационных технологий ф-та ВМК МГУ, e-mail: ryjov@mail.ru.

Ryjov Alexander Pavlovich — Sc.D. in engineering, Ph.D. in mathematics, MBA, professor, Moscow State University, Faculty of Computational Mathematics and Cybernetics, Chair of Intelligent Information Technology.



Рис. 1. Пример моделирования и оценки позы человека

раздел 3 посвящен анализу метрик качества, известных наборов данных и методов решения задач.

Обзор некоторых приложений систем НРЕ с упором на освоение физических навыков представлен в разделе 4; здесь же обсуждается задача персонализации обучения физическим навыкам с использованием методов оценки позы человека.

1.1. Постановка задачи

Целью оценки позы человека является определение расположения всех частей тела человека, учитывая отношения друг с другом и окружающей средой вокруг. Human pose estimation - это процесс выявления и классификации суставов человеческого тела и определения того, являются ли связи между этими суставами корректными. Другими словами, это процесс определения ключевых точек и достоверных пар связей между ними.

Ключевая точка - это 2 или 3-размерная координата в пространстве, описывающая расположение анатомического сустава, а также которая может быть использована для описания позы человека. Эти точки соединяются, образуя пары, при этом парой становится только существующее соединение. Связь, которая образуется между точками, должна быть осмысленной: любые две случайные точки не могут образовать пару.

Таким образом, методам НРЕ необходимо получить представление о человеческом теле подобное скелету, и затем использовать структуру для обработки человеческого тела для конкретных приложений. Именно это дает компьютерам возможность лучше понимать поведение человека.

1.2. Области применения

Уже существует ряд современных решений по применению технологий НРЕ в области фитнеса, реабилитации, физиотерапии, захвата движения, продвинутых игр, робототехники, видеонаблюдения и др.

1.2.1. Персональные тренеры на базе искусственного интеллекта

НРЕ может помочь поддерживать физическое самочувствие и достичь желаемого уровня физической подготовки без хорошего тренера. Пользователи могут открыть мобильное приложение, которое использует камеру для определения позы человека и оценивает, насколько точна поза во время определенного упражнения. Приложение может сообщить пользователю, что было сделано не правильно, и дать рекомендацию, как лучше тренироваться.

Например, Zenia [1] - это приложение для йоги на базе искусственного интеллекта, которое использует GPS, чтобы направлять пользователя к достижению правильной осанки во время занятий йогой.

1.2.2. Реабилитация и физиотерапия

Еще одной областью, где оценка позы человека может быть полезна, является физиотерапевтия. Это одно из наиболее перспективных приложений искусственного интеллекта для оценки позы в секторе здравоохранения. Цифровые консультации на дому повышают гибкость и доступность для пациента, а также производительность терапевта. Технологии ИИ [2] позволяют использовать новые и более сложные способы доставки лечения в цифровом виде.

1.2.3. Захват движения и дополненная реальность

Оценка позы человека была включена в виртуальную и дополненную реальности, например, в игры или фильмы для захвата движения. НРЕ может автоматически извлекать ключевые точки и визуализировать их для создания компьютерной графики. Использование данной технологии в AR и VR позволяет более естественно взаимодействовать с виртуальной средой.

1.2.4. Определение позы спортсмена

В наши дни почти все виды спорта в значительной степени зависят от анализа данных [3]. Определение позы может помочь игрокам улучшить свою технику и добиться лучших результатов. Помимо этого, НРЕ

может быть использовано для анализа и получения информации о преимуществах и недостатках противника.

1.2.5. Робототехника

Роботы являются неотъемлемой частью нашей повседневной жизни и их роль будет только возрастать. Существует множество областей реального мира, где робототехника, основанная на компьютерном зрении, значительно расширила возможности человека, например, медицинские вмешательства, такие как биопсия, автоматическое обнаружение и визуализация опухолей, применение роботов для обучения определенным навыкам и многое другое. С помощью компьютерного зрения и, в частности, оценки позы человека, есть шанс в будущем сделать робототехнические системы более отзывчивыми, гибкими и точными.

2. Вариации оценки позы человека

Оценка позы человека делится на два типа: оценка single-pose используется для оценки поз одного человека, в то время как оценка multi-pose при определении поз для нескольких человек в одном кадре. Последняя является более сложным из-за необходимости успешно идентифицировать каждого человека отдельно друг от друга.

Существует два способа с помощью, которых может быть выполнена оценка позы человека: двумерный (2D) и трехмерный (3D). 2D подход реализует методы извлечения признаков для прогнозирования ключевых точек человеческого тела по изображениям на основе значений пикселей. 3D подход связан с предсказанием пространственного положения конкретного человека или объекта по изображениям и видео аналогично 2D. Данный подход сложнее из-за фоновой сцены, условий освещения или других помех. Кроме того, имеется меньше доступных 3D наборов данных.

2.1. Представление человеческого тела

Существует 2 различных варианта моделирования человеческого тела для описания позы, позволяющие решать различные задачи на основе ее оценки .

2.1.1. Представление на основе ключевых точек

2D или 3D координаты ключевых точек человеческого тела являются простыми и интуитивно понятными представлениями для скелета тела, которые имеют несколько форм представления.

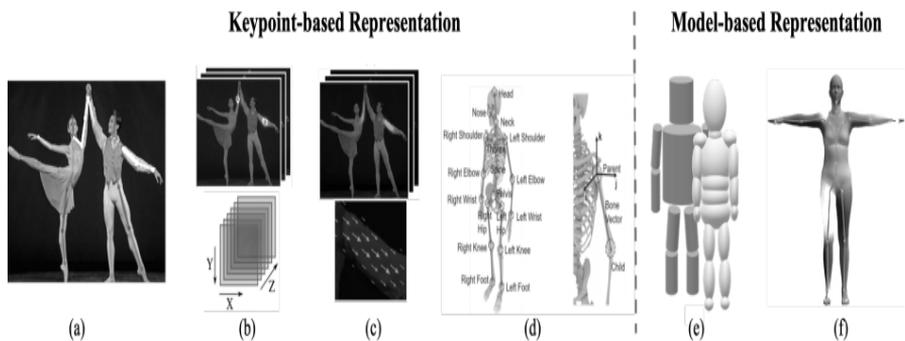


Рис. 2. Формы представления человеческого тела

- **2D/3D координаты ключевой точки.** Ключевые точки тела могут быть явно описаны 2D/3D координатами. Как показано на Рис. 2 (a), ключевые точки соединены в соответствии с присущей корпусу структурой.
- **2D/3D тепловые карты.** Чтобы сделать координаты более подходящими для регрессии сверточной нейронной сетью, многие методы представляют координаты ключевых точек в виде тепловой карты. Как показано на Рис. 2 (b), тепловая карта Гаусса каждой ключевой точки имеет высокое значение отклика на соответствующих 2D/3D координатах и низкое значение отклика в других положениях.
- **Ориентационные карты.** Этот метод использует карту ориентации ключевых точек тела в качестве вспомогательного представления тепловых карт с помощью «part affinity fields» (PAFs) для представления 2D ориентации между конечностями. Как показано на Рис. 2 (c), PAF представляет собой 2D векторное поле, которое связывает две ключевые точки конечности. Каждый пиксель в поле содержит 2D вектор, который указывает от одной конечности к другой.
- **Иерархические векторы моделирования костей.** 2D версия иерархического представления костей представляет собой комбинацию суставов и костных векторов. Она расширяется на трехмерный случай. Как показано на Рис. 2 (d), 3D человеческий скелет представлен набором костных векторов. Каждый вектор кости направлен от родительской ключевой точки к дочерней ключевой точке, следуя кинематическому дереву. Каждая родительская ключевая

точка связана с локальной сферической системой координат. Вектор кости может быть представлен сферическими координатами в этой системе.

2.1.2. Представление на основе модели

Представление на основе модели разработано в соответствии с присущими человеческому телу структурными характеристиками. Оно обеспечивает более информативное представление человеческого тела, по сравнению с предыдущим.

- **Объемная модель на основе деталей.** Объемная модель на основе деталей разрабатываются для решения реальных задач. Как показано на синей модели на Рис. 2 (е), каждая конечность представлена в виде цилиндра, модель была разработана для создания меток прегражденных частей. Каждый цилиндр находится путем совмещения центров верхней и нижней поверхностей с 3D ключевыми точками конечности. Как показано на розовой модели на Рис. 2 (е), предлагается модель эллипсоида, в которой эллипсоид используется в качестве базовой единицы частей тела для большей гибкости.
- **Подробная статистическая модель человеческого тела.** По сравнению с объемной моделью, основанной на деталях, статистическая трехмерная сетка человеческого тела описывает более подробную информацию, включая позу и форму тела.

Human pose estimation выполняется путем определения суставов тела. Различные алгоритмы могут идентифицировать различное количество ключевых точек человеческого тела. На Рис. 3 показан пример доступных ключевых точек.

2.2. Подходы к определению позы человека

2.2.1. «Снизу-вверх» и «сверху-вниз»

Существует 2 подхода к определению позы человека:

- Подходы **снизу-вверх** (bottom-up) сначала определяют каждый сустав тела, а затем группируют их, чтобы сформировать уникальную позу для разных людей, Рис. 4 (а).
- Подходы **сверху-вниз** (top-down) сначала определяют области ограничивающих рамок человека, а затем определяют суставы тела в пределах обнаруженных ограничивающих рамок для каждого человека по отдельности, Рис. 4 (b).

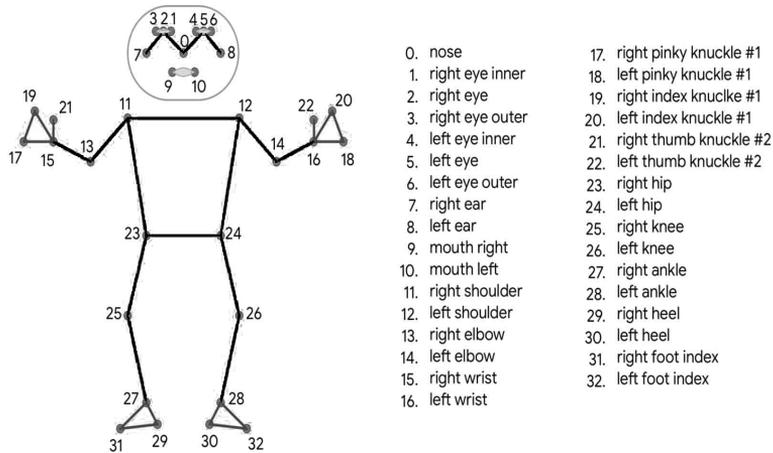


Рис. 3. Пример ключевых точек суставов человеческого тела

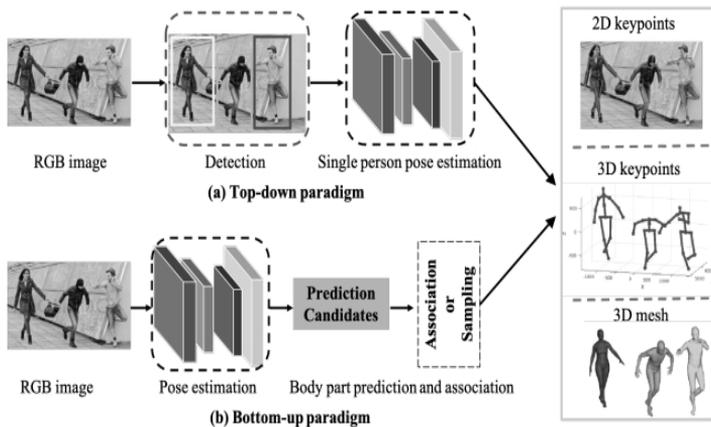


Рис. 4. Типы методов для оценки позы человека

2.2.2. Моделирование человеческого тела

Исследователи выделяют следующие модели человеческого тела:

Кинематическая модель (модель на основе скелета) используется для оценки 2D и 3D поз, Рис. 5 (а). Эта интуитивная модель человеческого тела содержит набор положений суставов и ориентаций конечностей для представления структуры человеческого тела.

Плоская Модель (модель на основе контуров) используется для оценки 2D позы, Рис. 5 (б). Эта модель используется для представления

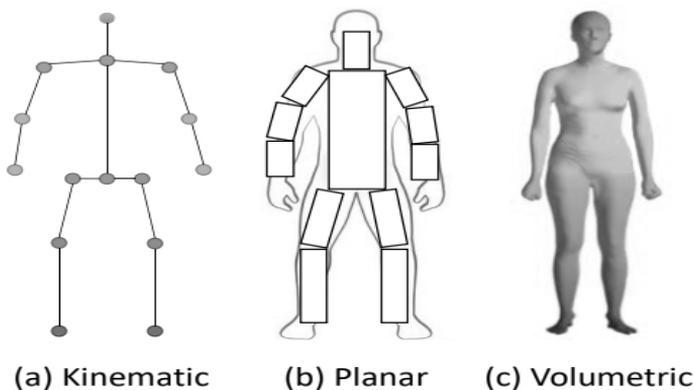


Рис. 5. Типы скелетона

внешнего вида и формы человеческого тела. Части тела представлены несколькими прямоугольниками, приблизительно повторяющими контуры человеческого тела.

Объемная модель используется для оценки позы в 3D, Рис. 5 (с). Эта модель предоставляет соответствующую информацию для оценки позы в 3D с помощью сканирования всего тела для получения сетки человеческого тела различных форм и поз.

3. Инструменты для задачи

3.1. Метрики оценки качества

Существует несколько показателей, используемых для оценки производительности моделей НРЕ:

Intersection Over Union (IoU) - это метрика которая определяет разницу между аннотациями истинными и прогнозируемыми ограничивающими рамками, Рис. 6. Удаляет все ненужные рамки на основе присвоенного порогового значения, которое обычно равно 0.5.

Percentage of Correct Parts (PCP) (процент правильных частей) - это мера, которая оценивает предсказания, чтобы сообщить о точности локализации конечностей. Локализация конечностей определяется, когда расстояние между прогнозируемым суставом и истинным суставом составляет менее доли длины конечности (от 0.1 до 0.5). Эта метрика используется для оценки single-person НРЕ. Однако, PCP не была широко

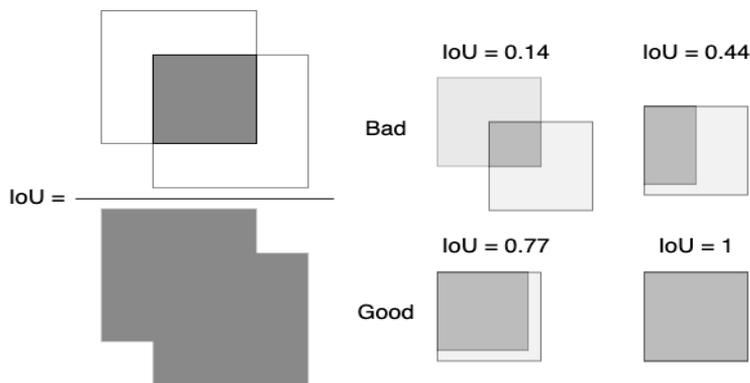


Рис. 6. Intersection over Union

применена в последних работах, потому что она наказывает конечности с короткой длиной, которые трудно обнаружить.

Percentage of Detected Joints (PDJ) (процент правильных суставов) - это метрика, в которой прогнозируемый сустав считается обнаруженным, если расстояние между прогнозируемым суставами и истинным суставом находится в пределах определенной доли диаметра туловища. Введена для того, чтобы устранить недостатки PCP

Percentage of Correct Key-points (PCK) (процент правильных ключевых точек) - используется для измерения точности локализации различных ключевых точек в пределах определенного порога, обычно устанавливаемым равным 50% от длины сегмента головы каждого тестового изображения.

Object Keypoint Similarity (OKS) (сходство ключевых точек объекта) - среднее сходство ключевых точек по всем ключевым точкам объекта. Рассчитывается на основе масштаба объекта и расстояния между прогнозируемой и истинной точками. Масштаба и константа ключевой точки необходимы для того, чтобы придать равную важность каждой ключевой точке.

Average Recall (AR) используется для измерения точности определения ключевых точек в соответствии с показателем *recall* - сколько из всех истинных положительных результатов было найдено моделью. Метрика представляет собой средние значения отзыва по всем значениям *recall* в диапазоне от 0 до 1 при различных пороговых значениях IoU. **Mean Average Recall (MAR)** - это среднее значение *recall* по всем классам при нескольких пороговых значениях по IoU всей модели.

Average Precision (AP) используется для измерения точности определения ключевых точек в соответствии с показателем *precision* - на

сколько точны прогнозы. Следовательно, показатель представляет собой средние значения точности по всем значениям precision в диапазоне от 0 до 1 при различных пороговых значениях IoU. **Mean Average Precision (MAP)** - это среднее значение precision по всем классам при нескольких пороговых значениях по IoU всей модели.

3.2. Наборы данных

Наборы данных являются одним из наиболее важных аспектов алгоритмов машинного обучения. Задача НРЕ не является исключением. Алгоритмы должны быть сначала обучены, а затем протестированы, чтобы убедиться что они правильно интерпретируют данные для выполнения поставленной задачи. Таким образом, необходимы хорошие обучающие и даже тестовые наборы данных. Существует несколько широко используемых наборов данных для задачи оценки позы человека.

Набор данных **Сосо** - наиболее широко используемый набор 2D данных, в первую очередь для multi-person НРЕ. Используется также для обнаружения объектов и содержит изображения, помеченные ключевыми точками, до 17 по всему телу. Существует 2 версии набора данных Сосо с разделением тренировочного, валидационного и тестового поднаборов данных. Разработан в 2014 году [16].

Набор данных **МРП** - 2D набор данных, используемый в основном для для single-pose моделирования. Он содержит изображения, помеченные вручную до 16 аннотированными суставами тела с расширенными обозначениями, такими как окклюзии частей тела и ориентации туловища, 3D ориентации торса и головы. Изображения охватывают 410 различных видов человеческой деятельности. Каждое изображение было взято из видео на платформе YouTube предыдущими и последующими кадрами, которые не были аннотированы. Разработан в 2014 году [17].

Набор данных **AI Challenger Human Keypoint Detection** - самый большой набор данных для двумерной оценки позы человека. Он содержит более 300 тысяч изображений с высоким разрешением, аннотированных по 14 ключевым точкам для обнаружения каждого человека и более 600 тысяч тестовых изображений. Все изображения были собраны из поисковых систем интернета и посвящены повседневной деятельности людей в различных позах [18].

PoseTrack - 2D набор данных о телосложении на основе видео, в основном используемый для multi-pose НРЕ, который содержит около 1356 видеорядов, 46 тысяч аннотированных кадров, и 276 тысяч аннотаций поз тела, содержащих до 15 ключевых точек. Предполагается, что у каждого пользователя есть уникальный идентификатор трека с аннотациями. Разработан в 2017 году [19].

Human3.6M - самый популярный и самый большой набор данных, используемый для 3D HPE. Он содержит 3.6 миллиона 3D поз человека, помеченных ключевыми 24 ключевыми точками, и использует 3 протокола с различным разделением данных для обучения и тестирования, а также 11 профессиональных актеров, выполняющих 17 действий с 4 разных точек обзора в закрытых лабораторных условиях. Разработан в 2014 году [20].

3.3. Популярные методы оценки позы человека

На сегодняшний день исследователи разработали следующие популярные модели оценки позы человека.

DeepPose [5] захватывает все соединения с помощью глубоких нейронных, добавляет слой объединения, слой свертки и полносвязный слой, чтобы сформировать части этих слоев.

OpenPose [6] - подход «снизу-вверх» для multi-person HPE с архитектурными функциями для режима реального времени. Преимущество Open Pose заключается в том, что это API, который предоставляет пользователям гибкость при выборе исходных изображений с камеры, веб-камер и других, что более важно для встроенных системных приложений.

DeepCut [7] - подход «снизу-вверх» для multi-person HPE. Модель работает путем определения количества людей на изображении, а затем прогнозирования совместных местоположений для каждого изображения. Deep Cut можно применить к видео или изображениям с несколькими объектами, например, футбол, баскетбол и другими.

Mask R-CNN - популярный алгоритм сегментации, архитектура которого может быть легко расширена для решения задач HPE. В этой модели используется сверточная нейронная сеть, которая извлекает объекты и их представление из входного изображения. Извлеченные объекты используются для генерации ряда возможных позиций для объекта, который известен как сеть региональных предложений (RPN).

MoveNet может определить 17 ключевых точек человеческого тела. Существует две версии модели. Lightning версия предназначена для приложений с низкими требованиями к задержке. Версия Thunder предназначена для приложений, которые нацелены на достижение более высокой точности. Обе модели способны обнаруживать данные в режиме реального времени и доказали свою эффективность для обнаружения приложений, работающих в режиме реального времени в фитнесе, спорте или здравоохранении. Совместно с MediaPipe достигли передовых результатов в предсказании как изображений, так и видео различных видов человеческой деятельности [15].

PoseNet [8] может использоваться для single-person или multi-person HPE для запуска на легких устройствах, таких как браузеры или мобильные устройства.

DCPose [9] или Deep Dual Consecutive Network определяет позу человека из нескольких кадров. Модель пытается преодолеть критические проблемы HPE, такие как размытость изображения при движении, расфокусированное видео и перекрытия, возникающие из-за зависимости от каждого видеокadra. Модель использует различные временные привязки между этими кадрами видео для облегчения точного определения ключевых точек.

DensePose [10] отображает различные пиксели на основе человека из RGB изображения относительно 3D поверхности человеческого тела. Модель использует крупный набор данных, содержащий аннотированную информацию от изображения к поверхности. Модель поддерживает single-person и multi-person HPE.

High-Resolution Net (HRNet) [11] находит ключевые точки относительно конкретного объекта или человека на изображении. Одним из преимуществ этой архитектуры является то, что большинство существующих моделей сопоставляют представления поз с высоким разрешением с представлениями с низким разрешением.

Regional Multi-Person Pose Estimation (AlphaPose) [12] - подход «сверху-вниз» для задачи HPE как для single-person так и multi-person случаев на изображении или видео. Модель используется для оценки человеческих поз по оптимально определенным ограничивающим рамкам при наличии их неточности.

TransPose [13] реализует основанный на сверточной нейронной сети подход к извлечению признаков, трансформер энкодер и возможности предсказания. Модель имеет встроенные функции, такие как преобразователь, который может собирать информацию из пространственных соотношений на большом расстоянии между различными ключевыми точками.

MediaPipe Pose предназначена для высокоточного отслеживания положения тела: доступно определение 33 трехмерных ориентиров и маски сегментации фона на всем теле из RGB видеокadров с использованием модели BlazePose [14], которая точно локализует ключевые точки, что делает его уникальным для фитнес приложений. Этот подход обеспечивает производительность в реальном времени на мобильных телефонах с процессорными вычислениями.

4. Применение задачи для обучения физическим навыкам

4.1. Приложения для фитнеса

Индустрия цифрового фитнеса переживает невероятный рост после COVID-19. Количество загрузок для здоровья и фитнеса выросло на 46% по всему миру 2020 году, и ожидается, что к 2025 году индустрия вырастет до \$16 млрд. [4].

Рассмотрим несколько компаний, которые разрабатывают подходы к фитнес приложениям с помощью искусственного интеллекта:

- 1) **Mirror** [21] программное обеспечение, анализирующее движения на тренировках и обеспечивающее обратную связь в режиме реального времени, основанную на предпочтениях и целях пользователя.
- 2) **Kaia Health** [22] использует компьютерное зрение и селфи камеру для предоставления звуковых подсказок и визуальной обратной связи для более эффективного выполнения упражнений.
- 3) **Vay** [23] предлагают приложение, анализирующее упражнения с помощью захвата камеры. Им удалось быстро расширить базу упражнений и добиться независимости от положения камеры.
- 4) **TwentyBN** [24] - это фитнес-киоск с ИИ тренером, который сочетает в себе видео понимание и NLP, ведет и корректирует тренировку в режиме реального времени. Их система искусственного интеллекта опирается на данные реальных пользователей и отслеживает различные движения при выполнении упражнений.

4.2. Персонализация

Современные фитнес решения с искусственным интеллектом основаны на анализе прогнозов алгоритмов оценки позы человека. Они получают координаты частей тела человека и сравнивают их с «эталонными». Но это не надежный и универсальный подход из-за отсутствия интерпретации разницы между двумя позами.

Чтобы преодолеть эту проблему, можно разработать решение, основанное на системе нечетких правил [25], которое может учитывать правила экспертной области для принятия решений, чтобы персонализированно следить за выполнением различных физических упражнений. Такие правила позволят обеспечивать более понятные и надежные решения. Например, правило может состоять из сравнения двух человеческих поз и установления фиксированных границ корректности позы.

Это может стать основой для системы рекомендаций, основанной на оценке позы человека, для различных видов деятельности от фитнеса до массовых профессий.

5. Заключение

В работе представлены достижения в области оценки позы человека, а также рассмотрены фундаментальные понятия постановки задачи, ее разновидности и существующие решения. Также проведен анализ применения данной технологии для обучения физическим навыкам на примере фитнес приложений.

Итоги исследования показывают, что решения в области оценки позы человека могут быть применены для построения инструмента персонализации обучения физическим навыкам. Такой инструмент может быть реализован в виде системы управления процессом тренировки различным физическим навыкам. Подобные персонализированные системы будут полезны в спорте (от фитнеса до подготовки спортсменов высшей квалификации) [26], здравоохранении (процессы реабилитации) [27], обучении выполнению физических навыков (рынок массовых профессий).

The problem of human pose estimation: tasks, methods, solutions Shergin I.A., Ryjov A.P.

The article deals with the problem of estimating a person's pose, which arises in many applications of computer vision systems. Various problem statements and approaches to solving the problem are considered, an overview of existing solution methods and data sets is provided. Additionally, the application of methods in the task of teaching physical skills is investigated.

Keywords: machine learning, deep learning, human posture estimation, computer vision.

References

- [1] Shubham Sharma, "Meet Zenia, an AI that can guide your yoga sessions", <https://www.newsbytesapp.com/news/science/zenia-ai-assistant-will-guide-your-yoga-sessions/story>.
- [2] Yuan Niu, Jinhua She, Chi Xu, "A Survey on IMU-and-Vision-based Human Pose Estimation for Rehabilitation", 2022 <http://dx.doi.org/10.23919/CCC55666.2022.9901525>.

- [3] Rui Liu, “A Study of Athlete Pose Estimation Techniques in Sports Game Videos Combining Multiresidual Module Convolutional Neural Networks”, 2021 <http://dx.doi.org/10.1155/2021/4367875>.
- [4] Konstantin Kalinin, “Pose Estimation: How to Track Human Movements in Real-Time Using Machine Learning”, 2021, <https://topflightapps.com/ideas/pose-estimation/>.
- [5] Alexander Toshev, Christian Szegedy, “DeepPose: Human Pose Estimation via Deep Neural Networks”, 2014 <http://dx.doi.org/10.1109/CVPR.2014.214>.
- [6] Zhe Cao, Tomas Simon, Shih-En Wei and Yaser Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, 2019.
- [7] Leonid Pishchulin, Eldar Insafutdinov, Siyu Tang, “DeepCut: Joint Subset Partition and Labeling for Multi-Person Pose Estimation”, 2016.
- [8] Pose Detection in the Browser: PoseNet Model <https://github.com/tensorflow/tfjs-models/tree/master/posenet>.
- [9] Zhenguang Liu, Haoming Chen, “Deep Dual Consecutive Network for Human Pose Estimation”.
- [10] Rıza Alp Guler, Natalia Neverova, Iasonas Kokkinos, “DensePose: Dense Human Pose Estimation In The Wild”, 2018.
- [11] Bowen Cheng, Bin Xiao, Jingdong Wang, “HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation”, 2020.
- [12] Hao-Shu Fang, Jiefeng L, “AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time”, 2022.
- [13] Sen Yang, Zhibin Quan, “TransPose: Keypoint Localization via Transformer”.
- [14] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, “BlazePose: On-device Real-time Body Pose tracking”, 2020 <https://doi.org/10.48550/arXiv.2006.10204>.
- [15] Jen-Li Chung, Lee-Yeng Ong and Meng-Chew Leow, “Comparative Analysis of Skeleton-Based Human Pose Estimation”, 2022.
- [16] Tsung-Yi Lin, “Coco dataset”, <https://cocodataset.org/home>.
- [17] Mykhaylo Andriluka, “2d human pose estimation: New benchmark and state of the art analysis”, 2014.

- [18] Jiahong Wu, “Large-scale datasets for going deeper in image understanding”, 2019.
- [19] Mykhaylo Andriluk, “Posetrack: A benchmark for human pose estimation and tracking”, 2018.
- [20] Catalin Ionescu, “Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments”, 2014.
- [21] DAMI LEE, “Mirror launches live training sessions that let coaches see you at home”, <https://www.theverge.com/2019/10/8/20899151/mirror-live-personal-training-interactive-home-gym-fitness-session-price>.
- [22] Kaia Health Team, “Digital Exercise Performance Tracking with Kaia Health’s Motion Coach”, <https://kaiahealth.com/blog/motion-coach-exercise-tracking>.
- [23] Robin Wauters, “University of Zurich spin-off Vay Sports has launched its AI-powered digital fitness coach in beta”, <https://tech.eu/2019/05/06/university-of-zurich-spin-off-vay-sports-has-launched-its-ai-powered-digital-fitness-coach-in-beta/>.
- [24] Twenty Billion Neurons Team, “2 fit-tech trends are leading us towards interactive AI fitness trainers”, <https://fityourself.club/2-fit-tech-trends-are-leading-us-towards-interactive-ai-fitness-trainers-2c74d4da4890>.
- [25] Malinka Ivanova, Petya Petkova, Nikolay Petkov, “Machine Learning and Fuzzy Logic in Electronics: Applying Intelligence in Practice”, 2021.
- [26] N. Mahendran, “Deep Learning for Fitness”, 2021.
- [27] J. Stenum, “Applications of Pose Estimation in Human Health and Performance across the Lifespan”, 2021.