УДК 004.008 **DOI** 10.15622/ia.25.3.5

И.В. Нетай , Е.П. Панкратов , К.И. Корнилов , А.Н. Голубинский

ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО КИНЕМАТИКЕ ХОДЬБЫ В ТРЁХМЕРНОМ ПРОСТРАНСТВЕ НА ОСНОВЕ ИНВАРИАНТНЫХ ПРИЗНАКОВ

Нетай И.,В., Панкратов Е.,П., Корнилов К.,И., Голубинский А.,Н. **Идентификация личности по кинематике ходьбы в трёхмерном пространстве на основе инвариантных признаков.**

Аннотация. Предлагается система сбора, предварительной обработки и анализа данных для идентификации личности по походке с использованием методов машинного обучения и анализа изображений. Система включает механизм предварительной обработки для нормализации биометрических данных движения человека и этап постобработки для выделения признаков, инвариантных относительно ортогональных преобразований пространства и изменения положения камеры, к перспективе. Реализован вычислительно эффективный метод распознавания походки с использованием одной стереокамеры. Проведен анализ пространства признаков для определения значимых характеристик, а также сравнительный анализ используемого набора признаков. Протестированы несколько архитектур машинного обучения, включая модели глубокого обучения, с анализом их точности и вычислительной эффективности. Экспериментально исследована производительность системы на различных вычислительных устройствах, измерены временные характеристики обработки данных. Результаты позволяют сравнить применимость и эффективность в зависимости от доступных вычислительных мощностей. Исходные данные формируются как временные ряды 3D-координат ключевых точек, реконструируемых по глубине на основе 2D-детекций позы (COCO, 17 точек), при этом пропуски детекций кодируются нулевым вектором. Для идентификации по походке точки головы исключаются как неинформативные, а пропуски восстанавливаются с последующим сглаживанием координат. Инвариантное признаковое пространство строится по геометрии скелета (длины сегментов, углы и их производные). Далее признаки агрегируются в единый вектор по скользящему окну, то есть каждый вектор описывает движение на интервале из нескольких последовательных кадров. На этапе идентификации личности используется фильтрация по порогу и агрегирование предсказаний на уровне кадров. Показано, что качество и скорость работы зависят от выбранной модели машинного обучения и доступных вычислительных ресурсов.

Ключевые слова: машинное обучение, компьютерное зрение, биометрия, нейронные сети, кинематика, 3D-видеообработка, инвариантные признаки

1. Введение. Результаты измерений крупной и мелкой моторики человека имеют широкий спектр применения в различных областях. Наиболее распространены маркерные системы измерений движений, используемые в компьютерной графике, биометрической идентификации и верификации, медицинской диагностике [1] и спортивной науке.

Маркерные системы обладают ограничениями [2]. Они требуют размещения отражающих индикаторов на теле, калибровки нескольких камер и контролируемой среды. Для работы, как

правило, требуются высокоскоростные камеры, специализированное программное обеспечение (ПО) и вычислительные ресурсы. Точность зависит от регулярной калибровки, что увеличивает трудоёмкость их эксплуатации.

Безмаркерные системы анализа движений уменьшают влияние некоторых недостатков. Хотя они по-прежнему уступают маркерным системам в точности, современные алгоритмы компьютерного зрения и глубокого обучения позволяют сократить этот разрыв благодаря развитию технологий трёхмерной оценки позы [3].

В данной работе предлагается метод безмаркерного анализа походки для идентификации личности с акцентом на построение инвариантного 3D-пространства признаков, отражающего уникальные кинематические характеристики каждого испытуемого. Предлагается система, сочетающая точность и простоту использования, с минимальными требованиями к оборудованию для съемки (одна стереокамера).

Предлагаемый метод основан на обработке видеозаписи походки с использованием моделей машинного обучения для оценки позы [4]. После записи и предобработки данных получаются 3D-координаты ключевых точек тела, что обеспечивает достаточно точное представление кинематики движений. Дальнейшая обработка включает выделение признаков из координат. Для обучения моделей вычисляются геометрические и кинематические ортогонально инвариантные признаки [5].

В отличие от более распространенных методов, работающих с 2D-данными или упрощенными 3D-моделями [6], данный подход обеспечивает пространственное и инвариантное представление движений. Использование ортогонально инвариантных параметров исключает возможность влияния на результаты положения и направления, угла съёмки камеры. Это потенциально повышает надёжность методов, применяемых в биометрической идентификации, медицинской диагностике и анализе двигательных нарушений.

2. Подготовка данных. Экспериментальные исследования проводились на собственном наборе видеоданных движений человека, разделённом на обучающую, валидационную и тестовую части. В эксперименте участвовали 19 испытуемых. Для каждого испытуемого было записано по три стереовидеофрагмента: две записи использовались для формирования обучающей и валидационной выборок, третья запись – в качестве тестовой. Таким образом, общий объём набора данных составил 57 видеозаписей (19×3).

Съёмка выполнялась стереоскопической камерой Intel RealSense Depth Camera D457 при разрешении 1280×720 и частоте 30 кадров/с [8]. Типичная продолжительность каждой из двух обучающих записей составляла около 90 с на одного испытуемого, тестовой записи – около 45 с. Во всех записях испытуемый двигался обычной походкой по освещённому коридору длиной 13 м, не приближаясь к камере ближе чем на 2 м. Для повышения воспроизводимости измерений и при этом обеспечения естественной вариативности каждая из трёх записей выполнялась в отдельную сессию через 2-5 недель после предыдущей; при этом допускались естественные изменения внешнего вида (например, одежда), а также небольшие вариации условий съёмки, неизбежные при повторной записи (незначительные изменения освещённости и фона, небольшие отличия в положении/наклоне камеры при установке). Такое разделение обеспечивало проверку устойчивости метода к межсессионным изменениям при сохранении сопоставимости условий [7].

Принцип разделения на обучающую/валидационную/тестовую части. Тестовая выборка формировалась целиком из третьей (отложенной) видеозаписи каждого испытуемого и не использовалась ни при настройке параметров, ни при выборе модели.

Для оценки обобщающей способности модели и выбора гиперпараметров применялась 5-кратная стратифицированная кросс-валидация. Каждая из двух обучающих видеозаписей каждого испытуемого разбивалась на 5 последовательных временных фрагментов (фолдов), внутри которых кадры сохраняли исходный временной порядок; в совокупности это давало 10 фолдов на одного испытуемого. Распределение фолдов между обучающей и валидационной подвыборками выполнялось с перемешиванием ($\text{seed} = 42$): в каждой итерации 2 из 10 фолдов составляли валидационную подвыборку, остальные 8 использовались для обучения. Сохранение временного порядка внутри фолдов уменьшает риск утечки информации между выборками при наличии сильной временной корреляции соседних кадров и при использовании скользящих окон, а перемешивание фолдов обеспечивает проверку устойчивости классификации: модель должна демонстрировать стабильное качество независимо от того, какие участки записей (начало, середина или конец движения) используются для обучения, а какие – для валидации.

Валидационная выборка применялась для:

– подбора и сравнения вариантов предварительной обработки (в частности, способов заполнения пропусков и параметров сглаживания);

– выбора архитектуры и гиперпараметров модели;

– настройки эвристик принятия решения (например, порога уверенности T при фильтрации кадров/окон);

– контроля переобучения при обучении нейросетевых моделей (ранняя остановка, планировщик шага обучения).

После фиксации всех настроек итоговая оценка качества проводилась однократно на тестовой выборке.

Процедура видеосъемки. Для обеспечения качества и сопоставимости данных соблюдались следующие условия видеосъемки:

– место съёмки должно быть хорошо освещено, в случае наличия отражающих поверхностей порог уверенности детектора YOLO (параметр `conf`) повышается до $\mathrm{conf} = 0.5$ для исключения ложных детекций (отражений);

– присутствие посторонних лиц в кадре во время формирования обучающих данных исключается; используемый метод рассчитан на распознавание одного человека в кадре;

– испытуемый должен на всём протяжении записи оставаться в пределах поля зрения видеокамеры; все части тела должны непрерывно находиться в кадре;

– движение должно начинаться при полностью видимом теле в кадре (испытуемый начинает движение, находясь целиком в поле зрения камеры);

– длительность обучающей записи должна быть не менее 30 с;

Указанные условия видеосъёмки обеспечивают получение устойчивых трёхмерных данных, необходимых для построения моделей движения и корректной оценки качества предложенного метода.

3. Последовательность обработки данных. Последовательность этапов обработки данных реализована в виде конвейера (Рис. 1) и получает на вход стереовидеозапись с камеры Intel RealSense D457, включающую RGB-поток (формат mp4) и синхронизированные данные глубины.

На первом этапе из RGB-потока формируется двумерный видеоряд, к которому применяется предобученная модель YOLOv11x-pose [9–11] для оценки позы и извлечения двумерных координат 17 ключевых точек в каждом кадре. Модель использовалась в режиме процедуры вывода без дообучения, с весами, предоставленными авторами библиотеки Ultralytics (предобучение на наборе данных COCO-pose). Все параметры детектора сохранены по умолчанию, за исключением порога уверенности, установленного на уровне $\mathrm{conf} = 0.3$ для обеспечения приемлемого баланса между полнотой обнаружения и уровнем ложных срабатываний в условиях реальной видеосъемки. Исходные кадры разрешением

1280×720 автоматически масштабировались к внутреннему размеру модели 640 пикселей по длинной стороне с сохранением пропорций; порог немаксимального подавления (NMS) $\mathrm{iou} = 0.7$. Координаты обнаруженных ключевых точек пересчитывались обратно в исходное разрешение.

На втором этапе для каждой обнаруженной суставной точки восстанавливаются трёхмерные координаты. Камера Intel RealSense D457 одновременно записывает RGB-поток и синхронизированный стереопоток, содержащий трёхмерные координаты каждого пикселя в системе координат камеры. Для перехода от двумерных детекций к трёхмерным координатам выполняется сопоставление пиксельных координат (u, v) суставных точек, полученных моделью YOLOv11x-pose из RGB-кадров, с соответствующими пикселями стереопотока, из которого извлекаются трёхмерные координаты (x, y, z).

Пространственное соответствие между RGB- и стереопотоками обеспечивается заводской калибровкой камеры: внутренние и внешние параметры хранятся в энергонезависимой памяти устройства, и выравнивание потоков выполняется автоматически средствами Intel RealSense SDK 2.0 [12]. Дополнительная калибровка камеры пользователем не проводилась.

Итоговое представление данных имеет вид массива размерности $N_{\text{кадров}} \times 17 \times 3$, содержащего трёхмерные координаты ключевых точек.

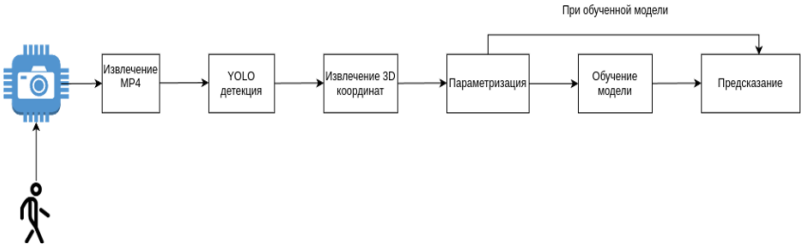

Рис. 1. Последовательность вычислений

После восстановления трёхмерных координат ключевых точек выполняется предварительная обработка, в ходе которой формируются инвариантные признаки: углы между смежными рёбрами графа ключевых точек, скорости, ускорения, длины градиентов и их первые и вторые производные. Для идентификации личности по построенным признакам используется многослойный персептрон с двумя скрытыми слоями, обеспечивающий удовлетворительное качество при малой вычислительной задержке и поддерживающий обработку в реальном

времени. При применении обученной модели выполняется та же предварительная обработка, после чего классификатор определяет личность по кинематическим признакам и выдаёт оценку уверенности. Далее приводится детальное описание всех этапов параметризации.

3.1. Предобработка трёхмерных координат и формирование признакового пространства. Исходные данные представляют собой временной ряд трёхмерных координат ключевых точек, полученных путём восстановления по данным глубины на основе двумерных детекций модели YOLOv11x-pose [11]. Данные организованы в виде массива размерности $N_{\text{кадров}} \times 17 \times 3$ в соответствии со схемой COCO (17 ключевых точек). Для задачи идентификации по походке исключаются ключевые точки головы (нос, глаза, уши) как неинформативные; в дальнейшем используется подмножество из 12 точек: плечи, локти, запястья, бёдра, колени и лодыжки. Исходные траектории содержат шум, обусловленный как внутренней вариативностью оценок детектора YOLO, так и внешними факторами, влияющими на качество детектирования [13]. Пропуски в данных (кодируемые вектором $[0, 0, 0]$) возникают в тех кадрах, где детектор не смог с достаточной уверенностью локализовать соответствующие ключевые точки. Такие случаи могут возникать при поворотах во время движения.

Предварительная обработка включает восстановление пропущенных значений, сглаживание координат, вычисление геометрических и кинематических признаков, а также их первых и вторых производных; далее выполняется повторное сглаживание, направленное на повышение устойчивости признаков. Для обеспечения инвариантности к жёстким преобразованиям сцены (поворотам, отражениям и переносам), а также к изменению положения и ориентации камеры, признаки формируются на основе относительных расстояний, углов и норм временных градиентов. Абсолютные координаты ключевых точек не используются, поскольку они зависят от глобального положения человека и камеры и, следовательно, не являются инвариантными.

3.2. Заполнение пропущенных значений. Их доля в общем объёме трёхмерных координат (после исключения ключевых точек головы – носа, глаз и ушей; используется подмножество из $K = 12$ точек) составляет 2.24%. Распределение пропусков неравномерно: наибольший процент наблюдается для суставных точек верхних конечностей – запястий (до 8.5%) и локтей (до 5.7%), что обусловлено махами рук во время ходьбы; для точек тазобедренных суставов и коленей доля пропусков не превышает 0.3%. Подробное распределение по ключевым точкам приведено в таблице 1.

Таблица 1. Доля пропусков детекции по ключевым точкам тела

Ключевая точка	Пропусков (кадров)	% от кадров
Левое плечо	283	0.77
Правое плечо	122	0.33
Левый локоть	2099	5.69
Правый локоть	1198	3.24
Левое запястье	3135	8.49
Правое запястье	2268	6.14
Левый тазобедр. сустав	81	0.22
Правый тазобедр. сустав	62	0.17
Левое колено	113	0.31
Правое колено	72	0.20
Левая лодыжка	236	0.64
Правая лодыжка	243	0.66

Далее рассматриваются пропуски в трёхмерных координатах ключевых точек, которые кодируются вектором $[0, 0, 0]$. Ключевые точки головы (нос, глаза, уши) исключаются из обработки как малозначимые для идентификации по походке. Обозначим через $\mathbf{p}_{t,k} \in \mathbb{R}^3$ координаты k-й точки в кадре t, где $t = 1, \ldots, T$, $k = 1, \ldots, K$. Введём бинарный индикатор достоверности координаты:

$$m_{t,k} = \begin{cases} 1, & \mathbf{p}_{t,k} \neq \mathbf{0}, \\ 0, & \mathbf{p}_{t,k} = \mathbf{0}, \end{cases}$$

где $\mathbf{0} = (0, 0, 0)^\top$. Восстановление выполнялось только для тех пар (t, k), для которых $m_{t,k} = 0$, с использованием одним из методов, описанных далее.

Базовый метод (forward fill). Наибольшую точность показало заполнение предыдущим валидным значением:

$$\mathbf{p}_{t,k} \leftarrow \begin{cases} \mathbf{p}_{t-1,k}, & \text{если } m_{t,k} = 0 \text{ и } m_{t-1,k} = 1, \\ \mathbf{p}_{t,k}, & \text{иначе}, \end{cases} \qquad t = 2, \ldots, T.$$

Для $t = 1$ заполнение не выполняется (первый кадр используется как исходный).

Альтернативные тестировавшиеся методы.

– **Кубическая сплайн-интерполяция по времени.** Для каждой ключевой точки k и каждой координаты $c \in \{x, y, z\}$ рассматривается одномерный временной ряд $\{p_{t,k}^{(c)}\}_{t=1}^{T}$. Обозначим через $V_k = \{t_1 < \ldots < t_m\}$ множество индексов кадров, где значение не равно $[0, 0, 0]$. Если $m \geqslant 2$, по точкам $\left(t_i,\, p_{t_i,k}^{(c)}\right)$ строится кубический сплайн $f_k^{(c)}(t)$, и для каждого пропущенного кадра $t_p \notin V_k$ координата восстанавливается как $p_{t_p,k}^{(c)} = f_k^{(c)}(t_p)$ (при $m < 2$ интерполяция для данной компоненты не выполняется).

– **KNN-заполнение по кадрам.** Каждый кадр t представляется плоским вектором $\mathbf{x}_t \in \mathbb{R}^D$, $D = 3K$, полученным конкатенацией координат всех ключевых точек: $\mathbf{x}_t = \left[\mathbf{p}_{t,1}^{\top}, \ldots, \mathbf{p}_{t,K}^{\top}\right]^{\top}$. Введём маску валидности координат в плоском представлении $\mathbf{q}_t \in \{0, 1\}^D$: для каждой точки k три соответствующие компоненты $\mathbf{q}_t$ равны $m_{t,k}$. Для кадра t выбираются n ближайших кадров $\{t_j\}_{j=1}^{n}$ по расстоянию:

$$d_j = \left\| \left(\mathbf{q}_t \odot \mathbf{q}_{t_j}\right) \odot \left(\mathbf{x}_t - \mathbf{x}_{t_j}\right) \right\|_2,$$

где $\odot$ – покомпонентное произведение. Затем отсутствующее значение $\mathbf{p}_{t,k}$ восстанавливается как взвешенное среднее по этим соседям [14]:

$$\mathbf{p}_{t,k} \leftarrow \frac{\sum_{j=1}^{n} w_j\, \mathbf{p}_{t_j,k}}{\sum_{j=1}^{n} w_j}, \qquad w_j = \frac{1}{d_j + \varepsilon},$$

где $n \in \mathbb{N}$ – число соседей, $\varepsilon > 0$ – малая константа для предотвращения деления на ноль.

– **Скользящее среднее по времени.** Пусть $w \in \mathbb{N}$ – ширина окна (в кадрах), $\lfloor \cdot \rfloor$ – целая часть. Для кадра t рассматривается окно индексов:

$$\mathcal{W}_t = \{t - \lfloor w/2 \rfloor, \ldots, t + \lfloor w/2 \rfloor\} \cap \{1, \ldots, T\}.$$

Из $\mathcal{W}_t$ выбираются только кадры с валидным значением точки k: $V = \{v \in \mathcal{W}_t : m_{v,k} = 1\}$. Пропуск в кадре t заполняется средним по этим значениям:

$$\mathbf{p}_{t,k} \leftarrow \frac{1}{|V|} \sum_{v \in V} \mathbf{p}_{v,k},$$

где $|V|$ – мощность множества V. В отличие от линейной интерполяции между двумя граничными точками пропуска, данный метод усредняет все доступные значения в окрестности и действует как локальное сглаживание траектории.

– **Итеративное модельное заполнение.** Последовательность рассматривается как матрица $\mathbf{X} \in \mathbb{R}^{T \times D}$ («кадры $\times$ координаты»), $D = 3K$, где пропуски заменены на NaN. Для каждого столбца $f \in \{1, \ldots, D\}$ с пропусками обучается регрессионная модель R_f, предсказывающая значения этого столбца по остальным признакам $\mathbf{X}_{-f}$ (матрица $\mathbf{X}$ без f-го столбца). Итерации продолжаются до сходимости по критерию:

$$\left\|\mathbf{X}^{(i)} - \mathbf{X}^{(i-1)}\right\|_F < \delta,$$

где $\mathbf{X}^{(i)}$ – матрица после i-й итерации, $\| \cdot \|_F$ – норма Фробениуса, $\delta > 0$ – порог остановки. В реализации в качестве R_f использовалось дерево решений, т.к. позволяет учитывать нелинейные взаимосвязи между координатами.

Выбор в рабочей конфигурации. Экспериментальная проверка показала, что на валидационной и тестовой выборках наибольшая точность идентификации достигается при использовании базового каузального способа заполнения пропусков (перенос последнего корректного значения вперёд, *forward fill*). Данный способ в меньшей степени вносит фазовые и амплитудные искажения, не приводит к переусреднению траектории, вычислительно малозатратен и не вызывает избыточного сглаживания данных.

Распределение длин непрерывных пропусков. Для оценки структуры пропусков была проанализирована длина каждого непрерывного пропуска – число последовательных кадров, в которых данная ключевая точка не была детектирована. Пусть для точки k на участке кадров $[t_{\text{start}}, t_{\text{end}}]$ выполняется $m_{t,k} = 0$ (пропуск), а в кадрах $t_{\text{start}} - 1$ и $t_{\text{end}} + 1$ значения валидны. Тогда длина пропуска определяется как:

$$g = t_{\text{end}} - t_{\text{start}} + 1. \quad (1)$$

Из 1374 зафиксированных непрерывных пропусков 37.2% составляют одиночные кадры ($g = 1$), медианная длина – 3 кадра, средняя – 7.2 кадра, 95-й перцентиль – 22 кадра. Максимальная длина пропуска составила 256 кадров (1 случай из 1374, связанный с длительным поворотом испытуемого во время движения). При частоте дискретизации 30 кадров/с одиночный пропуск соответствует интервалу 33 мс, за который координаты ключевой точки изменяются незначительно, что обосновывает выбор forward fill в качестве базового метода заполнения.

Абсолютные длины сегментов скелета. Для оценки качества трёхмерной реконструкции были вычислены средние длины сегментов скелетного графа $G = (V, E)$ в метрах. Пусть $e = (i, j) \in E$ – ребро скелета. Длина сегмента e в кадре t (после заполнения пропусков, до сглаживания) определяется как:

$$L_t^{(e)} = \|\mathbf{p}_{t,i} - \mathbf{p}_{t,j}\|_2 . \tag{2}$$

Из рассмотрения исключались кадры с нефизичными значениями длин ($L_t^{(e)} < 0.05$ м или $L_t^{(e)} > 2.0$ м), обусловленными выбросами стереореконструкции глубины. Обозначим множество валидных кадров для сегмента e и записи s как $\mathcal{V}_s^{(e)}$.

Среднее значение длины сегмента для записи s:

$$\bar{L}_s^{(e)} = \frac{1}{\left|\mathcal{V}_s^{(e)}\right|} \sum_{t \in \mathcal{V}_s^{(e)}} L_t^{(e)}. \tag{3}$$

Для испытуемых с несколькими записями значения $\bar{L}_s^{(e)}$ предварительно усреднялись:

$$\bar{L}_u^{(e)} = \frac{1}{|S_u|} \sum_{s \in S_u} \bar{L}_s^{(e)}, \tag{4}$$

где S_u – множество записей испытуемого u. Итоговые статистики вычислялись по $U = 19$ испытуемым. Среднее значение длины сегмента:

$$\mathrm{Mean}^{(e)} = \frac{1}{U}\sum_{u=1}^{U} \bar{L}_u^{(e)}. \tag{5}$$

Межсубъектное среднеквадратическое отклонение, характеризующее антропометрический разброс:

$$\mathrm{Std}^{(e)} = \sqrt{\frac{1}{U}\sum_{u=1}^{U}\left(\bar{L}_u^{(e)} - \mathrm{Mean}^{(e)}\right)^2}. \tag{6}$$

Результаты приведены в таблице 2. Средние длины сегментов соответствуют антропометрическим нормам: голень – 0.44 м, бедро – 0.45 м, торс (плечо–бедро) – 0.55 м, плечо – 0.32 м, предплечье – 0.29 м. Межсубъектное СКО составляет 0.03–0.04 м, что отражает естественный антропометрический разброс между испытуемыми и свидетельствует о том, что длины сегментов скелета несут индивидуальную идентификационную информацию.

Таблица 2. Средние длины сегментов скелета (м) (до сглаживания координат). Mean – среднее значение длины, усреднённое по испытуемым; Std – СКО средних длин между испытуемыми (антропометрический разброс)

Сегмент	Mean (м)	Std (м)
Лодыжка–колено (лев.)	0.442	0.034
Колено–бедро (лев.)	0.445	0.029
Лодыжка–колено (прав.)	0.442	0.033
Колено–бедро (прав.)	0.445	0.032
Бедро–бедро (таз)	0.233	0.033
Плечо–бедро (лев.)	0.547	0.037
Плечо–бедро (прав.)	0.547	0.042
Плечо–плечо	0.346	0.037
Плечо–локоть (лев.)	0.319	0.028
Плечо–локоть (прав.)	0.320	0.028
Локоть–запястье (лев.)	0.294	0.038
Локоть–запястье (прав.)	0.293	0.031

Оценка искажения длины сегментов. Для количественной оценки влияния заполнения пропусков на восстановление траекторий были рассчитаны относительные изменения длин сегментов скелета после

заполнения. Пусть $G = (V, E)$ – скелетный граф ключевых точек без головы, $|V| = K = 12$, а рёбра $E = \{(u_i, v_i)\}_{i=1}^{N}$, $N = 12$, задают сегменты скелета. Длину i-го сегмента в кадре t определим как:

$$L_t^i = \left\| \mathbf{p}_{t,u_i} - \mathbf{p}_{t,v_i} \right\|_2, \qquad t = 1, \ldots, T.$$

Относительное изменение длины сегмента вычисляется по последовательным кадрам:

$$\zeta_t^i = \frac{L_t^i - L_{t-1}^i}{L_{t-1}^i}, \qquad t = 2, \ldots, T.$$

Значения ζ_t^i учитывались только в тех кадрах, где обе длины определены, т.е. $m_{t,u_i} = m_{t,v_i} = m_{t-1,u_i} = m_{t-1,v_i} = 1$. Обозначим множество таких индексов кадров через:

$$\mathcal{T}_i = \left\{ t \in \{2, \ldots, T\} : \; m_{t,u_i} m_{t,v_i} m_{t-1,u_i} m_{t-1,v_i} = 1 \right\}, \qquad T_i = |\mathcal{T}_i|.$$

По всем полученным значениям $\{\zeta_t^i : \; t \in \mathcal{T}_i, \; i = 1, \ldots, N\}$ рассчитаны статистические характеристики (для способа *forward fill*, показавшего наилучшие результаты при классификации):

$$\mu = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \zeta_t^i = -7.8 \times 10^{-5},$$

$$\sigma^2 = \frac{1}{\sum_{i=1}^{N} T_i} \sum_{i=1}^{N} \sum_{t \in \mathcal{T}_i} \left(\zeta_t^i - \mu \right)^2 = 1.34 \times 10^{-4},$$

где μ – среднее относительное изменение, σ^2 – дисперсия. Малое значение μ указывает на отсутствие выраженного систематического смещения длин сегментов после заполнения пропусков. Дисперсия σ^2 (стандартное отклонение $\sigma \approx 0.0116$) характеризует высокую устойчивость восстановленных траекторий: в большинстве случаев относительные изменения длин сегментов не превышают $\pm 1.2\%$.

3.3. Сглаживание координат. После заполнения пропусков применяется фильтр экспоненциального забывания. Сглаживание выполняется покадрово для каждой ключевой точки и каждой координаты:

$$p'_{t,j,c} = \alpha\, p_{t,j,c} + (1-\alpha)\, p'_{t-1,j,c}, \quad t = 2, \ldots, T,\ j = 1, \ldots, K,\ c \in \{x, y, z\},$$

где T – число кадров в последовательности, $K = 12$ – число используемых ключевых точек, $p_{t,j,c}$ – исходная координата компоненты c точки j в кадре t, $p'_{t,j,c}$ – сглаженная координата, $\alpha \in (0, 1)$ – коэффициент сглаживания (в первичном сглаживании использовалось $\alpha = 0.007$). Инициализация задаётся как $p'_{1,j,c} = p_{1,j,c}$.

Для компенсации начальной предвзятости при коротких фрагментах может применяться поправка:

$$\widehat{p}'_{t,j,c} = \frac{p'_{t,j,c}}{1 - (1-\alpha)^t}, \quad t = 1, \ldots, T.$$

Связь коэффициента сглаживания с частотой среза. Связь α с постоянной времени τ при частоте кадров f_s ($T_s = 1/f_s$) задаётся:

$$\tau = -\frac{T_s}{\ln(1-\alpha)}, \qquad f_c \approx \frac{1}{2\pi\tau} = -\frac{\ln(1-\alpha)}{2\pi T_s} = -\frac{f_s\, \ln(1-\alpha)}{2\pi}.$$

Значения коэффициентов сглаживания определены исходя из физических свойств фильтруемых величин. При $f_s = 30$ кадр/с и $\alpha = 0{,}007$ получаем $\tau \approx 4{,}74$ с и $f_c \approx 0{,}034$ Гц. Данная частота среза существенно ниже частоты шагового цикла (0,8-1,0 Гц). Такой выбор параметра является намеренным: целью первичного сглаживания является не сохранение временной динамики координат, а стабилизация геометрии скелета для последующего извлечения инвариантных признаков – длин сегментов и углов между рёбрами скелетного графа. Для жёстких сегментов скелета длины должны быть приблизительно постоянными во времени, и агрессивное сглаживание координат подавляет покадровый шум 3D-реконструкции, обусловленный погрешностями оценки глубины камерой Intel RealSense D457 и вариативностью детекций позы. Динамическая информация о походке кодируется на следующем

этапе – через временные производные извлечённых признаков (углов и длин).

Дополнительно, после вычисления признаков применялось сглаживание признаковых рядов с коэффициентом $\alpha_f = 0{,}05$ (при той же частоте дискретизации), для которого $\tau \approx 0{,}65$ с и $f_c \approx 0{,}245$ Гц. Поскольку признаки извлекаются из уже сглаженных координат, остаточный шум в них невелик и достаточно мягкой фильтрации; при этом данный фильтр сохраняет динамику шагового цикла.

Таким образом, двухуровневая схема сглаживания реализует разделение задач: первое сглаживание ($f_c \approx 0{,}034$ Гц) стабилизирует геометрию скелета для извлечения инвариантных признаков, второе ($f_c \approx 0{,}245$ Гц) стабилизирует сами признаки при сохранении временной динамики движений.

Также были протестированы свёрточные методы: окно Ханна, медианный фильтр и скользящее среднее. Оконные свёртки требуют центрирования окна (что повышает задержку), и некоторые, использующие оконные функции с разрывами, подвержены краевым искажениям; медианный фильтр эффективен против импульсного шума, но искажает гладкие траектории и их производные. Фильтр экспоненциального забывания показал наилучший баланс между подавлением шумов и сохранением динамики: оно инвариантно по времени, линейно и даёт малую задержку при заданной степени сглаживания.

3.4. Формирование признаков. Первоначально рассматривался вариант прямого использования сглаженных трёхмерных координат ключевых точек без явной параметризации. Пусть $\mathbf{p}_{t,j} \in \mathbb{R}^3$ – вектор координат j-й ключевой точки в кадре t, где $t = 1, \ldots, T$, $j = 1, \ldots, K$. Результат, полученный после заполнения пропусков и сглаживания, обозначим $\mathbf{p}'_{t,j}$. Для базового варианта по последовательности $\{\mathbf{p}'_{t,j}\}_{t=1}^{T}$ формировались скользящие окна ширины L:

$$\mathbf{W}_t = \big[\mathbf{p}'_t,\, \mathbf{p}'_{t+1},\, \ldots,\, \mathbf{p}'_{t+L-1}\big],$$

где $\mathbf{p}'_t = [(\mathbf{p}'_{t,1})^\top, \ldots, (\mathbf{p}'_{t,K})^\top]^\top \in \mathbb{R}^{3K}$. Полученные окна подавались на вход классификатора в предположении, что модель самостоятельно выделит информативные характеристики движения. На практике такой способ представления обеспечил точность не выше 25%, что послужило основанием для перехода к явному построению инвариантных геометрико-кинематических признаков.

Далее используется скелетный граф ключевых точек $G = (V, E)$, заданный схемой COCO, с подмножеством точек без головы $|V| = K = 12$ и множеством связей E ($|E| = N_E = 12$ рёбер). Дополнительно введём множество троек $\mathcal{A} \subseteq V^3$ для вычисления углов между смежными рёбрами; $|\mathcal{A}| = N_\theta = 16$, а каждая тройка $(i, j, r) \in \mathcal{A}$ соответствует углу при вершине j между рёбрами (i, j) и (r, j).

Набор признаков. Признаки вычисляются покадрово, после чего агрегируются во временные окна (раздел 3.5).

1. **Длины сегментов.** Для каждого ребра $e = (i, j) \in E$ и кадра t длина сегмента определяется как:

$$\ell_{t,e} = \left\|\mathbf{p}'_{t,i} - \mathbf{p}'_{t,j}\right\|_2,$$

что даёт $N_E = 12$ признаков на кадр.

2. **Углы между смежными рёбрами.** Для каждой тройки $(i, j, r) \in \mathcal{A}$ положим $\mathbf{u}_{t,(i\to j)} = \mathbf{p}'_{t,i} - \mathbf{p}'_{t,j}$ и $\mathbf{u}_{t,(r\to j)} = \mathbf{p}'_{t,r} - \mathbf{p}'_{t,j}$. Тогда:

$$c_{t,(i,j,r)} = \frac{\mathbf{u}_{t,(i\to j)} \cdot \mathbf{u}_{t,(r\to j)}}{\|\mathbf{u}_{t,(i\to j)}\|_2 \, \|\mathbf{u}_{t,(r\to j)}\|_2}, \qquad \theta_{t,(i,j,r)} = \arccos\bigl(c_{t,(i,j,r)}\bigr),$$

что даёт $N_\theta = 16$ признаков на кадр.

3. **Динамические величины.** Для учёта динамики вводятся центральные разностные операторы первого и второго порядков по времени, применяемые к каждой точке покоординатно:

$$\Delta_1 \mathbf{p}'_{t,j} = \frac{1}{2}\bigl(\mathbf{p}'_{t+1,j} - \mathbf{p}'_{t-1,j}\bigr), \qquad \Delta_2 \mathbf{p}'_{t,j} = \mathbf{p}'_{t+1,j} - 2\mathbf{p}'_{t,j} + \mathbf{p}'_{t-1,j},$$

где $t \in \{2, \ldots, T-1\}$, $j = 1, \ldots, K$.

На границах $t \in \{1, T\}$ значения $\Delta_1 \mathbf{p}'_{t,j}$ и $\Delta_2 \mathbf{p}'_{t,j}$ задаются копированием ближайшего определённого значения (из кадра $t = 2$ или $t = T - 1$).

– *«Скорости» и «ускорения» сегментов.* Для каждого ребра $e = (i, j) \in E$ и $r \in \{1, 2\}$ определим:

$$\ell_{t,e}^{(r)} = \left\| \Delta_r \mathbf{p}'_{t,i} - \Delta_r \mathbf{p}'_{t,j} \right\|_2,$$

что даёт $N_E + N_E = 24$ признака на кадр.

– *Производные углов.* Вектор углов в кадре обозначим $\boldsymbol{\theta}_t \in \mathbb{R}^{N_\theta}$, где компоненты $\boldsymbol{\theta}_t$ равны $\theta_{t,(i,j,r)}$ для $(i,j,r) \in \mathcal{A}$. Тогда разностные аппроксимации первого и второго порядков по времени вычисляются как $\Delta_1 \boldsymbol{\theta}_t$ и $\Delta_2 \boldsymbol{\theta}_t$, что добавляет $N_\theta + N_\theta = 32$ признака на кадр.

4. **Маски и финальное сглаживание.** Пусть $m_{t,j} \in \{0,1\}$ – индикатор валидности координаты ключевой точки (раздел 3.2). Для каждого признака f в кадре t вводится маска $m_{t,f} \in \{0,1\}$, равная произведению индикаторов $m_{\tau,j}$ по всем точкам j и кадрам τ, которые используются при вычислении данного признака в кадре t. Например, для длины сегмента $e = (i,j)$ в кадре t: $m_{t,\ell_e} = m_{t,i} m_{t,j}$; для угла (i,j,r): $m_{t,\theta_{(i,j,r)}} = m_{t,i} m_{t,j} m_{t,r}$; для динамических признаков с центральной разностью в кадре t: дополнительно требуется валидность в кадрах $t-1$ и $t+1$.

Непосредственно перед финальным сглаживанием недостоверные компоненты зануляются покомпонентно:

$$\tilde{\mathbf{x}}_t = \mathbf{m}_t \odot \mathbf{x}_t, \qquad \mathbf{m}_t = [m_{t,1}, \ldots, m_{t,F}]^\top,$$

где $\mathbf{x}_t \in \mathbb{R}^F$ – вектор всех признаков кадра t, F – размерность признакового пространства. Далее к непрерывным признаковым рядам применяется фильтр экспоненциального забывания:

$$z'_{t,f} = \alpha_f \, \tilde{z}_{t,f} + (1 - \alpha_f) \, z'_{t-1,f}, \qquad f = 1, \ldots, F, \quad t = 2, \ldots, T,$$

с коэффициентом $\alpha_f = 0.05$ покомпонентно и инициализацией $z'_{1,f} = \tilde{z}_{1,f}$. Разностные ряды $\Delta_1(\cdot)$ и $\Delta_2(\cdot)$ вычисляются по времени из уже сглаженных базовых последовательностей.

Итоговое пространство признаков имеет размерность $F = 116$ (12 длин сегментов + 16 углов + 24 разностных признака длин + 32 разностных признака углов + 32 скалярные связи) и, за счёт использования относительных расстояний, углов и скалярных произведений, инвариантно к ортогональным преобразованиям и сдвигам сцены.

3.5. Оконная обработка для учёта временной динамики. Для учёта временной структуры движений применялись схемы скользящих окон, обеспечивающие агрегирование покадровых признаков во временные фрагменты фиксированной длины [15]. Пусть $\mathbf{x}_t \in \mathbb{R}^F$ – вектор из F признаков, вычисленных для кадра $t = 1, \ldots, T$. Маска валидности признаков $\mathbf{m}_t \in \{0,1\}^F$ определяется по индикаторам валидности координат $m_{t,j}$ (раздел 3.2) следующим образом: компонент $m_{t,f} = 1$ тогда и только тогда, когда для вычисления f-го признака в кадре t доступны (валидны) все требуемые ключевые точки и, при необходимости, соседние кадры (например, $t-1$ и $t+1$ для центральных разностей). При отсутствии масок полагаем $\mathbf{m}_t = \mathbf{1}$. Далее используется вектор $\tilde{\mathbf{x}}_t = \mathbf{m}_t \odot \mathbf{x}_t$, где $\odot$ – покомпонентное произведение.

Плотное скользящее окно (dense sliding window). Пусть $L \in \mathbb{N}$ – длина окна (в кадрах), $h \in \mathbb{N}$ – шаг между окнами. Для $t \in \{1,\, 1+h,\, \ldots,\, 1+(N_{\text{win}}^{\text{dense}}-1)h\}$ определим окно:

$$\mathbf{W}_t^{\text{dense}} = \big[\tilde{\mathbf{x}}_t, \tilde{\mathbf{x}}_{t+1}, \ldots, \tilde{\mathbf{x}}_{t+L-1}\big] \in \mathbb{R}^{F\times L}, \qquad \mathbf{w}_t^{\text{dense}} = \operatorname{vec}\big(\mathbf{W}_t^{\text{dense}}\big) \in \mathbb{R}^{LF},$$

где $\operatorname{vec}(\cdot)$ – оператор векторизации (конкатенация столбцов матрицы). Число окон равно:

$$N_{\text{win}}^{\text{dense}} = \left\lfloor \frac{T-L}{h} \right\rfloor + 1,$$

а определение корректно при условии $T \geqslant L$.

Прореженное окно (sparse sliding window). Пусть $n \in \mathbb{N}$ – число отобранных кадров в окне, $s \in \mathbb{N}$ – шаг прореживания внутри окна. Для начала окна в кадре t вводятся индексы:

$$\mathcal{I}_t = \{t,\, t+s,\, \ldots,\, t+(n-1)s\}, \qquad t+(n-1)s \leqslant T.$$

Тогда:

$$\mathbf{W}_t^{\text{dec}} = \big[\tilde{\mathbf{x}}_t, \tilde{\mathbf{x}}_{t+s}, \ldots, \tilde{\mathbf{x}}_{t+(n-1)s}\big] \in \mathbb{R}^{F\times n}, \qquad \mathbf{w}_t^{\text{dec}} = \operatorname{vec}\big(\mathbf{W}_t^{\text{dec}}\big) \in \mathbb{R}^{nF}.$$

При сдвиге начала окна на один кадр число таких окон равно:

$$N_{\text{win}}^{\text{dec}} = T - (n-1)s,$$

а определение корректно при условии $T \geqslant (n-1)s + 1$.

По результатам экспериментов на валидационной и тестовой выборках наилучшие показатели продемонстрировало плотное скользящее окно шириной $L = 30$ и шагом $h = 3$, а также прореженное окно при $s \in \{1, 2, 3\}$; с небольшим преимуществом лидировало плотное окно, которое было принято в качестве основного варианта в дальнейших экспериментах и обучении.

4. Обучение и тестирование моделей машинного обучения. Для решения задачи идентификации личности по кинематическим признакам походки рассматривались методы градиентного бустинга (CatBoost, XGBoost, LightGBM) и нейросетевые модели (MLP, LSTM) [16–18]; обучение и оценка выполнялись на объединённом корпусе с разделением на обучающие/тестовые выборки и последующей k-fold кросс-валидацией для робастной оценки качества.

Нормализация признаков и процедура обучения. Оконные векторы признаков всех обучающих записей объединялись в единую матрицу $X \in \mathbb{R}^{N \times D}$, где N – общее число окон, $D = LF$ – размерность векторизованного окна признаков (в рабочей конфигурации $L = 30$, $F = 116$, $D = 3480$). Масштабирование выполнялось стандартной нормализацией:

$$\mathbf{X}' = \frac{\mathbf{X} - \boldsymbol{\mu}}{\boldsymbol{\sigma}}, \quad \boldsymbol{\mu}, \boldsymbol{\sigma} \in \mathbb{R}^D.$$

Разделение на обучающую и валидационную подвыборки выполнялось в соответствии с процедурой 5-кратной кросс-валидации, описанной в разделе 2. На каждом фолде модель обучалась на обучающей подвыборке, а валидационная подвыборка использовалась для ранней остановки и сохранения лучшего состояния модели (по минимуму валидационной функции потерь).

По завершении всех пяти фолдов итоговой выбиралась модель того фолда, который обеспечил наибольшее значение Accuracy на своей валидационной подвыборке. Именно эта модель затем применялась к тестовому набору (третьи видеозаписи всех испытуемых).

В таблице 3 колонки «Обучение» и «Валидация» соответствуют обучающей и валидационной подвыборкам лучшего фолда, а колонка

«Тестирование» – результатам применения этой модели к отложенному тестовому набору.

Архитектура многослойного персептрона (MLP).

В качестве основной модели классификации использовался глубокий многослойный персептрон (MLP), реализованный на PyTorch. Модель принимает на вход нормализованный вектор признаков $\mathbf{X}' \in \mathbb{R}^{N \times D}$, где $D = 3480$ (116 признаков $\times$ 30 временных шагов после препроцессинга), и выполняет классификацию на C классов (число персон в задаче идентификации).

Архитектура состоит из четырёх линейных слоёв со следующими скрытыми размерностями:

– вход $\rightarrow$ 512 нейронов
– 512 $\rightarrow$ 256 нейронов
– 256 $\rightarrow$ 128 нейронов
– 128 $\rightarrow C$ (выходной слой)

Каждый скрытый блок имеет структуру:

$$\textbf{Linear} \rightarrow \textbf{BatchNorm1d} \rightarrow \textbf{GELU} \rightarrow \textbf{Dropout}$$

Вероятности dropout увеличиваются по глубине слоя:

$$d_0 = 0.10, d_1 = 0.15, d_2 = 0.30$$

Математически прямой проход описывается следующим образом:

$$\begin{aligned}
\mathbf{H}_0 &= \mathbf{X}', \\
\mathbf{H}_1 &= \mathrm{Dropout}_{0.10}\big(\mathrm{GELU}(\mathrm{BN}(\mathbf{H}_0\mathbf{W}_0 + \mathbf{b}_0))\big), \quad \mathbf{W}_0 \in \mathbb{R}^{D\times 512}, \\
\mathbf{H}_2 &= \mathrm{Dropout}_{0.15}\big(\mathrm{GELU}(\mathrm{BN}(\mathbf{H}_1\mathbf{W}_1 + \mathbf{b}_1))\big), \quad \mathbf{W}_1 \in \mathbb{R}^{512\times 256}, \\
\mathbf{H}_3 &= \mathrm{Dropout}_{0.30}\big(\mathrm{GELU}(\mathrm{BN}(\mathbf{H}_2\mathbf{W}_2 + \mathbf{b}_2))\big), \quad \mathbf{W}_2 \in \mathbb{R}^{256\times 128}, \\
\mathbf{O} &= \mathbf{H}_3\mathbf{W}_{\text{out}} + \mathbf{b}_{\text{out}}, \quad \mathbf{W}_{\text{out}} \in \mathbb{R}^{128\times C}.
\end{aligned}$$

Инициализация весов линейных слоёв выполнена методом **Kaiming Normal** с режимом `fan_in` и нелинейностью `relu` (рекомендуется для GELU). Смещения инициализированы нулями. Модель обучалась в режиме 5-fold стратифицированной кросс-валидации на всём доступном

обучающем наборе данных. Лучшая модель по валидационной потере сохранялась для последующей оценки.

Функция потерь. Основная функция потерь — многоклассовая кросс-энтропия:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N} \log p_{i,y_i}, \quad p_{i,c} = \frac{\exp(O_{i,c})}{\sum_{j=1}^{C}\exp(O_{i,j})}.$$

Также использовалась фокальная потеря (Focal Loss) для снижения вклада легко классифицируемых примеров и учёта дисбаланса классов [19]:

$$\mathcal{L}_{\mathrm{focal}} = -\frac{1}{N}\sum_{i=1}^{N} \alpha_{y_i}\,(1-p_{i,y_i})^{\gamma}\,\log p_{i,y_i},$$

где $\gamma = 2$, а α_c задавались пропорционально эмпирическим долям классов в текущем валидационном фолде. Однако результаты использования этой функции потерь не дали значительного прироста по точности предсказания.

Обучение, планировщик и ранняя остановка. Обучение выполнялось с использованием одного большого батча данных (размер батча $10\,000$ объектов), включающего наблюдения по всем испытуемым, при максимальном числе эпох 400. Фактическое число эпох определялось сочетанием правила адаптивного уменьшения шага обучения ReduceLROnPlateau и критерия ранней остановки. Шаг обучения автоматически уменьшался по правилу ReduceLROnPlateau (factor $= 0.5$, patience $= 3$, $\eta_{\min} = 10^{-5}$); при этом уменьшение происходило при отсутствии улучшения выбранной метрики на валидационной выборке в течение заданного числа эпох [20]. Критерий ранней остановки задавался как отсутствие улучшения значения функции потерь на валидационной выборке в течение 10 последовательных эпох.

Для обеспечения воспроизводимости результатов все источники случайности (инициализация весов, разбиение данных на фолды, порядок формирования пакетов данных) фиксировались при seed $= 42$.

4.1. Процедура вывода и агрегирование решений. При выполнении процедуры вывода к предварительно вычисленным признакам тестовой выборки применялся сохранённый преобразователь StandardScaler, после чего использовалась обученная модель MLP

в рабочей конфигурации. Пусть $\mathbf{x}_i \in \mathbb{R}^D$ – вектор признаков i-го экземпляра данных (окна или кадра) после нормализации, где $i = 1, \ldots, N$, N – общее число экземпляров данных в рассматриваемой выборке, D – размерность признакового вектора. Выходные значения сети (логиты) обозначим $\mathbf{o}_i \in \mathbb{R}^C$, где C – число классов (число испытуемых). Вероятности классов определяются как $\mathbf{p}_i = \mathrm{softmax}(\mathbf{o}_i)$, то есть $p_{i,c}$ – оценка вероятности класса c для экземпляра i, $c = 1, \ldots, C$.

Фильтрация по порогу разности вероятностей. Для каждого экземпляра данных определяется класс с максимальной вероятностью:

$$\hat{y}_i = \arg \max_{c \in \{1,\ldots,C\}} p_{i,c}.$$

и класс, занявший второе место по величине вероятности,

$$\tilde{y}_i = \arg \max_{c \in \{1,\ldots,C\} \setminus \{\hat{y}_i\}} p_{i,c}.$$

Вводится величина разности вероятностей (margin):

$$M_i = p_{i,\hat{y}_i} - p_{i,\tilde{y}_i}.$$

К дальнейшему анализу допускаются только те экземпляры, для которых выполняется условие:

$$\mathcal{I}(\tau) = \big\{ i \in \{1, \ldots, N\} : M_i \geqslant \tau \big\},$$

где $\tau \in [0, 1]$ – фиксированный порог уверенности; в базовой оценке использовалось $\tau = 0.9$.

Агрегирование по видеозаписи. Пусть $\mathcal{F}$ – множество тестовых видеозаписей (файлов), а $F \in \mathcal{F}$ – одна видеозапись. Обозначим через $I(F) \subseteq \{1, \ldots, N\}$ множество индексов экземпляров данных (окон/кадров), сформированных из видеозаписи F. Для каждой видеозаписи F накапливается взвешенная сумма по классам, рассчитанная только по отфильтрованным экземплярам:

$$S_c(F;\tau) \;=\; \sum_{\substack{i\in I(F)\cap\mathcal{I}(\tau)\\ \hat{y}_i=c}} M_i, \qquad c=1,\ldots,C,$$

а итоговое решение по видеозаписи определяется как:

$$\widehat{Y}(F) \;=\; \arg\max_{c\in\{1,\ldots,C\}} S_c(F;\tau).$$

Дополнительно вычисляется усреднённое распределение вероятностей:

$$\bar{\mathbf{p}}(F;\tau) \;=\; \frac{1}{|I(F)\cap\mathcal{I}(\tau)|} \sum_{i\in I(F)\cap\mathcal{I}(\tau)} \mathbf{p}_i,$$

по которому формируется список трёх наиболее вероятных классов для повышения интерпретируемости результатов.

Итоговое решение и показатели качества классификации. На уровне экземпляров данных (окон/кадров) формируется отчёт о классификации по множеству $\mathcal{I}(\tau)$ с метриками $\mathrm{Accuracy}$, $\mathrm{F1}_{\mathrm{weighted}}$, $\mathrm{Precision}_{\mathrm{weighted}}$, $\mathrm{Recall}_{\mathrm{weighted}}$. На уровне видеозаписей оценивается доля корректно распознанных записей:

$$\mathrm{Acc}_{\mathrm{files}} \;=\; \frac{\left|\left\{F\in\mathcal{F}:\ \widehat{Y}(F)=Y(F)\right\}\right|}{|\mathcal{F}|},$$

где $Y(F)\in\{1,\ldots,C\}$ – истинная метка личности для видеозаписи F.

Другие модели, используемые в эксперименте. Для сравнительной оценки в эксперименте были рассмотрены четыре модели машинного обучения, решающие задачу многоклассовой классификации походки по последовательностям предварительно вычисленных признаков.

Модель CatBoost представляет собой градиентный бустинг над решающими деревьями с поддержкой категориальных признаков. Лучшая конфигурация, полученная при случайном поиске гиперпараметров, включала 600 итераций, глубину деревьев 10, скорость обучения 0.01,

тип роста деревьев Depthwise, бутстрэп по схеме Bernoulli с subsample = 0.5, минимальное число объектов в листе min_data_in_leaf = 2, L_2-регуляризацию l2_leaf_reg = 1, random_strength = 0.1 и автоматическое взвешивание классов auto_class_weights = Balanced. Обучение выполнялось с использованием графического процессора.

Модель LightGBM также относится к семейству методов градиентного бустинга. Итоговая конфигурация включала 700 деревьев, num_leaves = 31, максимальную глубину 10, скорость обучения 0.01, min_data_in_leaf = 20, feature_fraction = 1.0, регуляризацию lambda_l1 = 0.1 и lambda_l2 = 0.1, subsample_for_bin = 200000 и тип бустинга gbdt. Как и в случае CatBoost, обучение выполнялось с использованием графического процессора.

Модель XGBoost настраивалась следующим образом: 600 деревьев, максимальная глубина 10, скорость обучения 0.01, subsample = 0.5, min_child_weight = 2, reg_lambda = 1, метод построения деревьев gpu_hist, целевая функция multi:softmax и функция качества mlogloss. Обучение также выполнялось с использованием графического процессора.

Модель LSTM представляет собой рекуррентную нейронную сеть с механизмом долгой краткосрочной памяти. Архитектура включала четыре двунаправленных слоя LSTM с размерностью скрытого состояния 186 и межслойным dropout 0.6. К выходу последнего временного шага применялась полносвязная классифицирующая часть: линейный слой, преобразующий конкатенацию прямого и обратного состояний ($186 \times 2 = 372$) в 128 нейронов с активацией ReLU и dropout = 0.3, а затем финальный линейный слой в пространство размерности N_classes.

Модели градиентного бустинга (CatBoost, LightGBM, XGBoost) обучались на покадровых признаках без явного учёта временной зависимости: каждый кадр рассматривался как независимый объект. В отличие от них, модель LSTM обрабатывала временные последовательности, что позволяло учитывать динамику изменения признаков и, как следствие, описывать временные закономерности походки.

Практические замечания. Порог T задаёт компромисс между точностью и полнотой на уровне кадров и влияет на устойчивость агрегирования по отдельным видеозаписям. В демонстрационных и базовых экспериментах использовалось фиксированное значение $T = 0.9$, тогда как в прикладных сценариях оно может подбираться по валидационной выборке.

Результаты экспериментов. Качество классификации оценивалось на тестовой выборке по двум основным показателям:

взвешенной F_1-мере ($\mathrm{F1}_{\mathrm{weighted}}$) и числу корректных идентификаций личности на уровне видеозаписей (TP – число истинно положительных решений). Для обучающей и валидационной подвыборок приводятся средние значения и стандартные отклонения по пяти фрагментам кросс-валидации; тестовые метрики являются точечными оценками, полученными однократным применением лучшей модели к отложенному тестовому набору. Сравнительные результаты для рассмотренных архитектур представлены в таблице 3.

Таблица 3. Метрики на обучающих, валидационных и тестовых данных (среднее $\pm$ std по 5 фолдам; тест – точечная оценка)

Модель	**Обучение**		**Валидация**		**Тестирование**	
	F1	**TP**	**F1**	**TP**	**F1**	**TP**
MLP	$0{,}99_{\pm 0{,}01}$	19/19	$0{,}98_{\pm 0{,}01}$	19/19	0,69	14/19
LSTM	$0{,}99_{\pm 0{,}01}$	19/19	$0{,}96_{\pm 0{,}02}$	19/19	0,59	11/19
Catboost	$0{,}97_{\pm 0{,}01}$	19/19	$0{,}89_{\pm 0{,}02}$	19/19	0,44	10/19
Xgboost	$0{,}96_{\pm 0{,}01}$	19/19	$0{,}88_{\pm 0{,}02}$	19/19	0,41	8/19
LightGBM	$0{,}95_{\pm 0{,}01}$	19/19	$0{,}88_{\pm 0{,}02}$	19/19	0,44	10/19

Оценка значимости различий между моделями. Стандартные отклонения $\mathrm{F1}_{\mathrm{weighted}}$ по фрагментам кросс-валидации не превышают $0{,}02$ для всех моделей, что свидетельствует о стабильности оценок. Различие между MLP и моделями градиентного бустинга на валидационной подвыборке составляет $0{,}09$–$0{,}10$ (при $\mathrm{std} \leqslant 0{,}02$), что превышает четыре стандартных отклонения и указывает на содержательную значимость преимущества MLP. На тестовой выборке разрыв ещё более выражен: $\mathrm{F1} = 0{,}69$ для MLP против $0{,}41$–$0{,}44$ для бустинговых моделей. Различия между тремя моделями градиентного бустинга (CatBoost, XGBoost, LightGBM) на валидации и тесте лежат в пределах наблюдаемой вариации ($\leqslant 0{,}03$) и не являются статистически значимыми.

Дополнительно были построены матрицы ошибок (confusion matrix) и ROC-кривые в подходе One-vs-Rest для всех моделей на тестовом наборе. Матрицы ошибок и ROC-кривые для лучшей модели (MLP) приведены на рисунках 2 и 3 соответственно.

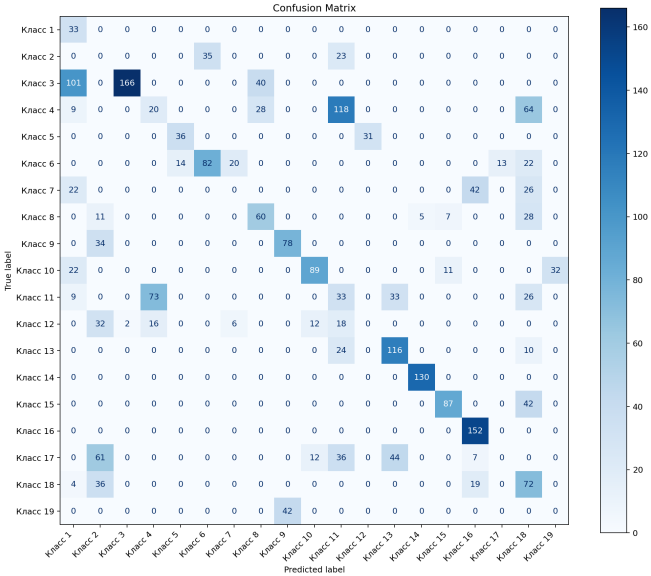

Рис. 2. Матрица ошибок для модели MLP на тестовом наборе

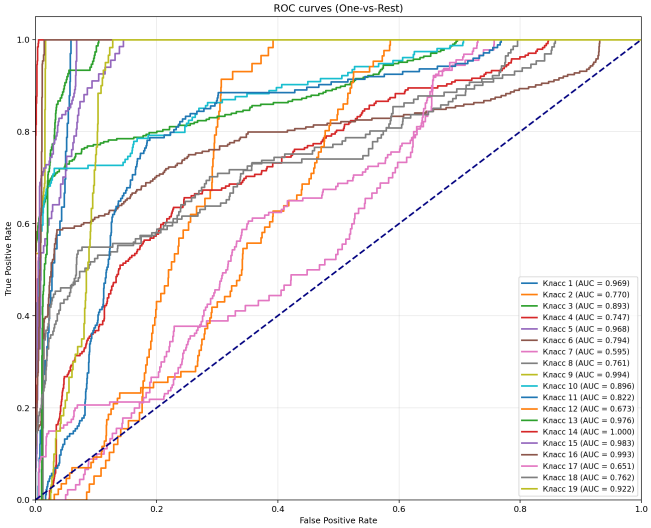

Рис. 3. ROC-кривые One-vs-Rest для модели MLP на тестовом наборе

Причины снижения качества на тестовой выборке. Снижение качества на тестовой выборке по сравнению с обучающей и валидационной (таблица 3) обусловлено тем, что валидация является внутрисессионной: она формируется из последних 30% кадров тех же двух видеозаписей, которые используются для обучения, поэтому распределения обучающих и валидационных данных близки. Тестовая выборка соответствует отдельной (третьей) сессии, выполненной через 2–5 недель, что приводит к межсессионному смещению распределения (изменение внешнего вида, небольшие изменения условий съёмки и установки камеры, вариативность темпа и манеры ходьбы); такие различия являются одной из основных причин деградации качества распознавания походки [6].

Дополнительным фактором выступает ограниченный объём данных (19 классов, по две обучающие записи на класс), из-за чего модель может частично подстраиваться под статистику конкретных сессий (характер шумов 3D-реконструкции, типичные пропуски ключевых точек, небольшие артефакты детекции), сохраняя высокие значения метрик на внутрисессионной валидации, но теряя качество при переносе на межсессионный тест.

Причины низкой производительности LSTM в сравнении с MLP. Обе модели обучались на одних и тех же оконных представлениях признаков и более низкое качество LSTM, по сравнению с MLP, может быть связано с тем, что рекуррентная архитектура оказалась избыточной и хуже «настроенной» под имеющийся объём данных. Признаки уже содержат существенную часть временной информации, поэтому дополнительное моделирование последовательности внутри окна даёт ограниченный выигрыш, а иногда и ухудшает обобщающую способность. При наличии локально шумных участков (ошибки трёхмерной реконструкции, несовершенство алгоритма заполнения пропусков) рекуррентная модель может сильнее накапливать влияние этих ошибок в скрытом состоянии, тогда как MLP, работающий с вектором окна, оказывается устойчивее.

Сравнение методов заполнения пропусков. Для основной модели MLP выполнено сравнение точности классификации в зависимости от выбора способа заполнения пропущенных значений. Результаты по показателю TP (число корректно идентифицированных личностей из 19 на обучающей выборке) приведены в таблице 4.

Таблица 4. Количество верно идентифицированных личностей (TP) для разных методов заполнения пропусков

Метод заполнения	TP (Обучающая выборка)
Forward fill	19/19
Кубическая сплайн-интерполяция	18/19
KNN-заполнение	18/19
Скользящее среднее по времени	19/19
Итеративное модельное заполнение	17/19

5. Сравнение производительности на различном аппаратном обеспечении. Оценка производительности системы проводилась при различных вычислительных ресурсах. Тестирование выполнялось на следующих платформах: NVIDIA A100 (80 ГБ), NVIDIA RTX 4070 Ti SUPER, NVIDIA Jetson Orin AGX 64 и NVIDIA Jetson TX2. Такой выбор позволяет оценить работу системы как на высокопроизводительных серверных решениях, так и на энергоэффективных встраиваемых вычислительных устройствах.

Измерения включали временные затраты на ключевые этапы обработки: вычисление признаков, параметризацию данных и выполнение процедуры вывода модели. На платформе NVIDIA Jetson TX2 этап подготовки данных не выполнялся, поскольку из-за ограниченного объёма внутренней памяти отсутствовала возможность загрузить стереовидеозаписи обучающей выборки. Все испытания проводились на одном и том же наборе данных для двух видеозаписей при фиксированных гиперпараметрах базовой модели, что обеспечивало сопоставимость результатов. Полученные значения времени обработки приведены в таблице 5 для каждого из тестируемых устройств. Анализ результатов позволяет выявить зависимость производительности системы от аппаратных характеристик используемой платформы.

Таблица 5. Скорость системы на устройствах разного типа

Устройства	Временные затраты		
	Подготовка данных, сек	Обучение модели на 19 классах, сек	Тестирование модели на 19 классах, сек
RTX 4070 TI SUPER	29	14	2
A100 80GB	43	31	11
Nvidia Jetson Orin AGX 64	62	35	13
Nvidia Jetson TX2	–	76	34

Программная реализация и окружение. Программная реализация выполнена на Ubuntu 22.04 с использованием Python 3.10. Для обеспечения воспроизводимости ниже приведены версии основных библиотек, использованных в экспериментальной части (обработка видеоданных, оценка позы, построение признаков и обучение моделей).

– **Базовые научные библиотеки:** numpy==1.26.4, scipy==1.14.0, pandas==2.2.2, numba==0.61.2.

– **Компьютерное зрение и работа с данными:** opencv-python==4.10.0.84, scikit-image==0.25.2, imageio==2.37.0, pyrealsense2==2.55.1.6486, ultralytics==8.3.54, ultralytics-thop==2.0.14.

– **Машинное обучение:** scikit-learn==1.5.1, scikit-learn-intelex==2024.6.0, torch==1.13.1, torchvision==0.14.1, torchaudio==0.13.1, catboost==1.2.7, lightgbm==4.5.0, xgboost==3.0.0.

– **Визуализация и интерпретация:** matplotlib==3.10.1, seaborn==0.13.2, plotly==5.23.0, shap==0.47.2, lime==0.2.0.1, tqdm==4.67.1.

– **GPU-библиотеки:** nvidia-cublas-cu11==11.10.3.66, nvidia-cuda-nvrtc-cu11==11.7.99, nvidia-cuda-runtime-cu11==11.7.99, nvidia-cudnn-cu11==8.5.0.96, nvidia-nccl-cu12==2.26.5.

6. Сравнение с опубликованными подходами.

6.1. Метрики качества. В работах по идентификации личности по походке широко используется *точность ранга-k* (Rank-k) – доля испытаний, в которых истинная личность попала в список из k наиболее близких кандидатов. Частный случай Rank-1 совпадает с обычной точностью классификации (правильный ответ на первом месте), Rank-5 – правильный ответ в первой пятёрке.

Метод авторов (трёхмерная кинематика, 1 стереокамера).

Задача: Идентификация личности по походке (биометрия).
Съёмка и ракурсы: 1 стереокамера RealSense D457 (стерео + карта глубины), 1 ракурс на испытуемого, $\sim$225 с на человека, 3 сессии, межсессионная проверка.
Вход: Трёхмерный скелет и инвариантные геометрико-кинематические признаки; классификатор – многослойный перцептрон.
Инвариантность: Явно вводится инвариантность признаков к геометрическим преобразованиям (перенос/поворот системы координат) за счёт конструкции признаков.
Качество: Rank-1 $= 14/19$ (73.7%) в межсессионном режиме.

Соколова А., Конушин А. [6]: двумерный оптический поток и сверточная сеть.

Задача: Идентификация личности по походке; метрики: Rank-1/Rank-5.
Съёмка и ракурсы: CASIA-B (Chinese Academy of Sciences, Institute of Automation, Dataset B): 124 человека, 11 ракурсов (0–180° с шагом 18°), на каждого человека по 10 видеопоследовательностей на *каждый* ракурс (6 – обычная ходьба, 2 – с сумкой, 2 – в верхней одежде), длительность 3–5 с; в «обобщаемом» протоколе фиксируется ракурс 90° (обучение на 60 людях, тест на 64).
Вход: Двумерный оптический поток $\rightarrow$ сверточная сеть.
Инвариантность: Явная инвариантность к изменению ракурса в приведённом протоколе не обеспечивается (ракурс фиксируется).
Качество: TUM-GAID (TUM Gait from Audio, Image and Depth): Rank-1 $= 97.52\%$, Rank-5 $= 99.89\%$; CASIA-B (90°): Rank-1 $= 74.93\%$.

Bari A.S.M.H., Gavrilova M.L. [8]: KinectGaitNet, трёхмерные суставы (RGB-D) и сверточная сеть.

Задача: Идентификация личности по походке; метрика: Accuracy (эквивалент Rank-1 для классификации).
Съёмка и ракурсы: 1 сенсор Microsoft Kinect (RGB-D); трёхмерные координаты суставов по циклу шага.
Данные: UPCV (University of Patras Computer Vision) – 30 участников, по 5 последовательностей; KGB (Kinect Gait Biometry) – 164 участника, по 5 последовательностей.
Инвариантность: Кинематический подход (работа по трёхмерным суставам), однако явная геометрическая инвариантность (к переносу/повороту системы координат) отдельно не выделена.
Качество: 96.91% (UPCV) и 99.33% (KGB).

Mao M., Song Y. [16]: графовая сверточная сеть по трёхмерному скелету.

Задача: Идентификация личности по походке; метрика: Rank-1.
Съёмка и ракурсы: Многоракурсная постановка (в т.ч. CASIA-B с 11 ракурсами).
Вход: Трёхмерные координаты суставов и костей $\rightarrow$ графовая сверточная сеть; объединение признаков; оптимизация: Softmax + центрирующая функция потерь (center loss).
Инвариантность: Строго кинематический подход (скелет); устойчивость к ракурсу достигается нормировкой скелета и обучаемым представлением, без явного введения инвариантных скалярных признаков.

Качество: Rank-1 варьируется в диапазоне от 66,3% до 87,7% на CASIA-B в зависимости от условий съёмки, что свидетельствует о чувствительности к ракурсу и внешним факторам.

Luo J., Xu B., Tjahjadi T., Yi J. [5]: параметрическая трёхмерная модель походки (3DGait), устойчивая к сумкам/одежде.
Задача: Идентификация личности по походке; основная метрика: Rank-1.
Съёмка и ракурсы: Оценка на CASIA-B (11 ракурсов) и др. наборах; в статье отдельно рассматриваются многоракурсная идентификация и вариации условий (сумка/одежда).
Вход: Дескрипторы 3DGait: форма (3D-Shape), поза/движение (3D-Pose) и внешние факторы (3D-eFactors); устойчивость к внешним условиям задаётся параметрической моделью и обучением отображения 2D→3D-дескрипторов.
Инвариантность: Авторы формулируют использование интерпретируемых трёхмерных дескрипторов как инвариантных признаков для устойчивости к переносимым предметам и вариативности одежды.
Качество (средние): CASIA-B, многоракурсная идентификация (среднее по тестовым ракурсам): 89,8%; CASIA-B, с вариациями условий: 86,5%.

Итоговые выводы. В отличие от большинства рассмотренных работ, наш метод изначально разрабатывался под применение стереокамеры, позволяющей получать трёхмерные координаты и строить признаки, устойчивые к геометрическим преобразованиям. Мы используем вычислительно экономичную архитектуру (классификатор на многослойном перцептроне), тогда как в ряде трёхмерных подходов применяются более тяжёлые нейросетевые методы. С точки зрения условий съёмки предлагаемый нами сценарий сложнее по ограничению наблюдения: одна камера и один ракурс на испытуемого, тогда как в типовых публичных наборах для сравнения часто доступны многоракурсные наблюдения и/или многокамерная регистрация. При этом наши записи на человека относительно длинные, но число сессий/примеров ограничено, что делает задачу более чувствительной к межсессионным изменениям и статистической недостаточности данных. Отдельно важно, что в исследовании оценка проводится именно в межсессионном режиме, и проблема деградации качества при смене сессии/условий наблюдается и в опубликованных работах (например, в экспериментах с временным разрывом на TUM GAID точность Rank-1 заметно ниже).

7. Заключение. В работе разработана и экспериментально исследована система идентификации личности по походке на основе

безмаркерной реконструкции трёхмерных координат ключевых точек и последующего построения инвариантного пространства признаков. Предложенный подход ориентирован на использование минимального набора оборудования (одна стереокамера) и на извлечение именно кинематической информации, без опоры на внешние признаки (лицо, одежда, фактура и т. п.), которые могут неявно повышать качество распознавания в лабораторных условиях, но снижают воспроизводимость и переносимость метода.

Ключевые научные выводы работы состоят в следующем:

– Показано, что прямое использование сглаженных трёхмерных координат ключевых точек без явной параметризации (окна из координат $\mathbf{W}_t$ на входе классификатора) обеспечивает точность не выше 25%, что указывает на недостаточность такого представления для надёжной идентификации и мотивирует переход к явному построению признаков.

– Построение геометрико-кинематических признаков, инвариантных к ортогональным преобразованиям и переносам, позволяет существенно повысить качество идентификации по сравнению с базисом из «сырых» координат.

– Использование скользящих окон признаков и последующего агрегирования решений по видеозаписи повышает устойчивость итогового результата по сравнению с покадровыми решениями, поскольку сглаживает влияние случайных ошибок детектора позы и локальных артефактов траекторий.

– Проведено сравнение нескольких моделей машинного обучения (градиентный бустинг и нейросетевые модели) и выполнена оценка вычислительных затрат на разных аппаратных платформах, что позволяет выбирать конфигурацию системы в зависимости от доступных вычислительных ресурсов и требований к времени обработки.

Дальнейшие исследования предполагают расширение набора данных за счёт дополнительных видеозаписей для каждого испытуемого, а также повышение точности на тестовой выборке посредством внедрения архитектур на основе трансформеров [21] и дополнительного обучения модели компьютерного зрения. В качестве направления развития рассматривается применение модели разметки с расширенным набором ключевых точек, описывающих положение и позу человека в пространстве (например, 35 точек вместо используемых 17) [22]. Кроме того, планируется переход от чисто классификационной постановки с кросс-энтропийной функцией потерь к обучению векторных представлений походки в скрытом пространстве, что позволит масштабировать число идентифицируемых персон без изменения архитектуры нейросетевой

модели и эффективнее использовать существенно более крупные выборки; при этом возможно расширение числа классов на основе таких векторных представлений без полного переобучения модели [23].

Литература

1. Ramesh S.H., Lemaire E.D., Tu A., Cheung K., Baddour N. Automated Implementation of the Edinburgh Visual Gait Score (EVGS) Using OpenPose and Handheld Smartphone Video // Sensors. 2023. vol. 23. pp. 1–36.
2. Moro M., Marchesi G., Hesse F., Odone F., Casadio M. Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study // Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22052011.
3. Han X., Guffanti D., Brunete M.A. A Comprehensive Review of Vision-Based Sensor Systems for Human Gait Analysis // Sensors. 2025. vol. 25. pp. 1–33. DOI: 10.3390/s25020498.
4. Sapkota R., Karkee M. YOLO11 and Vision Transformers Based 3D Pose Estimation of Immature Green Fruits in Commercial Apple Orchards for Robotic Thinning // arXiv preprint arXiv:2410.19846. 2024. pp. 1–18.
5. Luo J., Xu B., Tjahjadi T., Yi J. A Novel 3D Gait Model for Subject Identification Robust against Carrying and Dressing Variations // Computers, Materials Continua. 2024. vol. 82. no. 1. pp. 194–215. DOI: 10.32604/cmc.2024.050018.
6. Sokolova A., Konushin A. Gait Recognition Based on Convolutional Neural Networks // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017. vol. XLII-2/W4. pp. 207–212. DOI: 10.5194/isprs-archives-XLII-2-W4-207-2017.
7. Natraj S., Messmer T., Fujii Y., Suzuki K., Riener R., Eriks-Hoogland I., Paez-Granados D. 3D Pose Estimation for Scalable Remote Gait Kinematics Assessment // npj Digital Medicine. 2026. vol. 9. no. 37. DOI: 10.1038/s41746-025-02211-y.
8. Bari A.S.M.H., Gavrilova M.L. KinectGaitNet: Kinect-Based Gait Recognition Using Deep Convolutional Neural Network // Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22072631.
9. Cai S., Xu H., Cai W., et al. A human pose estimation network based on YOLOv8 with efficient multi-scale receptive field and expanded feature pyramid network // Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-00259-0.
10. Kwon J., Lee Y., Lee J. Comparative Study of Markerless Vision-Based Gait Analyses for Person Re-Identification // Sensors. 2021. vol. 21. pp. 1–25. DOI: 10.3390/s21248208.
11. Ultralytics YOLO11 // GitHub. URL: www.github.com/ultralytics/ultralytics (дата обращения: 10.11.2024).
12. Intel RealSense SDK 2.0 (librealsense) // GitHub URL: www.github.com/IntelRealSense/librealsense.
13. Yoon H., Jo E., Ryu S., Yoo J.-I., Kim M., Kim J.H. Noise-robust markerless video gait anomaly detection via two-stage acquisition and LSTM autoencoders // Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-26169-9.
14. Huo Z., Ji T., et al. DynImp: Dynamic Imputation for Wearable Sensing Data Through Sensory and Temporal Relatedness // arXiv preprint arXiv:2209.15415. 2022. pp. 1–25.
15. Jun K., Lee K., Lee S., Lee H., Kim M.-K. Hybrid Deep Neural Network Framework Combining Skeleton and Gait Features for Pathological Gait Recognition // Bioengineering. 2023. vol. 10. 1133 p. DOI: 10.3390/bioengineering10101133.

16. Mao M., Song Y. Gait Recognition Based on 3D Skeleton Data and Graph Convolutional Network // IEEE International Joint Conference on Biometrics (IJCB). 2020. DOI: 10.1109/IJCB48548.2020.9304916.
17. Rao H., Miao C. A Survey on 3D Skeleton-Based Person Re-Identification: Approaches, Designs, Challenges, and Future Directions // arXiv preprint arXiv:2401.15296. 2024.
18. Khaliluzzaman M., Uddin A., et al. Person Recognition Based on Deep Gait: A Survey // Sensors. 2023. vol. 23. pp. 1–36. DOI: 10.3390/s23104875.
19. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal Loss for Dense Object Detection // arXiv preprint arXiv:1708.02002. 2017. pp. 1–10.
20. Yun S., Jeong M., Kim R., Kang J., Kim H.J. Graph Transformer Networks // arXiv preprint arXiv:1911.06455. 2019. pp. 1–11.
21. Shi L.-F., Liu Z.-Y., Zhou K.-J., Shi Y., Jing X. Novel Deep Learning Network for Gait Recognition Using Multimodal Inertial Sensors // Sensors. 2023. vol. 23. pp. 1–17. DOI: 10.3390/s23020849.
22. Maji D., Nagori S., Mathew M., Poddar D. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss // arXiv preprint arXiv:2204.06806. 2022. pp. 1–10.
23. Netay I.V. Series of quasi-uniform scatterings with fast search, root systems and neural network classifications // arXiv preprint arXiv:2512.04865. 2025. pp. 1–13.

Нетай Игорь Витальевич — канд. физ.-мат. наук; старший научный сотрудник ИППИ РАН. Область научных интересов: алгебра, алгебраическая геометрия, теория представлений, теория инвариантов, численные методы, машинное обучение и нейросети, цифровая обработка сигналов, оптимизация вычислений. Число научных публикаций — 15. i.netay@kryptonite.ru, www.iitp.ru/ru/about; ИППИ РАН, Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, РФ; р.т. +7(903)778-6644.

Панкратов Евгений Павлович — младший научный сотрудник ИППИ РАН. Область научных интересов: машинное обучение, обработка естественного языка, большие языковые модели, нейросетевое моделирование. Число научных публикаций — 2. pankratov.ep@phystech.edu, www.iitp.ru/ru/about; ИППИ РАН, Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, РФ; р.т. +7(977)424-9237.

Корнилов Константин Игоревич — стажер-исследователь ИППИ РАН. Область научных интересов: генеративные модели, прогнозирование временных рядов, прикладные исследования больших языковых моделей. Число научных публикаций — 1. kornilov.ki@phystech.edu, www.iitp.ru/ru/about; ИППИ РАН, Большой Каретный переулок, д. 19, стр. 1, г. Москва, 127051, РФ; р.т. +7(910)531-8381.

Голубинский Андрей Николаевич — д-р техн. наук, доцент; начальник отдела, Российский научный фонд. Область научных интересов: машинное обучение, нейросетевое моделирование, автоматизированные системы управления с элементами искусственного интеллекта, обработка речевых сигналов. Число научных публикаций — 250. annikgol@mail.ru, www.rscf.ru; ул. Солянка, д. 14, стр. 3, г. Москва, 109240, РФ; р.т. +7(910)346-6537.

DOI 10.15622/ia.25.3.5

I. V. NETAY , E. P. PANKRATOV , K. I. KORNILOV , A. N. GOLUBINSKIY

IDENTIFICATION OF A PERSON BY GAIT KINEMATICS IN THREE-DIMENSIONAL SPACE BASED ON INVARIANT FEATURES

Netay I. V., Pankratov E. P., Kornilov K. I., Golubinskiy A. N. **Identification of a Person by Gait Kinematics in Three-Dimensional Space Based on Invariant Features.**

Abstract. A system for data acquisition, preprocessing, and analysis for gait-based personal identification using machine learning and image analysis methods is proposed. The system includes a preprocessing mechanism for normalizing biometric motion data and a post-processing stage for extracting features invariant to orthogonal transformations of space and changes in camera pose, including perspective effects. A computationally efficient gait recognition method using a single stereo camera is implemented. The feature space is analyzed to determine the most informative characteristics, and the employed feature set is comparatively assessed. Several machine-learning architectures, including deep learning models, are evaluated with respect to both accuracy and computational efficiency. System performance is experimentally studied on various computing devices, and processing time characteristics are measured. The results enable comparison of applicability and efficiency depending on available computational resources.

The input data are formed as time series of 3D coordinates of keypoints reconstructed from depth data using 2D pose detections (COCO, 17 keypoints), where missing detections are encoded by a zero vector. For gait identification, head keypoints are excluded as uninformative, while missing values are filled and the resulting trajectories are subsequently smoothed. An invariant feature space is built from the skeleton geometry (segment lengths, joint angles, and their derivatives); then these features are aggregated into a single vector using a sliding temporal window, i.e., each vector describes motion over an interval of several consecutive frames. At the identification stage, threshold-based filtering and aggregation of predictions at the frame level are applied. It is shown that both accuracy and processing speed depend on the chosen machine-learning model and the available computational resources.

Keywords: machine learning, computer vision, biometrics, neural networks, kinematics, 3D video processing, invariant features

References

1. Ramesh S.H., Lemaire E.D., Tu A., Cheung K., Baddour N. Automated Implementation of the Edinburgh Visual Gait Score (EVGS) Using OpenPose and Handheld Smartphone Video. Sensors. 2023. vol. 23. pp. 1–36.
2. Moro M., Marchesi G., Hesse F., Odone F., Casadio M. Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study. Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22052011.
3. Han X., Guffanti D., Brunete M.A. A Comprehensive Review of Vision-Based Sensor Systems for Human Gait Analysis. Sensors. 2025. vol. 25. pp. 1–33. DOI: 10.3390/s25020498.
4. Sapkota R., Karkee M. YOLO11 and Vision Transformers Based 3D Pose Estimation of Immature Green Fruits in Commercial Apple Orchards for Robotic Thinning. arXiv preprint arXiv:2410.19846. 2024. pp. 1–18.
5. Luo J., Xu B., Tjahjadi T., Yi J. A Novel 3D Gait Model for Subject Identification Robust against Carrying and Dressing Variations. Computers, Materials Continua. 2024. vol. 82. no. 1. pp. 194–215. DOI: 10.32604/cmc.2024.050018.

6. Sokolova A., Konushin A. Gait Recognition Based on Convolutional Neural Networks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2017. vol. XLII-2/W4. pp. 207–212. DOI: 10.5194/isprs-archives-XLII-2-W4-207-2017.
7. Natraj S., Messmer T., Fujii Y., Suzuki K., Riener R., Eriks-Hoogland I., Paez-Granados D. 3D Pose Estimation for Scalable Remote Gait Kinematics Assessment. npj Digital Medicine. 2026. vol. 9. no. 37. DOI: 10.1038/s41746-025-02211-y.
8. Bari A.S.M.H., Gavrilova M.L. KinectGaitNet: Kinect-Based Gait Recognition Using Deep Convolutional Neural Network. Sensors. 2022. vol. 22. pp. 1–15. DOI: 10.3390/s22072631.
9. Cai S., Xu H., Cai W., et al. A human pose estimation network based on YOLOv8 with efficient multi-scale receptive field and expanded feature pyramid network. Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-00259-0.
10. Kwon J., Lee Y., Lee J. Comparative Study of Markerless Vision-Based Gait Analyses for Person Re-Identification. Sensors. 2021. vol. 21. pp. 1–25. DOI: 10.3390/s21248208.
11. Ultralytics YOLO11. GitHub. Available at: www.github.com/ultralytics/ultralytics (accessed 10.11.2024).
12. Intel RealSense SDK 2.0 (librealsense). GitHub. Available at: www.github.com/IntelRealSense/librealsense.
13. Yoon H., Jo E., Ryu S., Yoo J.-I., Kim M., Kim J.H. Noise-robust markerless video gait anomaly detection via two-stage acquisition and LSTM autoencoders. Scientific Reports. 2025. vol. 15. DOI: 10.1038/s41598-025-26169-9.
14. Huo Z., Ji T., et al. DynImp: Dynamic Imputation for Wearable Sensing Data Through Sensory and Temporal Relatedness. arXiv preprint arXiv:2209.15415. 2022. pp. 1–25.
15. Jun K., Lee K., Lee S., Lee H., Kim M.-K. Hybrid Deep Neural Network Framework Combining Skeleton and Gait Features for Pathological Gait Recognition. Bioengineering. 2023. vol. 10. 1133 p. DOI: 10.3390/bioengineering10101133.
16. Mao M., Song Y. Gait Recognition Based on 3D Skeleton Data and Graph Convolutional Network. IEEE International Joint Conference on Biometrics (IJCB). 2020. DOI: 10.1109/IJCB48548.2020.9304916.
17. Rao H., Miao C. A Survey on 3D Skeleton-Based Person Re-Identification: Approaches, Designs, Challenges, and Future Directions. arXiv preprint arXiv:2401.15296. 2024.
18. Khaliluzzaman M., Uddin A., et al. Person Recognition Based on Deep Gait: A Survey. Sensors. 2023. vol. 23. pp. 1–36. DOI: 10.3390/s23104875.
19. Lin T.-Y., Goyal P., Girshick R., He K., Dollar P. Focal Loss for Dense Object Detection. arXiv preprint arXiv:1708.02002. 2017. pp. 1–10.
20. Yun S., Jeong M., Kim R., Kang J., Kim H.J. Graph Transformer Networks. arXiv preprint arXiv:1911.06455. 2019. pp. 1–11.
21. Shi L.-F., Liu Z.-Y., Zhou K.-J., Shi Y., Jing X. Novel Deep Learning Network for Gait Recognition Using Multimodal Inertial Sensors. Sensors. 2023. vol. 23. pp. 1–17. DOI: 10.3390/s23020849.
22. Maji D., Nagori S., Mathew M., Poddar D. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. arXiv preprint arXiv:2204.06806. 2022. pp. 1–10.
23. Netay I.V. Series of quasi-uniform scatterings with fast search, root systems and neural network classifications. arXiv preprint arXiv:2512.04865. 2025. pp. 1–13.

Netay Igor Vitalievich — Ph.D., Associate professor, Senior researcher, Federal State Budgetary Institution of Science A.A. Kharkevich Institute for Information Transmission Problems of the

Russian Academy of Sciences (IPPI RAS). Research interests: algebra, algebraic geometry, representation theory, invariant theory, numerical methods, machine learning and neural networks, digital signal processing, computational optimization. The number of publications — 15.
i.netay@kryptonite.ru; 19, Bolshoy Karetny Lane, 127051, Moscow, Russia; office phone: +7(903)778-6644.

Pankratov Evgeny Pavlovich — Junior research assistant, Federal State Budgetary Institution of Science A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences (IPPI RAS). Research interests: machine learning, natural language processing, large language models, neural network modeling. The number of publications — 2.
pankratov.ep@phystech.edu; 19, Bolshoy Karetny Lane, 127051, Moscow, Russia; office phone: +7(977)424-9237.

Kornilov Konstantin Igorevich — Intern researcher, Federal State Budgetary Institution of Science A.A. Kharkevich Institute for Information Transmission Problems of the Russian Academy of Sciences (IPPI RAS). Research interests: generative models, time series forecasting, applied research of large language models. The number of publications — 1.
kornilov.ki@phystech.edu; 19, Bolshoy Karetny Lane, 127051, Moscow, Russia; office phone: +7(910)531-8381.

Golubinskiy Andrey Nikolaevich — Ph.D., Dr. Sci., Associate professor, Head of Department, Russian Science Foundation (RSF). Research interests: machine learning, neural network modeling, automated control systems with artificial intelligence elements, speech signal processing. The number of publications — 250.
annikgol@mail.ru; 14, Solyanka St., 109240, Moscow, Russia; office phone: +7(910)346-6537.