УДК 004.722

Биометрическая идентификация, основанная на походке

Богданов М. Р., Шахмаметова Г. Р., Оськин Н. Н., Корнилов А. В.

***Постановка задачи**: рост экстремизма во многих странах мира делает актуальной проблему бесконтактной биометрической идентификации. Решение этой проблемы осложняется тем фактом, что участники беспорядков используют различные средства маскировки. В этих условиях биометрическая идентификация, основанная на распознавании лиц, становится неэффективной. В этой связи распознавание походки имеет ряд преимуществ, т.к. походку можно распознавать на большом расстоянии, ее трудно подделать, и она медленно изменяется с течением времени. Для распознавания походки можно использовать различные источники данных, например, видеопоток, звуки шагов или показания инерциальных датчиков мобильных телефонов, находящихся в карманах испытуемых. **Целью работы** является изучение факторов, влияющих на эффективность биометрической идентификации по походке, когда в качестве источника данных является видеопоток. **Используемые методы**: в качестве источника данных используется видеопоток. Из видеопотока извлекали ключевые точки тела человека с помощью нейронных сетей MoveNet (Lightning и Thunder), PoseNet и BlazePose (Lite и Heavy). Классификацию признаков проводили с помощью случайного леса и полносвязной нейронной сети. **Новизна:** для биометрической идентификации по походке были использованы методы, применяемые для решения задачи оценки позы. Изучено влияние на эффективность распознавания походки таких факторов, как разрешение видео, кадровая частота, алгоритм извлечения признаков из видеопотока и продолжительность съемки. **Результат**: было установлено, что: изменение частоты кадров и разрешения видео дает больший эффект при использовании моделей, ориентированных на высокую скорость в ущерб точности: для нейронных сетей MoveNet в версии Lightning и BlazePose в версии Lite, лучший результат достигается при частоте 60 FPS, худший – при 24 FPS, увеличение частоты кадров до 120 FPS не улучшает результат по сравнению с 60 FPS, разрешение видеопотока 1920 на 1080 позволяет получить лучшие результаты по сравнению с 3840 на 2160. Ансамблевые методы (случайный лес и градиентный бустинг показали более высокую эффективность классификации по сравнению с полносвязной нейронной сетью. **Практическая значимость**: было разработано программное обеспечение, предназначенное для биометрической идентификации по походке. Его можно применять в системах безопасности для биометрической идентификации и выявления паттернов агрессивного поведения.*

***Ключевые слова**: бесконтактная биометрическая идентификация, машинное обучение, информационная безопасность.*

Актуальность

Рост экстремизма во многих странах мира ставит перед правоохранительными органами новые задачи. С одной стороны, повсеместное внедрение камер видеонаблюдения делает потенциально возможным создание безопасной городской среды, с другой стороны нарушители порядка часто используют различные средства маскировки, такие как маски и балаклавы.

Библиографическая ссылка на статью:
Богданов М. Р., Шахмаметова Г. Р., Оськин Н. Н., Корнилов А. В. Биометрическая идентификация, основанная на походке // Системы управления, связи и безопасности. 2024. № 1. С. 191-212. DOI: 10.24412/2410-9916-2024-1-191-212

Reference for citation:
Bogdanov M. R., Shakhmametova G. R., Oskin N. N., Kornilov A. V. Biometric identification based on gait. *Systems of Control, Communication and Security*, 2024, no. 1, pp. 191-212 (in Russian). DOI: 10.24412/2410-9916-2024-1-191-212

Актуальной становится задача бесконтактной биометрической идентификации, устойчивой к применению средств маскировки. Одним из методов поведенческой биометрической идентификации является распознавание походки. Походка человека медленно меняется с течением времени, ее трудно скрыть или подделать. Походку можно распознавать на значительном расстоянии.

В качестве потенциальных источников данных о походке можно использовать видеопоток, звуки шагов или показания инерциальных датчиков сотовых телефонов, находящихся в карманах испытуемых. В предлагаемой работе для распознавания походки мы использовали последовательность видеокадров. В этом случае необходимо с одной стороны выделить признаки из силуэтов человека на кадрах, с другой стороны – решить задачу классификации этих признаков.

Решение проблемы бесконтактной идентификации по походке давно находится в фокусе внимания различных спецслужб. Так, в сентябре 2000 г. был запущен проект Управления перспективных исследовательских проектов Министерства обороны США HumanID, направленный на разработку автоматизированных мультимодальных биометрических технологий, способных обнаруживать, распознавать и идентифицировать людей на расстоянии [1] с использованием кинематики человека. Упор при этом делался на способность обнаруживать и идентифицировать людей, не желающих сотрудничать, в любой время дня и ночи, при любых погодных условиях, в условиях применения маскировки, действующих в одиночку или в составе групп.

Прогресс в области обнаружения человека на видео связан с успехом в области технологий компьютерного зрения. В начале 2000-х гг. ошибка распознавания изображений превышала 25%. Ситуация изменилась, когда в 2006 г. профессор университета в городе Урбана-Шампейн, штат Иллинойс Фей-Фей Ли предложила идею создания большого набора данных ImageNet. Ее идеи в чем-то перекликаются с работами Джорджа А. Миллера из Принстонского университета, который в 1985 г. начал работу над WordNet, лексической базой данных английского языка. Будучи чем-то средним между словарем и тезаурусом, она позволяла разрабатывать приложения для обработки естественного языка (natural language processing, NLP). В то время большинство исследователей считали, что алгоритмы важнее данных. Однако Ли была убеждена, что огромные объемы реальных данных сделают алгоритмы более точными. После встречи с членом проекта WordNet Кристианой Феллбаум, Ли решила использовать словесную базу и иерархию WordNet для своей базы данных изображений. Целью была разработка программного обеспечения для визуального распознавания объектов. При разработке ImageNet основной упор предполагалось делать не на модели, а на данные.

В июле 2008 г. в ImageNet не было изображений. К декабрю компания классифицировала 3 млн изображений по более чем 6000 классов. В апреле 2010 г. было более 11 млн изображений в более чем 15 тыс. классах. Работа стала возможной благодаря краудсорсингу на платформе Amazon Mechanical Turk.

В 2010 г. был организован первый конкурс ImageNet по крупномасштабному визуальному распознаванию ImageNet large scale visual recognition challenge (ILSVRC). Разработчики программ соревновалось в правильной классификации и обнаружении объектов и сцен. В настоящее время база данных содержит 14 млн аннотированных изображений.

В 2012 г. Алекс Крижевский предложил нейронную сеть AlexNet, состоящую из 5 сверточных и 3-х полносвязных слоев [3]. Сеть имела 60 млн параметров. Уровень ошибок снизился до 15,3%.

В 2013 г. профессор Нью-Йоркского университета Роб Фергюс и его студент Мэтью Д. Зейлер разработали сеть ZFNet с уровнем ошибок 11,2% [4].

ResNet была создана исследовательской командой Microsoft и выиграла конкурс ImageNet 2015 г. с уровнем ошибок 3,57% [5]. Нужно сказать, что уровень ошибки человека составляет 5,1%.

Дальнейший прогресс в области выделения объектов в визуальной сцене связан с разработкой набора данных Microsoft Common Objects in Context (COCO) и конкурсом COCO Challenge. Спецификой набора данных является то, что объекты находятся в реальном окружении, часть из них видна не полностью или выглядят неоднозначно. В 2018 г. в рамках конкурса COCO Challenge перед разработчиками ПО ставилась задача обнаружения ключевых точек человека в сложной неконтролируемой среде.

Цель исследования

Изучение факторов, влияющих на эффективность биометрической идентификации по походке при использовании видеопотока в качестве источника входных данных.

Предшествующие исследования

Использование циклически согласованных генеративно-состязательных сетей для распознавания походки описано в работе [6]. Работа проводилось на общедоступном наборе данных CASIA-B, предоставленном Китайской академией наук [7]. Угол обзора варьировался от 0° до 180° а испытуемые могли ходить с сумкой в руках или пальто. Выделение силуэтов в кадре осуществляется путем вычитания фона. После этого все изображения силуэтов конкретного человека усреднялись и выравнивались. Авторы добились точности 79%.

В 2005 г. был анонсирован проект HumanID Gait Challenge, включавший в себя набор данных, полученных в 12 экспериментах, и базовый алгоритм распознавания походки [8]. Базовый алгоритм извлекал человеческие силуэты путем вычитания фона и временной корреляции. В ходе экспериментов изучалось влияние на точность распознавания таких факторов, как угол обзора, тип обуви, наличие или отсутствие портфеля в руках, тип поверхности, а также время, прошедшее между сравниваемыми последовательностями. Точность распознавания варьировалась от 78% до 3% в зависимости от условий эксперимента. Наибольшее влияние оказывало напольное покрытие.

Интересной академической задачей является оценка позы человека (human pose estimation, HPE), позволяющая определять координаты ключевых точек человеческого тела, определяющих положение человеческого тела с точки зрения биомеханики (голова, руки, туловище и т.д.). Человеческое тело можно смоделировать в виде скелета, контура и объема. При этом ставится задача локализации суставов человеческого тела (например, коленей и запястий), также известных как ключевые точки, на изображениях или видео. В разных работах используется трехмерное (в виде облака точек), двухмерное (а виде совокупности прямоугольников) и одномерное (в виде шаростержневой модели) представление тела человека. Используя ключевые точки, создается скелетоподобное описание человеческого тела. Для решения проблемы HPE обычно используются методы компьютерного зрения.

В 2014 г. Тошев и соавторы впервые использовали для оценки позы человека сверточную нейронную сеть. Авторы назвали свою технику DeepPose [9].

В 2019 г. Чжэ Цао и др. предложили нейронную сеть OpenPose, которая сначала идентифицирует части тела или ключевые точки на изображении, а затем сопоставляет ключевые точки для формирования пар. Шаблоны извлекаются с использованием сверточной сети VGG-19 [10].

В 2016 г. Леонид Пищулин и соавторы предложили нейросеть DeepCut, которая могла бы решить задачу распознавания и определения положения тела [11]. Алгоритм распознает на изображении все возможные части тела с дальнейшей их маркировкой (голова, руки, ноги и так далее), тем самым получая скелетную модель человека.

Интересную технологию оценки позы предложила компания Google. Ее нейронная сеть MoveNet использует тепловые карты для точной локализации ключевых точек человека [12]. MoveNet обучалась на двух наборах данных: COCO и внутреннем наборе данных Google под названием Active. Нейронная сеть работает стабильно при очень высокой частоте кадров.

Существует ряд направлений технологий распознавания походки, например, персональные тренеры на базе искусственного интеллекта (Zenia – приложение для йоги), робототехника (хватательные движения и дополненная реальность), компьютерная графика в кинематографии (визуализация фантастических существ), распознавание поз спортсменов (анализ и изучение сильных и слабых сторон противника), отслеживание движений для игр (Microsoft Kinect использует трехмерное определение положения тела для отслеживания движений игроков), анализ движений младенца (анализ поведения ребенка для оценки его физического развития).

В исследовании [13] авторы анализировали походку человека с помощью уравнения Лагранжа.

Современные системы распознавания походки описывают каждый кадр видеоизображения с помощью дескрипторов, описывающих либо человека в целом, либо его отдельные части [14]. Для этой цели была разработана сверточная нейронная сеть GLConv, сочетающая в себе глобальные и локальные функции.

В исследовании [15] изучены временные характеристики движений суставов. Для распознавания походки авторы предложили подход, основанный на сверточной сети графов (graph convolutional networks, GCN).

Было обнаружено, что на точность распознавания сильно влияет одежда и способ ее ношения [16].

В исследовании [17] авторы предложили использовать модель MMGaitFormer, объединяющую пространственно-временную информацию скелетов и силуэтов.

В исследовании [18] авторы разработали свертку графов (frame-level topology refinement graph convolution, FTR-GC) для моделирования корреляций между соединениями на основе каждого кадра.

В исследовании [19] авторы предложили систему распознавания походки под названием Temporal Attention and Keypoint Embedding (GaitTAKE), которая объединяет глобальные и локальные особенности внешнего вида на основе временного внимания и временных агрегированных особенностей позы человека.

В исследовании [20] авторы предложили модель GaitMixer для изучения более различительного представления походки на основе данных последовательностей скелета.

В настоящем исследовании предлагается использовать технологии оценки позы человека для биометрической идентификации по походке.

Наборы данных походки

Для совершенствования алгоритмов распознавания позы человека был разработан ряд наборов данных [21]. Одним из них является CASIA-B. Набор данных содержит 13 640 образцов походки 124 испытуемых в разном контексте (разные углы обзора, разная одежда и стили ношения).

Набор данных Human ID Gait Challenge содержит 1,2 Тбайт видеофайлов походки 122 испытуемых при варьировании 32 факторов, влияющих на условия эксперимента.

Набор данных MPII Human Pose включает в себя около 25 тыс. изображений, содержащих более 40 тыс. человек с аннотированными суставами тела. Набор данных охватывает 410 видов деятельности человека, каждое изображение снабжено меткой вида деятельности. Каждое изображение было извлечено из видеохостинга YouTube и снабжено предыдущими и последующими неаннотированными кадрами. Тестовый набор включает более подробную аннотацию: окклюзию частей тела и трехмерную ориентацию туловища и головы [22].

Дейтасет Ten thousand People – Human Pose Recognition Data включает в себя сцены движения людей в помещении и на открытом воздухе. Он охватывает 10 тыс. мужчин и женщин китайского происхождения, преимущественно молодого и среднего возраста. В ходе измерений варьировали высоту съемки, возраст испытуемых, условия освещения, одежду в зависимости от времени года, а также варианты поз человека. Для каждого испытуемого были аннотированы пол, раса, возраст и одежда [23].

Дейтасет Microsoft COCO (Common objects in context) – это широко распространенный набор данных, предназначенный для стимулирования исследований по обнаружению объектов с упором на поиск незнакомых объектов, локализацию объектов на изображениях с точностью до пикселя и обнаружение объектов в сложных сценах. В рамках проекта проходят конкурсы на лучший алгоритм распознавания объектов, например, COCO 2018 Keypoint Detection Task, где требовалось локализовать ключевые точки человека в сложных, неконтролируемых условиях. Наборы данных для обучения содержали свыше 330 тыс. изображений испытуемых, помеченных ключевыми точками. Набор данных Microsoft COCO (Common Objects in Context) содержит 91 класс. Поза описывается с помощью 17 ключевых точек [24].

Метрики качества оценки положения тела человека

Задача предсказания местоположения ключевых точек имеет свою специфику, так как некоторые части тела человека могут быть скрыты от наблюдателя, что делает необходимым предсказание их наиболее вероятного местоположения с помощью специального алгоритма keypoint detector.

При решении задач, связанных с обнаружением объектов необходимо выделить последние с помощью рамки. В качестве метрики качества в этом случае обычно выступают точность (precision), полнота (recall) или их вариации. В основе этих показателей лежит мера сходства между истинным и прогнозируемым местоположением объектов. Применительно к оценке качества алгоритма предсказания ключевых точек используется метрика intersection over union (IoU), которая показывает, насколько хорошо ограничивающая рамка прогноза совпадает с истинной рамкой, ограничивающей объект.

Вычисление IoU основано на коэффициенте Жаккара и рассчитывается по формуле [25]:

$$K_J(A,B) = \frac{A \cap B}{A \cup B},$$

где: $K_J(A,B)$ – метрика IoU; A – область внутри истинной рамки объекта; B – область внутри спрогнозированной рамки объекта (рис. 1).

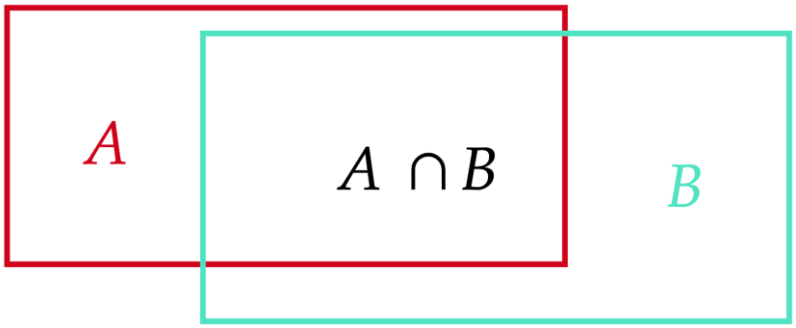

Рис. 1. Метрика IoU: A – область внутри истинной рамки объекта; B – область внутри спрогнозированной рамки объекта; $A \cap B$ – персечение областей A и B

Другой метрикой является «сходство ключевых точек объекта». Для каждого объекта истинные ключевые точки имеют вид:

$$\begin{pmatrix} x_1 & y_1 & v_1 & \cdots & x_k & y_k & v_k \end{pmatrix},$$

где: x, y – координаты ключевых точек; v – флаг видимости (0 – не помечен; 1 –помечен, но не виден; 2 – помечен и виден).

Каждый истинный объект имеет масштаб s (равный квадратному корню из площади сегмента объекта).

Для каждого объекта детектор ключевых точек должен предсказать местоположения и достоверность ключевых точек объекта. Прогнозируемые ключевые точки объекта должны иметь такой же формат, как и у истинных точек.

Мера сходства ключевых точек объекта (object keypoint similarity, OKS) определяется по формуле:

$$p = e^{\left(-\frac{d_i^2}{2s^2k_i^2}\right)},$$

где p – мера сходства ключевых точек объекта, d_i – евклидово расстояние между истинной и предсказанной точками, s – масштаб (отношение площади рамки выделения объекта к общей площади изображения), k_i – константа для каждой ключевой точки [26].

Модели глубокого обучения

Для выделения ключевых точек широко используются глубокие нейронные сети. При этом можно выделить два подхода:

– сверху вниз: сначала выполняется обнаружение человека, а затем в выбранной ограничивающей рамке производится поиск ключевых точек;
– снизу вверх: сначала производится локализация ключевых точек, потом они группируются в объекты, похожие на человека.

В нисходящих подходах обучение происходит на основе тепловых карт месторасположения ключевых точек.

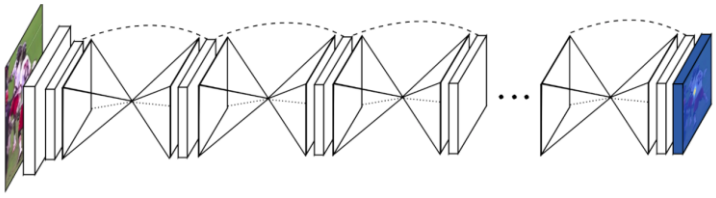

Рис. 2. Модель «Песочные часы»: последовательность модулей с одинаковой структурой, с постоянно уменьшающимся, а затем с увеличивающимся разрешением

В модели песочных часов (рис. 2) авторы предполагают использование нескольких модулей с одинаковой структурой. Каждый модуль обеспечивает повышающую и понижающую дискретизацию (схема напоминает песочные часы). Такая архитектура позволяет учитывать как локальный контекст (напри-

мер, расположение запястья), так и глобальный контекст (например, ориентацию тела) [27].

В модели HRNet (составные песочные часы) сеть состоит из параллельных подсетей с высоким и низким разрешением, с обменом информацией между подсетями с разным разрешением (многомасштабное слияние) (рис. 3). Горизонтальное и вертикальное направления соответствуют глубине сети и масштабу карт объектов соответственно. Стрелка вправо – свертка, стрелка вниз – понижающая дискретизация, стрелка вверх – повышающая дискретизация.

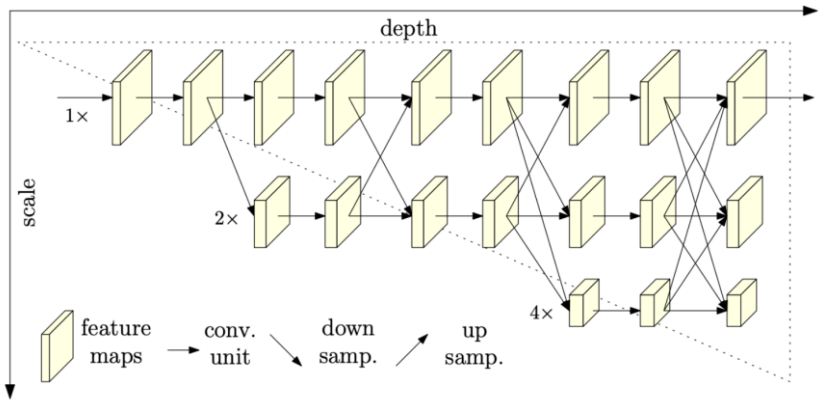

Рис. 3. Модель HRNet: feature maps – карты признаков;
conv. unit – модуль свертки; downsamp. – понижающая дискретизация;
upsamp. – повышающая дискретизация; depth – глубина; scale – масштаб;
1×, 2×, 4× – величина масштаба

Высокое разрешение поддерживается на протяжении всего процесса распознавания. Первоначально он начинается с высокого разрешения, но с каждым шагом глубины он создает больше одновременных шкал, которые получают информацию из более высокого, такого же или более низкого разрешения предыдущих шагов [28].

Модель ViTPose (рис. 4) содержит коллекцию блоков-трансформеров (каждый из которых представляет собой комбинацию слоя нормализации, модуля Multi-Head Self-Attention и нейронной сети прямой связи) и модуля декодера (в двух повторностях: слой деконволюции, за которым следуют пакетная нормализация и функция активации ReLU (rectified linear unit), а также слой линейного прогнозирования).Эта сеть довольно проста в масштабировании и не требует тщательного построения сверточных слоев с рассчитанным количеством параметров, но при этом дает хорошие результаты. Это решение, по-видимому, также хорошо работает для задачи оценки позы нескольких человек с серьезной окклюзией [29].

№1. 2024
ISSN 2410-9916

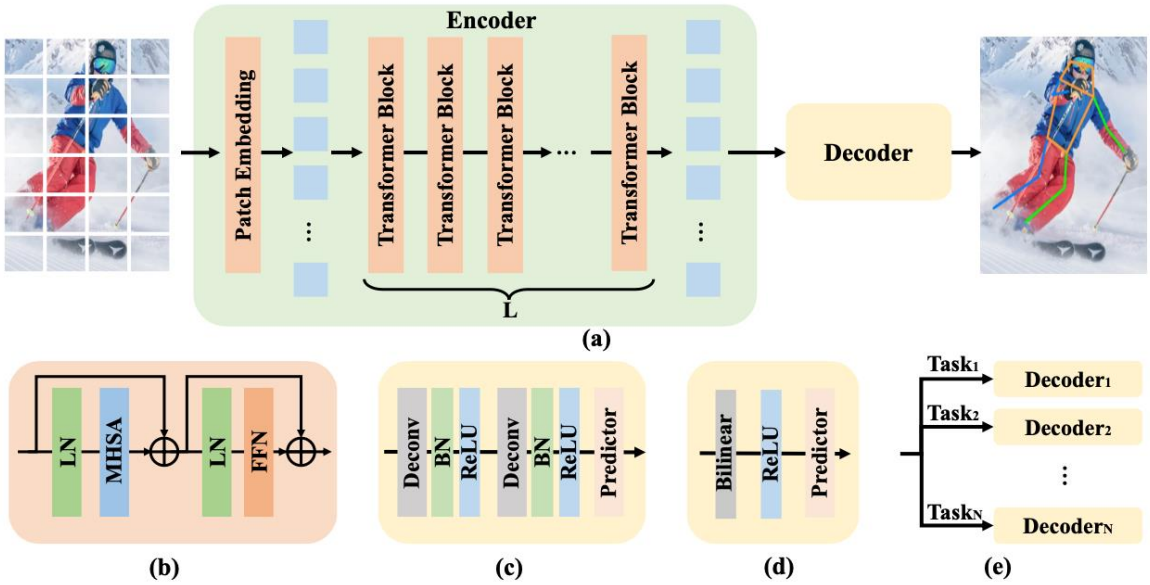

Рис. 4. Демонстрация работы модели ViTPose: модель ViTPose (a); блок трансформера (б); классический декодер (c); простой декодер (d); декодеры для нескольких наборов данных (e). Обозначения на рисунке: encoder – кодер; patch embedding – векторизация участка изображения; transformer block – блок трансформера; decoder – блок декодера; LN – слой нормализации; MHSA – слой Multi-Head Self-Attention (многократное сопоставление входных данных с целью получения более детальной информации); FFN – нейронная сеть прямой связи; deconv – слой деконволюции; BN – байесовский декодер; ReLU – функция активации ReLU; predictor – модуль прогнозирования; bilinear – билинейный слой; task – задача

Восходящие подходы создают несколько скелетных моделей человека одновременно, поэтому они обычно работают быстрее и лучше подходят для решений задач в режиме реального времени, особенно в сценах с толпой для оценки поз нескольких человек.

Модель OpenPose работает следующим образом (рис. 5):

1) вначале признаки извлекаются из нескольких первых слоев;
2) затем подключаются две ветви сверточных слоев: первая состоит из 18 карт достоверности, представляющих каждую конкретную часть скелета позы человека, а вторая имеет 38 полей сходства частей (ПСЧ), представляющих уровень связи между частями (двудольный граф со связями между ключевыми точками);
3) выполняется обрезка соединений между ключевыми точками, с низкой достоверностью от одного и того же экземпляра человека;
4) после этапа обрезки несколько человеческих поз регрессируются.

Данный подход позволяет оценивать позы нескольких человек в режиме реальном времени [30].

Модель OmniPose (рис. 6) начинается с двух сверток 3×3, за которыми следует блок ResNet Bottleneck (обеспечивает уменьшение размерности данных

DOI: 10.24412/2410-9916-2024-1-191-212
URL: http://sccs.intelgr.com/archive/2024-01/08-Bogdanov.pdf

перед их повторным расширением, подобно форме бутылочного горлышка). После этого следуют 3 блока HRNet (сверточная нейронная сеть высокого разрешения, предназначенная для решения задач в области семантической сегментации, обнаружения объектов и классификации изображений), каждый с обогащением модуляции гауссовой тепловой карты. Такой подход предполагает, что тепловая карта определенной ключевой точки соответствует распределению Гаусса, и пытается найти центр этого распределения [31].

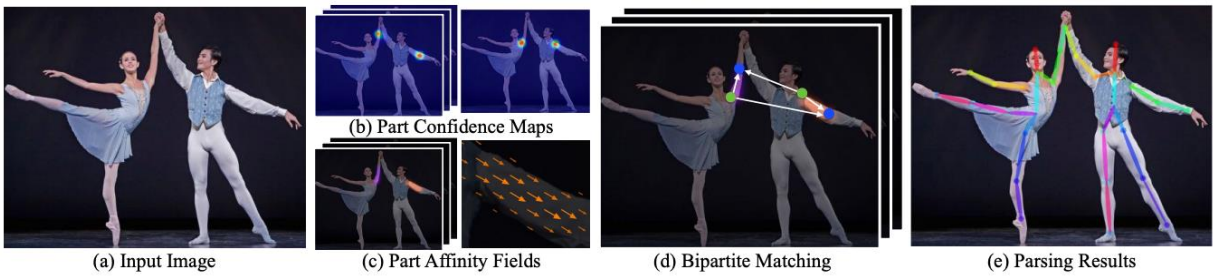

Рис. 5. Процесс создания скелетной модели человека с помощью модели OpenPose: input image – входное изображение (a); part confidence maps – карта согласованности деталей (b); part affinity fields – поле сходства деталей (c); bipartite matching – отображение на двудольный граф (d); parsing results – результат анализа (e)

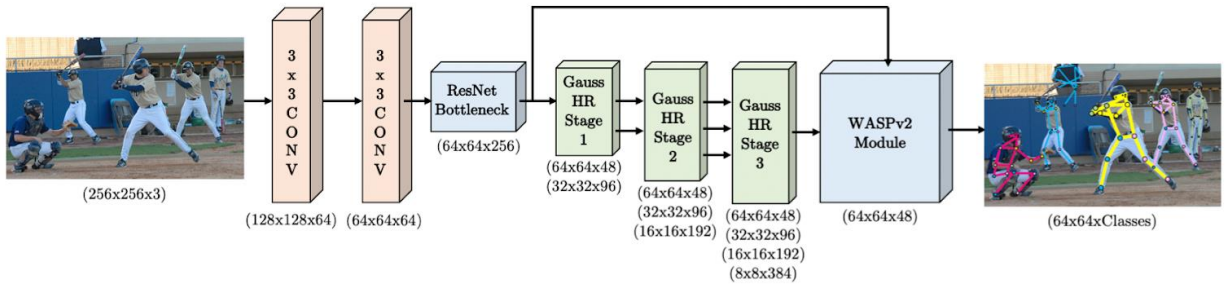

Рис. 6. Модель OmniPose: 3×3 CONV – модуль свертки 3×3; ResNet Bottleneck – блок «бутылочного горлышка» остаточной сети; Gauss HR Stage 1, 2, 3 – блоки сверточных сетей высокого разрешения с обогащением модуляции гауссовой тепловой карты 1, 2, 3 ступеней; WASPv2 Module – модуль каскада пространственного объединения сверток, обеспечивающих многомасштабное представление объектов); classes – распознанные классы объектов

MoveNet использует тепловые карты для точной локализации ключевых точек человеческого тела. Архитектура состоит из (рис. 7):

– экстрактора признаков MobileNetV2 с сетью пирамиды функций FPN;

– набора модулей прогнозирования.

Существуют четыре модуля прогнозирования:

– тепловая карта центра человека;

– регрессионное поле ключевых точек, которое прогнозирует ключевые точки человека;
– тепловая карта ключевых точек человека, которая прогнозирует все ключевые точки независимо от того, какому экземпляру человека они принадлежат;
– двумерное поле смещения для каждой ключевой точки, прогнозирующее смещение от каждого пикселя выходной карты объектов до местоположения каждой ключевой точки.

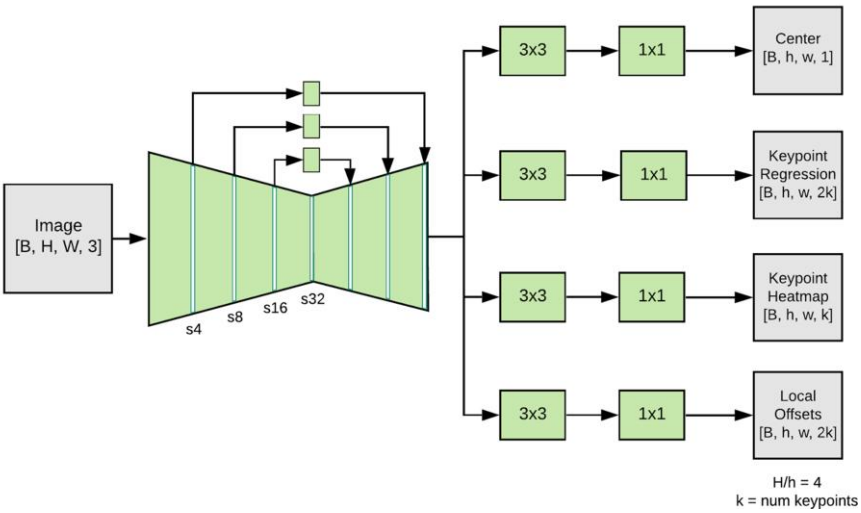

Рис. 7. Модель MoveNet: image – изображение; center – тепловая карта центра человека; keypoint regression – регрессионное поле ключевых точек; keypoint heatmap – тепловая карта ключевых точек; local offsets – двумерное поле смещения ключевых точек; H, h – высота; W, w – ширина; s_4, s_8, s_{16}, s_{32} – масштаб объекта; k – число ключевых точек

Эта архитектура была обучена на наборах данных Microsoft COCO и Google Activity, специализирующихся на сложных позах для фитнеса и йоги. Модель хорошо работает с окклюзиями и незнакомыми позами, реализована в двух версиях: Lightning с упором на скорость и Thunder с целью более высокой точности при сохранении частоты более 30 кадров в секунду [32].

Подведя итог, можно сказать, что при наличии большого объема данных лучше использовать модели ViTPose, OmniPose или HRNet. При работе над приложениями реального времени лучше использовать MoveNet или OpenPose. Для сценариев с недостатком данных, вероятно, лучше будут работать модели меньшего размера (например, MoveNet Lightning, HRNet-32 или OmniPose Lite) [33].

Эксперимент

Целью вычислительного эксперимента являлось изучение факторов, влияющих на эффективность биометрической идентификации по походке (метрики accuracy, precision, recall и F1-score). Были использованы различные варианты

методов машинного обучения для извлечения ключевых точек из видеопотока, частот кадров, разрешения видео и алгоритмов классификации признаков.

На первом этапе был сформирован обучающий набор данных. Для этой цели были задействованы 50 испытуемых из числа студентов и преподавателей Уфимского университета науки и технологий. Им было предложено пройти по замкнутому маршруту. При этом велась видеосъемка камерой Nikon Z6 с объективом Nikon NIKKOR Z 24-70 мм. Использовались следующие режимы видеозаписи:

– 24 FPS, 1920×1080;
– 30 FPS, 1920×1080;
– 60 FPS, 1920×1080;
– 120 FPS, 1920×1080;
– 24 FPS, 3840×2160;
– 30 FPS, 3840×2160.

Каждый видеосюжет можно представить в виде последовательности из отдельных кадров. Из этих кадров извлекали признаки (контрольные точки) с помощью нейронных сетей MoveNet (версий Lightning и Thunder), PoseNet и BlazePose (версий Lite и Heavy) и заносили их в таблицу атрибутов. В качестве меток классов использовали идентификаторы испытуемых.

В качестве признаков использовались в 17 ключевых точек в случае модели COCO и 33 точки в случае модели BlazePose (рис. 8) [33].

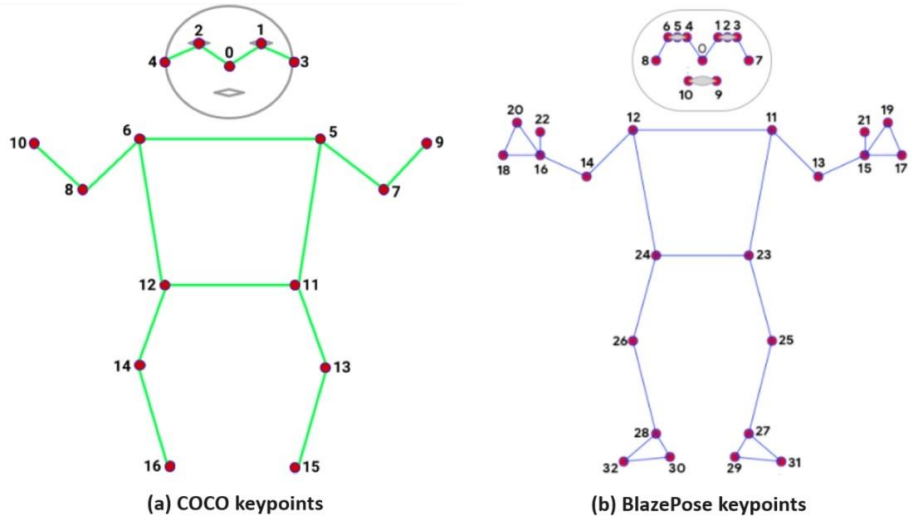

Рис. 8. Ключевые точки в скелетной модели человека:
модель COCO (a); модель BlazePose (b)

На втором этапе проводили обучение классификаторов случайный лес (random forest, RF) и полносвязная нейронная сеть (multilayer perceptron, MLP).

На третьем этапе оценивали эффективность биометрической идентификации по походке. В этом случае в качестве входного сигнала выступал видеофайл с неизвестным испытуемым. Программа извлекала из видеофайла признаки (контрольные точки), подавала их на вход классификатора и предсказывала идентификатор испытуемого (если он был ранее зарегистрирован) или выдавала сообщение об отсутствии испытуемого в базе данных.

Распознавание осуществлялось с помощью последовательности нескольких бинарных классификаций. Сами бинарные классификации проводились в двух вариантах: «один против всех» и «каждый против каждого».

Было разработано программное обеспечение на языке Python. Для вычислений использовался компьютер с процессором Intel Core i9, 256 Гбайт ОЗУ, видеоплатой NVIDIA RTX 3080Ti.

Кросс-валидация

В процессе кросс-валидации обучающую выборку случайным образом делили на две неравные части в соотношении 75:25. Использовали метрики качества accuracy, precision, recall и F1-score.

Методика

Как отмечалось выше, в рамках проводимых исследований была сформирована обучающая выборка, полученная в результате извлечения контрольных точек из кадров видео с испытуемыми. На этой обучающей выборке было обучено два классификатора: случайный лес и полносвязная нейронная сеть (метрика accuracy оказалась близка к 0,98). Далее, была проведена оценка эффективности биометрической идентификации. Для этой цели использовались видеофрагменты с разным разрешением и кадровой частотой, которые не использовались при обучении классификаторов, продолжительностью 10 с. Из этих видеофрагментов выделялись признаки, которые поступали на вход классификаторов. Ожидалось, что если испытуемый был зарегистрирован в системе, то мы получим его метку класса, в противном случае – получим сообщение о неуспехе распознавания. Классификацию проводили методом «один против всех» и «каждый против каждого».

В случае с нейронной сетью MoveNet использовалось 17 ключевых точек, каждая из которых содержала 3 признака (координаты x и y, а также вероятность правильной оценки) – всего 51 признак. В случае нейронной сети BlazePose использовалось 33 ключевые точки, каждая из которых содержала 3 вышеназванных признака, т.е. всего 99 признаков. Предположим, что частота кадров составляет 30 кадров в секунду, время записи – 10 с, то есть имеется 300 кадров, а после выделения признаков в нашей в таблице атрибутов имеется 300 строк по 51 признаку.

В случае классификации «один против всех» обучающая выборка состоит из двух классов. Целевому классу присвоена метка 1, альтернативный класс (метка 0) содержит признаки всех остальных испытуемых. Необходимо аугментировать целевой класс, чтобы избежать дисбаланса классов.

В случае классификации «каждый против каждого» признаки всех испытуемых сравнивались попарно. Вычислительные затраты резко возрастали как n!, где n – количество кадров, но точность увеличилась незначительно.

Результаты

Результаты сравнения производительности различных моделей в режиме классификации «один против всех» представлены в таблице 1, в режиме классификации «каждый против каждого» – в таблице 2.

Таблица 1 – Сравнение производительности различных моделей в различных режимах видеозаписи при использовании режима «один против всех»

Модель	24FPS 1080p	30FPS 1080p	60FPS 1080p	120FPS 1080p	24FPS 2160p	30FPS 2160p
MoveNet (Lightning), RF	0,97	0,98	1,00	0,98	0,95	0,96
MoveNet (Lightning), MLP	0,86	0,87	0,91	0,90	0,87	0,92
MoveNet (Thunder), RF	1,00	0,97	1,00	1,00	0,97	0,98
MoveNet (Thunder), MLP	0,80	0,78	0,89	0,79	0,77	0,84
PoseNet, RF	0,98	0,99	1,00	0,99	0,96	0,97
PoseNet, MLP	0,72	0,71	0,72	0,69	0,79	0,76
BlazePose (Lite), RF	1,00	0,99	1,00	1,00	0,98	0,98
BlazePose (Lite), MLP	0,88	0,89	0,93	0,90	0,89	0,93
BlazePose (Heavy), RF	1,00	0,99	1,00	1,00	1,00	0,99
BlazePose (Heavy), MLP	0,75	0,73	0,80	0,82	0,85	0,89

Таблица 2 – Сравнение производительности различных моделей в различных режимах видеозаписи при использовании режима «каждый против каждого»

Модель	24FPS 1080p	30FPS 1080p	60FPS 1080p	120FPS 1080p	24FPS 2160p	30FPS 2160p
MoveNet (Lightning), RF	0,99	0,99	1,00	1,00	0,99	0,98
MoveNet (Lightning), MLP	0,87	0,72	0,73	0,67	0,58	0,71
MoveNet (Thunder), RF	1,00	0,99	1,00	1,00	0,96	0,99
MoveNet (Thunder), MLP	0,68	0,57	0,64	0,63	0,65	0,79
PoseNet, RF	0,99	1,00	1,00	1,00	0,99	0,98
PoseNet, MLP	0,60	0,82	0,73	0,60	0,67	0,66
BlazePose (Lite), RF	1,00	0,99	1,00	1,00	0,99	1,00
BlazePose (Lite), MLP	0,82	0,73	0,79	0,75	0,67	0,70
BlazePose (Heavy), RF	1,00	1,00	1,00	1,00	1,00	1,00
BlazePose (Heavy), MLP	0,70	0,69	0,69	0,69	0,69	0,69

Как видно из таблиц 1, 2, ансамблевый метод случайного леса имеет преимущество перед моделью глубокого обучения. По-видимому, это объясняется малым размером вектора признаков. Было решено отказаться от использования мультиклассовой классификации, так как, по данным литературы, ее производительность падает с увеличением числа классов. Было установлено, что метод бинарной классификации «каждый против каждого» более эффективен по сравнению с методом «один против всех».

Выводы

Было установлено, что: изменение частоты кадров и разрешения видео дает больший эффект при использовании моделей, ориентированных на высокую скорость в ущерб точности: MoveNet в версии Lightning и BlazePose в версии Lite. При этом лучший результат достигается при частоте кадров 60 FPS, худший – при 24 FPS, увеличение частоты кадров до 120 FPS не улучшает результат по сравнению с 60 FPS. Разрешение видеопотока 1920×1080 позволяет получить лучшие результаты по сравнению с 3840×2160.

Ансамблевые метод случайного леса показал более высокую эффективность классификации по сравнению с полносвязной нейронной сетью. Сравнение двух вариантов бинарной классификации («один против всех» и «каждый против каждого») показало незначительное преимущество метода «каждый против каждого», однако вычислительная сложность неоправданно высока (n!, где n – количество классов).

Предложенный подход был реализован в виде программного обеспечения, предназначенного для биометрической идентификации по походке. Предварительные результаты вселяют сдержанный оптимизм. Технологию в дальнейшем можно будет использовать в системе «Безопасный город».

Исследование поддержано грантом Российского Научного Фонда № 22-19-00471.

Литература

1. Human Identification at a Distance // The IT Law Wiki [Электронный ресурс]. 24.02.2023. – URL: https://itlaw.fandom.com/wiki/Human_Identification_at_a_Distance (дата обращения: 24.02.2023).

2. ImageNet: A Pioneering Vision for Computers // History of Data Science [Электронный ресурс]. 24.02.2023. – URL: https://www.historyofdatascience.com/imagenet-a-pioneering-vision-for-computers/ (дата обращения: 24.02.2023).

3. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems. 2012. Vol. 25. P. 1097-1105.

4. Zeiler M. D., Fergus R. Visualizing and understanding convolutional networks // European Conference on Computer Vision (ECCV) (Zurich, 6-12 September 2014). – Cham, Springer International Publishing, 2014. Vol. 8689. P. 818-833. doi: 10.1007/978-3-319-10590-1_53.

5. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, 27–30 June 2016). – IEEE-USA, 2016. P. 770–778. doi: 10.1109/CVPR.2016.90.

6. Alsaggaf W. A., Mehmood I., Khairullah E. F., Alhuraiji S., Sabir M. F. S., Alghamdi A. S., Abd El-Latif A. A. A Smart Surveillance System for Uncooperative Gait Recognition Using Cycle Consistent Generative Adversarial Networks (CCGANs) // Hindawi. Computational Intelligence and Neuroscience. 2021. Vol. 2021.Article ID 3110416. doi: 10.1155/2021/3110416.

7. CASIA-B // GitHub [Электронный ресурс]. 24.02.2023. – URL: https://github.com/ShiqiYu/OpenGait/blob/master/datasets/CASIA-B/README.md (дата обращения: 24.02.2023).

8. Sarkar S., Phillips P. J., Liu Z., Vega I. R., Grother P., Bowyer K. W. The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005. Vol. 27. No. 2. P. 162-177. doi: 10.1109/tpami.2005.39.

9. Toshev A., Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks // IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Columbus, 23-28 June 2014). – IEEE-USA, 2014. P. 1653-1660. doi: 10.1109/CVPR.2014.214.

10. OpenPose // GitHub [Электронный ресурс]. 24.02.2023. – URL: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/README.md (дата обращения: 24.02.2023).

11. Pishchulin L., Insafutdinov E., Tang S., Andres B., Andriluka M., Gehler P. V., Schiele B. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation // IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, 27–30 June 2016). – IEEE-USA, 2016. P. 4929-4937. doi: 10.1109/CVPR.2016.533.

12. Inside MoveNet, Google's Latest Pose Detection Model // Analytics India Magazine [Электронный ресурс]. 24.02.2023. – URL: https://analyticsindiamag.com/inside-movenet-googles-latest-pose-detection-model/ (дата обращения: 24.02.2023).

13. Chai T., Li A., Zhang S., Li Z., Wang Y. Lagrange Motion Analysis and View Embeddings for Improved Gait Recognition // IEEE/CVF Conference on Computer Vision and Pattern Recognition (New Orleans, 18-24 June 2022). – IEEE-USA, 2022. P. 20249–20258. doi: 10.1109/CVPR52688.2022.01961.

14. Lin B., Zhang S., Yu X. Gait recognition via effective global-local feature representation and local temporal aggregation // IEEE/CVF International Conference on Computer Vision (ICCV) (Montreal, 10-17 Oct. 2021). – IEEE-USA, 2021. P. 14648–14656. doi: 10.1109/ICCV48922.2021.01438.

15. Teepe T., Gilg J., Herzog F., Hörmann S., Rigoll G. Towards a Deeper Understanding of Skeleton-based Gait Recognition // IEEE/CVF Conference on Computer Vision and Pattern Recognition. (New Orleans, 18-24 June 2022). – IEEE-USA, 2022. P. 1569-1577.

16. Wang L., Han R., Feng W. Combining the Silhouette and Skeleton data for Gait Recognition // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Singapore, 23-27 May 2022). – IEEE-USA, 2022. P. 1-5. doi: 10.1109/ICASSP49357.2023.10096986.

17. Cui Y, Kang Y. Multi-modal Gait Recognition via Effective Spatial-Temporal Feature Fusion // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Vancouver, 18-22 June 2023). – IEEE-USA, 2023. P. 17949-17957, doi: 10.1109/CVPR52729.2023.01721.

18. Wang L., Chen J., Liu Y. Frame-level refinement networks for skeleton-based gait recognition // Computer Vision and Image Understanding. 2022. Vol. 222, P. 103500. doi: 10.1016/j.cviu.2022.103500.

19. Hsu H.-M., Wang Y., Yang C.-Y, Hwang J.-N., Thuc H. L. U., Kim K.-J. GaitTAKE: Gait recognition by temporal attention and keypoint-guided embedding // IEEE International Conference on Image Processing (ICIP) (Bordeaux, 16-19 October 2022). – IEEE-USA, 2022. P. 2546–2550. doi: 10.1109/ICIP46576.2022.9897409.

20. Pinyoanuntapong E., Ali A., Wang P., Lee M., Chen C. GaitMixer: Skeleton-Based Gait Representation Learning Via Wide-Spectrum Multi-Axial Mixer // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). (Rhodes Island, 4-10 June 2023). – IEEE-USA, 2023. P. 1–5. doi: 10.1109/ICASSP49357.2023.10096917.

21. Shen C., Yu S., Wang J., Huang G. Q., Wang L. A Comprehensive Survey on Deep Gait Recognition: Algorithms, Datasets and Challenges // arXiv. 2023. doi: 10.48550/arXiv.2206.13732.

22. MPII Human Pose Dataset // Max Planck Institute for Informatics [Электронный ресурс]. 24.02.2023. – URL: http://human-pose.mpi-inf.mpg.de/ (дата обращения: 24.02.2023).

23. 10,000 People - Human Pose Recognition Data // Papers with Code [Электронный ресурс]. 24.02.2023. – URL: https://paperswithcode.com/dataset/10000-people-human-pose-recognition-data (дата обращения: 24.02.2023).

24. COCO 2018 Keypoint Detection Task // COCO – Common Objects in Context [Электронный ресурс]. 24.02.2023. – URL: https://cocodataset.org/#keypoints-2018 (дата обращения: 24.02.2023).

25. Intersection over Union (IoU) for object detection // SuperAnnotate [Электронный ресурс]. 24.02.2023. – URL: https://www.superannotate.com/blog/intersection-over-union-for-object-detection (дата обращения: 24.02.2023).

26. Common Objects in Context. Keypoint Evaluation // COCO – Common Objects in Context [Электронный ресурс]. 24.02.2023. – URL: https://cocodataset.org/#keypoints-eval (дата обращения: 24.02.2023).

27. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation // European Conference on Computer Vision (ECCV) (Amsterdam, 8-16 Oct 2016). – Cham, Springer International Publishing, 2016. Vol. 9912. P. 483-499. doi: 10.1007/978-3-319-46484-8_29.

28. Sun K., Xiao B., Liu D., Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Long Beach, 15-20 June 2019). – IEEE-USA, 2019. P. 5686-5696. doi: 10.1109/CVPR.2019.00584.

29. Xu Y., Zhang J., Zhang Q., Tao D. ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation // arXiv. 2022. doi: 10.48550/arXiv.2204.12484.

30. Cao Z., Hidalgo G., Simon T., Wei S.-E., Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Vol. 43. P. 172-186. doi: 10.1109/TPAMI.2019.2929257.

31. Artacho B., Savakis A. OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation // arXiv. 2021. doi: 10.48550/arXiv.2103.10180.

32. Votel R., Li N. Next-Generation Pose Detection with MoveNet and TensorFlow.js // TensorFlow Blog [Электронный ресурс]. 24.02.2023. – URL: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html (дата обращения: 24.02.2023).

33. Pose Detection // GitHub [Электронный ресурс]. 24.02.2023. – URL: https://github.com/tensorflow/tfjs-models/blob/master/pose-detection/README.md. (дата обращения: 24.02.2023).

References

1. Human Identification at a Distance. *The IT Law Wiki.* Available at: https://itlaw.fandom.com/wiki/Human_Identification_at_a_Distance (accessed 24 February 2023).

2. ImageNet: A Pioneering Vision for Computers. *History of Data Science*. Available at: https://www.historyofdatascience.com/imagenet-a-pioneering-vision-for-computers/ (accessed 24 February 2023).

3. Krizhevsky A., Sutskever I. and Hinton G. E. Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, 2012, vol. 25, pp. 1097-1105.

4. Zeiler M. D., Fergus R. Visualizing and understanding convolutional networks. *European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, 2014, vol. 8689, pp. 818-833. doi: 10.1007/978-3-319-10590-1_53.

5. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

6. Alsaggaf W. A., Mehmood I., Khairullah E. F., Alhuraiji S., Sabir M. F. S., Alghamdi A. S., Abd El-Latif A. A. A Smart Surveillance System for Uncooperative Gait Recognition Using Cycle Consistent Generative Adversarial Networks

(CCGANs). *Hindawi. Computational Intelligence and Neuroscience*, 2021, vol. 2021, article ID 3110416. doi: 10.1155/2021/3110416.

7. CASIA-B. *GitHub.* Available at: https://github.com/ShiqiYu/OpenGait/blob/master/datasets/CASIA-B/README.md (accessed 24 February 2023).

8. Sarkar S., Phillips P. J., Liu Z., Vega I. R., Grother P., Bowyer K. W. The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, vol. 27, No. 2, pp. 162-177. doi: 10.1109/tpami.2005.39.

9. Toshev A., Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 2014, pp. 1653-1660. doi: 10.1109/CVPR.2014.214.

10. OpenPose. *GitHub*. Available at: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/README.md (accessed 24 February 2023).

11. Pishchulin L., Insafutdinov E., Tang S., Andres B., Andriluka M., Gehler P. V., Schiele B. DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 4929-4937. doi: 10.1109/CVPR.2016.533.

12. Inside MoveNet, Google's Latest Pose Detection Model. *Analytics India Magazine*. Available at: https://analyticsindiamag.com/inside-movenet-googles-latest-pose-detection-model/ (accessed 24 February 2023).

13. Chai T., Li A., Zhang S., Li Z., Wang Y. Lagrange Motion Analysis and View Embeddings for Improved Gait Recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2022, pp. 20249–20258. doi: 10.1109/CVPR52688.2022.01961.

14. Lin B., Zhang S., Yu X. Gait recognition via effective global-local feature representation and local temporal aggregation. *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, BC, Canada, 2021, pp. 14648–14656. doi: 10.1109/ICCV48922.2021.01438.

15. Teepe T., Gilg J., Herzog F., Hörmann S., Rigoll G. Towards a Deeper Understanding of Skeleton-based Gait Recognition. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, pp. 1569-1577.

16. Wang L., Han R., Feng W. Combining the Silhouette and Skeleton data for Gait Recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, 2022, pp. 1-5. doi: 10.1109/ICASSP49357.2023.10096986.

17. Cui Y, Kang Y. Multi-modal Gait Recognition via Effective Spatial-Temporal Feature Fusion. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, pp. 17949-17957, doi: 10.1109/CVPR52729.2023.01721.

18. Wang L., Chen J., Liu Y. Frame-level refinement networks for skeleton-based gait recognition. *Computer Vision and Image Understanding*, 2022, vol. 222, pp. 103500. doi: 10.1016/j.cviu.2022.103500.

19. Hsu H.-M., Wang Y., Yang C.-Y, Hwang J.-N., Thuc H. L. U., Kim K.-J. GaitTAKE: Gait recognition by temporal attention and keypoint-guided embedding. *IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, 2022, pp. 2546–2550. doi: 10.1109/ICIP46576.2022.9897409.

20. Pinyoanuntapong E., Ali A., Wang P., Lee M., Chen C. GaitMixer: Skeleton-Based Gait Representation Learning Via Wide-Spectrum Multi-Axial Mixer. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10096917.

21. Shen C., Yu S., Wang J., Huang G. Q., Wang L. A Comprehensive Survey on Deep Gait Recognition: Algorithms, Datasets and Challenges, *arXiv*, 2023. doi: 10.48550/arXiv.2206.13732.

22. MPII Human Pose Dataset. *Max Planck Institute for Informatics.* Available at: http://human-pose.mpi-inf.mpg.de/ (accessed 24 February 2023).

23. 10,000 People - Human Pose Recognition Data. *Papers with Code.* Available at: https://paperswithcode.com/dataset/10000-people-human-pose-recognition-data (accessed 24 February 2023).

24. COCO 2018 Keypoint Detection Task. *COCO – Common Objects in Context.* Available at: https://cocodataset.org/#keypoints-2018 (accessed 24 February 2023).

25. Intersection over Union (IoU) for object detection. SuperAnnotate. Available at: https://www.superannotate.com/blog/intersection-over-union-for-object-detection (accessed 24 February 2023).

26. Common Objects in Context. Keypoint Evaluation. COCO – Common Objects in Context. Available at: https://cocodataset.org/#keypoints-eval (accessed 24 February 2023).

27. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation. *European Conference on Computer Vision (ECCV)*, Amsterdam, The Netherlands, 2016, vol. 9912, pp. 483-499. doi: 10.1007/978-3-319-46484-8_29.

28. Sun K., Xiao B., Liu D., Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 5686-5696. doi: 10.1109/CVPR.2019.00584.

29. Xu Y., Zhang J., Zhang Q., Tao D. ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. *arXiv*. 2022. doi: 10.48550/arXiv.2204.12484.

30. Cao Z., Hidalgo G., Simon T., Wei S.-E., Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, vol. 43, pp. 172-186. doi: 10.1109/TPAMI.2019.2929257.

31. Artacho B., Savakis A. OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation. *arXiv*. 2021. doi: 10.48550/arXiv.2103.10180.

32. Votel R., Li N. Next-Generation Pose Detection with MoveNet and TensorFlow.js. *TensorFlow Blog*. Available at: https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html (accessed 24 February 2023).

33. Pose Detection. GitHub Available at: https://github.com/tensorflow/tfjs-models/blob/master/pose-detection/README.md. (accessed 24 February 2023).

Статья поступила 10 января 2024 г.

Информация об авторах

Богданов Марат Робертович – кандидат биологических наук. Доцент кафедры вычислительной математики и кибернетики. Уфимский университет науки и технологий. Доцент кафедры прикладной информатики. Башкирский государственный педагогический университет имени М. Акмуллы. Область научных интересов: цифровая обработка сигналов, биометрическая идентификация. E-mail: bogdanov_marat@mail.ru

Шахмаметова Гюзель Радиковна – доктор технических наук, профессор. Заведующий кафедрой вычислительной математики и кибернетики. Уфимский университет науки и технологий. Область научных интересов: интеллектуальные системы принятия решений. E–mail: shakhgouzel@mail.ru

Оськин Николай Николаевич – Директор. ООО «Сибирская Телеметрическая Компания». Область научных интересов: цифровая обработка сигналов. E–mail: nonik2@mail.ru

Корнилов Андрей Владимирович – магистрант кафедры вычислительной математики и кибернетики. Уфимский университет науки и технологий. Область научных интересов: анализ данных. E–mail: kornilov_a_v@mail.ru

Адрес: 450000, Россия, г. Уфа, ул. Карла Маркса, д. 12, к. 6

Biometric identification based on gait

M. R. Bogdanov, G. R. Shakhmametova, N. N. Oskin, A. V. Kornilov

Statement of the problem*: the growth of extremism in many countries of the world makes the problem of contactless biometric identification urgent. Solving of this problem is complicated by the fact that rioters use various means of camouflage. Under these conditions, biometric identification based on facial recognition becomes ineffective. In this regard, gait recognition has a number of advantages, because gait can be recognized over a long distance, is difficult to fake, and changes slowly over time. To recognize gait, you can use various data sources, for example, a video stream, sounds of steps, or readings from inertial sensors of mobile phones located in the pockets of subjects.* ***The purpose of the work*** *is to study the factors influencing the effectiveness of biometric identification by gait when the video stream is the data source.* ***Methods used****: a video stream is used as a data source. Key points of the human body were extracted from the video stream using the MoveNet (Lightning and Thunder), PoseNet and BlazePose (Lite and Heavy) neural networks. Feature classification was carried out using a random forest and a fully connected neural network.* ***Novelty****: for biometric identification by gait, methods used to solve the problem of pose estimation were used. The influence on the efficiency of gait recognition of factors such as video resolution, frame rate, algorithm for extracting features from the video stream, and recording duration was studied.* ***Result****: It was found that: changing the frame rate and video resolution has a greater effect when using models that focus on high speed at the expense of accuracy: MoveNet in the Lightning version and BlazePose in the Lite version, the best result is achieved at a frame rate of 60 FPS, the worst - at 24 FPS, increasing the frame rate to 120 FPS does not improve the result compared to 60 FPS, the video stream resolution of 1920 by 080 allows*

you to get better results compared to 3840×2160. Ensemble methods (random forest and gradient boosting) showed higher classification efficiency compared to a fully connected neural network. ***Practical significance****: software was developed for biometric identification by gait. It can be used in security systems for biometric identification and identifying patterns of aggressive behavior.*

Key words: *contactless biometric identification, machine learning, information security.*

Information about Authors

Marat Robertovich Bogdanov – Ph.D. of Biological Sciences, Associate Professor at the Institute of Informatics, mathematics and robotics. Ufa University of Science and Technology. Associate Professor at the Institute of Physics, Mathematics, Digital and Nanotechnologies. M. Akmullah named after Bashkir State Pedagogical University. Field of research: digital signal processing, biometric identification. E-mail: bogdanov_marat@mail.ru

Gyuzel Radikovna Shakhmametova – Dr. habil. of Technical Sciences, Full Professor. Head of the Department of Computational Mathematics and Cybernetics. Ufa University of Science and Technology. Field of research: intelligent decision-making systems. E-mail: shakhgouzel@mail.ru

Nikolai Nikolaevich Oskin – CEO. Sibirskaya Telemetricheskaya Kompaniya ltd. Field of research: digital signal processing. E-mail: nonik2@mail.ru

Andrey Vladimirovich Kornilov – master's student at the Department of Computational Mathematics and Cybernetics. Ufa University of Science and Technology. Field of research: data analysis. E-mail: kornilov_a_v@mail.ru

Address: Russia, 450000, Ufa, ul. Karla Marksa, 12, k. 6