Методы идентификации человека по походке в видео

[1] А.И. Соколова <ale4kasokolova@gmail.com>
[2] А.С. Конушин <anton.konushin@grapics.cs.msu.ru>
[1] Национальный исследовательский университет Высшая школа экономики
101000, Россия, Москва, ул. Мясницкая, 20
[2] Московский государственный университет им. М.В. Ломоносова
119991, Россия, Москва, Ленинские горы

Аннотация. Походка – важный биометрический показатель, позволяющий идентифицировать человека на большом расстоянии и без непосредственного контакта. Благодаря этим качествам, отсутствующим у других популярных идентификаторов, таких как отпечатки пальцев и радужная оболочка глаза, распознавание человека по походке в наши дни стало очень распространено и востребовано в различных сферах, где возможно использование систем видеонаблюдения. С развитием методов компьютерного зрения появляется множество подходов к идентификации человека по движениям в видео, использующих как естественные биометрические характеристики (скелет человека, его силуэт, их изменение во время ходьбы), так и абстрактные признаки. Современные методы объединяют в себе классические алгоритмы анализа видео и изображений и новые подходы, показывающие высокие результаты в смежных задачах компьютерного зрения, таких как идентификация человека по лицу или распознавание действий. Однако из-за большого количества условий, влияющих на саму манеру движения человека и ее представление в видео, задача идентификации человека по походке до сих пор не имеет достаточно точного решения. Многие методы заточены исключительно под условия, присутствующие в базах данных, на которых они обучаются, что ограничивает их применимость в реальной жизни. В данной работе проводится обзор современных методов распознавания человека по походке, их анализ и сравнение на нескольких популярных видео коллекциях и для разных формулировок задачи распознавания, а также выявляются проблемы, препятствующие окончательному решению задачи идентификации по походке.

Ключевые слова: походка; биометрия; силуэт; нейронные сети; идентификация

DOI: 10.15514/ISPRAS-2019-31(1)-5

Для цитирования: Соколова А.И., Конушин А.С. Методы идентификации человека по походке в видео. Труды ИСП РАН, том 31, вып. 1, 2019 г., стр. 69-82. DOI: 10.15514/ISPRAS-2019-31(1)-5

1. Введение

Задача идентификации человека по походке особенно актуальна в современном мире. Согласно биометрическим исследованиям манера движения каждого человека индивидуальна, и фальсифицировать ее практически невозможно, что делает походку уникальным идентификатором, таким как отпечатки пальцев или радужная оболочка глаза. Однако в отличие от этих ``классических'' характеристик, походку можно наблюдать издалека, не контактируя с человеком напрямую, поэтому с развитием высококачественных систем видеонаблюдения именно походка становится наиболее подходящим показателем для распознавания.

Основной областью применения распознавания человека по походке является сфера безопасности, где часто есть необходимость идентифицировать человека, попадающего в

поле зрения видеокамер, например, для поимки преступников или контроля доступа на закрытые территории.

Задача распознавания человека по походке очень специфична в силу наличия множества факторов, меняющих походку визуально (наличие каблуков или неудобной обуви, переносимые тяжелые предметы, одежда, скрывающая части тела человека) или влияющих на внутреннее представление модели походки (ракурс, освещение, различные параметры камеры). Поэтому, несмотря на успехи современных методов компьютерного зрения, задачу идентификации по походке пока нельзя назвать решенной. В этой статье представлен обзор методов распознавания человека по походке в видео и их сравнение на популярных наборах данных.

На сегодняшний день существует два основных подхода к получению признаков походки и их классификации: построение признаков вручную и обучение признаков. Первый способ более традиционен и, как правило, основывается на вычислении различных свойств бинарных масок силуэта человека или на исследовании взаимного расположения суставов, относительных расстояний и скоростей, а также других кинетических показателей.

Обучение признаков характерно для искусственный нейронных сетей, набравших популярность в последние годы благодаря выдающимся результатам в решении многих задач компьютерного зрения, таким как классификация видео и изображений, сегментация изображений, детекция объектов, визуальный трекинг и другие. Признаки, обучаемые с помощью нейронных сетей, часто обладают более высоким уровнем абстракции, необходимым для качественного распознавания.

Кроме того, высокое качество идентификации достигается методами, комбинирующими два описанных подхода. На начальном этапе вручную вычисляются базовые характеристики походки, а на их основе обучается нейронная сеть, выделяющая более абстрактные признаки. Несмотря на успешность методов глубинного обучения, на данный момент наилучшего результата на некоторых наборах данных достигают неглубокие алгоритмы, поэтому оба глобальных подхода достойны внимания.

2. Базовые признаки походки

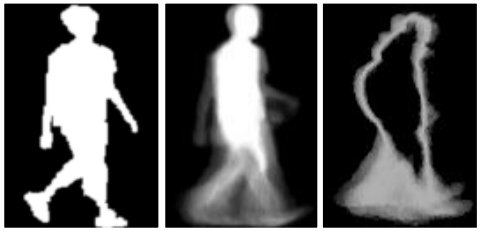
Рассмотрим сначала некоторые классические базовые подходы, в которых признаки походки извлекаются вручную из естественных соображений.

2.1 Бинарные силуэты человека

Наиболее распространенной характеристикой походки является изображения энергии походки (Gait Energy Image, GEI [11]). Такие изображения – усредненные по одному циклу походки бинарные маски силуэта движущегося человека. В предположении, что движения человека во время ходьбы периодически повторяются, вычисляется пространственно-временное описание походки человека. Полученные изображения характеризуют частоты нахождения человека в той или иной позе во время движения. Этот подход получил широкое распространение и лег в основу множества других методов распознавания походки.

Кроме того, многие подходы, не использующие изображения энергии напрямую, предлагают аналогичную агрегацию других базовых признаков. Например, распознавание возможно по изображениям энтропии походки (gait entropy image, GEnI [2]}), где вместо усреднения силуэтов вычисляется энтропия каждого пикселя, или по энергии разницы кадров (frame difference energy image, FDEI [4]), отражающей разности между силуэтами в последовательных кадрах видео.

Визуализация бинарных силуэтов и изображений энергии и энтропии походки представлена на рис. 1.



а) б) в)

Рис. 1. Примеры базовых дескрипторов походки: а) бинарный силуэт, б) изображение энергии походки, в) изображение энтропии походки

Fig. 1. Examples of basic gait descriptors: a) binary silhouette, b) gait energy image, c) gait entropy image

Несмотря на одинаковую простоту и естественность всех этих методов, именно изображения энергии используются и развиваются до сих пор. По таким изображениям, как и по обычным черно-белым, можно вычислять дальнейшие признаки, такие как гистограммы ориентированных градиентов (HOG-дескрипторы) [7, 16] или гистограммы оптического потока (HOF-дескрипторы) [14, 32], или строить более сложные алгоритмы классификации, использующие специфику задачи распознавания походки.

Так, одними из наиболее успешных многоракурсных подходов являются два неглубоких метода, использующих в качестве базовых признаков изображения энергии походки. Первый из них – байесовский подход, предложенный Ли в [15]. Авторы предлагают считать изображения энергии походки случайными матрицами, получающимися из собственно походки и независимого от нее шума, соответствующего различным условиям (таким как угол съемки, различная одежда или наличие переносимых в руках вещей), причем предполагается, что оба слагаемых – нормальные случайные величины. Рассмотрение совместного распределения двух представлений походки в предположении совпадения классов или их различия сводит проблему к задаче оптимизации ковариационных матриц, решаемую с помощью EM-алгоритма. Во втором подходе [19] предлагается обобщение метода линейного дискриминантного анализа [3], а именно – многоракурсный дискриминантный анализ. Для признаков походки, посчитанных для каждого угла обзора, обучается отдельное вложение, так чтобы внутриклассовый разброс был минимален, а межклассовый – максимален.

Идея уменьшения внутриклассовых расстояний и увеличения межклассовых наследуется и в более поздней работе [18], где предлагается единая структура для обучения метрики совместной интенсивности пары изображений и пространственной метрики. Поочередная оптимизация обеих метрик приводит к модели, превосходящей по качеству распознавания базовые модели дискриминантного анализа.

Описанные подходы интуитивно понятны и математически просты, что в дополнение к высоким результатам распознавания дает им преимущество перед многими другими более сложными методами. Общим недостатком методов, использующих GEI для многоракурсного распознавания, является необходимость вычислять изображение энергии для каждого угла обзора, присутствующего в выборке. Поэтому для каждого кадра видео нужно знать, под каким углом он был снят, что не всегда возможно в реальных данных.

Тем не менее, если информация об угле съемки все же известна, ее можно эффективно использовать, применяя модель преобразования ракурса, как это предложено в нескольких подходах [21, 22]. Для признаков походки одного человека, полученных с разных ракурсов, обучается преобразование, переводящее одни в другие. Благодаря таким преобразованиям дескрипторы, соответствующие разным углам съемки, можно вложить в единое пространство, в котором классификация оказывается точнее.

2.2 Поза человека

Другим важнейшим источником информации, помимо силуэтов, является скелет человека. Множество методов распознавания базируется на исследовании позы человека -- положении суставов и основных частей тела и их движении в видео при ходьбе. Подходы, основанные на позе, варьируются от полностью структурных (рассматривающих кинематические характеристики позы) до более сложных, объединяющих в себе кинематические и пространственно-временные характеристики.

К первой группе можно отнести работу [1], в которой распознавание происходит и по походке, и по внешности. В качестве основных характеристик походки при этом берутся абсолютные и относительные расстояния между суставами, а также признаки, основанные на смещениях ключевых точек фигуры между кадрами.

В подходе, предложенном в [30], также исследуется скелет человека, однако авторы вводят более сложную математическую модель, рассматривая семейство гладких Пуассоновских функций расстояния для построения изображения вариации скелета (Skeleton Variance Image, SVI).

Несколько других работ также предлагают модели, основанные на положении частей тела человека, но используют их в совокупности с признаками другой природы. Например, метод 2017 года [8] комбинирует кинематический подход с пространственным, рассматривая в качестве динамических признаков походки как траектории движения суставов, так и изменение формы силуэта во времени.

2.3 Траектории точек фигуры

Еще один неглубокий подход, показывающий высокое качество распознавания, предложен в [6], где рассматриваются траектории движения точек фигуры человека и по ним строятся дескрипторы движения Фишера, которые классифицируются методом опорных векторов.

3. Нейросетевые подходы

Несмотря на обилие структурных неглубоких подходов, сверточные нейронные сети (Convolutional Neural Networks, CNN) занимают прочную позицию во всех задачах компьютерного зрения и в том числе распознавании походки. За последние несколько лет было предложено множество нейросетевых методов идентификации по походке, отличающихся как технически (выбором архитектур сетей, функции потерь, способов обучения), так и идейно -- методом обработки данных и извлечения первичных признаков, подаваемых на вход сети.

В силу того, что фигура и образ человека могут меняться в зависимости от носимой одежды и освещения, важно, чтобы модель обращала внимание не столько на внешние параметры, сколько на само движение фигуры человека. Поэтому большинство методов классифицируют видео не напрямую по кадрам, а вычисляют всевозможные динамические характеристики походки и уже по ним распознают человека.

Одной из таких характеристик, дающих информацию о движении, является оптический поток -- векторное поле видимого движения точек сцены. Его преимущество состоит в том, что обученная на таких данных модель не обращает внимания на цвет, яркость или контрастность кадров видео. Влияние на распознавание оказывает только движение отдельных точек фигуры, а именно это и составляет походку человека.

В нескольких работах [5, 25], появившихся практически одновременно, предлагается рассматривать блоки карт оптического потока аналогично временной составляющей классической двухпоточной модели [24] для распознавания действий. Для нескольких идущих подряд пар соседних кадров вычисляется оптический поток и строится тензор – блок из нескольких карт потока. Для большей точности из этого блока вырезается часть, во всех

кадрах содержащая фигуру человека, и на таких блоках обучается нейронная сеть. На этапе тестирования сеть используется для извлечения признаков, которые потом можно классифицировать любым методом машинного обучения, например, методом опорных векторов (SVM) или методом ближайшего соседа (kNN).

Подход, предложенный в [26], развивает идеи [25], однако отказывается от блоков идущих подряд кадров. Вместо этого более детально исследуется движение точек около важных (как показывают эксперименты) частей тела. В ходе предобработки оценивается поза человека и оптический поток рассматривается в окрестности ступней человека, а также по отдельности в верхней и нижней частях тела (выше и ниже бедер, соответственно). Исследования показывают, что рассмотрение потока в большем масштабе вокруг более «влиятельных» частей дает заметный прирост качества распознавания.

Второй и наиболее популярный источник информации, на основе которого происходит обучение сетей, – бинарные маски силуэтов, о которых уже шла речь при рассмотрении неглубоких методов. В простейшем случае [37] сверточная архитектура обучается по отдельным силуэтам предсказывать человека, которому этот силуэт принадлежит. Как и предыдущих методах, сеть в дальнейшем используется для извлечения признаков, причем переход от дескрипторов отдельных кадров ко всему видео происходит путем выбора максимального отклика по циклу походки. Этот метод наиболее простой из всех глубоких подходов, т.к. при наличии масок силуэтов людей не требует практически никакой дополнительной предобработки. Длину цикла походки определяют, рассматривая автокорреляцию последовательности бинарных изображений. Вследствие того, что два кадра, отличающиеся на полный период походки, должны выглядеть похоже, корреляция таких кадров оказывается больше, чем у любой другой пары, что помогает вычислить длину цикла.

Еще один метод, использующий сами силуэты, предложен в [28]. Двухэтапный алгоритм сначала определяет угол съемки видео, а потом по изначальным данным и модели для найденного угла (своей для каждого ракурса) предсказывает человека. Чтобы учитывать не только пространственные, но и временные характеристики движения, на вход сети подаются не отдельные маски, а блоки из нескольких силуэтов, идущих подряд. Кроме того, изменчивость между кадрами учитывается благодаря архитектурам сетей в обеих подзадачах: авторы используют трехмерные сети, в которых свертки производятся не только в пространстве, но и во времени.

Отдельно стоит выделить множество методов, совмещающих «ручное» выделение признаков, и глубинное обучение. В качестве входных данных для сетей часто используют изображения энергии походки, упоминавшиеся выше. Основанные на этой идее модели варьируются от простейших [23], где неглубокая сеть предсказывает человека по получаемым изображениям энергии, вычисленным для различных ракурсов, до более сложных, как, например, [31], где определяется степень похожести пары GEI изображений и исследуются различные методы сравнения нейросетевых признаков походки, полученных из изображений энергии. Еще в нескольких работах [27, 36] также с помощью двух- и трехпоточных сиамских архитектур определяется, какие изображения походки близки и принадлежат одному человеку, а какие -- разным.

Стоит отметить, что большинство популярных в последние годы архитектур и подходов успешно применяются для распознавания по походке. Например, автокодировщики, используемые для генерации изображений и обучения представлений, также оказываются применимы для идентификации. Авторы [34] предлагают решать проблемы вариативности ракурсов и переносимых предметов с помощью множества автокодировщиков, каждый из которых, подобно модели [21], производит свое преобразование. Таким образом, проходя через последовательность «кодирующих» слоев, изображение приводится к боковому ракурсу, более простому для распознавания. Близкий подход предлагается и в [33], однако для преобразования изображения энергии походки к «базовому» виду (снятому сбоку, без

сумки и пальто) применяются генеративные состязующиеся модели (Generative Adversarial Models, GAN).

Метод [10] также базируется на противоборствующих моделях, однако решает задачу верификации, оценивая и трансформируя углы съемки для вычисления признаков, особенных для каждого ракурса. Кроме того, авторы строят изображение энергии периода походки (period energy image, PEI), усредняя силуэты по коротким временным промежуткам внутри цикла походки. Такой подход дает заметный прирост качества по сравнению с изображениями энергии походки, используемыми в большинстве методов.

Наряду с классическими сверточными сетями прямого распространения для распознавания по походке, как и для других задач анализа видео, используются рекуррентные нейронные сети. Такие архитектуры позволяют даже по простейшим данным (например, силуэты в отдельных кадрах [29] вычислять информативные динамические признаки походки. В более эффективном подходе [9] рекуррентные слои применяются к скелету человека, а именно к тепловым картам для суставов, полученным на предыдущих сверточных слоях из отдельных кадров. Такая модель делает предсказание на основании изменения позы человека, не опираясь напрямую на фигуру и силуэт человека, что делает ее более общей.

Многие из обсуждаемых подходов оцениваются на одних и тех же наборах данных при одних и тех же условиях, в этом обзоре мы приводим сравнение некоторых описанных методов и выделяем наиболее успешные решения.

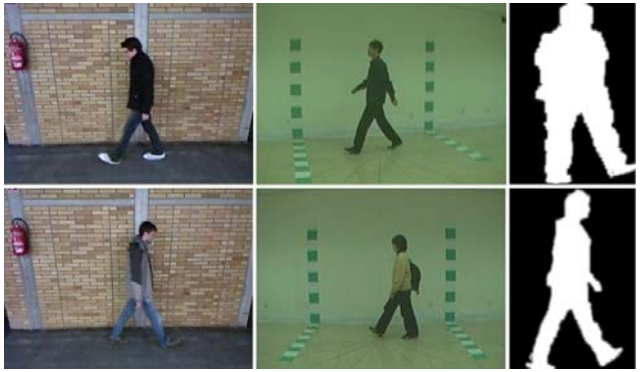


Рис. 2. Примеры кадров из баз данных походок:
TUM-GAID (слева), CASIA-B (посередине) и OULP (справа)
Fig. 2. Examples of frames from the gait database:
TUM-GAID (left), CASIA-B (middle), and OULP (right)

4. Базы данных для распознавания человека по походке

Наиболее широко используемыми в наше время сложными наборами данных для распознавания человека по походке являются базы TUM-GAID [12], OU-ISIR Large Population Dataset (OULP) [13] и CASIA Gait Dataset B [35]. Примеры кадров видео из этих баз можно найти на рис. 2.

Первая база данных используется для распознавания сбоку, все видео в ней сняты под углом 90º, она не очень большая (по 10 видео для 305 человек), однако состоит из полноценных цветных видео, что делает ее применимой для большого количества подходов. Кроме того, в этой базе присутствуют данные, снятые с разницей в полгода, что дает возможность проверить устойчивость алгоритмов к изменениям походки со временем.

Два других набора предназначены для многоракурсного распознавания. В то время, как CASIA – сравнительно небольшая по количеству человек база, но с очень большой

вариативностью ракурсов (11 различных углов съемки от 0 до 180 градусов для 124 человек), набор OULP состоит из видеопоследовательностей для более, чем 4000 человек, снятых двумя камерами, причем ракурс съемки плавно меняется от 55º до 85º. Данные из этой коллекции распространяются в виде масок силуэтов, поэтому не все описанные методы применимы к этой базе данных.

Многие из рассмотренных методов оценены именно на этих наборах данных, поэтому и мы приведем сравнение подходов на них.

5. Результаты работы методов

На описанных базах данных оценка качества алгоритмов происходит следующим образом. Сначала модель настраивается на части данных (как правило, это все видео для некоторого подмножества людей), а затем тестируется на другом наборе данных, состоящем из видео для остальных людей. Для баз TUM и OU-ISIR разделение на обучающую и тестовую выборки предоставлено авторами коллекций. В экспериментах с CASIA модель обучается на первых 24 людях и тестируется на всех оставшихся ста. В качестве метрики качества обычно рассматривается точность распознавания -- доля правильно классифицируемых видео.

Табл. 1. Сравнение результатов распознавания на базе TUM-GAID
Table 1. Comparison of recognition results with the base TUM-GAID

Метод	Точность
Sokolova, блоки ОП [25]	97.5%
Sokolova, части тела [26]	99.8%
Castro, SNN+ SVM [6]	98.0%
Marín-Jiménez [20]	98.9%
Castro, дескрипторы Фишера [6]	99.2%
Zhang [37]	97.7%

В табл. 1 приведено сравнение результатов распознавания на базе TUM-GAID. Наилучшего качества достигает нейросетевой подход [26], однако до его появления глубинные методы долгое время не могли превзойти по качеству метод [6], не использующий нейронные сети. Как будет видно ниже, и в задаче многоракурсного распознавания все еще продолжается борьба глубоких и неглубоких методов.

Табл. 2. Сравнение результатов распознавания видео, снятых в разные дни, на базе TUM-GAID
Table 2. Comparison of video recognition results taken on different days with the base TUM-GAID

Метод	Точность
Castro, CNN + SVM [5]	59.4%
Marín-Jiménez [20]	63.6%
Castro, дескрипторы Фишера [6]	60.4%

Интересно также рассмотреть устойчивость алгоритмов ко времени съемки видео. Оказывается, что качество идентификации сильно портится, если между временем первого попадания человека в поле зрения камер и моментов тестовой съемки и распознавания проходит много времени. В табл. 2 показаны результаты распознавания на базе TUM-GAID, когда между «обучающими» и «тестовыми» видео проходит полгода. Точность каждого из представленных алгоритмов падает примерно на 40%, что говорит о том, что выученные этими алгоритмами признаки плохо переносимы во времени.

Для многоракурсных баз сравнение, как правило, производится для всевозможных пар ракурсов: данные, снятые под некоторым «тестовым» углом классифицируются алгоритмом, при настройке которого используется другой, «обучающий» угол съемки.

Для сравнения различных алгоритмов на базе OU-ISIR используют два популярных протокола тестирования. Один из них, как уже было сказано, предоставлен авторами: из

коллекции выбрано 1912 человек, которые пятью способами разделяются пополам на обучающую и тестовую выборки, после чего качество моделей, обученных на этих разбиениях, усредняется. Второй протокол реализует кросс-валидацию, причем модели строятся на данных для 3844 людей, для которых в базе присутствуют данные с обеих камер. Для удобства результаты сравнения обычно агрегируют, рассматривая разности между «обучающим» и «тестовым» углами. В табл. 3 показана средняя точность методов для каждого из 4 возможных значений разности углов.

Табл. 3. Сравнение результатов распознавания на базе OU-ISIR
Table 3. Comparison of recognition results with the base OU-ISIR

Метод	0º	10º	20º	30º
Zhang [37]	94,1%	71,6%	21.8%	2,9%
Shiraga [23]	94,9%	93,9%	90,5%	80,65%
Li [15]	98,3%	98,2%	97,3%	94,6%
Mansur [19]	96,8%	96,3%	94,2%	90,3%

Самый простой метод [37] оказывается несостоятелен, когда углы съемки сильно отличаются, однако остальные подходы показывают достаточно высокое качество распознавания. Наилучших результатов при таких экспериментах также достигает метод, не использующий нейронные сети.

Табл. 4. Сравнение результатов кросс-валидации на базе OU-ISIR
Table 4. Tab. 4. Comparison of cross-validation results with the base OU-ISIR

Метод	0º	10º	20º	30º
Sokolova [26]	98,4%	98,2%	97,1%	94,1%
Shiraga [23]	96,5%	95,8%	92,5%	84,9%
Wu [31]	98,9%	95,5%	92,4%	85,3%
Takemura [27]	99,3%	99,2%	98,6%	96,9%
He [10]	-	96,7%	93,2%	94,2%

Однако при наличии большего количества данных для обучения и тестирования нейросетевые методы оказываются очень успешны. В табл. 4 приведены результаты сравнения алгоритмов при использовании данных для почти 4 тысяч людей.

Благодаря большему размеру обучающей выборки методы, использующие такой протокол тестирования, достигают более высокой точности. Отсутствие открытых реализаций и общего протокола тестирования не дает возможности сравнить все методы и найти оптимальный, однако даже имеющиеся результаты показывают, что на сегодняшний день глубокие и неглубокие подходы продолжают развиваться и показывают практически равное качество.

Табл. 5. Сравнение средних результатов для трех ракурсов базы CASIA
Table 5. Comparison of average results for three views of the CASIA base

Метод	54º	90º	126º
Sokolova [26]	77.8%	68.8%	74.7%
Wu [31]	77.8%	64.9%	76.1%
Feng [9]	52.2%	60.0%	61.9%
Yu, SPAE [34]	63.3%	62.1%	66.3%
Yu, GaitGAN [33]	64.5%	58.2%	65.7%

Для базы CASIA качество распознавания для различных ракурсов тоже, как правило, агрегируется. Следуя распространенному подходу, приведем в табл. 5 средние значения точности распознавания для тестовых ракурсов 54º, 90º и 126º (в качестве обучающих в каждом эксперименте используются оставшиеся 10 углов).

Кроме классической задачи классификации, в которой для человека в видео необходимо определить, кем из базы он является, задачу распознавания по походке часто формулируют в форме задачи верификации. Для пары видеопоследовательностей с двигающимся человеком требуется определить, разные ли люди в кадре или один и тот же.

Задача верификации интересна не только сама по себе, но и как дополнение к идентификации в случае, если человек впервые попадает в поле зрения камер и его еще нет в индексе. Даже отсутствующий в базе человек будет каким-то образом классифицирован идентификатором, и проверка уверенности модели в своём решении – отдельная сложная задача. Один из возможных подходов к решению – дополнительная верификация пары, состоящей из тестового видео и видео с кандидатом в ответы.

Для задачи верификации могут использоваться все описанные методы выделения признаков походки, однако вместо классификации или нахождения ближайшего объекта на последнем этапе оценивается близость пары дескрипторов и сравнивается с некоторым порогом. Достаточно близкие дескрипторы считаются принадлежащими одному человеку. Для оценки качества в такой задаче, как правило, строится ROC-кривая и считается площадь под ее графиком.

Интересно отметить, что несмотря на то, что по сути решается одна и та же задача распознавания по походке и вычисляются одни и те же дескрипторы, подходы, показывающие самые высокие результаты в задаче идентификации, могут оказаться не самыми успешными при оценке качества верификации. На рис. 3 изображены графики ROC-кривых для нескольких методов многоракурсного распознавания на базе OU-ISIR, предоставленные их авторами.

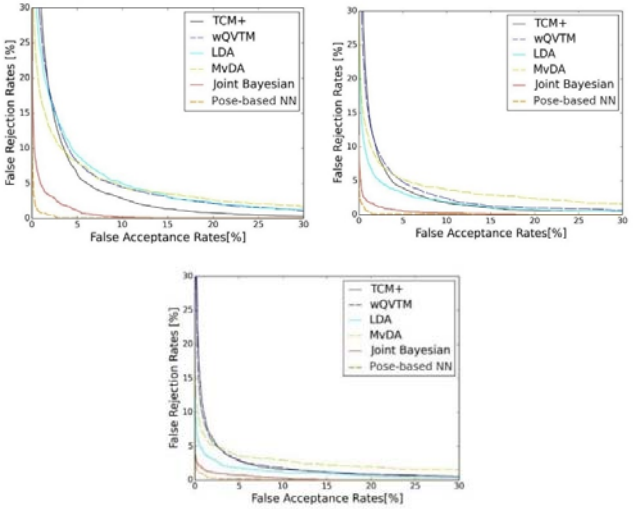


Рис. 3. Сравнение ROC-кривых различных методов для многоракурсной задачи верификации на базе OULP для тестового угла 85° и обучающих углов, равных 55°, 65°, and 75°, соответственно
Рис. 3. Comparison of ROC curves for various methods for a multi-view OULP-based verification task for a test angle of 85° and training angles of 55°, 65°, and 75° respectively

Кривая, которая соответствует подходу [26], отстающему в качестве идентификации от байесовского метода [15], оказывается ниже всех на всех графиках, что показывает, что этот алгоритм более точно определяет, совпадают ли люди в видео. Это в очередной раз

подтверждает, что одного идеального метода до сих пор не существует и разные подходы оказываются лучше при разных условиях и предположениях.

Результаты сравнения методов многоракурсного распознавания показывают, что различные дескрипторы не уступают друг другу в информативности. Методы, основанные на выделении и исследовании силуэтов, решают задачу практически с той же точностью, как и подходы, рассматривающие позу и движение точек в кадре, а иногда и лучше.

6. Заключение

Несмотря на все множество используемых признаков и разнообразие предлагаемых моделей и методов обучения, задача распознавания походки все еще не теряет актуальности: существующие решения пока не достигли идеальной точности идентификации. На представление движения влияет большое количество различных условий, а наборы данных, пригодные для этой задачи, ограничены по сравнению с другими проблемами компьютерного зрения, для которых собраны миллионы изображений лиц или десятки тысяч фигур для реидентификации. Базы данных, собранные на данный момент, пока не способны учесть все возможные вариации походки, что препятствует созданию совершенной модели.

Список литературы

[1] Arseev S., Konushin A., Liutov V. Human Recognition by Appearance and Gait. Programming and Computer Software, vol. 44, issue 4, 2018, pp. 258–265
[2] Khalid Bashir, Tao Xiang, Shaogang Gong. Gait recognition using gait entropy image. In Proc. of the 3rd international conference on crime detection and prevention, 2009, pp. 1–6.
[3] Belhumeur P.N., Hespanha J.P., Kriegman D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, 1997, pp. 711–720.
[4] Chen C., Liang J., Zhao H., Hu H., Tian J. Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters, vol. 30, no. 11, 2009, pp 977–984.
[5] Francisco Manuel Castro, Manuel J. Marín-Jiménez, Nicolás Guil, Nicolás Pérez de la Blanca. Automatic learning of gait signatures for people identification. Lecture Notes in Computer Science, vol. 10306, 2017. pp. 257–270.
[6] Francisco M. Castro, Manuel J. Marín-Jiménez, Rafael Medina Carnicer. Pyramidal Fisher Motion for multiview gait recognition. In Proc. of the 22nd International Conference on Pattern Recognition, 2014, pp. 1692–1697.
[7] Dalal N., Triggs B. Histograms of oriented gradients for human detection. In Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 01. 2005. pp. 886–893.
[8] Deng M., Wang C., Cheng F., Zeng W. Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning. Pattern Recognition, 2017, 67, pp. 186 – 200
[9] Feng Y., Li Y., Luo J. Learning effective gait features using LSTM. In Proc. of the 23rd International Conference on Pattern Recognition (ICPR), 2016, pp. 325–330.
[10] He Y., Zhang J., Shan H., Wang L. Multitask gans for view-specific feature learning in gait recognition. IEEE Transactions on Information Forensics and Security, vol. 14, no. 1, 2019, pp. 102–113.
[11] Han J., Bhanu B. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, issue 2, 2006, pp. 316–322.
[12] Hofmann M., Geiger J., Bachmann S., Schuller B., Rigoll G. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. Journal of Visual Communication and Image Representation, vol. 25, no. 1, 2014, pp.195 – 206.
[13] Iwama H., Okumura M., Makihara Y., Yagi Y. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. IEEE Transactions on Information Forensics and Security, vol. 7, issue 5, 2012, pp.1511–1521.
[14] Laptev I., Marszalek M., Schmid C., Rozenfeld B. Learning Realistic Human Actions from Movies. In Proc. of the CVPR 2008 – IEEE Conference on Computer Vision & Pattern Recognition, 2008, pp. 1–8
[15] Li C., Sun S., Chen X., Min X. Cross-view gait recognition using joint Bayesian. In Proc. of the Ninth International Conference on Digital Image Processing (ICDIP 2017), 2017.

[16] Liu Y., Zhang J., Wang C., Wang L. Multiple HOG templates for gait recognition. In Proc. of the 21st International Conference on Pattern Recognition (ICPR2012). 2012. pp. 2930–2933
[17] Makihara Y., Sagawa R., Mukaigawa Y., Echigo T., Yagi Y. Gait recognition using a view transformation model in the frequency domain. Lecture Notes in Computer Science, vol. 3953, 2006, pp. 151–163.
[18] Makihara Y., Suzuki A., Muramatsu D., Li X., Yagi Y. Joint intensity and spatial metric learning for robust gait recognition. In Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6786–6796
[19] Mansur A., Makihara Y., Muramatsu D., Yagi Y. Cross-view gait recognition using view-dependent discriminative analysis. In Proc. of the 2014 IEEE/IAPR International Joint Conference on Biometrics, 2014.
[20] M.J. Marín-Jiménez, F.M. Castro, N. Guil, F. de la Torre, R. Medina-Carnicer. Deep multitask learning for gait-based biometrics. In Proc. of the IEEE International Conference on Image Processing (ICIP), 2017.
[21] Muramatsu D., Makihara Y., Yagi Y. View transformation model incorporating quality measures for cross-view gait recognition. IEEE transactions on cybernetics, 2016, vol. 46, issue 7, pp. 1602–1615.
[22] Muramatsu D., Makihara Y., Yagi Y. Crossview gait recognition by fusion of multiple transformation consistency measures. IET Biometrics, vol. 4, issue 2, 2015, pp. 62–73.
[23] Shiraga K., Makihara Y., Muramatsu D., Echigo T., Yagi Y. GEINet: View-invariant gait recognition using a convolutional neural network. In Proc. of the 2016 International Conference on Biometrics (ICB), 2016, pp. 1–8.
[24] Simonyan K., Zisserman A. Two-stream convolutional networks for action recognition in videos. In Proc. of the of the 27th International Conference on Neural Information Processing Systems, vol. 1 of NIPS'14, 2014, pp. 568–576.
[25] Sokolova A., Konushin A. Gait recognition based on convolutional neural networks. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. XLII-2/W4, 2017. pp. 207–212.
[26] Sokolova A. and Konushin A. Pose-based Deep Gait Recognition. IET Biometrics, 2018
[27] Takemura N., Makihara Y., Muramatsu D. On input/output architectures for convolutional neural network-based crossview gait recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2017.
[28] Thapar D., Nigam A., Aggarwa, D., Agarwal P. VGR-net: A view invariant gait recognition network. In Proc. of the EEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA), 2018, pp. 1-8.
[29] Tong S., Fu Y., Ling H., Zhang E. Gait identification by joint spatial-temporal feature. Lecture Notes in Computer Science, vol. 10568, 2017, pp. 457–465.
[30] Whytock T., Belyaev A., Robertson N.M. Dynamic distance-based shape features for gait recognition. Journal of Mathematical Imaging and Vision, vol. 50, no. 3, 2014, pp. 314–326.
[31] Wu Z., Huang Y., Wang L., Wang X., Tan T. A comprehensive study on cross-view gait based human identification with deep cnns. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 2, 2016, pp. 209–226.
[32] Yang Y., Tu D., Li G. Gait recognition using flow histogram energy image. In Proc. of the 22nd International Conference on Pattern Recognition, 2014, pp. 444–449
[33] Yu S., Chen H., Reyes E. B. G., Poh, N. GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Network. In Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp 532–539
[34] Yu S., Chen H., Wang Q., Shen L., Huang Y. Invariant feature extraction for gait recognition using only one uniform model. Neurocomputing, vol. 239, 2017, pp. 81 – 93
[35] Yu S., Tan D., Tan T. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In Proc. of the 18th International Conference on Pattern Recognition (ICPR), vol. 4, 2006, pp. 441–444
[36] Zhang C. Liu W., Ma H., Fu H. Siamese neural network based gait recognition for human identification. In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 2832-2836.
[37] Zhang X., Sun S., Li C., Zhao X., Hu Y. Deepgait: A learning deep convolutional representation for gait recognition. Lecture Notes in Computer Science, vol. 10568, 2017, pp. 447–456.

Methods of gait recognition in video

[1] *A.I. Sokolova <nkgorelits@2100.gosniias.ru>*
[2] *A.S. Konushin <anton.konushin@grapics.cs.msu.ru>*
[1] *National research University Higher School of Economics, 20, Myasnitskaya st., Moscow, 101000, Russian*
[2] *Lomonosov Moscow State University Leninskie Gory, Moscow, 119991, Russia*

Abstract. Human gait is an important biometric index that allows to identify a person at a great distance without direct contact. Due to these qualities, which other popular identifiers such as fingerprints or iris do not have, the recognition of a person by the manner of walking has become very common in various areas where video surveillance systems can be used. With the development of computer vision techniques, a variety of approaches for human identification by movements in a video appear. These approaches are based both on natural biometric characteristics (human skeleton, silhouette, and their change during walking) and abstract features trained automatically which do not have physical justification. Modern methods combine classical algorithms of video and image analysis and new approaches that show excellent results in related tasks of computer vision, such as human identification by face and appearance or action and gesture recognition. However, due to the large number of conditions that can affect the walking manner of a person itself and its representation in video, the problem of identifying a person by gait still does not have a sufficiently accurate solution. Many methods are overfitted by the conditions presented in the databases on which they are trained, which limits their applicability in real life. In this paper, we provide a survey of state-of-the-art methods of gait recognition, their analysis and comparison on several popular video collections and for different formulations of the problem of recognition. We additionally reveal the problems that prevent the final solution of gait identification challenge.

Keywords: gait; biometrics; silhouette; neural networks; identification

DOI: 10.15514/ISPRAS-2019-31(1)-5

For citation: Sokolova A.I., Konushin A.S. Methods of gait recognition in video. Trudy ISP RAN/Proc. ISP RAS, vol. 31, issue 1, 2019. pp. 69-82 (in Russian). DOI: 10.15514/ISPRAS-2019-31(1)-5

References

[1] Arseev S., Konushin A., Liutov V. Human Recognition by Appearance and Gait. Programming and Computer Software, vol. 44, issue 4, 2018, pp. 258–265
[2] Khalid Bashir, Tao Xiang, Shaogang Gong. Gait recognition using gait entropy image. In Proc. of the 3rd international conference on crime detection and prevention, 2009, pp. 1–6.
[3] Belhumeur P.N., Hespanha J.P., Kriegman D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, 1997, pp. 711–720.
[4] Chen C., Liang J., Zhao H., Hu H., Tian J. Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters, vol. 30, no. 11, 2009, pp 977–984.
[5] Francisco Manuel Castro, Manuel J. Marín-Jiménez, Nicolás Guil, Nicolás Pérez de la Blanca. Automatic learning of gait signatures for people identification. Lecture Notes in Computer Science, vol. 10306, 2017. pp. 257–270.
[6] Francisco M. Castro, Manuel J. Marín-Jiménez, Rafael Medina Carnicer. Pyramidal Fisher Motion for multiview gait recognition. In Proc. of the 22nd International Conference on Pattern Recognition, 2014, pp. 1692–1697.
[7] Dalal N., Triggs B. Histograms of oriented gradients for human detection. In Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 01. 2005. pp. 886–893.
[8] Deng M., Wang C., Cheng F., Zeng W. Fusion of spatial-temporal and kinematic features for gait recognition with deterministic learning. Pattern Recognition, 2017, 67, pp. 186 – 200
[9] Feng Y., Li Y., Luo J. Learning effective gait features using LSTM. In Proc. of the 23rd International Conference on Pattern Recognition (ICPR), 2016, pp. 325–330.
[10] He Y., Zhang J., Shan H., Wang L. Multitask gans for view-specific feature learning in gait recognition. IEEE Transactions on Information Forensics and Security, vol. 14, no. 1, 2019, pp. 102–113.

[11] Han J., Bhanu B. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, issue 2, 2006, pp. 316–322.
[12] Hofmann M., Geiger J., Bachmann S., Schuller B., Rigoll G. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. Journal of Visual Communication and Image Representation, vol. 25, no. 1, 2014, pp.195 – 206.
[13] Iwama H., Okumura M., Makihara Y., Yagi Y. The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. IEEE Transactions on Information Forensics and Security, vol. 7, issue 5, 2012, pp.1511–1521.
[14] Laptev I., Marszalek M., Schmid C., Rozenfeld B. Learning Realistic Human Actions from Movies. In Proc. of the CVPR 2008 – IEEE Conference on Computer Vision & Pattern Recognition, 2008, pp. 1–8
[15] Li C., Sun S., Chen X., Min X. Cross-view gait recognition using joint Bayesian. In Proc. of the Ninth International Conference on Digital Image Processing (ICDIP 2017), 2017.
[16] Liu Y., Zhang J., Wang C., Wang L. Multiple HOG templates for gait recognition. In Proc. of the 21st International Conference on Pattern Recognition (ICPR2012). 2012. pp. 2930–2933
[17] Makihara Y., Sagawa R., Mukaigawa Y., Echigo T., Yagi Y. Gait recognition using a view transformation model in the frequency domain. Lecture Notes in Computer Science, vol. 3953, 2006, pp. 151–163.
[18] Makihara Y., Suzuki A., Muramatsu D., Li X., Yagi Y. Joint intensity and spatial metric learning for robust gait recognition. In Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6786–6796
[19] Mansur A., Makihara Y., Muramatsu D., Yagi Y. Cross-view gait recognition using view-dependent discriminative analysis. In Proc. of the 2014 IEEE/IAPR International Joint Conference on Biometrics, 2014.
[20] M.J. Marín-Jiménez, F.M. Castro, N. Guil, F. de la Torre, R. Medina-Carnicer. Deep multitask learning for gait-based biometrics. In Proc. of the IEEE International Conference on Image Processing (ICIP), 2017.
[21] Muramatsu D., Makihara Y., Yagi Y. View transformation model incorporating quality measures for cross-view gait recognition. IEEE transactions on cybernetics, 2016, vol. 46, issue 7, pp. 1602–1615.
[22] Muramatsu D., Makihara Y., Yagi Y. Crossview gait recognition by fusion of multiple transformation consistency measures. IET Biometrics, vol. 4, issue 2, 2015, pp. 62–73.
[23] Shiraga K., Makihara Y., Muramatsu D., Echigo T., Yagi Y. GEINet: View-invariant gait recognition using a convolutional neural network. In Proc. of the 2016 International Conference on Biometrics (ICB), 2016, pp. 1–8.
[24] Simonyan K., Zisserman A. Two-stream convolutional networks for action recognition in videos. In Proc. of the of the 27th International Conference on Neural Information Processing Systems, vol. 1 of NIPS'14, 2014, pp. 568–576.
[25] Sokolova A., Konushin A. Gait recognition based on convolutional neural networks. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. XLII-2/W4, 2017. pp. 207–212.
[26] Sokolova A. and Konushin A. Pose-based Deep Gait Recognition. IET Biometrics, 2018
[27] Takemura N., Makihara Y., Muramatsu D. On input/output architectures for convolutional neural network-based crossview gait recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2017.
[28] Thapar D., Nigam A., Aggarwa, D., Agarwal P. VGR-net: A view invariant gait recognition network. In Proc. of the EEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA), 2018, pp. 1-8.
[29] Tong S., Fu Y., Ling H., Zhang E. Gait identification by joint spatial-temporal feature. Lecture Notes in Computer Science, vol. 10568, 2017, pp. 457–465.
[30] Whytock T., Belyaev A., Robertson N.M. Dynamic distance-based shape features for gait recognition. Journal of Mathematical Imaging and Vision, vol. 50, no. 3, 2014, pp. 314–326.
[31] Wu Z., Huang Y., Wang L., Wang X., Tan T. A comprehensive study on cross-view gait based human identification with deep cnns. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 2, 2016, pp. 209–226.
[32] Yang Y., Tu D., Li G. Gait recognition using flow histogram energy image. In Proc. of the 22nd International Conference on Pattern Recognition, 2014, pp. 444–449
[33] Yu S., Chen H., Reyes E. B. G., Poh, N. GaitGAN: Invariant Gait Feature Extraction Using Generative Adversarial Network. In Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp 532–539
[34] Yu S., Chen H., Wang Q., Shen L., Huang Y. Invariant feature extraction for gait recognition using only one uniform model. Neurocomputing, vol. 239, 2017, pp. 81 – 93

[35] Yu S., Tan D., Tan T. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. In Proc. of the 18th International Conference on Pattern Recognition (ICPR), vol. 4, 2006, pp. 441–444
[36] Zhang C. Liu W., Ma H., Fu H. Siamese neural network based gait recognition for human identification. In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 2832-2836.
[37] Zhang X., Sun S., Li C., Zhao X., Hu Y. Deepgait: A learning deep convolutional representation for gait recognition. Lecture Notes in Computer Science, vol. 10568, 2017, pp. 447–456.