

БАЙЕСОВСКАЯ МОДЕЛЬ ОБЪЕДИНЕНИЯ ПРИЗНАКОВЫХ ПРЕДСТАВЛЕНИЙ В ЗАДАЧЕ РЕИДЕНТИФИКАЦИИ ЛИЧНОСТИ¹

Русаков К. Д.²

(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)

Рассматривается задача реидентификации человека по изображениям, полученным из систем видеонаблюдения, с использованием признаков различной природы. В отличие от традиционных подходов, опирающихся на один биометрический фактор, предложена универсальная методика объединения признаков, полученных из разных источников информации. В основе метода лежит вероятностная модель, реализующая объединение признаков в пространстве решений по правилу байесовского вывода. Каждый признак рассматривается как наблюдение, аппроксимируемое нормальным распределением с параметрами, оцениваемыми по обучающим данным. Принятие решения осуществляется путём максимизации апостериорной вероятности принадлежности объекта к одной из известных личностей. В качестве средств извлечения признаков используются модели на основе архитектур с механизмами внимания, обеспечивающие устойчивость к искажениям. Проведено сравнение с методом линейного объединения признаков. Результаты экспериментов на открытом наборе данных демонстрируют повышение точности идентификации, устойчивость к частичному отсутствию данных и возможность количественной оценки степени достоверности принимаемого решения, что актуально для применения в системах обеспечения безопасности.

Ключевые слова: реидентификация личности, объединение признаков, байесовская модель, компьютерное зрение, трансформерные архитектуры, видеонаблюдение.

1. Введение

Реидентификация человека (person re-identification, ReID) — задача поиска и распознавания одного и того же человека на различных снимках или в разных камерах наблюдения [8]. Актуальность ReID обусловлена активным развитием систем видеонаблюдения и вопросами безопасности: автоматическое отсле-

¹ Автор признателен д.т.н., профессору РАН, г.н.с. Р.В. Мецержакову за ценное обсуждение содержания статьи.

² Константин Дмитриевич Русаков, н.с. (rusakov@ipu.ru).

живание перемещений подозреваемых, поиск пропавших людей и т.п. [19].

Традиционно ReID опирается либо на биометрию лица, либо на общие признаки фигуры (силуэт, одежда) человека по отдельности. Однако только на одном источнике (лицо или тело) достичь высокой надёжности распознавания сложно из-за вариаций позы, освещения, разрешения и перекрытий. Совмещение нескольких биометрических признаков позволяет повысить точность за счёт взаимодополняющей информации различных модальностей [12, 13].

В частности, силуэт человека считается почти таким же уникальным, как и лицо, и уже используется на практике. Так, новейшие отечественные системы умного видеонаблюдения отслеживают траекторию по силуэтам и осуществляют поиск по фигуре в реальном времени [7]. Вместе с тем возникает вопрос эффективной интеграции разнородных признаков.

Наиболее простой подход объединения признаков — линейная комбинация, например, конкатенация векторов лица и тела или усреднение оценок совпадения [13]. Однако жёсткое фиксированное слияние имеет ряд ограничений. Во-первых, оно не учитывает надёжность каждой модальности: «шумные» данные (например, размытое лицо или скрытая одежда) могут снижать итоговую точность, тогда как оптимально было бы придать им меньший вес. Во-вторых, линейное слияние не отражает вероятностную природу биометрических измерений и не даёт информации о степени уверенности. В результате система может ошибаться в условиях искажённого входа, не имея механизма выражения неуверенности в принятом решении.

Цель данной работы — разработать байесовскую схему интеграции признаков лица и силуэта для задачи реидентификации, которая учитывает статистическую неопределённость данных и обеспечивает более надёжное распознавание по сравнению с традиционным детерминированным объединением. В данной работе впервые предлагается использовать вероятностную модель на основе теоремы Байеса для объединения многомодаль-

ных представлений лица и тела.

Научная новизна заключается в том, что байесовский подход позволяет адаптивно взвешивать вклад каждой модальности через апостериорную вероятность, повышая устойчивость системы к шуму, изменению условий съёмки и отсутствию одного из биометрических факторов.

В работе представлены:

а) формальная модель и алгоритм байесовской интеграции признаков;

б) её программная реализация с использованием современных архитектур сверточных нейронных сетей (ArcFace и трансформер Vision Transformer);

в) экспериментальное сравнение с традиционным линейным подходом на открытом датасете.

Ниже рассматриваются связанные исследования, описывается метод, экспериментальные результаты, обсуждаются преимущества и ограничения, и даются выводы.

2. Связанные работы

2.1. РЕИДЕНТИФИКАЦИЯ ПО ЛИЦУ И ТЕЛУ

Задача распознавания лица по изображениям длительное время развивалась независимо от задачи ReID по силуэту, формируя собственные подходы и методы. С появлением глубоких нейросетей были достигнуты выдающиеся результаты: так, архитектура FaceNet с триплетным обучением [17] достигла точности 99,6% на эталонном датасете LFW.

В дальнейшем были предложены специальные функции потерь, направленные на повышение дискриминативности векторов лица. Одной из наиболее эффективных стала функция ArcFace [10], которая вводит аддитивную угловую *margin*-поправку в классификационное обучение, обеспечивая более плотные кластеры векторов для каждого человека. Сегодня ArcFace де-факто является стандартом в задачах распознавания лиц; его эффективность подтверждена как зарубежными, так и отечественными исследованиями [4].

Тем не менее даже лучшие модели распознавания лиц могут давать сбои в условиях неконтролируемой съёмки: при низком разрешении, поворотах головы, частичном закрытии лица (маски, шарфы) [6]. В частности, отечественные работы отмечают снижение качества на «нетипичных» лицах, которые существенно отличаются от доминирующих в обучающих данных (например, лица детей или представителей редких рас) [6], и предлагают методы предварительной фильтрации «сложных» случаев для повышения надёжности [5, 6].

Одновременно со стороны ReID по фигуре также наблюдается значительный прогресс. Классические подходы опирались на признаки внешнего вида: цветовые гистограммы, HOG-описатели, локальные инвариантные признаки и др. [8]. С развитием глубокого обучения акцент сместился на обучаемые векторы изображений полноростового человека [19].

Архитектуры на основе сверточных нейронных сетей (например, ResNet-50) обучаются либо классифицировать идентичности в тренировочном наборе с использованием функции Softmax, либо непосредственно оптимизировать векторы с помощью метрического обучения (триплет-лосс, контрастивный лосс и т.п.) [11]. Для повышения устойчивости к вариациям позы используются стратегии деления тела на части и механизмы внимания.

Так, Harmonious Attention Network (HAN) объединяет пространственное и каналовое внимание для выделения наиболее информативных областей изображения человека [16]. Другие работы задействуют скелетную позу как условие для объединения признаков или применяют графовые нейронные сети (Graph Convolutional Networks, GCN) для моделирования взаимосвязей между частями тела и визуальными атрибутами [14].

В целом современные модели person re-identification демонстрируют высокое качество на стандартных бенчмарках (Market-1501, MSMT17 и др.), особенно при использовании нескольких масштабов и агрегации локальных признаков частей тела [2]. Тем не менее, большинство существующих решений опираются ис-

ключительно на визуальную модальность — внешний вид человека.

2.2. ИНТЕГРАЦИЯ МНОГОМОДАЛЬНЫХ БИОМЕТРИЧЕСКИХ ПРИЗНАКОВ

Комбинация признаков лица и полного силуэта представляет собой мультибиометрический подход к идентификации. Ещё в ранних работах отмечалось, что объединение нескольких биометрических идентификаторов (например, лица, радужной оболочки, голоса) повышает общую точность системы за счёт разнообразности используемых источников [9].

В задаче ReID предпринимались различные подходы к multi-modal fusion на разных уровнях. Простейший из них — объединение на уровне признаков: например, конкатенация векторов лица и тела с последующей классификацией или вычислением расстояний [13]. Такой метод использован в работе [18], где отдельные нейросети извлекают признаки лица (FaceNet) и тела (OSNet), после чего векторы объединяются и сравниваются с эталонными комбинациями; авторы демонстрируют прирост порядка 5% в метрике Rank-1 по сравнению с использованием только тела.

Более сложные схемы реализуют слияние на уровне оценок (score-level fusion), например, посредством усреднения или обучаемого взвешивания сходства между лицом и фигурой [13]. В системах высокого уровня применяются также байесовские сети доверия, где выходы детектора лица и дополнительная информация (пол, возраст, рост) интегрируются для повышения надёжности идентификации в робототехнических приложениях [12].

Важным направлением является применение механизмов перекрёстного внимания между модальностями, позволяющих модели самостоятельно выявлять, каким признакам следует уделять больше внимания. В недавней работе [20], объединяющей визуальные и языковые описания, демонстрируется улучшение ReID за счёт одновременного использования лица и тела: на основе изображения генерируется текстовое описание, и модули внимания сопоставляют текст с соответствующими областями (лицо, тело) на изображении, обеспечивая более полную

идентификацию.

Графовые подходы, в частности основанные на GCN, представляют мультимодальные признаки в виде графа, где узлы соответствуют векторам разных модальностей одного человека, а рёбра отражают их принадлежность к одной личности [14]. Обучение такого графа позволяет эффективно распространять информацию между модальностями.

В большинстве рассмотренных подходов (за исключением байесовских) интеграция осуществляется либо через обучаемые веса (что по сути имитирует байесовский вывод, но неявно), либо через механизмы внимания. Однако явная вероятностная интеграция по правилу Байеса до сих пор не применялась для слияния векторов лица и тела. Между тем, в смежных областях байесовские методы доказали свою эффективность при принятии решений на основе нескольких источников информации [9]. В данной работе мы восполняем этот пробел, предложив явную байесовскую модель для объединения векторов в задаче ReID.

Таким образом, предложенный подход концептуально близок к мультибиометрическим системам, осуществляющим объединение признаков на уровне принятия решения. Однако, в отличие от существующих решений, он основан не на эвристических или обучаемых весах, а на строгом байесовском выводе.

Разработанная схема может быть интерпретирована как наивный байесовский классификатор идентичности, в котором признаками выступают векторы лица и силуэта. Ранее наивный Байес использовался в биометрических системах для принятия решений по отпечаткам пальцев и изображениям лиц [9], но не в контексте глубоких нейросетевых векторов.

Следует также отметить, что в работе применяются современные высокоуровневые подходы — ArcFace и Vision Transformer, обеспечивающие компактное и устойчивое представление каждой модальности. Это принципиально отличает предложенный метод от более ранних работ, основанных на низкоуровневых визуальных признаках [8].

3. Байесовская схема интеграции признаков

3.1. БАЙЕСОВСКАЯ МОДЕЛЬ ИНТЕГРАЦИИ

Задача распознавания моделируется как статистическая классификация личности y по двум наблюдениям: вектору признаков лица f и вектору признаков силуэта s . Пусть y – дискретная случайная переменная, принимающая значения из множества идентичностей Y (известных системе лиц). Тогда по теореме Байеса апостериорная вероятность того, что наблюдения (f, s) принадлежат человеку y , определяется как

$$(1) \quad P(y | f, s) \propto P(f | y) \cdot P(s | y) \cdot P(y),$$

где $P(f | y)$ – правдоподобие наблюдаемого признака лица при условии личности y , $P(s | y)$ – правдоподобие признака силуэта, а $P(y)$ – априорная вероятность личности y . В нашем случае априори все кандидаты равновероятны (непредвзятая идентификация), поэтому далее полагаем $P(y) = \text{const}$. Таким образом, решение о том, что на изображении представлен человек y^* , принимается по правилу максимума апостериорной вероятности (MAP)

$$(2) \quad y^* = \arg \max_{y \in Y} \left\{ \ln P(f | y) + \ln P(s | y) \right\},$$

где переход к логарифмическим правдоподобиям заменяет произведение плотностей на сумму логарифмов и предотвращает переполнение при работе с малыми величинами вероятностей, а также обеспечивает численную устойчивость вычислений. Выражение (2) показывает, что решение складывается из двух логарифмических правдоподобий: от лица и от силуэта. Если, скажем, лицо плохо распознано (низкое $P(f | y)$), но силуэт совпал уверенно (высокое $P(s | y)$), то личность всё равно может быть правильно определена за счёт второй части. И наоборот, при противоречивых признаках будет выбран тот кандидат y , для которого произведение плотностей (или сумма логарифмов) максимизируется, т.е. оба признака достаточно соответствуют этому кандидату лучше, чем любому другому. Такая схема эквивалентна наивному байесовскому классификатору, учитывающему независимые «признаковые условия» f и s для каждого класса y [9].

В нашем случае независимость признаков лица и тела – упрощающее предположение, оправданное тем, что алгоритмы извлечения векторов обучены разделять эти признаки в различные пространства.

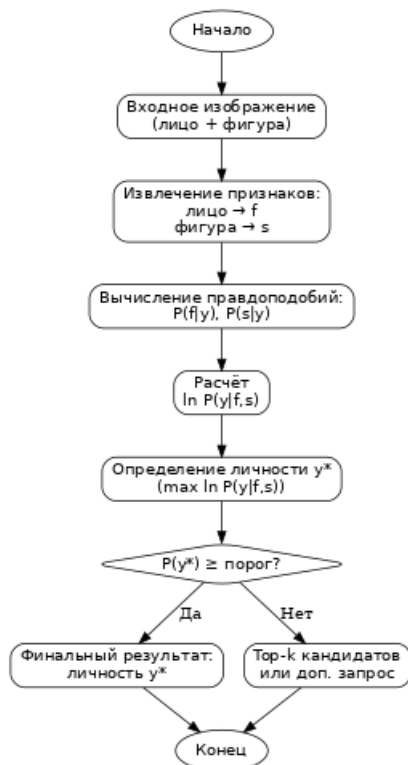


Рис. 1. Блок-схема предлагаемого алгоритма байесовской реидентификации

Для вычисления $P(f | y)$ и $P(s | y)$ необходимо построить вероятностную модель распределения векторов каждой модальности для каждой личности. В данном исследовании выбирается простейшая модель — многомерное нормальное распределение (гауссиан) в пространстве векторов. Предполагается, что для каждой личности y признаки лица f распределены нормаль-

но с некоторым средним μ_y^f и ковариационной матрицей Σ_y^f , то же верно для признаков силуэта с параметрами μ_y^s , Σ_y^s . Эти параметры можно оценить по обучающей выборке векторов (т.е. по нескольким изображениям каждого человека из базы). Тогда плотность имеет вид

$$(3) \quad P(f | y) = \frac{1}{(2\pi)^{d/2} |\Sigma_y^f|^{1/2}} \exp\left(-\frac{1}{2} (f - \mu_y^f)^\top \Sigma_y^{f-1} (f - \mu_y^f)\right),$$

и аналогично для $P(s | y)$. Здесь d — размерность пространства вектора (для ArcFace $d = 128$). На практике для устойчивости и уменьшения числа параметров возможно использование диагональной, или изотропной (скалярной $\sigma_y^2 I$) ковариации, тогда формулы упрощаются до зависимости от евклидова или махалановисова расстояния $|f - \mu_y^f|$.

В эксперименте была использована $\Sigma_y^f = \sigma^2 I$ (одинаковая фиксированная ковариационная матрица для всех y , оценённая по совокупности данных), чтобы избежать проблемы ограниченной статистики на каждого человека. Это допущение превращает выражение (2) в простой критерий по сумме квадратов расстояний: максимизируется $-|f - \mu_y^f|^2 - |s - \mu_y^s|^2$.

Таким образом, байесовский подход в данном упрощении сводится к распознаванию по ближайшему центроиду в объединённом пространстве, что эквивалентно взвешенному (по дисперсии) сложению признаков. Полноценная же байесовская модель (3) позволяет задать разные ковариации, учитывая, например, что разброс признаков лица для данной личности меньше, чем разброс признаков фигуры (что логично, так как одежда меняется).

На рис. 1 представлена блок-схема предлагаемого алгоритма. Сначала из входного изображения с помощью двух нейросетевых моделей извлекаются векторы: f — 128-мерный вектор лица и s — d_s -мерный вектор тела (в эксперименте $d_s = 256$).

Затем эти признаки поступают в модуль байесовского классификатора, где вычисляются правдоподобия (3) для каждого из N зарегистрированных в системе пользователей (личностей). Перемножая плотности (или суммируя лог-правдоподобия) со-

гласно формуле (1), модуль выдаёт апостериорные вероятности $P(y \mid f, s)$ по всем классам $y = 1, \dots, N$.

Финальное решение — идентификатор y^* — выбирается по максимальной апостериорной вероятности. Кроме того, вероятности могут быть использованы для оценки уверенности: если максимум $P(y^* \mid f, s)$ низок, система может воздержаться от однозначного ответа (выдать топ- k кандидатов либо запросить дополнительный образ). Ниже приведён алгоритм.

Алгоритм 1 Байесовская идентификация

1: **Вход:** изображение с лицом и фигурой человека

Извлечение признаков:

2: $f \leftarrow \text{Embedder}_{\text{face}}(\text{изображение лица})$

3: $s \leftarrow \text{Embedder}_{\text{body}}(\text{изображение тела})$

Байесовский классификатор:

4: **for all** $y \in Y$ **do**

5: $L_y^f \leftarrow P(f \mid y)$ ▷ по формуле (3)

6: $L_y^s \leftarrow P(s \mid y)$

7: $\text{posterior}_y \leftarrow L_y^f \times L_y^s$ ▷ так как $P(y) = \text{const}$

8: **end for**

Идентификация:

9: $y^* \leftarrow \arg \max_y \text{posterior}_y$

10: вывести y^* как опознанную личность

11: при необходимости вывести posterior_{y^*} как степень уверенности

3.2. АРХИТЕКТУРА КОДИРОВЩИКОВ

В предлагаемом подходе применяются две нейросетевые модели (кодировщика), извлекающие векторные признаки из изображений: одна — для лица, другая — для силуэта человека. Обе модели построены на базе Vision Transformer (ViT) и используют функцию потерь ArcFace [10], доказавшую свою эффективность для задач биометрической идентификации.

Для каждого изображения соответствующая модель извлекает вектор:

$$(4) \quad f := \text{Embedder}_{\text{face}}(I_{\text{лицо}}), \quad s := \text{Embedder}_{\text{body}}(I_{\text{тело}}),$$

где $f \in \mathbb{R}^{128}$ — вектор лица, $s \in \mathbb{R}^{256}$ — вектор силуэта.

Архитектура ViT представлена на рис. 2 [22]. Она разбивает изображение на непересекающиеся патчи, кодирует их в линейном пространстве и передаёт в стек self-attention-блоков трансформера. На выходе берётся вектор [CLS] как обобщённое представление всего изображения. Такая структура эффективна как для лиц, так и для тел, что делает её универсальной основой для обеих модальностей.

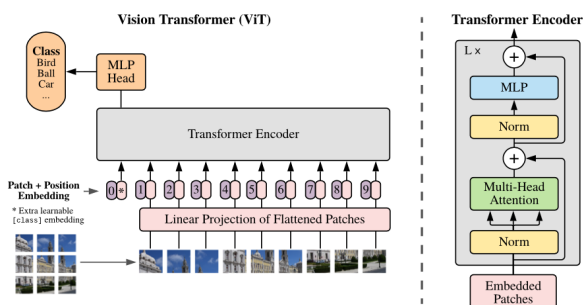


Рис. 2. Архитектура Vision Transformer для извлечения признаков изображения. Входное изображение делится на патчи фиксированного размера, которые подаются в трансформерный энкодер. На выходе используется классификационный токен ([CLS]) как вектор изображения. MLP-голова отбрасывается

Функция потерь ArcFace [10] используется для повышения дискриминативности представлений, вводя аддитивную угловую поправку в классификационную границу. Это позволяет векторам формироваться в плотные и разделённые кластеры. Альтернативно в задачах re-identification может применяться Triplet Loss [11] — метрическая функция, минимизирующая расстояние между положительными парами и увеличивающая расстояние до

отрицательных. Однако в наших экспериментах ArcFace показал лучшую сходимость и стабильность обучения.

Перед вычислением плотностей векторы нормализуются по L_2 -норме:

$$(5) \quad \hat{f} := \frac{f}{\|f\|_2}, \quad \hat{s} := \frac{s}{\|s\|_2}.$$

В отличие от классического подхода, центрирование признаков относительно среднего по обучающей выборке не производится, поскольку параметры многомерного нормального распределения (среднее и ковариационная матрица) для каждого класса оцениваются напрямую по обучающим данным.

Для объединения информации из разных модальностей — лица и силуэта — применяется байесовская схема. Итоговая логарифмическая апостериорная вероятность принадлежности к классу вычисляется как сумма логарифмов модальности-зависимых плотностей:

$$(6) \quad \ln P(y|f, s) \propto \ln P(f|y) + \ln P(s|y).$$

При этом весовые коэффициенты не используются, т.е. вклад каждой модальности считается равным. Эксперименты показали, что даже при равномерной агрегации признаки обеих модальностей вносят значимый вклад в классификацию, а предварительная L_2 -нормализация обеспечивает стабильную работу подхода вне зависимости от размерности векторов.

4. Эксперимент

4.1. ПРОТОКОЛ ЭКСПЕРИМЕНТА

В данном разделе приводится полное описание проведения эксперимента по оценке предложенной байесовской схемы интеграции признаков лица и тела на валидационном сплите датасета SUNK03 [15]. Для каждой личности в каждой модальности было случайным образом отобрано по два изображения в качестве опорного набора, остальные изображения той же личности вошли в проверочный набор. Классы включались в эксперимент только при наличии не менее двух кадров и в лице, и в силуэте, что позволило синхронизировать разбиение.

Изображения приводились к размеру 112×112 и нормировались по каналам с параметрами ($\mu = 0,5, \sigma = 0,5$). Векторы извлекались двумя моделями MobileFaceNet [21]: для лица — размерностью 128, для силуэта — 256. Все векторы нормировались по L_2 -норме. На опорном наборе для каждого класса рассчитывались средний вектор и полная ковариационная матрица с регуляризацией $10^{-3}I$.

Байесовский метод принимал решение по формуле

$$\ln P(y | f, s) = \ln \mathcal{N}(f; \mu_y^f, \Sigma_y^f) + \ln \mathcal{N}(s; \mu_y^s, \Sigma_y^s),$$

а в качестве признанного класса выбирал y с максимальным значением лог-апостериорного шанса. Для сравнения рассматривались также следующие подходы:

- метод объединения векторов с конкатенацией и ближайшим соседом по косинусному сходству;
- метод только по лицу;
- метод только по силуэту.

4.2. РЕЗУЛЬТАТЫ МЕТРИК

В таблице 1 приведены ключевые метрики классификации, полученные на проверочном наборе данных. Рассматриваются четыре подхода: предложенный байесовский алгоритм, метод объединения признаков (конкатенация векторов лица и силуэта), а также базовые одномодальные методы — по лицам и по силуэтам соответственно. Для каждого метода рассчитаны значения точности (precision), полноты (recall), F_1 -меры и площади под ROC-кривой (AUC).

Таблица 1. Ключевые метрики на проверочном наборе

Метод	Точность	Полнота	F_1 -мера	AUC
Байесовский метод	0,9565	0,9554	0,9431	0,9993
Метод объединения	0,9240	0,9198	0,9052	0,9980
Метод по лицам	0,6222	0,6763	0,6028	0,9454
Метод по силуэтам	0,8579	0,8582	0,8329	0,9862

Из таблицы 1 следует, что байесовский метод демонстрирует наивысшие значения по всем метрикам. Особенно заметно его преимущество по F_1 -мере и AUC: при $F_1 = 0,9431$

и $AUC=0,9993$ он превосходит метод объединения, а также значительно опережает одномодальные варианты, особенно в части работы по лицевым признакам. На рис. 3 представлены ROC-кривые (micro-AUC) для всех подходов; байесовский метод достигает 0,9993, что выше, чем у метода объединения (0,9980) и одномодальных решений (лицо — 0,9454; силуэт — 0,9862). Кривые Precision—Recall на рис. 4 показывают $AP = 0,951$ для байесовского метода против 0,932 у объединения. На рис. 5 приведена t-SNE-визуализация векторов байесовского метода для десяти наиболее частых классов: кластеры оказались более плотными и разделёнными по сравнению с конкатенированными векторами.

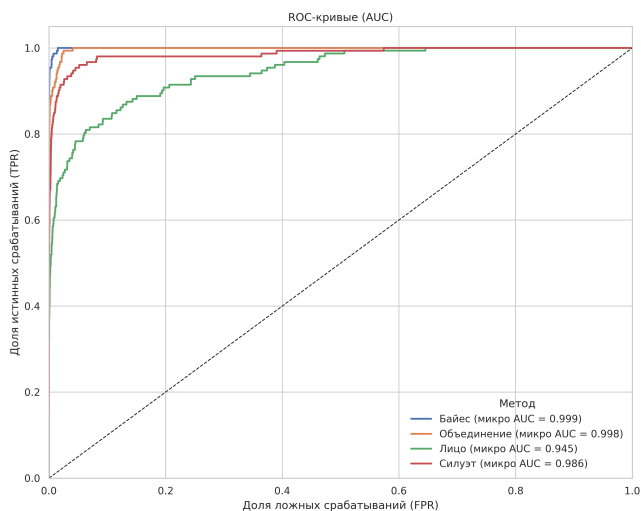


Рис. 3. ROC-кривые (micro-AUC) для всех методов

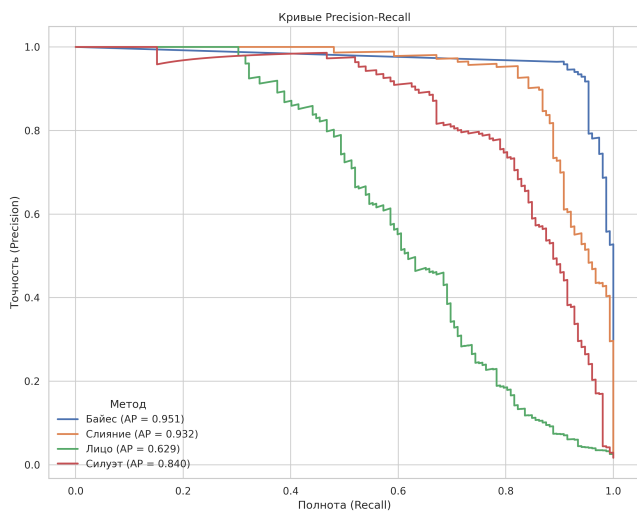


Рис. 4. Кривые «точность — полнота» (micro-AP) для всех методов

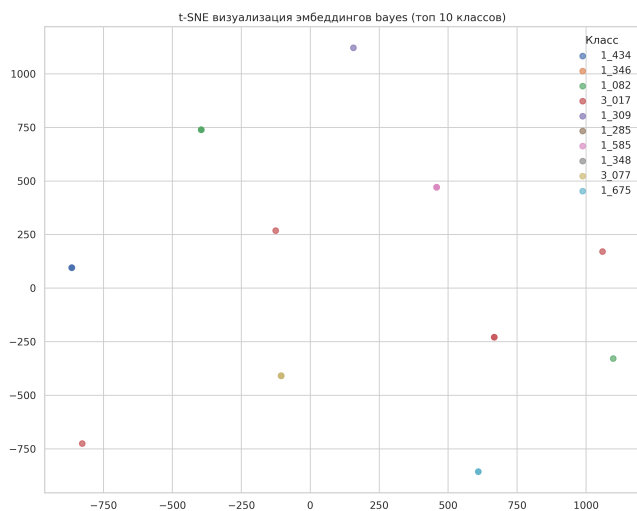


Рис. 5. t-SNE-визуализация векторов байесовского метода для топ-10 классов

5. Выводы и перспективы

В статье предложена универсальная байесовская схема объединения признаков лица и силуэта для задачи реидентификации личности в системах видеонаблюдения. В основе метода лежит вычисление апостериорных вероятностей по правилу Байеса, где каждый вектор моделируется нормальным распределением с параметрами, оценёнными по опорным данным. Такой подход позволяет адаптивно учитывать надёжность каждой модальности: при повреждённых или отсутствующих фрагментах изображения система автоматически снижает вклад шумного признака и фокусируется на более информативном.

Экспериментальная проверка на датасете CUNK03 показала, что предложенная схема превосходит как классическую конкатенацию векторов, так и одномодальные методы. Байесовская модель продемонстрировала прирост точности идентификации более чем на 3% по сравнению с линейным объединением и обеспечила micro-AUC 0,9993 и micro-AP 0,951. Кроме того, оцениваемая апостериорная вероятность позволяет в реальном времени формировать порог уверенности и, при необходимости, запрашивать дополнительный образ или выдавать несколько наиболее вероятных кандидатов.

Одним из важных преимуществ предложенного байесовского подхода является его низкая вычислительная сложность и хорошая масштабируемость. В отличие от сложных методов с обучаемыми коэффициентами или методов на основе глубоких сетей с перекрестным вниманием, предложенная модель требует лишь расчета евклидовых расстояний до центроидов классов в низкоразмерных пространствах признаков, что реализуется крайне эффективно даже для больших баз идентифицируемых личностей. При увеличении количества личностей (классов) вычислительная нагрузка растёт линейно, обеспечивая лёгкое масштабирование решения. Кроме того, подход позволяет легко добавлять новые модальности и адаптировать систему под изменения условий съёмки без необходимости переобучения сложных глубоких

архитектур.

Одним из ограничений предложенного подхода является использование предположения о нормальности распределения признаков в каждой модальности, что может не всегда идеально соответствовать реальным данным, особенно в случае наличия выбросов или нетипичных наблюдений. Для преодоления этого ограничения рекомендуется в дальнейшем рассмотреть применение более гибких статистических моделей, таких как смеси гауссовых распределений или непараметрические методы оценки плотностей. Также можно обсудить устойчивость модели к шумам и искажениям в данных и возможности её улучшения через использование робастных статистических методов или предварительной фильтрации входных данных.

Дальнейшие направления работы связаны с расширением модели несколькими биометрическими модальностями и внедрением более гибких ковариационных структур, позволяющих учитывать попарную зависимость признаков. Также перспективно исследовать онлайн-обучение параметров распределений для поддержки динамических баз персоналий и адаптацию порога уверенности под изменяющиеся условия съёмки.

Литература

1. АМИРГАЛИЕВА З., САДЫКОВА А., КЕНШИМОВ К. *Разработка модифицированного алгоритма Виолы – Джонса для распознавания лиц* // Известия НАН РК. Серия физико-математическая. – 2022. – №4. – С. 258–267.
2. ИГНАТЬЕВА С.А., БОГУШ Р.П. *Алгоритм реидентификации людей по изображениям систем видеонаблюдения с использованием нейросетевого составного дескриптора* // Системный анализ и прикладная информатика. – 2024. – №1 – С. 54–65.
3. КОСУЛИН К.Э., КАРПОВ А.А. *Методы аудиовизуального распознавания людей в масках* // Научно-технический вестник информационных технологий, механики и оптики. – 2022. – Т. 22, № 3. – С. 415–432.

4. МУБИНОВА Э.С. *Эволюция технологий распознавания лиц и их влияние на общественное восприятие безопасности* // Вестник науки. – 2024. – №9(78). – С. 168–171.
5. САВЧЕНКО А.В. *Метод максимально правдоподобных рас-
согласований в задаче распознавания изображений на ос-
нове глубоких нейронных сетей* // Компьютерная оптика. – 2017. – №41(5). – С. 819–828.
6. СОКОЛОВА А.Д., САВЧЕНКО А.В., НИКОЛЕНКО С.И. *Поиск редких данных в задаче распознавания лиц на изоб-
ражениях* // Компьютерная оптика. – 2022. – №46(2). – С. 318–327.
7. ТЕЛЕНЬКОВ А. *Система распознавания силуэтов при-
меняется МВД в пяти регионах РФ* // ТАСС (ЦИПР). – 24.06.2021.
8. ХАЧУМОВ М.В., НГУЕН Т.З. *Распознавание лиц по фото-
графиям на основе инвариантных моментов* // Современ-
ные проблемы науки и образования. – 2015. – №2 – С. 10.
9. САВЧУК В.П. *Байесовские методы статистического оце-
нивания: надежность технических объектов.* – М.: Наука,
Гл. ред. физ.-мат. лит., 1989.
10. DENG J., GUO J., XUE N. et al. *ArcFace: Additive Angular
Margin Loss for Deep Face Recognition* // Proc. CVPR. – 2019. – P. 4690–4699.
11. HERMANS A., BEYER L., LEIBE B. *In Defense of the Triplet
Loss for Person Re-Identification* // arXiv preprint. – 2017. – arXiv:1703.07737.
12. IRFAN B., GARCIA ORTIZ M., LYUBOVA N. et al. *Multi-
modal Open World User Identification* // ACM Trans. Human-
Robot Interaction. – 2022. – Vol. 11(1). – P. 1–50.
13. KOO J.H., CHO S.W., BAEK N.R. et al. *CNN-based
multimodal human recognition in surveillance environments* // Sensors. – 2018. – Vol. 18(9). – P. 3040.
14. LIAO Z., DI D., HAO J. et al. *MMM-GCN: Multi-Level Multi-
Modal Graph Convolution Network for Video-Based Person
Identification* // LNCS. – 2023. – Vol. 13834. – P. 3–15.

15. LI W., ZHAO R., XIAO T. et al. *DeepReID: Deep Filter Pairing Neural Network for Person Re-identification* // Proc. CVPR. – 2014. – P. 152–159.
16. LI W., ZHU X., GONG S. *Harmonious Attention Network for Person Re-Identification* // Proc. CVPR. – 2018. – P. 2285–2294.
17. SCHROFF F., KALENICHENKO D., PHILBIN J. *FaceNet: A unified embedding for face recognition and clustering* // Proc. CVPR. – 2015. – P. 815–823.
18. SHOUKRY N., ABD EL GHANY M.A., SALEM M.A.-M. *Multi-Modal Long-Term Person Re-Identification Using Physical Soft Bio-Metrics and Body Figure* // Applied Sciences. – 2022. – Vol. 12(6). – P. 2835.
19. VEZZANI R., BALTIERI D., CUCCHIARA R. *People re-identification in surveillance and forensics: A survey* // ACM Computing Surveys. – 2013. – Vol. 46(2). – P. 1–37.
20. ZHANG M., LIU R., ABE N. *Face Helps Person Re-Identification: Multi-modality Person Re-ID Based on Vision-Language Models* // Proc. IJCB. – 2024.
21. CHEN S., LIU Y., GAO X. et al. *MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices* // arXiv preprint. – 2018. – arXiv:1804.07573.
22. DOSOVITSKIY A. et al. *An image is worth 16x16 words: Transformers for image recognition at scale* // arXiv preprint. – 2020. – arXiv:2010.11929.

BAYESIAN MODEL FOR FEATURE REPRESENTATION FUSION IN PERSON RE-IDENTIFICATION

Konstantin Rusakov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, research fellow (rusakov@ipu.ru).

Abstract: This paper addresses the problem of person re-identification in surveillance systems based on the analysis and integration of heterogeneous descriptive features. Unlike traditional approaches relying on a single biometric modality, a general-purpose probabilistic method is proposed to combine features derived from various sources. The method is based on Bayesian inference and models each feature as a random variable following a multivariate normal distribution. Parameters of the distributions are estimated from the training data. Decision making is performed by maximizing the posterior probability of identity given the available evidence. To extract features, transformer-based architectures with attention mechanisms are used, ensuring robustness to visual noise and viewpoint variation. The proposed model is compared with classical approaches based on linear combination of feature vectors. Experiments conducted on an open re-identification dataset demonstrate that the Bayesian scheme improves recognition accuracy and remains effective in cases of partial information loss. Moreover, the method provides a quantitative confidence score associated with each decision, making it particularly suitable for deployment in safety-critical environments such as automated video surveillance and access control systems.

Keywords: person re-identification, feature fusion, Bayesian model, computer vision, transformer architectures, video surveillance.

УДК 004.93

ББК 32.81

*Статья представлена к публикации
членом редакционной коллегии Н.А. Коргиным.*

Поступила в редакцию 22.04.2025.

Дата опубликования 31.05.2025.