

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М. В. ЛОМОНОСОВА



На правах рукописи

Чернышов Виктор Геннадьевич

БИОМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО ИЗОБРАЖЕНИЮ ВНЕШНЕЙ
СТОРОНЫ ЛАДОНИ НА БАЗЕ МОБИЛЬНОГО УСТРОЙСТВА

05.13.17 — Теоретические основы информатики

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель —
доктор технических наук,
профессор Л. М. Местецкий

Москва

2019

Содержание

1	Введение	5
1.1	Актуальность темы	5
1.2	Цель исследования	7
1.3	Методы исследования	8
1.4	Предлагаемый подход к решению	8
1.5	Научные результаты, выносимые на защиту	9
1.6	Научная значимость	10
1.7	Практическая значимость	11
1.8	Достоверность результатов	11
1.9	Апробация работы	11
1.10	Обоснование специальности	12
1.11	Структура и объём	13
2	Обзор научной литературы	14
2.1	Обзор методов биометрической идентификации	14
2.2	Связь разрешения изображения ладони и извлекаемых текстурных признаков	17
2.3	Методы идентификации по ладони	19
2.3.1	По рисунку ладони	19
2.3.2	По форме ладони	22
2.3.3	По кожным складкам пальцев	24
2.3.4	Мультимодальные	25
2.4	Мобильные системы идентификации по ладони	27
2.4.1	По рисунку ладони	27
2.4.2	По форме ладони	28
2.4.3	По кожным складкам пальцев	29
2.5	Выводы к главе 2	30
3	Метод идентификации по ладони	33
3.1	Машинное обучение в задаче идентификации по ладони	33
3.1.1	Основные понятия	33
3.1.2	Модель алгоритмов и метод обучения	34
3.1.3	Функционал качества	35
3.1.4	Признаковое описание объектов	36

3.1.5	Цикл исследования	36
3.2	Задачи детектирования и распознавания	38
3.3	Корректность ладони	39
3.4	Метод детектирования	40
3.4.1	Сегментация	41
3.4.2	Получение геометрической информации о ладони	50
3.4.3	Семантическая разметка ладони	56
3.4.4	Определение расфокусированных кадров	61
3.4.5	Вычисление зёрен	64
3.4.6	Выводы к разделу 3.4	66
3.5	Метод распознавания	67
3.5.1	Сегментация	67
3.5.2	Генерация признаков формы	68
3.5.3	Локализация области кожных складок на пальце	74
3.5.4	Извлечение текстурных признаков на базе фильтра Габора	77
3.5.5	Сравнение на основе Габор-признаков	82
3.5.6	Сравнение на основе SIFT-признаков	82
3.5.7	Сравнение на основе текстурных признаков	84
3.5.8	Выводы к разделу 3.5	85
3.6	Выводы к главе 3	86
4	Экспериментальное исследование методов детектирования и распознавания	87
4.1	Оборудование	87
4.2	Базы изображений	87
4.2.1	База-1	88
4.2.2	База-2	89
4.2.3	База HGC	91
4.2.4	База PolyU HD	92
4.3	Вычислительные эксперименты	94
4.3.1	Тестирование метода детектирования	96
4.3.2	Тестирование метода распознавания	99
4.4	Выводы к главе 4	106

5	Мобильная система биометрической идентификации	107
5.1	Общая характеристика системы	107
5.2	Мобильное приложение	108
5.2.1	Принципы работы	108
5.2.2	Главный экран	114
5.2.3	Настройки	116
5.3	Сервер	118
5.4	Веб-интерфейс	121
5.5	Особенности разработки кода	124
5.6	Выводы к главе 5	125
6	Заключение	126
	Список литературы	128

1 Введение

1.1 Актуальность темы

Технологии биометрической идентификации личности широко эксплуатируются в системах контроля доступа и учета посещаемости. Разработано и опробовано множество подходов, использующих различные идеи, признаки, измерения. Основными методами, применяющими статические биометрические характеристики человека, являются: идентификация по папиллярному рисунку на пальцах, радужной оболочке глаза, геометрии лица, сетчатке глаза, рисунку вен ладони, геометрии и текстуре ладони (как внешней, так и внутренней её стороны). Также существует семейство методов, использующих динамические характеристики: идентификация по голосу, динамике рукописного почерка, сердечному ритму, походке. Далее будем называть *модальностями* категории биометрических характеристик человека.

Спрос на системы биометрической идентификации уверенно растёт на 10-15% в год на протяжении последнего десятилетия. В первую очередь их внедряют организации, которые заинтересованы в повышении уровня безопасности и улучшении имеющихся бизнес-процессов. Очень часто препятствием на пути внедрения подобных систем является их цена, которая на данный момент слишком высока для малого бизнеса и бюджетных учреждений с ограниченным финансированием. Также немаловажным сдерживающим фактором являются издержки, связанные с созданием и поддержанием необходимой инфраструктуры.

Каждая модальность обладает своими достоинствами и недостатками, определяющими сферы её применения. В частности, весьма перспективным направлением представляется идентификация человека по изображению внешней (тыльной) стороны его ладони.

Простота и удобство. Для идентификации достаточно расположить ладонь перед камерой, что предполагает минимальное взаимодействие со сканирующим оборудованием. Следует также отметить, что строение руки человека обеспечивает значительно более лёгкую демонстрацию ладони внешней стороной вверх (закрытое положение ладони) на горизонтальной плоскости, чем в противоположном случае (открытое положение ладони).

Социальная приемлемость. В отличие от идентификации по отпечатку пальца, как правило ассоциирующейся с криминалистикой, процедура идентификации по внешней стороне ладони является эмоционально нейтральной и не вызывает отторжения у людей. В то же время, демонстрация внутренней стороны ладони незнакомому человеку не является приемлемой в некоторых культурных традициях и религиозных верованиях.

Помимо этого, можно выделить несколько технических особенностей данной модальности.

1. Для неё не требуются специализированные сенсоры, нужна лишь цифровая камера. В то же время для систем, работающих с другими модальностями, требуется особое, зачастую дорогостоящее, оборудование: устройства сканирования радужной оболочки глаза, сканеры отпечатков пальцев, инфракрасные камеры для считывания рисунка вен и т. п.
2. Для получения изображения тыльной стороны ладони вполне подходят камеры мобильных телефонов и планшетов, которые на данный момент обладают достаточным разрешением (1280×720 пикселей и более), — и потому являются подходящими сенсорами в возникающей задаче биометрической идентификации.

В силу широкой распространенности мобильных устройств с качественными камерами можно считать, что идентификация по изображению ладони не требует специального оборудования.

Таким образом, выбор изображения внешней стороны ладони в качестве модальности создаёт предпосылки для создания на основе мобильного устройства интуитивно понятной и социально приемлемой системы биометрической идентификации, которая при этом технически доступна и недорога.

На сферу применения мобильной системы, использующей ладонь для распознавания, критическим образом влияет качество получаемых снимков. Оно варьируется в широких пределах — это определяется многообразием характеристик гаджетов и разнообразием условий съёмки. Довольно часто участки снимков, которые подвергаются анализу (в данной работе, например, это области кожных складок пальцев внешней стороны ладони), обладают низкой информативностью — характерные особенности едва различимы. Вследствие чего, построенные по ним признаки не позволяют достичь высокоточной идентификации, требуемой, к примеру, при контроле доступа на режимных объектах или в задачах криминалистики.

Мобильная система идентификации по тыльной стороне ладони должна применяться в случаях, не требующих большой точности распознавания. Таких, как контроль посещаемости в школах и университетах, контроль доступа в библиотеках, поликлиниках и прочих подобных учреждениях. Из этого следует, что база биометрических данных такого рода не представляет интереса для злоумышленников, и утечка информации из неё не способна нанести серьёзный вред. Затраты на хранение базы биометрических эталонов в этом случае невелики — нет необходимости обеспечивать высокий уровень их безопасности.

Тем не менее, вышесказанное не исключает возможности использования тыльной стороны ладони в качестве дополнительной модальности в системах с повышенными требованиями к точности распознавания.

1.2 Цель исследования

Целью исследования является создание метода идентификации человека по изображениям внешней стороны ладони, который может быть использован в системе распознавания на базе мобильного устройства. Неотъемлемой частью работы является реализация соответствующего программного комплекса, обосновывающего эффективность метода и его частей.

Выбор в качестве модальности **внешней** стороны ладони, а также желаемые показатели точности и скорости работы метода идентификации во многом продиктованы требованиями к системе распознавания в целом:

точность — небольшой уровень ошибок (т.е. приемлемый для предметной области, в которой планируется внедрение);

скорость — распознавание должно осуществляться в режиме реального времени;

доступность — система должна состоять из распространённых и недорогих компонентов;

социальная примлемость — система не должна вызывать негативную реакцию пользователей в силу социальных стереотипов или религиозных установок;

удобство — процесс идентификации должен быть интуитивно понятен для конечного пользователя и не вводить дополнительных ограничений (например, не требовать выравнивания ладони перед камерой), эксплуатация не должна быть сложна для владельца (например, минусом может являться сложный процесс разворачивания системы на новом месте);

гибкость — возможность работы в различных внешних условиях (например, при разном освещении и при съёмке «с руки»).

Представленная в работе мобильная система распознавания по внешней стороне ладони (по форме ладони и кожным складкам пальцев) разработана, в первую очередь, для идентификации в малых группах. Возможной областью её использования являются учреждения, где существует потребность в контроле посещаемости: вузы, школы, больницы, библиотеки.

На момент начала исследования (июнь 2013 года) в научной литературе (раздел 2) не было представлено систем идентификации по ладони в целом (и по внешней её стороне — в

частности) , одновременно удовлетворяющих всем вышеуказанным требованиям. Возможные причины этого кроются в сложности возникающих научно-технических задач, которые необходимо решить при создании такой системы. Разработке и проектированию системы биометрической идентификации по внешней стороне ладони сопутствует ряд неопределённых факторов, которые сводятся к нескольким принципиальным вопросам.

1. Насколько уникальными являются признаки формы ладони и рисунка кожных складок и в какой мере их достаточно для идентификации личности?
2. Существует ли возможность надёжного извлечения этих признаков из изображений, полученных с помощью общедоступных мобильных камер (без использования специального оборудования) при свободном расположении ладони во время съёмки?
3. Возможно ли решить задачу извлечения признаков и провести идентификацию на их основе в реальном времени при работе на мобильном устройстве в различных условиях применения?
4. Каким образом построить систему распознавания и организовать процесс идентификации, чтобы получившийся программный комплекс оказался практически востребованным?

Получение ответов на эти вопросы составляет основное содержание исследования, проведенного в данной работе.

1.3 Методы исследования

В исследовании использованы методы вычислительной геометрии, теории графов, теории анализа и распознавания изображений, компьютерной графики, анализа данных и машинного обучения. Для сбора данных и проведения вычислительных экспериментов использовались методы разработки многозвенных программных систем — в том числе, мобильных и веб-приложений.

1.4 Предлагаемый подход к решению

В работе предлагаются и обосновываются следующие принципы построения метода идентификации по изображению внешней стороны ладони и основанной на нём системы распознавания на базе мобильного устройства.

- Декомпозиция задачи идентификации на задачи детектирования и распознавания, их структурное распределение между клиентом и сервером:
 - задача клиента — детектирование, т. е. отбор подходящих для распознавания изображений ладони и последующая отправка их на сервер;
 - задача сервера — распознавание, т. е. вычисление признаков, сравнение с базой эталонов и выдача результата.
- Применяется клиент-серверная структура программного комплекса с мобильным устройством в качестве клиента.
- Для детектирования ладони и локализации областей кожных складок на пальцах используется непрерывное медиальное представление формы ладони.
- Формирование признакового описания ладони осуществляется с помощью двумерных фильтров Габора и SIFT-дескрипторов, которые применяются к областям изображения внешней стороны ладони, содержащим кожные складки вторых фаланг пальцев.
- Классификация изображений ладони происходит путём сравнения их признакового описания с базой эталонов — используется метод ближайшего соседа со специальной мерой сходства.

1.5 Научные результаты, выносимые на защиту

На защиту выносятся следующие результаты данной работы:

1. Метод биометрической идентификации личности по изображениям внешней стороны ладони со свободным положением руки в кадре, включающий детектирование ладони по признакам формы объекта и классификацию по текстурным признакам.
2. Метод детектирования ладони на основе непрерывного медиального представления изображения.
3. Метод формирования признакового описания ладони по изображениям кожных складок пальцев на базе двумерных фильтров Габора.
4. Мера сходства изображений ладони, основанная на признаках, полученных с помощью двумерных фильтров Габора и дескрипторов SIFT.

5. Обоснование работоспособности предложенных методов путём реализации программного комплекса на базе мобильного устройства. Экспериментально доказано, что рисунок кожных складок на внешней стороне средних фаланг пальцев является индивидуальным признаком личности человека, дающим возможность уверенно проводить идентификацию в малых группах.

Все результаты, выносимые на защиту, получены автором самостоятельно.

1.6 Научная значимость

Научный результат работы заключается в разработке метода идентификации по изображению внешней стороны ладони, а также в разработке подхода к созданию систем идентификации по ладони на базе представленного метода.

Ключевыми особенностями метода являются: разделение задачи идентификации на задачи детектирования и распознавания, использование непрерывного медиального представления формы ладони, применение методов машинного обучения.

Научная значимость предлагаемого решения определяется также универсальностью и гибкостью предложенного подхода:

- сфера применимости подхода не ограничена смартфонами и планшетами на базе ОС Android — аналогичное программное обеспечение может быть реализовано для другой подходящей аппаратной платформы: персонального компьютера с внешней камерой, мобильного устройства на базе иной ОС (например, iOS) или специализированного устройства с камерой и вычислительными ресурсами;
- разделение задачи идентификации на детектирование и распознавание не обязывает к физическому разнесению этих модулей между устройствами (клиентом и сервером) — возможно чисто структурное деление в рамках одного универсального устройства с нужными характеристиками;
- для построения информативного описания ладони в работе использованы области кожных складок пальцев на её тыльной стороне, однако в рамках предложенного подхода могут быть исследованы возможности построения системы, работающей с другими текстурными особенностями — например, с рисунком на внутренней стороне ладони.

1.7 Практическая значимость

Основной недостаток наиболее эффективных систем биометрической идентификации — необходимость использования специального оборудования (датчики, сканеры) и сопутствующей инфраструктуры. Как следствие, подобные системы дороги не только при покупке, но и при последующей эксплуатации, что делает их недоступными при небольшом финансировании.

Практическим результатом данной работы является программное обеспечение, решающее задачу идентификации по тыльной стороне ладони в малых группах, не выходя за рамки возможностей современных мобильных устройств. Иными словами, любой человек, имеющий смартфон или планшет на базе ОС Android, получает возможность на его основе за несколько минут внедрить для своего коллектива систему идентификации по форме и текстуре ладони.

1.8 Достоверность результатов

Достоверность результатов работы подтверждается корректностью постановок рассматриваемых задач исследования, применением математически обоснованных методов их решения, программной реализацией и вычислительными экспериментами, проведенными на реальных базах изображений человеческих ладоней. Реализованная в рамках исследования система биометрической идентификации по внешней стороне ладони прошла апробацию в качестве инструмента контроля посещаемости студентов и учёта посетителей на фестивале науки.

1.9 Апробация работы

Результаты исследования докладывались на следующих конференциях и семинарах:

1. 16-ая всероссийская конференция «Математические методы распознавания образов» ММРО-16, Казань, 2013;
2. 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding (OGRW 2014), Germany, Koblenz, 2014 [1];
3. 10th International Conference on Computer Vision Theory and Applications (VISAPP 2015), Germany, Berlin, 2015 [2];

4. 6-я научно-техническая конференция «Техническое зрение в системах управления-2015», Москва, 2015 [3];
5. 17-я всероссийская конференция «Математические методы распознавания образов» ММРО-17, Светлогорск, 2015 [4];
6. 26-я международная конференция по компьютерной графике и машинному зрению (ГрафиКон 2016), Нижний Новгород, 2016 [5];
7. 18-я всероссийская конференция «Математические методы распознавания образов» ММРО-18, Таганрог, 2017 [6];
8. семинар «Морфологический анализ данных» под руководством проф. Ю.П. Пытьева, Москва, 13 ноября 2014 г.;
9. совместный с Хуавэй международный научный семинар «Интеллектуальная обработка изображений и видео», Москва, 22 ноября 2017 г.;
10. XIII Московский Фестиваль науки NAUKA 0+, Москва, 12-14 ноября 2018 г.;
11. the 5th International Professor's Day on ICT Algorithm Design (ICTAD-2018), Russia, Moscow, November 28-29, 2018;
12. Zhejiang international workshop, Russia, Moscow, December 14, 2018.

Материалы исследования опубликованы в 9 печатных работах, из них: 2 статьи в рецензируемых журналах [7, 8], 4 статьи в сборниках трудов конференций [1, 2, 5, 9] и 3 тезиса докладов [3, 4, 6]. В изданиях, рекомендованных ВАК РФ, опубликовано 2 статьи [7, 8].

В рамках исследования разработана мобильная система идентификации личности по ладони, с помощью которой осуществлялся контроль посещаемости студентов двух групп ВМК МГУ. Также система применялась для учёта посетителей на XIII Московском фестивале науки Nauka 0+.

Описания отдельных результатов работы включены в отчёты по проектам РФФИ 14-01-00716, 16-57-45054, 17-01-00917.

1.10 Обоснование специальности

Данная работа по своей тематике и направленности полученных результатов соответствует следующим пунктам паспорта специальности 05.13.17 — «Теоретические основы информатики»:

5) Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

7) Разработка методов распознавания образов, фильтрации, распознавания и синтеза изображений, решающих правил. Моделирование формирования эмпирического знания.

1.11 Структура и объём

Работа состоит из введения, 4 глав, заключения и библиографии. Работа содержит 134 страницы основного текста, включая 60 иллюстраций и 12 таблиц. Перечень библиографических источников включает 71 наименование.

2 Обзор научной литературы

Изучение релевантных методов и анализ уже реализованных систем имеет для данной работы крайне важное значение, так как именно отсутствие готовых решений в рамках поставленной задачи (раздел 1.2) дало старт исследованию. Ниже приведена структура главы. В разделе 2.1 сделан общий обзор популярных методов биометрической идентификации. Следует за ним раздел 2.2, в котором содержатся важные замечания касательно связи разрешения изображения ладони и извлекаемых оттуда текстурных признаков. Раздел 2.3 посвящён методам, использующим минимум одну из модальностей из числа входящих в следующий список: форма ладони, рисунок ладони, кожные складки пальцев. Продолжает изложение раздел 2.4, в котором разбираются известные мобильные системы идентификации, базирующиеся на вышеприведенных модальностях. Завершает главу раздел 2.5 с выводами.

2.1 Обзор методов биометрической идентификации

Существует множество методов биометрической идентификации, обладающих своими достоинствами и недостатками [10, 11]. Краткое сравнение наиболее часто используемых в задаче идентификации биометрических признаков (модальностей) приведено в таблице 1, составленной на основе таблицы из работы [11] и собственного анализа изученной литературы. Значения столбцов следующие:

- точность — насколько уверенно человек может быть идентифицирован по этому признаку;
- универсальность — наличие данного признака у всех людей, использующих систему;
- уникальность — насколько данный признак отличается у разных людей;
- стабильность — насколько сильно изменяется данный признак с течением времени;
- бесконтактность — требуется ли соприкосновение со сканером, с сенсором или рабочей поверхностью; если указано «Варьируется», то встречаются различные варианты получения данных;
- стоимость — стоимость готовой биометрической системы, использующей данный признак;
- тип — оборудование, которое требуется, чтобы сконструировать систему, использующую данный признак.

Цель таблицы — дать некое качественное представление о той части предметной области, которая имеет отношение к данной работе. Следует отметить, что значение уникальности биометрического признака может зависеть от характеристик используемого оборудования — например, рисунок ладони и кожные складки пальцев имеют высокие показатели уникальности только при съёмке в высоком разрешении, а для распознавания по лицу наиболее эффективно одновременное применение нескольких камер для построения его 3d-модели. Считаем, что используется оптимальное оборудование.

Распознавание по рисунку радужной оболочки глаза. Характеризуется высокой точностью (близкой к 100% [12]), сравнение выполняется практически мгновенно. Радужная оболочка мало изменяется в течение жизни, распознавание не подвержено влиянию внешних факторов (съёмка производится с близкого расстояния в инфракрасном диапазоне). К недостаткам таких методов можно отнести высокую стоимость оборудования, трудность настройки сканера для использования несколькими людьми различного роста. Подобную систему можно обмануть, используя распечатанную качественную фотографию радужной оболочки. Во избежание этого могут применяться динамические тесты — например, отслеживание микродвижений зрачка. Изменчивость признаков в течение жизни небольшая, однако некоторые операции на глазах могут уменьшить вероятность корректного распознавания. Требуется специализированное оборудование.

Распознавание по лицу. Обладает сравнительно высокой точностью в идеальных условиях (более 98% [13]), сравнение производится быстро. Для некоторых систем не требуется кооперация с пользователем. Из недостатков можно отметить чувствительность к освещению, повороту головы, очкам, волосам и другим объектам, закрывающим лицо. Кроме того, распознавание чувствительно даже к изменению выражения лица. Обман системы возможен с помощью статических макетов, динамические тесты могут этому воспрепятствовать. Изменчивость признаков на протяжении жизни — от умеренной до высокой. Для получения данных требуется обычная цифровая камера. В системах, которые строят 3d-модель лица, их несколько.

Распознавание по голосу. Имеет сравнительно высокую точность (EER порядка 2% [14]), быстрое сравнение, устойчивость к внешним факторам (в т. ч. к шуму). Однако соответствующие методы не подходят для людей, имеющих проблемы с речевым аппаратом или немых. Также голос может быть искажен в результате болезни или травмы, что затрудняет распо-

Биометрический признак	Точность распознавания	Характеристики признака			Оборудование		
		Универсальность	Уникальность	Стабильность	Бесконтактность	Стоимость	Тип
Радужная оболочка глаза	Высокая	Высокая	Высокая	Высокая	Да	Высокая	Специализированное устройство
Лицо	Высокая	Высокая	Высокая	Средняя	Да	Варьируется	Камеры (1 и более)
Отпечаток пальца	Высокая	Средняя	Высокая	Высокая	Нет	Низкая	Специализированное устройство
Рисунок вен	Высокая	Высокая	Высокая	Высокая	Варьируется	Высокая	Специализированное устройство
Голос	Высокая	Средняя	Высокая	Средняя	Да	Низкая	Микрофон
Форма ладони	Средняя	Высокая	Средняя	Средняя	Варьируется	Низкая	Камера или сканер
Рисунок ладони	Высокая	Высокая	Высокая	Высокая	Варьируется	Низкая	Камера или сканер
Кожные складки пальцев	Высокая	Высокая	Высокая	Средняя	Да	Низкая	Камера или сканер

Таблица 1: Сравнение биометрических признаков.

знание. Аудиозапись голоса позволяет обмануть систему, одним из способов борьбы может служить постоянная смена произносимой контрольной фразы. Голос изменяется с возрастом, изменчивость может быть от умеренной до высокой. Требуется микрофон для получения данных. Наиболее совершенные методы устойчивы к шумам при записи.

Распознавание по рисунку вен. Демонстрирует высокую точность ($EER = 0.17\%$ [15]), сравнение осуществляется быстро. Устойчиво к внешним факторам. Существенное преимущество подобных методов в том, что из-за использования инфракрасного диапазона системе достаточно сложно обмануть — для распознавания необходима ладонь живого человека. Изменчивость признаков в течение жизни низкая. Требуется специализированное устройство с инфракрасной камерой.

Распознавание по отпечатку пальца. Является самым распространенным подходом к идентификации на данный момент. Для подобных методов характерны высокая точность ($EER = 0.67\%$ [16]), быстрое сравнение и независимость от внешних факторов. Обман такой системы затруднителен, однако возможен — например, через предъявление «универсального» отпечатка, составленного из наиболее часто встречающихся папиллярных узоров людей. Изменчивость низкая, но занятие некоторыми видами ручного труда может изменить отпечатки. Требуется специализированный сканер для получения данных.

2.2 Связь разрешения изображения ладони и извлекаемых текстурных признаков

Перед тем, как переходить к методам идентификации по изображению ладони, будет разумным пояснить некоторые моменты, связанные с информативностью снимков ладони в видимом диапазоне и разрешением матрицы цифровой камеры.

На интуитивном уровне достаточно очевидно, что изображения ладони, снятые одной и той же камерой с разных дистанций, отличаются по информативности — на более «удалённом» снимке мелкие детали неразличимы, и извлечь из них адекватное признаковое описание едва ли получится.

В растровой графике для обозначения *разрешающей способности* различных процессов преобразования изображений (сканирование, печать, растеризация и т. п.) популярен термин dpi (англ. dots per inch) — эта величина говорит о количестве точек на единицу длины: например, 300 dpi означает 300 точек на один дюйм). Для справки, 1 дюйм (1") равен 2.54 см.

Для устройств ввода, к которым относятся сканеры и цифровые камеры, для обозначения разрешающей способности вместо dpi часто используется аналогичный термин ppi (англ. pixels per inch) — количество пикселей на единицу длины.

Информация из буклета от компании-изготовителя перчаток (рис. 1) позволяет примерно оценить линейные параметры ограничивающего прямоугольника для ладони взрослого человека. Согласно наблюдениям, при оптимальном расположении ладони целиком внутри области съёмки (с некоторым запасом от краёв) минимальная сторона L_1 ограничивающего прямоугольника приблизительно в 2 раза превышает обозначенное на рисунке расстояние X («линейный» аналог обхвата ладони). Тем самым, X принимает значения из отрезка $[2.5'', 5'']$, а значения L_1 ограничены $[5'', 10'']$.

Не составляет труда определить фактическую разрешающую способность изображений ладони, входящих в базы (раздел 4.2), использованные автором при тестировании алгоритма распознавания по внешней стороне ладони (строки 2–4 в табл. 2).

Легко увидеть, что даже при наиболее удачных условиях съёмки (высокое разрешение камеры, физически маленькая ладонь, оптимальное расположение в кадре) фактический dpi изображения ладони остаётся достаточно низким — колеблется в пределах $96 \dots 240$ dpi.

Разрешение W (пикс.) x H (пикс.)	Пропорции изображения	мин. dpi (при $X=10''$)	макс. dpi (при $X=5''$)
640 x 480	4 : 3	48	96
1280 x 960	4 : 3	96	192
1600 x 1200	4 : 3	120	240
1920 x 1080	16 : 9	108	216

Таблица 2: Фактическое dpi изображений ладони при различных разрешениях камеры.

В исследовании [17] было показано, что для работы с текстурными особенностями внутренней стороны ладони уровня папиллярных линий необходима разрешающая способность не менее 400 dpi. Современные сканеры отпечатков пальцев, также работающие с папиллярным узором, имеют разрешающую способность от 500 dpi.

В то же время, для извлечения признаков из более крупных текстурных особенностей (уровня линий ладони и морщин на ней) зачастую достаточно меньшей (от 100 до 200 dpi) разрешающей способности [18]. Далее в работе в контексте текстурных особенностей ладони и её частей будем называть разрешающую способность до 200 dpi *низкой*, от 200 до 400 dpi — *средней*, от 400 dpi — *высокой*.

Вывод. В методах идентификации по ладони или её частям, использующих изображения в видимом диапазоне, разрешение снимков и механика их получения имеют ключевое значение, так как этим определяется разрешающая способность и, как следствие, характер извлекаемых текстурных признаков.

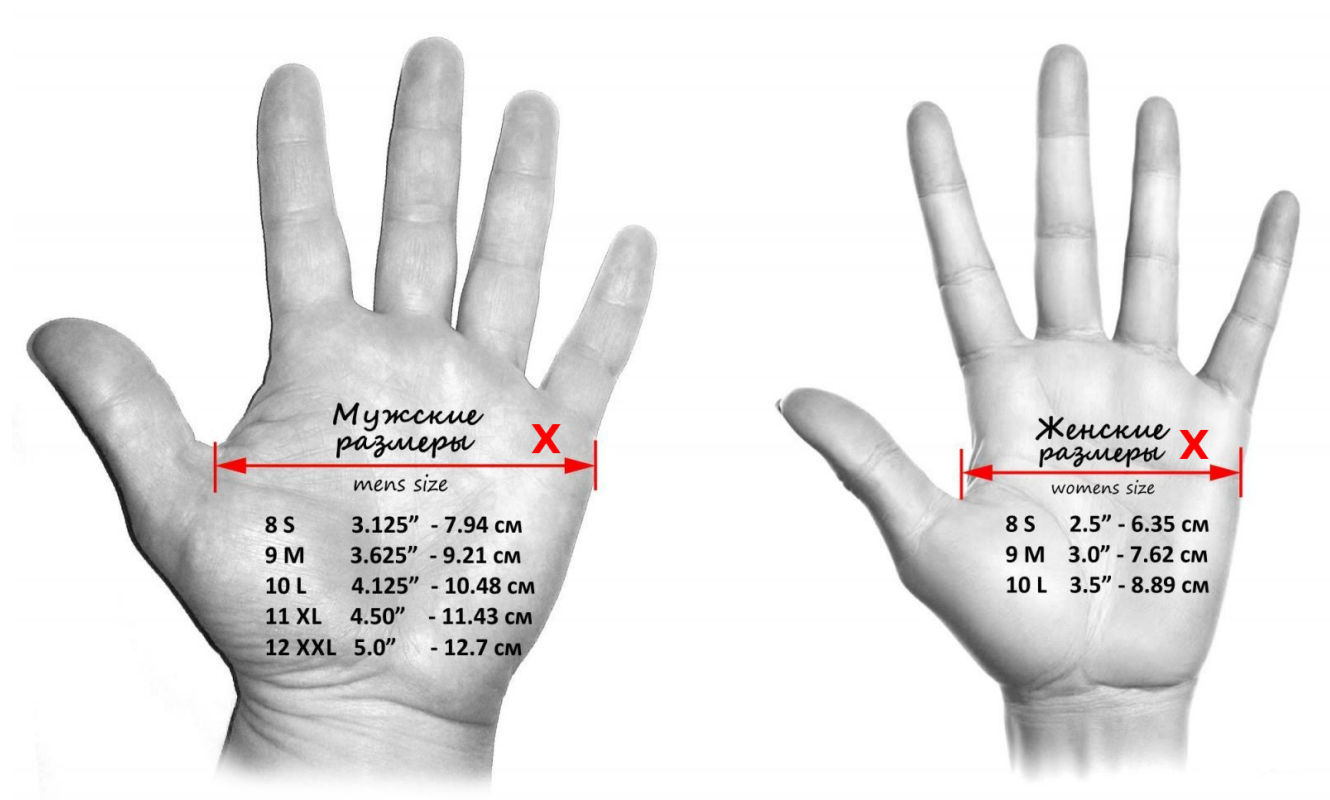


Рис. 1: Сводка размеров мужских и женских перчаток от производителя.

2.3 Методы идентификации по ладони

Помимо перечисленных в разделе 2.1 подходов к биометрической идентификации, существует ряд методов распознавания по ладони или её частям на основе изображений в видимом диапазоне. Рассмотрим их более подробно.

2.3.1 По рисунку ладони

Рисунком ладони называется узор основных линий ладони, а также морщин на внутренней стороне ладони. Изучение трудов, посвященных биометрической идентификации по рисунку внутренней стороны ладони, является важным этапом данного исследования, так как некоторые методы и идеи вполне применимы к распознаванию по кожным складкам пальцев на внешней стороне ладони.

Многие работы, исследующие распознавание по рисунку ладони, для тестирования используют представленную в [19] базу данных PolyU Palmprint [20]. База включает в себя изображения внутренней стороны ладони 193 людей с разрешением 1600×1200 пикселей, полученные в контролируемых условиях (равномерное искусственное освещение, чёрный фон, специальные упоры для фиксации ладони) с использованием оригинального устройства захвата изображений (схема приведена на рис. 2). Сбор изображений осуществлялся в две сессии, временной промежуток между ними составил 2 месяца (настройки устройства захвата во время второй сессии отличались, чтобы имитировать захват с помощью двух разных устройств). Во время каждой сессии собиралось по 10 изображений правой и левой ладоней каждого человека. Таким образом, каждый человек предоставил 40 изображений, а всего база содержит 7752 изображения 386 ладоней.

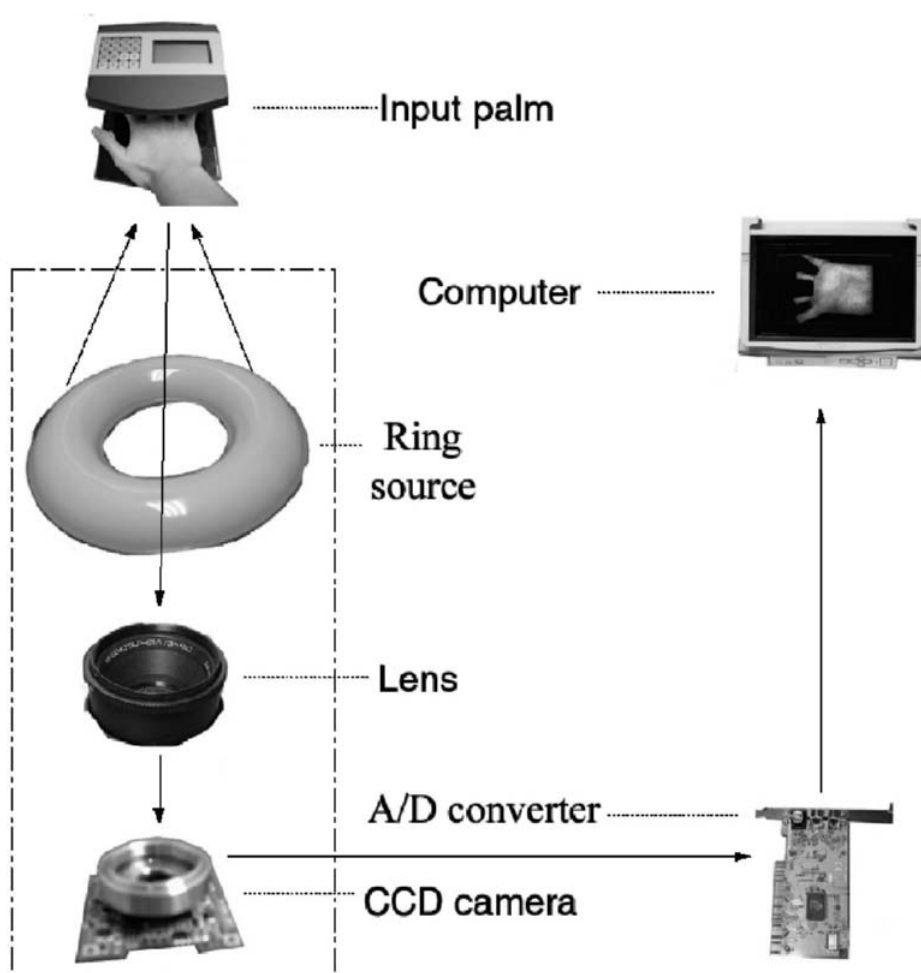


Рис. 2: Устройство для сбора изображений рисунка ладони, использованных для формирования базы PolyU Palmprint [20]. Здесь Ring source — лампа тороидальной формы, A/D converter — аналого-цифровой преобразователь, CCD Camera — цифровая камера с ПЗС-матрицей.

В системе используются направляющие для фиксации ладони внутри устройства. Кроме того, смонтирован круговой источник света для создания достаточного и равномерного освещения. В сочетании с тем, что съёмка производится на чёрном фоне, задача выделения области интереса существенно упрощается. А именно, достаточно выполнить простую пороговую бинаризацию, чтобы получить силуэт ладони. Затем выполняется поиск впадин между пальцами (между указательным и средним, между безымянным и мизинцем), чтобы по ним уже построить область интереса на ладони. Благодаря направляющим расположение пальцев на всех снимках одинаково, и поиск впадин не составляет труда.

Стоит заметить, что все изученные автором работы, использующие для тестирования базу PolyU Palmprint, применяют аналогичные несложные алгоритмы выделения области интереса.

Методы, основанные на векторе ориентации линий ладони в локальной окрестности. В работе [21] предложен метод кодирования особенностей линий ладони (как главных линий, так и морщин), названный Binary Orientation co-Occurrence Vector, или BOCV. Он является дальнейшим развитием зарекомендовавших себя методов, базирующихся на получении вектора ориентации в локальной окрестности, таких как: Palm Code [19], Fusion Code [22] и Competitive Code (CompCode) [23]. В основе алгоритма шесть двумерных фильтров Габора с разными ориентациями. Для получения значения вектора признаков в точке изображения объединяются бинаризованные по порогу результаты их применения. Благодаря использованию нескольких фильтров метод устойчив к небольшим поворотам и позволяет лучше извлекать локальные особенности в случае наличия в локальной окрестности двух и более пересекающихся линий ладони. Для тестирования использована база PolyU Palmprint [20]. Достигнуты значения $FAR = 3.3 * 10^{-6}\%$, $FRR = 0.2525\%$, $EER = 0.0189\%$ для задачи верификации, что превосходит показатели упомянутых «классических» методов.

В 2012 году в труде [24] предложено усовершенствование BOCV, названное E-BOCV. Вводится понятие «слабых» (fragile) битов, их позиции определяются как позиции точек, имеющих отклик фильтра меньше порога. При вычислении расстояния Хэмминга между BOCV-матрицами сравниваемых изображений слабые биты не учитываются. В плане точности E-BOCV незначительно улучшает результаты оригинального метода.

Метод, основанный на выделении основных линий. В статье [25] для выделения линий из рисунка ладони используется конечное преобразование Радона. По утверждению авторов, это позволяет извлекать основные линии быстро и эффективно, даже если изображение

содержит большое количество длинных и ярко выраженных морщин. В качестве признаков использовались бинарные изображения обнаруженных линий. Сравнение производилось с помощью метода пиксель-с-площадью (каждый пиксель первого изображения сравнивался с пикселями окрестности соответствующего пикселя второго изображения и наоборот). Для тестирования использована база PolyU Palmprint [20]. Достигнутые показатели FAR и FRR, равные 0.49% и 0.565% соответственно, подтверждают информативность основных линий в качестве признака. По утверждению авторов, полученный метод распознавания устойчив к небольшим сдвигам и поворотам ладони.

2.3.2 По форме ладони

Из рассмотрения исключены методы, работающие с 3d-снимками, так как соответствующие системы очевидно не удовлетворяют сформулированному ранее требованию доступности (раздел 1.2) из-за наличия 3d-сенсоров.

Методы на основе анализа контура. В работе [26] идентификация проводится по контуру ладони на бинарном изображении. Для получения признаков используются полиномы Цернике до 70-го порядка, авторами предложен собственный алгоритм вычисления таких полиномов. Для отбора признаков использован PCA (Principal Component Analysis), для сравнения применяется Евклидова метрика. Изображения ладони получены контактным способом с помощью камеры и прозрачного стола с подсветкой, за счет чего минимизировано влияние внешних условий на качество изображений. Упоры для позиционирования руки не используются, поэтому от пользователя требуется самостоятельно расставлять пальцы. Для сегментации применена бинаризация с фиксированным порогом. На базе из 400 изображений 40 людей достигнуто значение $EER = 2\%$.

В работе [27] наряду с прочими признаками формы также используется контур ладони. Изображения получают с помощью камеры и светодиодной подсветки, процедура бесконтактная. Контур, представляющий из себя простой многоугольник, получается путём применения детектора границ к результату пороговой бинаризации. Вычисляются несколько категорий признаков: основные характеристики ладони и пальцев (длина и ширина пальцев, ширина ладони, кривизна пальцев), углы (между парами соседних сегментов вдоль границы ладони), некоторые расстояния (между ключевыми точками и точками границы). Для сравнения векторов признаков применяется взвешенное расстояние Евклида. На базе из 4000 изображений 200 ладоней достигнуто значение $EER = 2.16\%$.

Методы на основе спектра. В статье [28] использован морфологический спектр в смысле Марагоса [29], который строится по бинаризованному изображению ладони. Для сравнения векторов признаков применяется евклидова метрика, в качестве классификатора выбран метод ближайшего соседа. База ладоней состоит из 400 изображений 40 людей, полученных с помощью сканера. Авторы отмечают высокую вычислительную сложность их способа вычисления спектра. В ходе тестирования достигнуто значение $EER = 2.85\%$.

В работе [30] рассмотрено формирование признакового описания по форме ладони не только на базе морфологического спектра, но и с помощью других видов спектра: медиального и дистанционного. В ходе проведенного тестирования различных метрик для сравнения векторов-признаков выбрана транспортная (EMD) метрика. Используемый классификатор kNN демонстрирует лучшие результаты при $k = 1$, т.е. оптимальным является метод ближайшего соседа. База ладоней состоит из 160 изображений 36 людей, полученных с помощью стационарной веб-камеры. Ладонь испытуемого в процессе съёмки прижата к тёмной горизонтальной поверхности непосредственно под камерой. Для морфологического, медиального и дистанционного спектров получено качество идентификации 93.7%, 94.5%, и 95.2% соответственно. Среднее время работы на изображении из базы разрешением 640×480 пикселей в системе с процессором Intel Core i5 2.67 ГГц составляет 2.548 с., 0.025 с. и 0.093 с. соответственно. Авторы делают вывод, что в прикладных задачах наиболее уместно использовать медиальный или дистанционный спектры, так как они могут вычисляться в режиме реального времени — например, при обработке видеопотока. Но из-за невысокого качества идентификации признаковое описание на базе спектра следует включать в мультимодальные системы наряду с другими признаками описания объекта.

Метод на основе гибких объектов. В работе [31] предложен метод распознавания на основе модели ладони в виде плоского графа, называемого скелетом. Скелет является непрерывной моделью ладони, с помощью его деформаций можно легко получить соответствующие деформации силуэта ладони. В базе хранятся скелеты для ладоней, снятых с расставленными пальцами без колец и браслетов. Далее выделяются пальцы и определяются точки их поворотов, после чего параметризуются допустимые деформации скелета: сдвиги и повороты всей ладони, повороты пальцев. При сравнении к эталонным моделям из базы применяются деформации так, чтобы привести пальцы в положение, минимизирующее расстояние до тестируемой модели по заданной метрике. В базе 22 человека и 97 изображений. В задаче верификации достигнуто значение $EER = 0.5\%$. В задаче идентификации метод показал се-

бя не лучшим образом, поэтому в данной работе предложено дополнительно использовать распознавание по голосу.

2.3.3 По кожным складкам пальцев

Методы, основанные на преобразовании Габора. Статья [32] является одной из самых цитируемых работ по распознаванию кожных складок пальцев внешней стороны ладони. Представлены программно-аппаратный комплекс для получения изображений суставов пальцев, включающий в себя устройство для съёмки с фиксацией пальца, и метод выделения области складок из полученного изображения. Устройство съёмки состоит из светодиодного источника света, упора для пальца (палец слегка согнут) и цифровой камеры (рис. 3).

Для выделения области кожных складок на пальце используется различие в направлении изгибов складок вокруг центра окна поиска. Извлечение признаков делается с помощью модификации ImCompCode&MagCode алгоритма Competitive Code [23], уже упомянутого выше при рассмотрении методов идентификации по рисунку ладони. Для каждого пикселя выполняется свёртка с несколькими двумерными фильтрами Габора разных ориентаций. Если нет доминирующего направления, то первый признак для данного пикселя получает некоторое специальное значение. В противном случае ему присваивается номер фильтра, дающего минимальное значение свертки (рассматривается её вещественная часть). Значение свёртки фильтра с максимальным откликом, подвергнутое локальной дискретизации, используется в качестве второго признака.

Для сравнения объектов предложена метрика, являющаяся взвешенной суммой метрик для первого и второго признаков. С помощью разработанного устройства собрана база изображений пальцев PolyU FKP [33], включающая 7920 изображений пальцев разрешением 768×576 пикселей от 165 человек, — по 4 пальца на человека (средний и указательный на обеих руках) и по 12 изображений на палец (в рамках двух сессий). Разрешающая способность снимков в базе составляет 400 dpi, но основные изложенные авторами результаты вычислительных экспериментов получены на изображениях с искусственно пониженным до 150 dpi — в целях демонстрации работоспособности метода на снимках с низкой разрешающей способностью. Достигнуто значение $EER = 1.48\%$.

Метод Local-Global Information Combination (LGIC) [34] представляет собой дальнейшее развитие идеи кодирования особенностей ориентации морщин. В самом деле, если в фильтре Габора увеличивать параметр амплитуды σ , то влияние локальных особенностей изображения на результат свёртки будет ослабевать. В предельном случае результат свёртки отражает

только глобальные особенности, а преобразование Габора по сути превращается в преобразование Фурье. Авторы публикации используют некоторое подмножество коэффициентов дискретного двумерного преобразования Фурье, применённого к области кожных складок пальца, в качестве дополнительного вектора признаков. Задействованный метод сопоставления объектов на основе признакового описания глобальной ориентации морщин был ранее представлен в [35] и получил название Band-Limited Phase-Only correlation (BLPOC). По утверждению авторов, метод LGIC в задаче верификации при тестировании на базе Poly FKP превосходит хорошо зарекомендовавшие себя методы BLPOC [35] и CompCode [23].

Методы, основанные на прочих преобразованиях. В работе [36] каждому пикселю области, содержащей кожные складки пальца, ставится в соответствие вектор признаков, задающий доминирующую ориентацию морщин в окрестности с центром в данном пикселе (KnuckleCode). Для кодирования используется модифицированное конечное дискретное преобразование Радона, которое в случае растрового изображения имеет значительно меньшую вычислительную сложность, чем двумерные фильтры Габора. Для расчёта расстояний между объектами ищется сдвиг матриц признаковых описаний, обеспечивающий минимальное расстояние по Хеммингу. Тестирование осуществлено на базе из работы [37], включающей 158 человек и 790 изображений. Достигнуто значение $EER = 1.14\%$.

В методе MonogenicCode [38] для извлечения локальных признаков, характеризующих пиксель изображения, используется моногенный сигнал, который в точке может быть представлен как вектор в сферической системе координат — с параметрами ориентации и фазы. На базе PolyU FKP для задачи верификации авторами получен результат $EER = 1.72\%$, что сравнимо с показателем CompCode. При этом скорость извлечения признакового описания у MonogenicCode почти в 3 раза выше, чем у CompCode.

2.3.4 Мультимодальные

В работе [39] используются текстурные признаки рисунка ладони наряду с признаками формы. Для сегментации области интереса на внутренней стороне ладони применяется карта расстояний внутренних точек ладони до контура. Текстурные признаки основаны на коэффициентах применённого к области интереса дискретного косинусного преобразования. Также используются 23 признака формы: длины и ширины пальцев, ширина и длина ладони, площадь ладони и ещё 7 признаков. Для выбора наиболее информативного подмножества признаков использован алгоритм, основанный на корреляции (correlation-based feature selection, CFS). Тестирование осуществлено на собственной базе из 100 человек и 1000 изображений.

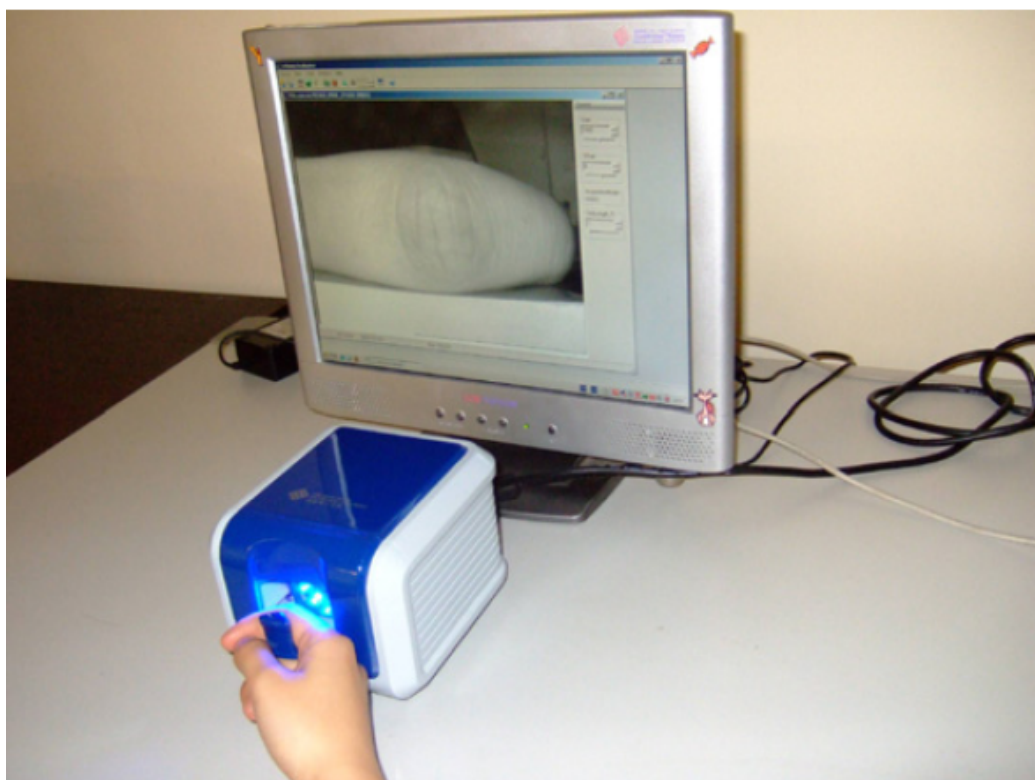


Рис. 3: Устройство для сбора изображений кожных складок пальцев внешней стороны ладони, использованное для создания базы PolyU FKP [33].

Для её создания применена обычная цифровая камера и поверхность с чёрным покрытием, к которой нужно приложить тыльную сторону руки. С применением SVM классификатора достигнута точность 98%.

Те же авторы в статье [40] вводят понятие *качества пользователя*, характеризующего возможность его корректной классификации. Оно определяется как минимальное расстояние между изображениями данного пользователя в обучающей выборке. Полученное значение используется для вычисления итогового расстояния между объектами в классификаторе. На базе из публикации [39] достигнуто значение $EER = 2.2\%$.

В работе [41] используется комбинация признаков на основе формы ладони, рисунка ладони и кожных складок пальцев внутренней стороны ладони. Признаки формы представлены «классическими» характеристиками размеров пальцев, а также параметрами их контуров. Для извлечения особых точек текстуры ладони применяется детектор углов Канаде-Лукаса-Томаси (использует матрицу вторых производных изображения). При генерации текстурных признаков пальцев использован оператор Собеля с последующим шумоподавлением на основе вейвлетов Добеши. Создана собственная база из 1900 изображений 95 человек, съёмка произведена внутри закрытой коробки в контролируемых условиях. Применение процеду-

ры иерархического сопоставления объектов позволило достичь при тестировании значений $FRR = 0.00898$ и $FAR = 2.52 * 10^{-6}$.

В статье [42] предлагается бимодальная система, использующая рисунок ладони и текстурные складки на внешней стороне пальцев. Для сравнения каждого из признаков использовалась функция фазовой корреляции [43] — значение пика этой функции принимается за расстояние. Объединение результатов сравнения осуществляется взвешенным суммированием, веса вычисляются как нормализованные значения EER каждого из признаков. Для идентификации применялась минимаксная нормализация расстояния. База для тестирования сформирована на основе баз PolyU Palmprint [20] и PolyU FKP [33] — по 12 изображений кожных складок пальцев и внутренней стороны ладони для 150 человек. Для идентификации достигнуто значение $EER = 0.003\%$, для верификации значение GAR (Genuine Acceptance Rate) составило 99.647%.

2.4 Мобильные системы идентификации по ладони

2.4.1 По рисунку ладони

В статье [44] представлена система верификации по внутренней стороне ладони, работающая на мобильном устройстве. Предполагается, что в центре изображения находятся пиксели ладони, по которым вычисляются пороги в пространстве nRGB (нормализованное пространство RGB, значение каждого канала поделено на сумму значений по трём каналам) для пороговой сегментации. В основе работы метод CompCode [23], немного модифицированный и ускоренный на этапе сопоставления признакового описания объектов. Особое внимание уделено разработке алгоритма устойчивой локализации области интереса, так как основной сценарий использования — съёмка «с руки». Собрана собственная база изображений из 30 человек, по 20 изображений на человека разрешением 640×480 пикселей. В качестве устройства использован iPhone 4, камера которого снабжена вспышкой, автофокусом, системой стабилизации и матрицей 5 мегапикселей. Мобильное приложение работает с отдельными фотографиями (не видеопотоком), на обработку каждой уходит порядка 0.5 с. без учёта времени сопоставления с объектами-эталоном. Примерные достигнутые показатели метода: $EER = 2.8\%$, $GAR = 0.968$, $FAR = 0.005$ (в публикации содержится только ROC-кривая).

В статье [45] предложена мобильная система, также работающая с изображениями внутренней стороны ладони. Сегментация ладони выполняется с помощью алгоритма роста области, начальное приближение для которого выбирается в середине ладони по цвету кожи в пространстве HSV. С помощью функции расстояний точек контура от некоторой рефе-

ренской точки на ладони определяются особые точки, по которым происходит локализация квадратной области интереса. Для снижения влияния искажений, неизбежно возникающих из-за вариаций углов съёмки при фотографировании «с руки», для каждой пары областей интереса «тестовое изображение – эталонное изображение» ищутся параметры аффинного преобразования, построенному по отобранным парам контрольных точек.

Далее на изображении внутри области интереса выделяются блоки, каждому из которых ставится в соответствие блок на сравниваемом изображении. Извлечение признаков и последующее сопоставление делается поблочно на основе функции фазовой корреляции методом Band-Limited Phase-Only Correlation (BLPOC) [43]. Итоговое расстояние между объектами определяется как пик функции фазовой корреляции, являющей усреднением по соответствующим функциям в блоках. Использован телефон HTC Nexus One, камера которого снабжена вспышкой, автофокусом, системой стабилизации и матрицей 5 мегапикселей. Мобильное изображение работает с отдельными фотографиями (не видеопотоком). Для каждого снимка генерация признаков занимает порядка 0.6 с., ещё 0.4 с. тратится на сопоставление с объектами-эталонами. На собственной базе изображений рисунков ладони (30 человек, 520 снимков, разрешение 640×480 пикселей, собрана при помощи смартфона) и на базе PolyU Palmprint [20] достигнуты значения EER равные 4.072% и 0.051% соответственно.

2.4.2 По форме ладони

Представленный в работе [46] метод использует контур ладони для распознавания. Для сегментации применяется алгоритм кластеризации k-средних с двумя классами, за которой следует постобработка морфологическими операциями для увеличения гладкости границы. Алгоритм сегментации разработан с расчётом на отсутствие ограничений к фону и освещению, но для более уверенного распознавания рекомендуется наличие однородного фона, контрастирующего с рукой. В качестве признаков использованы два вектора. Первый представляет собой набор углов наклона касательной к контуру в точках контура. Второй — набор расстояний от точек контура до его центроида. Для сравнения векторов признаков (как правило разной длины) применён алгоритм выравнивания последовательностей на базе нечёткой логики, обычно используемый при анализе ДНК-цепочек. База собрана с помощью мобильного устройства съёмкой «с руки» и включает в себя изображения 45 человек разрешением 1600×1200 пикселей. Достигнуто значение $EER = 3.7\%$. На момент публикации сама система на базе мобильного устройства авторами не реализована.

В более поздней работе с участием тех же авторов [47] предлагается другой метод распознавания по контуру. Алгоритм сегментации основан на представлении изображения в виде графа. Вершинам графа соответствуют пиксели, рёбра соединяют соседние пиксели (используется структура 4-связности), веса рёбер отражают различия в интенсивностях пикселей (b-канал в пространстве цветов CIE 1976 $L^*a^*b^*$). Пиксели объединяются в группы, имеющие похожие в смысле текстуры свойства. Далее последовательно выполняется агрегация полученных подграфов в более крупные структуры. Процесс прекращается, когда остается два подграфа, они представляют фон и ладонь. Для выделения пальцев применяется морфологическое открытие с большим радиусом. Результат вычитается из исходного бинарного изображения, и остаются области, соответствующие пальцам. После этого ищутся концы пальцев и точки между ними. Следующим шагом рассчитываются геометрические признаки каждого из пальцев (длина, кривизна, вектор параметров ширины) и объединяются в единый вектор признаков. В качестве классификатора выбран SVM с линейным ядром. База собрана с помощью смартфона Apple iPhone съёмкой «с руки» и включает в себя по 15 изображений 50 человек разрешением 400×300 пикселей. Достигнуто значение EER = 6%. В рассматриваемой статье сама система на базе мобильного устройства авторами не упоминается, но в публикации [48] на основе изложенного метода представлено Android приложение для задачи верификации. Мобильное приложение работает с отдельными фотографиями (не видеопотоком), на обработку каждого снимка уходит порядка 2.0 с., ещё около 0.4 с. тратится на сопоставление с объектами-эталоном.

2.4.3 По кожным складкам пальцев

В работе [49] для детектирования пальца на изображении применён алгоритм AdaBoost с признаками Хаара, при этом доля корректных распознаваний составляет 72%. Классификатор данного типа требует, чтобы фон и освещение мало отличались от тех, что присутствовали на обучающих изображениях. Для снятия этого ограничения владельцу смартфона предоставлена возможность самостоятельно указывать центр области локализации. После предобработки применяется детектор границ Канни для локализации области складок. Для минимизации влияния ориентации в сравнении с эталонами участвует не только оригинальное тестовое изображение пальца, но и несколько его копий, повернутых на небольшие углы. Генерация признаков основана на лог-Габор фильтрах (log-Gabor filters). Для определения минимального расстояния между векторами признаков к ним применяются битовые сдвиги друг относительно друга, вычисляется нормализованное расстояние Хэмминга и выбирается

наименьшее. Авторами написано мобильное приложение для операционной системы Android, которое работает с отдельными фотографиями (не видеопотоком). С его помощью собрана собственная база изображений пальцев, состоящая из 561 изображений пальцев 109 человек разрешением 768×1024 пикселей. Достигнуто значение EER = 9%. Для задачи верификации время распознавания на смартфоне HTC Desire HD A9191 составляет до 5 с.

2.5 Выводы к главе 2

Биометрические признаки на основе изображения ладони в видимом диапазоне становятся всё более востребованными в системах идентификации личности — многочисленные исследования в этой области и стремительное развитие цифрового оборудования делают распознавание всё точнее и быстрее.

В разделе 2.2 показана связь между разрешением изображения ладони (или её части) и характером извлекаемых текстурных особенностей через показатель разрешающей способности (dpi). Важным фактом является то, что использованные для тестирования большинством авторов базы прецедентов Poly FKP [33], PolyU Palmprint [20] и [37] состоят из изображений с dpi не более 400, т. е. для анализа доступны морщины и складки кожи, но не папиллярные линии.

Исторически большинство работ сконцентрировано на разработке методов биометрической идентификации по внутренней стороне ладони (раздел 2.3.1), но для массового использования данная модальность может оказаться не самым удачным вариантом. Рядовому пользователю удобно предъявлять ладонь тыльной стороной вверх, что влечёт за собой необходимость в специализированном сканирующем (или фотографирующем) оборудовании — а это противоречит сформулированному в разделе 1.2 требованию доступности системы распознавания. Более того, демонстрация внутренней стороны ладони с последующей съёмкой не для всех людей является допустимой и эмоционально нейтральной, что нарушает требование социальной приемлемости целевой системы.

Как было сказано выше, предъявление ладони тыльной стороной вверх физиологически просто и психологически комфортно для человека, при этом изображение внешней стороны ладони может быть получено обычной цифровой камерой (в том числе и мобильного устройства). С учётом вышесказанного, при разработке метода идентификации и связанной с ним системы автором использованы модальности формы ладони и рисунка кожных складок пальцев на внешней стороне ладони.

Хочется особо отметить похожесть методов, описанных авторами в трудах по распознаванию на основе кожных складок пальцев, с методами, работающими по рисунку внутренней стороны ладони. Это можно объяснить как общей природой морщин в обоих случаях, так и грамотным выбором математического аппарата. Примером таких «универсальных» подходов к генерации признакового описания являются преобразования Фурье, Габора, Радона и Рисса. Рассмотренные в разделе 2.3.3 методы на их основе стабильно демонстрируют показатели EER < 2% на внушительных публично доступных базах прецедентов PolyU FKP [33] и [37].

Отдельно стоят методы распознавания по форме ладони (раздел 2.3.2). Они требовательны к качеству сегментации, при этом не дают высоких показателей точности (EER обычно превышает 2%) — сказывается как вариабельность формы ладони в рамках одного человека, так и относительно высокая похожесть формы ладоней разных людей. Тем не менее, это не исключает использование формы ладони в качестве дополнительной модальности в некоторых системах биометрической идентификации. Почти во всех изученных работах предполагается стационарный сценарий получения изображения ладони — с фиксацией на стекле сканера или на поверхности перед объективом неподвижно закреплённой камеры (призвано уменьшить изменчивость формы в зависимости от угла съёмки). Исключением являются работы [46] и [47], где съёмка производится «с руки» с помощью мобильного устройства. Достигнутые значения EER в них ожидаемо невысоки — равны 3.7% и 6% соответственно.

Почти все приведенные в обзоре методы биометрической идентификации по кожным складкам внешней стороны пальцев ориентированы на работу с изображениями из базы PolyU FKP [33]. Эта база собрана с помощью специально разработанного устройства (рис. 3), в котором расположение пальца фиксировано и освещение постоянно. Как следствие, области кожных складок имеют высокое качество, одинаковый масштаб и единую ориентацию — работать с такими данными довольно удобно. Мультимодальные методы, в которые входят кожные складки пальцев, как правило используют в качестве второй модальности рисунок ладони [42]; работы, где дополнительно применяется форма ладони, не найдены.

Мобильные системы идентификации на основе ладони (по рисунку ладони, форме ладони, кожным складкам пальцев внешней стороны ладони) слабо представлены в литературе (раздел 2.4) и в силу своих характеристик не годятся для решения реальных прикладных задач. Среди них ещё менее распространены мобильные системы, работающие с кожными складками пальцев, — единственной известной автору релевантной работой в этом направлении является [49]. Точность распознавания предложенного программного комплекса невысо-

ка, разработанное приложение ориентировано для персональной аутентификации, а не для идентификации в малых группах.

Подводя итоги, можно утверждать, что в научной литературе не представлено системы биометрической идентификации человека по снимку внешней стороны ладони на базе мобильного устройства, удовлетворяющей требованиям из раздела 1.2. Последующие главы работы посвящены предложенному методу распознавания и особенностям реализации программного комплекса.

3 Метод идентификации по ладони

3.1 Машинное обучение в задаче идентификации по ладони

Машинное обучение (англ. machine learning, ML) — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться. Различают два типа обучения. *Обучение по прецедентам*, или *индуктивное обучение*, основано на выявлении общих закономерностей по частным эмпирическим данным. *Дедуктивное обучение* предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний. Дедуктивное обучение принято относить к области *экспертных систем*, поэтому термины машинное обучение и обучение по прецедентам можно считать синонимами.

3.1.1 Основные понятия

Пусть имеется множество объектов X , множество ответов Y , и существует целевая функция $y^*: X \rightarrow Y$, значения которой $y_i = y^*(x_i)$ известны только на конечном подмножестве объектов $\{x_1, \dots, x_l\} \subset X$. Пары «объект-ответ» (x_i, y_i) называются прецедентами. Совокупность пар $X^l = (x_i, y_i)_{i=1}^l$ называется обучающей выборкой.

Задача обучения по прецедентам заключается в том, чтобы восстановить функциональную зависимость между объектами и ответами, то есть построить отображение $a: X \rightarrow Y$, удовлетворяющее следующей совокупности требований:

- Отображение должно допускать эффективную компьютерную реализацию. По этой причине будем называть его *алгоритмом*.
- Алгоритм $a(x)$ должен воспроизводить на объектах выборки заданные ответы: $a(x_i) = y_i, i = 1, \dots, l$. Равенство здесь может пониматься как точное или как приближённое, в зависимости от конкретной задачи.
- Алгоритм $a(x)$ должен обладать *обобщающей способностью*, то есть достаточно точно приближать целевую функцию $y^*(x)$ не только на объектах обучающей выборки, но и на всём множестве X .
- На алгоритм $a(x)$ могут накладываться разного рода *априорные ограничения*, такие, как непрерывность, гладкость, монотонность, устойчивость, и т. д. В некоторых случаях задаётся функциональный вид (*модель*) алгоритма $a(x)$.

В данной работе задача идентификации человека по ладони сводится к разновидности задачи машинного обучения, называемой *классификацией*. При этом $Y = \{1, \dots, M\}$,

где M — число непересекающихся классов, и X также разбивается на классы $K_y = \{x \in X: y^*(x) = y\}$. Алгоритм $a(x)$ должен давать ответ на вопрос «какому классу принадлежит x ?».

В контексте исследования:

- каждому человеку взаимно-однозначно назначена метка класса y ;
- объектом x является изображение ладони человека;
- произвольный класс K_y множества объектов X сформирован из изображений ладони, принадлежащих человеку с меткой класса y ;
- искомый алгоритм классификации должен давать ответ на вопрос «какому человеку принадлежит изображение ладони x ?»

3.1.2 Модель алгоритмов и метод обучения

Моделью алгоритмов называется параметрическое семейство отображений A , из которого выбирается искомый алгоритм $a(x)$: $A = \{\varphi(x, \gamma) | \gamma \in \Gamma\}$, где $\varphi: X \times \Gamma \rightarrow Y$ — некоторая фиксированная функция, Γ — множество допустимых значений γ , называемое *пространством параметров* или *пространством поиска* (search space).

Процесс подбора параметров модели по обучающей выборке называют *настройкой* (fitting) или *обучением* (training, learning) алгоритма. В результате настройки выбирается единственный алгоритм $a \in A$, который должен приближать целевую зависимость.

Методом обучения называется отображение $\mu: (X \times Y)^l \rightarrow A$, которое произвольной конечной выборке X^l ставит в соответствие алгоритм $a: X \rightarrow Y$. Говорят также, что метод μ строит алгоритм a по выборке X^l . Метод обучения, как и сам алгоритм a , должен допускать эффективную программную реализацию.

Итак, в задачах обучения по прецедентам чётко различаются два этапа:

- на этапе обучения метод μ по выборке X^l строит алгоритм $a = \mu(X^l)$;
- на этапе применения алгоритму a подаются на вход новые объекты x , вообще говоря, отличные от обучающих, для получения ответов $y = a(x)$.

Этап обучения наиболее вычислительно сложен. Как правило, он сводится к поиску параметров модели, доставляющих оптимальное значение заданному функционалу качества. В данном исследовании примером μ может служить процедура настройки параметров классификатора 1NN, описанная в разделе 4.3.2.

Минимизацию эмпирического риска следует применять с известной долей осторожности, поскольку главным её недостатком является склонность к *переобучению*. Если алгоритм a доставляет минимум функционалу $Q(a, X^l)$ на заданной обучающей выборке X^l , это ещё не гарантирует, что он будет хорошо приближать целевую зависимость на произвольной *контрольной выборке* $X^k = (x'_i, y'_i)_{i=1}^k$. Когда качество работы алгоритма на новых объектах, не вошедших в состав обучения, оказывается существенно хуже, чем на обучающей выборке, говорят об эффекте *переобучения* (overtraining) или *переподгонки* (overfitting). При решении практических задач с этим явлением приходится сталкиваться очень часто.

3.1.3 Функционал качества

Функция потерь (loss function) — это неотрицательная функция $\mathcal{L}(a, x)$, характеризующая величину ошибки алгоритма a на объекте x . Если $\mathcal{L}(a, x) = 0$, то ответ $a(x)$ называется корректным. *Функционал качества* алгоритма a на выборке X^l :

$$Q(a, X^l) = \frac{1}{l} \sum_{i=1}^l \mathcal{L}(a, x_i).$$

Функционал Q называют также *функционалом средних потерь* или *эмпирическим риском*, так как он вычисляется по эмпирическим данным $(x_i, y_i)_{i=1}^l$. Функция потерь, принимающая только значения 0 и 1, называется бинарной. В этом случае $\mathcal{L}(a, x) = 1$ означает, что алгоритм a допускает ошибку на объекте x , а функционал Q называется *частотой ошибок* алгоритма a на выборке X^l .

Используемая в настоящей работе функция потерь:

$$\mathcal{L}(a, x) = [a(x) \neq y^*(x)]$$

— представляет собой индикатор несовпадения с правильным ответом.

Классический метод обучения, называемый *минимизацией эмпирического риска* (empirical risk minimization, ERM), заключается в том, чтобы найти в заданной модели A алгоритм a , доставляющий минимальное значение функционалу качества Q на заданной обучающей выборке X^l :

$$\mu(X^l) = \arg \min_{a \in A} Q(a, X^l).$$

Именно он и используется для обучения итогового классификатора в настоящей работе (раздел 4.3.2).

3.1.4 Признаковое описание объектов

Признаком (feature) называется отображение $f: X \rightarrow D_f$, описывающее результат измерения некоторой характеристики объекта, где D_f — заданное множество. В зависимости от множества допустимых значений D_f признаки делятся на следующие типы:

- бинарный признак: $D_f = \{0, 1\}$;
- номинальный признак: D_f — конечное множество;
- порядковый признак: D_f — конечное упорядоченное множество;
- количественный признак: $D_f = \mathbb{R}$.

Если все признаки имеют одинаковый тип, $D_{f_1} = \dots = D_{f_n}$, то исходные данные называются *однородными*, в противном случае — *разнородными*. Пусть имеется набор признаков f_1, \dots, f_n . Вектор $(f_1(x), \dots, f_n(x))$ называют *признаковым описанием объекта* $x \in X$. В дальнейшем мы не будем различать объекты из X и их признаковые описания, полагая $X = D_{f_1} \times \dots \times D_{f_n}$. Совокупность признаковых описаний всех объектов выборки X_l , записанную в виде таблицы размером $l \times n$, называют *матрицей объектов–признаков*:

$$F = \|f_j(x_i)\|_{l \times n} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_n) & \dots & f_n(x_l) \end{pmatrix} \quad (1)$$

Матрица объектов-признаков является стандартным и наиболее распространённым способом представления исходных данных. Однако на практике встречаются задачи, в которых данные устроены сложнее, например, описания объектов могут иметь переменную длину. В таких случаях по имеющимся исходным данным вычисляются преобразованные данные, имеющие стандартный вид (1). Этот приём называется *извлечением признаков* (features extraction) из данных.

Таким образом, признаки — это характеристики объектов, которые либо измеряются непосредственно, либо вычисляются по «сырым» исходным данным. Любое отображение из множества X можно рассматривать как признак. В текущем исследовании признаки извлекаются из изображений внешней стороны ладони (раздел 3.5).

3.1.5 Цикл исследования

Построение обучаемых алгоритмов в прикладных задачах (одной из которых, безусловно, является рассматриваемая задача идентификации личности по ладони) — это исследо-

вательская работа. Здесь нет универсальных рецептов, и решающую роль играют эксперименты на реальных данных. Приведенная ниже схема описывает наиболее типичный цикл исследований с комментариями в контексте данной работы. Выполнение каждого этапа может опровергнуть очередную гипотезу и вернуть исследователя к любому из предыдущих этапов. Процесс поиска решения предполагает постепенное углубление в суть задачи.

- *Постановка задачи.* На начальном этапе поставлена цель исследования, сформулированы исследовательские задачи и определены свойства, которыми должна обладать разрабатываемая система идентификации личности по ладони (раздел 1.2).
- *Сбор данных.* Сбор данных (раздел 4.2) проходил в несколько этапов. Определены начальные требования к съёмке и собрана «База-1» (раздел 4.2.1) изображений для тестирования алгоритмов детектирования ладони и распознавания по её форме. Далее в ходе разработки алгоритма распознавания создана обучающая выборка «База-2» — для её наполнения применён уже оттестированный алгоритм детектирования, а также уточнены условия съёмки. Более того, для дополнительного тестирования метода детектирования применена внешняя база HGC (раздел 4.2.3), а для тестирования метода распознавания — внешняя база PolyU (раздел 4.2.4).
- *Формулировка гипотез.* Одной из ранних гипотез было то, что для идентификации человека по ладони в малых группах достаточно лишь признаков, полученных на основе формы ладони. Эта гипотеза не оправдала себя на реальных данных, в силу чего сформулирована новая — и в метод распознавания были добавлены текстурные признаки. Проведенные вычислительные эксперименты показали значимый рост точности, ставший основанием для формулировки финальной гипотезы — для уверенной идентификации хватит только текстурных признаков.
- *Предобработка данных.* Метод детектирования (раздел 3.4) в разработанном решении идентификации по ладони по сути выполняет роль предобработки данных, так как отфильтровывает изображения ладони, непригодные для дальнейшего анализа (и извлечения признаков) в методе распознавания.
- *Разработка моделей алгоритмов.* В работе широко используются как классические методы машинного обучения (метрический алгоритм классификации 1NN, сегментация по Оцу, сегментация на основе разрезов в графах и т. д.), так и разработанные (модифицированные) автором: метод семантической разметки ладони (раздел 3.4.3), метод

извлечения (раздел 3.5.4) и сравнения (раздел 3.5.5) Габор-признаков, метод сравнения на основе SIFT-признаков (раздел 3.5.6) и т. д.

- *Оценка качества алгоритмов.* В рамках методологии тестирования обучаемых алгоритмов на каждом этапе исследования проводились вычислительные эксперименты (раздел 4.3) на реальных данных из обучающих выборок (раздел 4.2).
- *Отбор или коррекция алгоритмов.* Не все алгоритмы, протестированные или разработанные автором, вошли в итоговое решение. Примером может служить отбор подходящих для этапов детектирования и распознавания методов сегментации (раздел 3.4.1).
- *Опытная эксплуатация.* Разработанная мобильная система распознавания по внешней стороне ладони применена для контроля посещаемости студентов ВМК МГУ, а также для учёта гостей на Московском Фестивале науки, в ходе чего подтверждены заявленные свойства системы.
- *Автоматизированное принятие решений.* Текущая реализация системы распознавания на базе мобильного устройства не требует участия эксперта — как сбор образцов, так и процесс идентификации происходит при участии одного лишь испытуемого.

3.2 Задачи детектирования и распознавания

Для извлечения используемых в данной задаче признаков необходима качественная сегментация изображения из видеопотока камеры (то есть определение области, занимаемой ладонью), что, в сочетании с изменчивыми условиями съёмки, предъявляет весьма высокие требования к алгоритму сегментации. В то же время, существующие алгоритмы не могут обеспечить достаточно высокое качество сегментации при работе в реальном времени на мобильных устройствах из среднего ценового диапазона.

Кроме того, в биометрической системе необходимо хранить базу биометрических образцов. Хранилище при этом должно обеспечивать достаточный уровень безопасности, а также доступ к базе для всех клиентов. Для решения этой проблемы необходимо наличие центрального сервера. В качестве сервера может выступать как персональный компьютер, так и специализированное решение. Принимая во внимание также ограниченные вычислительные возможности мобильных устройств, предлагается перенести часть процесса идентификации на сервер, тем самым уменьшив нагрузку на мобильный клиент. А именно, разделить задачу идентификации личности человека по ладони на задачи детектирования и распознавания.

Задача детектирования — обнаружение объекта (в нашем случае ладони) на изображении и, в случае наличия, проверка «пригодности» изображения ладони для последующего распознавания.

Задача распознавания — генерация признакового описания по изображению ладони, прошедшей проверки на «пригодность», и непосредственно классификация ладони путём сопоставления с базой признакового описания ладоней-эталонов.

При этом детектирование выполняется на мобильном устройстве, а распознавание на удалённом сервере. Таким образом, осуществляется не только логическое, но и структурное разделения исходной задачи идентификации на детектирование и распознавание.

Связь между клиентом и сервером осуществляется с помощью беспроводной сети. Стоит также отметить, что задача распознавания является вычислительно трудной даже для персонального компьютера — как минимум потому, что одним из её этапов является «точная» сегментация. Это значит, что сервер не может обрабатывать большое число изображений в режиме реального времени, что вкупе с ограничениями сети предъявляет определённые требования к модулю детектирования.

Алгоритм детектирования должен допускать к дальнейшей обработке как можно меньше «непригодных» изображений, т. е. иметь низкое значение FAR (иначе сервер не справится с большим количеством изображений). Вместе с тем, при соблюдении распознаваемым пользователем процедуры предъявления ладони должно поддерживаться невысокое значение FRR (это обеспечит приемлемое общее время распознавания).

Далее рассмотрим подробнее каждую из *задач*.

3.3 Корректность ладони

Прежде чем переходить к предлагаемому в работе методу детектирования ладони, нужно сформулировать, что является *корректной ладонью* в контексте задачи детектирования:

- ладонь вместе с частью запястья находится целиком в кадре;
- посторонние предметы не закрывают ладонь, т. е. отсутствуют часы, кольца, браслеты, длинные рукава и т. п.;
- все пальцы расставлены (разведены).

Некорректная ладонь — ладонь, которая не является корректной в терминах определения выше. *Корректными изображениями (кадрами)* будем называть те изображения (кадры), которые содержат корректную ладонь. *Некорректные изображения (кадры)*, в свою очередь,

либо не содержат ладони вообще, либо содержат некорректную ладонь. Соответствующие проверки назовём проверками на *корректность*.

3.4 Метод детектирования

В алгоритме детектирования можно выделить ряд последовательных этапов:

1. обнаружение объекта-кандидата на роль ладони в кадре (задача «быстрой» сегментации — раздел 3.4.1);
2. создание скелетного представления ладони (раздел 3.4.2), построение семантической разметки и проверка ладони на корректность (раздел 3.4.3);
3. определение расфокусированных кадров (раздел 3.4.4);
4. в случае корректной ладони, находящейся в фокусе, — формирование данных для отправки в модуль распознавания (в частности, необходимых для работы «точной» сегментации — раздел 3.4.5). Такую ладонь, пригодную для последующего распознавания, назовём *валидной*.

Таким образом, корректность ладони является необходимым (но не достаточным) условием её валидности, и только валидные ладони участвуют в последующем распознавании.

Важно отметить, что алгоритм детектирования должен отрабатывать в режиме реального времени на кадрах из видеопотока камеры мобильного устройства, что накладывает достаточно жёсткие требования к производительности — 20-40 мс на обработку одного кадра.

Абсолютное большинство методов сегментации (например, на основе глубоких нейронных сетей или минимальных разрезов на графах) не укладывается в эти временные ограничения. В рамках поиска решения автором отобраны и опробованы перспективные «быстрые» методы сегментации, описанные разделе 3.4.1.

В результате, выбор сделан в пользу пороговой бинаризации изображения по Оцу (пункт «Пороговая бинаризация» в разделе 3.4.1). Для фонов, однородных по цвету и контрастных относительно ладони, метод даёт хорошее сочетание качества сегментации и скорости работы. Это позволяет быстро выделить участок, предположительно относящийся к ладони, а затем для выделенной области проверить условия корректности ладони, указанные в списке выше.

Автором проведено отдельное исследование по возможности использования пороговой бинаризации для произвольных фонов при условии предварительной локализации области

ладони с помощью метода Виолы-Джонса (пункт «Метод Виолы-Джонса с последующей пороговой бинаризацией» в разделе 3.4.1). Тестирование подхода в реальных условиях показало его чувствительность к ориентации предъявляемой ладони, что послужило причиной отказа от дальнейшего применения. На практике работа с однородным контрастным фоном при эксплуатации разработанной системы распознавания ладони не является значимым ограничивающим фактором.

3.4.1 Сегментация

Существует большое количество алгоритмов сегментации изображений. В контексте задачи нас интересует бинарная (двухклассовая) сегментация, в ходе которой происходит назначение метки («объект» или «фон») каждому пикселю. Среди методов бинарной сегментации можно выделить несколько обширных групп:

- Пороговая бинаризация по яркости. Пиксели делятся на две группы, в зависимости от того, больше ли их яркость некоего порога или нет.
- Вычитание фона. Предполагается, что изображение или последовательность изображений состоит из неподвижного фона и некоторого объекта. Строится модель фона и пиксели проверяются на соответствие этой модели.
- Кластеризация по цвету. Пиксели делятся на группы на основе своего цвета.
- Минимизация энергии. Разбиение пикселей на группы осуществляется путём оптимизации некой функции энергии, заданной на графе, построенном по изображению.
- Нейронные сети. Изображение подаётся на вход свёрточной нейронной сети, которая определяет принадлежность каждого пикселя к фону или объекту.

Рассмотрим каждую из групп подробнее.

Пороговая бинаризация. Один из самых простых методов двухклассовой сегментации изображений — пороговая бинаризация. Порог может быть как фиксированным ([26], [19]), так и вычисляться адаптивно [37]. Популярный и эффективный алгоритм вычисления порога — метод Оцу [50], минимизирующий внутриклассовую дисперсию и максимизирующий межклассовую (объект и фон). Данный метод работает очень быстро, но требует, чтобы фон и объект контрастировали друг с другом. Например, фон должен быть тёмным, а объект — светлым, или наоборот. В частности, в случае светлого фона и светлого объекта качество

сегментации может оказаться неудовлетворительным. Для решения этой проблемы автором предложено ввести дополнительный шаг, базирующийся на алгоритме Виолы-Джонса [51].

Метод Виолы-Джонса с последующей пороговой бинаризацией. Предлагаемый автором принцип работы заключается в следующем. Сначала следует выполнить локализацию ладони на изображении, т. е. получить прямоугольник, её содержащий. Это можно сделать, применяя подход, аналогичный методу Виолы-Джонса [51] для детектирования лиц.

Если ладонь не обнаружена, то обработку на этом можно прекратить, что сэкономит значительное количество времени.

При наличии ладони в кадре в случае светлого фона предполагается использовать то, что центральная часть полученного прямоугольника гарантированно принадлежит ладони. Для этого изображение переводится в цветовое пространство HSV (Hue, Saturation, Value). Далее строится цветовая модель кожи на основе выделенного прямоугольника. Один из вариантов такой модели — HS-гистограмма в указанном цветовом пространстве. Яркость (Value) при этом не учитывается, чтобы снизить влияние уровня освещения. После этого для каждого пикселя изображения в HSV-пространстве можно найти вероятность того, что пиксель принадлежит коже, на основании ранее построенной модели. Эта вероятность будет равна значению модельной гистограммы в ячейке, соответствующей H- и S-координатам данного пикселя в пространстве HSV. Сегментация исходного изображения тогда осуществляется пороговой бинаризацией полученных вероятностей.

В случае тёмного фона достаточно использовать простую пороговую бинаризацию по методу Оцу.

Метод Виолы-Джонса состоит из каскада классификаторов на основе модифицированных признаков Хаара (рис. 4). Признаки представляют собой бинарные маски определенных конфигураций (со значениями +1 и -1), значение такого признака в точке изображения получается свёрткой окрестности данной точки с маской.

Во время работы алгоритма по изображению движется окно поиска (обычно размера 24×24). Для каждого положения окна вычисляются значения признаков Хаара с различными положениями и ориентациями (порядка 162336 признаков). Для каждого признака обучается простой пороговый классификатор, который выносит решение, присутствует в данном окне объект или нет. Такие классификаторы объединяются в ансамбль с помощью алгоритма обучения AdaBoost (Adaptive Boosting), имеющий значительно большую разрешающую способность, чем каждый из простых классификаторов по отдельности.

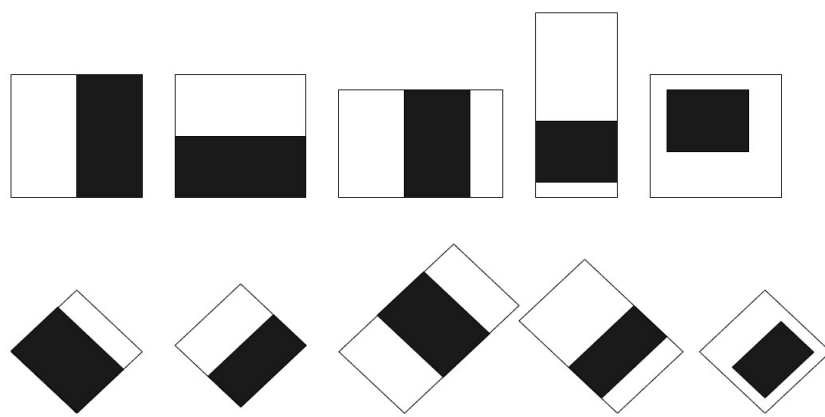


Рис. 4: Прimitives Хаара.

Ясно, что вычислять большое количество признаков для каждого положения окна поиска нецелесообразно. Необходимо отсеивать окна, заведомо не содержащие искомый объект. Авторами метода установлено, что простейший классификатор с двумя признаками может достичь почти 100% чувствительности (обнаруживать объект, когда он присутствует) при 50% доле ложноположительных срабатываний. Такой классификатор является первой ступенью т. н. каскада классификаторов — если он выдаёт отрицательный результат, то обработка данного окна прекращается. В противном случае начинает работать более сложный классификатор второй ступени (порядка 10 признаков) и т. д. Подобная цепочка усложняющихся классификаторов обеспечивает приемлемое быстродействие для приложений реального времени.

Автором проведено тестирование алгоритма Виолы-Джонса для детектирования левой руки в условиях меняющихся фона и освещения. Для обучения алгоритма необходимы две группы изображений: позитивные примеры (изображения, содержащие искомый объект) и негативные примеры (фоновые изображения без объекта на них). Алгоритму требуется большое количество примеров — как позитивных (рис. 5), так и негативных (рис. 6). Необходимые изображения извлечены из видеофайлов, снятых на мобильный телефон, и сведены в базу Хаар-1. Помимо этого составлена ещё одна база Хаар-2, сформированная из изображений базы Хаар-1, подвергнутых небольшим поворотам и наложенных на различные фоны.

Нужно отметить, что алгоритм Виолы-Джонса на основе признаков Хаара чувствителен к повороту объекта, поэтому детектируемые ладони должны иметь одну и ту же ориентацию относительно камеры (± 15 градусов).

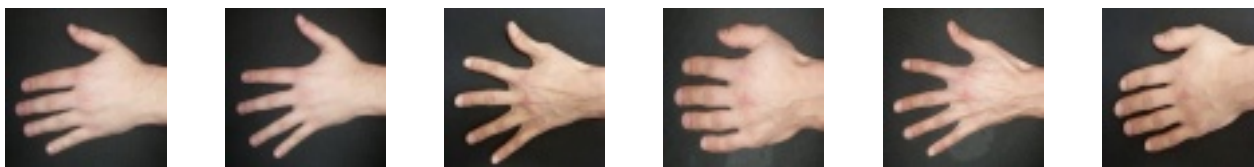


Рис. 5: Примеры изображений ладони на тёмном фоне.

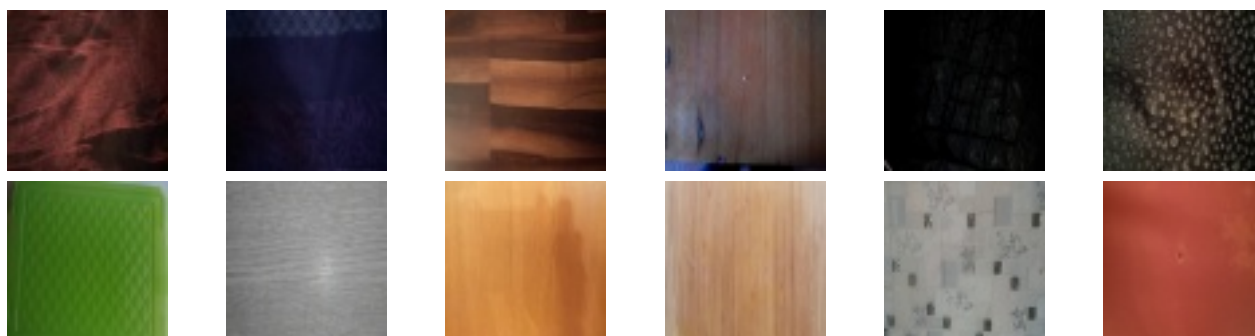


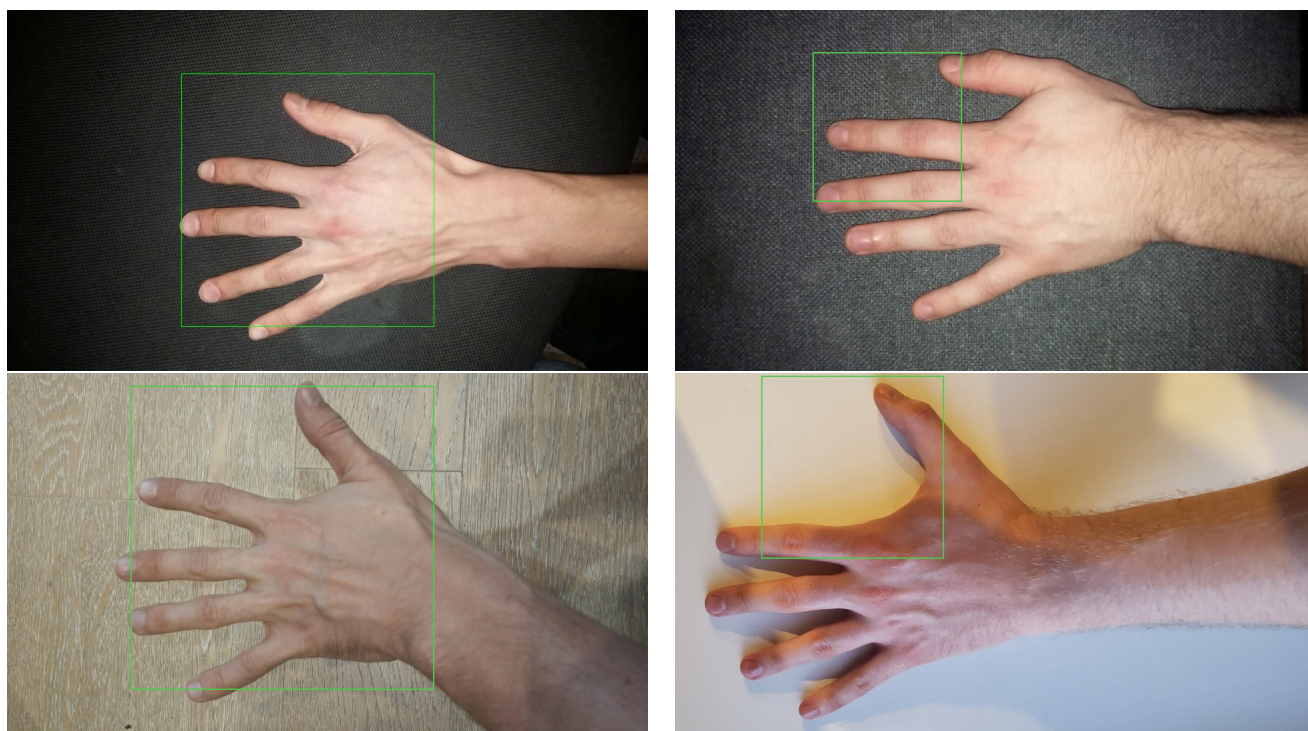
Рис. 6: Примеры изображений фона.

Номер эксперимента	Название базы	Цвет фона	Кол-во позитивных изображений	Ошибка 1-го рода, %	Кол-во негативных изображений	Ошибка 2-го рода, %	Ошибка общая, %
1	Хаар-1	тёмный	500	11.3	430	6.1	8.9
2	Хаар-1	светлый	500	18.1	430	4.3	11.7
3	Хаар-2	тёмный	3800	15.7	500	5.8	14.5
4	Хаар-2	светлый	4300	22.6	500	6.4	20.1

Таблица 3: Результаты детектирования ладони в кадре методом Виолы-Джонса.

Эксперименты показали, что алгоритм Виолы-Джонса далеко не всегда (табл. 3) верно обнаруживает область ладони как на тёмном, так и на светлом фоне — даже при соблюдении испытуемым нужной ориентации ладони (рис. 7). Вероятная причина в том, что сведение и разведение пальцев создаёт большое многообразие форм ладони даже в рамках одного человека. Более того, принудительное ориентирование ладони испытуемого вводит дополнительное ограничение в процесс идентификации, соблюдение которого зачастую является задачей более сложной, чем обеспечение наличия контрастного фона.

Исходя из этого, решено отказаться от использования текущей версии алгоритма детектирования области ладони на произвольном фоне на основе метода Виолы-Джонса.



(a) Корректное детектирование

(b) Некорректное детектирование

Рис. 7: Примеры работы алгоритма детектирования методом Виолы-Джонса.

Вычитание фона. Другой большой класс алгоритмов сегментации используется для выделения движущихся объектов на неподвижном фоне. Как правило, в таких методах строится модель фона (например, цветовое распределение), и проверяется, насколько каждый пиксель изображения соответствует этой модели (например, в смысле значения правдоподобия). Пиксели с низким соответствием считаются принадлежащими объекту, остальные — фону. Построение модели может осуществляться как непрерывно (с каждым новым поступившим кадром), так и только по первым кадрам видеопоследовательности. Большое число популярных алгоритмов вычитания фона содержится в библиотеке BGSLibrary [52].

Подобные методы предполагают, что камера неподвижна относительно фона. Но в реальных условиях эксплуатации разрабатываемой системы данное требование, как правило, не выполняется. В итоге, качество сегментации зачастую оказывается неудовлетворительным.

Для сценария съёмки «с руки» лучше подходят алгоритмы с динамической подстройкой модели фона. К сожалению, они имеют общий недостаток: объект, появившийся в кадре и недвигающийся длительное время, считается фоном спустя какое-то время. При использовании реализованной системы такие случаи весьма часты.

На рис. 8 приведены примеры работы алгоритмов вычитания фона при съёмке «с руки» на однородном контрастном фоне (движение камеры минимально). В левом столбце изображение с меньшим контрастом, в правом — с большим. Видно, что простая бинаризация по методу Оцу в указанных условиях работает либо так же, либо значительно лучше.

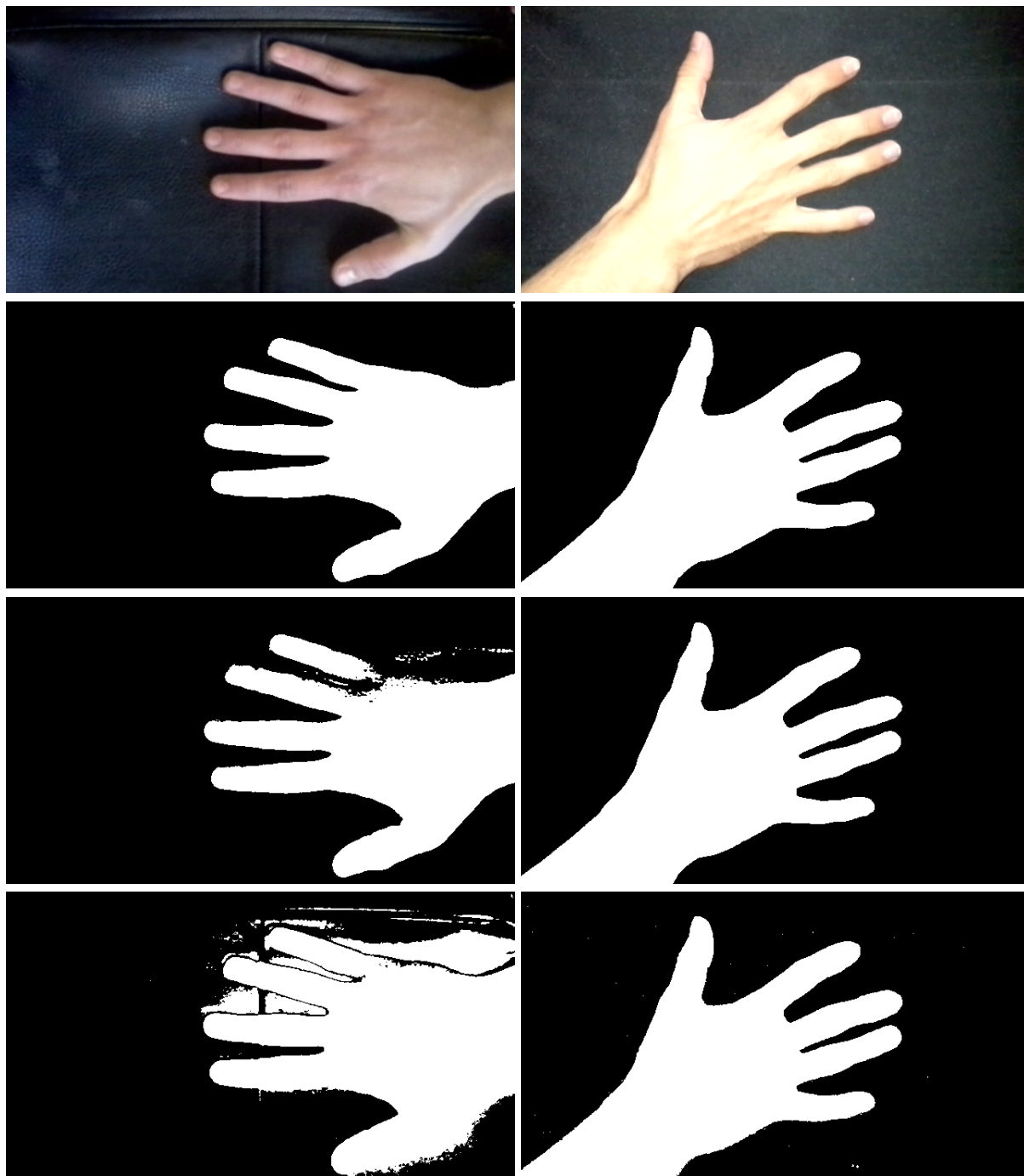


Рис. 8: Примеры работы алгоритмов сегментации из библиотеки BGSLibrary и бинаризации Оцу для двух ладоней. Первая строка — исходные изображения, вторая — бинаризация Оцу, третья — адаптивная медиана (BGSLibrary), четвёртая — смесь гауссиан (BGSLibrary).

В то же время стоит отметить, что съёмка с неподвижной камерой на одном и том же фоне является важным сценарием использования разрабатываемой системы, поэтому алгоритмы вычитания фона были добросовестно исследованы.

Выделение кожи в цветовом пространстве. Один из популярных способов сегментации основан на выделении пикселей кожи на снимке. Изображение переводится в какое-либо цветовое пространство (nRGB, YCbCr, CIE 1976 L*a*b*). Далее, как правило, выполняется отсечение по порогу [44]. Автором проведены тесты нескольких алгоритмов сегментации по цвету кожи, примеры их работы в сравнении с методом Оцу приведены на рис. 9. Анализ показал, что хотя сегментация по цвету кожи часто обеспечивает хороший результат, она является существенно менее устойчивой к изменению цвета источника освещения, чем метод Оцу.

Построение гауссовой модели цвета кожи. Для сегментации также может использоваться подход, основанный на гауссовой модели цвета кожи: по большому набору примеров кожи вычисляются параметры нормального распределения. Решение о принадлежности тестового пикселя к коже принимается на основании значения правдоподобия. Автором проведены эксперименты, в которых данные методы показали неудовлетворительные результаты, так как они чувствительны к освещению, а именно — к цвету источника. Это значит, что применение вспышки или смена помещения могут существенно влиять на качество сегментации.

Минимизация энергии и разрезы графов. Также существует обширный класс методов сегментации, основанных на минимизации энергии [53], [54]. Вводится некая функция энергии, определенная на изображении и зависящая от меток пикселей, — такая, что ее минимизация дает оптимальное разбиение пикселей. В работе [55] показано, что при определённых условиях задача минимизации энергии может быть сведена к задаче поиска минимального разреза графа.

Задача о минимальном разрезе применительно к сегментации формулируется следующим образом. Изображение представляется в виде графа. Вершины графа могут быть двух типов: вершины первого типа соответствуют пикселям изображения, вершин второго типа — две, они называются источником и стоком. Вершины первого типа соединены только со своими соседями (в смысле 8- или 4- связности), источником и стоком. Каждому ребру присвоен вес. Веса рёбер, соединяющих вершины первого типа между собой, определяют похожесть

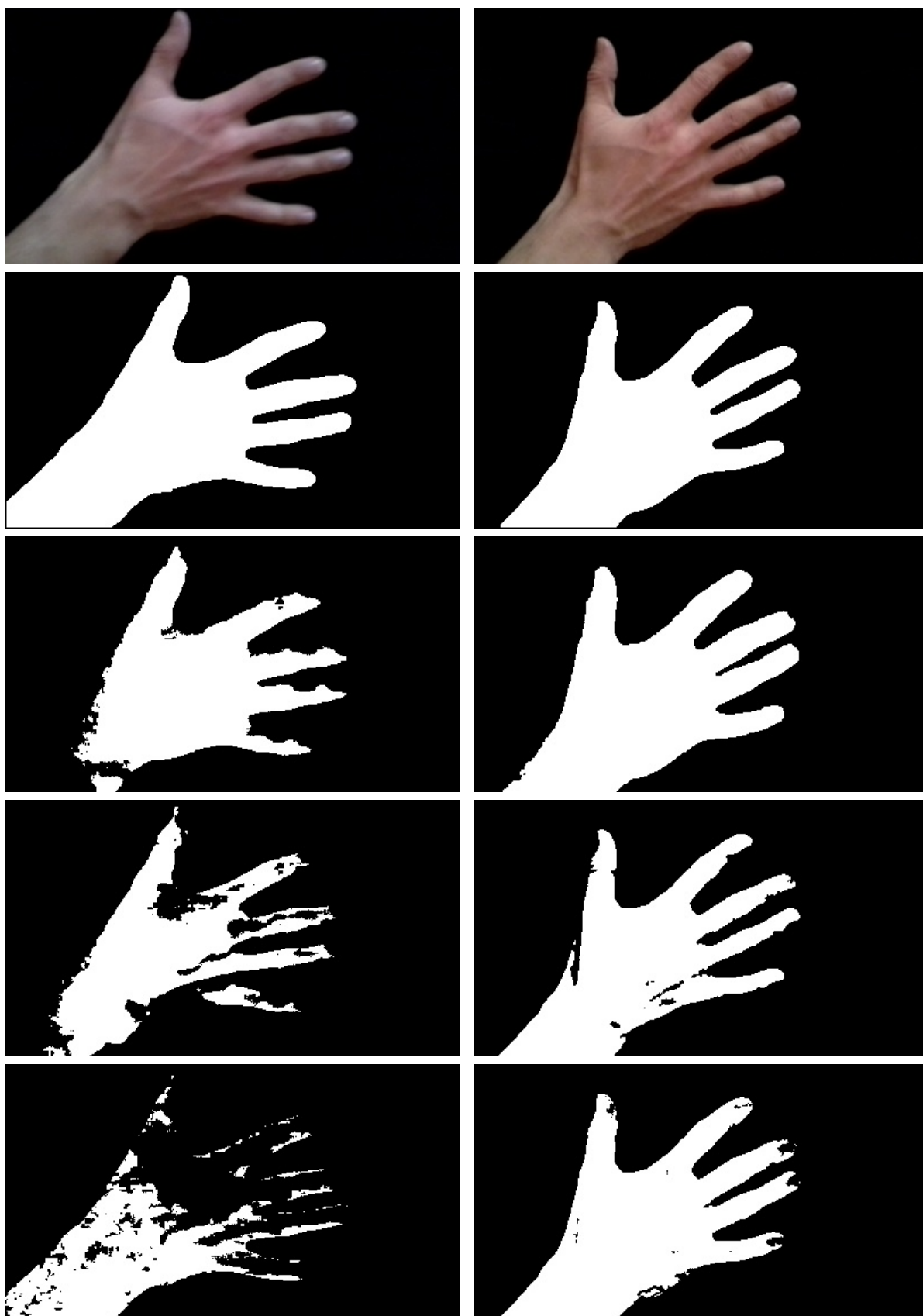


Рис. 9: Сегментация по цвету кожи и Оцу. Первая строка — исходные изображения, вторая — бинаризация Оцу, третья — Chai and Ngan, четвёртая — LUT, пятая — HSV.

пикселей по какой-либо метрике. Веса рёбер, соединяющие вершины первого типа с источником или стоком, задают степень соответствия пикселей моделям объекта и фона (правдоподобие). Такой граф называется **s-t** графом. Разрез графа представляет собой разбиение множества вершин на два подмножества. Стоимостью разреза называют суммарный вес рёбер, соединяющих вершины подмножеств разреза. Минимальный разрез имеет наименьшую стоимость. Пиксели изображения включённые в подмножество разреза, содержащее источник, помечаются как фон, во множество, содержащее сток — как объект.

Стоит отметить, что такие алгоритмы обладают большой вычислительной сложностью, быстро растущей с увеличением размера изображения. Например, для изображения размером 500×500 пикселей с жёсткими ограничениями, занимающими порядка 20% площади кадра, время вычисления сегментации на персональном компьютере средней мощности составляет порядка 0.6 сек, а в случае мобильного устройства время вычисления может быть существенно больше. Для решения данной проблемы разработаны различные методы, — например, интерактивная сегментация (пользователь помечает участки изображения как соответствующие объекту или фону) и алгоритмы кластеризации вершин графа (кластеризация с помощью водораздела).

Таким образом, алгоритмы, основанные на разрезах графов, не подходят для сегментации в режиме реального времени на мобильных устройствах из-за своей вычислительной сложности. Но вполне пригодны для использования на сервере (раздел 3.5).

Нейронные сети. В последнее время все большее внимание получают методы сегментации изображений на основе глубоких нейронных сетей. Как правило, в основе лежат сверточные сети. Подобные методы обеспечивают чрезвычайно высокое качество сегментации [56] в широком диапазоне условий (фон, освещение). Но они также обладают существенными недостатками, такими как высокая вычислительная сложность и необходимость большой обучающей выборки (сотни тысяч изображений). Как правило, для работы глубокой нейронной сети в реальном времени (время обработки одного кадра порядка 30 мс) требуется мощный графический ускоритель. Это ограничение не позволяет использовать методы этого класса для сегментации ладони как на мобильных устройствах (за исключением самых производительных моделей), так и на сервере.

Выводы. Сегментация ладони является первым и действительно важным этапом метода детектирования ладони на кадрах из видеопотока. От метода сегментации при этом требу-

ется не только производительность, но и точность. На основе проведенных экспериментов и изученной литературы сделаны следующие выводы.

Методы пороговой бинаризации работают достаточно быстро, но чувствительны к относительной яркости объекта и фона, т.е. применимы только в случаях, когда объект контрастирует с фоном. Автором предложен метод на основе алгоритма Виолы-Джонса с последующей пороговой бинаризацией, который ослабляет это ограничение. Подход показал хорошие результаты на модельных данных, но в реальных условиях эксплуатации системы его внедрение оказалось неоправданным.

Методы, основанные на вычитании фона, подходят для неподвижной относительно фона камеры. Тем не менее, даже в этом случае они могут испытывать проблемы с выделением ладони, если она продолжительное время находится в кадре без движения. Алгоритмы, использующие цвет кожи (на основе гауссовой модели или цветового пространства), чувствительны к освещению. Методы графовой сегментации вычислительно сложны, для приемлемой скорости работы и качества требуют априорных знаний об объекте и фоне. Глубокие нейронные сети требуют слишком больших вычислительных ресурсов, чтобы их можно было использовать на мобильных устройствах среднего класса.

Оптимальным вариантом остаются методы пороговой бинаризации, среди которых выбран метод Оцу — быстрый и дающий приемлемое качество. В качестве фона при этом рекомендуется использовать однородные поверхности, контрастирующие с ладонью. На практике это почти не ограничивает область применимости системы.

3.4.2 Получение геометрической информации о ладони

После того, как сегментация ладони выполнена (т.е. выделено множество пикселей, ей соответствующих), нужно определить, является ли выделенная область ладонью. И, если является, то корректна ли она в смысле условий, перечисленных в разделе 3.4. Для работы алгоритма проверки корректности необходимо извлечь дополнительную геометрическую информацию о ладони. Нас интересуют концы и основания пальцев, точки между пальцами, максимальный вписанный круг. Существует несколько методов получения такой информации.

Функция расстояния от точек контура до особой точки ладони. Один из популярных способов использует функцию расстояния до некой особой точки ладони (центроид, точка на запястье и т.д.), определенную на точках контура [37]. Примерный график такой функции показан на рис. 10.

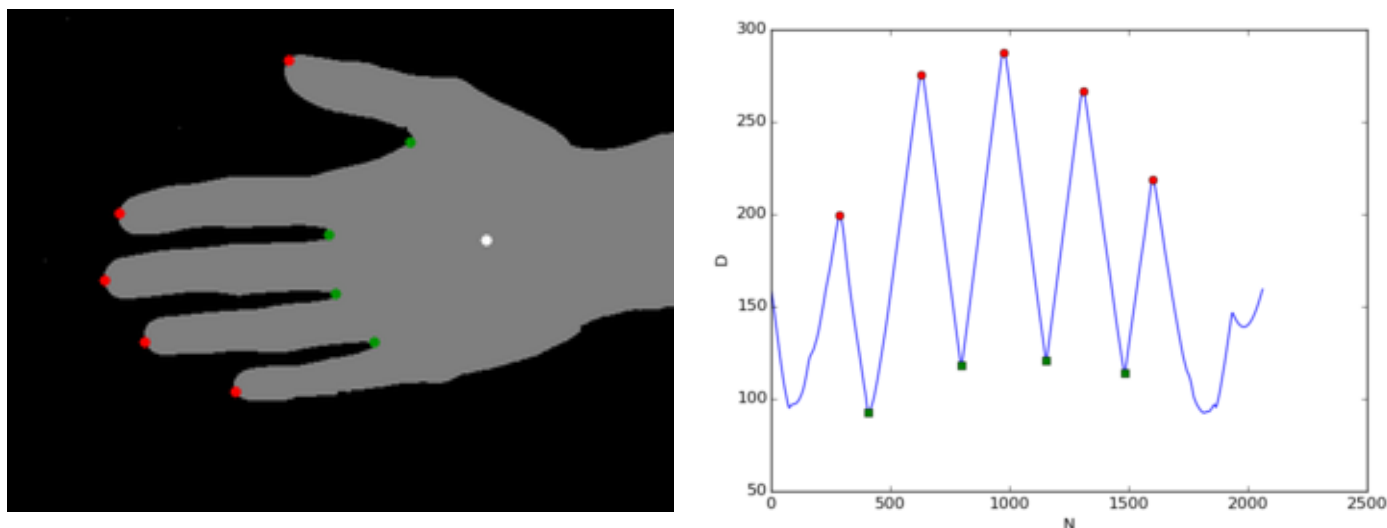


Рис. 10: Пример функции расстояния. На графике кружками отмечены концы пальцев, а квадратами — точки между пальцами. Расстояние вычисляется от точки, отмеченной белым кружком на левом изображении. На графике по оси X отложен номер точки на контуре ладони, а по оси Y — расстояние от данной точки до центра ладони

Этот способ позволяет уверенно находить точки концов пальцев и точки между пальцами. Однако с его помощью сложно отличить корректную ладонь от некорректной, так как информации, содержащейся в функции расстояния, для этого недостаточно. Более того, такой метод не способен отличить ладонь от любого другого объекта, имеющего нужное число выступов и впадин. В случае «неидеальной» сегментации (к примеру, если фон не матовый, присутствуют отражения) данный алгоритм также может давать ошибочные результаты. Примеры неверного определения пальцев приведены на рис. 11.



Рис. 11: Выделение трёх средних пальцев при «неидеальном контуре». Слева — метод, основанный на функции расстояния [37], справа — предложенный алгоритм, основанный на непрерывных скелетах. Нумерация пальцев начинается с большого пальца, имеющего индекс 0 (на рисунках не обозначен).

В связи с вышесказанным, в данной работе применяется подход к извлечению геометрической информации из формы ладони на основе циркулярных графов (непрерывных скелетов) [57]. *Скелетом* замкнутой фигуры называется множество центров её максимальных вписанных кругов. *Максимальным вписанным кругом* фигуры называется круг, не лежащий целиком ни в одном другом вписанном круге данной фигуры. Скелетное представление фигуры отражает её топологическую структуру, кроме того оно мало изменяется при деформациях фигуры. Построение скелета осуществляется следующим образом. Из бинарного изображения ладони извлекается контур, рассматриваемый как замкнутый многоугольник без самопересечений. На базе полученного многоугольника строится диаграмма Вороного, после чего выбирается подмножество рёбер диаграммы Вороного, формирующих скелет.

Данный метод позволяет сравнительно просто осуществить проверку ладони на корректность. Делается это с помощью выделения ветвей скелета, соответствующих пальцам ладони. Прежде, чем переходить к описанию процесса проверки, выделим основные свойства скелета ладони, которые будут использованы в дальнейшем.

- Скелет ладони делится на *внешний* (лежит вне контура ладони, связный граф) и *внутренний* (лежит внутри контура ладони, связный граф).
- Для каждой вершины скелета определен радиус соответствующего ей вписанного круга максимального радиуса. Это отображение множества вершин скелета во множество действительных чисел называется *радиальной функцией* и обозначается $R(v)$.
- Внутренний скелет ладони является деревом. Степень вершины может быть 1, 2 или 3.
- Внутренний скелет имеет вершину с *максимальным* радиусом вписанного круга R_{max} (среди *всех* вершин внутреннего скелета), данная вершина далее называется *корнем скелета*, а соответствующий круг — *максимальным кругом*.
- Внутренний скелет — ориентированный граф, рёбра направлены от корня скелета к вершинам степени 1.

Так как небольшие изменения контура могут приводить к существенным изменениям скелетного представления, необходимо проводить регуляризацию или *стрижку* построенного скелета [57]. Силуэтом скелета назовём объединение всех максимальных вписанных кругов с центрами в точках скелета. Силуэт скелета без регуляризации совпадает с исходной многоугольной фигурой. Если задан параметр ε , процесс стрижки можно описать так: последовательно удаляются вершины скелета степени 1, пока расстояние между силуэтом полученного скелета и исходной фигурой в метрике Хаусдорфа не превосходит ε . Далее в работе используется именно регуляризованный скелет, слово «регуляризованный» будем опускать.

Пример построения скелета приведен на рис. 12. Зеленым обозначен контур ладони, полученный по бинарному изображению. Жёлтым и розовым изображен внутренний скелет, бирюзовым и синим — внешний.

Используя в качестве входных данных скелетное представление ладони, автором предложен метод проверки ладони на корректность [8]. Осуществляется это с помощью отбора и анализа ветвей скелета, соответствующих пальцам ладони (раздел 3.4.3).

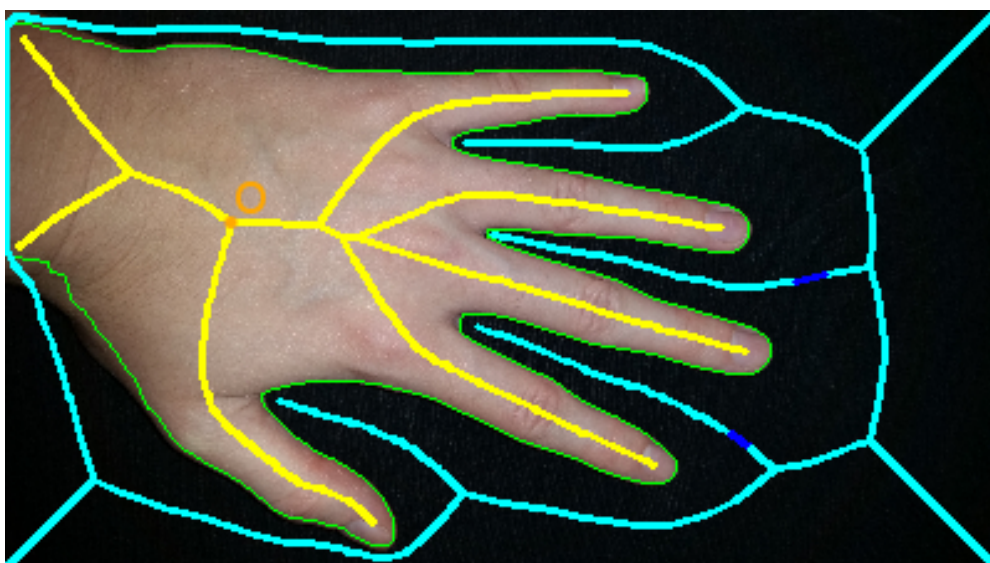
Выводы. Ранее были рассмотрены методы сегментации объекта относительно фона. В данном разделе описан подход к извлечению дополнительной геометрической информации о структуре объекта на основе циркулярных графов, который позволяет быстро и точно построить скелетное представление объекта-кандидата на роль ладони. Это необходимо для дальнейшего проведения семантической разметки ладони, т. е. выделения оснований и концов пальцев, точек между пальцами, определения максимального круга.



(a) Исходное изображение.



(b) Бинаризация по Оцу.



(c) Контур ладони и её скелет.

Рис. 12: Начальные этапы процедуры детектирования

3.4.3 Семантическая разметка ладони

Для последующего анализа необходимо по скелету ладони определить положение пальцев. Эта процедура состоит из нескольких последовательных шагов:

- выделение ветвей-кандидатов в пальцы;
- проверка радиальной функции, отсеивающая ветви-кандидаты, которые анатомически не могут соответствовать пальцам в силу своей «толщины»;
- выделение оснований и концов пальцев;
- базовые пороговые проверки расстояний, отсеивающие ветви-кандидаты, которые анатомически не могут являться пальцами в силу своей длины, расположения оснований и концов пальцев;
- исключение побочных ветвей-кандидатов, соответствующих одному пальцу;
- проверка троек, отсеивающая ветви, не лежащие в области пальцев (например, ветви из области запястья);
- определение большого пальца из оставшихся кандидатов;
- медианная проверка, отсеивающая изображения с близко сведёнными пальцами;
- поиск вершин внешнего скелета между пальцами.

Если на каком-то шаге не проходит проверка или остаётся меньше 5 ветвей-кандидатов в пальцы, то процедуры разметки досрочно завершается, и ладонь считается некорректной.

Выделение ветвей-кандидатов. В первую очередь, находятся ветви внутреннего скелета, заканчивающиеся в вершинах степени 1 и начинающихся в корне (обозначены жёлтым на рис. 12(с); корень скелета O выделен оранжевым).

Проверка радиальной функции. Так как радиальная функция пальцев имеет специфический вид, для каждой ветви-кандидата выполняется следующая проверка. Вычисляются значения нормализованной радиальной функции (или их линейная интерполяция) в 30 ключевых точках, равномерно распределённых вдоль ветви. Нормализованная радиальная функция определяется так: $R_n(v) = R(v)/R_{max}$. Таким образом, ее значения лежат на отрезке $[0, 1]$. По обучающей выборке рассчитываются границы, в которых могут лежать значения

функции в ключевых точках, совокупность этих границ называется граничным коридором. Граничный коридор общий для всех пальцев. Для каждой ветви-кандидата проверяется, лежит ли нормализованная радиальная функция в граничном коридоре. Это позволяет отсеять большое количество ложных ветвей, например, появившихся из-за искажений контура вследствие неточности сегментации.

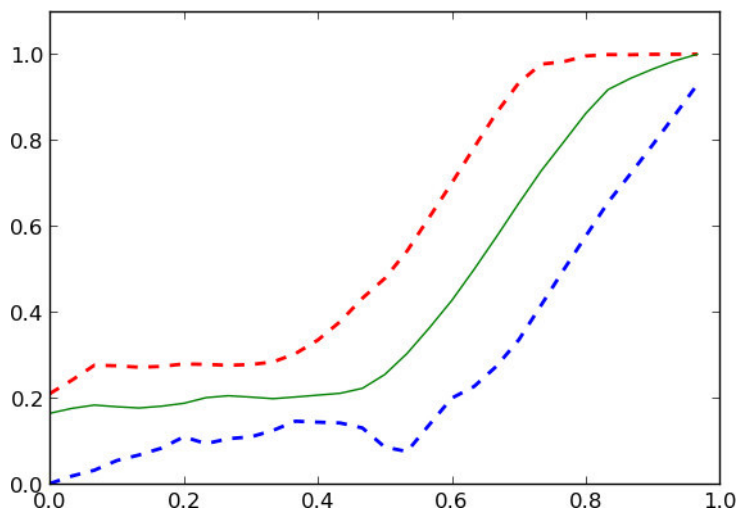


Рис. 13: Пример граничного коридора. По оси X нормализованное расстояние от вершины до корня, по Y — $R_n(v)$. Пунктирные линии — границы коридора, зелёная линия — нормализованная радиальная функция пальца.

Выделение оснований и концов пальцев. Для каждой ветви-кандидата находятся вершины, соответствующие *основанию* и *концу пальца*, аналогично [31]. Отрезок, соединяющий эти точки, будем называть *осью пальца*. Длину оси пальца, делённую на R_{max} , назовём *длиной пальца*. Участок ветви-кандидата между вершинами, определёнными в качестве основания и конца пальца, будем называть *пальцем-кандидатом*. Пример работы показан на рис. 14. Далее все значения, сравниваемые с порогами, полученными из обучающей выборки, считаем нормализованными на R_{max} .

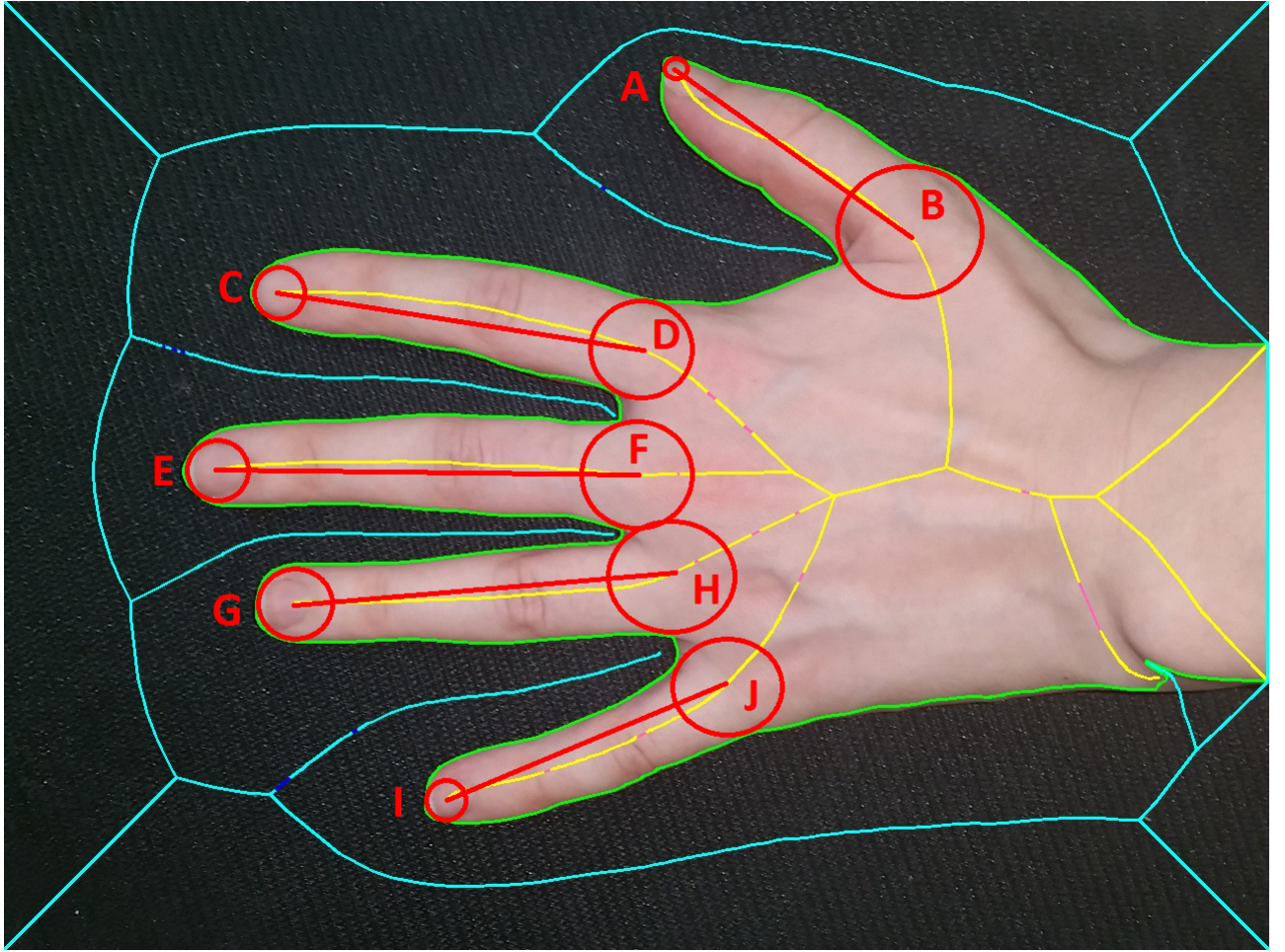


Рис. 14: Пример выделения оснований и концов пальцев. Вершины скелета B, D, F, H, J — основания пальцев, A, C, E, G, I — концы пальцев. Отрезки BA, DC, FE, HG, JI — оси соответствующих пальцев.

Базовые пороговые проверки расстояний. Для каждого пальца-кандидата проверяется, лежит ли его длина в неких фиксированных пределах $L_{min} < l < L_{max}$. Также проверяется, лежат ли основания и концы пальцев-кандидатов на определённом расстоянии от корня: $D_{min}^{orig} < d^{orig} < D_{max}^{orig}$ & $D_{min}^{top} < d^{top} < D_{max}^{top}$. Константы $L_{min}, L_{max}, D_{min}^{orig}, D_{max}^{orig}, D_{min}^{top}, D_{max}^{top}$ вычисляются по обучающей части выборки. Данные проверки удаляют ветви, которые не могут принадлежать пальцам по анатомическим причинам.

Исключение побочных ветвей-кандидатов. На этом этапе для основания каждого пальца-кандидата x ищется другой палец-кандидат y , чьё основание лежит внутри максимального вписанного круга, соответствующего основанию x . Если длина пальца-кандидата x меньше длины y , то ветвь пальца-кандидата y исключается из дальнейшего рассмотрения. Таким образом отсекаются побочные ветви-кандидаты, соответствующие одному и тому же пальцу (рис. 15).

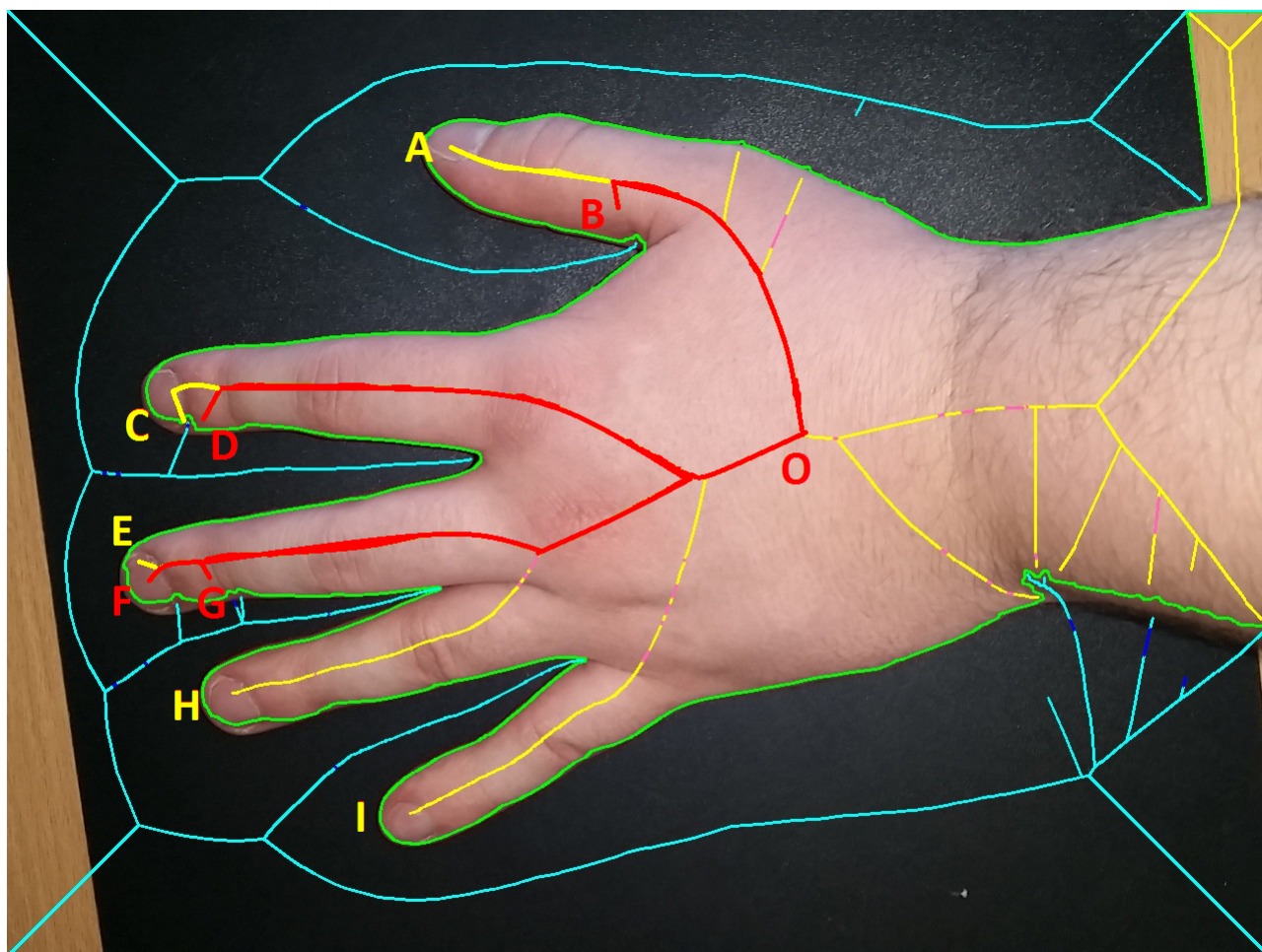


Рис. 15: Пример работы процедуры исключения побочных ветвей-кандидатов. Красным показаны ветви, которые будут удалены в ходе этой проверки: OB, OD, OF, OG. Ветви OA, OC, OE, OH, OI участвуют в следующих проверках. Прочие ветви на рисунке были отфильтрованы на предыдущих шагах.

Проверка троек. Пальцы-кандидаты упорядочиваются в порядке обхода контура и нумеруются, начиная с 0. Рассматриваются тройки последовательных пальцев-кандидатов (0-1-2, 1-2-3, 2-3-4, и т. д.), и для каждой из них строится отрезок, соединяющий середины осей первого и третьего пальцев из тройки (рис. 16). Если построенный отрезок пересекается с осью второго пальца из тройки, все три кандидата попадают в список для дальнейшей обработки. В противном случае не добавляется ни один.

Такая проверка позволяет отсеять пальцы-кандидаты, либо лежащие внутри ладони, либо имеющие малую длину, либо расположенные под большими углами к остальным кандидатам. Рис. 14 служит примером ладони, пальцы-кандидаты которой проходят проверку троек.

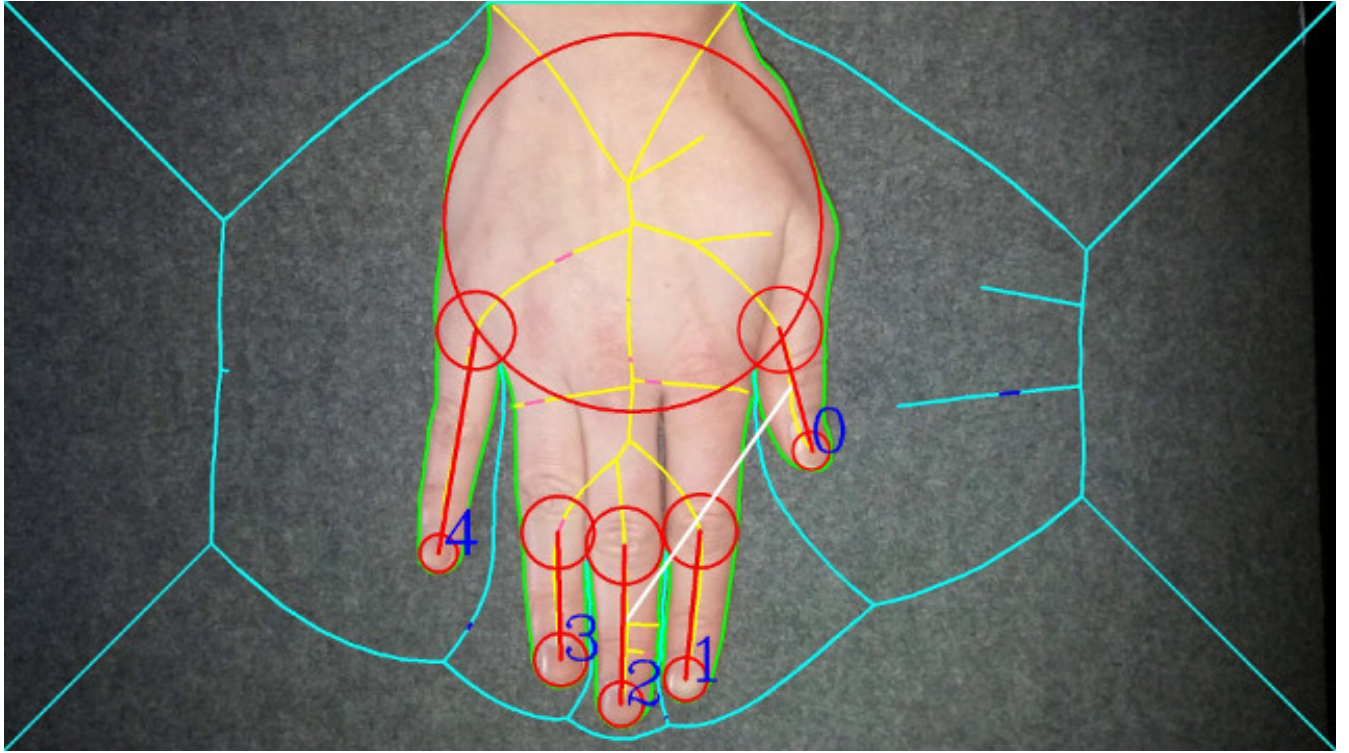


Рис. 16: 0-1-2 — пример тройки пальцев-кандидатов, не прошедших проверку троек: отрезок (обозначен белым), соединяющий середины осей большого (0) и среднего пальцев (2), не пересекает ось указательного (1).

Поиск большого пальца. Эксперименты показали, что приведённых выше проверок достаточно для отсеивания всех побочных ветвей. Таким образом, для корректной ладони на данном этапе должно остаться ровно 5 пальцев-кандидатов. Если это не так, ладонь считается некорректной.

В случае корректной ладони, для каждого пальца вычисляется евклидово расстояние между его основанием и основанием ближайшего пальца:

$$d_i = \min_{j \neq i} \{|orig_i - orig_j|\}, i, j \in \{0, \dots, 4\},$$

где $orig_k$ — координаты пикселя основания пальца k на изображении. Большой палец выбирается как кандидат с максимальным d_i . Пальцы переупорядочиваются в порядке обхода контура, начиная с большого, — его порядковый номер (индекс) принимается за 0.

Медианная проверка. Позволяет обнаружить случай «слипшихся» пальцев, т. е. когда пальцы (например, средний и безымянный) находятся слишком близко друг к другу, вследствие чего контур не проходит между ними, и скелет из-за этого меняет свою структуру. Вычисляются евклидовы расстояния:

$$\rho_i = \{|orig_i - root|\}, i \in \{1, \dots, 4\},$$

где $orig_i$ — координаты пикселя основания пальца i на изображении, $root$ — координаты пикселя, соответствующего корню скелета (центру максимального круга). Определим перестановку индексов (i_1, i_2, i_3, i_4) такую, что $\rho_{i_1} < \rho_{i_2} < \rho_{i_3} < \rho_{i_4}$. Выберем референсное значение следующим образом: $\rho_{ref} = \rho_{i_2}$. Далее, для каждого пальца вычисляются нормализованные отклонения от референсного значения:

$$\eta_i = |\rho_i - \rho_{ref}| / R_{max}, i \in \{1, \dots, 4\}$$

Если все расстояния η_i меньше некоторого порога, медианная проверка считается пройденной. В противном случае ладонь считается некорректной (рис. 17).

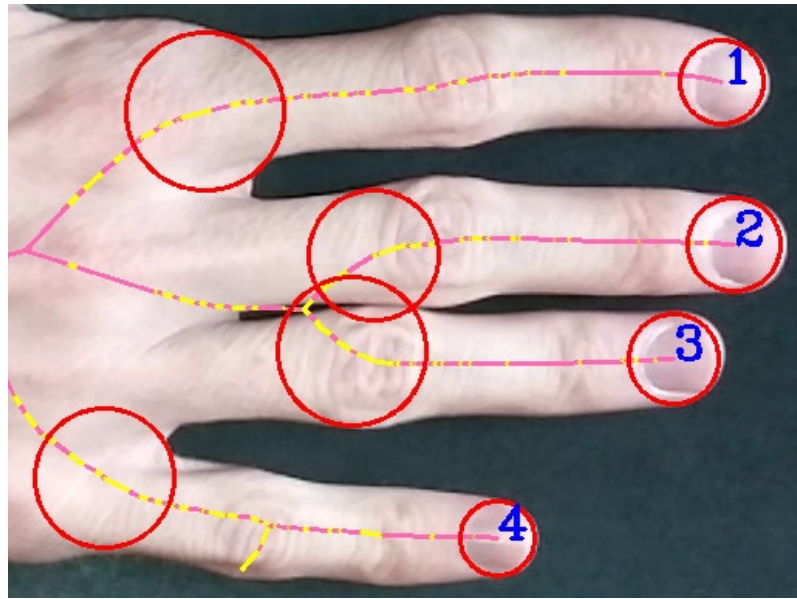


Рис. 17: Случай близко сведённых пальцев, который позволяет выявить медианная проверка.

Поиск вершин внешнего скелета между пальцами. В случае корректной ладони, выполняется поиск вершин внешнего скелета (раздел 3.4.2) степени 1, которые лежат между парами пальцев (0 и 1, 1 и 2, 2 и 3, 3 и 4). Если таких вершин для данной пары пальцев несколько — выбирается ближайшая к линии, соединяющей основания пальцев пары.

3.4.4 Определение расфокусированных кадров

Так как расстояние между ладонью и камерой может изменяться, возникает необходимость в постоянной фокусировке. Однако системы автофокуса в современных мобильных устройствах не всегда обеспечивают правильную фокусировку. Назовём *сфокусированным* такой кадр, где ладонь целиком находится в фокусе, т.е. её контур и пространство внутри него не смазаны. Примеры кадров с правильным и неправильным фокусами приведены на рис. 18.



(a) Правильная.



(b) Неправильная.

Рис. 18: Примеры фокусировки.

Стоит заметить, что приведенная выше процедура детектирования, основанная на форме ладони, не способна отличить сфокусированный кадр от расфокусированного. При этом для извлечения текстурных признаков пригодны только сфокусированные кадры. Таким образом, возникает необходимость в определении правильности фокусировки перед отправкой кадра на сервер.

Оптическое размытие из-за неправильной фокусировки может быть смоделировано с помощью фильтра Гаусса. Считаем, что расфокусированное изображение получено из оригинала применением к нему фильтра. На рис. 19 приведены примеры обработки сфокусированного и расфокусированного кадров.

Видно, что для сфокусированного кадра на бинаризованном изображении-разности (правый нижний столбец) очень много ненулевых пикселей, тогда как для расфокусированного изображения их существенно меньше. Причина в том, что Гауссово размытие действует как фильтр низких частот, отфильтровывая высокочастотные компоненты изображения. Поэтому при применении данного фильтра к уже размытому изображению, содержащему преимущественно низкие частоты, изменения будут минимальны.

Можно сформулировать основную идею так: если исходное изображение мало отличается от результата размытия по Гауссу, то кадр, вероятнее всего, расфокусирован. Однако необходимо учитывать следующие особенности. Фон с неоднородной текстурой может вносить существенные искажения в картину разности (нижнее изображение второго столбца на рис. 19), поэтому анализ проводится только для части кадра, лежащей внутри контура ладони и относящейся к пальцам (рис. 20).

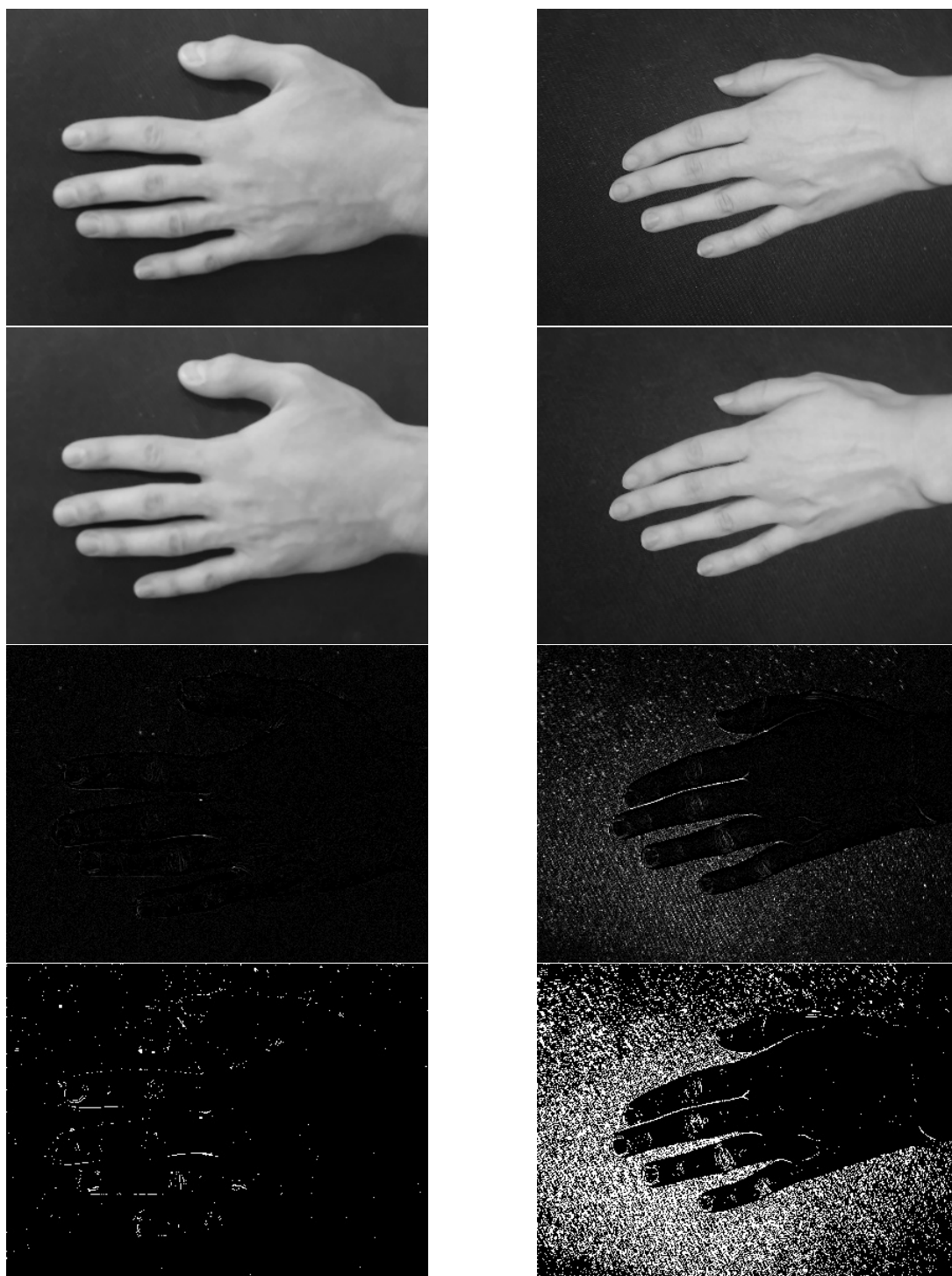


Рис. 19: Определение качества фокусировки. Первый столбец — расфокусированный кадр, второй — с правильным фокусом. Первая строка — исходное изображение, вторая — после применения размытия по Гауссу, третья — разность между исходным и размытым изображениями, четвёртая — пороговая бинаризация полученной разности.

Кроме того замечено, что вне зависимости от фокуса на изображении-разности много пикселей, имеющих минимально возможное ненулевое значение, которые представляют собой



Рис. 20: Выделение области интереса для определения фокуса. Область интереса находится внутри контура.

шум. Чтобы минимизировать его влияние, было решено отбирать пиксели, имеющие значение больше эмпирически подобранного порога.

В качестве количественной меры расфокусированности используется следующая величина. Пусть $S_{diff} = \{p_1, p_2, \dots, p_n\}$ — множество пикселей изображения разности внутри выделенной области интереса, где p_i — значение яркости пикселя, а n — количество пикселей в области интереса. Тогда мера расфокусированности будет вычисляться так:

$$D = \frac{|\{p_i | p_i > T, p_i \in S_{diff}\}|}{|\{p_i | p_i > 0, p_i \in S_{diff}\}|},$$

где T — эмпирически подобранное значение. Кадр признаётся расфокусированным, если D меньше заданного порога. Варьируя порог, можно изменять «уровень требований к качеству» фокусировки. Расфокусированные кадры исключаются из дальнейшего рассмотрения (не участвуют в распознавании).

3.4.5 Вычисление зёрен

Если ладонь корректна, строится ограничивающий прямоугольник, включающий в себя ладонь и пальцы (рис. 21). Это позволяет уменьшить вычислительную сложность последующего анализа, так как процедура распознавания оперирует только с областью внутри прямоугольника.

Для упрощения вычислений и улучшения качества при последующей «точной» сегментации необходимо задать т.н. жёсткие ограничения — два множества пикселей с фиксированными метками (объекта или фона), которые не изменяются в процессе работы алгоритма сегментации. Пиксели из указанных множеств также называют *зёрнами* (объекта или фона соответственно).

Зёрна могут быть получены на основе скелетного представления и построенной на его базе семантической разметки, описанных выше. Автором предложено использовать в каче-

стве зёрен объединение уменьшенных максимальных вписанных кругов для подмножества вершин скелета (рис. 22). Круги уменьшаются для того, чтобы даже в случае неточной сегментации по Оцу внутренние пиксели лежали внутри ладони, а внешние — принадлежали фону. Из внутренних вершин скелета в данное подмножество включаются те, которые лежат внутри ограничивающего прямоугольника — это зёрна объекта. Среди внешних вершин выбираются те, которые принадлежат ветвям, расположенным между пальцами и лежащим внутри прямоугольника, — это зёрна фона.

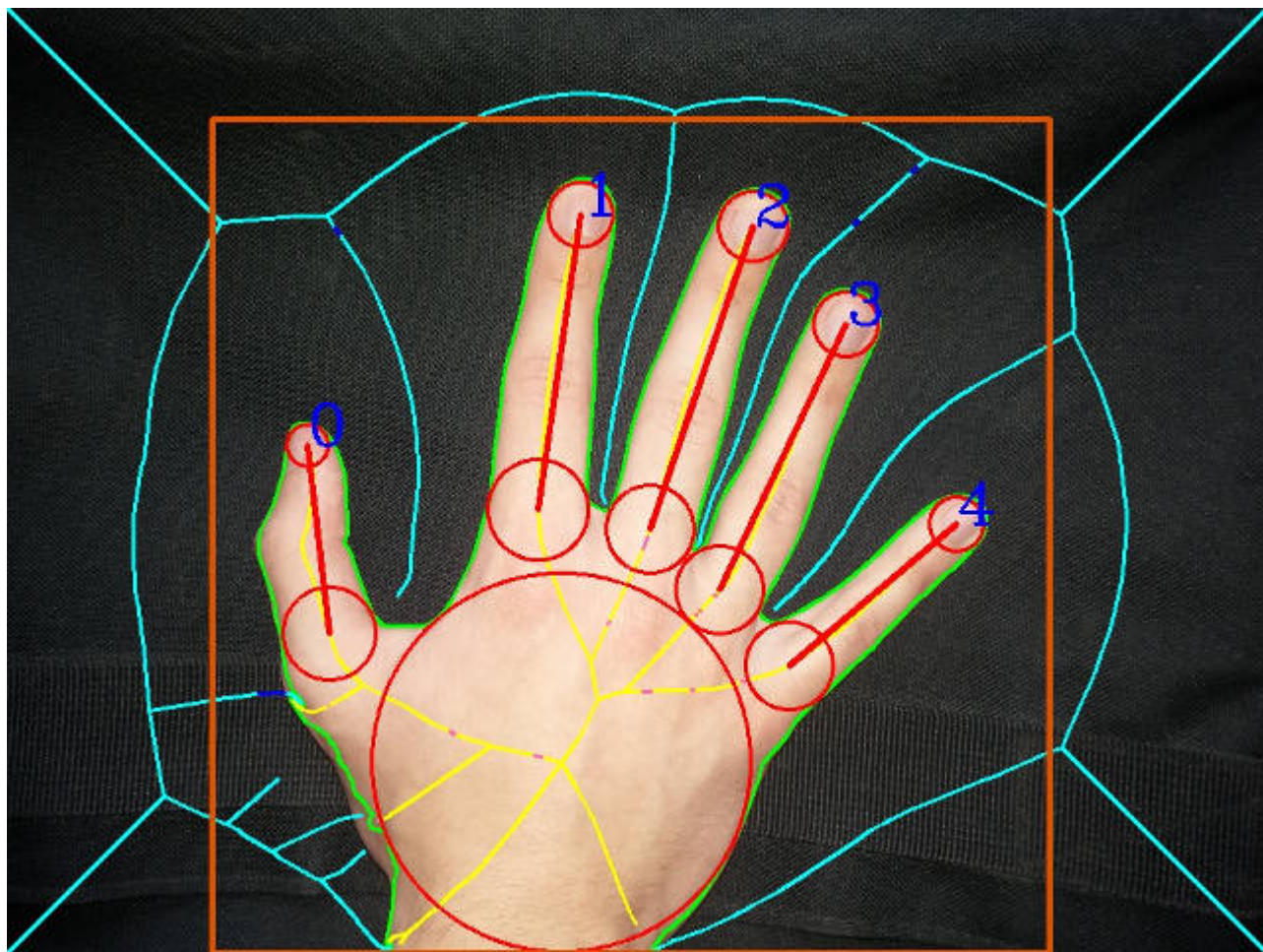


Рис. 21: Результат семантической разметки ладони. Оранжевым обозначен ограничивающий ладонь прямоугольник.

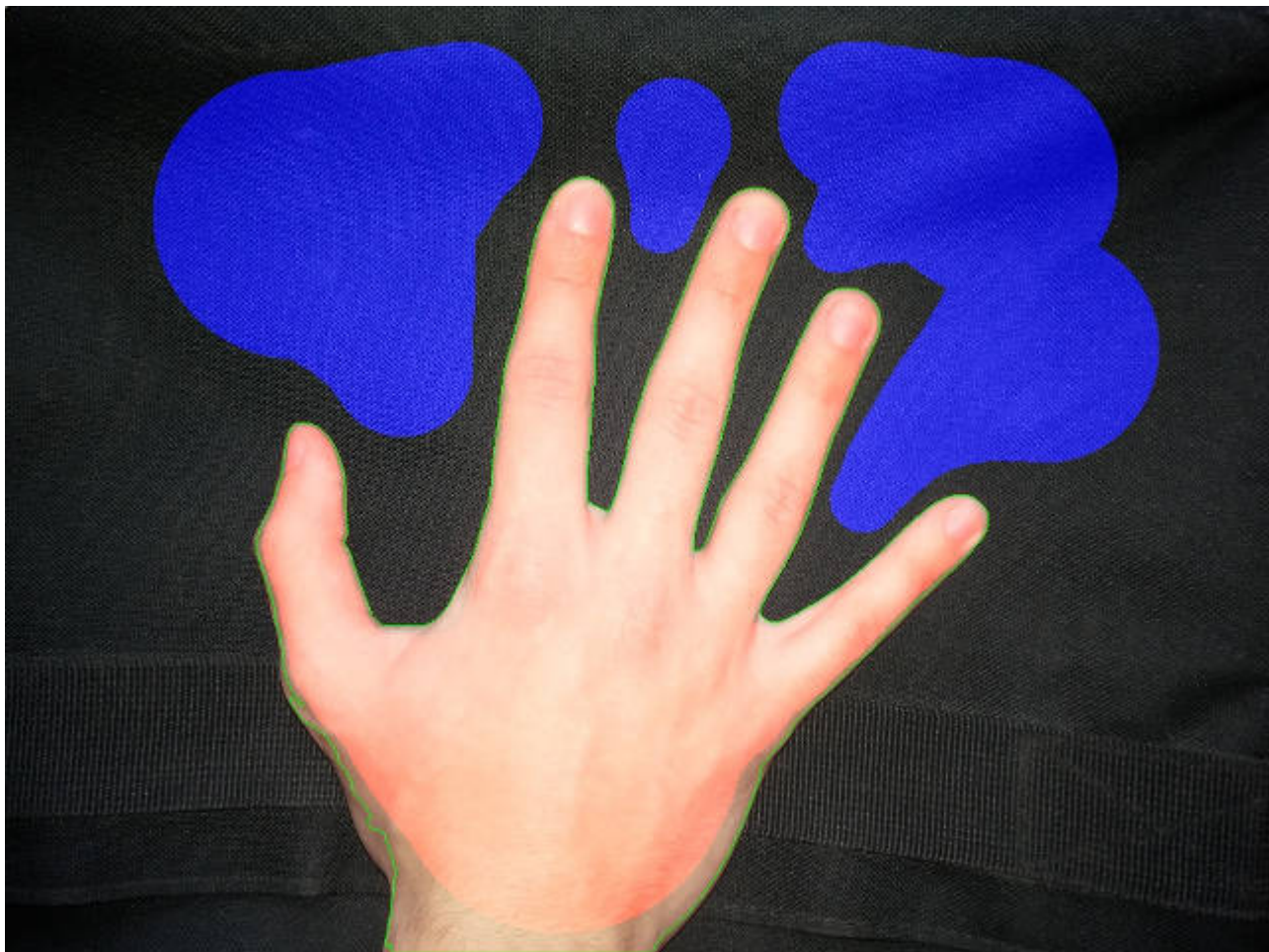


Рис. 22: Зёрна объекта обозначены красным цветом, зёрна фона — синим.

3.4.6 Выводы к разделу 3.4

Предложено разделить задачи детектирования и распознавания ладони, при этом задача детектирования разбивается на этапы сегментации и валидации. Для решения задачи первичной «быстрой» сегментации проведён анализ методов, часть из которых была протестирована. В качестве наиболее подходящего метода выбрана бинаризация по Оцу.

В рамках задачи валидации разработан метод семантической разметки ладони на основе непрерывных скелетов. Описан механизм поиска расфокусированных кадров, который существенно улучшает характеристики системы распознавания при реальной эксплуатации. Используя выход алгоритмов построения скелетного представления и полученной на его базе семантической разметки, представлен способ задания жёстких ограничений, необходимых для последующей «точной» сегментации.

Таким образом, зёрна фона и ладони, а также изображение ладони внутри ограничивающего прямоугольника передаются в модуль распознавания для дальнейшего анализа.

3.5 Метод распознавания

После того, как зёрна ладони и её изображение внутри ограничивающего прямоугольника переданы в модуль распознавания, можно приступать к генерации признакового описания. Первым шагом необходимо провести «точную» сегментацию (раздел 3.5.1).

3.5.1 Сегментация

Для «точной» сегментации применяется алгоритм минимизации энергии на основе разрезов графов OneCut [58] с использованием полученных ранее зёрен (раздел 3.4.5). Результатом, как правило, является очень качественная сегментация ладони. Полученное бинарное изображение (пиксели со значением 0 соответствуют фону, 1 — ладони) передается в модуль извлечения признаков формы, а также используется для выделения области интереса при извлечении текстурных признаков.

Подобные алгоритмы сегментации имеют высокую вычислительную сложность, быстро растущую с размером изображения, — так как они оперируют графами, число вершин в которых равно числу пикселей изображения. Автором предпринята попытка облегчить вычисления, используя в качестве узлов графа т. н. *суперпиксели*. Суперпиксель — связанное множество пикселей изображения, однородных в некотором смысле (например, по цвету). Это позволяет оперировать с суперпикселями принципиально так же, как и с исходными пикселями. Однако количество суперпикселей зачастую значительно меньше количества пикселей. К примеру, для изображения с разрешением 1980×1020 пикселей нормальное количество суперпикселей — порядка 200. При разбиении изображения на суперпиксели выполняется важное условие: границы между суперпикселями проходят вдоль границ объектов на исходном изображении. Примеры суперпикселей приведены на рис. 23.

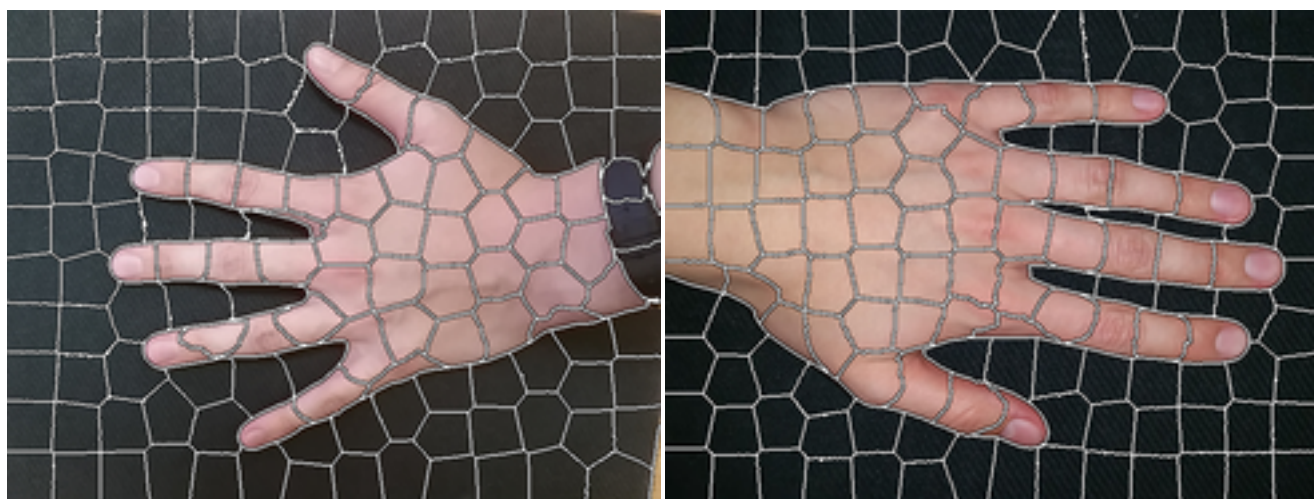


Рис. 23: Примеры суперпикселей

Проведённые эксперименты показали, что в данной задаче построение суперпикселей работает медленнее, чем вся сегментация с помощью разрезов графов. Выбранный алгоритм SLIC (Simple Linear Iterative Clustering) [59] выполняет итеративную кластеризацию пикселей в пятимерном пространстве: L , a , b из цветового пространства CIELAB и координаты пикселя x , y . Желаемое количество суперпикселей полагалось равным 100. Параметр, отвечающий за компактность, был равен 20 (чем он больше, тем больший вес имеет пространственное расстояние между пикселями и тем более компактными получаются суперпиксели). А именно, для изображений размером 400×300 пикселей время вычисления суперпикселей в среднем составило 0.37 с., в то время как для тех же изображений вычисление графовой сегментации с использованием зёрен в среднем заняло порядка 0.25 с. Причина этого в том, что зёрна, подаваемые на вход графовой сегментации, занимают сравнительно большую площадь изображения. Поэтому вычисление максимального потока в пиксельном графе выполняется существенно быстрее, чем обычно (при небольших областях жёстких ограничений). Таким образом, для данной задачи использование суперпикселей в варианте выше не оправдано.

Выводы. Для «точной» сегментации выбран метод на основе разрезов графов OneCut, показавший в ходе экспериментов оптимальное сочетание качества сегментации и скорости работы. Высокая эффективность метода сегментации обеспечена благодаря предложенному способу получения «больших» зёрен объекта и фона. Применение алгоритма построения суперпикселей SLIC не даёт преимуществ в скорости работы, поэтому его использование нецелесообразно.

3.5.2 Генерация признаков формы

Выходом алгоритма графовой сегментации (предыдущий раздел 3.5.1) является, как правило, точная маска ладони с гладким контуром. Это, в свою очередь, способствует точному извлечению признаков на основе формы. Актуальной задачей является подбор таких признаков, которые будут достаточно устойчивы и репрезентативны, чтобы применить их в задаче классификации пользователей по изображению ладони.

Генерация признаков формы происходит следующим образом. По бинарному изображению (маске ладони выше) строится скелет и проводится семантическая разметка ладони (рис. 24). Для этого используются методы, описанные ранее в разделах 3.4.2 и 3.4.3. После вычисляются следующие признаки:

1. Спектр (распределение медиальной ширины).

2. Ломаная, последовательно соединяющая вершины пальцев и точки между ними. В качестве признака используются нормализованные длины ее звеньев.
3. Кривизны пальцев. Для каждого пальца вычисляется максимальное расстояние между точками скелета и осью пальца по обе стороны от последней.

Остановимся подробнее на каждом из признаков.

Распределение медиальной ширины. Пусть имеется замкнутая фигура D и её скелет S . Для каждой точки $s \in S$ существует единственный максимальный вписанный круг $C(s) \subset D$ с центром в точке s и радиусом $r(s)$. Для каждой точки $s \in S$ определим множество $P(s) = \{p : p \in \partial D \cap C(s)\}$ ближайших к s граничных точек фигуры. Для каждой точки скелета $s \in S$ определено множество отрезков $\{sp : p \in P(s)\}$, соединяющих её с ближайшими точками границы фигуры. Эти отрезки назовём *спицами* (рис. 25).

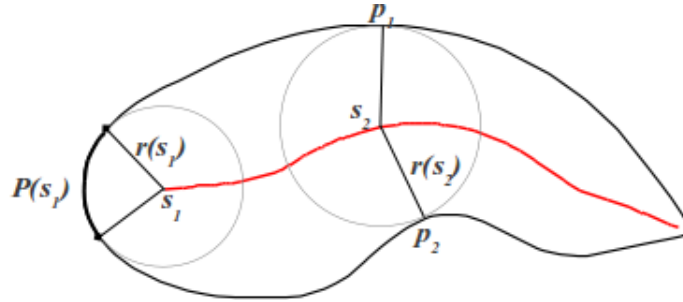


Рис. 25: Примеры спиц.

Для множества спиц фигуры выполняются следующие утверждения [60]:

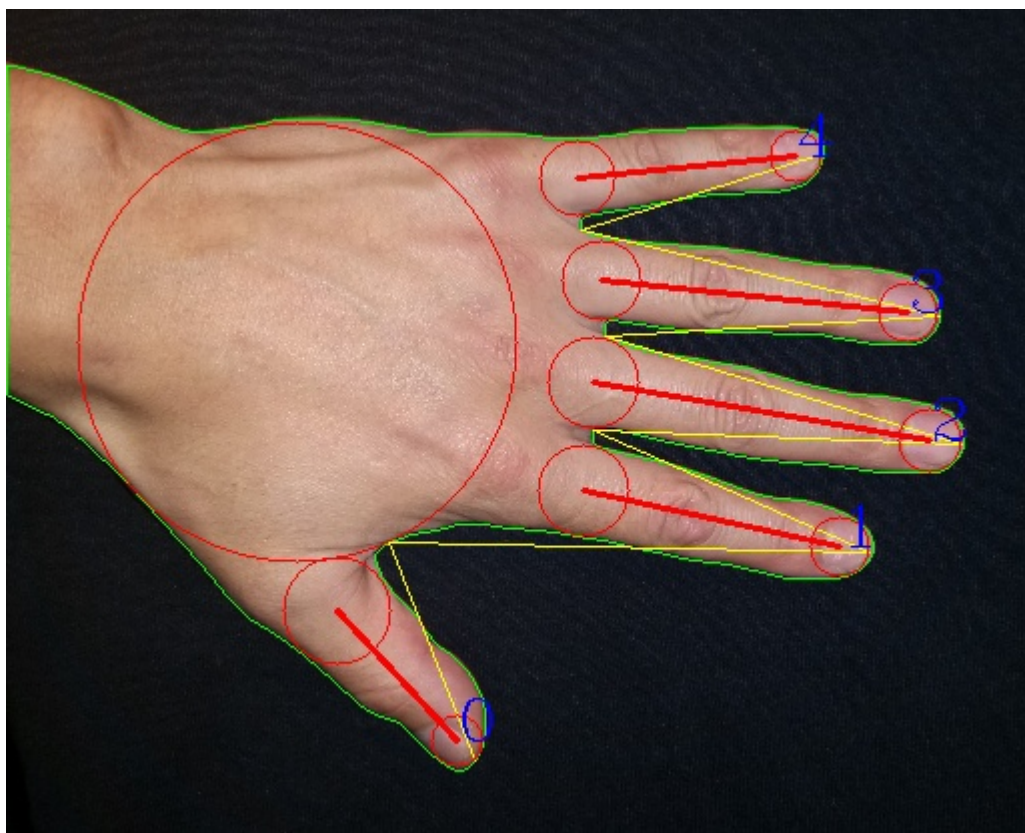
- две спицы могут пересекаться лишь в концевых своих точках, т. е. либо в точках скелета, либо в точках границы;
- через каждую точку фигуры проходит хотя бы одна спица;
- для каждой внутренней точки фигуры, не принадлежащей скелету, существует единственная спица, проходящая через эту точку;
- для каждой граничной точки фигуры, в которой имеется касательная к границе, инцидентная спица является единственной.

Медиальной шириной фигуры D в точке $x \in D$ назовём длину спицы, инцидентной данной точке. Обозначим эту функцию $M(x)$. Обозначим множество точек границы фигуры $L = \partial D$, имеющих медиальную ширину меньшую некоторого значения r ,

$$L(r) = \{x : x \in L, M(x) \leq r\}.$$



(а) Графовая сегментация методом OneCut.



(b) Семантическая разметка ладони.

Рис. 24: Выделение пальцев на основе точной сегментации

Тогда распределением медиальной ширины на точках границы фигуры называется функция

$$FL(r) = |L(r)|/|L|.$$

Назовем *медиальным спектром* фигуры функцию

$$f(r) = \frac{\partial FL(r)}{\partial r}.$$

Спектр имеет ограниченную область определения, а также представляет собой функцию распределения, т. е.

$$\int_0^{+\infty} f(r) dr = 1.$$

Спектр как характеристика формы является устойчивым, т. е. малое изменение формы фигуры влечет малое изменение спектра. В качестве признака используется набор значений спектра в натуральных точках числовой оси. Важную роль при сравнении спектров друг с другом играет выбор метрики. Так как спектр представляет собой одномерную нормированную гистограмму, остановимся подробнее на способах сравнения гистограмм.

Метрики для сравнения гистограмм. Пусть $H^{(1)}, H^{(2)}$ — дискретные спектры, N, M — количество элементов в них, $K = \min\{N, M\}$. Самая простая метрика — сумма поэлементных разностей:

$$d_{L_1} = \sum_{i=1}^K |H^{(1)}(i) - H^{(2)}(i)|.$$

Её достоинством является простота интерпретации. Чем меньше значение, тем больше сходство.

Следующая метрика — сумма пересечений гистограмм:

$$d_{intersection} = \sum_{i=1}^K \min\{H^{(1)}(i), H^{(2)}(i)\}.$$

Если обе гистограммы нормированы на единицу, то значение метрики 1 означает полное сходство, а 0 — полное различие.

Следующий метод сравнения предложен Карлом Пирсоном и носит названия метрики хи-квадрат:

$$d_{\chi^2} = \sum_{\substack{i=1 \\ H^{(1)}(i)+H^{(2)}(i) \neq 0}}^K \frac{(H^{(1)}(i) - H^{(2)}(i))^2}{H^{(1)}(i) + H^{(2)}(i)}.$$

Здесь чем меньше значение, тем больше сходство, 0 означает идеальное совпадение, максимальное значение не ограничено.

Ещё одна метрика называется расстоянием Бхаттачария:

$$d_B = \sqrt{1 - \sum_{i=1}^K \frac{\sqrt{H^{(1)}(i)H^{(2)}(i)}}{\sqrt{\sum_j H^{(1)}(j) \sum_j H^{(2)}(j)}}}.$$

Чем меньше значение, тем больше сходство, 0 означает полное совпадение, а 1 — полное несовпадение.

Важным классом метрик являются так называемые EMD (Earth Mover's Distance) метрики. Эти метрики отражают минимальную «стоимость преобразования» одной гистограммы в другую. Пусть имеется некоторая базовая метрика $d_E(i, j), i \in \{1, \dots, N\}, j \in \{1, \dots, M\}$. Значение i -го элемента гистограммы $H^{(1)}$ равно $h_i^{(1)}$, аналогично значение j -го элемента $H^{(2)}$ равно $h_j^{(2)}$. Тогда EMD-метрика определяется следующим образом:

$$\begin{aligned} EMD(H^{(1)}, H^{(2)}) &= \min_{h_{ij}} \sum_{i=1, \dots, N} \sum_{j=1, \dots, M} h_{ij} d_E(i, j), \\ \sum_{i=1, \dots, N} h_i^{(1)} &= 1, \quad \sum_{j=1, \dots, M} h_j^{(2)} = 1, \quad \sum_{i=1, \dots, N} \sum_{j=1, \dots, M} h_{ij} = 1, \\ \forall i, j: h_{ij} &\geq 0, h_i^{(1)} = \sum_{j=1, \dots, M} h_{ij}, h_j^{(2)} = \sum_{i=1, \dots, N} h_{ij}. \end{aligned}$$

Стоит отметить, что в общем случае вычисление данного расстояния требует решения задачи линейного программирования. Однако если гистограммы являются дискретными, одномерными и нормированными на единицу, как в данном случае, вычисления существенно упрощаются. Если в качестве базовой метрики выбрать L^1 -метрику, то расстояние EMD_{L^1} вычисляется следующим образом:

$$EMD_{L^1}(H^{(1)}, H^{(2)}) = \int_{-\infty}^{+\infty} |F(x) - G(x)| dx.$$

Здесь $F(x), G(x)$ — функции распределения для $H^{(1)}(x_i)$ и $H^{(2)}(y_j)$:

$$F(x) = \int_{-\infty}^x H^{(1)}(t) dt, \quad G(x) = \int_{-\infty}^x H^{(2)}(t) dt.$$

Таким образом, возможно вычисление данного расстояния за линейное относительно размеров гистограмм время.

Для сравнения спектров двух изображений с учётом результатов работы [30] автором выбрана EMD_{L^1} -метрика.

Ломаная. Соединяет особые точки контура ладони: 5 концов пальцев и 4 точки на контуре между пальцами (рис. 26). Выделение этих точек несложно делается на основе семантической

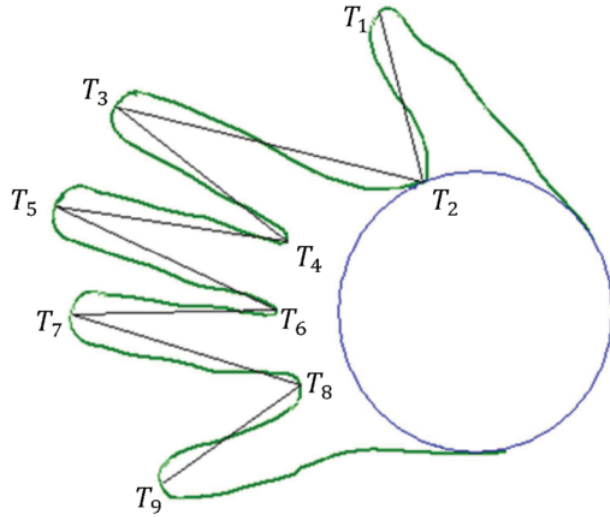


Рис. 26: Ломаная — признак на основе формы.

разметки ладони (раздел 3.4.3) и подробно описано в [61]. Для единообразия со статьёй будем считать, что пальцы нумеруются последовательно, начиная с 1.

Пусть T_1, T_2, \dots, T_9 — последовательные вершины ломаной, и $\gamma_i = |T_i T_{i+1}|, i = 1, \dots, 8$ длины соответствующих сегментов, $\gamma = \sum_{i=1}^8 \gamma_i$ — общая длина. Вектор признаков тогда определён как $V = \left(\frac{\gamma_1}{\gamma}, \frac{\gamma_2}{\gamma}, \dots, \frac{\gamma_8}{\gamma} \right)$. Мерой различия между двумя векторами признаков $V^{(1)}$ и $V^{(2)}$ является евклидово расстояние $\rho_\varepsilon(V^{(1)}, V^{(2)})$.

Кривизна пальцев. Мера кривизны пальцев строится следующим образом (рис. 27).

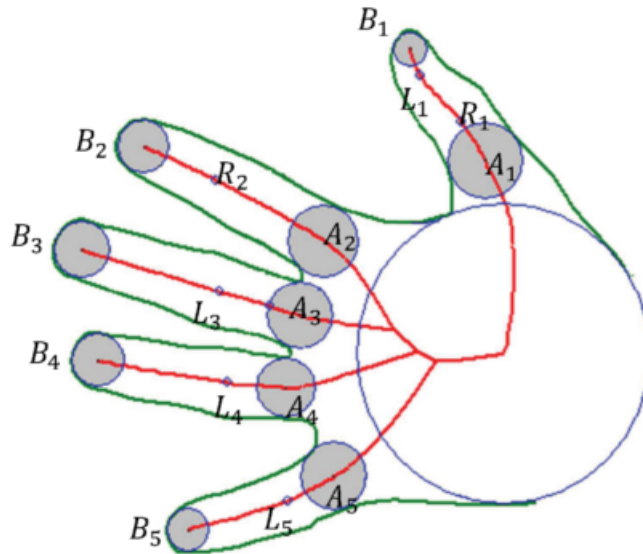


Рис. 27: Кривизна пальцев — признак на основе формы.

Продолжим использовать последовательную нумерацию пальцев, начиная с 1. Для каждого пальца с номером $i = 1, \dots, 5$ на непрерывном скелете находятся основание A_i и конец B_i (раздел 3.4.3). Затем, для ветви скелета A_iB_i находятся точки, лежащие на максимальном удалении от отрезка A_iB_i справа (точка R_i) и слева (точка L_i). Пусть δ_i — расстояние между A_iB_i и R_i , а ε_i — расстояние между A_iB_i и L_i и $\eta_i = |A_iB_i|$ — длина отрезка. Если правее (левее) отрезка A_iB_i нет точек, то соответствующие расстояния δ_i (ε_i) принимают нулевые значения. Тогда вектор $W = \left(\frac{\delta_1}{\eta_1}, \frac{\varepsilon_1}{\eta_1}, \frac{\delta_2}{\eta_2}, \frac{\varepsilon_2}{\eta_2}, \dots, \frac{\delta_5}{\eta_5}, \frac{\varepsilon_5}{\eta_5} \right)$ есть вектор кривизны пальцев.

Мерой различия между двумя векторами признаков $W^{(1)}$ и $W^{(2)}$ является евклидово расстояние $\rho_\varepsilon(W^{(1)}, W^{(2)})$.

3.5.3 Локализация области кожных складок на пальце

С помощью семантической разметки (раздел 3.4.3), полученной на этапе детектирования, на изображении ладони выделяются области, содержащие пальцы. А именно, для каждого пальца в качестве такой области выступает трапеция, которая строится следующим образом (рис. 28).

Через точку основания пальца проводится отрезок, перпендикулярный оси пальца и имеющий длину, равную диаметру круга основания. Через вершину пальца проводится параллельный ему отрезок с длиной 80% от длины первого отрезка. Такой выбор длины позволяет избежать искажения текстуры при дальнейших преобразованиях, и в то же время не захватывать большие участки фона. Эти отрезки формируют основания трапеции. Полученная трапеция далее подвергается перспективному преобразованию, приводящему ее к прямоугольнику фиксированного размера 96×480 пикселей.

После нормализации можно переходить непосредственно к поиску области кожных складок (рис. 29). Так как освещение может меняться в широких пределах (в частности, может быть неравномерным), применяется предобработка изображения, улучшающая контрастность и увеличивающая качество извлекаемых признаков. К полученному изображению применяется адаптивное выравнивание гистограммы яркости с ограниченным контрастом (Contrast Limited Adaptive Histogram Equalization, CLAHE) [62].

Далее изображение инвертируется, из результата вычитается медиана по всему изображению (пиксели с отрицательными значениями полагаются равными 0) и сглаживается фильтром Гаусса (рис. 29(b)–(d)). Это позволяет убрать шумы в однородных областях изображения. Для точного поиска области кожных складок используется детектор границ Канни (Canny) [63] — рис. 29(e). К полученному изображению применяется детектор границ, ре-

зультатом работы которого является бинарное изображение, ненулевые пиксели которого обозначают найденные границы. Назовём его *изображением границ*.

Для фиксированного размера окна поиска назовём *плотностью границ* количество ненулевых пикселей на изображении границ внутри данного окна. Поиск области складок проводится только в одном направлении — вдоль вертикали изображения границ (рис. 29(f)), из которого удалены «короткие» (менее 32 пикселей длиной) компоненты связности: по срединной оси (жёлтая линия) перемещается окно поиска (красный прямоугольник 24×144 пикселей), и вычисляется плотность границ в данном окне. Точка, являющаяся центром окна с максимальной плотностью границ (обозначена зелёным), выбирается в качестве центра искомой области. Во время поиска центр окна не выходит за пределы верхней половины изображения пальца — по анатомическим причинам, так как нас интересуют кожные складки сустава на стыке первой и второй фаланг пальца. Искомая область кожных складок размером $W_{fkr} \times H_{fkr} = 96 \times 144$ пикселей однозначно определяется положением центра окна поиска (зелёный прямоугольник на рис. 29(a)), после чего выделяется в отдельное изображение.

Сделано следующее наблюдение — если плотность границ меньше некоторого значения, то, как правило, это свидетельствует об изображении плохого качества (например, расфокусированном). Генерация признаков по такому изображению не имеет смысла. Поэтому, если хотя бы для одного пальца плотность границ при поиске области интереса меньше некоторого порога, всё изображение исключается из дальнейшей обработки. Стоит отметить, что такие изображения весьма редко встречаются при работе системы (от 0 до 4 изображений на 400 кадров) — проверка на расфокусированные кадры (раздел 3.4.4) на этапе детектирования кардинально уменьшает их количество.

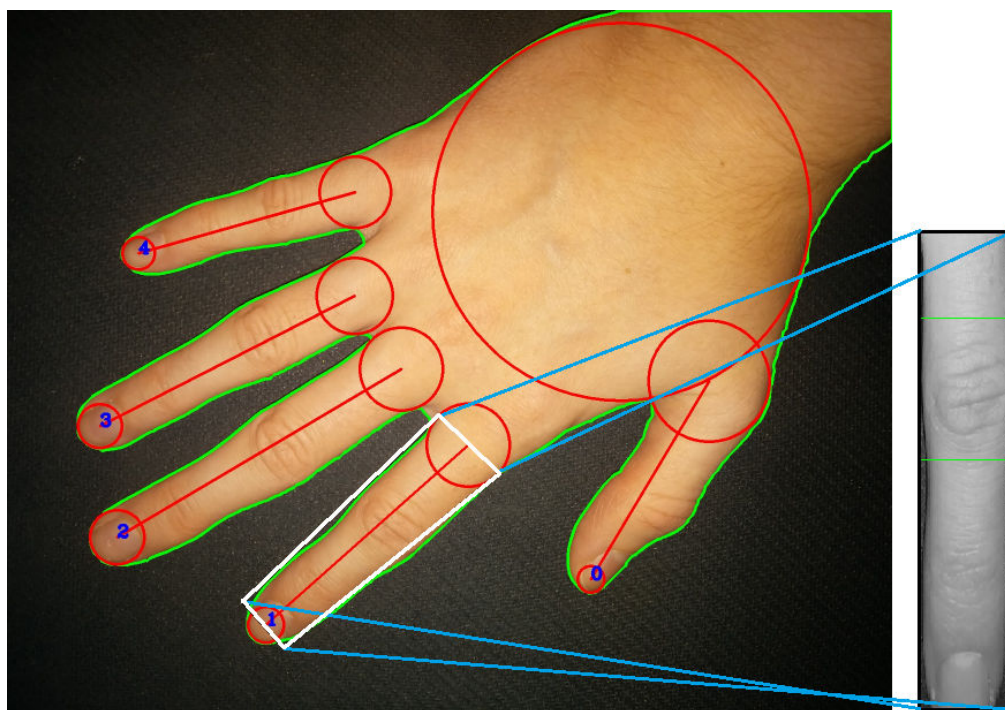


Рис. 28: Извлечение и нормализация области пальца. Область, ограниченная трапецией белого цвета, проективным преобразованием приводится к прямоугольнику 96×480 пикселей.

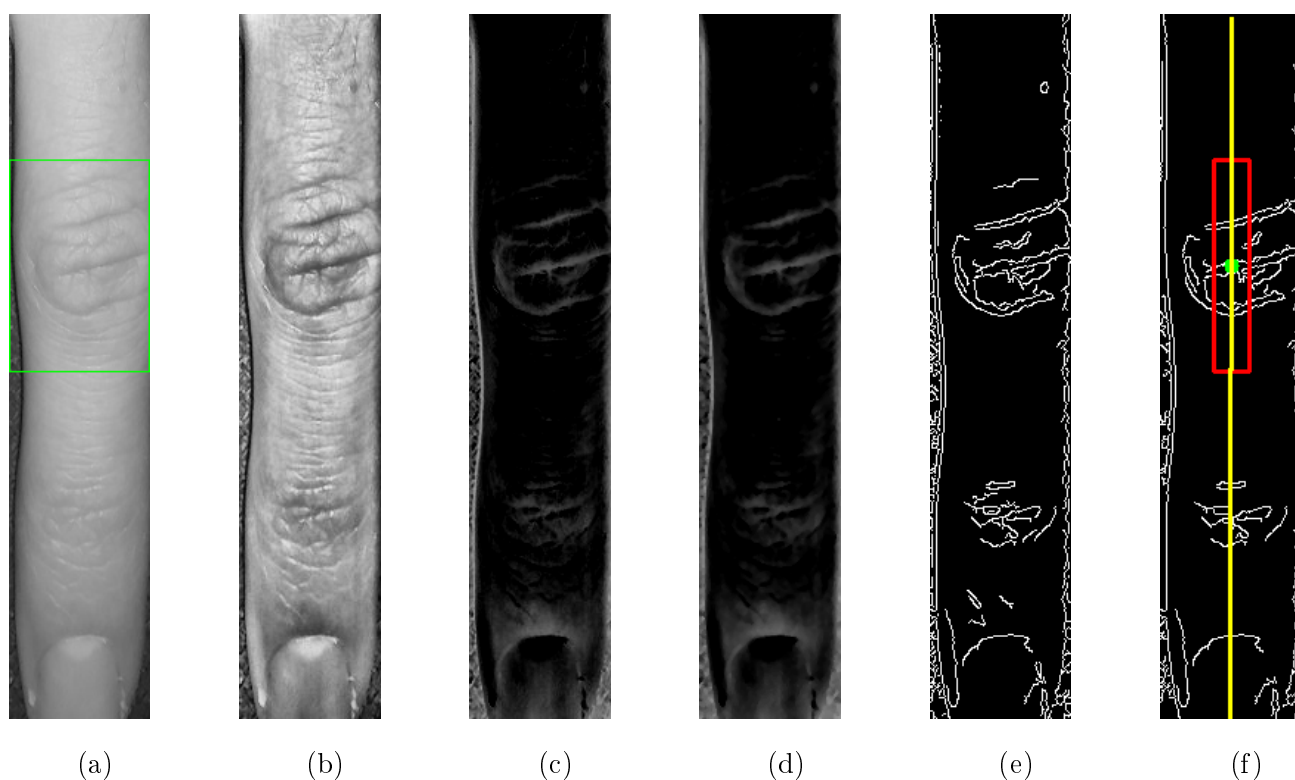


Рис. 29: Стадии процедуры локализации области кожных складок на пальце: а — исходное нормализованное изображение, б — выравнивание гистограммы яркости методом CLAHE, с — инверсия и вычитание медианы, d — сглаживание фильтром Гаусса, е — применение детектора границ Канни, f — поиск области интереса (красное окно с зелёным центром).

3.5.4 Извлечение текстурных признаков на базе фильтра Габора

Двумерный фильтр Габора широко используется в анализе изображений для извлечения локальных особенностей, имеющих квазипериодическую структуру. Классическим примером является его применение для распознавания по отпечаткам пальцев. Следует отметить, что кожные складки внешней стороны пальца также имеют квазипериодическую структуру, при этом они визуальны различимы на изображениях с низкой разрешающей способностью (раздел 2.2), с которыми идёт работа.

Метод кодирования локальной информации об ориентации. В настоящей работе используется та же основная идея по извлечению локальной информации об ориентации с помощью семейства двумерных фильтров Габора, что и в методе CompCode [23], предложенном для распознавания по рисунку ладони. Рассмотрим её подробнее. Фильтр Габора — линейный фильтр, импульсная характеристика которого имеет вид функции, представляющей собой произведение Гауссовой и гармонической функций:

$$G_R(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right),$$

где

$$x' = x \cos \theta + y \sin \theta, y' = -x \sin \theta + y \cos \theta.$$

Согласно другой интерпретации, выше представлена действительная часть комплексной функции Габора. Далее в работе под словами «фильтр Габора» и «функция Габора» мы будем понимать именно функцию G_R . Смысл параметров фильтра следующий:

- σ — стандартное отклонение Гауссова ядра, отвечает за амплитуду фильтра;
- λ определяет длину волны гармонической части фильтра;
- θ — угол поворота фильтра;
- γ определяет форму фильтра. Чем ближе значение к единице, тем менее вытянутую форму имеет фильтр;
- φ — фаза фильтра.

Также приведём вид одномерного фильтра Габора:

$$G(x; \lambda, \varphi, \sigma) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x}{\lambda} + \varphi\right).$$

Смысл параметров одномерного фильтра аналогичен двумерному случаю.

Указанная выше функция экспоненты обладает свойствами нормального распределения случайной величины. Согласно правилу трёх сигм, практически все значения экспоненты лежат в интервале $[-3\sigma, 3\sigma]$ — правило трёх сигм (рис. 31). В задачах анализа сигналов значения функции, зачастую, рассчитываются только в указанных пределах.

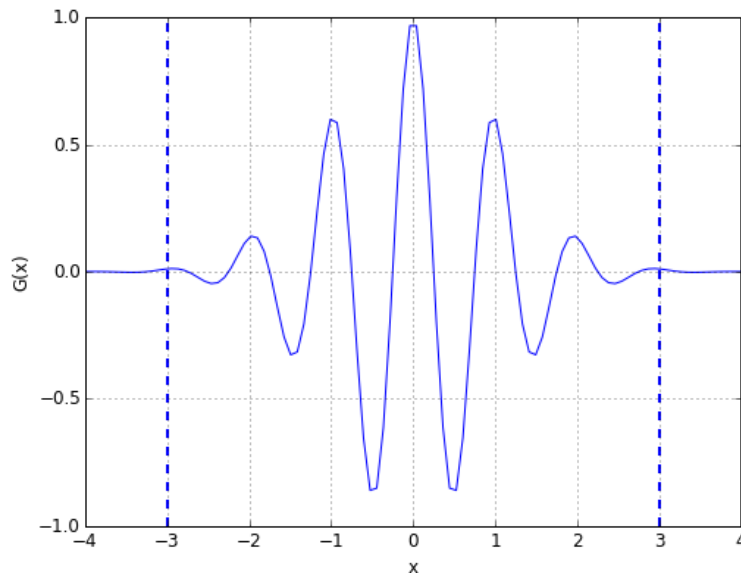


Рис. 30: Одномерный фильтр Габора с параметрами $\lambda = 1, \varphi = 0, \sigma = 1$. Вертикальные пунктирные линии ограничивают область $|x| < 3\sigma$.

Структура двумерного фильтра Габора позволяет адаптировать его для извлечения информации о характере кожных складок на изображении: их толщине и направлении. Выполним свёртку изображения с несколькими фильтрами, имеющими разные ориентации. Значением признака $F(x, y)$ в данной точке (x, y) изображения I является номер фильтра с минимальным откликом:

$$F(x, y) = \arg \min_j \{I(x, y) * G_R(x, y; \lambda, \theta_j, \varphi, \sigma, \gamma)\},$$

где $I(x, y)$ — окрестность изображения с центром в точке (x, y) ; θ_j — угол поворота фильтра, $\theta_j = j\pi/J, j = 0, \dots, J-1, J$ — количество различных ориентаций. При выборе $J = 6$ (обусловлено адаптацией теории вейвлетов к физиологическим особенностям человека [64]) при прочих фиксированных параметрах получается семейство из 6 разноориентированных фильтров (рис. 32(a) или (b)).

Модификация метода. Для адаптации изложенного метода кодирования локальной информации об ориентации под поставленную задачу идентификации личности по изображению внешней стороны ладони внесены следующие изменения.

1. Используется два семейства фильтров Габора с разными параметрами σ и λ , чтобы учесть текстурные особенности с разным характерным размером (с помощью «эпох» — рис. 32).
2. Вместо вектора признаков с длиной, равной количеству пикселей в области кожных складок пальца, используется значения $F(x, y)$ в вершинах равномерной прямоугольной сетки. Это позволяет существенно сократить размер вектора, и, как следствие, время извлечения и сравнения признаков, сохраняя приемлемый уровень точности классификации.
3. Для каждого изображения ладони вычисляются текстурные признаки областей кожных складок каждого из трёх пальцев. Сопоставление изображений областей кожных складок выполняется для каждого из пальцев отдельно, затем полученные значения расстояний объединяются с помощью функции MIN. Это значительно повышает качество классификации в сравнении с работой только по одному пальцу.

Далее остановимся подробнее на каждой модификации.

Эпохи. Кожные складки внешней стороны пальца имеют текстурные особенности разных характерных размеров. При фиксированных параметрах γ_0 и φ_0 семейство функций $\{G_R(x, y; \lambda_1, \theta_j, \varphi_0, \sigma_1, \gamma_0), j = 0, \dots, 5\}$ назовём *первой эпохой* (определяется параметрами σ_1, λ_1), $\{G_R(x, y; \lambda_2, \theta_j, \varphi_0, \sigma_2, \gamma_0), j = 0, \dots, 5\}$ — *второй эпохой* (определяется параметрами σ_2, λ_2).

Морщины пальцев обладают выраженной периодичностью, что описывается параметром λ , входящим в гармоническую часть функции Габора. Как правило, расстояния между крупными морщинами имеют характерный физический размер 2-3 мм., сами морщины обладают толщиной около 1-2 мм. Учитывая реальные физические размеры ширины пальцев (15–30мм) и ширину изображения кожных складок (96 пикселей), выделение этих особенностей обеспечивается эпохами с параметрами $\sigma_1 = 4.5, \lambda_1 = 2.4 * \sigma_1 * k_1, k_1 = 10$ и $\sigma_2 = 2, \lambda_2 = 2.4 * \sigma_2 * k_2, k_2 = 2$. Значения k_1 и k_2 были определены по Базе-2 (раздел 4.2.2) при обучении метода (раздел 4.3). На рис. 32 отображены эпохи, на рис. 31 приведены одномерные фильтры Габора с параметрами эпох.

Параметры $\gamma_0 = 0.5$ (морщины имеют вытянутую форму) и $\varphi_0 = 0$ (необходимости в сдвиге по фазе нет) заданы экспертно. Согласно правилу трёх сигм значения гауссианы близки к нулю на расстоянии большем 3σ от начала координат, поэтому размер окна свёртки

$S_x \times S_y$ для первой эпохи выбирается равным $[2 * 3\sigma_1] \times [2 * 3\sigma_1 * \gamma_0]$, а для второй — $[2 * 3\sigma_2] \times [2 * 3\sigma_2 * \gamma_0]$.

Таким образом, изображение описывают два вектора признаков F_1 и F_2 , где

$$F_i(x, y) = \arg \min_j \{I(x, y) * G_R(x, y; \lambda_i, \theta_j, \varphi_0, \sigma_i, \gamma_0)\}, \theta_j = j\pi/6, i \in \{1, 2\}, j = 0, \dots, 5.$$

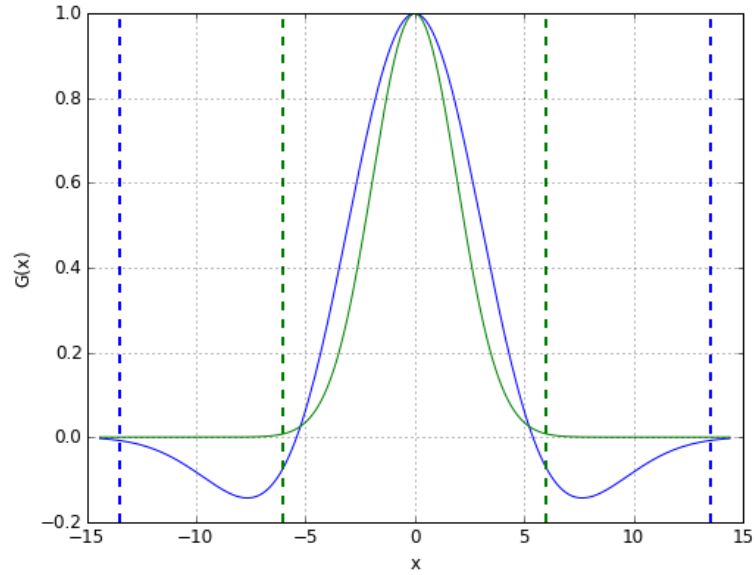
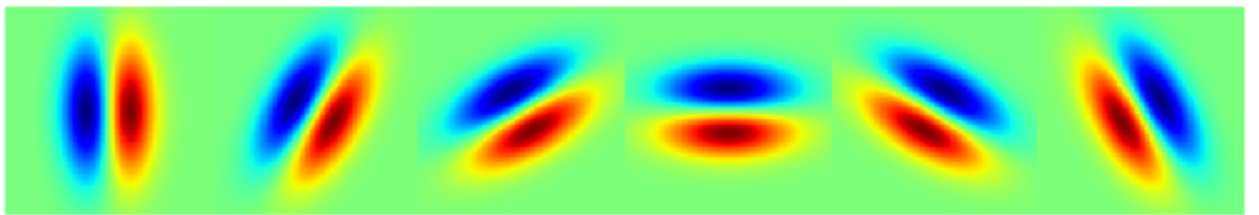
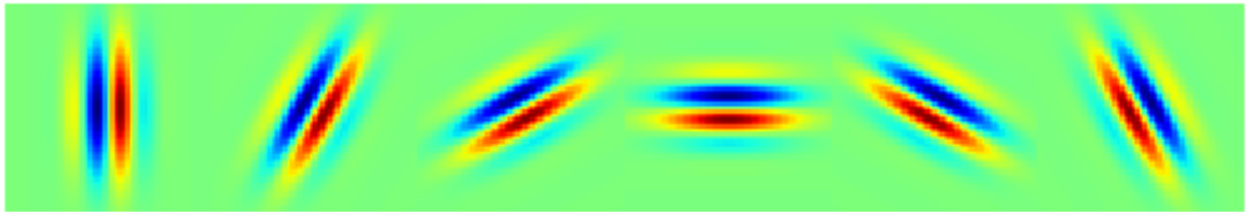


Рис. 31: Одномерные фильтры Габора с параметрами эпох: $\sigma_1 = 4.5$, $\lambda_1 = 4.8 * \sigma_1$ (синий) и $\sigma_2 = 2.0$, $\lambda_2 = 24 * \sigma_2$ (зелёный). Вертикальные пунктирные линии соответствующего цвета ограничивают области трёх сигм для каждого из графиков.



(a) Первая эпоха.



(b) Вторая эпоха.

Рис. 32: Визуализация используемых двумерных фильтров Габора.

Стоит также отметить, что для работы с локальными особенностями складок пальцев размером 1 мм. вполне достаточно DPI порядка 100, достигаемого при съёмке внешней стороны руки с разрешением 1280×960 пикселей (раздел 2.2).

Сетка. Вторая модификация состоит в том, что для получения векторов признаков каждой из эпох используются не все пиксели изображения кожных складок, а только их часть. А именно, пусть $S_x \times S_y$ пикселей — размер окна фильтра, (O_x, O_y) — стартовое смещение относительно начала координат изображения кожных складок. Тогда точки сетки, в которых берутся значения признака для конкретной эпохи, будут следующими:

$$P(j, i) = (O_x + j * S_x/2, O_y + i * S_y/2), i = 0, 1, \dots, \frac{2 * H_{fkp}}{S_y}; j = 0, 1, \dots, \frac{2 * W_{fkp}}{S_x},$$

где, как упоминалось выше, H_{fkp} и W_{fkp} — высота и ширина изображения кожных складок. Как показали эксперименты, такая модификация не ухудшает качество классификации, однако существенно уменьшает время сравнения пары признаков. Это является критически важным для системы реального времени.

Для первой эпохи мы получим $3 \times 11 = 33$ узла сетки (и такую же длину N_1 вектора признаков F_1), для второй — $8 \times 24 = 192$ узла сетки (и такую же длину N_2 вектора признаков F_2). Визуализация предлагаемого метода приведена на рис. 33. Для отображения значений $F(x, y)$ использованы оттенки серого от чёрного (0) к белому (1).

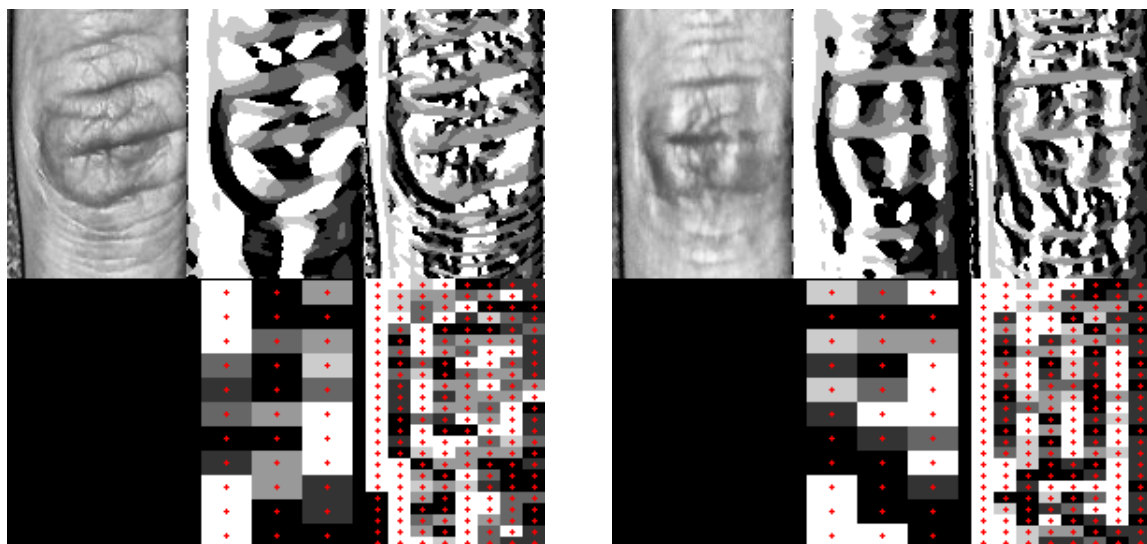


Рис. 33: Два примера формирования векторов признаков на базе эпох. Слева в верхем ряду — исходное изображение кожных складок; далее в верхнем ряду цветом пикселей показаны значения $F(x, y)$ для первой и второй эпох; в нижнем ряду красным отмечены узлы сетки для первой и второй эпох соответственно, цветом ячеек показаны значения $F(x, y)$ в узлах сетки.

3.5.5 Сравнение на основе Габор-признаков

При сравнении текстурных признаков применяется расстояние, учитывающее круговую природу ориентации в каждой точке. Пусть есть два вектора $f^{(1)}$ и $f^{(2)}$ размерности N , содержащие Габор-признаки на базе одной из эпох для двух сравниваемых изображений кожных складок. Расстояние между ними будет равно:

$$GDist(f^{(1)}, f^{(2)}) = \frac{1}{3N} \sum_{j=1}^N \min\{d_j, 6 - d_j\}, \quad d_j = |f_j^{(1)} - f_j^{(2)}|.$$

Чтобы компенсировать ошибки поиска области кожных складок, при сопоставлении векторов признаков конкретной эпохи выполняются сдвиги вверх-вниз точек сетки, в которых берутся значения признаков. В настоящей работе шаг сдвига равен 3 пикселя, используется 3 сдвига вверх и 3 вниз. В качестве результирующего расстояния по конкретной эпохе $GShiftDist(f^{(1)}, f^{(2)})$ берётся минимальное расстояние, полученное при попарных сдвигах векторов признаков относительно друг друга.

Обозначим за $GShiftDist_i^k$ расстояние между парой изображений кожных складок пальца i на базе эпохи k . Тогда расстояние по Габор-признакам между изображениями кожных складок пальца i вычисляется следующим образом:

$$GShiftDist_i = GShiftDist_i^1 + \alpha_g * GShiftDist_i^2, i \in \{1, 2, 3\},$$

где α_g — настраиваемый во время обучения параметр.

3.5.6 Сравнение на основе SIFT-признаков

Масштабно-инвариантная трансформация признаков (scale-invariant feature transform, SIFT) [65] — алгоритм выявления и описания локальных особенностей на изображении. Метод состоит из двух основных этапов: детектирования ключевых точек и их описания с помощью числовых векторов, называемых *дескрипторами*. Дескриптор SIFT инвариантен к масштабированию, изменению ориентации, изменению освещённости и частично инвариантен аффинным искажениям. В силу озвученных свойств его часто применяют в задачах машинного зрения (распознавание жестов, трекинг, трёхмерная реконструкция и т. д.).

В разделе 3.5.3 описан алгоритм локализации кожных складок на пальце, основным результатом которого является нормализованное изображение $W_{f_{kp}} \times H_{f_{kp}}$ пикселей найденной области. Далее на сравниваемых изображениях строятся дескрипторы SIFT (рис. 34), которые используются для вычисления меры различия между пальцами описанным ниже способом.

Пусть найдено n_1 ключевых точек на первом изображении I_1 области кожных складок пальца и n_2 — на втором изображении I_2 .

Обозначим $s_1 = \{(p_1^1, d_1^1), (p_2^1, d_2^1), \dots, (p_{n_1}^1, d_{n_1}^1)\}$ — множество пар ключевых точек и соответствующих дескрипторов для первого изображения, $s_2 = \{(p_1^2, d_1^2), (p_2^2, d_2^2), \dots, (p_{n_2}^2, d_{n_2}^2)\}$ — для второго изображения. Для каждого дескриптора d_i^1 ключевой точки p_i^1 на одном изображении вычисляется ближайший дескриптор на втором, в качестве метрики используется расстояние Евклида (в 128-мерном пространстве):

$$(p_{i*}^2, d_{i*}^2) = \arg \min_{(p^2, d^2) \in s_2} |d_i^1 - d^2|$$

Затем выполняется фильтрация полученных пар с целью отсеять ложные соответствия.

В первую очередь, для каждого дескриптора d_i^1 проверяется, насколько сильно расстояние $Dist_{1nn} = |d_i^1 - d_{i*}^2|$ до ближайшего соседа d_{i*}^2 отличается от расстояния $Dist_{2nn}$ до второго ближайшего:

$$Dist_{2nn} = \min_{\substack{(p^2, d^2) \in s_2; \\ d^2 \neq d_{i*}^2}} |d_i^1 - d^2|.$$

Если отношение этих расстояний больше 0.8 (определено эмпирически), то соответствие неустойчиво, и участвующие точки недостаточно отличаются от остальных. Такие точки исключаются из рассмотрения. Пусть после фильтрации осталось n пар ключевых точек с соответствующими дескрипторами.

В силу особенностей построения предполагается только линейный сдвиг нормализованных изображений кожных складок друг относительно друга — нормализация устраняет несоответствия в масштабе, а жёсткость пальца вдоль своей оси позволяет исключить влияние поворотов. Поэтому геометрические расстояния между координатами соответствующих ключевых точек на изображении $p_i^1 = (x_i^1, y_i^1)$ и $p_i^2 = (x_i^2, y_i^2)$ должны отличаться на одну и ту же величину. В качестве оценки этой величины выбрана медиана указанных расстояний по всем соответствиям:

$$M = median(\{|p_i^1 - p_{i*}^2|, i = 1, 2, \dots, n\})$$

Пары точек, абсолютное отклонение расстояния внутри которых $Dist_i^{geom} = |p_i^1 - p_{i*}^2|$ от медианы $|Dist_i^{geom} - M|$ больше 10 (определено эмпирически), отбрасываются.

Аналогично рассматриваются направления векторов $\mathbf{v}_i = p_i^1 - p_{i*}^2$. Проверяется, что все направления этих векторов примерно совпадают. Для этого вычисляется медиана всех направлений:

$$A = median(\{angle(\mathbf{v}_i), i = 1, 2, \dots, n\}),$$

где функция $angle(\mathbf{v})$ вычисляет направление вектора. После этого отбрасываются пары, у которых угол $angle(\mathbf{v}_i)$ отличается от медианы A больше $\pi/20$ (определено эмпирически).

Если после фильтрации осталось достаточное количество n_0 пар точек (дескрипторов), считается, что это изображения одного и того же пальца. Расстояние между двумя изображениями кожных складок пальцев тогда имеет следующий простой вид:

$$SDist(I_1, I_2) = \begin{cases} 0, & \text{если } n_0 \geq n_{pairs}; \\ 1, & \text{иначе.} \end{cases}$$

где n_{pairs} — пороговое значение, которое может выставляться экспертно (во время исследования в некоторых экспериментах использовано фиксированное значение 10) или настраиваться во время обучения на конкретной базе ладоней. За $SDist_i$ обозначим расстояние между двумя изображениями кожных складок пальца i .

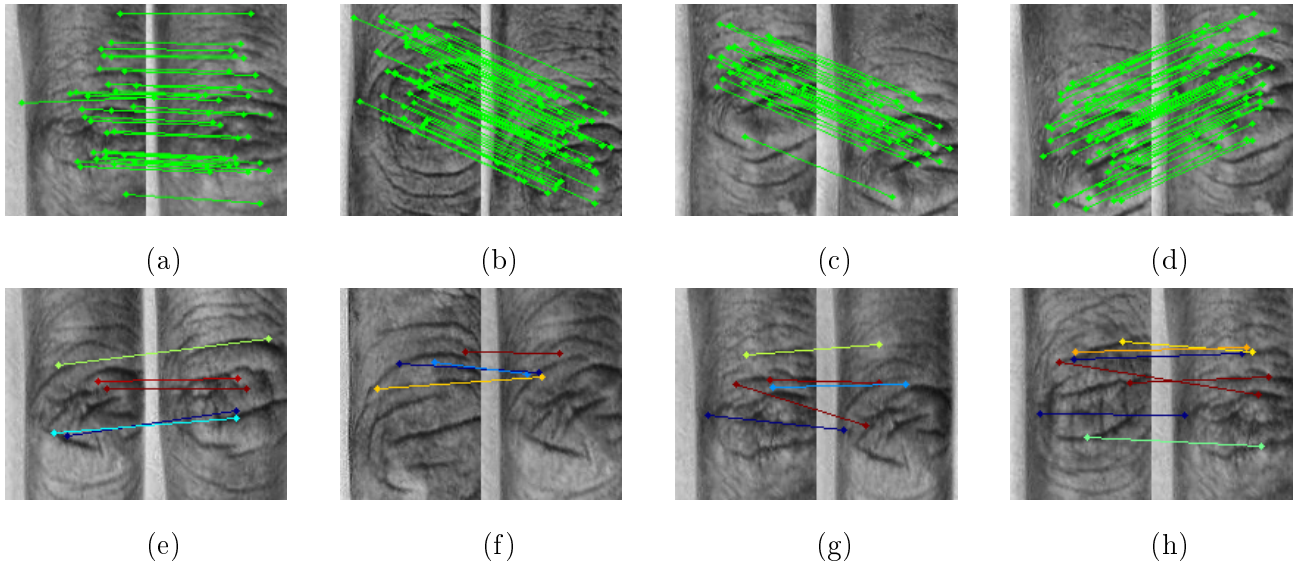


Рис. 34: Процедура сопоставления SIFT-дескрипторов для нормализованных изображений кожных складок пальцев. В верхнем ряду на любом изображении в поиске соответствия участвуют области кожных складок одного и того же человека, в нижнем ряду — разных людей.

3.5.7 Сравнение на основе текстурных признаков

Итоговое расстояние между парой изображений кожных складок пальца i с использованием Габор- и SIFT-признаков вычисляется так:

$$Dist_i = GShiftDist_i + \alpha_s SDist_i,$$

где α_s —настраиваемый во время обучения параметр. Итоговое расстояние между изображениями ладоней получается с помощью функции MIN из попарных расстояний между изоб-

ражениями кожных складок каждого из трёх пальцев:

$$Dist = \min\{Dist_1, Dist_2, Dist_3\}.$$

В качестве классификатора в работе используется метод ближайшего соседа (1NN) с функцией расстояния $Dist$.

3.5.8 Выводы к разделу 3.5

Предложен метод точного выделения маски ладони на снимке, использующий сегментацию на основе разрезов графов вместе с априорной информацией о пикселях фона и объекта, полученной на этапе детектирования.

Описаны признаки формы ладони (кривизна пальцев, ломаная, спектр), которые могут быть полезны при последующей классификации. Они вычисляются на основе точной сегментации. Приведены метрики для сравнения объектов на основе каждого из указанных признаков и их комбинации.

Представлен метод локализации области кожных складок на пальце, использующий семантическую разметку ладони.

Разработаны методы извлечения признакового описания из изображения области кожных складок внешней стороны пальца, а также предложены способы сравнения объектов на основе полученного признакового описания.

Первый метод для первичного кодирования информации о локальной ориентации текстурных особенностей в точке изображения использует семейство двумерных фильтров Габора с различными ориентациями. Можно выделить следующие особенности предложенного метода. Для учёта особенностей текстуры кожных складок пальца с разным характерным размером предложены два семейства фильтров Габора с различными параметрами σ , λ , называемых эпохами. С их помощью формируются два (по числу эпох) вектора признаков, значения элементов которых вычисляются в узлах равномерной сетки (размер ячейки равен размеру окна фильтра для соответствующей эпохи). В качестве вектора признаков берутся значения карты ориентаций в узлах сетки, размер ячейки которой равен размеру окна фильтров из соответствующей эпохи. При сравнении векторов признаков двух изображений кожных складок пальцев в рамках одной эпохи их сопоставление происходит с помощью специальной функции расстояний, учитывающей угловую природу признаков. Для минимизации влияния неточностей, связанных с локализацией областей кожных складок пальцев, выполняются попарные сдвиги (в контексте равномерной сетки) векторов признаков и

ищется минимальное расстояние. Итоговое Габор-расстояние между изображениями кожных складок является линейной комбинацией расстояний (с учётом сдвигов) по каждой из эпох.

Второй метод основан на дескрипторах ключевых точек SIFT. При сравнении двух изображений кожных складок пальца ищутся соответствия между дескрипторами. Если после окончания алгоритма поиска и фильтрации осталось достаточное количество пар, то SIFT-расстояние между изображениями принимается за 0 (принадлежат одному и тому же человеку), иначе — 1.

Для вычисления итогового расстояния между двумя изображениями внешней стороны ладони предложена функция расстояний, объединяющая результаты сравнения изображений кожных складок каждого из пальцев на основе Габор- и SIFT-признаков.

3.6 Выводы к главе 3

Задача биометрической идентификации личности по изображению внешней стороны ладони рассмотрена в рамках методологии машинного обучения.

Предложено разделить задачу идентификации по ладони на задачи детектирования и распознавания. Это связано как с ограниченными вычислительными возможностями мобильных устройств, так и с необходимостью обеспечить безопасность хранимых биометрических признаков.

Сформулировано понятие корректной ладони. Предложен метод детектирования ладони, состоящий из нескольких основных шагов: «быстрой» сегментации; построения скелетного представления, семантической разметки и проверки ладони на корректность; определения расфокусированных кадров; формирования данных для последующей отправки в модуль распознавания. Эффективная реализация позволяет работать с кадрами видеопотока на стороне клиента (мобильного устройства).

Разработан метод распознавания ладони, который выполняет извлечение признаков описания и сравнение его имеющимися в базе образцами. В него вошли: метод «точной» сегментации ладони на основе разрезов в графах и априорной информации с этапа детектирования; метод локализации области кожных складок на пальце на основе семантической разметки ладони; методы сравнения изображений кожных складок пальца на основе двумерных фильтров Габора и на базе дескрипторов SIFT; функция расстояний между двумя изображениями внешней стороны ладони на основе Габор- и SIFT-признаков отдельных пальцев.

4 Экспериментальное исследование методов детектирования и распознавания

4.1 Оборудование

Разработанная система идентификации по ладони тестировалась на мобильных устройствах из среднего ценового диапазона:

- планшет Samsung Galaxy Note 10.1 на базе операционной системы (ОС) Android 4.1, четырёхъядерного процессора Samsung Exynos 4412 1.4 ГГц и 2 Гб оперативной памяти (далее — Устройство-1);
- смартфон LG G2 на базе ОС Android 4.4, четырёхъядерного процессора Qualcomm Snapdragon 800 2.26 ГГц и 2 Гб оперативной памяти (далее — Устройство-2);
- смартфон Samsung Galaxy S5 на базе ОС Android 5.0, четырёхъядерного процессора Qualcomm Snapdragon 801 2.5 ГГц и 2 Гб оперативной памяти (далее — Устройство-3).

В качестве сервера использовался ноутбук на базе ОС Ubuntu 12.04, четырёхъядерного процессора Intel Core i7 2.4 ГГц и 6 Гб оперативной памяти.

4.2 Базы изображений

Для тестирования созданы собственные базы изображений тыльной стороны ладони.

База-1 собрана в рамках этапа исследования, посвящённому методу детектирования ладони из кадров видеопотока [2], [3], [8].

Во время сбора Базы-1 уделено недостаточное внимание резкости собираемых изображений, поскольку изначально предполагалась работа метода распознавания исключительно на основе признаков формы. В связи с этим, на этой базе методы распознавания по текстурам кожных складок пальцев показали слабые результаты. Как результат, автором создана новая база (далее — База-2), в которой требуемые участки ладони находятся в фокусе камеры. Изначально в базу вошли 280 корректных изображений (см. определение в разделе 3.3) тыльной стороны левой ладони 40 человек, при этом произведён итоговый экспертный отбор снимков — так, чтобы в рамках одного человека изображения ладони максимально попарно отличались друг от друга. На Базе-2 успешно опробованы методы распознавания по текстурам [5], [6].

К моменту проведения финальных вычислительных экспериментов, изложенных в текущей главе, База-2 дополнена до 502 изображений 48 человек.

Помимо Базы-1 качество детектирования проверено на базе HGC сканированных изображений внутренней стороны ладони.

Результаты распознавания получены не только на Базе-2, но также и на базе PolyU HD (база изображений тыльной стороны ладони от Гонконгского Политехнического Университета).

Далее приведено более детальное описание баз изображений ладони, использованных в ходе исследовательской работы.

4.2.1 База-1

База содержит 2322 изображения тыльной стороны правой ладони 50 человек, которые получены при разложении 80 видео на кадры с их последующим прореживанием и дальнейшей разметкой экспертом на корректные и некорректные. Кадры, сильно размытые из-за движений рукой испытуемым, из базы исключались. Отношение числа изображений с корректными ладоням к числу изображений с некорректными и отсутствующими — 1 к 1.

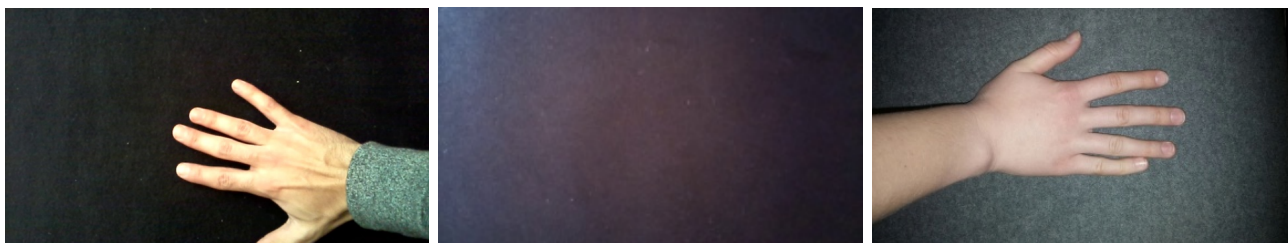
Съёмка видео осуществлена с помощью Устройства-1 и Устройства-2, разрешение — 1920×1080 или 1280×720 пикселей. Примеры изображений приведены на рис. 35. Данная база применялась для настройки структурных параметров метода детектирования — например, статистики граничного коридора, величин порогов в пороговых проверках (раздел 3.4.3). При съёмке требовалось соблюдение следующих условий:

1. в качестве фона выступает тёмная матовая поверхность;
2. ладонь достаточно освещена рассеянным источником — в противном случае в приложении необходимо активировать режим работы со встроенной вспышкой;
3. отсутствуют длинные рукава, кольца, часы, браслеты и других посторонние предметы;
4. ладонь прижата к поверхности, плоскость ладони ортогональна оси визирования (съёмки).

Пользователю разрешалось двигать ладонью и пальцами в кадре, избегая резких движений, — в частности, убирать из кадра ладонь и возвращать её обратно, сводить и разводить пальцы и т. д. Движения камерой оператором по возможности исключались — в частности, большинство видео сняты Устройством-1, закрепленным статично (рис. 36а).



(а) Корректные изображения.



(b) Некорректные изображения.

Рис. 35: Примеры изображений Базы-1.



(а) Съёмка со статичной камерой.



(b) Съёмка «с руки».

Рис. 36: Примеры экспериментальных установок системы идентификации по ладони.

4.2.2 База-2

База содержит 502 валидных изображения тыльной стороны левой ладони 48 человек (рис. 37 и 40а) — преимущественно студентов двух групп ВМК МГУ. Для валидации применялся алгоритм детектирования (раздел 3.4), настроенный на Базе-1, — т. е. помимо проверки на корректность присутствовала и проверка на расфокусированность. Съёмка выполнена разработанным мобильным приложением для идентификации по ладони (главу 5) с помощью

Устройства-2 и Устройства-3 в реальных условиях эксплуатации системы. Каждый человек участвовал минимум в трёх различных конфигурациях условий съёмки (освещение, фон, мобильное устройство). На финальном этапе экспертом исключены визуально похожие между собой снимки одного человека, чтобы максимизировать вариабельность образцов в пределах одного испытуемого. При съёмке требовалось соблюдение следующих условий:

1. в качестве фона выступает тёмная матовая поверхность;
2. ладонь достаточно освещена рассеянным источником — в противном случае в приложении необходимо активировать режим работы со встроенной вспышкой;
3. ладонь вместе с частью запястья расположена целиком в кадре, пальцы расставлены;
4. отсутствуют длинные рукава, кольца, часы, браслеты и других посторонние предметы;
5. камера при съёмке сфокусирована на средних фалангах пальцев;
6. движения ладонью и пальцами во время процедуры идентификации минимизированы — либо совершаются медленно, либо не совершаются вообще;
7. ладонь расслаблена и прижата к поверхности, плоскость ладони *примерно* (допустимы и приветствуются отклонения до 30°) ортогональна оси визирования (съёмки).

Условия 1, 2 обеспечивают приемлемое качество сегментации на этапе детектирования (раздел 3.4). Условия 3, 4 соответствуют аналогичным в определении корректной ладони (раздел 3.3), так как при отсутствии корректных кадров в видеопотоке модулю распознавания будет не с чем работать. Условия 5, 6 помогают понизить число расфокусированных («смазанных») снимков, которые непригодны для работы метода распознавания по текстурам внешней стороны пальцев. Такие кадры в некотором количестве всё равно будут присутствовать в видеопотоке, и для их отсева разработан специальный алгоритм (раздел 3.4.4). Заключительное условие 7 способствует уменьшению вариабельности особенностей текстуры и формы ладони в пределах одного человека в силу того, что сильное изменение угла наклона камеры кардинальным образом влияет на вид ладони на снимке (и, как следствие, на извлекаемые признаки).

Таким образом, принципы коллаборации, на которых построено взаимодействие пользователя и системы идентификации, позволяют организационными мерами уменьшить число непригодных для распознавания снимков. Тем самым, уменьшается время, необходимое для идентификации одного пользователя, и увеличивается точность распознавания.

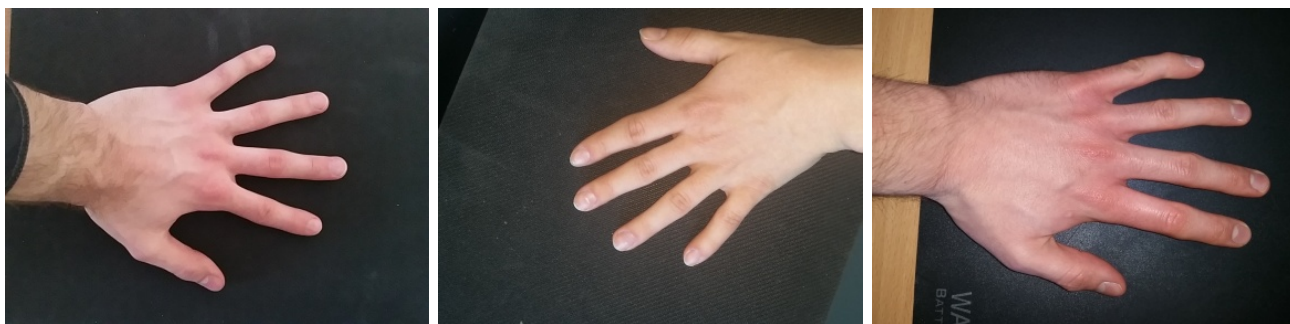


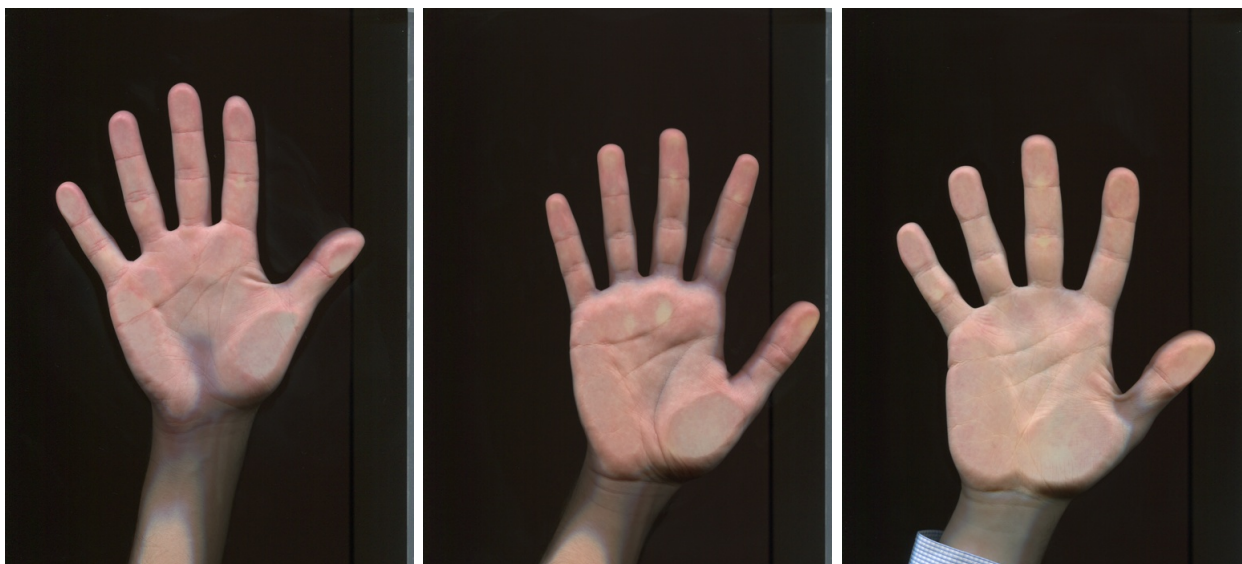
Рис. 37: Примеры изображений Базы-2.

4.2.3 База HGC

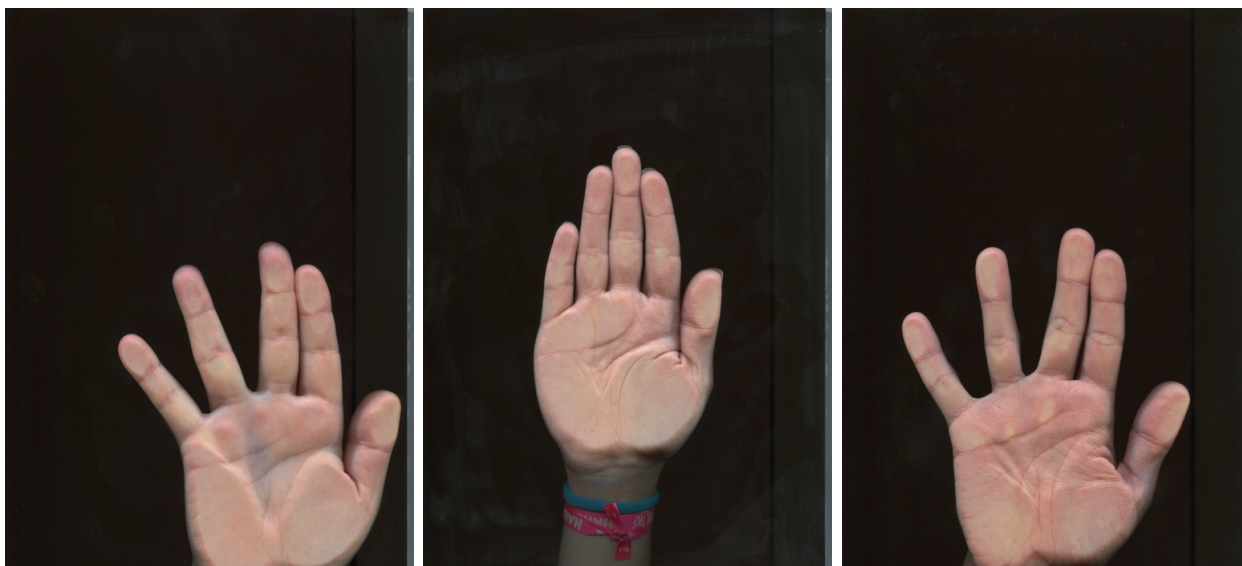
В качестве сторонней базы для оценки качества детектирования использовалась HGC [66], а именно — её тестовая часть. Изображения ладоней получены с помощью сканера. Фон везде тёмный и контрастирует со светлыми ладонями. Разрешение изображений составляет 510×702 пикселей. Стоит отметить, что для данной базы положение руки не фиксировано, поэтому присутствует много изображений с некорректной ладонью (таблица 4 и рис. 38).

Часть базы	Количество изображений	Некорректных изображений	Корректных изображений
Обучающая (Train)	300	155	145
Тестовая (Test)	160	64	96

Таблица 4: Состав базы HGC.



(a) Корректные изображения.



(b) Некорректные изображения.

Рис. 38: Примеры изображений базы HGC Test.

4.2.4 База PolyU HD

В качестве сторонней базы для оценки качества распознавания применена бесконтактная база изображений тыльной стороны ладони от Гонконгского Политехнического Университета (англ. The Hong Kong Polytechnic University Contactless Hand Dorsal Images Database) [67]. Для краткости назовём её PolyU HD. Изображения получены с помощью зафиксированной цифровой камеры и специального источника освещения, ось визирования ортогональна плоскости изображения. Фон везде светлый и контрастирует с тёмными ладонями. Разрешение изображений составляет 1600×1200 пикселей. Указанная выше публика-

ция посвящена распознаванию человека по текстуре кожных складок фаланг пальцев, при этом не на всех изображениях базы ладонь целиком присутствует в кадре.

Так как для работы разработанного автором метода распознавания требуется корректная ладонь, часть изображений из базы PolyU HD отброшена. В результате, для тестирования остаётся 165 человек и 931 изображение (рис. 39), от 3 до 12 изображений на человека (рис. 40). Следует отметить, что у алгоритмов сегментации (как быстрой, так и точной) не возникает проблем с выделением ладоней на базе PolyU HD.



(a) Корректные изображения



(b) Некорректные изображения

Рис. 39: Примеры изображений базы PolyU HD.

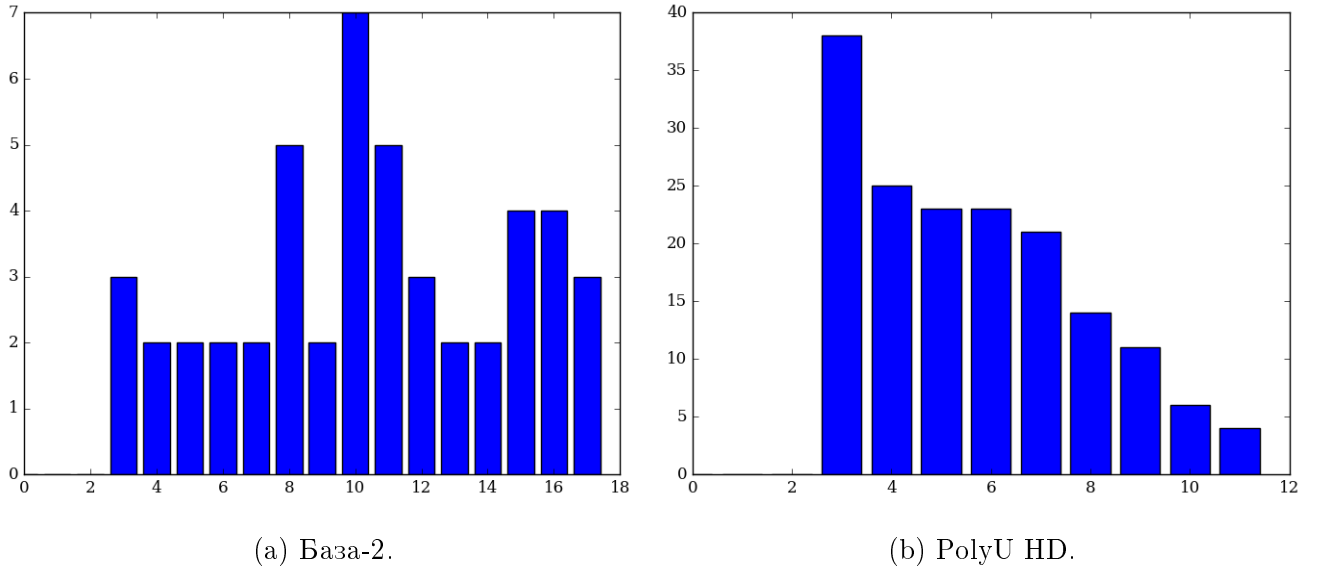


Рис. 40: Гистограммы количества картинок для баз ладоней. По горизонтальной оси показано количество картинок на одного человека, а по вертикальной — количество человек.

4.3 Вычислительные эксперименты

Общая схема тестирования. Применяемая в работе оценка качества какого-либо метода (равно реализующего его алгоритма $a = a(x, \gamma)$) системы биометрической идентификации по ладони построена на принципах машинного обучения (раздел 3.1) и выглядит следующим образом.

1. Исходная выборка (база) X^n случайно разбивается на обучающую X^l и тестовую непересекающиеся части X^k (принципы разбиения и соотношения частей могут отличаться).
2. По обучающей выборке X^l настраиваются параметры γ_{train} алгоритма a (если таковые имеются).
3. На тестовой выборке X^k вычисляется функционал качества $Q(a(X^l, (\gamma_{train}, \gamma_{struct})), X^k)$ — например, частота ошибок классификации для задачи идентификации или EER для задачи верификации.
4. Шаги 1-3 повторяются заданное число раз p . Для получения итоговой оценки качества метода значения функционала качества для каждого разбиения выборки суммируются и делятся на число разбиений p .

$\gamma = (\gamma_{train}, \gamma_{struct})$ — параметры алгоритма a : γ_{train} — вычисляемые по X^l , γ_{struct} — структурные (задаются экспертно или настраиваются внешней процедурой обучения). Разбиение исходной выборки на обучающую и тестовую части призвано уменьшить эффект переобучения

(раздел 3.1.2), а усреднение значений функционала качества по числу повторных запусков снижает зависимость итогового результата от конкретного разбиения.

При тестировании в данной работе разбиение на обучающую и тестовую части, как правило, осуществлялось в соотношении 2: 1, а усреднение выполнялось по $p = 10$ разбиениям.

Настройка структурных параметров. Предложенный метод распознавания (раздел 3.5) использует несколько параметров, определяющих поведение меры близости изображений, которые можно определить как структурные: α_g (раздел 3.5.5), n_{pairs} (раздел 3.5.6) и α_s (раздел 3.5.7).

Для их настройки может применяться следующая несложная схема. Обозначим для параметра p множество допустимых значений через $P \subset \mathbb{R}$. Пусть также $E(p)$ — функционал качества, зависящий от параметра p (величины других параметров фиксированы). Так как множество допустимых значений P , как правило, бесконечно, требуется сократить пространство поиска. Для этого применяется поиск по одномерной сетке с адаптивным шагом. Сначала множество параметров строится с экспоненциальным шагом:

$$P_{exp} = \{\exp(i), i = 0, 1, \dots, m\},$$

где m — разумно ограничивающее множество число. Далее применялся следующий алгоритм:

1. $P = P_{exp}$
2. $p_* = \arg \min_{p \in P} E(p)$
3. $left = p_*/2; right = p_* * 2$
4. Если $right - left < \varepsilon$ для некоторого малого ε , перейти на шаг 6, иначе перейти на шаг 5
5. $P = \{left + i * step, i = 0, 1, \dots, k\}$, где k — размер сетки на каждом шаге, а $step = (right - left)/k$ — шаг сетки. Перейти на шаг 2
6. Искомое значение параметра считается равным $(left + right)/2$, поиск завершён.

Параметры перебора m, k, ε выбираются, исходя из требуемого диапазона и желаемого времени перебора. Алгоритм перебора приведен для одномерного случая, для двух и более одновременно настраиваемых параметров можно использовать конструкцию из вложенных переборов.

Далее описаны эксперименты по оценке качества алгоритмов детектирования (раздел 4.3.1) и распознавания (раздел 4.3.2).

4.3.1 Тестирование метода детектирования

Эксперименты на Базе-1. Во время тестирования при случайном разбиении выборки по её обучающей части X^l настраиваются γ_{train} параметры метода детектирования: статистика граничного коридора, пороги длин пальцев и т.д. (раздел 3.4.3). В качестве функционала качества алгоритма детектирования выступает доля изображений ладони в тестовой выборке, для которых неверно определена корректность. В экспериментах по оценке качества детектирования из алгоритма (раздел 3.4) исключается процедура определения расфокусированных кадров (раздел 3.4.4), так как её наличие целесообразно при тестировании метода распознавания и работе всей системы в реальных условиях.

В проведённом вычислительном эксперименте все изображения были сжаты до разрешения 640×360 пикселей — этого разрешения вполне достаточно для уверенной работы метода детектирования. Результаты тестов приведены в таблице 5. В последнем столбце приведены значения затрат времени на кадр (Time Per Frame), включающие в себя всю процедуру детектирования, т.е. сумму времени бинаризации, построения скелета, семантической разметки и проверки на корректность ладони. Согласно показаниям *профилировщика* (программного инструмента для измерения времени выполнения программ или их частей), наиболее времязатратной частью алгоритма детектирования является построение диаграммы Вороного, необходимой для получения скелета, — порядка 60% времени.

FAR	FRR	Ошибка, %	TPF, мс
3.2	4.6	3.6	30.1

Таблица 5: Ошибки детектирования на Базе-1.

Предложенный в работе алгоритм показал низкие значения ошибок (как FAR, так и FRR) и достаточную скорость для работы в реальном времени, поэтому его вполне возможно использовать в задаче детектирования ладони на стороне клиента (в т.ч. на мобильном устройстве). Также алгоритм детектирования устойчив к понижению разрешения изображений — демонстрирует сравнимые результаты даже при сжатии исходных кадров до 320×180 пикселей.

Эксперименты на базе HGC. Процедура оценки качества алгоритма детектирования аналогична приведённой в предыдущем параграфе — за исключением того, что случайные разбиения не выполнялись, так как база уже разбита авторами на обучающую и тестовую части. Процедура семантической разметки (раздел 3.4.3) не предназначена для случаев све-

дённных пальцев, поэтому такие некорректные изображения были отброшены перед тестированием. Примеры работы процедуры детектирования приведены на рис. 41.

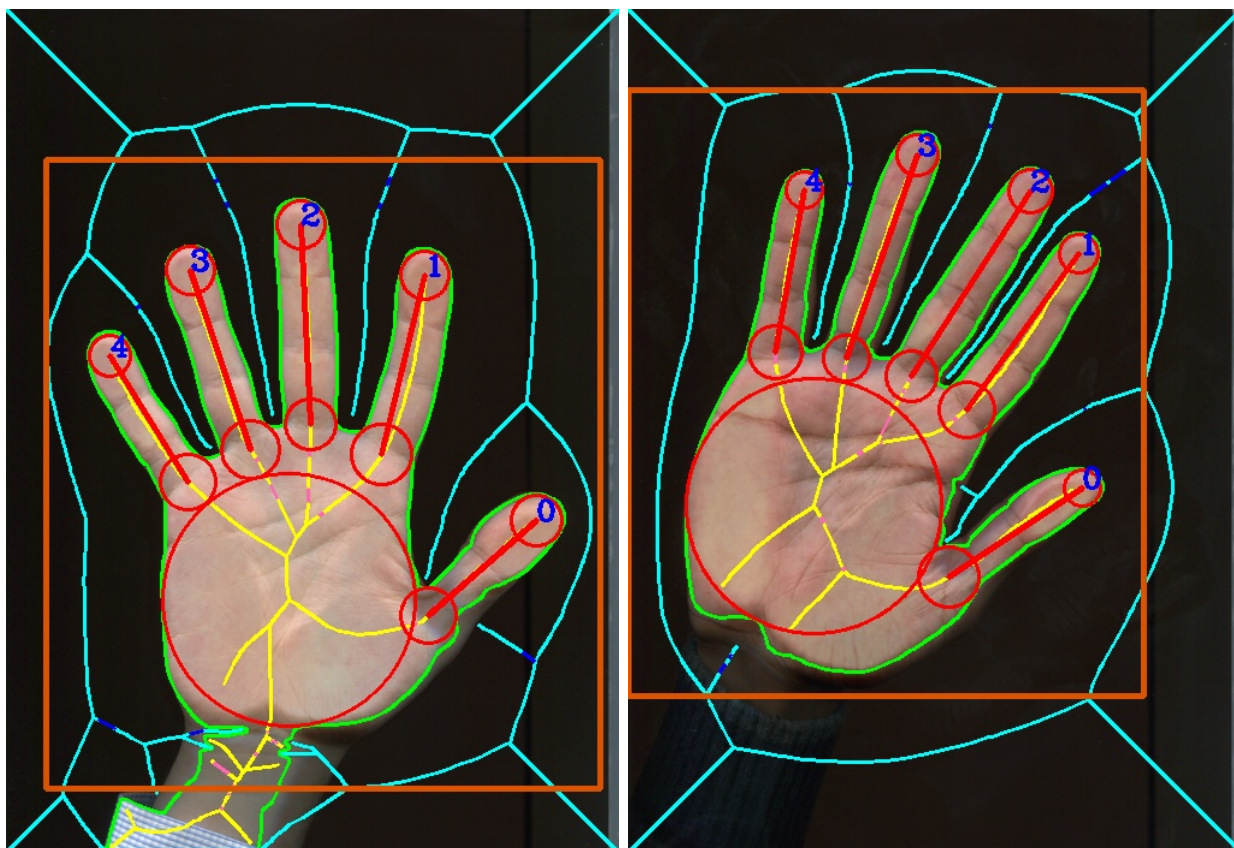


Рис. 41: Работа детектирования на базе HGC.

Количественное сравнение алгоритма детектирования на разных базах представлено в таблице 6. Результаты лучших алгоритмов на тестовой части базы HGC приведены в таблице 7, при этом предложенный автором метод детектирования показывает сравнимые с лидерами результаты.

База	FAR	FRR	Ошибка, %	TPF, мс
База-1	3.2	4.6	3.6	30.1
HGC [66]	3.8	5.1	4.3	45.4

Таблица 6: Результаты алгоритма детектирования на Базе-1 и HGC Test.

Выводы. Проведённые эксперименты показывают, что метод детектирования устойчиво и точно работает на различных данных, удовлетворяющих простым условиям съёмки (раздел 4.2.1). Малое время обработки (≈ 30 мс на ПК) одного кадра приемлемого разрешения

Авторы	Абсолютный штраф	Относительный штраф	Верная разметка (%)		Неверная разметка (%)	
			Концы пальцев	Точки между пальцами	Концы пальцев	Точки между пальцами
MCST [61]	22.00	0.14	99.38	97.34	0.63	2.66
Предложенный метод	15.00	0.16	99.17	97.14	0.83	2.86
ELSayed [68]	26.00	0.16	99.50	96.56	0.50	3.44
Alberto de Santos [69]	57.30	0.36	99.38	91.41	0.00	7.97
Diogo Martins [66]	77.90	0.49	97.00	88.91	0.63	4.84
scab12 [66]	98.10	0.61	96.13	86.56	0.25	7.97

Таблица 7: Сравнение различных алгоритмов детектирования на тестовой части базы HGS. Для предложенного метода учтены только результаты на части тестовых изображений, которые были определены экспертом как корректные (96 из 160). В таблице абсолютный штраф равен штрафу, начисленному по всем изображениям, на которых запускался тестируемый метод, а относительный — отношению этого штрафа к количеству обработанных изображений. Определение штрафа в работе [66].

(640×360 пикселей) позволяет проводить детектирование ладони в видеопотоке мобильного устройства с частотой 15-30 кадров в секунду, что подтверждается апробацией системы.

4.3.2 Тестирование метода распознавания

Влияние метода сегментации на уровень ошибок метода распознавания. В разделе 3.5.1 показано, каким образом результаты семантической разметки из модуля детектирования могут быть использованы для выполнения точной графовой сегментации ладони. В таблице 8 приведены результаты тестирования метода распознавания (Габор- и SIFT-признаки по 3 пальцам) на Базе-2 в зависимости от используемого алгоритма сегментации: «быстрого» Оцу [50] и «точного» OneCut [58]. Они обосновывают использование метода графовой сегментации в составе модуля распознавания.

Оцу, %	OneCut, %
0.9	2.3

Таблица 8: Качество идентификации для разных методов сегментации на Базе-2.

Анализ эпох. В разделе 3.5.4 для генерации признаков на основе двумерных фильтров Габора предложено использовать два семейства этих фильтров (две эпохи). В таблице 9 приведены результаты тестирования метода распознавания на Базе-2 при использовании только Габор-признаков для каждого из пальцев. Видно, что линейная комбинация эпох даёт более низкий уровень ошибок.

№ пальца	Эпоха 1, %	Эпоха 2, %	Обе эпохи, %
1	9.2	8.0	5.7
2	9.9	7.9	6.3
3	9.6	8.2	6.1

Таблица 9: Качество классификации для различных эпох и их комбинации на Базе-2.

Анализ признаков. Как и в случае прочих вычислительных экспериментов, для тестирования отдельных признаков и их комбинаций (табл. 10) применена общая схема, изложенная в начале раздела 4.3.

Для оценки классификатора в задаче биометрии часто строится ROC-кривая, которая характеризует зависимость доли верных положительных классификаций от доли ложных

положительных классификаций при варьировании порога решающего правила. Она предназначена для количественной оценки бинарной классификации, что в терминах биометрии соответствует задаче верификации. С помощью ROC-анализа могут быть получены такие важные характеристики метода распознавания, как AUC (англ. Area Under the Curve — агрегированная характеристика качества классификации, не зависящая от соотношения цен ошибок) и EER (англ. Equal Error Rate — коэффициент, при котором ошибки обоих родов эквивалентны).

ROC-кривые для классификаторов на основе признаков формы на Базе-2 приведены на рис. 42, на основе текстурных признаков на Базе-2 — на рис. 43, на основе текстурных признаков на базе Poly HD — на рис. 44.

Признаки формы продемонстрировали плохие результаты на Базе-2. Причина, по-видимому, состоит в том, что признаки формы ладони сильно зависят от угла съёмки. Более того, форма ладони сама по себе обладает высокой внутриклассовой вариабельностью (это гибкий объект) и невысокой межклассовой (несложно найти людей, чьи ладони визуальное сильно похожи).

Результаты на базе Poly HD значительно лучше, что согласуется с приближенными к идеальным условиями съёмки Poly HD (раздел 4.2.4) и свободным положением камеры при съёмке Базы-2 (раздел 4.2.2). При всём этом, спектр хорошо себя зарекомендовал на данных, полученных с помощью сканера [30].

Тестирование Габор-признаков, полученных по области кожных складок одного из пальцев, показало их хорошую информативность и устойчивость к изменению углу съёмки.

Классификация по Габор-признакам трёх пальцев (раздел 3.5.7, $\alpha_s = 0$) даёт существенное уменьшение ошибок по сравнению с классификацией по Габор-признакам лишь одного из пальцев. Совместное использование Габор- и SIFT-признаков трёх пальцев ещё немного уменьшает уровень ошибок.

С ростом числа людей (в Базе-2 содержатся изображения 48 людей, а в Poly HD — 165 людей) метод распознавания ожидаемо теряет в точности распознавания, сохраняя при этом показатели EER и средней ошибки идентификации менее 2%. Эти показатели сопоставимы с результатами лучших методов идентификации по кожным складкам внешней стороны пальцев, рассмотренных в обзорном разделе 2.3.3. Учитывая то, как была снята База-2 (раздел 4.2.2), это позволяет утверждать, что предложенный автором метод распознавания по изображению внешней стороны ладони на основе Габор- и SIFT-признаков уверенно проводит идентификацию в малых группах (<200 человек) с помощью камеры мобильного устройства при различных внешних условиях.

Предпринята попытка сконструировать функцию расстояния между изображениями ладони как линейную комбинацию расстояния на основе фильтров Габора и расстояния по одному из признаков формы. К сожалению, ни одно из таких комбинированных расстояний не даёт стабильного улучшения качества распознавания на Базе-2.

Оценка времени работы. Время выполнения основных этапов процедуры распознавания приведено в таблице 11. Стоит отметить, что для комфортного использования системы достаточно времени ответа от модуля распознавания в течение 1-2 секунд.

Если мы храним признаковое описание, построенное на основе 1000 изображений ладоней (этого количества достаточно для малой группы порядка 100 человек), то для сопоставления поступившего изображения со всеми эталонами с использованием трёх пальцев потребуется $204 + 30 + 3 * 44 + 3 * 0.3 * 1000 = 1266$ мс, что укладывается в обозначенные временные ограничения.

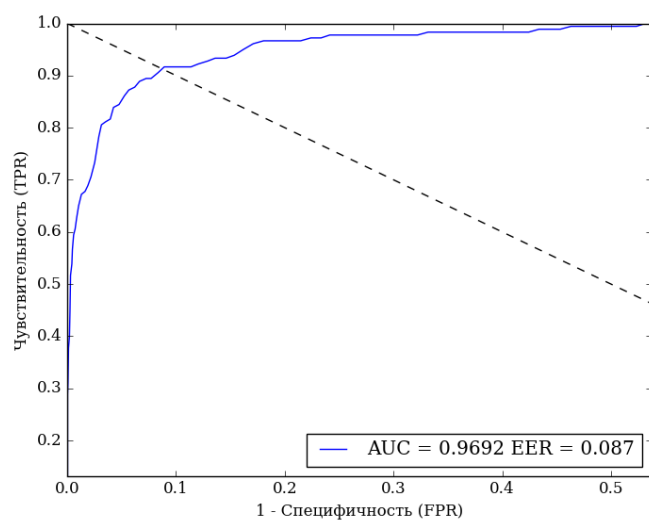
Этап	Время, мс
Точная сегментация ладони	204
Семантическая разметка ладони	30
Извлечение текстурных признаков (1 палец)	44
Сопоставления пары текстурных признаков (1 палец)	0.3

Таблица 11: Время работы основных этапов процедуры распознавания.

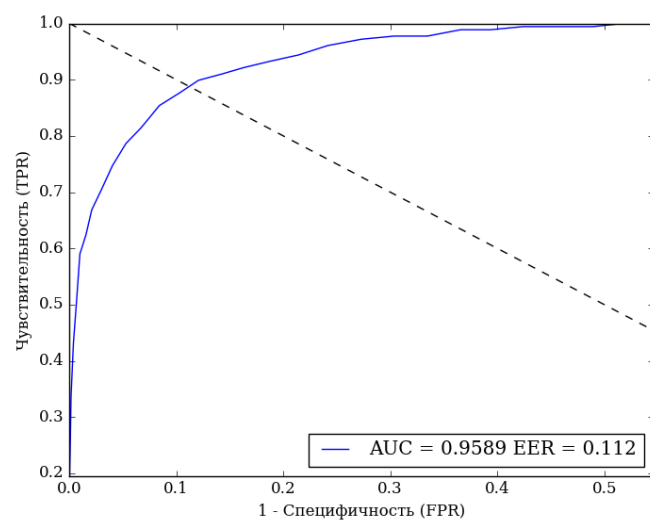
Пороги уверенной классификации. Алгоритм ближайшего соседа неустойчив к шумовым выбросам, поэтому в случае реальной системы при неуверенной классификации разумно отказаться от классификации вообще и запросить ещё одно изображение — такой сценарий реализован в мобильном приложении (раздел 5.2.1). Для этого используется порог уверенной классификации, выбор которого осуществляется следующим образом. Строится накопительная гистограмма попарных межклассовых расстояний для итоговой функции расстояний. В качестве порога выбирается x-координата последнего столбца, чьё значение не превышало заданный порог (выбрано $0.05 * S$, где S — сумма всех столбцов гистограммы).

Метод	Ошибка, % База-2	AUC База-2	EER, % База-2	Ошибка, % PolyU HD	AUC PolyU HD	EER, % PolyU HD
Кривизна	21.0	0.9692	8.7	10.6	0.9750	7.7
Ломаная	22.1	0.9589	11.2	12.6	0.9703	7.9
Спектр	24.3	0.9407	15.1	13.4	0.9672	9.8
Габор (палец 1)	5.7	0.9923	4.0	11.1	0.9734	6.5
Габор (палец 2)	6.3	0.9877	3.6	12.3	0.9723	8.4
Габор (палец 3)	6.1	0.9955	3.7	11.4	0.9757	7.1
Габор (три пальца)	1.4	0.9996	1.1	1.8	0.9990	1.3
Габор + SIFT (три пальца)	0.9	0.9997	0.7	1.4	0.9997	0.8

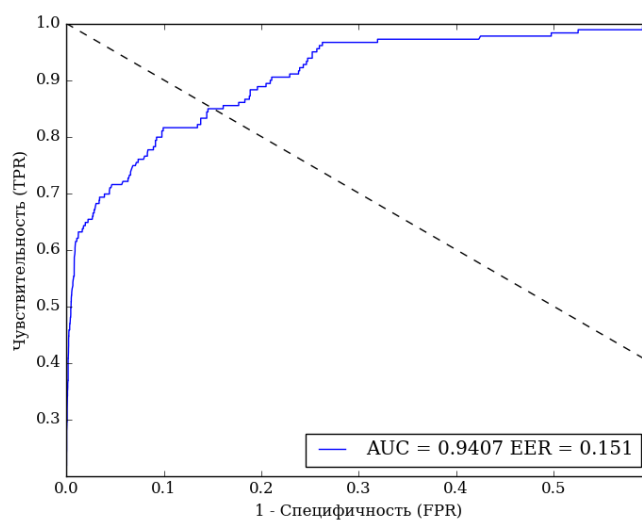
Таблица 10: Тестирование метода распознавания на Базе-2 и на базе PolyU HD при различных используемых признаках. Столбцы «Ошибка, %...» содержат ошибку идентификации, усреднённую по $p = 10$ случайным разбиениям базы на обучающую и тестовую части в соотношении 2: 1. Столбцы «EER, %...» и «AUC, %...» — содержат характеристики ROC-кривой, построенной по разбиению базы, на котором достигается медианное значение EER среди всех разбиений.



(a) Кривизна пальцев.

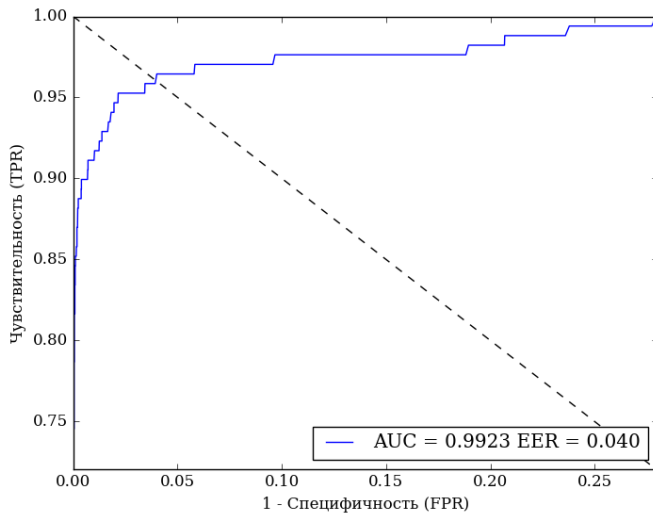


(b) Ломаная.

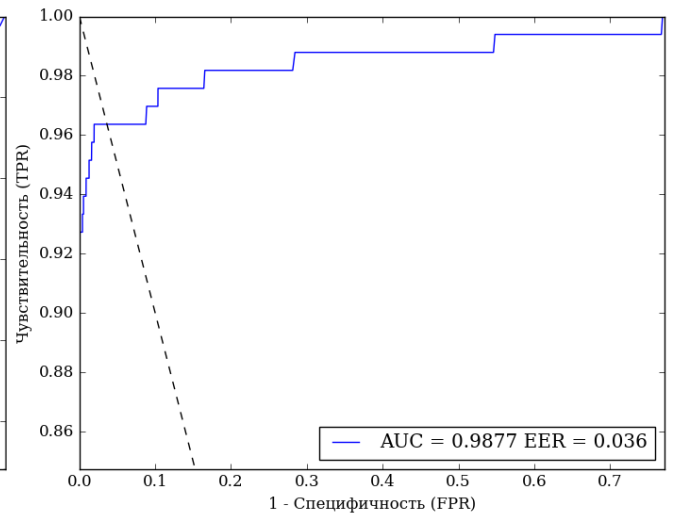


(c) Спектр.

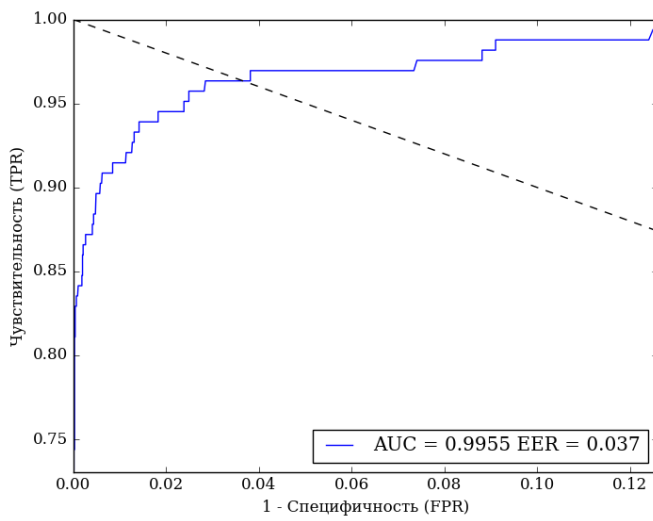
Рис. 42: ROC-кривые метода распознавания в зависимости от используемых признаков формы на Базе-2.



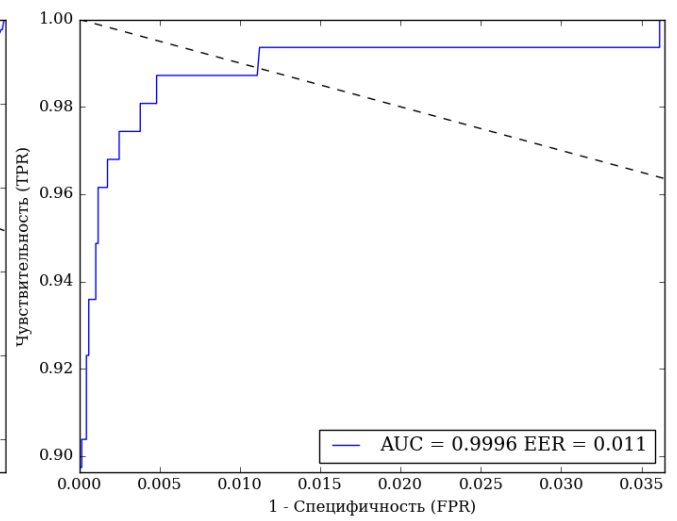
(a) Габор-признаки (палец 1).



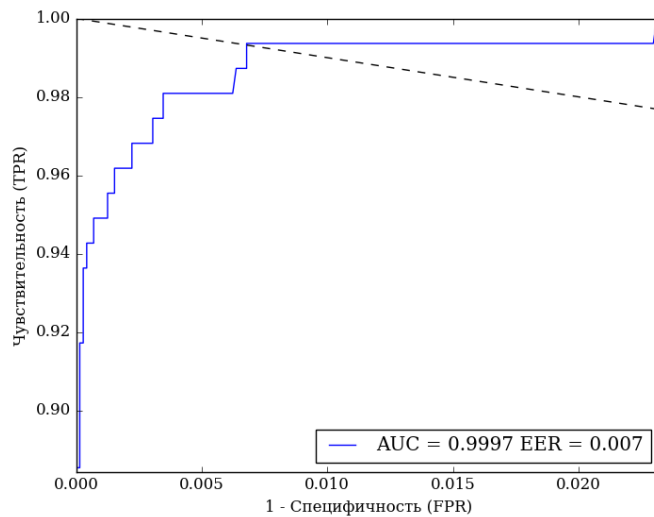
(b) Габор-признаки (палец 2).



(c) Габор-признаки (палец 3).

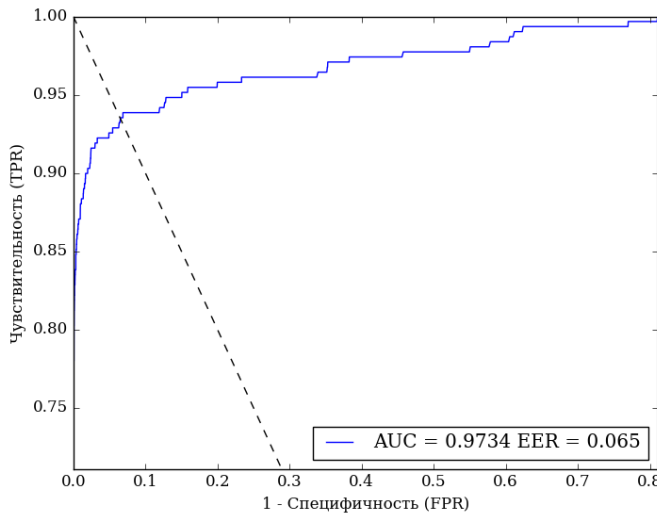


(d) Габор-признаки (три пальца).

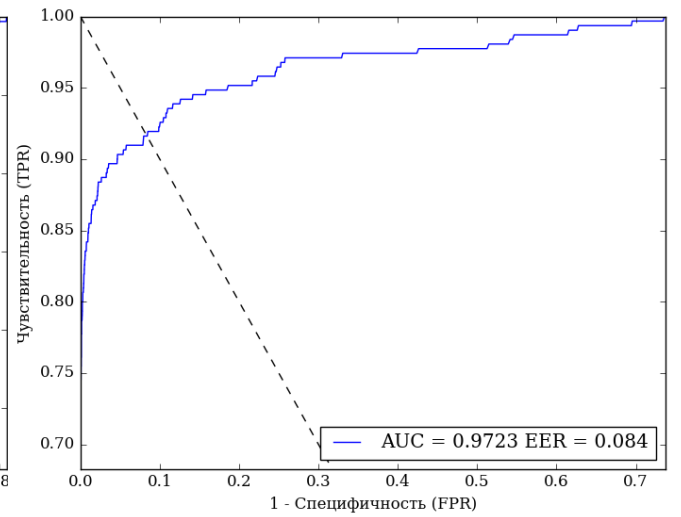


(e) Габор- и SIFT-признаки (три пальца).

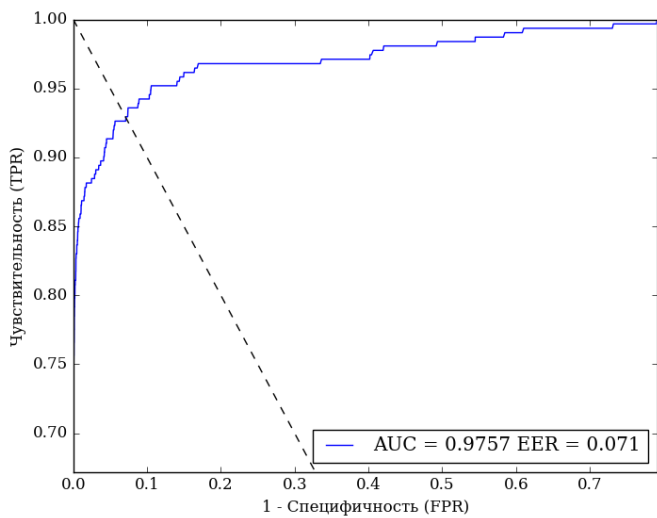
Рис. 43: ROC-кривые метода распознавания в зависимости от используемых текстурных признаков на Базе-2.



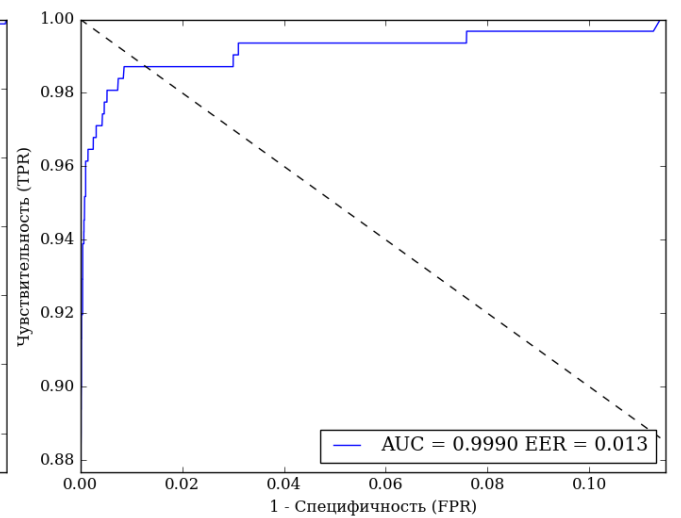
(a) Габор-признаки (палец 1).



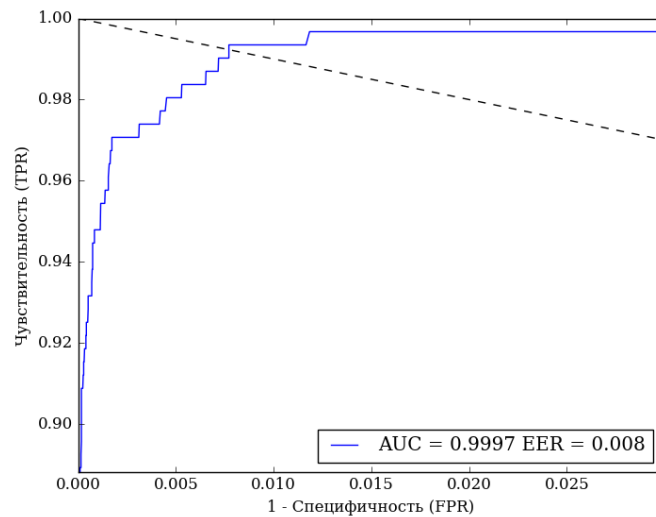
(b) Габор-признаки (палец 2).



(c) Габор-признаки (палец 3).



(d) Габор-признаки (три пальца).



(e) Габор- и SIFT-признаки (три пальца).

Рис. 44: ROC-кривые метода распознавания в зависимости от используемых текстурных признаков на базе PolyU HD.

4.4 Выводы к главе 4

Для вычислительных экспериментов с методом детектирования создана база изображений внешней стороны ладони (База-1), а также использована сторонняя база HGC. На обеих базах метод демонстрирует низкий уровень ошибок и высокую производительность, что делает его пригодным для работы с видеопотоком на мобильном устройстве в режиме реального времени. На базе HGC также проведено тестирование метода семантической разметки — полученные результаты сопоставимы с лучшими алгоритмами, участвовавшими в соответствующем состязании.

Для тестирования метода распознавания создана ещё одна база изображений внешней стороны ладони (База-2, съёмка «с руки»), дополнительно применена сторонняя база (PolyU HD, статичная съёмка). Проведённые вычислительные эксперименты показали высокую информативность и устойчивость предложенного автором признакового описания на основе фильтров Габора и дескрипторов SIFT.

Экспериментально обосновано использование двух семейств фильтров Габора для улучшения качества распознавания. Опытным путём доказано, что использование «точной» сегментации значительно увеличивает точность метода распознавания.

Достигаемые показатели точности метода распознавания сравнимы с результатами передовых методов идентификации по кожным складкам внешней стороны пальцев, рассмотренных в обзоре. Количество людей (48 и 165 соответственно) в использованных базах позволяет утверждать, что разработанный метод подходит для идентификации в малых группах.

5 Мобильная система биометрической идентификации

5.1 Общая характеристика системы

Практическим результатом исследовательской работы является разработка программного комплекса, реализующего предложенный в главе 3 метод биометрической идентификации личности по изображению внешней стороны ладони. При постановке цели исследования к системе было предъявлено ряд требований (раздел 1.2).

Наряду с **точностью** и **скоростью** (обеспечиваются характеристиками метода идентификации), одним из требуемых свойств системы является **доступность**. Это требование выполняется в силу использованных при проектировании системы компонентов: мобильное устройство на ОС Android сейчас есть практически у каждого, а арендовать обычный Linux-сервер можно просто и недорого.

Разбиение задачи идентификации по ладони на детектирование и распознавание (раздел 3.2) потребовало не только логическое, но и структурное разделение этих этапов между клиентом (мобильным устройством) и удалённым сервером. Связь между приложением и сервером осуществляется по беспроводному интернет-соединению. Подобное распределение задач позволило добиться идентификации в реальном времени при одновременном соблюдении всех прочих требований к системе идентификации.. Таким образом, разработанная система имеет клиент-серверную архитектуру, а мобильное устройство выступает в роли первичного сенсора.

Как характеристику, **гибкость** можно разделить на эксплуатационную гибкость и структурную гибкость.

В первом случае имеется ввиду стабильность процесса идентификации по отношению ко внешним условиям: работоспособность при разном освещении, съёмка как со штатива, так и с руки. Разработанная система может работать в обоих случаях благодаря использованию встроенной вспышки, аппаратного автофокуса и алгоритмов, устойчивых к масштабированию и поворотам изображения.

Структурная гибкость системы выражается в возможности распределения вычислительных модулей между мобильным приложением и сервером. Это позволяет выполнять сложные расчеты на более мощном сервере, оставляя на мобильном устройстве только пользовательский интерфейс и модуль детектирования. Благодаря такому распределению задач появляется возможность модифицировать алгоритмы распознавания, не изменяя мобильное приложение. Примером подобной модификации может служить добавление новых призна-

ков в модуль распознавания. Более того, данный программный комплекс несложно модифицировать для работы с другими модальностями на базе руки — например, с изображением внутренней стороны ладони. При дальнейшем развитии характеристик мобильных устройств вполне вероятно размещение на клиенте всех модулей системы идентификации (кроме базы биометрических признаков).

Что касается **удобства**, то тут ключевую роль играет интуитивно понятный пользовательский интерфейс мобильного приложения. Чтобы начать работу, необходимо лишь авторизоваться в системе и (по желанию) выбрать параметры работы приложения. Система может функционировать в двух режимах: обучения и тестирования (раздел 5.2.1). Процесс эксплуатации системы администратором предельно прост: подготовка приложения к работе с очередной группой распознаваемых занимает несколько секунд, сложности с переноской современных мобильных устройств в силу их скромных размеров также отсутствуют.

5.2 Мобильное приложение

В данном разделе даётся описание интерфейса мобильного приложения. В разделе 5.2.1 приводится общее описание процесса работы с системой. Далее в разделе 5.2.2 описан главный экран приложения и расположенные на нём элементы интерфейса. Описание меню настроек мобильного приложения приводится в разделе 5.2.3.

5.2.1 Принципы работы

Основные определения.

- *Съёмка* — получение изображений ладони с помощью камеры мобильного устройства. Используется видеопоток.
- *Пользователь* — человек, чьи изображения ладони участвуют в распознавании.
- *Сессия* — непрерывный и ограниченный по времени процесс съёмки ладони одного пользователя.
- *Результат сессии* — сообщение об итогах идентификации после окончания сессии. Может быть либо именем распознанного пользователя, либо отказом от классификации с указанием причины (невозможность уверенной классификации, отсутствие достаточного количества обработанных изображений и т. д.).

- *Группа* — именованное множество пользователей. Все пользователи, находящиеся в базе, должны состоять хотя бы в одной группе. Добавлять или удалять пользователей в список для распознавания можно только группами.
- *Администратор* — человек, имеющий учётную запись в системе. Администратор управляет мобильным приложением, также он имеет возможность создавать и редактировать группы и пользователей.
- *Валидный кадр* — кадр, содержащий валидную ладонь (раздел 3.4).
- *Невалидный кадр* — кадр, либо не содержащий ладони, либо содержащий невалидную ладонь.

Начало работы. Перед началом работы с системой администратору предлагается войти в свою учётную запись (при её отсутствии — зарегистрироваться), либо использовать существующий аккаунт Google (данная учётная запись необходима для установки любых приложений на мобильное устройство с ОС Android). Затем нужно создать хотя бы одну группу пользователей. Также можно настроить параметры работы приложения и камеры (раздел 5.2.3). После этого система готова к работе в режиме обучения (для функционирования режима тестирования необходимо наличие группы с хотя бы одним пользователем с ненулевым количеством изображений — раздел 5.2.1). Также стоит отметить, что администратор может использовать несколько устройств, все настройки синхронизируются между ними.

Сессии. Основным при работе с системой является понятие сессии. Сессия — это непрерывный и ограниченный по времени процесс съёмки ладони одного человека (раздел 5.2.1). Начать съёмку можно нажатием на кнопку «Старт» (раздел 5.2.2). Сессия, в свою очередь, начинается, когда в процессе съёмки встретилось N_{start} кадров ($N_{start} \approx 5$) подряд с валидной ладонью. Во время сессии выполняется определение наличия валидной ладони в каждом кадре, и, если такая ладонь найдена, изображение отправляется на сервер для распознавания. Экран приложения во время сессии показан на рис. 45. Концептуальная схема работы сессии приведена на рис. 46.

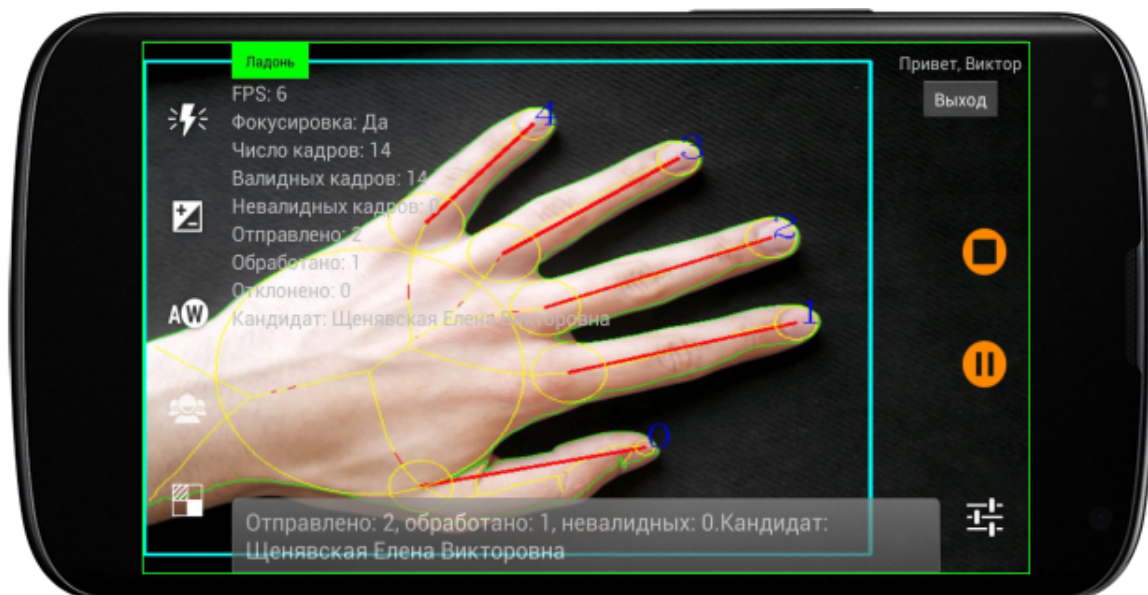


Рис. 45: Экран сессии.

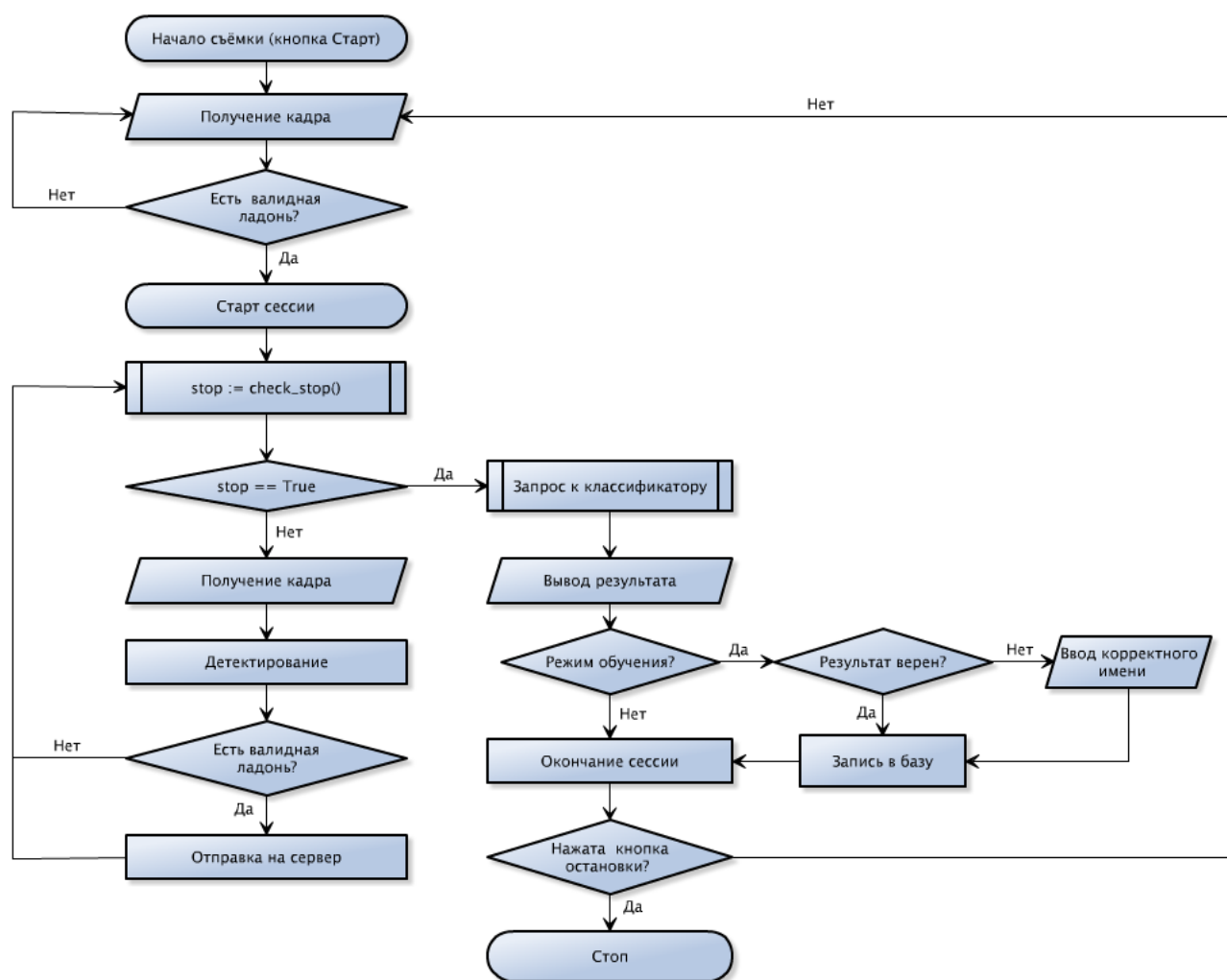


Рис. 46: Схема сессии. Здесь `check_stop()` — процедура проверки условий остановки сессии, она устанавливает флаг окончания сессии либо в положение `true` (сессия завершается) или `false` (сессия продолжается).

Сессия завершается при выполнении хотя бы одного из следующих условий:

- N_{conf} (≈ 3) изображений было успешно обработано сервером;
- на сервер отправлено N_{max} (≈ 3) изображений;
- встретилось подряд $N_{invalid}$ (≈ 30) невалидных кадров;
- с последнего валидного кадра прошло $T_{invalid}$ (≈ 4) секунд.

Схема проверки этих условий показана на рис. 47.

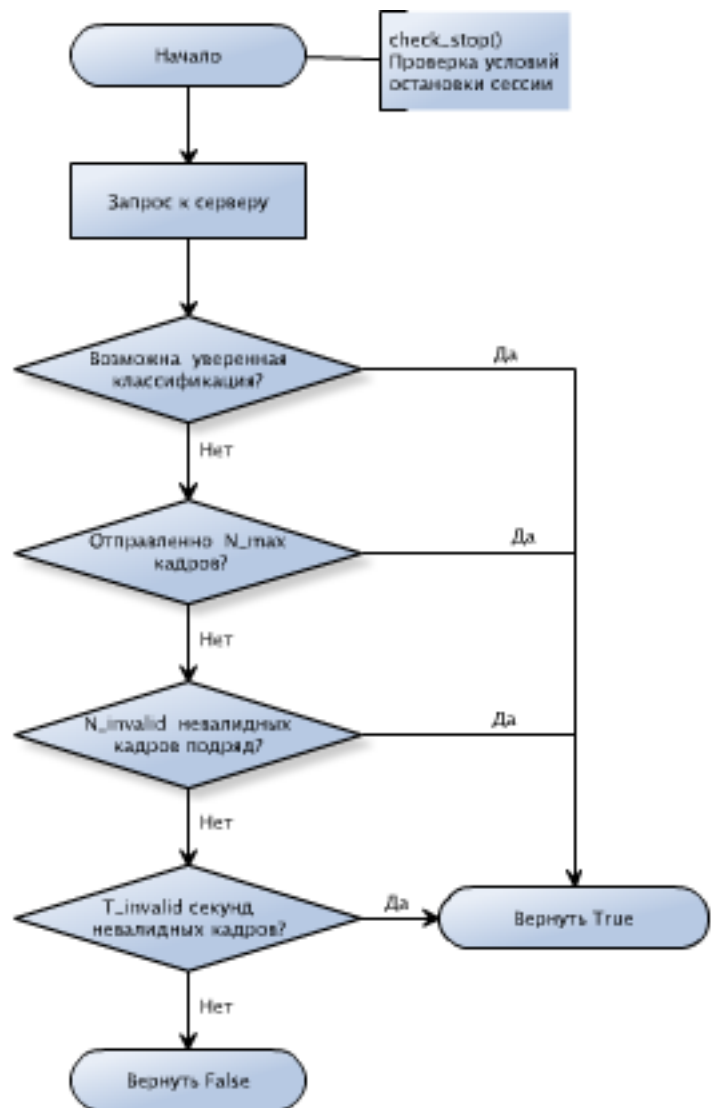


Рис. 47: Схема проверки условий остановки сессии.

Есть возможность приостановить текущую сессию, нажав на кнопку «Пауза сессии». Если сессия не была возобновлена спустя T_{pause} (≈ 2) минут, она завершается. Также можно принудительно завершить текущую сессию, нажав на кнопку «Остановка сессии». Когда

сессия завершена, администратору всегда выдаётся результат (раздел 5.2.1). После ответа администратора начинается новая сессия.

Результат сессии. Если классификация проведена успешно, система выдаёт имя распознанного пользователя. Существует несколько причин отказа от классификации:

- не было успешно обработано N_{conf} снимков;
- расстояние между снимками и их ближайшими соседями больше порога уверенной классификации.

В этом случае система выдаёт сообщение «Отказ от классификации» с указанием причины (рис. 48) и начинает новую сессию.

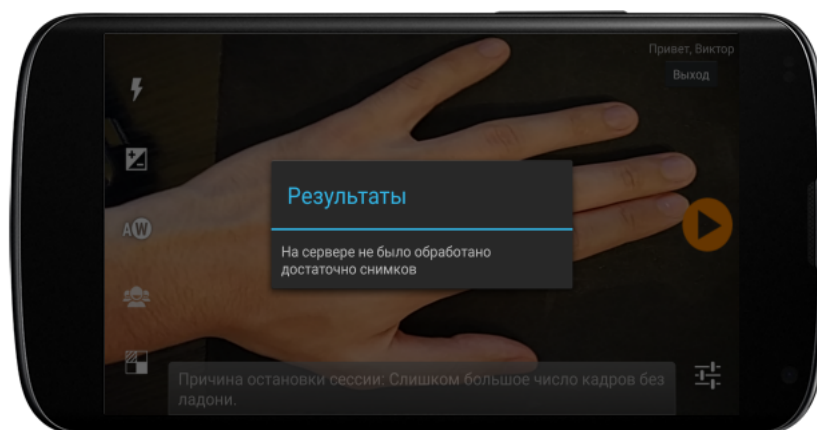


Рис. 48: Результат идентификации при отклонении большого количества кадров.

Далее, в режиме обучения (раздел 5.2.1) предлагается либо подтвердить правильность классификации, либо, если система ошиблась, ввести истинное имя пользователя (рис. 49 и 50). Чтобы предотвратить опечатки в именах пользователей и другие ошибки, администратору необходимо подтвердить свой выбор.

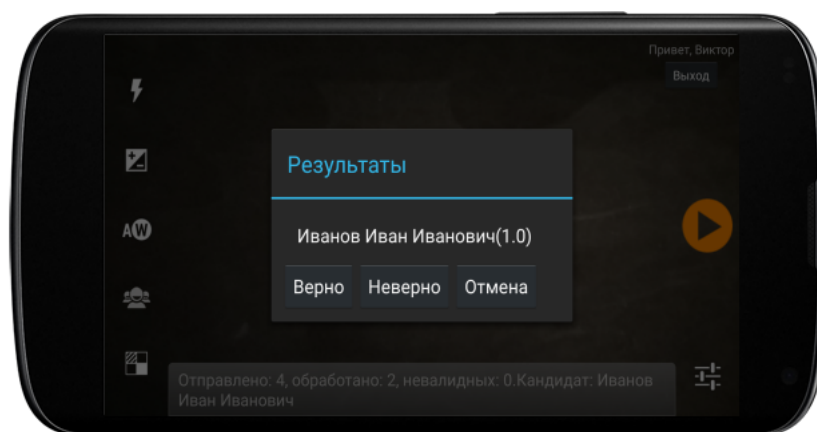


Рис. 49: Результат идентификации в режиме обучения.

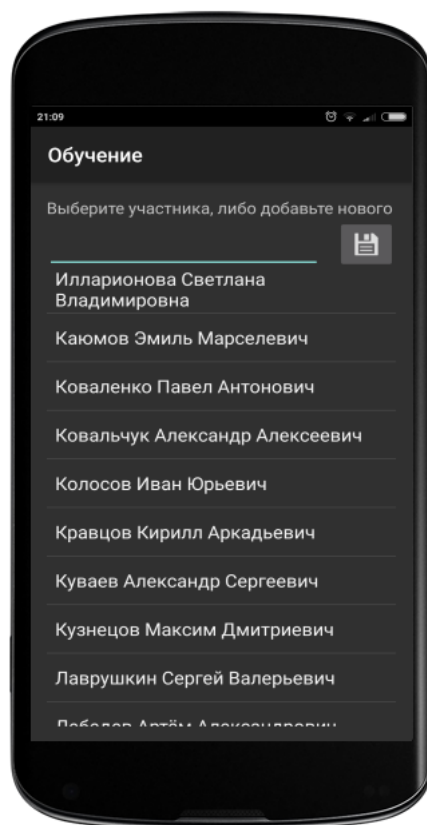


Рис. 50: Выбор другого пользователя.

В режиме тестирования действий администратора не требуется (рис. 51).

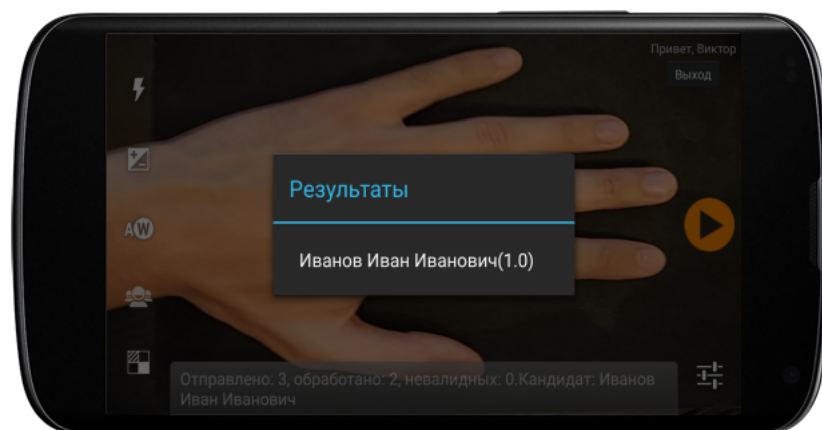


Рис. 51: Результат сессии в режиме тестирования.

Режимы. У системы существуют два основных режима: обучение и тестирование.

В режиме обучения происходит пополнение базы прецедентов (изображений ладони). Выполняется съёмка, при поступлении каждого нового изображения с валидной ладонью проводится классификация этого изображения на основе уже существующей (возможно пустой) базы эталонов, на экран выводится результат классификации. Если ладонь распознана правильно, администратор может добавить её в базу, иначе предлагается ввести имя другого

пользователя (в случае отсутствия его в базе, создаётся новый) и приписать это изображение ему.

Режим тестирования отличается от режима обучения тем, что нет возможности добавлять новых пользователей и снимаемые изображения ладоней не сохраняются в базе. Администратору лишь выводится результат идентификации.

Стоит отметить, что возможна идентификация по нескольким группам одновременно. В этом случае сравнение нового изображения происходит со всеми изображениями выбранных групп. В режиме обучения текущая картинка добавляется в базу для выбранного пользователя в каждой группе. Если этого пользователя нет в какой-либо из выбранных групп, создаётся новый пользователь.

5.2.2 Главный экран

На главном экране приложения расположены основные элементы интерфейса, также на него выводится изображение с камеры устройства и визуализация во время сессии. Когда сессия не начата, активны следующие кнопки (рис 52).

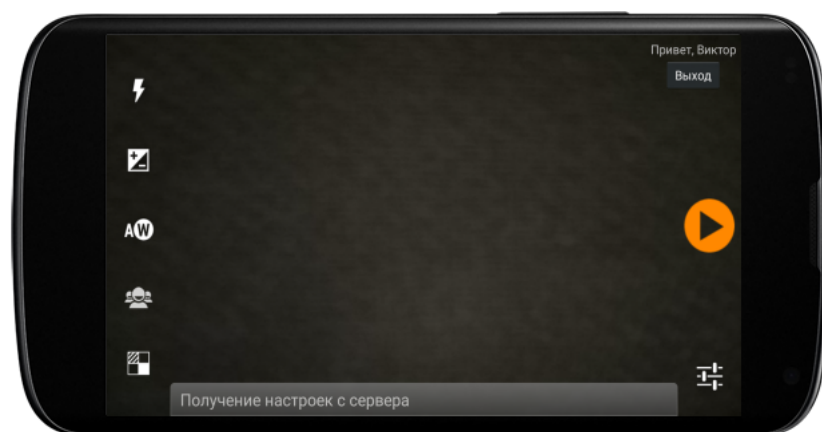


Рис. 52: Главный экран приложения.

Опции камеры. Позволяют настроить использование вспышки, коррекцию экспозиции и баланс белого.

Меню групп. Позволяет управлять группами. Можно выбрать одну или несколько групп из существующих либо создать новую (рис 53). Также есть возможность вывести список пользователей группы. Для каждого пользователя пишется количество изображений, участвующих в распознавании. Поддерживается редактирование пользователей (изменение имени, удаление изображений, удаление пользователя).



Рис. 53: Меню управления группами.

Выбор режима. Выбор между режимами тестирования и обучения (рис 54).

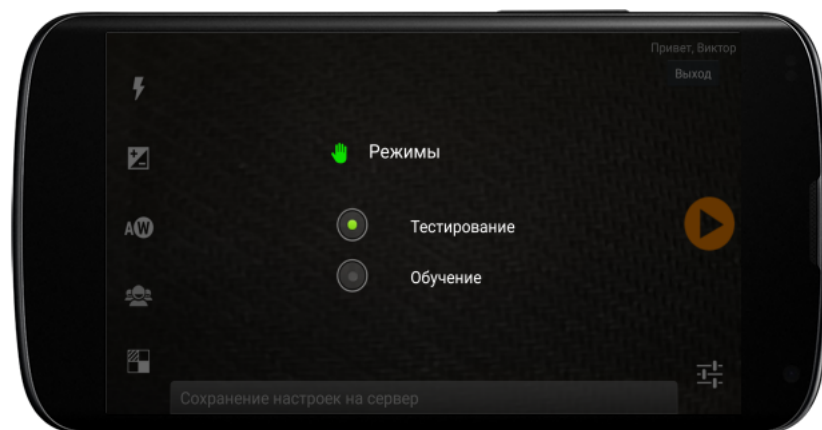


Рис. 54: Меню выбора режима.

Кнопка смены учётной записи. Возможность выйти из текущей учётной записи и зайти с помощью другой.

Старт. Кнопка начала сессии.

Меню настроек. Доступ к основным настройкам как системы в целом, так и мобильного приложения (раздел 5.2.3).

Когда идёт сессия, на главном экране отображается следующая информация:

- Количество кадров с валидной ладонью,
- Число отправленных на сервер кадров.

Помимо этого выводится изображение с камеры устройства, на которое наложена визуализация результатов процедуры детектирования (только в случае наличия валидной ладони), а также это изображение обрамлено в прямоугольник зелёного цвета в случае наличия валидной ладони и красного в противном случае. Также отображаются кнопки паузы и остановки сессии (рис. 45). Во время сессии доступ к выбору режима, управлению группами и к настройкам закрыт.

5.2.3 Настройки

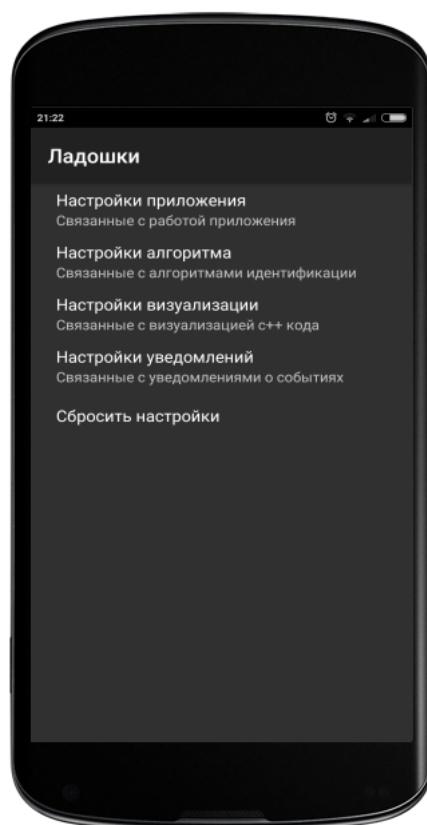


Рис. 55: Меню настроек приложения.

Все настройки синхронизированы с сервером таким образом, что каждый администратор имеет единые настройки для всех своих устройств. Меню настроек содержит несколько

подменю, отвечающих за параметры различных частей системы. Рассмотрим каждое из них подробнее.

Настройки приложения. В этом меню можно настроить таймауты сессии (время завершения сессии при сворачивании приложения) и диалоговых окон. Также есть возможность выбрать разрешения изображений, используемых для детектирования на устройстве и при распознавании на сервере.

Настройки алгоритма. Здесь настраиваются параметры сессии, такие как максимальное число отправляемых на сервер изображений, максимальное количество кадров без валидной ладони и т. д. (табл. 12).

Параметр	Значение по умолчанию	Описание
N_{start}	5	число подряд идущих кадров с корректной ладонью, необходимое для начала сессии
N_{conf}	3	количество изображений, необходимое для уверенной классификации
N_{max}	3	максимальное число изображений, отправляемое на сервер за время сессии
$N_{invalid}$	30	количество некорректных кадров подряд, при котором сессия завершается
$T_{invalid}$	4	время в секундах после последнего корректного кадра, после которого сессия завершается
T_{pause}	2	максимальное время приостановки сессии в минутах

Таблица 12: Параметры сессии.

Настройки визуализации. В этом подменю можно выбрать, визуализировать ли внутреннюю структуру алгоритма детектирования. Это снижает скорость работы приложения, но позволяет пользователю лучше расположить ладонь перед камерой для более точного распознавания.

Настройки уведомлений. Здесь можно настроить параметры уведомлений о различных событиях, таких как отправка снимка на сервер, начало сессии, окончание сессии и наличие

большого количества некорректных кадров в текущей сессии. Уведомления реализованы с помощью звуковых сигналов, что призвано упростить взаимодействие с приложением.

Также в меню есть кнопка «Сбросить настройки», возвращающая все настройки в положение по умолчанию.

5.3 Сервер

Вторым ключевым компонентом разработанной системы является сервер. К его задачам относятся классификация изображений ладони и хранение базы биометрических прецедентов. Общая схема работы сервера показана на рис. 56. Основными компонентами сервера являются: модуль распознавания, база данных, очередь задач, интерфейс взаимодействия с сервером. Рассмотрим подробнее каждый из них.

Модуль распознавания.

1. Сервер получает валидные изображения от мобильного приложения во время сессии посредством интернет-соединения.
2. Для каждого изображения извлекаются текстурные признаки (раздел 3.5).
3. Классификатор 1NN сравнивает признаковое описание поступившего кадра с описаниями из базы эталонов, после чего в мобильное приложение отсылается результат идентификации.

База данных. Так как данные в системе имеют заранее определённую структуру, использовалась реляционная база данных. Подобная модель данных обеспечивает высокое быстродействие, простоту поддержки и реализации. В качестве СУБД была выбрана PostgreSQL. ER-диаграмма таблиц базы данных приведена на рис. 57.

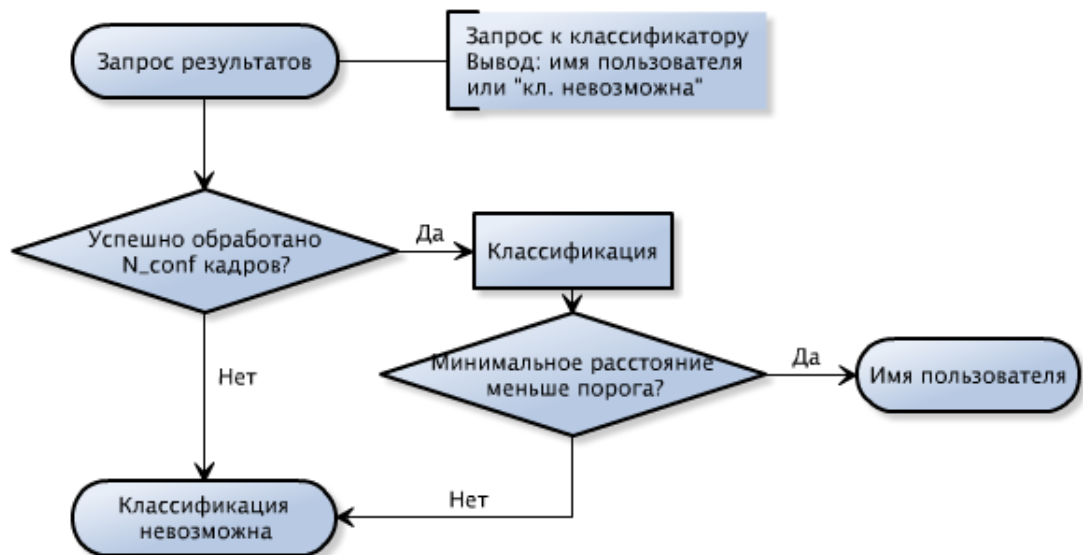
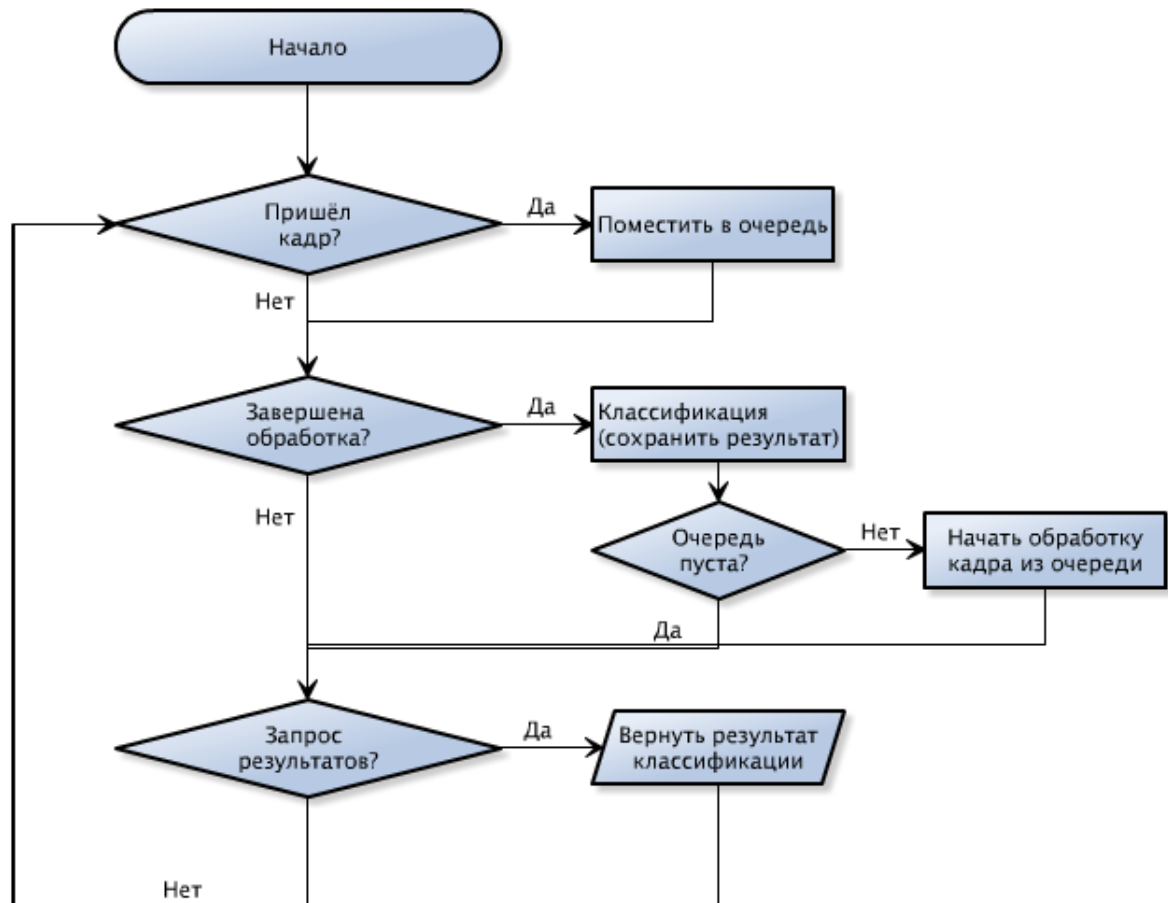


Рис. 56: Схема работы сервера.

api palmimage — содержит информацию об изображениях (пользователя, к которому относится изображение, вычисленные признаки), отправленных на сервер во время сессий;

api session — содержит информацию о сессиях;

custom user emailuser — содержит информацию об администраторах.

Очередь задач. Сервер является асинхронным, т. е. может выполнять несколько задач одновременно. Для этого используется очередь задач, распределяющая работу между несколькими вычислителями (каждый из них выполняется в отдельном потоке). Например, когда приходит новое изображение от мобильного приложения, задача его обработки помещается в очередь. Свободный вычислитель «забирает» себе эту задачу и выполняет её в своём потоке, в то время как сервер продолжает работу. По завершении задачи, вычислитель сохраняет результат выполнения в базу данных и готов выполнять новую.

Интерфейс взаимодействия с сервером (API). При разработке интерфейса сервера использовался современный подход, называемый REST (Representational State Transfer). Данный подход устанавливает жёсткие правила взаимодействия компонентов системы, что позволяет упростить архитектуру и разработку.

Один из примеров реализации REST API — отсутствие хранимого состояния сервера. Это значит, что вся информация, необходимая серверу для обработки запроса клиента, содержится в самом запросе. Таким образом, клиент не должен знать состояние сервера и наоборот. Этот метод делает сервер независимым от клиентов, что даёт возможность сравнительно просто добавить веб-интерфейс, не изменяя сервер. Также это позволяет свободно менять мобильное приложение, не внося изменений в код сервера.

Для обмена информацией между частями системы (сервер, клиент, веб-интерфейс) используется формат JSON. Это текстовый формат данных, представляющий собой набор пар ключ-значение. Он является универсальным способом передачи информации между разнородными модулями и не зависит от реализации каждого из них.

5.4 Веб-интерфейс

Вся биометрическая информация, полученная при работе с системой, хранится на сервере. Доступ администраторов к ней осуществляется через веб-интерфейс (рис. 58).

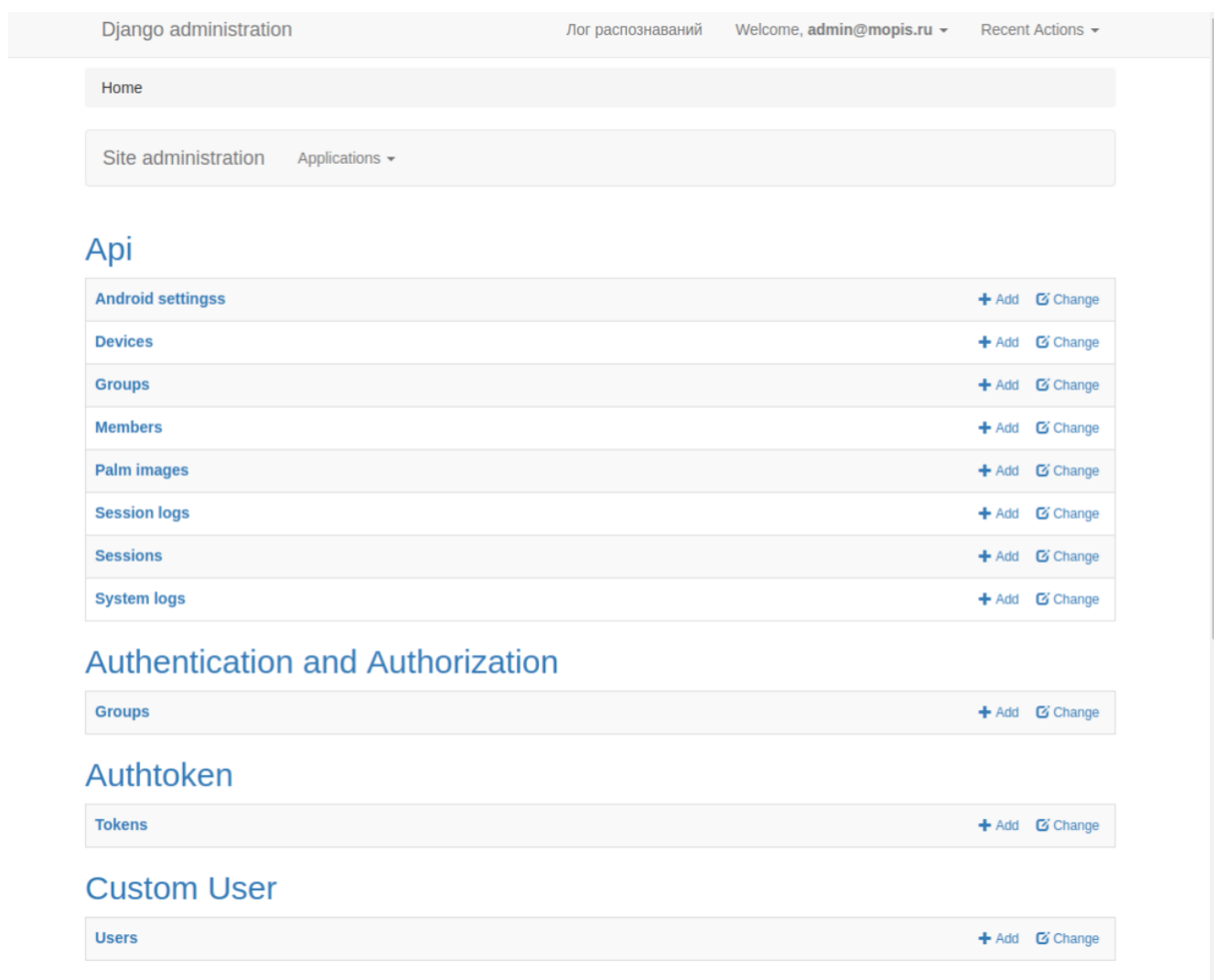


Рис. 58: Веб-интерфейс сервера. Главный экран.

С его помощью администраторы имеют возможность:

- редактировать информацию о своей учётной записи;
- редактировать пользователей (добавлять и удалять изображения, переименовывать и т. д.);
- создавать и удалять группы;
- редактировать группы (добавлять и удалять пользователей и т. д.);
- просматривать результаты сессий, в т. ч. отправленные на сервер изображения, визуализации и т. д.;
- изменять настройки приложения и алгоритма классификации.

Стоит отметить, что при добавлении изображений выполняется вычисление признаков и сохранение их в базу. В ходе этого процесса создаются визуализации, которые можно посмотреть в соответствующем меню (рис 59).

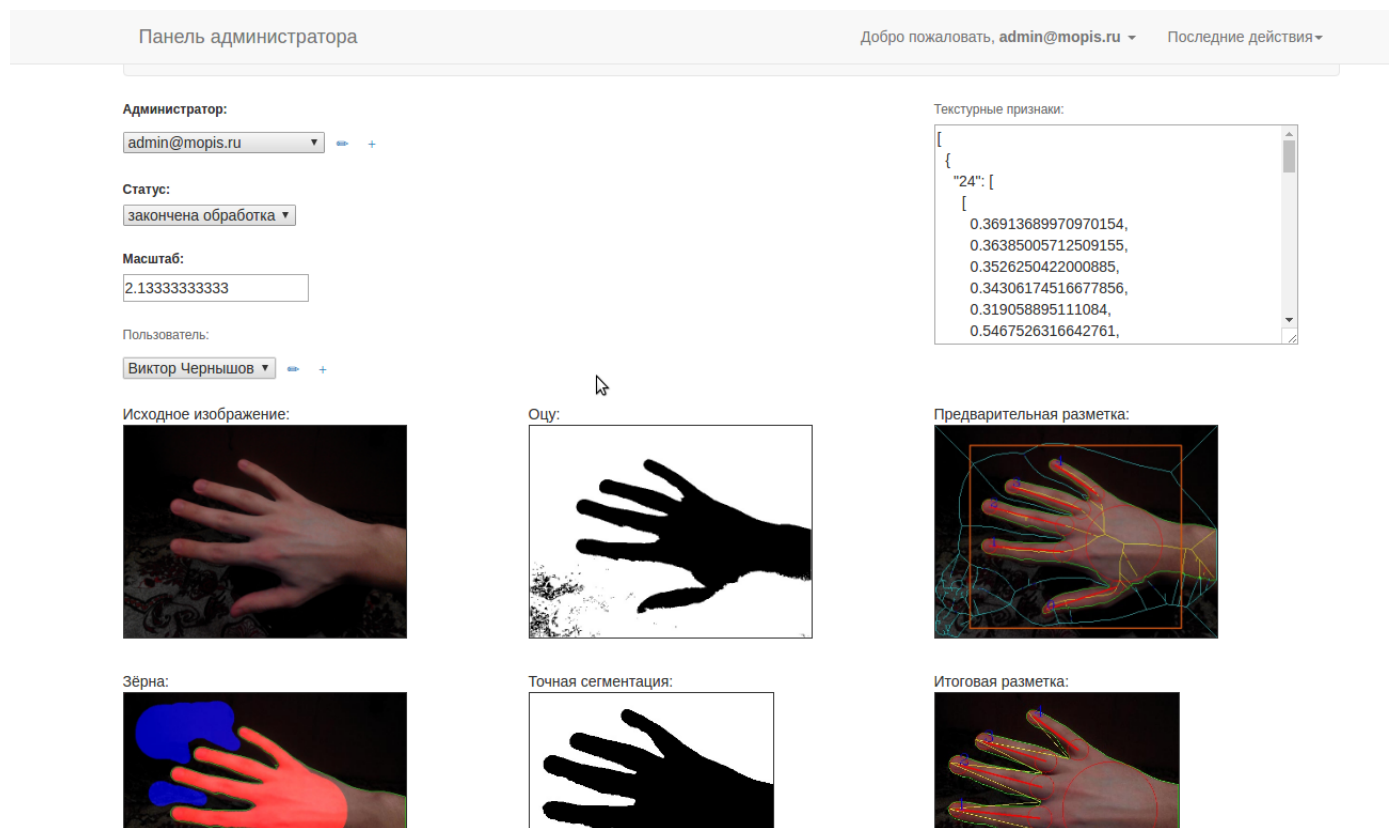


Рис. 59: Веб-интерфейс сервера. Редактирование изображений.

Благодаря системе учётных записей, администраторы могут изменять только свои группы и пользователей, не имея возможности изменять данные других администраторов.

Таблица посещений. Ключевым и самым часто используемым компонентом веб-интерфейса является таблица посещений. В ней отражается статистика распознаваний пользователей по датам. По строкам расположены пользователи, по столбцам — даты распознаваний. В ячейках таблицы первое число обозначает общее количество сессий распознавания для данного пользователя в данный день. Число в скобках — общее количество изображений в этих сессиях. Сессия считается успешной, если пользователь был распознан верно, и не успешной в противном случае. Цвет ячейки обозначает долю успешных сессий (если все успешны — зелёный, если все не успешны — красный, иначе — жёлтый). Пример части таблицы приведён на рис. 60. Также существует возможность фильтрации строк и столбцов таблицы по датам, группам и режимам.

Фильтрация

Пользователь	10.03.2016, Чт	17.03.2016, Чт	07.04.2016, Чт	20.04.2016, Ср	19.05.2016, Чт
Айсина Роза Мунеровна	3(5)	3(3)	1(1)	1(2)	1(1)
Алескин Александр Сергеевич	3(3)	1(1)	1(2)	1(2)	1(2)
Амир Мирас Сабыргалиулы	2(3)	1(1)	0(0)	1(1)	1(2)
Арсеев Сергей Петрович	2(2)	2(3)	1(1)	1(2)	2(4)
Белобородов Дмитрий Сергеевич	0(0)	2(3)	1(1)	1(2)	1(1)
Бобров Евгений Александрович	3(6)	1(2)	1(2)	1(2)	1(2)
Бротиковская Данута Станиславовна	2(2)	1(1)	1(1)	1(1)	2(3)
Быковец Евгений Владимирович	3(4)	2(4)	1(1)	1(2)	1(2)

Рис. 60: Таблица посещений.

5.5 Особенности разработки кода

В силу того, что компоненты системы запускаются в существенно различных средах, для разработки приложения и сервера использовались несколько языков программирования: C++, Java и Python. Мобильное приложение написано на Java, так как это основной язык разработки для ОС Android. При создании модуля детектирования для приложения использовался C++, так как он обеспечивает высокую скорость работы.

Для реализации сервера использовался Python из-за своей гибкости, удобства и простоты тестирования. Кроме того, хотя этот язык и медленнее таких языков как Java и C++, его скорости достаточно для работы сервера. Код модуля распознавания также написан на C++, так как в нём имеются вычислительно сложные этапы (например, точная сегментация).

Таким образом, возникла необходимость в компиляции C++-кода одновременно для системы Android и для Linux. Для этого использовалась кросс-платформенная система автоматизации сборки — CMake. Эта система скрывает низкоуровневые системно-зависимые детали процесса сборки, позволяя писать единые правила сборки для различных платформ.

Кроме того, модули детектирования и распознавания активно используют следующие библиотеки:

Boost [70] — для построения диаграмм Вороного, операций с графами, работы с файловой системой, измерения времени;

OpenCV [71] — для операций с изображениями, в т. ч. бинаризации и морфологических операций, а также для визуализации.

5.6 Выводы к главе 5

В связи с тем, что в научной литературе не представлено систем биометрической идентификации личности человека по изображению внешней стороны ладони (раздел 2.5), удовлетворяющих поставленным требованиям (раздел 1.2), был разработан соответствующий программный комплекс на базе мобильного Android-устройства.

Программное обеспечение состоит из нескольких основных компонентов: Android-приложения, сервера, веб-интерфейса для администратора.

Следует отметить, что код модулей детектирования (клиент) и распознавания (сервер) также использовался для вычислительных экспериментов (глава 4).

Эффективная реализация всех программных компонентов позволила создать систему, которая успешно прошла апробацию в качестве контроля посещаемости студентов МГУ и учёта посетителей на XIII Московском Фестивале науки NAUKA 0+.

6 Заключение

В работе представлен метод биометрической идентификации человека по изображению внешней стороны ладони, пригодный для использования в системе распознавания на базе мобильного устройства.

Проведена декомпозиция исходной задачи идентификации на задачи детектирования и распознавания. Предпосылками для этого являются ограниченная вычислительная мощность мобильных устройств и обеспечение безопасности хранимых биометрических признаков.

Ладони, пригодные для дальнейшего распознавания, определены через понятие корректности. В предложенном методе детектирования можно выделить ряд основных шагов:

1. обнаружение объекта-кандидата на роль ладони в кадре (на основе «быстрой» сегментации по Оцу);
2. создание непрерывного медиального представления формы ладони, построение семантической разметки и проверка ладони на корректность;
3. определение расфокусированных кадров;
4. формирование данных для отправки в модуль распознавания (если требуется).

Разработанный метод распознавания сводится к следующему:

1. «точная» сегментации ладони на основе разрезов в графах и априорной информации с этапа детектирования;
2. локализация областей кожных складок на пальцах на основе семантической разметки ладони;
3. извлечение текстурных признаков кожных складок пальцев на основе двух семейств двумерных фильтров Габора, позволяющих эффективно закодировать текстурные особенности разных ориентаций и характерных размеров;
4. построение специальной меры сходства изображений ладони, объединяющей результаты сопоставления областей кожных складок каждого из пальцев по SIFT-дескрипторам и Габор-признакам.

Предложенный метод идентификации воплощён в виде программного комплекса. Для тестирования метода детектирования собрана собственная база изображений внешней сто-

роны ладони (База-1), а также использована сторонняя (HGC). В обоих случаях метод продемонстрировал низкие показатели ошибок и высокую скорость работы, что подтверждает целесообразность его использования в мобильном клиенте при анализе видеопотока в реальном времени.

Для тестирования метода распознавания создана ещё одна база изображений внешней стороны ладони (База-2, съёмка «с руки»), дополнительно применена сторонняя (PolyU HD, статичная съёмка). Проведённые вычислительные эксперименты показали высокую точность метода распознавания как на задаче верификации, так и на задаче идентификации. Это доказывает эффективность предложенного метода распознавания, а также служит обоснованием того, что рисунок кожных складок внешней стороны пальцев является индивидуальным признаком личности человека, позволяющим проводить идентификацию в малых группах — даже в случае съёмки внешней стороны ладони «с руки» при помощи мобильного устройства.

В рамках исследования разработана сама система биометрической идентификации на базе мобильного устройства по изображению внешней стороны ладони, которая была использована на практике для контроля посещаемости студентов и учёта посетителей научного фестиваля. В ходе апробации продемонстрирована уверенная работа идентификации в режиме реального времени при съёмке «с руки» в малых группах (до 100 человек). Программное обеспечение состоит из нескольких основных компонентов: Android-приложения, сервера, веб-интерфейса для администратора. Получившаяся система является точной, быстрой, доступной, социально приемлемой и гибкой — т.е. удовлетворяет всем целевым характеристикам. Мобильные системы, работающие с кожными складками пальцев, крайне слабо представлены в научной литературе — и среди них нет тех, что удовлетворяют целевым характеристикам. Это позволяет утверждать, что подобная система была реализована впервые.

Можно выделить следующие направления дальнейшей работы:

1. исследование возможности надёжной локализации области кожных складок внешней стороны пальца без предварительной «точной» сегментации ладони;
2. поиск новых методов генерации информативного признакового описания для области кожных складок пальца;
3. перенос генерации признакового описания на клиент (мобильное устройство).

Список литературы

1. *Chernyshov Victor*. Real-Time Hand Detection Using Continuous Skeletons // OGRW 2014 - Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding, Koblenz, Germany, 1-5 December, 2014. — 2014. — Pp. 200–204.
2. *Chernyshov Victor*. Efficient Hand Detection on Client-server Recognition System // VISAPP 2015 - Proceedings of the 10th International Conference on Computer Vision Theory and Applications, Berlin, Germany, 11-14 March, 2015. — Vol. 2. — 2015. — Pp. 461–468.
3. *Чернышов В. Г.* Задачи детектирования и распознавания ладони в клиент-серверной системе биометрической идентификации // Сборник тезисов конференции «Техническое Зрение в Системах Управления-2015», г. Москва, Россия, 17-19 марта 2015 г.
4. *Чернышов В. Г., Местецкий Л. М.* Использование текстурных особенностей костяшек пальцев в клиент-серверной системе биометрической идентификации // Сборник тезисов 17-ой конференции «Математические Методы Распознавания Образов» (ММРО-17), г. Светлогорск, Россия, 19-25 сентября 2015 г.
5. *Чернышов В. Г.* Мобильная система контроля посещений по геометрии и текстуре ладони // Сборник трудов 26-ой международной конференции по компьютерной графике и зрению «ГрафиКон 2016», г. Нижний Новгород, Россия, 19-23 сентября 2016 г. — 2016. — С. 373–377.
6. *Чернышов В. Г.* Метод биометрической идентификации по изображению внешней стороны ладони // Сборник тезисов 18-ой конференции «Математические Методы Распознавания Образов» (ММРО-18), г. Таганрог, Россия, 9-13 октября 2017 г.
7. *Chernyshov V. G., Mestetskii L. M.* Mobile Computer Vision System for Hand-Based Identification // *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*. — 2015. — Vol. 25, no. 2. — Pp. 209–214.
8. *Chernyshov V., Mestetskiy L.* Real-time hand detection using continuous skeletons // *Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*. — 2016. — Vol. 26, no. 2. — Pp. 368–373.
9. *Chernyshov V., Mestetskiy L.* Mobile Machine Vision System for Palm-Based Identification // Proceedings of the 11th International Conference on Pattern Recognition and Image Analysis:

- New Information Technologies (PRIA-11-2013), Samara, Russia, 23-28 September, 2013. — Vol. 2. — IPSI RAS Samara, 2013. — Pp. 398–401.
10. *Mali Kalyani, Bhattacharya Samyita*. Comparative study of different biometric features // *International Journal of Advanced Research In Computer and Communication Engineering*. — 2013. — Vol. 2, no. 7.
 11. *Jaswal Gaurav, Kaul Amit, Nath Ravinder*. Knuckle Print Biometrics and Fusion Schemes—Overview, Challenges, and Solutions // *ACM Computing Surveys (CSUR)*. — 2016. — Vol. 49, no. 2. — P. 34.
 12. *Daugman John*. The importance of being random: statistical principles of iris recognition // *Pattern recognition*. — 2003. — Vol. 36, no. 2. — Pp. 279–291.
 13. *Karmakar Dhiman, Murthy CA*. Face Recognition using Face-Autocropping and Facial Feature Points Extraction // *Proceedings of the 2nd International Conference on Perception and Machine Intelligence / ACM*. — 2015. — Pp. 116–122.
 14. Recent developments in voice biometrics: Robustness and high accuracy / Nicolas Scheffer, Luciana Ferrer, Aaron Lawson et al. // *Technologies for Homeland Security (HST), 2013 IEEE International Conference on / IEEE*. — 2013. — Pp. 447–452.
 15. *Greitans Modris, Pudzs Mihails, Fuksis Rihards*. Palm vein biometrics based on infrared imaging and complex matched filtering // *Proceedings of the 12th ACM Workshop on Multimedia and Security / ACM*. — 2010. — Pp. 101–106.
 16. FVC2000: Fingerprint verification competition / Dario Maio, Davide Maltoni, Raffaele Cappelli et al. // *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. — 2002. — Vol. 24, no. 3. — Pp. 402–412.
 17. *Shi W, Rong G, Zhang David*. Automatic Palmprint Verification // *International Journal of Image and Graphics*. — 2001. — Vol. 1, no. 1. — Pp. 135–152.
 18. *Zhang Dapeng, Shu Wei*. Two novel characteristics in palmprint verification: datum point invariance and line feature matching // *Pattern Recognition*. — 1999. — Vol. 32, no. 4. — Pp. 691–702.
 19. Online Palmprint Identification / David Zhang, Wai-Kin Kong, Jane You, Michael Wong // *IEEE Trans. Pattern Anal. Mach. Intell.* — 2003. — sep. — Vol. 25, no. 9. — Pp. 1041–1050.

20. PolyU Palmprint Database 2006 [Online]. — <http://www.comp.polyu.edu.hk/~biometrics>.
21. Palmprint verification using binary orientation co-occurrence vector / Zhenhua Guo, David Zhang, Lei Zhang, Wangmeng Zuo // *Pattern Recognition Letters*. — 2009. — Vol. 30, no. 13. — Pp. 1219–1227.
22. Kong Adams Wai-Kin, Zhang David. Feature-Level Fusion for Effective Palmprint Authentication // ICBA. — Vol. 3072 of *Lecture Notes in Computer Science*. — 2004. — Pp. 761–767.
23. Kong Adams Wai-Kin, Zhang David. Competitive Coding Scheme for Palmprint Verification // Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04). — Vol. 1 of *ICPR '04*. — Washington, DC, USA: IEEE Computer Society, 2004. — Pp. 520–523.
24. Zhang Leiqi, Li Huaqing, Niu Jianwei. Fragile bits in palmprint recognition // *Signal Processing Letters, IEEE*. — 2012. — Vol. 19, no. 10. — Pp. 663–666.
25. Huang De-Shuang, Jia Wei, Zhang David. Palmprint verification based on principal lines // *Pattern Recogn.* — 2008. — apr. — Vol. 41, no. 4. — Pp. 1316–1328.
26. Peg-free hand shape verification using high order zernike moments / Gholamreza Amayeh, George Bebis, Ali Erol, Mircea Nicolescu // Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on / IEEE. — 2006. — Pp. 40–40.
27. A contactless hand shape identification system / Cui Xin, Xiangqian Wu, Zhao Qiushi, Tang Youbao // Advanced Computer Control (ICACC), 2011 3rd International Conference on / IEEE. — 2011. — Pp. 561–565.
28. Shape-based hand recognition approach using the morphological pattern spectrum / Juan Manuel Ramírez-Cortés, Pilar Gomez-Gil, Gabriel Sanchez-Perez, Cesar Prieto-Castro // *J. Electronic Imaging*. — 2009. — 01. — Vol. 18. — P. 013012.
29. Maragos P. Pattern spectrum and multiscale shape representation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. — 1989. — July. — Vol. 11, no. 7. — Pp. 701–716.
30. Местецкий Л. М., М. Зак Е., А. Ломов Н. Распознавание формы гибких объектов изображений на основе спектров ширины // *Приволжский научный журнал*. — 2014. — № 4. — С. 59–67.

31. *Bakina Irina*. Palm shape comparison for person recognition // Computer Vision, Imaging and Computer Graphics. Theory and Applications. — Springer, 2013. — Pp. 157–171.
32. Online finger-knuckle-print verification for personal authentication / Lin Zhang, Lei Zhang, David Zhang, Hailong Zhu // *Pattern recognition*. — 2010. — Vol. 43, no. 7. — Pp. 2560–2571.
33. The Hong Kong Polytechnic University Mobile Phone Finger Knuckle Database [Online]. — <http://www.comp.polyu.edu.hk/~csajaykr/knuckle.htm>. — 2012.
34. Ensemble of local and global information for finger-knuckle-print recognition / Lin Zhang, Lei Zhang, David Zhang, Hailong Zhu // *Pattern Recognition*. — 2011. — Vol. 44, no. 9. — Pp. 1990–1998.
35. *Zhang Lin, Zhang Lei, Zhang David*. Finger-Knuckle-Print Verification Based on Band-Limited Phase-Only Correlation // Computer Analysis of Images and Patterns. — 2009. — 09. — Pp. 141–148.
36. *Kumar Ajay, Zhou Yingbo*. Personal identification using finger knuckle orientation features // *Electronics Letters*. — 2009. — Vol. 45, no. 20. — Pp. 1023–1025.
37. *Kumar Ajay, Ravikanth Ch.* Personal authentication using finger knuckle surface // *IEEE Transactions on Information Forensics and Security*. — 2009. — Vol. 4, no. 1. — Pp. 98–110.
38. *Zhang Lin, Zhang Lei, Zhang David*. MonogenicCode: A Novel Fast Feature Coding Algorithm with Applications to Finger-Knuckle-Print Recognition. — Vol. 1. — 2010. — Pp. 1–4.
39. *Kumar Ajay, Zhang David*. Personal recognition using hand shape and texture // *Image Processing, IEEE Transactions on*. — 2006. — Vol. 15, no. 8. — Pp. 2454–2461.
40. *Kumar Ajay, Zhang David*. Improving biometric authentication performance from the user quality // *Instrumentation and Measurement, IEEE Transactions on*. — 2010. — Vol. 59, no. 3. — Pp. 730–735.
41. *Zhu Le-qing, Zhang San-yuan*. Multimodal biometric identification system based on finger geometry, knuckle print and palm print // *Pattern Recognition Letters*. — 2010. — Vol. 31, no. 12. — Pp. 1641–1649.
42. *Meraoumia Abdallah, Chitroub Salim, Bouridane Ahmed*. Fusion of finger-knuckle-print and palmprint for an efficient multi-biometric system of person recognition // Communications (ICC), 2011 IEEE International Conference on / IEEE. — 2011. — Pp. 1–5.

43. *Iitsuka Satoshi, Ito Koichi, Aoki Takafumi*. A Practical Palmprint Recognition Algorithm Using Phase Information. — 2009. — 01. — Pp. 1 – 4.
44. Palmprint verification on mobile phones using accelerated competitive code / Marc Franzgrote, Christian Borg, Benjamin J Tobias Ries et al. // Hand-Based Biometrics (ICHB), 2011 International Conference on / IEEE. — 2011. — Pp. 1–6.
45. A Contactless Palmprint Recognition Algorithm for Mobile Phones / Shoichiro Aoyama, Koichi Ito, Takafumi Aoki, Haruki Ota // *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences (Japanese Edition)*. — 2013. — Vol. J96-A, no. 5. — Pp. 250–263.
46. Silhouette-based hand recognition on mobile devices / Alberto de Santos Sierra, Javier Guerra Casanova, Carmen Sánchez Avila, Vicente Jara Vera // Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on / IEEE. — 2009. — Pp. 160–166.
47. An approach to hand biometrics in mobile devices / Alberto de Santos Sierra, Carmen Sánchez Ávila, Aitor Mendaza Ormaza, Javier Guerra Casanova // *Signal, Image and Video Processing*. — 2011. — Vol. 5, no. 4. — Pp. 469–475.
48. Hand Biometrics in Mobile Devices / Alberto de Santos-Sierra, Carmen Sánchez Ávila, Javier Casanova, Aitor Mendaza-Ormaza // Advanced Biometric Technologies. — 2011. — 08. — Pp. 368–382.
49. *Cheng KamYuen, Kumar Ajay*. Contactless finger knuckle identification using smartphones // 2012 BIOSIG - Proceedings of the International Conference of the Biometrics Special Interest Group. — 2012. — Pp. 1–6.
50. *Otsu Nobuyuki*. A threshold selection method from gray-level histograms // *Automatica*. — 1975. — Vol. 11, no. 285–296. — Pp. 23–27.
51. *Viola Paul, Jones Michael J*. Robust Real-Time Face Detection // *Int. J. Comput. Vision*. — 2004. — may. — Vol. 57, no. 2. — Pp. 137–154.
52. *Sobral Andrews*. BGSLibrary: An OpenCV C++ Background Subtraction Library // IX Workshop de Visão Computacional (WVC'2013). — Rio de Janeiro, Brazil: 2013. — Jun.
53. *Yi Faliu, Moon Inkyu*. Image segmentation: A survey of graph-cut methods // Systems and Informatics (ICSAI), 2012 International Conference on / IEEE. — 2012. — Pp. 1936–1941.

54. *Camilus K Santle, Govindan VK.* A review on graph based segmentation // *International Journal of Image, Graphics and Signal Processing (IJIGSP)*. — 2012. — Vol. 4, no. 5. — P. 1.
55. *Boykov Yuri Y, Jolly Marie-Pierre.* Interactive graph cuts for optimal boundary & region segmentation of objects in ND images // *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* / IEEE. — Vol. 1. — 2001. — Pp. 105–112.
56. *Vodopivec Tadej, Lepetit Vincent, Peer Peter.* Fine hand segmentation using convolutional neural networks // *arXiv preprint arXiv:1608.07454*. — 2016.
57. *Местецкий Леонид Моисеевич.* Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры. — М.: ФИЗМАТЛИТ, 2009. — С. 288.
58. Grabcut in One Cut / Meng Tang, Lena Gorelick, Olga Veksler, Yuri Boykov // *International Conference on Computer Vision*. — 2013. — December.
59. SLIC superpixels compared to state-of-the-art superpixel methods / Radhakrishna Achanta, Appu Shaji, Kevin Smith et al. // *IEEE transactions on pattern analysis and machine intelligence*. — 2012. — Vol. 34, no. 11. — Pp. 2274–2282.
60. *Местецкий Л. М.* Медиальная ширина фигуры - дескриптор формы изображений // *Машинное обучение и анализ данных*. — 2014. — Т. 1, № 9. — С. 1291–1318.
61. *Mestetskiy Leonid, Bakina Irina, Kurakin Alexey.* Hand geometry analysis by continuous skeletons // *Proceedings of the 8th international conference on Image analysis and recognition - Volume Part II*. — ICIAR'11. — 2011. — Pp. 130–139.
62. Adaptive histogram equalization and its variations / Stephen M Pizer, E Philip Amburn, John D Austin et al. // *Computer vision, graphics, and image processing*. — 1987. — Vol. 39, no. 3. — Pp. 355–368.
63. *Canny John.* A computational approach to edge detection // *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. — 1986. — no. 6. — Pp. 679–698.
64. *Lee Tai Sing.* Image Representation Using 2D Gabor Wavelets // *IEEE Trans. Pattern Anal. Mach. Intell.* — 1996. — oct. — Vol. 18, no. 10. — Pp. 959–971.
65. *Lowe David G.* Object recognition from local scale-invariant features // *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* / Ieee. — Vol. 2. — 1999. — Pp. 1150–1157.

66. HGC 2011-Hand Geometric Points Detection Competition Database / F Magalhaes, HP Oliveira, H Matos, A Campilho. — 2010.
67. *Kumar Ajay, Xu Zhihuan*. Personal identification using minor knuckle patterns from palm dorsal surface // *IEEE Transactions on Information Forensics and Security*. — 2016. — Vol. 11, no. 10. — Pp. 2338–2348.
68. Robust palm and knuckle ROI extraction in unconstrained environment / Ahmed S ELSayed, Hala M Ebeid, Mohamed Roushdy, Zaki T Fayed // *Pattern Analysis and Applications*. — 2017. — Pp. 1–23.
69. Invariant hand biometrics feature extraction / Alberto Santos Sierra, Carmen Sánchez Ávila, Javier Guerra Casanova, Gonzalo Bailador Pozo // *Biometric Recognition*. — 2011.
70. Boost C++ library. — <http://www.boost.org/>.
71. OpenCV library. — <http://opencv.org/>.