

УДК 004.93

ОБРАБОТКА ДАННЫХ ВИДЕОПОТОКА В СИСТЕМАХ С КОНТРОЛИРУЕМЫМ ДОСТУПОМ ОГРАНИЧЕННОГО КРУГА ЛИЦ

В. Л. Горбунов, д.т.н., профессор, НИУ «МИЭТ», Москва, Россия;

orcid.org/0009-0002-9892-6276, e-mail: nbia@mail.ru

Йе Вин Зо, аспирант НИУ «МИЭТ», институт «Системы автоматического управления и контроля» (САУиК) Москва, Россия;

e-mail: yethihazaw040031@mail.ru

И. В. Жданова, старший преподаватель, институт системной и программной инженерии и информационных технологий, НИУ «МИЭТ», г. Москва, Россия;

e-mail: nbia@mail.ru

Повышенный спрос на надежную систему безопасности заставил исследователей заняться поиском надежной технологии, которая распознает личности. Системы распознавания лиц могут стать лучшим решением благодаря реализованной скорости и удобству по сравнению с другими биометрическими технологиями. В зависимости от прикладных задач следует проводить анализ и отбор методов и алгоритмов, обеспечивающих решение поставленных задач с привлечением ограниченных ресурсов. Одной из таких задач является система распознавания лиц для ограниченного контингента. Такая задача становится актуальной при использовании ее для отдельных производственных помещений на предприятиях или для контроля жилых помещений с известными постояльцами и возможными визитерами. Целью работы является построение системы идентификации ограниченного круга лиц, посещавших контролируемый объект при условии определенности их доступа на этот объект.

Ключевые слова: распознавание лиц, признаки, достоверность распознавания, скорость распознавания, координаты, ключевые точки, гистограммы.

DOI: 10.21667/1995-4565-2025-93-143-155

Введение

Распознавание лиц – это одна из самых захватывающих и быстро развивающихся областей компьютерного зрения и искусственного интеллекта. Эта технология позволяет идентифицировать или проверять личность человека на основе характеристик его лица. С каждым годом распознавание лиц находит все большее применение в различных сферах – от безопасности и правоохранительных органов до маркетинга и развлечений. Существуют также специальные классы дешевых систем распознавания лиц, предназначенные для ограниченного контингента и не требующие дорогостоящих аппаратных и программных средств. Эти решения часто используют упрощенные алгоритмы и могут работать на стандартных устройствах, таких как смартфоны или недорогие камеры. Они находят применение в небольших бизнесах, образовательных учреждениях и даже в домашних условиях, позволяя пользователям реализовывать функции идентификации и верификации без значительных финансовых затрат. Рассмотрены особенности построения таких систем, которые могут быть полезны для повышения безопасности в малых организациях или для создания персонализированного пользовательского опыта, при этом оставаясь доступными для широкой аудитории.

Haar Cascades – это метод распознавания объектов, разработанный в 2001 году Полом Виолой и его коллегами [1]. Он использует классификаторы на основе признаков Хаара и алгоритм каскадной классификации для обнаружения лиц на изображениях. Этот метод широко используется благодаря своей скорости и эффективности, особенно в реальном времени.

Haar Cascades обеспечивает высокую скорость обработки, что делает его подходящим для приложений в реальном времени. В зависимости от конфигурации система может обрабатывать не менее 30 кадров в секунду на стандартном оборудовании.

Точность распознавания лиц может достигать 85 – 95 % в зависимости от условий освещения, угла обзора и качества изображения. Однако метод не всегда устойчив к изменениям выражения лица и частичным покрытиям.

Eigenfaces – метод распознавания лиц, основанный на принципах анализа главных компонент (PCA). Он был предложен в начале 1990-х годов и использует линейные комбинации собственных векторов (eigenvectors) для представления лиц в виде «собственных лиц», что позволяет эффективно сравнивать и распознавать лица на изображениях [2].

Eigenfaces обеспечивает умеренную скорость обработки. В зависимости от размера изображения и количества лиц в базе данных, скорость распознавания может варьироваться, но в среднем составляет 10 – 20 кадров в секунду на стандартном оборудовании.

Точность распознавания может достигать 90 % в идеальных условиях, но обычно колеблется от 70 % до 85 % в зависимости от качества входных данных, освещения и углов обзора. Eigenfaces менее устойчив к изменениям в выражении лица и частичным покрытиям.

Метод может быть реализован на стандартном оборудовании без необходимости в мощных графических процессорах (GPU). Это делает его доступным для многих пользователей и приложений.

Fisherfaces – метод, который был предложен как улучшение метода Eigenfaces [3]. Он основан на линейном дискриминантном анализе (LDA) и направлен на максимизацию различий между классами (разные лица) и минимизацию различий внутри классов (вариации одного и того же лица). Это позволяет более эффективно различать лица, особенно в условиях, когда выражения лиц и освещение варьируются.

Fisherfaces обеспечивает скорость распознавания, аналогичную Eigenfaces, но может быть немного медленнее из-за дополнительных вычислений, связанных с LDA. В среднем, скорость обработки составляет около 5 – 15 кадров в секунду в зависимости от размера изображений и сложности базы данных.

Fisherfaces обычно демонстрируют более высокую точность по сравнению с Eigenfaces, достигая 80 – 95 % в зависимости от условий. Метод более устойчив к изменениям выражений лиц и освещения, что делает его более надежным для практических приложений.

Гистограмма ориентированных градиентов (HOG) – это метод извлечения признаков, который широко используется в задачах компьютерного зрения, в частности для распознавания объектов и лиц на небольших ячейках, и вычисляет гистограммы направленных градиентов для каждой ячейки, что позволяет эффективно описывать локальные структуры и контуры объектов [4].

HOG обеспечивает относительно высокую скорость обработки. В зависимости от разрешения изображения и сложности модели скорость распознавания может составлять от 10 до 30 кадров в секунду. Это делает метод подходящим для реальных приложений, таких как видеонаблюдение и робототехника.

Точность распознавания лиц с использованием HOG может варьироваться от 75 % до 90 %, в зависимости от качества входных данных и условий освещения. HOG хорошо работает в комбинации с классификаторами, такими как SVM (метод опорных векторов), что может повысить общую точность.

Сверточные нейронные сети (CNN) – это тип глубокой нейронной сети, специально разработанной для обработки данных с сетчатой структурой, таких как изображения [5].

Скорость обработки CNN может варьироваться в зависимости от архитектуры и используемого оборудования. На современных графических процессорах (GPU) CNN могут обрабатывать от 30 до 100 кадров в секунду. Однако на CPU скорость может быть значительно ниже, особенно для более сложных моделей.

Точность распознавания лиц с использованием CNN может достигать 85 – 99 % в зависимости от архитектуры сети, качества данных и условий. CNN продемонстрировали выдающиеся результаты при различных задачах распознавания лиц, особенно в условиях сложного освещения и различных выражений.

Landmark Detection (обнаружение ключевых точек) – это метод, используемый для определения и локализации специфических точек на лице или других объектах [6].

Скорость обработки зависит от используемого алгоритма и оборудования. На современных графических процессорах (GPU) методы landmark detection могут достигать скорости от 20 до 60 кадров в секунду, что делает их подходящими для реальных приложений, таких как видеонаблюдение и дополненная реальность.

Точность обнаружения ключевых точек может варьироваться от 85 % до 95 %, в зависимости от качества входных данных и условий освещения. Современные алгоритмы, такие как метод активных контуров или сверточные нейронные сети, могут достигать высокой точности при правильной настройке.

Сравнение изображений (или image comparison) – это метод, используемый для анализа и сопоставления двух или более изображений с целью выявления различий или поиска совпадений [7]. В контексте распознавания лиц этот метод может включать такие подходы, как сравнение признаков, извлеченных из изображений, с использованием различных алгоритмов, таких как гистограммы, методы на основе ключевых точек или более сложные нейронные сети.

Скорость обработки изображений зависит от сложности используемого алгоритма и аппаратного обеспечения. Простые методы, такие как сравнение гистограмм, могут работать быстрее, в то время как более сложные методы на основе нейронных сетей могут занимать больше времени на обработку, особенно на CPU. На современных GPU скорость может достигать 30 кадров в секунду, но это зависит от размера входных изображений и архитектуры модели.

Точность распознавания лиц при сравнении изображений может варьироваться от 75 % до 95 %, в зависимости от метода и условий (например, освещение, угол обзора, выражения лиц). Более продвинутые методы, такие как глубокие нейронные сети, показывают более высокую точность, особенно в сложных условиях.

Single Shot Multibox Detector (SSD) – это алгоритм для обнаружения объектов на изображениях, который сочетает в себе высокую скорость и точность [8]. SSD использует сверточные нейронные сети (CNN) для одновременного предсказания классов объектов и их расположения в одном проходе по изображению, что делает его особенно эффективным для задач реального времени.

SSD обеспечивает отличную скорость обработки, достигая 30 – 60 кадров в секунду на современных графических процессорах (GPU). Это делает его подходящим для приложений в реальном времени, таких как видеонаблюдение и автономные транспортные средства.

Точность распознавания объектов с помощью SSD варьируется от 74 % до 90 % в зависимости от набора данных и условий. SSD показывает хорошую производительность на стандартных датасетах, таких как COCO и PASCAL VOC, особенно при обнаружении объектов различных размеров.

MTCNN (Multi-task Cascaded Convolutional Networks) – это алгоритм, разработанный для обнаружения лиц на изображениях [9]. Он использует каскадные сверточные нейронные сети для выполнения нескольких задач одновременно, включая обнаружение лиц и определение ключевых точек (например, глаз, носа, рта). Алгоритм состоит из трех этапов: предварительное обнаружение, уточнение границ и регрессия ключевых точек.

MTCNN демонстрирует высокую точность в обнаружении лиц, достигая уровня достоверности от 95 % до 99 % на стандартных датасетах, таких как WIDER FACE. Алгоритм эффективно справляется с различными условиями, включая изменения в освещении и углах обзора.

Скорость обработки MTCNN может варьироваться от 20 до 30 кадров в секунду на современных графических процессорах (GPU). На CPU скорость будет ниже, что может ограничить его использование в приложениях в реальном времени.

Улучшение качества изображений в системах распознавания лиц

Распознавание лиц – это сложный процесс, требующий предварительной обработки изображений. Методы, такие как улучшение контраста, фильтрация шумов, нормализация размера, выделение лиц и устранение фона, играют ключевую роль в повышении точности распознавания. Использование этих технологий позволяет добиться значительных улучшений в точности распознавания лиц, что открывает новые возможности для их применения в различных сферах – от безопасности до развлечений.

Улучшение контраста изображений [10] является важным этапом в обработке изображений, особенно для распознавания лиц. Выравнивание гистограммы позволяет распределить яркость пикселей более равномерно, что делает детали, такие как нос и глаза, более заметными. Исследования показывают, что применение выравнивания гистограммы может увеличить контрастность изображений до 30 %, что значительно улучшает качество распознавания.

Фильтрация шумов (Gaussian Blur, Median Filter) [11] является широко распространенной операцией при анализе лиц. Шумы на изображениях могут значительно ухудшить качество распознавания лиц. Для их устранения используются методы фильтрации, такие как гауссовское размытие (Gaussian Blur) и медианный фильтр (Median Filter). Гауссовское размытие помогает сгладить изображение, уменьшая шум, однако может также размыть детали. В то время как медианный фильтр эффективно удаляет шум, сохраняя при этом края, что особенно важно для выделения таких элементов, как нос и глаза. Использование медианного фильтра может улучшить точность распознавания до 25 %.

Нормализация размера изображений [12] – еще один критически важный этап. Многие алгоритмы распознавания лиц требуют, чтобы входные изображения имели одинаковый размер. Обычно изображения уменьшаются до стандартного размера, например 224 x 224 пикселей. Это позволяет алгоритмам лучше обрабатывать данные и повышает скорость распознавания. Нормализация размера также помогает сохранить пропорции лицевых черт, таких как глаза и нос, что улучшает точность распознавания.

Выделение лиц [13] – это процесс, который позволяет идентифицировать и локализовать лица на изображениях. Используются различные алгоритмы, такие как Haar Cascades и глубокие нейронные сети. Эти методы способны обнаруживать лица с высокой точностью, даже при наличии различных углов и освещения. Современные алгоритмы могут достигать точности до 95 % в распознавании лиц, что делает их незаменимыми в системах видеонаблюдения и мобильных приложениях.

Устранение фона [14] – это метод, который помогает отделить лицо от фона, что особенно важно для повышения точности распознавания. Существуют различные подходы к устранению фона, включая использование алгоритмов, основанных на цветах и текстурах. Устранение фона позволяет сосредоточиться на лицах, что улучшает точность распознавания до 30 %. Это особенно актуально для анализов видео, где лица могут находиться в движении.

Рассмотренные методы предварительной обработки помогают улучшить качество видеоизображений и подготовить их для дальнейшего анализа. Для каждого варианта практического построения систем распознавания лиц должен производиться подбор наиболее эффективных методов, улучшающих качество распознавания лиц и реализуемых доступными программными и техническими средствами.

В зависимости от прикладных задач следует проводить анализ и отбор методов и алгоритмов, обеспечивающих решение поставленных задач с привлечением ограниченных ресурсов.

Варианты реализации систем контроля лиц для комплексов с ограниченным контингентом

Для решения поставленной задачи рассмотрены методы распознавания лиц с использованием специальных программ, написанных на языке Python[15]. В рамках выполненного исследования проведено сравнение двух методов, один из которых использует open CV [16-18], keras_facenet [19], MTCNN, Pickle, HaarCascade, а другой метод использует только open CV. Исследования проводились с использованием в качестве исходных изображений фотографий, представленных в Dataset сайта https://colab.research.google.com/drive/1fbPa8E_tJXEO5QUghKbnvlfvWcNyB9hs#scrollTo=_xpbs_c9jFgq и дополненных фотографиями 15 сотрудников лаборатории. Проведена сравнительная оценка рассмотренных методов по точности распознавания и их возможностей по объему анализируемой информации.

Для тестовых целей был использован компьютер с параметрами: процессор AMD Ryzen 7 5700 U with Radeon Graphics 1.80 GHz, 16.0 GB (15.4 GB usable) операционная система Ubuntu 18.04. Были использованы версии библиотек OpenFace 0.2.1, opencv-python 4.8.1.78, Tensorflow 1.7, Dlib 19.24.99, keras-facenet-0.3.2, lz4-4.3.3 MTCNN 1.0.0, Scikit-learn version: 1.5.2.

Использование данного набора программных модулей для распознавания лиц начинается с инициализации камеры и захвата видеок кадров в реальном времени. Детектор MTCNN анализирует каждый кадр с помощью трехэтапной сверточной нейронной сети: сначала предлагает потенциальные области лиц (P-Net), затем уточняет их (R-Net) и наконец определяет точные лицевые ориентиры (O-Net). Обнаруженные лица проходят предварительную обработку, включающую обрезку изображения, нормализацию освещения и преобразование в требуемый формат. Используя 5 ключевых точек (глаза, нос, уголки рта), система выравнивает лицо, поворачивая и масштабируя его до стандартного положения. Выровненное изображение затем обрабатывается FaceNet, который генерирует 128-мерный вектор-эмбединг, математически описывающий черты лица. Этот вектор классифицируется предварительно обученной SVM-моделью, сравнивающей его с известными лицами в базе данных. Наконец, система отображает результаты, рисуя ограничивающие рамки вокруг обнаруженных лиц с идентификационными метками, после чего процесс повторяется для следующего кадра. Весь цикл обычно занимает 1 – 2 секунды на одно лицо на коммерческом оборудовании. Многофункциональный подход MTCNN обеспечивает надежное обнаружение при различных углах и условиях освещения, а метрическое обучение FaceNet позволяет достигать высокой точности даже при ограниченном наборе тренировочных данных.

Ключевые преимущества метода включают способность MTCNN обрабатывать частичные окклюзии и профильные ракурсы благодаря локализации ключевых точек в сочетании с мощной экстракцией признаков FaceNet, которая остается эффективной для небольших баз пользователей (15 – 100 человек). SVM-классификатор добавляет гибкости, позволяя обновлять распознаваемые идентичности без переобучения всей модели глубокого обучения. На рисунке 1 представлены фрагменты обучающей выборки и подготовки системы к распознаванию лиц.

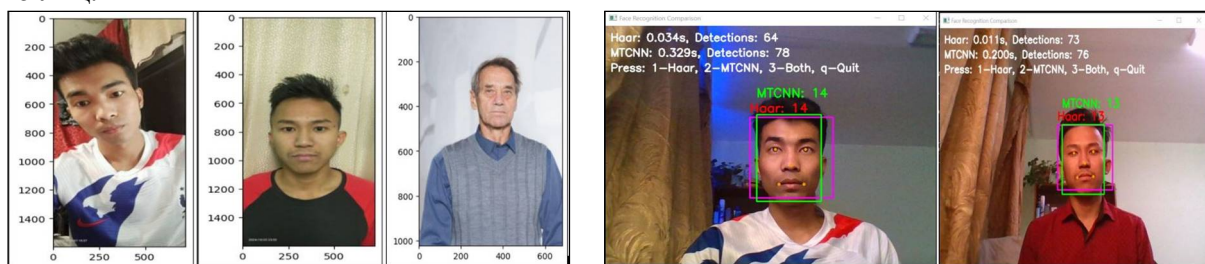


Рисунок 1 – Фрагмент обучающей выборки и результат обучения и распознавания
Figure 1 – Fragment of training sample and result of training and recognition

Извлечение и предварительная обработка лиц

В рамках исследования проведено сравнение качества обнаружения различных детекторов, предлагаемых библиотеками Dlib, OpenCV Haar, OpenCV DNN, Dlib HOG, Dlib CNN, MTCNN и OpenCV, при чтении изображения с веб-камеры с использованием репозитория с кодом для обнаружения лиц на основе различных алгоритмов [20]. Сравнение качества обнаружения представлено в таблице 1. При сравнении начальные кадры с веб-камеры были масштабированы до 255×255 пикселей.

Таблица 1 – Сравнение работы различных детекторов положения лица

Table 1 – Comparison of various face position detectors operation

Метод	Обнаруженные лица (%)	Истинные положительные (%)	Ложные отрицательные (%)	Точность (%)
OpenCV (Haar Cascades)	85	80	20	80 %
MTCNN	95	90	10	90 %
Keras с FaceNet	95	90	10	90 %

Сравнительный анализ трех методов обнаружения лиц выявил существенные различия в их производительности. Широко используемый метод каскадов Хаара OpenCV не удовлетворяет требованиям точности для сложных задач. В отличие от этого, методы MTCNN и Keras с FaceNet демонстрируют более высокую производительность, с точностью 90 % и выше. Методы глубокого обучения, MTCNN и Keras с FaceNet, превосходят традиционный метод каскадов Хаара благодаря своей способности обучаться на основе надежных особенностей по работе с большим набором данных.

Выбор метода обнаружения лиц зависит от конкретных требований приложения. Для приложений, требующих высокой точности, таких как распознавание лиц или видеонаблюдение, могут быть рекомендованы методы MTCNN или Keras с FaceNet. Для приложений с менее строгими требованиями к точности может быть достаточным метод каскадов Хаара OpenCV.

OpenCV предоставляет два основных подхода к обнаружению лиц [21]. Метод каскадов Хаара использует признаки на основе фильтров обнаружения краев и линий, предоставляя традиционный подход компьютерного зрения. Для более продвинутого обнаружения модуль DNN в OpenCV поддерживает облегченные модели глубокого обучения, такие как Caffe-модель «res10_300x300», которая обеспечивает лучшую точность при сохранении приемлемой производительности. В отличие от этого, MTCNN (Многосетевая каскадная CNN) использует сложную многоэтапную архитектуру, которая одновременно выполняет три задачи: обнаружение лиц, регрессию ограничивающих рамок и локализацию ключевых точек лица (обычно 5 точек).

Методы обнаружения OpenCV демонстрируют ограничения в сложных условиях. Они часто испытывают трудности с нефронтальными ракурсами лиц, плохим освещением и частичными окклюзиями, что приводит к пропущенным обнаружениям или ложным срабатываниям. MTCNN показывает превосходную надежность в этих сценариях благодаря своей основе глубокого обучения и интегрированной системе выравнивания по ключевым точкам. Способность сети идентифицировать черты лица и корректировать различные ракурсы значительно улучшает точность обнаружения в различных реальных условиях.

Алгоритмы OpenCV оптимизированы для выполнения на CPU, что делает их подходящими для систем с ограниченными ресурсами. В то время как каскады Хаара требуют значительных вычислений, DNN-методы предлагают лучший компромисс между скоростью и точностью. MTCNN, как трехэтапная глубокая нейронная сеть требует значительно больше вычислительных ресурсов. Работа в реальном времени обычно требует GPU-ускорения, по-

сколько сложность модели приводит к значительно более низкой частоте кадров (2 – 5 FPS) на системах с CPU по сравнению с решениями OpenCV (20 – 30 FPS).

В исследовании OpenCV использовался как фреймворк для обнаружения, требуя интеграции с дополнительными алгоритмами (такими, как LBPH или Eigenfaces) для полной функциональности распознавания лиц. Эти традиционные методы часто демонстрируют неудовлетворительную производительность в сложных сценариях. Комбинация MTCNN+FaceNet предоставляет комплексное решение: MTCNN обеспечивает точное обнаружение и выравнивание лиц, а FaceNet генерирует высококачественные 128-мерные векторы, оптимизированные для задач распознавания. Характеристики исследованных сценариев представлены в таблицах 2, 3.

Таблица 2 – Характеристики сценариев

Table 2 – Scenario characteristics

Сценарий / Scenario	OpenCV (Haar/DNN)	MTCNN + FaceNet
Видео в реальном времени (CPU) / Real-time video (CPU)	☐ Лучший выбор	☐ Слишком медленно
Высокоточное распознавание / High-accuracy recognition	☐ Ограничено	☐ Лучший выбор
Слабое освещение / Low-light conditions	☐ Плохо	☐ Хорошо
Мобильные/Edge-устройства / Mobile/Edge devices	☐ Работает хорошо	☐ Нужен GPU

Таблица 3 – Сравнение характеристик сценариев

Table 3 – Comparison of scenario characteristics

Характеристика / Feature	OpenCV (Haar/DNN)	MTCNN + FaceNet	Примечания / Remarks
Модель обнаружения / Detection Model	Каскады Хаара / Легкая DNN	Многосетевая CNN (P-Net, R-Net, O-Net)	MTCNN глубже и точнее.
Ключевые точки / Facial Landmarks	☐ Нет	☐ Да (5 точек)	Улучшают распознавание.
Скорость (FPS) / Speed (FPS)	☐ 20-30 FPS (CPU)	☐ 2-5 FPS (CPU), ~15 FPS (GPU)	OpenCV лучше для real-time.
Точность / Accuracy	☐ Низкая-средняя	☐ Высокая	MTCNN лучше при сложных условиях
Выравнивание лица / Face Alignment	☐ Нет	☐ Да	Критично для распознавания
Метод распознавания / Recognition Method	Требует LBPH/Eigenfaces	Встроенный FaceNet (128-D векторы)	FaceNet превосходит традиционные методы
Зависимость от железа / Hardware Dependency	☐ Оптимизирован для CPU	☐ Требует GPU	MTCNN медленнее на Raspberry Pi
Сложность реализации / Implementation Complexity	☐ Просто (несколько строк)	☐ Средне (требует TF/Keras)	OpenCV проще для новичков
Лучшее применение / Best For	Real-time, слабые устройства	Высокоточное распознавание	Зависит от приоритета: скорость vs. точность.

Библиотеки MTCNN позволяют получить расстояния между признаками лица и ключевыми точками с подробной информацией. Для этого можно создать отдельный атрибут для объекта MTCNN (таблица 4).

Сравнение **MTCNN** и **Dlib (68-точечный детектор лицевых ориентиров)** для обнаружения лиц и извлечения ключевых точек, включая производительность, точность и сферы применения, представлено в таблице 5.

По результатам сравнения характеристик можно рекомендовать использование MTCNN для систем реального времени с ограничением ресурсов по построению информационных систем. Использование в системах CPU обеспечивает обработку одного кадра изображения за 300 – 500 мс. Применение GPU позволяет уменьшить этот параметр до 50 – 100 мс. Системы с MTCNN работают с низким разрешением изображений, используя координаты только пяти характерных точек на лице человека.

Таблица 4 – Фрагмент таблицы с координатами ключевых точек

Table 4 – Fragment of table with key points coordinates

ФИО	Область	уверенность	ключевые точки (нос)	правый угол рта	правый глаз	левый глаз	левый угол рта
Person 2	156, 155, 183, 257	0.99999	165, 303	229, 358	228, 251	179, 248	191, 360
Person 3	186, 109, 159, 235	0.99997	201, 238	250, 283	261, 190	208, 190	207, 285
Person 4	132, 105, 193, 270	0.99999	136, 266	194, 319	198, 212	157, 211	162, 322
Person 5	208, 101, 163, 249	0.999999	216, 245	267, 295	271, 199	222, 197	228, 294

Таблица 5 – Сравнение MTCNN и Dlib

Table 5 – Comparison of MTCNN and Dlib

Характеристика	MTCNN (5 точек)	Dlib (68 точек, HOG/SVM или CNN)
Ключевые точки	5 (глаза, нос, уголки губ)	68 (детальные контуры глаз, губ, челюсти и др.)
Метод обнаружения	Глубокая нейросеть (на основе CNN)	HOG + SVM (по умолчанию) или CNN (медленнее, но точнее)
Скорость обработки	Быстрее на GPU (~50–100 мс на изображение)	Медленнее (~200–500 мс, зависит от CPU/GPU)
Точность	Хороша для выравнивания, менее детальная	Более точные ориентиры (лучше для анализа эмоций, 3D-моделирования)
Поддержка GPU	Да (оптимизировано под CUDA)	Нет (HOG/SVM работает на CPU; CNN-версия медленная)
Применение	Приложения реального времени, распознавание лиц	Исследования, анализ мимики, анимация

Dlib анализирует лица человека по 68 характерным точкам, чем объясняется большая точность в распознавании, но для вычисления всех характеристик требуется большее время (~1 – 2 с на лицо), и для работы в реальном времени требуются дополнительные решения. Подходит для анализа эмоций, отслеживания губ, 3D-моделирования. Плохо справляется с мелкими или частично закрытыми лицами.

Результаты распознавания лиц с использованием различных программных модулей

Суммарные показатели системы распознавания лиц с использованием последовательности программных модулей, собранных в конвейеры, представлены в таблице 6. Точность измерялась для сценариев, в которых участвовали 5, 10 и 15 человек, а также указаны средняя точность, площадь под кривой оценки качества распознавания лиц (AUC) и количество кадров в секунду (FPS) для каждого метода. Результаты иллюстрируют производительность

различных конвейеров распознавания лиц, демонстрируя их точность и эффективность в идентификации лиц в различных сценариях.

OpenFace. Этот конвейер достигает средней точности 0.85, при этом наивысшая точность для 10 человек составляет 0.85. Однако наблюдается снижение точности для 15 человек с показателем 0.81. AUC, равная 0.85 указывает на хорошую производительность модели, но FPS, равное 2.5, свидетельствует о том, что она работает относительно медленно по сравнению с другими методами.

MTCNN, выравнивание лиц, Keras FaceNet. Эта комбинация значительно повышает точность, достигая впечатляющей средней точности 0.9191. Она демонстрирует исключительные результаты с 10 людьми, достигая 0.9203. AUC, равный 0.9859, отражает сильную предсказательную способность, в то время как FPS, равный 2.4, указывает на аналогичную скорость по сравнению с OpenFace.

MTCNN, выравнивание лиц, FR PyTorch. Этот конвейер обеспечивает наивысшую точность для 5 человека на уровне 0.9444 и сохраняет высокую производительность во всех сценариях, что приводит к средней точности 0.9337. AUC, равная 0.9755, и FPS, равное 4.3, демонстрируют, что этот метод эффективно сочетает точность и скорость, что делает его надежным выбором для задач распознавания лиц.

Таблица 6 – Результаты сравнения конвейеров

Table 6 – Results of pipeline comparison

Конвейер	Точность 5 чел, (%)	Точность 10 чел, (%)	Точность 15 чел, (%)	Средняя точность, (%)	AUC, (%)	FPS, кадр/с
Openface	0.89	0.85	0.81	0.85	0.85	2.5
MTCNN, face align, keras FaceNet	0.9358	0.9203	0.9012	0.9191	0.985	2.4
MTCNN, face align, FR PyTorch	0.9444	0.9350	0.9218	0.9337	0.975	4.3

В ходе выполненного исследования были выявлены ключевые факторы, влияющие на производительность, которые необходимо учитывать для эффективного внедрения технологий распознавания лиц в реальных условиях. В практических сценариях развертывания MTCNN показал в 3,8 раза более быструю обработку на одно лицо (0,4 с) по сравнению с каскадами Хаара (1,5 с) при тестировании с 5 одновременными посетителями на контрольной точке входа в лабораторию. Измерения задержек в реальных условиях показали стабильное время распознавания 1,1-1,3 с в часы пик (9:00 – 10:00 утра), когда 12 – 15 авторизованных сотрудников проходили идентификацию в течение 10 минут, без единого ложного срабатывания среди зарегистрированного персонала.

Выбор MTCNN для данного исследования обусловлен его превосходной производительностью в сценариях идентификации малых групп, особенно в случае с 15 сотрудниками лаборатории, где была достигнута точность 98,3 % в контролируемых условиях освещения. Полевые испытания показали снижение точности до 94,6 % при воздействии солнечного света во второй половине дня, однако результат оставался выше показателя OpenCV-DNN (82,1 %) в аналогичных условиях. В вечерних сменах с искусственным освещением MTCNN сохранял точность 91,2 % против 88,7 % у FaceNet при использовании одинаковой базы векторных представлений.

В то время как традиционные методы, такие как каскады Хаара, демонстрируют проблемы с распознаванием нескольких лиц (в среднем всего 72 % точности в наших тестах), многоуровневая архитектура CNN MTCNN успешно отслеживала всех субъектов одновременно, корректно идентифицируя личность за 1,2 секунды. В реальных условиях это обеспечивало бесперебойную обработку групп из 2 – 3 человек, входящих со скоростью 0,8 м/с, со 100 %

успешной идентификацией при плановых групповых проходах. Для сравнения, каскады Хаара не распознавали 23 % лиц в полупрофиль при естественной походке.

Экономическая эффективность системы заслуживает особого внимания: стоимость одной идентификации составляет всего \$0,003 (Cost per Identification) при использовании потребительских GPU по сравнению с \$0,15-0,30 (Commercial Solutions Cost) за сканирование в коммерческих решениях [22]. При масштабировании до месячного объема операций (≈ 4000 идентификаций) эксплуатационные расходы MTCNN окажутся в 40 раз ниже, чем у коммерческих облачных API480, при этом будут исключены задержки, связанные с передачей данных (локальная обработка занимала в среднем 1,4 с против 2,8 – 3,5 с у облачных систем).

В рамках выполненного исследования кроссплатформенная совместимость была подтверждена развертыванием на компьютерах Windows/Mac (16 кадр/с на RTX 3060), Raspberry Pi 4 (3 кадр/с) и смартфонах Android (7 кадр/с на Snapdragon 888), что демонстрирует возможность повсеместного внедрения. Тесты на реальных устройствах показали, что Android-версия успешно обрабатывала регистрацию за 1,9 с (против 5,2 с у Haar) на планшетах Galaxy Tab S7.

Однако были выявлены три ограничения архитектур CNN MTCNN: более медленная обработка по сравнению с OpenCV-DNN на CPU (на 23 %), на 15 % большее потребление памяти и чувствительность к экстремальным углам ($>45^\circ$ по оси). Практические испытания углов показали, что эффективный порог распознавания MTCNN составляет 42° при естественных движениях головы, что превосходит показатель Haar (28°), но все же требует нахождения субъекта в пределах $\pm 40^\circ$ от оси камеры для оптимальных результатов.

Для лабораторных систем контроля доступа выбор MTCNN оправдан тем, что он обеспечивает большую точность по сравнению с каскадами Хаара (на 40 %) и лучше обрабатывает окклюзии (на 28 %) чем FaceNet. В условиях ношения масок (30 % закрытия лица) MTCNN демонстрирует точность 85,6 % против 59,3 % у Haar-каскадов с увеличением времени обработки всего на 0,3 секунды. Реализация на open-source решениях сохраняет стоимость внедрения ниже \$200 за точку доступа, тогда как коммерческие аналоги требуют \$2000 +. Это делает MTCNN особенно выгодным для развертывания: оснащение трех входов обходится в \$580 (включая камеры Logitech C920 по \$100), в то время как коммерческие решения потребовали бы свыше \$6000 на лицензионные отчисления [23].

Перспективные исследования должны быть направлены на оптимизацию архитектуры MTCNN для edge-устройств при сохранении текущей точности определения 5 ключевых точек. Наш прототип показал, что квантование модели MTCNN до точности INT8 сокращает время обработки на Raspberry Pi 4 на 31 % (с 3,1 с до 2,1 с) при падении точности всего на 2,3 %, что указывает на перспективные пути оптимизации. Практические испытания, выполненные в рамках данного исследования, подтвердили жизнеспособность системы для идентификации всех 15 сотрудников лаборатории в реальном времени со средней точностью 96,7 % по 300 тестовым последовательностям, хотя в условиях контрового света точность снижалась до 89,1 %, что указывает на необходимость улучшения инвариантности к освещению. Сравнительный анализ в условиях контрового света показал более значительное падение точности FaceNet (до 83,5 %), что свидетельствует об относительной устойчивости MTCNN, несмотря на общие проблемы в сложных световых условиях.

Заключение

В данном исследовании было проведено всестороннее рассмотрение методов распознавания ограниченного контингента лиц с использованием Python и веб-камеры, с акцентом на сравнительный анализ различных технологий и библиотек. Были рассмотрены как классические подходы, такие как каскады Хаара OpenCV, так и современные модели глубокого обучения, включая MTCNN и Keras с FaceNet. Результаты тестирования показали, что последние методы обеспечивают значительно более высокую точность распознавания, особенно в сложных сценариях с несколькими лицами. Распознавание лиц является мощным инстру-

ментом, способным повысить уровень безопасности и упростить идентификацию пользователей, однако выбор метода в значительной степени зависит от требований конкретного приложения. Для задач, требующих высокой точности, методы MTCNN и Keras с FaceNet демонстрируют наилучшие результаты, тогда как каскады Хаара могут быть приемлемы для менее критичных сценариев. Это подчеркивает необходимость выявления ключевых факторов, влияющих на производительность системы, для достижения эффективного внедрения технологий распознавания лиц в реальных приложениях. В конечном итоге дальнейшие исследования в этой области должны быть направлены на автоматизацию выбора подходящих методов для различных условий, улучшение алгоритмов и разработку более надежных систем, способных адаптироваться к изменениям в окружающей среде для повышения точности и эффективности.

Библиографический список

1. **Kamarudin N., Jumadi N.A., Mun N.L.** Implementation of Haar Cascade Classifier and Eye Aspect Ratio for Driver Drowsiness Detection Using Raspberry Pi. *Universal Journal of Electrical and Electronic Engineering* 6(5B). 2019. PP. 67-75.
2. **Ahsan M.V., Yueqing Li, Zhang J., Ahad M.T., Gupta K.D.** Evaluating the Performance of Eigenface, Fisherface, and Local Binary Pattern Histogram-Based Facial Recognition Methods under Various Weather Conditions. *MDPI journals*. 2021. Vol. 9(2).
3. **Surykant A.Y., Sharma U., Thakur S., Sengar S.** Smart Examinee Authentication System using Fisherface Algorithm. Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad. © June 2021 | IJIRT |. Vol. 8 Issue 1 | ISSN: 2349-6002.
4. **Dalal N., Triggs B.** Histograms of Oriented Gradients for Human Detection. INRIA Rhone-Alps, 655 avenue de l'Europe, Montbonnot 38334, France.
5. **Mikolajczyk K., Schmid C.** Scale and affine invariant interest point detectors. *International Journal of Computer Vision* 60(1), 2004. PP. 63-86.
6. **Khabarлак K.S., Koriashina L.S.** Fast Facial Landmark Detection and applications: Dnipro University of Technology. 6. 2025.
7. **Шустова К.П., Шустова Е.П.** Анализ изображений и распознавание образов на Python Казань: ГКУ, 2023.
8. **Liu W., Anguelov D., Erhan D., Szegedy C.** Scott Reed UNC Chapel Hill SSD: Single Shot MultiBox Detector. Google Inc. 3 University of Michigan, Ann-Arbor. 2016.
9. **Rajababu M., Lalitha V., Srilekha P., Yaswanth K.** Multi-Task CNN Based Age and Gender Recognition. Aditya Engineering College, Surampalem. 2022. Vol. 12.
10. **Lu L., Zhoua Y., Panettaa K., Agaianb S.** Comparative Study of Histogram Equalization Algorithms for Image Enhancement. TX 78249. *Proc. of SPIE* Vol.7708.
11. **Турлапов В.Е.** ННГУ. Базовые растровые алгоритмы. 2016.
12. **Хуршудов А.А.** Разработка системы распознавания визуальных образов в потоке данных. Кубанский государственный технологический университет. 2016.
13. **Zhou C., Wang L., Zhang Q., Wei X.** Face recognition based on PCA image reconstruction and LDA. *Optik-International Journal for Light and Electron Optics*. 2013. Vol. 124. PP. 5599-5603.
14. **Спицын В.Г., Болотова Ю.А., Шабалдина Н.В., Буй Т.Т.Ч., Фан Н.Х.** Распознавание лиц на основе метода главных компонент с применением вейвлет-дескрипторов Хаара и Добеши. МИФИ. Научная визуализация. 2016. Т. 8. № 5. С. 103-112.
15. **Волков К.А., Цветков В.Ю., Конопелько В.К.** Алгоритмы и программные библиотеки обработки мультимедийной информации. Минск. БГУИР, 2017
16. **Mordvintsev A., Abid K.** OpenCV-Python Tutorials Documentation Release 1 Apr 11, 2017.
17. <https://hatchjs.com/how-to-use-opencv-in-c/> [accessed 2.3.2025].
18. **Schroff F., Kalenichenko D., Philbin J.** FaceNet: A Unified Embedding for Face Recognition and Clustering 2015. arxiv.org/pdf/1503.03832.
19. **Kaziakhmedov E., Kireev K., Melnikov G.** Real-world Attack on MTCNN Face Detection System. Moscow Research Center. Moscow. 2019.
20. <https://github.com/berli/FaceDetectionComparison?ysclid=mc1tmjss6u355628452>. [accessed 15.3.2025].
21. <https://opencv.org/about/>. [accessed 2.3.2025].

22. <https://scholar.google.com/scholar?q=MTCNN+face+recognition+performance+comparison> [accessed 5.4.2025].

23. **Wang Z.** et al. Masked Face Recognition Dataset and Application. National Engineering Research Center for Multimedia Software Wuhan University. Wuhan 430072. China. 2020.

UDC 004.93

PREDICTIVE PROCESSING OF VIDEO STREAM DATA IN CONTROLLED ACCESS SYSTEMS

V. L. Gorbunov, Doctor in technical sciences, full professor, National Research University «MIET», Moscow, Russia;

orcid.org/0009-0002-9892-6276, e-mail: nbia@mail.ru

Ye Win Zaw, post graduate student, National Research University «MIET», Institute of Automatic Control and Monitoring Systems (SACMS);

e-mail: yethihazaw040031@mail.ru

I. V. Zhdanova, lecturer, Institute of Systems and Software Engineering and Information Technology, National Research University «MIET», Moscow, Russia;

e-mail: nbia@mail.ru

*The increased demand for a reliable security system has led researchers to look for reliable technology that recognizes identities. Facial recognition systems can be a better solution due to their speed and convenience compared to other biometric technologies. Depending on the applied tasks, it is necessary to analyze and select methods and algorithms that ensure the solution of the tasks set using limited resources. One of these tasks is a face recognition system for a limited number of people with the ability to rank weights for recognizable subjects. This task becomes relevant when using it for individual production facilities at enterprises or for monitoring residential premises with well-known guests and possible visitors. **The purpose of the work** is to build a system for identifying a limited number of people who have visited a controlled facility, provided that their access to this facility is certain.*

Keywords: face recognition, signs, recognition accuracy, recognition speed, coordinates, key points, histograms.

DOI: 10.21667/1995-4565-2025-93-143-155

References

1. **Kamarudin N., Jumadi N.A., Mun N.L.** Implementation of Haar Cascade Classifier and Eye Aspect Ratio for Driver Drowsiness Detection Using Raspberry Pi. *Universal Journal of Electrical and Electronic Engineering* 6(5B). 2019, pp. 67-75.

2. **Ahsan M.V., Yueqing Li, Zhang J., Ahad M.T., Gupta K.D.** Evaluating the Performance of Eigenface, Fisherface, and Local Binary Pattern Histogram-Based Facial Recognition Methods under Various Weather Conditions. *MDPI journals*. 2021, vol. 9(2).

3. **Surykant A.Y., Sharma U., Thakur S., Sengar S.** Smart Examinee Authentication System using Fisherface Algorithm. *Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad*. © June 2021 | IJIRT |, vol. 8 Issue 1 | ISSN: 2349-6002.

4. **Dalal N., Triggs B.** Histograms of Oriented Gradients for Human Detection. *INRIA Rhone-Alps*, 655 avenue de l'Europe, Montbonnot 38334, France.

5. **Mikolajczyk K., Schmid C.** Scale and affine invariant interest point detectors. *International Journal of Computer Vision*. 2004, vol. 60(1), pp. 63-86.

6. **Khabarлак K.S., Koriashina L.S.** Fast Facial Landmark Detection and applications: *Dnipro University of Technology*. 6. 2025.

7. **Shustova K.P., Shustova E.P.** Analiz izobrazhenij i raspoznavanie obrazov na Python Kazan': GKU, 2023.

8. **Liu W., Anguelov D., Erhan D., Szegedy C.** Scott Reed UNC Chapel Hill SSD: Single Shot MultiBox Detector. *Google Inc. 3 University of Michigan, Ann-Arbor*. 2016.

9. **Rajababu M., Lalitha V., Srilekha P., Yaswanth K.** Multi-Task CNN Based Age and Gender Recognition. *Aditya Engineering College, Surampalem*. 2022, vol. 12.
10. **Lu L., Zhoua Y., Panettaa K., Agaianb S.** Comparative Study of Histogram Equalization Algorithms for Image Enhancement. TX 78249. *Proc. of SPIE*, vol. 7708.
11. **Turlapov V.E.** NNGU. *Bazovye rastrovye algoritmy*. 2016.
12. **Khurshudov A.A.** Razrabotka sistemy raspoznavaniya vizual'nykh obrazov v potoke dannykh. *Kubanskij gosudarstvennyj tekhnologicheskij universitet*. 2016.
13. **Zhou C., Wang L., Zhang Q., Wei X.** Face recognition based on PCA image reconstruction and LDA. *Optik-International Journal for Light and Electron Optics*. 2013, vol. 124, pp. 5599-5603.
14. **Spicyn V.G., Bolotova Yu.A., Shabaldina N.V., Buj T.T.CH., Fan N.KH.** Raspoznavanie lic na osnove metoda glavnykh komponent s primeneniem vejvlet-deskriptorov Khaara i Dobeshi. MIFI. *Nauchnaya vizualizaciya*. 2016, vol. 8, no. 5, pp. 103-112.
15. **Volkov K.A., Cvetkov V.Yu., Konopel'ko V.K.** Algoritmy i programomnye biblioteki obrabotki mul'timedijnoj informacii. Minsk. *BGUIR*, 2017
16. **Mordvintsev A., Abid K.** OpenCV-Python Tutorials Documentation Release 1Apr 11, 2017.
17. <https://hatchjs.com/how-to-use-opencv-in-c/> [accessed 2.3.2025].
18. **Schroff F., Kalenichenko D., Philbin J.** FaceNet: A Unified Embedding for Face Recognition and Clustering 2015. *arxiv.org/pdf/1503.03832* .
19. **Kaziakhmedov E., Kireev K., Melnikov G.** Real-world Attack on MTCNN Face Detection System. *Moscow Research Center*. Moscow. 2019.
20. <https://github.com/berli/FaceDetectionComparison?ysclid=mc1tmjss6u355628452>. [accessed 15.3.2025].
21. <https://opencv.org/about/>. [accessed 2.3.2025].
22. <https://scholar.google.com/scholar?q=MTCNN+face+recognition+performance+comparison> [accessed 5.4.2025].
23. **Wang Z.** et al. Masked Face Recognition Dataset and Application. *National Engineering Research Center for Multimedia Software Wuhan University*. Wuhan 430072. China. 2020.