

На правах рукописи



Хтет Аунг

**АЛГОРИТМЫ РАСПОЗНАВАНИЯ И ОТСЛЕЖИВАНИЯ ЛИЦ
ПРИ ИХ ЧАСТИЧНОМ ЗАКРЫТИИ**

Специальность 2.3.1. Системный анализ, управление и обработка
информации, статистика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2024 г.

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана).

Научный руководитель:

Бобков Александр Валентинович, кандидат технических наук, доцент кафедры систем автоматического управления МГТУ им. Н.Э. Баумана.

Официальные оппоненты:

Дивеев Асхат Ибрагимович – доктор технических наук, профессор, главный научный сотрудник, руководитель отдела ФИЦ «Информатика и управление» РАН.

Шелагурова Марина Сергеевна – кандидат технических наук, Начальник сектора по индикации и картографии, АО «РПКБ»

Ведущая организация:

Общество с ограниченной ответственностью
«РТСофт-Смарт Грид»

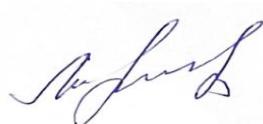
Защита состоится «22» октября 2024 г. в 14 час. 30 мин. на заседании диссертационного совета 24.2.331.14 при МГТУ им. Н.Э. Баумана по адресу: 105005, Москва, Госпитальный пер., 10, ауд. 613м.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана и на сайте <https://www.bmstu.ru/>.

Отзывы и замечания по автореферату в двух экземплярах, заверенные печатью, просьба высылать по адресу: 105005, Москва, ул. 2-я Бауманская, д. 5, стр. 1 на имя ученого секретаря диссертационного совета.

Автореферат разослан « » 2024 г.

Ученый секретарь
диссертационного совета
к.т.н., доцент



И. В. Муратов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. В работе рассматривается задача идентификации и отслеживания лиц в режиме реального времени при наличии их частичного закрытия с использованием комбинированного подхода на основе глубокого обучения и корреляционного фильтра.

За последние 20 лет было разработано множество биометрических методов для решения задач идентификации и верификации человека, таких как распознавание отпечатков пальца, радужной оболочки глаза и распознавание лиц. Среди них проблема распознавания лиц является одной из самых интересных и сложных. Способность идентифицировать личность человека, анализируя различные черты лица по видео изображение в режиме реального времени, обеспечивает огромные преимущества по сравнению с другими решениями биометрической безопасности.

Алгоритмы распознавания лиц анализируют различные аспекты лица человека, включая расстояние между глаз, форму линии подбородка и положение носа, чтобы создать уникальную цифровую подпись, которую можно использовать для идентификации человека в режиме реального времени. Автоматическое распознавание лиц широко применяется в различных отраслях, включая обеспечение безопасности, правоохранительные органы, банковское дело, здравоохранение, социальные сети и т.д.

Существует несколько проблем и ключевых факторов, которые могут существенно повлиять как на эффективность распознавания лиц, так и на результаты сопоставления. Эти проблемы, которые необходимо решить для повышения точности и надежности, можно сгруппировать в пять категорий: освещение, вариации выражений, вариации позы, различные окклюзии и низкое качество или разрешение изображения лица.

Одна из основных проблем в распознавании лиц заключается в том, что лицо человека может быть частично закрыто такими элементами, как медицинская маска, солнцезащитные очки, шарф, другими аксессуарами, а также естественными элементами (тень, элементы прически, усы, борода). Такое закрытие (окклюзия) оказывает значительное влияние на точность распознавания лиц и затрудняет их правильную идентификацию.

Во время всемирного эпидемия COVID-19, когда ношение маски для лица является обязательным для всех, распознавание лиц с маской становится очень важной проблемой. Существует много ситуаций, когда скрытое распознавание лиц очень важно, например, в аэропорту для проверки личности, в колледже и школе для проверки посещаемости студентов, посещаемости сотрудников в компании, когда можно распознавать лица с помощью биометрической машины, не снимая маски для лица. Поэтому проблемы повышения точности системы распознавания лиц в условиях окклюзии рассматриваются многими авторами во время эпидемии.

Задача автоматического высокоточного распознавания лиц при воздействии частичного закрытия (например, различные маски, солнцезащитные очки и руки), которое возникает в большом количестве важных практических

приложений, прежде всего – связанных с обеспечением безопасности и контроля доступа. Возможность распознавания лиц с высокой скоростью и надёжностью выводит эти приложения на новый уровень, существенно повышая оперативность работы.

В данной работе представлена система идентификации лиц с помощью трех основных этапов, включающую в себя модель обнаружения лица с частичной закрытой, модель распознавания лиц и модель отслеживания лиц.

Целью работы является разработка и исследование алгоритмов, способных обнаруживать и отслеживать лица в реальном времени в разных положениях, при разном освещении, с разным цветом кожи и другими различиями, а также распознавать их при частичном закрытии.

Задачи исследования:

1. Выполнить аналитический обзор существующих методов и алгоритмов, используемых в системах распознавания лиц;
2. Разработать алгоритм обнаружения лиц в условиях реальной окклюзии с использованием детектора YOLO;
3. Разработать систему распознавания частично закрытых лиц на основе сети ResNet50;
4. Разработать алгоритм отслеживания лиц на основе фильтра MOSSE (Минимальная Выходная Сумма Квадратов Ошибок);
5. Реализовать программное обеспечение по предлагаемым алгоритмам в среде MATLAB;
6. Провести эксперимента и выполнить анализ результатов.

Объект исследования

Объектом исследования диссертационной работы является система компьютерного зрения, осуществляющая идентификацию лиц на видео в режиме реального времени при наличии их частичного закрытия.

Предмет исследования

Предметом исследования является комбинированный подход к решению задачи идентификации и отслеживания лиц.

Методы исследования

В работе использовались методы теории глубокого обучения, автоматического управления, вероятностных моделей, математической статистики, цифровой обработки изображений и моделирование с использованием программного пакета MATLAB.

Научная новизна работы заключается в следующем:

1. Разработан комбинированный поход на основе сети YOLO и трекера MOSSE, способный обнаруживать и отслеживать лица на видео изображений в реальном времени;
2. Разработан комбинированный алгоритм с использованием существующих моделей сверточных нейронных сетей, позволяющий распознавать лица при их частичном закрытии;
3. Разработан новый алгоритм, основанный на применении предложенного комбинированного подхода, дающий возможность распознавания лиц на

видео изображении в реальном времени с высокой точностью при их частичном закрытии.

Научная ценность диссертационной работы заключается в предложенном оригинальном комбинированном подходе, основанном на использовании методов обнаружения лиц, отслеживания лиц, извлечение признаков и классификации признаков, повышающий качество системы идентификации лиц не только на изображениях, но и в видео реального времени с высокой точностью при условии частичного закрытия лица.

Практическая ценность диссертационной работы заключается в том, что разработанные алгоритмы в виде программного обеспечения могут быть использованы в системах контроля доступа для идентификации лиц, при наличии масок или других окклюзионных аксессуаров, таких как шарф, солнцезащитные очки и т.д., например, в условиях пандемии. Представленные алгоритмы распознавания частично закрытых лиц способных работать не только со статическими изображениями, но и с видеопоследовательностями в режиме реального времени.

Достоверность полученных результатов исследования подтверждается экспериментальными результатами моделирования предложенных алгоритмов и методов системы идентификации лиц.

Основные положения, выносимые на защиту:

1. Комбинированный алгоритм обнаружения и отслеживания лиц с использованием сети YOLO и фильтра MOSSE позволяет детектировать и отслеживать лица с высокой скоростью и при наличии частичного закрытия.
2. Предварительно обученная сетевая модель ResNet 50, использованная для извлечения признаков лица и идентификации с помощью потери кросс-энтропии. Инструменты моделирования с использованием языка программирования Matlab, включая computer vision toolbox, deep learning toolbox, image processing toolbox.
3. Оригинальные структура и алгоритмы функционирования системы распознавания лиц позволяют реализовать предложенный метод обеспечивают высокую точность распознавания не только при реальной окклюзии лица, но и в режиме реального времени.

Внедрение результатов

Разработанные варианты реализации системы распознавания лиц внедрены на кафедре информационных технологий машиностроения из технологического университета (г. Тханлин) под руководством министерства образования Союза Мьянмы, а также на кафедре системы автоматического управления МГТУ им. Н.Э. Баумана.

Апробация

1. 4-я Международная научно-практическая конференция, « проблемы лингвистики и лингводидактики в неязыковом вузе», (Москва, 2020).
2. Международная научно-техническая конференция «Пром-Инжиниринг», (Сочи, 2021).

3. 20-я Всероссийская конференция с международным участием «Математические методы распознавания образов», (Москва, 2021).
4. Международная научно-техническая конференция «Пром-Инжиниринг», (Сочи, 2022).
5. XLVII Академические чтения по космонавтике «Королёвские чтения 2023», (Москва, 2023).

Публикации

По результатам исследований опубликовано 10 научных работ, в числе которых 4 научные работы в журналах, входящих в список, утвержденный ВАК РФ по специальности 2.3.1, 4 работы, включенных в базу данных SCOPUS и 2 тезисы доклада на научных конференциях.

Структура работы

Диссертация состоит из списка сокращений, введения, трех глав, заключения, списка литературы, который содержит 121 наименование. Объем диссертации: 156 страниц, включая 50 рисунков, 10 таблиц и 5 листинг.

СОДЕРЖАНИЕ РАБОТЫ

Во введении содержится обоснование актуальности темы диссертации, определяются цели и задачи, указывается научная новизна и практическая значимость работы, формулируются положения, выносимые на защиту, содержатся сведения об апробации, внедрении результатов, публикациях, структуре и объеме диссертации.

В первой главе рассматриваются общие сведения о методах биометрического распознавания. Представлены различные проблемы и факторы, возникающие в задаче распознавания лиц и влияющие на качество распознавания. Анализируются методы и алгоритмы обнаружения и распознавания лиц в условиях окклюзии. При этом каждый из рассмотренных методов имеет свои особенности, влияющие на качество распознавания, и обсуждаются проблемы выбранного метода.

При разработке системы распознавания частично закрытых лиц необходимо учитывать факторы, связанные с отслеживанием лиц. Проанализированы различные методы и алгоритмы, обеспечивающие работу системы слежения за лицом. Рассмотрены существующие наборы данных для обнаружения и распознавания лиц в нормальных условиях и в условиях окклюзии.

Вторая глава посвящена алгоритмам и методам решения задач распознавания лиц. Приведена архитектура предлагаемой системы распознавания и отслеживания лиц, обеспечивающая возможность использования при частичном закрытии лица, а также в реальном времени (Рис. 1). Метод состоит из четырех этапов:

- Поиск закрытых лиц;
- Отслеживание лиц;
- Извлечение признаков;
- Классификация.

Приведено описание методов глубокого обучения, использованных при построении предлагаемой системы распознавания лиц.



Рис. 1. Основная архитектура предлагаемой работы

Разработан алгоритм на основе поисковой сети YOLO для обнаружения лиц при их частичной окклюзии. Алгоритм содержит три этапа:

- ✓ предварительная обработка данных обучающей выборки,
- ✓ извлечение признаков с помощью сети DarkNet53,
- ✓ обнаружение частично закрытых лиц на основе алгоритма YOLO.

На этапе предварительной обработки был использован открытый набор данных Fddb, а также добавлены различные реальные лица с окклюзией из сети Интернет (например, окклюзия ладони, маска, шарф, солнцезащитные очки). При создании обучающей выборки в среде MATLAB был использован инструментарий imgLabeler (Рис. 2). На этапе извлечения признаков использовались модель DarkNet53, состоящая из 53 сверточных слоев.

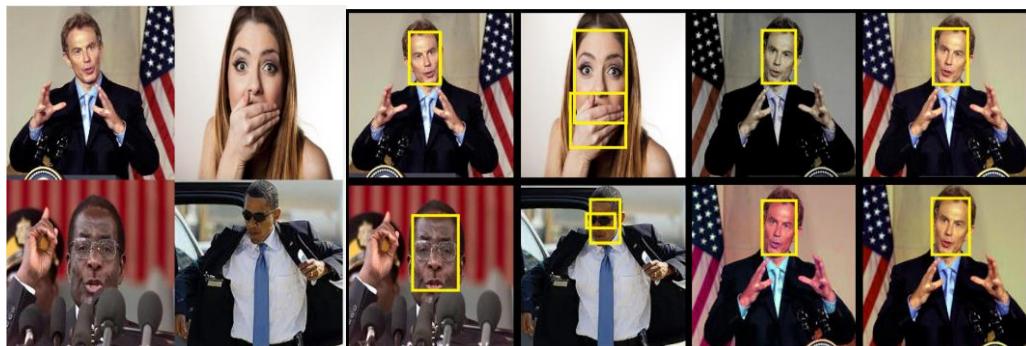


Рис. 2. Образцы исходных изображений, изображения обучающей выборки и аугментированные изображения

Предлагаемый метод обнаружения объекта YOLO выдает результат с размером ответа S^*S , где S – количество ячеек сетки. Каждая рамка имеет пять параметров: $(x, y, w, h \text{ и } Pr(obj))$, где x, y – координаты центра ограничивающей рамки, w, h – ширина и высота, а $Pr(obj)$ – вероятность нахождения объекта внутри рамки (Рис. 3). Показатель достоверности отражает вероятность включения целевого объекта в модель и точность блока обнаружения прогнозирования. Если объекта нет в ячейке, то $Pr(obj)$ должен быть равен нулю, а доверительное значение должно быть равно нулю: $C(obj) = 0$. Показатель достоверности $C(obj)$ определяется как:

$$C(\text{obj}) = \text{Pr}(\text{obj}) * \text{IoU}(\text{Pred}, \text{Gtruth})$$

где IoU – величина перекрытия найденной ограничивающей рамки и рамки из обучающей выборки, то есть отношение их пересечения и объединения:

$$\text{IoU}(\text{Pred}, \text{Gtruth}) =$$

перекрывающаяся область предсказанной рамки и рамки обучающей выборки

вся область предсказанной рамки и рамки обучающей выборки



Рис. 3. Метод обнаружения объектов YOLO

Уровень обнаружения YOLO используется для обнаружения на картах объектов трех различных размеров, имеющих шаги субдискретизации 32, 16, 8 соответственно. Это означает, что для входных изображений размером 316 x 316 будет выполняться обнаружение в масштабах 9 x 9, 19 x 19 и 39 x 39 (Рис. 4).

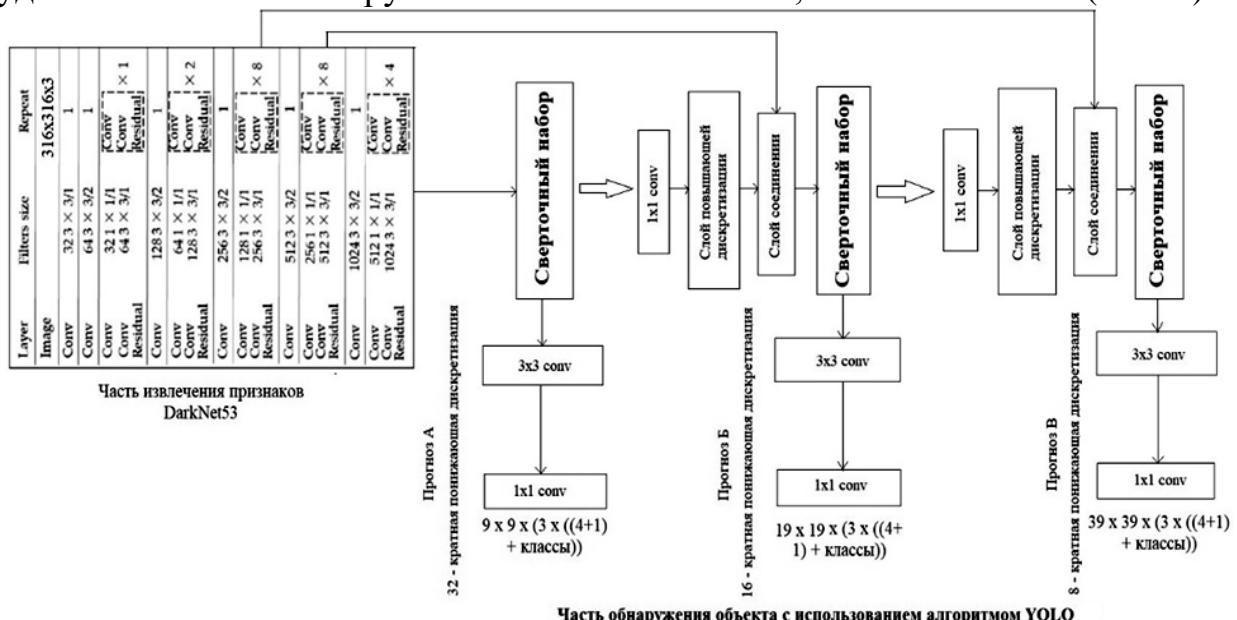


Рис. 4. Архитектура процесса обнаружения окклузированных лиц предлагаемой модели

Для обучения YOLO использует сумму квадратов ошибок между прогнозами и обучающими выборками для вычисления потери локализации. Для потери достоверности и потери классификации использовались потеря кросс энтропии. Функция потерь состоит из:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (2 - w_i \times h_i) \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \quad (1)$$

$$+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (2 - w_i \times h_i) \left[(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2 \right] \quad (2)$$

$$-\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \left[\hat{C}_i \log(C_i) + (1-\hat{C}_i) \log(1-C_i) \right] \quad (3)$$

$$-\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} \left[\hat{C}_i \log(C_i) + (1-\hat{C}_i) \log(1-C_i) \right] \quad (4)$$

$$-\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \sum_{c \in classes} \left[\hat{P}_i(c) \log(P_i(c)) + (1-\hat{P}_i(c)) \log(1-P_i(c)) \right] \quad (5)$$

где $I_{ij}^{obj} = 1$, если j -й граничный прямоугольник в i -й ячейке сетки отвечает за обнаружение объекта, то в противном случае 0;

λ_{coord} , λ_{noobj} – увеличение веса для потери в координатах граничного поля;

(x, y, w, h) – потери локализации между обучающей выборкой и предсказанной рамкой.

C и $P(c)$ – потери достоверности обнаружения и вероятности принадлежности к классу.

Предложен алгоритм быстрого и точного отслеживания лиц на последовательности видеоизображений в режиме реального времени, основанный на использовании фильтра MOSSE. В алгоритме отслеживания MOSSE есть две основные части: инициализация и отслеживание. Для инициализации шаблона лица используется рамка, ограничивающая его область, полученная с помощью детектора лица YOLO. Этот шаблон обрезается и центрируется для инициализации фильтра. Чтобы ускорить вычисления, корреляция вычисляется в области Фурье. На Рис. 5 представлена предварительная обработка шаблона для инициализации.

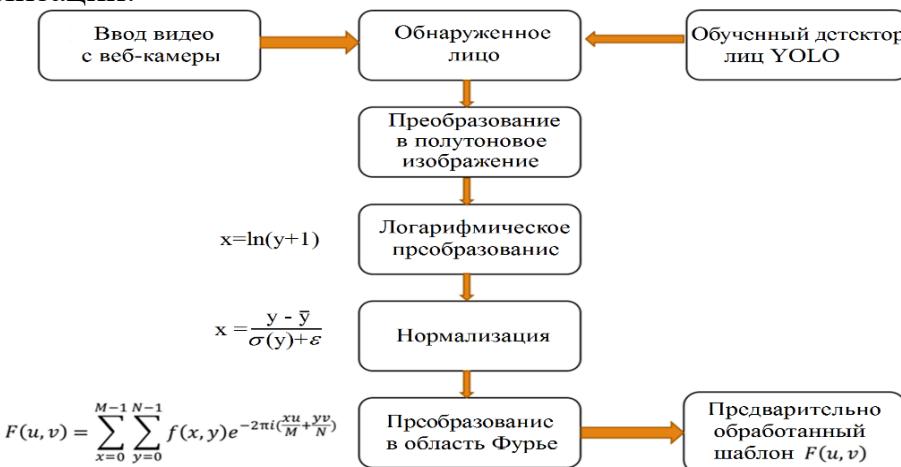


Рис. 5. Предварительная обработка шаблона для инициализации: $F(u, v)$ – шаблон лица в спектральной области, u, v – пространственные частоты, M, N – размер области лица, σ – дисперсия, ϵ – малое приращение

Для задания формы желаемой корреляционной функции сформируем синтетический объект в форме гауссойды с острым пиком. Синтетическое изображение объекта создается с использованием следующего уравнения:

$$g_i = \sum e^{-\frac{(x-x_j)^2 + (y-y_j)^2}{\sigma^2}} \quad (6)$$

где g_i – синтетический объект, x, y – расположение пикселей на изображении, x_j, y_j – местоположение центра объекта, σ – величина дисперсии, задающая радиус объекта (здесь $\sigma = 2$). После выполнения вышеуказанные получили точный фильтр, обеспечивающий значительный пик корреляционной функции в положении элемента и нулевые ее значения во всех других местоположениях при корреляции с шаблоном, на котором был инициализирован фильтр. Точный фильтр может быть построен следующим образом:

$$H^* = \frac{G_i \odot F_i^*}{F_i \odot F_i^* + \varepsilon} \quad (7)$$

где H^* – комплексное сопряжение фильтра, F_i – предварительно обработанный шаблон в области Фурье на i -том шаге, F_i^* – его комплексное сопряжение, G_i – синтетический объект в области Фурье, \odot – поэлементное умножение, $*$ – операция комплексного сопряжения, ε – коэффициент для исключения ошибки деления на ноль.

Из теоремы о спектре свертки следует, что матрицу фильтра H для i -того изображения можно найти в виде:

$$H = \frac{G_i}{F_i} \quad (8)$$

Данный фильтр должен минимизировать сумму квадратов ошибки между фактическим результатом свертки изображения с фильтром и желаемым результатом:

$$\min_{H^*} \sum_i |F_i \odot H^* - G_i|^2 \quad (9)$$

Эта функция имеет как комплексные, так и реальные значения. Решая уравнение MOSSE Поскольку все операции в области Фурье выполняются поэлементно, каждый элемент фильтра H может быть оптимизирован независимо. Чтобы сделать это, запишем ошибку как функцию от отдельных элементов фильтра H_{wv}^* . Затем приравняем частные производные по H_{wv}^* к нулю:

$$0 = \frac{\partial}{\partial H_{wv}^*} \sum_i |F_{wv} \odot H_{wv}^* - G_{wv}|^2 \quad (10)$$

Выражение замкнутой формы для фильтра MOSSE может быть найдено путем решения уравнения (10) относительно H^* :

$$H^* = \frac{\sum_i G_i \odot F_i^*}{\sum_i F_i \odot F_i^* + \varepsilon} \quad (11)$$

Обучающая выборка для первоначального обучения фильтра строится с использованием случайных аффинных преобразований исходного кадра для

генерации восьми изображений f_i . Выходные данные обучения g_i также генерируются в виде пиков, соответствующих искомому центру кадра.

Во время отслеживания объект может изменять внешний вид вследствие изменения ракурса съемки, условий освещенности, нежесткой деформации. Поэтому фильтры должны быстро адаптироваться, чтобы следить за объектами. Для подстройки фильтра будем использовать скользящее среднее:

$$H_i^* = \frac{P_i}{Q_i} \quad (12)$$

$$P_i = \eta(G_i \odot F_i^*) + (1-\eta)P_{i-1} \quad (13)$$

$$Q_i = \eta(F_i \odot F_i^* + \varepsilon) + (1-\eta)Q_{i-1} \quad (14)$$

где P и Q – числитель и знаменатель выражения (11) соответственно, η – скорость обучения ($\eta = 0.125$). Рекуррентная схема обновления фильтра (12-14) придает больший вес последним кадрам и экспоненциально со временем снижает веса предыдущих кадров. На Рис. 6 представлена блок-схема предлагаемого трекера MOSSE.

В качестве предварительно обученной модели извлечения признаков использована модель ResNet50. В качестве критерия качества классификации использована потеря кросс-энтропии. 50-слойная конструкция ResNet состоит из следующих компонентов (Рис. 7):

- Свертки размером 7×7 с 64 ядрами и шагом размером 2.
- Максимальный объединяющий слой с размером шага 2.
- Всего 9 сверточных слоев — $1 \times 1, 64, 3 \times 3, 64$ и $1 \times 1, 256$ ядер. Эти три сверточных слоя повторяются три раза.
- Четыре итерации еще на 12 слоях — были добавлены ядра $1 \times 1, 128, 3 \times 3, 128$ и $1 \times 1, 512$ в каждом.
- Шесть итераций выполнения еще на 18 слоях, каждый из которых имел размеры $1 \times 1, 256, 3 \times 3, 256$ и $1 \times 1, 1024$.
- Девять слоев с ядрами размером $1 \times 1, 512, 3 \times 3, 512$ и $1 \times 1, 2048$, которые повторяются три раза.

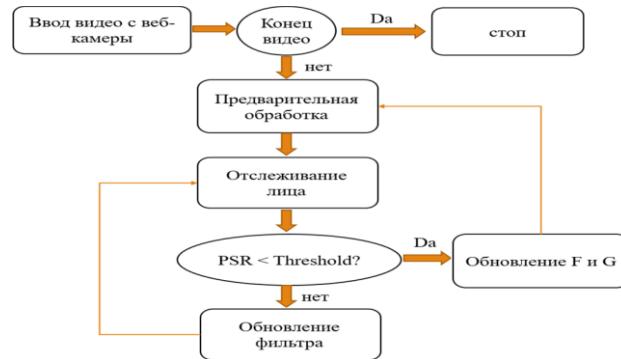


Рис. 6. Блок-схема предлагаемого трекера MOSSE

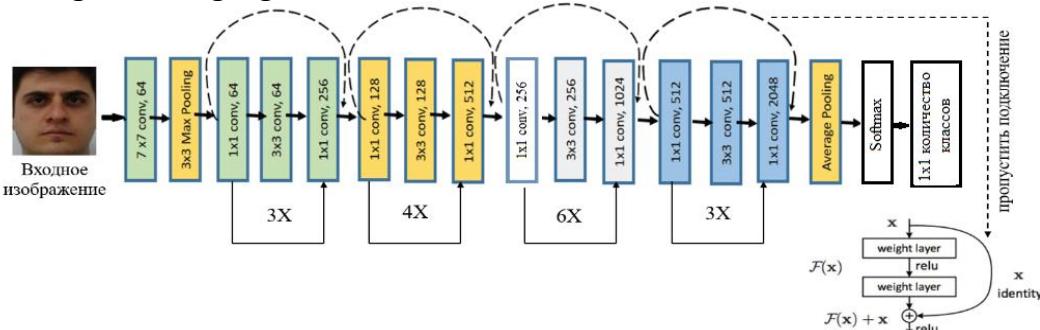


Рис. 7. Предварительно обученная сетевая модель ResNet50

Перед обучением модели осуществлена реализация блокировки на лице из базы данных с черным прямоугольником (например, 50%, 25 %, 20 % на различных местах лиц).

На Рис. 8 представлена схема предлагаемой системы распознавания лиц реального времени при их частичном закрытии.

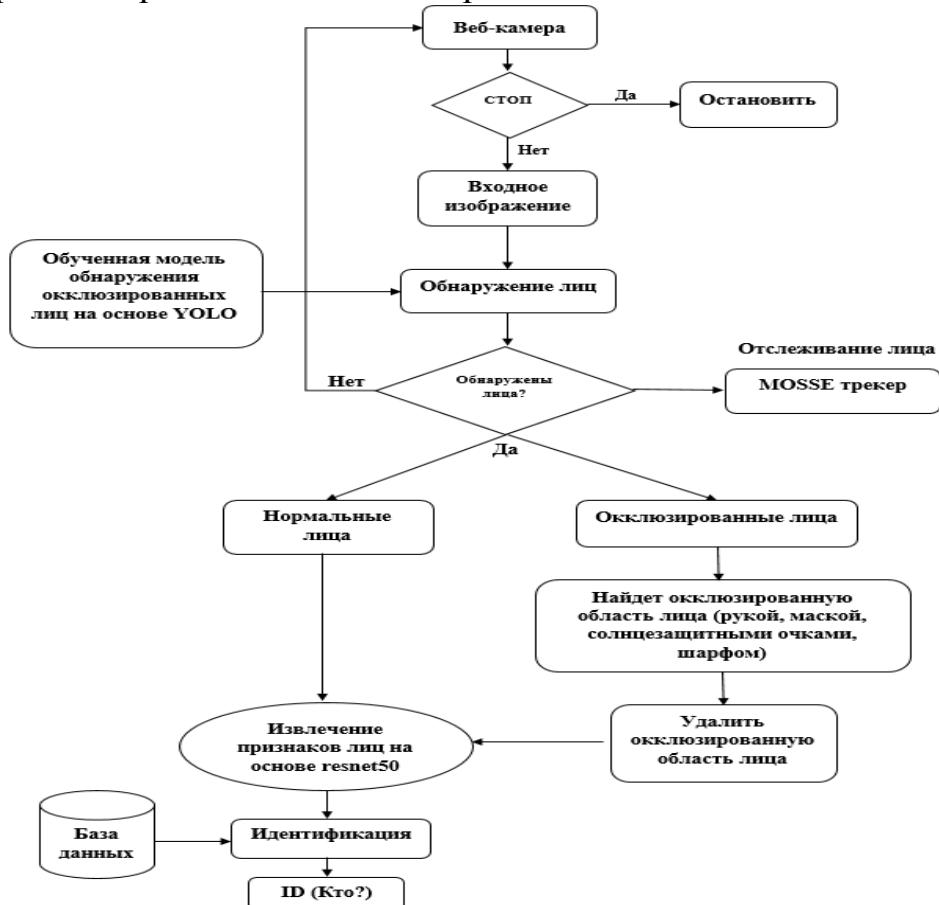


Рис. 8. Схема предлагаемой системы распознавания лиц реального времени при условии реальной окклюзии

В третьей главе представлены результаты экспериментальных исследований предложенных методов. Глава начинается с планирования эксперимента, включающего:

- Выбор средств, включающих не только аппаратное, но и программное обеспечение для реализации предлагаемой системы распознавания лиц;
- Выбор статических и динамических изображений для оценки системы распознавания лиц в условиях частичной окклюзии;
- Методы оценки предложенной системы;
- Эксперименты, сравнение результатов.

Для разработки предлагаемой системы использованы следующие инструменты: Matlab, Image processing toolbox, Computer vision toolbox, Deep learning toolbox. В наборе данных обучающей выборки содержится 5500 изображений из базы данных FDDB, представляющих разных людей под разными углами, в разных условиях освещения, разных цветов кожи, нескольких народов и с разными размерами лица. Кроме того, были дополнительны добавлены изображения лиц популярных людей с окклюзией из сети Интернет

(например, с частичным закрытием лица рукой, медицинской маской, шарфом, солнцезащитными очками) для предварительной обработки систем обнаружения окклюзионных лиц. Также для реализации части распознавания лиц были использованы база данных FEI, включающая 100 человек, а также лица двух наших студентов. При реализации проекта весь набор данных был делится на обучающую выборку (80%) и тестовую выборку (20%).

На Рис. 9 представлена средняя точность предлагаемой системы на основе детектора окклюзированных лиц YOLO и в Таблице 1 представлена средняя точность в режиме реального времени с кадром 300. Была протестирована обученная модель на тестовых данных и получена средняя точность 96 %. При работе с веб-камерой в реальном времени была получена точность 94% при перекрытии лица ладонью, и 100% при окклюзии другими объектами (Таблица 1).

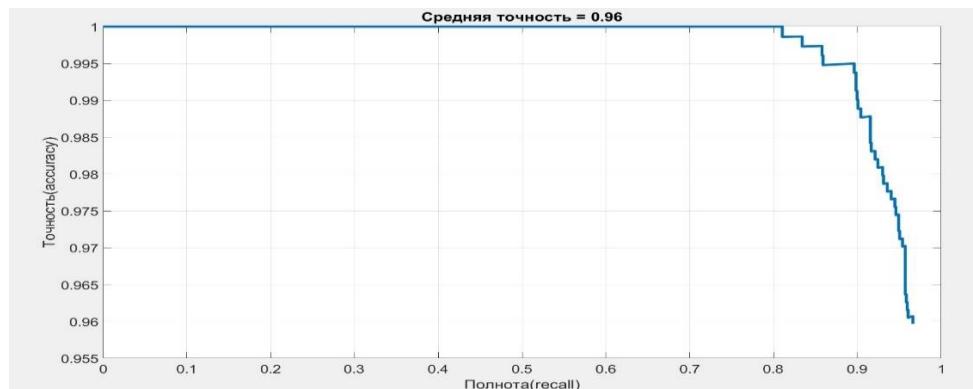


Рис. 9. Средняя точность предлагаемой системы на основе детектора окклюзированных лиц YOLO

Таблица 1.

Точность в режиме реального времени с кадром 300 на основе детектора окклюзированных лиц YOLO

Маска	Рука	Солнцезащитные очки	Шарф	Без окклюзии
100 %	94.37 %	100 %	100%	100 %

В Таблице 2 и на Рис. 10 представлены сравнение результатов на основе других предварительно обученных сетевых моделей. На Рис. 11 представлены некоторые результаты предлагаемой модели системы обнаружения лиц.

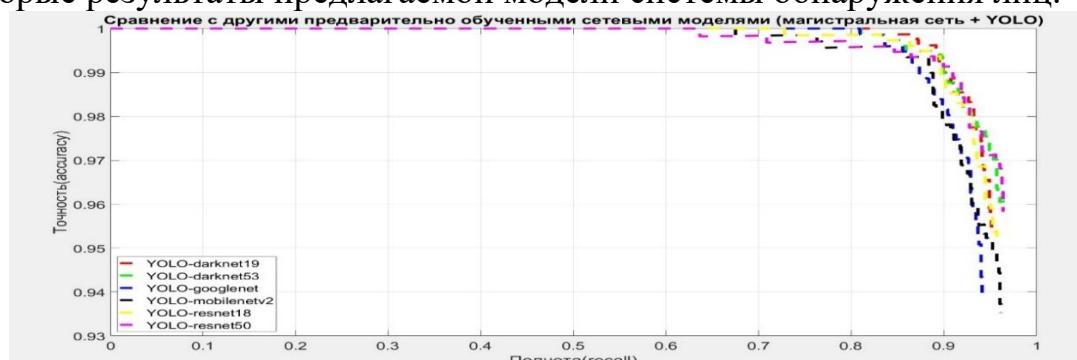


Рис. 10. Сравнение результатов на основе других предварительно обученных сетевых моделей

Таблица 2.

Сравнение результатов на основе других предварительно обученных сетевых моделей

Модель	Средняя точность набора тестовых изображений
Googlenet+YOLO	0.9381%
Mobilenetv2+YOLO	0.9515%
Resnet18+YOLO	0.9539%
Resnet50+YOLO	0.9573%
Darknet19+YOLO	0.9543%
Darknet53+YOLO (Предлагаемая модель)	0.9648%

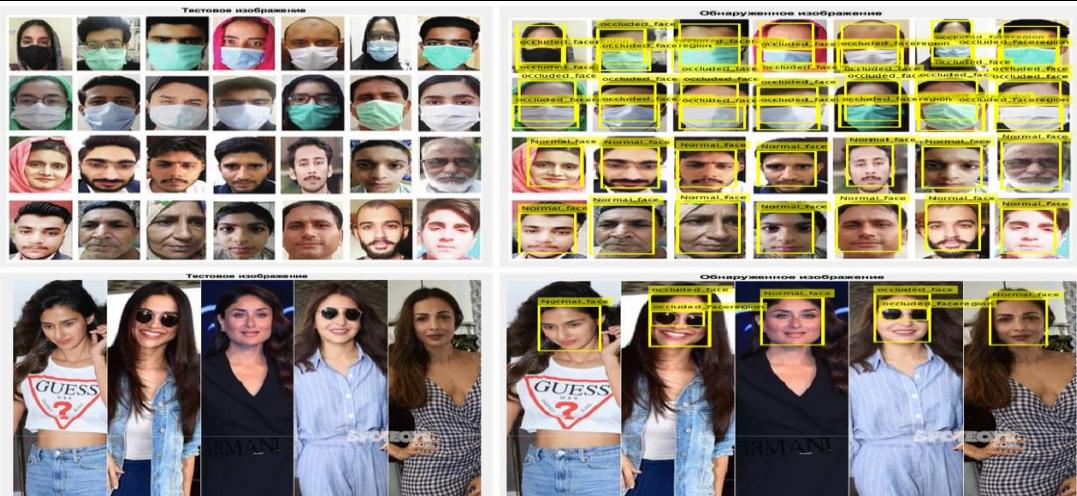


Рис. 11. Некоторые результаты предлагаемой модели для обнаружения окклюзированных лиц

Также при разработке трекера MOSSE, описанные алгоритмы были протестированы в режиме захвата изображения в режиме реального времени. На Рис. 12 представлен пример обработки нескольких различных кадров веб-камеры при помощи оригинального трекера MOSSE в режиме реального времени. Левый Рисунок в каждой паре – результат работы трекера, правый – результат работы детектора лица. Можно увидеть, что когда лицо находится слишком близко или слишком далеко от камеры, то у трекера не был изменен шаблон (ширина и высота). А также, когда лицо было закрыто, либо лицо быстро перемещалось, трекер терял объект и больше не обновлял фильтр.

Работа трекера была исследована на выборках нормального и частично закрытого изображения лица, по 500 последовательных кадров в каждой выборке. Сопровождение считалось успешным, если соотношение пика корреляционной функции к окрестности (PSR) превышало порог 12.

Количество успешных и неуспешных отслеживаний для нормального лица составило 335 и 165 соответственно. Количество успешных и неуспешных отслеживаний для частично закрытого лица составило 296 и 204. Достигнутая скорость обработки составила 195 кадров в секунду.



Рис. 12. Результат тестирования оригинального трекера MOSSE

На Рис. 13 представлен результат оригинального трекера MOSSE в режиме реального времени для нормальных лиц. Здесь точки над порогом показывают успешные отслеживания, а точки под порогом – обозначают потерю объекта.

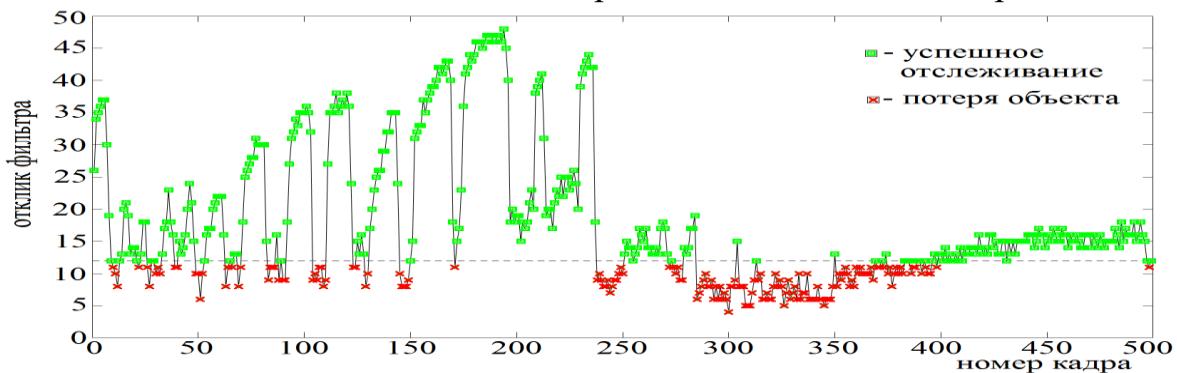


Рис. 13. Результат оригинального трекера MOSSE (с нормальным лицом).

На Рис. 14-15 представлены результаты тестирования предлагаемого трекера MOSSE в режиме реального времени. Условия эксперимента совпадают с предыдущим.



Рис. 14. Результаты тестирования предлагаемого трекера MOSSE

Количество успешных и неуспешных отслеживаний для предложенного трекера составило 483 и 17 соответственно для нормального лица, а для частично закрытого – 470 и 30. Этот результат существенно лучше, чем у оригинального трекера.

Скорость обработки при этом снизилась и составила 120 кадров в секунду: трекеру необходимо дополнительное время на повторное обнаружение лица в случае его потери.

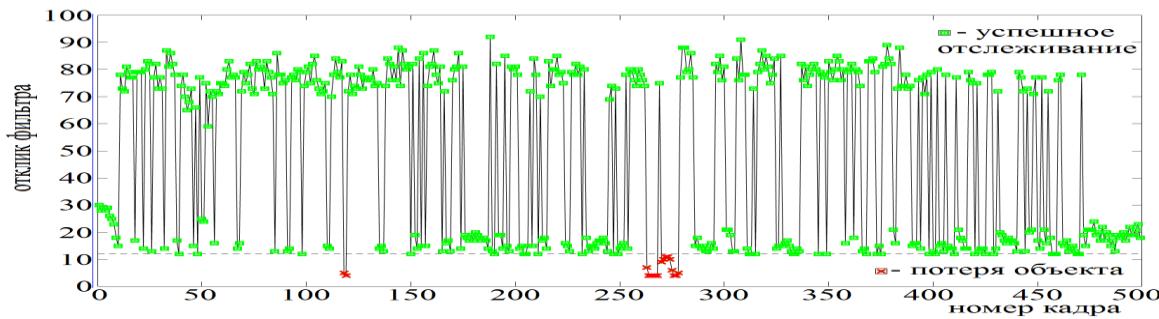


Рис. 15. Результаты тестирования предлагаемого трекера MOSSE для незакрытого лица

Приведём далее результаты комбинированных методов, представленных ранее, для решения задачи системы распознавания лиц при их частичном закрытии в режиме реального времени.

В Таблице 3 представлено сравнение точности результатов различных походов распознавания лиц на основе тестовой базы данных FEI.

Таблица 3.

Сравнение точности результатов различных походов распознавания лиц на основе тестовых базы данных FEI

Методы	Исходные лица	Блокирующий прямоугольник	Исходные лица + блокирующий прямоугольник
HOG+SVM	0.9093	1.0000	0.9515
HOG+FitcKNN	0.9096	1.0000	0.9159
LBP+SVM	0.9482	0.9900	0.9702
LBP+FitcKNN	0.9693	1.0000	0.9230
SURF+BoW+SVM	0.9781	1.0000	0.9934
ResNet50+FitcKNN	0.9900	0.9967	0.9991
ResNet50+SVM	0.9938	1.0000	1.0000
ResNet 50 (с потерей перекрестной энтропии)	0.9961	1.0000	0.9990

В Таблице 4 представлены результаты работы различных методов на основе естественных окклюзированных лиц (медицинских масок, солнцезащитных очков, рук и нормальных лиц). Судя по приведенным выше результатам, у них очень высокая точность на тестовом наборе данных. Однако другие известные методы имеют ещё меньшую точность при выполнении с реальными данными окклюзии. Сначала были подготовлены изображения с окклюзией с помощью маски, руки и солнцезащитных очков на изображениях лица. Затем была определена точность этих данных (всего 100 человек) на основе предложенной схемы. Здесь очевидно, что предложенный метод все еще сохраняет работоспособность не только на тестовых данных, но и в реальных данных с окклюзией. На Рис. 16-17 (левый – результаты трекера и правый – результаты распознавания) представлены результаты предложенного метода на изображении и в режиме реального времени при работе с веб-камерой.

Таблица 4.

Сравнение точности результатов различных походов распознавания лиц на основе реальных окклюзированных лиц

Точность/ Затраченное время – сек. (\approx)	Медицинская маска	Цвет- ная маска	Черн- ая маска	Маска N-95	Рука	Солн- цезащитные очки	Нормаль- ный
HOG+SVM	69/274	60/266	65/265	68/266	38/280	70/265	96/268
HOG+ FitcKNN	70/18	62/18	70/18	74/18	38/18	73/18	98/18
LBP+SVM	80/122	82/124	84/124	85/123	59/130	87/125	89/122
LBP+ FitcKNN	87/11	83/11	90/11	89/11	60/11	92/11	98/11
SURF+Bo W+SVM	37/19	29/19	39/18	24/19	64/20	88/18	99/8
ResNet50+ FitcKNN	99/27	99/30	99/28	93/27	96/27	98/28	100/28
ResNet50+ SVM	100/31	98/32	99/30	92/31	95/29	96/30 сек.	100/34
ResNet 50 (с потерей перекрест ной энтропии)	100/13	99/13	100/12	95/14	97/14	98/12	100/10

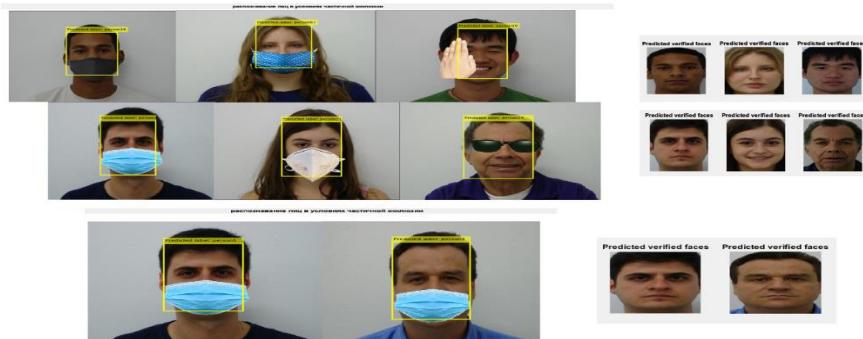


Рис. 16. Результаты предлагаемой системы

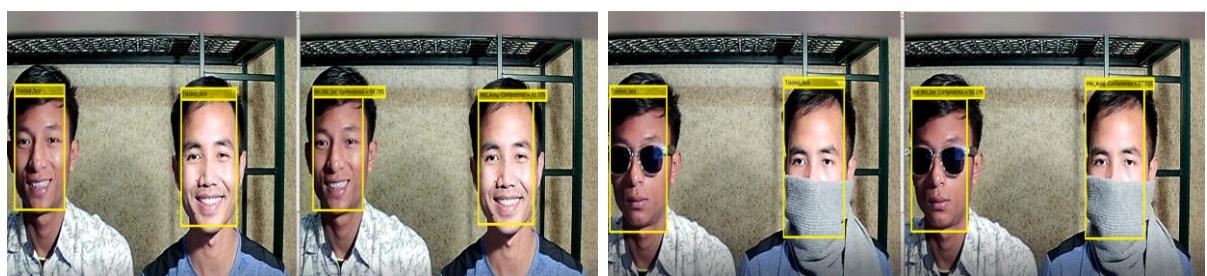


Рис. 17. Результаты по распознаванию лица с окклюзией в режиме реального времени

ЗАКЛЮЧЕНИЕ

В данной диссертационной работе решена актуальная научно-техническая задача исследования и разработки алгоритмов идентификации лиц на статических изображениях и видеопоследовательностях в условиях частичной закрытости лиц.

Выполненная диссертационная работа для достижения поставленных целей и задач содержит следующие результаты:

1. Выполнен анализ обзора существующих методов и алгоритмов, используемых в системах распознавания лиц, с целью предложить состав системы и подход к распознаванию лиц при их частичном закрытии.

2. Разработан алгоритм YOLO, основанный глубокого обучения, позволяющий обнаружить лица не только с высокой точностью, но и позволяет находить лицо на видео в режиме реального времени, с высокой скоростью, а также при реальной окклюзии. По сравнению с результатами других предварительно обученных сетевых моделей, предлагаемая модель имеет среднюю точность 96%. Это более эффективно и точно не только на нормальных базах данных, но и при частичном закрытии.

3. Разработан комбинированный подход на основе трекера MOSSE и детектора YOLO, дающий возможность отслеживать лица не только с высокой скоростью, но и в разных ракурсах, при разных условиях освещения и при окклюзии. По сравнению результатов с оригинальным трекером, предложенный комбинированный трекер получил количество 483 успешных и 17 неуспешных отслеживаний нормальных лиц на основе 500 кадров в режиме реального времени. А для частично закрытого – 470 и 30. Эти результаты существенно лучше, чем у оригинального трекера.

4. Для этапа распознавания была использована предварительно обученная сетевая модель ResNet50. Для 102 разных людей (100 человек базы FEI и двое наших студентов) с общим количеством изображений почти 4000, была получена точность более 98% как для исходных изображений, так и изображений с окклюзией. Далее проведено сравнение точности результатов различных подходов распознавания лиц на основе реальных окклюзированных лиц. Предложенный метод сохраняет работоспособность не только тестовых данных, но и в реальных данных с окклюзией, тогда как другие известные методы дают более низкую точность.

5. Разработана методика применения совокупности алгоритмов для идентификации и отслеживания различных персон на изображениях лиц и в видео, не только в условиях частичных закрытий (окклюзий), но и сложных фонов, различных цветов кожи, различных ракурсов с применением метода обнаружения лиц, метода корреляционного фильтра, метода извлечения признаков и классификации.

6. На основе предложенного комбинированного подхода разработано программное обеспечение для идентификации лиц путём захвата и распознавания изображения с видеокамеры в реальном времени, с учетом возможного частичного перекрытия области лица. Программное обеспечение и

отдельные его компоненты могут быть использованы в различных областях, таких как контроль доступа, связанный с обеспечением безопасности.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Aung H., Bobkov A. V., Tun N. L., Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network // International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russia, 2021. P. 697-702. (0,38 п.л./0,2 п.л.).
2. Aung H., Bobkov A. V., Aye B., Real-Time Face Tracking Based on the Kalman Filter // International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russian Federation, 2022. P. 842-846. (0,38 п.л./0,2 п.л.).
3. Tun N. L., Gavrilov A., Tun N. M., Trieu D. M., Aung H., Hyperspectral Remote Sensing Images Classification Using Fully Convolutional Neural Network // IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia, 2021. P. 2166-2170. (0,6 п.л./0,1 п.л.).
4. Tun N. L., Gavrilov A., Tun N. M., Trieu D. M., Aung H., Remote Sensing Data Classification Using A Hybrid Pre-Trained VGG16 CNN- SVM Classifier // IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia, 2021. P. 2171-2175. (0,6 п.л./0,1 п.л.).
5. Бобков А. В., Аунг Х., Идентификация человека по видеозображению в реальном времени на основе сетей YOLOv2 и VGG-16 // Автомат. и телемех., 2022, № 10, С. 94–104. (0,75 п.л./0,5 п.л.).
6. Бобков А. В., Аунг Х., Разработка системы отслеживания лиц в режиме реального времени на основе корреляционного фильтра MOSSE // Издательство «Автоматизация. Современные технологии», 2023. Т. 77. № 7. С. 324 - 329. (0,75 п.л./0,5 п.л.).
7. Тун Н. М., Гаврилов А. И., Тун Н. Л., Паинг П. П., Аунг Х., Автоматизированная система шумоподавления с использованием метода глубокого обучения для распознавания лиц // Современная наука: Актуальные проблемы теории и практики. Сер. Естественные и Технические Науки, 2020, №06, С. 136-142. (0,72 п.л./0,1 п.л.).
8. Бобков А. В., Аунг Х., Алгоритм распознавания лиц при их частичном закрытии // Издательство «Автоматизация. Современные технологии», 2024 . Т. 78. № 2. С. 77 - 87. (1,19 п.л./0,7 п.л.).
9. Бобков А. В., Аунг Х., Обнаружение лиц в реальном времени на основе алгоритма YOLO // статья в сборнике трудов конференции, источник: наука, технологии и бизнес, 2022. С. 184-191. (0,44 п.л./0,22 п.л.).
10. Бобков А.В., Хтет Аунг., Разработка системы обнаружения и отслеживания объектов в реальном времени // статья в сборнике трудов конференции, Королёвские чтения XLVII Академические чтения по космонавтике, 2023. С. 206-207. (0,25 п.л./0,13 п.л.).