

**УДК 004.01**

**КОМПЬЮТЕРНОЕ ЗРЕНИЕ. МЕТОДЫ РАСПОЗНАВАНИЯ ЛИЦ  
COMPUTER VISION. FACE RECOGNITION METHODS**

Р. Д. УЛЬЯНОВ - студент, Институт информационных технологий и радиоэлектроники, кафедра ИСПИ, группа ПРИ-118, E-mail: [mstr.calling@gmail.com](mailto:mstr.calling@gmail.com)

М. И. ОЗЕРОВА - научный руководитель, к.т.н, инж., проф. кафедры ИСПИ, Институт информационных технологий и радиоэлектроники, E-mail: [ozarovam@rambler.ru](mailto:ozarovam@rambler.ru)

R. D. Ulyanov - undergraduate, Vladimir state university, E-mail: [mstr.calling@gmail.com](mailto:mstr.calling@gmail.com)

M.I. OZEROVA - candidate of technical sciences, Vladimir state university, E-mail: [ozarovam@rambler.ru](mailto:ozarovam@rambler.ru)

**Аннотация:** В данной статье описывается компьютерное зрение, производится описание и сравнение различных методов распознавания лиц.

**Abstract:** This article describes computer vision, describes and compares various face recognition methods.

**Ключевые слова:** компьютерное зрение, метод.

**Key words:** computer vision, method.

### **Введение**

Компьютерное зрение[1] является одной из актуальных направлений в IT области. Данное направление представляет собой комплекс методов, реализующих обработку двумерных или трехмерных объектов, а также звуковой информации и применяется для обеспечения безопасности объектов и инженерных систем, поиска определенного человека по фото, в

бытовой сфере для контроля качества продуктов и автоматической уборки помещений, а также в военной сфере для создания беспилотных ракет и других самоуправляемых объектов, а также для организации автоматического управления транспортных средств.

### **Применение компьютерного зрения в современном мире**

Компьютерное зрение – это теория и технология создания машин, которые могут производить обнаружение, слежение и классификацию объектов. В настоящее время данная дисциплина является одним из важных направлений в информационной индустрии. Компьютерное зрение можно описать как дополнение к биологическому зрению. При помощи компьютерного зрения возможна реализация движения беспилотных автомобилей и роботов, идентификации человека по лицу, для управления компьютером при помощи движения глаз, в бытовых приборах, при создании цифровых вывесок, в камерах видеонаблюдения.

Раскрывая примеры применения компьютерного зрения, можно привести несколько примеров.

Компания NVIDIA выпустила суперкомпьютер NVIDIA Drive PX 2 с технологией DriveNet, служащий основой для беспилотных автомобилей автомобильных компаний Tesla, Audi, Volvo, Mercedes-Benz, BMW. Данная технология представляет собой самообучающееся компьютерное зрение, работающее на основе нейронных сетей. Система оснащена множеством датчиков и радаров, которые способны распознавать дорожную разметку, окружение, транспортные потоки. В 2020 году компания NVIDIA совместно с Audi планируют тест драйв полностью автономного автомобиля, основанного на суперкомпьютере Xavier AI.

NVIDIA представила продукт Co Pilot – искусственный интеллект, который способен считывать движение губ, распознавать лица, произнесенные слова и направления взгляда. Благодаря его интеграции в

машины управляемость им увеличится. В целесообразности интеграции данного программного решения стоит заметить, что искусственный интеллект во время движения автомобиля распознает примерно 90% сказанных слов, в то время как человек может слышать лишь 10%.

Компьютерное зрение также способно заботиться и о вашем здоровье. Компания Smarter выпустило устройство FridgeCam, стоимость которого составляет всего лишь \$100, но оно способно определять, что именно находится в холодильнике, срок годности товаров и подбирает список рецептов из перечня ингредиентов. Благодаря этому готовка становится в разы проще и безопаснее.

Благодаря компьютерному зрению возможно создание цифровых вывесок, изображение которых можно изменить. Благодаря этому возможно создание анимированной картинки на баннерах, флагах, вывесках. Компания Panasonic представила возможность проецирования изображения на флаги.

Компьютерное зрение также применяется и в мобильных устройствах. В ноябре 2019 года компания Lenovo выпустила первый смартфон, который поддерживает платформу Google Tango. Данная платформа включает в себя комбинацию датчиков и ПО совместно с компьютерным зрением. В этом же году компания Asus выпустила смартфон с поддержкой Google Tango и Daydream VR.

В заключении к приведенным примерам стоит сказать, что область применения компьютерного зрения достаточно обширна. В бытовой технике в последнее время активно используются технологии компьютерного зрения. В настоящее время созданы роботы пылесосы, которые способны убирать комнаты по заранее запрограммированному маршруту и обходить препятствия, дроны, способные летать без помощи человека. В военной отрасли компьютерное зрение применяется для конструирования самонаводящихся ракет, а также при конструировании

самоуправляемых роботов для разведки. В программах компьютерное зрение применяется при построении 3D изображения из 2D (AutoCAD), также при редактировании изображений для улучшения их качества. Это объясняет тот факт, что компьютерное зрение достаточно популярная область в IT сфере.

### **Методы компьютерного зрения. Технология распознавания лиц**

Методы компьютерного зрения способны обработать входную информацию любого формата: двумерные и трехмерные изображения, видео, аудио. Примером технологии обработки звука при помощи машинного обучения может служить сервис Sounds от компании Mail.Ru в рамках платформы Mail Cloud Solutions. Видео в процессе обработки сервисами компьютерного зрения преобразуется в массив кадров, каждый из которых обрабатывается отдельно. Примером сервиса для обработки видео с применением компьютерного зрения может служить облачный сервис от компании Google Cloud Videos Intelligent API. Данный сервис позволяет индексировать ролики, осуществить поиск определенного объекта в видео или посмотреть все его вхождения. Сервисом с применением машинного обучения, обрабатывающим изображения, может служить программный продукт Microsoft Azure от одноименной компании. Реализация методов компьютерного зрения ориентирована на ресурсы исполняющей машины, требуемой производительности и области применения. Системы компьютерного зрения включают в себя ряд общих функций [1]

1. *Получение объектов.* Это могут быть изображения 2D или 3D формата; звук; видео.
2. *Предварительная обработка.* Перед применением алгоритмов объекты необходимо обработать. Обработка входных данных включает следующие методы: удалить шум, настроить яркость и

контрастность для изображений для успешного обнаружения объекта, проверить правильность загрузки объекта, установить предельную громкость для звуковых файлов, масштабирование изображений для повышения их качества.

3. *Выделение деталей.* Детали различного уровня сложности выделяются из входного файла. Например, для изображения выделяются геометрические примитивы: линии, границы, кромки и локализованные точки интереса, такие как капли, точки или углы; детали, имеющее более сложное строение, относятся к структуре, форме или движению.
4. *Детектирование и сегментация.* На данном этапе обработке анализируются выбранные детали и принимается решение, какие из них представляют интерес для дальнейшей обработки. Примером может служить выделение группы точек или сегментация объекта на изображении.
5. *Высокоуровневая обработка.* На завершающем этапе обработки формируется список, содержащий данные о найденных объектах. Например, для изображений этот список может содержать набор точек или участок изображения, в котором предположительно находится объект.

На схеме ниже представлен список функций компьютерного зрения

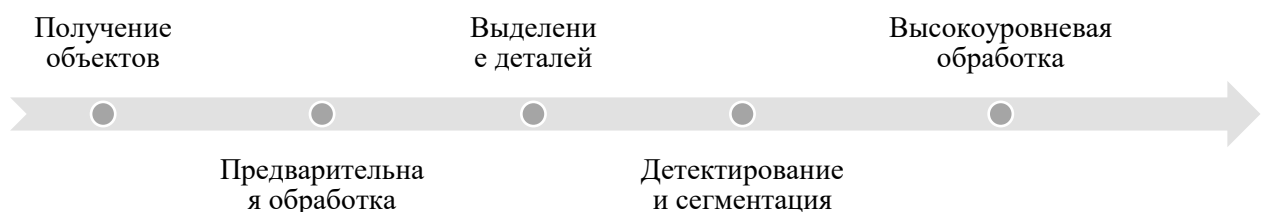


Рисунок 1 – Функции компьютерного зрения

В данной статье [1,3,5] рассматриваются 5 основных методов обработки изображений, осуществляющих распознавание лиц, а также

производится сравнение данных методов. Список рассматриваемых методов приведен ниже

1. Метод гибкого сравнения на графах (Elastic graph matching)
2. Нейронные сети
3. Скрытые марковские модели (СММ, НММ)
4. Метод главных компонент (principal component analysis)
5. Активные модели внешнего вида (Active Appearance Models, ААМ)

Ниже представлена схема методов компьютерного зрения, включающая описание преимуществ и недостатков каждого из них.

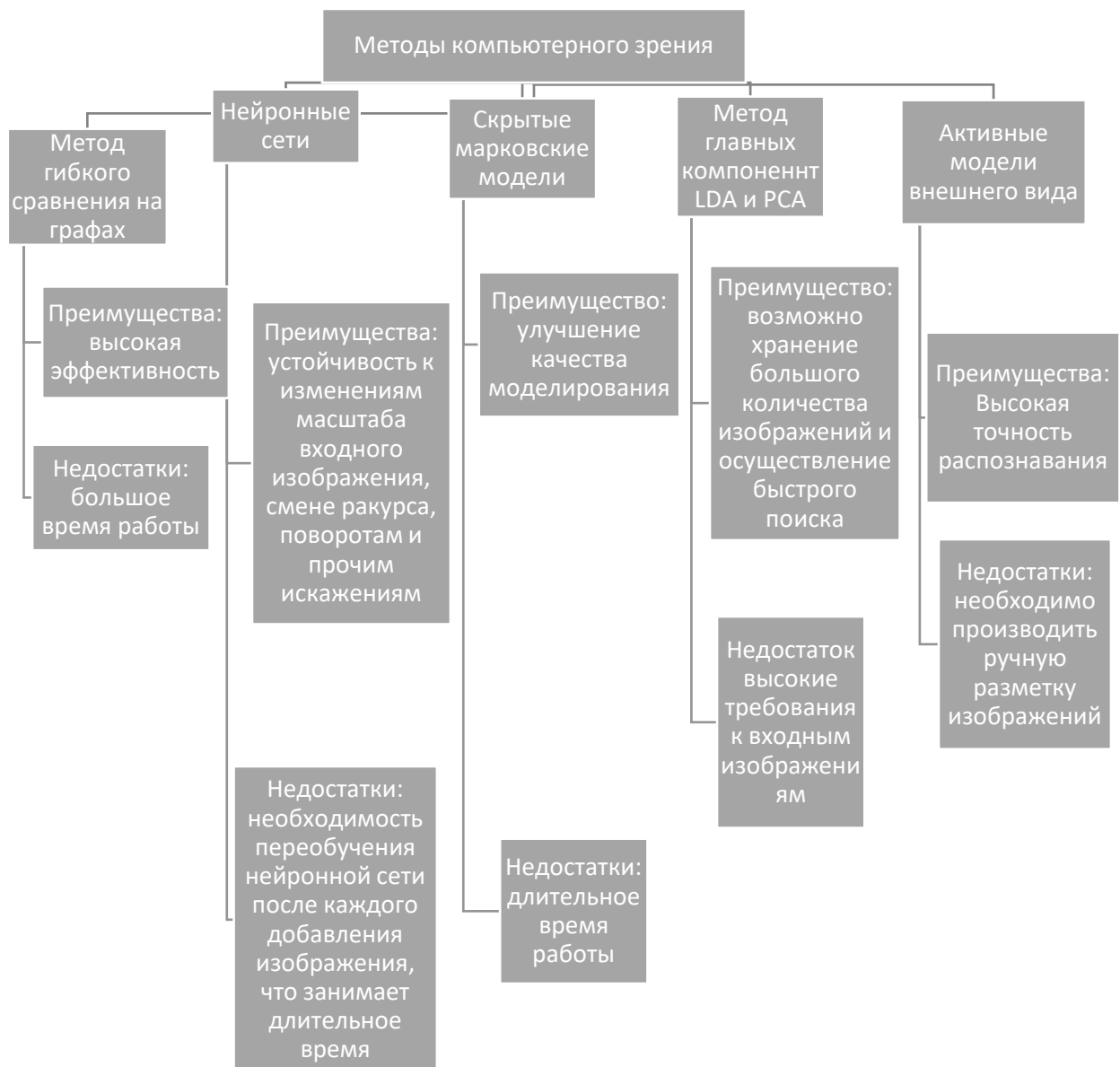


Рисунок 2 – Схема методов компьютерного зрения.

Эти методы не являются универсальными к применению, поскольку у каждого из них присутствуют преимущества и недостатки. Ниже представлено описание методов, где выявляются их «сильные» и «слабые» стороны, а также представлен оптимальный метод распознавания лиц, выявленный в результате сравнения.

### **Метод гибкого сравнения на графах**

Данный метод [1,3] основан на представлении лиц в виде графов, имеющих взвешенные вершины и ребра. В процессе распознавания один из графов остается неизменным, в то время как второй максимально подстраивается под первый граф. В данном методе графы могут быть представлены в двух вариантах: прямоугольной решетки или структуры, которая состоит из антропометрических точек лица – точек, позволяющих наиболее точно описать фигуру человека. В вершинах вычисляемого графа формируются значения признаков, для реализации вычислений применяется фильтры Габора или их упорядоченных наборов – Габровский вейвлет. Ребра графа взвешиваются расстояниями между смежными вершинами. Затем при помощи целевой функции деформации, учитывающую степень деформации ребер графа и различие между вычисленными в вершинах признаками, вычисляется различие между двумя графами. Вычисленная разность между графами называется дискриминационной характеристикой. Процесс деформации происходит в результате смещения каждой вершины графа на некоторое расстояние в определенном направлении, относительно заданной траектории. При этом выбирается позиция, в которой разница откликов фильтров Габора (значений признаков) в вершинах деформируемого графа и соответствующих им вершинах эталонного графа будет минимальной. Данная операция поочередно выполняется для всех ребер графа и

выполняется до тех пор, пока не будет достигнута минимальная сумма разностей всех вершин эталонного и деформируемого графов. Результат, возвращаемый целевой функцией деформации, будет являться степенью различия входного изображения и эталонного графа. Данный метод позволяет организовать поиск изображения из массива изображений, для этого исходное изображение (деформируемый граф) сравнивается со всеми изображениями (эталонными графами), возвращается минимальное значение из множества значений целевой функции деформации. Точность распознавания у данного метода, согласно измерительным тестам, составляет 95%-97%. Однако, данный метод имеет большое время работы. По данным измерений на ЭВМ с 23 процессорами оно составило около 25 секунд.

*Преимущество метода:* высокая эффективность. Успешное распознавание при различных эмоциональных выражениях лица и угла поворота до 15 градусов.

*Недостаток метода:* большое время работы.

### **Нейронные сети**

В настоящее время существует множество видов нейронных сетей [1,3]. Наиболее распространенной из всех является сети прямого распространения, в которых входной сигнал распространяется в прямом направлении, постепенно переходя от слоя к слою. Данные сети также называют многослойный персептрон, в основе их работы лежит многослойный персептрон Руммельхата. Нейронные сети обучаются при помощи наборов обучающих примеров, суть обучения нейронных сетей представляет собой настройку весов межнейронных связей при решении оптимизационной задачи при помощи метода градиентного спуска. В процессе обучения нейронных сетей автоматически извлекаются ключевые признаки, определения степени их важности и формировании



взаимосвязей между ними. В результате обученная нейронная сеть может применить полученный в процессе обучения опыт на неизвестные наборы данных для их решения. Наилучший результат при распознавании лиц среди нейронных сетей, согласно статистическим данным, имеет сверточная нейронная сеть (Convolutional Neural Network). Данная сеть была предложена французским ученым Я. Лекуновым в 1988 году. В персептроне, полносвязанной нейронной сети, каждый нейрон связан со всеми нейронами, находящимися в предыдущем слое, к тому же каждая связь имеет свой индивидуальный весовой коэффициент. В сверточной нейронной сети при выполнении операции свертки используется лишь небольшая матрица весов, которую сдвигают по всему обрабатываемому слою – входному изображению, генерируя после каждой итерации сигнал для активации нейрона следующего слоя с аналогичной позицией. Стоит отметить, что нейронные сети обладают высокой точностью распознавания изображений: при тестировании сверточной нейронной сети на группе изображений с различными искажениями, уровнями освещенности, масштабом и применении поворотов результат составил 96%.

*Преимущество метода:* устойчивость к изменениям масштаба входного изображения, смене ракурса, поворотам, смещениям и прочим искажениям. Высокая точность распознавания изображений.

*Недостатки метода:* при добавлении нового эталонного лица в данные для поиска нейронную сеть надо переобучать, используя измененный набор компонентов, что занимает от 1 часа до нескольких дней. При обучении сети требуется решение задач математических задач, таких как: попадание в локальный оптимум, выбор оптимизации шага оптимизации, переобучение. Сложная архитектура представления: множество нейронов, слоев и характер связи.

## **Скрытые марковские модели**

Скрытые марковские модели [1,3] являются одним из статистических методов распознавания лиц. При работе данный метод использует статистические сигналы и пространственные характеристики. Элементами данной модели являются: множество наблюдаемых состояний, множество скрытых состояний, матрица переходных вероятностей и начальная вероятность состояний. Каждому элементу данного метода соответствует своя марковская модель. В процессе распознавания изображения происходит перебор всех сгенерированных для объектов поиска марковских моделей и осуществляется поиск максимальной вероятности с учетом того, что последовательность наблюдений для данного объекта сгенерирована соответствующей моделью. На момент написания статьи данный метод в коммерческой сфере не используется. Точность распознавания изображений у данного метода, согласно результатам тестирования, составляет около 95%.

*Преимущество метода:* Высокая точность распознавания изображений.

*Недостатки метода:* в ходе выполнения алгоритма необходимо совершать множество вычислений, следовательно время работы увеличивается. Для оптимизации задачи разработаны алгоритмы прямого и обратного хода. Необходимо подбирать параметры модели для каждого набора элементов. По которым будет производиться поиск. Также данный алгоритм не обладает различающей способностью – данный алгоритм обучения лишь максимизирует отклик каждого изображения на свою модель, но не происходит минимизация отклика на другие модели.

## **Метод главных компонент**

Метод главных компонент [1,3], или principal component analysis, PCA, является одним из известных методов обработки объектов, в основе

которого лежит метод Карунена-Лоева. Изначально данный метод применялся в статистике для снижения пространства признаков без существенной потери информации. В технологии распознавания лиц его используют для представления лица в виде вектора малой размерности (главных компонент), который сравнивается с множеством векторов группы изображений, по которым осуществляется поиск. Главной целью данного метода является минимизация пространства признаков таким образом, чтобы она максимально описывала текущие образы, описывающие множество лиц. Благодаря использованию данного метода можно выявить различные изменения в обучающей выборке и описать ее при помощи нескольких ортогональных векторов, называемых собственными. Данный список векторов можно применять для кодирования других изображений лиц, которые являются взвешенной комбинацией этих собственных векторов. Применяя определенное число векторов, можно получить сжатую аппроксимацию входному изображению лица, которое впоследствии можно хранить в множестве данных для поиска, например в базе данных, в виде вектора коэффициентов, который также является ключом для поиска лиц в массиве данных. Алгоритм метода главных компонент состоит из ряда операций. Сначала обучающий набор преобразуется в матрицу данных, где каждая строка является один экземпляр отображения лица, представленный в строковом варианте. Все элементы матрицы должны быть приведены к одной длине (самого длинного элемента), а также иметь нормированные гистограммы. Следующим шагом осуществляется нормирование матрицы, вычисляется нулевое среднее и первая дисперсия, а также матрица ковариации. На заключительном этапе работы алгоритма осуществляется сортировка собственных векторов в порядке убывания их значений и оставляют первые  $k$  векторов по правилу  $\sum_{i=1}^k \left( \frac{a_i}{\sqrt{\lambda_i}} \right)^T$ , где  $a_i$  – векторы-

столбцы,  $\sqrt{\lambda_i}$  – корень из дисперсии вектора вдоль главной компоненты  $i$ , координата, которая разделяется для нормализации.

Данный метод активно используется на практике, но в случае изменения освещенности или выражении лица его эффективность снижается. Это объясняется тем, что данный метод выбирает пространство с такой целью, чтобы максимально аппроксимировать набор входных данных, а не осуществить дискриминацию между классами лиц. Для решения данной проблемы был предложен метод, использующий линейный дискриминант Фишера, в литературе имеет названия Fisherface, LDA. Данный метод выбирает область на изображении, которое максимизирует отношение  $\frac{\Phi^T S_b \Phi}{\Phi^T S_w \Phi}$ , где  $S_b = \sum_{i=1}^m N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$  – матрица межклассового разброса и  $S_w = \sum_{i=1}^m \sum_{x \in X_i} (x_i - \bar{x})(x_i - \bar{x})^T$  – матрица внутриклассового разброса,  $m$  – число классов в массиве данных поиска.

#### Отличия методов LDA и PCA

Оба этих метода выполняют уменьшение размерности входных данных при сохранении качества. Однако у них есть отличие.

Метод PCA ищет проекцию данных, при которых классы являются линейно сепарабельны

Метод PCA ищет проекцию данных, при которой будет максимизирован разброс по всей базе данных лиц без учета классов.

Из-за этого различия точность поиска метода LDA выше, чем у метода PCA и составляет 95% против 56%.

*Преимущества метода:* благодаря векторному представлению изображений в массиве данных возможно хранение большого количества изображений и осуществление довольно быстрого поиска. При наличии компонент для поиска, таких как раса, пол, освещение, изображения будут найдены по заданным компонентам.

*Недостаток метода:* к входным изображениям применяются высокие требования: высокая освещенность, одинаковый ракурс, перед применением алгоритмов необходимо провести качественную обработку изображений: удалить шум, настроить яркость и контрастность и т.д.

### **Активные модели внешнего вида**

Активные модели внешнего вида [3,4] – это статистические модели изображений, которые посредством различных деформаций подгоняются под изображение. Данный метод был предложен Т. Кутсом и К. Тейлором в 1998 году. Первоначально данный метод использовался в качестве оценки параметров изображений лиц. Данная модель содержит два типа параметров: параметры, связанные с пиксельной структурой изображения (параметры внешнего вида) и параметры, связанные с формой (параметры формы). Перед применением нужно обучить модель на заранее размеченных от руки изображениях. Каждая метка модели имеет свой номер и определяет базовую точку, которую обученная модель должна находить при анализе нового изображения. Алгоритм работы классического метода активных моделей внешнего вида состоит из следующих шагов. Вначале вычисляются все параметры формы и аффинные преобразования – отражения плоскости саму в себя. Вторым шагом вычисляется вектор ошибки  $\delta t$ . При помощи кусочной деформации изображения происходит извлечение текстуры анализируемого изображения. Следующим шагом вычисляется вектор возмущений  $\delta p = R\delta t$ . На четвертом шаге итерации производится обновление аффинных преобразований и вектора комбинированных параметров путем добавления к текущим значениям полученные значения вектора возмущения. На следующем шаге производится обновление текстуры и формы. Наконец, на последнем шаге итерации проверяется достижение сходимости, в противном случае осуществляется переход ко второй

итерации. Приведенный алгоритм решения классического метода имеет недостаток – медленную сходимость и большие вычисления в процессе работы, в результате чего он ограничен в применении. Данную проблему решает математически эффективный алгоритм активной модели внешнего вида композиции, предложенный Метьюсом и Бейкером, основанный на подходе Лукаса-Канаде. Данный алгоритм использует метод Ньютона для поиска минимума в функции ошибки. Данный алгоритм осуществляет поиск минимальной среднеквадратической ошибки между реальным изображением и шаблоном, в процессе вычислений шаблон деформируется с учетом вектора параметров, который отображает пиксели шаблона на пиксели реального изображения. Поскольку данный алгоритм является классической моделью внешнего вида с применением оптимизации, он наследует все свойства классической модели. Поэтому перед использованием данную модель необходимо обучить. Процесс обучения активной модели внешнего вида обратной композиции включает в себя вычисление группы изображений наискорейшего спуска. Одним из программных продуктов, реализующих данный метод, является библиотека AAMToolbox. Данная библиотека распространяется по лицензии и предназначена только для некоммерческого использования. Точность распознавания у данного метода, по результатам многочисленных тестов, составила 95%.

*Преимущества метода:* Высокая точность распознавания изображений

*Недостаток метода:* данный метод использует изображения частично; для работы алгоритма требуется производить разметку изображения; в процессе сравнения каждое изображение рассматривается как независимое, для него производится индивидуальный расчет, что замедляет время работы алгоритма.

## **Вывод**

В данной статье было приведено описание применения компьютерного зрения в современном мире и пяти основных методов распознавания лиц, а также описано сравнение данных методов по различным критериям. Ключевым параметром для сравнения являлось точность распознавания изображений и время работы метода. Согласно анализу, *наилучшим* методом для практического применения является *метод главных компонент*, использующий линейный дискриминант Фишера. Его точность распознавания составила 95%. Несмотря на то, что существует более точный алгоритм гибкого сравнения на графах, он ограничен в применении из-за длительного времени работы – для его использования вычисления необходимо производить на облачном сервере или использовать ЭВМ с соответствующими системными ресурсами. Также метод главных компонент обладает гибкой настройкой, позволяющей осуществлять поиск по полу, расе, освещению и другим критериям, благодаря чему данный метод выделяется на фоне других.

## **Список используемой литературы**

1. Л. Шапиро, Дж. Стокман. Компьютерное зрение/Computer Vision. — М.: Бином. Лаборатория знаний, 2006. — 752 с. — [ISBN 5-94774-384-1](#) (Дата обращения 17.02.20).
2. Дэвид Форсайт, Жан Понс. Компьютерное зрение. Современный подход/Computer Vision: A Modern Approach. — Москва: «Вильямс», 2004. — 928 с. — [ISBN 5-8459-0542-7](#) (Дата обращения 12.02.20).
3. В. Коломиец. Анализ существующих подходов к компьютерному зрению/Коломиец В. – научная статья – habr.com: интернет портал. – URL: <https://habr.com/ru/company/synesis/blog/238129/> (Дата обращения 12.02.20).

4. И. Губочкин. Активные модели внешнего вида/Губочкин И. – научная статья – habr.com: интернет портал. – Режим доступа URL: [https://habr.com/ru/post/155759/&post=5385365\\_18497/](https://habr.com/ru/post/155759/&post=5385365_18497/) (Дата обращения 12.02.20).
5. Брилюк. Д. Распознавание человека по изображению лица и нейросетевые методы/Д. Брилюк. – научная статья – Минск, 2001 – 83 с (Дата обращения 17.02.20)