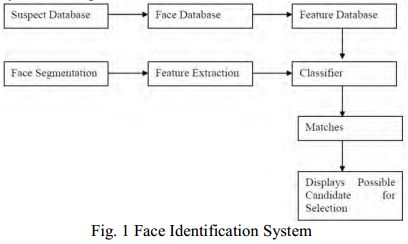
АНАЛЬТЕРНАТИВНЫЕ МЕТОДЫ РЕШЕНИЯ

Метод главных компонент

Метод главных компонент (МГК) включает математический метод, который преобразует ряд вероятностно коррелированных переменных в меньшее количество некоррелированных переменных, называемых основными компонентами. МГК является очень известным подходом, который используется для расчета набора функций для распознавания лица. Это может относиться к любому лицу:

1. Лицо может быть представлено минимальным набором характеристик.

2. Может быть построено с использование небольшой коллекции собственных фотографий. Изображение лица проецируется на некоторое количество шаблонов лиц, называемое собвственными лицами, затем вычисляется разница между этими лицами, которая будет считаться отличительной чертой для изображения. Когда набор собственных лиц вычислен, изображения лица может быть реконструировано взвешенной комбинацией собственных лиц. Веса образуют вектор признаков для представления лица и распознавания. Когда подается новое тестовое изображение, веса высчитаны благодяря проекции изображения на вектор собственных лиц. Классификация между изображениями выходит из сравнения расстояний между векторами весов тестового изображения и обучающего множества (входных данных). Как уже отмечено, можно реконструировать исходное изображение из собственных лиц так, что входное изображение должно точно соответствовать оригинальному изображению, используя все собственные лица, извлеченные из оригинальных изображений. МГК является математическим методом, который основан на преобразовании переменных, с помощью которого можно преобразовать набор коррелированных переменных в набор некоррелированных переменных. Переменные, которые некоррелируют друг с другом – основные компоненты. Количество основных компонентов означает, что некоррелированных должно быть меньше или равно количеству исходных переменных. В такого рода трансформации первые главные компоненты будут иметь самый высокий приоритет, который покажет максимальное отклонение. Это поможет вычислить точность данных. Если набор данных распределен нормально, только тогда главные компоненты могут быть независимыми от других переменных. Снижение размеров может привести к потере информации, но какое-то количество информации так или иначе останется.



Предобработка: (МГК) – на этом этапе сокращается размерность данных. Этап классификации - уменьшенные выходные векторы МГК идут на вход нейросети с обратным распространением ошибки для обучения данных, которые буду использованы для распознавания изображения. Распознавание лица – биометрическая технология используемая, к примеру, с целью наблюдения, поиска разыскиваемых преступников, пропавших без вести детей. Распознавание лица имеет три стадии:

1. обнаружение лица;
2. извлечение характеристик;
3. непосредственно классификация.

АЛГОРИТМЫ И ПРЕДЛОЖЕННЫЙ МЕТОД

Предобработка: (МГК) – на этом этапе сокращается размерность данных. Этап классификации - уменьшенные выходные векторы МГК идут на вход нейросети с обратным распространением ошибки для обучения данных, которые буду использованы для распознавания изображения. Распознавание лица – биометрическая технология используемая, к примеру, с целью наблюдения, поиска разыскиваемых преступников, пропавших без вести детей. Распознавание лица имеет три стадии:

1. обнаружение лица;
2. извлечение характеристик;
3. непосредственно классификация.

МГК-АЛГОРИТМ

Пусть имеется такой набор лиц для обучения: альтернативный текст Среднее лицо по набору может быть определено:

альтернативный текст

Каждое лицо отличается от среднего вектором признаков:

альтернативный текст

Матрица ковариаций строится благодаря вектору:

альтернативный текст

Где матрица альтернативный текст. Множество больших векторов затем подвергают МГК альтернативный текст. Получаем вектор весов альтернативный текст изображения лица. Лицо проецируется на пространство лиц:

альтернативный текст

Для k = 1, M', где M'<= M это номер собственного лица, используемого для распознавания. Веса формируют вектор альтернативный текст.

Линейный дискриминантный анализ

Линейный дискриминантный анализ (линейный дискриминант Фишера Linear Discriminant Analysis, {LDA}) использует такую проекцию пространства изображений на пространство признаков, которая минимизирует внутриклассовое и максимизирует межклассовое расстояние в пространстве признаков. В этих методах предполагается, что классы линейно разделимы.

Матрица для проецирования пространства изображения на пространство признаков выбирается из следующего условия:

Где - матрица межклассовой дисперсии, - матрица внутриклассовой дисперсии.

Может существовать до векторов составляющих базис пространства признаков, где - общее число классов. С помощью этих векторов пространство изображений переводится в пространство признаков.

Поскольку работа непосредственно с матрицей затруднительна из-за ее размерности, используется предварительное уменьшение размерности с помощью метода главных компонент, и затем вычисления производятся в пространстве меньшей размерности:

Где - матрица для проецирования в пространство меньшей размерности (пространство главных компонент).

Обычно тренировочный набор содержит изображения лиц при нескольких базовых условиях освещенности, на основе которых при помощи линейных комбинаций можно получить любые другие условия освещенности. Этот метод дает высокую точность распознавания (около 96{\%}) для широкого диапазона условий освещенности, различных выражений лица и наличия или отсутствия очков. Однако остаются невыясненными вопросы, применим ли этот метод для поиска в больших базах данных, может ли метод работать, когда в тренировочной выборке для некоторых лиц имеется изображение только в одних условиях освещенности. Для задачи детектирования лица с помощью LDA-классы лиц и "не лиц" разбивают на подклассы. Вышеописанный метод основывается на предположении о линейной разделимости классов в пространстве изображений. В общем случае такое предположение несправедливо. Инструмент для построения сложных разделяющих поверхностей предлагают нейросетевые методы.

Синтез объектов линейных классов

Данный метод позволяет синтезировать новые изображения объекта (и в частности, изображения лица) для разных ракурсов. Имеется тренировочный набор изображений лиц и только одно изображение нового объекта в определенном ракурсе. Тренировочный набор состоит из изображений объектов того же класса (класс лиц в данном случае), что и новый объект, и включает в себя изображения различных лиц, причем для каждого лица имеются его изображения в широком диапазоне ракурсов. Для нового объекта, имеющего изображение в ракурсе , осуществляется линейное разложение на изображения объектов из тренировочного набора в том же ракурсе, с вычислением коэффициентов , где - количество объектов в тренировочном наборе. Синтез изображения в новом ракурсе для нового объекта осуществляется сложением изображений из тренировочного набора в ракурсе с теми же коэффициентами: . Таким образом, метод позволяет синтезировать изображения нового объекта в различных ракурсах по изображению в одном ракурсе без привлечения сложных трехмерных моделей. Данный метод является перспективным для синтеза изображений в новых ракурсах без привлечения сложных трехмерных моделей, однако вопрос о качестве и количестве примеров в тренировочном наборе остается открытым.

Гибкие контурные модели

В данных методах распознавание производится на основе сравнения контуров лица. Контуры обычно извлекаются для линий головы, ушей, губ, носа, бровей и глаз. Контуры представлены ключевыми позициями, между которыми положение точек, принадлежащих контуру, вычисляются интерполированием. Для локализации контуров в различных методах используется как априорная информация, так и информация, полученная в результате анализа тренировочного набора. Обычно ключевые точки размещаются вручную на наборе тренировочных изображений. При поиске контуров нового лица используется метод симуляции отжига с целевой функцией из двух составляющих. Для первой из них ищется максимум при соответствии интенсивностей пикселов, извлеченных на перпендикулярной контуру линии, аналогичным пикселам из тренировочной выборки, для второй - при совпадении контура с формой контуров тренировочных примеров. Таким образом, извлекается контур черт лица. Для сравнения изображений используются значения главных компонент, вычисленные на наборе векторов, представляющих собой координаты ключевых точек. Главной задачей при распознавании по контурам является правильное выделение этих контуров. В общем виде эта задача по сложности сравнима непосредственно с распознаванием изображений.

Сравнение эластичных графов

В этом методе (Elastic Bunch Graph Matching) лицо представляется в виде графа, вершины которого расположены на ключевых точках лица таких, как контуры головы, губ, носы, и их крайних точках. Каждая грань помечена расстояниями между ее вершинами. В каждой такой точке вычисляются коэффициенты разложения по функциям Габора для пяти различных частот и восьми ориентаций. Набор таких коэффициентов называется *джетом* (jet). Джеты характеризуют локальные области изображений и служат для двух целей: во-первых, для нахождения точек соответствия в заданной области на двух различных изображениях; во-вторых - для сравнения двух соответствующих областей различных изображений. Каждый коэффициент для точек из одной области различных изображений характеризуется амплитудой , которая медленно меняется с изменением положения точки, и фазой , которая вращается со скоростью, пропорциональной частоте волнового вектора базисной функции. Поэтому в простейшем случае для поиска на новом изображении точки с аналогичными характеристиками в функции подобия фазу не учитывают:

Функция подобия с одним джетом в фиксированной позиции и другим с переменной позицией является достаточно гладкой, для того чтобы получить быструю и надежную сходимость при поиске с применением простейших методов таких, как диффузия или градиентный спуск. Более совершенные функции подобия привлекают информацию о фазе. Для различных ракурсов соответствующие ключевые точки отмечаются вручную на тренировочном наборе. Кроме того, чтобы для одного и того же лица представить различные вариации его изображения в одном и том же графе, для каждой точки используются несколько джетов, каждый из которых может соответствовать различным локальным характеристикам данной точки, например, открытому и закрытому глазу. Процесс распознавания неизвестного лица состоит в сравнении графа изображения лица со всеми остальными графами из набора при помощи функции подобия

Левая сумма характеризует подобие джетов, вычисленное с применением фазочувствительной функции, правая - топографическое соответствие, которое пропорционально квадрату разности расстояний между соответствующими вершинами сравниваемых изображений, - количество вершин, - количество граней, - коэффициент относительной важности топографической информации.

В представленном выше виде метод способен достаточно надежно распознавать при изменениях ракурса до 20 ; при больших углах точность распознавания резко уменьшается, функция подобия оказывается более чувствительной к ракурсу, чем к межклассовым различиям. Дальнейшее развитие метода заключается в извлечении коэффициентов важности на основе анализа обучающей выборки. Для каждого джета симплекс-методом вычисляется коэффициент важности, который затем используется в функции подобия. Коэффициенты важности вычисляются из условия максимизации функции подобия для одного и того же лица и минимизации - для различных лиц. Существуют также более ранние разновидности этого метода, которые не используют изначально определенные ключевые точки и структуры графа. Одни из них используют для сравнения решетки джетов, наложенные на изображение, рис. 3. В неизвестном изображении отыскиваются точки соответствия, и затем по найденным точкам строится искаженная решетка и измеряется мера ее искажения для определения наиболее похожего изображения. В других методах точки извлечения джетов изначально образуют решетку, а затем наименее пригодные для распознавания точки отсеиваются в процессе обучения.



Наложенная на изображение эластичная решетка и ее искаженная версия

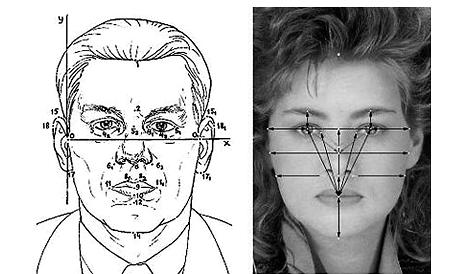
Методы, основанные на геометрических характеристиках лица

Один из самых первых методов - это анализ геометрических характеристик лица. Изначально он применялся в криминалистике и был там детально разработан. Потом появились компьютерные реализации этого метода. Суть его заключается в выделении набора ключевых точек (или областей) лица и последующем выделении набора признаков. Каждый признак является либо расстоянием между ключевыми точками, либо отношением таких расстояний. В отличие от метода сравнения эластичных графов, здесь расстояния выбираются не как дуги графов. Наборы наиболее информативных признаков выделяются экспериментально.

Ключевыми точками могут быть уголки глаз, губ, кончик носа, центр глаза и т. п. рис. 4. В качестве ключевых областей могут служить прямоугольные области, включающие в себя: глаза, нос, рот.

В процессе распознавания сравниваются признаки неизвестного лица с признаками, хранящимися в базе. Задача нахождения ключевых точек приближается к трудоемкости непосредственно распознавания, и правильное нахождение ключевых точек на изображении во многом определяет успех распознавания. Поэтому изображение лица человека должно быть без помех, мешающих процессу поиска ключевых точек. К таким помехам относят очки, бороды, украшения, элементы прически и макияжа. Освещение желательно равномерное и одинаковое для всех изображений. Кроме того, изображение лица должно иметь фронтальный ракурс, возможно, с небольшими отклонениями. Выражение лица должно быть нейтральным. Это связано с тем, что в большинстве методов нет модели учета таких изменений.

Таким образом, данный метод предъявляет достаточно строгие требования к условиям съемки и нуждается в надежном механизме нахождения ключевых точек для общего случая. Кроме того, требуется применение более совершенных методов классификации или построения модели изменений. В общем случае этот метод не является самым оптимальным, однако для некоторых специфических задач все же перспективен. К таким задачам можно отнести документный контроль, когда требуется сравнить изображение лица, полученного в текущий момент, с фотографией в документе. При этом других изображений этого человека не имеется, и, следовательно, механизмы классификации, основанные на анализе тренировочного набора, недоступны.



Идентификационные точки и расстояния: *а* - используемые при криминалистической экспертизе; *б* - наиболее часто применяемые при построении автоматизированных систем идентификации

Сравнение шаблонов.

Сравнение шаблонов (Template Matching) заключается в выделении областей лица на изображении рис. 5, и последующем сравнении этих областей для двух различных изображений. Каждая совпавшая область увеличивает меру сходства изображений. Это также один из исторически первых методов распознавания человека по изображению лица. Для сравнения областей используются простейшие алгоритмы вроде попиксельного сравнения.

Недостаток этого метода заключается в том, что он требует много ресурсов как для хранения участков, так и для их сравнения. Ввиду того, что используется простейший алгоритм сравнения, изображения должны быть сняты в строго установленных условиях: не допускается заметных изменений ракурса, освещения, эмоционального выражения и пр.



Области, входящие в шаблон лица

Скрытые Марковские модели.

Марковские модели являются мощным средством моделирования различных процессов и распознавания образов. По своей природе Марковские модели позволяют учитывать непосредственно пространственно-временные характеристики сигналов, и поэтому получили широкое применение в распознавании речи, а в последнее время - изображений (в частности, изображений лиц). Каждая модель , представляет собой набор состояний , между которыми возможны переходы. В каждый момент времени система находится в строго определенном состоянии. В наиболее распространенных Марковских моделях *первого* *порядка* первого порядка полагается, что следующее состояние зависит только от текущего состояния. При переходе в каждое состояние генерируется наблюдаемый символ, который соответствует физическому сигналу с выхода моделируемой системы. Набор символов для каждого состояния , количество символов . Выход, генерируемый моделью, может быть так же непрерывным. Существуют так же модели, в которых набор символов для всех состояний одинаков. Символ в состоянии в момент времени генерируется с вероятностью . Набор всех таких вероятностей составляет матрицу .

Матрица определяет вероятность перехода из одного состояния в другое: . Считается, что не зависит от времени. Если из каждого состояния можно достичь любого другого за один переход, то все , и модель называется эргодической. Модель имеет вероятность начальных состояний , где . Обычно в реальных процессах последовательность состояний является скрытой от наблюдения и остается неизвестной, а известен только выход системы, последовательность наблюдаемых символов , где каждое наблюдение - символ из , и - число наблюдений в последовательности. Поэтому такие модели называют *скрытыми* Марковскими моделями (Hidden Markov Models, ).

Модель с настроенными параметрами может быть использована для генерирования последовательности наблюдений. Для этого случайно, в соответствии с начальными вероятностями выбирается начальное состояние, затем на каждом шаге вероятность используется для генерации наблюдаемого символа, а вероятность - для выбора следующего состояния. Вероятность генерирования моделью последовательности состояний :

Где - последовательность состояний. Предполагается, что наблюдения статистически независимы.

В распознавании образов скрытые Марковские модели применяются так. Каждому классу соответствует своя модель . Распознаваемый образ (речевой сигнал, изображение и т. д.) представляется в виде последовательности наблюдений . Затем для каждой модели вычисляется вероятность того, что эта последовательность могла быть сгенерирована именно этой моделью. Модель , получившая наибольшую вероятность, считается наиболее подходящей, и образ относят к классу .

В связи с этим появляются несколько вопросов, называемых тремя основными задачами скрытых Марковских моделей.

1. Имея последовательность наблюдений

и настроенную модель , как оценить вероятность генерации этой моделью данной последовательности наблюдений? Эта задача называется задачей распознавания.

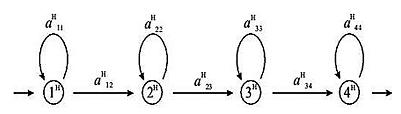
1. Имея последовательность наблюдений

и настроенную модель , как подобрать последовательность состояний , чтобы она была оптимальной (в соответствии с некоторым критерием, аналитически эта задача неразрешима)? Другими словами, это задача объяснения. Она нужна для последующей коррекции параметров модели.

1. Каким образом корректировать параметры модели

, для того чтобы максимизировать ? То есть как сделать так, чтобы модель больше соответствовала своему классу, одним из образов которого является данная последовательность наблюдений (или несколько различных последовательностей)? Это задача обучения.

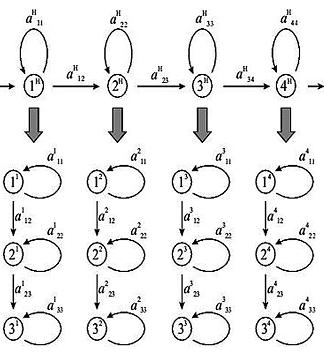
Первая задача имеет точное аналитическое решение, называемое процедурой прямого-обратного прохода. Последующие две задачи не имеют точного аналитического решения. Для решения второй задачи используется алгоритм Витерби, для третей - алгоритм Баума-Вельча. Оба этих метода являются разновидностями градиентного спуска и решаются оптимизационными методами.



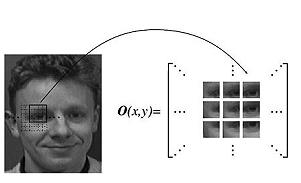
Линейная Марковская модель

Для того чтобы сократить вычисления, в распознавании речи используются линейные модели рис. 6. В таких моделях каждое состояние имеет только одно последующее, так же переход возможен обратно в то же состояние. Такие модели учитывают временные характеристики речевого сигнала: определенный порядок следования участков сигнала, их взаимное расположение, возможность локальных растяжений или сжатий. Это позволяет их применять и в распознавании изображений.

Суть двумерных Марковских моделей заключается в том, что, в отличие от одномерных линейных СММ, они позволяют моделировать искажения изображения и взаимное расположение участков не отдельно по горизонтали или вертикали, а в обоих направлениях одновременно. Для уменьшения вычислительной сложности применяются псевдодвумерные СММ (Pseudo-2D Hidden Markov Models, P2D-HMM). Такая модель состоит из нескольких линейных вертикальных моделей нижнего уровня и одной линейной горизонтальной модели верхнего уровня, на вход которой поступают выходы моделей нижнего уровня, рис. 7. Каждое



Псевдодвумерная скрытая Марковская модель



Извлечение участков-образцов наблюдения

состояние модели верхнего уровня включает в себя последовательность состояний соответствующей модели нижнего уровня. Модели нижнего уровня не связаны между собой. Изначально в модели верхнего уровня были вертикальными. В последующих работах модели верхнего уровня были сделаны горизонтальными (как это и изображено на рисунке), для того чтобы вертикальные модели нижнего уровня могли учесть тот факт, что глаза могут находиться на разной высоте. Таким образом, псевдодвумерная модель позволяет учесть локальные деформации и взаимное расположение участков изображений. Но в отличие от оптических потоков и других методов сопоставления деформациями, псевдодвумерная модель учитывает характер деформаций, а то, какими именно могут быть возможные деформации, псевдодвумерные СММ усваивают в процессе обучения. Другими словами, участок, соответствующий глазу, никогда не будет сопоставлен, например, участку на месте рта.

Пример функционирования СММ. Входом СММ являются квадратные участки изображений (рис. 8). Было обнаружено, что участки, извлекаемые с 75{\%} перекрытием друг с другом, дают наилучшую точность распознавания.

Для СММ важное значение имеет начальная инициализация модели. В качестве начальной инициализации всех моделей используются все изображения из тренировочного набора. Затем модель каждого класса настраивается на свои изображения.

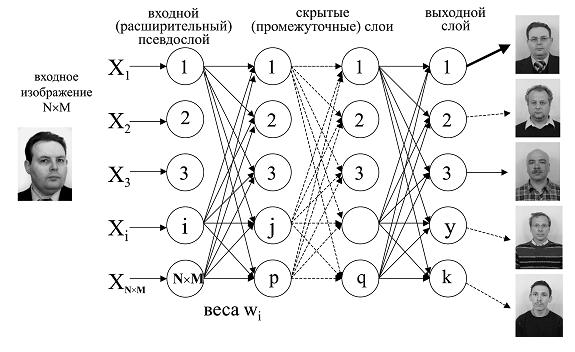
Недостатком СММ является то, что СММ не обладает различающей способностью, т. е. алгоритм обучения только максимизирует отклик каждой модели на свои классы, но не минимизирует отклик на другие классы и не выделяются ключевые признаки, отличающие один класс от другого. Таким образом, похожие классы могут оказаться слабо различимыми и при увеличении объема базы или использования в более широких условиях СММ может оказаться ненадежными.

Многослойные нейронные сети.

Архитектура многослойной нейронной сети (МНС) состоит из последовательно соединенных слоев, где нейрон каждого слоя своими входами связан со всеми нейронами предыдущего слоя, а выходами - следующего. НС с двумя решающими слоями может с любой точностью аппроксимировать любую многомерную функцию. НС с одним решающим слоем способна формировать линейные разделяющие поверхности, что сильно сужает круг задач, ими решаемых, в частности, такая сеть не сможет решить задачу типа "исключающее или". НС с нелинейной функцией активации и двумя решающими слоями позволяет формировать любые выпуклые области в пространстве решений, а с тремя решающими слоями - области любой сложности, в том числе и невыпуклой. При этом МНС не теряет своей обобщающей способности. Обучаются МНС при помощи алгоритма обратного распространения ошибки, являющегося методом градиентного спуска в пространстве весов с целью минимизации суммарной ошибки сети. При этом ошибки (точнее, величины коррекции весов) распространяются в обратном направлении от входов к выходам, сквозь веса, соединяющие нейроны. Простейшее применение однослойной НС (называемой автоассоциативной памятью) заключается в обучении сети восстанавливать подаваемые изображения. Подавая на вход тестовое изображение и вычисляя качество реконструированного изображения, можно оценить, насколько сеть распознала входное изображение. Положительные свойства этого метода заключаются в том, что сеть может восстанавливать искаженные и зашумленные изображения, но для более серьезных целей он не подходит. МНС также используется для непосредственной классификации изображений - на вход подается или само изображение в каком-либо виде, или набор ранее извлеченных ключевых характеристик изображения, на выходе нейрон с максимальной активностью указывает принадлежность к распознанному классу (рис. 9).

Нейрон с максимальной активностью (здесь первый) указывает принадлежность к распознанному классу. Если эта активность ниже некоторого порога, то считается, что поданный образ не относится ни к одному из известных классов. Процесс обучения устанавливает соответствие подаваемых на вход образов с принадлежностью к определенному классу. Это называется "обучением с учителем". В применении к распознаванию человека по изображению лица такой подход хорош для задач контроля доступа небольшой группы лиц. Он обеспечивает непосредственное сравнение сетью самих образов, но с увеличением числа классов время обучения и работы сети возрастает экспоненциально, и поэтому для таких задач, как поиск похожего человека в большой базе данных, требует извлечения компактного набора ключевых характеристик, на основе которых можно производить поиск.

В частности, МНС может использоваться для классификации изображений лиц на основе таких характеристик, как расстояния между некоторыми специфическими частями лица (нос, рот, глаза). Также существуют гибридные системы, например объединение с марковской моделью. В классической МНС межслойные нейронные соединения полносвязны, и изображение представлено в виде одномерного вектора, хотя оно двумерно. Архитектура сверточной НС направлена на преодоление этих недостатков. В ней используются локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов), общие веса (обеспечивают детектирование некоторых черт в любом месте изображения) и иерархическая организация с пространственными подвыборками (spatial subsampling). Сверточная НС (СНС) обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, искажениям. Архитектура СНС состоит из многих слоев, каждый из которых имеет несколько плоскостей, причем нейроны следующего слоя связаны только с небольшим числом нейронов предыдущего слоя из окрестности локальной области (как в зрительной коре человека). Веса в каждой точке одной плоскости одинаковы (сверточные слоя). За сверточным слоем следует слой, уменьшающий его размерность путем локального усреднения. Затем опять сверточный слой, и так далее. Таким образом, достигается иерархическая организация. Более поздние слои извлекают более общие характеристики, меньше зависящие от искажений изображения. Обучается СНС стандартным методом обратного распространения ошибки. Сравнение МНС и СНС показало существенные преимущества последней как по скорости, так и по надежности классификации. Полезным свойством СНС является и то, что характеристики, формируемые на выходах верхних слоев иерархии, могут быть применимы для классификации по методу ближайшего соседа (например, вычисляя евклидово расстояние), причем СНС может успешно извлекать такие характеристики и для образов, отсутствующих в обучающем наборе. Для СНС характерны высокая скорость обучения и работы. Тестирование СНС на базе данных ORL, содержащей изображения лиц с небольшими изменениями освещения, масштаба, пространственных поворотов, положения и различными эмоциями, показало приблизительно 98{\%} точность распознавания, причем для известных лиц предъявлялись варианты их изображений, отсутствующие в обучающем наборе. Такой результат делает эту архитектуру перспективной для дальнейших разработок в области распознавания изображений пространственных объектов. МНС применяются и для обнаружения объектов определенного типа. Помимо этого, любая обученная МНС в некоторой мере может определять принадлежность образов к "своим" классам, ее можно специально обучить надежному детектированию определенных классов. В этом случае выходными классами будут классы, принадлежащие и не принадлежащие к заданному типу образов.



Многослойная нейронная сеть для классификации изображений

Сети Габоровых вейвлетов (GWN).

Данный метод предназначен для слежения за лицом в реальном времени с помощью Gabor wavelet template (GWT), представляющего собой дискретную линейную комбинацию Габоровых вэйвлетов. Важно отметить, что точность представления лица регулируется числом Габоровых вэйвлетов в GWN, допуская представления множества лиц одним GWT. Такое представление допускает произвольные аффинные преобразования и быструю оценку аффинных параметров методом градиентного спуска. Тем самым, выполняя слежение за лицом, метод определяет и его ориентацию, однако с ограничением на повороты: допускаются, в основном, повороты в плоскости и незначительные пространственные вокруг вертикальной оси.

Нейронные сети Хопфилда.

НС Хопфилда (НСХ) является однослойной и полносвязной (связи нейронов на самих себя отсутствуют), ее выходы связаны со входами. В отличие от МНС, НСХ является релаксационной - т. е. будучи установленной в некое начальное положение функционирует до тех пор, пока не достигнет стабильного состояния, которое и будет являться ее выходным значением. НСХ применяются в качестве ассоциативной памяти и для решения оптимизационных задач. В первом случае НСХ обучается без учителя (например, по правилу Хебба), во втором случае веса между нейронами изначально кодируют решаемую задачу. НСХ бывают синхронными, когда одновременно пересчитываются все нейроны, и асинхронными, когда пересчитывается случайно выбранный нейрон. Для исследования динамики функционирования НСХ используются методы Ляпунова. Асинхронная НСХ всегда сходится к устойчивым точкам, а аттракторами синхронной НСХ являются устойчивые стационарные точки и предельные циклы длины два. Таким образом, НСХ из начального состояния сходится к ближайшему локальному минимуму энергии сети, состояние нейронов в котором и будет восстановленным образом для задач распознавания, и решением - для оптимизационных задач. Для поиска глобального минимума применительно к оптимизационным задачам используют стохастические модификации НСХ.

Применение НСХ в качестве ассоциативной памяти позволяет точно восстанавливать образы, которым сеть обучена, при подаче на вход искаженного образа. При этом сеть "вспомнит" наиболее близкий (в смысле локального минимума энергии) образ, и распознает его. Такое функционирование также можно представить как последовательное применение автоассоциативной памяти. В отличие от автоассоциативной памяти, НСХ идеально точно восстановит образ.

Хотя разработка систем идентификации по изображениям человеческих лиц ведется уже несколько десятилетий, задача создания эффективного алгоритма идентификации по изображениям человеческих лиц еще далека от завершения.

Для большинства современных систем автоматического распознавания лиц основной задачей является задача сравнения данного изображения лица с набором изображений лиц из базы данных. Характеристики систем автоматического распознавания лиц в этом случае оцениваются путем определения вероятностей ошибочного отказа в распознавании (для изображения лица, присутствующего в базе, принимается решение как о неопознанном лице) и ошибочного распознавания. В дополнение к вероятностям ошибок для оценки системы автоматического распознавания лиц часто используется оценка устойчивости к возмущению изображений, вызываемая комбинацией со сложными фонами, изменчивостью освещения, изменению прически, и т. д.

Выбор алгоритма или группы алгоритмов идентификации лиц для создания практических систем автоматического распознавания лиц должен основываться на системе оценок рабочих характеристик конкретной системы, соответствующей ограничениям, связанным с условиями эксплуатации данной конкретной системы, и определяется непосредственно на этапе проектирования.

Ссылки:

МГК <http://masters.donntu.org/2014/fknt/muradina/library/article10.htm>

В настоящее время в связи с возрастающими потребностями в осуществлении качественной автоматической верификации в охранных системах или криминалистической экспертизе большое внимание уделяется процессам, лежащим в основе визуального восприятия лиц. Поскольку человек видит предметы реального мира, не задумываясь над последовательностью операций, ведущих к формированию образа в сознании, восприятие объектов кажется нам простым и естественным. Но узнавание и распознавание объекта есть результат сложной «перцепторной» деятельности, основанной на работе центральной нервной системы.

Процесс зрительного восприятия человека [1] начинается с момента, когда возбуждения, возникающие на сетчатке глаза, доходят до первичной зрительной коры, где они, проецируясь на соматотопически организованных пунктах коры, распадаются на составляющие признаки (рис. 1). Такая организация процесса зрительного анализа обусловливается наличием в зрительной коре огромного числа высокодифференцированных нейронов, каждый из которых реагирует лишь на отдельный признак воспринимаемого объекта. Центральный процессор нервной системы обрабатывает информацию, поступающую с нейронов, создавая «перцепторную» гипотезу путем объединения существенных признаков в группы и исключения из рассмотрения несущественных. По результатам анализа данных из ряда альтернативных выбирается наиболее подходящая к данному объекту «перцепторная» гипотеза, и объект считается распознанным.

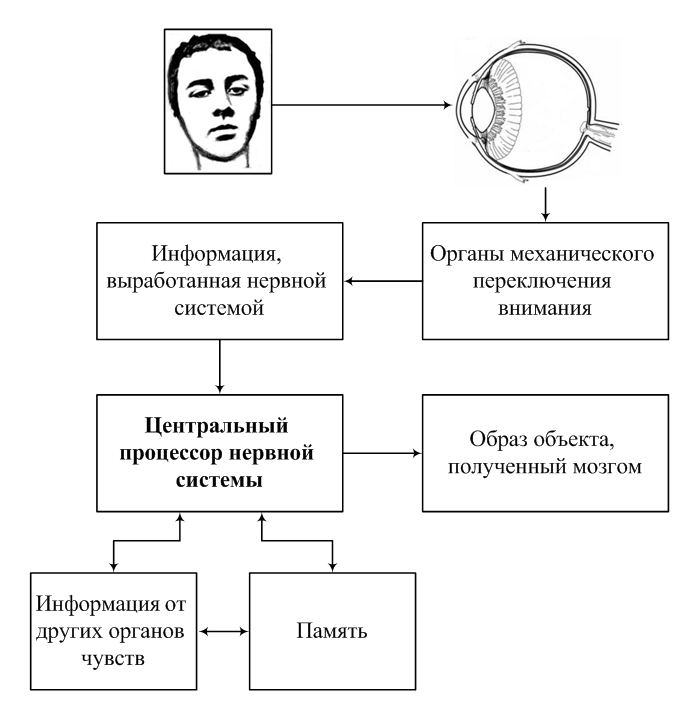


Рис. 1. Процесс зрительного восприятия человека

Работа центрального процессора нервной системы, на вход которого передаются выбранные признаки стационарного объекта, а на выходе выдается результат сравнения объекта с содержащейся в памяти информацией, представляется черным ящиком. Методы автоматической идентификации объектов строятся в попытке воспроизвести его работу и воссоздать зависимости, которым подчиняется анализ выбранных нейронами признаков. Поэтому принцип активного распознавания должен закладываться на этапе первичного проектирования систем кибернетического видения слабоконтрастных объектов, в частности, при распознавании человеческих лиц.

Имеют место закономерности, наблюдаемые при экспериментах, в которых фиксировался путь движения глаз при наблюдении лиц различных людей [1]. Их можно сформулировать в виде нескольких особенностей визуального распознавания таких, как:

1. Движение глазной мускулатуры от одной точки лица к другой. Данная особенность обусловлена тем фактом, что из-за крайней малости оптического угла четкого зрения глаз находится в постоянном движении, систематически обследуя отдельные детали объекта и останавливаясь на его ключевых точках. В этом процессе, имеющем рефлекторный характер, важную роль играют сенсорные и моторные аппараты, например, аппарат движения глаз, осуществляющий ориентировочно-исследовательскую деятельность.

2. Фиксация внимания на тональных различиях видимой картины. Поскольку распознаваемые человеком лица являются слабоконтрастными объектами, внимание зрителя непроизвольно фиксируется на градациях освещенности объекта.

3. Группировка зрительных элементов в простые понятия – гештальты [2]. Люди в процессе перцепторного восприятия окружающего мира подсознательно группируют полученные сенсорные данные в готовые объекты.

4. Динамический поиск лучшей интерпретации сенсорных данных. Центральная нервная система человека для принятия оптимального в текущий момент решения непрерывно анализирует полученные данные, соотнося их с накопленной в памяти информацией.

Для решения задачи распознавания объектов (в данном контексте под объектом понимается лицо человека) построено множество методов, которые сильно отличаются друг от друга по требованиям к настройке метода на базу объектов, входным данным, скорости работы и др. При этом сущность процесса распознавания в различных методах остается одной и той же. Соотнесение распознаваемого объекта с базой объектов, которые необходимо идентифицировать, проходит в три этапа:

- выделение того или иного признака объекта;

- объединение признаков в комплексы или классы;

- выбор предполагаемого значения из ряда альтернатив.

Существует целый ряд методов идентификации лиц таких, как метод главных компонент, линейный дискриминантный анализ, сравнение эластичных графов, анализ антропометрических характеристик, неокогнитрон, скрытые Марковские модели. Проведем сопоставление перечисленных методов с процессом визуального распознавания.

В методе главных компонентов признаками объекта являются главные компоненты, т. е. линейные коэффициенты, вычисленные на основе собственных векторов лицеподобной формы [3]. Первые *n* главных компонентов формируют представление в собственном пространстве, соответствующее отдельному объекту. Для сравнения набора компонентов, которые относятся к наблюдаемому объекту, с хранящимися в памяти наборами главных компонентов для других объектов используются методы сравнения такие, как метрики различного вида, нейронные сети (радиально-базисные, карты Конохена) и т.д. Конкретизация метода сравнения зависит от требуемой скорости вычислений, необходимого качества распознавания изображений лиц и поддерживаемых мощностей оборудования. Использование собственных векторов в методе главных компонентов отражает такую особенность человеческого распознавания, как фиксация на градациях изображения, при этом сами по себе вектора являются неким подобием гештальтов. Данный подход требует жестко фиксированных условий для изображений объекта, так как иначе будет невозможно сопоставить ему построенные по базе объектов собственные вектора.

В методе линейного дискриминантного анализа [4] объект представляется как проекция на пространство признаков, в котором базисные дискриминантные вектора близки по виду собственным в методе главных компонентов. При этом проекция выбирается для каждого объекта таким образом, чтобы обособить его от остальных объектов. Проецирование распознаваемого объекта, находящегося в пространстве изображений, на пространство признаков осуществляется с помощью линейного дискриминанта. Основное требование метода – возможность линейно разделить проекции, однако в общем случае данное условие может не выполняться, и метод станет выдавать ошибку. Линейный дискриминантный анализ так же как и метод главных компонентов, учитывает градации изображения лица, в связи с этим для обеспечения работы метода при изменяющихся условиях освещения необходимо использовать предобработку изображения, приводящую его к заданным стандартным условиям (см. далее).

Эластичные графы используются в методе распознавания лиц [5], где признаками идентифицируемого объекта являются вершины графа, расположенные на контурах головы, губ, глаз и др. В каждой вершине вычисляется джет, т.е. набор определенных заранее коэффициентов Габоровых функций, и конкретному лицу соответствует отдельная совокупность джетов для различных областей на лице. Тогда, если необходимо провести сравнение лиц, достаточно сопоставить джеты с использованием функции подобия для сравнения графов. Идея, лежащая в основе метода эластичных графов, напоминает такую особенность визуального распознавания, как движение глаза от одной точки объекта к другой, при этом в рассматриваемом методе не учитываются возможные градации изображения. Даже при использовании коэффициентов важности для джетов эластичные графы будут сравнивать лица лишь с помощью вычисления взвешенных геометрических искажений объекта, а большой объем информации, содержащейся в изменении оттенков цвета на коже лица, остается без внимания.

Следует особо отметить такой метод идентификации, как анализ антропометрических характеристик [6, 7], в котором в качестве признаков используются расстояния между уголками глаз, губ, центрами глаз или их отношения, сходные с вершинами графа в эластичных графах. В дальнейшем они группируются как известный набор ключевых точек или областей лицах [8], причем наиболее информативные признаки и их весовые коэффициенты выбираются экспериментально с привлечением различных математических методов. Сформированный для распознаваемого лица набор признаков сравнивается с базой, содержащей наборы для других лиц, с помощью методов сравнения количественных характеристик таких, как метод *k* ближайших соседей или многослойный персептрон. При сопоставлении метода на основе антропометрических характеристик с визуальным распознаванием можно заметить, что подобно сравнению эластичных графов, этот метод переносит в автоматическую идентификацию объектов последовательный обход «псевдоглазом» ключевых точек объекта. При этом необходимо иметь в виду, что для корректного анализа антропометрических характеристик предварительно устраняются те элементы на лице, которые вносят неточности при идентификации объекта.

В неокогнитроне [9] признаками объекта являются простые образы (линии и образованные ими углы), формирующие совокупность узлов неокогнитрона для каждого объекта из подготовленной базы объектов при различных уровнях абстракции данных в процессе обучения нейронной сети. В результате неокогнитрон, настроенный на конкретную базу объектов, последовательно сравнивает группу признаков для распознаваемого объекта с заложенными при обучении совокупностями признаков для объектов из базы. Все особенности визуального распознавания в том или ином виде отражаются в неокогнитроне, и его можно назвать наиболее эффективным для распознавания методом, если бы не большие вычислительные затраты [9] для обеспечения достаточной скорости распознавания, требующие применение нейрокомпьютеров. Из-за этого на современном уровне развития вычислительной техники широкое использование неокогнитрона затруднительно.

Для метода «скрытые Марковские модели» [10] признаками идентифицируемого объекта являются последовательности наблюдений объекта. Каждому объекту соответствует своя Марковская модель – набор состояний системы. При распознавании объекта проверяются сгенерированные для заданной базы объектов Марковские модели и ищется максимальная из наблюдаемых вероятность того, что последовательность наблюдений для данного объекта сгенерирована соответствующей моделью. Скрытые Марковские модели вобрали в себя все вышеперечисленные особенности человеческого восприятия и распознавания. При этом они не требуют больших вычислительных мощностей и позволяют менять состав сформированной базы без изменения всей совокупности настроенных Марковских моделей, хотя размер базы объектов, для которых обеспечивается корректное распознавание этим методом, ограничен.

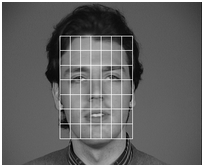
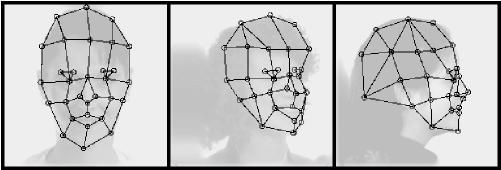
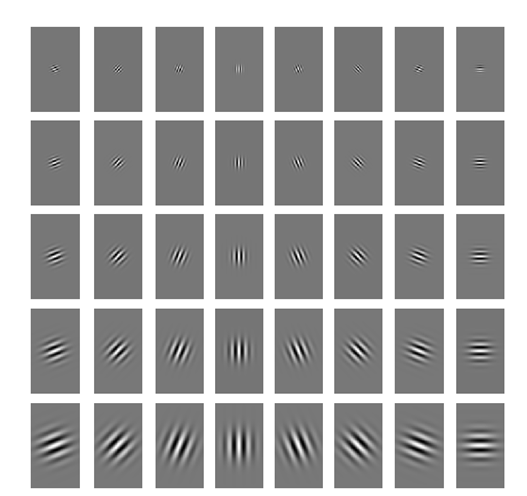
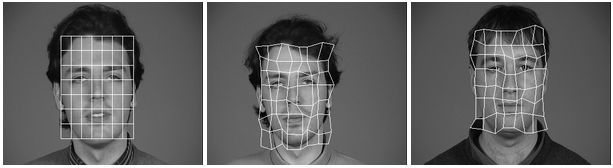
Анализ рассмотренных методов показывает, что они сформированы аналогично процессу визуального восприятия. При этом ни один из этих методов не достигает точности распознавания человеком, хотя и приближается к ней. Данный факт приводит к необходимости учитывать реальные условия для систем видеонаблюдения. Поэтому при реализации любого метода автоматической идентификации лиц с целью повышения его эффективности следует вводить предварительную обработку изображения лица, при которой должен приниматься во внимание ряд нижеприведенных факторов.

Один из факторов, влияющих на качество распознавания, характеризуется особенностями формирования изображения ПЗС-матрицей. В настоящее время в области видеонаблюдения прослеживается устойчивая тенденция перехода от аналоговых видеокамер к цифровым. Следует отметить, что приборы с зарядовой связью (ПЗС), которые применяются для получения цифровых изображений в таких камерах, обладают некоторыми свойствами. Так, сигнал, получаемый с ПЗС-матрицы, в дальнейшем преобразуется в дискретизированное и квантованное изображение. Он содержит шумовую составляющую, которая вызвана квантовым шумом, процессом считывания заряда, накопленного ПЗС, а также работой внешней электроники [11]. Для выделения полезного сигнала необходимо предварительно получить с видеокамеры изображение при равномерной засветке/затемнении и статистически выставить шумовой порог (*n*% максимального отклонения от реального значения сигнала).

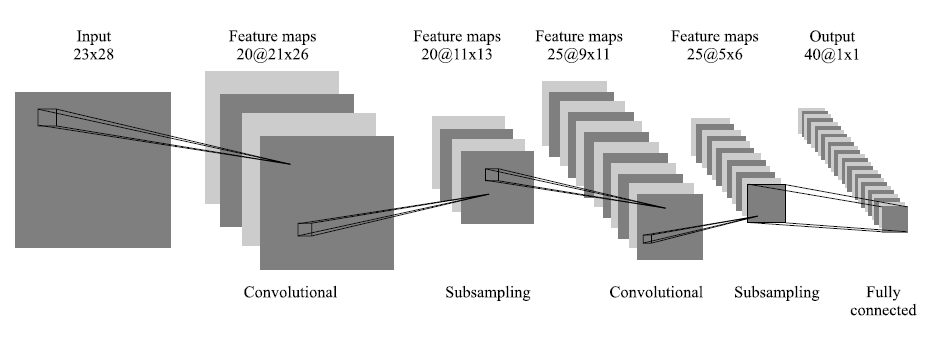
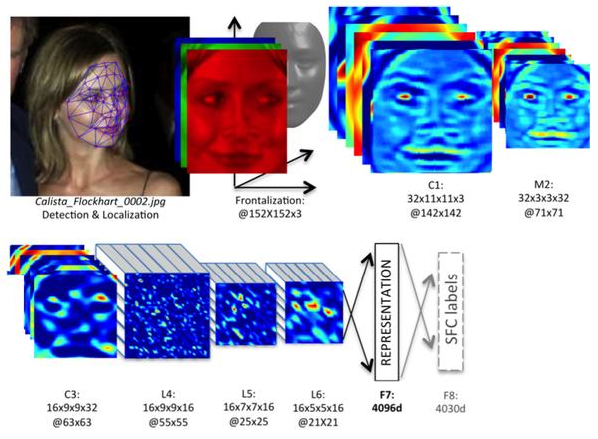
Другими факторами, от которых зависит качество распознавания людей, являются неоднородный яркостный фон изображения, слабая контрастность объекта, наличие мешающих распознаванию элементов таких, как части прически, борода, очки и др., нестандартное положение головы человека.

Таким образом, в результате проведенного исследования подготовлена платформа для разработки метода распознавания людей по изображениям лица с учетом визуального распознавания. Конфигурация метода включает в себя описанное выше выделение полезной составляющей сигнала; приведение изображения к равномерному освещению; удаление по цвету объектов, закрывающих часть лица, и замена скрытых ими участков на наиболее вероятные; нахождение угла поворота головы относительно осей ориентации камеры. Для сравнения анализируемых изображений лиц со сформированной базой предполагается использовать метод скрытых Марковских моделей.

##### **1. Метод гибкого сравнения на графах (Elastic graph matching) [13].**

Суть метода сводится к эластичному сопоставлению графов, описывающих изображения лиц. Лица представлены в виде графов со взвешенными вершинами и ребрами. На этапе распознавания один из графов – эталонный – остается неизменным, в то время как другой деформируется с целью наилучшей подгонки к первому. В подобных системах распознавания графы могут представлять собой как прямоугольную решетку, так и структуру, образованную характерными (антропометрическими) точками лица.   
  
а)   
  
б)   
  
*Пример структуры графа для распознавания лиц: а) регулярная решетка б) граф на основе антропометрических точек лица.*  
  
В вершинах графа вычисляются значения признаков, чаще всего используют комплексные значения фильтров Габора или их упорядоченных наборов – Габоровских вейвлет (строи Габора), которые вычисляются в некоторой локальной области вершины графа локально путем свертки значений яркости пикселей с фильтрами Габора.  
  
   
*Набор (банк, jet) фильтров Габора*  
  
   
*Пример свертки изображения лица с двумя фильтрами Габора*  
  
Ребра графа взвешиваются расстояниями между смежными вершинами. Различие (расстояние, дискриминационная характеристика) между двумя графами вычисляется при помощи некоторой ценовой функции деформации, учитывающей как различие между значениями признаков, вычисленными в вершинах, так и степень деформации ребер графа.  
Деформация графа происходит путем смещения каждой из его вершин на некоторое расстояние в определённых направлениях относительно ее исходного местоположения и выбора такой ее позиции, при которой разница между значениями признаков (откликов фильтров Габора) в вершине деформируемого графа и соответствующей ей вершине эталонного графа будет минимальной. Данная операция выполняется поочередно для всех вершин графа до тех пор, пока не будет достигнуто наименьшее суммарное различие между признаками деформируемого и эталонного графов. Значение ценовой функции деформации при таком положении деформируемого графа и будет являться мерой различия между входным изображением лица и эталонным графом. Данная «релаксационная» процедура деформации должна выполняться для всех эталонных лиц, заложенных в базу данных системы. Результат распознавания системы – эталон с наилучшим значением ценовой функции деформации.   
  
   
*Пример деформации графа в виде регулярной решетки*  
  
В отдельных публикациях указывается 95-97%-ая эффективность распознавания даже при наличии различных эмоциональных выражениях и изменении ракурса лица до 15 градусов. Однако разработчики систем эластичного сравнения на графах ссылаются на высокую вычислительную стоимость данного подхода. Например, для сравнения входного изображения лица с 87 эталонными тратилось приблизительно 25 секунд при работе на параллельной ЭВМ с 23 транспьютерами [15] (Примечание: публикация датирована 1993 годом). В других публикациях по данной тематике время либо не указывается, либо говорится, что оно велико.   
  
**Недостатки:** высокая вычислительная сложность процедуры распознавания. Низкая технологичность при запоминании новых эталонов. Линейная зависимость времени работы от размера базы данных лиц. 

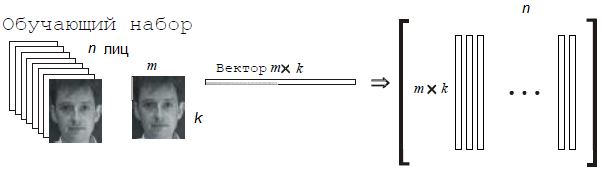
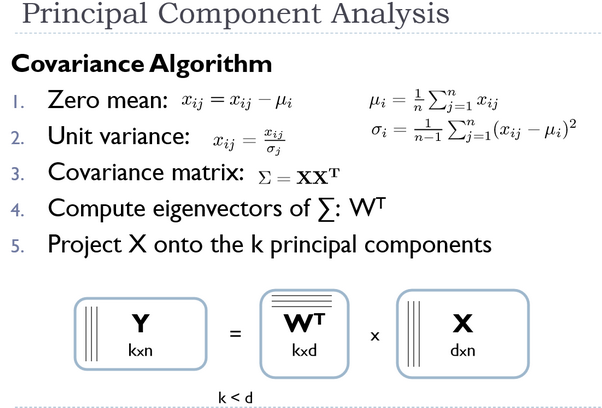
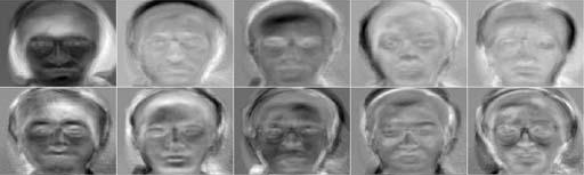
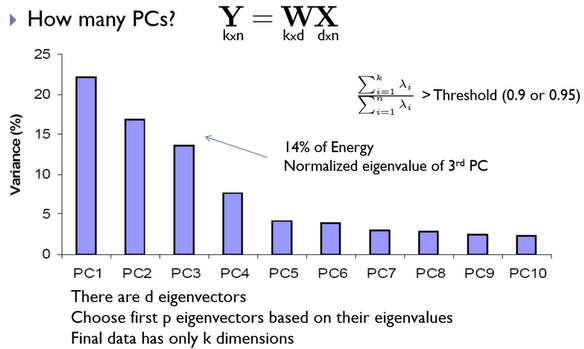
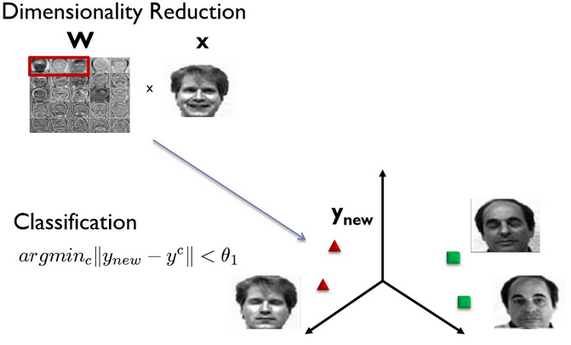
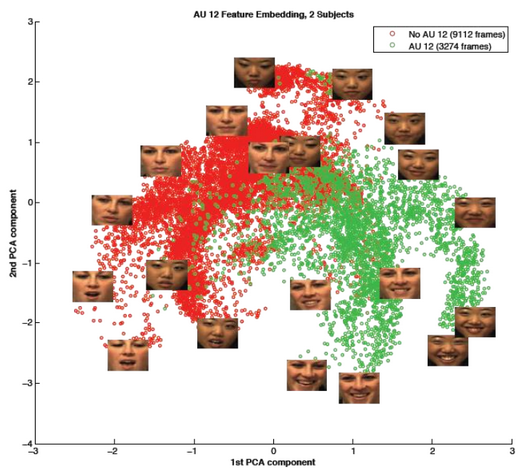
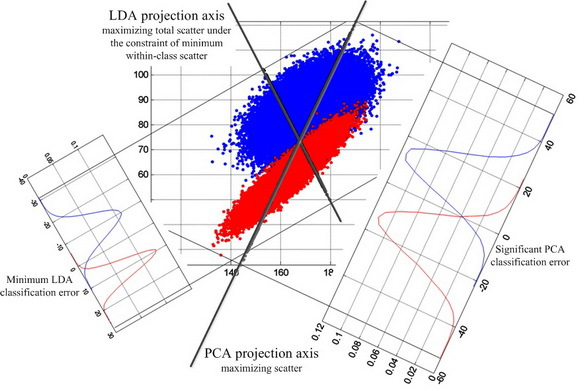
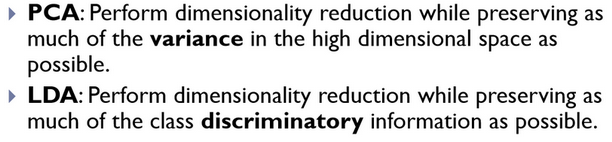
##### **2. Нейронные сети**

В настоящее время существует около десятка разновидности нейронных сетей (НС). Одним из самых широко используемых вариантов являться сеть, построенная на многослойном перцептроне, которая позволяет классифицировать поданное на вход изображение/сигнал в соответствии с предварительной настройкой/обучением сети.  
Обучаются нейронные сети на наборе обучающих примеров. Суть обучения сводится к настройке весов межнейронных связей в процессе решения оптимизационной задачи методом градиентного спуска. В процессе обучения НС происходит автоматическое извлечение ключевых признаков, определение их важности и построение взаимосвязей между ними. Предполагается, что обученная НС сможет применить опыт, полученный в процессе обучения, на неизвестные образы за счет обобщающих способностей.   
Наилучшие результаты в области распознавания лиц (по результатам анализа публикаций) показала Convolutional Neural Network или сверточная нейронная сеть (далее – СНС) [29-31], которая является логическим развитием идей таких архитектур НС как когнитрона и неокогнитрона. Успех обусловлен возможностью учета двумерной топологии изображения, в отличие от многослойного перцептрона.   
Отличительными особенностями СНС являются локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов), общие веса (обеспечивают детектирование некоторых черт в любом месте изображения) и иерархическая организация с пространственными сэмплингом (spatial subsampling). Благодаря этим нововведениям СНС обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.  
  
  
*Схематичное изображение архитектуры сверточной нейронной сети*  
  
Тестирование СНС на базе данных ORL, содержащей изображения лиц с небольшими изменениями освещения, масштаба, пространственных поворотов, положения и различными эмоциями, показало 96% точность распознавания.   
Свое развитие СНС получили в разработке DeepFace [47], которую приобрел   
Facebook для распознавания лиц пользователей своей соцсети. Все особенности архитектуры носят закрытый характер.   
  
   
*Принцип работы DeepFace*  
  
**Недостатки нейронных сетей:** добавление нового эталонного лица в базу данных требует полного переобучения сети на всем имеющемся наборе (достаточно длительная процедура, в зависимости от размера выборки от 1 часа до нескольких дней). Проблемы математического характера, связанные с обучением: попадание в локальный оптимум, выбор оптимального шага оптимизации, переобучение и т. д. Трудно формализуемый этап выбора архитектуры сети (количество нейронов, слоев, характер связей). Обобщая все вышесказанное, можно заключить, что НС – «черный ящик» с трудно интерпретируемыми результатами работы. 

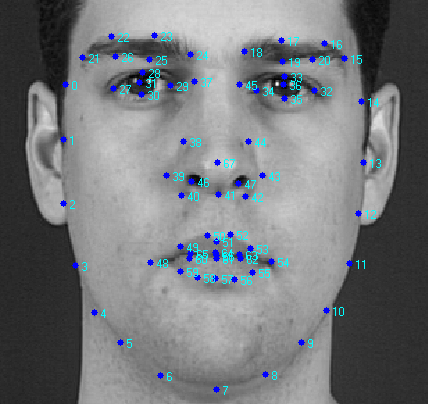
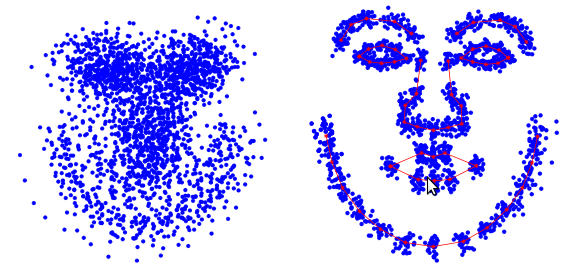
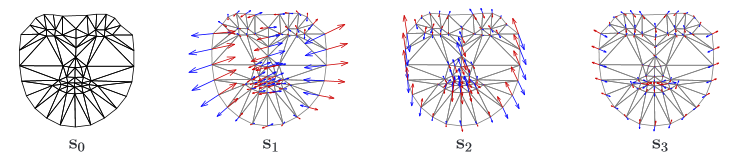
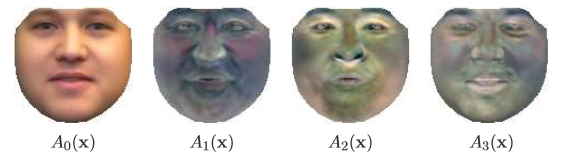
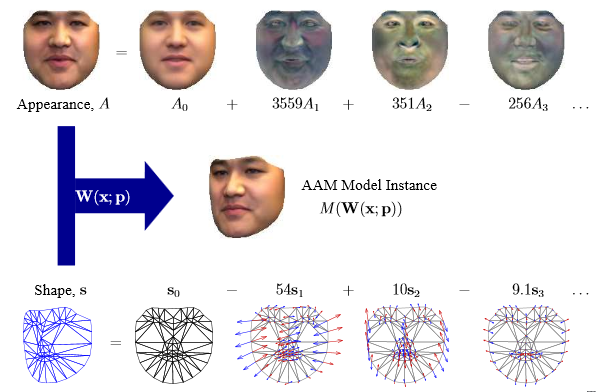
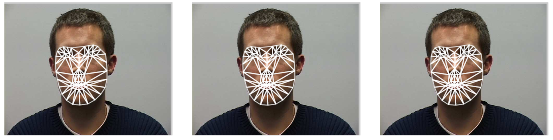
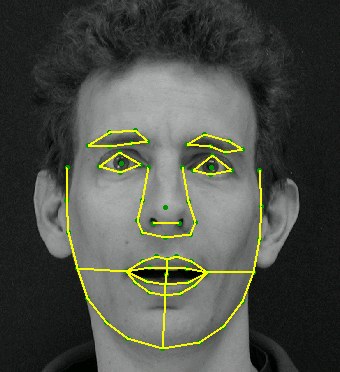
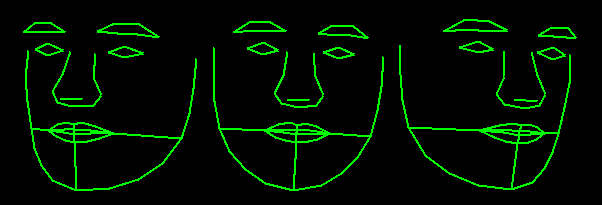
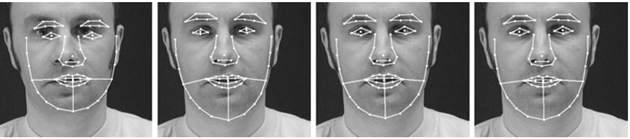
##### **3. Скрытые Марковские модели (СММ, HMM)**

Одним из статистических методов распознавания лиц являются скрытые Марковские модели (СММ) с дискретным временем [32-34]. СММ используют статистические свойства сигналов и учитывают непосредственно их пространственные характеристики. Элементами модели являются: множество скрытых состояний, множество наблюдаемых состояний, матрица переходных вероятностей, начальная вероятность состояний. Каждому соответствует своя Марковская модель. При распознавании объекта проверяются сгенерированные для заданной базы объектов Марковские модели и ищется максимальная из наблюдаемых вероятность того, что последовательность наблюдений для данного объекта сгенерирована соответствующей моделью.  
На сегодняшний день не удалось найти примера коммерческого применения СММ для распознавания лиц.   
  
**Недостатки:**   
— необходимо подбирать параметры модели для каждой базы данных;  
— СММ не обладает различающей способностью, то есть алгоритм обучения только максимизирует отклик каждого изображения на свою модель, но не минимизирует отклик на другие модели. 

##### **4. Метод главных компонент или principal component analysis (PCA) [11]**

Одним из наиболее известных и проработанных является метод главных компонент (principal component analysis, PCA), основанный на преобразовании Карунена-Лоева.   
Первоначально метод главных компонент начал применяться в статистике для снижения пространства признаков без существенной потери информации. В задаче распознавания лиц его применяют главным образом для представления изображения лица вектором малой размерности (главных компонент), который сравнивается затем с эталонными векторами, заложенными в базу данных.   
Главной целью метода главных компонент является значительное уменьшение размерности пространства признаков таким образом, чтобы оно как можно лучше описывало «типичные» образы, принадлежащие множеству лиц. Используя этот метод можно выявить различные изменчивости в обучающей выборке изображений лиц и описать эту изменчивость в базисе нескольких ортогональных векторов, которые называются собственными (eigenface).   
  
Полученный один раз на обучающей выборке изображений лиц набор собственных векторов используется для кодирования всех остальных изображений лиц, которые представляются взвешенной комбинацией этих собственных векторов. Используя ограниченное количество собственных векторов можно получить сжатую аппроксимацию входному изображению лица, которую затем можно хранить в базе данных в виде вектора коэффициентов, служащего одновременно ключом поиска в базе данных лиц.   
  
Суть метода главных компонент сводится к следующему. Вначале весь обучающий набор лиц преобразуется в одну общую матрицу данных, где каждая строка представляет собой один экземпляр изображения лица, разложенного в строку. Все лица обучающего набора должны быть приведены к одному размеру и с нормированными гистограммами.   
  
   
*Преобразования обучающего набора лиц в одну общую матрицу X*  
  
Затем производится нормировка данных и приведение строк к 0-му среднему и 1-й дисперсии, вычисляется матрица ковариации. Для полученной матрицы ковариации решается задача определения собственных значений и соответствующих им собственных векторов (собственные лица). Далее производится сортировка собственных векторов в порядке убывания собственных значений и оставляют только первые k векторов по правилу:   
  
https://habrastorage.org/files/077/2e6/884/0772e6884e704d4db05c2d35297669d5.png  
  
*Алгоритм РСА*  
  
  
*Пример первых десяти собственных векторов (собственных лиц), полученных на обучаемом наборе лиц*  
  
= 0.956\*-1.842\*+0.046 …  
  
*Пример построения (синтеза) человеческого лица с помощью комбинации собственных лиц и главных компонент*  
  
  
*Принцип выбора базиса из первых лучших собственных векторов*  
  
  
*Пример отображения лица в трехмерное метрическое пространство, полученном по трем собственным лицам и дальнейшее распознавание*  
  
  
  
Метод главных компонент хорошо зарекомендовал себя в практических приложениях. Однако, в тех случаях, когда на изображении лица присутствуют значительные изменения в освещенности или выражении лица, эффективность метода значительно падает. Все дело в том, что PCA выбирает подпространство с такой целью, чтобы максимально аппроксимировать входной набор данных, а не выполнить дискриминацию между классами лиц.   
  
В [22] было предложено решение этой проблемы с использование линейного дискриминанта Фишера (в литературе встречается название “Eigen-Fisher”, “Fisherface”, LDA). LDA выбирает линейное подпространство, которое максимизирует отношение:  
  
https://habrastorage.org/files/925/7f8/083/9257f808328d40b08f168e3de1f27b66.png  
  
где https://habrastorage.org/files/ae1/a74/d70/ae1a74d7025b4f3ea35c5d0f8aed274f.png  
  
матрица межклассового разброса, и  
https://habrastorage.org/files/741/c40/e07/741c40e07d4346f5afac59c2d8e32786.png  
  
Матрица внутриклассового разброса; m – число классов в базе данных.   
  
LDA ищет проекцию данных, при которой классы являются максимально линейно сепарабельны (см. рисунок ниже). Для сравнения PCA ищет такую проекцию данных, при которой будет максимизирован разброс по всей базе данных лиц (без учета классов). По результатам экспериментов [22] в условиях сильного бакового и нижнего затенения изображений лиц Fisherface показал 95% эффективность по сравнению с 53% Eigenface.   
  
  
*Принципиальное отличие формирования проекций PCA и LDA*  
  
Отличие PCA от LDA  
  


##### **5. Active Appearance Models (AAM) и Active Shape Models (ASM) (**[**Хабраисточник**](http://habrahabr.ru/post/155759/&post=5385365_18497/)**)**

**Active Appearance Models (AAM)**  
Активные модели внешнего вида (Active Appearance Models, AAM) — это статистические модели изображений, которые путем разного рода деформаций могут быть подогнаны под реальное изображение. Данный тип моделей в двумерном варианте был предложен Тимом Кутсом и Крисом Тейлором в 1998 году [17,18]. Первоначально активные модели внешнего вида применялись для оценки параметров изображений лиц.  
Активная модель внешнего вида содержит два типа параметров: параметры, связанные с формой (параметры формы), и параметры, связанные со статистической моделью пикселей изображения или текстурой (параметры внешнего вида). Перед использованием модель должна быть обучена на множестве заранее размеченных изображений. Разметка изображений производится вручную. Каждая метка имеет свой номер и определяет характерную точку, которую должна будет находить модель во время адаптации к новому изображению.  
  
   
*Пример разметки изображения лица из 68 точек, образующих форму AAM.*  
  
Процедура обучения AAM начинается с нормализации форм на размеченных изображениях с целью компенсации различий в масштабе, наклоне и смещении. Для этого используется так называемый обобщенный Прокрустов анализ.   
  
   
*Координаты точек формы лица до и после нормализации*  
  
Из всего множества нормированных точек затем выделяются главные компоненты с использованием метода PCA.   
  
   
*Модель формы AAM состоит из триангуляционной решетки s0 и линейной комбинации смещений si относительно s0*  
  
Далее из пикселей внутри треугольников, образуемых точками формы, формируется матрица, такая что, каждый ее столбец содержит значения пикселей соответствующей текстуры. Стоит отметить, что используемые для обучения текстуры могут быть как одноканальными (градации серого), так и многоканальными (например, пространство цветов RGB или другое). В случае многоканальных текстур векторы пикселов формируются отдельно по каждому из каналов, а потом выполняется их конкатенация. После нахождения главных компонент матрицы текстур модель AAM считается обученной.   
  
  
  
Модель внешнего вида AAM состоит из базового вида A0, определенного пикселями внутри базовой решетки s0 и линейной комбинации смещений Ai относительно A0  
  
  
  
Пример конкретизации AAM. Вектор параметров формы  
p=(p\_1,p\_2,〖…,p〗\_m )^T=(-54,10,-9.1,…)^T используется для синтеза модели формы s, а вектор параметров λ=(λ\_1,λ\_2,〖…,λ〗\_m )^T=(3559,351,-256,…)^Tдля синтеза внешнего вида модели. Итоговая модель лица 〖M(W(x;p))〗^ получается как комбинация двух моделей – формы и внешнего вида.  
  
Подгонка модели под конкретное изображение лица выполняется в процессе решения оптимизационной задачи, суть которой сводится к минимизации функционала  
  
https://habrastorage.org/files/ac4/33b/41e/ac433b41e1bd4ad3af3700fd967e662f.png  
  
методом градиентного спуска. Найденные при этом параметры модели и будут отражать положение модели на конкретном изображении.   
  
  
   
*Пример подгонки модели на конкретное изображение за 20 итераций процедуры градиентного спуска.*  
  
С помощью AAM можно моделировать изображения объектов, подверженных как жесткой, так и нежесткой деформации. ААМ состоит из набора параметров, часть которых представляют форму лица, остальные задают его текстуру. Под деформации обычно понимают геометрическое преобразование в виде композиции переноса, поворота и масштабирования. При решении задачи локализации лица на изображении выполняется поиск параметров (расположение, форма, текстура) ААМ, которые представляют синтезируемое изображение, наиболее близкое к наблюдаемому. По степени близости AAM подгоняемому изображению принимается решение – есть лицо или нет.   
  
***Active Shape Models (ASM)***  
  
Суть метода ASM [16,19,20] заключается в учете статистических связей между расположением антропометрических точек. На имеющейся выборке изображений лиц, снятых в анфас. На изображении эксперт размечает расположение антропометрических точек. На каждом изображении точки пронумерованы в одинаковом порядке.   
  
  
  
*Пример представления формы лица с использованием 68 точек*  
  
Для того чтобы привести координаты на всех изображениях к единой системе обычно выполняется т.н. обобщенный прокрустов анализ, в результате которого все точки приводятся к одному масштабу и центрируются. Далее для всего набора образов вычисляется средняя форма и матрица ковариации. На основе матрицы ковариации вычисляются собственные вектора, которые затем сортируются в порядке убывания соответствующих им собственных значений. Модель ASM определяется матрицей Φ и вектором средней формы s ̅.   
Тогда любая форма может быть описана с помощью модели и параметров:  
  
https://habrastorage.org/files/8a1/b92/6a1/8a1b926a18a842378dbd46ef8fd3d04c.png  
  
Локализации ASM модели на новом, не входящем в обучающую выборку изображении осуществляется в процессе решения оптимизационной задачи.   
  
   
а) б) в) г)  
*Иллюстрация процесса локализации модели ASM на конкретном изображении: а) начальное положение б) после 5 итераций в) после 10 итераций г) модель сошлась*  
  
Однако все же главной целью AAM и ASM является не распознавание лиц, а точная локализация лица и антропометрических точек на изображении для дальнейшей обработки.   
  
Практически во всех алгоритмах обязательным этапом, предваряющим классификацию, является выравнивание, под которым понимается выравнивание изображения лица во фронтальное положение относительно камеры или приведение совокупности лиц (например, в обучающей выборке для обучения классификатора) к единой системе координат. Для реализации этого этапа необходима локализация на изображении характерных для всех лиц антропометрических точек – чаще всего это центры зрачков или уголки глаз. Разные исследователи выделяют разные группы таких точек. В целях сокращения вычислительных затрат для систем реального времени разработчики выделяют не более 10 таких точек [1].  
  
Модели AAM и ASM как раз и предназначены для того чтобы точно локализовать эти антропометрические точки на изображении лица.