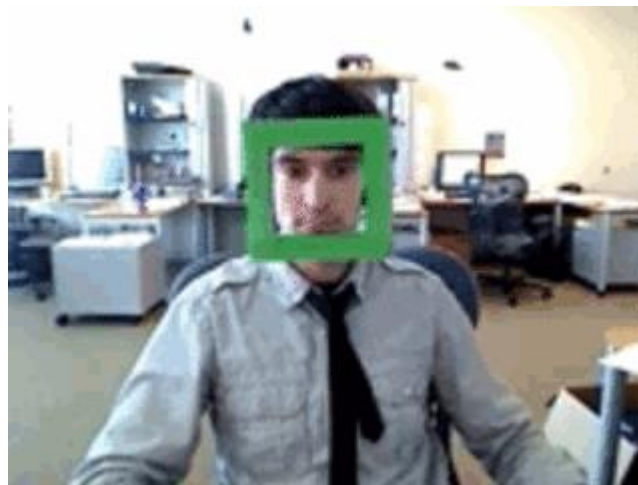


Определение лиц



Метод Виолы-Джонса (Viola-Jones)

Метод Виолы—Джонса (Viola–Jones object detection) — алгоритм, позволяющий обнаруживать объекты на изображениях в реальном времени.

- Разработан в 2001 году Полом Виолой и Майклом Джонсом.
- Используется в бюджетных цифровых фотоаппаратах.
- Является одним из лучших по соотношению показателей эффективности распознавания/скорость работы.
- Алгоритм находит лица с высокой точностью и низким количеством ложных срабатываний.
- Алгоритм хорошо работает и распознает черты лица под небольшим углом, примерно до 30 градусов.
- В составе библиотеки компьютерного зрения OpenCV реализован в функции `cvHaarDetectObjects()`.

Метод Виолы-Джонса. Основные принципы.

Основные принципы, на которых основан метод, таковы:

1. используются **изображения в интегральном представлении**, что позволяет вычислять быстро необходимые объекты;
2. используются **признаки Хаара**, с помощью которых происходит поиск нужного объекта;
3. используется **бустинг** (от англ. boost – улучшение, усиление) для выбора наиболее подходящих признаков для искомого объекта на данной части изображения;
4. все признаки поступают на вход **классификатора**, который даёт результат «верно» либо «ложь»;
5. используются **каскады признаков** для быстрого отбрасывания окон, где не найдено лицо.

1. Интегральное представление изображения

Интегральное представление изображения — это матрица, размерность которой совпадает с размерностью исходного изображения.

Каждый элемент интегрального изображения содержит в себе сумму значений всех пикселей левее и выше данного пикселя (x, y) .

Original

5	2	3	4	1
1	5	4	2	3
2	2	1	3	4
3	5	6	4	5
4	1	3	2	6

Integral

5	7	10	14	15
6	13	20	26	30
8	17	25	34	42
11	25	39	52	65
15	30	47	62	81

$$5 + 2 + 3 + 1 + 5 + 4 = 20$$

Original

5	2	3	4	1
1	5	4	2	3
2	2	1	3	4
3	5	6	4	5
4	1	3	2	6

Integral

5	7	10	14	15
6	13	20	26	30
8	17	25	34	42
11	25	39	52	65
15	30	47	62	81

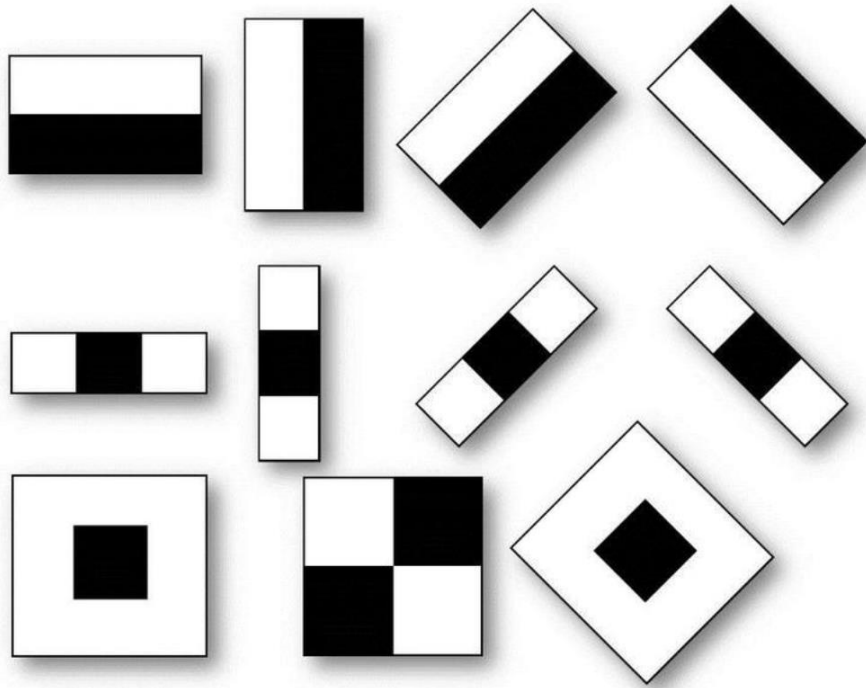
$$5 + 4 + 2 + 2 + 1 + 3 = 17$$

$$34 - 14 - 8 + 5 = 17$$

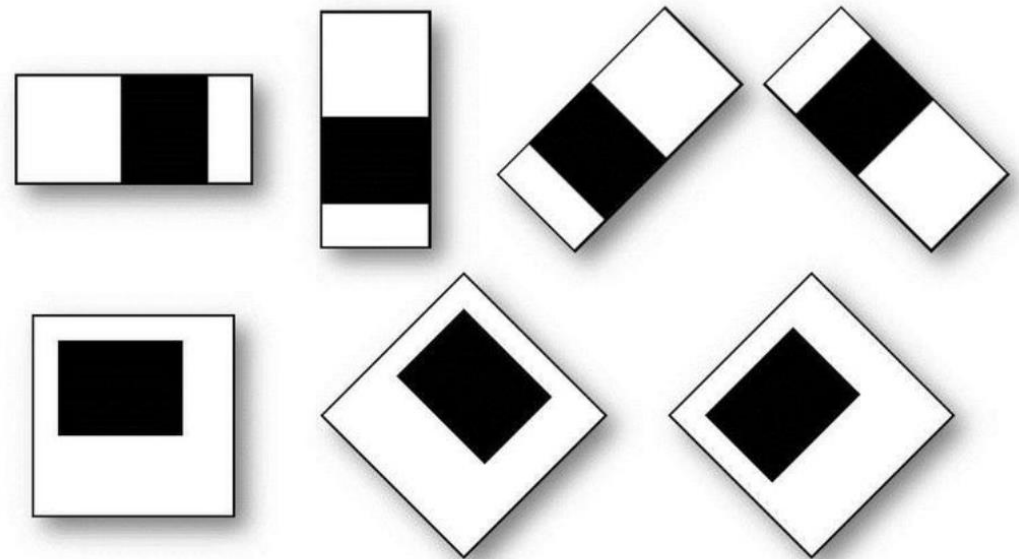
Интегральное представление позволяет **быстро вычислить сумму пикселей произвольного прямоугольника** с помощью четырех ссылок на массив.

2. Признаки Хаара

В методе Виолы-Джонса основу составляют примитивы Хаара, представляющие собой разбивку заданной прямоугольной области на наборы разнотипных прямоугольных подобластей:

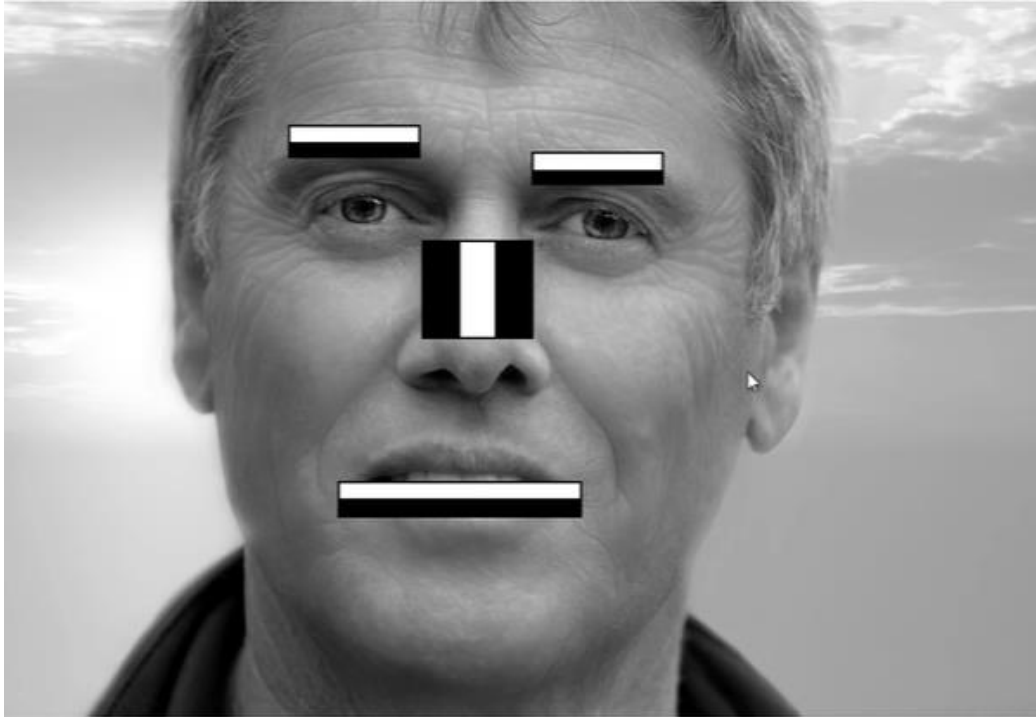


Стандартный метод



Расширенный метод

Вычисление признаков Хаара



Вычисляемым значением такого признака будет

$$F = X - Y$$

где X – сумма значений яркостей точек закрываемых светлой частью признака, а Y – сумма значений яркостей точек закрываемых темной частью признака.

Для их вычисления используется понятие интегрального изображения, рассмотренное выше.

3. Обучение классификатора в методе Виолы-Джонса

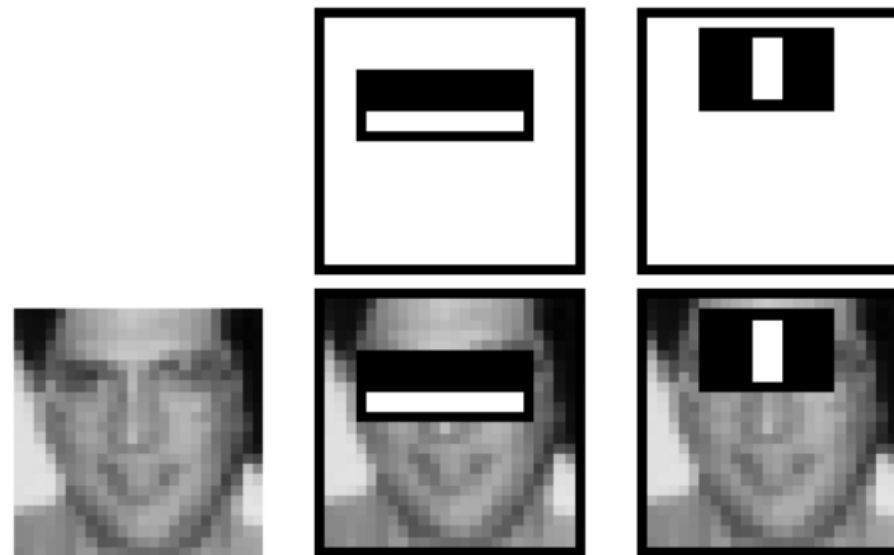
Бустинг означает дословно **«усиление» «слабых» моделей** – это процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов.

*Предложена Робертом Шапиром (Scharire) в конце 90-х годов, когда надо было найти решение вопроса о том, чтобы имея множество плохих (незначительно отличающихся от случайных) алгоритмов обучения, получить один хороший. В основе такой идеи лежит построение *цепочки (ансамбля) классификаторов*, который называется **каскадом**, каждый из которых (кроме первого) *обучается на ошибках предыдущего*.*

AdaBoost (*adaptive boosting*) (1999) может использовать произвольное число классификаторов и производить обучение на одном наборе примеров, поочередно применяя их на различных шагах.

4, 5. Каскад признаков

Из двух признаков Хаара строится первый каскад системы по распознаванию лиц, который имеет вполне осмысленную интерпретацию.



Для определения принадлежности к классу в каждом каскаде, находится сумма значений слабых классификаторов этого каскада. Каждый слабый классификатор выдает два значения в зависимости от того больше или меньше заданного порога значение признака, принадлежащего этому классификатору. В конце сумма значений слабых классификаторов сравнивается с порогом каскада и выносятся решения найден объект или нет данным каскадом.

Алгоритм сканирования окна с признаками



[Демо](#)

Обучение классификатора в методе Виолы-Джонса

В процессе поиска вычислять все признаки затратно. Следовательно, **классификатор** должен реагировать *только на определенное, нужное подмножество всех признаков*. Значит классификатор надо обучить нахождению лиц по данному определенному подмножеству.

Классификатор(classifier) — в задачах классификации это аппроксимирующая функция, выносящая решение, к какому именно классу данный объект принадлежит.

В случае алгоритма Виолы-Джонса для идентификации и распознавания лица классификация является *двухклассовой*.

Обучение OpenCV каскада Хаара

Весь процесс обучения выборки не требует навыков программирования. Для этого имеются уже готовые консольные программы, присутствующие в основной сборке OpenCV.

- **Фотографии предмета в реальной среде обитания.** Чем более похожа выборка будет на то, что мы будем распознавать, тем лучше будут результаты. Если обучать распознаватель лица по фотографиям людей из студии, то на улице уровень распознавания будет ниже, чем в студии. На это влияют как тени, одежда, так и выражение лица.
- **Выборка отрицательных фотографий, на которых нет объекта распознавания.** Фотографии должны быть сделаны в той же среде где будет распознавание. Если выборка контрпримеров будет сделана по фотографиям на северном полюсе, а распознавать будете в тропических джунглях, то ничего не заработает.

Обучение OpenCV каскада Хаара

Для того, чтобы начать обучение, нам нужно иметь **2 папки с примерами**. «Good» — папка с позитивными изображениями, «Bad» — с отрицательными.

Для каждой папки нужно иметь текстовый файл, в котором описаны используемые изображения. Назовём их «Good.dat» и «Bad.dat». ВАЖНО! Этот файл должен лежать на том же уровне файловой системы, на котором лежит папка.

```
\Good
    \1. bmp
    \2. bmp
    \... bmp
    \N. bmp

\Bad
    \1. bmp
    \2. bmp
    \... bmp
    \N. bmp

Good.dat
Bad.dat
```

Обучение OpenCV каскада Хаара

Файлы описания для отрицательных и положительных объектов имеют разную структуру.

```
Bad\1. bmp  
Bad\2. bmp  
Bad\.... bmp  
Bad\N. bmp
```

```
Good \0.bmp 1 0 0 414 148  
Good \1.bmp 1 0 0 568 164  
Good \....bmp 1 0 0 440 144  
Good \N.bmp 1 0 0 590 182
```

Само обучение происходит в два этапа:

- создается коллекция [*opencv_createsamples.exe*](#)
- расчет итогового каскада. Работает долго. Обучение каскада на 500-1000 объектов займёт почти целый день [*opencv_traincascade.exe*](#)

Branch: master ▼

opencv / data / haarcascades /

Create new file

Find file










History



StevenPuttemans fixing models to resolve XML violation issue

Latest commit 2ddbcc3 on 13 Jun

..

 haarcascade_eye.xml	some attempts to tune the performance	4 years ago
 haarcascade_eye_tree_eyeglasses.xml	some attempts to tune the performance	4 years ago
 haarcascade_frontalcatface.xml	fixing models to resolve XML violation issue	6 months ago
 haarcascade_frontalcatface_extende...	fixing models to resolve XML violation issue	6 months ago
 haarcascade_frontalface_alt.xml	some attempts to tune the performance	4 years ago
 haarcascade_frontalface_alt2.xml	some attempts to tune the performance	4 years ago
 haarcascade_frontalface_alt_tree.xml	some attempts to tune the performance	4 years ago
 haarcascade_frontalface_default.xml	some attempts to tune the performance	4 years ago
 haarcascade_fullbody.xml	fixing wrong model sizes	3 years ago

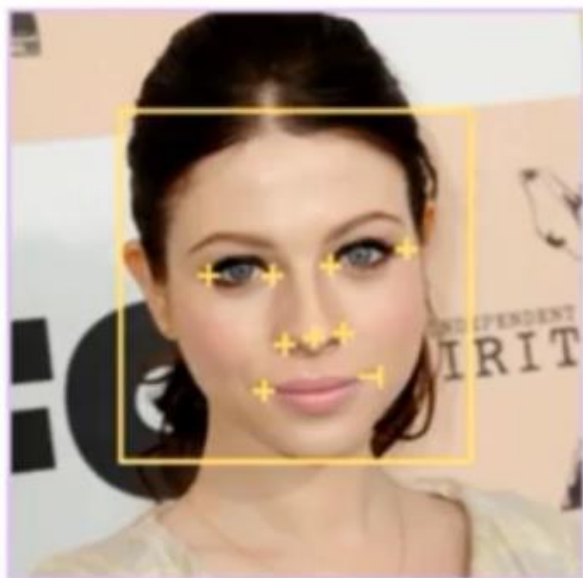


Распознавание лиц

Анализ существующих подходов

Этапы распознавания лица

1. Детектирование лица



2. Выравнивание лица



3. Извлечение дескриптора

$$f(\text{image}) = d_i$$

4. Вычисление схожести

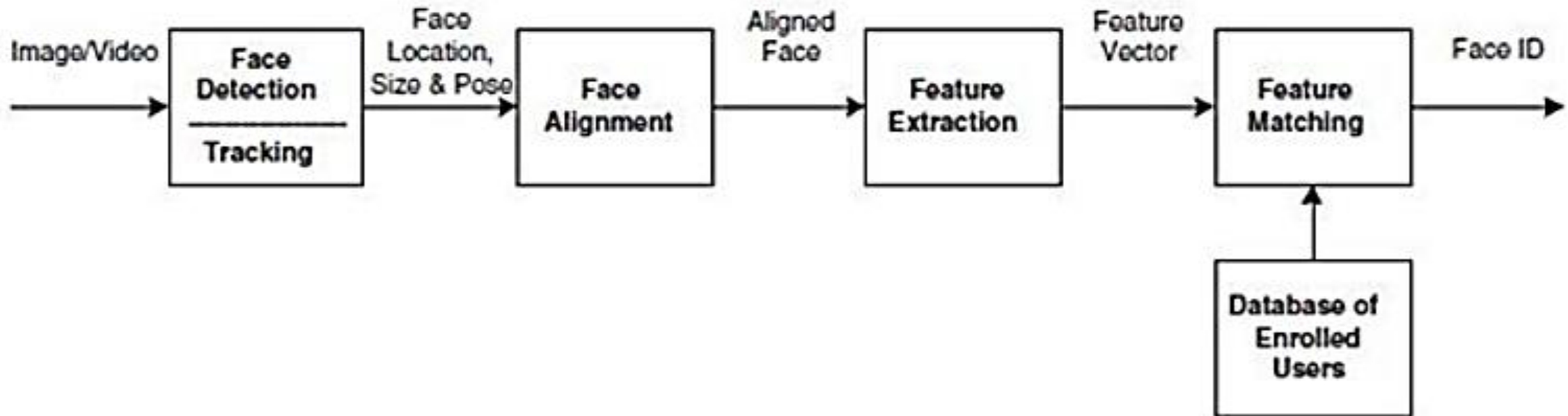
$$S_{ij} = k(d_i, d_j)$$

$$S(\text{Obama}, \text{Trump}) = 0$$

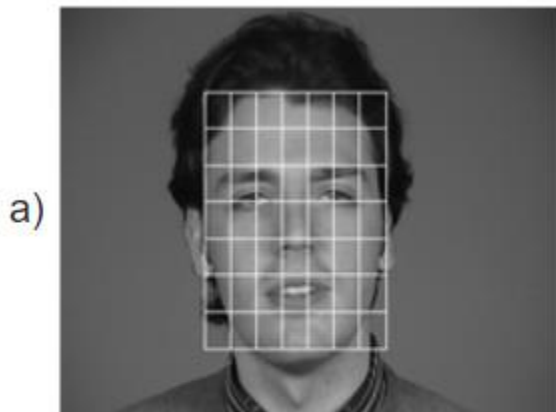
$$S(\text{Kunis}, \text{Kanye}) = 6$$

Распознавание лиц

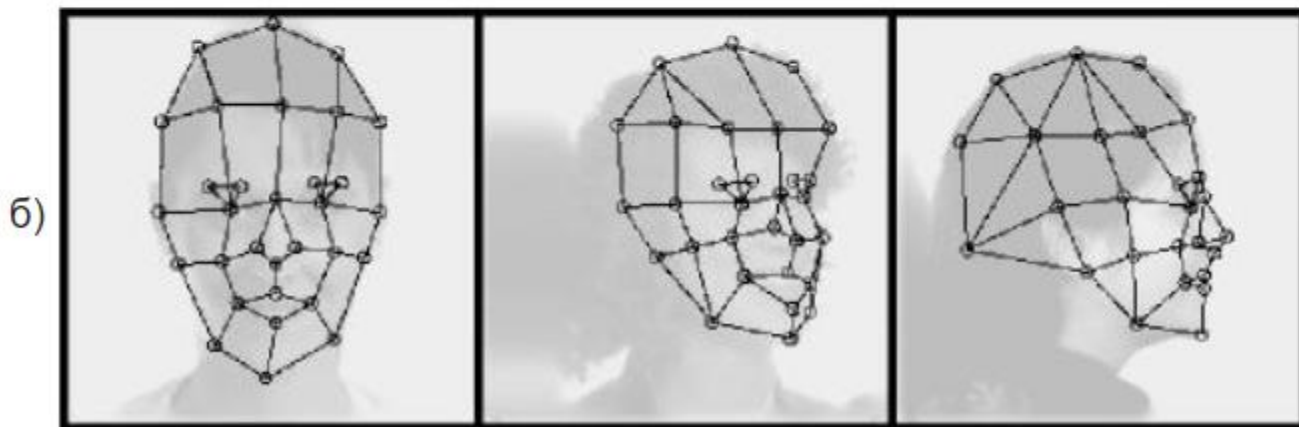
Несмотря на большое разнообразие представленных алгоритмов, можно выделить общую структуру процесса распознавания лиц:



1. Метод гибкого сравнения на графах (Elastic graph matching)

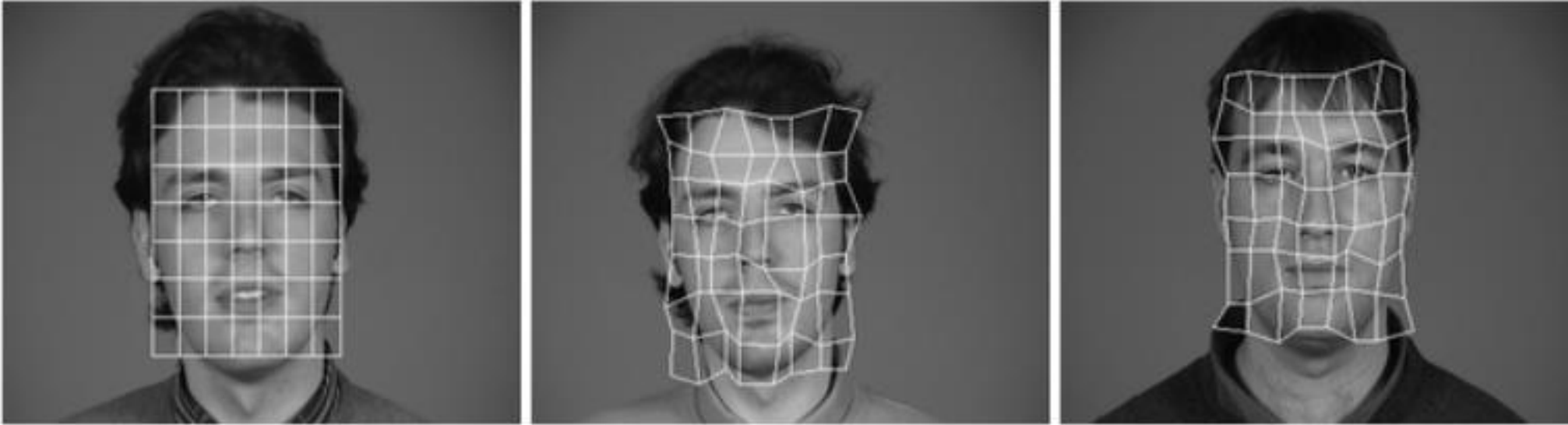


Суть метода сводится к эластичному сопоставлению графов, описывающих изображения лиц.



Графы могут представлять собой как прямоугольную решетку (а), так и структуру, образованную характерными (антропометрическими) точками лица (б).

1. Метод гибкого сравнения на графах



Пример деформации графа в виде регулярной решетки

На этапе распознавания один из графов – эталонный – остается неизменным, в то время как другой деформируется с целью наилучшей подгонки к первому.

Различие 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.

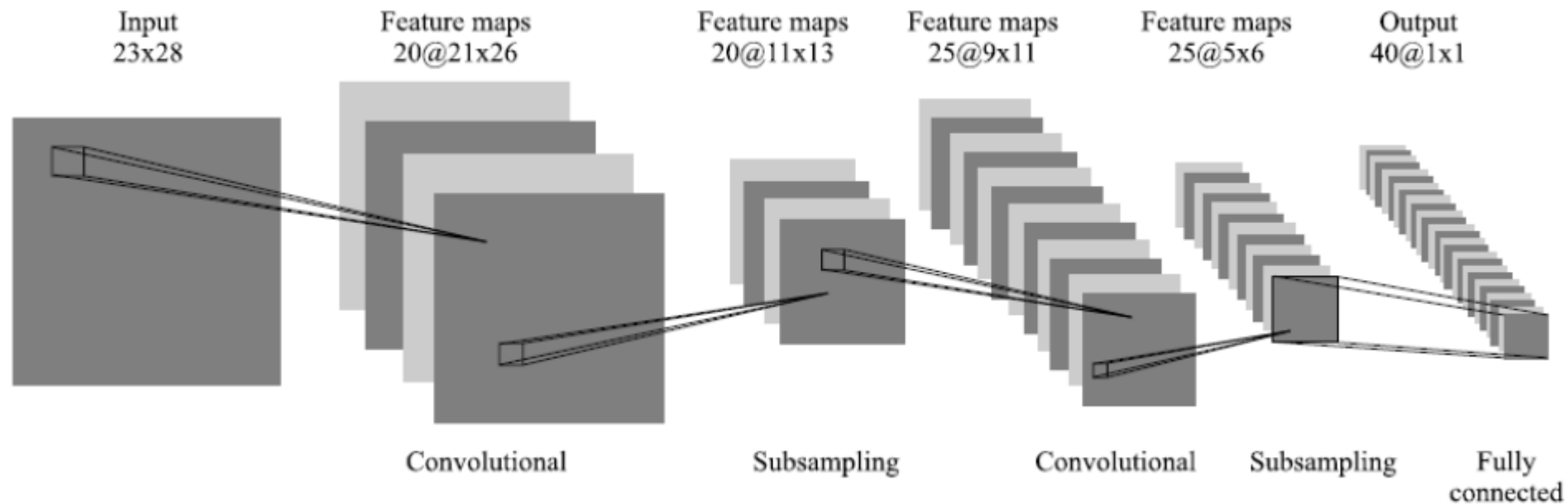
1. Метод гибкого сравнения на графах

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.

Недостатки:

- высокая вычислительная сложность процедуры распознавания
- низкая технологичность при запоминании новых эталонов
- линейная зависимость времени работы от размера базы данных лиц.

2. Нейронные сети

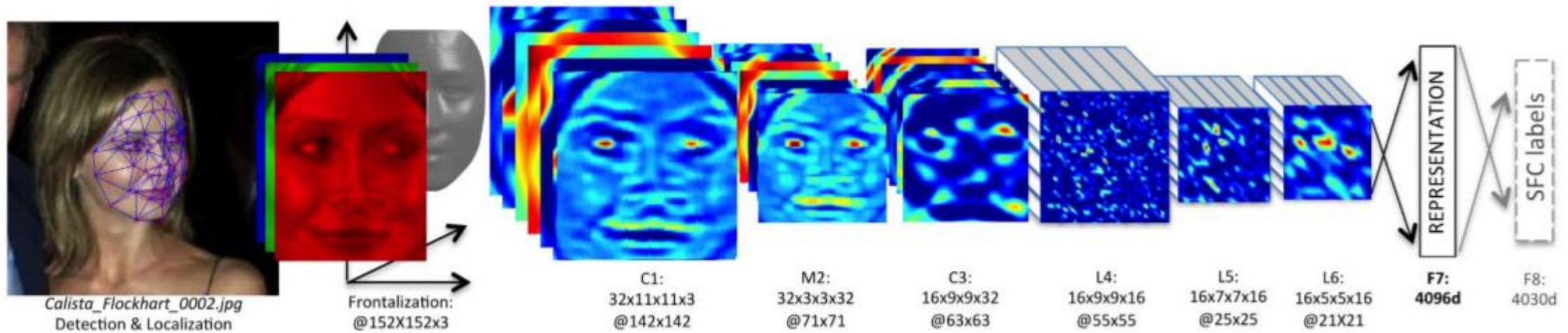


Наилучшие результаты в области распознавания лиц (по результатам анализа публикаций) показала Convolutional Neural Network или сверточная нейронная сеть (СНС).

СНС обеспечивает частичную устойчивость к изменениям масштаба, смещениям, поворотам, смене ракурса и прочим искажениям.

Нейронные сети. DeepFace

Тестирование СНС показало **96% точность** распознавания.



Свое развитие СНС получили в разработке DeepFace, которую приобрел Facebook для распознавания лиц пользователей своей соцсети.

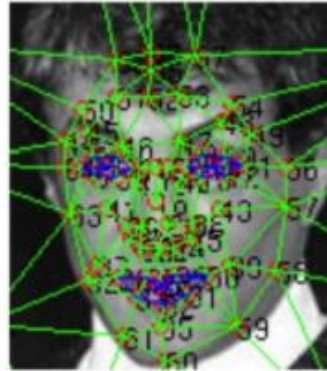
DeerFace. Выравнивание лица



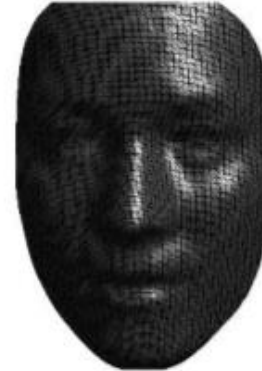
(a)



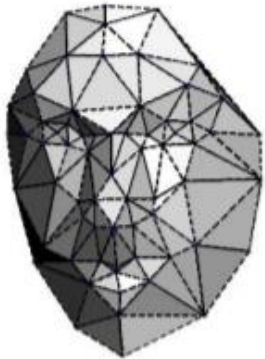
(b)



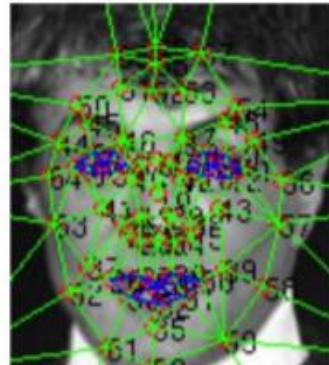
(c)



(d)



(e)



(f)



(g)



(h)

Detect 67 “fiducial points” (using SVR).

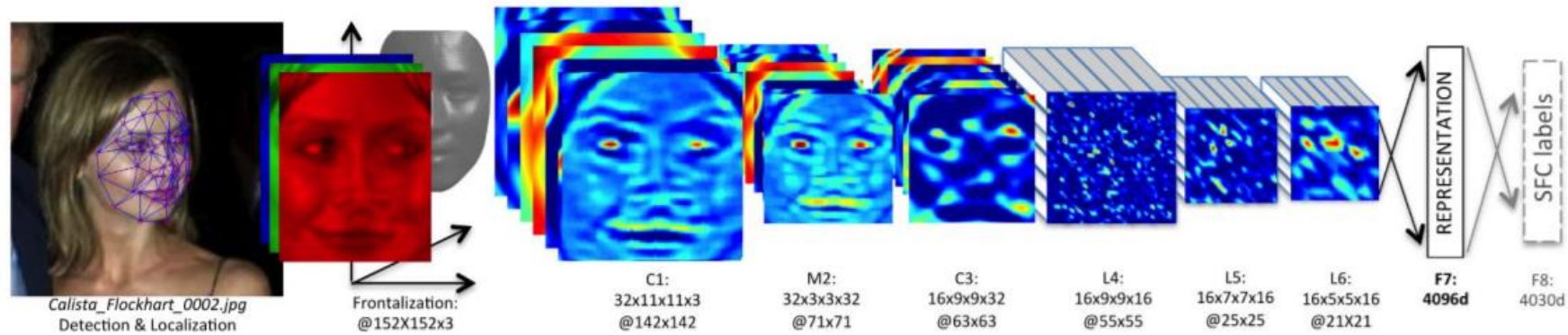
Compare them with a fixed 3D model of a generic face.

Reconstruct the 3D model of the detected face.

Generate the 3D-aligned version of the crop.

The final image is given to the DNN.

Нейронные сети.



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.

Метод главных компонент или principal component analysis (PCA).

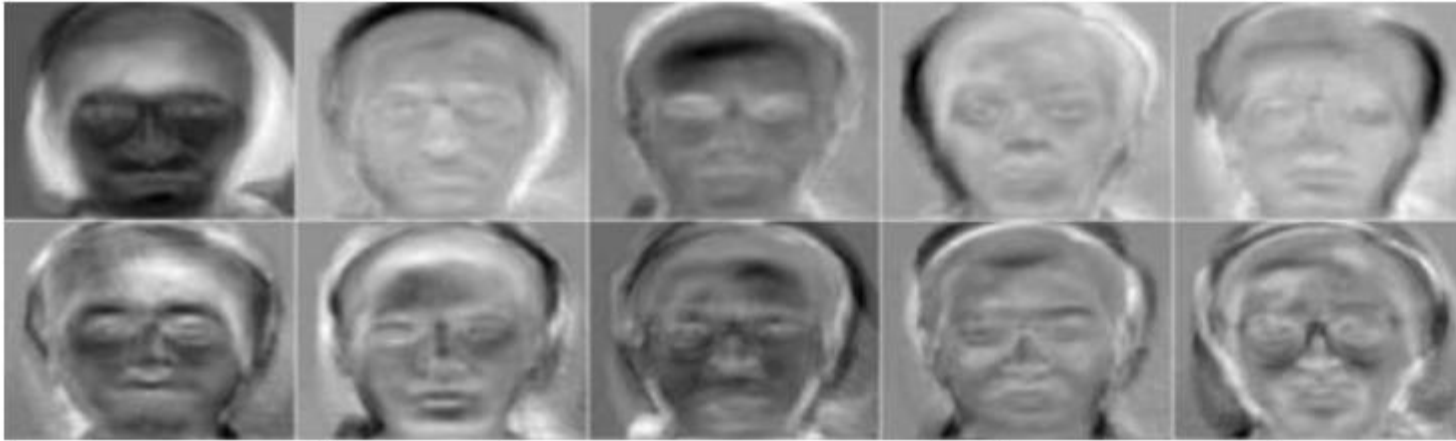
В задаче распознавания лиц его применяют главным образом для представления изображения лица вектором малой размерности (главных компонент), который сравнивается затем с эталонными векторами, заложенными в базу данных.

Главной целью метода главных компонент является значительное уменьшение размерности пространства признаков таким образом, чтобы оно как можно лучше описывало «типичные» образы, принадлежащие множеству лиц.

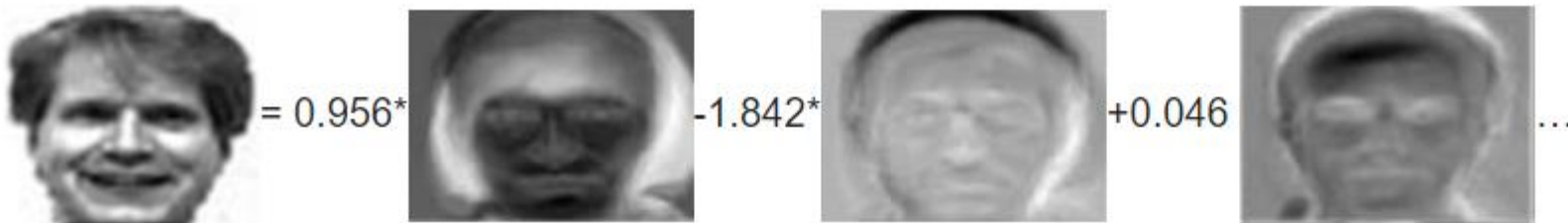
Полученный один раз на обучающей выборке изображений лиц набор собственных векторов используется для кодирования всех остальных изображений лиц, которые представляются взвешенной комбинацией этих собственных векторов.

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.

Метод главных компонент или principal component analysis (PCA).



Пример первых десяти собственных векторов (собственных лиц), полученных на обучаемом наборе лиц



Пример построения (синтеза) человеческого лица с помощью комбинации собственных лиц и главных компонент

Надежность систем распознавания

Надежность характеризуется:

- ✓ False Rejection Rate (FRR), ложный отказ в доступе (ошибка первого рода), т.е. доступ запрещён зарегистрированному в системе человеку;
- ✓ False Acceptance Rate (FAR) ложный пропуск (ошибка первого рода), т.е. система предоставляет доступ незарегистрированному пользователю.

Необходимо учитывать взаимосвязь этих показателей: искусственно снижая уровень «требуемости» системы (FAR), мы, как правило, уменьшаем процент ошибок FRR, и наоборот.

На сегодняшний день все биометрические технологии являются вероятностными, ни одна из них не способна гарантировать полное отсутствие ошибок FAR/FRR.

