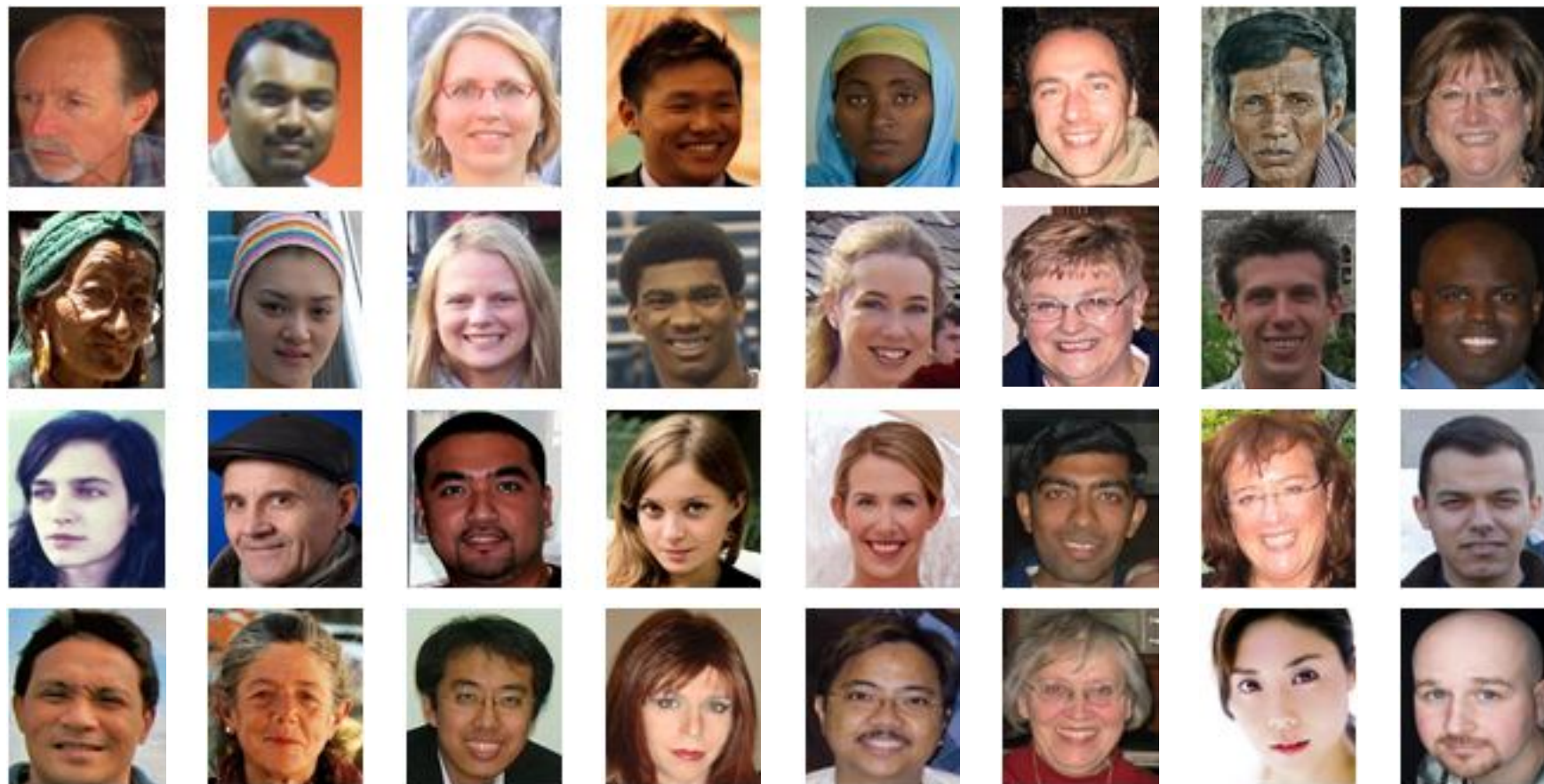




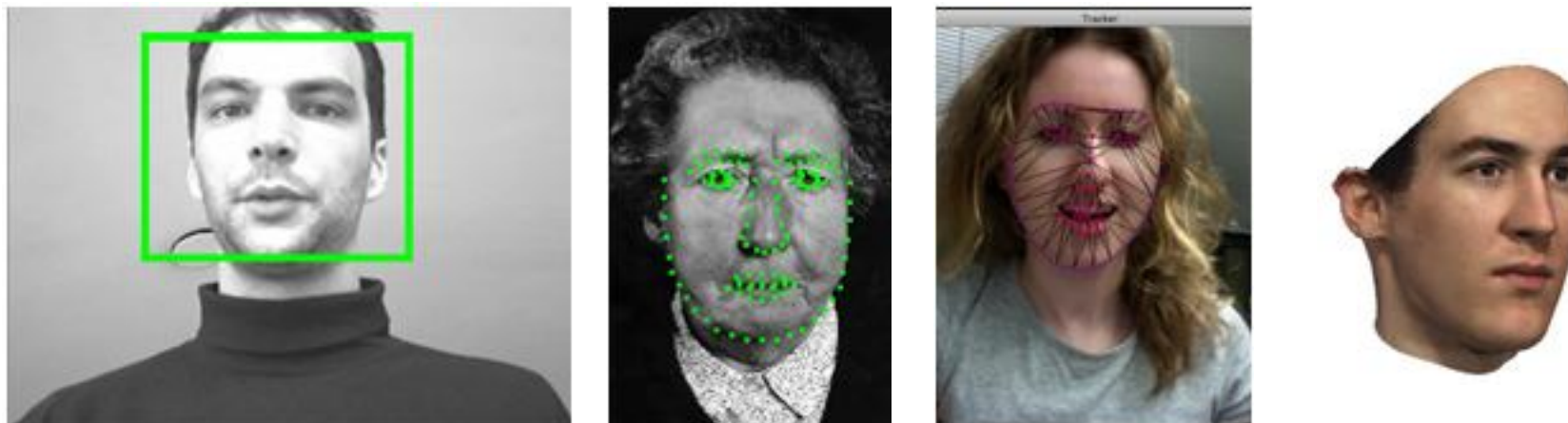
Распознавание и анализ лиц



Антон Конушин



Что предполагаем



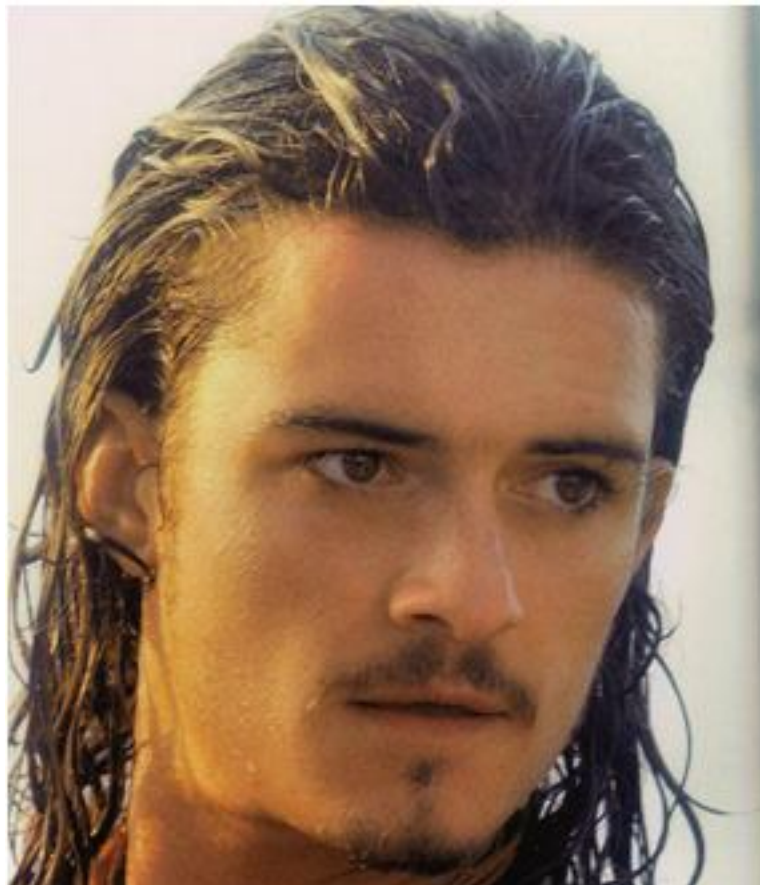
- Умеем:
 - Выделять лица в изображении
- Знаем, что можно (эти задачи рассматриваются в курсе «Доп. главы компьютерного зрения»)
 - Выделять контрольные (антропометрические) точки на изображении лица, такие как уголки глаз, и т.д.
 - Строить 2D сеточные модели лица по изображению (триангулирование сетки)



Что будем рассматривать

- Основные задачи и тестовые базы
- Распознавание лиц
- Распознавание атрибутов по лицу
- Фокусы с лицами

Верификация



На обоих изображениях один и тот же человек, или нет?

Базовая задача распознавания лиц, которую и человеку проще всего решать

Идентификация



- Есть фиксированный список людей (база) и тестовое изображение
- Нужно определить, кто из списка изображен на тестовой фотографии?

«Watch List»



- Есть список «подозрительных людей»
- Необходимо определить, входит ли человек в этот список по его фотографии
- Самая сложная постановка задачи

Аттрибуты



Мужчина

Азиат

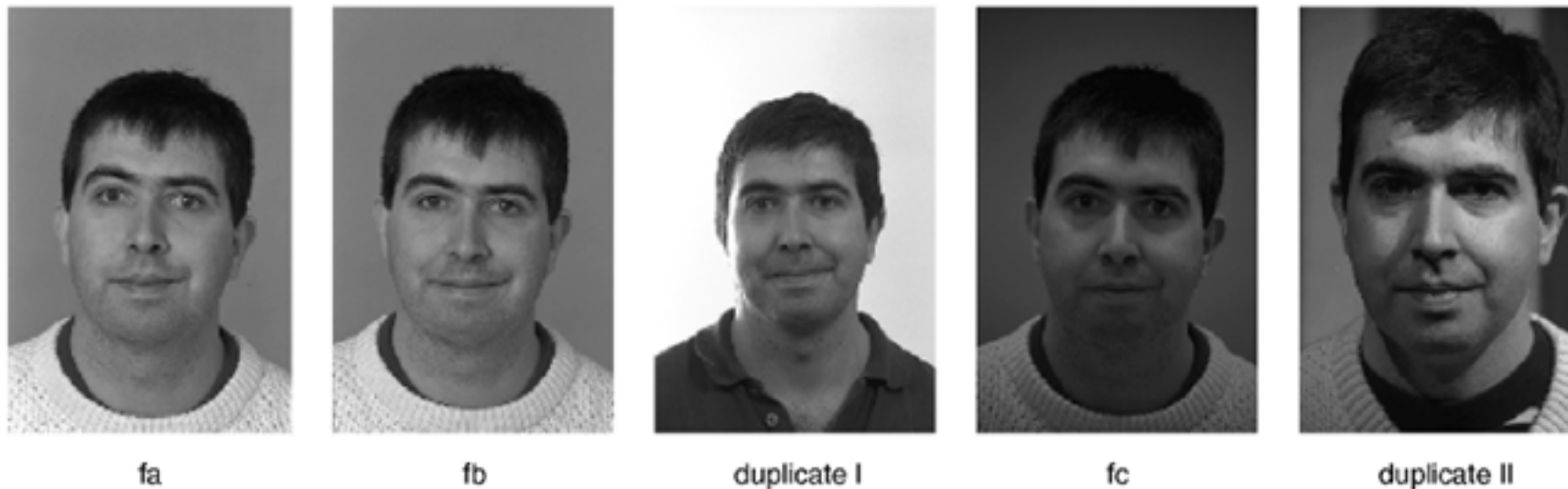
Борода

Улыбается

- Что мы можем сказать про этого человека?
- Атрибуты – «типичные» характеристики объекта, которые мы используем для его описания
- Для человека - пол, возраст, раса, борода, усы, улыбка, очки и т.д.



FERET (1995-2000)



- 14000 изображений, 1200 человек
- 2 фронтальных фотографии с разным выражением лица (fa, fb)
- Иногда одна фотография другой камерой и освещением
- Несколько дубликатов (фотографий того же человека через 1-2 года после первой съёмки)

P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "[The FERET evaluation methodology for face recognition algorithms](#)", PAMI 2000.

FRGC



(a)



- Face Recognition Grand Challenge (для Face Recognition Vendor Test)
- Идея: исследовать возможность повышения качества распознавания за счёт улучшенных данных разной природы (контролируемые, неконтролируемые, 3д), высокое разрешение
- 220 человек, 12К изображений в обучающих данных
- 465 человек в валидационной выборке (год спустя)

P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. **Overview of the Face Recognition Grand Challenge**. In CVPR, 2005.



Labeled Faces in the Wild (LFW)

Labeled Faces in the Wild <http://vis-www.cs.umass.edu/lfw>

Database by name, non-singleton

[A][B][C][D][E][F][G][H][I][J][K][L][M][N][O][P][Q][R][S][T][U][V][W][X][Y][Z]

Menu

- LFW Home
- UMass Vision

 Habib Rizieq (5)	 Hal Gehman (5)	 Hal Sutton (2)	 Halle Berry (16)	 Hamid Karzai (22)
 Hamzah Haz (2)	 Hanan Ashrawi (2)	 Hannah Stockbauer (2)	 Hans Blix (39)	 Hans Eichel (3)

Сбор коллекции изображений из интернета.

Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller.

Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. *University of Massachusetts, Amherst, Technical Report 07-49, 2007.*

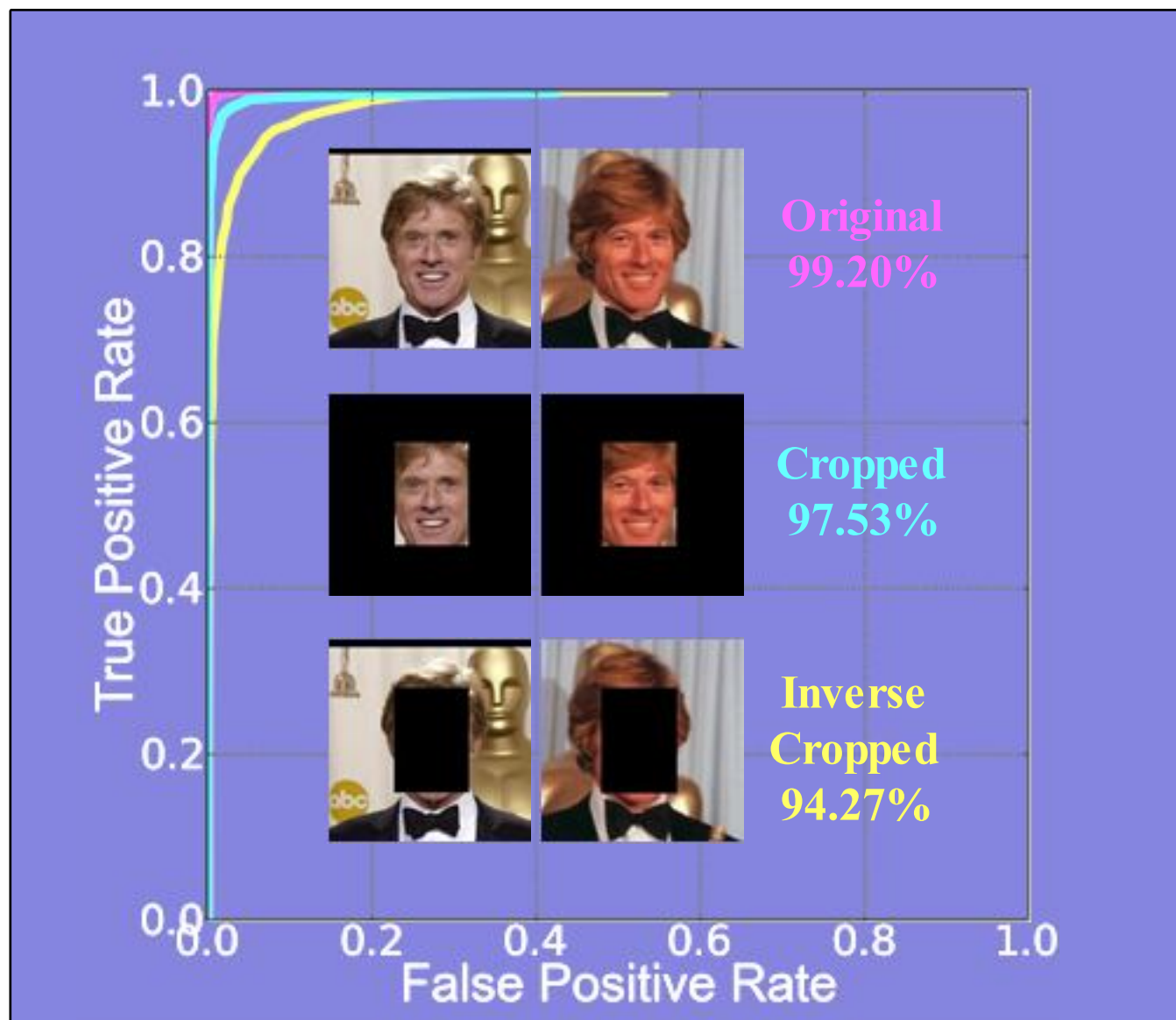


Labeled Faces in the Wild

- 5749 человек, 12K изображений, 1680 человек по 2 и более фотографии, остальные – по одной
- Фотографии разрешения 250*250, JPEG
- Неконтролируемые условия, очень разные фоны, позы, разное время съёмки
- “In the wild” сейчас очень популярный термин, обозначающий работу с реальными изображениями без ограничений
- На этой коллекции до текущего времени проводились сравнения алгоритмов распознавания лиц



Верификация человеком на LFW





Сравнение современных баз

Dataset	#Subjects	#Images
LFW	5,749	13,233
CelebFaces [1]	10,177	202,599
SFC [2]	4,030	4,400,000
CASIA-WebFace [3]	10,575	494,414

- Разметка за счёт фрилансеров
- Значительный рост размеров коллекций

1. Y. Sun, X. Wang, and X. Tang. “Deep learning face representation by joint identification-verification”. arXiv preprint arXiv:1406.4773, 2014
2. Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. “Deepface: Closing the gap to human-level performance in face verification”. In Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pages 1701–1708. IEEE, 2014
3. Dong Yi, Zhen Lei, Shengcai Liao and Stan Z. Li, “Learning Face Representation from Scratch”. arXiv preprint arXiv:1411.7923. 2014.



Нормализация изображения лица



Базовый вариант – по центрам глаз

- Выделяются центры глаз
- Вычисляется такое преобразование, чтобы центры глаз оказались на одной строке и с фиксированным расстоянием между ними



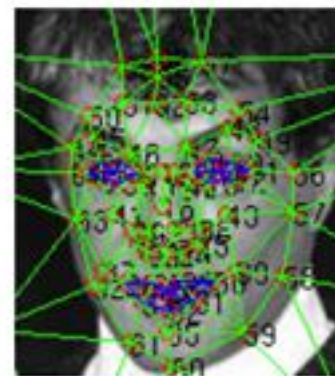
3D нормализация



(a)



(b)



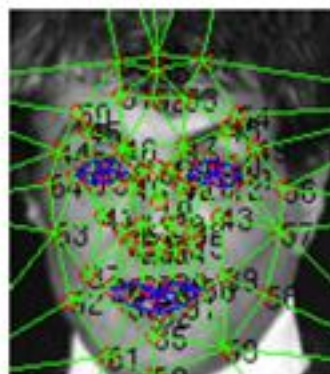
(c)



(d)



(e)



(f)



(g)



(h)

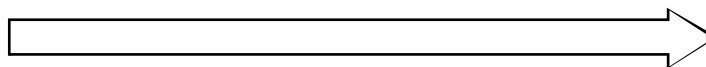
Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. **DeepFace: Closing the Gap to Human-Level Performance in Face Verification.** CVPR 2014.



Общая схема верификации

Изображения

Признаки



Верификация



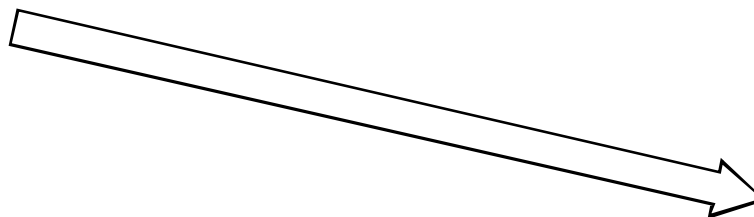
RGB

HOG

LBP

SIFT

Deep



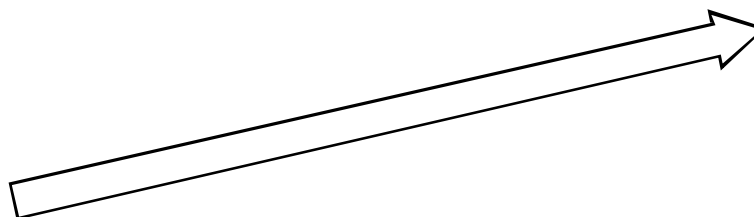
RGB

HOG

LBP

SIFT

Deep



Разные

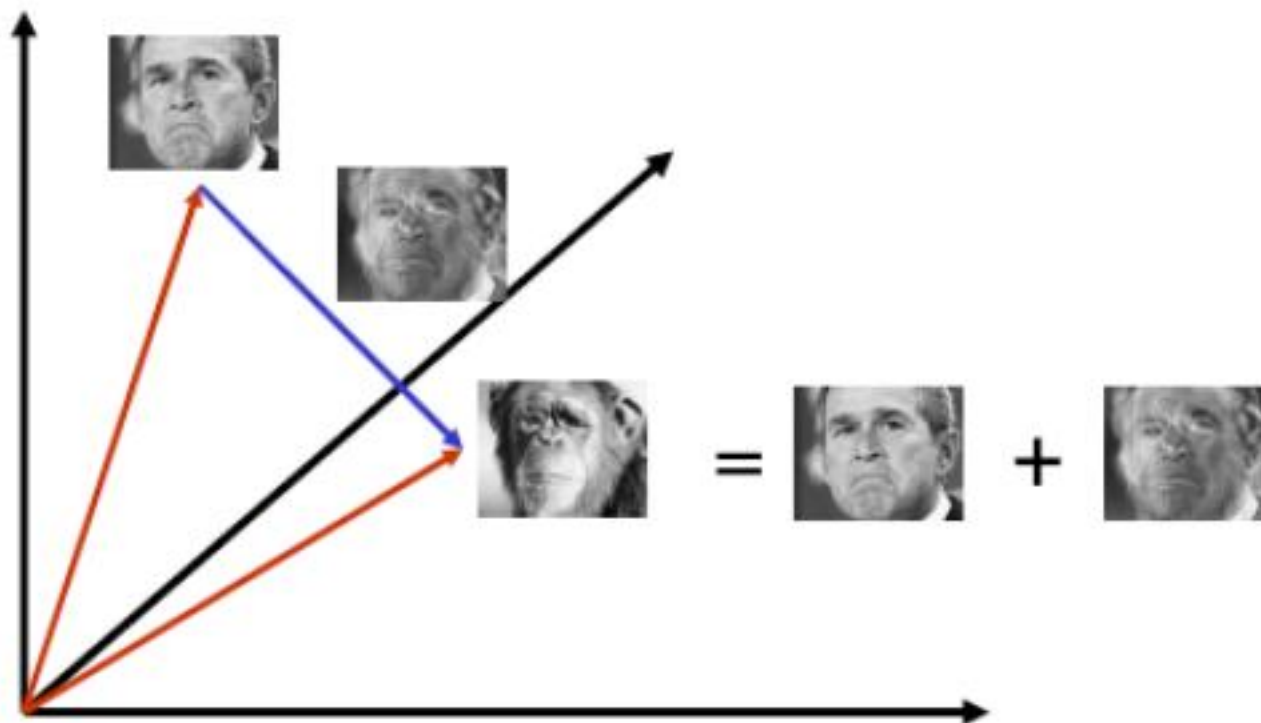


Что можно менять:

- Выбор признаков
 - RGB, HOG, биологические, LBP, обучаемые, нейросетевые
- Квантование, агрегация (pooling) признаков по изображению
- Построение дескриптора по признакам
- Сравнение дескрипторов
 - Метрика для сравнения дескрипторов
 - Обучение классификатора {разные, одинаковые}

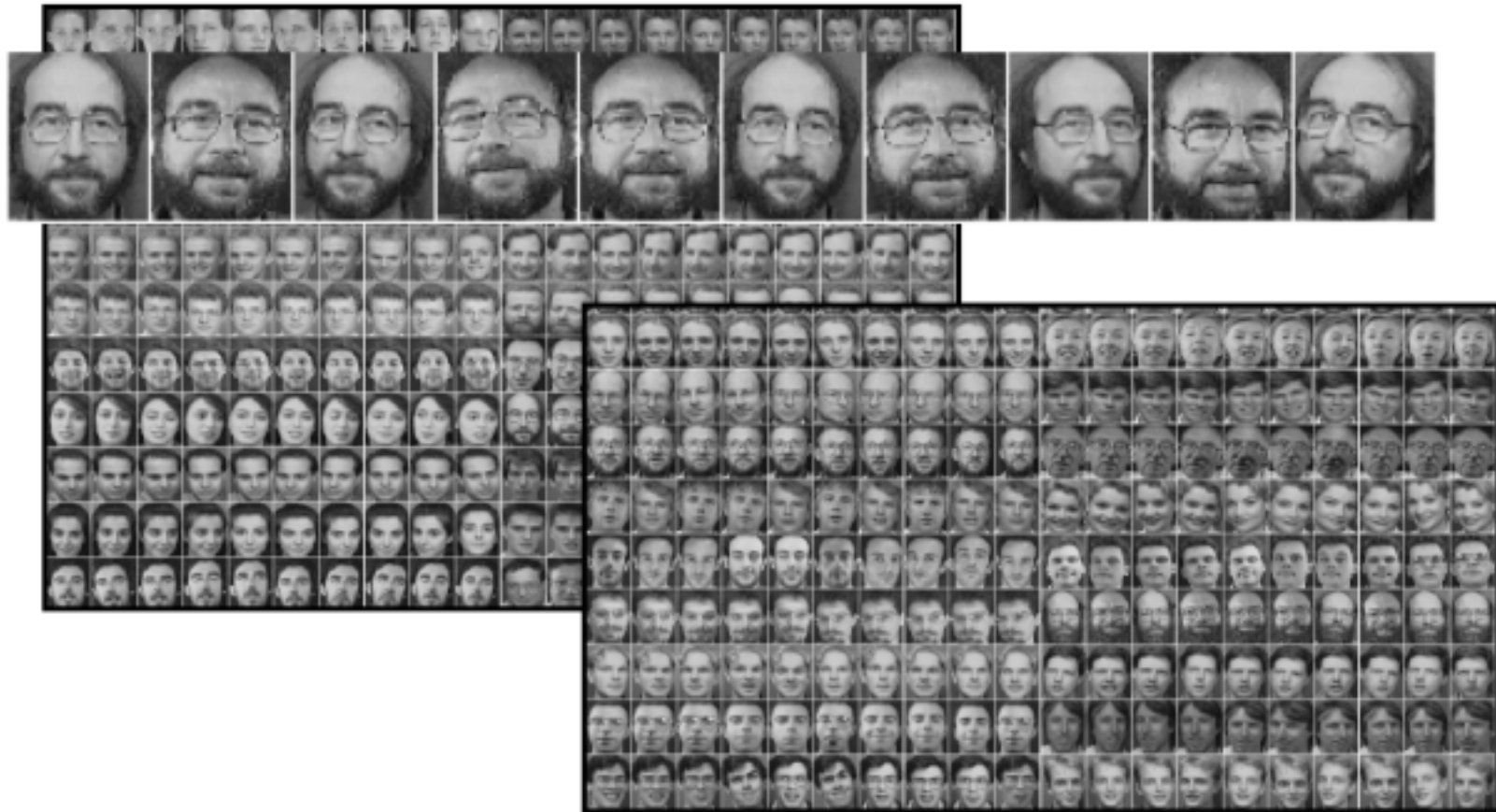


EigenFaces



M. Turk and A. Pentland ["Face recognition using eigenfaces"](#).
Proc. IEEE Conference on Computer Vision and Pattern Recognition. 1991 (более 4000 цитирований!)

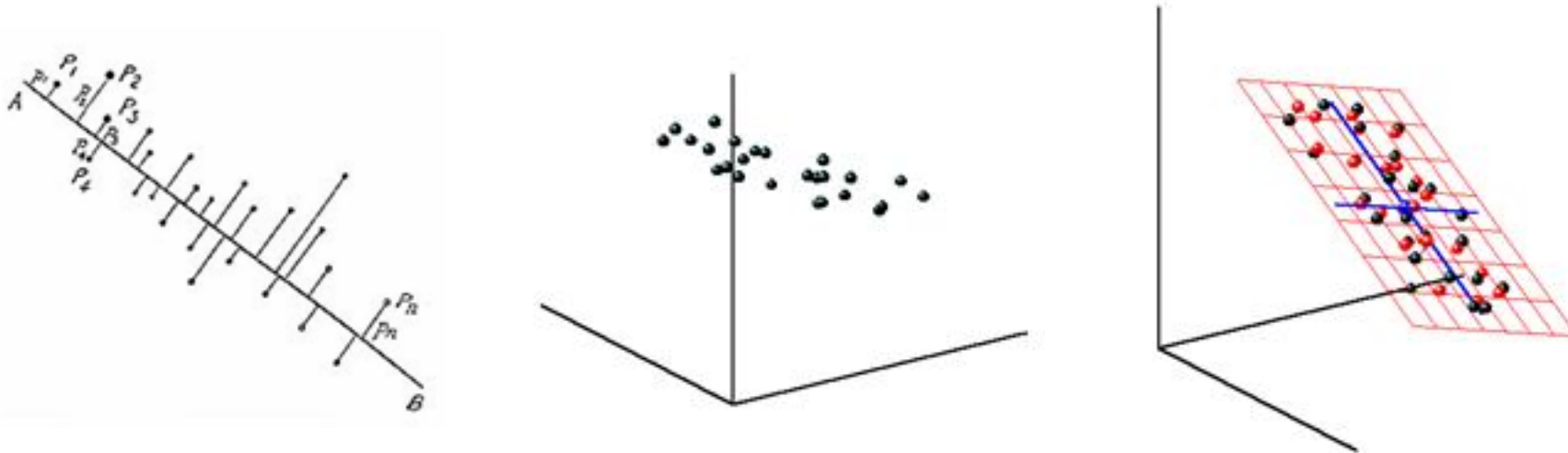
EigenFaces



Большая собственная выборка изображений лиц



Уменьшение размерности



- Dimension reduction – важная задача / группа методов
- Задача - аппроксимация данных векторами меньшей размерности
- В машинном обучении – уменьшение размерности вектор-признака
- Самый часто используемый метод – метод главных компонент (PCA – Principle Component Analysis), впервые описанный Пирсоном в 1901 году



Метод главных компонент

- Дана выборка $x_1, \dots, x_N$.
- Вычислим среднее $\mu = (1/N) \sum_{i=1}^N x_i$.

- Вычислим матрицу ковариации:

$$K = (1/N) \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T.$$

- Найдём собственные значения λ и собственные вектора e :

$$Ke = \lambda e.$$

- Упорядочим собственные значения по убыванию

$$\lambda_1 \geq \lambda_2 \geq \dots \lambda_N.$$

- Собственные вектора e_i задают новый базис в пространстве X

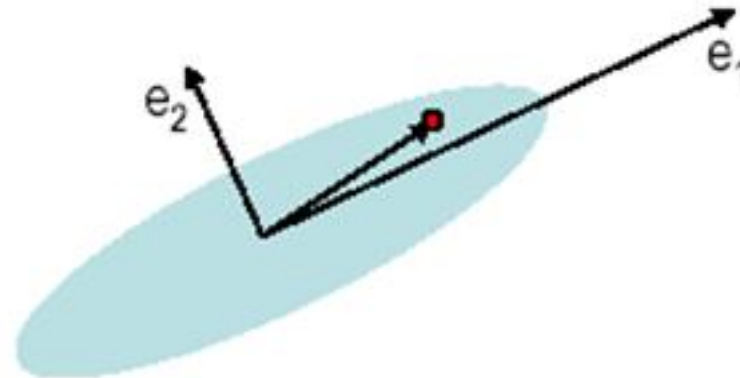


Метод главных компонент

- Значение λ_i указывает на разброс (variation) векторов вдоль направления e_i
- Большая часть λ_i может быть близка к нулю и ими можно пренебречь

Пример:

$$\lambda_1 \neq 0, \lambda_2 = 0.$$





Метод главных компонент

- Спроецируем x на первые M главных векторов:

$$x \mapsto \mu + \sum_{i=1}^M a_i e_i$$

где

$$a_i = (x - \mu) \cdot e_i.$$

- Вектор $(a_1, \dots, a_M)$ - это «сжатая» версия вектора X
 - Длина M вместо N
- Долю данных, покрываем первыми M главными векторами можно оценить как:

$$\frac{\sum_{i=1}^M \lambda_i}{\sum_{i=1}^N \lambda_i}$$



Метод главных компонент

- Если поставить задачу минимизировать сумму квадратов ошибки

$$J(\mu, a, e) = \sum_{k=1}^N ||(\mu + \sum_{i=1}^M a_{ki}e_i) - x_k||$$

- То можно показать, что решением этой задачи будет:

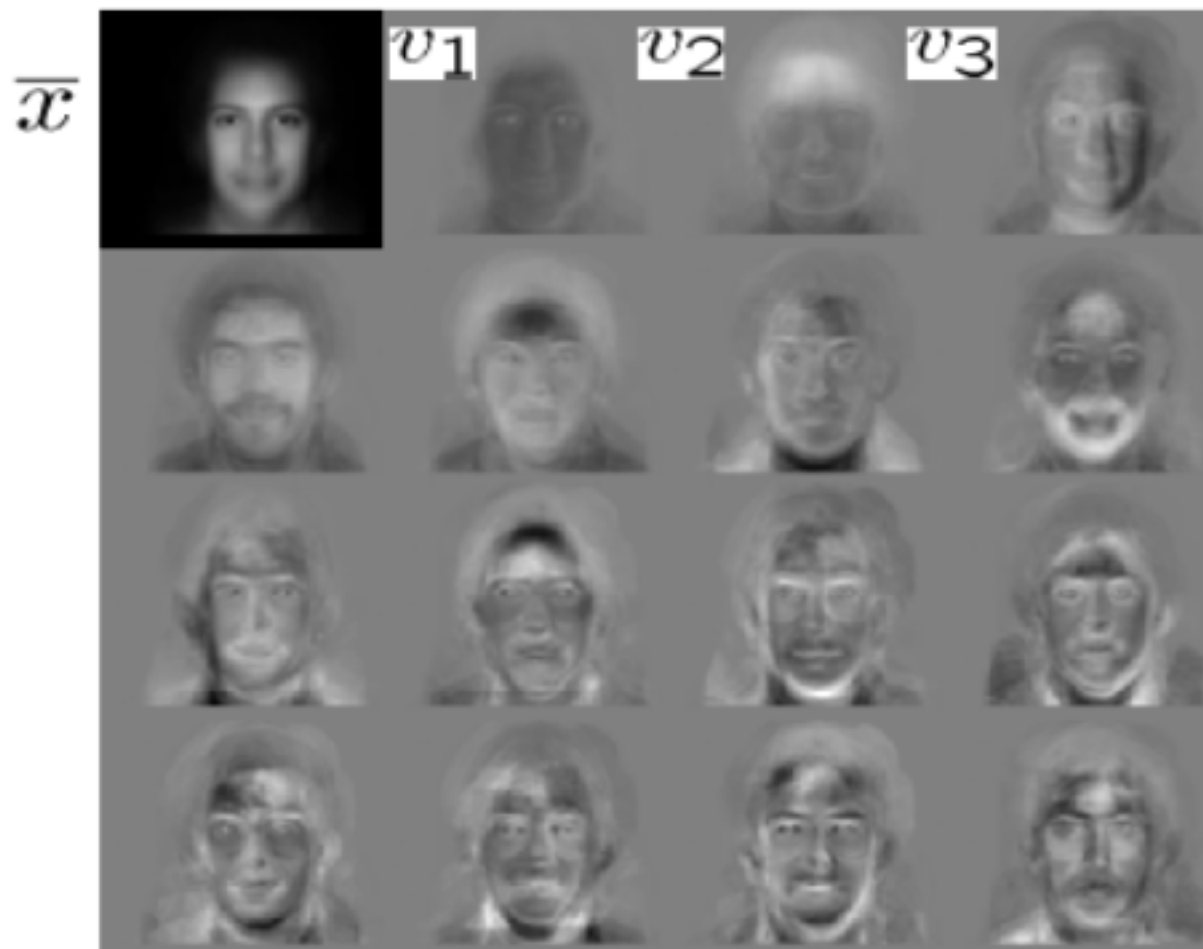
$$\mu = (1/N) \sum_{i=1}^N x_i.$$

- e_i – собственные вектора матрицы ковариации K
- a_i – проекция вектора x на собственные вектора e_i



Получение базиса лиц

$$\text{PCA: } \mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$$



Среднее лицо и базисные вектора

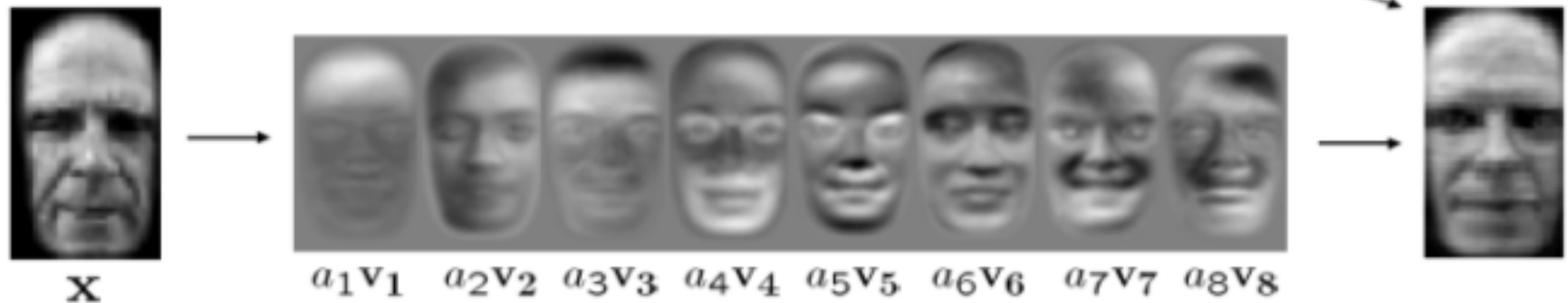


EigenFaces

Разложение лица по новому базису из PCA:

$$x \rightarrow (\underbrace{(x - \bar{x}) \cdot v_1}_{a_1}, \underbrace{(x - \bar{x}) \cdot v_2}_{a_2}, \dots, \underbrace{(x - \bar{x}) \cdot v_K}_{a_K})$$

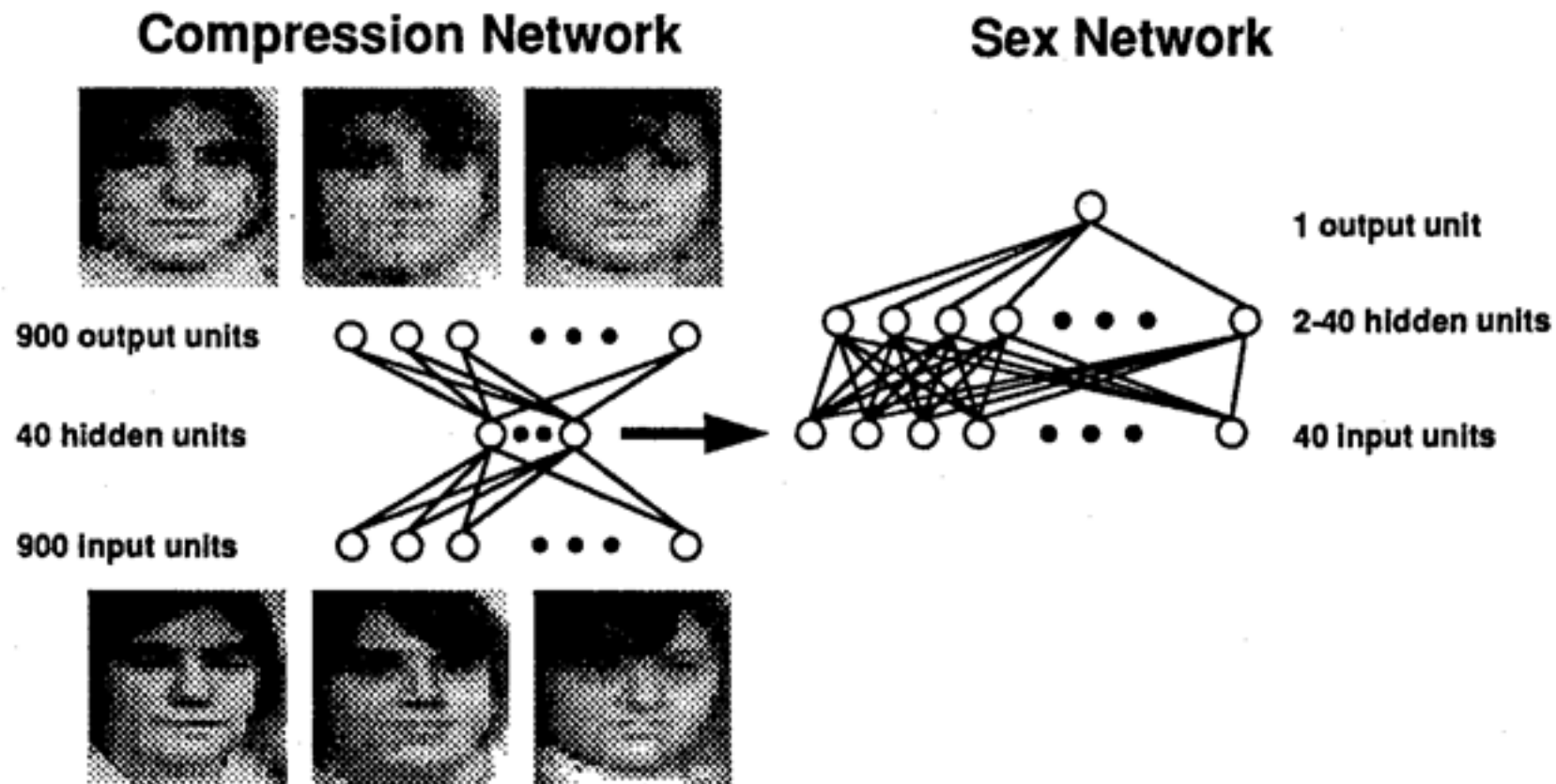
$$x \approx \bar{x} + a_1 v_1 + a_2 v_2 + \dots + a_K v_K$$



- Первые коэффициенты разложения лица используются как признаки в задачах идентификации личности
- Самые первые 3 компоненты в PCA считаем, отмечают за свет, поэтому при сравнении их отбрасываем



Sex Net – распознавание пола

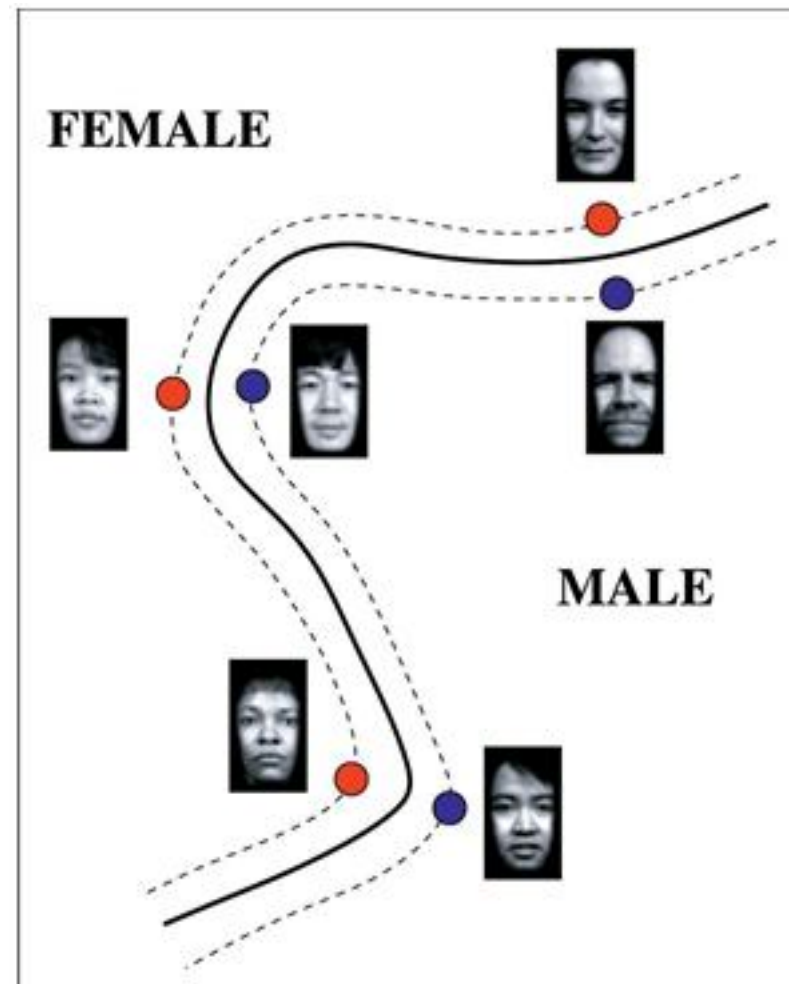


B. A. Golomb, D. T. Lawrence, T. J. Sejnowski SexNet: A neural network identifies sex from human faces, NIPS 1990



SVM-подход

- Уменьшенные изображения до 12*21 пикселей (thumbnails)
- SVM-RBF классификатор
- 300 опорных векторов
- FERET база
- Точность – 96.6%



B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *TPAMI*, vol. 24, no. 5, pp. 707–711, 2002



«Boosting Sex»



- Применим “boosting” подход для классификации пола лиц
- Опираемся на классификатор «позы лица», использующий сравнение интенсивности двух пикселей в изображении

S. Baluja and H. Rowley, “Boosting sex identification performance,” *IJCV*, 2007

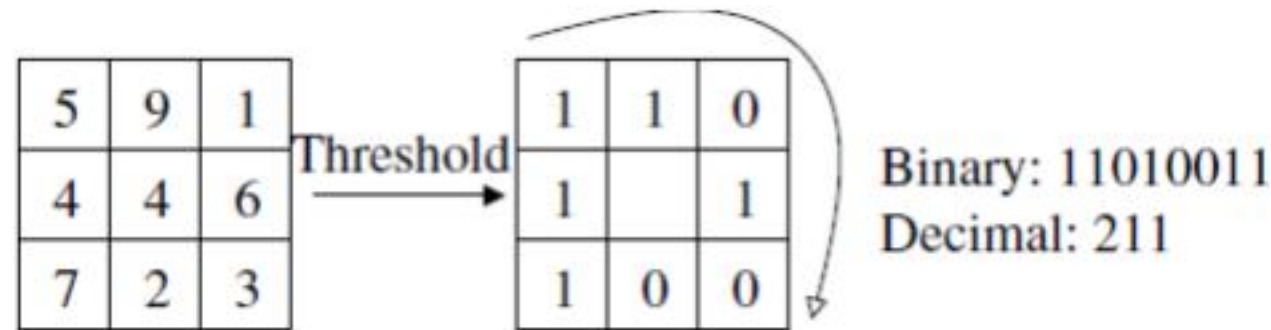


Результаты

- Обучали и тестировали на FERET
 - 10 сравнений – точность 80%
 - 20 сравнений – точность 90%
 - 1000 признаков – точность 96%
- Сравнение с SVM по скорости
 - В 25-50 раз быстрее

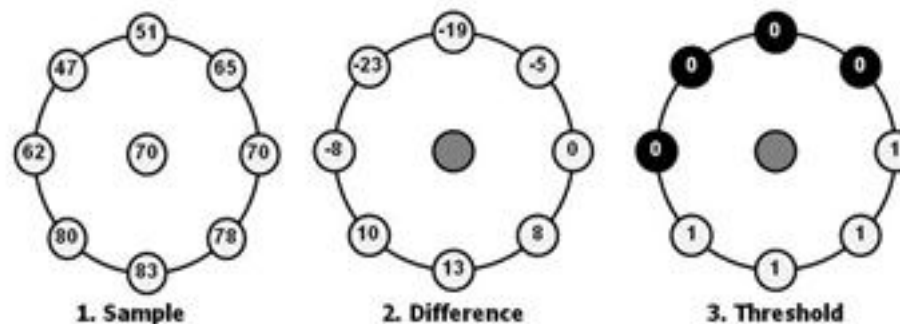


Local Binary Patterns



The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad s(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

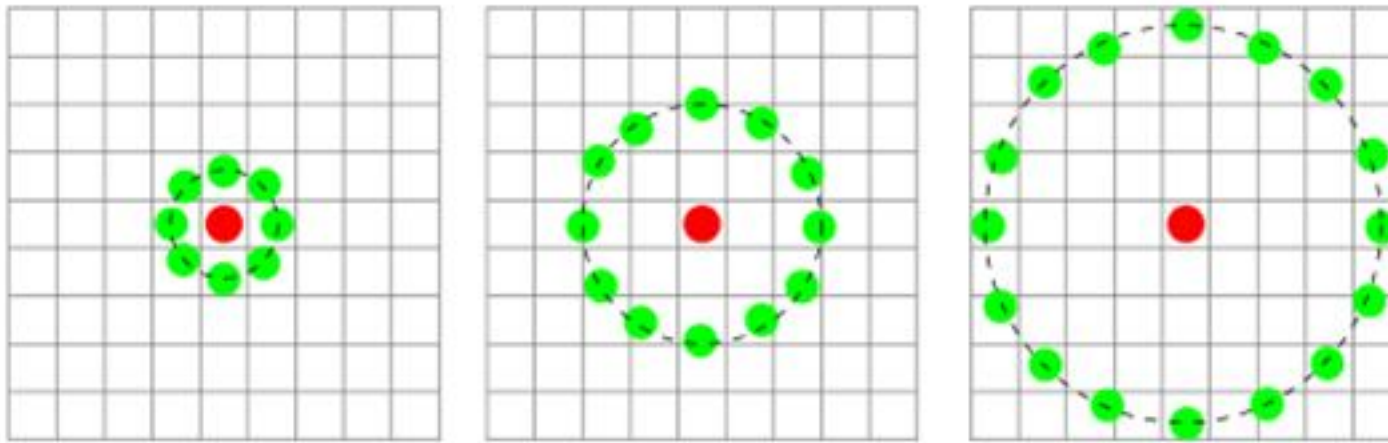


$$1 \cdot 1 + 1 \cdot 2 + 1 \cdot 4 + 1 \cdot 8 + 0 \cdot 16 + 0 \cdot 32 + 0 \cdot 64 + 0 \cdot 128 = 15$$

4. Multiply by powers of two and sum

T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", ICPR 1994

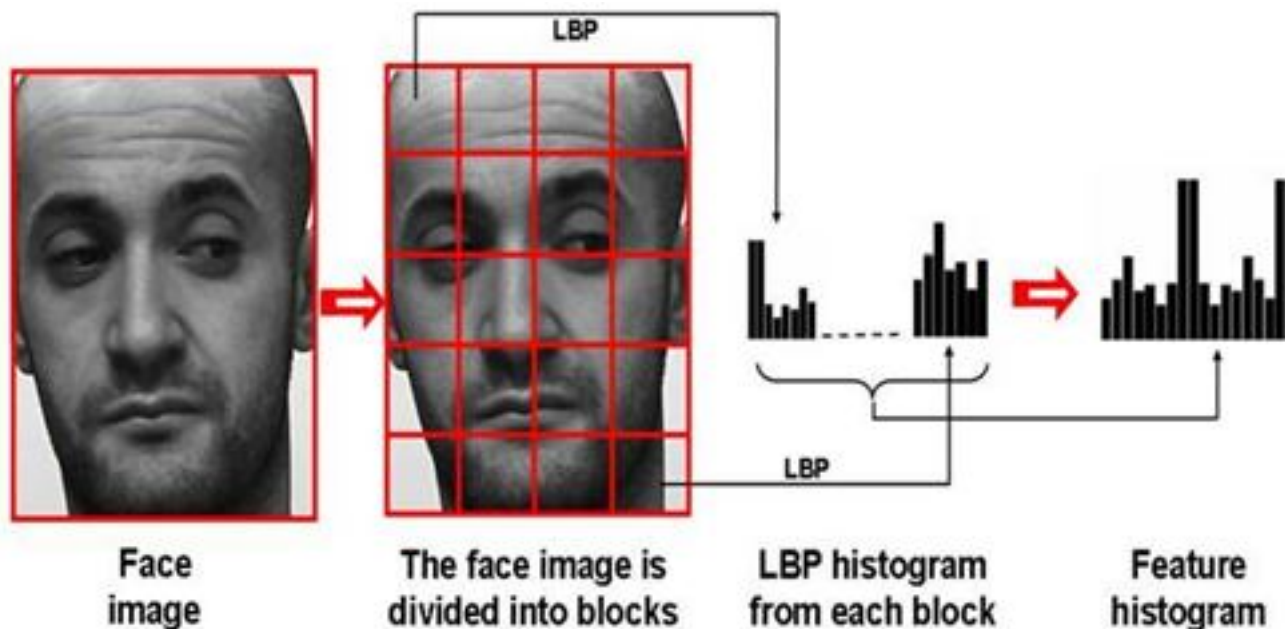
Расчёт LBP



Может потребоваться интерполировать значения в точках, необходимых для расчёта LBP



Применение LBP



- Изображение разбивается на области. В каждой области применяются LBP операторы к каждому пикселю. Строится гистограмма.
- Объединение гистограмм – LBP дескриптор для изображения.
- Для пары изображений считается разность дескрипторов по какой-нибудь метрике (например, Хи-квадрат)

Ahonen, T., Hadid, A. and Pietikäinen, M. (2006), Face Description with Local Binary Patterns: Application to Face Recognition. IEEE PAMI 28(12):2037-2041.

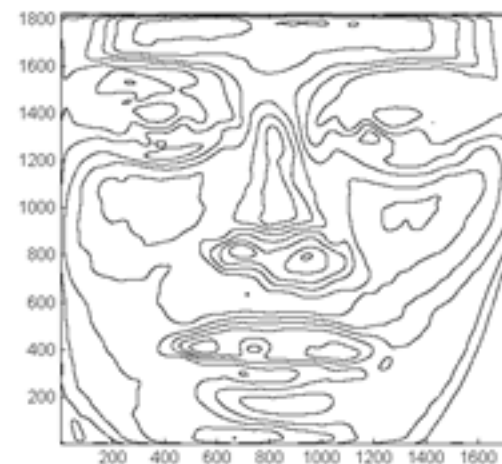
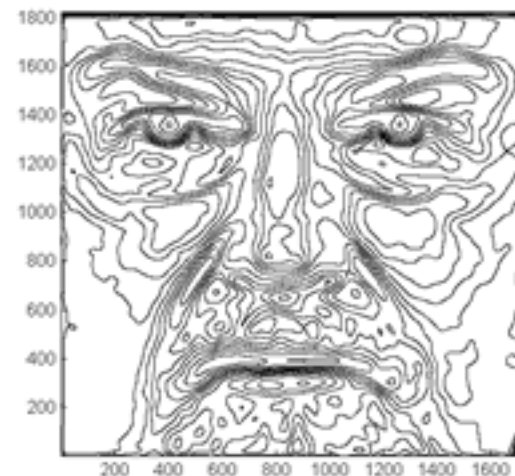


Uniform LBP

- Рассмотрим количество переходов:
 - 00000000 (0 transitions)
 - 01110000 (2 transitions)
 - 11001111 (2 transitions)
 - 11001001 (4 transitions)
 - 01010010 (6 transitions)
- Однородными LBP называются такие коды, в которых переходов ≤ 2
- Однородные составляют от 80 до 90% всех кодов
- Неоднородные коды иногда объединяют в один
- Для кода из 8 бит
 - 256 кодов всего
 - 58 однородных (59 разных меток)



Интерпретация однородных LBP

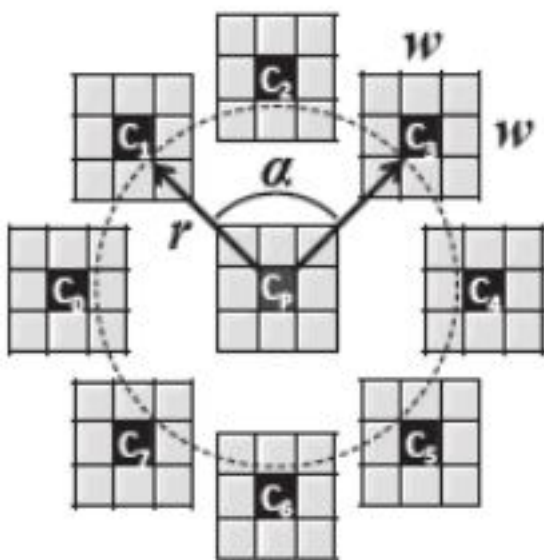


- Коды с 2мя переходами - кодирование изолиний



Развитие LBP

- Придумали два новых признаковых дескриптора: Three-Patch LBP и Four-Patch LBP - усовершенствованные LBP.
- Будем сравнивать не пиксели, а фрагменты изображения между собой (попиксельно по L2 и т.д.)
- $f(x) = 1$ если $x > T$, и 0, если наоборот
- Затем всё переводим

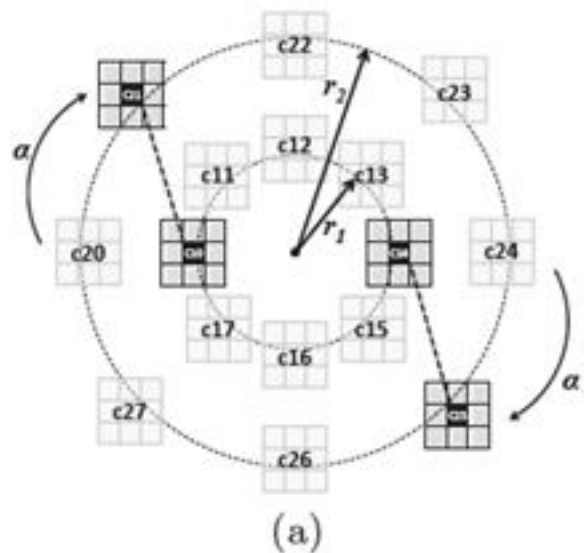


$$\begin{aligned} \text{TPLBP}_{r,8,3,2}(p) = & \\ & f(d(C_0, C_p) - d(C_2, C_p))2^0 + \\ & f(d(C_1, C_p) - d(C_3, C_p))2^1 + \\ & f(d(C_2, C_p) - d(C_4, C_p))2^2 + \\ & f(d(C_3, C_p) - d(C_5, C_p))2^3 + \\ & f(d(C_4, C_p) - d(C_6, C_p))2^4 + \\ & f(d(C_5, C_p) - d(C_7, C_p))2^5 + \\ & f(d(C_6, C_p) - d(C_0, C_p))2^6 + \\ & f(d(C_7, C_p) - d(C_1, C_p))2^7 \end{aligned}$$

Lior Wolf, Tal Hassner and Yaniv Taigman, "Descriptor Based Methods in the Wild," Faces in Real-Life Images workshop at the European Conference on Computer Vision (ECCV), 2008.



Four-Patch LBP



$$\text{FPLBP}_{r_1, r_2, 8, 3, 1}(p) =$$
$$f(d(C_{10}, C_{21}) - d(C_{14}, C_{25}))2^0 +$$
$$f(d(C_{11}, C_{22}) - d(C_{15}, C_{26}))2^1 +$$
$$f(d(C_{12}, C_{23}) - d(C_{16}, C_{27}))2^2 +$$
$$f(d(C_{13}, C_{24}) - d(C_{17}, C_{28}))2^3$$



Сравнение двух пар фрагментов



Обучение метрик сравнения

- Метрика COSINE:

$$CS(x, y) = \frac{x^T y}{\|x\| \|y\|}$$

- Обученная метрика

$$CS(x, y, A) = \frac{(Ax)^T (Ay)}{\|Ax\| \|Ay\|} = \frac{x^T A^T A y}{\sqrt{x^T A^T A x} \sqrt{y^T A^T A y}}$$

- Обучение метрики:

$$f(A) = \sum_{i \in Pos} CS(x_i, y_i, A) - \alpha \sum_{i \in Neg} CS(x_i, y_i, A) - \beta \|A - A_0\|^2$$



Алгоритм на основе COSINE

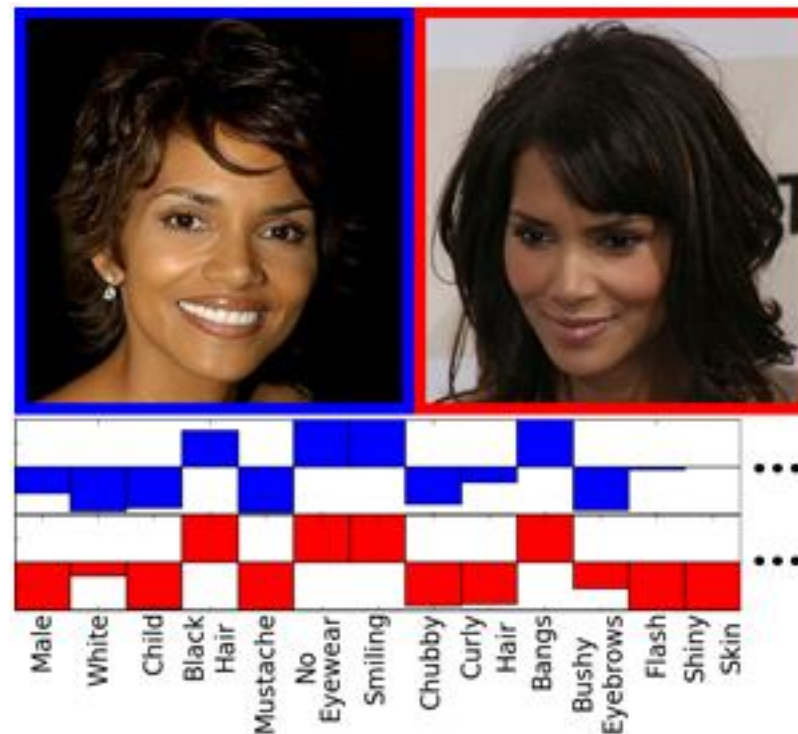
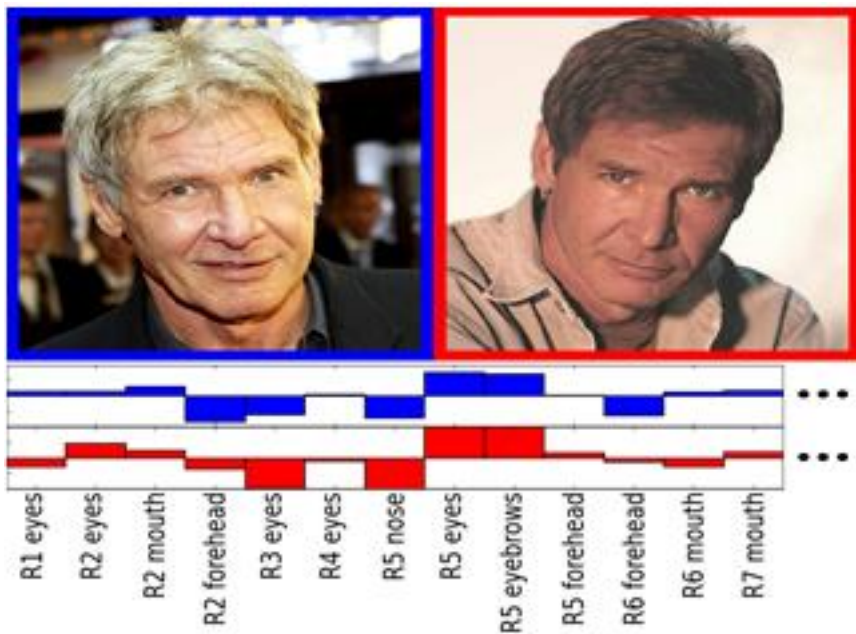
- Извлечь признаки
- Применить PCA для уменьшения размерности
- Обучить матрицу A для метрики
- Подобрать порог на валидационной выборке

Замечания:

- Если вектора нормализованы (длина = 1) то скалярное произведение = косинусная метрика
- Часто оказывается очень эффективной



Аттрибуты для верификации



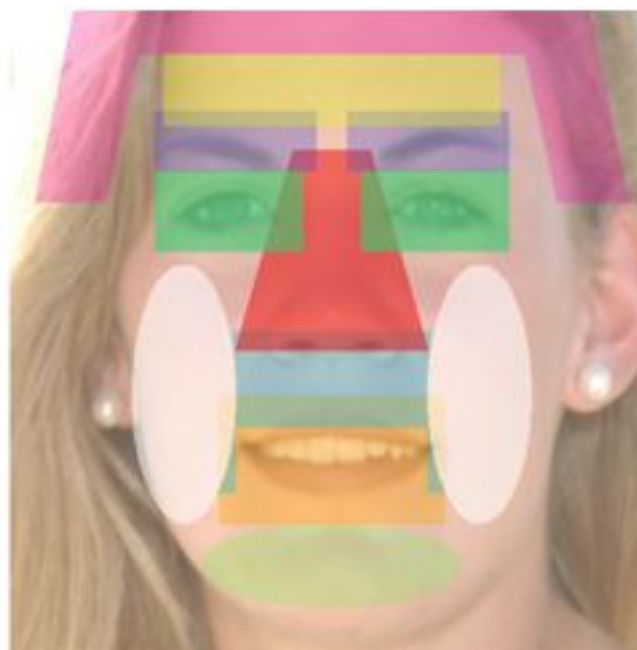
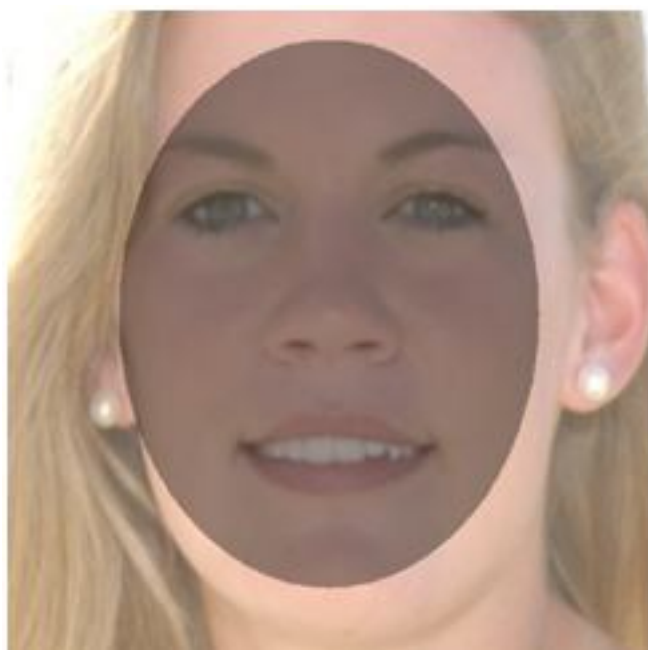
Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar.
Attribute and Simile Classifiers for Face Verification. /ICCV, 2009.

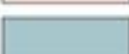


Построение классификатора

- Есть набор областей изображения
- Есть набор признаков
- Есть SVM + RBF
- Процедура
 - Обучим классификаторы на всех комбинациях (область, признак)
 - Выбираем наилучшую комбинацию (область, признак)
 - Выберем вторую наилучшую комбинацию, добавим в набор, обучим классификатор на наборе
 - Будем добавлять новые наилучшие комбинации в набор, пока не прекратиться рост качества (или до 6 комбинаций)

Области лица



-  Whole Face (W)
-  Hair (H)
-  Forehead (F)
-  Eyebrows (B)
-  Eyes (E)
-  Nose (N)
-  Cheeks (K)
-  Upper Lip (U)
-  Mouth (M)
-  Chin (C)



Виды особенностей

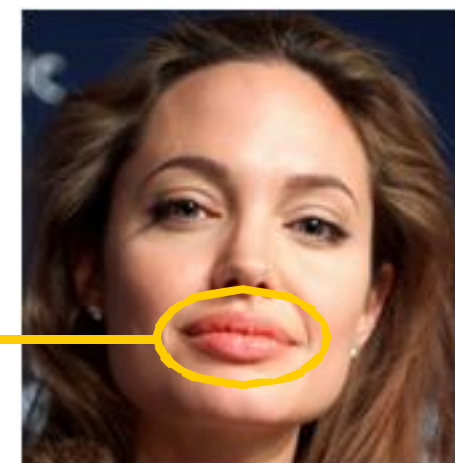
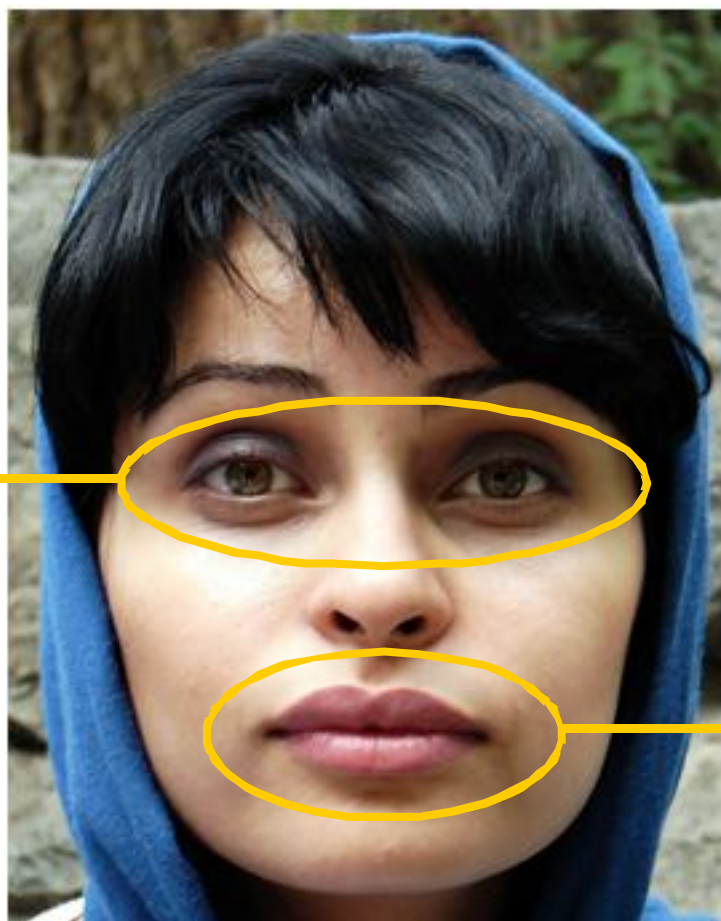
Pixel Value Type	Normalizations	Aggregation
RGB	None	None
HSV	Mean-Norm	Histogram
Image Intensity	Energy-Norm	Statistics
Edge Magnitude		
Edge Orientation		



«Similes» / «Подобия»



Penelope
Cruz



Angelina Jolie

«Подобия» - классификатор, похожа ли данная часть лица на тестовом изображении и у другого конкретного человека



«Эталонные люди»



Эталонный человек R1



Эталонный человек R2

Обучение «подобий»



Изображения Пенелопы Круз (её глаз)



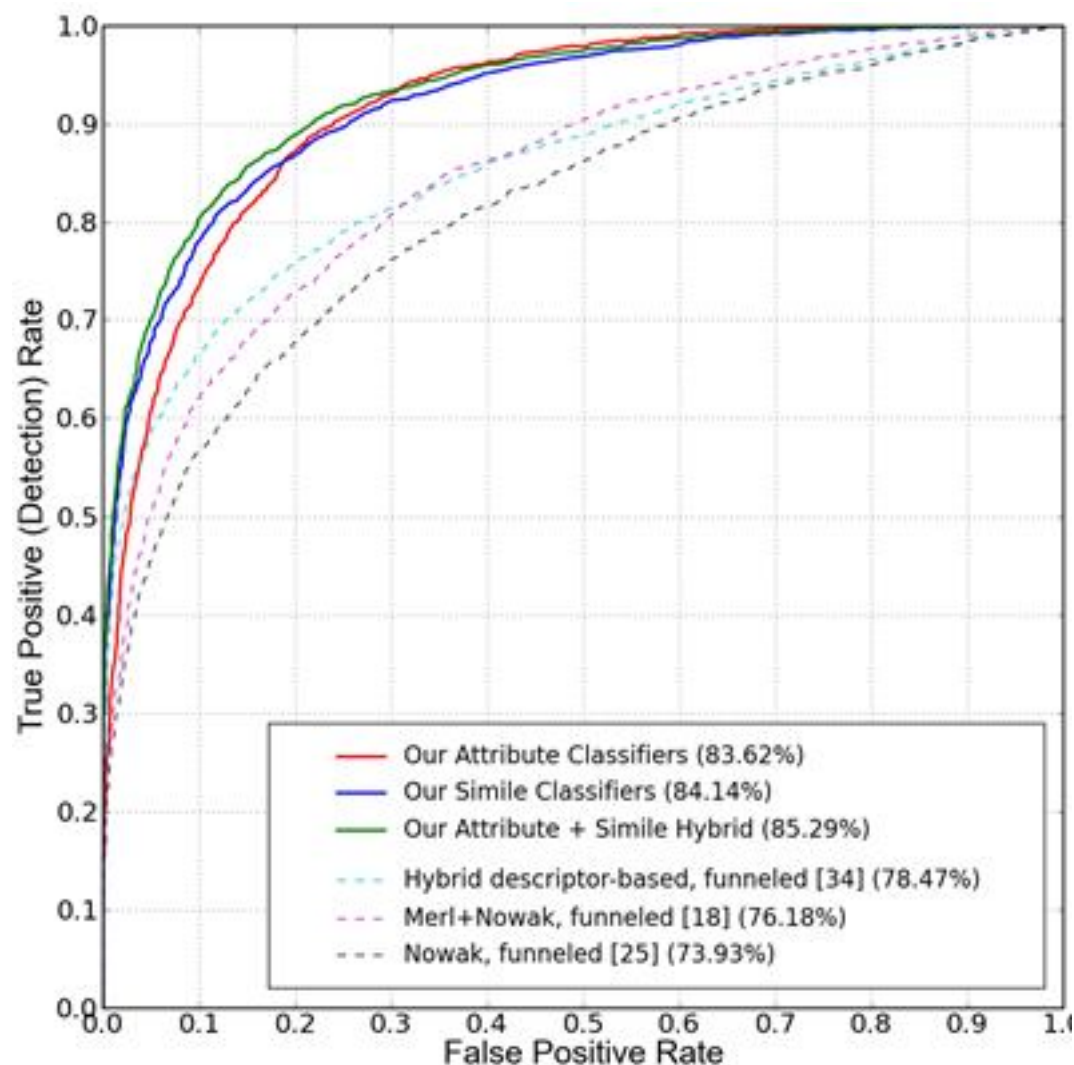
Изображения других людей (их глаз)

60 человек, 8 регионов, 6 типов особенностей (без выбора), для
каждого обучаем классификатор



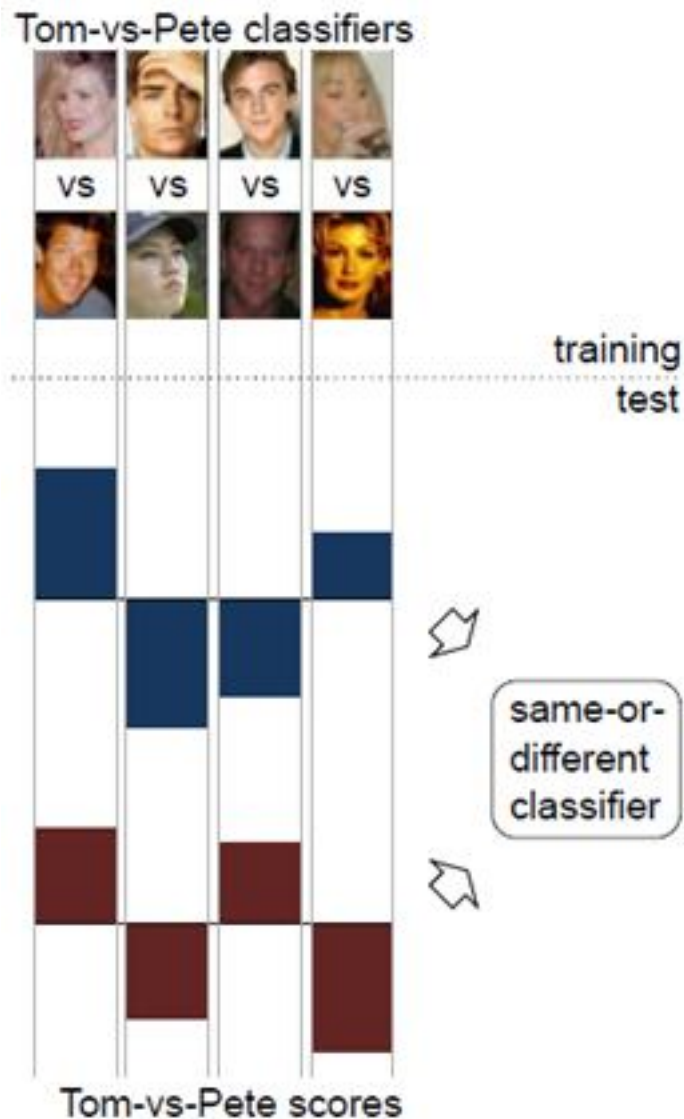
Алгоритм верификации

- Выходы всех классификаторов атрибутов и similes объединяются в один вектор признаков.
- Этот вектор подается на вход новому SVM + RBF классификатору.





Tom-vs-Pete

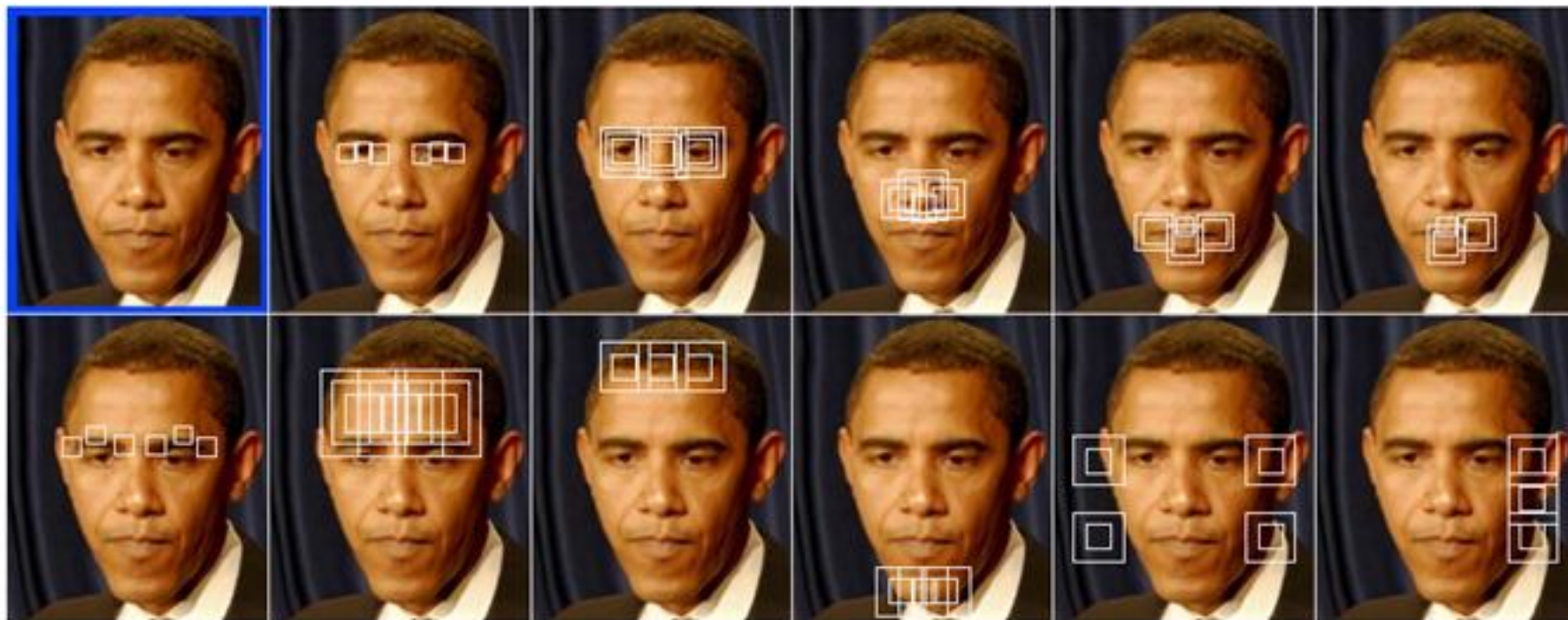


- Развитие «подобий»
- Берём пару людей
- Обучаем классификатор, на кого из пары больше похоже тестовое изображение

Thomas Berg and Peter N. Belhumeur.
Tom-vs-Pete Classifiers and Identity-Preserving Alignment for Face Verification.
British Machine Vision Conference (BMVC), 2012.



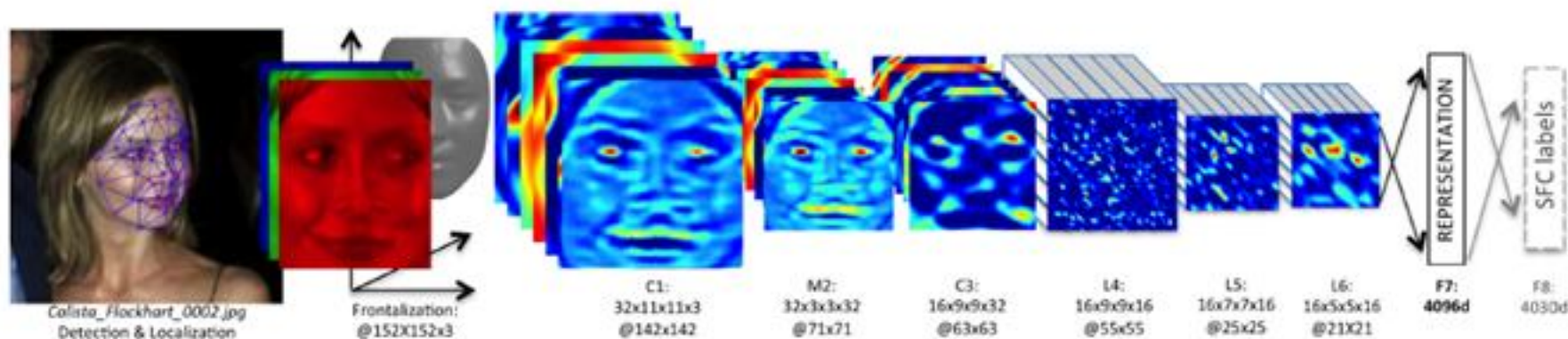
Детали Tom-vs-Pete



- 120 человек, 11 областей расчёта
- Базовый классификатор – HOG по области, линейный SVM
- Отбор 5000 Tom-vs-Pete классификаторов с помощью boosting-процедуры
- Конкатенация разностей выходов для получения общего дескриптора
- RBF SVM для принятия решения



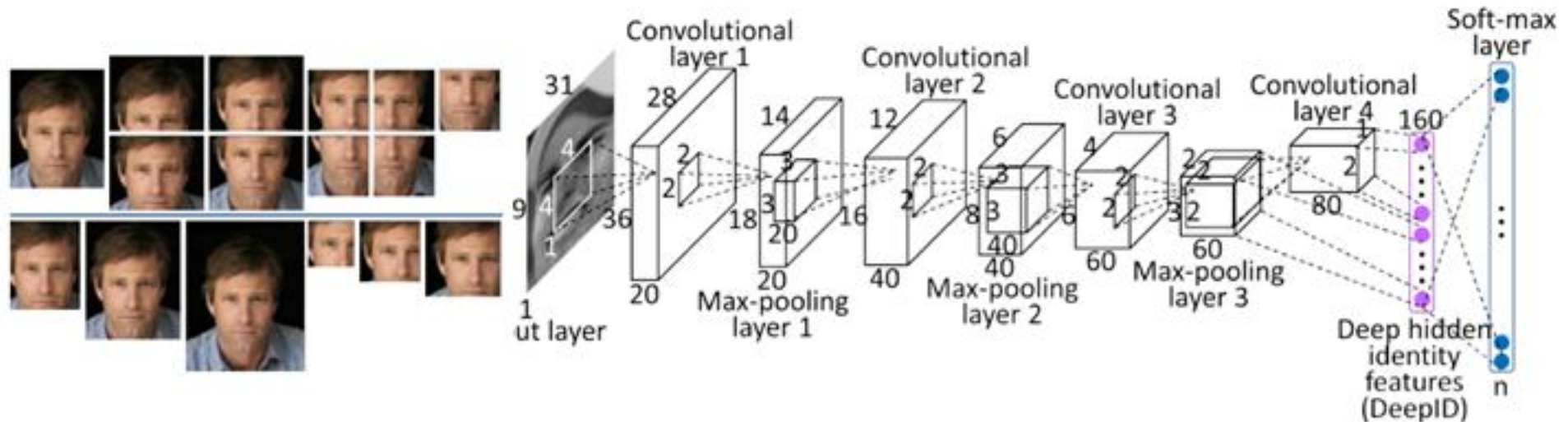
DeepFace – пример нейросетей



- Выборка из 4000 людей по 1000 изображений на каждого
- Аккуратное совмещение через 3D нормализацию, поэтому используют не только свёрточные, но и локально-свёрточные слои
- Обучение нейросети как классификатор 4000 людей
- Выход предпоследнего уровня = признаки для сравнения

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. **DeepFace: Closing the Gap to Human-Level Performance in Face Verification.** CVPR 2014.

DeepID



- Комитет нейросетей (до 25)
- Каждая смотрит свой фрагмент лица
- Обучают как классификатор 10000 классов
- Объединение верхних слоёв всех сетей = признак
- Итог – STAR-классификатор

Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep Learning Face Representation from Predicting 10,000 Classes. *CVPR*, 2014.



Текущие результаты

Attribute classifiers ¹¹	0.8525 ± 0.0060
Simile classifiers ¹¹	0.8414 ± 0.0041
Attribute and Simile classifiers ¹¹	0.8554 ± 0.0035
Multiple LE + comp ¹⁴	0.8445 ± 0.0046
Associate-Predict ¹⁸	0.9057 ± 0.0056
Tom-vs-Pete ²³	0.9310 ± 0.0135
Tom-vs-Pete + Attribute ²³	0.9330 ± 0.0128
combined Joint Bayesian ²⁶	0.9242 ± 0.0108
high-dim LBP ²⁷	0.9517 ± 0.0113
DFD ³³	0.8402 ± 0.0044
TL Joint Bayesian ³⁴	0.9633 ± 0.0108
face.com r2011b ¹⁹	0.9130 ± 0.0030
Face++ ⁴⁰	0.9727 ± 0.0065
DeepFace-ensemble ⁴¹	0.9735 ± 0.0025
ConvNet-RBM ⁴²	0.9252 ± 0.0038
POOF-gradhist ⁴⁴	0.9313 ± 0.0040
POOF-HOG ⁴⁴	0.9280 ± 0.0047
FR+FCN ⁴⁵	0.9645 ± 0.0025
DeepID ⁴⁶	0.9745 ± 0.0026
GaussianFace ⁴⁷	0.9852 ± 0.0066
DeepID2 ⁴⁸	0.9915 ± 0.0013



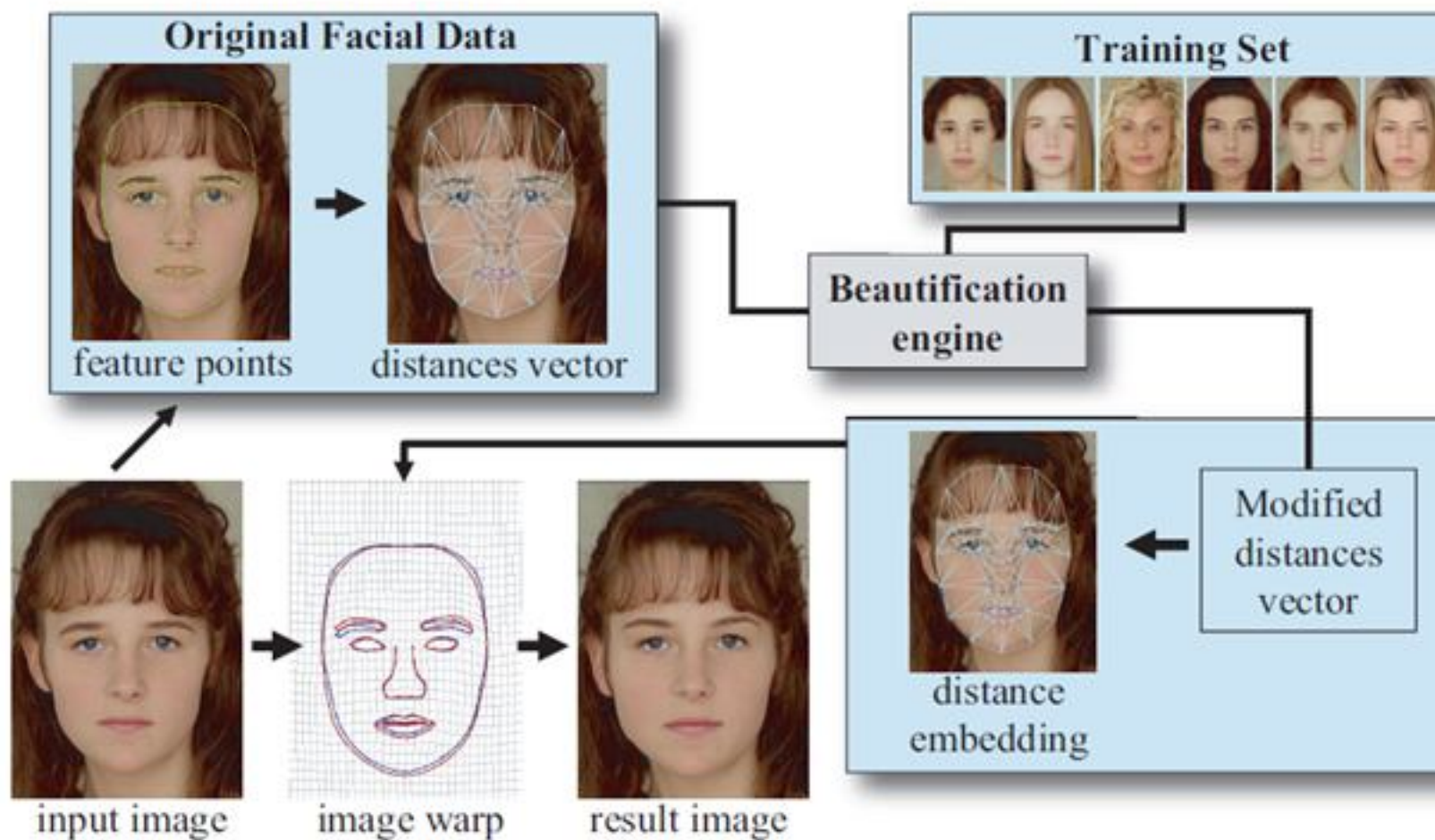
«Метод Тюбика»

- Эксперименты показали, что можно обучить классификатор SVM так, чтобы его результаты хорошо коррелировали с оценками людей (по шкале от 1 до 7 баллов)
- Попробуем чуть-чуть поправить все пропорции лица так, чтобы оно казалось красивее

EISENTHAL, Y., DROR, G., AND RUPPIN, E. 2006. Facial attractiveness: Beauty and the machine. *Neural Computation* 18, 1, 119–142.

[Tommer Leyvand](#), [Daniel Cohen-Or](#), [Gideon Dror](#) and [Dani Lischinski](#) Data-Driven Enhancement of Facial Attractiveness [ACM SIGGRAPH 2008](#)

Схема



- Вычисление черт на основе ASM моделей



Улучшение лица

- Из изображения извлекаем вектор v , описывающий геометрию модели.
- Нужно найти v' , близкий к исходному, но с более высокой оценкой красоты
- Подходы:

- K-NN

- Ранжируем все изображения в базе по
- Затем усредняем k ближайших

$$w_i = \frac{b_i}{\|v - v_i\|}$$

- Только выпуклые комбинации примеров из базы

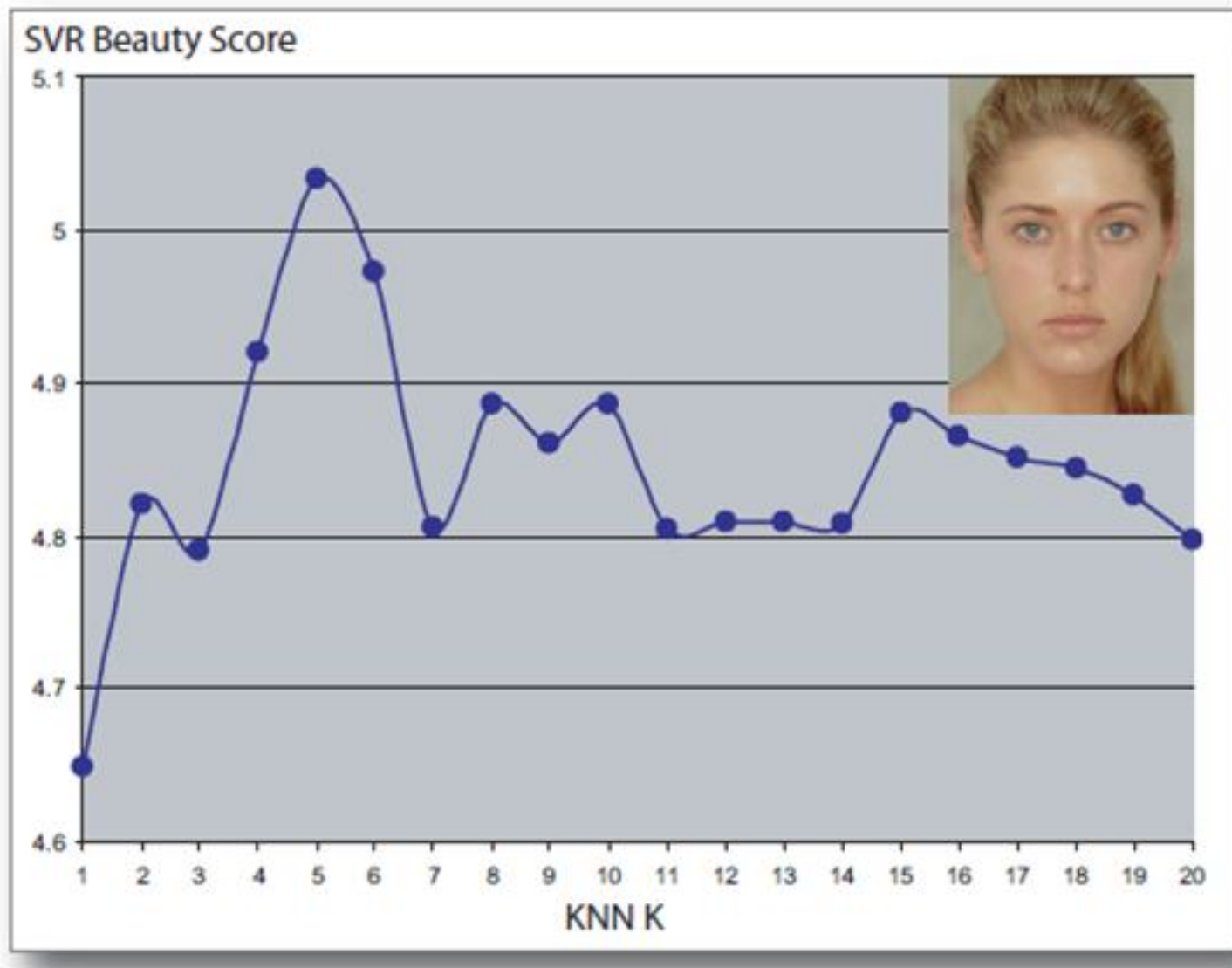
$$v' = \frac{\sum_{i=1}^K w_i v_i}{\sum_{i=1}^K w_i}$$

- SVR

- Оптимизация оценки красоты
- Проекция v через PCA (234 в 35)
- Регуляризация вероятностью модели (как в Active Shape)



Зависимость от К в К-NN



Сравнение подходов

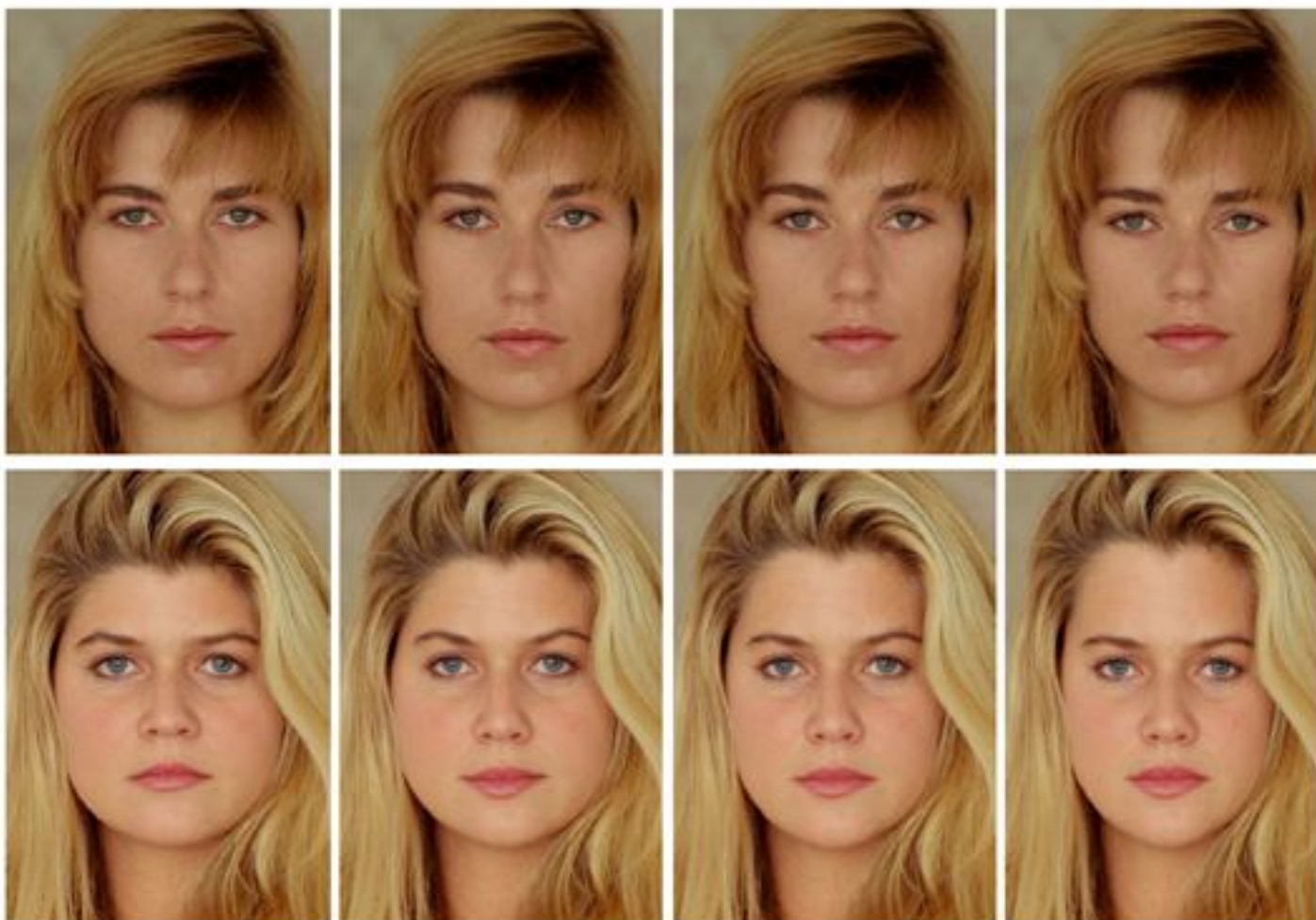
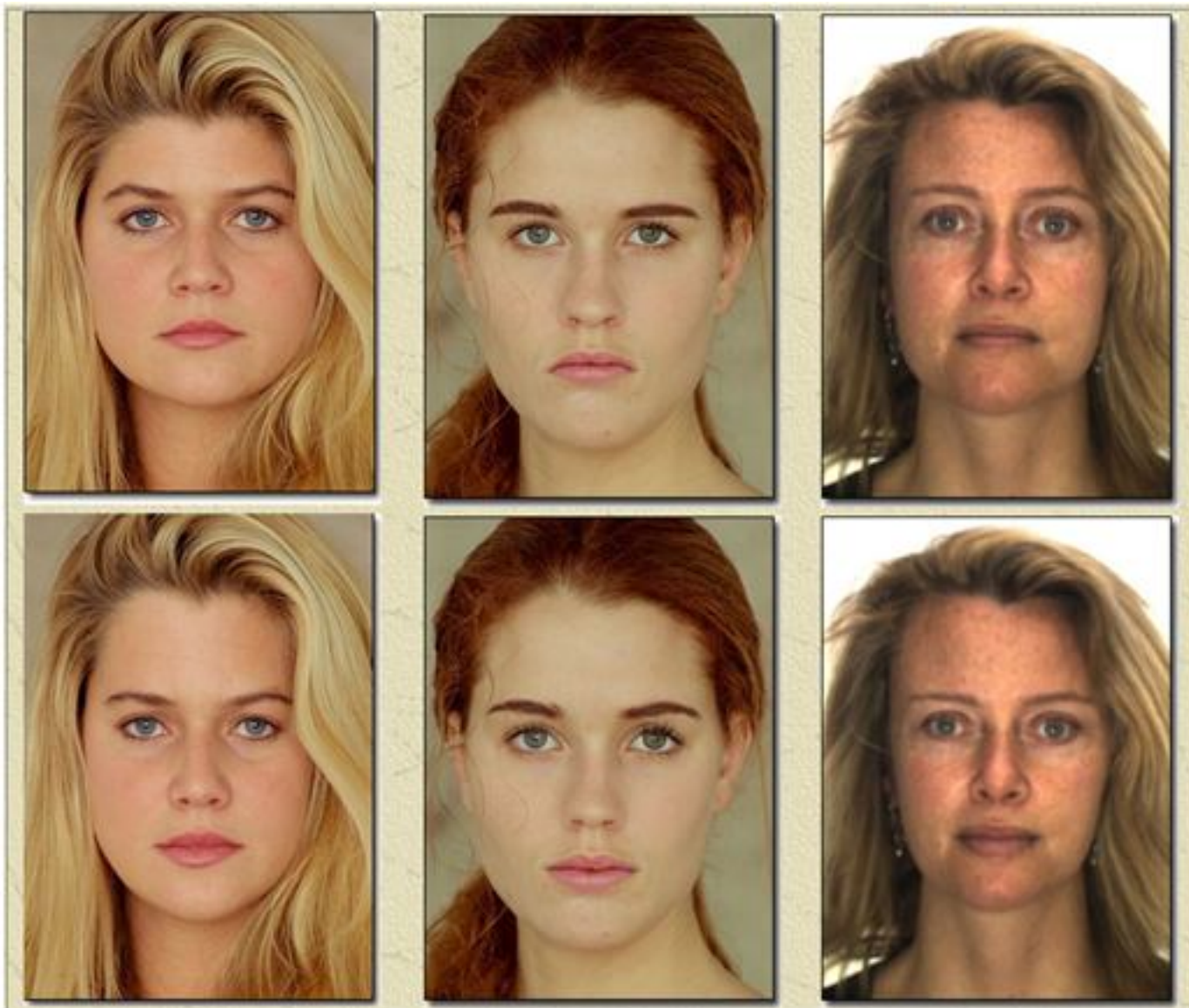


Figure 6: From left to right: original face, KNN-beautified with $K = 3$, KNN-beautified with optimal K , SVR-beautified.

Результаты



Результаты



Результаты





Data-Driven Enhancement of Facial Attractiveness

Tommer Leyvand
Tel Aviv University

Daniel Cohen-Or
Tel Aviv University

Gideon Dror
**Academic College of
Tel-Aviv-Yaffo**

Dani Lischinski
The Hebrew University



SIGGRAPH2008



Синтез анимации

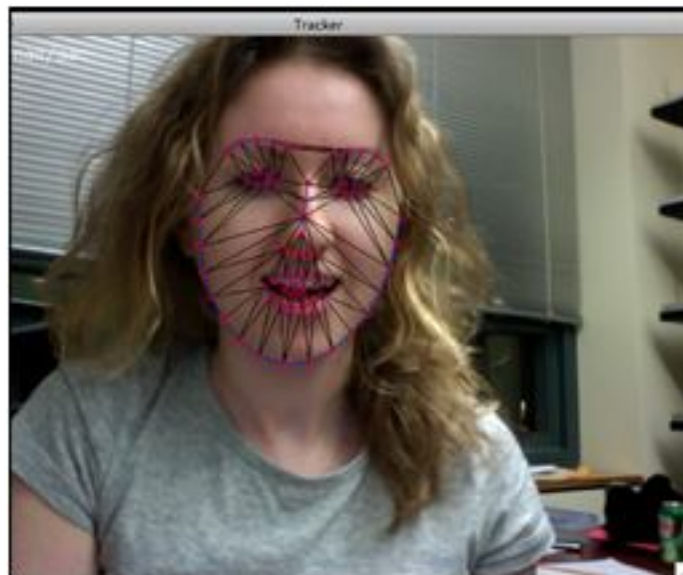


- Идея – для текущего изображения лица найдём в базе изображений / видео изображения с наиболее похожими выражениями

Ira Kemelmacher-Shlizerman, Aditya Sankar, Eli Shechtman, and Steven M. Seitz.
[Being John Malkovich](#), ECCV 2010



Выделение черт лица



- Любые методы подойдут
- Морфируемые модели дадут наиболее точный результат



Канонические лица



- Лицо отображется на каноническую позу. Таким образом, мы убираем разницу в позе и можем сравнивать выражение лица напрямую



Расстояние между выражениями



- Все изображение разбивается на блоки
- В каждом блоке считаем LBP
- Сравнивается область лица и область глаз

- Расстояние между блоками: $\chi^2(x, y) = 1/2 \sum_i (x_i - y_i)^2 / (x_i + y_i)$
- Для каждой области суммируем расстояние между всеми блоками
- Суммарное расстояние в выражении лиц – взвешенная сумма расстояний по области рта и области глаз:

$$d_{appear}(i, j) = \alpha^m d^m(i, j) + \alpha^e d^e(i, j)$$



Расстояние между изображениями

- Расстояние между позами:

$$d_{pose}(i, j) = L(|Y_i - Y_j|) + L(|P_i - P_j|) + L(|R_i - R_j|)$$

где $L(d)$ – логистическая функция

$$L(d) = \frac{1}{1 + e^{-\gamma(d-T)/\sigma}}$$

T отображается в 0.5, $T \pm \sigma$ в 0.01 и 0.99

- Суммарное расстояние:

$$D(i, j) = d_{appear}(i, j) + \alpha^p d_{pose}(i, j) + \alpha^t d_{appear}(i - 1, j)$$

- Учитывается сходство выражения между соседними кадрами



Примеры сопоставлений лиц



(a) Full measure

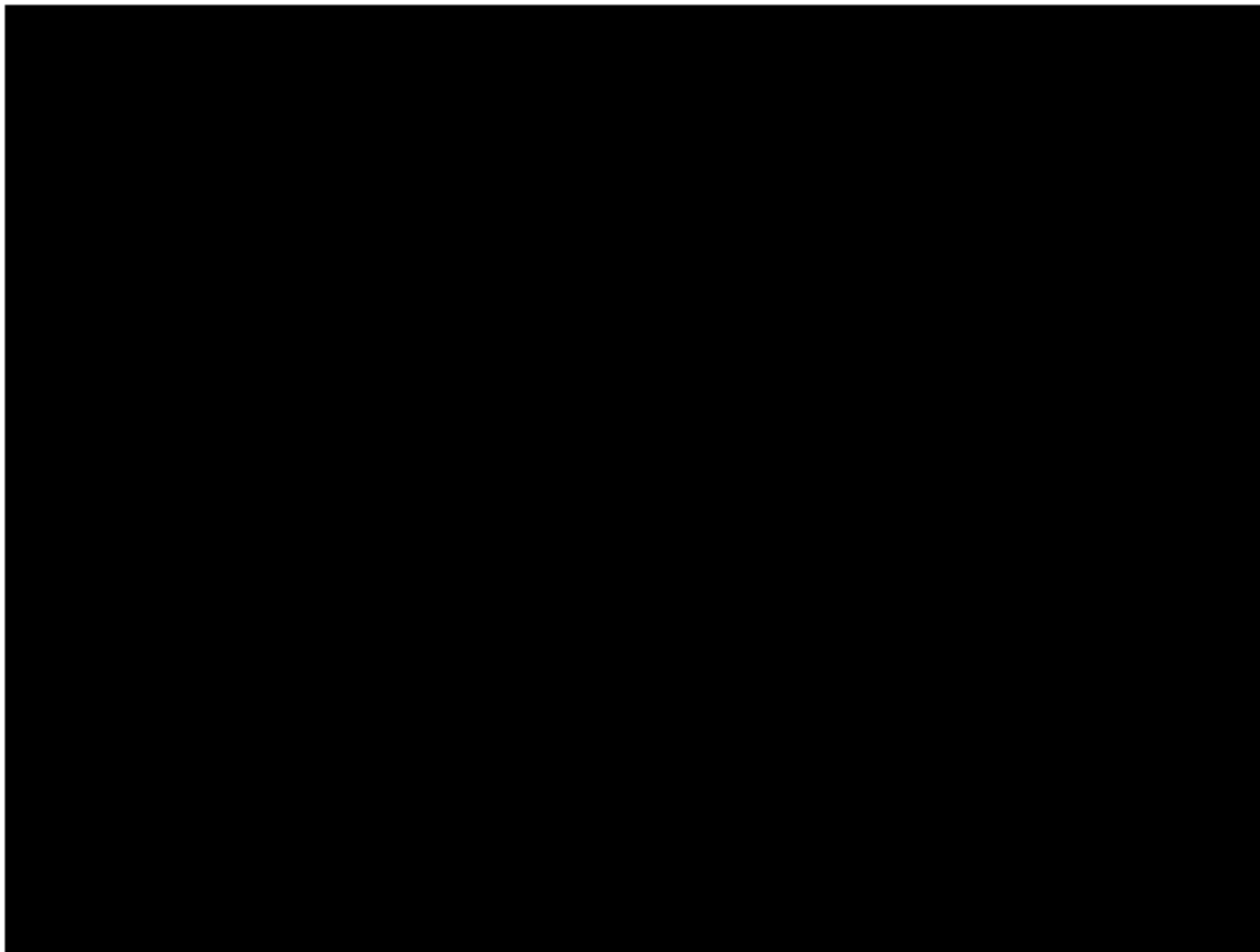


(b) Without mouth similarity



(c) Without eyes similarity

Видео





Резюме лекции

- Для предобработки изображения лица требуется знать минимум положения глаз
- Полезные инструменты:
 - Уменьшение размерности (PCA)
 - Текстурный признак LBP
 - Вычисление атрибутов и сравнений с «примерами» и использование их как высокоуровневых признаков
 - Нейросети