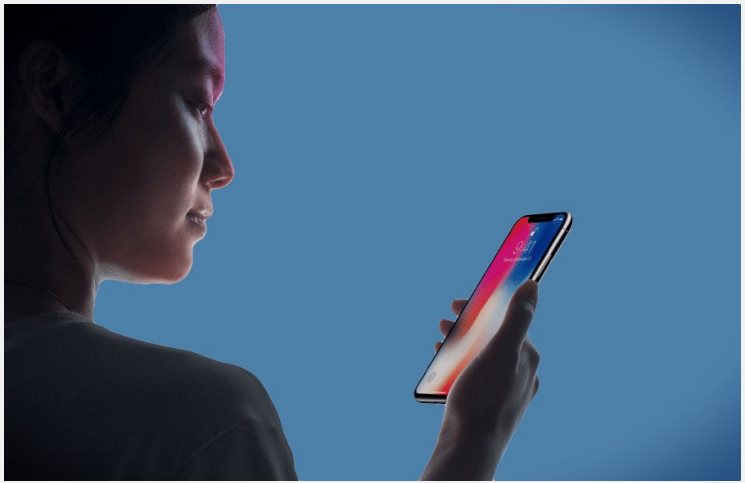
Imperial College London

Face Representations: Classical vs Modern

Binod Bhattarai
Ramesh Pathak, *Teaching Assistant*

Introduction

Source: *google image*



Source: *google image*

Imperial College London

Applications (Face verification for surveillance)

Credit: AFP/Getty Images

China's police have been testing sunglasses with built-in facial recognition since at least last month to catch suspects and those traveling under false identities. Now China is

Source: The verge

Source: google image

How facial recognition is taking over airports

CNN travel

(CNN) — Instead of scanning her boarding pass, the airport gate scanned her face.

In April 2019, traveler MacKenzie Fegan was left surprised and confused when she boarded a JetBlue flight from the United States to Mexico, without handing over her passport, or travel documents.

Face Analysis Tasks: Face retrieval

- Given a query image I_q, compute the ***similarity score*** between the query and the images in the database
- Rank them based on the ***sim_score***
- Retrieve **Top-K** ranked images

Face Analysis Tasks: Face Verification

Output whether the a pair of images are of a person

I_1 I_2

Same (+1)

I_2 I_3

Different (-1)

$f:(I_i, I_j) \rightarrow \{+1 , -1\}$

Challenges in Face Matching Algorithms

Illuminations: Same face looks differently at different illuminations

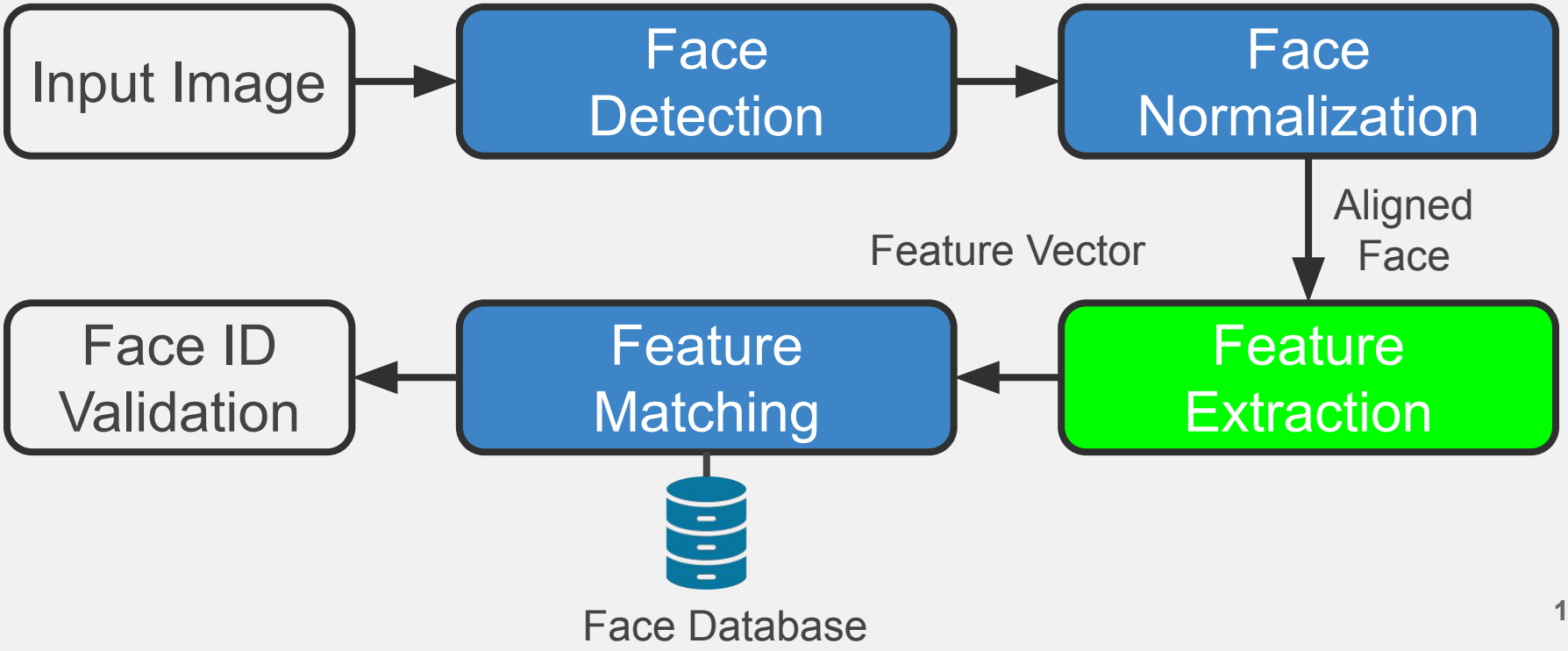
Challenges in Face Matching Algorithms

Poses: Same face looks differently at different poses

Challenges in Face Matching Algorithms

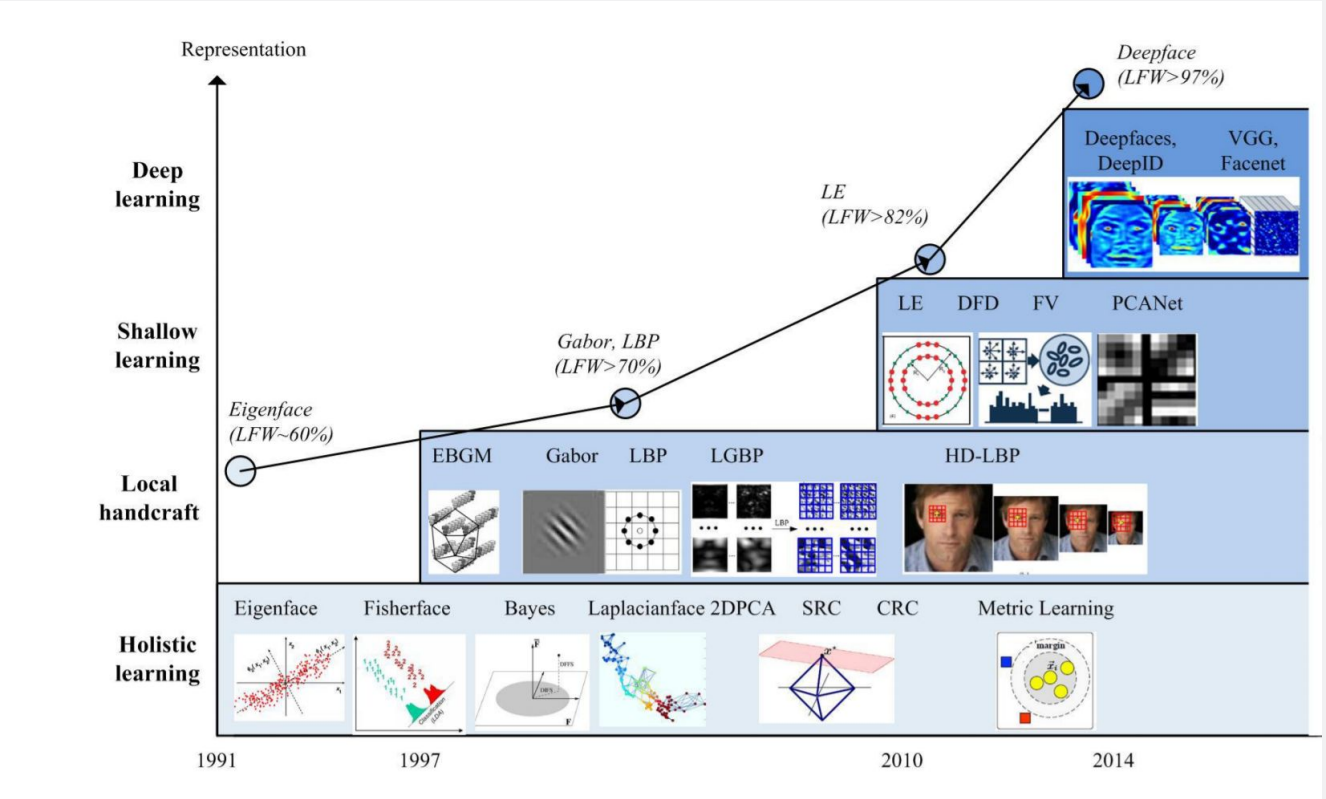
Expressions: Same face looks differently at different poses

Face Analysis Pipeline



Face Analysis Pipeline

How the face matching problem is addressed?



Naive Face Matching

1

Image Pixels to Vector (x)

2

y1,.....yn

Test if x matches some y_i in database

SSD: $(y_i - x)^2$

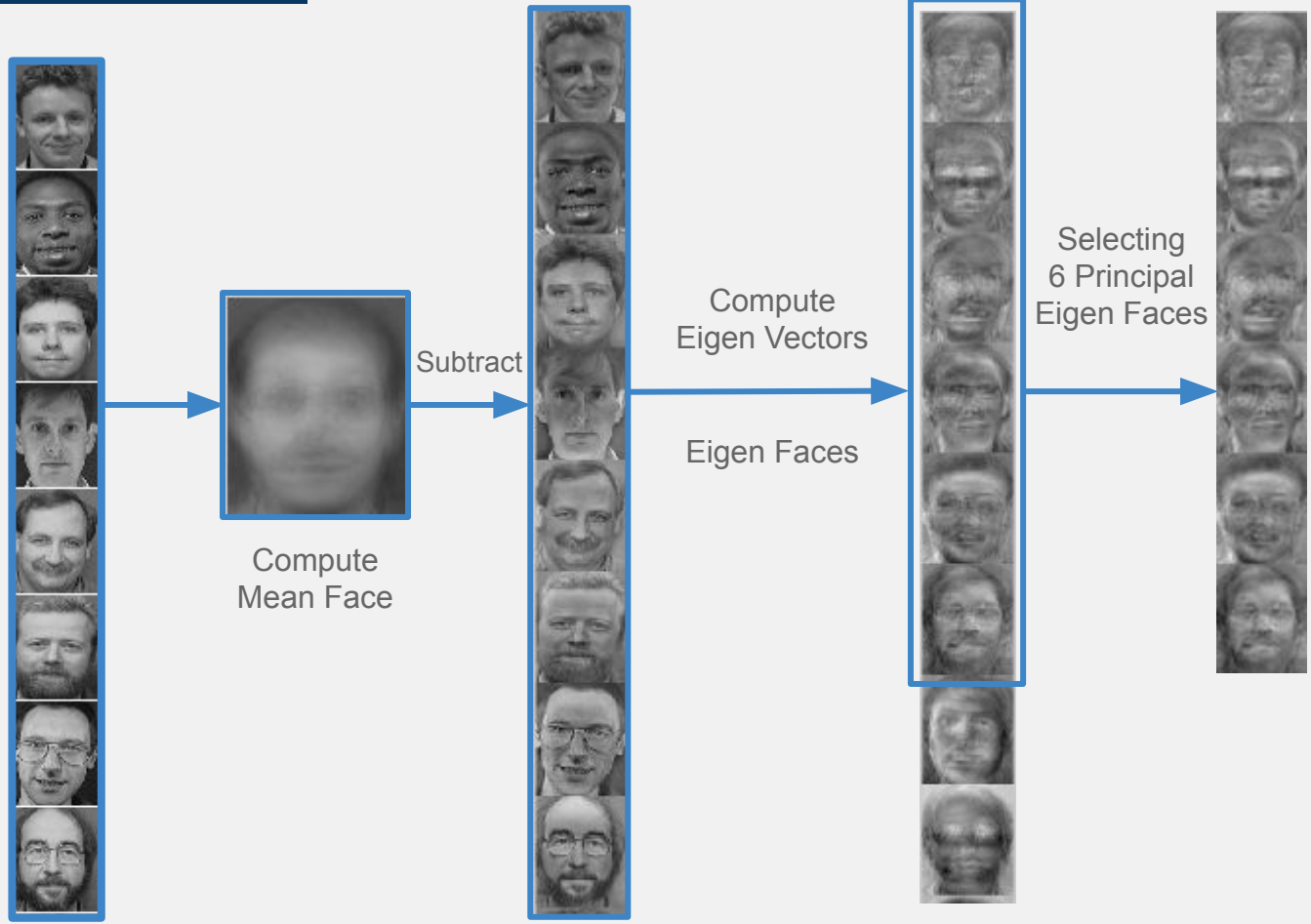
Zero Mean Normalized Cross Correlation (ZNCC):
The higher the ZNCC gets, the more are the two images correlated.

Problem of Naive Approach

- High dimensionality
- Not robust to illuminations, poses, expressions

Imperial College London

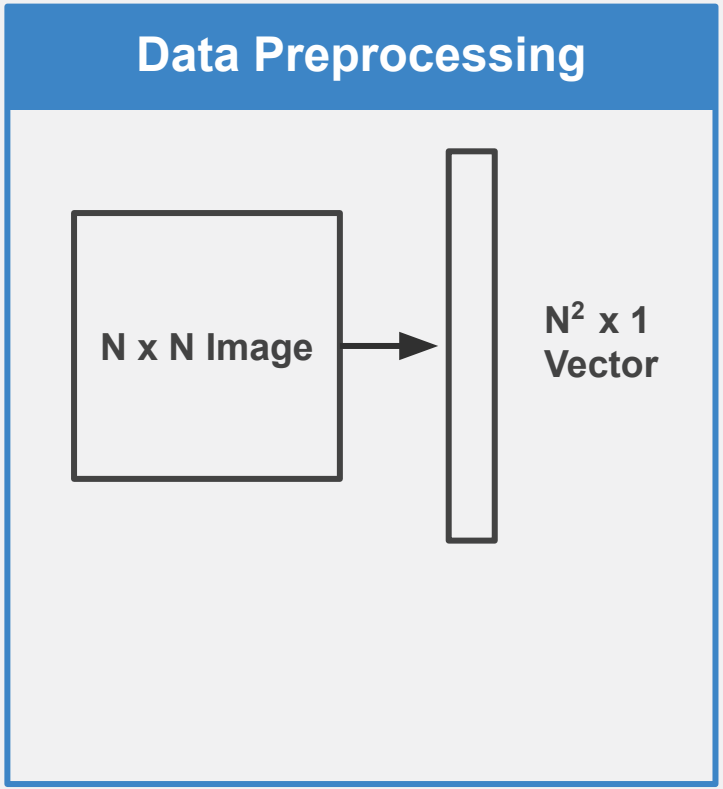
Eigen Faces



Imperial College London

Eigen Faces

Step 1



- Each sample image is rearranged into column vector of length **N x N = N^2 x 1**

  Eg: For a image of 112*92
  Column Vector length is:
  112*92=10304

- All images are put into a **matrix A** of size **N^2 x M**
  For a dataset of size M=400 images A is 10304x400

- Mean face is subtracted from each column.
- We apply PCA to compute eigen faces

Eigen Faces

Step 2

PCA

Find the eigenvectors of covariance matrix C of (face - mean face)

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T \quad (N^2 \text{x} N^2 \text{ matrix})$$

$$\text{where } A = [\Phi_1\ \Phi_2 \cdots \Phi_M] \quad (N^2 \text{x} M \text{ matrix})$$

Eigen Faces

Step 3

Projection

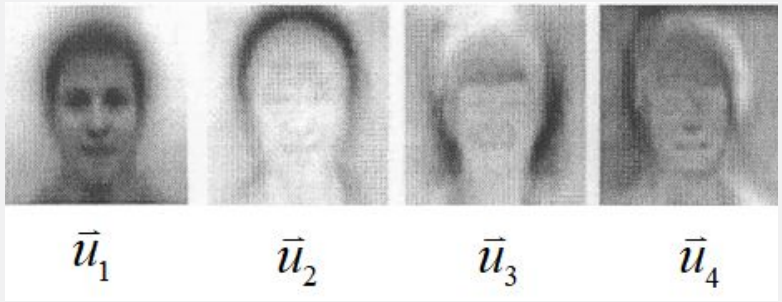
- Select top k eigenvectors with k largest eigenvalues **(k eigenfaces)**

- Do projection along these eigenfaces to find

Each face (minus the mean) in the training set can be represented as a linear combination of the best k eigenvectors:

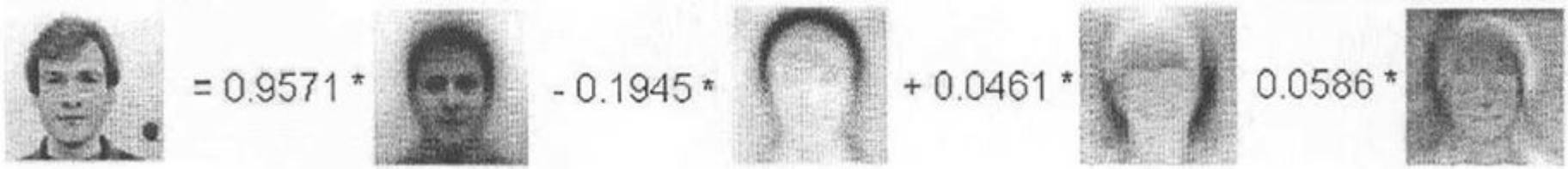
$$\hat{\Phi}_i - mean = \sum_{j=1}^{K} w_j u_j, \quad (w_j = u_j^T \Phi_i)$$

Typical eigenfaces when k=4:



Eigen Faces

Since $\{\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4\}$ is an orthonormal basis, any face (after mean subtraction) can be represented by this basis:



Each normalized training face Φi is represented in this basis by a vector:

PCA

- **Merits**
  - Removes redundancies and noise to make feature more discriminative
  - Transform the representations into compact form (dimensionality reduction)
  - Does not require label information

PCA

- **Demerits**
  - Does not embed category information
  - Large variations on illuminations or poses (not necessarily the axis with maximum variance contains discriminative features)

  - Not robust to illuminations, expressions, poses etc.

Fisher Faces

- Developed in 1997 by P.Belhumeur et al.
- Unlike PCA, uses class label information
- Works well even if different illumination
- Works well even if different facial expressions

How Fisherfaces Works?

$$y = w^T x \qquad y \in R^m,\ x \in R^n$$

Linear transformation mapping of
n-dim image space to m-dim Feature space

- N Sample Images $\{x_1, x_2, \ldots, x_N\}$
- N dimensional image space
- Each image belongs to one of c classes $\{X_1, X_2, \ldots, X_c\}$

Where $w \in R^{n \times m}$ is a matrix with orthonormal columns (i.e. m < n)

Note

We want to learn projection w that converts all points from x to a New space y.

How Fisherfaces Works?

Now we need an **objective function** to maximize between class scatter and minimize within class scatter

$$J(w) = max \frac{scatter\ between\ class}{scatter\ within\ class}$$

Between Class Scatter: $W^T S_B W$ Within Class Scatter: $W^T S_w W$

$$W_{opt} = argmax_w \left| \frac{W^T S_B W}{W^T S_w W} \right|$$

How Fisherfaces Works?

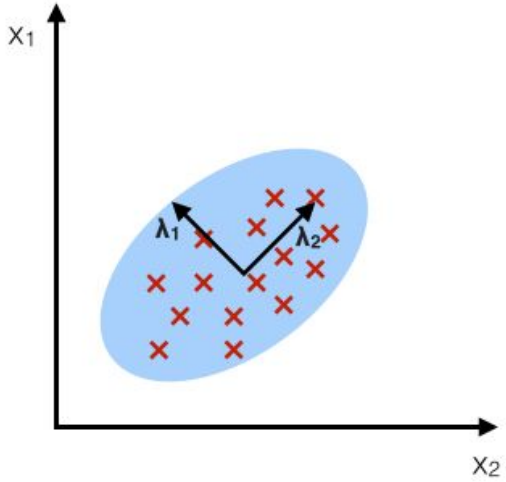
Scatter Matrices for c classes

- Scatter of class i: $$S_i = \sum_{x_k \in Y_i} (x_k - \mu_i)(x_k - \mu_i)^T$$
- Within class scatter: $$S_W = \sum_{i=1}^{c} S_i$$
- Between class scatter: $$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

Eigenface and Fisherface

PCA:
Component axes that Maximize the variance

LDA:
Maximize the component axes for class separation



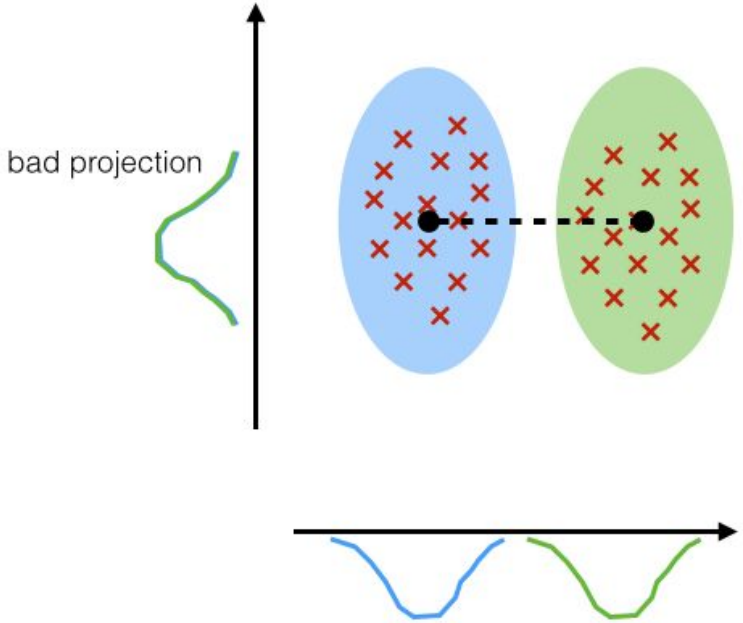
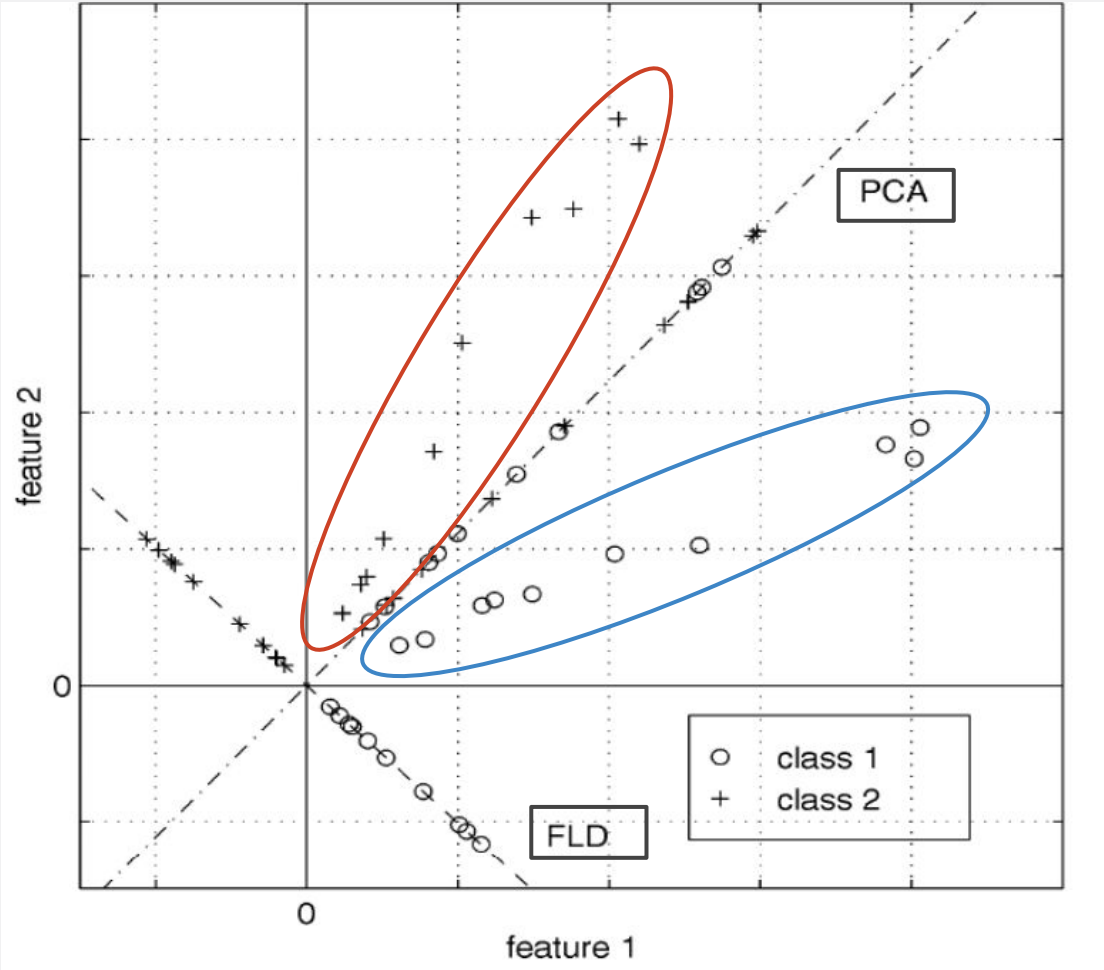


Image Source: https://sebastianraschka.com/Articles/2014_python_lda.html

Eigenface and Fisherface

- Eigenfaces attempt to **maximize the scatter** of the training images in face space.
- LDA seeks directions that are efficient for discrimination between the data
- Fisherfaces attempt to **maximize the between class scatter**, while **minimizing the within class scatter.**
- Moving Images of the same face closer together, while moving images of difference faces further apart.

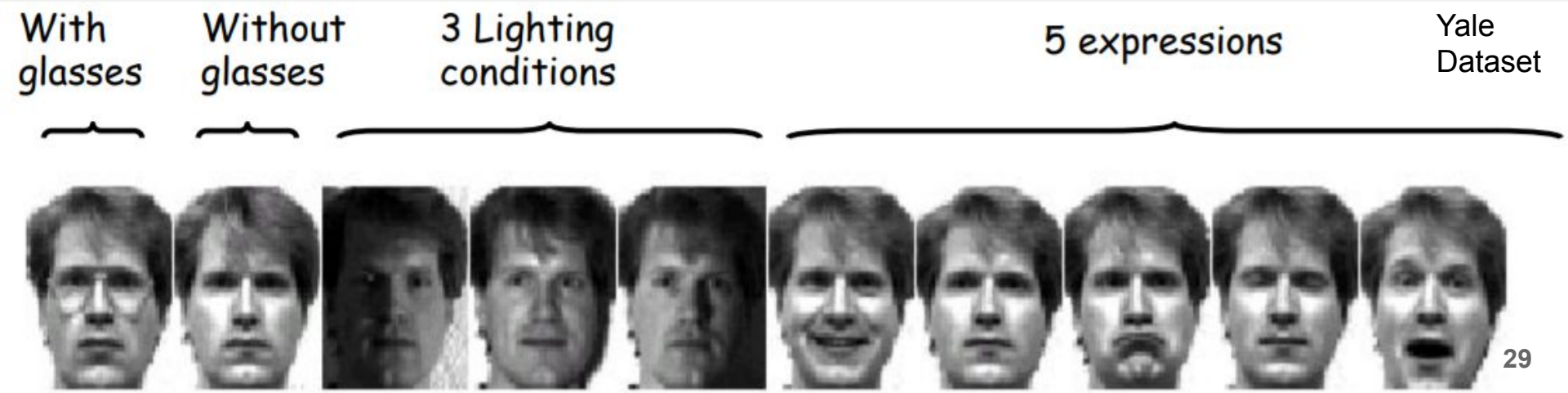
Eigenface and Fisherface



Eigenface Vs Fisherface

Input: 160 images of 16 people

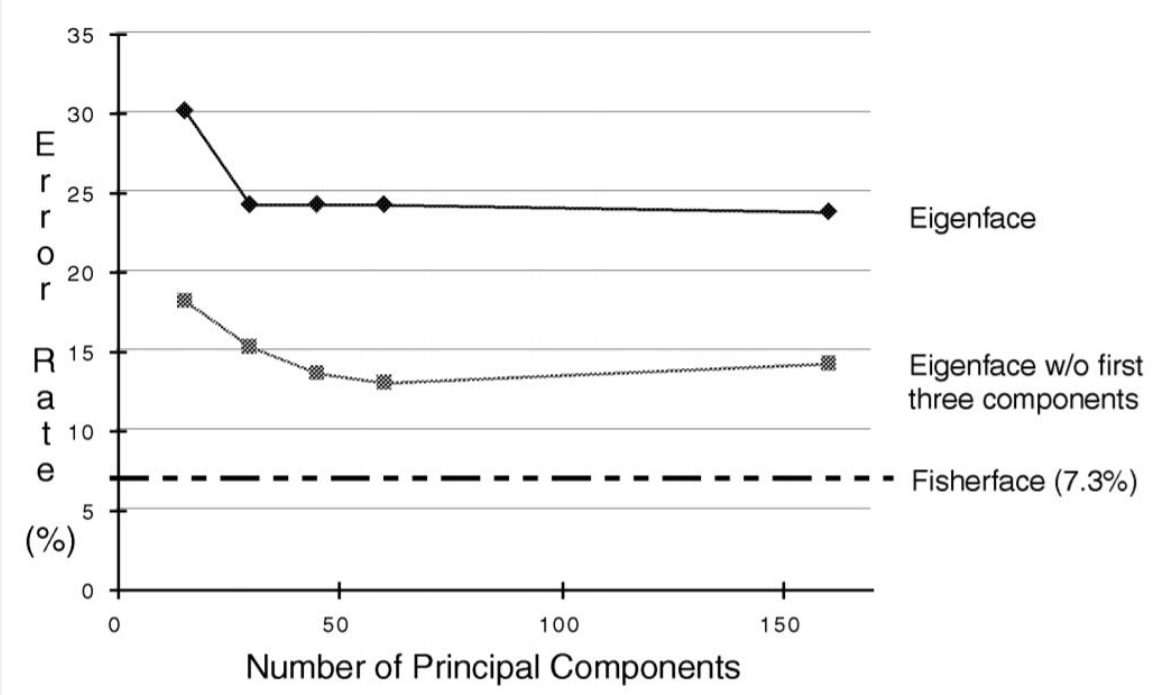
- Train: 159 images
- Test: 1 image
- Variation in Facial Expression, Eyewear, and Lighting



Fisher Face

Yale dataset

Test results demonstrated LDA/FDA is better than eigenface using linear PCA (1997).



Imperial College London

Local Binary Patterns

For finding good descriptor for face

Holistic methods: 1. PCA 2. LDA

Local Region methods:

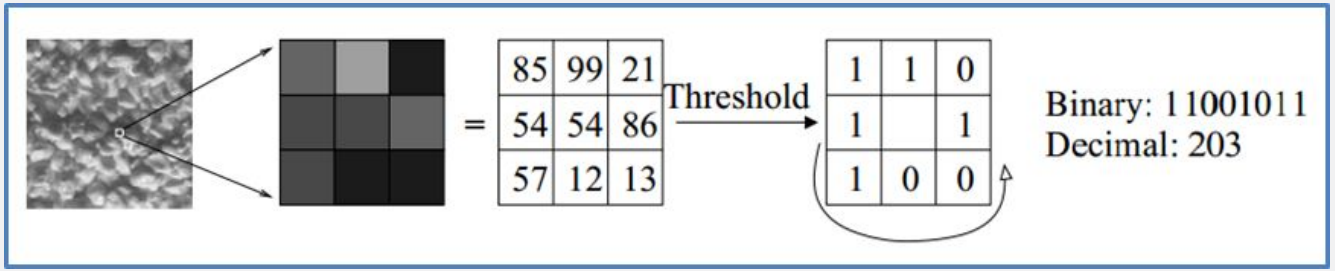
Facial image representation based on LBP texture features from local facial regions

Note

Local descriptors gained attention due to their robustness to challenges such as pose and illumination changes

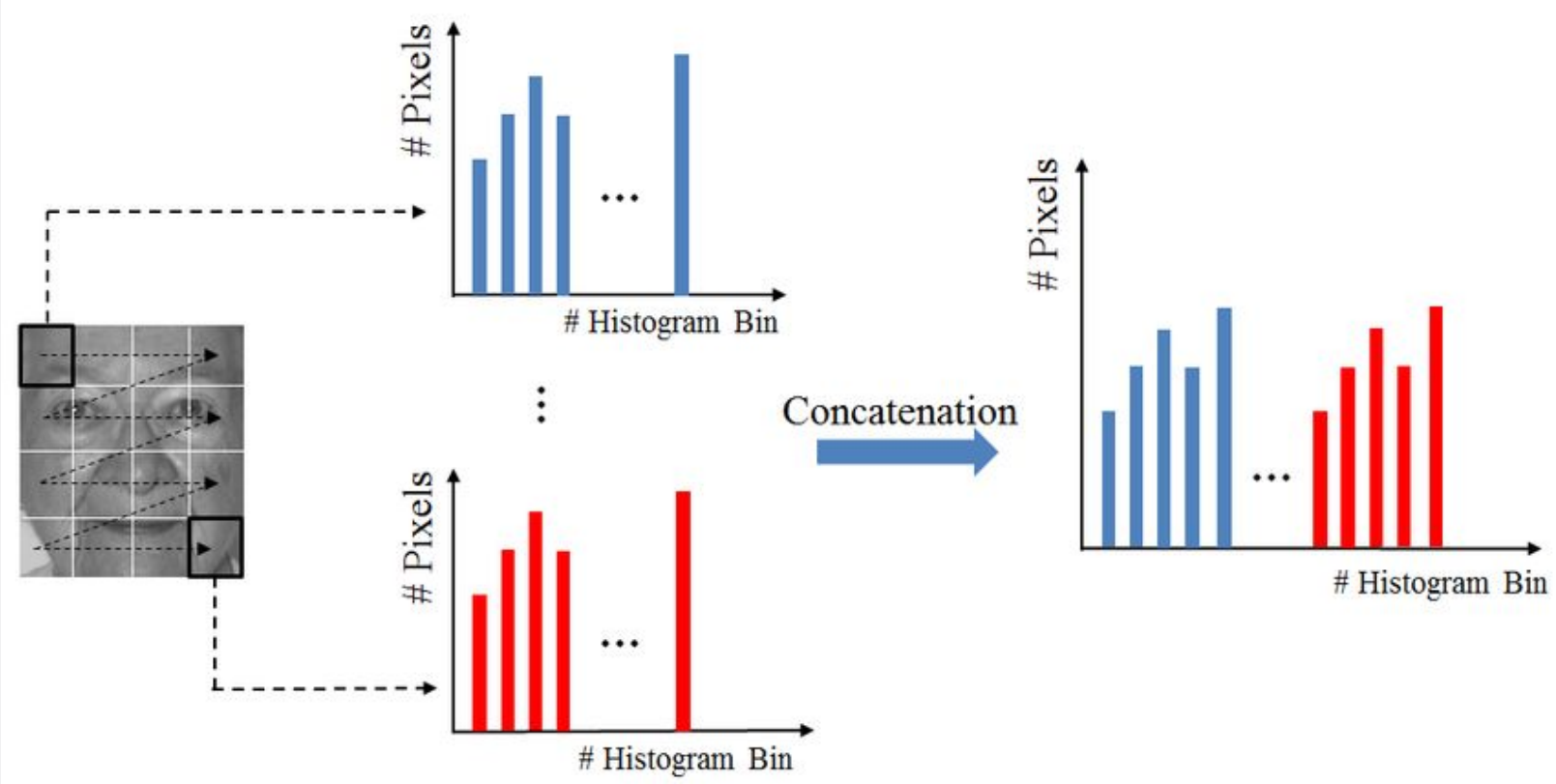
Local Binary Patterns

- One the best performing texture descriptors
- A label is assigned to every pixel
- Use center pixel value to threshold the 3x3 neighborhood
- Result in binary number
- Histogram of the labels is used as a texture descriptor



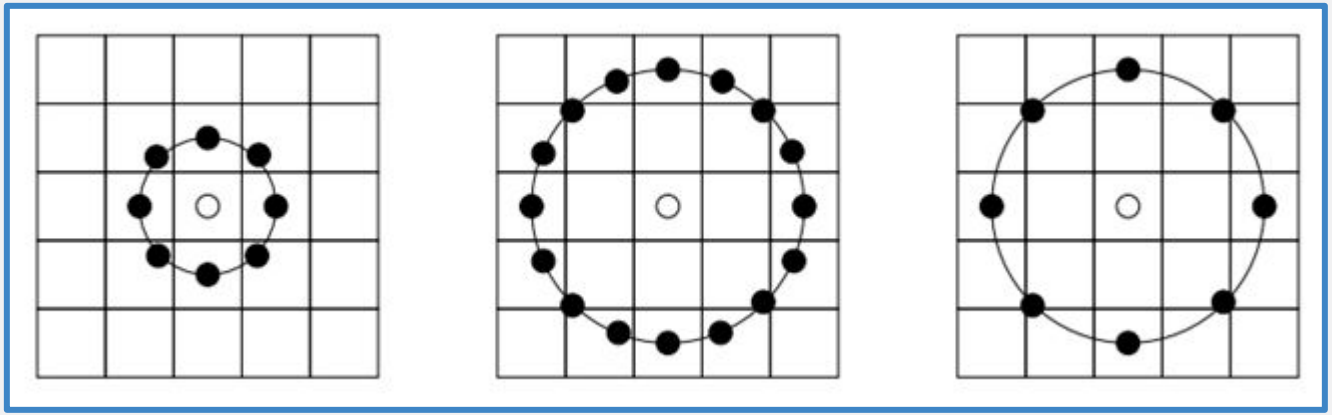
Basic LBP Operator

Linear Binary Patterns



Local Binary Patterns

- LBP is extended to use different sizes of neighborhoods.
- Local neighborhoods is defined as a set of sampling points.
- Points evenly spaced on a circle centered at the labeled pixel.
- **(P,R) , P = number of sampling points , R = radius**



The circular (8,1), (16,2) and (8,2) neighborhoods

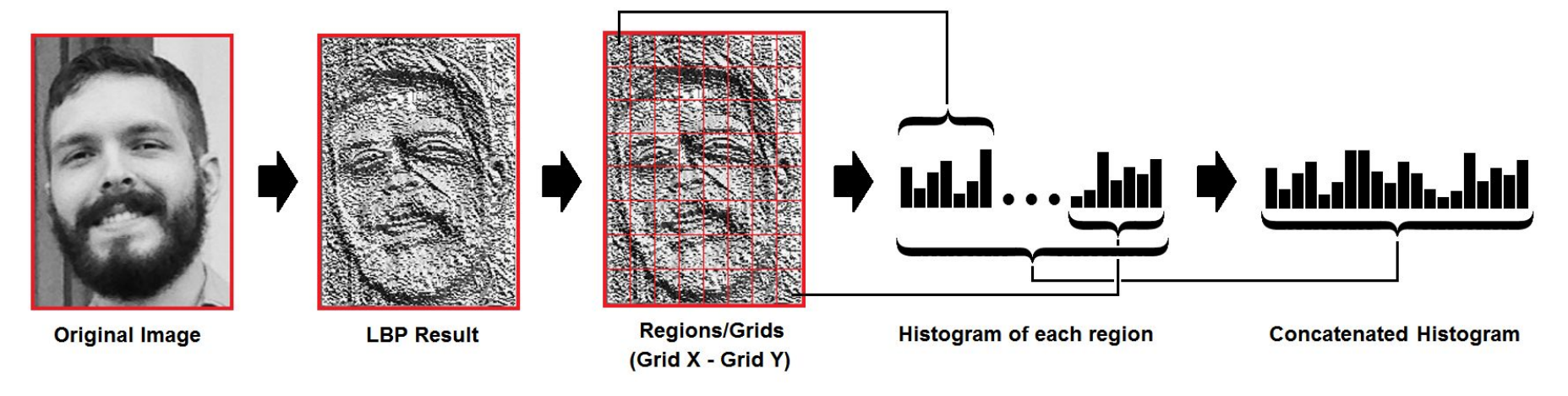
Description of the face on three levels of locality

1. The LBP labels for the histogram contain information about the patterns on a **pixel-level**

2. The labels are summed over a small region to produce information on a **regional level**

3. The regional histograms are concatenated to build a **global description of the face**.

Extracting the Histograms



Face Recognition with LBP

Steps

1. Build Gallery LBP Histograms
2. Build the Probe LBP Histogram

The recognition is performed using a **nearest neighbor classifier** in the computed feature space with **Chi square as a dissimilarity measure.**

LBP advantages over Linear representations

- Non-linear representations
- Illumination invariant
- Age Invariant
- Weighted Regions, better performance

Conclusion of LBP

- Facial images can be seen as micro-patterns(spots, edges, lines, etc.).
- This texture-based facial descriptor is based on dividing image into small regions.
- Texture description of each region can be computed using LBP

- Combining these descriptors into a spatially enhanced histogram or feature vector can be used for face recognition.

OH, WAIT…..Limitations

- They produce rather long histograms, which slow down the recognition speed especially on large-scale face database.
- Under some certain circumstance, they miss the local structure as they don't consider the effect of the center pixel.
- The binary data produced by them are sensitive to noise.

Slide credit: Jakob Verbeek

Metric Learning

- Euclidean or L2 distance is probably the most well known metric

$$d_{L2}(x,y) = (x-y)^T (x-y)$$

  - No parameter to learn
- Most common form of learned metrices are Mahalanobis

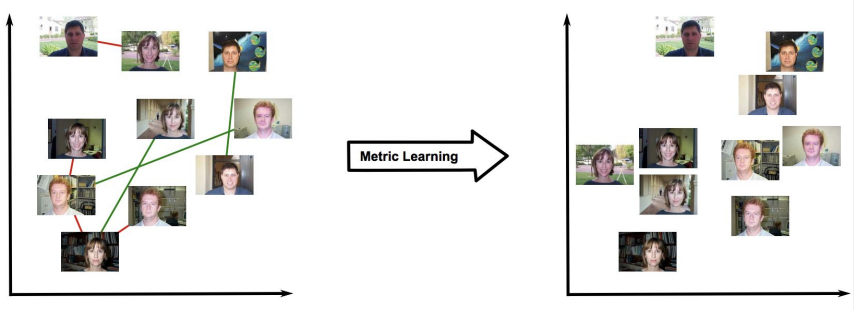
$$d_M(x,y) = (x-y)^T M(x-y)$$

  - $M \in R^{(D \times D)}$ is a semi-definite matrix as it measures distance
  - Generalization of Euclidean metric (setting M=I)
  - M can be decomposed into $L^T L$, $L \in R^{(D \times d)}$ $d<<D$
  - Corresponds to Euclidean metric after linear projection of data

$$d_M(x,y) = (x-y)^T M(x-y) = (x-y)^T L^T L(x-y) = d_{L2}(Lx,Ly)$$

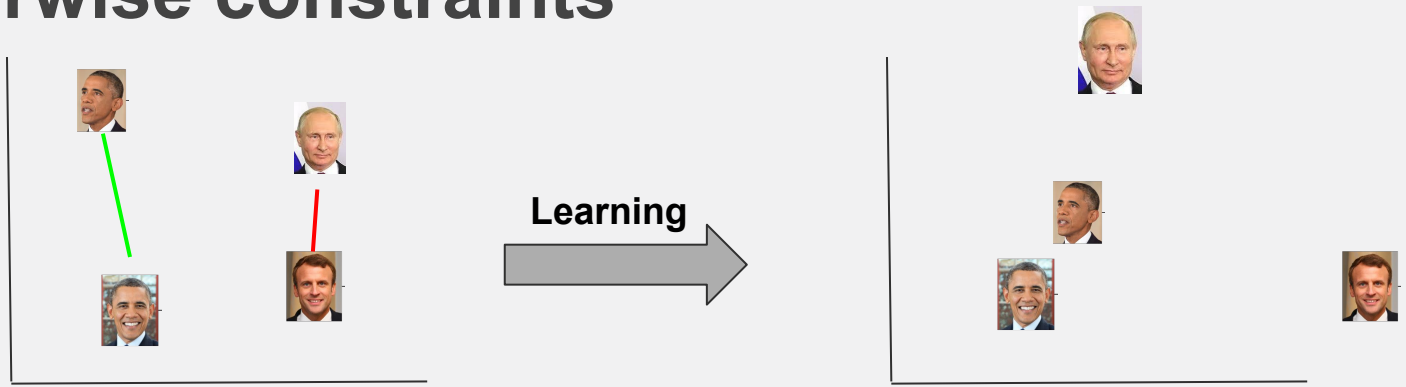
- Reduces the dimension by large-margin

Supervised Metric Learning (ML) for face verification



- Learn a projection matrix where the imposed constraints are **better satisfied**
- Commonly used constraints are: ***pairwise similarity and dissimilarity*** constraints and ***triplet constraints***

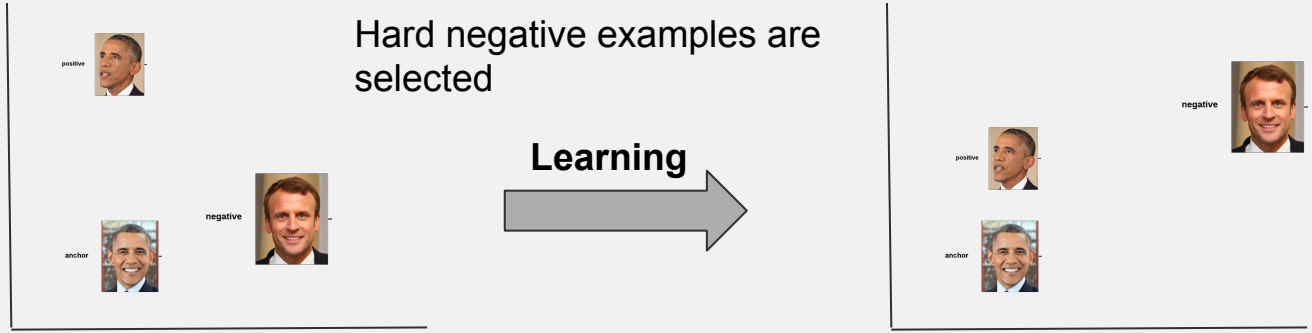
Pairwise constraints



- Must-link / cannot-link constraints (sometimes called positive / negative pairs):

$$\mathcal{S} = \{(x_i, x_j) : x_i \text{ and } x_j \text{ should be similar}\},$$
$$\mathcal{D} = \{(x_i, x_j) : x_i \text{ and } x_j \text{ should be dissimilar}\}.$$

Triplet constraints



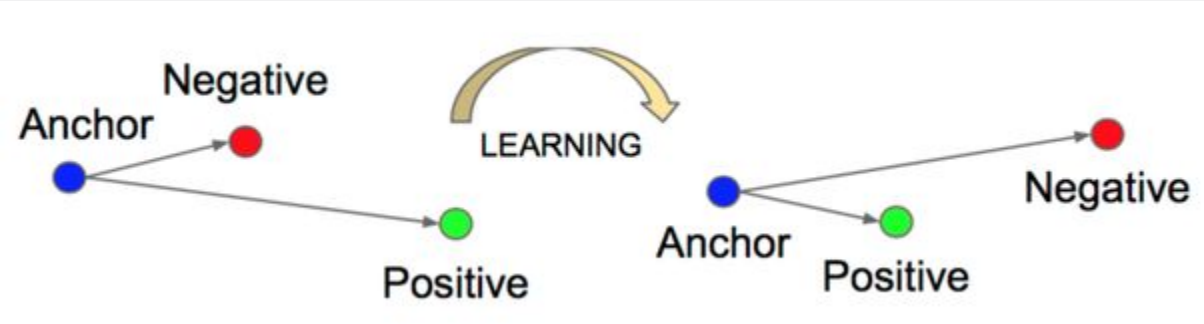
- Relative constraints (sometimes called training triplets):

$$\mathcal{R} = \{(x_i, x_j, x_k) : x_i \text{ should be more similar to } x_j \text{ than to } x_k\}.$$

Triplet Loss

FaceNet: In the FaceNet paper, a convolutional neural network architecture is proposed. For a loss function, FaceNet uses “triplet loss”. Triplet loss relies on minimizing the distance from positive examples, while maximizing the distance from negative examples.

$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$



Pairwise dis(similarity) vs triplet constraints

- Pairwise constraints are easy to collect (Weakly supervised)
- Eg. video frames
- Triplet ,requires (hard) negative examples, adds extra layer of difficulty

Learning projection matrix

- We minimize the max-margin objective function to learn the projection matrix satisfying pairwise dis(similar) constraints (y_{ij} = +/-1)

$$\underset{L}{\operatorname{argmin}} \sum_{t=1}^{t=n} \max\left(m - y_{ij}^{t}(b - d_L^2(x_i^t, x_j^t)),\ 0\right)$$

- Where $d_L^2(x_i, x_j) = ||Lx_i - Lx_j||^2$
- Pushes the examples s.t. Distance of negative pairs is larger by '*m*' than bias '*b*'

Learning the parameters of the projection matrix

- We use stochastic gradient descent

$$\frac{d_L^2(x_i, x_j)}{dL} = L(x_i - x_j)(x_i - xj)^T$$

- Update rule

if $y_{ij}(b - d_L^2(x_i, x_j)) < m$ **then**
  $L \leftarrow L - \eta y_{ij} L(x_i - x_j)(x_i - x_j)^T$
else
  no update
end if

Performance comparison

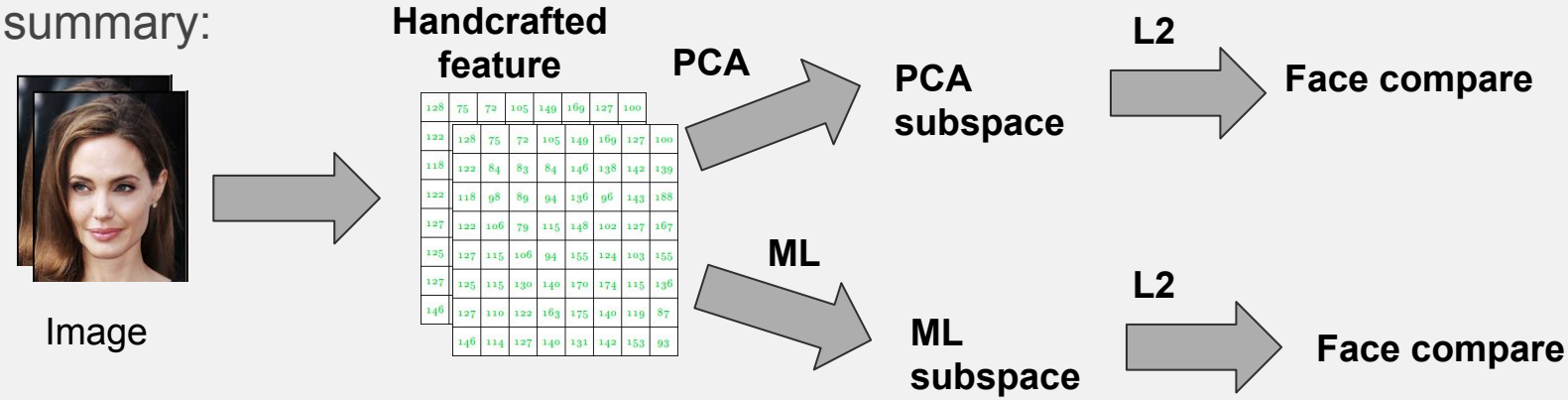
- Database
  - LFW: Labeled Faces in the Wild, contains *13K* of *5K* identities
  - Standard benchmark for face analysis task
- Performed face retrieval task
- Metric used is *1-call@K* (2, 5, 10)

Method	K=2	K=5	K=10	K=20
PCA	30.0	37.4	43.3	51.3
ML	**38.1**	**51.1**	**60.5**	**69.3**

Source: ***Bhattarai et al CVPR 2016***

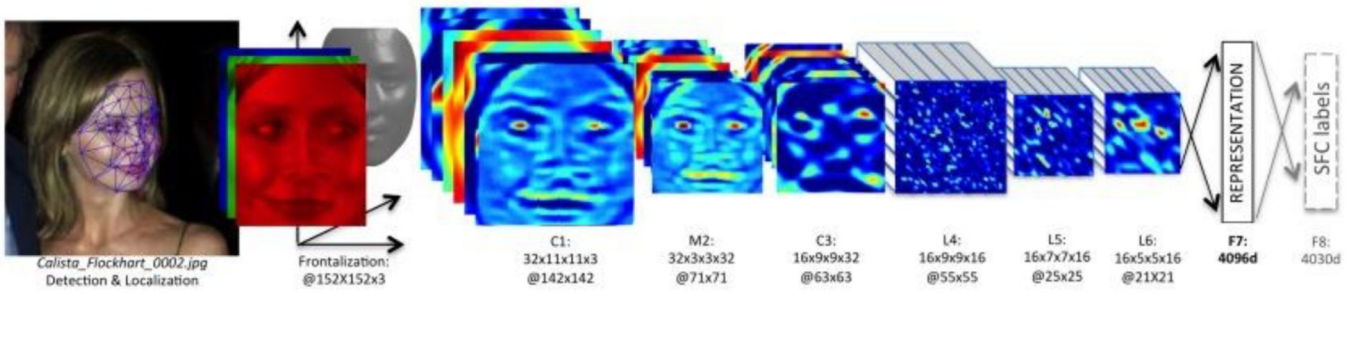
Limitations of parameterized distance

- In summary:



- Multi-stage approach
- Features are not optimal for end task (no feedback mechanism to propagate error to input)

Deepface



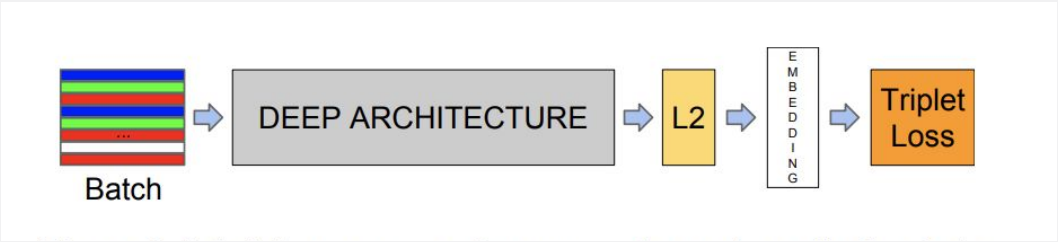
- Train set size: 4M images from 4K identities
- Minimize cross-entropy loss to learn the parameters

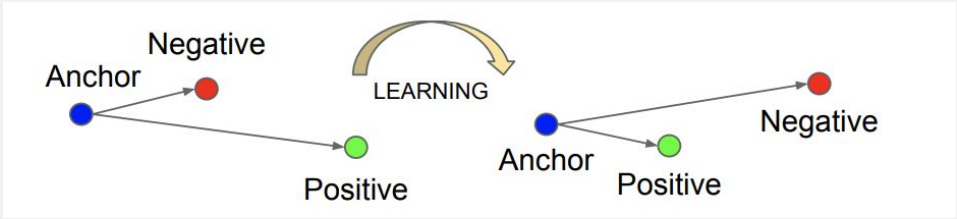
Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." ***CVPR 2014***

Experiments

- Dataset: LFW
- Accuracy: 0.9963 Vs 0.9735 (Deepface) Vs 0.975 (Human)

FaceNet



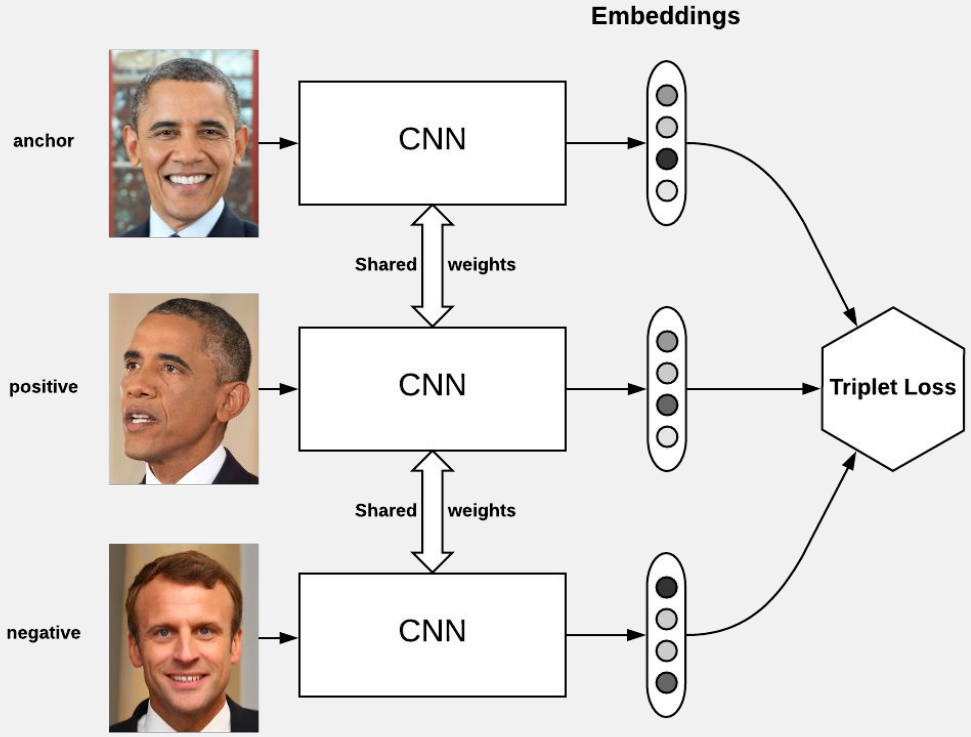


- Objective of this architecture is to minimize L2 distance between same identity's faces representations
- Directly transforms image representations at a low dimensional feature space (128D vs 4096D (Deepface)) rather than bottleneck intermediate representations

Schroff, Florian, Dmitry Kalenichenko, and James Philbin. **"Facenet: A unified embedding for face recognition and clustering."** *CVPR 2015*

APA

Triplet Loss



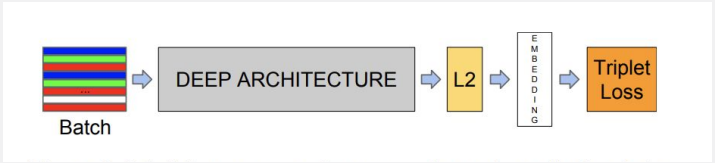
A, P, N

d(A, P) + alpha <= a(A, N)

Choose triplets that are hard to train on

d(A, P) close to d(A, N)

FaceNet



- Uses triplet loss
- Minimize the max-margin objective

$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

s.t.

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 ,$$

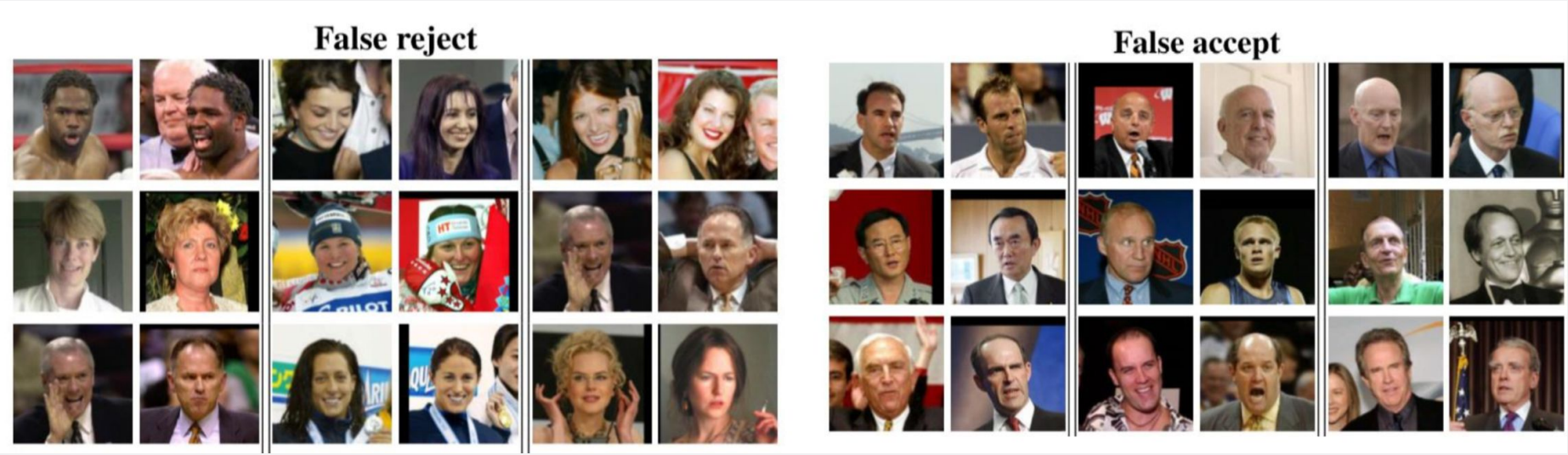
$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} .$$

- This ensures all positive examples are nearer than negative examples
- Very useful for clustering of faces

Schroff, Florian, Dmitry Kalenichenko, and James Philbin. **"Facenet: A unified embedding for face recognition and clustering."** *CVPR 2015*

APA

Mis-classified examples

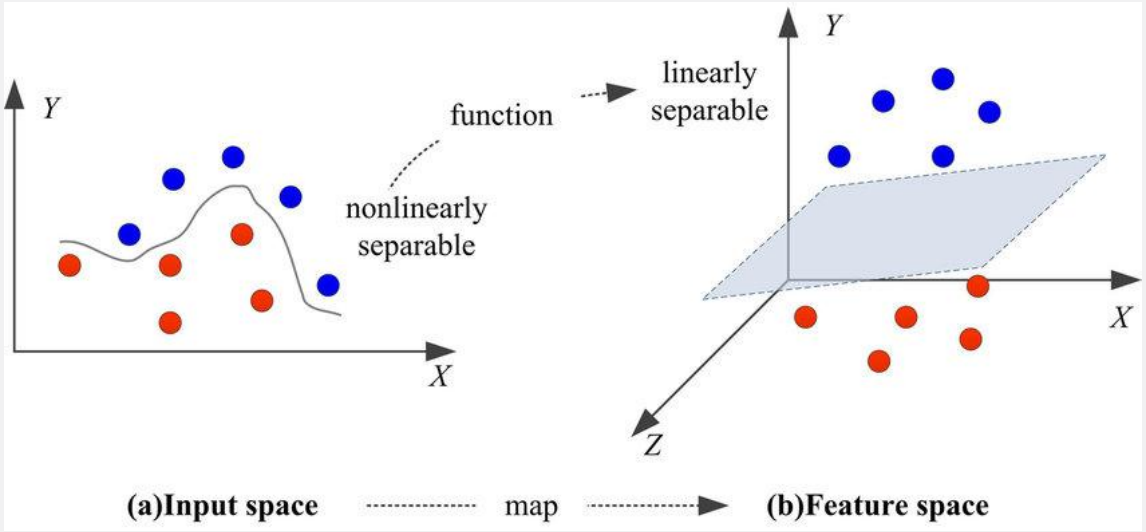


Bottlenecks in face analysis using deep learning

- Computing resource
- Data hungry

Additional Slides

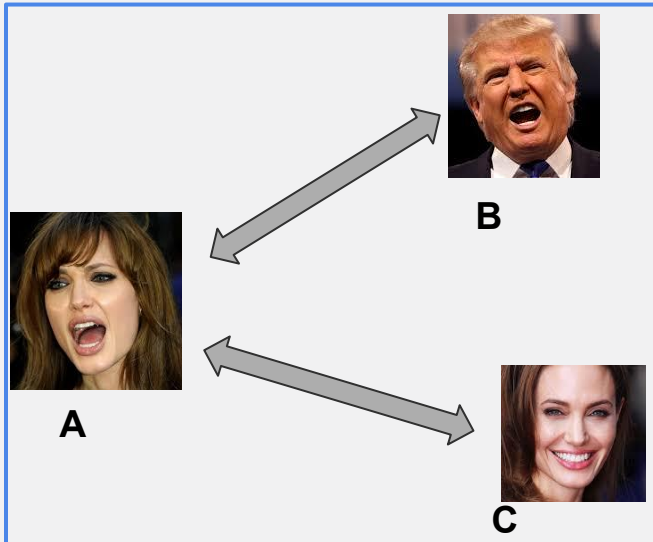
Feature Representations



Why Face Recognition is a Nonlinear Problem?

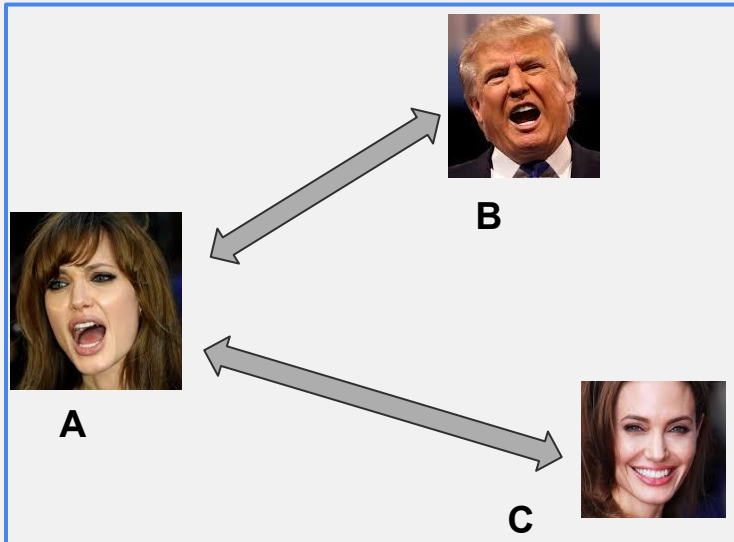
- **Pose (Out-of-Plane Rotation):** frontal, 45 degree, profile, upside down
- **Presence or absence of structural components:** beards, mustaches, and glasses
- **Facial expression:** face appearance is directly affected by a person's facial expression
- **Occlusion:** faces may be partially occluded by other objects
- **Orientation (In-Plane Rotation):** face appearance directly vary for different rotations about the camera's optical axis
- **Imaging conditions:** lighting (spectra, source distribution and intensity) and camera characteristics (sensor response, gain control, lenses), resolution

Similarity between the Faces



- **Which pair is more similar?**
  - (A, B)
  - (A, C)

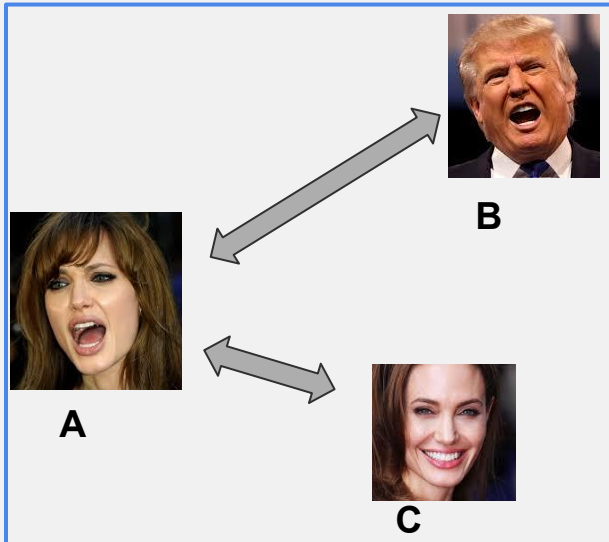
Similarity between the Faces



- When we care about expression

$$d_{(mood)}(A, B) < d_{mood}(A, C)$$

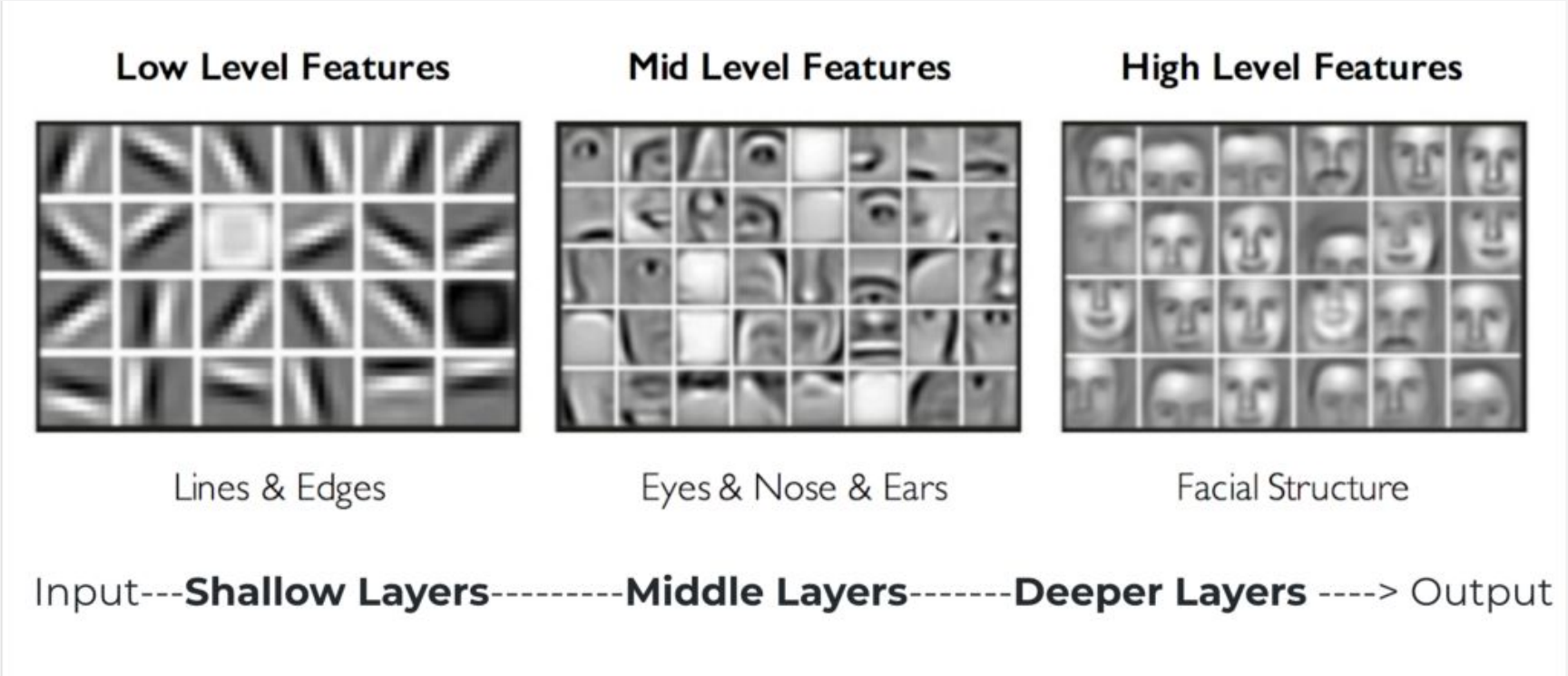
Similarity between the Faces



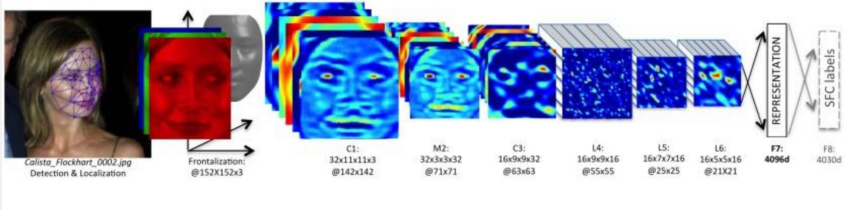
- When we care about identity

$$d_{(id)}(A, B) > d_{(id)}(A, C)$$

- Similarity between the images depends on the **the task we care about**
- **Need a methodology to learn such metric**


Low Level Features
Mid Level Features
High Level Features
Lines & Edges
Eyes & Nose & Ears
Facial Structure
Input---Shallow Layers---------Middle Layers-------Deeper Layers ----> Output

Deepface



- F8 calculates probability with softmax $p_k = \exp(o_k) / \sum_h \exp(o_h)$
- Cross-entropy loss function: $L = -\sum_k \log(p_k)$
- Computed using SGD and performs backpropagation

Experiments

- DeepFace was evaluated in LFW
- Human cropped: 97.5% vs Deepface 97.35%

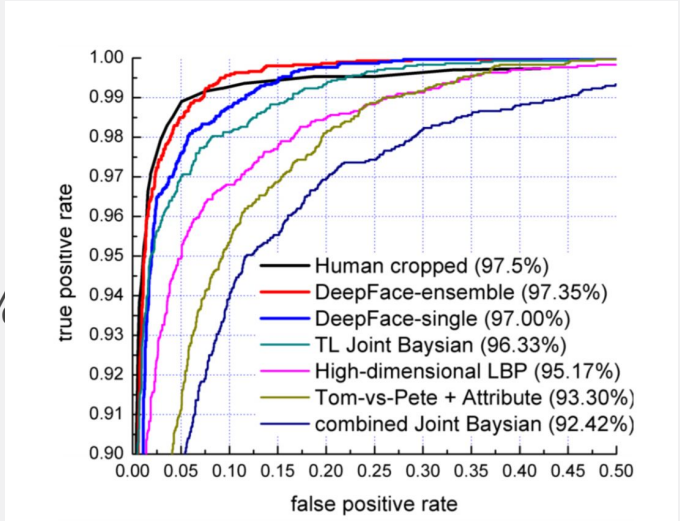
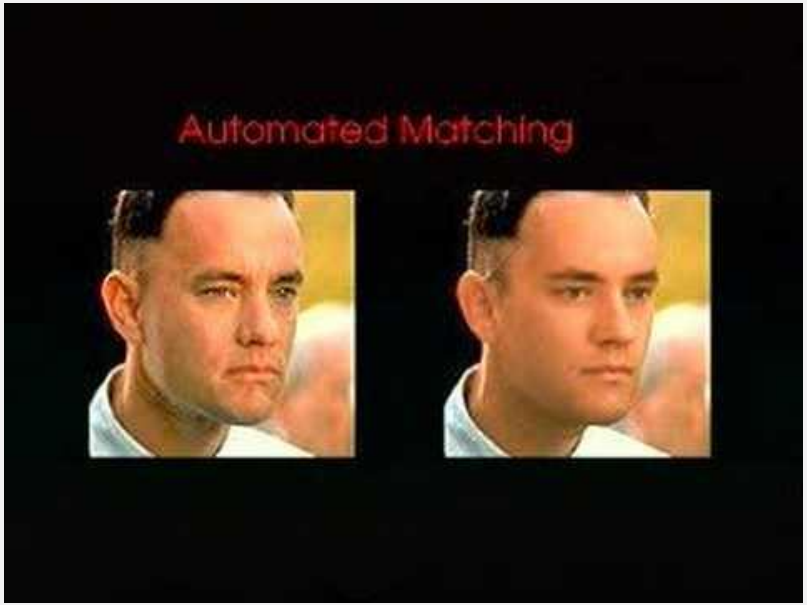


Fig. Roc Curve on LFW

3D Morphable Model



Synthetic Face Images by 3DMM

- + Easy to manipulate attributes such as identity, pose, expression, and lighting
- + Can generate millions of images with controlled attributes
- - Domain gap with real face images

Synthetic images generated by 3DMM

Problem Definition

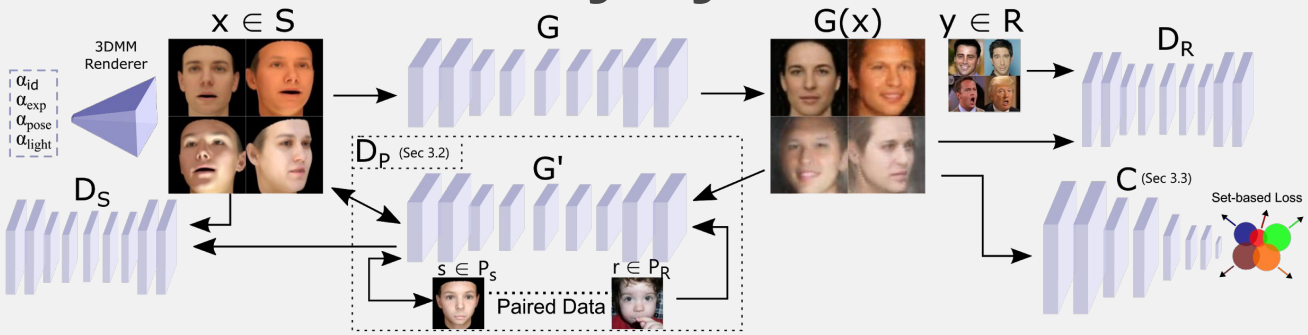
- Generating photorealistic face images 3DMM rendered faces of new identities with arbitrary poses, expressions, and illuminations

Problem Definition

- Generating photorealistic face images 3DMM rendered faces of new identities with arbitrary poses, expressions, and illuminations
- We formulate this problem as domain adaptation problem/ style transfer problem (3DMM -> Real)
  - *Pixel2Pixel (Isola et al 2017)*
  - *CycleGAN(Zhu et al 2017)*

- How can we benefit small amount of paired data in unsupervised style transfer GAN?
- How to prove identity consistency of generated images?

Photorealistic identity synthesis (Gecer, Bhattarai, Kittler, and Kim *ECCV' 2018*)



- Randomly generated 3DMM images with random pose, expression and lighting attributes for the new IDs.
- Unsupervised training with forward cycle consistency.
- Adversarial Pair Matching network G' by the help of a limited number of paired data.
- ID preservation by a set-based supervision through a pre-trained classification network C.

Experiments (Quantitative)

VGG(%100)	1.8M	-	96×96	94.8
VGG(%100) + GANFaces-500K	1.8M	500K	96×96	**94.9**
VGG(%100) + GANFaces-5M	1.8M	5M	96×96	**95.2**

Tab. Verification accuracy on LFW benchmark

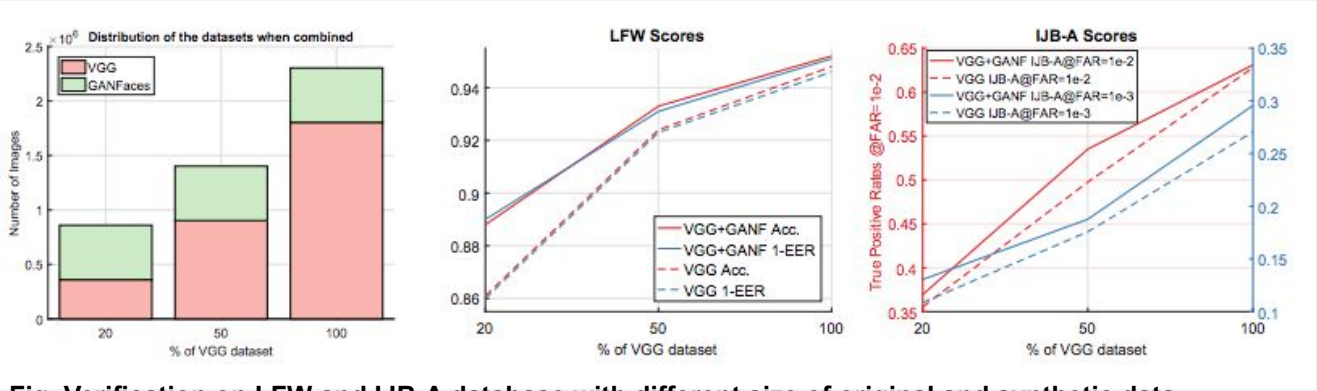


Fig. Verification on LFW and IJB-A database with different size of original and synthetic data

Experiments (Qualitative)

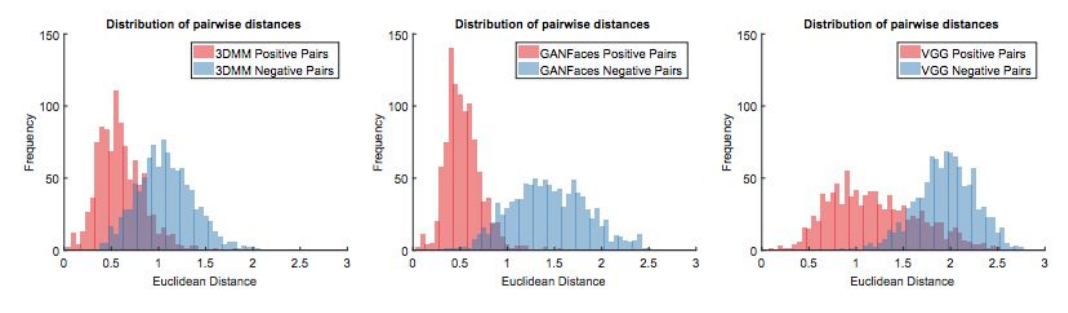
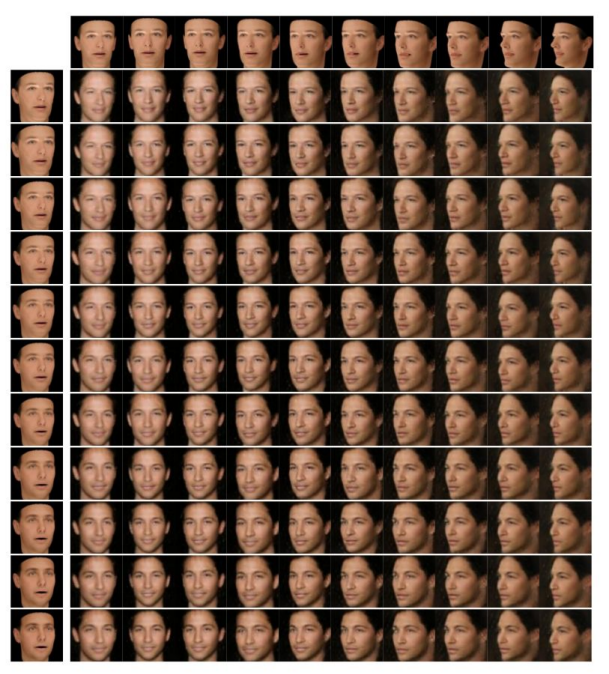


Fig. Face pairs euclidean distance distribution

Experiments (Qualitative)

Fig. Face generated by the proposed method conditioned on 3DMM identity, expression and pose parameters

Experiments (Qualitative)



- Interpolation in identity space
- Smooth transition from one identity space to another identity space shows that manifold of image generator is smooth.

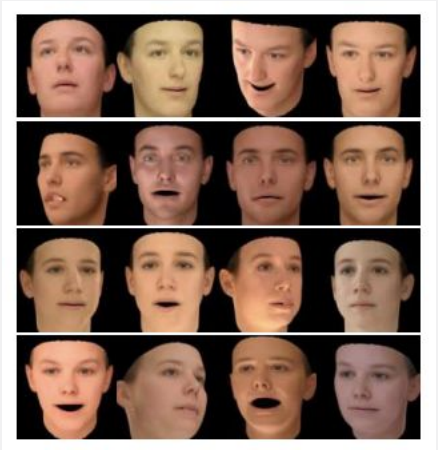
Experiments (illumination preservation)

Experiments(Qualitative)

- The nearest images from the training set in terms of identity features for the images
- Variation in the nearest images shows diversity of GANFaces in the embedding space while they bear similar higher order attributes such as gender, shape of face etc.

Comparison with existing methods

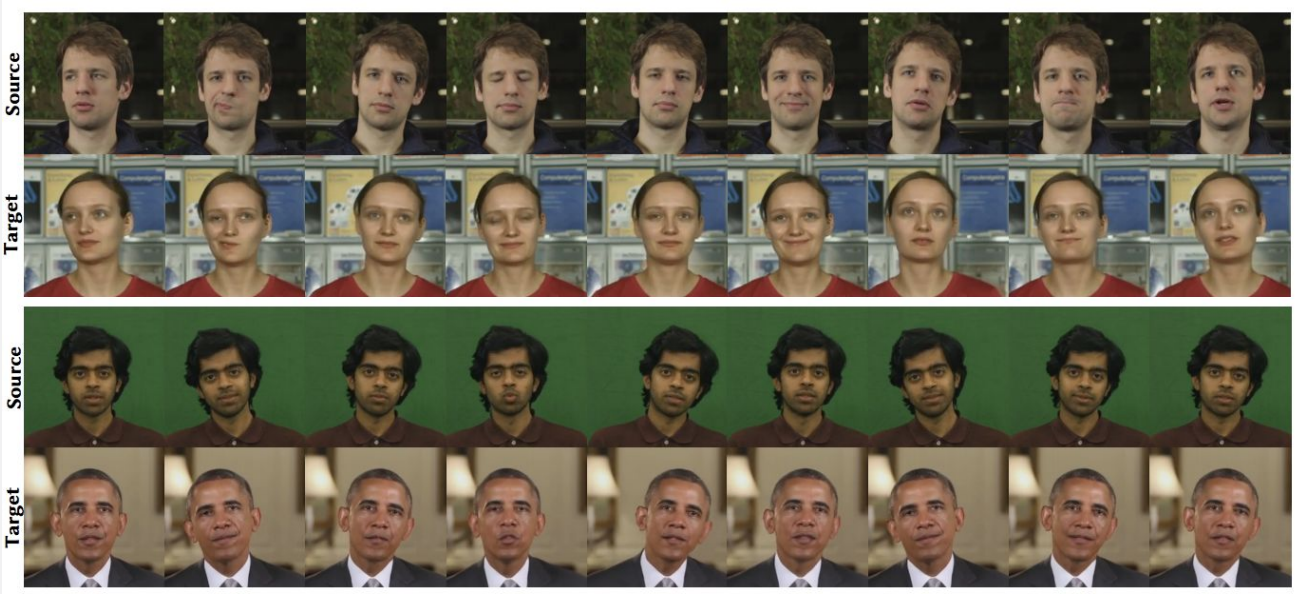
A) 3DMM



B) CycleGAN

C) Our approach

Qualitative results



Properties of pseudo-distance

1. $d_M(\boldsymbol{x}, \boldsymbol{x}') \geq 0$ (nonnegativity),
2. $d_M(\boldsymbol{x}, \boldsymbol{x}) = 0$ (identity),
3. $d_M(\boldsymbol{x}, \boldsymbol{x}') = d(\boldsymbol{x}', \boldsymbol{x})$ (symmetry),
4. $d_M(\boldsymbol{x}, \boldsymbol{x}'') \leq d(\boldsymbol{x}, \boldsymbol{x}') + d(\boldsymbol{x}', \boldsymbol{x}'')$ (triangle inequality).

References

1. Ojala, Timo, Matti Pietikäinen, and Topi Mäenpää. "Gray scale and rotation invariant texture classification with local binary patterns." *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2000.
2. Ahonen, Timo, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12 (2006): 2037-2041.