



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
FACULTY OF ENGINEERING

# Computer Vision Project – Winter Term 2024/25

## Face Recognition

Dr.-Ing. Thomas Köhler

Pattern Recognition Lab, Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg

e.solutions GmbH, Erlangen

December 9, 2024



# Face Recognition Example Use Cases

Unlock smart phones via face recognition<sup>1</sup>



<sup>1</sup> <https://www.apple.com/iphone-x/#face-id>

# Face Recognition Example Use Cases

Identification of occupants and personalization of vehicle settings



## Scope of this Project

### In the lecture:

- Learn how to design and evaluate the core of current facial recognition systems from a technical point of view
- Overview of modern machine learning methods in this field
- Discussion of their strengths and limitations

### In the exercise:

- Implementation of a simple system comprising basic functionality for face verification, identification, and clustering
- Evaluation of face recognition algorithms
- Design own recognition method and participate in a challenge

## Organization

- Exercises will be held virtually (start: CW 51/2024)
  - We will propose a couple of available time slots at the beginning of the week
  - Please drop me an e-mail with your availability
  - We will send you an invitation
  - **Be well prepared (only 15 - 30 minutes per group/week)!**
- Final submission via StudOn
  - Upload your source codes to StudOn **before the submission deadline!**
  - We will schedule a meeting to review your final solution
  - You will present your solution and we will ask questions
  - All exercise partners shall contribute equally to all tasks and shall be able to answer questions on each of the exercises!
  - **No certificate without presentation, review, and answering questions!**
- Additional exercise
  - Required to gain 10 ECTS
  - Will be a challenge with winners (details to be announced) :-)

# Outline

## Introduction

## Face Representation

- Eigenfaces

- Fisherfaces

- Deep Features

## Selected Topics in Face Recognition

- Distance Measures and Face Verification

- Face Identification

- Face Clustering

- Evaluating Face Recognition Systems

## Summary



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
FACULTY OF ENGINEERING

# Introduction



# The Face as a Biometric Marker

Biometric markers for identification<sup>2</sup>:

- Fingerprint
- Iris
- Speech
- Face

Face vs. other markers:

- Face the only marker that can be captured at large distances
- Can work with low-cost hardware

---

<sup>2</sup>Jain, Anil K., and Stan Z. Li. Handbook of face recognition. New York, Springer, 2011.



# Challenges

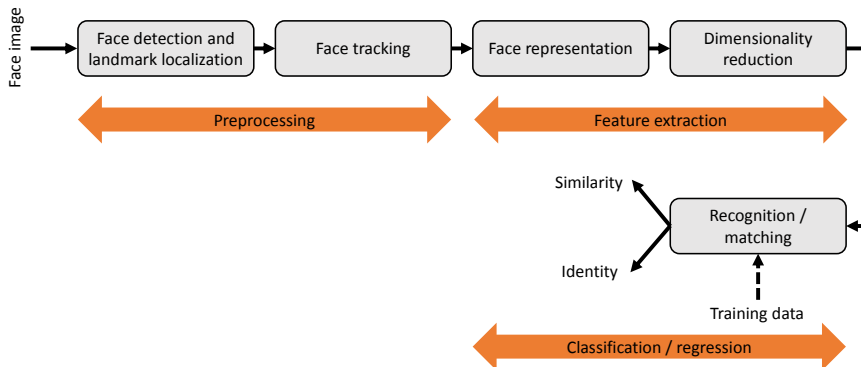
## Intrinsic conditions:

- High intra-subject variation: expressions, changes in facial hair or pose, aging
- Low inter-subject variation: similar skin or hair color, similar eye glasses

## Extrinsic conditions:

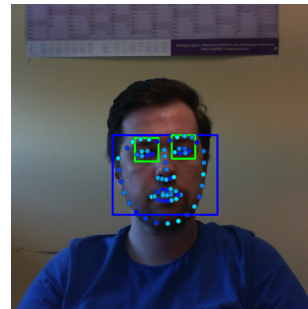
- Varying illumination: images captured at day and at night
- Image quality: facial data processed with different image or video codecs

# Face Recognition Pipeline



# Face Detection, Landmark Localization, and Tracking

- Detection of face region
- Detection of eye regions
- Extracting landmarks to model pose and facial expression
- Alignment using detected landmarks
  - Compensate for different head poses
  - Eyes, nose, and mouth at predefined positions
- Track bounding box and landmarks in video data



# Face Representation

## Geometric approach:

- Distances, areas, or angles between salient points (eyes, nose, mouth)
- Obtained by feature detection algorithms
- Requires hand-crafted features



## Data-driven approach:

- Image intensities form raw features
- Learn suitable representation from exemplars
- Requires training from large datasets


$$\begin{pmatrix} 124 \\ 122 \\ 116 \\ 130 \\ 132 \\ \dots \end{pmatrix}$$

# Face Recognition Problem Statements

## Verification (one-to-one matching):

- Given one probe image and one gallery image
- Check if both images show the same identity

## Identification (one-to-many matching):

- Given one probe image and a set of gallery images with identity labels
- Retrieve identity of probe image

## Clustering (many-to-many matching):

- Given a set of unlabeled face images
- Cluster according to the identities captured in the data



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
FACULTY OF ENGINEERING

# Face Representation



# Data-Driven Face Representations

How to represent faces in digital images?

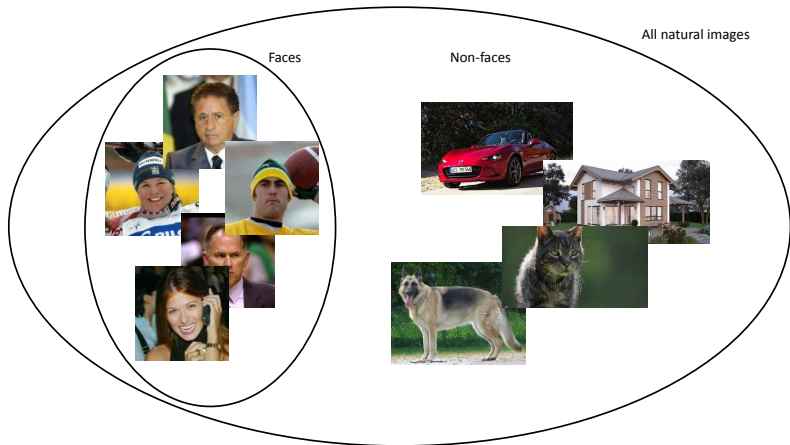


$$\begin{pmatrix} 124 \\ 122 \\ 116 \\ 130 \\ 132 \\ \dots \end{pmatrix}$$

- Grayscale image with  $M \times N$  pixels and  $b$  bits per pixel (e. g.  $64 \times 64$ , 8 bit)  
→  $(2^b)^{M \cdot N}$  different images (e. g.  $256^{4096} \gg 10^{1000}$ )
- Only a small fraction of images corresponds to valid faces

# Manifold of Face Images

Faces are a small subset of the natural image manifold





## Learning Face Representations: Eigenfaces

- Let  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}_i \in \mathbb{R}^D$  be a set of  $n$  face images with mean:

$$\mu_x = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad (1)$$

and covariance:

$$\Sigma_x = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \mu_x)(\mathbf{x}_i - \mu_x)^\top \quad (2)$$

- Project each  $\mathbf{x}_i$  to a  $M$ -dimensional space ( $M \ll D$ ):

$$\mathbf{z}_i = \mathbf{W} \mathbf{x}_i \in \mathbb{R}^M \quad (3)$$

- Seek linear transform  $\mathbf{W} \in \mathbb{R}^{M \times D}$  that maximizes the variance of  $\mathbf{z}_1, \dots, \mathbf{z}_n$ :

$$\begin{aligned} \Sigma_z &= \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \mu_z)(\mathbf{y}_i - \mu_z)^\top \\ &= \mathbf{W}^\top \Sigma_x \mathbf{W} \end{aligned} \quad (4)$$

→ Principal Component Analysis (principal components  $\equiv$  Eigenfaces)

## Learning Face Representations: Fisherfaces

- Exploit two constraints using class labels (one unique face  $\equiv$  one class):
  - Between-class scatter shall be maximum
  - Within-class scatter shall be minimum
- Seek the transform  $\mathbf{W} \in \mathbb{R}^{M \times D}$  maximizing Fisher's linear discriminant:

$$J(\mathbf{W}) = \frac{\mathbf{W}^\top \Sigma_{\text{inter}} \mathbf{W}}{\mathbf{W}^\top \Sigma_{\text{intra}} \mathbf{W}} \quad (5)$$

Intra-class and inter-class scatter for  $c$  classes:

$$\Sigma_{\text{intra}} = \sum_{i=1}^c \frac{1}{|\mathcal{X}_i|} \sum_{\mathbf{x}_k \in \mathcal{X}_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^\top \quad (6)$$

$$\Sigma_{\text{inter}} = \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu)^\top \quad (7)$$

$\mu_i$  is the mean of all faces  $\mathcal{X}_i$  in the  $i$ -th class and  $\mu$  is the overall mean face

# Learning Face Representations: Deep Features

Limitations of the methods discussed so far:

- Determine face representations under a linear model
- Limited robustness against pose or illumination variations

Extension:

- Use a non-linear face representation  $f_{\theta}(\mathbf{x})$ :

$$f_{\theta}(\mathbf{x}) \tag{8}$$

- The transform  $f_{\theta}(\mathbf{x})$  is implemented by a deep neural network
- Provides a face representation (aka. embedding)
- Parameters  $\theta$  are learned from exemplars

## Recap: Convolutional Neural Networks (CNNs)

Neural network:

- Computation graphs comprising neurons
- Propagation of input signals through the network to obtain outputs

Network design with input, output, and hidden layers:

- Convolutional layer: convolution of input neuron activation with filter kernel
- Pooling layer: fusion of clusters of input neuron activation
- Locally/fully connected layer: weighted sum of input neuron activation

$\theta$ : Parameters of all layers in the network

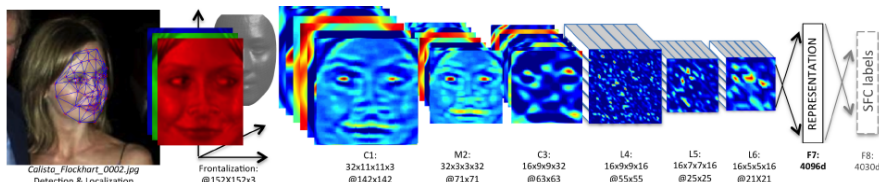
## Face Representation via Classification

- Design neural network to classify face images according to their identity to one out of  $c \geq 2$  classes
- Fully connected output layer ( $\mathbf{W}, \mathbf{b}$ ) to model probability distribution:

$$p_i = \frac{\exp(\mathbf{w}_i^\top \mathbf{x} + b_i)}{\sum_{j=1}^c \exp(\mathbf{w}_j^\top \mathbf{x} + b_j)}, \quad i = 1, \dots, c \quad (9)$$

- Train network parameters on face exemplars by minimizing misclassification error (e.g. cross entropy loss)
- Activation of fully connected layer  $\equiv$  embedding

## Example: Facebooks DeepFace (trained on 4M faces)<sup>3</sup>



- Face frontalization for preprocessing
- Eight layer network: convolutional (C1, C3), max-pooling (M2), locally connected (L4, L4, L6), and fully connected layers (F7, F8)
- Feature maps describe face from high-level to low-level
- Activation of F7 fully connected layer used as face representation

<sup>3</sup>Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2014.

# Face Representation via Contrastive Learning

Discussion of the classification-based approach:

- Classification requires class labels (identities)
- Models discriminative features for face recognition only implicitly
- Learned features are not necessarily optimal

Extension to regression-based / contrastive losses:

- Can be used for pre-training or fine-tuning
- No full labeling of dataset required

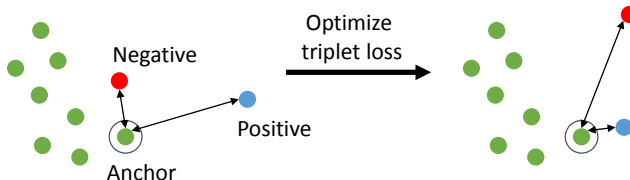
## Triplet Loss

- Let  $\mathbf{x}_i^a$  be an anchor face,  $\mathbf{x}_i^p$  a face of the same subject (positive) and  $\mathbf{x}_i^n$  be a face of different subject (negative)
- Ensure for all triplets  $(\mathbf{x}_i^a, \mathbf{x}_i^p, \mathbf{x}_i^n)$  with margin  $\alpha$ :

$$\underbrace{\|f_\theta(\mathbf{x}_i^a) - f_\theta(\mathbf{x}_i^p)\|_2^2}_{\text{anchor-to-positive distance}} + \alpha < \underbrace{\|f_\theta(\mathbf{x}_i^a) - f_\theta(\mathbf{x}_i^n)\|_2^2}_{\text{anchor-to-negative distance}} \quad (10)$$

- Penalize distances of positive and negative samples w.r.t. the anchor

$$\mathcal{L}_{\text{triplet}}(\theta) = \sum_{i=1}^n \|f_\theta(\mathbf{x}_i^a) - f_\theta(\mathbf{x}_i^p)\|_2^2 - \|f_\theta(\mathbf{x}_i^a) - f_\theta(\mathbf{x}_i^n)\|_2^2 + \alpha \quad (11)$$





## Practical Considerations on Deep Features

### Learning:

- Learn  $f_{\theta}(\mathbf{x})$  on large datasets (Facebook: 4M faces, Google: 200M faces)
- Computationally very demanding
- Implemented on graphics processing units (GPU)

### Inference:

- Determine face representation  $f_{\theta}(\mathbf{x})$  by forward pass
- Efficient to compute using additional optimizations (weight quantization)
- Use face representation within lightweight classification or clustering models
- Can be implemented on embedded devices



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
FACULTY OF ENGINEERING

# Selected Topics in Face Recognition



## Face Verification

- Given two face images  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , is the image pair of the same person?
- Define a suitable distance measure  $d(\mathbf{x}_1, \mathbf{x}_2)$ , e.g., cosine distance:

$$d(\mathbf{x}_1, \mathbf{x}_2) = \frac{\|f_\theta(\mathbf{x}_1) - f_\theta(\mathbf{x}_2)\|_2^2}{\|f_\theta(\mathbf{x}_1)\|_2^2 + \|f_\theta(\mathbf{x}_2)\|_2^2} \quad (12)$$

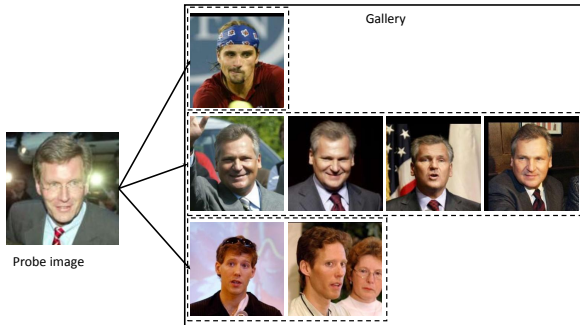
- Verification using the distance measure and threshold  $\tau$ :

$$v(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} \text{same} & \text{if } d(\mathbf{x}_1, \mathbf{x}_2) \leq \tau \\ \text{not same} & \text{if } d(\mathbf{x}_1, \mathbf{x}_2) > \tau \end{cases} \quad (13)$$

- Alternatively, use similarity measure  $s(\mathbf{x}_1, \mathbf{x}_2) = -d(\mathbf{x}_1, \mathbf{x}_2)$

## Face Identification

- Given a probe image  $\mathbf{x}_{\text{probe}}$  of unknown identity
- Determine identity  $I(\mathbf{x}_{\text{probe}})$  from labeled images in a gallery



### Two protocols:

- Closed-set: all possible identities of probes are contained in the gallery
- Open-set: identities of some probes are missing in the gallery

## Face Identification with Closed-Set Protocol

Repeated face verification using k-nearest neighbors (k-NN) classifier:

- For the probe image  $\mathbf{x}_{\text{probe}}$ , find the  $k$  closest gallery images  $\mathbf{x}_1, \dots, \mathbf{x}_k$  according to a distance  $d(\mathbf{x}_{\text{probe}}, \cdot)$
- Identity of  $\mathbf{x}_{\text{probe}}$  is the majority (mode) in  $\mathbf{x}_1, \dots, \mathbf{x}_k$ :

$$l(\mathbf{x}_{\text{probe}}) = \text{mode}(l(\mathbf{x}_1), \dots, l(\mathbf{x}_k)) \quad (14)$$

- Alternatively, we can consider similarities  $s(\mathbf{x}_{\text{probe}}, \cdot) = -d(\mathbf{x}_{\text{probe}}, \cdot)$

Other discriminative classification models:

- Support vector machine (SVM)
- Random forests
- Boosting methods

## Face Identification with Open-Set Protocol

Handle open space using thresholded nearest neighbors:

- Extract best matching gallery image  $\mathbf{x}_{\text{match}}$  with minimum distance to  $\mathbf{x}_{\text{probe}}$
- Possibly assign "unknown" identity to  $\mathbf{x}_{\text{probe}}$ :

$$I(\mathbf{x}_{\text{probe}}) = \begin{cases} \text{mode}(I(\mathbf{x}_1), \dots, I(\mathbf{x}_k)) & \text{if } d(\mathbf{x}_{\text{probe}}, \mathbf{x}_{\text{match}}) \leq \tau \\ \text{unknown} & \text{if } d(\mathbf{x}_{\text{probe}}, \mathbf{x}_{\text{match}}) > \tau \end{cases} \quad (15)$$

$\tau$  is the face verification threshold

# Learning Open-Set Models using Known Unknowns

Extend the training setup of open-set models:

- Training set with KCs and known unknown classes (KUCs)

$$\mathcal{T} = \{(\mathbf{x}_i, y_i)\}_{i=1}^T \quad (16)$$

- Labels  $y_i$  can encode any KC ( $\mathcal{C}_K = y_1, y_2, \dots, y_n$ ) or unknowns  $u$

$$\mathcal{C} = \mathcal{C}_K \cup u \quad (17)$$

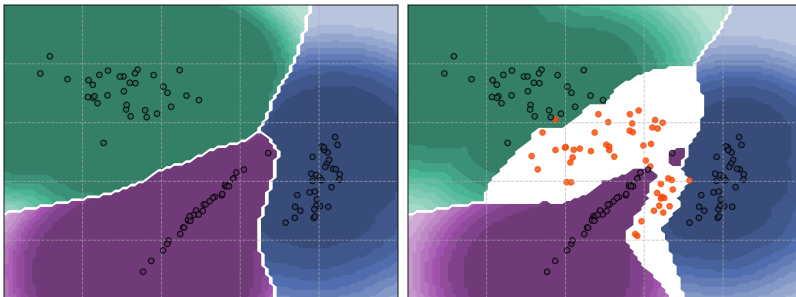
- Learn a likelihood function  $h(\mathbf{x}; \mathcal{T})$  from the entire training set  $\mathcal{T}$

Decision function to predict a label  $\hat{y}$  with maximum likelihood:

$$\begin{cases} \hat{y} & \max_{y \in \mathcal{C}} h(\mathbf{x}; \mathcal{T}) \geq \tau \text{ and } \hat{y} \neq u \\ u & \text{otherwise} \end{cases} \quad (18)$$

# Pseudo Labeling for Open-Set Learning

Idea: introduce pseudo labels for KUCs and learn from modified training set



Open-set decision boundary without KUCs (left) and with consideration of KUCs during training (right)<sup>4</sup>

<sup>4</sup>Koch, T., Riess, C., Köhler, T. (2023). LORD: Leveraging Open-Set Recognition with Unknown Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4386-4396).



# Pseudo Labeling for Open-Set Learning

Single pseudo label (SPL):

- Assumption: all KUC samples form a large background class
- KUCs are treated as a single class with pseudo-label  $u$  resulting in the label set:

$$\mathcal{C} = \{\underbrace{y_1, y_2, \dots, y_n}_{n \text{ KCs}}, u\} \quad (19)$$

- Predicts unknowns directly independent of a decision threshold
- Can be used with any open-set / closed-set backbone to learn likelihood function

# Pseudo Labeling for Open-Set Learning

Multi pseudo label (MPL):

- Assumption: every KUC sample models a different class
- KUCs are treated as separate classes with a single sample per pseudo label resulting in the label set:

$$\mathcal{C} = \underbrace{\{y_1, y_2, \dots, y_n\}}_{n \text{ KCs}}, \underbrace{\{y_{n+1}, y_{n+2}, \dots, y_{n+m}\}}_{m \text{ KUCs}} \quad (20)$$

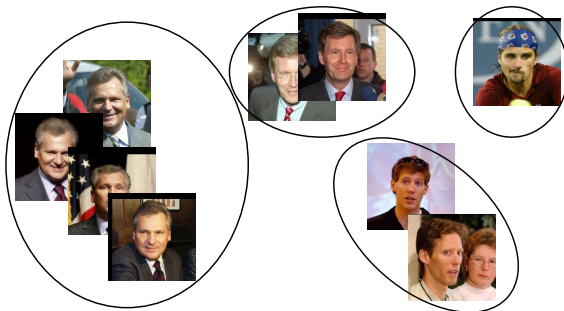
- Prediction of any pseudo label is mapped to the unknown class
- Can be used with any open-set / closed-set backbone to learn likelihood function but limited due to computational complexity

# Face Clustering

- Given  $n$  face images  $\mathbf{x}_1, \dots, \mathbf{x}_n$ , cluster them into  $k \leq n$  clusters  $\mathcal{C}_1, \dots, \mathcal{C}_k$
- Minimize with cluster centers  $\mu_i$  and distance measure  $d(\cdot, \cdot)$ :

$$(\mathcal{C}_1, \dots, \mathcal{C}_k) = \operatorname{argmin}_{\mathcal{C}_1, \dots, \mathcal{C}_k} \sum_{i=1}^k \sum_{\mathbf{x} \in \mathcal{C}_i} d(\mathbf{x}, \mu_i) \quad (21)$$

- $k$ -means clustering:  $d(\mathbf{x}, \mu_i) = \|\mathbf{x} - \mu_i\|_2^2$  and  $\mu_i \equiv$  cluster mean



## Clustering using $k$ -Means Algorithm

Iterative algorithm:

1. Assignment: assign each face to the cluster with closest center

$$\mathcal{C}_i^t = \{ \mathbf{x} : \|\mathbf{x} - \mu_i\|_2^2 \leq \|\mathbf{x} - \mu_j\|_2^2 \text{ for all } i \neq j \} \quad (22)$$

2. Update: re-calculate cluster centers from current assignment  $\mathcal{C}_1^t, \dots, \mathcal{C}_k^t$

$$\mu_i^{t+1} = \frac{1}{|\mathcal{C}_i^t|} \sum_{\mathbf{x} \in \mathcal{C}_i^t} \mathbf{x} \quad (23)$$

# Evaluating Face Recognition Systems

Face identification with closed-set protocol:

- Rank of a probe image  $\mathbf{x}_{\text{probe}}$  for gallery  $\mathcal{G}$  with true matching image  $\mathbf{x}_{\text{match}}$ :

$$\text{Rank}(\mathbf{x}_{\text{probe}}) = \left| \left\{ \mathbf{x}' \in \mathcal{G} : s(\mathbf{x}', \mathbf{x}_{\text{probe}}) \geq s(\mathbf{x}_{\text{match}}, \mathbf{x}_{\text{probe}}) \right\} \right| \quad (24)$$

$\text{Rank}(\mathbf{x}_{\text{probe}}) = 1$  if  $\mathbf{x}_{\text{probe}}$  is correctly associated with  $\mathbf{x}_{\text{match}}$

- Rank- $k$  Identification rate on a test set  $\mathcal{T}$ :

$$\text{IR}(r) = \frac{\left| \left\{ \mathbf{x}' \in \mathcal{T} : \text{Rank}(\mathbf{x}') \leq r \right\} \right|}{|\mathcal{T}|} \quad (25)$$

For  $r = 1$  it is equivalent to the accuracy

## Evaluating Face Recognition Systems

Face identification with open-set protocol:

- Trade-off between true identifications and unknowns that are incorrectly detected as knowns (false alarms) depending on similarity threshold  $\theta$
- Detection and identification rate for test set of knowns  $\mathcal{K}$ :

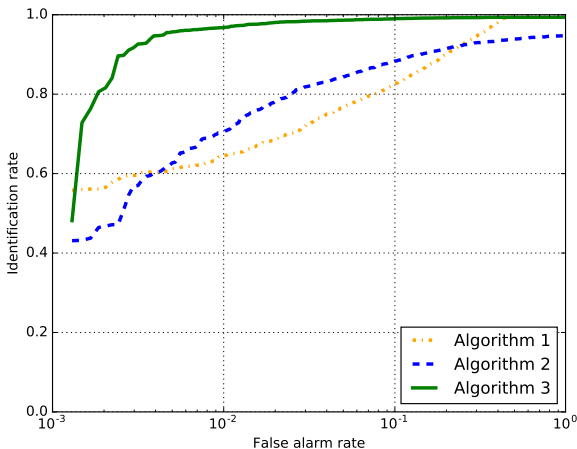
$$\text{DIR}(\theta) = \frac{|\{ \mathbf{x}' \in \mathcal{K} : s(\mathbf{x}', \mathbf{x}_{\text{gallery}}) \geq \theta \text{ and Rank}(\mathbf{x}') = 1 \}|}{|\mathcal{K}|} \quad (26)$$

- False alarm rate for complementary test set of unknowns  $\mathcal{U}$ :

$$\text{FAR}(\theta) = \frac{|\{ \mathbf{x}' \in \mathcal{U} : s(\mathbf{x}', \mathbf{x}_{\text{gallery}}) \geq \theta \text{ for any } \mathbf{x}_{\text{gallery}} \in \mathcal{G} \}|}{|\mathcal{U}|} \quad (27)$$

# DIR Curve for Comparison of Face Recognition Algorithms

Depict  $\text{DIR}(\theta)$  at different  $\text{FAR}(\theta)$  with semi-logarithmic axes:





FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
FACULTY OF ENGINEERING

# Summary





## Take Home Messages

- Widespread application domains of face recognition
- Still a hard problem under uncontrolled conditions (e. g. difficult poses)
- Face representation is a key component for modern face recognition systems
  - Learned on large training datasets
  - Different methodologies: Eigenfaces, Fisherfaces, deep features
- Recognition tasks (verification, identification, clustering) solved by common machine learning algorithms
  - Based on suitable face representation
  - Today also applicable on embedded devices

## Further Readings

Overview on face image analysis and recognition techniques:

Anil K. Jain and Stan Z. Li. "Handbook of face recognition". Springer, 2011

Eigenfaces, Fisherfaces, and other classical methods:

Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection." IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 1997, 711-720.

Deep learning based methods:

- Yaniv Taigman *et al.* "Deepface: Closing the gap to human-level performance in face verification." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014
- Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015
- Yandong Wen *et al.* "A discriminative feature learning approach for deep face recognition." European Conference on Computer Vision, 2016.



Thanks for listening.  
**Any questions?**