Face Recognition –Part II

Dr.-Ing. Saquib Sarfraz

November 29, 2013

Outline

Recent Advances

- Local Appearance-based Face Recognition (already last lecture)
- Face recognition accross pose
  - Face normalization using AAMs
  - Video-based face recogntion
  - Face Recognition using 3D Morphable Models
- Face Recogntion in the Wild

Databases

Problem: Matching across face pose

Problem: Different view-point / head orientation

Recognition results degrade, when images of different head orientation have to be matched

Karlsruhe Institute of Technology

Three major directions to address the face recognition across pose Probelm

- Geometric pose normalization (image affine warps)
- 2D specific pose models, image rendering at pixel or feature level – 2D+3D approaches
- 3D face Model fitting

Pose-Normalization

Alignment using just eye-positions is not sufficient

Idea:

- Find several facial features (mesh)
- Use complete mesh to normalize face

Here: 2D Active Appearance Models

- A texture and shape-based parametric model
- Efficient fitting algorithm:

Model and Fitting

Independent Shape and Appearance Model

$$s = (x_1, y_1, x_2, y_2, \cdots, x_v, y_v)^T = s_0 + \sum_{i=1}^{n} p_i s_i$$

$$A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) \qquad \forall x \in s_0$$

Fitting Goal:

$$\arg\min_{p,\lambda} \sum_{x \in s_0} \left[A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) - I(W(x;p)) \right]^2$$

Instances of Shape and Texture

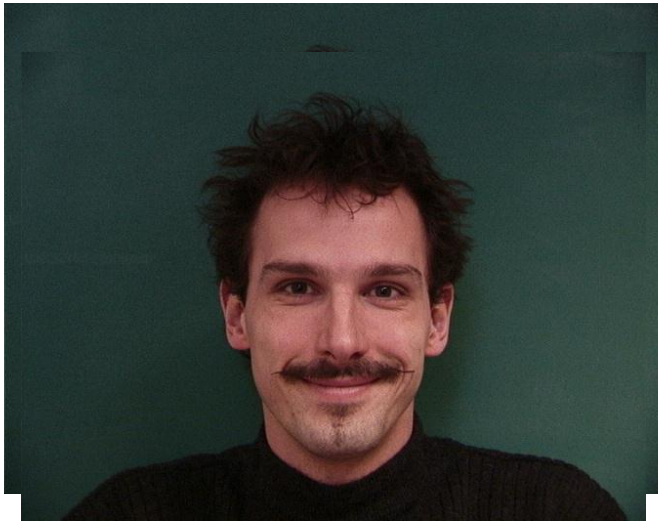
s_0

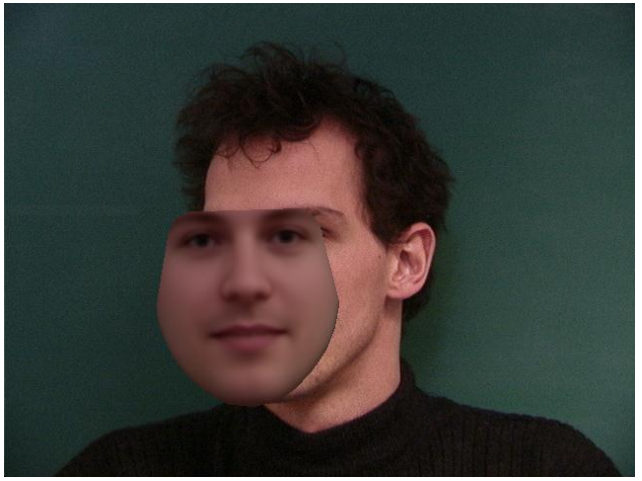
$A_0(x)$

$s_0 + p_1 s_1$ $s_0 + p_2 s_2$ $s_0 + p_3 s_3$

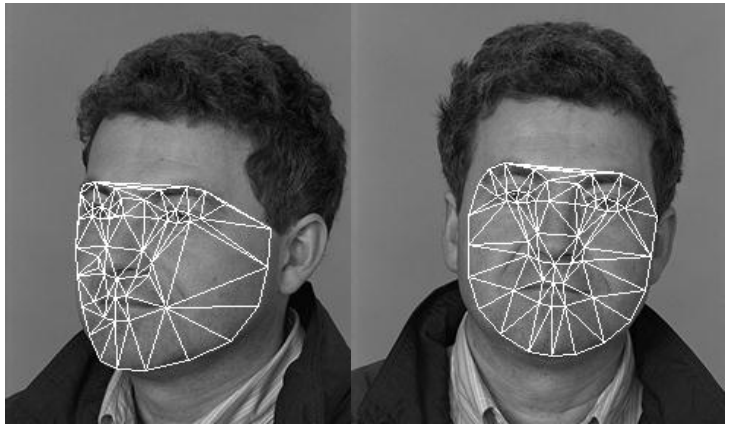
$A_0(x) + \lambda_1 A_1(x)$ $A_0(x) + \lambda_2 A_2(x)$ $A_0(x) + \lambda_3 A_3(x)$

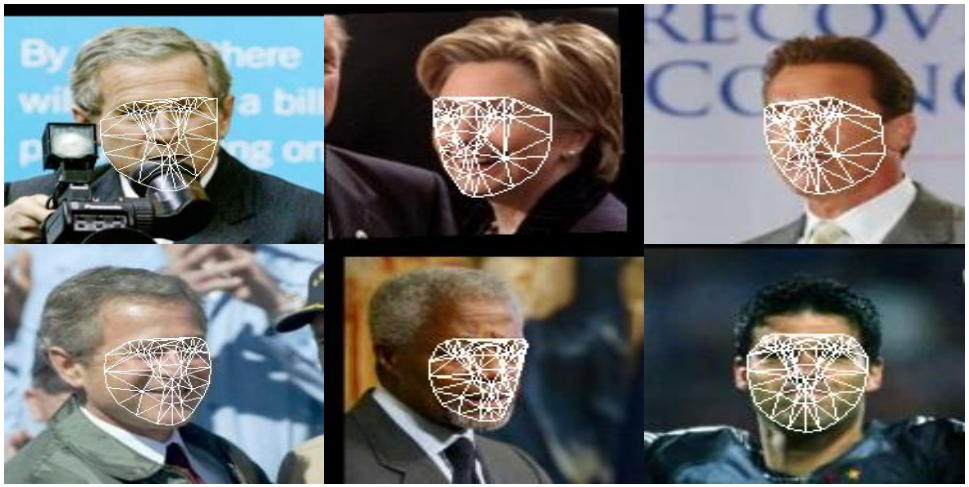
Alignment with AAMs



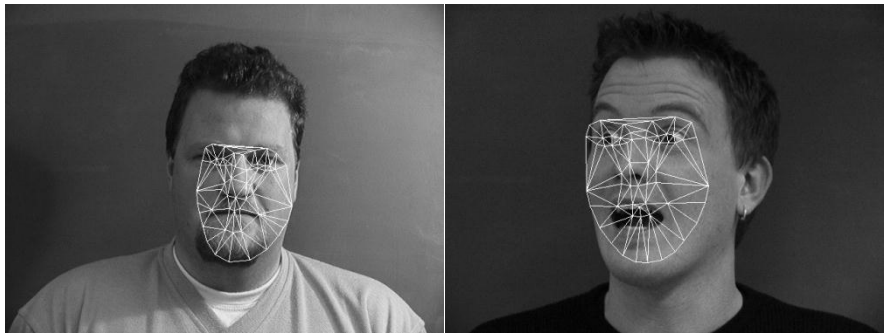


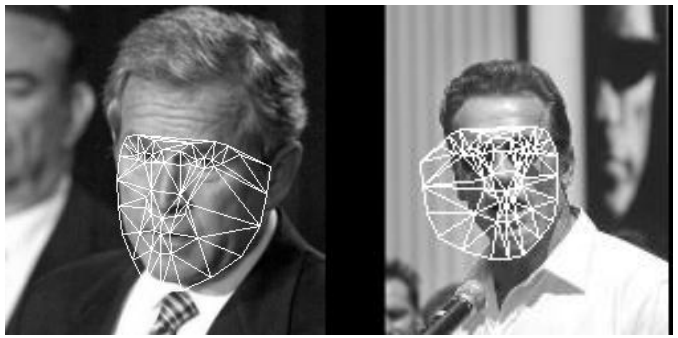
Fitting Examples





Fitted mesh





Mismatched mesh

Pose normalization

Fitted modal can be used to warp image to frontal pose

E.g. using piecewise affine transformation of mesh triangles

Faces with different poses from FERET data base and their pose-aligned images

Results (2)

Much better results under pose variations compared to simple affine transform:

Probe set	*bb*	*bc*	*bd*	*be*	*bf*	*bg*	*bh*	*bi*
With pose correction	44.0%	81.5%	93.0%	97.0%	98.5%	91.5%	78.5%	52.5%
Simple Affine Transf.	0.0%	5.5%	26.0%	62.5%	78.5%	26.5%	4.0%	1.0%

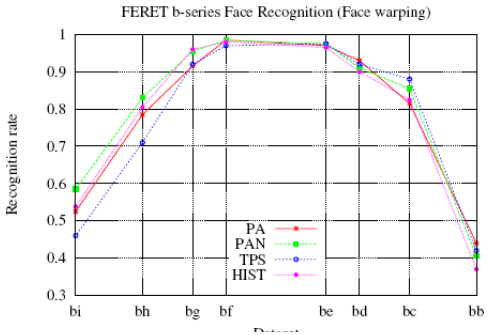
Different warping functions can be used

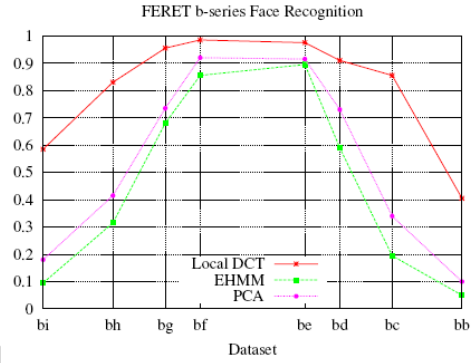
- Piecewise affine transformation worked best

Approach works well with local-DCT-based approach

- but not so well with holistic approaches, such as Eigenfaces (PCA)

(Gao, Ekenel, Stiefelhagen, ICB'09)

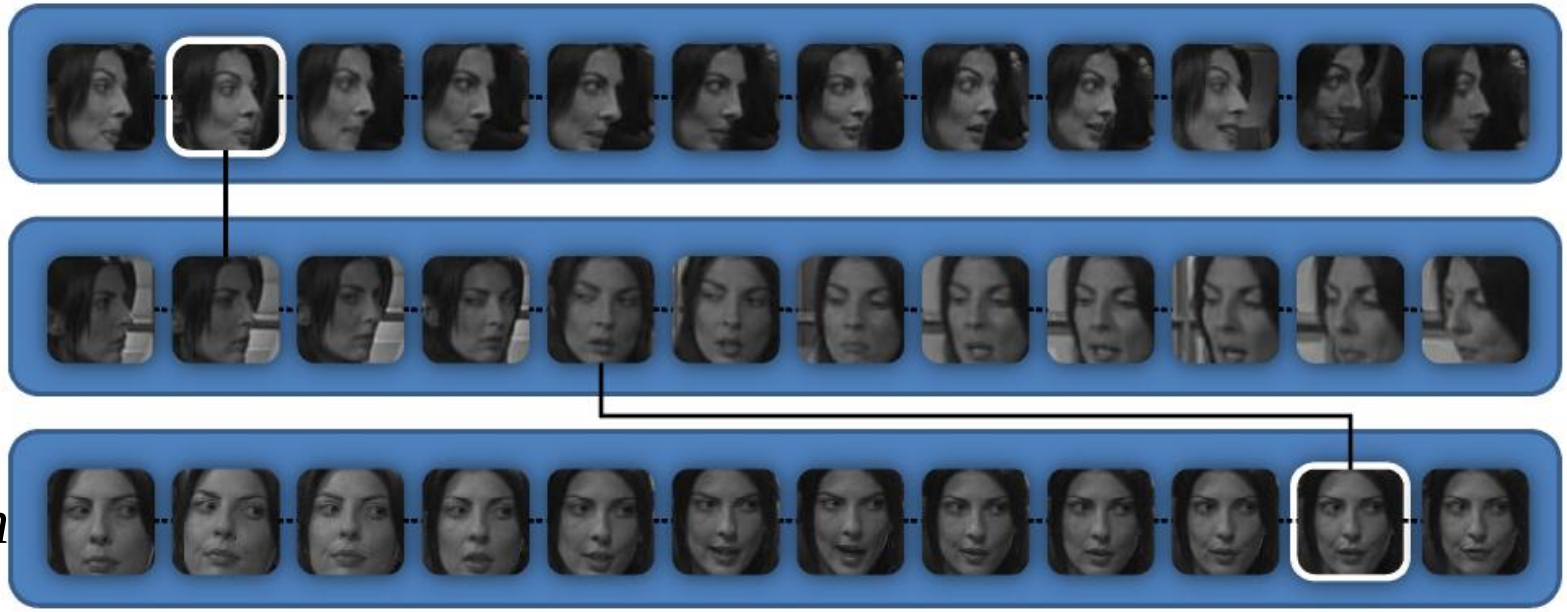




Video-based face recognition

- Temporal information (face tracking) can help us to match faces under different poses
  - problem with AAMs: need good resolution …
- Investigated in a person-retrieval scenario
  - Goal is to find shots with a specific actor
- Approach
  - Pre-segment shots via shot boundary detection
  - Track all faces in each shot -> “face tracks”
  - Find best matches between “face track” of selected face and all other face “tracklets”

Matching face tracks



- Three different tracks of a person
- Faces under different pose can be matched using the temporal association from the different tracks and the associations by face recognition (solid lines)

ace Recognition Based on Fitting a 3D lorphable Model (Blanz & Vetter, 2003)

A method for face recognition across variations in pose and illumination.

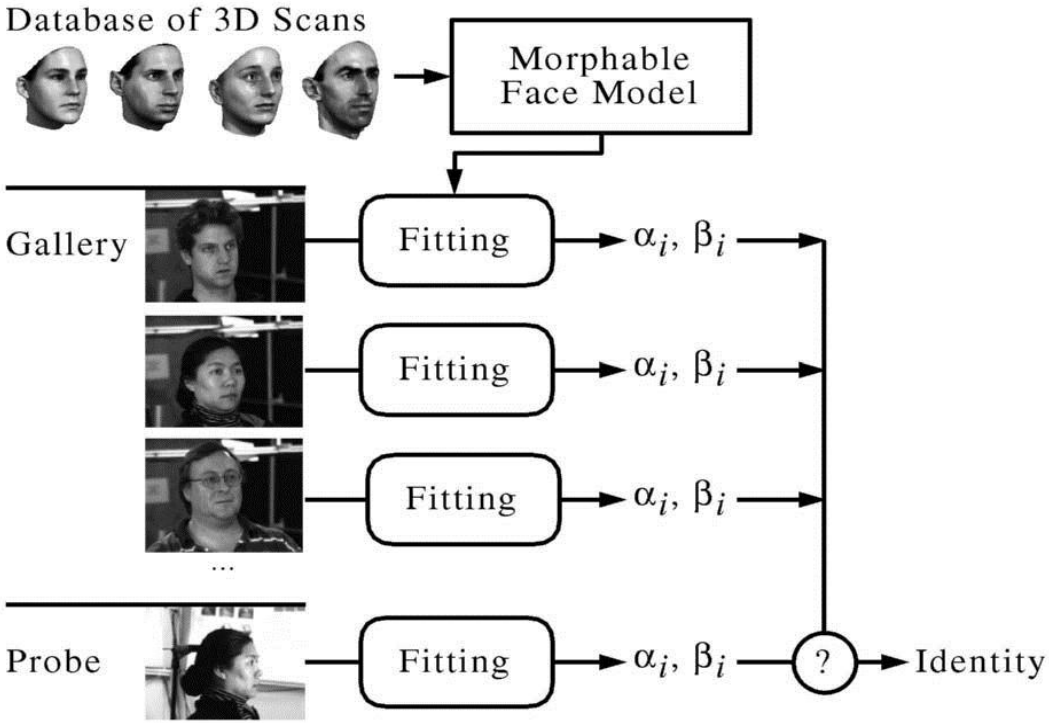
Simulates the process of image formation in 3D space.

Estimates 3D shape and texture of faces from single images by fitting a statistical morphable model of 3D faces to images.

Faces are represented by model parameters for 3D shape and texture.

Face Recognition Based on Fitting a 3D Morphable Model

Model-based Recognition



A Morphable Model of 3D Faces

The morphable face model is based on a vector space representation of faces that is constructed such that any combination of shape and texture vectors S_i and T_i describes a realistic human face:

$$S = \sum_{i=1}^{m} a_i S_i \qquad T = \sum_{i=1}^{m} b_i T_i$$

Database of 3D Laser Scans

3D scans of 100 males and 100 females were used to derive the morphable model.

The scans represents face shape in cylindrical coordinates.

The device measures radius *r*, and red, green, blue{R,G,B} components of surface texture.

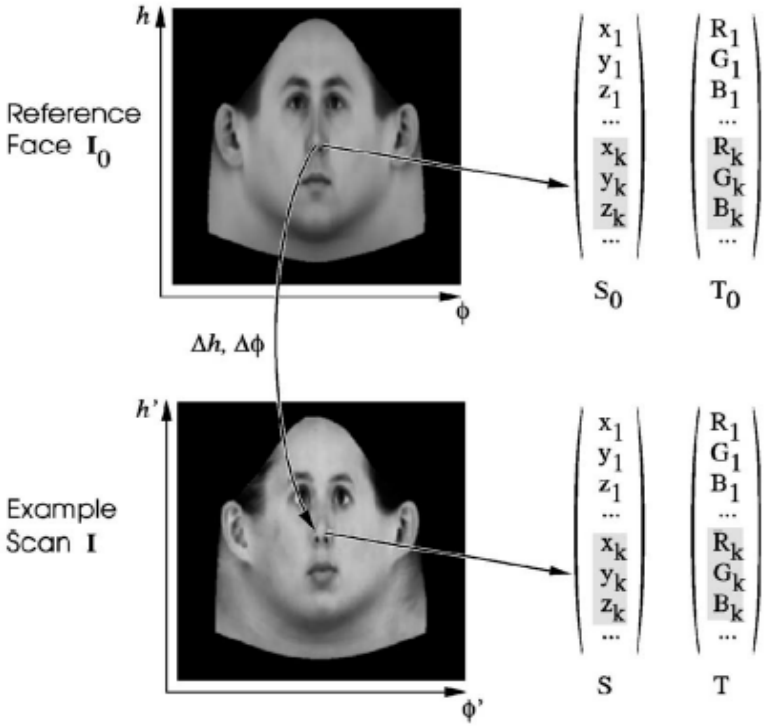
$$I(h,\phi) = (r(h,\phi), R(h,\phi), G(h,\phi), B(h,\phi))^T,$$

$$h,\phi \in \{0,...,511\}.$$

h: vertical steps, ϕ : angular steps

Preprocessing / Alignment

- Some preprocessing is necessary
  - Filling holes
  - Some trimming along edges
  - Removal of back of the head and shoulders
- 3D Alignment of faces
  - Establish point-to-point correspondence between reference and new face
  - Based on optical flow



Face Vectors

The definition of shape and texture vectors is based on a reference face $\mathbf{I_0}$.

The location of the vertices of the mesh in Cartesian coordinates is (x_k, y_k, z_k) with colors (R_k, G_k, B_k)

Reference shape and texture vectors are defined by:

$$S_0 = (x_1, y_1, z_1, x_2, ..., x_n, y_n, z_n)^T$$

$$T_0 = (R_1, G_1, B_1, R_2, ..., R_n, G_n, B_n)^T$$

To encode a novel scan $\mathbf{I}$, the flow field from $\mathbf{I_0}$ to $\mathbf{I}$ is computed.

Principal Component Analysis

PCA is performed on the set of shape and texture vectors separately.

Eigenvectors form an orthogonal basis:

$$\mathbf{S} = \overline{\mathbf{s}} + \sum_{i=1}^{m-1} \alpha_i \cdot \mathbf{s}_i, \quad \mathbf{T} = \overline{\mathbf{t}} + \sum_{i=1}^{m-1} \beta_i \cdot \mathbf{t}_i$$

Principal Component Analysis

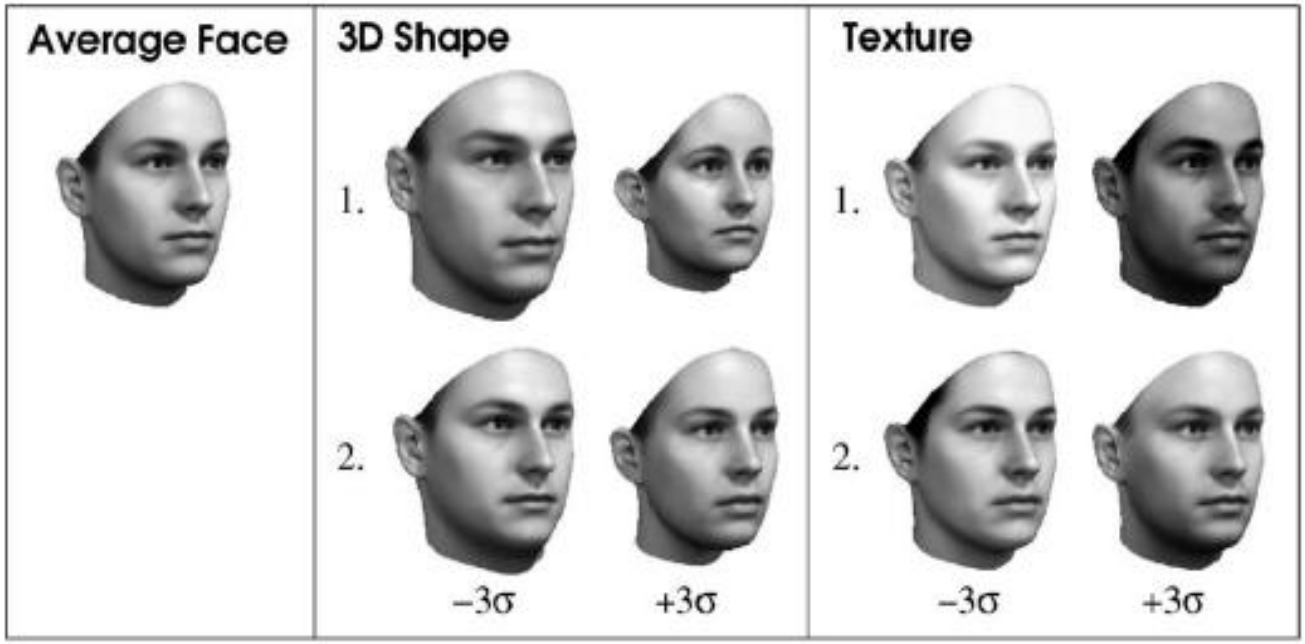


Fig. 4. The average and the first two principal components of a data set of 200 3D face scans, visualized by adding $\pm 3\sigma_{S,i}\mathbf{s}_i$ and $\pm 3\sigma_{T,i}\mathbf{t}_i$ to the average face.

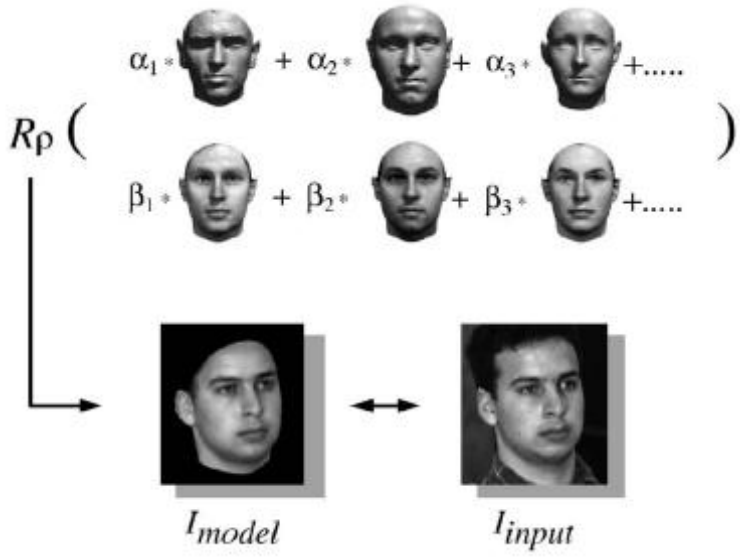
Videos –Building a Morphable Model

Video

Model-based Image Analysis

The goal of the fitting process is to find shape and texture coefficients describing a 3D face model such that rendering produces an image $\mathbf{I}_{\mathbf{model}}$ that is as similar as possible to $\mathbf{I}_{\mathbf{input}}$.

For initialization 7 facial feature points, such as the corners of the eyes or tip of the nose should be labelled manually.



Model Fitting

Goal: Minimize $$E_I = \sum_{x,y} \left\| \mathbf{I}_{input}(x,y) - \mathbf{I}_{model}(x,y) \right\|^2.$$

Shape, texture, transformation, and illumination are optimized for the entire face and refined for each segment.

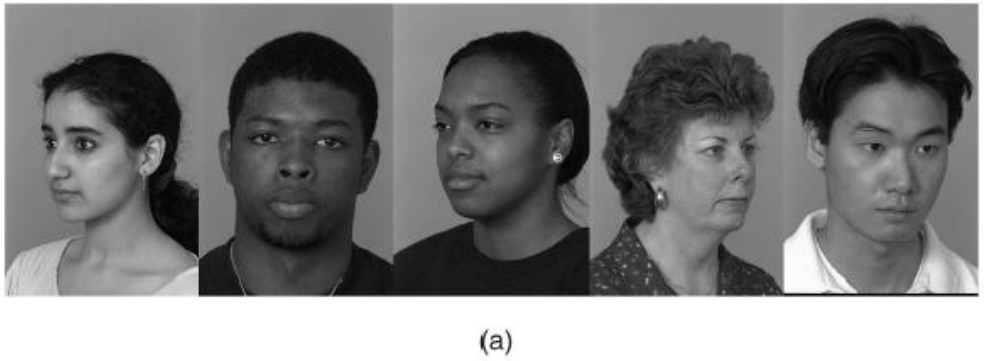
- 99 coefficients α_i, β_i
- 22 rendering parameters (pose angles, translation, focal length, light intensities, illumination direction, …)

Complex iterative optimization procedure

- Processing time: 4.5 minutes per image (2 GHz, P4)

3D Face Reconstruction from a Single Image

Fitting Results

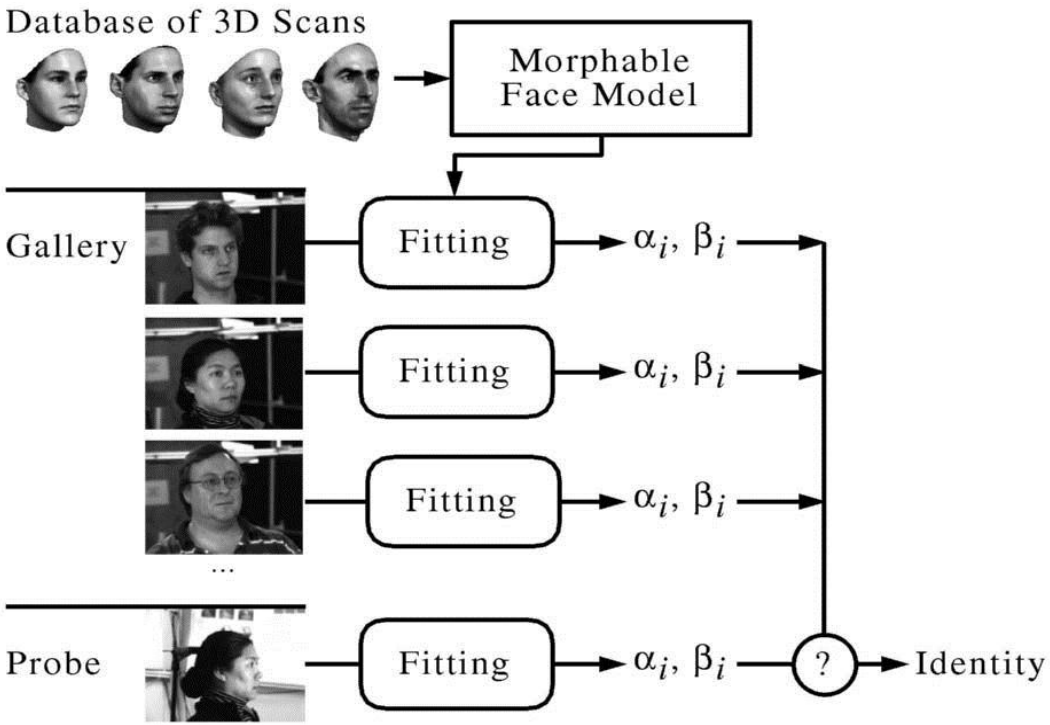


Original Image

Fitted Model

Novel views

Model-based Recognition



Results

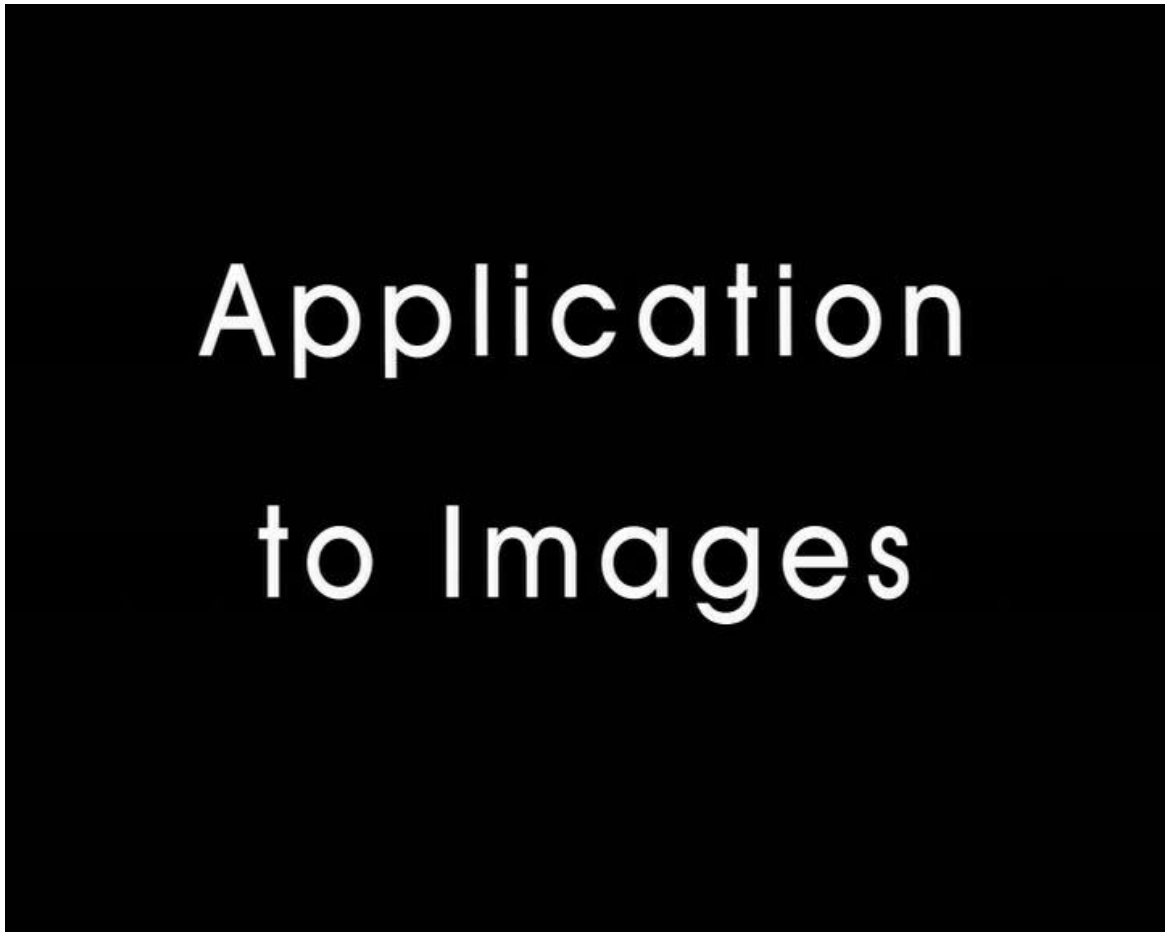
FERET: used 194 individuals, 10 poses, different illuminations conditions (only in one pose)

CMU-PIE: used 68 individuals, three view-points (front, side, profile), 22 illuminations

Database	d_M	d_A	d_W
CMU-PIE	87.2%	94.2%	**95.0%**
FERET	80.3%	92.2%	**95.9%**

Results for different distance measures

Videos –Application to Images



Video

Face Recognition in the Wild – State-of-the-art

1. **Probabilistic Elastic Matching for Pose Variant Face Verification (CVPR 2013)**

2. **Blessing of Dimensionality: High-dimensional Feature and Its Efficient Compression for Face Verification (CVPR 2013)**

Some Background

- High Dimensional Features / Bag of Features
- Some Mid-level representation e.g. GMM to encode
- Subspace dimensionality Reduction
- Matching directly or by Metric Learning (SVM etc)
- 6-7 % improvement over single scale normal features?

1. Probabilistic Elastic Matching

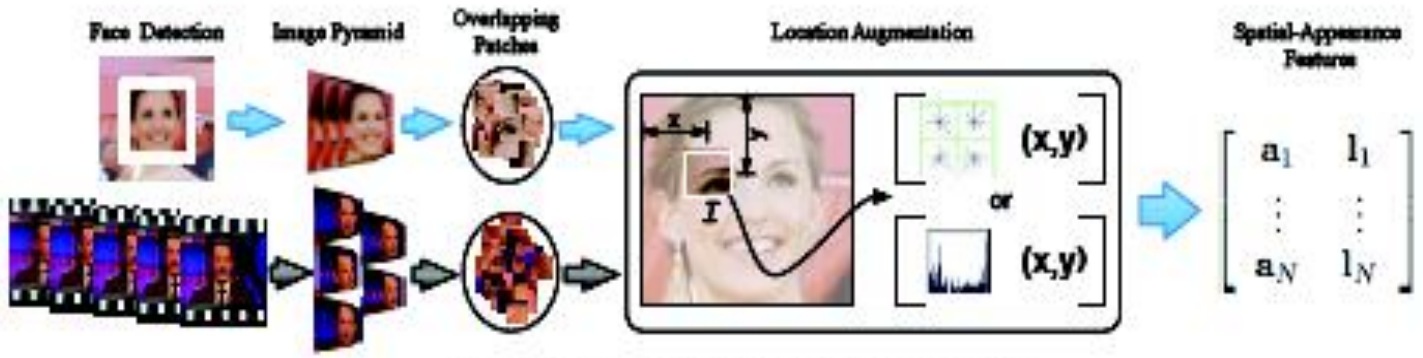


Figure 1. Spatial-appearance feature extraction pipeline.

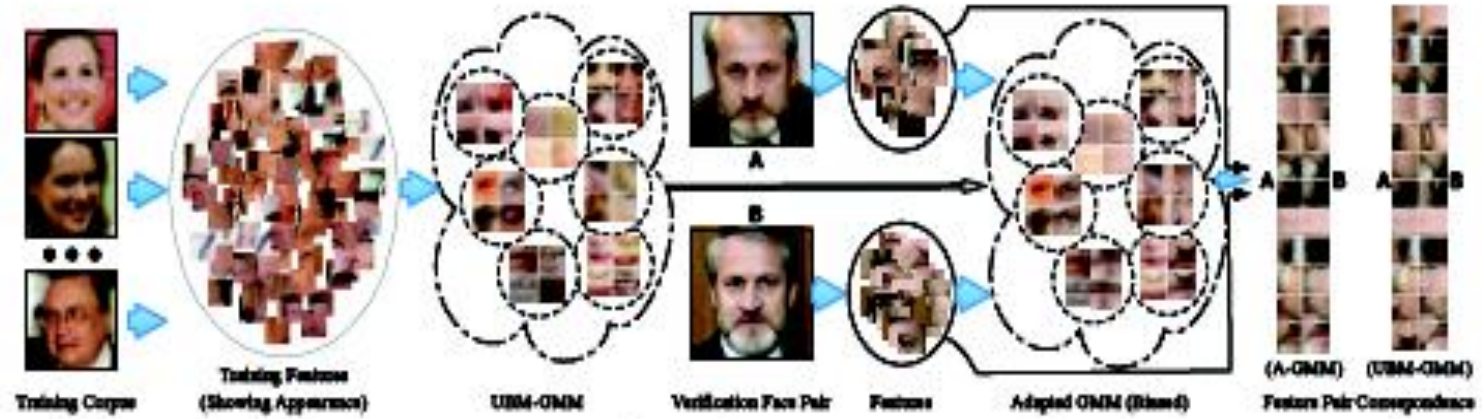


Figure 2. Our pipeline to build feature pair correspondence.

- 3-level image pyramid
- Dense overlapping patches / Feat Ext. (SIFT/LBP) with location embedding– Bag of Feature model
- **Training** : Universal GMM with 1024 spherical Gaussians
- As opposed to use for encoding they use this to establish direct correspondences in a face pair.
- **Test**: assign one most likely feature vector to each Gaussian component, thereby getting a 1024 features in each image in the pair. (Correspondence)
- This correspondence is improved by adapting the UB-GMM to each pair. (APEM)
- The difference vector is then classified sing SVM.

Karlsruhe Institute of Technology

Adapted GMM

- The Universal Background Model (UBM) parameters are used as priors to get the adapted GMM

- For a pair of face image the features X_P are the union of the feature set from both images.

Matching

- Assign most likely feature to each Gaussian component.
- Represent each image as K M-dim vector
- Classify the difference vector $\mathbf{d}_i = [\Delta\mathbf{a}_{g_1}\ \Delta\mathbf{a}_{g_2}\ \ldots\ \Delta\mathbf{a}_{g_K}]^T,$ (12)

(a) Feature correspondences built through UBM-GMM

(b) Feature correspondences built through A-GMM

Figure 4. In both figures, the row above shows local patches from face A shown in Figure 2, while the bottom ones are from face B. Each column shows a pair of features captured by one Gaussian component in the GMM.

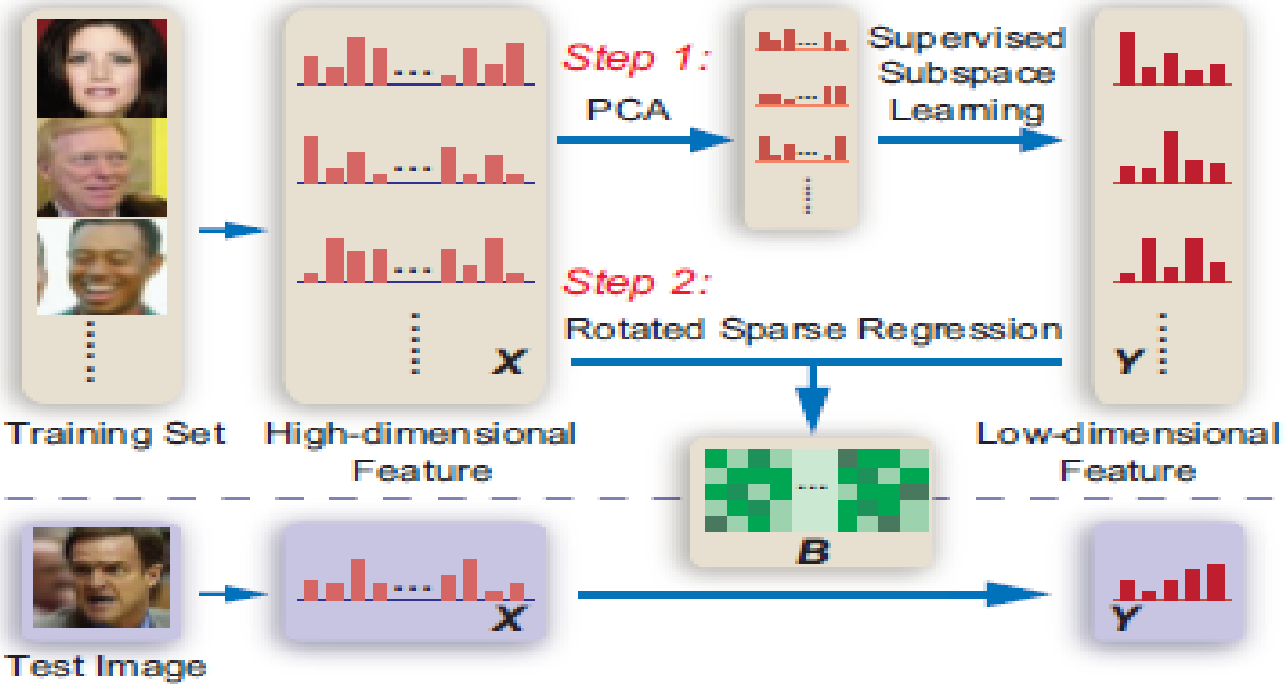
Results

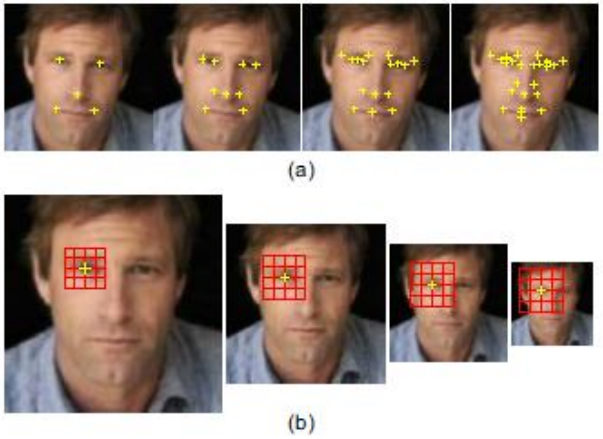
Table 1. Performance comparison on the most restricted LFW

Algorithm	Accuracy ± Error(%)
Nowak[20]	73.93 ± 0.49
Hybrid descriptor-based[28]	78.47 ± 0.51
V1/MKL[21]	79.35 ± 0.55
Baseline (fusion)	77.30 ± 1.59
PEM (LBP)	81.10 ± 1.71
PEM (SIFT)	81.38 ± 0.98
PEM (fusion)	82.93 ± 1.18
APEM (LBP)	81.97 ± 1.90
APEM (SIFT)	81.88 ± 0.94
APEM (fusion)	$\mathbf{84.08 \pm 1.20}$
APEM (fusion, unaligned)	81.70 ± 1.78

2. Blessing of Dimensionality: High-dimensional Feature and Its Efficient Compression for Face Verification

- Extract dense feature on 27 landmarks across 5 scales (simple resized)
- Instead of using GMM etc. to encode the chunk it uses learned subspace mapping to directly project the concatenated vector onto a lower dimensional space.





- Computes 5 different desriptors LBP, SIFT HOG, Gabor and LE from each patch (4x4 cell). Concatenating each descriptor from all patches/sclaes give rise to the high dim. feature
- Training: PCA (400) and supervised subspace LDA, joint Bayesian
- Linear Sparse Regression to learn the direct projection matrix

Experiments

- Supervised setting (unrestricted LFW or using WDRef as outside training data)

	Baseline	High dimension
LE	88.78%	**92.92%**
LBP	88.33%	**93.18%**
SIFT	85.95%	**91.77%**
HOG	87.90%	**91.10%**
Gabor	84.93%	**90.97%**

Table 1. The comparison between the high-dimensional feature and the baseline feature under LFW unrestricted protocol.

	Baseline	High dimension
LE	90.28%	**94.89%**
LBP	89.39%	**95.17%**
SIFT	86.85%	**93.21%**
HOG	88.93%	**93.40%**
Gabor	87.38%	**92.83%**

Table 3. The comparison between the high-dimensional feature and the baseline feature. Training is on WDRef and testing is on LFW.

- Unsupervised (just using LFW restricted)

	LFW		Multi-PIE	
	Baseline	High dim	Baseline	High dim
LE	81.05%	**84.58%**	83.27%	**87.23%**
LBP	80.05%	**84.08%**	80.60%	**83.92%**
SIFT	77.17%	**83.03%**	79.30%	**83.97%**
HOG	80.08%	**84.98%**	82.98%	**87.08%**
Gabor	74.97%	**82.02%**	81.05%	**85.12%**

Table 4. The comparison between the high-dimensional feature and the baseline feature on LFW and Multi-PIE database under unsupervised setting.

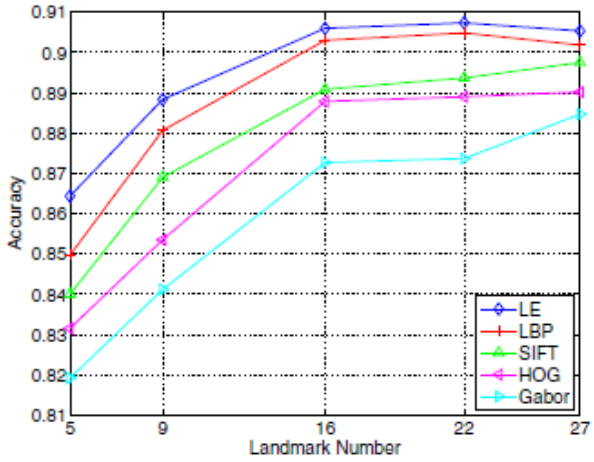


Figure 4. The effect of landmark number on performance.

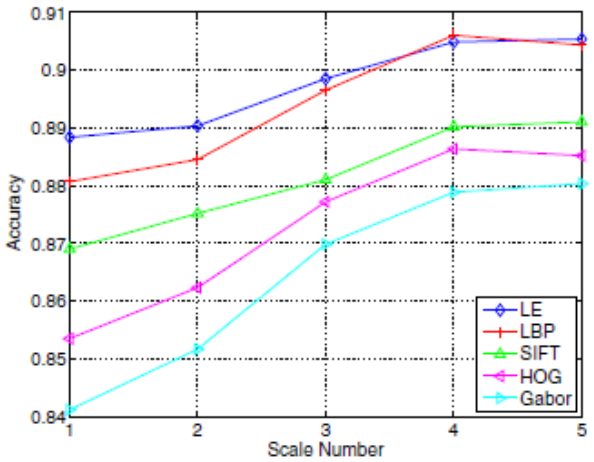


Figure 5. This figure shows the effect of multi-scale representation.

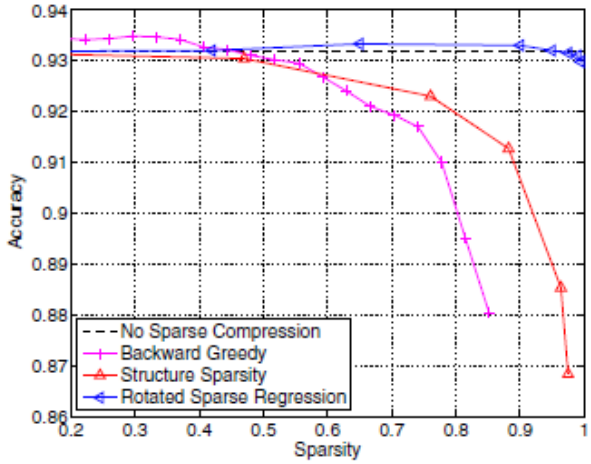


Figure 6. This figure compares the rotated sparse regression and two feature selection methods

Sparsity	Compression Ratio	Sparse Regression	Rotated Sparse Regression
0.95	20	93.18%	93.18%
0.98	50	92.93%	93.18%
0.99	100	92.05%	93.09%
0.995	200	91.43%	92.98%

Table 5. The comparison of the sparse regression and rotated s-parse regression under various sparsity.

Further reading

- M. Dixit, N. Rasiwasia, and N. Vasconcelos. Adapted gaussian models for image classification. In CVPR, 2011.
- D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun. Bayesian face revisited: A joint formulation. In European Conference on Computer Vision, pages 566–579, 2012.
- L. Clemmensen, T. Hastie, D.Witten, and B. Ersboll. Sparse discriminant analysis. Technometrics, 2011

References

Volker Blanz, Thomas Vetter, Face Recognition Based on Fitting a 3D Morphable Model, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.9, September 2003.

Chen D., X. Cao, L. Wang, F. Wen, and J. SunJ. (2013): Blessing of High Dimensionalty: High-dimensional Feature and Its Efficient Compression for Face Verification. In CVPR.

Li H., Hua G., Lin Z., Brandt J. , Yang J.,(2013): Probabilistic Elastic Matching for Pose Variant Face Verification. In CVPR.

Additional

Sarfraz M. Saquib , Siddique M. A. ,Stiefelhagen R. (2013) "**RPM, Random Points Matching for pair-wise Face-Similarity**", In proceedings British Machine Vision Conference BMVC

Hua Gao, Hazım Kemal Ekenel, and Rainer Stiefelhagen.
Pose Normalization for Local Appearance-Based Face Recognition.
3rd Int.l Conference on Biometrics (ICB 2009), LNCS 5558, pp. 32–41, 2009.

M. Bäuml, K. Bernardin, M.Fischer, H.K. Ekenel, R.Stiefelhagen
Multi-Pose Face Recognition for Person Retrieval in Camera Networks
7th International Conference on Advanced Video and Signal-Based Surveillance, Boston, August 2010

HiWi

Chance to work on some ongoing multi partner funded project

Your job is to evaluate the developed algorithms on the dataset as per instructions.

You need to be good at C/C++, openCV.

Contact

Saquib Sarfraz (saquib.sarfraz@kit.edu)

Karlsruhe Institute of Technology

Databases

Face Recognition Evaluations & Databases

Databases

- are necessary to develop and improve algorithms
- provide common testbeds and benchmarks which allow for comparing different approaches
- different databases focus on different problems
  - illumination, pose, expression, number of individuals etc.

Well-known databases for face recognition

- FERET:12000 individuals, different illumination & expression
- FRVT: 37437 individuals, still images + video
- FRGC: ~700 subjects, high-resolution images, 3D data
- CMU-PIE: 68 subjects, different illuminations and head poses
- BANCA: multimodal data for verification, 208 individuals, 4 languages
- XM2VTS: multimodal, 295 subjects, various head poses
- Yale, Harvard, MIT, Olivetti, ... (many more!)

Face Recognition Technology (FERET) Program

The goal of the program:

to develop automatic face recognition capabilities that could be employed to assist security, intelligence, and law enforcement personnel in the performance of their duties.

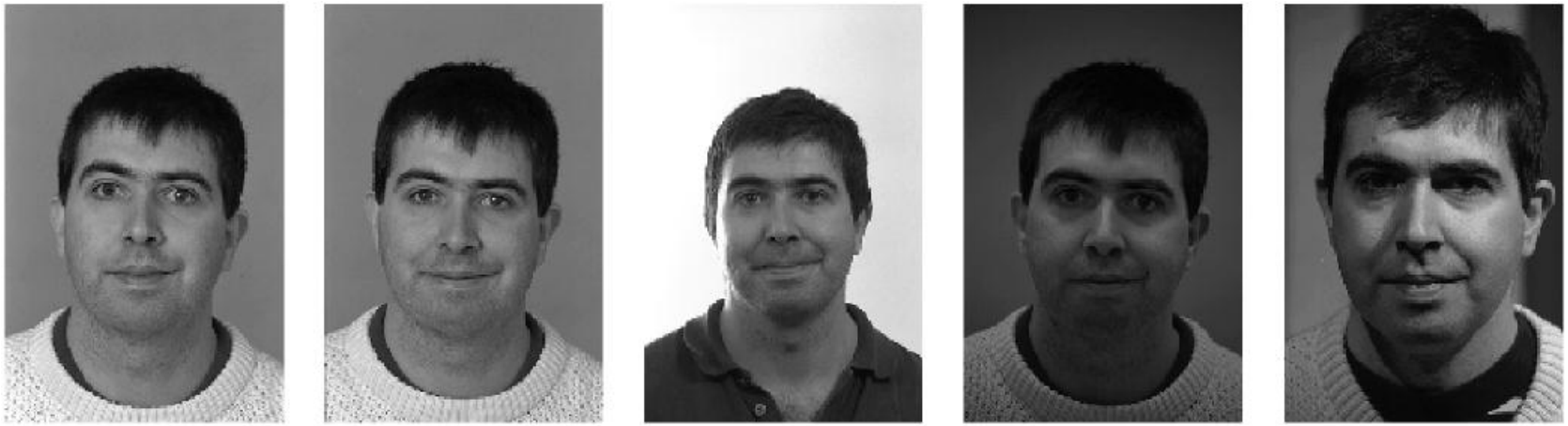
The elements of the program

Sponsoring research

Collecting the FERET database (14126 facial images of 1199 individuals)

Performing the FERET evaluations (1994, 1995, 1996)

Examples of face image categories in the FERET Database

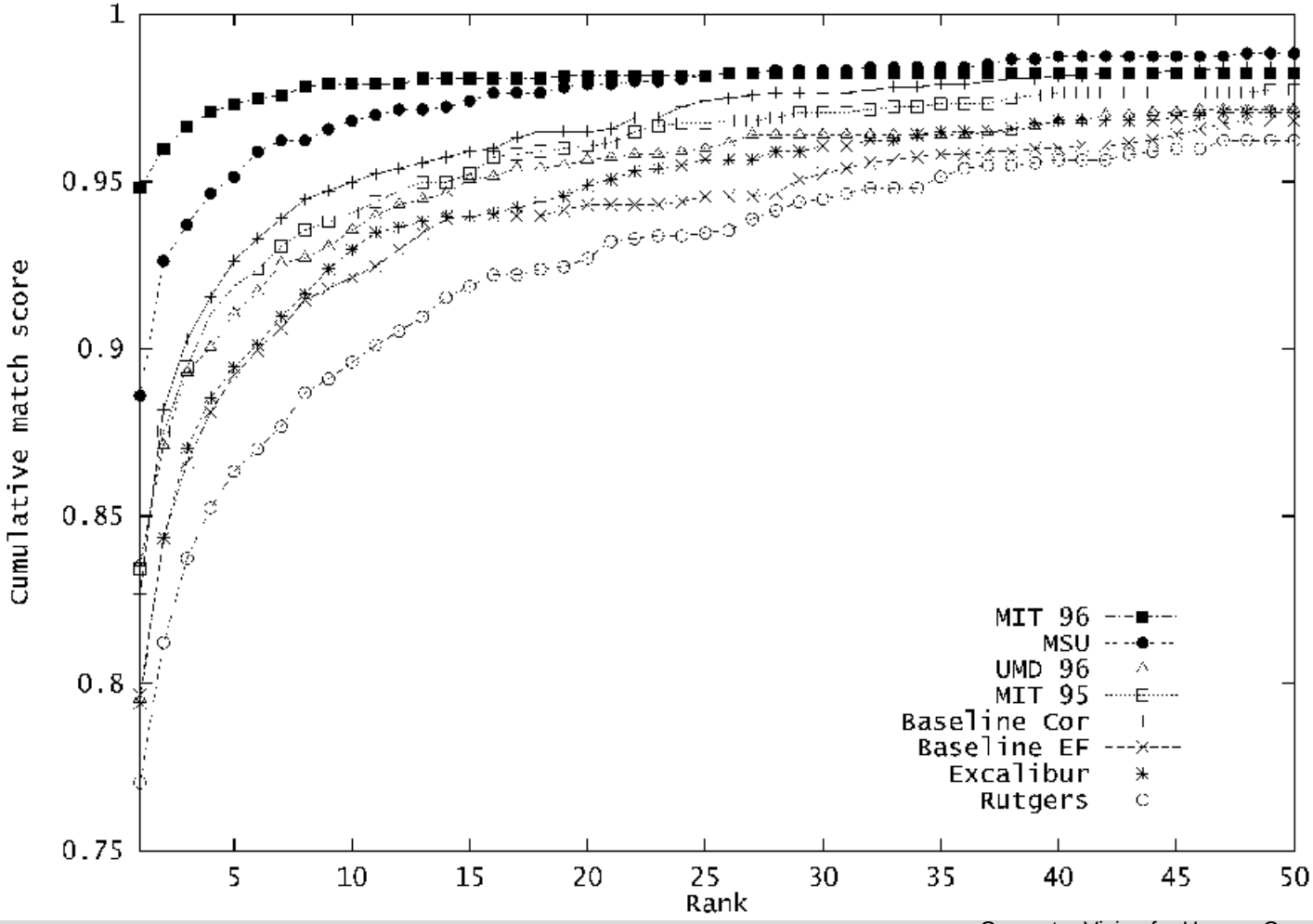


- **fa:** Face image under controlled lighting, normal expression
- **fb:** Face image under controlled lighting, different expression
- **fc:** Face image under different lighting, normal expression
- **duplicate I** image was taken within one year of the fa image
- **duplicate II** and fa images were taken at least one year apart

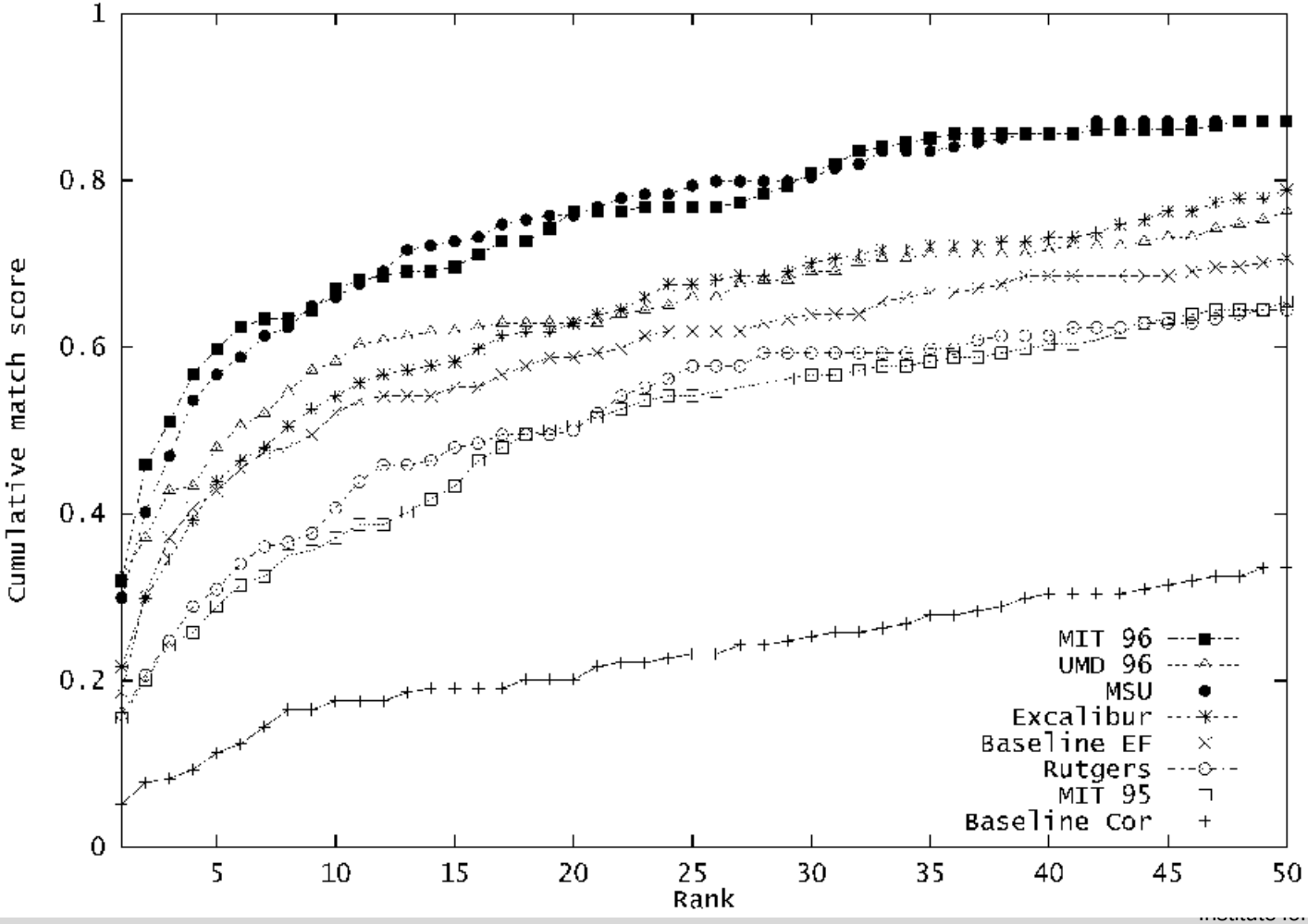
Representation and Similarity Metric for Algoritms Evaluated (1996)

Algorithm	Represenation	Similarity measure
Excalibur Co.	Unknown	Unknown
MIT Media Lab 95	PCA	L_2
MIT Media Lab 96	PCA-difference space	MAP Bayesian Statistic
Michigan St. U.	Fischer discriminant	L_2
Rutgers U.	Greyscale projection	Weighted L_1
U. of So. CA.	Dynamic Link Architecture (Gabor Jets)	Elastic graph matching
U. of MD 96	Fischer discriminant	L_2
U. of MD 97	Fischer discriminant	Weighted L_2
Baseline	PCA	L_1
Baseline	Correlation	Angle

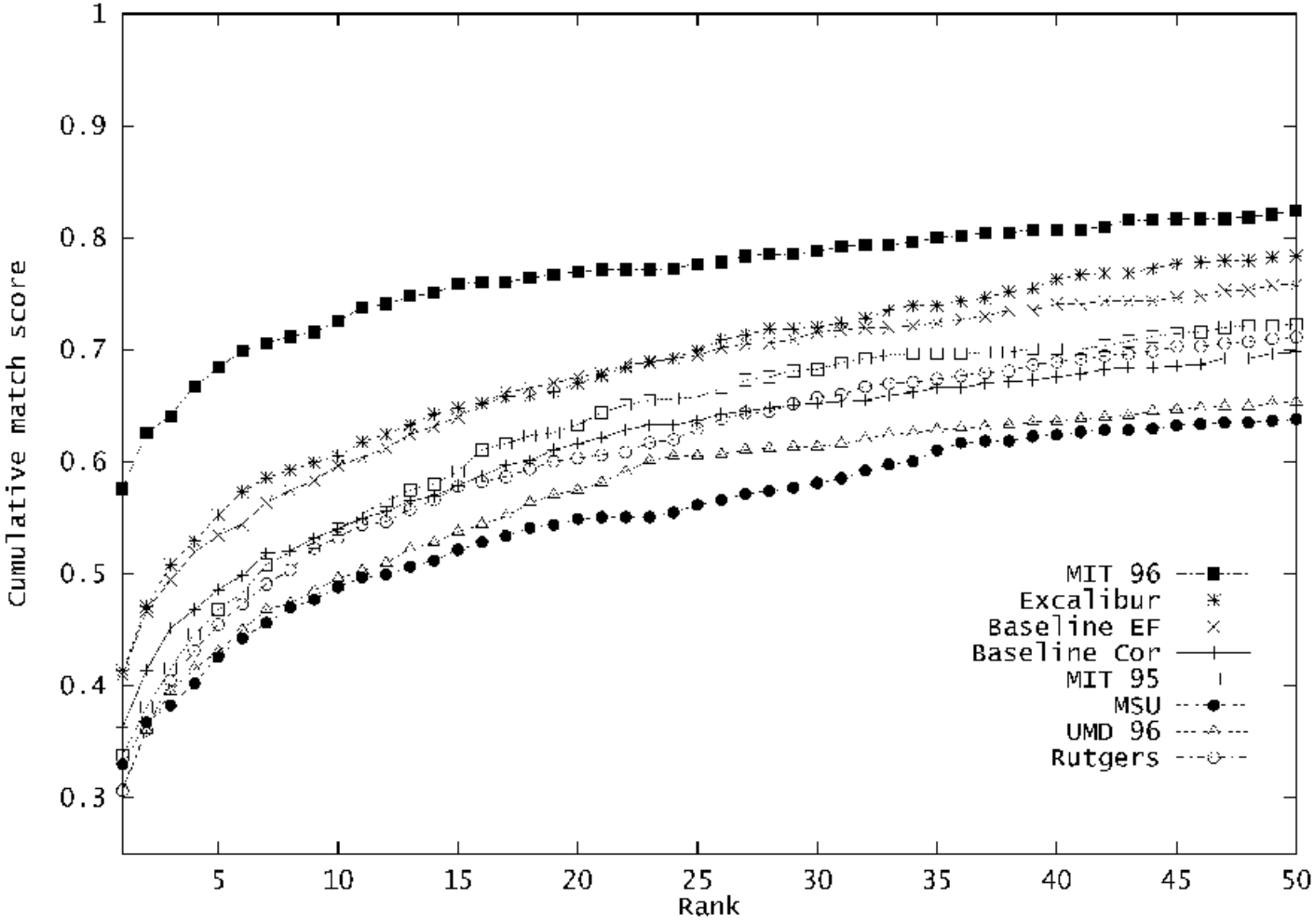
dentification performance against “fb“ images different expressions)



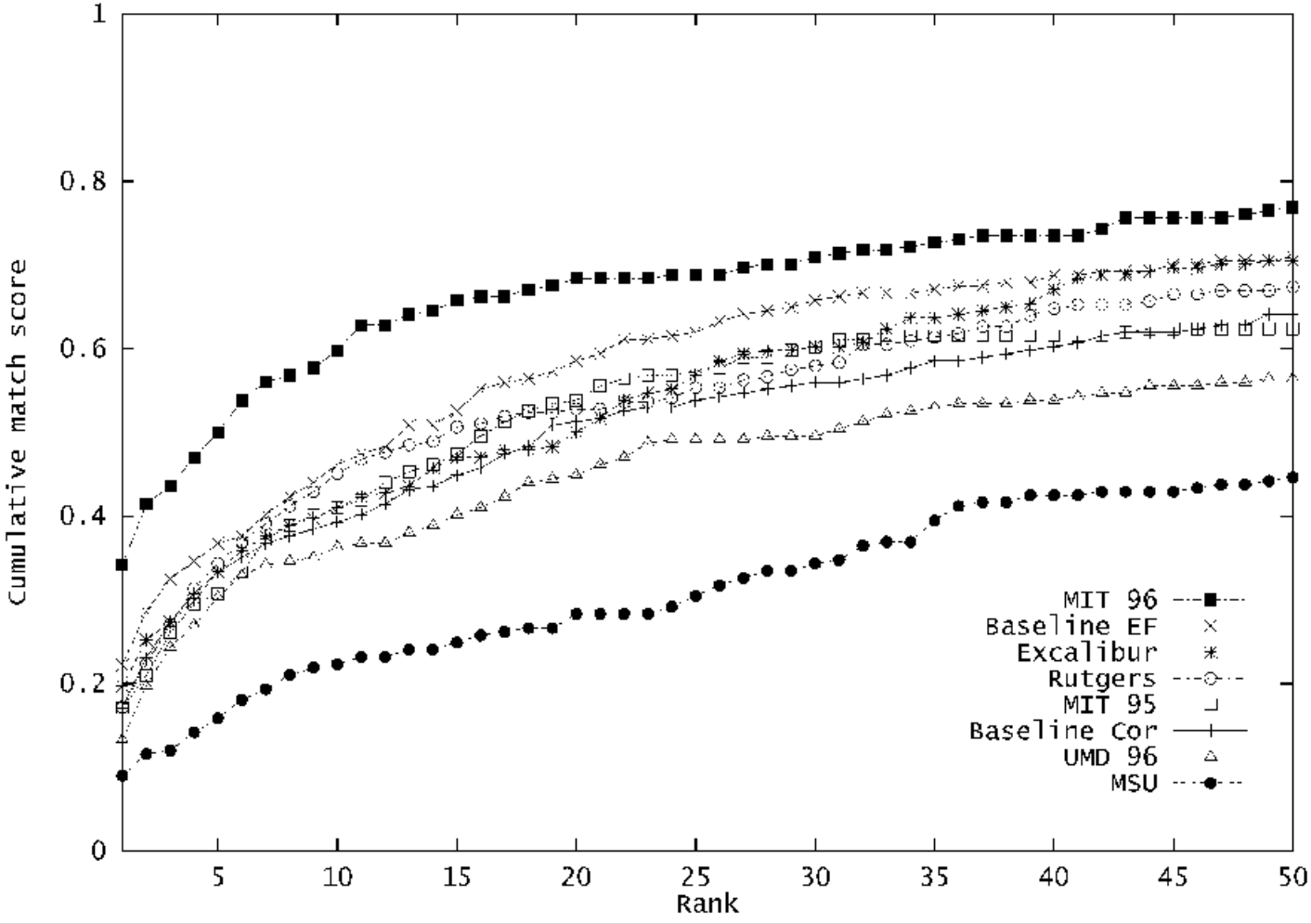
dentification performance against “fc“ images different illumination)



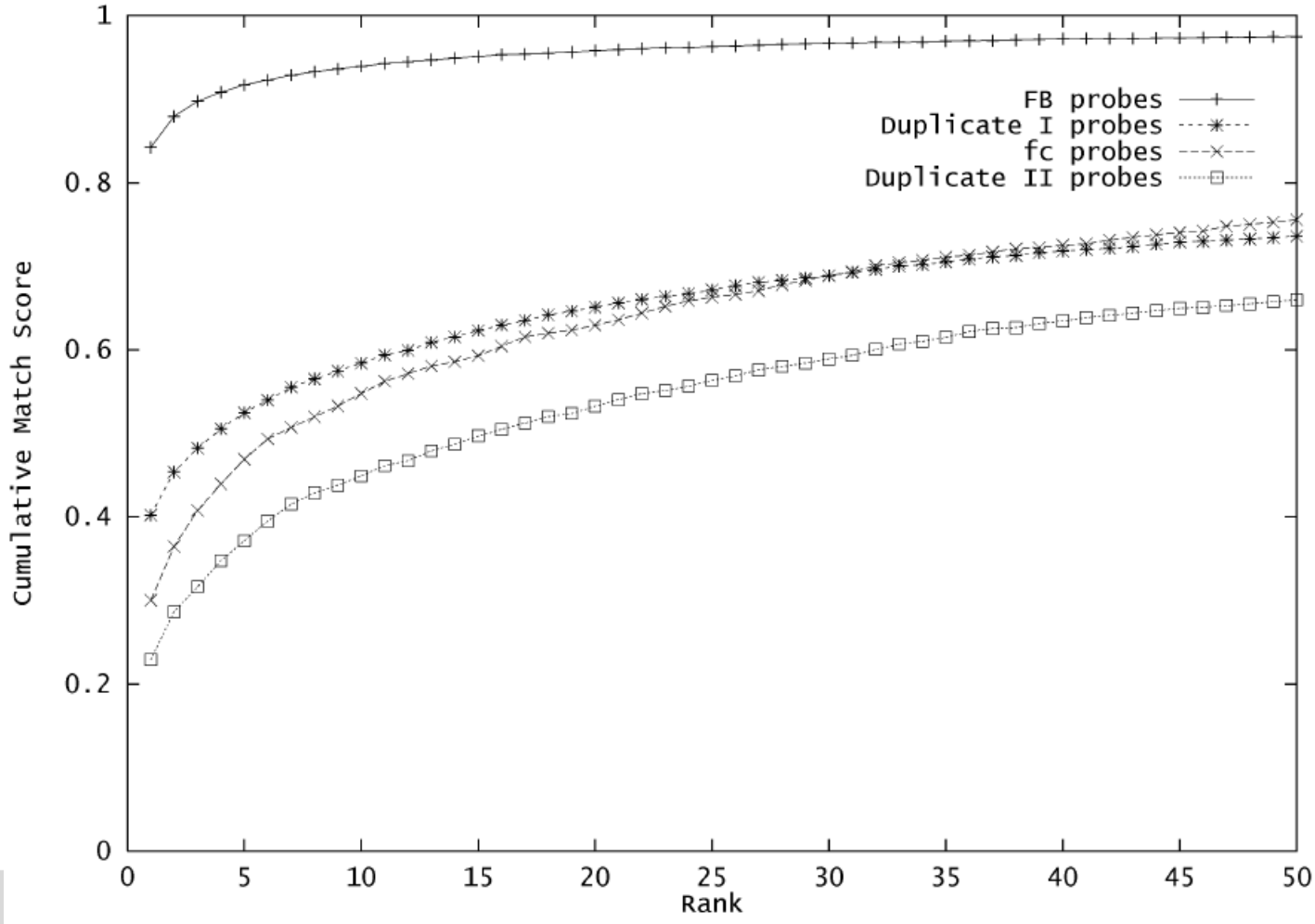
dentification performance against “duplicate “ images



dentification performance against “duplicate “ images



Average Identification performance on each image category



Face Recognition Vendor Test (FRVT) (2000, 2002, 2006)

Provides independent government evaluations of commercially available and prototype face recognition technologies.

Designed to provide U.S. Government and law enforcement agencies with information to assist them in determining where and how facial recognition technology can best be deployed.

The results also help to identify future research directions for the face recognition community.

Face Recognition Vendor Test (FRVT) (2002)

The database:

- 121589 facial images of 37437 individuals

10 commercial firms participated

The impact of three recent techniques for improving face recognition are also assessed:

- 3D morphable models
- Normalization of similarity scores
- Face recognition from video sequences

Face Recognition Vendor Test (FRVT) (2002)

Evaluation parameters:

- Identification performance as a function of database size.
- The variability in performance for different groups of people.
- Performance as a function of elapsed time between enrolled and new images of a person.
- The effect of demographics on performance.

Face Recognition Grand Challenge (FRGC) (2004-2006)

The goal:

to promote and advance face recognition technology designed to support existing face recognition efforts in the U.S. Government.

The ways to improve the face recognition algorithms:

- High resolution images
- 3D facial data
- New preprocessing techniques

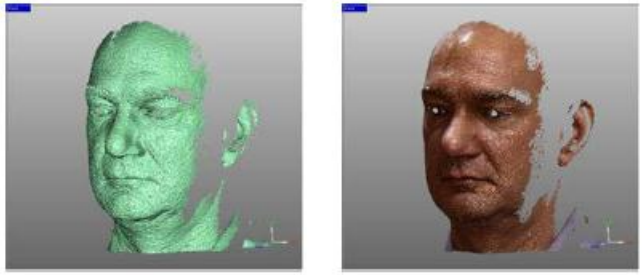
Karlsruhe Institute of Technology

Sample Images FRGC

Controlled conditions

Uncontrolled conditions

3D Shape and Texture



Observations

One 3-D image is *more powerful* for face recognition than one 2-D image.

One high resolution 2-D image is *more powerful* for face recognition than one 3-D image.

Using 4 or 5 well-chosen 2-D face images is *more powerful* for face recognition than one 3-D face image or multi-modal 3D+2D face.

CHIL/CLEAR Face Recognition Evaluations (2004-2007)

Video-based face recognition evaluations

Low resolution, multi-camera data

7 individuals, single-frame-to-video matching (2004, 2005)

> 25 individuals, video-to-video matching (2006, 2007)

Fully-automatic system comparison (Automatic detection of faces, features) (2004, 2005)

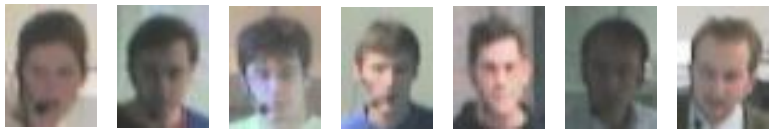
Semi-automatic system comparison (Face bounding boxes and feature locations are provided) (2006, 2007)

CHIL/CLEAR Evaluations (2004-2007)

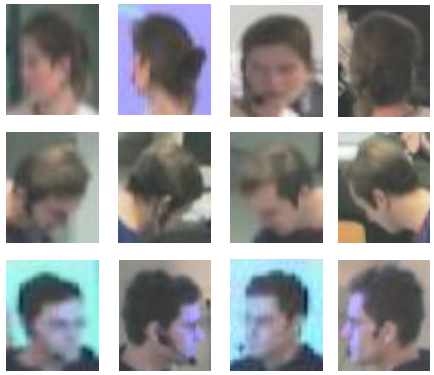
Sample data, 4 camera views

Sample Images from CHIL DB

Sample training face images



Test samples at the same instant from different cameras



Sub-sampled testing sequence from a single camera