Face Recognition: Methods and Practice

Stan Z. Li

Center for Biometrics and Security Research (CBSR) &
National Lab of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences

ICB 2012 Tutorial
Delhi, India

Outline

- Introduction to Face Recognition
- Subspace Analysis
  - Linear Methods
  - Nonlinear Methods
  - Face Grand Challenges from Subspace Viewpoint
- Face Recognition Methods
  - Face Detection
  - Face Pose Estimation
  - Face Alignment
  - Face Recognition
- Face Recognition by Fusion of 2D+3D
- Face Recognition Using Near Infrared Images
- Heterogeneous Face Recognition
- Face Biometric Anti-spoofing
- Demos – AuthenMetric Face Recognition Systems

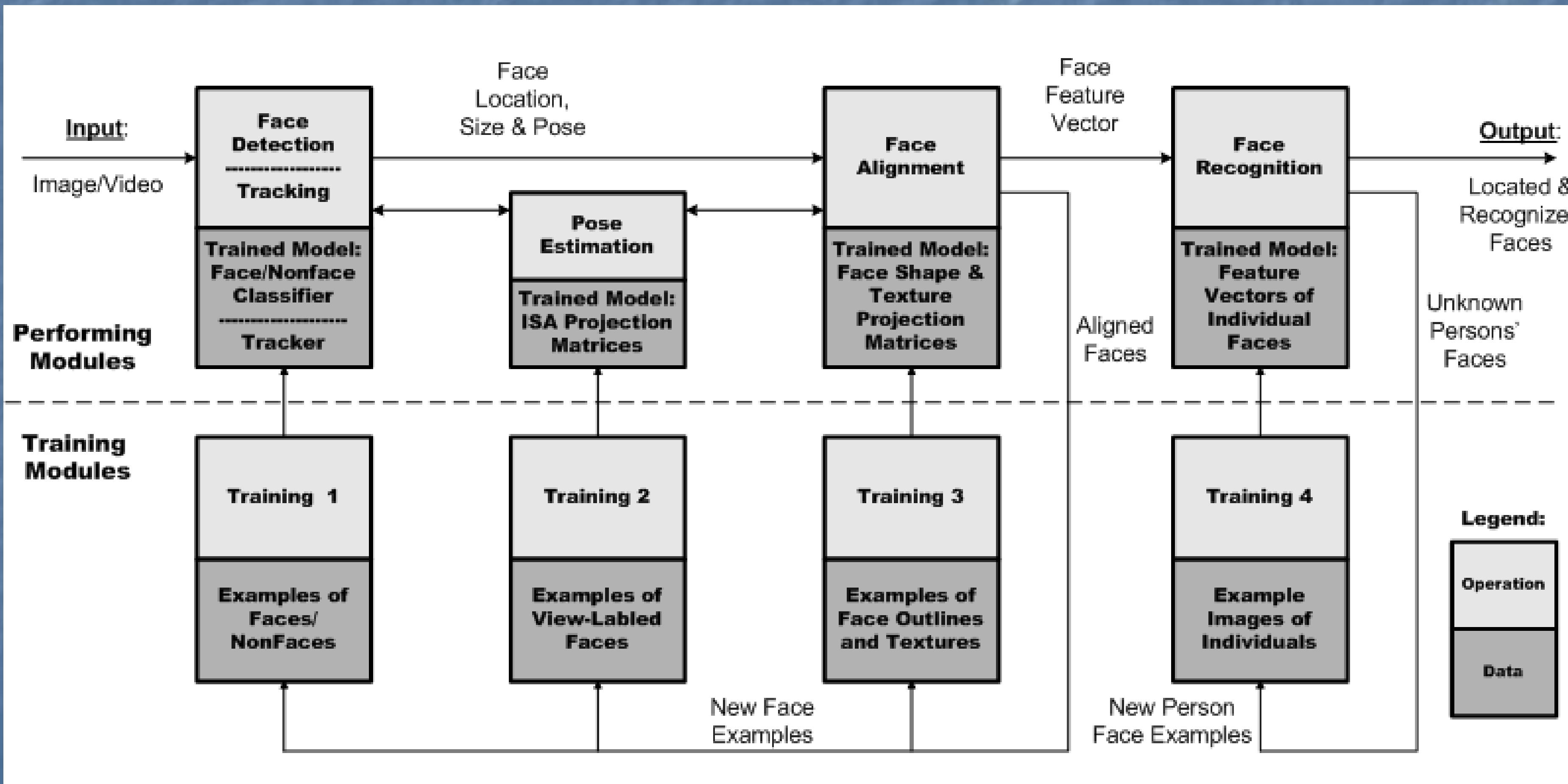
Introduction

Face Recognition Process

1. Face Detection
2. Face Tracking
3. Face Alignment
4. Face Recognition

@ Microsoft Techfest 2002

Face Recognition System



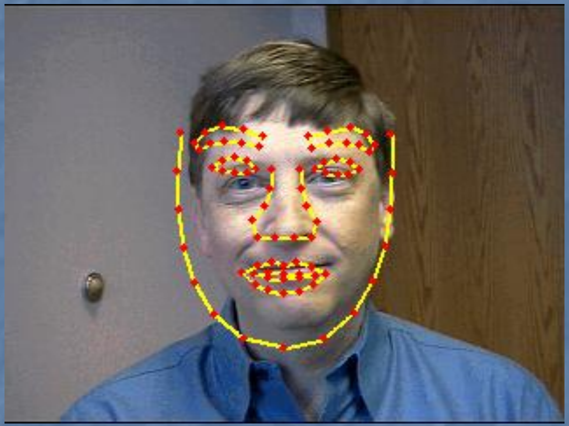
Face Tracking (MSR)

- Input:
  - Video containing moving faces
- Output:
  - Locations, scales, and poses of tracked face

@ Microsoft Techfest 2002

Face Alignment

- Input:
  - Face detection/tracking output (location, scale, and pose)
- Output:
  - Accurate localization of facial landmarks / outline
- Purpose:
  - For geometric normalization towards accurate facial feature extraction



Face Matching

Face Input

- Feature Extraction
- Matching
- Decision (Identification, Verification)

Face Database

Face as Compared to Other Biometrics

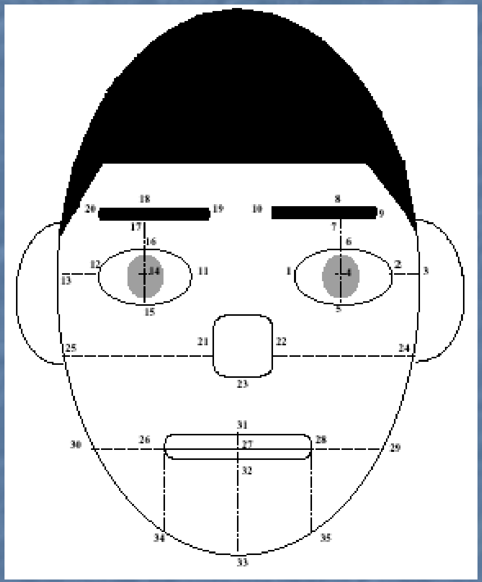
- Universality -- H
- Acceptance -- H
- Easy to acquire-- H
- Permanence -- M
- Reliability -- M-H
- Uniqueness -- L

Face Recognition R&D

- Applied Basic Research
  - Image processing
  - Vision – pose, lighting
  - Pattern recognition
    - Statistical learning
    - Subspaces & manifolds
- Algorithm Research
  - Face detection
  - Face tracking
  - Face alignment
  - Feature extraction
  - Face Matching

- System development
- Application Development

History (60-70's):
Geometric Feature Based Approach



Feature	Distance
1	0.5 * ((1,2) + (11,12))
2	0.5 * ((5,6) + (15,16))
3	(3,13)
4	(24,25)
5	(29,30)
6	(34,35)
7	(26,34)
8	(28,35)
9	(26,28)
10	(27,31)
11	(27,32)
12	(32,33)
13	(23,31)
14	(21,22)
15	0.5 * ((13,25) + (3,24))
16	0.5 * ((25,30) + (24,29))
17	0.5 * ((30,34) + (29,35))
18	0.5 * ((1,22) + (11,21))
19	(10,19)
20	0.5 * ((2,9) + (12,20))
21	0.5 * ((9,10) + (19,20))
22	0.5 * ((11,19) + (1,10))
23	0.5 * ((6,7) + (16,17))
24	0.5 * ((7,8) + (17,18))
25	0.5 * ((18,19) + (8,10))
26	0.5 * ((18,20) + (8,9))
27	(11,23)
28	(1,23)
29	0.5 * ((1,28) + (11,26))
30	0.5 * ((12,13)+(2,3))

Table 1: The 30-dimensional feature vector.

- In traditional AI-CV framework
- Image features pre-specified
- Features=

  {type, locations, distances}

History (90's -):
Learning-Based, Subspace Analysis Approach

- Different from the AI-CV approach
  - Example-based
  - Features Learned
  - Dimension reduction
  - Linear mapping from high-dim to low-dim spaces
- Linear Subspace Methods: Eigenface (PCA) and Others
  - Face Representation: Kirby & Sirovich. 1990.
  - Face Recognition: Turk & Pentland. 1991.
- Nonlinear Methods
  - (More contemporary work).

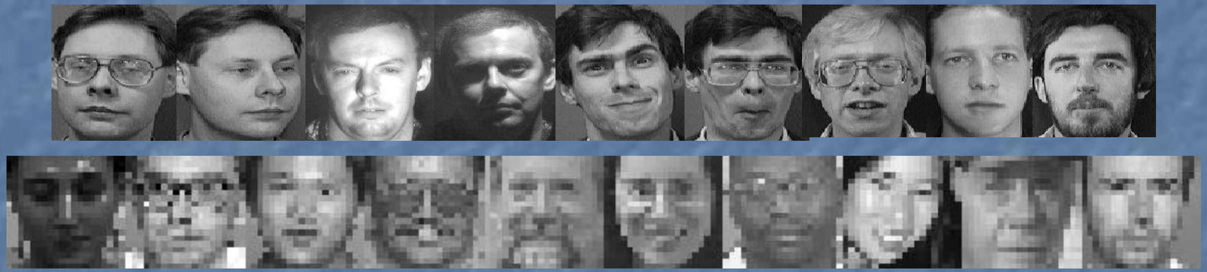
- ....

Year 2002: EyeCU at MSR Techfest

Year 2005: AuthenMetrics of CBSR

- E-Passport at China-Hong Kong Boarder
- E-Passport at China-Macau Boarder
- Access-control in “Beijing”
- Others

Challenges in Face Recognition



- Image Changes due to Variations in
  - Geometry (Head pose, Facial expression)
  - Photometry (Illumination, Camera properties)
- Other Variation Factors
  - Aging, Facial hair, Cosmetics, Accessories (eyeglasses, etc)

Outline

1. Introduction to Face Recognition
2. Linear Subspace Analysis
3. Nonlinear Subspace Analysis
4. Research in MSRA Face Group
5. Face Recognition Evaluation
6. Future Perspectives

Linear Subspace Analysis

Subspace Modeling
Dimension Reduction
Feature Extraction

- Eg: Images of size 64x64
- Dimensionality of image space: 64x64=4096 (pixels)
- Pixel values in {0,…,255}
- 256^4096 > 10^9864 possible configurations in 4096-dim hypercube
- Face pattern living in low dim subspace
- Dimension reduction (features = projected coordinates)

Dimension Reduction

Given $\{x_i \text{ in } \mathbb{R}^N \mid i=1,\ldots,K\}$, find

1. A space $\mathbb{R}^n$
2. Dimension Reduction Mapping
   $y=F(x): \mathbb{R}^N \rightarrow \mathbb{R}^n$
3. Reconstruction Mapping (Smooth, Non-Singular)
   $x=f(y): \mathbb{R}^n \rightarrow \mathbb{M}$ (a manifold in $\mathbb{R}^N$)

Such that

1. $n<N$ as small as possible
2. $\mathbb{M}$ approximately contains $\{x_i\}$
   (reconstruction error is small)

Note that f(F(x)) needs not be x (identity mapping)

Linear Subspace Analysis

- Linear projection: n dim → m dim, m<n
  - h=Px   x: nx1, P: mxn, h: mx1
- Reconstruction x=Bh,   B: nxm
- Matrix Factorization
  - $X \approx BH, \quad X = (x_1, \ldots, x_N), H = (h_1, \ldots, h_N)$
    where X: nxN, B: nxm, H: mxN
  - Principal component analysis (PCA)
  - Vector Quantization (VQ)
  - Independent component analysis (ICA)
  - Non-Negative Matrix Factorization (NMF, LNMF)
  - Linear Discriminant Analysis (LDA)

PCA, VQ, NMF, and LNMF

$$X \approx BH$$

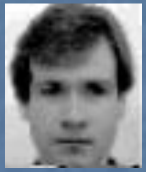
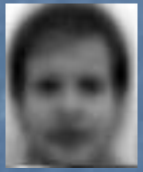
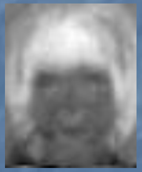
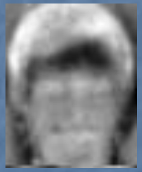
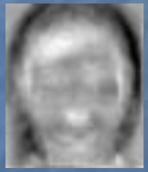
Method	Constraints
PCA	b orthonormal vectors
VQ	h unary vectors
ICA	h independent
NMF	b,h non-negative vectors
LNMF	b,h non-negative + h sparse → b ***really*** part-based

PCA Representation

- Basis vectors = Principal eigenfaces

1st 2nd 3rd 4th 5th 6th 7th

- Face as linear combination of eigenfaces

 = 0.9569 X  - 0.1945 X  + 0.0461 X  + 0.0573 X  - ....

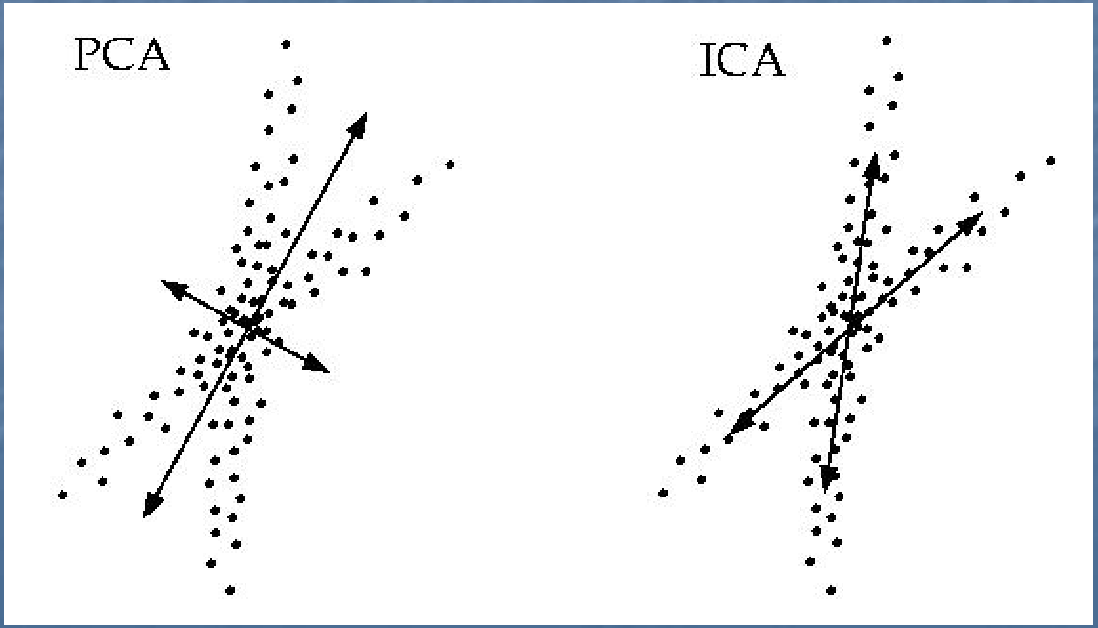
Original Face 1st 2nd 3rd 4th

Y=[0.9569, - 0.1945 , 0.0461, 0.0573, …,]

Independent Component Analysis

$$X \approx BH, \quad X = (x_1, \ldots, x_N), H = (h_1, \ldots, h_N)$$

H components as independent as possible



Learning View-Subspaces by Using PCA, ICA, ISA, TICA (Li et al 2001)

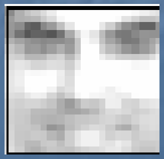
	PCA	ICA	ISA	TICA
view-specific	n	Y	Y	Y
View-grouping	n	n	Y	Y
View-ordering	n	n	n	Y

Non-negative Matrix Factorization

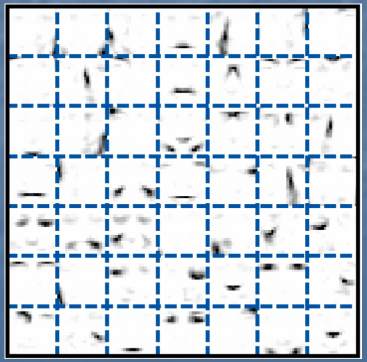
- Papers:
  - Lee and Seung, *Nature* , 1999
  - Lee and Seung, *NIPS*, 2001.
- Non-negative Matrix Factorization $X \approx BH$

  min D(X||BH), s.t. B,H >=0 and $\sum_i b_{ij} = 1$ for all *j*

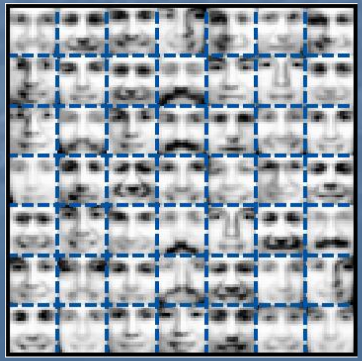
Basis Components learned by different methods



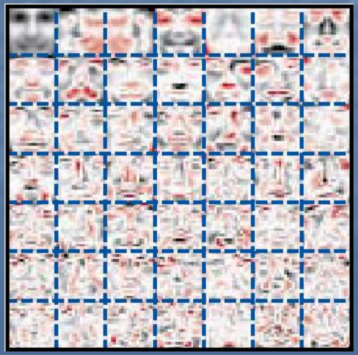
Training Example



NMF



VQ

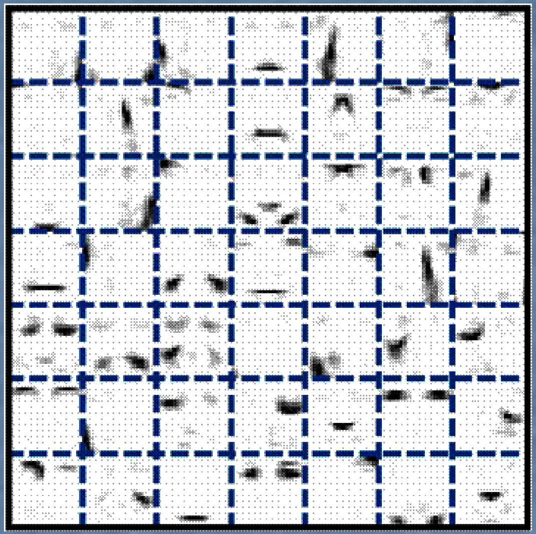


PCA

Problems with NMF

NMF Results Learned From:

Lee-Seung's Data



ORL Data

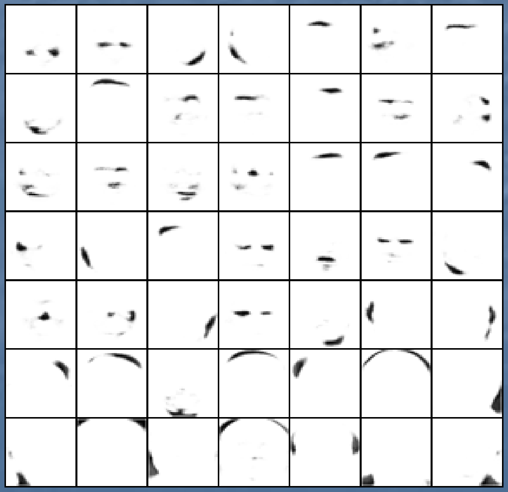
Our Data

1. Learned components not really localized, part-based
2. Face recognition not very good

Local Non-negative Matrix Factorization

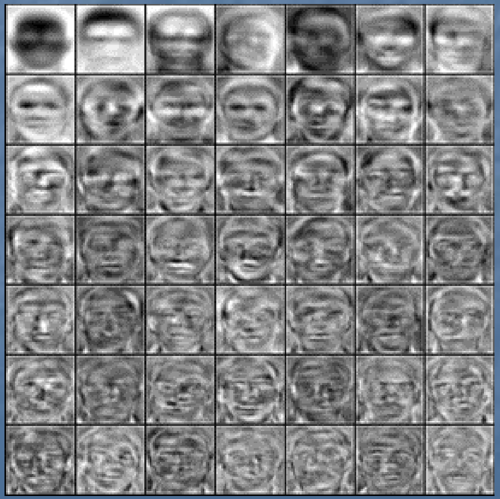
- Additional constraints imposed on NMF
  for spatially localized, part-based representation

Comparative results learned from ORL data:



LNMF

NMF



PCA

Additional Constraints

$\mathrm{X} \approx \mathrm{BH}$. Let $U = [u_{ij}] = B^T B,\ V = [v_{ij}] = HH^T$

1. Maximum Sparsity in *H*. A basis component $\mathbf{b}_j$ should not be further decomposed into more components. Given $\sum_i b_{ij} = 1$, $\|\mathbf{b}_j\| = u_{jj} = \sum_i b^2{}_{ij}$ should be minimized for all *j:*

$$\sum_j \|\mathbf{b}_j\| = \sum_j u_{jj} = \min$$

2. Maximum total activity. Retain most expressive components:

$$\sum_i \sum_j h^2{}_{ij} = \sum_i v_{ii} = \max$$

3. Orthogonality of basis:

$$\sum_{i \neq j} u_{ij} = \min$$

Learning by Constrained Optimization

- NMF

  min $D(X \| BH) = \sum_{i,j} X_{ij} \log X_{ij} / (BH)_{ij} - X_{ij} + (BH)_{ij}$

  s.t. B,H>=0, $\sum_i b_{ij} = 1$

- LNMF = NMF + localization constraints

  min $L(X, BH) = \sum_{i,j} X_{ij} \log X_{ij} / (BH)_{ij} + \alpha \sum_{i,j} (B'B)_{ij} - \beta \sum_i H_{ii}^2$

  s.t. B,H>=0, $\sum_i b_{ij} = 1$

LNMF Learning Algorithm

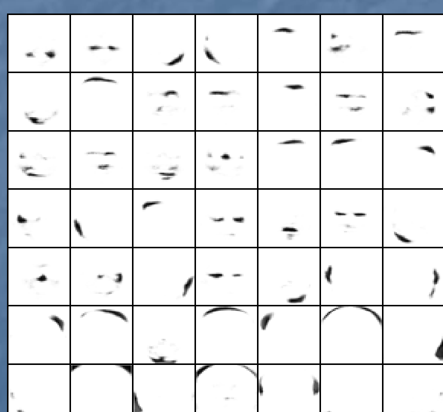
1. $$h_{kl} = \sqrt{h_{kl} \sum_i x_{il} \frac{b_{ik}}{\sum_k b_{ik} h_{kl}}}$$

2. $$b_{kl} = \frac{b_{kl} \sum_j x_{kj} \frac{h_{lj}}{\sum_k b_{kl} h_{lj}}}{\sum_j h_{lj}}$$

3. $$b_{kl} = \frac{b_{kl}}{\sum_k b_{kl}}$$

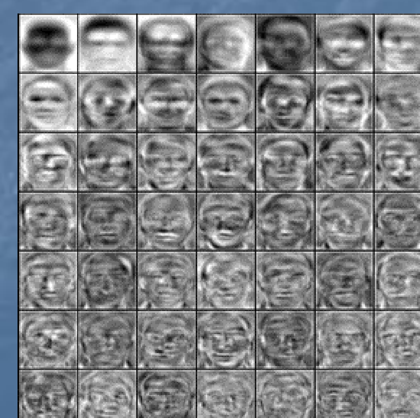
Convergence proved.

Comparison of PCA, NMF and LNMF Bases



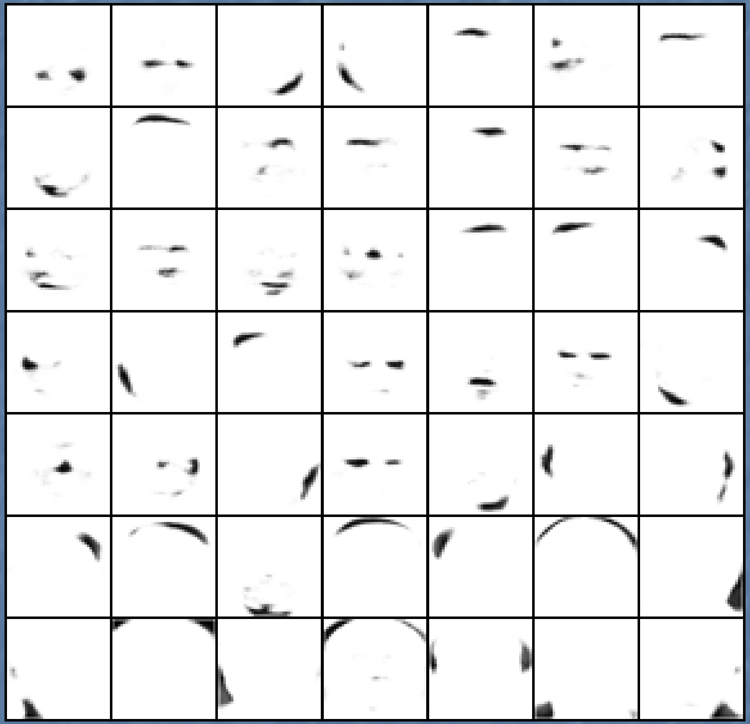
LNMF

NMF



PCA

LNMF vs NMF



LNMF

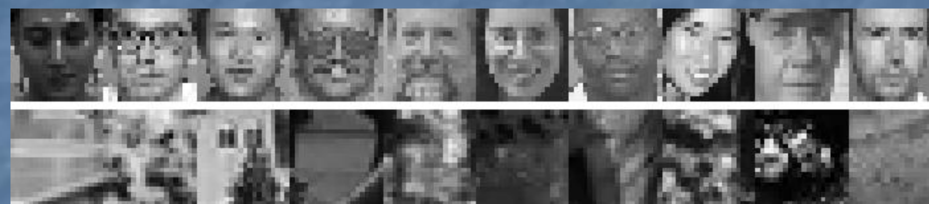
NMF

Nonlinear Subspace Analysis

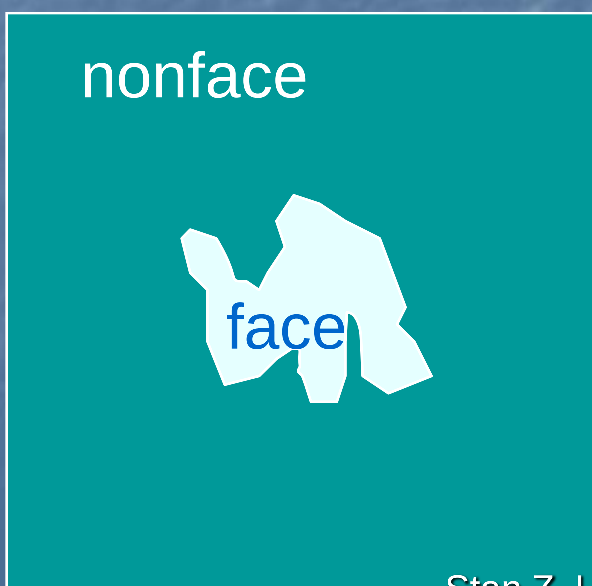
Face Detection and Recognition - From Manifold Viewpoint

Faces

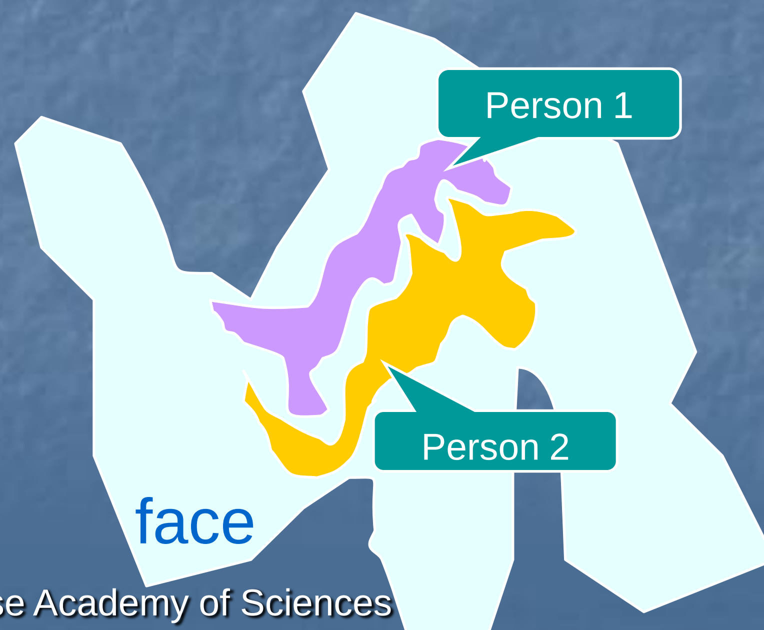
Nonfaces



Detection



Recognition



Beyond Linear Subspaces

- Face subspaces are nonlinear manifolds
  - Manifolds of faces and nonfaces (detection)
  - Manifolds of different persons (recognition)
- Nonlinear Separability of face manifolds
  - Faces/Nonfaces are separable in image space
  - Face 1 / Face 2 / … / Face N are also separable
  - Yet, highly nonlinear and interweaving

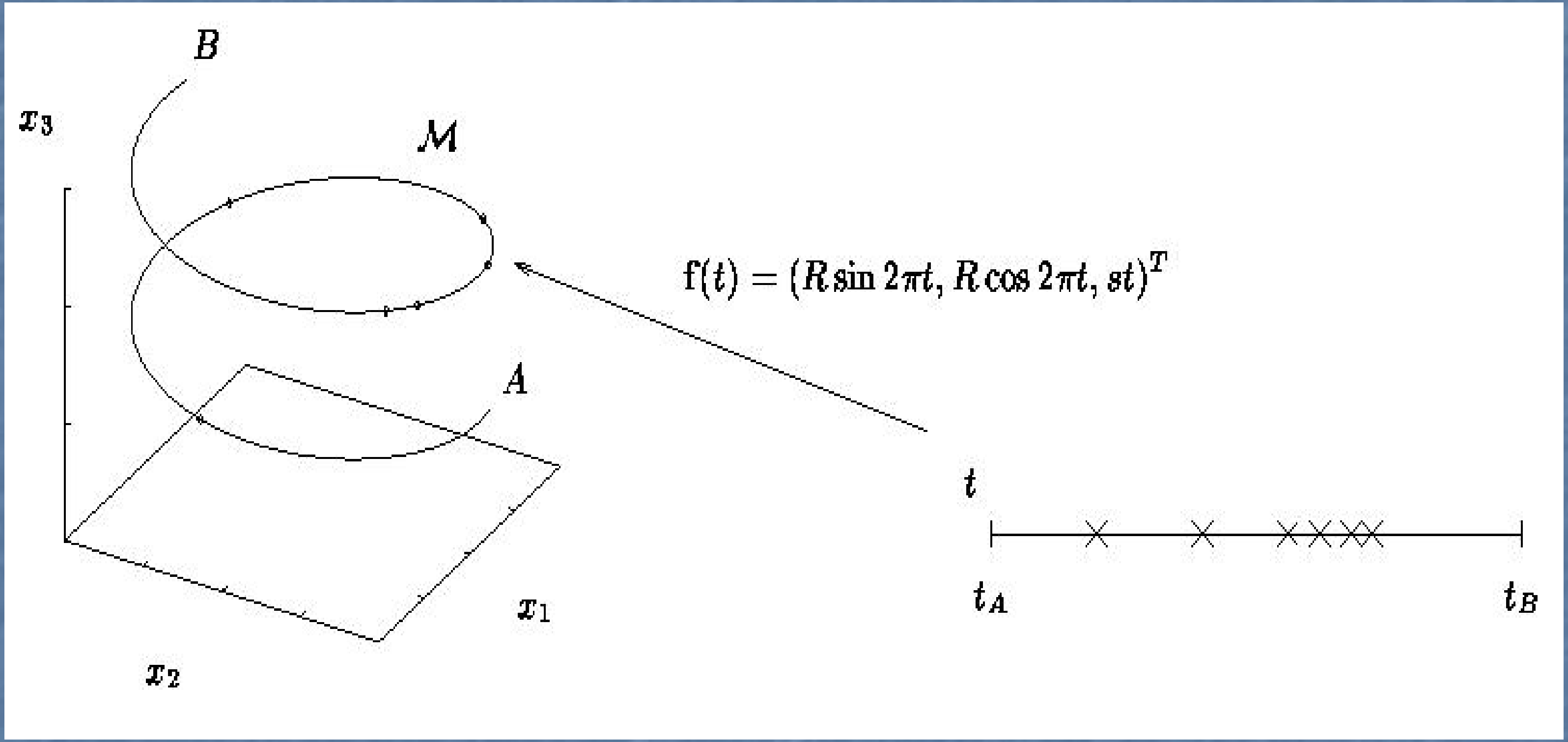
Dimensionality of Nonlinear Subspace

- Linear Space: Spanning dimension (basis dimension)
- Intrinsic dimension (latent dimension): the smallest number of parameters to model the data without loss. If it is *d*, then the observation (of dimension *n*>*d*) is generated by

$$\mathbf{X}=\mathbf{H}(v_1,\ldots,v_d)$$

- Topological space (local dimension): the basis dimension of the local linear approximation of the hypersurface on which the data resides.

Nonlinear Subspace Illustration:



Linear projection onto the x_1-x_2 subspace (2-D) as a circle.

Nonlinear projection onto the t subspace (1-D) as a straight line.

Intrinsic Dimension Estimation

- Using Neighborhood Information (Jain, Dubes and Students. IEEE-PAMI. 1979).
- Packing Number Methods (Kegl 2002)

correlation dimension *of S is defined as*

$$D_{\text{corr}} = \lim_{r \to 0} \frac{\log C(r)}{\log r}.$$

Definition 3 *The* **scale-dependent correlation dimension** *of a finite set* $\mathcal{S}_n = \{x_1, \ldots, x_n\}$ *is*

$$\widehat{D}_{\text{corr}}(r_1, r_2) = \frac{\log C(r_2) - \log C(r_1)}{\log r_2 - \log r_1}.$$

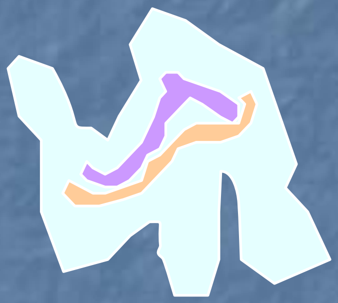
Definition 4 *The* **capacity dimension** *of a subset $\mathcal{S}$ of a metric space $\mathcal{X}$ is*

$$D_{\text{cap}} = -\lim_{r \to 0} \frac{\log N(r)}{\log r}.$$

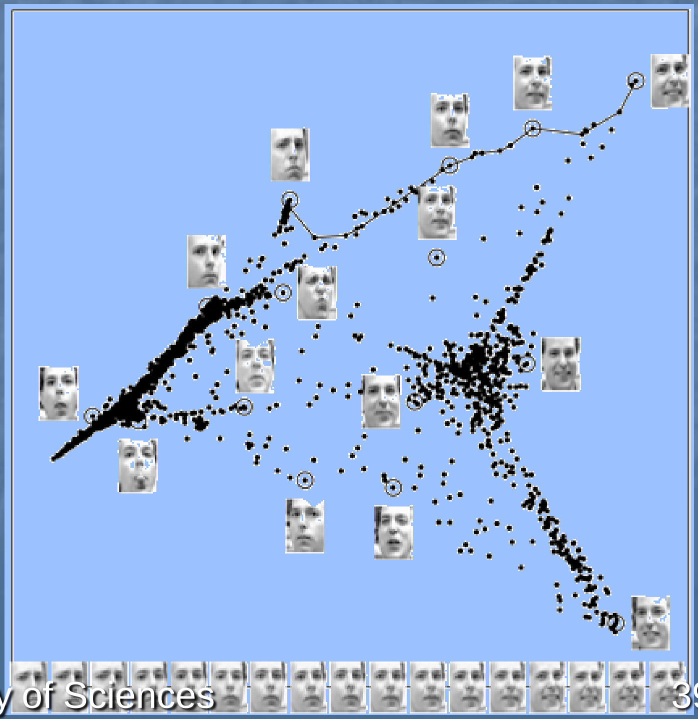
Definition 5 *The* **scale-dependent capacity dimension** *of a finite set* $\mathcal{S}_n = \{x_1, \ldots, x_n\}$ *is*

$$\widehat{D}_{\text{cap}}(r_1, r_2) = -\frac{\log M(r_2) - \log M(r_1)}{\log r_2 - \log r_1}.$$

Nonlinear Subspace Analysis



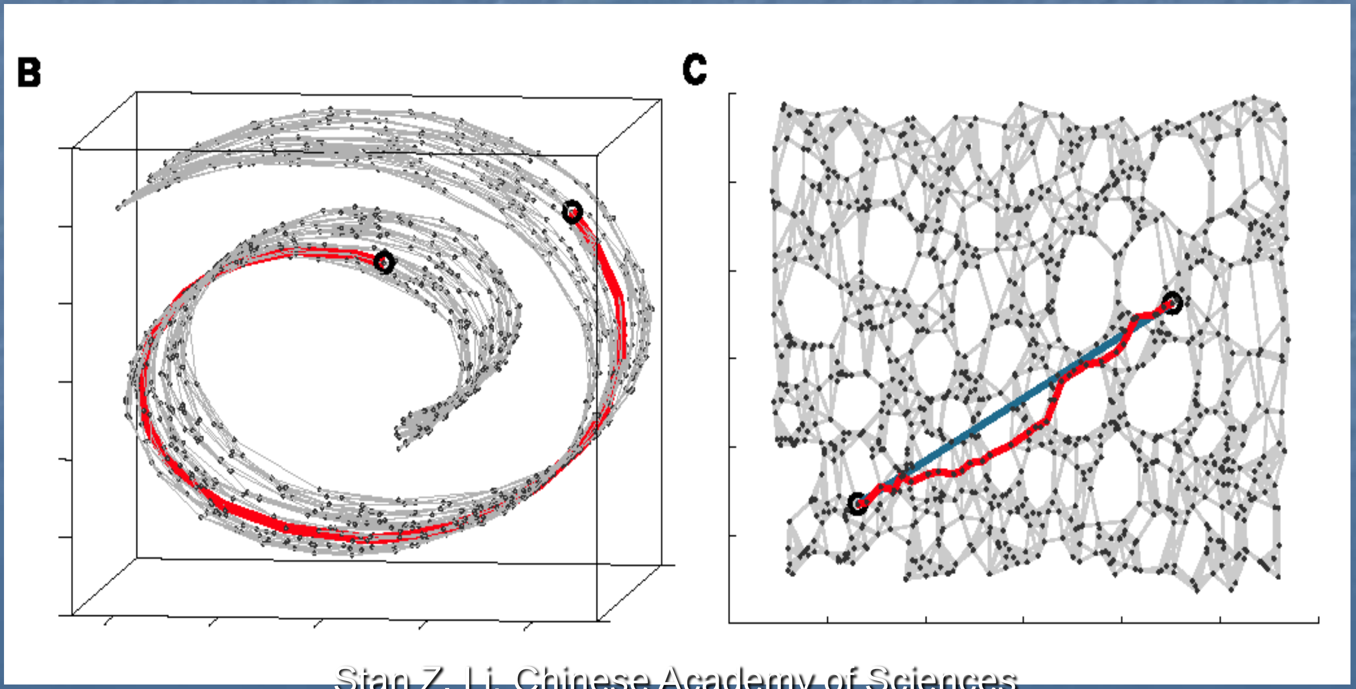
- Recent Advances
  - ISOMAP (*Science*, 2000)
  - LLE (*Science*, 2000)
  - Laplacian EigenMap (*NIPS*, 2001)
- Properties
  - Count for interaction within neighborhood
  - Non-orthogonal projections
- Advantages
  - Lead more sensible modeling
  - Discover intrinsic dimensions
- Disadvantages
  - Increased computation
  - Need to work out a mapping



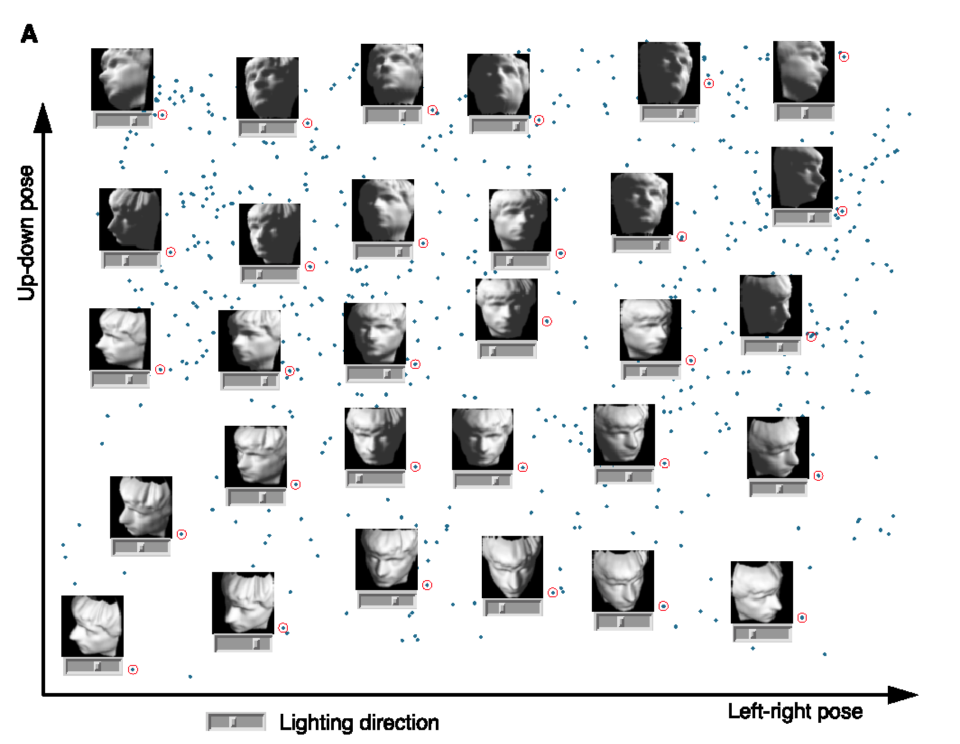
ISOMAP

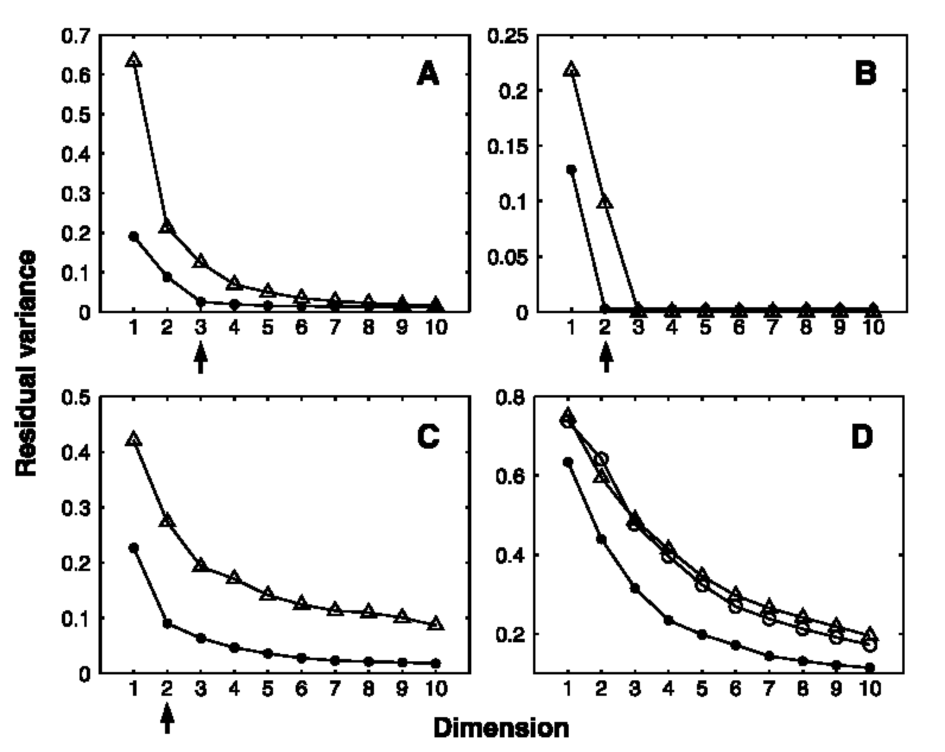
- ISOMAP = Geodesic dist + MDS

• Metric MDS is used to recover parametrizations in lower dimensional space



ISOMAP Results





Locally Linear Embedding (LLE)

- In the high dim space find W

$$\vec{X}_i = \sum_j W_{ij} \vec{X}_j$$

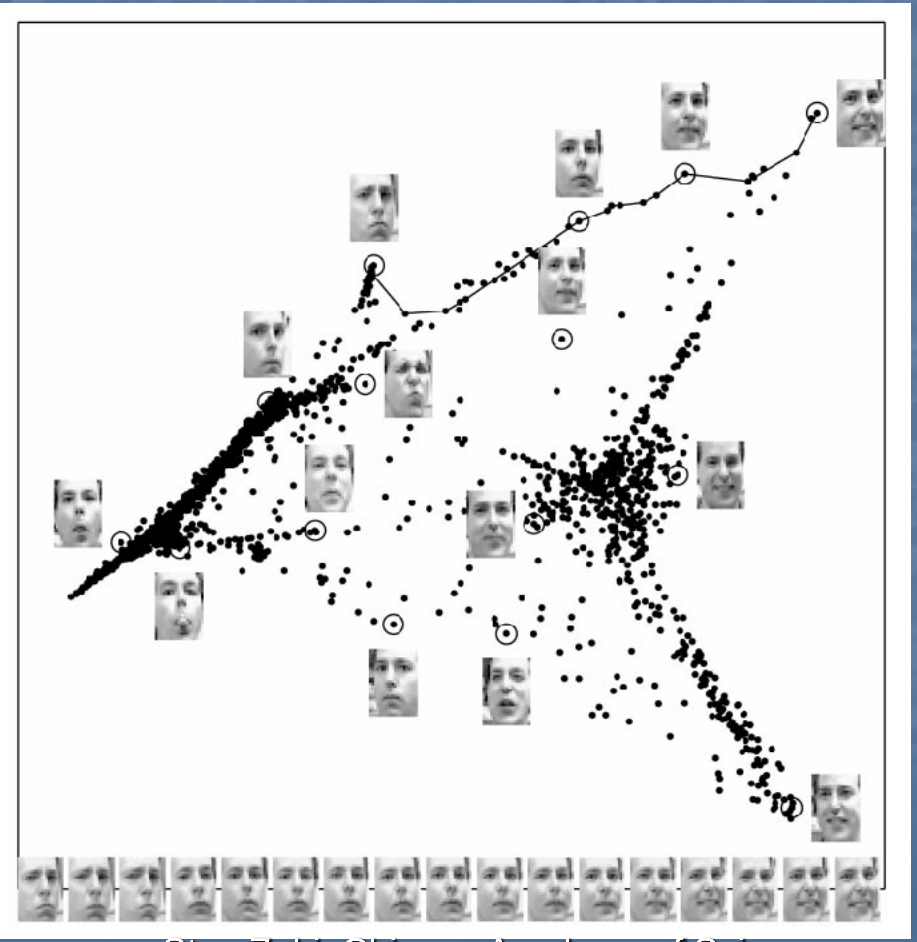
$$\min_W \varepsilon(W) = \sum_i | \vec{X}_i - \sum_j W_{ij} \vec{X}_j |^2$$

s.t. $\sum_i W_{ij}=1$

- In the low dim space find Y

$$\Phi(Y) = \sum_i | \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j |^2$$

LLE Result: Pose and Expression Dimensions



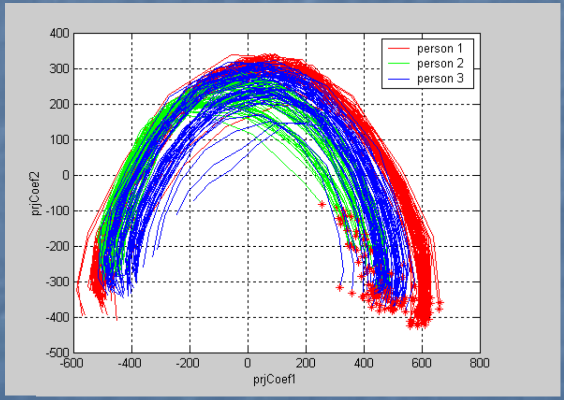
Face Grand Challenges
- From Subspace Viewpoint

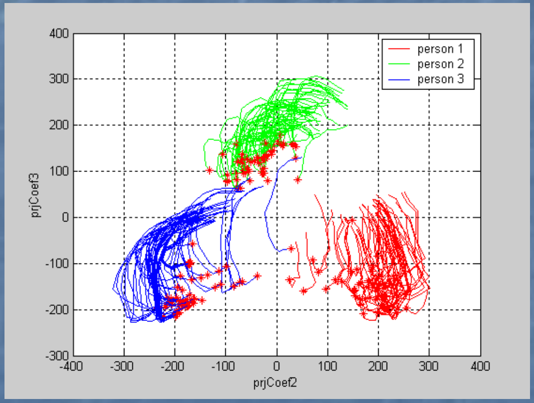
Challenges in Face Recognition

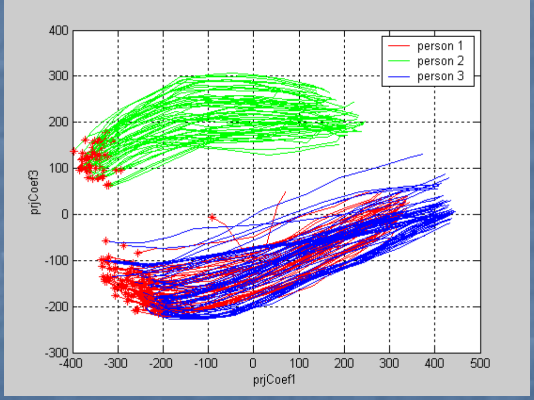
- Complexity of nonlinear face manifolds
- Problem in Generalizing
  - Limited Training Data
  - When lighting changes
  - When pose changes
  - Daily changes and aging
  - When Camera property change
- Euclidean Geometry Inappropriate in image space

Rotated Faces

in PCA Subspace

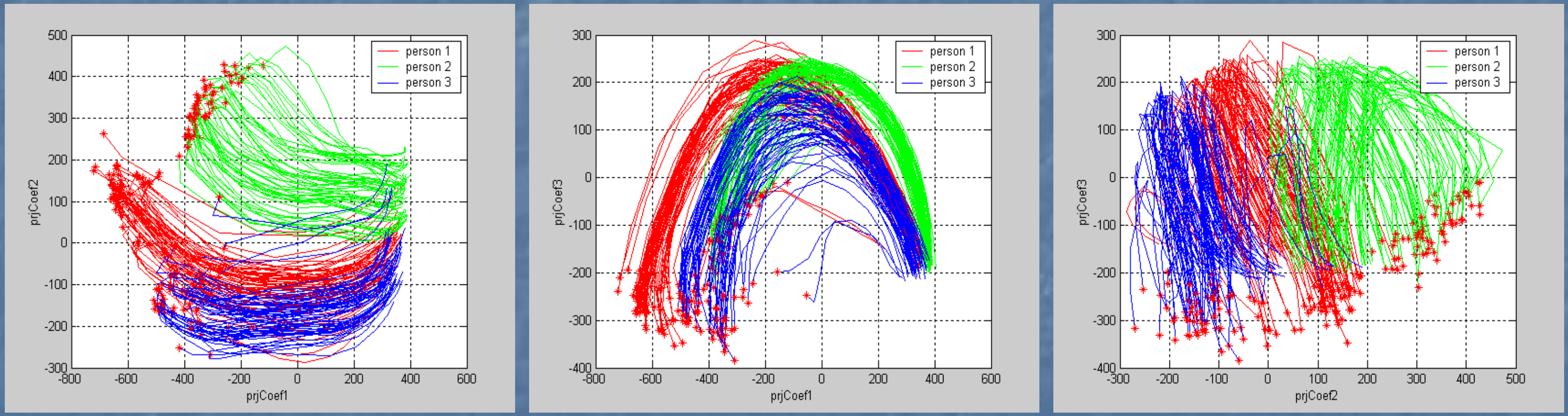






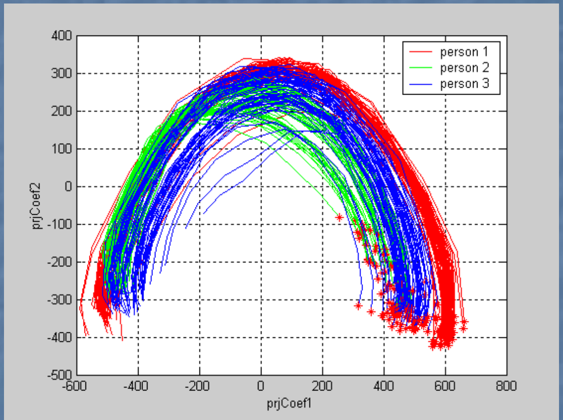
Scaled Faces

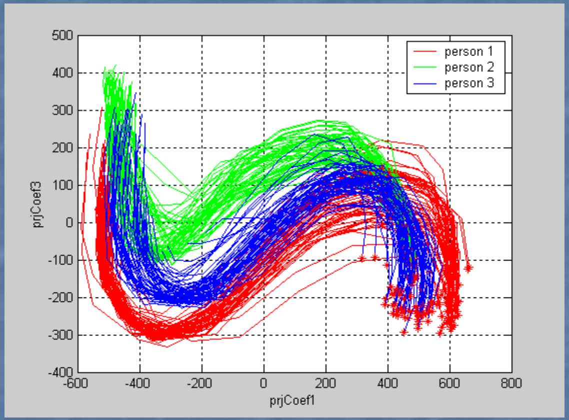
in PCA Subspace

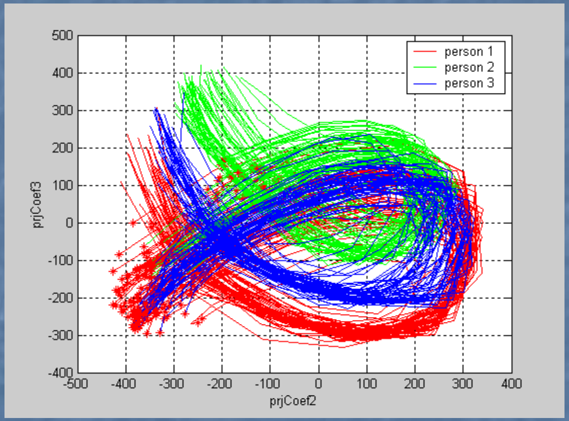


Translated Faces

in PCA Subspace

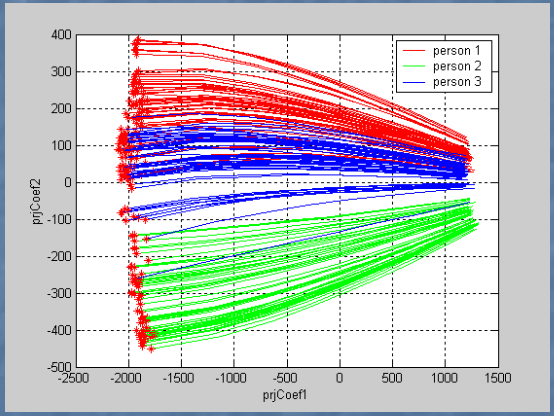


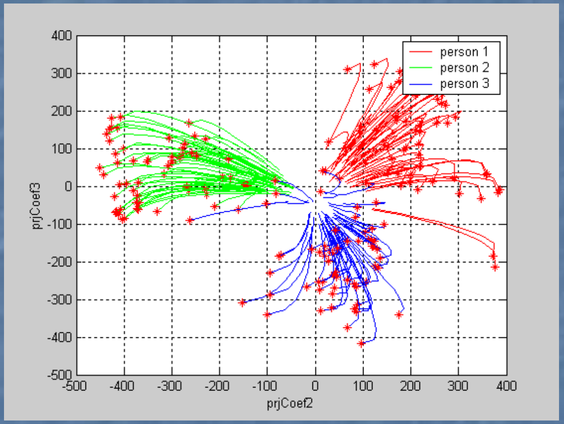


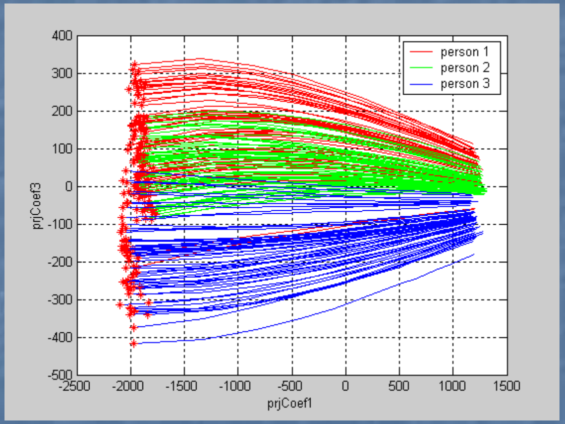


Manifolds are *Folding* and *Interweaving*

PCA Subspace of “Re-Lighted” Faces



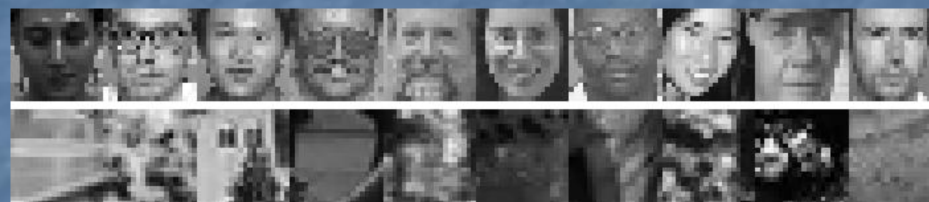




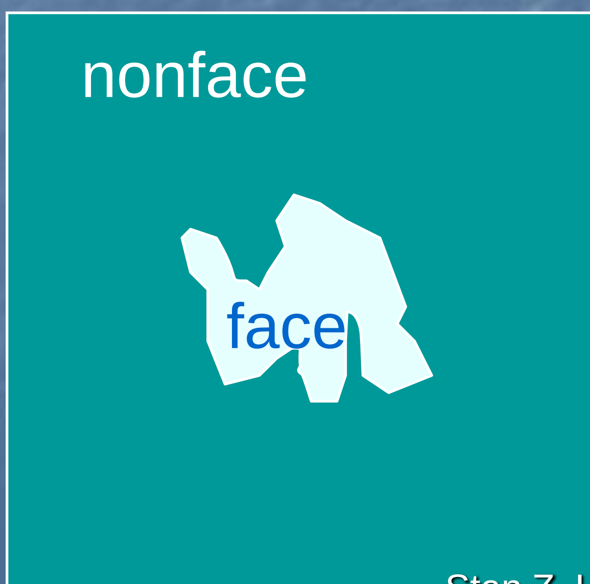
Subspaces in Detection and Recognition

Faces

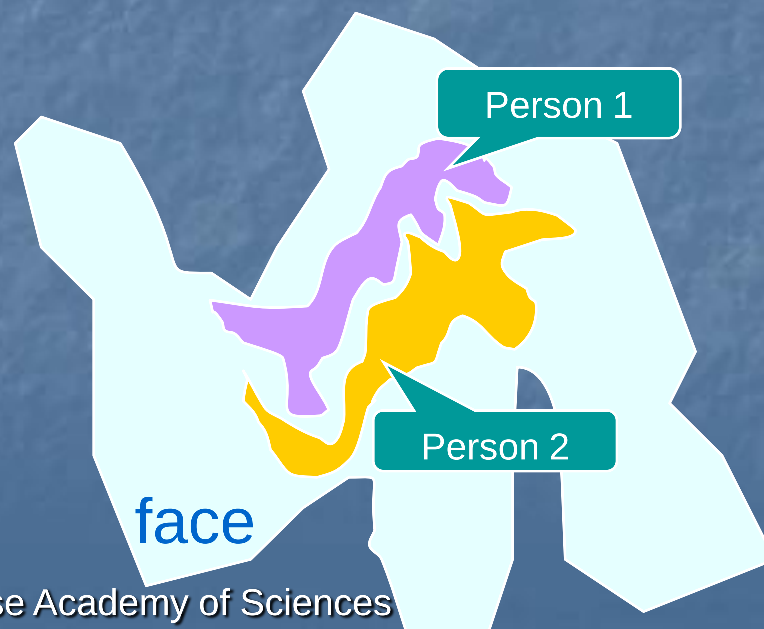
Nonfaces



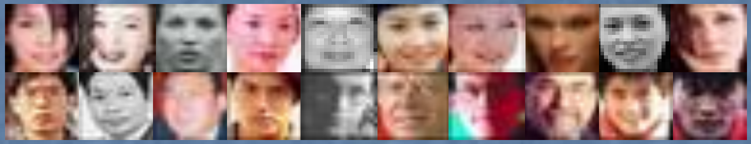
Detection

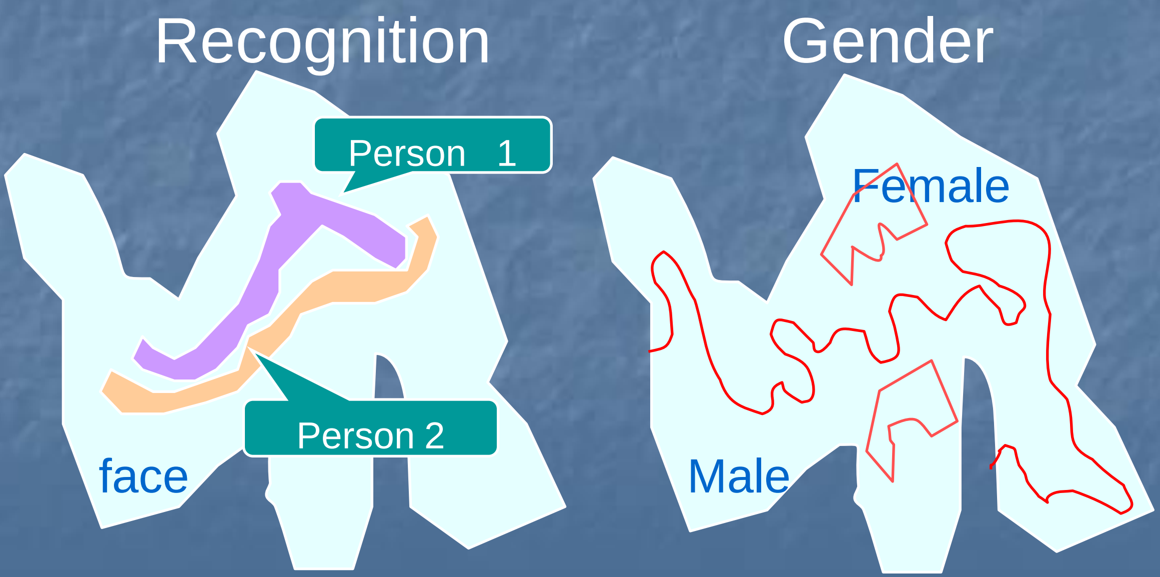


Recognition



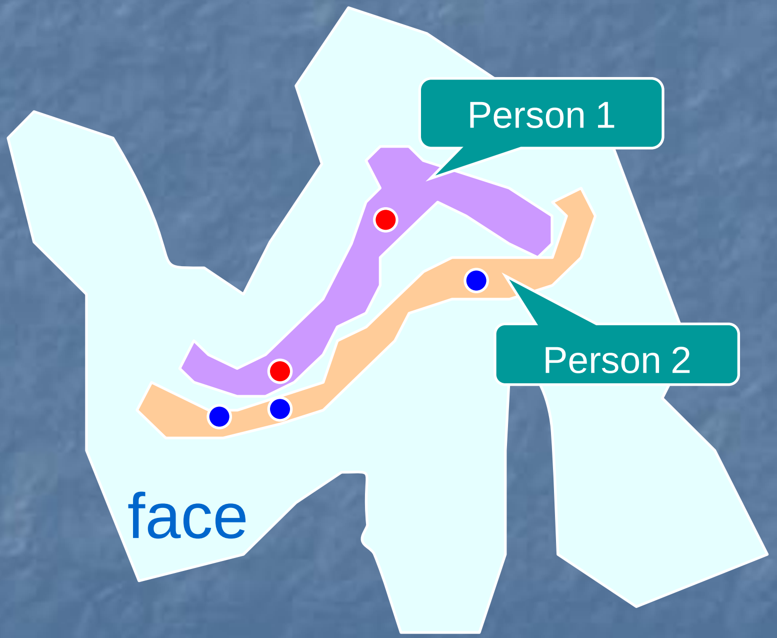
Subspaces in Face Recognition and Gender Classification





Non-Euclidean Geometry

- Euclidean Geometry Inappropriate
- Need to model manifolds in Non-Euclidean Space
- Geodesic distance



Separability in Image and Feature Spaces

- Individual faces Separable in image space
  - Complex, but separable
- Difficult to separate in feature space
  - Overlapping in feature space due to information loss

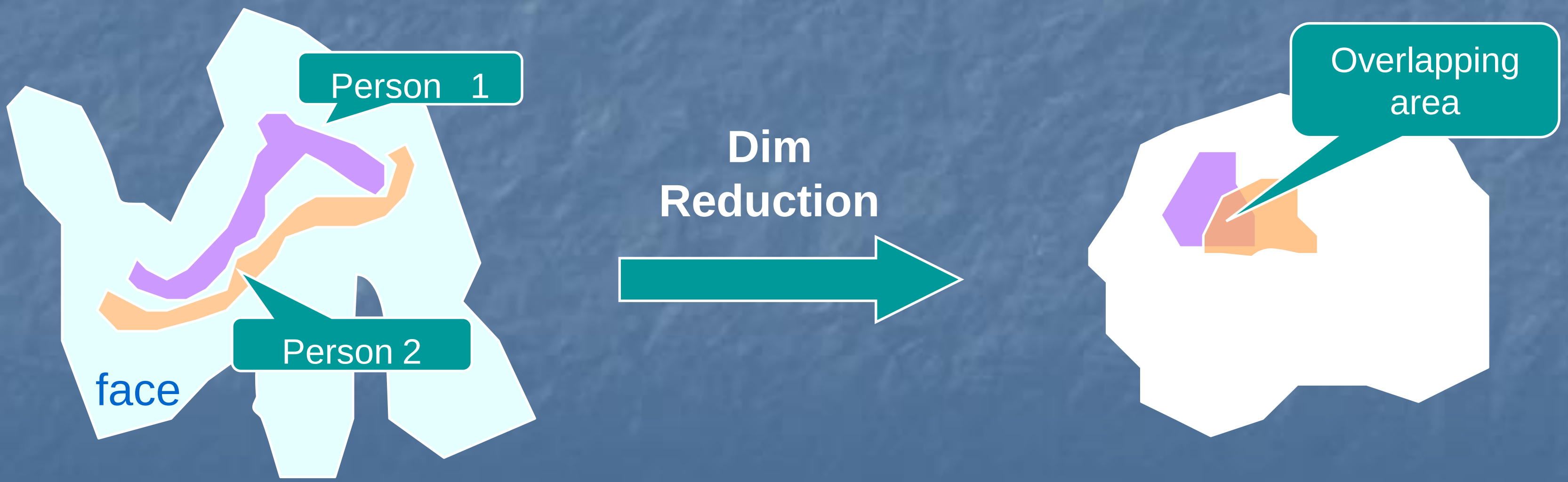


Image Space

Feature Space

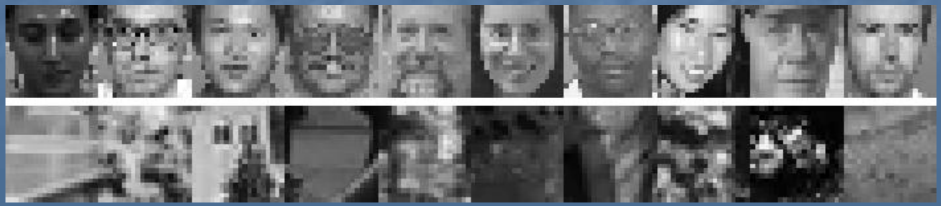
Face Detection

Face Detection: Approach

- Scan the image with subwindows of varying size and location
- Classify a subwindow x into face/nonface
  - Need a "strong classifier" for accurate classification
- Post-processing: Merge multiple detects

Faces

Nonfaces

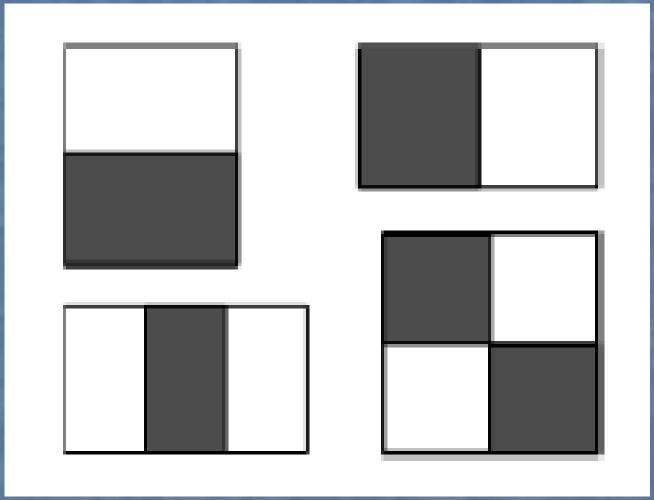


State-of-the-Art Methods: Local Features + Boosting

- Viola & Jones, 2001
  - Haar Features + AdaBoost + Cascade
- Schneiderman & Kanade, 2000
  - Wavelet Histograms
- Li, et al, 2002
  - Extended Haar Features + FloatBoost + Pyramid
- Haizhou Ai, et al, 2003-2005
  - Omni-view face detection, Haar feature + Boosting + More advanced architecture

AdaBoost Method
(Viola & Jones)

Simple Haar features (Viola & Jones)



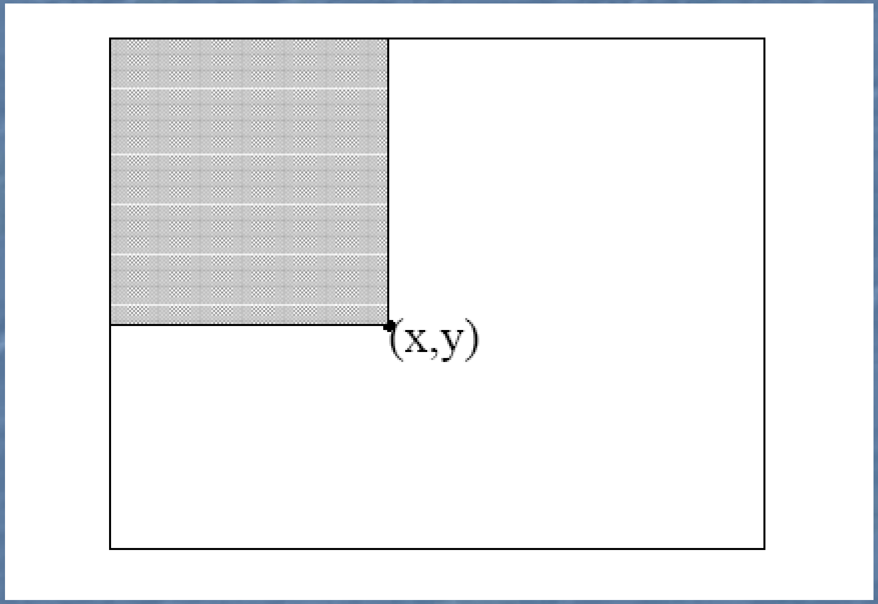
3 rectangular features types:

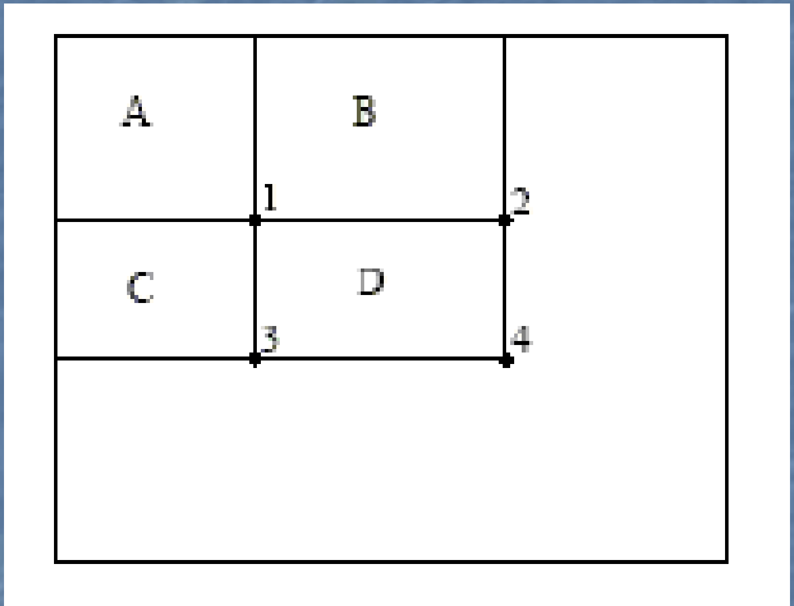
- *two-rectangle feature* type (horizontal/vertical)
- *three-rectangle feature* type
- *four-rectangle feature* type

These rectangular features, as opposed to more expressive steerable filters, can be computed very efficiently using integral images.

Using 24x24 windows → 49,396 features.

Integral Images





AdaBoost Learning

- Proposed by Freund et al 1997, 1998
- Task: Given $\{(x_i, y_i)\}$, learns $H_M(x)$ so that $y_i = \text{sign}(H_M(x))$
- Learns and combines a sequence of weak classifiers $h_m(x)$ into a strong classifier

$$H_M(x) = \sum_{m=1}^{M} \alpha_m h_m(x)$$

- $h_m(x)$ are learned in stages to minimize error bound (see later)

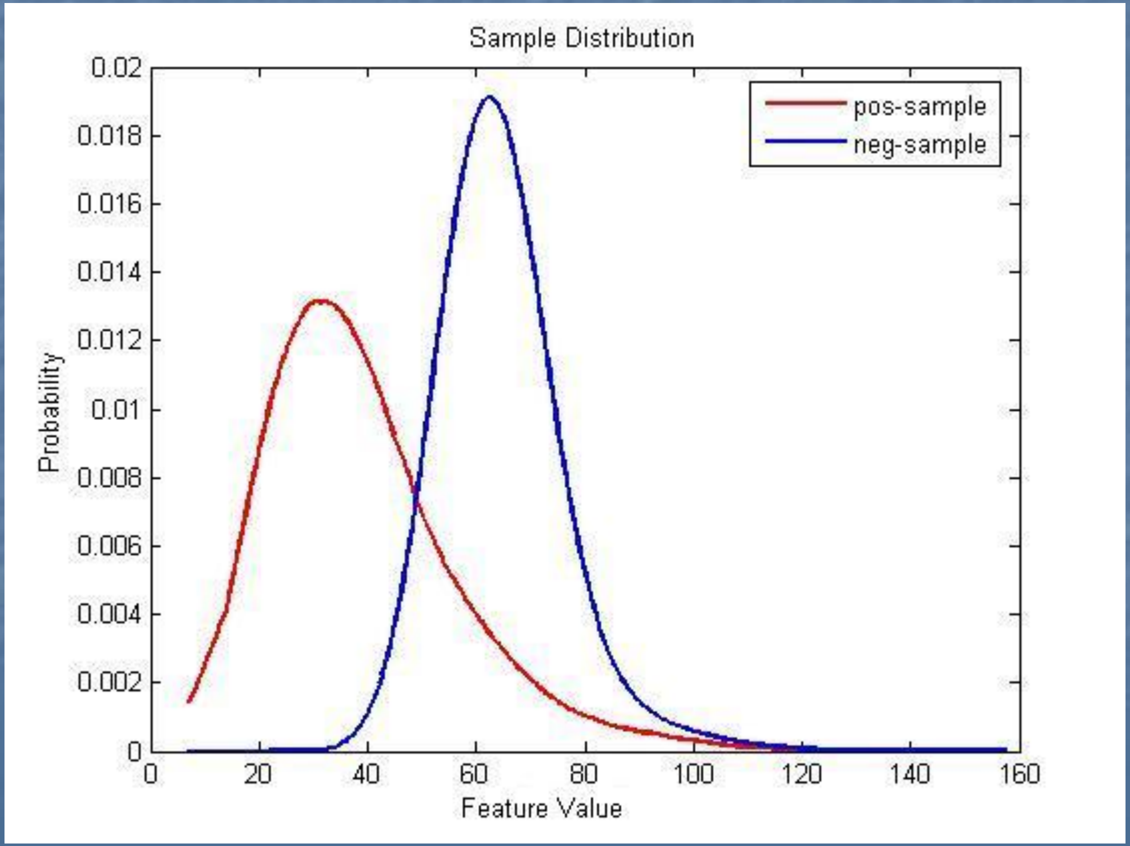
$$J(H_M(x)) = \sum_i \mathrm{e}^{-y_i H_M(x_i)}$$

- Associate (x_i, y_i) with weight w_i and reweight after each iteration (see formula later)

Weak Classifiers

- One WC for a scalar Haar feature
- WC outputs face/nonface by comparing the scalar value with a threshold
- Best threshold obtained by examining the weighted histogram

Learning Weak Classifiers Based on Weighted Histogram

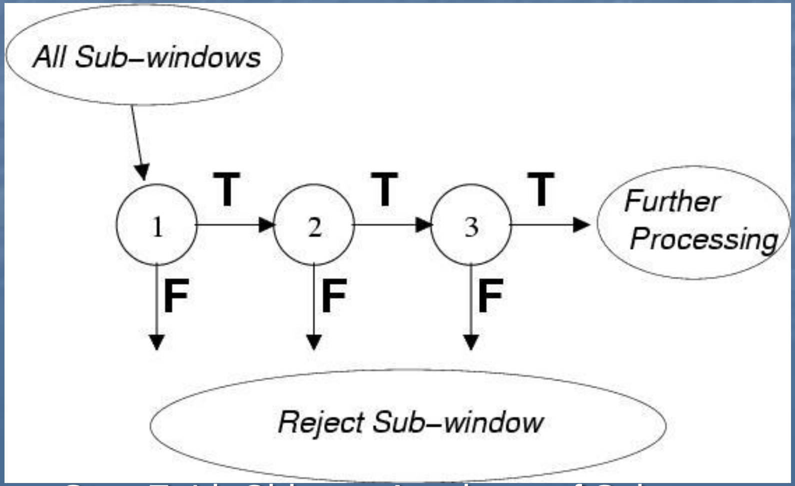


Best Features Learned


- First features selected by AdaBoost are meaningful and have high discriminative power
- By varying the threshold of the final classifier one can construct a two-feature classifier which has a detection rate of 1 and a false positive rate of 0.4.

Speed-up through Cascade

- Simple, boosted classifiers can reject many of negative sub-windows while detecting all positive instances.
- Series of such simple classifiers can achieve good detection performance while eliminating the need for further processing of negative sub-windows.



FloatBoost Method

Li, et al

AdaBoost: Advantages

- Provably effective provided that h_m are "good enough"
- Generally does not overfit
  - Does overfit when data contains outliers
  - a less complex classifier (combining fewer weak classifiers) is preferred
- Simple and easy to program
- Almost no parameters to tune (except M, #WC)

AdaBoost: Problems

1. A sequential, local minimizer
2. May overfit when too many weak classifiers are combined (recent studies)
3. "Detachment" between cost function and error rate
4. Need methods for learning weak classifiers

FloatBoost Project: Objectives

- Better boosting learning: To address (1-3) by incorporating Floating Search (Pudil, Novovicova & Kittler, 1994)
- Weak classifier: For (4), to derive formula for efficient approximation of weak classifier
- Fast multi-view face detection: System

FloatBoost = AdaBoost + FloatingSearch

- Procedure
  1. Boosting to add one weak classifier
  2. If removing a weak classifier leads to a maximum improvement (eg in error rate), remove the weak learner and go to 2
  3. If termination condition not satisfied, go to 1
- Results in a strong classifier of less complexity with improved performance

Learning Weak Classifiers

- RealBoost learns a strong classifier of the form

  $H_M(x) = h_1(x) + h_2(x) + \ldots + h_{M-1}(x) + h_M(x)$

  in stages to minimize the error bound:

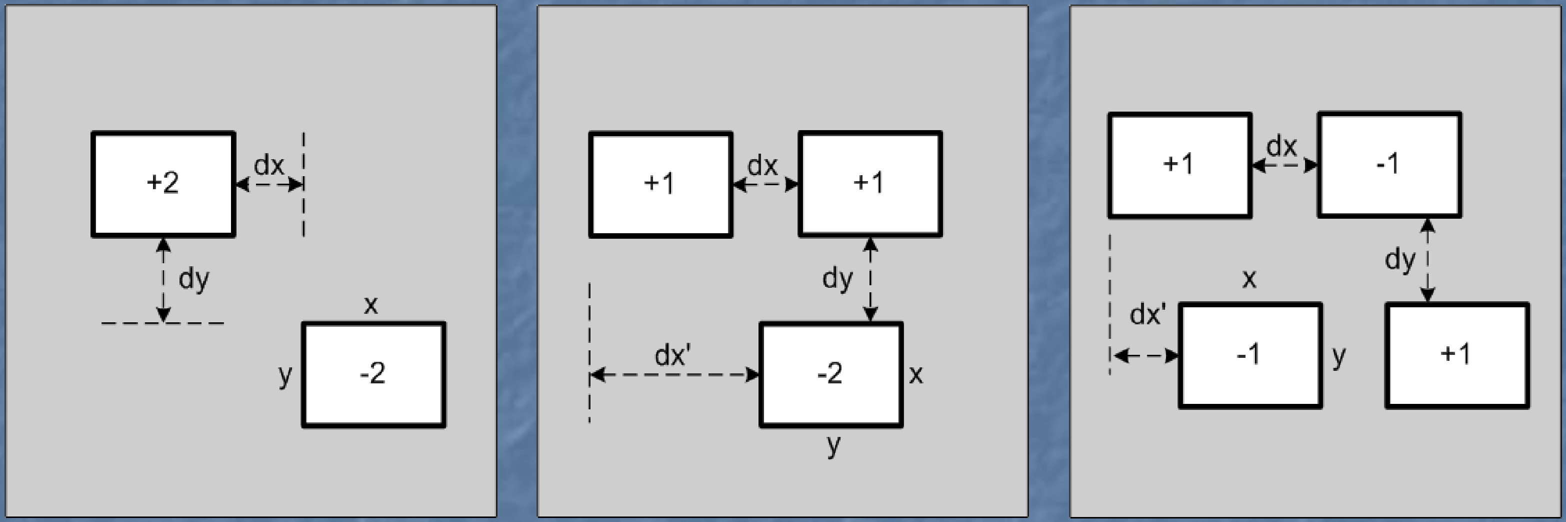
$$J(H_M(x)) = \sum_i e^{-y_i H_M(x_i)}$$

- Given the first M -1 weak classifiers, the best, ideal M-th is derived as

$$h_M(x) = \frac{1}{2} \log \frac{P(y=+1 \mid x, w^{(M-1)})}{P(y=-1 \mid x, w^{(M-1)})} \quad \text{with} \quad w_i^{M-1} = w_i^{M-2}\, e^{-y_i H_{M-1}(x_i)}$$

Extended Haar Features

- Three types of f_k



- A total of K>400,000 such features
- They are overcomplete for representing x

Weak Classifiers

$$h_M(x) = \frac{1}{2}\log\frac{P(y=+1 \mid x, w^{(M-1)})}{P(y=-1 \mid x, w^{(M-1)})} = L_M(x) - T$$

where

$$L_M(x) = \frac{1}{2}\log\frac{p(x \mid y=+1, w^{(M-1)})}{p(x \mid y=-1, w^{(M-1)})}$$

$$T = \frac{1}{2}\log\frac{P(y=+1)}{P(y=-1)}$$

- Problem: estimation of $p(x \mid y, w)$ or $L_M(x)$ is difficult for high dimensional data x

Approximating $p(x \mid y, w)$

- Design weak classifiers in 1-space instead of 400-D space
- Design a dictionary of candidate scalar features of x:

  $\{ \boldsymbol{f_k}(x) \mid k=1,\ldots,K \}$ -- see later

- Given $f^{(1)}(x),\cdots,f^{(M-1)}(x)$ selected by previous stages, approximate

$$p(x \mid y, w^{(M-1)}) \approx p(f^{(1)}(x),\cdots,f^{(M-1)}(x),\boldsymbol{f_k}(x) \mid y, w^{(M-1)})$$
$$= p(\boldsymbol{f_k}(x) \mid y, w^{(M-1)})\, p(f^{(M-1)}(x) \mid y, w^{(M-1)}) \cdots p(f^{(1)}(x) \mid y, w^{(0)})$$

Uni-Variate Weak Classifiers

- Construct a dictionary of candidate weak classifiers

$$h_k^{(M)}(x) = \frac{1}{2}\log\frac{p(f_k(x) \mid y=+1, w^{(M-1)})}{p(f_k(x) \mid y=-1, w^{(M-1)})} - T_k^{(M)}$$

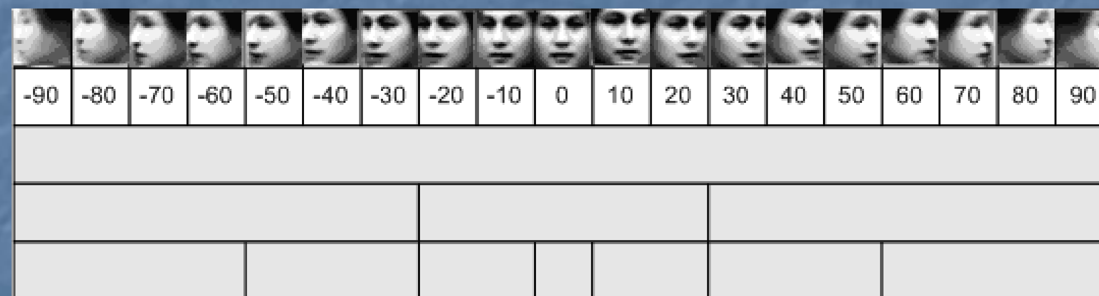
where

$$T_k^{(M)} = \frac{1}{2}\sum_{m=1}^{M-1}\log\left(\frac{p(f^{(m)}(x) \mid y=+1, w^{(m-1)})}{p(f^{(m)}(x) \mid y=-1, w^{(m-1)})}\right)$$

- Find the feature *k* so that $h_k^{(M)}(x)$ best fits $h_M(x)$ w.r.t. training data $\{x_i\}$, and take $h_M(x) = h_k^{(M)}(x)$

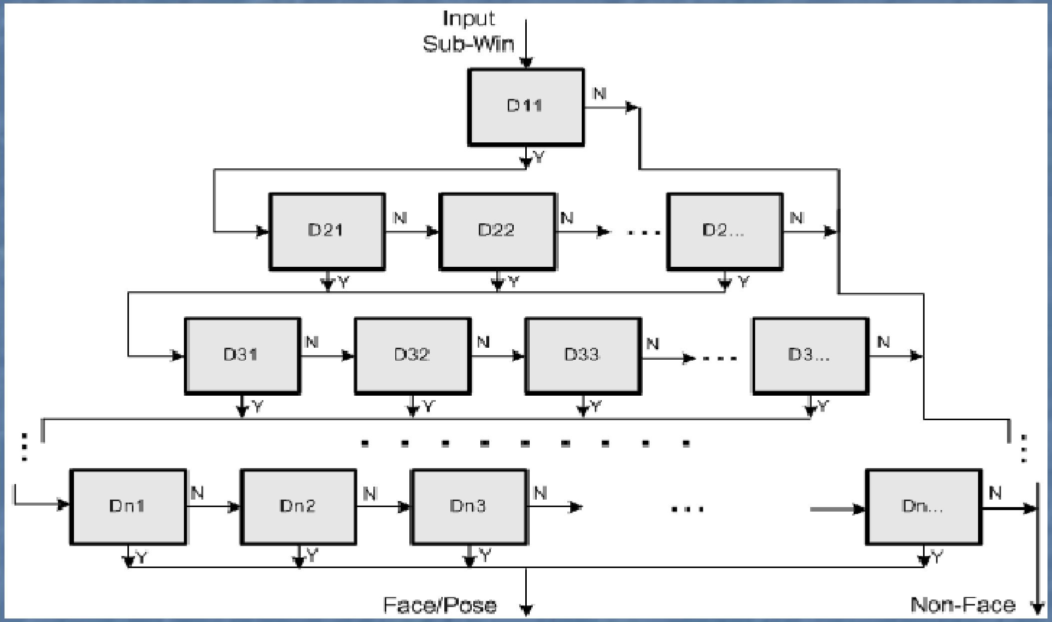
Dealing with Out-of-Plane Rotation

- Coarse to fine view partition:



combine

Detector-Pyramid

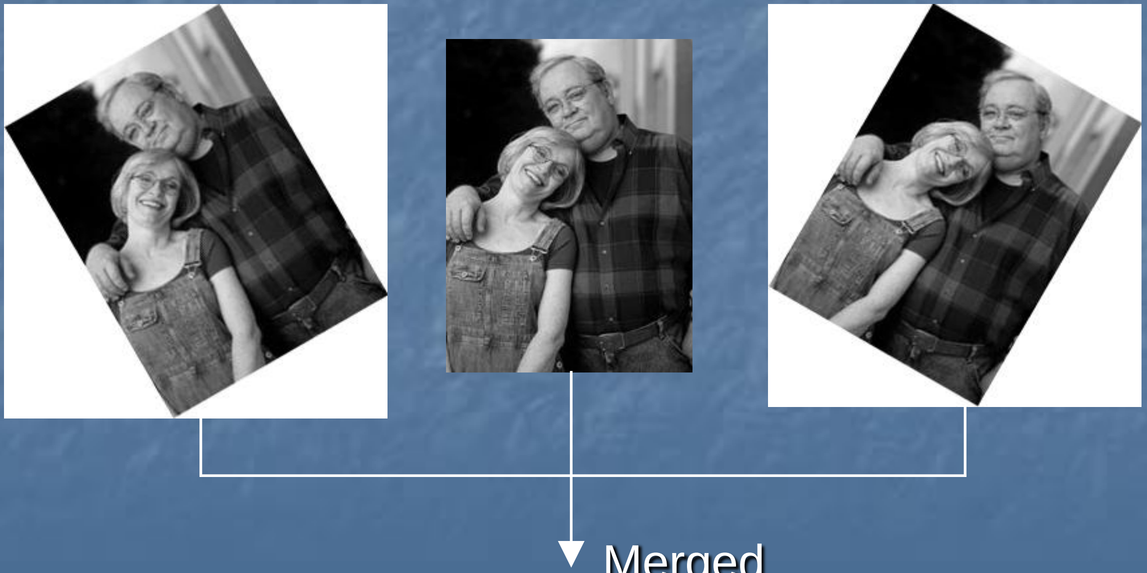


Level	1	2	3	Total
Time (ms)	110	77	15	202

Method	V-based	Det.Pyr
Time (ms)	967	202

Dealing with In-Plane Rotation

- Boosted face detector covers +- 15 deg
- Rotating the image by +-30 deg
- As result, +-45 deg can be covered



Multi-View Face Detection

- Fast MV FD system reported (5 fps)
- Gives pose estimate while detecting faces
- Face detection and recognition demo tomorrow

Conclusions

- FloatBoost learns a strong classifier of less complexity than AdaBoost (hence less overfitting)
- Formula for uni-variate approximation of ideal weak classifiers
- Fast multi-view face detection system
- Future work:
  - Dealing with outliers in learning
  - Improving training efficiency by sub-sampling

Face Pose Estimation

Facial Pose Estimation

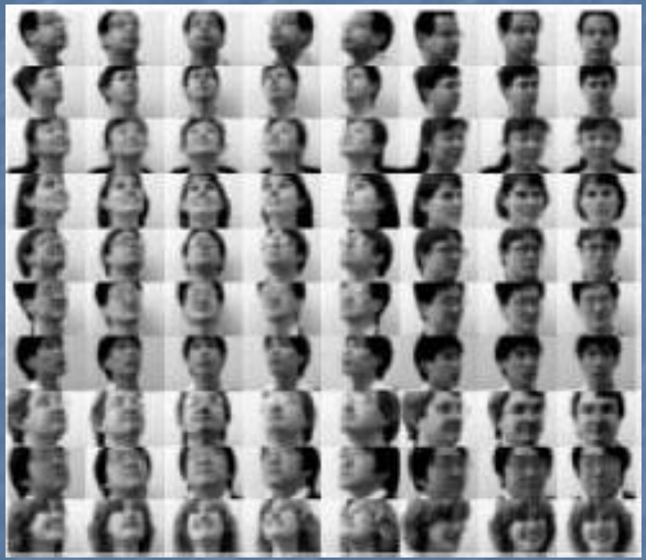
- Approximately 75 percent of the faces in home photos are non-frontal
- Task: to estimate the angle of head rotation

out-of-plane rotation

Approaches

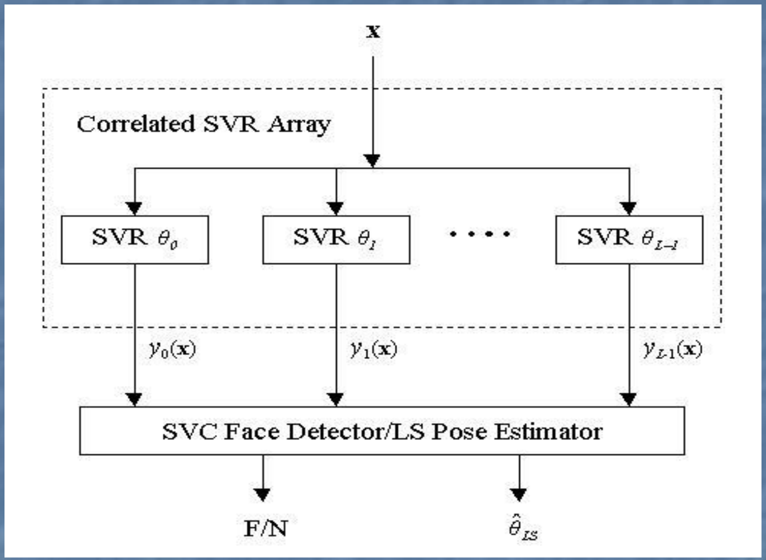
- Unsupervised learning (eg using ICA)
  - Pose clustering and pose classification learned using pose-unlabeled face data
- Supervised learning
  - Pose clustering and pose classification learned using pose-labeled face data

Supervised Learning of Nonlinear Mapping for Pose Estimation



$$\theta_0 = 0^o, \theta_1 = 10^o, \ldots, \theta_9 = 90^o$$

Regardless of illumination and identity

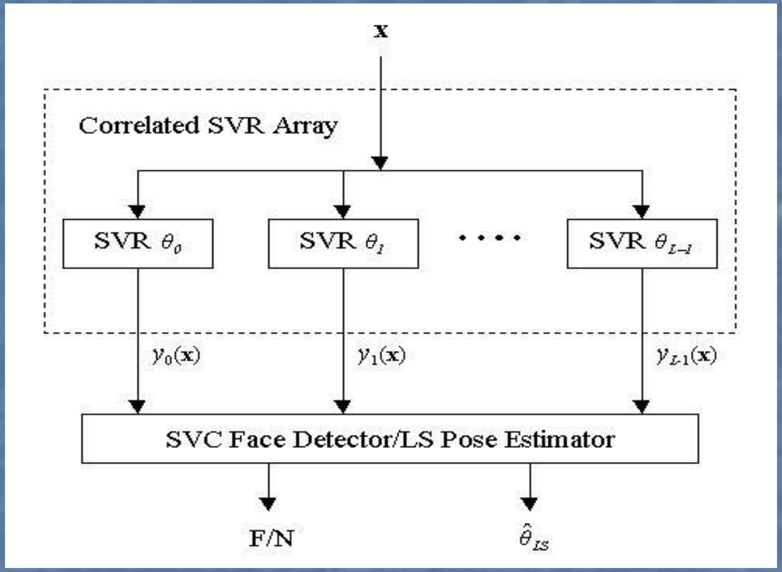


x: 400-D→ y: 10-D→ θ: 1-D

SVR training objective:

$$y_i(x) = \cos(\theta(x) - \theta_j)$$

Pose Estimation Using SVR Array (Li et al ICCV'01)



View-Specific SVR's:

$$\theta_0 = 0^{o}, \theta_1 = 10^{o}, ..., \theta_9 = 90^{o}$$

SVR θ_j is trained to output:

$$y_j = \cos(\theta(x) - \theta_j)$$

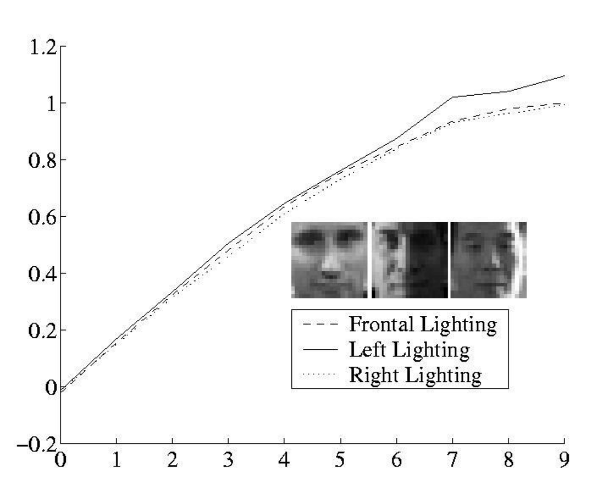
- Supervised Learning of Nonlinear View-Subspaces
  - From View Labeled Training Data
  - Illumination-invariant

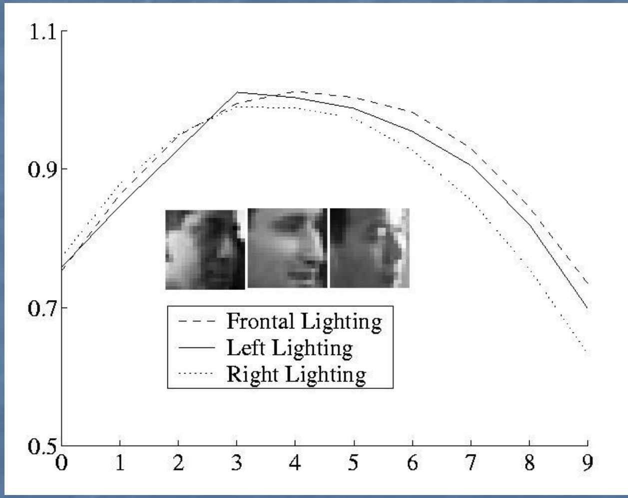
Illumination-Invariant, View-Specific Signature

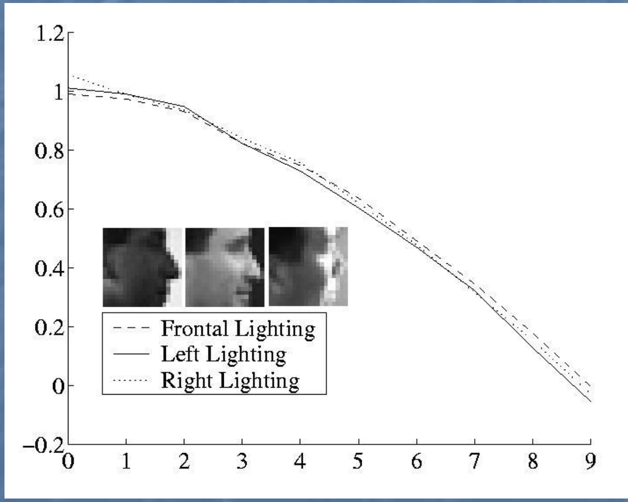
a=90 degree

a=40 degree

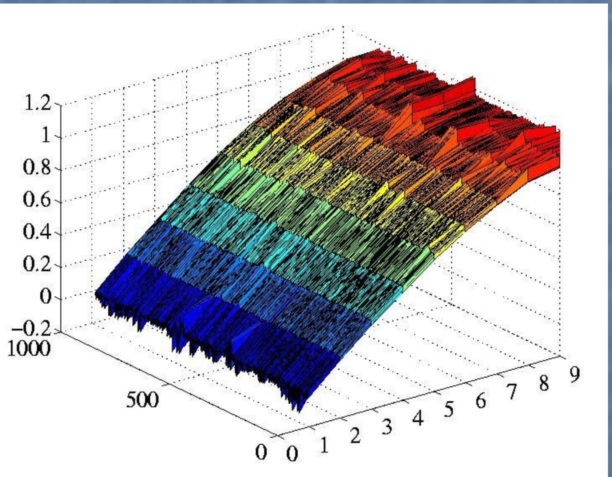
a=0 degree

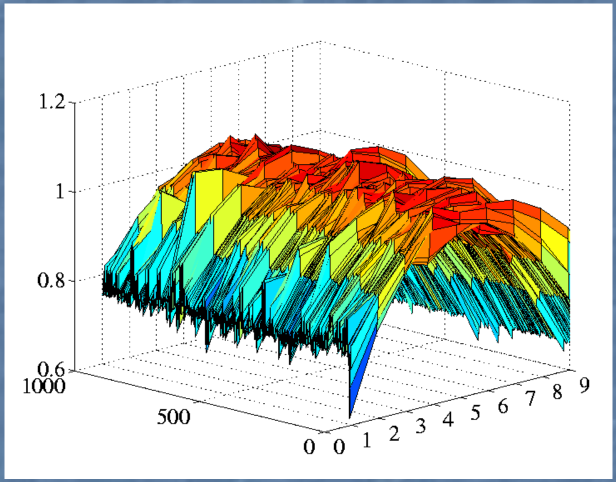


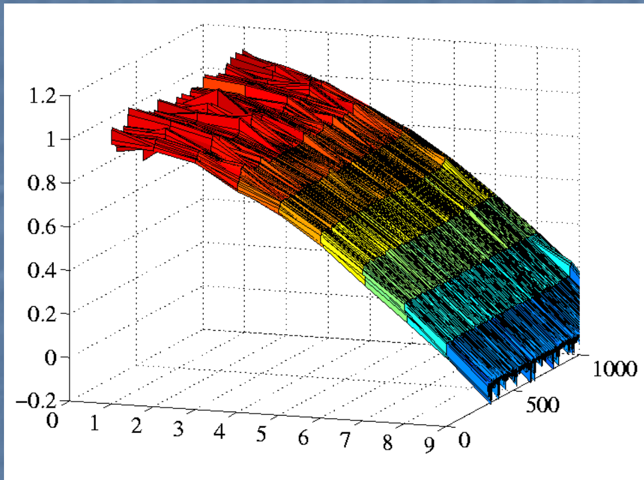




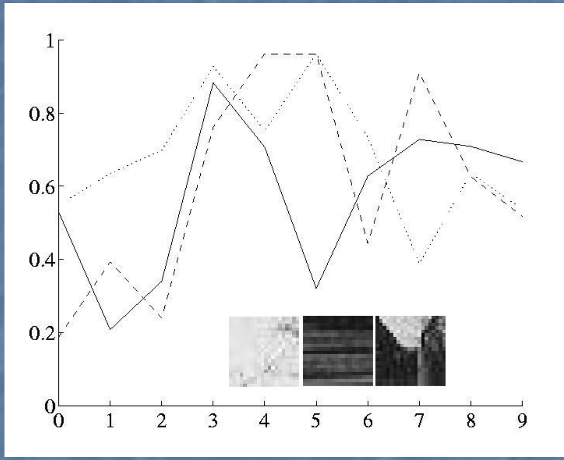
Results with 2000 Test Samples Each View

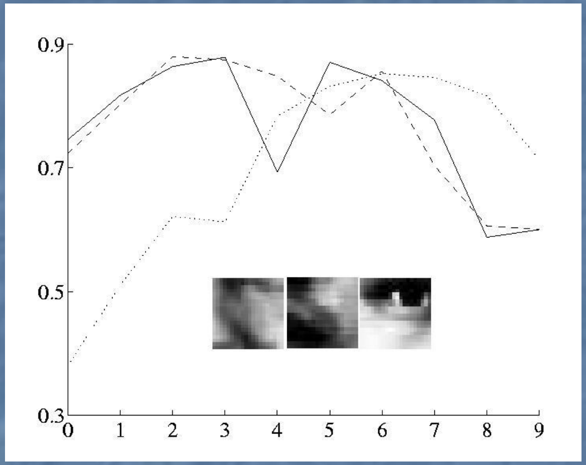






SVR Output for Nonfaces

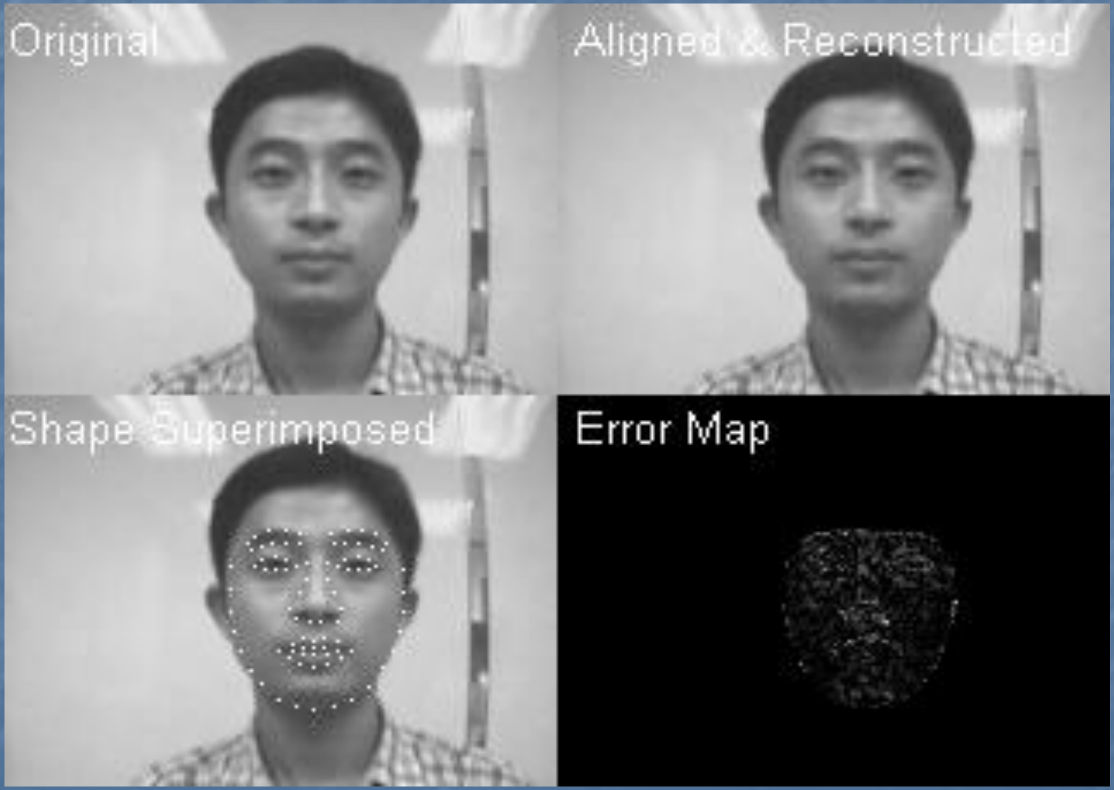




Face Alignment

ASM, AAM, DAM,TC-ASM

AAM/DAM


Original
Aligned & Reconstructed
Shape Superimposed
Error Map

Active Shape Models (ASM)

- Developed by Cootes, Taylor, *et al.*
  - The solution space is constrained by PDM, namely the global shape space.
  - Local appearance models derived at the landmarks converge to the local image evidence.

Formulation of ASM

- Global Shape Model: $S = \bar{S} + Us$
- Local Appearance Models:

$$(x, y) = \underset{(x,y)\in N(x_i^n, y_i^n)}{\arg\min} \| g_i(x, y) - \bar{g}_i \|^2_{\Sigma_i^g}$$

Where $\bar{g}_i$ is the average profile around the i-th landmark, and Σ_i^g is the covariance matrix of the sample profiles for the i-th landmark.

Formulation of ASM

- In each iteration, S_{lm} is obtained from the refinement of the local appearance models, the solution shape s is derived by maximizing the likelihood probability:

$$s = \arg\max_{s} p(S_{lm} \mid s) = \arg\min_{s} Eng(S_{lm}; s)$$

where

$$Eng(S_{lm}; s) = \lambda \| S_{lm} - S_{lm}^{'} \|^2 + \| s - s_{lm} \|_{\Lambda}^2$$

Active Appearance Models(AAM)

- Cootes proposed and developed the Active Appearance Model (AAM)
  - Built based on PDM.
  - Shape and texture are combined for the *appearance* modeling.
  - Alignment is guided by minimizing the texture difference between model and ground truth.

Formulation of AAM

- Shape Model: $S = \bar{S} + Us$
- Texture Model: $T = \bar{T} + Vt$
- Appearance Model:

$$A = \begin{pmatrix} \Lambda s \\ t \end{pmatrix} \qquad A = Wa$$

- The search strategies are based on the linear regression assumptions:

$$\delta a = A_a \delta T \qquad \delta p = A_p \delta T$$

Direct Appearance Models

Shape and Texture Subspaces in AAM

- Shape is represented as *s* in PCA shape subspace:

$$S = \overline{S} + Us$$

- Texture represented as *t* in PCA texture subspace

$$T = \overline{T} + Vt$$

- Appearance represented as a in appearance subspace

$$A = \begin{pmatrix} \Lambda s \\ t \end{pmatrix} = Wa$$

Problems with AAM

- Shortcoming 1: In most case, dim (S_a) < dim (S_t). Therefore, some admissible textures are not modeled in appearance subspace
- Shortcoming 2: *a* and *T* are high dim vectors. So, very large memory is required in learning Aa in $\delta a = A_a \delta T$

DAM Modeling

- To rectify the shortcomings, depend *s* entirely on *t*, ie

$$s = Rt + \varepsilon$$

- Reasons
  - Intuitively, the same shape can enclose different textures, however, the reverse is not true.
  - The dimension of texture space is much higher than that of shape space.

DAM Searching

1. Given current *p* and shape *s*, get texture *T*;
2. Use the principle components of δT to predict the position displacement

$$\delta p = R_p \delta T' = R_p H^T \delta T$$

3. Use warped texture *T* to predict next shape

$$s = Rt = RV^T T$$

4. Goto 1;

- DAM Learning: R_p, H^T, and R

Advantages of DAM

- DAM subspace includes the textures previously unseen by AAM.
- The convergence and accuracy are improved.
- The memory requirement is cut down to a large extent.

Experiment Results

	$E(\|\delta T\|^2)$	$E(\|\delta p\|^2)$	**Converge Rate**
DAM (Training)	**0.156572**	**0.986815**	**100%**
AAM (Training)	**0.712095**	**2.095902**	**70%**
DAM (Test)	**1.114020**	**2.942606**	**85%**
AAM (Test)	**2.508195**	**4.253023**	**62%**

* The convergence is judged by the satisfactions of two conditions: $\|\delta T\|^2 < 0.5$ and $\|\delta p\|^2 < 3$.

Texture-Constrained ASM

Motivation

- ASM:
  - Local information statistics enable good landmark localization (pro)
  - Solution often sub-optimal, depending on the initialization (con)
- AAM:
  - Incorporate global texture evidence (pro)
  - Linear assumption about texture variation to appearance and position variation make it affected by illumination variation.
- Texture constrained ASM: Inherit pros + Rectify cons + New optimization strategy

TC- ASM

- Use the local appearance model of ASM for landmark localization - less sensitive to illumination variation.
- Use global texture to constrain the shape - for more accurate estimation of shape parameters in optimization process.

TC-ASM

- Texture-constrained shape model
  - For the edge or contour landmark, position uncertainty exists given the texture, whilst there are correlations between the shape and the texture for the face pattern.
  - The conditional distribution of the shape s given texture t is assumed Gaussian:

$$p(s \mid t) \propto N(s_t, \Sigma_t)$$

  - The shape s_t can be derived from the texture t directly, and is assumed linearly dependent on t:

$$s_t = Rt$$

TC-ASM

- Search based on Bayesian framework
  - S_{lm} is obtained from local appearance models as in ASM;
  - The texture t is extracted from the shape S_{lm} and shape s_t is derived from $s_t = Rt$.
  - The posterior(MAP) estimation of the solution shape s given S_{lm} and s_t :

$$s = \arg\max_s p(s \mid S_{lm}, s_t)$$

$$= \arg\max_s \frac{p(S_{lm} \mid s_t, s)\, p(s \mid s_t)\, p(s_t)}{p(S_{lm}, s_t)}$$

TC-ASM

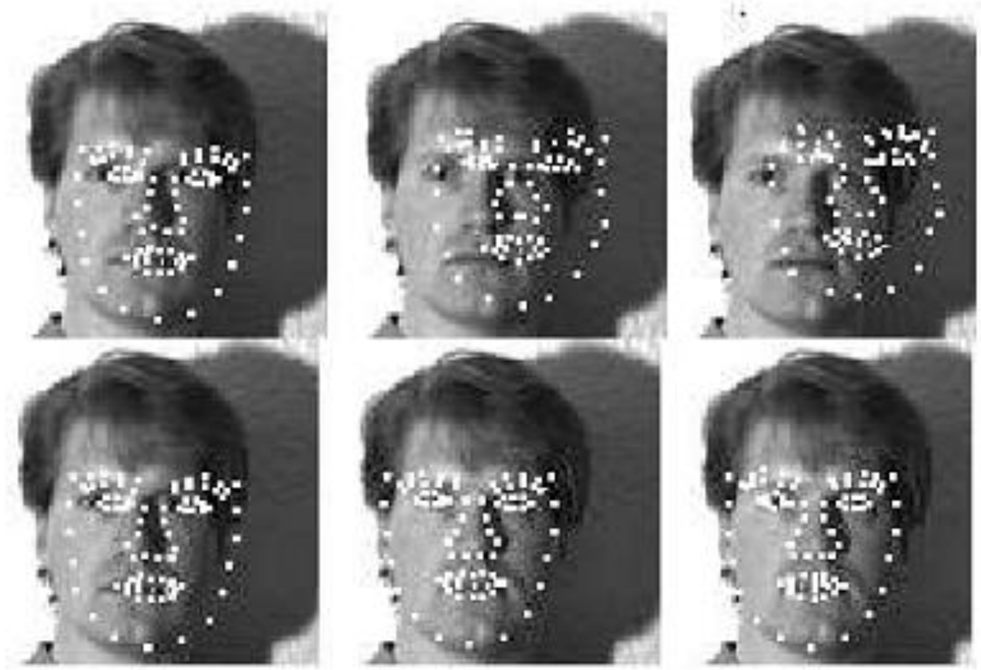
- Assuming S_{lm} is independent to s_t , given s , we obtain:

$$s = \arg\max_{s} p(S_{lm} \mid s)\, p(s \mid s_t)$$

$$= \arg\min_{s} \{Eng(S_{lm}; s) + Eng(s; s_t)\}$$

$$= (\Lambda^{-1} + \Sigma_t^{-1})^{-1}(\Lambda^{-1} s_{lm} + \Sigma_t^{-1} s_t)$$

Comparing AAM & TC-ASM under Illumination Change

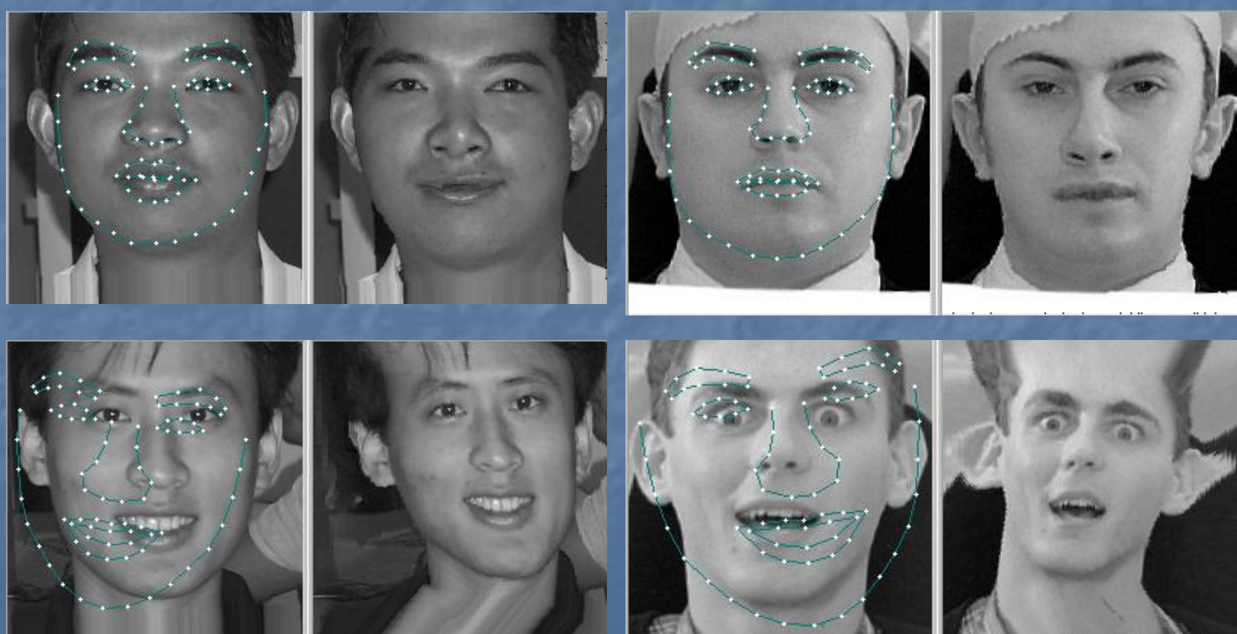


Sensitivities of AAM (upper) and TC-ASM (lower) to illumination condition not seen in the training data. From left to right are the results obtained at the 0-th, 2-th, and 10-th iterations.(Result in different level of image pyramid is scaled back to the original scale)

Evaluation of ASM Alignment Results

Learning Evaluation Function

- Using AdaBoost classifier output as quantitative measure of alignment quality

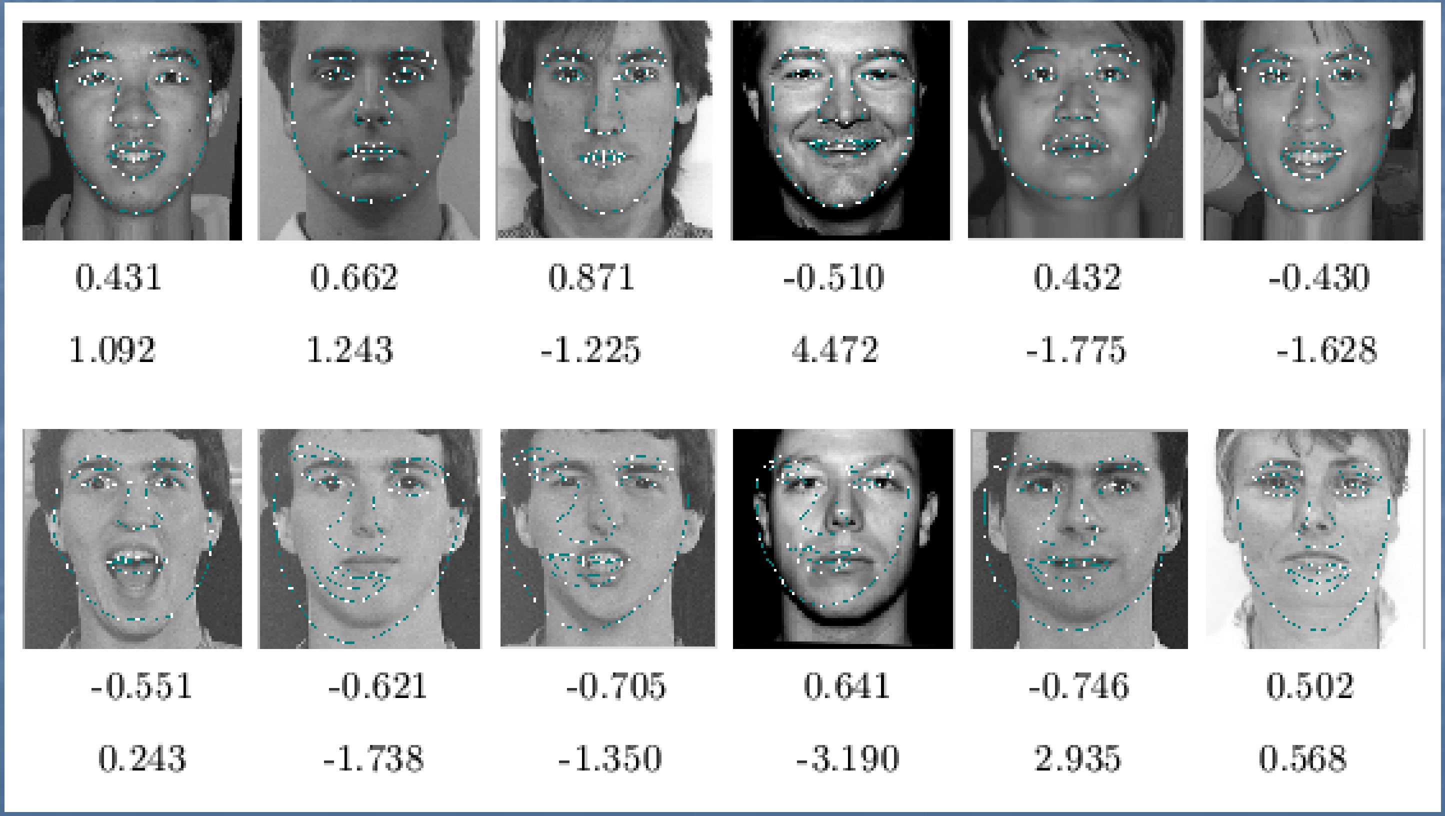


Good alignment

Bad alignment

Evaluation Results

- AdaBoost Output vs. Reconstruction Error



-- Learning
-- Reconstr

-- Learning
-- Reconstr

Face Recognition

Local Features + AdaBoost Learning

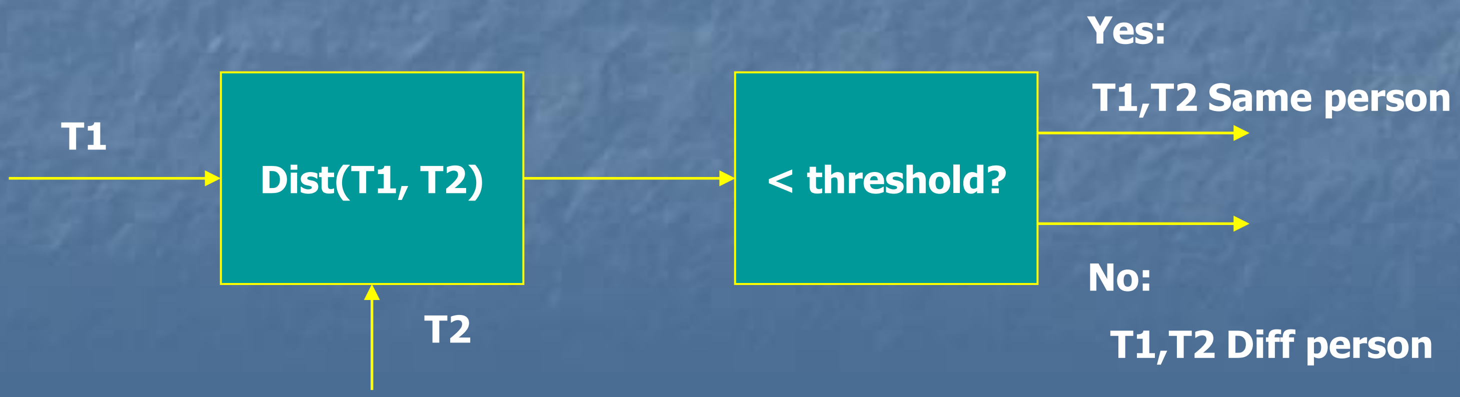
Framework

- Local Features
  - Eg: Haar, Gabor, LBP, Ordinal, etc
  - Having good properties
  - Form a High-Dim Space
- Intra vs Extra Representation for Multi-class Problem
- Statistical Learning
  - 2-Class Classification
  - Training on pos and neg samples
  - Nonlinear classifier: Eg AdaBoos, SVM
  - Learning for
    - Dim reduction (feature selction)
    - Classifier construction

Intra vs Extra Representation: N Class → Two Class
(Baback Moghaddam)

- N persons

- Compare 2 templates



Intra- and Extra- personal Variations in Image Space

(Baback Moghaddam)

Representative Works

- Viola & Jones, papers 2001,2002
- M. Jones, MERL TechReport 2003
- Li & Students, papers 2001-2005

Local Features

Good Features

- Reduce extrinsic factors while keeping intrinsic factors unchanged
- Simpler in shape than in image space
- Individual faces are still separable
  - By a metric matching of templates
  - Separable by a nonlinear boundary

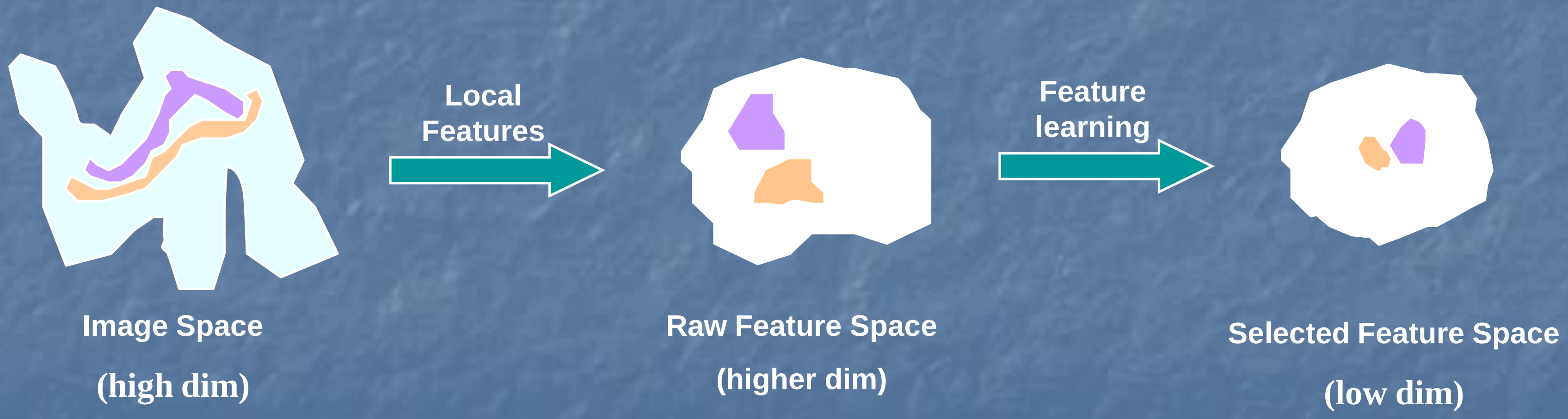
Local Features

- Haar
- Gabor wavelets
- Local binary patterns (LBP)
- Ordinal Features, etc

Dim expansion: More local features than pixels.

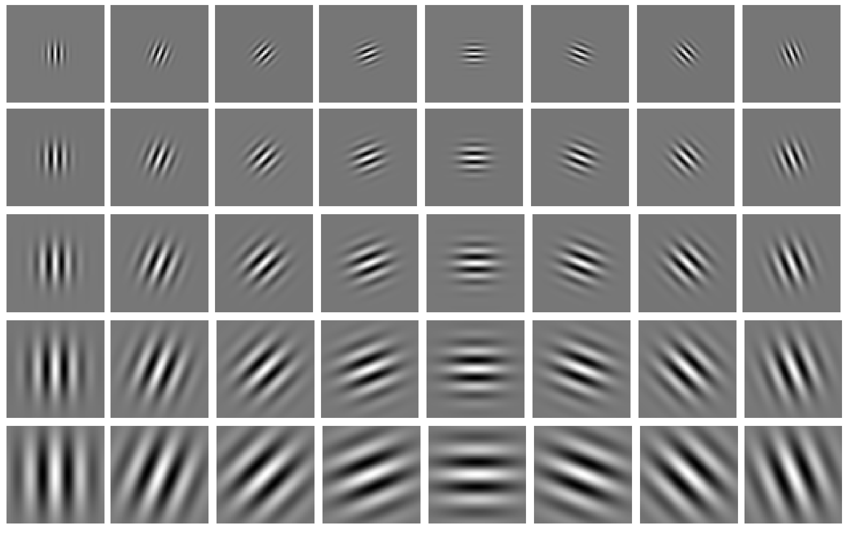
Working in Good Feature Space

- Map input image to a higher dim local feature space
- Reduce dim by learning good features

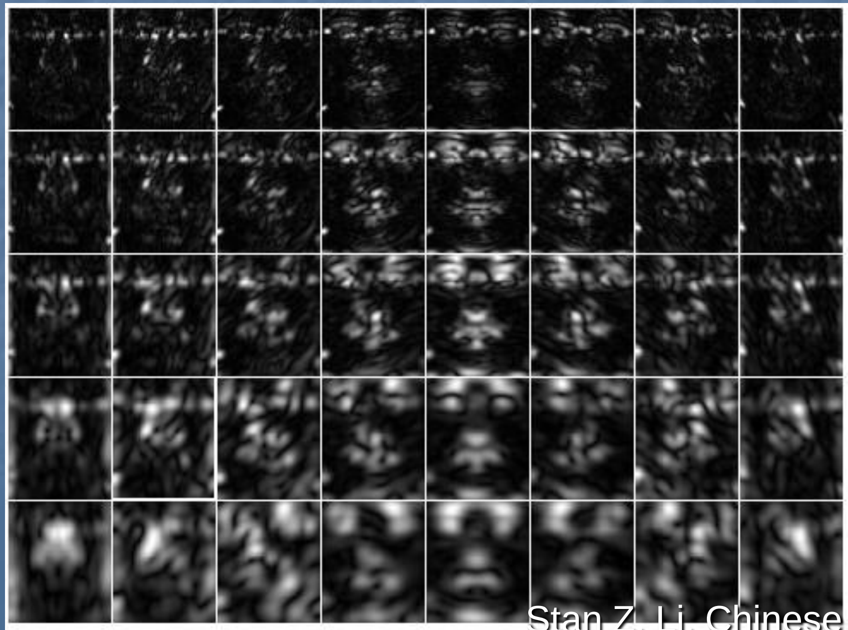


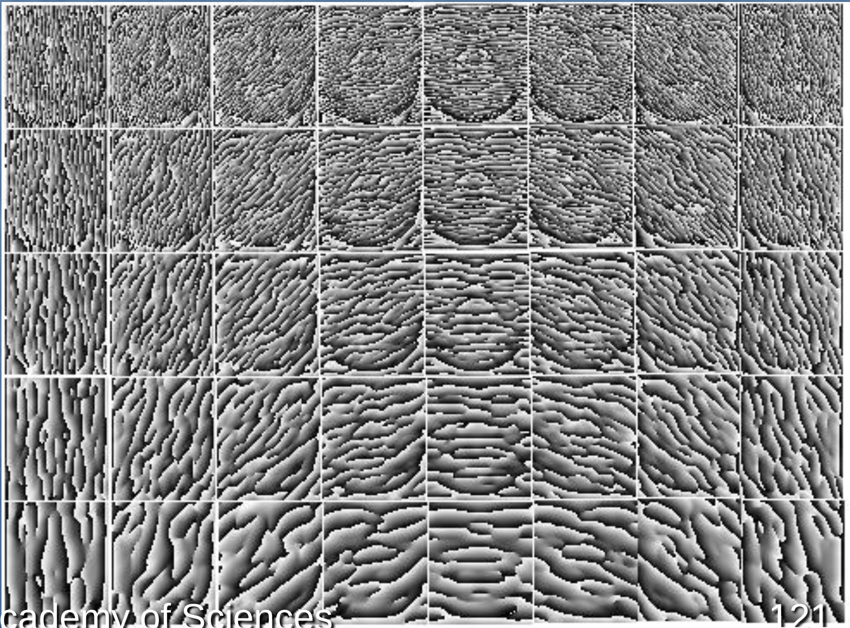
Gabor Features

Real Parts:

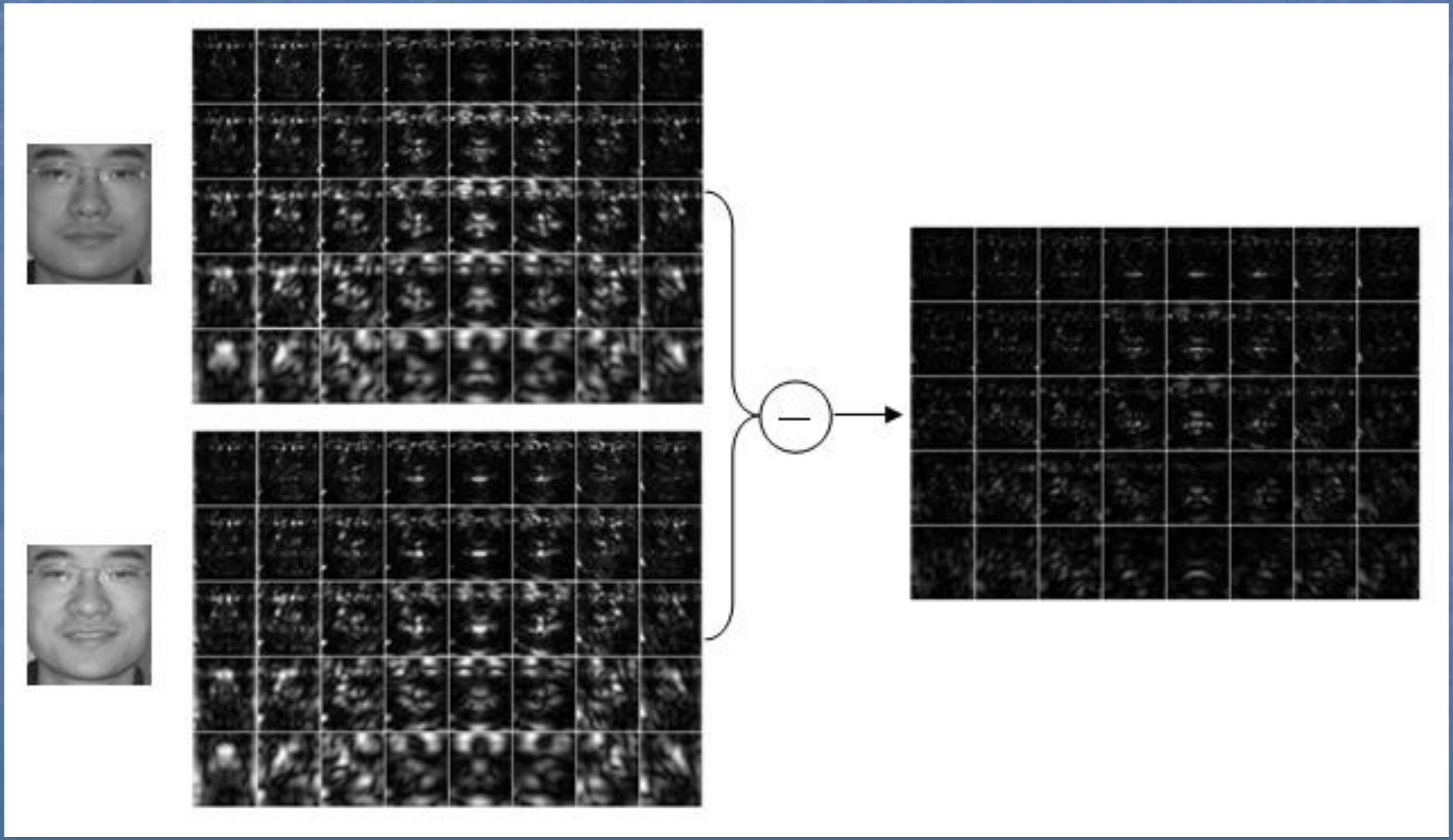


Magnitude & Phase

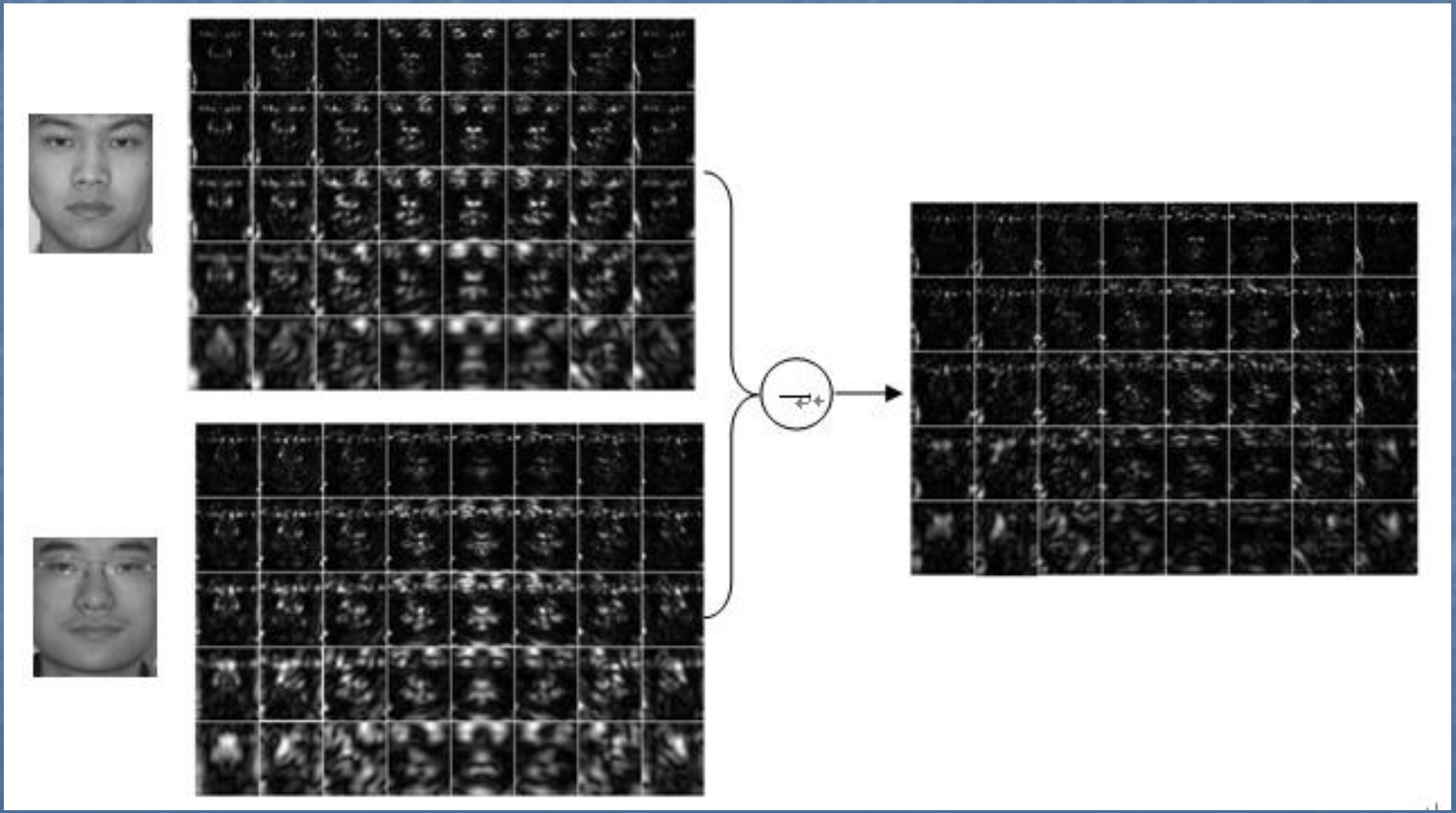




Intra-Personal Variation - Gabor

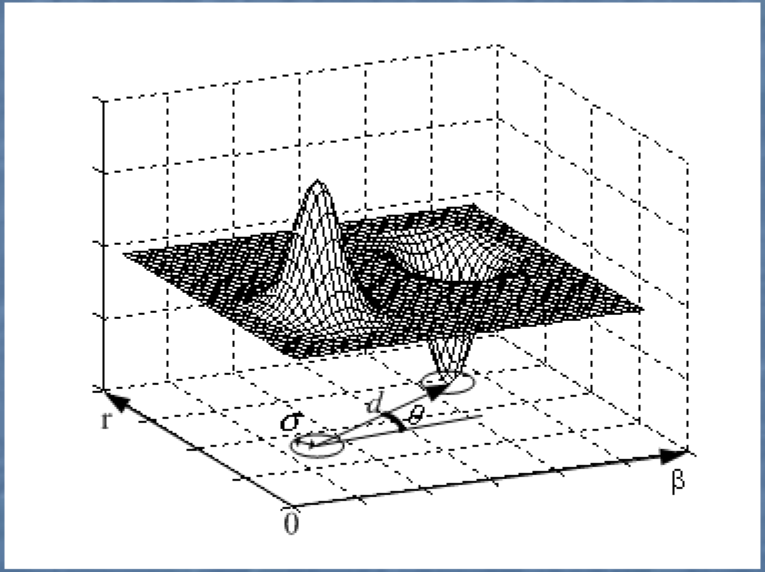


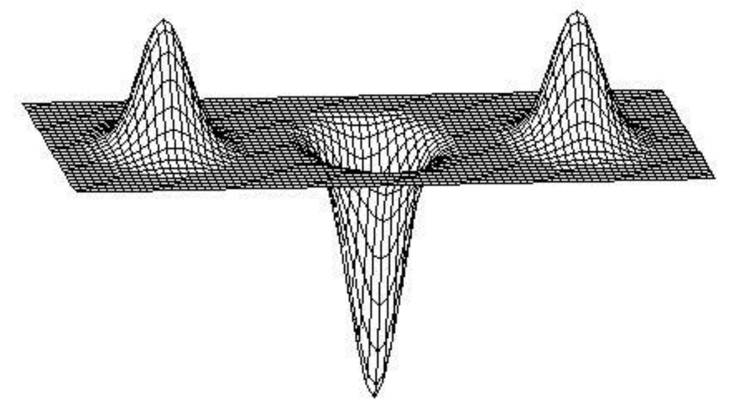
Extra-Personal Variation - Gabor

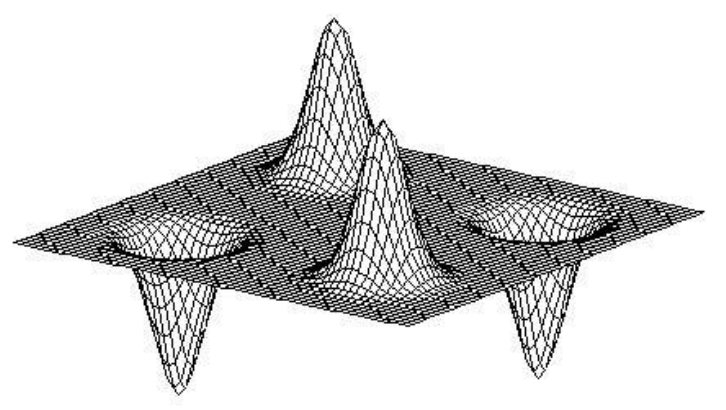


Ordinal Features (Liao, et al, ICB 2006)

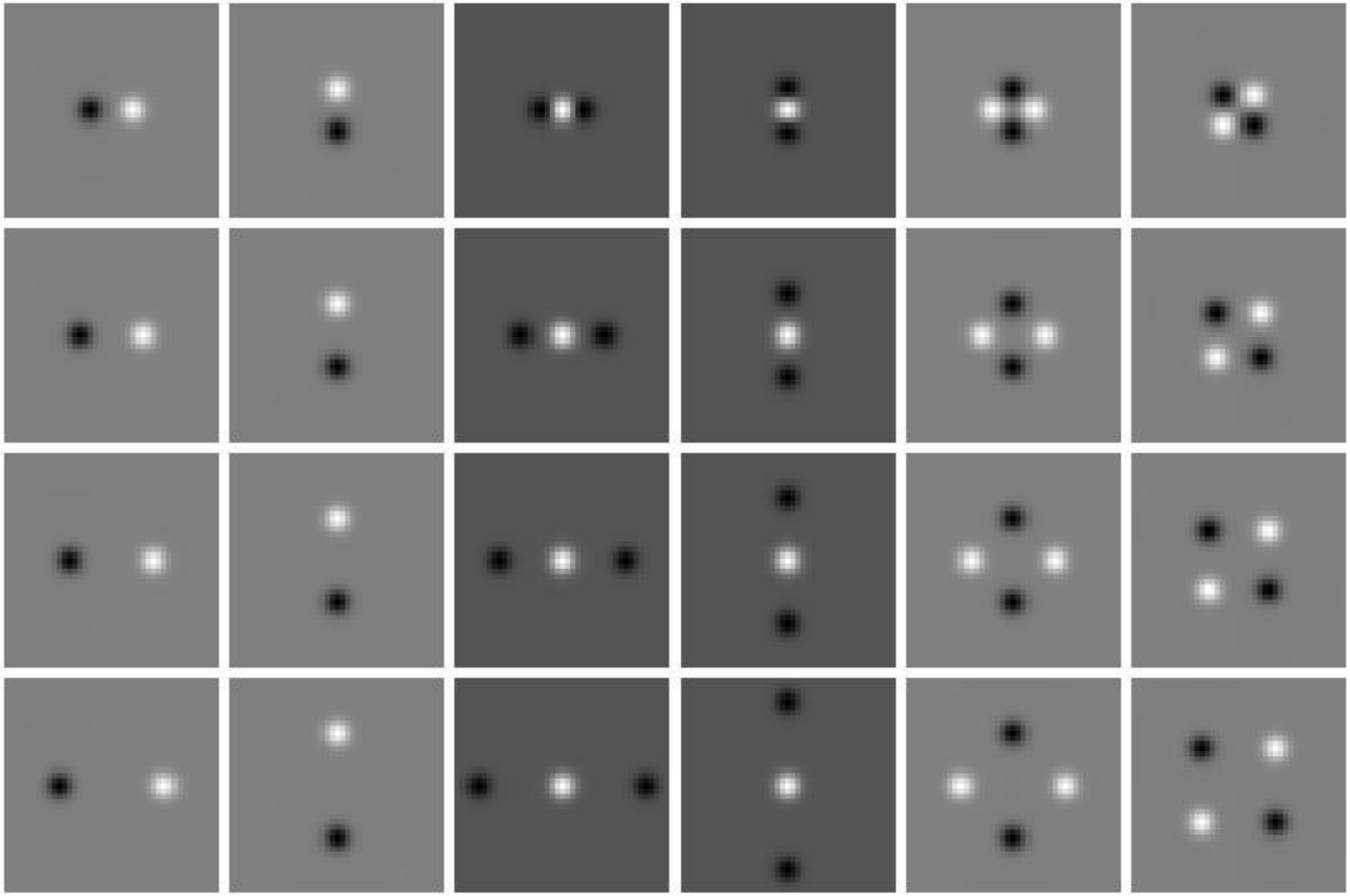
2-pole,3-pole,4-pole Filters







Parameters:σ,d,θ

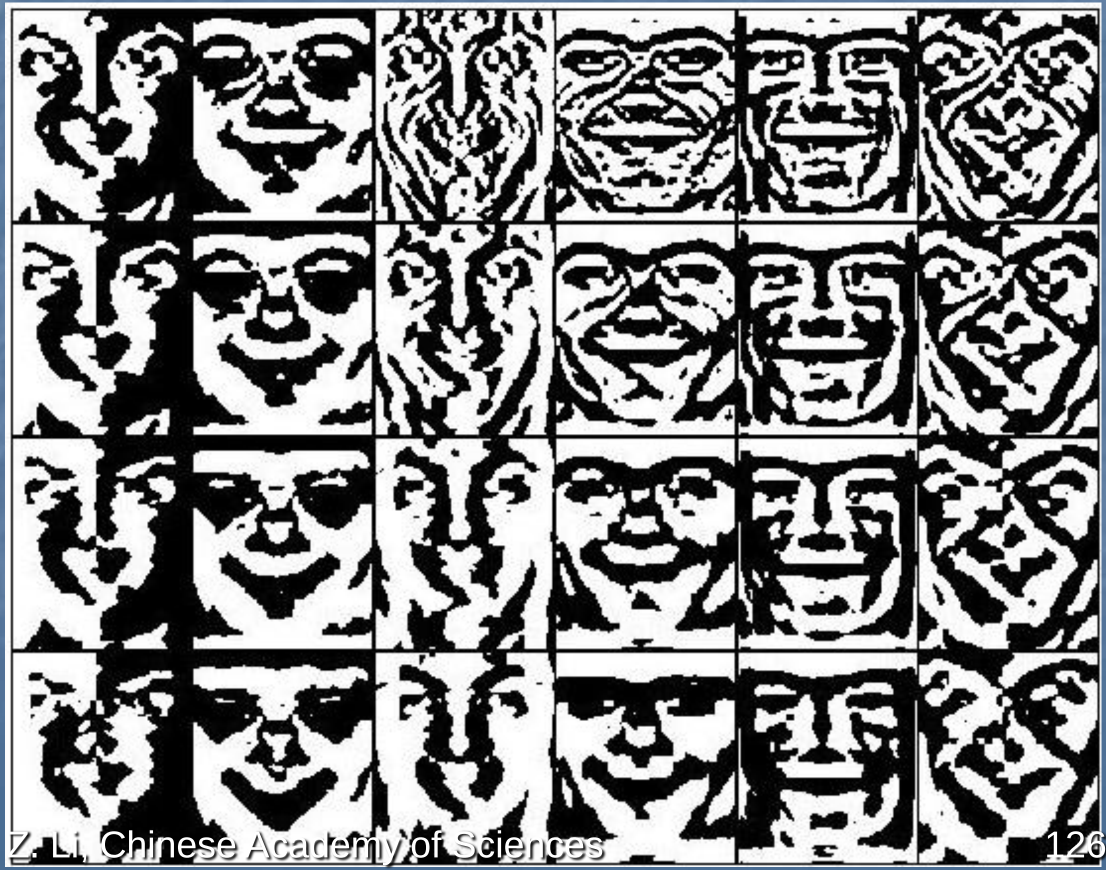


24 ordinal filters used in the experiments

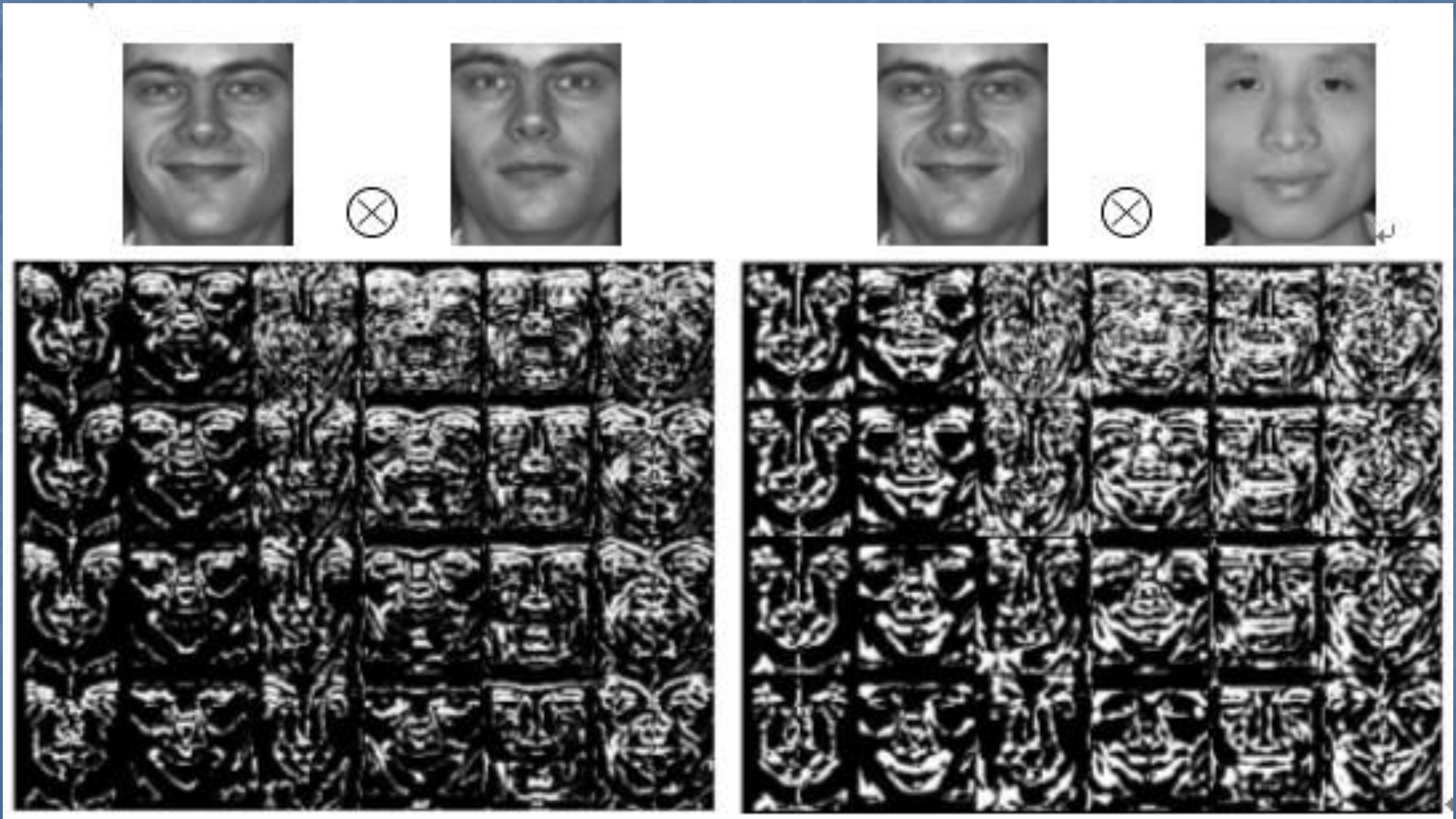
Ordinal Encoding of Face

filter → Threshold at 0 → binary image

→



Differences of Ordinal Maps



Intra-Difference Extra-Difference

Learning Good Local Features and Local Feature Based Classifiers

Feature Selection By Statistical Learning

- Goals:
  - Select good features from a large pool
  - Learn a sequence of weak classifiers using good features
  - Combine them into a strong classifier (the output result)
- Advantages
  - Few parameters to adjust
  - Trained classifier works fast

As Result of AdaBoost Learning

- Effective features are selected
- A weak classifier is constructed for each feature
- The weak classifiers are combined into a strong one
- Fusion at both feature and decision levels

Successful Applications

- 2D Face Detection & Recognition
  - Viola & Jones, Haar + boosting for face detection & recognition
  - Li and students

    Haar / Gabor / LBP / Oridinal + boosting, 2001-2005
- 3D and 3D+2D Fusion
  - Li and students, 2005
- NIR Face Recognition
  - Li and students, 2004-2005

Learning Fusion of 3D+2D at Feature and Decision Levels

Motivation

- 2D and 3D modals
  - are both useful
  - but contribute in different ways
- 2D+3D fusion
  - Performs better than 2D or 3D alone
  - currently done (mostly) at decision level
- 2D+3D fusion at feature level could be advantageous (Bowyer, Chang, Flynn 2004)
- 2D+3D fusion could be even better if fusion at both feature and decision levels (this paper)
- We do this in the framework of "local feature + AdaBoost learning"

Background

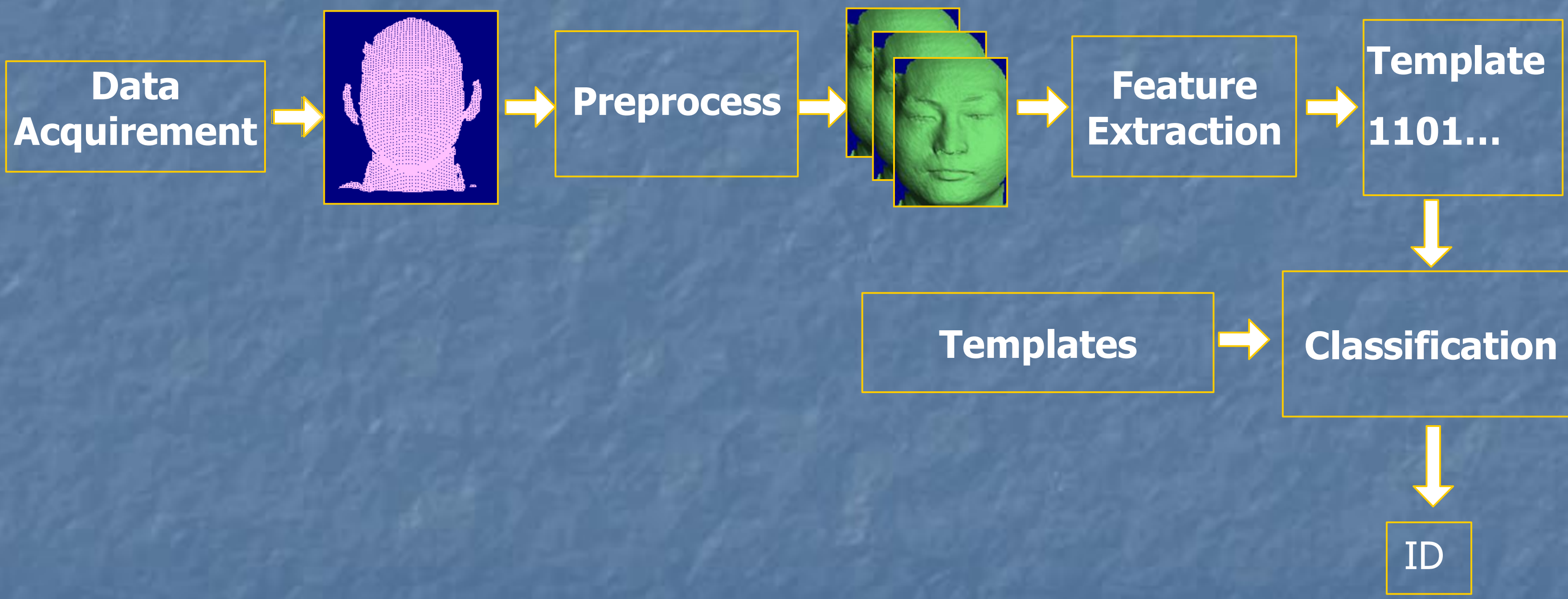
- PCA, LDA, ICA, EGBM, etc
- Local Feature + Boosted Classifiers (2D)
- 3D range imaging (Besl and Jain, 1985)
- 2D+3D multi-model face biometrics (Bowyer, Chang, and Flynn, ICPR 2004)
- 2D+3D fusion performs better than using either 3D or 2D alone (Chang, Bowyer and Flynn, PAMI 2005)

3D Face Recognition

Using 3D information： shape, depth

2D face recognition problem		3D face recognition resolution
Illumination	→	Laser scanner
Pose	→	Exact registration
Expression	→	Select robust region

Processing 3D Face Recognition

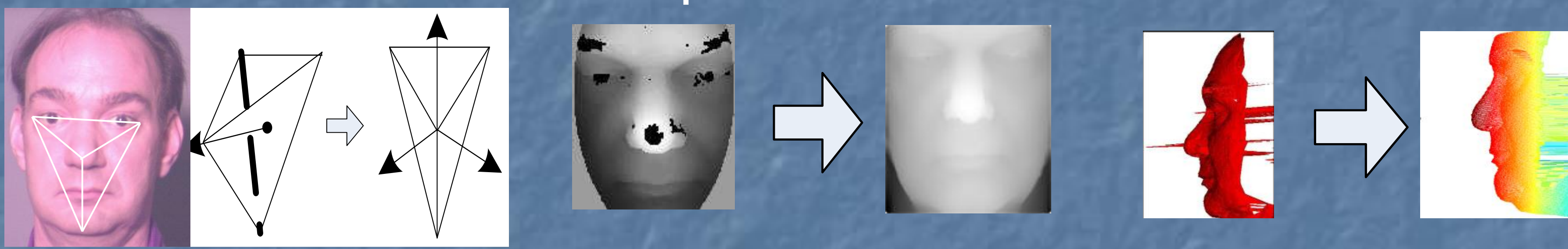


Preprocess for 3D

Shift & Rotation

Cubic interpolation

Nearest filter



Preprocess for 2D

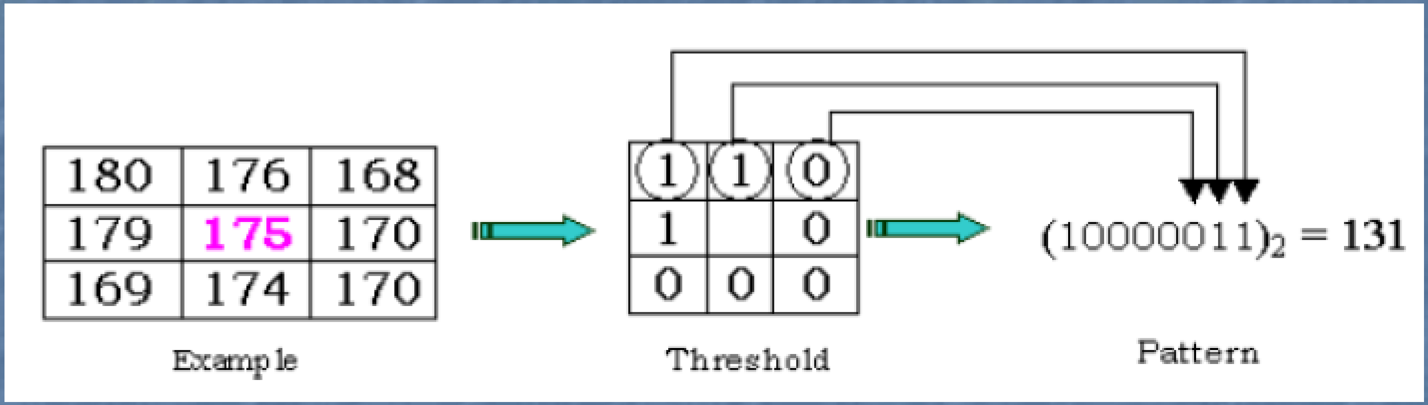
Alignment

Histogram equalization

Mask

Preprocessing Results

LBP Local Features



- Extracted for every pixel location in 3D and 2D images

LBP Histograms for Sub-Windows

- An LBP Histogram
  - 256 bins for un-restricted LBP code
  - 59 bins for uniform LBP code
  - Subject to sub-window size
- An LBP histogram computed for each location with a given window size
- Distance btw 2 histogram – Eq.(2)
- Naïve distance btw 2 LBP feature templates: summing up for all locations and window sizes

3D+2D Fusion

3D data

2D data

- Generate a large number of Local Features
- Boosting learning to select best features
- Fusion at both feature and decision level
- Better 3D results and 3D+2D results

Experiments

- Compared Methods
  - CBF (Chang, Bowyer, Flynn 2005) fusion
    - 2D, 3D Metrics distance in PCA spaces
    - Weight ＝ (dist2－dist1)/(dist3-dist1)
      (dist1,2,3 are ranked distances for a probe)
  - Score fusion
    - Compute scores of boosted classifiers for 3D & 2D
    - Addition of the boosted scores (better than CBF addition)
  - Feature+Score fusion (proposed method)

Data Sets

3D Data	Num. of Images	Num. of Persons
Train	945	246
Gallery	252	252
Probe	1108	252

2D Data	Num. of Images	Num. of Persons
Train	945	246
Gallery	252	252
Probe	1108	252

Alignment Results

Features Learned for 3D or 2D

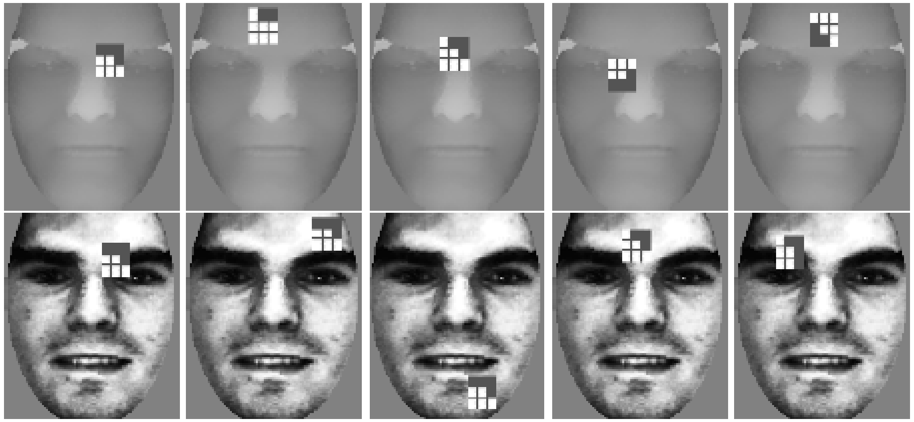
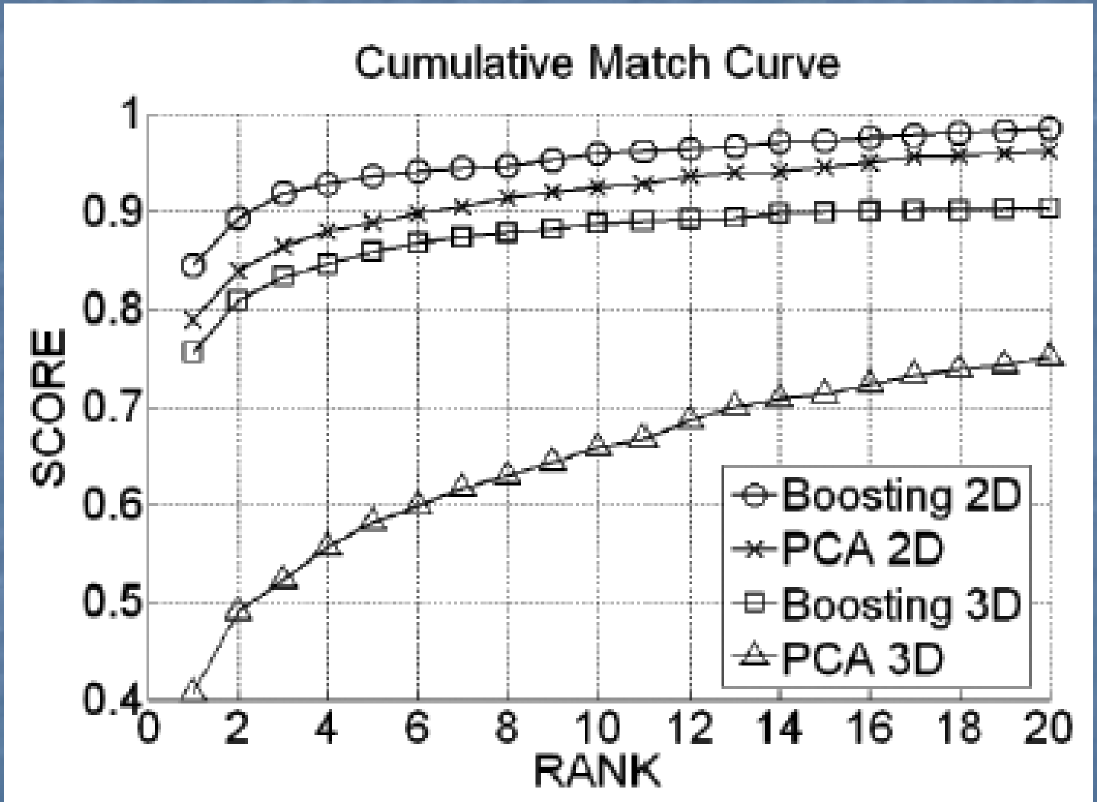


Fig. 5. The first 5 features for 3D (top) and for 2D (bottom) learned by AdaBoost.

3D or 2D Boosted Classifiers vs. PCA



Features Learned for 3D+2D

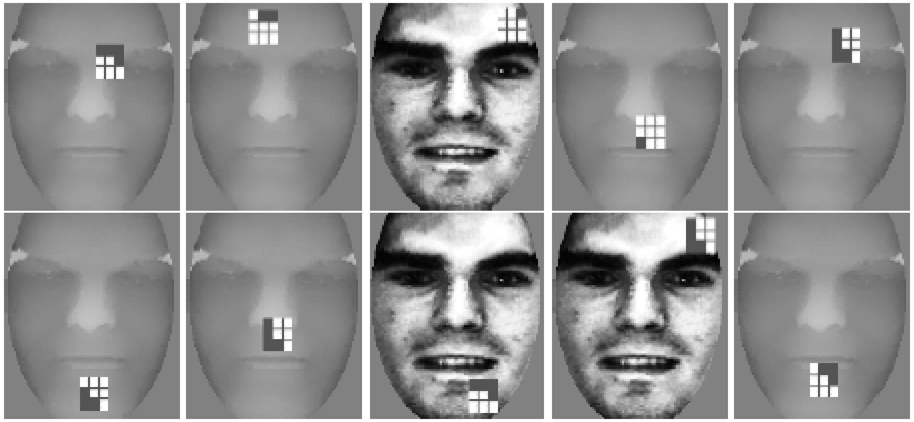
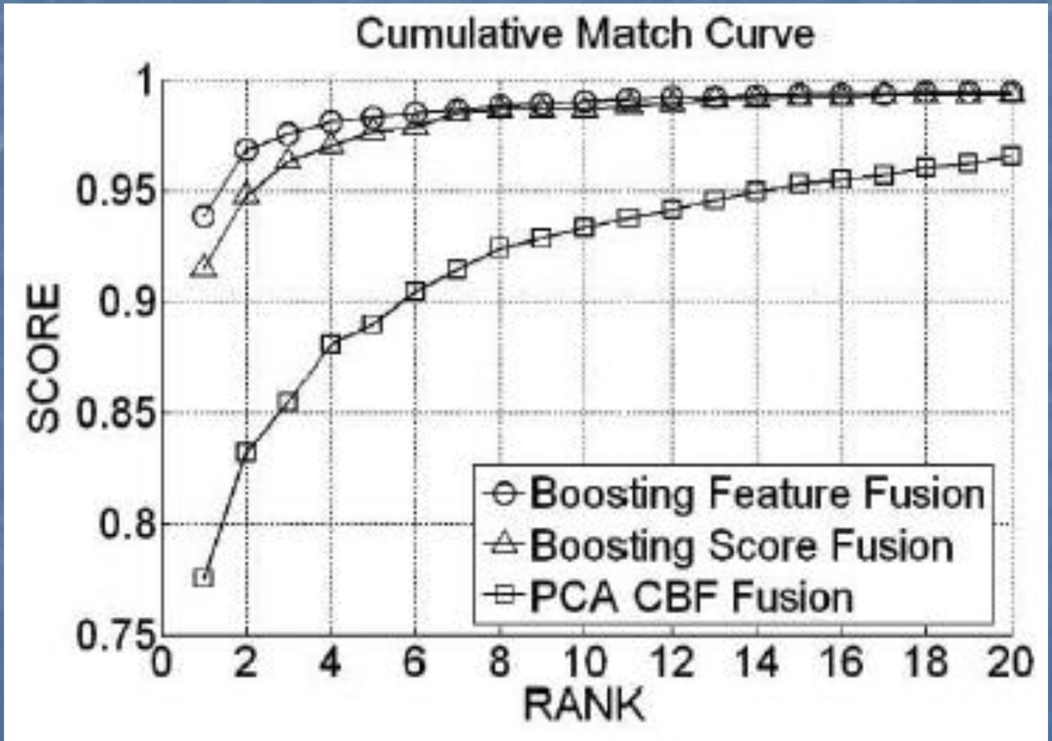


Fig. 7. The first 10 LBP features learned by boosted fusion of 3D+2D, ranked 1 to 10 from left to right, from top to bottom. Among these top 10, 7 features are from 3D data and 3 from 2D.

Comparison of 3 Fusion Methods



Conclusion

- Novelties
  - First using LBP Feature on 3D face recognition
  - First Adaboost learning for 3D Features
  - First fusion of 2D+3D at both feature and decision levels, using Adaboost learning
- Advantages demonstrated

Face Recognition Using Near Infrared Images

Intrinsic vs. Extrinsic Factors

- Extrinsic Variations
  - illumination
  - facial expression
  - head pose
  - facial hair, cosmetics
  - accessories (eyeglasses, etc)
  - image size and quality

- Intrinsic Info
  - (1) Info specific to faces (for face/nonface classification)
  - (2) Info specific for identity classification (identity dimension, for face recognition)
  - Immune from extrinsic factors

Imaging Models

- Face is a 3D
- Physical Imaging Model

$$I(x,y) = \rho(x,y)\, n^T(x,y)\, s$$

(Lambertian Model)

- Imaging Factors
  - Shape $n(x,y)$ – intrinsic factor
  - Albedo $\rho(x,y)$ – intrinsic factor
  - Illumination $s = (s_1, s_2, s_3)$ – extrinsic factor

Intrinsic Part of Face Image

- Face Image Model (Lambertian)

$$I(x,y) = \rho(x,y)\, n^T(x,y)\, s$$
$$= F^T(x,y)\, s$$

$F^T = \rho\, n^T$ -- the intrinsic factor of face identity

Strategies for Better Accuracy

- Remove Extrinsic Factors from Images
  - 3 Modeling - Vision Techniques
  - Morphable Models – Image and Learning
  - Advanced Sensors – This work
- Recognition Based on Intrinsic Factors only
  - Capturing both 3D information and reflectance of facial surfaces – This work

Near Infrared Face Recognition

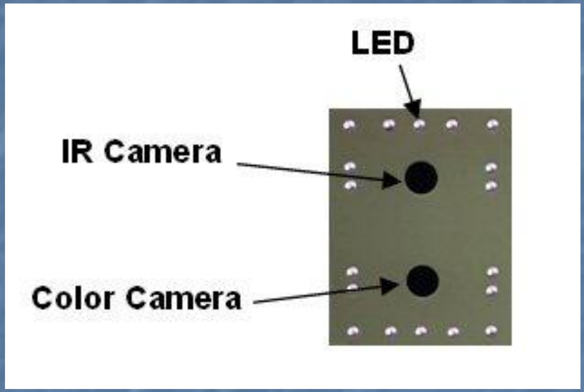
- Advantage
  - Illumination invariant face recognition Method

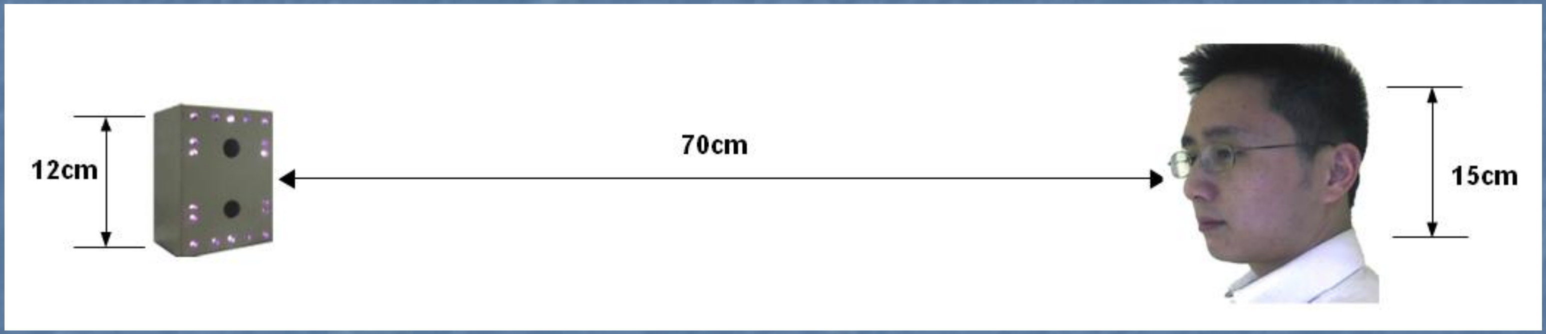
  - Highly accurate and fast
  - Can work in dark environment

AuthenMetric NIR Face Recogniton System

- For Cooperative Applications
  - Access control, E-Passport, ATM, etc
- Features
  - Novel NIR image capture device to minimizes influence of environmental lighting
  - Recognition Classifier learned using LBP features + AdaBoost
- Performance
  - Stable in environmental lighting of 0-50,000 Lux
  - Accurate and fast system in "Scenario Tests"

NIR Imaging Hardware

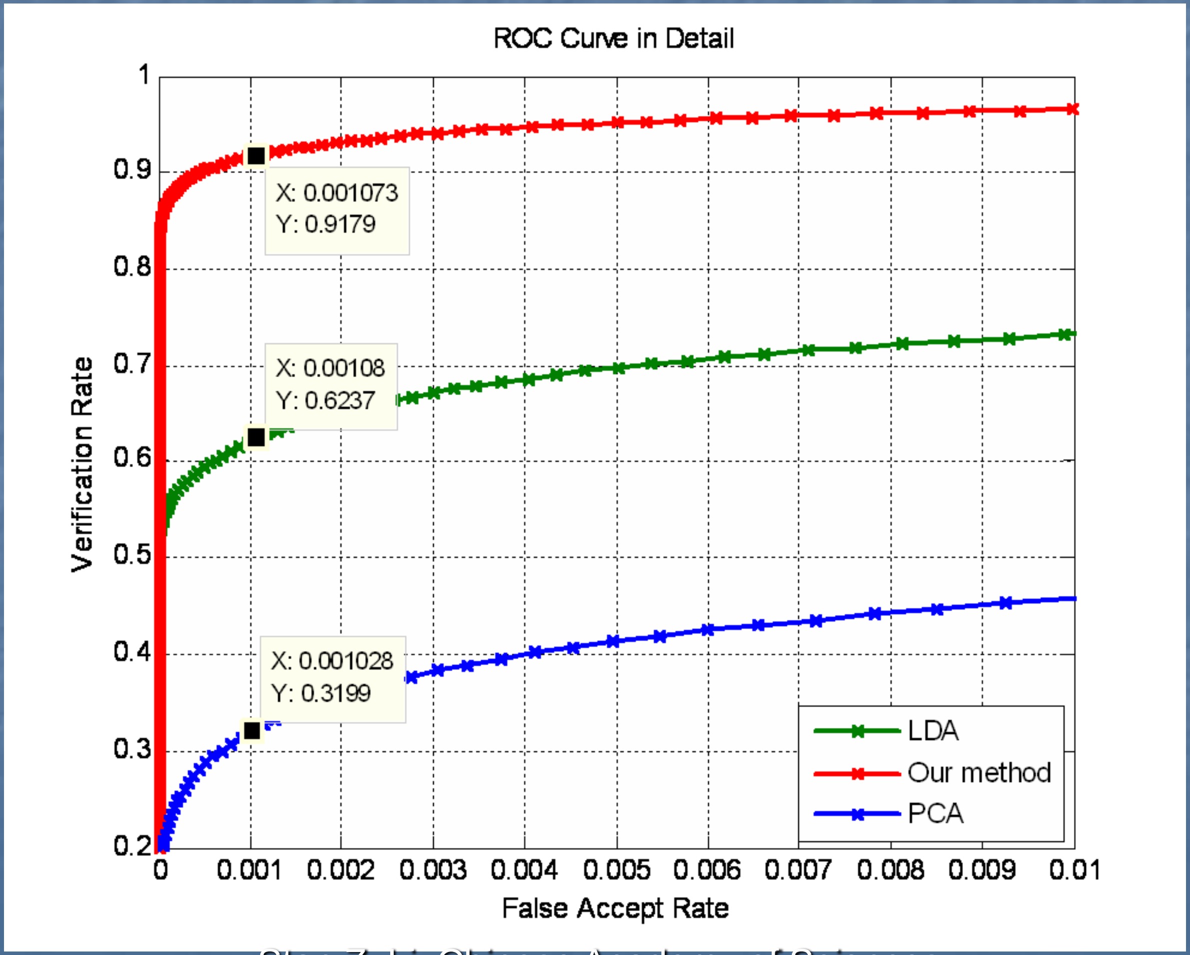




$$I(x,y) = \rho(x,y)\, n^T(x,y)\, s$$
$$= F^T(x,y)\, s$$

with $s = (0,0,1)$

Performance Comparison:
LBP+Boosting vs PCA vc LDA



Scenario Tests

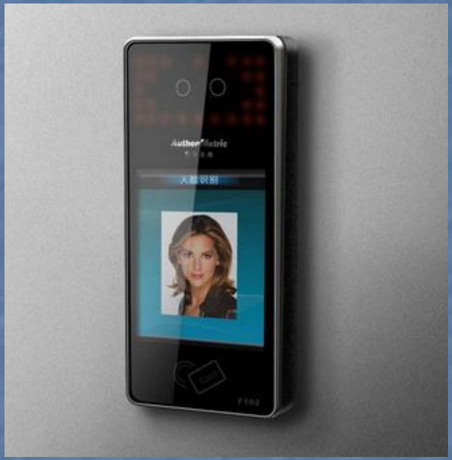
Development NIR Face System in 2005

	# Persons	# Sessions	# Sessions Accepted	# Sessions Rejected	Success Rate
Clients	100	4500	4500	0	100%,
Imposters	10+visitors	3300	3	3297	99.85%
Background	1460	-	-	-	-

Table 1. Scenario evaluation statistics.

NIR Face Products

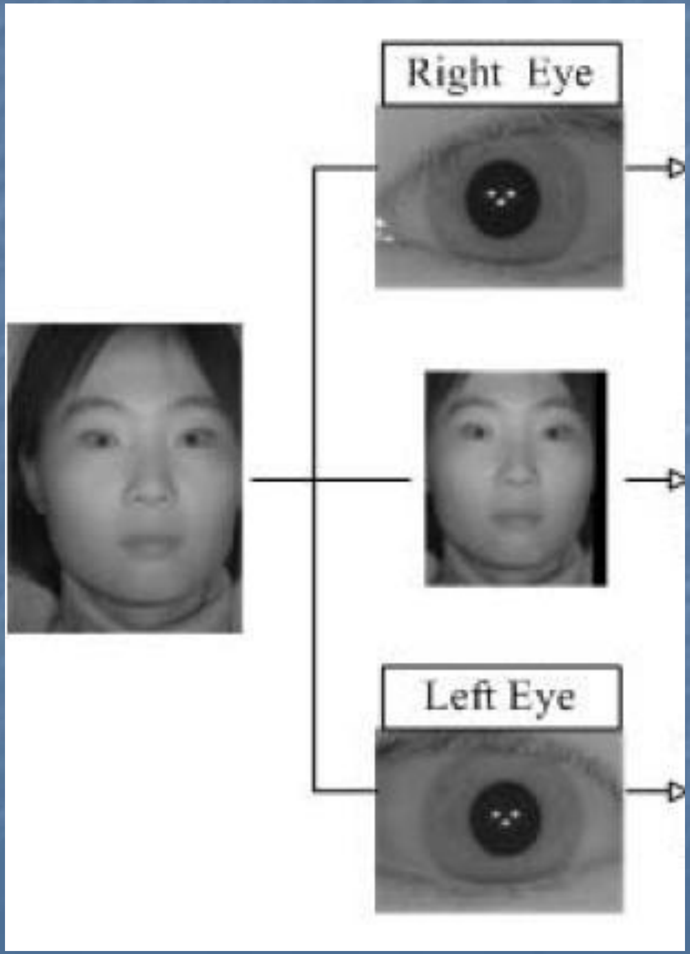
- Platform: PC based and Embedded



- Working Mode: Online, offline, networked

NIR Face+Iris Multimodality

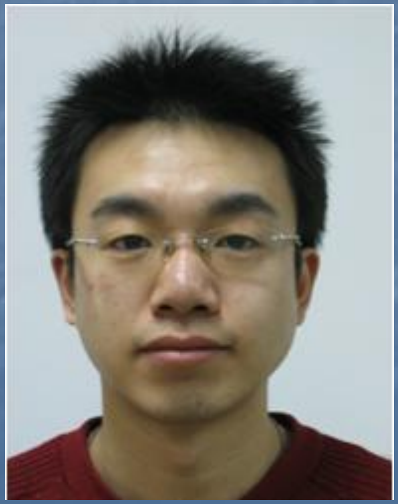
- Face + Iris Unified
  - Multimodality
  - In a single shot
  - NIR imaging for both
  - Non-intrusive
- Iris as part of face
- Challenge
  - Effortless Imaging

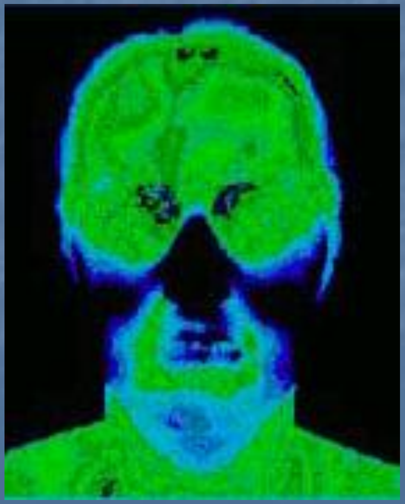


Heterogeneous Face Recognition

Heterogeneous Face Biometrics (HFB)

- Heterogeneous Types of Face Images:
  Visible vs. NIR vs. 3D vs. Thermal NIR
- Face Matching across Heterogeneous Types





HFBs in Broad Sense

- VIS type of face images
  - CCD vs. CMOS sensors,
  - photo scan,
  - face sketch,
  - under different illumination conditions,
  - of different image resolutions,
  - of different image quality.

Significance of HFB Research

- As standalone face biometric technology
- As an added module for multimodal face recognition
- Addressing underlying issues in existing face biometrics
- Research and development on HFBs investigate problems caused by heterogeneities in homogeneous face biometrics and may lead to better solutions
- Provides new directions for face based biometrics, image analysis, pattern recognition and machine learning

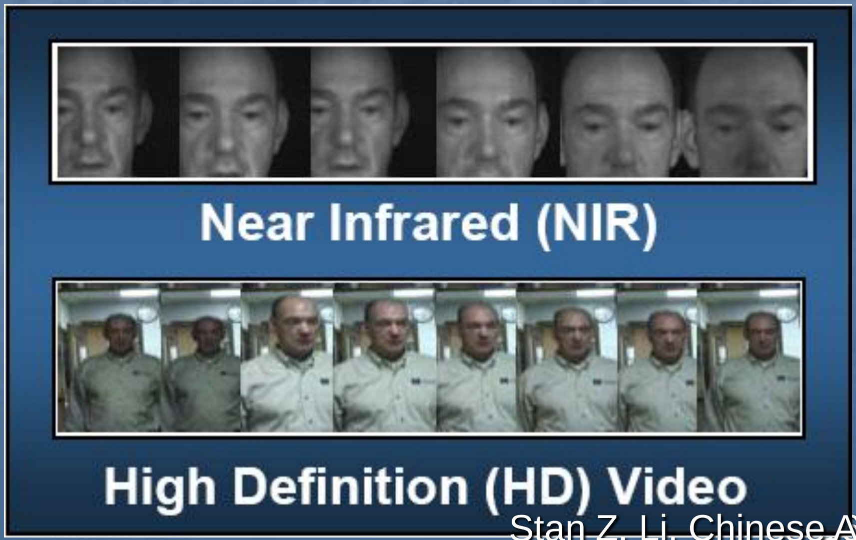
Research Issues

- Understanding heterogeneous image formation models
- Discovering relations between heterogeneous images
- Formulating transformation of one type to another
- Common feature extraction
- Matching across heterogeneous images
- CASIA Heterogeneous Face Biometrics (HFB) database

  http://www.cbsr.ia.ac.cn/english/HFB%20Databases.asp

Influence on NIST Projects

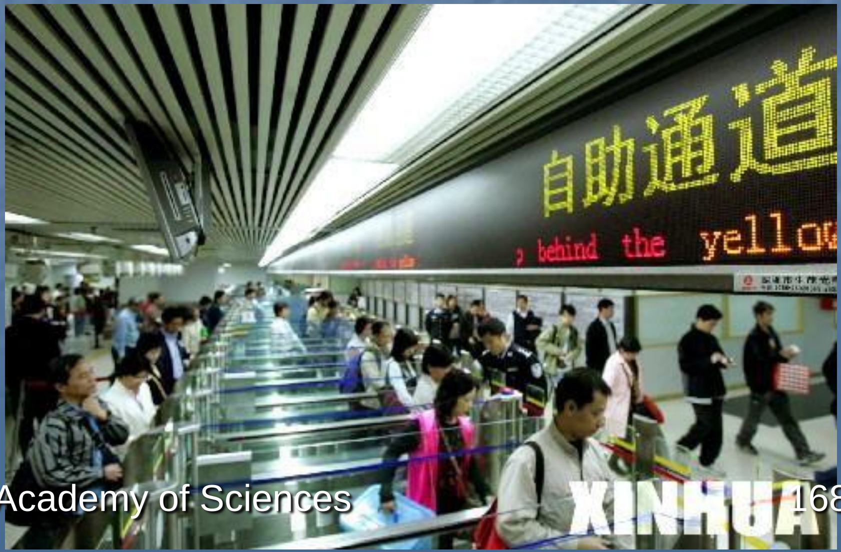
- NIST（National Institute of Standards and Technology）Multiple Biometric Grand Challenge （MBGC 2008）
  - Include NIR Face Video and VIS Color Face images
  - Input: NIR face image
  - Enrollment: VIS face image
  - NIR vs VIS
  - Partial Face Matching



Some More Applications

Biometric Border-Crossing: ShenZhen – HongKong

- 400,000 border-crossings every day
- Two scenarios： Passengers & Vehicle Drivers
- 150 gates deployed by now
- Two Modalities: Face & Fingerprint
- 1,600,000 people enrolled.
- Verification Speed: 6 sec / crossing

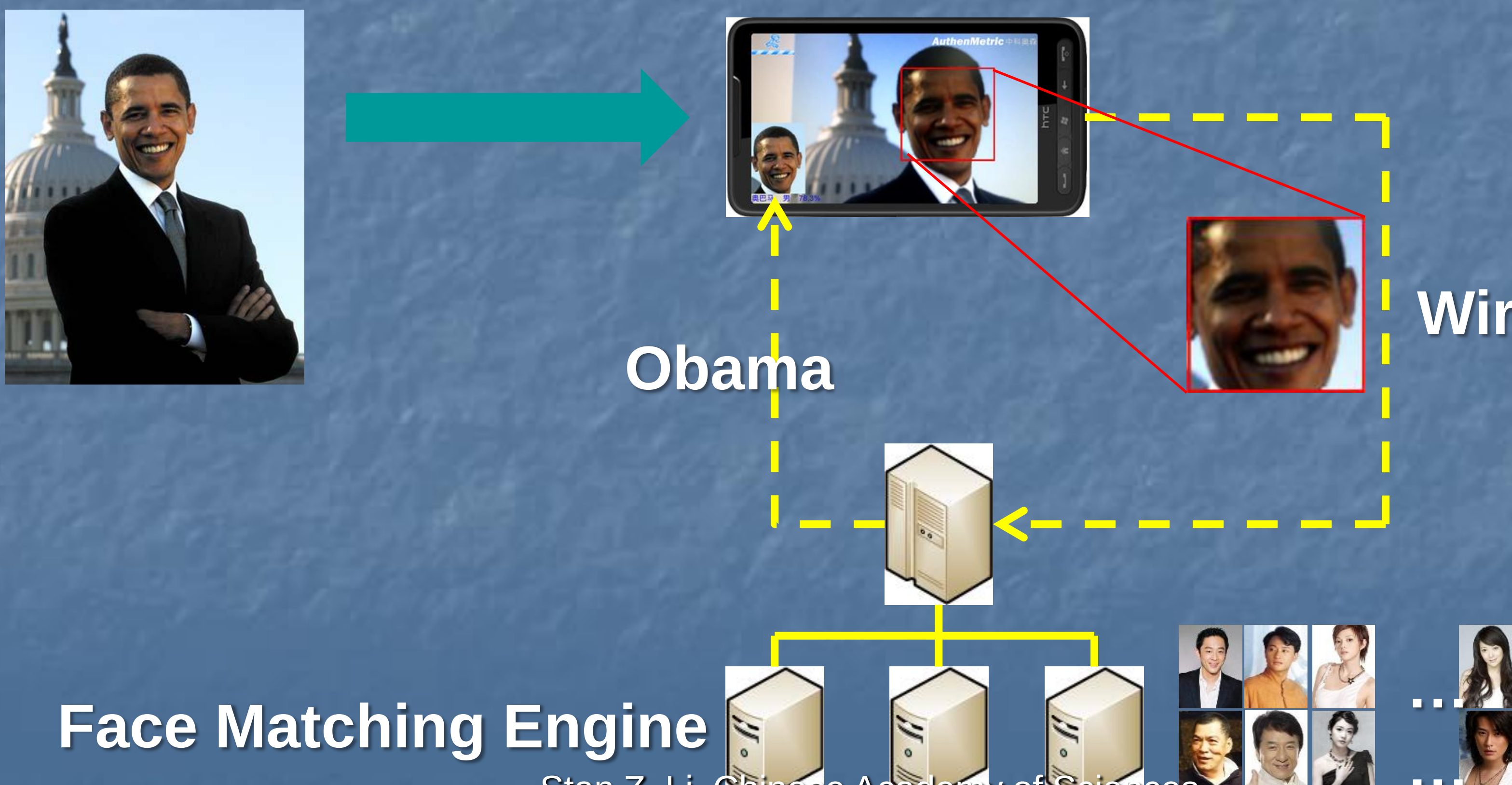


Beijing Olympic 2008

- RFID tickets associated with identities
- Verification of identity
- Video capture vs photo scan

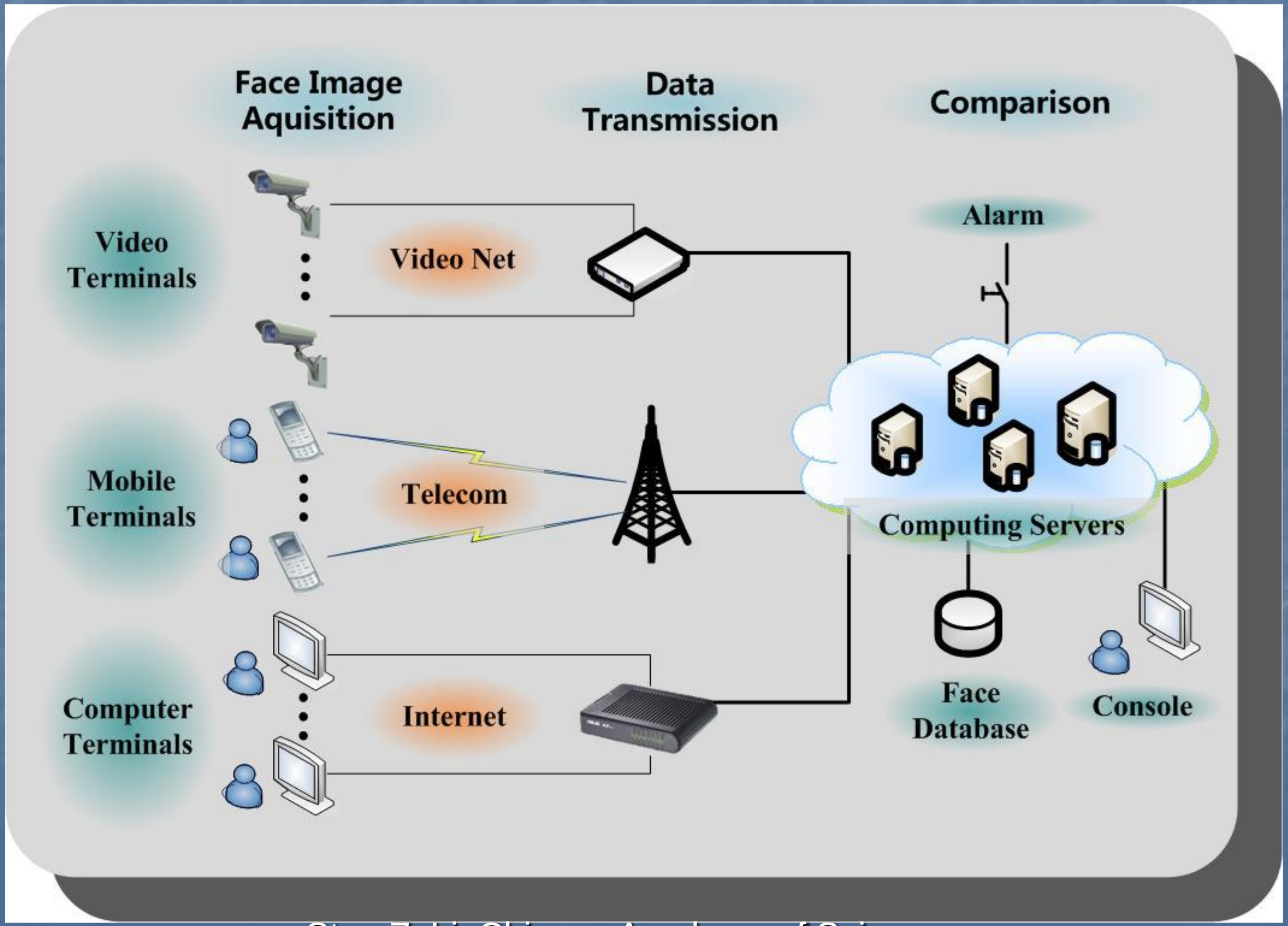
Face Recognition on Mobile

Ubiquitous face recognition



Mobile Recognition Results

Ubiquitous Face Recognition -- A unified platform



Face Surveillance and Identification

- Fuse face recognition, object tracking and Id.
- Comparison with watch-list

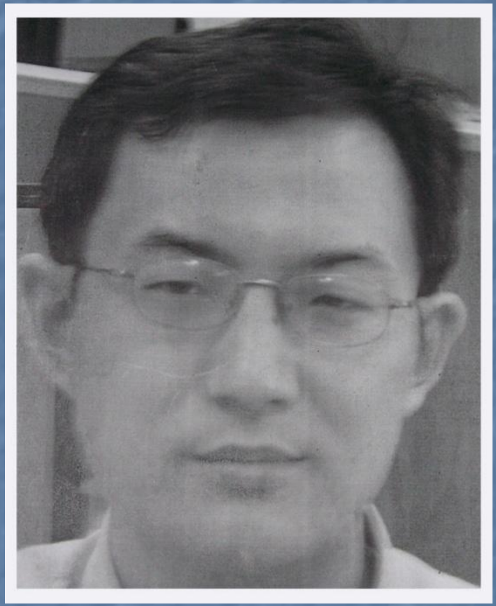
Discovery Report

Deployed

More Challenges

Spoofing Attacks

- Printed Photo, Video Replay, 3D Model



Reported Cases

Case 1. A young man, disguised as an old, cheated Canadian Airline security, 10-29-2010.

Case 2. Google phone "Face Recognition Unlock" function can be easily spoofed by photos, 2011.

Face Anti-spoofing

- To classify a live face from fakes
- HCI-based methods: Challenge-Response
  - Computer requires the subjects to exhibit specific facial motions, the detection of which determines the liveness.

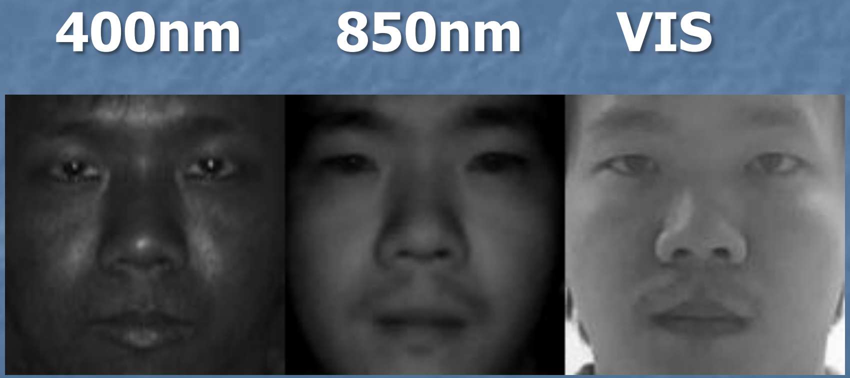
- Multi-modality methods:
  - Face, voice, gesture modalities, ...

Spoofing Attacks

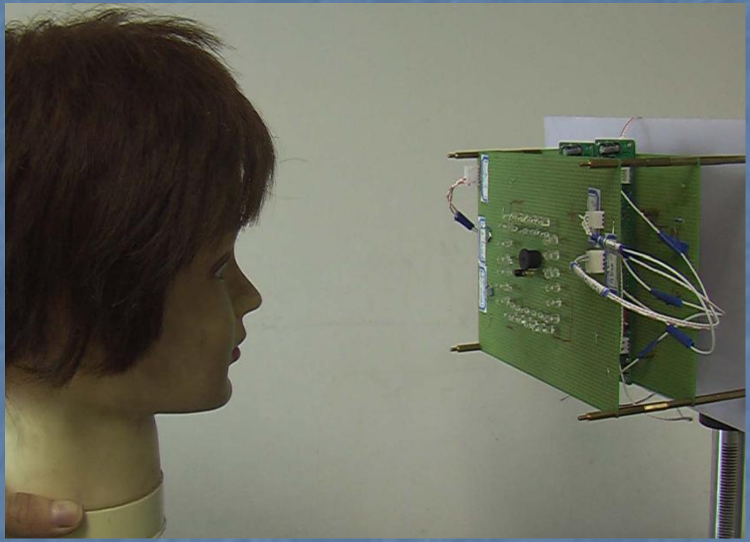
Face spoofing can be more than one might think …

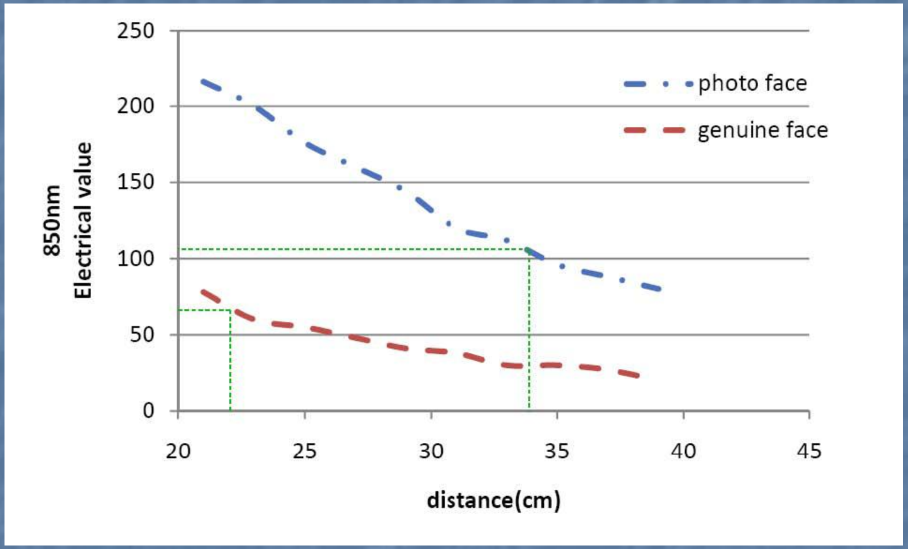
Nonintrusive Solution: Multispectral Techs

- **Imaging faces beyond visible spectrum**
- **Analyze reflectance property**
- **Train statistical model**



Previous Results

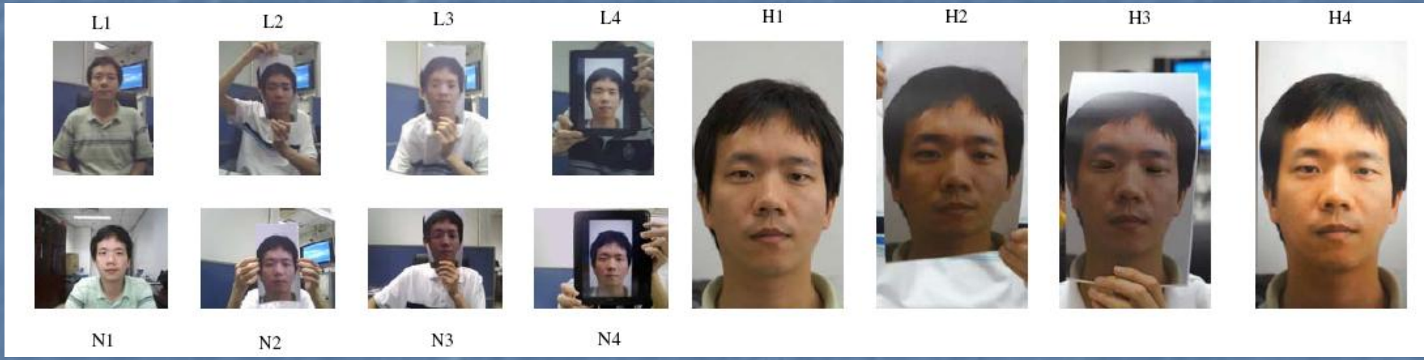




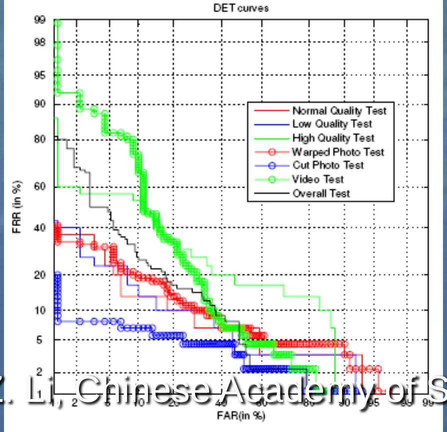
Attack Detection	Accuracy
Genuine vs Photo	92.2%
Genuine vs Video Replay	100%
Genuine vs Mask	89.2%

CASIA Face Anti-spoofing Database

A diverse and comprehensive database for evaluating anti-spoofing techniques



Baseline evaluation under different scenarios – provided



Unsolved Problems

Pose

Make up

Facial Wear

Aging

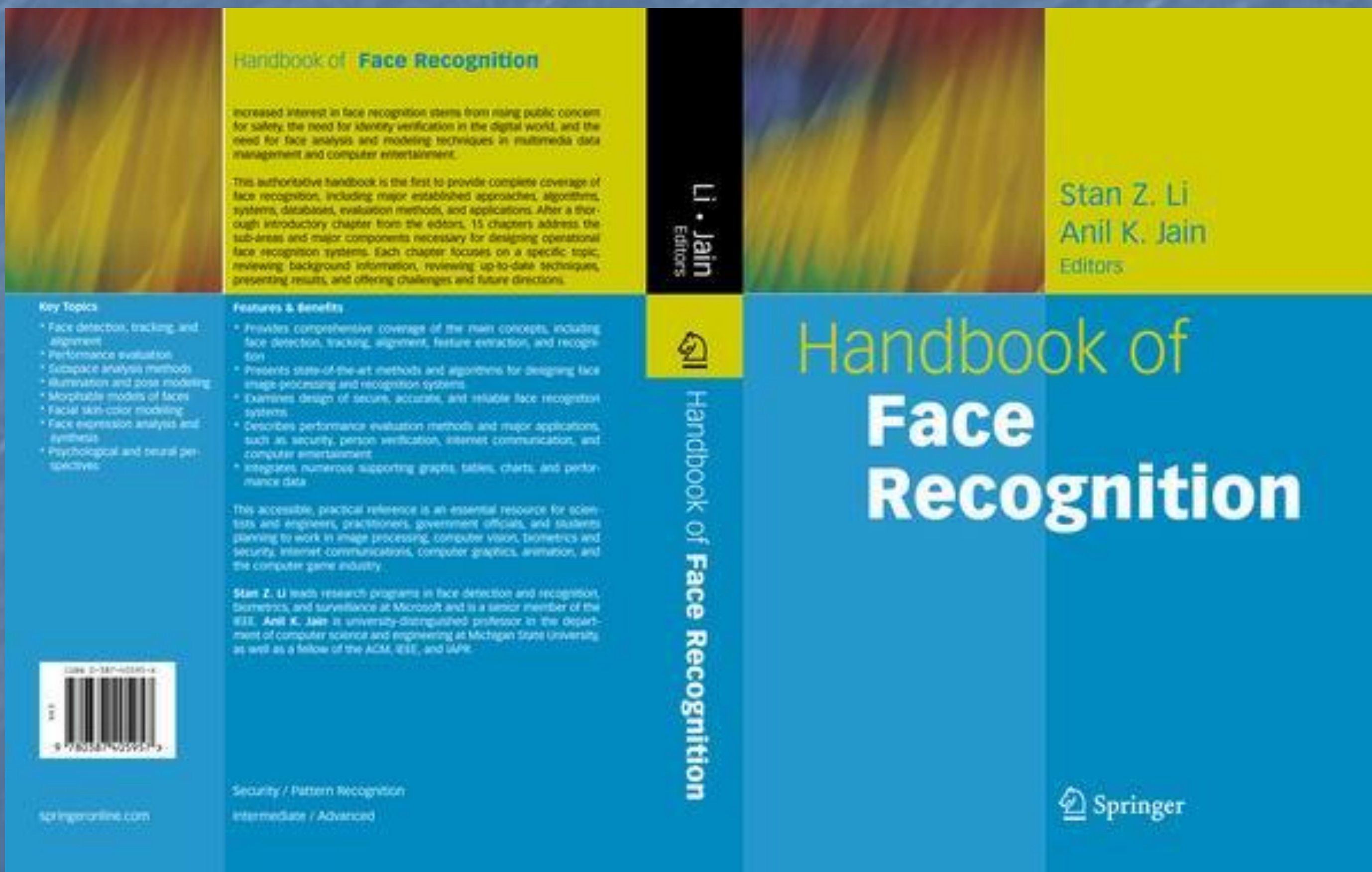
Center for Biometrics and Security Research

Institute of Automation, Chinese Academy of Sciences
www.cbsr.ia.ac.cn

Goal: To achieve excellence in R & D

- Biometrics
  - Face
  - Iris
  - Fingerprint
  - Palmprint
  - Gait
  - Signature
- Intelligent surveillance
  - Security surveillance
  - Traffic surveillance
  - Object detection, classification, tracking
  - Abnormal event detection

Commercial: The Face Handbook



Thank you