Distance Metric Learning for Set-based Visual Recognition

Ruiping Wang

Institute of Computing Technology (ICT),
Chinese Academy of Sciences (CAS)

June 7, 2015

中国科学院计算技术研究所
Institute of Computing Technology,Chinese Academy of Sciences

Outline

- Background
- Literature review
- Evaluations
- Summary

Face Recognition with Single Image

- Identification
  - Typical applications
    - Photo matching (1:N)
    - Watch list screening (1:N+1)
  - Performance metric
    - FR(@FAR)

- Verification
  - Typical applications
    - Access control (1:1)
    - E-passport (1:1)
  - Performance metric
    - ROC: FRR+FAR

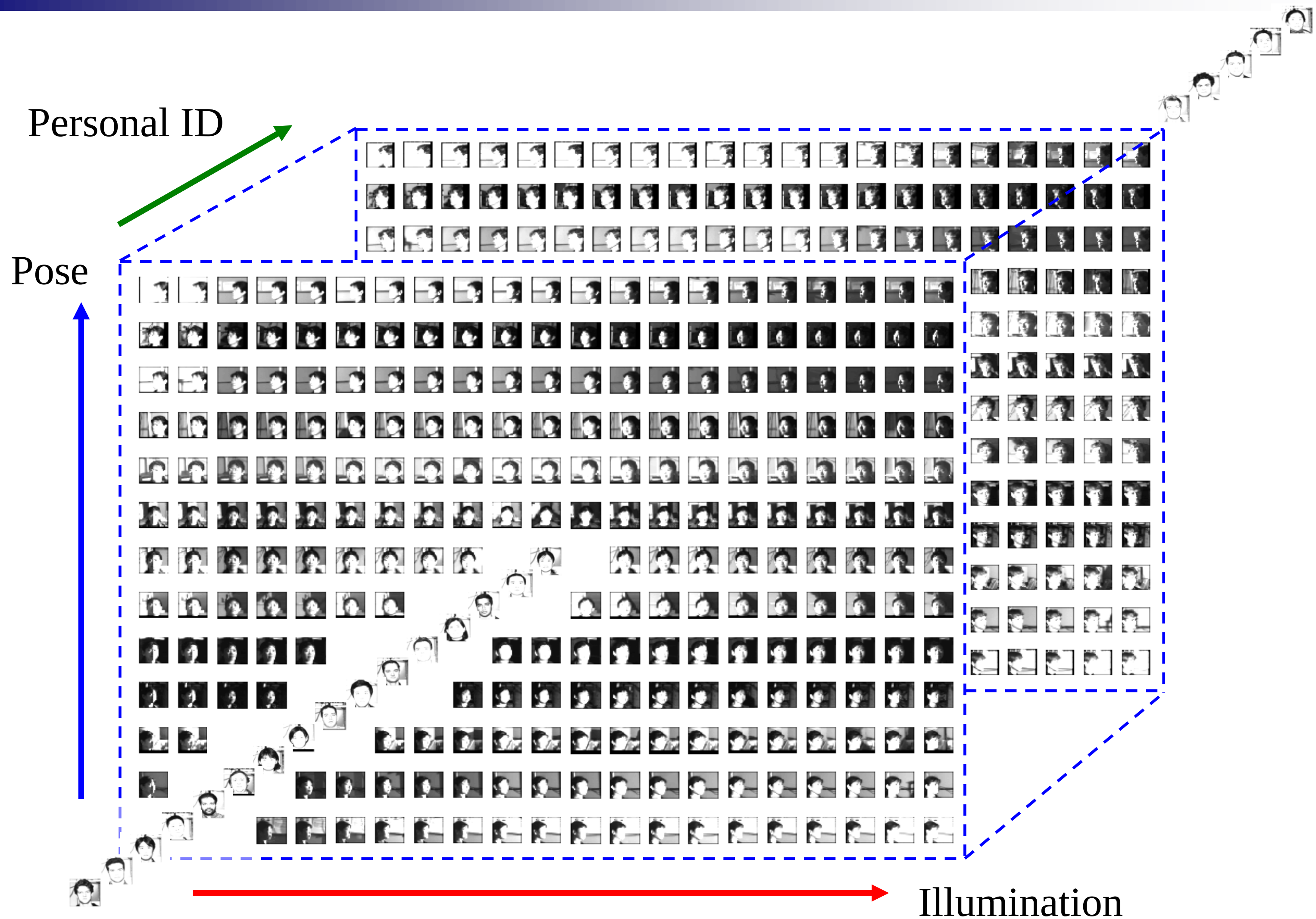
Who is this celebrity?

Are they the same guy?

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

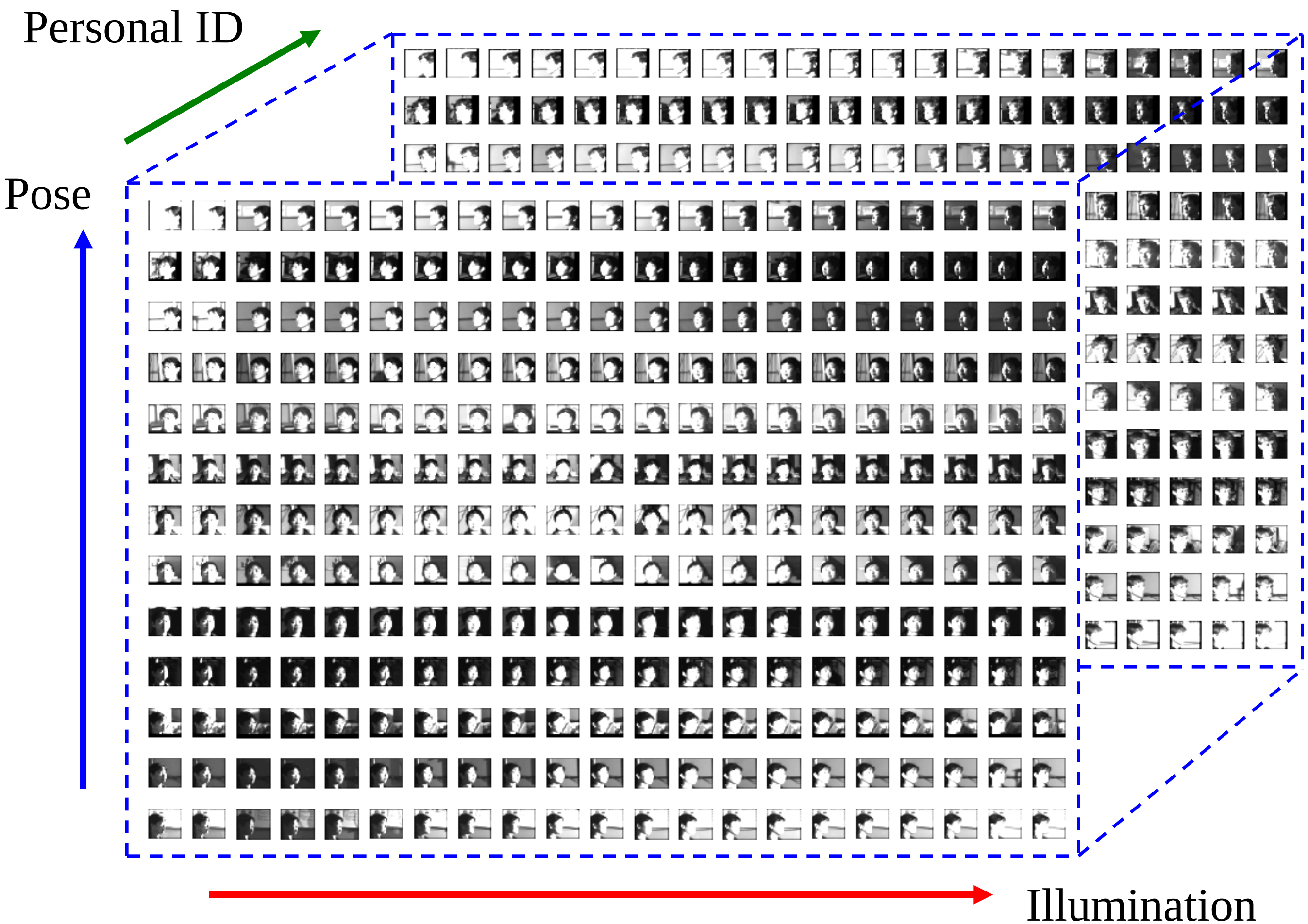
Challenges

Challenges



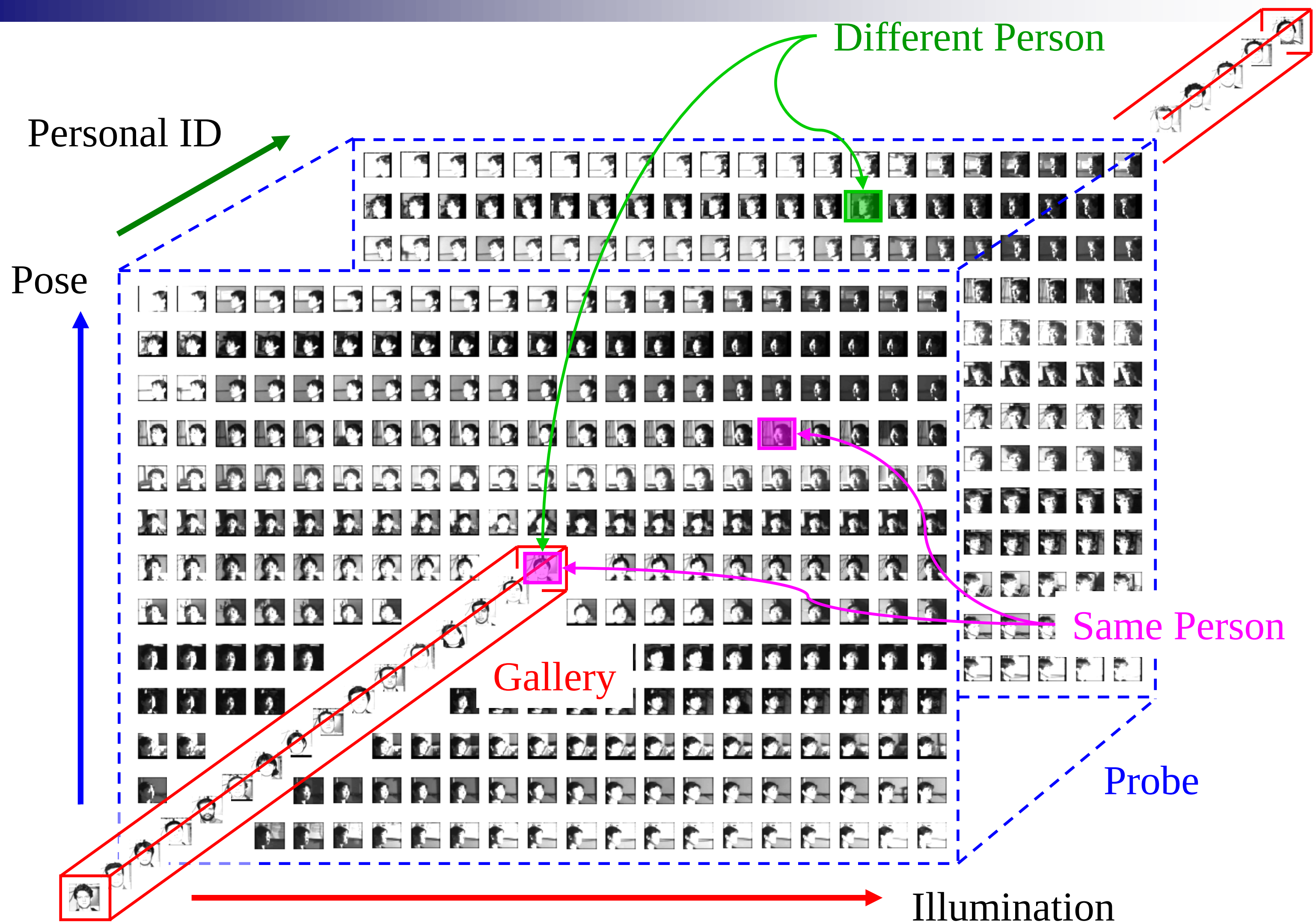
中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Challenges



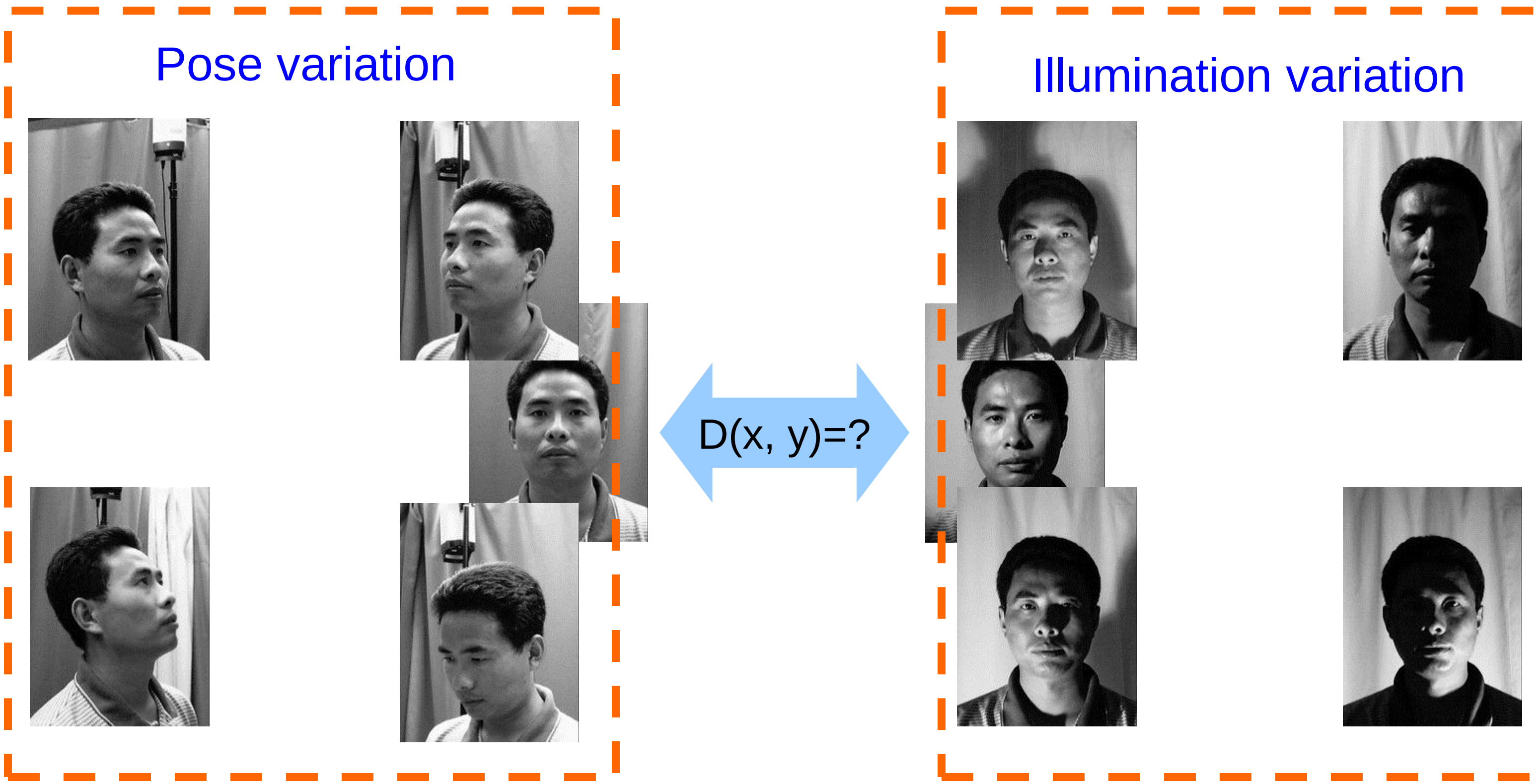
中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Challenges


Different Person
Personal ID
Pose
Same Person
Gallery
Probe
Illumination

Challenges

- Intra-class variation vs. inter-class variation
  - Distance measure →semantic meaning
  - Sample-based metric learning is made even harder



Face Recognition with Videos

- Video surveillance

Seeking missing children

Searching criminal suspects

http://www.youtube.com/watch?v=M80DXI932OE
http://www.youtube.com/watch?v=RfJsGeq0xRA#t=22

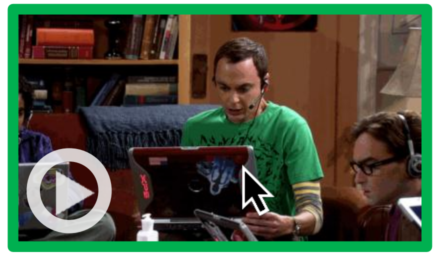
中科院计算所

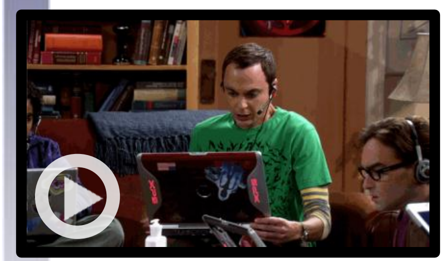
Institute of Computing Technology, Chinese Academy of Sciences

Face Recognition with Videos

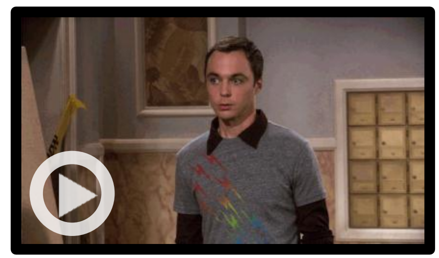
- Video shot retrieval

Smart TV-Series Character
Shots Retrieval System
"the Big Bang Theory"




S01E01: 10'48"

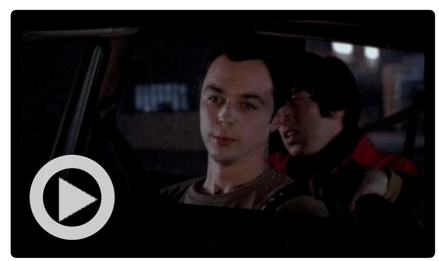

S01E06: 05'22"

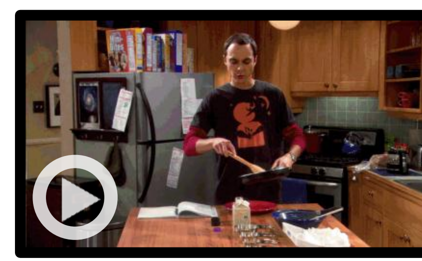

S01E02: 04'20"


S01E02: 00'21"


S01E03: 08'06"


S01E05: 09'23"


S01E01: 22'20"


S01E04: 03'42"

1 2 3 4 5 6 7 8 9 Next Page

Treating Video as Image Set

- A new & different problem
  - Unconstrained acquisition conditions
  - Complex appearance variations
  - Two phases: set modeling + set matching

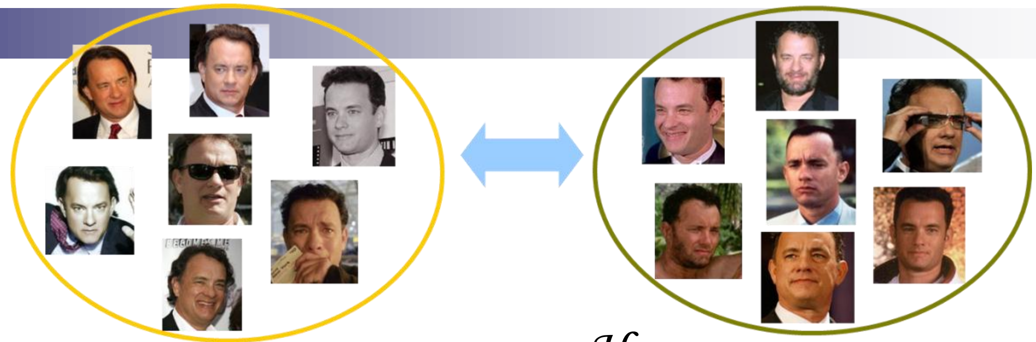
New paradigm:
set-based metric learning

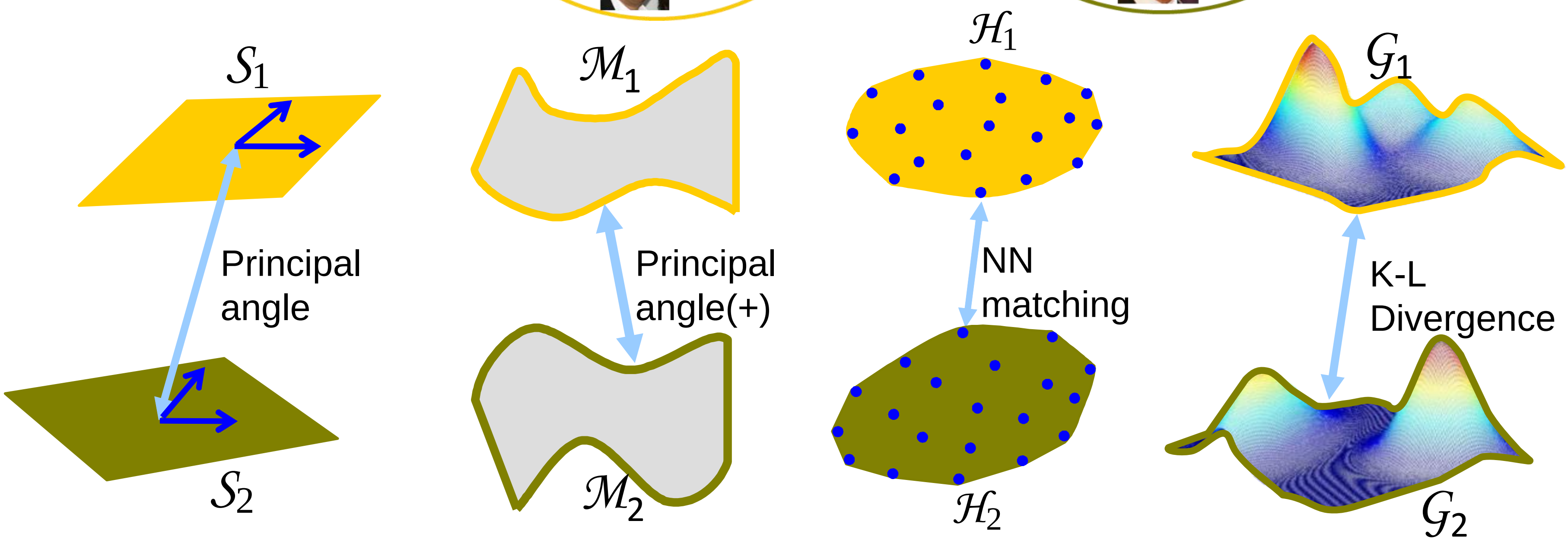
Outline

- Background
- Literature review
- Evaluations
- Summary

Overview of previous works

From the view of set modeling





◆Linear subspace
[Yamaguchi, FG'98]
[Kim, PAMI'07]
[Hamm, ICML'08]
[Huang, CVPR'15]

◆Nonlinear manifold
[Kim, BMVC'05]
[Wang, CVPR'08]
[Wang, CVPR'09]
[Chen, CVPR'13]
[Lu, CVPR'15]

◆Affine/Convex hull
[Cevikalp, CVPR'10]
[Hu, CVPR'11]
[Zhu, ICCV'13]

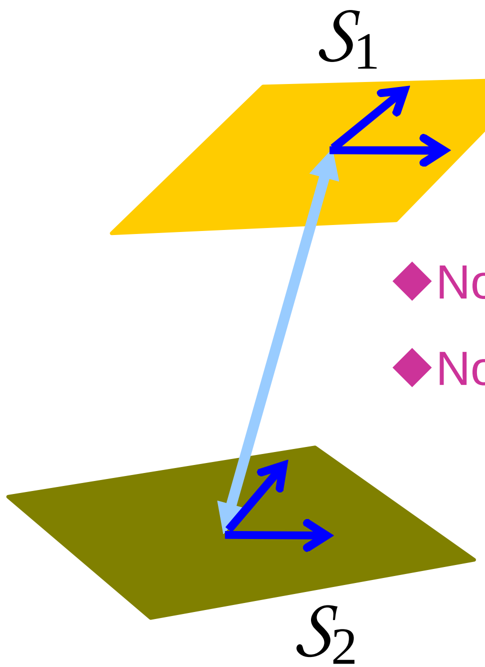
◆Statistics
[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]
[Wang, CVPR'12]
[Harandi, ECCV'14]
[Wang, CVPR'15]

Overview of previous works

- Set modeling
  - Linear subspace→Nonlinear manifold
  - Affine/Convex Hull (affine subspace)
  - Parametric PDFs → Statistics
- Set matching—basic distance
  - Principal angles-based measure
  - Nearest neighbor (NN) matching approach
  - K-L divergence→SPD Riemannian metric…
- Set matching—metric learning
  - Learning in Euclidean space
  - Learning on Riemannian manifold

Set model I: linear subspace

- Properties
  - PCA on the set of image samples to get subspace
  - Loose characterization of the set distribution region
  - Principal angles-based measure discards the varying importance of different variance directions
- Methods
  - MSM [FG'98]
  - DCC [PAMI'07]
  - GDA [ICML'08]
  - GGDA [CVPR'11]
  - PML [CVPR'15]
  - …



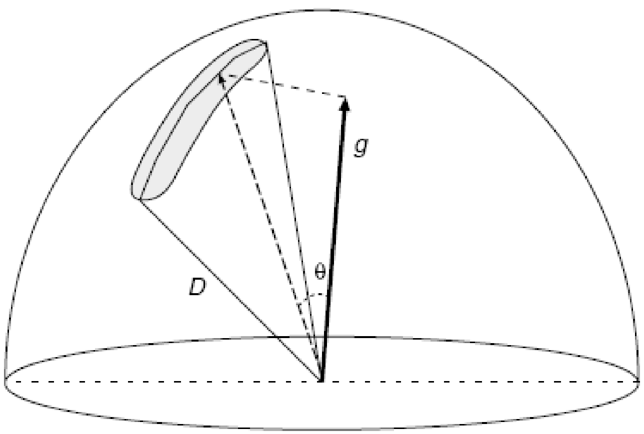
◆No distribution boundary

◆No direction difference

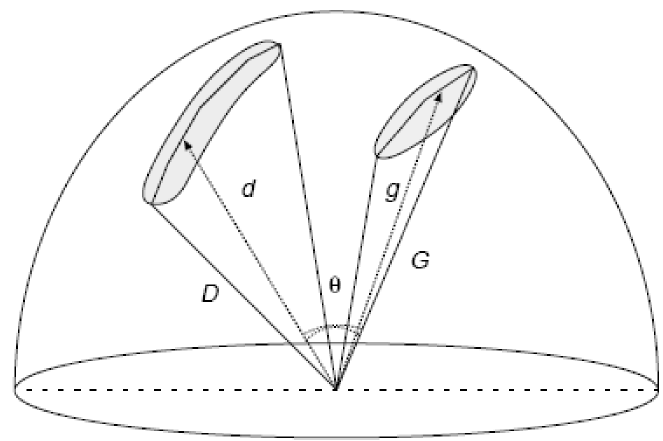
Set model I: linear subspace

- MSM (Mutual Subspace Method) [FG'98]
  - Pioneering work on image set classification
  - First exploit principal angles as subspace distance
  - Metric learning: N/A

$$cos^2\theta = \sup_{d\in D, g\in G, ||d||\neq 0, ||g||\neq 0} \frac{|(d,g)|^2}{||d||^2||g||^2}$$



subspace method



Mutual subspace method

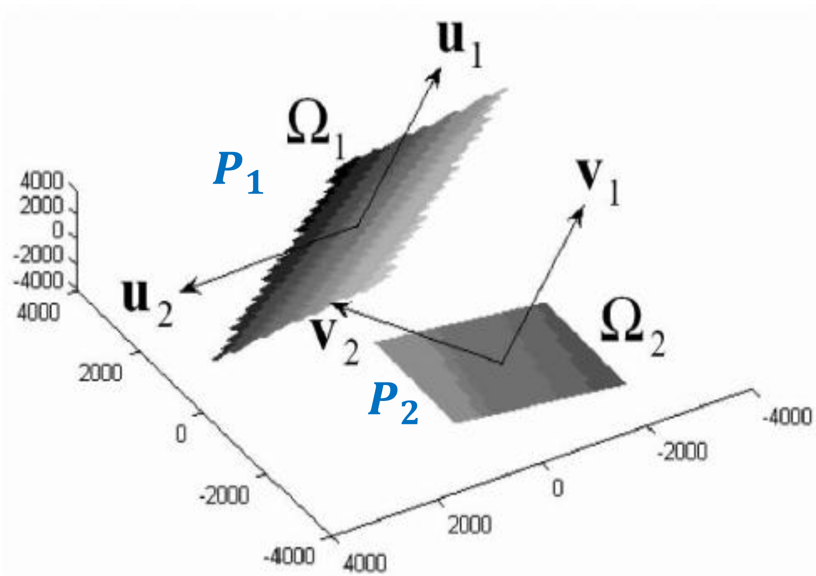
[1] O. Yamaguchi, K. Fukui, and K. Maeda. Face Recognition Using Temporal Image Sequence. *IEEE FG 1998.*

Set model I: linear subspace

- DCC (Discriminant Canonical Correlations) [PAMI'07]
  - Metric learning: in Euclidean space

Set 1: $\boldsymbol{X}_1$

Set 2: $\boldsymbol{X}_2$



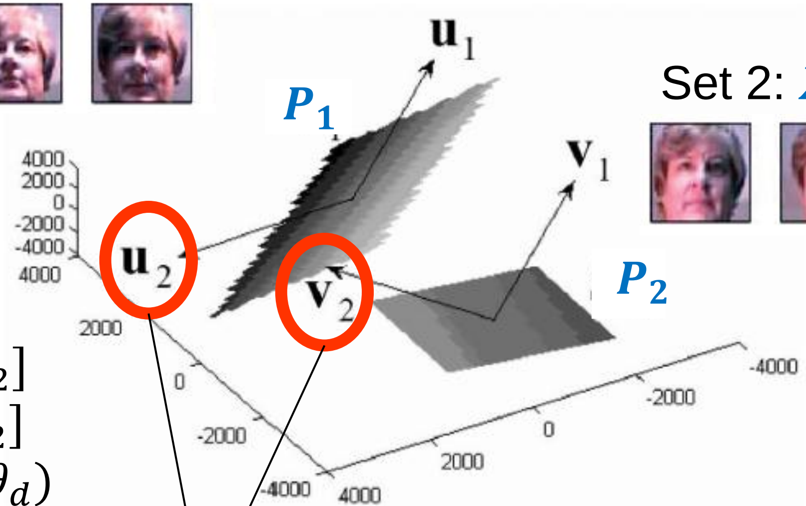
Linear subspace by:
orthonormal basis matrix
$\boldsymbol{X}_i\boldsymbol{X}_i^T \simeq \boldsymbol{P}_i\boldsymbol{\Lambda}_i\boldsymbol{P}_i^T$

[1] T. Kim, J. Kittler, and R. Cipolla. Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations. *IEEE T-PAMI, 2007.*

DCC

- Canonical Correlations/Principal Angles
  - Canonical vectors→common variation modes

Set 1: $\boldsymbol{X_1}$



Set 2: $\boldsymbol{X_2}$

$$\boldsymbol{P}_1^T\boldsymbol{P}_2 = \boldsymbol{Q}_{12}\boldsymbol{\Lambda}\boldsymbol{Q}_{21}^T$$
$$\boldsymbol{U} = \boldsymbol{P}_1\boldsymbol{Q}_{12} = [\boldsymbol{u}_1, \dots, \boldsymbol{u}_2]$$
$$\boldsymbol{V} = \boldsymbol{P}_2\boldsymbol{Q}_{21} = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_2]$$
$$\boldsymbol{\Lambda} = diag(\cos\theta_1, \dots, \cos\theta_d)$$

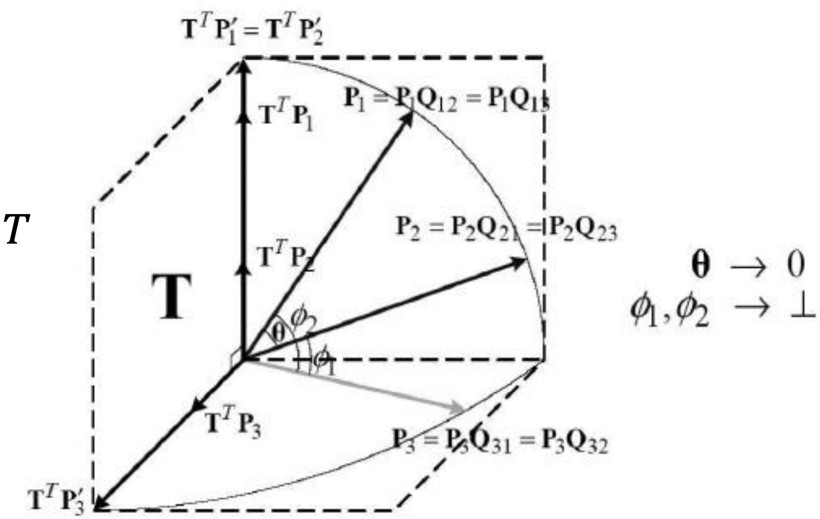
Canonical Correlation: $\cos\theta_i$

Principal Angles: θ_i

Canonical vectors

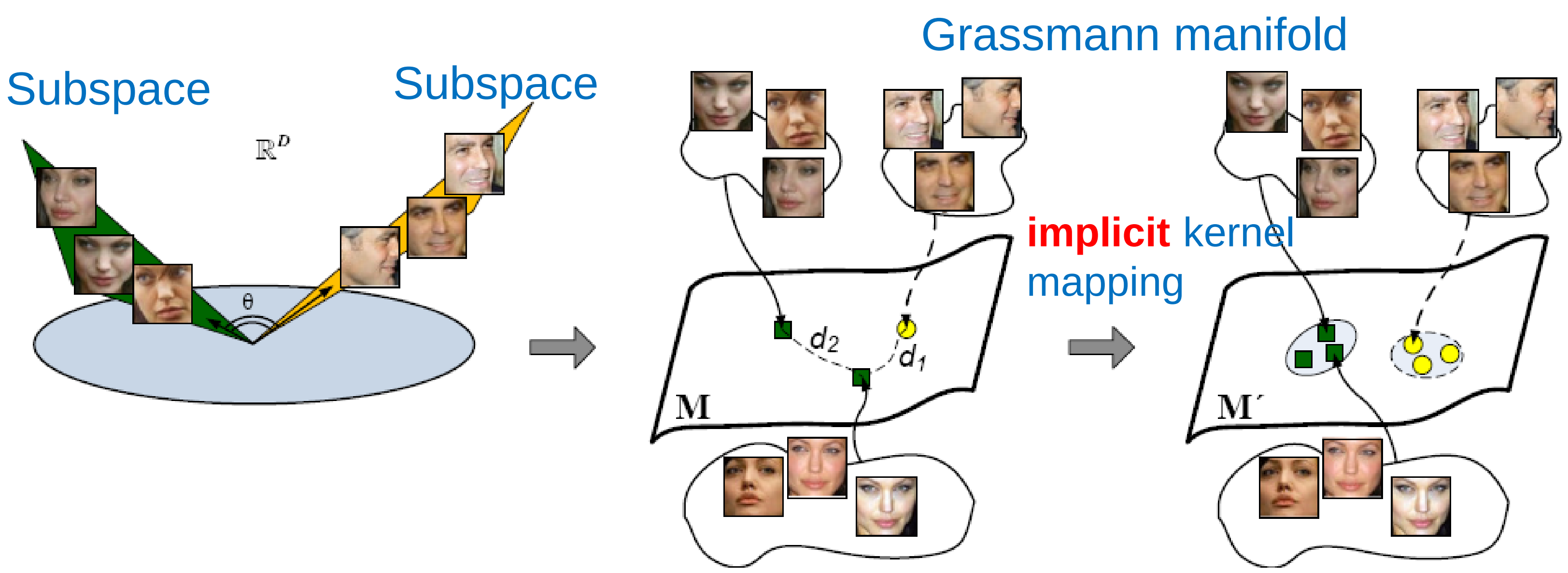
DCC

- Discriminative learning
  - Linear transformation
    - $\boldsymbol{T}$: $\boldsymbol{X}_i \longrightarrow \boldsymbol{Y}_i = \boldsymbol{T}^T\boldsymbol{X}_i$
  - Representation
    - $\boldsymbol{Y}_i\boldsymbol{Y}_i^T = (\boldsymbol{T}^T\boldsymbol{X}_i)(\boldsymbol{T}^T\boldsymbol{X}_i)^T$
      $\simeq (\boldsymbol{T}^T\boldsymbol{P}_i)\boldsymbol{\Lambda}_i(\boldsymbol{T}^T\boldsymbol{P}_i)^T$
  - Set similarity
    - $F_{ij} = \max_{\boldsymbol{Q}_{ij},\boldsymbol{Q}_{ji}} tr(M_{ij})$
      - $M_{ij} = \boldsymbol{Q}_{ij}^T\boldsymbol{P'}_i^T\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{P}_j^T\boldsymbol{Q}_{ji}$
  - Discriminant function
    - $\boldsymbol{T} = \max_{arg\,\boldsymbol{T}} tr(\boldsymbol{T}^T\boldsymbol{S}_b\boldsymbol{T})\,/tr(\boldsymbol{T}^T\boldsymbol{S}_w\boldsymbol{T})$



Set model I: linear subspace

- GDA [ICML'08] / GGDA [CVPR'11]
  - Treat subspaces as points on Grassmann manifold
  - Metric learning: on Riemannian manifold



[1] J. Hamm and D. D. Lee. Grassmann Discriminant Analysis: a Unifying View on Subspace-Based Learning. *ICML 2008*.

[2] M. Harandi, C. Sanderson, S. Shirazi, B. Lovell. Graph Embedding Discriminant Analysis on Grassmannian Manifolds for Improved Image Set Matching. *IEEE CVPR 2011.*

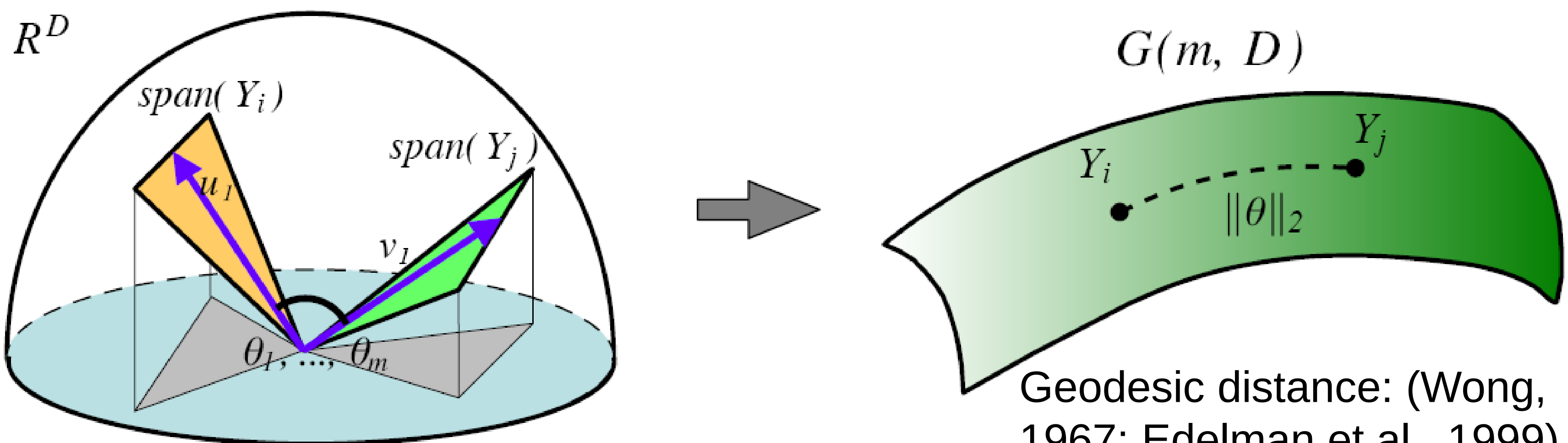
中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

GDA/GGDA

- Projection metric
  - $d_P(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = (\sum_i \sin^2 \theta_i)^{1/2} = 2^{-1/2} \| \boldsymbol{Y}_1 \boldsymbol{Y}_1^T - \boldsymbol{Y}_2 \boldsymbol{Y}_2^T \|_F$

θ_i: Principal angles



Geodesic distance: (Wong, 1967; Edelman et al., 1999)

$$d_G^2(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \sum_i \theta_i^2$$

GDA/GGDA

- Projection kernel
  - Projection embedding (isometric)
    - $\Psi_P: \mathcal{G}(m, D) \longrightarrow \mathbb{R}^{D\times D},\ span(\boldsymbol{Y}) \longrightarrow \boldsymbol{Y}\boldsymbol{Y}^T$
  - The inner-product of $\mathbb{R}^{D\times D}$
    - $tr((\boldsymbol{Y}_1\boldsymbol{Y}_1^T)(\boldsymbol{Y}_2\boldsymbol{Y}_2^T)) = \left\|\boldsymbol{Y}_1^T\boldsymbol{Y}_2\right\|_F^2$
  - Grassmann kernel (positive definite kernel)
    - $k_P(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \left\|\boldsymbol{Y}_1^T\boldsymbol{Y}_2\right\|_F^2$

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

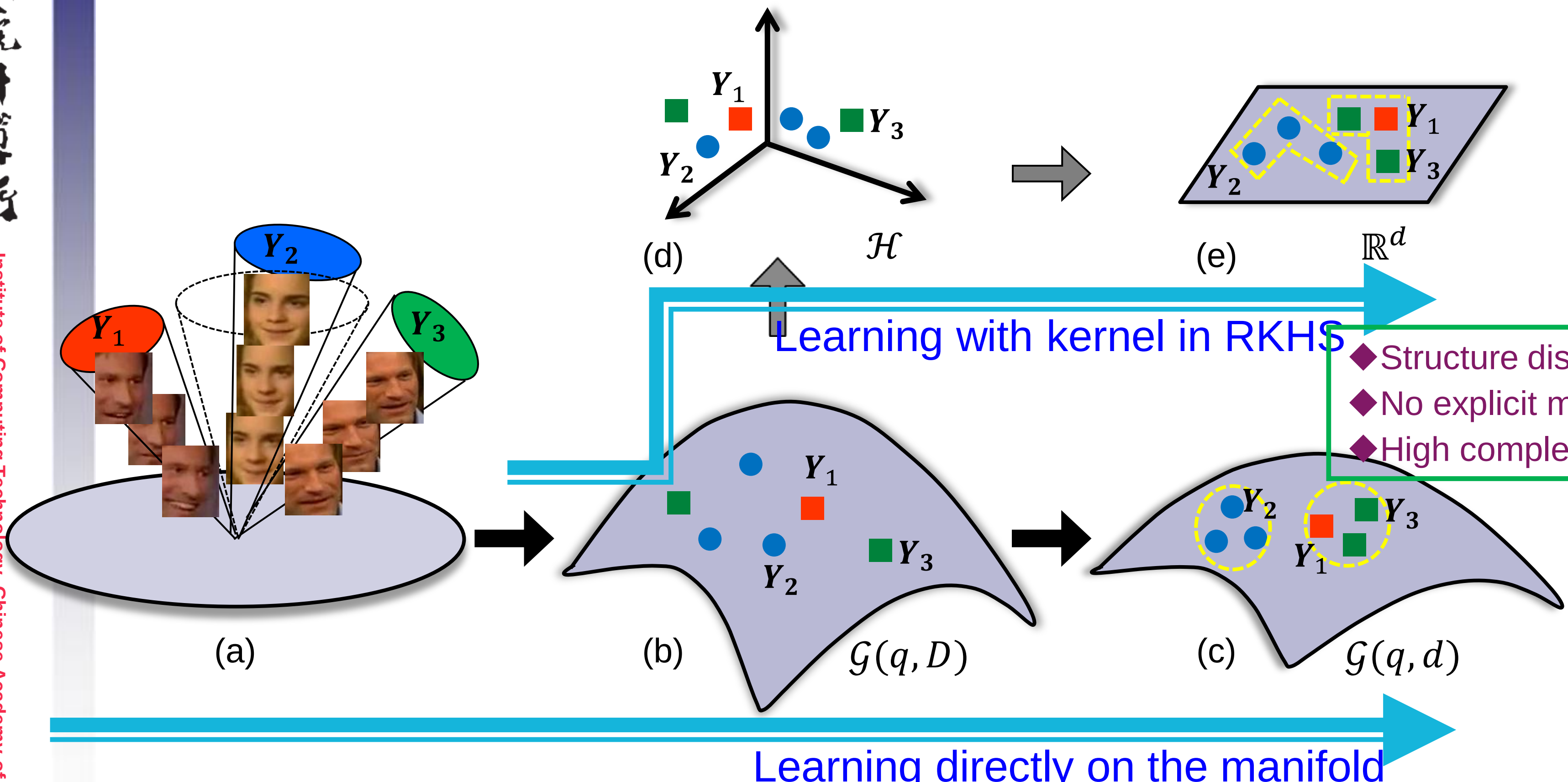
GDA/GGDA

- Discriminative learning
  - Classical kernel methods using the Grassmann kernel
    - e.g., Kernel LDA / kernel Graph embedding
  - $$\boldsymbol{\alpha}^{*} = \arg\max_{\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}^{T}\boldsymbol{K}\boldsymbol{W}\boldsymbol{K}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{T}\boldsymbol{K}\boldsymbol{K}\boldsymbol{\alpha}}$$

Grassmann kernel

Set model I: linear subspace

- PML (Projection Metric Learning) [CVPR'15]
  - Metric learning: on Riemannian manifold

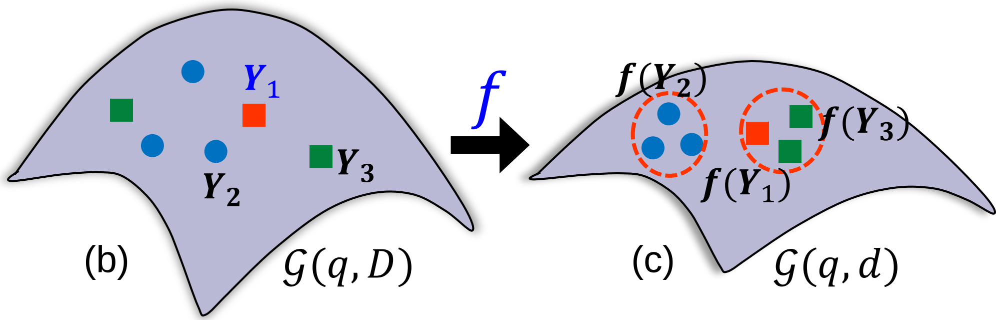


◆Structure distortion
◆No explicit mapping
◆High complexity

[1] Z. Huang, R. Wang, S. Shan, X. Chen. Projection Metric Learning on Grassmann Manifold with Application to Video based Face Recognition. *IEEE CVPR 2015.*

PML

- Explicit manifold to manifold mapping
  - $f(\boldsymbol{Y}) = \boldsymbol{W}^T\boldsymbol{Y} \in \mathcal{G}(q,d), \boldsymbol{Y} \in \mathcal{G}(q,D),\ d \leq D$



- Projection metric on target Grassmann manifold $\mathcal{G}(q,d)$
  - $d_p^2\left(f(\boldsymbol{Y}_\mathrm{i}), f(\boldsymbol{Y}_\mathrm{j})\right) = 2^{-1/2}\left\|(\boldsymbol{W}^T\boldsymbol{Y}_i')(\boldsymbol{W}^T\boldsymbol{Y}_i')^T - (\boldsymbol{W}^T\boldsymbol{Y}_j')(\boldsymbol{W}^T\boldsymbol{Y}_j')^T\right\|_F^2 = 2^{-1/2} tr\left(\boldsymbol{P}^T\boldsymbol{A}_{ij}\boldsymbol{A}_{ij}P\right)$
    - $\boldsymbol{A}_{ij} = (\boldsymbol{Y}_i'\boldsymbol{Y}_i'^T - \boldsymbol{Y}_j'\boldsymbol{Y}_j')^T$, $\boldsymbol{P} = \mathbf{WW}^T$ is a rank-d symmetric positive semidefinite (PSD) matrix of size $D \times D$ (similar form as Mahalanobis matrix)
    - $\boldsymbol{Y}_\mathrm{i}$ needs to be normalized to $\boldsymbol{Y}_i'$ so that the columns of $\boldsymbol{W}^T\boldsymbol{Y}_\mathrm{i}$ are orthonormal

PML

- Discriminative learning
  - Discriminant function
    - Minimize/Maximize the projection distances of any within-class/between-class subspace pairs
    - $J = \min \sum_{l_i = l_j} tr\left(\boldsymbol{P}^T \boldsymbol{A}_{ij} \boldsymbol{A}_{ij} \boldsymbol{P}\right) - \lambda \sum_{l_i \neq l_j} tr\left(\boldsymbol{P}^T \boldsymbol{A}_{ij} \boldsymbol{A}_{ij} \boldsymbol{P}\right)$

      within-class

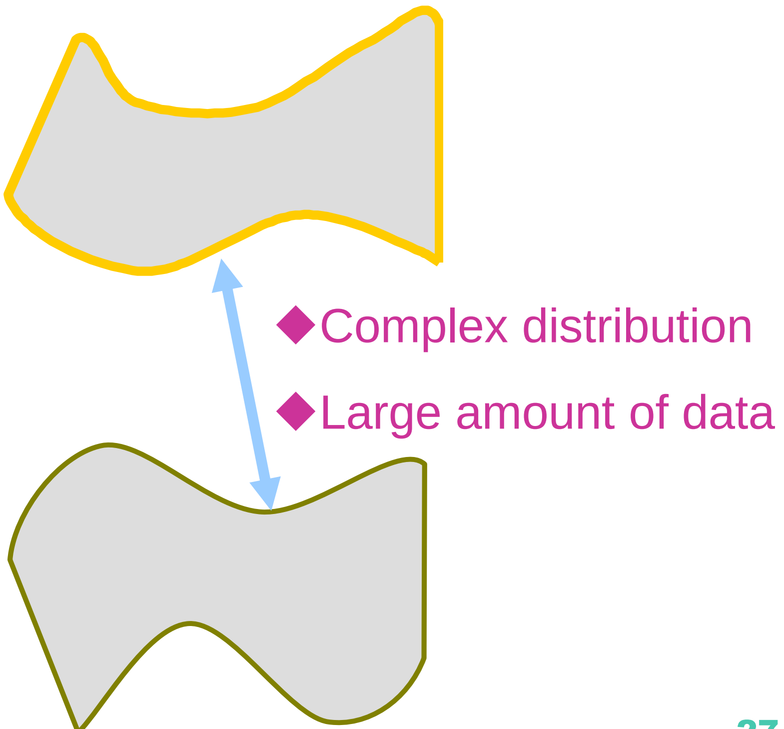
      between-class
  - Optimization algorithm
    - Iterative solution for one of $\boldsymbol{Y}'$ and $\boldsymbol{P}$ by fixing the other
    - Normalization of $\boldsymbol{Y}$ by QR-decomposition
    - Computation of $\boldsymbol{P}$ by Riemannian Conjugate Gradient (RCG) algorithm on the manifold of PSD matrices

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

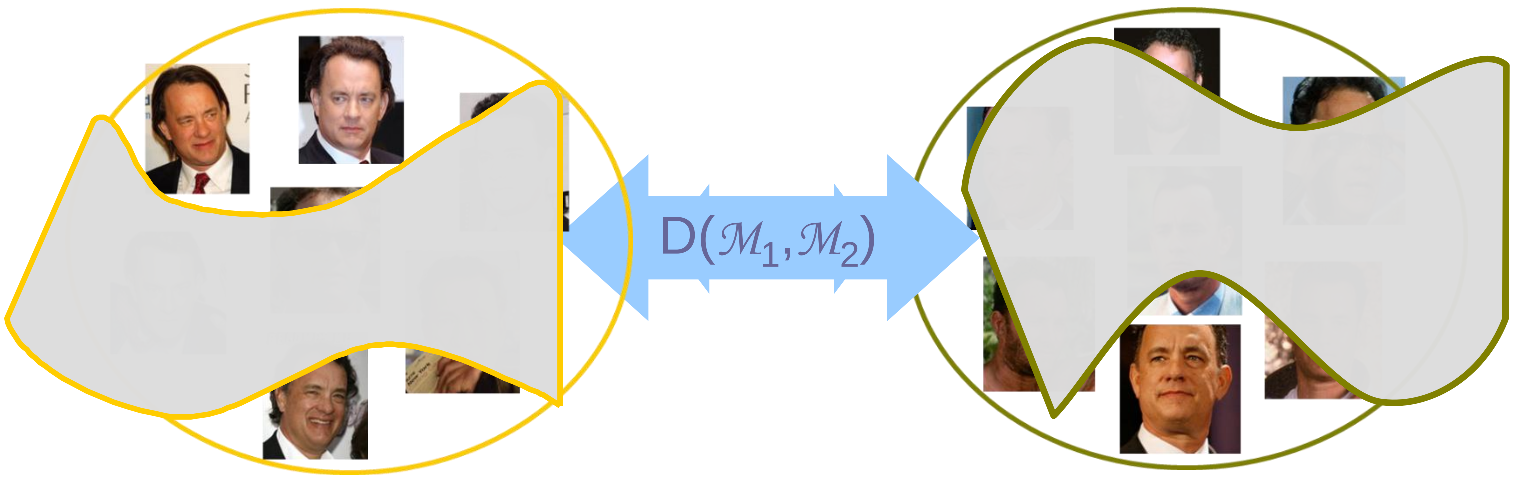
Set model II: nonlinear manifold

- Properties
  - Capture nonlinear complex appearance variation
  - Need dense sampling and large amount of data
  - Less appealing computational efficiency
- Methods
  - MMD [CVPR'08]
  - MDA [CVPR'09]
  - BoMPA [BMVC'05]
  - SANS [CVPR'13]
  - MMDML [CVPR'15]
  - …



Set model II: nonlinear manifold

- MMD (Manifold-Manifold Distance) [CVPR'08]
  - Set modeling with nonlinear appearance manifold
  - Image set classification →distance computation between two manifolds
  - Metric learning: N/A

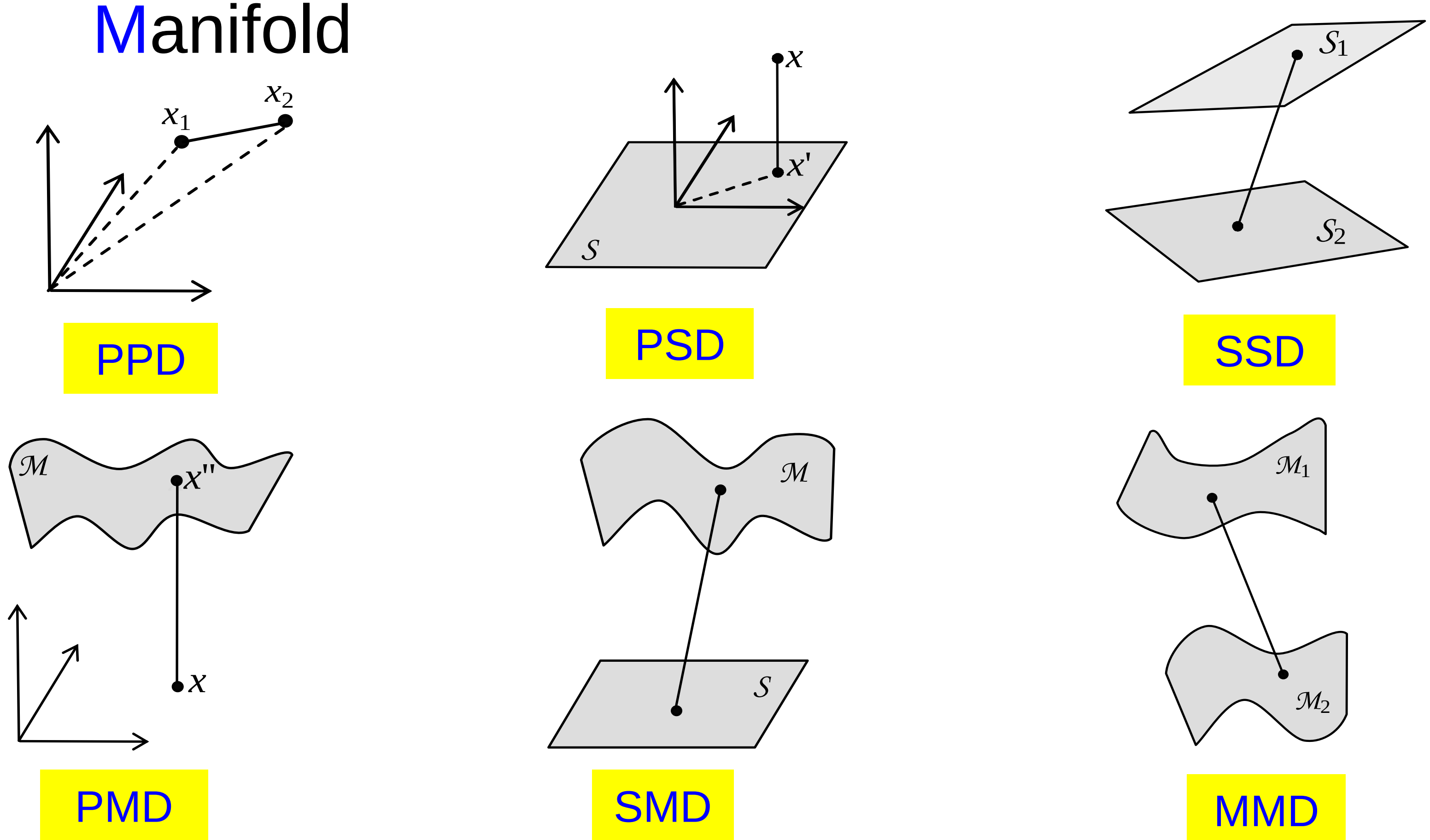


[1] R. Wang, S. Shan, X. Chen, W. Gao. Manifold-Manifold Distance with Application to Face Recognition based on Image Set. *IEEE CVPR 2008*. (Best Student Poster Award Runner-up)

[2] R. Wang, S. Shan, X. Chen, Q. Dai, W. Gao. Manifold-Manifold Distance and Its Application to Face Recognition with Image Sets. *IEEE Trans. Image Processing*, 2012.

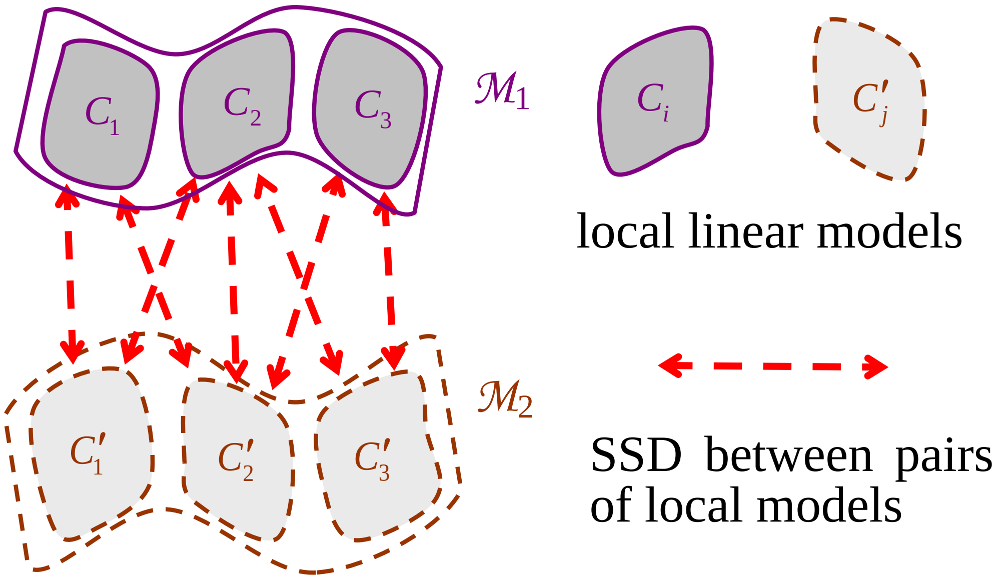
MMD

- Multi-level MMD framework
  - Three pattern levels: Point->Subspace->Manifold



MMD

Formulation & Solution



Three modules

- Local model construction
- Local model distance
- Global integration of local distances

$$d(\mathcal{M}_1, \mathcal{M}_2) = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d(C_i, C'_j)$$

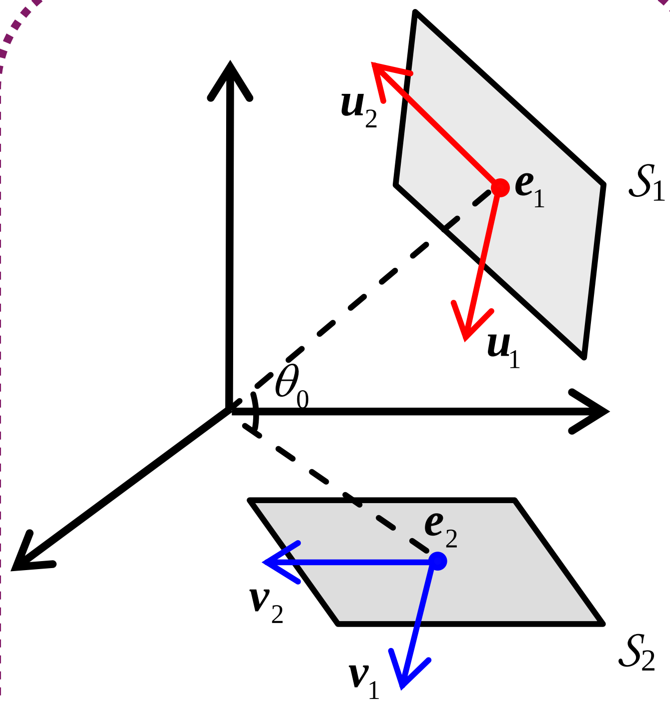
$$s.t. \quad \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = 1, \quad f_{ij} \geq 0$$

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

MMD

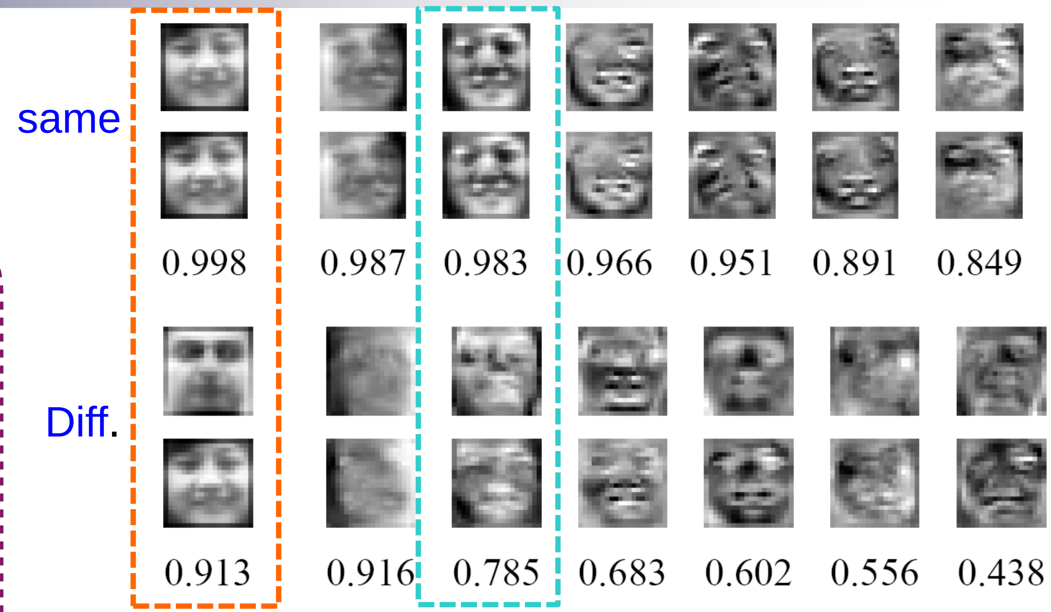
- Subspace-subspace distance (SSD)



mean (cosine correlation)

+

variance (canonical correlation)



$$d_E(S_1, S_2) = (1 - \cos^2\theta_0)^{1/2} = \sin\theta_0$$

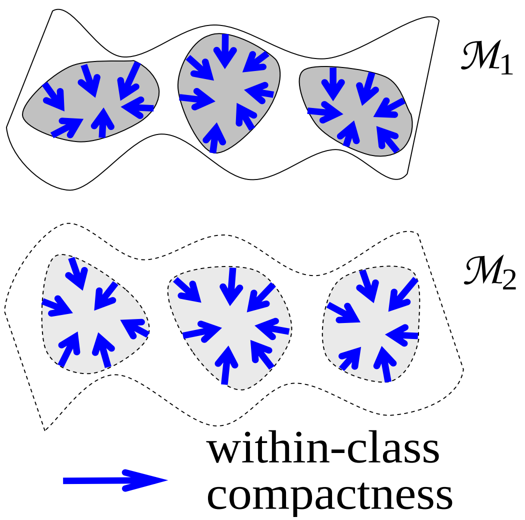
$$d_P(S_1, S_2) = (\sum_{k=1}^{r} \sin^2\theta_k)^{1/2} = (r - \sum_{k=1}^{r} \cos^2\theta_k)^{1/2}$$

$$d(S_1, S_2) = (\sin^2\theta_0 + \frac{1}{r}\sum_{k=1}^{r} \sin^2\theta_k)^{1/2}$$

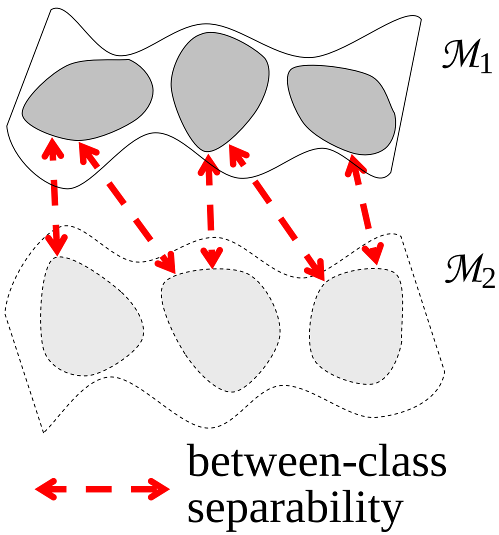
$$= (2 - \cos^2\theta_0 - \frac{1}{r}\sum_{k=1}^{r} \cos^2\theta_k)^{1/2}$$

Set model II: nonlinear manifold

- MDA (Manifold Discriminant Analysis) [CVPR'09]
  - Goal: maximize "manifold margin" under Graph Embedding framework
  - Metric learning: in Euclidean space
    - Euclidean distance between pair of image samples



Intrinsic graph



Penalty graph

[1] R. Wang, X. Chen. Manifold Discriminant Analysis. *IEEE CVPR 2009*.

MDA

- Optimization framework
  - Min. within-class scatter

Intrinsic graph

$$S_w = \sum_{m,n} \left\| \mathbf{v}^T \mathbf{x}_m - \mathbf{v}^T \mathbf{x}_n \right\|^2 w_{m,n} = 2\mathbf{v}^T X(D - W) X^T \mathbf{v}$$

  - Max. between-class scatter

Penalty graph

$$S_b = \sum_{m,n} \left\| \mathbf{v}^T \mathbf{x}_m - \mathbf{v}^T \mathbf{x}_n \right\|^2 w'_{m,n} = 2\mathbf{v}^T X(D' - W') X^T \mathbf{v}$$

  - Objective function
    - Global linear transformation

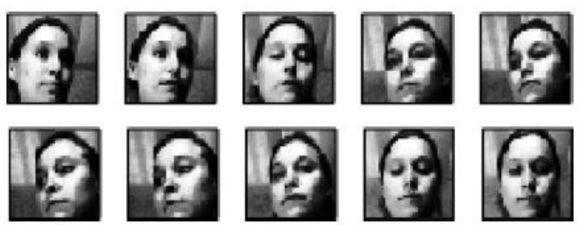
$$\text{Maximize } J(\mathbf{v}) = \frac{|S_b|}{|S_w|} = \frac{\mathbf{v}^T X(D' - W') X^T \mathbf{v}}{\mathbf{v}^T X(D - W) X^T \mathbf{v}}$$

    - Solved by generalized eigen-decomposition

Set model II: nonlinear manifold

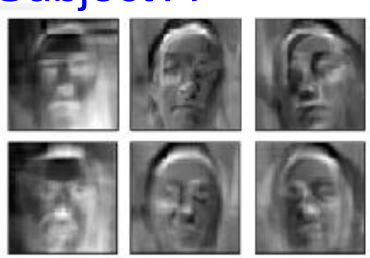
- BoMPA (Boosted Manifold Principal Angles) [BMVC'05]
  - Goal: optimal fusion of different principal angles
  - Exploit Adaboost to learn weights for each angle

$$\mathcal{M}(\Theta) = \mathrm{sign}\left[\sum_{i=1}^{N} w_i \mathcal{M}(\theta_i) - \frac{1}{2}\sum_{i=1}^{N} w_i\right]$$



Subject *A*

Subject *B*



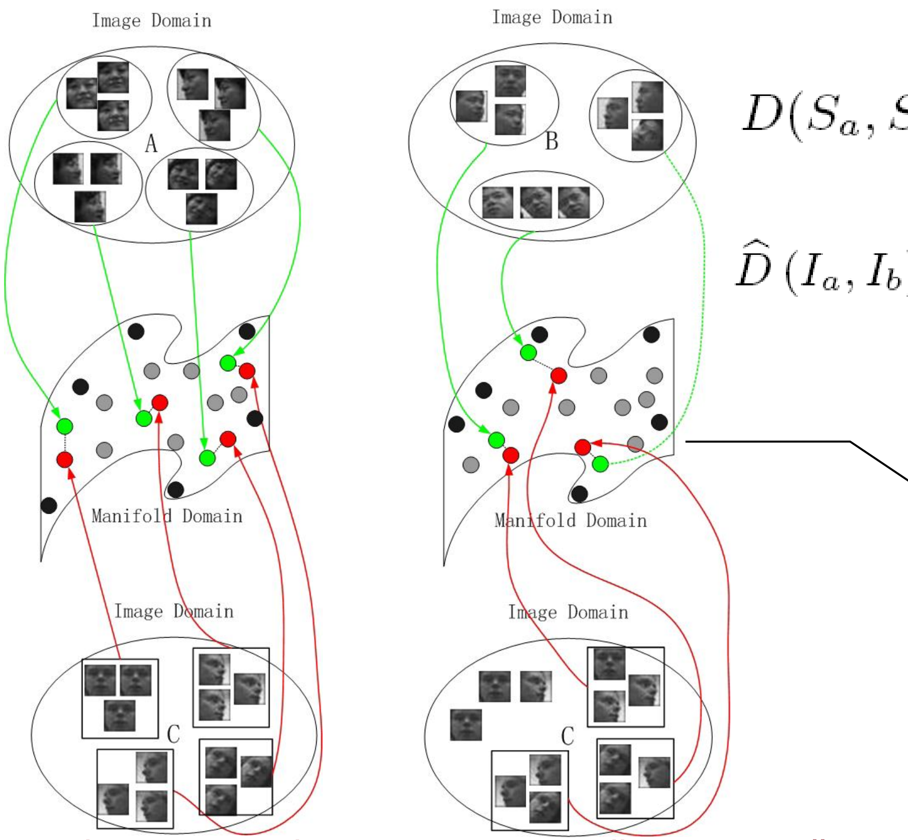
PA→similar (common illumination)

BoMPA→dissimilar

[1] T. Kim O. Arandjelovic, R. Cipolla. Learning over Sets using Boosted Manifold Principal Angles (BoMPA). *BMVC 2005*.

Set model II: nonlinear manifold

- SANS (Sparse Approximated Nearest Subspaces) [CVPR'13]
  - Goal: adaptively construct the nearest subspace pair
  - Metric learning: N/A



$$D(S_a, S_b) = ||\boldsymbol{U}_a - \boldsymbol{U}_b \boldsymbol{U}_b' \boldsymbol{U}_a||_F^2$$

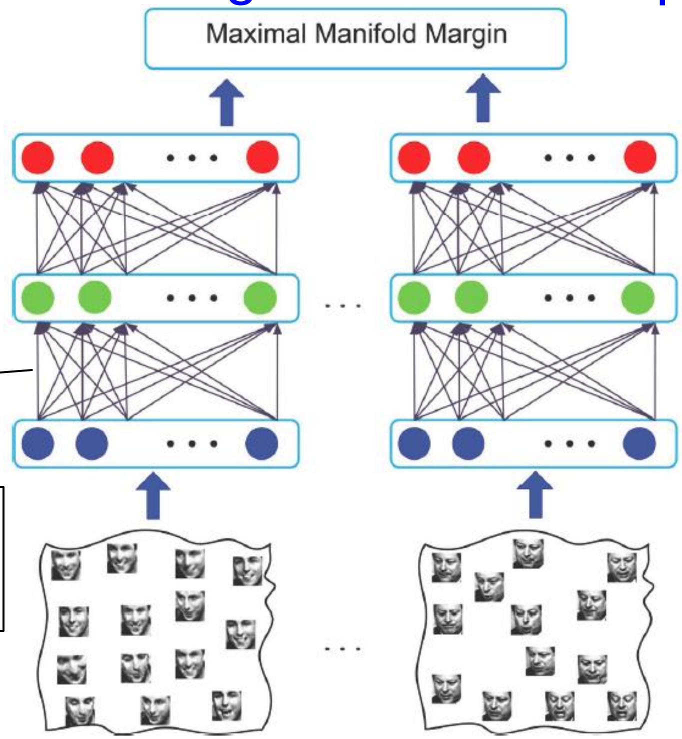
$$\widehat{D}\left(I_a, I_b\right) = \frac{1}{N_c} \sum_{k=1}^{N_c} D\left(s_k^a, s_k^b\right), \quad k \in [1, N_c]$$

◆**SSD** (Subspace-Subspace Distance): Joint sparse representation (JSR) is applied to approximate the nearest subspace over a Grassmann manifold.

[1] S. Chen, C. Sanderson, M.T. Harandi, B.C. Lovell. Improved Image Set Classification via Joint Sparse Approximated Nearest Subspaces. *IEEE CVPR 2013*.

Set model II: nonlinear manifold

- MMDML (Multi-Manifold Deep Metric Learning) [CVPR'15]
  - Goal: maximize "manifold margin" under Deep Learning framework
  - Metric learning: in Euclidean space



◆Class-specific DNN
◆Model nonlinearity

Objective function

$$D_1(h_{ci}^L) = \frac{1}{K_1}\sum_{p=1}^{K_1}\|h_{ci}^L - h_{cip}^L\|_2^2$$

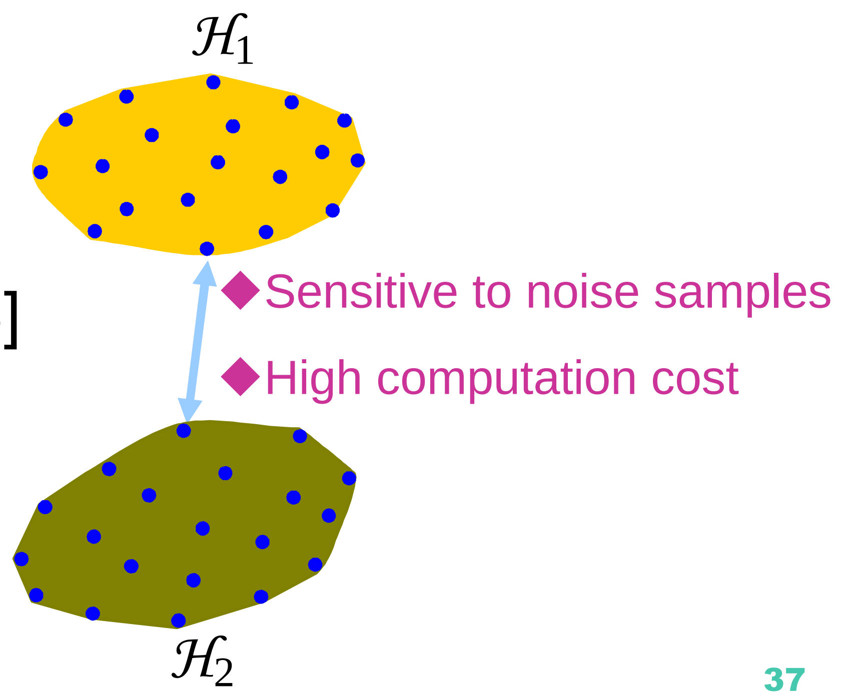
$$D_2(h_{ci}^L) = \frac{1}{K_2}\sum_{q=1}^{K_2}\|h_{ci}^L - h_{ciq}^L\|_2^2$$

$$\min_{f_c}\sum_{i=1}^{N_c}\left(D_1(h_{ci}^L) - D_2(h_{ci}^L)\right)$$

[1] J. Lu, G. Wang, W. Deng, P. Moulin, and J. Zhou. Multi-Manifold Deep Metric Learning for Image Set Classification. *IEEE CVPR 2015*.

Set model III: affine subspace

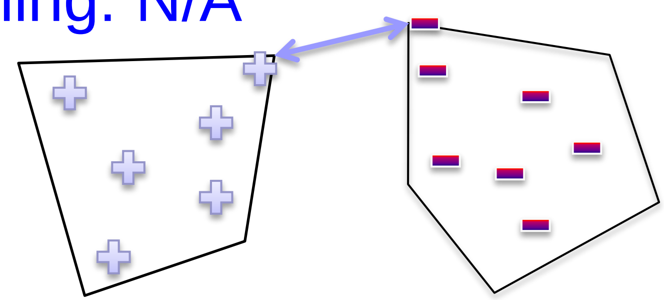
- Properties
  - Linear reconstruction using: mean + subspace basis
  - Synthesized virtual NN-pair matching
  - Less characterization of global data structure
  - Computationally expensive by NN-based matching
- Methods
  - AHISD/CHISD [CVPR'10]
  - SANP [CVPR'11]
  - PSDML/SSDML [ICCV'13]
  - …



中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Set model III: affine subspace

- AHISD/CHISD [CVPR'10], SANP [CVPR'11]
  - NN-based matching using sample Euclidean distance
  - Metric learning: N/A



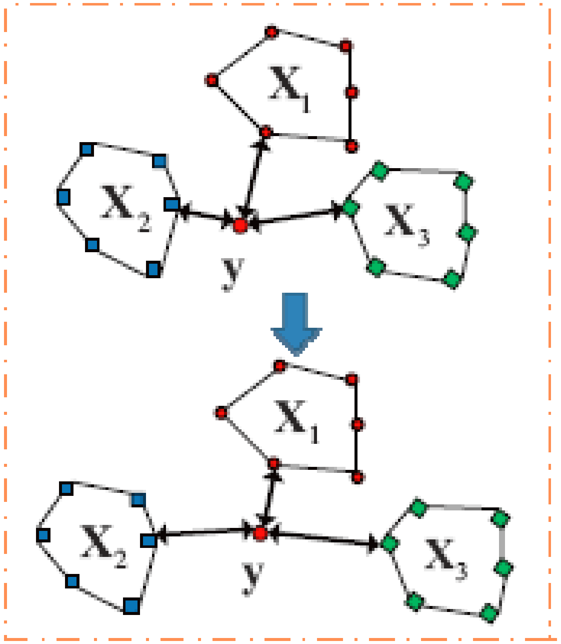
  - Subspace spanned by all the available samples $\boldsymbol{D} = \{d_1, \ldots, d_n\}$ in the set
    - Affine hull
      - $H(\boldsymbol{D}) = \{\boldsymbol{D\alpha} = \sum d_i \alpha_i \mid \sum \alpha_i = 1\}$
    - Convex hull
      - $H(\boldsymbol{D}) = \{\boldsymbol{D\alpha} = \sum d_i \alpha_i \mid \sum \alpha_i = 1, 0 \le \alpha_i \le 1\}$

[1] H. Cevikalp, B. Triggs. Face Recognition Based on Image Sets. *IEEE CVPR 2010.*

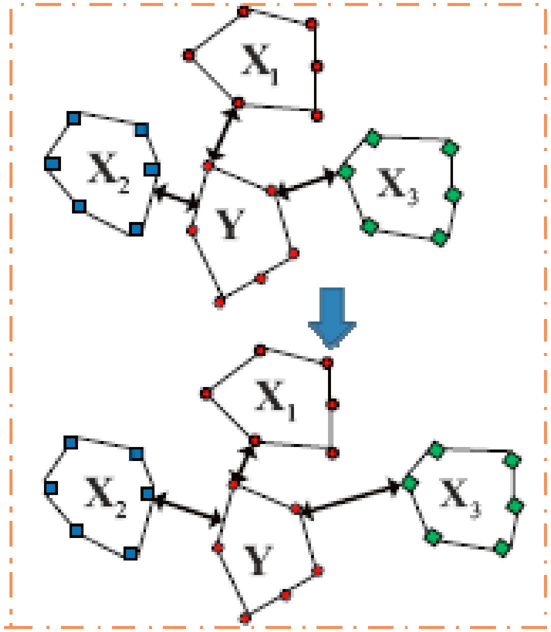
[2] Y. Hu, A.S. Mian, R. Owens. Sparse Approximated Nearest Points for Image Set Classification. *IEEE CVPR 2011.*

Set model III: affine subspace

- PSDML/SSDML [ICCV'13]
  - Metric learning: in Euclidean space



point-to-set distance metric learning (PSDML)



set-to-set distance metric learning (SSDML)

[1] P. Zhu, L. Zhang, W. Zuo, and D. Zhang. From Point to Set: Extend the Learning of Distance Metrics. *IEEE ICCV 2013.*

Institute of Computing Technology, Chinese Academy of Sciences
中科院计算所

PSDML/SSDML

- Point-to-set distance
  - Basic distance
    - $d^2(\boldsymbol{x}, \boldsymbol{D}) = min_{\boldsymbol{\alpha}} \|\boldsymbol{x} - H(\boldsymbol{D})\|_2^2 = min_{\boldsymbol{\alpha}} \|\boldsymbol{x} - \boldsymbol{D\alpha}\|_2^2$
      - Solution: Least Square Regression or Ridge Regression
  - Mahalanobis distance
    - $d_{\boldsymbol{M}}^2(\boldsymbol{x}, \boldsymbol{D}) = min \|\boldsymbol{P}(\boldsymbol{x} - \boldsymbol{D\alpha})\|_2^2 = (\boldsymbol{x} - \boldsymbol{D\alpha})^T \boldsymbol{P}^T \boldsymbol{P}(x - \boldsymbol{D\alpha}) = (\boldsymbol{x} - \boldsymbol{D\alpha})^T \boldsymbol{M}(\boldsymbol{x} - \boldsymbol{D\alpha})$

Projection matrix

PSDML/SSDML

- Point-to-set distance metric learning (PSDML)
  - SVM-based method

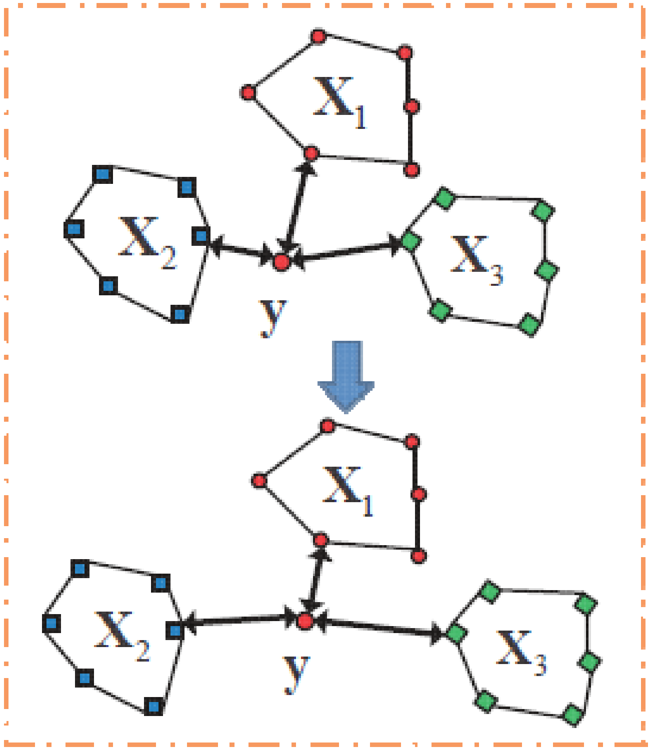
$$\min_{\boldsymbol{M},\boldsymbol{\alpha}_{l(\boldsymbol{x}_i)},\boldsymbol{\alpha}_j,\xi_{ij}^N,\xi_i^P,b} \|\boldsymbol{M}\|_F^2 + \nu\Big(\sum_{i,j}\xi_{i,j}^N + \sum_i \xi_i^P\Big)$$

$$s.t.\ d_M\big(\boldsymbol{x}_i,\boldsymbol{D}_j\big) + b \geq 1 - \xi_{ij}^N, j \neq l(\boldsymbol{x}_i);\ \text{(-)}$$

$$d_M\big(\boldsymbol{x}_i,\boldsymbol{D}_{l(\boldsymbol{x}_i)}\big) + b \leq -1 + \xi_i^P;\ \text{(+)}$$

$$\boldsymbol{M} \succcurlyeq 0, \forall i,j, \xi_{ij}^N \geq 0, \xi_i^P \geq 0$$

$$\begin{aligned} d_{\boldsymbol{M}}^2(\boldsymbol{x},\boldsymbol{D}) &= \min\|\boldsymbol{P}(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{\alpha})\|_2^2 \\ &= (\boldsymbol{x}-\boldsymbol{D}\boldsymbol{\alpha})^T\boldsymbol{P}^T\boldsymbol{P}(x-\boldsymbol{D}\boldsymbol{\alpha}) \\ &= (\boldsymbol{x}-\boldsymbol{D}\boldsymbol{\alpha})^T\boldsymbol{M}(\boldsymbol{x}-\boldsymbol{D}\boldsymbol{\alpha}) \end{aligned}$$



PSDML/SSDML

- Set-to-set distance
  - Basic distance
    - $d^2(\boldsymbol{D}_1, \boldsymbol{D}_2) = min_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} \|H(\boldsymbol{D}_1) - H(\boldsymbol{D}_1)\|_2^2 = min_{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2} \|\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2\|_2^2$
      - Solution: **AHISD/CHISD [Cevikalp, CVPR’10]**
  - Mahalanobis distance
    - $d_{\boldsymbol{M}}^2(\boldsymbol{D}_1, \boldsymbol{D}_2) = min\|\boldsymbol{P}(\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2)\|_2^2 = (\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2)^T\boldsymbol{M}(\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2)$

PSDML/SSDML

- Set-to-set distance metric learning (SSDML)
  - SVM-based method

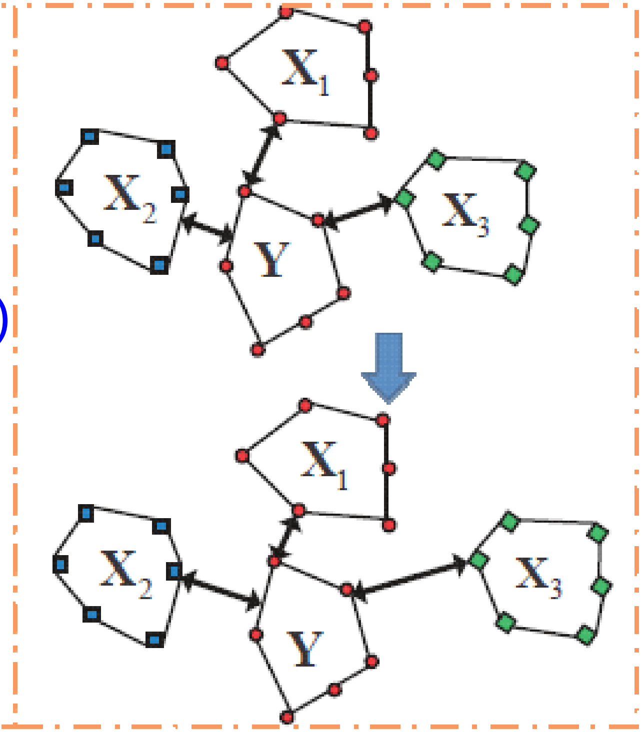
$$\min_{\boldsymbol{M},\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j,\boldsymbol{\alpha}_k,\xi_{ij}^N,\xi_{ik}^P,b} \|\boldsymbol{M}\|_F^2 + \nu\left(\sum_{i,j} \xi_{ij}^N + \sum_{i,k} \xi_{ik}^P\right)$$

$$s.t.\ d_M\left(\boldsymbol{D}_i, \boldsymbol{D}_j\right) + b \geq 1 - \xi_{ij}^N, l(\boldsymbol{D}_i) \neq l\left(\boldsymbol{D}_j\right); \text{(-)}$$

$$d_M(\boldsymbol{D}_i, \boldsymbol{D}_k) + b \leq -1 + \xi_{ik}^P, l(\boldsymbol{D}_i) = l(\boldsymbol{D}_k); \text{(+)}$$

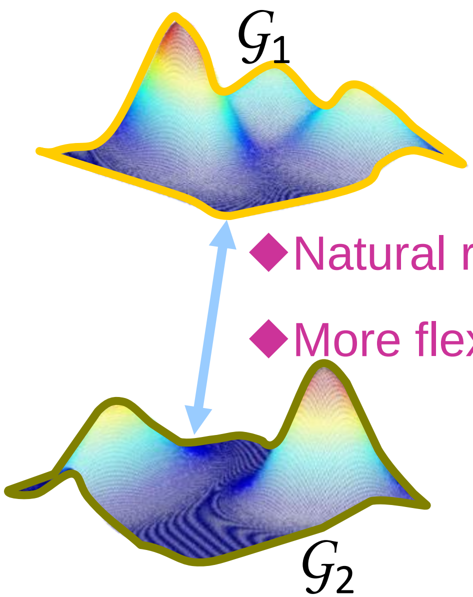
$$\boldsymbol{M} \succcurlyeq 0, \forall i, j, k, \xi_{ij}^N \geq 0, \xi_{ik}^P \geq 0$$

$$d_{\boldsymbol{M}}^2(\boldsymbol{D}_1, \boldsymbol{D}_2) = \min\|\boldsymbol{P}(\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2)\|_2^2$$
$$= (\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2)^T\boldsymbol{M}(\boldsymbol{D}_1\boldsymbol{\alpha}_1 - \boldsymbol{D}_2\boldsymbol{\alpha}_2)$$



Set model IV: statistics (COV+)

- Properties
  - The natural raw statistics of a sample set
  - Flexible model of multiple-order statistical information
- Methods
  - CDL [CVPR'12]
  - LMKML [ICCV'13]
  - DARG [CVPR'15]
  - SPD-ML [ECCV'14]
  - LEML [ICML'15]
  - LERM [CVPR'14]
  - HER [CVPR'15]
  - …

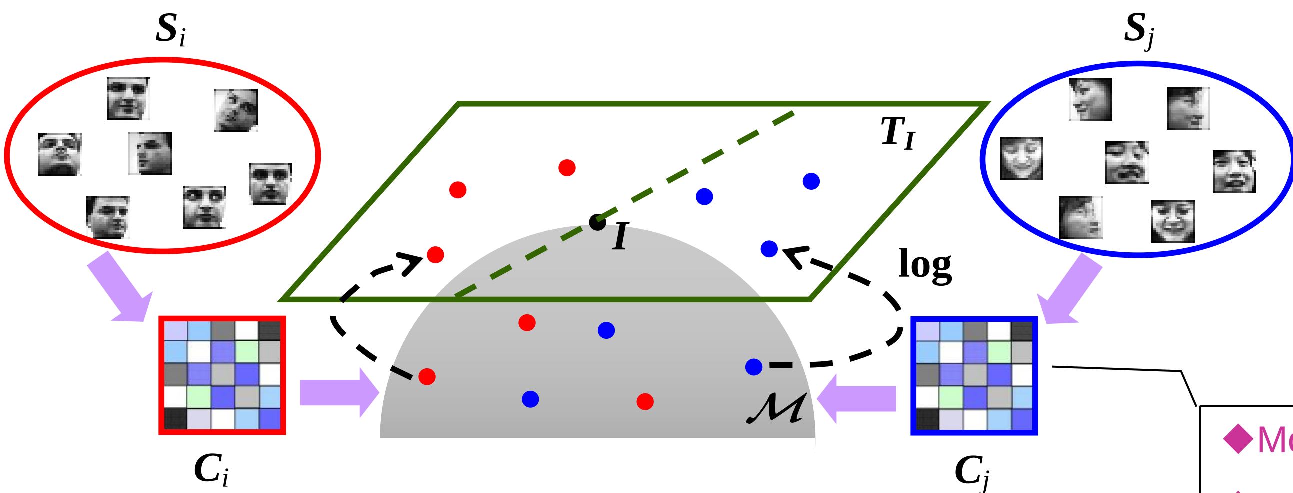


[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Set model IV: statistics (COV+)

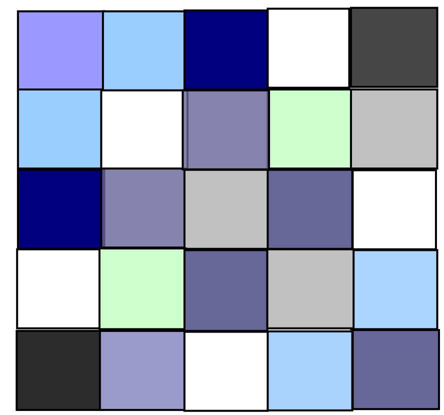
- CDL (Covariance Discriminative Learning) [CVPR'12]
  - Set modeling by Covariance Matrix (COV)
    - The 2nd order statistics characterizing set data variations
    - Robust to noisy set data, scalable to varying set size
  - Metric learning: on the SPD manifold



◆Model data variations
◆No assum. of data distr.

[1] R. Wang, H. Guo, L.S. Davis, Q. Dai. Covariance Discriminative Learning: A Natural and Efficient Approach to Image Set Classification. *IEEE CVPR 2012*.

Set modeling by Covariance Matrix



◆Image set: *N* samples with *d*-dimension image feature

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]_{d \times N}$$

◆COV: *d*d* symmetric positive definite (SPD) matrix*

$$\boldsymbol{C} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T$$

*: *use regularization to tackle singularity problem*

CDL

- Set matching on COV manifold
  - Riemannian metrics on the SPD manifold
    - Affine-invariant distance (AID) [1]

$$d^2(C_1, C_2) = \sum_{i=1}^{d} \ln^2 \lambda_i(C_1, C_2)$$

or

$$d^2(C_1, C_2) = \left\| \log_I (C_1^{-1/2} C_2 C_1^{-1/2}) \right\|_F^2$$

    - Log-Euclidean distance (LED) [2]

$$d(C_1, C_2) = \left\| \log_I (C_1) - \log_I (C_2) \right\|_F$$

High computational burden

More efficient, more appealing

[1] W. Förstner and B. Moonen. A Metric for Covariance Matrices. *Technical Report* 1999.

[2] V. Arsigny, P. Fillard, X. Pennec and N. Ayache. Geometric Means In A Novel Vector Space Structure On Symmetric Positive-Definite Matrices. *SIAM J. MATRIX ANAL. APPL*. Vol. 29, No. 1, pp. 328-347, 2007.

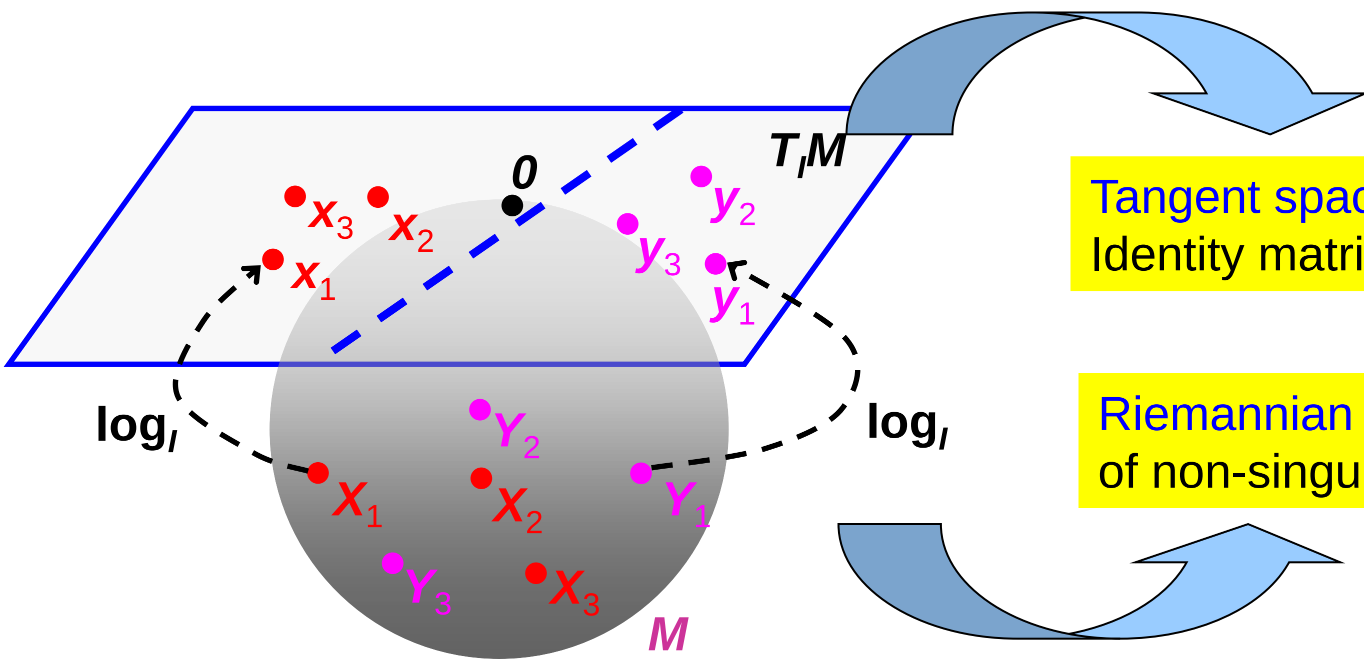
CDL

- Set matching on COV manifold (cont.)
  - Explicit Riemannian kernel feature mapping with LED

$$\Psi_{log}: \quad C \to \log_I(C), \quad (\mathcal{M} \mapsto R^{d\times d})$$

$$k_{log}(C_1, C_2) = trace[\log_I(C_1)\cdot \log_I(C_2)]$$

Mercer's theorem

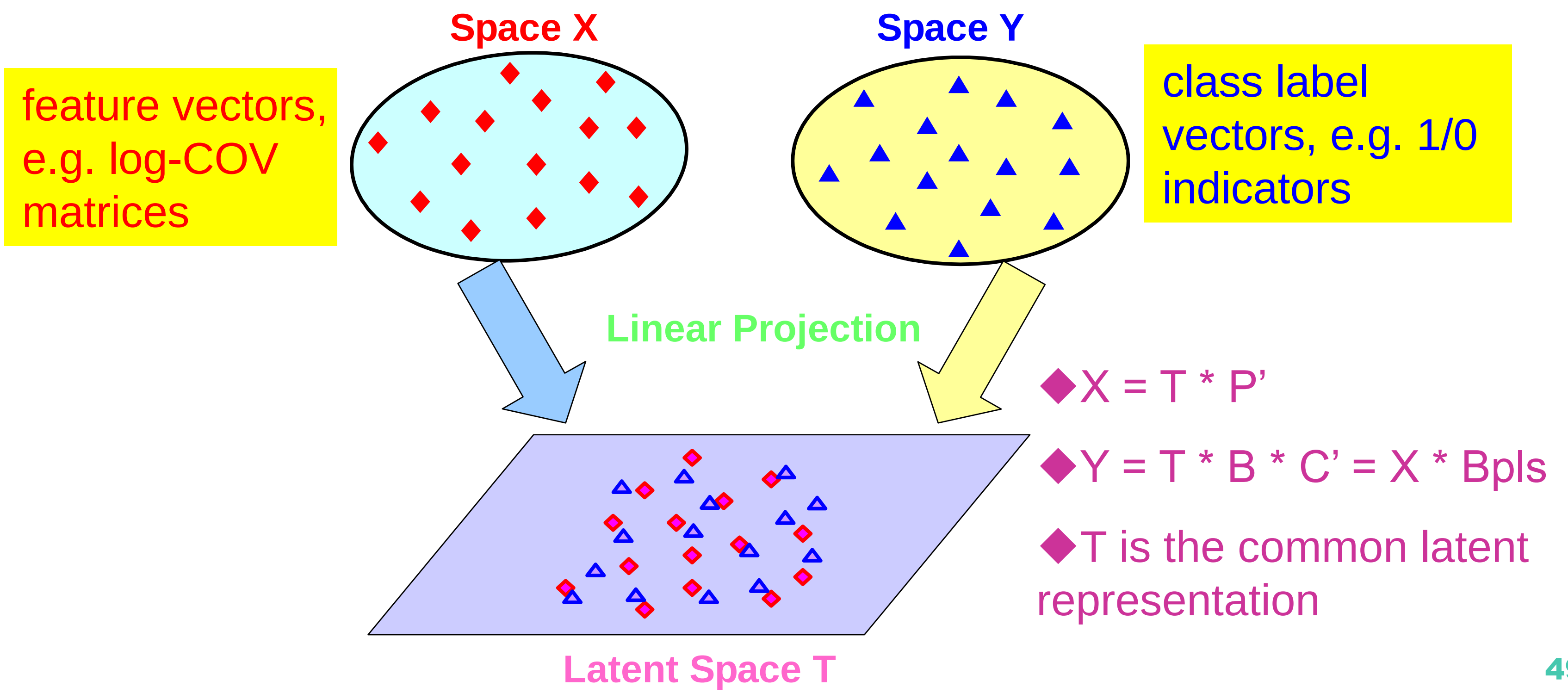


Tangent space at Identity matrix ***I***

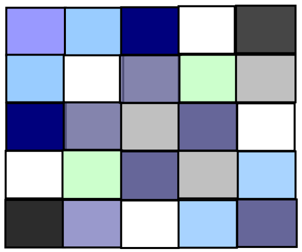
Riemannian manifold of non-singular COV

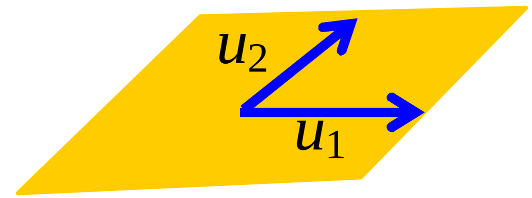
CDL

- Discriminative learning on COV manifold
  - Partial Least Squares (PLS) regression
  - Goal: Maximize the covariance between observations and class labels



CDL vs. GDA





- COV→SPD manifold
  - Model

$$C = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T$$

  - Metric

$$d(C_1, C_2) = \left\| \log_I(C_1) - \log_I(C_2) \right\|_F$$

$$\Rightarrow \log_I(C) = U \log_I(\Lambda) U^T$$

  - Kernel

$$\Psi_{log} : \ C \rightarrow \log_I(C), \ (\mathcal{M} \mapsto R^{d\times d})$$

- Subspace→Grassmannian
  - Model

$$C = U \Lambda U^T$$

$$\Rightarrow U = [u_1, u_2, \ldots, u_m]_{D\times m}{*}$$

  - Metric

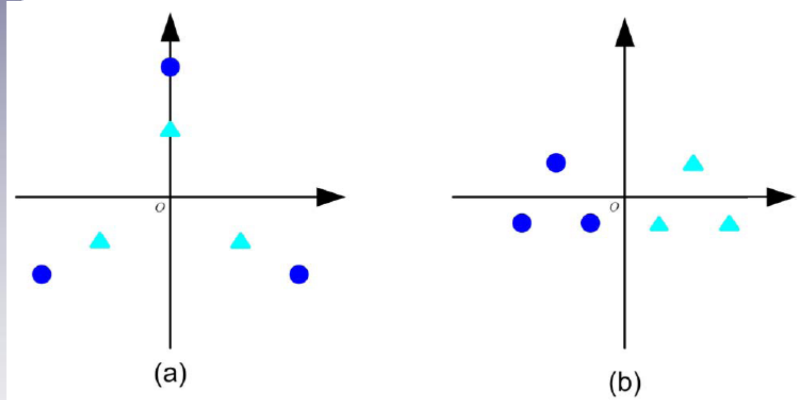
$$d_{proj}(U_1, U_2) = 2^{-1/2} \left\| U_1 U_1^T - U_2 U_2^T \right\|_F$$

  - Kernel

$$\Psi_{proj} : \ U \rightarrow UU^T, \ \mathcal{G}(m, D) \rightarrow R^{d\times d}$$

Set model IV: statistics (COV+)

- LMKML (Localized Multi-Kernel Metric Learning) [ICCV'13]
  - Exploring multiple order statistics
    - Data-adaptive weights for different types of features
    - Ignoring the geometric structure of 2nd/3rd-order statistics
  - Metric learning: in Euclidean space



Complementary information
(mean vs. covariance)

1st / 2nd / 3rd-order statistics

$$m = \frac{1}{n}\sum_{i-1}^{n} x_i \qquad C = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - m)(x_j - m)^T$$

$$\mathcal{T} = C \otimes m$$

Objective function

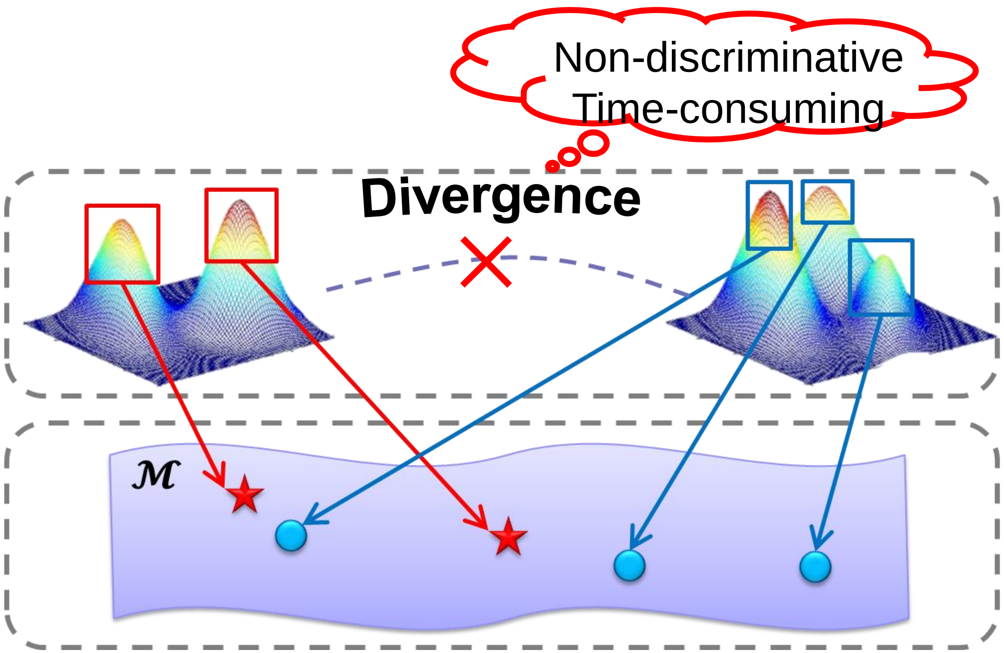
$$d(S_i, S_j) = \sum_{p=1}^{P} \eta_p(\phi_i^p)(\phi_i^p - \phi_j^p)^T M(\phi_i^p - \phi_j^p)\eta_p(\phi_j^p)$$

$$\max_M J = \sum_{\substack{i,j=1 \\ (S_i,S_j)\in C^-}}^{N} \frac{d(S_i, S_j)}{N_{C-}} - \sum_{\substack{i,j=1 \\ (S_i,S_j)\in C^+}}^{N} \frac{d(S_i, S_j)}{N_{C+}}$$

[1] J. Lu, G. Wang, and P. Moulin. Image Set Classification Using Holistic Multiple Order Statistics Features and Localized Multi-Kernel Metric Learning. *IEEE ICCV 2013*.

Set model IV: statistics (COV+)

- DARG (Discriminant Analysis on Riemannian manifold of Gaussian distributions) [CVPR'15]
  - Set modeling by mixture of Gaussian distribution (GMM)
    - Naturally encode the 1st order and 2nd order statistics
  - Metric learning: on Riemannian manifold



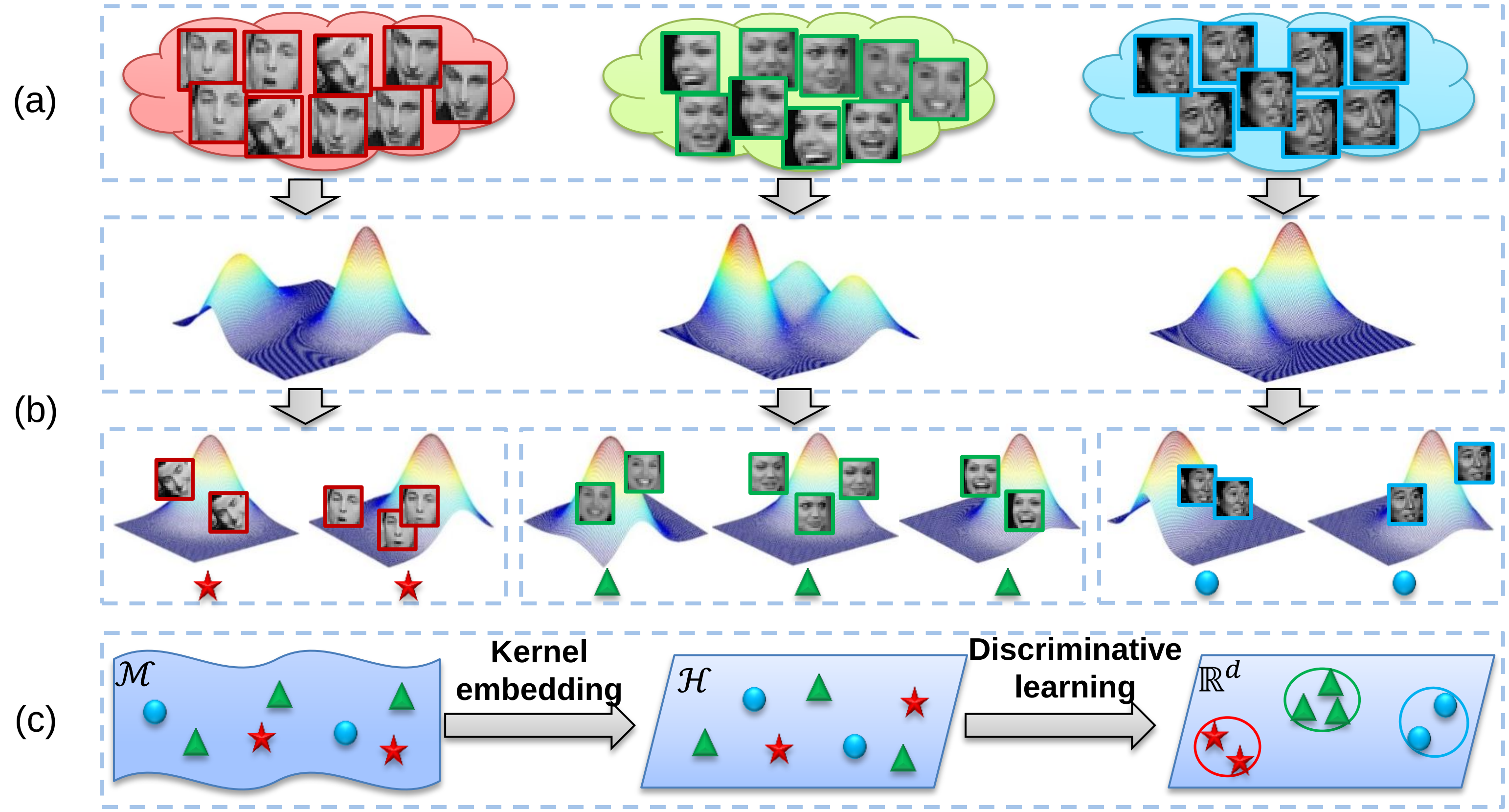
[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]

$\mathcal{M}$: Riemannian manifold of Gaussian distributions

[1] W. Wang, R. Wang, Z. Huang, S. Shan, X. Chen. Discriminant Analysis on Riemannian Manifold of Gaussian Distributions for Face Recognition with Image Sets. *IEEE CVPR 2015*.

DARG

■ Framework



$\mathcal{M}$: Riemannian manifold of Gaussian distributions
$\mathcal{H}$: high-dimensional reproducing kernel Hilbert space (RKHS)
$\mathbb{R}^d$: target lower-dimensional discriminant Euclidean subspace

- Kernels on the Gaussian distribution manifold
  - kernel based on Lie Group
    - Distance based on Lie Group (LGD)

$$LGD(P_i, P_j) = \left\| \log(P_i) - \log(P_j) \right\|_F ,$$

SPD matrix according to information geometry

$$g \sim N(x|\mu, \Sigma) \longmapsto P = |\Sigma|^{-\frac{1}{d+1}} \begin{pmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{pmatrix}$$

    - Kernel function

$$K_{\mathrm{LGD}}(g_i, g_j) = \exp\left(-\frac{LGD^2(P_i, P_j)}{2t^2}\right)$$

DARG

- Kernels on the Gaussian distribution manifold
  - kernel based on Lie Group
  - kernel based on MD and LED
    - Mahalanobis Distance (MD) between mean

$$MD(\mu_i,\mu_j)=\sqrt{(\mu_i-\mu_j)^T(\Sigma_i^{-1}+\Sigma_j^{-1})(\mu_i-\mu_j)}$$

    - LED between covariance matrix

$$LED(\Sigma_i,\Sigma_j)=\|\log(\Sigma_i)-\log(\Sigma_j)\|_F$$

    - Kernel function

$$K_{MD+LED}(g_i,g_j)=\gamma_1 K_{MD}(\mu_i,\mu_j)+\gamma_2 K_{LED}(\Sigma_i,\Sigma_j)$$

$$K_{MD}(\mu_i,\mu_j)=\exp\left(-\frac{\mathrm{MD}^2(\mu_i,\mu_j)}{2t^2}\right)$$

$$K_{LED}(\Sigma_i,\Sigma_j)=\exp\left(-\frac{LED^2(\Sigma_i,\Sigma_j)}{2t^2}\right)$$

γ_1,γ_2 are the combination coefficients

DARG

- Discriminative learning
  - Weighted KDA (kernel discriminant analysis)
    - incorporating the weights of Gaussian components

$$J(\alpha) = \frac{|\alpha^T \boldsymbol{B} \alpha|}{|\alpha^T \boldsymbol{W} \alpha|}$$

$$\boldsymbol{W} = \sum_{i=1}^{C} \frac{1}{\omega_i} \sum_{j=1}^{N_i} w_j^{(i)} (k_j^i - m_i)(k_j^i - m_i)^T$$

$$\boldsymbol{B} = \sum_{i=1}^{C} N_i (m_i - m)(m_i - m)^T$$

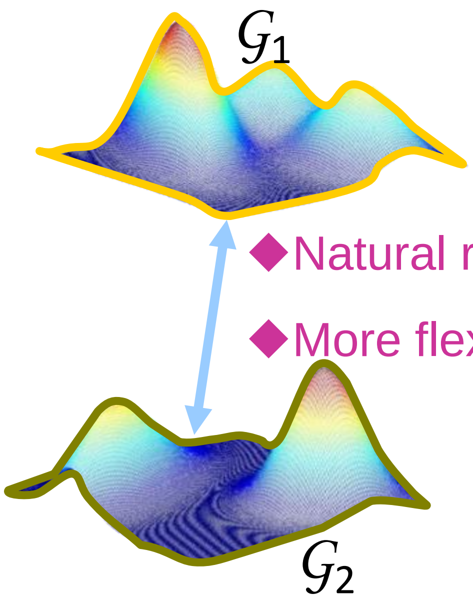
$$m_i = \frac{1}{N_i \omega_i} \sum_{j=1}^{N_i} w_j^i k_j^i \,, \mathrm{m} = \frac{1}{N_i} \sum_{i=1}^{C} \frac{1}{\omega_i} \sum_{j=1}^{N_i} w_j^i k_j^i$$

Institute of Computing Technology, Chinese Academy of Sciences

中科院计算所

Set model IV: statistics (COV+)

- Properties
  - The natural raw statistics of a sample set
  - Flexible model of multiple-order statistical information
- Methods
  - CDL [CVPR'12]
  - LMKML [ICCV'13]
  - DARG [CVPR'15]
  - SPD-ML [ECCV'14]
  - LEML [ICML'15]
  - LERM [CVPR'14]
  - HER [CVPR'15]
  - …



◆Natural raw statistics

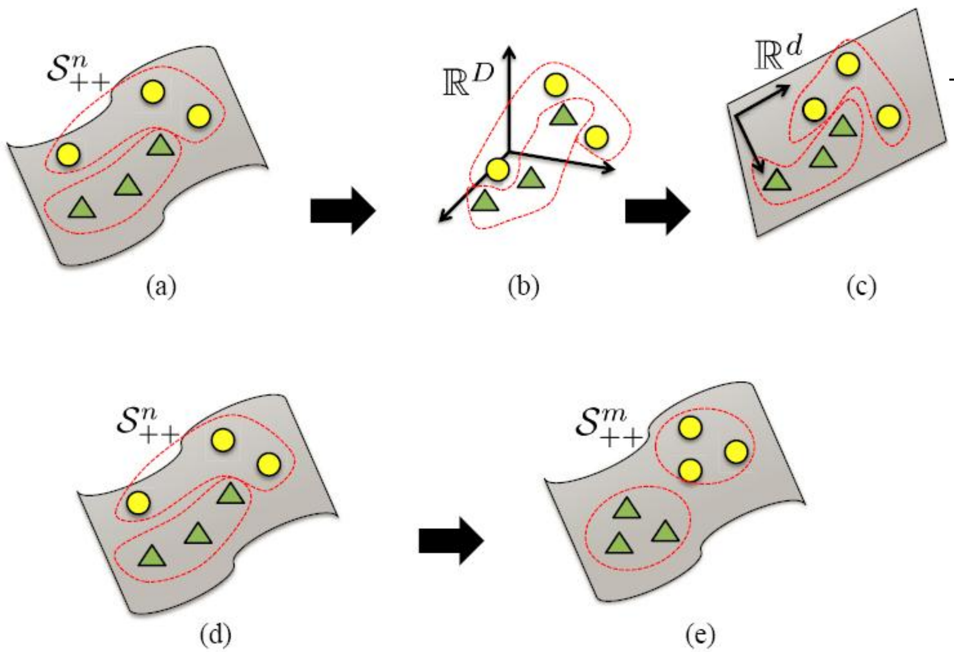
◆More flexible model

[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Set model IV: statistics (COV+)

- SPD-ML (SPD Manifold Learning) [ECCV'14]
  - Pioneering work on explicit manifold-to-manifold dimensionality reduction
  - Metric learning: on Riemannian manifold

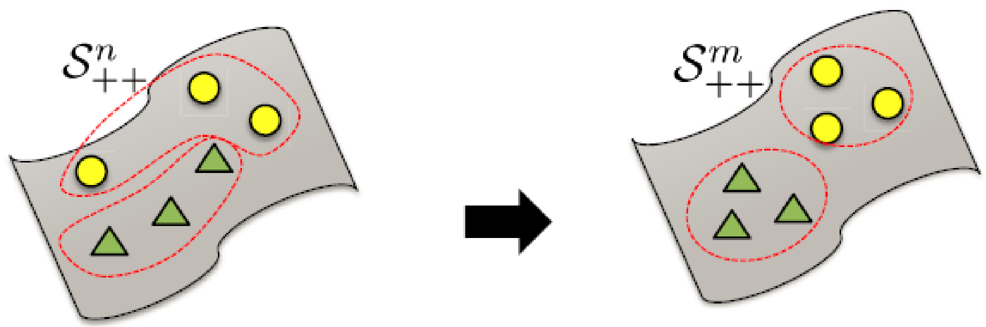


◆Structure distortion
◆No explicit mapping

[1] M. Harandi, M. Salzmann, R. Hartley. From Manifold to Manifold: Geometry-Aware Dimensionality Reduction for SPD Matrices. *ECCV 2014.*

SPD-ML

- SPD manifold dimensionality reduction
  - Mapping function: $f: \mathcal{S}_{++}^{n} \times \mathbb{R}^{n\times m} \rightarrow \mathcal{S}_{++}^{m}$
    - $f\left(\boldsymbol{X}, \widetilde{\boldsymbol{W}}\right) = \widetilde{\boldsymbol{W}}^{T}\boldsymbol{X}\widetilde{\boldsymbol{W}} \in \mathcal{S}_{++}^{m} \succ 0,\ \boldsymbol{X} \in \mathcal{S}_{++}^{n},\ \widetilde{\boldsymbol{W}} \in \mathbb{R}^{n\times m}$ (full rank)



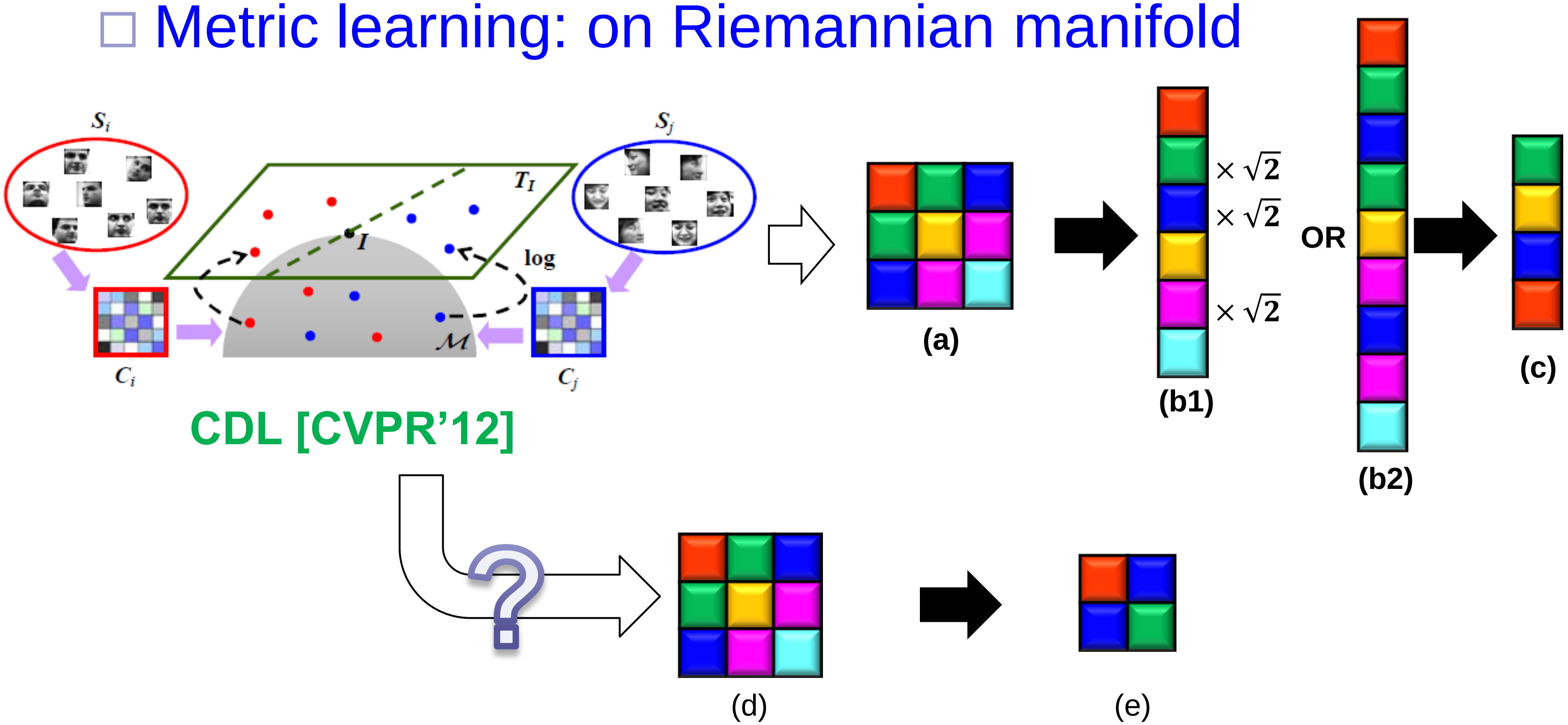
- Affine invariant metrics: AIRM / Stein divergence on target SPD manifold $\mathcal{S}_{++}^{m}$
  - $\delta^{2}\left(\widetilde{\boldsymbol{W}}^{T}\boldsymbol{X}_{i}\widetilde{\boldsymbol{W}}, \widetilde{\boldsymbol{W}}^{T}\boldsymbol{X}_{i}\widetilde{\boldsymbol{W}}\right) = \delta^{2}\left(\boldsymbol{W}^{T}\boldsymbol{X}_{i}\boldsymbol{W}, \boldsymbol{W}^{T}\boldsymbol{X}_{j}\boldsymbol{W}\right)$
    - $\widetilde{\boldsymbol{W}} = \boldsymbol{M}\boldsymbol{W},\ \boldsymbol{M} \in GL(n),\ \boldsymbol{W} \in \mathbb{R}^{n\times m},\ \boldsymbol{W}^{T}\boldsymbol{W} = \boldsymbol{I}_{m}$

SPD-ML

- Discriminative learning
  - Discriminant function
    - Graph Embedding formalism with an affinity matrix that encodes intra-class and inter-class SPD distances
    - $\min L(\boldsymbol{W}) = \min \sum_{ij} \boldsymbol{A}_{ij} \delta^2(\boldsymbol{W}^T \boldsymbol{X}_i \boldsymbol{W}, \boldsymbol{W}^T \boldsymbol{X}_j \boldsymbol{W})$
      - $s.t. \boldsymbol{W}^T \boldsymbol{W} = \boldsymbol{I}_m$ (orthogonality constraint)
  - Optimization
    - Optimization problems on Stiefel manifold, solved by nonlinear Conjugate Gradient (CG) method

Set model IV: statistics (COV+)

- LEML (Log-Euclidean Metric Learning) [ICML'15]
  - Learning tangent map by preserving matrix symmetric structure
  - Metric learning: on Riemannian manifold



[1] Z. Huang, R. Wang, S. Shan, X. Li, X. Chen. Log-Euclidean Metric Learning on Symmetric Positive Definite Manifold with Application to Image Set Classification. *ICML 2015*.

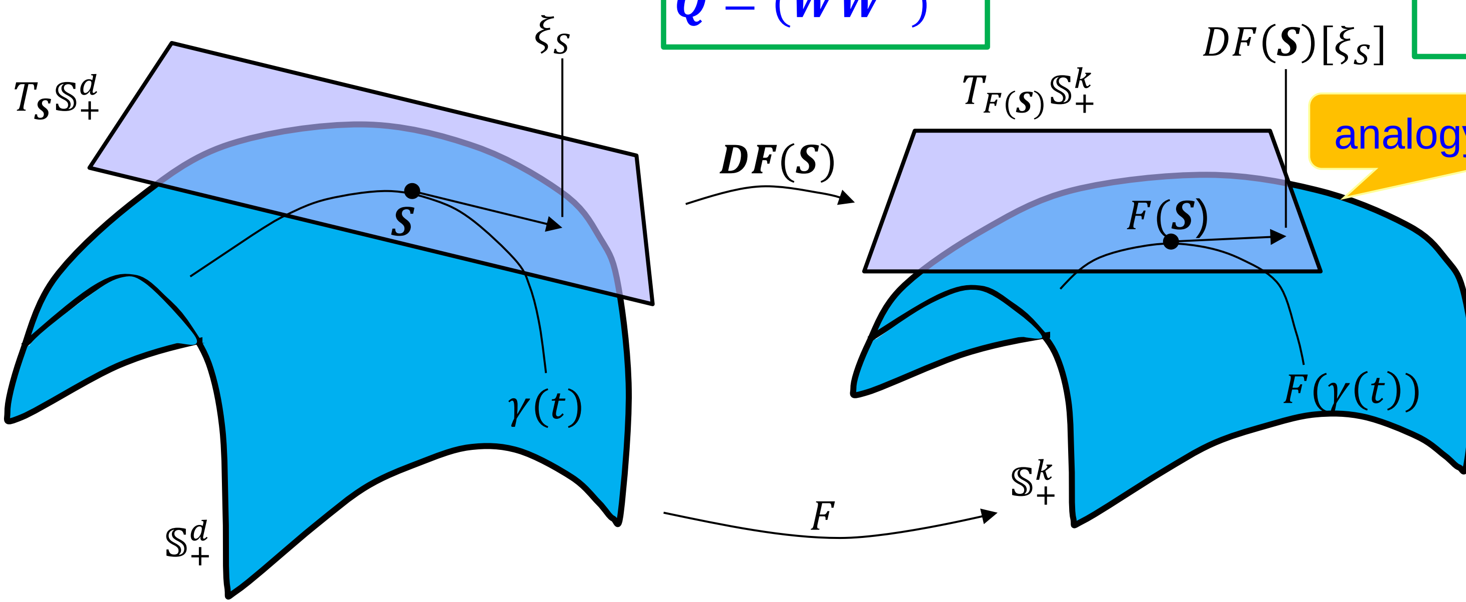
LEML

- SPD tangent map learning
  - Mapping function: DF: $f(\log(\boldsymbol{S})) = \boldsymbol{W}^T \log(\boldsymbol{S})\boldsymbol{W}$
    - $\boldsymbol{W}$ is column full rank
  - Log-Euclidean distance in the target tangent space
    - $d_{LED}\left(f(\boldsymbol{T}_i), f(\boldsymbol{T}_j)\right) = ||\boldsymbol{W}^T\boldsymbol{T}_i\boldsymbol{W} - \boldsymbol{W}^T\boldsymbol{T}_j\boldsymbol{W}||_F$

$$= tr(\boldsymbol{Q}(\boldsymbol{T}_i - \boldsymbol{T}_j)(\boldsymbol{T}_i - \boldsymbol{T}_j))$$

$\boldsymbol{Q} = (\boldsymbol{W}\boldsymbol{W}^T)^2$

$\boldsymbol{T} = \log(\boldsymbol{S})$



analogy to 2DPCA

LEML

- Discriminative learning
  - Objective function
    - $\arg\min_{Q,\xi} D_{ld}(\boldsymbol{Q}, \boldsymbol{Q_0}) + \eta D_{ld}(\boldsymbol{\xi}, \boldsymbol{\xi_0})$

$$\text{s.t.}, \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{A}_{ij}^{T}\mathbf{A}_{ij}\right) \leq \xi_{\boldsymbol{c}(i,j)}, (i,j) \in \boldsymbol{S}$$

$$\operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{A}_{ij}^{T}\mathbf{A}_{ij}\right) \geq \xi_{\boldsymbol{c}(i,j)}, (i,j) \in \boldsymbol{D}$$

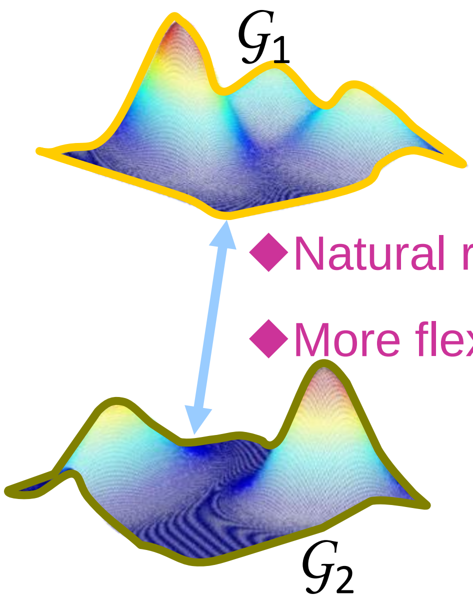
    - $\mathbf{A}_{ij} = \log(\boldsymbol{C}_i) - \log(\boldsymbol{C}_j)$, D_{ld}: LogDet divergence
  - Optimization
    - Cyclic Bregman projection algorithm [Bregman’1967]

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Set model IV: statistics (COV+)

- Properties
  - The natural raw statistics of a sample set
  - Flexible model of multiple-order statistical information
- Methods
  - CDL [CVPR'12]
  - LMKML [ICCV'13]
  - DARG [CVPR'15]
  - SPD-ML [ECCV'14]
  - LEML [ICML'15]
  - LERM [CVPR'14]
  - HER [CVPR'15]
  - …



◆Natural raw statistics

◆More flexible model

[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]

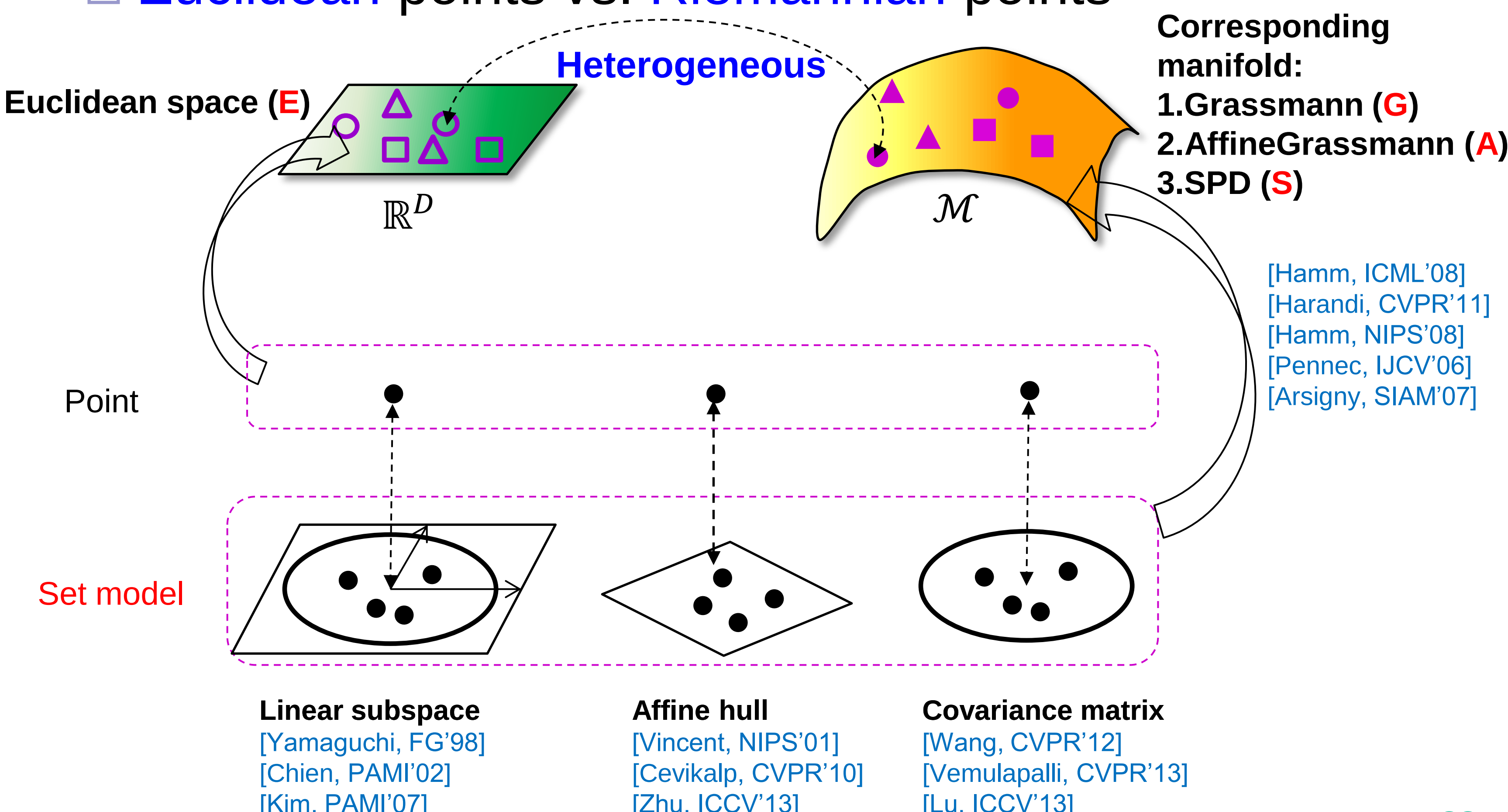
Set model IV: statistics (COV+)

- LERM (Learning Euclidean-to-Riemannian Metric) [CVPR'14]
  - Application scenario: still-to-video face recognition
  - Metric learning: cross Euclidean space and Riemannian manifold

[1] Z. Huang, R. Wang, S. Shan, X. Chen. Learning Euclidean-to-Riemannian Metric for Point-to-Set Classification. *IEEE CVPR 2014*.

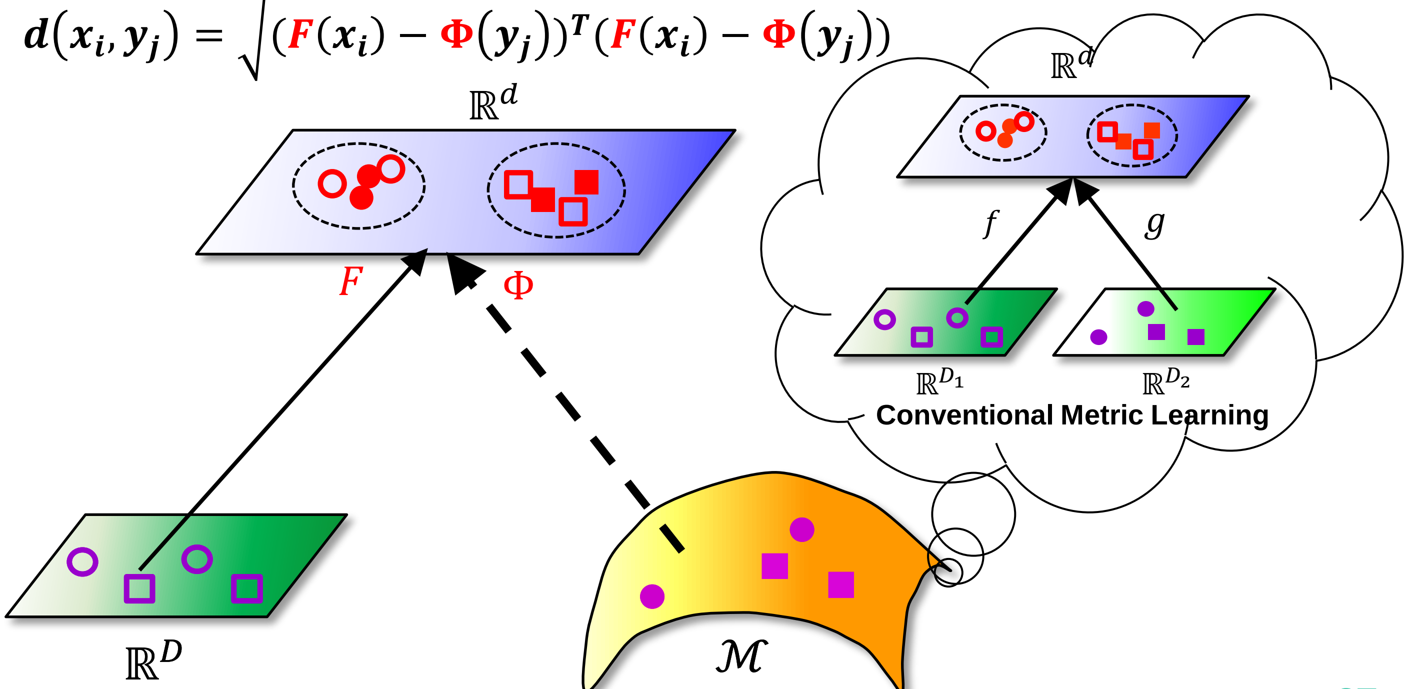
LERM

- Point-to-Set Classification
  - Euclidean points vs. Riemannian points



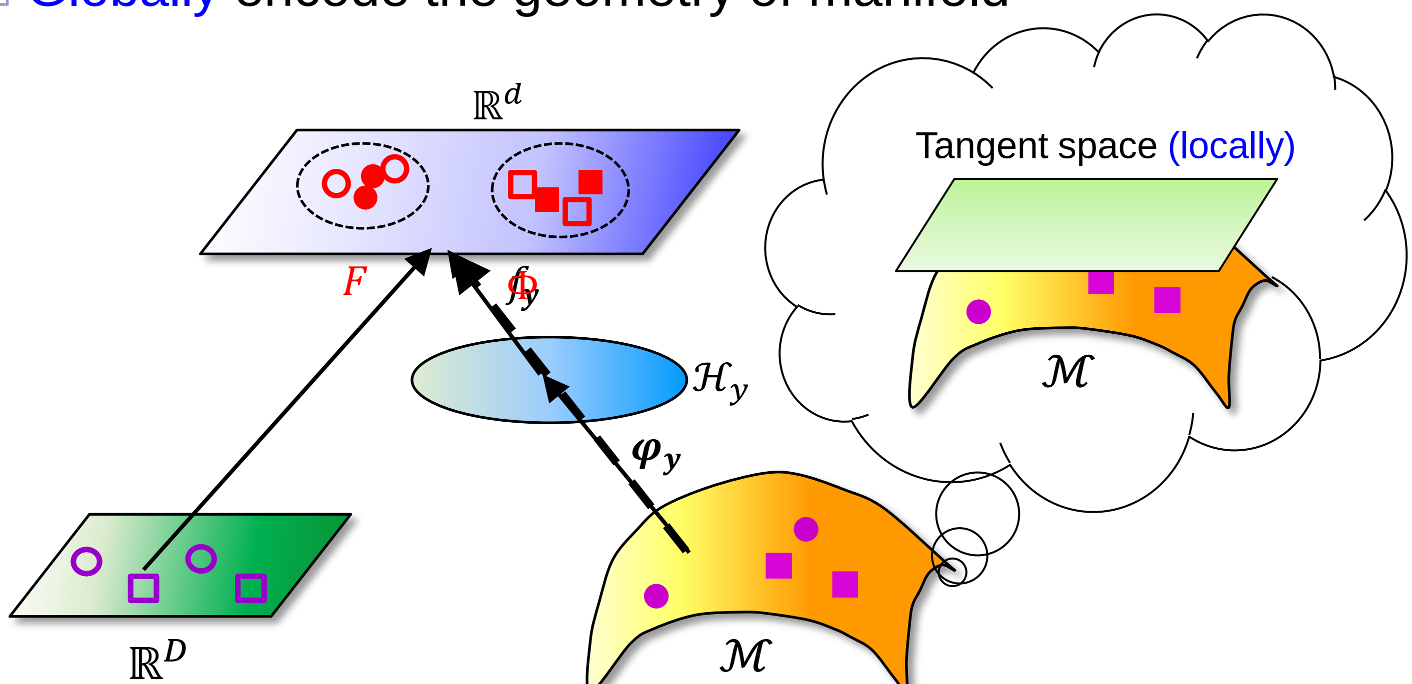
LERM

- Basic idea
  - Reduce Euclidean-to-Riemannian metric to classical Euclidean metric
    - Seek maps F, Φ to a common Euclidean subspace
      - $d(x_i, y_j) = \sqrt{(F(x_i) - \Phi(y_j))^T (F(x_i) - \Phi(y_j))}$



LERM

- Basic idea
  - Bridge Euclidean-to-Riemannian gap
    - Hilbert space embedding
      - Adhere to Euclidean geometry
      - Globally encode the geometry of manifold



LERM

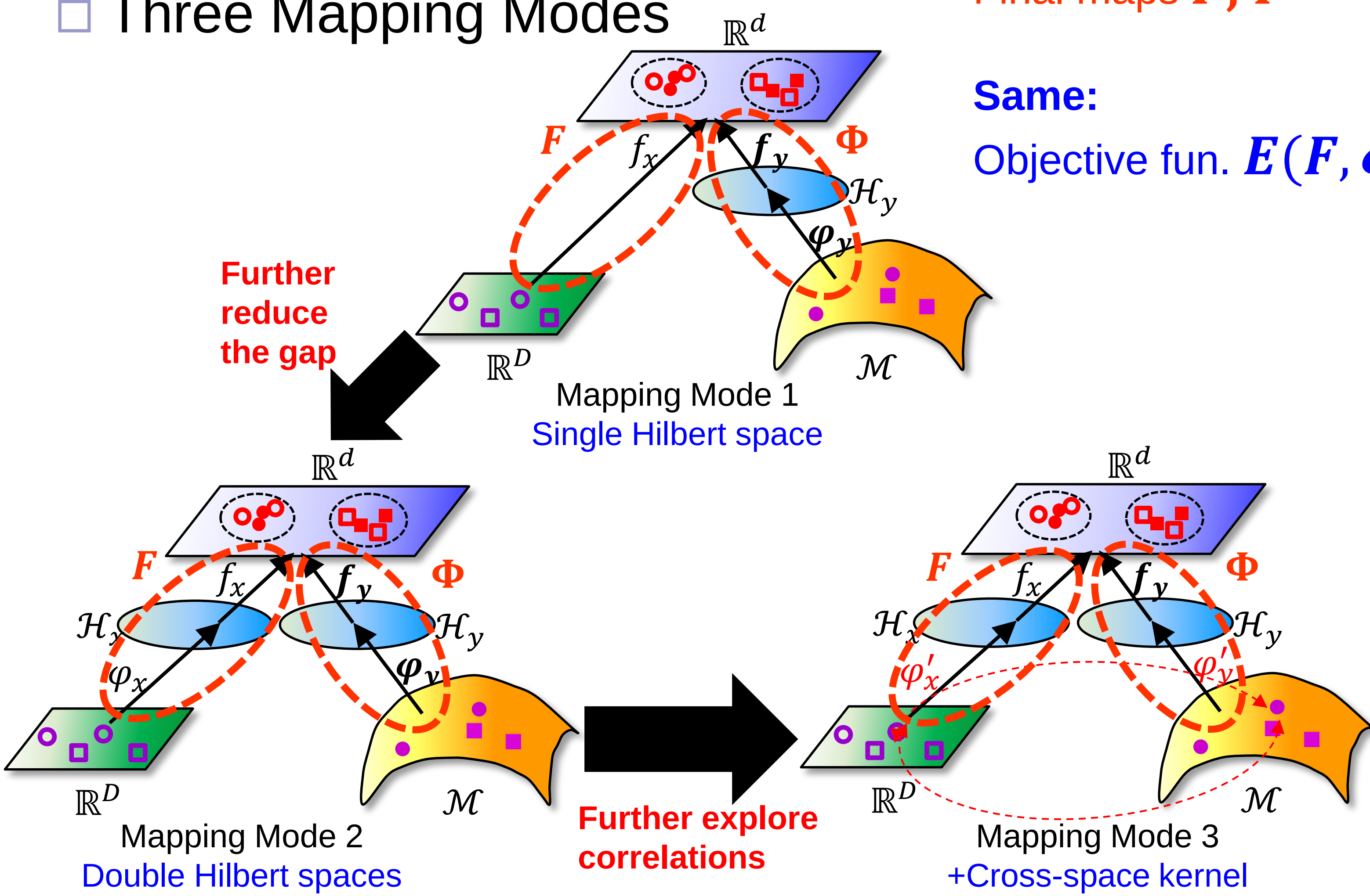
- Formulation
  - Three Mapping Modes

Different:

Final maps $\boldsymbol{F}, \boldsymbol{\Phi}$

Same:

Objective fun. $\boldsymbol{E}(\boldsymbol{F}, \boldsymbol{\varphi})$

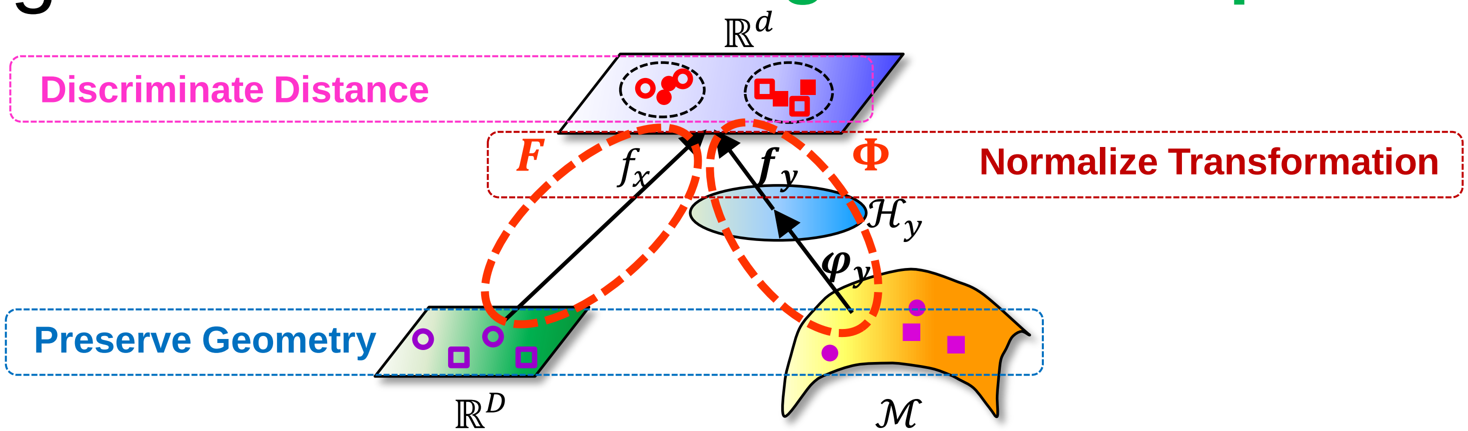


LERM

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

- e.g. Mode 1

Single Hilbert space



Final maps:

$$F = f_x = W_x^T X$$

$$\Phi = f_y \circ \varphi_y = W_y^T K_y$$

$$\langle \varphi_{y_i}, \varphi_{y_j} \rangle = K_y(i,j)$$

$$K_y(i,j) = \exp(-d^2(y_i, y_j)/2\sigma^2)$$

Riemannian metrics [ICML'08, NIPS'08, SIAM'06]

Distance metric:

$$d(x_i, y_j) = \sqrt{(F(x_i) - \Phi(y_j))^T (F(x_i) - \Phi(y_j))}$$

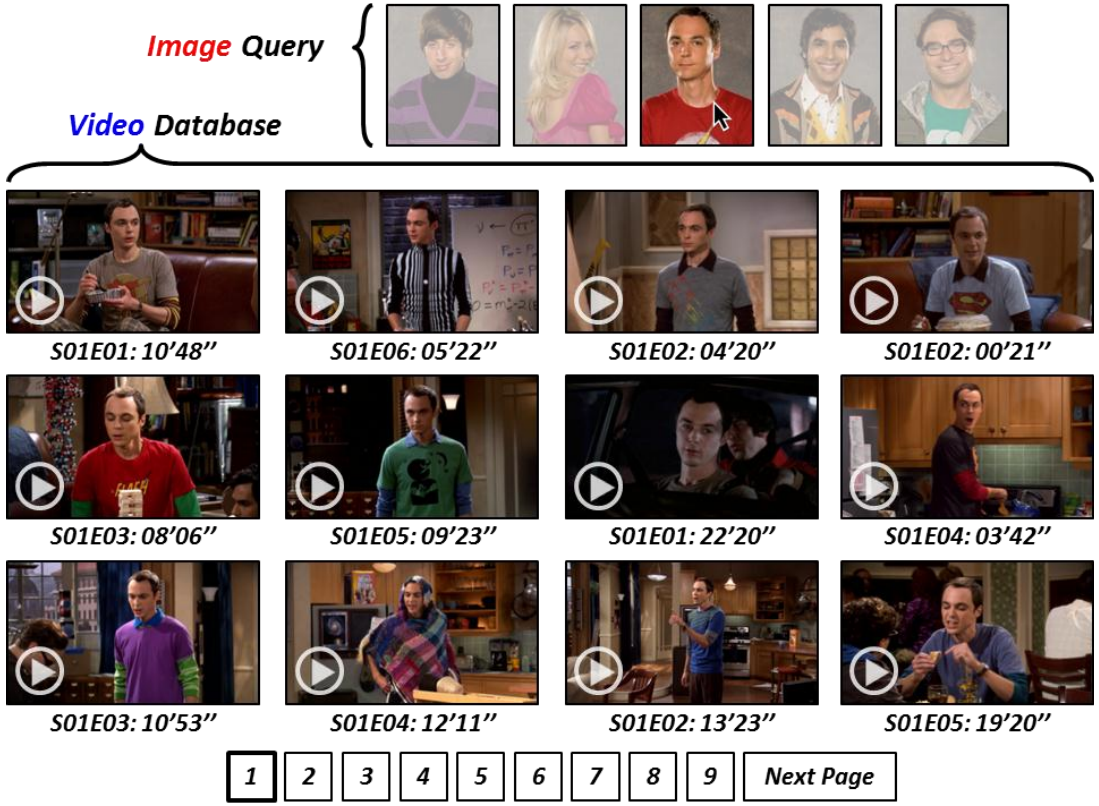
Objective function: $E(F, \varphi)$

$$\min_{F,\Phi} \{ \boxed{D(F,\Phi)} + \lambda_1 \boxed{G(F,\Phi)} + \lambda_2 \boxed{T(F,\Phi)} \}$$

Distance Geometry Transformation

Set model IV: statistics (COV+)

- HER (Hashing across Euclidean and Riemannian) [CVPR'15]
  - Application scenario: video-based face retrieval
  - Metric learning: hamming distance learning across heter. spaces

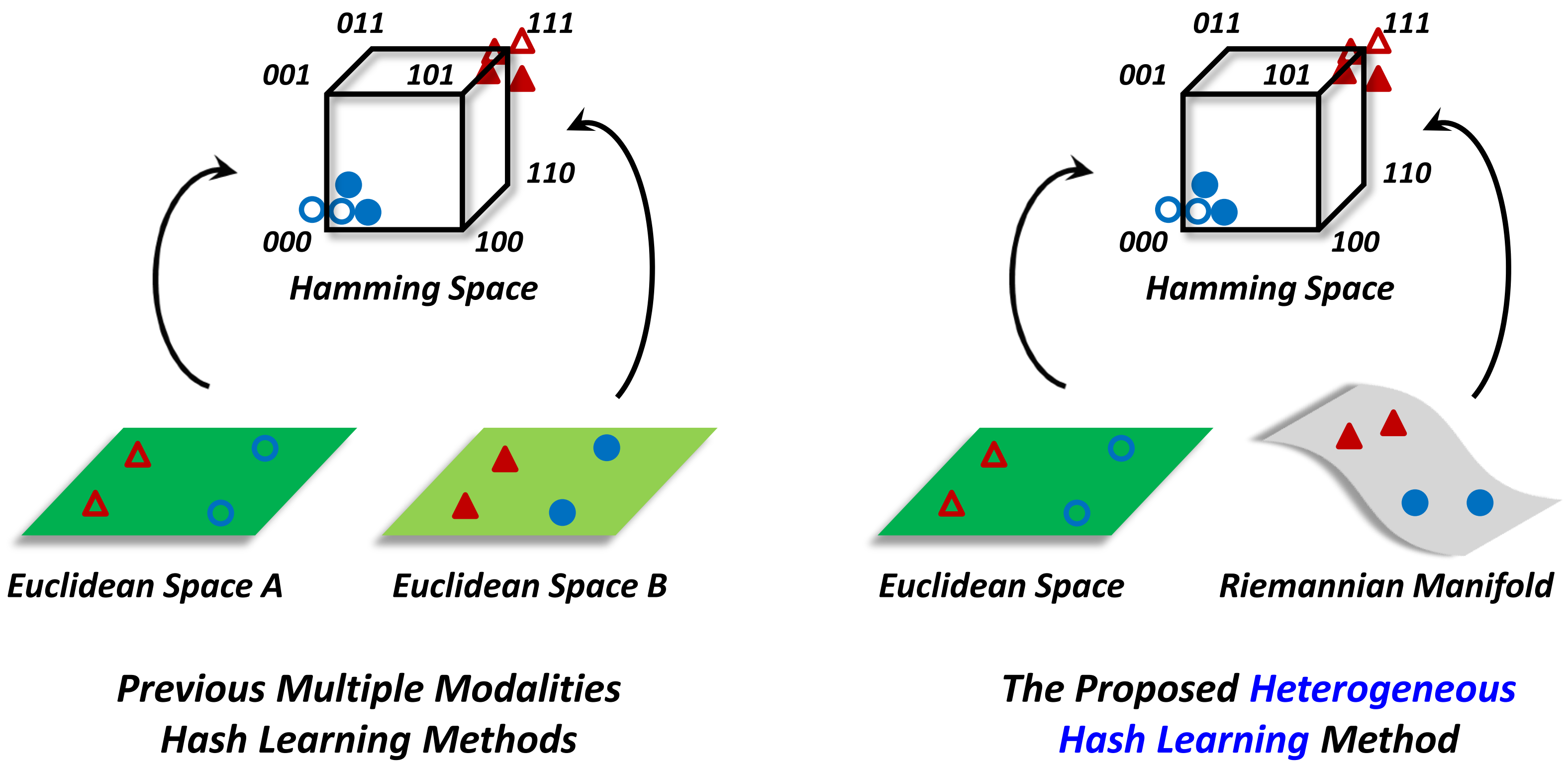


[1] Y. Li, R. Wang, Z. Huang, S. Shan, X. Chen. Face Video Retrieval with Image Query via Hashing across Euclidean Space and Riemannian Manifold. *IEEE CVPR 2015*.

中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

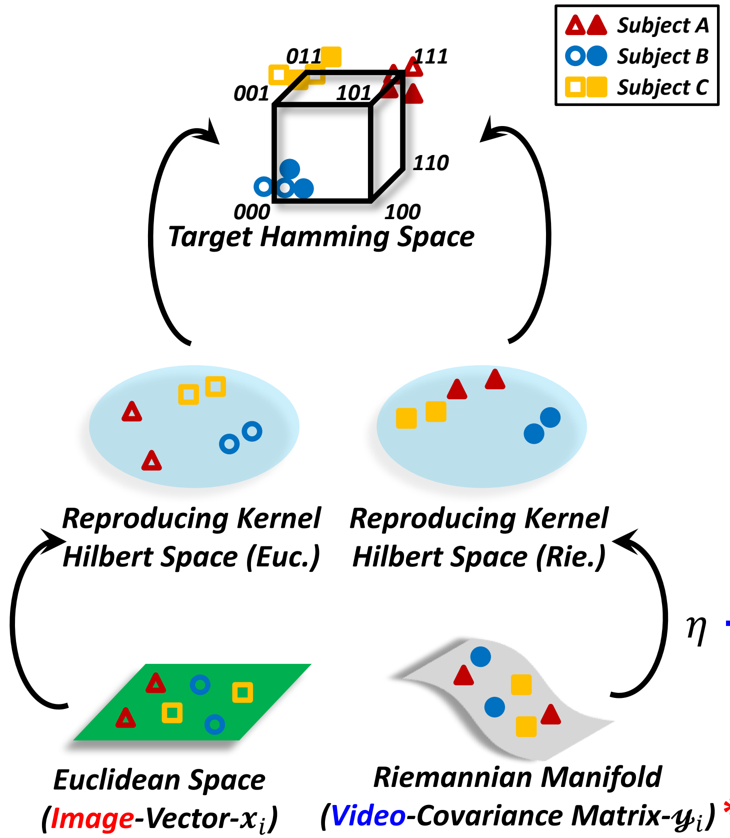
HER

- Heterogeneous Hash Learning



HER

- Two-Stage Architecture
  - Stage-1: kernel mapping



Euclidean Kernel (Image)

$$K_e^{ij} = \exp(-\frac{\|x_i - x_j\|_2^2}{2\sigma_e^2})$$

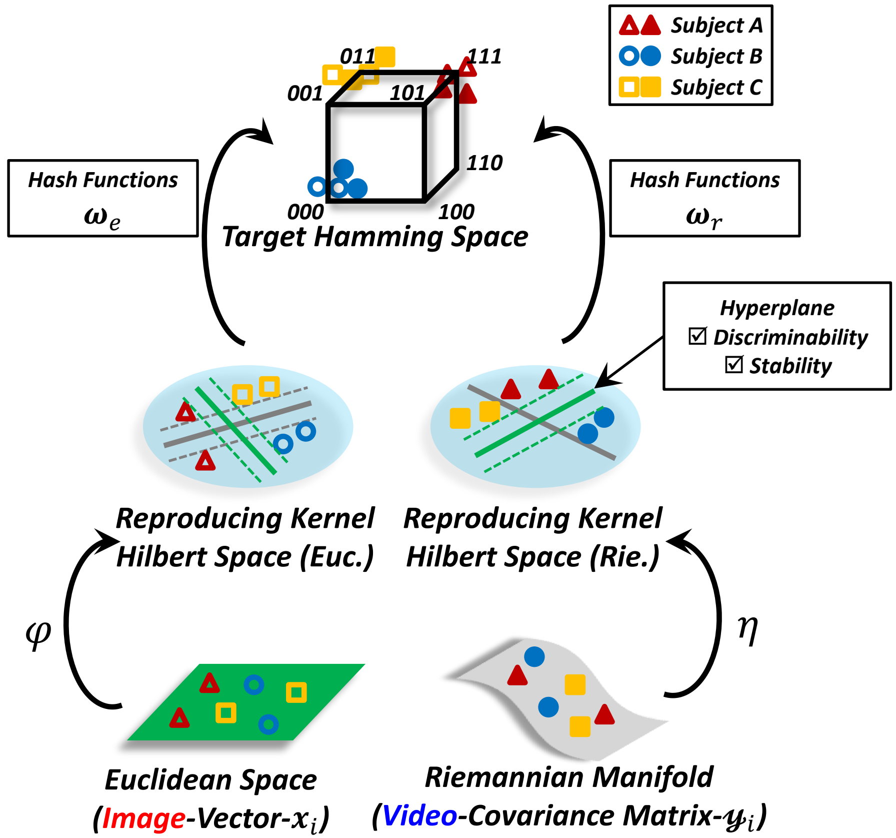
Riemannian Kernel (Video)

$$K_r^{ij} = \exp(-\frac{\|\log(\mathcal{Y}_i) - \log(\mathcal{Y}_j)\|_F^2}{2\sigma_r^2})$$

[*] Y. Li, R. Wang, Z. Cui, S. Shan, X. Chen. Compact Video Code and Its Application to Robust Face Retrieval in TV-Series. *BMVC 2014*. (**: **CVC** method for video-video face retrieval)

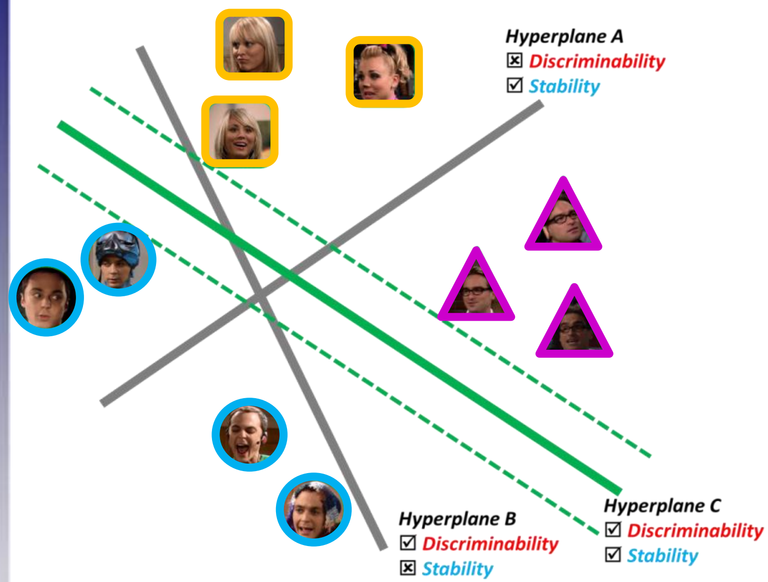
HER

- Two-Stage Architecture
  - Stage-2: binary encoding



HER

- Binary encoding [Rastegari, ECCV'12]



Discriminability

$$E_e = \sum_{c\in\{1:C\}} \sum_{m,n\in c} d(\boldsymbol{B}_e^m, \boldsymbol{B}_e^n) - \lambda_e \sum_{\substack{c_1\in\{1:C\}\\ p\in c_1}} \sum_{\substack{c_2\in\{1:C\}\\ c_1\neq c_2, q\in c_2}} d\left(\boldsymbol{B}_e^p, \boldsymbol{B}_e^q\right)$$

$$E_r = \sum_{c\in\{1:C\}} \sum_{m,n\in c} d(\boldsymbol{B}_r^m, \boldsymbol{B}_r^n) - \lambda_r \sum_{\substack{c_1\in\{1:C\}\\ p\in c_1}} \sum_{\substack{c_2\in\{1:C\}\\ c_1\neq c_2, q\in c_2}} d\left(\boldsymbol{B}_r^p, \boldsymbol{B}_r^q\right)$$

$$E_{er} = \sum_{c\in\{1:C\}} \sum_{m,n\in c} d(\boldsymbol{B}_e^m, \boldsymbol{B}_r^n) - \lambda_{er} \sum_{\substack{c_1\in\{1:C\}\\ p\in c_1}} \sum_{\substack{c_2\in\{1:C\}\\ c_1\neq c_2, q\in c_2}} d\left(\boldsymbol{B}_e^p, \boldsymbol{B}_r^q\right)$$

Objective Function

$$\min_{\omega_e,\omega_r,\xi_e,\xi_r,\boldsymbol{B}_e,\boldsymbol{B}_r} \boxed{\lambda_1 E_e + \lambda_2 E_r + \lambda_3 E_{er}} + \boxed{\gamma_1 \sum_{k\in\{1:K\}} \left\|\boldsymbol{\omega}_e^k\right\|^2 + C_1 \sum_{\substack{k\in\{1:K\}\\ i\in\{1:N\}}} \xi_e^{ki}} + \boxed{\gamma_2 \sum_{k\in\{1:K\}} \left\|\boldsymbol{\omega}_r^k\right\|^2 + C_2 \sum_{\substack{k\in\{1:K\}\\ i\in\{1:N\}}} \xi_r^{ki}}$$

s.t.

$$\boldsymbol{B}_e^{ki} = sgn\left(\boldsymbol{\omega}_e^{k^T} \varphi(\boldsymbol{x}_i)\right), \forall k \in \{1{:}K\}, \forall i \in \{1{:}N\}$$

$$\boldsymbol{B}_r^{ki} = sgn\left(\boldsymbol{\omega}_r^{k^T} \eta(\boldsymbol{y}_i)\right), \forall k \in \{1{:}K\}, \forall i \in \{1{:}N\}$$

$$\boldsymbol{B}_r^{ki}\left(\boldsymbol{\omega}_e^{k^T} \varphi(\boldsymbol{x}_i)\right) \geq 1 - \xi_e^{ki}, \xi_e^{ki} > 0, \forall k \in \{1{:}K\}, \forall i \in \{1{:}N\}$$

$$\boldsymbol{B}_e^{ki}\left(\boldsymbol{\omega}_r^{k^T} \eta(\boldsymbol{y}_i)\right) \geq 1 - \xi_r^{ki}, \xi_r^{ki} > 0, \forall k \in \{1{:}K\}, \forall i \in \{1{:}N\}$$

Discriminability (LDA)

Stability (SVM)

Compatibility (Cross-training)

Outline

- Background
- Literature review
- Evaluations
- Summary

Evaluations

- Two YouTube datasets
  - YouTube Celebrities (YTC) [Kim, CVPR'08]
    - 47 subjects, 1910 videos from YouTube
  - YouTube FaceDB (YTF) [Wolf, CVPR'11]
    - 3425 videos, 1595 different people

YTC

YTF

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Evaluations

- COX Face [Huang, ACCV'12/TIP'15]
  - 1,000 subjects
    - each has 1 high quality images, 3 unconstrained video sequences

Images

Videos

http://vipl.ict.ac.cn/resources/datasets/cox-face-dataset/cox-face

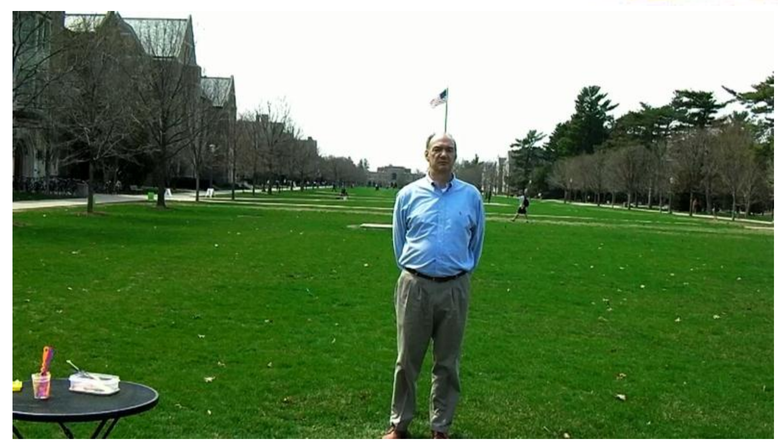
Evaluations

- PaSC [Beveridge, BTAS’13]
  - Control videos
    - 1 mounted video camera
    - 1920*1080 resolution
  - Handheld videos
    - 5 handheld video cameras
    - 640*480~1280*720 resolutic

Table 2. Summary of Video PaSC Data.

Number of Subjects	265
Total Videos	2,802
Total Control Videos	1,401
Total Handheld Videos	1,401
Control Videos per Subject	4 to 7
Handheld Videos per Subject	4 to 7
Number of Locations	6

Control video



Handheld video

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Evaluations

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

■ Results (*reported in our DARG paper**)

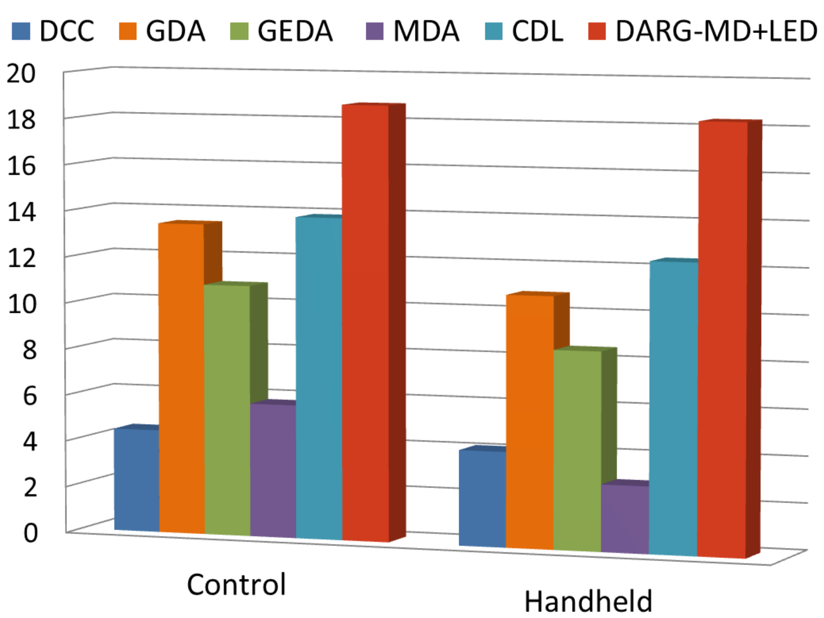
Method	YTC	COX					
		COX-11	COX-12	COX-23	COX-21	COX-31	COX-32
CHISD [CVPR'10]	66.46	56.87	30.10	14.80	44.37	26.44	13.68
GDA [CVPR'08]	65.91	72.26	80.70	74.36	71.44	81.99	77.57
GGDA [CVPR'11]	66.83	76.73	83.80	76.59	72.56	82.84	79.99
MMD [CVPR'08]	65.30	38.29	30.34	15.24	34.86	22.21	11.44
MDA [CVPR'09]	66.98	65.82	63.01	36.17	55.46	43.23	29.70
SGM [ECCV'02]	52.00	26.74	14.32	12.39	26.03	19.21	10.50
MDM [CVPR'05]	62.12	30.70	24.98	14.30	28.90	31.72	19.30
CDL [CVPR'12]	69.70	78.37	85.25	79.74	75.59	85.83	81.87
DARG-KLD	72.21	71.93	80.11	73.65	70.87	81.03	76.99
DARG-LGD	68.72	76.74	84.99	78.02	72.93	83.88	81.54
DARG-MD+LED	**77.09**	**83.71**	**90.13**	**85.08**	**81.96**	**89.99**	**88.35**

[*] W. Wang, R. Wang, Z. Huang, S. Shan, X. Chen. Discriminant Analysis on Riemannian Manifold of Gaussian Distributions for Face Recognition with Image Sets. *IEEE CVPR 2015*.

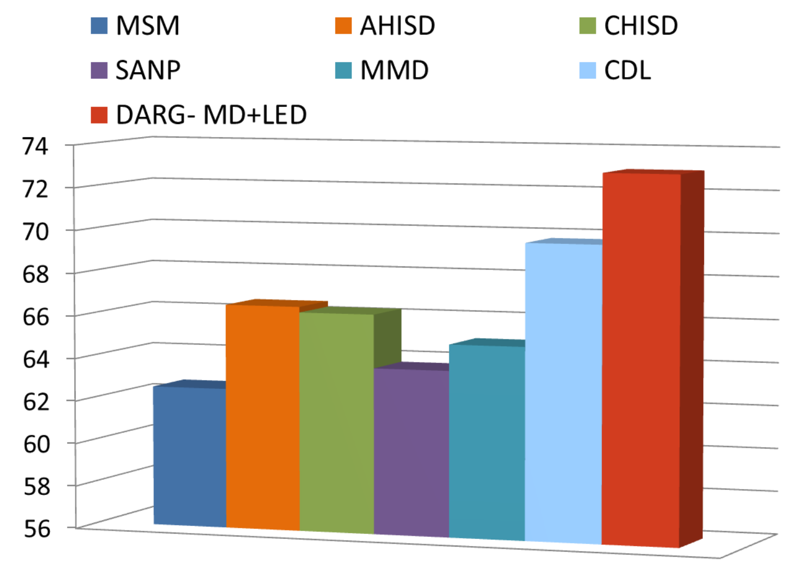
Evaluations

- Results (*reported in our DARG paper**)

VR@FAR=0.01 on PaSC



AUC on YTF

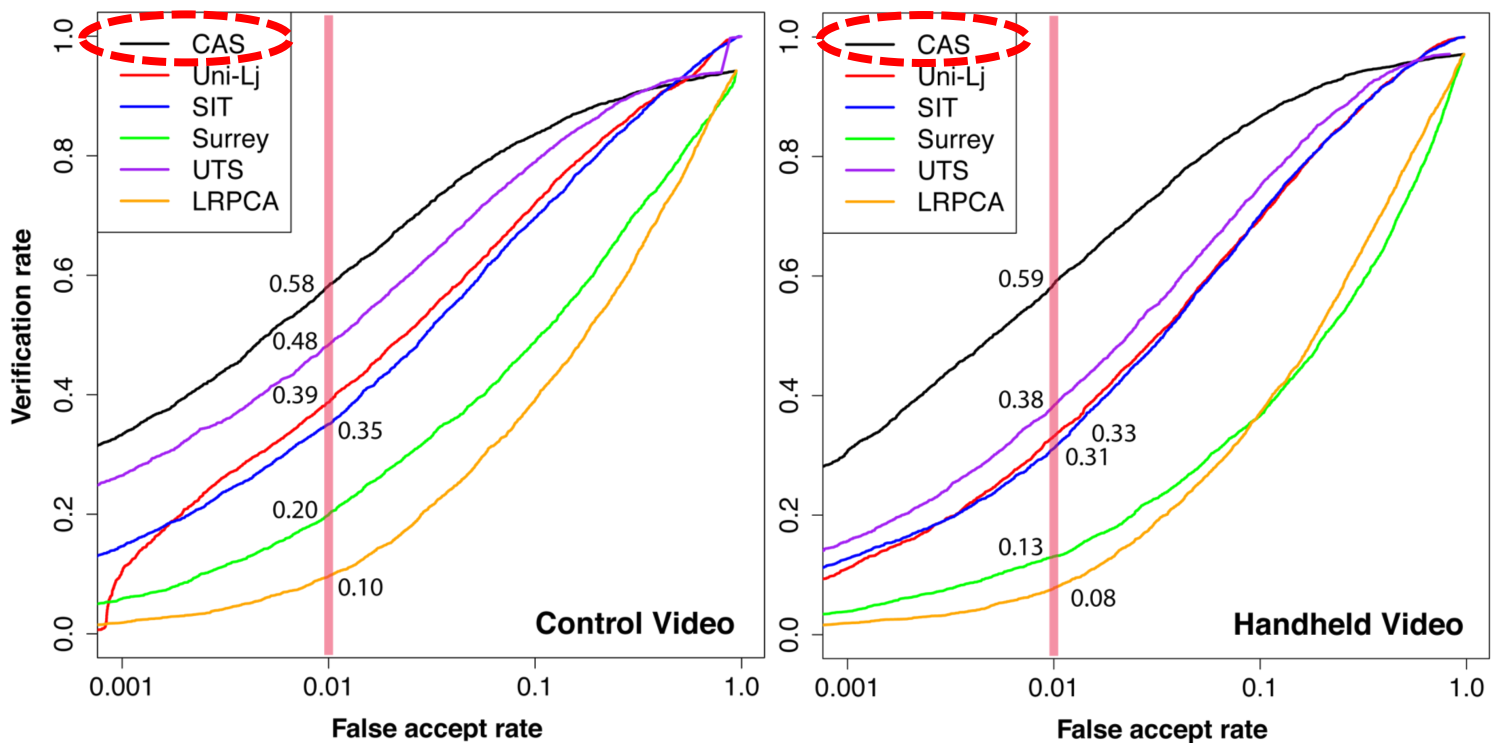


中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Evaluations

- Performance on PaSC Challenge (IEEE FG'15)
  - HERML-DeLF
    - DCNN learned image feature
    - Hybrid Euclidean and Riemannian Metric Learning*



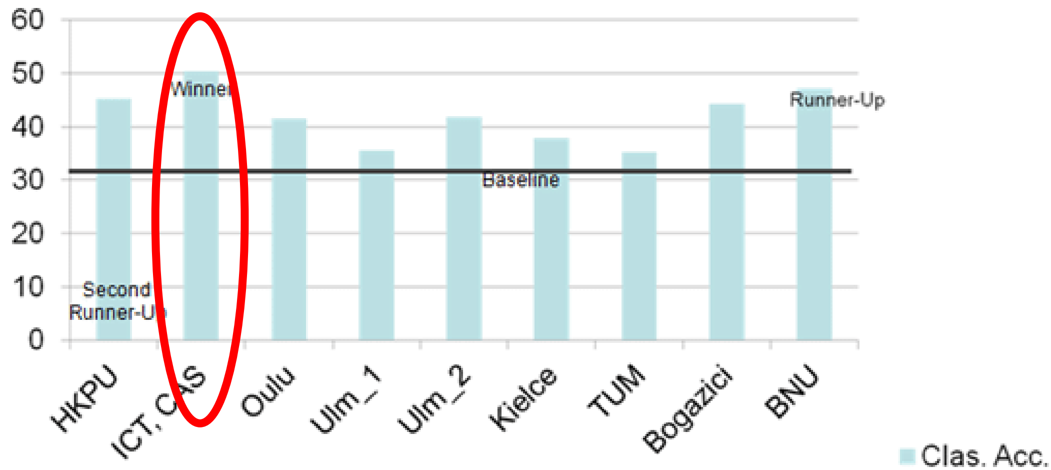
[*] Z. Huang, R. Wang, S. Shan, X. Chen. Hybrid Euclidean-and-Riemannian Metric Learning for Image Set Classification. *ACCV 2014*. (**: the key reference describing the method used for the challenge)

Evaluations

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- Performance on EmotiW Challenge (ACM ICMI'14)*
  - Combination of multiple statistics for video modeling
  - Learning on the Riemannian manifold

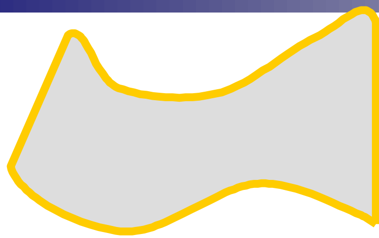


[*] M. Liu, R. Wang, S. Li, S. Shan, Z. Huang, X. Chen. Combining Multiple Kernel Methods on Riemannian Manifold for Emotion Recognition in the Wild. *ACM ICMI 2014*.

Outline

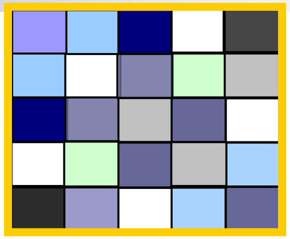
- Background
- Literature review
- Evaluations
- Summary

Route map of our methods



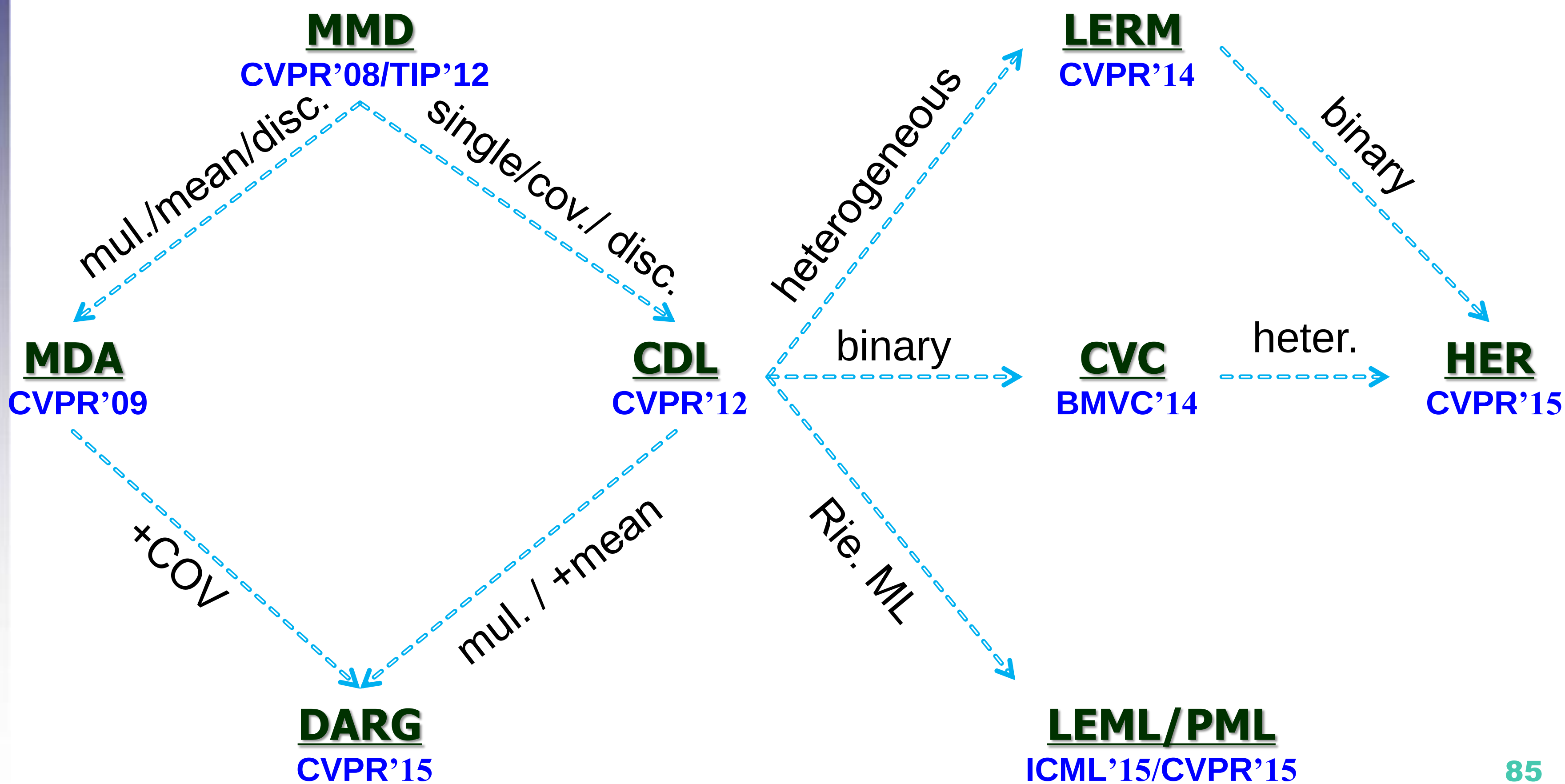
Appearance manifold

- Complex distribution
- Large amount of data



Covariance matrix

- Natural raw statistics
- No assum. of data dist.



中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Summary

- What we learn from current studies
  - Set modeling
    - Linear(/affine) subspace → Manifold → Statistics
  - Set matching
    - Non-discriminative → Discriminative
  - Metric learning
    - Euclidean space → Riemannian manifold
- Future directions
  - More flexible set modeling for different scenarios
  - Multi-model combination
  - Learning method should be more efficient
  - Set-based vs. sample-based?

Additional references (not listed above)

- [Arandjelović, CVPR'05] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face Recognition with Image Sets Using Manifold Density Divergence. *IEEE CVPR 2005*.
- [Chien, PAMI'02] J. Chien and C. Wu. Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE T-PAMI 2002*.
- [Rastegari, ECCV'12] M. Rastegari, A. Farhadi, and D. Forsyth. Attribute discovery via predictable discriminative binary codes. *ECCV 2012*.
- [Shakhnarovich, ECCV'02] G. Shakhnarovich, J. W. Fisher, and T. Darrell. Face Recognition from Long-term Observations. *ECCV 2002*.
- [Vemulapalli, CVPR'13] R. Vemulapalli, J. K. Pillai, and R. Chellappa. Kernel learning for extrinsic classification of manifold features. *IEEE CVPR 2013*.
- [Vincent, NIPS'01] P. Vincent and Y. Bengio. K-local hyperplane and convex distance nearest neighbor algorithms. *NIPS 2001*.

Thanks, Q & A

Zhiwu Huang

Yan Li

Wen Wang

Mengyi Liu

Ruiping Wang

Shiguang Shan

Xilin Chen

Lab of Visual Information Processing and Learning (VIPL) @ICT@CAS

Codes of our methods available at: http://vipl.ict.ac.cn/resources/codes