Riemannian Metric Learning and its Vision Applications

Ruiping Wang

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

Nov. 2, 2019

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

Outline

- Background
  - Visual feature aggregation
  - Metric learning
- Literature review
  - Image set classification
  - Image recognition (fine-grained)
- Summary

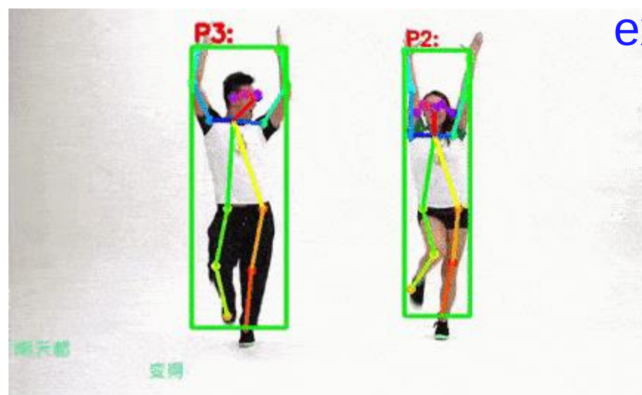
Video classification

Searching criminal suspects

Smart TV-Series Character Shots Retrieval

Automating facial expression analysis

Gesture analysis



Action analysis

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

Video as Image Set

- Image set classification
  - Unconstrained acquisition conditions
  - Complex appearance variations

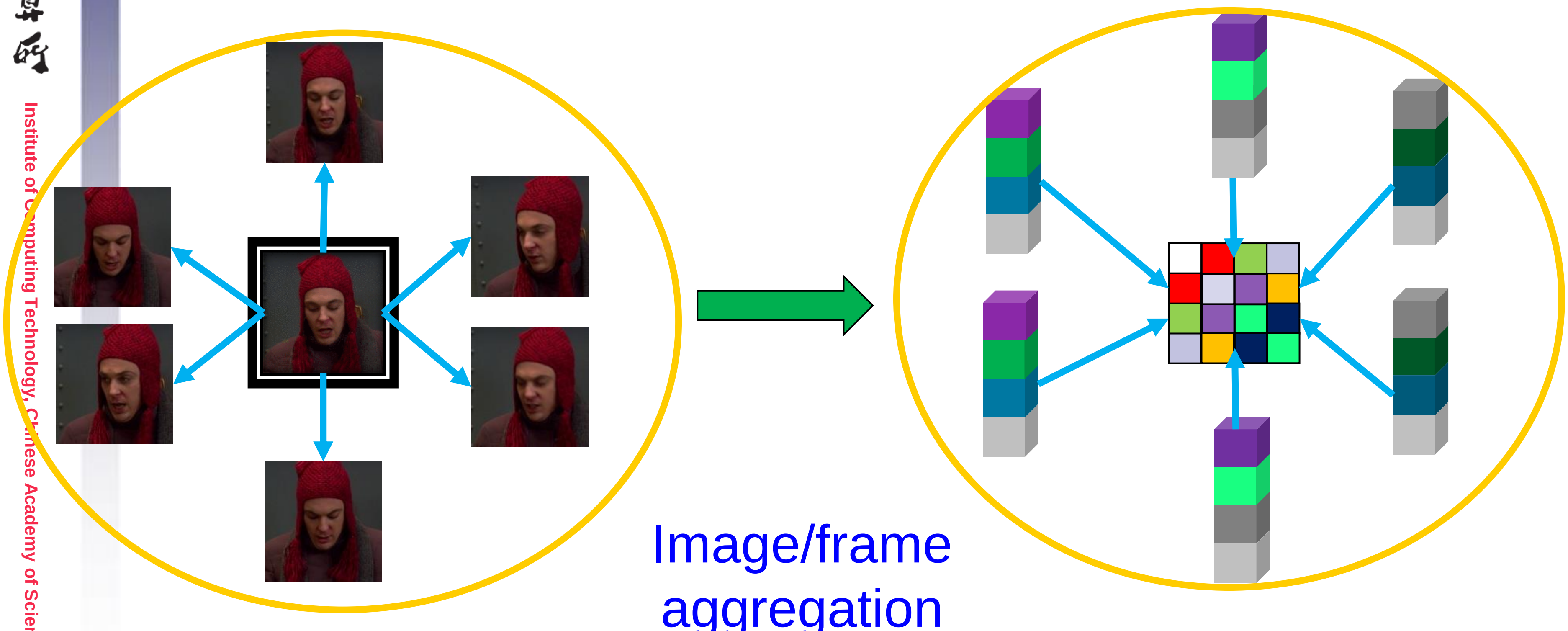
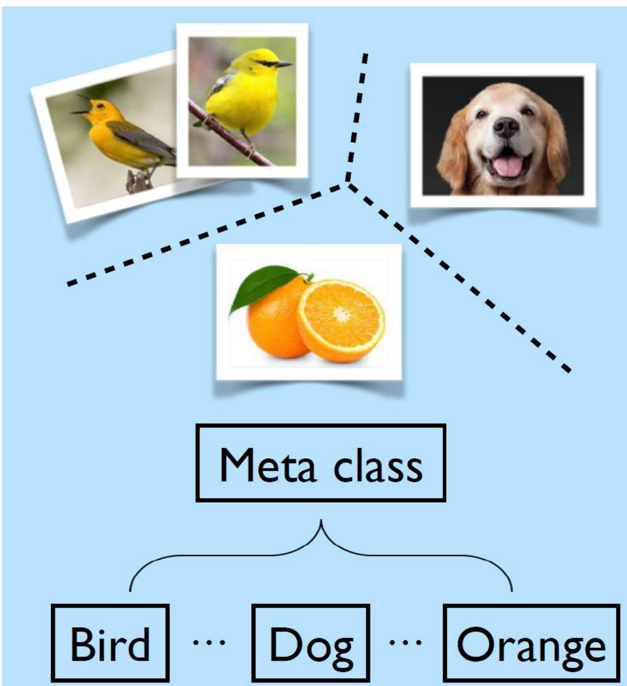
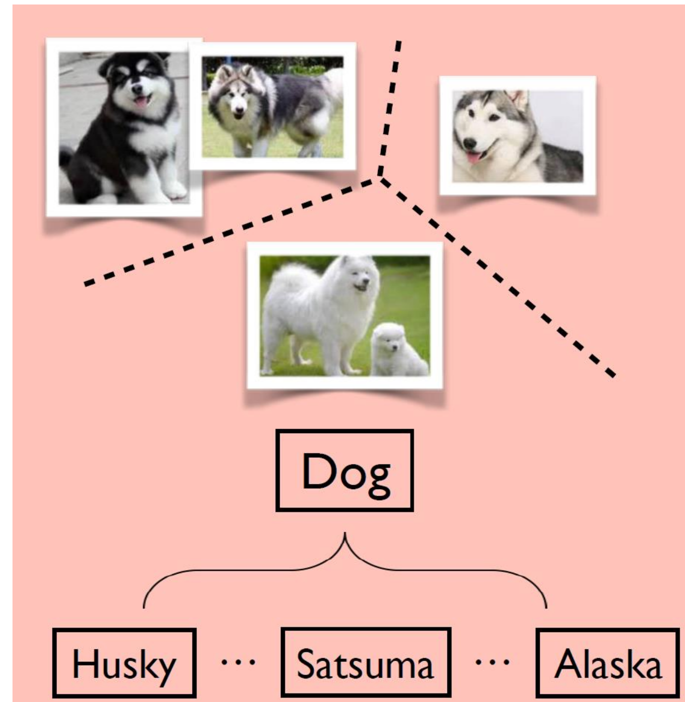


Image recognition

- General recognition



- Fine-grained recognition

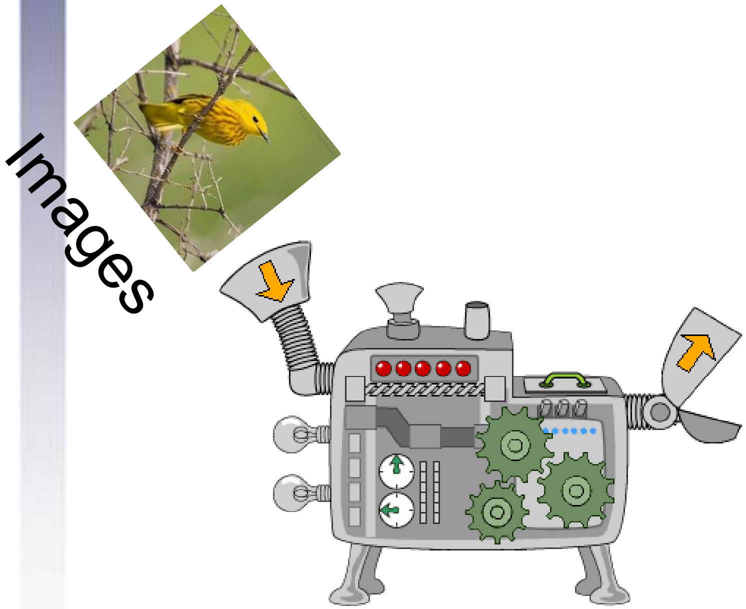


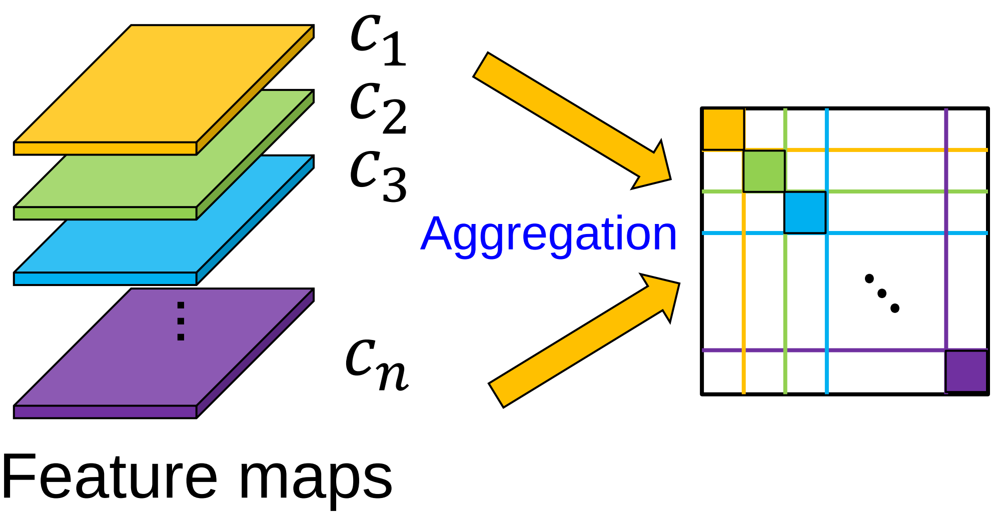
Figures courtesy: Wei *et al.* arXiv1907.03069

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

Image recognition

- Highly complicated features extraction
  - Deep and multi-channel
  - Aggregation for classifiers

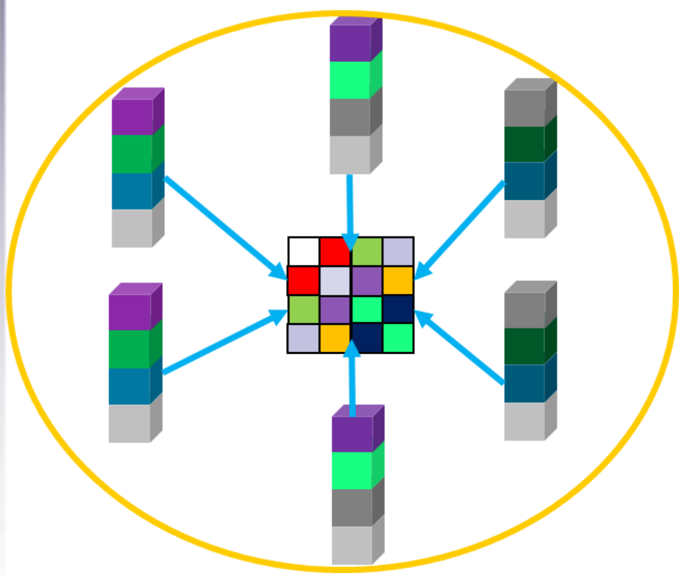




中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Common representation paradigm

- Aggregation to obtain informative feature
  - Covariance matrix
  - Gaussian distribution
  - …



Video representation

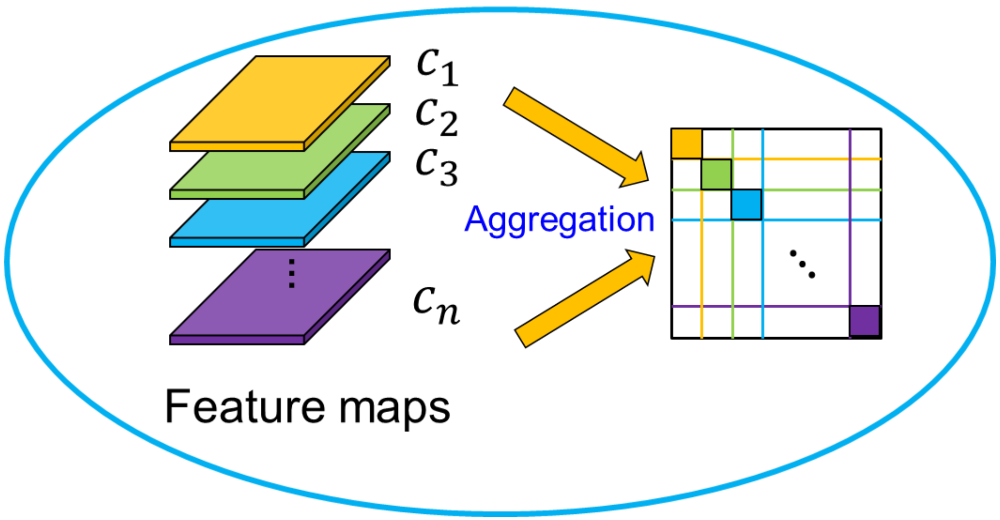


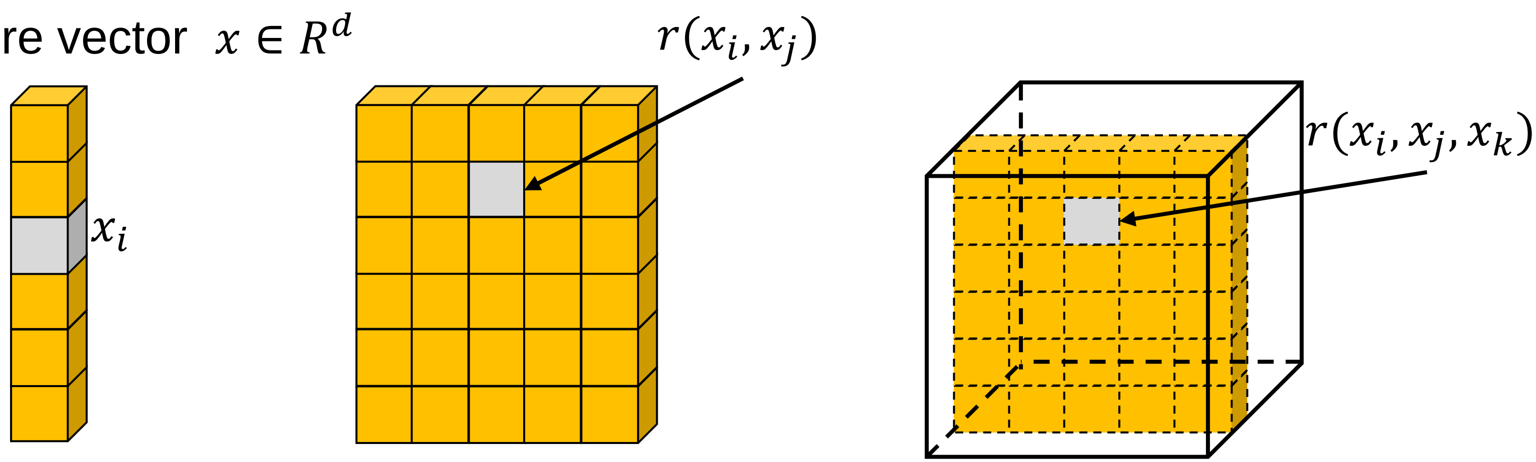
Image representation

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

Second-order and higher

- What is second-order?

Feature vector $x \in R^d$



1st order

$R = x$

2nd order

$$R = \frac{1}{|c|}\sum_{x \in C} x \otimes x$$

3rd order

$$R = \frac{1}{|c|}\sum_{x \in C} x \otimes x \otimes x$$

Metric learning

- Sample similarity
  - Euclidean distance

$$d(\mathbf{x}_1, \mathbf{x}_2) = \left\| \mathbf{x}_1 - \mathbf{x}_2 \right\|_2 = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T (\mathbf{x}_1 - \mathbf{x}_2)}$$

  - Mahalanobis distance

$$d_M(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{(\mathbf{x}_1 - \mathbf{x}_2)^T \Sigma^{-1} (\mathbf{x}_1 - \mathbf{x}_2)}$$

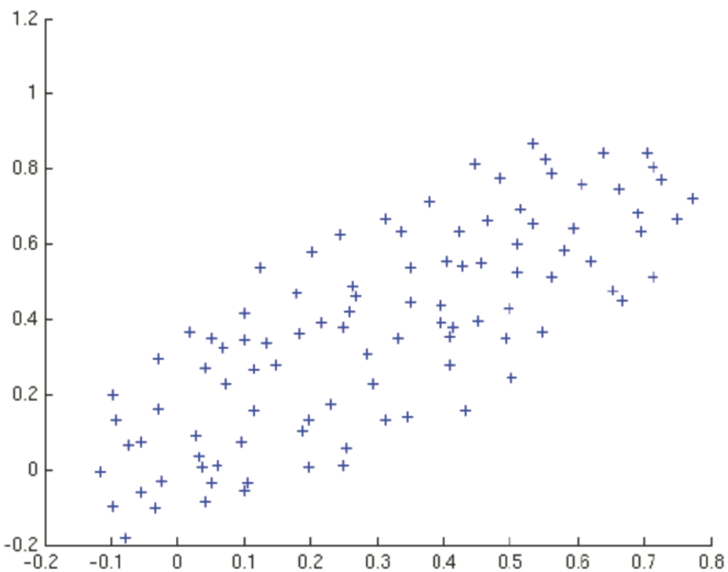
where

$$\Sigma = \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T$$

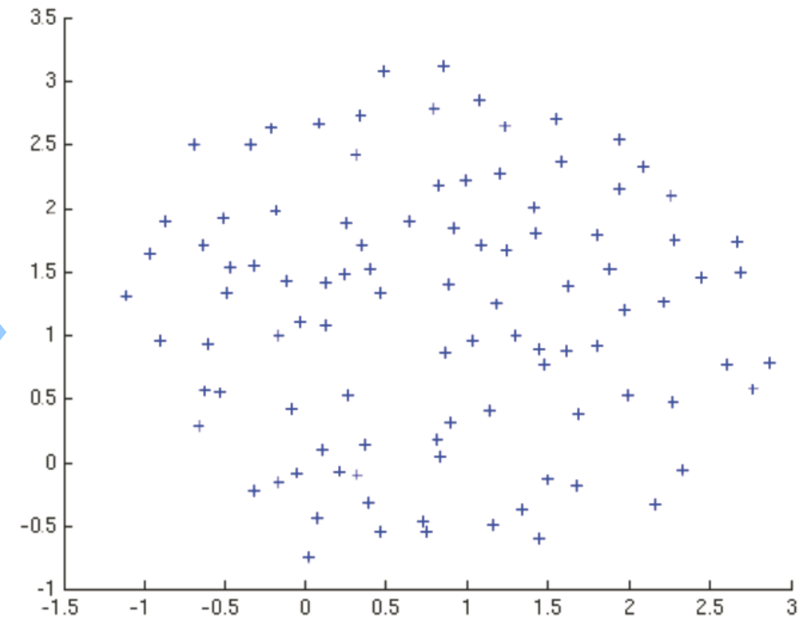
$$\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$$

Metric learning

- Sample similarity
  - Euclidean distance
  - Mahalanobis distance



Euclidean distance



Mahalanobis distance

Metric learning

- Basic formulation
  - Applying Mahalanobis distance to learn a semi-positive semi-definite (PSD) matrix

$$d_M(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$$

  - Relationship with subspace learning

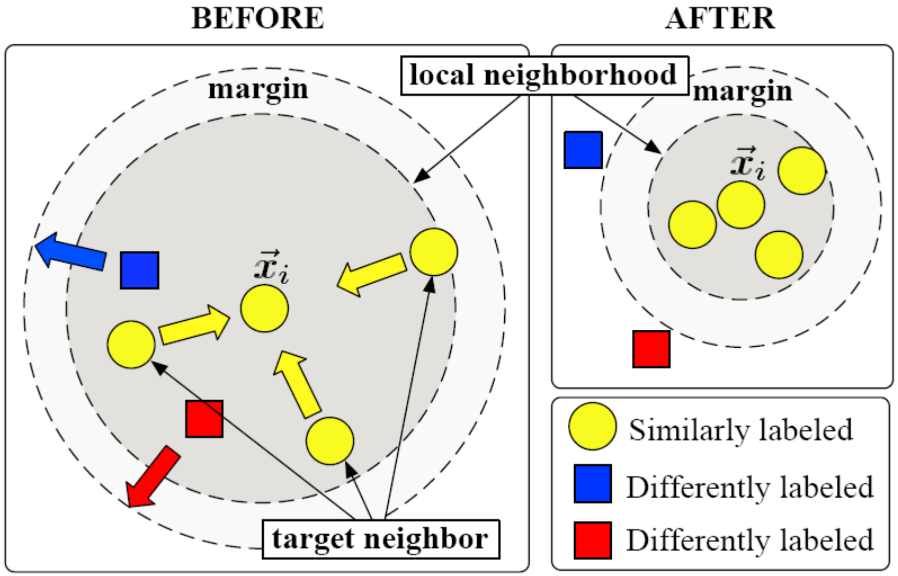
$$d_M(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$$

$$= \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{W}^T \boldsymbol{W} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$$

$$= \left\| \boldsymbol{W}\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{x}_j \right\|_2$$

where $\boldsymbol{M} = \boldsymbol{W}^T \boldsymbol{W}$

Metric learning



- Large Margin Nearest
  - Cost function

$$\epsilon(\mathbf{L}) = \sum_{ij} \eta_{ij} \left\| \mathbf{L}(\boldsymbol{x}_i - \boldsymbol{x}_j) \right\|^2$$

$$+ c \sum_{ijl} \eta_{ij} (1 - y_{il}) \left[1 + \left\| \mathbf{L}(\boldsymbol{x}_i - \boldsymbol{x}_j) \right\|^2 - \left\| \mathbf{L}(\boldsymbol{x}_i - \boldsymbol{x}_l) \right\|^2 \right]_+$$

  - Objective function: semidefinite programming (SDP)

Minimize $\sum_{ij} \eta_{ij} (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \mathbf{M} (\boldsymbol{x}_i - \boldsymbol{x}_j) + c \sum_{ijl} \eta_{ij} (1 - y_{il}) \xi_{ijl}$ **subject to**

(1) $(\boldsymbol{x}_i - \boldsymbol{x}_l)^T \mathbf{M} (\boldsymbol{x}_i - \boldsymbol{x}_l) - (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \mathbf{M} (\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 1 - \xi_{ijl}$

(2) $\xi_{ijl} \geq 0$

(3) $\mathbf{M} \succcurlyeq 0$

[1] K. Q. Weinberger, J. Blitzer and L. K. Saul. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *NIPS 2005.*

Metric learning

- Information-Theoretic Metric Learning (ITML)
  - Distance metric learning problem

$$\begin{aligned} \min_{A} \quad & \mathrm{KL}\left(p(\boldsymbol{x}; A_0) \parallel p(\boldsymbol{x}; A)\right) \\ \text{subject to} \quad & d_A\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \le u \quad (i,j) \in S, \\ & d_A\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \ge l \quad (i,j) \in D. \end{aligned}$$

where $\mathrm{KL}\left(p(\boldsymbol{x}; A_0) \parallel p(\boldsymbol{x}; A)\right) = \int p(\boldsymbol{x}; A_0) \log \frac{p(\boldsymbol{x}; A_0)}{p(\boldsymbol{x}; A)} dx$

  - Optimization problem can be reformulated as

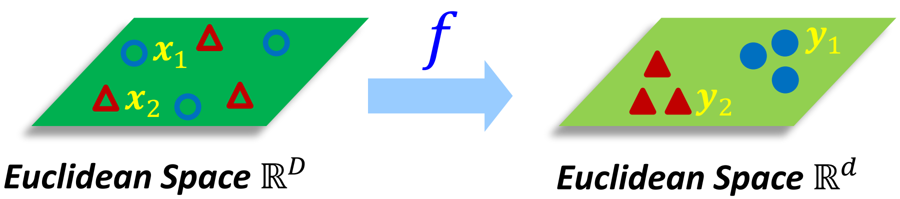
$$\begin{aligned} \min_{A \succcurlyeq 0} \quad & D_{ld}(A, A_0) \\ \text{s.t.} \quad & \mathrm{tr}\left(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T\right) \le u \quad (i,j) \in S, \\ & \mathrm{tr}\left(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T\right) \ge l \quad (i,j) \in D. \end{aligned}$$

where $D_{ld}(A, A_0) = \mathrm{tr}(AA_0^{-1}) - \log\det(AA_0^{-1}) - n$

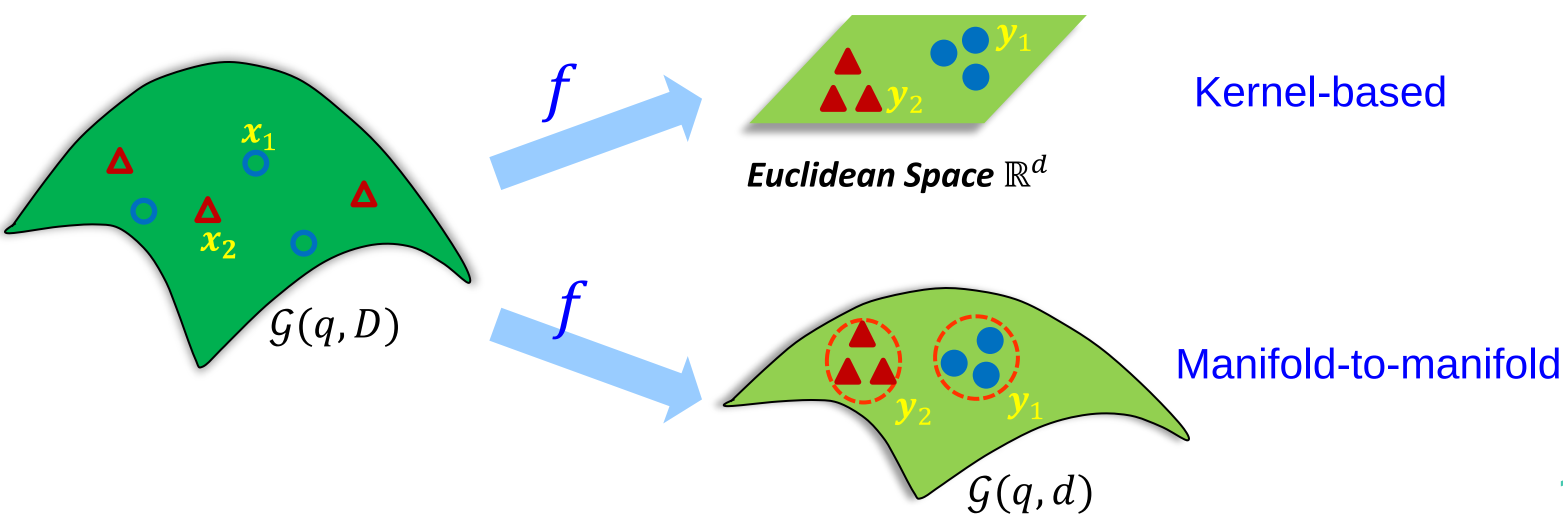
[1] J.V. Davis, B. Kulis, P. Jain, S. Sra, and I.S. Dhillon. Information-Theoretic Metric Learning. *ICML 2007.*

Metric learning: linear vs. nonlinear

- Linear
  - $f: \boldsymbol{x} \rightarrow \boldsymbol{y}, \boldsymbol{y} = W\boldsymbol{x}\ (\boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{y} \in \mathbb{R}^d, W \in \mathbb{R}^{d \times D})$

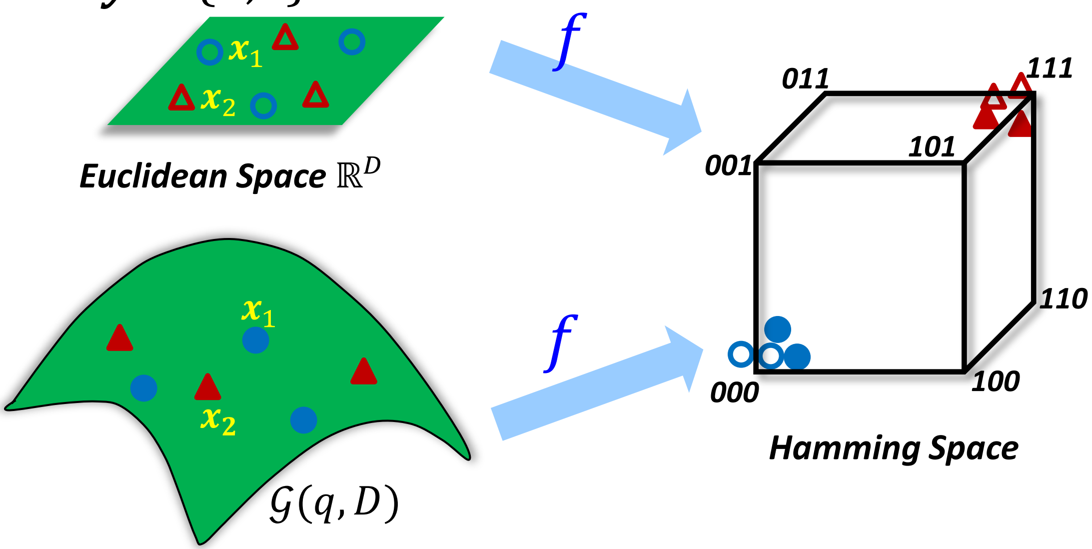


- Nonlinear
  - $f(.)$ is nonlinear mapping, or $\boldsymbol{x}, \boldsymbol{y}$ is in non-Euclidean space
  - Riemannian metric learning
    - $\boldsymbol{x} \in M$ is element on some Riemannian manifold M



Metric learning: linear vs. nonlinear

- Linear
  - $f: \boldsymbol{x} \rightarrow \boldsymbol{y}, \boldsymbol{y} = W\boldsymbol{x}$ $(\boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{y} \in \mathbb{R}^d, W \in \mathbb{R}^{d \times D})$
- Nonlinear
  - $f(.)$ is nonlinear mapping, or $\boldsymbol{x}$, $\boldsymbol{y}$ is in non-Euclidean space
  - Riemannian metric learning
    - $\boldsymbol{x} \in M$ is element on some Riemannian manifold M
  - Hash learning (*a.k.a.* binary code learning)
    - $\boldsymbol{y} \in \{0,1\}^K$ is element in *K*-dimensional Hamming space



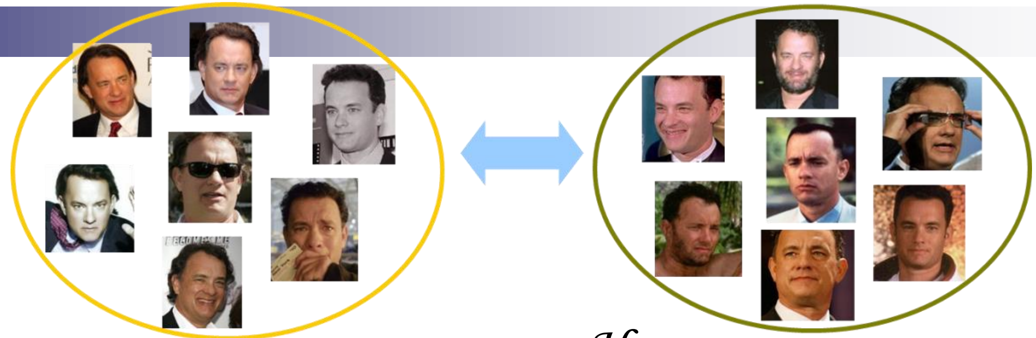
中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

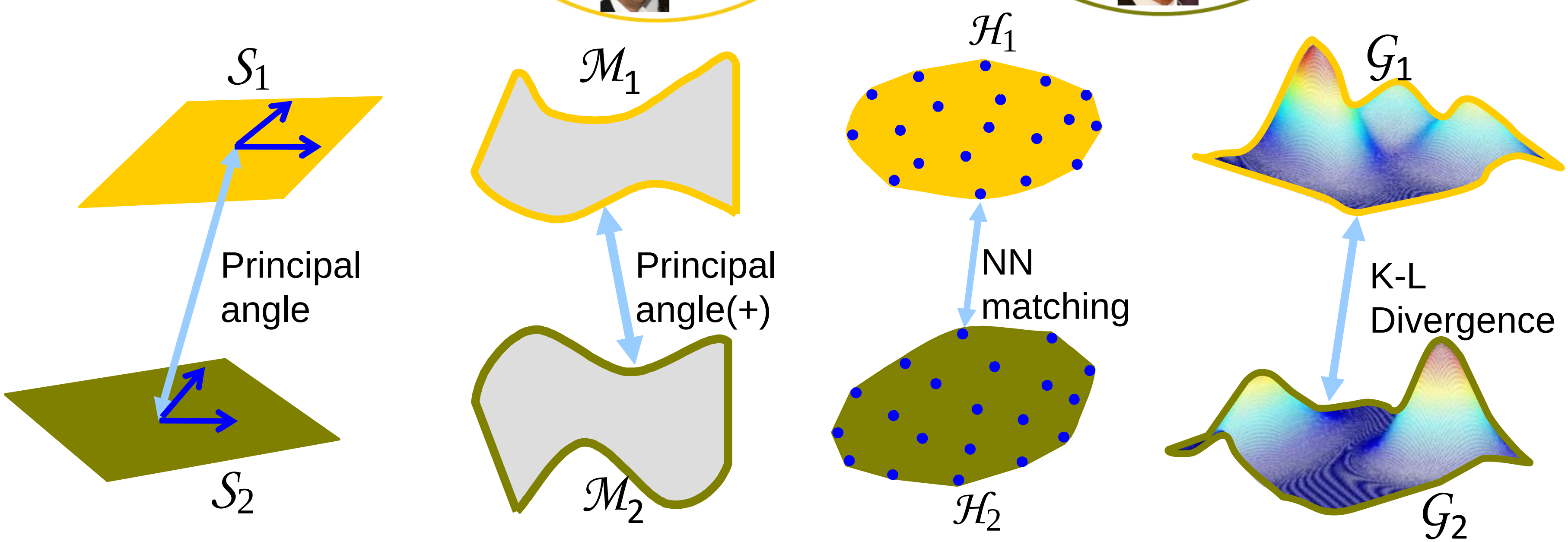
Outline

- Background
- Literature review
- Evaluations
- Summary

Task 1: Image set classification

From the view of set modeling





◆Linear subspace
[Yamaguchi, FG'98]
[Kim, PAMI'07]
[Hamm, ICML'08]
[Harandi, CVPR'11]
[Huang, CVPR'15]

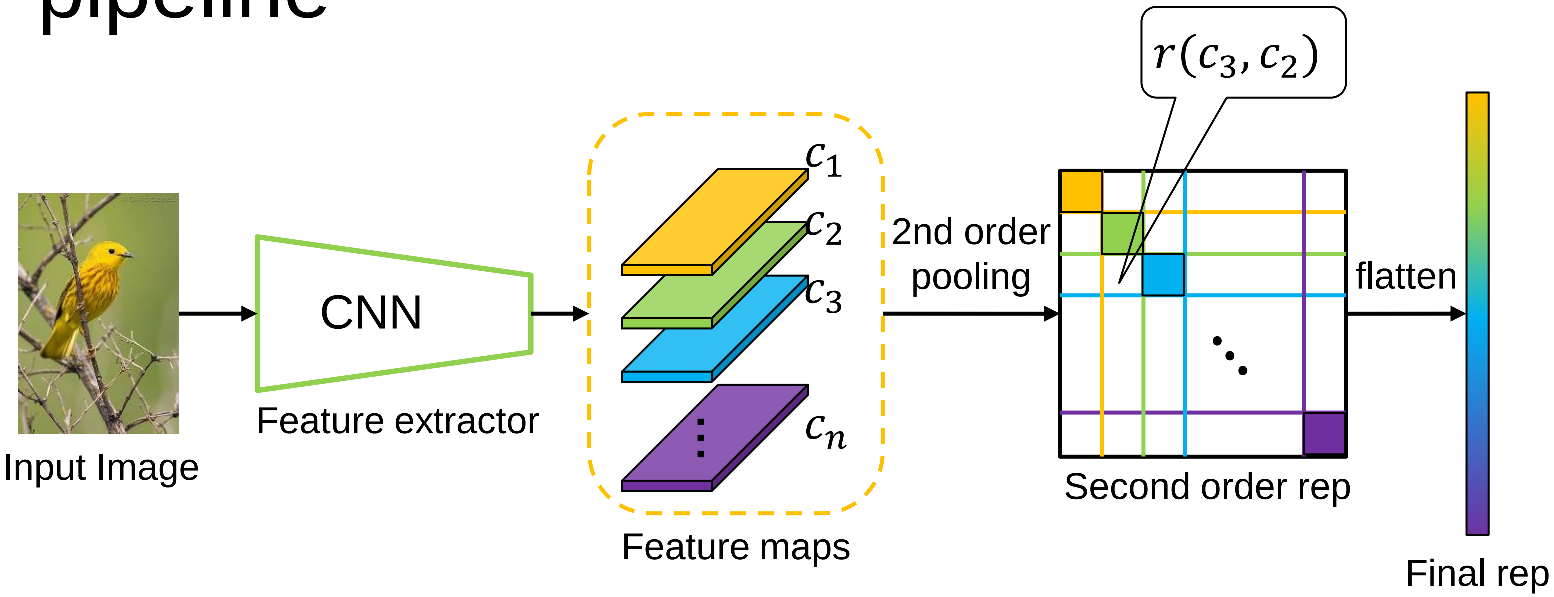
◆Nonlinear manifold
[Hadid, FG'04]
[Kim, BMVC'05]
[Wang, CVPR'08/09]
[Chen, CVPR'13]
[Lu, CVPR'15]

◆Affine/Convex hull
[Cevikalp, CVPR'10]
[Hu, CVPR'11]
[Yang, FG'13]
[Zhu, ICCV'13]
[Wang, ACCV'16]

◆Statistics
[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]
[Wang, CVPR'12]
[Harandi, ECCV'14/ICCV'15]
[Wang, CVPR'15/CVPR'17]

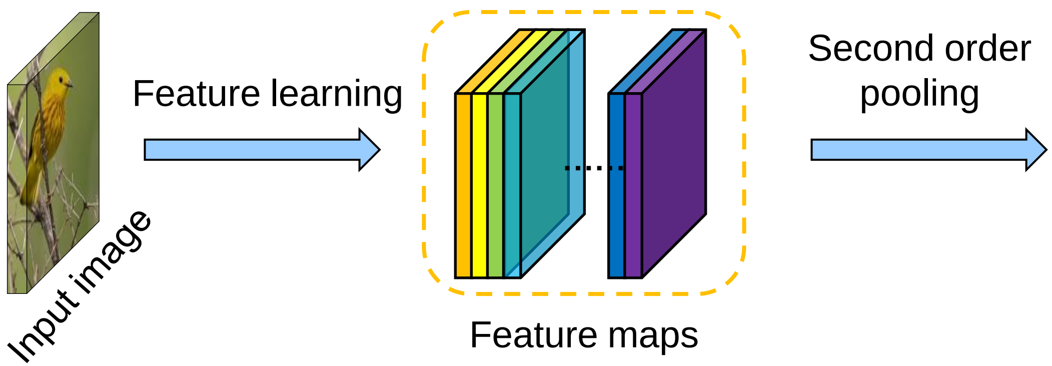
Task 2: Image recognition

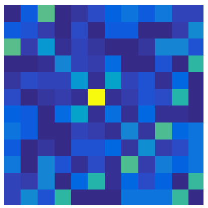
- Second order representation learning pipeline



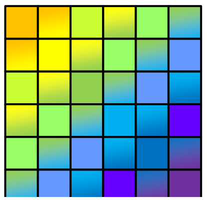
Task 2: Image recognition

From the view of channel-set modeling

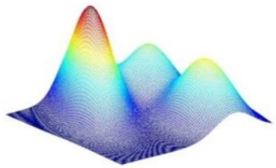




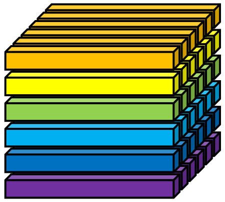
- Bilinear
  [Lin, ICCV’15]
  [Lin, BMVC’17]



- Covariance
  [Li, ICCV’17]
  [Li, CVPR’18]



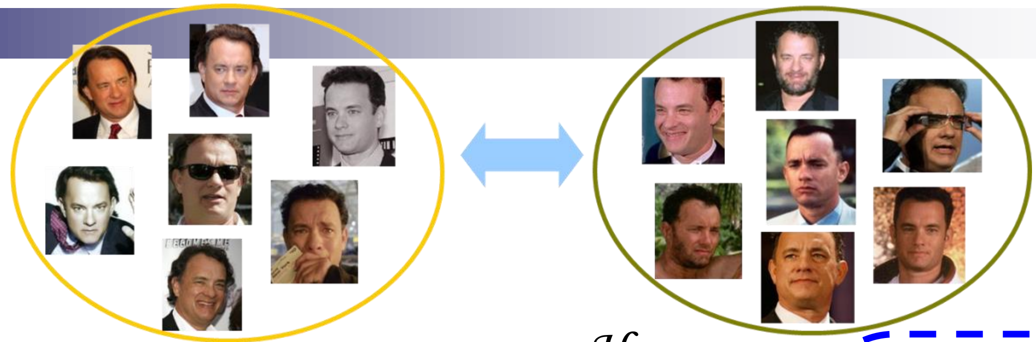
- Gaussian
  [Wang, ICCV’15]

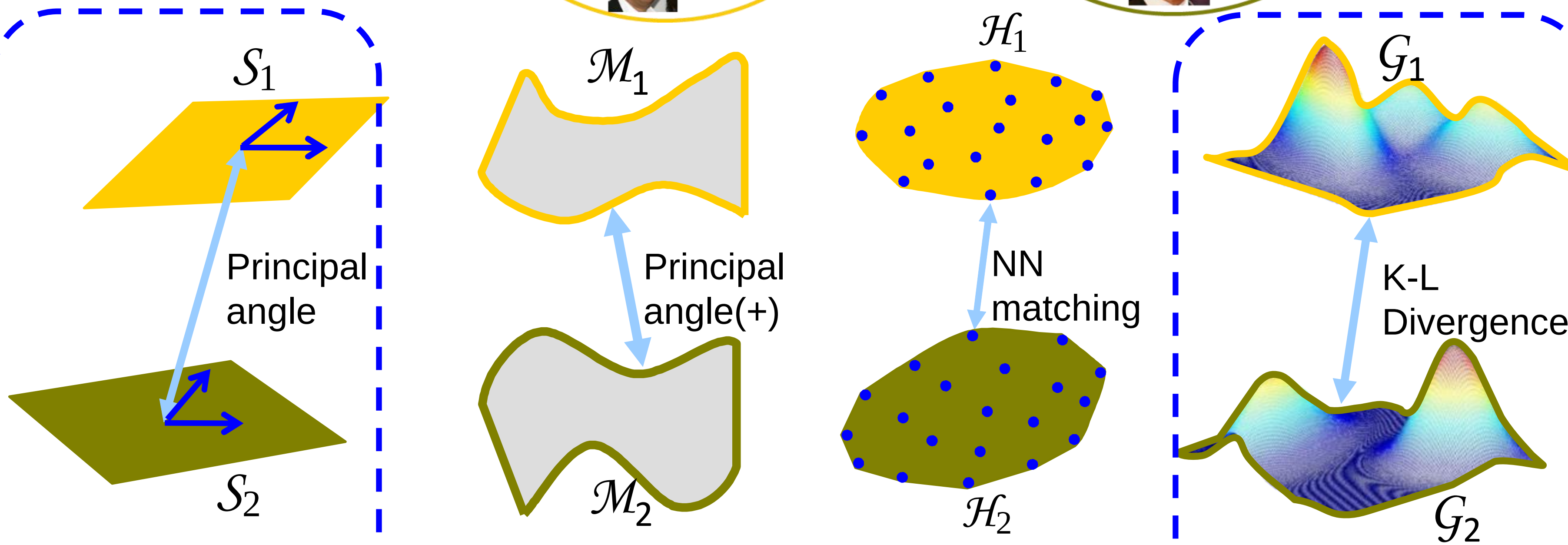


- Tensor sketch
  [Gao, CVPR’16]

Task 1: Image set classification

From the view of set modeling





◆Linear subspace
- [Yamaguchi, FG'98]
- [Kim, PAMI'07]
- [Hamm, ICML'08]
- [Harandi, CVPR'11]
- [Huang, CVPR'15]

◆Nonlinear manifold
- [Hadid, FG'04]
- [Kim, BMVC'05]
- [Wang, CVPR'08/09]
- [Chen, CVPR'13]
- [Lu, CVPR'15]

◆Affine/Convex hull
- [Cevikalp, CVPR'10]
- [Hu, CVPR'11]
- [Yang, FG'13]
- [Zhu, ICCV'13]
- [Wang, ACCV'16]

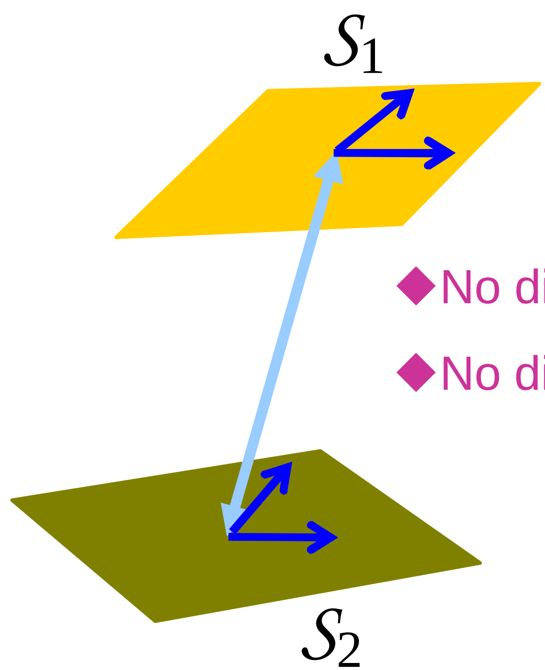
◆Statistics
- [Shakhnarovich, ECCV'02]
- [Arandjelović, CVPR'05]
- [Wang, CVPR'12]
- [Harandi, ECCV'14/ICCV'15]
- [Wang, CVPR'15/CVPR'17]

Overview of previous works

- Set modeling
  - Linear subspace→Nonlinear manifold
  - Affine/Convex Hull (affine subspace)
  - Parametric PDFs → high-order 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]
  - LieNet [CVPR'17]



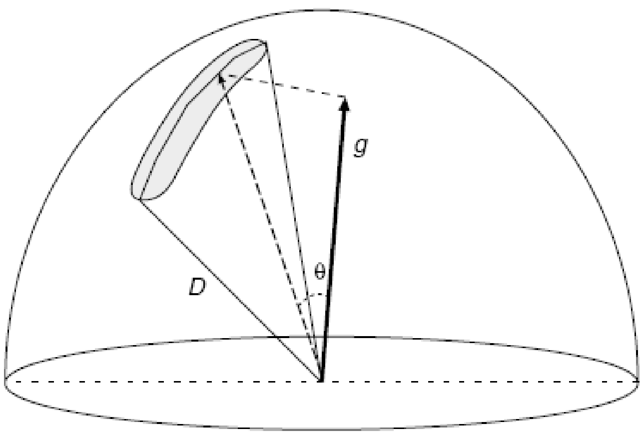
◆No distribution boundary

◆No direction difference

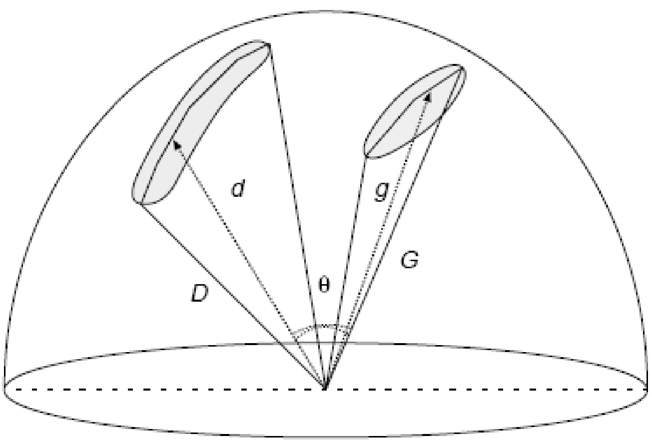
Set model I: linear subspace

- 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

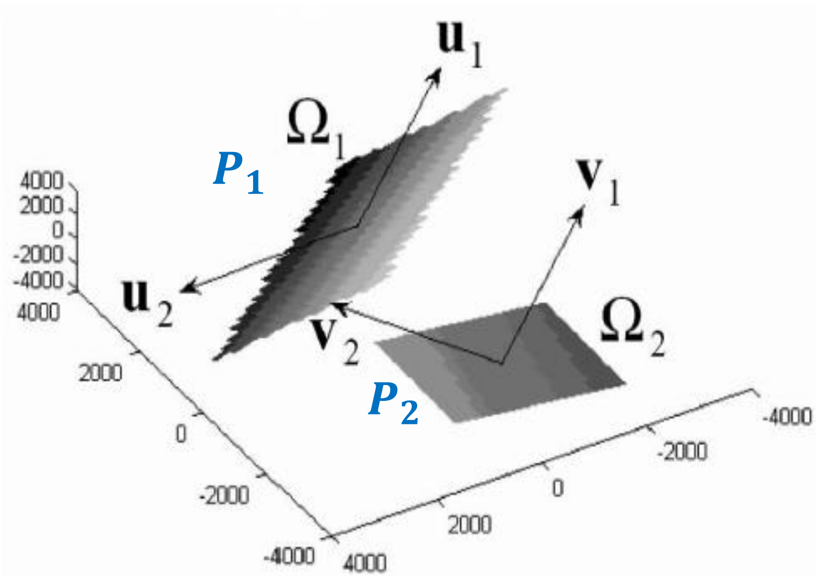
[1] O. Yamaguchi, K. Fukui, and K. Maeda. Face Recognition Using Temporal Image Sequence. *IEEE FG 1998.*

Set model I: linear subspace

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

Set 1: $\boldsymbol{X}_1$

Set 2: $\boldsymbol{X}_2$



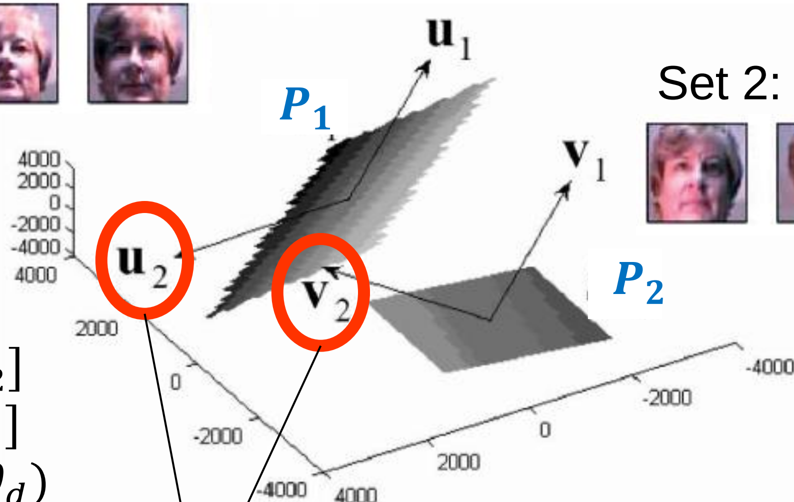
Linear subspace by:
orthonormal basis matrix
$\boldsymbol{X}_i\boldsymbol{X}_i^T \simeq \boldsymbol{P}_i\boldsymbol{\Lambda}_i\boldsymbol{P}_i^T$

[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: X_1



Set 2: X_2

$$P_1^T P_2 = Q_{12} \Lambda Q_{21}^T$$
$$U = P_1 Q_{12} = [u_1, \ldots, u_2]$$
$$V = P_2 Q_{21} = [v_1, \ldots, v_2]$$
$$\Lambda = diag(\cos\theta_1, \ldots, \cos\theta_d)$$

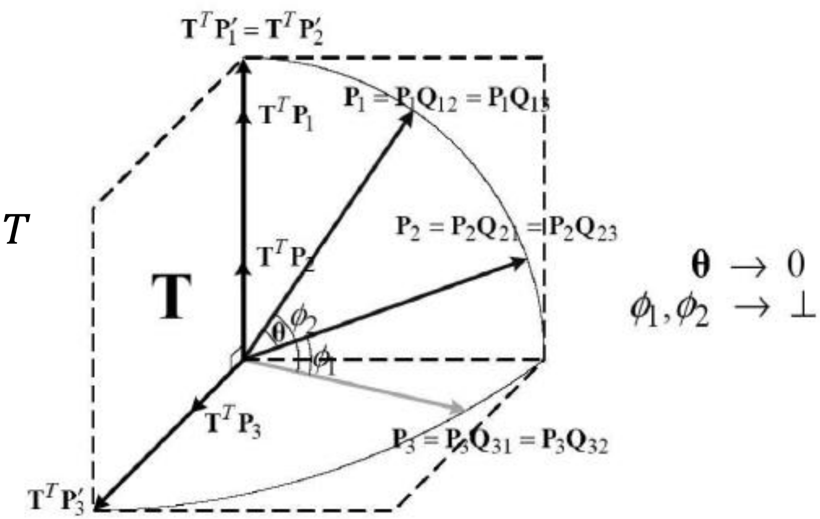
Canonical Correlation: $\cos\theta_i$

Principal Angles: θ_i

Canonical vectors

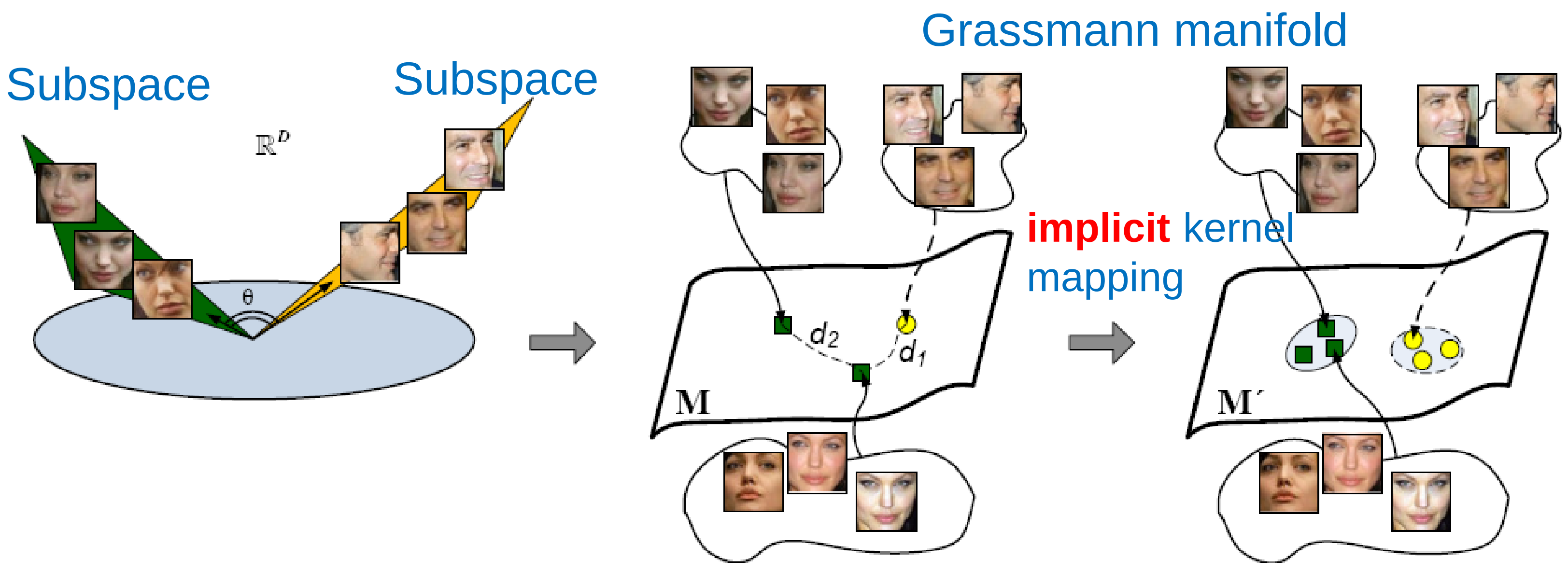
DCC

- 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
    - $F_{ij} = \max_{\boldsymbol{Q}_{ij},\boldsymbol{Q}_{ji}} tr(M_{ij})$
      - $M_{ij} = \boldsymbol{Q}_{ij}^T\boldsymbol{P}'^T_i\boldsymbol{T}\boldsymbol{T}^T\boldsymbol{P}_j^T\boldsymbol{Q}_{ji}$
  - 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*.

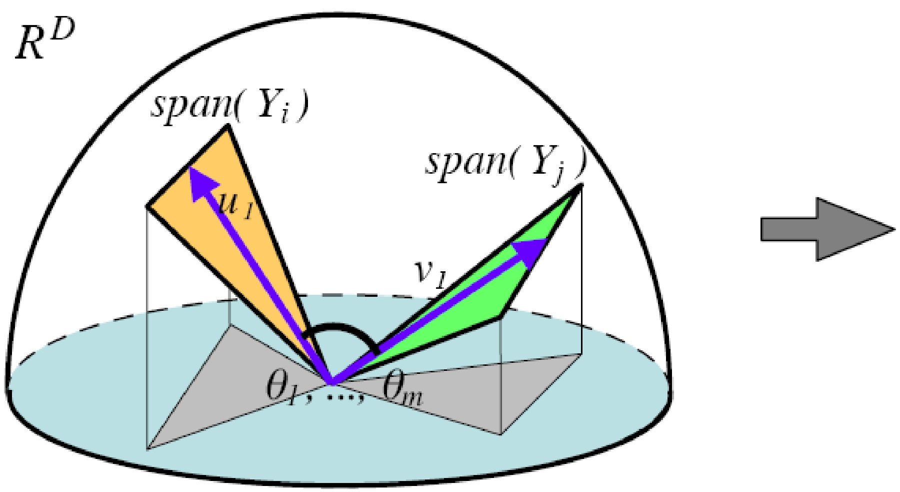
[2] M. Harandi, C. Sanderson, S. Shirazi, B. Lovell. Graph Embedding Discriminant Analysis on Grassmannian Manifolds for Improved Image Set Matching. *IEEE CVPR 2011.*

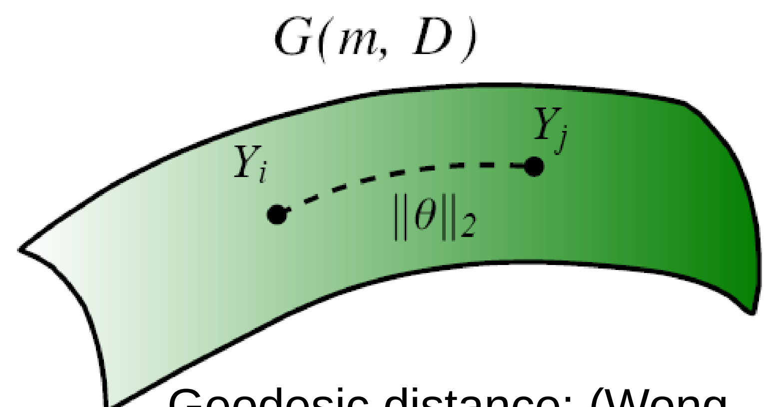
GDA/GGDA

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

- 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)
    - $k_P(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = \left\|\boldsymbol{Y}_1^T\boldsymbol{Y}_2\right\|_F^2$

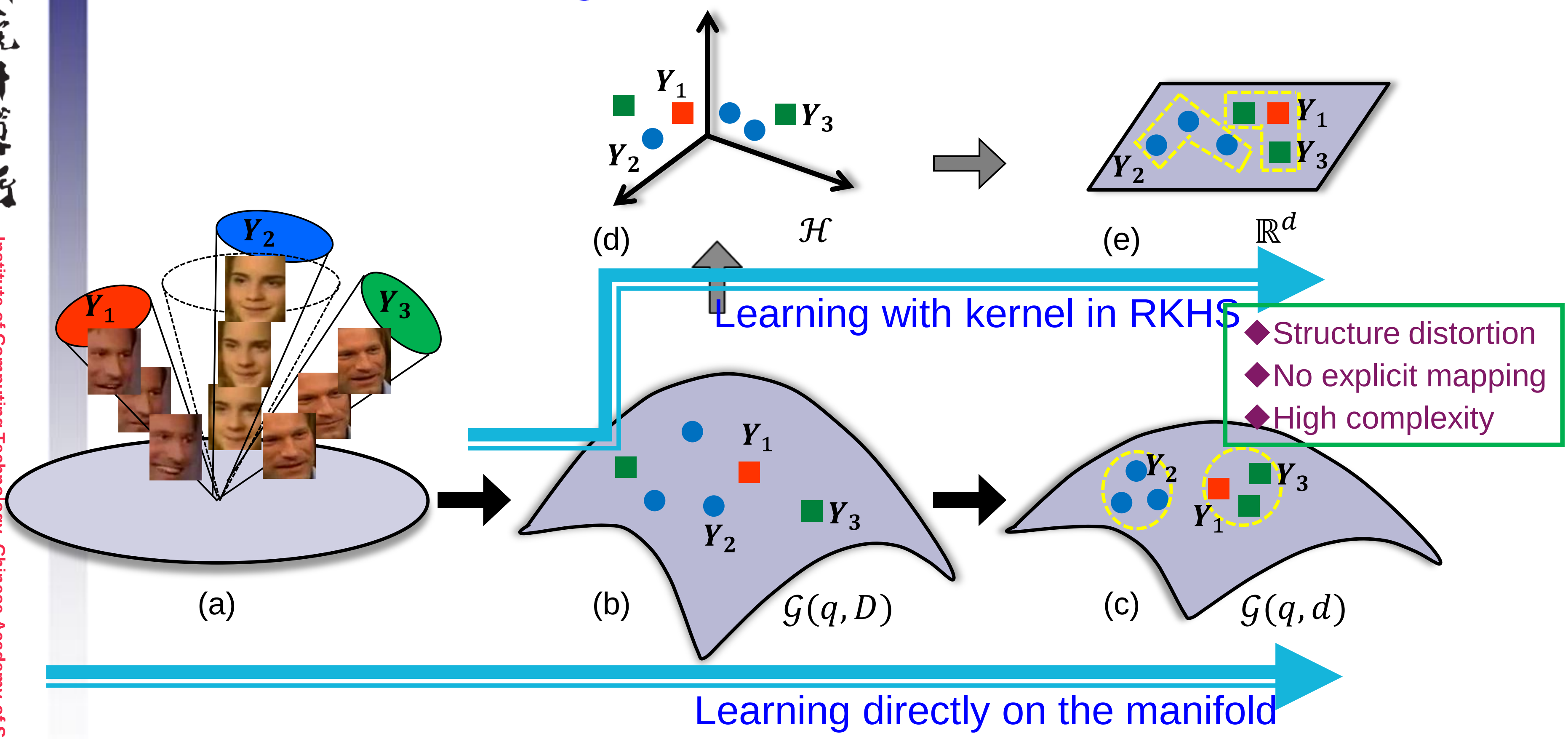
GDA/GGDA

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

Grassmann kernel

Set model I: linear subspace

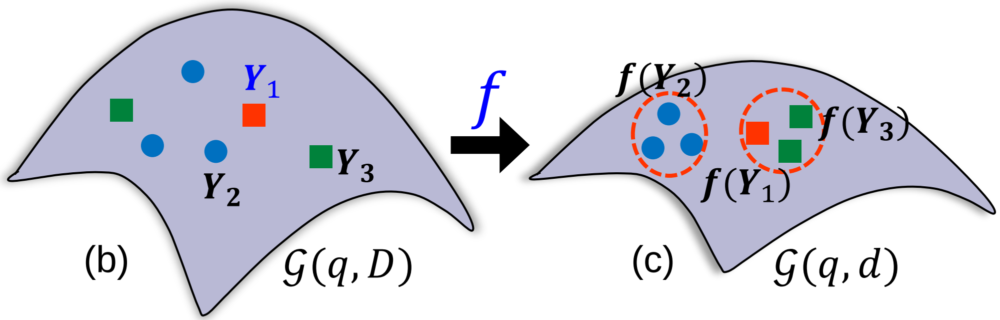
- PML (Projection Metric Learning) [CVPR'15]
  - Metric learning: on Riemannian manifold



[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
  - $f(\boldsymbol{Y}) = \boldsymbol{W}^T\boldsymbol{Y} \in \mathcal{G}(q,d), \boldsymbol{Y} \in \mathcal{G}(q,D),\ d \le D$



- Projection metric on target Grassmann manifold $\mathcal{G}(q,d)$
  - $d_p^2\left(f(\boldsymbol{Y}_\mathrm{i}), f(\boldsymbol{Y}_\mathrm{j})\right) = 2^{-1/2}\left\|(\boldsymbol{W}^T\boldsymbol{Y}_i')(\boldsymbol{W}^T\boldsymbol{Y}_i')^T - (\boldsymbol{W}^T\boldsymbol{Y}_j')(\boldsymbol{W}^T\boldsymbol{Y}_j')^T\right\|_F^2 = 2^{-1/2} tr\left(\boldsymbol{P}^T\boldsymbol{A}_{ij}\boldsymbol{A}_{ij}P\right)$
    - $\boldsymbol{A}_{ij} = (\boldsymbol{Y}_i'\boldsymbol{Y}_i'^T - \boldsymbol{Y}_j'\boldsymbol{Y}_j')^T$, $\boldsymbol{P} = \mathbf{W}\mathbf{W}^T$ 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

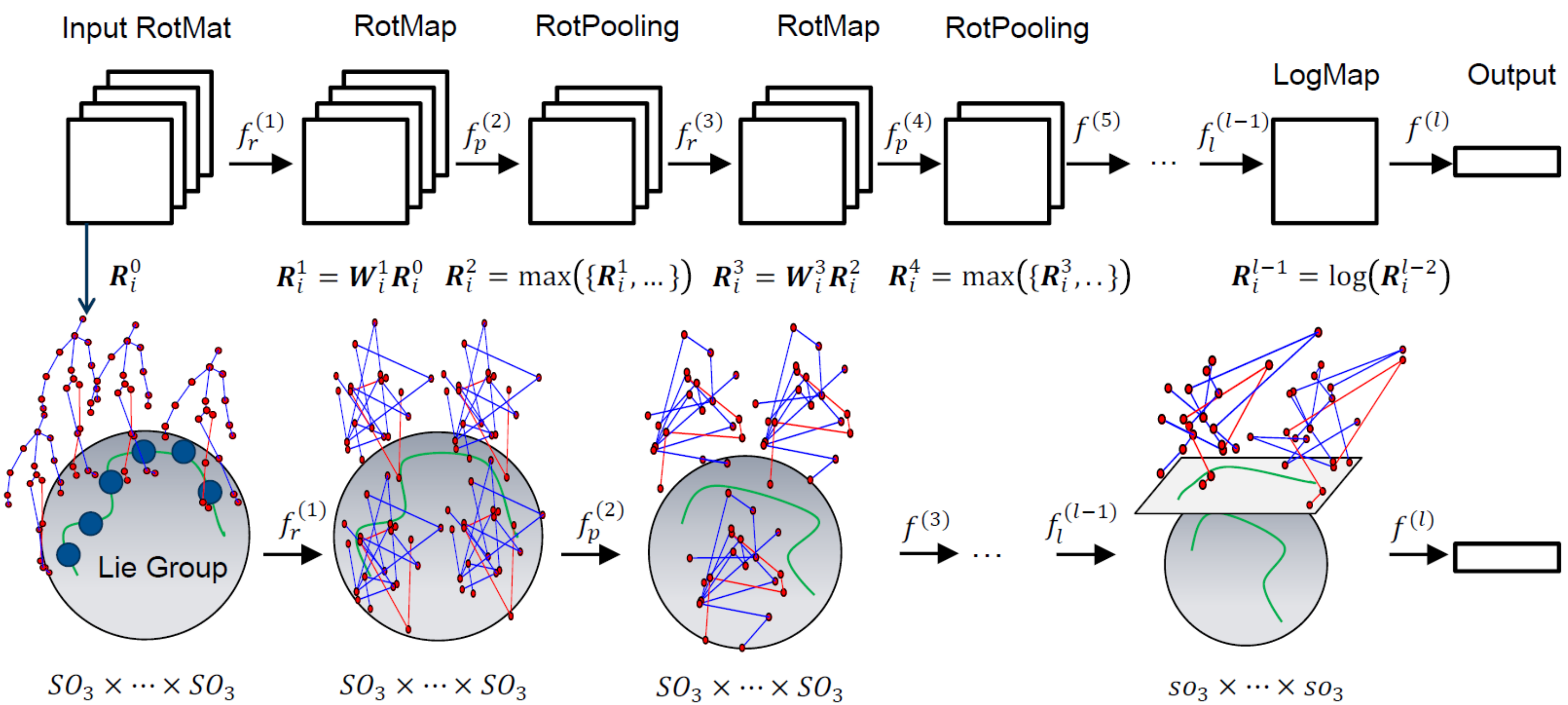
PML

- 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

Set model I: linear subspace

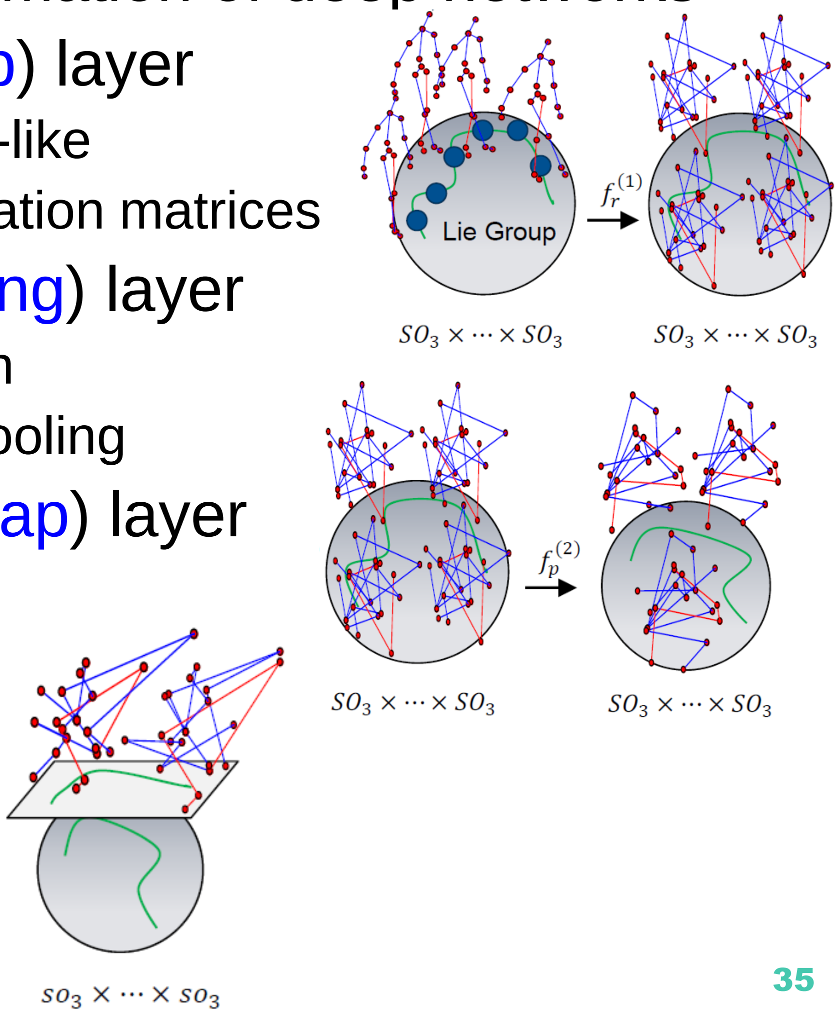
- LieNet [CVPR'17]
  - Metric learning: Lie group nonlinear learning in deep networks



[1] Z. Huang, C. Wan, T. Probst, L. Van Gool. Deep Learning on Lie Groups for Skeleton-based Action Recognition. *IEEE CVPR 2017.*

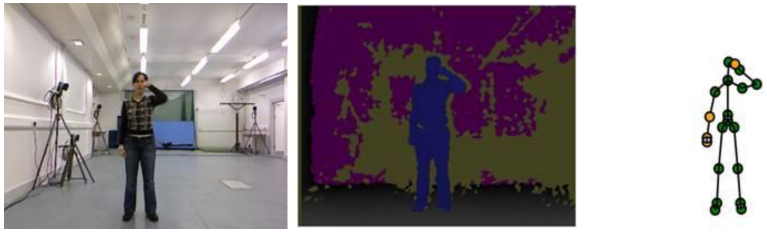
LieNet

- Basic idea: respect manifold property of Lie group structure in nonlinear transformation of deep networks
  - Rotation mapping (RotMap) layer
    - Fully connected convolution-like
    - Transform and align the rotation matrices
  - Rotation pooling (RotPooling) layer
    - Reduce Lie group dimension
    - Both spatial and temporal pooling
  - Logarithm mapping (LogMap) layer
    - Manifold to Euclidean space

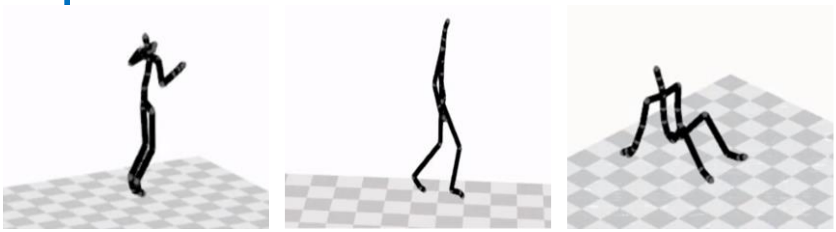


Evaluations

- G3D-Gaming [Bloom, CVPR workshop'12]
  - 20 motions, 663 sequences
  - 3D coordinates of 20 joints



- HDM05 [Muller, Tech. Rep.'07]
  - 130 actions, 2,337 sequences
  - 3D coordinates of 31 joints



- NTU RGB+D [Shahroudy, CVPR'16]
  - 60 actions, 56,000 sequences
  - 3D coordinates of 25 joints

Evaluations

- Performance comparisons

Method	G3D-Gaming
RBM+HMM [32]	86.40%
SE [41]	87.23%
SO [42]	87.95%
LieNet-0Block	84.55%
LieNet-1Block	85.16%
LieNet-2Blocks	86.67%
LieNet-3Blocks	**89.10%**

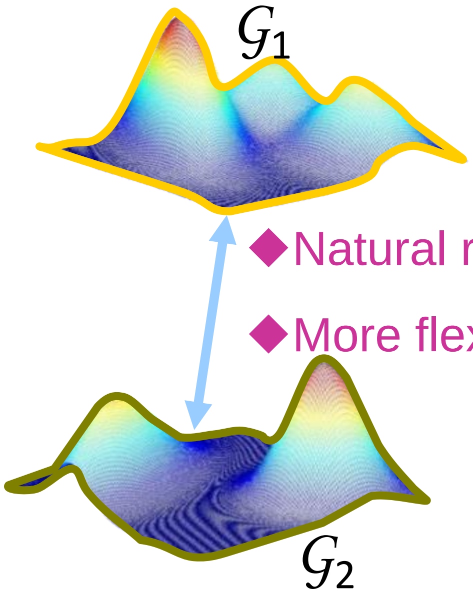
Method	HDM05
SPDNet [18]	61.45%±1.12
SE [41]	70.26%±2.89
SO [42]	71.31%±3.21
LieNet-0Block	71.26%±2.12
LieNet-1Block	73.35%±1.14
LieNet-2Blocks	**75.78%±2.26**

Method	RGB+D-subject	RGB+D-view
HBRNN [13]	59.07%	63.97%
Deep RNN [37]	56.29%	64.09%
Deep LSTM [37]	60.69%	67.29%
PA-LSTM [37]	62.93%	70.27%
ST-LSTM [26]	**69.2%**	**77.7%**
SE [41]	50.08%	52.76%
SO [42]	52.13%	53.42%
LieNet-0Block	53.54%	54.78%
LieNet-1Block	56.35%	60.14%
LieNet-2Blocks	58.02%	62.52%
LieNet-3Blocks	61.37%	66.95%

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]
  - B. Gauss [ICCV'15]
  - SPD-ML [ECCV'14]
  - LEML [ICML'15]
  - DCRL [CVPR'17]
  - SPDNet [AAAI'17]
  - DHH [TIP'19]

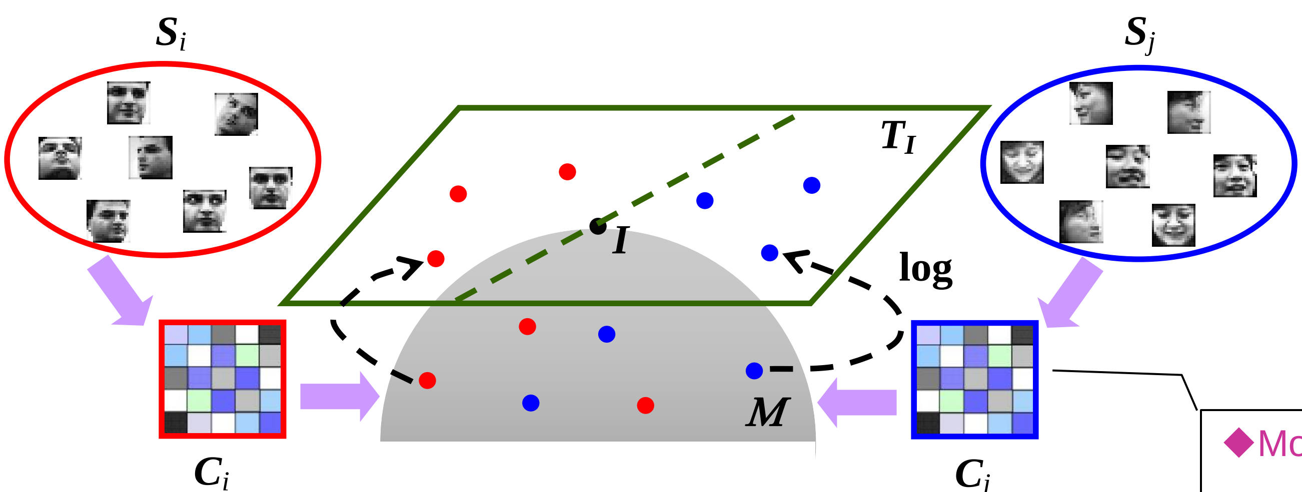


[Shakhnarovich, ECCV'02]
[Arandjelović, CVPR'05]

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



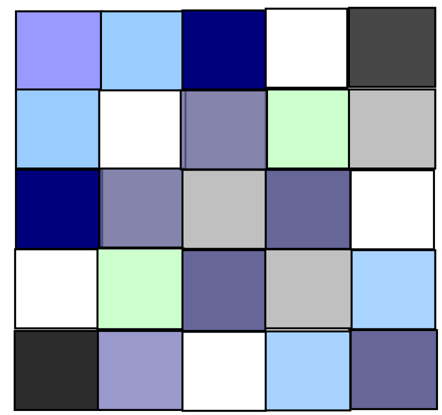
◆Model data variations

◆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*.

CDL

- 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.

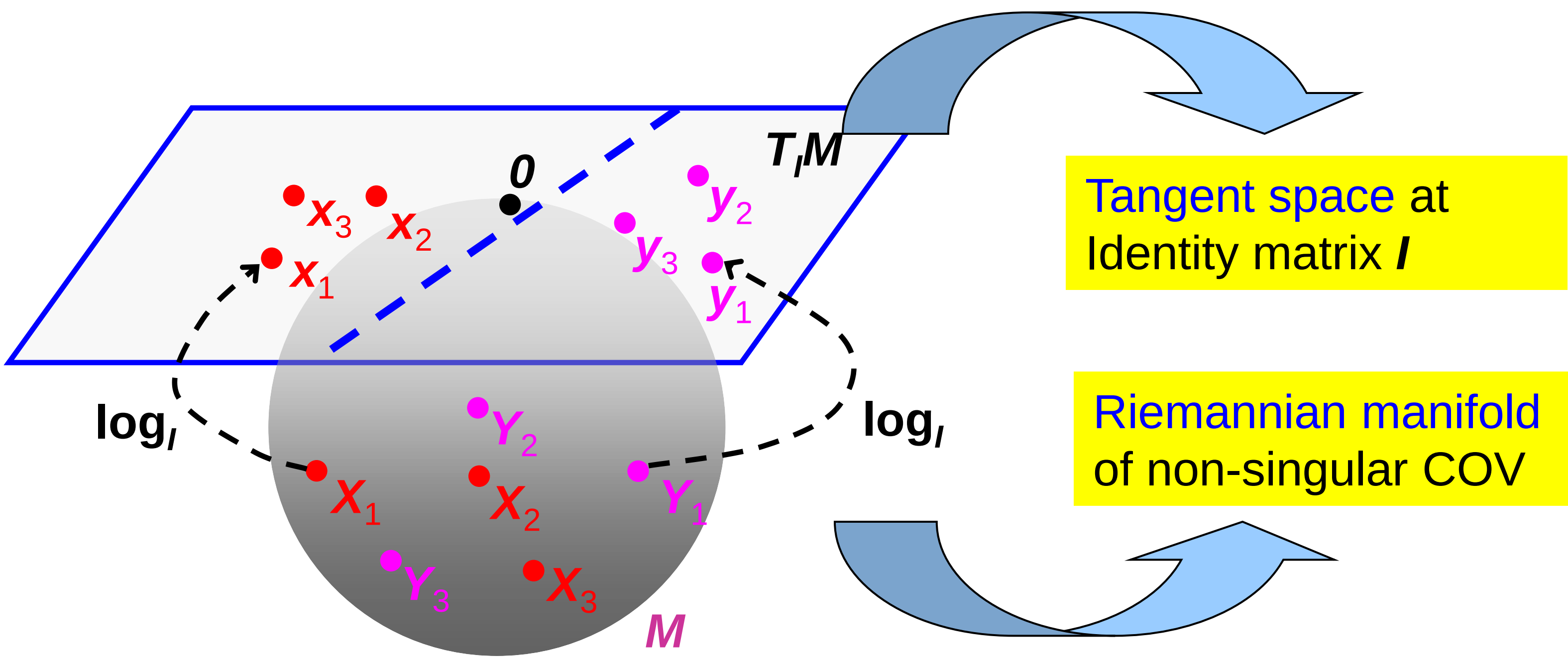
CDL

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

$$\Psi_{log} : \ \boldsymbol{C} \rightarrow \log_{\boldsymbol{I}}(\boldsymbol{C}), \ \ (M \mapsto R^{d\times d})$$

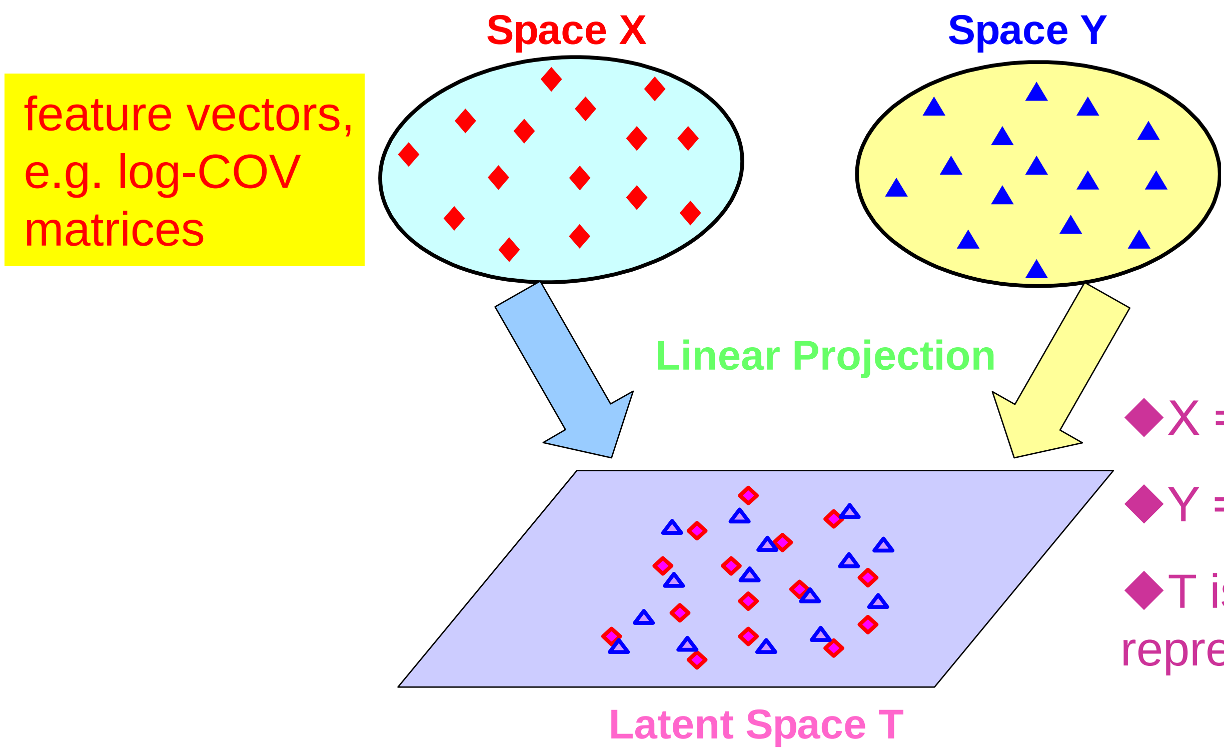
$$k_{log}(\boldsymbol{C}_1, \boldsymbol{C}_2) = trace[\log_{\boldsymbol{I}}(\boldsymbol{C}_1)\cdot \log_{\boldsymbol{I}}(\boldsymbol{C}_2)]$$

Mercer's theorem



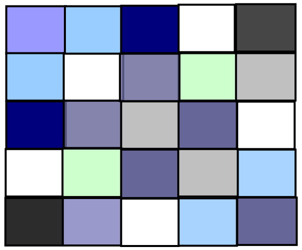
CDL

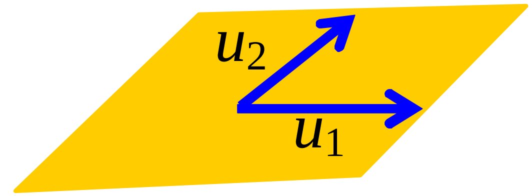
- Discriminative learning on COV manifold
  - Partial Least Squares (PLS) regression
  - Goal: Maximize the covariance between observations and class labels



- X = T * P’
- Y = T * B * C’ = X * Bpls
- T is the common latent representation

CDL vs. GDA





- COV→SPD manifold
  - Model

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

  - Metric

$$d(\boldsymbol{C}_1, \boldsymbol{C}_2) = \left\| \log_{\boldsymbol{I}}(\boldsymbol{C}_1) - \log_{\boldsymbol{I}}(\boldsymbol{C}_2) \right\|_F$$

$$\Rightarrow \log_{\boldsymbol{I}}(\boldsymbol{C}) = \boldsymbol{U} \log_{\boldsymbol{I}}(\boldsymbol{\Lambda}) \boldsymbol{U}^T$$

  - Kernel

$$\Psi_{log} : \ \boldsymbol{C} \to \log_{\boldsymbol{I}}(\boldsymbol{C}), \ (M \mapsto R^{d \times d})$$

- Subspace→Grassmannian
  - Model

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$$

$$\Rightarrow \boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m]_{D \times m}{}^{*}$$

  - Metric

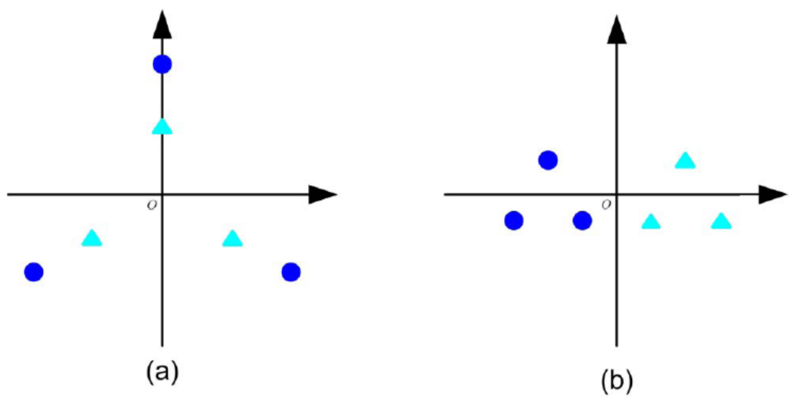
$$d_{proj}(\boldsymbol{U}_1, \boldsymbol{U}_2) = 2^{-1/2} \left\| \boldsymbol{U}_1\boldsymbol{U}_1^T - \boldsymbol{U}_2\boldsymbol{U}_2^T \right\|_F$$

  - Kernel

$$\Psi_{proj} : \ \boldsymbol{U} \to \boldsymbol{U}\boldsymbol{U}^T, \ \ G(m, D) \to R^{d \times d}$$

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

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

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

Objective function

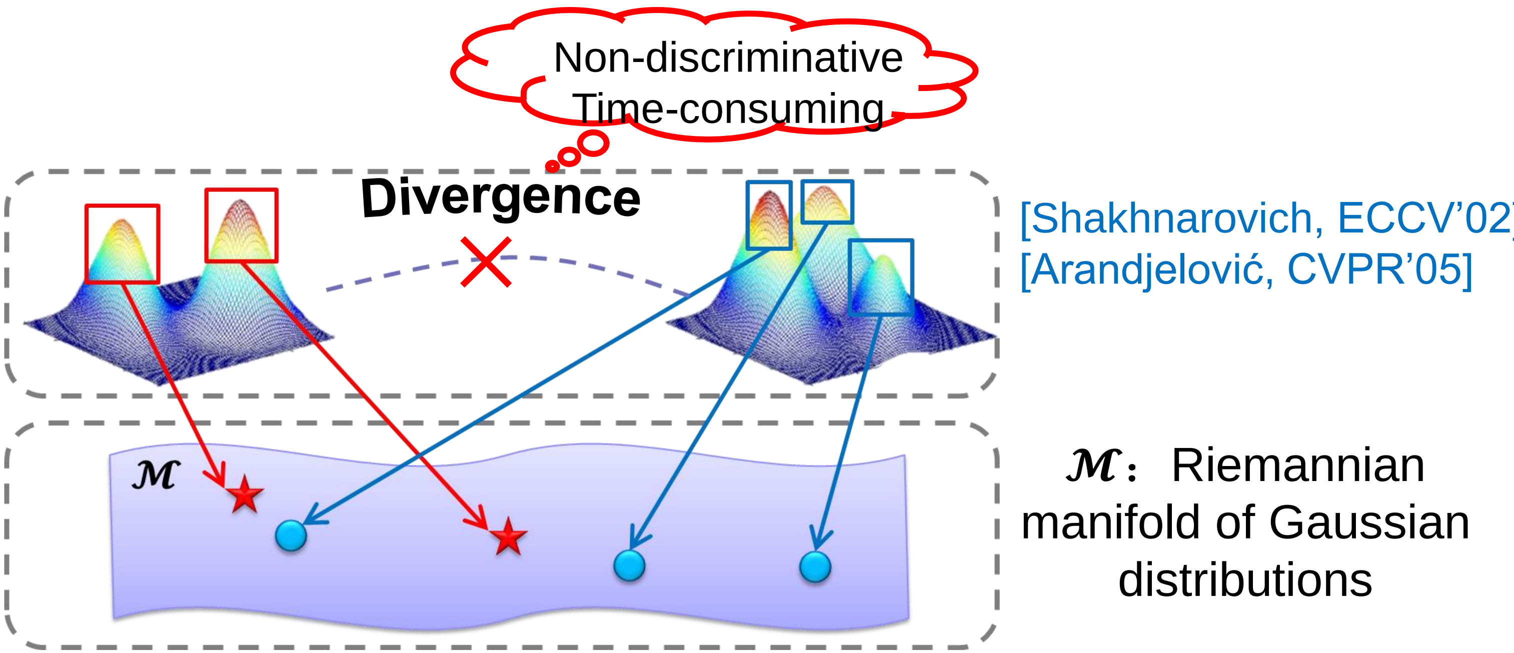
$$d(S_i, S_j) = \sum_{p=1}^{P} \eta_p(\phi_i^p)(\phi_i^p - \phi_j^p)^T M(\phi_i^p - \phi_j^p)\eta_p(\phi_j^p)$$

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

[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+)

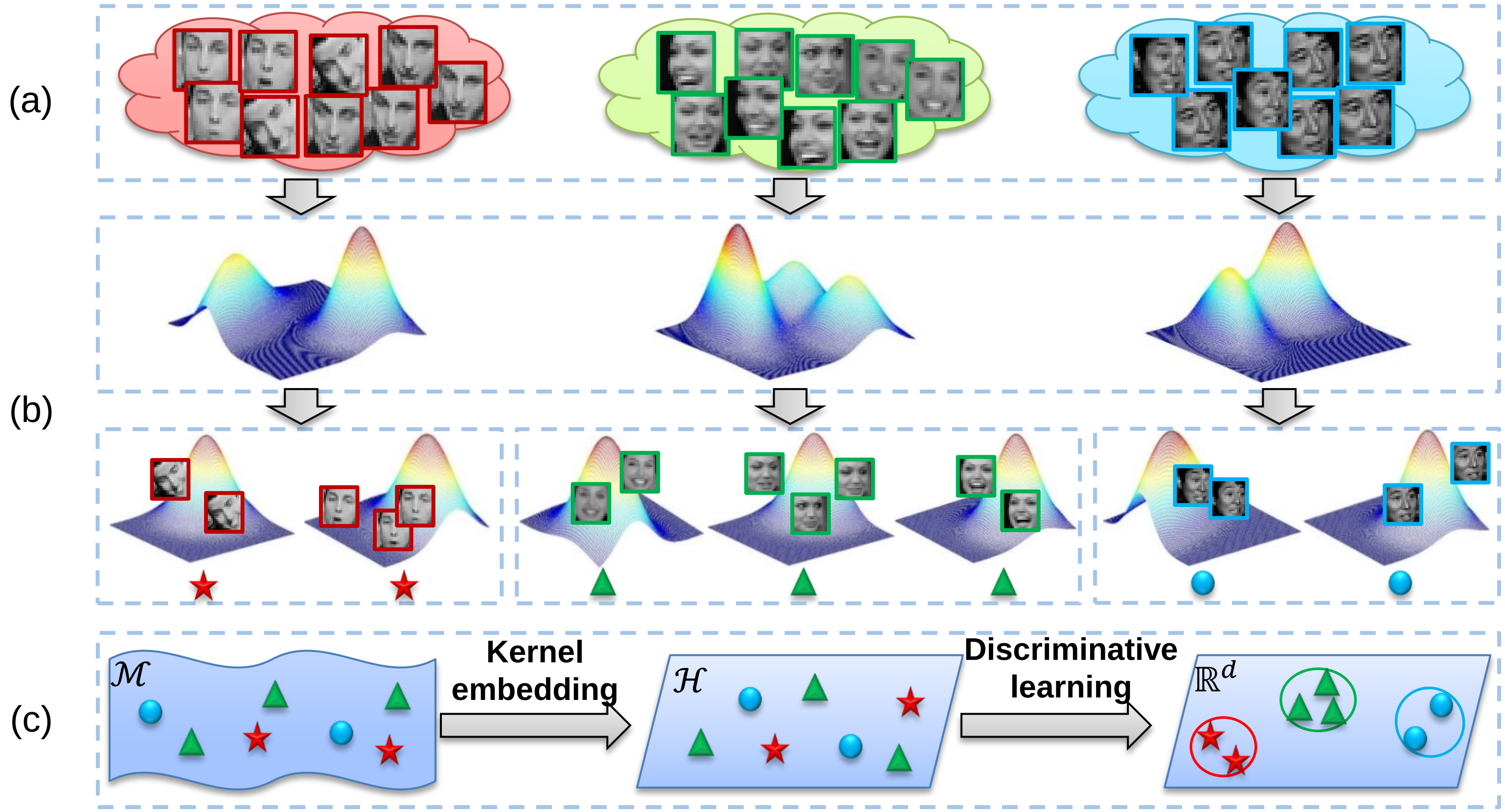
- DARG (Discriminant Analysis on Riemannian manifold of Gaussian distributions) [CVPR'15]
  - Set modeling by mixture of Gaussian distribution (GMM)
    - Naturally encode the 1st order and 2nd order statistics
  - Metric learning: on Riemannian manifold



[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

DARG

- 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

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

    - 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

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

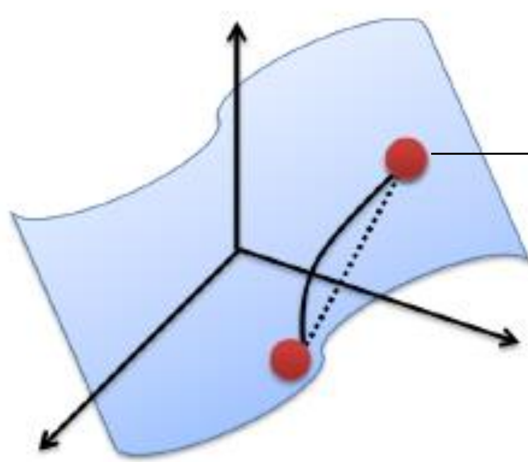
$$\boldsymbol{W} = \sum_{i=1}^{C} \frac{1}{\omega_i} \sum_{j=1}^{N_i} w_j^{(i)} (k_j^i - m_i)(k_j^i - m_i)^T$$

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

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

Set model IV: statistics (COV+)

- Beyond Gauss [ICCV'15]
  - Set modeling by probability distribution functions (PDFs)
    - More general than Gaussian assumption
    - non-parametric, data-driven kernel density estimator (KDE)
  - Metric learning: on Riemannian manifold



◆PDFs form a Riemannian manifold, i.e., the **statistical manifold**.

◆Csiszár ***f*-divergences** are exploited to measure the geodesic distance.

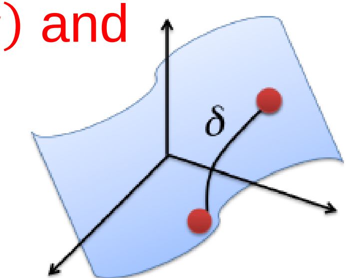
[1] M. Harandi, M. Salzmann, and M. Baktashmotlagh. Beyond Gauss: Image-Set Matching on the Riemannian Manifold of PDFs. *IEEE ICCV 2015*.

Beyond Gauss

- Set modeling with PDFs
  - ***Kernel Density Estimation (KDE)***

$$\hat{p}(x) = \frac{1}{n\sqrt{\det(2\pi\Sigma)}}\sum_{i=1}^{n}\exp\left(-\frac{1}{2}(x-x_i)^T\Sigma^{-1}(x-x_i)\right)$$

  - Given two image sets $\{x_i^p\}_{i=1}^{n_p}$ and $\{x_i^p\}_{i=1}^{n_p}$ with estimated PDFs $p(x)$ and $q(x)$, how to compare two PDFs $p(x)$ and $q(x)$?



- Empirical estimation of *f*-Divergences
  - ***Hellinger distance***

$$\hat{\delta}_H^2(p||q) = \frac{1}{n_p}\sum_{i}^{n_p}\left(\sqrt{T(x_i^p)} - \sqrt{1-T(x_i^p)}\right)^2 + \frac{1}{n_q}\sum_{i}^{n_q}\left(\sqrt{T(x_i^q)} - \sqrt{1-T(x_i^q)}\right)^2$$

  - ***Jeffrey divergence***

$$\hat{\delta}_H^2(p||q) = \frac{1}{n_p}\sum_{i}^{n_p}\left(2T(x_i^p)-1\right)\ln\frac{T(x_i^p)}{1-T(x_i^p)} + \frac{1}{n_q}\sum_{i}^{n_q}\left(2T(x_i^q)-1\right)\ln\frac{T(x_i^q)}{1-T(x_i^q)}$$

$$T(x) = \frac{p(x)}{p(x)+q(x)}$$

Beyond Gauss

中科院计算所

Institute of Computing Technology, Chinese Academy of Sciences

- Kernels on the Statistical Manifold
  - Hellinger Kernel

$$K_H(p,q) = \exp\left(-\sigma\delta_H^2(p,q)\right)$$

  - Laplace Kernel

$$K_L(p,q) = \exp\left(-\sigma\delta_H(p,q)\right)$$

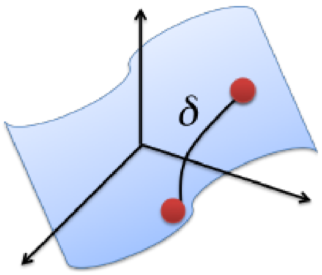
  - Jeffrey Kernel

$$K_J(p,q) = \exp\left(-\sigma\delta_J(p||q)\right)$$

- Dimensionality Reduction

$$W^* = \underset{W}{\operatorname{argmin}}\, L(W)\,, s.t. W^T W = I_d$$

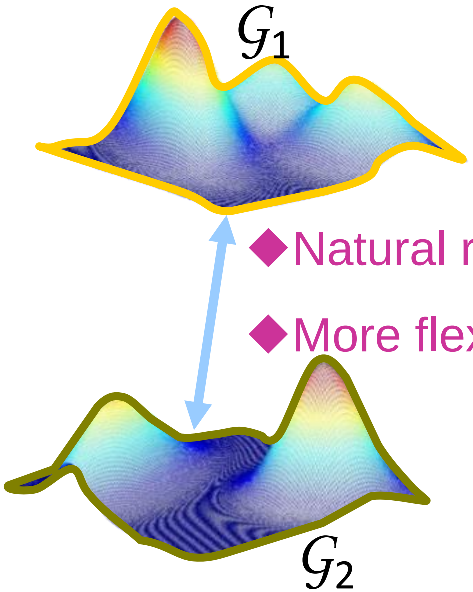
$$L(W) = \sum_{i,j} \underbrace{a(X_i, X_j)}_{\text{Affinity}} \cdot \delta(W^T X_i, W^T X_j)$$



  - High affinity $a(X_i, X_j) \rightsquigarrow$ small distance after mapping
  - Low/negative affinity $a(X_i, X_j) \rightsquigarrow$ large distance after mapping
  - Optimization by conjugate gradient on a Grassmann manifold.

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]
  - B. Gauss [ICCV’15]
  - SPD-ML [ECCV’14]
  - LEML [ICML’15]
  - DCRL [CVPR’17]
  - SPDNet [AAAI’17]
  - DHH [TIP’19]



◆Natural raw statistics

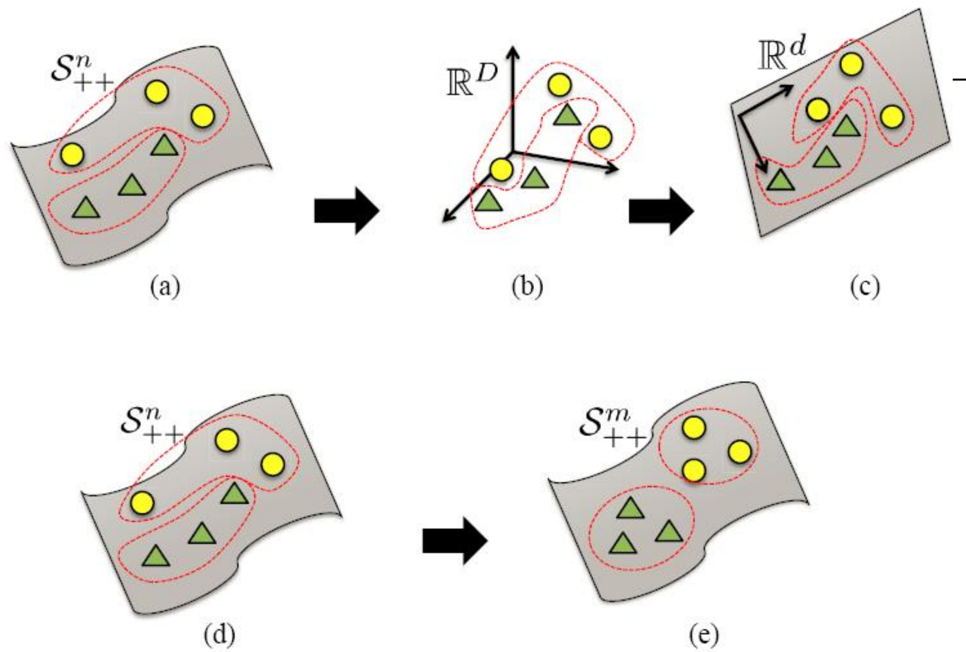
◆More flexible model

[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

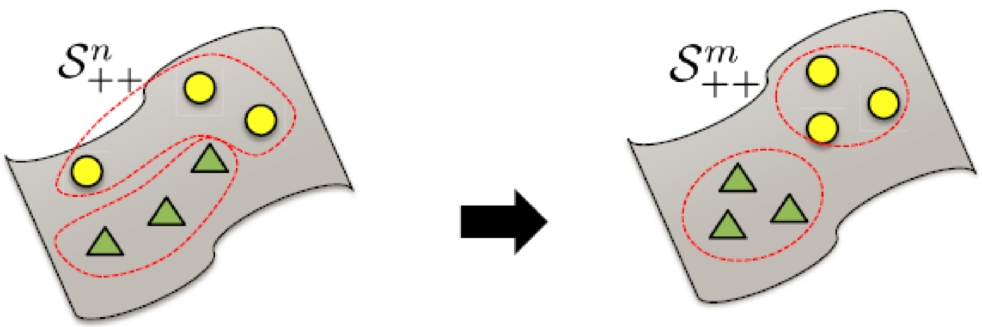


- Structure distortion
- No explicit mapping

[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: $f: \mathcal{S}_{++}^{n} \times \mathbb{R}^{n\times m} \rightarrow \mathcal{S}_{++}^{m}$
    - $f(\boldsymbol{X}, \widetilde{\boldsymbol{W}}) = \widetilde{\boldsymbol{W}}^{T}\boldsymbol{X}\widetilde{\boldsymbol{W}} \in \mathcal{S}_{++}^{m} \succ 0, \boldsymbol{X} \in \mathcal{S}_{++}^{n}, \widetilde{\boldsymbol{W}} \in \mathbb{R}^{n\times m}$ (full rank)



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

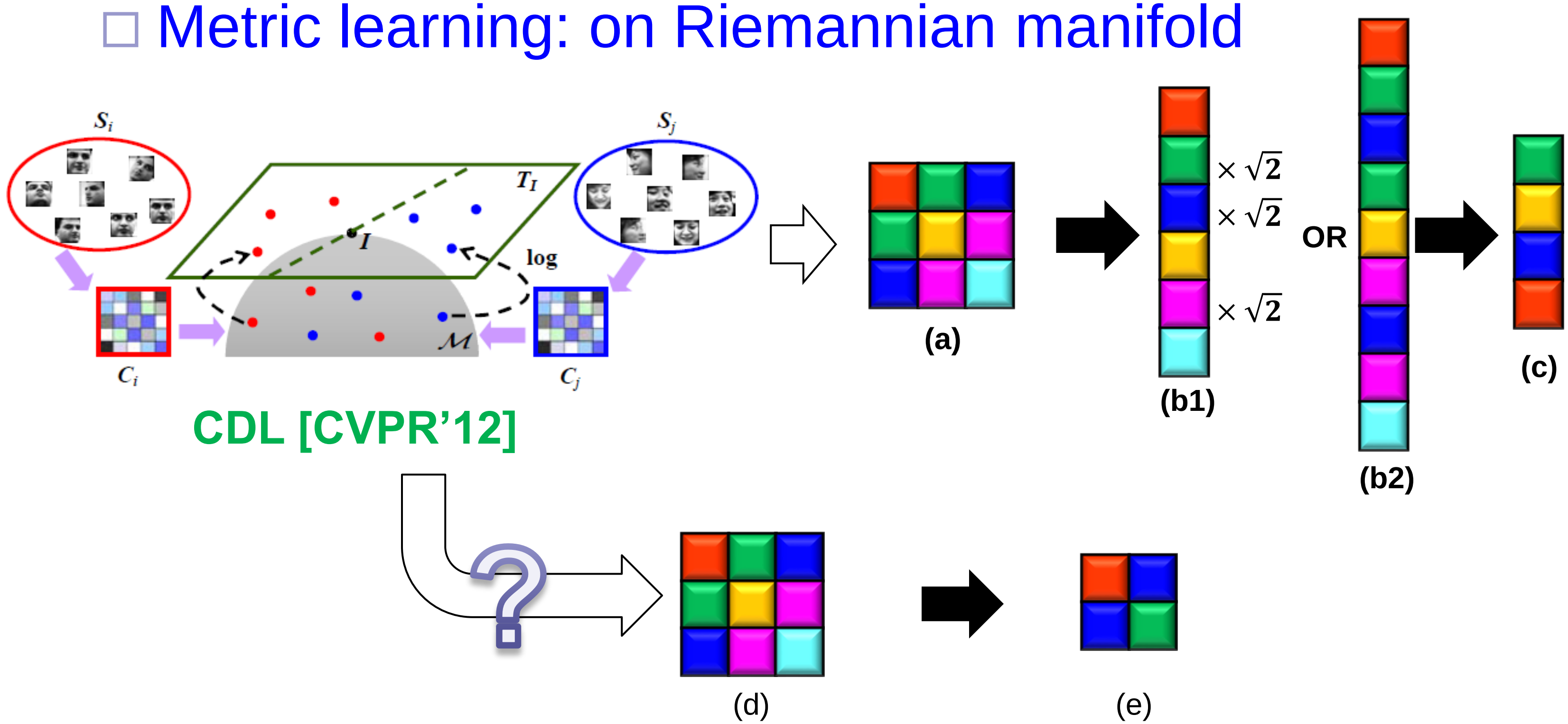
SPD-ML

- 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

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

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*.

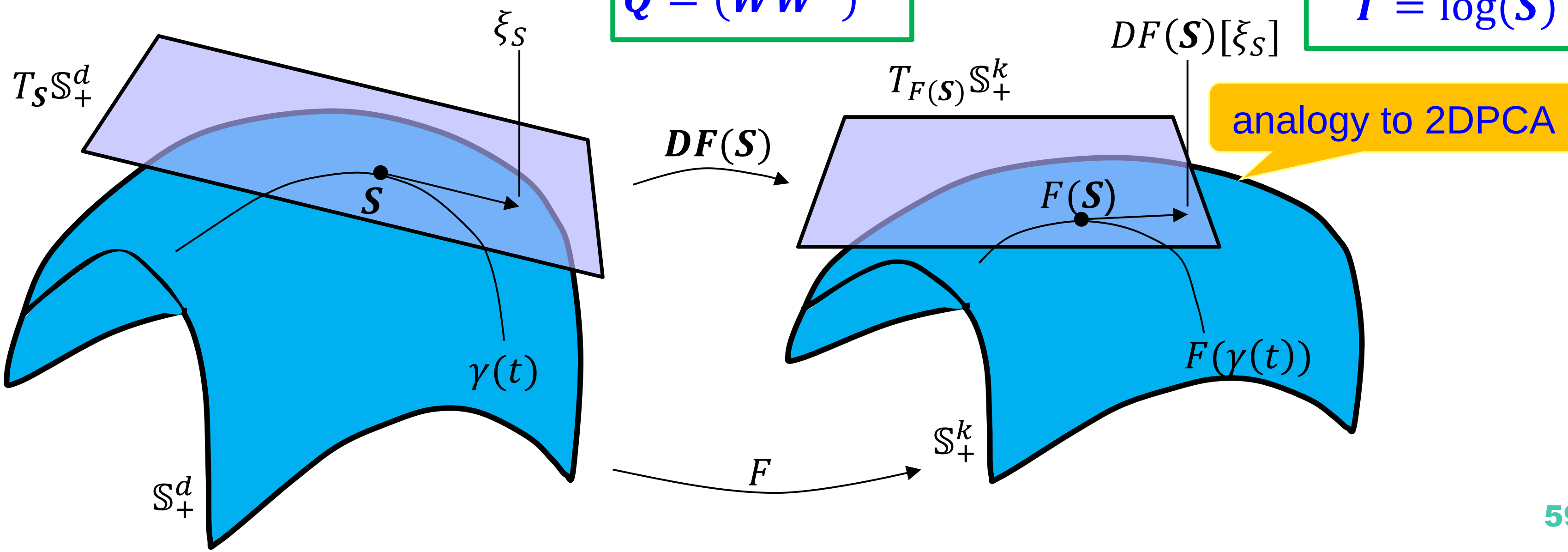
LEML

- 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
    - $d_{LED}\left(f(\boldsymbol{T}_i), f(\boldsymbol{T}_j)\right) = ||\boldsymbol{W}^T\boldsymbol{T}_i\boldsymbol{W} - \boldsymbol{W}^T\boldsymbol{T}_j\boldsymbol{W}||_F$

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

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

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



LEML

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

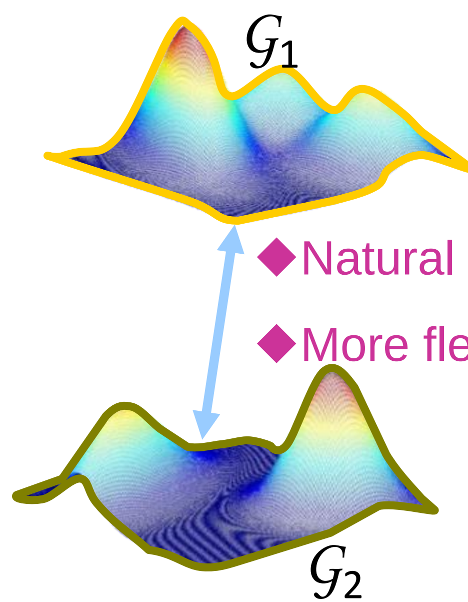
$$\text{s.t.}, \operatorname{tr}\left(\boldsymbol{Q}\boldsymbol{A}_{ij}^T\mathbf{A}_{ij}\right) \le \xi_{\boldsymbol{c}(i,j)}, (i,j) \in \boldsymbol{S}$$

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

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

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]
  - □ B. Gauss [ICCV'15]
  - □ SPD-ML [ECCV'14]
  - □ LEML [ICML'15]
  - □ DCRL [CVPR'17]
  - □ SPDNet [AAAI'17]
  - □ DHH [TIP'19]



◆Natural raw statistics

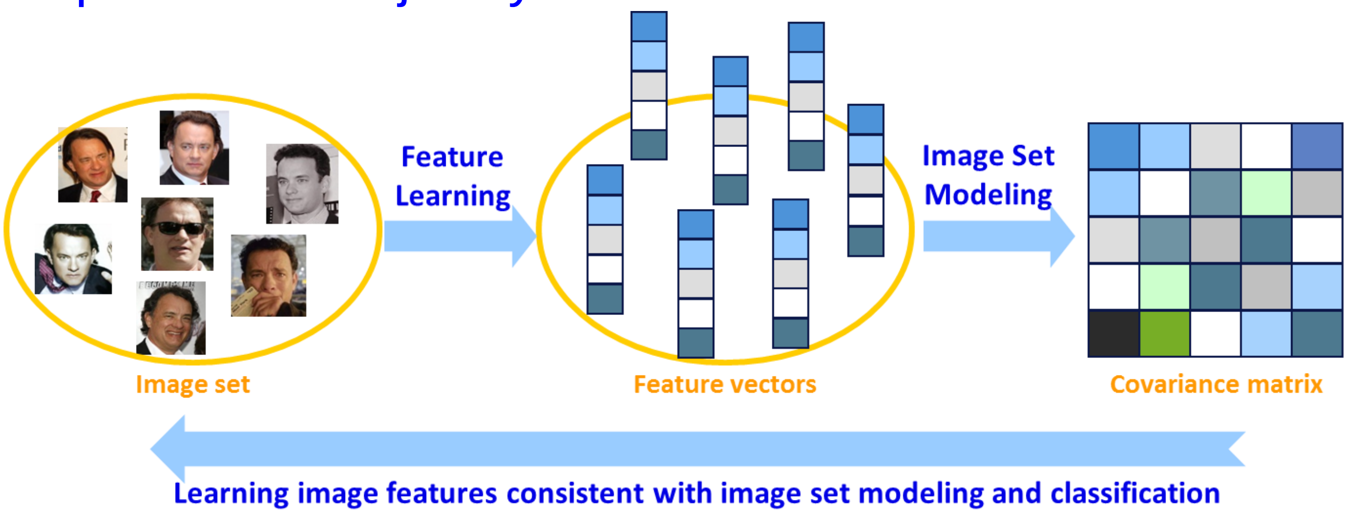
◆More flexible model

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

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

Set model IV: statistics (COV+)

- DCRL (Discriminative Covariance Oriented Representation Learning) [CVPR'17]
  - Image feature learning that facilitates image set modeling and classification
    - Image feature learning: Deep learning networks, e.g., CNN
    - Image set modeling: Set covariance matrices
  - Metric learning: on set covariance matrices with image feature space learned jointly



[1] W. Wang, R. Wang, S. Shan, X. Chen. Discriminative Covariance Oriented Representation Learning for Face Recognition with Image Sets. *IEEE CVPR 2017.*

DCRL

- Formulation
  - Given n training image sets $\{X_i\}_{i=1}^{n}$, where X_i contains original feature vectors of N_i images
  - Image feature learning
    - $X_i \mapsto h_i = \phi_\Theta(X_i)$
  - Image set modeling
    - $C_i = \hat{h}_i^T \hat{h}_i$, where $\hat{h}_i$ is the centered h_i
  - Network optimization
    - Formulate the discrimination of set covariance matrices by some loss function
    - Optimize the feature learning network to minimize such loss function

DCRL

Graph Embedding Scheme

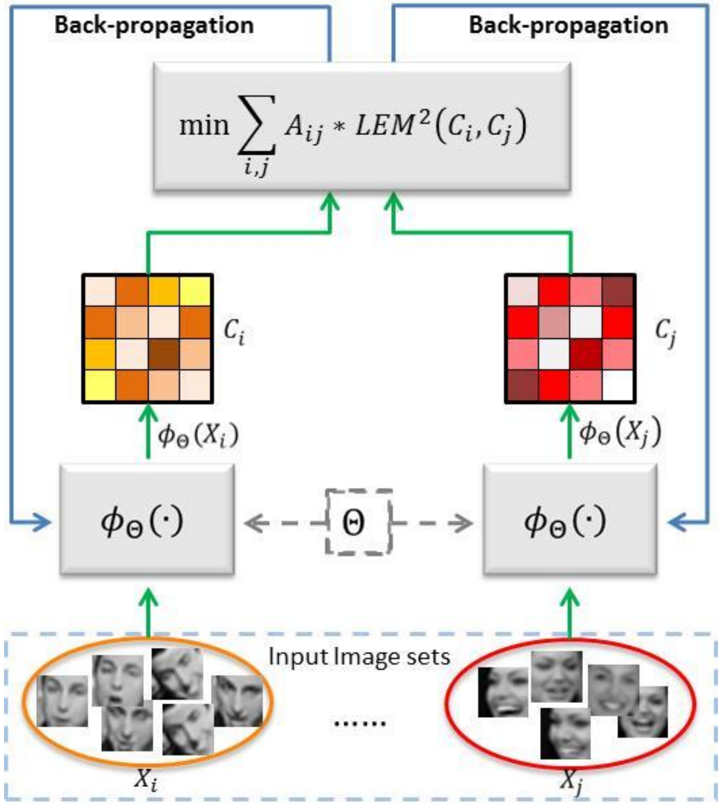
Loss function

$$J(\Theta) = \frac{1}{4}\sum_{i,j} A_{ij} LEM^2(C_i, C_j)$$

where

$$LEM(C_i, C_j) = \left\| \log_I(C_i) - \log_I(C_j) \right\|_F$$

is the Log-Euclidean Metric (LEM)*



[1] 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.

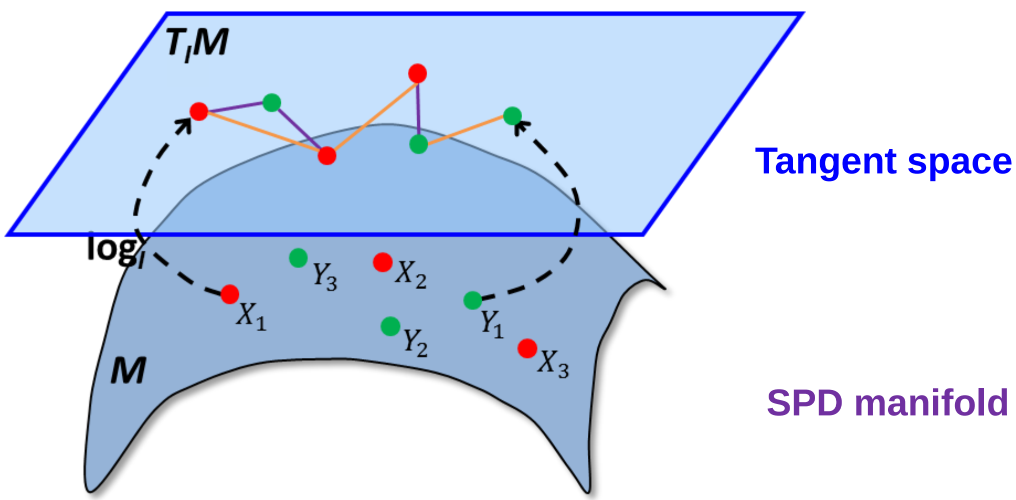
Graph Embedding Scheme

Loss function

- Adjacency Graph: Encode the data structure and semantic relationship of set covariance matrices

$$A_{ij} = \begin{cases} d_{ij} & if\ X_i \in N_w(X_j)\ or\ X_j \in N_w(X_i) \\ -d_{ij}, & if X_i \in N_b(X_j)\ or\ X_j \in N_b(X_i) \\ 0 & otherwise \end{cases}$$

$$d_{ij} = \exp(-LEM^2(C_i, C_j)/\sigma^2)$$



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

- Softmax Regression Scheme
  - Loss function
    - Softmax regression

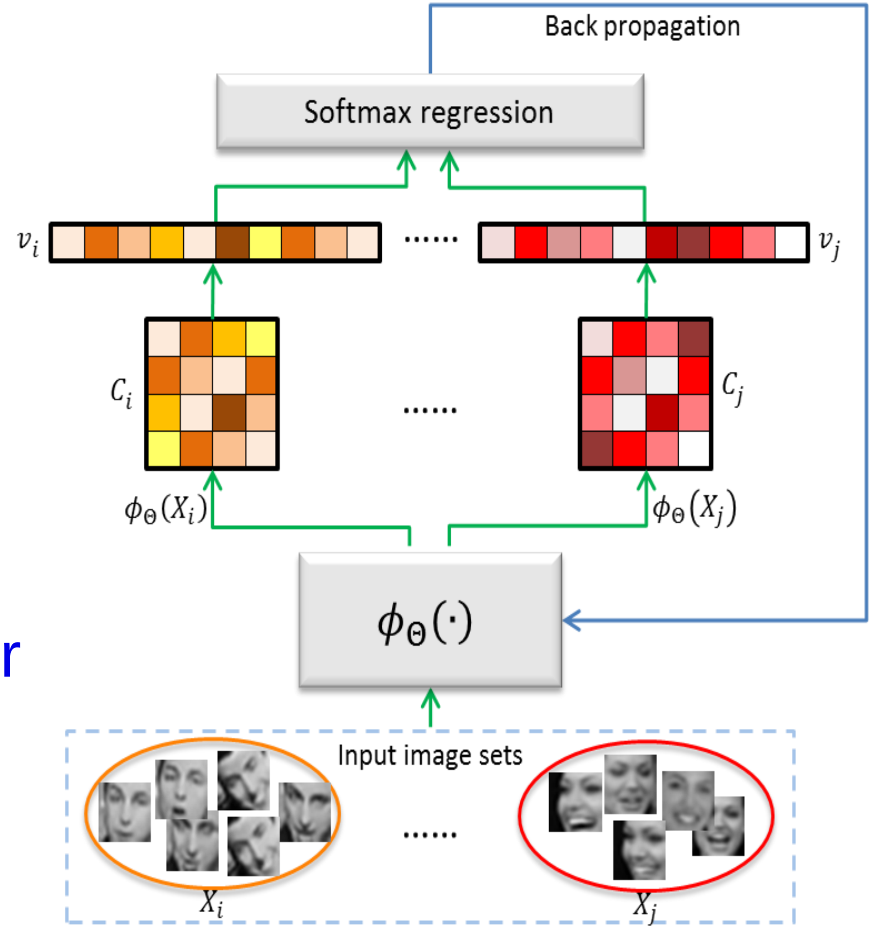
$$J(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} 1\{y_i = j\}\log(o_{ij})$$

$$1\{true\} = 1, 1\{false\} = 0;$$

$$o_{ij} = P(y_i = j | v_i; W, b)$$

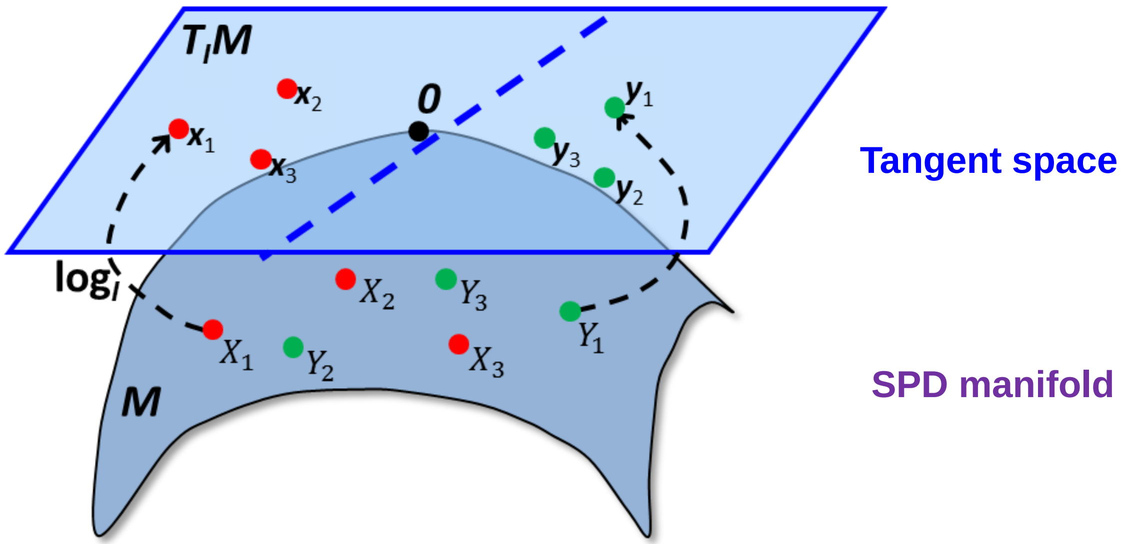
  - log-covariance vector

$$v_i = vec(\log(C_i))$$



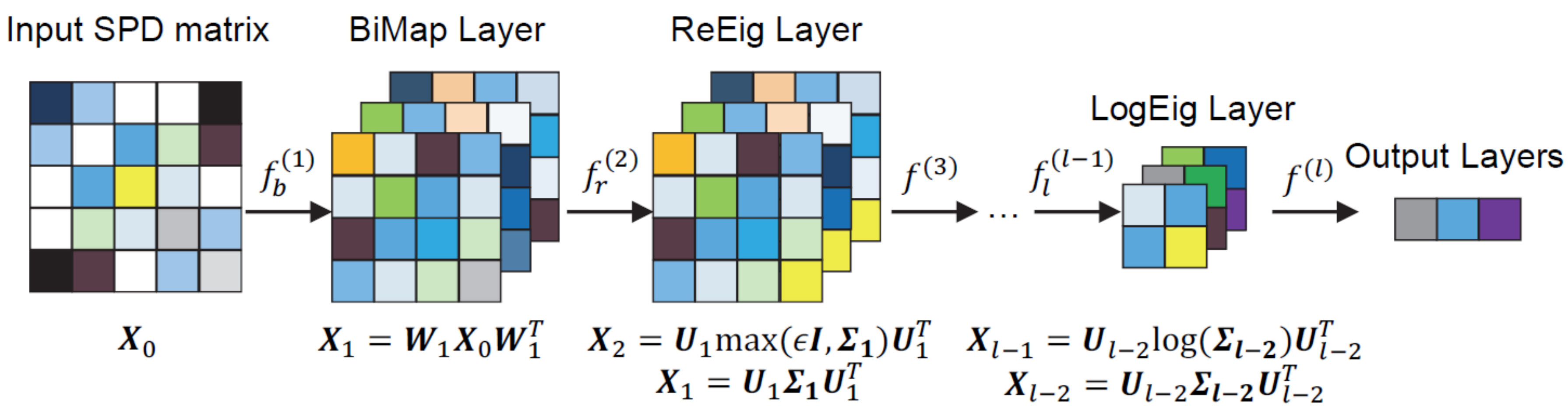
DCRL

- Softmax Regression Scheme
  - Loss function
    - Train a Softmax classifier to discriminate the set covariance matrices on a flat tangent space



Set model IV: statistics (COV+)

- SPDNet (SPD matrix network) [AAAI'17]
  - Metric learning: SPD matrices nonlinear learning in deep networks



[1] Z. Huang, L. Van Gool. A Riemannian Network for SPD Matrix Learning. *AAAI 2017*.

SPDNet

- Basic idea: respect Riemannian geometry of SPD manifold in nonlinear transformation of deep networks
  - Bilinear mapping (BiMAP) Layer
    - Fully connected convolution-like
  - Eigenvalue rectification (ReEig) Layer
    - Rectified linear units (ReLU)-like
  - Eigenvalue logarithm (LogEig) Layer
    - Manifold to Euclidean space

BiMap Layer

$X_1 = W_1 X_0 W_1^T$

ReEig Layer

$X_2 = U_1 \max(\epsilon I, \Sigma_1) U_1^T$
$X_1 = U_1 \Sigma_1 U_1^T$

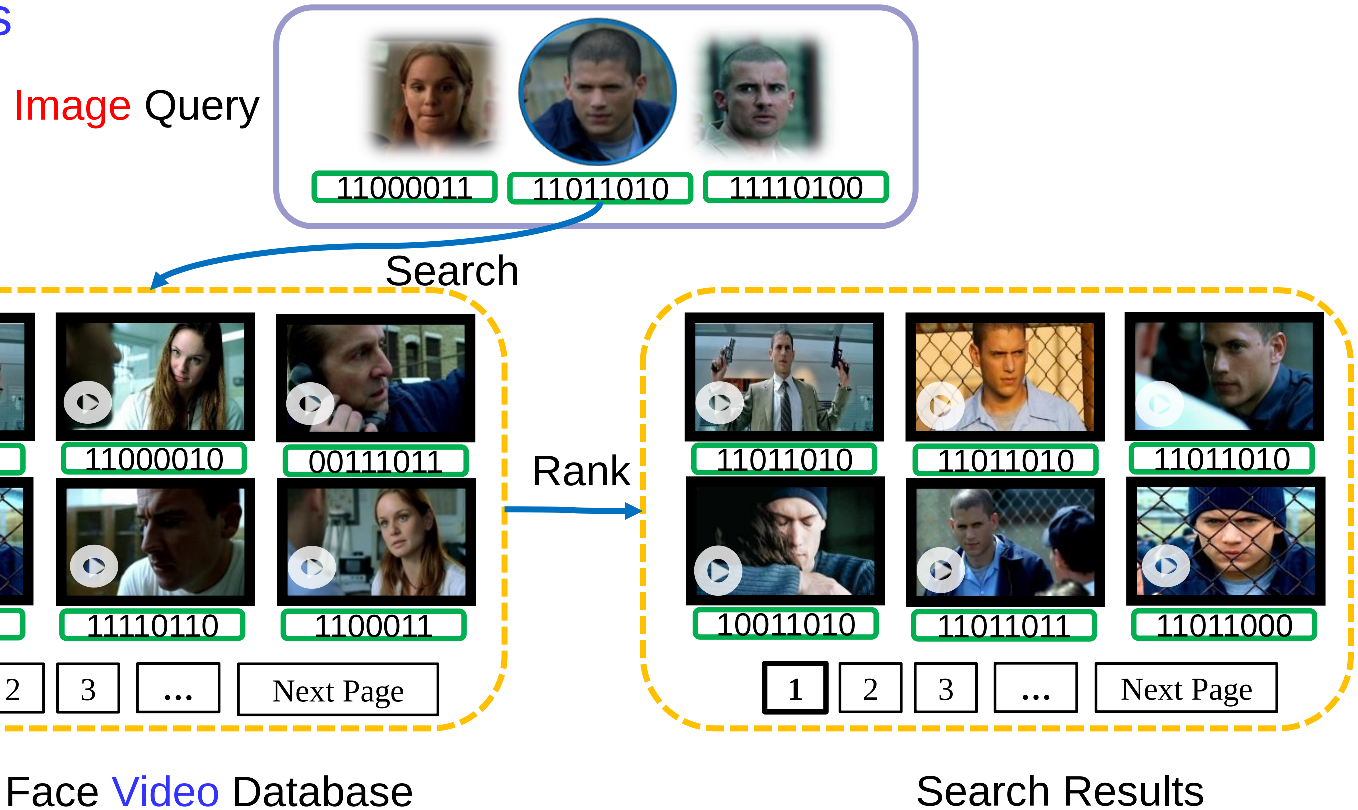
LogEig Layer

$X_{l-1} = U_{l-2} \log(\Sigma_{l-2}) U_{l-2}^T$
$X_{l-2} = U_{l-2} \Sigma_{l-2} U_{l-2}^T$

Formulation analogous to LieNet

Set model IV: statistics (COV+)

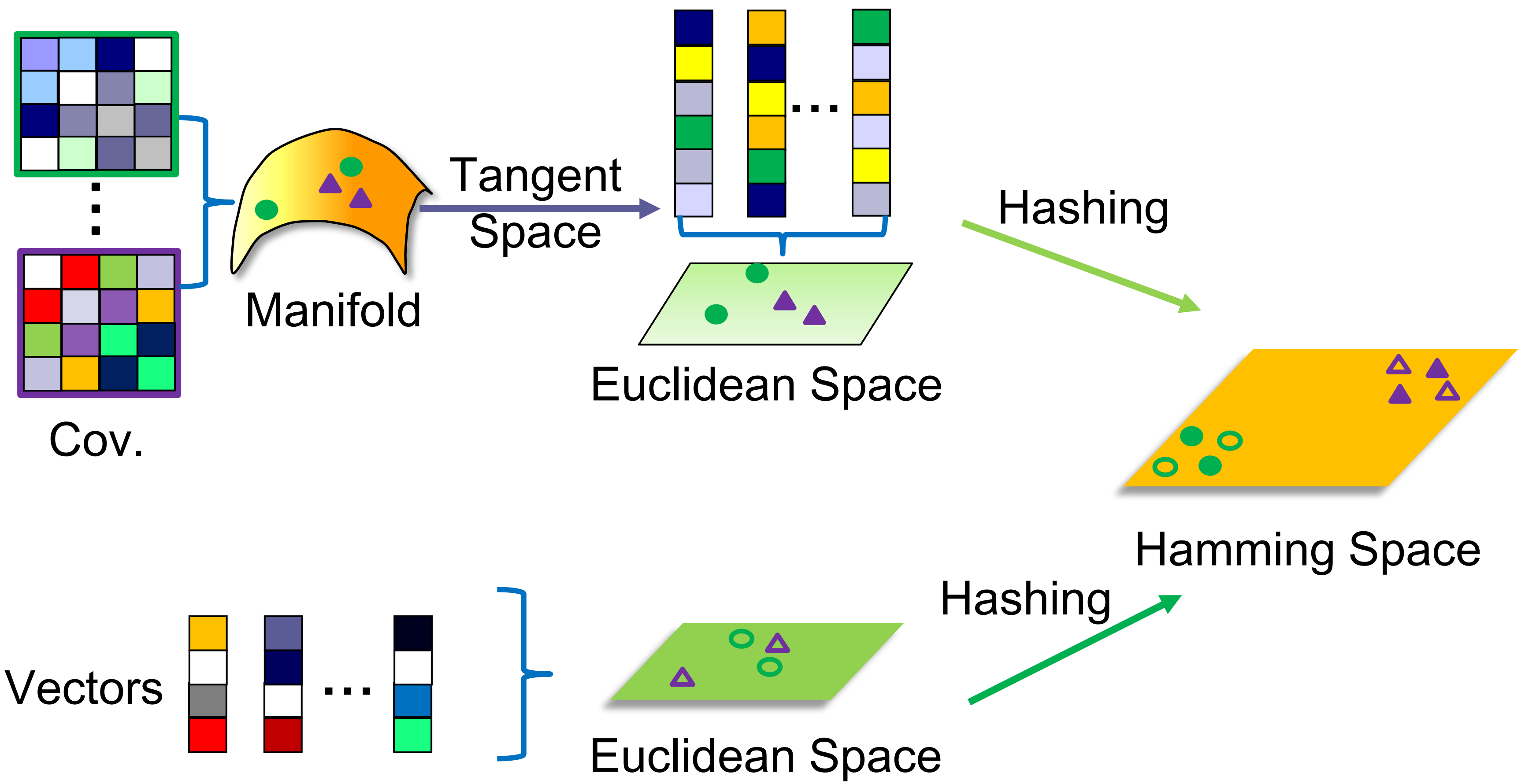
- DHH (Deep Heterogeneous Hashing) [TIP'19]
  - Application scenario: image-video face retrieval
  - Metric learning: hamming distance learning across deep heter. spaces



[1] S. Qiao, R. Wang, S. Shan, X. Chen. Deep Heterogeneous Hashing for Face Video Retrieval. *IEEE TIP 2019*.

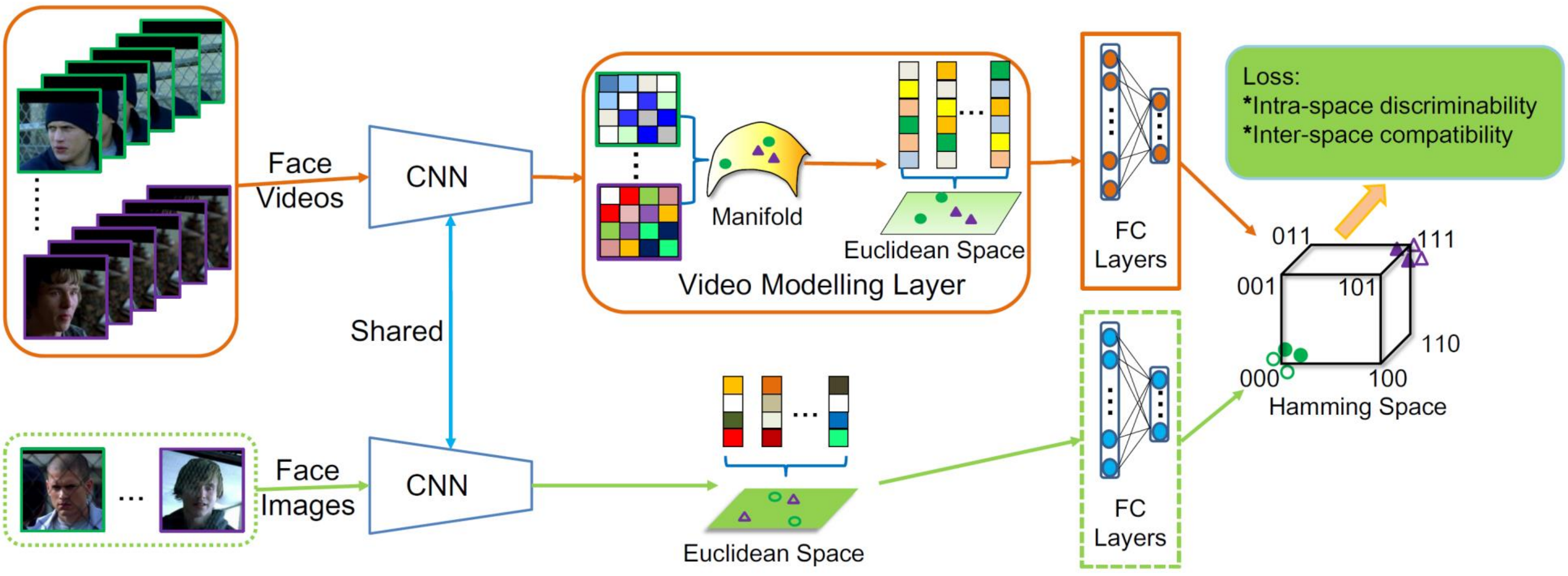
DHH

- Heterogeneous Hash Learning



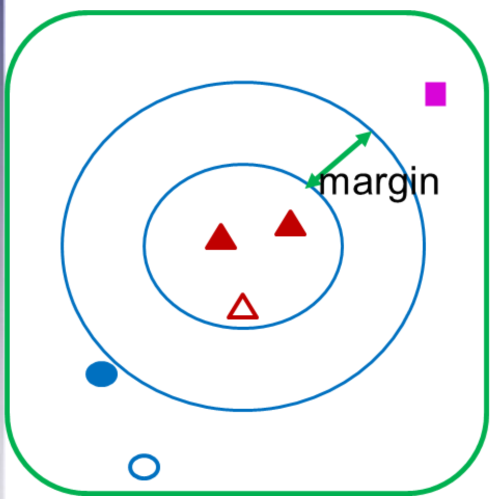
DHH

- Framework
  - Jointly learn deep hashing in homogeneous and across heterogeneous spaces



DHH

- Framework
  - Jointly learn deep hashing in homogeneous and across heterogeneous spaces



Triplet Loss Function

$$J_{i,j,k} = \max\{0, \alpha + d_h(b_i, b_j) - d_h(b_i, b_k)\}$$

$$\text{s.t.} \quad b_i, b_j, b_k \in \{0,1\}^K,$$

$$(i,j) \in positive, \quad (i,k) \in \text{negative}$$

Objective Function

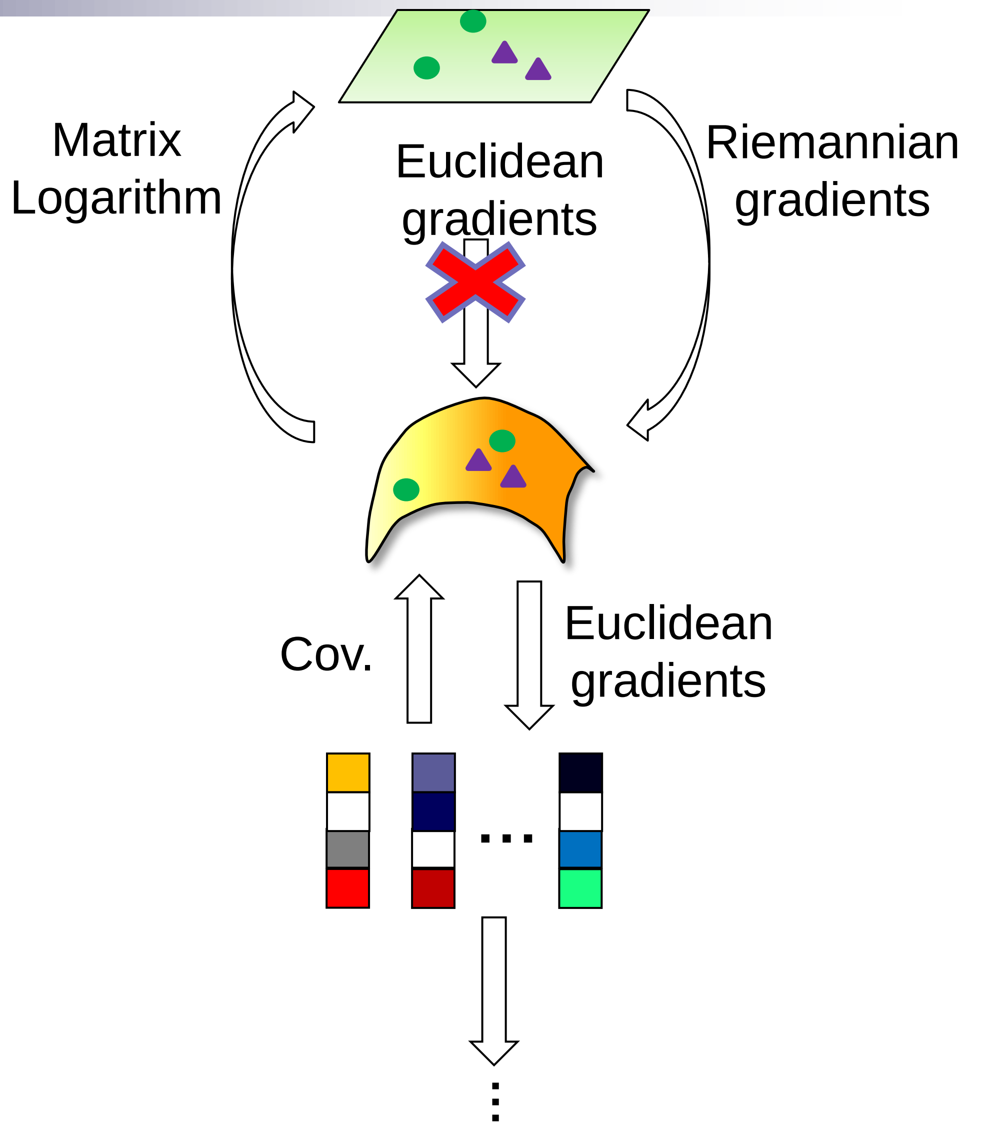
$$J = J_{er} + \lambda_1 J_e + \lambda_2 J_r$$

Across heter. spaces (J_{er})

Within homo. spaces ($\lambda_1 J_e + \lambda_2 J_r$)

DHH

- Riemannian matrix backpropagation



DHH

- Riemannian matrix backpropagation

Structured gradients between two adjacent layers

$$\frac{\partial L^{(k)}}{\partial X_{k-1}}:dX_{k-1}=\frac{\partial L^{(k+1)}}{\partial X_k}:dX_k$$

Riemannian matrix gradients of video second-order modeling

$$Input: X_{k-1}, SVD(X_{k-1})=U\Sigma V^T, \qquad Output:\ X_k=V logm^{(\Sigma^T\Sigma+\varepsilon I)}V^T$$

$$D=\Sigma_m^{-1}\left(\frac{\partial L^{(k')}}{\partial V}\right)_1^T-\Sigma_m^{-1}V_1^T\left(\frac{\partial L^{(k')}}{\partial V}\right)_2 V_2^T, K_{i,j}=\begin{cases}0 & i=j\\ \frac{1}{\sigma_j^2-\sigma_i^2} & i\neq j\end{cases}$$

$$\frac{\partial L^{(k)}}{\partial X_{k-1}}=UD+U\left(\frac{\partial L^{(k')}}{\partial \Sigma}-DV\right)_{diag}V^T+2U(K\circ(-DV\Sigma^T))_{sym}\Sigma V^T$$

$$\frac{\partial L^{(k')}}{\partial V}=2\left(\frac{\partial L^{(k+1)}}{\partial X_k}\right)_{sym} V logm^{(\Sigma^T\Sigma+\varepsilon I)},$$

$$\frac{\partial L^{(k')}}{\partial \Sigma}=2\Sigma(\Sigma^T\Sigma+\varepsilon I)^{-1}V^T\left(\frac{\partial L^{(k+1)}}{\partial X_k}\right)_{sym}V$$

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

Outline

- Background
- Literature review
- Evaluations
- Summary

Evaluations: video face recognition

- 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

Evaluations: video face recognition

- 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

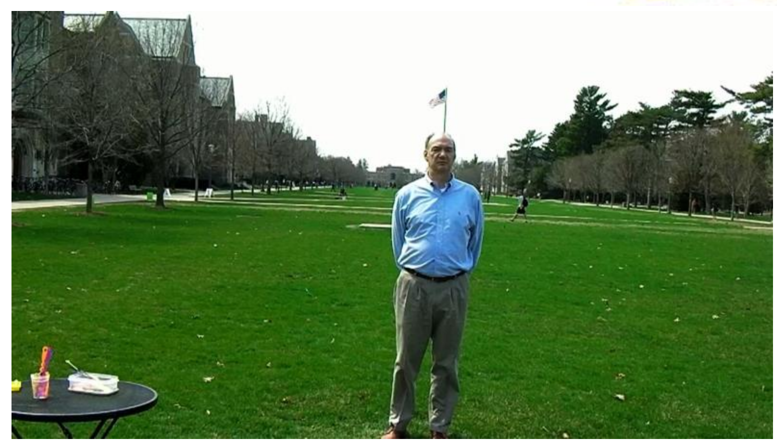
Evaluations: video face recognition

- 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 video



Handheld video

Evaluations

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

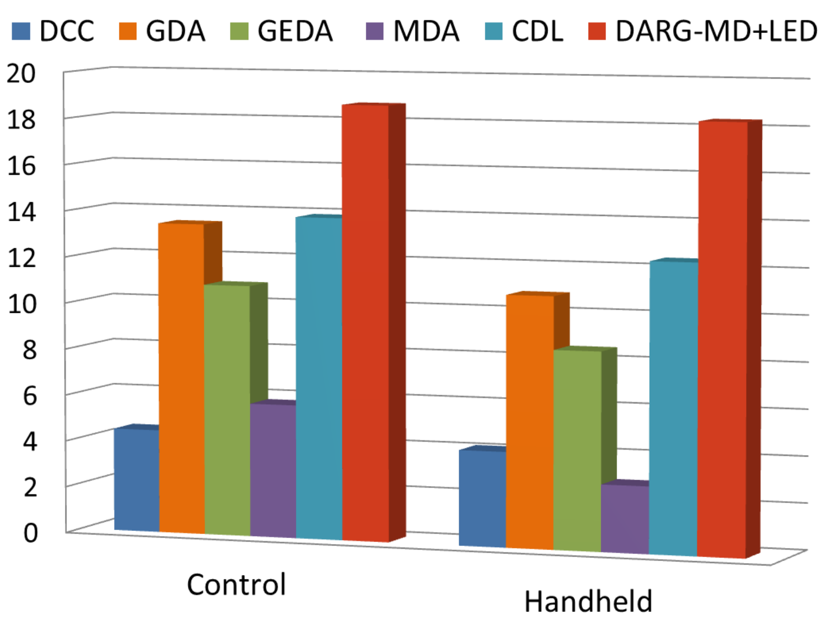
Method	YTC	COX					
		COX-11	COX-12	COX-23	COX-21	COX-31	COX-32
CHISD [CVPR'10]	66.46	56.87	30.10	14.80	44.37	26.44	13.68
GDA [CVPR'08]	65.91	72.26	80.70	74.36	71.44	81.99	77.57
GGDA [CVPR'11]	66.83	76.73	83.80	76.59	72.56	82.84	79.99
MMD [CVPR'08]	65.30	38.29	30.34	15.24	34.86	22.21	11.44
MDA [CVPR'09]	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*.

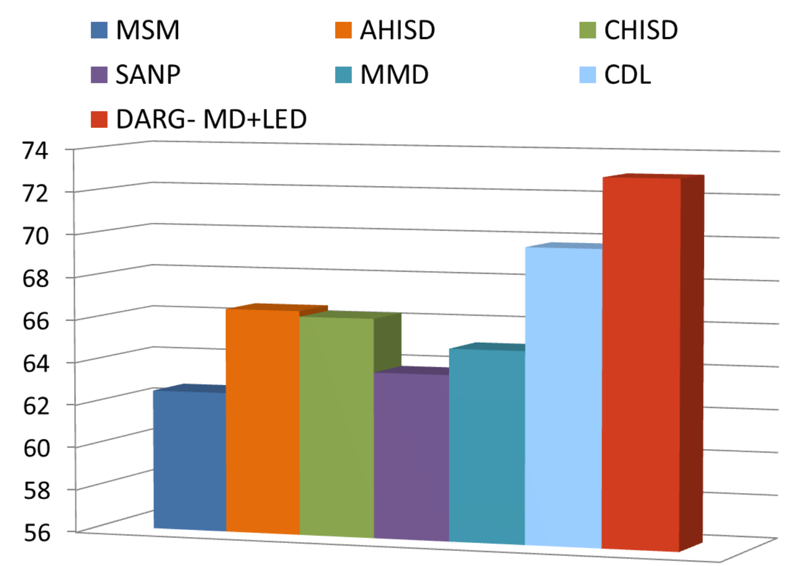
Evaluations

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

VR@FAR=0.01 on PaSC

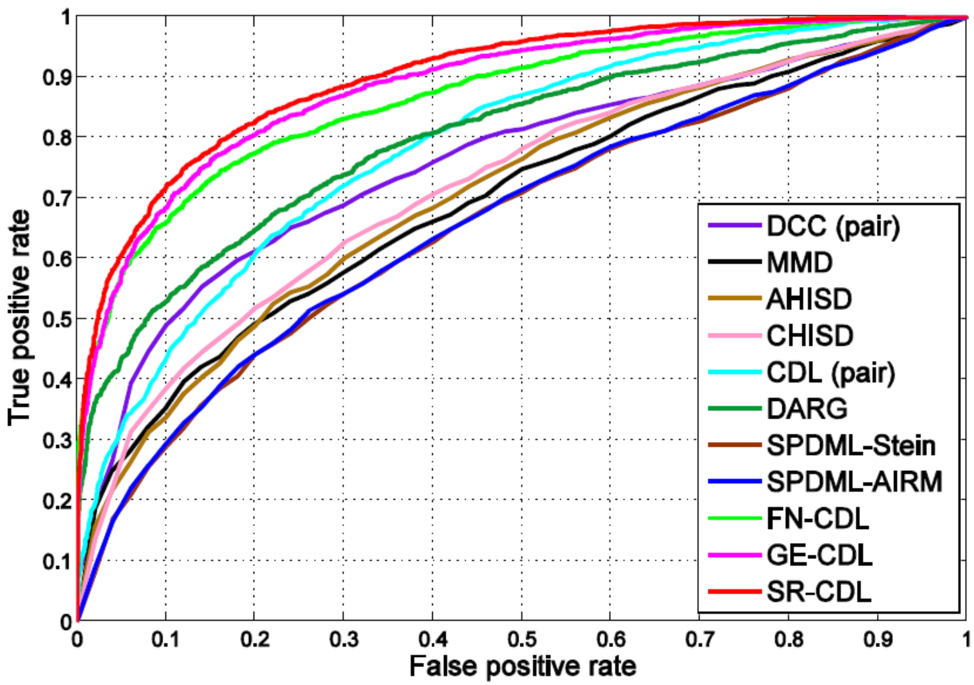


AUC on YTF



Evaluations

- Results (*reported in our DCRL paper**)
  - Verification task (on YTF dataset)

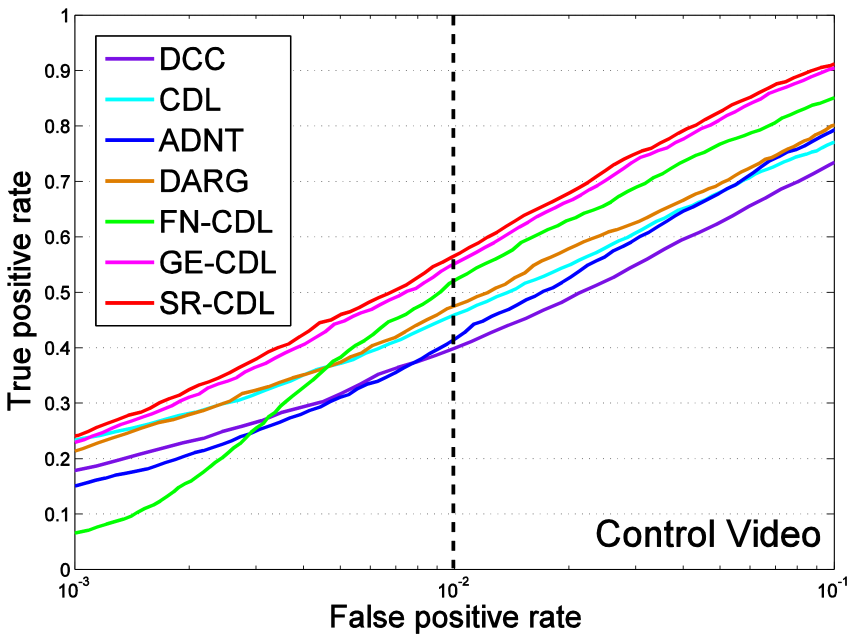


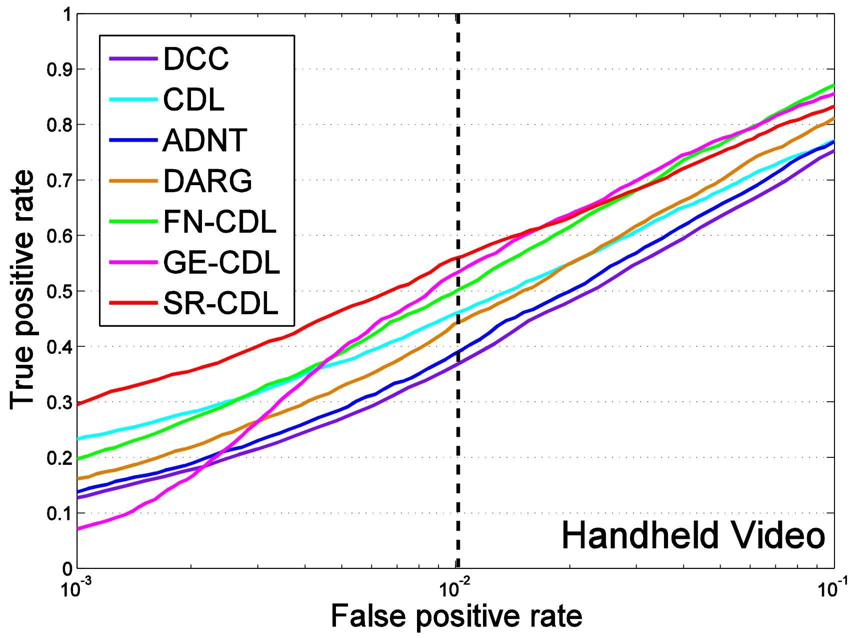
GE: Graph embedding scheme
SR: Softmax regression scheme
FN: Baseline Deep ID net with single CNN

[*] W. Wang, R. Wang, S. Shan, X. Chen. Discriminative Covariance Oriented Representation Learning for Face Recognition with Image Sets. *IEEE CVPR 2017*.

Evaluations

- Results (*reported in our DCRL paper**)
  - Verification task (on PaSC dataset)

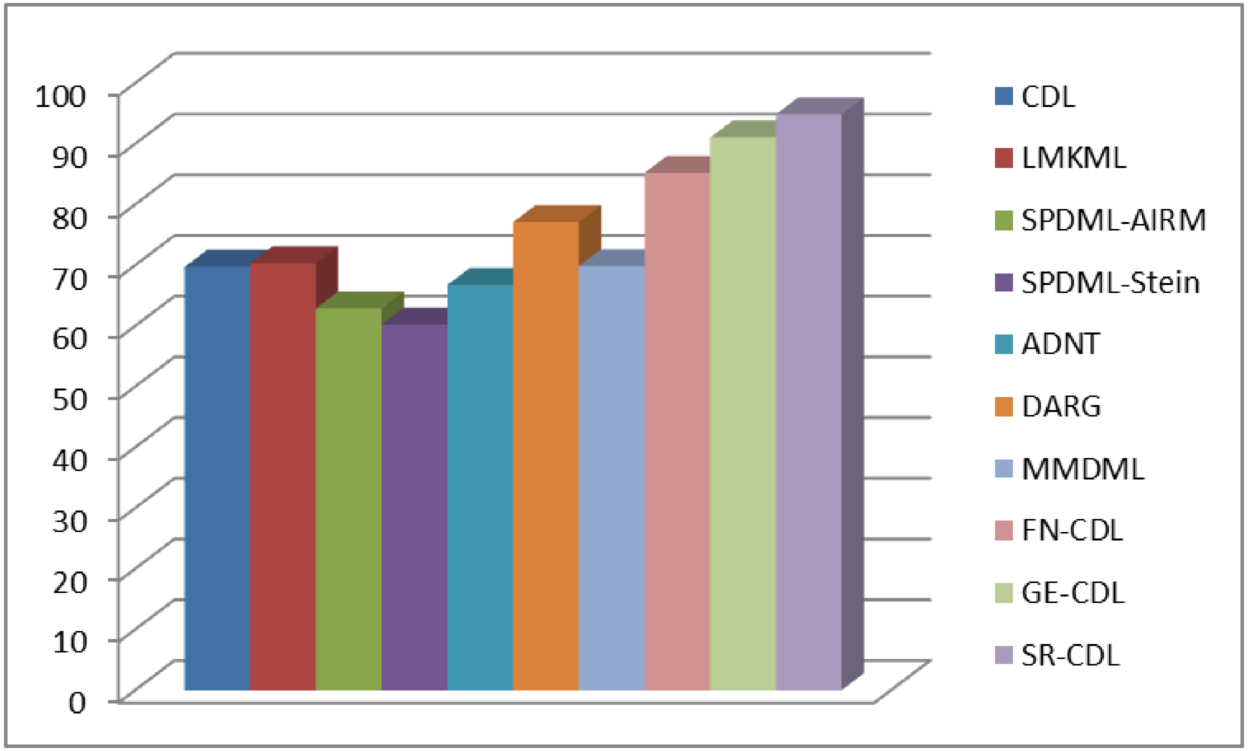




中科院计算所
Institute of Computing Technology, Chinese Academy of Sciences

Evaluations

- Results (*reported in our DCRL paper**)
  - Identification task (on YTC dataset)



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

Evaluations: video face retrieval

- Datasets
  - YouTube Celebrities (YTC) [Kim, CVPR'08]
  - Prison Break (PB) [Li, FG.'15]
  - UMDFaces-200 [Bansal, IJCB'17]

Dataset	YTC	PB	UMDFaces
Training Video #	7,190	2,415	6,614
Test Video #	3,101	10,495	3,422
Training Image #	21,570	7,245	19,842
Test Image #	3,101	10,495	3,422
Video # per subject	219.0±114.6	679.5±710.6	50.2±19.5

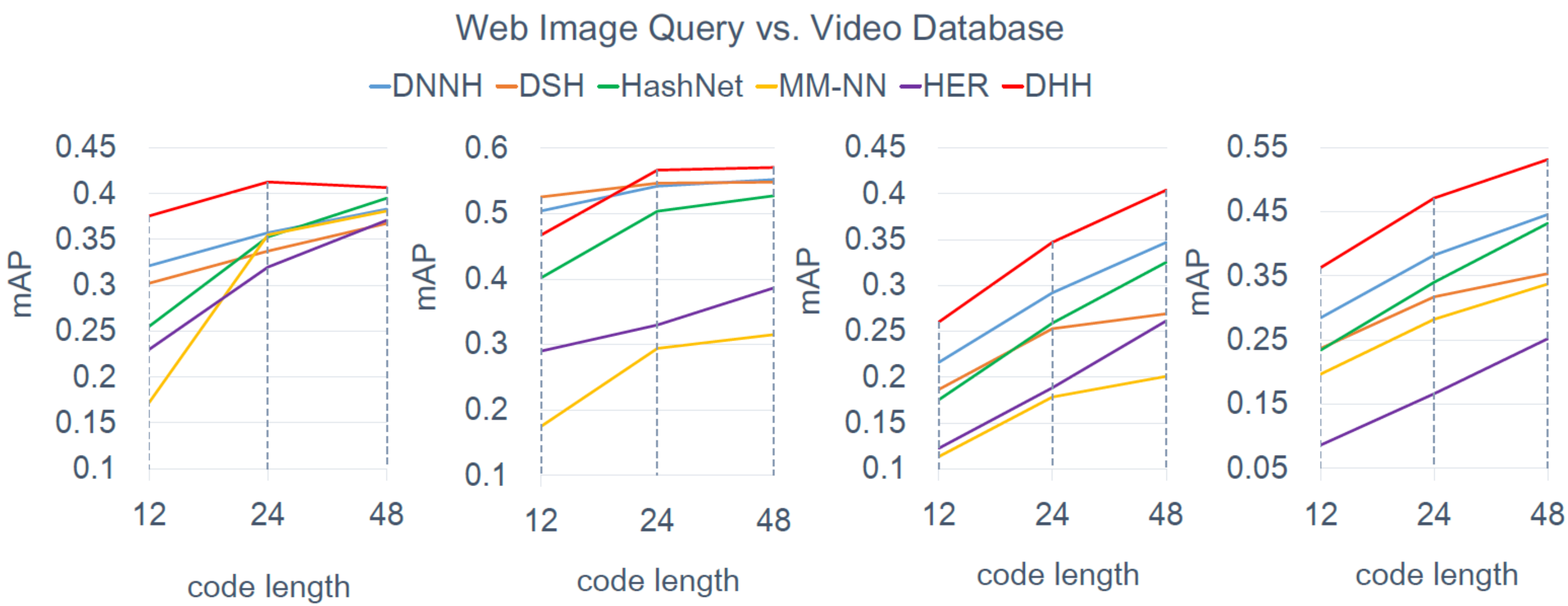
Evaluations: video face retrieval

- Retrieval mAP results

	Method	YouTube Celebrities			the Prison Break			UMDFaces		
		12-bit	24-bit	48-bit	12-bit	24-bit	48-bit	12-bit	24-bit	48-bit
SMH	LSH [4]	0.1105	0.1504	0.2042	0.2346	0.2649	0.4046	0.0600	0.1079	0.1804
	SH [5]	0.2262	0.2726	0.2814	0.3132	0.3089	0.2930	0.1369	0.2073	0.2470
	SSH [33]	0.2811	0.3324	0.3068	0.4102	0.3574	0.2931	0.1701	0.2405	0.2803
	ITQ [6]	0.3461	0.4424	0.4596	0.6666	0.7061	0.6911	0.1905	0.2791	0.3477
	DBC [35]	0.4244	0.5017	0.5478	0.7234	0.8034	0.8051	0.1509	0.2182	0.2825
	KSH [7]	0.3973	0.4917	0.5709	0.7576	0.8168	0.8451	0.1911	0.2741	0.3329
	DNNH [8]	0.4868	0.5467	0.5701	**0.9334**	**0.9480**	0.9529	0.2592	0.3563	0.4260
	DSH [46]	0.4657	0.5305	0.5432	0.9303	0.9467	0.9432	0.2443	0.3184	0.3423
	HashNet [50]	0.3965	0.5302	0.5865	0.8858	0.9372	0.9411	0.2172	0.3202	0.4190
MMH	CMSSH [52]	0.1082	0.1703	0.2005	0.2242	0.2564	0.3492	0.0586	0.1014	0.1398
	CVH [53]	0.2081	0.2371	0.2693	0.3290	0.3143	0.2712	0.1092	0.1647	0.2146
	PLMH [55]	0.1755	0.1959	0.2065	0.3130	0.3083	0.2797	0.0826	0.1370	0.1925
	PDH [10]	0.2719	0.3809	0.4190	0.5395	0.6059	0.6523	0.1128	0.1606	0.2047
	MLBE [54]	0.4641	0.4438	0.5287	0.6297	0.6281	0.6234	0.0800	0.2238	0.3265
	MM-NN [56]	0.2791	0.5218	0.5595	0.4856	0.8261	0.8468	0.1617	0.2247	0.2568
	HER [58]	0.3600	0.5045	0.5756	0.7094	0.7930	0.8421	0.1544	0.2329	0.3126
	DHH	**0.5406**	**0.5802**	**0.6120**	0.9029	0.9470	**0.9563**	**0.3037**	**0.4101**	**0.4736**

Evaluations: video face retrieval

- Retrieval using still images from internet



(a) YTC (b) PB (c) UMDFaces (d) UMDFaces*

Outline

- Background
- Literature review
- Evaluations
- Summary

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*.

中科院计算所

Thanks, Q & A

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

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