Nonlinear metric learning for video-based face recognition and retrieval

Ruiping Wang

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

Jul. 1, 2017 @ CQU

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

Outline

- Problem
- Motivation
- Our methods
- Summary

Face Recognition with Single Image

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

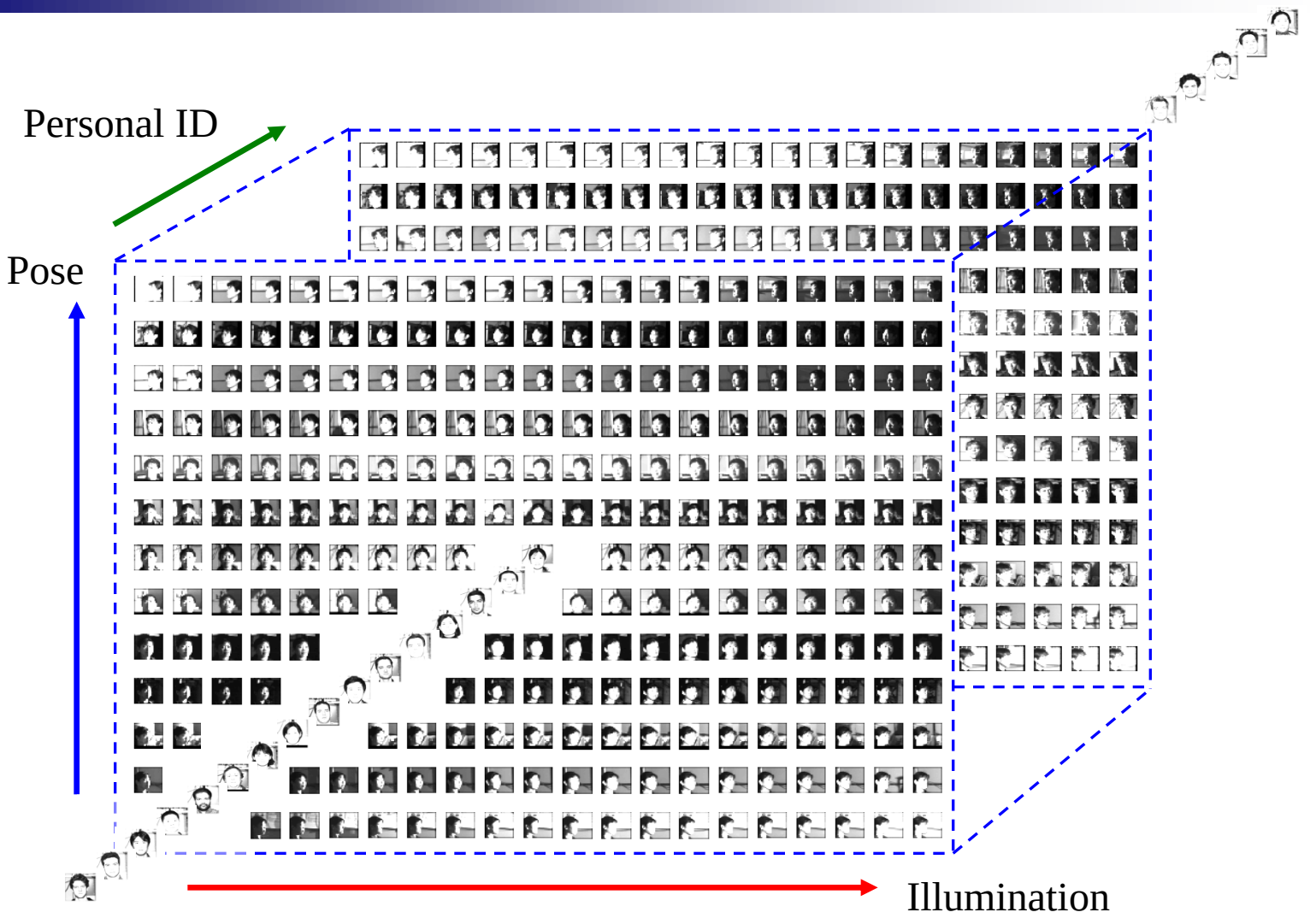
Who is this celebrity?

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

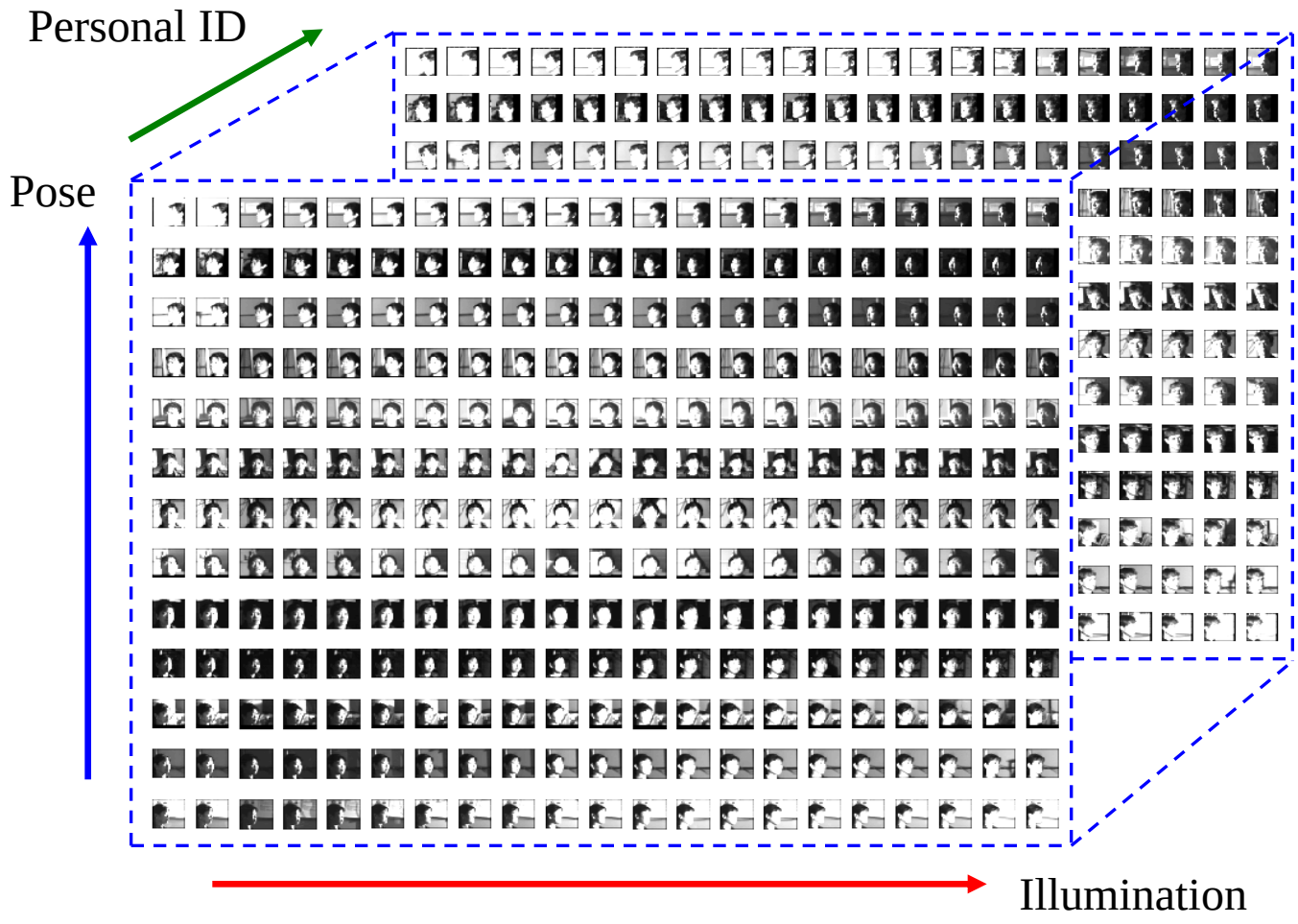
Are they the same guy?

Challenges

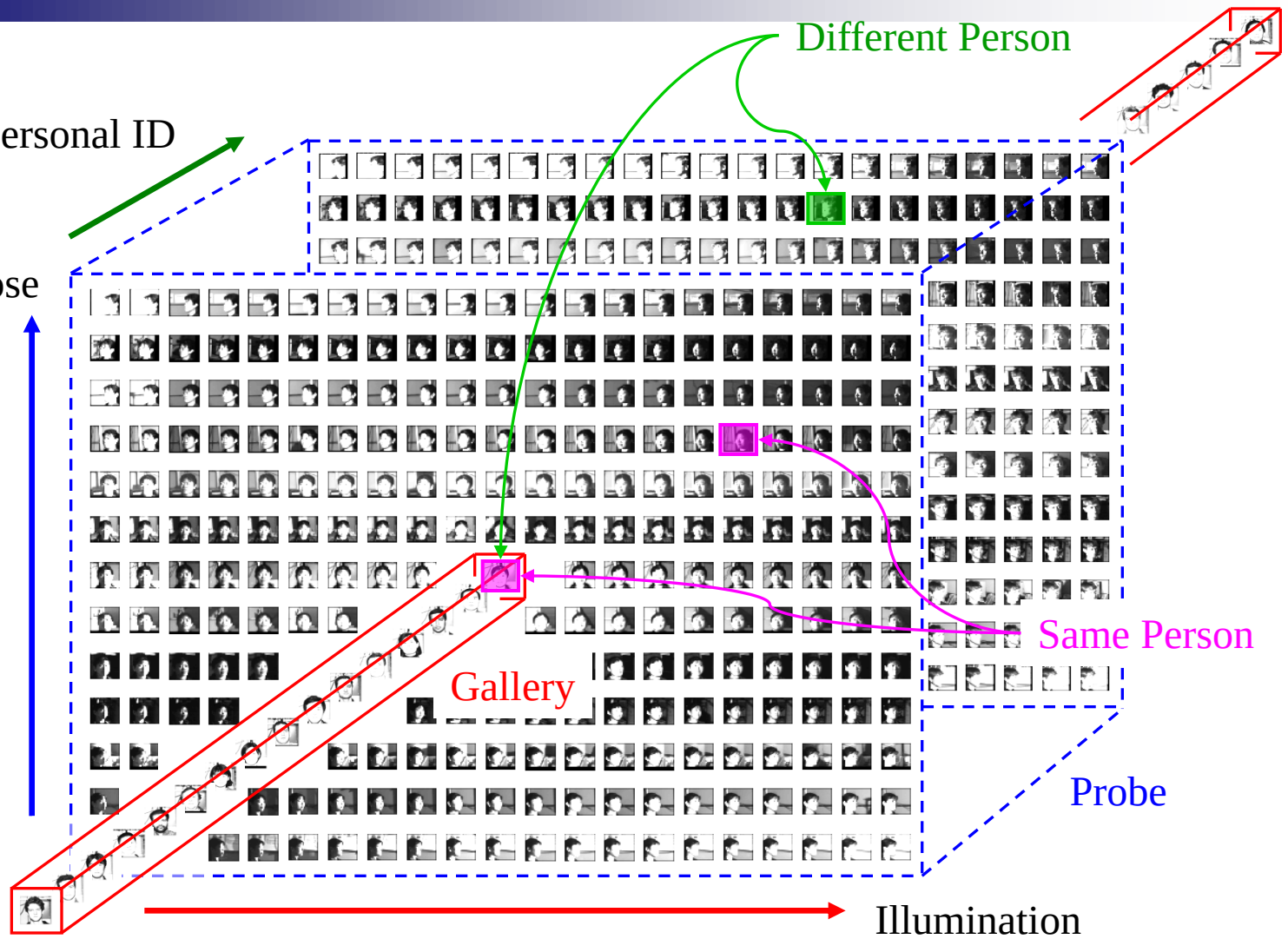
Challenges



Challenges

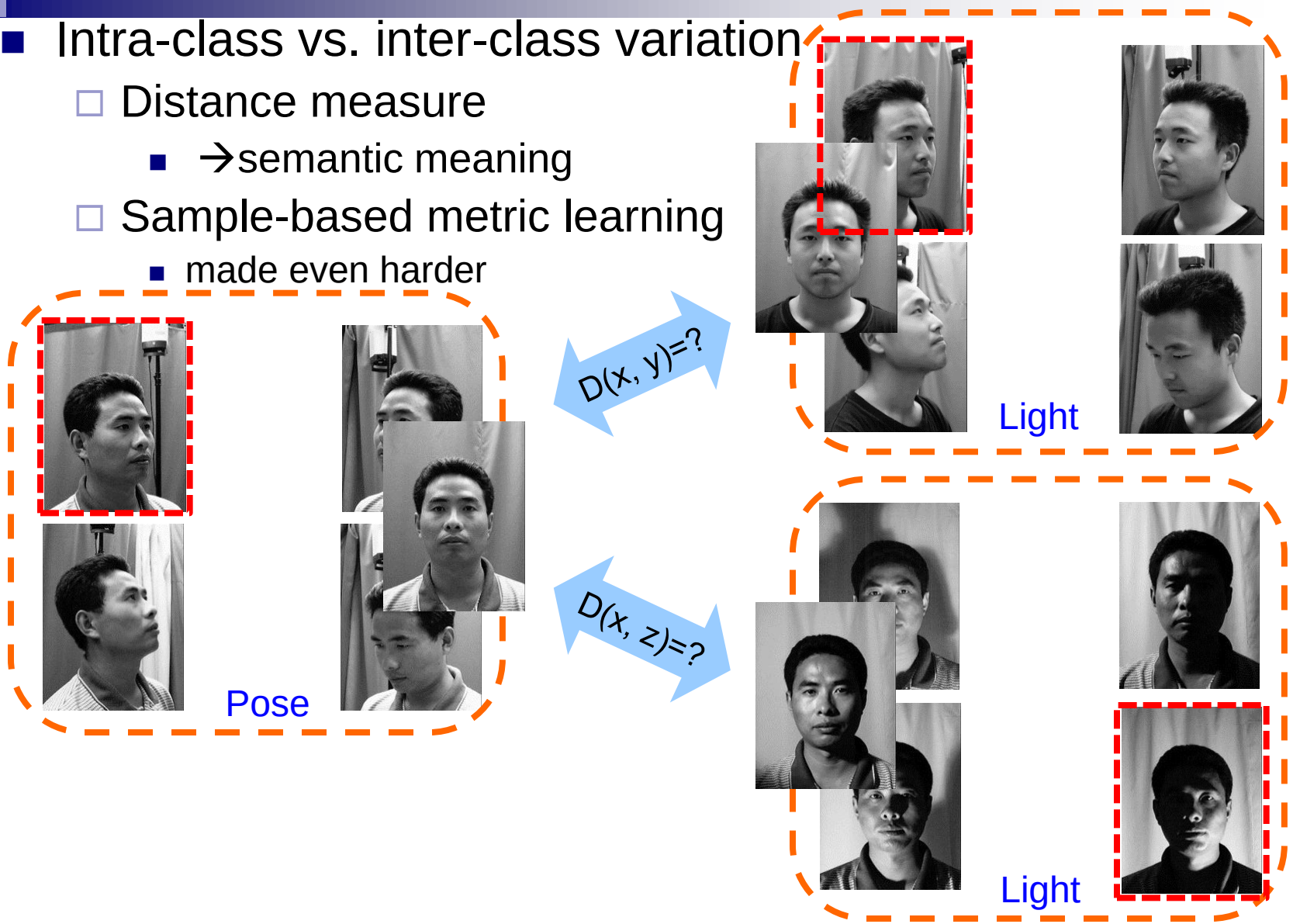


Challenges



Challenges

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



Face Recognition with Videos

- Video surveillance

Seeking missing children

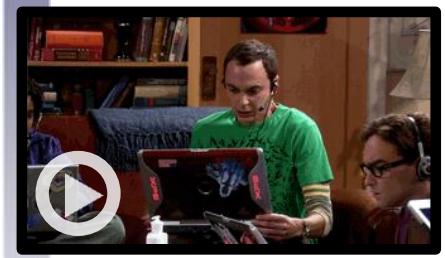
Searching criminal suspects

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

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"

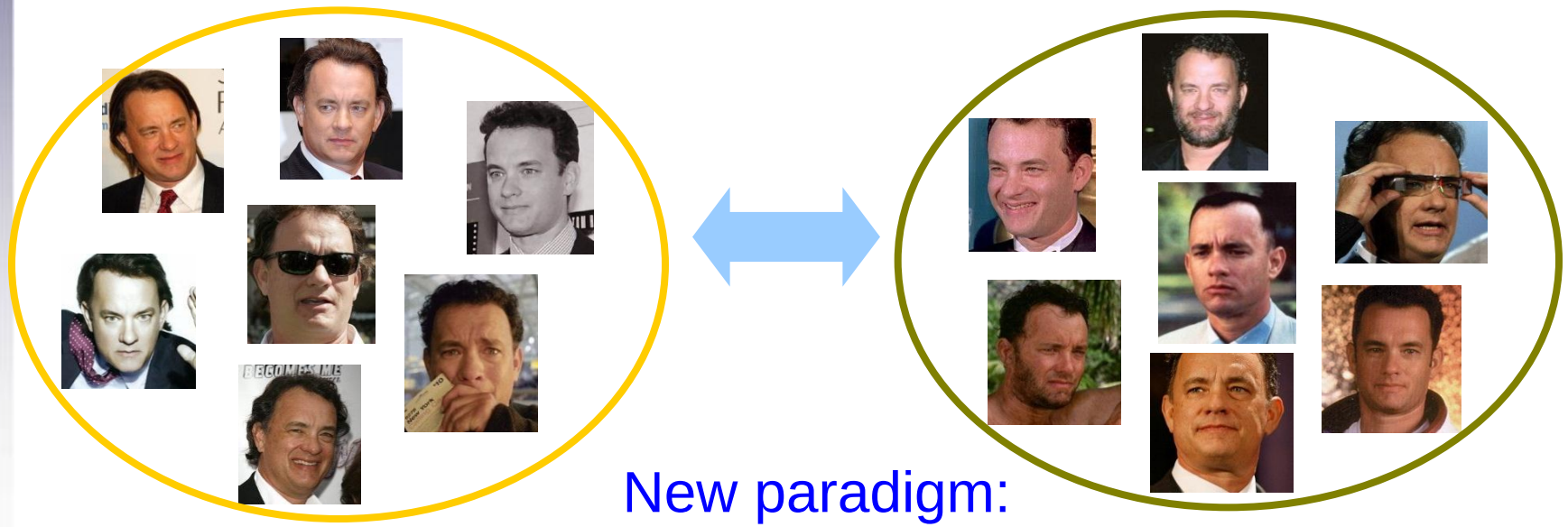

S01E04: 03'42"

Outline

- Problem
- Motivation
- Our methods
- Summary

Treating Video as Image Set

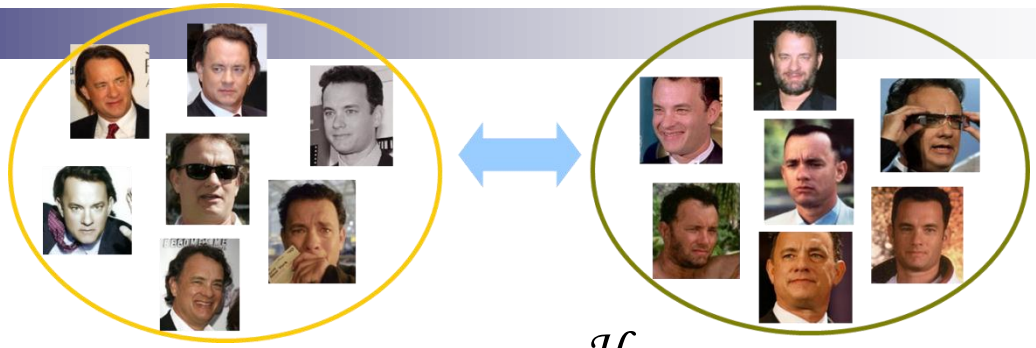
- A new & different problem
  - Unconstrained acquisition conditions
  - Complex appearance variations
  - Two phases: set modeling + set matching

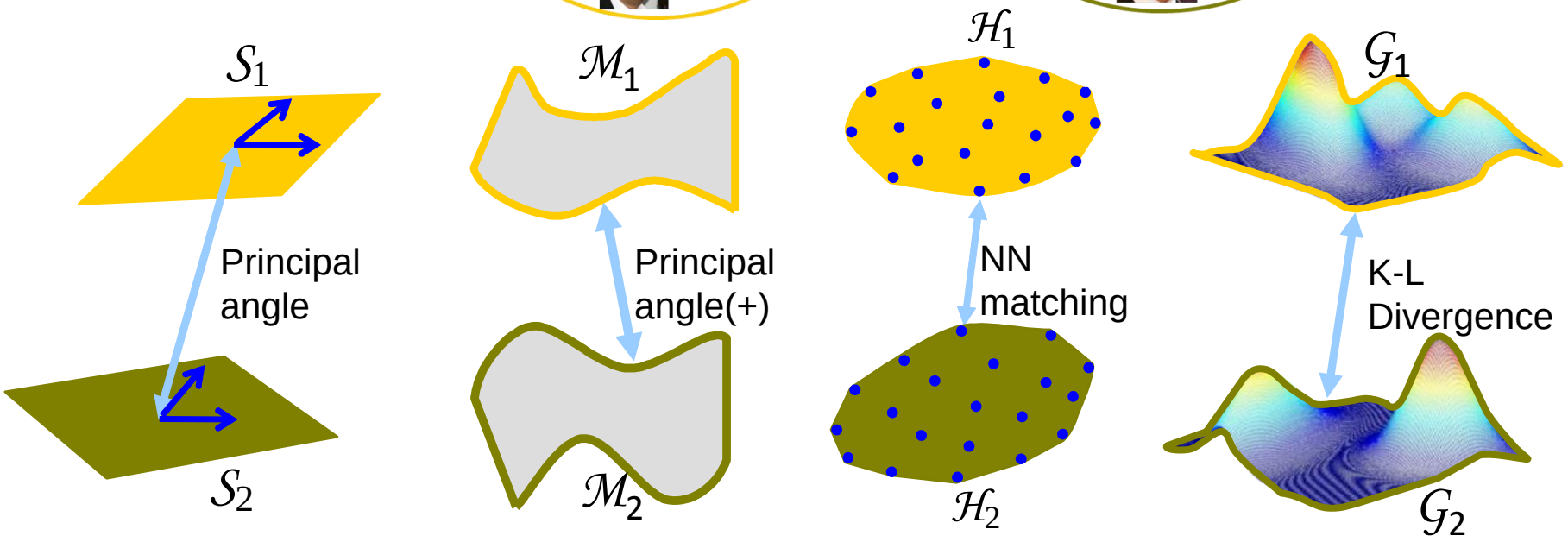


New paradigm:
set-based metric learning

Overview of existing works

From the view of set modeling





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

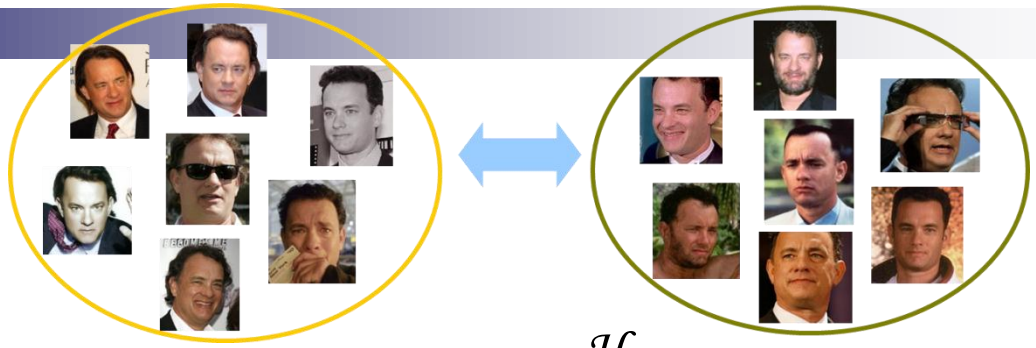
◆Nonlinear manifold
[Hadid, FG'04]
[Kim, BMVC'05]
[Wang, CVPR'08/09]
[Chen, CVPR'13]
[Lu, CVPR'15]

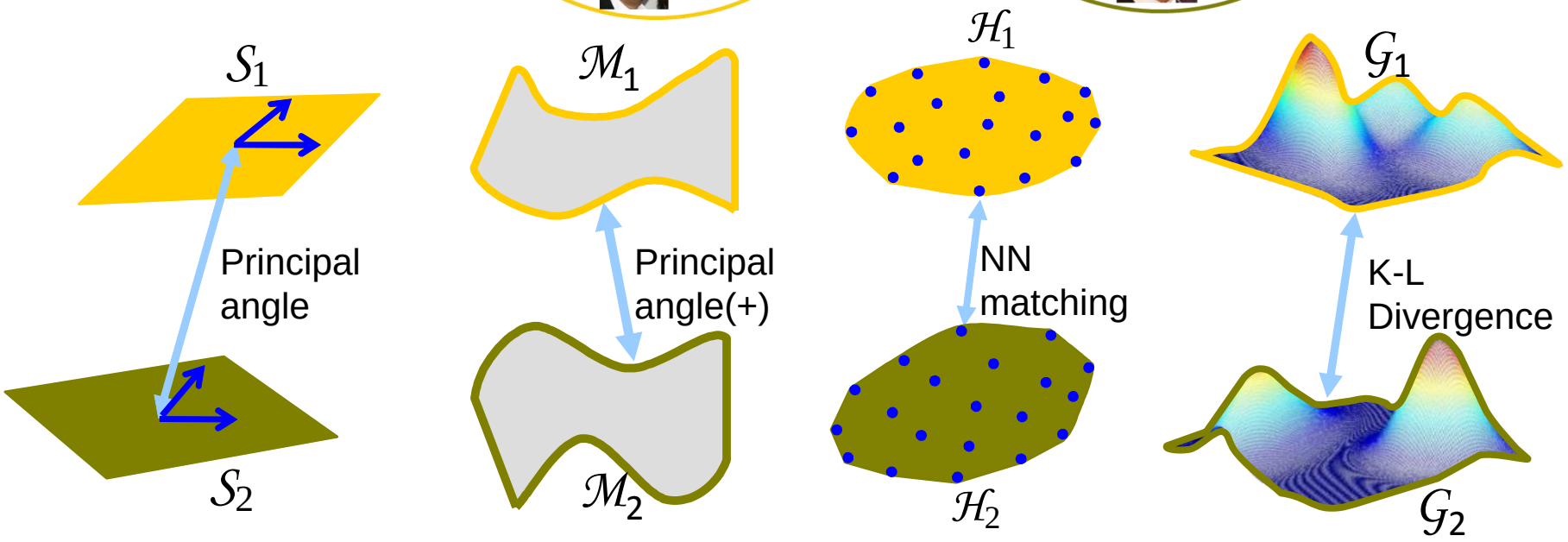
◆Affine/Convex hull
[Cevikalp, CVPR'10]
[Hu, CVPR'11]
[Yang, FG'13]
[Zhu, ICCV'13]
[Wang, ACCV'16]

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

Overview of existing works

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]

Outline

- Problem
- Motivation
- Our methods
- Summary

Background

- Sample similarity
  - Euclidean distance

$$d(x_1, x_2) = \|x_1 - x_2\|_2 = \sqrt{(x_1 - x_2)^T (x_1 - x_2)}$$

  - Mahalanobis distance

$$d_M(x_1, x_2) = \sqrt{(x_1 - x_2)^T \Sigma^{-1} (x_1 - x_2)}$$

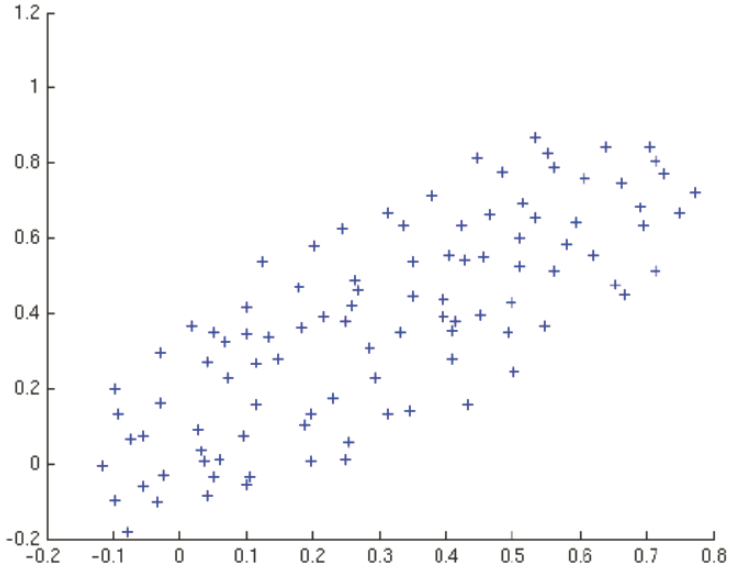
where

$$\Sigma = \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T$$

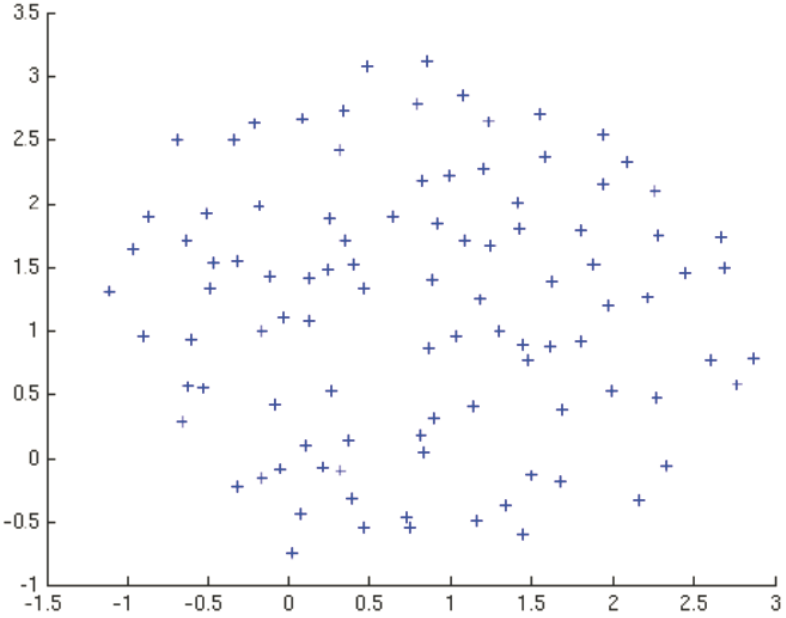
$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Background

- Sample similarity
  - Euclidean distance
  - Mahalanobis distance



Euclidean distance



Mahalanobis distance

Background

- Metric learning
  - 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

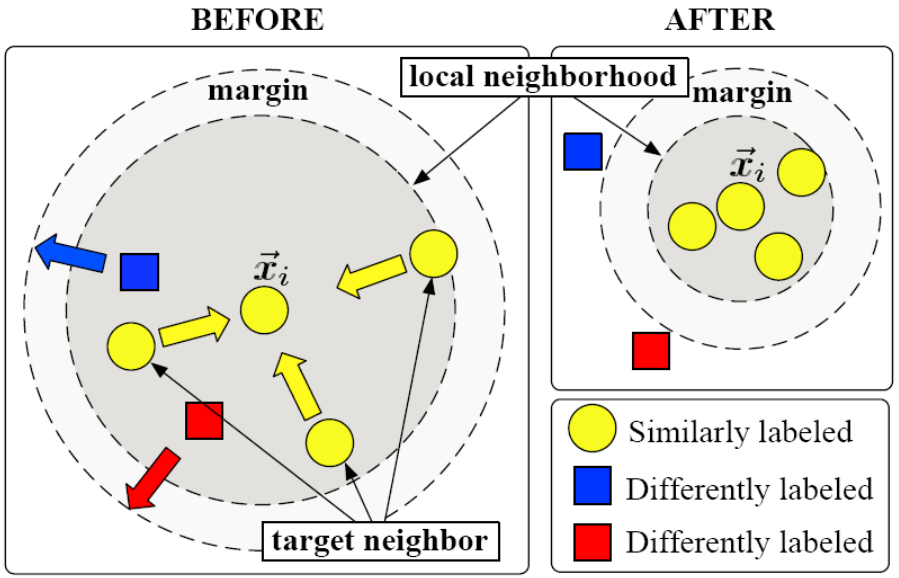
$$d_M(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$$

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

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

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

Background



- Large Margin Nearest I
  - Cost function

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

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

  - Objective function: semidefinite programming (SDP)

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

(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.*

Background

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

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

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

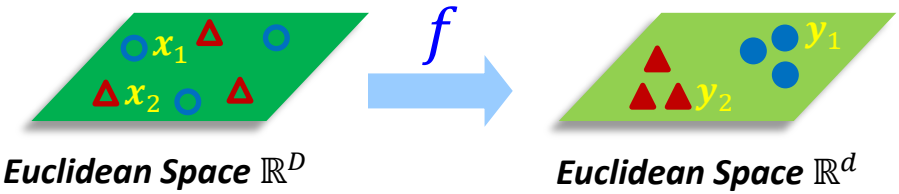
$$\begin{aligned} \min_{A \succcurlyeq 0} \quad & D_{ld}(A, A_0) \\ \text{s.t.} \quad & \mathrm{tr}(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T) \le u \quad (i,j) \in S, \\ & \mathrm{tr}(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T) \ge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

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

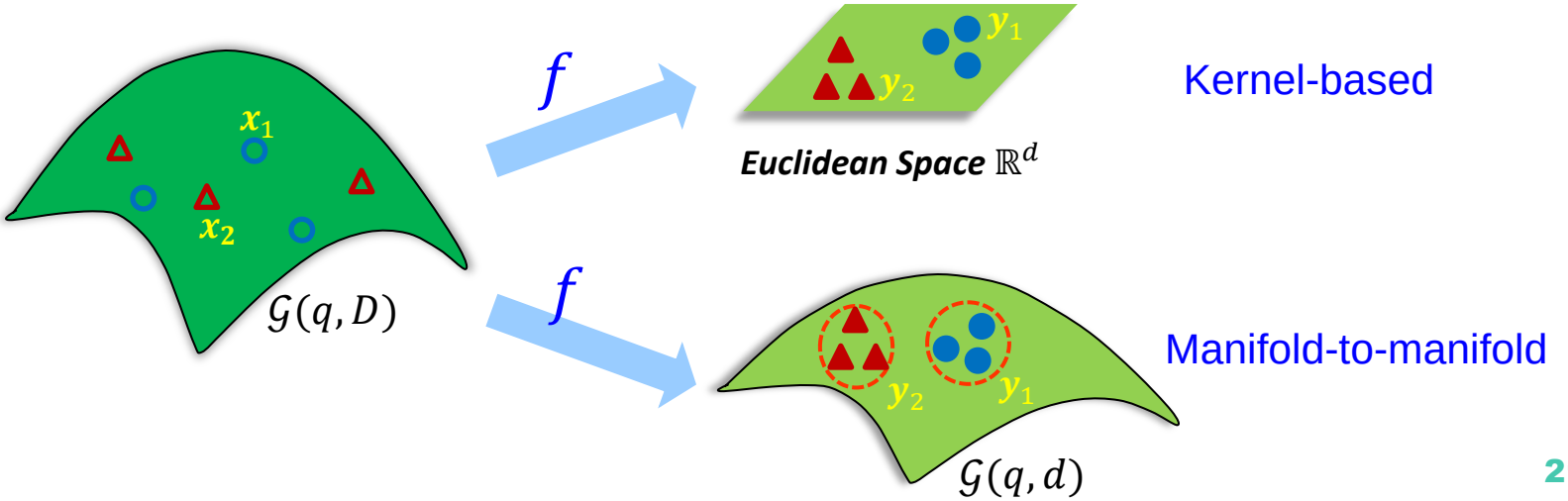
[1] J.V. Davis, B. Kulis, P. Jain, S. Sra, and I.S. Dhillon. Information-Theoretic Metric Learning. *ICML 2007.*

Metric learning: linear vs. nonlinear

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

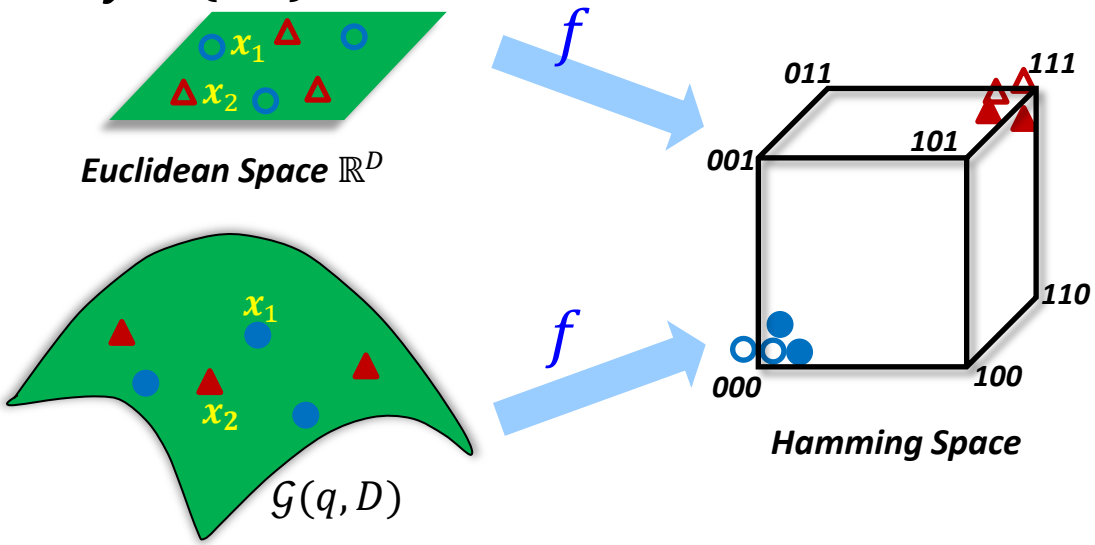


- Nonlinear
  - $f(.)$ is nonlinear mapping, or $\boldsymbol{x}$, $\boldsymbol{y}$ is in non-Euclidean space
  - Riemannian metric learning
    - $\boldsymbol{x} \in \mathcal{M}$ is element on some Riemannian manifold $\mathcal{M}$

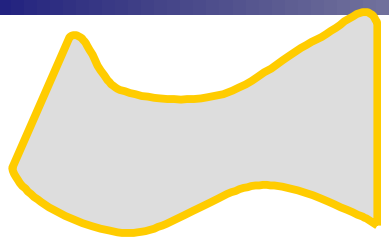


Metric learning: linear vs. nonlinear

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

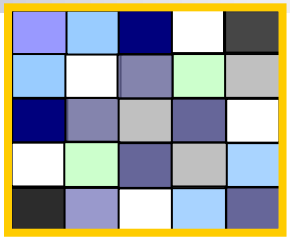


Route map



Appearance manifold

- Complex distribution
- Large amount of data



Covariance matrix

- Natural raw statistics
- No assumption of data

MMD
CVPR'08/TIP'12

mul./mean/disc.

single/cov./ disc.

MDA
CVPR'09

CDL
CVPR'12

heterogeneous

LERM
CVPR'14

binary

HER
CVPR'15

binary

CVC
BMVC'14

heterogeneous

+COV

mul. / +mean

Rie. ML

deep

deep

DARG
CVPR'15

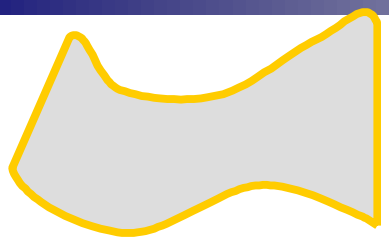
LEML/PML
ICML'15/CVPR'15

DCRL
CVPR'17

binary

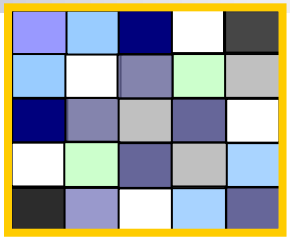
DHH
ICCV'17

Route map



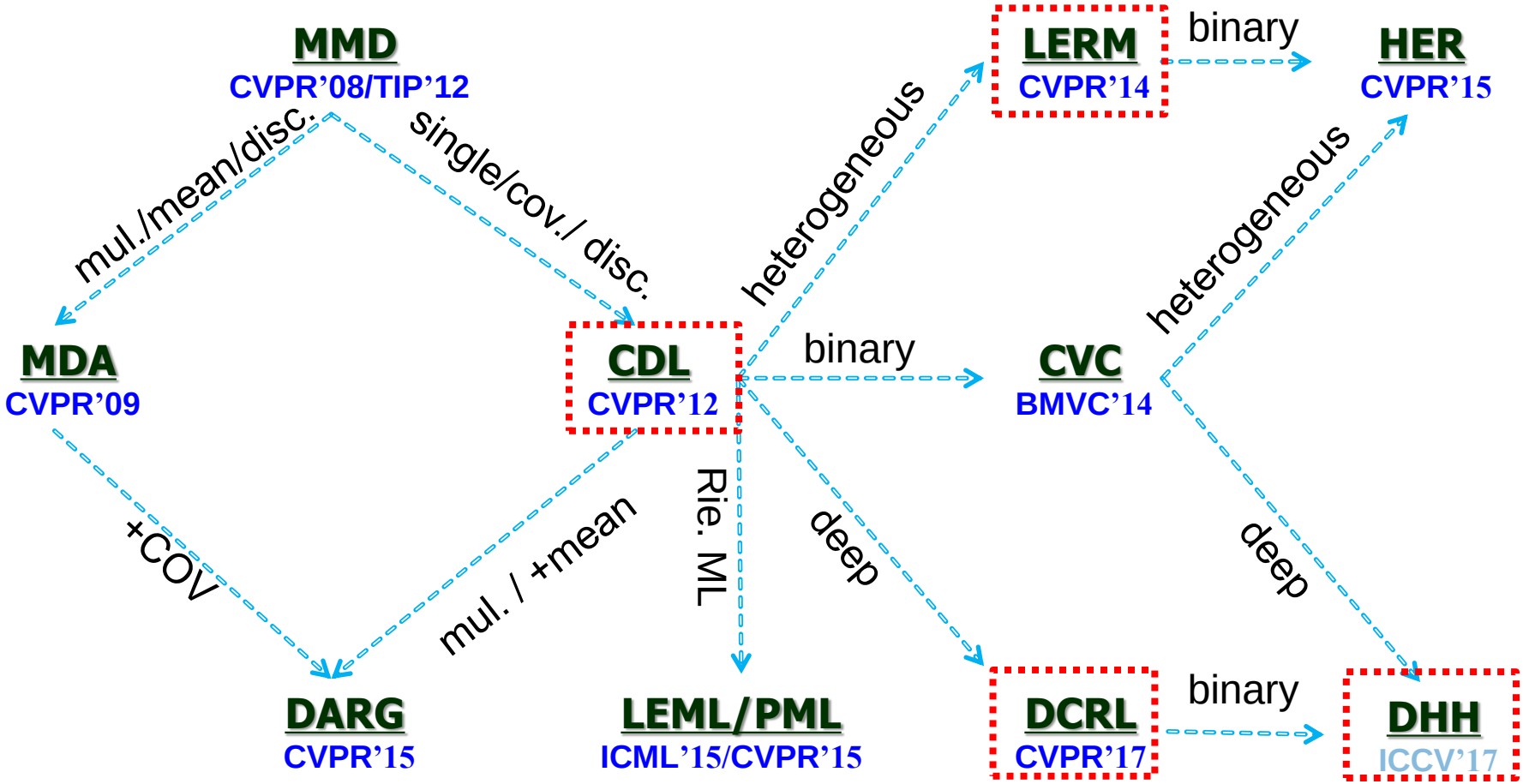
Appearance manifold

- Complex distribution
- Large amount of data



Covariance matrix

- Natural raw statistics
- No assumption of data



Covariance?

Set modeling by Covariance Matrix

- Image set: *N* image samples
- Feature vector: *D*-dimension, any type of features
  - Intensity, Gabor, LBP, HOG, SIFT, etc.
- COV: *D*D* symmetric positive definite(SPD) matrix
  - *N>D*? OR *N<D*?

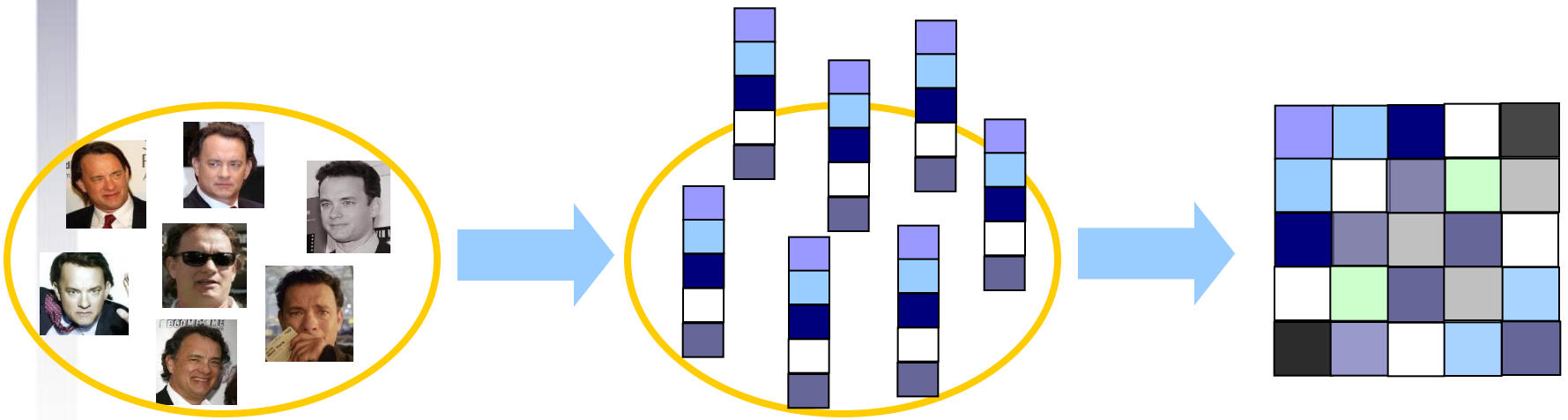


image set　　feature vectors　　covariance matrix

Covariance?

The power of Covariance Matrix (COV)

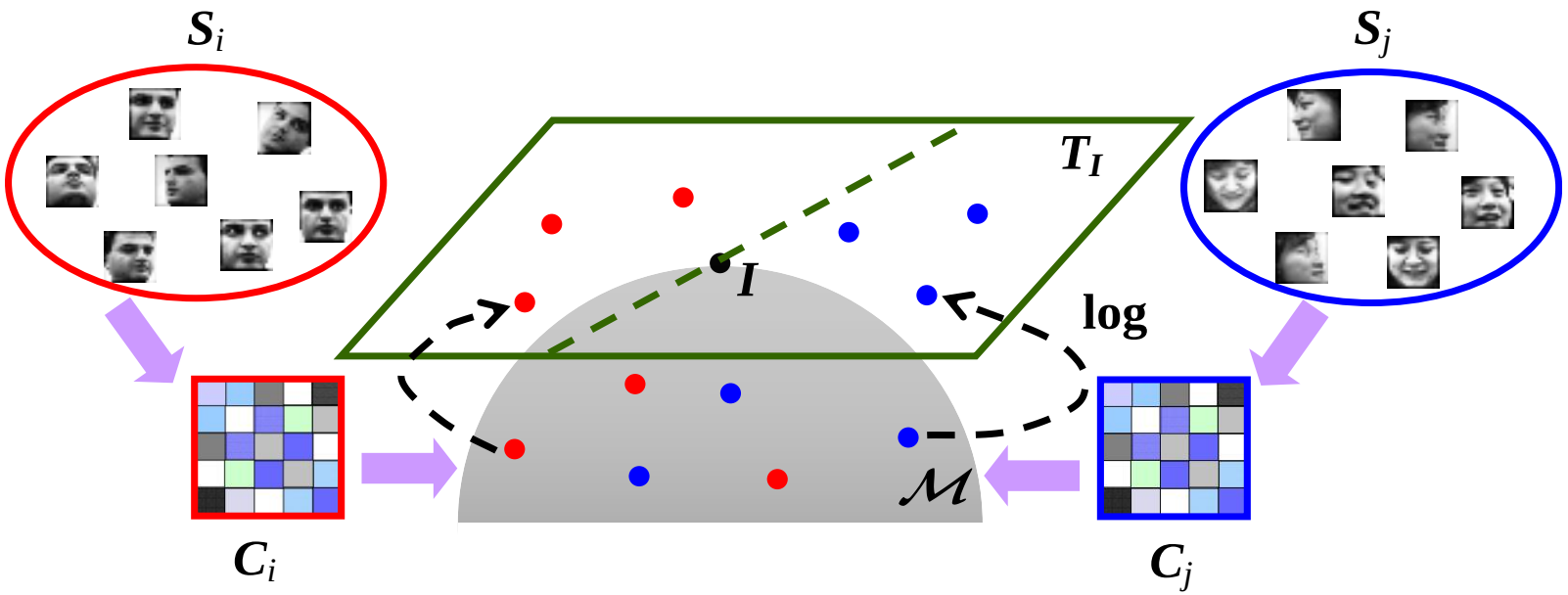
- The natural 2nd-order statistics of a sample set
- Makes NO assumption of the data distribution
- Robust to noisy set data
- Scalable to varying set size
- As region descriptor, COV has been successfully applied in texture classification, object detection & tracking, etc.*

[1] O. Tuzel, F. Porikli, and P. Meer. Region Covariance: A Fast Descriptor for Detection and Classification. *ECCV 2006.*

[2] O. Tuzel, F. Porikli, and P. Meer. Pedestrian Detection via Classification on Riemannian Manifolds. *IEEE Trans. PAMI 2008.*

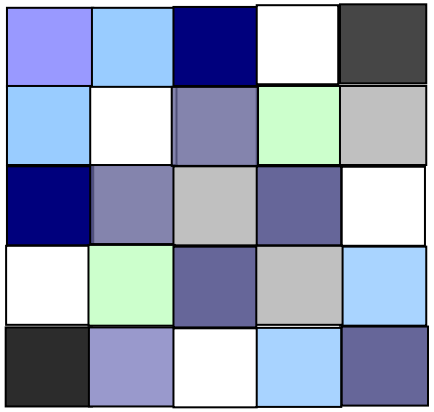
Covariance Discriminative Learning (CDL)*

- Problem
  - Video-to-Video face recognition
  - Classification on SPD Riemannian manifold



[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 intensity feature

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

◆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]

$$d^2(\boldsymbol{C}_1,\boldsymbol{C}_2)=\sum_{i=1}^{d}\ln^2\lambda_i(\boldsymbol{C}_1,\boldsymbol{C}_2)$$

or

$$d^2(\boldsymbol{C}_1,\boldsymbol{C}_2)=\left\|\log_{\boldsymbol{I}}(\boldsymbol{C}_1^{-1/2}\boldsymbol{C}_2\boldsymbol{C}_1^{-1/2})\right\|_F^2$$

High computational burden

    - Log-Euclidean distance (LED) [2]

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

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.

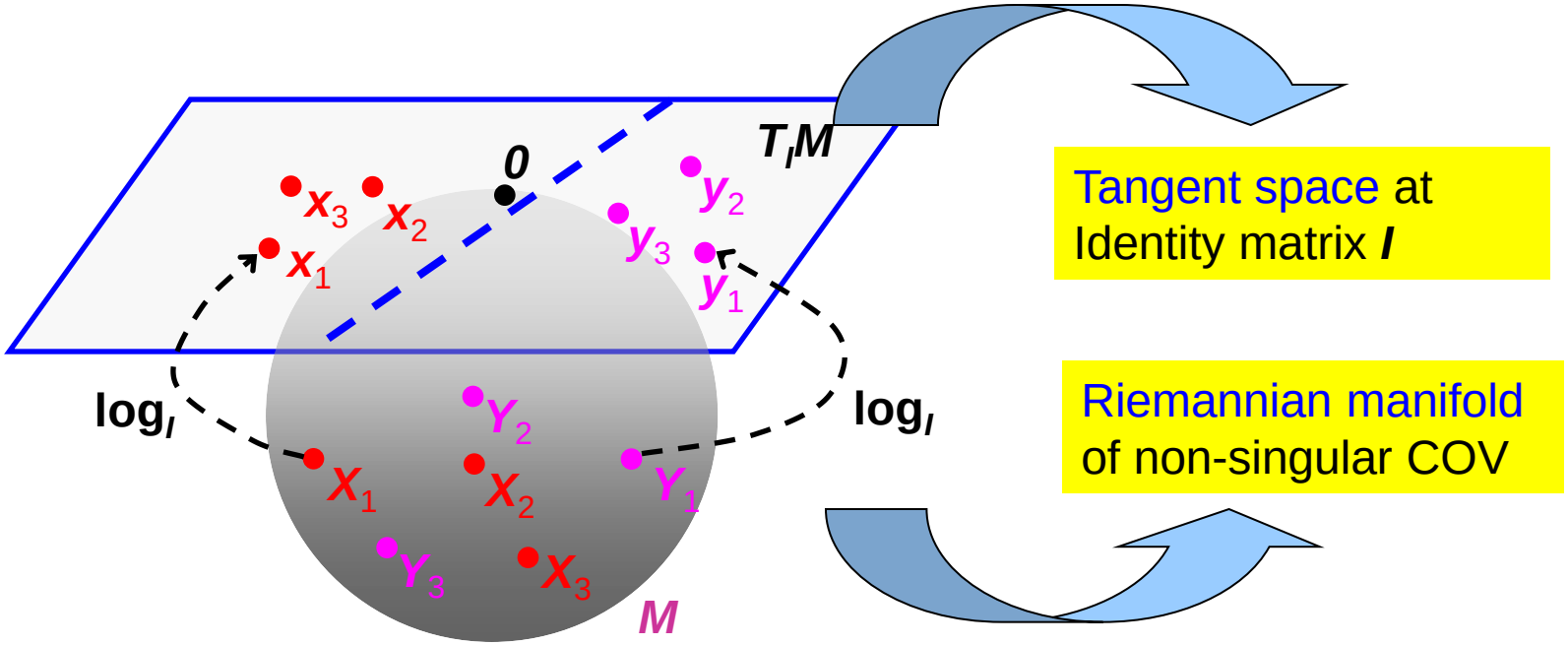
CDL

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

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

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

Mercer's theorem

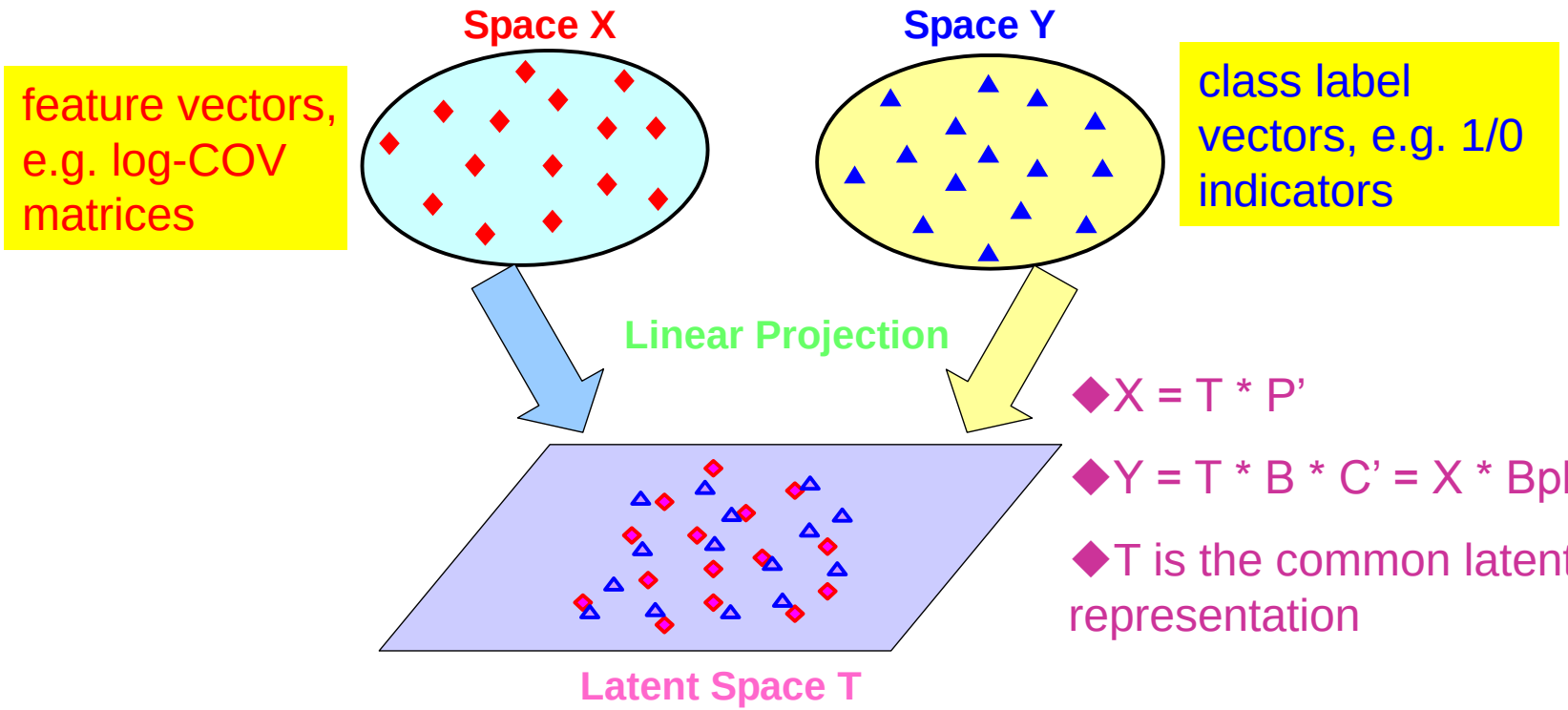


Tangent space at Identity matrix ***I***

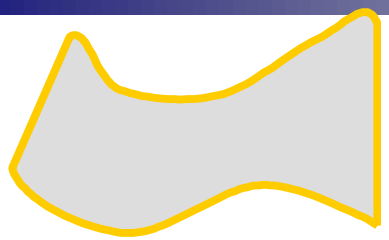
Riemannian manifold of non-singular COV

CDL

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

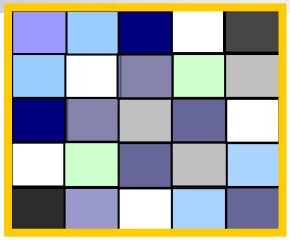


Route map



Appearance manifold

- Complex distribution
- Large amount of data



Covariance matrix

- Natural raw statistics
- No assumption of data

MMD
CVPR'08/TIP'12

MDA
CVPR'09

CDL
CVPR'12

LERM
CVPR'14

HER
CVPR'15

CVC
BMVC'14

DARG
CVPR'15

LEML/PML
ICML'15/CVPR'15

DCRL
CVPR'17

DHH
ICCV'17

mul./mean/disc.

single/cov./ disc.

heterogeneous

binary

heterogeneous

+COV

mul. / +mean

Rie. ML

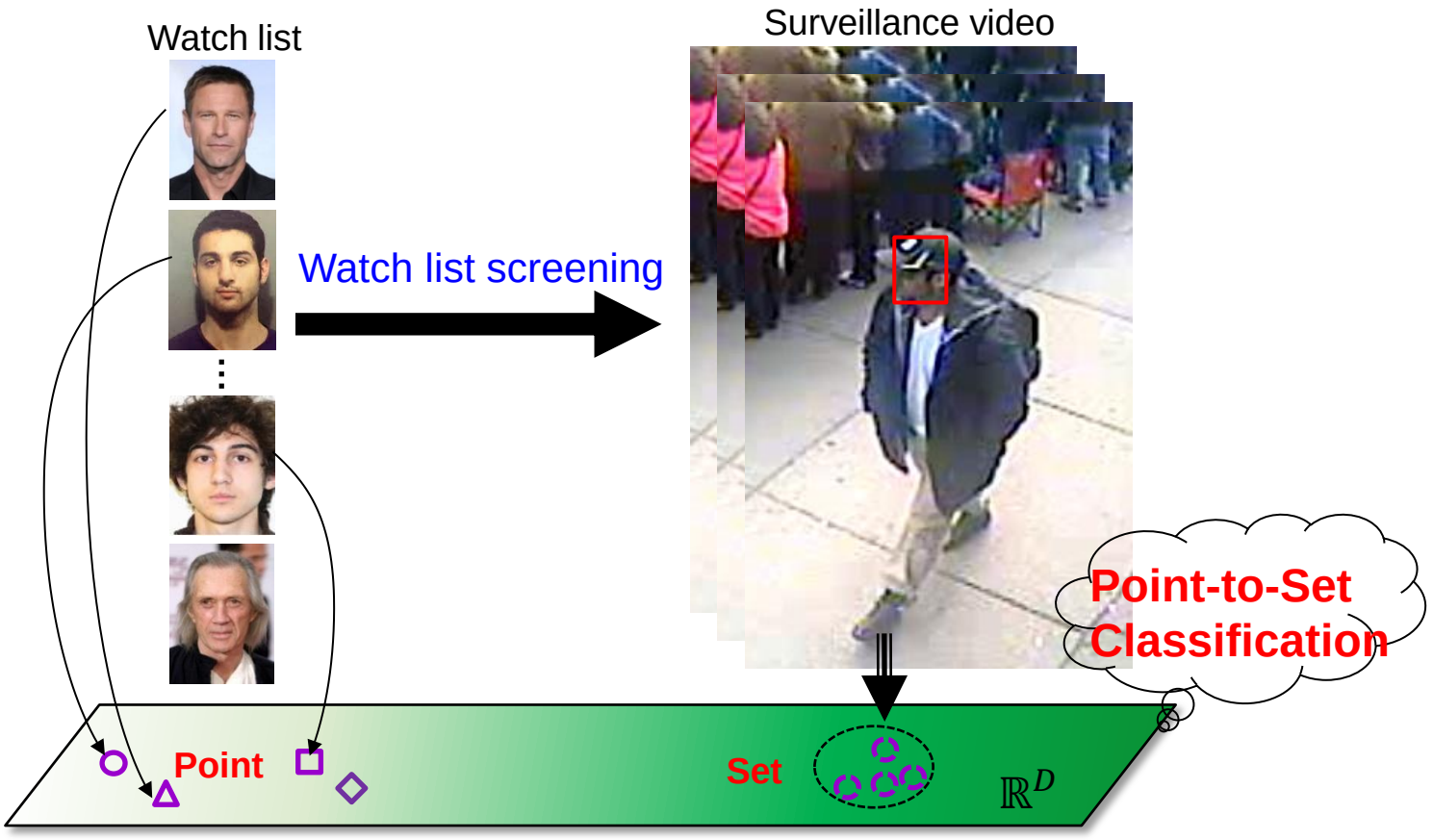
deep

deep

binary

Learning Euclidean-to-Riemannian Metric (LERM)*

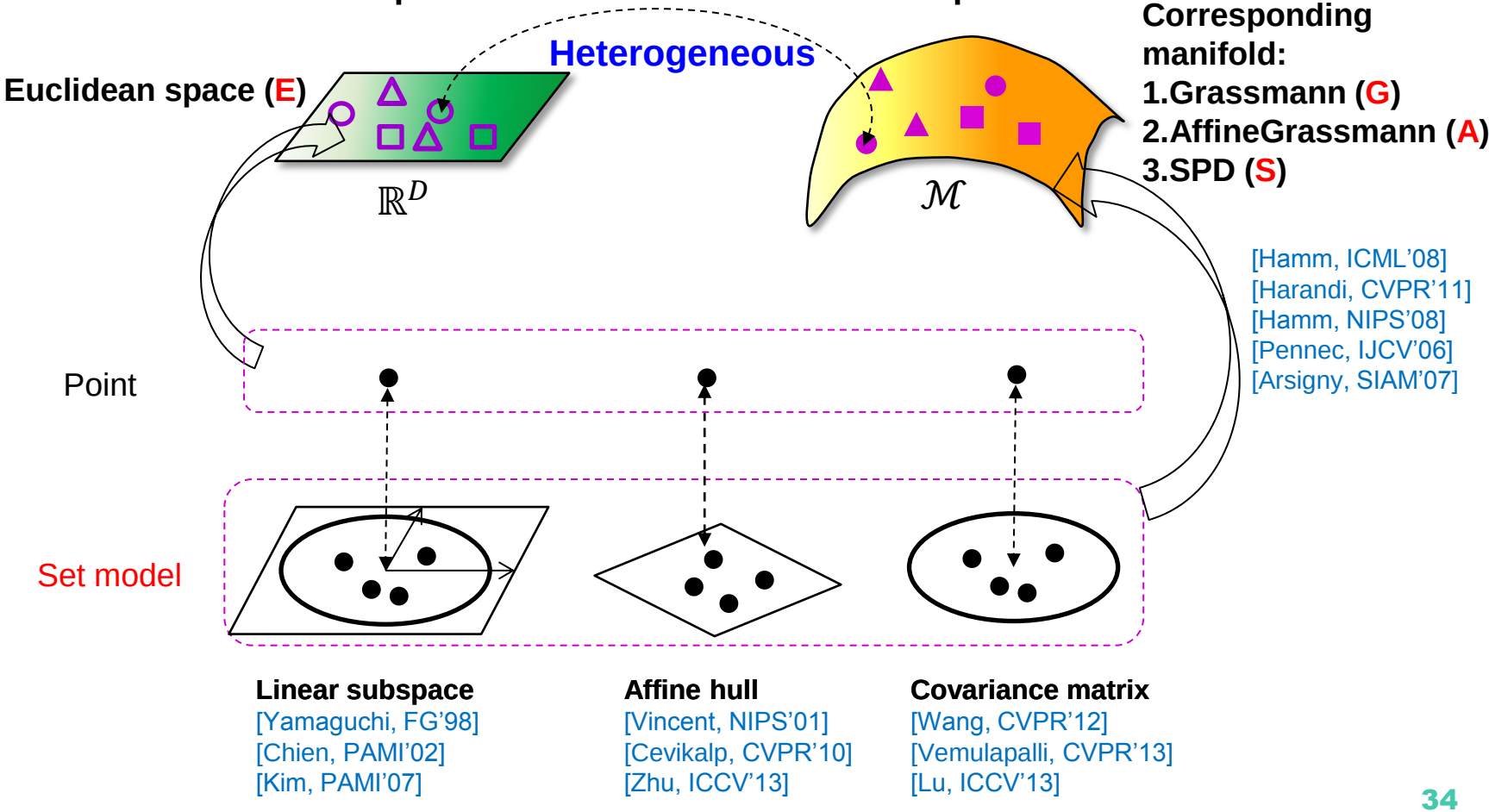
- Problem: Still-to-Video face recognition



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

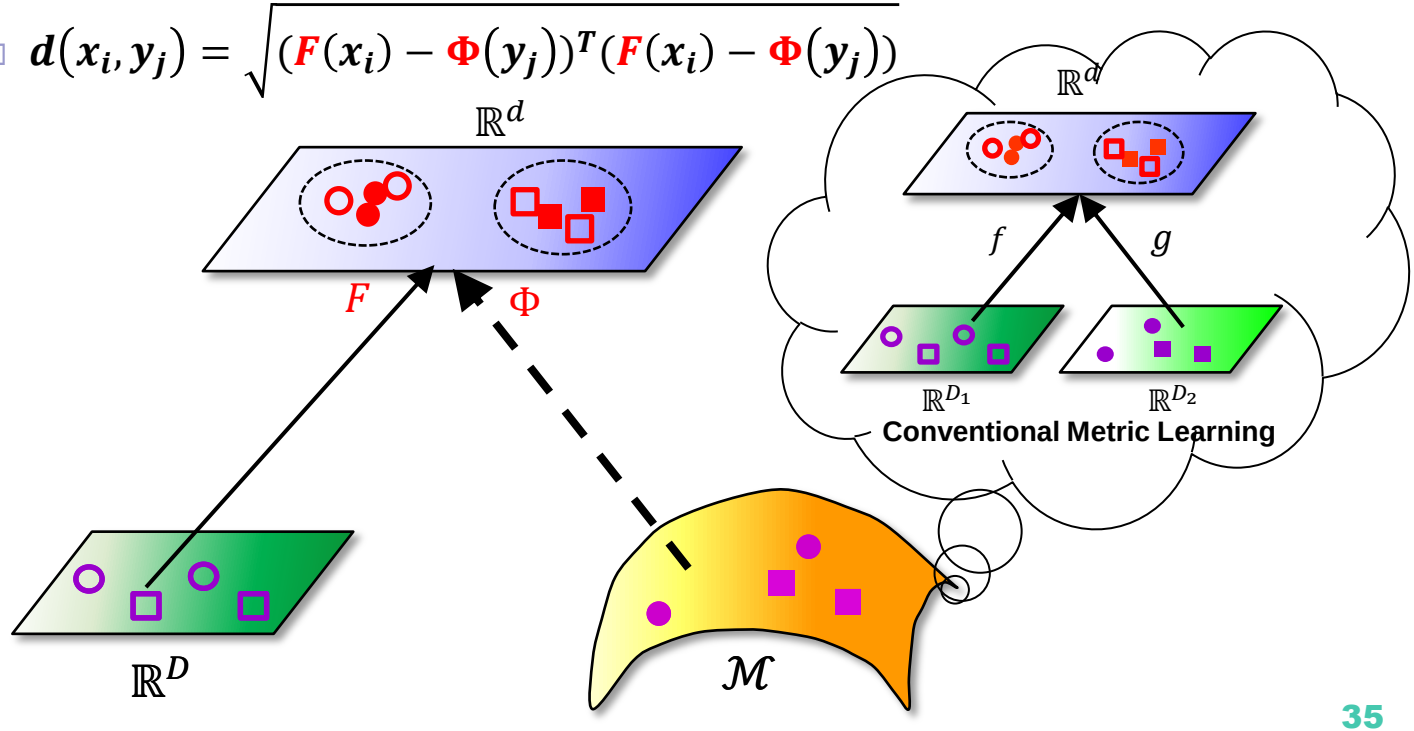
LERM

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



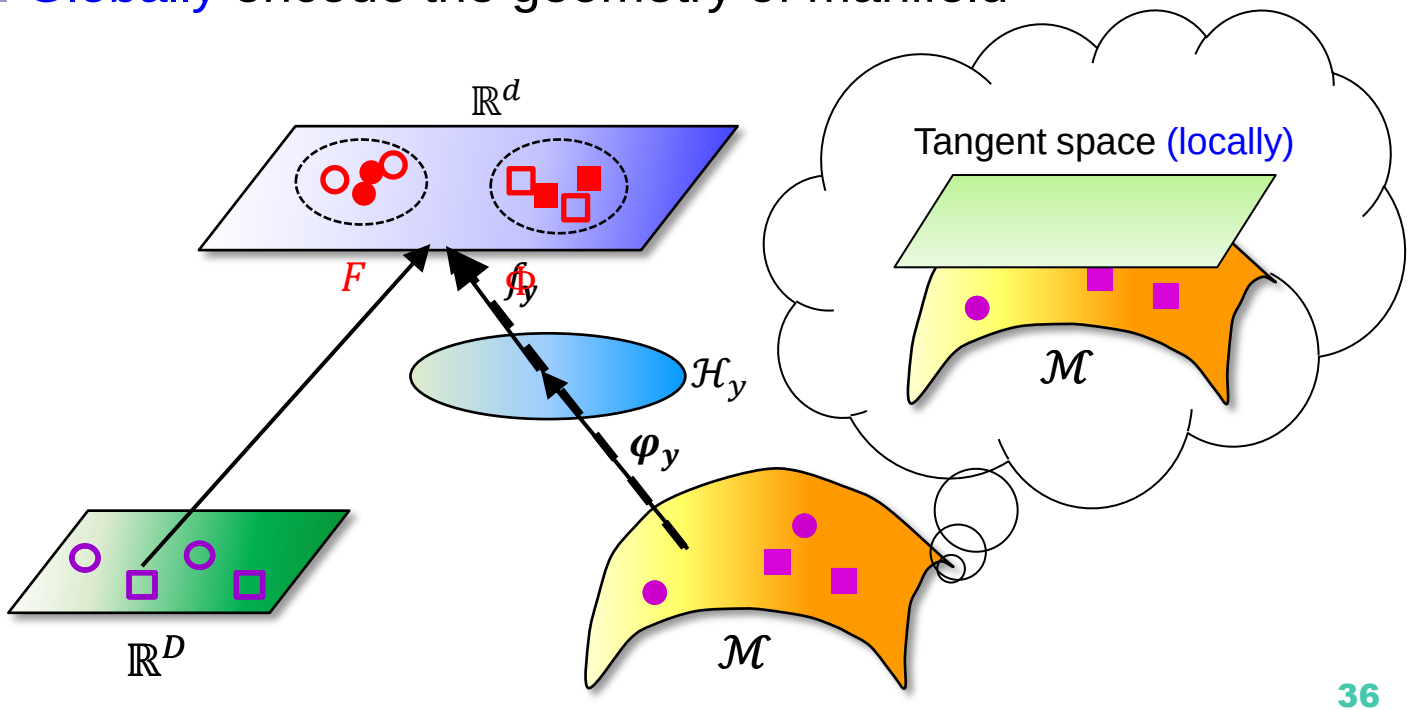
LERM

- 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

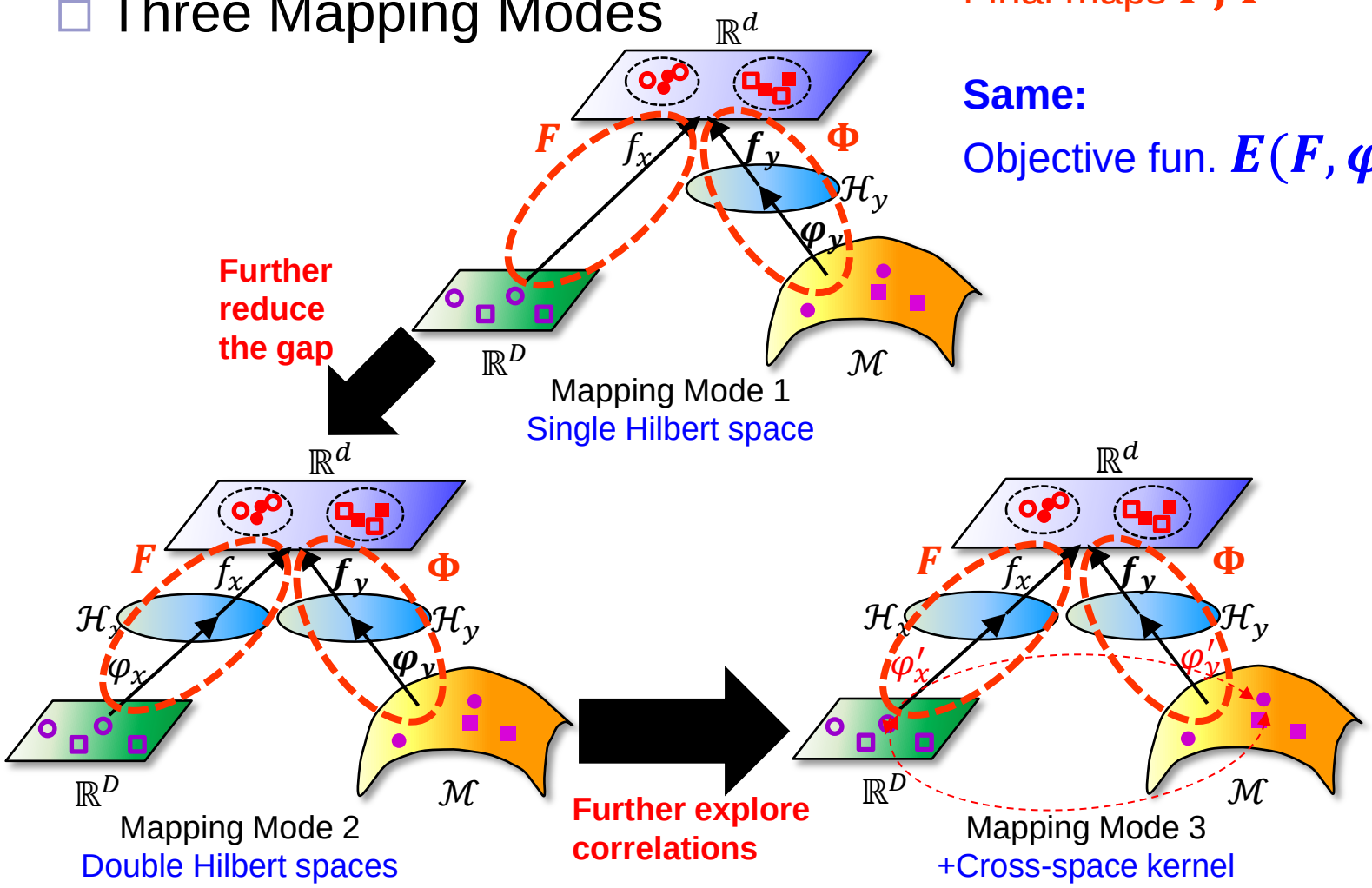
- Formulation
  - Three Mapping Modes

Different:

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

Same:

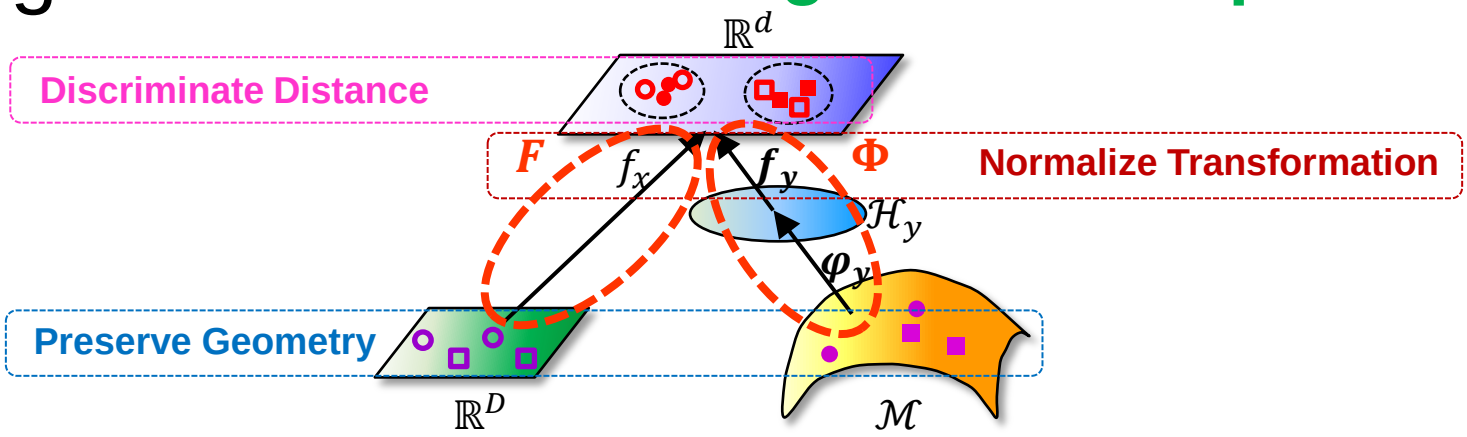
Objective fun. $\boldsymbol{E(F, \varphi)}$



LERM

- e.g. Mode 1

Single Hilbert space



Final maps:

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

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

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

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

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

Distance metric:

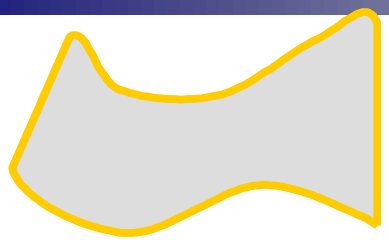
$$d(x_i, y_j) = \sqrt{(F(x_i) - \Phi(y_j))^T (F(x_i) - \Phi(y_j))}$$

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

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

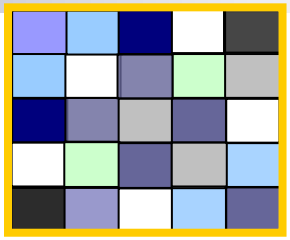
Distance Geometry Transformation

Route map



Appearance manifold

- Complex distribution
- Large amount of data



Covariance matrix

- Natural raw statistics
- No assumption of data

MMD
CVPR'08/TIP'12

MDA
CVPR'09

CDL
CVPR'12

LERM
CVPR'14

HER
CVPR'15

CVC
BMVC'14

DARG
CVPR'15

LEML/PML
ICML'15/CVPR'15

DCRL
CVPR'17

DHH
ICCV'17

mul./mean/disc.

single/cov./ disc.

heterogeneous

binary

heterogeneous

+COV

mul. / +mean

Rie. ML

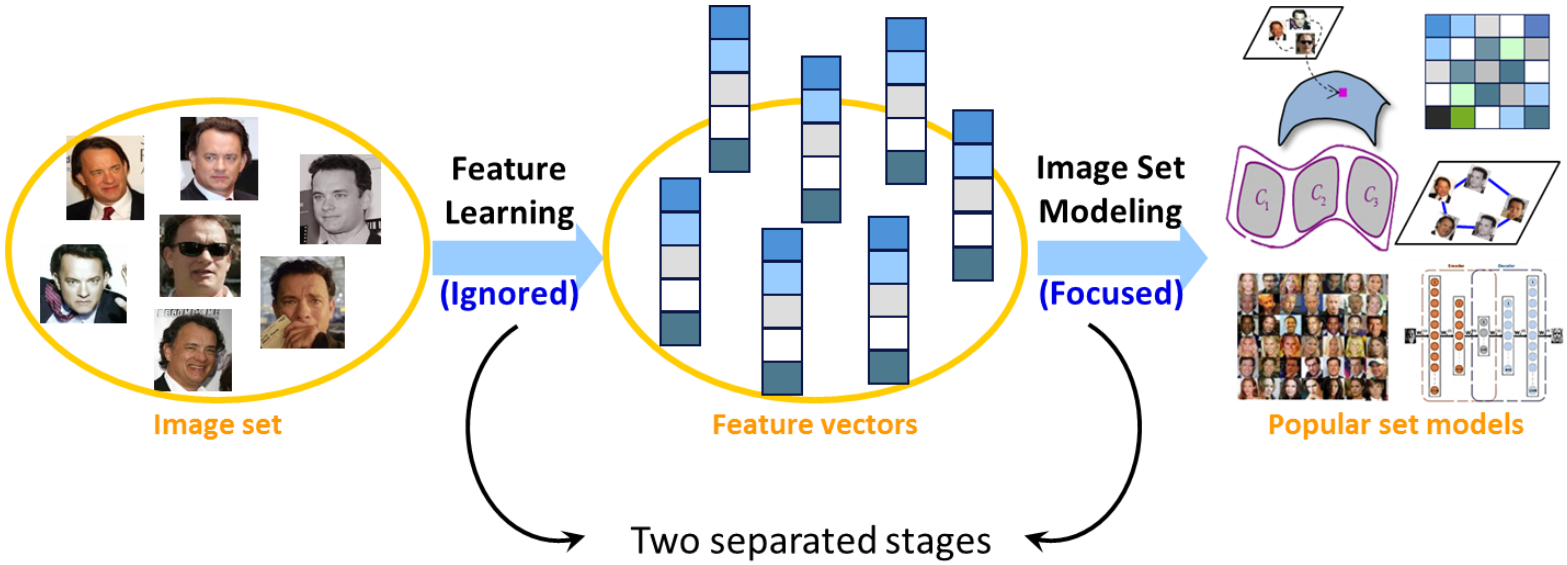
deep

deep

binary

Discriminative Covariance oriented Representation Learning (DCRL)*

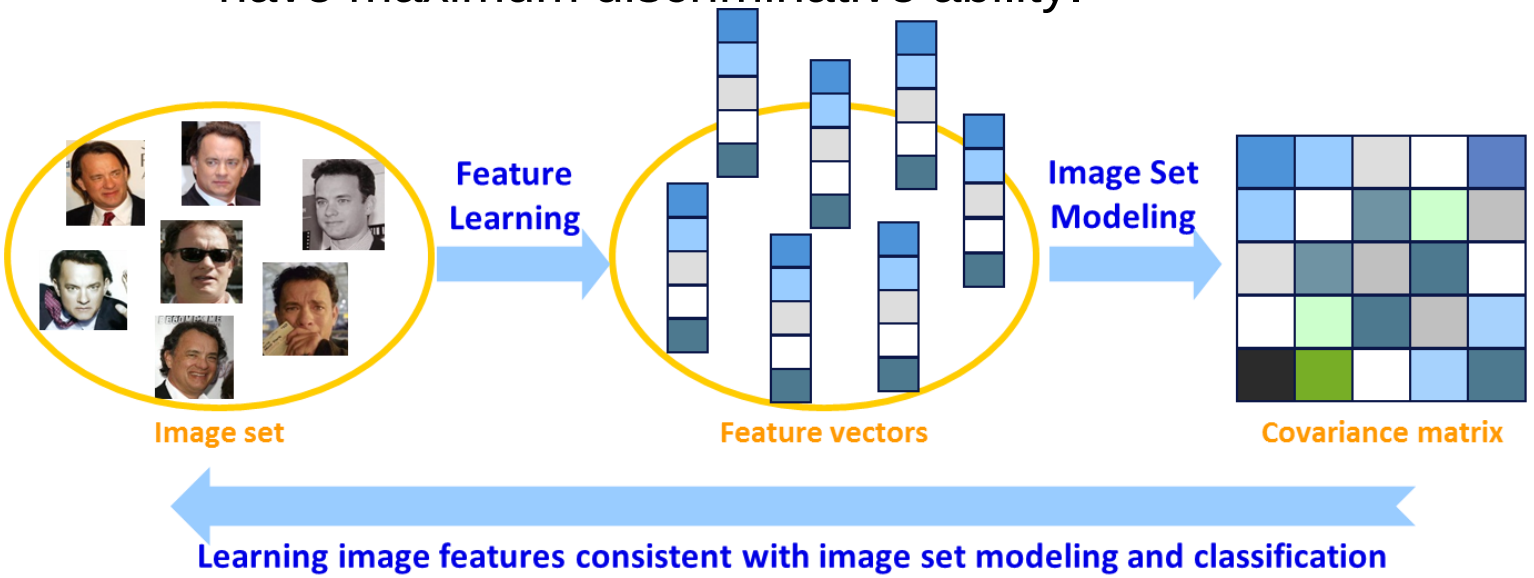
- Problem
  - Video-to-Video face recognition
  - Image feature learning that facilitates image set modeling and classification



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

DCRL

- Basic idea
  - Image feature learning
    - Deep learning networks, e.g., CNN
  - Image set modeling
    - Set covariance matrices
  - Objective
    - In the learned image feature space, set covariance matrices have maximum discriminative ability.



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

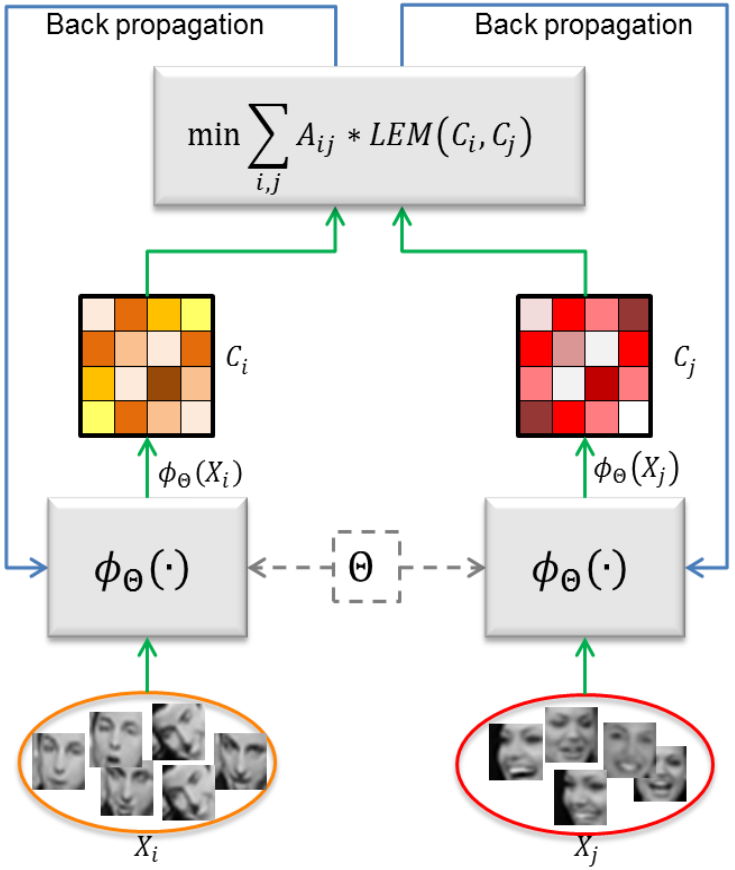
- Graph Embedding Scheme
  - Loss function

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

where

$$LEM(\boldsymbol{C}_i, \boldsymbol{C}_j) = \left\| \log_{\boldsymbol{I}}(\boldsymbol{C}_i) - \log_{\boldsymbol{I}}(\boldsymbol{C}_j) \right\|_F$$

is the Log-Euclidean Metric (LEM)*



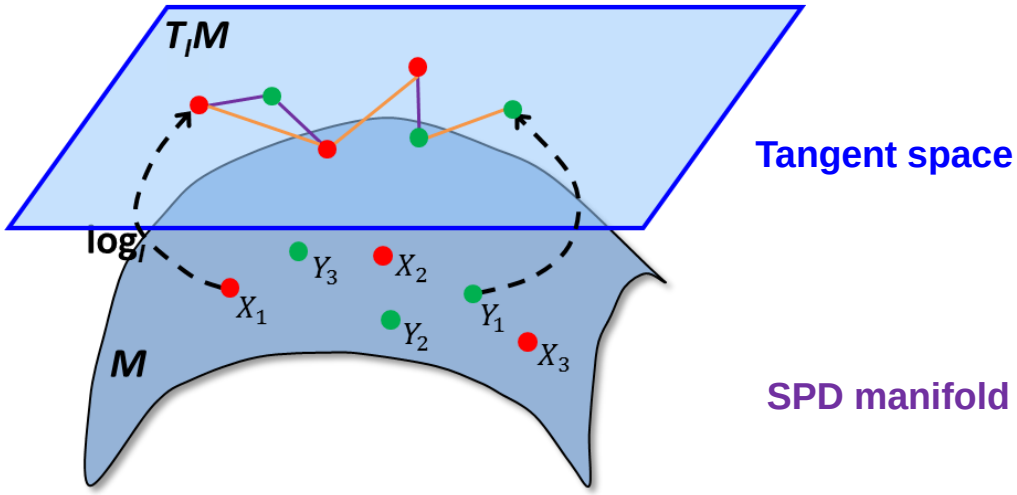
[1] V. Arsigny, P. Fillard, X. Pennec and N. Ayache. Geometric Means In A Novel Vector Space Structure On Symmetric Positive-Definite Matrices. *SIAM J. MATRIX ANAL. APPL*. Vol. 29, No. 1, pp. 328-347, 2007.

DCRL

- Graph Embedding Scheme
  - Loss function
    - Adjacency Graph: Encode the data structure and semantic relationship of set covariance matrices

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

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



DCRL

- Softmax Regression Scheme
  - Loss function
    - Softmax regression

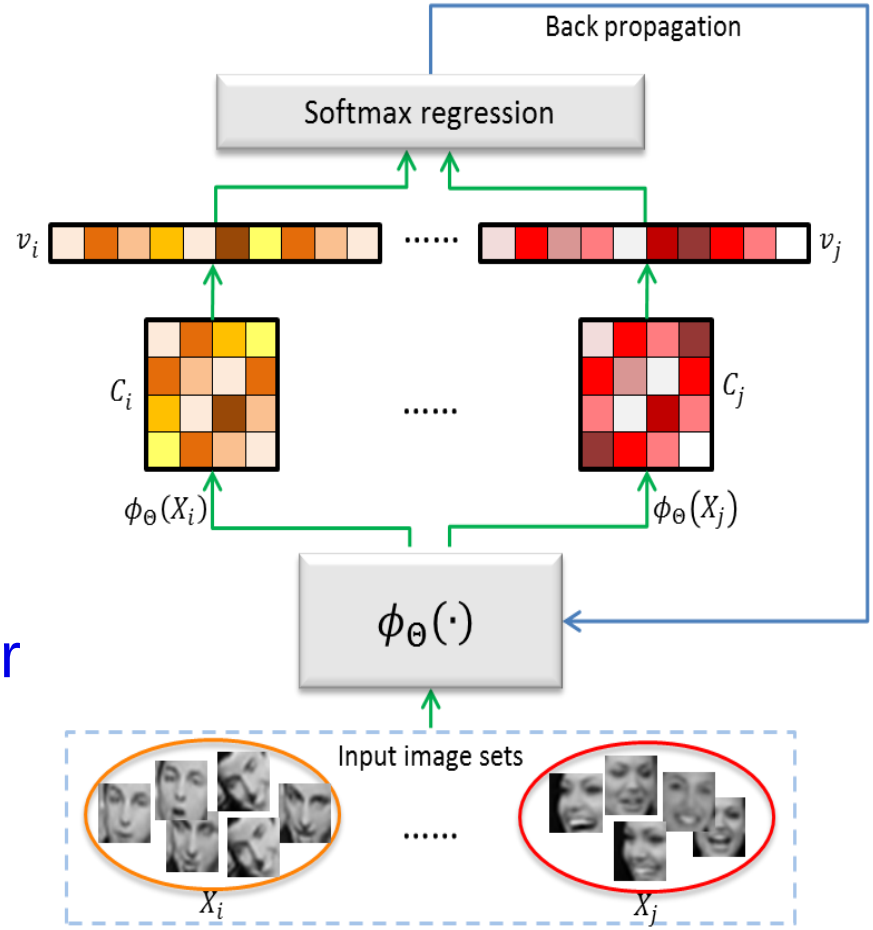
$$J(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} 1\{y_i = j\}\log(o_{ij})$$

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

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

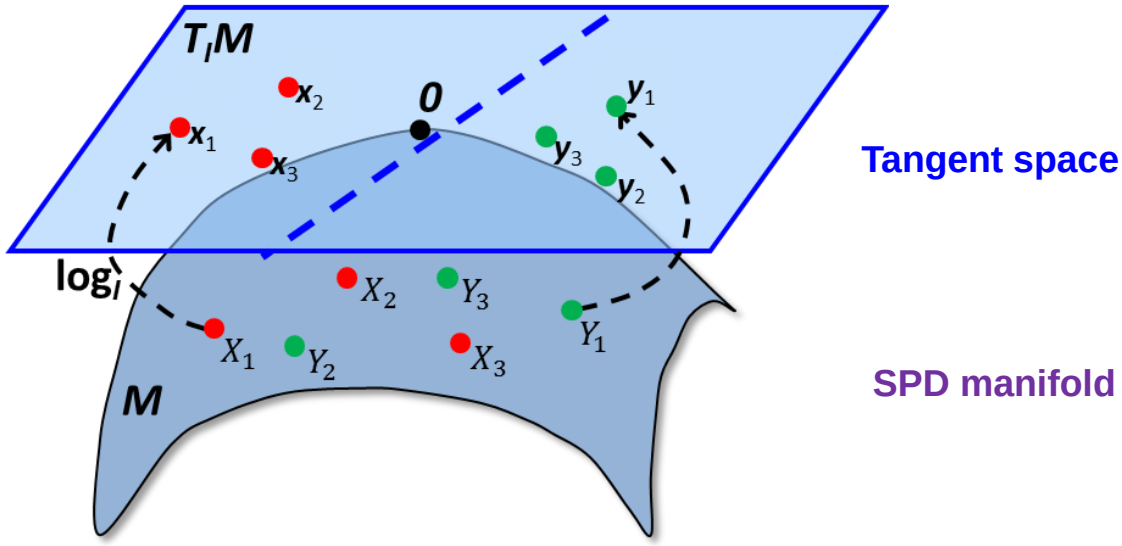
    - log-covariance vector

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



DCRL

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



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

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.

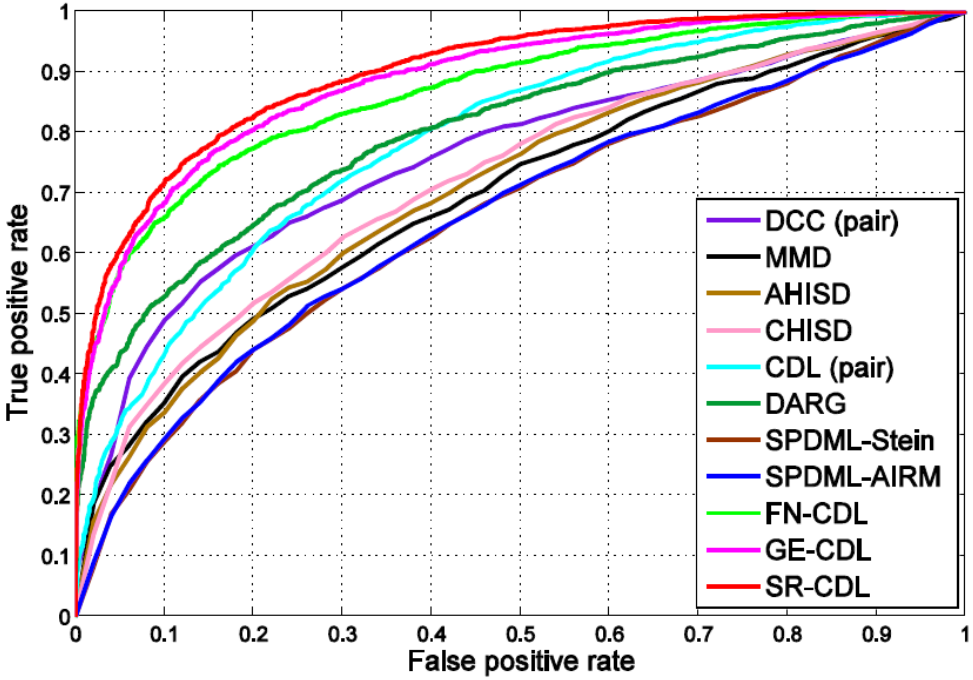
Number of Subjects	265
Total Videos	2,802
Total Control Videos	1,401
Total Handheld Videos	1,401
Control Videos per Subject	4 to 7
Handheld Videos per Subject	4 to 7
Number of Locations	6

Control video

Handheld video

Evaluations

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

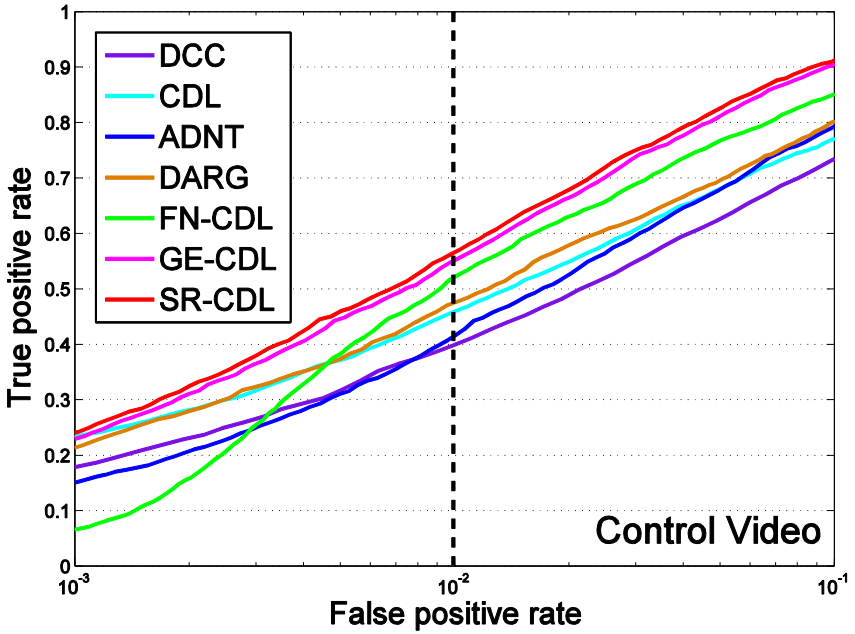


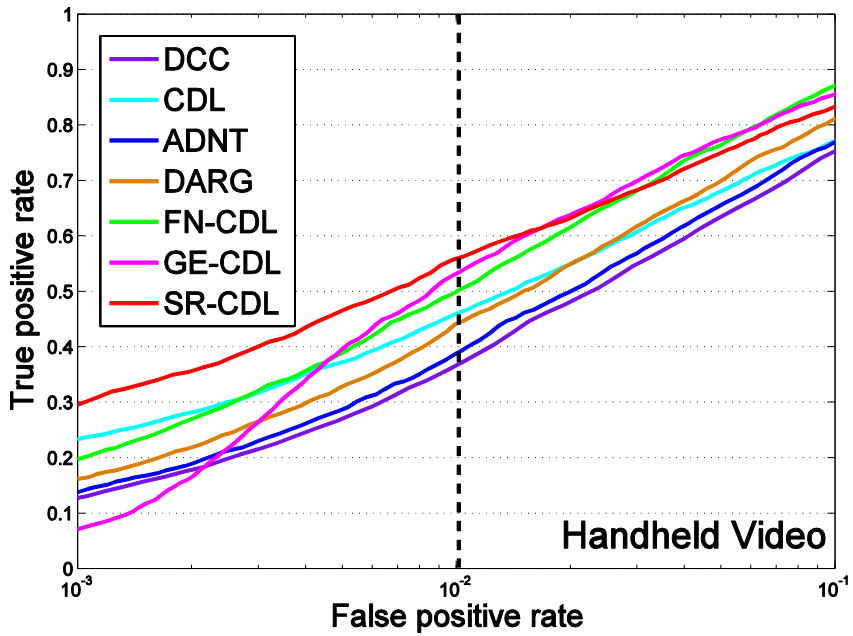
GE: Graph embedding scheme
SR: Softmax regression scheme
FN: Baseline Deep ID net with single CNN

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

Evaluations

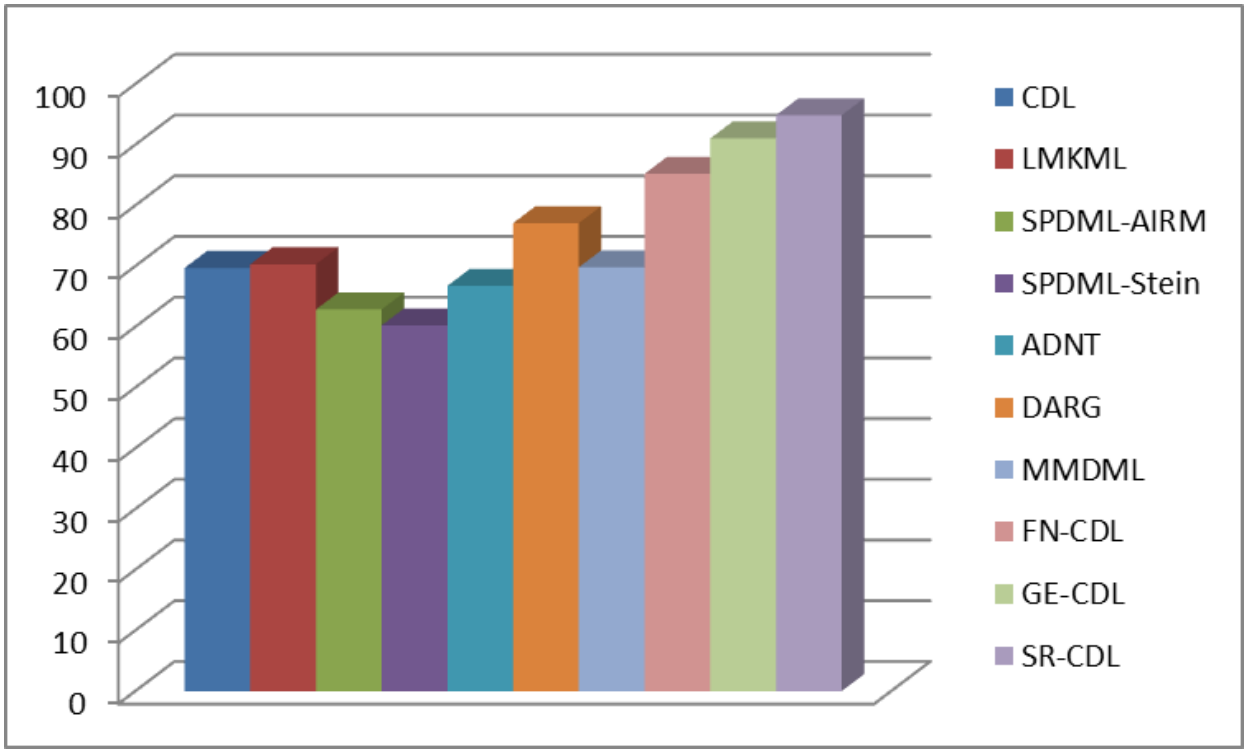
- Results (*reported in our DCRL paper**)
  - Verification task (on PaSC dataset)





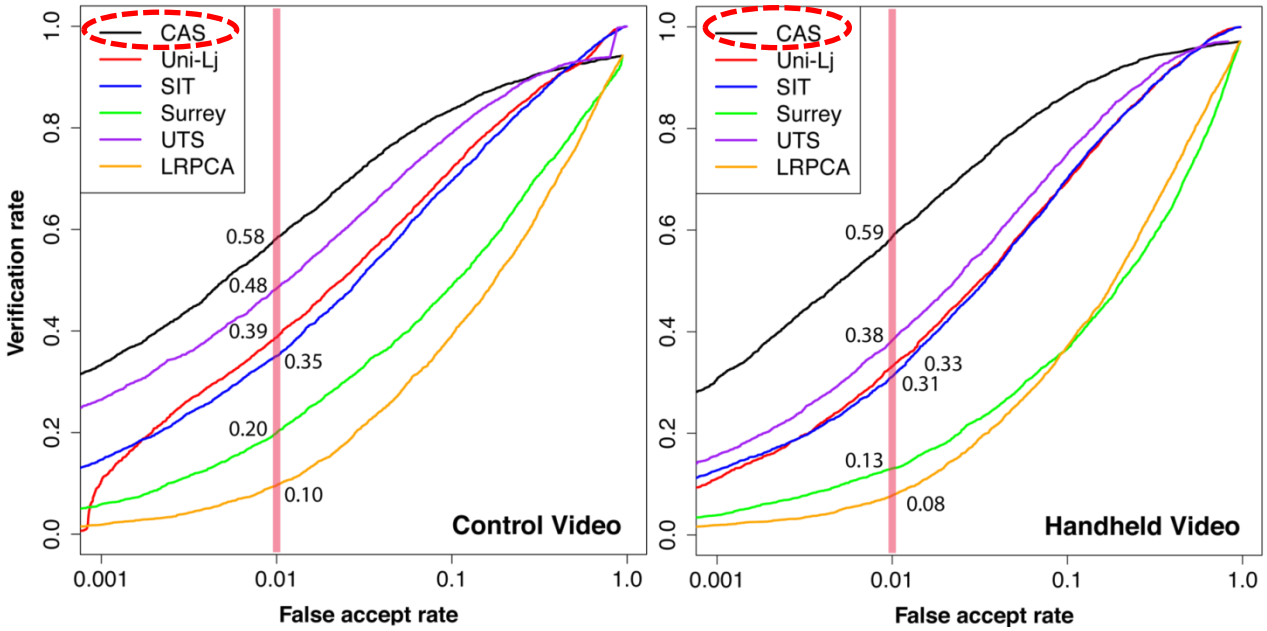
Evaluations

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



Evaluations

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

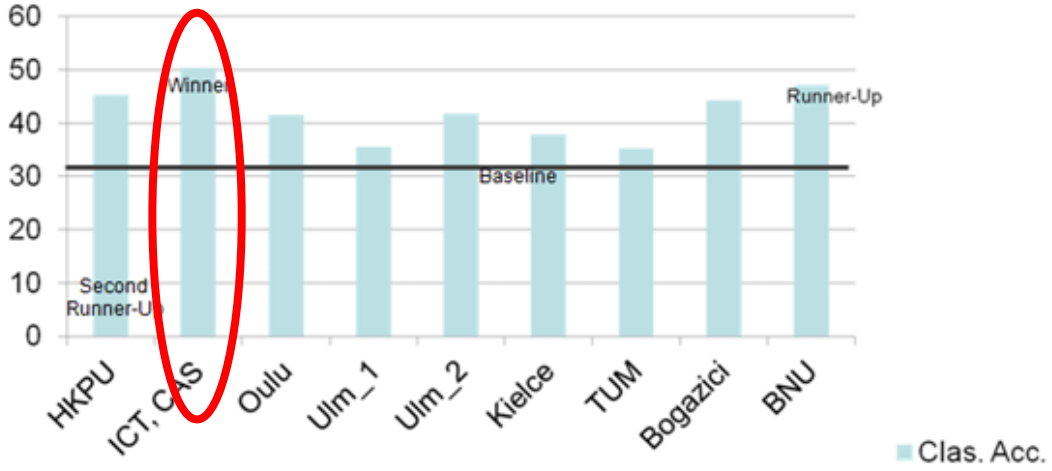


[*] J. Ross Beveridge *et al*. Report on the FG 2015 Video Person Recognition Evaluation. *IEEE FG 2015*.

[**] 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

- 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*.

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?

From face to object

- Content-based Image Retrieval
  - Performance metric: mAP, precision-recall

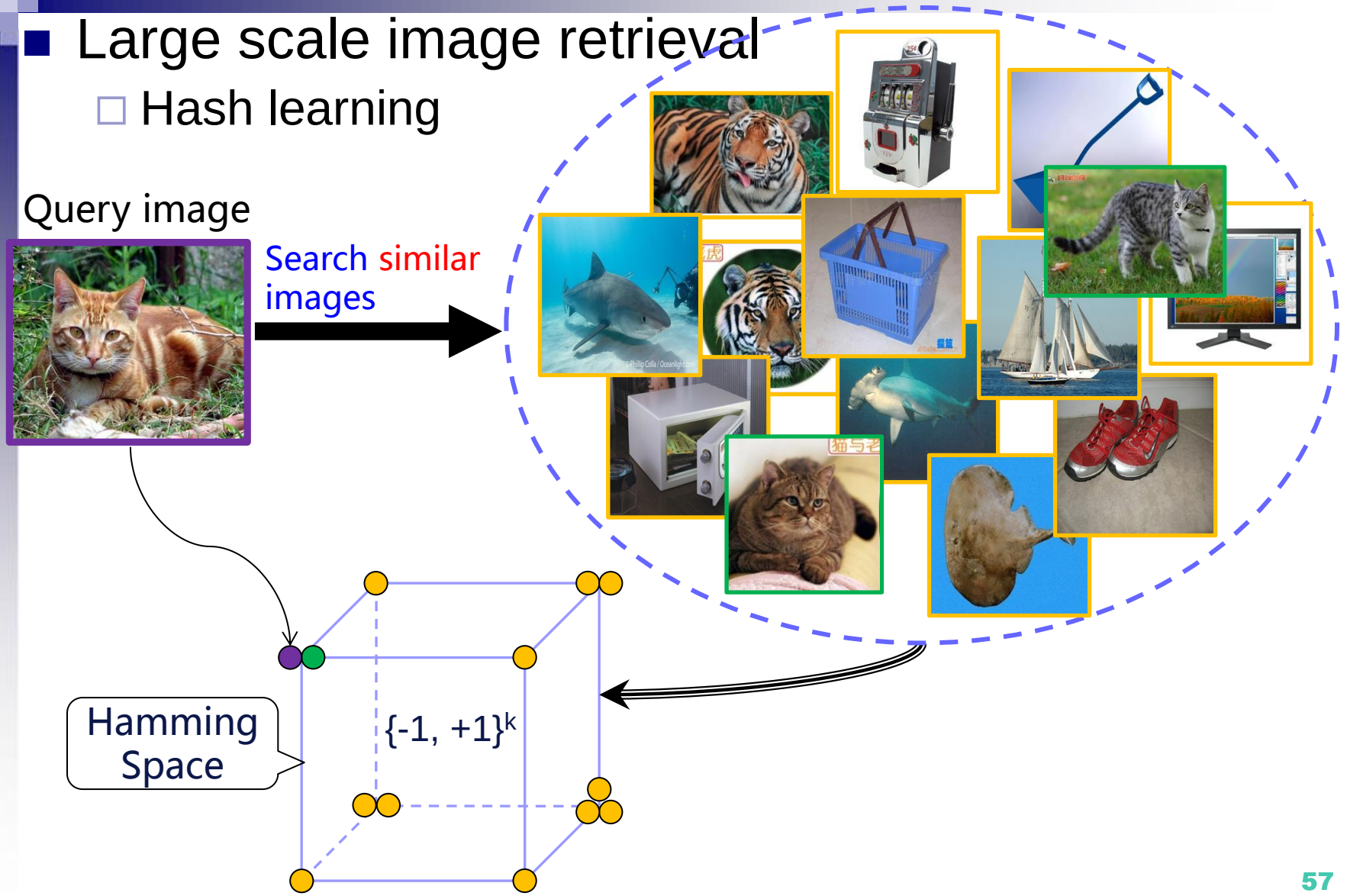
Query image

Image dataset

Top retrieved image

From face to object

- Large scale image retrieval
  - Hash learning

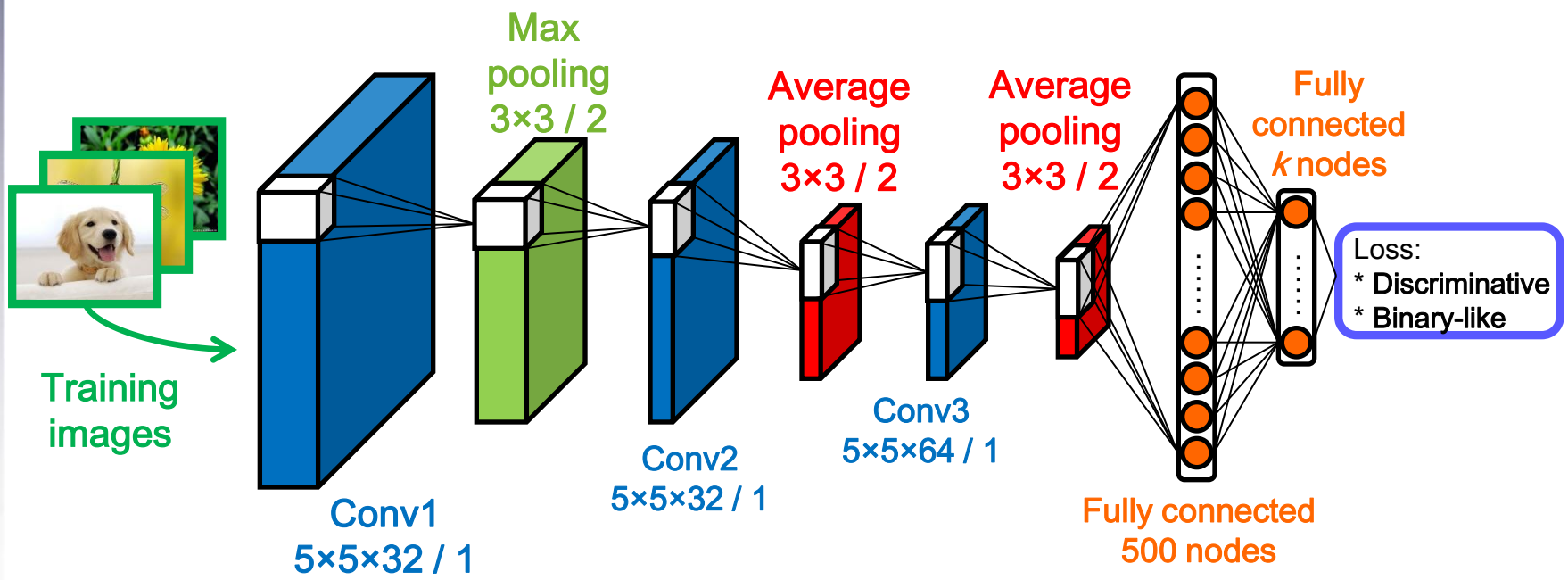


CNN-based Deep Hashing

- Low storage cost
  - Compact binary codes
- Fast retrieval
  - Hash table lookup
  - Highly efficient bit operations
- Robustness
  - Non-linear hash functions
  - Feature learning using CNNs

Deep Supervised Hashing (DSH)*

- Basic idea
  - Train CNN models with pairwise similarity constraints to learn discriminative binary-like image representations.
  - Quantize network outputs to obtain binary codes.



[1] H. Liu, R. Wang, S. Shan, X. Chen. Deep Supervised Hashing for Fast Image Retrieval. *IEEE CVPR 2016.*

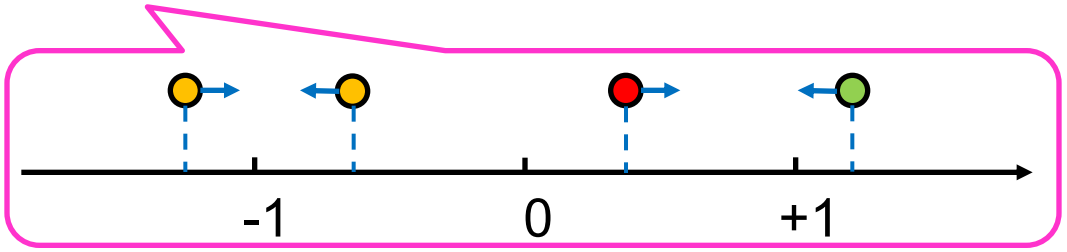
DSH

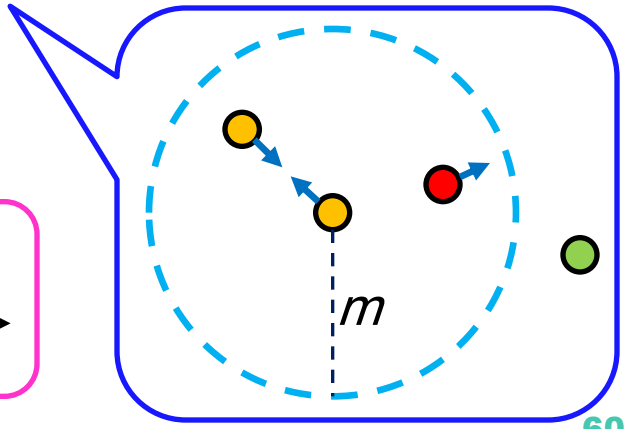
- Formulation of loss function
  - Similarity-preserving & minimizing quantization error

k	Code length	$\mathbf{b}_i$	Network outputs of the i-th image
y	$y=0$ if two images are similar, and 1 otherwise	$\|\cdot\|_{2(1)}$	L2 (L1) norm of vector
m	Margin parameter	$\|\cdot\|$	Element-wise absolute value operation

$$L_r\left(\mathbf{b}_1,\mathbf{b}_2,y\right)=\frac{1}{2}\left(1-y\right)\left\|\mathbf{b}_1-\mathbf{b}_2\right\|_2^2+\frac{1}{2}y\max\left(m-\left\|\mathbf{b}_1-\mathbf{b}_2\right\|_2^2,0\right)$$

$$+\alpha\left(\left\|\,|\mathbf{b}_1|-\mathbf{1}\right\|_1+\left\|\,|\mathbf{b}_2|-\mathbf{1}\right\|_1\right)$$



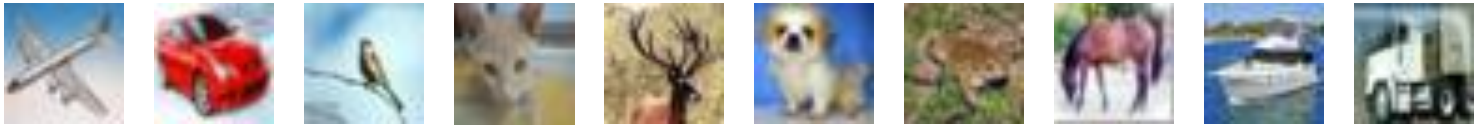


DSH

- Design principles
  - Loss function
    - Parameter *m*： avoiding collapsed solution
    - Regularizer: approximating Hamming space while avoiding the gradient vanish problem of *sigmoid or tanh* activations
- Implementation details
  - Construct training pairs on-the–fly for lower storage demand
  - Fine-tune the trained models for another code length to reduce computation cost and alleviate overfitting

DSH

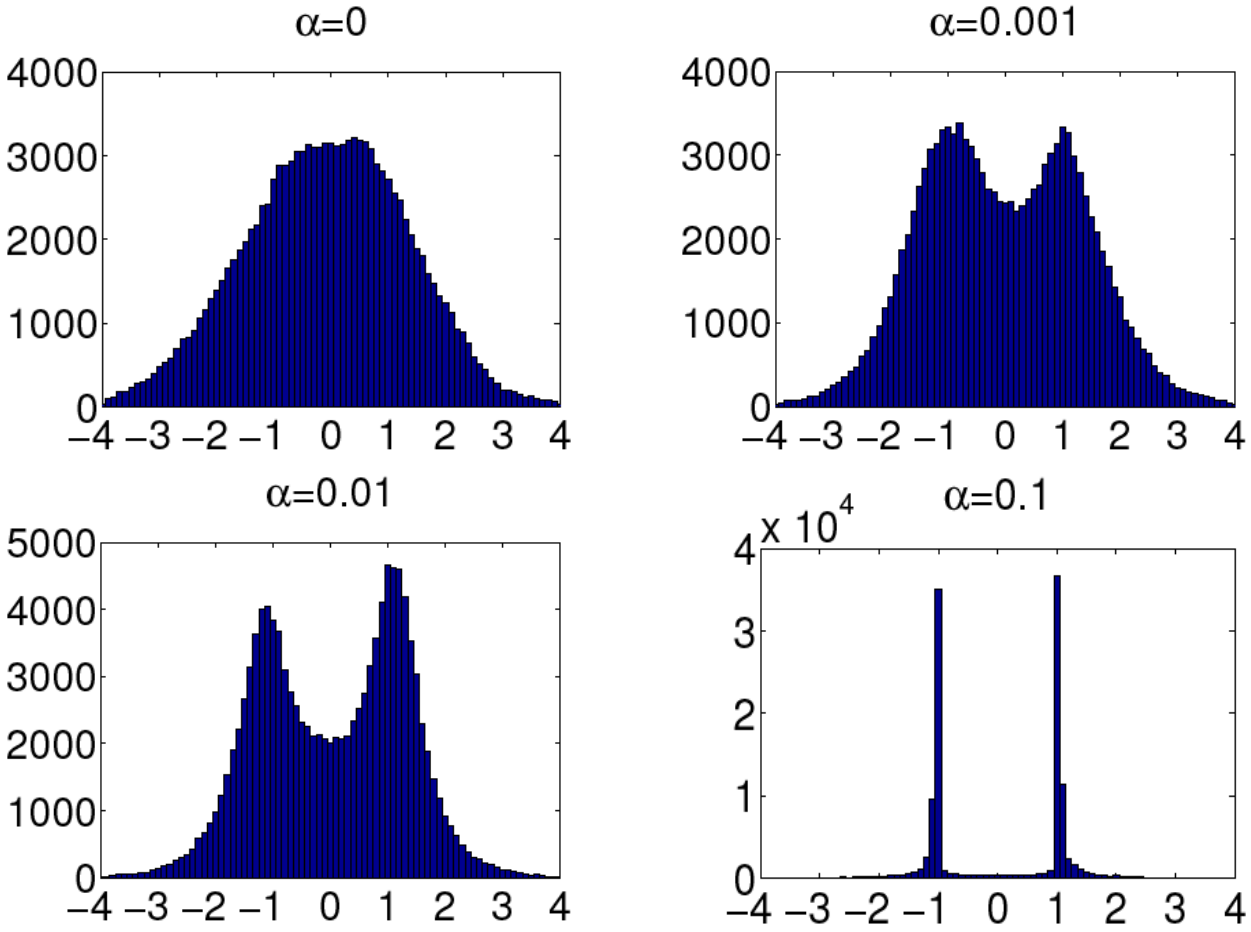
- Evaluation dataset
  - CIFAR-10
    - 10 classes, 6K images from each class



  - NUS-WIDE
    - About 180K images, 21 labels for each image

DSH

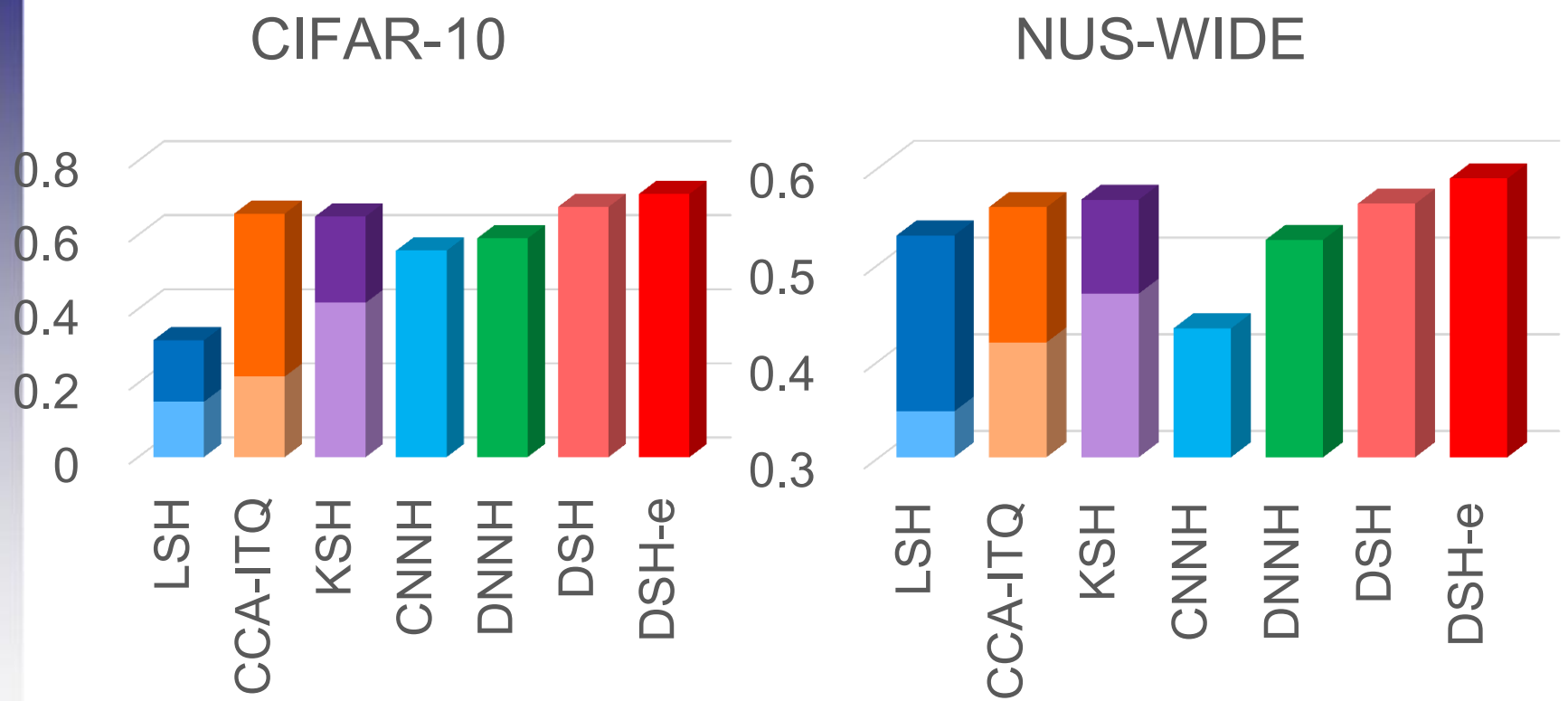
- Effectiveness of the proposed regularizer
  - Distribution of network outputs (CIFAR-10, 12-bit)



Param.	mAP
α=0	0.5497
α=0.001	0.6100
α=0.01	0.6157
α=0.1	0.4337

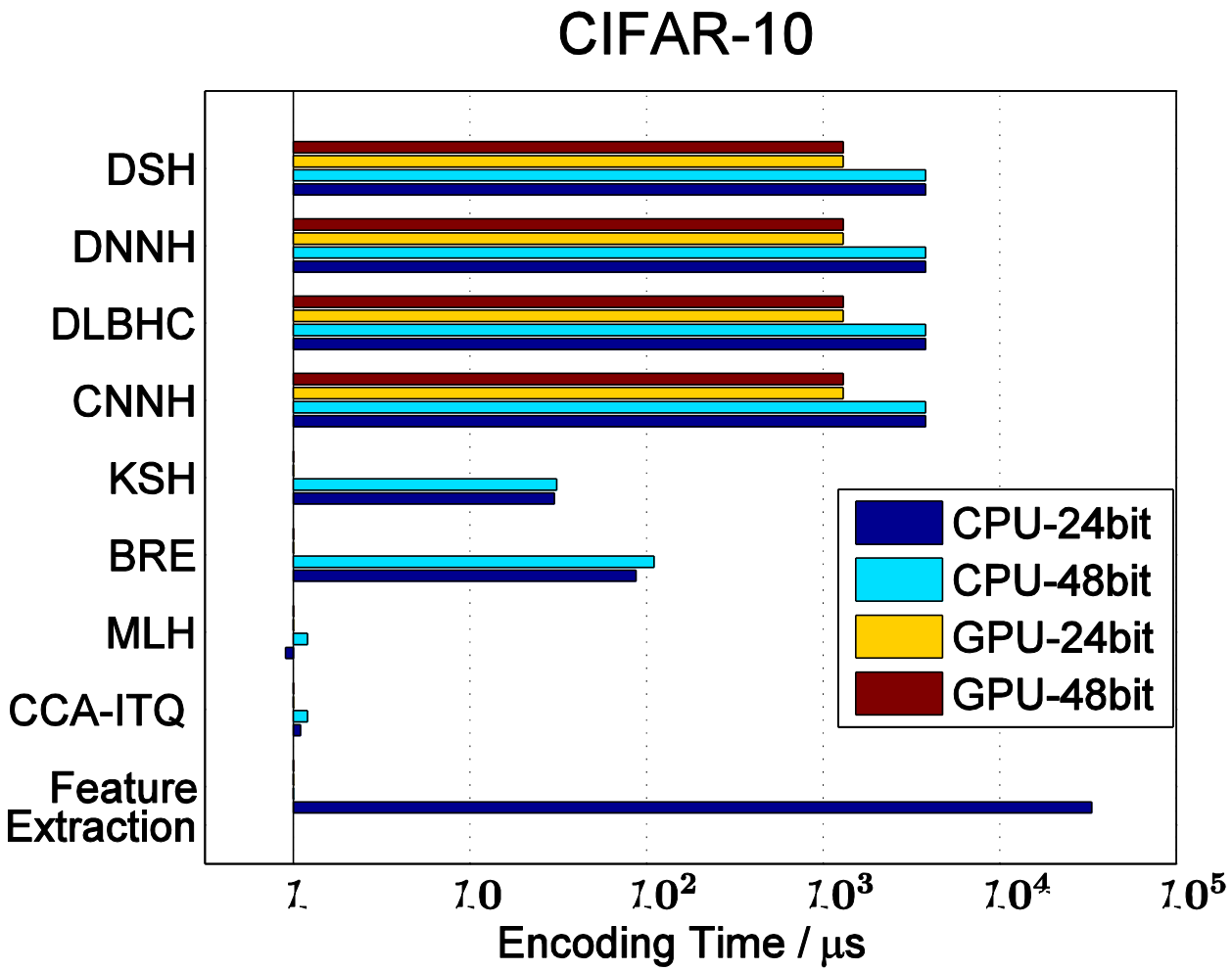
DSH

- Comparison with state of the arts



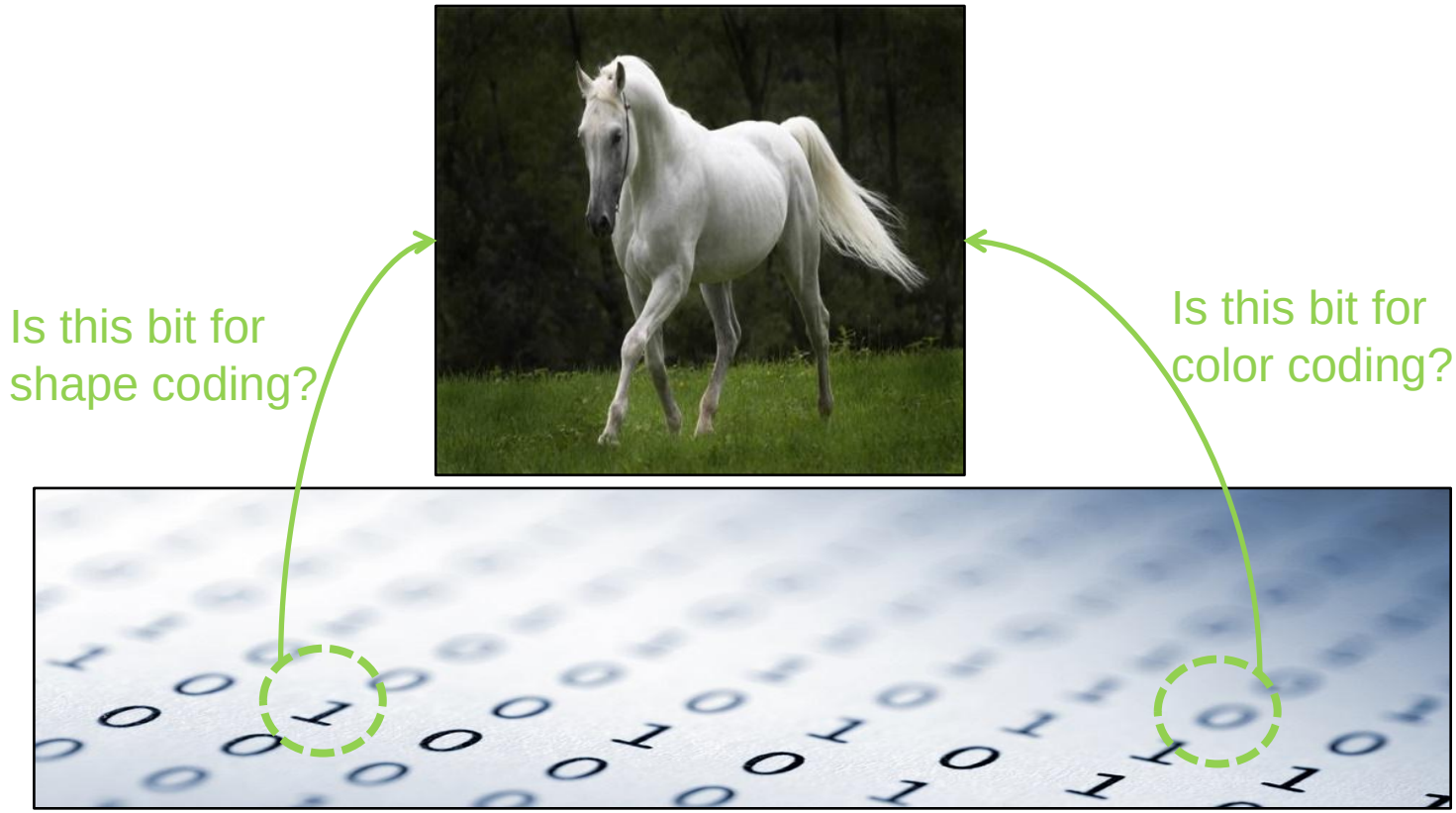
DSH

- Encoding time



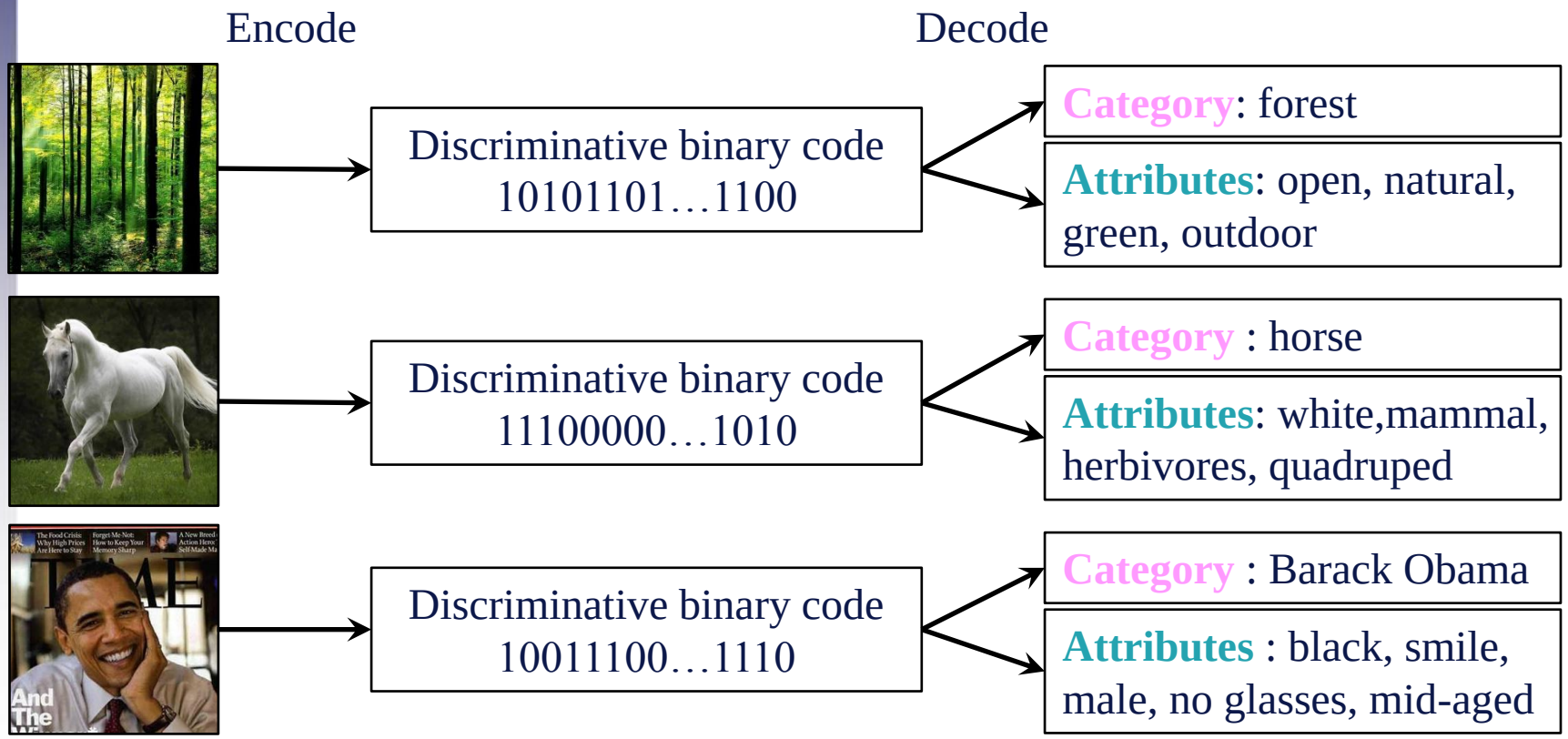
Binary code

- Limitation of discriminative binary code
  - Only contain category information
  - Human-incomprehensible, i.e., nobody knows what a zero-one string means



Binary code+?

Is it possible to encode other information into binary codes, e.g., visual attributes?

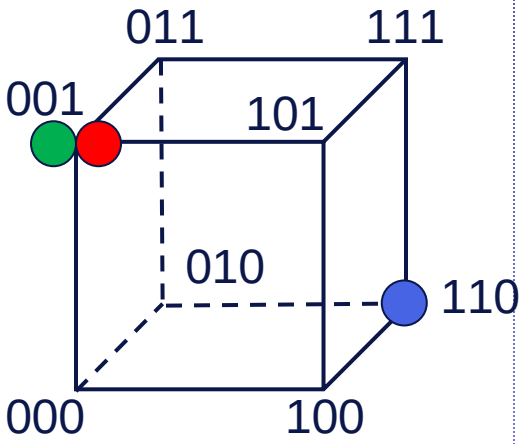


Towards Multi-functional Binary Codes

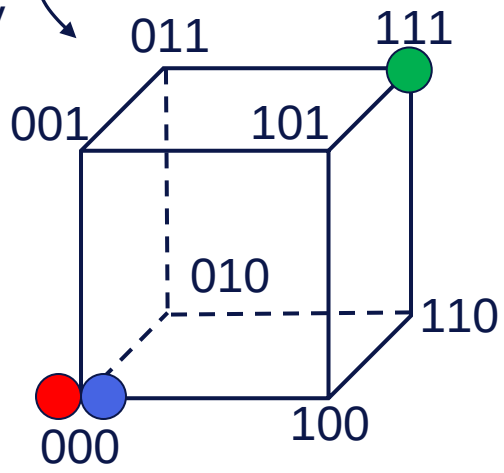
How to measure semantic similarity ?

Cat
Cat
Tiger

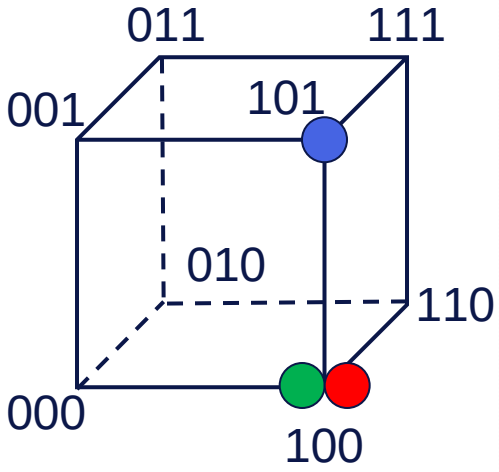
Category similarity



Appearance similarity

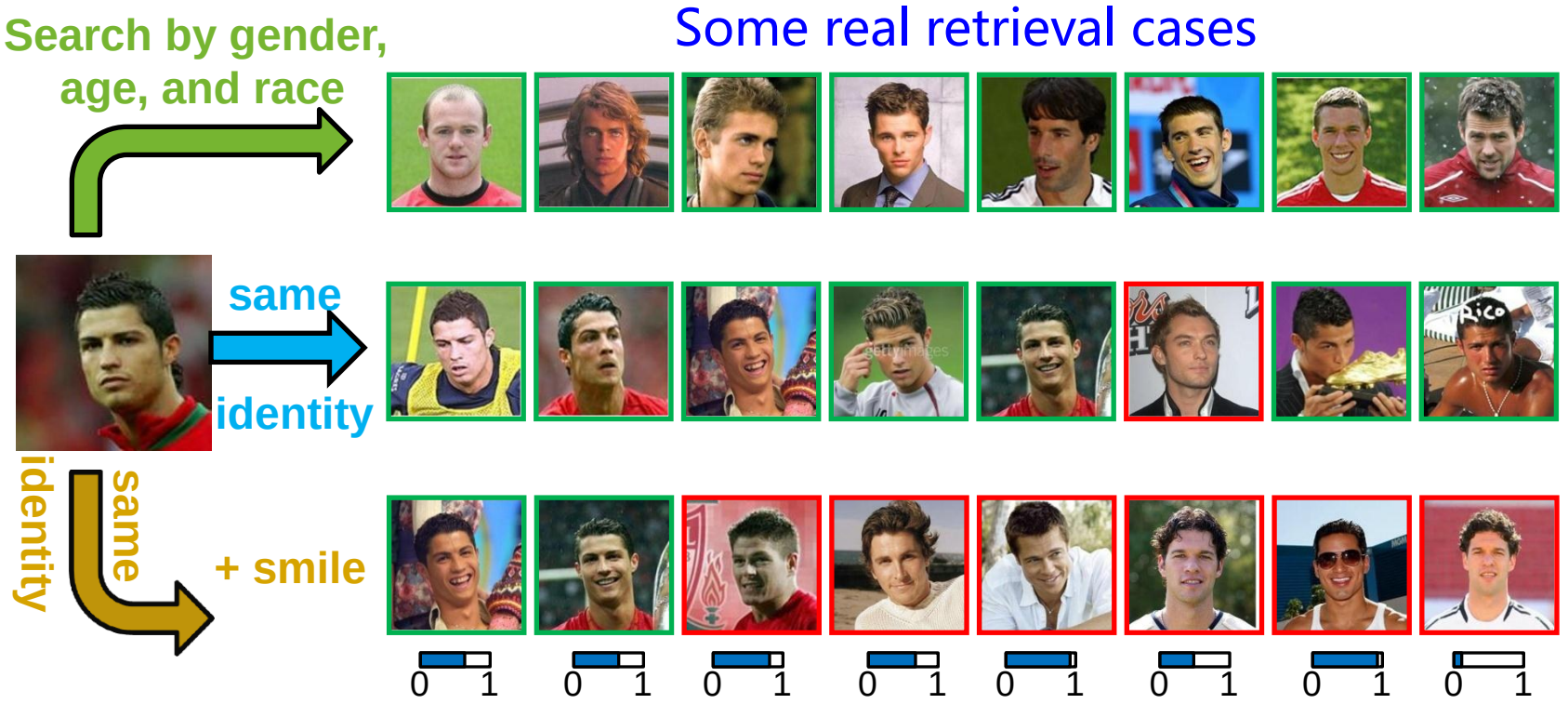


Both similarities



Dual Purpose Hashing (DPH)*

- Problem
  - Unified framework for multiple retrieval tasks



[1] H. Liu, R. Wang, S. Shan, X. Chen. Learning Multifunctional Binary Codes for Both Category and Attribute Oriented Retrieval Tasks. *IEEE CVPR 2017.*

DPH

- Motivation
  - Class labels: high-level semantic descriptor
  - Visual attributes: visual appearance descriptor
  - Commonly-seen image data:

	Black	Round	Stripe	Wooden	...	Tail	Class
	√	×	√	×	...	√	White Shark
	?	?	?	?	...	?	Balloon
	×	×	×	√	...	×	?
	?	×	√	×	...	√	?

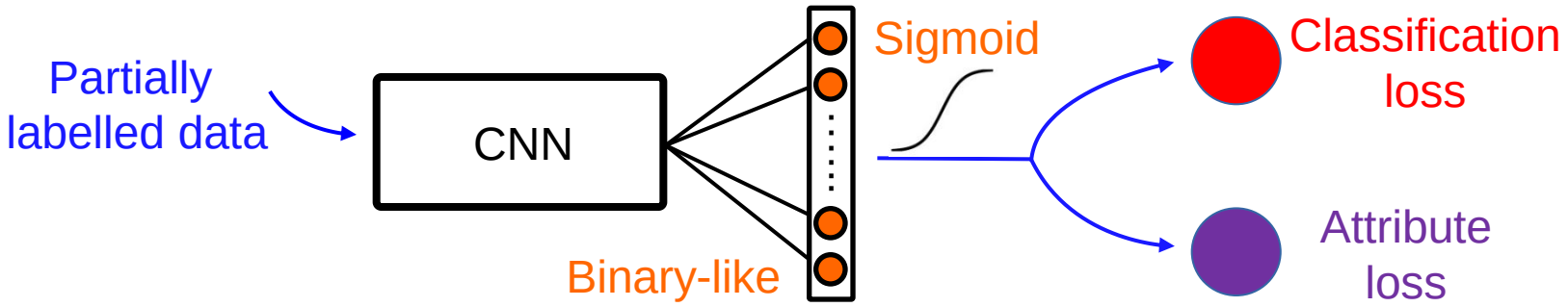
scarce

abundant (partially labelled)

√: Positive × : Negative ? : Unavailable

DPH

- Framework
  - Making use of partially labelled data



class label: y softmax output: $\mathbf{s}$

#classes: C

Cls. loss:

$$1\{\cdot\} = \begin{cases} 1, & if \ \cdot \ is \ true \\ 0, & otherwise \end{cases}$$

$$-\sum_{c} 1\{y=c\} \log \frac{\mathbf{s}_c}{\sum_{l=1}^{C} \mathbf{s}_l}$$

attr. label: $\mathbf{a}$ attr. predictions: $\mathbf{p}$

#attributes: m

Different weights for pos./neg. samples

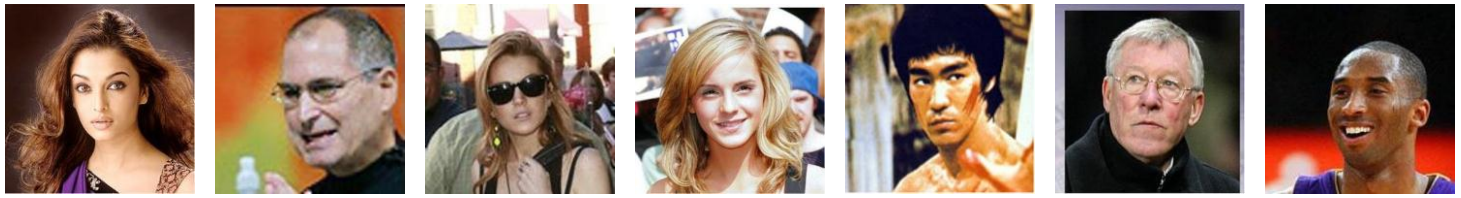
Attr. loss:

$$-\sum_{j=1}^{m} 1\{\mathbf{a}_j \neq ?\}[w_j \mathbf{a}_j \log(\mathbf{p}_j) + (1-w_j)(1-\mathbf{a}_j)\log(1-\mathbf{p}_j)]$$

DPH

- Evaluation
  - CFW-60K
    - 500 identities, 29~184 images each

Data partition	#images/person	has attribute label	has class label
Train-Base	2	√	√
Train-Attribute	8	√	
Train-Category	9~164		√
Test	10	√	√



DPH

- Evaluation
  - ImageNet-150K
    - 1000 classes, 150 images from each class

Data partition	#images/person	has attribute label	has class label
Train-Base	5	√	√
Train-Attribute	43	√	
Train-Category	100		√
Test	2	√	√

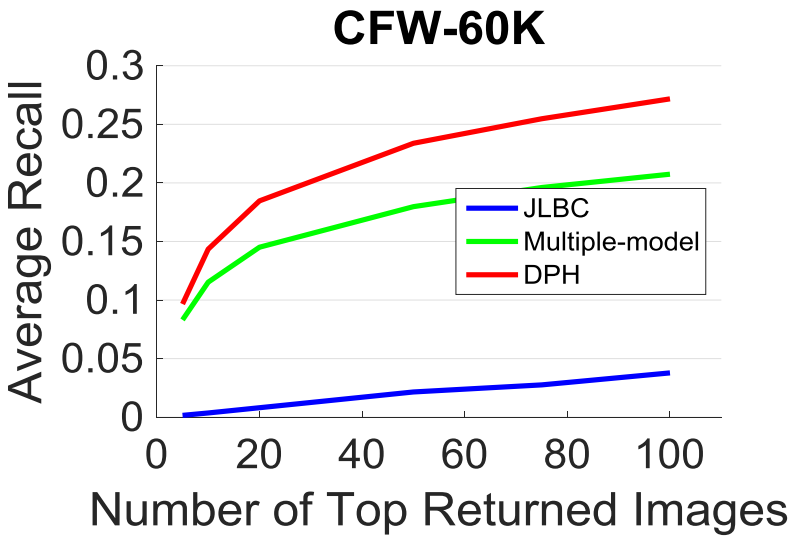


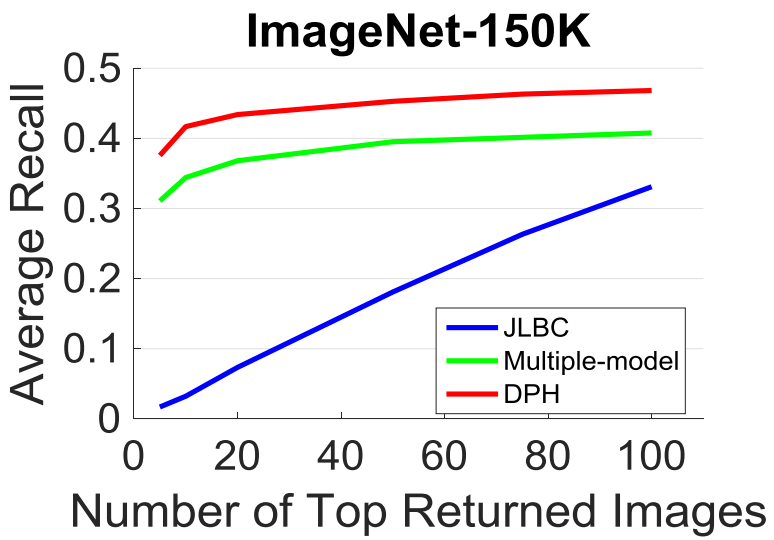





DPH

- Results
  - Retrieval by both class and attribute
    - Multiple-model: 2 models, one for attributes and the other for classes
    - JLBC [ICCV'15]*

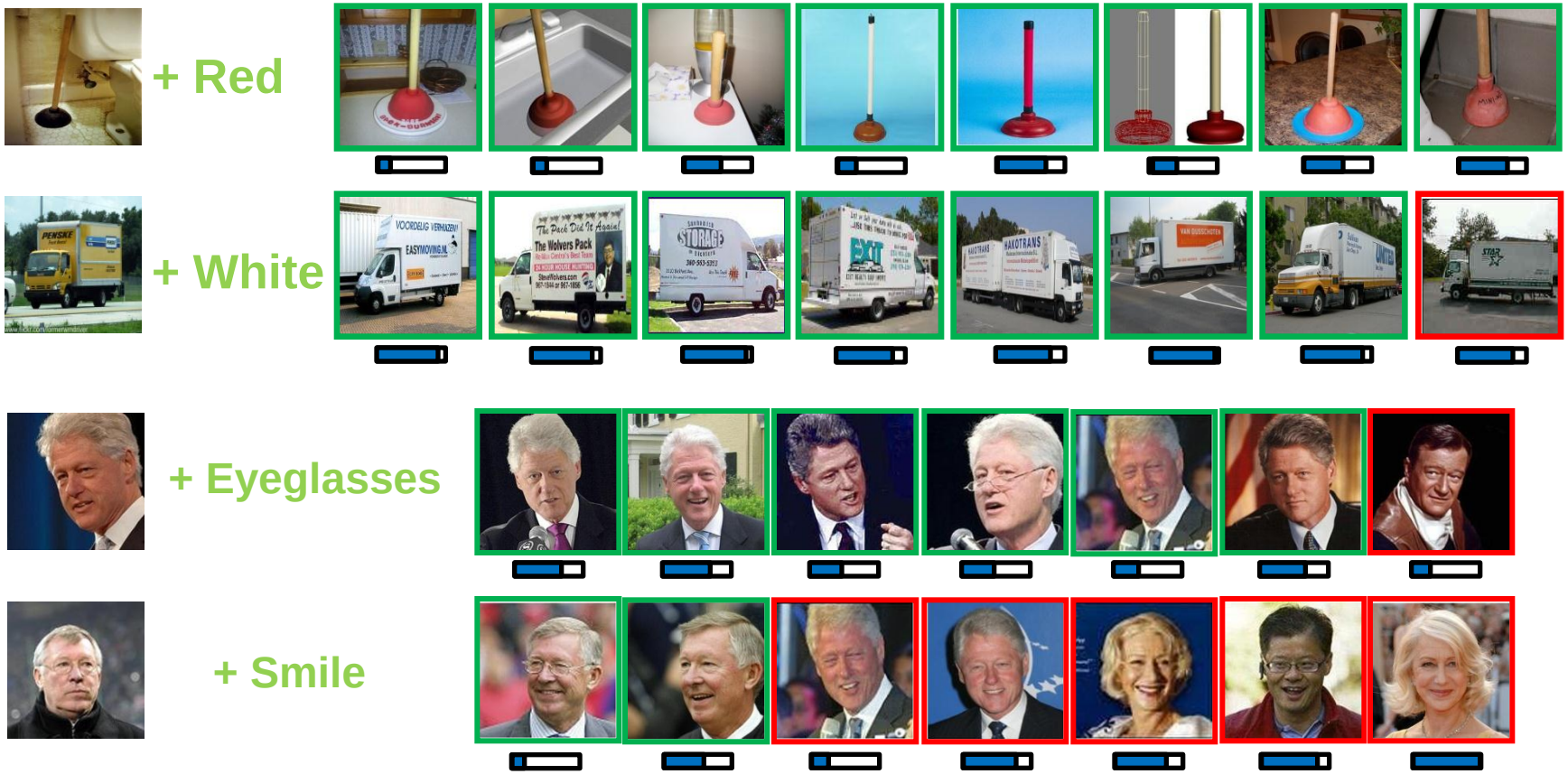




[1] Y. Li, R. Wang, H. Liu, H. Jiang, S. Shan, X. Chen. Two Birds, One Stone: Jointly Learning Binary Code for Large-scale Face Image Retrieval and Attributes Prediction. *IEEE ICCV 2015.*

DPH

- Real retrieval cases



DPH

- Demo video

Summary

- What we learn from current studies
  - Learning image representation together with hash functions significantly improves performance
  - Respecting the structure of Hamming space helps (regularizer)
  - Joint learning of attributes and classes is possible
- Future directions
  - More suitable losses, e.g. triplet loss
  - Specifically designed network structure

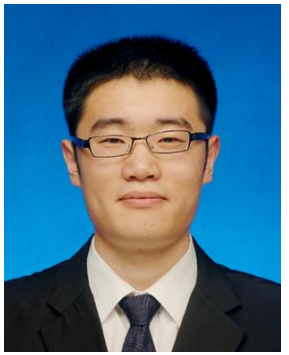
Thanks, Q & A

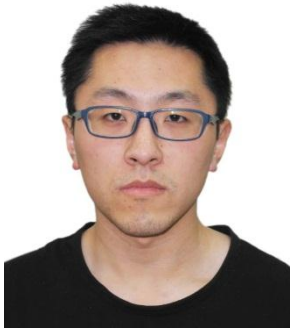

Zhiwu Huang


Yan Li


Wen Wang


Mengyi Liu


Shishi Qiao


Haomiao 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*