Learning Euclidean-to-Riemannian Metric for Point-to-Set Classification

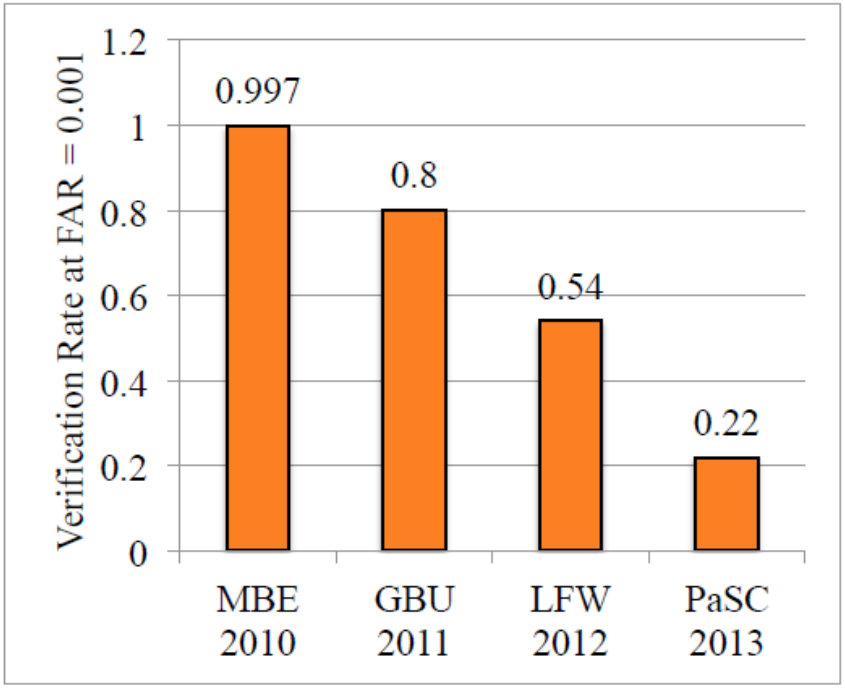
Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xilin Chen
Institute of Computing Technology, Chinese Academy of Sciences

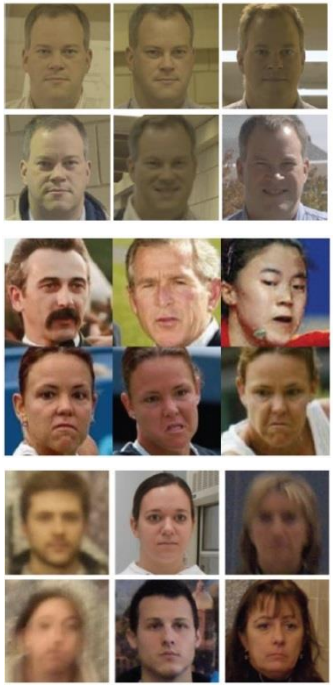
VALSE QQ Webinar, 2014.12.11

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

Problem(1/3)

- Face Recognition
  - From image-based to video-based setting





GBU (image)

LFW (image)

PaSC (video)

Figure: Performance progressively drops when shifting from controlled scenarios (**Image-based**) to uncontrolled point-and-shoot conditions (**video-based**). **[Beveridge, BTAS'13]**

Problem(2/3)

Institute of Computing Technology

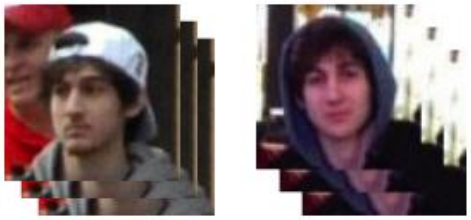
- Video-based FR
  - Video surveillance

Video frame of pair
suspected in Boston attack

Query Target

Video-to-Still (V2S)

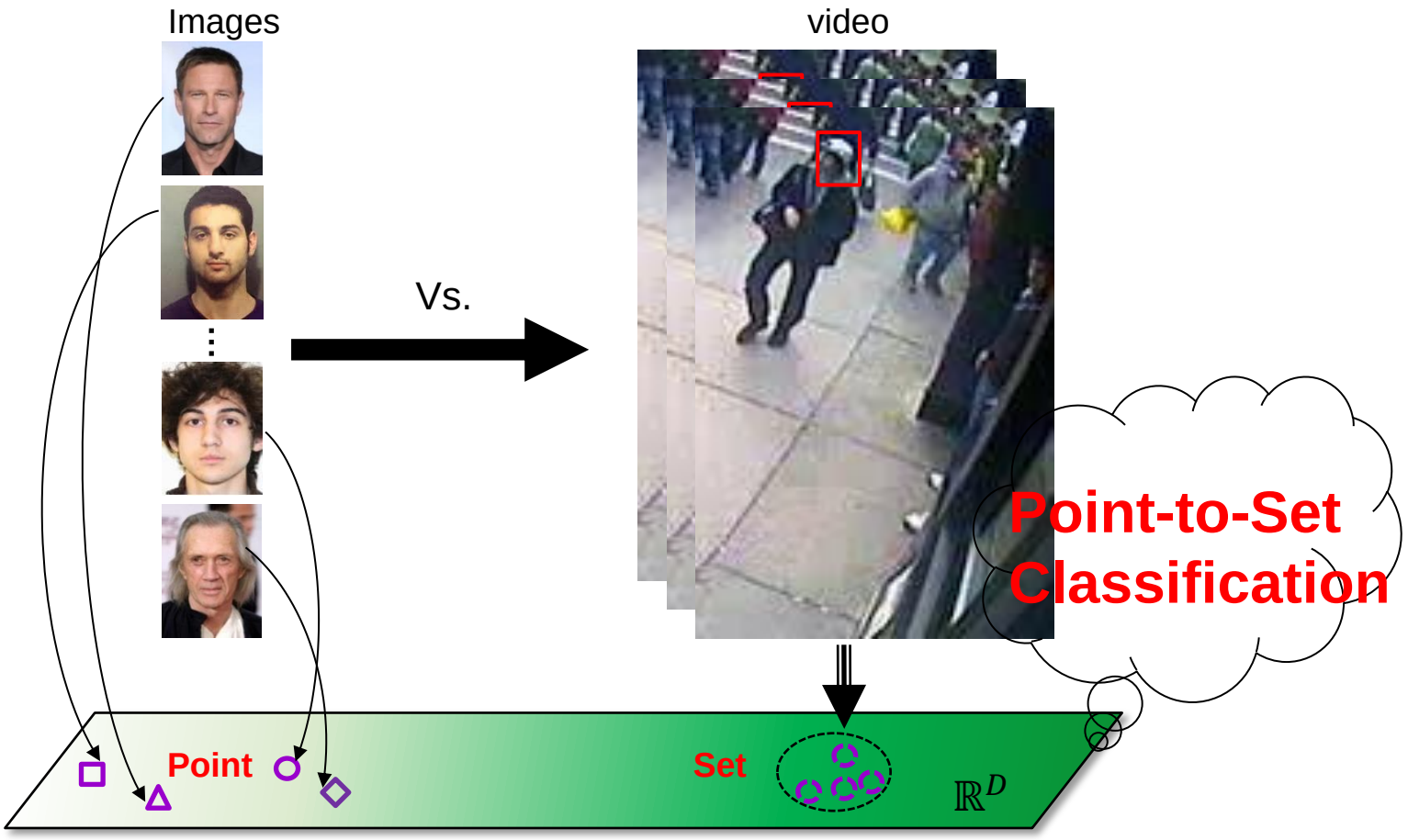
Still-to-Video (S2V)



Video-to-Video (V2V)

Problem(3/3)

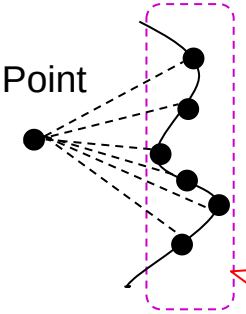
- V2S/S2V FR



Motivation(1/2)

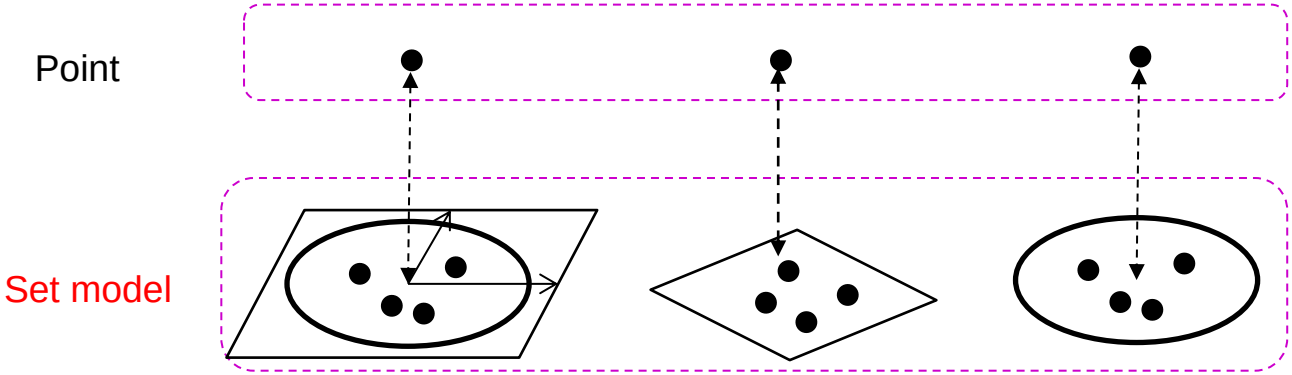
- Point-to-Set Classification
  - Match points against sets
    - Compare point with all points in set
      - 1-N Point to Point (P2P) matching

Hausdorff dist.



Time / space consuming

    - Compare point with set model
      - 1-1 Point to Set (P2S) matching



More efficient

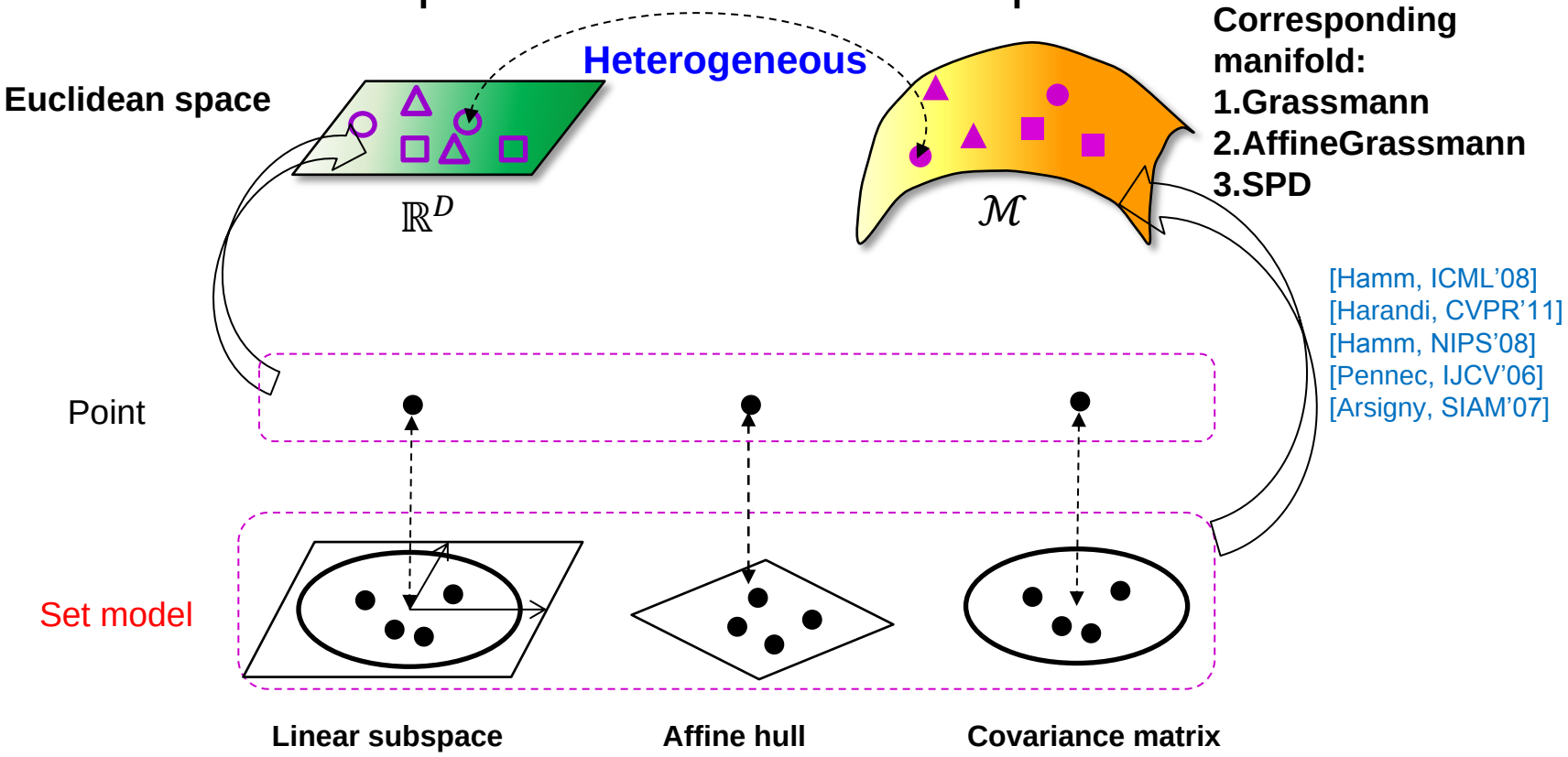
Linear subspace
[Yamaguchi, FG'98]
[Chien, PAMI'02]
[Kim, PAMI'07]

Affine hull
[Vincent, NIPS'01]
[Cevikalp, CVPR'10]
[Zhu, ICCV'13]

Covariance matrix
[Wang, CVPR'12]
[Vemulapalli, CVPR'13]
[Lu, ICCV'13]

Motivation(2/2)

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



Related Work(1/3)

- Grassmann manifold $\mathcal{G}(D', D)$
  - The space of linear subspaces
    - Each is represented by an orthonormal matrix $U \in \mathbb{R}^{D \times D'}$
  - Riemannian metric (projection metric) [Hamm et al., ICML'08]
    - $d^2(U_i, U_j) = 2^{-1/2} \left\| U_i U_i^T - U_j U_j^T \right\|_F$
- Affine Grassmann manifold $\mathcal{AG}(D', D)$
  - The space of affine subspaces
    - Each is represented by $U \in \mathbb{R}^{D \times D'}$ with mean value u
  - Riemannian metric (affine projection metric) [Hamm et al., NIPS'08]
    - $d^2(U_i, U_j) = 2^{-1/2} (\left\| U_i U_i^T - U_j U_j^T \right\|_F + \left\| (I - U_i U_i^T) u_i - (I - U_j U_j^T) u_j \right\|_F)$
- Symmetric Positive Define (SPD) manifold $\mathbb{S}_{++}^D$
  - The space of SPD matrices $C \in \mathbb{R}^{D \times D}$
  - Riemannian metric (Log-Euclidean metric) [Arsigny et al., SIAM'07]
    - $d^2(C_i, C_j) = \left\| \log(C_i) - \log(C_j) \right\|_F$

Related Work(2/3)

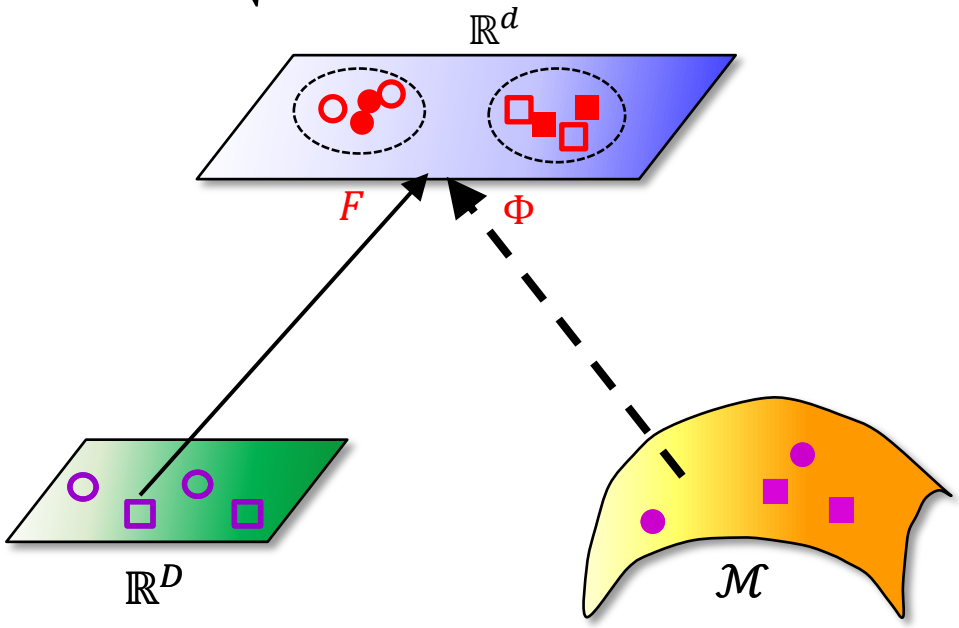
- Euclidean-to-Riemannian distance
  - Euclidean-to-Grassmannian
    - Point to Linear Subspace [Chien et al., PAMI'02]
      - $d(x_i, \mathrm{U}_j) = \|x_i - U_j U_j^T x_i\|_F$
  - Euclidean-to-AffGrassmannian
    - Point to Affine Subspace [Vincent et al., NIPS'01]
      - $d(x_i, \mathrm{A}_j) = \min_{\alpha} \|(U_j\alpha + u_j) - x_i\|_F$, α is a vector of free parameters that provides coordinates for points within the subspace
  - Euclidean-to-SPD
    - Point to Covariance Matrix (Mahalanobis distance)
      - $d^2(x_i, C_j) = \sqrt{(x_i - u_j)^T C_j^{-1} (x_i - u_j)}$

Related Work(3/3)

- Point-to-set distance metric learning [zhu et al. , ICCV'13]
  - Basic distance
    - $d^2(\boldsymbol{x}, \boldsymbol{D}) = \min\|\boldsymbol{x} - H(\boldsymbol{Y})\|_2^2$
      - Affine subspace $H(\boldsymbol{Y}) = \boldsymbol{Y\alpha}$ spanned by all the available samples $\boldsymbol{Y} = \{y_1, \ldots, y_n\}$ in the set, s.t. $\sum \alpha_i = 1$
      - Solution: Least Square Regression or Ridge Regression
  - Mahalanobis distance
    - $d_{\boldsymbol{M}}^2(\boldsymbol{x}, \boldsymbol{D}) = \min\|\boldsymbol{W}(\boldsymbol{x} - \boldsymbol{Y\alpha})\|_2^2 = (\boldsymbol{x} - \boldsymbol{Y\alpha})^T\boldsymbol{W}^T\boldsymbol{W}(x - \boldsymbol{Y\alpha}) = (\boldsymbol{x} - \boldsymbol{Y\alpha})^T\boldsymbol{M}(\boldsymbol{x} - \boldsymbol{Y\alpha})$

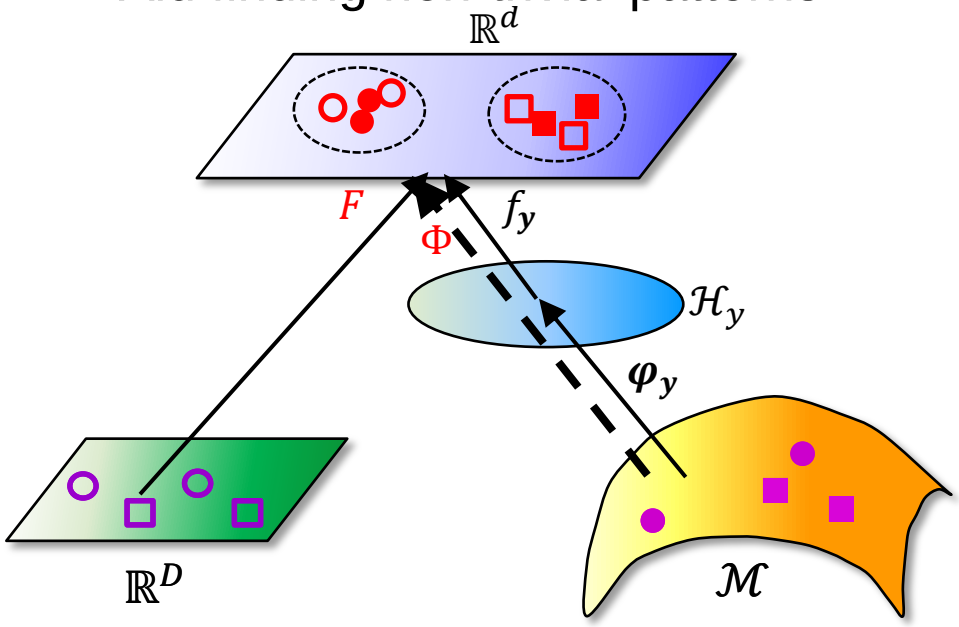
Our Method(1/10)

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



Our Method(2/10)

- Basic idea
  - Bridge Euclidean-to-Riemannian gap
    - Hilbert space embedding
      - Adhere to Euclidean geometry
      - Yield much richer data representation
        - Aid finding non-trivial patterns



Our Method(3/10)

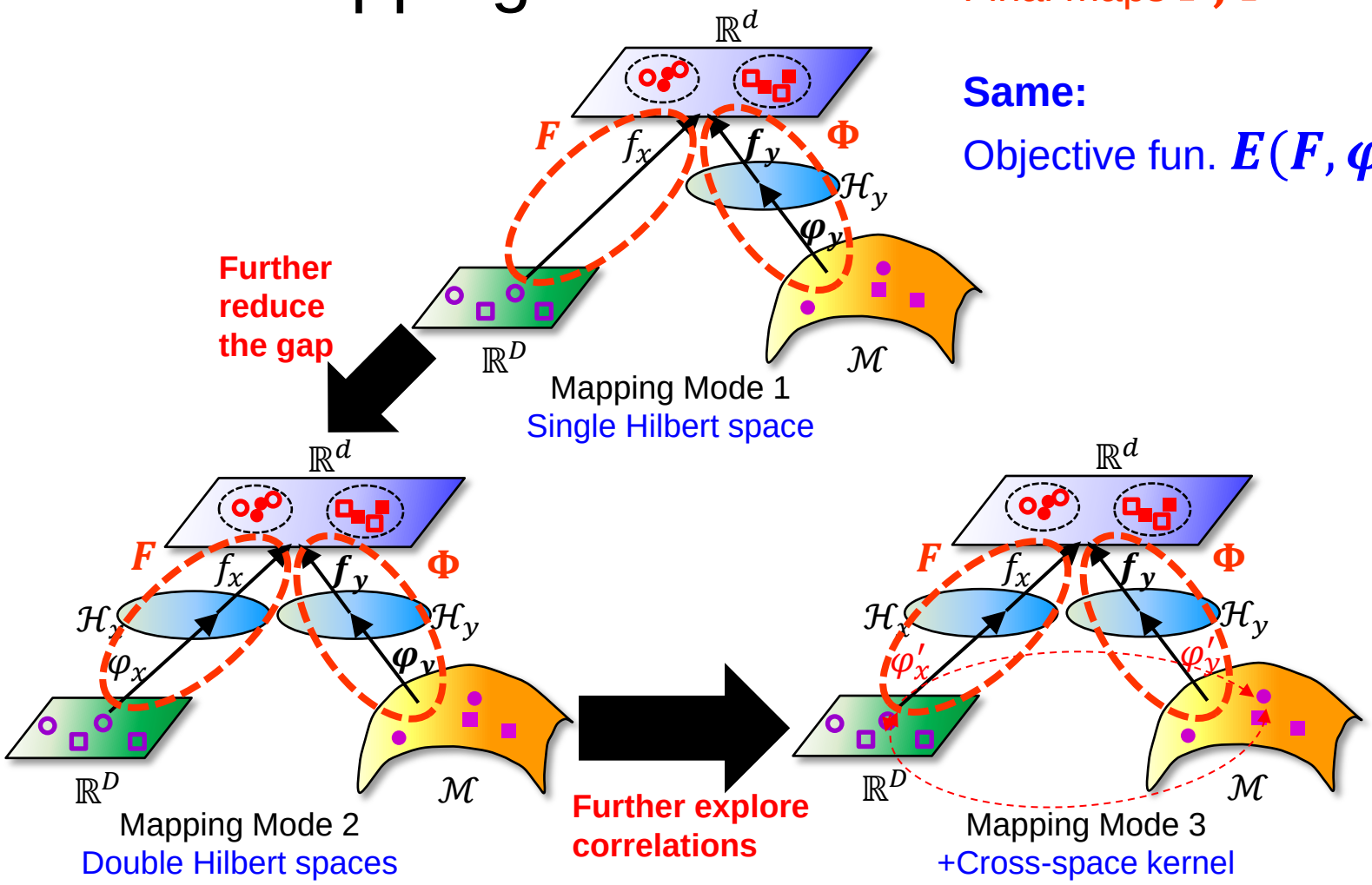
- Three Mapping Modes

Different:

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

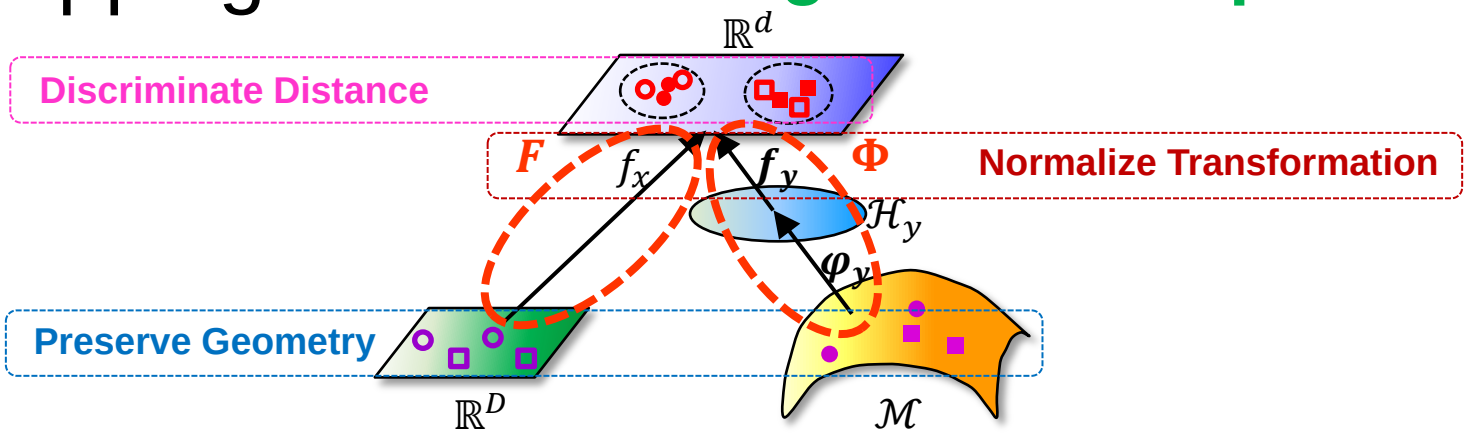
Same:

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



Our Method(4/10)

- Mapping 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:

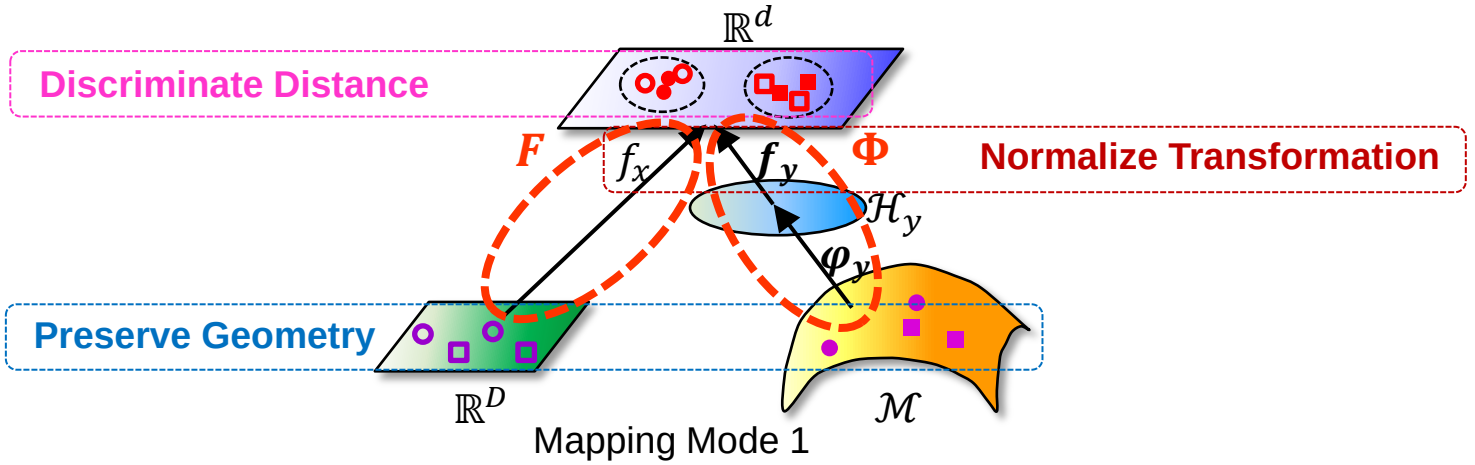
$$d(x_i, y_j) = \sqrt{(F(x_i) - \Phi(y_j))^T (F(x_i) - \Phi(y_j))}$$

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

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

Distance Geometry Transformation

Our Method(5/10)



Mapping Mode 1

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

Objective fun.: $E(F, \Phi) = \min_{F,\Phi} \{ D(F, \Phi) + \lambda_1 G(F, \Phi) + \lambda_2 T(F, \Phi) \}$

Distance | Geometry | Transformation

$$D = \frac{1}{2} \sum \sum sgn(l_i^x, l_j^y) d^2(x_i, y_j)$$

$$G_x = \frac{1}{2} \sum \sum e^{\frac{\|x_i - x_j\|^2}{\sigma_x^2}} d_x^2(x_i, x_j)$$

$$T = \frac{1}{2} (\|F(X)\|_F^2 + \|\Phi(Y)\|_F^2)$$

Our Method(6/10)

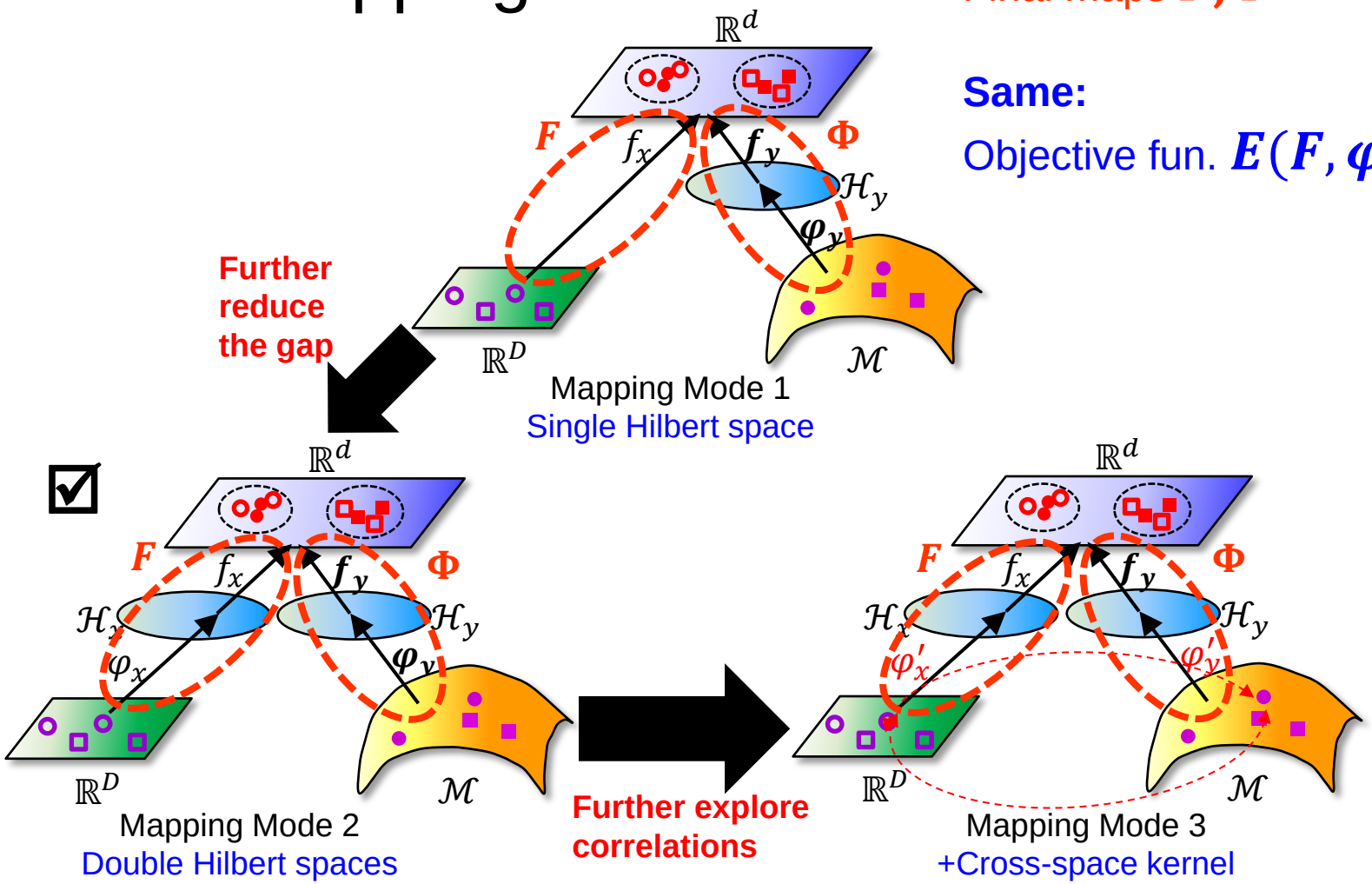
- Three Mapping Modes

Different:

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

Same:

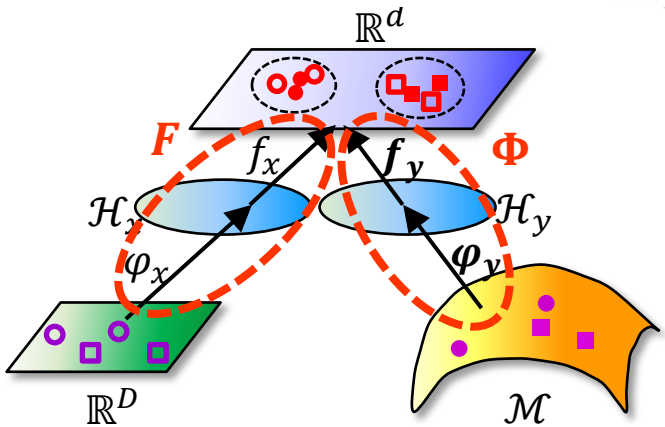
Objective fun. $\boldsymbol{E(F, \varphi)}$



Our Method(7/10)

- Mapping Mode 2

Further reduce the E-R gap with double Hilbert spaces



Mapping Mode 2

Final maps:

$$F = f_x \circ \varphi_x = W_x^T K_x$$
$$\Phi = f_y \circ \varphi_y = W_y^T K_y$$

$$K_x(i,j) = \exp(-\|x_i - x_j\|_2^2 / 2\sigma^2)$$

Classical Euclidean distance

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

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

Our Method(8/10)

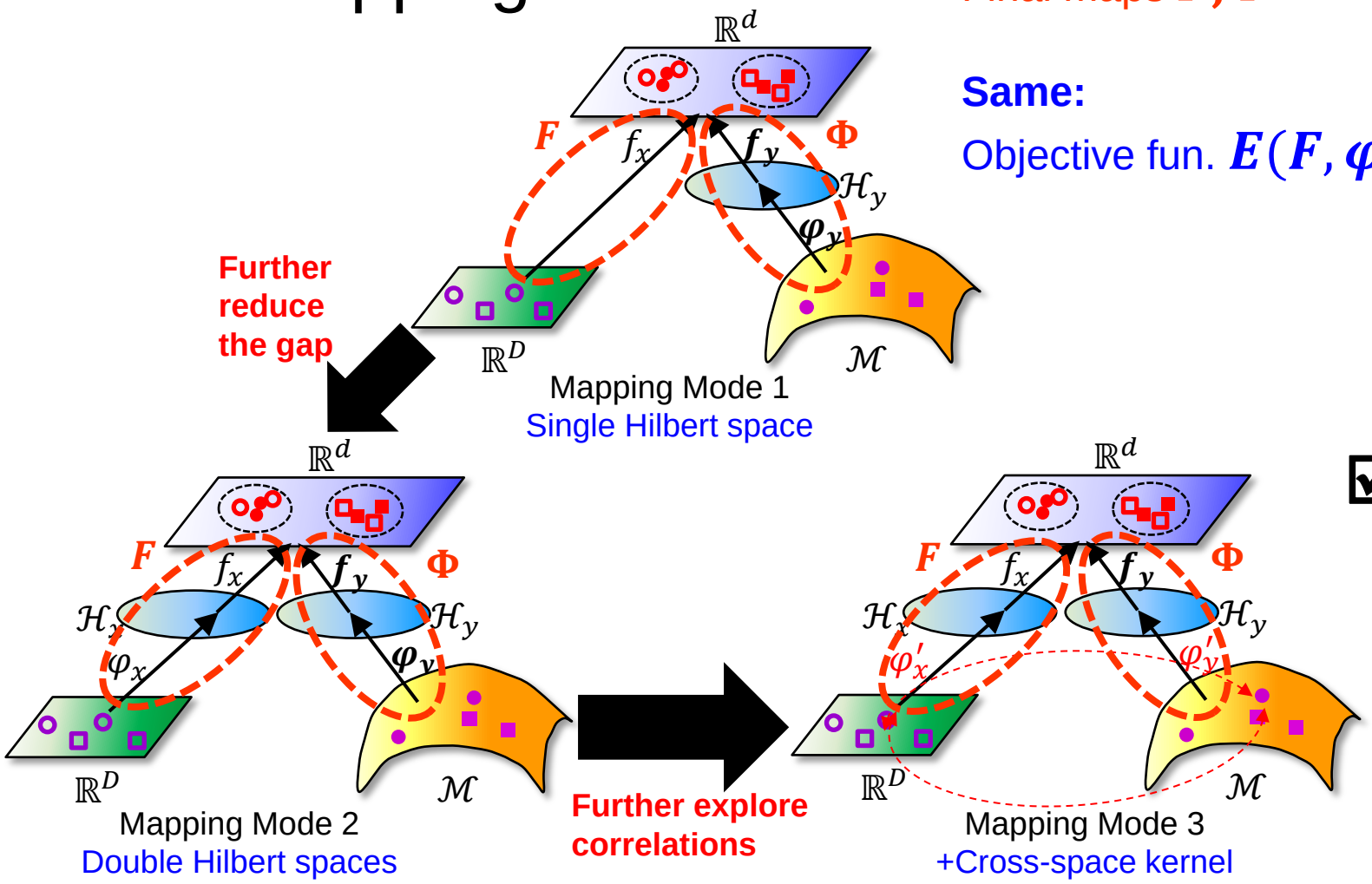
- Three Mapping Modes

Different:

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

Same:

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



☑

Our Method(9/10)

- Mapping Mode 3

Final Maps:

$$F = f_x \circ \varphi'_x = W_x^T[K_x, K_{xy}]$$

$$\Phi = f_y \circ \varphi'_y = W_y^T[K_y, K_{xy}^T]$$

$$K_x(i,j) = \exp(-\|x_i - x_j\|_2^2 / 2\sigma^2)$$

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

$$K_{xy}(i,j) = \exp(-d^2(x_i, y_j)/2\sigma^2)$$

Further correlate the double Hilbert spaces with cross-space kernel

Point-to-Set (E-R) distance
[PAMI'02, NIPS'01, Mahalanobis]

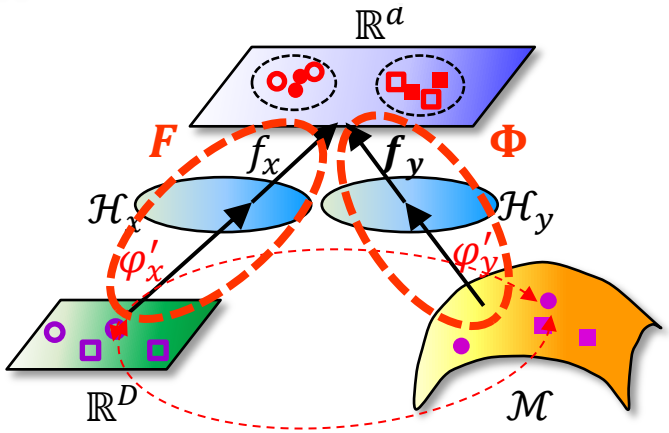
Distance metric:

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

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

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

Distance Geometry Transformation



Mapping Mode 3

Our Method(10/10)

- Optimization of the objective functions
  - Iterative optimization
    - CCA-like Initialization
      - $\max_{W_x, W_y} \{D^b(W_x, W_y) + \lambda_1 G^b(W_x, W_y)\}$
      - $s.t. D^w(W_x, W_y) + \lambda_1 G^w(W_x, W_y) = 1$
    - Alternately updating (gradient descent)
      - Fix W_y to update W_x
        - W_x =?
      - Fix W_x to update W_y
        - W_y =?

D^b, G^b: the between-class template of D, G

D^w, G^w: the within-class template of D, G

$$\boldsymbol{F = f_x \circ \varphi_x = W_x^T K_x}$$
$$\boldsymbol{\Phi = f_y \circ \varphi_y = W_y^T K_y}$$

Institute of Computing Technology

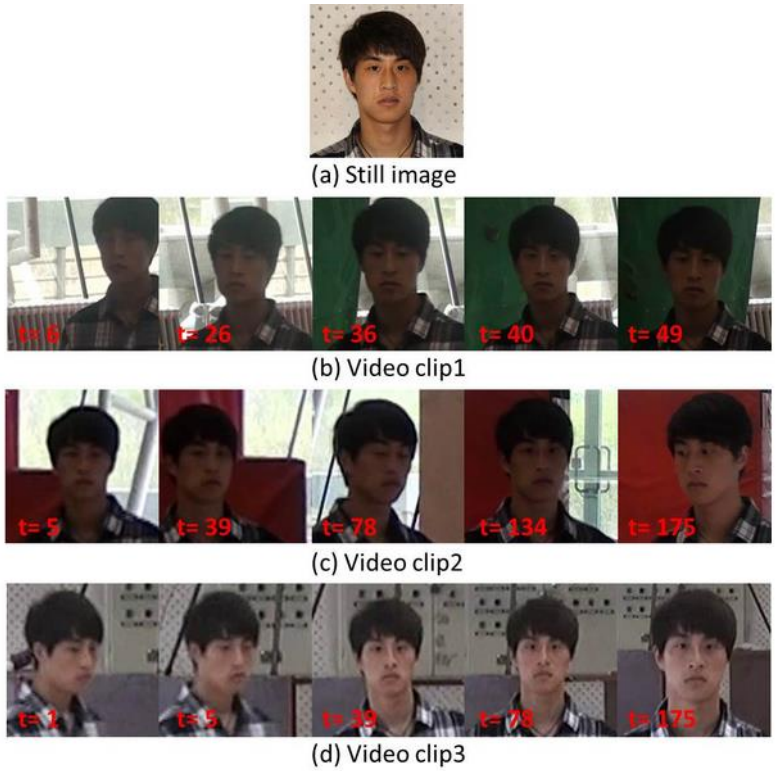
Experiments(1/7)

- Video face datasets
  - **YouTube Celebrities** [Kim, CVPR'08]
    - 47 subjects
      - 1,910 videos from YouTube, images selected from videos

Images Videos

Experiments(2/7)

- COX video face database
  - http://vipl.ict.ac.cn/resources/datasets/cox-face-dataset
- Features of COX
  - 1000 subjects, each
    - 1 high quality still image
    - 3 low quality video clips from 3 camcorders
  - (Intended to) simulate video surveillance
  - Evaluation protocols



Experiments(3/7)

- Comparative Methods
  - Point-to-Point (**P2P**) metric learning methods
    - NCA [Goldberger, NIPS'04]
    - ITML [Davis, ICML'07]
    - LMNN [Weinberger, JMLR'09]
  - Point-to-Set (**P2S**) matching methods
    - NFS [Chien, PAMI'02]
    - HKNN [Vincent, NIPS'01]
    - PSDML [Zhu, ICCV'13]
  - Kernelized Multiview Learning (**KML**) methods
    - KPLS [Sharma, CVPR'11]
    - KCCA [Hardoon, Neural Comp.'04]
    - KGMLDA [Sharma,CVPR'12]

Experiments(4/7)

- Still-to-Video face recognition
  - YouTube dataset

Methods		Probe-Gallery; Average rank-1 recognition rate	
		Video-to-Still	Still-to-Video
P2P	NCA [NIPS'04]	51.74±3.11	60.35±3.09
	ITML [ICML'07]	47.62±1.73	59.72±4.27
	LMNN [JMLR'09]	55.02±2.71	70.99±3.25
P2S	NFS [PAMI'02]	53.27±2.75	60.21±5.99
	HKNN [NIPS'01]	36.94±2.71	48.01±4.80
	PSDML [ICCV'13]	55.30 ±1.90	61.21±5.24
KML	KPLS [CVPR'11]	54.02±2.61	64.89±5.18
	KCCA [Neural Comp.'04]	55.80±3.12	67.80±3.71
	KGMLDA [CVPR'12]	**58.19±4.00**	**89.36±6.88**
Ours	**LERM**	**69.11±2.99**	**96.60±3.36**

Experiments(5/7)

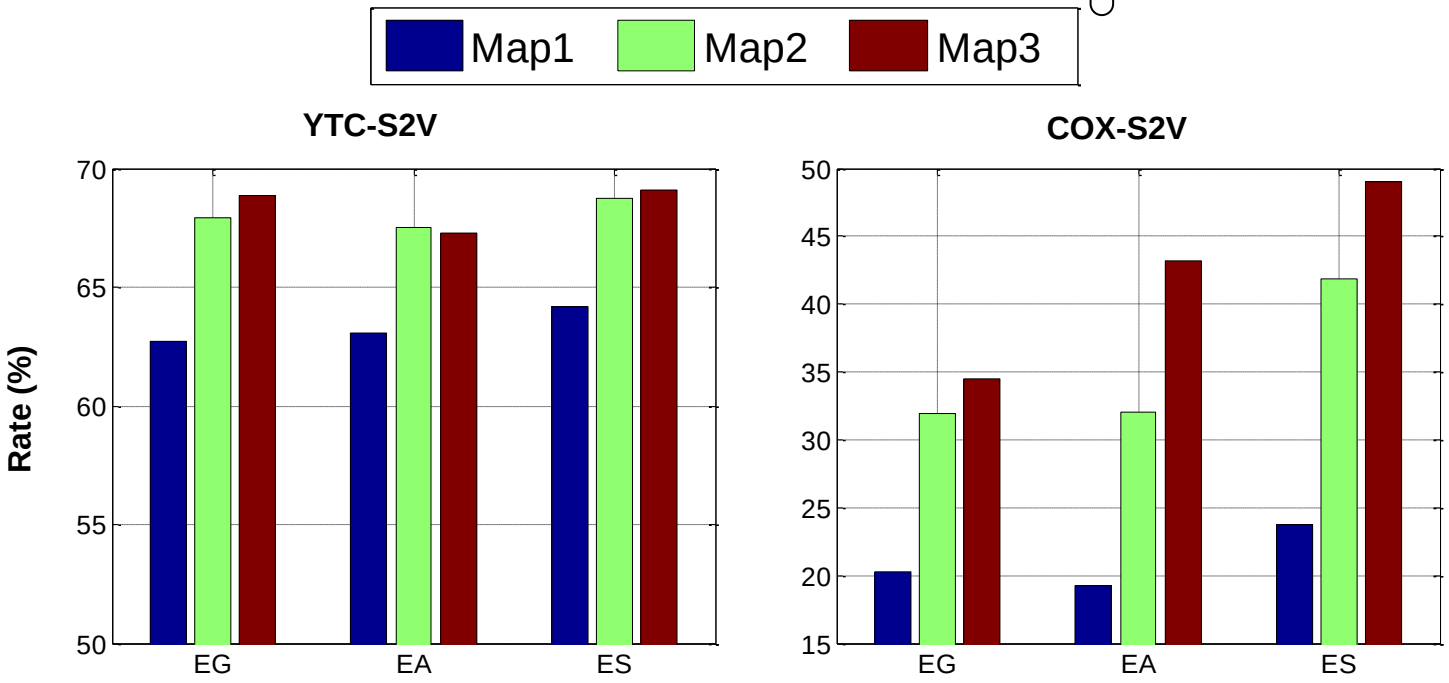
- Still-to-Video face recognition
  - COX Face dataset

Methods		Probe-Gallery; Average rank-1 recognition rate					
		V1-S	V2-S	V3-S	S-V1	S-V2	S-V3
P2P	NCA [NIPS'04]	39.14	31.57	57.57	37.71	32.14	58.86
	ITML [ICML'07]	19.83	18.20	36.63	26.66	25.21	47.57
	LMNN [JMLR'09]	34.44	30.03	**58.06**	37.84	**35.77**	**63.33**
P2S	NFS [PAMI'02]	9.99	5.90	22.23	11.64	6.51	31.67
	HKNN [NIPS'01]	4.70	3.70	12.70	6.34	4.64	20.41
	PSDML [ICCV'13]	12.14	9.43	25.43	7.04	4.14	29.86
KML	KPLS [CVPR'11]	20.21	16.21	27.23	14.83	11.61	23.99
	KCCA [Neural Comp.'04]	38.60	33.20	53.26	36.39	30.87	50.96
	KGMLDA [CVPR'12]	**41.89**	**38.29**	52.87	**38.03**	33.29	50.06
Ours	**LERM**	**45.71**	**42.80**	**58.37**	**49.07**	**44.16**	**63.83**

Experiments(6/7)

- Different mapping modes

Map3 > Map2 > Map1



Set models:

- EG: Euclidean-to-Grassmannian (linear subspace)
- EA: Euclidean-to-AffineGrassmannian (affine subspace)
- ES: Euclidean-to-SPD (covariance matrix)

Experiments(7/7)

- Running time (seconds) on YouTube

Methods		Training	Test
P2P	NCA [NIPS'04]	7761	0.165
	ITML [ICML'07]	523.2	0.394
	LMNN [JMLR'09]	282.4	0.162
P2S	PSDML [ICCV'13]	**55.78**	**0.016**
	LERM	**3.615**	**0.032**

Note: The running time in test is average per testing video

Conclusion

- V2S/S2V FR →Point-to-set classification

⇨Euclidean-to-Riemannian matching

- Conventional metric learning

⇨Euclidean-to-Riemannian Metric Learning

- Impressively improvement and much higher efficiency

Future work

- This framework may be also applied in many other CV tasks, e.g., image-to-set object categorization, image-to-video face retrieval
- This framework can be extended for metrics between sets with application to V2V FR

Take home message

- Learning Euclidean-to-Riemannian Metric (LERM) for point-to-set classification

Get a copy of COX face dataset for video-based (V2S/S2V/V2V) face recognition!

Data and code are available online!

http://vipl.ict.ac.cn/resources

Thank you!

Q&A

References

- O. Yamaguchi, K. Fukui, K. Maeda. Face recognition using temporal image sequence. In FG, 1998.
- J. Chien and C. Wu. Discriminant waveletfaces and nearest feature classifiers for face recognition. IEEE T-PAMI, 24(12):1644–1649, 2002.
- T. Kim, J. Kittler, R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. IEEE T-PAMI 29(6), 1005-1018, 2007.
- P. Vincent and Y. Bengio. K-local hyperplane and convex distance nearest neighbor algorithms. In NIPS, 2001.
- H. Cevikalp, B. Triggs. Face recognition based on image sets. In CVPR, 2010.
- P. Zhu, L. Zhang, W. Zuo, and D. Zhang. From point to set: extend the learning of distance metrics. In ICCV, 2013.
- R. Wang, H. Guo, L. S. Davis, and Q. Dai. Covariance discriminative learning: A natural and efficient approach to image set classification. In CVPR, 2012.
- R. Vemulapalli, J. K. Pillai, and R. Chellappa. Kernel learning for extrinsic classification of manifold features. In CVPR, 2013.
- J. Lu, G. Wang, P. Moulin. Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning. In ICCV, 2013.
- J. Hamm and D. D. Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In ICML, 2008.
- M.T. Harandi, C. Sanderson, S. Shirazi, B.C. Lovell. Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching. In CVPR, 2011.
- J. Hamm and D. D. Lee. Extended grassmann kernels for subspace-based learning. In NIPS, 2008.
- X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. IJCV, 66(1):41–66, 2006.
- 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 Analysis and Applications, 29(1):328–347, 2007.

References

- M. Kim, S. Kumar, V. Pavlovic, and H. Rowley. Face tracking and recognition with visual constraints in real-world videos. In CVPR, 2008.
- Z. Huang, S. Shan, H. Zhang, H. Lao, A. Kuerban, and X. Chen. Benchmarking still-to-video face recognition via partial and local linear discriminant analysis on COX-S2V dataset. In ACCV, 2012.
- J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood components analysis. In NIPS, 2004.
- J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In ICML, 2007.
- K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. JMLR, 10:207–244, 2009.
- A. Sharma and D. Jacobs. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. In CVPR, 2011.
- A. Sharma, A. Kumar, H. Daume, and D. Jacobs. Generalized multiview analysis: A discriminative latent space. In CVPR, 2012.