ECCV2016 Tutorial on “Distance Metric Learning for Computer Vision” Part-3

Distance Metric Learning for Set-based Visual Recognition

Ruiping Wang
Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS)
Oct. 8, 2016
Outline

- Background
- Literature review
- Evaluations
- Summary
Face Recognition with Single Image

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

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

- Personal ID
- Pose
- Illumination
Challenges

- Personal ID
- Pose
- Illumination
Challenges

Personal ID

Pose

Gallery

Same Person

Different Person

Illumination

Probe
Challenges

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

Pose variation

Illumination variation

$D(x, y) = ?$
Face Recognition with Videos

- Video surveillance

Seeking missing children

[Video link 1](http://www.youtube.com/watch?v=M80DXI932OE)

Searching criminal suspects

[Video link 2](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”
S01E06: 00’21”
S01E03: 08’06”
S01E05: 09’23”
S01E01: 22’20”
S01E04: 03’42”
A new & different problem

- Unconstrained acquisition conditions
- Complex appearance variations
- Two phases: set modeling + set matching

New paradigm: set-based metric learning
Outline

- Background
- Literature review
- Evaluations
- Summary
Overview of previous works

From the view of set modeling

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

- **Nonlinear manifold**
  - [Kim, BMVC’05]
  - [Wang, CVPR’08]
  - [Wang, CVPR’09]
  - [Chen, CVPR’13]
  - [Lu, CVPR’15]

- **Affine/Convex hull**
  - [Cevikalp, CVPR’10]
  - [Hu, CVPR’11]
  - [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]
Overview of previous works

- Set modeling
  - Linear subspace $\rightarrow$ Nonlinear manifold
  - Affine/Convex Hull (affine subspace)
  - Parametric PDFs $\rightarrow$ Statistics

- Set matching—basic distance
  - Principal angles-based measure
  - Nearest neighbor (NN) matching approach
  - K-L divergence $\rightarrow$ SPD Riemannian metric...

- Set matching—metric learning
  - Learning in Euclidean space
  - Learning on Riemannian manifold
Set model I: linear subspace

- Properties
  - PCA on the set of image samples to get subspace
  - Loose characterization of the set distribution region
  - Principal angles-based measure discards the varying importance of different variance directions

- Methods
  - MSM [FG’98]
  - DCC [PAMI’07]
  - GDA [ICML’08]
  - GGDA [CVPR’11]
  - PML [CVPR’15]
  - …

- No distribution boundary
- No direction difference
Set model I: linear subspace

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

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

Set model I: linear subspace

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

Set 1: $X_1$

Set 2: $X_2$

Linear subspace by:
*orthonormal basis matrix*

$X_iX_i^T \approx P_i\Lambda_iP_i^T$

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: $\theta_i$

Canonical vectors
DCC

- Discriminative learning
  - Linear transformation
    - \( T: X_i \rightarrow Y_i = T^T X_i \)
  - Representation
    - \( Y_i Y_i^T = (T^T X_i)(T^T X_i)^T \)
    - \( \simeq (T^T P_i) \Lambda_i (T^T P_i)^T \)
  - Set similarity
    - \( F_{ij} = \max_{Q_{ij}, Q_{ji}} tr(M_{ij}) \)
    - \( M_{ij} = Q_{ij}^T P'_i T T^T P'_j Q_{ji} \)
- Discriminant function
  - \( T = \max_T \frac{tr(T^T S_b T)}{tr(T^T S_w T)} \)
Set model I: linear subspace

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


Projection metric

\[ d_P(Y_1, Y_2) = \left( \sum_i \sin^2(\theta_i) \right)^{1/2} = 2^{-1/2} \left\| Y_1 Y_1^T - Y_2 Y_2^T \right\|_F \]

\( \theta_i \): Principal angles

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

\[ d_G^2(Y_1, Y_2) = \sum_i \theta_i^2 \]
GDA/GGDA

- Projection kernel
  - Projection embedding (isometric)
    - $\Psi_P: G(m, D) \rightarrow \mathbb{R}^{D \times D}$, $\text{span}(Y) \rightarrow YY^T$
  - The inner-product of $\mathbb{R}^{D \times D}$
    - $tr((Y_1Y_1^T)(Y_2Y_2^T)) = \|Y_1^TY_2\|^2_F$

- Grassmann kernel (positive definite kernel)
  - $k_P(Y_1, Y_2) = \|Y_1^TY_2\|^2_F$
GDA/GGDA

- Discriminative learning
  - Classical kernel methods using the Grassmann kernel
    - e.g., Kernel LDA / kernel Graph embedding

\[ \alpha^* = \arg \max_\alpha \frac{\alpha^T KWK\alpha}{\alpha^T KK\alpha} \]
Set model I: linear subspace

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

Explicit manifold to manifold mapping

- \( f(Y) = W^T Y \in G(q, d), Y \in G(q, D), \ d \leq D \)

Projection metric on target Grassmann manifold \( G(q, d) \)

- \( d_p^2(f(Y_i), f(Y_j)) = 2^{-1/2} \left\| (W^T Y_i')(W^T Y_j')^T - (W^T Y_j)(W^T Y_j')^T \right\|_F^2 = 2^{-1/2} \text{tr}(P^T A_{ij} A_{ij} P) \)
  - \( A_{ij} = (Y_i' Y_j')^T - Y_j' Y_j')^T \), \( P = WW^T \) is a rank-\( d \) symmetric positive semidefinite (PSD) matrix of size \( D \times D \) (similar form as Mahalanobis matrix)
  - \( Y_i \) needs to be normalized to \( Y_i' \) so that the columns of \( W^T Y_i \) are orthonormal
PML

- Discriminative learning
  - Discriminant function
    - Minimize/Maximize the projection distances of any within-class/between-class subspace pairs
    - $J = \min \sum_{l_i=l_j} \text{tr}(P^T A_{ij} A_{ij} P) - \lambda \sum_{l_i \neq l_j} \text{tr}(P^T A_{ij} A_{ij} P)$
  - Optimization algorithm
    - Iterative solution for one of $Y'$ and $P$ by fixing the other
    - Normalization of $Y$ by QR-decomposition
    - Computation of $P$ by Riemannian Conjugate Gradient (RCG) algorithm on the manifold of PSD matrices
Set model II: nonlinear manifold

- Properties
  - Capture **nonlinear** complex appearance variation
  - Need **dense sampling** and **large amount of data**
  - Less appealing computational efficiency

- Methods
  - MMD [CVPR'08]
  - MDA [CVPR'09]
  - BoMPA [BMVC'05]
  - SANS [CVPR'13]
  - MMDML [CVPR'15]
  - …
Set model II: nonlinear manifold

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


MMD

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

![Diagram showing multi-level MMD framework](image-url)
Formulation & Solution

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

$$d(M_1, M_2) = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d(C_i, C'_j)$$

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

- Subspace-subspace distance (SSD)

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

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

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

\[ = (2 - \cos^2 \theta_0 - \frac{1}{r} \sum_{k=1}^{r} \cos^2 \theta_k)^{1/2} \]
Set model II: nonlinear manifold

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

MDA

- Optimization framework
  - Min. within-class scatter
    \[ S_w = \sum_{m,n} \left\| v^T x_m - v^T x_n \right\|^2 w_{m,n} = 2v^T X (D - W) X^T v \]
  - Max. between-class scatter
    \[ S_b = \sum_{m,n} \left\| v^T x_m - v^T x_n \right\|^2 w'_{m,n} = 2v^T X (D' - W') X^T v \]
- Objective function
  - Global linear transformation
    \[ \text{Maximize } J(v) = \frac{|S_b|}{|S_w|} = \frac{v^T X (D' - W') X^T v}{v^T X (D - W) X^T v} \]
  - Solved by generalized eigen-decomposition
Set model II: nonlinear manifold

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

\[ M(\Theta) = \text{sign} \left[ \sum_{i=1}^{N} w_{i} M(\theta_{i}) - \frac{1}{2} \sum_{i=1}^{N} w_{i} \right] \]

Set model II: nonlinear manifold

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

\[
D(S_a, S_b) = \|U_a - U_b U_b' U_a\|_F^2
\]

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

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

Set model II: nonlinear manifold

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

Class-specific DNN

Model nonlinearity

Objective function

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

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

\[ \min_{f_c} \sum_{i=1}^{N_c} (D_1(h_{ci}^L) - D_2(h_{ci}^L)) \]

Set model III: affine subspace

- **Properties**
  - Linear reconstruction using: mean + subspace basis
  - Synthesized virtual NN-pair matching
  - Less characterization of global data structure
  - Computationally expensive by NN-based matching

- **Methods**
  - AHISD/CHISD [CVPR’10]
  - SANP [CVPR’11]
  - RNP [FG’13]
  - PSDML/SSDML [ICCV’13]
  - PDL [ACCV’16]
  - ...

- Sensitive to noise samples
- High computation cost
Set model III: affine subspace

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

- Subspace spanned by all the available samples
  \[ D = \{d_1, \ldots, d_n\} \] in the set
  - Affine hull
    \[ H(D) = \{D\alpha = \sum d_i \alpha_i | \sum \alpha_i = 1\} \]
  - Convex hull
    \[ H(D) = \{D\alpha = \sum d_i \alpha_i | \sum \alpha_i = 1, 0 \leq \alpha_i \leq 1\} \]

Set model III: affine subspace

  - Improve AHISD/CHISD by imposing **different regularizations** on the linear representation coefficients
  - Metric learning: N/A

SANP, L1-norm regularization

\[ x_i = \mu_i + U_i \cdot \psi_i \]

\[ x_j = \mu_j + U_j \cdot \psi_j \]

SANP, L1-norm regularization

Minimize L2 Distance \( D(x_i, x_j) \)

Maximize Sparsity on \( \alpha \) and \( \beta \)

RNP, L2-norm regularization


Set model III: affine subspace

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

Point-to-set distance

- Basic distance
  \[ d^2(x, D) = \min_\alpha \| x - H(D) \|^2 = \min_\alpha \| x - D\alpha \|^2 \]
  Solution: Least Square Regression or Ridge Regression

- Mahalanobis distance
  \[ d_M^2(x, D) = \min \| P(x - D\alpha) \|^2 = (x - D\alpha)^T P^T P (x - D\alpha) = (x - D\alpha)^T M (x - D\alpha) \]

Projection matrix
Point-to-set distance metric learning (PSDML)

- SVM-based method

\[
\min_{M, \alpha_i(x_i), \alpha_j, \xi_{i,j}, \xi_i, b} \|M\|_F^2 + \nu \left( \sum_{i,j} \xi_{i,j}^N + \sum_i \xi_i^P \right)
\]

s. t. \( d_M(x_i, D_j) + b \geq 1 - \xi_{i,j}^N, j \neq l(x_i); (-) \)
\( d_M(x_i, D_{l(x_i)}) + b \leq -1 + \xi_i^P; (+) \)

\( M \succeq 0, \forall i, j, \xi_{i,j}^N \geq 0, \xi_i^P \geq 0 \)

\[
d^2_M(x, D) = \min \|P(x - D\alpha)\|_2^2
= (x - D\alpha)^T P^T P (x - D\alpha)
= (x - D\alpha)^T M (x - D\alpha)
\]
Set-to-set distance

Basic distance

\[ d^2(D_1, D_2) = \min_{\alpha_1, \alpha_2} \| H(D_1) - H(D_1) \|_2^2 = \min_{\alpha_1, \alpha_2} \| D_1 \alpha_1 - D_2 \alpha_2 \|_2^2 \]

Solution: AHISD/CHISD [Cevikalp, CVPR’10]

Mahalanobis distance

\[ d_M^2(D_1, D_2) = \min \| P(D_1 \alpha_1 - D_2 \alpha_2) \|_2^2 = (D_1 \alpha_1 - D_2 \alpha_2)^T M(D_1 \alpha_1 - D_2 \alpha_2) \]
Set-to-set distance metric learning (SSDML)

**SVM-based method**

\[
\min_{M,\alpha_i,\alpha_j,\alpha_k,\xi_{ij},\xi_{ik},b} \|M\|_F^2 + \nu\left(\sum_{i,j} \xi_{ij}^N + \sum_{i,k} \xi_{ik}^P\right)
\]

s. t. \(d_M(D_i, D_j) + b \geq 1 - \xi_{ij}^N, l(D_i) \neq l(D_j); (-)\)

\(d_M(D_i, D_k) + b \leq -1 + \xi_{ik}^P, l(D_i) = l(D_k); (+)\)

\(M \succeq 0, \forall i, j, k, \xi_{ij}^N \geq 0, \xi_{ik}^P \geq 0\)

\[
d_M^2(D_1, D_2) = \min \|P(D_1 \alpha_1 - D_2 \alpha_2)\|_2^2
\]

\[= (D_1 \alpha_1 - D_2 \alpha_2)^T M (D_1 \alpha_1 - D_2 \alpha_2)\]
Set model III: affine subspace

- PDL (Prototype Discriminative Learning) [ACCV’16]
  - Goal: jointly learn prototypes and linear transformation
    - In the target subspace, for any image in each set, its NN prototype from the same class is closer than that from different classes
  - Metric learning: in Euclidean space

Training process of one sample

The target subspace and prototypes

 Prototype Representation

- Original image sets
  \[ X_c = \{x_{c,1}, \ldots, x_{c,n_c}\} \in \mathbb{R}^{d \times n_c}, c = 1, \ldots, C \]

- Affine hull representation [Cevikalp, CVPR’10]
  - The smallest affine subspace containing affine combinations of sample images in an image set.
  \[ H_c = \left\{ x = \sum_{i=1}^{n_c} \alpha_{c,i} \cdot x_{c,i} \left| \sum_{i=1}^{n_c} \alpha_{c,i} = 1 \right. \right\}, c = 1, \ldots, C \]

- Assumption: Prototypes \( P_c = \{p_{c,1}, \ldots, p_{c,m_c}\} \subseteq H_c \)
  \[ p_{c,i} = \mu + U_c v_{c,i}, v_{c,i} \in \mathbb{R}^l \]
Objective function

\[ \{w^*, P_1^*, ..., P_C^*\} = \arg\min_{W, P_1, ..., P_C} J(W, P_1, ..., P_C) \]

\[ J(W, P_1, ..., P_C) = \sum_{c=1}^{C} \sum_{x \in X_c} S_\beta (Q_x) \]

\[ Q_x = \frac{d(y, nn_w^c (W^T x))}{d(y, nn_b^c (W^T x))} \]

A smooth approximation of the step function:

\[ S_\beta (z) = \frac{1}{1 + e^{\beta (1-z)}} \]

Nearest neighbor prototype from the same class

Nearest neighbor prototype from different classes
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]
  - LERM [CVPR’14]
  - HER [CVPR’15]
  - …

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

---

Set modeling by Covariance Matrix

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

$$X = [x_1, x_2, ..., x_N]_{d \times N}$$

- **COV:** $d \times d$ symmetric positive definite (SPD) matrix

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

*: use regularization to tackle singularity problem
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
    \[ d^2(C_1, C_2) = \| \log_I(C_1^{-1/2}C_2C_1^{-1/2}) \|_F^2 \]
  - Log-Euclidean distance (LED) \([2]\)
    \[ d(C_1, C_2) = \| \log_I(C_1) - \log_I(C_2) \|_F \]

---


Set matching on COV manifold (cont.)

- **Explicit Riemannian kernel feature mapping with LED**

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

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

- 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

feature vectors, e.g. log-COV matrices

class label vectors, e.g. 1/0 indicators

\[ X = T \cdot P' \]
\[ Y = T \cdot B \cdot C' = X \cdot B_{pls} \]

\( T \) is the common latent representation
CDL vs. GDA

- **COV → SPD manifold**
  - **Model**
    \[ C = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \]
  - **Metric**
    \[ d(C_1, C_2) = \| \log_I(C_1) - \log_I(C_2) \|_F \]
  - **Kernel**
    \[ \Psi_{log} : C \to \log_I(C), \ (\mathcal{M} \to \mathbb{R}^{d \times d}) \]

- **Subspace → Grassmannian**
  - **Model**
    \[ C = U \Lambda U^T \]
    \[ \Rightarrow U = [u_1, u_2, \ldots, u_m] \in \mathbb{R}^{D \times m} \]
  - **Metric**
    \[ d_{proj}(U_1, U_2) = 2^{-1/2} \left\| U_1 U_1^T - U_2 U_2^T \right\|_F \]
  - **Kernel**
    \[ \Psi_{proj} : U \to UU^T, \ \mathcal{G}(m, D) \to \mathbb{R}^{d \times d} \]
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 2\textsuperscript{nd}/3\textsuperscript{rd}-order statistics
- Metric learning: in Euclidean space

![Diagram showing 1\textsuperscript{st} / 2\textsuperscript{nd} / 3\textsuperscript{rd}-order statistics and objective function](image)

Objective function

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

\[
T = C \otimes m
\]

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

\[
\max_{M} J = \sum_{i,j=1}^{N} \frac{d(S_i, S_j)}{N_{C^-}} - \sum_{i,j=1}^{N} \frac{d(S_i, S_j)}{N_{C^+}}
\]

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

Non-discriminative
Time-consuming

DARG

**Framework**

- **(a)**
  - ![Image](image1.png)

- **(b)**
  - ![Image](image2.png)

- **(c)**
  - ![Image](image3.png)

**Mathematical Definitions**

- $\mathcal{M}$: Riemannian manifold of Gaussian distributions
- $\mathcal{H}$: high-dimensional reproducing kernel Hilbert space (RKHS)
- $\mathbb{R}^d$: target lower-dimensional discriminant Euclidean subspace
Kernels on the Gaussian distribution manifold
- **kernel based on Lie Group**
- **Distance based on Lie Group (LGD)**

\[ LGD(P_i, P_j) = \| \log(P_i) - \log(P_j) \|_F, \]

\[ g \sim N(x|\mu, \Sigma) \iff P = |\Sigma|^{-1/a+1} \begin{pmatrix} \Sigma + \mu \mu^T & \mu \\ \mu^T & 1 \end{pmatrix} \]

**Kernel function**

\[ K_{LGD}(g_i, g_j) = \exp \left( -\frac{LGD^2(P_i, P_j)}{2t^2} \right) \]
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
  \[ K_{MD+LED}(g_i, g_j) = \gamma_1 K_{MD}(\mu_i, \mu_j) + \gamma_2 K_{LED}(\Sigma_i, \Sigma_j) \]
  \[ K_{MD}(\mu_i, \mu_j) = \exp\left( -\frac{MD^2(\mu_i, \mu_j)}{2t^2} \right) \]
  \[ K_{LED}(\Sigma_i, \Sigma_j) = \exp\left( -\frac{LED^2(\Sigma_i, \Sigma_j)}{2t^2} \right) \]

\( \gamma_1, \gamma_2 \) are the combination coefficients
DARG

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

\[ J(\alpha) = \frac{|\alpha^T B \alpha|}{|\alpha^T W \alpha|} \]

\[
W = \sum_{i=1}^{C} \frac{1}{\omega_i} \sum_{j=1}^{N_i} w_j^{(i)} (k_i^j - m_i)(k_i^j - m_i)^T
\]

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

\[
m_i = \frac{1}{N_i \omega_i} \sum_{j=1}^{N_i} w_j^{(i)} k_i^j, m = \frac{1}{N} \sum_{i=1}^{C} \frac{1}{\omega_i} \sum_{j=1}^{N_i} w_j^{(i)} k_i^j
\]
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.

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**
    \[
    \hat{\delta}_H^2(p||q) = \frac{1}{n_p} \sum_{i}^{n_p} (2T(x_i^p) - 1) \ln \frac{T(x_i^p)}{1 - T(x_i^p)} + \frac{1}{n_q} \sum_{i}^{n_q} (2T(x_i^q) - 1) \ln \frac{T(x_i^q)}{1 - T(x_i^q)}
    \]
    \[
    T(x) = \frac{p(x)}{p(x) + q(x)}
    \]
Beyond Gauss

- Kernels on the Statistical Manifold
  - Hellinger Kernel
    \[ K_H(p, q) = \exp(-\sigma \delta_H^2(p, q)) \]
  - Laplace Kernel
    \[ K_L(p, q) = \exp(-\sigma \delta_H(p, q)) \]
  - Jeffrey Kernel
    \[ K_J(p, q) = \exp(-\sigma \delta_J(p||q)) \]

- Dimensionality Reduction
  \[ W^* = \arg\min_W L(W), \text{s.t. } W^T W = I_d \]
  \[ L(W) = \sum_{i,j} a(X_i, X_j) \cdot \delta(W^T X_i, W^T X_j) \]

- High affinity \( a(X_i, X_j) \Rightarrow \text{small distance after mapping} \)
- Low/negative affinity \( a(X_i, X_j) \Rightarrow \text{large distance after mapping} \)
- 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]
  - LERM [CVPR’14]
  - HER [CVPR’15]
  - …

[Natural raw statistics]
[More flexible model]

[Shakhnarovich, ECCV’02]
[Arandjelović, CVPR’05]
SPD-ML (SPD Manifold Learning) [ECCV’14]
- Pioneering work on explicit manifold-to-manifold dimensionality reduction
- Metric learning: on Riemannian manifold

Set model IV: statistics (COV+)

SPD-ML

- SPD manifold dimensionality reduction
  - Mapping function: \( f: S_{++}^n \times \mathbb{R}^{n \times m} \to S_{++}^m \)
    - \( f(X, \overrightarrow{W}) = \overrightarrow{W}^T X \overrightarrow{W} \in S_{++}^m > 0, X \in S_{++}^n, \overrightarrow{W} \in \mathbb{R}^{n \times m} \) (full rank)

- Affine invariant metrics: AIRM / Stein divergence on target SPD manifold \( S_{++}^m \)
  - \( \delta^2(\overrightarrow{W}^T X_i \overrightarrow{W}, \overrightarrow{W}^T X_i \overrightarrow{W}) = \delta^2(\overrightarrow{W}^T X_i \overrightarrow{W}, \overrightarrow{W}^T X_j \overrightarrow{W}) \)
    - \( \overrightarrow{W} = MW, M \in GL(n), W \in \mathbb{R}^{n \times m}, \overrightarrow{W}^T \overrightarrow{W} = I_m \)
SPD-ML

- Discriminative learning
  - Discriminant function
    - Graph Embedding formalism with an affinity matrix that encodes intra-class and inter-class SPD distances
    - \( \min L(W) = \min \sum_{ij} A_{ij} \delta^2(W^T X_i W, W^T X_j W) \)
      - s. t. \( W^T W = I_m \) (orthogonality constraint)
  - Optimization
    - Optimization problems on Stiefel manifold, solved by nonlinear Conjugate Gradient (CG) method
Set model IV: statistics (COV+)

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

LEML

- SPD tangent map learning
  - Mapping function: $DF: f(\log(S)) = W^T \log(S) W$
    - $W$ is column full rank
  - Log-Euclidean distance in the target tangent space
    - $d_{LED}(f(T_i), f(T_j)) = \|W^T T_i W - W^T T_j W\|_F$
      $$= \text{tr}(Q(T_i - T_j)(T_i - T_j))$$
      - $Q = (WW^T)^2$
      - $T = \log(S)$
      - Analogy to 2DPCA
LEML

- Discriminative learning
  - Objective function
    \[
    \arg \min_{Q, \xi} D_{ld}(Q, Q_0) + \eta D_{ld}(\xi, \xi_0)
    \]
    s. t.
    \[
    \text{tr}(QA_{ij}^T A_{ij}) \leq \xi_{c(i,j)}, (i, j) \in S
    \]
    \[
    \text{tr}(QA_{ij}^T A_{ij}) \geq \xi_{c(i,j)}, (i, j) \in D
    \]
    \[
    A_{ij} = \log(C_i) - \log(C_j), D_{ld}: \text{LogDet divergence}
    \]
  - 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]
- LERM [CVPR’14]
- HER [CVPR’15]
- …

Natural raw statistics
More flexible model

[Shakhnarovich, ECCV’02]
[Arandjelović, CVPR’05]
Set model IV: statistics (COV+)

LERM (Learning Euclidean-to-Riemannian Metric) [CVPR’14]

- Application scenario: still-to-video face recognition
- Metric learning: cross Euclidean space and Riemannian manifold

Watch list

Surveillance video

Watch list screening

Point-to-Set Classification

Point

Set

Point-to-Set Classification

- Euclidean points vs. Riemannian points

Corresponding manifold:
1. Grassmann (G)
2. Affine Grassmann (A)
3. SPD (S)

Set model:
- 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]
Basic idea

- Reduce Euclidean-to-Riemannian metric to classical Euclidean metric

Seek maps $F, \Phi$ to a common Euclidean subspace

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

Same:
Objective fun. $E(F, \phi)$

Mapping Mode 1
Single Hilbert space

Further reduce the gap

Mapping Mode 2
Double Hilbert spaces

Further explore correlations

Mapping Mode 3
+Cross-space kernel
**LERM**

- e.g. Mode 1

**Single Hilbert space**

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

**Objective function:**
\[
\min_{F,\Phi} \{ D(F, \Phi) + \lambda_1 G(F, \Phi) + \lambda_2 T(F, \Phi) \}
\]

**Final maps:**
\[
F = f_x = W^T_x X \\
\Phi = f_y \circ \varphi_y = W^T_y K_y \\
K_y(i, j) = \exp(-d^2(y_i, y_j)/2\sigma^2)
\]

Riemannian metrics [ICML’08, NIPS’08, SIAM’06]
Set model IV: statistics (COV+)

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

Image Query

Video Database

S01E01: 10’48”
S01E06: 05’22”
S01E02: 04’20”
S01E02: 00’21”
S01E03: 08’06”
S01E05: 09’23”
S01E01: 22’20”
S01E04: 03’42”
S01E03: 10’53”
S01E04: 12’11”
S01E02: 13’23”
S01E05: 19’20”

Heterogeneous Hash Learning

Previous Multiple Modalities Hash Learning Methods

The Proposed Heterogeneous Hash Learning Method
Two-Stage Architecture

- **Stage-1**: kernel mapping

- **Stage-2**: feature mapping

Euclidean Kernel (Image)

\[ K_e^{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \]

Riemannian Kernel (Video)

\[ K_r^{ij} = \exp\left(-\frac{\|\log(y_i) - \log(y_j)\|_F^2}{2\sigma^2}\right) \]

## Two-Stage Architecture

- **Stage-2**: binary encoding

**Hash Functions**

\[ \omega_e \]

**Target Hamming Space**

**Hash Functions**

\[ \omega_r \]

**Reproducing Kernel Hilbert Space (Euc.)**

\[ \varphi \]

**Euclidean Space**

(\textbf{Image-Vector-}x_i)

**Reproducing Kernel Hilbert Space (Rie.)**

\[ \eta \]

**Riemannian Manifold**

(\textbf{Video-Covariance Matrix-}y_i)

**Hyperplane**

- Discriminability
- Stability

- Subject A
- Subject B
- Subject C
**HER**

- Binary encoding [Rastegari, ECCV’12]

**Objective Function**

\[
\min_{\omega_e, \omega_r, \xi_e, \xi_r, B_e, B_r} \lambda_1 E_e + \lambda_2 E_r + \lambda_3 E_{er} + \gamma_1 \sum_{k \in \{1:K\}} \|\omega^k_e\|^2 + C_1 \sum_{k \in \{1:K\}} \xi^k_e + \gamma_2 \sum_{k \in \{1:K\}} \|\omega^k_r\|^2 + C_2 \sum_{k \in \{1:K\}} \xi^k_r
\]

**st.**

- **Discriminability (LDA)**
  
  \[
  B^k_i = \text{sgn} \left( \omega^T_e \varphi(x_i) \right), \forall k \in \{1:K\}, \forall i \in \{1:N\}
  \]

- **Stability (SVM)**
  
  \[
  B^k_i = \text{sgn} \left( \omega^T_r \eta(y_i) \right), \forall k \in \{1:K\}, \forall i \in \{1:N\}
  \]

- **Compatibility (Cross-training)**
  
  \[
  B^k_i \left( \omega^T_e \varphi(x_i) \right) \geq 1 - \xi^k_e, \xi^k_e > 0, \forall k \in \{1:K\}, \forall i \in \{1:N\}
  \]

  \[
  B^k_i \left( \omega^T_r \eta(y_i) \right) \geq 1 - \xi^k_r, \xi^k_r > 0, \forall k \in \{1:K\}, \forall i \in \{1:N\}
  \]
DL extension for video hash learning

- **DVC (Deep Video Code) [ACCV’16]**
  - Application scenario: video-video face retrieval
  - Metric learning: hamming distance learning

An end-to-end video hashing framework

- A multi-branch CNN for frame-level feature extraction
- Temporal pooling for fusing complementary feature
- Learning to hash with upper bounded triplet loss function

Outline

- Background
- Literature review
- Evaluations
- Summary
Evaluations

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

---

**Table 2. Summary of Video PaSC Data.**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Subjects</td>
<td>265</td>
</tr>
<tr>
<td>Total Videos</td>
<td>2,802</td>
</tr>
<tr>
<td>Total Control Videos</td>
<td>1,401</td>
</tr>
<tr>
<td>Total Handheld Videos</td>
<td>1,401</td>
</tr>
<tr>
<td>Control Videos per Subject</td>
<td>4 to 7</td>
</tr>
<tr>
<td>Handheld Videos per Subject</td>
<td>4 to 7</td>
</tr>
<tr>
<td>Number of Locations</td>
<td>6</td>
</tr>
</tbody>
</table>
Evaluations

Results (*reported in our DARG paper*)

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

Evaluations

Results (*reported in our DARG paper*)

VR@FAR=0.01 on PaSC

AUC on YTF
Evaluations

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

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

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

Route map of our methods

- **Complex distribution**
- **Large amount of data**

**Covariance matrix**
- Natural raw statistics
- No assum. of data dist.

**MMD**
- CVPR’08/TIP’12
- mul./mean/disc.
- single/cov./disc.

**MDA**
- CVPR’09
- +COV

**CDL**
- CVPR’12
- mul. / +mean

**MDA**
- CVPR’09

**CDL**
- CVPR’12

**MDA**
- CVPR’09

**CDL**
- CVPR’12

**MDA**
- CVPR’09

**CDL**
- CVPR’12

**LERM**
- CVPR’14
- heterogeneous

**LERM**
- CVPR’14
- binary

**HER**
- CVPR’15
- heter.

**LEML/PML**
- ICML’15/CVPR’15
- binary

**DARG**
- CVPR’15

**CVC**
- BMVC’14
- binary

**DARG**
- CVPR’15

**CVC**
- BMVC’14
- binary

**DARG**
- CVPR’15

**CVC**
- BMVC’14
- binary

**DARG**
- CVPR’15

**HER**
- CVPR’15

**LEML/PML**
- ICML’15/CVPR’15

**HER**
- CVPR’15
Summary

- What we learn from current studies
  - Set modeling
    - Linear(/affine) subspace $\rightarrow$ Manifold $\rightarrow$ Statistics
  - Set matching
    - Non-discriminative $\rightarrow$ Discriminative
  - Metric learning
    - Euclidean space $\rightarrow$ 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)

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