Person Identification in Large Scale Camera Networks

Wei-Shi Zheng (郑伟诗)

Sun Yat-sen University

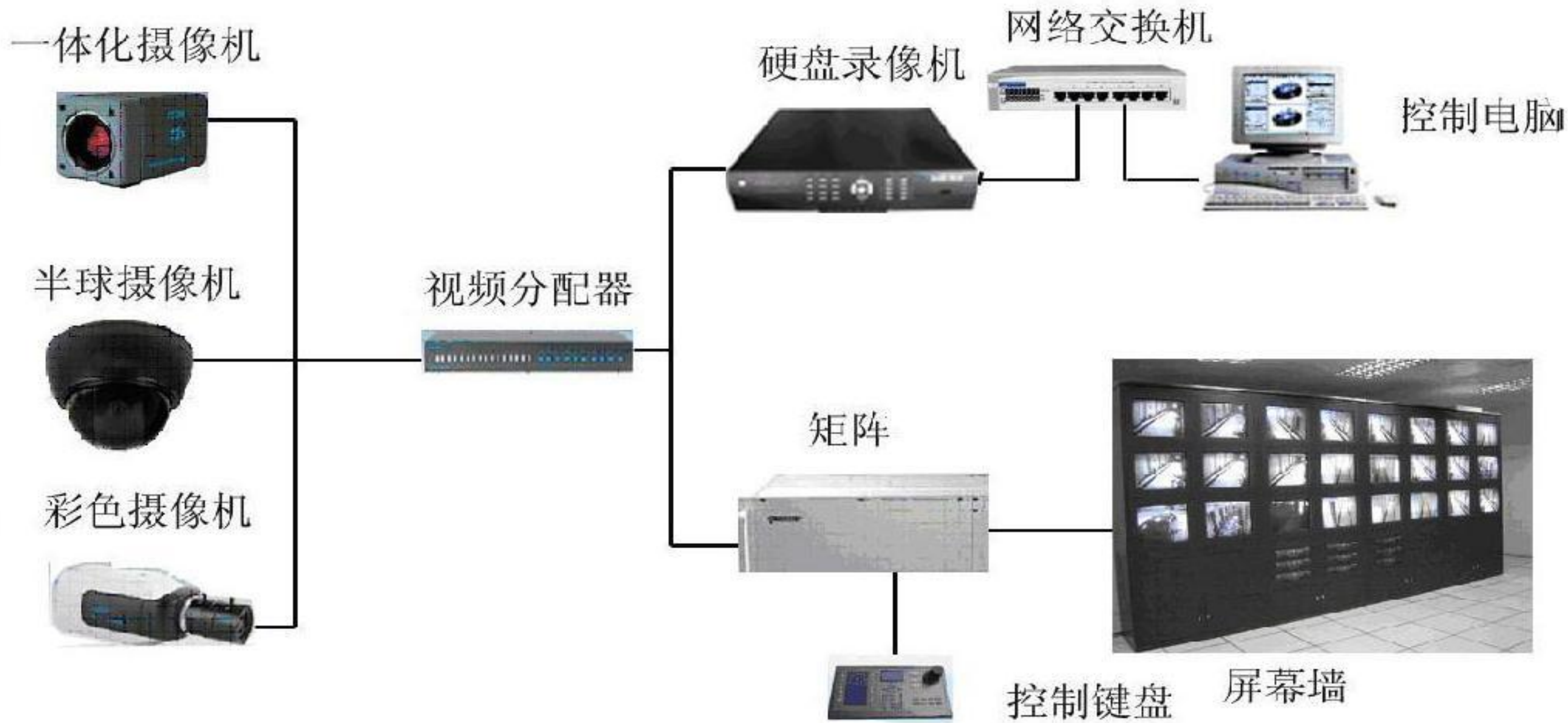
中山大學
SUN YAT-SEN UNIVERSITY

Visual Surveillance

中山大學
SUN YAT-SEN UNIVERSITY

Visual Surveillance

A Typical System



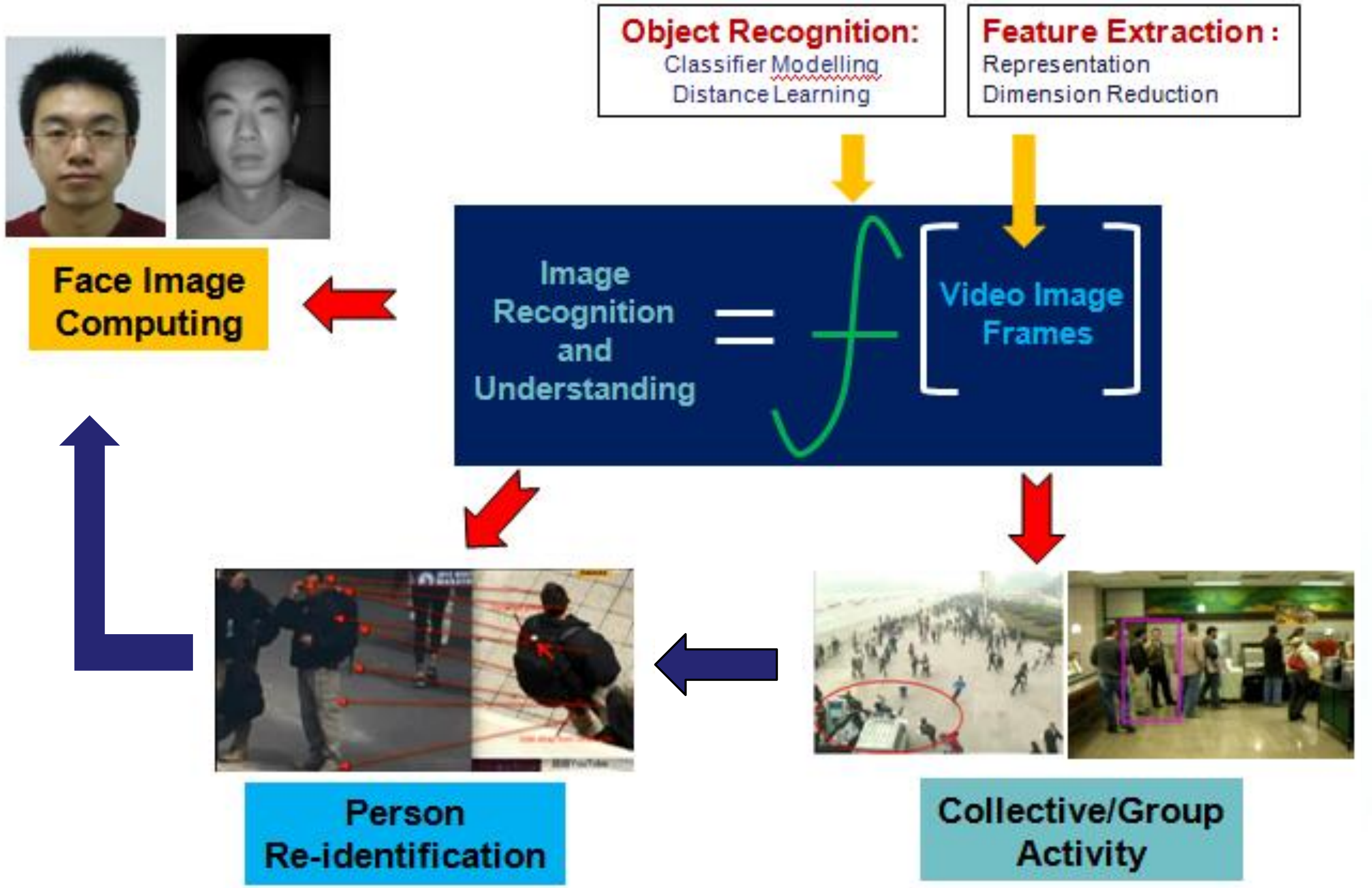
Visual Surveillance

- **Real-world**

Visual Surveillance

- **Real-world**

Our Research



中山大學
SUN YAT-SEN UNIVERSITY

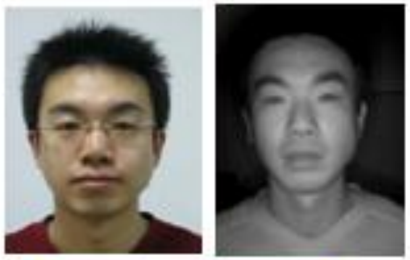
Large Scale Person Identification

- Large in Understanding Group Activity
- Large in Tracking Group and People Across Disjoint Camera View
- Large in Making the Identification Efficiently Deployed in Many Scenarios
- Large in Processing Data and Exploring Share Information Across Different Types of Data

中山大學 SUN YAT-SEN UNIVERSITY

Relation to Human Identification

- Concern the person who is joining an activity
- Tracking him/her across camera-views
- Identifying him/her when we can capture him/her face very well
- Recognising/Searching face images in a Large Dataset



Face Image Computing

Person Re-identification

Collective/Group Activity

Outline

- **Group Activity & Recognition**
- **Cross View Tracking: Person Re-identification**
  - Robust Metric
  - View Change Invariant Features
  - Cross Scenario Transfer
  - Open-world Modelling
- **Online Classification for Identification**

Group Activity: Interaction Modelling

Xiaobin Chang (student), Wei-Shi Zheng*, and Jianguo Zhang. Learning Person-Person Interaction in Collective Activity Recognition. IEEE Transactions on Image Processing, vol. 24, no. 6, pp. 1905-1918, 2015.

Introduction to Group Activity

- **Why Learning Interaction Activity**

We are interacting with others everyday

中山大學 SUN YAT-SEN UNIVERSITY

Introduction to Group Activity

Terrorist Attack in Kunming

BUS Explosion in Guangzhou

中山大學
SUN YAT-SEN UNIVERSITY

Introduction to Group Activity

中山大學
SUN YAT-SEN UNIVERSITY

Introduction to Group Activity

- ## The Challenges

Talking?? **Queuing??**

Individual Action

≈

Local does not mean global

Collective Activity

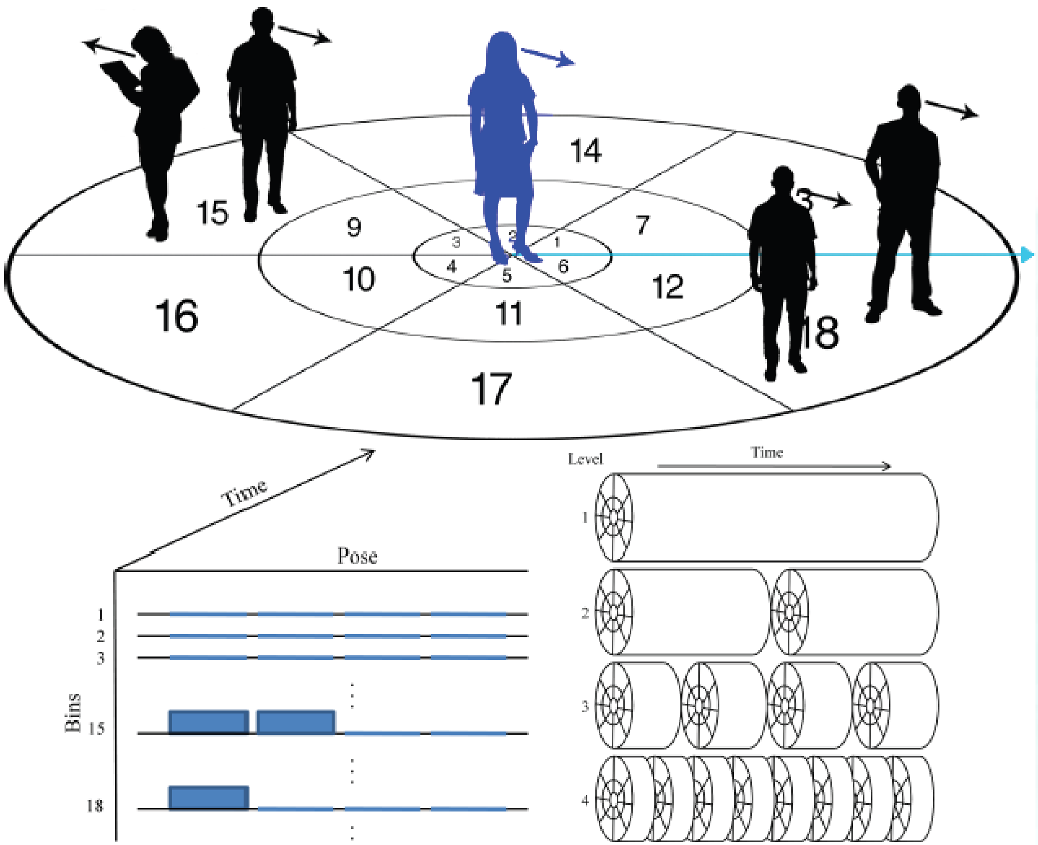
≠

Talking!! **Queuing!!**

SUN YAT-SEN UNIVERSITY 中山大學

Learning Person–Person Interaction

- **Related Work: Spatial Temporal Model**



Choi et al 09'ICCVW

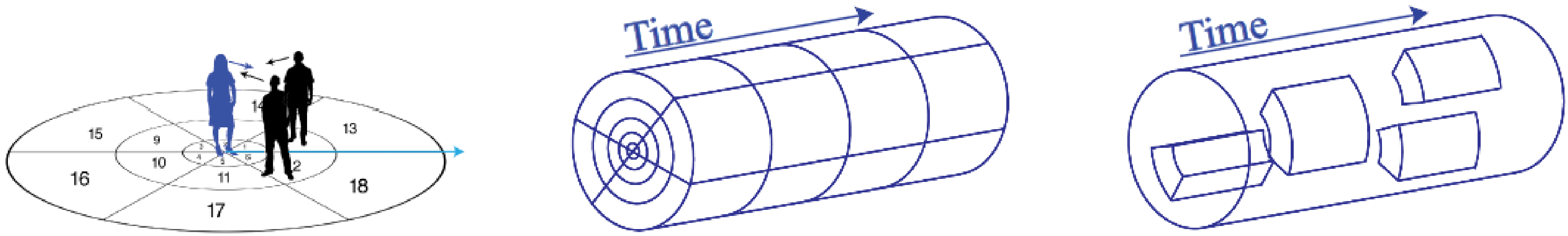
1.Capturing the Spatial Distribution of Collective Activity.

2.Capturing the Temporal Variation of the Spatial Distribution .

中山大學 SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction

- **Related Work: Spatial Temporal Model**



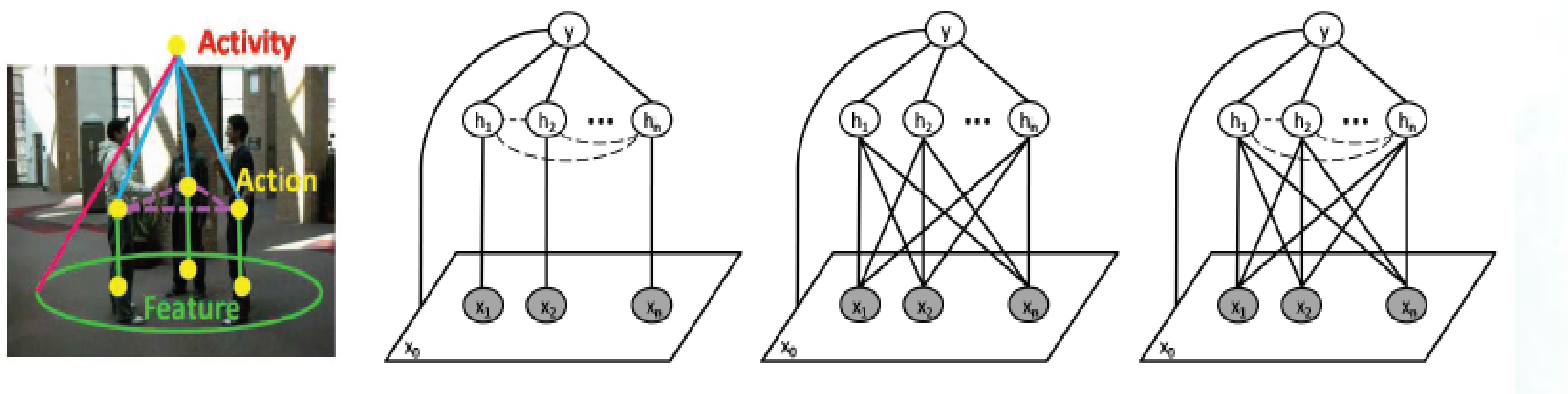
(a) (b) (c)

Choi et al 11' CVPR

Capturing the Spatial Temporal Information and finding out the most Discriminative ones for Collective Activity Recognition as well.

Learning Person–Person Interaction

Related Work: Hierarchical Model



(a) (b) (c) (d)

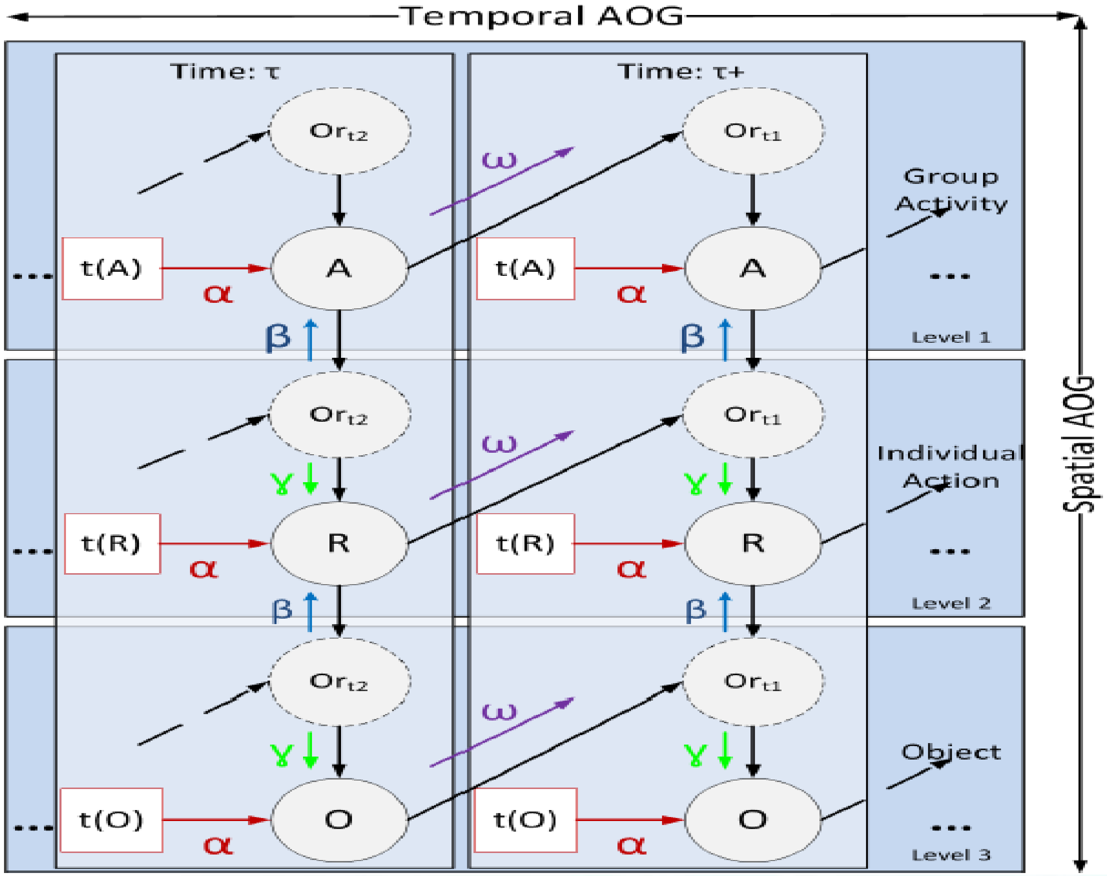
Lan et al 12' TPAMI

1.Collective Activity is based on the action of each person.

2.The connections among people can be inferred as latent variables.

Learning Person–Person Interaction

Related Work: Hierarchical Model



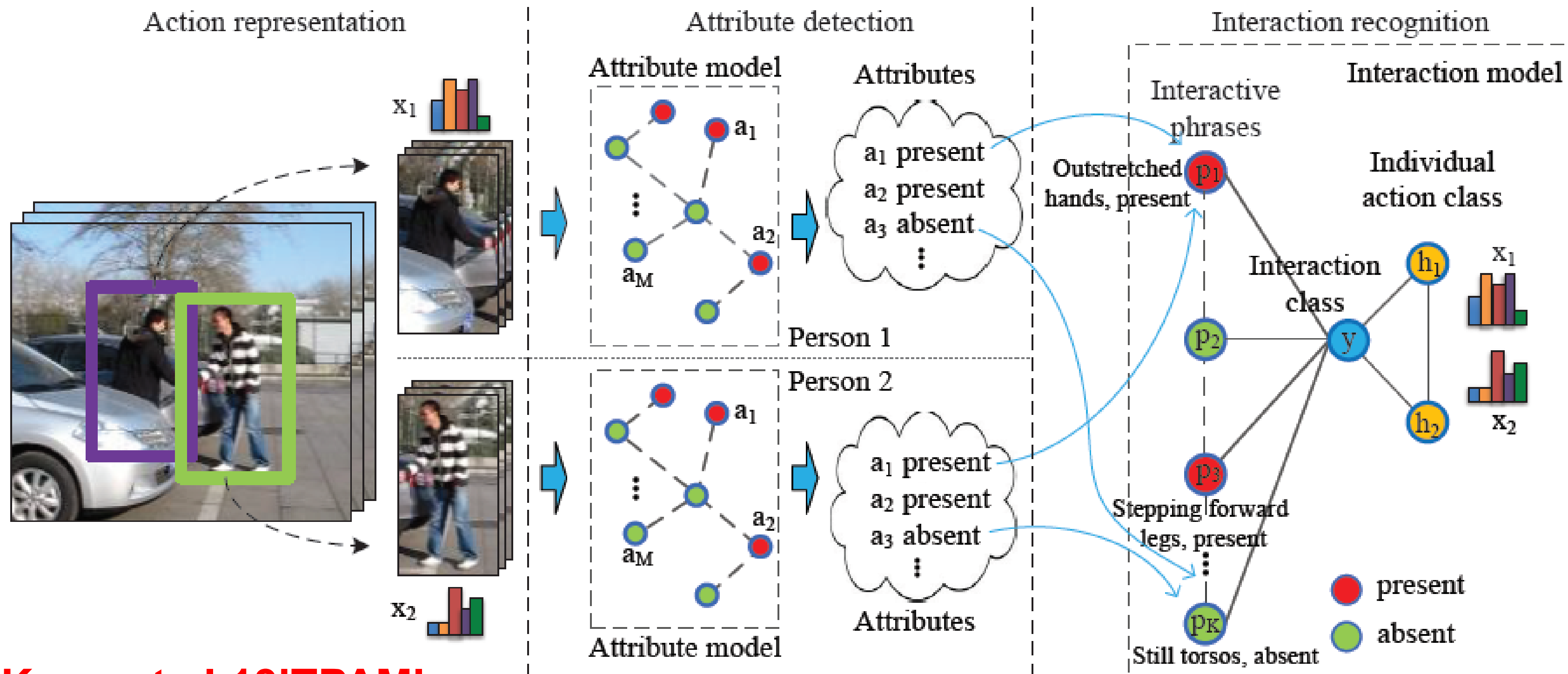
Amer et al 13'ICCV

1. Three layers are used. They are Object level, Individual Action level and Collective Activity Level, from bottom to top.

2. An And-Or Graph is used for modelling these three layers.

Learning Person–Person Interaction

Related Work: Interactive Phrase



Kong et al 13'TPAMI

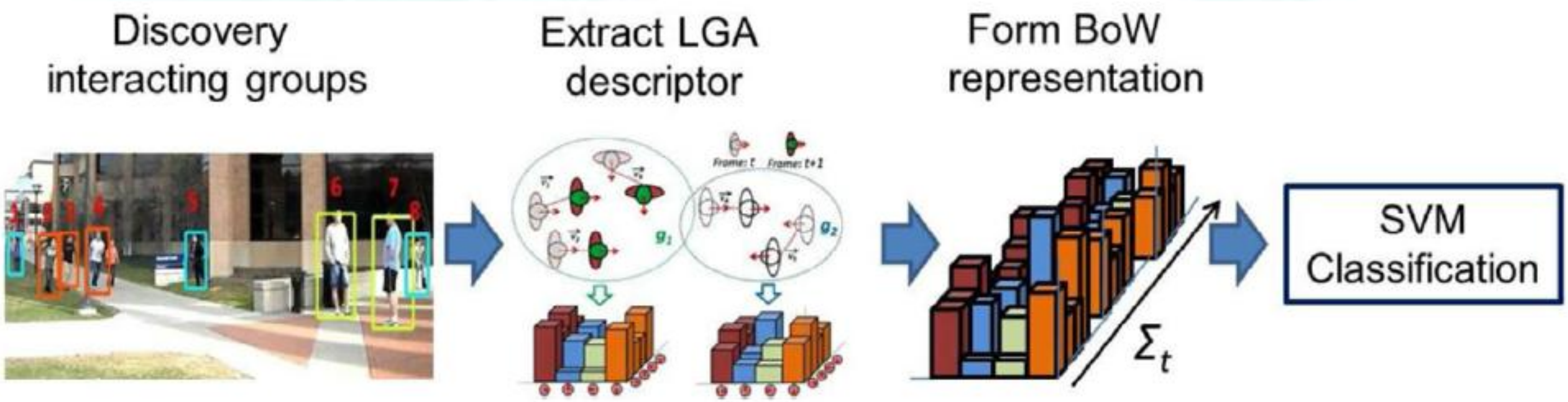
1.Describing the person-person interaction by capturing the interaction patterns by exploiting motion relationships between body parts.

2. The interaction is inexplicitly captured by the model.

中山大學 SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction

Related Work: Interactive Descriptor

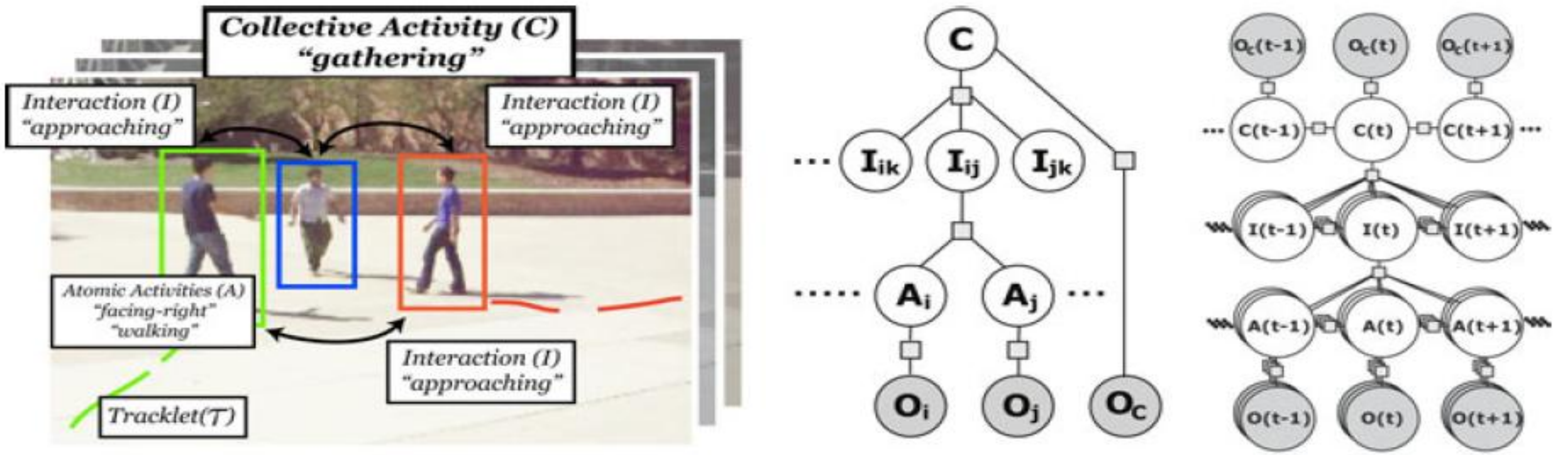


Tran et al 14' Pattern Recognition Letters

A Descriptor called LGA is used to capture the interactions among people for Collective Activity Recognition.

Learning Person–Person Interaction

Related Work: Combine Model



Choi et al 14'TPAMI

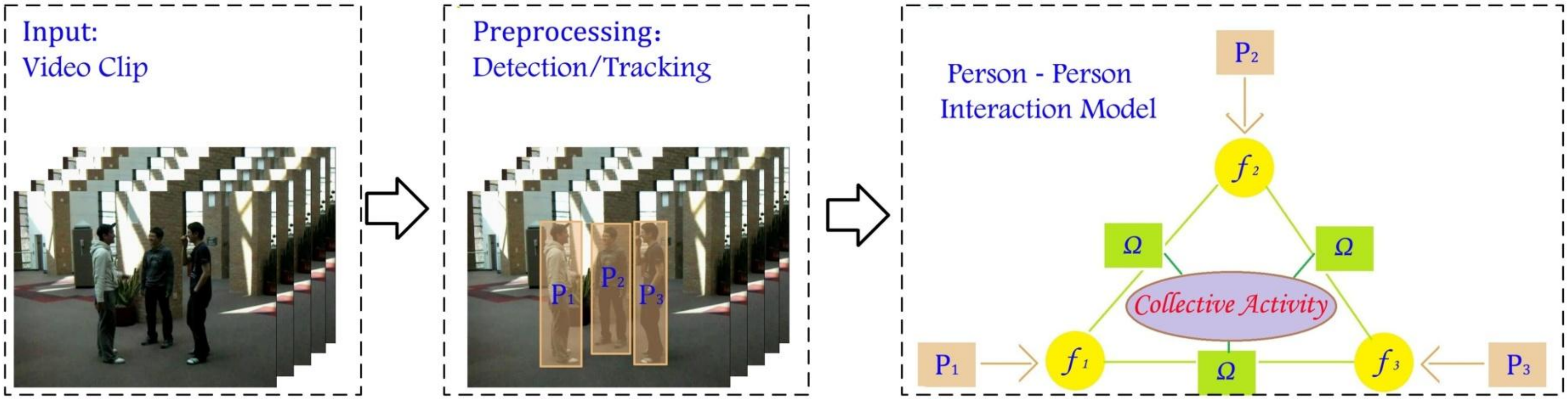
1. This model combines Different Tasks together:
Collective Activity Recognition, Interaction Recognition, Individual Action Recognition, as well as Multi-people Tracking.

2. It believes different tasks can benefit from each others during learning procedure.

3.Hard to be optimised & Require many manual labels.

Learning Person–Person Interaction

A Complete Learning Approach



Short Video Clips (~15 frames)

Spatial-Temporal Feature Of Each Person's Action

Focus on Modeling Person-Person Interaction

Two connected atomic activities in one collective activity are either:

1) quite similar and spatially close to each other to form a meaningful collective activity (e.g. two people are walking together);

2) not quite similar but are strongly interacting to each other (e.g. facing each other when two people are talking, or fighting).

SUN YAT-SEN UNIVERSITY

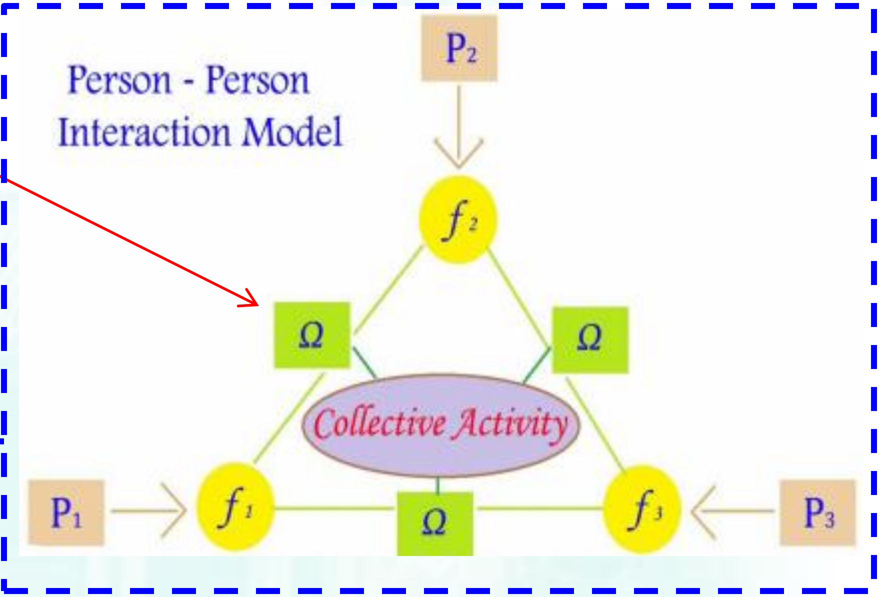
Learning Person–Person Interaction

- **Inference**

Interaction

person-person interaction under the collective activity m

$$R_{m,q} = \sum_{i,j=1,\ i\neq j}^{N_q} \boxed{f_{i,q}^{'} \Omega_m f_{j,q}}$$



Summarising all the person-person interactions

Inference:

$$\hat{l}_q = \arg\max_{m\in\gamma} R_{m,q}$$

The Interaction Response should be maximised under the ground-truth collective activity

Learning Person–Person Interaction

- **Learning**

$$\min_{\Omega_m} \frac{1}{2}\|\Omega_m\|_F^2 + C\sum_{t=1}^{T} max(0, 1 - y_t^m(\sum_{i,j=1,i\neq j}^{N_t} f_{i,t}'\Omega_m f_{j,t}))^2$$

$$s.t\ rank(\Omega_m) < v$$

Matrix Factorisation:

$$\Omega_m = L_m * L_m'$$

$$L_m \in R^{\ell\times d} \quad d \ll \ell$$

Advantages:
1. More Effective
2. Low-rank Representation

$$\min_{L_m} J(L_m) = \min_{L_m} \frac{1}{2}\|L_m\|_F^2 - \boxed{\frac{\beta}{2} log\ det(L_m' L_m)}$$
$$+ C\sum_{t=1}^{|T|} \max(0, 1 - y_t^m(\sum_{i,j=1}^{N_t} f_{i,t}' L_m L_m' f_{j,t}))^2$$

- *log det* regularisation Term: *Lm* is of full rank. (Avoid the redundant problem)

where $\beta \geqslant 0$, $C \geqslant 0$.

Learning Person–Person Interaction

Multi-task Extension

(a)过马路（Crossing） (b)等待（Waiting） (c)排队（Queuing）

(d)行走（Walking） (e)交谈（Talking）

Different Collective Activities are *Different but Related*.

- **Class-Specific: Global Interaction is different;**
- **Shared Aspects: Local Interaction is sometimes similar**

  Person Actions(standing, walking),

  Spatial Distribution, etc.

中山大學 SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction

Multi-task Extension

$$R_{m,q} = \sum_{i,j=1,\ i\neq j}^{N_q} f_{i,q}^{'} \Omega_m f_{j,q}$$

α controls the balance between shared variable and class-specific variable in Ω_m

$$\Omega_m = (1 - \alpha)\bar{\Omega}_0 + \alpha\bar{\Omega}_m,\ \alpha \in [0,1]$$

$\bar{\Omega}_m$ for modelling Class-Specific Information

$\bar{\Omega}_0$ for modelling Shared Information

Learning Person–Person Interaction

Multi-task Extension

$$\min_{L_m} J(L_m) = \min_{L_m} \frac{1}{2}\|L_m\|_F^2 - \frac{\beta}{2} log\ det(L_m' L_m)$$

$$+ C \sum_{t=1}^{|T|} \max(0, 1 - y_t^m (\sum_{i,j=1}^{N_t} f_{i,t}' L_m L_m' f_{j,t}))^2$$

$$\Omega_m = (1 - \alpha)\overline{\Omega}_0 + \alpha\overline{\Omega}_m, \ \alpha \in [0,1]$$

$$\overline{\Omega}_m = \overline{L}_m * \overline{L}_m', \overline{L}_m \in R^{l\times d}$$

$$\overline{\Omega}_0 = \overline{L}_0 * \overline{L}_0', \overline{L}_0 \in R^{l\times d}$$

$$\min_{\overline{L}_0, \overline{L}_m} J(\overline{L}_0, \overline{L}_m)$$

$$= \frac{1}{2} \sum_{k=0,m} \{\|\overline{L}_k\|_F^2 - \frac{\beta}{2} log\ det(\overline{L}_k' \overline{L}_k)\}$$

$$+ C \sum_{t=1}^{|T|} Loss(\overline{L}_0, \overline{L}_m)$$

Gradient Descent & Interactive Optimization

$$Loss(\overline{L}_0, \overline{L}_m)$$

$$= \max(0, 1 - y_t^m (\sum_{i,j=1}^{N_t} f_{i,t}'((1-\alpha)\overline{L}_0 \overline{L}_0' + \alpha \overline{L}_m \overline{L}_m') f_{j,t}))^2$$

Learning Person–Person Interaction

Two Benchmark Datasets

1. Collective Activity Dataset (CAD)

- 44 video sequences; 5 activities (crossing, waiting, queuing, walking, talking);
- Exp. Setting: random splits 1/4 of the dataset for testing and the rest for training.

(a) Crossing (b) Waiting (c) Queuing (d) Walking (e) Talking

2. Choi's Dataset

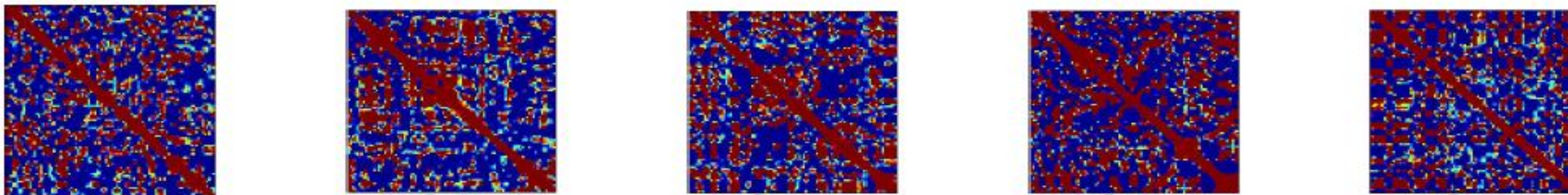
- 32 video sequences; 6 activities(gathering, talking, dismissal, walking together, chasing, and queuing);
- Exp. Setting: the standard experimental protocol of the 3-fold cross validation.

(f) Gathering (g) Talking (h) Dismissal (i) Walking (j) Chasing (k) Queuing

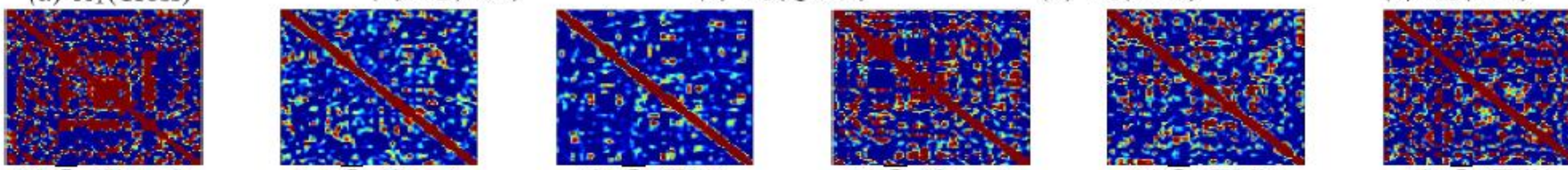
Learning Person–Person Interaction

- **Visualization of the Learned Ω**

CAD dataset:

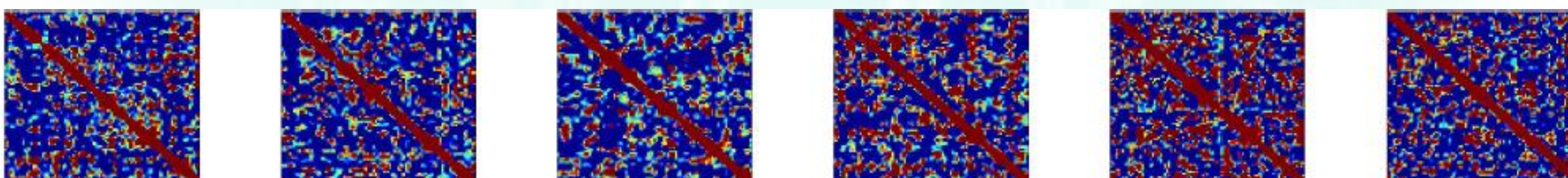


(a) Ω_1(Cross) *(b)* Ω_2(Wait) *(c)* Ω_3(Queue) *(d)* Ω_4(Walk) *(e)* Ω_5(Talk)

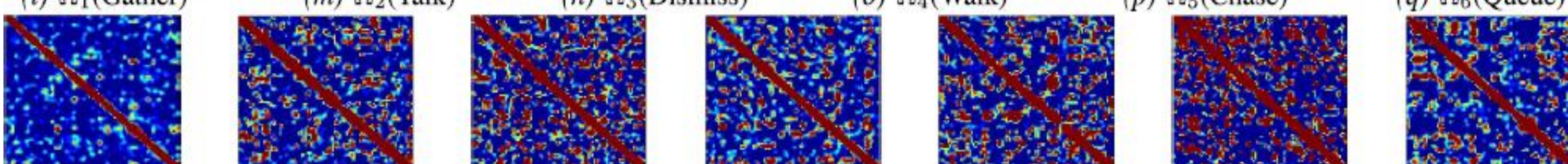


(f) Ω_0(Shared) *(g)* Ω_1(Cross) *(h)* Ω_2(Wait) *(i)* Ω_3(Queue) *(j)* Ω_4(Walk) *(k)* Ω_5(Talk)

Choi's dataset:



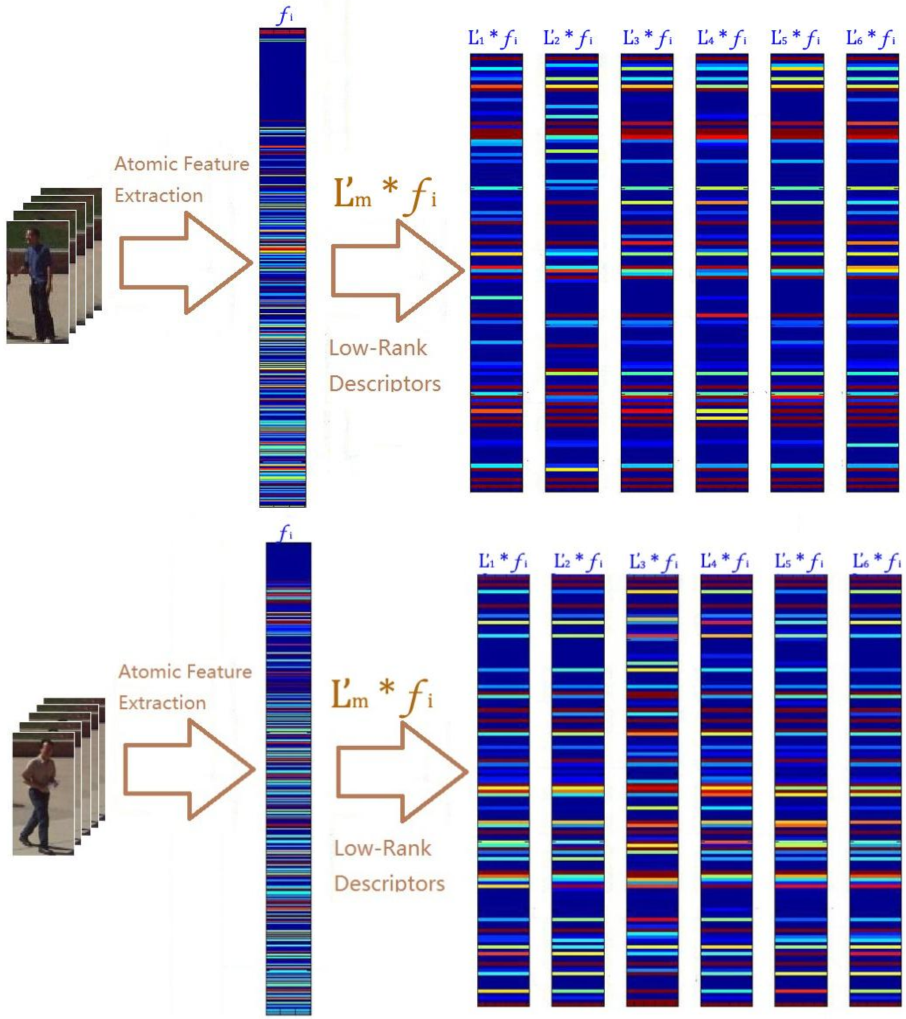
(l) Ω_1(Gather) *(m)* Ω_2(Talk) *(n)* Ω_3(Dismiss) *(o)* Ω_4(Walk) *(p)* Ω_5(Chase) *(q)* Ω_6(Queue)



(r) Ω_0(Shared) *(s)* Ω_1(Gather) *(t)* Ω_2(Talk) *(u)* Ω_3(Dismiss) *(v)* Ω_4(Walk) *(w)* Ω_5(Chase) *(x)* Ω_6(Queue)

Learning Person–Person Interaction

Low-Rank Representation



For Each Collective activity *m*, each individual feature $f_i \in R^l$ has the corresponding low-rank representation:

$$r_{i,m} = L'_m f_i, \quad r_{i,m} \in R^d$$

$$d \ll l$$

$r_{i,m}$ **is:**

1.Class-Specified Feature

2. Learned Individual Representation for Person-Person Interaction

SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction

Results On Two Benchmark Datasets

CAD:

Class	Baseline	[24]	[23]	[26]	IR	MIR
Crossing	62.3	68.0	65.0	**77.0**	72.3	65.9
Waiting	55.5	69.0	60.0	63.0	76.3	**82.2**
Queuing	**98.6**	76.0	96.0	70.0	90.0	91.9
Walking	66.8	80.0	68.0	73.0	77.5	**81.4**
Talking	91.9	**99.0**	**99.0**	88.0	93.3	95.2
Average	75.0	78.4	77.5	74.2	81.9	**83.3**

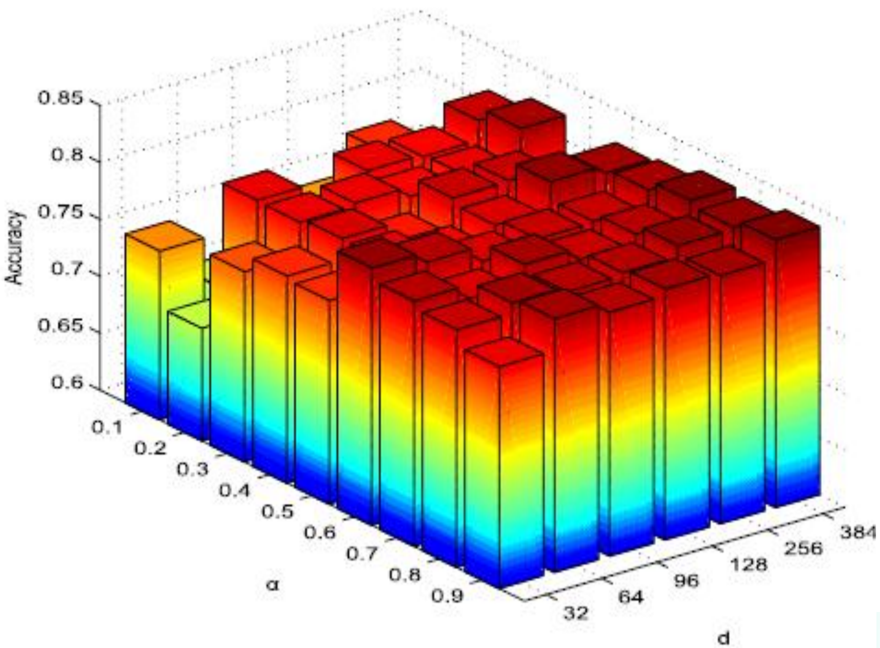
Choi's Dataset:

Class	Baseline	[12]	[11]	[4]	IR	MIR
Gathering	**64.1**	50.0	43.5	48.1	55.2	59.9
Talking	96.5	72.2	82.2	81.3	94.3	**97.0**
Dismissal	76.4	49.2	77.0	55.3	**91.8**	90.5
Walking	90.4	83.2	87.4	89.1	93.4	**94.3**
Chasing	21.6	95.2	91.9	**95.9**	42.2	53.9
Queuing	78.7	95.9	93.4	**96.7**	84.3	86.3
Average	71.3	74.3	79.2	77.7	76.9	**80.3**

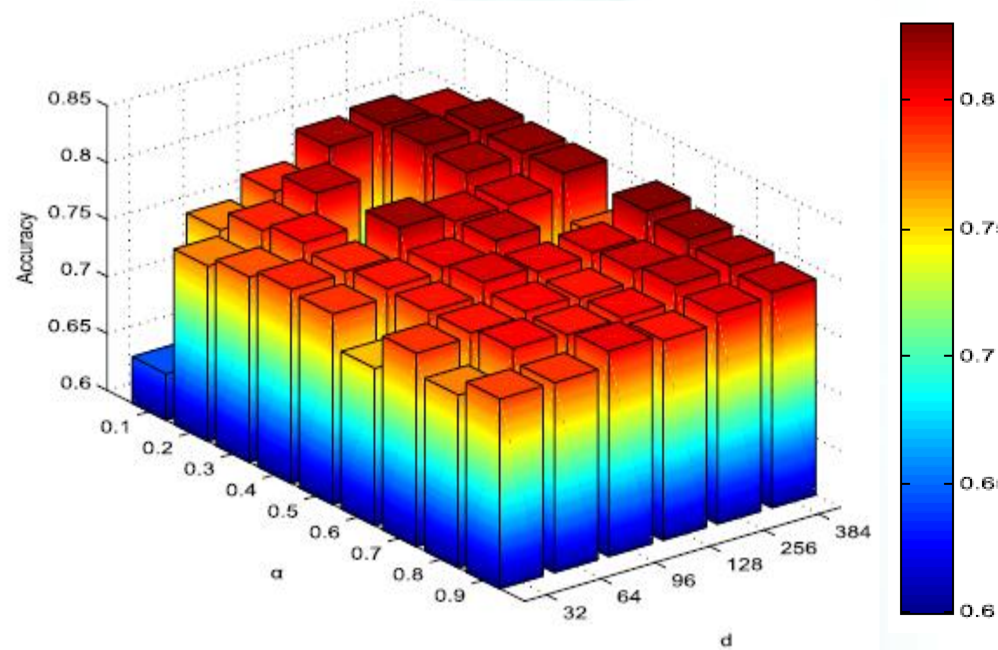
Learning Person–Person Interaction

- Parameter Evaluations:

The impacts of *α* and *d* on CAD



The impacts of *α* and *d* on Choi's Dataset



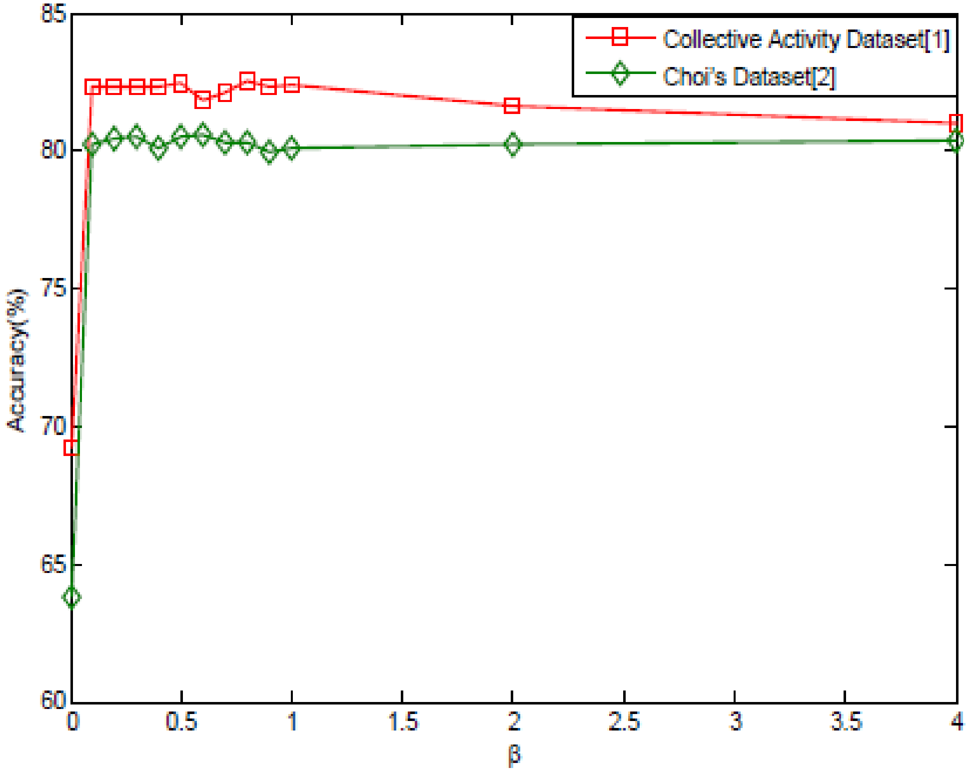
***α* varies from 0.1~0.9; *d* = {32,64,96,128,256,384}**

SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction

Effect of *logdet* regularization

The impact of *β* on both datasets



$$\min_{\overline{L}_0,\overline{L}_m} J(\overline{L}_0,\overline{L}_m) = \frac{1}{2}\sum_{k=0,m}\{\|\overline{L}_k\|_F^2 - \frac{\beta}{2}log\ det(\overline{L}_k'\overline{L}_k)\} + C\sum_{t=1}^{|T|} Loss(\overline{L}_0,\overline{L}_m)$$

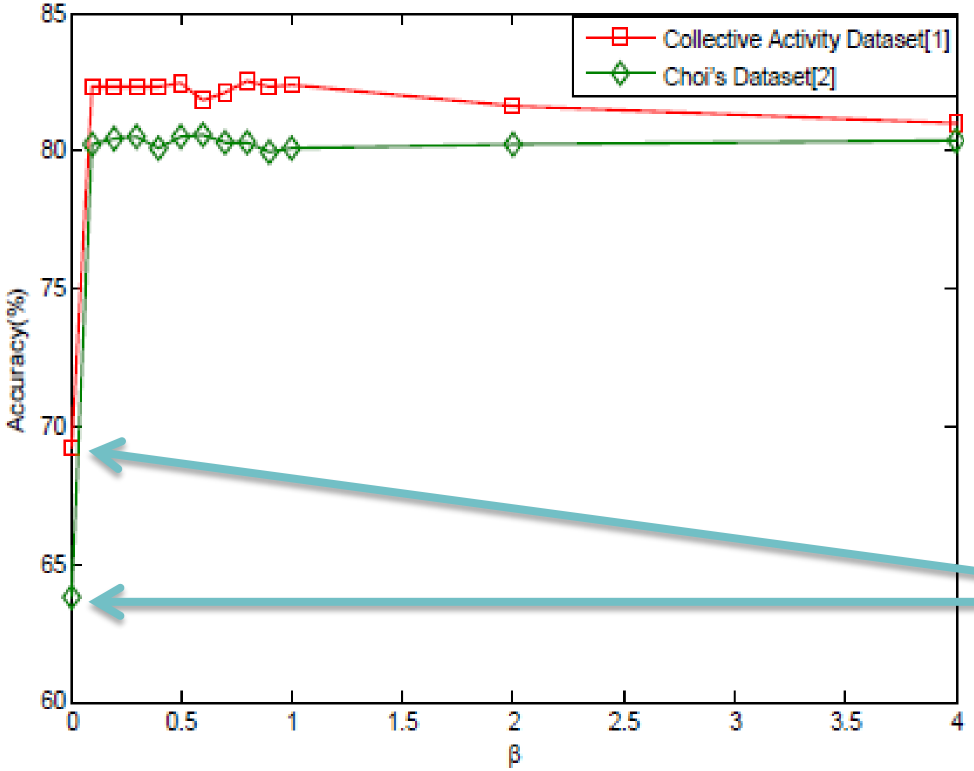
- **The performances fall obviously without –*logdet* regularization(*β* = *0*);**
- **The performances become stable when *β* = *0.3***

中山大學
SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction

Effect of *logdet* regularization

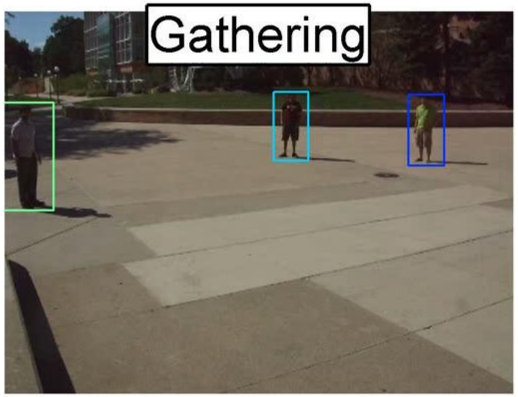
The impact of β on both datasets

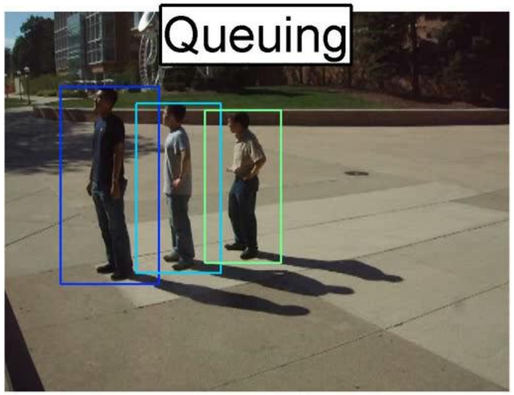


$$\min_{\overline{L}_0,\overline{L}_m} J(\overline{L}_0,\overline{L}_m) = \frac{1}{2}\sum_{k=0,m}\{\|\overline{L}_k\|_F^2 - \frac{\beta}{2} log\ det(\overline{L}_k'\overline{L}_k)\} + C\sum_{t=1}^{|T|} Loss(\overline{L}_0,\overline{L}_m)$$

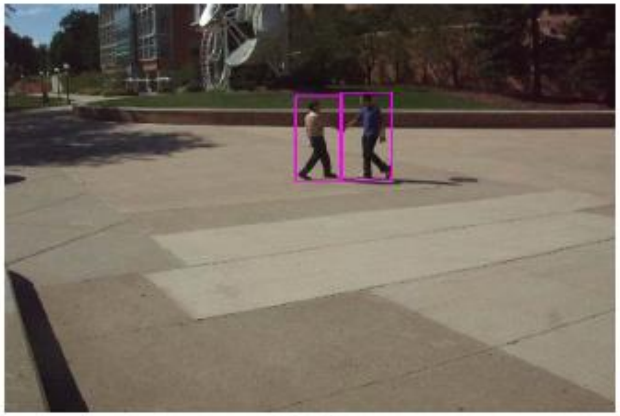
- ➢The performances **fall** obviously without *–logdet* regularization(***β = 0***);
- ➢The performances become **stable** when ***β = 0.3***

Learning Person–Person Interaction





Some Wrong Prediction

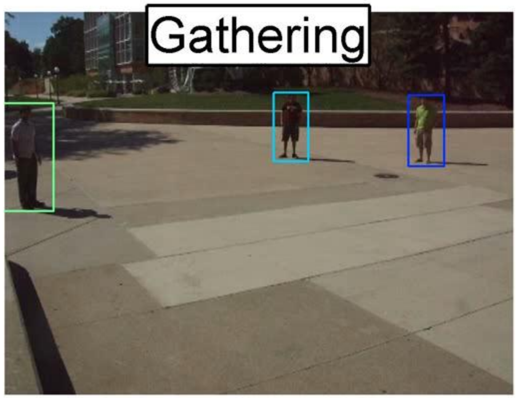


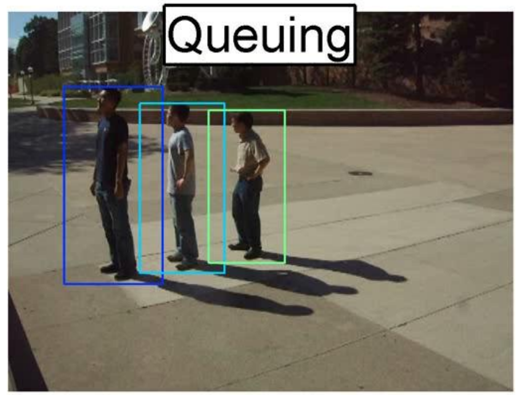
(a)

(b)

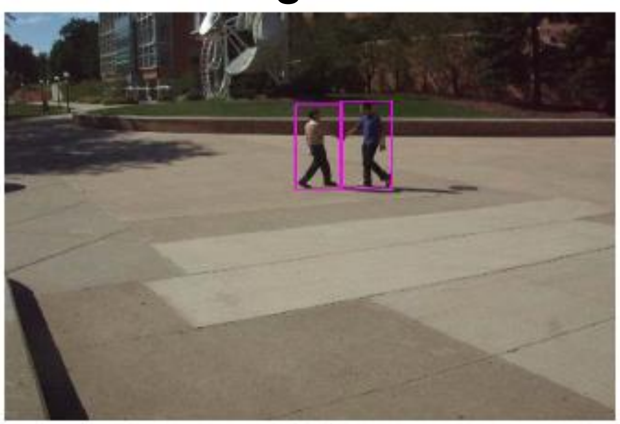
中山大學
SUN YAT-SEN UNIVERSITY

Learning Person–Person Interaction





Some Wrong Prediction


(a)


(b)

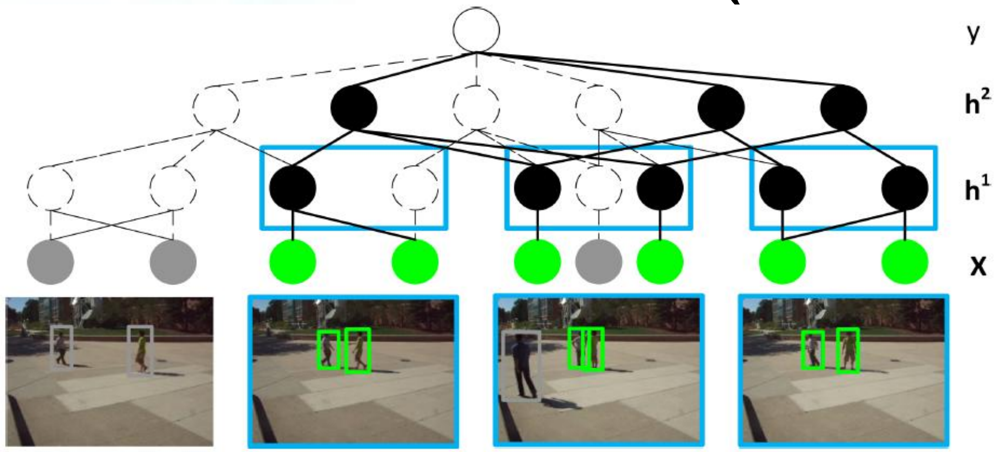
Predict: talking
GroundTruth: gathering

Predict: walking
GroundTruth: chasing

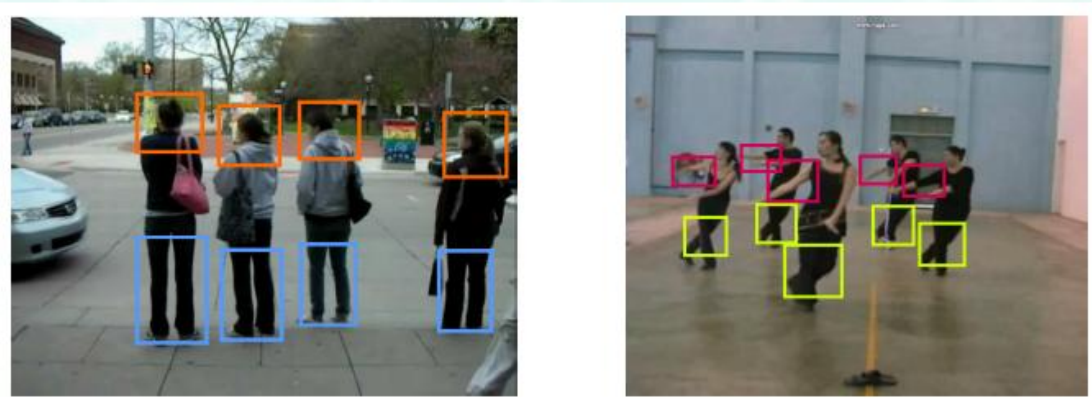
SUN YAT-SEN UNIVERSITY 中山大學

Other Learning Based Methods

- **Hierarchical Random Field (ECCV'2014)**



- **Part-based Learning (ECCV'2014)**
  - Learning Latent Constituents for Recognition of Group Activities in Video



Tracking Across Disjoint Views:

Person Re-identification

Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Towards Open-World Person Re-Identification by One-Shot Group-based Verification. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2015. (DOI: 10.1109/TPAMI.2015.2453984)

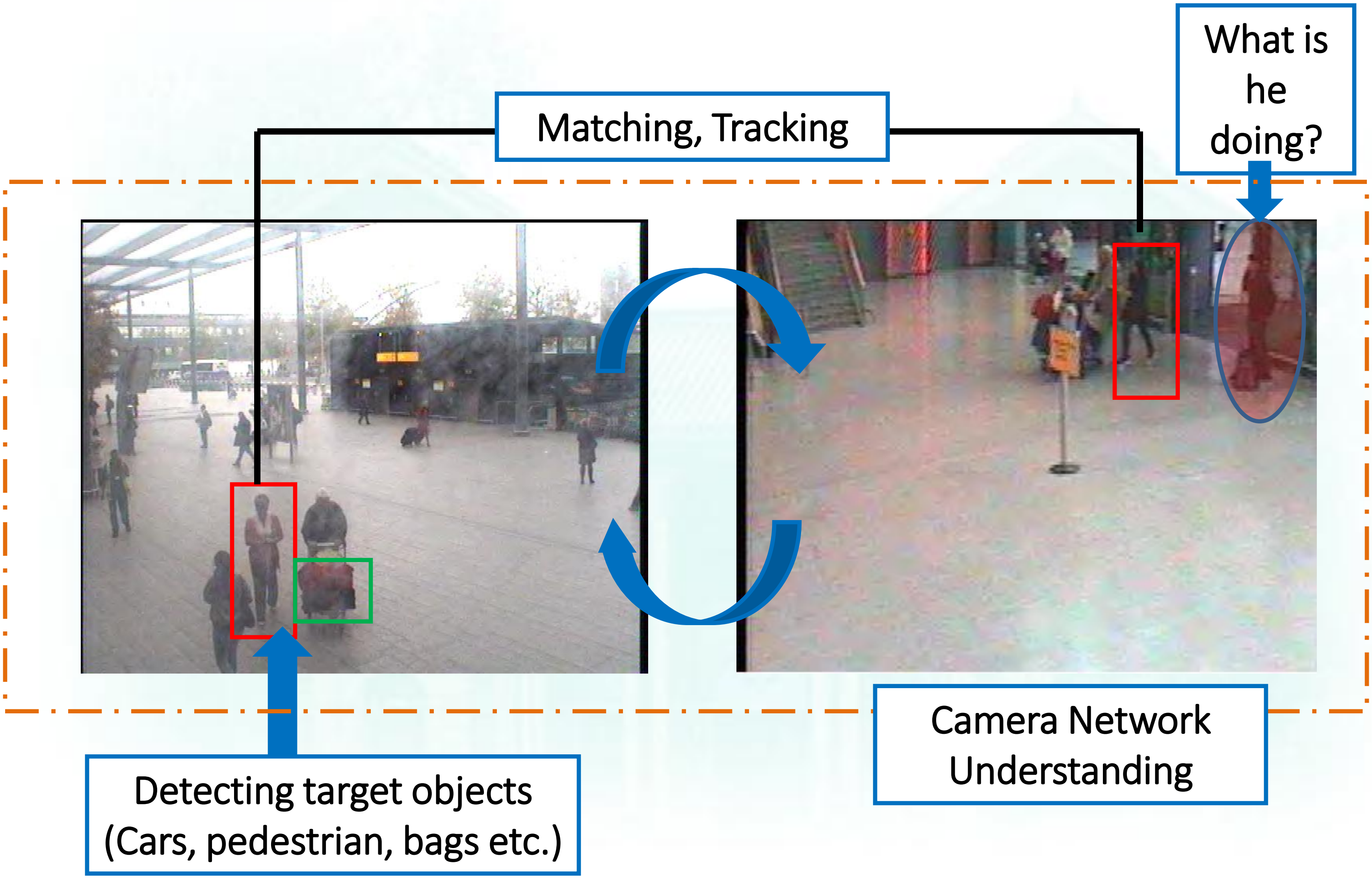
Wei-Shi Zheng et al., “Re-identification by Relative Distance Comparison”, IEEE Trans. on PAMI, 2013

Ying-Cong Chen et al.,”Mirror Representation for Modeling View-specific Transform in Person Re-identification”, IJCAI, 2015

Xiaojuan Wang et al., “Cross-scenario Transfer Person Re-identification”, IEEE Trans. on CSVT, to appear

中山大學 SUN YAT-SEN UNIVERSITY

Person Re-identification



Person Re-identification

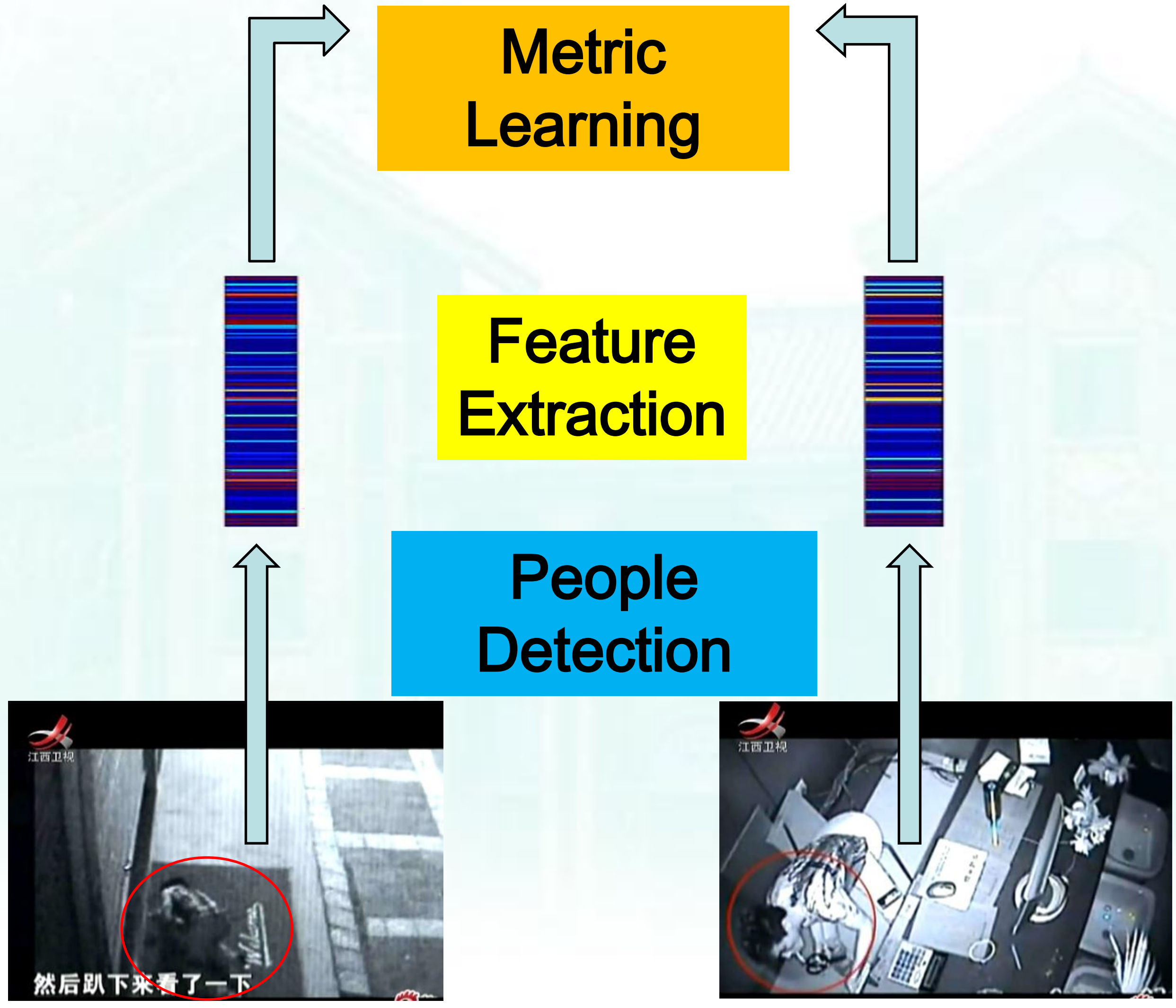
- A key component to track people across disjoint views

Suspect, Bomb in Boston, USA (2013)

Suspect, Terrorist Attack, Kunming, China (2014)

中山大學
SUN YAT-SEN UNIVERSITY

The Main Processing



Person Re-identification: Challenges

View **Lighting** **Occlusion** **Low Resolution**

……………….

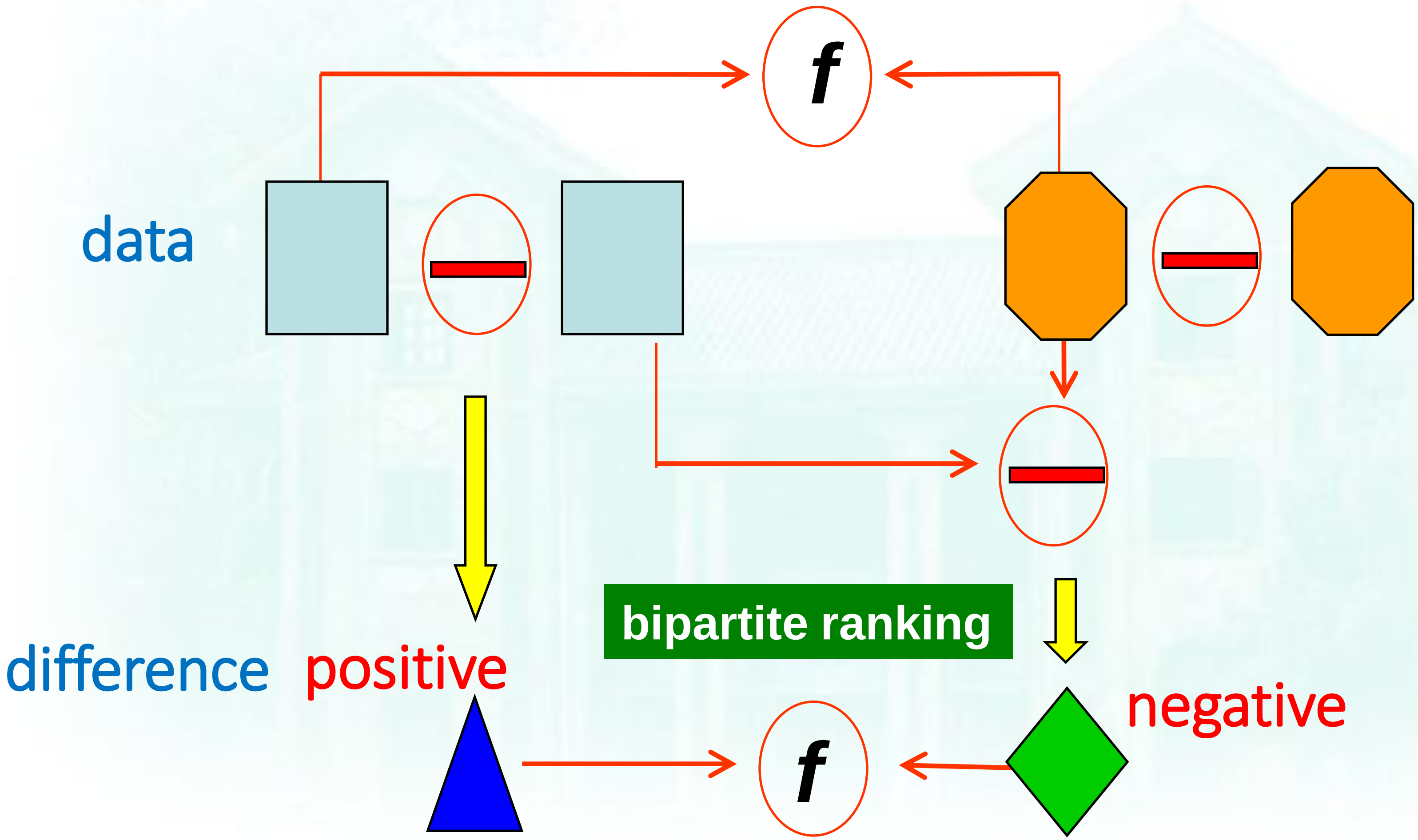
How to measure the differences between two person images

Wei-Shi Zheng et al., "Re-identification by Relative Distance Comparison", IEEE Trans. on PAMI, 2013

中山大學
SUN YAT-SEN UNIVERSITY

Triple based Learning: Bipartite Ranking

Main Idea



Triple based Learning: Bipartite Ranking

Idea from RankSVM

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 + \beta \sum_{i=1}^{|O|} \max\left(0, 1 - \mathbf{w}^T\left(\mathbf{x}_i^p - \mathbf{x}_i^n\right)\right)^2$$

$\mathbf{x}_i^p$ positive data difference

$\mathbf{x}_i^n$ related negative data difference

- Maximising the margin between difference souces of data difference
- Quantifying first-order feature vectors
- Sensitive to parameter

Triple based Learning: Bipartite Ranking

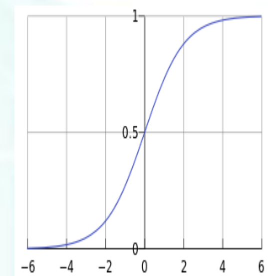
- A Relative Distance Comparison Model

OBJECTIVE

positive difference vector $\underline{f(\mathbf{x}_i^p) < f(\mathbf{x}_i^n)}$ negative difference vector

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{M} \mathbf{x}, \ \ \mathbf{M} \succeq 0$$

$$\mathbf{M} = \mathbf{A} \mathbf{\Lambda} \mathbf{A}^T = \mathbf{W} \mathbf{W}^T, \quad \mathbf{W} = \mathbf{A} \mathbf{\Lambda}^{\frac{1}{2}}$$



soft margin measure

$$\min_{\mathbf{W}} r(\mathbf{W}, \mathbb{O}), \ \ s.t. \ \ \mathbf{w}_i^T \mathbf{w}_j = 0, \ \forall i \neq j$$

$$r(\mathbf{W}, \mathbb{O}) = \sum\nolimits_{\mathbb{O}_i} \log(1 + \underline{\exp\{||\mathbf{W}^T \mathbf{x}_i^p||^2 - ||\mathbf{W}^T \mathbf{x}_i^n||^2\}})$$

- Reduce the sensitivity for comparison
  Enhance the performance (↑ i-LIDS, VIPeR)

Wei-Shi Zheng et al. Re-identification by Relative Distance Comparison. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI). 2013.

中山大學 SUN YAT-SEN UNIVERSITY

Triple based Learning: Bipartite Ranking

Learn the projection vectors each by each

$$\mathbf{w}_{\ell+1} = \arg\min_{\mathbf{w}} r_{\ell+1}(\mathbf{w}, \mathbb{O}^{\ell+1}),$$

where

$$r_{\ell+1}(\mathbf{w}, \mathbb{O}^{\ell+1}) = \sum_{\mathbb{O}_i^{\ell+1}} \log\left(1 + a_i^{\ell+1} \exp\left\{\left\|\mathbf{w}^T \mathbf{x}_i^{p,\ell+1}\right\|^2 - \left\|\mathbf{w}^T \mathbf{x}_i^{n,\ell+1}\right\|^2\right\}\right).$$

$$a_i^{\ell+1} = \exp\left\{\sum_{j=0}^{\ell} \|\mathbf{w}_j^T \mathbf{x}_i^{p,j}\|^2 - \|\mathbf{w}_j^T \mathbf{x}_i^{n,j}\|^2\right\}$$

$$\mathbf{x}_i^{s,\ell} = \mathbf{x}_i^{s,\ell-1} - \tilde{\mathbf{w}}_{\ell-1}\tilde{\mathbf{w}}_{\ell-1}^T \mathbf{x}_i^{s,\ell-1}, \quad s \in \{p, n\},\ i = 1, \ldots, |\mathbb{O}|,$$

$$\tilde{\mathbf{w}}_{\ell-1} = \mathbf{w}_{\ell-1}/\|\mathbf{w}_{\ell-1}\|$$

$\mathbf{x}_i^{s,0} = \mathbf{x}_i^s$, $s \in \{p, n\}$, and $\tilde{\mathbf{w}}_0 = \mathbf{0}$

中山大學 SUN YAT-SEN UNIVERSITY

Triple based Learning: Bipartite Ranking

- **Convergence**

Theorem 1. *The learned vectors* $\mathbf{w}_\ell$, $\ell = 1, \ldots, L$, *are orthogonal to each other.*

Theorem 2. $r(\mathbf{W}^{\ell+1}, \mathbb{O}) \leq r(\mathbf{W}^{\ell}, \mathbb{O})$, *where* $\mathbf{W}^{\ell} = (\mathbf{w}_1, \ldots, \mathbf{w}_\ell)$, $\ell \geq 1$. *That is, the algorithm iteratively decreases the objective function value.*

Triple based Learning: Bipartite Ranking

Entry-wise Absolute Difference Vector

$$\mathbf{x} = d(\mathbf{z}, \mathbf{z}') = \left|\mathbf{z} - \mathbf{z}'\right|, \ \mathbf{x}(k) = \left|\mathbf{z}(k) - \mathbf{z}'(k)\right|$$

$$f(|\mathbf{x}_{ij}|) = |\mathbf{z}_i - \mathbf{z}_j|^T \mathbf{M} |\mathbf{z}_i - \mathbf{z}_j| = ||\mathbf{W}^T |\mathbf{x}_{ij}| \ ||^2$$

$$\left|\left||\mathbf{x}_{ij}| - |\mathbf{x}_{ij'}|\right|\right| \leq \left|\left|\mathbf{x}_{ij} - \mathbf{x}_{ij'}\right|\right|$$

$$upper(\left|\left|\mathbf{W}^T(|\mathbf{x}_{ij}| - |\mathbf{x}_{ij'}|)\right|\right|) \leq upper(\left|\left|\mathbf{W}^T(\mathbf{x}_{ij} - \mathbf{x}_{ij'})\right|\right|)$$

Relative Distance Learning can be more robust in the absolute distance space

Triple based Learning: Bipartite Ranking

Ensemble Metric Learning

Ensemble RDC: Motivation

- **RDC: Large space complexity**

$$O(q\cdot((\frac{1}{L}-\frac{1}{L^2})\cdot N^3+(\frac{1}{L}-1)\cdot N^2)) \Rightarrow O(q\cdot((\frac{b^2}{L}-\frac{b}{L^2})\cdot N^3+(\frac{b}{L}-b^2)\cdot N^2))$$

- **RDC: Trapped in locally optimal solution**

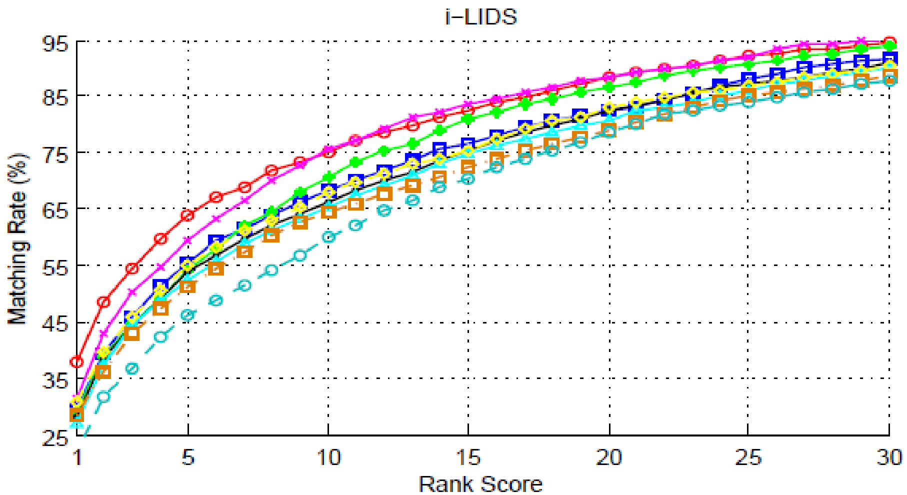
Ensemble RDC: Modelling

- **Randomly dividing the set into small groups**
- **Learning a set of weak RDC models**
- **Boosting them**

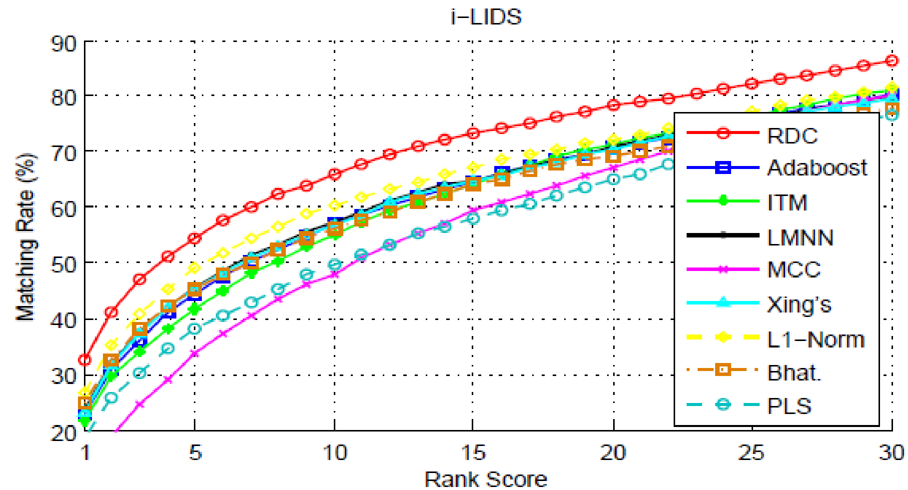
Wei-Shi Zheng et al., “Re-identification by Relative Distance Comparison” IEEE Trans. on PAMI, 2013

Triple based Learning: Bipartite Ranking

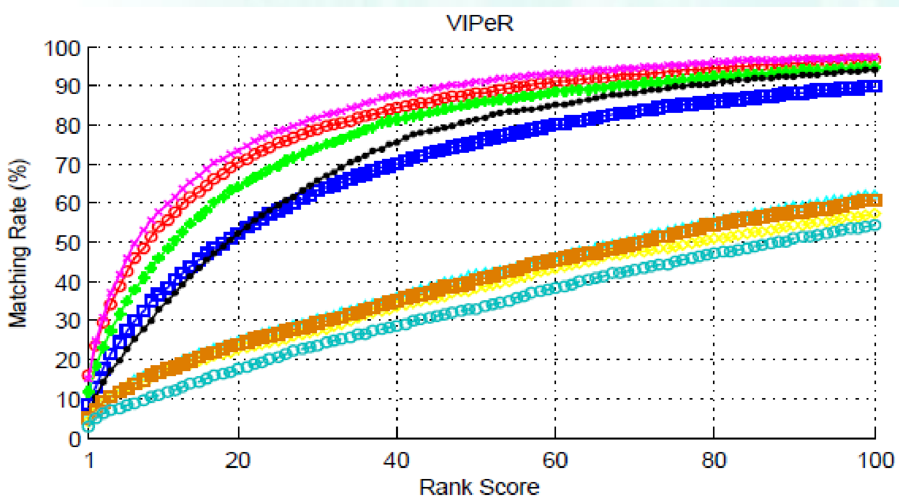
Re-identification (i-LIDS&VIPeR)



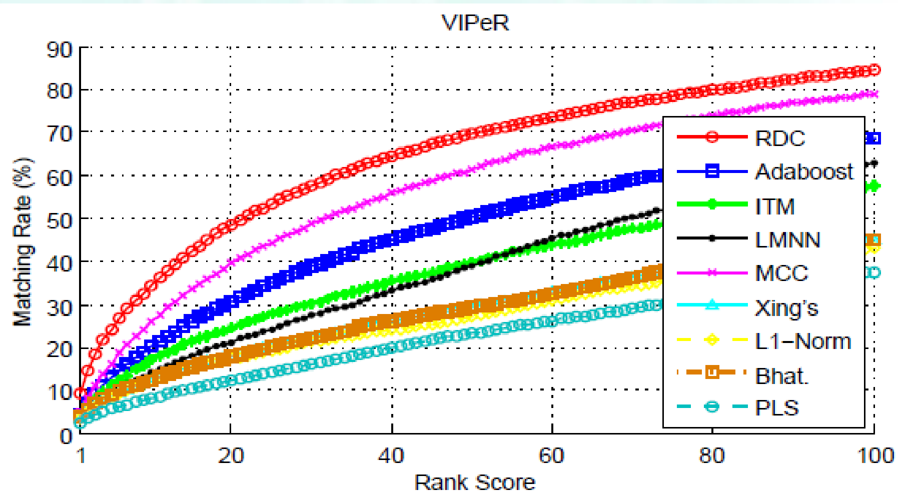
(a) $p = 50$



(b) $p = 80$



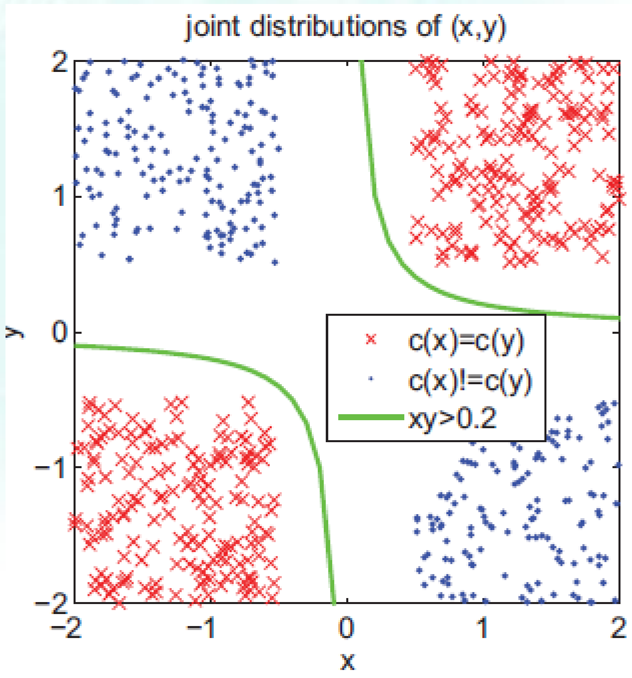
(a) $p = 316$



(b) $p = 532$

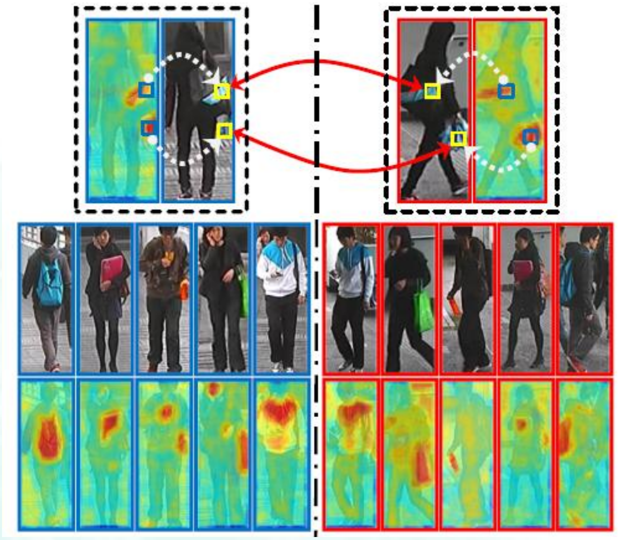
Other Distance Model for RE-ID

- **Subspace: LFDA (CVPR'12)**
  - Local Fisher Discriminant Analysis for Pedestrian Re-identification
- **KISSME (CVPR'12)**
  - Large Scale Metric Learning from Equivalence Constraints
- **Local Boundary (CVPR'13)**
  - Learning Locally-Adaptive Decision Functions for Person Verification

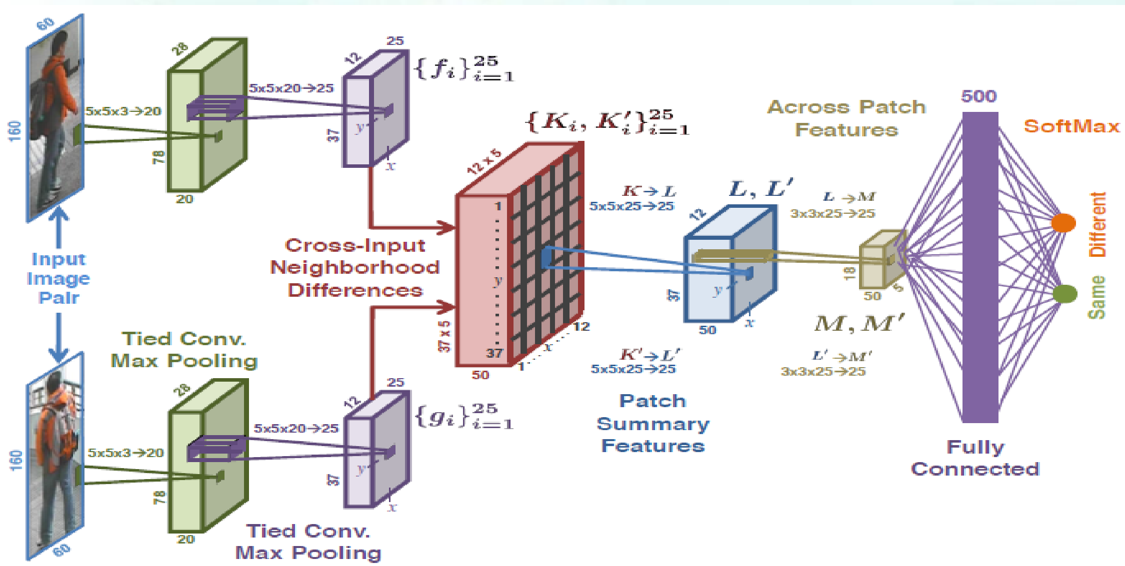


Other Distance Model for RE-ID

- **Saliency Matching (CVPR'13, ICCV'13)**
  - Unsupervised Salience Learning for Person Re-identification



- **Deep Metric (CVPR'15)**
  - An Improved Deep Learning Architecture for Person Re-Identification

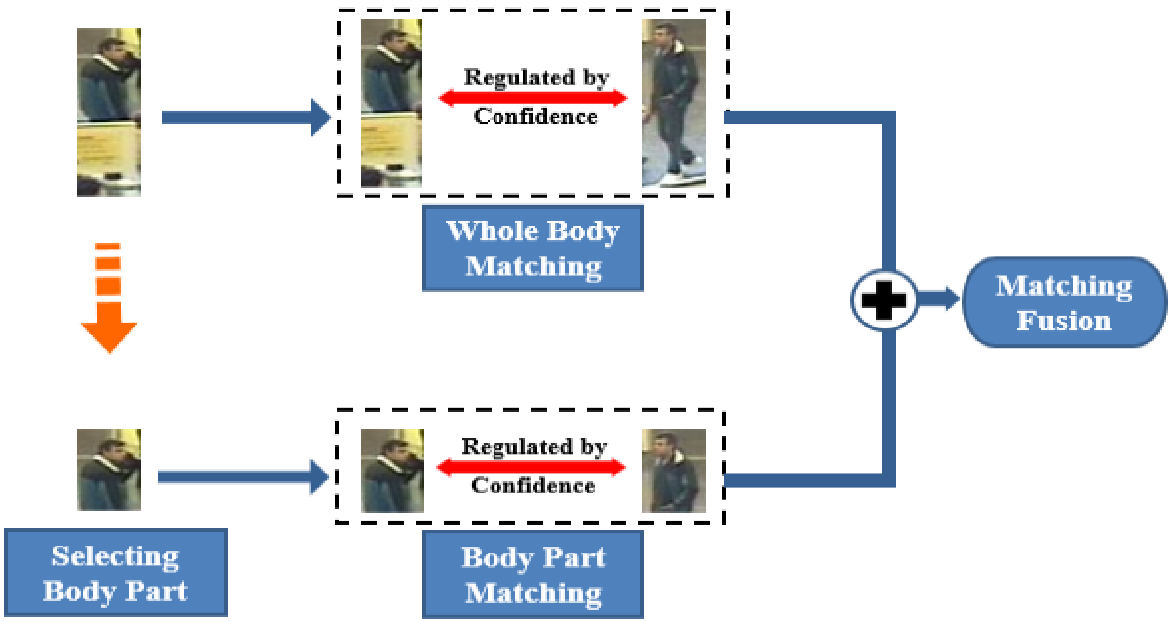


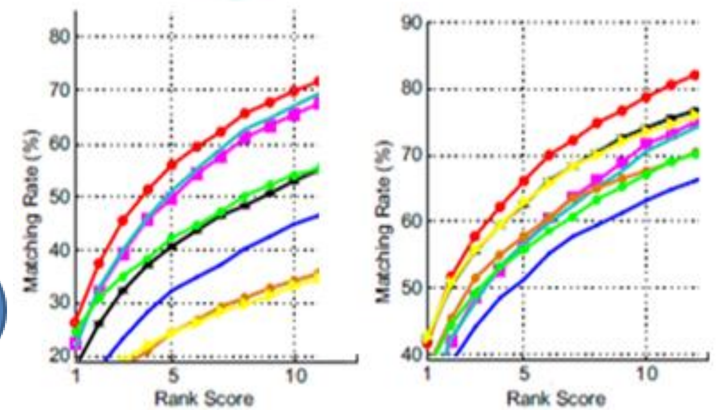
SUN YAT-SEN UNIVERSITY
中山大學

Other Distance Model for RE-ID

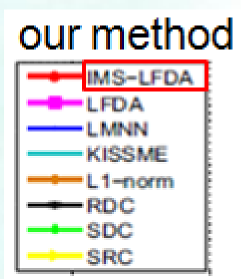
The Integrated Matching Scheme (IMS):(ISBA 2015)

"Towards More Reliable Matching for Person Re-identification"

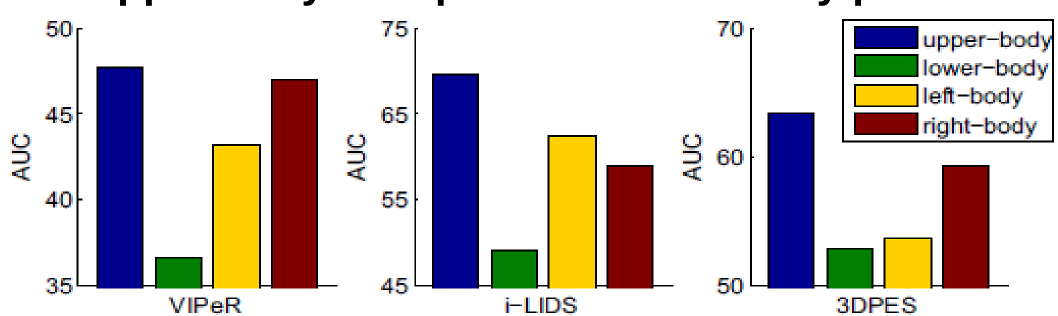




(a) VIPeR (b) i-LIDS



The upper-body is superior to other body parts



中山大學 SUN YAT-SEN UNIVERSITY

Other Distance Model for RE-ID

- **Local Relative Distance Comparison** **(Zheng et al., TPAMI'15)**
  - Towards Open-World Person Re-Identification by One-Shot Group-based Verification

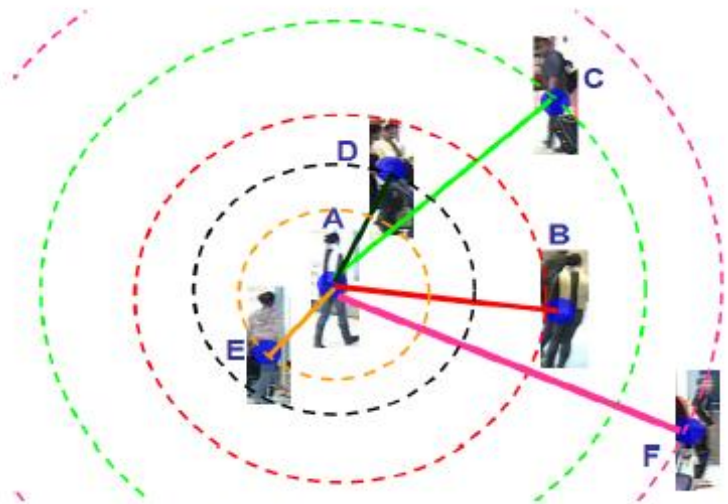


Fig. 3. Illustration of our local relative comparison. Among the six images, A and B belong to the same person whilst the other four are of four other people. See text for more details.

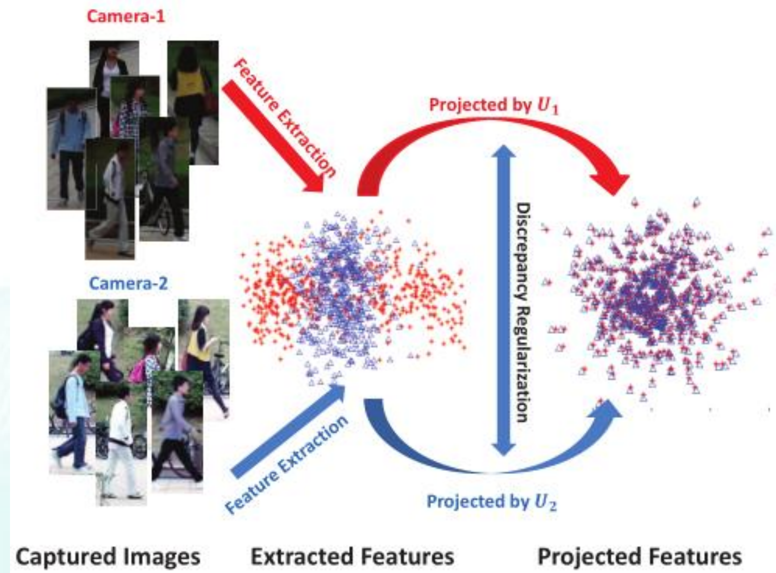
	t-LRDC	t-RDC	t-RankSVM	RDC	RankSVM
Sensitive to PCA	×	√	√	√	√
Max Memory Cost	∼0.7 G	∼16.9G	∼16.9G	∼16.1G	∼16.1G

TABLE III
COST COMPARISON: RELATIVE COMPARISON LEARNING ON VIPER

中山大學
SUN YAT-SEN UNIVERSITY

What is Wrong with Current Metrics

- The view label Information is not explicitly used



- The distributions of person images across camera views are different

- Existing metrics are learned for each scenario and cannot generalize very well

When View Labels are available, how to model the view transform more accurately

Ying-Cong Chen et al.,"Mirror Representation for Modeling View-specific Transform in Person Re-identification", IJCAI, 2015

Mirror Representation

Usefulness of View Label Information

- Illumination, viewpoint or camera features vary across views, and distributions of each view are different.
- View-Specific Mappings can be adopted to correct different distributions of views.

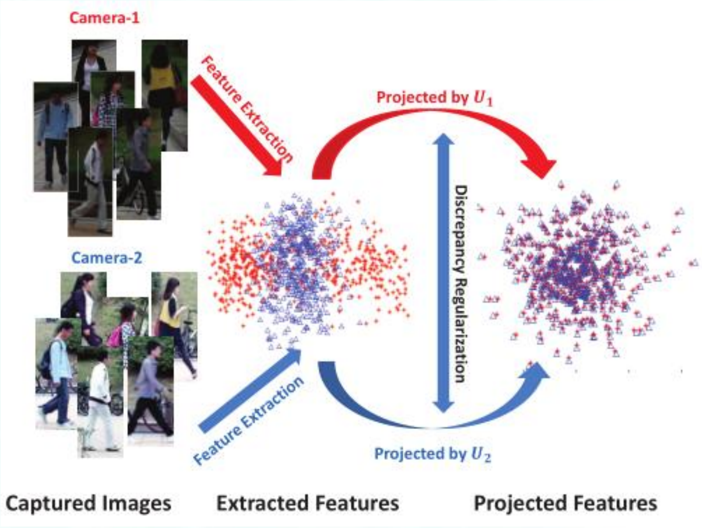
中山大學 SUN YAT-SEN UNIVERSITY

Mirror Representation

- Augmenting original feature with zeros
  - $X^a \rightarrow [X^a, 0]$
  - $X^b \rightarrow [0, X^b]$

Zero-Padding Augmentation

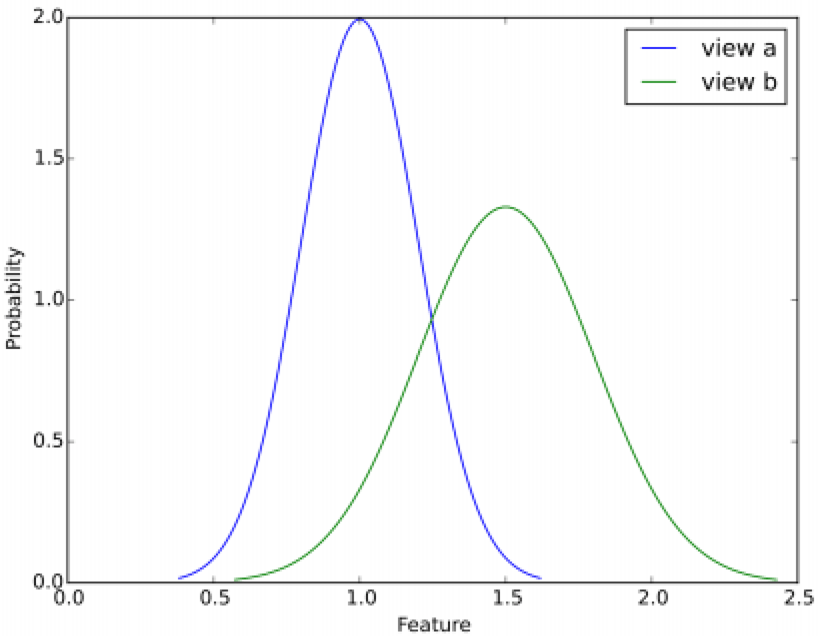
- Learning projection bases with augmented features
  - $\min_{U} f(U^T X) \rightarrow U = [U_1; U_2]$
- View-specific projection
  - $f_a(X^a) = U^T[X^a, 0] = U_1^T X^a$
  - $f_b(X^b) = U^T[0, X^b] = U_2^T X^b$



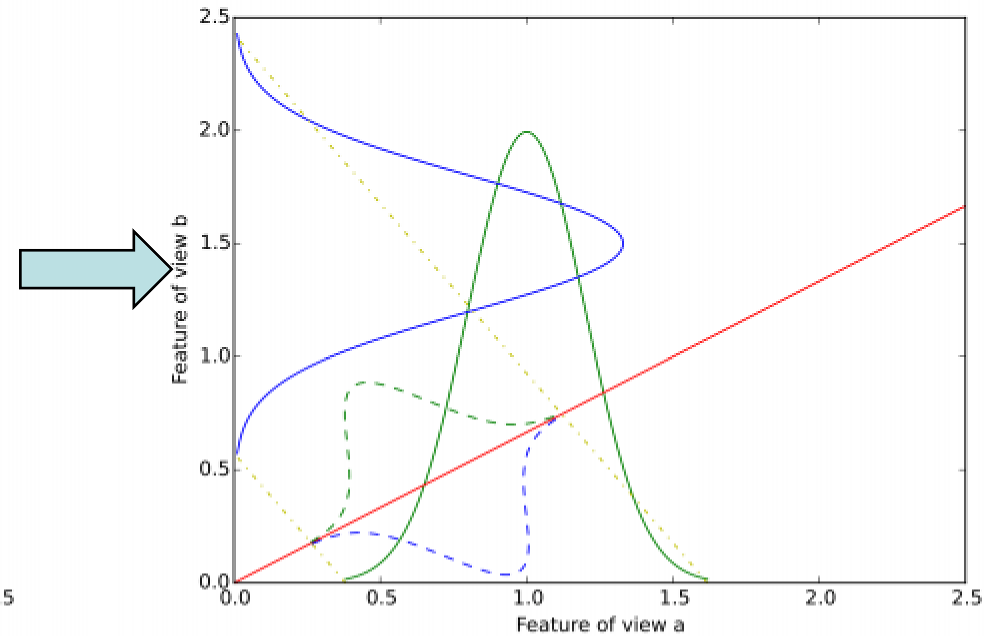
SUN YAT-SEN UNIVERSITY

Mirror Representation

Illustration of Zero-Padding Augmentation



(a) Mismatched Distribution



(b) Zero-padding

SUN YAT-SEN UNIVERSITY

Limitation of Zero-Padding

- In person re-identification, features of different camera views are often related, thus U_1 and U_2 should also be related.

- By using zero-padding, one loses directly control of the relation between U_1 and U_2.

中山大學
SUN YAT-SEN UNIVERSITY

Reformulation of Zero-Padding

- $X^a_{aug} = [I, 0]X^a, X^b_{aug} = [0, I]X^b$
  - $f_a = [U_1 + 0U_2]X^a = U_1X^a$
  - $f_b = [0U_1 + U_2]X^b = U_2X^b$

generalize

- $X^a_{aug} = [R, M]X^a, X^b_{aug} = [R, M]X^b$
  - $f_a = [RU_1 + MU_2]X^a$
  - $f_b = [MU_1 + RU_2]X^b$

control the discrepancy of $\boldsymbol{f}_a$ and $\boldsymbol{f}_b$

SUN YAT-SEN UNIVERSITY

A Feature-Level Discrepancy Modeling

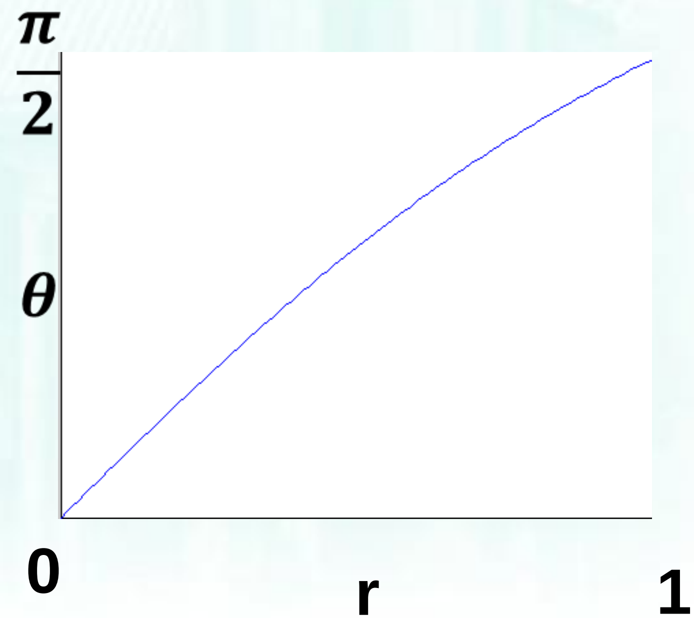
- $R = \frac{1-r}{z} I$
- $M = \frac{1+r}{z} I$

$$\boldsymbol{Z = \sqrt{(1-r)^2 + (1+r)^2}}$$

$$\boldsymbol{0 < r < 1}$$

Discrepancy of $[R, M]$ and $[M, R]$ is measured by the principle angles:

$$\theta = arccos(\frac{2}{r^2+1} - 1)$$



A Transformation-Level Discrepancy Modeling

We directly impose $||U_1 - U_2||^2$ as a regularization term, which can be easily integrated into ridge regularization

$$||U_1 - U_2||^2 = U^T B U, U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}, B = \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}$$

$$U^T B U + \lambda U^T U = U^T C U \,, C = \begin{bmatrix} I & -\beta I \\ -\beta I & I \end{bmatrix}$$

A Transformation-Level Discrepancy Modeling

$$\min_{\boldsymbol{U}} \quad f(\boldsymbol{U}^T \boldsymbol{X}_{aug}) + \lambda \boldsymbol{U}^T \boldsymbol{C} \boldsymbol{U}$$
$$s.t. \quad g_i(\boldsymbol{U}^T \boldsymbol{X}_{aug}) \quad i = 1, 2, \cdots, c$$

$$\boldsymbol{U} = \boldsymbol{P}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{H}, \boldsymbol{U}^T\boldsymbol{C}\boldsymbol{U} = \boldsymbol{H}^T\boldsymbol{H}$$

Mirror Representation

$$\min_{\boldsymbol{H}} \quad f(\boldsymbol{H}^T \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{P}^T \boldsymbol{X}_{aug}) + \lambda \boldsymbol{H}^T \boldsymbol{H}$$
$$s.t. \quad g_i(\boldsymbol{H}^T \boldsymbol{\Lambda}^{-\frac{1}{2}} \boldsymbol{P}^T \boldsymbol{X}_{aug}) \quad i = 1, 2, \cdots, c$$

can be solved by traditional metric learning (with ridge regularization)

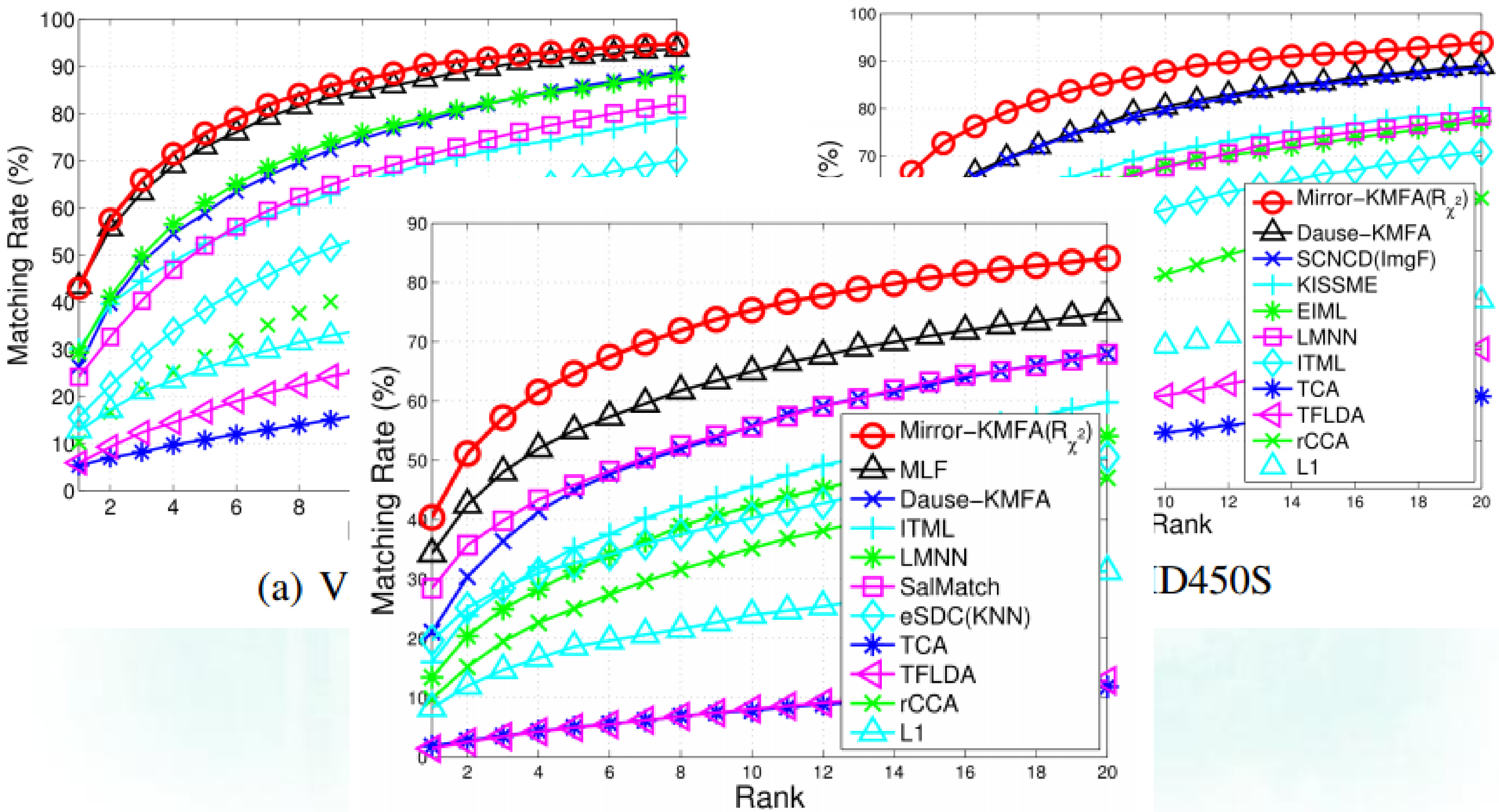
Effectiveness of Mirror Representation

	Representation	Mirror Representation				Original Feature				Zero-Padding			
	Rank	1	5	10	20	1	5	10	20	1	5	10	20
VIPeR	$KMFA(R_{\chi^2})$	**42.97**	**75.82**	**87.28**	**94.84**	37.37	71.23	84.72	93.45	33.67	67.66	82.31	91.87
	$KMFA(\chi^2)$	**39.62**	**71.36**	**84.18**	**93.23**	35.57	67.34	81.14	91.74	30.28	63.54	77.88	89.15
	$KPCCA(R_{\chi^2})$	**32.88**	**67.91**	**82.03**	**91.77**	29.05	62.94	78.26	89.68	21.84	52.44	67.37	79.40
	$KPCCA(\chi^2)$	**29.37**	**64.11**	**78.96**	**90.63**	25.63	59.78	76.27	87.78	18.77	51.17	66.77	82.31
	MFA	**33.48**	**63.10**	**75.60**	**86.55**	30.76	59.43	73.61	85.41	21.87	52.06	66.58	81.39
	$PCCA$	**27.56**	**60.57**	**75.66**	**87.37**	25.47	56.96	71.08	85.25	22.53	55.60	71.30	86.36
CUHK01	$KMFA(R_{\chi^2})$	**40.40**	**64.63**	**75.34**	**84.08**	34.98	60.16	71.27	81.50	33.53	59.00	70.20	80.24
	$KMFA(\chi^2)$	**37.31**	**61.11**	**71.36**	**81.25**	32.34	56.14	67.52	77.73	31.35	56.71	67.56	78.18
	$KPCCA(R_{\chi^2})$	**29.57**	**56.53**	**69.21**	**79.40**	25.30	52.40	64.61	76.76	17.84	41.53	53.95	67.83
	$KPCCA(\chi^2)$	**26.69**	**54.40**	**66.88**	**77.87**	22.79	48.65	62.10	74.06	17.84	41.53	53.95	67.83
	MFA	**25.47**	**48.38**	**58.86**	**69.19**	20.71	41.51	52.42	63.21	14.13	33.12	43.10	54.07
	$PCCA$	**19.74**	**40.96**	**52.44**	**65.00**	16.79	38.13	49.29	61.35	3.89	9.02	12.32	16.28
PRID450S	$KMFA(R_{\chi^2})$	**55.42**	**79.29**	**87.82**	**93.87**	52.76	77.56	84.71	91.56	46.18	74.13	84.31	92.40
	$KMFA(\chi^2)$	**53.42**	**77.29**	**85.82**	**91.51**	51.02	75.29	82.80	89.47	41.82	71.29	81.82	90.04
	$KPCCA(R_{\chi^2})$	**41.51**	**71.51**	**81.42**	**91.24**	40.09	68.76	79.73	90.13	33.60	65.78	78.18	88.00
	$KPCCA(\chi^2)$	**39.82**	**68.31**	**80.22**	**89.82**	37.60	66.18	78.49	88.62	28.27	58.71	72.40	85.60
	MFA	**40.58**	**77.56**	**67.47**	**86.58**	38.22	63.42	73.87	83.64	21.16	50.00	62.98	76.84
	$PCCA$	**38.40**	**68.40**	**79.51**	**88.31**	36.76	65.69	76.22	85.16	32.80	64.62	76.98	87.38

The best is marked red, and the second best is marked blue.

中山大學 SUN YAT-SEN UNIVERSITY

Performance

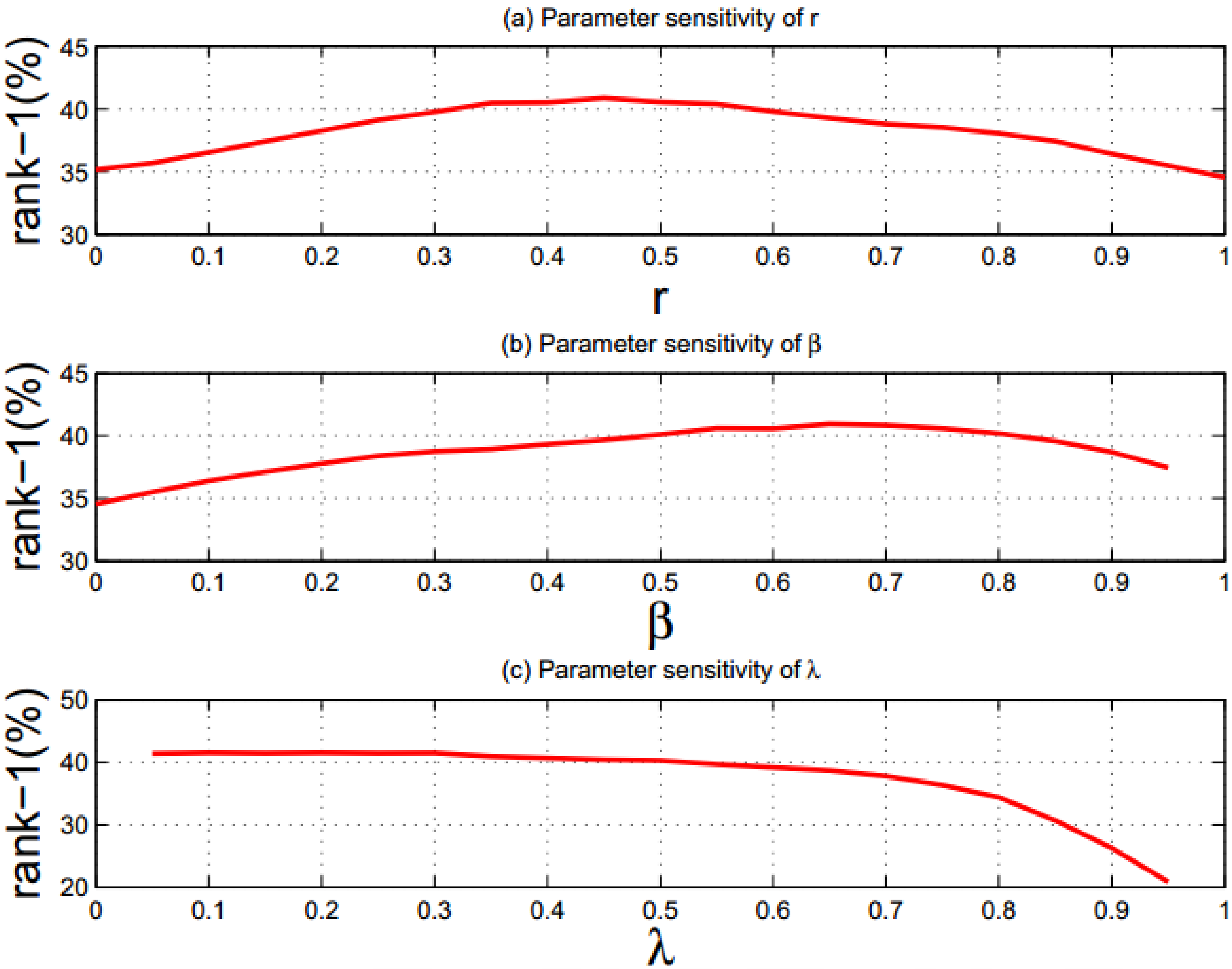


(a) V

D450S

(c) CUHK01

Parameter Analysis



中山大學
SUN YAT-SEN UNIVERSITY

Person Re-identification

target scenario

Labeling images is costly and even prohibitive in some scenarios

Xiaojuan Wang et al., “Cross-scenario Transfer Person Re-identification”, IEEE Trans. on CSVT, to appear

Person Re-identification

Person Re-identification

source scenario

Is it possible to use collected images in other scenarios to boost the learning in the target scenario?

Cross-scenario Transfer Person Re-identification

Xiaojuan Wang et al., “Cross-scenario Transfer Person Re-identification” IEEE Trans. on CSVT, to appear

Framework

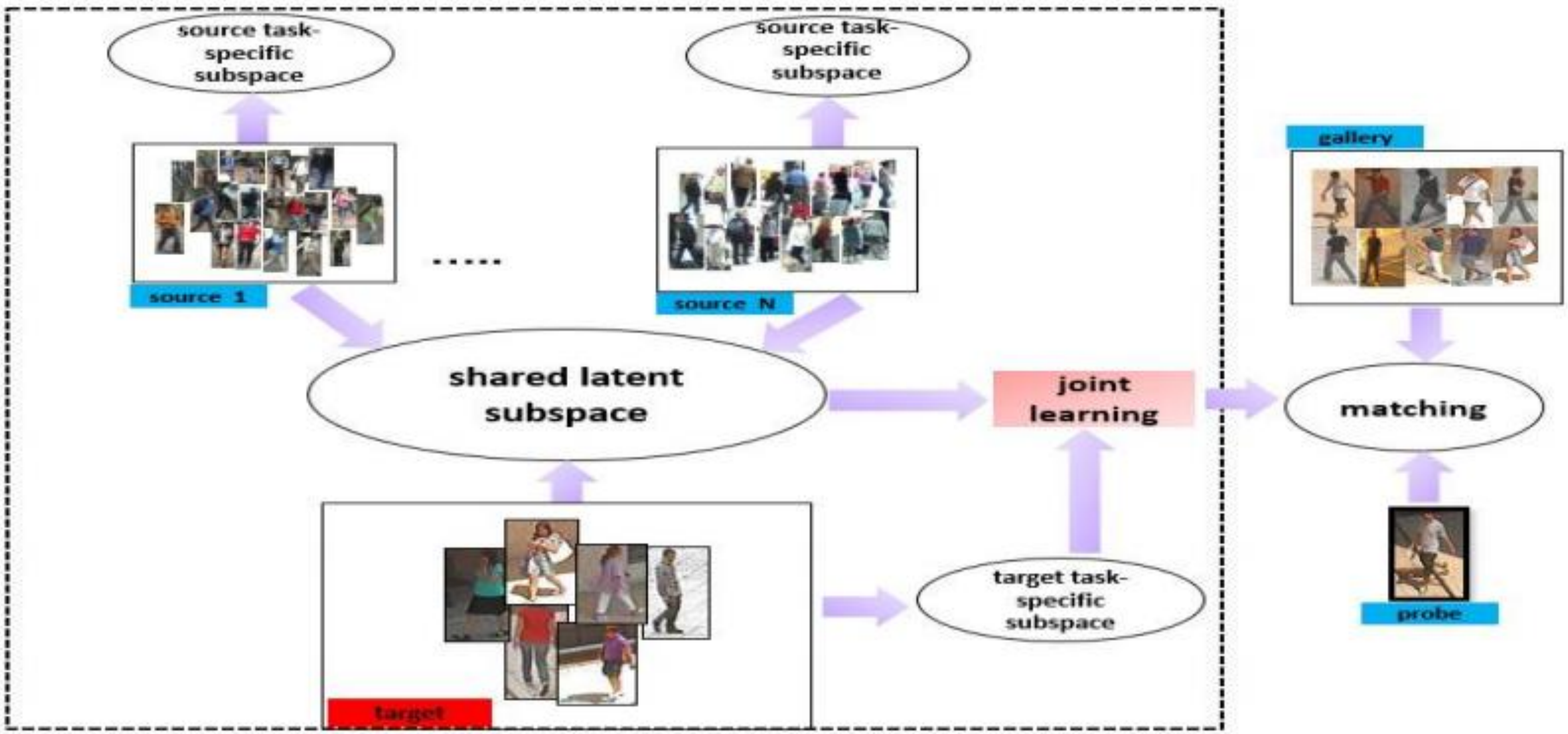


Figure 1: Asymmetric Multi-task Person Re-identification system: diagrams inside the large dash rectangle box indicate the principle of training.

Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets

Modeling

shared latent subspace: $\mathbf{W}_0 \in \mathrm{R}^{d \times r}$

source task-specific subspace: $\mathbf{W}_s \in \mathrm{R}^{d \times r}$

target task-specific subspace: $\mathbf{W}_t \in \mathrm{R}^{d \times r}$

The projection of a target sample $\mathbf{x}_t$ $\mathbf{z}_t = ((1-\beta)\mathbf{W}_0 + \beta \mathbf{W}_t)' \mathbf{x}_t$

joint learning

The projection of a source sample $\mathbf{x}_s$ $\mathbf{z}_s = ((1-\beta)\mathbf{W}_0 + \beta \mathbf{W}_s)' \mathbf{x}_s$

$$0 \leq \beta \leq 1$$

中山大學 SUN YAT-SEN UNIVERSITY

Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets

Modeling

To maximize local inter-class variances and meanwhile to minimize the local intra-class variances in both tasks

$0 \leq \gamma \leq 1$

$$\max_{\mathbf{W}_1,\mathbf{W}_2} (1-\gamma)\frac{\mathrm{tr}(\mathbf{W}_1'\mathbf{S}_b^s\mathbf{W}_1)}{\mathrm{tr}(\mathbf{W}_1'\mathbf{S}_w^s\mathbf{W}_1)} + \gamma\frac{\mathrm{tr}(\mathbf{W}_2'\mathbf{S}_b^t\mathbf{W}_2)}{\mathrm{tr}(\mathbf{W}_2'\mathbf{S}_w^t\mathbf{W}_2)}, \quad (1)$$

$$\mathbf{W}_1 = (1-\beta)\mathbf{W}_0 + \beta\mathbf{W}_s)$$
$$\mathbf{W}_2 = (1-\beta)\mathbf{W}_0 + \beta\mathbf{W}_t)$$

non-convex

relaxation

$$\max_{\mathbf{W}_1,\mathbf{W}_2} \frac{\mathrm{tr}((1-\gamma)\mathbf{W}_1'\mathbf{S}_b^s\mathbf{W}_1 + \gamma\mathbf{W}_2'\mathbf{S}_b^t\mathbf{W}_2)}{\mathrm{tr}((1-\gamma)\mathbf{W}_1'\mathbf{S}_w^s\mathbf{W}_1 + \gamma\mathbf{W}_2'\mathbf{S}_w^t\mathbf{W}_2)}, \quad (2)$$

Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets

Insight

$$
\begin{aligned}
&\mathrm{tr}(\mathbf{W}_1'\mathbf{S}_b^s\mathbf{W}_1) \\
&= \frac{1}{2}\sum_{i,j=1}^{n}\overline{\mathbf{A}}_{i,j}^{b}\sum_{k=1}^{r}\mathbf{W}_1(:,k)'(\mathbf{x}_i^s-\mathbf{x}_j^s)(\mathbf{x}_i^s-\mathbf{x}_j^s)'\mathbf{W}_1(:,k) \\
&= \frac{1}{2}\sum_{i,j=1}^{n}\overline{\mathbf{A}}_{i,j}^{b}\sum_{k=1}^{r}\left[(1-\beta)\mathbf{W}_0(:,k)'(\mathbf{x}_i^s-\mathbf{x}_j^s)+\beta\mathbf{W}_s(:,k)'(\mathbf{x}_i^s-\mathbf{x}_j^s)\right]^2
\end{aligned}
$$

adding those measures together gives us a stronger cue on overall discriminativeness

Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets

Optimization

$$\mathbf{W} = [\mathbf{W}_0; \mathbf{W}_s; \mathbf{W}_t] \in \mathbb{R}^{3d \times r}$$

$$\mathbf{\Theta}_s = [(1-\beta)\mathbf{I}_d, \beta\mathbf{I}_d, \mathbf{O}_{d\times d}] \in \mathbb{R}^{d\times 3d} \quad \mathbf{\Theta}_t = [(1-\beta)\mathbf{I}_d, \mathbf{O}_{d\times d}, \beta\mathbf{I}_d] \in \mathbb{R}^{d\times 3d}$$

$$\mathbf{A} = (1-\gamma)(\mathbf{\Theta}_s^{'}\mathbf{S}_b^s\mathbf{\Theta}_s) + \gamma(\mathbf{\Theta}_t^{'}\mathbf{S}_b^t\mathbf{\Theta}_t) \tag{3a}$$

$$\mathbf{B} = (1-\gamma)(\mathbf{\Theta}_s^{'}\mathbf{S}_w^s\mathbf{\Theta}_s) + \gamma(\mathbf{\Theta}_t^{'}\mathbf{S}_w^t\mathbf{\Theta}_t) \tag{3b}$$

Eq.(2) is equal to

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{\mathrm{tr}(\mathbf{W}'\mathbf{A}\mathbf{W})}{\mathrm{tr}(\mathbf{W}'\mathbf{B}\mathbf{W})} \tag{4}$$

$$\mathbf{A}\mathbf{W} = \lambda\mathbf{B}\mathbf{W}$$

generalized eignenvalue problem, global solution guaranteed

Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets

task-specific projection for each source dataset: $\mathbf{W}_s^i$

by redefining: $\mathbf{W} = [\mathbf{W}_0; \mathbf{W}_s^1; \ldots; \mathbf{W}_s^i; \ldots; \mathbf{W}_s^m; \mathbf{W}_t] \in \mathbb{R}^{(m+2)d \times r}$

$$\mathbf{\Theta}_s^i = [(1-\beta)\mathbf{I}_d, \ldots, \beta \mathbf{I}_d, \ldots, \mathbf{O}_{d\times d}] \in \mathbb{R}^{d\times(m+2)d}$$

$$\mathbf{\Theta}_t = [(1-\beta)\mathbf{I}_d, \mathbf{O}_{d\times d}, \ldots, \mathbf{O}_{d\times d}, \beta \mathbf{I}_d] \in \mathbb{R}^{d\times(m+2)d}$$

$$\mathbf{A} = (1-\gamma)(\frac{1}{m}\sum_{i=1}^{m}(\mathbf{\Theta}_s^i)'\mathbf{S}_b^{s,i}\mathbf{\Theta}_s^i) + \gamma(\mathbf{\Theta}_t'\mathbf{S}_b^t\mathbf{\Theta}_t) \tag{5a}$$

$$\mathbf{B} = (1-\gamma)(\frac{1}{m}\sum_{i=1}^{m}(\mathbf{\Theta}_s^i)'\mathbf{S}_w^{s,i}\mathbf{\Theta}_s^i) + \gamma(\mathbf{\Theta}_t'\mathbf{S}_w^t\mathbf{\Theta}_t). \tag{5b}$$

solution could be obtained by Eq. (4)

Constrained Asymmetric Multi-task Component Analysis

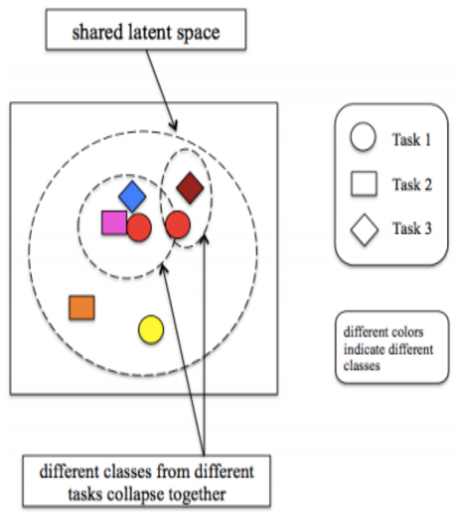


Figure 2: A schematic illustration of the motivation behind CTDD. Different shapes represent data from different tasks, and different colors represent different classes. In the shared latent space, different classes from different tasks could collapse together.

instance *i* from task *k*

instance *j* from task *l*

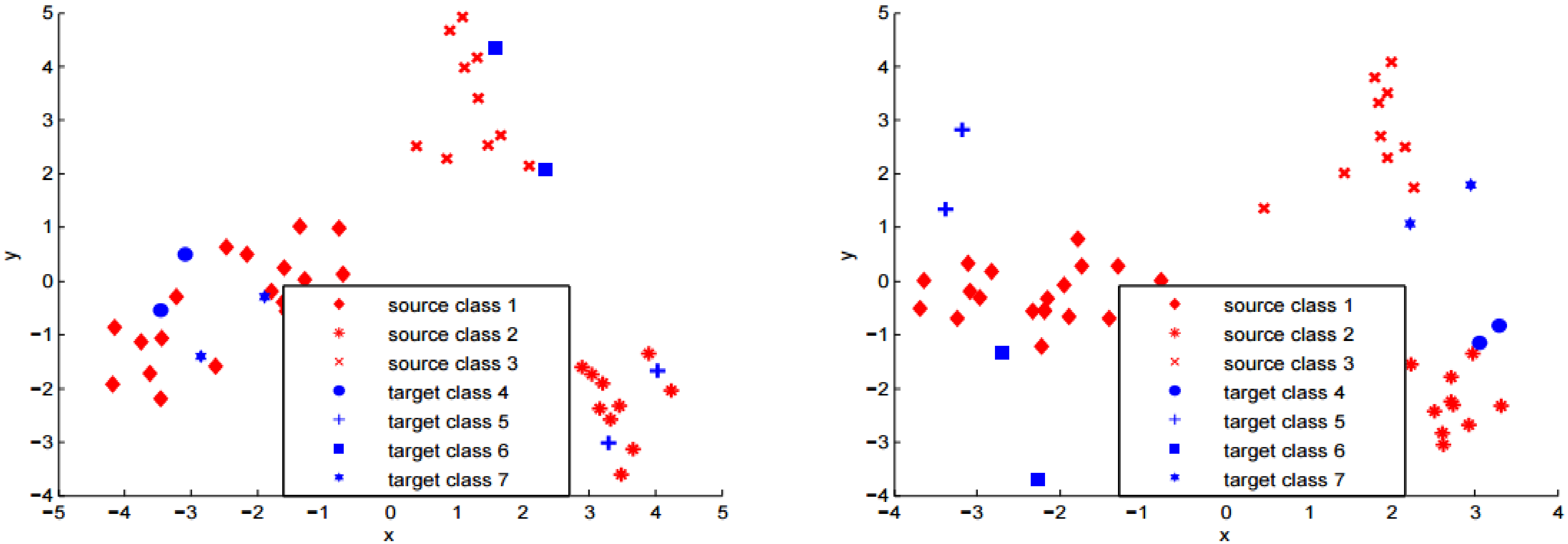
$$CTDD(\mathbf{W}_0) = \frac{1}{N}\,\mathrm{tr}\Big(\mathbf{W}_0'\Big\{\sum_{k \neq l}\sum_{i,j}(\mathbf{x}_i^k - \mathbf{x}_j^l)(\mathbf{x}_i^k - \mathbf{x}_j^l)'\Big\}\mathbf{W}_0\Big) \qquad (6)$$

separate data from different tasks

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{\mathrm{tr}(\mathbf{W}'\mathbf{A}\mathbf{W}) + \alpha CTDD(\mathbf{W}_0)}{\mathrm{tr}(\mathbf{W}'\mathbf{B}\mathbf{W})} \qquad (7)$$

中山大學 SUN YAT-SEN UNIVERSITY

Constrained Asymmetric Multi-task Component Analysis



(a) Without CTDD (b) With CTDD

Figure 3: Illustration of the effect of CTDD in the transfer from CAVIAR (source) to i-LIDS (target), where three source classes (in red) and four target classes (in blue) are used for demonstration. Different markers indicate different persons (classes). The x-axis and y-axis are the first two PCA scores of the samples in the shared latent space. When there is no CTDD, blue circles and blue hexagrams collapse with red diamonds, blue plus signs collapse with red asterisks. However, after imposing CTDD, data from different tasks are well separated.

中山大學 SUN YAT-SEN UNIVERSITY

Experiment

- Datasets
- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA

dataset	VIPeR	3DPeS	i-LIDS	CAVIAR
number of persons	632	192	119	72
number of images	1264	1011	476	1220
location (scenario)	street	campus	airport	shopping mall

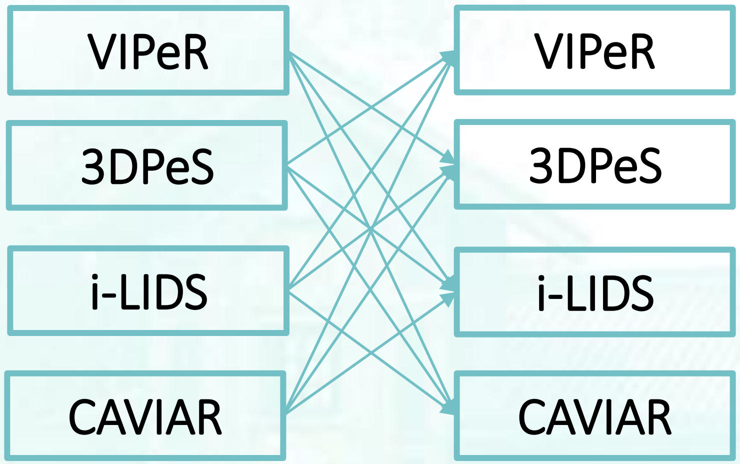
Table 1: Summary of datasets used in the experiments.

(a) VIPeR (b) 3DPeS (c) i-LIDS (d) CAVIAR

Figure 4: Illustration of person images of the four datasets. Images in the same column are from the same person.

Experiment

- Datasets`
- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA



Single transfer : 12 cases
Multiple transfer: 16 cases

Feature representation: concatenated color (RGB, YCbCr, HS), HoG, LBP features extracted from sub-blocks of images

Default parameter setting: $\beta = 0.1$, $\gamma = 0.8$, $\alpha = 1 - \beta$

Experiment

Datasets
Transfer setting
Compared methods
- single-task methods
- multi-task + domain adaptation methods

Further evaluation of cAMT-DCA

Compared methods: LFDA (Pedagadi et al.), LMNN (Weinberger et al.), KISSME (Kostinger et al.), LADF (Li et al.), PCCA (Mignon et al.), RDC (Zheng et al.)

trained in three ways
- using source data only (e.g. LFDA_T)
- using target data only (e.g. LFDA_S)
- using a pooled set of source data and target data (e.g. LFDA-Mix)

Experiment

- Datasets
- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA

Methods	VIPeR→i-LIDS				3DPeS→i-LIDS				CAVIAR→i-LIDS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**36.47**	**60.59**	**72.13**	**84.17**	**33.79**	**54.96**	**67.89**	**81.38**	**33.85**	**57.46**	**69.79**	**81.27**
LFDA_T	30.32	51.81	64.46	79.86	30.32	51.81	64.46	79.86	30.32	51.81	64.46	79.86
LMNN_T	27.14	46.61	56.41	74.00	27.14	46.61	56.41	74.00	27.14	46.61	56.41	74.00
KISSME_T	20.31	40.95	53.43	70.11	20.31	40.95	53.43	70.11	20.31	40.95	53.43	70.11
LADF_T	14.20	36.49	49.60	69.59	14.20	36.49	49.60	69.59	14.20	36.49	49.60	69.59
PCCA_T	13.48	34.14	50.30	71.01	13.48	34.14	50.30	71.01	13.48	34.14	50.30	71.01
RDC_T	30.42	51.19	61.88	77.10	30.42	51.19	61.88	77.10	30.42	51.19	61.88	77.10

Methods	VIPeR→CAVIAR				3DPeS→CAVIAR				i-LIDS→CAVIAR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**34.39**	**59.84**	**72.63**	**90.67**	**33.54**	**57.76**	**73.61**	**91.88**	**35.39**	**60.68**	**75.53**	**92.23**
LFDA_T	28.41	49.91	63.79	82.19	28.41	49.91	63.79	82.19	28.41	49.91	63.79	82.19
LMNN_T	24.41	39.71	55.78	79.40	24.41	39.71	55.78	79.40	24.41	39.71	55.78	79.40
KISSME_T	20.28	35.21	52.32	77.18	20.28	35.21	52.32	77.18	20.28	35.21	52.32	77.18
LADF_T	20.68	46.07	62.23	81.55	20.68	46.07	62.23	81.55	20.68	46.07	62.23	81.55
PCCA_T	16.45	37.98	53.81	76.30	16.45	37.98	53.81	76.30	16.45	37.98	53.81	76.30
RDC_T	28.75	45.86	58.55	75.25	28.75	45.86	58.55	75.25	28.75	45.86	58.55	75.25

Methods	VIPeR→3DPeS				i-LIDS→3DPeS				CAVIAR→ 3DPeS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**31.88**	**53.49**	**63.94**	**75.08**	**30.19**	**52.59**	**63.37**	**74.56**	**29.51**	**51.03**	**62.29**	**74.32**
LFDA_T	26.57	48.90	61.42	72.35	26.57	48.90	61.42	72.35	26.57	48.90	61.42	72.35
LMNN_T	23.68	43.91	55.45	67.88	23.68	43.91	55.45	67.88	23.68	43.91	55.45	67.88
KISSME_T	13.96	31.90	44.04	58.68	13.96	31.90	44.04	58.68	13.96	31.90	44.04	58.68
LADF_T	15.53	35.48	49.27	65.28	15.53	35.48	49.27	65.28	15.53	35.48	49.27	65.28
PCCA_T	8.56	25.13	37.55	54.12	8.56	25.13	37.55	54.12	8.56	25.13	37.55	54.12
RDC_T	25.58	44.74	54.59	65.07	25.58	44.74	54.59	65.07	25.58	44.74	54.59	65.07

Methods	i-LIDS → VIPeR				CAVIAR→VIPeR				3DPeS→ VIPeR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**23.39**	**52.75**	**67.12**	**81.14**	**22.18**	**50.44**	**64.94**	**80.32**	**21.61**	**50.92**	**66.27**	**81.36**
LFDA_T	20.89	48.39	63.96	78.51	20.89	48.39	63.96	78.51	20.89	48.39	63.96	78.51
LMNN_T	8.13	21.80	31.52	44.65	8.13	21.80	31.52	44.65	8.13	21.80	31.52	44.65
KISSME_T	20.25	48.01	63.23	79.81	20.25	48.01	63.23	79.81	20.25	48.01	63.23	79.81
LADF_T	9.72	29.53	44.34	61.14	9.72	29.53	44.34	61.14	9.72	29.53	44.34	61.14
PCCA_T	16.65	44.24	61.27	78.45	16.65	44.24	61.27	78.45	16.65	44.24	61.27	78.45
RDC_T	17.78	40.66	52.88	67.18	17.78	40.66	52.88	67.18	17.78	40.66	52.88	67.18

Table 2: Matching rate(%): cAMT-DCA vs. single-task methods. '_T' indicates the single-task methods are learned on target datasets only. Two sample images ($p = 2$) are used for each target person.

Experiment

- Datasets
- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA

Methods	VIPeR→i-LIDS				3DPeS→i-LIDS				CAVIAR→i-LIDS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**36.47**	**60.59**	**72.13**	**84.17**	**33.79**	**54.96**	**67.89**	81.38	**33.85**	**57.46**	**69.79**	**81.27**
LFDA_S	31.32	51.93	62.56	79.24	28.42	49.25	62.35	79.58	31.50	53.99	66.71	78.18
LMNN_S	29.16	50.41	63.96	79.19	27.52	46.61	60.32	76.38	29.43	52.37	62.84	76.33
KISSME_S	32.22	51.87	63.34	80.97	27.86	49.46	65.81	**81.65**	30.73	54.23	68.61	80.80
LADF_S	14.16	35.21	49.04	66.44	10.85	34.58	52.99	71.75	9.28	33.66	46.35	64.14
PCCA_S	22.83	40.97	54.41	71.25	23.55	46.44	61.45	80.02	19.64	43.20	59.31	76.77

Methods	VIPeR→CAVIAR				3DPeS→CAVIAR				i-LIDS→CAVIAR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**34.39**	**59.84**	**72.63**	**90.67**	**33.54**	**57.76**	**73.61**	**91.88**	**35.39**	**60.68**	**75.53**	**92.23**
LFDA_S	32.43	51.82	64.73	83.66	30.09	52.70	67.94	84.80	33.91	53.14	67.02	87.38
LMNN_S	28.01	48.40	64.56	84.16	27.18	47.59	63.04	83.57	28.97	48.09	64.04	84.04
KISSME_S	30.19	52.45	67.62	84.37	30.60	52.81	67.86	84.03	30.69	53.58	70.26	88.08
LADF_S	25.08	50.17	65.04	82.02	18.65	46.27	60.33	83.46	25.48	51.52	67.65	84.13
PCCA_S	23.07	41.67	57.47	83.27	24.04	46.79	61.91	83.51	20.78	50.12	69.50	85.64

Methods	VIPeR→3DPeS				i-LIDS→3DPeS				CAVIAR→ 3DPeS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**31.88**	**53.49**	**63.94**	**75.08**	**30.19**	**52.59**	**63.37**	**74.56**	**29.51**	**51.03**	**62.29**	**74.32**
LFDA_S	26.85	46.18	55.88	66.36	25.41	43.75	53.66	65.30	26.48	45.49	54.50	65.32
LMNN_S	26.93	47.04	56.12	66.72	24.43	43.20	52.04	63.00	25.72	44.57	53.94	64.74
KISSME_S	27.64	47.48	56.14	67.28	25.74	45.60	56.35	68.36	26.91	46.33	55.52	66.24
LADF_S	12.23	32.28	43.32	57.83	11.85	28.90	41.05	56.51	6.49	17.84	27.33	42.63
PCCA_S	19.67	39.70	51.11	63.93	17.03	35.72	47.90	63.09	16.53	35.31	46.30	61.86

Methods	i-LIDS → VIPeR				CAVIAR→VIPeR				3DPeS→ VIPeR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**23.39**	**52.75**	**67.12**	**81.14**	**22.18**	**50.44**	**64.94**	**80.32**	**21.61**	**50.92**	**66.27**	**81.36**
LFDA_S	8.16	22.47	33.32	44.59	8.23	21.11	30.06	43.26	8.64	22.18	33.61	48.10
LMNN_S	7.06	23.01	34.59	46.30	7.63	20.82	31.20	44.97	6.46	18.23	27.85	40.38
KISSME_S	8.13	22.15	31.96	44.78	9.87	20.00	29.37	41.65	6.87	20.95	29.43	42.94
LADF_S	2.72	10.35	17.85	28.35	1.08	4.94	9.91	16.71	3.04	11.11	19.68	31.68
PCCA_S	5.57	16.58	23.26	33.39	5.57	13.29	20.57	31.23	5.54	16.77	26.71	39.18

Table 3: Matching rate(%): cAMT-DCA vs. single-task methods. '_S' indicates the single-task methods are learned on source datasets only. Two sample images ($p = 2$) are used for each target person.

Experiment

- Datasets
- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA

Methods	VIPeR→i-LIDS				3DPeS→i-LIDS				CAVIAR→i-LIDS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**36.47**	**60.59**	**72.13**	**84.17**	**33.79**	**54.96**	**67.89**	**81.38**	**33.85**	**57.46**	**69.79**	**81.27**
LFDA-Mix	31.82	51.59	63.96	80.24	30.10	51.26	63.30	78.86	30.53	49.62	62.39	79.03
LMNN-Mix	30.15	51.20	63.57	79.98	27.69	47.84	60.33	75.95	29.26	49.30	62.23	76.17
KISSME-Mix	35.24	54.95	67.54	83.32	26.87	45.22	58.38	75.17	27.35	44.65	57.27	73.60
LADF-Mix	16.18	38.51	52.00	69.85	11.67	38.72	57.41	76.09	14.55	38.12	52.56	68.60
PCCA-Mix	23.96	47.39	62.06	77.85	18.02	44.51	61.40	78.92	20.04	45.74	59.78	74.94

Methods	VIPeR→CAVIAR				3DPeS→CAVIAR				i-LIDS→CAVIAR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**34.39**	**59.84**	**72.63**	**90.67**	**33.54**	**57.76**	**73.61**	**91.88**	**35.39**	**60.68**	**75.53**	**92.23**
LFDA-Mix	32.32	53.39	65.44	85.22	31.12	50.99	65.60	85.64	33.70	53.66	69.41	87.56
LMNN-Mix	27.80	49.62	65.00	85.17	27.05	46.87	62.15	83.45	27.94	47.20	62.07	82.55
KISSME-Mix	32.11	53.30	67.96	85.89	27.64	45.61	60.50	81.59	30.76	50.89	67.51	86.65
LADF-Mix	25.85	50.85	66.59	84.38	25.85	50.85	66.59	84.38	30.41	56.04	70.28	88.67
PCCA-Mix	25.63	48.43	64.26	85.79	24.72	49.69	67.73	87.64	26.38	52.26	69.20	88.01

Methods	VIPeR→3DPeS				i-LIDS→3DPeS				CAVIAR→ 3DPeS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**31.88**	**53.49**	**63.94**	**75.08**	**30.19**	**52.59**	**63.37**	**74.56**	**29.51**	**51.03**	**62.29**	**74.32**
LFDA-Mix	27.38	48.48	58.79	69.59	26.82	48.85	60.21	71.79	23.79	43.43	54.59	66.57
LMNN-Mix	27.44	47.92	58.01	69.42	24.92	45.64	55.59	67.28	25.29	45.15	55.62	67.75
KISSME-Mix	28.94	49.82	60.66	71.28	26.31	47.00	59.51	71.50	22.34	39.81	51.20	63.26
LADF-Mix	13.13	34.15	47.76	63.35	9.25	27.55	41.86	59.53	10.29	26.32	39.80	54.82
PCCA-Mix	22.39	45.66	58.18	71.89	22.36	44.23	56.63	71.97	19.32	40.26	52.38	67.84

Methods	i-LIDS → VIPeR				CAVIAR→VIPeR				3DPeS→ VIPeR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**23.39**	**52.75**	**67.12**	**81.14**	**22.18**	**50.44**	**64.94**	**80.32**	**21.61**	**50.92**	**66.27**	**81.36**
LFDA-Mix	19.24	45.44	59.18	75.25	16.68	40.73	56.61	72.47	16.90	42.63	58.23	74.78
LMNN-Mix	8.13	21.93	33.45	46.17	7.72	21.17	31.36	46.27	7.78	20.89	31.30	44.46
KISSME-Mix	15.03	35.47	49.34	64.46	9.05	20.66	29.72	39.94	12.22	32.18	44.15	58.48
LADF-Mix	6.61	21.11	33.58	49.08	6.30	21.42	33.45	49.15	8.96	28.70	42.85	58.83
PCCA-Mix	14.34	41.61	56.71	72.37	-	-	-	-	14.37	39.94	55.60	72.34

Table 4: Matching rate(%): cAMT-DCA vs. single-task methods. '-Mix' indicates the single-task methods are learned on a pooled set of source and target datasets. Two sample images ($p=2$) are used for each target person.

Two observations:

- Only using source dataset for the chosen metric learning algorithms often results in better performance than only using limited target data (except for the case with VIPeR as target dataset).

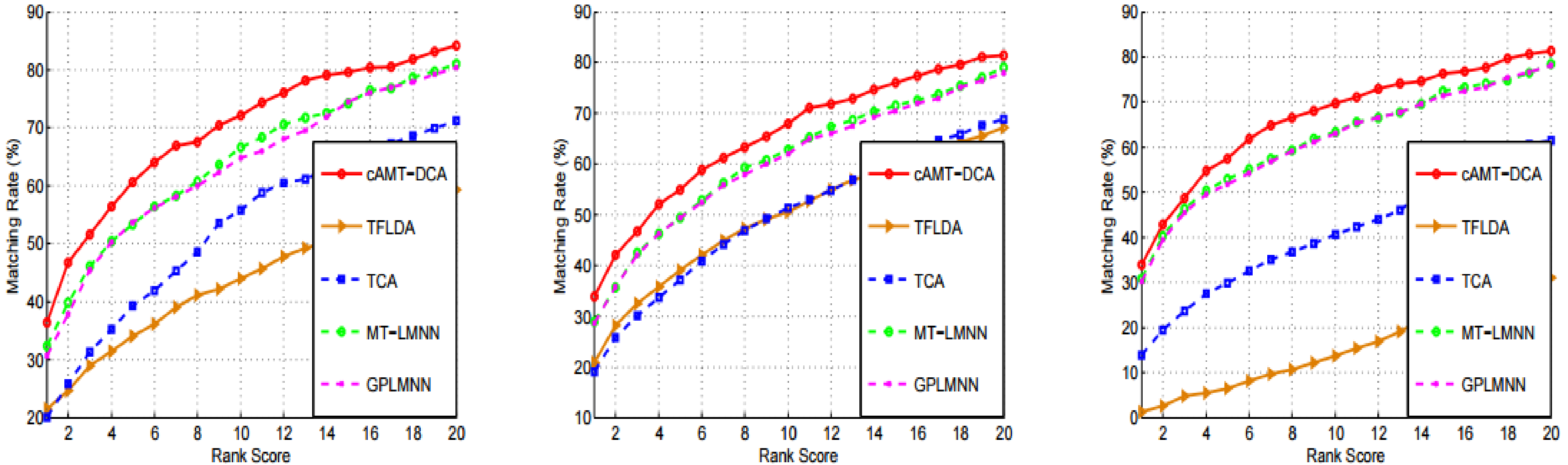
- Using the pooled set of source and target data for the chosen metric learning methods almost performs almost the same as using only source data and sometimes even worse.

Experiment

Datasets
Transfer setting
Compared methods
- single-task methods
- multi-task + domain adaptation methods

Further evaluation of cAMT-DCA

Compared methods: TCA (Pan et al.), TFLDA (Si et al.), MT-LMNN (Parameswaran et al.), GPLMNN (Yang et al,)



(a) VIPeR→i-LIDS (b) 3DPeS→i-LIDS (c) CAVIAR→i-LIDS

Figure 5: Matching rates of cAMT-DCA, multi-task methods and domain adaptation methods, with i-LIDS as target dataset. Two sample images ($p = 2$) are used for each target person.

Experiment

- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA

With CTDD vs. Without CTDD

Methods	VIPeR→i-LIDS				3DPeS→i-LIDS				CAVIAR→i-LIDS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**36.47**	**60.59**	**72.13**	84.17	**33.79**	54.96	67.89	**81.38**	**33.85**	**57.46**	**69.79**	**81.27**
AMT-DCA	35.75	57.85	70.51	**84.39**	32.84	**55.13**	**68.11**	80.53	32.55	53.89	66.48	79.80
Methods	VIPeR→CAVIAR				3DPeS→CAVIAR				i-LIDS→CAVIAR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**34.39**	**59.84**	**72.63**	**90.67**	**33.54**	**57.76**	**73.61**	**91.88**	**35.39**	**60.68**	**75.53**	92.23
AMT-DCA	33.45	55.42	70.81	89.52	33.14	56.18	71.14	91.72	33.87	58.75	73.35	**92.52**
Methods	VIPeR→3DPeS				i-LIDS→3DPeS				CAVIAR→ 3DPeS			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**31.88**	**53.49**	**63.94**	**75.08**	**30.19**	**52.59**	**63.37**	**74.56**	**29.51**	**51.03**	**62.29**	**74.32**
AMT-DCA	30.48	52.45	62.49	73.72	29.43	51.23	62.63	73.66	27.59	48.26	59.16	71.08
Methods	i-LIDS→VIPeR				CAVIAR→VIPeR				3DPeS→ VIPeR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**23.39**	**52.75**	**67.12**	**81.14**	**22.18**	**50.44**	**64.94**	**80.32**	**21.61**	**50.92**	**66.27**	**81.36**
AMT-DCA	21.36	50.54	66.20	81.08	20.35	48.13	62.94	77.12	20.13	49.68	65.25	79.11

Table 5: Matching rate(%): With and Without CTDD in cAMT-DCA. The AMT-DCA is exactly cAMT-DCA without using CTDD. Two sample images ($p = 2$) are used for each target person.

Experiment

Transfer setting
Compared methods
- single-task methods
- multi-task + domain adaptation methods

Further evaluation of cAMT-DCA

Parameter evaluation

varying β in [0 : 0.1 :1], γ in [0.1:0.1:1]

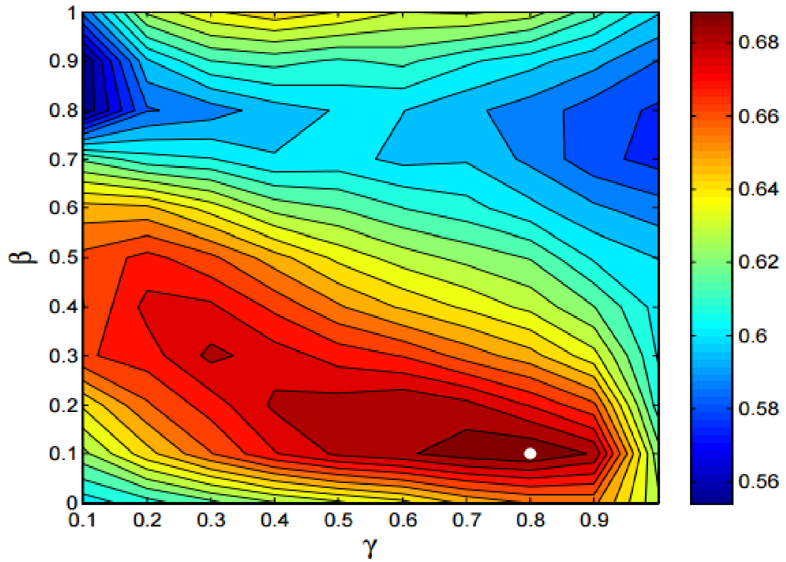


Figure 6: Visualization of AUC contour parameterized by β and γ in VIPeR→i-LIDS, the highest AUC value is highlighted by the white spot in the figure. Two sample images are used for each target person.

The optimal values are β = 0.1, γ = 0.8, and same result could be observed in other transfer cases

Experiment

Transfer setting
Compared methods
- single-task methods
- multi-task + domain adaptation methods

Further evaluation of cAMT-DCA

Increase number of target training samples

Methods	$p=3$				$p=4$				$p=5$			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**36.63**	**58.72**	**74.12**	**89.94**	**37.09**	59.90	**75.34**	91.39	**39.22**	**61.88**	**76.55**	89.88
LFDA-Mix	32.73	53.61	66.82	84.78	33.72	54.47	68.03	86.19	34.03	55.10	68.39	86.57
LMNN-Mix	28.34	47.60	62.25	82.87	29.99	47.98	61.53	83.05	28.63	46.81	61.57	83.21
KISSME-Mix	32.89	58.58	72.13	88.02	33.97	**60.62**	74.51	**92.64**	36.91	60.84	74.92	**91.18**
LADF-Mix	23.46	51.65	67.81	83.68	20.76	49.90	67.01	87.26	26.20	56.31	72.38	88.33
PCCA-Mix	27.68	53.52	68.11	87.64	25.81	54.86	72.25	90.15	27.93	55.61	71.37	89.36
TCA	19.10	37.13	50.84	73.94	19.68	36.27	49.59	77.96	19.05	36.58	51.70	74.40
TFLDA	18.67	33.43	49.09	70.68	20.05	33.40	49.27	70.94	19.81	33.42	48.76	71.16
MT-LMNN	29.85	52.90	68.40	85.64	30.92	49.71	64.11	84.55	29.00	51.05	66.22	85.90
GPLMNN	30.04	52.98	68.58	86.37	30.16	49.51	63.63	84.91	29.52	49.83	64.93	85.86

Table 6: cAMT-DCA vs. others: matching rate(%) in "VIPeR→CAVIAR", with respect to different number p of target training images for each person.

中山大學
SUN YAT-SEN UNIVERSITY

Experiment

- Transfer setting
- Compared methods
  - single-task methods
  - multi-task + domain adaptation methods
- Further evaluation of cAMT-DCA

Multiple transfer

Methods	VIPeR+CAVIAR→i-LIDS				VIPeR+3DPeS→CAVIAR				VIPeR+i-LIDS→3DPeS				CAVIAR+i-LIDS→VIPeR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**35.64**	**58.86**	**70.45**	**83.72**	**33.70**	**58.20**	**75.68**	**93.45**	**31.86**	**52.37**	**63.06**	**73.29**	**20.35**	**48.26**	**63.01**	**77.94**
LFDA-Mix	30.27	51.20	64.57	80.07	32.14	53.10	64.73	85.28	28.00	49.50	58.68	70.15	17.25	42.94	58.20	73.13
LMNN-Mix	30.15	51.09	63.57	77.34	26.99	47.49	61.60	83.66	25.93	45.41	54.99	67.05	8.32	21.14	32.44	46.36
KISSME-Mix	26.58	48.57	59.94	75.89	30.58	49.76	61.84	83.94	27.64	48.58	58.65	70.07	8.77	21.68	31.08	41.84
LADF-Mix	18.86	42.35	55.90	71.46	22.76	51.48	68.80	90.11	9.40	27.93	40.79	58.88	6.80	20.76	32.15	47.63
MT-LMNN	31.39	54.44	66.88	81.48	29.14	51.75	65.87	88.42	28.43	49.12	60.19	71.49	16.33	43.61	58.04	72.56
GPLMNN	32.00	52.98	65.70	80.98	29.47	50.70	63.47	87.69	27.26	48.20	59.05	70.75	16.20	43.07	57.37	72.72
TCA	14.51	32.70	44.65	64.55	21.87	41.25	53.79	74.61	14.84	27.85	37.89	50.22	5.89	16.55	24.43	37.06
TFLDA	21.04	41.79	52.88	68.72	18.54	34.71	49.04	73.45	18.28	35.57	46.10	57.66	5.16	13.54	19.68	29.21

Table 7: cAMT-DCA vs. others: matching rate(%) with i-LIDS, CAVIAR, 3DPeS and VIPeR as target dataset each, and two of others are used as sources for transfer. Two sample images ($p=2$) are used for each target person.

Methods	VIPeR+CAVIAR+3DPeS→i-LIDS				VIPeR+i-LIDS+3DPeS→CAVIAR				VIPeR+CAVIAR+i-LIDS→3DPeS				CAVIAR+i-LIDS+3DPeS→VIPeR			
	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$	$r=1$	$r=5$	$r=10$	$r=20$
cAMT-DCA	**36.53**	**59.31**	**70.50**	**84.39**	**36.22**	**59.82**	**75.16**	**92.58**	**30.52**	**51.68**	**61.43**	**72.28**	**19.49**	**46.77**	**61.87**	**77.72**
LFDA -Mix	32.51	53.43	66.97	81.37	33.07	53.95	66.49	85.43	26.04	45.18	56.76	68.78	16.80	42.56	56.93	72.28
LMNN-Mix	31.21	50.69	63.01	77.51	28.30	48.30	62.57	83.37	26.22	45.18	55.25	66.94	8.45	22.34	32.15	46.27
KISSME-Mix	29.21	48.06	63.40	78.13	31.85	51.76	66.49	85.30	25.96	44.16	53.79	66.62	9.87	22.18	31.87	44.72
LADF-Mix	16.94	42.97	57.86	72.86	22.27	52.31	71.17	89.36	13.45	34.43	47.57	63.78	6.65	19.08	30.54	44.84
MT-LMNN	31.78	55.56	66.60	81.37	30.58	52.92	67.54	87.22	29.14	49.70	60.57	70.97	16.36	42.28	56.20	72.31
GPLMNN	32.79	53.26	64.86	81.42	30.16	50.81	65.15	87.82	27.89	48.72	59.19	70.94	16.30	41.84	55.57	71.96
TCA	16.20	35.01	46.50	65.32	20.75	38.16	55.02	78.78	14.61	28.49	37.69	50.64	4.56	14.11	21.65	30.89
TFLDA	24.41	41.53	55.98	70.58	19.03	34.21	49.88	73.04	19.30	34.09	43.75	56.12	4.91	13.20	21.58	31.71

Table 8: cAMT-DCA vs. others: matching rate(%) with i-LIDS, CAVIAR, 3DPeS and VIPeR as target dataset each, and the other three are used as sources for transfer. Two sample images ($p=2$) are used for each target person.

中山大學
SUN YAT-SEN UNIVERSITY

Discussion

using more source datasets does not mean a better improvement

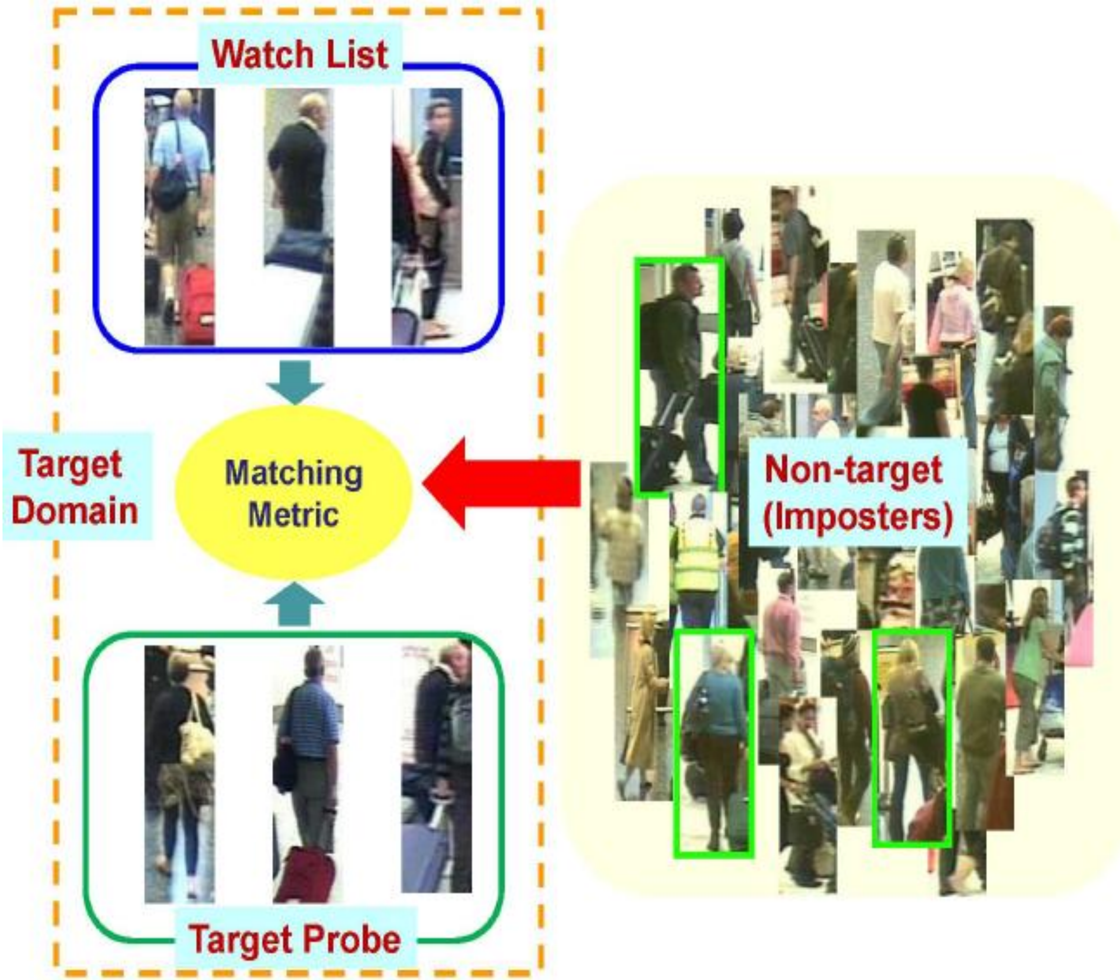
how to select suitable source or source sample pairs for transfer

In real world, there are quite a lot of imposters, and only a few guys are target to track

Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Towards Open-World Person Re-Identification by One-Shot Group-based Verification. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2015. (DOI: 10.1109/TPAMI.2015.2453984)

One-Shot Open-World Group-based Re-id

Motivation



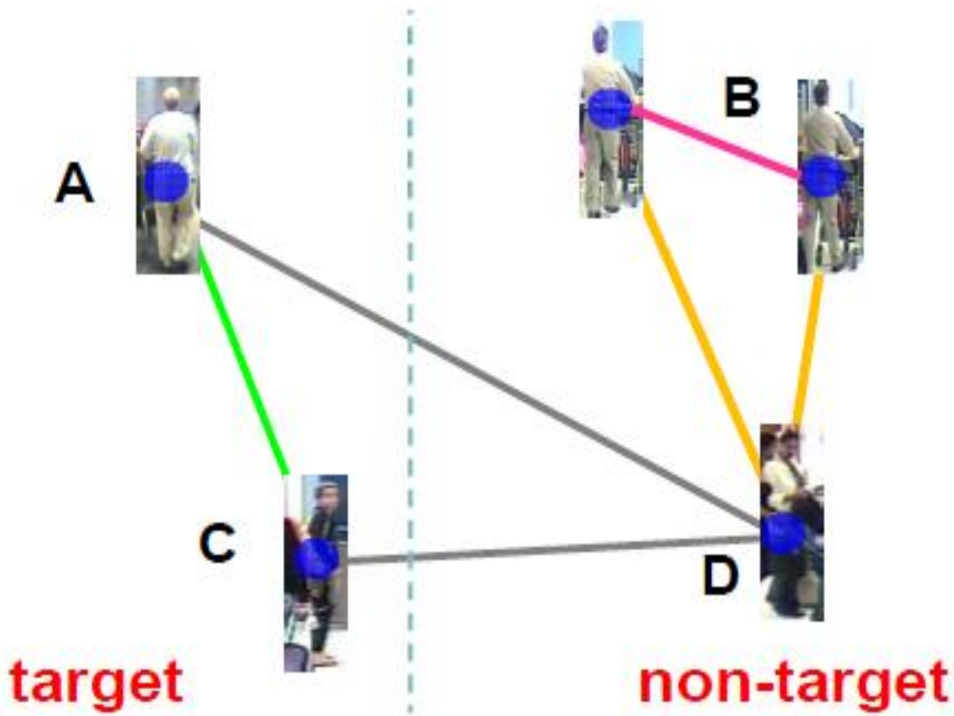
Open-world person re-identification setting

1) **A large amount of non-target imposters captured along with the target people on the watch list.**

2) **Their images will also appear in the probe set and some of them will look visually similar to the target people**

One-Shot Open-World Group-based Re-id

- **Knowledge to transfer**



Enrich intra-class variation

Approximate target intra-inter class pair (magenta line and green line)

Enrich inter-class variation

Target specific non-target intra-inter class pair
(magenta line and yellow line)

Enrich group separation

Group separation intra-inter class pair (green line and grey line)

One-Shot Open-World Group-based Re-id

- **Criterion**

$$\min_{\mathbf{M}\succeq 0} f(\mathbf{M})$$

$$f(\mathbf{M}) = \frac{1-\alpha}{\#\mathbb{O}_g}\sum_{t=1}^{N_T}\sum_{(\mathbf{x}_{t_j},\mathbf{x}_s,\mathbf{x}_{t'})\in\mathbb{O}_g(\mathbf{x}_t)} \ell(d(\mathbf{x}_{t_j},\mathbf{x}_s) < d(\mathbf{x}_t,\mathbf{x}_{t'}))$$

$$+\frac{\alpha}{\#\mathbb{O}_a+\#\mathbb{O}_b}\Big(\sum_{t=1}^{N_T}\sum_{(\mathbf{x}_s,\mathbf{x}_{s'},\mathbf{x}_{s''})\in\mathbb{O}_a(\mathbf{x}_t)} \ell(d(\mathbf{x}_s,\mathbf{x}_{s'}) < d(\mathbf{x}_s,\mathbf{x}_{s''}))$$

$$+\beta\sum_{t=1}^{N_T}\sum_{(\mathbf{x}_{t'},\mathbf{x}_s)\in\mathbb{O}_b(\mathbf{x}_t)} \ell(d(\mathbf{x}_t,\mathbf{x}_{t'}) < d(\mathbf{x}_t,\mathbf{x}_s))\Big).$$

Enrich intra-class variation

Approximate target intra-inter class pair (magenta line and green line)

Enrich inter-class variation

Target specific non-target intra-inter class pair
(magenta line and yellow line)

Enrich group separation

Group separation intra-inter class pair (green line and grey line)

中山大學
SUN YAT-SEN UNIVERSITY

Other Distance Model for RE-ID

Local Relative Distance Comparison

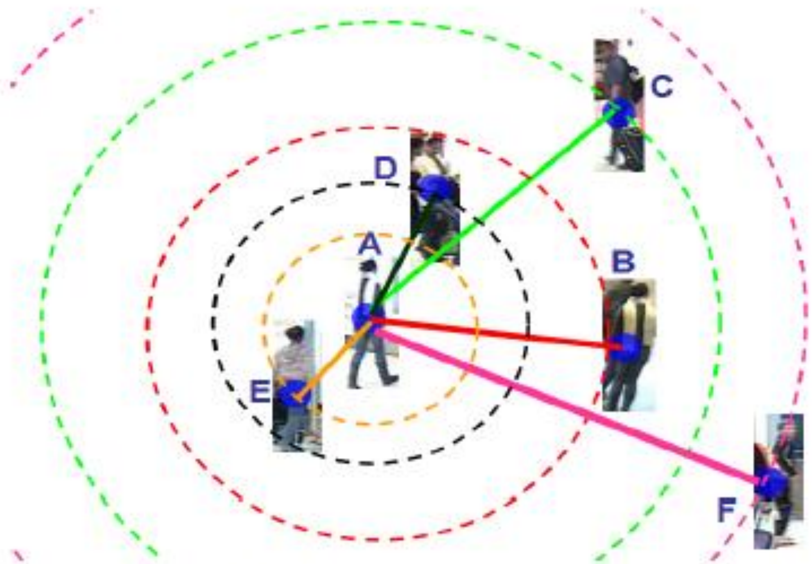


Fig. 3. Illustration of our local relative comparison. Among the six images, A and B belong to the same person whilst the other four are of four other people. See text for more details.

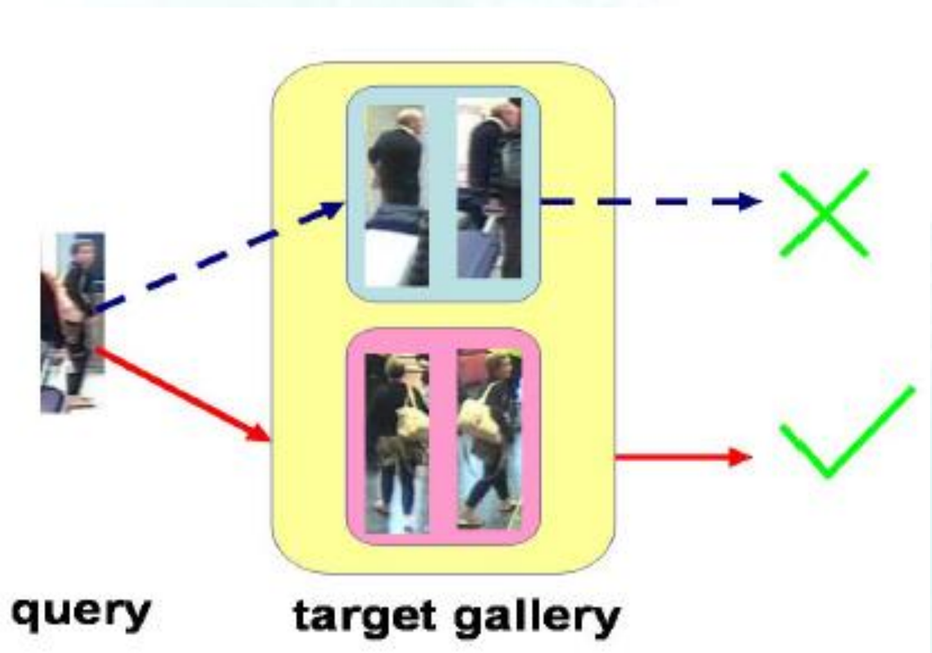
$$d(\mathbf{x}_i, \mathbf{x}_j) < d(\mathbf{x}_i, \mathbf{x}_m) - \rho, \ \rho > 0$$
$$\text{when } (\mathbf{x}_m - \mathbf{x}_i) \in \mathcal{N}_k^{\mathrm{P}}(\mathbf{x}_i, \mathcal{D}), \ (\mathbf{x}_j - \mathbf{x}_i) \in \mathcal{D}$$

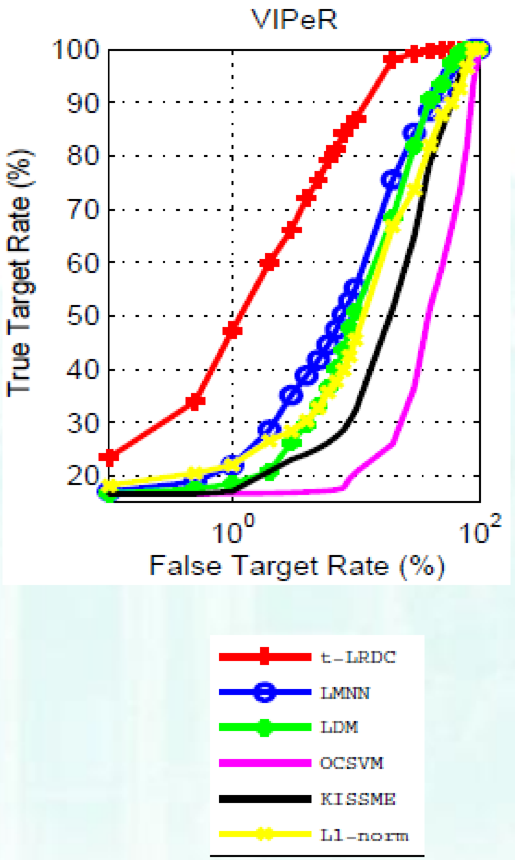
1) $\mathcal{D}_{y_i}^{+}(\mathbf{x}_i)$ denotes all the intra-class difference vectors related to $\mathbf{x}_i$ within class y_i, i.e. $\mathcal{D}_{y_i}^{+}(\mathbf{x}_i) = \{(\mathbf{x}_q - \mathbf{x}_i) \mid y_q = y_i\}$;
2) $\mathcal{D}_{y_i}^{-}(\mathbf{x}_i)$ denotes all the inter-class difference vectors between $\mathbf{x}_i$ and any other image out of class y_i but still from one of the target classes, i.e. $\mathcal{D}_{y_i}^{-}(\mathbf{x}_i) = \{(\mathbf{x}_q - \mathbf{x}_i) \mid y_q \neq y_i \ \& \ 1 \leq q \leq N_T\}$;

SUN YAT-SEN UNIVERSITY

One-Shot Open-World Group-based Re-id

Group-based verification





中山大學
SUN YAT-SEN UNIVERSITY

One-Shot Open-World Group-based Re-id

Individual Verification

Database	i-LIDS						ETHZ					
FTR	0.1%	1%	5%	10%	20%	30%	0.1%	1%	5%	10%	20%	30%
t-LRDC	14.58	**32.03**	**48.36**	**61.64**	74.57	81.65	45.13	65.62	83.82	89.86	95.20	97.67
t-LRDC(Global)	13.45	30.94	47.35	61.07	**76.66**	**87.30**	42.22	62.72	79.95	86.69	92.45	96.89
t-RDC	**16.78**	30.98	45.31	57.12	72.07	81.91	**54.14**	**76.29**	**88.07**	91.91	96.02	98.38
t-RankSVM	14.31	27.12	42.06	55.10	70.86	77.31	50.49	74.70	87.82	**92.72**	**96.60**	**99.09**
t-RDC-PCA	10.85	24.49	39.39	49.64	63.57	70.92	42.33	61.57	76.23	82.76	89.54	92.94
t-RankSVM-PCA	7.44	17.06	36.76	46.76	60.31	70.05	35.13	55.75	74.98	81.72	87.54	91.36
RDC [50]	15.16	28.04	44.89	57.53	70.89	79.99	53.16	75.07	87.30	91.67	95.16	97.63
RankSVM [31]	12.09	23.66	40.97	56.07	69.26	77.76	47.87	72.40	86.62	91.56	95.96	98.82
OCSVM [33]	6.00	6.34	11.78	17.87	28.59	36.25	0.56	2.23	11.62	18.36	28.11	35.12
KISSME [17]	11.77	25.46	36.74	44.92	61.00	67.79	46.49	61.21	76.31	85.33	93.06	96.94
LMNN [42]	8.61	20.81	41.43	49.92	58.00	68.85	41.80	58.65	75.43	82.59	90.74	93.43
LDM [44]	8.51	18.24	39.08	48.80	61.65	72.96	29.76	49.80	69.37	78.05	86.01	90.83
LADF [22]	7.86	20.72	39.88	53.80	69.29	79.89	20.23	53.14	76.67	85.86	93.67	96.42
LFDA [29]	7.22	13.43	24.72	35.47	50.11	63.74	27.49	43.98	60.96	73.52	84.83	89.23
Salience [46]	6.00	6.10	8.07	11.81	20.40	29.48	26.87	44.76	55.85	63.09	71.80	79.92
L1-norm	8.42	19.90	43.50	53.22	60.53	69.29	42.39	60.47	77.45	84.45	89.52	92.97

Database	CAVIAR						VIPeR					
FTR	0.1%	1%	5%	10%	20%	30%	0.1%	1%	5%	10%	20%	30%
t-LRDC	**15.45**	**28.13**	**40.76**	**50.78**	60.80	69.99	**23.47**	**47.27**	75.41	86.88	98.04	99.17
t-LRDC(Global)	13.78	25.85	39.87	49.01	63.18	71.58	19.63	39.04	69.25	84.13	96.17	98.13
t-RDC	14.08	24.40	39.30	49.13	**63.59**	**71.82**	19.38	40.72	73.18	88.42	97.85	98.71
t-RankSVM	10.64	20.90	35.67	45.50	57.83	68.23	22.73	45.95	76.12	**89.44**	97.86	98.99
t-RDC-PCA	12.48	23.38	36.55	45.43	57.65	66.49	18.79	24.73	40.03	54.54	76.71	85.61
t-RankSVM-PCA	12.41	20.00	33.37	42.23	55.73	63.98	17.53	21.60	34.38	44.12	68.77	79.58
RDC [50]	14.61	23.40	37.32	47.08	59.40	69.15	19.27	43.98	**77.95**	88.62	96.00	99.89
RankSVM [31]	6.33	16.64	31.43	42.04	56.25	64.13	20.27	44.97	77.41	89.16	96.70	**100**
OCSVM [33]	1.85	2.56	5.75	11.04	22.99	33.12	16.66	16.69	17.12	20.68	26.03	36.62
KISSME [17]	13.40	23.60	33.96	43.45	54.47	64.25	16.93	29.97	68.92	79.80	93.50	98.73
LMNN [42]	13.78	23.01	36.50	43.65	54.69	63.22	17.11	21.98	41.73	55.23	75.51	84.26
LDM [44]	9.48	17.65	29.86	39.72	53.96	62.74	16.76	18.54	33.18	50.81	68.42	81.82
LADF [22]	6.04	17.63	38.19	49.72	63.46	74.8	18.59	27.43	68.40	84.37	**99.41**	**100**
LFDA [29]	9.39	16.15	27.20	36.37	47.15	56.54	16.66	20.54	28.65	41.42	56.89	67.06
Salience [46]	13.45	24.05	34.99	43.53	52.63	59.45	16.67	16.84	17.81	19.03	25.46	35.32
L1-norm	13.57	24.27	35.95	44.93	53.54	62.83	16.96	21.13	38.11	46.94	63.45	76.55

TABLE I

ONE-SHOT INDIVIDUAL VERIFICATION RESULTS: TRUE TARGET RATE (TTR) IN % AGAINST FALSE TARGET RATE (FTR).

One-Shot Open-World Group-based Re-id

Individual Verification

Database	i-LIDS						ETHZ					
FTR	0.1%	1%	5%	10%	20%	30%	0.1%	1%	5%	10%	20%	30%
t-LRDC	9.65	18.75	**33.55**	**43.18**	**51.75**	**63.36**	36.63	47.90	62.81	71.38	80.42	85.43
t-LRDC(Global)	8.10	16.97	32.62	39.43	48.26	58.69	31.34	46.34	60.84	67.28	76.17	81.40
t-RDC	**10.82**	**20.73**	32.24	37.70	48.73	58.08	**38.97**	**59.66**	**74.86**	**81.48**	**86.84**	90.25
t-RankSVM	8.82	16.29	26.73	34.18	46.86	56.92	33.24	57.10	72.82	80.10	86.54	90.19
t-RDC-PCA	6.98	13.55	25.37	34.18	44.49	53.78	32.38	46.48	59.61	68.21	76.04	80.81
t-RankSVM-PCA	7.19	10.33	18.63	24.98	42.94	54.04	22.59	38.75	54.60	61.99	72.16	78.94
RDC [50]	7.72	17.32	28.63	38.13	47.73	58.63	38.76	57.64	73.79	80.76	86.67	90.18
RankSVM [31]	7.20	14.48	23.40	31.99	47.57	59.40	31.09	53.63	70.81	78.88	85.13	**90.65**
OCSVM [33]	6.02	7.05	13.27	18.22	29.28	36.44	1.01	3.34	12.89	18.95	28.48	35.56
KISSME [17]	8.88	14.68	26.83	35.23	41.36	48.91	34.83	49.35	59.94	68.16	76.87	84.63
LMNN [42]	6.84	10.03	21.88	32.83	46.00	54.15	33.61	43.91	58.51	66.56	76.21	82.81
LDM [44]	7.13	10.12	19.87	25.30	41.92	56.19	21.58	32.47	49.26	58.39	69.34	77.26
LADF [22]	7.25	11.66	23.24	31.35	44.68	56.88	10.23	26.88	51.49	61.75	73.12	81.26
LFDA [29]	6.59	8.51	15.73	21.28	30.28	42.29	19.80	31.42	44.64	52.37	63.01	69.96
Salience [46]	6.00	6.08	7.11	10.52	17.55	26.41	16.51	35.92	53.25	61.25	73.00	83.20
L1-norm	7.19	9.58	18.92	30.54	48.17	57.71	31.39	46.06	60.13	66.88	77.15	83.31

Database	CAVIAR						VIPeR					
FTR	0.1%	1%	5%	10%	20%	30%	0.1%	1%	5%	10%	20%	30%
t-LRDC	**11.65**	16.74	**29.92**	**36.72**	**47.53**	58.04	16.66	**27.43**	**45.99**	**63.27**	**75.11**	**88.62**
t-LRDC(Global)	11.09	15.40	26.37	34.24	45.82	56.67	18.48	20.63	37.20	54.26	67.42	78.49
t-RDC	9.44	**16.89**	26.50	34.51	46.22	56.44	17.11	21.20	38.08	50.97	72.28	79.22
t-RankSVM	5.45	13.94	23.08	29.48	41.89	53.57	**18.96**	23.83	42.49	58.38	72.82	83.62
t-RDC-PCA	9.00	13.87	24.65	33.90	46.28	56.70	16.66	19.21	24.55	32.75	49.96	58.70
t-RankSVM-PCA	8.31	15.20	22.69	29.48	39.72	51.48	16.66	18.60	22.20	27.08	39.67	53.08
RDC [50]	10.18	16.66	26.89	34.83	47.22	**58.07**	16.79	22.57	41.24	56.02	69.29	83.11
RankSVM [31]	3.35	10.14	20.20	28.16	42.42	53.56	16.92	23.17	42.45	57.36	72.87	82.21
OCSVM [33]	2.17	2.75	6.06	11.31	23.60	33.42	16.66	16.70	17.13	20.85	26.07	36.72
KISSME [17]	9.36	16.39	25.41	32.35	41.59	50.70	16.68	20.24	30.37	52.22	74.25	83.83
LMNN [42]	9.50	15.15	25.49	34.00	46.61	55.62	16.76	17.62	21.93	31.96	52.70	62.87
LDM [44]	6.39	11.28	19.12	27.56	39.55	49.93	16.66	17.53	23.17	30.31	46.06	62.19
LADF [22]	4.0	8.75	19.33	28.68	43.52	51.99	17.21	18.92	26.25	44.35	65.93	82.37
LFDA [29]	7.73	11.72	20.26	26.51	36.91	48.48	16.66	16.77	23.00	31.09	44.12	51.28
Salience [46]	10.13	15.15	25.58	32.74	44.89	52.73	16.67	16.73	17.44	18.54	21.60	25.88
L1-norm	10.48	15.58	26.38	34.55	45.12	54.87	16.72	17.24	20.81	33.80	48.20	61.58

TABLE II

ONE-SHOT SET VERIFICATION RESULTS: TRUE TARGET RATE (TTR) IN % AGAINST FALSE TARGET RATE (FTR).

Conclusions

- We formulate a relative distance comparison learning for person re-identification.
- We introduce a Mirror Representation for Modeling View-specific Transform in Person Re-identification
- We develop a *Constrained Asymmetric Multi-task Discriminant Component* model (cAMT-DCA) , the first to attempt to address the problem of *Cross-scenario Transfer Person Re-identification* with a model of asymmetric multi-task learning.
- We introduce a person verification model to avoid the effect of imposter for re-identification

中山大學
SUN YAT-SEN UNIVERSITY

Large Scale Person Identification

Local Online Learning

Zhaoze Zhou, Wei-Shi Zheng, Jian-Fang Hu, Yong Xu, Jane You, One-pass Online Learning: A Local Approach", Pattern Recognition(PR), 2015, to appear

中山大學
SUN YAT-SEN UNIVERSITY

Background: Challenge of Large Scale Data

- **Speed of data generation >> Speed of data processing**
- **Bottlenecks of classical methods (CPU/Memory/Disk Space)**
- **Known techniques:**
  - Parallelizing classical methods with data parallelization, such as distributed version of SVM
  - Design algorithms targeted at distributed computing paradigm, such as Map-Reduce
  - ONLINE LEARNING: Effective & Efficient machine learning algorithm!

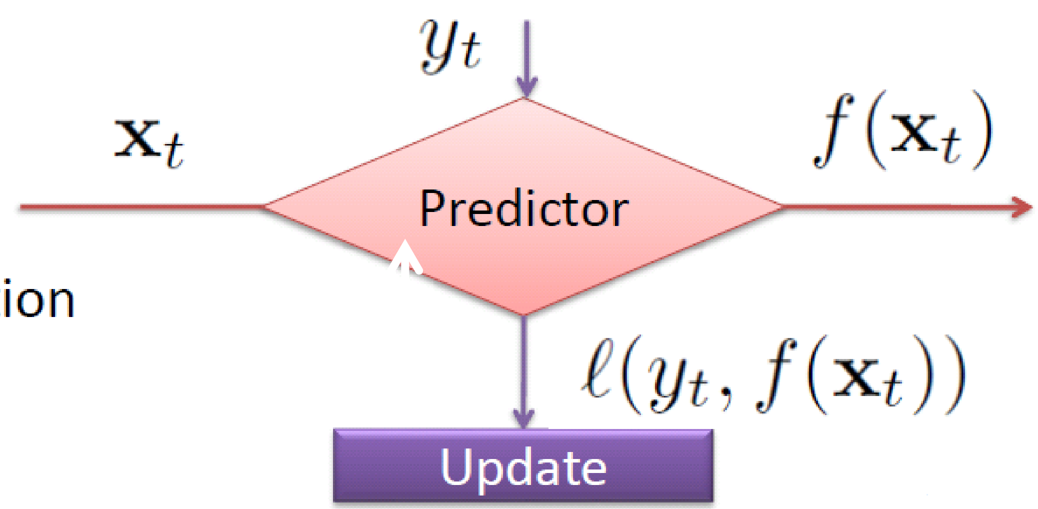
What is online learning?

Offline Learning vs. Online Learning

Learn a model with all or a batch of training data and then make predictions with learned model.

Learn model from sample sequence (data stream), prediction and learning are performed in the same time.

Example:
Online Classification



Advantages of Online Learning algorithm

- **Avoiding re-computation over all the samples when a new sample is observed.**
- **Saving memory and avoiding random access.**
- **Adaptation to variability of data stream.**

Our target problem

Online Learning of linear model

- For Instance, *Perceptron* is a well known online learning method, which make prediction and update its model when prediction is wrong

- Advantages: simple, efficient

- One sample is processed at a time

Passive Aggressive (PA) online learning

Core idea: searching for a closet hyperplane to the current one to satisfy the minimum margin constraint

Object function: $\mathbf{w}_{t+1} = \arg\min_{\mathbf{w}} \frac{1}{2} \boxed{\|\mathbf{w} - \mathbf{w}_t\|^2} + C\xi,$

$$\boxed{s.t.\ \ell^{pa}(\mathbf{w}; (\mathbf{x}_t, y_t)) \leq \xi \quad \text{and} \quad \xi \geq 0}$$

in which, $\ell^{pa}(\mathbf{w}; (\mathbf{x}_t, y_t)) = max\{0, 1 - y_t \cdot \mathbf{w}^T \mathbf{x}_t\}$

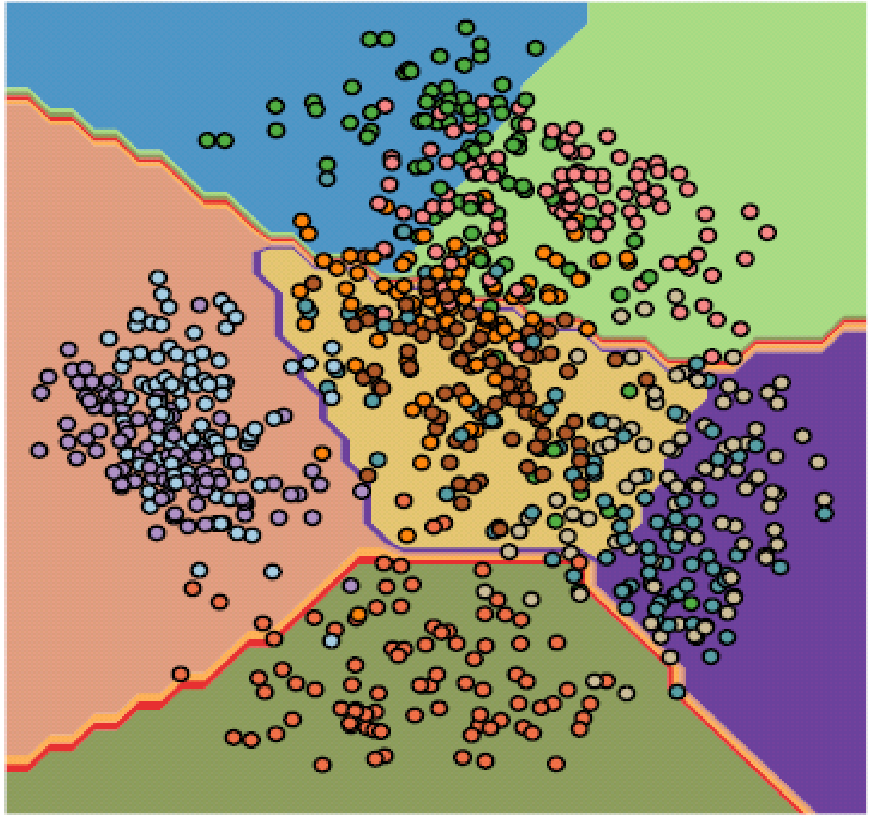
with the final update form:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta_t y_t \mathbf{x}_t, \quad \eta_t = \min\{C, \frac{\ell^{pa}}{\|\mathbf{x}_t\|^2}\}$$

PA is a globally linear model!

Problem of Linearly Nonseparable Data

- When we meet a large scale of data, it is hard to find a hyperplane to fit all (most) samples.



SUN YAT-SEN UNIVERSITY

Kernel method/Kernel approximation

1)Direct application of kernel method, causing memory overflow! ☹

2)Improvement: Budget based kernel method, keeping a limited subset of support vectors

3) Mapping to high-dimension space with kernel approximation methods:

$$k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle \approx \mathbf{z}(\mathbf{x})'\mathbf{z}(\mathbf{y}).$$

- Such as random Fourier feature

* High computational cost!

Computational Efficiency vs. Classification Accuracy

	Advantages	Disadvantages
Online (linear) learning	Constant memory, high efficiency☺	Low performance when data is not linearly separable☹
Kernel Method	Implicitly fit the non-linear nature, high accuracy when parameters are well set	High computational cost and increase of memory

Main idea and contributions

- **Though it is not separable globally, maybe it is separable locally**
- **Locally linear or Piece-wise linear**
- ~~**Straight forward idea: data parallelism, i.e. training independent models on each subset (usually by clustering) . However the models lose global information**~~☹。
- **Key idea of our model:**
  - assuming local hyperplanes share a common component, which leverages information between hyperplanes. ☺

Formula of our model

Model:

$$f(\mathbf{x}) = \sum_{i=1}^{k} f_i(\mathbf{x}) \cdot \mathbf{1}(\mathbf{x} \in \mathcal{D}(\mathcal{P}_i)) \qquad f_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x}$$

Define common component and local components: $\mathbf{w}_i = \mathbf{w} + \mathbf{u}_i$

Objective function:

$$\ell^{lol}(\mathbf{w}, \mathbf{u}_1, \ldots, \mathbf{u}_k; (\mathbf{x}_t, y_t)) = \begin{cases} 0 & y_t \cdot f(\mathbf{x}_t) \geq 1 \\ 1 - y_t \cdot f(\mathbf{x}_t) & \text{otherwise} \end{cases}$$

No. of local hyperplanes

$$\mathbf{w}_{t+1} = \arg \min_{\mathbf{w}, \mathbf{u}_1,} \ldots \boxed{\mathbf{u}_{i,t}\|^2} + C\xi$$

Balancing parameter

$$s.t. \quad \ell^{lol}(\mathbf{w}, \mathbf{u}_1, \ldots, \mathbf{u}_k; (\mathbf{x}_t, y_t)) \leq \xi, \;\; \xi \geq 0$$

Solving the problem

Define the following mapping for simplicity:

$$\tilde{\mathbf{x}}_t = \left[\frac{\mathbf{x}_t}{\sqrt{\lambda}}^T, \mathbf{0}^T, \mathbf{0}^T, ..., \mathbf{x}_t{}^T, ..., \mathbf{0}^T\right]^T$$

$$\tilde{\mathbf{w}} = \left[\sqrt{\lambda}\mathbf{w}^T, \mathbf{u}_1^T, \mathbf{u}_2^T, ..., \mathbf{u}_i^T, ..., \mathbf{u}_k^T\right]^T$$

$$\Rightarrow \tilde{\mathbf{w}}^T\tilde{\mathbf{x}}_t = (\mathbf{w}^T + \mathbf{u}_i^T)\mathbf{x}_t = \mathbf{w}_i^T\mathbf{x}_t$$

$$\tilde{\mathbf{w}}_{t+1} = \arg\min_{\tilde{\mathbf{w}}} \left(\frac{1}{2}\|\tilde{\mathbf{w}} - \tilde{\mathbf{w}}_t\|^2 + C\xi\right)$$

$$s.t. \quad \ell^{lol}(\tilde{\mathbf{w}}; (\tilde{\mathbf{x}}_t, y_t)) \leq \xi, \ \xi \geq 0.$$

$$\Rightarrow \quad \tilde{\mathbf{w}}_{t+1} = \tilde{\mathbf{w}}_t + \eta_t\tilde{\mathbf{x}}_t$$

$$\eta_t = \min\{C, \frac{\ell^{lol}}{\|\tilde{\mathbf{x}}_t\|^2}\}$$

∗i.e. PA online learning is a specific instance of our model, with No. of prototypes = 1

Assign a sample to a local hyperplane

- **Using clustering**
  - Our model adopts a sequential K-Means

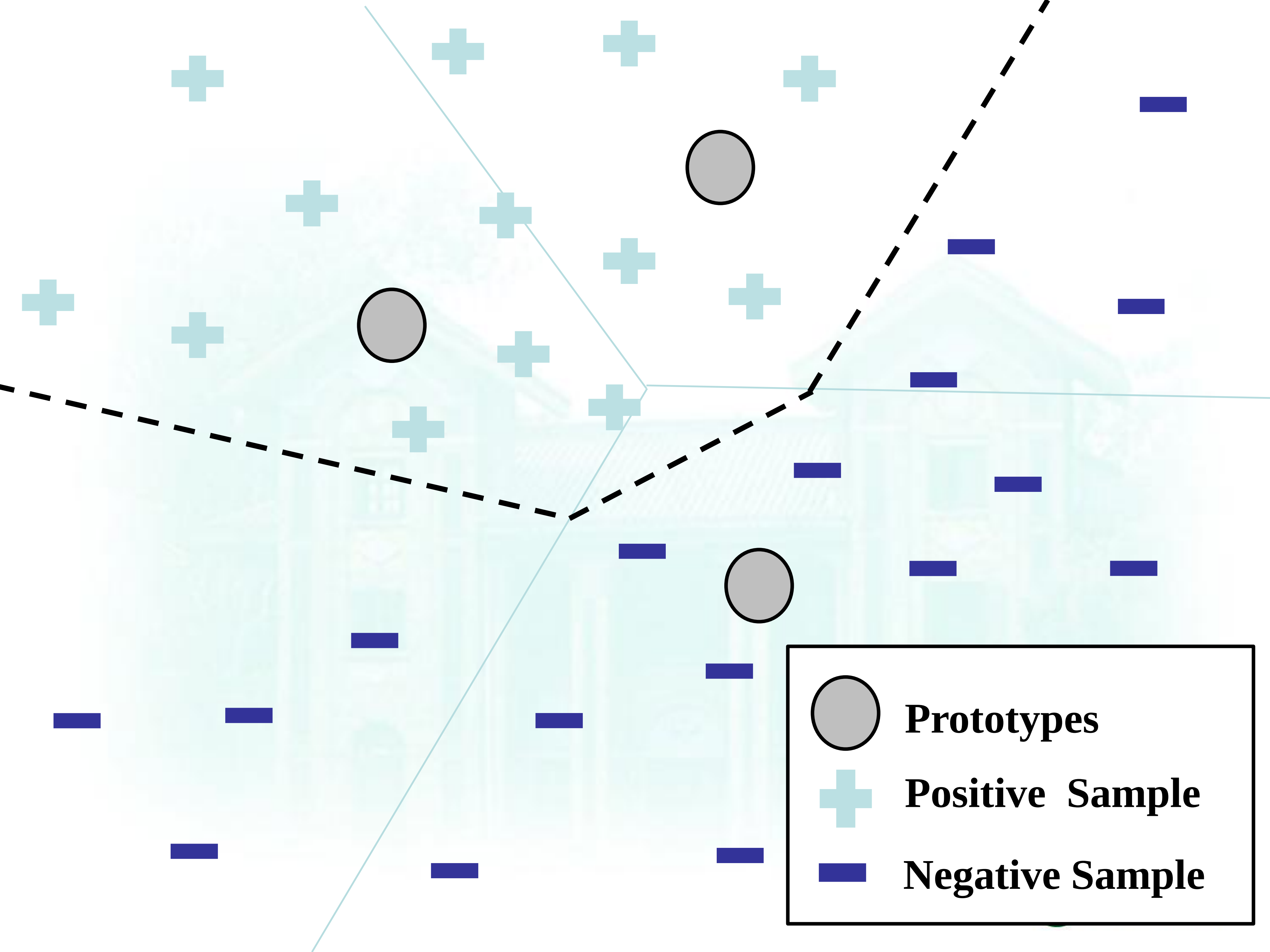
Pseudo code:

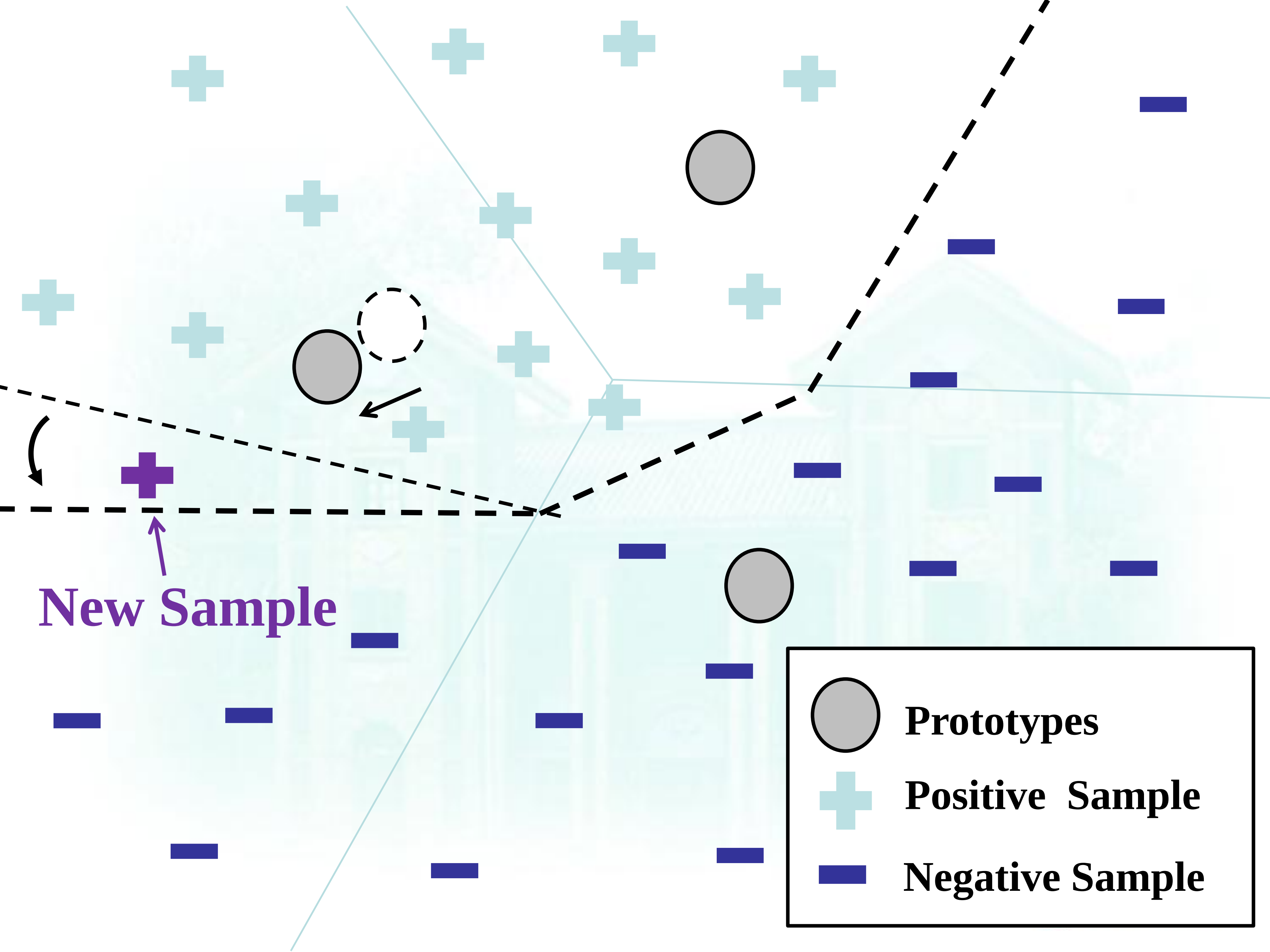
For t in [1, 2, …, T]

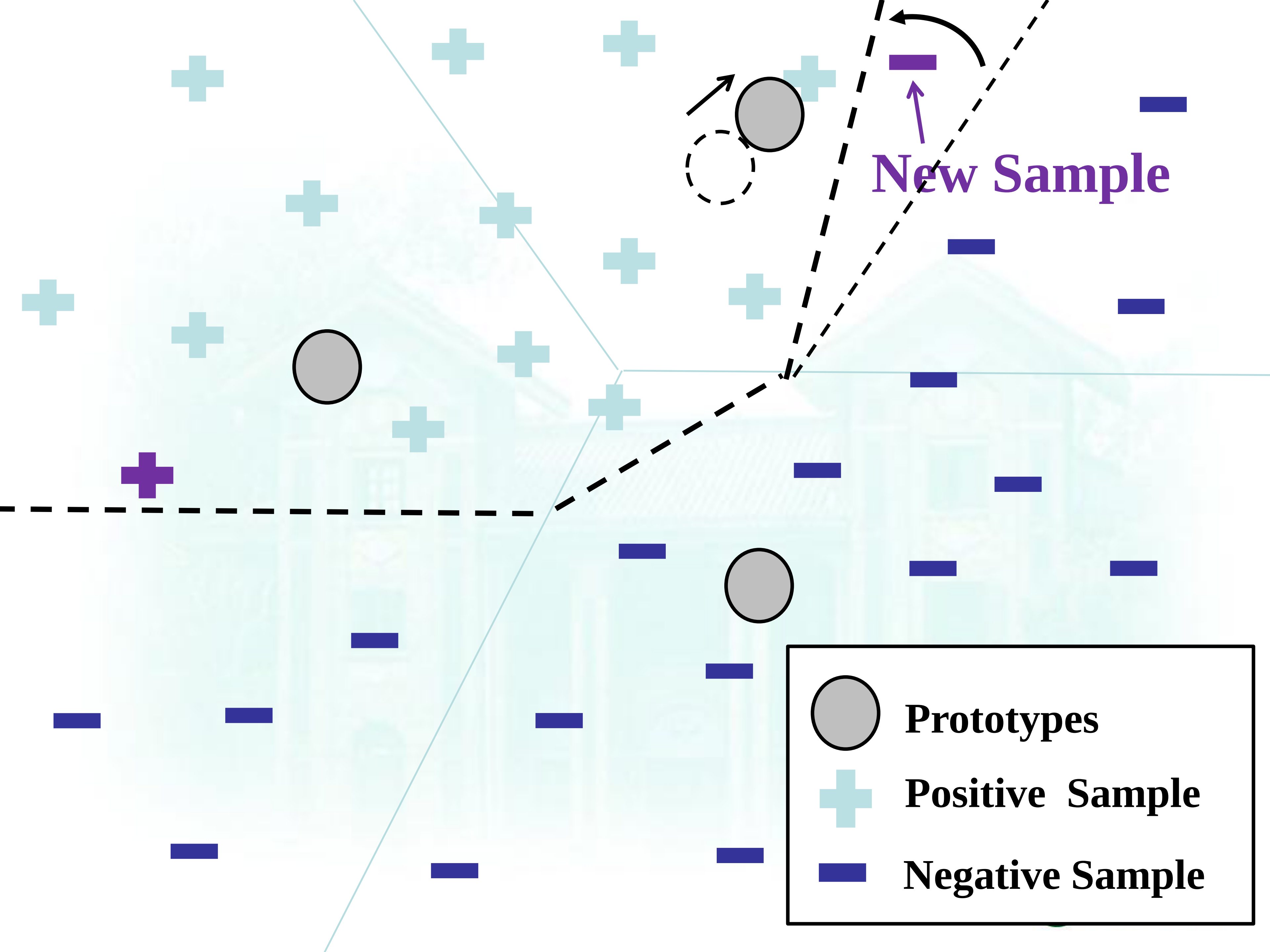
- Acquire current sample x_t
- If P_i is the nearest prototype (centroid) of sample x_t
  - $n_i += 1$
  - $P_i = P_i + \frac{1}{n_i}(x_t - P_i)$ // update the nearest prototype

Learning Process:

A Demonstration


Prototypes
Positive Sample
Negative Sample


New Sample
Prototypes
Positive Sample
Negative Sample


New Sample
Prototypes
Positive Sample
Negative Sample

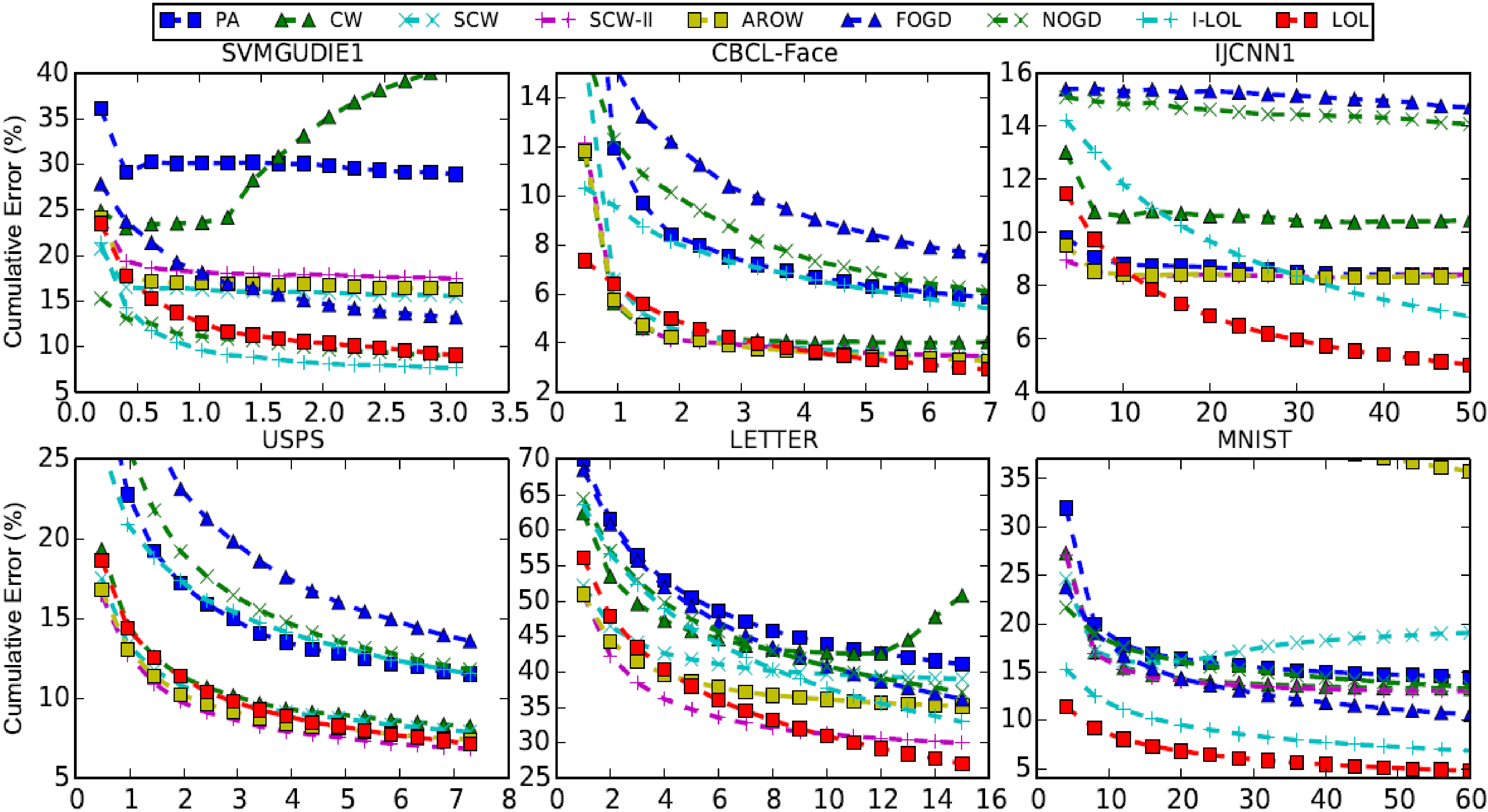
Theoretical Analysis

Cumulative loss of Passive Aggressive online learning.

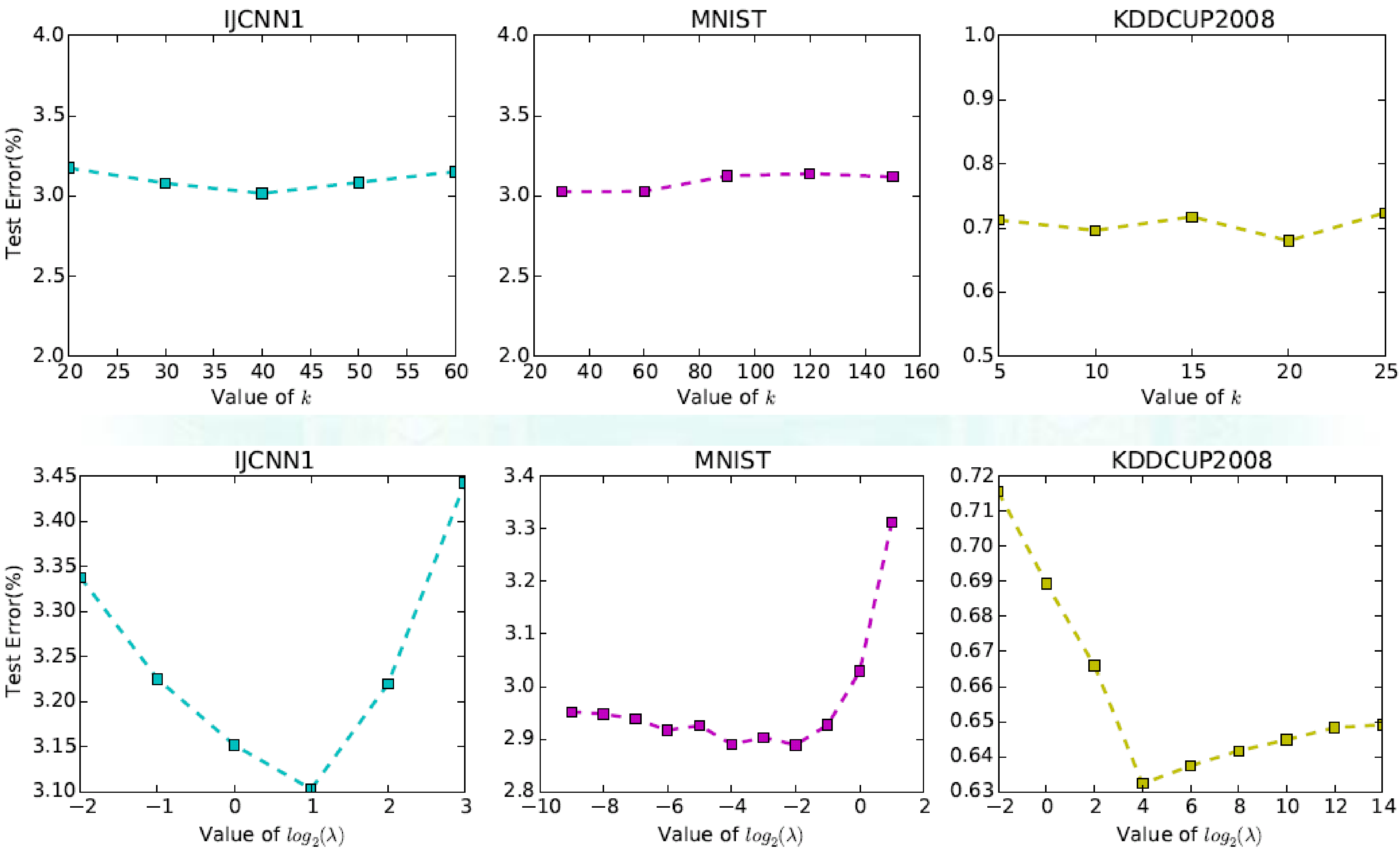
$$\sum_{t=1}^{T} \boxed{\left(\ell_t^{lol}\right)^2} \leq \frac{1}{2}\left(1+\frac{1}{\lambda}\right)R^2\left(\| \tilde{\mathbf{u}}^{lol} \|^2 + \sum_{t=1}^{T} \frac{\boxed{\left(\ell_t^{pa}\right)^2}}{\|\tilde{\mathbf{x}}_t\|^2}\right)$$

Cumulative loss of local online learning.

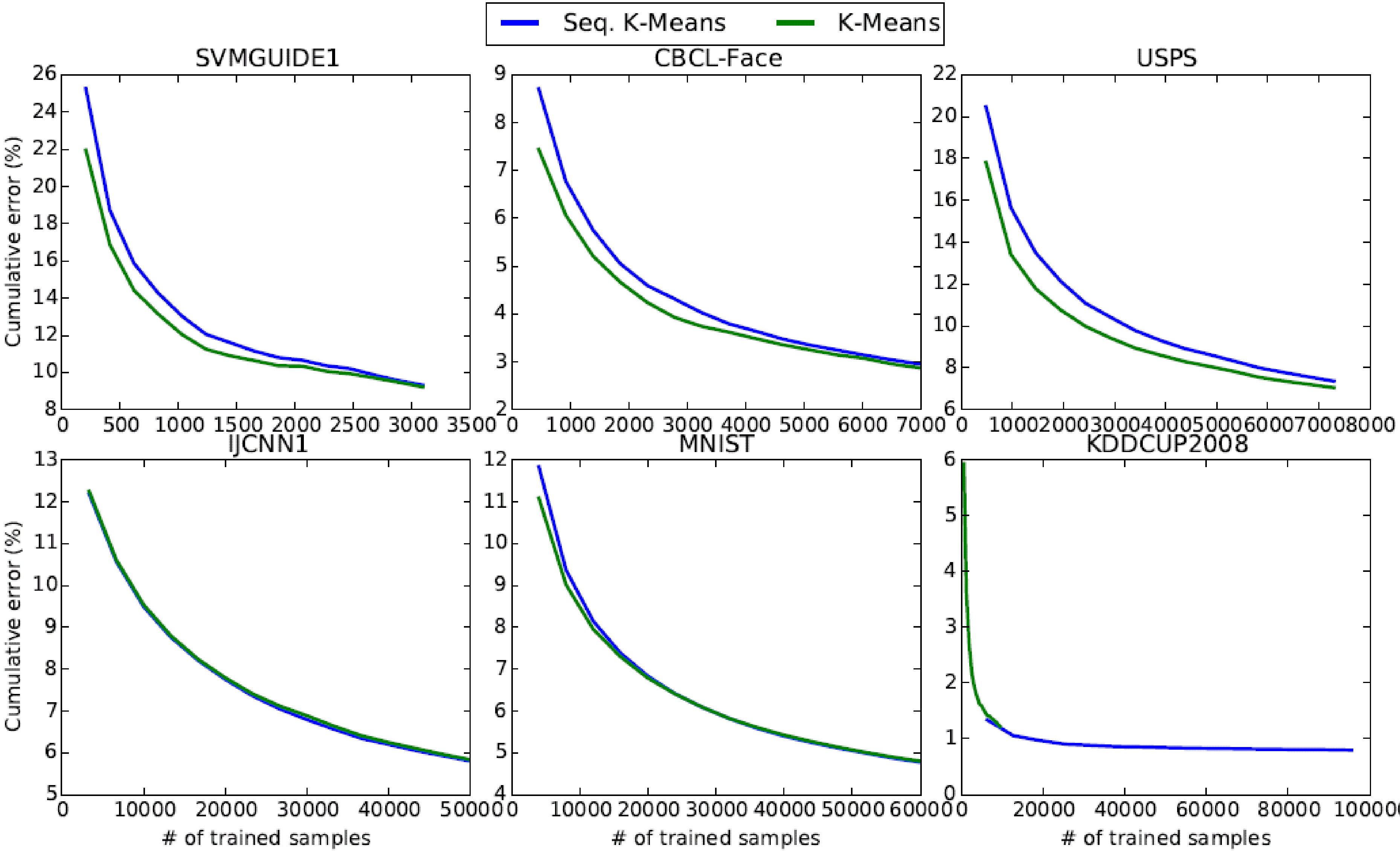
Empirical Result： Comparison of our model (in red) and the other methods



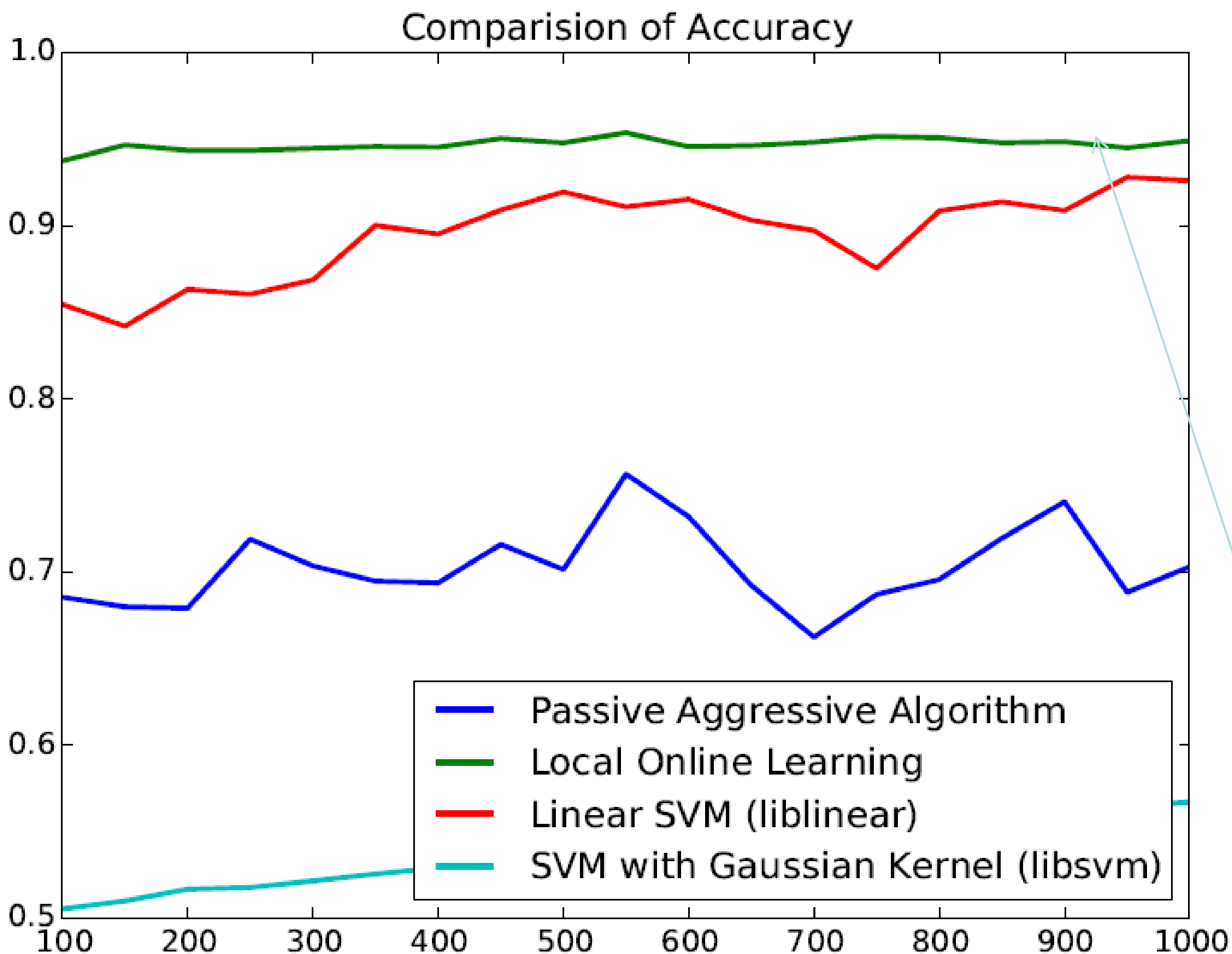
Empirical Result: Evaluation of parameters



Empirical Result: Evaluation of the effect of Sequential K-Means (in green)



Computation Speed and Accuracy on SVMGUIDE1



Application on Large Scale Face Recognition

Pose classification & Identity Recognition

- Multiple PIE: >140,000 face images

method	Multi-PIE Pose	Multi-PIE Identity
PA	4.22%±0.003	43.64%±0.003
Pegasos	15.01%±0.025	83.35%±0.013
CW	4.26%±0.001	31.17%±0.001
SCW-I	4.37%±0.000	78.43%±0.004
SCW-II	3.54%±0.000	29.04%±0.002
AROW	6.63%±0.001	31.49%±0.002
BSGD	8.99%±0.000	46.18%±0.102
FOGD	3.58%±0.004	18.17%±0.007
NOGD	10.78%±0.012	79.45%±0.014
I-LOL	3.48%±0.002	31.95%±0.009
LOL	**2.15%±0.001**	**4.62%±0.003**

Summary

1. **Tackle the nonlinear separation problem**
2. **Proposed a multi-hyperplane model with a shared component**
3. **Joint optimization on shared and local components.**
4. **Theoretical analysis for performance guarantee.**

Take Home Messages

- **Large Scale Person Identification involves multiple aspects of researches**
  - Group activity
  - Person re-identification across disjoint views
  - Large scale face/person image recognition

- **Large scale**
  - not only at the computational level
  - but also at the problem level (/vision level)

FOR MORE INFO.

VISIT MY HOME PAGE

http://sist.sysu.edu.cn/~zhwshi/

EMAIL ME：wszheng@ieee.org

中山大學
SUN YAT-SEN UNIVERSITY