# Machine Learning for Person Identification

**Wei-Shi Zheng**（郑伟诗）

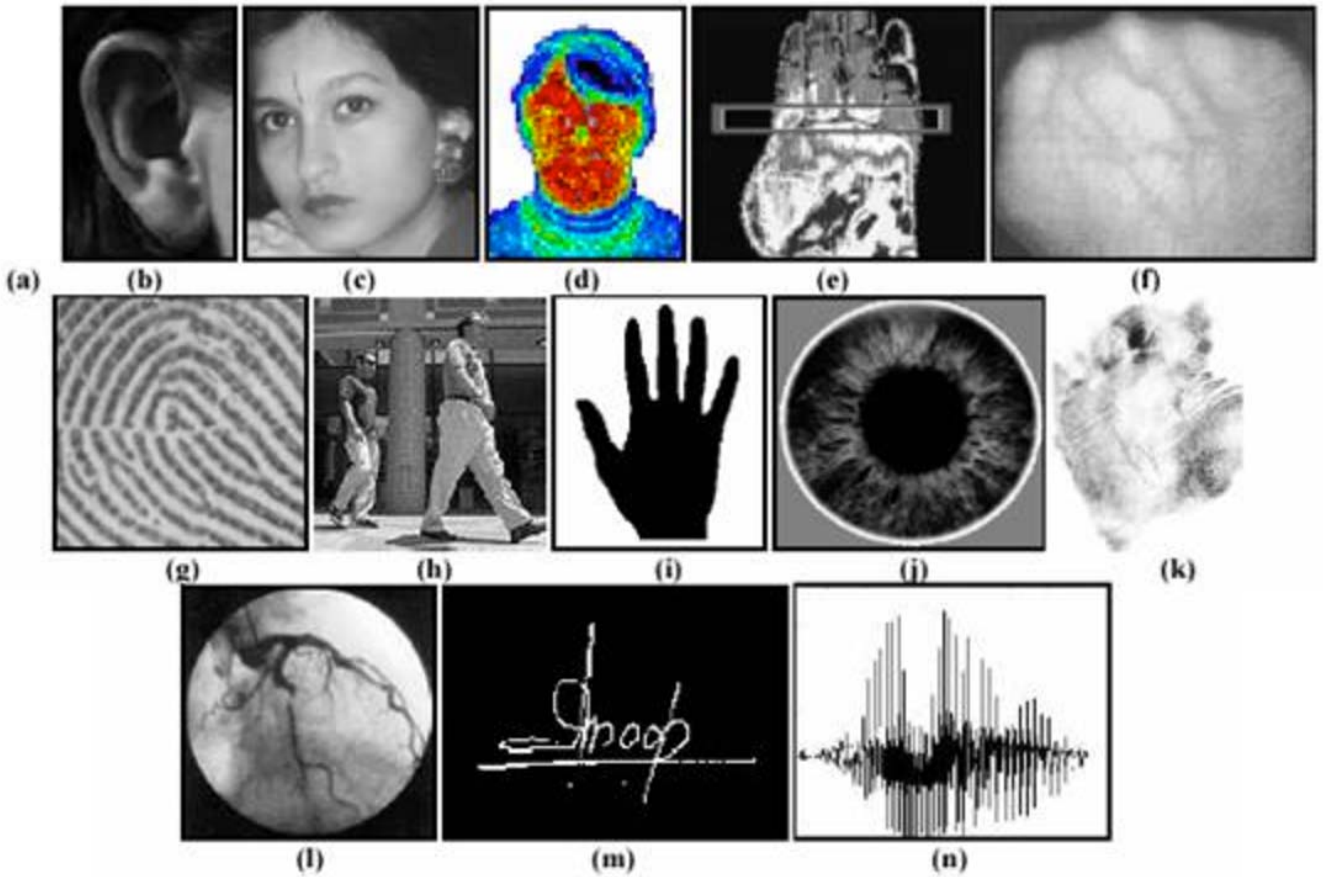
机器智能与先进计算
教育部重点实验室

# Outline

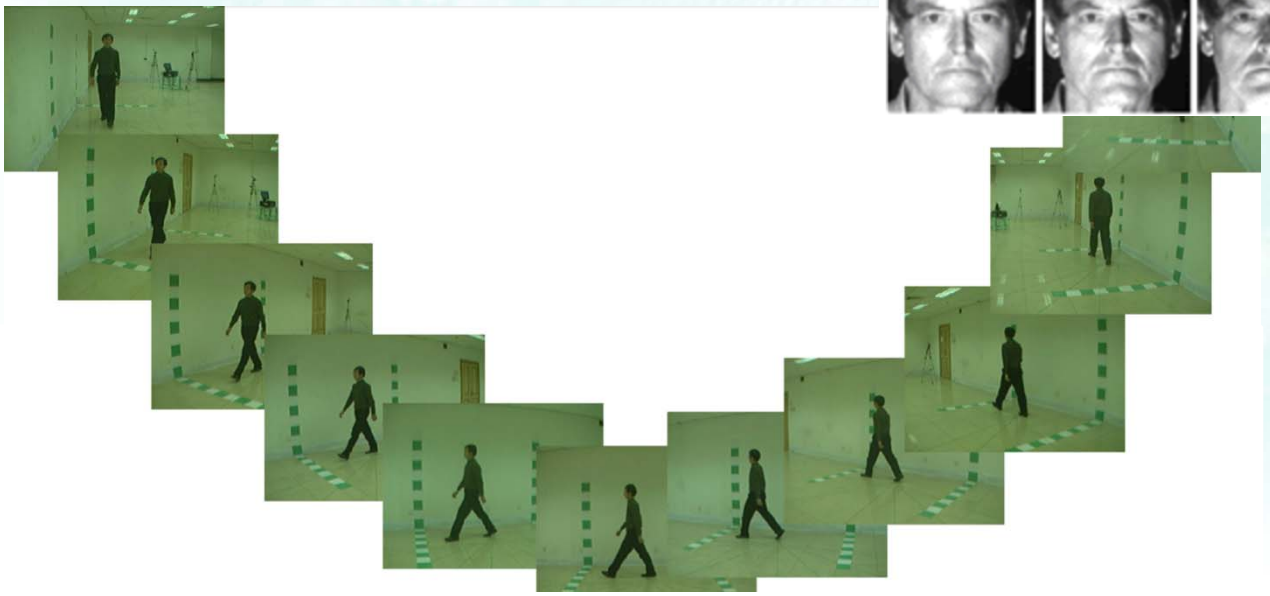
- **Brief Introduction of ML for Biometrics**
- **ML for Person Re-identification**
  - Distance Metric Learning
  - View Change Invariant Features
  - Partial Re-id
  - Low Resolution
  - Video-based Re-id
  - Cross Scenario Transfer
  - Open-world Modelling
  - Depth Re-identification
- **Summary**

# A BRIEF INTRODUCTION ON MACHINE LEARNING FOR PERSON IDENTIFICATION

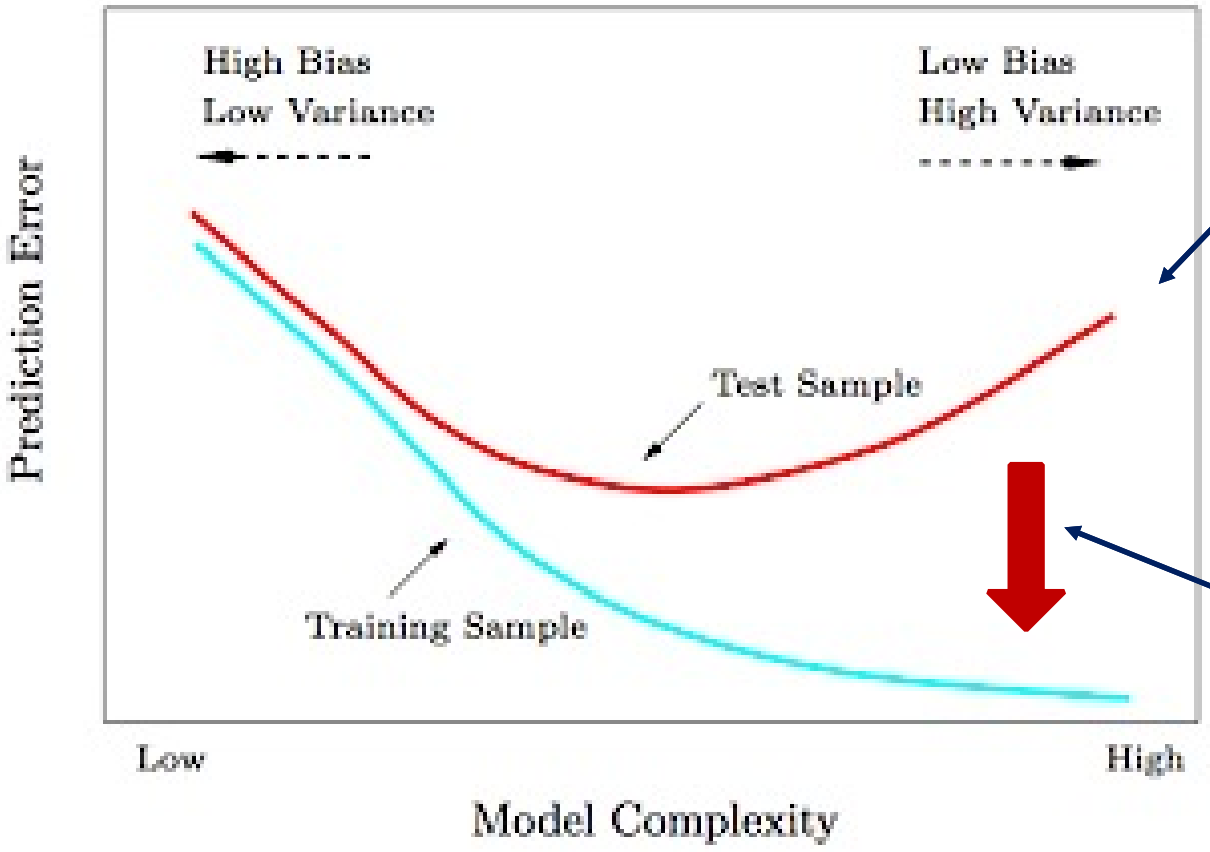
# Biometrics



# Why Machine Learning is Needed

# Why Machine Learning is Needed




Small sample size

Large-scale sample size

# Preprocessing

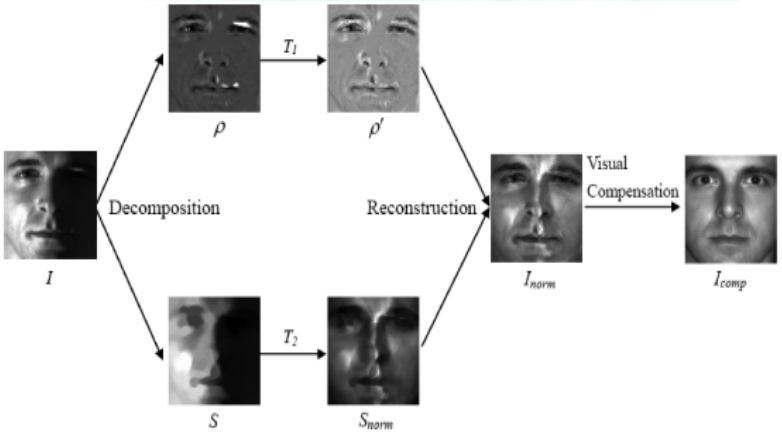
- Propose a two-step framework
- Propose a weakly supervised penalty: guide the learning

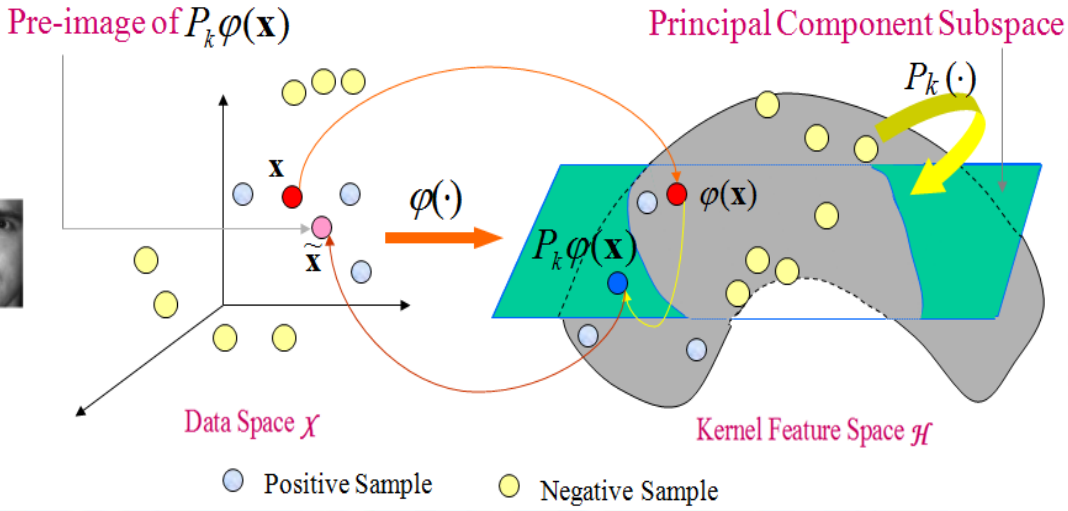
$$\hat{\mathbf{x}} = \boldsymbol{\Psi}_{s,\mathbf{x}}\mathbf{w}_{\mathbf{x}} = \sum_{i=1}^{s} w_i^{\mathbf{x}}\tilde{\mathbf{x}}_i$$

weakly supervised penalty

$$\mathbf{w}_{\mathbf{x}} = \underset{\mathbf{w}=(w_1,\ldots,w_s)^T\in\Re^s}{\arg\min} \left\{ \mathbf{w}^T\boldsymbol{\Psi}_{s,\varphi(\mathbf{x})}^T\boldsymbol{\Psi}_{s,\varphi(\mathbf{x})}\mathbf{w} - 2(\boldsymbol{\Psi}_{\varphi}\boldsymbol{\gamma}^{\mathbf{x}})^T\boldsymbol{\Psi}_{s,\varphi(\mathbf{x})}\mathbf{w} + \lambda\cdot F(\mathbf{w}) \right\}$$

$$\text{s.t.} \sum_{i=1}^{s} w_i^{\mathbf{x}} = 1 \quad \text{and} \quad w_i^{\mathbf{x}} \geq 0, \qquad i = 1,\ldots,s$$



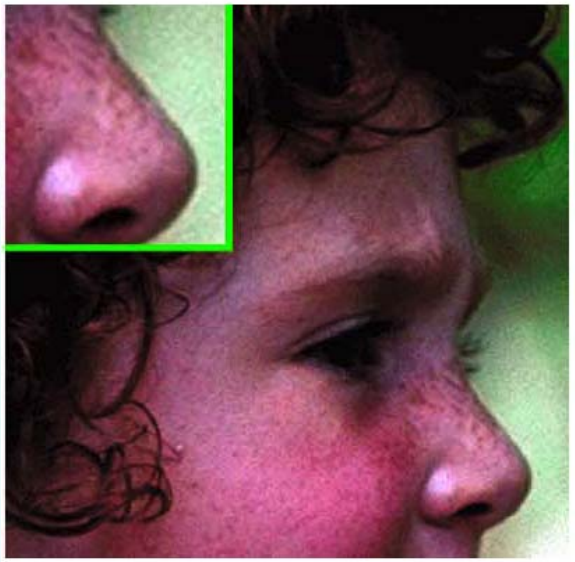


- J. T. Kwok and I. W. Tsang, "The pre-image problem in kernel methods," IEEE Trans. Neural Netw., vol. 15, no. 6, pp. 1517–1525, Nov. 2004.
- Wei-Shi Zheng, JianHuang Lai, and Pong C. Yuen, "Penalized Pre-image Learning in Kernel Principal Component Analysis," IEEE Trans. on Neural Networks, vol. 21, no. 4, pp. 551-570, 2010.

# Super-resolution

- **Sparse Coding**

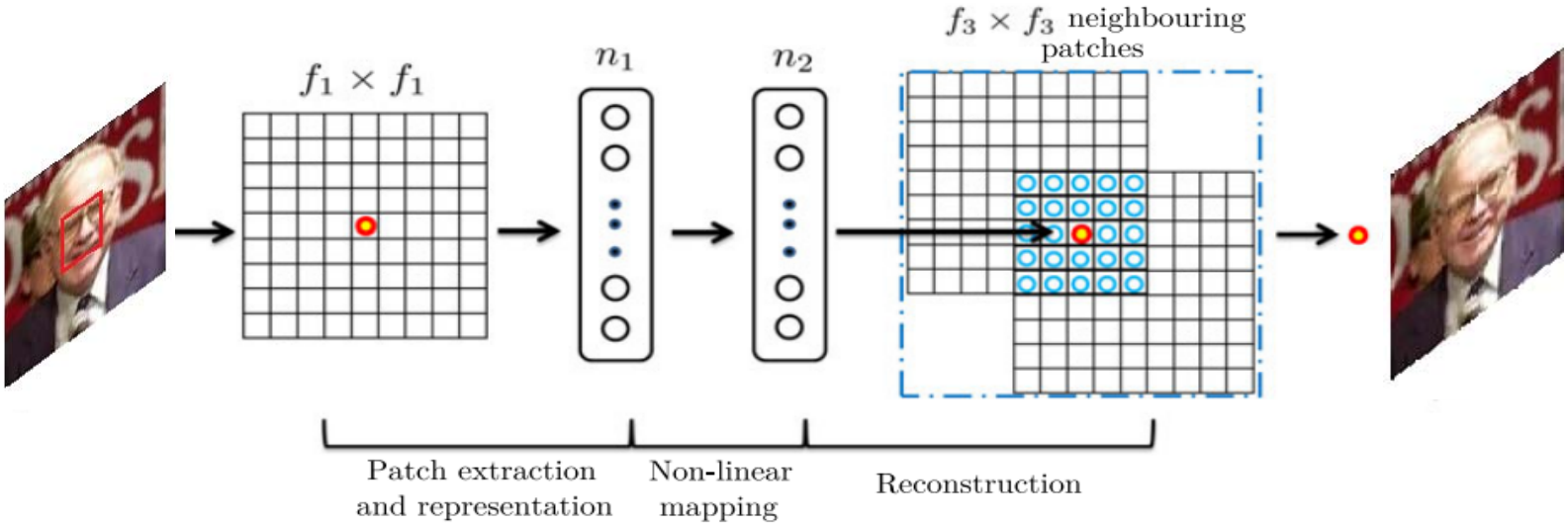
$$\boldsymbol{X}^* = \arg\min_{\boldsymbol{X}} \|SH\boldsymbol{X} - \boldsymbol{Y}\|_2^2 + c\|\boldsymbol{X} - \boldsymbol{X}_0\|_2^2$$



Jianchao Yang et al. Image Super-Resolution Via Sparse Representation. IEEE Trans. on Image Processing, 2010.

# Super-resolution

- **Deep Processing**




Patch Extraction and Representation $F_1(\mathbf{Y}) = \max(0, W_1 * \mathbf{Y} + B_1),$

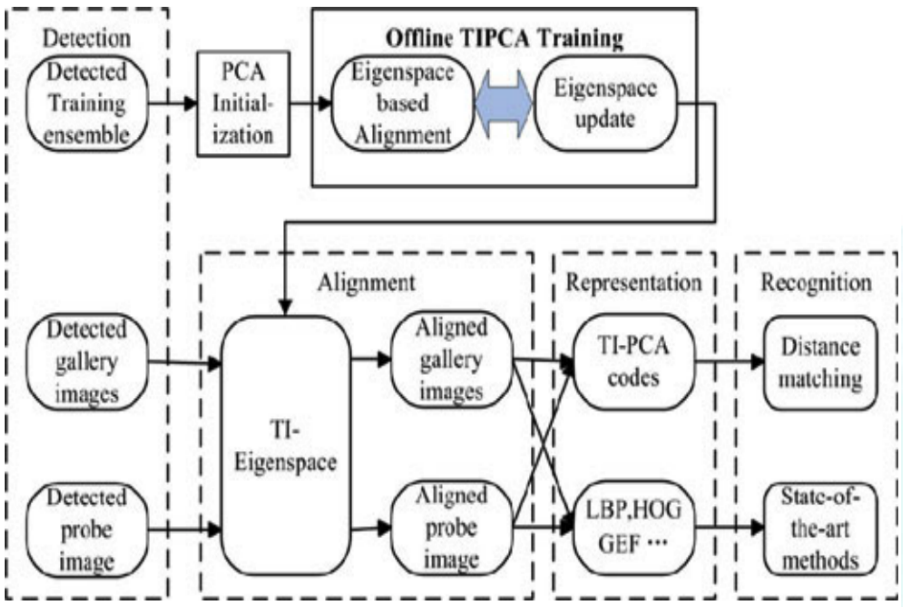
Non-Linear Mapping $F_2(\mathbf{Y}) = \max(0, W_2 * F_1(\mathbf{Y}) + B_2).$

Reconstruction $F(\mathbf{Y}) = W_3 * F_2(\mathbf{Y}) + B_3.$

**Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Image Super-Resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.**

# Alignment

## PCA Alignment




$$I(\mathbf{W}(\mathbf{x};\mathbf{p})) = \mu(\mathbf{x}) + \sum_{j=1}^{m} a_j \phi_j(\mathbf{x}) + e(\mathbf{x}),$$

$$\arg\min_{\mu,\,\phi_j} \frac{1}{N} \sum_{i=1}^{N} \left\{ \min_{\mathbf{p}^i,\,\mathbf{a}^i} \sum_{\mathbf{x}} \left[ e^i(\mathbf{x}) \right]^2 \right\}, \text{ where}$$

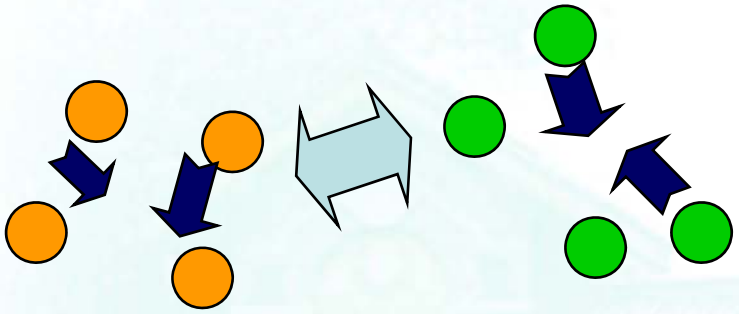
$$e^i(\mathbf{x}) = I^i(\mathbf{W}(\mathbf{x};\mathbf{p}^i)) - \left[ \mu(\mathbf{x}) + \sum_{j=1}^{m} a_j^i \phi_j(\mathbf{x}) \right].$$

$$\min_{\mathbf{p},\,\mathbf{a}} \sum_{\mathbf{x}} \left\{ I(\mathbf{W}(\mathbf{x};\mathbf{p})) - \left[ \mu(\mathbf{x}) + \sum_{j=1}^{m} a_j \phi_j(\mathbf{x}) \right] \right\}^2.$$

**Weihong Deng, Jiani Hu, Jiwen Lu, Jun Guo. Transform-Invariant PCA: A Unified Approach to Fully Automatic FaceAlignment, Representation, and Recognition. IEEE TPAMI, 2014.**

# Feature Extraction

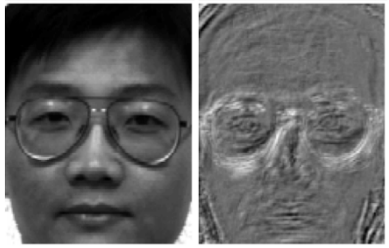
- **Subspace Learning**



Class 1

Class 2

Fisherface



$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{trace(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{trace(\mathbf{W}^T \mathbf{S}_w \mathbf{W})}. \quad \Leftrightarrow \quad \mathbf{S}_w^{-1}\mathbf{S}_b\mathbf{W} = \mathbf{W}\Lambda$$

Between-class covariance matrix →

$$\mathbf{S}_b = \frac{1}{N}\sum_{k=1}^{L} N_k(\mathbf{u}_k - \mathbf{u})(\mathbf{u}_k - \mathbf{u})^T = \Phi_b \Phi_b^T.$$

Within-class covariance matrix →

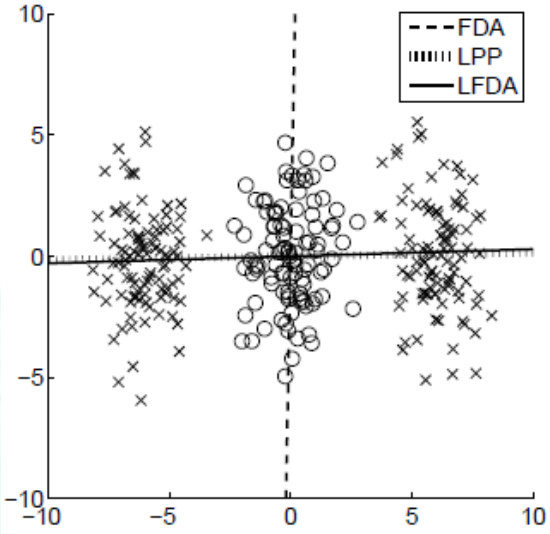
$$\mathbf{S}_w = \frac{1}{N}\sum_{k=1}^{L}\sum_{i=1}^{N_k}(\mathbf{x}_i^k - \mathbf{u}_k)(\mathbf{x}_i^k - \mathbf{u}_k)^T = \Phi_w \Phi_w^T,$$

**P.N. Belhumeur, J. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.**

# Feature Extraction

- **LFDA**

Distribution of pedestrian features is multi-modal.




$$\overline{\boldsymbol{S}}^{(w)} = \frac{1}{2}\sum_{i,j=1}^{n} \overline{\boldsymbol{A}}_{i,j}^{(w)} (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top, \qquad \overline{\boldsymbol{A}}_{i,j}^{(w)} = \begin{cases} \boldsymbol{A}_{i,j}/n_c & \text{if } y_i = y_j = c, \\ 0 & \text{if } y_i \neq y_j, \end{cases}$$

$$\overline{\boldsymbol{S}}^{(b)} = \frac{1}{2}\sum_{i,j=1}^{n} \overline{\boldsymbol{A}}_{i,j}^{(b)} (\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top, \qquad \overline{\boldsymbol{A}}_{i,j}^{(b)} = \begin{cases} \boldsymbol{A}_{i,j}(1/n - 1/n_c) & \text{if } y_i = y_j = c, \\ 1/n & \text{if } y_i \neq y_j. \end{cases}$$

**S. Pedagadi, J. Orwell, S. Velastin and B. Boghossian, Local Fisher Discriminant Analysis for Pedestrian Re-identification, 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3318-3325.**

# Feature Extraction

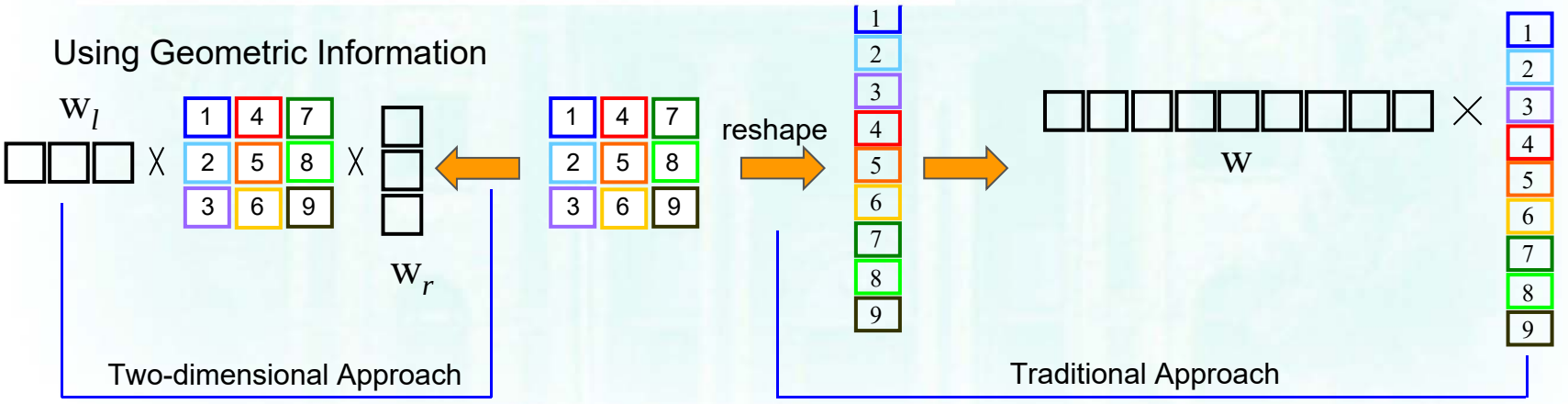
- **Subspace Learning**

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{trace(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{trace(\mathbf{W}^T \mathbf{S}_w \mathbf{W})}.$$

$$\mathbf{S}_b = \frac{1}{N}\sum_{k=1}^{L} N_k(\mathbf{u}_k - \mathbf{u})(\mathbf{u}_k - \mathbf{u})^T = \Phi_b \Phi_b^T.$$

$$\mathbf{S}_w = \frac{1}{N}\sum_{k=1}^{L}\sum_{i=1}^{N_k}(\mathbf{x}_i^k - \mathbf{u}_k)(\mathbf{x}_i^k - \mathbf{u}_k)^T = \Phi_w \Phi_w^T,$$

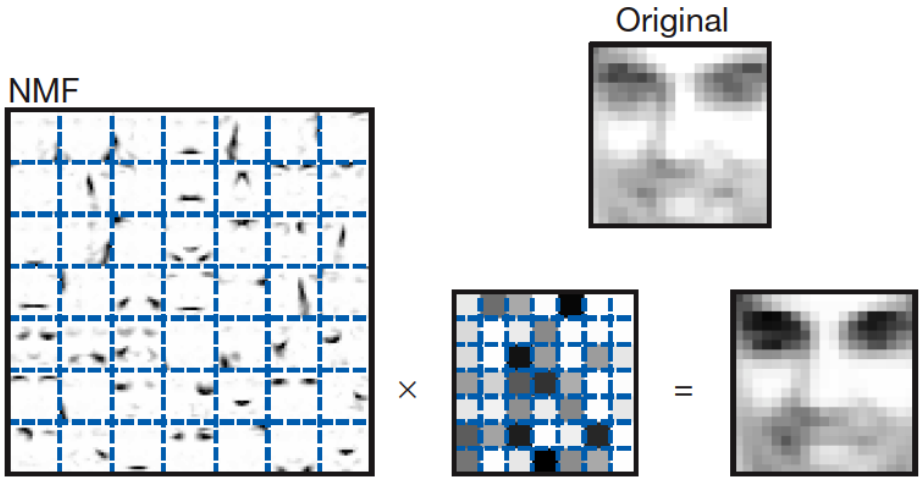
2D-LDA could lose the cross-covariance information between rows or columns




**Wei-Shi Zheng, JianHuang Lai, and Stan Z. Li, "1D-LDA versus 2D-LDA: When Is Vector-based Linear Discriminant Analysis Better than Matrix-based?" Pattern Recognition, vol. 41, no. 7, pp. 2156-2172, 2008.**

# Feature Extraction

- **Non-negativity Matrix Factorization**




$$\min_{\mathbf{W} \in \mathbb{R}^{d \times L}, \mathbf{h}_i \in \mathbb{R}^L} \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{W}\mathbf{h}_i\|_F^2$$

$$\text{s.t.} \quad \mathbf{W} \geq \mathbf{0}, \ \mathbf{h}_i \geq \mathbf{0}.$$

D.D. Lee, H.S. Seung. Learning the parts of objects by non-negative matrix factorization, Nature, 1999.

# Feature Extraction

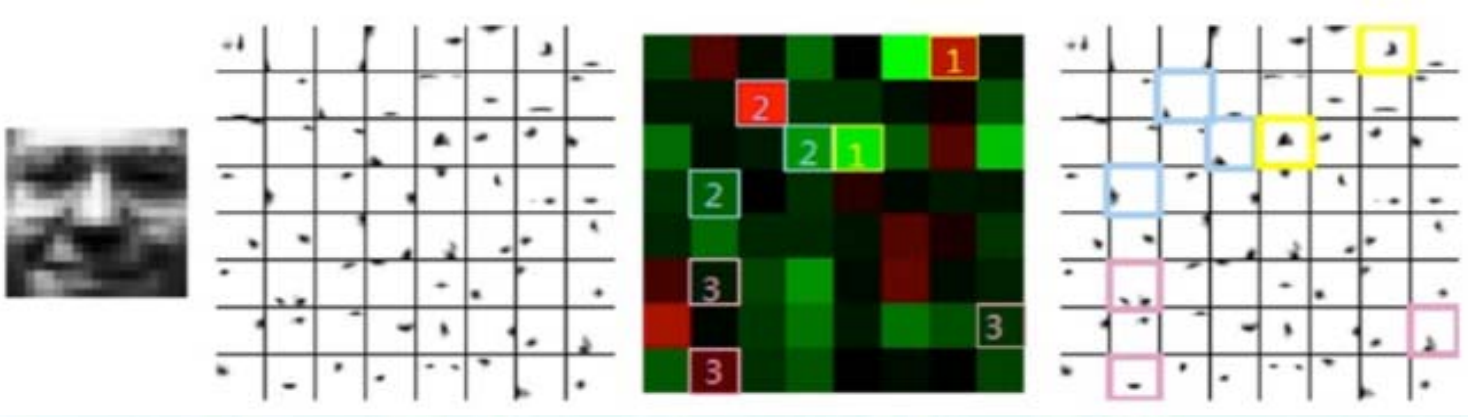
## Pixel Dispersion Penalty

$$\min_{\mathbf{W}\in\mathbb{R}^{d\times L},\mathbf{h}_i\in\mathbb{R}^L} f_{\mathbf{E}_l}(\mathbf{W},\mathbf{H}) = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{x}_i-\mathbf{W}\mathbf{h}_i\|_F^2+\frac{\lambda}{L}\,trace(\mathbf{W}^T\mathbf{E}_l\mathbf{W}),\ \lambda\geq 0.$$

$$\text{s.t.}\quad \mathbf{W}\geq\mathbf{0},\ \left|\mathbf{h}_i(j)\right|\leq c_0,$$

Pixel Dispersion Penalty

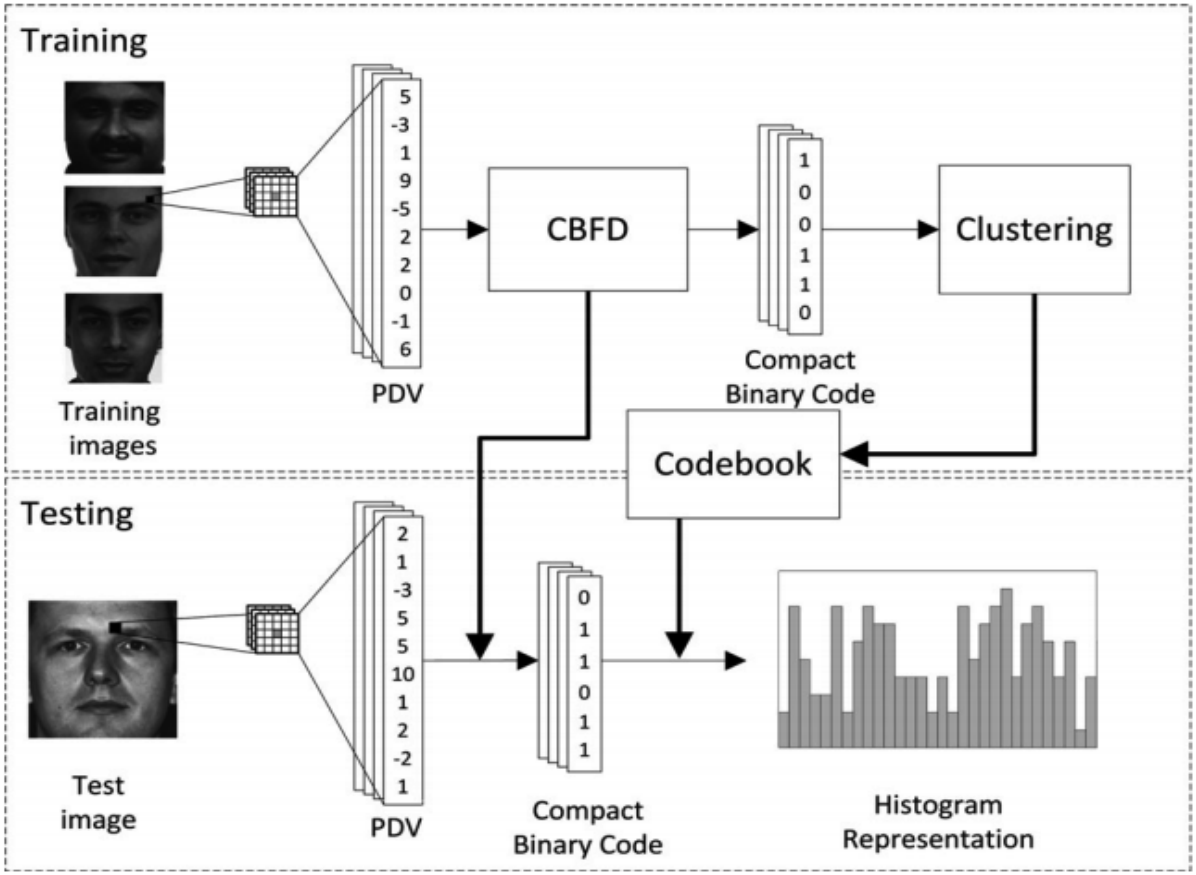
$$D(\mathbf{w}_i)=\sum_{x=1}^{a}\sum_{y=1}^{b}\left|\mathbf{w}_i^{2D}(y,x)\right|\times\left\{\sum_{x'=1}^{a}\sum_{y'=1}^{b}\left|\mathbf{w}_i^{2D}(y',x')\right|d_{y,x}(y'-y,x'-x)\right\}$$




Wei-Shi Zheng, JianHuang Lai, Shengcai Liao, and Ran He, "Extracting Non-negative Basis Images Using Pixel Dispersion Penalty," Pattern Recognition, vol. 45, no. 8, pp. 2912-2926, 2012.

# Feature Extraction

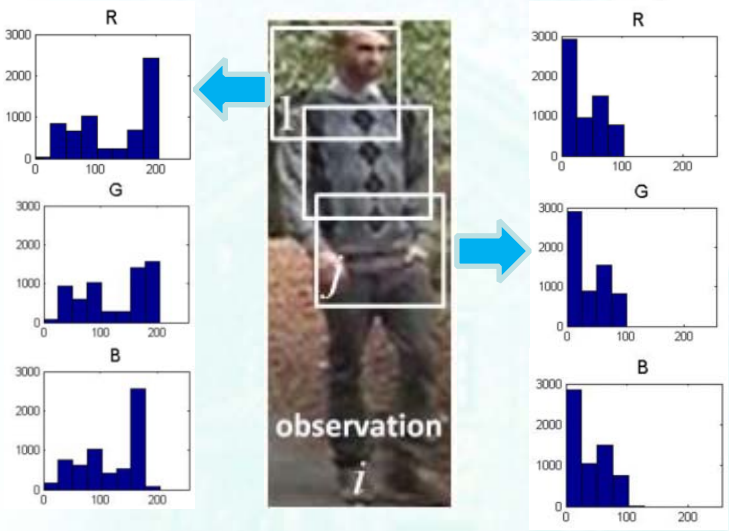
- **Binary Coding**



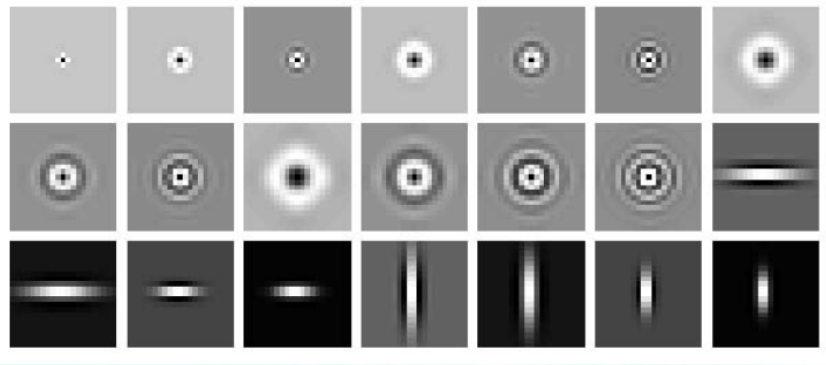

Jiwen Lu et al., Learning Compact Binary Face Descriptor for Face Recognition. IEEE TPAMI, 2015.

# Feature Selection

- Ensemble of localized features




Color histograms



Textures

Ensemble of localized features exploit color histograms and textures.

**D. Gray, H. Tao. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. European Conference on Computer Vision (ECCV), 2008**

# Feature Selection

- **Floatboost**

(Conditional Exclusion)

(1) $h' = \arg\min_{h \in \mathcal{H}_M} \varepsilon(H_M - h)$;

(2) If $\varepsilon(H_M - h') < \varepsilon^{\min}_{M-1}$, then

(a) $\mathcal{H}_{M-1} = \mathcal{H}_M - h'$;

$\varepsilon^{\min}_{M-1} = \varepsilon(H_M - h')$; $M = M - 1$;

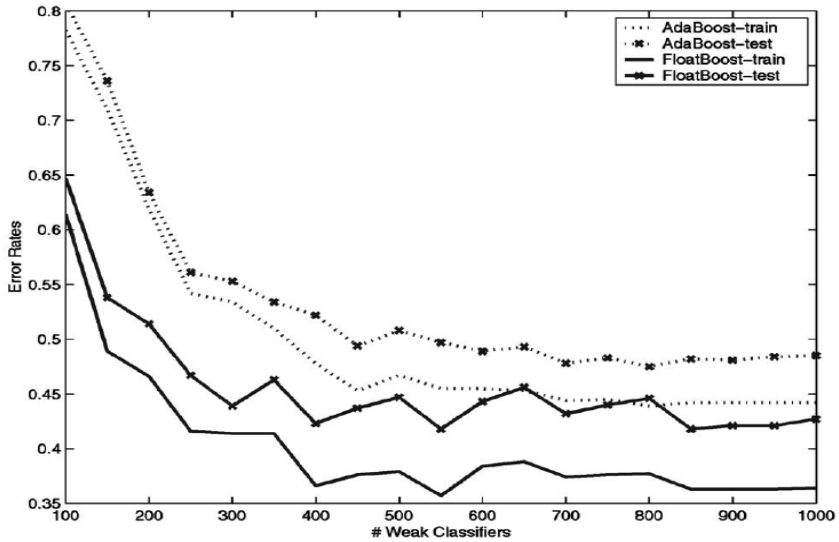
(b) $H_M = \sum_{h \in \mathcal{H}_M} h$;

(c) goto 3.(1);

(3) else

(a) if $M = M_{\max}$ or $J(\mathcal{H}_M) < J^*$, then goto 4;

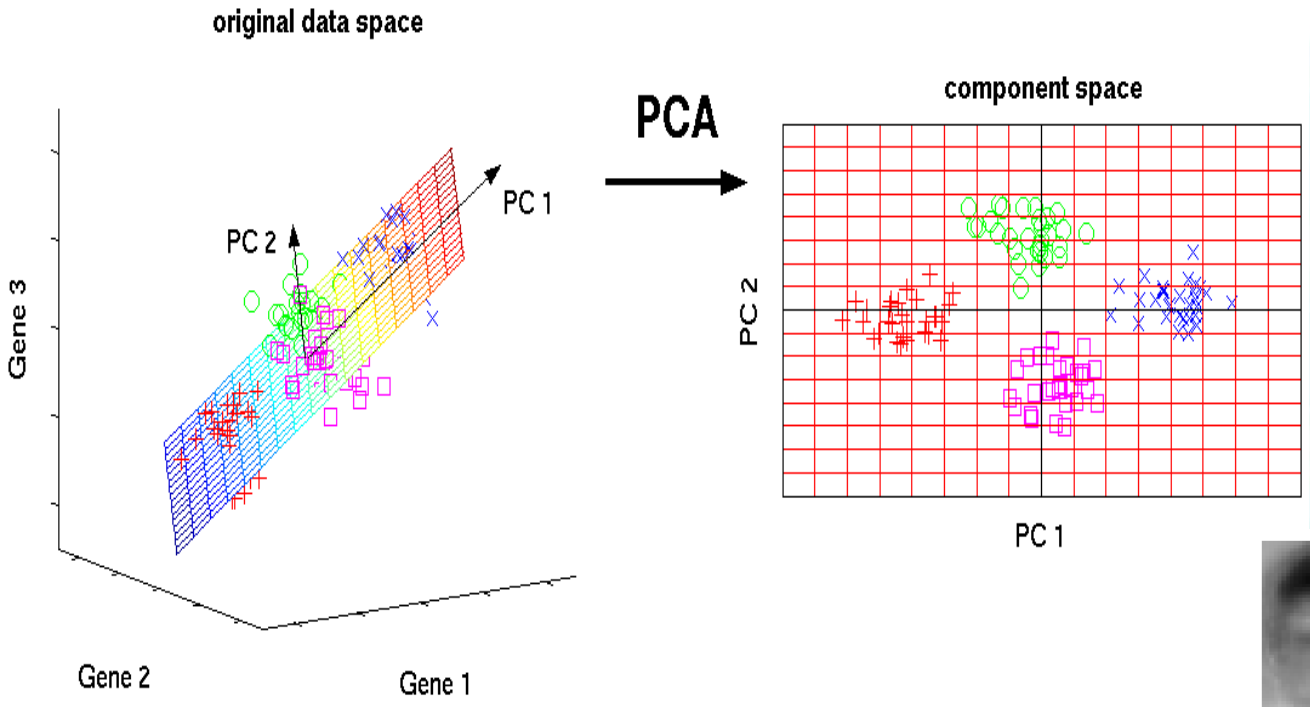
(b) $w_i^{(M)} \leftarrow \exp[-y_i H_M(x_i)]$; goto 2.(1);

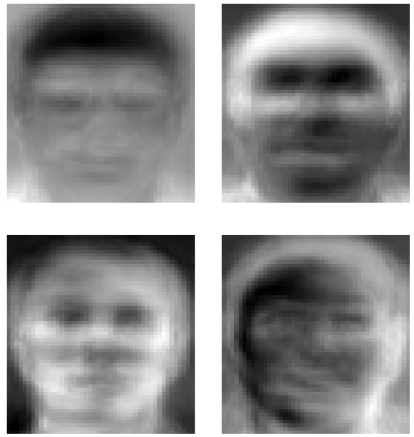



Stan Li et al. FloatBoost Learning and Statistical Face Detection. IEEE TPAMI, 2004.

# Dimension Reduction

- **Subspace Learning**



Turk, Matthew A and Pentland, Alex P. *Face recognition using eigenfaces*. IEEE CVPR, 1991.

# Classification

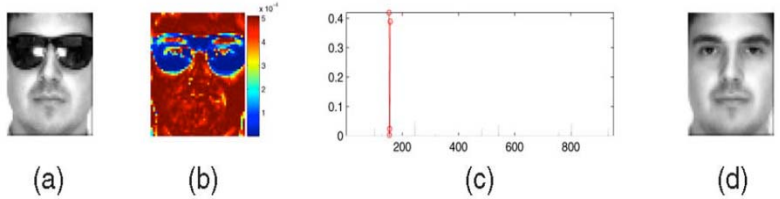
## Sparse Representation-based Classifier

SRC：J. Wright et al.

$$\min \|\beta\|_0 \quad s.t. \quad y = X\beta,$$

Non-negativity：He et al.

$$J_{CESR} = \max_{\beta} \sum_{j=1}^{m} g\left(y_j - \sum_{i=1}^{n} x_{ij}\beta_i\right) - \lambda \sum_{i=1}^{n} \beta_i \quad s.t.\ \beta_i \geq 0.$$



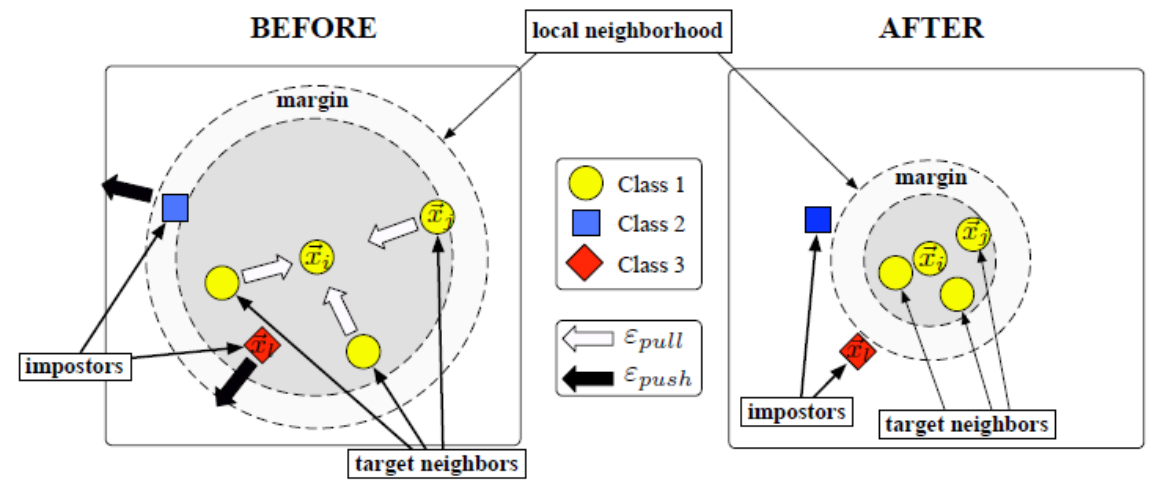


- J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210-227, Feb. 2009.

- Ran He, Wei-Shi Zheng, and BaoGang Hu. Maximum Correntropy Criterion for Robust Face Recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1561 - 1576, 2011.

# Classification

## Distance Metric Learning




Similarity is measured by Mahalanobis distance

$$\mathcal{D}_{\mathbf{M}}(\vec{x}_i,\vec{x}_j)=(\vec{x}_i-\vec{x}_j)^{\top}\mathbf{M}(\vec{x}_i-\vec{x}_j)$$

$$\varepsilon(\mathbf{M})=(1-\mu)\sum_{i,j\rightsquigarrow i}\mathcal{D}_{\mathbf{M}}(\vec{x}_i,\vec{x}_j)+\mu\sum_{i,j\rightsquigarrow i}\sum_{l}(1-y_{il})\left[1+\mathcal{D}_{\mathbf{M}}(\vec{x}_i,\vec{x}_j)-\mathcal{D}_{\mathbf{M}}(\vec{x}_i,\vec{x}_l)\right]_+.$$

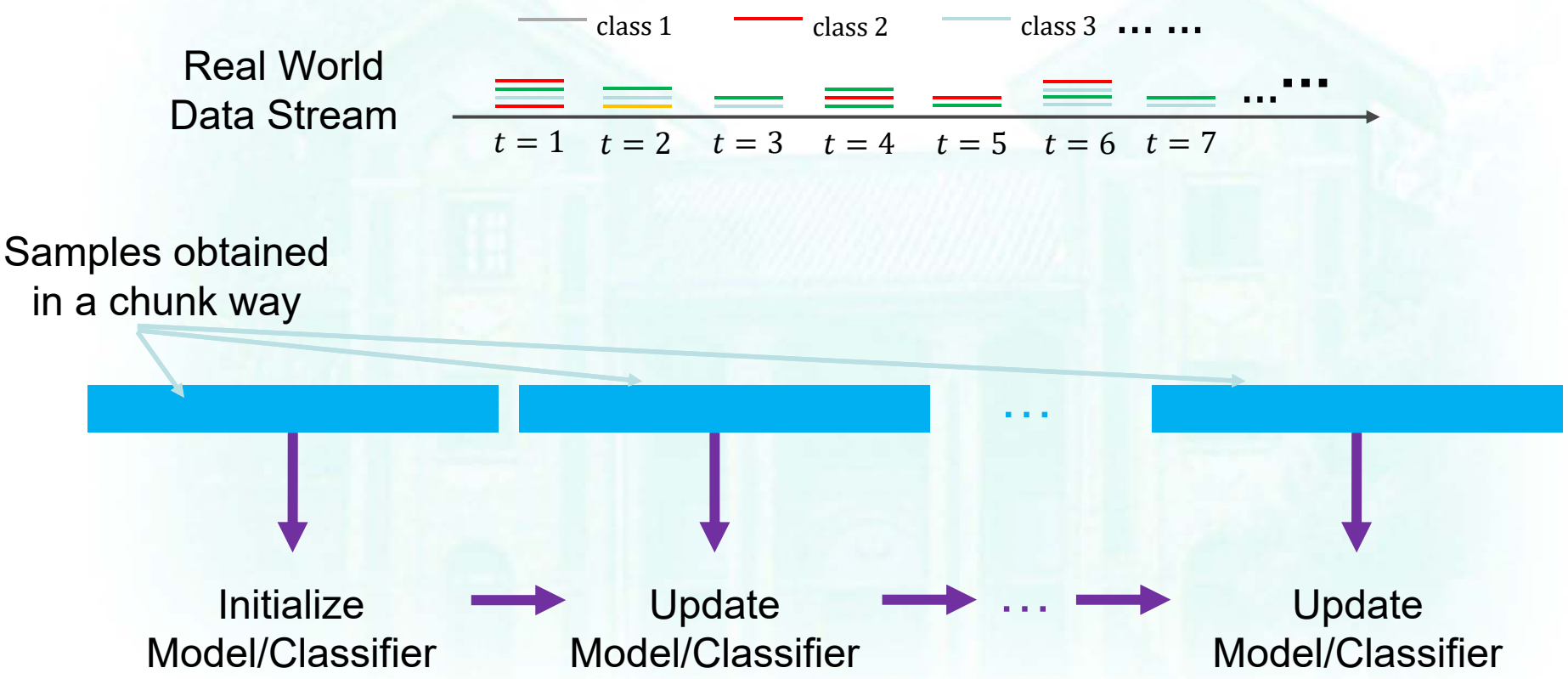
Pulling positive pairs by minimizing intra-class distances

Pushing negative pairs by enlarging inter-class distances

**Weinberger, Kilian Q., and Lawrence K. Saul. "Distance metric learning for large margin nearest neighbor classification." Journal of Machine Learning Research, 2009**

# Online Learning

- **Incremental Learning**

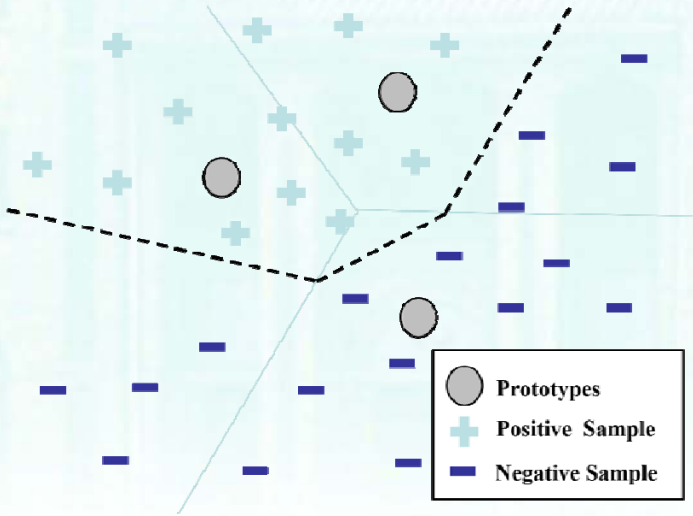


# Online Learning

- **One-pass Learning**

$$\mathbf{w}_{t+1}, \mathbf{u}_{1,t+1}, \ldots, \mathbf{u}_{k,t+1} = \arg\min_{\mathbf{w},\mathbf{u}_1,\ldots,\mathbf{u}_k} \frac{\lambda}{2}\|\mathbf{w}-\mathbf{w}_t\|^2 + \frac{1}{2}\sum_{i=1}^{k}\|\mathbf{u}_i-\mathbf{u}_{i,t}\|^2 + C\xi \quad \text{s.t. } \ell^{lol}(\mathbf{w},\mathbf{u}_1,\ldots,\mathbf{u}_k;(\mathbf{x}_t,y_t)) \le \xi,\ \xi \ge 0$$

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



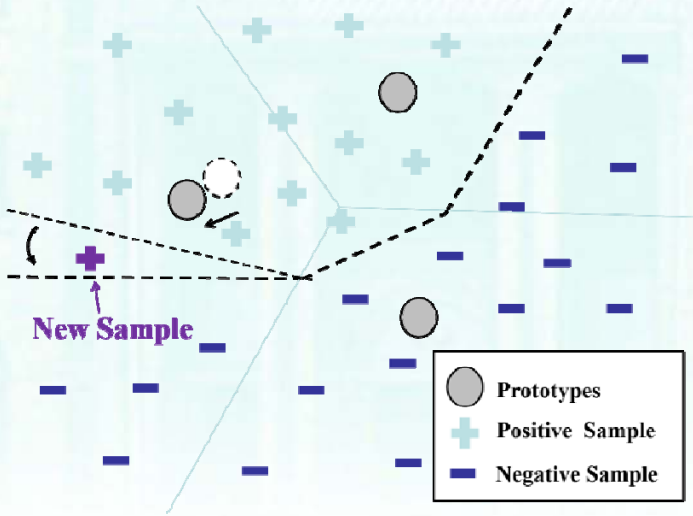

Zhaoze Zhou, Wei-Shi Zheng, et al. One-pass online learning: A local approach. Pattern Recognition, 2016.

# Online Learning

- **One-pass Learning**

$$\mathbf{w}_{t+1}, \mathbf{u}_{1,t+1}, \ldots, \mathbf{u}_{k,t+1} = \arg\min_{\mathbf{w},\mathbf{u}_1,\ldots,\mathbf{u}_k} \frac{\lambda}{2} \| \mathbf{w} - \mathbf{w}_t \|^2 + \frac{1}{2} \sum_{i=1}^{k} \| \mathbf{u}_i - \mathbf{u}_{i,t} \|^2 + C\xi \quad \text{s.t. } \ell^{lol}(\mathbf{w}, \mathbf{u}_1, \ldots, \mathbf{u}_k; (\mathbf{x}_t, y_t)) \le \xi, \ \xi \ge 0$$

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



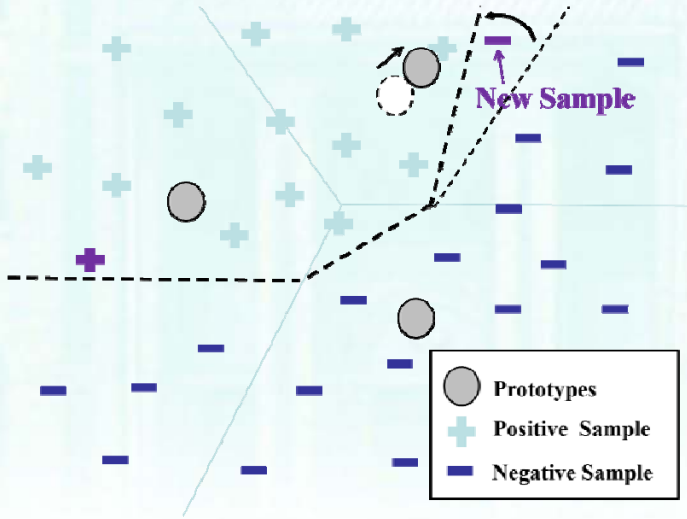

Zhaoze Zhou, Wei-Shi Zheng, et al. One-pass online learning: A local approach. Pattern Recognition, 2016.

# Online Learning

- **One-pass Learning**

$$\mathbf{w}_{t+1}, \mathbf{u}_{1,t+1}, \ldots, \mathbf{u}_{k,t+1} = \arg\min_{\mathbf{w},\mathbf{u}_1,\ldots,\mathbf{u}_k} \frac{\lambda}{2}\|\mathbf{w}-\mathbf{w}_t\|^2 + \frac{1}{2}\sum_{i=1}^{k}\|\mathbf{u}_i-\mathbf{u}_{i,t}\|^2 + C\xi \quad \text{s.t. } \ell^{lol}(\mathbf{w},\mathbf{u}_1,\ldots,\mathbf{u}_k;(\mathbf{x}_t,y_t)) \le \xi,\ \xi \ge 0$$

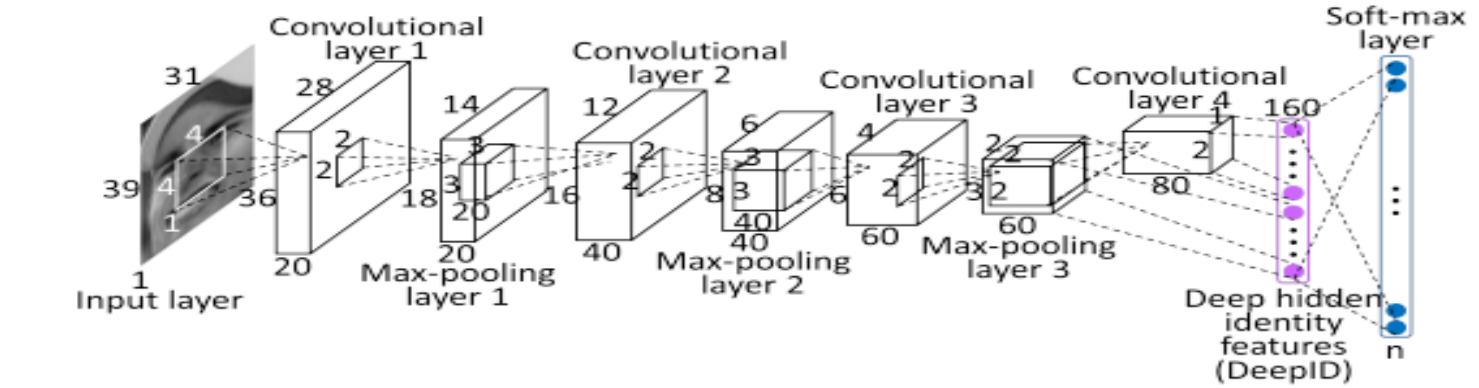
$$\ell^{lol}(\mathbf{w},\mathbf{u}_1,\ldots,\mathbf{u}_k;(\mathbf{x}_t,y_t)) = \begin{cases} 0 & y_t \cdot f(\mathbf{x}_t) \ge 1 \\ 1 - y_t \cdot f(\mathbf{x}_t) & \text{otherwise} \end{cases}$$

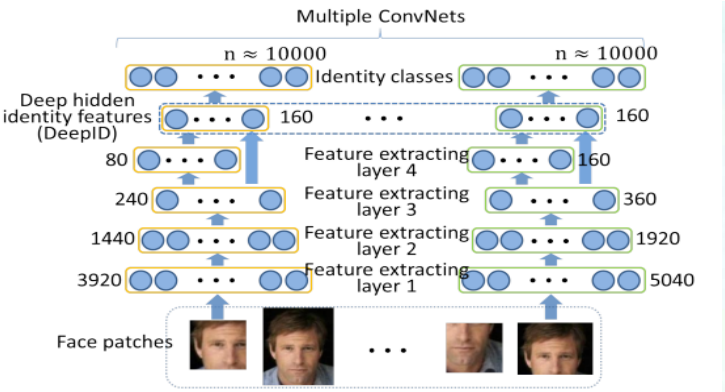



Zhaoze Zhou, Wei-Shi Zheng, et al. One-pass online learning: A local approach. Pattern Recognition, 2016.

# Deep Learning

- **DeepID**

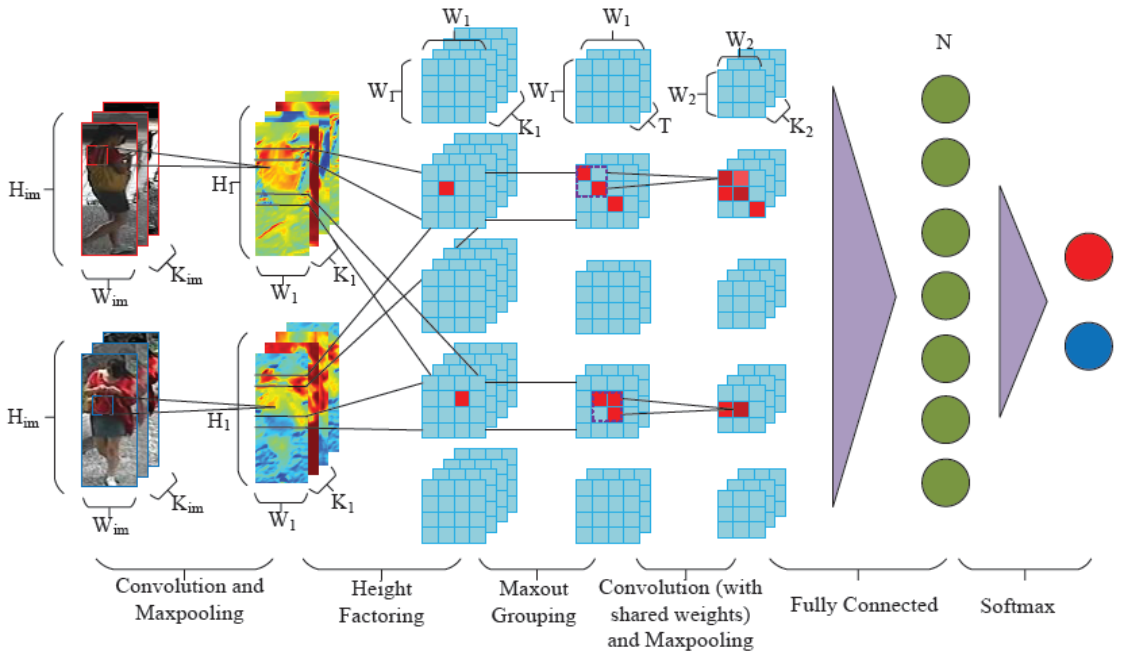



Yi Sun et al. Deep learning face representation from predicting 10,000 classes. In CVPR, 2014

# Deep Feature

- Deep Re-id

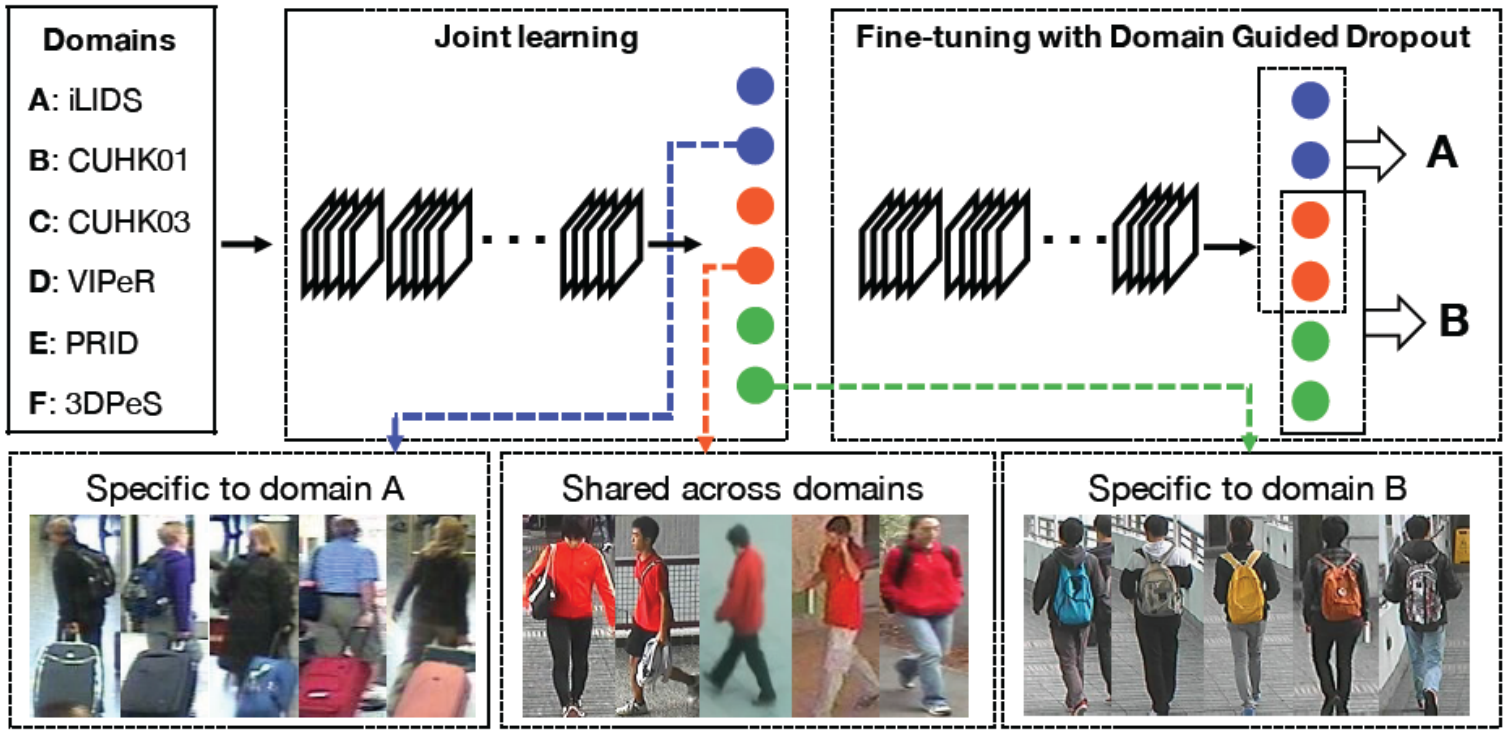



The filter pairing neural network (FPNN) jointly handles misalignment, photometric and geometric transforms, occlusions and background clutter.

**W. Li, R. Zhao, T. Xiao and X. Wang, "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014**

# Deep Feature

- Deep Learning with Domain Guided Dropout




Learning deep feature representations from multiple domains with Convolutional Neural Networks (CNNs).

**T. Xiao, H. Li, W. Ouyang and X. Wang. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification. IEEE International Conference on Computer Vision, 2016**

# Deep Learning

- **Deep RE-ID+Mirror**

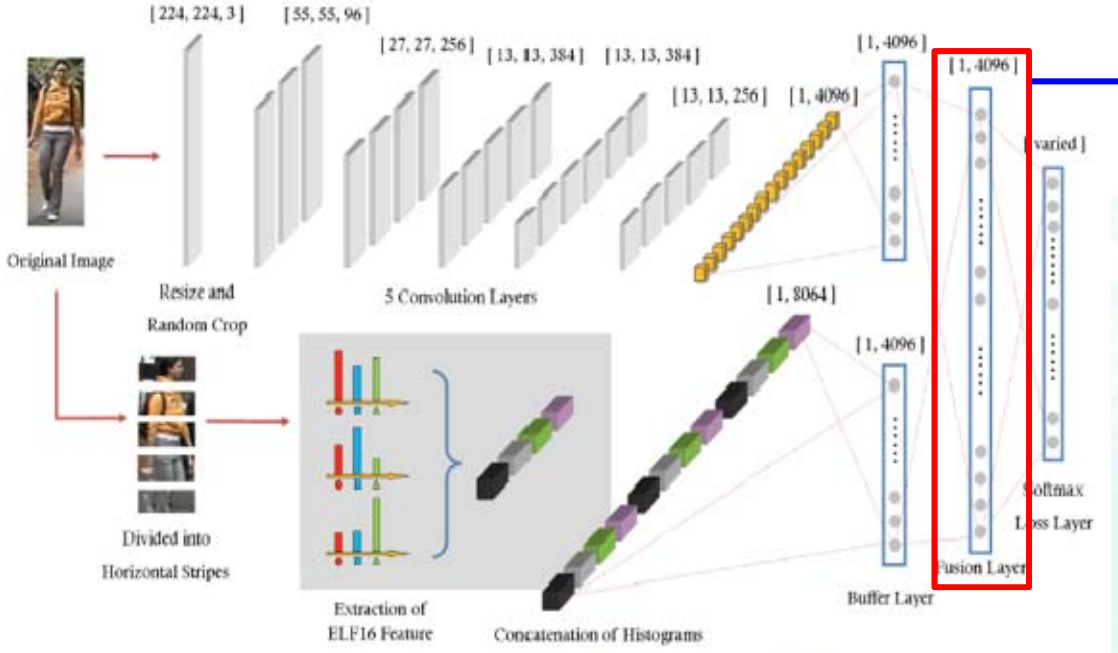
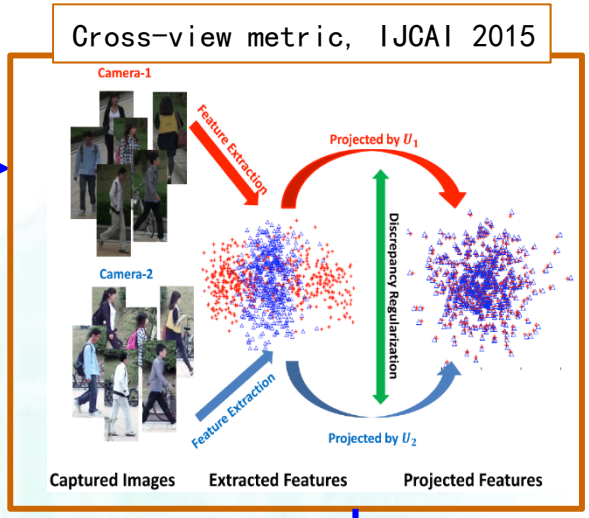



Figure 2: Fusion Feature Net (FFN) for ELF16 feature and CNN feature.



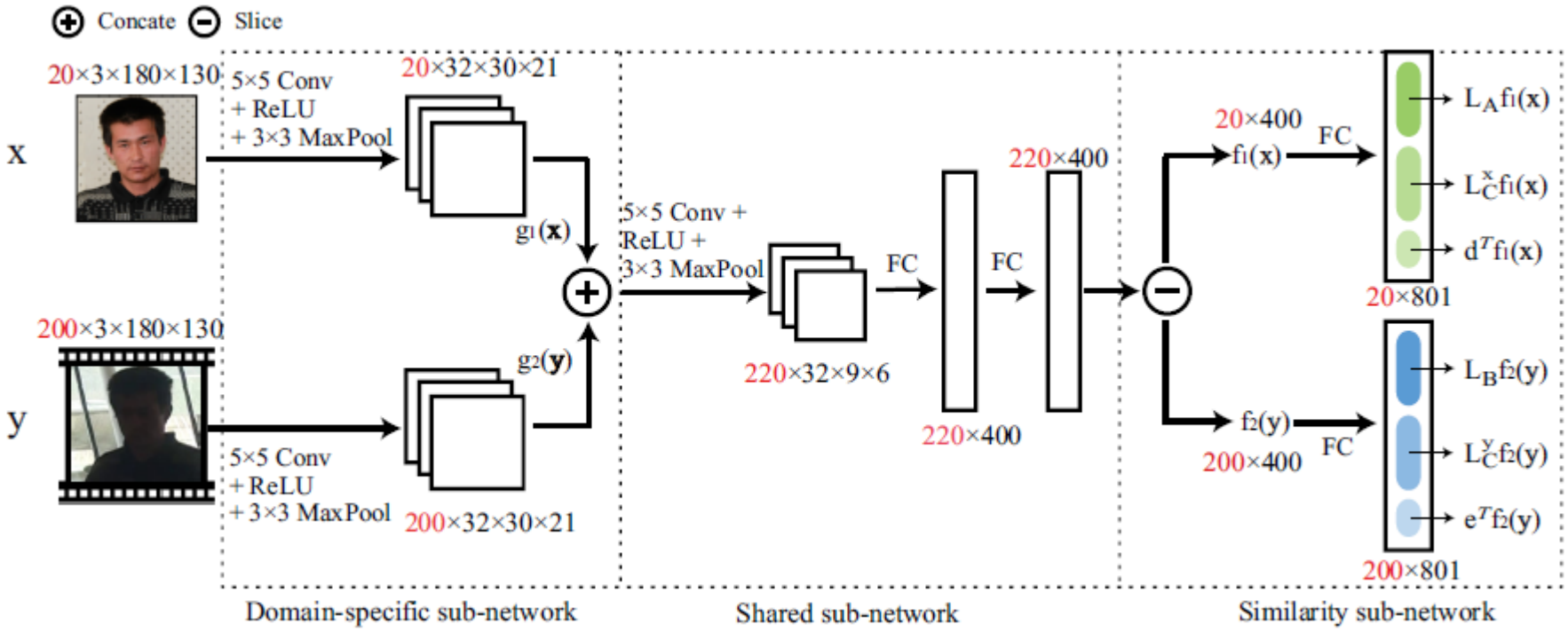

| Rank | 1 | 5 |
|---|---|---|
| Our Model | **51.06** | **81.01** |
| Deep Feature Learning[6] | 40.50 | 60.80 |
| LOMO+XQDA [18] | 40.00 | 67.40 |
| Mirror KMFA($R_{\chi 2}$) [3] | 42.97 | 75.82 |
| mFilter+LADF [30] | 43.39 | 73.04 |
| mFilter [30] | 29.11 | 52.10 |
| SalMatch [28] | 30.16 | 52.31 |
| LFDA [23] | 24.18 | 52.85 |
| LADF [17] | 29.34 | 61.04 |
| RDC [33] | 15.66 | 38.42 |
| KISSME [12] | 24.75 | 53.48 |
| LMNN-R [5] | 19.28 | 48.71 |
| PCCA [21] | 19.28 | 48.89 |
| $L_2 - norm$ | 10.89 | 22.37 |
| $L_1 - norm$ | 12.15 | 26.01 |

Shangxuan Wu, Ying-Cong Chen, Xiang Li, An-Cong Wu, Jin-Jie You, and **Wei-Shi Zheng***.
An Enhanced Deep Feature Representation for Person Re-identification. WACV, 2016.

Yingcong Chen, **Wei-Shi Zheng***, and Jian-Huang Lai
"Mirror Representation for Modeling View-specific Transform in Person Re-identification," IJCAI, 2015.

# Cross Modality Learning

## Generalized-Similarity-based Feature Learning



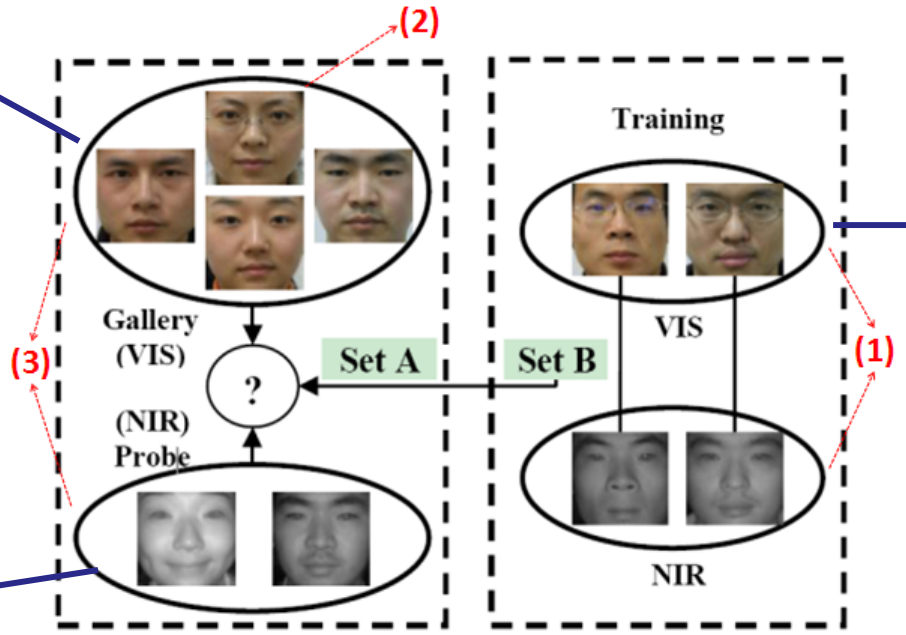

Learning deep feature representations for two modalities with domain-specific and shared sub-networks and a generalized similarity measure.

$$S(\mathrm{x},\mathrm{y}) = [\mathrm{x}^T\ \mathrm{y}^T\ 1] \begin{bmatrix} \mathrm{A} & \mathrm{C} & \mathrm{d} \\ \mathrm{C}^T & \mathrm{B} & \mathrm{e} \\ \mathrm{d}^T & \mathrm{e}^T & f \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ \mathrm{y} \\ 1 \end{bmatrix}$$

L. Lin; G. Wang; W. Zuo; F. Xiangchu; L. Zhang, "Cross-Domain Visual Matching via Generalized Similarity Measure and Feature Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence , 2016

# Cross Modality Learning

Existing VIS based face recognition system




Labelled pair-wise VIS-NIR training samples

Newly captured unlabelled NIR images

**Fig.1.** Illustration of VIS-NIR face matching

(1) Assume a set of VIS-NIR pairs of training people is available
(2) Guide the learned VIS-NIR matching upon training to facilitate the matching for target ones.

Jun-Yong Zhu (student), Wei-Shi Zheng*, Jian-Huang Lai, Stan Z. Li. Matching NIR Face to VIS Face using Transduction. IEEE Transactions on Information Forensics and Security, vol. 9, no. 3, pp. 501-514, 2014.

# TOO A LOT

- Manifold Learning
- Subspace: ICA, CCA
- Dictionary Learning
- Semi-supervised Learning
- Other Classifiers:
  - Bayes, Adaboost, Random Forest
- Active Learning
- Unsupervised Discriminant Learning

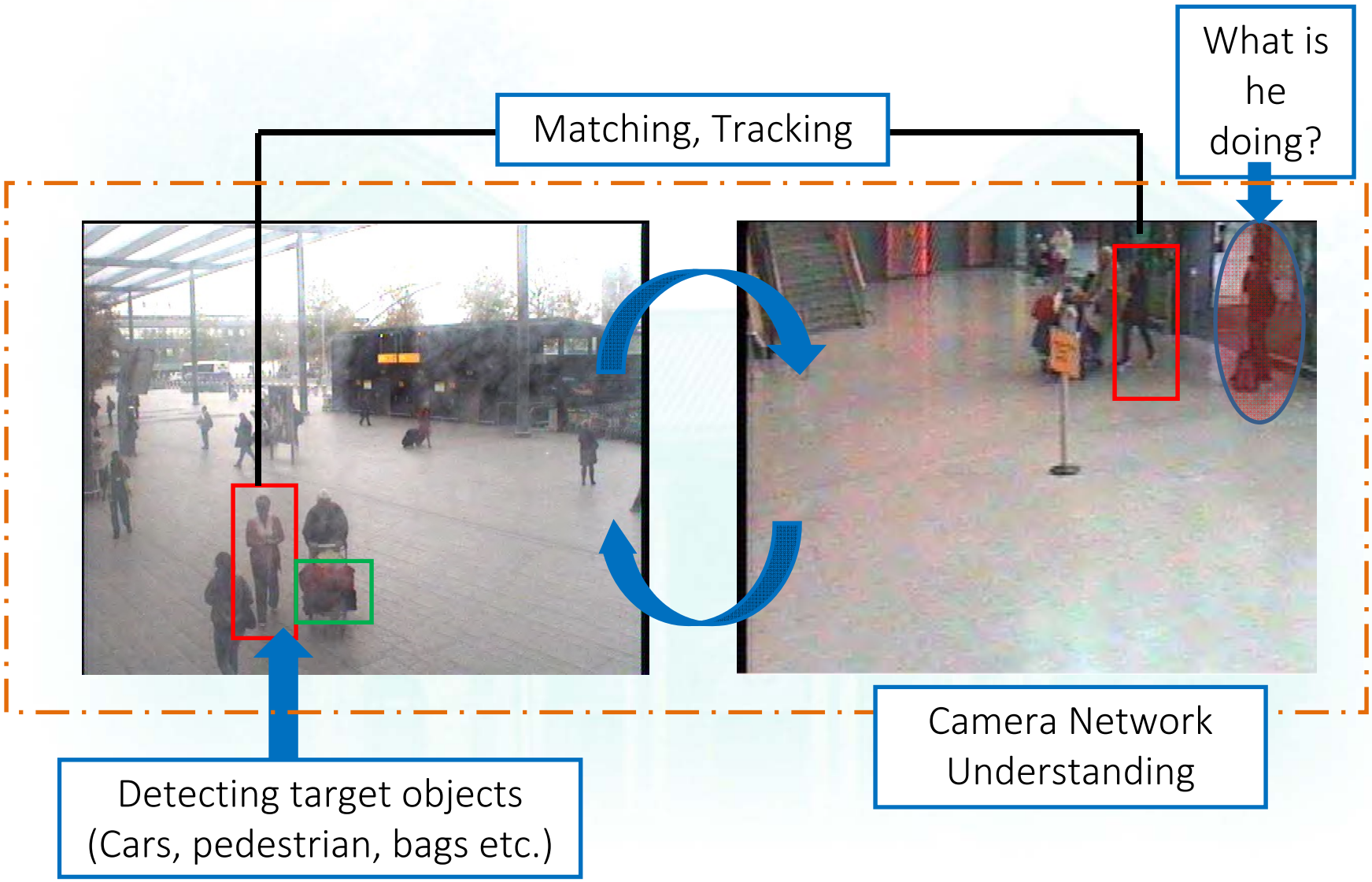

# MACHINE LEARNING FOR PERSON RE-IDENTIFICATION

# Outline

- Brief Introduction of ML for Biometrics
- **ML for Person Re-identification**
  - Distance Metric Learning
  - View Change Invariant Features
  - Partial Re-id
  - Low Resolution
  - Video-based Re-id
  - Cross Scenario Transfer
  - Open-world Modelling
  - Depth Re-identification
- Summary

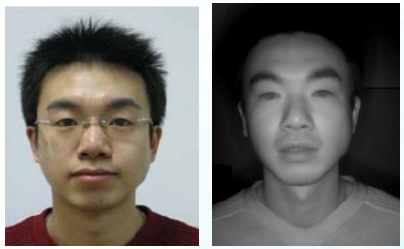
# Background: Visual Surveillance


# Person Re-identification



# Person Re-identification

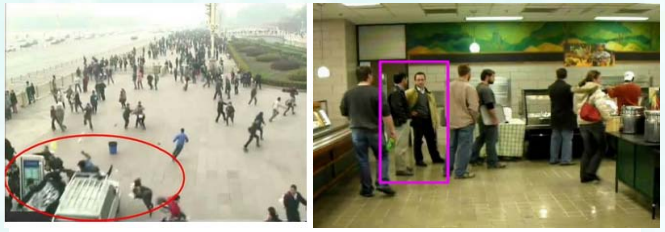
- Concern the person who is joining an activity



**Face Image Computing**

- Tracking him/her across camera-views
- Identifying him/her when we can capture his/her face very well
- Recognising/Searching face images in a Large Dataset

**Person Re-identification**



**Activity**

# Person Re-identification

- A key component to track people across disjoint views



Suspect, Bomb in Boston, USA（2013）




Suspect, Terrorist Attack, Kunming, China（2014）

# Person Re-identification

- kidnapping



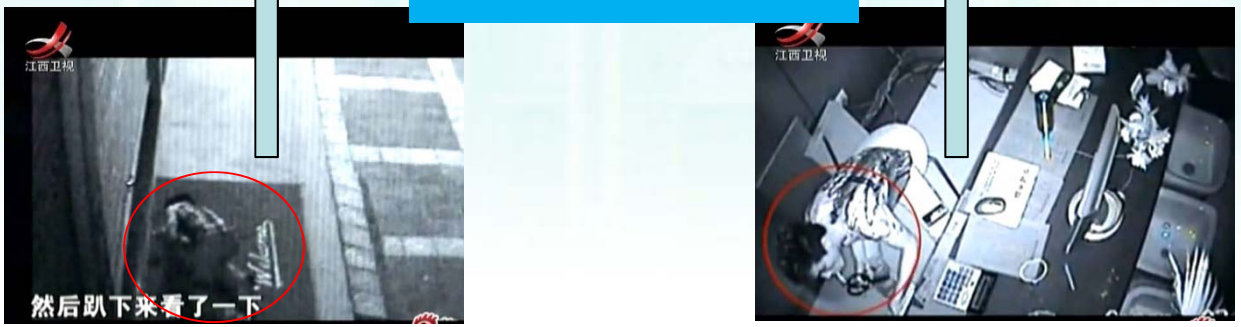

# Person Re-identification

$$\mathcal{D}_{\mathbf{M}}(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^\top \mathbf{M} (\vec{x}_i - \vec{x}_j)$$

Distance Learning

Feature Extraction

People Detection



# Person Re-identification: Challenges

- **Main Variations**


View | Lighting | Occlusion | Low Resolution | Cloth Change

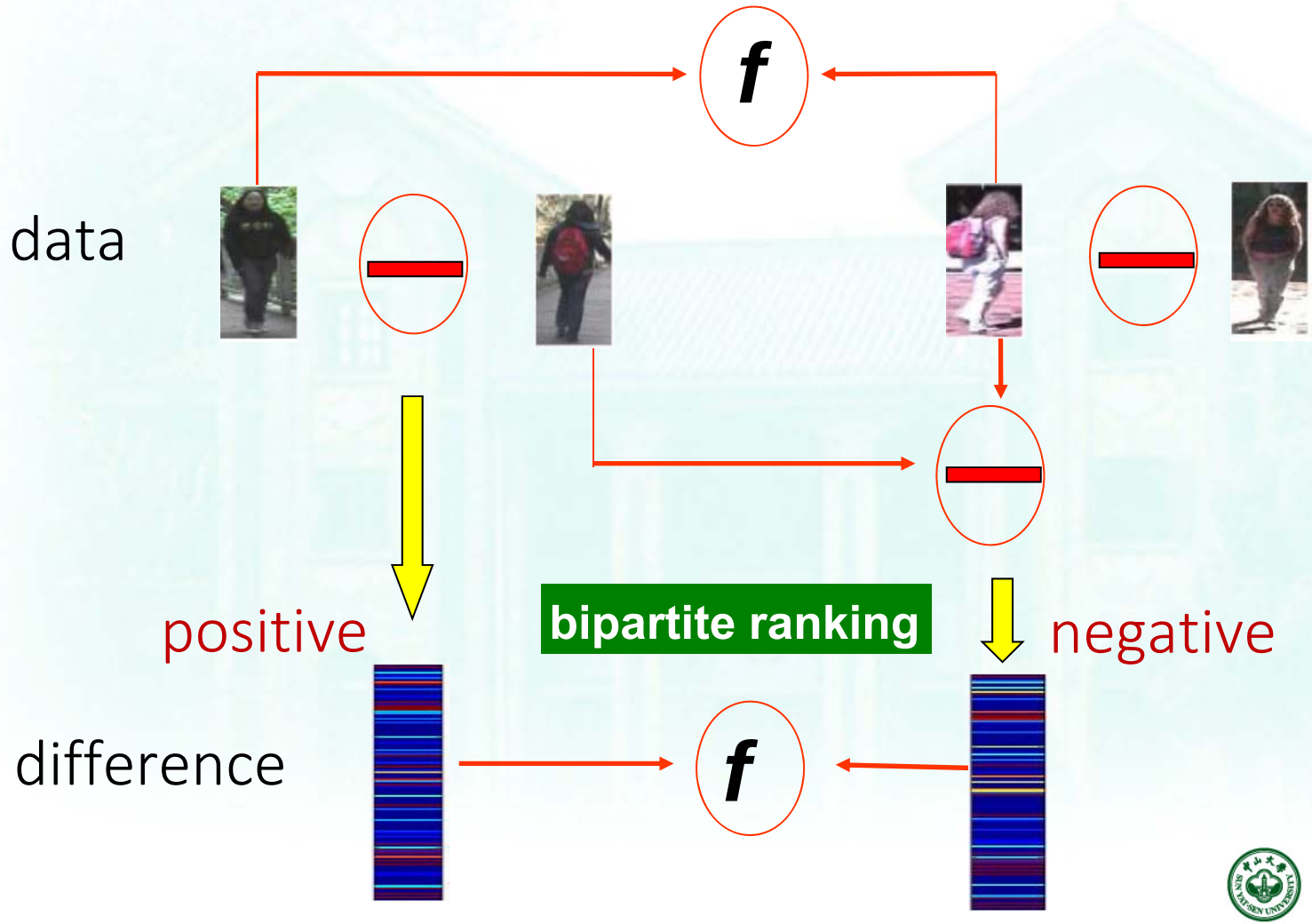


# How to measure the differences between two person images

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

# Triple based Learning: Bipartite Ranking

## Our Idea



# Triple based Learning: Bipartite Ranking

- A Relative Distance Comparison Model

OBJECTIVE

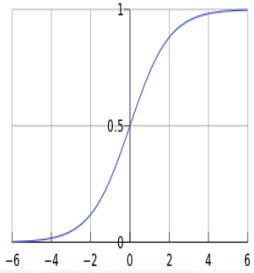
positive difference vector $f(\mathbf{x}_i^p) < f(\mathbf{x}_i^n)$ negative difference vector

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

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

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

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




soft margin measure

- Reduce the sensitivity for comparison
  Enhance the performance (↑ 20~30%,i-LIDS, VIPeR)

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

# Triple based Learning: Bipartite Ranking

## Entry-wise Absolute Difference Vector

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

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

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

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

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

# Triple based Learning: Bipartite Ranking

Learn the projection vectors each by each

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

where

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

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

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

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

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

# 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

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

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

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

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

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**

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

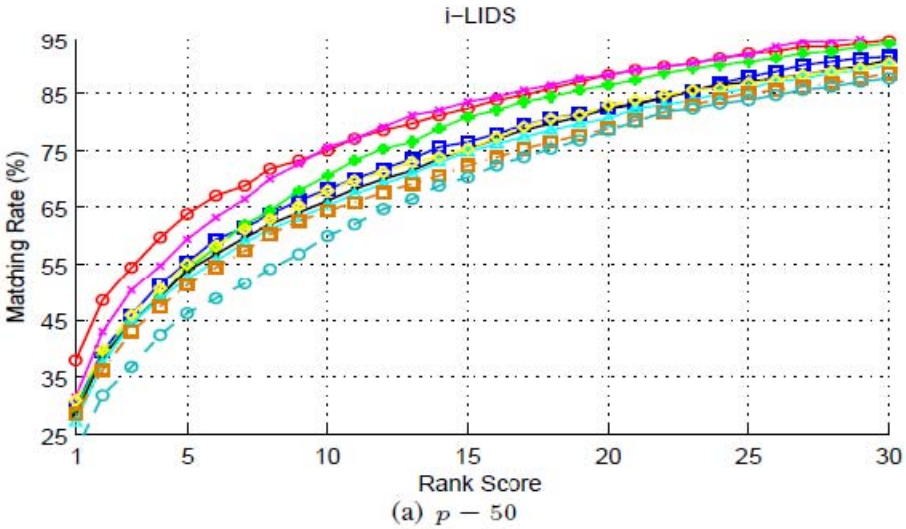
- **RDC: Trapped in locally optimal solution**

### Ensemble RDC: Modelling

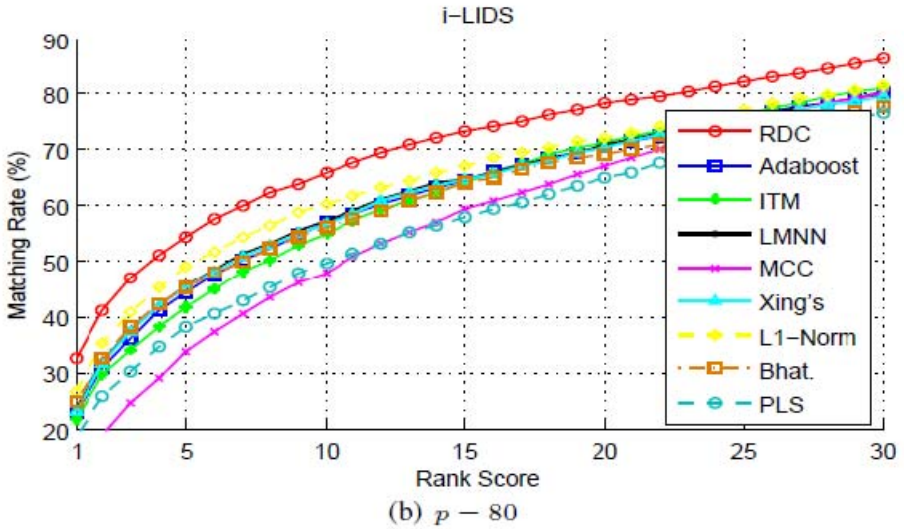
- **Randomly dividing the set into small groups**
- **Learning a set of weak RDC models**
- **Boosting them**

# Triple based Learning: Bipartite Ranking

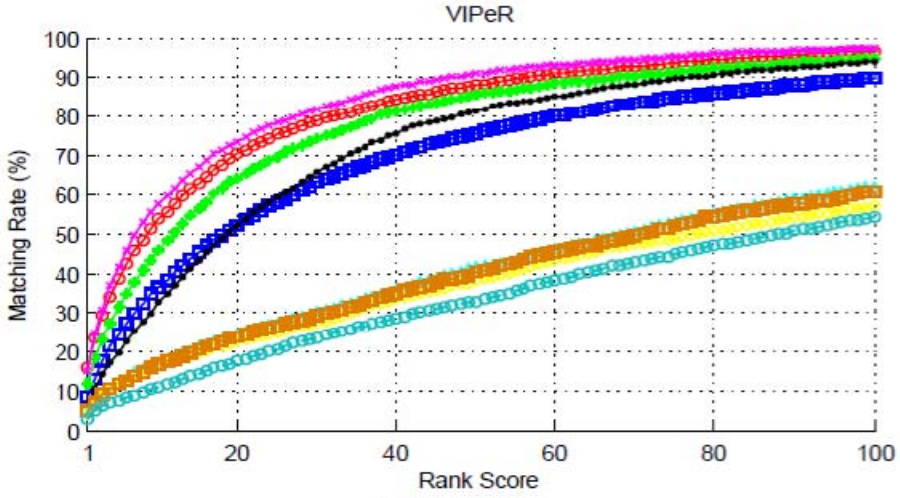
## Re-identification (i-LIDS&VIPeR)



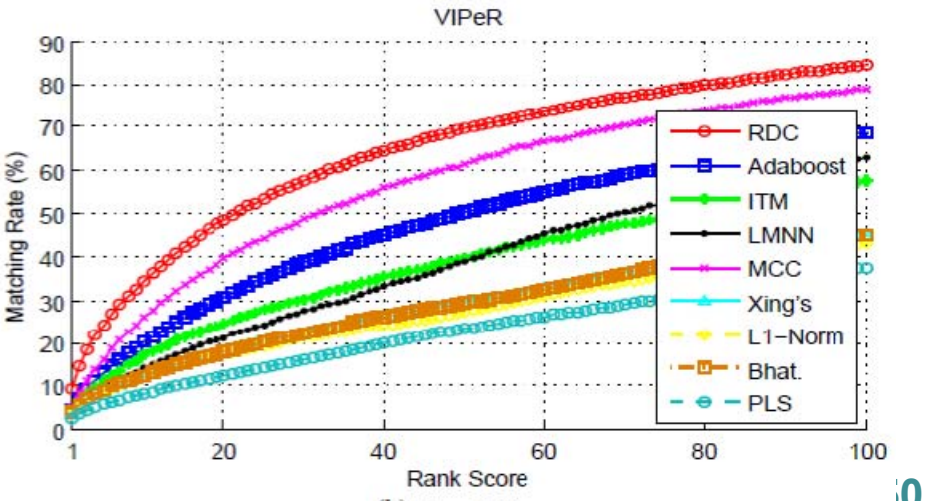
(a) $p - 50$



(b) $p - 80$



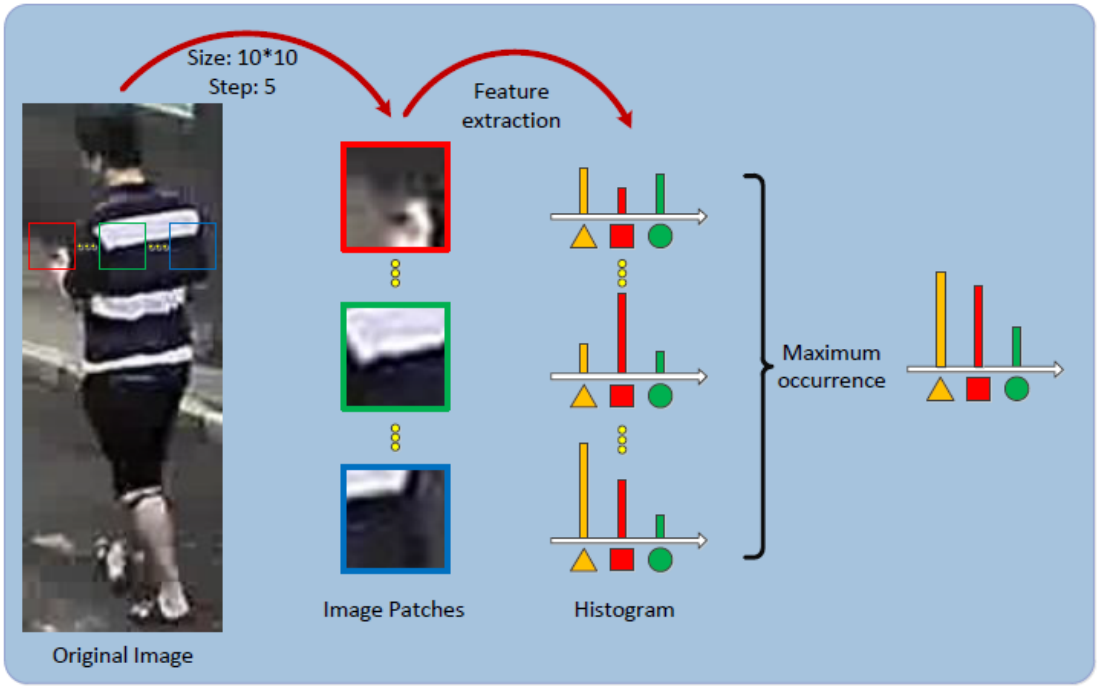
(a) $p = 316$



(b) $p = 532$

# XQDA

- Local Maximal Occurrence Representation (CVPR2015)

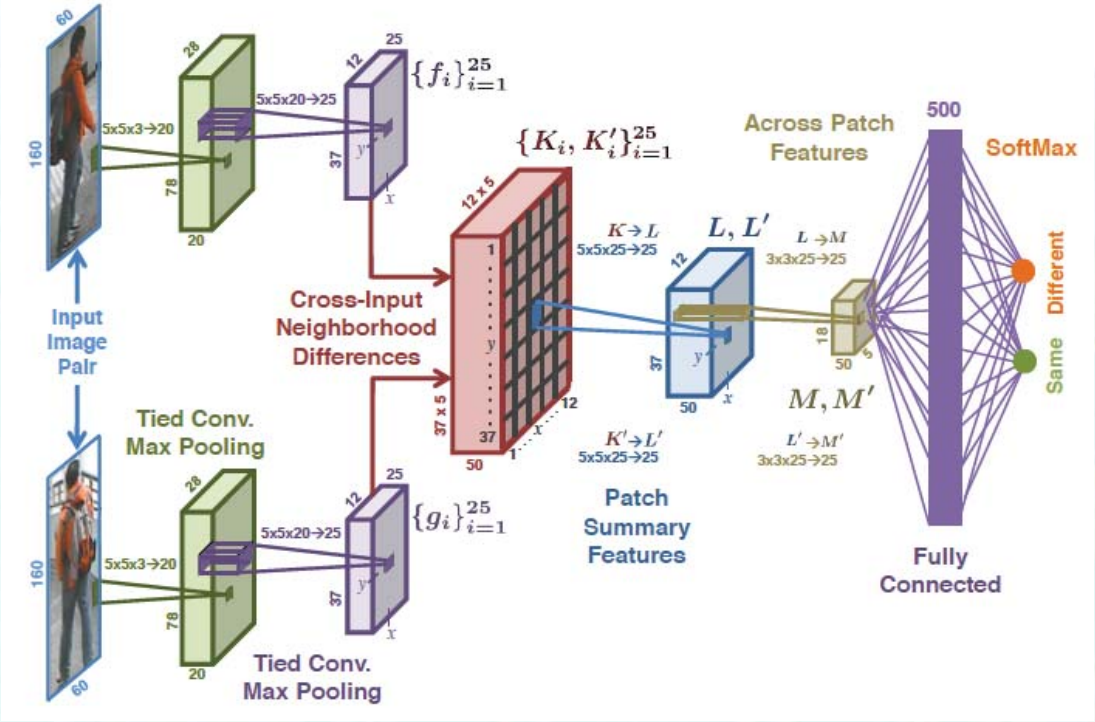



An effective handcrafted feature and a distance metric are proposed.

**Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning. CVPR, 2015.**

# Deep Distance

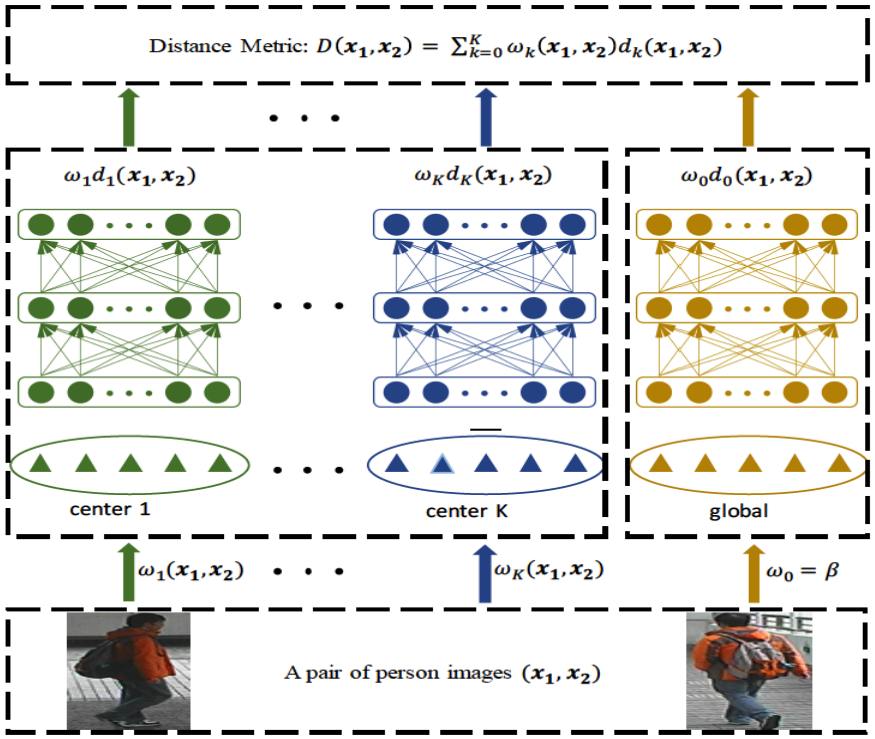
- Improved Deep Learning Architecture (CVPR2014)




A network for simultaneously learning features and a corresponding similarity metric for person re-identification.

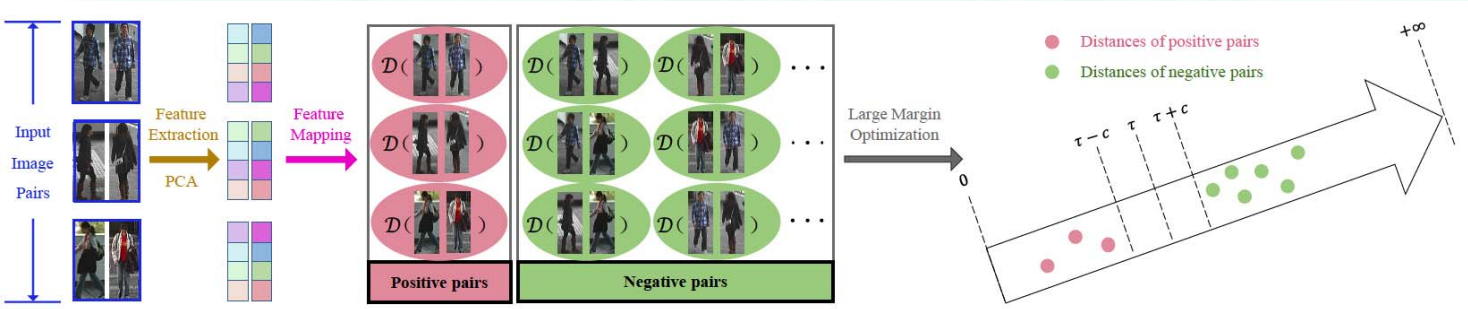
**E. Ahmed, M. Jones and T. K. Marks, "An improved deep learning architecture for person re-identification," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014**

# Deep Distance



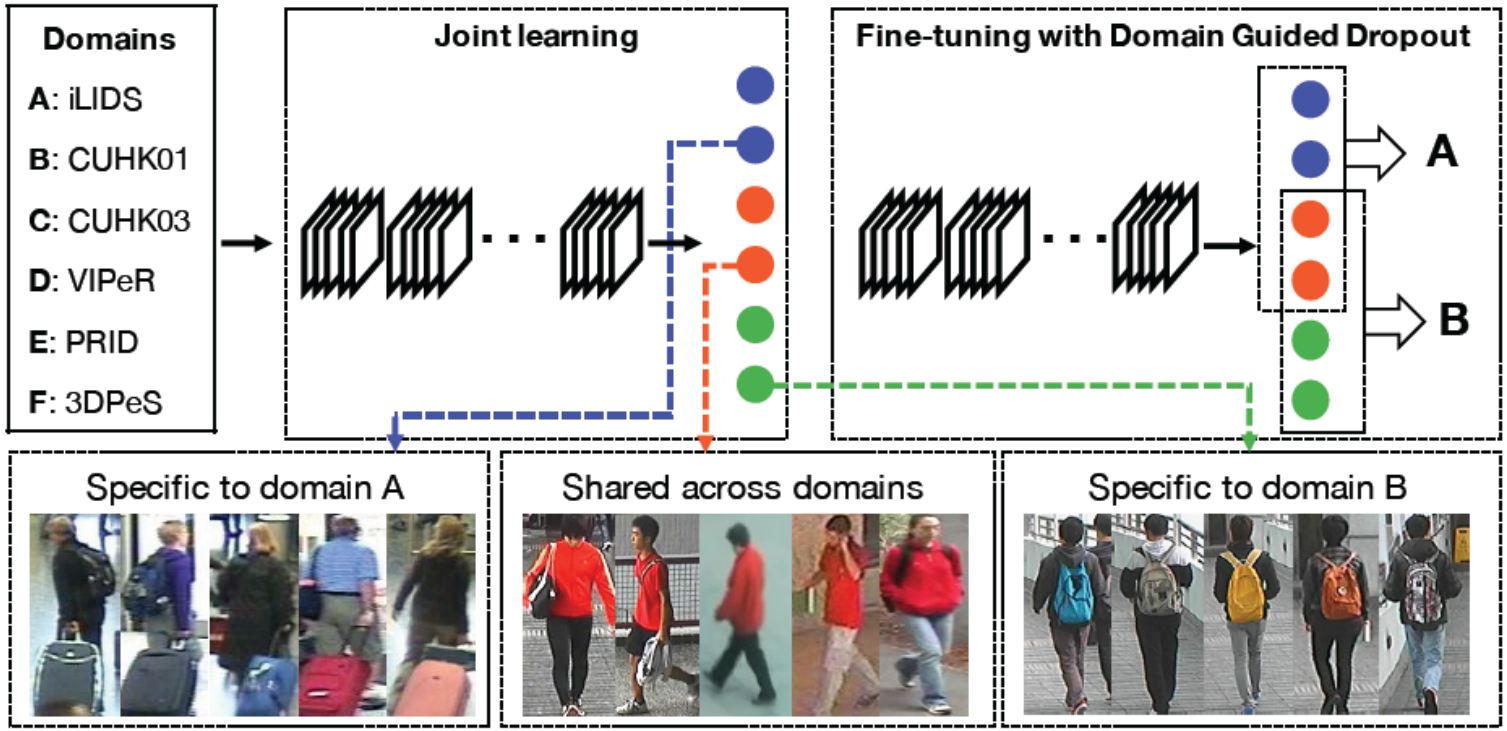

Nonlinear Local Metric Learning for Person Re-identification

By Siyuan Huang, Jiwen Lu, et al.



# Deep Feature

- Deep Learning with Domain Guided Dropout (CVPR2016)

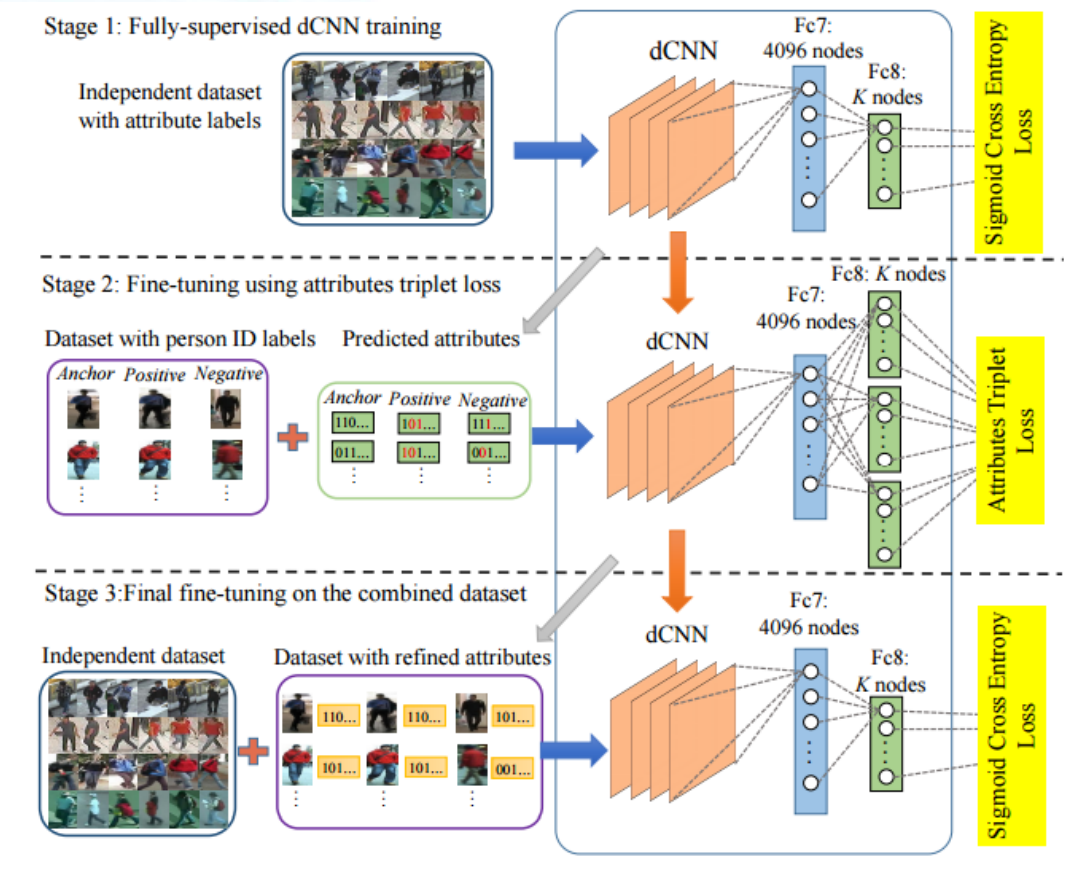



Learning deep feature representations from multiple domains with Convolutional Neural Networks (CNNs).

**T. Xiao, H. Li, W. Ouyang and X. Wang, "Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification," IEEE International Conference on Computer Vision (CVPR), 2016**

# Learning-based Mid-level Feature

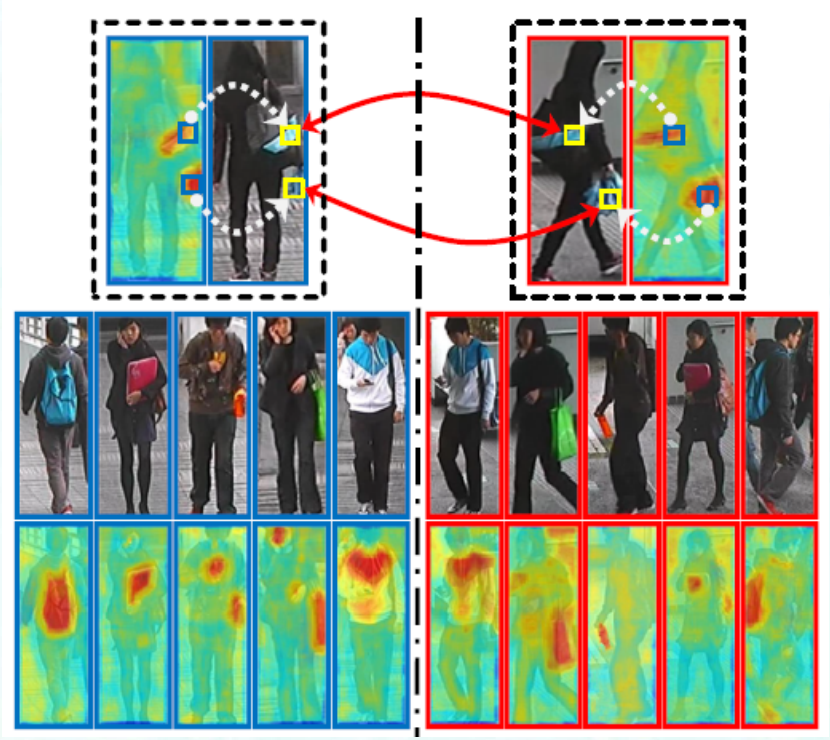
- Deep Attribute Learning (ECCV2016)




**Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, Qi Tian, "Deep Attributes Driven Multi-Camera Person Re-identification", European Conference on Computer Vision (ECCV), 2016.**

# Learning-based Mid-level Feature

- Salience Learning for Re-ID (CVPR2013)



Small salient regions are exploited to match persons.

**R. Zhao, W. Ouyang and X. Wang, "Unsupervised Salience Learning for Person Re-identification," Computer Vision and Pattern Recognition (CVPR), 2013**

# Unsupervised Learning

- Person Re-Identification by Unsupervised L1 Graph Learning

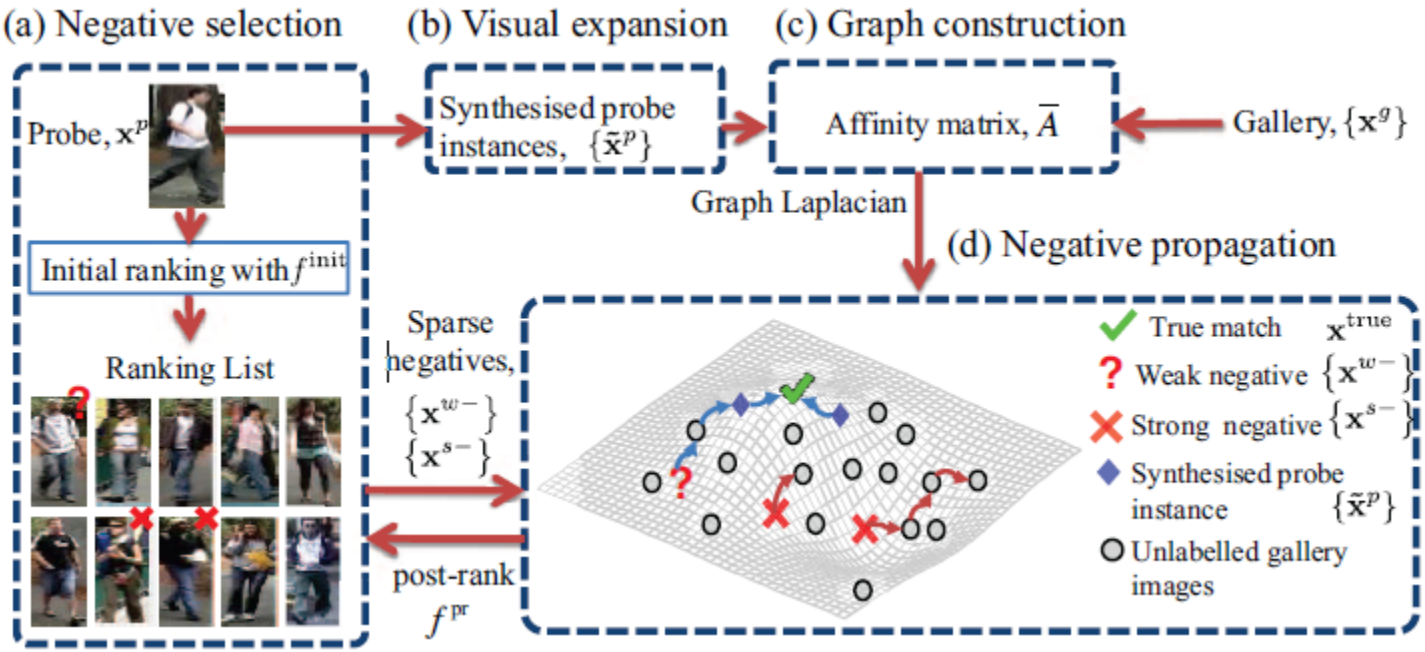
$$\min_{\mathbf{D},\mathbf{W},\mathbf{Y}} \frac{1}{2}\|\mathbf{X}-\mathbf{D}\mathbf{Y}\|_F^2 + \lambda_1\|\mathbf{Y}\mathbf{A}_\mathbf{W}\|_1 + \lambda_2\|\mathbf{W}\|_F^2$$
$$s.t. \quad \|\mathbf{d}_i\|_2^2 \le 1,\ \mathbf{W}_i^T\mathbf{1} = 1,\ \mathbf{W}_i \ge 0.$$

The unsupervised Re-ID problem is formulated by graph regularized dictionary learning method.

**E. Kodirov, T. Xiang, Z. Fu, S. Gong, "Person Re-Identification by Unsupervised L1 Graph Learning", ECCV, 2016**

# Post-rank Search

- Re-ranking Re-ID (ICCV2013)



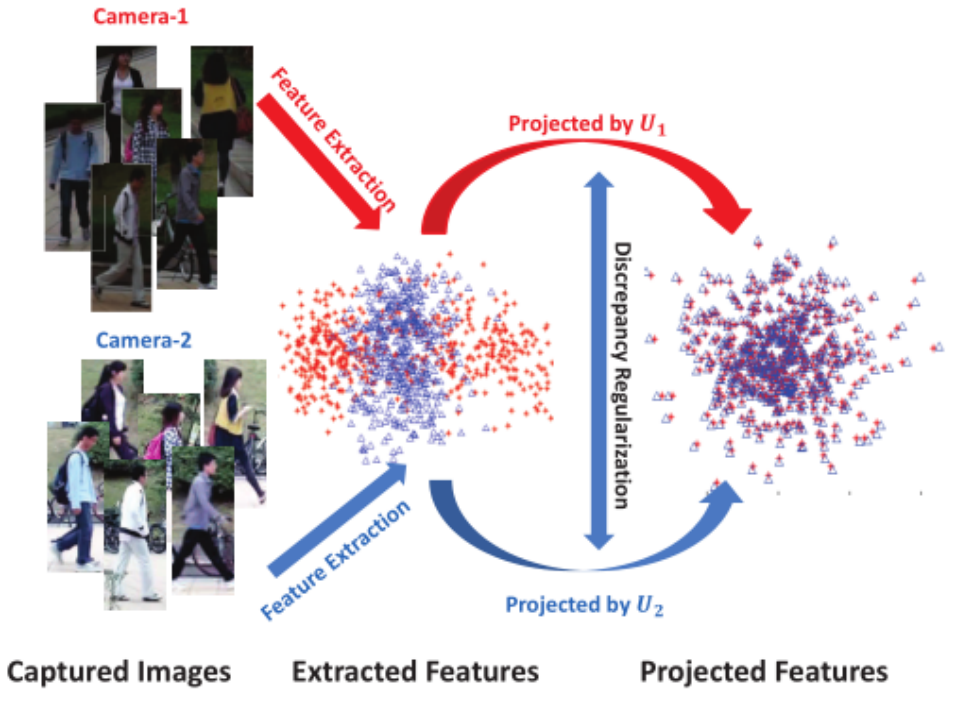

Strong negatives are labeled by human operator in the re-ranking process.

**C. Liu, C. C. Loy, S. Gong and G. Wang, "POP: Person Re-identification Post-rank Optimisation," IEEE International Conference on Computer Vision (ICCV), 2013**

# 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

Yingcong Chen, Wei-Shi Zheng*, and Jian-Huang Lai, "Mirror Representation for Modeling View-specific Transform in Person Re-identification," International Joint Conference on Artificial Intelligence (IJCAI), 2015.

Ying-Cong Chen, Wei-Shi Zheng*, Jianhuang Lai, Pong C. Yuen. An Asymmetric Distance Model for Cross-view Feature Mapping in Person Re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 2016.

中山大學 SUN YAT-SEN UNIVERSITY

# Mirror Representation

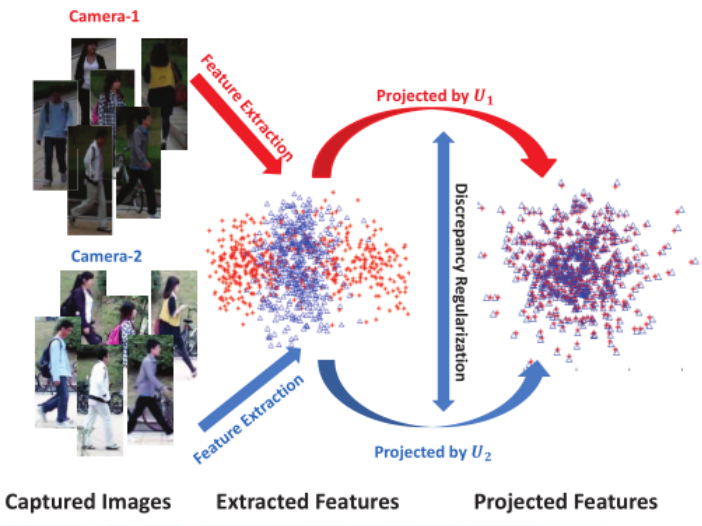
## Usefulness of View Label Information

- 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.

# Mirror Representation

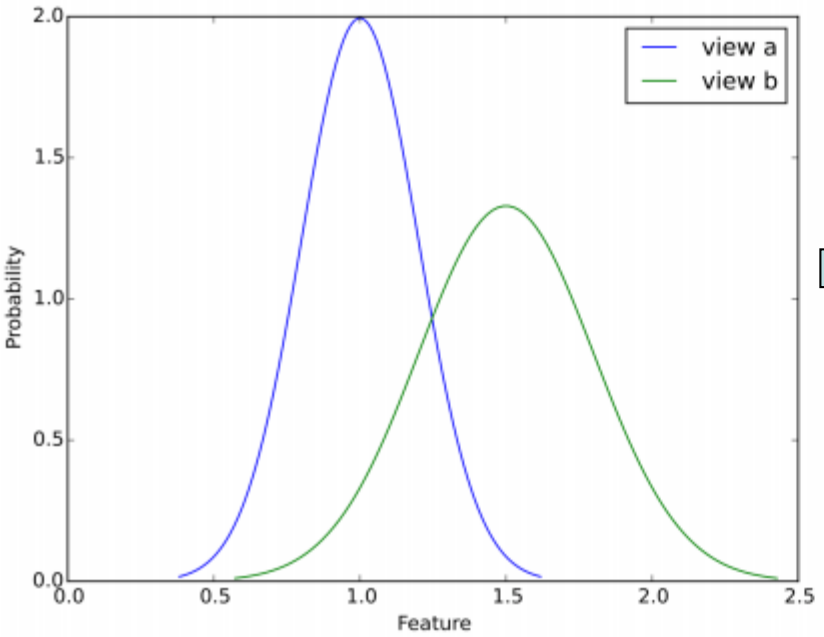
- Augmenting original feature with zeros
  - $X^a \rightarrow [X^a, 0]$
  - $X^b \rightarrow [0, X^b]$
- Learning projection bases with augmented features
  - $\min_U f(U^T X) \rightarrow U = [U_1; U_2]$
- View-specific projection
  - $f_a(X^a) = U^T[X^a, 0] = U_1^T X^a$
  - $f_b(X^b) = U^T[0, X^b] = U_2^T X^b$

Zero-Padding Augmentation

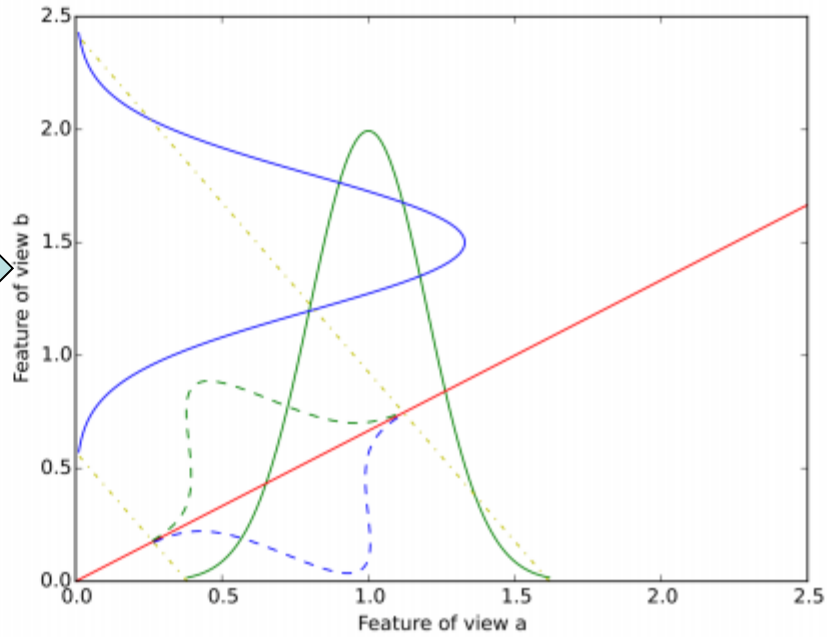


# Mirror Representation

## Illustration of Zero-Padding Augmentation




(a) Mismatched Distribution




(b) Zero-padding

# 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 $U_2$.

# Reformulation of Zero-Padding

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

generalise

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

control the discrepancy of $f_a$ and $f_b$

# A Feature-Level Discrepancy Modeling

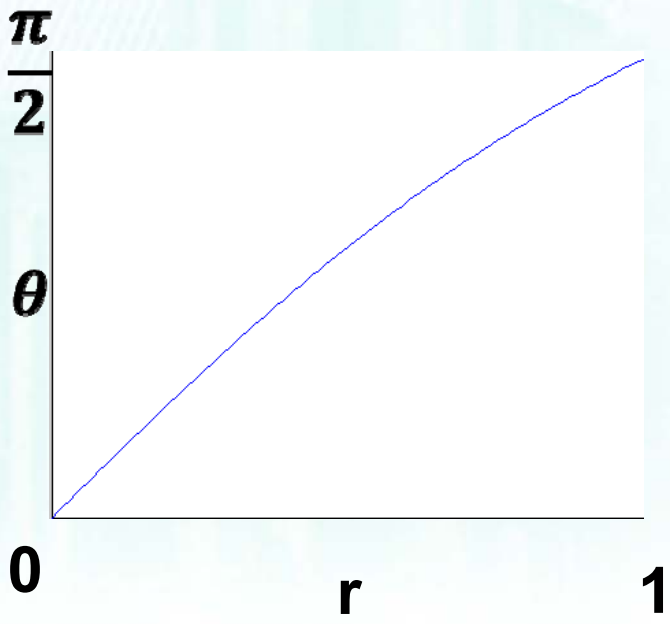
- $R = \frac{1-r}{z} I$
- $M = \frac{1+r}{z} I$

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

$$0 < r < 1$$

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

- **Is it Not Optimal?**

$$\boldsymbol{R} = \frac{2-\omega}{\varpi}\boldsymbol{I}_{d\times d}, \qquad \boldsymbol{M} = \frac{\omega}{\varpi}\boldsymbol{I}_{d\times d},$$

$$\boldsymbol{R}^{\text{opt}} = \boldsymbol{R} + \nabla\boldsymbol{R}, \quad \boldsymbol{M}^{\text{opt}} = \boldsymbol{M} + \nabla\boldsymbol{M}$$

$$\begin{aligned} f_a(\tilde{\boldsymbol{X}}^a_{\text{cca}}) &= \boldsymbol{W}^\top \tilde{\boldsymbol{X}}^a_{\text{cca}} \\ &= (\boldsymbol{R}^\top(\boldsymbol{W}^a)^\top + \boldsymbol{M}^\top(\boldsymbol{W}^b)^\top)\boldsymbol{X}^a \\ &= \underbrace{\frac{2-\omega}{\varpi}(\boldsymbol{W}^a)^\top}_{\text{specificity}}\boldsymbol{X}^a + \underbrace{\frac{\omega}{\varpi}(\boldsymbol{W}^b)^\top}_{\text{adaptiveness}}\boldsymbol{X}^a, \\ f_b(\tilde{\boldsymbol{X}}^b_{\text{cca}}) &= \boldsymbol{W}^\top \tilde{\boldsymbol{X}}^b_{\text{cca}} \\ &= (\boldsymbol{M}^\top(\boldsymbol{W}^a)^\top + \boldsymbol{R}^\top(\boldsymbol{W}^b)^\top)\boldsymbol{X}^b \\ &= \underbrace{\frac{\omega}{\varpi}(\boldsymbol{W}^a)^\top}_{\text{adaptiveness}}\boldsymbol{X}^b + \underbrace{\frac{2-\omega}{\varpi}(\boldsymbol{W}^b)^\top}_{\text{specificity}}\boldsymbol{X}^b, \end{aligned}$$

$$\begin{aligned} &\boldsymbol{W}^a(\boldsymbol{R}+\nabla\boldsymbol{R}) + \boldsymbol{W}^b(\boldsymbol{M}+\nabla\boldsymbol{M}) \\ &\quad = (\boldsymbol{W}^a + \nabla\boldsymbol{W}^a)\boldsymbol{R} + (\boldsymbol{W}^b + \nabla\boldsymbol{W}^b)\boldsymbol{M} \\ &\boldsymbol{W}^a(\boldsymbol{M}+\nabla\boldsymbol{M}) + \boldsymbol{W}^b(\boldsymbol{R}+\nabla\boldsymbol{R}) \\ &\quad = (\boldsymbol{W}^a + \nabla\boldsymbol{W}^a)\boldsymbol{M} + (\boldsymbol{W}^b + \nabla\boldsymbol{W}^b)\boldsymbol{R} \end{aligned}$$

# 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

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

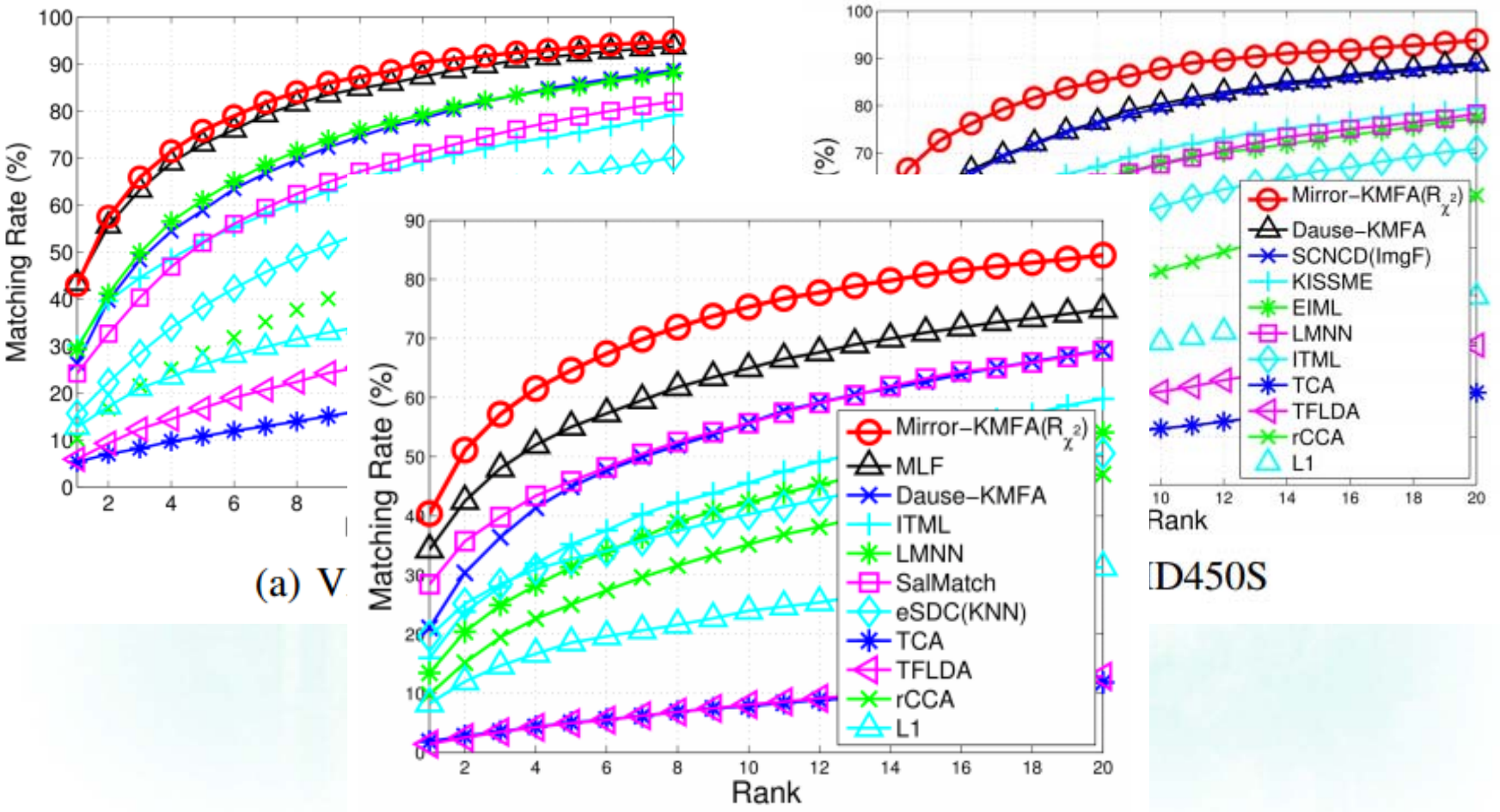
can be solved by traditional metric learning (with ridge regularization)

# 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.

# Performance



# Deep RE-ID+Mirror

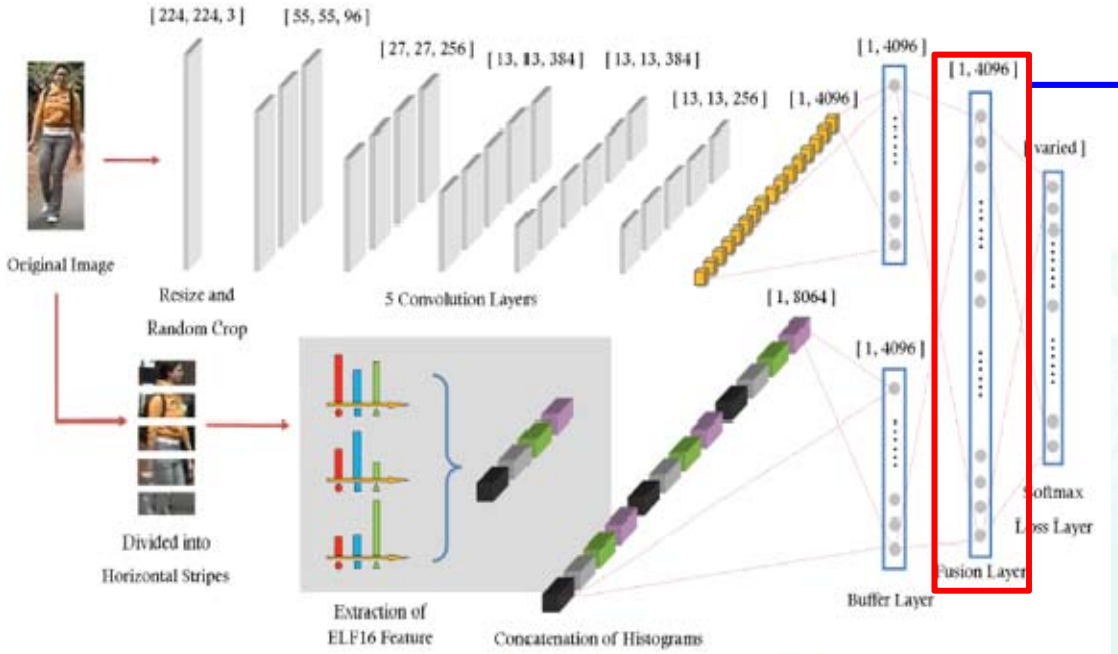
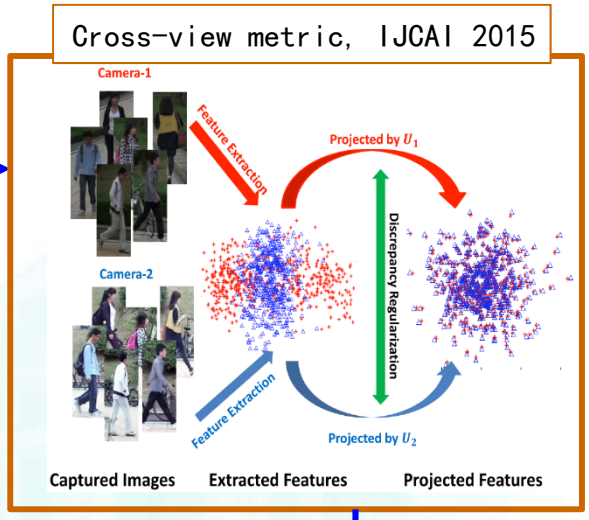



Figure 2: Fusion Feature Net (FFN) for ELF16 feature and CNN feature.




| Rank | 1 | 5 |
|---|---|---|
| Our Model | **51.06** | **81.01** |
| Deep Feature Learning[6] | 40.50 | 60.80 |
| LOMO+XQDA [18] | 40.00 | 67.40 |
| Mirror KMFA($R_{\chi 2}$) [3] | 42.97 | 75.82 |
| mFilter+LADF [30] | 43.39 | 73.04 |
| mFilter [30] | 29.11 | 52.10 |
| SalMatch [28] | 30.16 | 52.31 |
| LFDA [23] | 24.18 | 52.85 |
| LADF [17] | 29.34 | 61.04 |
| RDC [33] | 15.66 | 38.42 |
| KISSME [12] | 24.75 | 53.48 |
| LMNN-R [5] | 19.28 | 48.71 |
| PCCA [21] | 19.28 | 48.89 |
| $L_2 - norm$ | 10.89 | 22.37 |
| $L_1 - norm$ | 12.15 | 26.01 |

**Shangxuan Wu, Ying-Cong Chen, Xiang Li, An-Cong Wu, Jin-Jie You, and Wei-Shi Zheng*.**
An Enhanced Deep Feature Representation for Person Re-identification. WACV 2016.

Yingcong Chen (student), Wei-Shi Zheng*, and Jian-Huang Lai "Mirror Representation for Modeling View-specific Transform in Person Re-identification," IJCAI, 2015.

# Deep RE-ID+Mirror

(a) VIPeR

(b) CUHK01

(c) PRID450s

| Rank | 1 | 5 | 10 | 20 |
|---|---|---|---|---|
| Our Model | **51.06** | **81.01** | **91.39** | **96.90** |
| Deep Feature Learning[6] | 40.50 | 60.80 | 70.40 | 84.40 |
| LOMO+XQDA [18] | 40.00 | 67.40 | 80.51 | 91.08 |
| Mirror KMFA($R_{\chi 2}$) [3] | 42.97 | 75.82 | 87.28 | 94.84 |
| mFilter+LADF [29] | 43.39 | 73.04 | 84.87 | 93.70 |
| mFilter [29] | 29.11 | 52.10 | 67.20 | 80.14 |
| SalMatch [27] | 30.16 | 52.31 | 65.54 | 79.15 |
| LFDA [23] | 24.18 | 52.85 | 67.12 | 78.96 |
| LADF [17] | 29.34 | 61.04 | 75.98 | 88.10 |
| RDC [32] | 15.66 | 38.42 | 53.86 | 70.09 |
| KISSME [12] | 24.75 | 53.48 | 67.44 | 80.92 |
| LMNN-R [5] | 19.28 | 48.71 | 65.49 | 78.34 |
| PCCA [21] | 19.28 | 48.89 | 64.91 | 80.28 |
| $L_2 - norm$ | 10.89 | 22.37 | 32.34 | 45.19 |
| $L_1 - norm$ | 12.15 | 26.01 | 32.09 | 34.72 |

Table 2: Top Matching Rank on VIPeR.

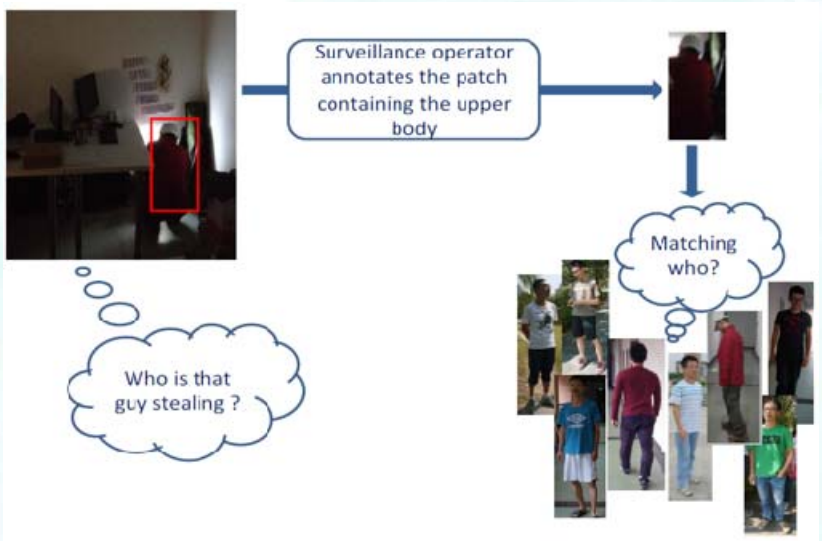
| Rank | 1 | 5 | 10 | 20 |
|---|---|---|---|---|
| Our Model | **66.62** | **86.84** | **92.84** | **96.89** |
| Mirror KMFA($R_{\chi 2}$) [3] | 55.42 | 79.29 | 87.82 | 93.87 |
| Ahmed's Deep Re-id [1] | 34.81 | 63.72 | 76.24 | 81.90 |
| ITML [4] | 24.27 | 47.82 | 58.67 | 70.89 |
| LFDA [23] | 36.18 | 61.33 | 72.40 | 82.67 |
| KISSME [12] | 36.31 | 65.11 | 75.42 | 83.69 |
| LMNN-R [5] | 28.98 | 55.29 | 67.64 | 78.36 |
| $L_2 - norm$ | 11.33 | 24.50 | 33.22 | 43.89 |
| $L_1 - norm$ | 25.50 | 25.33 | 51.73 | 53.07 |

Table 4: Top Matching Rank on PRID450s.

| Rank | 1 | 5 | 10 | 20 |
|---|---|---|---|---|
| Our Model | **55.51** | **78.40** | **83.68** | **92.59** |
| Mirror KMFA($R_{\chi 2}$) [3] | 40.40 | 64.63 | 75.34 | 84.08 |
| Ahmed's Deep Re-id [1] | 47.53 | 72.10 | 80.53 | 88.49 |
| mFilter [29] | 34.30 | 55.12 | 64.91 | 74.53 |
| SalMatch [27] | 28.45 | 45.85 | 55.67 | 67.95 |
| DeepReID [16] | 27.87 | 64.01 | 82.50 | 87.36 |
| ITML [4] | 15.98 | 35.22 | 45.60 | 59.81 |
| eSDC [28] | 19.67 | 32.72 | 40.29 | 50.58 |
| LFDA [23] | 22.08 | 41.56 | 53.85 | 64.51 |
| KISSME [12] | 14.02 | 32.20 | 44.44 | 56.61 |
| LMNN-R [5] | 13.45 | 31.33 | 42.25 | 54.11 |
| $L_2 - norm$ | 5.63 | 16.00 | 22.89 | 30.63 |
| $L_1 - norm$ | 10.80 | 15.51 | 37.57 | 35.57 |

Table 3: Top Matching Rank on CUHK01.

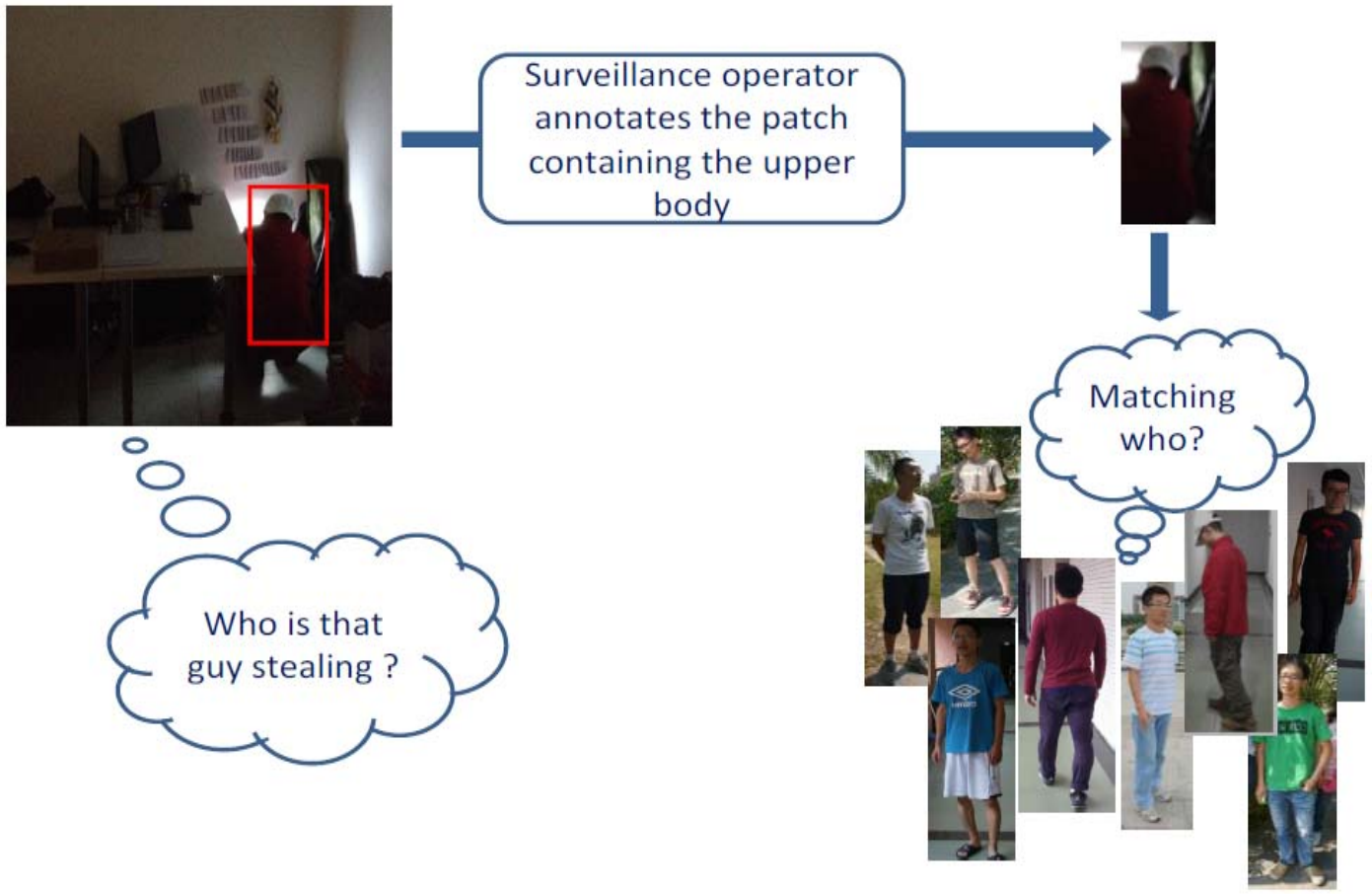
# Partial Re-identification



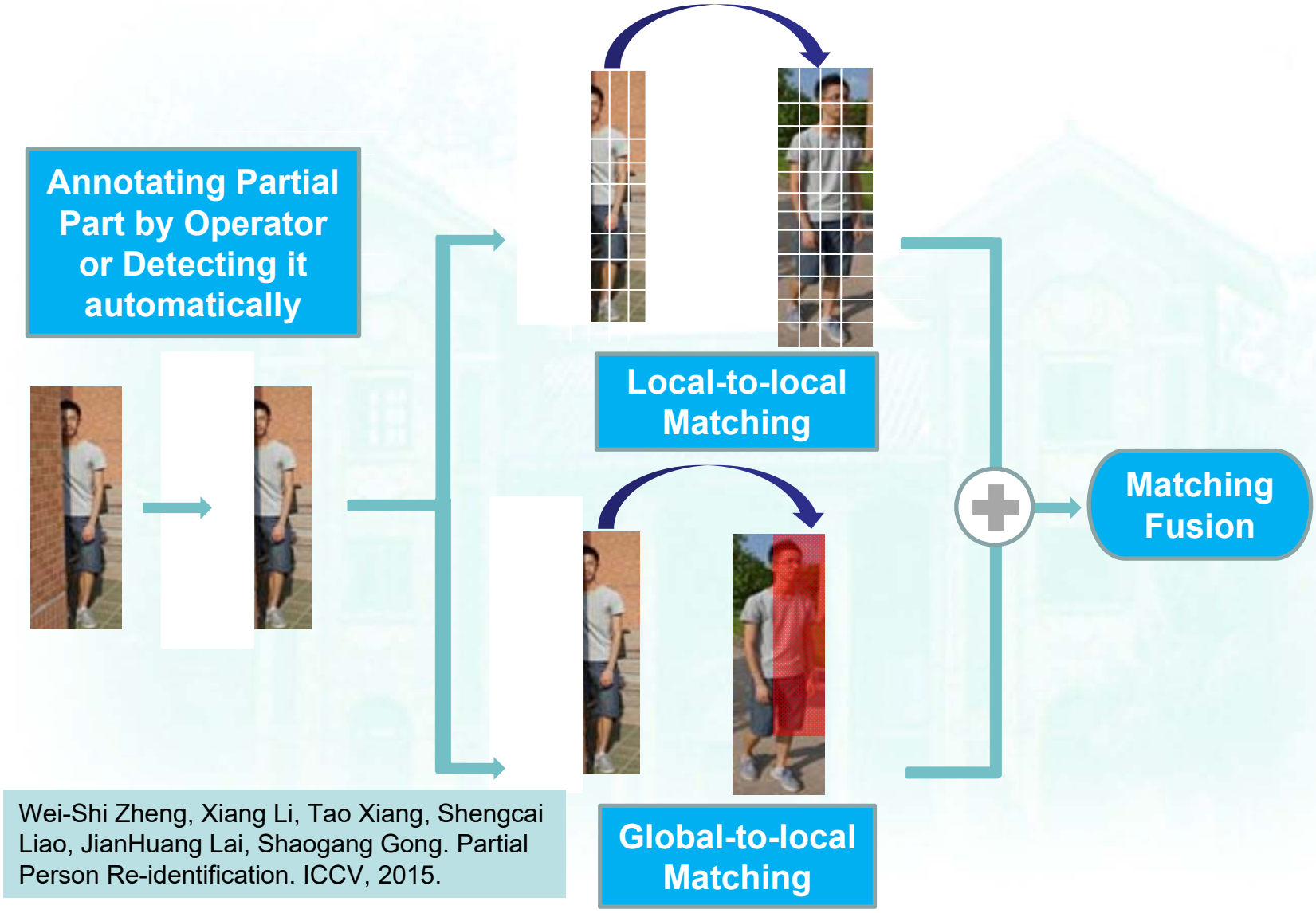

**Wei-Shi Zheng**, Xiang Li, Tao Xiang, Shengcai Liao, JianHuang Lai, Shaogang Gong. Partial Person Re-identification. In IEEE Conf. on Computer Vision (ICCV), 2015 (oral)

中山大學 SUN YAT-SEN UNIVERSITY

# Partial Observation



# Partial Re-ID




Wei-Shi Zheng, Xiang Li, Tao Xiang, Shengcai Liao, JianHuang Lai, Shaogang Gong. Partial Person Re-identification. ICCV, 2015.

# Partial Re-ID

## Local-to-local Matching

### — Ambiguity-Sensitive Matching classifier (AMC)

Constructing patch level descriptors from gallery person images to form a dictionary

$\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_C]$ where $\mathbf{D}_c = [\mathbf{d}_{c_1}, \mathbf{d}_{c_2}, \cdots, \mathbf{d}_{c_{k_c}}]$

probe patch feature

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2,$$

Ambiguity Score

$\mathbf{pa}_i = [pa_{i_1}, pa_{i_2}, \cdots, pa_{i_K}]^T,\ i = 1, 2, \cdots, n,$

where $pa_{i_j} = -G_\sigma(D(\mathbf{y}_i, \mathbf{d}_j)),\ j = 1, 2, \cdots, K,$

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \alpha \mathbf{pa}_i^T \mathbf{x}_i, i = 1, \cdots, n,$$

Ambiguity constraint

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \alpha \mathbf{pa}_i^T \mathbf{x}_i + \beta \|\mathbf{x}_i\|_1, i = 1, \cdots, n,$$

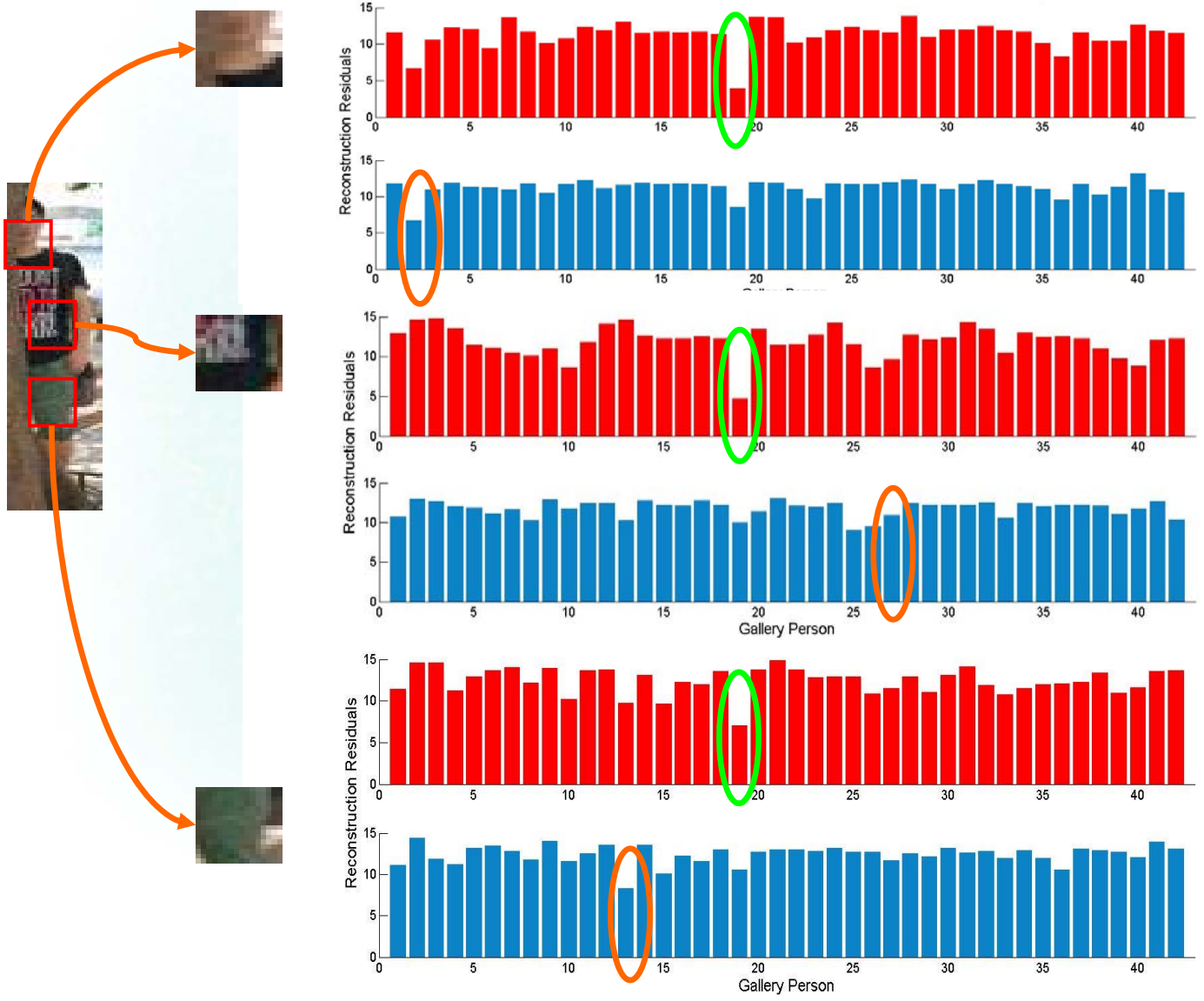
Sparsity constraint

Classifying a probe partial image $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]$ :

$$\hat{c} = \arg\min_c r_c(\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{D}_c \delta_c(\mathbf{x}_i)\|_2^2$$

# Partial Re-ID

## Example of AMC used for partial person matching



# Partial Re-ID

## Global-to-local Matching

### — Sliding Window Matching (SWM)

- Set up a sliding window of the same size as the probe image.
- Search for the most similar image region within each gallery image.
- Use L1-norm to measure the distance.

## Fusion Matching

### — AMC-SWM

$$\mathbf{S}_{dist} = \gamma \mathbf{R}_{dist} + (1-\gamma)\mathbf{L}_{dist}$$

$\mathbf{R}_{dist}$ → AMC distance

$\mathbf{L}_{dist}$ → SMM distance

# Partial Re-ID

## A New Partial REID dataset

**(new collection, released now: http://isee.sysu.edu.cn/resource):**

- 600 images of 60 people
- 5 full-body images and 5 partial images per person

Fig. Examples of partial person images (first row), and the input partial part annotated by an operator for recognition (second row) and the corresponding full-body images (third row).

# Partial Re-ID

## Two Simulated datasets: P-iLIDs and P-CAVIAR

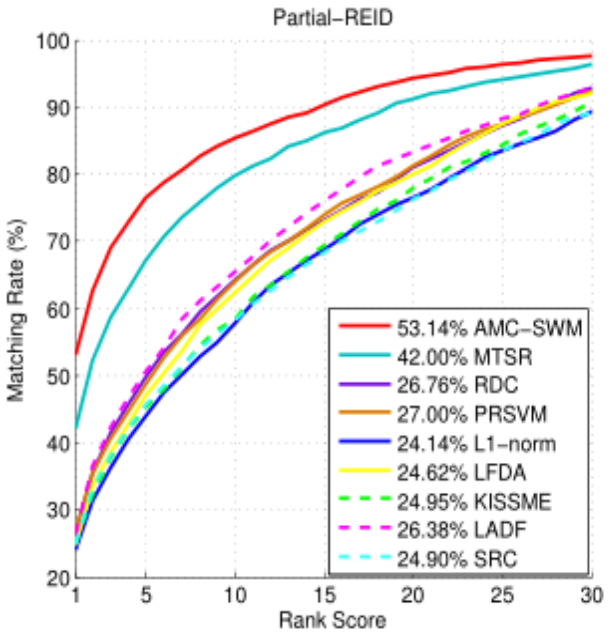
- Based on i-LIDS (476 images of 119 people) & CAVIAR (1220 images of 72 people).
- Randomly selected half of all the images of each person and replaced them with the partial images.

**Fig. Examples of partial person images (firstrow) and the corresponding full images (second row). From left to right, columns 1–3 are from P-i-LIDS, and columns 4–6 from P-CAVIAR.**

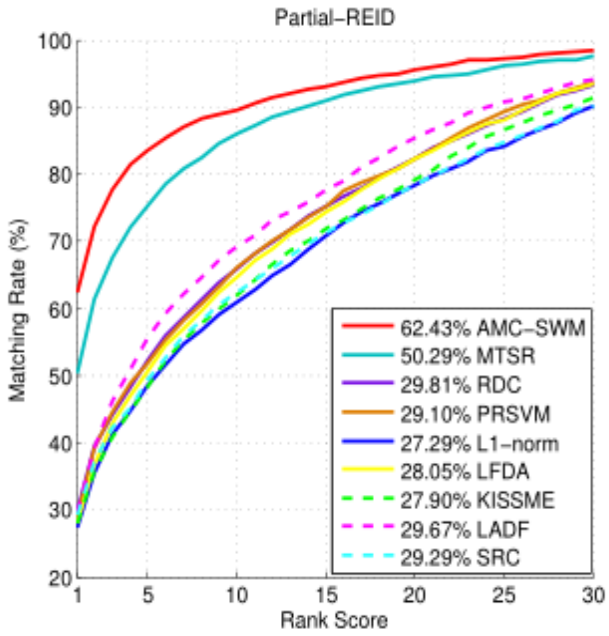
# Partial Re-ID

## Results on Partial REID dataset

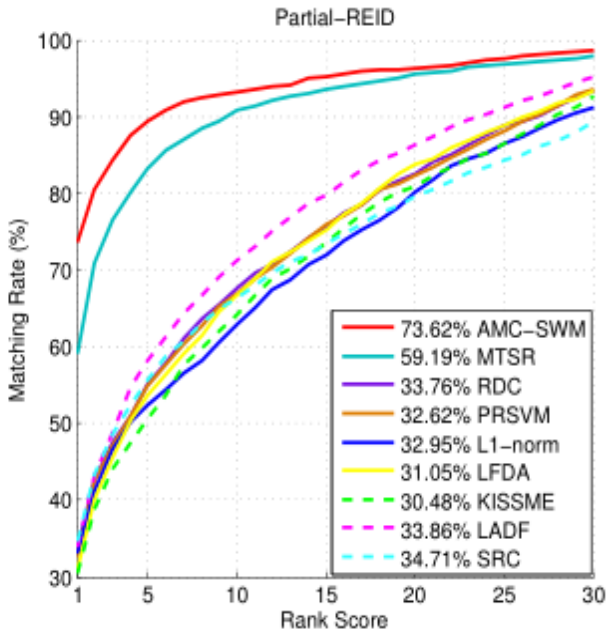
- The test sets were randomly selected using 70% of the individuals.
- Both single-shot and multi-shot experiments were conducted.




(a) N=1,CMC



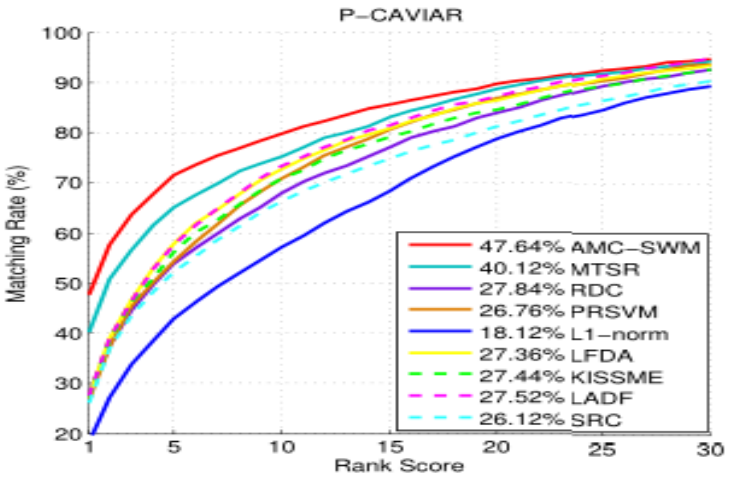

(b) N=2,CMC



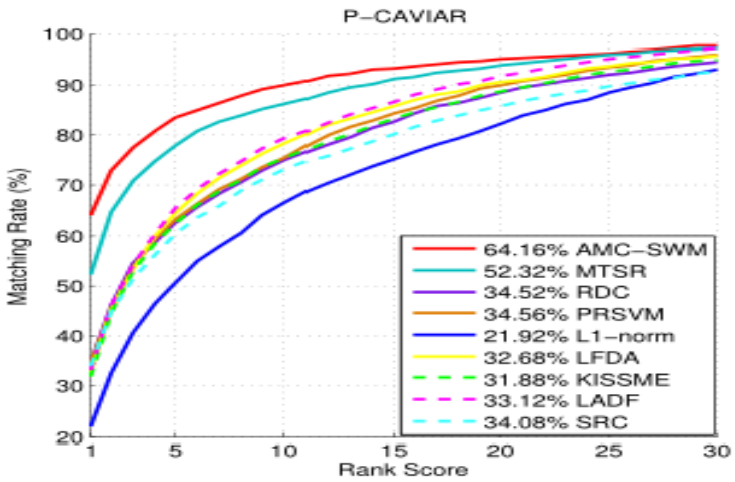

(c) N=5,CMC

# Partial Re-ID

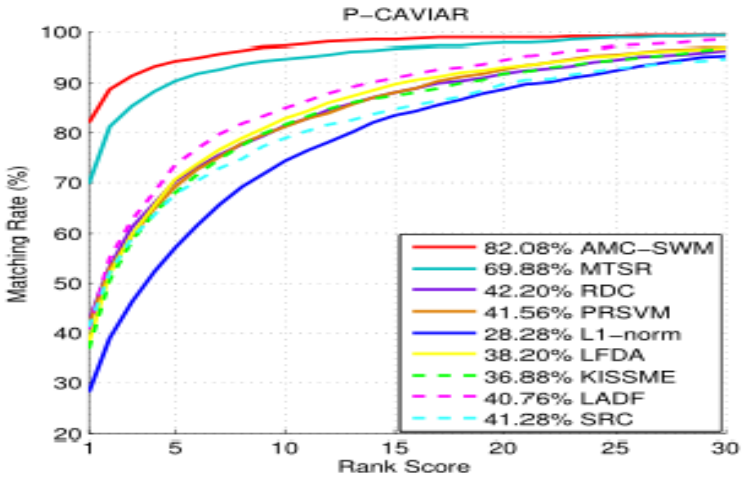
## Results on two simulated datasets



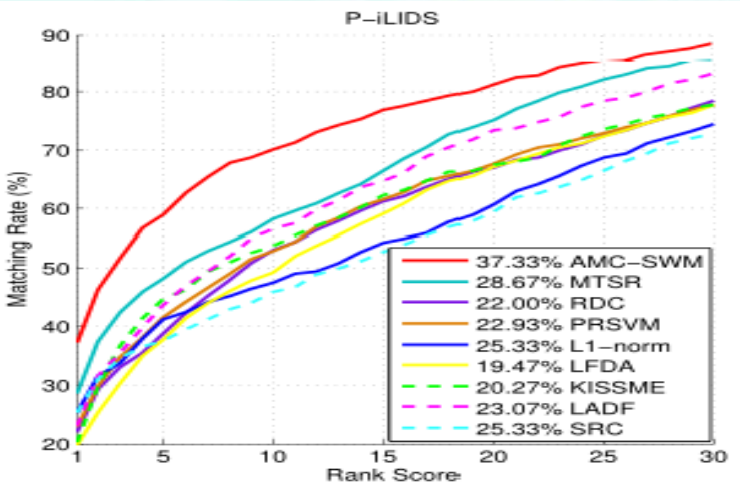

(a) N=1,CMC




(b) N=2,CMC




(c) N=5,CMC



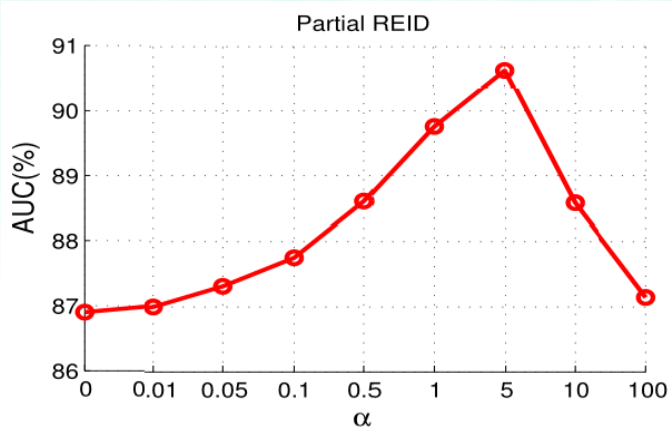

(a) N=1,CMC

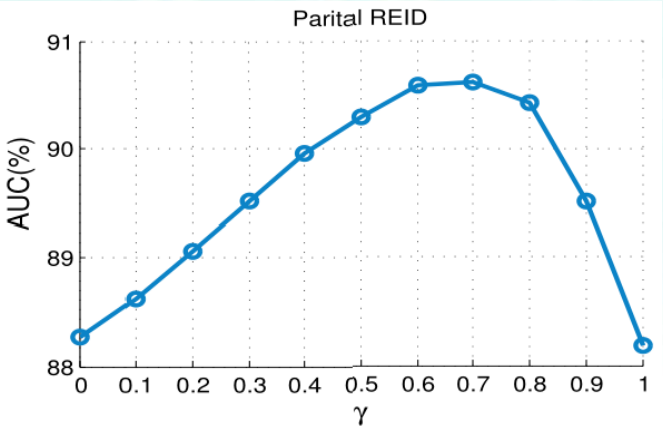
# Partial Re-ID

## Evaluation of the two matching components

| Methods | N=1 | | | N=2 | | | N=5 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $r=1$ | $r=5$ | $r=10$ | $r=1$ | $r=5$ | $r=10$ | $r=1$ | $r=5$ | $r=10$ | |
| AMC-SWM | **53.14** | **76.43** | **85.29** | **62.43** | **83.62** | **89.57** | **73.62** | **89.43** | **93.29** | Partial REID |
| AMC | 45.19 | 70.29 | 81.00 | 53.48 | 77.43 | 87.19 | 63.86 | 86.57 | 91.48 | |
| SWM | 47.24 | 71.05 | 80.57 | 56.00 | 77.67 | 86.10 | 65.24 | 85.10 | 90.91 | |
| AMC-SWM | **47.64** | **71.52** | **79.80** | **64.16** | **83.44** | **89.92** | **82.08** | **94.40** | **97.24** | P-CAVIAR |
| AMC | 44.72 | 67.52 | 78.32 | 59.48 | 82.00 | 88.68 | 78.84 | 92.20 | 95.56 | |
| SWM | 44.16 | 65.08 | 74.72 | 58.88 | 78.12 | 85.96 | 76.92 | 91.28 | 95.32 | |
| AMC-SWM | **37.33** | **59.07** | **70.13** | - | - | - | - | - | - | P-iLIDS |
| AMC | 31.87 | 52.80 | 59.60 | - | - | - | - | - | - | |
| SWM | 35.73 | 54.53 | 65.87 | - | - | - | - | - | - | |

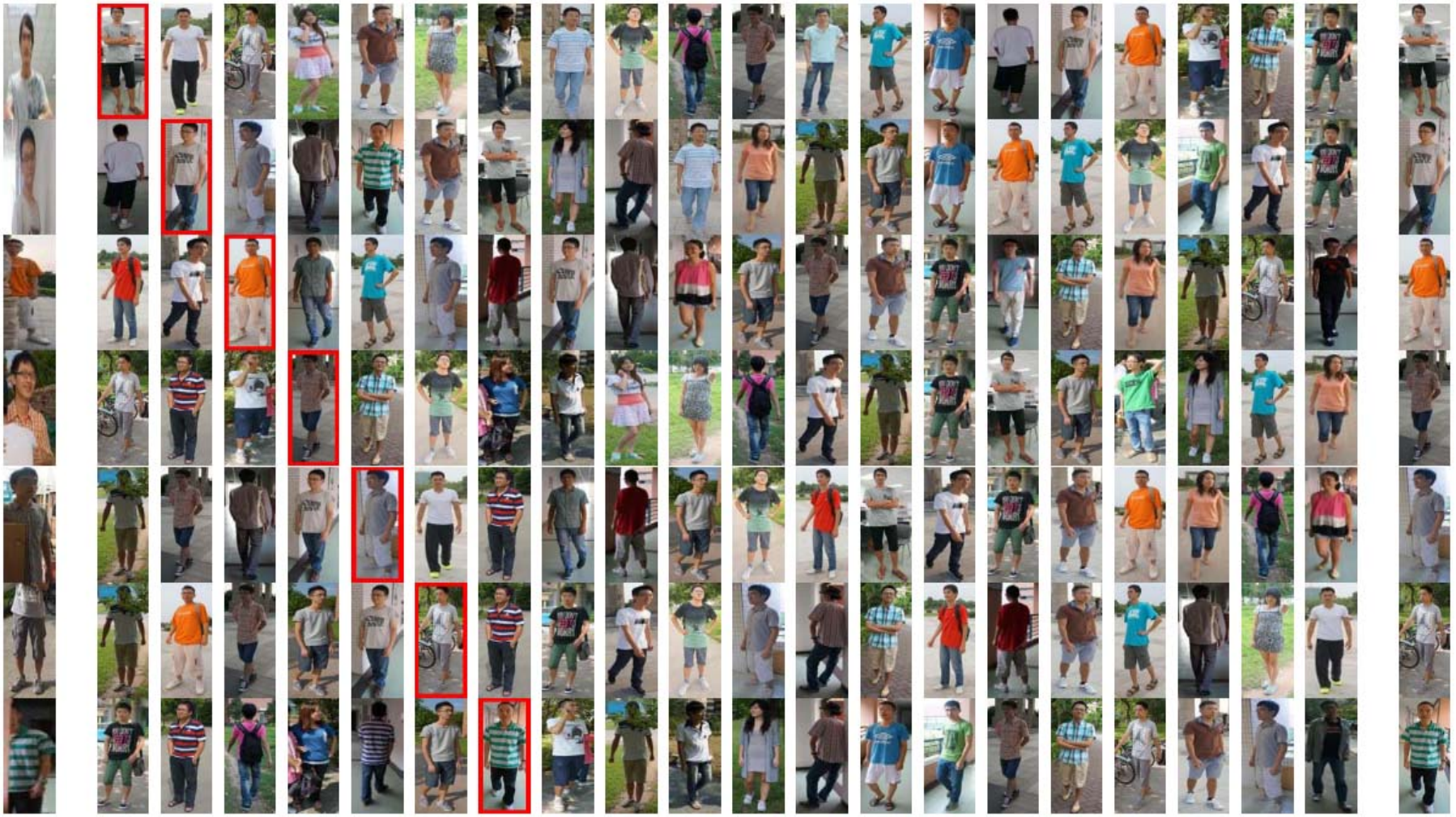
## Parameter Evaluation



# Partial Re-ID

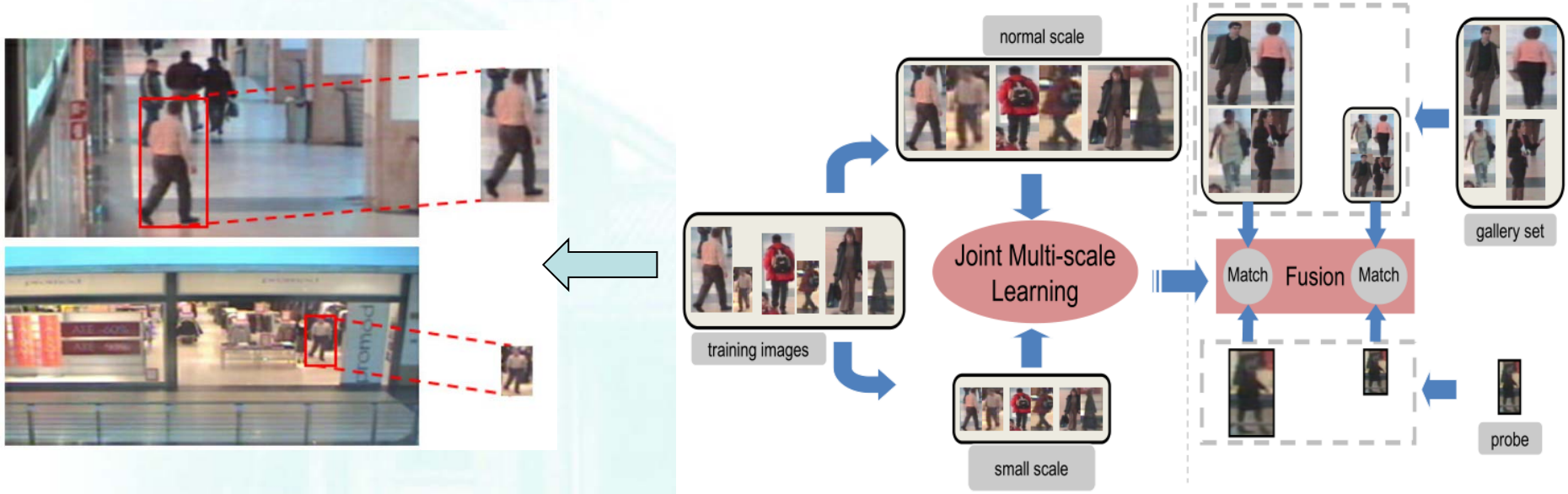
## Illustration of Matching Examples on Partial REID dataset



# Re-identification under Low Resolution

# Low Resolution

- **Our Proposed Multi-scale Learning Model**

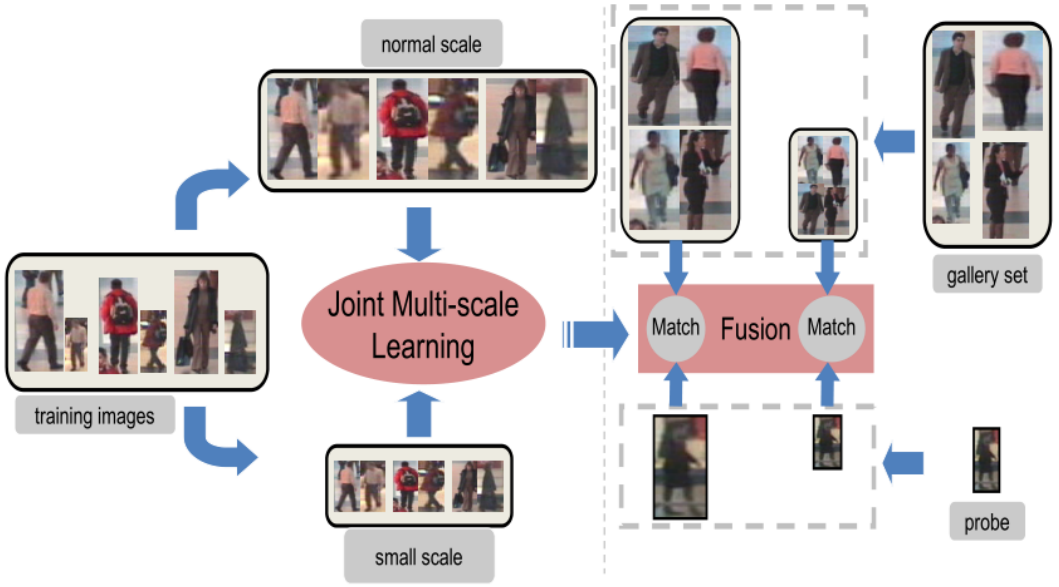



**JUDEA : joint multi-scale discriminant component analysis**

Xiang Li, Wei-Shi Zheng*, Xiaojuan Wang, Tao Xiang, Shaogang Gong. Multi-scale Learning for Low-resolution Person Re-identification. IEEE Conf. on Computer Vision (ICCV), 2015.

# Low Resolution

- **Or Proposed Multi-scale Learning Model**




$$\max_{W_h, W_s} \begin{cases} \mathrm{HCMD}(W_h, W_s)^{-1} \\ \frac{tr(W_h^T S_b^h W_h)}{tr(W_h^T S_w^h W_h)}, \\ \frac{tr(W_s^T S_b^s W_s)}{tr(W_s^T S_w^s W_s)}. \end{cases}$$

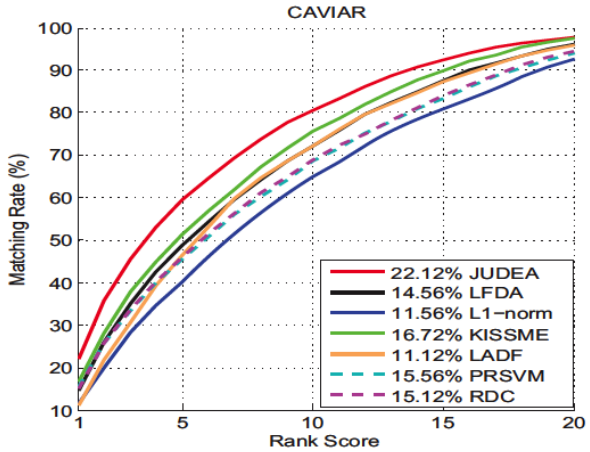
$$\min_{W_h, W_s} \mathrm{HCMD}(W_h, W_s) = \frac{1}{C} \sum_{i=1}^{C} \| W_h^T u_i^h - W_s^T u_i^s \|_2^2$$

Cross-scale Image Domain Alignment

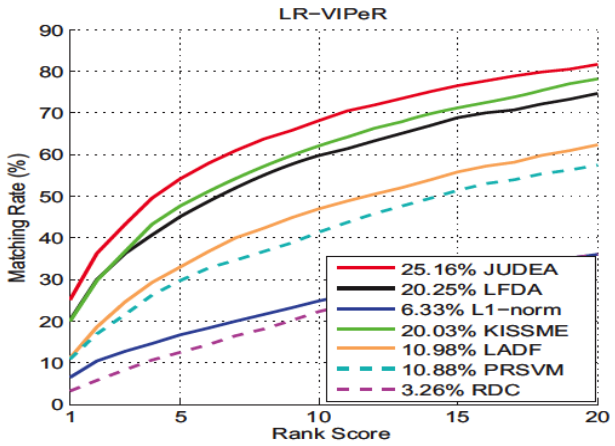
$$\max_{W_h, W_s} \frac{\mathrm{tr}(W_h^T S_b^h W_h + W_s^T S_b^s W_s)}{\mathrm{tr}(W_h^T S_w^h W_h + W_s^T S_w^s W_s) + \alpha \mathrm{HCMD}(W_h, W_s)}$$

# Low Resolution

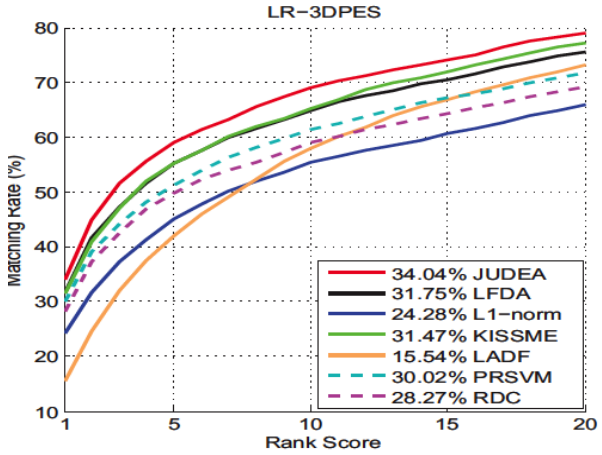
(a) CAVIAR (b) LR-VIPeR (c) LR-3DPES




CMC

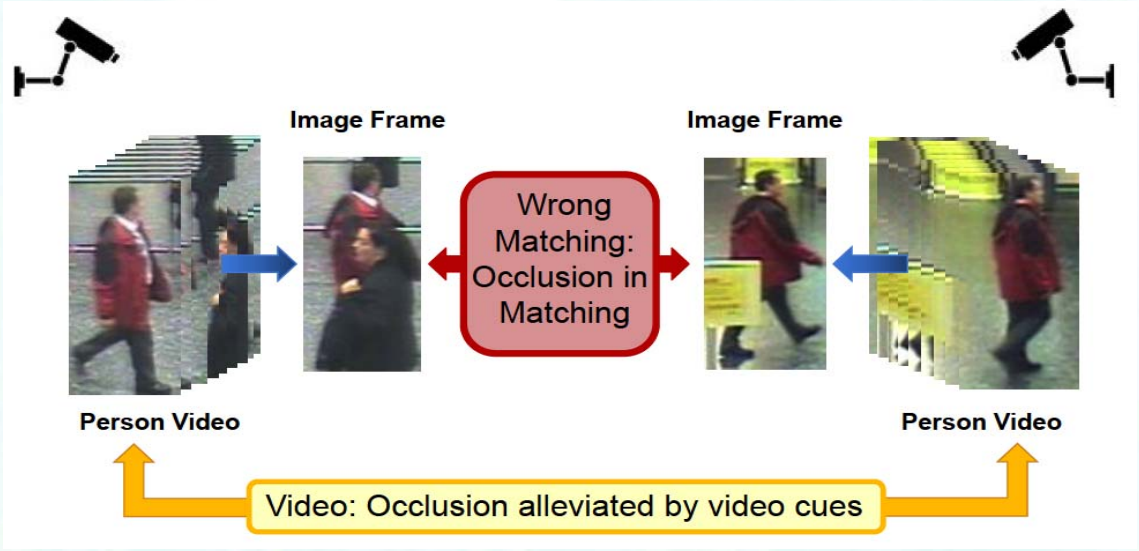



CMC: LR-VIPeR




CMC: LR-3DPES

# Video-based Re-identification




Jinjie You, Ancong Wu, Xiang Li, **Wei-Shi Zheng***. Top-push Video-based Person Re-identification. In IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.

中山大學 SUN YAT-SEN UNIVERSITY

# TOP-PUSH Distance Metric Learning

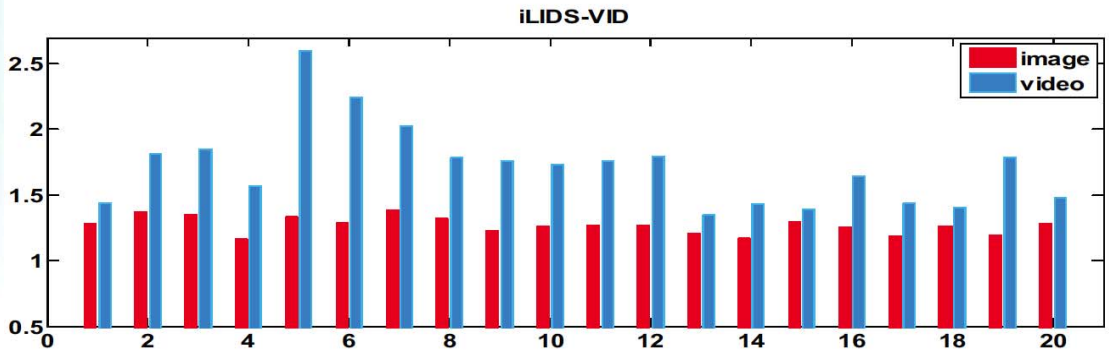
The goal is to learn a Mahalanobis metric:

$$\mathcal{D}(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^\top \mathbf{M}(\vec{x}_i - \vec{x}_j)$$

A triplet hinge loss function:

$$\mathcal{D}(\vec{x}_i, \vec{x}_j) + \rho < \min_{y_k \neq y_i} \mathcal{D}(\vec{x}_i, \vec{x}_k),\ y_i = y_j$$

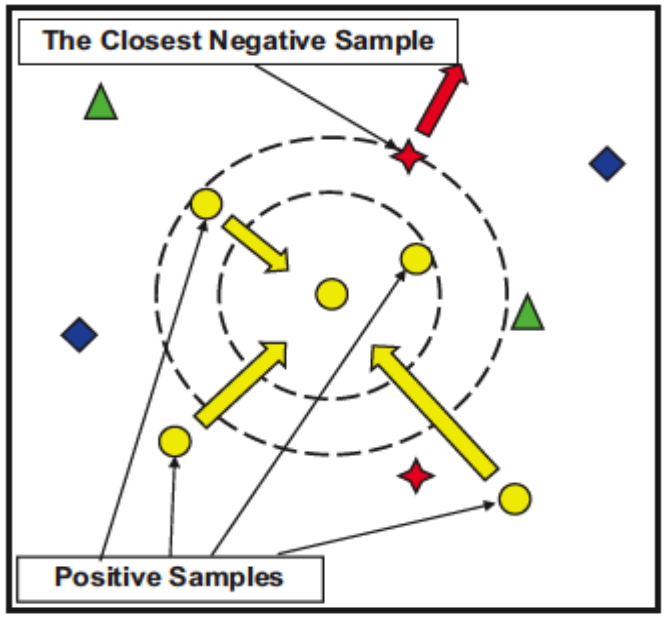

largest intra-class distance $D_w$ / the smallest inter-class distance $D_b$





**Jinjie You, Ancong Wu, Xiang Li, Wei-Shi Zheng*. Top-push Video-based Person Re-identification. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2016.**

# TOP-PUSH Distance Metric Learning

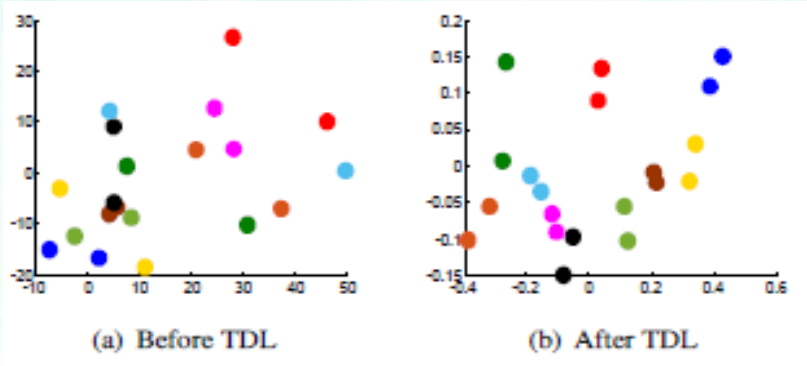



Top-push Distance Metric Model

$$f(D) = (1-\alpha) \sum_{\vec{x}_i, \vec{x}_j, y_i = y_j} \mathcal{D}(\vec{x}_i, \vec{x}_j)$$

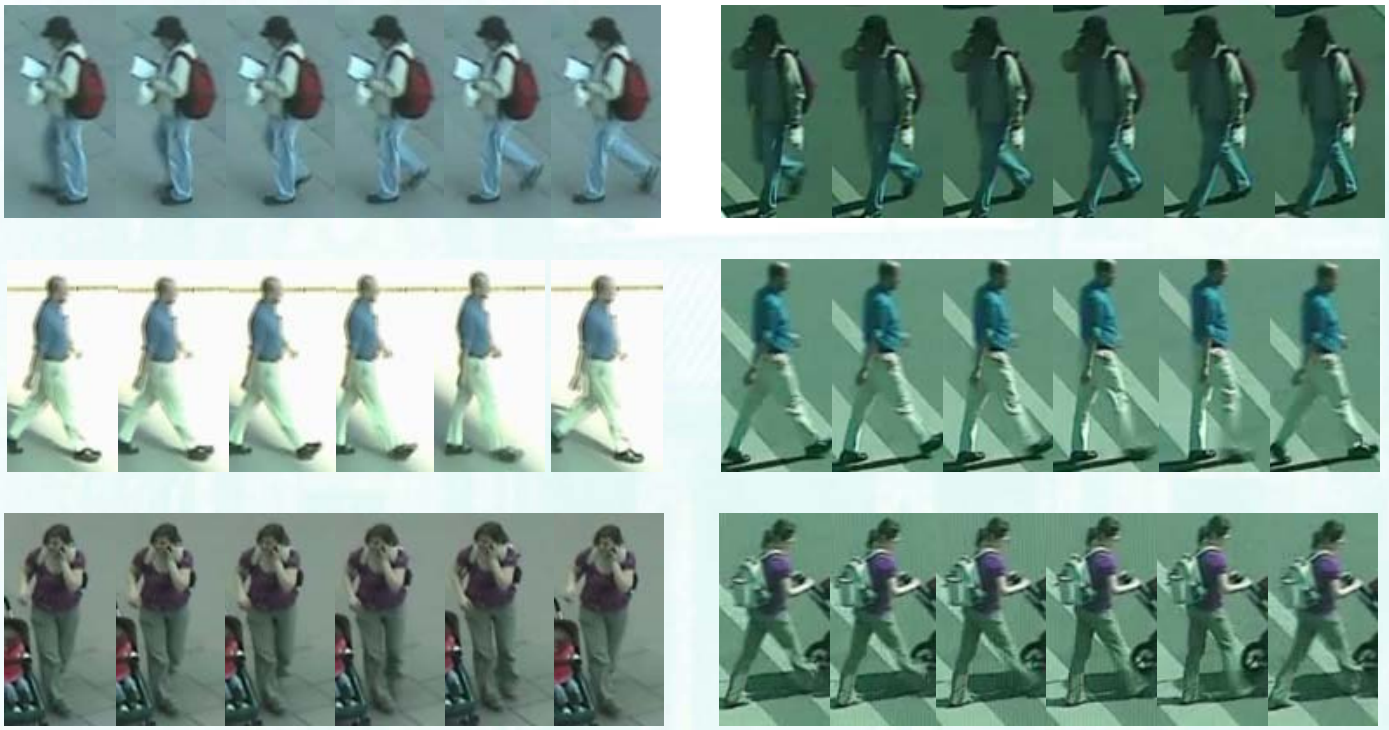
$$+\alpha \sum_{\vec{x}_i, \vec{x}_j, y_i = y_j} \max\{\mathcal{D}(\vec{x}_i, \vec{x}_j) - \min_{y_k \neq y_i} \mathcal{D}(\vec{x}_i, \vec{x}_k) + \rho, 0\}$$

Top-push constraint



# TOP-PUSH Distance Metric Learning

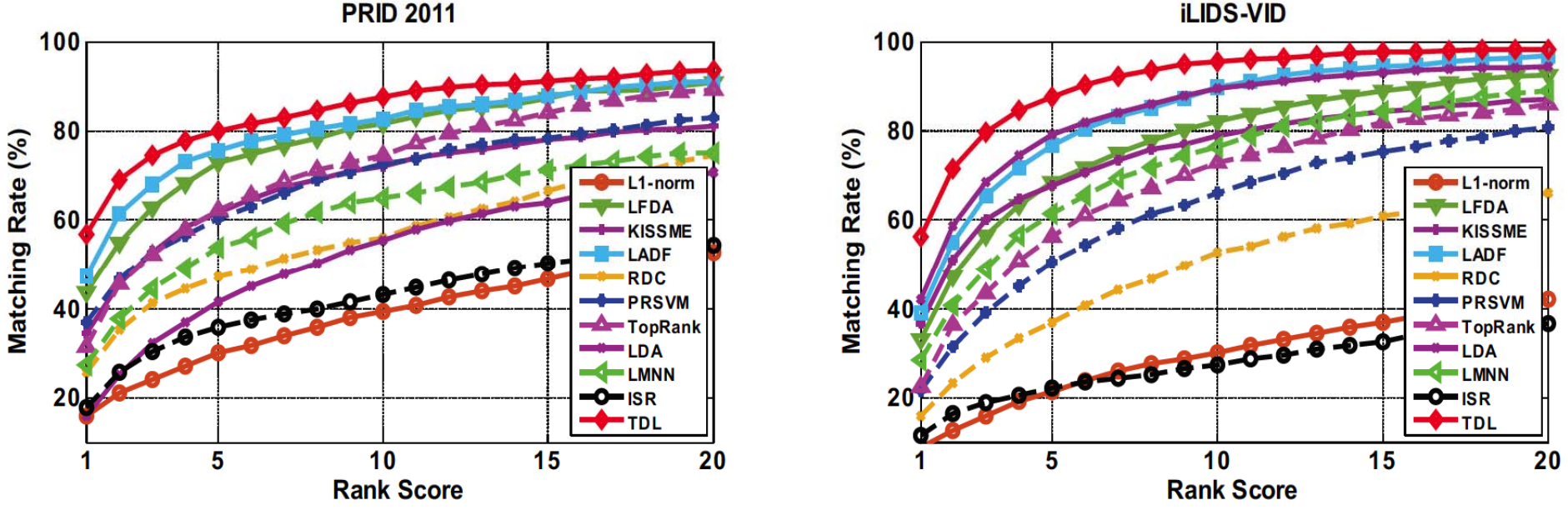
## PRID 2011



# TOP-PUSH Distance Metric Learning

## iLIDS-VID (Extracted from the i-LIDS dataset )

# TOP-PUSH Distance Metric Learning



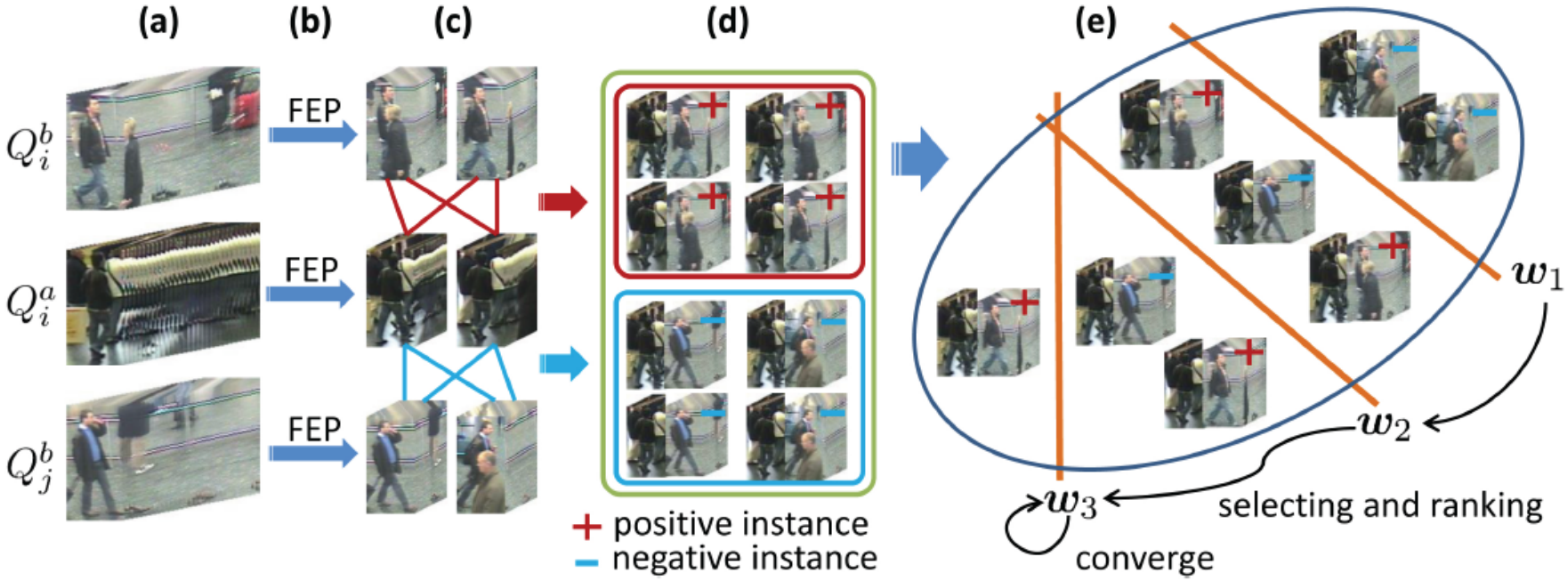

(a) PRID 2011 (b) iLIDS-VID

| Settings | Methods | PRID 2011 | | | | iLIDS-VID | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank | Rank-1 | Rank-5 | Rank-10 | Rank-20 | Rank-1 | Rank-5 | Rank-10 | Rank-20 |
| | LDA [5] | 27.04 | 58.09 | 69.66 | 82.47 | 10.27 | 27.40 | 39.80 | 55.27 |
| | LMNN [34] | 14.38 | 38.09 | 50.22 | 67.19 | 4.47 | 13.20 | 21.60 | 35.47 |

Table 2. Comparison with related methods on PRID 2011 and iLIDS-VID datasets. The matching rate (%) at Rank $i$ means the accuracy of the matching within the top $i$ gallery classes.

# Video-based Re-ID

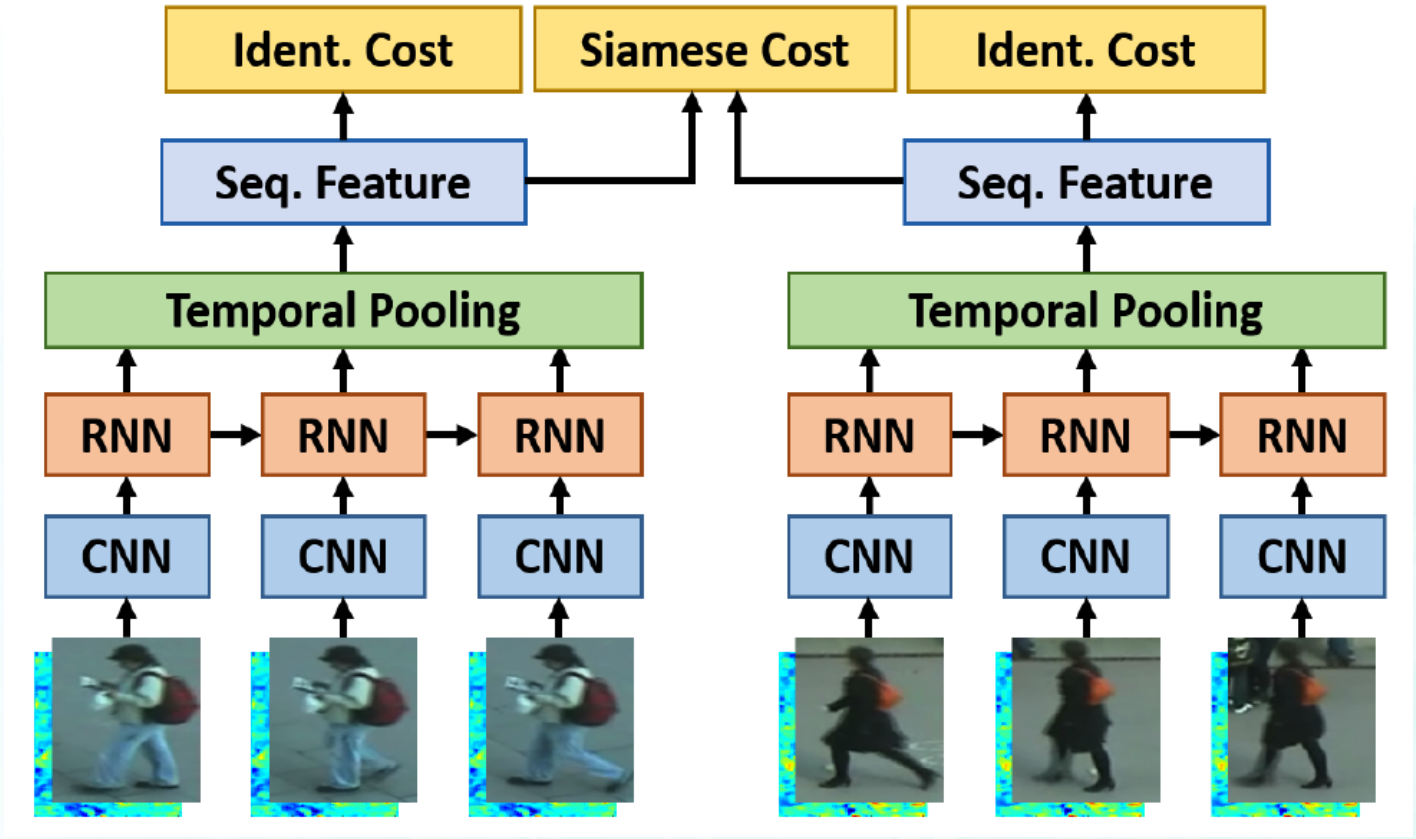
- Discriminative Video fragments selection and Ranking (ECCV2014)




The video-based model automatically selects the most discriminative video fragments and learns a ranking function simultaneously.

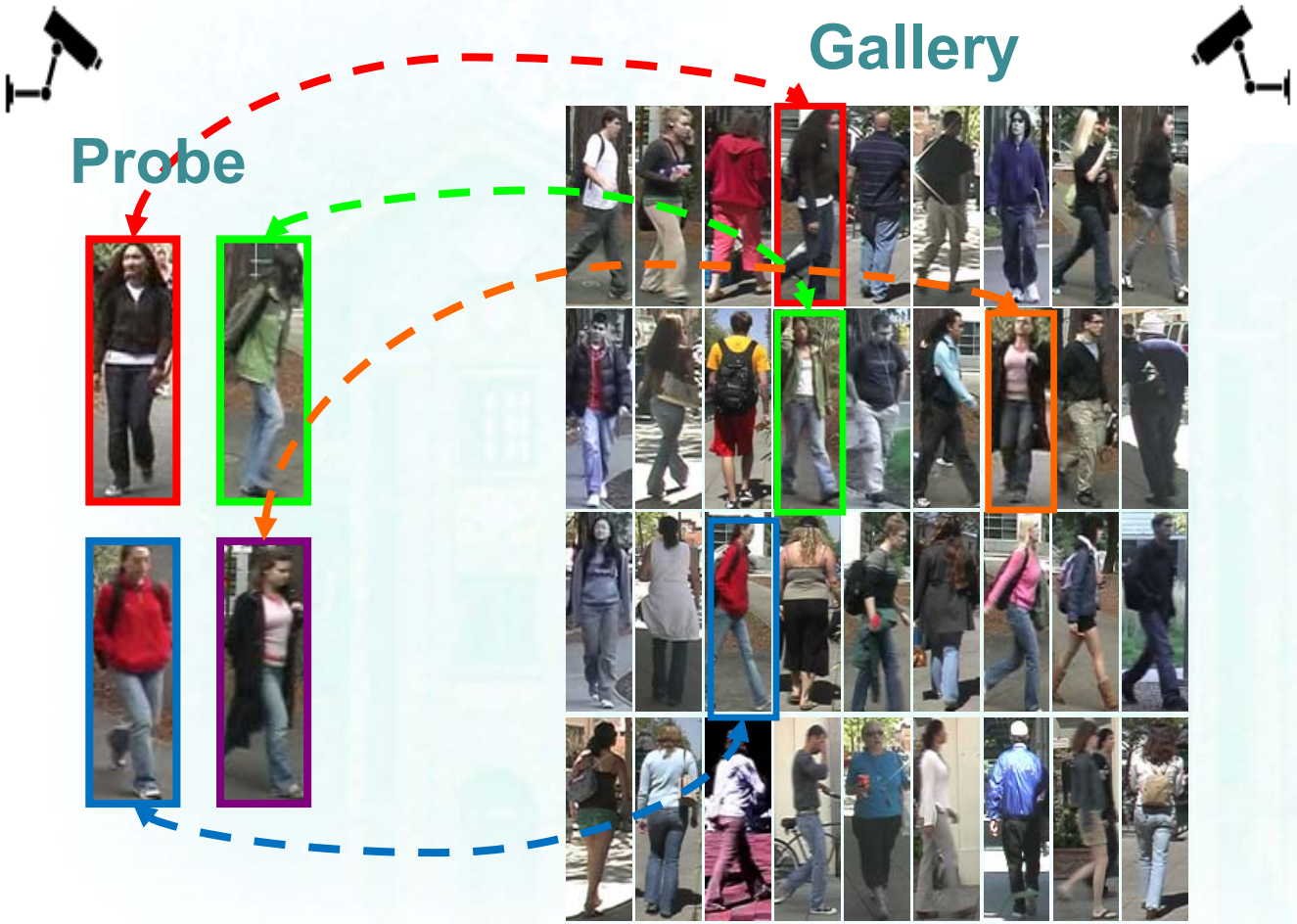
**T. Wang, S. Gong, X. Zhu and S. Wang, "Person Re-Identication by Video Ranking,“ European Conference on Computer Vision (ECCV), 2014**

# Video RE-ID




Niall McLaughlin, Jesus Martinez del Rincon, Paul Miller. Recurrent Convolutional Network for Video-based Person Re-Identification. CVPR 2016

# Person Re-identification: Labelling




**Labelling images across camera views is costly**

Labeling images is costly and even prohibitive in some scenarios

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

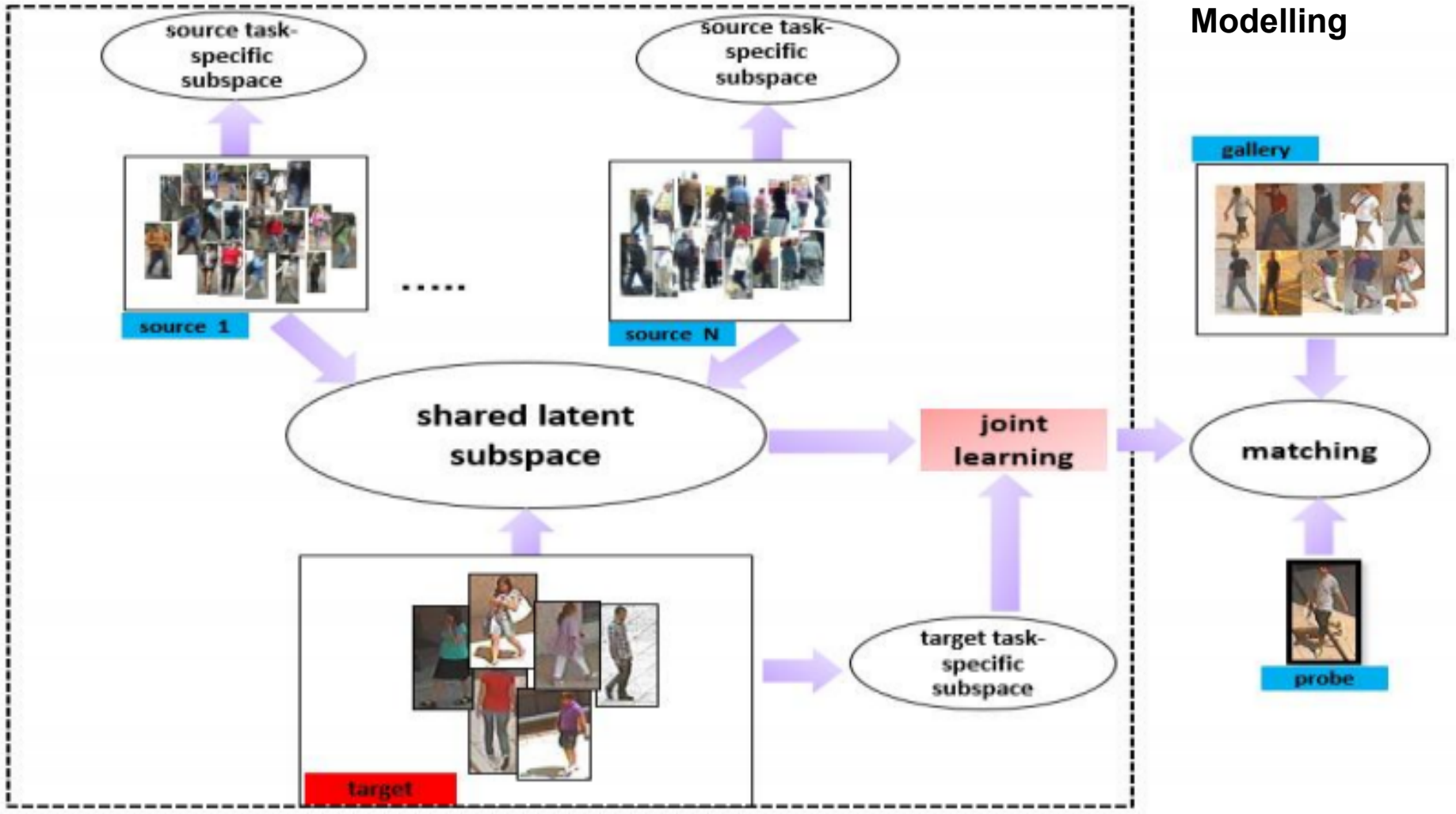
*Cross-scenario Transfer Person Re-identification*

Xiaojuan Wang, Wei-Shi Zheng*, Xiang Li, and Jianguo Zhang. Cross-scenario Transfer Person Re-identification. IEEE Transactions on Circuits and Systems for Video Technology, DOI: 10.1109/TCSVT.2015.2450331, 2015.

中山大學 SUN YAT-SEN UNIVERSITY

# Framework

**An Asymmetric Multi-task Modelling**



# Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets

**source task-specific subspace:**

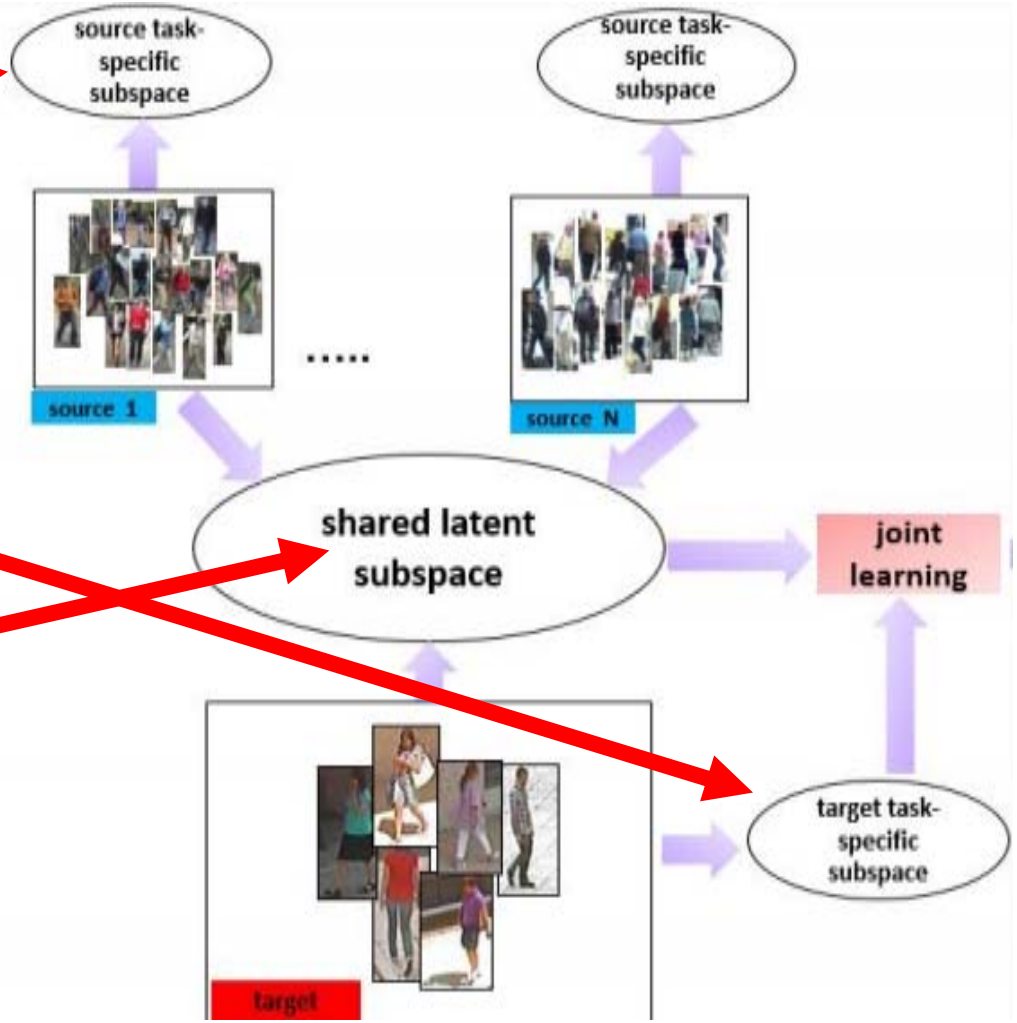
$$\mathbf{W}_s \in \mathrm{R}^{d\times r}$$

**target task-specific subspace:**

$$\mathbf{W}_t \in \mathrm{R}^{d\times r}$$

**shared latent subspace:**

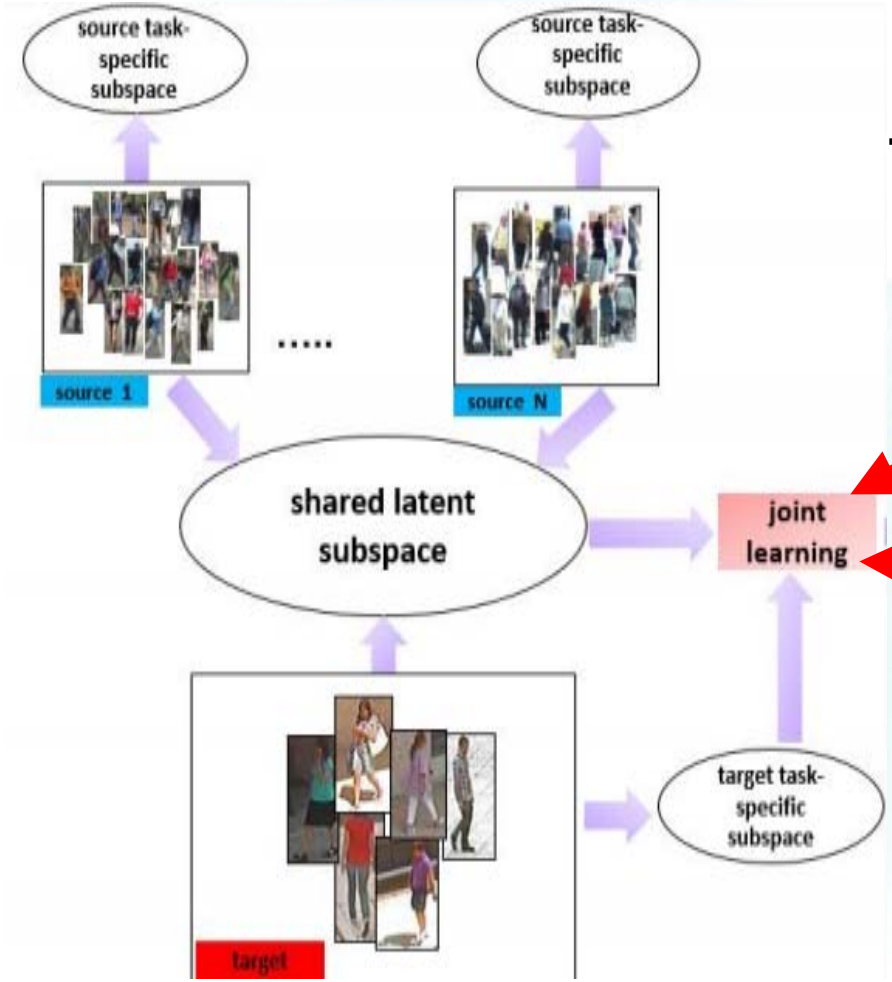
$$\mathbf{W}_0 \in \mathrm{R}^{d\times r}$$



# Cross-scenario Transfer Modeling

Transfer one source dataset

Transfer multiple source datasets




**The projection of a target sample** $\mathbf{x}_t$

$$\mathbf{z}_t = ((1-\beta)\mathbf{W}_0 + \beta\mathbf{W}_t)'\mathbf{x}_t$$

**The projection of a source sample** $\mathbf{x}_s$

$$\mathbf{z}_s = ((1-\beta)\mathbf{W}_0 + \beta\mathbf{W}_s)'\mathbf{x}_s$$

# 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 task

$$\max_{\mathbf{W}_1,\mathbf{W}_2} (1-\gamma)\frac{\mathrm{tr}(\mathbf{W}_1'\mathbf{S}_b^s\mathbf{W}_1)}{\mathrm{tr}(\mathbf{W}_1'\mathbf{S}_w^s\mathbf{W}_1)} + \gamma\frac{\mathrm{tr}(\mathbf{W}_2'\mathbf{S}_b^t\mathbf{W}_2)}{\mathrm{tr}(\mathbf{W}_2'\mathbf{S}_w^t\mathbf{W}_2)}, \quad 0 \le \gamma \le 1 \qquad (1)$$

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

non-convex

relaxation

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

## Insight

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

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

## 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

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

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

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

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

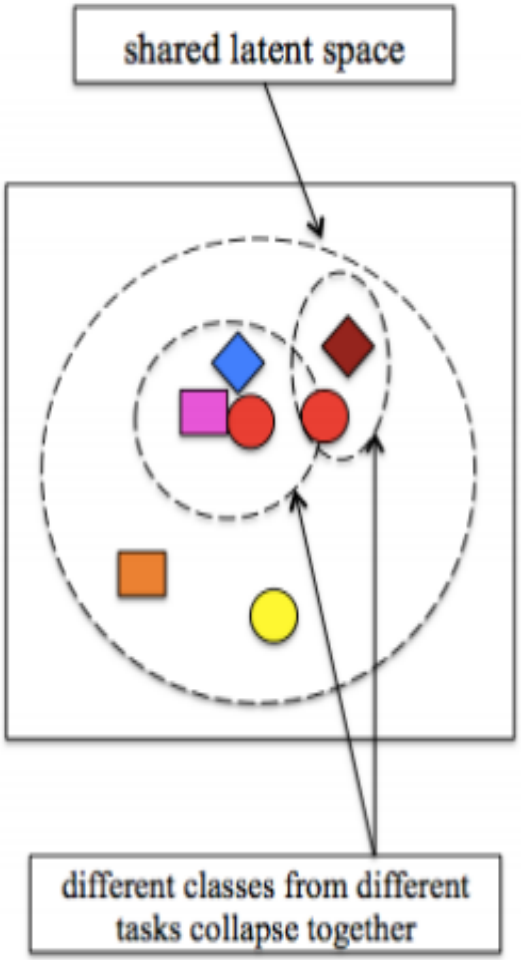
$$\mathbf{A} = (1-\gamma)(\frac{1}{m}\sum_{i=1}^{m}(\mathbf{\Theta}_s^i)'\mathbf{S}_b^{s,i}\mathbf{\Theta}_s^i) + \gamma(\mathbf{\Theta}_t'\mathbf{S}_b^t\mathbf{\Theta}_t) \quad (5a)$$

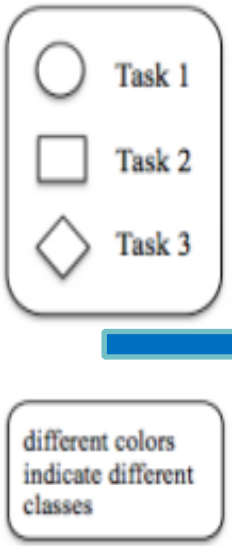
$$\mathbf{B} = (1-\gamma)(\frac{1}{m}\sum_{i=1}^{m}(\mathbf{\Theta}_s^i)'\mathbf{S}_w^{s,i}\mathbf{\Theta}_s^i) + \gamma(\mathbf{\Theta}_t'\mathbf{S}_w^t\mathbf{\Theta}_t). \quad (5b)$$

solution could be obtained by Eq. (4)

# Constrained Asymmetric Multi-task Component Analysis

In the shared latent space, different classes from different tasks could collapse together.




instance *i* from task *k*

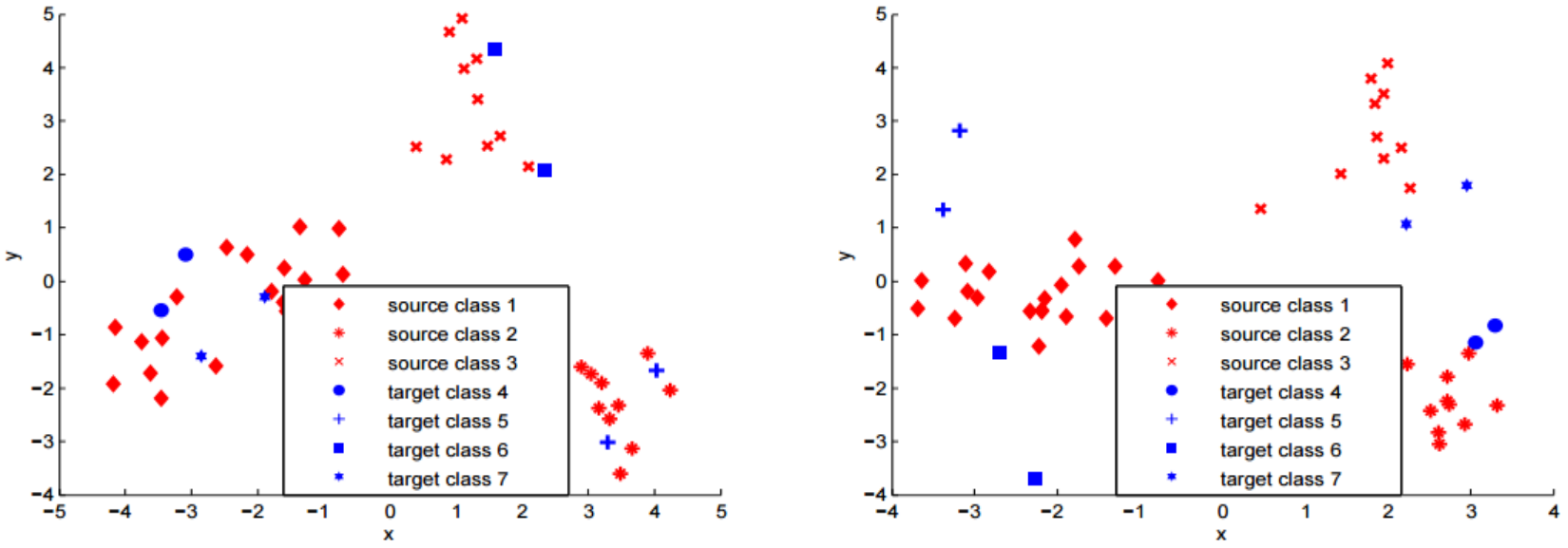
instance *j* from task *l*

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

separate data from different tasks

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

# 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.

# 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 |

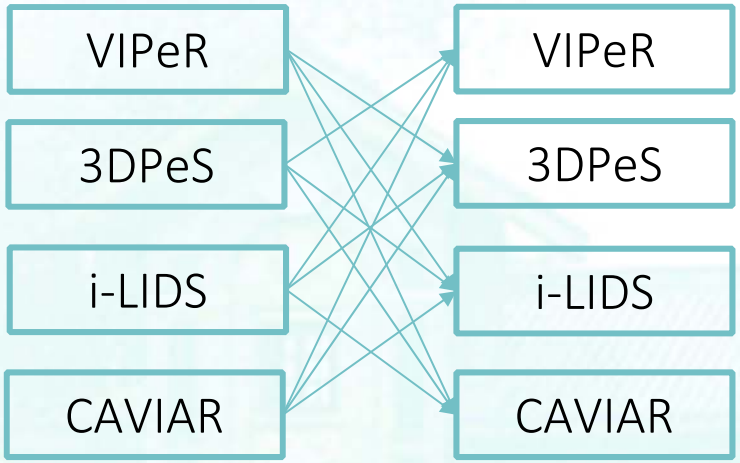
Table 1: Summary of datasets used in the experiments.

(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 target data only (e.g. LFDA_T)
- using source 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:

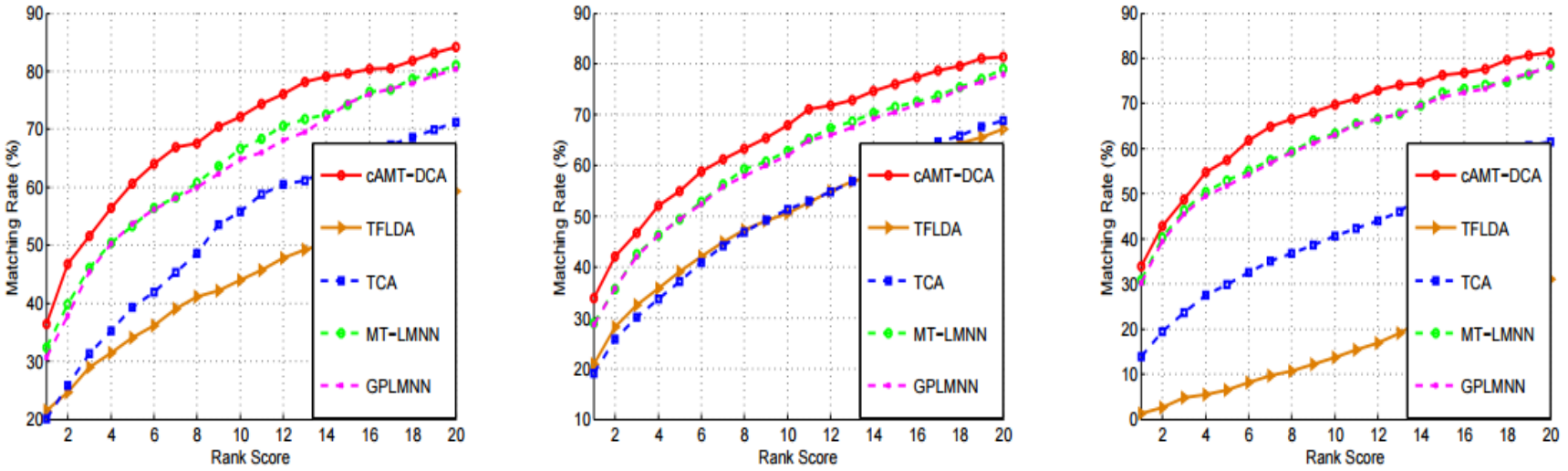
- Only using source dataset for the chosen metric learning algorithms often results in better performance than only using limited target data (except for the case with VIPeR as target dataset).

- 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.

# 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.

# 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.

# Transfer Learning

- Unsupervised Cross-Dataset Transfer Learning for Person Re-identification

$$
\begin{aligned}
&[D^s, D_T^u, D_1^r, \cdots, D_T^r] = \arg\min \\
&\sum_{t=1}^{T-1} \eta_t^2 \{\|X_t - D^s A_t^s\|_F^2 + \|X_t - D^s A_t^s - D_t^r A_t^r\|_F^2\} + \\
&\|X_T - D^s A_T^s\|_F^2 + \|X_T - D^s A_T^s - D_T^u A_T^u\|_F^2 + \\
&\|X_T - D^s A_T^s - D_T^u A_T^u - D_T^r A_T^r\|_F^2 + \\
&\lambda \sum_{t=1}^{T} \sum_{i,j=1}^{N_t} w_{t,i,j} \left\|a_{t,i}^s - a_{t,j}^s\right\|^2 + \lambda \sum_{i,j=1}^{N_T} w_{T,i,j} \left\|a_{T,i}^u - a_{T,j}^u\right\|^2, \\
&s.t.\ \|d_i^s\|_2^2 \le 1,\ \left\|d_{T,i}^u\right\|_2^2 \le 1,\ \left\|d_{t,i}^r\right\|_2^2 \le 1,\ \forall i, t
\end{aligned}
$$

A Dictionary-learning-based model for cross-dataset transfer.

**P. Peng, T. Xiang et al., "Unsupervised Cross-Dataset Transfer Learning for Person Re-identification", CVPR, 2016**

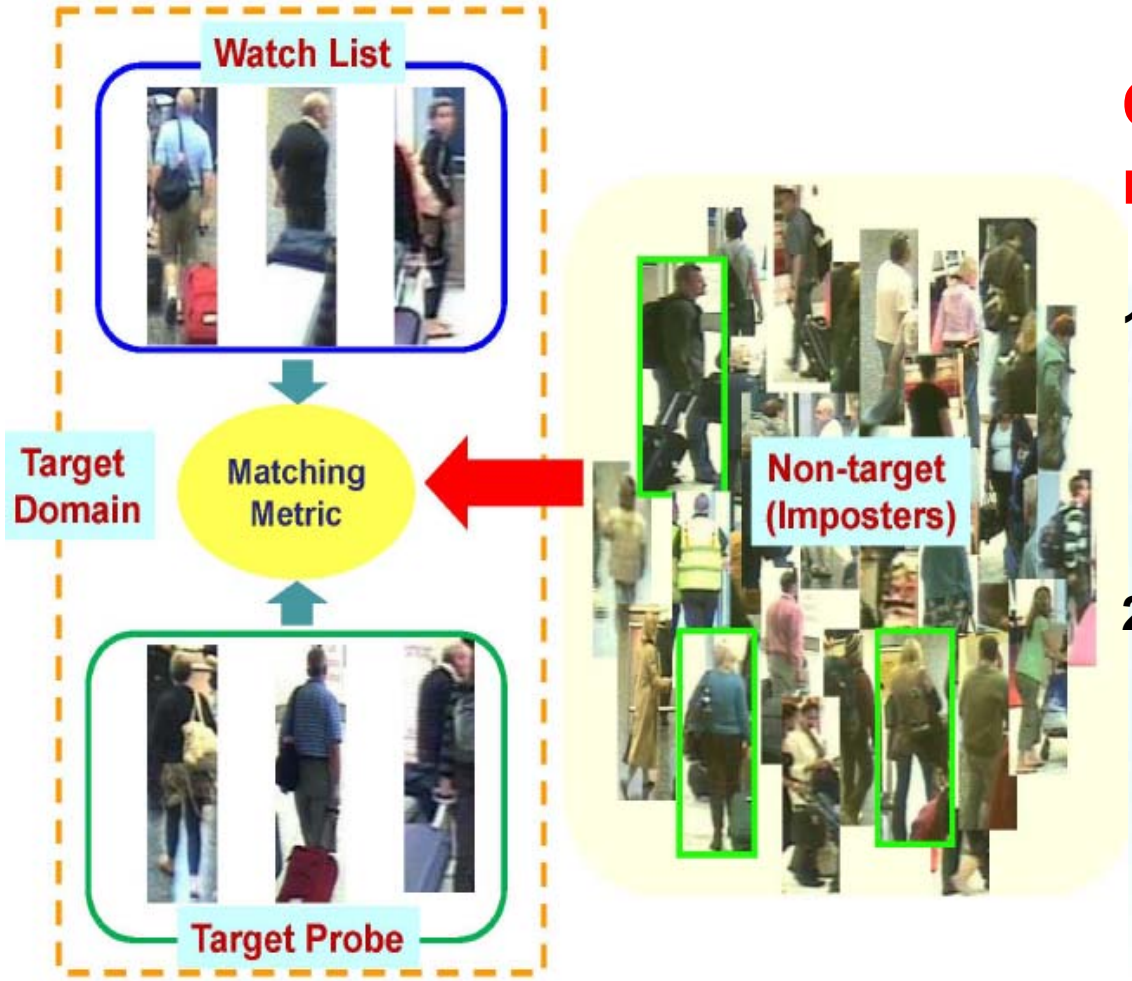
# In real world, there are quite a lot of imposters, and only a few guys are target to track


**Wei-Shi Zheng**, Shaogang Gong, and Tao Xiang. Towards Open-World Person Re-Identification by One-Shot Group-based Verification. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), vol. 38, no. 3, pp. 591-606, 2016.

中山大學 SUN YAT-SEN UNIVERSITY

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

- Motivation



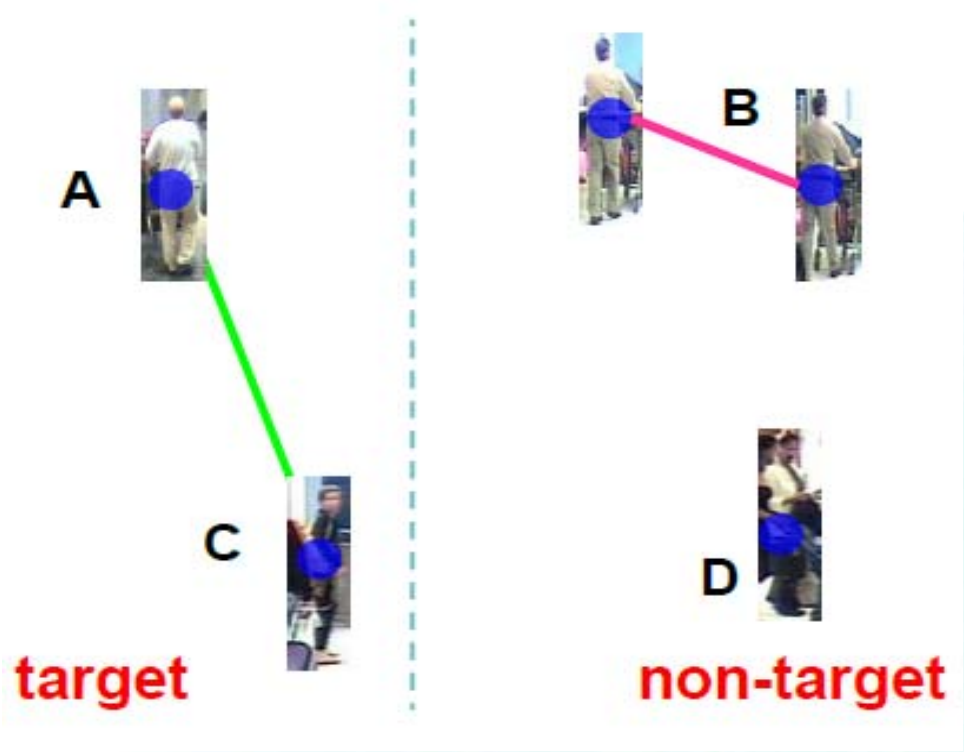

**Open-world person re-identification setting**

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**

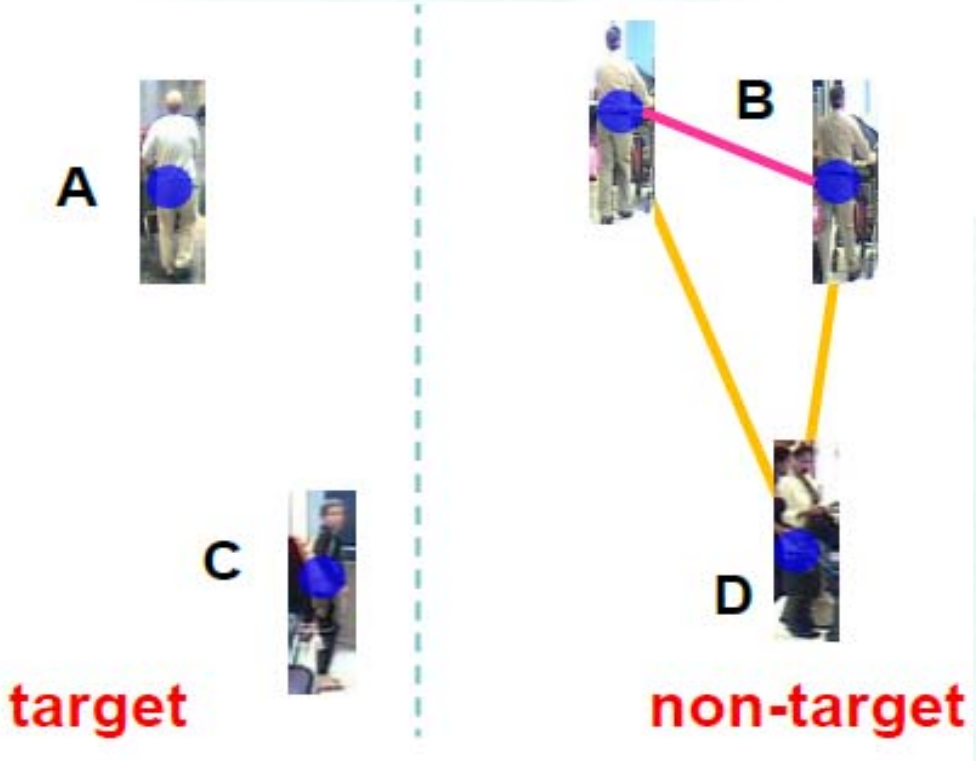



**Enrich intra-class variation**

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

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

- **Knowledge to transfer**

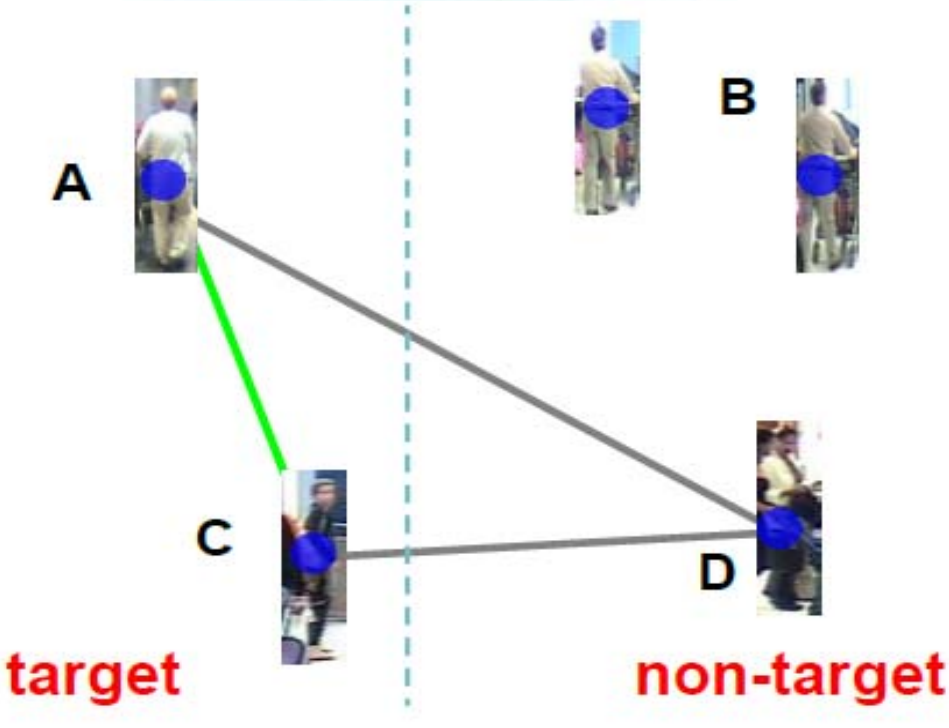



**Enrich inter-class variation**

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

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

- **Knowledge to transfer**

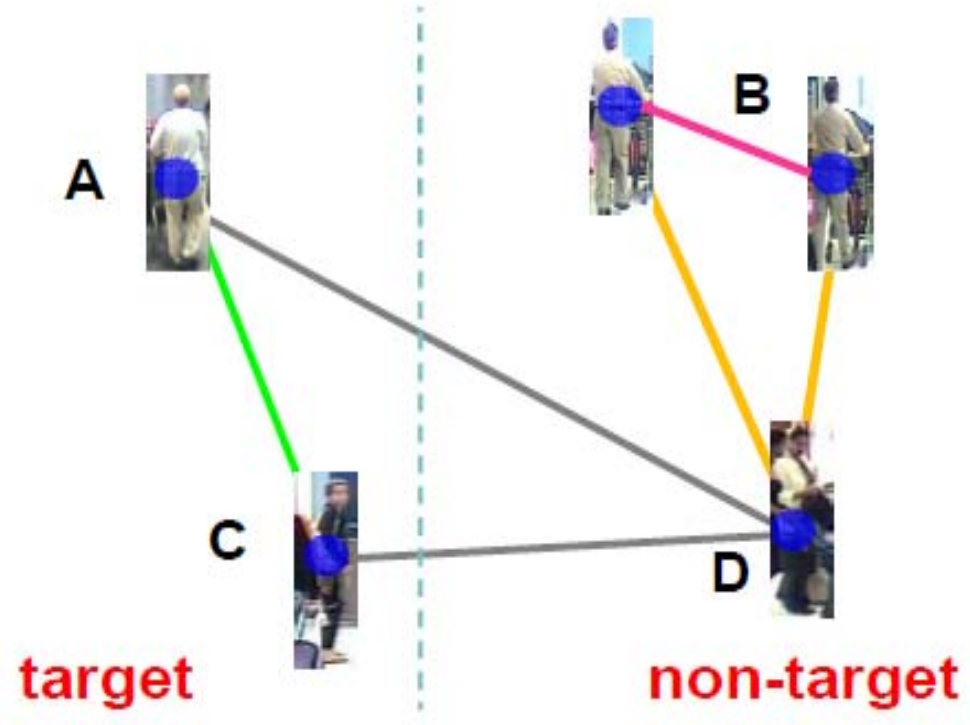



**Enrich group separation**

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

# 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)**

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

- **Criterion**

Similar source person image

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

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

Source intra-class

Target inter-class

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

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

$$\mathbb{O}_g(\mathbf{x}_t) = \{(\mathbf{x}_{t_j},\mathbf{x}_s,\mathbf{x}_{t'})| \; g(\mathbf{x}_t,\mathbf{x}_{t_j}) = 1,\; y_{t_j} = y_s, y_t \neq y_{t'},\; 1 \le t' \le N_T,\; N_T+1 \le t_j, s \le N\}$$

$$\mathbb{O}_a(\mathbf{x}_t) = \{(\mathbf{x}_s,\mathbf{x}_{s'},\mathbf{x}_{s''})|g(\mathbf{x}_t,\mathbf{x}_s) = 1,\; y_{s'} = y_s,\; y_{s''} \neq y_s,\; N_T+1 \le s,s',s'' \le N\}$$

$$\mathbb{O}_b(\mathbf{x}_t) = \{(\mathbf{x}_{t'},\mathbf{x}_s)|g(\mathbf{x}_t,\mathbf{x}_s) = 0,\; y_t \neq y_{t'},\; y_s \neq y_t,\; 1 \le t' \le N_T,\; N_T+1 \le s \le N\}$$

$$g(\mathbf{x}_s,\mathbf{x}_t) = \begin{cases} 1, & s(\mathbf{x}_s,\mathbf{x}_t) \ge h; \\ 0, & s(\mathbf{x}_s,\mathbf{x}_t) < h. \end{cases}$$

$$s(\mathbf{x}_s,\mathbf{x}_t) = \frac{|\mathbf{x}_s^T\mathbf{x}_t|}{||\mathbf{x}_s||||\mathbf{x}_t||}$$

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

## Local Relative Distance Comparison

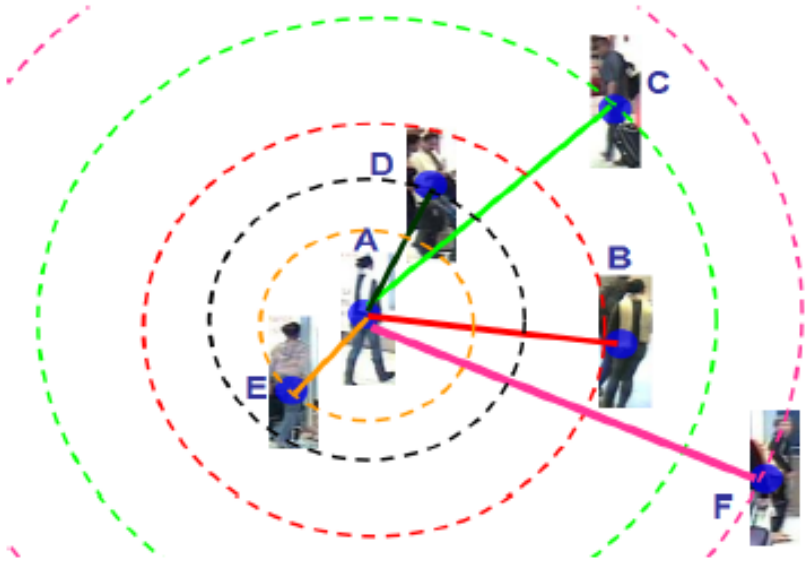



Fig. 3. Illustration of our local relative comparison. Among the six images, A and B belong to the same person whilst the other four are of four other people. See text for more details.

**constraining all the relative distance comparisons around the neighbourhood of a difference dataset**

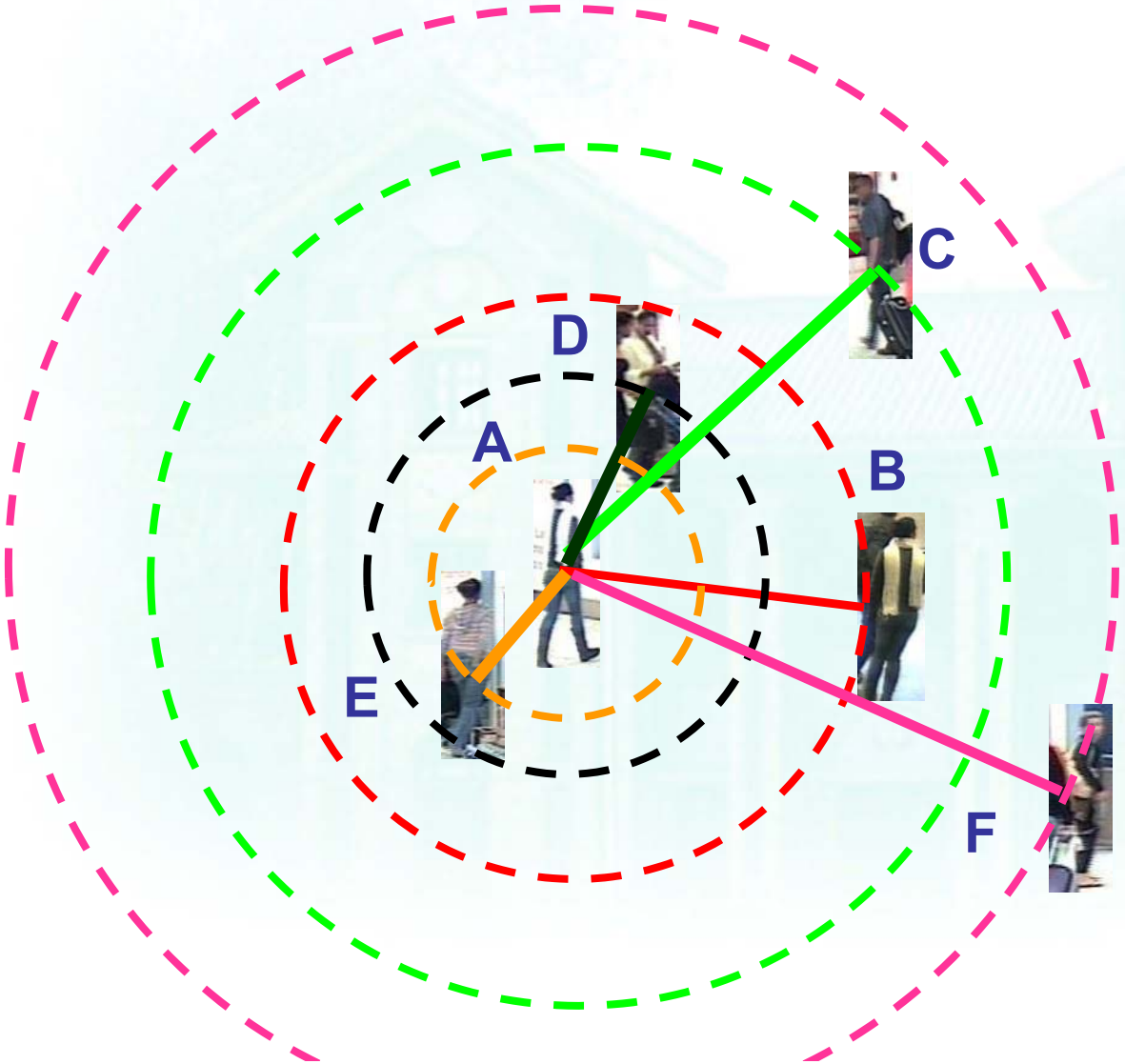
$\mathcal{D}^{+}_{y_i}(\mathbf{x}_i)$ **or** $\mathcal{D}^{-}_{y_i}(\mathbf{x}_i)$

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

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. $\mathcal{D}^{-}_{y_i}(\mathbf{x}_i) = \{(\mathbf{x}_q - \mathbf{x}_i) \mid y_q \neq y_i \ \& \ 1 \leq q \leq N_T\}$;

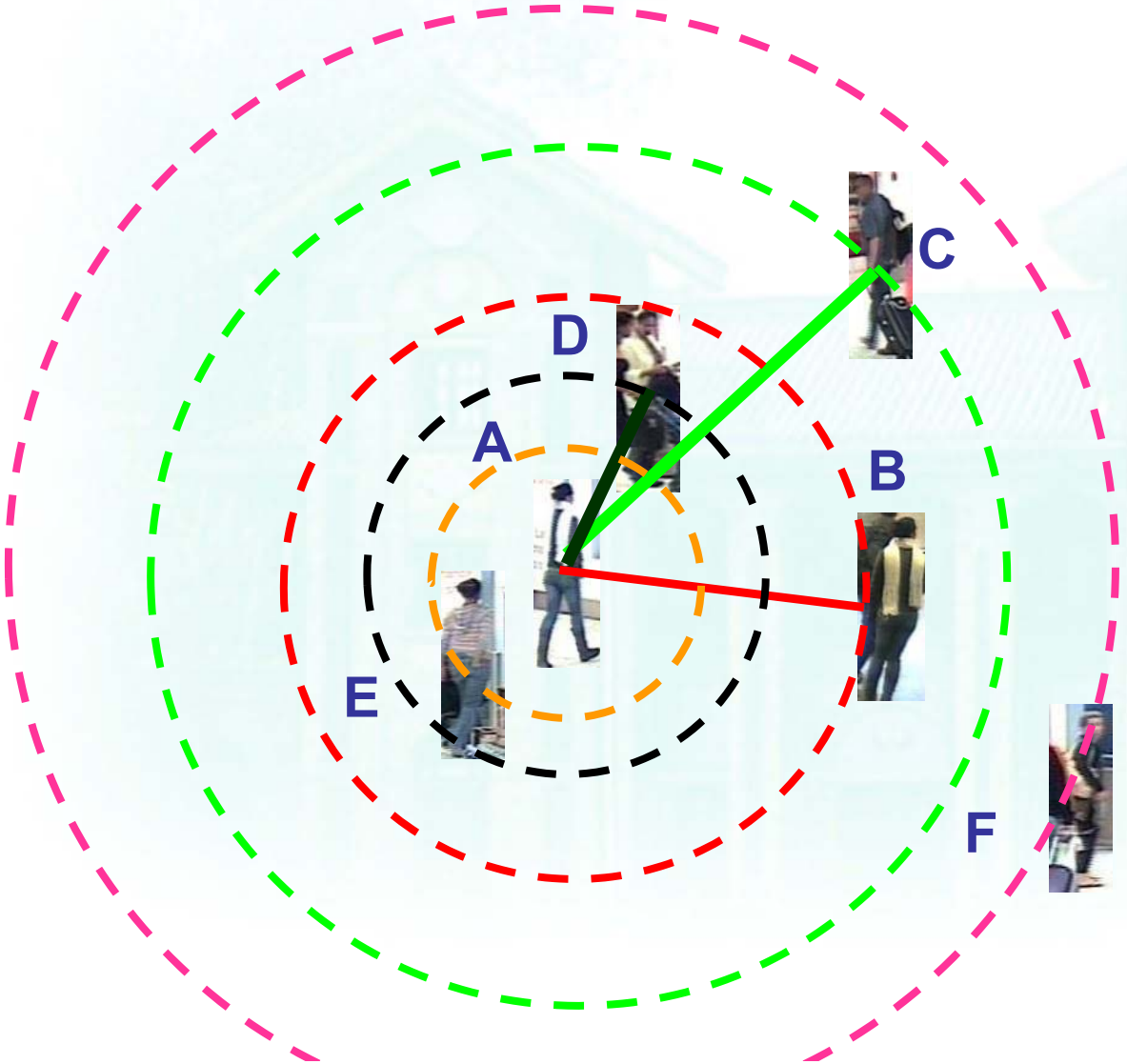
# One-Shot Open-World Group-based Re-id

- **Local Modelling: Remained Comparison**



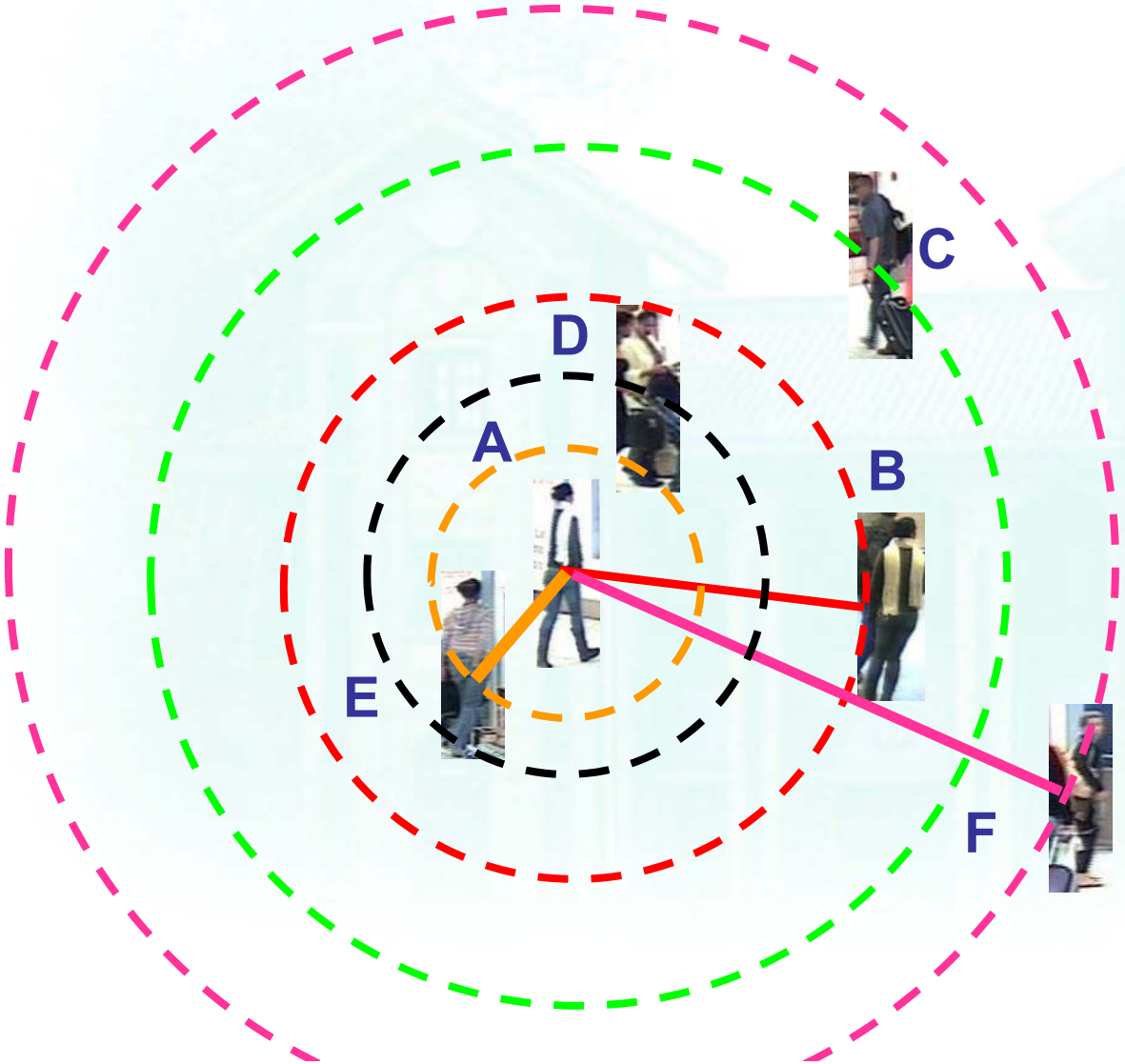
# One-Shot Open-World Group-based Re-id

- **Local Modelling: Remained Comparison**

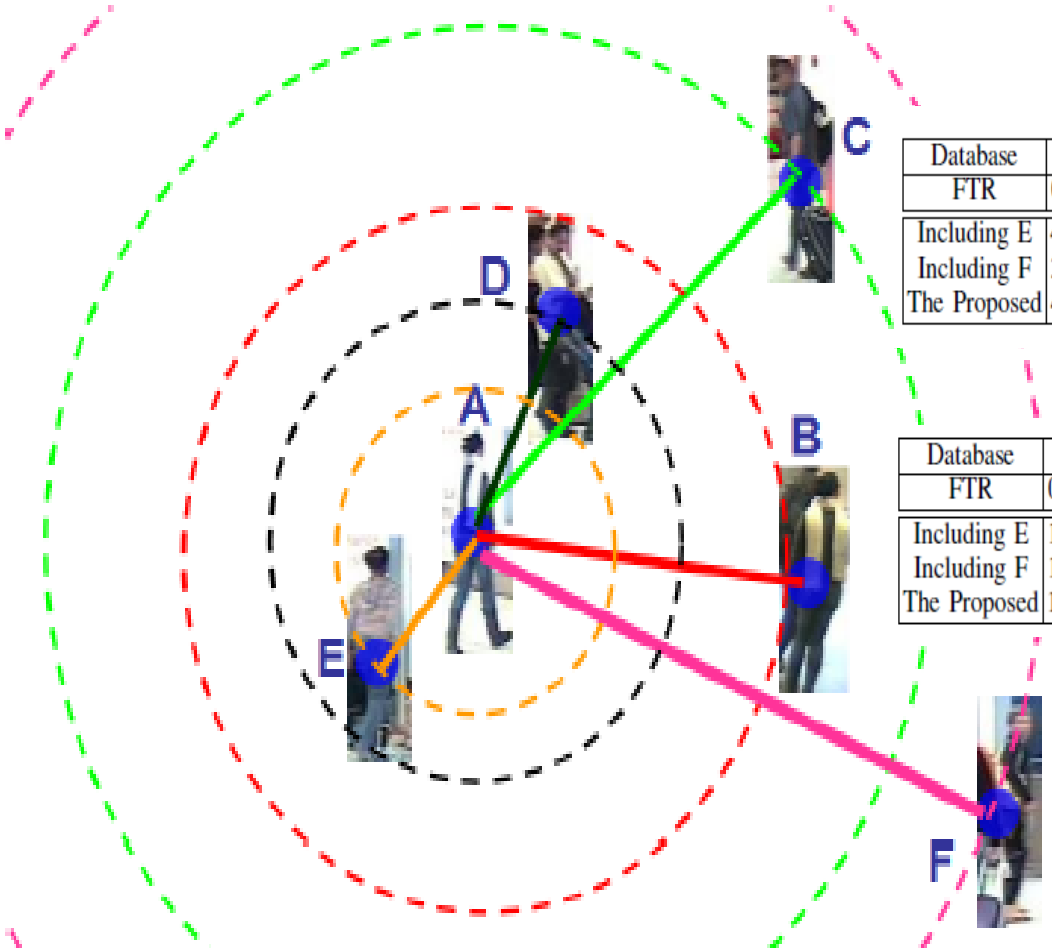


# One-Shot Open-World Group-based Re-id

- **Local Modelling: Removed Comparison**



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




ETHZ

| Database | Individual Verification | | | | | | Set Verification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FTR | 0.1% | 1% | 5% | 10% | 20% | 30% | 0.1% | 1% | 5% | 10% | 20% | 30% |
| Including E | 43.07 | 62.54 | 80.04 | 87.45 | 93.58 | 96.51 | 33.25 | 46.03 | 60.08 | 68.03 | 76.08 | 81.61 |
| Including F | 38.66 | 59.94 | 80.28 | 86.76 | 92.98 | 96.87 | 30.08 | 42.60 | 57.98 | 65.99 | 76.38 | 81.98 |
| The Proposed | **45.13** | **65.62** | **83.82** | **89.86** | **95.20** | **97.67** | **36.63** | **47.90** | **62.81** | **71.38** | **80.42** | **85.43** |

CAVIAR

| Database | Individual Verification | | | | | | Set Verification | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FTR | 0.1% | 1% | 5% | 10% | 20% | 30% | 0.1% | 1% | 5% | 10% | 20% | 30% |
| Including E | 14.38 | 25.45 | 40.85 | 49.19 | 61.67 | 70.00 | 10.14 | 16.29 | 25.43 | 34.19 | 46.01 | 56.80 |
| Including F | 14.23 | 23.73 | **41.06** | 50.19 | **63.07** | **73.48** | 9.40 | 16.35 | 23.80 | 32.13 | 46.18 | 55.91 |
| The Proposed | **15.45** | **28.13** | 40.76 | **50.78** | 60.80 | 69.99 | **11.65** | **16.74** | **29.92** | **36.72** | **47.53** | **58.04** |

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

- **A stochastic gradient algorithm**

**Algorithm 1:** Learning Procedure for t-LRDC model

Data: Dataset $\{(\mathrm{x}_i, y_i)\}_{i=1}^{N}$, $\mathrm{x}_i \in \mathbb{R}^D$, Maximum Iteration $P$, $\varepsilon > 0$
begin
  $\mathrm{M}_0 \longleftarrow D^{-1} \cdot \mathrm{I}$;
  $n \longleftarrow 0$;
  while $n \leq P$ do
    **Active Set:**
      Compute active sets $\mathbb{O}_g^{\ell}(\mathrm{x}_t, n), \mathbb{O}_a^{\ell}(\mathrm{x}_t, n), \mathbb{O}_b^{\ell}(\mathrm{x}_t, n)$;
      by Eqs. (21)-(23);
    **Gradient Descent:**
      Compute the gradient matrix $\Delta_{\mathrm{M}}$ by Eq. (27);
      $\overline{\mathrm{M}_{n+1}} \longleftarrow \mathrm{M}_n - \eta_n \cdot \Delta_{\mathrm{M}}$, $\eta_n = \frac{1}{n+1}$;
    **Projection:**
      Project $\overline{\mathrm{M}_{n+1}}$ onto $\mathbb{O}^+$ by Eq. (29)
      and obtain $\mathrm{M}_{n+1}$;
    if $||\mathrm{M}_n - \mathrm{M}_{n+1}||_F^2 < \varepsilon$ then
      break;
    end
    $n \longleftarrow n + 1$;
  end
end
Output: $\mathrm{M} = \mathrm{M}_{n+1}$

$$\min_{\mathrm{M} \succeq 0} f(\mathrm{M}) + \frac{\lambda}{2(1-\lambda)}||\mathrm{M}||_F^2$$

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

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

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

**Compute Active Set**

**The local neighbourhood sets are updated at each step**

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

## Individual Verification

| 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 |

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

## Individual Verification

| 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 |

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

## Individual Verification

| 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** |

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

## Individual Verification

| 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 |

# 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 |

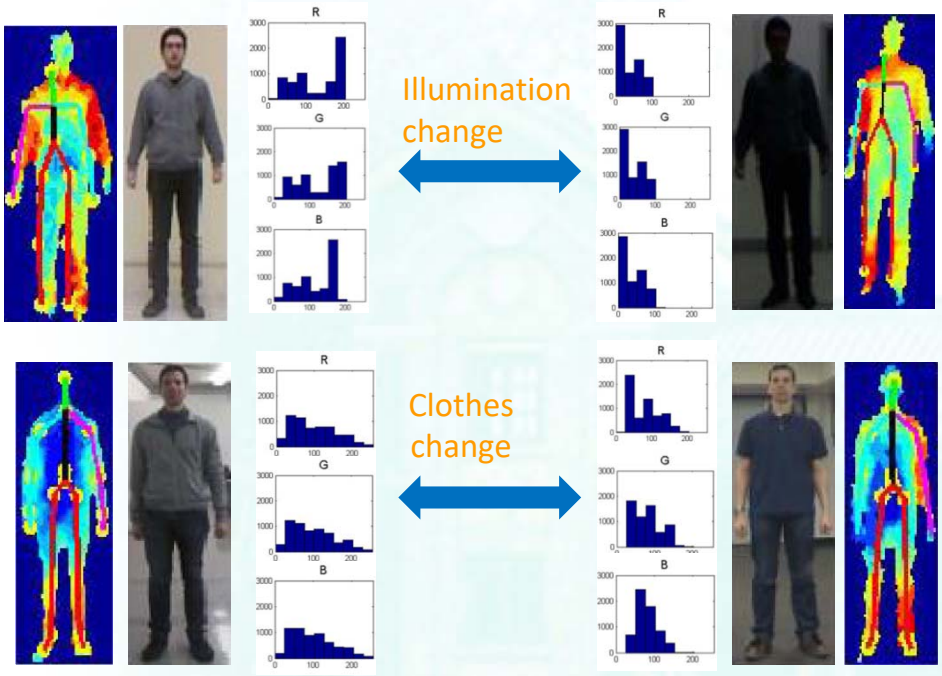
# When Clothing Change?
# Bad Lighting?

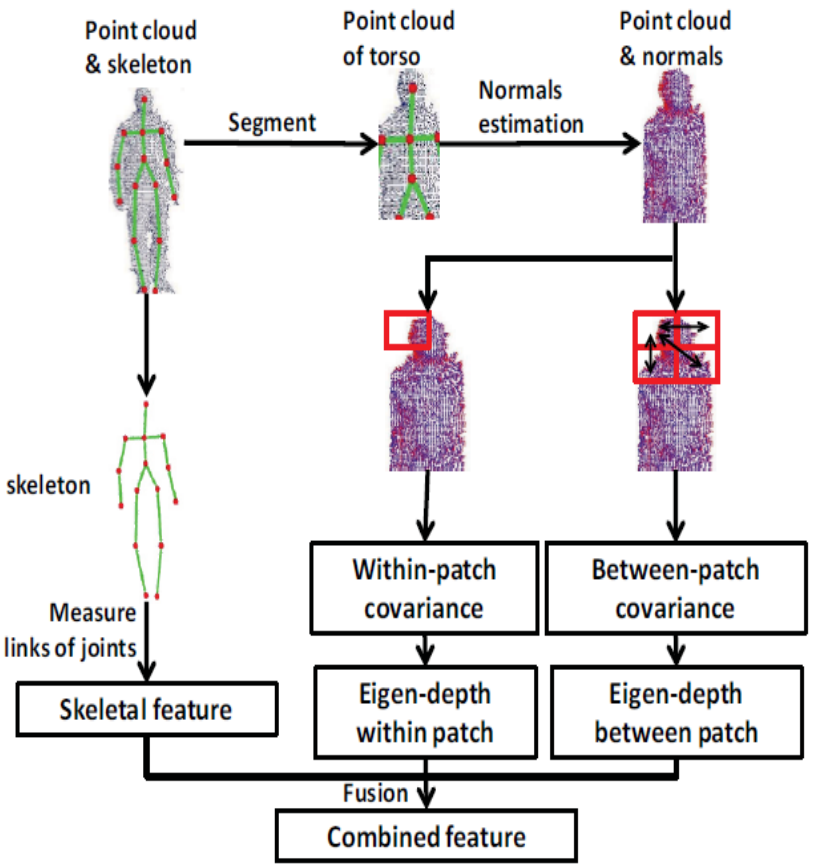


中山大學
SUN YAT-SEN UNIVERSITY

# Depth RE-ID

- **Depth-based Re-identification**




**In these cases, appearance cues are not reliable.**

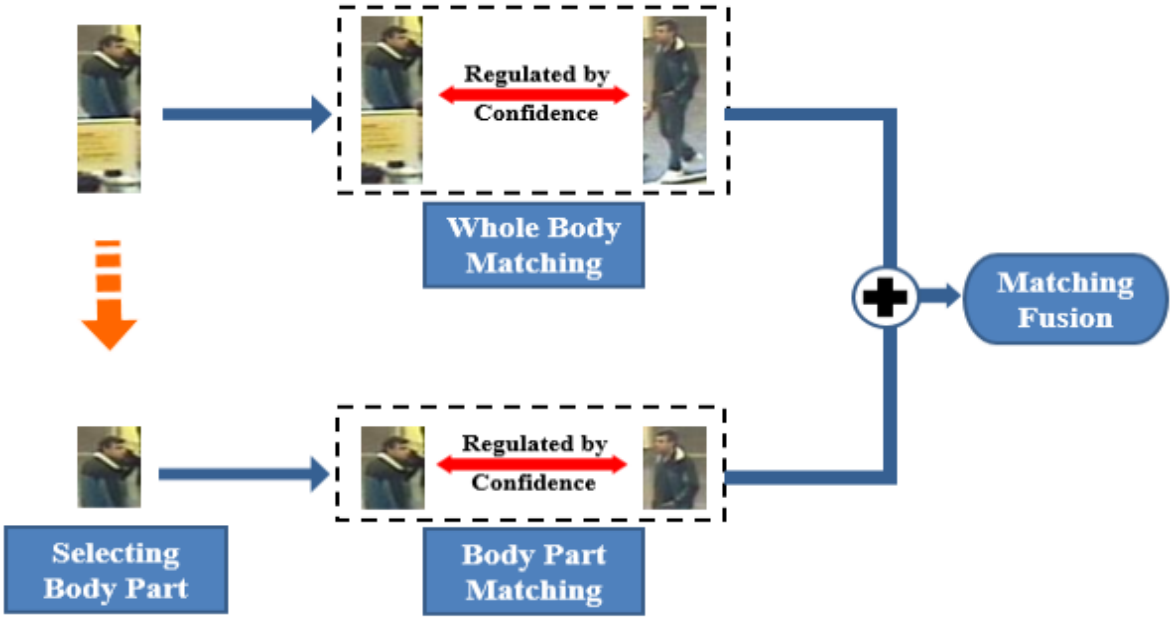
Ancong Wu, Wei-Shi Zheng*, and Jian-Huang Lai. Depth-based Person Re-identification. Asian Conference on Pattern Recognition, 2015, oral.

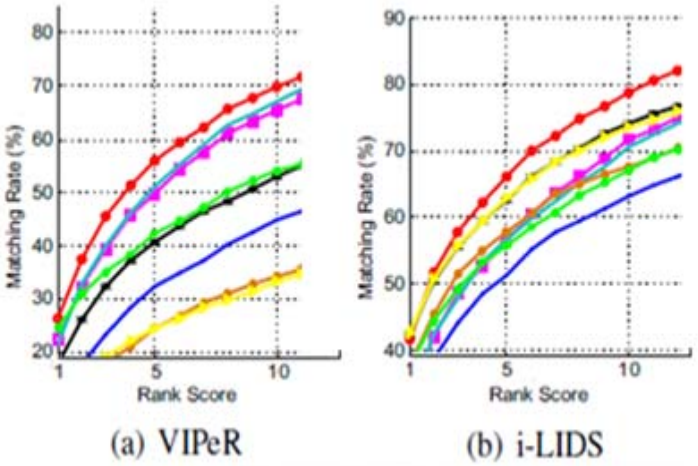
Ancong Wu, Wei-Shi Zheng*, and Jian-Huang Lai. Robust Depth-based Person Re-identification. Submitted to IEEE Transactions on Image Processing, 2016. (Minor Revision)

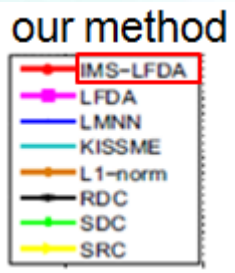
# Which Body Part is Important

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

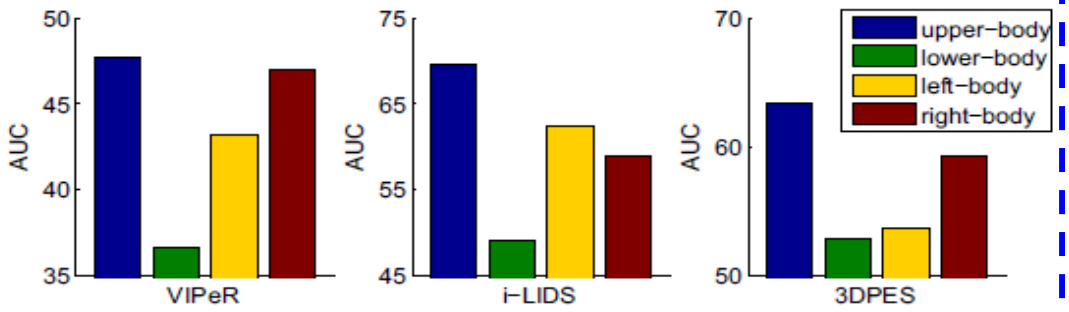
**"Towards More Reliable Matching for Person Re-identification"**




**The upper-body is superior to other body parts**



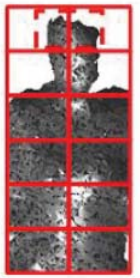
# Depth RE-ID

- **Depth-based Re-identification**

- **Within-patch Covariance**

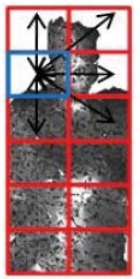
$$C_1 = \frac{1}{m-1}\sum_{i=1}^{m}(f_{1i} - \mu_1)(f_{1i} - \mu_1)^T,$$

where $\mu_1$ is the mean of the feature vectors of $R_1$.



- **Between-patch Covariance**

$$C_{12} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(f_{1i} - f_{2j})(f_{1i} - f_{2j})^T.$$



- **Eigen-depth feature**

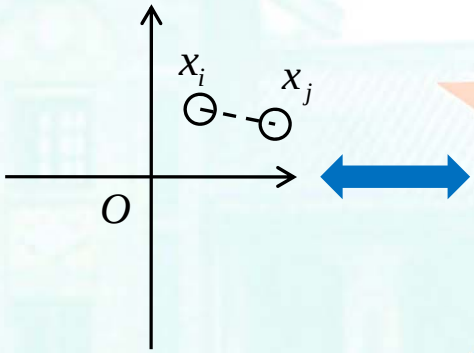
$$\mathbf{x}_p = [\ln\lambda_{p,1}\ \ln\lambda_{p,2}\ \ldots\ \ln\lambda_{p,6}]^T$$

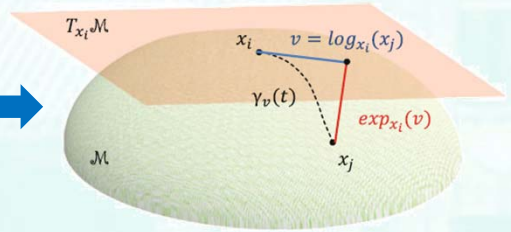
**Eigen-depth feature is rotation invariant.**

# Depth RE-ID

- **Depth-based Re-identification**

**Theorem** $\|X_2 - X_1\|_2 = dist(C_1, C_2^N)$




$$dist(C_1, C_2^N) = \sqrt{\sum_{k=1}^{6} ln^2 \lambda_k(C_1, C_2^N)}$$

**Extracting Eigen-depth feature converts covariance matrices on Riemannian manifold to feature vectors in Euclidean space.**

# Depth RE-ID

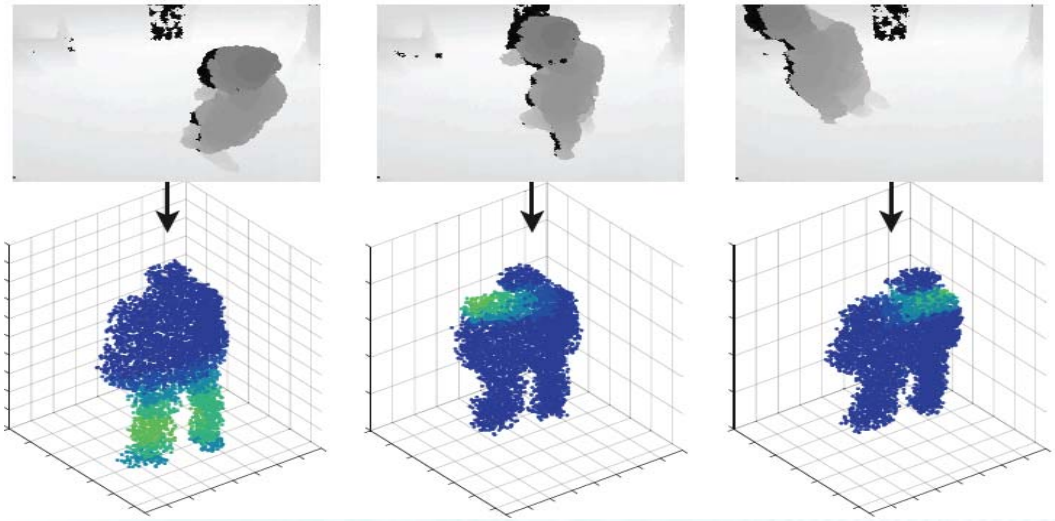

Fig. 10. Examples captured from multipl changed clothes and environment.

3DPES AND CAVIAR4REID: EVALUATION OF THE TRANSFERRED DEPTH FEATURE WHEN COMBINED WITH RGB-BASED APPEARANCE FEATURES USING DIFFERENT METRICS (%).

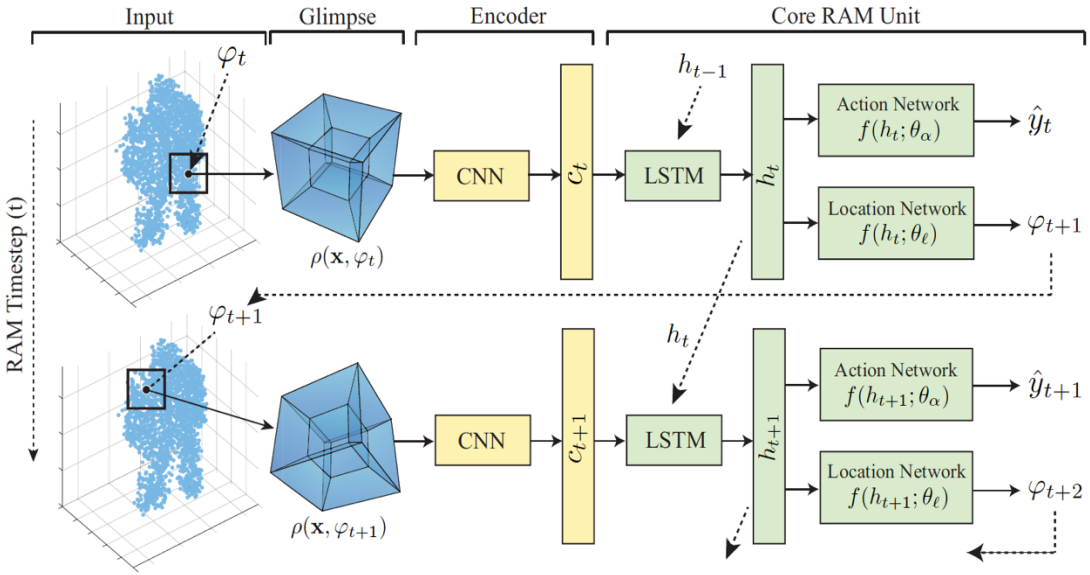
| Dataset | 3DPeS | | | | | CAVIAR4REID | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| TED | **16.0** | **21.7** | **26.4** | **29.0** | **32.1** | **27.8** | **35.2** | **39.7** | **43.6** | **46.8** |
| D-CCA | 6.5 | 11.0 | 15.0 | 17.8 | 20.5 | 24.2 | 31.9 | 36.6 | 40.1 | 42.9 |
| D-SPA | 2.7 | 4.5 | 6.4 | 7.5 | 9.0 | 5.7 | 9.2 | 11.8 | 14.7 | 17.9 |
| LFDA metric | | | | | | | | | | |
| ELF18 | 30.3 | 40.5 | 46.4 | 51.5 | 55.3 | 32.6 | 42.9 | 49.6 | 55.2 | 59.0 |
| ELF18+TED | **34.7** | **45.2** | **51.3** | **56.5** | **60.0** | **37.0** | **45.8** | **52.0** | **56.9** | **60.8** |
| ELF18+D-CCA | 30.0 | 40.5 | 47.2 | 52.7 | 56.7 | 35.7 | 44.0 | 48.9 | 53.3 | 56.6 |
| ELF18+D-SPA | 30.3 | 40.5 | 47.1 | 51.5 | 55.6 | 32.2 | 41.9 | 47.6 | 53.0 | 57.5 |
| LOMO | 41.4 | 53.4 | 60.4 | 64.3 | 68.0 | 40.2 | 50.1 | 56.7 | 61.8 | 65.6 |
| LOMO+TED | **43.8** | **54.9** | **61.2** | **65.7** | **68.8** | **42.2** | **51.4** | **56.9** | **62.1** | **65.7** |
| LOMO+D-CCA | 41.2 | 52.8 | 59.6 | 64.0 | 67.2 | 40.9 | 49.3 | 54.9 | 59.5 | 63.3 |
| LOMO+D-SPA | 40.2 | 51.8 | 59.2 | 63.8 | 66.9 | 38.8 | 47.9 | 53.8 | 59.0 | 62.7 |
| MLAPG metric | | | | | | | | | | |
| ELF18 | 35.5 | 47.1 | 54.2 | 59.1 | 62.8 | 34.5 | 46.4 | 54.0 | 60.0 | 65.1 |
| ELF18+TED | **38.6** | **49.7** | **56.6** | **61.8** | **65.1** | **38.5** | **49.3** | **55.7** | **60.8** | **65.7** |
| ELF18+D-CCA | 33.9 | 45.9 | 52.3 | 57.5 | 61.6 | 36.9 | 46.1 | 52.1 | 56.8 | 60.3 |
| ELF18+D-SPA | 34.5 | 46.7 | 53.3 | 59.1 | 62.9 | 34.2 | 45.0 | 52.2 | 58.4 | 62.5 |
| LOMO | 47.1 | 58.5 | 64.5 | 68.5 | 71.7 | 40.6 | 51.8 | 59.4 | 65.2 | 69.4 |
| LOMO+TED | **48.4** | **58.7** | **64.6** | **68.8** | **72.0** | **42.8** | **52.9** | **59.8** | **65.3** | **69.6** |
| LOMO+D-CCA | 43.7 | 55.2 | 62.3 | 66.5 | 69.7 | 41.6 | 50.0 | 56.3 | 60.9 | 64.8 |
| LOMO+D-SPA | 44.3 | 55.8 | 62.3 | 66.5 | 69.3 | 39.0 | 48.8 | 55.3 | 60.9 | 65.2 |
| KISSME metric | | | | | | | | | | |
| ELF18 | 32.4 | 42.8 | 48.9 | 53.5 | 57.0 | 33.3 | 42.6 | 48.7 | 53.5 | 57.7 |
| ELF18+TED | **35.3** | **45.4** | **52.1** | **56.7** | **59.8** | **36.3** | **45.6** | **50.9** | **55.3** | **59.5** |
| ELF18+D-CCA | 32.6 | 42.6 | 49.7 | 54.4 | 57.8 | 35.9 | 43.7 | 48.4 | 52.8 | 56.0 |
| ELF18+D-SPA | 32.4 | 42.6 | 48.5 | 53.5 | 57.0 | 33.1 | 42.1 | 47.8 | 52.4 | 56.7 |
| LOMO | 44.3 | 54.6 | 61.3 | 65.2 | 68.7 | 42.7 | 52.7 | 59.4 | 64.3 | 69.0 |
| LOMO+TED | **45.2** | **55.3** | **61.6** | **65.8** | **69.5** | **43.3** | **53.2** | **59.9** | **64.5** | **69.0** |
| LOMO+D-CCA | 43.3 | 54.3 | 61.2 | 65.3 | 68.9 | 42.4 | 51.8 | 58.1 | 62.9 | 66.9 |
| LOMO+D-SPA | 44.1 | 54.5 | 60.7 | 65.2 | 68.9 | 41.9 | 51.3 | 57.8 | 62.9 | 67.2 |

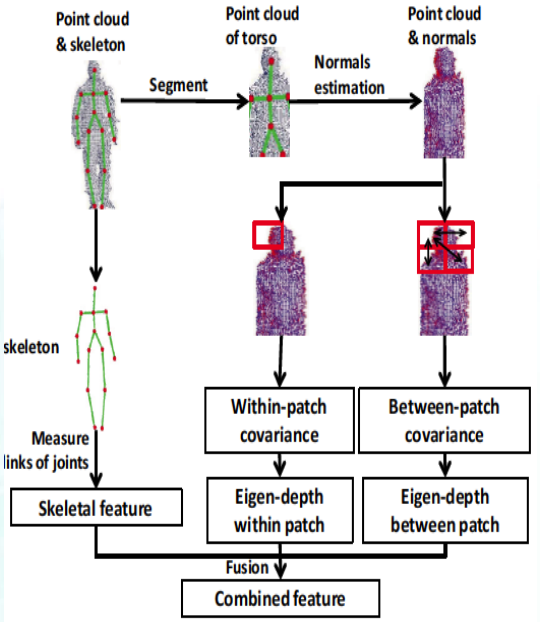
lking2" in PAVIS. Most n "Walking1".

# Depth (Identification/ Re-identification)



Attention in the Dark: A Recurrent Attention Model for Person Identification, by Albert Haque Alexandre Alahi Li Fei-Fei @ CVPR 2016




Depth-based Person Re-identification, by Ancong Wu , Wei-Shi Zheng* , and Jian-Huang Lai, ACPR 2015. Journal Version Submitted

# Outline

- **Brief Introduction of ML for Biometrics**
- **ML for Person Re-identification**
  - Distance Metric Learning
  - View Change Invariant Features
  - Partial Re-id
  - Low Resolution
  - Video-based Re-id
  - Cross Scenario Transfer
  - Open-world Modelling
  - Depth Re-identification
- **Summary**

# Summary

**Cross-scenario RE-ID**
(TCSVT 2016)

**Relative Distance Comparison**
(TPAMI 2013, CVPR2011)

**Depth RE-ID**
(ACPR2015, TIP Minor)

**Multi-scale Method**
(ICCV2015)

**RE-ID Specific Distance Metric Learning**

**OPEN-WOLRD RE-ID**
(TPAMI 2016/CVPR 2012)

**Partial RE-ID**
(ICCV2015)

**DEEP RE-ID**
(WACV2016
TPAMI Minor)

**Video-based RREID**
(CVPR2016/PR2011)

# Summary

- **Some thoughts of RE-ID**
  - **Not just a topic about image-image recognition**
  - **Not just about a conventional classification problem**
  - **Not just about a conventional retrieval problem**
  - **Not just a machine learning task**
  - **Interaction with operators: Human in the loop**
  - **Long-term ……**
  - **Unsupervised Learning**
  - **How to select the person you want to track?**
    - **Activity/action of our works: ICCV 2013, CVPR 2015, ECCV 2016, TIP 2015 (small group, early prediction, RGB-D)**

# MORE INFO.

VISITING MY HOME PAGE

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

http://isee.sysu.edu.cn

EMAIL ME: wszheng@ieee.org