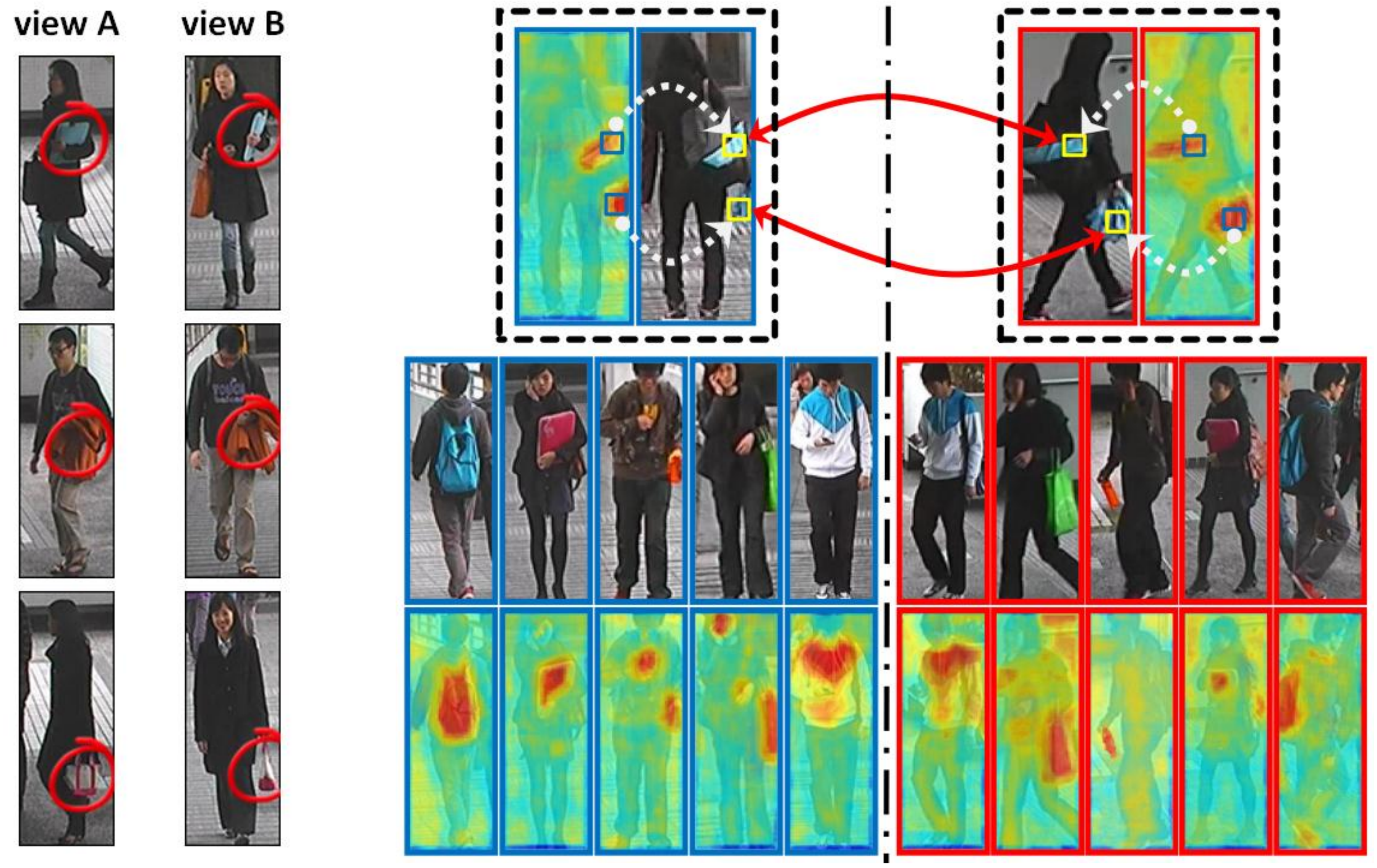
Unsupervised Salience Learning for Person Re-identification

Rui Zhao Wanli Ouyang Xiaogang Wang
Department of Electronic Engineering, The Chinese University of Hong Kong

香港中文大學
The Chinese University of Hong Kong



Motivation:

- We can recognize persons across camera views from their local distinctive regions
- Human salience
  - can identify important local features
  - is robust to the change of view points
  - itself is a useful descriptor for pedestrian matching
- Distinct patches are considered as salient only when they are matched and distinct in both camera views
- **These regions are discarded as outliers by existing methods or have little effect on person matching because of small sizes**

Contribution:

- An unsupervised framework to extract distinctive features for person re-identification.
- Patch matching is utilized with adjacency constraint for handling the misalignment problem caused by viewpoint change and pose variation.
- Human salience is learned in an unsupervised way.

Code is available at
http://mmlab.ie.cuhk.edu.hk/projects/project_salience_reid/index.html

Dense Correspondence:

- **Features**: dense color histogram + dense SIFT
- **Adjacency constrained search**: simple patch matching

Unsupervised Salience Learning:

- **Definition**: Salient regions are ***discriminative*** in making a person standing out from their companions, and ***reliable*** in finding the same person across camera views.
- **Assumption**: fewer than half of the persons in a reference set share similar appearance if a region is salient. Hence, we set k = Nr/2. Nr is the number of images in reference set.

• K-Nearest Neighbor Salience:

$$\mathbf{X}_{nn}(x_{m,n}^{A,p}) = \{x \mid \operatorname*{argmax}_{\hat{x}\in\hat{\mathcal{S}}_{p,q}} s(x_{m,n}^{A,p}, \hat{x}), q = 1, 2, ..., N_r\}$$

$$\mathbf{score}_{knn}(x_{m,n}^{A,p}) = D_k(\mathbf{X}_{nn}(x_{m,n}^{A,p}))$$

• One-Class SVM Salience:

$$\min_{R\in\mathbb{R},\xi\in\mathbb{R}^l,c\in F} R^2 + \frac{1}{vl}\sum_i \xi_i$$

$$s.t. \|\Phi(X_i) - c\|^2 \leq R^2 + \xi_i, \quad \forall i \in \{1, ...l\} : \xi_i \geq 0$$

$$f(X) = R^2 - \|\Phi(X) - c\|^2$$

$$\mathbf{score}_{ocsvm}(x_{m,n}^{A,p}) = d(x_{m,n}^{A,p}, x^*), \quad x^* = \operatorname*{argmax}_{x\in\mathbf{X}_{nn}(x_{m,n}^{A,p})} f(x)$$

Matching for Re-identification

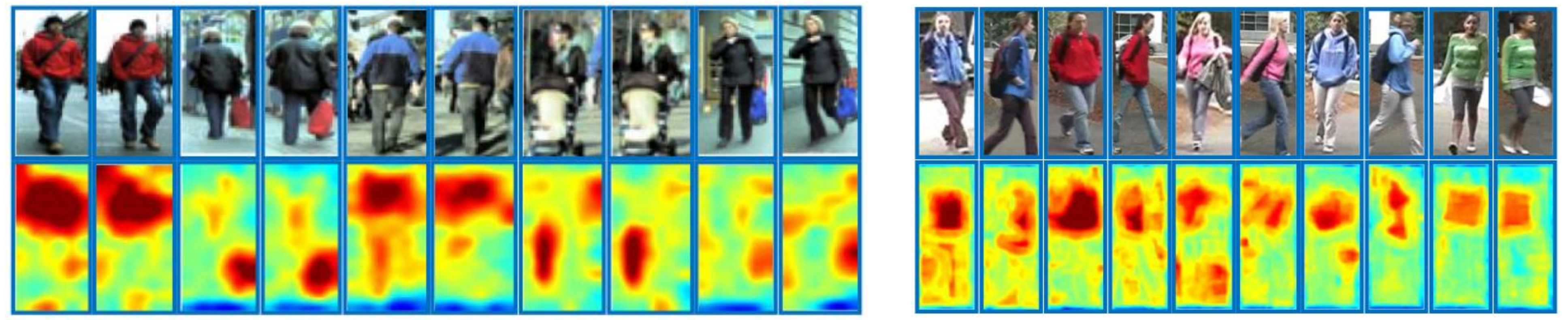
- **Bi-directional Weighted Matching**

$$\mathbf{Sim}(\mathbf{x}^{A,p}, \mathbf{x}^{B,q}) = \sum_{m,n} \frac{\mathbf{score}_{knn}(x_{m,n}^{A,p}) \cdot s(x_{m,n}^{A,p}, x_{i,j}^{B,q}) \cdot \mathbf{score}_{knn}(x_{i,j}^{B,q})}{\alpha + |\mathbf{score}_{knn}(x_{m,n}^{A,p}) - \mathbf{score}_{knn}(x_{i,j}^{B,q})|}$$

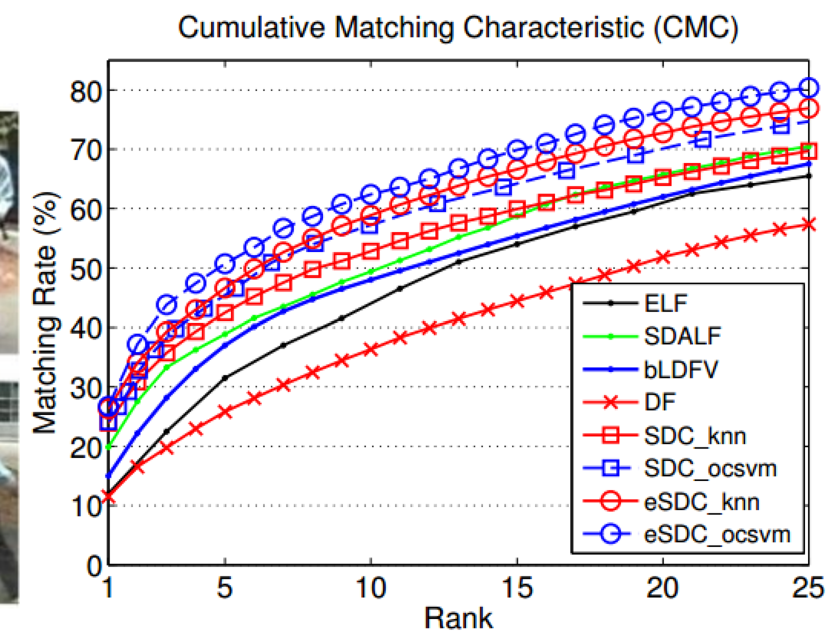
- **Complementary Combination**

$$d_{eSDC}(I_p^A, I_q^B) = \sum_i \beta_i \cdot \mathbf{s}_i(F_i(I_p^A), F_i(I_q^B)) - \beta_{SDC} \cdot \mathbf{Sim}(\mathbf{x}^{A,p}, \mathbf{x}^{B,q})$$

Experimental Results:

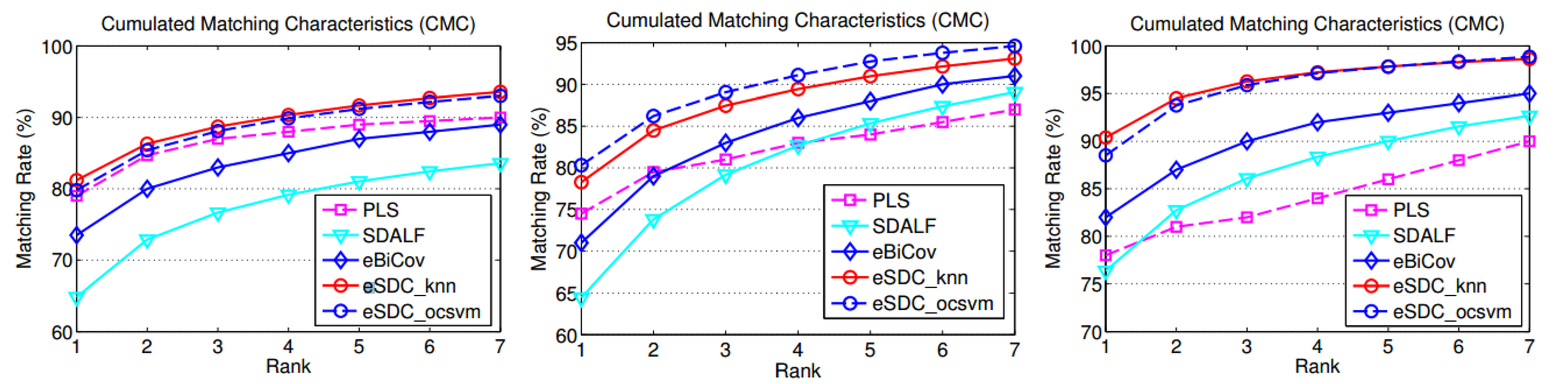


• VIPeR Dataset



Method	r=1	r=5	r=10	r=20
LMNN[29]	6.23	19.65	32.63	52.25
ITML [29]	11.61	31.39	45.76	63.86
PRDC[29]	15.66	38.42	53.86	70.09
aPRDC[19]	16.14	37.72	50.98	65.95
PCCA [24]	19.27	48.89	64.91	80.28
ELF[13]	12.00	31.00	41.00	58.00
SDALF [10]	19.87	38.89	49.37	65.73
CPS [7]	21.84	44.00	57.21	71.00
eBiCov [21]	20.66	42.00	56.18	68.00
eLDFV [22]	22.34	47.00	60.04	71.00
eSDC_knn	**26.31**	**46.61**	**58.86**	**72.77**
eSDC_ocsvm	**26.74**	**50.70**	**62.37**	**76.36**

• ETHZ Dataset



SEQ.#1 SEQ.#2 SEQ.#3