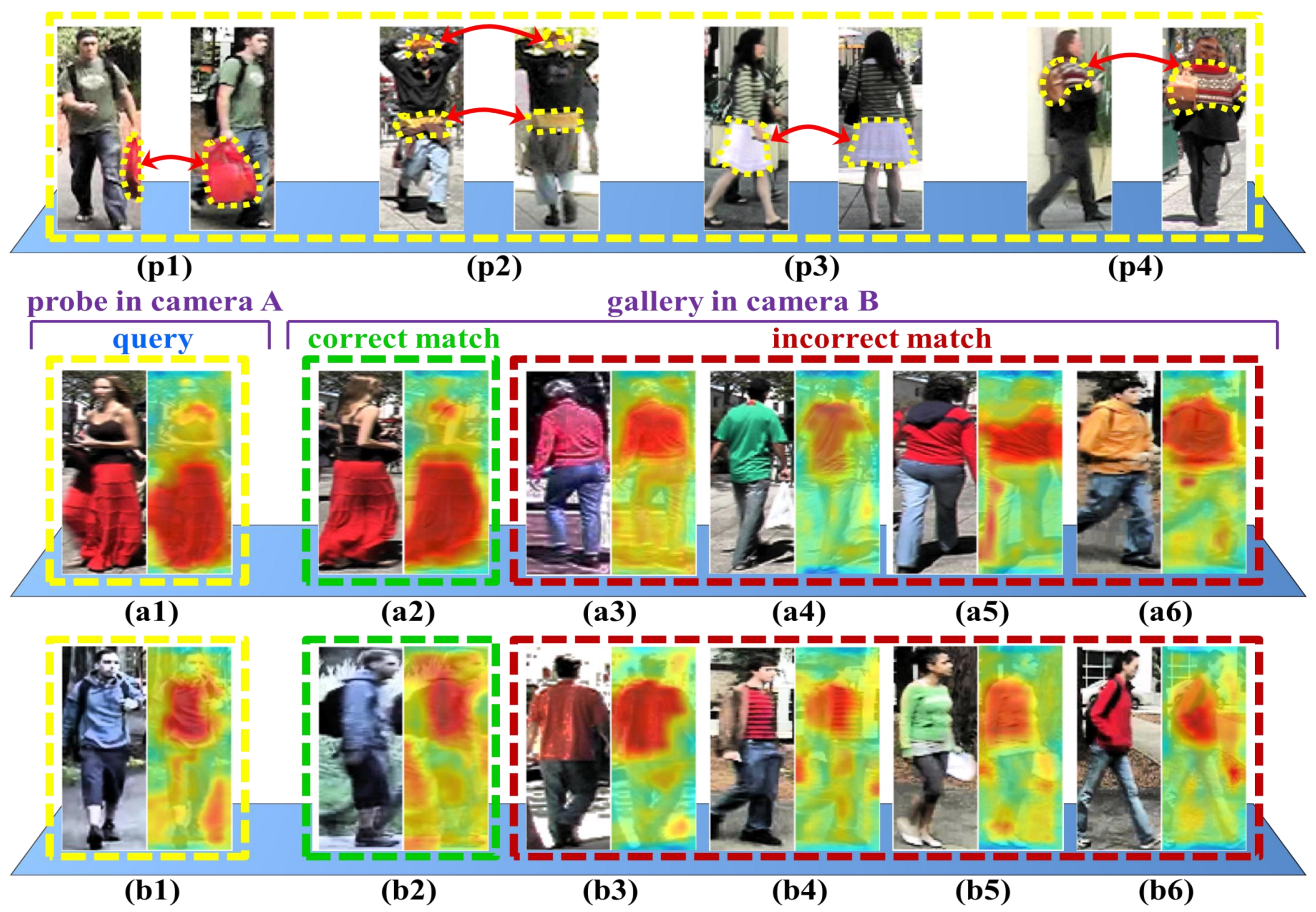
Person Re-identification by Salience Matching

Rui Zhao Wanli Ouyang Xiaogang Wang

The Chinese University of Hong Kong

香港中文大學 The Chinese University of Hong Kong

International Conference on Computer Vision Sydney 2013



Motivation

- Misalignment is caused by variations of viewpoints and poses, which commonly exist in person re-identification.
- Some local patches are more distinctive and reliable when matching two persons.
- Images of the same person captured from different camera views have some invariance property on their spatial distribution on salience.

Contribution

- A probabilistic distribution of salience is reliably estimated with our approach.
- We formulate person re-identification as a salience matching problem.
- Salience matching and patch matching are tightly integrated into a unified structural RankSVM learning framework.

Unsupervised Human Salience Learning

Algorithm 1 Compute human salience.

Input: image $\mathbf{x}^{A,u}$ and a reference image set $\mathcal{R} = \{\mathbf{x}^{B,v}, v = 1, \dots, N_r\}$

Output: salience probability map $P(l^{A,u}_{m,n} = 1 \mid x^{A,u}_{m,n})$

1: **for** each patch $x^{A,u}_{m,n} \in X$ **do**

2: compute $X_{NN}(x^{A,u}_{m,n})$ with Eq. (1)

3: compute $score(x^{A,u}_{m,n})$ with Eq. (2)

4: compute $P(l^{A,u}_{m,n} = 1 \mid x^{A,u}_{m,n})$ with Eq. (3)

5: **end for**

Step 1: construct nearest-neighbor patch set

$$X_{NN}(x^{A,u}_{m,n}) = \{x \mid \underset{x^{B,v}_{i,j}}{\operatorname{argmin}}\, d(x^{A,u}_{m,n}, x^{B,v}_{i,j}), \quad x^{B,v}_{i,j} \in \mathfrak{S}(x^{A,u}_{m,n}, \mathbf{x}^{B,v}), v = 1, \dots, N_r\}, \tag{1}$$

Step 2: compute k-NN distances as salience score

$$score(x^{A,u}_{m,n}) = d_k(X_{NN}(x^{A,u}_{m,n})), \tag{2}$$

Step 3: convert to salience probability

$$P(l^{A,u}_{m,n} = 1 \mid x^{A,u}_{m,n}) = 1 - exp(-score(x^{A,u}_{m,n})^2/\sigma_0^2), \tag{3}$$

Supervised Salience Matching Framework

Matching based on Salience:

$$f_z(\mathbf{x}^A, \mathbf{x}^B, \mathbf{l}^A, \mathbf{l}^B; \mathbf{p}, \mathbf{z}) = \sum_{p_i} \Big\{ z_{p_i,1} l^A_{p_i} l^B_{p'_i} + z_{p_i,2} l^A_{p_i}(1 - l^B_{p'_i}) + z_{p_i,3}(1 - l^A_{p_i}) l^B_{p'_i} + z_{p_i,4}(1 - l^A_{p_i})(1 - l^B_{p'_i}) \Big\},$$

- Hidden Salience label: $\mathbf{l}^A = \{l^A_{p_i} \mid l^A_{p_i} \in \{0,1\}\} \quad \mathbf{l}^B = \{l^B_{p'_i} \mid l^B_{p'_i} \in \{0,1\}\}$
- Matching score: $\mathbf{z} = \{z_{p_i,k}\}_{k=1,2,3,4}$

$$z_{p_i,k} = \alpha_{p_i,k} \cdot s(x^A_{p_i}, x^B_{p'_i}) + \beta_{p_i,k}, \quad s(x^A_{p_i}, x^B_{p'_i}) = \exp\Big(-\frac{d(x^A_{p_i}, x^B_{p'_i})^2}{2\sigma_0^2}\Big),$$

Marginalization:

$$f^*(\mathbf{x}^A, \mathbf{x}^B; \mathbf{p}, \mathbf{z}) = \sum_{\mathbf{l}^A, \mathbf{l}^B} f_z(\mathbf{x}^A, \mathbf{x}^B, \mathbf{l}^A, \mathbf{l}^B; \mathbf{p}, \mathbf{z}) p(\mathbf{l}^A, \mathbf{l}^B \mid \mathbf{x}^A, \mathbf{x}^B) = \sum_{p_i} \sum_{k=1}^{4} \Big[\alpha_{p_i,k} \cdot s(x^A_{p_i}, x^B_{p'_i}) + \beta_{p_i,k}\Big] c_{p_i,k}(x^A_{p_i}, x^B_{p'_i}),$$

Probabilistic salience matching

$$c_{p_i,k}(x^A_{p_i}, x^B_{p'_i}) = \begin{cases} P(l^A_{p_i} = 1 \mid x^A_{p_i}) P(l^B_{p'_i} = 1 \mid x^B_{p'_i}), & k = 1, \\ P(l^A_{p_i} = 1 \mid x^A_{p_i}) P(l^B_{p'_i} = 0 \mid x^B_{p'_i}), & k = 2, \\ P(l^A_{p_i} = 0 \mid x^A_{p_i}) P(l^B_{p'_i} = 1 \mid x^B_{p'_i}), & k = 3, \\ P(l^A_{p_i} = 0 \mid x^A_{p_i}) P(l^B_{p'_i} = 0 \mid x^B_{p'_i}), & k = 4. \end{cases}$$

Final Formulation:

$$f^*(\mathbf{x}^A, \mathbf{x}^B; \mathbf{p}, \mathbf{z}) = \mathbf{w}^{\mathrm{T}} \Phi(\mathbf{x}^A, \mathbf{x}^B; \mathbf{p}) = \sum_{p_i} w^{\mathrm{T}}_{p_i} \phi(x^A_{p_i}, x^B_{p'_i}),$$

- Weighting parameters: $w_{p_i} = [\{\alpha_{p_i,k}\}_{k=1,2,3,4}, \{\beta_{p_i,k}\}_{k=1,2,3,4}]$.
- Matching feature: $\phi(x^A_{p_i}, x^B_{p'_i})$

$$\phi(x^A_{p_i}, x^B_{p'_i}) = \begin{bmatrix} s(x^A_{p_i}, x^B_{p'_i}) P(l^A_{p_i} = 1 \mid x^A_{p_i}) P(l^B_{p'_i} = 1 \mid x^B_{p'_i}) \\ s(x^A_{p_i}, x^B_{p'_i}) P(l^A_{p_i} = 1 \mid x^A_{p_i}) P(l^B_{p'_i} = 0 \mid x^B_{p'_i}) \\ s(x^A_{p_i}, x^B_{p'_i}) P(l^A_{p_i} = 0 \mid x^A_{p_i}) P(l^B_{p'_i} = 1 \mid x^B_{p'_i}) \\ s(x^A_{p_i}, x^B_{p'_i}) P(l^A_{p_i} = 0 \mid x^A_{p_i}) P(l^B_{p'_i} = 0 \mid x^B_{p'_i}) \\ P(l^A_{p_i} = 1 \mid x^A_{p_i}) P(l^B_{p'_i} = 1 \mid x^B_{p'_i}) \\ P(l^A_{p_i} = 1 \mid x^A_{p_i}) P(l^B_{p'_i} = 0 \mid x^B_{p'_i}) \\ P(l^A_{p_i} = 0 \mid x^A_{p_i}) P(l^B_{p'_i} = 1 \mid x^B_{p'_i}) \\ P(l^A_{p_i} = 0 \mid x^A_{p_i}) P(l^B_{p'_i} = 0 \mid x^B_{p'_i}) \end{bmatrix}.$$

Ranking by Partial Order

- **Task** – finding a good ranking:

$$\mathbf{y}^{A,u}_* = \underset{\mathbf{y}^{A,u} \in \mathcal{Y}^{A,u}}{\operatorname{argmax}}\, \mathbf{w}^{\mathrm{T}} \Psi_{po}(\mathbf{x}^{A,u}, \mathbf{y}^{A,u}; \{\mathbf{x}^{B,v}\}^V_{v=1}, \{\mathbf{p}^{u,v}\}^V_{v=1}),$$

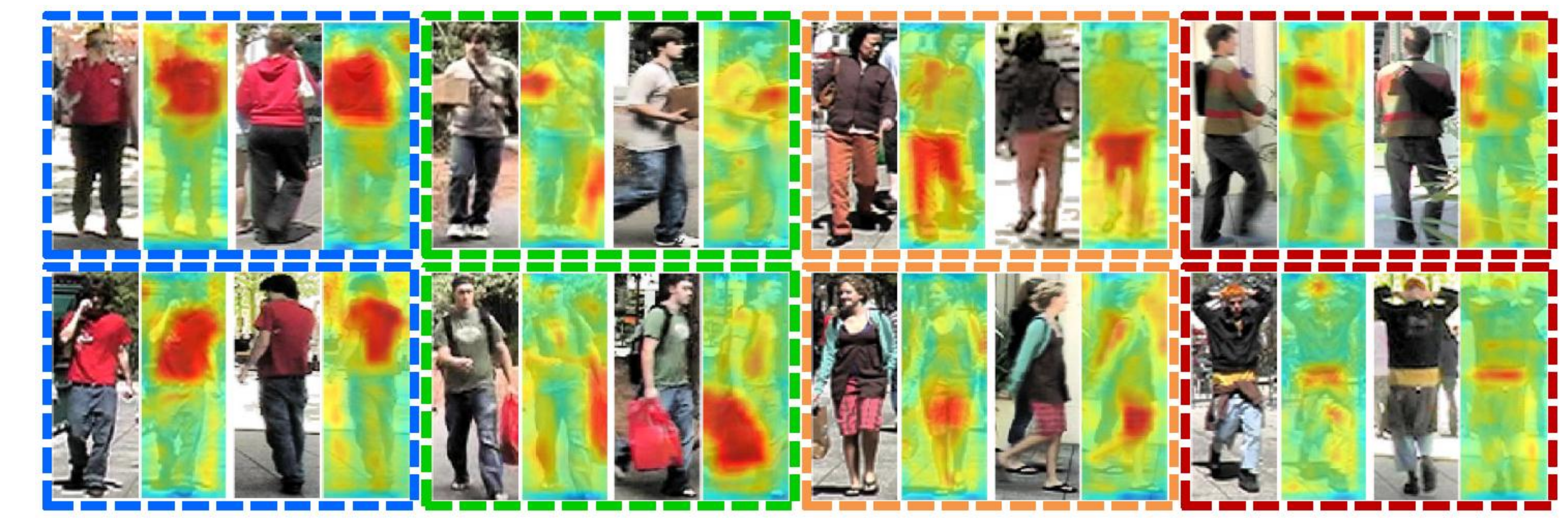
Partial order feature for structural RankSVM training:

$$\Psi_{po}(\mathbf{x}^{A,u}, \mathbf{y}^{A,u}; \{\mathbf{x}^{B,v}\}^V_{v=1}, \{\mathbf{p}^{u,v}\}^V_{v=1}) = \sum_{\mathbf{x}^{B,v} \in S^+_{\mathbf{x}^{A,u}}} \sum_{\mathbf{x}^{B,v'} \in S^-_{\mathbf{x}^{A,u}}} y^{A,u}_{v,v'} \frac{\Phi(\mathbf{x}^{A,u}, \mathbf{x}^{B,v}; \mathbf{p}^{u,v}) - \Phi(\mathbf{x}^{A,u}, \mathbf{x}^{B,v'}; \mathbf{p}^{u,v'})}{|S^+_{\mathbf{x}^{A,u}}| \cdot |S^-_{\mathbf{x}^{A,u}}|},$$

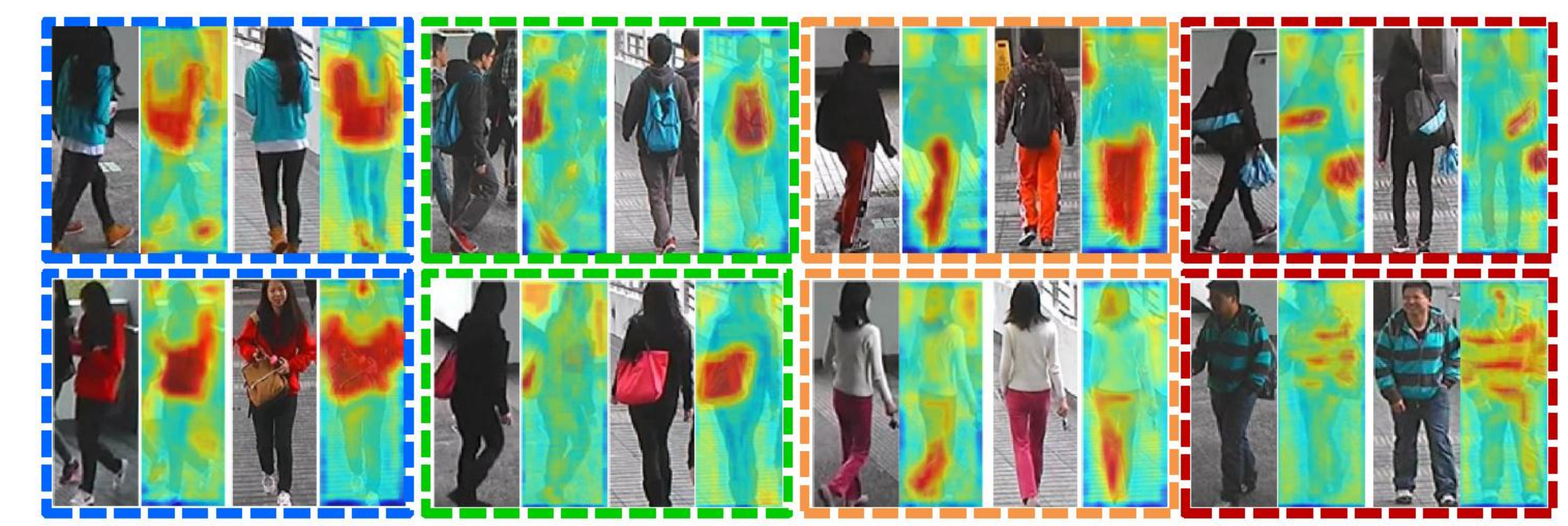
- **Solution** – sorting gallery by $\{\mathbf{w}^{\mathrm{T}}\Phi(\mathbf{x}^{A,u}, \mathbf{x}^{B,v}; \mathbf{p}^{u,v})\}_v$ in descending order.

Experimental Results

Learned Human Salience

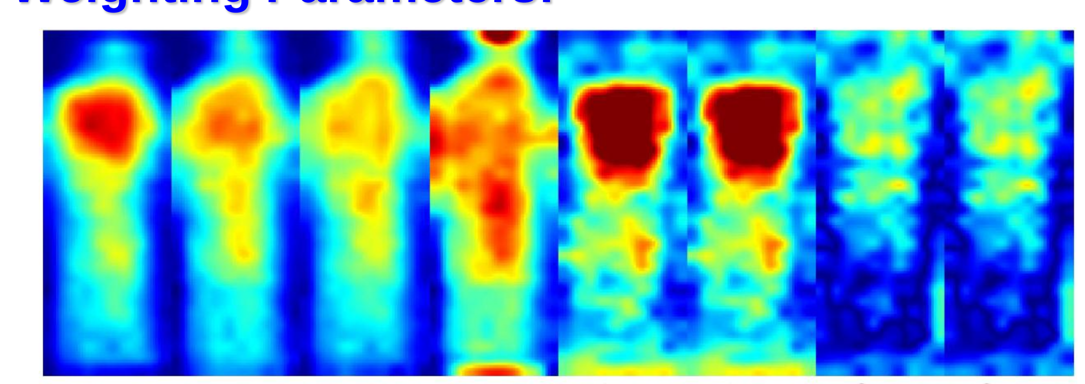


(a) VIPeR dataset



(b) CUHK Campus dataset

Learned Weighting Parameters:

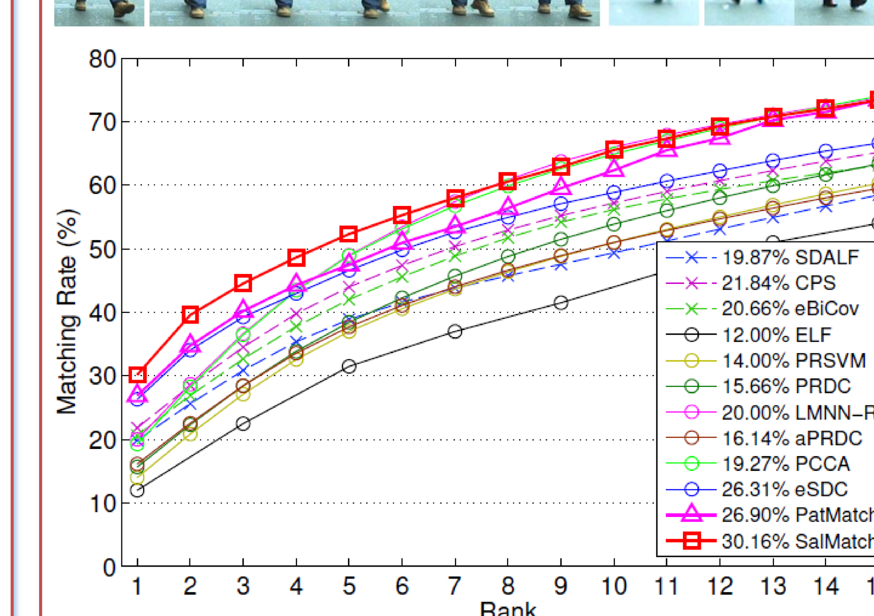


$\{\alpha_{p_i,k}\}_{k=1,2,3,4}$ correspond to the first four terms of matching features based on salience matching with visual similarity, and $\{\beta_{p_i,k}\}_{k=1,2,3,4}$ correspond to the last four terms only depending on salience matching.

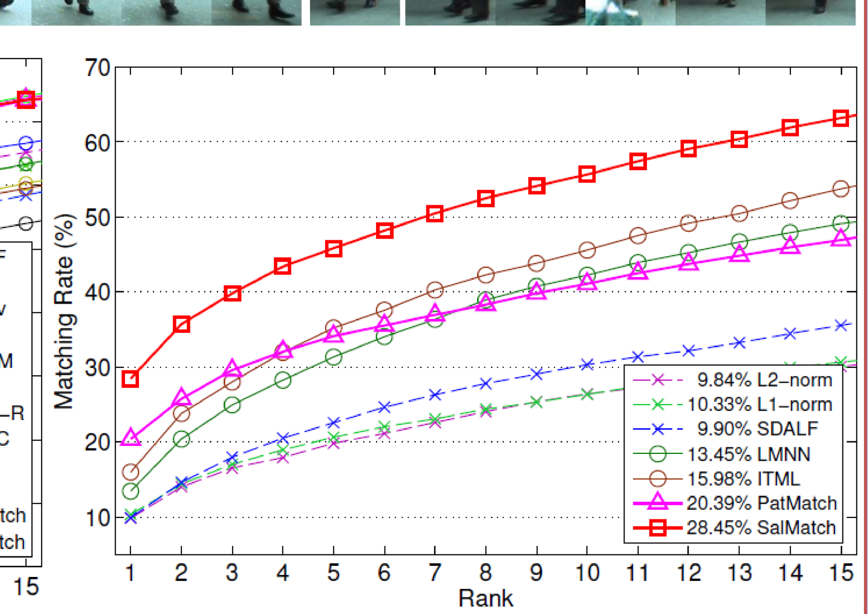
Comparison with Sate-of-the-Arts:

VIPeR dataset

CUHK Campus dataset



(a) VIPeR dataset



(b) CUHK Campus dataset