Dense Invariant Feature Based Support Vector Ranking for Person Re-identification

Shoubiao Tan
Key Laboratory of Intelligent Computing & Signal Processing,
Ministry of Education,
Anhui University
Hefei 230039, China
Email: starry226@ahu.edu.cn

Feng Zheng
Department of Electronic and Electrical Engineering,
The University of Sheffield
Mappin Street, Sheffield,
S1 3JD, UK
Email: cip12fz@sheffield.ac.uk

Ling Shao
Department of Computer Science and Digital Technologies,
Northumbria University
Newcastle upon Tyne,
NE1 8ST, UK
Email: ling.shao@ieee.org

This research was supported by the Anhui Provincial Natural Science Foundation (Grant No. 1508085MF120) and the numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Anhui University.

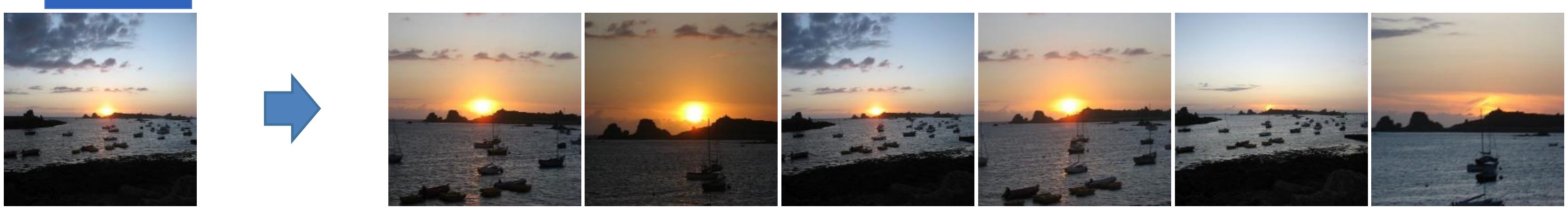
GlobalSIP 2015

Person Re-identification (ReID)

- Image Search

query

query



Person Re-identification (ReID)

- **Person Re-identification**

Difficulty

Non-overlapping Camera Views

Irrelevant negative samples, difficult to train classifiers

Difficulty

View/Pose Changes

Difficulty

Occlusions

Carried objects occlude the person appearance

Difficulty

Illumination Changes

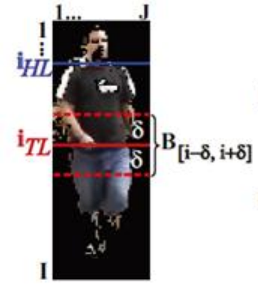
Need illumination-invariant features or light-amending process

Difficulty

Large Intra-class Variations & Limited Samples for Learning

Literature

Finding Correspondence by Segmentation

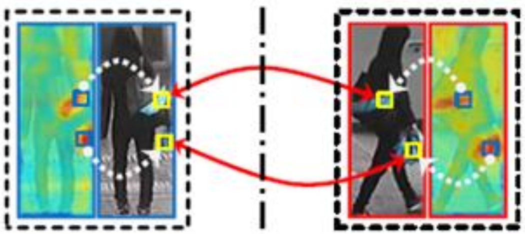


M. Farenzena et al., “Person Re-Identification by Symmetry-Driven Accumulation of Local Features”, CVPR 2010.

Finding Correspondence by Detection

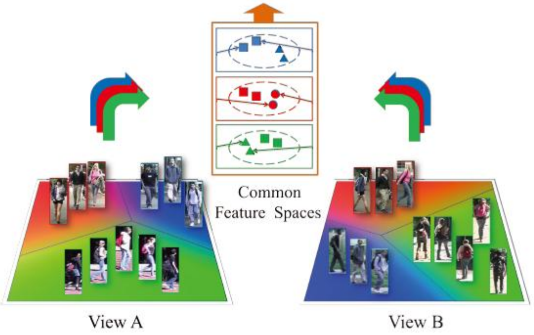
D.S. Cheng, M. Cristani, et al., “Custom pictorial structures for re-identification”, BMVC 2011.

Finding Correspondence by Salience



R. Zhao et al., "Unsupervised Salience Learning for Person Re-identification “, CVPR 2013.

Learning Transformation



W.S. Zheng et al., "Person re-identification by support vector ranking ", BMVC 2010.
W.S. Zheng et al., “Re-identification by Probabilistic Relative Distance Comparison”, TPAMI 2012.
W. Li et al., "Locally Aligned Feature Transforms across Views “, CVPR 2013.

Motivation

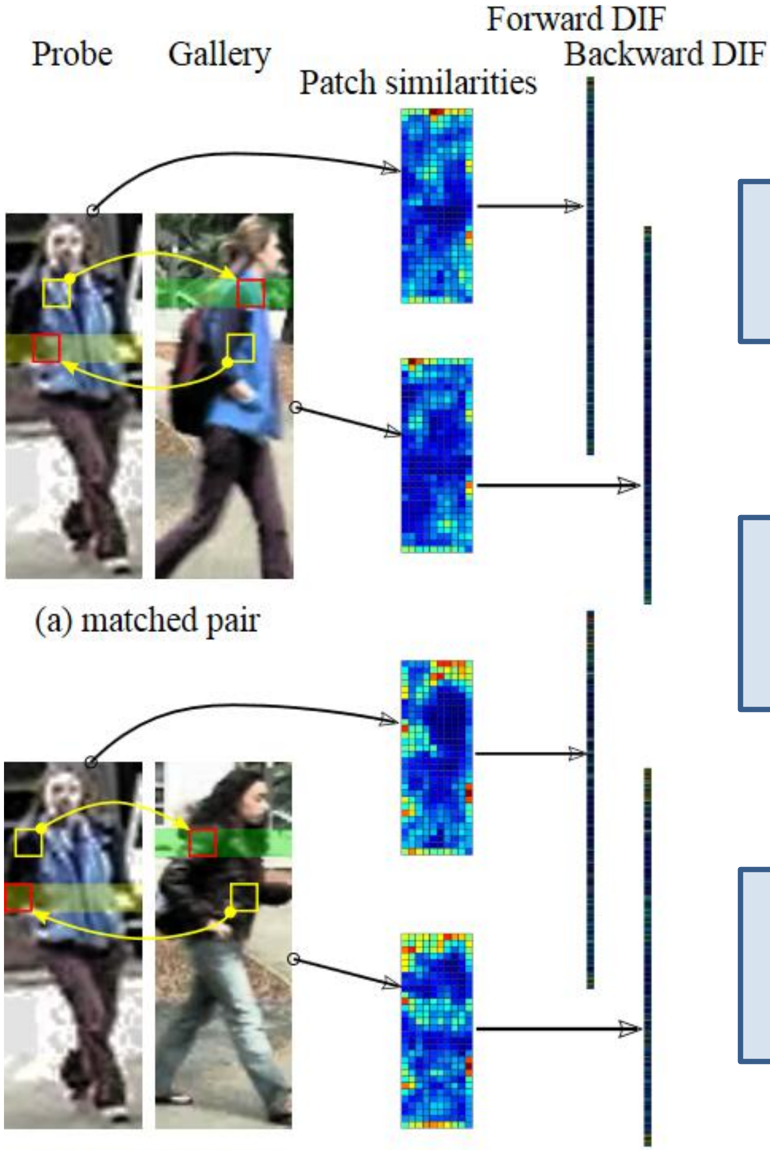
- A pose-invariant feature is important in representation of similarity of images in different views.
- Same features in different views usually located in adjacent area.
- Transformation learning is also important to person re-identification.

Flowchart of the Alg.

- Feature Extraction – dColorSIFT*
- Feature Expression – Dense Invariant Feature (DIF)
- Similarity Evaluation – Support Vector Ranking (SVR)

*R. Zhao, W. Ouyang, and X. Wang, “Unsupervised salience learning for person re-identification,” in CVPR, 2013, pp. 3586–3593

The Dense Invariant Feature



(a) matched pair

(b) unmatched pair

Densely-sample the images into patches.

⬇

Find the most similar patch of each patch of an image in the surrounding area of the other image of a pair.

⬇

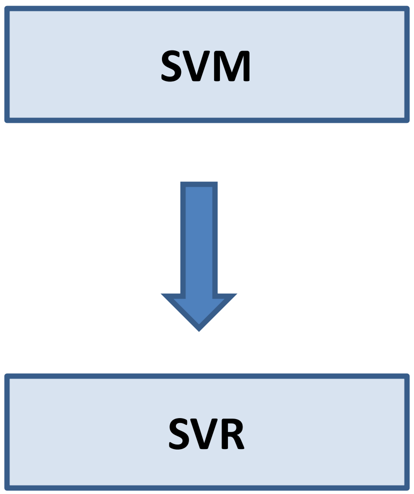
Assemble the largest similarity of each patch into a feature for a pair of images.

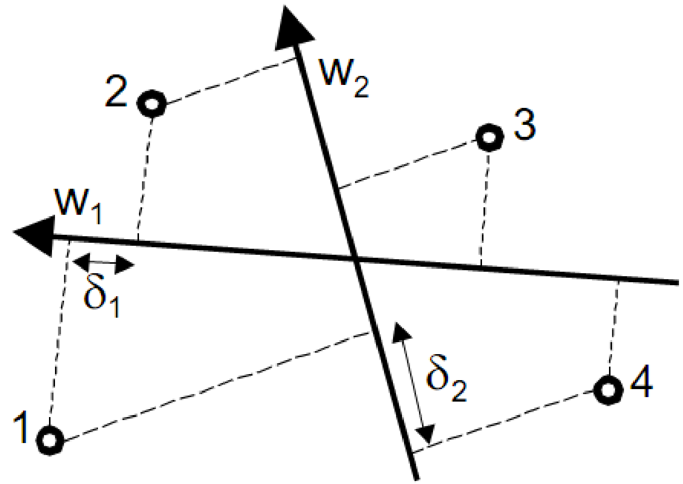
The Dense Invariant Feature

Form a feature by finding the most similar patches from an image of view B for an image of view A. → Forward DIF

Form a feature by finding the most similar patches from images of view A for images of view B. → Backward DIF

The SVR Algorithm*





$$\textit{minimize:} \quad V(\vec{w},\vec{\xi}) = \frac{1}{2}\,\vec{w}\cdot\vec{w} + C\sum \xi_{i,j,k}$$

subject to:

$$\forall (d_i, d_j) \in r_1^* : \vec{w}\Phi(q_1, d_i) \geq \vec{w}\Phi(q_1, d_j) + 1 - \xi_{i,j,1}$$

$$\ldots$$

$$\forall (d_i, d_j) \in r_n^* : \vec{w}\Phi(q_n, d_i) \geq \vec{w}\Phi(q_n, d_j) + 1 - \xi_{i,j,n}$$

$$\forall i \forall j \forall k : \xi_{i,j,k} \geq 0$$

*T. Joachims, “Optimizing search engines using clickthrough data,” in KDD, 2002, pp. 133–142.

Feature-Fusion for Ranking

Forward DIF: $\boldsymbol{F}_{ij}^{x}$ Backward DIF: $\boldsymbol{F}_{ij}^{y}$

i means image of person No. i of view A and j means image of person No. j of view B

- Feature Fusion: Project Backward DIF with a vector *P* to the space of Forward DIF and merge them into a new feature.

- The ranking objectives: Learning the projection *P* and a linear weight vector *ω* to make the best ranking of training data.

- Ranking objective function:

$$f((\boldsymbol{F}_{ii}^{x}, \boldsymbol{P}^{T}\boldsymbol{F}_{ii}^{y})) > f((\boldsymbol{F}_{ij}^{x}, \boldsymbol{P}^{T}\boldsymbol{F}_{ij}^{y})), i \neq j.$$

$$\begin{aligned} &(\boldsymbol{w}^{*}, \boldsymbol{P}^{*}) = \min_{(\boldsymbol{w}, \boldsymbol{P})} \|(\boldsymbol{w}^{x}, \boldsymbol{P}\boldsymbol{w}^{y})\|^{2} + C\sum \xi_{ij} \\ &s.t. \quad f((\boldsymbol{F}_{ii}^{x}, \boldsymbol{P}^{T}\boldsymbol{F}_{ii}^{y})) > f((\boldsymbol{F}_{ij}^{x}, \boldsymbol{P}^{T}\boldsymbol{F}_{ij}^{y})) + 1 - \xi_{ij}. \end{aligned}$$

Experiment

- Datasets

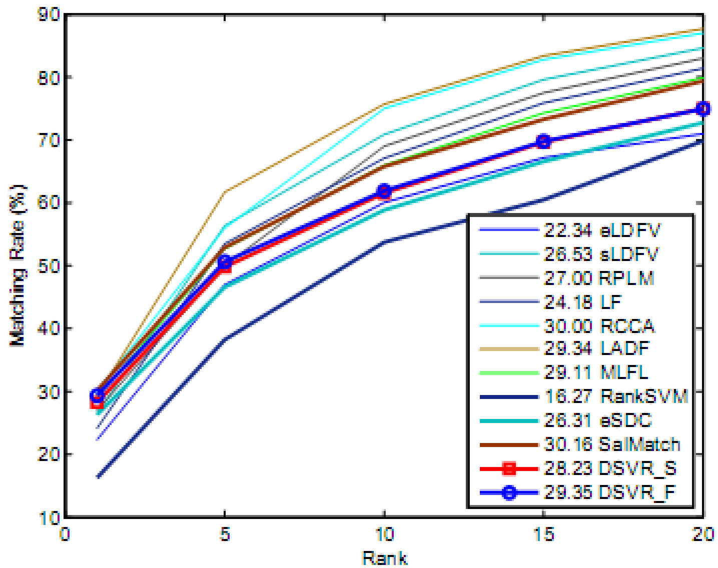
Dataset	# views	# persons	# images	Image size
VIPeR	2	632	1264	128×48
CAMPUS	2	971	3884	160×60

- Basic Feature

Densely sampled (with the size of 10×10 and an overlap of 6×6) dColorSIFT

Experiment

- Results



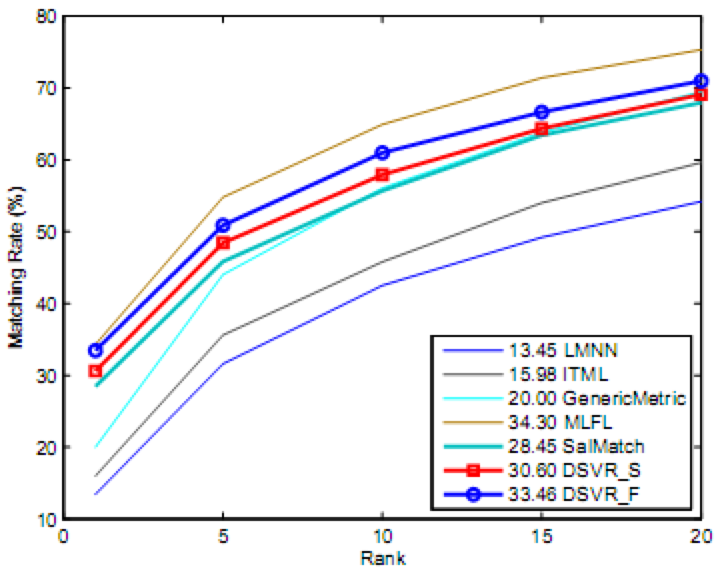
Method	r=1	r=5	r=10	r=20
sLDFV [15]	26.53	56.4	70.88	84.63
RPLM [3]	27	50	69	83
LF [8]	24.18	53.5	67.12	81.38
RCCA [7]	30	56	75	87
LADF [16]	29.34	61.7	75.7	87.7
MLFL [17]	29.11	52.7	66	79.9
RankSVM [1]	16.27	38.23	53.73	69.87
eSDC [18]	26.31	46.61	58.86	72.77
SalMatch [19]	30.16	52.9	65.8	79.4
DSVR_S[1]	27.56	50.15	61.74	75.09
DSVR_SA	28.23	49.81	61.55	75
DSVR_F[2]	28.35	50.69	61.99	74.74
DSVR_FA	29.35	50.66	61.93	74.94

TOP RANKED RATES WITH 316 PERSONS ON THE VIPER DATASET

*DSVR_SA: Our ranking method with a single forward DIF.
*DSVR_FA: Our ranking method with the fused feature.

Experiment

- Results



Method	r=1	r=5	r=10	r=20
LMNN [14]	13.45	31.6	42.5	54.2
ITML [14]	15.98	35.6	45.8	59.6
GenericMetric [14]	20	44.02	56.07	69.47
MLFL [17]	34.3	54.8	64.9	75.3
SalMatch [19]	28.45	45.85	55.67	67.95
DSVR_S[1]	30.04	48.97	58.88	69.69
DSVR_SA	30.6	48.44	57.9	69.09
DSVR_F[2]	32.82	51.5	61.31	71.33
DSVR_FA	33.46	50.88	60.97	70.97

TOP RANKED RATES WITH 486 PERSONS ON THE CUHK CAMPUS DATASET

*DSVR_SA: Our ranking method with a single forward DIF.
*DSVR_FA: Our ranking method with the fused feature.

Summary

- A novel ranking method which fuses the dense invariant features has been presented in this paper to model the relationship between an image pair across different camera views to solve the challenging REID problem effectively.
- The designed DIF is a good descriptor of an image pair with a large improvement on ranking performance.
- The fusion of bidirectional DIFs in the ranking process further improves the performance due to the reduction of the noise.

Future Work

- Test other feature fusion method with the SVR. alg.
- Try to apply our method to cross-modal person re-identification problem.

Thank-You.