Scalable Person Re-identification on Supervised Smoothed Manifold

Song Bai[1], Xiang Bai[1], Qi Tian[2]

[1]Huazhong University of Science and Technology, [2]University of Texas at San Antonio

Abstract

Existing person re-identification algorithms either extract robust visual features or learn discriminative metrics for person images. However, the underlying manifold which those images reside on is rarely investigated. That raises a problem that the learned metric is not smooth with respect to the local geometry structure of the data manifold.

In this paper, we study person re-identification with manifold-based affinity learning, which did not receive enough attention from this area. An unconventional manifold-preserving algorithm is proposed, which can 1) make the best use of supervision from training data, whose label information is given as pairwise constraints; 2) scale up to large repositories with low on-line time complexity; and 3) be plunged into most existing algorithms, serving as a generic post-processing procedure to further boost the performance. Extensive experimental results on five popular person re-identification benchmarks consistently demonstrate the effectiveness of our method. Especially, on the largest CUHK03 and Market-1501, our method outperforms the state-of-the-art alternatives by a large margin with high efficiency, which is more appropriate for practical applications.

Introduction

❑ Task

Person re-identification (ReID): an active task driven by the applications of visual surveillance, which aims to identify person images from the gallery that share the same identity as the given probe.

❑ Motivation

Current research interests can be coarsely divided into two mainstreams: those focus on designing robust visual descriptors and those seek for a discriminative metric. Unlike those methods performed in the metric space, we assume that person images reside on an underlying data manifold. The learned relationships between instances should be smooth with respect to the local geometry of the manifold.

❑ Potential Solution

	Semi-supervised learning	Unsupervised manifold learning
Representatives	Label Propagation [1], Local and Global Consistency [2], etc.	Manifold Ranking [3], Diffusion Process [4], etc.
1st Limitation	can only predict the labels of unlabeled data	ignore the beneficial influence from the labeled training data
2nd Limitation	high algorithmic complexity (graph-based)	

❑ Our Contribution

Supervised Smoothed Manifold (SSM): the similarity value between two instances is estimated in the context of other pairs of instances, thus the learned similarity well reflects the geometry structure of the underlying manifold.

SSM further has three merits specifically customized for ReID, as

1) Supervision: take advantage of the supervision in pairwise constraints, which is easily accessible in this task.
2) Efficiency: affinity learning is performed only with database instances off-line.
3) Generalization: a post-processing procedure (or a generic tool) to further boost the identification accuracies of most existing algorithms.

Methods

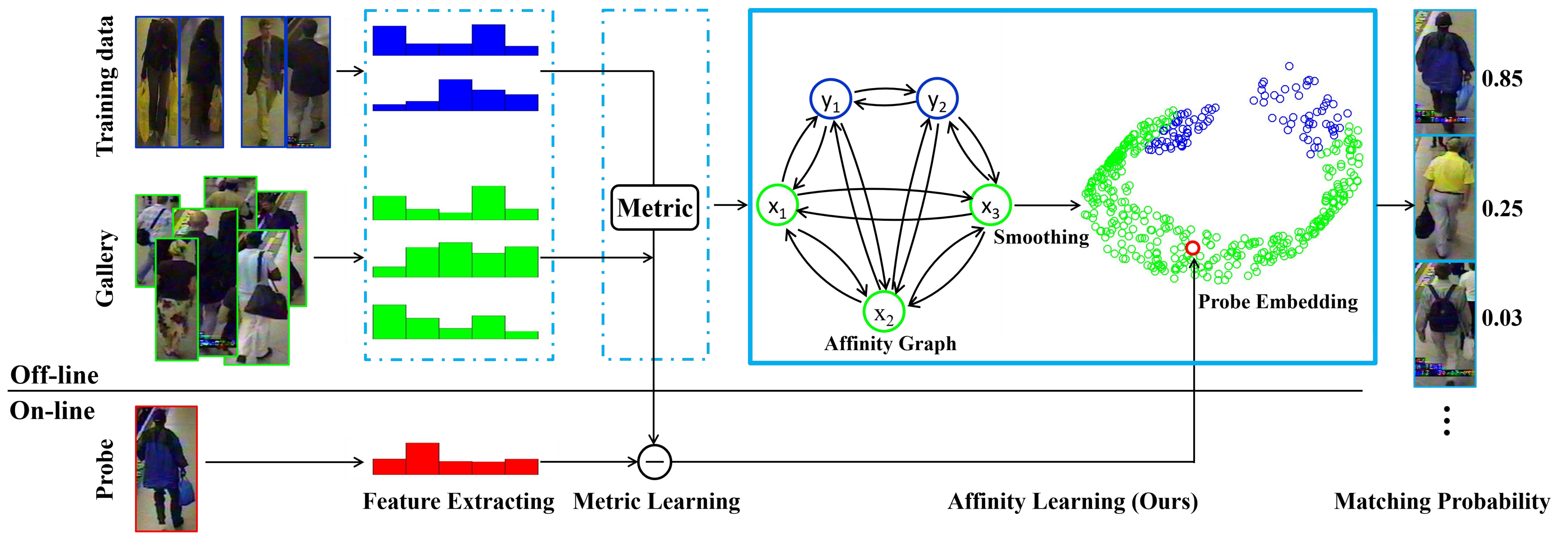


Figure 1. The pipeline of a person re-identification system. The blue, green and red color indicate training data, gallery and probe, respectively. Previous works focus on feature extracting and metric learning, marked with dashed boxes. Our work can be the post-processing procedure about affinity learning, marked with a solid box.

Basic Methodology

❑ Goal

Given the probe **p**, the testing gallery $\mathbf{X}=\{x_1,x_2,\dots,x_{N_g}\}$, we aim at learning a similarity $\mathbf{Q}\in\mathbb{R}^{N\times N}$, with the help of the labeled training set $\mathbf{Y}=\{y_1,y_2,\dots,y_{N_l}\}$.

Construct the affinity graph **G={V,W}**, where $\mathbf{V}=p\cup X\cup Y$, and $\mathbf{W}\in\mathbb{R}^{N\times N}$ is the adjacency matrix. $\mathbf{N=1+N_g+N_l}$.

❑ Observation

For the label set **L** used in ReID, if v_i and v_j belong to the same identity, then $L_{ij}=1$, otherwise **0**. For the similarity **Q**, Q_{ij} should be larger if v_i and v_j are similar, and should be zero if dissimilar. Hence, affinity learning can be done by propagating the pairwise constraint label **L** with tuples as primitive data.

❑ Supervised Similarity Propagation

Let (v_k,v_i) and (v_l,v_j) be two tuples, the propagation is defined as

$$Q_{ki}^{(t+1)} = \alpha \sum_{l,j}^{N} \mathcal{P}(ki \to lj) Q_{lj}^{(t)} + (1-\alpha) L_{ki}$$

We hold the *product rule* to calculate $\mathcal{P}(ki \to lj) = P(k \to l)P(i \to j) = P_{kl}P_{ij}$.

The iteration converges to $Q = vec^{-1}\left((1-\alpha)(I-\alpha\mathcal{P})^{-1}\vec{L}\right)$.

❑ Similarity Crop

Since **Q** can be divided into $Q = \begin{bmatrix} Q_{pp} & Q_{pX} & Q_{pY} \\ Q_{Xp} & Q_{XX} & Q_{XY} \\ Q_{Yp} & Q_{YX} & Q_{YY} \end{bmatrix}$, we can obtain the matching probabilities between the probe **p** and the gallery **X**, *i.e.*, Q_{pX} by cropping **Q**.

Computationally expensive, since the iteration has to be done

1) **on all the enumerated tuples → $O(TN^4)$.**
2) **once a new probe is observed → $O(TN_pN^4)$.**

Re-identification on-the-fly

❑ Iteration transform

The iteration degenerates to

$$Q^{(t+1)} = \alpha P Q^{(t)} P^{\mathrm{T}} + (1-\alpha)L$$

Consequently, the time complexity of the iteration is reduced to $O(TN^3)$.

❑ Probe embedding

A useful equivalence: the converged solution can be derived by solving

$$\min_Q \Phi(Q) + \frac{1-\alpha}{\alpha}\Omega(Q),$$

where $\underbrace{\Phi(Q) = \frac{1}{2}\sum_{i,j,k,l}^{N} P_{ij}P_{kl}(Q_{ki}-Q_{lj})^2}_{\text{Smoothness criterion}}$ and $\underbrace{\Omega(Q) = \sum_{k,i=1}^{N}(Q_{ki}-L_{ki})^2.}_{\text{Fitting term}}$

Compared with the large database (testing gallery and labeled data), there is only one probe **p** at each testing time. We hold two assumptions that

1) the database itself constitutes an underlying manifold;
2) when **p** is embedded into the manifold smoothly, it will not alter its geometry structure.

Therefore, we can first perform affinity learning off-line with only database instances, then do the probe embedding on-line following the smoothness criterion

$$\frac{\partial\Phi(Q)}{\partial Q_{pi}} = \sum_{j,l=1}^{N_g+N_l} P_{ij}P_{pl}(Q_{pi}-Q_{lj}).$$

By setting it to zero and varying $v_i \in \mathbf{X}$, we can have

$$Q_{pX} = \begin{bmatrix} P_{pX} & P_{pY} \end{bmatrix} \begin{bmatrix} Q_{XX} & Q_{XY} \\ Q_{YX} & Q_{YY} \end{bmatrix} \begin{bmatrix} P_{XX}^{\mathrm{T}} \\ P_{XY}^{\mathrm{T}} \end{bmatrix}.$$

The final time complexity is reduced to $O(N_p(N_g+N_l)N_g)$.

References

[1] X. Zhu and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical report, 2002.
[2] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. In *NIPS*, 2003.
[3] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Scholkopf. Ranking on data manifolds. In *NIPS*, 2004.
[4] M. Donoser and H. Bischof. Diffusion processes for retrieval revisited. In *CVPR*, 2013.
[5] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In CVPR, 2015.
[6] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato. Hierarchical gaussian descriptor for person re-identification. In CVPR, 2016.
[7] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification: What features are important? In ECCV, 2012.

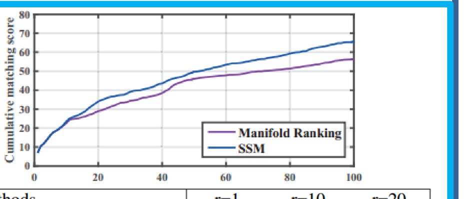
Experiments

❑ Datasets: GRID, VIPeR, PRID450S, CUHK03 and Market-1501.
❑ Features: LOMO [4], GOG [5], EFL6 [6].
❑ Metric: Euclidean and XQDA [4].

Results on GRID

Feature	Metric	Affinity	r=1	r=10	r=20
ELF6	Euclidean	×	4.64	19.60	28.00
ELF6	Euclidean	√	6.96	23.44	34.08
ELF6	XQDA	×	10.48	38.64	**52.56**
ELF6	XQDA	√	**11.04**	**40.72**	51.76
LOMO	Euclidean	×	15.20	30.80	36.40
LOMO	Euclidean	√	16.00	33.68	41.60
LOMO	XQDA	×	16.56	41.84	52.40
LOMO	XQDA	√	**18.96**	**44.16**	**55.92**
GOG	Euclidean	×	13.28	33.76	44.40
GOG	Euclidean	√	14.40	36.80	44.48
GOG	XQDA	×	24.80	58.40	68.88
GOG	XQDA	√	**26.16**	**59.20**	**70.40**
Fusion	Euclidean	×	14.72	35.44	45.84
Fusion	Euclidean	√	17.76	37.60	44.48
Fusion	XQDA	×	27.04	59.36	70.00
Fusion	XQDA	√	**27.20**	**61.12**	**70.56**
Fusion*	Euclidean	×	14.80	35.60	46.24
Fusion*	Euclidean	√	15.92	35.60	46.40
Fusion*	XQDA	×	27.20	61.12	71.20
Fusion*	XQDA	√	**27.60**	**62.56**	**71.60**

Table 2. The comparison with baselines on GRID dataset. √ indicates SSM is used and × indicates not used.



Methods	r=1	r=10	r=20
ELF6+RankSVM [39]	10.24	33.28	43.68
ELF6+PRDC [64]	9.68	32.96	44.32
ELF6+RankSVM+MR [30]	12.24	36.32	46.56
ELF6+PRDC+MR [30]	10.88	35.84	46.40
ELF6 + XQDA [25]	10.48	38.64	52.56
LOMO + XQDA [25]	16.56	41.84	52.40
MLAPG [26]	16.64	41.20	52.96
NLML [18]	24.54	43.53	55.25
PolyMap [9]	16.30	46.00	57.60
SSDAL [46]	22.40	48.00	58.40
MtMCML [33]	14.08	45.84	59.84
LSSCDL [58]	22.40	51.28	61.20
KEPLER [35]	18.40	50.24	61.44
DR-KISS [47]	20.60	51.40	62.60
SCSP [8]	24.24	54.08	65.20
GOG+XQDA [36]	24.80	58.40	68.88
SSM (Ours)	27.20	61.12	70.56

Table 3. The comparison with state-of-the-art on GRID dataset. The best and second best performances are marked in red and blue, respectively.

VIPeR, PRID450S and CUHK03

The default setup: the concatenation of LOMO and GOG under XQDA metric.

Methods	Ref	r=1	r=10	r=20
Local Fisher [38]	CVPR2013	24.18	67.12	-
eSDC [60]	CVPR2013	26.74	62.37	76.36
SalMatch [59]	ICCV2013	30.16	-	-
Mid-Filter [61]	CVPR2014	29.11	65.95	79.87
SCNCD [54]	ECCV2014	37.80	81.20	90.40
ImprovedDeep [1]	CVPR2015	34.81	-	-
PolyMap [9]	CVPR2015	36.80	83.70	91.70
XQDA [25]	CVPR2015	40.00	80.51	91.08
Semantic [44]	CVPR2015	41.60	86.20	95.10
MetricEmsemb. [37]	CVPR2015	45.90	88.90	95.80
QALF [63]	ICCV2015	30.17	62.44	73.81
CSL [42]	ICCV2015	34.80	82.30	91.80
MLAPG [26]	ICCV2015	40.73	82.34	92.37
MTL-LORAE [45]	ICCV2015	42.30	81.60	89.60
DCIA [13]	ICCV2015	63.92	87.50	-
DGD [51]	CVPR2016	38.60	-	-
LSSCDL [58]	CVPR2016	42.66	84.27	91.93
TPC [10]	CVPR2016	47.80	84.80	91.10
GOG [36]	CVPR2016	49.72	88.67	94.53
Null [56]	CVPR2016	51.17	90.51	95.92
SCSP [8]	CVPR2016	53.54	91.49	96.65
S-CNN [48]	ECCV2016	37.80	66.90	-
Shi *et al.* [43]	ECCV2016	40.91	-	-
ℓ1-graph [20]	ECCV2016	41.50	-	-
S-LSTM [49]	ECCV2016	42.40	79.40	-
SSDAL [46]	ECCV2016	43.50	81.50	89.00
TMA [34]	ECCV2016	48.19	87.65	93.54
SSM (Ours)		53.73	91.49	96.08

Table 5. The comparison with state-of-the-art on VIPeR dataset.

Methods	Ref	r=1	r=10	r=20
SCNCD [7]	ECCV2014	41.60	79.40	87.80
Semantic [6]	CVPR2015	44.90	77.50	86.70
CSL [5]	ICCV2015	44.40	82.20	89.80
XQDA [1]	CVPR2015	61.38	91.02	95.33
TMA [2]	ECCV2016	52.89	85.78	93.33
LSSCDL [8]	CVPR2016	60.49	88.58	93.60
GOG [3]	CVPR2016	68.40	94.50	97.80
SSM (Ours)		72.98	96.76	99.11

Table 1. The comparison with state-of-the-art on PRID450S.

Methods	Labeled			Detected		
	r=1	r=5	r=10	r=1	r=5	r=10
DeepReID [23]	20.7	51.7	68.3	19.9	49.0	64.3
XQDA [25]	52.2	-	-	46.3	-	-
ImprovedDeep [1]	54.7	88.3	93.3	45.0	75.7	83.0
LSSCDL [58]	57.0	-	-	51.2	-	-
MLAPG [26]	58.0	-	-	51.2	-	-
Shi *et al.* [43]	61.3	-	-	52.0	-	-
MetricEmsemb. [37]	62.1	89.1	94.3	-	-	-
Null [56]	62.5	90.0	94.8	54.7	84.7	94.8
S-LSTM [49]	-	-	-	57.3	80.1	88.3
S-CNN [48]	-	-	-	61.8	80.9	88.3
GOG [36]	67.3	91.0	96.0	65.5	88.4	93.7
DGD [51]	75.3	-	-	-	-	-
SSM (Ours)	76.6	94.6	98.0	72.7	92.4	96.1

Table 6. The comparison with state-of-the-art on CUHK03 dataset.

On Market-1501, we report mAP **68.80** with single query and **76.18** with multiple query. Please refer to our paper for the details.

As SSM manages transferring the graph-based affinity learning to off-line, the off-line cost is increased especially on larger datasets. In on-line stage, the extra indexing time brought by SSM only occupies a small percentage on all the datasets.

Datasets	Off-line		On-line	
	#M	#A	#M	#A
GRID	0.90*s*	+2.38*s*	0.17*s*	+10.3*ms*
VIPeR	2.19*s*	+2.22*s*	0.19*s*	+10.2*ms*
PRID450S	1.21*s*	+0.78*s*	0.12*s*	+3.80*ms*
CUHK03	789.6*s*	+1952*s*	0.09*s*	+0.516*s*
Market1501	-	+2769*s*	146.11*s*	+21.68*s*

Table 8. #M denotes the initial time cost of metric learning using XQDA. #A denotes the extra cost brought by the proposed SSM.

Conclusions

In this paper, we do not design robust features or metrics that are superior to others in person re-identification. Instead, we contribute a generic tool called Supervised Smoothed Manifold (SSM), upon which most existing algorithms can easily boost their performances further. SSM is very easy to implement. It can also handle the special kind of labeled data and has potential capacity in large scale ReID. Comprehensive experiments on five benchmarks demonstrate that SSM not only achieves the best performances, but more importantly, incurs acceptable extra on-line cost. In the future, we will investigate how to effectively fuse multiple features and apply the proposed SSM to other datasets.

Contact

Name: Song Bai Email: songbai@hust.edu.cn
Homepage: https://sites.google.com/site/songbaihust/

*************** I am looking for a job or postdoc *******************