# 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
{songbai,xbai}@hust.edu.cn, qi.tian@utsa.edu

## Abstract

*Most 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 postprocessing procedure to further boost the identification accuracies. 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.*

## 1. Introduction

Person re-identification (ReID) is 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. Due to the large intra-class variations in viewpoint, pose, illumination, blur and occlusion, person re-identification is still a rather challenging task, though extensively studied in recent years.

Current research interests can be coarsely divided into two mainstreams: 1) those focus on designing robust visual descriptors [55, 13, 37, 26, 64] to accurately model the appearance of person; 2) those seek for a discriminative metric [65, 25, 54, 28, 18], under which instances of the same identity should be closer while instances of different identities are far away.

Unlike those methods performed in the metric space, we investigate person re-identification task from another perspective, *i.e.*, taking into account the manifold structure [42]. Since existing methods only analyze the pairwise distances between instances, the underlying data manifold, which those images reside on, is more or less neglected. It results in that the learned relationships (similarities or dissimilarities) between instances are not smooth with respect to the local geometry of the manifold.

To overcome this issue, potential solutions can be semi-supervised [68, 66] or unsupervised [67, 5, 58, 12, 6] algorithms about manifold learning. However, directly applying such algorithms to person re-identification might be problematic for two reasons. First, semi-supervised algorithms (*e.g.*, label propagation [68]) can only predict the labels of unlabeled data, but fail to depict the relationship between the probe and gallery instances. Moreover, they require category labels, while supervision in ReID is given as pairwise (equivalence) constraints [22]. Meanwhile, unsupervised algorithms (*e.g.*, manifold ranking [67], graph transduction [7]) totally ignore the beneficial influence from the labeled training data. Second, since most manifold learning algorithms operate on graph models, their algorithmic complexity is usually high. Therefore, the heavy computational cost hinders their promotions in this field, especially in recent years researchers begin to attach more importance to the scalability issue [24, 63]. In summary, due to the above factors, those conventional manifold learning algorithms are inadequate to derive a more faithful similarity for person re-identification.

In this paper, we tackle person re-identification task on the data manifold by proposing a novel affinity learning algorithm called Supervised Smoothed Manifold (SSM). Compared with existing algorithms, the primary contribution of SSM is that 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.

Moreover, SSM is customized specifically for person re-identification, which further possesses three merits (as illustrated in Fig. 1) as follows: i) **supervision**: instead of

*Corresponding author.

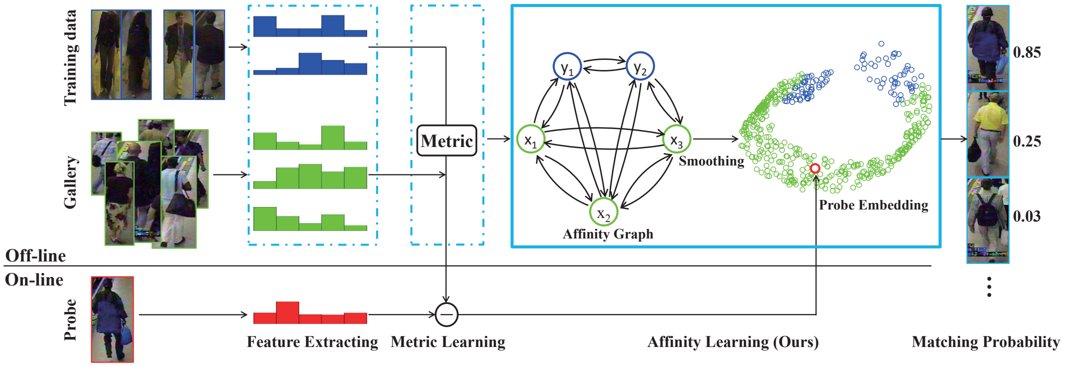

Figure 1. The pipeline of a person re-identification system. The blue, green and red color indicate training data, gallery and probe, respectively. Previous works concentrate on feature extracting and metric learning, marked with dashed boxes. Our work can be the postprocessing procedure about affinity learning, marked with a solid box. Sample images come from GRID dataset [33].

considering each instance individually, we propose to learn the similarity with instance pairs. By doing so, SSM can take advantage of the supervision in pairwise constraints, which is easily accessible in this task; ii) **efficiency**: to overcome the limitation of high time complexity of SSM, two improvements are proposed to accelerate its on-line person matching. Consequently, the affinity learning is performed only with database instances off-line, and SSM can be applied to the scenario on large scale person re-identification; and iii) **generalization**: different from most existing algorithms performed in metric space, SSM focuses on affinity learning between instances. Hence, SSM can be deemed as a postprocessing procedure (or a generic tool) to further boost the identification accuracies of those algorithms.

The rest of the paper is organized as follows. In Sec. 2, we present the differences between SSM and relevant works. The basic affinity learning framework of SSM is introduced in Sec. 3, and significantly accelerated in Sec. 4. Experiments are presented in Sec. 5. Conclusions and future works are given in Sec. 6.

## 2. Related Work

The manifold structure has been observed by several works. Motivated by the fact that pedestrian data are distributed on a highly curved manifold, a sampling strategy for training neural network called Moderate Positive Mining (MPM) is proposed in [44]. However, considering the data distribution is hard to define, MPM does aim at estimating the geodesic distances along the manifold. From this point of view, SSM explicitly learns the geodesic distances between instances, which can be directly used for re-identification.

Manifold ranking [67] is introduced by [31] to person re-identification. Through a random walk [2] on the affinity graph, it propagates the probe label to the gallery iteratively assuming that the probe is the only labeled data. Despite the ignorance of labeled training data as analyzed above, manifold ranking encounters severe obstacles when handling larger databases, since the graph-based iteration has to be run each time a new probe is observed. In this aspect, SSM also learns the similarities via iterative propagation. Nevertheless, it enables a highly-efficient on-line matching.

Post-ranking techniques have not drawn much attention in this field. Most of them require human feedback in-the-loop [3, 17], such as Post-rank OPtimisation (POP) [29], Human Verification Incremental Learning (HVIL) [51]. Meanwhile, several works [4, 23] operate in an unsupervised manner. For example, Discriminant Context Information Analysis (DCIA) [14] focuses on the visual ambiguities shared between the first ranks, where the true match is supposed to be located. In comparison, SSM does not need human interaction or hold the “rank-1” assumption. Instead, its essence is to learn a smooth similarity measure, supervised by the special kind of labels in pairwise constraints.

At the first glance, affinity learning in our work appears the same as similarity learning (*e.g.*, PolyMap [10]). Unlike similarity learning on polynomial feature map [9] which connects to Mahalanobis distance metric and bilinear similarity, affinity learning in SSM does not rely on the definition of metric (non-metric can be also used). Therefore, they are inherently different. Finally, it is acknowledged that those metric learning methods (*e.g.*, KISSME [22], XQDA [26]) are also relevant, but take effects prior to SSM in a person re-identification system as Fig. 1 shows.

## 3. Proposed Method

Given a probe $p$ and a testing gallery $X = \{x_1, x_2, \dots, x_{N_g}\}$, we aim at learning a smooth similarity $Q \in \mathbb{R}^{N \times N}$ with the help of the labeled training set $Y = \{y_1, y_2, \dots, y_{N_l}\}$, where $N = N_g + N_l + 1$. The data manifold is modeled as a weighted affinity graph $\mathcal{G} = \{V, W\}$. The vertex set $V = \{v_1, v_2, \dots, v_N\}$ is equivalent to the union of the probe $p$ and the database instances (gallery $X$ and labeled set $Y$). $W \in \mathbb{R}^{N \times N}$ is the adjacency matrix of $\mathcal{G}$, with $W_{ij}$ measuring the similarity between vertex $v_i$ and $v_j$. To facilitate a random walk [2] on the graph $\mathcal{G}$, a transition matrix $P \in \mathbb{R}^{N\times N}$ is usually needed. The transition probability from vertex $v_i$ to $v_j$ can be calculated as

$$P(i \rightarrow j) = P_{ij} = \frac{W_{ij}}{\sum_{j'=1}^{N} W_{ij'}}. \tag{1}$$

Thus, $P$ is a row stochastic matrix.

### 3.1. Supervised Similarity Propagation

The label set $L \in \mathbb{R}^{N\times N}$ used in person re-identification is given in pairwise constraints, *i.e.*, if $v_i$ and $v_j$ belong to the same identity, $L_{ij} = 1$, otherwise $L_{ij} = 0$. Meanwhile, in the ideal case, the learned similarity $Q_{ij}$ should be larger if $v_i$ and $v_j$ belong to the same identity, and $Q_{ij}$ should be close to 0 otherwise. Therefore, we can conclude that both $L$ and $Q$ provide a probabilistic interpretation to the likelihood of the tuple $(v_i, v_j)$ being a true matching pair. The difference is that $L_{ij}$ is a discrete binary variable, indicating exactly matching or not, while $Q_{ij}$ is a continuous variable, specifying a matching degree. Such an observation motivates us that affinity learning can be done by propagating the pairwise constraint label $L$ with tuples as primitive data. In other words, similarities are spread from the most confident tuples generated from the labeled set $Y$ to the unexplored tuples generated from the testing gallery $X$.

Let $(v_k, v_i)$ and $(v_l, v_j)$ be two tuples, the propagation step in the $t$-th iteration is defined as

$$Q_{ki}^{(t+1)} = \alpha \sum_{l,j}^{N} \mathcal{P}(ki \rightarrow lj) Q_{lj}^{(t)} + (1-\alpha) L_{ki}, \tag{2}$$

where $\mathcal{P}(ki \rightarrow lj)$ is the transition probability from tuple $(v_k, v_i)$ to tuple $(v_l, v_j)$, and $0 < \alpha < 1$. Eq. (2) reveals that at each iteration, the tuple $(v_k, v_i)$ absorbs a fraction of label information from the rest tuples with probability $\alpha$, then retains its initial label $L_{ki}$ with probability $1-\alpha$. Assuming the independence within tuples, we hold the *product rule* to calculate $\mathcal{P}(ki \rightarrow lj)$, as

$$\mathcal{P}(ki \rightarrow lj) = P(k \rightarrow l) P(i \rightarrow j) = P_{kl} P_{ij}. \tag{3}$$

Afterwards, Eq. (2) can be rewritten in matrix form

$$\vec{Q}^{(t+1)} = \alpha \mathcal{P} \vec{Q}^{(t)} + (1-\alpha)\vec{L}. \tag{4}$$

To prove this, we need two identical coordinate transformations, that is $\mu \equiv N(i-1)+k$ and $\nu \equiv N(j-1)+l$. Then $Q$ can be vectorized to $\vec{Q} = vec(Q) \in \mathbb{R}^{N^2\times 1}$, with the element correspondence $\vec{Q}_\mu = Q_{ki}$. Let $\mathcal{P} \in \mathbb{R}^{N^2\times N^2}$ be the Kronecker product of $P$ with itself, *i.e.*, $\mathcal{P} = P \otimes P$. Then, the correspondence between $\mathcal{P}$ and $P$ is given as $\mathcal{P}_{\mu\nu} = P_{ij}P_{kl}$. Eventually, Eq. (2) can be expressed as

$$\vec{Q}_\mu^{(t+1)} = \alpha \sum_{\nu=1}^{N^2} \mathcal{P}_{\mu\nu} \vec{Q}_\nu^{(t)} + (1-\alpha)\vec{L}_\mu. \tag{5}$$

The proof is complete.

### 3.2. Convergence Proof

By running the iteration for $t$ times, Eq. (4) can be expanded as

$$\vec{Q}^{(t+1)} = (\alpha\mathcal{P})^t \vec{Q}^{(1)} + (1-\alpha)\sum_{i=0}^{t-1} (\alpha\mathcal{P})^i \vec{L}. \tag{6}$$

$\mathcal{P}$ is also a row stochastic matrix, since

$$\sum_\nu \mathcal{P}_{\mu\nu} = \sum_{l,j} P_{ij}P_{kl} = \sum_j P_{ij} \sum_l P_{kl} = 1. \tag{7}$$

Therefore, according to *Perron-Frobenius Theorem*, we can obtain that spectral radius of $\mathcal{P}$ is bounded by 1, the maximum value of its row sums. Considering that $0 < \alpha < 1$, we have

$$\lim_{t\rightarrow\infty} (\alpha\mathcal{P})^t = 0, \quad \lim_{t\rightarrow\infty} \sum_{i=0}^{t-1} (\alpha\mathcal{P})^i = (I - \alpha\mathcal{P})^{-1}, \tag{8}$$

where $I$ is an identity matrix in appropriate size. Consequently, Eq. (6) converges to

$$\lim_{t\rightarrow\infty} \vec{Q}^{(t+1)} = (1-\alpha)(I-\alpha\mathcal{P})^{-1}\vec{L}. \tag{9}$$

Then $Q$ can be obtained by reshaping $\vec{Q}$ to matrix form as $Q = vec^{-1}(\vec{Q})$.

### 3.3. Basic Pipeline

Intuitively, person re-identification using the above affinity learning algorithm can be accomplished in three steps. First, each time a probe instance $p$ is observed, the affinity graph $\mathcal{G}$ is constructed. Second, a new similarity $Q$ is learned by either running Eq. (4) until convergence or directly using the closed-form solution in Eq. (9). At last, since $Q$ can be divided into

$$Q = \begin{bmatrix} Q_{pp} & Q_{pX} & Q_{pY} \\ Q_{Xp} & Q_{XX} & Q_{XY} \\ Q_{Yp} & Q_{YX} & Q_{YY} \end{bmatrix}, \tag{10}$$

we can obtain the matching probabilities between the probe $p$ and the gallery $X$, that is $Q_{pX} \in \mathbb{R}^{1\times N_g}$. Note that $W$ and $P$ also have such a division.

We draw readers' attention that when the probe $p$ is used for testing, the other probe instances are invisible to users. Therefore, one cannot simultaneously include all the probe instances to constitute $\mathcal{G}$ for a global probe search.

However, this pipeline is computationally too demanding in practice. First, affinity learning itself is computationally expensive. It requires time complexity $O(TN^4)$ and space complexity $O(N^4)$ to run the iteration in Eq. (4), where $T$ is the iteration number. Alternatively, using the closed-form solution in Eq. (9) requires time complexity

$O(N^6)$ and space complexity $O(N^4)$, since we need to invert and store a huge matrix of size $N^2 \times N^2$.

Second, adapting new probe instances is computationally expensive. As our method is algorithmically graph-based, we need to discard the old probe and do the affinity learning at each time a new probe is observed. Assume we have $N_p$ probe instances in total, we at least need time complexity $O(TN_pN^4)$ to finish the whole probe search. Note that constructing the affinity graph is computationally cheap due to the fact that the similarities between database instances can be pre-computed off-line for once and reused consistently.

## 4. Re-identification on-the-fly

In this section, we propose two modifications to decrease the high complexity of the basic pipeline in Sec. 3, such that person re-identification can be done on-the-fly.

### 4.1. Iteration Transform

Our first improvement focuses on affinity learning itself. We observe the following useful identity

$$\mathcal{P}\vec{Q} = (P \otimes P)vec(Q) = vec(PQP^{\mathrm{T}}). \tag{11}$$

So, Eq. (4) can be transformed into

$$Q^{(t+1)} = \alpha P Q^{(t)} P^{\mathrm{T}} + (1-\alpha)L. \tag{12}$$

As a result, the time and space complexity of affinity learning are reduced to $O(TN^3)$ and $O(N^2)$, respectively.

### 4.2. Probe Embedding

Our second improvement concentrates on improving the efficiency in adapting new probe instances. First, we prove that the closed-form solution in Eq. (9) can be derived from

$$\min_{Q} \Phi(Q) + \frac{1-\alpha}{\alpha}\Omega(Q), \tag{13}$$

where

$$\begin{aligned}\Phi(Q) &= \frac{1}{2}\sum_{i,j,k,l}^{N} P_{ij}P_{kl}(Q_{ki}-Q_{lj})^2,\\ \Omega(Q) &= \sum_{k,i=1}^{N}(Q_{ki}-L_{ki})^2.\end{aligned} \tag{14}$$

Using the two identical coordinate transformations, Eq. (13) can be vectorized, where

$$\Phi(\vec{Q}) = \frac{1}{2}\sum_{\mu,\nu}^{N^2}\mathcal{P}_{\mu\nu}(\vec{Q}_\mu - \vec{Q}_\nu)^2,\ \ \Omega(\vec{Q}) = \|\vec{Q}-\vec{L}\|_2^2. \tag{15}$$

$\Phi(\vec{Q})$ measures the smoothness of $\vec{Q}$ with respect to the local manifold structure, and $\Omega(\vec{Q})$ measures the fitness of $\vec{Q}$ to the given label $\vec{L}$.

The derivative of $\Phi(\vec{Q})$ with respect to $\vec{Q}$ is

$$\frac{\partial\Phi(\vec{Q})}{\partial\vec{Q}} = \left((I-\mathcal{P}) + (I-\mathcal{P})^{\mathrm{T}}\right)\vec{Q}. \tag{16}$$

According to [8], Eq. (16) can be approximated by $2(I-\mathcal{P})\vec{Q}$. So, one can easily induce the derivative of Eq. (13) with respect to $\vec{Q}$

$$2(I-\mathcal{P})\vec{Q} + \frac{2(1-\alpha)}{\alpha}(\vec{Q}-\vec{L}). \tag{17}$$

By setting Eq. (17) to zero and applying $vec^{-1}$ operator, we can get the closed-form solution of Eq. (13)

$$Q = vec^{-1}\left((1-\alpha)(I-\alpha\mathcal{P})^{-1}\vec{L}\right), \tag{18}$$

which is equivalent to Eq. (9). The proof is complete.

Compared with the large database (testing gallery and labeled data), there is only one probe $p$ at each testing time. Therefore, 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. With these prerequisites, we can first perform affinity learning off-line with only database instances, then do the probe embedding on-line.

Of course, the embedding of the probe should also follow the smoothness criterion $\Phi(Q)$. After the pairwise similarities between database instances are smoothed, the partial derivative of $\Phi(Q)$ with respect to $Q_{pi}$ is

$$\frac{\partial\Phi(Q)}{\partial Q_{pi}} = \sum_{j,l=1}^{N_g+N_l} P_{ij}P_{pl}(Q_{pi}-Q_{lj}). \tag{19}$$

Setting it to zero, the similarity between the probe $p$ and a certain database instance $v_i$ can be calculated

$$Q_{pi} = \sum_{j,l=1}^{N_g+N_l} P_{pl}Q_{lj}P_{ij}. \tag{20}$$

By varying $v_i \in X$, Eq. (20) can be rewritten in matrix form

$$Q_{pX} = \begin{bmatrix}P_{pX} & P_{pY}\end{bmatrix}\begin{bmatrix}Q_{XX} & Q_{XY}\\ Q_{YX} & Q_{YY}\end{bmatrix}\begin{bmatrix}P_{XX}^{\mathrm{T}}\\ P_{XY}^{\mathrm{T}}\end{bmatrix}. \tag{21}$$

### 4.3. Complexity Analysis

The final pipeline of the proposed SSM is rather simple, summarized in Alg. 1. As can be seen, affinity learning is done only with database instances. The computational cost still seems a bit heavy, since there are $(N_g + N_l)$ vertices

```
Algorithm 1: Supervised Smoothed Manifold.
  Input: The probe $p$, the testing gallery $X$, the labeled
    data $Y$, the training label $L$.
  Output: The matching probability $Q_{pX}$.
  begin
      Off-line:
      begin
          Construct the affinity graph with $X$ and $Y$;
          Affinity learning with label $L$ using Eq. (12).
          return $Q_{XX}$, $Q_{XY}$, $Q_{YX}$, $Q_{YY}$
      On-line:
      begin
          for each probe $p$ do
              Do pedestrian matching using Eq. (21);
              return $Q_{pX}$.
```

| Methods | Time Complexity | Space Complexity |
|---|---|---|
| Standard | $O(TN_pN^4)$ | $O(N^4)$ |
| Accelerated | $O\left(N_p(N_g+N_l)N_g\right)$ | $O\left((N_g+N_l)^2\right)$ |

Table 1. The complexity comparison between the standard solution and the accelerated solution of SSM. Recall that $N_p$ is the number of probe, $N_g$ is the number of gallery, $N_l$ is the number of labeled data, and $T$ is the number of iterations. $N = N_g + N_l + 1$.

in the graph. However, those operations can be done offline, and reused with different probe instances. The learned similarities can all be maintained dynamically as long as new database instances are added or distance matrices are changed.

In Table 1, we present the on-line complexity comparison between the standard solution in Sec. 3 and the accelerated solution in Sec. 4. Eq. (21) reveals that on-line indexing for $N_p$ probe instances involves the multiplication of three matrices. Whereas the multiplication of the right two can be also computed off-line, the on-line time complexity is only $O\left(N_p(N_g+N_l)N_g\right)$. Furthermore, the space complexity is dominated by the storage of the learned similarity, requiring $O\left((N_g+N_l)^2\right)$.

## 5. Experiments

The proposed Supervised Smoothed Manifold (SSM) is evaluated on five popular benchmarks, including GRID [33, 32], VIPeR [15], PRID450S [41], CUHK03 [24] and Market-1501 [63]. In the implementations of SSM, we do not carefully tune parameters, but fix $\alpha = 0.1$ and the number of iterations $T = 30$ throughout our experiments. The affinity graph is constructed by applying self-tuning [56] Gaussian kernel to pairwise distances following [31].

### 5.1. QMUL GRID

QMUL underGround Re-IDentification (GRID) [33, 32] is a challenging dataset, which has gradually become popular. The variations in the pose, colors and illuminations of pedestrians, as well as the poor image quality, make it very difficult to yield high matching accuracies.

GRID dataset consists of 250 identities, with each identity having two images seen from different camera views. Besides, 775 additional images that do not belong to the 250 identities are used to enlarge the gallery. Sample images can be found in Fig. 1. A fixed training/testing split with 10 trials is provided. For each trial, 125 image pairs are used for training. The remaining 125 image pairs and the 775 background images are used for testing. To evaluate the performances, we employ Cumulated Matching Characteristics (CMC) curves and the cumulated matching accuracy at selected ranks.

To obtain the image representations, we utilize two representative descriptors, *i.e.*, Local Maximal Occurrence (LOMO) [26] and Gaussian Of Gaussian (GOG) [37]. In addition, ELF6 feature [30], provided along with the dataset, is also tested to ensure the fair comparison.

**Comparison with Baselines.** In Table 2, we present the performances before and after SSM is used. Besides the three individual visual features, two types of fused features are also used. *Fusion* means the concatenation of LOMO and GOG, while *Fusion*$^\star$ means the concatenation of all the three features, both with equal weights. The pairwise distances between instances are computed in metric space. In our experiments, besides the natural choice of Euclidean metric, we also evaluate Cross-view Quadratic Discriminant Analysis (XQDA) [26] which is taken as a representative of metric learning techniques.

As can be drawn, SSM leads to considerable performance gains against the baselines. For example, with ELF6 in Euclidean metric, the improvement of identification rate brought by SSM is 2.32 at rank-1, 3.84 at rank-10 and 6.08 at rank-10. Meanwhile, by integrating XQDA, SSM can still boost the performances further. For example, the rank-1 accuracy of LOMO with XQDA is originally 16.56, then increased to 18.96 after the proposed SSM is used. Those experimental results suggest that most existing visual features or metric learning algorithms in person re-identification are compatible with SSM. In other words, after visual features are given, person re-identification systems can be improved with two steps, *i.e.*, applying metric learning first, and applying SSM next.

As a related work to ours, manifold ranking [31] reports an identification rate of 30.96 at rank-20 using ELF6 and Euclidean metric, which is significantly lower than 34.08 achieved by SSM. It clearly demonstrates that it is beneficial to exploit the supervision information in affinity learning step. To avoid the performance uncertainty (though rather tiny) led by different implementation details, we compare SSM with manifold ranking using exactly the same affinity graph. The results are given in Fig. 2. As we can see,

| 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$^\star$ | Euclidean | × | 14.80 | 35.60 | 46.24 |
| Fusion$^\star$ | Euclidean | √ | 15.92 | 35.60 | 46.40 |
| Fusion$^\star$ | XQDA | × | 27.20 | 61.12 | 71.20 |
| Fusion$^\star$ | XQDA | √ | **27.60** | **62.56** | **71.60** |

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

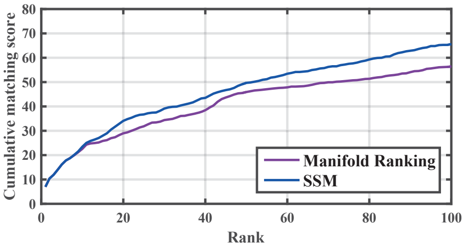

Figure 2. The comparison between the proposed SSM and Manifold Ranking with the same affinity graph.

| Methods | r=1 | r=10 | r=20 |
|---|---|---|---|
| ELF6+RankSVM [40] | 10.24 | 33.28 | 43.68 |
| ELF6+PRDC [65] | 9.68 | 32.96 | 44.32 |
| ELF6+RankSVM+MR [31] | 12.24 | 36.32 | 46.56 |
| ELF6+PRDC+MR [31] | 10.88 | 35.84 | 46.40 |
| ELF6 + XQDA [26] | 10.48 | 38.64 | 52.56 |
| LOMO + XQDA [26] | 16.56 | 41.84 | 52.40 |
| MLAPG [27] | 16.64 | 41.20 | 52.96 |
| NLML [19] | 24.54 | 43.53 | 55.25 |
| PolyMap [10] | 16.30 | 46.00 | 57.60 |
| SSDAL [47] | 22.40 | 48.00 | 58.40 |
| MtMCML [34] | 14.08 | 45.84 | 59.84 |
| LSSCDL [59] | 22.40 | 51.28 | 61.20 |
| KEPLER [36] | 18.40 | 50.24 | 61.44 |
| DR-KISS [48] | 20.60 | 51.40 | 62.60 |
| SCSP [9] | 24.24 | 54.08 | 65.20 |
| GOG+XQDA [37] | **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.

the rank-1 accuracy of both manifold ranking and SSM is 6.96. However, SSM outperforms manifold ranking by a large margin at higher ranks. The difference of accuracy is nearly 10 at rank-100.

One of the most important properties of SSM is its high time efficiency in on-line pedestrian matching. Here, we omit the overhead in constructing the affinity graph, which can be done off-line. Under the same computing platform, manifold ranking takes 14.40 seconds in total to fulfill searching 125 probe instances, while SSM only needs $9.52ms$. One can easily find that SSM is 3 orders of magnitude faster than manifold ranking. The reason behind is that the iteration of manifold ranking is conducted each time a new probe is observed. In comparison, SSM proposes to do affinity learning only off-line, and embed the probe smoothly into the manifold. As a result, although SSM has to do affinity learning on a much larger graph (to leverage the supervision from training data), the on-line cost can be controlled so that SSM has the potential ability of handling large-scale person re-identification. We will further discuss this aspect below.

From Table 2, failure cases of SSM can be also observed. As it suggests, the rank-20 accuracy of ELF6 under XQDA metric is originally 52.56, then is decreased slightly by SSM to 51.76. The reason behind such abnormal phenomena is that the principle of SSM is to obtain a global similarity measure between each two instances, which varies smoothly with respect to the local geometry of the underlying manifold. The learned similarity cannot guarantee that the identification rate at specifical ranks will be improved. But in general cases, the overall performances will be refined.

**Comparison with State-of-the-art.** In Table 3, we give a thorough comparison with other state-of-the-art methods. The performances of the proposed SSM are reported by using *Fusion* feature (the concatenation of LOMO and GOG) under XQDA metric, which is a default configuration used in our later experiments.

Previous state-of-the-art performances are achieved by Spatially Constrained Similarity function on Polynomial feature map (SCSP) [9] and GOG [37]. Chen *et al.* [9] impose spatially constraints to the similarity learning on polynomial feature map [10], and report rank-1 accuracy 24.24 by fusing 6 visual cues. GOG [37] is a powerful descriptor proposed recently, which captures the mean and the covariance information of pixel features. With XQDA metric, it reports the best performances on GRID dataset, *i.e.*, rank-1 accuracy 24.80. Benefiting from *Fusion* feature and XQDA metric, SSM easily sets a new state-of-the-art performance, outperforming the previous by 2.40 in rank-1 accuracy.

We emphasize that SSM is not restricted by the used descriptor and metric. Table 2 presents that SSM can achieve higher performances with *Fusion*$^\star$ feature.

## 5.2. VIPeR, PRID450S and CUHK03

VIPeR [15] is a widely-accepted benchmark for person re-identification containing 632 identities, and PRID450S [41] consists of 450 identities, both captured by two disjoint cameras. The widely adopted experimental protocol on two datasets is that a random selection of half persons is used for training and the rest for testing. The procedure is repeated for 10 times, then the average performances are reported.

CUHK03 [24] is among the largest public available benchmarks nowadays. It includes 13, 164 images of 1, 360 persons, with each person having 4.8 images on average. Besides manually cropped images, auto detected images are also provided. Following the conventional experimental setup [24, 26, 27, 37, 57], 1, 160 persons are used for training and 100 persons are used for testing. The experiments are conducted in single-shot setting with 20 random splits.

In Table 4, we present the performances of SSM and the baselines, where distances are calculated under XQDA metric. Consistent to previous experiments, SSM can easily boost the performances of baselines by around 2.5 percent on average. In particular, the performance improvements are more dramatic on CUHK03. For example, the rank-1 accuracy of *Fusion* is increased by 4.76 on CUHK03 labeled dataset, and by 4.65 on CUHK03 detected dataset. The preference of SSM on larger datasets stems from the fact that the manifold structure can be better sampled given more data points.

**Comparison on VIPeR.** Since enormous algorithms have reported results on VIPeR dataset, it is less realistic to exhibit all of them. Hence, we only include those published in recent 3 years or have close relationships with our work.

The comparison is given in Table 5. As can be seen, SSM yields the best rank-10 accuracy 91.49, which is the same as SCSP [9]. Meanwhile, SSM also achieves the second best performances at rank-1 and rank-20. To our best knowledge now, the best rank-1 accuracy is achieved by Discriminant Context Information Analysis (DCIA) [14]. The superiority of DCIA at rank-1 lies in that it tries to remove the visual ambiguities between the probe and its true match, which is supposed to be located at the first rank. By contrast, SSM does not hold such assumptions, which seem to be a bit strict in realistic settings. Thus, one can also observe that SSM outperforms DCIA by 3.99 at rank-10. Considering their inherent difference of principles, it can be anticipated that SSM and DCIA can benefit from each other, and a proper ensemble of them can lead to better performances.

**Comparison on PRID450S.** On PRID450S dataset, SSM provides the state-of-the-art performances on all the three evaluation metrics, *i.e.*, 72.98 at rank-1, 96.76 at rank-10, and 99.11 at rank-20. Due to space limitation, please refer to the supplementary material for the detailed comparison.

| Methods | Ref | r=1 | r=10 | r=20 |
|---|---|---|---|---|
| Local Fisher [39] | CVPR2013 | 24.18 | 67.12 | - |
| eSDC [61] | CVPR2013 | 26.74 | 62.37 | 76.36 |
| SalMatch [60] | ICCV2013 | 30.16 | - | - |
| Mid-Filter [62] | CVPR2014 | 29.11 | 65.95 | 79.87 |
| SCNCD [55] | ECCV2014 | 37.80 | 81.20 | 90.40 |
| ImprovedDeep [1] | CVPR2015 | 34.81 | - | - |
| PolyMap [10] | CVPR2015 | 36.80 | 83.70 | 91.70 |
| XQDA [26] | CVPR2015 | 40.00 | 80.51 | 91.08 |
| Semantic [45] | CVPR2015 | 41.60 | 86.20 | 95.10 |
| MetricEmsemb. [38] | CVPR2015 | 45.90 | 88.90 | 95.80 |
| QALF [64] | ICCV2015 | 30.17 | 62.44 | 73.81 |
| CSL [43] | ICCV2015 | 34.80 | 82.30 | 91.80 |
| MLAPG [27] | ICCV2015 | 40.73 | 82.34 | 92.37 |
| MTL-LORAE [46] | ICCV2015 | 42.30 | 81.60 | 89.60 |
| DCIA [14] | ICCV2015 | **63.92** | 87.50 | - |
| DGD [52] | CVPR2016 | 38.60 | - | - |
| LSSCDL [59] | CVPR2016 | 42.66 | 84.27 | 91.93 |
| TPC [11] | CVPR2016 | 47.80 | 84.80 | 91.10 |
| GOG [37] | CVPR2016 | 49.72 | 88.67 | 94.53 |
| Null [57] | CVPR2016 | 51.17 | **90.51** | 95.92 |
| SCSP [9] | CVPR2016 | 53.54 | **91.49** | **96.65** |
| S-CNN [49] | ECCV2016 | 37.80 | 66.90 | - |
| Shi *et al.* [44] | ECCV2016 | 40.91 | - | - |
| $\ell$1-graph [21] | ECCV2016 | 41.50 | - | - |
| S-LSTM [50] | ECCV2016 | 42.40 | 79.40 | - |
| SSDAL [47] | ECCV2016 | 43.50 | 81.50 | 89.00 |
| TMA [35] | 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 | Labeled | | | Detected | | |
|---|---|---|---|---|---|---|
| | r=1 | r=5 | r=10 | r=1 | r=5 | r=10 |
| DeepReID [24] | 20.7 | 51.7 | 68.3 | 19.9 | 49.0 | 64.3 |
| XQDA [26] | 52.2 | - | - | 46.3 | - | - |
| ImprovedDeep [1] | 54.7 | 88.3 | 93.3 | 45.0 | 75.7 | 83.0 |
| LSSCDL [59] | 57.0 | - | - | 51.2 | - | - |
| MLAPG [27] | 58.0 | - | - | 51.2 | - | - |
| Shi *et al.* [44] | 61.3 | - | - | 52.0 | - | - |
| MetricEmsemb. [38] | 62.1 | 89.1 | 94.3 | - | - | - |
| Null [57] | 62.5 | 90.0 | 94.8 | 54.7 | 84.7 | **94.8** |
| S-LSTM [50] | - | - | - | 57.3 | 80.1 | 88.3 |
| S-CNN [49] | - | - | - | 61.8 | 80.9 | 88.3 |
| GOG [37] | 67.3 | **91.0** | **96.0** | **65.5** | **88.4** | 93.7 |
| DGD [52] | **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.

**Comparison on CUHK03.** The comparison on CUHK03 dataset is given in Table 6. As it shows, the rank-1 identification rate of SSM is 72.7 with automatically detected bounding boxes, which is the first work reporting rank-1 accuracy larger than 70.

In [57], Zhang *et al.* overcome the small sample size (SSS) problem by matching people in a discriminative null space of the training data, which report the second best performance 94.8 at rank-10 with automatically detected

| Methods | VIPeR | | | PRID450S | | | CUHK03 (labeled) | | | CUHK03 (detected) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | r=1 | r=10 | r=20 | r=1 | r=10 | r=20 | r=1 | r=5 | r=10 | r=1 | r=5 | r=10 |
| LOMO | 40.00 | 80.51 | 91.08 | 61.38 | 91.02 | 95.33 | 50.85 | 81.38 | 91.14 | 44.45 | 78.70 | 87.65 |
| LOMO++SSM | 42.22 | 83.54 | 92.82 | 62.84 | 92.62 | 96.49 | 52.50 | 84.53 | 92.49 | 49.05 | 81.25 | 90.30 |
| GOG | 49.72 | 88.67 | 94.53 | 68.00 | 94.36 | 97.64 | 68.47 | 90.69 | 95.84 | 64.10 | 88.40 | 94.30 |
| GOG+SSM | 50.73 | 89.97 | 95.63 | 68.49 | 95.73 | 98.53 | 71.82 | 92.54 | 96.64 | 68.20 | 90.30 | 94.10 |
| Fusion | 53.26 | 90.95 | 95.73 | 72.04 | 95.82 | 98.49 | 71.87 | 92.64 | 96.80 | 68.05 | 90.15 | 94.95 |
| Fusion+SSM | **53.73** | **91.49** | **96.08** | **72.98** | **96.76** | **99.11** | **76.63** | **94.59** | **97.95** | **72.70** | **92.40** | **96.05** |

Table 4. The comparison with baselines on VIPeR, PRID450S and CUHK03 dataset.

bounding boxes. Nevertheless, the performance gap with SSM becomes larger at lower ranks. For instance, SSM makes a significant improvement of 18.0 in rank-1 accuracy over Null [57] with detected bounding boxes.

GOG remains to be one of the most robust descriptors on this dataset. Under XQDA metric, it achieves the second best performances at most ranks. As analyzed above, SSM can be deemed as a generic tool for those visual descriptors and metric learning techniques. Thus, SSM can further enhance their discriminative power.

### 5.3. Market-1501

Market-1501 [63] is the largest benchmark in person re-identification up to present, which is comprised of 1501 identities. 750 identities ($12,936$ images) are used for training and 751 identities ($19,732$ images) are used for testing. $3,368$ images are taken as the probe. Both CMC scores and mean average precision (mAP) are used for evaluation.

Thanks to plenty of training images provided, training deep neural networks becomes feasible on this dataset and preferred by most previous works [47, 50, 49]. Following this trend, we introduce Residual Network (ResNet) [16] to person re-identification. More specifically, we fine-tune a 50-layer ResNet with classification loss on training images, and extract activations of its last fully connected layer. The $L_2$ normalized activations are taken as visual features and Euclidean metric is utilized to measure the distances between images. The baseline performances are mAP 61.12 with single query (SQ) and 70.82 with multiple query (MQ), respectively.

In Table 7, we present the experimental comparisons. As can be seen, SSM improves the baseline by mAP 7.68 for SQ and 5.30 for MQ. Moreover, SSM outperforms the previous state-of-the-art [49] by a very large margin, with the improvement of mAP 29.25 for SQ and 27.73 for MQ.

### 5.4. Time Analysis

As an additional improvement over metric learning, SSM introduces extra time cost without doubt as analyzed in Sec. 4.3. In Table 8, we present the extra execution time of SSM over XQDA metric. As SSM manages transferring the graph-based affinity learning to off-line, the off-line cost is increased especially on larger datasets (*e.g.*, CUHK03 and Market-1501). In on-line stage, the extra indexing time brought by SSM only occupies a small percentage on all the datasets except CUHK03. On CUHK03 dataset, indexing using XQDA metric only requires $0.09s$, since CUHK03 has a small gallery. As SSM takes into account the larger training data provided by CUHK03, the extra indexing cost is $0.516s$. Nevertheless, the overall indexing time is still within 1 second.

| Methods | Single Query | | Multiple Query | |
|---|---|---|---|---|
| | Rank-1 | mAP | Rank-1 | mAP |
| SSDAL [47] | 39.40 | 19.60 | 49.00 | 25.80 |
| WARCA [20] | 45.16 | - | - | - |
| SCSP [9] | 51.90 | 26.35 | - | - |
| S-LSTM [50] | - | - | 61.60 | 35.31 |
| Null [57] | 61.02 | 35.68 | 71.56 | 46.03 |
| S-CNN [49] | **65.88** | **39.55** | **76.04** | **48.45** |
| SSM (Ours) | **82.21** | **68.80** | **88.18** | **76.18** |

Table 7. The comparison with state-of-the-art on Market-1501.

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

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

## 6. Conclusion

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 furture, we will investigate how to effectively fuse multiple features [38] and apply SSM to other datasets [53].

**Acknowledgements.** This work was supported in part by NSFC 61573160 and 61429201, in part to Dr. Qi Tian by ARO grants W911NF-15-1-0290 and Faculty Research Gift Awards by NEC Laboratories of America and Blippar.

## References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *CVPR*, pages 3908–3916, 2015. 7

[2] D. Aldous and J. A. Fill. Reversible markov chains and random walks on graphs, 2002. 2

[3] S. Ali, O. Javed, N. Haering, and T. Kanade. Interactive retrieval of targets for wide area surveillance. In *ACM international conference on Multimedia*, pages 895–898, 2010. 2

[4] L. An, M. Kafai, S. Yang, and B. Bhanu. Person reidentification with reference descriptor. *TCSVT*, 26(4):776–787, 2016. 2

[5] S. Bai and X. Bai. Sparse contextual activation for efficient visual re-ranking. *TIP*, 25(3):1056–1069, 2016. 1

[6] S. Bai, X. Bai, Q. Tian, and L. J. Latecki. Regularized diffusion process for visual retrieval. In *AAAI*, 2017. 1

[7] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu. Learning context-sensitive shape similarity by graph transduction. *TPAMI*, 32(5):861–874, 2010. 1

[8] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003. 4

[9] D. Chen, Z. Yuan, B. Chen, and N. Zheng. Similarity learning with spatial constraints for person re-identification. In *CVPR*, pages 1268–1277, 2016. 2, 6, 7, 8

[10] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang. Similarity learning on an explicit polynomial kernel feature map for person re-identification. In *CVPR*, pages 1565–1573, 2015. 2, 6, 7

[11] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *CVPR*, pages 1335–1344, 2016. 7

[12] M. Donoser and H. Bischof. Diffusion processes for retrieval revisited. In *CVPR*, pages 1320–1327, 2013. 1

[13] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, pages 2360–2367, 2010. 1

[14] J. Garcia, N. Martinel, C. Micheloni, and A. Gardel. Person re-identification ranking optimisation by discriminant context information analysis. In *ICCV*, pages 1305–1313, 2015. 2, 7

[15] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, pages 262–275, 2008. 5, 7

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 8

[17] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, pages 91–102, 2011. 2

[18] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*, pages 780–793, 2012. 1

[19] S. Huang, J. Lu, J. Zhou, and A. K. Jain. Nonlinear local metric learning for person re-identification. *arXiv:1511.05169*, 2015. 6

[20] C. Jose and F. Fleuret. Scalable metric learning via weighted approximate rank component analysis. In *ECCV*, pages 875–890, 2016. 8

[21] E. Kodirov, T. Xiang, Z. Fu, and S. Gong. Person re-identification by unsupervised $\ell 1$ graph learning. In *ECCV*, pages 178–195, 2016. 7

[22] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, pages 2288–2295, 2012. 1, 2

[23] Q. Leng, R. Hu, C. Liang, Y. Wang, and J. Chen. Person re-identification with content and context re-ranking. *Multimedia Tools and Applications*, 74(17):6989–7014, 2015. 2

[24] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, pages 152–159, 2014. 1, 5, 7

[25] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, pages 3610–3617, 2013. 1

[26] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, pages 2197–2206, 2015. 1, 2, 5, 6, 7

[27] S. Liao and S. Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *ICCV*, pages 3685–3693, 2015. 6, 7

[28] G. Lisanti, I. Masi, A. D. Bagdanov, and A. Del Bimbo. Person re-identification by iterative re-weighted sparse ranking. *TPAMI*, 37(8):1629–1642, 2015. 1

[29] C. Liu, C. Change Loy, S. Gong, and G. Wang. Pop: Person re-identification post-rank optimisation. In *ICCV*, pages 441–448, 2013. 2

[30] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person re-identification: What features are important? In *ECCV*, pages 391–401, 2012. 5

[31] C. C. Loy, C. Liu, and S. Gong. Person re-identification by manifold ranking. In *ICIP*, pages 3567–3571, 2013. 2, 5, 6

[32] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *CVPR*, pages 1988–1995, 2009. 5

[33] C. C. Loy, T. Xiang, and S. Gong. Time-delayed correlation analysis for multi-camera activity understanding. *IJCV*, 90(1):106–129, 2010. 2, 5

[34] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *TIP*, 23(8):3656–3670, 2014. 6

[35] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury. Temporal model adaptation for person re-identification. In *ECCV*, pages 858–877, 2016. 7

[36] N. Martinel, C. Micheloni, and G. L. Foresti. Kernelized saliency-based person re-identification through multiple metric learning. *TIP*, 24(12):5645–5658, 2015. 6

[37] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato. Hierarchical gaussian descriptor for person re-identification. In *CVPR*, pages 1363–1372, 2016. 1, 5, 6, 7

[38] S. Paisitkriangkrai, C. Shen, and A. van den Hengel. Learning to rank in person re-identification with metric ensembles. In *CVPR*, pages 1846–1855, 2015. 7, 8

[39] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*, pages 3318–3325, 2013. 7

[40] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person re-identification by support vector ranking. In *BMVC*, page 6, 2010. 6

[41] P. M. Roth, M. Hirzer, M. Koestinger, C. Beleznai, and H. Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification*, pages 247–267. 2014. 5, 7

[42] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000. 1

[43] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang. Person re-identification with correspondence structure learning. In *ICCV*, pages 3200–3208, 2015. 7

[44] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, and S. Z. Li. Embedding deep metric for person re-identification: A study against large variations. In *ECCV*, pages 732–748, 2016. 2, 7

[45] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *CVPR*, pages 4184–4193, 2015. 7

[46] C. Su, F. Yang, S. Zhang, Q. Tian, L. S. Davis, and W. Gao. Multi-task learning with low rank attribute embedding for person re-identification. In *ICCV*, pages 3739–3747, 2015. 7

[47] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian. Deep attributes driven multi-camera person re-identification. In *ECCV*, pages 475–491, 2016. 6, 7, 8

[48] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, and Y. Y. Tang. Person re-identification by dual-regularized kiss metric learning. *TIP*, 25(6):2726–2738, 2016. 6

[49] R. R. Varior, M. Haloi, and G. Wang. Gated siamese convolutional neural network architecture for human re-identification. In *ECCV*, pages 791–808, 2016. 7, 8

[50] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang. A siamese long short-term memory architecture for human re-identification. In *ECCV*, pages 135–153, 2016. 7, 8

[51] H. Wang, S. Gong, X. Zhu, and T. Xiang. Human-in-the-loop person re-identification. In *ECCV*, pages 405–422, 2016. 2

[52] T. Xiao, H. Li, W. Ouyang, and X. Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *CVPR*, pages 1249–1258, 2016. 7

[53] T. Xiao, S. Li, , B. Wang, L. Lin, and X. Wang. Joint detection and identification feature learning for person search. In *CVPR*, pages 1249–1258, 2017. 8

[54] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*, pages 1–16, 2014. 1

[55] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *ECCV*, pages 536–551, 2014. 1, 7

[56] L. Zelnik-manor and P. Perona. Self-tuning spectral clustering. In *NIPS*, pages 1601–1608, 2005. 5

[57] L. Zhang, T. Xiang, and S. Gong. Learning a discriminative null space for person re-identification. In *CVPR*, 2016. 7, 8

[58] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas. Query specific rank fusion for image retrieval. *TPAMI*, 37(4):803–815, 2015. 1

[59] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan. Sample-specific svm learning for person re-identification. In *CVPR*, 2016. 6, 7

[60] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, pages 2528–2535, 2013. 7

[61] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, pages 3586–3593, 2013. 7

[62] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, pages 144–151, 2014. 7

[63] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, pages 1116–1124, 2015. 1, 5, 8

[64] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian. Query-adaptive late fusion for image search and person re-identification. In *ICCV*, pages 1741–1750, 2015. 1, 7

[65] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *TPAMI*, 35(3):653–668, 2013. 1, 6

[66] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS*, pages 321–328, 2003. 1

[67] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. In *NIPS*, pages 169–176, 2004. 1, 2

[68] X. Zhu and Z. Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical report, Citeseer, 2002. 1