Efficient Online Local Metric Adaptation via Negative Samples for Person Re-Identification

Jiahuan Zhou, Pei Yu, Wei Tang and Ying Wu
Electrical Engineering and Computer Science, Northwestern University, US
{jzt011, pyi980, wtt450, yingwu}@eecs.northwestern.edu

Abstract

Many existing person re-identification (PRID) methods typically attempt to train a faithful global metric offline to cover the enormous visual appearance variations, so as to directly use it online on various probes for identity matching. However, their need for a huge set of positive training pairs is very demanding in practice. In contrast to these methods, this paper advocates a different paradigm: part of the learning can be performed online but with nominal costs, so as to achieve online metric adaptation for different input probes. A major challenge here is that no positive training pairs are available for the probe anymore. By only exploiting easily-available negative samples, we propose a novel solution to achieve local metric adaptation effectively and efficiently. For each probe at the test time, it learns a strictly positive semi-definite dedicated local metric. Comparing to offline global metric learning, its computational cost is negligible. The insight of this new method is that the local hard negative samples can actually provide tight constraints to fine tune the metric locally. This new local metric adaptation method is generally applicable, as it can be used on top of any global metric to enhance its performance. In addition, this paper gives in-depth theoretical analysis and justification of the new method. We prove that our new method guarantees the reduction of the classification error asymptotically, and prove that it actually learns the optimal local metric to best approximate the asymptotic case by a finite number of training data. Extensive experiments and comparative studies on almost all major benchmarks (VIPeR, QMUL GRID, CUHK Campus, CUHK03 and Market-1501) have confirmed the effectiveness and superiority of our method.

1. Introduction

Person re-identification (PRID) generally refers to evaluating the similarity of a probe image from an unknown person against a set of gallery images with known identities. The gallery images may be obtained from different cameras at a different time. This has been a critical yet very challenging task in video surveillance [25]. The difficulties are mainly due to the large and complex variations in the visual appearances of a person under various views, poses, illumination and occlusion conditions.

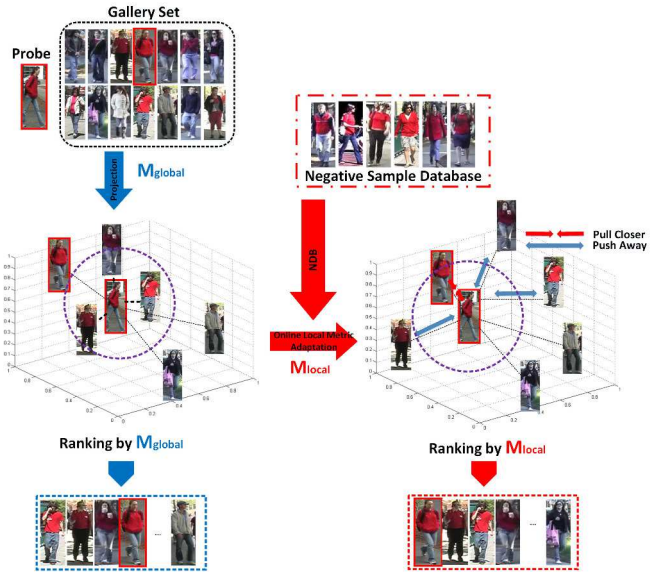


Figure 1. The overall idea of our proposed online local metric adaptation approach. Unlike existing methods that learn a single global metric for all probes, we exploit negative samples to learn a dedicated local metric for each online probe.

Recent attempts were based on learning a visual metric to better capture the visual similarities [15, 12, 14, 20, 35, 36], and reported encouraging results. The training data for such metric learning are generally sample pairs: a *positive* pair refers to two images of the same identity, and a *negative* pair otherwise. These methods typically attempt to train a faithful global metric offline, hoping to cover the enormous visual appearance variations so as to directly use it online for all test probes. Thus they demand a huge set of positive training pairs. Unfortunately, in practice, although it is relatively easy to collect negative pairs, it is in general difficult to obtain many positive pairs for a specific person. There-

fore, the metrics learned from insufficient positive training data are likely to be biased. In addition, it is computationally intensive to learn a strictly positive semi-definite (PSD) global metric, while ignoring the PSD constraint leads to unstable and noisy metrics [15].

In contrast to these methods, this paper advocates a different paradigm: shifting part of the metric learning to online local metric adaptation. Specifically, for each online probe at the test time, our new approach learns a dedicated local metric with a nominal computational cost. Combining a global metric with local metric adaptation achieves an adaptive nonlinear metric. In our approach, its online learning is special, because there are no positive training pairs available at all for the probe, as its identity is unknown.

An attractive property of our new method is that it only uses negative data from a negative sample database (NDB). We call it **OL-MANS** for short of online local metric adaptation via negative samples. For a given test probe, a subset of samples from NDB are selected to form informative negative pairs with this test probe. These selected samples from NDB are visually similar to the probe, but are guaranteed to have different identities from the probe (at least with a very large probability). These negative samples provide effective local discrimination for further constraining the local metric tuning, by pushing away local false positives, as illustrated in Fig. 1. For each probe, our new method learns a strictly positive semi-definite local metric efficiently, via solving a kernel SVM problem. Comparing to offline global metric learning, the computational cost of the proposed online learning is negligible. Moreover, our method is generally applicable, and can be used on top of any global metric.

A significant property of our new method is that it is justified and backed up with a theoretical guarantee to improve the performance of the global metric. This paper gives in-depth theoretical analysis to well justify the proposed method. We first prove that this new method guarantees the reduction of classification error asymptotically when there are an infinite number of training data. Then we pursue the best approximation of the asymptotic case by using a finite number of training data. We prove that the learning objective of the proposed local metric adaptation is equivalent to the optimal approximation of the asymptotic case. In addition, we also provide consistency and sample complexity analysis. This indicates that the learned local metric is bound to improve the PRID performances. These properties have been confirmed to be very effective and practical by our extensive experiments and comparative studies on several PRID benchmarks, including VIPeR, QMUL GRID, CUHK Campus, CUHK03 and Market-1501.

2. Related Work

Existing metric learning-based PRID methods either learn a single global metric or a local discrimination. Zheng

Figure 2. The improvement of ranking by our OL-MANS on VIPeR [7]. BLUE boxes: input probes, RED: gallery targets. For each case, top row is the ranking result from the baseline [15], and bottom row is our ranking result. (Best view in color and enlarged)

et al. [35] proposed a relative distance comparison (PRDC) method to maximize the probability of a positive pair to have a smaller distance than a negative pair. Hirzer *et al.* [8] relaxed the PSD constraint to simplify the computation. Liao *et al.* [14] learned a discriminant subspace and a global distance metric simultaneously for dimension reduction and optimal dimensionality. A logistic metric learning called MLAPG was proposed by Liao *et al.* [15] for a global PSD metric via an asymmetric sample weighting strategy.

There were methods based on local strategies. Zhang *et al.* [29] formulated the PRID problem as a local distance comparison problem to handle the multi-modal distributions of the visual appearances. Li *et al.* [12] proposed the Locally-Adaptive Decision Functions (LADF) which integrates a traditional distance metric with a local decision rule. Pedagadi *et al.* [22] employed the Local Fisher Discriminant Analysis (LFDA) which combines the fisher discriminant analysis (FDA) and Local Preserving Projections (LPP) to exploit the local geometrical information of samples. Liong *et al.* [16] developed a regularized local metric learning (RLML) method to combine global and local metrics, so as to utilize the local data distribution to alleviate over-fitting. Zhang *et al.* [31] proposed LSSCDL to learn a specific SVM classifier for each training sample, then the weight parameters of a new sample can be inferred. A novel multi-task maximally collapsing metric learning (MtMCML) model was proposed by Ma *et al.* [20].

In contrast to methods learning a global metric, our proposed method is mainly focused on learning local metrics specifically adaptive to individual test probes. Different from RLML that requires clustering in advance to obtain the local data distributions, our new approach does not need clustering but is rather instance-based learning, and thus avoiding the risk of inaccurate clustering results. Also note that MtMCML learning still follows the global manner although it learns different metrics for different cameras. In contrast to LADF that needs a large number of positive sample pairs to drive the local decision function learning, our new approach only uses negative sample pairs which are much easier to obtain. LSSCDL also requires a lot of positive training pairs for offline learning, but ours performs online learning per probe without the requirement of posi-

tive pairs.

3. Online Local Metric Adaptation via Negative Samples (OL-MANS)

3.1. Problem Setup

A single-shot PRID dataset consists of n pairs of identity images $\{(x_i^p, x_i^g)\}_{i=1}^n$ collected from two different disjoint cameras: x_i^p is from the probe camera and x_i^g is from the gallery camera. The index $i = \{1, 2, ..., n\}$ represents the identity label of n different persons. For training and testing in PRID, all identity pairs can be divided into two disjoint subsets $\{u_1, u_2, ..., u_{m'}\}$ and $\{v_1, v_2, ..., v_m\}$ where $n = m + m'$ and

$$\begin{aligned} \mathbf{X}_{train} &= \mathbf{X}_{train}^p \cup \mathbf{X}_{train}^g = \left\{x_{u_i}^p\right\}_{i=1}^{m'} \cup \left\{x_{u_i}^g\right\}_{i=1}^{m'} \\ \mathbf{X}_{test} &= \mathbf{X}_{test}^p \cup \mathbf{X}_{test}^g = \left\{x_{v_i}^p\right\}_{i=1}^{m} \cup \left\{x_{v_i}^g\right\}_{i=1}^{m} \end{aligned} \quad (1)$$

So that $\mathbf{X}_{train}$ is used as the training set and $\mathbf{X}_{test}$ is the test set. In our algorithm, an additional negative sample database, denoted by $\mathbf{Y}^{neg} = \{y_i\}_{i=1}^k$, is needed, and will be discussed shortly in Sec. 3.3.

3.2. Conventional Global Metric Learning

Conventional learning-based PRID methods [14, 35, 26, 15] aim to learn a single global Mahalanobis distance metric $\mathbf{M}_G$ by using the training set $\mathbf{X}_{train}$. The learned metric $\mathbf{M}_G$ projects the original samples into another feature space, and the matching between one probe x_i^p and one gallery image x_j^g at test stage is measured by:

$$d_{\mathbf{M}_G}\left(x_i^p, x_j^g\right) = \left\|x_i^p - x_j^g\right\|_{\mathbf{M}_G}^2 = \left(x_i^p - x_j^g\right)^T \mathbf{M}_G \left(x_i^p - x_j^g\right) \quad (2)$$

where $\mathbf{M}_G = \mathbf{W}^T\mathbf{W} \succeq 0$ needs to be positive semi-definite, as $\mathbf{W}$ is the learned projection. Different methods adopt different loss functions to learn $\mathbf{M}_G$, and a good solution to $\mathbf{M}_G$ should align the similarity structure in the projected feature space, so as to pull the samples from the same identity group closer and to make different identities more discriminative. Due to the fact that the global metric does not aim to fit the local distributions for all the samples specifically, it may lead to large biases and distortions in some places in the feature space. As illustrated in Fig. 1, our new approach puts an instance-based online local metric adaptation on top of the global metric.

3.3. Instance-based OL-MANS

In this paper, we propose an online local metric adaptation algorithm called *OL-MANS* to adaptively adjust the metric dedicated to specific test probes with minimum online training by utilizing only negative training samples.

Specifically, for a probe image $x_{v_i}^p$ in the probe set $\mathbf{X}_{test}^p$, we aim to learn a local Mahalanobis distance $\mathbf{M}_L^i$ only using

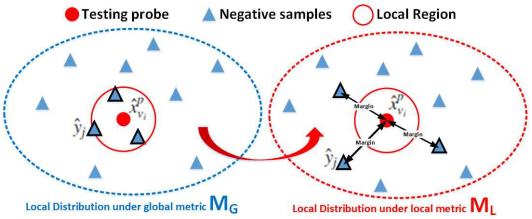


Figure 3. The local metric $\mathbf{M}_L^i$ for $\hat{x}_{v_i}^p$ can push the closest negative sample $\hat{y}_j$ of $\hat{x}_{v_i}^p$ away from the local region $\Omega(\hat{x}_{v_i}^p)$

the samples in a negative sample database $\mathbf{Y}^{neg}$ as training data. This negative sample database provides rather faithful negative samples to the tests with a large probability. There are many ways to collect $\mathbf{Y}^{neg}$, e.g., data from a different benchmark can be used, or false positive matches from images that do not contain humans. The insight here is that all such negative samples are "hard negatives" for the probes. In this research, we have investigated how $\mathbf{Y}^{neg}$ influences the performance. This study is presented as the supplementary material due to the page limit.

As the global projection $\mathbf{W}$ learned by the global metric learning maps $\mathbf{X}_{test}^p$ to a low dimensional subspace $\hat{\mathbf{X}}_{test}^p = \mathbf{W}\mathbf{X}_{test}^p = \left\{\hat{x}_{v_i}^p\right\}_{i=1}^m$, we propose to further adjust the local similarity for each specific $\hat{x}_{v_i}^p$ by an online learned local metric $\mathbf{M}_L^i$ which is solely learned from $\mathbf{Y}^{neg}$.

We propose to pursue an optimal PSD Mahalanobis metric $\mathbf{M}_L^i$ for the local adaptation, by maximizing the distance to the closest (or "hardest" conceptually) negative sample of $\hat{x}_{v_i}^p$, as shown in Fig. 3:

$$\mathbf{M}_L^i = \arg\max_{\mathbf{M}_L^i \succeq 0} \left(\min_{1 \le j \le k} \left(\hat{x}_{v_i}^p - \hat{y}_j\right)^T \mathbf{M}_L^i \left(\hat{x}_{v_i}^p - \hat{y}_j\right) \right) \quad (3)$$

where $\hat{y}_j = \mathbf{W}y_j$ is the projected negative sample based on the global metric. We regularize $\mathbf{M}_L^i$ for a stable solution. This can be done via minimizing the norm under a fixed margin constraint, instead of maximizing the margin under a fixed norm constraint [6], so the alternative objective is:

$$\begin{aligned} &\mathbf{M}_L^i = \arg\min_{\mathbf{M}_L^i} \frac{1}{2}\left\|\mathbf{M}_L^i\right\|^2 \\ &sub\ to: \ \left(\hat{x}_{v_i}^p - \hat{y}_j\right)^T \mathbf{M}_L^i \left(\hat{x}_{v_i}^p - \hat{y}_j\right) \ge 2, \ \ \forall 1 \le j \le k \\ &\qquad\qquad \mathbf{M}_L^i \succeq 0 \end{aligned} \quad (4)$$

where the constant 2 is arbitrary only for manipulation convenience. While this is a convex semi-definite programming problem, it can be very slow for high dimensional data, even for the state-of-the-art PSD solvers.

In the proposed OL-MANS approach, we relax the PSD constraint requiring $\mathbf{M}_L^i \succeq 0$, but we prove below that the relaxed objective is equivalent to a kernel SVM problem with a quadratic kernel. And thus the solution is still a PSD metric. In addition, it can be readily solved with off-the-

shelf SVM solvers such as LIBSVM [2]. More importantly, we also prove that this learning objective is equivalent to the best approximation to the asymptotic classification error, which is proved to be lower than the global metric (details see Sec. 4 and supplementary materials).

Theorem 1 *The solution to Eqn. 4 is equivalent to a kernel SVM with* $k(x,y) = \langle x,y\rangle^2$ *on* $\{\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_k\}$ *where* $\tilde{y}_j = \hat{x}^p_{v_i} - \hat{y}_j$ *(for* $j \geq 1$*), and* $\tilde{y}_0 = \hat{x}^p_{v_i} - \hat{x}^p_{v_i} = 0$.

Proof 1 *Define auxiliary labels by:*

$$\zeta_j = \begin{cases} -1, & j = 0 \\ 1, & j \neq 0 \end{cases} \tag{5}$$

so the objective Eqn. 4 can be rewritten as:

$$\begin{aligned} &\boldsymbol{M}^i_L = \arg\min_{\boldsymbol{M}^i_L} \frac{1}{2}\left\|\boldsymbol{M}^i_L\right\|^2 \\ &sub\ to: \ \ \zeta_j\left(\tilde{y}_j^T \boldsymbol{M}^i_L \tilde{y}_j - 1\right) \geq 1, \forall\ 0 \leq j \leq k \end{aligned} \tag{6}$$

Eqn. 6 is exactly an SVM problem with quadratic kernel and with bias fixed to one. Next we prove the solution to objective Eqn. 6 is exactly the same as that to the original objective Eqn. 4. Consider the dual of the SVM, the optimal solution $\boldsymbol{M}^i_L$ *has the form:*

$$\boldsymbol{M}^i_L = \sum_{j=0}^{k} \alpha_j \zeta_j \tilde{y}_j \tilde{y}_j^T, \ \ \alpha_j \geq 0 \tag{7}$$

Since $\tilde{y}_j\tilde{y}_j^T$ *is PSD for* $j \geq 1$ *(* $\tilde{y}_0\tilde{y}_0^T = 0$ *) and* $\zeta_j = 1$ *for* $j \geq 1$*, so we have:*

$$\boldsymbol{M}^i_L = \sum_{j=0}^{k} \alpha_j \zeta_j \tilde{y}_j \tilde{y}_j^T = \sum_{j=1}^{k} \alpha_j \tilde{y}_j \tilde{y}_j^T \succeq 0 \tag{8}$$

It is obvious that the positive semi-definiteness of $\mathbf{M}^i_L$ is guaranteed even if no PSD constraint is explicitly imposed in our learning objective Eqn. 6.

3.4. Person Re-identification via OL-MANS

At the online test stage, for a probe $x^p_{v_i}$ from $\mathcal{X}^p_{test}$ and one gallery image $x^g_{v_j}$ from $\mathcal{X}^g_{test}$, our method combines a global metric $\mathbf{M}_G$ (with flexible choices) with our local metric adaptation $\mathbf{M}^i_L$ to achieve an adaptive nonlinear metric:

$$\begin{aligned} &d(x^p_{v_i}, x^g_{v_j}) = d_{\mathbf{M}_G}(x^p_{v_i}, x^g_{v_j}) + \lambda d_{\mathbf{M}_L}(x^p_{v_i}, x^g_{v_j}) \\ &= (x^p_{v_i} - x^g_{v_j})^T \mathbf{W}^T (\mathbf{I} + \lambda \mathbf{M}^i_L) \mathbf{W} (x^p_{v_i} - x^g_{v_j}) \end{aligned} \tag{9}$$

where $\mathbf{M}_G = \mathbf{W}^T\mathbf{W}$ is an learned global metric and $\mathbf{M}^i_L$ is the local metric adaptation specific for $x^p_{v_i}$. λ is the weighting parameter which can be decided by cross-validation. In this paper, we set λ by Eqn. 10 in all the experiments (details please see the supplementary materials).

$$\lambda = \max_{1\leq j\leq m'} \left(d_{\mathbf{M}_G}(x^p_{v_i}, y^g_{v_j})\right) / \max_{1\leq j\leq m'} \left(d_{\mathbf{M}_L}(x^p_{v_i}, y^g_{v_j})\right) \tag{10}$$

4. Theoretical Analysis and Justification

We first prove that the asymptotic error by using the proposed OL-MANS is bound to be lower than that without. When the negative samples are truly hard negative ones, the asymptotic error by using OL-MANS can be very close to the Bayesian error (Sec. 4.1). Besides this theoretically meaningful result, we prove that this strong asymptotic error can actually best approximated by using finite data, which is practically also meaningful. More importantly, we prove that this approximation is actually achieved by OL-MANS (Sec. 4.2). We also present its consistency and sample complexity analysis in Sec. 4.3.

4.1. Asymptotic Error is Reduced

The core of PRID is indeed a two-class (ω_+ and ω_-) 1-Nearest neighbor (NN) classification problem by using the gallery set $\mathcal{D}$. If there is infinite number of data, it is well-known that its asymptotic error $\mathcal{P}(e|x)$ is bounded by 2 times the Bayesian error [4]:

$$\mathcal{P}^* \leq \mathcal{P}(e|x) = 2P(\omega_+|x)P(\omega_-|x) \leq 2\mathcal{P}^* \tag{11}$$

where $\mathcal{P}^*$ is the Bayesian error. In our work, we prove that by adding the hard negative samples x_a to $\mathcal{D}$ to form an augmented dataset $\mathcal{D}^a$, the asymptotic error $\mathcal{P}^a(e|x)$ by using $\mathcal{D}^a$ is always smaller than $\mathcal{P}(e|x)$:

$$\mathcal{P}^a(e|x) \leq \mathcal{P}(e|x) \tag{12}$$

Theorem 2 *For an input* x*, its NN is* x' *in* $\mathcal{D}^a$*. Define the probability that* x' *is an augmented data* x_a*, i.e.,* $x' \sim x_a$ *as* $P(x' \sim x_a) = q$*; otherwise,* x' *is not an augmented data* x_a*, i.e.,* $x' \neg x_a$*,* $P(x' \neg x_a) = 1 - q$*, where* $0 \leq q \leq 1$*. The asymptotic error* $\mathcal{P}^a(e|x)$ *by using* $\mathcal{D}^a$ *is:*

$$\mathcal{P}^a(e|x) = \frac{(2-q)\mathcal{P}(e|x)}{2 - 2q\mathcal{P}(e|x)} \leq \mathcal{P}(e|x) \tag{13}$$

Proof 2 *Due to the page limit, please see the proof in the supplementary materials.*

Since q is the probability of $P(x' \sim x_a)$, $0 \leq q \leq 1$. If $q = 0$ which indicates that the augmented negative data are useless, then we have $P^a(e|x) = P(e|x)$. Another extreme is when $q = 1$ implying the negative data are abundant and effective to constrain the classification, then we have [1]

$$\mathcal{P}^a(e|x) = \frac{\mathcal{P}(e|x)}{2[1 - \mathcal{P}(e|x)]} \leq \mathcal{P}(e|x) \tag{14}$$

In this case, when $\mathcal{P}(e|x)$ is very small, we have

$$\mathcal{P}^a(e|x) \simeq \frac{\mathcal{P}(e|x)}{2} \simeq \mathcal{P}^*(e) \tag{15}$$

The asymptotic error of our negative-augmented approach can be very close to the Bayesian error.

[1] $\mathcal{P}(e|x) \leq \frac{1}{2}$ is always true.

4.2. Finite Approximation to $\mathcal{P}^a(e|x)$

The asymptotic error $\mathcal{P}^a(e|x)$ in Eqn. 13 is only meaningful when the sample size is infinite, $n \to \infty$. However, in practice, only finite number of samples are available. To make it practically meaningful, we prove that it can be best approximated by the practical error rate $\mathcal{P}_n(e|x)$ (n is finite) by finding a local metric $\mathbf{M}_L$. And this local metric turns out to be the one for the proposed OL-MANS.

Still consider the 2-class 1-NN rule scenario (on the negative-augmented data $\mathcal{D}^a$). To make the notation less cluttered, here we use $\mathcal{P}(e|x)$ to indicate $\mathcal{P}^a(e|x)$ without confusion. Given a sample x and its nearest neighbor x' from the finite dataset containing n samples. The probability of error for x is:

$$\begin{aligned}\mathcal{P}_n(e|x) &= P(\omega_+|x)P(\omega_-|x') + P(\omega_-|x)P(\omega_+|x') \\ &= \mathcal{P}(e|x) + [P(\omega_+|x) - P(\omega_-|x)][P(\omega_+|x) - P(\omega_+|x')]\end{aligned}$$

Our goal is to find a best local metric $\mathbf{M}_x$ for x such that the conditional MSE $\min_{\mathbf{M}_x} \mathbb{E}\{[\mathcal{P}_n(e|x) - \mathcal{P}(e|x)]^2|x\}$ is minimized. Since $[P(\omega_+|x) - P(\omega_-|x)]$ is constant for a given x, so the minimization is equal to:

$$\min_{\mathbf{M}_x} \mathbb{E}\{[P(\omega_+|x) - P(\omega_+|x')]^2|x\} \tag{16}$$

Because $P(\omega_+|x') \simeq P(\omega_+|x) + \nabla P(\omega_+|x)^T(x' - x)$, Eqn. 16 is approximately equivalent to:

$$\min_{\mathbf{M}_x} \mathbb{E}\{\|\nabla P(\omega_+|x)^T(x' - x)\|^2|x\} \tag{17}$$

The core here is to compute the gradient of posterior $\nabla P(\omega_+|x)$. Recall our proposed OL-MANS approach, a local linear classifier $\mathbf{w}$ where $\mathbf{M}_x = \mathbf{w}\mathbf{w}^T$ is learned for sample x via a standard kernel SVM framework. So the posterior of x in a logistic sigmoid function form is:

$$P(\omega_+|x) = \frac{1}{1 + e^{\zeta_x(\mathbf{w}^T x + b) - \gamma}}, P(\omega_-|x) = 1 - P(\omega_+|x) \tag{18}$$

The gradient of $P(\omega_+|x)$ can be easily computed:

$$\nabla P(\omega_+|x) = \zeta_x P(\omega_+|x)P(\omega_-|x)\mathbf{w} \tag{19}$$

Substituting Eqn. 19 for $\nabla P(\omega_+|x)$ in Eqn. 17 gives us:

$$\begin{aligned}&\min_{\mathbf{M}_x} \mathbb{E}\{\|\zeta_x P(\omega_+|x)P(\omega_-|x)\mathbf{w}^T(x' - x)\|^2|x\} \\ &= \min_{\mathbf{M}_x} (x' - x)^T \mathbf{w}\mathbf{w}^T(x' - x)\end{aligned} \tag{20}$$

Recall our optimization objective Eqn. 6, for the positive samples, we have $1 - (x' - x)^T\mathbf{M}_x(x' - x) \geq 1$ which is equal to $(x' - x)^T\mathbf{M}_x(x' - x) \leq 0$. On the other hand, $(x - x')^T\mathbf{M}_x(x - x') \geq 0$ is always true for a PSD $\mathbf{M}_x$, so $(x' - x)^T\mathbf{M}_x(x' - x) \equiv 0$ always holds. It is obvious Eqn. 20 is always optimized by adopting the local metric $\mathbf{M}_x$ learned by our algorithm Eqn. 6.

4.3. Consistency and Sample Complexity Analysis

A set of samples $\{x_0, x_1, ..., x_k\}$ is identically drawn from a D-dimensional space $\mathbb{D} \in \mathbb{R}^D$ where l_i is the label of x_i, then a paired sample set $S_k^{pair} = \{s_i\}_{i=1}^k = \{(x_0, x_i)\}_{i=1}^k$ of size k is formed. For our proposed learning objective Eqn. 6, the true risk over the whole distribution $\mathbb{D}$ and the empirical error based on S_k^{pair} are defined as:

$$\begin{aligned}Err^\lambda(\mathbf{M}_L, \mathbb{D}) &= \mathbb{E}_{x_i, x_j \sim \mathbb{D}} \phi^\lambda(\mathbf{M}_L, (x_i, x_j)) \\ Err^\lambda(\mathbf{M}_L, S_k^{pair}) &= \frac{1}{k}\sum_{i=1}^k \phi^\lambda(\mathbf{M}_L, s_i)\end{aligned}$$

where $\phi^\lambda(\mathbf{M}_L, s_i)$ is the hinge loss function:

$$\phi^\lambda(\mathbf{M}_L, s_i) = \lambda[\zeta_i\left((x_i - x_0)^T\mathbf{M}_L(x_i - x_0)\right) - \gamma_{\zeta_i}]_+$$

where $\zeta_i = -1$ if $l_i = l_0$ and 1 otherwise, $[A]_+ = \max(0, A)$ is the hinge loss and γ_{ζ_i} is the desired margin between samples. The empirical risk minimizing metric based on S_k^{pair} can be readily defined as $\mathbf{M}_L^* = \arg\min_{\mathbf{M}_L} Err^\lambda(\mathbf{M}_L, S_k^{pair})$. Our goal is to compare the generalization performance of $\mathbf{M}_L^*$ over the unknown $\mathbb{D}$.

Theorem 3 *Let $\phi^\lambda(\boldsymbol{M}_L, s_i)$ be a distance-based loss function that is λ-Lipschitz in the first argument. Then with probability at least $1-\delta$ over $\{s_1, ..., s_k\}$ from an unknown B-bounded-support (each $(x, l) \sim \mathbb{D}, ||x|| \leq B$) distribution $\mathbb{D}$, we have:*

$$\begin{aligned}&\underset{\boldsymbol{M}_L \in \mathcal{M}}{Sup}\left[Err^\lambda(\boldsymbol{M}_L, \mathbb{D}) - Err^\lambda(\boldsymbol{M}_L, S_k^{pair})\right] \\ &\leq O\left(\lambda B^2\sqrt{D\ln(1/\delta)/k}\right)\end{aligned} \tag{21}$$

Theorem. 3 proves that to achieve an estimation error rate ϵ, $k = \Omega\left((\lambda B^2/\epsilon)^2 D\ln(1/\delta)\right)$ samples are sufficient.

Theorem 4 *Let $\boldsymbol{M}_L$ be any class of weighting metrics on the feature space $X = \mathbb{R}^D$, and define $d := Sup_{\boldsymbol{M}_L \in \mathcal{M}} \|\boldsymbol{M}_L\|_F^2$. Following the same parameter setting in Theorem. 3, we have:*

$$\begin{aligned}&\underset{\boldsymbol{M}_L \in \mathcal{M}}{Sup}\left[Err^\lambda(\boldsymbol{M}_L, \mathbb{D}) - Err^\lambda(\boldsymbol{M}_L, S_k^{pair})\right] \\ &\leq O\left(\lambda B^2\sqrt{d\ln(1/\delta)/k}\right)\end{aligned} \tag{22}$$

From Theorem. 4, we observe that if the learned metric $\mathbf{M}_L$ has a low metric learning complexity $d \ll D$, it can help sharpen the sample complexity result, yielding a dataset-dependent bound. Recall our objective Eqn. 6, $d := Sup_{\mathbf{M}_L \in \mathcal{M}} \|\mathbf{M}_L\|_F^2$ is already optimized via our proposed learning objective. Therefore, the bound is further tighter under the same number of samples.

5. Experiments

5.1. Experiment Settings

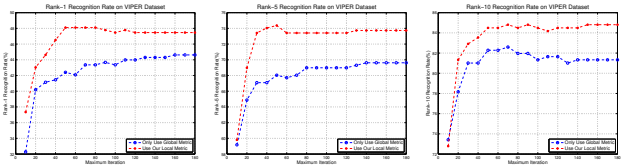
Data & Evaluation. We have performed thorough experiments and comparative studies to evaluate our method on five most widely-used benchmark datasets: VIPeR [7], QMUL GRID [19], CUHK Campus [10], CUHK03 [11] and Market-1501 [34]. The last three large-scale datasets are pretty challenging due to the extremely complicated variance of person appearance and abundant distractors. For a fair comparison, the training data of each dataset are used as the negative training samples for itself $\mathbf{Y}^{neg} = \mathbf{X}_{train}$, so no more extra information is utilized in the experiment. For all the experiments, the single-shot evaluation setting (except for the CUHK Campus dataset where the multi-shot matching setting is applied) is adopted and all the results are shown in the form of Cumulated Matching Characteristic (CMC) curves. Due to space limitation, we only report the cumulated matching accuracy at selected ranks in tables. The plot of the full curve is included in the supplementary materials.

Feature. The recently proposed high-dimensional feature LOMO [14] is adopted as the visual feature representation. Since it is not practical to directly use such a high dimensional feature in metric learning, we employ principal component analysis (PCA) to reduce the feature dimension.

Baselines. For fair comparisons, several global metric learning approaches [15, 14, 30] whose code is available to access and the feature can be replaced are compared to our proposed method under the same experiment setting and using the same LOMO feature. Besides, the most recent state-of-the-art published results are also reported for a thorough comparison. For all the experiments, the global metric learner, MLAPG [15] is chosen as the underlying baseline so that our online local metric adaptation algorithm is applied on top of it.

5.2. Influence of Global Metric Learning Quality

Our proposed OL-MANS algorithm is applied on top of a global metric $\mathbf{M}_G$, thus its overall performance may depend on the learning quality of adopted global metric learner. In order to verify whether our OL-MANS can always be helpful, global metrics obtained at various learning stages of a global metric learner [15] are tested, as in general the performance of a global metric learner improves with more training (e.g., more training iterations). As shown in Fig. 4, even the learned global metric does perform poorly (in its early training stages), our online local metric adaptation is able to consistently and significantly improve the performances by a large margin. This is because the local discriminative information introduced by hard negative samples is able to capture the specific crux of one identity which is quite helpful for identification.



(a) Rank-1 Evaluation (b) Rank-5 Evaluation (c) Rank-10 Evaluation

Figure 4. Demonstration about the influence of the quality of global metric. The x-axis means the maximum iteration time for global metric learning and the y-axis is the identification rate.

Methods	GRID		VIPeR	
	R=1	R=20	R=1	R=20
Euc	9.12	29.76	15.32	50.66
Euc+OL-MANS	20.88	45.12	21.99	56.11
XQDA[14]	12.96	43.52	38.99	91.94
XQDA+OL-MANS	29.20	50.96	43.54	92.15
MLAPG[15]	17.60	56.08	40.28	93.39
MLAPG+OL-MANS	30.16	59.36	44.97	93.64
DNSL[30]	15.12	53.12	40.19	93.54
DNSL+OL-MANS	28.96	56.96	43.67	93.61

Table 1. Comparison of identification rate with/without proposed OL-MANS on VIPeR and GRID. All the experiments are under the same setting and use the same **LOMO** feature. **+OL-MANS** means implementing our OL-MANS on the original global metric learner. Red represents the better results.

5.3. Influence of Global Metric Learner Choice

An interesting question is whether our OL-MANS can always work for any global metric learners as promised. To verify it, we conduct the following experiment that different kinds of global metric learners, Euclidean distance, XQDA [14], MLAPG [15] and DNSL [30] are adopted as the underlying global metric that our OL-MANS algorithm will be readily applied on. For each learner, we compare the identification rates without and with our online local metric adaptation. The 10-run-average results on VIPeR and GRID datasets are reported in Table. 1, as well as the complete CMC curves (included in the supplementary materials). We observe that for all the learners, our proposed online local metric adaptation algorithm is able to boost the identification performance with a significantly improvement, even double the identification accuracy (on GRID). Even for the most state-of-the-art global metric learner [30], applying our OL-MANS to it can still achieve a non-trivial improvement.

5.4. Training Costing Analysis and Comparison

Although every test probe needs to learn a local Mahalanobis metric at the test stage, solving a kernel SVM problem instead of solving the original PSD problem makes the learning efficient and largely reduces the training time. Ta-

Method	ITML [5]	MLAPG [15]	LADF [12]
Ave Time	20.5	25.8	31.7
Method	LMNN [27]	PRDC [36]	**OL-MANS**
Ave Time	152.9	394.6	4.8

Table 2. Average training time (seconds) on VIPeR.

Method	XQDA [30]	MLAPG [15]	MFA [28]
Training	3233.8	2732.8	437.8
Method	kLFDA [28]	DNSL [30]	**OL-MANS**
Training	995.2	3149.7	19.60

Table 3. Training time (seconds) on the Market-1501.

ble. 2 [2] provides a thorough comparison of average training time of various state-of-the-art metric learning-based methods on VIPeR dataset. Besides, Table. 3 shows the training time of different advanced global metric learners on a large-scale dataset, Market-1501. All the experiments are conducted on a remote server with an Intel i7-5930K @3.50GHz CPU and 32G memory. The total average training time of our method on VIPeR is only 4.81 seconds for the adaptation of all the 316 probes, much shorter than learning a single global metric in 25.82 seconds. For the large scale dataset Market-1501, the efficiency advantage of ours is much more pronounced. Our local metric adaptation time is $10 \sim 100$ times less than the other global metric learners. So the extra time spent in our local metric adaptation is indeed nominal compared with learning a global metric.

5.5. Extensive Comparisons on Benchmarks

Experiments on VIPeR: The VIPeR dataset [7] is a widely used benchmark dataset for PRID. It contains 632 pedestrian image pairs taken from 2 different cameras in an outdoor environment. We follow the widely adopted experimental protocol on VIPeR: 632 pairs are randomly divided into half for training and the other half for testing, and use 10-run-average for performance. We conducted the comparison experiment under the same experiment setting and using the same LOMO feature and the results are reported in Table. 4. Our proposed algorithm achieves the best performances on all the ranks. For the important Rank-1 evaluation, our performance 44.97% outperforms the second best approach LSSCDL by 2.31%. This promising performance indicates that the proposed local metric adaptation method is consistently effective, several representative examples are shown in Fig. 2. Besides, more results on state-of-the-art comparison are included in the supplementary materials.

Experiments on QMUL GRID: The QMUL underGround Re-IDentification (GRID) dataset [19] contains 250

[2]The total learning time of OL-MANS includes the local metric adaptation time and gallery ranking time for all probes.

Method	R=1	R=5	R=10	R=20
Ours	44.97	74.43	84.97	93.64
LSSCDL[31]	42.66	-	84.27	91.93
DNSL[30]	42.28	71.46	82.94	92.06
MLAPG[15]	40.73	69.94	82.34	92.37
XQDA[14]	40.00	68.13	80.51	91.08
TMA[21]	39.88	-	81.33	91.46
KISSME[9]	34.81	60.44	77.22	86.71
ITML[5]	24.64	49.78	63.04	78.39
LMNN[27]	29.43	59.78	73.51	84.91
kCCA[17]	30.16	62.69	76.04	86.80
MFA[28]	38.67	69.18	80.47	89.02
kLFDA[28]	38.58	69.15	80.44	89.15

Table 4. Comparison results on VIPeR (P = 316). **All the methods use the same LOMO feature**. RED color is the best result and BLUE color is the second best one.

Method	R=1	R=5	R=10	R=20
Ours	30.16	42.64	49.20	59.36
LSSCDL(LOMO)[31]	22.40	-	51.28	61.20
DNSL(LOMO)[30]	15.12	31.92	40.72	53.12
MLAPG(LOMO)[15]	17.60	33.52	43.36	56.08
XQDA(LOMO)[14]	12.96	26.80	34.56	43.52
EPKFM[3]	16.30	35.80	46.00	57.60
MtMCML[20]	14.08	34.64	45.84	59.84
RQDA[13]	15.20	30.08	39.20	49.28
M-RankSVM[18]	12.24	27.84	36.32	46.56
M-PRDC[18]	11.12	26.08	35.76	46.56
PRDC[36]	9.68	22.00	32.96	44.32

Table 5. Comparison results on GRID (P = 900).

pedestrian image pairs taken from 8 disjoint camera views and 775 additional images that do not belong to the 250 persons. GRID is also a pretty tough dataset because of the large viewpoint variations and the low-resolution image quality. The experimental protocol for GRID is the same as [15, 3, 20]: we randomly divide the 250 identities into half for training and the other half for testing as well as the extra 775 images are used as distractors to enlarge the gallery set. The average performance of 10 random trials is provided in Table. 5. It can be clearly observed that the proposed algorithm outperforms all the existing algorithms at Rank-1 by a very significant 7.76% improvement on the identification rate. Although the GRID dataset is more challenging than VIPeR, our proposed algorithm can still handle it well by adapting the local similarity structure of each probe.

Experiments on CUHK Campus: The CUHK Campus dataset [10] consists of 971 persons captured from two camera views in a campus environment, two images per person in each camera view. We split the set to 485 for training and 486 for testing and multi-shot matching scenario is applied to CUHK Campus dataset for evaluation [14, 1, 23, 28]. We

Method	R=1	R=5	R=10	R=20
Ours	68.44	87.16	92.67	95.88
LSSCDL[31]	65.97	-	-	
DNSL(LOMO)[30]	64.98	84.96	89.92	94.36
MLAPG(LOMO)[15]	64.24	85.41	90.84	94.92
XQDA(LOMO)[14]	63.21	83.89	90.04	94.16
kFLDA(LOMO)[28]	54.63	80.45	86.87	92.02
MFA(LOMO)[28]	54.79	80.08	87.26	92.72
kCCA(LOMO)[17]	54.63	80.45	86.87	92.02
IDLA[1]	47.53	-	-	-
Mid-L-F[33]	34.30	-	64.96	74.94
TSRPR[23]	32.70	51.20	64.40	76.30
SalMatch[32]	28.45	-	55.67	67.95
K-Ensb2[28]	24.00	38.90	46.70	55.40

Table 6. Comparison results on CUHK Campus (P = 486).

Method	R=1	R=5	R=10	R=20
Ours	61.68	88.39	95.23	98.47
MLAPG(LOMO) [15]	57.96	87.09	94.74	98.00
XQDA(LOMO) [14]	52.20	82.23	92.14	96.25
DNSL(LOMO) [30]	58.90	85.60	92.45	96.30
DeepReID[11]	20.65	51.50	66.50	80.00
Im-Deep[1]	54.74	86.50	93.88	98.10

Table 7. Comparison results on CUHK03 **Labeled** (P=100).

Method	R=1	R=5	R=10	R=20
Ours	62.71	87.59	93.80	97.55
MLAPG(LOMO)[15]	51.15	83.55	92.05	96.90
XQDA(LOMO)[14]	46.25	78.90	88.55	94.25
DNSL(LOMO)[30]	53.70	83.05	93.00	94.80
DeepReID[11]	19.89	50.00	64.00	78.50
Im-Deep[1]	44.96	76.01	83.47	93.15

Table 8. Comparison results on CUHK03 **Detected** (P = 100).

evaluate the performance by fusing scores of all the probe images of the same identity. As shown in Table. 6, the proposed method consistently outperforms other state-of-the-art methods in all identification rates.

Experiments on CUHK03: The CUHK03 dataset [11] contains 13164 images of 1360 pedestrians. All the images are captured by six surveillance cameras. Each person is observed by two disjoint camera views with an average of 4.8 images in each view. Two kinds of data are provided: manually cropped pedestrian images and images detected with a pedestrian detector. We follow the same experimental protocol [11, 15, 14]: splitting all the pedestrians into a training set of 1160 persons and a test set of 100 persons. The results in Table. 7 and Table. 8 show that for both two datasets, the proposed algorithm achieves the best performances at all ranks. It outperforms the second best approach by almost 10% in Rank-1 rate, even for the data under such a complicated practical situation, which is very significant.

Experiments on Market-1501: Market-1501 [34] is the largest image-based PRID benchmark dataset to date which contains 32668 bboxes of 1501 identities. Each person is recorded by six cameras at most, and two at least. For training and testing, the given fixed training and test set are utilized and both single-shot and multi-shot settings are used for evaluation. As shown by Table. 9, directly using Euclidean distance without metric learning and a state-of-the-art deep embedding-based method [24] are compared as baselines. We perform our OL-MANS to different global metric learners [15, 14, 30] based on the same LOMO features and experiment setting. As shown by the result, for all the global metric learners even for the Euclidean distance, a significant improvement on Rank-1 can be achieved by performing our OL-MANS algorithm to it, no matter under the single-shot or multi-shot evaluation setting.

Methods	Single-Q		Multi-Q	
	R=1	R=20	R=1	R=20
Baseline[34]	35.84	67.64	44.36	73.25
Kissme(LOMO)[34]	40.50	N/A	N/A	N/A
MFA-χ^2(LOMO)[28]	45.67	N/A	N/A	N/A
kLFDA(LOMO)[28]	51.37	N/A	52.67	N/A
Hist-Loss[24]	59.47	91.09	N/A	N/A
Euc(LOMO) [34]	32.93	63.87	40.33	69.40
Euc+OL-MANS	40.93	74.06	51.45	80.98
MLAPG(LOMO)[15]	43.87	88.40	61.33	96.40
MLAPG+OL-MANS	44.93	89.20	62.40	94.27
XQDA(LOMO)[14]	45.87	81.73	56.27	85.07
XQDA+OL-MANS	51.87	84.40	74.00	94.00
DNSL(LOMO)[30]	51.73	88.67	57.70	88.59
DNSL+OL-MANS	60.67	91.87	66.80	92.19

Table 9. Comparison results on the Market-1501 database under both **single-shot** and **multiple-shot** evaluation settings. Red represents the better result.

6. Conclusions

In this paper, we proposed a novel online local metric adaptation algorithm to learn a dedicated Mahalanobis metric for each probe at the test stage. Our new approach only uses negative samples for metric adaptation, which is practical in real situation. It largely reduces the demand for a large number of positive training data as in existing PRID methods, and it only incurs minimum computational costs to perform online training. In-depth theoretical analysis well justifies our algorithm and extensive experiments also demonstrate that our new approach consistently and significantly outperforms the state-of-the-art methods.

Acknowledgements

This work was supported in part by National Science Foundation grant IIS-1217302, IIS-1619078, and the Army Research Ofice ARO W911NF-16-1-0138.

References

[1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. *Differences*, 2015. 7, 8

[2] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM TIST*, 2011. 4

[3] 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*, 2015. 7

[4] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE TIT*, 1967. 4

[5] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *ICML*, 2007. 7

[6] E. Fetaya and S. Ullman. Learning local invariant mahalanobis distances. *Proceedings of The 32st International Conference on Machine Learning*, 2015. 3

[7] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *PETS*, 2007. 2, 6, 7

[8] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*. 2012. 2

[9] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012. 7

[10] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *ACCV*, 2012. 6, 7

[11] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014. 6, 8

[12] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013. 1, 2, 7

[13] S. Liao, Y. Hu, and S. Z. Li. Joint dimension reduction and metric learning for person re-identification. *CoRR*, 2014. 7

[14] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*, 2015. 1, 2, 3, 6, 7, 8

[15] S. Liao and S. Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *ICCV*, 2015. 1, 2, 3, 6, 7, 8

[16] V. E. Liong, J. Lu, and Y. Ge. Regularized local metric learning for person re-identification. *PR Letters*, 2015. 2

[17] G. Lisanti, I. Masi, and A. Del Bimbo. Matching people across camera views using kernel canonical correlation analysis. In *ICDSC*, 2014. 7, 8

[18] C. C. Loy, C. Liu, and S. Gong. Person re-identification by manifold ranking. In *ICIP*, 2013. 7

[19] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *CVPR*, 2009. 6, 7

[20] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *IEEE TIP*, 2014. 1, 2, 7

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

[22] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*, 2013. 2

[23] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a semantic representation for person re-identification and search. In *CVPR*, 2015. 7, 8

[24] E. Ustinova and V. Lempitsky. Learning deep embeddings with histogram loss. In *NIPS*, 2016. 8

[25] R. Vezzani, D. Baltieri, and R. Cucchiara. People reidentification in surveillance and forensics: A survey. *ACM CSUR*, 2013. 1

[26] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by video ranking. In *ECCV*. 2014. 3

[27] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, 2005. 7

[28] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*. 2014. 7, 8

[29] G. Zhang, Y. Wang, J. Kato, T. Marutani, and K. Mase. Local distance comparison for multiple-shot people re-identification. In *ACCV*. 2013. 2

[30] L. Zhang, T. Xiang, and S. Gong. Learning a discriminative null space for person re-identification. In *CVPR*, 2016. 6, 7, 8

[31] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan. Sample-specific svm learning for person re-identification. In *CVPR*, 2016. 2, 7, 8

[32] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013. 8

[33] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, 2014. 8

[34] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. Scalable person re-identification: A benchmark. In *ICCV*, 2015. 6, 8

[35] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *CVPR*, 2011. 1, 2, 3

[36] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *IEEE TPAMI*, 2013. 1, 7