# Efficient PSD Constrained Asymmetric Metric Learning for Person Re-identification

Shengcai Liao and Stan Z. Li
Center for Biometrics and Security Research, National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, China
{scliao,szli}@nlpr.ia.ac.cn

## Abstract

*Person re-identification is becoming a hot research topic due to its value in both machine learning research and video surveillance applications. For this challenging problem, distance metric learning is shown to be effective in matching person images. However, existing approaches either require a heavy computation due to the positive semidefinite (PSD) constraint, or ignore the PSD constraint and learn a free distance function that makes the learned metric potentially noisy. We argue that the PSD constraint provides a useful regularization to smooth the solution of the metric, and hence the learned metric is more robust than without the PSD constraint. Another problem with metric learning algorithms is that the number of positive sample pairs is very limited, and the learning process is largely dominated by the large amount of negative sample pairs. To address the above issues, we derive a logistic metric learning approach with the PSD constraint and an asymmetric sample weighting strategy. Besides, we successfully apply the accelerated proximal gradient approach to find a global minimum solution of the proposed formulation, with a convergence rate of $O(1/t^2)$ where $t$ is the number of iterations. The proposed algorithm termed MLAPG is shown to be computationally efficient and able to perform low rank selection. We applied the proposed method for person re-identification, achieving state-of-the-art performance on four challenging databases (VIPeR, QMUL GRID, CUHK Campus, and CUHK03), compared to existing metric learning methods as well as published results.*

## 1. Introduction

Person re-identification is a technique to search a desired person from a large set of gallery. This task is very challenging because there exist complex intra-class variations in illumination, pose or viewpoint, blur, and occlusion. Many approaches have been proposed for person re-identification [4], which greatly advance this field.

Among existing approaches, the metric learning methods are shown to be effective in matching person images [33, 9, 10, 13]. The Mahalanobis distance learning proposed by [27] is a relatively simple but effective approach, which has been widely studied. This method tries to learn a Mahalanobis distance function parameterized by a PSD matrix to separate the positive sample pairs from the negative sample pairs. However, existing Mahalanobis metric learning methods either require a heavy computation when the PSD constraint is applied, or ignore the PSD constraint and learn a free distance function that makes the learned metric potentially noisy. We argue in this paper that the PSD constraint provides a useful regularization to smooth the solution of the metric, and hence the learned metric is more robust than without the PSD constraint.

Alternatively, the PSD constraint can also be achieved by representing the Mahalanobis metric with a product of two matrices [26, 33]. However, the quadratic form makes the objective function more complex. Besides, a rank parameter has to be given beforehand for optimization. For a different rank, the metric has to be learned again, which is inconvenient. What's more, there is no principle to select the most informative dimensions from the learned projection matrix. In practice, most existing methods apply the Principle Component Analysis (PCA) dimension reduction method before metric learning, of which the effect is unclear. Considering this, it is better to perform metric learning in a higher dimensional space while allowing low rank selection with the learned metric.

Another issue in applying metric learning for person re-identification is that, the numbers of positive and negative sample pairs are largely unbalanced. Especially, most existing person re-identification datasets are relatively small (e.g. VIPeR [5] and QMUL GRID [17]), resulting in very limited number of positive sample pairs. Therefore,

the learning process can be dominated by a large number of negative sample pairs, making the limited positive pairs easily be neglected and resulting in a weak metric. This issue is even more important than the PSD constraint according to the study of this paper.

To address the above issues, we derive a logistic metric learning approach with the PSD constraint and an asymmetric sample weighting strategy. Besides, we successfully apply the widely used Accelerated Proximal Gradient (APG) [21, 25, 2] approach to find a global minimum solution of the proposed formulation, with a convergence rate of $O(1/t^2)$ where $t$ is the number of iterations. We applied the proposed method termed MLAPG to the person re-identification problem, achieving state-of-the-art performance on four challenging databases (VIPeR [5], QMUL GRID [17], CUHK Campus [11], and CUHK03 [12]), compared to existing metric learning methods as well as published results. In addition, we also reveal some nice properties of the proposed method, such as the fast convergence, efficient computation, and the ability of low rank selection[1].

## 2. Related Work

Many Mahalanobis distance learning approaches have been proposed following [27]. Among them, the Large Margin Nearest Neighbor Learning (LMNN) [26], Information Theoretic Metric Learning (ITML) [3] and Logistic Discriminant Metric Learning (LDML) [7] are three representative methods and regarded as the state of the art [10]. The LMNN algorithm tries to learn a Mahanalobis distance metric to improve the k-nearest neighbor classifier, where the goal is to pull samples of the same class lying within the k-nearest neighbors, while push samples from different classes by a large margin. The ITML algorithm considers minimizing the differential relative entropy between two multivariate Gaussians for learning the Mahalanobis distance function. It is further formulated as a Bregman optimization problem that minimizes the LogDet divergence subject to linear constraints. The LDML algorithm applies the logistic discriminant function to model the probability of samples being the same class or not. It can be converted as a linear logistic discriminant model, and a general gradient descent procedure is applied to maximize the log-likelihood and learn the metric.

The metric learning approach has been applied to computer vision problems in recent years and shown to be effective [26, 7, 33, 9, 10, 13, 14]. Particularly, the Keep It Simple and Straightforward Metric Learning (KISSME) [10], Locally-Adaptive Decision Functions (LADF) [13], and Cross-view Quadratic Discriminant Analysis (XQDA) [14] algorithms have shown to be achieving the state of the art for face recognition and person re-identification. The KISSME algorithm considers a log likelihood ratio test of two Gaussian distributions, and a simplified and very efficient solution can be obtained accordingly. The LADF method is a joint model of a distance metric and a locally adapted thresholding rule, which are combined to form a unified quadratic classifier. The XQDA algorithm learns a discriminant subspace as well as a distance metric simultaneously, and it is able to perform dimension reduction and select the optimal dimensionality.

Though not explicitly addressing the problem of unbalanced positive and negative sample pairs, several metric learning approaches are immune to this issue. For example, the LMNN algorithm only considers impostors within the k-nearest neighbors, the KISSME and XQDA algorithms formulate the metric learning problem as separately estimating two gaussian distributions for the two classes, and the PRDC algorithm [33] considers ranking one positive pair and one negative pair at a time which is balanced.

The proposed approach is mostly related to LDML [7] and PRDC [33], which propose two similar logistic metric learning formulations. Though sharing the same superiority due to the soft-margin loss function, there are notable differences with the proposed method: 1) we explicitly model the PSD constraint, which brings a notable accuracy gain; 2) we introduce an asymmetric sample weighting strategy in the loss function, which is important and also brings impressive performance gain; and 3) we present an efficient approach based on APG to solve the PSD constrained metric learning problem, which has a fast convergence rate. Detail analysis of the three components can be found in the experiments. In contrast, the LDML algorithm does not apply the PSD constraint, does not consider largely unbalanced samples, and the LDML solver is based on the general gradient descent algorithm which is known to be slow (O(1/t)). The PRDC algorithm is quite different, which formulates the metric learning problem as a ranking problem through relative distance comparison of a positive pair and a negative pair at a time. Therefore, it needs to compute $O(n^2)$ difference vectors and $O(n^3)$ relative difference comparisons, which becomes intractable with large dataset.

Regarding optimization, the APG is a general first-order method but its limitation is that one needs to find a closed-form solution for the constraint iteratively. In this paper, we show that the PSD constrained metric learning problem can be successfully solved by APG, and we give a closed-form solution for this constraint in each iteration, resulting in efficient learning and some interesting findings such as the low-rank representation and selection. To our knowledge, such solution has not been found for metric learning.

There are many other approaches beyond metric learning for person re-identification, such as the ensemble of local features (ELF) [6], SDALF [1], RankSVM [23], kBiCov

[1]Code available at http://www.cbsr.ia.ac.cn/users/scliao/projects/mlapg/

[18], salience match [30], mid-level filter [32], local fisher discriminant analysis (LF) [22], to name a few. Interested readers please refer to [4] for a survey.

## 3. Cross-view Logistic Metric Learning

Suppose we have a cross-view training set $\{\mathbf{X}, \mathbf{Z}, \mathbf{Y}\}$, where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \in \mathbb{R}^{d\times n}$ contains $n$ samples in a $d$-dimensional space from one view, $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_m) \in \mathbb{R}^{d\times m}$ contains $m$ samples in the same $d$-dimensional space but from the other view, and $\mathbf{Y} \in \mathbb{R}^{n\times m}$ is the matching label between $\mathbf{X}$ and $\mathbf{Z}$, with $y_{ij} = 1$ indicating that $\mathbf{x}_i$ and $\mathbf{z}_j$ are from the same class, and $y_{ij} = -1$ otherwise. We call $(\mathbf{x}, \mathbf{z})$ a positive sample pair if $y = 1$, and a negative sample pair otherwise. The cross-view matching problem arises from many applications, for example, heterogeneous face recognition and viewpoint invariant person re-identification. Note that $\mathbf{Z}$ is the same with $\mathbf{X}$ in the single-view matching scenario.

The task is to learn a Mahalanobis distance function [27]

$$D^2_{\mathbf{M}}(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|^2_{\mathbf{M}} = (\mathbf{x} - \mathbf{z})^T\mathbf{M}(\mathbf{x} - \mathbf{z}) \quad (1)$$

to measure the distance between the cross-view samples, where $\mathbf{M} \succeq 0$ is a PSD matrix so that $D_{\mathbf{M}}$ satisfies the nonnegativity and the triangle inequality. To learn such a metric, a smooth and convex loss function can be applied. We consider a log-logistic loss function similar as in [7, 33]

$$f_{\mathbf{M}}(\mathbf{x}, \mathbf{z}) = log\left(1 + e^{y\left(D^2_{\mathbf{M}}(\mathbf{x},\mathbf{z}) - \mu\right)}\right), \quad (2)$$

where $\mu = ED^2_{\mathbf{I}}(\mathbf{x}, \mathbf{z})$ is a constant positive bias, which is applied considering that $D$ has a lower bound of zero. The logistic function provides a soft margin to separate the two classes, which is particularly useful for classification problems, for example, in the traditional logistic regression formulation. Accordingly, the overall loss function is

$$F(\mathbf{M}) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} f_{\mathbf{M}}(\mathbf{x}_i, \mathbf{z}_j), \quad (3)$$

where $w_{ij} = \frac{1}{N_{pos}}$ if $y_{ij} = 1$, and $\frac{1}{N_{neg}}$ otherwise, and $N_{pos}$ and $N_{neg}$ are the total number of positive and negative sample pairs, respectively. This asymmetric weighting is important because $N_{pos}$ and $N_{neg}$ are heavily unbalanced.

As a result, the cross-view logistic metric learning problem is formulated as

$$\min_{\mathbf{M}} F(\mathbf{M}), \quad s.t.\ \mathbf{M} \succeq 0. \quad (4)$$

## 4. Accelerated Proximal Gradient Solution

### 4.1. Proximal Operator

The problem (4) contains a nonlinear, but smooth and convex objective function, and a closed convex conic constraint. Therefore, it has a unique global minimum solution. Considering this structure, we apply the widely used APG [21, 25, 2] approach to solve (4). APG is a first-order optimization method that achieves the optimal convergence rate $O(1/t^2)$ where t is the number of iterations [21]. Given a solution path $\{\mathbf{M}_t\}_{t\geq 0}$, the APG optimization procedure constructs an aggregation sequence $\{\mathbf{V}_t\}_{t\geq 1}$ by linearly combining the two most recent solution $\mathbf{M}_{t-1}$ and $\mathbf{M}_{t-2}$ at each iteration to accelerate the optimization, that is

$$\mathbf{V}_t = \mathbf{M}_{t-1} + \frac{\alpha_{t-1} - 1}{\alpha_t}(\mathbf{M}_{t-1} - \mathbf{M}_{t-2}), \quad (5)$$

where $\alpha_t = (1 + \sqrt{4\alpha^2_{t-1} + 1})/2$ following [2].

With the aggregation forward matrix $\mathbf{V}_t$, the gradient of the objective function $F(\mathbf{V})$ at iteration $t$ is computed as

$$\begin{aligned}\nabla F(\mathbf{V}_t) &= \frac{\partial F(\mathbf{V})}{\partial \mathbf{V}}|_{\mathbf{V}=\mathbf{V}_t} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} g^{(t)}_{ij}(\mathbf{x}_i - \mathbf{z}_j)(\mathbf{x}_i - \mathbf{z}_j)^T \\ &= \mathbf{X}\mathbf{A}_t\mathbf{X}^T - \mathbf{X}\mathbf{G}_t\mathbf{Z}^T - (\mathbf{X}\mathbf{G}_t\mathbf{Z}^T)^T + \mathbf{Z}\mathbf{B}_t\mathbf{Z}^T,\end{aligned} \quad (6)$$

where

$$g^{(t)}_{ij} = \frac{w_{ij}y_{ij}}{1 + e^{-y_{ij}\left(D^2_{\mathbf{V}_t}(\mathbf{x}_i,\mathbf{z}_j) - \mu\right)}}, \quad (7)$$

and $\mathbf{A}$ and $\mathbf{B}$ are two diagonal matrices with the main diagonal containing the row sum and column sum of $\mathbf{G}$, respectively. This matrix computation is more efficient than previous methods which compute all $(\mathbf{x}_i - \mathbf{z}_j)$.

As a result, a proximal operator can be constructed to linearize the objective function at the search point $\mathbf{V}_t$,

$$\begin{aligned}&P_{\eta_t}(\mathbf{M}, \mathbf{V}_t) = \\ &F(\mathbf{V}_t) + \langle \mathbf{M} - \mathbf{V}_t, \nabla F(\mathbf{V}_t)\rangle + \frac{1}{2\eta_t}\|\mathbf{M} - \mathbf{V}_t\|^2_F,\end{aligned} \quad (8)$$

where $< \mathbf{A}, \mathbf{B} >= Tr(\mathbf{A}^T\mathbf{B})$ is the matrix inner product, $\|A\|_F$ is the Frobenius norm of a matrix, and $\eta_t > 0$ is the step size. With a proper step size $\eta_t$, the proximal operator $P_{\eta_t}(\mathbf{M}, \mathbf{V}_t)$ is an upper bound of $F(\mathbf{M})$ [21]. Therefore, at the $t$-th iteration, minimizing (4) is equivalent to solving

$$\min_{\mathbf{M}} P_{\eta_t}(\mathbf{M}, \mathbf{V}_t), \quad s.t.\ \mathbf{M} \succeq 0. \quad (9)$$

### 4.2. Solution

**Theorem 1.** *The solution to the problem (9) is*

$$\mathbf{M}_t = \mathbf{U}_t\mathbf{\Lambda}^+_t\mathbf{U}^T_t, \quad (10)$$

*where $\mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}^T_t$ is the singular value decomposition (SVD) of a symmetric matrix $\mathbf{C}_t = \mathbf{V}_t - \eta_t\nabla F(\mathbf{V}_t)$, with $\mathbf{U}^T_t\mathbf{U}_t = \mathbf{I}$ and $\mathbf{\Lambda}_t$ being a diagonal matrix containing the singular values of $\mathbf{C}_t$ in the main diagonal, and $\mathbf{\Lambda}^+_t = \max\{0, \mathbf{\Lambda}_t\}$.*

*Proof.* By ignoring the constant term $F(\mathbf{V}_t)$ and adding another constant term $\frac{\eta_t}{2}\|\nabla F(\mathbf{V}_t)\|_F^2$, solution of (9) is equivalent to minimizing

$$\frac{1}{2\eta_t}\|\mathbf{M}-(\mathbf{V}_t-\eta_t\nabla F(\mathbf{V}_t))\|_F^2=\frac{1}{2\eta_t}\|\mathbf{M}-\mathbf{C}_t\|_F^2. \quad (11)$$

Since $\{\mathbf{M}_k\}_{k=0}^{t-1}$ is a sequence of symmetric PSD matrices, we can infer from (5) and (6) that $\mathbf{C}_t$ is a symmetric matrix. Therefore, $\mathbf{C}_t$ can be decomposed as $\mathbf{C}_t=\mathbf{U}_t\mathbf{\Lambda}_t\mathbf{U}_t^T$ by SVD, where $\mathbf{U}_t^T\mathbf{U}_t=\mathbf{I}$, and $\mathbf{\Lambda}_t$ is a diagonal matrix containing the eigenvalues $\{\lambda_i\}_{i=1}^d$ of $\mathbf{C}_t$ in the main diagonal. As a result, finding the optimal solution $\mathbf{M}_t$ in (11) satisfying $\mathbf{M}_t\succeq 0$ is known as the nearest PSD matrix approximation problem under the Frobenius norm [24, 8]. In fact, considering that $\mathbf{U}_t^T\mathbf{U}_t=\mathbf{I}$, we have

$$\|\mathbf{M}-\mathbf{C}_t\|_F^2=\|\mathbf{U}_t^T\mathbf{M}\mathbf{U}_t-\mathbf{\Lambda}_t\|_F^2, \quad (12)$$

and $\mathbf{S}=\mathbf{U}_t^T\mathbf{M}\mathbf{U}_t$ must also be a PSD matrix because $\mathbf{M}$ is required to be PSD. Thus, we have $s_{ii}\geq 0$ for all $i$, and

$$\begin{aligned}\|\mathbf{M}-\mathbf{C}_t\|_F^2&=\|\mathbf{S}-\mathbf{\Lambda}_t\|_F^2\\&=\sum\nolimits_{i,j}s_{ij}^2+\sum\nolimits_i(s_{ii}-\lambda_i)^2\\&\geq\sum\nolimits_{\lambda_i<0}(s_{ii}-\lambda_i)^2\geq\sum\nolimits_{\lambda_i<0}\lambda_i^2\end{aligned} \quad (13)$$

The equality is uniquely achieved for $\mathbf{S}=\mathbf{\Lambda}_t^+$. Therefore, the solution to the problem (9) is given by (10). □

Note that the operation of (10) was early noticed by [24], where it is called a smoothing procedure to regularize a matrix to be PSD. This kind of smoothing operation is useful to regularize the learning procedure and derive a robust metric, as will be demonstrated later through experiments.

### 4.3. Line Search of Step Size

According to [21], the working step size is $\eta\leq L$ where $L$ is the Lipschitz constant of the gradient function $\nabla F(\mathbf{M})$. However, L is unknown and not easy to estimate. Alternatively, we start with a large $\eta_0$ and do a line search similar as in [2], which iteratively check whether

$$F(\mathbf{M}_t)\leq P_{\eta_t}(\mathbf{M}_t,\mathbf{V}_t) \quad (14)$$

is satisfied. If the condition (14) is not satisfied, adapt $\eta_t$ to be $\eta_t/\gamma$ for a constant factor $\gamma>1$ and repeat the condition checking until it is satisfied. This procedure adaptively ensures that the step size is suitable and the convergence is guaranteed. In practice, we empirically found that the adaptation occurs only in the initialization of the algorithm. Once a suitable step size is found, the condition (14) is found to be always satisfied. This is supported in an experiment shown in Fig. 1, where it is observed that $\eta$ could be as large as 128 (given that features were normalized to unit length), and once initialized, it was unchanged. Therefore, the line search procedure of the proposed algorithm does not require much computation.

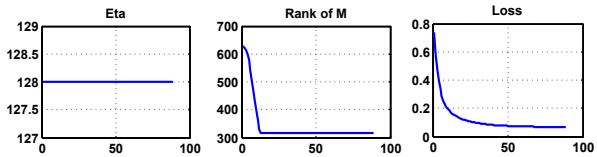



Figure 1. A demonstration of $\eta_t$, $r_t$, and $F(\mathbf{M}_t)$ as a function of the iteration number $t$ during the training on the VIPeR dataset [5]. The training data contained $c=316$ classes with $d=631$ dimensions, the $\eta$ was initialized to $2^8$, and $\gamma=2$.

### 4.4. Dimension Reduction

Since the solution of the proposed algorithm is always found by SVD as in (10), a low-rank structure of the solution $\mathbf{M}_t$ is naturally followed. Therefore, we are able to decompose $\mathbf{M}_t$ as $\mathbf{M}_t=\mathbf{P}_t\mathbf{P}_t^T$, where

$$\mathbf{P}_t=\mathbf{U}_t^+(\mathbf{\Lambda}_t^{++})^{1/2}, \quad (15)$$

and $\mathbf{U}_t^+$ and $\mathbf{\Lambda}_t^{++}$ are sub matrices of $\mathbf{U}_t$ and $\mathbf{\Lambda}_t^+$, respectively, by removing dimensions corresponding to zero diagonal elements of $\mathbf{\Lambda}_t^+$. This way, $\mathbf{P}_t\in\mathbb{R}^{d\times r_t}$ is a projection matrix, where $r_t$ is the number of nonzero diagonal elements of $\mathbf{\Lambda}_t^+$. Therefore, the Mahalanobis distance defined in (1) can be reduced to an Euclidean distance as follows

$$D_{\mathbf{P}_t}^2(\mathbf{x},\mathbf{z})=\|\mathbf{P}_t^T\mathbf{x}-\mathbf{P}_t^T\mathbf{z}\|_2^2, \quad (16)$$

where $\mathbf{P}_t$ is applied for dimension reduction. As demonstrated in Fig. 1, for a training data of 316 classes, we found that the proposed algorithm was able to reduce the rank of $\mathbf{M}$ from 631 to 315 after 15 iterations.

Furthermore, by removing dimensions corresponding to small eigenvalues of $\mathbf{C}_t$, the dimensions of the projection matrix $\mathbf{P}_t$ can be further reduced. The influence of recognition performance by this kind of dimension reduction will be analyzed in the experiment section.

### 4.5. Convergence Analysis

The following theorem states that the convergence rate of the proposed algorithm, termed MLAPG, is $O(1/t^2)$, where $t$ is the number of iterations.

**Theorem 2.** *Let $\{\mathbf{M}_t\}$ and $\{\mathbf{V}_t\}$ be the sequences generated by the MLAPG algorithm. Then $\forall t\geq 1$ we have*

$$F(\mathbf{M}_t)-F(\mathbf{M}^*)\leq\frac{2\gamma L\|\mathbf{M}_0-\mathbf{M}^*\|_F^2}{(t+1)^2}, \quad (17)$$

*where $\mathbf{M}^*$ is the optimal solution to (4).*

This theorem can be similarly proofed following [2], and it is omitted here. The example shown in Fig. 1 indicates that the algorithm converges with 89 iterations, measured by $|\frac{F(\mathbf{M}_t)-F(\mathbf{M}_{t-1})}{F(\mathbf{M}_{t-1})}|\leq 0.001$.

## 5. Experiments

We evaluated the proposed algorithm on four challenging person re-identification databases, VIPeR [5], QMUL GRID [17], CUHK Campus [11], and CUHK03 [12]. Several state-of-the-art metric learning algorithms with the same feature representation were compared, and the state-of-the-art published results on the four datasets were also compared. We detail the experimental description below.

### 5.1. Feature Representation

We utilized the Local Maximal Occurrence (LOMO) feature proposed in [14] for person representation. The LOMO feature is proved to be robust against illumination variations and viewpoint changes, and it is also discriminant, which captures local region characteristics of a person. The LOMO feature first applies a multiscale Retinex transform for image preprocessing, resulting in a consistent color representation. Then, a set of sliding windows are extracted, and both color and texture histograms are computed, with each histogram bin represents the occurrence probability of one pattern in a subwindow. To overcome the difficulty in viewpoint changes, the maximal occurrence of each pattern among all subwindows at the same horizontal location is computed. The final descriptor has 26,960 dimensions.

### 5.2. Baseline Metric Learning Algorithms

We evaluated several state-of-the-art metric learning algorithms, including LMNN v2.5[26], ITML [3], LDML [7], PRDC [33], KISSME [10], LADF [13], and XQDA [14], where LMNN and ITML are two popularly used metric learning algorithms, while KISSME, LADF, and XQDA are recent methods that have shown state-of-the-art results for person re-identification. In particular, the LDML and PRDC algorithms also apply the logistic loss function for metric learning. Brief introductions of these algorithms have been described in Section 2. For all algorithms, the PCA was first applied but all energies were reserved. This step reduces the computation of metric learning but does not affect the performance, since the feature dimensions of LOMO are much larger than the number of samples in the three databases. Then, the LMNN, ITML, KissMe, and LADF algorithms were applied with the first 100 PCA components. The XQDA, PRDC, LDML, and MLAPG algorithms were applied with all PCA components, since they were able to further learn a low-rank projection matrix. The learned projection matrices of XQDA and MLAPG were truncated to 100 dimensions for a fair comparison. Parameters used for the proposed method were $\eta_0 = 2^8$ and $\gamma = 2$, which were not critical for the performance of the proposed method, because they only affected the convergence rate of the proposed method to find the global minimum. Besides, we set the maximal iterations to 300, with a stopping criterion by $|\frac{F(\mathbf{M}_t)-F(\mathbf{M}_{t-1})}{F(\mathbf{M}_{t-1})}| \leq 10^{-4}$.

(a) VIPeR [5] (b) GRID [17]

Figure 2. Example pairs of images from the VIPeR and GRID databases. Images in the same column represent the same person.

Table 1. Comparison of state-of-the-art metric learning algorithms with the same feature on the VIPeR dataset (P=316). The subspace dimensions were truncated to 100.

| Method | rank = 1 | rank = 10 | rank = 20 |
|---|---|---|---|
| MLAPG | **39.21** | **81.42** | **92.50** |
| XQDA | 38.23 | 81.14 | 92.18 |
| KISSME | 33.54 | 79.30 | 90.47 |
| LMNN | 28.42 | 72.31 | 85.32 |
| LADF | 27.63 | 75.47 | 88.29 |
| ITML | 19.02 | 52.31 | 67.34 |
| LDML | 13.99 | 38.64 | 48.73 |
| PRDC | 12.15 | 35.82 | 48.26 |

### 5.3. Experiments on VIPeR

VIPeR [5] is a challenging person re-identification dataset that has been widely used for benchmark evaluation. It contains 632 pairs of person images, captured by a pair of cameras in an outdoor environment. Images in VIPeR contain large variations in background, illumination, and viewpoint. Fig. 2(a) shows some example pairs of images from this dataset. All images were scaled to $128 \times 48$ pixels. The widely adopted experimental protocol on this database is to randomly divide the 632 pairs of images into half for training and the other half for test, and repeat the procedure 10 times to get an average performance.

Table 1 shows the results with the same LOMO feature representation. It is clear that the proposed MLAPG algorithm outperforms the other existing metric learning methods. Especially, MLAPG achieves a 39.21% rank-1 identification rate, outperforming the second-best one XQDA by about 1%. It can also be seen that the performances of XQDA and KISSME are impressive; especially, they achieve a comparable performance as MLAPG at larger ranks.

The success of MLAPG is mainly due to the soft-margin loss function, the PSD regularization, and the asymmetric sample weighting strategy. Note that LDML and PRDC also apply a logistic loss function for metric learning, but their performances are not so good. For LDML, the reason may be the absence of the PSD constraint, and the learning

Table 2. Comparison of state-of-the-art results on the VIPeR database (P=316). MLAPG dimensions were not truncated.

| Method | rank=1 | rank=10 | rank=20 | Reference |
|---|---|---|---|---|
| MLAPG | **40.73** | **82.34** | **92.37** | Proposed |
| XQDA | 40.00 | 80.51 | 91.08 | 2015 CVPR [14] |
| SCNCD | 37.80 | 81.20 | 90.40 | 2014 ECCV [29] |
| Kernel Ensb 2 | 36.1 | 80.1 | 85.6 | 2014 ECCV [28] |
| kBiCov | 31.11 | 70.71 | 82.45 | 2014 IVC [18] |
| LADF | 30.22 | 78.92 | 90.44 | 2013 CVPR [13] |
| SalMatch | 30.16 | 65.54 | 79.15 | 2013 ICCV [30] |
| Mid-level Filter* | 29.11 | 65.95 | 79.87 | 2014 CVPR [32] |
| MtMCML | 28.83 | 75.82 | 88.51 | 2014 TIP [19] |
| RPLM | 27.00 | 69.00 | 83.00 | 2012 ECCV [9] |
| SSCDL | 25.60 | 68.10 | 83.60 | 2014 CVPR [15] |
| LF | 24.18 | 67.12 | 82.00 | 2013 CVPR [22] |
| SDALF | 19.87 | 49.37 | 65.73 | 2013 CVIU [1] |
| KISSME | 19.60 | 62.20 | 77.00 | 2012 CVPR [10] |
| PCCA | 19.27 | 64.91 | 80.28 | 2012 CVPR [20] |
| PRDC | 15.66 | 53.86 | 70.09 | 2013 TPAMI [33] |
| ELF | 12.00 | 44.00 | 61.00 | 2008 ECCV [6] |

*Note that [32] reports a 43.39% rank-1 accuracy by fusing their method with LADF [13]. Fusing different methods generally improves the performance. In fact, we also tried to fuse our method with LADF, and got a 47.88% rank-1 identification rate.

is largely affected by excessive negative sample pairs. For PRDC, our comparison here is not very fair, because PRDC could not directly handle the high-dimensional LOMO feature and we had to apply PRDC on the PCA subspace, but it was empirically shown that the sample difference based ranking worked not well with transformed features[2].

We also compare the performance of the proposed approach (without truncation of the learned projection matrix) to the state-of-the-art results reported on the VIPeR database using the same protocol. The results are summarized in Table 2. From Table 2 it can be observed that the proposed algorithm achieves the new state of the art, 40.73% at rank 1. Compared to the second best one XQDA, the improvement by MLAPG is 0.73%, 1.83%, and 1.29%, respectively, at rank 1, 10, and 20. This promising result may indicate that the proposed MLAPG algorithm is effective in learning a robust metric for viewpoint invariant person re-identification. It should be noted that the success of the proposed method is partially due to the robust feature representation, as can be observed from Table 1 that the existing algorithms KISSME and XQDA can also achieve impressive performance with the LOMO feature.

### 5.4. Experiments on QMUL GRID

The QMUL underGround Re-IDentification (GRID) Dataset [17] is another challenging person re-identification test bed but have not been largely noticed. The GRID dataset was captured from 8 disjoint camera views in a underground station. It contains 250 pedestrian image pairs, with each pair being two images of the same person from different camera views. Besides, there are 775 additional images that do not belong to the 250 persons which can be used to enlarge the gallery. Sample images from GRID can be found in Fig. 2(b). It can be seen that these images have poor image quality and low resolutions, and contain large illumination and viewpoint variations.

[2]This has been confirmed by the authors of PRDC [33].

Table 3. Comparison of metric learning algorithms with the same feature representation on the QMUL GRID database (P=900). The subspace dimensions were truncated to 100.

| Method | rank = 1 | rank = 10 | rank = 20 |
|---|---|---|---|
| XQDA | **16.56** | **41.44** | **52.48** |
| MLAPG | 15.60 | 40.48 | **52.48** |
| LMNN | 10.80 | 34.24 | 45.76 |
| KISSME | 10.64 | 31.60 | 43.20 |
| ITML | 9.44 | 27.04 | 35.20 |
| LDML | 8.16 | 22.24 | 27.36 |
| PRDC | 7.52 | 23.84 | 31.44 |
| LADF | 6.00 | 27.36 | 41.28 |

An experimental setting of 10 random trials is provided for the GRID dataset. For each trial, 125 image pairs are used for training, and the remaining 125 image pairs, as well as the 775 background images are used for test.

Performance comparison of the metric learning algorithms applied with the same LOMO feature representation is shown in Table 3. Compared to Table 1, it can be observed that the GRID database is more challenging than the VIPeR database, because the GRID database has 8 underground camera views, while VIPeR contains only two camera views. The MLAPG algorithm achieves comparable accuracy to the best performer XQDA; both of them show quite better performance than other existing algorithms. Interestingly, LADF performs good on the VIPeR database, but it is not able to handle the complex viewpoint changes involved in the GRID database. In contrast, MLAPG and XQDA seems to be more robust than the existing algorithms in addressing such challenges of the GRID database.

Next, we compare the performance of the proposed method (without truncation of the learned projection matrix) to state-of-the-art results reported on the GRID database following the same protocol, as shown in Table 4. It can be observed that the proposed MLAPG algorithm outperforms all existing algorithms at rank 1, though it is only slightly better than XQDA. This indicates that the new algorithm is also promising in handling complex viewpoint changes and poor image conditions as involved in the GRID database. Note that the MtMCML algorithm [19] utilizes the camera network information available from the GRID dataset, and learns specific metrics for each individual pair of camera views. Therefore, it successfully improves the re-identification accuracy at higher ranks.

Table 4. Comparison of state-of-the-art results on the QMUL GRID database (P=900). MLAPG dimension was not truncated.

| Method | rank=1 | rank=10 | rank=20 |
|---|---|---|---|
| MLAPG | **16.64** | 41.20 | 52.96 |
| XQDA [14] | 16.56 | 41.84 | 52.40 |
| MtMCML [19] | 14.08 | **45.84** | **59.84** |
| MRank-RankSVM [16] | 12.24 | 36.32 | 46.56 |
| MRank-PRDC [16] | 11.12 | 35.76 | 46.56 |
| RankSVM [23] | 10.24 | 33.28 | 43.68 |
| PRDC [33] | 9.68 | 32.96 | 44.32 |
| L1-norm [16] | 4.40 | 16.24 | 24.80 |

Table 5. Comparison of state-of-the-art multi-shot results on the CUHK Campus database (P=486,M=2). MLAPG dimensions were not truncated.

| Method | rank=1 | rank=10 | rank=20 |
|---|---|---|---|
| MLAPG | **64.24** | **90.84** | **94.92** |
| XQDA [14] | 63.21 | 90.04 | 94.16 |
| Mid-level Filter [32] | 34.30 | 64.96 | 74.94 |
| SalMatch [30] | 28.45 | 55.67 | 67.95 |
| GenericMetric [11] | 20.00 | 56.04 | 69.27 |
| eSDC [31] | 19.67 | 40.29 | 50.58 |

### 5.5. Experiments on CUHK Campus

The CUHK Campus dataset [11] contains two camera views captured in a campus environment. Different from the VIPeR and GRID datasets, images in this dataset are of higher resolution. This dataset includes 971 persons, two images per person in each camera view. The persons were split to 485 for training and 486 for test. All the images were normalized to $160 \times 60$ for evaluation. Multi-shot matching scenario was applied, which fused scores of multiple images of the same person by the sum rule.

We compare the performance of the proposed method (without truncation of the learned projection matrix) to the state-of-the-art results reported on the CUHK Campus database following the same protocol. As shown in Table 5, the proposed method improves the rank-1 identification rate by 1.03% over XQDA, while both of them largely outperform the other existing state of the art methods.

### 5.6. Experiments on CUHK03

Compared to XQDA, MLAPG has a 1%-2% improvement for most experiments, which is not a big improvement considering that XQDA is simpler. However, the XQDA model is based on covariance estimation, which follows the Gaussian assumption and may have a limitation with complex data distributions. To demonstrate this, we did an experiment on the CUHK03 dataset, which is much larger and more complex. The CUHK03 dataset [12] includes 13,164 images of 1,360 pedestrians. It was captured with 6 surveillance cameras over months, with each person observed by two disjoint camera views and having an average of 4.8 images in each view. Beyond manually cropped pedestrian images, samples detected with a state-of-the-art pedestrian detector are also provided. This is a more realistic setting considering misalignment, occlusions and part missing.

Table 6. Comparison of state-of-the-art rank-1 identification rates (%) on the CUHK03 database [12] with both labeled and detected setting (P=100). MLAPG dimensions were not truncated. Most of the compared results are from [12].

| | Labeled | Detected |
|---|---|---|
| LOMO+MLAPG | **57.96** | **51.15** |
| LOMO+XQDA [14] | 52.20 | 46.25 |
| DeepReID [12] | 20.65 | 19.89 |
| KISSME [10] | 14.17 | 11.70 |
| LDML [7] | 13.51 | 10.92 |
| eSDC [31] | 8.76 | 7.68 |
| LMNN [26] | 7.29 | 6.25 |
| ITML [3] | 5.53 | 5.14 |
| SDALF [1] | 5.60 | 4.87 |

We run our algorithm with the same setting of [12]. That is, the dataset was partitioned into a training set of 1,160 persons and a test set of 100 persons. The experiments were conducted with 20 random splits and all results were computed with the single-shot setting. The rank-1 identification rates of various algorithms in both labeled and detected setting are shown in Table 6. The proposed method achieved 57.96% and 51.15% rank-1 identification rates with the labeled bounding boxes and the automatically detected bounding boxes, respectively, which clearly outperform the state-of-the-art method XQDA [14], with an improvement of 5.76% for the labelled setting, and 4.9% for the detected setting. This indicates a better capability of MLAPG in learning from complex data.

### 5.7. Analysis of the Proposed Algorithm

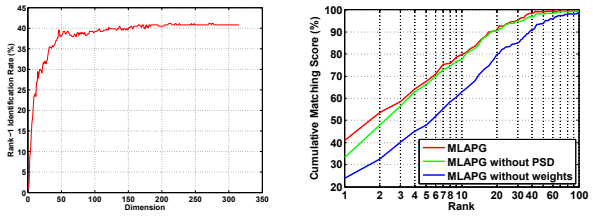
In this subsection, we will analyze the proposed algorithm in several aspects: dimension reduction effects, PSD regularization effects, asymmetric weighting effects, and training time. The analysis was performed on the VIPeR database, by randomly sampling a training set of 316 persons, and a test set of the remaining persons.

#### 5.7.1 Influence of Dimensions

To understand the influence of the dimensions, we performed the matching with different dimensions of $P$ in Eq. (16). The result is shown in Fig. 3(a). It can be observed that: i) all dimensions of the learned projection matrix are useful for distance matching, because the identification rate almost increases with the increasing dimensions; and ii) the low-rank representation is effectively learned, observing that the proposed method with a very low dimension 20 can achieve about 30% rank-1 identification rate, and with 50

Table 7. Comparison of average training time (seconds).

| **Method** | KISSME | LDML | XQDA | ITML | MLAPG | LADF | LMNN | PRDC |
|---|---|---|---|---|---|---|---|---|
| **Training Time** | **1.3** | 1.4 | 1.9 | 19.3 | 25.1 | 29.2 | 141.3 | 356.3 |
| **MEX Function** | No | Yes | No | No | No | Yes | Yes | No |




(a) (b)

Figure 3. (a) Rank-1 identification rate versus the dimensionality of the projection matrix $\mathbf{P}$; and (b) CMC curves of the proposed method with and without the PSD constraint and the asymmetric sample weighting.

dimensions the performance is almost comparable to other larger dimensions. Therefore, selection of a discriminant low-rank subspace is possible.

#### 5.7.2 Influence of PSD Regularization

For the influence of the PSD regularization, Fig. 3(b) shows an example of the proposed method with and without the PSD constraint. As can be observed, MLAPG with the PSD constraint is much better than without the PSD constraint at rank-1 (40.82% vs. 33.23%), though the two methods show similar performance with larger ranks. As noticed in [24], projecting the solution to a PSD cone helps to derive a smooth and robust estimate. Our study also shows that this helps to derive a smooth and robust metric, so that it generalizes better, resulting in more correct retrievals at rank 1. This phenomena also explains why the proposed method performs better at rank-1 in Table 1.

#### 5.7.3 Influence of Asymmetric Weights

We did another experiment comparing the proposed method with and without the asymmetric sample weighting in Eq. (3). The results are also shown in Fig. 3(b). It can be observed that the proposed MLAPG algorithm with the asymmetric sample weighting is much better than without the asymmetric sample weighting (40.82% vs. 23.73% at rank 1). This finding, as well as the benefit of the PSD regularization, explains why the proposed method performs much better than the LDML algorithm where a logistic metric learning formulation is also proposed. Besides, by the comparison shown in Fig. 3(b), we can find that the asymmetric sample weighting is even more important than the PSD constraint, which has not been paid much attention to in the metric learning literature.

#### 5.7.4 Training Time

Table 7 shows a comparison of average training time on the VIPeR dataset for 10 random trials. All algorithms are implemented in MATLAB, with some algorithms having MEX functions implemented in C or C++ to accelerate the computation. The training was performed on a desktop PC with an Intel i5-2400 @3.10GHz CPU. From the comparison in Table 7, KISSME and XQDA are shown to be very efficient, which have closed-form solutions. The LDML algorithm is also very efficient, thought with the help of MEX functions. The training speed of the proposed MLAPG algorithm is comparable to that of ITML and LADF, but it is much faster than LMNN and PRDC. Note that the LMNN algorithm also learns a metric that requiring PSD, but it took nearly 5 times longer than MLAPG for training, despite that MEX functions were applied in LMNN for acceleration.

The overall training time of MLAPG on this database is reasonable, though it requires SVD per iteration. However, the feature dimension in this example is 631 after a full-energy PCA, which is relatively small. Considering that the complexity of SVD is $O(d^3)$, the MLAPG algorithm may still have a difficulty in directly addressing high-dimensional data. Therefore, working in the PCA subspace with a reasonable dimension may still be a better choice in practice, as with many other metric learning methods.

## 6. Conclusion

We have proposed a logistic metric learning algorithm with the PSD constraint and an asymmetric sample weighting strategy, which can be efficiently solved by APG. The proposed method termed MLAPG is shown to be fast in convergence, and with low-rank representation. We have applied it to the person re-identification problem, achieving state-of-the-art performance on four challenging databases, compared to existing metric learning methods as well as published results. Due to the general APG optimization framework, other additional smooth or non-smooth constraints may be studied in future research for robust metric learning. Besides, the proposed method may also have potential values for other applications, e.g. face recognition.

## Acknowledgements

This work was supported by NSFC #61203267, #61375037, #61473291, #61572501, National Science and Technology Support Program #2013BAK02B01, and CAS Project #KGZD-EW-102-2.

## References


[1] L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117(2):130–144, 2013. 2, 6, 7

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009. 2, 3, 4

[3] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007. 2, 5, 7

[4] S. Gong, M. Cristani, S. Yan, and C. C. Loy. *Person Re-Identification*. Springer, 2014. 1, 3

[5] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International workshop on performance evaluation of tracking and surveillance*, 2007. 1, 2, 4, 5

[6] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, 2008. 2, 6

[7] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In *International Conference on Computer Vision*, 2009. 2, 3, 5, 7

[8] N. J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103(0):103 – 118, 1988. 4

[9] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *European Conference on Computer Vision*. 2012. 1, 2, 6

[10] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 1, 2, 5, 6, 7

[11] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *Assian Conference on Computer Vision*, 2012. 2, 5, 7

[12] W. Li, R. Zhao, T. Xiao, and X. Wang. DeepReID: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 2, 5, 7

[13] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 1, 2, 5, 6

[14] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 2, 5, 6, 7

[15] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, and J. Bu. Semi-supervised coupled dictionary learning for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 6

[16] C. C. Loy, C. Liu, and S. Gong. Person re-identification by manifold ranking. In *IEEE International Conference on Image Processing*, volume 20, 2013. 7

[17] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1988–1995. IEEE, 2009. 1, 2, 5, 6

[18] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014. 3, 6

[19] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *IEEE Transactions on Image Processing*, 2014. 6, 7

[20] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 6

[21] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2003. 2, 3, 4

[22] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 3, 6

[23] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person re-identification by support vector ranking. In *BMVC*, 2010. 2, 7

[24] N. Schwertman and D. Allen. Smoothing an indefinite variance-covariance matrix. *Journal of Statistical Computation and Simulation*, 9(3):183–194, 1979. 4, 8

[25] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*, 2008. 2, 3

[26] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. *Advances in neural information processing systems*, 2006. 1, 2, 5, 7

[27] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. J. Russell. Distance metric learning with application to clustering with side-information. In *Proceedings of Neural Information Processing Systems*, 2002. 1, 2, 3

[28] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision*. 2014. 6

[29] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *Proceedings of the European Conference on Computer Vision*, 2014. 6

[30] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *International Conference on Computer Vision*, 2013. 3, 6, 7

[31] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 7

[32] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014. 3, 6, 7

[33] W.-S. Zheng, S. Gong, and T. Xiang. Reidentification by relative distance comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):653–668, 2013. 1, 2, 3, 5, 6, 7