Pattern Recognition 94 (2019) 218–229

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Towards fast and kernelized orthogonal discriminant analysis on person re-identification

Min Cao[a,b], Chen Chen[a,b,*], Xiyuan Hu[a,b], Silong Peng[a,b,c]

[a] Institute of Automation Chinese Academy of Sciences Beijing, China
[b] University of Chinese Academy of Sciences, Beijing, China
[c] Beijing ViSystem Corporation Limited China

ARTICLE INFO

Article history:
Received 6 May 2018
Revised 1 April 2019
Accepted 26 May 2019
Available online 28 May 2019

Keywords:
Person re-identification
Metric learning
Singularity problem
Orthogonal discriminant analysis

ABSTRACT

Recognizing a person across different non-overlapping camera views, is the task of person re-identification. For achieving the task, an effective way is to learn a discriminative metric by minimizing the within-class variance and maximizing the between-class variance simultaneously. However, the dimension of sample feature vector is usually greater than the number of training samples, as a result, the within-class scatter matrix is singular and the metric cannot be learned. In this paper, we propose to solve the singularity problem by employing the pseudo-inverse of the within-class scatter matrix and learning an orthogonal transformation for the metric. The proposed method can be effectively solved with a closed-form solution and no parameters required to tune. In addition, we develop a kernel version against non-linearity in person re-identification, and a fast version for more efficient solution. In experiments, we prove the validity and advantage of the proposed method for solving the singularity problem in person re-identification, and analyze the effectiveness of both kernel version and fast version. Extensively comparative experiments on VIPeR, PRID2011, CUHK01 and CUHK03 person re-identification benchmark datasets, show the state-of-the-art results of the proposed method.

1. Introduction

The aim of person re-identification (re-id) is to identify the same person, captured from different non-overlapping camera views. Because of its abroad application in video surveillance system such as tracking criminals and behavior analysis, person re-id is favored by the academe in recent years. Despite of years of efforts by researchers, it is still a challenging problem due to the large intra-class variations caused by the change in illumination, person pose and occlusion across views. In addition, the imperfect pedestrian detection result and the similarity in appearance among different people further increase its difficulty in real applications.

Existing research on this topic has mainly concentrated on two aspects: feature representation and metric learning. Researchers focus on developing an effective feature representation [1,2] against the variation in appearance of people across different camera views. The feature dimensionality is usually large for a more-refined representation, which leads to *curse of dimensionality*. Moreover, in most cases, features are extracted in the unsupervised setting so that the identification ability is weak. Whereas, the metric learning based method mainly concentrates on learning a proper distance function (or a discriminative low-dimensional space) in the supervised setting. Out of the many metric learning approaches, Linear Discriminant Analysis (LDA) is the most classical and popular approach due to its characteristics of no-parameters tuning required and closed-form solution. LDA aims to learn an optimal linear transformation by which the new low-dimensional feature space is obtained. In this space, the within-class variance is minimized and the between-class variance is maximized simultaneously. The within-class scatter matrix to be non-singular is a requirement in the classical LDA. However, in person re-id, the dimension of sample feature vector is generally greater than the sample size, as a result, all scatter matrices are singular. This means that person re-id is a singularity (or undersampled) problem and classical LDA cannot be directly applied. To address this limitation, several efforts have been devoted to handling the singularity problem in person re-id. Regularized LDA (RLDA) in which a scalar multiple of the identity matrix is added to the within-class scatter matrix has been applied to person re-id [2,3], with parameters required to tune. However, the solutions may not be optimal since they suffer from the degenerate eigenvalue

* Corresponding author at: 95 Zhongguancun East Road, Beijing 100190, China.
E-mail addresses: caomin2014@ia.ac.cn (M. Cao), chen.chen@ia.ac.cn (C. Chen).

https://doi.org/10.1016/j.patcog.2019.05.035
0031-3203/

Table 1
The objective function of various variants of LDA. S_w and S_b denote the within-class scatter matrix and the between-class scatter matrix, respectively. $(S)^+$ denotes the pseudo-inverse of the matrix S. L^* denotes the optimal transformation.

Method	Objective function
LDA	$L^* = \arg\max trace\{(L^T S_w L)^{-1} L^T S_b L\}$
Regularized LDA	$L^* = \arg\max trace\{(L^T (S_w + \lambda I) L)^{-1} L^T S_b L\}$
Null space LDA	$L^* = \arg\max_{L^T S_w L = 0} trace\{L^T S_b L\}$
Pseudo-inverse LDA	$L^* = \arg\max trace\{(L^T S_w L)^{+} L^T S_b L\}$

problem, i.e several eigenvectors share the same eigenvalue. Zhang et al. [4] proposed to apply Null Space LDA (NLDA) to person re-id. NLDA aims to compute the discriminant vectors in the null space of the within-class scatter matrix, but it may not be applicable for low-dimensional data since the null space of the within-class scatter matrix is empty in the low-dimensional data.

In this context, we propose Pseudo-inverse LDA (PLDA) to deal with the singularity problem in person re-id. PLDA employs the pseudo-inverse of the within-class scatter matrix to overcome the singularity problem, and is equivalent to approximating the solution using a least-squares solution method. Compared with the variants of LDA mentioned above, PLDA not only solves the singularity problem, but also avoids the above shortcomings. Table 1 shows the objective functions of these variants of LDA. For the solution of PLDA, the traditional solution is to apply Generalized Singular Value Decomposition (GSVD) [5], yet with the high computational cost. Instead, we solve PLDA based on the simultaneous diagonalization of within-class matrix, between-class matrix and total scatter matrix [6]. The resulting optimal discriminant vectors are orthogonal to each other, known as Orthogonal LDA (OLDA). OLDA is specifically designed for the singularity problem, with complete theoretical analysis [6]. Besides, there are some important characteristics about OLDA: 1) closed-form solution, 2) no-parameters tuning required and 3) better robustness for the noise in the data, which are advantageous for person re-id.

Furthermore, in consideration of the non-linear distribution of data in person re-id, we develop a non-linear version via the kernel trick, which combines the strengths of both OLDA and kernel-based learning techniques, to boost the performance of person re-id. In addition, it is imperative to note here that the eigen-decomposition is involved in solving the model. Compared with the QR decomposition, the eigen-decomposition is computationally expensive for the high-dimensional data. Hence, we present a fast OLDA algorithm [7] for solving the singularity problem in person re-id. The fast version is implemented by using the QR decomposition rather than the eigen-decomposition. Its computational complexity, run time and performance will be analyzed in detail in the experiment section.

In summary, the contributions of this paper are the following three-folds: **1)** For the singularity problem in person re-id, we propose to learn an orthogonal transformation with the pseudo-inverse of the within-class scatter matrix for the first time. **2)** Against the non-linear distribution of data in person re-id, we develop a kernel version for learning the orthogonal transformation, thereby boosting the performance of person re-id. **3)** In order to improve the solving efficiency, we present a fast version with the unchanged performance of person re-id.

The rest of this paper is organized as follows. Section 2 reviews the related works. In section 3, the proposed method, and the corresponding kernel version and fast version are described in detail, respectively. In Section 4, we present a thorough comparative evaluation of the proposed method with respect to the state-of-the-art methods on four benchmark datasets, and a detail analysis of kernel version and fast version for the proposed method. Finally, in Section 5, we conclude this paper.

2. Related works

In this paper, we explore a novel way to solve the singularity problem in person re-id, which is closely pertinent to LDA approach. In the following, we will discuss the relevant works on LDA, person re-id and LDA based person re-id.

2.1. Linear discriminant analysis

LDA is a classical supervised metric learning method, aiming to seek a linear transformation in the training data that maximizes the between-class variance and minimizes the within-class variance, simultaneously. It has been applied successfully in many application areas for decades. However, there is a main disadvantage for LDA: the within-class scatter matrix must be nonsingular, which is not met in many applications. In order to overcome the limitation and make LDA applicable in a wider range of applications, researchers have proposed many methods to extend the classical LDA. Specifically, a kind of methods to overcome the limitation are to project the original data to a lower-dimensional space by Principal Component Analysis (PCA) approach resulting in a full-rank within-class scatter matrix before LDA, known as PCA+LDA (or two-stage LDA) [8]. Another kind of methods are to modify the within-class scatter matrix by adding a perturbation term, known as RLDA [9]. In addition, Chen et al. [10] proposed to compute the most discriminant transformation by modifying the Fisher's criterion of the classical LDA: the within-class variance equals zero and the between-class variance is maximized, so that the singularity problem is implicitly avoided, known as NLDA. Furthermore, there are a kind of approaches to avoid the singularity problem: PLDA in which the inverse of within-class scatter matrix is replaced with the pseudo-inverse of the one in the Fisher's criterion. The methods based on this approach include LDA/GSVD [5], Uncorrelated LDA (ULDA) [6] and OLDA [6]. The LDA/GSVD applies the GSVD to compute the optimal transformation. For ULDA, the solution is obtained by diagonalizing the three scatter matrices simultaneously and the resulting features in the reduced space are uncorrelated to each other. After ULDA, an orthogonalization step is applied to the transformation, deducing the OLDA. The discriminant vectors in OLDA are orthogonal to each other. Some methods mentioned above are closely related, for example, NLDA is a special OLDA under some condition [7].

2.2. Person re-identification

In the current field of person re-id, researchers usually assume that pedestrian detection has been completed and the cropped person image is taken as input of person re-id system. In this system, feature extraction and metric learning are the essential steps. The discriminative ability through extracting features in person image is limited, therefore learning a discriminative distance metric is important for closing the gap between different features of same person and is the focus of this paper. The metric learning based method can be divided into two groups: iterative-learning based [11,12] and closed-form solution based [2–4]. In the iterative-learning based method, the objective function is constructed based on criteria that a pair of true match should have a smaller distance than that of a wrongly matched pair, and the optimal metric function is obtained by using the iterative optimization method such as gradient descent method. Compared with the iterative-learning based method, the closed-form solution based method is simple and efficient, which is usually related to LDA technology, such as cross-view quadratic discriminant analysis (XQDA) [2]. In this paper, we propose to learn an optimal transformation based on the LDA technology and the closed-form solution can be derived. Recently, various deep learning architectures [13,14] have also been

proposed to either address visual ambiguity of people across views, or learn better metric function. However, convolution neural network (CNN) architectures used in person re-id are not competitive on public datasets with small size.

2.3. Linear discriminant analysis based person re-identification

Because of a large number of theoretical researches on LDA and its characteristics of non-parameter tuning required and closed-form solution, some metric learning based person re-id methods have been proposed to learn an optimal metric distance function based on LDA. Pedagadi et al. [3] adopt two-steps to learn the low-dimensional discriminative feature space: PCA and then Regularized Local Fisher Discriminant Analysis (RLFDA), which is the graph-based LDA approach with regularization. Since the feature dimensionality is large compared to the number of samples after the first-step PCA, the singularity problem still exists in the second-step and RLFDA is therefore applied to solve the problem. Liao et al. [2] proposed to learn a subspace with the training data, and at the same time learn a distance function in the subspace by applying RLDA. For RLDA, choosing an appropriate perturbation is critical, since a large perturbation results in the loss of information in the scatter matrix and a small perturbation may not be effective in solving the singularity problem. The main disadvantage of RLDA is that the optimal amount of the perturbation is difficult to determine. To this end, NLDA has been developed to learn the optimal metric distance function in person re-id [4]. However, one limitation is that NLDA may not be applicable for low-dimensional data. In this paper, we propose to solve the singularity problem in person re-id with the aid of OLDA, which is applicable for both low-dimensional data and high-dimensional data.

3. Methodology

3.1. Problem description

Given a probe person captured from one camera view and a candidate set in which people are captured from another camera view, we aim to obtain an ordered list of candidate IDs for the probe person, based on the distances between the probe person and the candidates in ascending order. In the training phase, an optimal feature transformation is learned based on the proposed method. In the test phase, the distance between the probe person and the candidate is computed by the Euclidean distance in the new feature space obtained based on the learned transformation.

In consideration of many notations used in the rest of the paper, we present some important notations in Table 2 for convenience.

3.2. Classical linear discriminant analysis

We start with briefly introducing the LDA before describing the proposed person re-id method.

Given a feature matrix $X = [\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}] \in \mathbb{R}^{m\times n}$ with n people in the training set and m-dimensional feature vector for each person, classical LDA computes a linear transformation $L \in \mathbb{R}^{m\times d}$ $(m > d)$ that maps the feature vector $\boldsymbol{x_i}$ of i-th person to a new feature vector $\boldsymbol{y_i}$ in the d-dimensional space:

$$\boldsymbol{y_i} = L^T \boldsymbol{x_i}. \tag{1}$$

Define the matrices

$$\begin{aligned} H_w &= \frac{1}{\sqrt{n}}\left[X_1 - \boldsymbol{c^{(1)}}(\boldsymbol{e^{(1)}})^T, \ \ldots, \ X_k - \boldsymbol{c^{(k)}}(\boldsymbol{e^{(k)}})^T\right], \\ H_b &= \frac{1}{\sqrt{n}}\left[\sqrt{n_1}(\boldsymbol{c^{(1)}} - \boldsymbol{c}), \ \ldots, \sqrt{n_k}(\boldsymbol{c^{(k)}} - \boldsymbol{c})\right], \\ H_t &= \frac{1}{\sqrt{n}}\left(X - \boldsymbol{c}\boldsymbol{e^T}\right), \end{aligned} \tag{2}$$

where $X_i \in \mathbb{R}^{m\times n_i}$ $(i = 1, \ldots, k)$ is the feature matrix of i-th class sample and $\sum_{i=1}^{k} n_i = n$, $\boldsymbol{e^{(i)}} = [1, \ldots, 1]^T \in \mathbb{R}^{n_i}$ and $\boldsymbol{e} = [1, \ldots, 1]^T \in \mathbb{R}^n$, $\boldsymbol{c^{(i)}} = \frac{1}{n_i} X_i \boldsymbol{e^{(i)}}$ and $\boldsymbol{c} = \frac{1}{n} X \boldsymbol{e}$ are the centroid of the i-th class sample and the global centroid, respectively.

Then, the within-class scatter matrix S_w, the between-class scatter matrix S_b and the total scatter matrix S_t are expressed as

$$S_w = H_w H_w^T, \quad S_b = H_b H_b^T, \quad S_t = H_t H_t^T, \tag{3}$$

where it is easy to verify that $S_t = S_b + S_w$.

The goal of LDA is to find an optimal transformation L^* such that the within-class samples are much closer while the between-class samples farther apart in the reduced-dimensional space:

$$L^* = \arg\max_L trace\{(L^T S_w L)^{-1} L^T S_b L\}. \tag{4}$$

The optimization problem is equivalent to solve the generalized eigen-problem $S_b \boldsymbol{l} = \lambda S_w \boldsymbol{l}$. The transformation L^* can be obtained as the eigenvectors corresponding to the $k-1$ largest eigenvalues of $S_w^{-1} S_b$, if S_w is nonsingular. However, in person re-id, it is usually the case that $m > n$ resulting in a singular S_w and the classical LDA cannot be directly applied. Therefore, we propose to solve the singularity problem by considering the pseudo-inverse instead of the inverse of S_w. The optimal solution is computed based on the simultaneous diagonalization of the three scatter matrices, and resulting in an orthogonal transformation.

3.3. Learning an orthogonal transformation based on pseudo-inverse LDA

The new optimization problem is defined as

$$L^* = \arg\max_L trace\{(L^T S_t L)^{+} L^T S_b L\}, \tag{5}$$

where $(L^T S_t L)^+$ denotes the pseudo-inverse and is equivalent to $(L^T S_w L)^+$ due to $S_t = S_b + S_w$.

We solve the above maximization problem based on the simultaneous diagonalization of the three scatter matrices.

Theorem 1. *Let the SVD of H_t as $H_t = U\Sigma V^T$, where $\Sigma = \begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix}$ and $\Sigma_t \in \mathbb{R}^{r_t \times r_t}$ with $r_t = rank(S_t)$. Divide U into $U = [U_1, U_2]$ with $U_1 \in \mathbb{R}^{m\times r_t}$ and $U_2 \in \mathbb{R}^{m\times(m-r_t)}$. Denote $T = \Sigma_t^{-1} U_1^T H_b$ and compute its SVD as $T = P\widetilde{\Sigma}Q^T$. The simultaneous diagonalization of the three scatter matrices is achieved by the matrix $G = U_1 \Sigma_t^{-1} P$.*

Table 2
The description of important notations used in the paper.

Notations	Descriptions	Notations	Descriptions
m	Feature dimension in the original space	n	Sample size in the training set
d	Feature dimension in the reduced space	X	Feature matrix in the training set
k	Number of classes	S_w	Within-class scatter matrix
S_b	Between-class scatter matrix	S_t	Total scatter matrix
r_b	Rank of the matrix S_b	r_t	Rank of the matrix S_t
L	Transformation matrix	–	–

Proof. According to the SVD of H_t, we have

$$S_t = H_t H_t^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V \Sigma^T U^T = U\begin{bmatrix}\Sigma_t^2 & 0\\ 0 & 0\end{bmatrix}U^T,$$

$$\Rightarrow \quad U^T S_t U = \begin{bmatrix}U_1\\ U_2\end{bmatrix} S_t \begin{bmatrix}U_1 & U_2\end{bmatrix} = \begin{bmatrix}\Sigma_t^2 & 0\\ 0 & 0\end{bmatrix},$$

$$\Rightarrow \quad U_1^T S_t U_1 = \Sigma_t^2,$$

$$\Rightarrow \quad \Sigma_t^{-1} U_1^T S_t U_1 \Sigma_t^{-1} = I_t. \tag{6}$$

Since $S_t = S_w + S_b$, we have

$$\Sigma_t^{-1} U_1^T S_w U_1 \Sigma_t^{-1} + \Sigma_t^{-1} U_1^T S_b U_1 \Sigma_t^{-1} = I_t. \tag{7}$$

Let $T = \Sigma_t^{-1} U_1^T H_b$ and compute the SVD as $T = P\widetilde{\Sigma} Q^T$. Then

$$\Sigma_t^{-1} U_1^T S_b U_1 \Sigma_t^{-1} = \Sigma_t^{-1} U_1^T H_b H_b^T U_1 \Sigma_t^{-1} = P\widetilde{\Sigma} Q^T Q \widetilde{\Sigma}^T P^T = P\Sigma_b P^T, \tag{8}$$

where $\Sigma_b = \widetilde{\Sigma}^2$. Since P is an orthogonal matrix, it follows that

$$P^T \Sigma_t^{-1} U_1^T S_b U_1 \Sigma_t^{-1} P = \Sigma_b. \tag{9}$$

Accordingly, from Eqs. (7) and (8), we have

$$\Sigma_t^{-1} U_1^T S_w U_1 \Sigma_t^{-1} = I_t - P\Sigma_b P^T,$$

$$\Rightarrow \quad P^T \Sigma_t^{-1} U_1^T S_w U_1 \Sigma_t^{-1} P = I_t - \Sigma_b = \Sigma_w. \tag{10}$$

Considering the orthogonal matrix P, the last line in Eq. (6) can be rewritten as

$$P^T \Sigma_t^{-1} U_1^T S_t U_1 \Sigma_t^{-1} P = I_t. \tag{11}$$

Now, combining Eqs. (9), (10) and (11) together, we define $G = U_1 \Sigma_t^{-1} P$ and it follows that

$$G^T S_t G = I_t, \quad G^T S_b G = \Sigma_b, \quad G^T S_w G = \Sigma_w. \tag{12}$$

Then all three scatter matrices are diagonalized by the matrix $G = U_1 \Sigma_t^{-1} P$. □

According to Theorem 1, we rewrite $L^T S_t L$ and $L^T S_b L$ as

$$L^T S_t L = L^T (G^{-1})^T (G^T S_t G) G^{-1} L = \widetilde{L}^T I_t \widetilde{L} = \widetilde{L}^T \widetilde{L},$$
$$L^T S_b L = L^T (G^{-1})^T (G^T S_b G) G^{-1} L = \widetilde{L}^T \Sigma_b \widetilde{L}, \tag{13}$$

where

$$\widetilde{L} = G^{-1} L. \tag{14}$$

Then, the following equation holds:

$$F = trace\left\{(L^T S_t L)^+ L^T S_b L\right\} = trace\left\{(\widetilde{L}^T \widetilde{L})^+ \widetilde{L}^T \Sigma_b \widetilde{L}\right\}$$
$$= trace\left\{(\widetilde{L}\widetilde{L}^+)^T \Sigma_b (\widetilde{L}\widetilde{L}^+)\right\}. \tag{15}$$

We compute the SVD of $\widetilde{L}$ as:

$$\widetilde{L} = M\begin{bmatrix}\Sigma_l & 0\\ 0 & 0\end{bmatrix}N^T. \tag{16}$$

where $\Sigma_l \in \mathbb{R}^{r_l \times r_l}$ with $r_l = rank(\widetilde{L})$. Since both M and N are orthogonal matrices and Σ_l is a diagonal matrix, so

$$\widetilde{L}\widetilde{L}^+ = M\begin{bmatrix}\Sigma_l & 0\\ 0 & 0\end{bmatrix}N^T N\begin{bmatrix}\Sigma_l^{-1} & 0\\ 0 & 0\end{bmatrix}M^T = M\begin{bmatrix}I_l & 0\\ 0 & 0\end{bmatrix}M^T. \tag{17}$$

Then, Eq. (15) can be rewritten as:

$$F = trace\left\{M\begin{bmatrix}I_l & 0\\ 0 & 0\end{bmatrix}M^T \Sigma_b M\begin{bmatrix}I_l & 0\\ 0 & 0\end{bmatrix}M^T\right\}$$
$$= trace\left\{\begin{bmatrix}I_l & 0\\ 0 & 0\end{bmatrix}M^T \Sigma_b M\begin{bmatrix}I_l & 0\\ 0 & 0\end{bmatrix}\right\}$$
$$= trace\{M_l^T \Sigma_b M_l\}, \tag{18}$$

where M_l is generated by the first r_l columns of M.

Now, the optimization problem in Eq. (5) becomes the maximization of $trace\{M_l^T \Sigma_b M_l\}$. Next, we introduce a lemma [15] for solving the maximization problem.

Lemma 1. *For any matrix $A \in \mathbb{R}^{m\times q}$ $(q \le m)$ with $A^T A = I$, and the positive semi-definite matrix $J \in \mathbb{R}^{m\times m}$ with $\boldsymbol{h_i}$ be the eigenvector of J corresponding to the i-th largest eigenvalue λ_i, we have $trace(A^T J A) \le \lambda_1 + \cdots + \lambda_q$. The equation holds when $A = [\boldsymbol{h_1}, \ldots, \boldsymbol{h_q}]E$, where $E \in \mathbb{R}^{q\times q}$ is an arbitrary orthogonal matrix.*

The proof of the lemma 1 is not provided here for the sake of space and interested readers can refer to [15].

According to Lemma 1, we have $F = trace\{M_l^T \Sigma_b M_l\} \le \lambda_1 + \cdots + \lambda_{r_b}$ with $r_b = rank(S_b)$. The maximization of F is reached if $M_l = \begin{bmatrix}E\\ 0\end{bmatrix}$, where $E \in \mathbb{R}^{r_b \times r_b}$ is an arbitrary matrix with $E^T E = I$. We fed M_l into Eq. (16) with $r_b = r_l$ and obtain $\widetilde{L} = \begin{bmatrix}E\Sigma_l N^T\\ 0\end{bmatrix}$. Since both E and N are the orthogonal matrices and Σ_l is the diagonal matrix, we denote $A = E\Sigma_l N^T$ as an arbitrary nonsingular matrix.

Now, with Eq. (14), we solve the optimization problem in Eq. (5) by $L^* = GA$.

There is a problem for selection of the arbitrary nonsingular matrix A. The simplest choice is $A = I$ and correspondingly $L^* = G$. With Eq. (12), we can find that $L^{*T} S_t L^* = I_t$. It means that the discriminant vectors are S_t-orthogonal to each other. As a result, the new features obtained by the optimal transformation L^* are uncorrelated to each other, which is named as ULDA. However, ULDA involves the minimal redundant information in the new feature space, so that it is likely to lead to overfitting and be sensitive to the noise in the data.

Therefore, we make the QR decomposition of G as $G = Q_g R_g$ and choose $A = R_g^{-1}$. Correspondingly, $L^* = Q_g$ with $L^{*T} L^* = I$, which is named as OLDA. Compared with ULDA, the redundancy is introduced into the transformed space and the overfitting problem can be effectively avoided in OLDA.

In summary, in the training set, an optimal transformation L^* is learned in which the discriminant vectors are orthogonal to each other, based on the pseudo-inverse LDA. The learned transformation L^* works on the feature vectors in the test set by Eq. (1), resulting in the new discriminative feature vectors. Algorithm 1 formalizes the proposed person re-id method.

Algorithm 1 Learning the Orthogonal Transformation for Person Re-Identification.

Input:

The feature matrix X and its label information in the training set.

The feature vectors $\boldsymbol{x_i^{test}}$ and $\boldsymbol{x_j^{test}}$ in the test set.

Output:

The distance $D(\boldsymbol{x_i^{test}}, \boldsymbol{x_j^{test}})$ between $\boldsymbol{x_i^{test}}$ and $\boldsymbol{x_j^{test}}$.

1: Construct the matrices H_b and H_t by Eq. 2.

2: Compute SVD of H_t as $H_t = U\begin{bmatrix}\Sigma_t & 0\\ 0 & 0\end{bmatrix}V^T$.

3: Denote $T = \Sigma_t^{-1} U_1 H_b$, where U_1 is generated by the first r_t columns of U with $r_t = rank(H_t H_t^T)$, and compute SVD as $T = P\widetilde{\Sigma} Q^T$.

4: Denote $G = U_1 \Sigma_t^{-1} P$.

5: Compute the QR decomposition of G as $G = Q_g R_g$.

6: Denote $L^* = Q_g$, which is the optimal solution of Eq.5.

7: **return** $D(\boldsymbol{x_i^{test}}, \boldsymbol{x_j^{test}}) = \left\| L^{*T} \boldsymbol{x_i^{test}} - L^{*T} \boldsymbol{x_j^{test}} \right\|^2$;

3.4. Kernel version for learning the orthogonal transformation

An optimal linear transformation is learned in Algorithm 1, which however is limited to solve the linear problem. However, in most cases person re-id is a non-linear problem due to the non-linearity in person's appearance. In this section, a non-linear orthogonal discriminant analysis based on the kernel technique is developed, to better fit the person re-id problem.

Let the original feature space be mapped into a new high-dimensional kernel space through a non-linear mapping function Φ: $\boldsymbol{x} \in \mathbb{R}^m \rightarrow \boldsymbol{\Phi}(\boldsymbol{x}) \in \mathbb{R}^r$ ($r \gg m$). The new feature matrix is denoted as $\Phi(X) = [\boldsymbol{\Phi}(\boldsymbol{x_1}), \ \ldots, \boldsymbol{\Phi}(\boldsymbol{x_n})] \in \mathbb{R}^{r\times n}$.

Denote the three scatter matrices in the kernel space by S_w^Φ, S_b^Φ and S_t^Φ, respectively. Then, we can obtain

$$
\begin{aligned}
S_w^\Phi &= \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_i}\left[\boldsymbol{\Phi}(\boldsymbol{x_i^j}) - \boldsymbol{\Phi}(\boldsymbol{c^{(j)}})\right]\left[\boldsymbol{\Phi}(\boldsymbol{x_i^j}) - \boldsymbol{\Phi}(\boldsymbol{c^{(j)}})\right]^T \\
S_b^\Phi &= \frac{1}{n}\sum_{j=1}^{k} n_i\left[\boldsymbol{\Phi}(\boldsymbol{c^{(j)}}) - \boldsymbol{\Phi}(\boldsymbol{c})\right]\left[\boldsymbol{\Phi}(\boldsymbol{c^{(j)}}) - \boldsymbol{\Phi}(\boldsymbol{c})\right]^T \\
S_t^\Phi &= \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{\Phi}(\boldsymbol{x_i}) - \boldsymbol{\Phi}(\boldsymbol{c})][\boldsymbol{\Phi}(\boldsymbol{x_i}) - \boldsymbol{\Phi}(\boldsymbol{c})]^T,
\end{aligned} \tag{19}
$$

where $\boldsymbol{\Phi(c^{(j)})}$ and $\boldsymbol{\Phi(c)}$ are the centroid of the j-th class sample and the global centroid in the kernel space, respectively.

Explicitly computing the mapping $\boldsymbol{\Phi(x)}$ and then performing Algorithm 1 is intractable and computationally expensive. Instead, the data can be implicitly embedded by rewriting the algorithm in terms of dot products, i.e. the kernel trick is replaced by the kernel function $k(\boldsymbol{x_i}, \boldsymbol{x_j}) = \boldsymbol{\Phi}(\boldsymbol{x_i}) \cdot \boldsymbol{\Phi}(\boldsymbol{x_j})$.

First we notice that the discriminant vector $\boldsymbol{l^\Phi}$ satisfies $\boldsymbol{l^\Phi} = \sum_{i=1}^{n}\alpha_i\boldsymbol{\Phi}(\boldsymbol{x_i})$, where the coefficient vector $\boldsymbol{\alpha} = [\alpha_1, \ \ldots, \ \alpha_n]^T$. Then, we have

$$
\begin{aligned}
(\boldsymbol{l^\Phi})^T S_b^\Phi \boldsymbol{l^\Phi} &= \boldsymbol{\alpha}^T[K(B-O)(B-O)^TK]\boldsymbol{\alpha} \\
(\boldsymbol{l^\Phi})^T S_t^\Phi \boldsymbol{l^\Phi} &= \boldsymbol{\alpha}^T[K(I-O)(I-O)^TK]\boldsymbol{\alpha},
\end{aligned} \tag{20}
$$

where $K = \Phi(X)^T\Phi(X)$, $B = diag(B_1, \cdots, B_k) \in \mathbb{R}^{n\times n}$ and $B_i \in \mathbb{R}^{n_i\times n_i}$ ($i = 1, \ldots, k$) with all terms equal to $\frac{1}{n_i}$, $O \in \mathbb{R}^{n\times n}$ with all terms equal to $\frac{1}{n}$ and I is a $n \times n$ identity matrix[1].

Accordingly, the three scatter matrix can be kernelized as

$$
\begin{aligned}
&K_w = K(I-B)(I-B)^TK, \quad K_b = K(B-O)(B-O)^TK, \\
&K_t = K(I-O)(I-O)^TK.
\end{aligned} \tag{21}
$$

We now can rewrite the optimization problem in Eq. (5) as

$$
\mathcal{A}^* = \arg\max_{\mathcal{A}} trace\left\{(\mathcal{A}^TK_t\mathcal{A})^+\mathcal{A}^TK_b\mathcal{A}\right\} \tag{22}
$$

where $\mathcal{A} = [\boldsymbol{\alpha_1}, \ \ldots, \ \boldsymbol{\alpha_d}]$.

Then the d orthogonal discriminant vectors in kernel space are obtained according to Algorithm 1, which need not be repeated here. In particular, in the test phase, the new discriminative feature is obtained as $\boldsymbol{y_i^{test}} = (L^\Phi)^T \cdot \boldsymbol{\Phi}(\boldsymbol{x_i^{test}}) = \mathcal{A}^{*T}k(X, \boldsymbol{x_i^{test}})$.

3.5. Fast version for learning the orthogonal transformation

In this section, we present a fast version for solving the optimization problem in Eq. (5) with QR decomposition replacing eigen-decomposition in the computation.

The feature matrix X is rewritten as $X = [X_1, \ X_2, \ \ldots, \ X_k]$, where $X_i \in \mathbb{R}^{m\times n_i}$ ($i = 1, \ldots, k$) denotes the feature matrix of i-th class sample.

We firstly compute the economy-size QR decomposition of X as $X = Q_xR_x$ with $Q_x \in \mathbb{R}^{m\times n}$ and $R_x \in \mathbb{R}^{n\times n}$. Compared with the full QR decomposition $X = QR$ with $Q \in \mathbb{R}^{m\times m}$ and $R \in \mathbb{R}^{m\times n}$, the economy-size one only computes the first n columns of Q and the first n rows of R.

Denote Z as a permutation matrix generated by exchanging the i-th column and the $(\sum_{j=1}^{i-1} n_j + 1)$-th column of the identity matrix $I \in \mathbb{R}^{n\times n}$, H_i ($i = 1, \ldots, k$) and H as the Householder transformations of vectors $[1, \ \ldots, \ 1]^T \in \mathbb{R}^{n_i}$ and $[\sqrt{n_1}, \ \ldots, \ \sqrt{n_k}]^T$, respectively. Then, we express

$$
R_x\begin{bmatrix} H_1 & & \\ & \ddots & \\ & & H_k \end{bmatrix} Z\begin{bmatrix} H & \\ & I \end{bmatrix} = [R_1, \ R_2, \ R_3], \tag{23}
$$

where $R_1 \in \mathbb{R}^{n\times 1}$, $R_2 \in \mathbb{R}^{n\times(k-1)}$ and $R_3 \in \mathbb{R}^{n\times(n-k)}$.

We compute the rank-revealing QR decomposition (also called economic QR decomposition with column pivoting) of $[R_2, R_3]$ as

$$
[R_2, \ R_3]P_{qr} = \widetilde{Q}\widetilde{R} = \left[\widetilde{Q}_1, \ \widetilde{Q}_2, \ \widetilde{Q}_3\right]\begin{bmatrix} \widetilde{R}_{11} & \widetilde{R}_{12} \\ 0 & \widetilde{R}_{22} \\ 0 & \widetilde{R}_{32} \end{bmatrix}, \tag{24}
$$

where P_{qr} is a permutation matrix, $\widetilde{R}_{11}$ is an upper triangular matrix and $\widetilde{R}_{32} \approx 0$ is numerically negligible. $\widetilde{Q}_1 \in \mathbb{R}^{n\times r_2}$, $\widetilde{Q}_2 \in \mathbb{R}^{n\times r_3}$ and $\widetilde{Q}_3 \in \mathbb{R}^n$ with $r_2 = rank(R_2)$ and $r_3 = rank(R_3)$. $\widetilde{R}_{11} \in \mathbb{R}^{r_2\times(k-1)}$ and $\widetilde{R}_{12} \in \mathbb{R}^{r_2\times(n-k)}$. Let

$$
\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \widetilde{R}_{11} & \widetilde{R}_{12} \\ 0 & \widetilde{R}_{22} \\ 0 & \widetilde{R}_{32} \end{bmatrix} P_{qr}^T. \tag{25}
$$

Then we obtain the QR decomposition of $[R_2, R_3]$.

Accordingly, the QR decomposition of $\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix}R_{22}^T$ is computed as

$$
\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix}R_{22}^T = \bar{Q}\begin{bmatrix} \bar{R} \\ 0 \end{bmatrix}. \tag{26}
$$

The optimization problem of Eq. (5) is solved by $L^* = Q_x\left[\widetilde{Q}_1, \ \widetilde{Q}_2\right]\bar{Q}$. For the detailed proof, the interested reader can refer to [7].

As stated above, QR decomposition is used for solving Eq. (5) instead of eigen-decomposition. Algorithm 2 formalizes the

Algorithm 2 Fast Version for Learning the Orthogonal Transformation.

Input:

The feature matrix X and its label information in the training set.

The feature vectors $\boldsymbol{x_i^{test}}$ and $\boldsymbol{x_j^{test}}$ in the test set.

Output:

The distance $D(\boldsymbol{x_i^{test}}, \boldsymbol{x_j^{test}})$ between $\boldsymbol{x_i^{test}}$ and $\boldsymbol{x_j^{test}}$.

1: Compute the economy-size QR decomposition of X as $X = Q_xR_x$.

2: Construct the matrices Z, H and H_i ($i = 1, \ldots, k$) and compute $[R_1, \ R_2, \ R_3]$ according to Eq.2.

3: Compute the rank-revealing QR decomposition of $[R_2, \ R_3]$ and obtain the corresponding QR decomposition $[R_2, \ R_3] = \left[\widetilde{Q}_1, \ \widetilde{Q}_2, \ \widetilde{Q}_3\right]\begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix}$.

4: Compute the QR decomposition of $\begin{bmatrix} R_{12} \\ R_{22} \end{bmatrix}R_{22}^T = \bar{Q}\begin{bmatrix} \bar{R} \\ 0 \end{bmatrix}$.

5: Denote $L^* = Q_x\left[\widetilde{Q}_1, \ \widetilde{Q}_2\right]\bar{Q}$, which is the optimal solution of Eq.5.

6: **return** $D(\boldsymbol{x_i^{test}}, \boldsymbol{x_j^{test}}) = \left\|L^{*T}\boldsymbol{x_i^{test}} - L^{*T}\boldsymbol{x_j^{test}}\right\|^2$;

[1] Detailed derivation of Eq. (20) is given in Appendix.

fast version for our method, and has lower computational complexity than Algorithm 1. The detailed analysis about the computational complexity will be shown in the experiment section.

4. Experiments

4.1. Datasets and settings

We carry out experiments on four widely used benchmark datasets: **VIPeR** [16], **PRID2011** [17], **CUHK01** [18] and **CUHK03** [19].

VIPeR contains 632 image pairs of pedestrians from two disjoint outdoor cameras, with large variations in viewpoint, illumination and poses. Following the standard settings, we normalize all the images as 128×48 pixels and randomly partition the dataset into training and test sets using the ration 1: 1.

PRID2011 is collected from two disjoint surveillance cameras, containing 385 and 749 identities, respectively, with only 200 identities appearing in both cameras. All the images have been normalized to 128×64 pixels. We use the single shot version of the dataset. For the training set, 100 image pairs are randomly chosen from the 200 identities in both cameras. For the test set, the probe set is formed by the remaining 100 identities appearing in both cameras and the gallery set is formed by the remaining ones and 649 single images. This dataset is very challenging since there are many disturbed images in the gallery set for test data.

CUHK01 is a larger dataset in person re-id, including 3,884 images of 971 identities captured by two disjoint cameras in a campus environment, with each identity having two images under each camera. All the images have been normalized to 160×60 pixels. In the one camera, the side view of identities is captured. And in the another camera, the front or back view is captured. We use the same protocol with [4]: the 485/486 training/test with multi-shot setting.

CUHK03 is one of the largest person re-id datasets. It consists of 13,164 images of 1,360 identities under six surveillance cameras in campus, with each identity captured by two cameras and having 4.8 images at each viewpoint. The dataset provides both manually annotated bounding boxes and bounding boxes detected by the Deformable-Part-Model (DPM) [20], denoted as CUHK03(manual) and CUHK03(detected) dataset, respectively. The latter type of annotation is more approach to the real-world person re-id system. 1,260 identities and 100 ones are used for training and testing under the single-shot setting, respectively. And all the images are normalized as 128×48 pixels.

Evaluation protocol. We use Cumulative Matching Characteristics (CMC) for evaluation where Rank k matching rate is the expectation of correct match at rank k. The splitting of training set and test set is repeated 10 times for all datasets except for CUHK03 dataset, in which the procedure is repeated 20 times as provided in [19]. Rank 1, Rank 5 and Rank 10 in the average CMC curve are reported.

Parameter setting. There is not any parameter required to tuned in our method. However, due to the kernelisation of our method, it is necessary to select the proper parameter for the kernel function. We set these parameters via 2-fold cross validation. Specifically, for each splitting of training set and test set, we randomly select 90% samples in the training set as the new training set and the other 10% in the training set are used as the validation set. Moreover, there are PCA technology involved for the comparison purpose in the experiment and we perform PCA with holding 90% energy.

Feature descriptor. In order to perform an overall evaluation for our method, we use three feature descriptors in the experiment: Histogram of Intensity Pattern & Histogram of Ordinal Pattern (HIPHOP) [21], Local Maximal Occurrence (LOMO) [2] and Gaussian Of Gaussian (GOG) [1]. The HIPHOP descriptor[2] is deep person appearance representation by exploit the AlexNet [22]. The LOMO descriptor analyzes the horizontal occurrence of local features with robustness to viewpoint changes. The GOG descriptor represents the pedestrian by hierarchical Gaussian distribution with both means and covariances, and is extracted from the alternative color channels as GOG_{RGB}, GOG_{Lab}, GOG_{HSV} and GOG_{RnG}. Fig. 1 summarizes the statistics of dimensionality of feature and sample size. For all feature descriptors mentioned above on all datasets, the dimension of feature vector is greater than the sample size, and the singularity problem does exist in person re-id. In order to further verify the performance of our method in the case that there is no singularity problem in person re-id, and the performance of NLDA in low-dimensional data, PCA dimensionality reduction is performed in GOG feature descriptor, denoted as GOG_{PCA}, also used as the feature descriptor and compared in the experiment. We can see from Fig. 1 that the dimension of feature vector is less than the sample size only for GOG_{PCA} on all datasets, representing the non-singularity case; while the rest resulting in the singularity problem usually met in person re-id context.

4.2. Performance comparison

4.2.1. Comparison with variants of LDA

We compare ours with various variants of LDA on VIPeR, PRID2011, CUHK01 and CUHK03(manual) datasets. The compared methods include PCA+LDA [8], RLDA [9], ULDA [6] and NLDA [10]. For an overall comparison, HIPHOP, LOMO, GOG and GOG_{RGB} feature descriptors are used to represent the pedestrian in the experiment, respectively. In these cases, person re-id is a singularity problem. Moreover, we also use GOG_{PCA} as the feature descriptor in the experiment, to demonstrate the non-singularity case for NLDA in person re-id.

Fig. 2 and Table 3 summarize the comparison results. We can see that: **1)** Our method beats PCA+LDA and ULDA for all feature descriptors on all four datasets. It shows that the ability of our method in solving person re-id is superior to these two methods. **2)** Our method is obvious better than RLDA in most cases except for low-dimensional data such as GOG_{RGB}. Nevertheless, in RLDA, the parameter has to be tuned carefully to obtain the highest results. Conversely, there is not any parameter required to tuned in our method. **3)** NLDA achieves almost the same performance with our method in the vast majority of cases. However, for low-dimensional data such as GOG_{PCA}, the performance of NLDA is inferior to our method. In particular, NLDA performs poorly in the large dataset CUHK01 and CUHK03(manual) with 0.2% and 1.0% at Rank 1, respectively. It verifies the NLDA's inapplicability to low-dimensional data. **4)** Our method achieves the best performance in three datasets VIPeR, CUHK01 and CUHK03(manual) for GOG_{PCA} descriptor, which indicates the advantage of our method in the non-singularity case in person re-id. **5)** The performance of our method is variable for different features. This is reasonable because different features have different discrimination properties, and as a result the performance of the method based on these features may be affected.

4.2.2. Comparison with kernel version

We investigate the performances of different kernel functions applying to our method on VIPeR, PRID2011, CUHK01 and CUHK03(manual) datasets. HIPHOP, LOMO and GOG feature descriptors are used to represent the pedestrian in the experiment,

[2] Following the standard protocol [21], all the images are normalized to 227×227 pixels when extracting HIPHOP descriptor in the experiment.

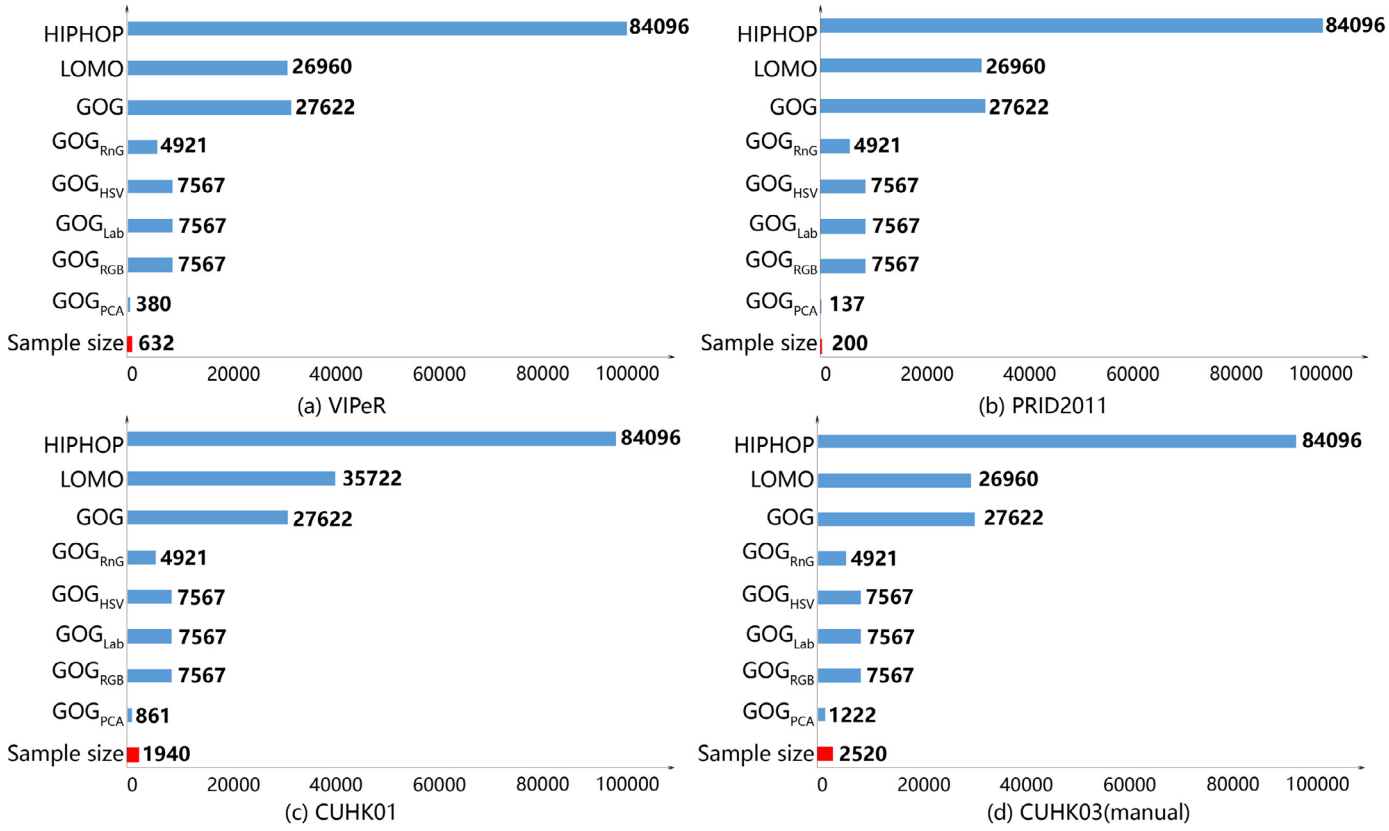


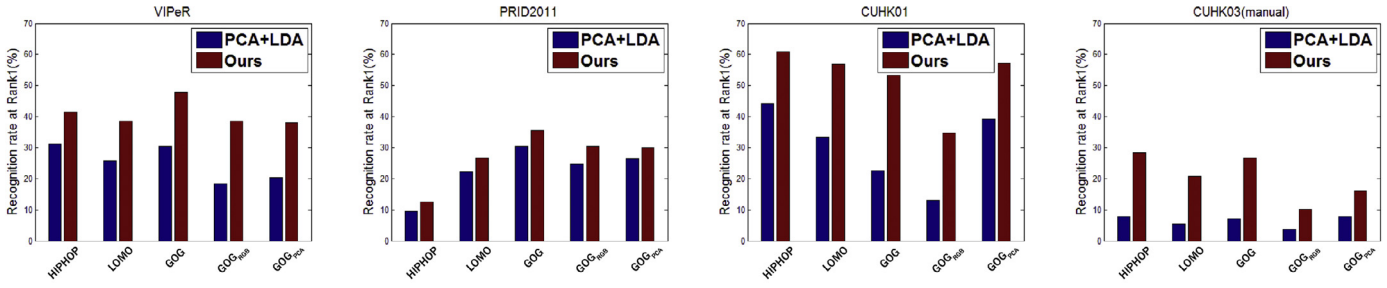
Fig. 1. The statistics of dimensionality of feature and sample size on four datasets. Better viewed in colour.

Table 3
Comparison with various variants of LDA. The best results are shown in boldface.

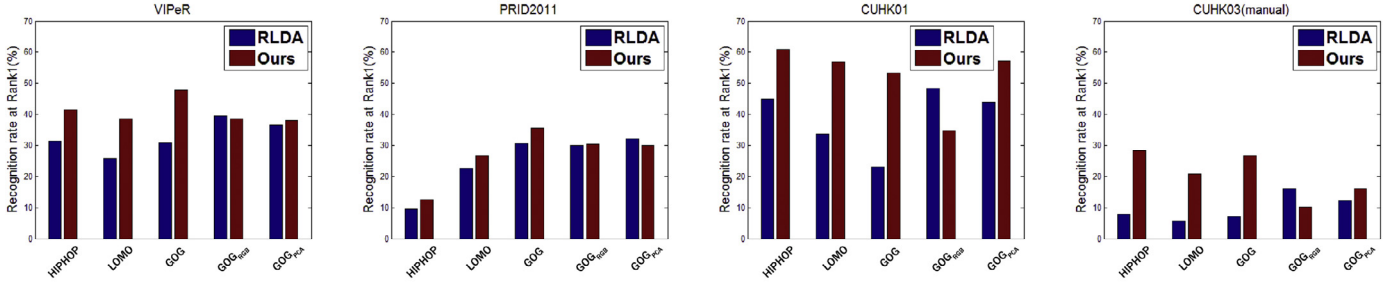
	Methods	VIPeR			PRID2011			CUHK01			CUHK03(manual)		
		$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
HIPHOP	PCA+LDA	31.3	54.9	65.7	9.6	23.2	30.0	44.1	61.9	69.5	7.9	23.0	32.0
	RLDA	31.4	55.0	65.8	9.5	23.3	30.8	44.9	62.9	70.5	7.8	22.9	32.2
	ULDA	32.5	56.6	67.1	9.9	24.9	32.5	43.1	61.0	69.0	9.1	22.5	33.1
	NLDA	**41.5**	69.7	80.6	**12.5**	27.9	38.4	60.7	81.2	87.0	**28.3**	52.6	64.0
	Ours	**41.5**	69.7	80.6	**12.5**	27.9	38.4	**60.9**	81.1	87.1	**28.3**	52.6	64.0
LOMO	PCA+LDA	25.7	50.0	61.2	22.3	41.8	50.3	33.2	51.2	59.8	5.4	17.4	28.9
	RLDA	25.7	50.4	61.8	22.6	41.3	50.7	33.7	52.0	60.5	5.6	17.2	29.2
	ULDA	25.8	49.5	61.0	23.4	41.4	50.7	29.2	48.2	56.1	4.9	16.0	26.1
	NLDA	**38.5**	68.3	79.3	**26.9**	50.0	61.7	**56.9**	78.0	83.7	**20.9**	47.7	61.9
	Ours	**38.5**	68.3	79.3	26.7	50.2	61.7	56.8	78.1	83.8	20.8	47.7	61.8
GOG	PCA+LDA	30.4	55.0	64.7	30.3	50.3	60.5	22.6	39.6	49.1	7.1	17.2	26.9
	RLDA	30.9	55.8	65.2	30.5	50.3	60.5	23.1	39.9	49.8	7.1	17.3	27.0
	ULDA	31.1	55.3	65.7	30.0	50.7	59.9	19.1	34.3	43.9	6.6	16.6	24.9
	NLDA	**47.8**	78.6	88.4	**35.6**	57.8	69.0	53.2	74.3	81.6	**26.7**	54.7	69.8
	Ours	**47.8**	78.6	88.4	**35.6**	57.8	69.0	**53.3**	74.5	81.8	**26.7**	54.7	69.8
GOG_{RGB}	PCA+LDA	18.3	37.8	48.5	24.7	45.6	55.4	12.9	25.8	34.4	3.8	11.4	19.4
	RLDA	**39.4**	72.1	83.4	30.0	51.7	61.8	**48.4**	70.6	78.9	**16.0**	35.8	50.5
	ULDA	18.4	36.8	46.5	26.5	43.8	53.3	7.7	17.2	24.7	2.5	9.8	16.5
	NLDA	38.6	70.7	82.1	**30.3**	52.3	61.9	34.5	56.5	66.6	10.2	27.1	38.8
	Ours	38.6	70.7	82.1	**30.3**	52.3	61.9	34.6	56.9	66.7	10.2	27.1	38.8
GOG_{PCA}	PCA+LDA	20.3	40.3	50.5	26.5	46.1	56.1	39.2	60.9	70.4	7.9	19.8	30.5
	RLDA	36.5	67.7	80.1	**32.0**	55.4	64.9	44.0	65.5	73.9	12.2	29.0	40.4
	ULDA	24.2	44.7	54.7	27.1	45.6	56.5	22.8	38.9	48.5	6.4	15.6	25.0
	NLDA	33.7	64.6	77.4	25.9	48.8	60.1	0.2	1.0	2.1	1.0	5.0	10.0
	Ours	**37.9**	68.8	80.5	30.0	49.9	61.7	**57.2**	78.1	85.1	**16.0**	39.9	53.4

respectively. Moreover, for verifying our method's ability of score-level fusion with different feature descriptors, we present the result of our method with fusion pattern in which we add up the distance values obtained by three different feature descriptors HIPHOP, LOMO and GOG as the final distance value. The related kernel functions are introduced as follows:

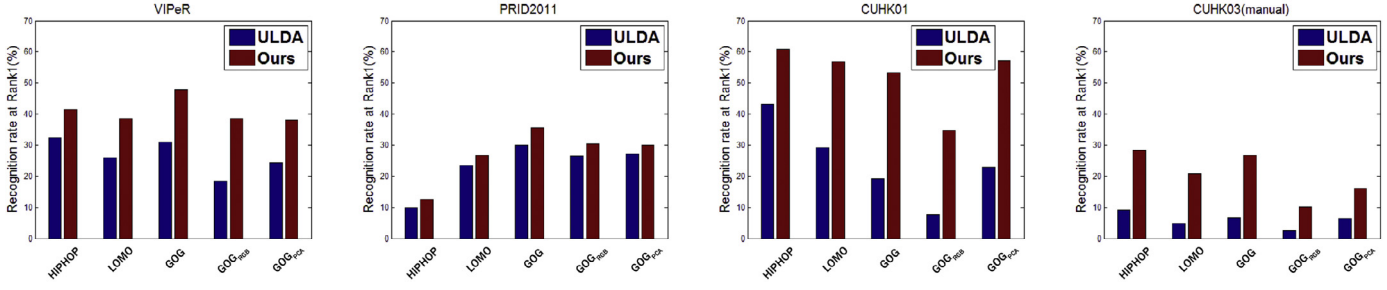
- Gaussian: A classical radial basis kernel with good anti-interference for the noise in the data and the mathematical form $k(\boldsymbol{x},\boldsymbol{y}) = e^{-\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{2\sigma^2}}$.
- Sigmoid: An "S" shape kernel function widely applied in the field of deep learning now, with the mathematical form $k(\boldsymbol{x},\boldsymbol{y}) = \tanh(\alpha\langle\boldsymbol{x},\boldsymbol{y}\rangle + \beta)$.



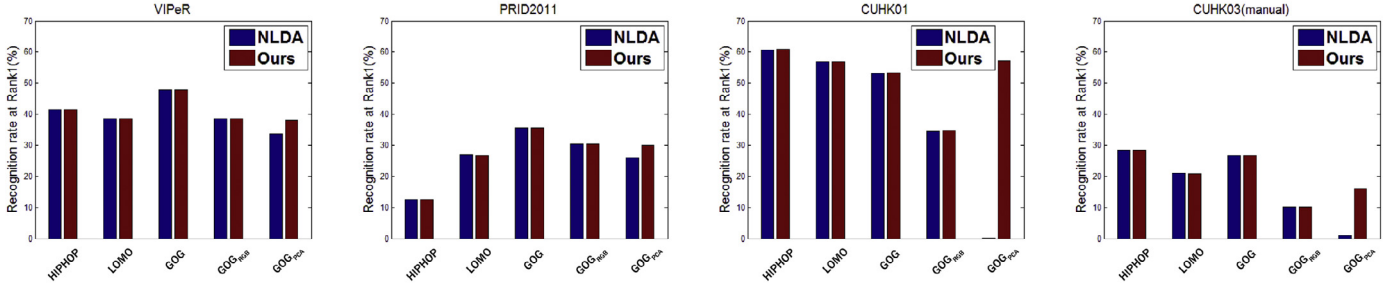
(a) comparison with PCA+LDA



(b) comparison with RLDA



(c) comparison with ULDA



(d) comparison with NLDA

Fig. 2. Comparison of Rank 1 with various variants of LDA. Better viewed in colour.

- Cauchy: a kernel function coming from Cauchy distribution that can be applied in very high-dimensional data due to the wide domain of definition, with the mathematical form $k(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\|\boldsymbol{x}-\boldsymbol{y}\|^2/\sigma+1}$.
- ANOVA: A radial basis kernel, with the mathematical form $k(\boldsymbol{x},\boldsymbol{y}) = e^{(-\sigma(\boldsymbol{x}-\boldsymbol{y})^2)^\gamma}$.
- Multiquadric (MQ): A rational quadratic kernel, with the mathematical form $k(\boldsymbol{x},\boldsymbol{y}) = \sqrt{\|\boldsymbol{x}-\boldsymbol{y}\|^2 + c^2}$.
- Inverse Multiquadric (IM): An inverse multi quadric kernel in which the kernel matrix is full rank, with the mathematical form $k(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{\sqrt{\|\boldsymbol{x}-\boldsymbol{y}\|^2+c^2}}$.

The results are given in Table 4. We have the following observation: **1)** The performance can be boosted by introducing kernelisation into our method in most cases. Among them, our method with Gaussian kernel function and fusion pattern achieves the largest increase 33.7% at Rank 1 on CUHK03(manual) dataset. However, on

Table 4
Comparison of our methods for non-kernel version and kernel version with different kernel functions.

	Methods		VIPeR			PRID2011			CUHK01			CUHK03(manual)		
			$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
GOG	Ours(w/o)		47.8	78.6	88.4	35.6	57.8	69.0	53.3	74.5	81.8	26.7	54.7	69.8
	Ours(kernel)	Gaussian	49.1	80.6	89.3	33.3	57.9	67.1	70.8	88.3	92.8	59.8	87.7	94.5
		Sigmoid	47.2	80.2	88.9	34.9	58.4	67.7	70.8	88.4	92.9	58.0	86.6	94.0
		Cauchy	49.2	80.8	89.4	32.7	56.7	66.2	71.3	88.7	93.0	58.8	86.9	94.8
		ANOVA	48.6	79.9	89.2	34.7	58.0	67.7	71.1	88.6	92.9	57.8	86.6	93.8
		MQ	48.9	80.1	89.3	33.4	57.8	67.7	71.2	88.6	93.0	59.2	87.1	94.4
		IM	48.8	80.0	89.4	33.3	58.1	67.1	71.2	88.8	93.0	59.4	86.9	94.8
LOMO	Ours(w/o)		38.5	68.3	79.3	26.7	50.2	61.7	56.8	78.1	83.8	20.8	47.7	61.8
	Ours(kernel)	Gaussian	41.7	71.2	83.3	27.3	50.2	62.0	68.0	88.1	93.1	48.4	80.4	89.6
		Sigmoid	42.5	73.5	85.2	27.2	50.4	61.1	68.1	88.7	93.3	49.3	81.7	91.6
		Cauchy	41.4	71.1	83.3	26.6	49.7	61.5	67.6	87.7	92.7	47.9	80.2	90.2
		ANOVA	41.5	71.4	83.5	27.2	50.3	61.4	67.9	87.8	92.7	49.3	81.8	91.4
		MQ	41.1	70.8	82.6	26.9	50.6	61.8	66.8	86.3	91.5	45.2	77.5	87.3
		IM	41.5	71.1	82.7	27.2	50.4	61.1	67.7	87.5	92.4	49.3	80.8	91.1
HIPHOP	Ours(w/o)		41.5	69.7	80.6	12.5	27.9	38.4	60.9	81.1	87.1	28.3	52.6	64.0
	Ours(kernel)	Gaussian	44.2	73.2	83.3	12.3	27.4	37.1	65.4	84.9	90.8	44.8	75.8	86.1
		Sigmoid	44.4	72.9	82.9	13.0	29.3	38.6	64.9	83.9	90.0	43.5	72.8	83.5
		Cauchy	44.2	73.3	83.1	13.0	28.3	37.4	65.3	84.7	90.6	43.0	75.7	86.9
		ANOVA	44.0	73.4	83.2	13.6	28.6	37.8	65.0	84.1	90.2	46.2	76.8	85.6
		MQ	44.1	72.8	82.9	12.7	28.5	37.1	64.9	84.0	90.1	44.2	74.6	84.4
		IM	44.2	73.0	83.0	12.7	28.5	37.1	63.8	85.0	90.7	46.2	78.2	86.4
Fusion	Ours(w/o)		52.0	82.0	90.5	40.6	65.6	76.7	55.0	81.6	90.3	34.8	70.4	82.2
	Ours(kernel)	Gaussian	52.8	82.8	91.3	37.8	63.2	73.4	79.7	94.7	97.3	68.5	92.7	97.0
		Sigmoid	44.2	74.6	85.9	36.0	59.0	68.6	76.1	93.0	96.8	57.8	87.1	94.2
		Cauchy	52.9	82.0	91.2	36.9	62.3	73.0	79.8	94.7	97.6	64.4	91.0	95.9
		ANOVA	50.9	80.4	89.6	35.5	60.6	69.2	80.0	94.5	97.4	66.4	91.7	85.6
		MQ	53.6	82.7	91.8	38.3	63.1	73.1	78.1	93.6	97.0	65.1	89.6	95.7
		IM	49.3	77.8	87.4	32.8	57.6	67.1	73.4	91.3	95.3	62.1	88.2	94.2

VIPeR and PRID2011 datesets, there is the case of performance reduction when introducing kernelisation into our method, for example, our method with Sigmoid kernel function and GOG feature descriptor decreases the Rank 1 by 0.7% on PRID2011 dataset. It might be due to the influence of different description for non-linearity on datasets of different sizes. For large dataset, there are very large number of pedestrians resulting in a more accurate description for non-linearity in person re-id, so that the performance is improved by learning model in kernel space; in contrast, the description for non-linearity may be inaccurate on the small dataset, so that kernelisation has no effect on performance. **2)** Our method with each kernel function is roughly equal on performance for GOG, LOMO and HIPHOP feature descriptors on all datasets, indicating the robustness of the proposed method against kernel functions. **3)** Compared with our method with single feature descriptor, the one with the fusion of these descriptors can significantly improve performance. The increases are 4.4%, 5.0%, 8.7% and 8.0% for the best results from single descriptor to the fusion of descriptors on VIPeR, PRID2011, CUHK01, CUHK03(manual) datasets, respectively.

4.2.3. Comparison with state-of-the-art methods

We compare the performance of ours against state-of-the-art person re-id methods. For comparison against metric learning based person re-id methods, since most of methods with LOMO feature are reported in papers, the same feature descriptor LOMO is used in the proposed method for a fair comparison. For comparison against other state-of-the-art methods, the results of our method with fusion pattern (ours(fusion) for short in the following) are presented. Gaussian kernel is utilized for our method in all the comparison experiments[3]

[3] By 2-fold cross validation, the parameter σ in Gaussian kernel is set as 1 for VIPeR dataset, 2 for PRID2011 dataset and 0.5 for CUHK01 and CUHK03 datasets, respectively.

Table 5
Comparison with state-of-the-art methods on VIPeR dataset. The best results are shown in boldface.

	Methods	$r=1$	$r=5$	$r=10$
LOMO	ITML [23]	24.7	49.8	63.0
	LMNN [24]	29.4	59.8	73.5
	KISSME [25]	34.8	60.4	77.2
	kCCA [26]	30.2	62.7	76.0
	MFA [27]	38.7	69.2	80.5
	kLFDA [27]	38.6	69.2	80.4
	XQDA [2]	40.0	68.1	80.5
	MLAPG [28]	40.7	–	82.3
	CRAFT [21]	**42.3**	**74.7**	**86.5**
	Ours	41.7	71.2	83.3
	DMLLV [29]	50.4	80.5	88.7
	PML&LSL [30]	46.5	69.3	80.7
	PDC [31]	51.3	74.1	84.2
	MuDeep [14]	43.0	74.4	85.8
	DM^3 [32]	42.7	74.3	85.1
	CSPL [33]	51.3	81.7	90.2
	Ours(Fusion)	**52.8**	**82.8**	**91.3**

VIPeR. For VIPeR dataset, we first compare our method with the classical and state-of-the-art metric learning based methods. As shown in Table 5, our method obtains the second best accuracy and performs slightly worse than CRAFT [21] approach that focuses on learning view-specific feature transformations. However, in comparison to other state-of-the-art methods, our method achieves the best accuracy at Rank 1, Rank 5 and Rank 10.

PRID2011. Similarly, we compared our method to metric learning based person re-id methods and other state-of-the-art methods, respectively, on PRID2011 dataset. The results are reported in Table 6. Our method performs the best among all compared methods. Specifically, our method achieves 15.8% gain at Rank 1 com-

Table 6
Comparison with state-of-the-art methods on PRID2011 dataset. The best results are shown in boldface.

	Methods	$r=1$	$r=5$	$r=10$
LOMO	kCCA [26]	14.3	37.4	47.6
	MFA [27]	22.3	45.6	57.2
	kLFDA [27]	22.4	46.5	58.1
	XQDA [2]	26.7	49.9	61.9
	Ours	**27.3**	**50.2**	**62.0**
	Metric Ensembles [35]	17.9	39.0	50.0
	M3TCP [34]	22.0	–	47.0
	CrowdPSE [36]	21.1	46.7	60.0
	DMLLV [29]	27.8	48.4	59.5
	Ours(Fusion)	**37.8**	**63.2**	**73.4**

Table 7
Comparison with state-of-the-art methods on CUHK01 dataset. The best results are shown in boldface.

	Methods	$r=1$	$r=5$	$r=10$
LOMO	kCCA [26]	56.3	80.7	87.9
	MFA [27]	54.8	80.1	87.3
	kLFDA [27]	54.6	80.5	86.9
	XQDA [2]	63.2	83.9	90.0
	MLAPG [28]	64.2	–	90.8
	CRAFT [21]	65.4	85.3	90.5
	Ours	**68.0**	**88.1**	**93.1**
	Metric Ensembles [35]	53.4	76.4	84.4
	M3TCP [34]	53.7	84.3	91.0
	DMLLV [29]	65.0	85.6	91.1
	PML&LSL [30]	53.5	82.5	91.2
	DM^3 [32]	49.7	77.3	86.1
	CSPL [33]	72.0	88.6	92.8
	SGLE [11]	70.9	89.8	93.5
	Ours(Fusion)	**79.7**	**94.7**	**97.3**

pared with M3TCP[4] [34] that learns both the global full-body and local body-parts features via CNN. It verifies the fact that currently the deep learning based method cannot work well on small dataset for person re-id.

CUHK01. Compared with the above two datasets, there are more number of identities on CUHK01 dataset. Nevertheless, the singularity problem still exists in the dataset (refer to Fig. 1). We compare the classical and state-of-the-art results shown in Table 7. The results show clearly that both for metric learning based methods and state-of-the-art methods, our method achieves the best results at Rank 1, Rank 5 and Rank 10. Specifically, our method outperforms deep learning based method M3TCP [34] with a large margin of 26.0% at Rank 1.

CUHK03. It is one of the largest datasets in person re-id but there still exists the singularity problem in this dataset (refer to Fig. 1). We conduct experiments on both CUHK03(manual) and CUHK03(detected) datasets, with comparison to several traditional and deep learning based state-of-the-art methods. The experimental results are shown in Table 8. Compared with the traditional methods, our method achieves the best accuracy at all Rank on CUHK03(manual) dataset and the best accuracy at Rank 10 on CUHK03(detected) datasets. However, deep learning based methods show better performance than traditional methods. In deep learning models, almost ten millions of parameters are learned and more information from data can be obtained so that the model has better generalization ability. Without doubt, it is coming from a premise that the model is trained on the large dataset. In addition, various hyper-parameters such as batch size and learning rate require to be set. Compared with deep learning based methods, our method is much simpler, fewer number of parameters are learned and there is only one kernel hyper-parameter, meanwhile the accuracy at Rank 10 can achieve 97% that only decreases by 1.4% compared to the best result from MuDeep [14] on CUHK03(manual) dataset.

[4] In M3TCP, more than one million of parameters are learned and three hype-parameters require to be set, in contrast, there is only about 0.06 million parameters for learning and one kernel parameter for setting in the proposed method with fusion pattern.

Table 8
Comparison with state-of-the-art methods on CUHK03 dataset. The best results are shown in boldface.

	Methods	CUHK03(manual)			CUHK03(detected)		
		$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
Deep learning	IDLA [37]	54.7	87.6	94.0	54.7	86.5	93.9
	EDM [13]	61.3	88.9	96.4	52.1	82.9	91.8
	LSTM S-CNN [38]	–	–	–	57.3	80.1	88.3
	MSCAN [39]	74.2	94.3	97.5	68.0	91.0	95.4
	DSPL [12]	73.16	92.26	96.54	–	–	–
	MuDeep [14]	**76.9**	**96.1**	**98.4**	**75.6**	**94.4**	**97.5**
Traditional	kLFDA [27]	48.2	59.3	66.4	–	–	–
	XQDA [2]	52.2	82.2	92.1	46.3	78.9	88.6
	MLAPG [28]	58.0	87.1	94.7	51.2	83.6	92.1
	SSSVM [40]	57.0	84.4	90.9	–	–	–
	GOG [1]	67.3	91.0	96.0	**65.5**	**88.4**	93.7
	Ours	**68.5**	**92.7**	**97.0**	60.8	86.6	**94.2**

Table 9
Comparison with fast version for our method in computational complexity.

	Algorithm1	Algorithm2
Step1	$O(mn)$	$4mn^2-\frac{4}{3}n^3$
Step2	$14mn^2-2n^3$	$O(n^2)$
Step3	$2mr_tk+14r_tk^2-2k^3$	$2nr_t^2+\frac{2}{3}r_t^3$
Step4	$2mr_t^2$	$2r_tk(n-k)+4kr_t^2+\frac{2}{3}k^3-2r_tk^2$
Step5	$4mr_b^2-\frac{4}{3}r_b^3$	$2(m+r_t)nr_b$
Overall	$18mn^2+n^3$	$5mn^2+6n^3$

Table 10
Comparison with fast version for our method in accuracy and run time.

Datasets	Methods	$r=1$	$r=5$	$r=10$	Time(s)
VIPeR	Ours	47.8	78.6	88.4	2.1952
	$Ours_{fast}$	47.9	78.6	88.3	1.2910
PRID2011	Ours	35.6	57.8	69.0	0.4919
	$Ours_{fast}$	35.4	57.7	69.2	0.2484
CUHK01	Ours	53.3	74.5	81.8	13.4156
	$Ours_{fast}$	53.3	74.5	81.8	7.8118
CUHK03(manual)	Ours	26.7	54.7	69.8	25.4500
	$Ours_{fast}$	26.6	54.8	69.6	12.7273

4.2.4. Comparison with fast version

A fast version for our method is introduced in Section 3.5. In this section, we analyze its computational complexity, accuracy and run time. Comparison in the computational complexity[5] is given in Table 9. Ignoring low order terms, we compute the overall computational complexity of normal version and fast version about $18mn^2+n^3$ and $5mn^2+6n^3$, respectively. Compared with the normal version, the computation complexity of the fast version is reduced by 50%. Table 10 presents the comparison in the accuracy and run time. The experiments are conducted with GOG feature descriptor on a PC with Intel Core CPU (4GHz × 8) and 16GB RAM. It can be seen that the fast version has the same accuracy with a halving of run time.

5. Conclusion

In this paper, a pseudo-inverse LDA based person re-id method is proposed to solve the singularity problem in person re-id by

[5] The interested reader can refer to [7] for more detailed theoretical analysis.

computing an optimal orthogonal transformation based on the simultaneous diagonalization of three scatter matrices. Furthermore, in consideration of the non-linearity in person re-id, we develop a non-linear model by learning the orthogonal transformation in kernel space, thus improve the effectiveness of the proposed method; we also develop a fast version for learning the transformation aiming to improve the efficiency of the proposed method. A closed-form solution is provided in the proposed method in which there is no parameters required to tune, so the proposed method has advantages of simplicity and high-efficiency. Extensive experiments on four datasets show that the proposed method outperforms the state-of-the-art methods. We also conduct the comparison experiment with kernel version and fast version for the proposed method, validating the superiority of these versions.

The singularity problem commonly exists in image classification and retrieval. Since the feature vectors extracted from the image are generally high dimensional for more accurately representing appearance of content in image, and the number of training samples is limited by the time-consuming of sample labeling and is generally far less than the dimension of feature vector, which is very much similar to the situation in person re-id. The proposed method focusing on solving the singularity problem in person re-id could give some inspiring ideas about solving this common problem existing in image classification and retrieval. By means of labeled samples, the optimal orthogonal transformation is learned in the proposed method. However, in some cases, there are very small numbers of labeled samples and large numbers of unlabeled samples in reality. In view of this, in the future work, we will focus on the semi-supervised way to solve the singularity problem in person re-id. For example, we can employ label propagation technology to propagate the label information from labeled samples to unlabeled samples and further develop the proposed method into a semi-supervised method.

Acknowledgement

This work is supported by the National Key R&D Program of China under Grant 2017YFC0803505.

Appendix

For Eq. (20), we now present a detail derivation. According to Eq. (19), we have

$$\begin{aligned}(\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_w^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} &= (\boldsymbol{l}^{\boldsymbol{\Phi}})^T \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left[\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{j}}) - \boldsymbol{\Phi}(\boldsymbol{c}^{(j)})\right] \\ &\quad \times \left[\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{j}}) - \boldsymbol{\Phi}(\boldsymbol{c}^{(j)})\right]^T \boldsymbol{l}^{\boldsymbol{\Phi}}, \\ (\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_t^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} &= (\boldsymbol{l}^{\boldsymbol{\Phi}})^T \sum_{i=1}^{n}[\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}) - \boldsymbol{\Phi}(\boldsymbol{c})][\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}) - \boldsymbol{\Phi}(\boldsymbol{c})]^T \boldsymbol{l}^{\boldsymbol{\Phi}}. \end{aligned} \tag{27}$$

We feed $\boldsymbol{l}^{\boldsymbol{\Phi}} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}})$ into the above equation and have

$$\begin{aligned}(\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_w^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} &= \boldsymbol{\alpha}^T \Phi(X)^T \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left[\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{j}}) - \boldsymbol{\Phi}(\boldsymbol{c}^{(j)})\right] \\ &\quad \times \left[\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{j}}) - \boldsymbol{\Phi}(\boldsymbol{c}^{(j)})\right]^T \Phi(X)\boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^T \Gamma_w^{\Phi} \Gamma_w^{\Phi^T} \boldsymbol{\alpha}, \\ (\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_t^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} &= \boldsymbol{\alpha}^T \Phi(X)^T \sum_{i=1}^{n}[\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}) - \boldsymbol{\Phi}(\boldsymbol{c})][\boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}) - \boldsymbol{\Phi}(\boldsymbol{c})]^T \Phi(X)\boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}^T \Gamma_t^{\Phi} \Gamma_t^{\Phi^T} \boldsymbol{\alpha}, \end{aligned} \tag{28}$$

where

$$\begin{aligned}\Gamma_w^{\Phi} &= \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left[\Phi(X)^T \boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{j}}) - \frac{1}{n_i}\sum\nolimits_{q=1}^{n_i} \Phi(X)^T \boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{q}}^{\boldsymbol{j}})\right] \\ &= \sum_{j=1}^{k}\sum_{i=1}^{n_i}\left[k(X, \boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{j}}) - \frac{1}{n_i}\sum_{q=1}^{n_i} k(X, \boldsymbol{x}_{\boldsymbol{q}}^{\boldsymbol{j}})\right] = KI - KB, \\ \Gamma_t^{\Phi} &= \sum_{i=1}^{n}\left[\Phi(X)^T \boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{i}}) - \frac{1}{n}\sum_{q=1}^{n} \Phi(X)^T \boldsymbol{\Phi}(\boldsymbol{x}_{\boldsymbol{q}})\right] \\ &= \sum_{i=1}^{n}\left[k(X, \boldsymbol{x}_{\boldsymbol{i}}) - \frac{1}{n}\sum_{q=1}^{n} k(X, \boldsymbol{x}_{\boldsymbol{q}})\right] \\ &= KI - KO. \end{aligned} \tag{29}$$

and $I \in \mathbb{R}^{n\times n}$ is a identity matrix, $B = diag(B_1, \cdots, B_k) \in \mathbb{R}^{n\times n}$ for which $B_i \in \mathbb{R}^{n_i \times n_i}$ with all terms equal to $\frac{1}{n_i}$, and $O \in \mathbb{R}^{n\times n}$ with all terms equal to $\frac{1}{n}$.

Then, we have

$$\begin{aligned}(\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_w^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} &= \boldsymbol{\alpha}^T [K(I-B)(I-B)^T K]\boldsymbol{\alpha}, \\ (\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_t^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} &= \boldsymbol{\alpha}^T [K(I-O)(I-O)^T K]\boldsymbol{\alpha}. \end{aligned} \tag{30}$$

Since $S_t^{\Phi} = S_b^{\Phi} + S_w^{\Phi}$, it follows that

$$(\boldsymbol{l}^{\boldsymbol{\Phi}})^T S_b^{\Phi} \boldsymbol{l}^{\boldsymbol{\Phi}} = \boldsymbol{\alpha}^T [K(B-O)(B-O)^T K]\boldsymbol{\alpha}. \tag{31}$$

References

[1] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical gaussian descriptor for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1363–1372.
[2] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
[3] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: Computer Vision and Pattern Recognition, 2013, pp. 3318–3325.
[4] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: Computer Vision and Pattern Recognition, 2016, pp. 1239–1248.
[5] J. Ye, R. Janardan, C.H. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, IEEE Trans. Pattern Anal. Mach.Intell. 26 (8) (2004) 982–994.
[6] J. Ye, Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems., J. Mach. Learn. Res. 6 (1) (2005) 483–502.
[7] D. Chu, S.T. Goh, A new and fast orthogonal linear discriminant analysis on undersampled problems, SIAM J. Sci. Comput. 32 (4) (2010) 2274–2297.
[8] P. Howland, H. Park, Generalizing discriminant analysis using the generalized singular value decomposition, IEEE Trans. Pattern Anal. Mach.Intell. 26 (8) (2004) 995–1006.
[9] D.Q. Dai, P.C. Yuen, Regularized discriminant analysis and its application to face recognition, Pattern Recognit. 36 (3) (2003) 845–847.
[10] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, G.J. Yu, A new lda-based face recognition system which can solve the small sample size problem, Pattern Recognit. 33 (10) (2000) 1713–1726.
[11] D. Cheng, Y. Gong, X. Chang, W. Shi, A. Hauptmann, N. Zheng, Deep feature learning via structured graph laplacian embedding for person re-identification, Pattern Recognit. 82 (2018) 94–104.
[12] S. Zhou, J. Wang, D. Meng, X. Xin, Y. Li, Y. Gong, N. Zheng, Deep self-paced learning for person re-identification, Pattern Recognit. 76 (2018) 739–751.
[13] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, S.Z. Li, Embedding deep metric for person re-identification: astudy against large variations, in: European Conference on Computer Vision, 2016, pp. 732–748.
[14] X. Qian, Y. Fu, Y.-G. Jiang, T. Xiang, X. Xue, Multi-scale deep learning architectures for person re-identification, in: IEEE International Conference on Computer Vision, 2017.
[15] A. Edelman, T.A. Arias, S.T. Smith, The geometry of algorithms with orthogonality constraints, in: SIAM J. MATRIX ANAL. APPL, 1998, pp. 303–353.
[16] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), 2007.
[17] M. Hirzer, C. Beleznai, P.M. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, in: Scandinavian Conference on Image Analysis, 2011, pp. 91–102.
[18] W. Li, R. Zhao, X. Wang, Human reidentification with transferred metric learning, in: Asian Conference on Computer Vision, 2012, pp. 31–44.
[19] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: deep filter pairing neural network for person re-identification, in: Computer Vision and Pattern Recognition, 2014, pp. 152–159.

[20] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans.Pattern Anal.Mach.Intell. 32 (9) (2010) 1627–1645.
[21] Y.-C. Chen, X. Zhu, W.-S. Zheng, J.-H. Lai, Person re-identification by camera correlation aware feature augmentation, IEEE Trans.Pattern Anal.Mach.Intell. 40 (2) (2018) 392–408.
[22] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: International Conference on Neural Information Processing Systems, 2012, pp. 1097–1105.
[23] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: International Conference on Machine Learning, 2007, pp. 209–216.
[24] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (1) (2009) 207–244.
[25] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2288–2295.
[26] G. Lisanti, I. Masi, A. Del Bimbo, Matching people across camera views using kernel canonical correlation analysis, in: Proceedings of the International Conference on Distributed Smart Cameras, ACM, 2014, p. 10.
[27] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: European Conference on Computer Vision, 2014, pp. 1–16.
[28] S. Liao, S.Z. Li, Efficient PSD constrained asymmetric metric learning for person re-identification, in: IEEE International Conference on Computer Vision, 2015, pp. 3685–3693.
[29] C. Sun, D. Wang, H. Lu, Person re-identification via distance metric learning with latent variables, IEEE Trans. Image Process. 26 (1) (2017) 23–34.
[30] Z. Zhao, B. Zhao, F. Su, Person re-identification via integrating patch-based metric learning and local salience learning, Pattern Recognit. 75 (2018) 90–98.
[31] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, Q. Tian, Pose-driven deep convolutional model for person re-identification, in: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, 2017, pp. 3980–3989.
[32] Z. Wang, R. Hu, C. Chen, Y. Yu, J. Jiang, C. Liang, S. Satoh, Person reidentification via discrepancy matrix and matrix metric, IEEE Trans. Cybernet. PP (99) (2017) 1–15.
[33] J. Dai, Y. Zhang, H. Lu, H. Wang, Cross-view semantic projection learning for person re-identification, Pattern Recognit. 75 (2018) 63–76.
[34] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng, Person re-identification by multi-channel parts-based CNN with improved triplet loss function, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1335–1344.
[35] S. Paisitkriangkrai, C. Shen, A. van den Hengel, Learning to rank in person re-identification with metric ensembles, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1846–1855.
[36] S.M. Assari, H. Idrees, M. Shah, Human re-identification in crowd videos using personal, social and environmental constraints, in: European Conference on Computer Vision, Springer, 2016, pp. 119–136.
[37] E. Ahmed, M. Jones, T.K. Marks, An improved deep learning architecture for person re-identification, in: Computer Vision and Pattern Recognition, 2015, pp. 3908–3916.
[38] R.R. Varior, B. Shuai, J. Lu, D. Xu, G. Wang, A siamese long short-term memory architecture for human re-identification, in: European Conference on Computer Vision, 2016, pp. 135–153.
[39] D. Li, X. Chen, Z. Zhang, K. Huang, Learning deep context-aware features over body and latent parts for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 384–393.
[40] Y. Zhang, B. Li, H. Lu, A. Irie, R. Xiang, Sample-specific SVM learning for person re-identification, in: Computer Vision and Pattern Recognition, 2016, pp. 1278–1287.

Min Cao received her B.S. degree in electronic information engineering, Tianjin Polytechnic University in 2014. She is currently a Ph.D. candidate in the Institute of Automation, Chinese Academy of Sciences, Beijing, China. Her research interest is in person re-identification.

Chen Chen received her M.Sc. and Ph.D. degree in Computer Science from University of Copenhagen, Denmark in 2011 and 2013. She is currently an assistant professor in Institute of Automation, Chinese Academy of Sciences (CASIA), China. Her research interests focus on pattern recognition and machine learning.

Xiyuan Hu received his B.S. and M.E. degrees in computer science from Nanjing University of Science and Technology, Nanjing, China, in 2005 and 2007, respectively; and the Ph.D. degree in pattern recognition and intelligent systems from Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2011. In July 2011, he became a member of the Institute of Automation, Chinese Academy of Sciences, where he is currently an Associate Professor. He was a visiting scholar with the ReyLab in Harvard Medical School, Boston, US, in 2014. His research interests include adaptive signal processing, digital image processing and compression.

Silong Peng received the B.S. degree in mathematics from the Anhui University in 1993, and the M.S. and Ph.D. degrees in mathematics from Institute of Mathematics, Chinese Academy of Sciences (CAS), in 1995 and 1998, respectively. From 1998 to 2000, he worked as a postdoctoral researcher in the Institute of Automation, CAS. During this period, he was also a visiting scholar with Department of Mechanics and Mathematics, Lomonosov Moscow State University, Russia. In 2000, he became a full professor of signal processing and pattern recognition in Institute of Automation, CAS. His research interests include wavelets, multi-rate signal processing, and digital image processing.