Person re-identification based on kernel local Fisher discriminant analysis and Mahalanobis distance learning

Anonymous ICCV CHI Workshop submission

Paper ID ****

Abstract

In person re-identification (Re-ID), robust descriptors and metric are essential to improve accuracy. Mahalanobis distance based metric learning is a popular method for metric learning. However, for descriptors with high dimensionality (thousands or more), it is intractable to learn a high dimensional semi-definite positive (SPD) matrix without dimension reduction. Many metric learning methods have been proposed to learn a subspace that preserves those discriminative information. However, few works have been done to study metric learning on those subspaces obtained after an initial dimension reduction. In this paper, a two-level structure of metric learning is proposed. The kernel local Fisher discriminant analysis (KLFDA) is used to reduce dimension taking advantage of the idea that kernelization method can greatly improve Re-ID performance [20]. Then a Mahalanobis distance metric is learned on the lower dimensional descriptors based on the constraint that the intraclass distance must be at least one unit smaller than the minimum interclass distance. This method turns out to have state-of-the-art performance compared with other advanced metric learning methods.

1. Introduction

Person re-identification (Re-ID) has received increasing attention in recent years. Re-ID is very challenging due to several factors such as low image resolution, occlusion, background noise and different camera color responses. In the single-shot Re-ID problem, where only one image is provided in each camera for each person, confusion often occurs when different people have similar pose or clothes (shown in Figure 1). In the multi-shot case, there might exist significant difference in different frames of the same person showing different pose and illuminations (shown in Figure 2). To overcome these challenges most previous works either try to find better feature representation [8, 9, 16, 5] or learn better metrics [7, 13, 11, 22, 25, 17, 15, 18, 20, 23, 3].

VIPeR

Figure 1. Different pedestrians may look the same under the same viewpoint when wearing similar clothes and have similar postures. In this figure, individuals in each column are different while they look the same under the same viewpoint.

Prid2011

Figure 2. Pedestrians may look different under different angles, illumination and postures. In this figure, images in the first and second row are of the same person under different viewpoints and posture.

A good descriptor is supposed to be robust to illumination change and occlusions. Though much progress has been made, there still exists some challenges caused by classical problems like small sample problem (SSS) and high computation complexity on large datasets.

For descriptors with high dimensionality, it is hard to directly learn an SPD matrix $\boldsymbol{M}$, especially when the small sample size $n(n << d)$ is small. A popular method is to use principal component analysis (PCA) to reduce dimension. PCA is very popular for dimensionality reduction, but

PCA is a global dimension reduction scheme. As a result, much interclass discriminative information will be lost after dimensionality reduction. In this paper kernel local Fisher discriminant analysis (KLFDA) [20] is used to reduce dimensionality. This supervised dimension reduction combines linear discriminant analysis and locality preserving projection. Moreover, the kernelization version of LFDA proves to improve performance and reduces computation cost. However, the use of metric learning on dimensionality reduced vectors by KLFDA hasn't been fully studied.

The contributions of this paper are as follows. (1) KLFDA and metric learning are combined together to improve Re-ID performance. Previous works mainly used KLFDA [20] as a subspace learning method. Euclidean distance is then used to measure the similarity of dimension reduced descriptors. (2) Inspired by [22], we propose to learn a Mahalanobis distance matrix based on the constraint that intraclass distance is at least one unit smaller than the minimum interclass distance. Therefore, metric learning in the lower dimensional space is transformed into an optimization problem solved by an iterative process using gradient descent method. We compare the intraclass distance only with the minimum interclass distance in each iteration thus avoiding the need to compute every possible possible positive and negative pairs. Extensive experiments are performed on VIPeR, CUHK01, Prid_450s and GRID dataset. These ones demonstrated that the proposed method can produce state-of-the-art performance on some of these datasets.

2. Related work

Descriptor design and metric learning are two core components in people Re-ID. In descriptor design, color, texture and their statistical properties are exploited to characterize individuals. In [5] Symmetry-Driven Accumulation of Local Features (SDALF) divides the human silhouette into head, torso and legs and extract features according to their symmetry and asymmetry axis. In [8] Local Maximal Occurrence (LOMO) uses overlapping samplings and creates local histograms of pixel features extracting maxima value along horizontal stripes. In [9] hierarchical Gaussian descriptor constructs a two-level model from pixel features to patch features and from patch features to region features.

For metric learning, [8] represents the within-class and between-class difference individually with a Gaussian model. The problem to distinguish different classes is transformed into maximize the probability ratio of between-class and within-class Gaussian distribution. In [23] Null Foley-Sammon transform (NFST) is proposed to find a null space so that with this space the intraclass points collapse to a same point in the null space while interclass points are projected to different points.

Convolutional neural networks (CNN) have also been exploited in Re-ID. In [10], the author proposes a recurrent neural network layer and temporal pooling to combine all time-steps data to generate a feature vector from a video sequence. In [2], the author proposes a multi-channel layer based neural network to jointly learn both local body parts and whole body information from input person images. In [19], a convolutional neural network learning deep feature representations from multiple domains is proposed, and this work also proposes a domain-guided dropout algorithm when learning from different datasets.

There are many other works based on convolutional neural networks. However, person re-identification may be one of the areas where CNN may not perform as well as regular machine learning methods because of the small sample size problem (SSS). In most datasets, the sample size of each pedestrian is quite small. Especially in single-shot Re-ID only one frame is provided in each view for each person. This is why Re-ID more often relies on classical machine learning.

In this paper, the Mahalanobis distance learning is motivated by the Top-push Distance Learning method described in [22]. But first, dimension reduction is applied to a high-dimensional descriptor.

3. Dimension reduction based on kernel local Fisher discriminant analysis

Here we briefly review the definition of KLFDA. KLFDA is the kernel version of local Fisher discriminant analysis (LFDA). For a set of d-dimensional observations $\boldsymbol{x}_i$, where $i \in \{1, 2, \cdots, n\}$, and a set of class labels $l_i \in \{1, 2, \cdots, l\}$, two matrix are defined as the intraclass scatter matrix $\boldsymbol{S}^{(w)}$ and interclass matrix $\boldsymbol{S}^{(b)}$ as follows,

$$\begin{aligned} \boldsymbol{S}^{(w)} &= \sum_{i=1}^{l} \sum_{j:l_j=i} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T, \\ \boldsymbol{S}^{(b)} &= \sum_{i=1}^{l} n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \end{aligned} \tag{1}$$

where the $\boldsymbol{\mu}_i$ is the mean of samples whose label is i, and $\boldsymbol{\mu}$ is the mean of all samples,

$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{l_j=i} \boldsymbol{x}_j, \boldsymbol{\mu} = \frac{1}{n} \sum \boldsymbol{x}_i. \tag{2}$$

The Fisher discriminant analysis transform matrix $\boldsymbol{T}_{FDA}$ can be represented as

$$\boldsymbol{T}_{FDA} = \arg\max_{T} \frac{\boldsymbol{T}^T \boldsymbol{S}^{(b)} \boldsymbol{T}}{\boldsymbol{T}^T \boldsymbol{S}^{(w)} \boldsymbol{T}}. \tag{3}$$

Fisher discriminant analysis minimizes the intraclass scatter matrix while maximize the interclass scatter matrix. $\boldsymbol{T}$ is

computed through an eigenvalue decomposition and $\boldsymbol{T}_{FDA}$ is represented as the set of all the corresponding eigenvectors, as $\boldsymbol{T}_{FDA} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_k]$.

FDA dimension reduction however has poor performance when dealing with multimodal classes. In [6] locality preserving projection (LPP) is proposed to exploit data locality. In LPP an affinity matrix is created to record the affinity of sample $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Typically the range of elements in $\boldsymbol{A}_{i,j}$ is $[0, 1]$. There are many ways to define a $n \times n$ affinity matrix $\boldsymbol{A}$. Usually two sample points at a small distance have a larger affinity value than more distant point pairs. In this case if $\boldsymbol{x}_i$ is within k-nearest neighbours of $\boldsymbol{x}_j$ then $\boldsymbol{A}_{i,j} = 1$ otherwise $\boldsymbol{A}_{i,j} = 0$. LFDA combines FDA and LPP and has better performance [15]. The key in LFDA is it assigns weights to elements in $\boldsymbol{A}^{(w)}$ and $\boldsymbol{A}^{(b)}$, so that,

$$\begin{aligned} \boldsymbol{S}^{(w)} &= \frac{1}{2}\sum_{i=1}^{l}\sum_{j:y_j=i} \boldsymbol{A}^{w}_{i,j}(\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T, \\ \boldsymbol{S}^{(b)} &= \frac{1}{2}\sum_{i=1}^{l} \boldsymbol{A}^{b}_{i,j}(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \end{aligned} \tag{4}$$

where

$$\begin{aligned} \boldsymbol{A}^{(w)}_{i,j} &= \begin{cases} \boldsymbol{A}_{i,j}/n_c & y_i = y_j \\ 0 & else \end{cases}, \\ \boldsymbol{A}^{(b)}_{i,j} &= \begin{cases} (\frac{1}{n} - \frac{1}{n_c})\boldsymbol{A}_{i,j} & y_i = y_j \\ \frac{1}{n} & else \end{cases}. \end{aligned} \tag{5}$$

with y_i being the class label of sample point $\boldsymbol{x}_i$.

When applying LFDA to high dimensional descriptors, computational cost becomes an issue. Suppose the vector data has dimension d, we have to extract the eigenvalue of a matrix of dimension $d \times d$. For some descriptors, d could be more than 20000 and thus the cost is prohibitive.

Kernelization is a proven solution to this explosion in dimensionality. Moreover, in [20] it has been demonstrated that kernel-based metric learning methods has better performance than those without kernelization. Kernelization is a projection from low dimensional space to high dimensional space, which makes classification and clustering much more accurate. The difference of KLFDA is that the interclass and intraclass scatter matrix will be kernelised and the eigenvalue decomposition will be operated on kernel space. Suppose a set of sample points $\boldsymbol{x}_i, i \in \{1, 2, \cdots, n\}$, can be mapped to a implicit higher feature space by a function $\phi(\boldsymbol{x}_i)$. The kernel function is implicit and only the inner product of mapped vectors $\phi(\boldsymbol{x}_i)$ and $\phi(\boldsymbol{x}_j)$ needs to be known. A 'kernel trick' is applied by defining a function $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = < \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) >$, where the $< \cdot >$ is the inner product. There are many types of kernels such as polynomial kernel and radial basis function (RBF) kernel. In this paper the RBF kernel is adopted. A RBF kernel is defined as

$$k_{RBF}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp^{(-\gamma||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)}, \tag{6}$$

where γ is a constant term.

4. Metric learning on dimension reduced space by gradient descent optimization

Top-push Distance metric learning is proposed in [22]. This paper combines KLFDA and Top-push Distance metric learning to produce a robust re-identification method. We briefly review the Top-push Distance metric learning in this section. Since Re-ID is a ranking problem, it is desired that the rank 1 descriptor should be the right match. Instead of comparing all the possible positive and negative pairs, a simplified version was proposed in [22] in which the intraclass distance should be at least one unit smaller than inter distance. Given a Mahalanobis matrix $\boldsymbol{M}$, for dimension reduced sample points $\boldsymbol{x}_i, i = 1, 2, 3, \cdots, n$, n is the number of all samples. The requirement is that distance between positive pair should be at least one unit smaller than the minimum of all negative distance. This can be denoted as

$$D(\boldsymbol{x}_i, \boldsymbol{x}_j) + \rho < \min D(\boldsymbol{x}_i, \boldsymbol{x}_k), y_i = y_j, y_i \neq y_k, \tag{7}$$

ρ is a slack variable and $\rho \in [0, 1]$. This equation can be transformed into an optimization problem with respect to descriptor $\boldsymbol{x}_i$ as

$$\arg\min \sum_{y_i=y_j} \max\{D(\boldsymbol{x}_i, \boldsymbol{x}_j) - \min_{y_i \neq y_k} D(\boldsymbol{x}_i, \boldsymbol{x}_k) + \rho, 0\}. \tag{8}$$

However, the equation above only penalizes small interclass distance. Another term is needed to penalize large intraclass distance. That is, we want to make the sum of intraclass distance as small as possible. These two terms are combined using a ratio factor α. The target function can then be denoted as

$$\begin{aligned} f(\boldsymbol{M}) = (1-\alpha) \sum_{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{y}_i = y_j} D(\boldsymbol{x}_i, \boldsymbol{x}_j) + \\ \alpha \sum_{\boldsymbol{x}_i, \boldsymbol{x}_j, y_i = y_j} \max\{D(\boldsymbol{x}_i, \boldsymbol{x}_j) - \min_{y_i \neq y_k} D(\boldsymbol{x}_i, \boldsymbol{x}_k) + \rho, 0\}. \end{aligned} \tag{9}$$

This way the problem is transformed into an optimization problem. Notice that $D(\boldsymbol{x}, \boldsymbol{y})$ can be denoted as

$$D(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{M} (\boldsymbol{x} - \boldsymbol{y}) = Tr(\boldsymbol{M}\boldsymbol{X}_{i,j}), \tag{10}$$

where $\boldsymbol{X}_{i,j} = (\boldsymbol{x} - \boldsymbol{y}) * (\boldsymbol{x} - \boldsymbol{y})^T$, and Tr is matrix trace.

Therefore, Eq. (9) can be transformed as follow,

$$f(\boldsymbol{M}) = (1-\alpha)\sum_{y_i=y_j} Tr(\boldsymbol{M}\boldsymbol{X}_{i,j}) + \alpha \sum_{y_i=y_j} \max\{Tr(\boldsymbol{M}\boldsymbol{X}_{i,j}) - \min_{y_i \neq y_k} Tr(\boldsymbol{M}\boldsymbol{X}_{i,k}) + \rho, 0\}. \tag{11}$$

To minimize Eq. (11), the gradient descent method is used. The gradient with respect to $\boldsymbol{M}$ is computed as

$$\boldsymbol{G} = \frac{\partial f}{\partial \boldsymbol{M}} = (1-\alpha)\sum_{y_i=y_j} \boldsymbol{X}_{i,j} + \alpha \sum_{y_i=y_j, y_i \neq y_k} (\boldsymbol{X}_{i,j} - \boldsymbol{X}_{i,k}), \tag{12}$$

the iteration process is summarized as in Table 1. Basically the set of all intraclass distances remain constant during this optimization process while the set of interclass distances is recomputed at each iteration in order to extract the minimal distances.

Table 1. Optimization algorithm of Mahalanobis distance matrix learning

Gradient optimization algorithm for target function
Input Descriptors of training person pairs
Output An SPD matrix
Initialization
Initialize $\boldsymbol{M}_0$ with eye matrix $\boldsymbol{I}$;
Initialize target function value f_0 with $\boldsymbol{M}_0$ using Eq. (11);
Initialize Iteration count $t = 0$;
while(not converge)
• Update $t = t + 1$;
• Find corresponding $\boldsymbol{x}_k$ for all $\boldsymbol{x}_i$, where $y_i \neq y_k$, so that $\boldsymbol{x}_i$ and $\boldsymbol{x}_k$ has minimal interclass distance;
• Update gradient $\boldsymbol{G}_{t+1}$ with Eq. (12) with the found $\boldsymbol{X}_{i,k}$;
• Update $\boldsymbol{M}$ with equation : $\boldsymbol{M}_{t+1} = \boldsymbol{M}_t - \lambda \boldsymbol{G}_t$;
• Project $\boldsymbol{M}_{t+1}$ to the semi-positive definite space;
• Update the target value $f\|_{\boldsymbol{M}=\boldsymbol{M}_{t+1}}$;
end while
return $\boldsymbol{M}$

After each iteration, to make sure the updated $\boldsymbol{M}$ is an SPD matrix, first a eigenvalue decomposition is performed on $\boldsymbol{M}$, and we have

$$\boldsymbol{M} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T. \tag{13}$$

Here $\boldsymbol{\Lambda}$ is a diagonal matrix, and its diagonal elements are eigenvalues. The negative eigenvalues in $\boldsymbol{V}$ are set to zeros, then $\boldsymbol{M}$ is restored by Eq. (13).

5. Experiment

In this paper, we apply the proposed metric based on KLFDA and Mahalanobis distance learning to the GOG descriptor. First, KLFDA is applied to GOG descriptor to reduce the dimension to $C-1$, where C is the number of person pairs in the training dataset. A transformation matrix $\boldsymbol{T}$ is thus generated. Second, from the dimension-reduced descriptors, a $(C-1)\times(C-1)$-dimensional SPD matrix $\boldsymbol{M}$ is learned based on the relative distance comparison. A gradient descent method is used to minimize the target function denoted by the Eq. (11). With the trained SPD matrix $\boldsymbol{M}$, the similarity scores between person image pairs can be computed.

The hierarchical GOG descriptor [9] describes a local region in an image with hierarchical Gaussian distribution in which the mean and the covariance are all included in the parameters. The local region are divided into many overlapping patches at first. The pixels in each patch are first described with a vector $\boldsymbol{f} = (y, M_0, M_{90}, M_{180}, M_{270}, R, G, B)$. The patch is modelled with a multivariate Gaussian function based on all the pixels inside. The mean and covariance parameter are transformed into a one-dimensional vector so that the patches is described with this vector. A same process is repeated so that the region is modelled based on all the patches inside the region.

There are two versions of this Gaussian of Gaussian descriptor. The first one is extracted from RGB color space only; it is denoted as GOG_{rgb}. The second one is extracted from four color space {RGB, HSV, Lab, nRGB}. nRGB is the normalized RGB color space. For Gaussian of Gaussian descriptor in all the four color spaces, the dimensions are {7567,7567,7567,4921}. Therefore, the descriptor concatenating all four color spaces has a dimension of 27622.

5.1. Datasets and evaluation settings

VIPeR dataset is the most used dataset in person Re-ID. In this dataset there are 632 different individuals and for each person there are two outdoor images from different viewpoints. All the images are scaled to 48×128 pixels. In this experiment the we randomly select 316 individuals from camera a and camera b as the training set, the rest of the images in camera a are used as probe images and those in camera b as gallery images. This process is repeated 10 times to compute average performance values.

CUHK1 dataset contains 971 identities from two disjoint camera views. The cameras are static in each pair of view and images are listed in the same order. For each individual, there are two images in each view. All images are scaled to 60×160 pixels. In this paper, we randomly select 485 image pairs as training data and the rest person pairs are used for test data.

Prid_450s dataset contains 450 image pairs recorded from

two different, static surveillance cameras. Additionally, the dataset also provides an automatically generated, motion-based foreground/background segmentation as well as a manual segmentation of person parts. The images are stored in two folders that represent the two camera views. In this test, we randomly selected 225 person pairs from each of two camera views as the training set, and the remaining persons are left as gallery and probe images.

GRID defines two camera views. A probe folder contains 250 probe images captured in one view. The gallery folder contains 250 true match images of the probes. Besides, in gallery folder there are a total of 775 additional images that do not belong to any of the probes. We randomly selected 125 persons from those 250 persons appearing in both camera views as training pairs, and the remaining persons in probe folder are used as probe images. The remaining 125 persons and those 775 additional persons from gallery folder are used as gallery images. In [9], mean removal and L_2 normalization is shown to improve performance by 5.1%. The reason for this is that mean removal and normalization can reduce the impact of extrema in a single descriptor. For fair comparisons, mean removal and L_2 normalization are also adopted in this experiment.

5.2. Parameters setting for metric learning

The metric learning process includes a number of parameters that control the iterative gradient descent. These are the slack variable ρ, maximal iteration T, gradient step λ, the interclass and intraclass limitation factor α and the updating ratio β. The slack variable ρ is initialized to one to ensure the minimum interclass distance is at least one unit larger than intraclass distance. The step size of gradient updating λ is initialized at 0.01. When target value f increases, λ is scaled by a factor of 0.5, and λ is scaled by 1.01 when target value f decreases. To judge if target value converges, the threshold β is defined as the ratio target function value change versus previous target function value, that is, $\beta = \frac{(f_{t+1}-f_t)}{f_t}$. Based on our experiments, when it satisfies $\beta = 10^{-5}$, the target value has converged and the iteration is stopped. The maximal iteration times t is set to 100 since the target value f will converge in around 15 iterations. The last parameter for the iteration is factor α that assigns weight to interclass distance comparison. We tested 11 different values for α ranging from 0 to 1 with a step of 0.1. We found that the rank 1 and rank 5 scores reach a maxima inside the interval $[0.7, 0.8]$. Then another ten trials with alpha ranging from $[0.7, 0.8]$ with a step of 0.01 revealed that a value at 0.76 for α led to the highest top-rank score. The best α value should have as large top rank scores as possible and at last we find that the optimal value for α is 0.76.

Table 2. Parameters setting

Parameters	α	β	λ	t	ρ
Values	0.76	10^{-5}	0.01	100	1

5.3. Performance analysis

We compared our proposed metric with other state-of-the-art metrics including NFST [23], XQDA [8]. NFST is a metric which learns a null space for descriptors so that the same class descriptors will be projected to a single point to minimize intraclass scatter matrix while different classes are projected to different points. This metric is a good solution to small sample problem in Re-ID. XQDA learns a projection matrix W and then a Mahalanobis SPD matrix M is learned in the subspace. Those two metrics proved to have state-of-the-art performance compared with many other methods. The GOG_{rgb} in all tables means the hierarchical Gaussian descriptor in RGB color space while $\text{GOG}_{\text{fusion}}$ means the one that combines four different color spaces {RGB, Lab, HSV, nRnG}.

In [9], it has been shown that GOG + XQDA outperforms many other combinations, including Metric ensemble [12], SCNCD [21], Semantic method [14]. In [23], it has been shown that LOMO + NFST outperforms other metrics such as LMNN [17], KCCA [18], ITML [3], KLFDA [15], MFA [20], KISSME [7], Similarity learning [1], SCNCD [21], Mid-level filters [24] and Improved deep learning [4]. Based on the result that XQDA and NFST outperform other metrics, only XQDA and NFST are used in this thesis to compare with our proposed metric learning.

VIPeR A comparison form is given in Table 3. Some of recent results are also included in this form. We can find that the rank scores are better than those of NFST and XQDA in terms of both GOG_{rgb} and $\text{GOG}_{\text{fusion}}$. More specifically, the rank 1, rank 5, rank 10, rank 15 and rank 20 scores of proposed metric learning are 0.76%, 0.92%, 1.39%, 1.08%, 1.52% higher than those of GOG_{rgb} + XQDA. The rank 1, rank 5, rank 10, rank 15 and rank 20 $\text{GOG}_{\text{fusion}}$ scores of proposed metric learning are 0.35%, -0.54%, 0.98%, 0.66%, 0.79% higher than $\text{GOG}_{\text{fusion}}$ + XQDA respectively. Also we can see that the proposed metric learning has a better performance than NFST.

Table 3. Performance of different metrics on VIPeR

	Rank(%)				
Methods	1	5	10	15	20
GOG_{rgb}+NFST	43.23	73.16	83.64	89.59	92.88
GOG_{rgb}+XQDA	43.01	73.92	83.86	89.24	92.37
GOG_{rgb}+Proposed	**43.77**	**74.84**	**85.25**	**90.32**	**93.89**
$\text{GOG}_{\text{fusion}}$+NFST	47.15	76.39	87.31	91.74	94.49
$\text{GOG}_{\text{fusion}}$+XQDA	47.97	**77.44**	86.80	91.27	93.70
$\text{GOG}_{\text{fusion}}$+Proposed	**48.32**	76.90	**87.78**	**91.93**	**94.49**

CUHK1 We can find that the rank 1, rank5, rank 10, rank 15, rank 20 score of GOG_{rgb} combined with proposed metric are 5.4%, 4.18%,3.31%,2.16%,1.46% higher

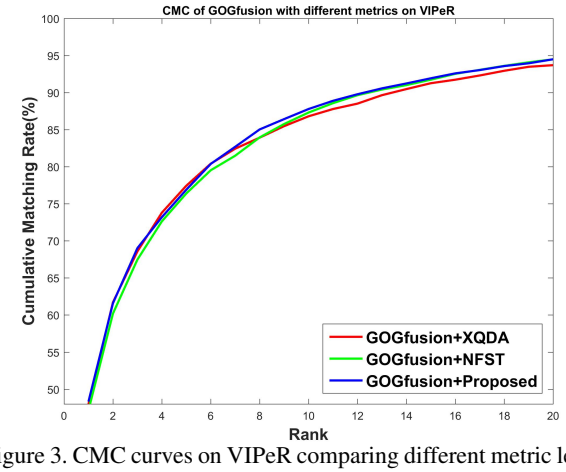

Figure 3. CMC curves on VIPeR comparing different metric learning

Table 4. Performance of different metrics on CUHK1

	Rank(%)				
Methods	1	5	10	15	20
GOG_{rgb}+NFST	55.60	83.02	89.07	91.98	93.56
GOG_{rgb}+XQDA	50.51	80.06	87.10	90.99	93.21
GOG_{rgb}+Proposed	**55.91**	**84.24**	**90.41**	**93.15**	**94.67**
GOG_{fusion}+NFST	56.26	83.66	89.63	92.22	93.70
GOG_{fusion}+XQDA	52.10	81.85	88.81	91.98	94.01
GOG_{fusion}+Proposed	**56.67**	**84.49**	**90.51**	**93.31**	**94.84**

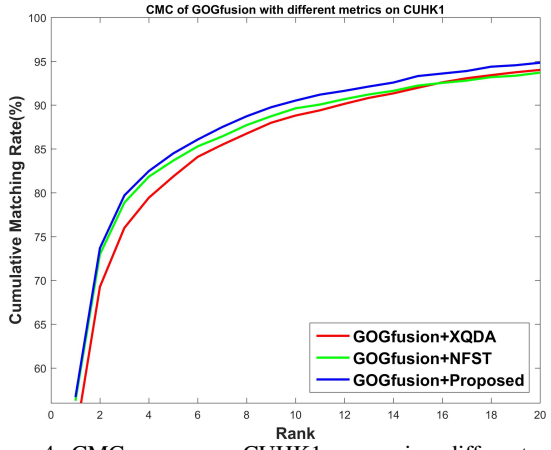

Figure 4. CMC curves on CUHK1 comparing different metric learning

than XQDA, and 0.31%, 1.22%,1.34%, 1.17%, 1.11% than NFST. Also the rank 1, rank5, rank 10, rank 15, rank 20 score of GOG_{fusion} combined with proposed metric are 4.57%, 2.64%, 0.70%, 1.33%, 0.83% higher than GOG_{fusion} combined with XQDA, and 0.41%, 0.83%, 0.88%, 1.09%, 1.14% than GOG_{fusion} combined with NFST.

Prid_450s In this dataset, we can find the rank 1 score of XQDA and NFST is higher than proposed metric, but they have almost the same rank 5, rank 10, rank 15, and rank 20 scores with respect to both kinds of descriptors.

Table 5. Performance of different metrics on prid_450s

	Rank(%)				
Methods	1	5	10	15	20
GOG_{rgb}+NFST	61.96	84.98	90.53	94.09	96.09
GOG_{rgb}+XQDA	**65.29**	**85.02**	91.13	**94.76**	**96.49**
GOG_{rgb}+Proposed	60.71	84.53	**91.29**	94.13	96.27
GOG_{fusion}+NFST	64.53	86.62	92.93	**95.78**	**97.42**
GOG_{fusion}+XQDA	**68.40**	**87.42**	**93.47**	95.69	97.02
GOG_{fusion}+Proposed	62.80	86.58	92.36	95.29	96.89

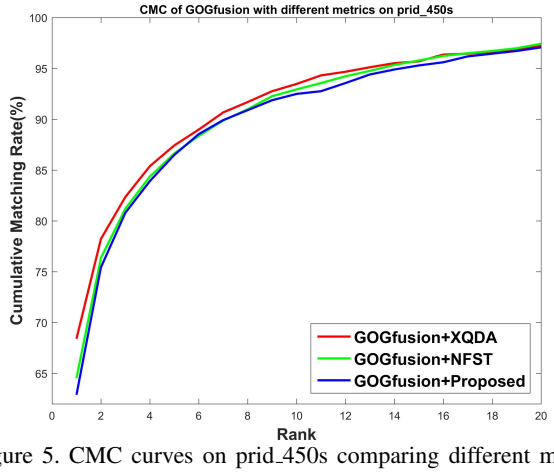

Figure 5. CMC curves on prid_450s comparing different metric learning

Table 6. Performance of different metrics on GRID

	Rank(%)				
Methods	1	5	10	15	20
GOG_{rgb}+NFST	21.84	41.28	50.96	57.44	62.88
GOG_{rgb}+XQDA	**22.64**	**43.92**	**55.12**	**61.12**	**66.56**
GOG_{rgb}+Proposed	**22.64**	43.68	52.00	59.04	65.04
GOG_{fusion}+NFST	23.04	44.40	54.40	61.84	66.56
GOG_{fusion}+XQDA	23.68	**47.28**	**58.40**	**65.84**	**69.68**
GOG_{fusion}+Proposed	**23.92**	44.64	54.88	62.32	66.40

GRID We can see that the rank 1 score of proposed metric are 0.24% higher than XQDA and 0.88% higher than NFST in terms of GOG_{fusion}, but XQDA outperforms proposed metric on rank 5, rank 10, rank 15 and rank 20 scores. Besides, proposed metric outperforms NFST on rank 5, rank 10, rank 15 scores.

In summary, the Re-ID performance is improved in VIPeR, CUHK01 dataset, and has almost the same performance with NFST and XQDA on prid_450s dataset. Specifically, proposed metric learning has the best rank 1 score in GRID dataset and its performance is only second to XQDA.

5.4. Computational cost

The training time increases when the dataset size increases. For the CUHK1 dataset which has the largest number of training pairs, it takes half an hour to train the metric on a desktop PC with a 16GB RAM, Intel i5 processor. For datasets with fewer training pairs like Prid_2011 and GRID,

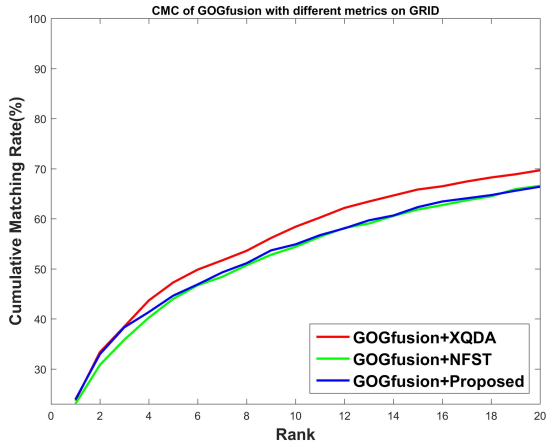

Figure 6. CMC curves on GRID comparing different metric learning

metric learning takes no more than five minutes.

6. Conclusion

In this paper we combined KLFDA with gradient descent method based metric learning. An SPD matrix is learned on the lower dimension space after dimensionality reduction by KLFDA.

In summary, our proposed metric improved the Re-ID accuracy in VIPeR and CUHK1 datasets, and has almost the same performance with NFST and XQDA in the prid_450s dataset. Furthermore, the proposed metric learning has the best rank 1 score in the GRID dataset and its performance is only second to XQDA. The proposed metric has superior performance for the following reasons: (1) dimension reduction by KLFDA exploits the nonlinearity, and the loss of discriminant information between classes is minimized; (2) the simplified relative distance constraint optimization helps to confine the Mahalanobis distance matrix $\boldsymbol{M}$ to discriminate different classes.

References

[1] D. Chen, Z. Yuan, G. Hua, and N. Z. a. Wang. Similarity Learning on an Explicit Polynomial Kernel Feature Map for Person Re-Identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, Apr. 2015. 5

[2] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng. Person Re-Identification by Multi-Channel Parts-Based CNN With Improved Triplet Loss Function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–10, July 2016. 2

[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, June 2007. 1, 5

[4] M. J. Ejaz Ahmed and T. K. Marks. An Improved Deep Learning Architecture for Person Re-Identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. 5

[5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person Re-Identification by Symmetry-Driven Accumulation of Local Features. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1–8. IEEE, Mar. 2016. 1, 2

[6] X. He and P. Niyogi. Locality Preserving Projections. In *NIPS*, pages 1–8, Nov. 2003. 3

[7] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large Scale Metric Learning from Equivalence Constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2288–2295. IEEE, Apr. 2012. 1, 5

[8] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–10, Apr. 2015. 1, 2, 5

[9] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato. Hierarchical Gaussian Descriptor for Person Re-Identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1363–1372, Dec. 2016. 1, 2, 4, 5

[10] N. McLaughlin, J. M. del Rincon, and P. Miller. Recurrent Convolutional Network for Video-Based Person Re-Identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–10, July 2016. 2

[11] A. Mignon and F. Jurie. PCCA: A New Approach for Distance Learning from Sparse Pairwise Constraints. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1–7. IEEE, Apr. 2012. 1

[12] S. Paisitkriangkrai, C. Shen, and 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*, pages 1–10, Jan. 2015. 5

[13] S. Pedagadi, J. Orwell, Velastin, Sergio, and B. Boghos. Local fisher discriminant analysis for pedestrian re-identification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3318–3325. IEEE, Jan. 2013. 1

[14] Z. Shi, T. M. Hospedales, and T. Xiang. Transferring a Semantic Representation for Person Re-Identification and Search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–10, Apr. 2015. 5

[15] M. Sugiyama. Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction. In *Proceedings of the 23rd international conference on Machine learning*. ACM, Dec. 2006. 1, 3, 5

[16] O. Tuzel, F. Porikli, and P. Meer. Region Covariance: A Fast Descriptor for Detection and Classification. In *European conference on computer vision*, pages 1–14. Springer, Dec. 2016. 1

[17] Weinberger, K. Q, Saul, and L. K. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *Journal of Machine Learning Research*, 10:207–244, Feb. 2009. 1, 5

[18] M. Welling. Kernel Canonical Correlation Analysis. pages 1–3, Mar. 2005. 1, 5

[19] T. Xiao, Hongsheng, Li, Wanli, Ouyang, and X. Wang. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification . In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–10, Apr. 2016. 2

[20] F. Xiong, M. Gou, O. Camps, and S. Mario. Person Re-Identification using Kernel-based Metric Learning Methods. In *European conference on computer vision*, pages 1–16. Springer, July 2014. 1, 2, 3, 5

[21] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In *European Conference on Computer Vision*, 2014. 5

[22] J. You, A. Wu, X. Li, and W.-S. Zheng. Top-push Video-based Person Re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, Jan. 2016. 1, 2, 3

[23] L. Zhang, T. Xiang, and S. Gong. Learning a Discriminative Null Space for Person Re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1239–1248, Mar. 2016. 1, 2, 5

[24] R. Zhao, W. Ouyang, and X. Wang. Learning Mid-level Filters for Person Re-identification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 144–151. IEEE, May 2014. 5

[25] W. Zheng, S. Gong, and T. Xiang. Person Re-identification by Probabilistic Relative Distance Comparison. In *Computer vision and pattern recognition (CVPR), 2011 IEEE conference on*, pages 1–8. IEEE, Nov. 2016. 1