Person re-identification based on and kernel local fisher discriminant analysis and Mahanalobis distance learning

Qiangsen He
School of Electrical Engineering and Computer Science, University of Ottawa
Ottawa, Ontario, qiangsenhe@gmail.com

Abstract—**In person re-identification(re-ID) it's very important to choose robust descriptors and metric learning to improve accuracy. Mahanalobis distance based metric learning is a popular method for metric learning. However, since directly extracted descriptors usually have high dimension(thousands or more), it's intractable to learn a high dimensional semi-definite positive(SPD) matrix without dimension reduction. Many subspace learning metrics have been proposed to learn a subspace while preserving those discriminative information as much as possible. However, few work has been done to probe the metric learning on those subspace after the first-time metric learning. In this paper the KLFDA [1] is used to reduce dimension given that kernelization method can greatly improve re-identification performance for nonlinearity. Then a Mahanalobis distance metric is learned on lower dimensional descriptors based on the limitation that the intraclass distance is at least 1 unit smaller than the minimum interclass distance. By comparing the intraclass distance only with the interclass distance the computation complexity is reduced. This method turns to have excellent performance compared with other advanced metric learning.**

1. Introduction

Person re-identification(re-ID) has received increasing attention in recent years. This task is very challenging caused by many factors like low image resolution, occlusion, background noise and different camera color response, etc. In the single shot re-ID problem, since only one image is provided in each camera for each person, it might be quite confusing when different people have similar pose or clothes. Also, in the multi-shots case, there might exist quite much difference even in different frames of the same person for different pose and illuminations [2]. Therefore, good descriptors are supposed to be robust to illumination change and occlusions.

Most previous work try to find better feature representation [3–6] and metric learning [1, 2, 7–15], or a subspace learning. Most descriptors are color and texture based descriptors. Distance metric learning has been fully studied in [16], most existing metric learning are based on Mahanalobis distance to learn a metric which outputs small intraclass distance and large interclass distance. Though much progress has been made, there still exists some challenges caused by classical problems like small sample problem(SSS), computation complexity on large datasets.

Among all those metric learning methods, fisher discriminant analysis(FDA) proves to be one of advanced metrics to improve re-ID performance. It tries to maximize the ration of interclass scatter matrix versus intraclass scatter distance and this problem is transformed as a eigenvalue decomposition problem. Local fisher discriminant analysis combines locality preserving projection(LPP) with FDA to exploit locality of sample points. Besides, kernel method [13] has been shown to improve re-ID performance since it exploits the nonlinearity of data.

For descriptors with high dimension($d \geq 1000$), it's hard to directly learn a Mahanalobis distance matrix M for the small sample size $n(n << d)$. A popular method is to use principal component analysis(PCA) to reduce dimension. PCA is a very popular preprocessing method which uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. One problem is PCA doesn't consider the information between classes thus many discriminant information will be lost after dimension reduction. In this paper kernel local fisher discriminant Analysis(KLFDA) [1] is used to reduce dimension, this supervised dimension reduction combines the linear discriminant analysis and locality preserving projection. Moreover, the kernelization version of LFDA proves to improve performance and reduce computation cost. However, the metric learning on dimension reduced vectors by KLFDA hasn't been fully probed.

In this paper, KLFDA is only used to reduce dimension instead of being used as an unique subspace learning method. Previous work uses Euclidean distance to measure the similarity after dimension reduction with KLFDA, and the metric learning on the projected vector space by KLFDA has not been fully probed. In this paper for the dimension reduced data, a Mahanalobis distance matrix M is learned based on the limitation that intra distance is at least 1 unit smaller than interclass distance. This iterative metric learning method is inspired by the [2] paper. A target function respect M is created to penalize big intraclass distance and small interclass distance. So we transformed re-ID into a optimization problem in the lower dimension space. With the target function, the gradient descent method is used to get a optimal M.

2. Related work

Previous work focus on find more discriminative descriptors and better metric learning. A good descriptor is robust to problems like illumination, low resolution and viewpoint, etc. In [6] the symmetry and asymmetry property of pedestrian foreground is considered. But this descriptor depends much on foreground extraction performance. In [17] the image is divided into a few horizontal strides, for each slide the color histogram is extracted, then the color histograms are concatenated together as the whole image's descriptor. This descriptor is simple but has low performance for it doesn't consider the texture and pixel spatial distribution. In [5] the covariance descriptors of local patches are used to represent people. In [3] densely sampled overlapping small windows are adopted to overcome viewpoint variation. In each sample window the scale invariant local ternary pattern(SILTP) and local binary pattern(LBP) is extracted, then the local maximal occurrence of patterns of all windows are extracted to consist of LOMO descriptor. The LOMO descriptor succeeds in characterizing images with viewpoint changes. In [4] a two-level multivariate gaussian descriptors are proposed to exploit the stochastic property of the color, texture. By a Riemannian manifold based mapping, this descriptors embeds d dimensional multivariate gaussian function to a $d+1$ dimensional semi-positive definite space. Then by matrix vectorization we can get the gaussian of gaussian descriptor(GOG). The combination of GOG and cross view quadratic discriminative analysis outperforms most advanced works.

Other works try to use metric learning to improve re-ID performance. In [10] the distance between positive pairs must be smaller than the distance between negative pairs. Since it has to compare possible positive and negative pairs, computation complexity will be quite huge. In [14] the author proposes to to learn a null space in which the descriptors of the same class will collapse into the same point while descriptors of different classed are projected onto different points. In [18] the author exploits the combination of different metrics by assigning weights to different metrics and optimize the weights to maximize probability that any of these top k matches are correct. In [19] the author proposed a earth mover's distance(EMD) based metric learning for descriptors of gaussian mixture model(GMM). It's more complex to compute similarity of GMM model. However, one problem this EMD-based distance metric is it's huge time complexity, which limits its real-time application. In [20] the author presents a new semantic attribute learning approach for person re-identification and search but this method suffers from low performance. In [3] the author extends the Bayesian face and keep it simple and straightforward(KISSME) approaches to learn a discriminant low dimensional subspace by cross-view quadratic discriminant analysis(XQDA), this work has a top performance in most datasets when combined with LOMO descriptor and GOG descriptors.

There are also some works which tries to use convolutional neural network [21, 22] to improve accuracy. However, person re-identification may be one of the area which CNN won't work for the small sample size(SSS) problem. 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. So re-ID will more rely on classical machine learning.

3. Dimension reduction based on kernel local fisher discriminant analysis

3.1. Background of kernel LFDA

Fisher discriminant analysis is a supervised dimension reduction algorithm, whose input includes the original descriptors and the class labels. Here a brief review of Fisher linear analysis and LPP is given. For a set of d-dimensional observations $\boldsymbol{x}_i$, where $i \in \{1, 2, \cdots, n\}$, the label $l_i \in \{1, 2, \cdots, l\}$. Two matrix are defined as the intraclass scatter matrix $\boldsymbol{S}^{(w)}$ and between class matrix $\boldsymbol{S}^{(b)}$,

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

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

$$\begin{aligned} \boldsymbol{\mu}_i &= \frac{1}{n_i} \sum \boldsymbol{x}_i \\ \boldsymbol{\mu} &= \frac{1}{n} \sum \boldsymbol{x}_i \end{aligned} \tag{2}$$

The Fisher Discriminant Analysis transform matrix $\boldsymbol{T}$ can be represented as

$$\boldsymbol{T} = \arg\max \frac{\boldsymbol{T}^T \boldsymbol{S}^{(b)} \boldsymbol{T}}{\boldsymbol{T}^T \boldsymbol{S}^{(w)} \boldsymbol{T}} \tag{3}$$

Fisher discriminant analysis tries to minimize the within-class distance while maximize the between class distance. The $\boldsymbol{T}$ is computed by the eigenvalue decomposition so that the between class scatter is maximized and the intraclass scatter matrix is minimized. $\boldsymbol{T}$ can be represented as the set of all the corresponding eigenvectors, as $\boldsymbol{T} = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \cdots, \boldsymbol{\phi}_k)$.

FDA analysis has a form similar with signal and noise ratio, however, the FDA dimension reduction may have poor performance for it doesn't consider the locality of data. In[Hexiaofei] locality preserving projection 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 manners to define a $n \times n$ affinity matrix $\boldsymbol{A}$, usually the two sample points with a smaller distance measured by Euclidean or other distance has a higher affinity value than those with bigger distance value. One of them is if $\boldsymbol{x}_i$ is within k-nearest neighbours of $\boldsymbol{x}_j$ then $\boldsymbol{A}_{i,j} = 1$ otherwise $\boldsymbol{A}_{i,j} = 0$. Another diagonal matrix D can be defined that each diagonal element is the sum of corresponding column in $\boldsymbol{A}$,

$$\boldsymbol{D}_{i,i} = \sum_{j=1}^{n} \boldsymbol{A}_{i,j} \tag{4}$$

then the LPP transform matrix is defined as follow,

$$\boldsymbol{T}_{LPP} = \underset{\boldsymbol{T} \in \boldsymbol{R}^{d \times m}}{\arg\min} \frac{1}{2} \sum_{i,j=1}^{n} \boldsymbol{A}_{i,j} || \boldsymbol{T}^T \boldsymbol{x}_i - \boldsymbol{T}^T \boldsymbol{x}_j || \tag{5}$$

so that $\boldsymbol{T}^T \boldsymbol{X} \boldsymbol{D} \boldsymbol{X}^T \boldsymbol{T} = \boldsymbol{I}$. Suppose the subspace has a dimension of m, then LPP transform matrix T can be represented as

$$\boldsymbol{T}_{LPP} = \{ \boldsymbol{\phi}_{d-m+1} | \boldsymbol{\phi}_{d-m+1} | \cdots \boldsymbol{\phi}_d \}$$

And each $\boldsymbol{\phi}$ in T is the eigenvector of following fomula,

$$\boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T \boldsymbol{\phi} = \gamma \boldsymbol{X} \boldsymbol{D} \boldsymbol{X}^T \tag{6}$$

where γ is corresponding eigenvalue of $\boldsymbol{\phi}$, and $L = D - A$. But the LPP dimension reduction is still not discriminant enough, LFDA combines FDA and LPP and have a more strong performance. The key in LFDA is it assigns weights to elements in $\boldsymbol{A}^{(w)}$ and $\boldsymbol{A}^{(b)}$, so that,

$$\begin{aligned} \boldsymbol{S}^{(w)} &= \frac{1}{2} \sum_{i=1}^{l} \sum_{j:l_j = i} \boldsymbol{A}^{w}_{i,j} (\boldsymbol{x}_j - \boldsymbol{\mu}_i)(\boldsymbol{x}_j - \boldsymbol{\mu}_i)^T \\ \boldsymbol{S}^{(b)} &= \frac{1}{2} \sum_{i=1}^{l} \boldsymbol{A}^{b}_{i,j} (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \end{aligned} \tag{7}$$

where

$$\begin{aligned} \boldsymbol{A}^{(w)}_{i,j} &= \begin{cases} \boldsymbol{A}_{i,j}/n_c & y_i = y_j \\ 0 & else \end{cases} \\ \boldsymbol{A}^{(b)}_{i,j} &= \begin{cases} (\frac{1}{n} - \frac{1}{n_c}) \boldsymbol{A}_{i,j} & y_i = y_j \\ \frac{1}{n} & else \end{cases} \end{aligned} \tag{8}$$

where y_i is the class label of sample point $\boldsymbol{x}_i$.

When applying the LFDA to original high dimensional descriptors, one problem is the computation cost. Suppose the vector data has a dimension of d, LFDA has to solve the eigenvalue a matrix with dimension $d \times d$. In some descriptors the d could be more than 20000 and thus the cost is not trivial. It may takes a few days to compute even on a computer with good configurations. For the huge complexity of LFDA, the kernel LFDA is introduced to shorten running time.

3.2. Kernel LFDA

Kernelization is proved to greatly improve performance since the non-linearity is exploited. In [13] it has been demonstrated that kernelization improves the performance of many dimension reduction and metric learning. Kernelization is a projection from low dimension to high dimension, which may make classification and clustering much more accurate. The difference of kernel version LFDA is that the between class and intraclass scatter matrix will be transformed into kernel space 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 $\phi(\boldsymbol{x}_i)$. It has been proved that kernel function can be implicit and only the inner product of mapped vectors $\phi(\boldsymbol{x}_i)$ and $\phi(\boldsymbol{x}_j)$ need to be known. The kernel trick is proposed to solve this problem by defining a function $k(\boldsymbol{x}_i, \boldsymbol{x}_j) =< \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) >$, the $< \cdot >$ is the inner product. There are many kinds of kernel like linear kernel, polynomial kernel and radial basis function(RBF) kernel. In this paper the RBF kernel is adopted. A RBF kernel is defined as $k_{RBF}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp^{(-\gamma ||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)}$.

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

The Mahalanobis distance based metric learning has received much attention in similarity computing. The Mahanalobis distance of two observations $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$D(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{M} (\boldsymbol{x} - \boldsymbol{y}), \tag{9}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$are $d \times 1$ observation vectors, $\boldsymbol{M}$ is a positive-semidefinite matrix. Since $\boldsymbol{M}$ is positive-semidefinite, $\boldsymbol{M}$ can be decomposed as $\boldsymbol{M} = \boldsymbol{W}^T \boldsymbol{W}$, and Mahanalobis distance can also be written as

$$D(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{W}^T \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{y}) = ||\boldsymbol{W}(\boldsymbol{x} - \boldsymbol{y})|| \tag{10}$$

Therefore, Mahanalobis distance can be regarded as a variant of Euclidean distance. Since re-identification is a problem of ranking, it is desired that the rank-1 descriptor should be the right match. In this paper, instead compare all the possible positive and negative pairs, a simplified version is proposed that the intraclass distance should be at 1 unit smaller than inter distance. This will decrease computation complexity quite much. Given a Mahanalobis matrix $\boldsymbol{M}$, for samples $\boldsymbol{x}_i, i = 1, 2, 3, \cdots, n$, n is the number of all samples, the requirement is distance between positive pair should be smaller than the minimum of all negative distance. This can be denoted as

$$D(\boldsymbol{x}_i, \boldsymbol{x}_j) + \rho < \min D(\boldsymbol{x}_i, \boldsymbol{x}_k), y_i = y_j, y_i \neq y_k. \tag{11}$$

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

$$\arg\min \sum_{y_i = y_j} \max\{ D(\boldsymbol{x}_i, \boldsymbol{x}_j) - \min_{y_i \neq y_k} D(\boldsymbol{x}_i, \boldsymbol{x}_k) + \rho \}. \tag{12}$$

Table 1. Optimization algorithm on dimension reduced vectors

```
Gradient optimization algorithm for target function
Input Descriptors of training person pairs
Output A SPD matrix
Initialization
Initialize $\boldsymbol{M}$ with eye matrix $\boldsymbol{I}$;
Compute the initial target function value $f_0$ with $\boldsymbol{M}_0$;
Iteration count $t = 0$;
while(not converge)
Update $t = t + 1$;
Update gradient $\boldsymbol{G}_{t+1}$ with equation 24;
Update $\boldsymbol{M}$ with equation : $\boldsymbol{M}_{t+1} = \boldsymbol{M}_t - \lambda \boldsymbol{G}_t$
Project $\boldsymbol{M}_{t+1}$ to the positive semi-definite space
by $\boldsymbol{M}_{t+1} = \boldsymbol{V}_{t+1}\boldsymbol{S}_{t+1}\boldsymbol{V}_{t+1}^T$;
Update the target value $f|_{\boldsymbol{M}=\boldsymbol{M}_{t+1}}$;
end while
return $\boldsymbol{M}$
```

However, the equation above only penalize small interclass distance. Another term is needed to penalize large intraclass distance. That is, to make the sum of intraclass distance as small as possible. This term is denoted as

$$\min \sum D(\boldsymbol{x}_i, \boldsymbol{x}_j), y_i = y_j. \tag{13}$$

To combine equations above, a ratio factor α is assigned to term 12 so that the target function can be denote as

$$f(\boldsymbol{M}) = (1-\alpha) \sum_{\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{y}_i = y_j} D(\boldsymbol{x}_i, \boldsymbol{x}_j) + \\ \alpha \sum_{\boldsymbol{x}_i, \boldsymbol{x}_j, y_i = y_j} \max\{D(\boldsymbol{x}_i, \boldsymbol{x}_j) - \min_{y_i \neq y_k} D(\boldsymbol{x}_i, \boldsymbol{x}_k) + \rho, 0\} \tag{14}$$

In this way the problem is transformed to an optimization problem. Notice that equation 9 can be denoted as

$$D(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{M} (\boldsymbol{x} - \boldsymbol{y}) = Tr(\boldsymbol{M}\boldsymbol{X}_{i,j}) \tag{15}$$

where $\boldsymbol{X}_{i,j} = \boldsymbol{x}_i * \boldsymbol{x}_j^T$, and Tr is to compute matrix trace. Therefore, equation 14 can be transformed as follow,

$$f(\boldsymbol{M}) = (1-\alpha) \sum_{y_i = y_j} Tr(\boldsymbol{M}\boldsymbol{X}_{i,j}) \\ + \alpha \sum_{y_i = y_j, y_i \neq y_k} \max\{Tr(\boldsymbol{M}\boldsymbol{X}_{i,j}) - Tr(\boldsymbol{M}\boldsymbol{X}_{i,k}) + \rho, 0\} \tag{16}$$

To minimize equation 23, the gradient descent method is used. The gradient respect to $\boldsymbol{M}$ is computed as

$$\boldsymbol{G} = \frac{\partial f}{\partial \boldsymbol{M}} = (1-\alpha) \sum_{y_i = y_j} \boldsymbol{X}_{i,j} \\ + \alpha \sum_{y_i = y_j, y_i \neq y_k} (\boldsymbol{X}_{i,j} - \boldsymbol{X}_{i,k}) \tag{17}$$

The iteration process is summarized as in Table 1;

5. Experiment

The hierarchical gaussian descriptors [4] are used in this paper. There are two versions of gaussian of gaussian descriptor. The first one is extracted only in RGB color space, denoted as GOGrgb. While the second one is extracted from four color space {RGB, HSV, Lab, nRGB}. nRGB means normalized RGB color space by equation

$$nR = \frac{R}{R+G+B}, nG = \frac{G}{R+G+B}, nB = \frac{B}{R+G+B}. \tag{18}$$

since nB component can be computed by nR and nG, only those first two components are adopted to reduce redundancy. Besides, the cumulative matching curve(CMC) is used to measure metric performance in this paper.

5.1. Datasets and evaluation settings

VIPeR 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 into 48×128. In this experiment the we randomly select 316 individuals from cam a and cam b as the training set, the rest images in cam a are used as probe images and those in cam b as gallery images. This process is repeated 10 times to reduce error.

CUHK1 CUHK01 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 into 60×160. In this paper, we randomly select 485 image pairs as training data and the rest person pairs are used for test data.

Prid_2011 The dataset consists of images extracted from multiple person trajectories recorded from two different, static surveillance cameras. Images from these cameras contain a viewpoint change and a stark difference in illumination, background and camera characteristics. Camera view A shows 385 persons, camera view B shows 749 persons. The first 200 persons appear in both camera views. In this paper, we randomly select 100 persons that appeared in both camera views as training pairs, and the remaining 100 persons of camera A is used as probe set while the 649 remaining persons from camera B are used for gallery images.

Prid_450s The 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 parts of a person. The images are stored in two folders that represent the two camera views. In this test, we randomly select 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 There are two camera views in this dataset. Folder probe contains 250 probe images captured in one view. 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. In this paper, we randomly select 125 persons from those 250 persons appeared in both camera views as training pairs, and the remaining persons in probe folder is used as probe images while the remaining 125 persons and those 775 additional persons from gallery folder are used as gallery images.

5.2. The influence of mean removal and L_2 normalization

In [4], mean removal and L_2 normalization is found to improve performance by 5.1%. The reason for this is mean removal and normalization can reduce the impact of extrema in a single descriptor. A comparison between performance of original descriptors and preprocessed descriptors is shown in Tables [2, 3, 4, 5, 6], all those datasets are tested by proposed metric. Original GOG means no mean removal and normalization. It shows that the mean removal and normalization has a slight improvement around 0.5% on the performance on all five datasets. In [4] the mean removal and normalization is adopted, to compare with results in this paper, the mean removal and normalization are also adopted in this experiment.

Table 2. The influence of data preprocessing on VIPeR

	Rank(%)				
Terms	1	5	10	15	20
Original GOG	43.01	74.91	84.87	89.81	93.32
Preprocessed GOGrgb	43.77	74.84	85.25	90.32	93.89
Original GOGfusion	48.77	77.47	87.41	91.52	94.27
Preprocessed GOGfusion	48.32	76.90	87.78	91.93	94.49

Table 3. The influence of data preprocessing on CUHK1

	Rank(%)				
Terms	1	5	10	15	20
Original GOGrgb	56.11	83.77	90.10	92.65	94.28
Preprocessed GOGrgb	55.91	84.24	90.41	93.15	94.67
Original GOGfusion	57.10	84.65	90.35	92.88	94.65
Preprocessed GOGfusion	56.67	84.49	90.51	93.31	94.84

Table 4. The influence of data preprocessing on prid_2011

	Rank(%)				
Terms	1	5	10	15	20
Original GOGrgb	24.80	52.10	63.20	69.90	72.90
Preprocessed GOGrgb	23.80	52.20	63.50	70.20	73.50
Original GOGfusion	32.20	56.60	67.00	73.10	77.70
Preprocessed GOGfusion	32.30	57.40	66.30	73.40	78.00

Table 5. The influence of data preprocessing on prid_450s

	Rank(%)				
Terms	1	5	10	15	20
Original GOGrgb	60.93	84.31	91.29	94.00	96.18
Preprocessed GOGrgb	60.71	84.53	91.29	94.13	96.27
Original GOGfusion	63.07	86.67	92.53	95.20	96.98
Preprocessed GOGfusion	62.80	86.58	92.36	95.29	96.89

Table 6. The influence of data preprocessing on GRID

	Rank(%)				
Terms	1	5	10	15	20
Original GOGrgb	22.96	41.92	51.68	58.72	64.64
Preprocessed GOGrgb	22.64	43.68	52.00	59.04	65.04
Original GOGfusion	24.32	44.56	54.80	62.40	66.64
Preprocessed GOGfusion	23.92	44.64	54.88	62.32	66.40

5.3. Parameters setting of gradient descent iteration

In this experiment, there are a few parameters for the iteration computing including slack variable ρ, maximal iteration T, gradient step λ, the inter and intraclass limitation factor α and the updating ratio β. Firstly the slack variable ρ is initialized as 1 to ensure the minimum interclass distance is 1 larger than intraclass distance at least. The step size of gradient updating λ is initialized as 0.01. When target value f increases, λ is scaled by a factor 0.5, and λ is scaled by 1.01 when target value f decreases. To judge if target value converges, the thresh β is defined as the ratio target value change versus previous target value, that is, $\beta = \frac{(f_{t+1}-f_t)}{f_t}$. According many experiment trials, when it satisfies $\beta = 10^{-5}$, the target value converges and the iteration is stopped. The maximal iteration times is set to 100 since the target value f will converge in around 15 iterations. The last parameter for the iteration is α, to know the best value for α, we tried 11 different values ranges from 0 to 1 with a step of 0.1, and find that the rank 1 and rank 5 scores reach maxima at interval $[0.7, 0.8]$. Then another ten trials with alpha ranging from $[0.7, 0.8]$ with a step of 0.01. 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 7. Parameters setting

Paramters	α	thresh	step	Max iteration	slack variable
Values	0.76	10^{-5}	0.01	100	1

5.4. Performance analysis

In this paper, we compare proposed metric with other state-of-the-art metrics including NFST [14], XQDA [3]. NFST is a metric which learn a null space for descriptors so that the 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 problems in person re-identification. XQDA is quite similar with many other metrics, which learns a projection matrix W and then a Mahanalobis SPD matrix M is learned in the subspace. Those two metric are proved to have state-of-the-art performance compared with many other methods. The GOGrgb in all forms stands for the hierarchical gaussian descriptor in RGB color space while GOGfusion stands for the one in four different color spaces {RGB, Lab, HSV,

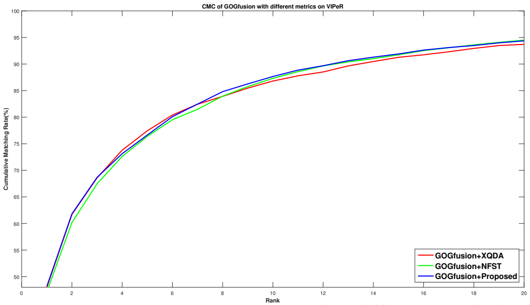

Figure 1. CMC curves on VIPeR comparing different metric learning

nRnG}.

VIPeR A comparison form is given in Table 8. 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 GOGrgb and GOGfusion. 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 GOGrgb+XQDA, and the rank 1, rank 5, rank 10, rank 15 and rank 20 GOGfusion scores of proposed metric learning are 0.35%, -0.54%, 0.98%, 0.66%, 0.79% higher than GOGfusion + XQDA respectively. Also we can see that the proposed metric learning has a better performance than NFST.

Table 8. Performance of different metrics on VIPeR

	Rank(%)				
Methods	1	5	10	15	20
GOGrgb+NFST	43.23	73.16	83.64	89.59	92.88
GOGrgb+XQDA	43.01	73.92	83.86	89.24	92.37
GOGrgb+Proposed	43.77	74.84	85.25	90.32	93.89
GOGfusion+NFST	47.15	76.39	87.31	91.74	94.49
GOGfusion+XQDA	47.97	77.44	86.80	91.27	93.70
GOGfusion+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 GOGrgb combined with proposed metric are 5.4%, 4.18%,3.31%,2.16%,1.46% higher 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 GOGfusion combined with proposed metric are 4.57%, 2.64%, 0.70%, 1.33%, 0.83% higher than GOGfusion combined with XQDA, and 0.41%, 0.83%, 0.88%, 1.09%, 1.14% than GOGfusion combined with NFST.

Table 9. Performance of different metrics on CUHK1

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

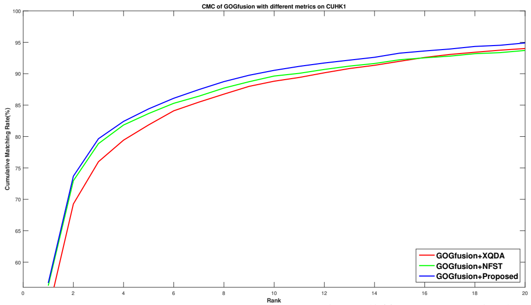

Figure 2. CMC curves on CUHK1 comparing different metric learning

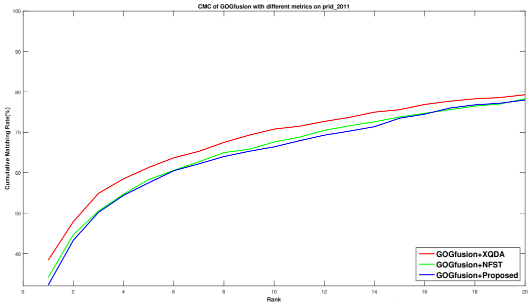

Figure 3. CMC curves on prid_2011 comparing different metric learning

Table 10. Performance of different metrics on prid_2011

	Rank(%)				
Methods	1	5	10	15	20
GOGrgb+NFST	26.60	53.80	62.90	71.30	75.40
GOGrgb+XQDA	31.10	55.70	66.10	72.40	76.10
GOGrgb+Proposed	23.80	52.20	63.50	70.20	73.50
GOGfusion+NFST	34.10	58.30	67.60	73.80	78.30
GOGfusion+XQDA	38.40	61.30	70.80	75.60	79.30
GOGfusion+Proposed	32.30	57.40	66.30	73.40	78.00

Prid_2011 The rank 1, rank5, rank 10, rank 15, rank 20 score of GOGfusion combined with proposed metric are 6.1%, 3.9%, 4.5%, 2.2% and 1.3% lower than GOGfusion combined with XQDA. The performance of NFST is slightly better than proposed metric. Also in terms of GOGrgb XQDA and NFST has better performance than the proposed one. So in this dataset the proposed metric has worse performance than XQDA and NFST.

Table 11. Performance of different metrics on prid_450s

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

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.

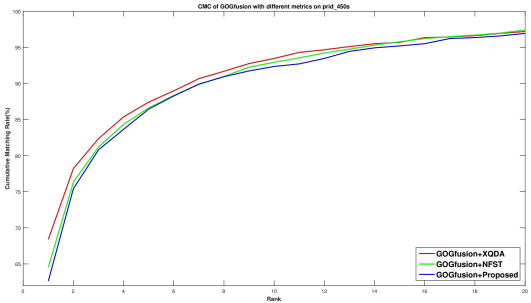

Figure 4. CMC curves on prid_450s comparing different metric learning

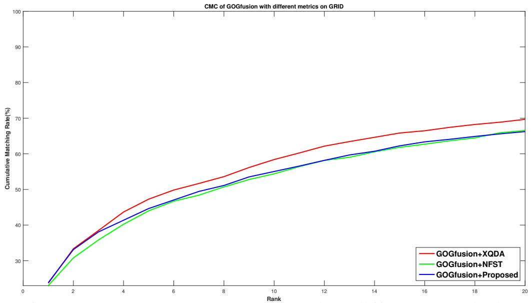

Figure 5. CMC curves on GRID comparing different metric learning

Table 12. Performance of different metrics on GRID

	Rank(%)				
Methods	1	5	10	15	20
GOGrgb+NFST	21.84	41.28	50.96	57.44	62.88
GOGrgb+XQDA	22.64	43.92	55.12	61.12	66.56
GOGrgb+Proposed	22.64	43.68	52.00	59.04	65.04
GOGfusion+NFST	23.04	44.40	54.40	61.84	66.56
GOGfusion+XQDA	23.68	47.28	58.40	65.84	69.68
GOGfusion+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 GOGfusion, 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.

6. Conclusion

In this paper we combined KLFDA with gradient descent method based metric learning. A semi-positive definite(SPD) matrix is learned on the lower dimension space after dimension reduction by kernel local fisher discriminative analysis. By analysis we can find the proposed metric has better performance than NFST and XQDA on VIPeR and CUHK1 datasets, but XQDA and NFST outperforms the proposed metric learning on Prid_2011 and Prid_450s, and the proposed metric learning has better rank 1 score than NFST and its performance is only second to XQDA on GRID dataset.

References

[1] M. Sugiyama, “Local Fisher Discriminant Analysis for Supervised Dimensionality Reduction,” Dec. 2016.

[2] J. You, A. Wu, X. Li, and W.-S. Zheng, “Top-push Video-based Person Re-identification,” pp. 1–9, Jan. 2017.

[3] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, “Person Re-identification by Local Maximal Occurrence Representation and Metric Learning.” CVPR, Apr. 2015, pp. 1–10.

[4] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, “Hierarchical Gaussian Descriptor for Person Re-Identification,” Dec. 2016, pp. 1–10.

[5] O. Tuzel, F. Porikli, and P. Meer, “Region Covariance: A Fast Descriptor for Detection and Classification,” Dec. 2016, pp. 1–14.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person Re-Identification by Symmetry-Driven Accumulation of Local Features.” CVPR, Mar. 2016, pp. 1–8.

[7] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large Scale Metric Learning from Equivalence Constraints,” pp. 1–8, Apr. 2012.

[8] 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)*. IEEE, Jan. 2013, pp. 3318–3325.

[9] A. Mignon and F. Jurie, “PCCA: A New Approach for Distance Learning from Sparse Pairwise Constraints,” pp. 1–7, Apr. 2012.

[10] W. Zheng, S. Gong, and T. Xiang, “Person Re-identification by Probabilistic Relative Distance Comparison,” pp. 1–8, Nov. 2016.

[11] Weinberger, K. Q, Saul, and L. K, “Distance Metric Learning for Large Margin Nearest Neighbor Classification,” pp. 1–38, Feb. 2009.

[12] M. Welling, “Kernel Canonical Correlation Analysis,” pp. 1–3, Mar. 2005.

[13] F. Xiong, M. Gou, O. Camps, and S. Mario, “Person Re-Identification using Kernel-based Metric Learning Methods,” pp. 1–16, Jul. 2014.

[14] L. Zhang, T. Xiang, and S. Gong, “Learning a Discriminative Null Space for Person Re-identification,” pp. 1–10, Mar. 2016.

[15] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-theoretic metric learning,” pp. 1–8, Jun. 2007.

[16] L. Yang, “Distance Metric Learning: A Comprehensive Survey,” pp. 1–51, May 2006.

[17] S. G. C. C. L. Chunxiao Liu and X. Lin, “LNCS 7583 - Person Re-identification: What Features Are Important?” pp. 1–11, Nov. 2015.

[18] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, “Learning to Rank in Person Re-Identification With Metric Ensembles,” pp. 1–10, Jan. 2015.

[19] P. Li, Q. Wang, and L. Zhang, “A Novel Earth Mover’s

Distance Methodology for Image Matching with Gaussian Mixture Models," in *2013 IEEE International Conference on Computer Vision (ICCV)*. IEEE, Dec. 2013, pp. 1689–1696.

[20] Z. Shi, T. M. Hospedales, and T. Xiang, "Transferring a Semantic Representation for Person Re-Identification and Search," pp. 1–10, Apr. 2015.

[21] L. Wu, C. Shen, and A. van den Hengel, "PersonNet: Person Re-identification with Deep Convolutional Neural Networks," pp. 1–7, Jun. 2016.

[22] N. McLaughlin, J. M. del Rincon, and P. Miller, "Recurrent Convolutional Network for Video-Based Person Re-Identification," pp. 1–10, Jul. 2016.