

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

Qiangsen He

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

**Abstract**—Person re-identification has been an intense research area recently. It's very important to choose robust descriptors and metric learning to improve accuracy. Mahalanobis based metric learning is a popular method to measure similarity. However, since directly extracted descriptors usually have high dimension, it's intractable to learn a high dimensional Mahalanobis matrix. Dimension reduction are used to project high dimensional descriptors to lower dimension space while preserving those discriminative information as much as possible. In this paper the kernel LFDA is used to reduce dimension given that kernelization method can greatly improve re-identification performance for nonlinearity. Then a metric matrix is learned on lower dimensional descriptors based on the limitation that the within class distance is at least 1 unit smaller than the minimum inter class distance. 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. The task of re-ID is to match a given person with a list of identities with known labels. This problem is very challenging caused by many factors like low image resolution, occlusion, background noise and different camera color response, etc. According to sample size of each person, re-ID can be divided into single shot and multi-shot re-ID. 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 multishots case, there might exist quite much difference even in different frames of the same person for different pose and illuminations. Therefore, good descriptors are supposed to be robust to illumination change and occlusions.

Since most datasets are well images with well cropped bounding boxes, classical re-ID focus on descriptors extraction and metric learning. It's important to select a proper descriptor for images. Most descriptors are color and texture based descriptors. One simple descriptor is the histogram descriptor. In [] 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 and is not robust to person with similar color but different texture. In [SDALF] the symmetry and asymmetry of foreground is considered. The maximally stable color region is also exploited in this paper. In [covariance] the covariance descriptors of local patches are used to represent people.

Given descriptors of individuals, a matching algorithm is needed to re-identify people. To compute similarity, the Mahalanobis distance based metric learning is very popular. Many works have been focusing on learning a Mahalanobis based matrix to improve performance. That is, suppose there are two d-dimensional vectors of two persons,  $x, y$ , a matrix  $M$  is used to calculate the distance  $D(x, y) = (x - y)^T M (x - y)$ . The matrix  $M$  can be learned by many limitations like LMNN[].

For descriptors with high dimension( $d \geq 1000$ ), it's hard to directly learn a Mahalanobis distance matrix  $M$  for the small sample size  $n(n \ll d)$ . 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. Linear discriminant analysis(LDA) is also used to reduce dimension. There are a few differences between LDA and PCA, LDA is a supervised method while PCA is unsupervised, and LDA is more used independently while PCA is more used as a preprocessing method. The advantage of LDA is it distinguishes different classes by maximize the ratio of inter-class scatter matrix versus intra-class scatter matrix. The local fisher discriminant analysis(LDFA) combines LDA and local projection preserving(LPP) to exploit the correlation of neighbor points of sample points. Moreover, the kernelization version of LFDA proposed to improve performance.

In this paper, instead of using KLFDA as a unique subspace learning method, we use KLFDA to reduce dimension. Again with the lower dimension data, a Mahalanobis distance matrix  $M$  is learned based on the limitation that intra distance is at least 1 unit smaller than inter class

distance. This is inspired by the [TDL] paper. A target function respect  $M$  is created to penalize big intra class distance and small inter class 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 [] the images are divided into a few horizontal stripes and the color histogram on HSV color space is extracted for each stripe. This descriptor is mainly useful for viewpoint invariant datasets, but since dividing image into many stripes may lose certain spatial information, its robustness is much weakened. In [SDALF] the symmetry and asymmetry property of pedestrian foreground is considered. But this descriptor depends much on foreground extraction performance. In [LOMO] densely sampled overlapping small windows are adopted to concur 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 [GOG] 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 [PRDC] 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 [NFST] the author proposes 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 [Metric ensemble] 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 [EMD] 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 [semantic] the author presents a new semantic at-tribute learning approach for person re-identification and search but this method suffers from low performance. In [XQDA] the author proposed to learn a Mahanalobis distance matrix on the subspace, 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 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

## 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  $\mathbf{x}_i$ , where  $i \in \{1, 2, \dots, n\}$ , the label  $l_i \in \{1, 2, \dots, l\}$ . Two matrix are defined as the within class scatter matrix  $\mathbf{S}^{(w)}$  and between class matrix  $\mathbf{S}^{(b)}$ ,

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

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

$$\begin{aligned}\boldsymbol{\mu}_i &= \frac{1}{n_i} \sum \mathbf{x}_i \\ \boldsymbol{\mu} &= \frac{1}{n} \sum \mathbf{x}_i\end{aligned}\quad (2)$$

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

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

Fisher discriminant analysis tries to minimize the within-class distance while maximize the between class distance. The  $\mathbf{T}$  is computed by the eigenvalue decomposition so that the between class scatter is maximized and the within class scatter matrix is minimized.  $\mathbf{T}$  can be represented as the set of all the corresponding eigenvectors, as  $\mathbf{T} = (\phi_1, \phi_2, \dots, \phi_k)$ .

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[Hexiaofoei] locality preserving projection is proposed to exploit data locality. In LPP an affinity matrix is created to record the affinity of sample  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , typically the range of elements in  $\mathbf{A}_{i,j}$  is  $[0, 1]$ . There are many manners to define a  $n \times n$  affinity matrix  $\mathbf{A}$ , 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  $x_i$  is within k-nearest neighbours of  $x_j$  then  $A_{i,j} = 1$  otherwise  $A_{i,j} = 0$ . Another diagonal matrix  $D$  can be defined that each diagonal element is the sum of corresponding column in  $A$ ,

$$D_{i,i} = \sum_{j=1}^n A_{i,j} \quad (4)$$

then the LPP transform matrix is defined as follow,

$$T_{LPP} = \arg \min_{T \in \mathbb{R}^{d \times m}} \frac{1}{2} \sum_{i,j=1}^n A_{i,j} \|T^T x_i - T^T x_j\| \quad (5)$$

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

$$T_{LPP} = \{\phi_{d-m+1} | \phi_{d-m+1} | \dots | \phi_d\}$$

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

$$X L X^T \phi = \gamma X D X^T \phi \quad (6)$$

where  $\gamma$  is corresponding eigenvalue of  $\phi$ , 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  $A^{(w)}$  and  $A^{(b)}$ , so that,

$$\begin{aligned} S^{(w)} &= \frac{1}{2} \sum_{i=1}^l \sum_{j:l_j=i} A_{i,j}^w (x_j - \mu_i)(x_j - \mu_i)^T \\ S^{(b)} &= \frac{1}{2} \sum_{i=1}^l A_{i,i}^b (\mu_i - \mu)(\mu_i - \mu)^T \end{aligned} \quad (7)$$

where

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

where  $y_i$  is the class label of sample point  $x_i$ .

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 in [ ] since the non-linearity is exploited. In [ ] 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 within class scatter matrix will be transformed into kernel space and the eigenvalue decomposition will be operated on kernel space. Suppose a set of sample points  $x_i, i \in \{1, 2, \dots, n\}$ , can be mapped to a implicit higher feature space by a function  $\phi(x_i)$ . It has been proved that kernel function can be implicit and only the inner product of mapped vectors  $\phi(x_i)$  and  $\phi(x_j)$  need to be known. The kernel trick is proposed to solve this problem by defining a function  $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$ , the  $\langle \cdot \rangle$  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  $k_{RBF}(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ .

### 4. Mahalanobis metric learning on vectors with lower dimension

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

$$D(x, y) = (x - y)^T M (x - y), \quad (9)$$

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

$$D(x, y) = (x - y)^T W^T W (x - y) = \|W(x - y)\| \quad (10)$$

Therefore, Mahalanobis 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 intra class distance should be at 1 unit smaller than inter distance. This will decrease computation complexity quite much. Given a Mahalanobis matrix  $M$ , for samples  $x_i, i = 1, 2, 3, \dots, n$ ,  $n$  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

$$D(x_i, x_j) + \rho < \min_{y_i = y_j, y_i \neq y_k} D(x_i, x_k), \quad (11)$$

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

$$\arg \min_{y_i = y_j} \sum \max\{D(x_i, x_j) - \min_{y_i \neq y_k} D(x_i, x_k) + \rho\}. \quad (12)$$

Table 1. Optimization algorithm on dimension reduced vectors

Gradient optimization algorithm for target function
<b>Input</b> Descriptors of training person pairs
<b>Output</b> A SPD matrix
<b>Initialization</b>
Initialize $M$ with eye matrix $I$ ;
Compute the initial target function value $f_0$ with $M_0$ ;
Iteration count $t = 0$ ;
<b>while</b> (not converge)
Update $t = t + 1$ ;
Update gradient $G_{t+1}$ with equation 24;
Update $M$ with equation : $M_{t+1} = M_t - \lambda G_t$
Project $M_{t+1}$ to the positive semi-definite space
by $M_{t+1} = V_{t+1} S_{t+1} V_{t+1}^T$ ;
Update the target value $f M=M_{t+1}$ ;
<b>end while</b>
<b>return</b> $M$

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

$$\min \sum D(x_i, x_j), y_i = y_j. \quad (13)$$

To combine equations above, a ratio factor  $\alpha$  is assigned to equation [] so that the target function can be denote as

$$f(M) = (1 - \alpha) \sum_{x_i, x_j, y_i = y_j} D(x_i, x_j) + \alpha \sum_{x_i, x_j, y_i = y_j} \max\{D(x_i, x_j) - \min_{y_i \neq y_k} D(x_i, x_k) + \rho, 0\} \quad (14)$$

In this way the problem is transformed to an optimization problem. Notice that equation 16 can be denoted as

$$D(x, y) = (x - y)^T M (x - y) = Tr(M X_{i,j}) \quad (15)$$

where  $X_{i,j} = x_i * x_j^T$ , and  $Tr$  is to compute matrix trace. Therefore, equation 21 can be transformed as follow,

$$G = f(M) = (1 - \alpha) \sum_{y_i = y_j} Tr(M X_{i,j}) + \alpha \sum_{y_i = y_j, y_i \neq y_k} \max\{Tr(M X_{i,j}) - Tr(M X_{i,k}) + \rho, 0\} \quad (16)$$

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

$$\frac{\partial f}{\partial M} = (1 - \alpha) \sum_{y_i = y_j} X_{i,j} + \alpha \sum_{y_i = y_j, y_i \neq y_k} (X_{i,j} - X_{i,k}) \quad (17)$$

The iteration process can be summarized as in table [];

## 5. Experiment

The hierarchical gaussian descriptors [] 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

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

since nB component can be computed by nR and nG, only those first two components are adopted to prevent redundancy.

### 5.1. Datasets and evaluation settings

**VIPeR** VIPeR dataset is the most used dataset in person re-identification. 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  $48 \times 128$ . 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  $60 \times 160$ . 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 persons 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. Folder gallery 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.

Table 2. Testing setting for different datasets

Dataset	training	probe	gallery	cam_a	cam_b
VIPeR	316	316	316	632	632
CUHK1	485	486	486	971	971
PRID_2011	100	100	649	385	749
PRID_450s	225	225	225	450	450
GRID	125	125	900	250	1025

## 5.2. The influence of mean removal and $L_2$ normalization

In [GOG], 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 extremas of descriptors. When testing proposed metric learning, we find the mean removal can slightly improve performance. 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. The 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. Since preprocessing are required to test XQDA, the mean removal and normalization are operated on descriptors in this experiment.

Table 3. The influence of data preprocessing on VIPeR

Terms	Rank(%)				
	1	5	10	15	20
Original GOG	43.32	74.78	85.00	89.94	93.39
Preprocessed GOGrgb	43.73	74.75	85.41	90.28	93.86
Original GOGfusion	48.67	77.41	87.41	91.65	94.34
Preprocessed GOGfusion	48.10	76.90	87.59	91.90	94.40

Table 4. The influence of data preprocessing on CUHK1

Terms	Rank(%)				
	1	5	10	15	20
Original GOGrgb	56.15	83.79	90.08	92.63	94.26
Preprocessed GOGrgb	55.86	84.28	90.45	93.09	94.65
Original GOGfusion	57.00	84.55	90.37	92.82	94.69
Preprocessed GOGfusion	56.69	84.40	90.53	93.27	94.90

Table 5. The influence of data preprocessing on prid\_2011

Terms	Rank(%)				
	1	5	10	15	20
Original GOGrgb	24.70	51.80	63.30	69.60	72.70
Preprocessed GOGrgb	23.80	52.10	63.50	70.20	73.50
Original GOGfusion	32.40	56.80	66.80	73.10	77.70
Preprocessed GOGfusion	32.20	57.50	66.40	73.50	78.00

Table 6. The influence of data preprocessing on prid\_450s

Terms	Rank(%)				
	1	5	10	15	20
Original GOGrgb	61.02	84.22	91.33	94.09	96.22
Preprocessed GOGrgb	60.44	84.44	91.33	94.00	96.13
Original GOGfusion	62.89	86.62	92.53	95.29	96.89
Preprocessed GOGfusion	62.62	86.44	92.36	95.20	96.93

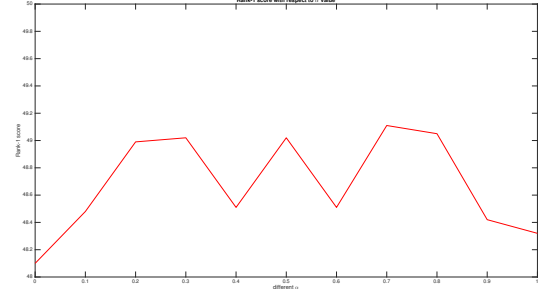
Figure 1. Rank 1 scores with respect to  $\alpha$  on VIPeR

Table 7. The influence of data preprocessing on GRID

Terms	Rank(%)				
	1	5	10	15	20
Original GOGrgb	22.96	42.00	51.76	58.72	64.64
Preprocessed GOGrgb	22.80	43.76	52.08	59.04	65.12
Original GOGfusion	24.32	44.40	54.96	62.40	66.56
Preprocessed GOGfusion	23.84	44.64	55.04	62.24	66.24

## 5.3. Parameters setting of gradient descent iteration

In this experiment, there are a few parameters for the iteration computing including slack variable  $\rho$ , maximal iteration  $T$ , gradient step  $\lambda$ , the inter and intra class limitation factor  $\alpha$  and the updating ratio  $\beta$ . Firstly the slack variable  $\rho$  is initialized as 1 to ensure the minimum inter class distance is 1 larger than intra class distance at least. The step size of gradient updating  $\lambda$  is initialized as 0.01. When target value  $f$  increases,  $\lambda$  is scaled by a factor 0.5, and  $\lambda$  is scaled by 1.01 when target value  $f$  decreases. To judge if target value converges, the thresh  $\beta$  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  $\alpha$ , to know the best value for  $\alpha$ , we tried 11 different values ranges from 0 to 1 with a step of 0.1, and the rank-1 and rank-5 scores of responding  $\alpha$  is shown in figure []. The best  $\alpha$  value should have as large top rank scores as possible. By comparison,  $\alpha$  is set as 0.7. A form of all parameters are shown in Form 7.

Table 8. Parameters setting

Paramters	$\alpha$	thresh	step	Max iteration	slack variable
Values	0.7	$10^{-5}$	0.01	100	1

**Performance measuring** The cumulative matching curve is used to measure the descriptor performance. The score means the probability that the right match is within the top  $n$  samples. A perfect CMC curve is expected to have a high rank-1 value and reaches 1 as fast as possible.

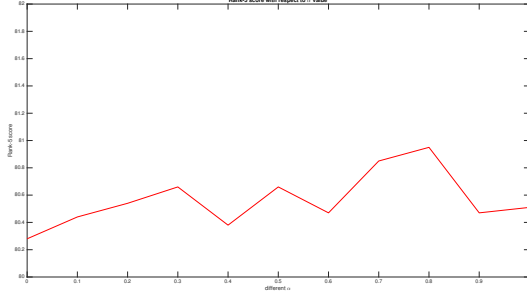


Figure 2. Rank 5 scores with respect to  $\alpha$  on VIPeR

#### 5.4. Performance analysis

In this paper, we compare proposed metric with other state-of-the-art metrics including NFST[], XQDA[]. 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 within class 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 Mahalanobis SPD matrix  $M$  is learned in the subspace. Those two metric are proved to have state-of-the-art performance 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,nRnR}.

**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 GOGrgb scores of proposed metric learning are 0.44%, 0.72%, 1.27%, 1.3%, 1.47% 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 0.19%, -0.79%, 0.86%, 0.63%, 0.67% higher than GOGfusion + XQDA respectively. Also we can see that the proposed metric learning has a way more better performance than NFST. We can infer that the performance of KLFDA is better than XQDA, with its rank 1 score 1.6% higher.

Table 9. Performance of different metrics on VIPeR

Methods	Rank(%)				
	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	43.48	74.59	85.35	90.47	93.67
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	48.29	76.84	87.88	91.90	94.62

**CUHK1** We can find that the rank 1, rank5, rank 10, rank 15, rank 20 score of GOGrgb combined with proposed

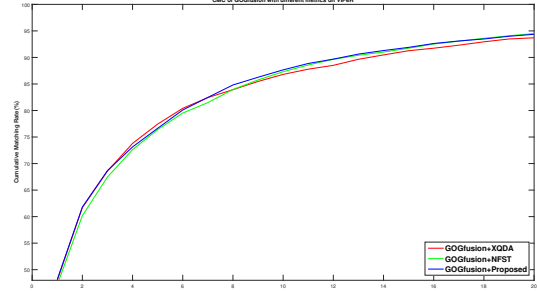


Figure 3. CMC curves on VIPeR comparing different metric learning

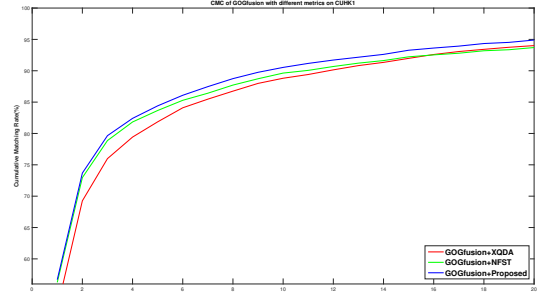


Figure 4. CMC curves on CUHK1 comparing different metric learning

metric are 5.35%, 4.22%, 3.35%, 2.1%, 1.44% higher than XQDA, and 0.26%, 1.26%, 1.38%, 1.11%, 1.09% than NFST. Also the rank 1, rank5, rank 10, rank 15, rank 20 score of GOGfusion combined with proposed metric are 4.59%, 2.55%, 0.72%, 1.29%, 0.89% higher than GOGfusion combined with XQDA, and 0.4%, 0.74%, 0.6%, 1.05%, 1.2% than GOGfusion combined with NFST.

Table 10. Performance of different metrics on CUHK1

Methods	Rank(%)				
	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	55.86	84.28	90.45	93.09	94.65
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	56.69	84.40	90.53	93.27	94.90

Table 11. Performance of different metrics on prid\_2011

Methods	Rank(%)				
	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	52.10	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	32.20	57.50	66.40	73.50	78.00

**Prid\_2011** The rank 1, rank5, rank 10, rank 15, rank 20 score of GOGfusion combined with proposed metric are 6.2%, 3.8%, 4.4%, 2.1% and 1.3% lower than GOGfusion combined with XQDA. The performance of NFST

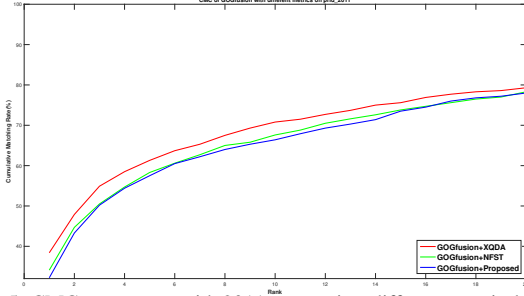


Figure 5. CMC curves on prid\_2011 comparing different metric learning

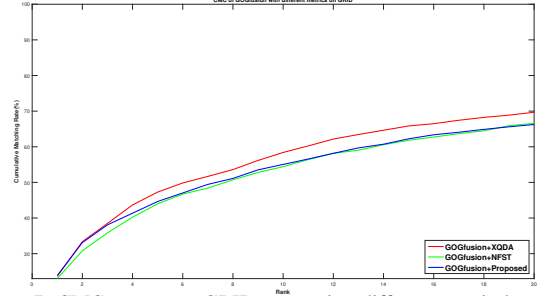


Figure 7. CMC curves on GRID comparing different metric learning

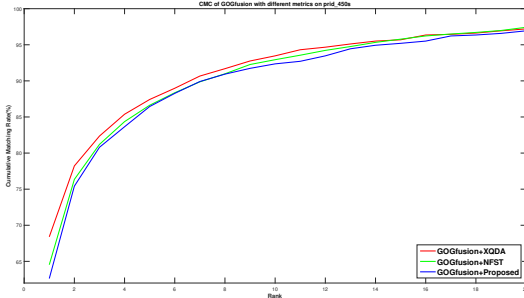


Figure 6. CMC curves on prid\_450s comparing different metric learning

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 12. Performance of different metrics on prid\_450s

Methods	Rank(%)				
	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	60.44	84.44	91.33	94.00	96.13
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	62.89	86.49	92.49	95.29	97.07

**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 descriptors.

Table 13. Performance of different metrics on GRID

Methods	Rank(%)				
	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	22.80	43.76	52.08	59.04	65.12
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	23.84	44.64	55.04	62.24	66.24

**GRID** We can see that the rank 1 score of proposed metric are slightly higher than XQDA and 0.8% higher than NFST in terms of GOGfusion, but XQDA outperforms

proposed metric on rank 5, rank 10, rank 15 and rank 20 scores. But proposed metric outperforms NFST on rank 5, rank 10, rank 15 and rank 20 scores.

## 6. Conclusion

In this paper a 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 XQDA on GRID dataset.

## References

- [1] "Hierarchical Gaussian Descriptor for Person Re-Identification," pp. 1–10, Dec. 2016.