Person re-identification based on 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*—In person re-identification (Re-ID) it's very important to choose robust descriptors and metric learning to improve accuracy. Mahalanobis distance based metric learning is a popular method for metric learning. However, directly extracted descriptors usually have high dimensionality (thousands or more), it's 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 while preserving those discriminative information. However, few work has been done to study the metric learning on those subspace after the first-time metric learning. In this paper the kernel local Fisher discriminant analysis (KLFDA) [1] is used to reduce dimension given that kernelization method can greatly improve Re-ID performance for nonlinearity. Then a Mahalanobis distance metric is learned on lower dimensional descriptors based on the limitation that the intraclass distance is at least one unit smaller than the minimum interclass distance. By comparing the intraclass distance only with the minimum interclass pair the computation complexity of metric learning is reduced. This method turns to have excellent 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 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 confusing when different people have similar pose or clothes. Also, in the multi-shot case, there might exist 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] or metric learning [1, 2, 7–15]. Most descriptors are color and texture based descriptors. Distance metric learning has been fully studied in [**?**] and most existing metric learning are based on Mahalanobis distance to learn a metric which has 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) and high computation complexity on large datasets.

For descriptors with high dimensionality, it's hard to directly learn an SPD matrix $\boldsymbol{M}$ for the small sample size $n(n << d)$. A popular method is to use principal component analysis (PCA) to reduce dimension. PCA is very popular for dimensionality reducing, but one problem is PCA doesn't consider the discriminant information between different classes thus many discriminant information will be lost after dimensionality reduction. In this paper kernel local Fisher discriminant analysis (KLFDA) [1] is used to reduce dimensionality, 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 dimensionality reduced vectors by KLFDA hasn't been fully probed.

The contributions of this paper are as follows. (1) KLFDA and metric learning are combined together to improve Re-ID performance. Previous works mainly use KLFDA as a subspace learning method. Euclidean distance is used to measure the similarity of dimension reduced descriptors. The Mahalanobis distance metric learning on the projected space by KLFDA has not been fully studied. In this paper, metric learning is performed by an iterative computation based on gradient optimization. (2) Inspired by [2], in this work we propose to learn a Mahalanobis distance matrix based on the limitation 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. It's important to compare the intraclass distance only with the minimum interclass distance in each iteration to reduce computation complexity because there is no need to compute every possible possible positive and negative pairs. (3) Extensive experiments are performed on VIPeR, CUHK01, prid_450s and GRID dataset. It proves the proposed work has advanced performance on those datasets.

2. Dimension reduction based on kernel local Fisher discriminant analysis

KLFDA is the kernel version of local Fisher discriminant analysis (LFDA). Here a brief review of LFDA 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 interclass 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,

$$\boldsymbol{\mu}_i = \frac{1}{n_i} \sum \boldsymbol{x}_i, \boldsymbol{\mu} = \frac{1}{n} \sum \boldsymbol{x}_i \tag{2}$$

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 minimizes the intraclass scatter matrix while maximize the interclass scatter matrix. $\boldsymbol{T}$ is computed by the eigenvalue decomposition and $\boldsymbol{T}$ 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 has a similar form with signal and noise ratio. However, the FDA dimension reduction has poor performance for it doesn't consider the locality of data when dealing with multimodality. In [16] 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 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 larger 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 $\boldsymbol{A}_{i,j} = 0$. LFDA combines FDA and LPP and has a more strong performance. The key in LFDA is it assigns weights to elements in $\boldsymbol{A}^{(w)}$ and $\boldsymbol{A}^{(b)}$, so that,

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

where

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

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 computational cost. Suppose the vector data has dimension of d, we have 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.

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 dimensional space to high dimensional space, which makes classification and clustering much more accurate in high dimensional space. 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 can be implicit and only the inner product of mapped vectors $\phi(\boldsymbol{x}_i)$ and $\phi(\boldsymbol{x}_j)$ needs 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 kernels 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 $k_{RBF}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp^{(-\gamma||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)}$, γ is a constant term.

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

Since Re-ID is a problem of ranking, it is desired that the rank 1 descriptor should be the right match. In this paper, instead of comparing all the possible positive and negative pairs, a simplified version is proposed that the intraclass distance should be at least one unit smaller than inter distance. This will decrease computation complexity quite much. 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 distance between positive pair should be at least one unit smaller than the minimum of all negative distance. This can be denoted as

$$D(\boldsymbol{x}_i, \boldsymbol{x}_j) + \rho < \min D(\boldsymbol{x}_i, \boldsymbol{x}_k), y_i = y_j, y_i \neq y_k. \tag{6}$$

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

$$\arg\min \sum_{y_i=y_j} \max\{D(\boldsymbol{x}_i, \boldsymbol{x}_j) - \min_{y_i \neq y_k} D(\boldsymbol{x}_i, \boldsymbol{x}_k) + \rho, 0\}. \tag{7}$$

However, the equation above only penalizes 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

$$\min \sum_{y_i=y_j} D(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{8}$$

To combine equations above, a ratio factor α is assigned to equation 7 so that the target function can be denoted as

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

In this way the problem is transformed to 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, Equation 9 can be transformed as follow,

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

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

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

The iteration process is summarized as in Table 1;

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 $\boldsymbol{I}$;
Compute the initial target function value f_0 with $\boldsymbol{M}_0$;
Iteration count $t = 0$;
while(not converge)
Update $t = t + 1$;
Find $\boldsymbol{x}_k$ for all sample points $\boldsymbol{x}_i$, where $y_i \neq y_k$;
Update gradient $\boldsymbol{G}_{t+1}$ with Equation 12;
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-positive definite space;
Update the target value $f\|_{\boldsymbol{M}=\boldsymbol{M}_{t+1}}$;
end while
return $\boldsymbol{M}$

4. 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 GOG_{rgb}. While the second one is extracted from four color space {RGB, HSV, Lab, nRGB}. nRGB means normalized RGB color space.

4.1. 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 into 48×128. In this experiment the we randomly select 316 individuals from camera a and camera b as the training set, the rest 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 value.

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 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_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.

4.2. The influence of mean removal and L_2 normalization

In [4], mean removal and L_2 normalization is shown 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. Original GOG means no mean removal and normalization. It shows that the mean removal and L_2 normalization has an 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 L_2 normalization are also adopted in this experiment.

4.3. 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 interclass and intraclass limitation factor α and the updating ratio β. Firstly the slack variable ρ is initialized as one to ensure the minimum interclass distance is one 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 threshold β is defined as the ratio target function value change versus previous target function value, that is, $\beta = \frac{(f_{t+1}-f_t)}{f_t}$. According to many experiment trials, when it satisfies $\beta = 10^{-5}$, the target value converges 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 to assign weight to intercalss distance comparison. 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 2. Parameters setting

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

4.4. 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 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 is 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 metrics are proved to have state-of-the-art performance compared with many other methods. The GOG_{rgb} in all forms stands for the hierarchical Gaussian descriptor in RGB color space while GOG_{fusion} stands for the one in four different color spaces {RGB, Lab, HSV, nRnG}.

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 GOG_{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 GOG_{fusion} scores of proposed metric learning are 0.35%, -0.54%, 0.98%, 0.66%, 0.79% higher than GOG_{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
GOG_{fusion}+NFST	47.15	76.39	87.31	91.74	94.49
GOG_{fusion}+XQDA	47.97	77.44	86.80	91.27	93.70
GOG_{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 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.

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

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

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 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. The proposed metric has superior performance for following reasons: (1) dimension reduction by KLFDA exploits the nonlinearity and the loss of discriminant information between classes are minimized. (2) the simplified relative distance limitation optimization helps to confine the Mahalanobis distance matrix $\boldsymbol{M}$ to discriminate different classes.

5. Conclusion

In this paper we combined KLFDA with gradient descent method based metric learning. A SPD matrix is learned on the lower dimension space after dimensionality reduction by KLFDA. By comparison and 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. On GRID dataset the proposed metric learning has better rank 1 score than NFST and its performance is only second to XQDA.

References

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

[2] 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*, Jan. 2016, pp. 1–9.

[3] 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*, Apr. 2015, pp. 1–10.

[4] 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*, Dec. 2016, pp. 1363–1372.

[5] O. Tuzel, F. Porikli, and P. Meer, "Region Covariance: A Fast Descriptor for Detection and Classification," in *European conference on computer vision*. Springer, 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," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 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," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, Apr. 2012, pp. 2288–2295.

[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," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, Apr. 2012, pp. 1–7.

[10] 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*. IEEE, Nov. 2016, pp. 1–8.

[11] Weinberger, K. Q, Saul, and L. K, "Distance Metric Learning for Large Margin Nearest Neighbor Classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 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," in *European conference on computer vision*. Springer, Jul. 2014, pp. 1–16.

[14] 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*, Mar. 2016, pp. 1239–1248.

[15] 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*. ACM, Jun. 2007, pp. 209–216.

[16] X. He and P. Niyogi, "Locality Preserving Projections," in *NIPS*, Nov. 2003, pp. 1–8.