16th International Conference on Machine Vision Applications (MVA)
National Olympics Memorial Youth Center, Tokyo, Japan, May 27-31, 2019.

Kernelized Cross-view Quadratic Discriminant Analysis for Person Re-Identification

Tetsu Matsukawa and Einoshin Suzuki
Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motooka Nishi-ku Fukuoka, 819-0395, Japan
{matsukawa, suzuki}@inf.kyushu-u.ac.jp

Abstract

In person re-identification, Keep It Simple and Straightforward MEtric (KISSME) is known as a practical distance metric learning method. Typically, kernelization improves the performance of metric learning methods. Nevertheless, deriving KISSME on a reproducing kernel Hilbert space is a non-trivial problem. Nyström method approximates the Hilbert space in low-dimensional Euclidean space, and the application of KISSME is straightforward, yet it fails to preserve discriminative information. To utilize KISSME in a discriminative subspace of the Hilbert space, we propose a kernel extension of Cross-view Discriminant Analysis (XQDA) which learns a discriminative low-dimensional subspace, and simultaneously KISSME in the learned subspace. We show with the standard kernel trick, the kernelized XQDA results in the case when the empirical kernel vector is used as the input of XQDA. Experimental results on benchmark datasets show the kernelized XQDA outperforms XQDA and Nyström-KISSME.

1 Introduction

Person re-identification (re-id) [1] is a challenging problem of finding the same persons from images captured in different camera views, which often cause substantial appearance changes, *e.g.*, view angle/body pose/illumination changes, and background clutter.

One of the most critical steps in person re-id is the distance calculation for matching the feature descriptors extracted from person images.

Distance metric learning obtains a robust and discriminative metric from training data. Until now, researchers have developed various metric learning methods and shown promising results in person re-id [2, 3, 4, 5, 6]. Among them, Keep It Simple and Straightforward MEtric (KISSME) [3] is one of the most popular metrics due to its simplicity and effectiveness.

KISSME [3] considers two independent generation processes for modeling the differences of feature descriptors for similar and dissimilar pairs. In particular, it assumes a zero-mean Gaussian distribution for each pair. As in the Quadratic Discriminant Analysis (QDA) in statistics, KISSME performs a likelihood-ratio test of the two Gaussian distributions to judge if an input sample pair is similar or dissimilar. From the likelihood-ratio test, KISSME derives a Mahalanobis metric straightforwardly. In computation, it requires only two inverse covariance matrix estimations and a projection operation of a matrix into a cone of Positive Semi-Definite (PSD) matrices.

The feature descriptors for person re-id tend to be high-dimensional. Besides, the number of training samples is small because obtaining a sufficient number of the samples are mostly infeasible due to the requirement of tedious labeling efforts for human operators. Consequently, KISSME may yield poor estimation of covariance matrices, thus leading to poor generalization ability. A common practice to circumvent this problem was the projection of feature descriptors into a Principal Component Analysis (PCA) subspace before applying KISSME [3]. Instead of PCA, Cross-view Quadratic Discriminant Analysis (XQDA) [5] applies discriminative subspace learning to keep discriminative information in the original feature space, which typically produces better performance with lower-dimensional subspace.

Distance metric learning on Reproducing Kernel Hilbert Space (RKHS), namely, kernelized metric learning method, often produces superior results in person re-id [7]. Zhao et al. [8] showed that the distribution of sample difference on existing person re-id features is irregular and undulant while the difference on the kernel transformed feature vectors tends to obey a Gaussian distribution. Based on this analysis, they used empirical kernel vectors as the input of KISSME and showed improved results.

However, compared to other metric learning methods, kernelizing KISSME is a non-trivial problem because it has difficulty in handling the inverse covariance matrix and the projection operation in RKHS [9, 10]. The corresponding KISSME algorithm in the Hilbert space is unclear for the kernelized KISSME which only uses the empirical kernel vectors as input [8]. Faraki et al. [9] proposed a kernelized version of KISSME via infinite dimensional covariance matrices, yet the strict kernelized version adopts an iterative optimization on a Riemannian manifold. Consequently, it is much more complicated than KISSME in the original feature space. Nyström method [11] approximates RKHS in low-dimensional Euclidean space and maintains the simplicity of KISSME, yet fails to preserve discriminative information in RKHS.

To utilize KISSME in a discriminative subspace of Hilbert space, we propose a kernel extension of XQDA. We show that with the standard kernel trick, the kernelized XQDA coincidences with XQDA when the empirical kernel vector is the input of XQDA. We think the reason the kernel version of XQDA remains unexplored is probably due to the lack of theoretical understanding behind XQDA. As a supplement of this problem, we also examine the suitability of the discriminative subspace of XQDA for any of Mahalanobis metrics.

2 Review of XQDA

We briefly review XQDA, which is built on two components; KISSME and discriminative subspace learning through maximization of generalized Rayleigh quotient criterion. At a glance, the suitability of the subspace for the KISSME might be not obvious because the generalized Rayleigh criterion is unaware of the Mahalanobis metric. The following subsection describes the respective components of XQDA. We then examine the suitability of generalized Rayleigh quotient criterion for any of Mahalanobis metrics

2.1 KISSME

Assume we have a training dataset $\{\boldsymbol{x}_i, l_i, c_i\}_{i=1}^N$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ is a d dimensional sample, $l_i \in \{1, ..., L\}$ is a person ID and L is the number of persons, $c_i \in \{1, ..., C\}$ is a camera ID and C is the number of camera views and N is the number of training samples. Let $S = \{(i,j)|l_i = l_j, c_i \neq c_j\}$, $D = \{(i,j)|l_i \neq l_j, c_i \neq c_j\}$ be sample index sets of similar and dissimilar pairs on the training dataset, respectively [1].

The squared Mahalanobis distance between samples $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is defined as

$$d^2_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j), \tag{1}$$

where the Mahalanobis matrix $\boldsymbol{M} \succeq 0$ is a PSD matrix.

In KISSME [3] and XQDA [5], $\boldsymbol{M}$ is learned based on a likelihood-ratio test of two zero-mean Gaussian distributions, of similar and dissimilar pairs. The dissimilarity hypothesis is defined as follows:

$$\epsilon(i,j) = \log \left[\frac{\frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_D|}} \exp(-\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{\Sigma}_D^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)}{2})}{\frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}_S|}} \exp(-\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{\Sigma}_S^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)}{2})} \right], \tag{2}$$

where $\boldsymbol{\Sigma}_D = \frac{1}{N_D} \sum_{(i,j)\in D} (\boldsymbol{x}_i - \boldsymbol{x}_i)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T$ and $\boldsymbol{\Sigma}_S = \frac{1}{N_S} \sum_{(i,j)\in S} (\boldsymbol{x}_i - \boldsymbol{x}_i)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T$ are covariance matrices of sets S and D, respectively.

By taking the log and omitting constant terms, the Mahalanobis matrix is obtained as $\boldsymbol{M} = \mathrm{Proj}(\boldsymbol{\Sigma}_S^{-1} - \boldsymbol{\Sigma}_D^{-1})$, where $\mathrm{Proj}(\cdot)$ means the projection into the cone of PSD matrices for the guarantee of a valid metric.

2.2 Discriminative subspace learning

The objective of dimension reduction of XQDA is to seek optimal linear projection $\boldsymbol{W} = [\boldsymbol{w}_1, ..., \boldsymbol{w}_r] \in \mathbb{R}^{d \times r} (r < d)$ of the data, so that discriminative properties are preserved in the projected data $\boldsymbol{y} \in \mathbb{R}^r = \boldsymbol{W}^T \boldsymbol{x}$ when the KISSME is applied in the project space. In the subspace, the covariance matrices become $\boldsymbol{\Sigma}'_S = \boldsymbol{W}^T \boldsymbol{\Sigma}_S \boldsymbol{W}$ and $\boldsymbol{\Sigma}'_D = \boldsymbol{W}^T \boldsymbol{\Sigma}_D \boldsymbol{W}$. Thus, the KISSME on the subspace becomes $\boldsymbol{M}' = \mathrm{Proj}(\boldsymbol{\Sigma}_S^{'-1} - \boldsymbol{\Sigma}_D^{'-1})$.

Given a projection matrix $\boldsymbol{W}$ and Mahalanobis metric $\boldsymbol{M}'$, the distance function becomes as follows:

$$d^2_{\boldsymbol{W},\boldsymbol{M}'}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{W} \boldsymbol{M}' \boldsymbol{W}^T (\boldsymbol{x}_i - \boldsymbol{x}_j). \tag{3}$$

Note that the Mahalanobis matrix on the original feature space is obtained as $\boldsymbol{M} = \boldsymbol{W} \boldsymbol{M}' \boldsymbol{W}^T$.

[1] XQDA admits sample pairs only drawn from different camera views.

A direct optimization of $d^2_{\boldsymbol{W},\boldsymbol{M}'}$ is difficult due to the inclusion of $\boldsymbol{W}$ in two inverse covariance matrices [5]. Instead, XQDA greedily optimizes each projection basis $\boldsymbol{w}$ so that the ratio of variances of the similar and dissimilar pairs, σ_S, σ_D is maximized. Because $\sigma_S(\boldsymbol{w}) = \boldsymbol{w}^T \Sigma_S \boldsymbol{w}$ and $\sigma_D(\boldsymbol{w}) = \boldsymbol{w}^T \Sigma_D \boldsymbol{w}$, it corresponds to the generalized Rayleigh quotient criterion $J(\boldsymbol{w}) = \frac{\sigma_D(\boldsymbol{w})}{\sigma_S(\boldsymbol{w})} = \frac{\boldsymbol{w}^T \Sigma_D \boldsymbol{w}}{\boldsymbol{w}^T \Sigma_S \boldsymbol{w}}$. Thereby, XQDA reduces to the generalized eigendecomposition problem, similar to Fisher's discriminant analysis [12, 13].

2.3 Generalized Rayleigh quotient criterion in metric embedded space

A greedy maximization of generalized Rayleigh quotient criterion is equivalent to the maximization of the ratio trace objective $J(\boldsymbol{W}) = \mathrm{Tr}\left((\boldsymbol{\Sigma}'_S)^{-1}(\boldsymbol{\Sigma}'_D)\right)$ where $\mathrm{Tr}(\cdot)$ is the trace of matrix [14].

Because $\boldsymbol{M}'$ is a positive-semidefinite matrix, there exists matrix square-root $\boldsymbol{M}'^{\frac{1}{2}}$. The transformation of the samples as $\boldsymbol{y}' \in \mathbb{R}^r = \boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{W}^T \boldsymbol{x}$ emebeds the Mahalobis metric $\boldsymbol{M}$ into a Euclidean space. We consider the covariance matrices on the metric embedded space $\boldsymbol{\Sigma}_{M,D} = \boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{W}^T \boldsymbol{\Sigma}_D \boldsymbol{W} \boldsymbol{M}'^{\frac{1}{2}} = \boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{\Sigma}'_D \boldsymbol{M}'^{\frac{1}{2}}$, $\boldsymbol{\Sigma}_{M,S} = \boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{\Sigma}'_S \boldsymbol{M}'^{\frac{1}{2}}$ and examine the maximization of the ratio trace in metric embedded space $J(\boldsymbol{W}, \boldsymbol{M}) = \mathrm{Tr}\left((\boldsymbol{\Sigma}_{M,S})^{-1}(\boldsymbol{\Sigma}_{M,D})\right)$. The following proposition holds:

Proposition.1 *The objective function for ratio trace maximization is invariant under the embedded metric.*

Proof. We see that

$$\begin{aligned} J(\boldsymbol{W}, \boldsymbol{M}) &= \mathrm{Tr}\left((\boldsymbol{\Sigma}_{M,S})^{-1}(\boldsymbol{\Sigma}_{M,D})\right) \\ &= \mathrm{Tr}\left((\boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{\Sigma}'_S \boldsymbol{M}'^{\frac{1}{2}})^{-1} (\boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{\Sigma}'_D \boldsymbol{M}'^{\frac{1}{2}})\right) \\ &= \mathrm{Tr}\left(\boldsymbol{M}'^{-\frac{1}{2}} (\boldsymbol{\Sigma}'_S)^{-1} \boldsymbol{M}'^{-\frac{1}{2}} \boldsymbol{M}'^{\frac{1}{2}} \boldsymbol{\Sigma}'_D \boldsymbol{M}'^{\frac{1}{2}}\right) \\ &= \mathrm{Tr}\left((\boldsymbol{\Sigma}'_S)^{-1}(\boldsymbol{\Sigma}'_D)\right) = J(\boldsymbol{W}). \end{aligned} \tag{4}$$

□

In this way, the existence of any Mahalanobis metric on the projected subspace has no influence on the ratio trace objective function [2]. This proposition holds true regardless of whether $\boldsymbol{W}$ is included in $\boldsymbol{M}'$ or not. As such, XQDA maximizes the generalized Rayleigh quotient criterion on the embedded space of KISSME.

3 Kernelized XQDA

We present a kernelized version of XQDA (KXQDA) with the standard kernel trick [15]. The idea of kernel trick is to map input samples by non-linear mapping $\phi(\cdot)$ from $\mathbb{R}^d$ to RKHS $\mathcal{H}$ where the inner product of sample $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ is defined by a kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_i)$. In $\mathcal{H}$, the covariance matrices would be defined as $\boldsymbol{\Sigma}_{\mathcal{H},D} = \frac{1}{N_D} \sum_{(i,j)\in D} (\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j))(\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j))^T$ and $\boldsymbol{\Sigma}_{\mathcal{H},S} = \frac{1}{N_S} \sum_{(i,j)\in S} (\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j))(\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j))^T$. Though KISSME would be obtained as $\boldsymbol{M}_{\mathcal{H}} = \mathrm{Proj}_{\mathcal{H}}(\boldsymbol{\Sigma}_{\mathcal{H},S}^{-1} - \boldsymbol{\Sigma}_{\mathcal{H},D}^{-1})$, the derivations of the

[2] The Fukunaga's book [12] also shows the invariance of the ratio trace problem under the non-singular square matrix multiplications to the covariance matrices (Chapter10, Eq.(10.15)).

inverse covariance matrices and the projection operation in $\mathcal{H}$ are not obvious.

Let us consider learning the linear projection bases $\boldsymbol{W}_{\mathcal{H}} = [\boldsymbol{w}_{\mathcal{H},1}, ..., \boldsymbol{w}_{\mathcal{H},r}]$ which map feature $\phi(\boldsymbol{x})$ in $\mathcal{H}$ to $\mathbb{R}^r$ as $\boldsymbol{W}_{\mathcal{H}}^T\boldsymbol{\phi}(\boldsymbol{x})$. The likelihood-ratio test in the subspace derives the following Mahalanobis distance:

$$d^2_{\boldsymbol{W}_{\mathcal{H}},\boldsymbol{M}'_{\mathcal{H}}}(\boldsymbol{x}_i,\boldsymbol{x}_j) \quad (5)$$
$$= (\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j))^T \boldsymbol{W}_{\mathcal{H}}\boldsymbol{M}'_{\mathcal{H}}\boldsymbol{W}_{\mathcal{H}}^T (\phi(\boldsymbol{x}_i) - \phi(\boldsymbol{x}_j)),$$

where $\boldsymbol{M}'_{\mathcal{H}} = \text{Proj}(\boldsymbol{\Sigma}'^{-1}_{\mathcal{H},S} - \boldsymbol{\Sigma}'^{-1}_{\mathcal{H},D})$ is the Mahalanobis matrix on the subspace in RKHS. Here $\boldsymbol{\Sigma}'_{\mathcal{H},S} = \boldsymbol{W}_{\mathcal{H}}^T\boldsymbol{\Sigma}_{\mathcal{H},S}\boldsymbol{W}_{\mathcal{H}}$ and $\boldsymbol{\Sigma}'_{\mathcal{H},D} = \boldsymbol{W}_{\mathcal{H}}^T\boldsymbol{\Sigma}_{\mathcal{H},D}\boldsymbol{W}_{\mathcal{H}}$. Note that the Mahalanobis matrix in $\mathcal{H}$ is obtained as $\boldsymbol{M}_{\mathcal{H}} = \boldsymbol{W}_{\mathcal{H}}\boldsymbol{M}'_{\mathcal{H}}\boldsymbol{W}_{\mathcal{H}}^T$.

Because the linear projections learnt from the training samples lie in the space spanned by the samples, a projection basis becomes:

$$\boldsymbol{w}_{\mathcal{H}} = \textstyle\sum_{i=1}^N \alpha_i\phi(\boldsymbol{x}_i) = \boldsymbol{\Phi}\boldsymbol{\alpha}, \quad (6)$$

where $\boldsymbol{\Phi} = [\phi(\boldsymbol{x}_1), ..., \phi(\boldsymbol{x}_N)]$ and $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_N]^T$. Thus, the r projection bases become $\boldsymbol{W}_{\mathcal{H}} = \boldsymbol{\Phi}\boldsymbol{A}$, where $\boldsymbol{A} = [\boldsymbol{a}_1, ..., \boldsymbol{a}_r]$, and the linear projection of $\phi(\boldsymbol{x})$ becomes $(\boldsymbol{W}_{\mathcal{H}})^T\phi(\boldsymbol{x}) = \boldsymbol{A}^T\boldsymbol{\Phi}^T\boldsymbol{\phi}(\boldsymbol{x}_i) = \boldsymbol{A}^T\boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{\Phi}^T\phi(\boldsymbol{x}) = [k(\boldsymbol{x}_1,\boldsymbol{x}), ..., k(\boldsymbol{x}_N,\boldsymbol{x})]^T \in \mathbb{R}^N$ is an empirical kernel vector.

We see that $d^2_{\boldsymbol{W}_{\mathcal{H}},\boldsymbol{M}'_{\mathcal{H}}}(\boldsymbol{x}_i,\boldsymbol{x}_j)$ in Eq.(5) is equivalent to the following distance:

$$d^2_{\boldsymbol{A},\boldsymbol{M}'_{\mathcal{K}}}(\boldsymbol{z}_i,\boldsymbol{z}_j) = (\boldsymbol{z}_i - \boldsymbol{z}_j)^T\boldsymbol{A}\boldsymbol{M}'_{\mathcal{K}}\boldsymbol{A}^T(\boldsymbol{z}_i - \boldsymbol{z}_j), \quad (7)$$

where $\boldsymbol{M}'_{\mathcal{K}} = \text{Proj}(\boldsymbol{\Sigma}'^{-1}_{\mathcal{K},S} - \boldsymbol{\Sigma}'^{-1}_{\mathcal{K},D})$ and $\boldsymbol{\Sigma}'_{\mathcal{K},D} = \boldsymbol{A}^T\boldsymbol{\Sigma}_{\mathcal{K},D}\boldsymbol{A}$, $\boldsymbol{\Sigma}'_{\mathcal{K},S} = \boldsymbol{A}^T\boldsymbol{\Sigma}_{\mathcal{K},S}\boldsymbol{A}$, $\boldsymbol{\Sigma}_{\mathcal{K},D} = \frac{1}{N_D}\sum_{(i,j)\in D}(\boldsymbol{z}_i - \boldsymbol{z}_j)(\boldsymbol{z}_i - \boldsymbol{z}_j)^T$ and $\boldsymbol{\Sigma}_{\mathcal{K},S} = \frac{1}{N_S}\sum_{(i,j)\in S}(\boldsymbol{z}_i - \boldsymbol{z}_j)(\boldsymbol{z}_i - \boldsymbol{z}_j)^T$. Note that the Mahalanobis matrix for the empirical kernel vectors $\{\boldsymbol{z}\}_{i=1}^N$ is obtained as $\boldsymbol{M}_{\mathcal{K}} = \boldsymbol{A}\boldsymbol{M}'_{\mathcal{K}}\boldsymbol{A}^T$.

Eq.(7) corresponds to Eq.(3) when we replace the input vector $\boldsymbol{x}$ to the empirical kernel vector $\boldsymbol{z}$ and the projection vector $\boldsymbol{W}$ to $\boldsymbol{A}$. Thereby, we can implement the kernelized XQDA straightforwardly by using the empirical kernel vectors $\{\boldsymbol{z}\}_{i=1}^N$ instead of $\{\boldsymbol{x}\}_{i=1}^N$ to the input of XQDA algorithm. In the test process, we also transform test samples into the empirical kernel vector and measure the distance by Eq.(7).

4 Experiments

4.1 Datasets and Settings

Datasets and evaluation protocols. We evaluate the proposed KXQDA on two public datasets: CUHK Campus [2] and Market-1501 [16]. The CUHK Campus dataset contains 3,884 images of 971 persons. There are two images of each person in each camera view. We evaluate with both *single-shot* and *multi-shot* matchings. In the multi-shot setting, we calculate the distances between two persons by averaging the corresponding cross-view image pairs. We report the average of 10 random person splits for the training/test sets so that each set includes half person identities. We calculate the Cumulative Matching Characteristic (CMC) curve, which gives an expectation of finding the correct person in the top r matches. From the CMC curve, we especially report when $r = 1$ (rank-1 rate). For a measure to evaluate the whole CMC curve, we also report the Proportion of Uncertainty Removed (PUR), which represents the uncertain reduction by a given algorithm from the random ranking [4].

The Market-1501 dataset contains 32,668 bounding boxes of 1,501 persons. Each person is captured by 2-6 cameras. During testing, for each person, one query image in each camera is selected. We use a fixed 750/751 person split for the training/test set. We report the results of both *single-query* and *multi-query* evaluations on 3,386 query persons. For the multi-query setting, we averaged the feature descriptors of different query images of the same person. For the Market-1501 dataset, we report the mean Average Precision (mAP), which considers both the precision and recall of the retrieval process [16] because the gallery contains multiple images of one person.

Feature representation. We use Gaussian Of Gaussians (GOG) descriptor [17], which is known as an effective descriptor for person re-id [1]. The dimensionality of the descriptor is 27,622. As suggested, we apply a mean removal and L2 normalization for each feature.

Kernel. We use the kernel of Radial Basis Function (RBF) with Euclidean distance. As used in various articles, we use the mean distance of the training data as the variance of the RBF kernel.

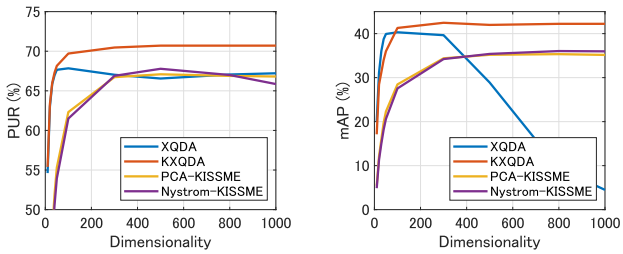
Regularization parameter. In practice, XQDA adds a small regularizer ϵ to diagonal elements of $\boldsymbol{\Sigma}_S$. In particular, we use $\epsilon = 10^{-3}$ for XQDA in the original feature space because it works well when the features are normalized in unit length [5]. KXQDA uses the empirical kernel vector as the input, and the norm of the empirical kernel vector is unnormalized. We found that the effective regularization parameter for KXQDA is smaller than XQDA and use $\epsilon = 10^{-12}$.

4.2 Results

Comparison with baseline methods. We compare the performance of KQDA with the following baselines: PCA-KISSME [3], XQDA [5] and Nyström-KISSME [9] [3]. PCA-KISSME refers a method in which the input features are projected into low-dimensional subspace by PCA before applying KISSME. Nyström-KISSME firstly embeds an original feature into low-dimension Euclidean space which approximate RKHS by Nyström method [11]. Fig. 1 shows the results on various subspace dimensionalities.

We see that KISSME decreases its performance when the feature vectors are projected into a lower dimension subspace by PCA. XQDA outperforms PCA-KISSME in the lower dimensional spaces. These results are because the subspace projection of XQDA considers discriminative information.

We also see that the KXQDA consistently outperforms XQDA in the original feature space. These results confirm the improvement by kernelizing the XQDA. KXQDA also outperforms Nyström-KISSME. The reason for these results is the same as XQDA outperforms PCA-KISSME. Namely, Nyström method obtains the Euclidean space without supervised information. In contrast, KXQDA maximizes the discriminative criterion, and thus more discriminative information is retained.



(a) CUHK Campus (single) (b) Market-1501 (single)

Figure 1. Comparison with baseline methods in various subspace dimensionalities.

Table 1. Comparison of metric learning methods.

Methods	CUHK Campus				Market-1501			
	Single-shot		Multi-shot		Single-query		Multi-query	
	r=1	PUR	r=1	PUR	r=1	mAP	r=1	mAP
PCA-KSSME [3]	57.0	66.7	67.8	75.2	60.2	35.2	67.8	44.3
Nyström-KISS. [9]	58.6	67.8	68.3	75.8	58.3	36.0	66.0	43.8
XQDA [5]	57.9	66.8	67.1	74.8	65.0	40.2	74.0	50.1
KNFST [6]	60.8	69.4	70.3	77.1	66.6	**42.4**	74.8	**52.7**
KXQDA	**62.2**	**70.7**	**71.5**	**78.0**	**66.7**	42.2	**75.1**	52.3

We summarize the performance of the compared methods in Table 1. In the table, the dimensionalities of XQDA/KXQDA are automatically set by selecting all eigenvectors greater than 1. For the dimensionalities of PCA-KISSME and Nyström-KISSME, 500 and 1000 are set for the CUHK Campus and Market-1501 datasets, respectively.

Comparison with another metric learning method. Finally, we compare the performance of KXQDA with another metric learning method: Kernelized Null Foley-Sammon Transform (KNFST) [6]. KNFST seeks a subspace in which within class distances are zero (null space), and between class distances are positive. The subspace dimensionality is automatically determined as the number of the null space, and there is no free parameter. We use the same RBF kernel to the KXQDA. Note that the application of KNFST to person re-id is more recently proposed than other metric learning methods, e.g., Local Fisher Discriminant Analysis (LFDA) [4], a kernel extension of regularized Pairwise Constrained Component Analysis (rPCCA) [7], Kernelized LFDA (KLFDA) [7] [4], Kernelized Marginal Fisher Analysis (KMFA) [7, 18] [4] and XQDA [5].

Table 1 includes the results. We see that KXQDA outperforms KNFST in both rank-1 rates and PUR on the CUHK Campus dataset, and in rank-1 rates on the Market-1501 dataset. These results are probably because that KNFST projects feature vectors into a null space of within-class distances, whereas it only considers the subspace which has positive between-classes. Thus, KNFST mostly focuses on the minimization of within-class distances. In contrast, XQDA/KXQDA maximize the ratio of these distances. Thus, the subspace which has a better discriminative ability of different classes could be obtained. Also, cross-view restrictions in the covariance matrices and KISSME in the projected space could be the cause of better performance of KXQDA. The superior performance of KNFST to XQDA is because KNFST is kernelized. By kernelizing XQDA, KXQDA outperforms KNFST on these datasets.

[3] Another kernel version of KISSME expects much slower than KXQDA because it adopts an iterative optimization on a Riemannian manifold [9]. The recently proposed kernel-KISSME [10] is simpler than [9], yet it does not accompany with the dimension reduction. Thus, the matching process is expected slow.

[4] Compared to KXQDA, KLFDA [7] and KMFA [7] require an additional computation step to obtain neighborhood graphs for covariance matrices. Besides, only KXQDA uses KISSME on the projected space. With the same RBF kernel and regularization parameter to KXQDA, PUR scores on the CUHK Campus dataset (single-shot) with 500 subspace dimensionality were 70.5% and 70.7%, respectively for KLFDA and KMFA.

5 Conclusion

We have proposed a kernel extension of XQDA which obtains a low-dimensional discriminative subspace in RKHS, and simultaneously KISSME in the learned subspace. In the experiments on person re-id datasets, the kernelized XQDA showed improved performances of XQDA and also outperformed Nyström-KISSME.

References

[1] S. Karanam, M. Gou, Z. Wu, A. Rates-Borras, O. Camps and R. J. Radke, A systematic evaluation and benchmark for person re-identification: feature, metrics, and datasets, *IEEE Trans. on PAMI*, vol.41, no.3, pp.523–536, 2019.

[2] W. Li, R. Zhao, and X. Wang, Human reidentification with transferred metric learning, In Proc. ACCV, 2012.

[3] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth and H. Bischof, Large scale metric learning from equivalence constraints, In Proc. CVPR, 2012,

[4] S. Pedagadi, J. Orwell, S. Velastin and B. Boghossian, Local Fisher discriminant analysis for pedestrian re-identification, In Proc. CVPR, 2013.

[5] S. Liao, Y. Hu, X. Zhu and S. Z. Li, Person Re-identification by local maximal occurrence representation and metric learning, In Proc. CVPR, 2015.

[6] L. Zhang, T. Xiang and S. Gong, Learning a discriminative null space for person re-identification, In Proc. CVPR, 2016.

[7] F. Xiong, M. Gou, O. Camps and M. Sznaier, Person re-identification using kernel-based metric learning methods, In Proc. ECCV, 2014.

[8] C. Zhao, Y. Chen, X. Wang, W. K. Wong, D. Miao and J. Lei, Kernelized random KISS metric learning for person re-identification, *Neurocomputing*, vol.275, pp.403-417, 2018.

[9] M. Faraki, M. T. Harandi and F. Porikli, No fuss metric learning, a Hilbert space scenario, *Pattern Recognition Letters*, vol.98, pp.83-89, 2017.

[10] B. Nguyen and B. D. Baets, Kernel distance metric learning using pairwise constraints for person re-identification, *IEEE Trans. on Image Processing*, vol.28, no.2, pp.589-600, 2019.

[11] C. T. H. Baker, The numerical treatment of integral equations, Clarendon press, Oxford, 1977.

[12] K. Fukunaga, Introduction to statistical pattern recognition, Academic press, San Diego, second edition, 1991.

[13] M. Kan, S. Shan, D. Xu and X. Chen, Side-Information based Linear Discriminant Analysis for Face Recognition, In Proc. BMVC, 2011.

[14] J. P. Cunningham and Z. Ghahramani, Linear dimensionality reduction: survey, insights, and generalizations, *Journal of Machine Learning Research*, vol.16, pp.2859–2900, 2015.

[15] J. Shawe-Taylor and N. Cristianini, Kernel methods for pattern analysis, Cambridge university press, 2004.

[16] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wand and Q. Tian, Scalable person re-identification: A benchmark, In Proc. ICCV, 2015.

[17] T. Matsukawa, T. Okabe, E. Suzuki and Y. Sato, Hierarchal Gaussian descriptor for person re-identification, In Proc. CVPR, 2016.

[18] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang and S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, *IEEE Trans. on PAMI*, vol.29, no.1, pp.40–51, 2007.

Appendix

An insight on the Mahalanobis matrix of XQDA

An eigenvalue λ of $\mathbf{\Sigma}_{\mathrm{S}}^{-1}\mathbf{\Sigma}_{\mathrm{D}}$ corresponds to $\sigma_{\mathrm{D}}/\sigma_{\mathrm{S}}$. Because the subspace $\sigma_{\mathrm{D}} < \sigma_{\mathrm{S}}$ may provide no discriminative information, a reasonable criterion to determine the subspace dimensionality is to use all eigenvectors whose corresponding eigenvalues are greater than 1 [5].

To have a valid metric on the subspace, the Mahalanobis matrix $\boldsymbol{M}'$ must be positive-semidefinite. The following proposition shows that we need no projection operation for $\boldsymbol{M}'$, as far as we select subspace bases by the aforementioned criterion:

Proposition.2 *The Mahalanobis matrix $\boldsymbol{M}'$ in the XQDA is diagonal and PSD matrix if we select the eigenvectors corresponding eigenvalues λ of $\mathbf{\Sigma}_{\mathrm{S}}^{-1}\mathbf{\Sigma}_{\mathrm{D}}$ are greater than 1.*

Proof. This is because the eigendecomposition of the generalized Rayleigh quotient maximization is obtained as a result of simultaneous diagonalization of matrices $\mathbf{\Sigma}_D$ and $\mathbf{\Sigma}_S$ [12]. Namely, the eigenvectors $\boldsymbol{W}$ diagonalize $\mathbf{\Sigma}_D$ and $\mathbf{\Sigma}_S$ as $\boldsymbol{W}^T\mathbf{\Sigma}_D\boldsymbol{W} = \mathrm{diag}(\sigma_{D,1}, \cdots, \sigma_{D,r})$ and $\boldsymbol{W}^T\mathbf{\Sigma}_S\boldsymbol{W} = \mathrm{diag}(\sigma_{S,1}, \cdots, \sigma_{S,r})$. Thus,

$$
\begin{aligned}
\boldsymbol{M}' &= \begin{bmatrix} \sigma_{S,1}^{-1} & & \\ & \ddots & \\ & & \sigma_{S,r}^{-1} \end{bmatrix} - \begin{bmatrix} \sigma_{S,1}^{-1} & & \\ & \ddots & \\ & & \sigma_{S,r}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\sigma_{S,1}} & & \\ & \ddots & \\ & & \frac{1}{\sigma_{S,r}} \end{bmatrix} - \begin{bmatrix} \frac{1}{\sigma_{S,1}} & & \\ & \ddots & \\ & & \frac{1}{\sigma_{S,r}} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\sigma_{D,1}-\sigma_{S,1}}{\sigma_{S,1}\sigma_{D,1}} & & \\ & \ddots & \\ & & \frac{\sigma_{D,r}-\sigma_{S,r}}{\sigma_{S,r}\sigma_{D,r}} \end{bmatrix}.
\end{aligned}
$$

If we select the eigenvectors whose corresponding eigenvalue $\lambda = \frac{\sigma_D}{\sigma_S} \geq 1$ then $\sigma_D - \sigma_S \geq 0$. □

We note that replacing Singular Value Decomposition (SVD) applied to $(\mathbf{\Sigma}_S^{-1})\mathbf{\Sigma}_D$ in the Matlab code[5] (line 148) to the eigendecomposition is necessary to validate this proposition. In fact, using SVD here produces no longer the solution of the generalized Rayleigh quotient maximization of XQDA. This is because for symmetric matrix $\boldsymbol{A}$, each of the decomposed matrices of SVD is given as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{U}^T$, where $\boldsymbol{D}$ is the diagonal with the square-root of the eigenvalue of $\boldsymbol{A}\boldsymbol{A}^T$, $\boldsymbol{V}$ and $\boldsymbol{U}^T$ are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$. Therefore the SVD corresponds to the eigendecomposition of $(\mathbf{\Sigma}_S^{-1})\mathbf{\Sigma}_D((\mathbf{\Sigma}_S^{-1})\mathbf{\Sigma}_D))^T$. We have fixed this misuse and conducted all experiments in this paper.

Computation of covariance matrices

A direct computation of $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_D$ costs $O(N_S d^2)$ and $O(N_D d^2)$ floating-point operations. Because most of the sample pairs are dissimilar pairs, N_D grows with the order $O(N^2)$. We can use a practical computation method with the cost of $O(Nd^2)$ floating-point operations [5] only for the case of two-camera views. When multiple cameras exist, we need to divide all samples into two views of all camera combinations and combine their covariance matrices.

For handling the multi-camera setting naturally, we propose to compute the covariance matrices using graph Laplacian [18]. For each of the sample index sets of similar/dissimilar pairs, an (i, j) element of an affinity matrix is set to 1 or 0 based on the existence of the pair (i, j) in the set. The computational costs using the graph Laplacian for $\mathbf{\Sigma}_S$ and $\mathbf{\Sigma}_D$ are both $O(\max(N^2 d, Nd^2))$ floating-point operations, equivalent to the practical computation [5] when $N \leq d$.

Table 2 shows the empirical time for constructing covariance matrices, measured by a computer equipped with an Intel Xeon E5-2687W v3 @3.1GHz. The graph Laplacian shows faster empirical times than the practical computation method [5]. Relatively significant time reduction by the graph Laplacian on the Market-1501 dataset is because it has 6 cameras, whereas the CUHK Campus dataset has 2 cameras.

Table 2. Empirical time of covariance matrix computations for XQDA/KXQDA (sec.)

Methods	CUHK Campus		Market-1501	
	XQDA	KXQDA	XQDA	KXQDA
Practical computation [5]	0.502	0.412	128.3	129.2
Graph Laplacian	**0.366**	**0.369**	**51.9**	**51.8**

[5] https://github.com/liangzheng06/PRW-baseline/blob/master/utils/LOMO_XQDA/code/XQDA.m