Joint Dimension Reduction and Metric Learning for Person Re-identification

Shengcai Liao, Yang Hu, Stan Z. Li

Center for Biometrics and Security Research
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, China
{sc.liao,y.hu,szli}@nlpr.ia.ac.cn

Abstract. Person re-identification is an important technique towards automatic search of a person’s presence in a surveillance video. Among various methods developed for person re-identification, the Mahalanobis metric learning approaches have attracted much attention due to their impressive performance. In practice, many previous papers have applied the Principal Component Analysis (PCA) for dimension reduction before metric learning. However, this may not be the optimal way for metric learning in low dimensional space. In this paper, we propose to jointly learn the discriminant low dimensional subspace and the distance metric. This is achieved by learning a projection matrix and a Restricted Quadratic Discriminant Analysis (RQDA) model. We show that the problem can be formulated as a Generalized Rayleigh Quotient, and a closed-form solution can be obtained by the generalized eigenvalue decomposition. We also present a practical computation method for RQDA, as well as its regularization. For the application of person re-identification, we propose a Retinex and maximum occurrence based feature representation method, which is robust to both illumination and viewpoint changes. Experiments on two challenging public databases, VIPeR and QMUL GRID, show that the performance of the proposed method is comparable to the state of the art.

Keywords: dimension reduction, metric learning, quadratic discriminant analysis, Rayleigh quotient, regularization, person re-identification

1 Introduction

Person re-identification is a problem of finding a person from a gallery who has the same identity to a probe. This is a challenging problem because of big intra-class variations in illumination, pose or viewpoint, and occlusion. Many approaches have been proposed for person re-identification [1,2], which greatly advance this field.

A typical forensic application of person re-identification is to search a suspect from a video or a set of videos. This application naturally involves a big
number of unseen classes. Therefore, the traditional way of learning several discriminant functions to classify a set of known classes is no longer suitable for the rapid demand of real applications. The Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) \cite{3} are two examples working in this way.

Metric learning is a kind of technique to address the above problem. Instead of classifying known classes, metric learning tries to learn a Mahalanobis distance function which is able to distinguish a pair of samples being the same person or not. Learning a distance or similarity function is a more general way for identification or verification applications where the class of the test sample is usually not included in the learning process. If the learned distance function has a good generalization ability, it can be easily applied for a general scoring of various pairs of samples, regardless the actual identity of the samples.

Many metric learning algorithms have been proposed for person re-identification \cite{4,5,6,7,8,9,10,2}. In practice, many previous papers \cite{4,6,5,8} show a two-stage processing for metric learning, that is, the Principle Component Analysis (PCA) is first applied for dimension reduction, then metric learning is performed on the PCA subspace. Zheng et al. \cite{7} found that, for some metric learning algorithms like LMNN \cite{4} and ITML \cite{6}, dimension reduction by PCA must be performed, otherwise they becomes intractable given the high dimensional feature space. However, this two-stage processing may not be the optimal way for metric learning in low dimensional space, because samples of different classes may already be cluttered after the first stage.

In this paper, we propose to jointly learn the discriminant low dimensional subspace and the distance metric. This is achieved by learning a projection matrix and a Restricted Quadratic Discriminant Analysis (RQDA) model. We show that the problem can be formulated as a Generalized Rayleigh Quotient, and a closed-form solution can be obtained by the generalized eigenvalue decomposition. We also present a practical computation method for RQDA, as well as its regularization. For the application of person re-identification, we propose a Retinex and maximum occurrence based feature representation method, which is robust to both illumination and viewpoint changes. Experiments on two challenging public databases, VIPeR \cite{11} and QMUL GRID \cite{12}, show that the performance of the proposed method is comparable to the state of the art.

The remainder of this paper is organized as follows. In Section 2, we briefly review the related work. In Section 3, we present the proposed RQDA algorithm. In Section 4, we introduce the proposed feature representation method. Experimental results are shown in Section 5 and finally we conclude this paper in Section 6.

2 Related Work

Many existing person re-identification approaches try to build a robust feature representation which is both distinctive and stable for describing a person’s appearance under various conditions \cite{13,14,15,16}. Farenzena et al. \cite{17} proposed
the Symmetry-Driven Accumulation of Local Features (SDALF) method, where multiple features were combined, and the symmetry and asymmetry property in pedestrian images were considered to handle viewpoint variations. Malocal et al. [18] turned local descriptors into the Fisher Vector to produce a global representation of an image. Cheng et al. [19] utilized the Pictorial Structures for person re-identification, where part-based color information and color displacement within the whole body were extracted. Recently, the use of saliency information has been seen in person re-identification [20,21,22], leading to a novel feature extraction scheme.

Besides the research of robust features, metric learning approaches have been widely used in person re-identification [4,5,6,7,8,9,10,2]. Weinberger et al. [4] proposed the LMNN algorithm to learn a Mahalanobis distance metric for improving the k-nearest neighbor (kNN) classification, which can be solved by semidefinite programming. Subsequently, in [5], a method similar to LMNN called Large Margin Nearest Neighbor with Rejection (LMNN-R) was proposed, which achieved an impressive improvement. Davis et al. [6] presented an approach called ITML, which considered minimizing the differential relative entropy between two multivariate Gaussians for learning the Mahalanobis distance function. Zheng et al. [7] proposed the PRDC algorithm, which formulates person re-identification as a distance learning problem to maximize the probability that a pair of true match samples has a smaller distance than a nonmatch pair. Koestinger et al. [8] proposed the KISSME method which considered a log likelihood ratio test of two Gaussian distributions, and a simplified solution was derived. Li at al. [9] proposed the learning of Locally-Adaptive Decision Functions (LADF) for person verification, which can be viewed as a joint model of a distance metric and a locally adapted thresholding rule.

There are also other person re-identification methods that deserve mentioning. Gray and Tao [23] proposed to use AdaBoost to select good features out of a set of color and texture features. Prosser et al. [24] formulated the person re-identification problem as a ranking problem, and applied the Ensemble RankSVM to learn a subspace where the potential true match gets the highest rank. In [25], complex cross-view transformations are considered for person re-identification across views. These transformations can be used to assign images to different local experts, and then they can be projected to a common feature space and matched with a locally learned discriminative metric. Liu et al. [26] proposed a Post-rank OPtimisation (POP) approach, which allows a user to quickly refine their search during a re-identification process. This kind of feedback optimisation demonstrates a significant performance improvement for person re-identification.

3 Restricted Quadratic Discriminant Analysis

3.1 RQDA

Suppose $X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^{d \times n}$ contains $n$ samples in a $d$-dimensional space. $y = (y_1, y_2, \ldots, y_n)$ contains the corresponding class labels, where $y_i \in$
Consider the sample difference $\Delta = x_i - x_j$. $\Delta$ is called the intrapersonal difference if $y_i = y_j$, while it is called the extrapersonal difference if $y_i \neq y_j$ \cite{27}. Accordingly, two classes of variations can be defined: the intrapersonal variations $\Omega_I$ and the extrapersonal variations $\Omega_E$. Therefore, in this way the multi-class classification problem can be solved by distinguishing the above two classes.

For the classification of $\Omega_I$ and $\Omega_E$, Moghaddam et al. \cite{27} proposed to model each of the two classes with a multivariate Gaussian distribution. This corresponds to a QDA model with the defined $\Omega_I$ and $\Omega_E$ as two classes. Furthermore, it was noticed in \cite{27} that both $\Omega_I$ and $\Omega_E$ have zero mean. This is because for each $\Delta = x_i - x_j$, there exists a $\Delta = x_j - x_i$. Kostinger et al. also derived a similar approach called KISSME in \cite{8} via the log likelihood ratio test of the two Gaussian distributions. In this paper, this model is called Restricted QDA to reflect its connection to QDA and the restriction of zero mean.

Formally, the RQDA model is formulated as follows. Under the zero-mean Gaussian distribution, the likelihoods of observing $\Delta$ in $\Omega_I$ and $\Omega_E$ are defined as

$$P(\Delta | \Omega_I) = \frac{1}{(2\pi)^{d/2} |\Sigma_I|^{1/2}} e^{-\frac{1}{2} \Delta^T \Sigma^{-1}_I \Delta},$$

$$P(\Delta | \Omega_E) = \frac{1}{(2\pi)^{d/2} |\Sigma_E|^{1/2}} e^{-\frac{1}{2} \Delta^T \Sigma^{-1}_E \Delta},$$

where

$$\Sigma_I = \frac{1}{n_I} \sum_{y_i = y_j} (x_i - x_j)(x_i - x_j)^T,$$

$$\Sigma_E = \frac{1}{n_E} \sum_{y_i \neq y_j} (x_i - x_j)(x_i - x_j)^T,$$

are the covariance matrices of $\Omega_I$ and $\Omega_E$, respectively, and $n_I$ and $n_E$ denotes the number of samples in the two classes. According to the Bayesian rule, the posterior probability of assigning an observation $\Delta$ to $\Omega_I$ and $\Omega_E$ are

$$P(\Omega_I | \Delta) = \frac{P(\Delta | \Omega_I) P(\Omega_I)}{P(\Delta)},$$

$$P(\Omega_E | \Delta) = \frac{P(\Delta | \Omega_E) P(\Omega_E)}{P(\Delta)}$$

Therefore, the decision function is obtained by comparing $P(\Omega_E | \Delta) - P(\Omega_I | \Delta)$. Further taking the log, and discarding some constants that can be merged to the decision threshold, the decision function can be simplified as

$$f(\Delta) = \Delta^T (\Sigma^{-1}_I - \Sigma^{-1}_E) \Delta,$$

and so the RQDA distance function between $x_i$ and $x_j$ is

$$d(x_i, x_j) = (x_i - x_j)^T (\Sigma^{-1}_I - \Sigma^{-1}_E)(x_i - x_j).$$

Therefore, learning the distance function corresponds to estimating the covariant matrices $\Sigma_I$ and $\Sigma_E$. 

[1, c], i = 1, 2, \ldots , n, and c is the number of classes.
3.2 Joint Subspace and Metric Learning

Usually, the original feature dimensions \( d \) is large, and it is preferred to perform the classification in a low dimensional space \( \mathbb{R}^r \), with \( r < d \). In [27], it is suggested to decompose \( \Sigma_I \) and \( \Sigma_E \) separately to reduce the dimensions. In [8], PCA was applied, then \( \Sigma_I \) and \( \Sigma_E \) were estimated in the PCA subspace. However, both methods are not optimal because the dimension reduction does not consider the distance metric learning.

In this paper, we consider to learn a subspace \( W = (w_1, w_2, \ldots, w_r) \in \mathbb{R}^{d \times r} \), and at the same time learn a RQDA distance function in the \( r \) dimensional subspace. Each sample \( x \) is projected in the subspace as \( x' = W^T x \). Then, according to Eq. (8), the RQDA distance function in the \( r \) dimensional subspace is computed as

\[
d_W(x_i, x_j) = (x_i - x_j)^T W (\Sigma'_I^{-1} - \Sigma'_E^{-1}) W^T (x_i - x_j),
\]

where

\[
\Sigma'_I = \frac{1}{n_I} \sum_{y_i = y_j} W^T (x_i - x_j)(x_i - x_j)^T W = W^T \Sigma_I W,
\]

\[
\Sigma'_E = \frac{1}{n_E} \sum_{y_i \neq y_j} W^T (x_i - x_j)(x_i - x_j)^T W = W^T \Sigma_E W.
\]

Therefore, we needs to learn a kernel matrix \( M(W) = W(\Sigma'_I^{-1} - \Sigma'_E^{-1})W^T \).

![Fig. 1. Sample distributions of \( \Omega_I \) and \( \Omega_E \) in one projected dimension.](image)

However, directly optimizing \( d_W \) is difficult because \( W \) is contained in two inverse matrices. Recall that \( \Omega_I \) and \( \Omega_E \) have zero mean, then given a basis \( w \), the projected samples of the two classes will center at zero, but may have different variance, as shown in Fig. [1]. In this case, the Fisher criterion used to derive LDA is no longer suitable because the two classes have the same mean. However, the variances \( \sigma_I \) and \( \sigma_E \) can still be used to distinguish the two classes. Therefore, we can optimize the projection direction \( w \) such that \( \sigma_E(w)/\sigma_I(w) \) is maximized.
Notice that
\[
\sigma_I(w) = \frac{1}{n_I} \sum_{y_i = y_j} (w^T x_i - w^T x_j)(w^T x_i - w^T x_j)^T = w^T \Sigma_I w,
\] (12)
\[
\sigma_E(w) = \frac{1}{n_E} \sum_{y_i \neq y_j} (w^T x_i - w^T x_j)(w^T x_i - w^T x_j)^T = w^T \Sigma_E w.
\] (13)

Therefore, the objective \(\sigma_E(w)/\sigma_I(w)\) corresponds to the Generalized Rayleigh Quotient
\[
J(w) = \frac{w^T \Sigma_E w}{w^T \Sigma_I w}.
\] (14)

The maximization of \(J(w)\) is equivalent to
\[
\max_w w^T \Sigma_E w, \quad s.t. \quad w^T \Sigma_I w = 1,
\] (15)
which can be solved by the generalized eigenvalue decomposition problem. That is, the largest eigenvalue of \(\Sigma_I^{-1} \Sigma_E\) is the maximum value of \(J(w)\), and the corresponding eigenvector \(w_1\) is the solution. Furthermore, the solution orthogonal to \(w_1\) and corresponding to the second largest value of \(J(w)\) is the eigenvector of the second largest eigenvalue of \(\Sigma_I^{-1} \Sigma_E\), and so on. Therefore, with \(W = (w_1, w_2, \ldots, w_r)\) we learn a discriminant subspace, as well as a RQDA distance function in the learned subspace, as defined in Eq. (9).

### 3.3 Practical Computation

The computation of the two covariance matrices \(\Sigma_I\) and \(\Sigma_E\) defined in Eqs. (3) and (4) require \(O(mnd^2)\) and \(O(n^2d^2)\) multiplication operations, respectively, where \(m = n/c\) represents the average number of images in each class. To reduce the computation, we rewrite Eq. (3) as
\[
n_I \Sigma_I = \frac{1}{2} \sum_{k=1}^{c} \sum_{y_i = k} \sum_{y_j = k} (x_i - x_j)(x_i - x_j)^T
\]
\[
= \frac{1}{2} \sum_{k=1}^{c} \sum_{y_i = k} \sum_{y_j = k} (x_i x_i^T + x_j x_j^T - x_i x_j^T - x_j x_i^T)
\]
\[
= \frac{1}{2} \sum_{k=1}^{c} \left( n_k \sum_{y_i = k} x_i x_i^T + n_k \sum_{y_j = k} x_j x_j^T - \sum_{y_i = k} x_i \sum_{y_j = k} x_j^T - \sum_{y_j = k} x_j \sum_{y_i = k} x_i^T \right)
\]
\[
= \sum_{k=1}^{c} \left( n_k \sum_{y_i = k} x_i x_i^T - \sum_{y_i = k} x_i \sum_{y_j = k} x_j^T \right)
\]
\[
= \sum_{k=1}^{c} n_k^2 \left( \Sigma_k - \mu_k \mu_k^T \right),
\] (16)
where $\mu_k$ is the sample mean of class $k$, $\Sigma_k$ is the sample covariance of class $k$, and $n_k$ is the number of images in class $k$. Similarly, we can obtain
\[
n_E \Sigma_E = n \sum_{i=1}^{n} x_i x_i^T - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^T - n_I \Sigma_I
\]
\[
= n^2 (\Sigma - \mu \mu^T) - n_I \Sigma_I,
\]
where $\mu$ and $\Sigma$ are the mean and covariance of all samples, respectively. The above simplification shows that the computations of $\Sigma_I$ and $\Sigma_E$ are reduced to $O(nd^2)$ and $O(nd^2)$, respectively. It can be observed that, $\Sigma_I$ and $\Sigma_E$ defined in Eqs. (3) and (4) can be computed directly from sample mean and covariance of each class and all classes, so there is no need to actually compute the $n^2$ pairs of sample differences.

Another practical issue is that, $\Sigma_I$ may be singular, resulting that $\Sigma_I^{-1}$ cannot be computed. Therefore, it is useful to add a small regularizer to the diagonal elements of $\Sigma_I$, as usually done in similar problems like LDA. This will make the estimation of $\Sigma_I$ more smooth and robust. Empirically we find that, when all samples are normalized to unit length, a value of 0.001 as regularizer can be commonly applied to improve the result.

Finally, there is a remaining issue of setting the dimensionality of the derived RQDA subspace. In real applications, there should be a consideration to have a low dimensional subspace to ensure the processing speed. Beyond this consideration, we find that having the selected eigenvalues of $\Sigma_I^{-1} \Sigma_E$ larger than 1 is a useful signature to determine the dimensions. This is because the eigenvalue of $\Sigma_I^{-1} \Sigma_E$ corresponds to $\sigma_E/\sigma_I$ in Fig. 1 and $\sigma_E < \sigma_I$ may not provide useful discriminant information.

4 Feature Representation

Person re-identification is a challenging problem due to big intra-class variations in illumination, pose or viewpoint, and occlusion. Therefore, robust feature representation is important for the description of person images [28]. In this paper, we propose a robust feature representation for person re-identification, which addresses the illumination and viewpoint changes, as described below.

4.1 Illumination Variations

Color is an important feature for describing person images. However, the illumination conditions across cameras can be very different. For example, Fig. 2 shows some sample images from the VIPeR database [11]. It can be observed that images of the same person across the two camera views may have a large difference in illumination.

In this paper, we propose to apply the Retinex algorithm [29,30,31] to pre-process person images. Retinex considers human lightness and color perception. It aims at producing a color image that is consistent to human observation of
the scene. The restored image usually contains vivid color information, especially enhanced details in shadowed regions.

We have implemented the multiscale Retinex algorithm \cite{31} in C++. The code runs fast, usually taking several milliseconds to process a $128 \times 48$ person image on a i5-2400 desktop PC.

Fig. 3 shows some examples of the processed images by our implementation of Retinex. Comparing to Fig. 2 it can be observed that the Retinex images of the same person across cameras have a better consistency in lighting and color. This makes person re-identification easier than using the original images. With the Retinex images, we apply the HSV color histogram to extract color features.

In addition to color description, we also apply the Scale Invariant Local Ternary Pattern (SILTP) \cite{32} descriptor for illumination invariant texture de-
4.2 Viewpoint Changes

Pedestrians under different cameras usually appear in different viewpoint. For example, a person with frontal view in a camera may appear in back view under another camera. Therefore, matching persons in different viewpoints is also difficult. To address this, \cite{7,28} proposed to equally divide a person image into six horizontal stripes, and a single histogram is computed in each stripe. This representation is considered to be robust to pose or viewpoint changes. However, it may also lose spatial details within a stripe.

In this paper, we propose to use sliding windows to describe local details of a person image. Specifically, we use a subwindow size of $16 \times 16$, with an overlapping step of 8 pixels to locate local patches in $128 \times 48$ images. Within each subwindow, we extract an 81-bin SILTP histogram, and an $8 \times 8 \times 8$ joint HSV histogram. Each histogram bin represents the occurrence probability of one pattern in a subwindow. To deal with the viewpoint change, we check all subwindows at the same horizontal location, and compute the maximum occurrence of each pattern (i.e. the same histogram bin) among these subwindows. The resulting histogram achieves some invariance to viewpoint change, and at the same time captures local region characteristics of a person. Fig. 4 shows the procedure of the maximum occurrence based feature representation.

![Fig. 4. Illustration of the maximum occurrence based feature extraction method.](image)

By a mixture of the SILTP and HSV features, our final descriptor has 17,197 dimensions. Since we only use simple SILTP and HSV features, the proposed feature extraction method is efficient. Our MATLAB code running on a i5-2400 desktop PC requires about 20 milliseconds to process a $128 \times 48$ person image.

5 Experiments

The proposed algorithm was evaluated on two challenging public databases, VIPeR \cite{11} and QMUL GRID \cite{12}. Two available features provided by \cite{8} and
for person re-identification were evaluated and compared to the proposed feature. Several existing metric learning algorithms were also evaluated for comparison, and the performance of the proposed method is also compared to the state of the art on the VIPeR and QMUL GRID databases. We detail the experimental description below.

5.1 Experiments on VIPeR

VIPeR [11] is a challenging person re-identification database that has been widely used for benchmark evaluation. It contains 632 pairs of person images, captured by a pair of cameras in an outdoor environment. Images in VIPeR contains large variations in background, illumination, and viewpoint. Fig. 2 shows some example pairs of images from the VIPeR database. All images are scaled to $128 \times 48$ pixels. The widely adopted experimental protocol on this database is to randomly divide the 632 pairs of images into half for training and the other half for testing, and repeat the procedure 10 times to get an average performance. We followed this procedure in our experiments.

Comparison of Metric Learning Algorithms

We used a software package provided by [8] to evaluate several metric learning algorithms, including Identity (i.e. Euclidean distance), MAHAL (i.e. Mahalanobis), LMNN [4], ITML [6] and KISSME [8]. A feature set extracted from the VIPeR database is also provided by [8], which is a mixture of HSV, Lab, and LBP features. We first used this public feature set to evaluate the proposed RQDA algorithm, with a comparison to the above metric learning algorithms. For the compared algorithms, according to [8], PCA was first applied to reduce the feature dimensionality to 34. The proposed RQDA algorithm also learned a 34-dimensional subspace. The resulting Cumulative Matching Characteristic (CMC) curves are shown in Fig. 5. It can be seen that the proposed method is better than the compared metric learning algorithms. This indicates that RQDA successfully learns a better subspace than PCA for metric learning.

Influence of Subspace Dimensions

We further tried the above experimental setting with varying subspace dimensions. The corresponding results, measured by rank-10 identification rates, are shown in Fig. 6. Clearly, RQDA and ITML are more stable than other algorithms. Therefore, determining a proper number of subspace dimensionality is not too difficult for RQDA.

Comparison of Features

Next, all compared algorithms were applied with features extracted by the proposed method. For the existing algorithms, PCA was applied to reduce the feature dimensionality to 34. The proposed RQDA algorithm also learned a 34-dimensional subspace. The resulting CMC curves are depicted in Fig. 7. It can also be seen that RQDA does a better job in jointly learning the discriminant subspace and distance function, compared to other two-stage based learning methods.
Fig. 5. CMC curves on the VIPeR database [11] (p=316). Algorithms were run with features provided by [8]. The original feature dimensionality was reduced to 34 according to [8].

Fig. 6. Rank-10 identification rates with different dimensionality on the VIPeR database [11] (p=316). Algorithms were run with features provided by [8]. The original feature dimensionality was reduced to 34, 50, 70, and 90.

Fig. 7. CMC curves on the VIPeR database [11] (p=316). Algorithms were run with features extracted by the proposed method. The original feature dimensionality was reduced 34 according to [8].
Furthermore, the rank-10 identification rates of using the feature set of [8] and the proposed feature are summarized in Table 1. The results show that all algorithms achieve a performance gain using the proposed feature. Especially, the RQDA, KISSME, and LMNN algorithms obtain an improvement larger than 10%. Since the two kinds of features are similar in fusing color and texture information, the improvement made by the proposed feature is mainly due to the specific consideration of handling illumination and viewpoint changes.

**Table 1.** Rank-10 identification rates (%) with different features on the VIPeR database (p=316).

| Method       | RQDA  | KISSME | ITML  | LMNN  | MAHAL | IDENTITY |
|--------------|-------|--------|-------|-------|-------|----------|
| Feature [8]  | 65.66 | 62.16  | 53.01 | 55.85 | 52.37 | 32.91    |
| Proposed feature | **76.42** | 72.63  | 59.97 | 65.98 | 63.45 | 34.49    |

**Comparison to the State of the Art** Finally, we shall compare the performance of the proposed approach to the state-of-the-art results reported on the VIPeR database. We used the proposed features, and applied the RQDA dimension reduction and metric learning. To avoid tuning the subspace dimensionality, in the following experiments we determined this value automatically by accepting all eigenvalues of $\Sigma^{-1}_I \Sigma_E$ that are larger than 1, as discussed earlier. The results are summarized in Table 2. The LADF [9] is a recent metric learning algorithm which reports the best performance on the VIPeR database to date. The proposed algorithm achieves a small improvement over LADF, indicating that it is promising for addressing the challenging person re-identification problem.

**Table 2.** Comparison of state-of-the-art results on the VIPeR database (p=316). The cumulative matching scores (%) at rank 1, 5, 10, and 20 are listed.

| Method           | rank = 1 | rank = 5 | rank = 10 | rank = 20 |
|------------------|----------|----------|-----------|-----------|
| ELF [18]         | 12.00    | 31.00    | 41.00     | 58.00     |
| SDALF [20]       | 19.87    | 38.89    | 49.37     | 65.73     |
| PRDC [20]        | 15.66    | 38.42    | 53.86     | 70.09     |
| CPS [20]         | 21.84    | 44.00    | 57.21     | 71.00     |
| eLDFV [20]       | 22.34    | 47.00    | 60.04     | 71.00     |
| eSDC_knn [20]    | 26.31    | 46.61    | 58.86     | 72.77     |
| eSDC_ocsvm [20]  | 26.74    | 50.70    | 62.37     | 76.36     |
| LADF             | 32.59    | 64.24    | 78.16     | 89.24     |
| RQDA             | **34.65** | **65.35** | **78.64** | **89.56** |

Note: References in the first column indicate where the corresponding results are from. The performance of LADF [9] is obtained by the code released by the author.
5.2 Experiments on QMUL GRID

The QMUL underGround Re-IDentification (GRID) Dataset [12] is another challenging person re-identification test bed but have not been largely noticed. The GRID dataset was captured from 8 disjoint camera views in a underground station. It contains 250 pedestrian image pairs, with each pair contains two images of the same person from different camera views. Besides, there are 775 additional images that do not belong to the 250 persons which can be used to enlarge the gallery. Sample images from GRID can be found in Fig. 8. It can be seen that these images have poor image quality and low resolutions, and contain large variations of illumination and viewpoint.

An experimental setting of 10 random trials is provided for the GRID dataset. For each trial, 125 image pairs are used for training, and the remaining 125 image pairs, as well as the 775 background images are used for test. A feature set described in [28] is also available for developing machine learning algorithms. This feature is computed from histograms in six equally divided horizontal stripes. 8 colour channels (RGB, HSV and YCbCr) and 21 texture filters (8 Gabor filters and 13 Schmid filters) are used for the histogram representation, resulting in a 2784-dimensional feature vector for each image. This feature is considered to be effective for person re-identification [28] and has demonstrated good performance in several papers [34,28,35].

![Fig. 8. Example pairs of images from the GRID database [12]. Images in the same column represent the same person.](image)

We first applied the proposed method on the provided feature set of GRID. This leads to results of “RQDA+feature” in listed in Table 3. We compared available results from [35] where the same feature set was used. Results shown in Table 3 indicates that the proposed joint dimension reduction and metric learning approach outperforms other distance learning algorithms such as RankSVM [24], PRDC [7], and MRank [35], except that the rank-1 accuracy of RQDA is slightly worse than MRank-RankSVM.
Table 3. Comparison of state-of-the-art results on the GRID database (p=900)

| Method                      | rank=1 | rank=5 | rank=10 | rank=15 | rank=20 |
|-----------------------------|--------|--------|---------|---------|---------|
| L1-norm                     | 4.40   | 11.68  | 16.24   | 19.12   | 24.80   |
| RankSVM                     | 10.24  | 24.56  | 33.28   | 39.44   | 43.68   |
| PRDC                        | 9.68   | 22.00  | 32.96   | 38.96   | 44.32   |
| MRank-RankSVM               | 12.24  | 27.84  | 36.32   | 42.24   | 46.56   |
| MRank-PRDC                  | 11.12  | 26.08  | 35.76   | 41.76   | 46.56   |
| RQDA+feature [28]           | 11.04  | 28.08  | 38.72   | **45.92** | **52.64** |
| RQDA+proposed feature       | **15.20** | **30.08** | **39.20** | **44.72** | **49.28** |

Note: The results of the compared methods are from [35]. Red numbers represents the best results, while blue numbers are the second best ones.

We also tried the proposed feature extraction method, and applied the same RQDA algorithm for metric learning. This corresponds to the results of the last row in Table 3. The comparison shows that the new feature improves the performance at rank 1-10. Especially, a 4% performance gain can be obtained for the rank-1 accuracy. This indicates that the new feature helps to reduce intra-class variations, so that the same person can be recognized at a higher rank.

6 Conclusions

In this paper, we have presented an efficient approach to jointly learn the discriminant low dimensional subspace and the distance metric. We have shown that a projection matrix can be considered in a Restricted Quadratic Discriminant Analysis model for the joint learning. By formulating the Generalized Rayleigh Quotient as the objective, we reach a closed-form solution given by the generalized eigenvalue decomposition. Practical computation issues for RQDA have been discussed, including a simplified computation and the regularization for the covariance estimation. We have also proposed a Retinex and maximum occurrence based feature extraction method, which is robust to both illumination and viewpoint changes. Our experimental results on two challenging public databases, VIPeR and QMUL GRID, indicate that the performance of the proposed method is comparable to the state of the art.

References

1. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. ACM Computing Surveys (CSUR) 46(2) (2013) 29
2. Gong, S., Cristani, M., Yan, S., Loy, C.C.: Person Re-Identification. Springer (2014)
3. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer (2009)
4. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. Advances in neural information processing systems 18 (2006) 1473
5. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: Computer Vision–ACCV 2010. Springer (2011) 501–512
6. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: Proceedings of the 24th international conference on Machine learning, ACM (2007) 209–216
7. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: CVPR. (2011) 649–656
8. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012) 2288–2295
9. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2013) 3610–3617
10. Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? metric learning approaches for face identification. In: Computer Vision, 2009 IEEE 12th International Conference on, IEEE (2009) 498–505
11. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International workshop on performance evaluation of tracking and surveillance, Citeseer (2007)
12. Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1988–1995
13. Hu, Y., Liao, S., Lei, Z., Yi, D., Li, S.Z.: Exploring structural information and fusing multiple features for person re-identification
14. Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: CVPR (2). (2006) 1528–1535
15. Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B.: Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In: ICDSC. (2008) 1–6
16. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.H.: Shape and appearance context modeling. In: ICCV. (2007) 1–8
17. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR. (2010) 2360–2367
18. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer (2012) 413–422
19. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC. Volume 2. (2011) 6
20. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2013) 3586–3593
21. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by salience matching, ICCV (2013)
22. Liu, Y., Shao, Y., Sun, F.: Person re-identification based on visual saliency. In: Intelligent Systems Design and Applications (ISDA), 2012 12th International Conference on, IEEE (2012) 884–889
23. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV (1). (2008) 262–275
24. Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC. Volume 1. (2010) 5
25. Li, W., Wang, X.: Locally aligned feature transforms across views. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE (2013) 3594–3601
26. Liu, C., Loy, C.C., Gong, S., Wang, G.: Pop: Person re-identification post-rank optimisation. In: International Conference on Computer Vision. (2013)
27. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognition 33(11) (2000) 1771–1782
28. Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: what features are important? In: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer (2012) 391–401
29. Land, E.H., McCann, J.: Lightness and retinex theory. JOSA 61(1) (1971) 1–11
30. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: Properties and performance of a center/surround retinex. Image Processing, IEEE Transactions on 6(3) (1997) 451–462
31. Jobson, D.J., Rahman, Z.U., Woodell, G.A.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. Image Processing, IEEE Transactions on 6(7) (1997) 965–976
32. Liao, S., Zhao, G., Kellokumpu, V., Pietikäinen, M., Li, S.Z.: “Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes”. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA (June 2010)
33. Ojala, T., Pietikäinen, M., Harwood, D.: “A comparative study of texture measures with classification based on feature distributions”. Pattern Recognition 29(1) (January 1996) 51–59
34. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 649–656
35. Loy, C.C., Liu, C., Gong, S.: Person re-identification by manifold ranking. In: IEEE International Conference on Image Processing. Volume 20. (2013)