Pattern Recognition Letters 107 (2018) 66–74

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# Person re-identification by kernel null space marginal Fisher analysis

Husheng Dong[a,b], Ping Lu[b], Chunping Liu[a,d,e], Yi Ji[a], Yonggang Li[a], Shengrong Gong[c,a,f,*]

[a] School of Computer Science and Technology, Soochow University, Suzhou, 215006, China
[b] Suzhou Institute of Trade and Commerce, Suzhou, 215009, China
[c] Changshu Institute of Science and Technology, Changshu, 215500, China
[d] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, 130012, China
[e] Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, 210046, China
[f] School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, China

ARTICLE INFO

*Article history:*
Available online 6 December 2017

*MSC:*
41A05
41A10
65D05
65D17

*Keywords:*
Person re-identification
Null space
Marginal Fisher analysis
Kernel method

ABSTRACT

Distance metric learning has been widely applied for person re-identification. However, the typical Small Sample Size (SSS) problem, which is induced by high dimensional feature and limited training samples in most re-identification datasets, may lead to a sub-optimal metric. In this work, we propose to embed samples into a discriminative null space based on Marginal Fisher Analysis (MFA) to overcome the SSS problem. In such a null space, multiple images of the same pedestrian are collapsed to a single point, resulting the extreme Marginal Fisher Criterion. We theoretically analyze the null space and derive its closed-form solution which can be computed very efficiently. To deal with the heavy storage burden in computation, we further extend the proposed method to kernel version, which is called Kernel Null Space Marginal Fisher Analysis (KNSMFA). Experiments on four challenging person re-identification datasets show that KNSMFA uniformly outperforms state-of-the-art approaches.



## 1. Introduction

Person re-identification is the task of matching people observed from non-overlapping camera views, which is a fundamental task in video surveillance. For applications such as human retrieval [1], cross-view tracking [2], and group behavior analysis [3], person re-identification serves as a crucial step. Although great efforts have been devoted for years, person re-identification remains an unsolved problem due to large variations in illumination, body pose, viewpoint and occlusion. The appearance of one person usually changes significantly in different camera views. As a result, it is rather difficult to measure the similarity of cross-view image pairs.

To address person re-identification, varieties of feature representations have been developed to handle the inter-camera visual differences, such as Symmetry-Driven Accumulation of Local Features (SDALF) [4], Biologically Inspired Features (BIF) [5], Salient Color Names (SCN) [6], and Local Maximal Occurrence (LOMO) [7]. Although impressive advancements have been made, designing a more discriminative visual descriptor is still an open problem.

On the other hand, many methods focus on learning an optimal distance metric to measure the similarity between cross-view image pairs [7–11]. With the learned metric, the distance between positive image pairs (images of the same person) should be as small as possible, while the distance between negative image pairs (images of different persons) should be as large as possible. In practice, distance metric learning methods have shown great success in improving the re-identification performance. However, they are still limited by the Small Sample Size (SSS) problem [12,13] which also exists in face recognition as well as other pattern recognition tasks. The SSS problem refers to the within-class scatter matrix is singular when the number of training samples is much smaller than the feature dimension. Since the within-class scatter matrix always appears in the form of matrix inverse, a singular matrix would make many metric learning algorithms run into numerical problems.

In person re-identification, most datasets are relatively small, typically with only hundreds of samples (e.g. VIPeR [14] and PRID450S [15]). In contrast, to capture rich appearance information of pedestrian, the feature representations are usually of rather high dimension, typically as high as thousands or even tens of thousands. As a result, many re-identification distance metric learning algorithms suffer SSS problem heavily. Although some techniques such as matrix regularization or Principle Component Analysis (PCA) can alleviate this, they may make the learned distance metric sub-optimal and less discriminative [12,13,16]. By learning the null space of within-class scatter matrix, the Null Foley–Sammon Transform (NFST) [13,17] and null space based

* Corresponding author.
*E-mail address:* shrgong@suda.edu.cn (S. Gong).

https://doi.org/10.1016/j.patrec.2017.10.032
0167-8655/

Linear Discriminant Analysis (NLDA) [12] have been demonstrated to address SSS problem effectively. Nevertheless, they rely on the assumption that the data of each class is of Gaussian distribution, a foundation of many algorithms of LDA family [7,17,18].

In this paper, we propose an efficient subspace learning method called Null Space Marginal Fisher Analysis (NSMFA) to solve the SSS problem in distance metric learning for person re-identification. As an extension of the Marginal Fisher Analysis (MFA) [19,20], NSMFA can work well without any assumption about the data distribution. To cope with the heavy storage burden in computation, a kernel version of NSMFA is further developed.

The contributions of this work are 2-folds: (1) We propose to solve the SSS problem in re-identification distance metric learning by learning a discriminative null space based on MFA. With rigorous theoretical analysis, we derive a closed-form solution of the null space. (2) The proposed method is further extended to nonlinear case via the kernel method, which is called Kernel Null Space Marginal Fisher Analysis (KNSMFA). We apply the proposed method to person re-identification problem, achieving state-of-the-art performance on four challenging datasets, VIPeR [14], GRID [1], PRID450S [15], and 3DPeS [21].

The rest of this paper is organized as follows. In Section 2, we briefly review the related work. In Section 3, the proposed NSMFA and its kernelized version are presented. In Section 4, the experimental evaluations are demonstrated. Finally, we draw some conclusions in Section 5.

## 2. Related work

Most existing approaches for person re-identification are carried out from two aspects: (1) developing powerful feature representations, (2) learning discriminative distance metrics. Here we briefly review both. For comprehensive surveys, please refer to [22], [23], and [24].

### 2.1. Feature representation

Many approaches try to build distinctive and robust feature representations to capture the invariance of pedestrian's appearance in different camera views. In general, the features are extracted from either horizontal stripes or dense blocks. For stripe-based features, Gray and Tao [14] proposed the Ensemble of Local Features (ELF) which is computed from six horizontal stripes with the same height. Chen et al. [25] extended ELF to ELF18 by computing local features from eighteen stripes. Yang et al. [6] proposed the SCN based on a bag-of-words model. Lisanti et al. [26] proposed the Weighted Histograms of Overlapping Stripes (WHOS) which is extracted from a stripe-based pyramid space. In [27], Lisanti et al. further proposed to extract multiple channel features of color and texture from both stripes and three regions of the pedestrian image.

One drawback of stripe-based features is the fine appearance information of local patches cannot be well explored. A remedy is to compute features from dense blocks. Ma et al. [5] proposed the BIF to present pedestrian images via covariance descriptor, which obtained good robustness against illumination changes and background clutter. Zhao et al. [28] proposed to learn Mid-level filters from the clusters of dense patches. Composed of joint HSV and Scale Invariant Local Ternary Pattern (SILTP) [29] histograms, the LOMO descriptor [7] has shown impressive robustness against viewpoint changes. By describing a local region of pedestrian image via hierarchical Gaussian distributions in which each Gaussian distribution representing the appearance of a local patch, the Gaussian of Gaussian (GOG) descriptor [30] achieved even better performance than LOMO. However, just as one coin has two sides, the dense-block-based features have the shortcoming of failing to capture holistic information of large area. We argue that features should be computed from stripes and dense blocks simultaneously, such that the complementary coarse information and fine details can be both utilized for re-identification.

Recently, there are also some works try to learn feature representations based on deep learning models [31–33]. However, they suffer the bottleneck of small training sample size in most re-identification datasets.

### 2.2. Distance metric learning

Distance metric learning has shown impressive performance in re-identification. Some works learn a Mahalanobis form metric following [34], while others learn a subspace instead. Although seemingly different, they are closely related. By denoting the linear projection of sample $\boldsymbol{x}_i \in \mathbb{R}^d$ as $\boldsymbol{y}_i = \mathbf{W}^\top \boldsymbol{x}_i$, where $\mathbf{W} \in \mathbb{R}^{d\times m} (m < d)$ is the subspace, the Euclidean distance between projected samples is equivalent to the Mahalanobis distance in the original space. This is because $\left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|_2^2 = \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)^\top \mathbf{W}\mathbf{W}^\top \left(\boldsymbol{x}_i - \boldsymbol{x}_j\right)$, where $\mathbf{W}\mathbf{W}^\top$ is just the Mahalanobis distance metric.

There are two most representative methods that learn Mahalanobis distance metric directly, the Large Margin Nearest Neighbor (LMNN) [35] and the Information Theoretic Metric Learning (ITML) [36]. The LMNN tried to learn a metric to separate positive pairs from negative ones by a large margin. In ITML, the metric is learned by minimizing the LogDet divergence. The Locally-Adaptive Decision Functions (LADF) [37] jointly learned the distance metric with a locally adapted thresholding rule for each cross-view image pair. The Keep It Simple and Straightforward Metric Learning (KISSME) [8] derived an efficient closed-form metric by considering a log-likelihood ratio of positive and negative image pairs. As an improvement of KISSME, the Cross-view Quadratic Discriminant Analysis (XQDA) [7] learned a more discriminant metric accompanied by a low-dimensional subspace, which can effectively address the deficiency of two-stage processing shared by most metric learning methods. In [38] the metric matrix was integrated with latent variables to tackle the cross-view misalignment problem. Different from the methods that learn metric from all training samples, Li et al. [39] proposed to learn discriminative dissimilarities on the neighborhood structure manifold.

Among the methods that learn discriminative subspaces, Pedagadi et al. [18] employed the Local Fisher Discriminant Analysis (LFDA) to learn a transformation which can maximize between-class separability and preserve the multi-class modality. Mignon and Jurie [40] proposed the Pairwise Constrained Component Analysis (PCCA) to address the situations when only a sparse set of pairwise constraints on high-dimensional data points is given. From the ranking view, Zheng et al. [11] proposed the PRDC algorithm which learns a transformation from relative distance comparison. Xiong et al. [10] applied a ranking ensemble voting scheme to multiple kernel-based metrics, including regularized PCCA, kernel LFDA, and kernel MFA.

In practice, most subspace learning methods suffer the SSS problem [12,13]. From the connection between Mahalanobis distance metric and subspace learning, it is not difficult to find many Mahalanobis distance metric learning methods suffer SSS problem too. To address this, some works have to perform dimension reduction before learning [8,11,18,35,36]. However, this two-stage processing is sub-optimal because samples of different classes may be cluttered in the dimension reduction phase [7]. To avoid numerical problems, some works resort to matrix regularization such that the within-class scatter matrix is invertible [7,10,18]. But the regularization may lead to the degenerate eigenvalue problem in turn [13]. As an exception, learning the null space of within-class scatter matrix can avoid two-stage processing and degenerate eigenvalue problem effectively. The NFST [17] successfully addressed SSS prob-

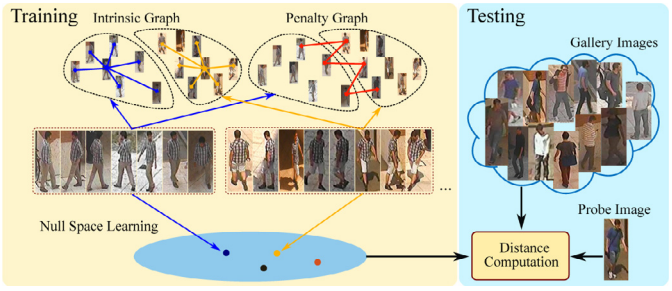


**Fig. 1.** Illustration of the proposed re-identification system. A discriminative null space is learned based on the intrinsic graph and penalty graph defined in MFA. Images of the same person are collapsed to a single point in the learned null space.

lem by employing the null Foley–Sammon transformation. Another superiority of learning a null space is that the samples of the same class are collapsed into a single point in the transformed subspace. Since the within-class variance is zero, the Fisher Criterion will approach infinity if the between-class variance is non-zero. Thus it is optimal to map samples to the null space for pattern recognition. The idea of collapsing samples is also exploited by NLDA [12] and Maximally Collapsing Metric Learning (MCML) [41].

In this work, we propose to learn a discriminative null space for person re-identification by extending MFA to the framework of null space learning. An illustration of the re-identification system is shown in Fig. 1. The proposed method is mostly related to MCML, NLDA and NFST, which share the idea of collapsing samples of the same class to a single point. However, there are notable differences with the proposed method: (1) Our method learns the metric from only the neighborhood manifold by defining an intrinsic graph and a penalty graph, while MCML, NLDA and NFST consider all the training samples. (2) Different from our method that relies on the Laplacian matrices of the defined graphs, there are no Laplacian constraints in MCML, NLDA and NFST. (3) The NLDA and NFST inherit the assumption of LDA that each class is of a Gaussian distribution. In contrast, the proposed method needs no assumptions on the data distribution and thus has wider applicability.

## 3. The proposed approach

Given a training set composed of $N$ images that belong to $C$ pedestrians, we denote the feature matrix as $\mathbf{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) \in \mathbb{R}^{d\times N}$, with each column $\boldsymbol{x}_i \in \mathbb{R}^d$ representing the $i$th image, where $d$ is the feature dimension which is typically very high in person re-identification. Considering $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ that are captured from different camera views, if they both represent the $c$th pedestrian, i.e., $l(\boldsymbol{x}_i) = l(\boldsymbol{x}_j) = c$, then $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a positive pair, and a negative pair otherwise. Our objective is to learn a projection matrix $\mathbf{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m) \in \mathbb{R}^{d\times m} (m < d)$ to embed samples into a discriminative subspace. In such subspace, the Euclidean distance between a positive pair should be shrunk while that of a negative pair should be enlarged. We learn a null space of training samples to achieve this based on MFA.

### 3.1. Review of Marginal Fisher Analysis

As a graph embedding algorithm, MFA employs an intrinsic graph $\mathbf{G}_w = \{\mathbf{X}, \mathbf{A}_w\}$ and a penalty graph $\mathbf{G}_b = \{\mathbf{X}, \mathbf{A}_b\}$ to characterize the within-class compactness and between-class separability [19,20], where $\mathbf{X}$ is the vertex set (i.e., the sample set), $\mathbf{A}_w \in \mathbb{R}^{N\times N}$ and $\mathbf{A}_b \in \mathbb{R}^{N\times N}$ are the similarity matrices defined as

$$\begin{aligned} \mathbf{A}_w(i,j) &= \begin{cases} 1, & \text{if } i \in N^+_{k_w}(j) \text{ or } j \in N^+_{k_w}(i) \\ 0, & \text{else} \end{cases} \\ \mathbf{A}_b(i,j) &= \begin{cases} 1, & \text{if } i \in N^-_{k_b}(l(\boldsymbol{x}_j)) \text{ or } j \in N^-_{k_b}(l(\boldsymbol{x}_i)) \\ 0, & \text{else} \end{cases} \end{aligned} \tag{1}$$

Here, $N^+_{k_w}(i)$ indicates the closest $k_w$ within-class neighbors of sample $\boldsymbol{x}_i$, and $N^-_{k_b}(l(\boldsymbol{x}_i))$ indicates the closest $k_b$ between-class samples of class $l(\boldsymbol{x}_i)$. The task of MFA is to learn a low-dimensional representation of $\mathbf{X}$ that preserves the similarities and differences between vertex pairs in $\mathbf{G}_w$ and $\mathbf{G}_b$.

Considering a projection direction $\boldsymbol{w}$, the within-class compactness is measured as the total distances of within-class sample pairs:

$$\begin{aligned} d_w &= \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{A}_w(i,j) \left\| \boldsymbol{w}^\top \boldsymbol{x}_i - \boldsymbol{w}^\top \boldsymbol{x}_j \right\|_2^2 \\ &= 2\boldsymbol{w}^\top \mathbf{X}\mathbf{L}_w\mathbf{X}^\top \boldsymbol{w}. \end{aligned} \tag{2}$$

Here $\mathbf{L}_w$ is the Laplacian matrix of $\mathbf{G}_w$, which is defined as $\mathbf{L}_w = \mathbf{D}_w - \mathbf{A}_w$, and $\mathbf{D}_w$ is a diagonal matrix with elements $\mathbf{D}_w(i,i) = \sum_{j\neq i} \mathbf{A}_w(i,j), \forall i$.

Using the penalty graph $\mathbf{G}_b$ that illustrates the between-class marginal point adjacency relationship, the between-class separability can be characterized by the following term

$$\begin{aligned} d_b &= \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbf{A}_b(i,j) \left\| \boldsymbol{w}^\top \boldsymbol{x}_i - \boldsymbol{w}^\top \boldsymbol{x}_j \right\|_2^2 \\ &= 2\boldsymbol{w}^\top \mathbf{X}\mathbf{L}_b\mathbf{X}^\top \boldsymbol{w}, \end{aligned} \tag{3}$$

where $\mathbf{L}_b = \mathbf{D}_b - \mathbf{A}_b$ is the Laplacian matrix of $\mathbf{G}_b$, and $\mathbf{D}_b$ is a diagonal matrix with $\mathbf{D}_b(i,i) = \sum_{j\neq i} \mathbf{A}_b(i,j), \forall i$.

By defining within-class scatter matrix $\mathbf{S}_w = \mathbf{X}\mathbf{L}_w\mathbf{X}^\top$ and between-class scatter matrix $\mathbf{S}_b = \mathbf{X}\mathbf{L}_b\mathbf{X}^\top$, the Marginal Fisher Criterion can be formulated as

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \frac{\boldsymbol{w}^\top \mathbf{S}_w \boldsymbol{w}}{\boldsymbol{w}^\top \mathbf{S}_b \boldsymbol{w}}. \tag{4}$$

Eq. (4)can be solved by the generalized eigenvalue decomposition

$$\mathbf{S}_w\boldsymbol{w} = \lambda \mathbf{S}_b \boldsymbol{w}. \tag{5}$$

The final projection directions $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m$ are the eigenvectors corresponding to the $m$ largest eigenvalues of $\mathbf{S}_w^{-1}\mathbf{S}_b$. However, this only holds when $\mathbf{S}_w$ is non-singular. In the case of small sample size, $\mathbf{S}_w$ would be singular, and MFA runs in numerical problems. To address this, MFA has to perform PCA first and apply regularization to $\mathbf{S}_w$ [20]. In this work, we derive a more principled way to overcome this by learning the null space of $\mathbf{S}_w$.

### 3.2. Null Space Marginal Fisher Analysis (NSMFA)

Following the idea of null space learning [13,16,17], we aim to learn a discriminative subspace such that the samples of each class are collapsed to a single point. To this end, we first rewrite the criterion of MFA in Eq. (4) to an equivalent form

$$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^\top \mathbf{S}_b \boldsymbol{w}}{\boldsymbol{w}^\top \mathbf{S}_w \boldsymbol{w}}. \tag{6}$$

It can be obviously found that when $\boldsymbol{w}^\top \mathbf{S}_b \boldsymbol{w} \neq 0$ and $\boldsymbol{w}^\top \mathbf{S}_w \boldsymbol{w} = 0$, the Marginal Fisher Criterion approaches infinity and this leads to the optimal projection $\boldsymbol{w}^*$ automatically. As a result, our objective can be formulated as

$$\begin{aligned} &\max_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_m} \ \boldsymbol{w}_i^\top \mathbf{S}_b \boldsymbol{w}_i \\ &s.t.\ \boldsymbol{w}_i^\top \mathbf{S}_w \boldsymbol{w}_i = 0, \boldsymbol{w}_i^\top \mathbf{S}_b \boldsymbol{w}_i > 0, i = 1, 2, \ldots, m. \end{aligned} \tag{7}$$

Before deriving the solution of $\mathbf{W} = (\boldsymbol{w}_1^*, \boldsymbol{w}_2^*, \ldots, \boldsymbol{w}_m^*)$, we first give some related theorems to show where the projection directions lie. Hereafter, the null space of matrix $\mathbf{A}$ is denoted as $\mathbf{N}(\mathbf{A})$, i.e., $\mathbf{N}(\mathbf{A}) = \{\boldsymbol{x} | \mathbf{A}\boldsymbol{x} = \mathbf{0}\}$, and the orthogonal complement of $\mathbf{N}(\mathbf{A})$ is denoted as $\mathbf{N}(\mathbf{A})^\perp$.

**Theorem 1.** *Let* $\mathbf{A}$ *be a positive semidefinite matrix, then* $\boldsymbol{x}^\top\mathbf{A}\boldsymbol{x}=0$ *iff* $\mathbf{A}\boldsymbol{x}=\mathbf{0}$.

**Proof.** The necessity is obvious, so we only prove the sufficiency here. Since $\mathbf{A}$ is positive semidefinite, we can obtain $\mathbf{A}=\mathbf{L}\mathbf{L}^\top$ by Cholesky decomposition. Then, we have $\boldsymbol{x}^\top\mathbf{A}\boldsymbol{x}=\|\mathbf{L}^\top\boldsymbol{x}\|_2^2=0$, and this gives $\mathbf{L}^\top\boldsymbol{x}=\mathbf{0}$. Thus, $\mathbf{L}\mathbf{L}^\top\boldsymbol{x}=\mathbf{A}\boldsymbol{x}=\mathbf{0}$. This proves the theorem. □

**Theorem 2.** *Let* $\mathbf{W}$ *be a projection matrix that maps samples into* $\mathbf{N}(\mathbf{S}_w)$ *which is spanned by the orthonormal set of* $\mathbf{W}$. *Then the mapped samples of the same class are collapsed to a single point in* $\mathbf{N}(\mathbf{S}_w)$, *i.e., the new within-class scatter matrix* $\tilde{\mathbf{S}}_w$ *is a zero matrix.*

**Proof.** Let $\mathbf{Y}$ be the mapped result of $\mathbf{X}$ by $\mathbf{W}$. That is, $\mathbf{Y}=\mathbf{W}^\top\mathbf{X}$. According to the definition of within-class scatter matrix $\mathbf{S}_w$ and the similarity preserving property of MFA [19,20], we have

$$\tilde{\mathbf{S}}_w=\mathbf{Y}\mathbf{L}_w\mathbf{Y}^\top=\mathbf{W}^\top\mathbf{X}\mathbf{L}_w\left(\mathbf{W}^\top\mathbf{X}\right)^\top=\mathbf{W}^\top\mathbf{S}_w\mathbf{W}.$$

Since $\mathbf{N}(\mathbf{S}_w)$ is spanned by the orthonormal set of $\mathbf{W}$, we have $\mathbf{S}_w\mathbf{W}=\mathbf{0}$. According to Theorem 1, we can obtain $\tilde{\mathbf{S}}_w=\mathbf{0}$. □

**Theorem 3.** *Let* $\mathbf{S}_t=\sum_{i=1}^N(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^\top$ *be the total scatter matrix, where* $\boldsymbol{\mu}=\frac{1}{N}\sum_{i=1}^N\boldsymbol{x}_i$ *is the mean vector of all samples. When the samples are centered, the projection directions* $\boldsymbol{w}_1^*,\boldsymbol{w}_2^*,\ldots,\boldsymbol{w}_m^*$ *which are the solutions of Eq. (7), lie in the space of* $\mathbf{N}(\mathbf{S}_t)^\perp\cap\mathbf{N}(\mathbf{S}_w)\cap\mathbf{N}(\mathbf{S}_b)^\perp$.

**Proof.** For simplification, we denote $\overline{\mathbf{X}}$ the matrix consisting of centered vectors $\boldsymbol{x}_i-\boldsymbol{\mu}$, then $\mathbf{S}_t=\overline{\mathbf{X}}\,\overline{\mathbf{X}}^\top$. We first prove the projection directions lie in $\mathbf{N}(\mathbf{S}_t)^\perp$.

Let $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_t)$ be any one vector selected from $\mathbf{N}(\mathbf{S}_t)$, then $\mathbf{S}_t\boldsymbol{w}=\mathbf{0}$. According to Theorem 1, we have $\boldsymbol{w}^\top\mathbf{S}_t\boldsymbol{w}=\boldsymbol{w}^\top\overline{\mathbf{X}}\,\overline{\mathbf{X}}^\top\boldsymbol{w}=0$, thus $\overline{\mathbf{X}}^\top\boldsymbol{w}=\mathbf{0}$. When the samples are centered, we have $\mathbf{S}_w=\overline{\mathbf{X}}\mathbf{L}_w\overline{\mathbf{X}}^\top$ and $\mathbf{S}_b=\overline{\mathbf{X}}\mathbf{L}_b\overline{\mathbf{X}}^\top$. From the definition of $\mathbf{L}_w$ and $\mathbf{L}_b$, we know $\mathbf{S}_w$ and $\mathbf{S}_b$ are both positive semidefinite matrices. Since $\overline{\mathbf{X}}^\top\boldsymbol{w}=\mathbf{0}$, we have $\boldsymbol{w}^\top\mathbf{S}_w\boldsymbol{w}=0$ and $\boldsymbol{w}^\top\mathbf{S}_b\boldsymbol{w}=0$. So for any projection direction $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_t)$, it has no discriminative power, this in turn proves that $\boldsymbol{w}$ should be selected from $\mathbf{N}(\mathbf{S}_t)^\perp$.

Next, we can easily obtain the desired $\boldsymbol{w}$ lies in $\mathbf{N}(\mathbf{S}_w)$ from Eq. (7) and Theorem 2.

At last, we prove that $\boldsymbol{w}$ lies in $\mathbf{N}(\mathbf{S}_b)^\perp$. Let $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_b)$ be an arbitrary projection direction selected from $\mathbf{N}(\mathbf{S}_b)$, then we have $\boldsymbol{w}^\top\mathbf{S}_b\boldsymbol{w}=0$, that is, $\boldsymbol{w}$ is not discriminative at all. On the contrary, the discriminative power can be guaranteed if $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_b)^\perp$. This is because $\mathbf{S}_b$ is positive semidefinite from above analysis, when $\boldsymbol{w}\notin\mathbf{N}(\mathbf{S}_b)$, $\boldsymbol{w}^\top\mathbf{S}_b\boldsymbol{w}>0$ always holds.

From the above analysis, we obtain the projection direction $\boldsymbol{w}$ should satisfy $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_t)^\perp$, $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_w)$, and $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_b)^\perp$, simultaneously. Thus the desired projections $\boldsymbol{w}_1^*,\boldsymbol{w}_2^*,\ldots,\boldsymbol{w}_m^*$ lie in the space of $\mathbf{N}(\mathbf{S}_t)^\perp\cap\mathbf{N}(\mathbf{S}_w)\cap\mathbf{N}(\mathbf{S}_b)^\perp$. □

From Theorem 3, we know where $\boldsymbol{w}_1^*,\boldsymbol{w}_2^*,\ldots,\boldsymbol{w}_m^*$ lie, the next problem is how to compute them. Let $\{\boldsymbol{w}\}$ be the set of arbitrary projection directions, we first derive how to choose one that satisfies $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_t)^\perp$. From the proof of Theorem 3, we find that $\mathbf{N}(\mathbf{S}_t)=\{\mathbf{z}\in\mathbb{R}^d|\mathbf{S}_t\mathbf{z}=0\}=\{\mathbf{z}\in\mathbb{R}^d|\overline{\mathbf{X}}^\top\mathbf{z}=0\}$. Therefore, $\mathbf{N}(\mathbf{S}_t)^\perp$ is the subspace spanned by columns of $\overline{\mathbf{X}}$. Denoting the orthonormal basis of $\overline{\mathbf{X}}$ as $\mathbf{U}=(\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_{N-1})$, then $\boldsymbol{w}$ can be represented as

$$\boldsymbol{w}=v_1\boldsymbol{u}_1+v_2\boldsymbol{u}_2+\ldots+v_{N-1}\boldsymbol{u}_{N-1}=\mathbf{U}\boldsymbol{v},\tag{8}$$

where $\boldsymbol{v}=(v_1,v_2,\ldots,v_{N-1})^\top$. Note that $\mathbf{U}\in\mathbb{R}^{d\times(N-1)}$ because the rank of $\mathbf{S}_t$ is $N-1$, and $\mathbf{U}$ can be easily obtained by applying QR decomposition to $\overline{\mathbf{X}}$.

The next step is to make $\boldsymbol{w}$ satisfy $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_w)$. By substituting Eq. (8) into Eq. (7), we can obtain

$$\begin{aligned}&\max_{\boldsymbol{v}}\ \boldsymbol{v}^\top\mathbf{S}_b'\boldsymbol{v}\\&s.t.\boldsymbol{v}^\top\mathbf{S}_w'\boldsymbol{v}=0,\boldsymbol{v}^\top\mathbf{S}_b'\boldsymbol{v}>0,\end{aligned}\tag{9}$$

where $\mathbf{S}_b'=\mathbf{U}^\top\mathbf{S}_b\mathbf{U}$ and $\mathbf{S}_w'=\mathbf{U}^\top\mathbf{S}_w\mathbf{U}$. From $\boldsymbol{v}^\top\mathbf{S}_w'\boldsymbol{v}=0$ and Theorem 1, we know that $\boldsymbol{v}$ should be chosen from the eigenvectors corresponding to the zero eigenvalue of $\mathbf{S}_w'$. Let $\mathbf{V}=(\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_r)$ be the matrix composed of possible $\boldsymbol{v}$ (without loss of generality, we assume that $r\geq1$), then the columns of $\mathbf{UV}$, denoted as $(\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\ldots,\boldsymbol{\eta}_r)$, are just those projection directions that lie in $\mathbf{N}(\mathbf{S}_t)^\perp\cap\mathbf{N}(\mathbf{S}_w)$.

The last problem is to make $\boldsymbol{w}$ satisfy the constraint of $\boldsymbol{w}^\top\mathbf{S}_b\boldsymbol{w}>0$ in Eq. (7), which is same as the constraint of $\boldsymbol{v}^\top\mathbf{S}_b'\boldsymbol{v}>0$ in Eq. (9). It is worth noting that due to the randomness of $\boldsymbol{w}$, $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\ldots,\boldsymbol{\eta}_r$ form a basis of $\mathbf{N}(\mathbf{S}_t)^\perp\cap\mathbf{N}(\mathbf{S}_w)$, thus for any $\boldsymbol{w}\in\left(\mathbf{N}(\mathbf{S}_t)^\perp\cap\mathbf{N}(\mathbf{S}_w)\right)\cap\mathbf{N}(\mathbf{S}_b)^\perp$, it can be linearly represented by $\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\ldots,\boldsymbol{\eta}_r$, that is

$$\boldsymbol{w}=\sum_{i=1}^r\boldsymbol{\eta}_ih_i=\mathbf{U}\mathbf{V}\boldsymbol{h},\tag{10}$$

where $\boldsymbol{h}=(h_1,\ldots,h_r)^\top$. Substituting Eq. (10) into Eq. (7), we get

$$\begin{aligned}&\max_{\boldsymbol{h}}\ \boldsymbol{h}^\top\mathbf{S}_b''\boldsymbol{h}\\&s.t.\ \boldsymbol{h}^\top\mathbf{S}_b''\boldsymbol{h}>0.\end{aligned}\tag{11}$$

Here $\mathbf{S}_b''=\mathbf{V}^\top\mathbf{S}_b'\mathbf{V}$, and we ignore the constraint of $\boldsymbol{w}^\top\mathbf{S}_w\boldsymbol{w}=0$ since it has already been satisfied. From $\boldsymbol{w}\in\mathbf{N}(\mathbf{S}_b)^\perp$, we know that $\boldsymbol{h}$ should be selected from the eigenvectors corresponding to the non-zero eigenvalues of $\mathbf{S}_b''$. Let $\mathbf{H}=(\boldsymbol{h}_1,\boldsymbol{h}_2,\ldots,\boldsymbol{h}_s)$ be the eigenvectors corresponding to descendingly sorted non-zero eigenvalues (without loss of generality, we assume that $s\geq1$), then the desired projections $\boldsymbol{w}_1^*,\boldsymbol{w}_2^*,\ldots,\boldsymbol{w}_m^*$ are just the first $m$ columns selected from $\mathbf{UVH}$. That is, the projection matrix $\mathbf{W}$ can be represented as

$$\mathbf{W}=(\boldsymbol{w}_1^*,\boldsymbol{w}_2^*,\ldots,\boldsymbol{w}_m^*)=(\mathbf{UVH})_{\cdot,1:m},\tag{12}$$

where $(\mathbf{A})_{\cdot,1:m}$ indicates the first $m$ columns of $\mathbf{A}$. Furthermore, considering $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{H}$ are all orthogonal matrices, thus we have $\mathbf{W}^\top\mathbf{W}=\mathbf{I}$.

From the above derivation, we know the learning of NSMFA can be solved efficiently by performing QR decomposition once and eigenvalue decomposition twice, and the obtained projection matrix $\mathbf{W}$ has the property of orthogonality.

### 3.3. Kernel Null Space Marginal Fisher Analysis (KNSMFA)

Although NSMFA can solve the SSS problem theoretically, it faces a practical problem. When the feature dimension is too high, the $\mathbf{S}_t$, $\mathbf{S}_w$, and $\mathbf{S}_b$ cannot be computed directly due to heavy burden of storage. We can tackle this problem by employing the kernel method. In the following, we detail the kernelization of NSMFA.

Let $\Phi(\boldsymbol{x}_i)$ be a function that maps sample $\boldsymbol{x}_i$ to an implicit space $\mathcal{H}$ of higher dimension, and $\Phi(\mathbf{X})=(\Phi(\boldsymbol{x}_1),\Phi(\boldsymbol{x}_2),\ldots,\Phi(\boldsymbol{x}_N))$. Then in $\mathcal{H}$, the within-class scatter matrix $\mathbf{S}_w^\Phi$, the between-class scatter matrix $\mathbf{S}_b^\Phi$, and the total scatter matrix $\mathbf{S}_t^\Phi$ can be represented as follows

$$\begin{aligned}\mathbf{S}_w^\Phi&=\Phi(\mathbf{X})\mathbf{L}_w^\Phi\Phi(\mathbf{X})^\top\\\mathbf{S}_b^\Phi&=\Phi(\mathbf{X})\mathbf{L}_b^\Phi\Phi(\mathbf{X})^\top\\\mathbf{S}_t^\Phi&=\Phi(\mathbf{X})(\mathbf{I}-\mathbf{M})(\mathbf{I}-\mathbf{M})^\top\Phi(\mathbf{X})^\top,\end{aligned}\tag{13}$$

where $\mathbf{L}_w^\Phi$ and $\mathbf{L}_b^\Phi$ are the Laplacian matrices of intrinsic graph and penalty graph in $\mathcal{H}$ respectively, $\mathbf{I}$ is the $N\times N$ identity matrix, and

$\mathbf{M}$ is a $N \times N$ matrix with all terms equal to $1/N$. Our objective is

$$\max_{\boldsymbol{w}_1^\Phi,\ldots,\boldsymbol{w}_m^\Phi} \left(\boldsymbol{w}_i^\Phi\right)^\top \mathbf{S}_b^\Phi \boldsymbol{w}_i^\Phi \quad s.t.\ \left(\boldsymbol{w}_i^\Phi\right)^\top \mathbf{S}_w^\Phi \boldsymbol{w}_i^\Phi = 0, \left(\boldsymbol{w}_i^\Phi\right)^\top \mathbf{S}_b^\Phi \boldsymbol{w}_i^\Phi > 0. \tag{14}$$

Similar to NSMFA, the desired $\boldsymbol{w}_i^\Phi$ $(i = 1, \ldots, m)$ lies in the space of $\mathbf{N}\left(\mathbf{S}_t^\Phi\right)^\perp \cap \mathbf{N}\left(\mathbf{S}_w^\Phi\right) \cap \mathbf{N}\left(\mathbf{S}_b^\Phi\right)^\perp$. From [42], we know that any projection direction $\boldsymbol{w}^\Phi$ can be represented as

$$\boldsymbol{w}^\Phi = \sum_{i=1}^{N} \left(\Phi(\boldsymbol{x}_i) - \boldsymbol{\mu}^\Phi\right)\hat{u}_i = \Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\boldsymbol{u}}, \tag{15}$$

where $\boldsymbol{\mu}^\Phi = \frac{1}{N}\sum_{i=1}^{N} \Phi(\boldsymbol{x}_i)$ and $\hat{\boldsymbol{u}} = (\hat{u}_1, \ldots, \hat{u}_N)^\top$. Given a kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}_j) \rangle$, Eq. (14) can be reformulated as

$$\max_{\hat{\boldsymbol{u}}} \hat{\boldsymbol{u}}^\top \mathbf{S}_b^{\mathbf{K}} \hat{\boldsymbol{u}} \quad s.t.\ \hat{\boldsymbol{u}}^\top \mathbf{S}_w^{\mathbf{K}} \hat{\boldsymbol{u}} = 0, \hat{\boldsymbol{u}}^\top \mathbf{S}_b^{\mathbf{K}} \hat{\boldsymbol{u}} > 0, \tag{16}$$

where $\mathbf{S}_w^{\mathbf{K}} = (\mathbf{I} - \mathbf{M})^\top \mathbf{K} \mathbf{L}_w^\Phi \mathbf{K}^\top (\mathbf{I} - \mathbf{M})$, $\mathbf{S}_b^{\mathbf{K}} = (\mathbf{I} - \mathbf{M})^\top \mathbf{K} \mathbf{L}_b^\Phi \mathbf{K}^\top (\mathbf{I} - \mathbf{M})$, and $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the kernel matrix with $\mathbf{K}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. To obtain $\mathbf{L}_w^\Phi$ and $\mathbf{L}_b^\Phi$, the similarity matrices $\mathbf{A}_w^\Phi$ and $\mathbf{A}_b^\Phi$ in $\mathcal{H}$ are computed similar as $\mathbf{A}_w$ and $\mathbf{A}_b$ in Eq. (1), but the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ used for selecting nearest neighbors is computed as

$$d^2(\boldsymbol{x}_i, \boldsymbol{x}_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_i) + k(\boldsymbol{x}_j, \boldsymbol{x}_j) - 2k(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{17}$$

The task of obtaining $\boldsymbol{w}^\Phi \in \mathbf{N}\left(\mathbf{S}_t^\Phi\right)^\perp$ is equivalent to make $\hat{\boldsymbol{u}} \in \mathbf{N}\left(\mathbf{S}_t^{\mathbf{K}}\right)^\perp$, where $\mathbf{S}_t^{\mathbf{K}}$ is

$$\mathbf{S}_t^{\mathbf{K}} = \mathbf{K}(\mathbf{I} - \mathbf{M})(\mathbf{I} - \mathbf{M})^\top \mathbf{K}^\top. \tag{18}$$

We can obtain the orthonormal basis of $\mathbf{N}\left(\mathbf{S}_t^{\mathbf{K}}\right)^\perp$ by performing eigenvalue decomposition of $\mathbf{S}_t^{\mathbf{K}} = \hat{\mathbf{U}}\boldsymbol{\Lambda}\hat{\mathbf{U}}^\top$, where $\boldsymbol{\Lambda}$ is a diagonal matrix with $N-1$ non-zero eigenvalues sorted in descending order, and $\hat{\mathbf{U}} = \left(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \ldots, \hat{\boldsymbol{u}}_{N-1}\right)$ is composed of the corresponding eigenvectors. Then, we compute $\hat{\mathbf{U}}' = \hat{\mathbf{U}}\boldsymbol{\Lambda}^{-1/2}$ to guarantee $\boldsymbol{w}^\Phi$ is normalized.

Let $\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2, \ldots, \hat{\boldsymbol{\xi}}_{N-1}$ be the columns of $\Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'$. Note that they also form the orthonormal basis vectors of $\mathbf{N}\left(\mathbf{S}_t^\Phi\right)^\perp$. So for any $\boldsymbol{w}^\Phi \in \mathbf{N}\left(\mathbf{S}_t^\Phi\right)^\perp \cap \mathbf{N}\left(\mathbf{S}_w^\Phi\right)$, there exist coefficients $\hat{\boldsymbol{v}} = (\hat{v}_1, \ldots, \hat{v}_{N-1})^\top$ such that

$$\boldsymbol{w}^\Phi = \sum_{i=1}^{N-1} \hat{\boldsymbol{\xi}}_i \hat{v}_i = \Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'\hat{\boldsymbol{v}}. \tag{19}$$

Substituting Eq. (19) into $\left(\boldsymbol{w}^\Phi\right)^\top \mathbf{S}_w^\Phi \boldsymbol{w}^\Phi = 0$, we obtain $\hat{\boldsymbol{v}}^\top \tilde{\mathbf{S}}_w^{\mathbf{K}} \hat{\boldsymbol{v}} = 0$, where $\tilde{\mathbf{S}}_w^{\mathbf{K}} = (\hat{\mathbf{U}}')^\top \mathbf{S}_w^{\mathbf{K}} \hat{\mathbf{U}}'$. So $\hat{\boldsymbol{v}}$ should be chosen from the eigenvectors corresponding to the zero eigenvalue of $\tilde{\mathbf{S}}_w^{\mathbf{K}}$. Let $\hat{\mathbf{V}} = (\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_{\hat{r}})$ be these eigenvectors, then the columns of $\Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'\hat{\mathbf{V}}$ denoted as $\left(\hat{\boldsymbol{\eta}}_1, \hat{\boldsymbol{\eta}}_2, \ldots, \hat{\boldsymbol{\eta}}_{\hat{r}}\right)$ are those projection directions that lie in $\mathbf{N}\left(\mathbf{S}_t^\Phi\right)^\perp \cap \mathbf{N}\left(\mathbf{S}_w^\Phi\right)$.

At last, we solve the constraint of $\left(\boldsymbol{w}^\Phi\right)^\top \mathbf{S}_b^\Phi \boldsymbol{w}^\Phi > 0$, i.e., making the obtained $\boldsymbol{w}^\Phi$ lie in $\mathbf{N}\left(\mathbf{S}_b^\Phi\right)^\perp$. Similar to $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_r$ that form a basis of $\mathbf{N}(\mathbf{S}_t)^\perp \cap \mathbf{N}(\mathbf{S}_w)$, $\hat{\boldsymbol{\eta}}_1, \hat{\boldsymbol{\eta}}_2, \ldots, \hat{\boldsymbol{\eta}}_{\hat{r}}$ form a basis of $\mathbf{N}\left(\mathbf{S}_t^\Phi\right)^\perp \cap \mathbf{N}\left(\mathbf{S}_w^\Phi\right)$. So any $\boldsymbol{w}^\Phi$ lies in $\left(\mathbf{N}(\mathbf{S}_t^\Phi)^\perp \cap \mathbf{N}(\mathbf{S}_w^\Phi)\right) \cap \mathbf{N}(\mathbf{S}_b^\Phi)^\perp$ can be linearly represented by $\hat{\boldsymbol{\eta}}_1, \hat{\boldsymbol{\eta}}_2, \ldots, \hat{\boldsymbol{\eta}}_{\hat{r}}$, that is

$$\boldsymbol{w}^\Phi = \sum_{i=1}^{\hat{r}} \hat{\boldsymbol{\eta}}_i \hat{h}_i = \Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'\hat{\mathbf{V}}\hat{\boldsymbol{h}}, \tag{20}$$

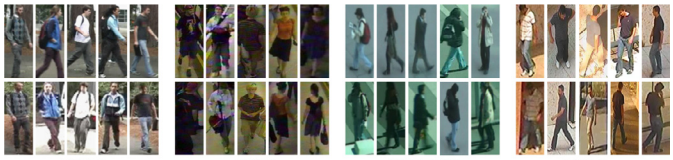


**Fig. 2.** Example image pairs from VIPeR, GRID, PRID450S, and 3DPeS datasets. Images in each column are captured from the same person by different cameras.

where $\hat{\boldsymbol{h}} = (\hat{h}_1, \ldots, \hat{h}_{\hat{r}})^\top$. Substituting Eq. (20) into Eq. (14), we obtain

$$\max_{\hat{\boldsymbol{h}}} \hat{\boldsymbol{h}}^\top \hat{\mathbf{S}}_b^{\mathbf{K}} \hat{\boldsymbol{h}} \quad s.t.\ \hat{\boldsymbol{h}}^\top \hat{\mathbf{S}}_b^{\mathbf{K}} \hat{\boldsymbol{h}} > 0. \tag{21}$$

Here, $\hat{\mathbf{S}}_b^{\mathbf{K}} = \hat{\mathbf{V}}^\top (\hat{\mathbf{U}}')^\top \mathbf{S}_b^{\mathbf{K}} \hat{\mathbf{U}}'\hat{\mathbf{V}}$, and we ignore $\left(\boldsymbol{w}_i^\Phi\right)^\top \mathbf{S}_w^\Phi \boldsymbol{w}_i^\Phi = 0$ since it has already been satisfied. Let $\hat{\mathbf{H}} = (\hat{\boldsymbol{h}}_1, \hat{\boldsymbol{h}}_2, \ldots, \hat{\boldsymbol{h}}_{\hat{s}})$ be the eigenvectors corresponding to the descendingly sorted non-zero eigenvalues of $\hat{\mathbf{S}}_b^{\mathbf{K}}$, then the desired $\boldsymbol{w}_1^\Phi, \boldsymbol{w}_2^\Phi, \ldots, \boldsymbol{w}_m^\Phi$ are just the first $m$ columns selected from $\Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'\hat{\mathbf{V}}\hat{\mathbf{H}}$. That is, the projection matrix $\mathbf{W}^\Phi$ can be represented as

$$\mathbf{W}^\Phi = \left(\Phi(\mathbf{X})(\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'\hat{\mathbf{V}}\hat{\mathbf{H}}\right)_{\cdot,1:m}. \tag{22}$$

Similar to NSMFA, the obtained $\mathbf{W}^\Phi$ is also an orthogonal matrix. Given a new sample $\boldsymbol{x}$, we compute its projected result in space $\mathcal{H}$ as

$$\hat{\boldsymbol{x}} = \left(\mathbf{W}^\Phi\right)^\top \Phi(\boldsymbol{x}) = \left(\left((\mathbf{I} - \mathbf{M})\hat{\mathbf{U}}'\hat{\mathbf{V}}\hat{\mathbf{H}}\right)_{\cdot,1:m}\right)^\top \mathbf{K}_{\boldsymbol{x}}, \tag{23}$$

where $\mathbf{K}_{\boldsymbol{x}} = (k(\boldsymbol{x}_1, \boldsymbol{x}), \ldots, k(\boldsymbol{x}_N, \boldsymbol{x}))^\top$. Given a gallery image $\boldsymbol{x}_g$ and a probe image $\boldsymbol{x}_p$, the distance between them is then computed using the Euclidean distance metric, $d^2(\boldsymbol{x}_g, \boldsymbol{x}_p) = \left\|\hat{\boldsymbol{x}}_g - \hat{\boldsymbol{x}}_p\right\|_2^2$.

## 4. Experiments

### 4.1. Experimental settings

We evaluate the proposed method on four widely used person re-identification datasets: VIPeR [14], GRID [1], PRID450S [15], and 3DPeS [21]. Fig. 2 shows some example image pairs from these datasets. We adopt the single-shot evaluation protocol for all datasets. Each dataset is separated into non-overlapping training set and test set first, then the images in test set are further split into probe set and gallery set that consisted of the images captured from different camera views. Images in the gallery set are matched with everyone in the probe set, and then they are ranked according to the distances. To obtain a robust result, the procedures are repeated 10 times to get an average performance. We employ the Cumulative Matching Characteristics (CMC) curve as our evaluation metric which represents the expectation of finding the right match in the top $r$ matches. For better comparison with other published results, we also report the mean cumulated matching rate at some selected top ranks in tables.

There are three hyper parameters in the proposed method, the within-class/between-class nearest neighbor numbers $k_w$, $k_b$, and the subspace dimension $m$. In experiments, we set $k_w$ to the minimum within-class number of training samples. The values of $k_b$ and $m$ are set to 20 and 100 which are obtained by cross-validation. The effects of $k_w$, $k_b$, and $m$ on re-identification performance are further discussed in Section 4.7. For KNSMFA, we test three widely used kernels including Linear, RBF, and $\chi^2_{rbf}$ on the VIPeR dataset. The kernel that gives the highest performance is selected and applied to all datasets. For the compared models, their hyper parameters are carefully tuned to report the best results.

**Table 1**
Performance comparison (%) of KNSMFA with different kernels and NSMFA on VIPeR dataset. The standard deviations are also presented.

| Method | $r = 1$ | $r = 5$ | $r = 10$ | $r = 20$ |
|---|---|---|---|---|
| KNSMFA + Linear | **55.57** ± 3.20 | **83.29** ± 1.75 | **91.14** ± 1.67 | **97.12** ± 0.74 |
| KNSMFA + RBF | 53.58 ± 2.63 | 81.08 ± 2.17 | 90.22 ± 1.82 | 96.80 ± 0.93 |
| KNSMFA + $\chi^2_{rbf}$ | 50.57 ± 3.12 | 78.70 ± 2.62 | 88.86 ± 1.80 | 95.79 ± 0.88 |
| NSMFA | 53.99 ± 3.16 | 81.84 ± 1.57 | 90.28 ± 1.69 | 95.82 ± 0.71 |

### 4.2. Visual descriptor

We fuse the dense-block-based LOMO with a new stripe-based representation as our visual descriptor. In such way, the fine information from dense blocks and the coarse appearance information from stripes work cooperatively to enhance the discrimination of descriptor. The LOMO feature has been widely used for person re-identification [7,9,17,43], it achieves stability against viewpoint changes by applying max pooling to the patterns among subwindows at the same horizontal location.

For the new designed stripe-based feature, it is computed from a set of overlapping horizontal stripes with the same height. The foregrounds of human images are first obtained by Pedestrian Parsing [44] if no masks provided. Then, eight stripes are divided from the foreground image. After abandoning 1/2 height of top and bottom stripes, the cropped foreground is divided into seven stripes again. We extract four types of basic features from each stripe: SCN [6], joint histograms of RGB and HSV, and SILTP. The SCN feature is extracted as in [6]. The RGB/HSV histograms and SILTP are computed with the same settings of HSV and SILTP in LOMO, but there are no max pooling operations. Our final visual descriptor is the concatenation of LOMO and the features computed from fifteen stripes, with a total dimension of $26960+(8^3 + 8^3 + 3^4 \times 2+16) \times 15 = 44990$. We call our visual descriptor enhanced LOMO (eLOMO). Due to the high dimension of eLOMO, the SSS problem is serious on all tested datasets.

### 4.3. Experiments on VIPeR

There are 632 persons in VIPeR dataset [14] and each has only one pair of images captured from two different cameras in outdoor environment. All images are scaled to 128 × 48 pixels. Due to significant variations in viewpoint, pose and illumination, VIPeR is rather challenging for re-identification. We follow the widely adopted experiment protocol of splitting the dataset equally. That is, 316 persons are randomly selected for training, and the left 316 persons are used for testing.

#### 4.3.1. Kernel selection

Similar to other kernel methods, kernel selection may have a great impact on the performance of KNSMFA. Here, we test KNSMFA with three widely used kernels: Linear, RBF, and $\chi^2_{rbf}$. The parameters of kernel width in RBF and $\chi^2_{rbf}$ are set to mean pairwise distance of training samples. The results of KNSMFA with different kernels as well as NSMFA are summarized in Table 1. Due to the high dimension of eLOMO, the computation of $\mathbf{S}_t$, $\mathbf{S}_b$, and $\mathbf{S}_w$ has to bear heavy memory burden, so the PCA is applied before learning NSMFA. From Table 1, we can observe that the performance of KNSMFA + Linear kernel is the best, which gives 55.57% mean re-identification accuracy on rank-1, whilst only 53.58% and 50.57% are obtained by RBF kernel and $\chi^2_{rbf}$ kernel. Although the RBF and $\chi^2_{rbf}$ kernels are supposed to be more effective [10,25] due to non-linear transformations, they don't perform well with eLOMO. We believe this maybe because the feature dimensions were only 2580 and 8064 in [10] and [25], while it is as high as 44,990 for eLOMO. It is supposed that non-linear kernels may lead to the over-fitting problem in this case. Another observation from Table 1 is that NSMFA also leads to an encouraging result which is even better than that of RBF and $\chi^2_{rbf}$ kernels. Nevertheless, it is lower than KNSMFA + Linear kernel. Due to the higher performance and efficiency of Linear kernel, we will only report the results of KNSMFA + Linear kernel in the following.

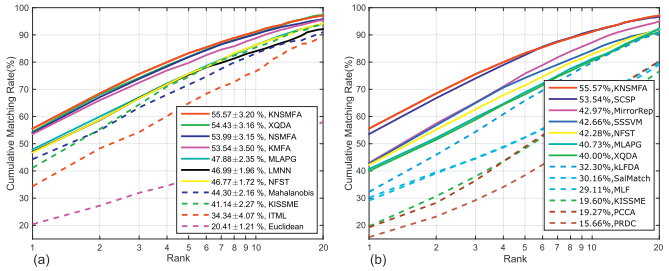


**Fig. 3.** Comparison of CMC curves and mean matching rates at rank-1 on VIPeR dataset. (a) KNSMFA/NSMFA with other metric/subspace learning algorithms using the same descriptor, the subspace dimensions are truncated to 100. The standard deviations at rank-1 are shown in the legend. (b) KNSMFA with other state-of-the-art results.

#### 4.3.2. Comparison of metric/subspace learning algorithms

We then compare the proposed NSMFA and KNSMFA with several state-of-the-art metric learning and subspace learning algorithms, including XQDA [7], KISSME [8], MLAPG [9], LMNN [35], ITML [36], KMFA [10], and NFST [17]. Brief introductions of them have been described in Section 2. All mentioned methods are code-available, so we can evaluate them with the same descriptor. The Euclidean distance and Mahalanobis distance trained with genuine pairs are also evaluated as baselines. For KISSME, LMNN, ITML, Euclidian, and Mahalanobis, the PCA is applied first to reduce the descriptor dimension to 100. For MLAPG and NSMFA, full-energy PCA is applied. The XQDA, NFST, KMFA, and KNSMFA are tested with the original descriptor directly, because they can handle high-dimensional data directly by learning a low-rank subspace. But the learned projection matrices of MLAPG, NSMFA, XQDA, NFST, KMFA, and KNSMFA are truncated to 100 dimensions for a fair comparison.

From the comparison shown in Fig. 3(a), we can find that the proposed KNSMFA outperforms all other algorithms. The obtained rank-1 matching rate of KNSMFA is 55.57%, outperforming the second-best XQDA by 1.14%. Compared to 53.54% of KMFA, there is an improvement of 2.03%, which demonstrates the superiority of matching sample pairs in the null space. Another finding is that all compared algorithms obtain impressive performance with eLOMO. The rank-1 matching rates of KNSMFA, XQDA, NSMFA, and KMFA are 55.57%, 54.43%, 53.99%, and 53.54%, respectively, all exceed 50%. In general, the results of the algorithms without two-stage processing are better than those need to apply PCA first, and this is consistent with the findings in [7].

#### 4.3.3. Comparison to the state of the art

Finally, we compare the performance of KNSMFA with a number of state-of-the-art results reported on VIPeR dataset, including SCSP [45], MPCNN [31], SSSVM [43], NFST [17], MLAPG [9], XQDA [7], MirrorRep [25], KISSME [8], MtMCML [46], kLFDA [10], SCNCD [6], MLF [28], SalMatch [47], PCCA [40], and PRDC [11]. The comparison results are shown in Table 2 and Fig. 3(b). We can see that the proposed KNSMFA outperforms all competitors. Compared to the previous best rank-1 matching rate 53.54% reported by SCSP [45], the improvement is 2.03%. On rank-10, our method is slightly lower than 91.49% of SCSP. Note that the scores of SCSP were obtained by fusing the results of global similarity (computed from the

**Table 2**
Comparison of cumulative matching rates (%) at rank 1, 5, 10 and 20 on VIPeR dataset. For KNSMFA, the standard deviations are also reported.

| Method | $r = 1$ | $r = 5$ | $r = 10$ | $r = 20$ |
|---|---|---|---|---|
| KNSMFA (Ours) | **55.57** ± 3.20 | **83.29** ± 1.75 | 91.14 ± 1.67 | **97.12** ± 0.74 |
| SCSP [45] | 53.54 | 82.59 | **91.49** | 96.65 |
| MPCNN [31] | 47.8 | 74.7 | 84.8 | 91.1 |
| MirrorRep [25] | 42.97 | 75.82 | 87.28 | 94.84 |
| SSSVM [43] | 42.66 | 74.21 | 84.27 | 91.93 |
| NFST [17] | 42.28 | 71.46 | 82.94 | 92.06 |
| MLAPG [9] | 40.73 | 69.94 | 82.34 | 92.37 |
| XQDA [7] | 40.00 | 68.13 | 80.51 | 91.08 |
| SCNCD [6] | 37.80 | 68.50 | 81.20 | 90.40 |
| kLFDA [10] | 32.30 | 65.80 | 79.70 | 90.90 |
| SalMatch [47] | 30.16 | 52.32 | 65.54 | 79.15 |
| MLF [28] | 29.11 | 52.34 | 65.95 | 79.87 |
| KISSME [8] | 19.60 | 49.37 | 62.20 | 77.00 |
| PCCA [40] | 19.27 | 48.89 | 64.91 | 80.28 |
| PRDC [11] | 15.66 | 38.42 | 53.86 | 70.09 |
| MLF + LADF [28] | 43.39 | 73.04 | 84.47 | 93.70 |
| ME [48] | 45.90 | 77.50 | 88.90 | 95.80 |
| NFST (fusion) [17] | 51.17 | 82.09 | 90.51 | 95.52 |

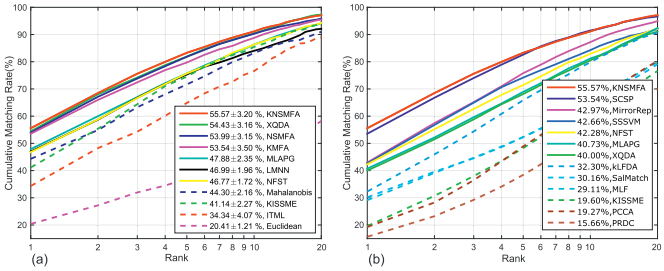


**Fig. 4.** Comparison of CMC curves and mean matching rates at rank-1 on GRID dataset. (a) KNSMFA/NSMFA with other metric/subspace learning algorithms using the same descriptor, the subspace dimensions are truncated to 100. The standard deviations at rank-1 are also shown in the legend. (b) KNSMFA with other state-of-the-art results.

whole image) and local similarities (computed from local regions), the rank-1 matching rate obtained by global similarity was 48.10% [45].

We also compare our approach with MLF + LADF [28], ME [48], and NFST (fusion) [17], that fused different features or different models. From the results shown in the last three rows in Table 2, we can find our method outperforms the nearest rival NFST (fusion) by 4.4% on rank-1, showing much better performance.

### 4.4. Experiments on GRID

The GRID dataset [1] was captured by 8 disjoint cameras in an underground station, with images organized in a probe set and a gallery set. In both probe set and gallery set, there are 250 persons that each has one image pair captured from different camera views. Besides, there are additional 750 images do not match any probe image in the gallery set. From Fig. 2, we can find the images in this dataset are of poor quality and low resolutions, which makes the matching task rather difficult. We scale all images to 128 × 64 pixels for experiment. Following the dataset partition in [7], the 250 persons are equally split into two halves, one for training, and the other half along with the 775 extra images are used for testing. So there are 125 probe images and 900 gallery images in testing phase.

Similar to experiments on VIPeR dataset, we first evaluate the proposed KNSMFA/NSMFA and several metric/subspace learning algorithms with the same eLOMO descriptor. For a fair comparison, the subspace dimensions are truncated to 100 for all evaluated algorithms. From the performance comparison shown in Fig. 4(a),

**Table 3**
Comparison of cumulative matching rates (%) at rank 1, 5, 10 and 20 on GRID dataset. For KNSMFA, the standard deviations are also reported.

| Method | $r = 1$ | $r = 5$ | $r = 10$ | $r = 20$ |
|---|---|---|---|---|
| KNSMFA (Ours) | **24.72** ± 2.73 | **47.12** ± 3.07 | **58.40** ± 3.35 | **70.48** ± 4.05 |
| SCSP [45] | 24.24 | 44.56 | 54.08 | 65.20 |
| SSSVM [43] | 22.40 | 40.40 | 51.28 | 61.20 |
| MLAPG [9] | 16.64 | 33.12 | 41.20 | 52.96 |
| XQDA [7] | 16.56 | 33.84 | 41.84 | 52.40 |
| PolyMap [49] | 16.30 | 35.80 | 46.00 | 57.60 |
| MtMCML [46] | 14.08 | 34.64 | 45.84 | 59.84 |
| Mrank-PRDC [50] | 11.12 | 26.08 | 35.76 | 46.56 |
| Mrank-RankSVM [50] | 12.24 | 27.84 | 36.32 | 46.56 |

**Table 4**
Comparison of cumulative matching rates (%) at rank 1, 5, 10 and 20 on PRID450S dataset. For KNSMFA, the standard deviations are also reported.

| Method | $r = 1$ | $r = 5$ | $r = 10$ | $r = 20$ |
|---|---|---|---|---|
| KNSMFA (Ours) | **78.67** ± 2.60 | **94.49** ± 1.68 | **97.33** ± 1.20 | **98.98** ± 0.56 |
| LOMO+XQDA [7] | 61.42 | – | 90.84 | 95.33 |
| SSSVM [43] | 60.49 | 82.93 | 88.58 | 93.60 |
| MirrorRep [25] | 55.42 | 79.29 | 87.82 | 93.87 |
| SCNCD [6] | 42.44 | 69.22 | 79.56 | 88.44 |
| ECM [51] | 41.90 | 66.30 | 76.90 | 84.90 |
| KISSME [8] | 33.47 | 59.82 | 70.84 | 79.47 |
| EIML [52] | 34.71 | 57.73 | 67.91 | 77.33 |

we can observe that the re-identification task on GRID dataset is much more difficult than that on VIPeR dataset. This is because GRID is captured from 8 camera views while VIPeR contains only two. The KNSMFA consistently achieves higher accuracy than other algorithms, the rank-1 matching rate is as high as 24.72%, outperforming the second best 23.68% of XQDA by 1.04%. Interestingly, although KISSME performs well on VIPeR dataset, it cannot handle the complex viewpoint changes and only obtains 12.80% rank-1 accuracy. In contrast, KNSMFA is more robust than compared algorithms in addressing such challenges.

Next, we compare the proposed KNSMFA with eight state-of-the-art methods, including SCSP [45], SSSVM [43], XQDA [7], MLAPG [9], Polymap [49], MtMCML [46], Mrank-PRDC [50], and Mrank-RankSVM [50]. The results are shown in Table 3 and Fig. 4(b). We can find the proposed KNSMFA outperforms all previous state-of-the-arts. Compared to the previous best 24.24% of SCSP [45], the improvement on rank-1 is 0.48%, and about average 4% on larger ranks. This indicates KNSMFA is promising in handling the poor image quality and complex viewpoint changes in GRID dataset.

### 4.5. Experiments on PRID450S

The PRID450S dataset [15] consists of 450 persons and each one has a single-shot image pair captured by two disjoint cameras. The images in different camera views take serious variations in viewpoint, partial occlusion and background interference, making it also a challenging dataset. In experiments, the images are normalized to 128 × 64 pixels and the provided masks are used for extracting stripe-based feature in eLOMO. We randomly select half of the image pairs for training and the left are used as test set.

Seven approaches that have reported results on PRID450S are compared with the proposed KNSMFA, the comparison of the results is shown in Table 4. We can find the proposed method achieves 78.67% rank-1 matching rate, outperforming the existing best 61.42% by 17.25%.

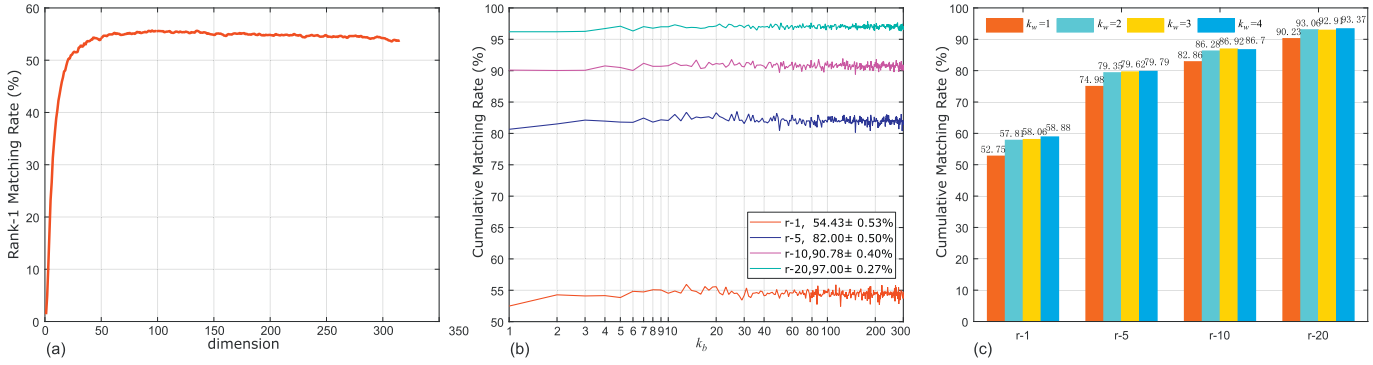


**Fig. 5.** Analysis of KNSMFA: (a) Rank-1 matching rate w.r.t subspace dimension, (b) cumulative matching rates (%) at rank 1, 5, 10, and 20 w.r.t $k_b$, the numbers in the legend indicate the mean values and corresponding standard deviations, (c) performance comparison with different $k_w$. (a) and (b) are carried out on VIPeR, (c) is performed on 3DPeS dataset.

**Table 5**
Comparison of cumulative matching rates (%) at rank 1, 5, 10 and 20 on 3DPeS dataset. For KNSMFA, the standard deviations are also reported.

| Method | $r = 1$ | $r = 5$ | $r = 10$ | $r = 20$ |
|---|---|---|---|---|
| KNSMFA (Ours) | **58.88** ± 1.66 | **79.79** ± 2.70 | **86.70** ± 2.49 | **93.37** ± 1.96 |
| SCSP [45] | 57.29 | 78.97 | 86.02 | 91.51 |
| DGD [33] | 56.0 | – | – | – |
| ME [48] | 53.33 | 76.79 | 84.95 | 92.78 |
| kLFDA [10] | 54.02 | 77.74 | 85.90 | 92.38 |
| LF [18] | 33.47 | – | 69.98 | – |

### 4.6. Experiments on 3DPeS

The 3DPeS dataset [21] contains 1011 images of 192 persons captured from 8 cameras in an academic campus. Each person has 2–26 images. The images were collected during different times of the day, resulting in large variations of lighting conditions as can be observed in Fig. 2. In experiments, we scale all images to 128 × 64 pixels and use the provided masks to extract stripe-based feature.

The comparison results are summarized in Table 5. It can be observed that the proposed KNSMFA achieves the highest accuracy on all ranks. On the most important rank-1, KNSMFA achieves an accuracy of 58.88%, outperforming the previous best 57.29% of SCSP by 1.59%. The impressive result of KNSMFA indicates it has better capability in learning from complex data.

### 4.7. Analysis

**Influence of subspace dimensions.** To understand the influence of the subspace dimensions, we evaluate KNSMFA on VIPeR with different values of $m$ in Eq. (22). From the result shown in Fig. 5(a), we can observe that: 1) even with a dimension as low as 50, a rank-1 matching rate of higher than 50% can be obtained. This means that a discriminative low-rank representation can be learned effectively; and 2) the matching rate keeps increasing with increasing dimensions, and it achieves the highest at about 100. Then it generally remains stable and there is a slight decrease with higher $m$. Therefore, we choose $m = 100$ as the subspace dimension. **Influence of** $k_b$ and $k_w$**.** Similar to MFA, the parameters $k_b$ and $k_w$ also play an important role in KNSMFA. The $k_b$ controls the between-class marginal point adjacency relations. Larger $k_b$ means more negative image pairs are considered in training. Whilst in [53], it was pointed out that not all negative pairs help to enhance the discrimination of the learned metric. So it is not a good choice to set $k_b$ too large. Here we investigate the cumulative matching rates at rank 1, 5, 10, and 20 w.r.t $k_b$ on the VIPeR dataset. From the results shown in Fig. 5(b), we can find the matching rates generally increase with increasing $k_b$ at the initial stage. When $k_b > 10$, the matching rates begin to fluctuate within small ranges. In general, it is reasonable to set $k_b$ to 20 for the balance of computational cost and performance.

**Table 6**
Training time (seconds) of different algorithms.

| Method | KISSME | XQDA | NFST | KNSMFA | LMNN | MLAPG | ITML |
|---|---|---|---|---|---|---|---|
| Time | 3.7 | 3.5 | 1.4 | 1.5 | 269.3 | 62.1 | 195.8 |

As for $k_w$, it controls the number of within-class samples used for training. In person re-identification, the number of cross-view positive image pairs is always much smaller than negative pairs. This leads to rather imbalanced training data. To make the best use of positive pairs, the $k_w$ should be set as large as possible such that more within-class samples are involved. However, the numbers of within-class samples are not always the same for every class (e.g. 3DPeS dataset). So we set $k_w$ to the minimal within-class number of training samples. Note that $k_w = 1$ on VIPeR, GRID, and PRID450S, because there is only one image pair for each person in these datasets. On the 3DPeS dataset, we compare the results obtained with $k_w$ in the range of 1 to the minimal within-class number. From Fig. 5 (c) we can see that, the re-identification performance keeps rising as $k_w$ increases. This confirms that more positive pairs can benefit the null space learning.

**Training time.** Table 6 gives a training time (including the time of computing kernel matrix and PCA) comparison of KNSMFA with other metric/subspace learning algorithms. The training time of all algorithms are obtained by averaging 10 random trials on the VIPeR dataset with the same eLOMO descriptor. All algorithms are implemented in MATLAB and run on a desktop PC with Intel i7-3720 @2.6 GHz CPU. From the comparison we can find that our KNSMFA is as efficient as NSFT, and faster than KISSME and XQDA that have closed-form solutions too, not to mention the rest. The LMNN, MLAPG, and ITML are rather time consuming, because they need iterative optimization.

## 5. Conclusion and future work

In this paper, we have proposed a novel subspace learning algorithm called NSMFA for person re-identification. With rigorous theoretical analysis, we derive its closed-form solution and further extend it to the kernel version. The obtained KNSMFA not only well addresses the SSS problem in re-identification distance metric learning, but could also avoid the heavy storage burden in the computation of NSMFA. Experiments on four challenging datasets show that the proposed method achieves better performance compared to state-of-the-art approaches. In future work, we

will consider applying the proposed method to cross-dataset re-identification. As the proposed KNSMFA relies on the existence of SSS problem incurred by high dimensional features, we will also explore its sensitiveness to the noise in high dimensional data.

## Acknowledgment

This work was partially supported by National Natural Science Foundation of China (NSFC grant no. 61170124, 61272258, 61301299, 61272005, 61572085), Provincial Natural Science Foundation of Jiangsu (grant no. BK20151254, BK20151260), Six Talents Peaks project of Jiangsu (DZXX-027), Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University (grant no. 93K172016K08), and Collaborative Innovation Center of Novel Software Technology and Industrialization.

## References

[1] C.L. Chen, T. Xiang, S. Gong, Multi-camera activity correlation analysis, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009, pp. 1988–1995.
[2] B. Wang, G. Wang, K.L. Chan, L. Wang, Tracklet association with online target-specific metric learning, in: Computer Vision and Pattern Recognition, 2014, pp. 1234–1241.
[3] W.S. Zheng, S. Gong, T. Xiang, Group association: assisting re-identification by visual context, Person Re-identification 14 (2014) 183–201.
[4] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2360–2367.
[5] B. Ma, Y. Su, F. Jurie, Covariance descriptor based on bio-inspired features for person re-identification and face verification, Image. Vis. Comput. 32 (6) (2014) 379–390.
[6] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S.Z. Li, Salient color names for person re-identification, in: Computer Vision–ECCV 2014, 2014, pp. 536–551.
[7] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
[8] M. Köestinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2288–2295.
[9] S. Liao, S.Z. Li, Efficient psd constrained asymmetric metric learning for person re-identification, in: IEEE International Conference on Computer Vision, 2015, pp. 3685–3693.
[10] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: Computer Vision–ECCV 2014, Springer, 2014, pp. 1–16.
[11] W.-S. Zheng, S. Gong, T. Xiang, Reidentification by relative distance comparison, Pattern Anal. Mach. Intell. IEEE Trans. 35 (3) (2013) 653–668.
[12] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, G.J. Yu, A new lda-based face recognition system which can solve the small sample size problem, Pattern Recognit. 33 (10) (2000) 1713–1726.
[13] Z. Wenming, Z. Li, Z. Cairong, Foley-sammon optimal discriminant vectors using kernel approach, IEEE Trans. Neural Netw. 16 (1) (2005) 1–9.
[14] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: Computer Vision–ECCV 2008, Springer, 2008, pp. 262–275.
[15] P.M. Roth, M. Hirzer, M. Kostinger, C. Beleznai, H. Bischof, Mahalanobis Distance Learning for Person Re-identification, 2014.
[16] J. Ye, T. Xiong, Null space versus orthogonal linear discriminant analysis, in: International Conference, 2006, pp. 1073–1080.
[17] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1239–1248.
[18] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3318–3325.
[19] D. Xu, S. Yan, D. Tao, S. Lin, H.J. Zhang, Marginal fisher analysis and its variants for human gait recognition and content- based image retrieval., IEEE Trans. Image Process. 16 (11) (2007) 2811.
[20] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, Pattern Anal. Mach. Intell. IEEE Trans. 29 (1) (2007) 40–51.
[21] D. Baltieri, R. Vezzani, R. Cucchiara, Sarc3d: A new 3d body model for people tracking and re-identification, in: International Conference on Image Analysis and Processing, 2011, pp. 197–206.
[22] G. Doretto, T. Sebastian, P. Tu, J. Rittscher, Appearance-based person reidentification in camera networks: problem overview and current approaches, J. Ambient Intell. Humaniz. Comput. 2 (2) (2011) 127–151.
[23] S. Gong, M. Cristani, S. Yan, C.C. Loy, Person Re-identification, Springer, 2014.
[24] S.W. Henggeler, S.K. Schoenwald, J.G. Liao, E.J. Letourneau, D.L. Edwards, A survey of approaches and trends in person re-identification, Image. Vis. Comput. 32 (4) (2014) 270–286.
[25] Y.C. Chen, W.S. Zheng, J. Lai, Mirror representation for modeling view-specific transform in person re-identification, in: International Conference on Artificial Intelligence, 2015, pp. 3402–3408.
[26] G. Lisanti, I. Masi, A.D. Bagdanov, A.D. Bimbo, Person re-identification by iterative re-weighted sparse ranking, Pattern Anal. Mach. Intell. IEEE Trans. 37 (8) (2015) 1629–1642.
[27] G. Lisanti, S. Karaman, I. Masi, Multichannel-kernel canonical correlation analysis for cross-view person reidentification, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 13 (2) (2017) 13.
[28] R. Zhao, W. Ouyang, X. Wang, Learning mid-level filters for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 144–151.
[29] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes, in: IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 1301–1306.
[30] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical gaussian descriptor for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1363–1372.
[31] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng, Person re-identification by multi-channel parts-based cnn with improved triplet loss function, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1335–1344.
[32] S.Z. Chen, C.C. Guo, J.H. Lai, Deep ranking for person re-identification via joint representation learning, IEEE Trans. Image Process. 25 (5) (2016) 2353.
[33] T. Xiao, H. Li, W. Ouyang, X. Wang, Learning deep feature representations with domain guided dropout for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[34] E.P. Xing, A.Y. Ng, M.I. Jordan, S. Russell, Distance metric learning, with application to clustering with side-information, Adv. Neural Inf. Process. Syst. 15 (2002) 505–512.
[35] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244.
[36] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Proceedings of the 24th international conference on Machine learning, ACM, 2007, pp. 209–216.
[37] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, J. Smith, Learning locally-adaptive decision functions for person verification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3610–3617.
[38] C. Sun, D. Wang, H. Lu, Person re-identification via distance metric learning with latent variables, IEEE Trans. Image Process. 26 (1) (2017) 23–34.
[39] W. Li, Y. Wu, J. Li, Re-identification by neighborhood structure metric learning, Pattern Recognit. 61 (2017) 327–338.
[40] A. Mignon, F. Jurie, Pcca: A new approach for distance learning from sparse pairwise constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2666–2672.
[41] A. Globerson, S.T. Roweis, Metric learning by collapsing classes., Adv. Neural Inf. Process. Syst. 18 (2005) 451–458.
[42] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[43] Y. Zhang, B. Li, H. Lu, A. Irie, X. Ruan, Sample-specific svm learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1278–1287.
[44] P. Luo, X. Wang, X. Tang, Pedestrian parsing via deep decompositional network, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2648–2655.
[45] D. Chen, Z. Yuan, B. Chen, N. Zheng, Similarity learning with spatial constraints for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1268–1277.
[46] L. Ma, X. Yang, D. Tao, Person re-identification over camera networks using multi-task distance metric learning, IEEE Trans. Image Process. 23 (8) (2014) 3656–3670. A Publication of the IEEE Signal Processing Society
[47] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2013, pp. 3586–3593.
[48] S. Paisitkriangkrai, C. Shen, V.D.H. Anton, Learning to rank in person re-identification with metric ensembles, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1846–1855.
[49] D. Chen, Z. Yuan, G. Hua, N. Zheng, J. Wang, Similarity learning on an explicit polynomial kernel feature map for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1565–1573.
[50] C.L. Chen, C. Liu, S. Gong, Person re-identification by manifold ranking, in: IEEE International Conference on Image Processing, 2013, pp. 3567–3571.
[51] X. Liu, H. Wang, Y. Wu, J. Yang, An ensemble color model for human re-identification, in: Applications of Computer Vision, 2015, pp. 868–875.
[52] M. Hirzer, P.M. Roth, H. Bischof, Person re-identification by efficient impostor-based metric learning, in: IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, 2012, pp. 203–208.
[53] J. García, N. Martinel, A. Gardel, I. Bravo, G.L. Foresti, C. Micheloni, Discriminant context information analysis for post-ranking person re-identification, IEEE Trans. Image Process. 26 (4) (2017) 1650–1665.