Pattern Recognition Letters 000 (2017) 1–9

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 xxx

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-

Please cite this article as: H. Dong et al., Person re-identification by kernel null space marginal Fisher analysis, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.032