Non-negative Dual Graph Regularized Sparse Ranking for Multi-shot Person Re-identification

Aihua Zheng, Hongchao Li, Bo Jiang(✉), Chenglong Li, Jin Tang, and Bin Luo

School of Computer Science and Technology, Anhui University, Hefei, China
{ahzheng214,tj,luobin}@ahu.edu.cn, {lhc950304,lcl1314}@foxmail.com,
zeyiabc@163.com

Abstract. Person re-identification (Re-ID) has recently attracted enthusiastic attention due to its potential applications in social security and smart city surveillance. The promising achievement of sparse coding in image based recognition gives rise to a number of development on Re-ID especially with limited samples. However, most of existing sparse ranking based Re-ID methods lack of considering the geometric structure on the data. In this paper, we design a non-negative dual graph regularized sparse ranking method for multi-shot person Re-ID. First, we enforce a global graph regularizer into the sparse ranking model to encourage the probe images from the same person generating similar coefficients. Second, we enforce additional local graph regularizer to encourage the gallery images of the same person making similar contributions to the reconstruction. At last, we impose the non-negative constraint to ensure the meaningful interpretation of the coefficients. Based on these three cues, we design a unified sparse ranking framework for multi-shot Re-ID, which aims to simultaneously capture the meaningful geometric structures within both probe and gallery images. Finally, we provide an iterative optimization algorithm by Accelerated Proximal Gradient (APG) to learn the reconstruction coefficients. The ranking results of a certain probe against given gallery are obtained by accumulating the redistributed reconstruction coefficients. Extensive experiments on three benchmark datasets, i-LIDS, CAVIARA4REID and MARS with both hand-crafted and deep features yield impressive performance in multi-shot Re-ID.

Keywords: Person re-identification · Sparse ranking
Dual graph regularization · Non-negativity

1 Introduction

Person re-identification (Re-ID), which aims to identify person images from the gallery that shares the same identity as the given probe, is an active task driven by the applications of visual surveillance and social security. Despite of years of

J.-H. Lai et al. (Eds.): PRCV 2018, LNCS 11256, pp. 108–120, 2018.
https://doi.org/10.1007/978-3-030-03398-9_10

extensive efforts [2,5,30–32], it still faces various challenges due to the changes of illumination, pose, camera view and occlusions.

From the data point of view, Re-ID task fails into two categories: (1) Single-shot Re-ID, where only a single image is recorded for each person under each camera view. Despite of extensive studied in recent years [16,21,29], the performance is restrained by the limited information in a single person image. (2) Multi-shot Re-ID, where multiple frames are recorded for each person, is more realistic in real-life applications with more visual aspects. We focus on multi-shot Re-ID in this paper. The main stream of solving Re-ID problem devote to two aspects or both: (1) Appearance modeling [2,3,29,30], which develops a robust feature descriptor to leverage the various changes and occlusions between cameras. (2) Learning-based methods [16,18,21,31], which learns a metric distance to mitigate the appearance gaps between the low-level features and the high-level semantics. Recently, deep neural networks have made a remarkable progress on feature learning for Re-ID [14,17,24,28,34]. However, most of existing methods require large labor of training procedure.

Sparse ranking [22], as a powerful subspace learning and representation technique, has been successfully applied to extensive image based applications which gives rise to a number of development on Re-ID. The basic idea is to characterize the probe image as a linear combination of few items/images from an over-complete dictionary gallery. Liu et al. [19] proposed to learn two coupled dictionaries for both probe and gallery from both labeled and unlabeled images to transfer the features of the same person from different cameras. Karanam et al. [12] learnt a single dictionary for both gallery and probe images to overcome the viewpoint and associated appearance changes and then discriminatively trained the dictionary by enforcing explicit constraints on the associated sparse representations. Zheng et al. [33] proposed a weight-based sparse coding approach to reduce the influence of abnormal residuals caused by occlusion and body variation. Lisanti et al. [18] proposed to learn a discriminative sparse basis expansions of targets in terms of a labeled gallery of known individuals followed by a soft- and hard- re-weighting to redistribute energy among the most relevant contributing elements. Jing et al. [11] proposed a semi-coupled low-rank discriminant dictionary learning with discriminant term and a low-rank regularization term to characterize intrinsic feature space with different resolution for Re-ID.

However, most of existing sparse ranking based Re-ID methods encoded the probe images from the same person independently therefore failed to take advantage of their intrinsic geometric structure information, especially in multi-shot Re-ID. As we observed that, the same person under the same camera are generally with similar appearance. Therefore, we argue to preserve this geometrical structure embedded in both probe and gallery images. Inspired by the great superiority of graph regularized sparse coding in image based applications [9,25,26], we propose to explore the intrinsic geometry in multi-shot Re-ID via a non-negative dual graph regularized sparse ranking approach in this paper. After rendering the Re-ID task as sparse coding based multi-class classification problem, we first explore the global geometrical structure by enforcing the

smoothness between the coefficients referring the images from the same person in probe. Then, we explore the local geometrical structure by encouraging the images from the same person in the gallery making similar contributions while reconstructing a certain probe image. The optimized coefficients considering both global and local information are obtained via iterative optimization by Accelerated Proximal Gradient (APG) [20]. The final rankings of the certain probe against given gallery are achieved by accumulating the reconstruction coefficients.

2 Problem Statement

Given $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in R^{d \times n}$, where n denotes the number of images of a person in probe, where $\mathbf{x}_j \in R^{d \times 1}, j = \{1, \ldots, n\}$ denotes the corresponding d-dimensional feature. While $\mathbf{D} = [\mathbf{D}^1, \mathbf{D}^2, \ldots, \mathbf{D}^G] \in R^{d \times M}$ denotes the total M images of G persons in gallery, where $\mathbf{D}^p = [\mathbf{d}_1^p, \mathbf{d}_2^p, \ldots, \mathbf{d}_{g_p}^p] \in R^{d \times g_p}, p = \{1, \ldots, G\}$ represents the matrix of g_p basis feature vectors for the p-th person, g_p denotes the number of images of the p-th person in gallery. Obviously, $M = \sum_{\mathrm{p}=1}^{G} g_p$. The basic idea of sparse ranking based Re-ID is to reconstruct a testing probe image $\mathbf{x}_j$ with linear spanned training gallery images of G persons:

$$\begin{aligned} \mathbf{x}_j &\approx \sum_{p=1}^{G} \mathbf{D}^p \mathbf{c}_j^p \\ &= \mathbf{D}\mathbf{c}_j \end{aligned} \tag{1}$$

where $\mathbf{c}_j^p = [\mathbf{c}_{j,1}^p, \mathbf{c}_{j,2}^p, \ldots, \mathbf{c}_{j,g_p}^p]^T \in R^{g_p \times 1}$ represents the coding coefficients of the p-th person against the probe instance $\mathbf{x}_j$. The dictionary $\mathbf{D}$ can be highly overcomplete. In order to concentratively reconstruct the probe via relatively few dictionary atoms from the gallery, we can impose the sparsity constraint into above formulation as an ℓ_1-norm regularized least squares problem:

$$\min_{\mathbf{c}_j} \|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2^2, \ s.t. \ \|\mathbf{c}_j\|_1 \leq \epsilon, \quad \epsilon > 0 \tag{2}$$

where ϵ is error bound of the sparsity. It is equivalent to the LASSO problem [7], which could be formulated as

$$\min_{\mathbf{c}_j} \|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2^2 + \lambda \|\mathbf{c}_j\|_1 \tag{3}$$

where λ controls the tradeoff between minimization of the ℓ_2 reconstruction error and the ℓ_1-norm of the sparsity used to reconstruct $\mathbf{x}_j$.

It worth noting that Eq. (3) reconstructs each probe image independently while ignoring the intrinsic geometry within the probe images. Moreover, it lacks of considering the dependency of the dictionary atoms in gallery when reconstructing a certain probe.

3 Non-negative Dual Graph Regularized Sparse Ranking

Based on above discussion, we design a non-negative dual graph regularized sparse ranking (NNDGSR) to simultaneously exploit the global and local geometric structures in both probe and gallery for multi-shot Re-ID.

3.1 Dual Graph Regularized Sparse Ranking

Global Graph Regularization. On the one hand, we argue that the feature vectors derived from the multiple images of the same person tend to have similar geometric distribution. To exploit the intrinsic geometric distribution among the probe images, we first enforce a global graph regularizer over the reconstruction coefficients:

$$\min_{\mathbf{c}_j} \sum_{j=1}^{n} \|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2^2 + \lambda\|\mathbf{c}_j\|_1 + \frac{1}{2}\beta \sum_{i,j\in\{1,\ldots,n\}} \|\mathbf{c}_i - \mathbf{c}_j\|_2^2 \mathbf{S}_{i,j}, \tag{4}$$

$\{\mathbf{c}_i, \mathbf{c}_j\} \in R^{M\times 1}$ is the reconstruction coefficients of images $\mathbf{x}_i$ and $\mathbf{x}_j$ from the same person over gallery dictionary $\mathbf{D}$ respectively. β is a balance parameter controlling the contribution of the regularizer. The similarity matrix $\mathbf{S} \in R^{n\times n}$ is defined as:

$$\mathbf{S}_{i,j} = exp(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma_1^2}), \tag{5}$$

where σ_1 is a parameter fixed as 0.2 in this paper. The global regularizer in Eq. (4) encourages the probe images from the same person with higher similarity to generate closer coefficients during reconstruction.

Local Graph Regularization On the other hand, we further argue that the multiple images of the same person in gallery fail into similar geometry. To exploit the intrinsic geometry among the gallery images, we further enforce a local graph regularizer over the reconstruction coefficients:

$$\begin{aligned}\min_{\mathbf{c}_j} \sum_{j=1}^{n} \|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2^2 + \lambda\|\mathbf{c}_j\|_1 + \frac{1}{2}\beta \sum_{i,j\in\{1,\ldots,n\}} \|\mathbf{c}_i - \mathbf{c}_j\|_2^2 \mathbf{S}_{i,j} \\ + \frac{1}{2}\gamma \sum_{p=1}^{G}\sum_{j=1}^{n} \sum_{k,l\in\{1,\ldots,g_p\}} (\mathbf{c}_{j,k}^p - \mathbf{c}_{j,l}^p)^2 \mathbf{B}_{k,l}^p,\end{aligned} \tag{6}$$

where the $\mathbf{c}_j^p = [\mathbf{c}_{j,1}^p, \mathbf{c}_{j,2}^p, \ldots, \mathbf{c}_{j,g_p}^p]^T \in R^{g_p\times 1}$ represents the coefficients to reconstruct $\mathbf{x}_j$ for the p-th person. γ is a parameter to signify the local regularizer. The similarity matrix $\mathbf{B} = diag\{\mathbf{B}^1, \mathbf{B}^2, \ldots, \mathbf{B}^G\} \in R^{M\times M}$, and each element $\mathbf{B}^p \in R^{g_p\times g_p}$ is defined as:

$$\mathbf{B}_{k,l}^p = exp(\frac{-\|\mathbf{d}_k^p - \mathbf{d}_l^p\|_2^2}{2\sigma_2^2}), \tag{7}$$

where σ_2 is a parameter fixed as 0.2 in this paper. $\mathbf{D} = [\mathbf{D}^1, \mathbf{D}^2, \ldots, \mathbf{D}^G] \in R^{d \times M}$ denotes the total M images of G persons in gallery, where $\mathbf{D}^p = [\mathbf{d}_1^p, \mathbf{d}_2^p, \ldots, \mathbf{d}_{g_p}^p] \in R^{d \times g_p}, p = \{1, \ldots, G\}$ represents the matrix of g_p basis feature vectors for the p-th person. The local regularizer in Eq. (6) encourages the higher similarity between the gallery images from the same person, the closer contribution to the reconstruction. With simple algebra, Eq. (6) can be rewritten as:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{DC}\|_F^2 + \lambda \|\mathbf{C}\|_1 + \beta tr(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma tr(\mathbf{C}^T\mathbf{L}_2\mathbf{C}). \tag{8}$$

where $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n] \in R^{M \times n}$, $\mathbf{L}_1 = \mathbf{H} - \mathbf{S}$ is the graph Laplacian matrix, $\mathbf{H} = \text{diag}\{\sum_j \mathbf{S}_{1,j}, \sum_j \mathbf{S}_{2,j}, \cdots\}$ is the degree matrix of $\mathbf{S}$, and $\text{diag}\{\cdots\}$ indicates the diagonal operation, $tr\{\cdots\}$ indicates the trace of a matrix. Analogously, $\mathbf{L}_2 = \mathbf{T} - \mathbf{B}$ is the graph Laplacian matrix, and $\mathbf{T} = \text{diag}\{\sum_j \mathbf{B}_{1,j}, \sum_j \mathbf{B}_{2,j}, \cdots\}$ is the degree matrix of $\mathbf{B}$.

3.2 Non-negative Dual Graph Regularized Sparse Ranking

Thinking that the reconstruction coefficients are meaningless while representing similarity measures between probe and gallery, we further enforce the non-negative constraint on the reconstruction coefficients in the proposed model, and the final formulation is as follows:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{DC}\|_F^2 + \lambda \|\mathbf{C}\|_1 + \beta tr(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma tr(\mathbf{C}^T\mathbf{L}_2\mathbf{C}), \ s.t. \ \mathbf{C} \geq \mathbf{0}. \tag{9}$$

which is named NNDGSR in this paper. The non-negative constraint ensures that the probe image should be represented by the gallery images in a non-subtractive way.

3.3 Model Optimization

Due to the non-negativeness of the elements in $\mathbf{C}$, Eq. (10) can be written as:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{DC}\|_F^2 + \lambda \mathbf{1}^T\mathbf{C}\mathbf{1} + \beta tr(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma tr(\mathbf{C}^T\mathbf{L}_2\mathbf{C}), \ s.t. \ \mathbf{C} \geq \mathbf{0}, \tag{10}$$

where $\mathbf{1}$ denotes the vector that its all elements are 1. To solve Eq. (10), we convert it to an unconstrained form as:

$$\min_{\mathbf{C}} \|\mathbf{X} - \mathbf{DC}\|_F^2 + \lambda \mathbf{1}^T\mathbf{C}\mathbf{1} + \beta tr(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma tr(\mathbf{C}^T\mathbf{L}_2\mathbf{C}) + \psi(\mathbf{C}), \tag{11}$$

where

$$\psi(\mathbf{c}_{j,k}^p) = \begin{cases} 0, & if \ \mathbf{c}_{j,k}^p \geq 0, \\ \infty, & otherwise. \end{cases} \tag{12}$$

In this paper, we utilize the accelerated proximal gradient (APG) [20] approach to optimize efficiently. We denote:

$$\begin{aligned} F(\mathbf{C}) &= \min_{\mathbf{C}} \|\mathbf{X} - \mathbf{DC}\|_F^2 + \lambda \mathbf{1}^T\mathbf{C}\mathbf{1} + \beta tr(\mathbf{C}\mathbf{L}_1\mathbf{C}^T) + \gamma tr(\mathbf{C}^T\mathbf{L}_2\mathbf{C}) \\ Q(\mathbf{C}) &= \psi(\mathbf{C}) \end{aligned} \tag{13}$$

Obviously, $F(\mathbf{C})$ and $Q(\mathbf{C})$ are a differentiable convex function and a non-smooth convex function, respectively. Therefore, according to the APG method, we obtain:

$$\mathbf{C}_{k+1} = \min_{\mathbf{C}} \frac{\xi}{2}\|\mathbf{C} - \mathbf{K}_{k+1} + \frac{\nabla F(\mathbf{K}_{k+1})}{\xi}\|_F^2 + Q(\mathbf{C}), \tag{14}$$

where k indicates the current iteration time, and ξ is the Lipschitz constant. $\mathbf{K}_{k+1} = \mathbf{C}_k + \frac{\rho_{k-1}-1}{\rho_k}(\mathbf{C}_k - \mathbf{C}_{k-1})$, where ρ_k is a positive sequence with $\rho_0 = \rho_1 = 1$. Equation (14) can be solved by:

$$\mathbf{C}_{k+1} = \max(0, \mathbf{K}_{k+1} - \frac{\nabla F(\mathbf{K}_{k+1})}{\xi}). \tag{15}$$

Algorithm 1 summarizes the whole optimization procedure.

Algorithm 1. Optimization Procedure to Eq. (13)

Input: query feature matrix $\mathbf{X}$, dictionary/gallery feature matrix $\mathbf{D}$, Laplacian matrix $\mathbf{L}_1$ and $\mathbf{L}_2$, parameters λ, β and γ;
Set $\mathbf{C}_0 = \mathbf{C}_1 = \mathbf{0}, \xi = 1.8 \times 10^3, \varepsilon = 10^{-4}, \rho_0 = \rho_1 = 1, maxIter = 150, k = 1$
Output: $\mathbf{C}$
1: While not converged **do**
2: Update $\mathbf{K}_{k+1}$ by $\mathbf{K}_{k+1} = \mathbf{C}_k + \frac{\rho_{k-1}-1}{\rho_k}(\mathbf{C}_k - \mathbf{C}_{k-1})$;
3: Update $\mathbf{C}_{k+1}$ by Eq. (15);
4: Update $\rho_{k+1} = \frac{1+\sqrt{1+4\rho_k^2}}{2}$;
5: Update k by $k = k + 1$;
6: The convergence condition: maximum number of iterations reaches $maxIter$ or the maximum element change of $\mathbf{C}$ between two consecutive iterations is less than ε.
7: **end While**

4 Ranking Implementation for Multi-shot Re-ID

Due to the sparsity of the reconstruction coefficients, the majority of which collapse to zero after few higher coefficients. Therefore, we can not support ranking for all the individuals in gallery. To cope this issue, we develop an error distribution technique. First, we can obtain the normalized reconstruction error for current probe $\mathbf{x}_j$ according to coefficients as:

$$e_j = \frac{\|\mathbf{x}_j - \mathbf{D}\mathbf{c}_j\|_2}{\|\mathbf{x}_j\|_2}. \tag{16}$$

Then, we re-distribute the reconstruction errors into the gallery individuals according to their similarity to the current probe image $\mathbf{x}_j$ as:

$$\mathbf{W}_{j,k}^p = \frac{1/dis(\mathbf{x}_j, \mathbf{d}_k^p)}{\sum_{p=1}^{G}\sum_{k=1}^{g_p}(1/dis(\mathbf{x}_j, \mathbf{d}_k^p))}, \ k = \{1, \ldots, g_p\}, \tag{17}$$

where $\mathbf{d}_k^p$ represents the feature of the k-th image from the p-th person in gallery/dictionary $\mathbf{D}$, $dis(\mathbf{x}_j, \mathbf{d}_k^p)$ denotes the Euclidean distance between probe $\mathbf{x}_j$ and each element $\mathbf{d}_k^p$ in gallery. $\mathbf{W}_{j,k}^p$ indicates the similarity/weight of $\mathbf{d}_k^p$ relative to $\mathbf{x}_j$.

In this paper, we employ the reconstruction coefficients as the similarity measures, and define the accumulated all reconstruction coefficients from the p-th person as a part of the ranking value of the probe person with n images against the p-th person. Moreover, we use the reconstruction residues to make the p-th category whose reconstruction coefficients are all zeros has ranking value. Therefore, the final ranking value of the probe person with n images against the p-th person in gallery is defined as follows:

$$\mathbf{r}^p = \sum_{j=1}^{n} \sum_{k=1}^{g_p} \mathbf{c}_{j,k}^p + \mathbf{W}_{j,k}^p * e_j, \ p = \{1, ..., G\}. \tag{18}$$

The higher similarity of $\mathbf{d}_k^p$ relative to $\mathbf{x}_j$, the higher value distributed to $\mathbf{c}_{j,k}^p$. Since e_j is usually small, the value distributed to $\mathbf{c}_{j,k}^p$ is also very small, which will not change the ranks of the non-zero coefficients but will reorder the zero coefficients according to Euclidean distance.

Our final decision rule is :

$$class(\mathbf{X}) = \arg\max_p \mathbf{r}^p. \tag{19}$$

5 Experimental Results

We evaluate our method on three benchmark datasets including i-LIDS [30], CAVIAR4REID [6] and MARS [28] comparing to the state-of-the-art algorithms for multi-shot Re-ID. We use the standard measurement named Cumulated Match Characteristic (CMC) curve to figure out the matching results, where the matching rate at rank-n indicates the percentage of correct matchings in top n candidates according to the learnt ranking function Eq. (18).

5.1 Datasets and Settings

i-LIDS [30] is composed by 479 images of 119 people, which was captured at an airport arrival hall under two non-overlapping camera views with almost two images each person per camera views. This dataset consists challenging scenarios with heavy occlusions and pose variance.

CAVIAR4REID [6] contains 72 unique individuals with averagely 11.2 images per person extracted from two non-overlapping cameras in a shopping center: 50 of which with both the camera views and the remaining 22 with only one camera view. The images for each camera view have variations with respect to resolution changes, light conditions, occlusions and pose changes.

MARS [28] is the largest and newly collected dataset for video based Re-ID. It is collected from six near-synchronized cameras in the campus of Tsinghua

University. MARS consists of 1261 pedestrians each of which appears at least two cameras. It contains 625 identities with 8298 tracklets for training and 636 identities with 12180 tracklets for testing. Different from the other datasets, it also consists of 23380 junk bounding boxes and 147743 distractors bounding boxes in the testing samples.

Parameters. There are three important parameters in our model: λ controls the tradeoff between minimization of the ℓ_2 reconstruction error and the ℓ_1 sparsity of the coefficients. β controls the global regularizer in queries. γ controls the local regularizer in gallery. We empirically set: $\{\lambda, \beta, \gamma\} = \{0.2, 0.5, 0.5\}$.

5.2 Evaluation on Benchmarks

The performance of the proposed approach on the three benchmark datasets comparing with the state-of-the-art algorithms is reported in this section. We evaluate the proposed NNDGSR on both hand-crafted features and deep features. Followed by the protocol in [18], we use WHOS feature [18] as hand-crafted features. As for deep features, we generate APR [17] features based on ResNet-50, which is pre-trained on large Re-ID dataset Market-1501 [29] for i-LIDS [30] and CAVIAR4REID [6], while utilize IDE feature [28] for MARS [28] as provided.

Comparison on i-LIDS. Evaluation results on i-LIDS dataset are shown in Table 1 and Fig. 1 (a). From which we can see, Our approach significantly outperforms the state-of-the-arts. The rank-1 accuracies of our approach achieve 84.3% and 78.4% on hand-crafted and deep features respectively, which improve 21.4% and 1.2% than the second best method ISR [18]. It is worth noting that: (1) The limited number of samples in i-LIDS compromises the performance of deep learning. (2) Our NNDGSR significantly improves the ranking results on both hand-crafted and deep features.

Comparison on CAVIAR4REID. Evaluation results on CAVIAR4REID [6] are shown in Table 1 and Fig. 1 (b). We evaluate our method with APR [17] deep features in the same manner as on i-LIDS and adopt the same experimental protocols as ISR [18] by 50 random trials. Clearly, our approach significantly outperforms the state-of-the-art algorithms on both hand-crafted and deep features. Specifically, the Rank-1 accuracies with $N = 5$ achieve 93.2% and 89.0% on hand-crafted features and deep features respectively. Together with the results on i-LIDS, it suggests that the proposed method achieves impressive performance on small size datasets.

Comparison on MARS. In this dataset, the query trackelets are automatically generated from the testing samples. For each query trackelet, we construct two feature vectors via max pooling and average pooling respectively on the provide deep features, IDE [28]. For the remaining testing trackelets, since there are multiple trackelets for each person under a certain camera, we conduct the max pooling for each trackelet to construct the multiple feature vectors followed by the state-of-the-art methods on MARS [28]. Note that, our method

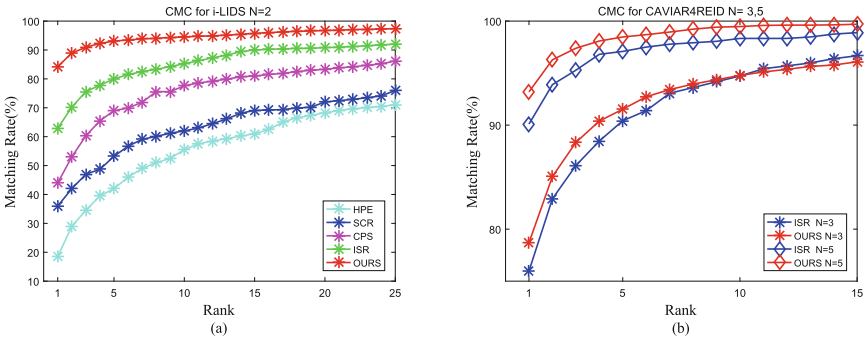

Fig. 1. The cumulative match characteristic curves on i-LIDS and CAVIRA4REID with hand-crafted feature comparing with the state-of-the-arts.

Table 1. Comparison results at Rank-1 on i-LIDS and CAVIAR4REID (in %)

Features	Methods:	i-LIDS	CAVIAR4REID		References:
		N = 2	N = 3	N = 5	
Hand-craft features	HPE [2]	18.5	-	-	ICPR2010
	AHPE [3]	32	7.5	7.5	PRL2012
	SCR [4]	36	-	-	ICAVSS2010
	MRCG [1]	46	-	-	ICAVSS2011
	SDALF [8]	39	8.5	8.3	CVPR2010
	CPS [6]	44	13	17.5	BMVC2011
	COSMATI [5]	44	-	-	ECCV2012
	WHOS + ISR [18]	62.9	75.1	90.1	PAMI2015
	WHOS [18] **+NNDGSR**	**84.3**	**78.7**	**93.2**	-
Deep features	APR [17] + EU [17]	67.7	44.3	53.8	Arxiv2017
	APR [17] + ISR [18]	77.2	65.7	80.7	Arxiv2017+PAMI2015
	APR [17] + **NNDGSR**	**78.4**	**70.4**	**89.0**	-

doesn't require any training therefore only the testing set containing with the query set is utilized. The performance of our method against different metrics is reported in Table 2. As we can see: (1) CNN based methods generally outperforms the traditional metric learning methods on hand-crafted features. (2) The sparse ranking based method outperforms on the powerful deep feature comparing with the traditional Euclidian distance. (3) By introducing the non-negative dual graph regularized into the sparse ranking framework, our method can significantly boost the performance by increasing 9.5% at Rank-1 accuracy.

Table 2. Comparison with baseline on MARS dataset (in %)

Features	Methods:	Rank-1	Rank-5	Rank-20	References:
Hand-craft Features	HOG3D [13]+ KISSME [21]	2.6	6.4	12.4	BMVC2010+CVPR2012
	GEI[10]+ KISSME[21]	1.2	2.8	7.4	PAMI2005+CVPR2012
	HistLBP[23]+ XQDA[16]	18.6	33.0	45.9	ECCV2014+CVPR2015
	BoW[29]+ kissme[21]	30.6	46.2	59.2	ICCV2015+CVPR2012
	LOMO+ XQDA[16]	30.7	46.6	60.9	CVPR2015
Deep features	ASTPN [24]	44	70	81	ICCV2017
	LCAR [27]	55.5	70.2	80.2	Arxiv2017
	SATPP [15]	69.7	84.7	92.8	Arxiv2017
	SFT [34]	70.6	90	97.6	CVPR2017
	MSCAN [14]	71.8	86.6	93.1	CVPR2017
	IDE+EU [28]	58.7	77.1	86.8	ECCV2016
	IDE [28] + ISR [18]	63	77.1	85.6	ECCV2016+PAMI2015
	IDE[28] + NNDGSR	**72.50**	**88.0**	**93.30**	-

Table 3. Evaluation on individual component on CAVIAR4REID dataset with N =5 on APR deep features (in %)

Components:	Rank-1	Rank-5	Rank-10	Rank-20
ISR	80.7	95.8	97.9	**99.4**
SR+NN	84.3	94.9	97.3	98.6
SR+NN+GG	87.8	96.7	**98.2**	**99.4**
SR+NN+GG+LG (NNDGSR)	**89.0**	**96.8**	**98.2**	99.3

5.3 Component Analysis

To verify the contribution of the proposed non-negative dual graph regularized sparse ranking for multi-shot Re-ID, we further evaluate the components of our method on CAVIAR4REID [6] with APR features [17] and report the results in Table 3, where: SR indicates the original sparse ranking without any non-negative or regularization as Eq. (3). NN, GG and LG denotes introducing the Non-negative constraint, global graph regularizer and local graph regularizer respectively. From which we can see: (1) Both non-negative constraint and the dual graph regularizers play important roles. (2) By enforcing the non-negative constraint on the coefficients, it can improve 3.6% at rank-1 accuracy. (3) Global and local graph regularizers can further improve the performance by 3.5% and 1.2% respectively in Rank-1, which demonstrates the contribution of the components of non-negative dual graph regularized sparse ranking.

6 Conclusion

In this paper, we have proposed a novel sparse ranking based multi-shot person Re-ID approach. In order to simultaneously capture the intrinsic geometric structures in both probe and gallery, we design a non-negative dual graph regularized sparse ranking method for multi-shot Re-ID. Then we provide a fast optimization for the proposed unified sparse ranking framework. Experiments on three challenging multi-shot person Re-ID datasets demonstrate the promising performance of the proposed method especially on small size datasets where the performance of deep learning is compromised. In the future, we will investigate the effective way of fusing key-feature information from video-based Re-ID.

Acknowledgement. This study was funded by the National Nature Science Foundation of China (61502006, 61602001, 61702002, 61872005, 61860206004) and the Natural Science Foundation of Anhui Higher Education Institutions of China (KJ2017A017).

References

1. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Multiple-shot human re-identification by mean riemannian covariance grid. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance, pp. 179–184 (2011)
2. Bazzani, L., Cristani, M., Perina, A., Farenzena, M., Murino, V.: Multiple-shot person re-identification by HPE signature. In: International Conference on Pattern Recognition, pp. 1413–1416 (2010)
3. Bazzani, L., Cristani, M., Perina, A., Murino, V.: Multiple-shot person re-identification by chromatic and epitomic analyses. Pattern Recogn. Lett. **33**(7), 898–903 (2012)
4. Bk, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 435–440 (2010)
5. Charpiat, G., Thonnat, M.: Learning to match appearances by correlations in a covariance metric space. In: European Conference on Computer Vision, pp. 806–820 (2012)
6. Dong, S.C., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: British Machine Vision Conference, pp. 68.1–68.11 (2011)
7. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. **32**(2), 407–451 (2004)
8. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Computer Vision and Pattern Recognition, pp. 2360–2367 (2010)
9. Feng, X., Wu, S., Tang, Z., Li, Z.: Sparse latent model with dual graph regularization for collaborative filtering. Neurocomputing (2018)
10. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2005)
11. Jing, X.Y., et al.: Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. IEEE Trans. Image Process. **26**(3), 1363–1378 (2017)

12. Karanam, S., Li, Y., Radke, R.J.: Person re-identification with discriminatively trained viewpoint invariant dictionaries. In: IEEE International Conference on Computer Vision, pp. 4516–4524 (2015)
13. Klaser, A.: A spatiotemporal descriptor based on 3D-gradients. In: British Machine Vision Conference, September 2010
14. Li, D., Chen, X., Zhang, Z., Huang, K.: Learning deep context-aware features over body and latent parts for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
15. Li, J., Zhang, S., Wang, J., Gao, W., Tian, Q.: LVreID: person re-identification with long sequence videos. arXiv preprint arXiv:1712.07286 (2017)
16. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
17. Lin, Y., Zheng, L., Zheng, Z., Wu, Y., Yang, Y.: Improving person re-identification by attribute and identity learning. arXiv preprint arXiv:1703.07220 (2017)
18. Lisanti, G., Masi, I., Bagdanov, A.D., Bimbo, A.D.: Person re-identification by iterative re-weighted sparse ranking. IEEE Trans. Pattern Anal. Mach. Intell. **37**(8), 1629–1642 (2015)
19. Liu, X., Song, M., Tao, D., Zhou, X., Chen, C., Bu, J.: Semi-supervised coupled dictionary learning for person re-identification. In: Computer Vision and Pattern Recognition, pp. 3550–3557 (2014)
20. Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**(3), 127–239 (2014)
21. Roth, P.M., Wohlhart, P., Hirzer, M., Kostinger, M., Bischof, H.: Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2288–2295 (2012)
22. Wright, J., Ganesh, A., Zhou, Z., Wagner, A., Ma, Y.: Demo: robust face recognition via sparse representation. In: IEEE International Conference on Automatic Face Gesture Recognition, pp. 1–2 (2009)
23. Xiong, F., Gou, M., Camps, O., Sznaier, M.: Person re-identification using kernel-based metric learning methods. In: European Conference on Computer Vision, pp. 1–16 (2014)
24. Xu, S., Cheng, Y., Gu, K., Yang, Y., Chang, S., Zhou, P.: Jointly attentive spatial-temporal pooling networks for video-based person re-identification. arXiv preprint arXiv:1708.02286 (2017)
25. Yankelevsky, Y., Elad, M.: Dual graph regularized dictionary learning. IEEE Trans. Signal Inf. Process. Netw. **2**(4), 611–624 (2017)
26. Yin, M., Gao, J., Lin, Z., Shi, Q., Guo, Y.: Dual graph regularized latent low-rank representation for subspace clustering. IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. **24**(12), 4918–4933 (2015)
27. Zhang, W., Hu, S., Liu, K.: Learning compact appearance representation for video-based person re-identification. arXiv preprint arXiv:1702.06294 (2017)
28. Zheng, L., et al.: MARS: a video benchmark for large-scale person re-identification. In: European Conference on Computer Vision, pp. 868–884 (2016)
29. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
30. Zheng, W.S., Gong, S., Xiang, T.: Associating groups of people. Active Range Imaging Dataset for Indoor Surveillance (2009)
31. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by Relative Distance Comparison. IEEE Computer Society (2013)

32. Zheng, W., Gong, S., Xiang, T.: Towards open-world person re-identification by one-shot group-based verification. IEEE Trans. Pattern Anal. Mach. Intell. **38**(3), 591–606 (2016)
33. Zheng, Y.W., Hao, S., Zhang, B.C., Zhang, J., Zhang, X.: Weight-based sparse coding for multi-shot person re-identification. Sci. China Inf. Sci. **58**(10), 100104–100104 (2015)
34. Zhou, Z., Huang, Y., Wang, W., Wang, L., Tan, T.: See the forest for the trees: joint spatial and temporal recurrent neural networks for video-based person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 6776–6785 (2017)