2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Region-specific Metric Learning for Person Re-identification

Min Cao[13], Chen Chen[13], Xiyuan Hu[13], Silong Peng[123]
1 Institute of Automation Chinese Academy of Sciences Beijing, China
2 Beijing Visytem Co. Ltd
3 University of Chinese Academy of Sciences, Beijing, China
{caomin2014,chen.chen,xiyuan.hu,silong.peng}@ia.ac.cn

***Abstract*—Person re-identification addresses the problem of matching individual images of the same person captured by different non-overlapping camera views. Distance metric learning plays an effective role in addressing the problem. With the features extracted on several regions of person image, most of distance metric learning methods have been developed in which the learnt cross-view transformations are region-generic, i.e all region-features share a homogeneous transformation. The spatial structure of person image is ignored and the distribution difference among different region-features is neglected. Therefore in this paper, we propose a novel region-specific metric learning method in which a series of region-specific sub-models are optimized for learning cross-view region-specific transformations. Additionally, we also present a novel feature pre-processing scheme that is designed to improve the features' discriminative power by removing weakly discriminative features. Experimental results on the publicly available VIPeR, PRID450S and QMUL GRID datasets demonstrate that the proposed method performs favorably against the state-of-the-art methods.**

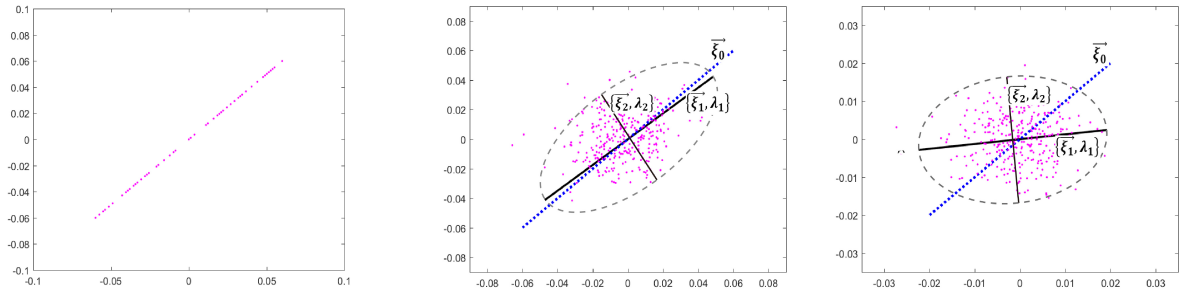
I. Introduction

Inter-camera person association, known as person re-identification (re-id), enables tracking of the same person across non-overlapping camera views. It plays an important role in many practical applications, such as criminals tracking and crowd movements analysis. It also a rather challenging task due to visual ambiguities caused by illumination changes, viewpoint and pose variations, and so forth. To address these challenges, most of existing methods focus on feature extraction [10][11] and distance metric learning [6][17], which are the critical steps in person re-id system.

The input data of person re-id system is usually some cropped images of pedestrians. The feature is firstly extracted from the raw images based on the aim of the robustness of feature against the variations in persons' appearance across camera views [5][21]. With the extracted features used as the input of distance metric learning, there are two problems: 1) the extracted feature dimension is usually large, and 2) the features are flawed for low-quality person images and imperfect feature extraction. Therefore, the feature pre-processing is necessary for efficient computation and improving accuracy. Principal Component Analysis (PCA) is often performed for dimension reduction. However, it is a unsupervised dimensionality reduction method and the features' discriminative power has not been improved (related to the problem in 2)). In this paper, we propose a novel feature pre-processing scheme focusing on improving the features' discriminative power. Clearly, an ideally discriminative feature should be consistent for intra-class and different for inter-class, as shown in Fig 1(a). In fact, the features have different discriminative powers for each dimension compared with the ideally discriminative feature, as shown in Fig 1(b). We argue that removing features of some dimensions with poorly discriminative power is beneficial for improving the features' discriminative power. Specifically, after PCA-based dimensional-reduction, features with poorly discriminative power are removed if its distribution is quite different from the distribution of the ideally discriminative feature. As a result, both problems in 1) and 2) can be alleviated in the proposed feature pre-processing scheme.

After feature extraction, a suitable distance metric is necessary for matching inter-camera people. Instead of matching people in the original feature space with unsupervised-based distance metric (such as Euclidean distance), many existing methods [18][19] aim to learn distance metrics to maximize the inter-class variations and minimize the intra-class variations, in which Mahalanobis distance receives the most attention. Mahalanobis distance between features $\mathbf{z}_i$ and $\mathbf{z}_j$ is equivalent to the Euclidean distance on the space projected by a projection matrix L, namely, $d(\mathbf{z}_i, \mathbf{z}_j) = (\mathbf{z}_i - \mathbf{z}_j)^T M(\mathbf{z}_i - \mathbf{z}_j) = \left\|\mathbf{z}'_i - \mathbf{z}'_j\right\|_2^2$ where $M = L^T L$, $\mathbf{z}'_i = L\mathbf{z}_i$ and $\mathbf{z}'_j = L\mathbf{z}_j$. With the input features extracted on several regions of person image, most of methods learn the projection matrix L based on the equal treatment of features from each region of person image. As the input data of metric learning-based method, the robust features are extracted with consideration for the spatial structure of person image. However, it cannot be taken into account in above metric learning-based methods. Moreover, the distribution difference among different region-features is also neglected. Accordingly, we propose a novel effective model named *Region-Specific metric Learning* (RSL) in which each of image regions are treated independently. Specifically, for the feature $\mathbf{z} = [\hat{\mathbf{z}}_1 \dots \hat{\mathbf{z}}_S]$, where $\hat{\mathbf{z}}_s$ $(s = 1, \dots, S)$ is the features after pre-processing on s-th image region, we learn a projection matrix $L = diag(L_{11}, L_{22}, \dots, L_{SS})$ with block-diagonal structure, where L_{ss} is a sub-projection matrix over s-th image region. Fig 2 illustrates the process of feature mapping in an intuitive way.

To summarize, the contributions of the paper are that **1)** we



(a) Ideally discriminative feature

(b) The extracted features

Fig. 1. The distributions of (a) the ideally discriminative feature and (b) GOG descriptor [11] on VIPeR dataset [8] in one dimension. In (b), the distribution at left is about the feature with good discriminative power, the one at right is about the feature with poorly discriminative power. The x-axis and y-axis of purple dots represent the features of the same people from two different cameras, respectively. The explanation about the notations in (b) is given in section III. Better viewed in colour.

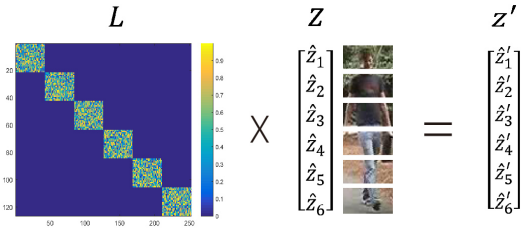


Fig. 2. The illustration of the proposed region-specific metric learning model. We assume that the person image is divided into 6 horizontal strips for convenience. Better viewed in colour.

show a novel feature pre-processing scheme that not only improves the features' discriminative power and also reduces the feature dimension; **2)** consider the spatial structure of person image and the distribution difference among different region-features, we propose a novel region-specific metric learning model. For convenience, we name the proposed method as RSL+Pre below.

II. Related Work

Feature extraction and distance metric learning are two critical steps in person re-id system. In this section, we will review the related works.

A. Feature Extraction-based Method

Low-level features such as color, texture and gradient are most commonly used for person representation. However, these features are not powerful enough to discriminate the same person from different camera views. This category of person re-id mainly focuses on extracting visual features by combining these low-level features together to build both distinctive and stable features under changing conditions between different cameras, such as symmetry-driven accumulation of local features (SDALF) [7], local maximal occurrence representation (LOMO) [10], hierarchal Gaussian descriptor (GOG) [11] and histogram of intensity and ordinal pattern (HIPHOP) [5], and so forth. These methods have been mainly conducted on unsupervised settings of which performance is inferior to the supervised methods, so the extracted features are usually provided as input to the distance metric learning-based methods. Certainly, the dimension of the extracted features is usually large and PCA is often applied for dimension reduction before metric learning step. Moreover, the features have different impacts on the performance on the training set and feature selection is usually performed to improve the features' discriminative power, resulting in greater accuracy of person re-id. Compared with these feature selection methods [16][24] in which the weights on different features are learned, we select features by removing weakly discriminative features based on comparison of the distribution between the ideally discriminative feature and the features in each dimension.

B. Metric Learning-based Method

In metric learning methods, they usually use training data to learn effective distance functions that can minimize intra-class variation whilst maximizing intra-class variation. Jurie et al. [9] built a low-dimensional space obtained by learning a transformation matrix L from sparse pairwise similarity/dissimilarity constraints. Liao et al. [10] considered the dimension reduction into the distance metric learning. With the features extracted on several regions of person images, most of metric learning-based methods learn the distance metric based on all region-features together. Considering the spatial structure of person image and the distribution difference among different region-features, some methods [4][6] proposed to learn multiple distance metrics over each image region. Chen et al. [4] argued that each region in person image has its own similarity measurement that excels at handling the special intra-person variation within it, and a similarity function consisting of multiple sub-similarity measurements is proposed and optimized by ADMM algorithm. Chong et al. [6] proposed a latent metric learning method for modeling vertical misalignments, horizontal misalignments and leg postures, and a pedestrian was represented as the mixture of a holistic model and a number of flexible models. They are the heavy and complicated works. Instead, we propose a novel simple

and effective region-specific metric learning method that the projection matrix L is learned with a block-diagonal structure.

Furthermore, a trend of re-ranking for person re-id has emerged recently. An initial ranking list is obtained by feature extraction and distance metric learning steps, which are only based on information from the person image. And then, the ranking list can be optimized through the context information or some customized rules about the ranking list. For example, Cao et al. [1] improved the performance of person re-id by combing the individual information with the group information. Chen et al. [3] proposed a key person aided person re-id framework, where outstanding contextual pedestrians are used to revise the ranking list. Zhong et al. [22] argued that a gallery image is similar to the probe in the k-reciprocal nearest neighbors and a k-reciprocal feature is used for re-ranking. These re-ranking methods can be used in the proposed RSL+Pre method to achieve performance improvement.

III. Methodology

A. Feature Pre-processing

Given a probe set of N samples denoted as $X^p \in \mathbb{R}^{d_x \times N}$. Each column $\mathbf{x}_i^p$ of the data descriptor matrix X^p is a feature vector representing the i-th training sample in probe set. Similarly, a gallery set is denoted as $X^g \in \mathbb{R}^{d_x \times N}$ and $(\mathbf{x}_i^p, \mathbf{x}_i^g)$ is a correct matching pair. The feature dimension d_x is usually large and we perform the feature dimension reduction via the PCA method. $\mathbf{y}$ denotes the new feature from d_y dimension space ($d_y < d_x$), in which each dimension has different discriminative powers for person re-id. With the aim of improving features' discriminative power, we select d_z dimension features with most discriminative information from d_y dimension features ($d_z < d_y$).

We call $\mathbf{y}$ as the ideally discriminative features if satisfy $\mathbf{y}_i^p = \mathbf{y}_i^g \ (i = 1, \ldots, N)$ and $\mathbf{y}_i^p \neq \mathbf{y}_j^g \ (i \neq j)$. The distribution of the scatter plot of $\{(\mathbf{y}_i^p(k), \mathbf{y}_i^g(k))\}_{i=1}^N$ for any k-th dimension feature is in a straight line with 45° slope (Fig 1(a)). However in the real world, due to low-quality person images and imperfect feature extraction and so on, $Y^p = Y^g$ rarely holds. The scatter plot of $\{(\mathbf{y}_i^p(k), \mathbf{y}_i^g(k))\}_{i=1}^N$ is usually an ellipse cluster (Fig 1(b)). The features whose scatter plot for each dimension is a flattened ellipse cluster and the major axis of ellipse cluster is close to the straight line with 45° slope, are similar to the ideally discriminative features, and vice versa. We select features based on the observation.

For k-th dimension feature, we apply PCA to $P = \{(\mathbf{y}_i^p(k), \mathbf{y}_i^g(k))\}_{i=1}^N$ and obtain the eigenvectors $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ and corresponding eigenvalues λ_1, λ_2 ($\lambda_1 > \lambda_2$). The eigenvector and the eigenvalue reflect the axis of ellipse and its length, respectively (Fig 1(b)). The flat grade of ellipse is measured by *flattening* [15] defined as

$$f_k(\lambda_1, \lambda_2) = \frac{\lambda_1 - \lambda_2}{\lambda_1}, \tag{1}$$

where $f_k \in [0, 1]$ and $f = 0$ for a circle. In Fig 1(b), $f_k = 0.37$ for the left and $f_k = 0.15$ for the right. The angle between the major axis of ellipse and the straight line with 45° slope denoted as $\boldsymbol{\xi}_0$ is measured by

$$g_k(\boldsymbol{\xi}_0, \boldsymbol{\xi}_1) = \arccos \frac{\langle \boldsymbol{\xi}_0, \boldsymbol{\xi}_1 \rangle}{\|\boldsymbol{\xi}_0\| \, \|\boldsymbol{\xi}_1\|}, \tag{2}$$

where g_k is represented by the radian. In Fig 1(b), $g_k = 0.07$ for the left and $g_k = 0.66$ for the right.

Formally, we define the *discrimination* measurement of k-th dimension feature as

$$\rho_k = \frac{f_k(\lambda_1, \lambda_2)}{g_k(\boldsymbol{\xi}_0, \boldsymbol{\xi}_1) + \varepsilon}, \tag{3}$$

where ε is a small value for avoiding a zero denominator.

If ρ_k is more than the threshold τ, we select k-th dimension feature. The new discriminative feature is denoted by $\mathbf{z}$. The threshold τ will be discussed in the experiment section.

B. Region-Specific Metric Learning

Let $D(\mathbf{z}_i^p, \mathbf{z}_j^g) = \left\| L(\mathbf{z}_i^p - \mathbf{z}_j^g) \right\|_2^2$ be the distance between feature vectors $\mathbf{z}_i^p$ and $\mathbf{z}_j^g$. We aim to lean an optimal projection matrix L that satisfies the following constraints:

$$D(\mathbf{z}_i^p, \mathbf{z}_j^g) \begin{cases} < C & i = j \\ > C & i \neq j \end{cases} \quad i, j = 1, \ldots, N. \tag{4}$$

where C is the parameter of hyperplane and is set to 1 in the experiments.

Specifically, L is learned by minimizing the following function:

$$\min_L \sum_{i=1,j=1}^{N} l_\beta(k_{ij}(D(\mathbf{z}_i^p, \mathbf{z}_j^g) - 1)) + \alpha \|L\|_F^2, \tag{5}$$

where $k_{ij} = 1$ and $k_{ij} = -1$ indicate that $(\mathbf{z}_i^p, \mathbf{z}_i^g)$ is a correct matching pair and incorrect matching pair, respectively. α is the regularization parameter. $l_\beta(x) = \frac{1}{\beta} \log(1 + e^{\beta x})$ is the smooth approximation of the hinge loss function $h(x) = \max\{x, 0\}$ [1]. α and β are set to 0.01 and 3 in the experiments.

For handling spatial misalignment and better describing of spatial details in the person image, the features are usually extracted on several regions of person image, i.e. $\mathbf{z} = [\hat{\mathbf{z}}_1 \ldots \hat{\mathbf{z}}_S]$, where $\hat{\mathbf{z}}_s \ (i = s, \ldots, S)$ is the extracted features on s-th region. We re-write $L = \begin{bmatrix} L_{11} & \ldots & L_{1S} \\ \vdots & \ddots & \vdots \\ L_{S1} & \ldots & L_{SS} \end{bmatrix}$ represented by a block matrix, and $D(\mathbf{z}_i^p, \mathbf{z}_j^g) = \left\| L(\mathbf{z}_i^p - \mathbf{z}_j^g) \right\|_2^2 = \left\| \mathbf{z'}_i^p - \mathbf{z'}_j^g \right\|_2^2$. The new discriminative features $\mathbf{z'} = \left[\hat{\mathbf{z}}'_1 \ldots \hat{\mathbf{z}}'_S\right]$, in which $\hat{\mathbf{z}}'_i = \sum_{s=1}^{S} L_{is} \hat{\mathbf{z}}_i$ is obtained by a series of linear combinations of all region-features $\{\hat{\mathbf{z}}_i\}_{i=1}^S$. This means that the information of spatial structure in the new features $\mathbf{z'}$ is lost. Moreover, the model in Eq.5 implicitly assumes that all region-features share a homogeneous projection matrix L and the fact that different region-features are varied in distribution is ignored. Therefore, we propose a region-specific metric learning

[1] $\lim_{\beta \to \infty} l_\beta(x) = h(x)$

Algorithm 1 RSL+Pre based person re-id

Input:

The training set $\{(\mathbf{x}_i^p, \mathbf{x}_j^g), y_{ij}\}_{i,j=1}^N$.

The testing set $\{(\mathbf{u}_i^p, \mathbf{u}_j^g)\}_{i,j=1}^{N_t}$.

Output:

$D(\mathbf{u}_i^p, \mathbf{u}_j^g)$.

Initialize $s = 1$.
while $s \leq S$ **do**
  (I) Feature pre-processing (Section III-A)
  Perform PCA-based dimension reduction for s-th region-feature.
  Initialize $k = 1$.
  while $k \leq d_y^s$ **do**
    Apply PCA to $P = \{(\hat{\mathbf{y}}_s)_i^p(k), (\hat{\mathbf{y}}_s)_j^g(k)\}_{i,j=1}^N$
    Compute ρ_k by Eq.3.
    If $\rho_k < \tau$, k-th dimension feature is removed.
    $k = k + 1$.
  end
  Output the training set $\{((\hat{\mathbf{z}}_s)_i^p, (\hat{\mathbf{z}}_s)_j^g), y_{ij}\}_{i,j=1}^N$ and the testing set $\{(\hat{\mathbf{v}}_s)_i^p, (\hat{\mathbf{v}}_s)_j^g\}_{i,j=1}^N$.
  (II) Region-specific metric learning (Section III-B)
  Compute L_{ss} by Eq.5 with $\{(\hat{\mathbf{z}}_s)_i^p, (\hat{\mathbf{z}}_s)_j^g\}_{i,j=1}^N$.
  $s = s + 1$.
end
Compute $L = diag(L_{11}, L_{22}, \ldots, L_{SS})$.
Compute $D(\mathbf{u}_i^p, \mathbf{u}_j^g) = \left\| L(\mathbf{v}_i^p - \mathbf{v}_j^g) \right\|_2^2$.
return $D(\mathbf{u}_i^p, \mathbf{u}_j^g)$;

model in which the learnt projection matrix L is a block-diagonal structure matrix, i.e. $L = diag(L_{11}, L_{22}, \ldots, L_{SS})$, where L_{ss} is a sub-projection matrix over s-th image region. As a result, we introduce a new constraint with a block-diagonal structure L into Eq.5:

$$\begin{aligned} &\min_{L} \sum_{i=1,j=1}^{N} l_\beta(k_{ij}(D(\mathbf{z}_i^p, \mathbf{z}_j^g) - 1)) + \alpha \left\| L \right\|_F^2 \\ &s.t \quad L_{ij} = \mathbf{O},\ i \neq j,\ i, j = 1, \ldots, S \end{aligned} \tag{6}$$

By doing so, the sub-projection matrixes between different regions are independent of each other. The new feature $\mathbf{z}'$ is also region-connected and holds the structure of the original feature. The parameters learned are reduced so that the over-fitting problem to some extend can be alleviated.

In the experiments, as the constraint in Eq.6 is nonconvex, the block-diagonal projection matrix L is obtained through minimizing Eq.5 applied on each image region at the same time. In addition, because of the non-linearity in person's appearance, we use the "kernel trick" in Eq.5. Specifically, we use the Gaussian kernel function $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$.

C. Person Re-Identification via Region-Specific Learning

In summary, the proposed method is a region-specific metric learning method, and the input features are pre-processed by removing weakly discriminative features. Algorithm 1 formalizes the proposed RSL+Pre method. A specific sub-projection matrix is learned over each image region by the model in Eq.5, which is the base model of our method. In fact, most of region-generic metric learning methods can be used as the base model of our method, resulting in better performance.

IV. Experiments

A. Datasets and Settings

Datasets. We evaluate the proposed RSL+Pre method on three publicly available datasets: **VIPeR** [8], **PRID450S** [13] and **QMUL GRID** [2]. The VIPeR dataset includes 632 identities and the PRID450S dataset contains 450 identities, both collected from two outdoor cameras with different illumination conditions and pose variation. Each identity has one image per camera. The GRID dataset is captured in 8 disjoint cameras. It consists of 250 identities that have two images from different camera views and 775 identities that are only from one cameras to enlarge the gallery.

Feature. We extract the feature similar to those employed in [9]. Person image is divided into 14 non-overlapping regions. For each region, color histograms in RGB, HSV and YCrCb color spaces and texture histograms based on Local Binary Patterns (LBP) are computed and are normalized to unit L_1 norm. By concatenating all the histograms to a single vector, the final descriptor has 6020 dimensions and is used as the input of the proposed RSL+Pre. Besides, we also use the Gaussian Of Gaussian (GOG) descriptor [11] for person representation. GOG descriptor modeled both the mean and the covariance information of pixel features, described by color, texture and pixel location. The descriptor is proposed based on 7 horizontal strips of the images and the final descriptor has 27622 dimensions.

Settings. In the experiments, all the images are resized to 128×48 pixels for normalization. Following the commonly used evaluation protocol in [10], we use the Cumulative Matching Characteristics (CMC) curve to evaluate the performance of the proposed method. The whole procedure is repeated 10 times and rank1, rank10, rank20 of the average CMC curve are reported. The threshold parameter τ is set to 0.1 for color+texture histogram descriptor and GOG descriptor by cross-validation and will be discussed in section IV-C.

B. Performance Comparison

We compare the performance of the proposed RSL+Pre method with the state-of-the-art person re-id methods on all the three datasets. Among them, CSL [14], LOMO+XQDA [10], Kernel HPCA [12], LSSCDL [19], SCSP [4], OL-MANS [23] and LML [6] are metric learning-based methods, and PAR [21] and Spindle [20] utilized deep learning method to extract feature. The comparison results are shown in Table I. There are two variants of our method. "RSL+Pre†" means that we use GOG descriptor for person representation. "RSL+Pre‡" denotes that GOG descriptor is for person representation and the projection matrix L is learned by null space LDA (NLDA) [18] instead of the model in Eq.5.

TABLE I

PERFORMANCE COMPARISON WITH THE STATE-OF-THE-ART METHODS ON VIPER, PRID450S AND GRID DATASETS. THE BEST AND SECOND BEST RESULTS (%) ARE RESPECTIVELY SHOWN IN RED AND BLUE. BETTER VIEWED IN COLOUR.

		VIPeR			PRID450S			GRID		
Methods	**Reference**	r=1	r=10	r=20	r=1	r=10	r=20	r=1	r=10	r=20
CSL	ICCV2015[14]	34.8	82.3	91.8	44.4	82.2	89.8	-	-	-
LOMO+XQDA	CVPR2015[10]	40.0	80.5	91.1	61.4	91.0	95.3	16.6	41.8	52.4
Kernel HPCA	ICPR2016[12]	39.4	85.1	93.5	52.2	92.8	94.4	-	-	-
LSSCDL	CVPR2016[19]	42.7	84.3	91.9	60.5	88.6	93.6	22.4	51.3	61.2
SCSP	CVPR2016[4]	53.5	91.5	96.7	58.8	91.3	96.3	24.2	54.1	65.2
OL-MANS	ICCV2017[23]	45.0	85.0	93.6	-	-	-	30.2	49.2	59.4
PAR	ICCV2017[21]	48.7	85.1	93.0	-	-	-	-	-	-
LML	TIP2017[6]	50.4	88.7	95.0	64.5	92.1	96.0	19.8	46.5	58.1
Spindle	CVPR2017[20]	53.8	83.2	92.1	67.0	89.0	92.0	-	-	-
RSL+Pre	**Ours**	31.8	75.3	87.2	45.6	83.2	91.2	13.8	43.2	56.4
RSL+Pre†	**Ours**	43.3	85.7	93.1	55.6	89.7	95.6	18.1	49.6	60.0
RSL+Pre‡	**Ours**	49.8	89.3	95.3	68.1	94.0	97.3	25.1	57.2	68.1

It can be seen from Table I that, RSL+Pre‡ outperforms most of compared methods on all the three dataset expect SCSP, OL-MANS and Spindle. SCSP performs the best at rank10 and rank20 on VIPeR dataset, but RSL+Pre‡ achieves better performance than SCSP on other two datasets. The performance of RSL+Pre‡ is better than OL-MANS on all the dataset expect that the rank1 of OL-MANS is better by 5.1% on GRID dataset. Similarly, Spindle achieves best performance at rank1 on VIPeR dataset, but beyond that RSL+Pre‡ has better performances. Moreover, RSL+Pre and RSL+Pre† are less competitive than other methods. This is due to weaker feature description and metric learning. A better performance can be achieved by RSL+Pre‡ with GOG descriptor and NLDA-based metric learning, which proves the effectiveness of region-specific metric learning.

C. Discussion

Performance of feature pre-processing. The benefit from the proposed feature pre-processing scheme lies in two points: 1) the improvement of the feature's discriminative power and 2) the reduce of the feature dimension. We compare the proposed method RSL+Pre† and the method without feature pre-processing denoted by RSL†. The comparison shown from Table II on VIPeR dataset shows that the matching accuracies are improved and the feature dimension is reduced with the proposed feature pre-processing scheme. It follows that the proposed method can save the memory and running time, meanwhile improve the performance of person re-id.

Analysis of parameter τ. Choosing an appropriate parameter τ is a critical issue in feature pre-processing. A large τ may result in more useful features being removed, while a small τ may not be effective in improving the accuracy of person re-id. We study how the value of τ affects the performance. We conduct the experiment on VIPeR dataset: randomly partition the training dataset into two parts, one for model learning and the remaining for validation. The performances for various values of $\tau = 0.05, 0.1, 0.5, 0.9, 1.3$, are given in Table III, which shows RSL+Pre† with $\tau = 0.1$ achieves the best result.

Comparison with region-specific/generic metric learning. As described in former, the distribution may vary among different region-features. We show that region-specific metric learning can boost the performance than region-generic one. For a fair comparison with RSL+Pre†, the region-generic metric learning (denoted by RGL+Pre†) is modeled in Eq.5 and GOG descriptor is used for person representation. The comparison results are reported in Table IV. It can be seen that region-specific metric learning greatly improves the performance and rank1 result increases by 17.5%.

Analysis of the number of blocks. For the model in Eq.6, the number of blocks in L is determined by the number of regions S. If person image is divided into S regions when extracting features, S sub-projection matrixes $\{L_{ss}\}_{s=1}^{S}$ need to be learned. Compared with learning L without the structure of block, the model with a block-diagonal structure L considers the spatial structure of person image and the distribution difference among different region-features. Similarly, the sub-projection matrix L_{ss} can also be learned with a block-diagonal structure, i.e $L_{ss} = diag(\{L_{ss}\}_{11}, \{L_{ss}\}_{22}, \ldots, \{L_{ss}\}_{pp})$. The experiments are conducted with various values of $p = 2, 4$. We use the GOG descriptor for person representation and the descriptor is proposed based on 7 regions of the images, so $p = 2$ and

TABLE II

PERFORMANCE COMPARISON WITH THE METHODS BASED ON DIFFERENT VARIANTS OF FEATURE PRE-PROCESSING ON VIPER DATASET.

	r=1	r=10	r=20	dimension
RSL†	44.5	86.1	92.7	4417
RSL+Pre†	44.5	87.0	93.0	3953

TABLE III

PERFORMANCE COMPARISON WITH THE PROPOSED RSL+PRE METHODS BASED ON DIFFERENT THRESHOLD τ ON VIPER DATASET.

τ	r=1	r=10	r=20
0.05	44.5	86.9	92.8
0.1	44.5	87.0	93.0
0.5	42.2	85.7	92.9
0.9	39.6	83.5	91.9
1.3	38.7	81.5	90.5

TABLE IV
PERFORMANCE COMPARISON WITH THE METRIC LEARNING METHODS BASED ON DIFFERENT NUMBER OF BLOCKS LEARNED IN PROJECTION MATRIX L ON VIPER DATASET.

	r=1	r=10	r=20
RGL+Pre†	27.0	74.5	86.0
RSL+Pre†	44.5	87.0	93.0
RSL+Pre†_14	41.5	85.5	92.8
RSL+Pre†_28	34.9	79.2	89.2

$p = 4$ represent the number of regions S equaling to 14 (denoted by RSL+Pre†_14) and 28 (denoted by RSL+Pre†_28), respectively. Experimental results are shown in Table IV. The performance cannot be improved with more number of block in L. It is because that the projection matrix L is divided into $S = 14$ and $S = 28$ blocks without following the actual structure of feature descriptor of person, and the distribution of features within the same region is similar.

V. CONCLUSION

In this paper, we develop a novel region-specific metric learning method with input of improved features by removing weakly discriminative features for person re-id. The proposed RSL+Pre method firstly performs PCA to the extracted features for dimensional reduction and selects features with most discriminative information from the low-dimensional feature space. Specifically, the features are selected by computing the *discrimination* measurement of each dimension feature, based on comparing the distribution of features with the ideally discriminative feature that should be consistency for intra-class and difference for inter-class. After that, in consideration of the spatial structure of person image and the difference on the distribution of region-features among different regions of person image, a metric learning algorithm is proposed by learning a projection matrix L with a block-diagonal structure constraint, i.e. for each region-features, a specific sub-projection matrix is learned. Numerous experiments on VIPeR, PRID450S and GRID datasets are conducted. The results validate the effectiveness of the proposed method over state-of-the-art methods.

ACKNOWLEDGMENT

This work is supported by the National Key R&D Program of China under Grant 2017YFC0803505.

REFERENCES

[1] Min Cao, Chen Chen, Xiyuan Hu, and Silong Peng. From groups to co-traveler sets: Pair matching based person re-identification framework. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2573–2582, 2017.

[2] Change Loy Chen, Tao Xiang, and Shaogang Gong. Multi-camera activity correlation analysis. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1988–1995, 2009.

[3] Chen Chen, Min Cao, Xiyuan Hu, and Silong Peng. Key person aided re-identification in partially ordered pedestrian set. In *Conference the British Machine Vision Conference*, 2017.

[4] Dapeng Chen, Zejian Yuan, Badong Chen, and Nanning Zheng. Similarity learning with spatial constraints for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1268–1277, 2016.

[5] Ying Cong Chen, Xiatian Zhu, Wei Shi Zheng, and Jian Huang Lai. Person re-identification by camera correlation aware feature augmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, PP(99):1–1, 2017.

[6] Sun Chong, Wang Dong, and Huchuan Lu. Person re-identification via distance metric learning with latent variables. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 26(1):23–34, 2016.

[7] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Computer Vision and Pattern Recognition*, pages 2360–2367, 2010.

[8] Doug Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007.

[9] F. Jurie and A. Mignon. Pcca: A new approach for distance learning from sparse pairwise constraints. In *Computer Vision and Pattern Recognition*, pages 2666–2672, 2013.

[10] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *Computer Vision and Pattern Recognition*, pages 2197–2206, 2015.

[11] Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, and Yoichi Sato. Hierarchical gaussian descriptor for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1363–1372, 2016.

[12] Raphael Felipe Prates and William Robson Schwartz. Kernel hierarchical pca for person re-identification. In *International Conference on Pattern Recognition*, 2017.

[13] Peter M. Roth, Martin Hirzer, Martin Kostinger, Csaba Beleznai, and Horst Bischof. Mahalanobis distance learning for person re-identification. In *Person Re-Identification. Springer London*, pages 247–267, 2014.

[14] Yang Shen, Weiyao Lin, Junchi Yan, Mingliang Xu, Jianxin Wu, and Jingdong Wang. Person re-identification with correspondence structure learning. In *IEEE International Conference on Computer Vision*, pages 3200–3208, 2016.

[15] John Parr Snyder. Map projections – a working manual. *Geological Survey Professional Paper*, 1395, 1987.

[16] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z. Li. Salient color names for person re-identification. In *European Conference on Computer Vision*, pages 536–551, 2014.

[17] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. Deep metric learning for person re-identification. In *International Conference on Pattern Recognition*, pages 34–39, 2014.

[18] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In *Computer Vision and Pattern Recognition*, pages 1239–1248, 2016.

[19] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Ruan Xiang. Sample-specific svm learning for person re-identification. In *Computer Vision and Pattern Recognition*, pages 1278–1287, 2016.

[20] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 907–915, 2017.

[21] Liming Zhao, Xi Li, Jingdong Wang, and Yueting Zhuang. Deeply-learned part-aligned representations for person re-identification. 2017.

[22] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Computer Vision and Pattern Recognition*, 2017.

[23] Jiahuan Zhou, Pei Yu, Wei Tang, and Ying Wu. Efficient online local metric adaptation via negative samples for person re-identification. In *IEEE International Conference on Computer Vision*, pages 2439–2447, 2017.

[24] Qin Zhou, Shibao Zheng, Hua Yang, Yu Wang, and Hang Su. Joint instance and feature importance re-weighting for person reidentification. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1546–1550, 2016.