Pattern Recognition 97 (2020) 107014

Contents lists available at ScienceDirect

Pattern Recognition

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

Similarity learning with joint transfer constraints for person re-identification

Cairong Zhao[a,1,*], Xuekuan Wang[a,1], Wangmeng Zuo[b], Fumin Shen[c], Ling Shao[d], Duoqian Miao[a]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, China
[b] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
[c] School of Computer Science and Engineering, University of Electronic Science and Technology, Chengdu, China
[d] Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

ARTICLE INFO

Article history:
Received 7 December 2018
Revised 4 June 2019
Accepted 17 August 2019
Available online 28 August 2019

Keyword:
Person re-identification
Feature extraction
Similarity learning

ABSTRACT

The inconsistency of data distributions among multiple views is one of the most crucial issues which hinder the accuracy of person re-identification. To solve the problem, this paper presents a novel similarity learning model by combining the optimization of feature representation via multi-view visual words reconstruction and the optimization of metric learning via joint discriminative transfer learning. The starting point of the proposed model is to capture multiple groups of multi-view visual words (*MvVW*) through an unsupervised clustering method (i.e. *K-means*) from human parts (e.g. head, torso, legs). Then, we construct a joint feature matrix by combining multi-group feature matrices with different body parts. To solve the inconsistent distributions under different views, we propose a method of joint transfer constraint to learn the similarity function by combining multiple common subspaces, each in charge of a sub-region. In the common subspaces, the original samples can be reconstructed based on *MvVW* under low-rank and sparse representation constraints, which can enhance the structure robustness and noise resistance. During the process of objective function optimization, based on confinement fusion of multi-view and multiple sub-regions, a solution strategy is proposed to solve the objective function using joint matrix transform. Taking all of these into account, the issue of person re-identification under inconsistent data distributions can be transformed into a consistent iterative convex optimization problem, and solved via the inexact augmented Lagrange multiplier (*IALM*) algorithm. Extensive experiments are conducted on three challenging person re-identification datasets (i.e., VIPeR, CUHK01 and PRID450S), which shows that our model outperforms several state-of-the-art methods.

1. Introduction

The central theme of person re-identification (Re-ID) is to match two pedestrian images undergoing significant human appearance changes in viewpoint, illumination and pose across camera views (see Fig. 1). To address this challenge, many algorithms have been proposed, and the researches can be divided into two major directions.

One of the research directions is to develop robust feature descriptor for representing human appearance [10,44]. Currently, most appearance-based Re-ID methods use low-level visual features as feature representations of pedestrian images such as color features [16] and texture features [9]. To improve the performance of Re-ID, a wide variety of fusion methods have been designed [41], such as deep context-aware features [26], CRAFT [3], joint global and local feature learning [30], hierarchical Gaussian descriptor [10], local maximal occurrence representation [6], cross-modality feature [27] and salience matching [23]. Besides, deep learning is also a noteworthy category of methods which has exhibited promising performance in learning feature representation [17]. However, it remains very challenging to design a feature representation that is discriminative, reliable and invariant to severe changes and misalignment across disjoint views.

Another research direction, e.g., metric learning [28,37], tries to learn a similarity function or a robust distance metric to optimize the matching score. Typical metric learning methods include Local Fisher Discriminant Analysis (LFDA) [11], kernel-based method [19], Cross-view Quadratic Discriminant Analysis (XQDA) [6], supervised smoothed manifold [25], domain adaptation [39], reference constraints [40], ranking [2] and deep metric learning [14]. Although these metric learning based methods outperform most

* Corresponding author.
E-mail address: zhaocairong@tongji.edu.cn (C. Zhao).
[1] The authors contribute equally to this work.

https://doi.org/10.1016/j.patcog.2019.107014
0031-3203/

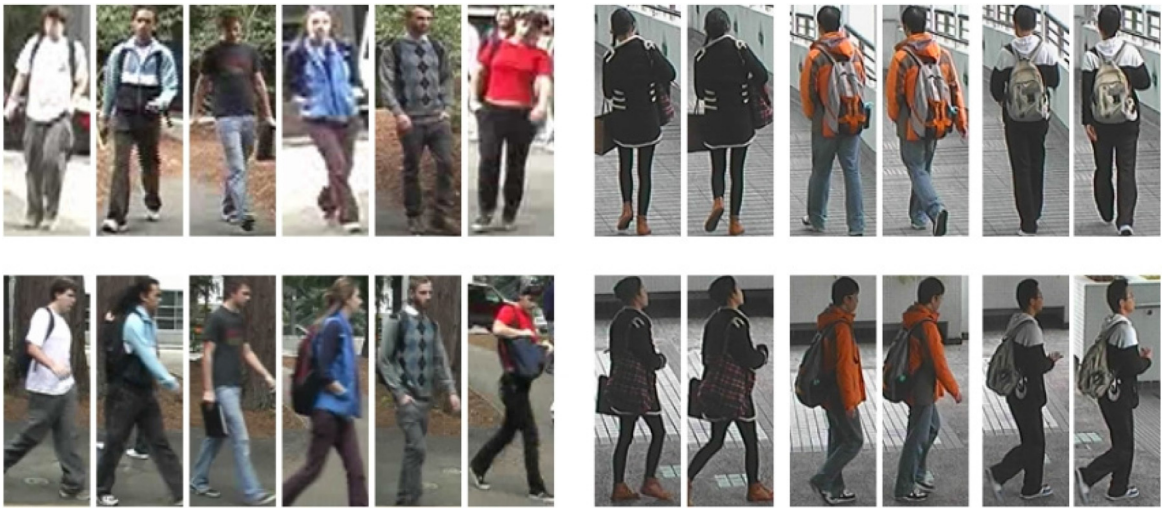

Fig. 1. Examples of person re-identification datasets.

Re-ID approaches, they are nevertheless limited by some classical problems, such as the inconsistent distributions between multiple views and the small sample size (***SSS***) issue for model learning.

To address these problems, we propose a novel similarity learning approach under joint transfer constraints, in which four groups of multi-view visual words (***MvVW***) can be captured, including three groups of local features and one group of global features via an unsupervised clustering method (***K-means***), to effectively describe the structure of human body. Also, the ***MvVW*** has the ability to integrate multi-view information. Based on the ***MvVW*** representation, we learn to reconstruct the original samples with the assistance of transformation matrix, reconstruction coefficient matrix and noise matrix. Note that for the sake of ensuring consistent distributions of sample data, we utilize transfer learning [13] to obtain a common subspace, denoted as the transformation matrix. Meanwhile, we impose joint low-rank and sparse constraints on the reconstruction coefficient matrix and noise matrix in order that more relevant samples from different domains are interlaced, in comparison to irrelevant samples in these domains [8]. Furthermore, we apply discriminative analysis to transfer matrix and obtain discriminative low-level transfer features, and then utilize the learned transfer matrix to compute the reconstruction coefficient matrix which is defined as the mid-level features in our model. To get the consistent optimal solutions, we combine the discriminative analysis with the mid-level features and transfer learning, and then produce the solutions via the proposed method of similarity learning function which can maintain the consistency of representation and metric learning [5]. In addition, by employing light weighting method, max and min operator, we can expand the samples to suppress the adverse effect of the SSS problem on Re-ID. Compared to deep learning based methods, the training process of our method does not require a large number of samples, thus our method can better cope with the SSS problem. When the number of samples is sufficient, the features extracted by our method may not be robust enough compared to deep features, but when the number of samples is small, the performance of our method is much higher than the ones using deep features.

Finally, we describe the motivation and contribution of this paper as follows:

1.1. Motivation

Although considerable progress has been made in person Re-ID, there remains some limitations for most existing methods:

(1) Most approaches assume that the data distributions under multi-camera views are consistent. However, this assumption is one-sided because important attributes of each camera view are different in practice. In our approach, we apply transfer-learning method to seek a common subspace for different camera views, and obtain the mid-level features for similarity metric;
(2) Traditional descriptors are mainly based on low-level features. However, mid-level features are also helpful for person Re-ID. In our approach, we combine transfer learning, discriminant analysis and sparse constraint to learn mid-level features, and then consider multi-level feature for similarity metric;
(3) Most metric learning methods suffer from the SSS problem and it is difficult to obtain an optimal solution. To address the problems above, we propose a novel similarity learning method under joint transfer constraints for multi-view and multi-region person Re-ID. In particular, it should be noted that the relaxed loss term considering sample pairs instead of single sample can alleviate the ***SSS*** issue.

1.2. Contribution

The main contributions of our work are summarized as follows:

(1) We propose a novel similarity learning method by considering joint transfer constraints which can learn a discriminative subspace with consistent data distributions and perform better than the competing methods for multi-view person Re-ID;
(2) The mid-level features are introduced by defining the reconstruction matrix, solved via the inexact augmented Lagrange multiplier (***IALM***) algorithm, and then integrated with low-level features and discriminative transfer features to describe the appearance of pedestrian images;
(3) In order to fuse the local and global features, we design a joint transfer constraint to solve the optimal function. For this optimization problem, a new solution strategy is presented by using joint matrix transform. Furthermore, the proposed method is shown to be effective and efficient through person Re-ID experiments on three public datasets.

2. Related work

In complicated real-world tasks, the data taken from different domains have different feature spaces and different types of data

Table 1
Notations.

Notation	Description
X_{im}	feature matrix
$\tilde{X}$	the joint feature matrix
P_i	transfer matrix for different with different regions
$\tilde{P}$	the joint transfer matrix
D_i	the multi-view word matrices with different regions
$\tilde{D}$	the joint multi-view word matrix
Z_i	the reconstruction coefficient matrices with different regions
$\tilde{Z}$	the joint reconstruction coefficient matrix
E_i	the noise matrices with different regions
$\tilde{E}$	the joint noise matrix
$\alpha, \beta, \gamma, \eta, \rho, \mu, \lambda, \sigma$	model parameters

distributions [13]. To address the problem of inconsistent distributions, numerous approaches based on transfer learning have been proposed and applied for various visual tasks [20,42,43].

2.1. Transfer learning for person Re-ID

For person Re-ID, one of the essential requirements is to build a robust matching model which can always work well from one type of scene to another under the challenges of camera viewing angle differences, pose variation, occlusion change, etc. [18]. Accordingly, the transfer learning methods have been exploited to address the challenges of cross-scenario transfer. In [1], Tamar et al. proposed the approach of Implicit Camera Transfer (ICT) to model the binary relation by training a (non-linear) binary classifier with concatenations of vector pairs captured from different camera views. Similarly, considering the consistency of cross view, Wang et al. [18] combined the learning of the shared latent subspace and the learning of the corresponding task specific subspace to get the similarity measurement for each task in cross-scenario transfer person Re-ID. Furthermore, Zheng et al. [24] formulated a transfer local relative distance comparison (t-LRDC) model to address the open world person Re-ID problem. In addition, Shi et al. [15] suggested a framework to learn a semantic attribute model from the existing fashion datasets, and adapted the resulting model to facilitate person Re-ID. Different from the above-mentioned methods, Lv et al. [31] considered unsupervised cross-dataset and utilize Bayes analysis for fusing spatial-temporal patterns for person Re-ID under different domains. Wang et al. [29] also investigated the problem of unsupervised person Re-ID by learning transferable joint attribute-identity feature.

2.2. Transfer subspace learning

In preliminary works [38], we assume that the original samples can be linearly represented by transfer learning in a common subspace. According to Shao et al. [13], we can reconstruct the two domain samples (X, Y) using the coefficient matrix Z and transfer learning (ensuring the consistency of distributions), as follows:

$$P^T X = P^T Y Z \tag{1}$$

where P denotes the transfer matrix, which can be used to obtain a common subspace and can minimize the divergence between the distributions of both domains. However, due to the fact that n samples belong to c different classes and $n \gg c$, these samples should be drawn from $\boldsymbol{c}$ different subspace.

Therefore, the coefficient matrix Z is expected to be of low rank [20], and the F-norm constraint can be further incorporated to preserve the local structure of data such that each source sample can be well reconstructed from a few samples. Thus the transfer matrix and coefficient matrix are obtained by solving the following optimization problem,

$$\min_{P,Z} rank(Z) + \alpha \|Z\|_F^2, s.t. \quad P^T X = P^T Y Z \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm, $rank(\cdot)$ is a nonconvex function, and α is the penalty parameter. In order to alleviate the effect of noise, we introduce the matrix E with sparse constraint to model noise, resulting in the following model,

$$\min_{P,Z} rank(Z) + \alpha \|Z\|_F^2 + \beta \|E\|_1, s.t. \quad P^T X = P^T Y Z + E \tag{3}$$

We adopt the nuclear norm to relax the rank function [20], and the modified model can be written as

$$\min_{P,Z} \|Z\|_* + \alpha \|Z\|_F^2 + \beta \|E\|_1, s.t. \quad P^T X = P^T Y Z + E \tag{4}$$

where $\|Z\|_*$ is the nuclear norm of matrix Z.

3. Our approach

In this section, we first revisit the polynomial feature map. Based upon the map, we present a novel framework of transfer learning for multiple features by a constrained similarity function, and formulate the learning problem specifically designed for person Re-ID. The abbreviations of main variables and parameters used in this paper are summarized in Table 1.

3.1. Multi-view visual words by K-means

To capture structure information and multi-view information, we propose a descriptor called multi-view Visual words (***MvVW***) using an unsupervised clustering method of ***K-means***. Firstly, we divide a pedestrian image (x_i) into five horizontal stripes, along the vertical direction of human body consistently. Next, we define each low-level feature histogram as a visual word, and then capture three groups of local visual words from three horizontal stripes $\frac{1}{5}$, $\frac{2}{5}$, $\frac{2}{5}$ and one group of global visual words from the whole person images, as shown in Fig. 2. And these three local areas usually include the head, torso and legs of the human body structure. Furthermore, we employ k-means to fuse the multi-view information and obtain seven groups of ***MvVW***. Note that, a light weighting method, as well as the max and min operators, is employed to expand the sample data for reducing the effect of the SSS problem.

In the following, we define $\boldsymbol{MvVW} = \{D_i\}, \ i \in \{l0, \ l1, \ l2, \ g\}$, where $\{D_i\}$ represents the i th group of ***MvVW***, $\{D_0, D_1, D_2\}$ are the local multi-view visual words obtained from five horizontal stripes of pedestrian images and D_3 is the global multi-view visual word obtained from the whole pedestrian images. Then, we use each group of ***MvVW*** to reconstruct the corresponding region of multi-view person sample data X. It is worth noting that the head of a human body is most probably represented by the other

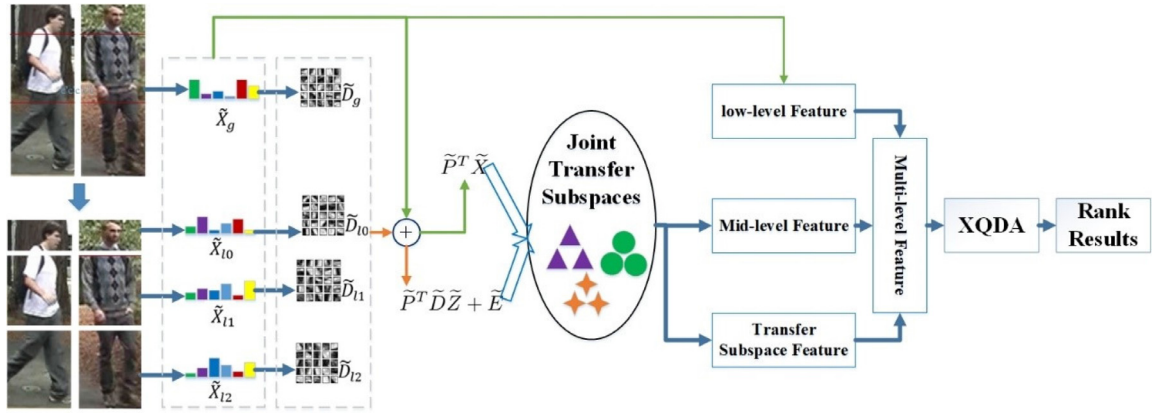

Fig. 2. The framework of our proposed method: 1) Capture the global feature $\tilde{X}_g$, local feature $\tilde{X}_{l0}, \tilde{X}_{l1}, \tilde{X}_{l2}$ from different regions and obtain the multi-view visual words $\tilde{D}_g$, $\tilde{D}_{l0}, \tilde{D}_{l1}, \tilde{D}_{l2}$. 2) Joint the multi-group features capturing from three local regions and one global region, and learn the joint transfer subspace with consistent distribution constraints. Then combine sparse and low-rank constraints (shown in Eq. (10)) in the joint transfer subspace and solve this optimal function with a new approach described in section IV. 3) Considering the advantages of multi-level features, we combine low-level and mid-level features and utilize the metric method of XQDA to obtain the final rank results for person Re-ID.

heads with similar structures. We can therefore formulate the reconstruction problem as:

$$\min_{P,Z} \|Z\|_* + \alpha\|Z\|_F^2 + \beta\|E\|_1, s.t.\ \ P^T X = P^T DZ + E \tag{5}$$

where Z is the reconstruction coefficient matrix and can be captured from the low-level features, denoted as the mid-level features for person Re-ID. Considering two different domains of D and X, they have different data distributions. To address this problem, we will utilize transfer learning to seek a subspace with consistent data distribution.

3.2. Multiple transfer features function

In our proposed method, we consider three kinds of local features and one kind of global features, and different features have inconsistent distributions. Thus, to combine multiple descriptors, we reformulate the optimal function based on Eq. (5) as:

$$\min_{Z_i,P_i,E_i} \sum_{i=0}^{n} \left(\|Z_i\|_* + \|Z_i\|_F^2 + \|E_i\|_1\right),\ \text{s.t. } P_i^T X_i = P_i^T D_i Z_i + E_i \tag{6}$$

where X_i represents the set of the i th feature and n is the number of the group of features.

3.2.1. Discriminant term

We combine the discriminant analysis for the transfer matrix of P and define the discriminant term as:

$$\min\left(-P_i^T \Sigma_{E_i} P_i\right)\ s.t. P_i^T \Sigma_{I_i} P_i = 1 \tag{7}$$

where Σ_{I_i} and Σ_{E_i} are the covariance matrices of the intrapersonal variations and the inter-personal variations for the sample of X_i. Furthermore, according to Lagrange operator, we can rewrite the discriminative term as:

$$\hat{J}(P_i) = -P_i^T \Sigma_{E_i} P_i + \eta\left(P_i^T \Sigma_{I_i} P_i - 1\right) \tag{8}$$

3.2.2. Relaxed loss term

The training data for person Re-ID can be organized as follows. Given the descriptors of probe images $X_i = \{x_{i0}, x_{i1}, \ldots, x_{im}, \ldots, x_{iM}\}$, $i = [0, 1, 2, 3]$ represents the descriptors with different body parts. M is the number of probe images. x_{im} is associated with two sets of gallery images: a positive set X_{im}^+ composed of the descriptors about the same person and a negative set X_{im}^- composed of the descriptors about different persons. To enforce the relative comparison, we adopt a relaxed loss term [32]:

$$L(P_i) = \frac{1}{N}\sum_{i=0}^{N}\left[1 - \frac{\sum_{x_{ip}\in X_{im}^+,\ x_{iq}\in X_{im}^-} s\left(x_{ip}, x_{iq}, x_{im}\right)}{\left|X_{im}^+\right| \cdot \left|X_{im}^-\right|}\right]_+ \tag{9}$$

Where $s(x_{ip}, x_{iq}, x_{im}) = \|P_i^T x_{iq} - P_i^T x_{im}\|_F^2 - \|P_i^T x_{ip} - P_i^T x_{im}\|_F^2$ and $[\cdot]_+$ denotes the hinge loss. N is the number of sample pairs. Given a probe descriptor, instead of forcing every positive pair to achieve a higher score than negative pairs, we require the average score of positive pairs should be higher than the average score of the negative pairs at least by a margin 1, representing as [1-...]. The relaxed loss term only consists of N constraints, largely accelerating the training in comparison with the non-relaxed one.

3.2.3. Objective function

According to Eqs. (6) and (9), the overall model for person Re-ID is given by:

$$\min_{Z_i,P_i,E_i} \sum_{i=0}^{n} \left(\|Z_i\|_* + \|Z_i\|_F^2 + \|E_i\|_1 + L(P_i) + \hat{J}(P_i)\right)\ s.t.\ P_i^T X_i = P_i^T D_i Z_i + E_i \tag{10}$$

4. Optimization

4.1. Solution

To clarify the notation, we first concatenate the multiple feature descriptors in each sub-region together:

$$\tilde{X} = \begin{matrix} X_0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & X_i & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & X_{n-1} \end{matrix}$$

where $n = 4$.

And, we define the multiple visual words matrix as follows:

$$\tilde{D} = \begin{matrix} D_0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & D_i & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & D_{n-1} \end{matrix}$$

Furthermore, with $\tilde{P} = [P_0, \ldots, P_i, \ldots, P_n]$, the similarity function of Eq. (10) can be reformulated as:

$$\min_{\tilde{P},\tilde{Z},\tilde{E}} \|\tilde{Z}\|_* + \alpha\|\tilde{Z}\|_F^2 + \beta\|\tilde{E}\|_1 + \gamma L(\tilde{P}) + \lambda J(\tilde{P})\ s.t. \tilde{P}^T\tilde{X} = \tilde{P}^T\tilde{D}\tilde{Z} + \tilde{E} \tag{11}$$

In the light of the non-convexity of Eq. (11), we adopt the inexact ALM (IALM) algorithm [20] to solve this optimization problem. First, we introduce variables $\widetilde{Z}_1$, $\widetilde{Z}_2$ and impose two constraints on $\tilde{Z}$ to relax the original problem,

$$\min_{\tilde{P},\tilde{Z},\widetilde{Z}_1,\widetilde{Z}_2,\tilde{E}} \left\|\widetilde{Z}_1\right\|_* + \alpha\left\|\widetilde{Z}_2\right\|_F^2 + \beta\|\tilde{E}\|_1 + \gamma L(\tilde{P}) + \lambda J(\tilde{P}) \; s.t. \tilde{P}^T\tilde{X} = \tilde{P}^T\tilde{D}\tilde{Z} + \tilde{E},\ \tilde{Z} = \widetilde{Z}_1 = \widetilde{Z}_2 \tag{12}$$

More specifically, the function of Eq. (12) can be written as:

$$\begin{aligned}\min_{\tilde{P},\tilde{Z},\widetilde{Z}_1,\widetilde{Z}_2,\tilde{E},\mathcal{L}_1,\mathcal{L}_2,\mathcal{L}_3} & \left\|\widetilde{Z}_1\right\|_* + \alpha\left\|\widetilde{Z}_2\right\|_F^2 + \beta\left\|\tilde{E}_1\right\| + \gamma L(\tilde{P}) \\ & + \lambda J(\tilde{P}) + \left\langle \mathcal{L}_1, \tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\right\rangle \\ & + \left\langle \mathcal{L}_2, \tilde{Z} - \widetilde{Z}_1\right\rangle + \left\langle \mathcal{L}_3, \tilde{Z} - \widetilde{Z}_2\right\rangle \\ & + \frac{\mu}{2}\left(\left\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\right\|_F^2 + \left\|\tilde{Z} - \widetilde{Z}_1\right\|_F^2 \right. \\ & \left. + \left\|\tilde{Z} - \widetilde{Z}_2\right\|_F^2\right)\end{aligned} \tag{13}$$

where $\mu > 0$ and $\gamma > 0$ are penalty parameters. $\mathcal{L}_1 \in R^{m\times n}$, $\mathcal{L}_2 \in R^{m\times p}$, $\mathcal{L}_3 \in R^{m\times n}$ are Lagrange multipliers. The main steps of solving Eq. (13) are given as follows and all steps have closed-form solutions.

Step 1 (Update $\tilde{P}$): $\tilde{P}$ can be updated by solving the following optimization problem,

$$\min_{\tilde{P}} \frac{\mu}{2}\left\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E} + \frac{\mathcal{L}_1}{\mu}\right\|_F^2 + \gamma L(\tilde{P}) + \lambda J(\tilde{P}) \tag{14}$$

Then, we can obtain the closed-form solution of P^*.

$$P^* = \left(\mu G_1 G_1^T - \lambda\Sigma_E + \eta\Sigma_I + \sigma I\right)^{-1}\left(\mu G_1 G_2^T - \gamma\Phi(\tilde{P})\right) \tag{15}$$

where $G_1 = \tilde{X} - \tilde{D}\tilde{Z}$ and $G_2 = \tilde{E} - \frac{\mathcal{L}_1}{\mu}$. $\Phi(\tilde{P})$ represent the partial derivatives of P.

Step 2 (Update $\tilde{Z}$): $\tilde{Z}$ is updated by solving the optimization problem,

$$\min_{\tilde{Z}} \left\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E} + \frac{\mathcal{L}_1}{\mu}\right\|_F^2 + \left\|\tilde{Z} - \widetilde{Z}_1 + \frac{\mathcal{L}_2}{\mu}\right\|_F^2 + \left\|\tilde{Z} - \widetilde{Z}_2 + \frac{\mathcal{L}_2}{\mu}\right\|_F^2 \tag{16}$$

Then, we can obtain the closed-form solution of Z^*.

$$Z^* = \left(\mu\tilde{D}^T\tilde{P}\tilde{P}^T\tilde{D} + 2\mu I\right)^{-1}\left(G_4 + G_5 - \tilde{D}^T\tilde{P}G_3\right) \tag{17}$$

where $G_3 = \tilde{P}^T\tilde{X} - \tilde{E} + \frac{\mathcal{L}_1}{\mu}$, $G_4 = \widetilde{Z}_1 - \frac{\mathcal{L}_2}{\mu}$, $G_5 = \widetilde{Z}_2 - \frac{\mathcal{L}_3}{\mu}$.

Step 3 (Update $\widetilde{Z}_1$): $\widetilde{Z}_1$ is updated by solving optimization problem,

$$\min_{\widetilde{Z}_1} \left\|\widetilde{Z}_1\right\|_* + \frac{\mu}{2}\left\|\tilde{Z} - \widetilde{Z}_1 + \frac{\mathcal{L}_2}{\mu}\right\|_F^2 \tag{18}$$

The closed-form solution of $\widetilde{Z}_1^*$ is

$$\widetilde{Z}_1^* = \theta_{\frac{1}{\mu}}\left(\tilde{Z} + \frac{\mathcal{L}_2}{\mu}\right) \tag{19}$$

where $\theta_\lambda(A) = US_\lambda(\Sigma)V^T$ is a singular value thresholding operator with respect to a singular value λ; $S_\lambda(\Sigma) = sign(\Sigma)\max(0, |\Sigma - \lambda|)$ is the soft-thresholding operator. $A = U\Sigma V^T$ defines the singular value decomposition of A.

Step 4 (Update $\widetilde{Z}_2$): $\widetilde{Z}_2$ is updated by solving the optimization problem,

$$\min_{\widetilde{Z}_2} \frac{\mu}{2}\left\|\tilde{Z} - \widetilde{Z}_2 + \frac{\mathcal{L}_3}{\mu}\right\|_F^2 \tag{20}$$

And its closed-form solution is obtained by,

$$\widetilde{Z}_2 = \tilde{Z} + \frac{\mathcal{L}_3}{\alpha\mu} \tag{21}$$

Step 5 (Update $\tilde{E}$): $\tilde{E}$ is updated by solving the optimization problem,

$$\min_{\tilde{E}} \beta\|\tilde{E}\|_1 + \mathcal{L}_1, \tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E} + \frac{\mu}{2}\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\|_F^2 \tag{22}$$

with the shrinkage operator [20], the above problem has the following closed-form solution

$$E^* = shrink\left(\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} + \frac{\mathcal{L}_1}{\mu}, \frac{\beta}{\mu}\right) \tag{23}$$

where $shrink(x, a) = \mathrm{sgn}(x)\max(|x| - a, 0)$

Step 6: Multipliers $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$ and iteration step-size $\rho(\rho > 0)$ are updated,

$$\begin{cases}\mathcal{L}_1 = \mathcal{L}_1 + \mu\left(\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\right) \\ \mathcal{L}_2 = \mathcal{L}_2 + \mu\left(\tilde{Z} - \widetilde{Z}_1\right) \\ \mathcal{L}_3 = \mathcal{L}_3 + \mu\left(\tilde{Z} - \widetilde{Z}_2\right) \\ \mu = \min(\rho\mu,\ \mu_{max})\end{cases} \tag{24}$$

Finally, the process of solving Eq. (12) is summarized in Algorithm 1.

Algorithm 1 Solving problem of Eq. (25) by IALM.

Input: $\tilde{X}$,$\tilde{D}$,α, β, γ, η, λ, σ, ρ, μ, μ_{max}
Initialization: $\tilde{Z} = \widetilde{Z}_1 = \widetilde{Z}_2$, $\tilde{E} = 0$, $\mathcal{L}_1 = 0$, $\mathcal{L}_2 = 0$, $\mathcal{L}_3 = 0$, $\alpha = 0.07$, $\beta = 0.2$, $\gamma = 0.1$, $\eta = 0.06$, $\lambda = 0.06$, $\sigma = 0.3$, $\rho = 1.05$, $\mu = 0.4$, $\mu_{max} = 10^7$
Begin:
While not converged
Update $\tilde{P}$ by solving Eq. (15).
Update $\tilde{Z}$ by solving Eq. (17).
Update $\widetilde{Z}_1$ by solving Eq. (19).
Update $\widetilde{Z}_2$by solving Eq. (21).
Update $\tilde{E}$ by solving Eq. (23).
Update the multipliers and parameters by solving Eq. (24).
Given others fixed.
Check the convergence condition:
$\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\|_\infty < \varepsilon$, $\|\tilde{P}^T\tilde{P} - I_p\|_\infty < \varepsilon$, $\|\tilde{Z} - \widetilde{Z}_1\|_\infty < \varepsilon$, $\|\tilde{Z} - \widetilde{Z}_2\|_\infty < \varepsilon$
End while
Output:$\tilde{Z}$, $\tilde{P}$, $\tilde{E}$

4.2. Multi-level descriptor

With the optimal solution of Eq. (12), we can compute the transfer matrix P, and then obtain the transfer subspace features by P^TX. Thus, we can obtain the mid-level descriptor of the construction matrix of Z as follows:

$$\begin{aligned}\min_{\tilde{Z},\widetilde{Z}_1,\widetilde{Z}_2} & \left\|\widetilde{Z}_1\right\|_* + \alpha\left\|\widetilde{Z}_2\right\|_F^2 + \beta\left\|\tilde{E}\right\|_1 + \left\langle \mathcal{L}_1, \tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\right\rangle \\ & + \left\langle \mathcal{L}_2, \tilde{Z} - \widetilde{Z}_1\right\rangle + \left\langle \mathcal{L}_3, \tilde{Z} - \widetilde{Z}_2\right\rangle \\ & + \frac{\mu}{2}\left(\left\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\right\|_F^2 + \left\|\tilde{Z} - \widetilde{Z}_1\right\|_F^2 + \left\|\tilde{Z} - \widetilde{Z}_2\right\|_F^2\right)\end{aligned} \tag{25}$$

The above problem can also be solved using the IALM algorithm, as given in Algorithm 2.

4.3. Metric learning

In our approach, we first get the low-level feature of local maximal occurrence feature (LOMO) [6] and hierarchical Gaussian descriptor (GOG) [10]. Then, we obtain the mid-level features via the aforementioned method, defined respectively as $\tilde{Z}_{LOMO}$ and $\tilde{Z}_{GOG}$, which all include seven reconstruction coefficient matrices. Furthermore, we combine the low-level features ($F_{LOMO} \in R^{d_{LOMO}\times n}$, $F_{GOG} \in R^{d_{GOG}\times n}$) and the mid-level features

Algorithm 2 Solving problem of Eq. (12) by IALM.

Input: $\tilde{X}, \tilde{D}, \tilde{P}, \alpha, \beta, \gamma, \eta, \sigma, \mu, \mu_{max}$
Initialization: $\tilde{Z} = \tilde{Z}_1 = \tilde{Z}_2$, $\tilde{E} = 0$, $\mathcal{L}_1 = 0$, $\mathcal{L}_2 = 0$, $\mathcal{L}_3 = 0$, $\alpha = 0.07$, $\beta = 0.2$, $\gamma = 0.1$, $\eta = 0.06$, $\lambda = 0.06$, $\sigma = 0.3$, $\rho = 1.05$, $\mu = 0.4$, $\mu_{max} = 10^7$
Begin:
While not converged
Update $\tilde{Z}$ by solving Eq. (17).
Update $\tilde{Z}_1$ by solving Eq. (19).
Update $\tilde{Z}_2$ by solving Eq. (21).
Update $\tilde{E}$ by solving Eq. (23).
Update the multipliers and parameters by solving Eq. (24).
Given others fixed.
Check the convergence condition:

$$\|\tilde{P}^T\tilde{X} - \tilde{P}^T\tilde{D}\tilde{Z} - \tilde{E}\|_\infty < \varepsilon,$$
$$\|\tilde{Z} - \tilde{Z}_1\|_\infty < \varepsilon, \|\tilde{Z} - \tilde{Z}_2\|_\infty < \varepsilon$$

End while
Output: $\tilde{Z}$, $\tilde{E}$

($\tilde{Z}_{LOMO} \in R^{m\times n}, \tilde{Z}_{GOG} \in R^{m\times n}$) to formulate our descriptor. Note that, in order to reduce the dimension of our descriptor, we define the new low-level features as $\tilde{F}_{LOMO} \in R^{n\times n}$ and $\tilde{F}_{GOG} \in R^{n\times n}$ by ***PCA***. Therefore, the final dimension of our descriptor is $(2n + 2 \times 4m)$. Finally, we apply the metric learning method of XQDA [6] to measure the similarity for person Re-ID.

4.4. Complexity analysis

For complexity analysis, we can consider two aspects: time complexity and spatial complexity. In our approach, we utilize IALM algorithm to obtain the optimal solution and most of the time computational effort is concentrated on solving inverse matrices, especially when the dimension of sample feature increases. Besides, the spatial complexity is also related to the dimension of sample feature and the number of samples. In addition, our approach concatenates the multiple feature descriptors in each subregion and it leads to an increase in the complexity of the algorithm. This is also a disadvantage of our algorithm and our future work will try to solve this problem.

5. Experiments

5.1. Experimental setting

5.1.1. Datasets

We consider three datasets to train and evaluate the proposed method: VIPeR [4], CUHK01 [23] and PRID450S [10]. VIPeR is one of the most challenging dataset for person Re-ID, due to that the images of the 632 people are taken in different poses, from different viewpoints. CUHK-01 dataset was captured from two camera views, with higher resolution, containing 971 persons, and each person has two images in each view. PRID450S contains 450 image pairs recorded from two different static surveillance cameras. All images are scaled to 128×48 pixels.

5.1.2. Evaluation

For these datasets, we randomly divide all of the images into two equal-size subsets for training and testing, respectively. To quantitatively evaluate the experimental results, the widely used cumulative match curve (CMC) metric is adopted in our experiments. For each query image, we first compute the distance between the query image and each image in the gallery set, then return the top n gallery images with the smallest distance. If the returned list contains at least one image belonging to the same person as the query image, this query is considered as success of top n. Top 1, 5, 10 and 20 are used in our experiments. The experiments are repeated 10 times, and the average rate is used as the evaluation result.

Table 2
The recognition results of our model and other the state-of-the-art methods on VIPeR dataset at rank-1, 5, 10, 20.

Method	Rank=1	Rank=5	Rank=10	Rank=20
Ours	56.32	83.03	90.01	95.76
CRAFT+XQDA [3]	47.82	77.53	87.78	94.84
GOG+XQDA [10]	49.68	79.71	88.67	94.52
LSSL [21]	47.86	78.03	87.63	94.05
LOMO+MLAPG [7]	39.46	70.04	82.41	92.84
LOMO+XQDA [6]	40.00	68.13	80.51	91.08
KCCA+XQDA [35]	33.53	62.31	74.43	85.25
FFN4096+XQDA [34]	28.86	55.35	68.13	81.14
ELF16+XQDA [33]	23.64	47.78	62.5	75.60
ResNet+XQDA [36]	22.66	52.97	67.78	83.70
kLFDA [19]	22.17	47.23	60.27	76.01
MFA [19]	20.46	48.97	63.35	76.08
KISSME [1]	22.53	49.57	64.11	78.15
SVMML [12]	25.41	54.75	70.28	83.50
LFDA [11]	18.34	44.64	57.25	72.96

5.1.3. Parameters

In our model, the parameters include mainly α, β, γ, η, λ, σ, μ and ρ. We obtain the optimal parameters through a method of adjusting one parameter while fixing other parameters. Note that, a large value for μ is adopted for the sake of fast convergence.

5.2. Comparison on the VIPeR dataset

We evaluated our proposed method against 14 existing methods on VIPeR dataset and randomly choose 316 pairs of images for training and leave the rest for testing. These methods consider low-level descriptor, such as LOMO, GOG, CRAFT or deep features, such as ResNet [36], and learn the metric function, such as XQDA, LSSL, kLFDA and so on. For our proposed method, we try to learn the mid-level features and utilize the metric function of XQDA for Re-ID.

5.2.1. Comparison to the state-of-the-art methods

We utilize the K-means method to obtain 4×100 multi-view visual words (***MvVW***) including 3 groups of local and 1 group of global features. Table 2 clearly shows the clear performance superiority of our proposed method over the competing methods. The results of CMC curves are shown in Fig. 3.

It can be seen that our proposed method is obviously better than other state-of-the-art methods. Specifically, our proposed method, achieving a rate of 56.32%, outperforms the 2nd best model (i.e. GOG+XQDA) by 6.64% at rank=1. Furthermore, our proposed method also outperforms other methods at rank>1 from Fig. 3. From these results, we can see that the consideration of the multi-view information and applying the discriminative transfer learning to a common subspace with consistent contributions are necessity for person Re-ID. It further proves our model, capturing the mid-level features, can effectively improve the performance of person Re-ID.

5.2.2. Comparison with the metric learning methods

We evaluate the proposed method with different metric learning methods, including L1-Norm distance, kLFDA and XQDA. The results of CMC curves are shown in Fig. 4 and Table 3. It can be seen that the proposed method with XQDA is better than the other metric learning algorithms, with a gain of 23.49%, in comparison with kLFDA. This indicates that our model with XQDA performs favorably in learning a discriminative transfer subspace as well as an effective metric.

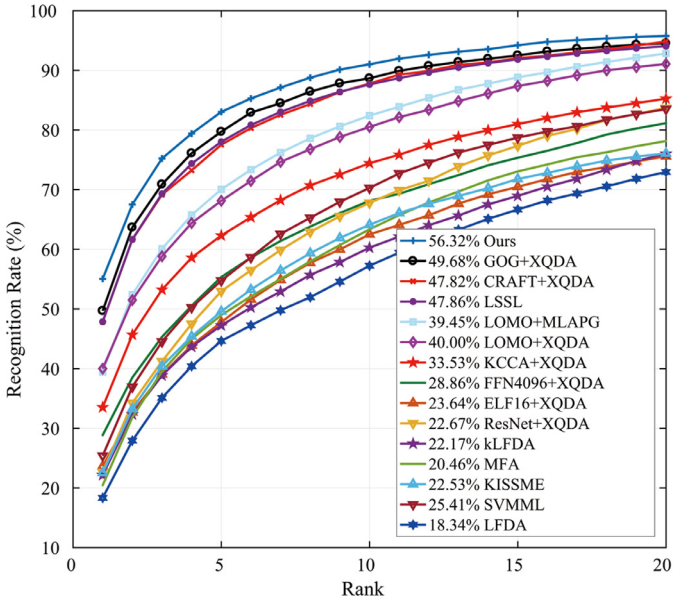

Fig. 3. The CMC curves and rank-1 matching rates on the VIPeR dataset.

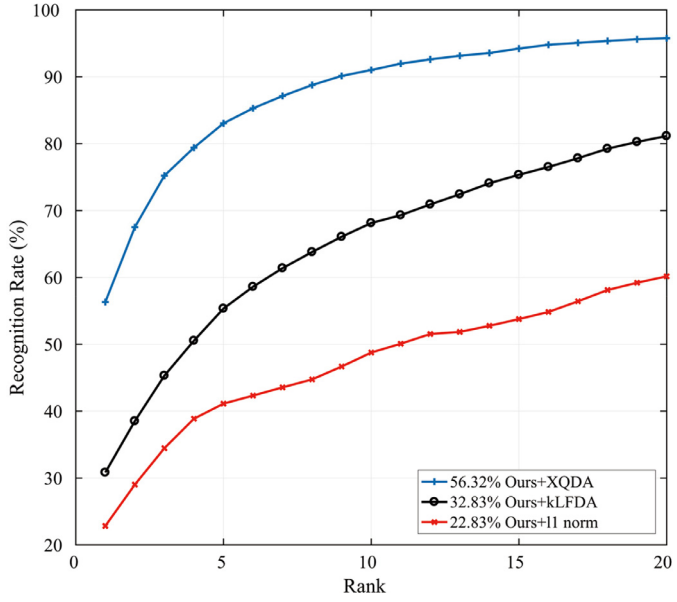

Fig. 4. The CMC curves and rank-1 matching rates by different metric learning methods on the VIPeR dataset.

Table 3
The recognition results of our model with different metric methods on the VIPeR dataset at rank-1, 10, 20.

Method	Rank=1	Rank=10	Rank=20
Ours+XQDA	56.32	90.01	95.76
Ours+kLFDA	22.53	49.57	76.50
Ours+L1-Norm	9.18	24.68	60.75

5.2.3. Effect of the number of multi-view visual words

We compare the performances with different numbers of multi-view visual words (**MvVW**) obtained by K-means, and the results are shown in Fig. 5 and Table 4. It is obvious that our method with the number of (100, 150 and 200) can do better than other models. It can also be observed that our proposed method performs consistently the best with all of **MvVW**. Especially, we can obtain the best result of 57.05% at rank-1 with $m = 150$, which is

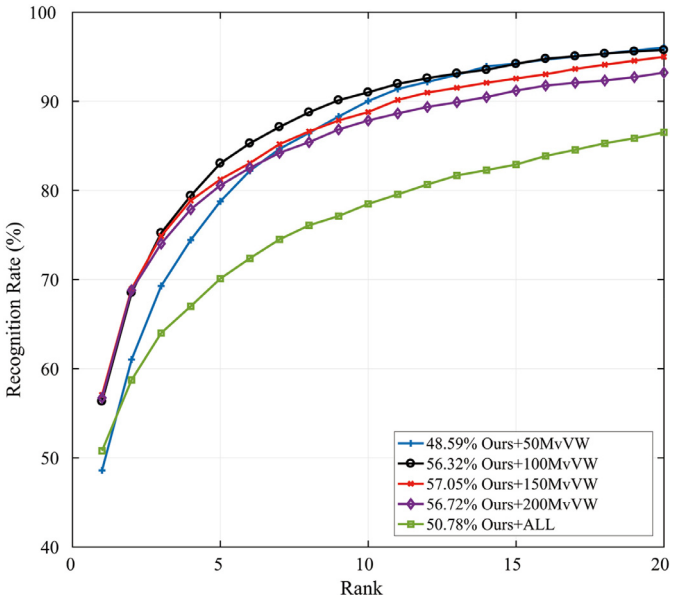

Fig. 5. The CMC curves and rank-1 matching rates on the VIPeR dataset with $m = 50$, 100, 150, 200 and all.

Table 4
The results of comparison with different numbers of multi-view visual words ($m = 50$, 100, 150, 200, All).

Method	Rank=1	Rank=10	Rank=20
Ours(50-*MvVW*)	48.59	78.47	90
Ours(100-*MvVW*)	56.32	83.03	90.5
Ours(150-*MvVW*)	57.05	81.56	89
Ours(200-*MvVW*)	56.6	80.04	88.56
Ours(All-*MvVW*)	49.69	71.56	78.48

6.27%, higher than the visual words without K-means (All-**MvVW**). The result indicates that the original visual words have more redundant information and the **MvVW**, fusing multiview information with K-means, can achieve a better recognition rate. Nonetheless, we should also ensure that the available information is sufficient, so we set $m = 100$ on VIPeR dataset.

5.2.4. Contribution of each region

It is interesting to investigate which region is more effective in our proposed method. At the testing stage, we only use the similarities measurement for a single region and set the similarity scores of other regions to be 0. The CMC curves in Fig. 6 show that the similarity measurement of the whole region evidently outperforms any individual local region. For local similarity measurements, the ones for upper body are more effective than those for lower body. In particular, the measurement of Region2 including the torso achieves better performances with the low rank value.

5.2.5. Effect of parameter selection

In this experiment, we compare the performances with different parameters and describe the method of parameters selection. In our model, the parameters include mainly α, β, γ, η, λ, σ, μ and ρ. We provide the results of our model with different parameters at rank-1 in Fig. 7 where the scale of horizontal ordinate is 10^{-2}, 10^{-1}, 10^{-1}, 10^{-2}, 10^{-2}, 10^{-1}, 10^{-1}, 10^{-0}. As we can see in this figure, our proposed model is insensitive to the setting on these parameters, performing the best with a small change for person Re-ID. In our model, we obtain the optimal parameters through a method of adjusting one parameter while fixing other parameters, and set the values of α, β, γ, η, λ, σ, μ and ρ as 0.07, 0.2,

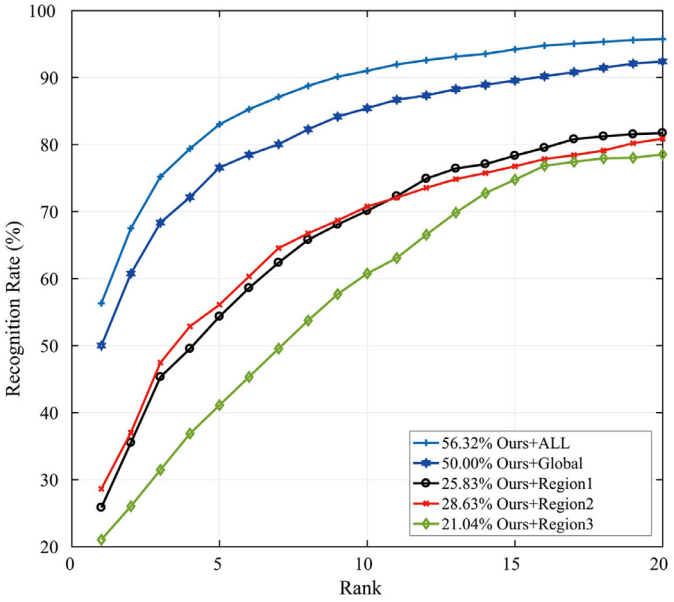

Fig. 6. The CMC curves and rank-1 matching rates on the VIPeR dataset with different regions.

Table 5
The recognition results of our model and other the state-of-the-art methods on CUHK01 dataset at Rank-1, 5, 10, 20.

Method	Rank=1	Rank=5	Rank=10	Rank=20
Ours	68.44	86.24	93.65	96.8
GOG+XQDA [10]	65.33	84.13	90.25	94.61
LOMO+MLAPG [7]	64.74	86.60	91.55	95.40
LOMO+XQDA [6]	63.02	83.33	90.47	94.56
FFN4096+XQDA [19]	39.69	60.05	68.43	75.79
kLFDA [19]	35.91	52.71	61.05	69.77
MFA [19]	35.44	55.10	64.11	72.09
KISSME [1]	30.20	47.66	57.54	68.16
SVMML [12]	31.07	56.04	67.27	78.30
LFDA [11]	34.86	50.91	59.91	68.03

Table 6
The recognition results of our model and other the state-of-the-art methods on PRID450S dataset at Rank-1, 10.

Method	Rank=1	Rank=10
Ours	72.15	94.62
GOG+XQDA [10]	67.9	94.4
LOMO+XQDA [6]	52.3	84.6
SCNCD [22]	41.6	79.4
Semantic [15]	43.1	78.2

0.1, 0.06, 0.06, 0.3, 0.4 and 1.05 when $m = 100$. Note that, if we need fast convergence speed, we can set a larger value for μ.

5.3. Experiments on the CUHK01 dataset

The CUHK-01 dataset was captured from two camera views, with higher resolution, containing 971 persons, and each person has two images in each view. We randomly choose 486 pairs of images for training and leave the rest for testing. And we utilize the K-means method to obtain 4×200 ***MvVW***. The rank-1, rank-5, rank-10 and the rank-20 matching rates are described in Table 5 and the CMC curves are drawn in Fig. 8. As we can see in the Table 5 and Fig. 8, our method outperforms the competing methods, achieving the best rank-1 matching rate of 68.44% with a gain of 3.11%, in comparison with the best result of 65.33% obtained by GOG+XQDA. Similar to the experimental results on the VIPeR dataset, the experimental results on the CUHK01 dataset also show that our method can achieve a better performance on small sample size dataset, which further verifies the robustness of our method.

5.4. Experiments on the PRID450S dataset

The PRID450S dataset contains 450 image pairs recorded from two different static surveillance cameras. In this experiment, we randomly choose 250 pairs of images for training and leave the rest for testing. And we utilize the K-means method to obtain 4×100 ***MvVW***. The rank-1, rank-10 matching rates are reported in Table 6. As we can see in this table, our proposed method achieves 72.15% rank-1 matching rate and 94.62% rank-10 matching rate on the PRID450S dataset, which improves the state-of-the-art rank-1,10 matching rates by 4.15% and 0.22%, respectively. The results also verify the robustness and effectiveness of our method.

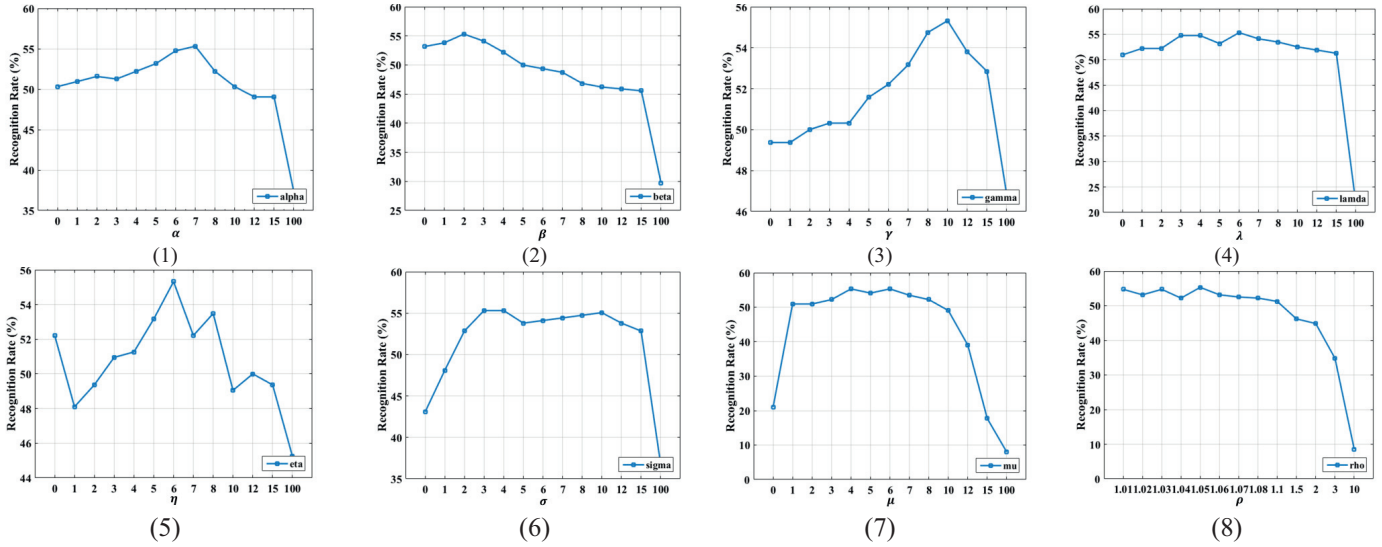

Fig. 7. The CMC curves and rank-1 matching rates with different parameters on the VIPeR dataset. (1) α (2) β (3) γ (4) λ (5) η (6) σ (7) μ (8) ρ.

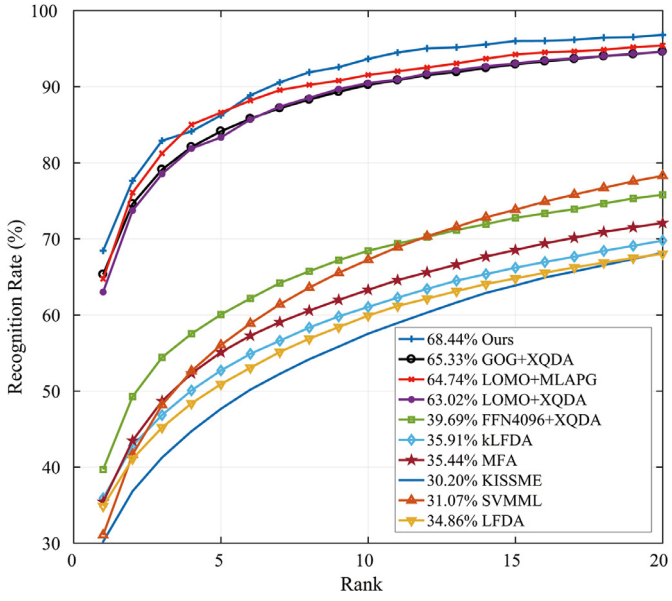

Fig. 8. The CMC curves with different metric learning methods on the CUHK01 dataset.

6. Conclusion

In this paper, we have proposed a novel similarity learning model that formulating the person Re-ID problem as a consistent iterative multi-view joint transfer learning optimal problem, and then solved this optimal problem using IALM algorithm. By adding the transfer, low-rank, and sparse representation constraints, the gap between multi-view images was greatly eliminated and the small sample size problem was effectively alleviated. The experimental results on three challenging person Re-ID benchmark datasets prove that our proposed model achieves state-of-the-art performance and is robust against inconsistent data distributions in terms of viewpoint changes and illumination variations. However, as a major difficulty in person re-identification, the problem of imbalance between positive and negative samples still affect the performance of our method. Besides, for large datasets or more difficult scenes, the features may not be robust. In future, we will study alternative schemes for choosing the proper samples to train the model, and combine with deep learning methods. In addition, we will try to solve the computational complexity problem caused by the dimension of features and blocking strategy.

Acknowledgment

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work was supported by the National Natural Science Foundation of China (NSFC) under grant nos. 61673299, 61203247, 61573259, and 61573255. This work was also supported by the Fundamental Research Funds for the Central Universities and the Open Project Program of the National Laboratory of Pattern Recognition (NLPR).

References

[1] T. Avraham, I. Gurvich, M. Lindenbaum, S. Markovitch, Learning implicit transfer for person re-identification, Eur. Conf. Comput. Vis. (ECCV) (2012) 381–390.
[2] J. Wang, S. Zhou, J. Wang, Q. Hou, Deep ranking model by large adaptive margin learning for person re-identification, Pattern Recognit. 74 (2018) 241–252.
[3] Y.C. Chen, X. Zhu, W.S. Zheng, J.H. Lai, Person reidentification by camera correlation aware feature augmentation, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2) (2018) 392–408.
[4] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, in: Proceedings of the IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), 3, 2007, pp. 1–7.
[5] Z. Jiang, Z. Lin, L.S. Davis, Label consistent k-svd: learning a discriminative dictionary for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2651–2664.
[6] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2197–2206.
[7] S. Liao, S.Z. Li, Efficient psd constrained asymmetric metric learning for person re-identification, in: IEEE International Conference on Computer Vision (ICCV), 2015, pp. 3685–3693.
[8] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[9] Y. Guo, G. Zhao, M. Pietikälnen, Discriminative features for texture description, Pattern Recognit. 45 (10) (2012) 3834–3843.
[10] 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 (CVPR), 2016, pp. 1363–1372.
[11] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3318–3325.
[12] P.M. Roth, M. Hirzer, M. Koestinger, C. Beleznai, H. Bischof, Mahalanobis distance learning for person re-identification, Pers. Re-Identif. (2014) 247–267.
[13] L. Shao, F. Zhu, X. Li, Transfer learning for visual categorization: a survey, IEEE Trans Neural Netw Learn Syst 26 (5) (2015) 1019–1034.
[14] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, S.Z. Li, Embedding deep metric for person re-identification: a study against large variations, in: European Conference on Computer Vision (ECCV), 2016, pp. 732–748.
[15] Z. Shi, T.M. Hospedales, T. Xiang, Transferring a semantic representation for person re-identification and search, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4184–4193.
[16] C. Patruno, R. Marani, G. Cicirelli, E. Stella, T. D'Orazio, People re-identification using skeleton standard posture and color descriptors from RGB-D data, Pattern Recognit. 89 (2019) 77–90.
[17] C. Zhao, K. Chen, Z. Wei, Y. Chen, D. Miao, W. Wang, Multilevel triplet deep learning model for person re-identification, Pattern Recognit. Lett 117 (2019) 161–168.
[18] X. Wang, W.S. Zheng, X. Li, J. Zhang, Cross-scenario transfer person re-identification, IEEE Trans. Circuits Syst. Video Technol. 28 (8) (2015) 1447–1460.
[19] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: European Conference on Computer Vision (ECCV), 2014, pp. 1–16.
[20] Y. Xu, X. Fang, J. Wu, X. Li, D. Zhang, Discriminative transfer subspace learning via low-rank and sparse representation, IEEE Trans. Image Process. 25 (2) (2016) 850–863.
[21] Y. Yang, S. Liao, Z. Lei, S.Z. Li, Large scale similarity learning using similar pairs for person verification, Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[22] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S.Z. Li, Salient color names for person re-identification, in: European Conference on Computer Vision (ECCV), 2014, pp. 536–551.
[23] Z. Zhao, B. Zhao, F. Su, Person re-identification via integrating patch-based metric learning and local salience learning, Pattern Recognit. 75 (2018) 90–98.
[24] W.S. Zheng, S. Gong, T. Xiang, Towards open-world person reidentification by one-shot group-based verification, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2016) 591–606.
[25] S. Bai, X. Bai, Q. Tian, Scalable person re-identification on supervised smoothed manifold, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 6–7.
[26] D. Li, X. Chen, Z. Zhang, K. Huang, Learning deep context-aware features over body and latent parts for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 384–393.
[27] A. Wu, W.S. Zheng, H.X. Yu, S. Gong, J. Lai, RGB-Infrared cross-modality person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 5380–5389.
[28] S. Li, M. Shao, Y. Fu, Person re-identification by cross-view multi-level dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 40 (12) (2018) 2963–2977.
[29] J. Wang, X. Zhu, S. Gong, W. Li, Transferable joint attribute-identity deep learning for unsupervised person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 2275–2284.
[30] L. Wu, Y. Wang, J. Gao, X. Li, Deep adaptive feature embedding with local sample distributions for person re-identification, Pattern Recognit. 73 (2018) 275–288.
[31] J. Lv, W. Chen, Q. Li, C. Yang, Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2018, pp. 7948–7956.
[32] D. Chen, Z. Yuan, B. Chen, N. Zheng, Similarity learning with spatial constraints for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2016, pp. 1268–1277.
[33] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: European Conference on Computer Vision (ECCV), 2008, pp. 262–275.

[34] S. Wu, Y.C. Chen, X. Li, A.C. Wu, J.J. You, W.S. Zheng, An enhanced deep feature representation for person re-identification, in: IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1–8.
[35] G. Lisanti, I. Masi, A. Del Bimbo, Matching people across camera views using kernel canonical correlation analysis, in: Proceedings of the International Conference on Distributed Smart Cameras, 2014, p. 10.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2016, pp. 770–778.
[37] M. Koestinger, 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 (CVPR), 2012, pp. 2288–2295.
[38] C. Zhao, X. Wang, Y. Chen, C. Gao, W. Zuo, D. Miao, Consistent iterative multi-view transfer learning for person re-identification, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV workshop), 2017, pp. 1087–1094.
[39] B. Yang, A.J. Ma, P.C. Yuen, Learning domain-shared group-sparse representation for unsupervised domain adaptation, Pattern Recognit. 81 (2018) 615–632.
[40] J. Zhou, B. Su, Y. Wu, Easy identification from better constraints: multi-Shot person re-identification from reference constraints, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5373–5381.
[41] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, B. Du, Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding, Pattern Recognit. 48 (10) (2015) 3102–3112.
[42] L. Zhang, L. Zhang, B. Du, J. You, D. Tao, Hyperspectral image unsupervised classification by robust manifold matrix factorization, Inf Sci (Ny) 485 (2019) 154–169.
[43] J. Han, K.N. Ngan, M. Li, H. Zhang, Unsupervised extraction of visual attention objects in color images, IEEE Trans. Circuits Syst. Video Technol. 16 (1) (2005) 141–145.
[44] F. Zhang, B. Du, L. Zhang, Saliency-Guided unsupervised feature learning for scene classification, IEEE Trans. Geosci. Remote Sensing 53 (4) (2015) 2175–2184.

Cairong Zhao is currently an associate professor at Tongji University. He received the Ph.D. degree from Nanjing University of Science and Technology, M.S. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, and B.S. degree from Jilin University, in 2011, 2006 and 2003, respectively. He is the author of more than 30 scientific papers in pattern recognition, computer vision and related areas. His research interests include computer vision, pattern recognition and visual surveillance.