

Article

# Sparse and Low-Rank Joint Dictionary Learning for Person Re-Identification

Jun Sun <sup>1,\*</sup> , Lingchen Kong <sup>1</sup> and Biao Qu <sup>2</sup><sup>1</sup> Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, China; lchkong@bjtu.edu.cn<sup>2</sup> Institute of Operations Research, Qufu Normal University, Rizhao 276826, China; qubiao001@163.com

\* Correspondence: junsun2017@bjtu.edu.cn

**Abstract:** In the past decade, the scientific community has become increasingly interested in the re-identification of people. It is still a challenging problem due to its low-quality images; occlusion between objects; and huge changes in lighting, viewpoint and posture (even for the same person). Therefore, we propose a dictionary learning method that divides the appearance characteristics of pedestrians into a shared part, which comprises the similarity between different pedestrians, and a specific part, which reflects unique identity information. In the process of re-identification, by removing the shared part of a pedestrian's visual characteristics and considering the unique part of each person, the ambiguity of the pedestrian's visual characteristics can be reduced. In addition, considering the structural characteristics of the shared dictionary and special dictionary, low-rank,  $l_0$  norm and row sparsity constraints instead of their convex-relaxed forms are introduced into the dictionary learning framework to improve its representation and recognition capabilities. Therefore, we adopt the method of alternating directions to solve it. The experimental results of several commonly used datasets show the effectiveness of our proposed method.

**Keywords:** re-identification; dictionary learning; sparsity constraints; low rank constraints



**Citation:** Sun, J.; Kong, L.; Qu, B. Sparse and Low-Rank Joint Dictionary Learning for Person Re-Identification. *Mathematics* **2022**, *10*, 510. <https://doi.org/10.3390/math10030510>

Academic Editor: Paolo Crippa

Received: 29 December 2021

Accepted: 28 January 2022

Published: 5 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Pedestrian re-identification (re-ID) aims to identify specific pedestrians through cameras at different locations, that is, to establish the correspondence between people at different visual ranges. It is a key task in most surveillance and security applications [1–3], and has attracted increasing attention from the computer vision community [4,5]. However, in a real complex environment, such as different camera resolutions, viewing angles and background changes, lighting changes, occlusions and person pose changes can adversely affect pedestrian recognition, increase the difficulty of successful pedestrian recognition and make it face many technical challenges. Furthermore, there is still a big gap between the current person re-identification technology and practical applications.

This research direction has attracted the attention of a large number of scholars and research institutions. Aiming to solve the problem of person re-identification, the research mainly focuses on the following two aspects: the expression of pedestrian characteristics [6–12] and similarity measurement learning [13–18]. The feature descriptors try to determine how to select visual features with good discrimination and robustness for pedestrian image matching. As high-dimensional visual features usually do not capture the invariant factors under sample variance, a distance metric is introduced into pedestrian re-identification. The main concept of metric learning is that the visual characteristics of different pedestrians should be more separate and the visual characteristics of the same pedestrian under different perspectives should be as similar as possible in the embedded space. Since sparse dictionary learning is a special case of metric learning, it has been successfully applied in computer vision fields such as face recognition [19,20], and is now applied to pedestrian re-recognition.

In 2010, Cong et al. first introduced dictionary learning into person re-identification [21]. They first built a dictionary through a camera, and then the pedestrians of the other camera were represented by the dictionary sparsely and linearly. Khedher et al. [22] used the surf descriptor to extract features from each person's pictures and then constructed a known dictionary using reference SURFs. They used the sparse representation model to learn a coefficient and then determined the identity information. Karanam et al. [23] learned a single view invariant dictionary for different cameras. They also improved the discriminative ability of the dictionary by adding explicit constraints on the sparse codes, which made the Euclidean distance between the coding coefficients of the same pedestrian under different views smaller than that between the sparse coding coefficients of different pedestrians. Jing et al. [24] proposed a novel semi-coupled low-rank discriminant dictionary learning approach for high- and low-resolution images. Karanam et al. [25] proposed a block sparse representation method based on dictionary learning. An et al. [26] used canonical correlation analysis (CCA) to learn a subspace in which the goal is to maximize the correlation between data from different cameras but corresponding to the same people. Then, they jointly learned the dictionaries for each camera view in the CCA subspace. Zhou et al. [27] proposed a novel joint learning framework that unifies representative feature dictionary learning and discriminative metric learning. Xu et al. [28] proposed to separate the images of the same pedestrian observed from different camera views into view-shared components and view-specific components so as to improve the discriminating performance of the learned dictionary. Peng et al. [29] proposed a novel dictionary learning model which divides the dictionary space into three parts corresponding to semantic, latent discriminative and latent background attributes, respectively. Li [30] proposed a discriminative semicoupled projective dictionary learning (DSPDL) model that employs an efficient projection dictionary learning strategy and jointly learns a pair of dictionaries and a mapping function to model the correspondence of cross-view data. Li et al. [31] proposed a person re-ID method to divide a pedestrian's appearance features into different components. They developed a framework for learning a pair of commonality and specificity dictionaries, while introducing a distance constraint to force the particularities of the same person over the specificity dictionary to have the same coding coefficients and the coding coefficients of different pedestrians to have weak correlation. Li et al. [32] considered novel joint fusion and super-resolution framework based on discriminative dictionary learning. They jointly learned two pairs of low-rank and sparse dictionaries and a conversion dictionary, which are used to represent the low-rank and sparse components of low-resolution images, and to reconstruct a high-resolution fused result. However, to accurately characterize the sparsity and low rank, it is suggested to impose the sparsity and low-rank constraints directly instead of using the approximations/regularizations.

In 2014, Li et al. [33] first used deep learning methods for person re-identification research, and since then, an increasing number of researchers have tried to combine deep learning methods with person re-identification research. Deep learning can integrate feature extraction and metric learning into a unified learning framework and is mainly focused on extracting global identity features from pedestrian images. He et al. [34] proposed to use the Spatial Pyramid structure to extract sample features. Huang et al. [35] used a deep neural network to learn different representation features of different parts of pedestrian appearance images, and then calculated the similarity of the corresponding parts of the image. Then, three sub-networks were constructed for each part to learn the differences between images, feature maps and spatial changes, and the results of the three sub-networks were combined. Wu et al. [36] introduced a deep architecture that combines Fisher vectors and deep neural networks to learn a mixture of nonlinear transformations of pedestrian images into a deep space where the data can be linearly separated. Tao et al. [37] utilized Cross-view Quadratic Discriminant Analysis (XQDA) metric learning for person recognition in order to achieve simultaneous spatial localization and feature representation. Compared with images, there are not only spatial dependencies, but also temporal order relationships between frames in video sequences. Reasonable use of the temporal features of videos can reflect the

motion characteristics of pedestrians and improve the recognition accuracy. Therefore, for video-based pedestrian re-identification, the spatiotemporal features of videos are often extracted for recognition. Gao et al. [38] proposed a temporally aligned pooling representation method, which uses the periodic characteristics of walking to divide the video sequence into independent walking cycles, and selects the cycle that best matches the characteristics of the sinusoidal signal to represent the video sequence. Rahmani et al. [39] proposed a deep fully connected neural network by finding the nonlinear transformations of a set of connected views, which learn from 2D projections of the dense trajectories of synthetic 3D human models fitted to real motion capture data. Using the spatiotemporal motion characteristics of human walking, Khan et al. [40] proposed a novel view-invariant gait representation deep fully connected neural network for cross-view gait recognition. However, spatiotemporal features are susceptible to factors such as viewing angle, scale and speed. With the substantial increase in pedestrians, the motion similarity between pedestrians also increases, which greatly reduces the ability to distinguish spatiotemporal features. At the same time, the large number of cameras in large datasets increases the pose differences and motion differences of the same pedestrian. Obviously, these all limit the role of spatiotemporal features in pedestrian re-identification.

In this paper, we propose a new special and shared dictionary learning model with structure characteristic constraints, which has stronger interpretability. We divide the learning dictionary into two parts. One is a shared dictionary, which represents some features shared by all pedestrians in the camera, such as the same background. The other is a special dictionary, which represents the unique characteristics of each pedestrian. Then, only the unique part that represents the identity of the pedestrian is considered in the recognition process, which can reduce the ambiguity caused by some other unnecessary visual feature factors. The main contributions of the paper are summarized as follows:

- (I) The shared dictionary part, whose features are shared by all people, have a strong correlation, so the shared dictionary must be low rank; then, we directly impose the low-rank constraint. Next, we impose a  $l_0$  norm constraint to the special dictionary, which has strong sparsity and contains only information unique to each person.
- (II) In order to better describe the shared information of pedestrians and force the commonality of different pedestrians to have the same coding coefficients in the shared dictionary, we introduce the  $l_{2,0}$  norm constraint on the coding coefficients  $Z_s$ .
- (III) Due to the  $l_0$  norm and low-rank constraints, the dictionary learning model with structure characteristics constraints is highly nonconvex and computationally NP-hard in general; therefore, we adopt the method of alternating directions to solve it. When dealing with each subproblem, we directly deal with the original problems with the  $l_0$  norm and rank constraints instead of their convex relaxed form. Numerical experiments performed on some real datasets show that our method is superior to traditional methods, and even better than some deep learning methods on some datasets.

The rest of this paper is organized as follows. The joint dictionary learning model is presented in Section 2, while Section 3 is devoted to optimization algorithm for the special and shared dictionary learning model and the re-identification process. In Section 4, the computational experiments are reported. Finally, we conclude the paper with future work in Section 5.

## 2. Joint Dictionary Learning Model

We know that the general cameras are fixed in a place, so the picture of each person in the camera contains part of the same elements that do not help in recognition. What is useful is the part of unique features that represents information about each person. Assuming that we have two camera views,  $a$  and  $b$ , and  $Y_l = [y_{l,1}, y_{l,2}, \dots, y_{l,N}] \in R^{m \times N}$  ( $l = a, b$ ) is a set of training samples composed of  $N$  individuals images from  $l$ -th view,  $Y_l$  can be divided into two parts

$$Y_l = Y_s + Y_{l,\mu}.$$

Considering the actual situation, we mainly study the following dictionary learning model for the person re-identification problem

$$\begin{aligned}
 \min_{D_s, D_u, Z_s, Z_{a,u}, Z_{b,u}} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 \\
 & + \lambda_1 \|Z_{a,u}^T Z_{b,u}\|_F^2 + \lambda_2 \|I - Z_{a,u}^T Z_{b,u}\|_F^2 + \lambda_3 \|Z_{a,u} - Z_{b,u}\|_F^2 \\
 \text{s.t.} & \|D_s(i)\|_2^2 \leq 1, \|D_u(i)\|_2^2 \leq 1, \\
 & \text{rank}(D_s) \leq k, \|D_u\|_0 \leq s_2, \|Z_s\|_{2,0} \leq s_1,
 \end{aligned} \tag{1}$$

where  $Z_{a,u}$  and  $Z_{b,u}$  are the coding coefficients of person-specific components under camera views  $a$  and  $b$ ;  $Z_s = [z_{s,1}, z_{s,2}, \dots, z_{s,N}]$  is the coding coefficient of person-shared components under different camera views.  $D_s$  is the dictionary for the person-shared elements, and  $D_u$  is the dictionary for the person-specific elements.  $\lambda_1, \lambda_2, \lambda_3$  are penalty parameters and  $s_1, s_2, k$  are three integers representing the prior information on the upper bounds of the sparsity and the rank, respectively.  $\|D_u\|_0$  is the zero norm of  $D_u$ , representing the number of its nonzero elements.  $\text{rank}(D_s)$  represents the rank of the matrix  $D_s$ .  $\|Z_s\|_{2,0}$  is the zero norm of the rows of the matrix  $Z_s$ , representing the sparsity of the rows of the matrix  $Z_s$ .

We know that different people have different features, so the coding coefficients of different pedestrians should be largely irrelevant. The same pedestrian has a greater similarity under different cameras, i.e., one person under different views should have the same coding coefficient.  $\text{cor}(Z_{a,u}(i), Z_{b,u}(j)) \leq \varepsilon_1, i \neq j, \text{cor}(Z_{a,u}(i), Z_{b,u}(i)) \geq \varepsilon_2$ , where  $0 < \varepsilon_1 < \varepsilon_2 \leq 1$  are the given correlation parameters, and  $\text{cor}$  is the correlation function. However, the correlation coefficient is more difficult to calculate; similar to article [19], we transformed the correlation coefficient constraint into the following form:  $\|Z_{a,u}^T Z_{b,u}\|_F^2, \|I - Z_{a,u}^T Z_{b,u}\|_F^2$ . The same elements play the same role for each pedestrian under the camera, and the shared features are only a small part of all features. So,  $\|Z_s\|_{2,0} \leq s_1$  was added. Generally, the common part has a strong correlation. For example, two cameras may have a part of the same background, and the picture background often has a low-rank structure. So, the shared dictionary should have a low-rank structure. At the same time, the unique information for each pedestrian is different, so it should have a sparse structure, i.e.,  $\text{rank}(D_s) \leq k, \|D_u\|_0 \leq s_2$ . The identity information of the same pedestrian under different cameras is the same and should be as similar as possible. Therefore, the same pedestrian should have the same coefficient under different cameras; that is,  $Z_{a,u} - Z_{b,u} = 0$ .

### 3. Algorithm Implementation

In this section, we describe the algorithm of dictionary learning and the process of re-identification.

#### 3.1. Dictionary Learning Algorithm

First, we propose the algorithm for the problem (1). Due to the  $l_0$  norm, low-rank constraints and the properties of the objective function of dictionary learning, we adopted the alternating directions scheme and optimized one variable at a time by fixing the other variables.

(1) *Update  $D_s$ .* With fixed  $D_u, Z_s, Z_{a,u}, Z_{b,u}$ , we solved the following minimization problem:

$$\begin{aligned}
 \min_{D_s} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 \\
 \text{s.t.} & \|D_s(i)\|_2^2 \leq 1, \text{rank}(D_s) \leq k.
 \end{aligned} \tag{2}$$

To facilitate optimization, a relaxation variable  $\widetilde{D}_s$  was introduced to simplify the solving process. Then, the optimization problem in (2) can be written as follows:

$$\begin{aligned} \min_{D_s, \widetilde{D}_s} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 + \|D_s - \widetilde{D}_s\|_F^2 \\ \text{s.t.} & \quad \|\widetilde{D}_s(i)\|_2^2 \leq 1, \text{rank}(D_s) \leq k. \end{aligned} \tag{3}$$

First, we fixed  $D_s$  and updated  $\widetilde{D}_s$ . This problem can be optimized by the Lagrange dual [41].

$$\begin{aligned} \min_{\widetilde{D}_s} & \|D_s - \widetilde{D}_s\|_F^2 \\ \text{s.t.} & \quad \|\widetilde{D}_s(i)\|_2^2 \leq 1. \end{aligned} \tag{4}$$

Then, we fixed  $\widetilde{D}_s$ , and updated  $D_s$  by solving:

$$\begin{aligned} \min_{D_s} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 + \|D_s - \widetilde{D}_s\|_F^2 \\ \text{s.t.} & \quad \text{rank}(D_s) \leq k. \end{aligned} \tag{5}$$

The above problem can be solved by [42].

(2) Update  $D_u$ . With fixed  $D_s, Z_s, Z_{a,u}, Z_{b,u}$ , we solved the following minimization problem:

$$\begin{aligned} \min_{D_u} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 \\ \text{s.t.} & \quad \|D_u(i)\|_2^2 \leq 1, \|D_u\|_0 \leq s_2. \end{aligned} \tag{6}$$

For easier calculation, the relaxation variable  $\widetilde{D}_u$  was introduced to simplify the solving process. The optimization problem in (6) can be written as follows:

$$\begin{aligned} \min_{D_u, \widetilde{D}_u} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 + \|D_u - \widetilde{D}_u\|_F^2 \\ \text{s.t.} & \quad \|\widetilde{D}_u(i)\|_2^2 \leq 1, \|D_u\|_0 \leq s_2. \end{aligned} \tag{7}$$

We fixed  $D_u$ , and updated  $\widetilde{D}_u$  by solving. It can be optimized in the same way as in (4)

$$\begin{aligned} \min_{\widetilde{D}_u} & \|D_u - \widetilde{D}_u\|_F^2 \\ \text{s.t.} & \quad \|\widetilde{D}_u(i)\|_2^2 \leq 1. \end{aligned} \tag{8}$$

With the updated  $\widetilde{D}_u$ , we can update  $D_u$  by solving

$$\begin{aligned} \min_{D_u} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 + \|D_u - \widetilde{D}_u\|_F^2 \\ \text{s.t.} & \quad \|D_u\|_0 \leq s_2. \end{aligned} \tag{9}$$

This problem can be solved via the gradient support projection algorithm (GSPA) [43].

(3) Update  $Z_s$ . With fixed  $D_s, D_u, Z_{b,u}, Z_{b,u}$ , we solved the following minimization problem:

$$\begin{aligned} \min_{Z_s} & \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 \\ \text{s.t.} & \quad \|Z_s\|_{2,0} \leq s_1. \end{aligned} \tag{10}$$

Since the  $L_{2,0}$  norm can be regarded as a special  $L_0$  norm, it can be solved similar to in (9).

(4) *Update  $Z_{a,u}$ .* With fixed  $D_s, D_u, Z_s, Z_{b,u}$ , we solved the following minimization problem:

$$\min_{Z_{a,u}} \|Y_a - D_s Z_s - D_u Z_{a,u}\|_F^2 + \lambda_1 \|Z_{a,u}^T Z_{b,u}\|_F^2 + \lambda_2 \|I - Z_{a,u}^T Z_{b,u}\|_F^2 + \lambda_3 \|Z_{a,u} - Z_{b,u}\|_F^2. \tag{11}$$

This is a smooth convex problem; hence, we could obtain the following closed-form solution:

$$Z_{a,u} = \left( D_u^T D_u + 2\lambda_2 Z_{b,u} Z_{b,u}^T + \lambda_3 I \right)^{-1} \left( D_u^T Y_a + \lambda_2 Z_{b,u} I^T - D_u^T D_s Z_s + \lambda_3 Z_{b,u} \right). \tag{12}$$

(5) *Update  $Z_{b,u}$ .* With fixed  $D_s, D_u, Z_s, Z_{a,u}$ , we solved the following minimization problem:

$$\min_{Z_{b,u}} \|Y_b - D_s Z_s - D_u Z_{b,u}\|_F^2 + \lambda_1 \|Z_{a,u}^T Z_{b,u}\|_F^2 + \lambda_2 \|I - Z_{a,u}^T Z_{b,u}\|_F^2 + \lambda_3 \|Z_{a,u} - Z_{b,u}\|_F^2. \tag{13}$$

This is a smooth convex problem, so that the closed-form solution is as follows:

$$Z_{b,u} = \left( D_u^T D_u + 2\lambda_2 Z_{a,u} Z_{a,u}^T + \lambda_3 I \right)^{-1} \left( D_u^T Y_b + \lambda_2 Z_{a,u} I^T - D_u^T D_s Z_s + \lambda_3 Z_{a,u} \right). \tag{14}$$

A detailed description of the above learning is summarized in the following Algorithm 1.

---

**Algorithm 1** Dictionary Learning Algorithm

---

**Input:** samples matrices  $Y_a, Y_b$  and initial point  $D_s, D_u, Z_s, Z_{a,u}$  and  $Z_{b,u}$  randomly.

- 1: **while** not converged **do**
- 2:   Fix other variables, update  $Z_s, Z_{a,u}$  and  $Z_{b,u}$  via solving (10), (11) and (13), respectively.
- 3:   Update dictionary  $D_s$  by solving (3).
- 4:   Update dictionary  $D_u$  by solving (7).
- 5: **end while**

**Output:** dictionaries  $D_u$  and  $D_s$ .

---

3.2. *Re-Identification*

Given the gallery feature vector matrices and the learned dictionaries  $D_s, D_u$ , we propose the following steps to re-identify a person:

1. The coding coefficients of the shared components of pedestrians can be obtained by:

$$Z_{l,s} = \arg \min_{Z_{l,s}} \left\{ \|Y_l - D_s Z_{l,s} - D_u Z_{l,u}\|_F^2 + \lambda_4 \|Z_{l,s}\|_{2,1} \right\}, \tag{15}$$

where  $\lambda_4$  is the scalar parameter;  $Y_l = [Y_a, Y_b]$ ,  $Z_{l,s} = [Z_{a,s}, Z_{b,s}]$  and  $Z_{l,u} = [Z_{a,u}, Z_{b,u}]$ ;  $Z_{l,s}$  denotes the dictionary coefficient corresponding to the part shared by the pedestrians under different cameras;  $Z_{l,u}$  denotes the dictionary coefficient corresponding to the part unique to each person under different cameras.

2. The dictionary coefficient corresponding to the part unique of each person under different cameras can be obtained as follows:

$$Z_{a,u} = \arg \min_{Z_{a,u}} \left\{ \|Y_a - D_s Z_{a,s} - D_u Z_{a,u}\|_F^2 + \beta_1 \|Z_{a,u}\|_1 \right\}, \tag{16}$$

$$Z_{b,u} = \arg \min_{Z_{b,u}} \left\{ \|Y_b - D_s Z_{b,s} - D_u Z_{b,u}\|_F^2 + \beta_1 \|Z_{b,u}\|_1 \right\}, \quad (17)$$

where  $\beta_1$  is the scalar parameter.

3. We took  $z_{a,u,i}$  and  $z_{b,u,j}$  to be the dictionary coefficients corresponding to the individual special parts of the  $i$ -th pedestrian and the  $j$ -th pedestrian under cameras  $a$  and  $b$ , respectively. Then, we computed the Euclidean distance between  $z_{a,u,i}$  and  $z_{b,u,j}$  for personal identification matching

$$\text{sim}(z_{a,u,i}, z_{b,u,j}) = \|z_{a,u,i} - z_{b,u,j}\|_2^2, \quad (18)$$

where  $i = 1, 2, \dots, N$  and  $j = 1, 2, \dots, M$ .

## 4. Numerical Experiments

### 4.1. Datasets

In this section, we evaluate the proposed method. All the codes were written in MATLAB, and all the computations were performed on a LENONVE ideapad with Windows 10 Inter(R) Core(TM)i5-6200U CPU @2.30 GHz, 2.40 GHz and 4 GB memory. We empirically validated the method proposed in this paper using five publicly available multi-shot re-ID datasets:

VIPeR [44]: the VIPeR dataset contains 632 pedestrians; each pedestrian captures one image from each of the two cameras; it contains two images of each pedestrian.

PRID 2011 [45]: The PRID 2011 dataset contains images of 200 people. These images were taken with two non-overlapping cameras in an uncrowded outdoor environment with significant point-of-view and lighting variations.

QMUL-GRID [46]: The GRID dataset contains 250 images of pedestrians. Each pedestrian image contains two images seen from different camera views, both of which come from eight non-intersecting camera views installed in busy subway stations. The dataset is challenging due to variations in pose, color and lighting, as well as poor image quality due to low spatial resolution.

CUHK01 [47]: The CUHK01 dataset contains contains 971 photos of pedestrians. Each pedestrian consists of two images from a pair of disjoint cameras. The image quality is relatively good.

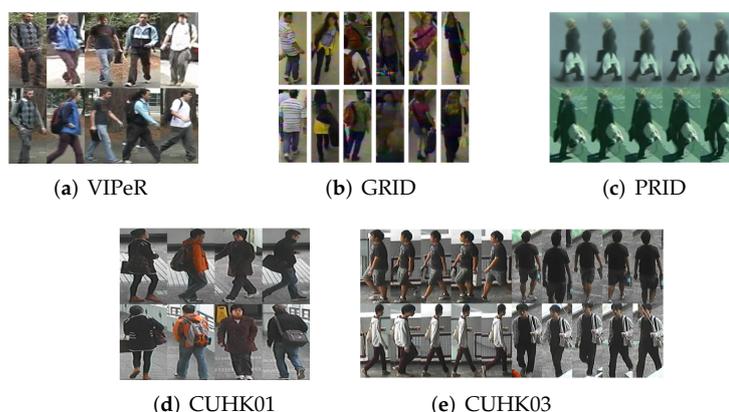
CUHK03 [48]: CUHK03 is one of the largest personal identification datasets, and it contains 1467 pedestrians. It provides two types of data, one obtained from manually labeled pedestrian bounding boxes and the other from automatically detected bounding boxes. The detected CUHK03 is more challenging than the labeled CUHK03 dataset due to incorrectly detected bounding boxes.

In our experiments, we randomly divided VIPeR, PRID 2011, QMUL-GRID, CUHK01 and CUHK03 into two parts, one as the training data and the other as the testing data. More details are given in Table 1. For the experiments on each dataset, the above procedure was repeated 10 times. The average of the 10 measurements was considered as the final experimental result.

**Table 1.** The details and settings of person re-ID datasets.

Dataset	VIPeR	PRID 2011	QMUL-GRID	CUHK01	CUHK03
Cams	2	2	8	6	6
IDs	632	200	250	971	1467
TrainIDs	316	100	125	871	1367
TestIDs	316	100	125	100	100
Interfering Img	0	549	775	0	0
Labeled	1264	949	1275	1942	14,096
Detected	0	0	0	0	14,096

Some pedestrian image pairs selected from the five datasets are presented in Figure 1.



**Figure 1.** Sample images.

**Feature Selection:** In our experiment, we used the GOG feature method proposed by Matsukawa et al. [11] to deal with the original image feature. GOG feature describes local regions in an image through a hierarchical Gaussian distribution and shows strong robustness against changes in pedestrian body pose, illumination, background clutter, and picture quality.

#### 4.2. Experiment on VIPeR

To illustrate the effectiveness of our proposed model as well as the method, several alternative state-of-the-art re-ID methods were selected: KISSME (2012) [16], DGD (2016) [49], KCVDCa (2017) [50], JDL (2017) [28], JLML (2017) [48], MVLDMML (2018) [51], MPML (2019) [52], DIMN (2019) [53], DLA (2020) [31] and VS-SSL (2020) [54]. The numerical results of the above methods on dataset VIPeR are reported in Table 2. From Table 2, the matching rates of our method on ranks 1, 5 and 10 are 51.23%, 80.73% and 90.56%, respectively, which are higher than those of the other methods. Although the recognition accuracy of our method in rank 1 is the same as that of DIMN method, the recognition accuracy of our method in ranks 2–5 is higher than that of the DIMN method, meaning our method is significantly better than DIMN in rank 5, which also shows the superior performance of our method. The results show that the recognition accuracy of our approach on the VIPeR dataset is better than that of some other existing methods.

**Table 2.** Comparison of top-ranked matching rates (%) with the state-of-the-art methods on the VIPeR dataset. Bold number represents the best accuracy. ‘-’ denotes that there is no reported result.

Methods	Rank 1	Rank 5	Rank 10	Rank 20
KISSME [16]	19.60	48.00	62.20	77.00
DGD [49]	38.60	-	-	-
KCVDCa [50]	43.29	72.66	83.51	92.18
JDL [28]	44.30	72.66	82.50	92.90
JLML [48]	50.20	74.20	84.30	91.60
MVLDMML [51]	50.03	79.20	88.54	94.65
MPML [52]	50.54	78.16	87.63	95.00
DIMN [53]	<b>51.23</b>	70.19	75.98	-
DLA [31]	50.89	80.03	89.33	94.87
VS-SSL [54]	44.80	72.30	79.30	86.10
Ours	<b>51.23</b>	<b>80.73</b>	<b>90.56</b>	<b>95.02</b>

#### 4.3. Experiment on PRID 2011

The third experiment was conducted on the dataset PRID 2011. Similarly, we also compared it with some of the most advanced methods, including PPLM (2012) [55], XQDA+LOMO (2015) [12], XQDA+GOG (2016) [11], M+DMLV (2017) [56], DMLV+LOMO (2017) [57], JDL (2017) [28], MTL-LOREA (2018) [58], APDL (2018) [59], DLA (2020) [31] and VS-SSL (2020) [54]. The experimental results of the approaches are reported Table 3. As shown in the table, the recognition rates of the proposed method are 36.80%, 63.80%, 73.20% and 83.30% at different ranks, respectively. This further proves that the proposed approach has a better performance on the dataset with interfering images than that of other traditional approaches.

**Table 3.** Comparison of top-ranked matching rates (%) with the state-of-the-art methods on the PRID 2011 dataset. Bold number represents the best accuracy. ‘-’ denotes that there is no reported result.

Methods	Rank 1	Rank 5	Rank 10	Rank 20
PPLM [55]	15.00	32.00	42.00	54.00
XQDA+LOMO [12]	26.70	49.90	61.90	73.80
XQDA+GOG [11]	35.90	60.10	68.50	78.10
M+DMLV [56]	15.20	36.10	48.30	-
DMLV+LOMO [57]	27.80	48.40	59.50	83.20
JDL [28]	26.50	53.60	63.20	73.00
MTL-LOREA [58]	18.00	37.40	50.10	66.60
APDL [59]	25.00	54.00	67.00	82.00
DLA [31]	36.50	63.00	72.90	82.70
VS-SSL [54]	35.40	57.80	66.30	77.90
Ours	<b>36.80</b>	<b>63.80</b>	<b>73.20</b>	<b>83.30</b>

#### 4.4. Experiment on QMUL-GRID

The second experiment was conducted on dataset QMUL-GRID to illustrate the effectiveness of our proposed method, and several alternative state-of-the-art re-ID methods were selected: MtMCML (2014) [60], LSSCDL (2016) [61], DR-KISS (2016) [11], Multi-HG (2017) [62], SCRWI (2017) [63], JLML (deep-learning) (2017) [48], CSPL+GOG (2018) [64], MPML (2019) [52], SRRTC (2019) [65], DIMN (2019) [53], DLA (2020) [31] and KISS+ (2021) [66]. The numerical results of the above methods for dataset QMUL-GRID are reported in Table 4. As shown in Table 4, the deep learning method JLML has the best performance in the recognition rate of rank 1 and rank 5, but the KISS+ method has the highest accuracy in the accuracy rates of rank 10 and rank 20. Our proposed method is second only to these two methods. This is also due to the fact that the GRID dataset is relatively small, and we only used 125 markers to train our model, while JLML first pre-trained on large-scale ImageNet [67], Market-1501 [68] and CUHK03, and then selected 125 identifiers from QMUL-GRID to fine-tune the pre-trained model. The KISS+ method used an orthogonal basis vector to generate virtual samples to deal with the small sample size problem. Our proposed approach has better performance than most of the other traditional non-deep learning methods, which proves the strong applicability of the proposed method.

**Table 4.** Comparison of top-ranked matching rates (%) with the state-of-the-art methods on the QMUL-GRID dataset. Bold number represents the best accuracy. ‘-’ denotes that there is no reported result.

Methods	Rank 1	Rank 5	Rank 10	Rank 20
MtMCML [60]	14.08	34.64	45.84	59.84
LSSCDL [61]	22.40	-	51.28	61.20
DR-KISS [11]	20.60	39.30	51.40	62.60
Multi-HG [62]	19.84	40.48	56.88	62.32
SCRWI [63]	24.80	45.40	54.10	68.90
JLML [48]	<b>37.50</b>	<b>61.40</b>	69.40	77.40
CSPL+GOG [64]	25.84	46.48	58.40	70.16
MPML [52]	24.32	47.92	58.40	68.08
SRRTC [65]	26.56	46.32	56.16	66.80
DIMN [53]	29.28	53.28	65.84	-
DLA [31]	28.16	49.04	59.52	70.48
KISS+ [66]	33.58	58.58	<b>79.37</b>	<b>82.55</b>
Ours	33.46	56.64	68.13	75.44

#### 4.5. Experiment on CUHK01

The fourth experiment was conducted on the dataset PRID2011. Similarly, the proposed approach was compared with other excellent methods, including KISSME (2012) [16], SalMatch (2013) [69], Mid-Filter (2014) [70], XQDA+LOMO (2015) [12], XQDA+GOG (2016) [11], JLML (2017) [48], LADF+DMLV (2017) [56], Multi-HG (2017) [62], DMLV+LOMO (2017) [57], MVLDML (2018) [51], MPML (2019) [52], DLA (2020) [31] and MLAPG (2020) [71]. The experimental results of different methods on ranks 1, 5, 10 and 20 are reported in Table 5. Similar to the numerical results of the QMUL-GRID dataset, our method is second only to the deep learning method JLML, while outperforming the other traditional methods, as JLML first pre-trained on large-scale ImageNet, Market-1501 and CUHK03, and then selected part of the identifiers from CUHK01 to fine-tune the pre-trained model. On rank 1, the recognition rate of our proposed method achieved 72.36%, which is 5% higher than MPML, and on rank 5, it reached 88.43%, which is a 2.7% improvement over DLA. On rank 10 and rank 20, the recognition accuracy of MLAPG and MVLDML improved from 89.50% and 95.85% to 92.43% and 96.06%, respectively.

**Table 5.** Comparison of top-ranked matching rates (%) with the state-of-the-art methods on the CUHK01 dataset. Bold number represents the best accuracy. ‘-’ denotes that there is no reported result.

Methods	Rank 1	Rank 5	Rank 10	Rank 20
KISSME [16]	10.30	27.20	37.50	49.70
SalMatch [69]	28.50	45.90	55.70	68.00
Mid-Filter [70]	34.30	55.10	65.00	74.90
XQDA+LOMO [12]	63.20	83.90	90.00	94.40
XQDA+GOG [11]	57.80	79.10	86.20	92.10
JLML [48]	<b>76.70</b>	<b>92.60</b>	<b>95.60</b>	<b>98.10</b>
LADF+DMLV [56]	63.20	83.90	90.00	94.40
Multi-HG [62]	64.37	-	90.56	94.56
DMLV+LOMO [57]	65.00	85.60	91.10	95.10
MVLDML [51]	61.37	82.74	88.88	93.85
MPML [52]	67.18	86.89	92.06	95.85
DLA [31]	70.19	85.78	90.87	94.57
MLAPG [71]	57.09	85.58	89.50	-
Ours	72.36	88.43	92.43	96.06

#### 4.6. Experiment on CUHK03

The last experiment was performed on the labeled CUHK03 and the detected CUHK03 datasets. To illustrate the effectiveness of our method, we compared the performance of our approach with some state-of-the-art methods, specifically, XQDA+LOMO (2015) [12], XQDA+GOG (2016) [11], DGD (2016) [49], BTloss (2017) [72], JLML (2017) [48], MGN (2018) [73], JSLA (2018) [29], MLFN (2018) [74], DDPM (2018) [75], RAPMR (2018) [76], GLAD (2019) [77], PLNET (2019) [78], SAN (2019) [79], DLA (2020) [31], Deep-Person (2020) [80] and Gconv (2020) [81]. For the labeled and detected CUHK03, the experimental results of different methods are listed in Tables 6 and 7.

From Table 6, we can see that on the labeled CUHK03, our proposed method is second only to the deep learning method Deep-Person on rank 1, rank 5 and rank 10, and second only to the deep learning method JLML on rank 20, and it performs better than some existing traditional methods, as well as deep learning methods, as the Deep-Person method applied long short-term memory (LSTM) in an end-to-end fashion to model pedestrians, treating it as a head-to-toe sequence of body parts. It exploited the complementary information between local and global features to better align with the whole person. From Table 7, we can see that on the detected CUHK03, our proposed method achieved the best recognition accuracy on all ranks, which is better than some existing traditional and deep learning methods, such as GLAD, Deep-Person and Gconv. This proves the effectiveness and competitiveness of our method.

**Table 6.** Comparison of top-ranked matching rates (%) with the state-of-the-art methods on the labeled CUHK03 dataset. Bold number represents the best accuracy. ‘-’ denotes that there is no reported result.

Methods	Rank 1	Rank 5	Rank 10	Rank 20
XQDA+LOMO [12]	52.20	82.20	92.10	96.30
XQDA+GOG [11]	67.30	91.00	96.00	-
DGD [49]	75.30	-	-	-
BTloss [72]	75.53	95.15	99.16	-
JLML [48]	83.20	98.00	99.40	<b>99.80</b>
MGN [73]	68.80	-	-	-
JSLA [29]	77.50	92.40	96.50	99.20
MLFN [74]	54.70			
GLAD [77]	86.00	98.10	99.20	99.70
PLNET [78]	82.75	96.59	98.60	-
SAN [79]	88.30	-	-	-
DLA [31]	90.26	97.98	99.17	99.72
Deep-Person [80]	<b>91.50</b>	<b>99.00</b>	<b>99.50</b>	-
Gconv [81]	85.90	98.30	99.30	99.70
Ours	90.31	98.20	99.46	99.78

From the above experiments on different datasets, it can be seen that our method is the best on some datasets and outperforms the existing traditional methods. Although not the best on some datasets, it is only second to one or two deep learning methods. In conclusion, our proposed method performs well overall.

**Table 7.** Comparison of top-ranked matching rates (%) with the state-of-the-art methods on the detected CUHK03 dataset. Bold number represents the best accuracy. ‘-’ denotes that there is no reported result.

Methods	Rank 1	Rank 5	Rank 10	Rank 20
XQDA+LOMO [12]	46.30	78.90	83.50	93.20
XQDA+GOG [11]	65.50	88.40	93.70	-
JLML [48]	80.60	96.90	98.70	99.20
MGN [73]	66.80	-	-	-
JSLA [29]	64.20	89.10	93.40	96.10
MLFN [74]	82.80	-	-	-
DDPM [75]	75.90	-	-	-
RAPMR [76]	70.60	-	-	-
GLAD [77]	83.30	96.10	97.70	98.80
SAN [79]	84.30	-	-	-
DLA [31]	87.62	97.32	99.46	99.65
Deep-Person [80]	89.40	98.20	99.10	-
Gconv [81]	83.10	96.40	98.00	98.90
Ours	<b>88.42</b>	<b>98.33</b>	<b>99.48</b>	<b>99.66</b>

## 5. Conclusions and Discussion

In this paper, we propose a new special and shared dictionary learning model with structure characteristic constraints, including sparse, low-rank and row-sparse constraints. Here, we divided the dictionary into two, one to represent features shared by all pedestrians and the other to represent features unique to each individual. Then, only the unique part that represents the identity of the pedestrian was considered in the recognition process, which can reduce the ambiguity caused by some other unnecessary visual feature factors in the recognition process. In order to improve the accuracy of matching and to better characterize the structural characteristics of the dictionary, on the basis of original dictionary learning,  $l_0$  norm and low-rank constraints were directly added instead of their convex regular form. We used the method of alternating directions to solve the optimization model, and when solving each sub-problem, we also directly solved the problem with constraints. Finally, experiments on different datasets showed that our algorithm has a high accuracy rate.

Since the objective function of dictionary learning, as well as the highly nonconvex and computationally NP-hard of  $l_0$  norm and rank constraints, is bilinear, the optimality condition of the model and the convergence of the algorithm were not established here. It is difficult to not only explore the impact of feature extraction methods, but also compare the performance of metric learning methods. So our research focuses on the comparison of metric learning methods. We think that evaluating the GOG features using view transformation model (VTM) based approaches is a good attempt. These will be addressed in our future work.

**Author Contributions:** Data collection and analysis, J.S.; validation, J.S., L.K. and B.Q.; writing—original draft, J.S., L.K. and B.Q.; writing—review and editing, J.S., L.K. and B.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Natural Science Foundation of China (12071022) and the Natural Science Foundation of Shandong Province (ZR2018MA019).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available from <http://robustsystems.coe.neu.edu/sites/robustsystems.coe.neu.edu/files/systems/projectpages/reiddataset.html#panel-element-549143>, accessed on 26 December 2021.

**Acknowledgments:** The authors would like to thank the Associate Editor and the anonymous referee for their helpful suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Camps, O.; Gou, M.; Hebble, T.; Karanam, S.; Lehmann, O.; Li, Y.; Radke, R.J. From the lab to the real world: Re-identification in an airport camera Network. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 540–553. [[CrossRef](#)]
2. Li, Y.; Wu, Z.; Karanam, S.; Radke, R. Real-world re-identification in an airport camera network. In Proceedings of the International Conference on Distributed Smart Cameras, Venezia, Italy, 4–7 November 2014; pp. 1–6.
3. Tian, B.; Morris, B.T.; Tang, M.; Liu, Y.; Yao, Y.; Gou, C.; Shen, D.; Tang, S. Hierarchical and networked vehicle surveillance in its: A survey. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 25–48. [[CrossRef](#)]
4. Olszewska, J.I. Automated face recognition: Challenges and solutions. In *Pattern Recognition: Analysis and Applications*; IntechOpen: London, UK, 2016; pp. 59–79.
5. Pang, Y.; Cao, J.; Wang, J.; Han, J. JCS-net: Joint classification and super-resolution network for smallscale pedestrian detection in surveillance images. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 3322–3331. [[CrossRef](#)]
6. Patruno, C.; Marani, R.; Cicirelli, G.; Stella, E.; D’Orazio, T. People re-identification using skeleton standard posture and color descriptors from RGB-D data. *Pattern Recognit.* **2019**, *89*, 77–90. [[CrossRef](#)]
7. Bedagkar-Gala, A.; Shah, S.K. A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **2014**, *32*, 270–286. [[CrossRef](#)]
8. Leng, Q.; Ye, M.; Tian, Q. A survey of open-world person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 1092–1108. [[CrossRef](#)]
9. Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. Person re-identification by symmetry-driven accumulation of local features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2360–2367.
10. Ma, B.; Su, Y.; Jurie, F. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image Vis. Comput.* **2014**, *32*, 379–390. [[CrossRef](#)]
11. Matsukawa, T.; Okabe, T.; Suzuki, E.; Sato, Y. Hierarchical Gaussian descriptor for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1363–1372.
12. Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 2197–2206.
13. Zou, G.; Fu, G.; Peng, X.; Liu, Y.; Gao, M.; Liu, Z. Person re-identification based on metric learning: A survey. *Multimed. Tools Appl.* **2021**, *80*, 26855–26888. [[CrossRef](#)]
14. Yang, Y.; Liao, S.; Lei, Z.; Li, S.Z. Large scale similarity learning using similar pairs for person verification. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 3655–3661.
15. Liao, S.; Li, S.Z. Efficient psd constrained asymmetric metric learning for person re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2016; pp. 3685–3693.
16. Kostinger, M.; Hirzer, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Large scale metric learning from equivalence constraints. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2288–2295.
17. Davis, J.V.; Kulis, B.; Jain, P.; Sra, S.; Dhillon, I.S. Informationtheoretic metric learning. In Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, USA, 20–24 June 2007; pp. 209–216.
18. Chen, Y.C.; Zhu, X.; Zheng, W.S.; Lai, J.H. Person re-identification by camera correlation aware feature augmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 392–408. [[CrossRef](#)]
19. Chen, C.F.; Wei, C.P.; Wang, Y.C. Low-rank matrix recovery with structural incoherence for robust face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2619–2625.
20. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Projective dictionary pair learning for pattern classification. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 793–801.
21. Cong, D.; Achard, C.; Khoudour, L. People re-identification by classification of silhouettes based on sparse representation. In Proceedings of the IEEE 2010 2nd International Conference on Image Processing Theory Tools and Applications, Paris, France, 7–10 July 2010; pp. 60–65.
22. Khedher, M.I.; Mounîm, A.E.; Dorizzi, B. Multi-shot SURF-based person re-identification via sparse representation. In Proceedings of the 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, Krakow, Poland, 27–30 August 2013; pp. 159–164.
23. Karanam, S.; Li, Y.; Radke, R.J. Person re-identification with discriminatively trained viewpoint invariant dictionaries. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 4516–4524.

24. Jing, X.Y.; Zhu, X.; Wu, F.; You, X.; Liu, Q.; Yue, D.; Hu, R.; Xu, B. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 December 2015; pp. 695–704.
25. Karanam, S.; Li, Y.; Radke, R.J. Person re-identification with block sparse recovery. *Image Vis. Comput.* **2016**, *60*, 75–90. [[CrossRef](#)]
26. An, L.; Chen, X.; Yang, S.; Bhanu, B. Sparse representation matching for person re-identification. *Inf. Sci.* **2016**, *355–356*, 74–89. [[CrossRef](#)]
27. Zhou, Q.; Zheng, S.; Ling, H.; Su, H.; Wu, S. Joint dictionary and metric learning for person re-identification. *Pattern Recognit.* **2017**, *72*, 196–206. [[CrossRef](#)]
28. Xu, Y.; Guo, J.; Huang, Z. Joint dictionary learning for person re-identification. In Proceedings of the 2017 IEEE Second International Conference on Data Science in Cyberspace, Shenzhen, China, 26–29 June 2017; pp. 505–510.
29. Peng, P.; Tian, Y.; Xiang, T.; Wang, Y.; Pontil, M.; Huang, T. Joint semantic and latent attribute modelling for cross-class transfer learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1625–1638. [[CrossRef](#)]
30. Li, K.; Ding, Z.; Li, S.; Fu, Y. Toward resolution-invariant person reidentification via projective dictionary learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 1896–1907. [[CrossRef](#)]
31. Li, H.; Xu, J.; Yu, Z.; Luo, J. Jointly learning commonality and specificity dictionaries for person re-identification. *IEEE Trans. Image Process.* **2020**, *29*, 7345–7358. [[CrossRef](#)]
32. Li, H.; Yang, M.; Yu, Z. Joint image fusion and super-resolution for enhanced visualization via semi-coupled discriminative dictionary learning and advantage embedding. *Neurocomputing* **2021**, *422*, 62–84. [[CrossRef](#)]
33. Li, W.; Zhao, R.; Xiao, T.; Wang, X. DeepReID: Deep filter pairing neural network for person re-identification. In Proceedings of the IEEE International conference of Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 152–159.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
35. Huang, Y.; Sheng, H.; Zheng, Y.; Zhang, X. DeepDiff: Learning deep difference features on human body parts for person re-identification. *Neurocomputing* **2017**, *241*, 191–203. [[CrossRef](#)]
36. Wu, L.; Shen, C.; Van Den Hengel, A. Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *Pattern Recognit.* **2017**, *65*, 238–250. [[CrossRef](#)]
37. Tao, D.; Guo, Y.; Yu, B.; Pang, J.; Yu, Z. Deep multi-view feature learning for person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2657–2666. [[CrossRef](#)]
38. Gao, C.; Wang, J.; Liu, L.; Yu, J.; Sang, N. Temporally aligned pooling representation for video-based person re-identification. In Proceedings of the IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016; pp. 4284–4288.
39. Rahmani, H.; Mian, A.; Shah, M. Learning a deep model for human action recognition from novel viewpoints. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 667–681. [[CrossRef](#)]
40. Khan, M.; Farid, M.; Grzegorzec, M. A non-linear view transformations model for cross-view gait recognition. *Neurocomputing* **2020**, *402*, 100–111. [[CrossRef](#)]
41. Lee, H.; Battle, A.; Raina, R.; Ng, A. Efficient sparse coding algorithms. In *Proceedings of the Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2007; pp. 801–808.
42. Schneider, R.; Uschmajew, A. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM J. Optim.* **2015**, *25*, 622–646. [[CrossRef](#)]
43. Pan, L.L.; Xiu, N.H.; Zhou, S.L. Gradient support projection algorithm for affine feasibility problem with sparsity and nonnegativity. *Mathematics* **2014**, *42*, 1439–1444.
44. Gray, D.; Tao, H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 262–275.
45. Hirzer, M.; Belezni, C.; Roth, P.M.; Bischof, H. Person re-identification by descriptive and discriminative classification. In *Proceedings of the Scandinavian Conference on Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 91–102.
46. Loy, C.; Liu, C.; Gong, S. Person re-identification by manifold ranking. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia, 15–18 September 2013; pp. 3567–3571.
47. Li, W.; Zhao, R.; Wang, X. Human reidentification with transferred metric learning. In *Asian Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 31–44.
48. Li, W.; Zhu, X.; Gong, S. Person re-identification by deep joint learning of multi-loss classification. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 2194–2200.
49. Xiao, T.; Li, H.; Ouyang, W.; Wang, X. Learning deep feature representations with domain guided dropout for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1249–1258.
50. Chen, Y.C.; Zheng, W.S.; Lai, J.H.; Yuen, P.C. An asymmetric distance model for cross-view feature mapping in person reidentification. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *27*, 1661–1675. [[CrossRef](#)]
51. Yang, X.; Wang, M.; Tao, D. Person re-identification with metric learning using privileged information. *IEEE Trans. Image Process.* **2018**, *27*, 791–805. [[CrossRef](#)] [[PubMed](#)]

52. Hu, H.M.; Fang, W.; Li, B.; Tian, Q. An adaptive multi-projection metric learning for person re-identification across non-overlapping cameras. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 2809–2821. [[CrossRef](#)]
53. Song, J.; Yang, Y.; Song, Y.Z.; Xiang, T.; Hospedales, T.M. Generalizable person reidentification by domain-invariant mapping network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 719–728.
54. Jia, J.; Ruan, Q.; Jin, Y.; An, G.; Ge, S. View-specific subspace learning and re-ranking for semi-supervised person re-identification. *Pattern Recogn.* **2020**, *108*, 107568. [[CrossRef](#)]
55. Hirzer, M.; Roth, P.M.; Köstinger, M.; Bischof, H. Relaxed pairwise learned metric for person re-identification. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 780–793.
56. Lin, W.; Shen, Y.; Yan, J.; Xu, M.; Wu, J.; Wang, J.; Lu, K. Learning correspondence structures for person re-identification. *IEEE Trans. Image Process.* **2017**, *26*, 2438–2453. [[CrossRef](#)]
57. Sun, C.; Wang, D.; Lu, H. Person re-identification via distance metric learning with latent variables. *IEEE Trans. Image Process.* **2017**, *26*, 23–34. [[CrossRef](#)]
58. Su, C.; Yang, F.; Zhang, S.; Tian, Q.; Davis, L.S.; Gao, W. Multi-task learning with low rank attribute embedding for multi-camera person reidentification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1167–1181. [[CrossRef](#)]
59. Li, H.; Zhu, J.; Tao, D. Asymmetric projection and dictionary learning with listwise and identity consistency constraints for person re-identification. *IEEE Access* **2018**, *6*, 37977–37990. [[CrossRef](#)]
60. Ma, L.; Yang, X.; Tao, D. Person re-identification over camera networks using multi-task distance metric learning. *IEEE Trans. Image Process.* **2019**, *23*, 3656–3670.
61. Zhang, Y.; Li, B.; Lu, H.; Irie, A.; Ruan, X. Sample-specific SVM learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1278–1287.
62. An, L.; Chen, X.; Yang, S.; Li, X. Person re-identification by multihypergraph fusion. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2763–2774. [[CrossRef](#)]
63. Guo, J.; Zhang, Y.; Huang, Z.; Qiu, W. Person re-identification by weighted integration of sparse and collaborative representation. *IEEE Access* **2017**, *5*, 21632–21639. [[CrossRef](#)]
64. Dai, J.; Zhang, Y.; Lu, H.; Wang, H. Cross-view semantic projection learning for person re-identification. *Pattern Recognit.* **2018**, *75*, 63–76. [[CrossRef](#)]
65. Lei, J.; Niu, L.; Fu, H.; Peng, B.; Huang, Q.; Hou, C. Person re-identification by semantic region representation and topology constraint. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 2690–2702. [[CrossRef](#)]
66. Han, H.; Zhou, M.; Shang, X.; Cao, W.; Abusorrah, A. KISS+ for rapid and accurate pedestrian re-identification. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 394–403. [[CrossRef](#)]
67. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE International Conference on Computer Vision, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
68. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
69. Zhao, R.; Ouyang, W.; Wang, X. Person re-identification by saliency matching. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2528–2535.
70. Zhao, R.; Ouyang, W.; Wang, X. Learning mid-level filters for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 144–151.
71. Rao, W.; Xu, M.; Zhou, J. Improved metric learning algorithm for person re-identification based on asymmetric metric. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications, Dalian, China, 27–29 June 2020; pp. 212–216.
72. Chen, W.; Chen, X.; Zhang, J.; Huang, K. Beyond triplet loss: A deep quadruplet network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1320–1329.
73. Wang, G.; Yuan, Y.; Chen, X.; Li, J.; Zhou, X. Learning discriminative features with multiple granularities for person re-identification. In Proceedings of the 26th ACM international conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 274–282.
74. Chang, X.; Hospedales, T.; Xiang, T. Multi-level factorisation net for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2109–2118.
75. Bak, S.; Carr, P. Deep deformable patch metric learning for person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 2690–2702. [[CrossRef](#)]
76. Wang, Y.; Wang, L.; You, Y.; Zou, X.; Chen, V.; Li, S.; Huang, G.; Hariharan, B.; Weinberger, K. Resource aware person reidentification across multiple resolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8042–8051.
77. Wei, L.; Zhang, S.; Yao, H.; Gao, W.; Tian, Q. GLAD: Global-local alignment descriptor for scalable person re-identification. *IEEE Trans. Multimed.* **2019**, *21*, 986–999. [[CrossRef](#)]
78. Yao, H.; Zhang, S.; Hong, R.; Zhang, Y.; Xu, C.; Tian, Q. Deep representation learning with part loss for person re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 2860–2871. [[CrossRef](#)] [[PubMed](#)]

79. Shen, C.; Qi, G.; Jiang, R.; Jin, Z.; Yong, H.; Chen, Y.; Hua, X. Sharp attention network via adaptive sampling for person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 3016–3027. [[CrossRef](#)]
80. Bai, X.; Yang, M.; Huang, T.; Dou, Z.; Yu, R.; Xu, Y. Deep-person: Learning discriminative deep features for person re-identification. *Pattern Recognit.* **2020**, *98*, 107036. [[CrossRef](#)]
81. Yuan, Y.; Zhang, J.; Wang, Q. Deep Gabor convolution network for person re-identification. *Neurocomputing* **2020**, *378*, 387–398. [[CrossRef](#)]