Journal of Electronic Imaging

JElectronicImaging.org

Person reidentification by semisupervised dictionary rectification learning with retraining module

Hongyuan Wang
Zongyuan Ding
Ji Zhang
Suolan Liu
Tongguang Ni
Fuhua Chen

Hongyuan Wang, Zongyuan Ding, Ji Zhang, Suolan Liu, Tongguang Ni, Fuhua Chen, "Person reidentification by semisupervised dictionary rectification learning with retraining module," *J. Electron. Imaging* **27**(4), 043043 (2018), doi: 10.1117/1.JEI.27.4.043043.

Person reidentification by semisupervised dictionary rectification learning with retraining module

Hongyuan Wang,[a,*] Zongyuan Ding,[a] Ji Zhang,[a] Suolan Liu,[a] Tongguang Ni,[a] and Fuhua Chen[b]
[a]Changzhou University, School of Information Science and Engineering, Changzhou, Jiangsu, China
[b]West Liberty University, College of Science, West Liberty, West Virginia, United States

Abstract. At present, in the field of person reidentification (re-id), the commonly used supervised learning algorithms require a large amount of labeled samples, which is not conducive to the model promotion. On the other hand, the accuracy of unsupervised learning algorithms is lower than supervised algorithms due to the lack of discriminant information. To address these issues, we make use of a small amount of labeled samples to add discriminant information in the basic dictionary learning. Moreover, the sparse coefficients of dictionary learning are decomposed into a projection problem of the original features, and the projection matrix is trained by labeled samples, which is transformed into a metric learning problem. It thus integrates the advantages of the two methods through combining dictionary learning and metric learning. After the data are trained, a projection matrix is used to project the unlabeled features into a feature subspace and the labels of the samples are reconstructed. The semisupervised learning problem is then transformed to a supervised learning problem with a graph regularization term. Experiments on different public pedestrian datasets, such as VIPeR, PRID, iLIDS, and CUHK01, show that the recognition accuracy of our method is better than some other existing person re-id methods. [DOI: 10.1117/1.JEI.27.4.043043]

Keywords: person reidentification; semisupervised learning; dictionary learning; metric learning; Laplace term.

Paper 180365 received Apr. 27, 2018; accepted for publication Jul. 24, 2018; published online Aug. 11, 2018.

1 Introduction

Person reidentification (re-id) is a key problem in surveillance and security applications. It is getting more and more attention in the field of pattern recognition and artificial intelligence. Many new solutions have been developed in recent years, which can be classified into two groups: the feature-based methods[1–6] and the model training-based methods.[7–18] Among the model training-based methods, metric learning,[7–10] multitask learning,[11–13] and dictionary learning (also known as sparse coding)[14–18] are studied largely. Apart from the above-mentioned, methods based on deep learning have achieved satisfying accuracy in computer vision, including re-id problem,[19–21] whether it is based on the method of extracting deep features or the end-to-end method. However, there are still many challenges about person re-id due to change of viewpoints, change of illuminations, and different resolutions. What's more, despite excellent performance of deep learning for re-id problem, there are some common obstacles for deep learning, such as the need for large samples and long training time. Furthermore, they require a significant computational resource.

Among the aforementioned methods, metric learning-based methods received even more attention and achieved better results. These methods use labeled images to train the model and to find an optimal metric, in which labels are used to generate discriminant information, and the discriminant information are then used to instruct the learning. This kind of learning is called supervised learning. Therefore, a key step for supervised learning is to label as many images as possible. But in reality, getting a large number of labels for pedestrian images is quite tedious and unpractical, which makes the supervised learning-based person re-id not conducive to promotion. An alternative way is to use unsupervised learning methods for person re-id. Dictionary learning has been proven to be very successful in unsupervised learning and has achieved a lot of success in face recognition and visual tracking.[22–24] Currently, dictionary learning has also been introduced into person re-id,[17,18,25] but the recognition rate is still very low. One major reason is that such methods lack discriminant information. In face recognition, the features of face images are stable and robust. Therefore, a lot of discriminant information can be extracted for training. Different from face recognition, the features of pedestrian images usually vary hugely due to different cameras, scenes, resolutions, and lightening. To address this problem, some researchers add a graph regularization term in their model[17,26] to improve the recognition rate. However, the Laplace matrix of graph regularization term is built on the original feature space that is also not very unreliable.

To solve these issues, we propose a semisupervised dictionary rectification learning with retraining module (SSDRL-r)-based method for person re-id (see Fig. 1), which integrates the benefits from both supervised methods and unsupervised methods. On the one side, it uses the dictionary learning to reduce the requirement of large amount of labeled images; on the other side, it only uses a small number of labeled images to collect discriminant information. As we know, it is practical to collect the labels of few samples (notice that all the labeled samples are of known classes, because we perform experiments in this paper using specific

*Address all correspondence to: Hongyuan Wang, E-mail: hywang@cczu.edu.cn

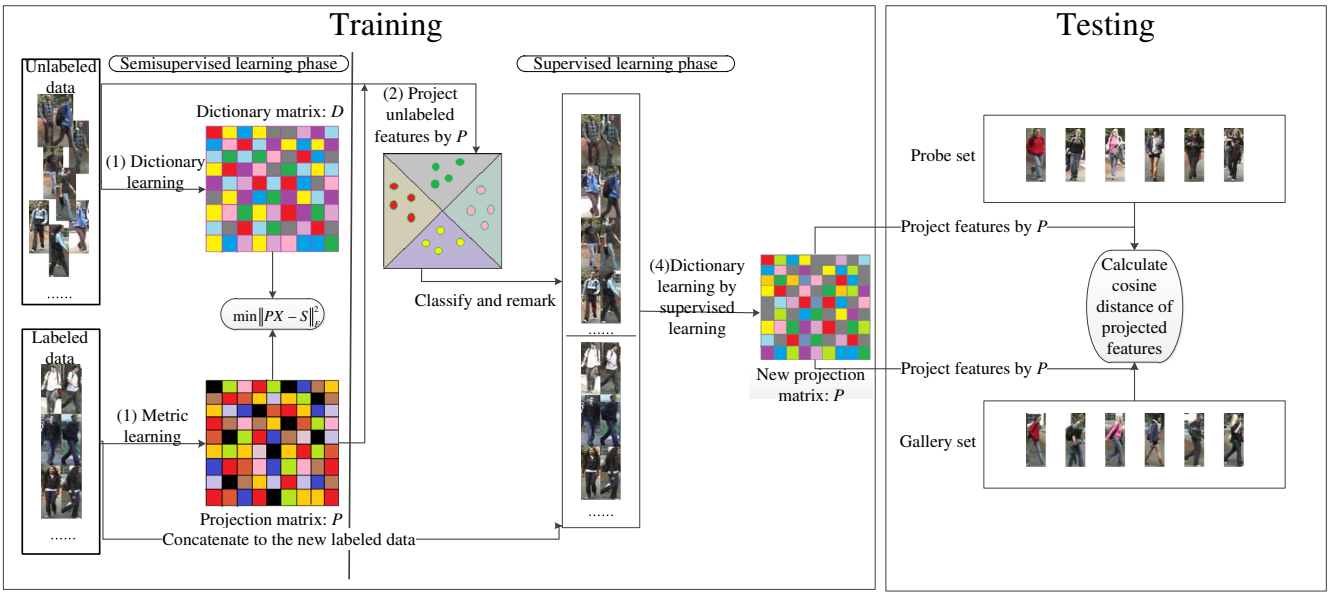

Fig. 1 Flowchart of SSDRL for person re-id. (In the phase of testing, we classify different pedestrians by *k*NN classifier after calculating cosine distance of projected features.)

public datasets). More detailed, we first combine dictionary learning and metric learning by forcing the projected features obtained from metric learning and the features obtained from dictionary learning as close as possible. The dictionary matrix is learned by all pedestrian images and the projection matrix is learned by labeled pedestrian images. At this stage, the training process is semisupervised learning. After the semisupervised learning, we use a projection matrix to project the original features to a new feature space and then reconstruct pseudolabels of unlabeled features by the k-nearest neighbors (kNN) classifier. (Here, we choose the optimal number of neighbors by several experiments.) Since all features have "labels," we can transform the model to a supervised learning. Then at the second stage, a Laplace term is introduced into the model to retrain the dictionary learning and make the dictionary matrix and the projection matrix more robust. The Laplace term here is constructed by the projected label information, not the original features, which can make the model more credible. Our major contribution in this paper is to investigate a method to integrate the benefits of dictionary learning and the benefits of metric learning such that less labeled images are used but more discriminant information is obtained.

2 Model Development

2.1 *Dictionary Learning for Person Reidentification*

In this paper, we denote $X \in R^{n\times m}$ as the pedestrian feature matrix, where n represents the dimension of pedestrian vectors and m represents the number of pedestrian images. As in usual dictionary learning, we denote $D \in R^{n\times k}$ and $S \in R^{k\times m}$ as the dictionary matrix and the sparse coefficient matrix, respectively. The goal of dictionary learning is to find a sparse dictionary representation so that such a representation of a feature vector is close to the original feature vectors. To make coefficients sparse, an l_1 norm is introduced to constrain the coefficients. The primitive function of dictionary learning can be shown as

$$\min_{D,S} \|X - DS\|_F^2 + \lambda_1 \|S\|_1, \tag{1}$$

where $\|\cdot\|_F$ represents Frobenius norm, $\|\cdot\|_1$ represents l_1 norm, and λ_1 is a weight to balance the coding term and the sparse term.

Equation (1) is hard to be solved because the l_1 norm is not always differentiable. In addition, the model must be solved during the testing stage, which is time-consuming and not practical. To improve the model, we add a projection term to help find an optimal projection matrix P by making PX and S close to each other (by massive training of dictionary learning, optimal sparse coefficient matrix S has learnt, so it can get optimal P by this method in the this stage), where $P \in R^{k\times n}$ is the projection matrix. Although the improved model still needs to solve an l_1 norm problem during the training stage, however, we just need to calculate the Euclidean distance after projection in the testing stage, so it is more time-saving. The model is shown as

$$\min_{D,P,S} \|X - DS\|_F^2 + \lambda_1 \|S\|_1 + \lambda_2 \|PX - S\|_F^2, \tag{2}$$

where λ_2 is a parameter used to balance among the three terms.

Equation (2) is not practical and may lead to singular solutions. The general solution is to add constraints to force the Frobenius norm of each of the two matrices to be no more than one. As stated in Sec. 2.4, it is clear that dictionary matrix D can be calculated by adding disturbance term $\alpha_1 I$, otherwise term SS^T has a great possibility of irreversibility, which makes the problem impossible to solve. After two constrains are added, the model is given as

$$\min_{D,P,S} \|X - DS\|_F^2 + \lambda_1 \|S\|_1 + \lambda_2 \|PX - S\|_F^2$$
$$\text{s.t. } \|D\|_F^2 \le 1, \|P\|_F^2 \le 1. \tag{3}$$

2.2 *Semisupervised Dictionary Learning for Person Reidentification*

Equation (3) is a general formula for dictionary learning and no supervised information (label information) is needed. For the specific person re-id problem, original feature space is not trustworthy. So the recognition rate of this model is often lower than supervised learning methods. From Eq. (3), we can see that the projection matrix P is only trained on the unlabeled pedestrian features, which may mislead the training of sparse coefficients S. For this reason, we make full use of the labeled pedestrian features in our framework and transform the process of training projection matrix to a supervised learning problem.

First, as in many relative distance-based methods, we build the positive pairs and negative pairs $\mathbb{S} = \{\mathbb{S}_t = (p_t^{\text{pos}}, p_t^{\text{neg}}) | t = 1,2,\ldots,l\}$, where l represents the number of labeled pedestrians. More specifically, the positive pairs are built based on the images from same pedestrians, and the negative pairs are built based on images from different pedestrians. After building the positive pairs and negative pairs, we compute the difference of any two images of positive pairs and that of negative pairs, which are denoted as $d^{\text{pos}} = \{d_t^{\text{pos}} | t = 1,2,\ldots,l\}$ and $d^{\text{neg}} = \{d_t^{\text{neg}} | t = 1,2,\ldots,l\}$, respectively. The aim of the projection is to make the distances of positive pairs shorter than that of negative pairs. Or equivalently, we want to minimize the quantity Y defined below

$$Y = \frac{1}{l}\left(\sum_{t=1}^{l} d_t^{\text{pos}} - \sum_{t=1}^{l} d_t^{\text{neg}}\right). \tag{4}$$

By adding the distance term Y into Eq. (3), we obtain the following model:

$$\begin{aligned} O_{\text{semi}}(D,P,S) = \min_{D,P,S} & \|X - DS\|_F^2 + \lambda_1 \|S\|_1 \\ & + \lambda_2 \|PX - S\|_F^2 + \gamma \|PY\|_F^2 \\ & \text{s.t. } \|D\|_F^2 \le 1, \|P\|_F^2 \le 1. \end{aligned} \tag{5}$$

This is the final model of our semisupervised dictionary rectification learning (SSDRL) for person re-id. By training with some labeled images, the projection matrix P can be more credible.

2.3 *Retraining by Supervised Learning for Person Reidentification*

It is noted that only using a small size of labeled samples in previous introduced semisupervised learning algorithm cannot guarantee the efficiency of the re-id framework. (It is better to use the full-label information to train the model.) To make the model more robust, after training of semisupervised learning [i.e., Eq. (5)], we can get the optimal projection matrix P, then we use the obtained projection matrix P to project the unlabeled pedestrian features to a new feature space, classify the pedestrians in the new feature space using the kNN classifier, and then remark the samples. In this way, the learning process in this stage becomes supervised learning. The number of classes in the kNN classifier is determined by multitrials in the training stage.

Inspired by the excellent performance of those dictionary learning methods with graph regularization term, we further proposed SSDRL-r, in more detail, except the step of training model of Eq. (5), we retrain the projection matrix and dictionary matrix by adding a graph regularization term to the model with pseudolabel information. Unlike some existing methods, we use the labels as an instruction other than neighborhood information to construct the Laplace matrix of graph regularization term. In this paper, the Laplace matrix is denoted by L and is written as

$$L = Q - W, \tag{6}$$

where $W \in R^{n_l \times n_l}$ consists of w_{ij} in the i'th row and the j'th column, n_l is the number of all pedestrian images. $Q \in R^{n_l \times n_l}$ is a diagonal matrix whose entries are the column sum of W, i.e., $Q_{ii} = \sum_j w_{ij}$. Here, w_{ij} is the semantic affinity metric of two pedestrians x_i and x_j, which is defined as follows:

$$w_{ij} = \begin{cases} 1, & \text{if } x_i \text{ and } x_j \text{ are the same person} \\ 0 & \text{otherwise} \end{cases}. \tag{7}$$

To constraint the similarity of pedestrian images that belong to a same class in the common semantic space, we can minimize the following function:

$$\begin{aligned} O(S) &= \frac{1}{2}\sum_{i,j=1}^{n_l} w_{ij}\|s_i - s_j\|^2 \\ &= \text{tr}[S(Q-W)S^T] \\ &= \text{tr}(SLS^T), \end{aligned} \tag{8}$$

where $\text{tr}(\cdot)$ represents the trace of a matrix.

In this stage, the training process is a supervised learning. The sparse term is dropped in this stage so that the training process can be faster than the first stage. On the other hand, the graph regularization term here can constrain the intraclass distance very well. It can be seen from Eq. (7) that the semantic affinity metric of the same pedestrian is 1, and the semantic affinity metric of the nonpeer is 0, so as is seen in Eq. (8), minimizing the graph regularization term, the distance between the same pedestrians can be minimized, which further enhances the robustness of the model. The final model of the SSDRL-r for person re-id is transformed to

$$\begin{aligned} O_{\text{sup}}(D,P,S) = \min_{D,P,S} & \|X - DS\|_F^2 + \beta_1 \text{tr}(SLS^T) \\ & + \beta_2 \|PX - S\|_F^2 \\ & \text{s.t. } \|D\|_F^2 \le 1, \|P\|_F^2 \le 1, \end{aligned} \tag{9}$$

where β_1 and β_2 are the parameters used to balance among different terms.

2.4 *Optimization of the Model of SSDRL*

Equation (5) is nonconvex with three matrix variables D, P, and S. Moreover, the l_1 norm in Eq. (5) is not always differentiable. However, the model is convex with respect to any one of the three variables when the remaining two variables are treated as constants. So the model can be solved using the alternating direction method of multipliers (ADMM)[27] by repeating the three steps described as follows until convergence.

Step 1: By fixing S and P, the objective function reduces to

$$\min_D \|X - DS\|_F^2 \quad \text{s.t.} \|D\|_F^2 \le 1.$$

To solve this, we use the Lagrange dual method as in Ref. 28, so we can convert this constraint minimization problem into an unconstrained minimization problem, the objective function can be written as

$$\min_D \|X - DS\|_F^2 + \alpha_1(\|D\|_F^2 - 1),$$

where α_1 is the Lagrange multiplier, and the analytical solution of D can be computed as

$$D = XS^T(SS^T + \alpha_1 I)^{-1}. \tag{10}$$

Step 2: Similar to the step 1, by fixing S and D, P can be solved explicitly as

$$P = \lambda_2 SX^T(\lambda_2 XX^T + \gamma YY^T + \alpha_2 I)^{-1}, \tag{11}$$

where I is the identity matrix and α_2 is the Lagrange multiplier.

Step 3: Fix D and P, update S. Since the last term of Eq. (5) has nothing to do with S, the objective function can be simplified to

$$O_{\text{semi}}(S) = \min_S \|X - DS\|_F^2 + \lambda_1\|S\|_1 + \lambda_2\|PX - S\|_F^2. \tag{12}$$

So the ADMM form of the above equation is

$$\begin{aligned} &\text{minimize} \|X - DS\|_F^2 + \lambda_1\|M\|_1 + \lambda_2\|PX - S\|_F^2 \\ &\text{s.t. } S - M = 0. \end{aligned} \tag{13}$$

According to the algorithm of ADMM, we can solve it alternatingly with the following three steps with respect to S and M, respectively. First, for given M^k and U^k, solve the following objective function to estimate S

$$S^{k+1} = \arg\min_S (\|X - DS\|_F^2 + \lambda_2\|PX - S\|_F^2 + \frac{\rho}{2}\|S - M^k + U^k/\rho\|_F^2),$$

where U is the Lagrange multiplier defined as in Ref. 27 and ρ is a penalty parameter.

Since each term in this objective function is quadratic, we can take partial derivative and set it to zero to get S^{k+1}

$$S^{k+1} = [2D^TD + (2\lambda_2 + \rho)I]^{-1}(2D^TX + 2\lambda_2 PX + \rho M^k - U^k). \tag{14}$$

Second, for given S^{k+1} and U^k, solve the following objective function to estimate M:

$$M^{k+1} = \arg\min_M (\lambda_1\|M\|_1 + \frac{\rho}{2}\|S^{k+1} - M + U^k/\rho\|).$$

We can use the soft-thresholding operator to get M^{k+1}

$$M^{k+1} = \text{sign}(S^{k+1} + U^k/\rho)\max(|S^{k+1} + U^k/\rho| - \lambda_1/\rho). \tag{15}$$

For Lagrange multiplier U, we update it using the following way:

$$U^{k+1} = U^k + \rho(S^{k+1} - M^{k+1}). \tag{16}$$

We summarize all steps of SSDRL for person re-id in Algorithm 1.

2.5 *Optimization of Supervised Learning with Graph Regularization Term*

Equation (9) is also convex with respect to any one of the three variables when the other variables are treated as constants. So we can still use the alternating iterative strategy to solve the problem. For the specific matrix variable D, due to

Algorithm 1 SSDRL for person re-id

Input: training samples X, labeled feature difference vector Y, balance factors λ_1, λ_2, γ, Lagrange multipliers α_1, α_2, penalty parameter ρ, accuracy controller ε, maximum iteration T

Output: dictionary matrix D, projection matrix P

```
Initialization: D, P, iteration index k1, k2, sparse coefficient S, auxiliary variables U, M
Compute f^{k1} using Eq. (5)
while f^{k1} − f^{k1+1} > ε do
for k1 = 1, 2, ... T do
    Update D, P using Eqs. (10) and (11)
    Compute f^{k2} using Eq. (13)
    while f^{k2} − f^{k2+1} > ε do
      for k2 = 1, 2, ... T do
        Update S, M, U using Eqs. (14)–(16)
        Compute f^{k2+1} using Eq. (13)
        f^{k2+1} ← f^{k2}
      end
    end
    Compute f^{k1+1} using Eq. (5)
    f^{k1+1} ← f^{k1}
  end
end
```

Algorithm 2 Relabeling supervised learning for person re-id

Input: training samples X, balance factors β_1, β_2, Lagrange multipliers α_1, α_2, accuracy controller ε, maximum iteration T

Output: dictionary matrix D, projection matrix P

```
Initialization: D, P, iteration index k3, sparse coefficient S
Compute f^{k3} using Eq. (5)
while f^{k3} − f^{k3+1} > ε do
  for k3 = 1, 2, ..., T do
    Update D, P using Eqs. (6) and (12)
    Update S using lyap function of MATLAB in Eq. (13)
    Compute f^{k3+1} using Eq. (5)
    f^{k3+1} ← f^{k3}
  end
end
```

the similar form to that in the semisupervised learning stage, we update D using Eq. (10).

Accordingly, the optimal solution of P can be solved explicitly as

$$P = \beta_2 S X^T (\beta_2 X X^T + \alpha_3 I)^{-1}. \tag{17}$$

For the matrix variable S, by fixing P and D and letting $\frac{\partial O_{\text{sup}}}{\partial S} = 0$, we obtain

$$AS + SB + C = 0, \tag{18}$$

where $A = 2D^T D + 2\beta_2 I$, $B = \beta_1 (L + L^T)$, and $C = -2(D^T + \beta_2 P)X$. Equation (18) is a Sylvester equation[29] and can be solved using the "lyap" function in MATLAB.

We summarize all steps of relabeling supervised learning for person re-id in Algorithm 2.

2.6 *Reidentification*

After learning the projection matrix P using training datasets, we can test the efficiency of the framework. Given a pair of test samples x_i^a and x_i^b, we use the projection matrix to project the original feature vectors as below to obtain new features y_i^a and y_i^b

$$y_i^a = P x_i^a \quad y_i^b = P x_i^b.$$

After obtaining projected feature vectors y_i^a and y_i^b, their matching is done by computing the cosine distance between their respective projected feature vectors. The distance is used to measure the visual similarity for re-id. Hence, our model is very efficient in the stage of testing.

3 Experiments

3.1 *Datasets and Settings*

3.1.1 *Datasets*

As shown in Fig. 2, several public datasets are used for the experiments. VIPeR[30] contains two cameras, each of which captures one image per person. It also provides the viewpoint angle of each image. Although it has been tested by many researchers, it is still one of the most challenging datasets. We split the dataset into two subsets of 316 image pairs, one for training and the other for testing. Sixteen image pairs of training set are labeled and the rest are unlabeled. PRID[31] is different from other existing datasets in that the gallery and probe sets have different numbers of people. In our experiments, we use the single-shot version of the dataset. Only 200 people appear in both views of this dataset. In each data split, 100 out of these 200 people are chosen randomly for training while the remaining 100 are used for testing. In the training set, only 20 out of 100 people's images are used as labeled images. The dataset iLIDS[32] contains 476 images of 119 people. We randomly choose 83 people's images for training and the remaining for testing, and 20 people's images of the training set are used as labeled images. CUHK01[33] consists of 971 people with two images per person per camera view. We set 486 people's images for training and the rest for testing, and 86 people's images of the training set are used as labeled images. Except for the above-mentioned small- and medium-scaled datasets, we also perform the experiments on large-scaled dataset Market-1501,[34] which contains 32,668 images of 1501 people. We randomly divided the people into two parts, that is, 1000 people for training and 501 people for testing. Among the training set, we randomly selected 100 people's images as labeled images and rest as unlabeled images. All the experiments in this paper are carried out by cross-validation mechanism, that is, the datasets are randomly divided into specified ways introduced already, and multiple sets of experiments are performed, then the mean value of the matching rate and optimal parameters are finally obtained.

3.1.2 *Features*

The features performed in the experiments are the same as in Ref. 35, which are computed consisting of three kinds: color histogram using RGB, HS, and lab color spaces (2880-D); HOG (1040-D);[36] and LBP (1218-D).[37] The final image feature vector, 5138-D, is obtained by concatenating these three kinds of features.

3.1.3 *Evaluation metric*

We adopt conventional cumulative matching characteristics (CMC) curves for our models and other models with codes available. The CMC curve represents the expectation of finding the correct match in the top n matches. However, to compare with a wider range of baselines, for which no code is available, we report cumulative matching accuracies at different ranks, which correspond to key points on the CMC curves. All the matching rates at top 20 ranks are displayed in form of tables.

3.1.4 *Parameter setting*

The parameters of our model were set to the following: balance factor $\lambda_1, \lambda_2, \gamma, \beta_1$, and β_2 are 1, 1000, 1000, 100, and

Fig. 2 Different datasets (from left to right are VIPeR, PRID, iLIDS, CUHK01, and Market-1501, respectively).

1000, respectively. First, these parameters were set by experience [e.g., to emphasis the weight of semisupervised learning, we need to set a large weight of projection term in Eq. (5)], and later, we set these parameters using the ideal of dichotomy. The Lagrange multipliers α_1, α_2, and α_3 are all 1. Penalty parameter ρ in Eq. (16) is 1000. Accuracy control parameters in Algorithms 1 and 2 are both 0.001. In addition, the number of column k in the dictionary matrix D is one half of the numbers of rows. All the matrix variables (for example, dictionary matrix, sparse coefficients, and projection matrix) are initialized using the rand function in MATLAB.

3.2 *Evaluation of Unsupervised Learning-Based Person Reidentification*

3.2.1 *Competitors*

Under this setting, we compared our method (SSDRL-r) with three categories of methods: (1) the hand-crafted feature-based methods, including SDALF[5] and CPS,[6] in which the features are designed to be view invariant; (2) the saliency learning-based eSDC[1] and GTS;[2] and (3) the sparse representation classification-based ISR[16] and GL.[17]

3.2.2 *Analysis*

Comparison between SSDRL-r and other unsupervised learning methods is shown in Tables 1–5, "—" means no result reported. The numbers in bold in Tables 1–5 and 8 indicate that the algorithm's matching rate at specific rank is higher than other algorithms. From Tables 1–5, we can see that SSDRL-r is superior to other unsupervised learning-based person re-id methods on the datasets. In our experiments, we relabel the pedestrian features by the kNN classifier, and the number of neighbors here is the average number of images owned by each pedestrian in the labeled samples. Especially for the dataset VIPeR and CUHK01, since the number of each pedestrian's images is the same, we can get better classification of these samples after projecting the original unlabeled samples. Based on the more accurate label information, better experimental results can be achieved at the stage of supervised learning. We can see that SSDRL-r algorithm can also get better performance on large scale dataset Market-1501. From Table 2, although matching rate at rank1 of SSDRL-r is less than that of ISR, we can see

Table 1 Performance of different methods on dataset VIPeR (%).

Methods	Rank1	Rank5	Rank10	Rank20
SDALF	19.9	38.9	49.4	65.7
CPS	22.0	44.7	57.0	71.0
eSDC	26.7	50.7	62.4	76.4
GTS	25.2	50.0	62.5	75.8
ISR	27.0	49.8	61.2	73.0
GL	33.5	52.3	64.8	75.2
SSDRL-r	**38.3**	**60.0**	**66.7**	**76.7**

Table 2 Performance of different methods on dataset iLIDS (%).

Methods	Rank1	Rank5	Rank10	Rank20
SDALF	28.2	46.3	56.5	66.4
CPS	29.8	52.6	62.2	73.0
eSDC	—	—	—	—
GTS	—	—	—	—
ISR	**39.7**	56.8	67.5	77.3
GL	—	—	—	—
SSDRL-r	37.8	**57.4**	**72.1**	**85.3**

SSDRL-r's matching rate is still higher than other methods at rank5, rank10, and rank20, so does the dataset PRID.

3.3 *Comparison with Other Learning Models-Based Person Reidentification*

Apart from comparison with other common unsupervised learning, in this part, we perform our model with other

Table 3 Performance of different methods on dataset PRID (%).

Methods	Rank1	Rank5	Rank10	Rank20
SDALF	16.3	29.6	38.0	48.7
CPS	—	—	—	—
eSDC	—	—	—	—
GTS	—	—	—	—
ISR	17.0	34.4	42.0	54.3
GL	**25.0**	44.5	56.3	72.7
SSDRL-r	20.3	**46.3**	**60.9**	**75.4**

Table 4 Performance of different methods on dataset CUHK01 (%).

Methods	Rank1	Rank5	Rank10	Rank20
SDALF	9.9	23.4	29.8	—
CPS	—	—	—	—
eSDC	26.6	—	—	—
GTS	—	—	—	—
ISR	53.2	72.3	80.5	87.5
GL	41.0	68.6	79.9	90.2
SSDRL-r	**55.8**	**75.4**	**83.7**	**91.7**

Table 5 Performance of different methods on dataset Market-1501 (%).

Methods	Rank1	Rank5	Rank10	Rank20
SDALF	33.5	46.3	78.6	89.3
CPS	—	—	—	—
eSDC	—	—	—	—
GTS	36.2	53.3	82.0	90.1
ISR	40.3	67.6	85.7	92.5
GL	—	—	—	—
SSDRL-r	**54.2**	**73.5**	**86.1**	**95.5**

learning models based person re-id, which are comprised of original dictionary learning [i.e., Eq. (1)], SSDRL model [i.e., Eq. (5)]. So we can justify that SSDRL-r is superior to original dictionary learning and SSDRL model.

Table 6 Results compared to the SSDRL-r and dictionary learning, measured by rank1 accuracies (%).

Datasets	VIPeR	iLIDS	PRID	CUHK01	Market-1501
Dic	29.9	30.5	14.3	49.3	50.2
SSDRL-r	38.3	37.8	20.3	55.8	54.2

Table 7 Results compared to the SSDRL-r and SSDRL, measured by rank1 accuracies (%).

Datasets	VIPeR	iLIDS	PRID	CUHK01	Market-1501
SSDRL	30.6	35.5	16.4	50.9	38.3
SSDRL-r	38.3	37.8	20.3	55.8	54.2

Table 8 Results compared to the deep learning reported in literatures, measured by rank1 accuracies (%).

Datasets	VIPeR	iLIDS	PRID	CUHK01
Cheng 2016	**47.8**	**60.4**	**22.0**	**53.7**
SSDRL-r	40.3	39.8	21.3	52.8

3.3.1 *Dictionary learning*

First, we compare SSDRL-r model with dictionary learning (denote as Dic), as is shown in Eq. (1), the dictionary learning is a total unsupervised learning model, so the training process might be misled because of the lack of discriminative information. We mainly observe matching rate at rank1 on different datasets. From Table 6, we can see SSDRL-r has higher accuracy than dictionary learning on these five datasets. Therefore, a small amount of label information and retraining strategy can provide better guidance for the training process.

3.3.2 *SSDRL*

On the other hand, we need to verify the effectiveness of retraining process in SSDRL-r. Therefore, we need to focus on the experimental results of SSDRL and SSDRL-r. Because the strategy of remarking the unlabeled images is introduced in SSDRL, more discriminative information can be used to guild the construction of a more robust model. The experimental results of SSDRL-r and SSDRL are shown in Table 7. We can see that the matching rates of SSDRL-r at rank1 are higher than SSDRL on five dataset, which verify that more labels can improve the robust of the learning model.

However, as we all know, deep learning has achieved great performance in the field of computer vision, including person re-id. We also compare our model with the deep learning-based method in Ref. 21 (denoted as Cheng 2016, there is no report on dataset Market-1501). Under the same set of size of sample in different datasets, we redo the experiment on different datasets by SSDRL-r, and the results of the

experiment are shown in Table 8. Clearly, deep learning-based method is superior to our method, this is the weakness of all nondeep learning models, but deep learning also has some inherent shortcomings, such as the need for large samples, long training time, and dependence on high-performance devices, and the difference between the accuracy of our method and the deep learning algorithm is not too big on dataset PRID and CUHK01.

4 Conclusion

In this paper, we proposed a semisupervised re-id model based on dictionary rectifying learning. The key contribution of our method is the integration of dictionary learning and metric learning so that the information of a small size of labeled sample can be reused to rectify the dictionary matrix. In addition, classifying the unlabeled features is developed, which remarks the unlabeled features and converts the model to a supervised learning problem. The graph regularization term is introduced in the second stage to further improve the efficiency of the proposed model to deal with outlying samples in person re-id data. Experiments on the four benchmark datasets show that the proposed method significantly outperforms the existing unsupervised methods. However, it still has some shortcomings with our model, for example, after first stage training, we use kNN algorithm to classify the unlabeled features, but kNN algorithm is of high central processing unit overhead, and the accuracy of our model is lower than deep learning-based method. In the future, we will do more research on developing a more effective classifier to class the unlabeled images. What's more, we can expand our model to the field of generic tasks, such as image classification.

Disclosures

This paper's original version has been listed in the proceedings of 2018 SPIE Commercial + Scientific Sensing and Imaging (SI18C), volume DL10670.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61502058 and 61572085 and the Jiangsu Joint Research Project of Industry, Education and Research under Grant No. BY2016029-15.

References

1. R. Zhao, W. L. Ouyang, and X. G. Wang, "Unsupervised salience learning for person re-identification," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3586–3593 (2013).
2. H. Wang, G. Gong, and T. Xiang, "Unsupervised learning of generative topic saliency for person re-identification," in *Proc. British Machine Vision Conf.* (2014).
3. S. Y. Ding et al., "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognit.* **48**(10), 2993–3003 (2015).
4. K. Liu et al., "A spatio-temporal appearance representation for video-based pedestrian re-identification," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 3810–3818 (2015).
5. M. Farenzena et al., "Person re-identification by symmetry-driven accumulation of local features," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. **23**, pp. 2360–2367 (2010).
6. D. S. Cheng et al., "Custom pictorial structures for re-identification," in *Proc. British Machine Vision Conf.*, Vol. **68**, pp. 1–11 (2011).
7. M. Hirzer, "Large scale metric learning from equivalence constraints," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2288–2295 (2012).
8. S. Pedagadi et al., "Local Fisher discriminant analysis for pedestrian re-identification," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. **9**, pp. 3318–3325 (2013).
9. B. He and S. Yu, "Ring-push metric learning for person reidentification," *J. Electron. Imaging* **26**(3), 033005 (2017).
10. J. X. Chen, Z. X. Zhang, and Y. H. Wang, "Relevance metric learning for person re-identification by exploiting global similarities," in *Int. Conf. on Pattern Recognition*, Vol. **24**, pp. 1657–1662 (2014).
11. A. J. Ma et al., "Cross-domain person reidentification using domain adaptation ranking SVMs," *IEEE Trans. Image Process.* **24**(5), 1599–1613 (2015).
12. L. Y. Ma, X. K. Yang, and D. C. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Trans. Image Process.* **23**(8), 3656–3670 (2014).
13. C. Su et al., "Multi-task learning with low rank attribute embedding for person re-identification," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 3739–3747 (2015).
14. S. Li, M. Shao, and Y. Fu, "Cross-view projective dictionary learning for person re-identification," in *Proc. AAAI*, pp. 2155–2161 (2015).
15. S. Karanam, Y. Li, and R. J. Radke, "Person re-identification with discriminatively trained viewpoint invariant dictionaries," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 4516–4524 (2016).
16. G. Lisanti et al., "Person re-identification by iterative re-weighted sparse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(8), 1629–1642 (2015).
17. E. Kodirov et al., "Person re-identification by unsupervised ℓ1 graph learning," *Lect. Notes Comput. Sci.* **9905**, 178–195 (2016).
18. E. Kodirov, T. Xiang, and S. Gong, "Dictionary learning with iterative Laplacian regularization for unsupervised person re-identification," in *Proc. British Machine Vision Conf.*, pp. 1–12 (2015).
19. N. McLaughlin, J. Del, and P. C. Miller, "Person reidentification using deep convnets with multitask learning," *IEEE Trans. Circuits Syst. Video Technol.* **27**(3), 525–539 (2017).
20. J. Wang et al., "DeepList: learning deep features with adaptive listwise constraint for person reidentification," *IEEE Trans. Circuits Syst. Video Technol.* **27**(3), 513–524 (2017).
21. D. Cheng et al., "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1335–1344 (2016).
22. X. Jia, H. Lu, and M. H. Yang, "Visual tracking via adaptive structural local sparse appearance model," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1822–1829 (2012).
23. S. Zhang et al., "Sparse coding based visual tracking: review and experimental comparison," *Pattern Recognit.* **46**(7), 1772–1788 (2013).
24. L. Fagot-Bouquet et al., "Improving multi-frame data association with sparse representations for robust near-online multi-object tracking," *Lect. Notes Comput. Sci.* **9912**, 774–790 (2016).
25. Y. Huang et al., "Person re-identification by unsupervised color spatial pyramid matching," *Lect. Notes Comput. Sci.* **9403**, 799–810 (2015).
26. J. Tang, K. Wang, and L. Shao, "Supervised matrix factorization hashing for cross-modal retrieval," *IEEE Trans. Image Process.* **25**(7), 3157–3166 (2016).
27. S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.* **3**(1), 1–122 (2010).
28. B. Schölkopf, J. Platt, and T. Hofmann, *Efficient Sparse Coding Algorithms*, pp. 801–808, MIT Press, Massachusetts (2007).
29. S. G. Lee and Q. P. Vu, "Simultaneous solutions of Sylvester equations and idempotent matrices separating the joint spectrum," *Linear Algebra Appl.* **435**(9), 2097–2109 (2011).
30. D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," *Lect. Notes Comput. Sci.* **5302**, 262–275 (2008).
31. M. Hirzer et al., "Person re-identification by descriptive and discriminative classification," *Lect. Notes Comput. Sci.* **6688**, 91–102 (2011).
32. W. S. Zheng, S. G. Gong, and T. Xiang, "Associating groups of people," in *Proc. British Machine Vision Conf.*, pp. 1–11 (2009).
33. W. Li, R. Zhao, and X. G. Wang, "Human reidentification with transferred metric learning," *Lect. Notes Comput. Sci.* **7724**, 31–44 (2012).
34. L. Zheng et al., "Scalable person re-identification: a benchmark," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 1116–1124 (2015).
35. L. Giuseppe, M. Iacopo, and D. B. Alberto, "Matching people across camera views using kernel canonical correlation analysis," in *Proc. of the Int. Conf. on Distributed Smart Cameras*, Vol. **10**, pp. 1–6 (2014).
36. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. **1**, pp. 886–893 (2005).
37. T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12), 2037–2041 (2006).

Hongyuan Wang received his PhD from Nanjing University of Science and Technology in 2004. He is a professor and a master supervisor at Changzhou University. His main research interests

include image processing, artificial intelligence, and pattern recognition.

Zongyuan Ding received his MS degree from Changzhou University in 2018. Currently, he is a PhD student at Nanjing University of Science and Technology. His main research interests include image processing and pattern recognition.

Ji Zhang received his MS degree from Nanjing University of Science and Technology in 2006. Currently, he is a lecturer at Changzhou University. His research interests include computer vision, image processing, and pattern recognition.

Biographies for the other authors are not available.