Pattern Recognition 75 (2018) 99–111

Contents lists available at ScienceDirect

Pattern Recognition

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

Feedback mechanism based iterative metric learning for person re-identification

Yutao Ren [a,b], Xuelong Li [b], Xiaoqiang Lu [b,*]

[a] Wuhan University of Technology, Wuhan 430070, PR China
[b] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China

ARTICLE INFO

Article history:
Received 15 November 2016
Revised 6 April 2017
Accepted 12 April 2017
Available online 18 April 2017

Keywords:
Person re-identification
Metric learning
Feedback
Iteration

ABSTRACT

Person re-identification problem is targeting to match people in the views of non-overlapped camera networks. It is an important task in the fields of computer vision and video surveillance. It shows great value in applications like surveillance and action recognition. Existing metric learning based methods measure the similarity of sample pairs by learning a metric space in which the positive pairs are closer than negative pairs. However, the appearance features undergo with drastic variation. Person re-identification is a typical small sample problem. It is hard to learn a robust projection of metric subspace that takes all the situations into consideration. The learned metric subspace is usually over-fitting to training dataset due to the strict metric learning constraint. And the hard pairs in training dataset will weaken the discrimination of matching pairs' similarity. To address these problems, a feedback mechanism based iterative metric learning method is proposed. The proposed method introduces a mean distance of multi-metric subspace to deal with the over-fitting problem. The joint discriminant optimal model on feedback top ranks matching pairs will enhance the discrimination of matching pairs' similarity. It is a robust and discriminative distance metric which measures the matching pairs similarity with distances of multiple metric projections learned by a set of training datasets. Aiming to learn the multi-metric subspace, the proposed method gives a feedback mechanism based approach which back propagates the top ranks identification results as pseudo training datasets. The effectiveness of proposed mean distance of multi-metric projection is analyzed and proved theoretically. And extensive experiments on three challenging datasets, VIPeR, GRID and CUHK01 are conducted. The results show that the proposed method achieves the best performance and improves the state-of-the-art rank-1 identification rates by 18.48%, 2.00% and 5.41% on three datasets respectively.

1. Introduction

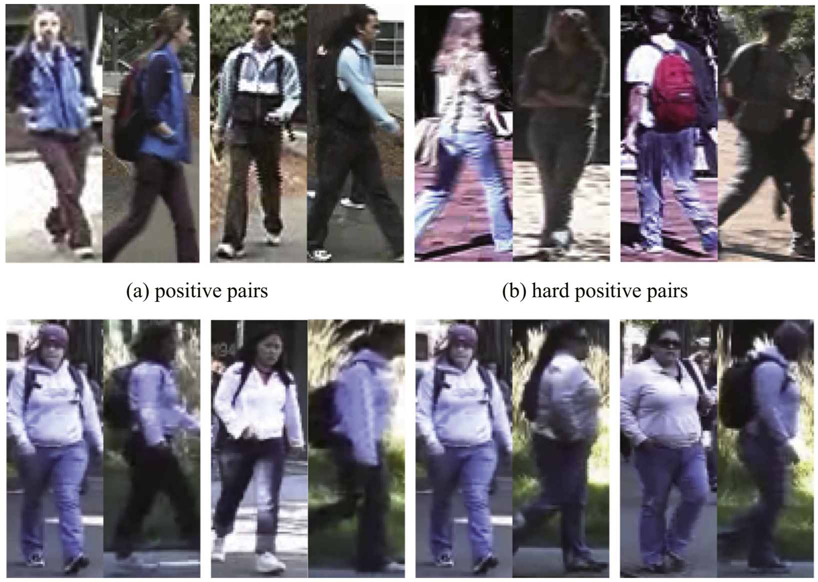
Person re-identification is targeting for matching pedestrians who disappeared in the view of one camera and then appeared in the view of another camera. The event happened in a non-overlapping camera network or in the view of single camera across different times. It is one of important directions in video surveillance that shows a promising prospect in the security applications and objects behavior understanding in multi-camera network, such as across cameras pedestrian tracking, multi-camera event detection and pedestrian retrieval. These important applications make it a hot spot in the field of computer vision [1,2]. With the help of an efficient person re-identification system, it will greatly raise the efficiency of surveillance video search for criminal suspect whom the police are interested in. More generally, it can be helpful to keep the integrity of object detecting and tracking when there is object moving out of view happening.

Given a pair of person images taken from different cameras, re-identification is the process of images matching. Person images across cameras are different from location, shooting time, visual angle, illumination and gait (see Fig. 1). These differences lead to the appearance changes on color and distribution of pixels belonging to the body structure. Therefore, it is difficult to learn a robust metric for person re-identification task with unstable appearance features.

To address the appearance feature model challenges, a plenty of methods have been proposed to design a robust feature of appearance representation and distance metric model. Research on appearance representation model aims to develop a method of image description which is robust to the changes on individuals' appear-

* Corresponding author.
E-mail address: luxq666666@gmail.com (X. Lu).

http://dx.doi.org/10.1016/j.patcog.2017.04.012
0031-3203/



(a) positive pairs (b) hard positive pairs

(c) similar negative pairs

Fig. 1. Person re-identification matching pairs cross camera views in the VIPeR dataset. (a) are normal positive pairs which are similar in appearance. While (b) are non-normal positive pairs which are different in appearance. (c) are negative pairs with similar appearance.

ances and discriminative for identification. Color based features [3–7] and texture based features [7–11] are commonly used in person appearance representation. Traditional features such as Local Binary Patterns (LBP) [8,9] and 21 texture filters feature [10] had been successfully applied to address the person re-identification problem at the early stage. The color information seems to be a discriminative feature in addressing the person re-identification problem due to the limit of low-resolution. Color histogram is the most commonly used color based feature [3–5]. While, traditional color histogram feature is not satisfactory due to the challenges of illumination variations, shadows, gait and visual angle changes. The recently proposed features like ELF [3], SCNCD [6], and LOMO [7] show good performance in person re-identification. Although many appearance feature models have been proposed, it is still not reaching the accuracy of identification demand.

Distance metric method is another crucial issue that draws researchers' attention in person re-identification. Many efforts have paid on the study of metric learning for image classification [5,7,8,12–17]. Distance metric model for person re-identification aims to learn a distance metric under which the positive pairs are closer than negative pairs. There are many metric models have been proposed, such as LDML [12], PRDC [13], KISSME [14], RPML [8], LFDA [5] and NFST [15] etc. Besides, some supervised and unsupervised algorithms have been proposed to address the person re-identification task by selecting the most discriminative features, such as SVM (*Support Vector Machine*) model [18] and Color-based Ranking Aggregation [19].

In this paper, we focus on the distance metric model. In the person re-identification task, metric learning based methods commonly learn a distance metric by forcing the intra-class distance to be much smaller than the inter-class distance. Due to the changes of illumination, visual angle, gait and background, the appearance of person across cameras undergo great variation. There are hard positive pairs that could be quite different both in visual appearance and distance metric (see Fig. 1(b)). These pairs could barely be visually recognized. The negative pair could be quite visual similar even closer than the positive pair (see Fig. 1(c)). The hard positive pairs in Fig. 1(b) are typical visual non-identifiable samples.

However, the metric learning based methods for person re-identification are generally supervised which learn the distance metric with the labeled matching pairs. Existing metric learning based methods (e.g. LFDA and NFST) learn parameters with the positive pairs and negative pairs following the label information strictly. Therefore, the hard positive pairs and similar negative pairs shown in Fig. 1(b) and (c) will cause over-fitting to training dataset. With learned metric model, there is sharp difference of metric distance between negative pairs and positive pairs (see Fig. 2(a)). The positive pairs' metric distances shown in Fig. 2(a) have almost the same distances of quite small values. It is not in conformity with the actual. Therefore, it is difficult to learn a robust metric model to discriminate the similarity of the positive pairs and difference of the negative pairs following sample pairs' visual similarity.

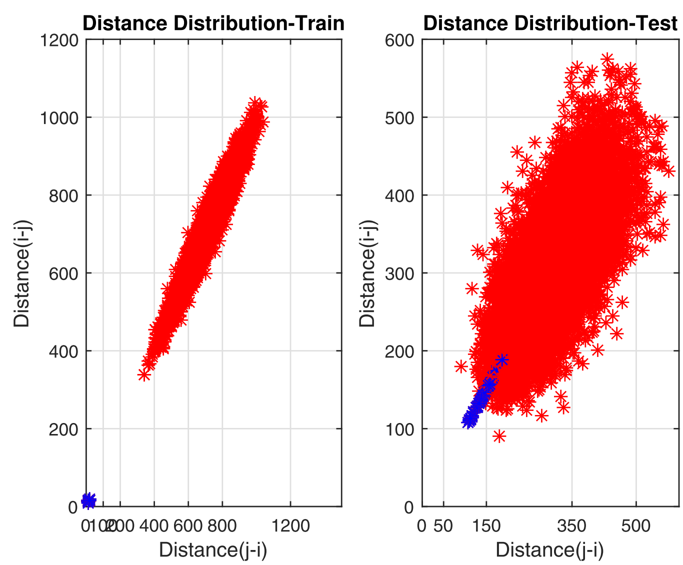


Fig. 2. Illustration of sample pairs distance distribution. Each point stands for the distance of a matching pair. The blue points are positive pairs and the red ones are negative pairs. The value of x coordinate is the distance between the jth person gallery image and the ith person probe image. The value of y coordinate is the distance between the ith person gallery image and the jth person probe image. (a) is sample pairs distance distribution of training data. (b) is sample pairs distance distribution of test data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this paper, we believe that existing metric learning based methods are over-fitting to the training dataset. As shown in Fig. 2, the learned distance metric perform perfect on training dataset. It is not in conformity with the complexity of persons' view changes cross cameras. Traditional metric learning based methods learn the subspace and Mahalanobis distance on a single training dataset. The feature of person is a complex representation with high dimension. It is impossible to take all the conditions into consideration with a limited size dataset. These challenges cause over-fitting problem in metric learning of person re-identification.

The motivation of proposed method is that the positive pairs have relative more stable performance than negative pairs under different metric subspaces learned with existing metric learning based methods. According to the Gaussian hypothesis, the distances of both positive pairs and negative pairs follow Gaussian distributions of different variances. The distribution of positive pairs' distances have smaller variance than negative pairs in the metric subspace. Metric subspaces learned with different training datasets are independent of each other. The distances of negative pairs undergo with great variation under different metric subspaces. To solve the over-fitting problem of traditional metric learning based methods, a mean distance of multi-metric projection is proposed to give a more robust metric approach. The proposed method gives a feedback mechanism and improves the metric learning approach in iterative way. The distance measure results on test dataset are back propagated to join into the metric learning.

Existing metric learning based methods are weak discriminative with the small sample problem and over-fitting problem. While the top ranks identification results contain most of the positive pairs and many similar negative pairs. In this paper, the proposed method aims to learn a metric following samples' visual similarity. The similar negative pairs are more useful than the hard positive pairs for metric learning. There is rarely a method takes the measure information of test dataset into consideration. While, a feedback mechanism based method is proposed in this paper. The measure results of existing metric learning based method are utilized to generate a set of pseudo training datasets. They are back propagated to learn a set of metric subspaces. The image pairs of test dataset are measured by the proposed mean distance of multi-metric projection again. Then the new distance metric results are back propagated for a new set of metric subspaces learning. With the feedback information, a multi-metric subspace is learned and there will be a set of new distance measure results. The new generated distance measure results are back propagated for a new multi-metric subspace learning again. Hence, the proposed method learns the distance metric in iterative way.

The feedback information give a way to deal with the small sample problem. On the other hand, the pseudo training datasets are generated with the top ranks matching pairs. They are true matching pairs and similar pairs. Therefore, the proposed mean distance of multi-metric projection is more powerful in visual similarity metric to against the over-fitting problem and hard pairs problem. The contributions of this paper would be summarized as follows:

(1) The proposed method uses a mean distance of multi-metric projection for similarity measure. The metric learning based method for person re-identification is typical small sample problem with high dimension feature. Existing metric learning based method is over-fitting to single training dataset. The proposed method measures the matching pair's distance with a mean distance of multiple metric subspaces.
(2) A feedback mechanism based iterative metric learning approach is proposed. The top ranks matching pairs are utilized to form a set of multi-metric pseudo training datasets for metric subspaces learning. On the one hand, the feedback mechanism enlarges the training dataset to deal with the small sample problem. On the other hand, the feedback matching pairs are visually similar. They are utilized to learn a multi-projection metric following images' visual similarity to deal with the hard positive pair problem and over-fitting problem.
(3) Extensive experiments are taken on three open datasets, VIPeR, QMUL GRID and CUHK01. Compared with other state-of-the-art metric learning based methods, our method achieves the best performance. The experimental results verify that the proposed method is robust and discriminative for person similarity measure across cameras and more robust to the decrease of training dataset size.

The rest of this paper is organized as follows: in Section 2, a brief introduction of person re-identification is given. The proposed method is given in Section 3. In Section 4, experimental results on three benchmark dataset, VIPeR, GRID and CUHK01 are presented and analyzed. Finally, we make a conclusion in Section 5.

2. Previous works

Person re-identification task is a typical classification problem. To address this problem, many methods have been proposed. Existing methods are roughly divided into appearance feature, metric learning and deep learning based method.

2.1. Appearance feature based methods

As the challenges aforementioned, it is difficult to find a robust appearance representation to be the feature of individual description. Researches on person re-identification seek a representation which could describe the similarity of positive pairs and the difference of the negative pairs. Traditional color based features and texture based features, like LBP, Gablor filters, Schmid fiters are not robust to appearance changes. Liu et al. [20] proposed a Attribute-Restricted Latent Topic Model for person re-identification. They designed an intermediate representation for human appearance modeling. Conventional color and texture based features are clustered to form codebooks. Yang et al. [6] used the color names as the appearance feature which is more robust. They proposed a novel representation, SCNCD, to describe colors of person appearance. The color names feature based method had a much better matching result in person re-identification task than a fusion feature combined with HSV and Lab color histogram and texture feature, LBPs [8,9]. Liao et al. [7] proposed the LOMO feature which combined the color histogram features with the texture histogram features of different scales. It is now a common feature used in person re-identification task.

Except for color and texture based features, the body structure is utilized to adapt to the appearance changes. Dong et al. [21] designed a novel appearance model based on pictorial structures information of human body. And the human body was represented by 6 joint rectangles. When extracting features, they weighted the body structure with the saliency regions. Furthermore, researchers proposed methods matching person across cameras by the saliency detection directly [22–25]. Zhao et al. [22] matched the person by a SVM structure with the learned saliency patches. And they also proposed a novel saliency learning based method [23] which was unsupervised. Adjacency constrained is used to deal with the misalignment problem in this method. And then the pedestrian image saliency was learned to patch match.

2.2. Deep learning based methods

Recently, deep learning has drawn much attention in the area of computer vision and machine learning due to its successes in various pattern recognition and artificial intelligence related tasks. It is powerful in feature representation. Compared to metric learning based methods, the deep learning based method is end to end that works on the raw data directly. It learns the semantic features and recognition model in an uniform deep structure. In the former researches, Bromley et al. [26] proposed a siamese architecture to matching image pairs. It was used in the fields like face recognition firstly. And the early deep learning researches on person re-identification task were commonly based on this architecture. Recently, many deep learning based methods [26–31] have been proposed. Li et al. [27] proposed a filter pairing neural network for person re-identification, while the parameter learning of deep network needs large sample size, Li et al. built the largest dataset for person re-identification in their work. The deep network proposed by Ahmed [28] is similar to the filter pairing neural network, but different in structure and filter size. There is a cross-input neighborhood differences structure in the middle of this deep network. And this novel structure of deep network reached a good recognition rate. The deep learning network proposed by Dong et al. [29] was a siamese convolutional neural network, SCNN. In the framework of Yi et al. [29], the images were cut in an overlapped way to three body structures. These three different overlapped parts were learned in three same network jointly. This way of learning sample with joint networks of overlapped body structures is now a common framework in person re-identification task.

In 2015, a bilinear CNN (B-CNN) framework [32] had been proposed in deep learning based method research on fine-grained visual recognition. Roychowdhury et al. [30] then proposed a multiregion B-CNN network for person re-identification. In this framework, the sample was also cut into three parts and learned with a joint networks. A multiregion bilinear architecture was proposed as the layer before full connection layer in the B-CNN based network for person re-identification.

Generally, the deep learning network based method is a promising direction of research on person re-identification problem. While the deep learning based methods need large sample size. And it costs large computational quantity.

2.3. Metric learning based methods

Metric learning based methods were the best performance and commonly used methods in person re-identification before deep learning models being proposed to solve the person re-identification problem. Many methods like LDML [12], PRDC [13], KISSME [14], Relaxed Pairwise Metric Learning(RPML) [8], LFDA [5], XQDA [7], NFST [15] have been proposed. Metric learning based methods aim to learn a Mahalanobis distance metric for positive and negative pairs classification. Zheng et al. [13,33] proposed the RDC and PRDC in re-identification problem which are the state-of-the-art algorithms. RDC treated the person re-identification task as a relative distance comparison problem to set up optimization criterion of relative distance. And an iterative method was given to learn the projection matrix. PRDC improved RDC by transforming the relative distance comparison optimization problem of minimizing intra-class distance and maximizing inter-class distance into a matching probability problem based on the relative distance. Then, Kostinger et al. [14] proposed a novel method based on a Gaussian hypothesis. The Gaussian hypothesis assumed that the positive and negative pairs followed two different multivariate Gaussian distributions and both zero-mean. Then they proposed the KISSME method which programed the metric matrix by the probability defined by the Gaussian hypothesis. The XQDA [7] algorithm proposed by Liao et al. was same to KISSME that was following the Gaussian hypothesis. The XQDA solved the re-identification task based on a designed handcraft feature, LOMO. XQDA combined the FDA algorithm and KISSME to learn the distance Metric. It learned a subspace projection similar to the FDA algorithm and the KIEEME metric model was introduced to learn a Mahalanobis distance. The XQDA is state-of-the-art algorithm in person re-identification task. Zhang et al. [15] proposed a novel null foleysammon transform based method(NFST). The NFST based method solved the re-identification problem by learning a discriminative null space which was similar to the FDA method. Finally, the NFST method was improved to be a non-linear model by the kernelisation process. The NFST kernelisation model is more appropriate in person re-identification distance measure due to the non-linearity of person appearance features. Besides, many other methods have been proposed, including the dictionary learning based methods [34,35] and learning to rank based methods [19,36]. The dictionary learning based methods could be treat as feature transform methods. They assumed that the images of same person in views across cameras have same representation. With the same representation, the raw images could be reconstructed with different dictionary. Then the appearance variations of sample pairs would be bridged together with the learned dictionary and representation. Liu et al. [34] proposed a semi-supervised dictionary learning based method. It learnt the two dictionaries for sample pairs in the views of two non-overlapped cameras in one unite framework. The Least Square Semi-Coupled Dictionary Learning algorithm [35] proposed by Ying Zhang et al. was based on the SCDL method. And the authors combined the dictionary learned by LSSCDL algorithm with the weight vectors learned by SVM. Learning to rank based methods work directly on the recognition results. The CBRA algorithm [19] proposed by Prates et al. combined ranking results of several features and learns a ranking model by ranking aggregation method.

3. Feedback based iteration metric learning

In this section, the formulation and solution of proposed method are given (see Fig. 3). Firstly, the person re-identification problem is defined in Section 3.1 Secondly, the way of metric learning is introduced in Section 3.2. Thirdly, the over-fitting problem and hard pairs problem are described in Section 3.3. Finally, the formulation of proposed method is introduced in Section 3.4.

3.1. Definition of person re-identification problem

The basic theory of person re-identification problem is descripted as following:

Let x_i^g denotes the ith person image feature of the gallery set and x_i^p denotes the ith person image feature of the probe set. g means gallery image set and p means probe set. A matching pair is consist of two images, one gallery image and one probe image. Given training dataset $X \in \mathbf{R}^{2d \times n}$, d denotes the dimension of image feature and N denotes matching pairs' number of training dataset.

Assuming that there are N samples of c classes in gallery image set $\{x_i^g\}$ which means that there are N images belonging to c persons under camera A. There are $n_1, n_2, n_3, \ldots, n_c$ samples for each class respectively, $n = n_1 + n_2 + n_3 + \ldots + n_c$. Under the single-shot situation all the numbers of classes are equal to 1, $n_1 = n_2 = n_3 = \ldots = n_c$. The probe image set $\{x_i^p\}$ is the same situation with gallery image set.

In the task of person re-identification, the image feature of person appearance is a typical high dimension vector. The sample difference of the ith person between gallery set and probe set could be denoted as $\triangle = x_i^g - x_i^p$. (x_i^g, x_i^p) is a positive pair belonging to

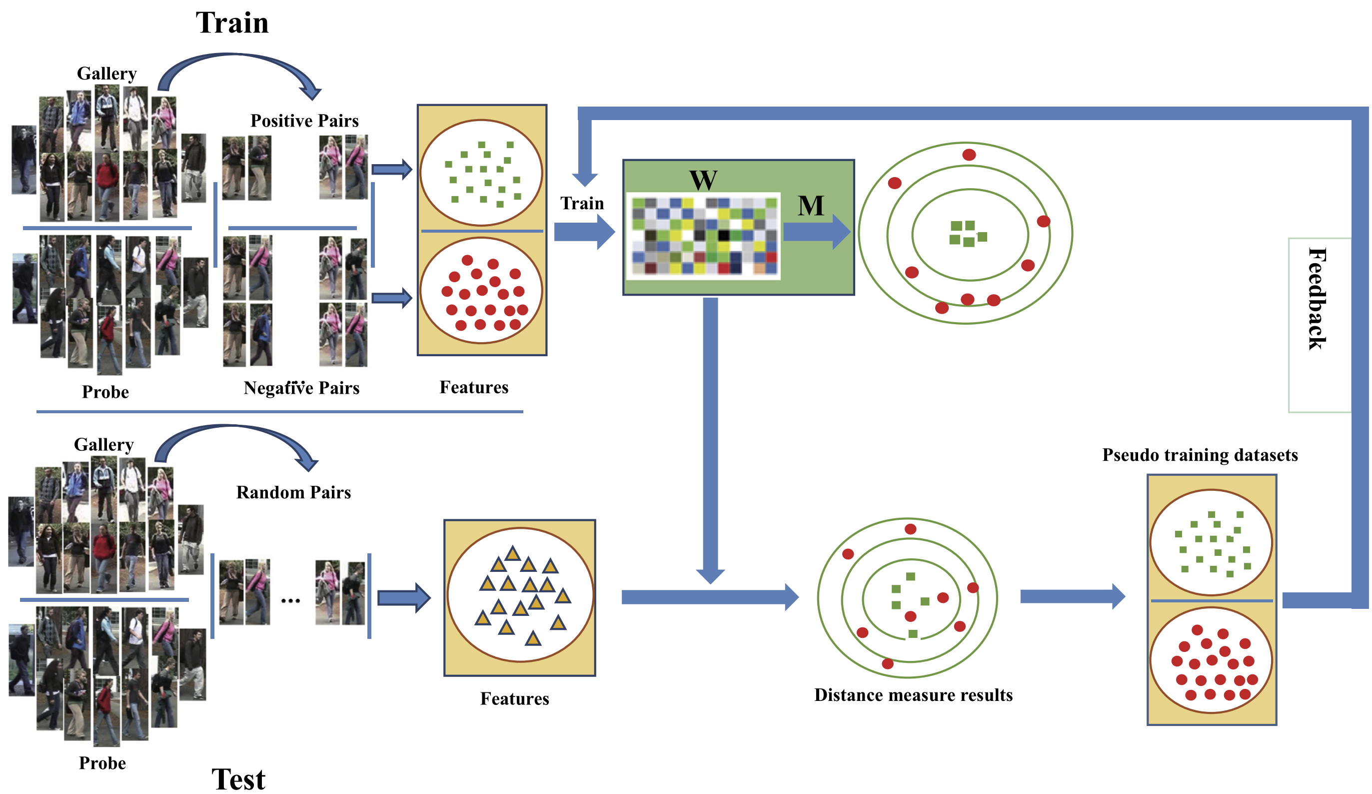


Fig. 3. Diagrammatic sketch of the proposed iterative metric learning method. The training set is used to learn metric subspace. Then test set is measured by the learned metric subspace and the top ranks results are used to form the pseudo training datasets. They will be back propagated to join into multi-metric subspace learning.

same person. Generally, the sample difference of random pairs (x_i^g, x_i^p) is denoted as $\triangle_{ij} = x_i^g - x_i^p$. The distance of sample pairs (x_i^g, x_i^p) is a measure of the difference vector. Simply, the distance is measured by squared Euclidean distance as following:

$$d(x_i^g, x_j^p) = (x_i^g - x_j^p)^T (x_i^g - x_j^p). \quad (1)$$

Metric learning based methods rely on the effects of two factors, representation of features and discrimination of distance measure. squared Euclidean distance (1) is not satisfied in solving person re-identification problem. Researchers on metric learning based method are working on learning a Mahalanobis distance. Therefore, the distance function of sample pair measure is:

$$d(x_i^g, x_j^p) = (x_i^g - x_j^p)^T \mathbf{M} (x_i^g - x_j^p). \quad (2)$$

where distance (2) equal to (1) when $\mathbf{M} = \mathbf{I}$.

3.2. Metric learning

According to the definition of person re-identification problem in 3.1, the task of metric learning is learning $\mathbf{M}$. The metric matrix $\mathbf{M}$ in (2) is positive semi-definite. The matrix $\mathbf{M}$ could be decomposed with eigenvalue decomposition as following:

$$\mathbf{M} = \mathbf{P}\mathbf{\Lambda}\mathbf{P^T}. \quad (3)$$

where $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues as main diagonal element. The column elements of $\mathbf{P}$ are orthonormal eigenvectors. Let $\mathbf{W} = \mathbf{P}\mathbf{\Lambda}^{\frac{1}{2}}$, the decomposition of metric matrix $\mathbf{M}$ could be also defined as following:

$$\mathbf{M} = \mathbf{W}\mathbf{W^T}. \quad (4)$$

Therefore, the Mahalanobis distance of matching pairs are transformed into the form as following:

$$d(x_i^g, x_j^p) = (x_i^g - x_j^p)^T \mathbf{W}\mathbf{W^T} (x_i^g - x_j^p). \quad (5)$$

where $\mathbf{W}$ is treated as a projection to a metric subspace. The person re-identification metric learning problem is transformed into a subspace learning problem. The subspace learning aims to learn the projection matrix $\mathbf{W}$ by minimizing the intra-class distance and maximizing the inter-class distance. Targeting at the optimization criterion of minimizing intra-class distance and maximizing inter-class distance, a Foley-sammon Transfer (FST) method is exploited to learning the subspace.

Given the training set cross view $X = (x_i^g, x_j^p)$, a subspace $W = (w_1, w_2, w_3, \ldots, w_c) \in \mathbf{R}^{d\times r}$ is learned. The intra-class distance and inter-class distance are defined as following:

$$S_w = \sum_{i=1}^{c} \frac{n_i}{n} S_i. \quad (6)$$

$$S_b = \sum_{i=1}^{c} \frac{n_i}{n} (u_i - u)(u_i - u)^T. \quad (7)$$

where S_w is within-class scatter and S_b is inter-class scatter. S_i denotes the covariance of the ith class.

The subspace learning is a discriminant optimal problem by maximizing the Fisher discriminant criterion:

$$\max_w J(w) = \frac{w^T S_b w}{w^T S_w w}. \quad (8)$$

The Fisher optimal criterion (8) is equivalent to:

$$\max_w w^T S_b w, \ \ s.t. \ w^T S_w w = 1. \quad (9)$$

The optimization problem (9) could be solved by Lagrange multiplier method. The optimization of formula (9) then can be transformed to be the problem of solving generalized eigenvalue:

$$\lambda w = S_w^{-1} S_b w. \quad (10)$$

Then, the optimization problem (9) is formulated as a generalized matrix eigenvalue problem. The eigenvector w_1, which is corresponding to the maximum eigenvalue of $S_w^{-1}S_b$, is the maximum solution of optimal criterion (9). And the eigenvector corresponding to the second eigenvalue of $S_w^{-1}S_b$ is the sub-optimal solution of optimal criterion (9). Therefore the subspace could be learned by the solution of former m eigenvectors $\mathbf{W} = (w_1, w_2, w_3, \ldots, w_m)$.

Furthermore, Liao et al. introduce the KISSME approach into the person re-identification problem and propose the XQDA method. The KISSME approach learns the Mahalanobis distance with the positive pairs and negative pairs covariances of training daset based on the Gaussian hypothesis. In the person re-identification problem, the matching pair features $\triangle = x_i^g - x_i^p$ are defined with two multivariate Gaussian distributions. $\triangle \sim N(\boldsymbol{\mu}_p, \ \boldsymbol{\Sigma}_p)$ when (x_i^g, x_j^p) is positive pair. While $\triangle \sim N(\boldsymbol{\mu}_n, \ \boldsymbol{\Sigma}_n)$ when (x_i^g, x_j^p) is negative pair. Both $\boldsymbol{\mu}_p$ and $\boldsymbol{\mu}_n$ are mean of feature difference vectors of same high dimension. $\boldsymbol{\Sigma}_p$ and $\boldsymbol{\Sigma}_n$are covariance matrices of same large size. XQDA method introduces another metric matrix which is learned by the covariance of training set in the subspace. The distance in XQDA is as following:

$$d(x_i^g, x_j^p) = (x_i^g - x_j^p)^T \mathbf{W}\mathbf{H}\mathbf{W}^\mathbf{T}(x_i^g - x_j^p). \tag{11}$$

where $\mathbf{H} = ((\mathbf{W}\boldsymbol{\Sigma}_\mathbf{p}\mathbf{W}^\mathbf{T})^{-\mathbf{1}} - (\mathbf{W}\boldsymbol{\Sigma}_\mathbf{n}\mathbf{W}^\mathbf{T})^{-\mathbf{1}})$, $\boldsymbol{\Sigma}_p$ denotes the covariance of positive training pairs. $\boldsymbol{\Sigma}_n$ denotes the covariance of negative training pairs.

3.3. High-dimension small sample metric learning problem for person re-identification

In person re-identification task, over-fitting problem and hard positive pairs problem are two main difficulties for metric learning. The over-fitting problem is caused by small sample size and high feature dimension. The metric subspace, (**W, H**), learned with a small training dataset is incomplete. The learned metric subspace will over fit to the training dataset due to the strict constraint of metric learning model. Hard positive pairs problem is caused by the positive matching pairs which are unrecognizable. The metric learning based methods learn distance metrics following the label information strictly. The hard positive pairs will weaken the discrimination of matching pairs' visual similarity. In this paper, a mean distance of multi-metric projection based method is proposed to deal with the over-fitting problem. Meanwhile, a joint discriminant optimal model is proposed which learns a multi-metric projection with feedback top ranks matching pairs to deal with the hard positive pairs problem. The motivation of proposed method is to separate the negative pairs with multiple distances learned by several independent metric subspaces. The joint discriminant optimal model will enhance the discrimination of matching pairs' visual similarity because of that the feedback top ranks matching pairs are visually similar.

As shown in Fig. 2, the distribution of positive pairs distances vary smaller variation than negative pairs under the incomplete metric subspace. The distances of the positive pairs are much more stable than the negative pairs. Assume that *X, Y* denote the training dataset and test dataset respectively. And there is a population *T* of all kinds of persons views that $X, Y \subset T$. $\triangle$ denotes the matching feature vector of train dataset, $\boldsymbol{\Delta} = \{\Delta = x_i^g - x_j^p | x_i^g, x_j^p \in X\}$. O denotes the matching feature vector of test dataset, $\mathbf{O} = \{O = y_i^g - y_j^p | y_i^g, y_j^p \in Y\}$. Since training and testing datasets follow the same distribution, for arbitrary real number $b > 0$, $\exists c \in R$, $c > 0$, $\forall O_k \in \mathbf{O}$, if there is $\Delta_k \in \boldsymbol{\Delta}$ meets that $\|O_k - \Delta_k\|_2 < c$, $|d(O_k) - d(\Delta_m)| < b$. Where $d(X) = X^T\mathbf{W}\mathbf{H}\mathbf{W}^\mathbf{T}\mathbf{X}$, $\mathbf{W}$ is the projection matrix to the metric subspace learned by training dataset *X*. That is to say, with the metric subspace (**W, H**) learned with training dataset, positive or negative pair of test dataset would be measured with a discriminative distance if there is a pair in training dataset similar to it.

But the person re-identification problem is a small sample problem with high-dimension feature. It is impossible to get the complete dataset that considers all the person appearances and view changes to learn the person metric space. As shown in Fig. 2, the distribution of test dataset matching pairs $\boldsymbol{\Delta}$, is quite different from the distribution of training dataset matching pairs $\mathbf{O}$, under the metric subspace.

Noticed that $w^T S_w w = 1$ in fisher criterion (9) is equivalent to a less determined when the elements of projection vector are treated as variables. The projection vector w is over-fitting to training dataset with the strict optimal criterion (9) due to its high dimension. Then we call the subspace (**W, H**) incomplete metric subspace. Since the complexity of person re-identification pairs, overfitting to the weak pairs like hard positive pairs and similar negative pairs leads to a weak distance measure on test dataset. The test positive pairs distance values are different from training positive pairs. Moreover, there are some negative pairs with a better distance than positive pairs. In this paper, we assume that the distances of positive pairs and negative pairs follow two different univariate Gaussian distributions,

$$d_p(\Delta) \sim N(\mu_1, \sigma_1), d_n(\Delta) \sim N(\mu_2, \sigma_2). \tag{12}$$

where d_p denotes the positive pair distance and d_n denotes the negative pair distance. $\mu_1 \ll \mu_2$, $\sigma_1 \ll \sigma_2$.

As the training dataset is select by randomly sampling and these samples are independent of each other, the metric subspace parameters (**W, H**) learned with different training datasets are independent. Assume that a set of training datasets are given and $(\mathbf{W}_1, \mathbf{H}_1), (\mathbf{W}_2, \mathbf{H}_2), \ldots, (\mathbf{W}_m, \mathbf{H}_m)$ are corresponding metric subspace parameters. The distances of person matching pair under these metric subspace parameters are independent and following the same distribution. Then the similarity of matching pair is measured with a mean distance of a set of metric subspaces:

$$\begin{aligned} d_{\{\mathbf{W}\}}(x_i^g, x_i^p) &= \frac{1}{m}\sum_{i=1}^{m} d_i(x_i^g, x_i^p) \\ d_i(x_i^g, x_i^p) &= (x_i^g - x_i^p)^T \mathbf{W}_i\mathbf{H}_i\mathbf{W}_{i^T}(x_i^g - x_i^p) \end{aligned} \tag{13}$$

Since $d_i (1 \le i \le m)$ are independent and identically distributed, $d_{\{W\}}$ obeys normal distribution. For the distance defined by formula (13), when m is large enough, $\forall \Delta_p = x^g - x^p$, $\forall \Delta_n = x^g - x^p$, the relationship of positive pairs and negative pairs is defined in (14). Therefore, the mean distance of multi-metric projection based method could solve the over-fitting problem if there were enough sub-training datasets.

$$\begin{aligned} d_{\{\mathbf{W}\}}(\Delta_p) &= \frac{1}{m}\sum_{i=1}^{m} d_{p,i}(\Delta_p) \approx \mu_1 \ll \mu_2 \approx \frac{1}{m}\sum_{i=1}^{m} d_{n,i}(\Delta_n) \\ &= d_{\{\mathbf{W}\}}(\Delta_n). \end{aligned} \tag{14}$$

However, the person re-identification is a typical small sample problem. It is impossible to get more training datasets for multiple metric learning with a given datasets. In this paper, feedback mechanism based method is proposed to learn the multi-metric subspaces. The top ranks matching pairs are utilized to form a set of pseudo training datasets for multiple metric learning.

Furthermore, there is a certain percentage of hard positive pairs in the training dataset. The training dataset could be divided into $X = \{X_1, X_2\}$. Where $X_1 = \{(x_k^g, x_j^p), k = 1, \ldots, m_1\}$, $X_2 = \{(x_l^g, x_j^p), l = 1, \ldots, m_2\}$, $m_1 + m_2 = N$. X_1 is consist of the normal positive pairs and corresponding negative pairs. X_2 is consist of the hard positive pairs and corresponding negative pairs. The optimization problem (9) can be transformed to be the form shown in (15):

$$\max_{w} w^T S_b w, \ \ s.t. \ w^T(S_{w1}+S_{w2})w = 1. \tag{15}$$

Where S_{w1}, S_{w2} are within-class scatters corresponding to X_1, X_2. Existing methods identify persons across camera views based on the appearance feature. The hard positive pairs are visually unrecognizable. The second part of the constraint in (15), $w^T S_{w2} w = 1$, will weaken the discrimination of metric learning model. And the strict constraint will cause over-fitting problem to the training dataset.

3.4. Feedback based iterative metric learning

As it is mentioned before, the over-fitting problem leads to a weak identification ability of the learned metric subspace. In Section 3.3, a mean distance of multi-metric projection based method is proposed and proved to solve this problem. While, it is hard to get enough training datasets for person re-identification task which is a typical small sample problem. The size of complete person views dataset is endless in theory. It is impossible to learn a mean distance of multi-metric projection with the given small size training dataset. In this paper, a feedback based iterative metric learning approach is proposed to achieve a better re-identification performance. The test dataset distance measure results are used to generate pseudo training datasets which are back propagated to improve metric learning.

(1) Feedback Mechanism based Pseudo Training Datasets Learning: In the proposed method, $X = \{x_i^g, x_j^p\}$ and $Y = \{y_i^g, y_j^p\}$ are defined as training dataset and test dataset respectively. The samples in training dataset are ordered which means that (x_i^g, x_j^p) correspond to same person when $i = j$. While the matching information of Y is blind. The samples (y_i^g, y_j^p) in test dataset are unordered. The feedback based approach firstly learns the initial metric subspace parameters ($\mathbf{W}_0$, $\mathbf{H}_0$) with training dataset $X = \{x_i^g, x_j^p\}$. Secondly, the test dataset $Y = \{y_i^g, y_j^p\}$ is measured with the initial metric subspace. The distance matrix is denoted as $D(i, j) = d(y_i^g, y_j^p|\mathbf{W}_0, \mathbf{H}_0)$. The distance matrix D then used to form pseudo training datasets. The top m ranks matching results are used to form m pseudo training datasets as following.

$$\begin{aligned} &Y_1^{\cdot} = \{y_i^g, y_{1,j}^p\}, y_{1,j}^p = \arg\ rank(D(i,j)) = 1 \\ &Y_2^{\cdot} = \{y_i^g, y_{2,j}^p\}, y_{2,j}^p = \arg\ rank(D(i,j)) = 2 \\ &\qquad\ldots\ldots \qquad\qquad\qquad \ldots\ldots \\ &Y_m^{\cdot} = \{y_i^g, y_{m,j}^p\}, y_{m,j}^p = \arg\ rank(D(i,j)) = m \end{aligned} \tag{16}$$

where $Y_1^{\cdot}, Y_2^{\cdot}, \ldots, Y_m^{\cdot}$ denote m pseudo training datasets. They are back propagated for metric learning enhancement. $rank(\bullet)$ denotes the rank calculation of matching pair. $y_{m,j}^p$ denotes the rank-m matching image corresponding to y_i^p. The pseudo training dataset $Y_m^{\cdot}$ is consist of all the ranks-m matching pairs of test dataset.

(2) A Mean Distance of Multi-metric Projection Learning: The proposed method aims to enhance the similarity discrimination of metric learning model and ease the over-fitting problem to single training dataset. It generates a set of pseudo training datasets,$(Y_1^{\cdot}, Y_2^{\cdot}, \ldots, Y_m^{\cdot})$, with the top ranks matching pairs. The similarity distance is measured with a mean distance shown in (17). It is defined on a set of metric subspaces learned with the pseudo training datasets.

$$d_{\{\mathbf{W}\}}(\Delta_p) = \frac{1}{m}\sum_{i=0}^{m} O^T \mathbf{W}_i \mathbf{H}_i \mathbf{W}_i^T O. \tag{17}$$

These metric subspaces are learned with a joint discriminant optimal problem over a set of training datasets. A joint Fisher discriminant criterion is defined in (18).

$$\max_{\{\mathbf{W}\}} J(w_0, \ldots, w_m) = \frac{w_0^T S_b w_0}{w_0^T S_w w_0} + \sum_{i=1}^{m} \frac{w_i^T S_{b,i} w_i}{w_i^T S_{w,i} w_i} \tag{18}$$

where $S_{b,\,i}$ and $S_{w,\,i}$ are inter-class scatter and within-class scatter corresponding to the ith training datasets respectively.

The joint Fisher discriminant criterion (18) then can be transformed to (19).

$$\begin{aligned} \max_{w}\ \ & w_0^T S_b w_0 + \sum_{i=1}^{m} w_i^T S_{b,i} w_i, \\ & w_0^T (S_{w1}+S_{w2}) w_0 = 1. \\ & w_1^T S_{w,1} w_1 = 1. \\ s.t.\ \ & w_2^T S_{w,2} w_2 = 1. \\ & \qquad\ldots\ldots \\ & w_m^T S_{w,m} w_m = 1. \end{aligned} \tag{19}$$

Then, a set of metric subspaces, $(\mathbf{W}_1, \mathbf{H}_1), (\mathbf{W}_2, \mathbf{H}_2), \ldots, (\mathbf{W}_m, \mathbf{H}_m)$ are learned with the pseudo training datasets. The Mahalanobis distances are learned with corresponding projection and datasets independently by (10).

As shown in (19), the first constraint on training dataset are transformed to $w^T(S_{w1}+S_{w2})w = 1$ due to the hard positive pairs problem in (15). The proposed method learns a multi-metric subspace projection with the feedback pseudo training datasets. These feedback pseudo training datasets are top ranks matching pairs under initial metric subspace (**W, H**). That is to say, they are all of visually similar pairs. Therefore, the proposed method enhances the discrimination on visual similarity and ease the effect of hard positive pairs. Besides, the metric subspaces learned with the pseudo training datasets over-fitting to different datasets which are visually similar. The mean distance of multi-metric projection will ease the over-fitting to single training dataset. So, the proposed mean distance of multi-metric projection is more powerful in similarity distance metric dealing with the over-fitting problem.

Notice that the independence of training datasets for multi-metric subspaces learning is an important assumption of proposed method. For person re-identification task, the problem is addressed on the relationship of image pairs. Every image pair represents an independent appearance changing. Therefore, different image pairs are independent with each other. Meanwhile, the feedback training datasets are generated from randomly selected persons. The matching results are determined by the visual similarity under initial metric subspace.

(3) Iterative Multi-metric Learning: A new distance matrix $\mathbf{D}_{new}$ will be calculated by solving the joint discriminant optimal problem in (18). A new set of pseudo training datasets are generated by (16). Therefore, $(\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_m, \mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_m)$ can be learned in an iterative way defined in (20).

$$\begin{aligned} &Y_{r,k}^{\cdot} = \{y_i^g, y_{r,i}^p\}, y_{r,i}^p = \arg\ rank(\mathbf{D}_k(i,j)) = r \\ &\mathbf{W}_{r,k} = g(Y_{r,k}^{\cdot}) \\ &\mathbf{H}_{r,k} = ((\mathbf{W}_{r,k}^T \boldsymbol{\Sigma}_{r,p} \mathbf{W}_{r,k})^{-1} - (\mathbf{W}_{r,k}^T \boldsymbol{\Sigma}_{r,n} \mathbf{W}_{r,k})^{-1}) \\ &\mathbf{D}_{k+1} = f(Y|\mathbf{W}_{1,k}, \ldots, \mathbf{W}_{m,k}, \mathbf{H}_{1,k}, \ldots, \mathbf{H}_{m,k}) \end{aligned} \tag{20}$$

where $Y_{r,k}^{\cdot}$ denote the pseudo training datasets which is generated with similarity distance matrix $\mathbf{D}_k$ in the kth iteration. $\mathbf{D}_k$ is the similarity distance matrix in the kth iteration. $\boldsymbol{\Sigma}_{r,\,p}$ and $\boldsymbol{\Sigma}_{r,\,n}$ are covariance of positive pairs and negative pairs corresponding to the rth rank pseudo training dataset in the kth iteration. $\mathbf{w}_{r,\,k}$ and $\mathbf{H}_{r,\,k}$ are multi-metric subspace parameters in the kth iteration.

Summarization of the proposed method is shown in Algorithm 1.

Algorithm 1 Iterative learning method.

Input: **X**: training set; **Y**: test set;
M: top ranks number for feedback pseudo training dataset.
Initialization: $\mathbf{W}_0, \mathbf{H}_0$
1: initialize N=5, M=10,$\mathbf{W}_0$=0, $\mathbf{H}_0$=0 and $\mathbf{D}_0$=0;
2: learn and with X by (10) and (11);
3: measure sample pairs similarity over Y to calculate the distance matrix $\mathbf{D}_0$ by (11);
4: initialize i=0;
5:**Repeat**
6: generate pseudo training dataset, $Y_1^{\cdot}, Y_2^{\cdot}, \ldots, Y_m^{\cdot}$, with $\mathbf{D}_i$;
7: calculate the multi-metric subspaces $(\mathbf{W}_1, \mathbf{H}_1), \ldots, (\mathbf{W}_m, \mathbf{H}_m)$ by(18);
8: calculate similarity distance matrix $\mathbf{D}_{i+1}$ by (16);
9: i = i+1;
10: **until** Maximum iteration number or convergence achieves.

4. Experiments

Experiments on three datasets, VIPeR [37], GRID [38] and CUHK01 [39] are conducted to validate the effectiveness of proposed method. There are three parts in this sections, datasets, competition methods and experiment results. In the first part, the two single-shot datasets chosen in this paper are introduced. In the second part, seven state-of-the-art methods are introduced to make a convincing comparison of re-identification effects. In the third part, we give the detailed experimental results compared with state-of-the-art metric learning based methods.

4.1. Datasets and settings

To evaluate the validity and efficiency of proposed method, the feedback mechanism based iterative metric learning method is test on three publicly datasets as following:

VIPeR is an open datasets that is widely used in validation of algorithm effectiveness for person re-identification problem. This dataset includes 1264 pictures of 632 persons captured in the views of two non-overlapped cameras outdoor. All the images are normalized in the same size of 128× 48. There are two images for each person, one gallery image and one probe image that are shown in Fig. 1.

GRID is a typical dataset with small sample size. It is a very challenging dataset in person re-identification problem. This dataset is captured in a subway station by eight cameras. The whole dataset could divided into two parts, the first part is consist of 500 images for 250 persons captured in the views of two non-overlapped cameras that is similar to the VIPeR. The second part is consist of 775 images for 775 persons. There is only one image for one person. The images in second part are used as interference in the experiment of algorithm test.

CUHK01 is a multi-shot dataset. There are more than one image for each person in the view of one camera. The images in CUHK01 are normalized into the same size of 160× 60. There are 971 persons in this dataset. Each person has two images in one camera view.

4.2. Comparison algorithms

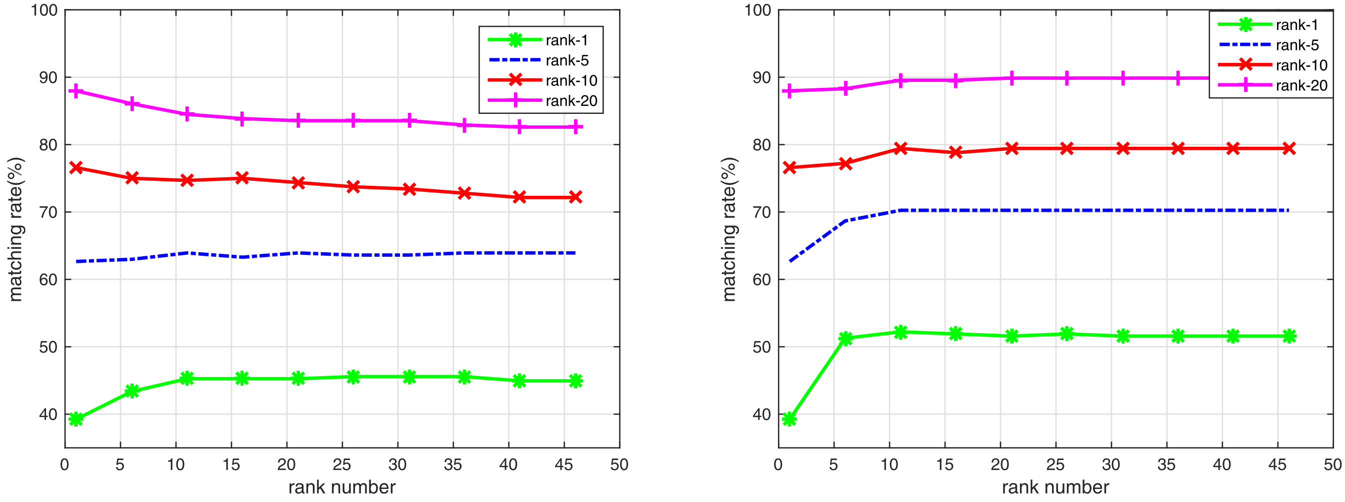
To verify the effectiveness of the proposed method, comparison experiments will be done with existing state-of-the-art methods. The comparison algorithms chosen in this paper are mainly metric learning based method for the reason that the proposed method is also metric learning based. Some state-of-the-art metric learning based methods are introduced as following:

1. **RDC** [33] is a typical metric learning based method in early researches for person re-identification problem. It is commonly used in the comparison experiments. RDC solve the re-identification task by transform the distance measure problem into a subspace learning problem. And an iterative optimization algorithm was proposed to learn a set of orthogonal projection vectors that form the subspace.
2. **ITML** [40] formulate the similarity of person images distance with information-theoretic approach. It also solve the person re-identification problem by learning Mahalanobis distance. Different from traditional metric learning methods, ITML learn a set of Manalanobis distances to approach to the existing Mahalanobis distance defined by the covariance matrix of a multivariate Gaussian distribution by an information-theoretic setting.
3. **KISSME** [14] is based on a distribution hypothesis. This method learning a Mahalanobis distance for person re-identification problem based on a distribution hypothesis which assumes the difference vector of positive sample pairs and positive sample pairs features are following two different Gaussian distribution.
4. **LMNN** [39] is a classic metric learning based method which learns a Mahalanobis distance for classification problems. This method is proposed for face recognition task. The distance is learned by forcing the samples close to its k nearest neighbours and be far away with the rest samples.
5. **Improved Deep** [28] is a classic deep learning based method. It improves the SCNN deep structure with a neighborhood differences structure. The structure utilize a neighborhood difference operation to generate the similarity map for identification features learning. The final deep features are measured by softmax function.
6. **MLAPG** [41] considers the metric learning for person re-identification from the smoothness of the metric. It learn the Mahalanobis distance by formulate a log-logistic loss function with positive semi-define constraint. Besides, considering the balance between positive pairs and negative pairs, MLAPG method give different weights to positive pairs and negative pairs with an asymmetric weighting strategy.
7. **QXDA** [7] is one of the best performance metric learning based methods. The authors propose a handcraft feature and a metric distance and achieve a good performance for person re-identification problem. The proposed metric distance method introduce LDA and KISSME approaches.

4.3. Appearance features

Appearance feature is crucial to the re-identification rate. Persons visual appearance is changing in spatial distribution for the reasons of visual angle and gait variations (visual angle and gait variations will cause the changing of persons body structure which leads to the changing of person appearance spatial distribution in pixel level). The color and texture based statistical features with horizontal strips or patches structure are better performed in appearance representation to against the geometric transform.

In this paper, the LOMO feature designed by Liao et al. [7], is extracted for person appearance description. LOMO is a combined feature with color and texture feature. The color based features are 8× 8 × 8-bin joint HSV histograms of overlapping patches. The texture based features are two SILTP histogram features of different scales. The same features are extracted in the down-sample image of average pooling. That is to say, the LOMO extracts features in multi-scale by building a pyramid representation. The multi-scale pyramid representation enable LOMO to get stable features in different scales. The proposed method and comparison methods are applied to the LOMO appearance representation extracted by the features aforementioned.



(a) Results over iteration numbers

(b) Results over top ranks numbers

Fig. 4. The sensitivity analysis over parameters on dataset VIPeR. (a) is over different iteration numbers and (b) is over different top ranks numbers.

4.4. Experimental results

(1) **Experiments on parameters**: In this section, we first conduct experiments over different parameter settings. The sensitivity experiments over parameters are test on the VIPeR dataset. The dataset is divided into two equal size parts by random sampling. Half of the sample set is used for training and the other half is used for test. Iteration number and top ranks number are two main parameters of proposed method. Re-identification rates with different iteration numbers are shown in Fig. 4(a). The value of feedback top ranks number is fixed to 1. Fig. 4(a) shows the rank-1, rank-5, rank-10 and rank-20 identification rates changing curves with the iteration numbers range from 1 to 46. As shown in the figure, the rank-1 identification rates is increasing along with the increase of iteration number. When iteration number is larger than 10, the rank-1 re-identification is stable. While the rank-5 re-identification rate is stable to the iteration number changing that there is little change of rank-5 identification rate along with the increase of iteration number. Rank-10 and rank-20 identification rates descend slightly and they are stable too when the iteration number is larger than 10. The top ranks re-identification rates, especially the rank-1 re-identification rate, are the most important thing in person re-identification task. The proposed method is effective and improves the matching results significantly as shown in Fig. 4(a).

Top ranks number is another crucial parameter of the proposed method. The sensitivity over feedback rank number are shown in Fig. 4(b). Rank-1, ranl-5, rank-10 and rank-20 re-identification rate curves are plotted in Fig. 4(b) along with the feedback rank number increasing. The iteration number is set to 1. As shown in Fig. 4(b), rank-1 re-identification rate increases obviously when the feedback top ranks number increases from 1 to 10. And it is stable when feedback top ranks number is larger than 10.

Iteration number and top ranks number have great influence on the identification results. While there is obvious difference between these two parameters. The re-identification rates over top ranks number performs better than iteration number. For the reason that the top ranks matching pairs under existing distance metric are visually similar pairs. With the feedback rank number increasing, the proposed method brings massive visually similar pairs into metric learning. Nearly all true positive pairs join into the metric learning together with massive similar negative pairs.

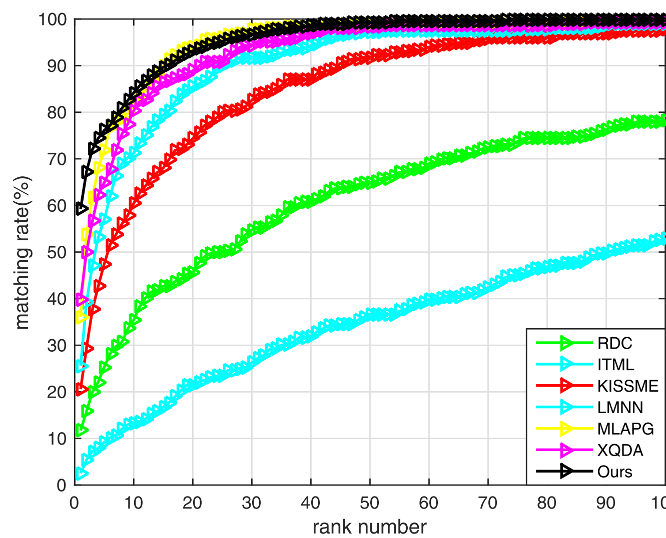


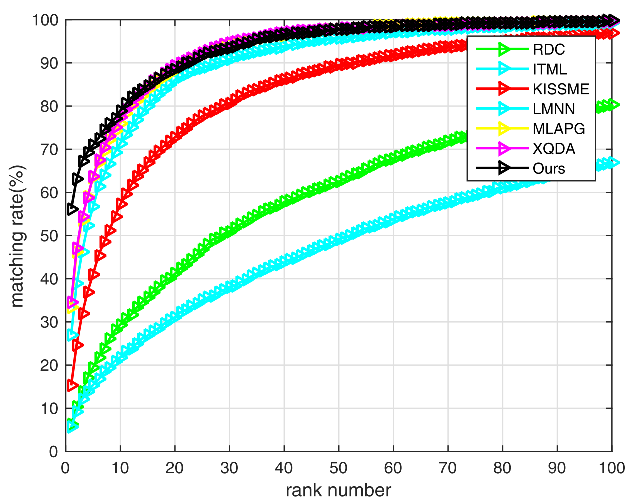
Fig. 5. Comparing experiments results on dataset VIPeR. The training dataset size is 316. There are seven curves of different colors which stand for the six comparing methods and the proposed method respectively.

The proposed method improves existing method by enhancing its visual similarity metric power. Top ranks number's increasing gives more visually similar pairs than iteration number. Then, top ranks number's increasing performs better than iteration number.

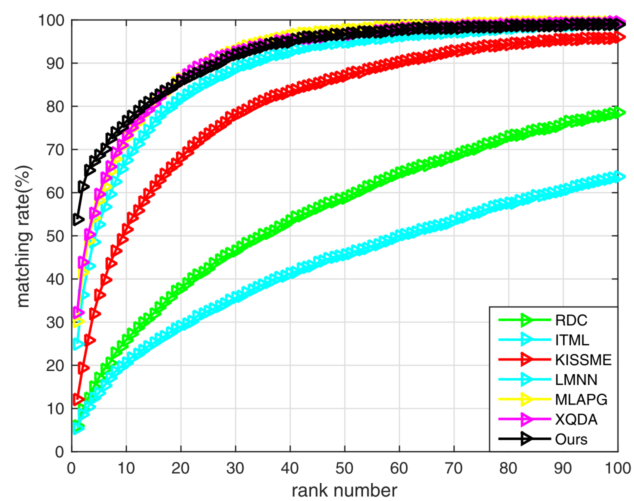
(2) **Identification performance on different datasets**: In this section, the identification performance of proposed method on the chosen dataset is displayed with the cumulated matching curves and cumulated identification rates of different ranks. Moreover, extensive experiments are conducted on the VIPeR dataset with different training dataset size P. The iteration number and top ranks number are set 5 and 10 respectively. To verify the effectiveness of proposed method notably, comparing experiments are taking on the methods [7,33,40,42] and [41] with the same appearance feature representation under the same settings. Refs. [7,33,40–42] are state-of-the-art metric learning based methods. Besides, the proposed method is also compared with the performance of deep learning based method [28] and dictionary representation based method [43]. The experiments conducting on VIPeR dataset are setting with three different training dataset size which are

Table 1
Comparison results on the VIPeR dataset(P = 316).

Methods	$P = 316$				$P = 200$				$P = 150$			
	r = 1	r = 5	r = 10	r = 20	r = 1	r = 5	r = 10	r = 20	r = 1	r = 5	r = 10	r = 20
RDC-LOMO [33]	11.71	25.32	35.44	45.57	4.75	17.72	25.95	36.39	5.70	17.09	22.78	35.13
LMNN-LOMO [42]	25.63	56.96	71.20	85.44	26.90	53.80	72.78	85.13	25.95	54.43	67.09	81.01
ITML-LOMO [40]	7.91	18.03	23.42	33.54	7.28	15.82	22.78	31.65	2.53	9.17	13.29	21.52
KISSME [14]	18.67	47.15	61.71	75.63	17.09	43.67	57.28	73.10	13.61	37.03	50.95	68.67
SLDDL [43]	16.86	41.22	58.06	**95.57**	–	–	–	–	–	–	–	–
Improved Deep [28]	34.81	63.5	75.00	80.00	–	–	–	–	–	–	–	–
KCCA [44]	30.16	62.69	76.04	86.80	–	–	–	–	–	–	–	–
XQDA-LOMO [7]	40.38	68.26	80.76	91.11	36.39	61.39	75.32	87.03	32.03	61.01	73.92	86.39
MLAPG-LOMO [41]	40.73	69.94	82.34	92.37	36.07	63.29	75.63	88.29	32.28	59.49	72.78	83.54
Ours-LOMO	**59.21**	**76.11**	**84.21**	93.20	**56.11**	**70.95**	**78.99**	**88.35**	**53.83**	**68.80**	**76.52**	**85.82**



(a) Identification results on VIPeR (P=200)



(b) Identification results on VIPeR (P=150)

Fig. 6. Identification results of comparing experiments on dataset VIPeR under two different training dataset sizes, P = 200 and P = 150. There are seven curves of different colors which stand for the six comparing methods and the proposed method respectively.

316, 200 and 150 respectively. The experimental results on VIPeR dataset are displayed in Fig. 5 and Table 1.

As shown in Fig. 5, the rank-1 identification rate of our proposed method has a outstanding high matching rate value compared with other methods. The rank-1 identification rate of proposed method is improved by 18.48% compared with the suboptimal comparing method, MLAPG [41]. It outperforms all the other comparing methods at the top 10 ranks. The detailed experimental results on VIPeR dataset are shown in Table 1. The experimental results shown in Fig. 5 and Table 1 have verified the effectiveness of our proposed method powerfully. Then the size of training dataset is reduced to verify the performance of our proposed method under P = 200 and P = 150. The experimental results are shown in Fig. 6 and Table 1. As the identification rates shown in Table 1, our proposed achieves a even better improvement performance with the decrease of training dataset size. The identification rate of rank-1 is improved by 20.04% under P = 200 and 21.55% under P = 150 compared with the suboptimal comparing method. Moreover, it reaches the best performance at all the listed ranks. It proves that our proposed method is more robust to the training dataset size and more powerful in person re-identification task.

According to the identification results shown in Figs. 5 and 6, it should be noticed that the improvement of identification performance of our proposed method mainly focus on the top ranks. It is for the reason that there are a certain proportion of hard positive pairs, the identification rate could not continue to increase to 100%. From another point of view, it verify that our proposed method has overcome the over-fitting problem. As aforementioned, there are hard positive pairs and similar negative pairs in training dataset. It leads to the over-fitting problem by the strictly constraint of metric learning. The pseudo training pairs of feedback mechanism are similar in projection subspace. A set of metric subspaces are learned with the pseudo training pairs following visual similarity. Therefore, the proposed mean distance of traditional metric subspace and feedback metric subspaces is in accordance with the visual similarity.

Table 2
Comparison results on the GRID dataset.

Methods	$P = 125$		
	r = 1	r = 10	r = 20
RDC [33]	9.68	32.96	44.32
RankSVM [28]	10.24	33.28	43.68
XQDA [7]	17.68	47.84	58.00
MLAPG [41]	16.64	41.2	52.96
Ours	**19.68**	**49.04**	**59.52**

Moreover, in order to verify the effectiveness of our proposed method, experimental experiments on other two datasets, GRID and CUHK01, are conducted. The identification performance of our proposed method on datasets GRID and CUHK01 are displayed in Tables 2 and 3 respectively. The iteration number and top ranks number for feedback are also set to 5 and 10 respectively.

As shown in Table 2, our proposed method outperforms all the state-of-the-art metric learning based methods that are chosen for comparing experiment. It reaches the best identification rate at all

Table 3
Comparison results on the CUHK01 dataset.

Methods	P = 485			
	r = 1	r = 5	r = 10	r = 20
Improved Deep [28]	47.53	70.00	80.00	–
KCCA [44]	56.30	80.66	87.94	93.00
XQDA-LOMO [7]	63.21	83.89	90.04	94.14
MLAPG-LOMO [41]	64.24	85.41	90.84	94.92
Ours-LOMO	**69.65**	**88.33**	**92.98**	**96.03**

Table 4
Comparison results on VIPeR dataset for algorithms training time (seconds).

Methods	KISSME	RDC	ITML	LMNN	XQDA	Ours
Time	13.8	360.1	547.6	354.2	4.3	97.6

the ranks displayed in table of comparing results on GRID dataset. More concretely, Our proposed method improves the state-of-the-art identification rates, rank-1, rank-10 and rank-20 by 2.00%, 1.20% and 1.52% respectively.

Table 3 displays the comparing identification results on CUHK01 dataset. Although the identification improvements at the top ranks are not as that much significant as it is on VIPeR dataset, there is a great increase of identification rates at all the displayed ranks. Our proposed method improves the state-of-the-art identification ranks at rank-1, rank-5, rank-10 and rank-20 by 5.41%, 2.92%, 2.14% and 1.09%, respectively. The improvement is more significant than it is on GRID dataset. For the reason that the identification performance on CUHK01 dataset is much better than the identification performance on VIPeR dataset, the improvements on CUHK01 dataset are lower than it is on VIPeR dataset.

(3) **Running Time**: The running time is another important standard for algorithm evaluation. The comparing training time results are displayed in Table 4. The training time test of all methods are conducted on the same computer with an Intel i3-2130 3.40GHz CPU. Several state-of-the-art metric learning based methods such as, KISSME, RDC, ITML, LMNN and XQDA, are tested to compare with our proposed method. According the test results shown in tale 4, we can see that our proposed method take 97.6 seconds for metric learning. That is because the feedback mechanism and iteration metric learning method increase the computational burden. While the training time could be reduced by decrease the size of pseudo training dataset and iteration number in practical application if needed.

5. Conclusion

In this paper, a feedback mechanism based method is proposed for person re-identification problem. Due to the complexity of appearance features changes, it is impossible to learn a metric subspace that take all the situations of into consideration with a small size sample. Generally, existing metric learning based method learn a subspace which is over-fitting to training dataset with the strict fisher optimization criterion. To solve this problem, the proposed method firstly learn the metric subspace to measure the distance of test dataset by the existing method. Then, the top ranks matching pairs are treated as pseudo positive pairs to be back propagated to join into the metric subspace learning. Hence a multiple projection metric subspace will be learned for test dataset distance measure. The new measure results of test dataset will be back propagated to learn the metric subspace in an iterative way. The proposed method learns the metric subspace with a multiple projection approach which makes it more robust and discriminative. Experiments on three datasets in Section 4 show that the proposed method in this paper improves the rank-1 re-identification rates significantly.

References

[1] A. Iosifidis, A. Tefas, I. Pitas, View-invariant action recognition based on artificial neural networks, IEEE Trans. Neural Netw. Learn. Syst. 23 (3) (2012) 412–424, doi:10.1109/TNNLS.2011.2181865.

[2] S.C. Huang, B.H. Chen, Highly accurate moving object detection in variable bit rate video-based traffic monitoring systems, IEEE Trans. Neural Netw. Inproc. Learn. Syst. 24 (12) (2013) 1920–1931, doi:10.1109/TNNLS.2013.2270314.

[3] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: European Conference on Computer Vision, 2008, pp. 262–275, doi:10.1007/978-3-540-88682-2_21.

[4] I. Kviatkovsky, A. Adam, E. Rivlin, Color invariants for person reidentification, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2013) 1622–1634, doi:10.1109/TPAMI.2012.246.

[5] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3318–3325, doi:10.1109/CVPR.2013.426.

[6] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S.Z. Li, Salient color names for person re-identification, in: European Conference on Computer Vision, 2014, pp. 536–551, doi:10.1007/978-3-319-10590-1_35.

[7] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, IEEE Conf. Comput. Vis. Pattern Recogn. 8 (4) (2015) 2197–2206, doi:10.1109/CVPR.2015.7298832.

[8] M. Hirzer, P.M. Roth, M. Köstinger, H. Bischof, Relaxed pairwise learned metric for person re-identification, in: European Conference on Computer Vision, 2012, pp. 780–793, doi:10.1007/978-3-642-33783-3_56.

[9] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987, doi:10.1109/TPAMI.2002.1017623.

[10] C. Liu, S. Gong, C.C. Loy, X. Lin, Person re-identification: what features are important? in: IEEE International Conference on Computer Vision, 2012, pp. 391–401, doi:10.1007/978-3-642-33863-2_39.

[11] J. Lu, V.E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2041–2056, doi:10.1109/TPAMI.2015.2408359.

[12] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? Metric learning approaches for face identification, in: IEEE International Conference on Computer Vision, 2009, pp. 498–505, doi:10.1109/ICCV.2009.5459197.

[13] Z. Weishi, G. Shaogang, X. Tao, Person re-identification by probabilistic relative distance comparison, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 649–656, doi:10.1109/CVPR.2011.5995598.

[14] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: European Conference on Computer Vision, 2012, pp. 2288–2295, doi:10.1109/CVPR.2012.6247939.

[15] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1239–1248, doi:10.1109/CVPR.2016.139.

[16] J. Lu, X. Zhou, Y.P. Tan, Y. Shang, J. Zhou, Neighborhood repulsed metric learning for kinship verification, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 331–345, doi:10.1109/TPAMI.2013.134.

[17] Y. Luo, T. Liu, D. Tao, C. Xu, Decomposition-based transfer distance metric learning for image classification., IEEE Trans. Image Process. 23 (9) (2014) 3789, doi:10.1109/TIP.2014.2332398.

[18] B. Prosser, W. Zheng, S. Gong, T. Xiang, Q. Mary, Person re-identification by support vector ranking, in: British Machine Vision Conference, 2010, pp. 1–11, doi:10.5244/C.24.21.

[19] R.F.D.C. Prates, W.R. Schwartz, Cbra: color-based ranking aggregation for person re-identification, in: IEEE International Conference on Image Processing, 2015, doi:10.1109/ICIP.2015.7351146.

[20] X. Liu, M. Song, Q. Zhao, D. Tao, C. Chen, J. Bu, Attribute-restricted latent topic model for person re-identification, Pattern Recognit. 45 (12) (2012) 4204–4213, doi:10.1016/j.patcog.2012.05.019.

[21] S.C. Dong, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification, in: British Machine Vision Conference, 2011, pp. 68.1–68.11, doi:10.5244/C.25.68.

[22] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience matching, in: IEEE International Conference on Computer Vision, 2013, pp. 2528–2535, doi:10.1109/ICCV.2013.314.

[23] H. Wang, S. Gong, T. Xiang, Unsupervised learning of generative topic saliency for person re-identification, Brit. Mach. Vis. Assoc. Bmva (2014), doi:10.5244/C.28.48.

[24] S. Iodice, A. Petrosino, Salient feature based graph matching for person re-identification, Pattern Recognit. 48 (4) (2015) 1074–1085, doi:10.1016/j.patcog.2014.09.011.

[25] R. Zhao, W. Oyang, X. Wang, Person re-identification by saliency learning, IEEE Trans. Pattern Anal. Mach. Intell. (2017) 356–370, doi:10.1109/TPAMI.2016.2544310.

[26] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a siamese time delay neural network, in: NIPS, 1993, pp. 737–744, doi:10.1142/S0218001493000339.

[27] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: deep filter pairing neural network for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 152–159, doi:10.1109/CVPR.2014.27.
[28] E. Ahmed, M. Jones, T.K. Marks, An improved deep learning architecture for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3908–3916, doi:10.1109/CVPR.2015.7299016.
[29] D. Yi, Z. Lei, S.Z. Li, Deep metric learning for person re-identification, in: International Conference on Pattern Recognition, 2014, pp. 34–39, doi:10.1109/ICPR.2014.16.
[30] E. Ustinova, Y. Ganin, V. Lempitsky, Multiregion bilinear convolutional neural networks for person re-identification, Comput. Sci. 48 (10) (2015) 2993–3003.
[31] S. Ding, L. Lin, G. Wang, H. Chao, Deep feature learning with relative distance comparison for person re-identification, Pattern Recognit. 48 (10) (2015) 2993–3003, doi:10.1016/j.patcog.2015.04.005.
[32] T.Y. Lin, A. Roychowdhury, S. Maji, Bilinear cnn models for fine-grained visual recognition, in: IEEE International Conference on Computer Vision, 2015, pp. 1449–1457, doi:10.1109/ICCV.2015.170.
[33] W. Zheng, S. Gong, T. Xiang, Reidentification by relative distance comparison, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 653–668, doi:10.1109/TPAMI.2012.138.
[34] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, J. Bu, Semi-supervised coupled dictionary learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3550–3557, doi:10.1109/CVPR.2014.454.
[35] Y. Zhang, B. Li, H. Lu, A. Irie, X. Ruan, Sample-specific svm learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1278–1287, doi:10.1109/CVPR.2016.143.
[36] G. Lisanti, I. Masi, A.D. Bagdanov, A.D. Bimbo, Person re-identification by iterative re-weighted sparse ranking, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1629–1642, doi:10.1109/TPAMI.2014.2369055.
[37] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2007.
[38] C. Liu, S. Gong, C.L. Chen, On-the-fly feature importance mining for person re-identification, Pattern Recognit. 47 (4) (2014) 1602–1615, doi:10.1016/j.patcog.2013.11.001.
[39] W. Li, R. Zhao, X. Wang, Human reidentification with transferred metric learning, in: Asian Conference on Computer Vision, 2012, pp. 31–44, doi:10.1007/978-3-642-37331-2_3.
[40] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, in: Machine Learning, Proceedings of the Twenty-Fourth International Conference, 2007, pp. 209–216, doi:10.1145/1273496.1273523.
[41] S. Liao, S.Z. Li, Efficient psd constrained asymmetric metric learning for person re-identification, in: IEEE International Conference on Computer Vision, 2015, pp. 3685–3693, doi:10.1109/ICCV.2015.420.
[42] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (1) (2009) 207–244, doi:10.1109/LSP.2014.2301037.
[43] X.Y. Jing, X. Zhu, F. Wu, X. You, Q. Liu, D. Yue, R. Hu, B. Xu, Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 695–704, doi:10.1109/CVPR.2015.7298669.
[44] G. Lisanti, I. Masi, A. Del Bimbo, Matching people across camera views using kernel canonical correlation analysis, in: Proceedings of the International Conference on Distributed Smart Cameras, 2014, pp. 1–6, doi:10.1145/2659021.2659036.

Yutao Ren is currently working toward the Ph.D. degree in college of science, Wuhan University of Technology, Wuhan, China. His current research interests include pattern recognition computer vision and machine learning.

Xuelong Li is a Full Professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xian Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xian, China.

Xiaoqiang Lu is currently a Full Professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xian, China. His current research interests include pattern recognition, machine learning, hyperspectral image analysis, cellular automata, and medical imaging.