IET Research Journals

*Submission Template for IET Research Journal Papers*

IET Journals
The Institution of Engineering and Technology

# Metric Learning by Simultaneously Learning Linear Transformation Matrix and Weight Matrix for Person Re-identification

ISSN 1751-8644
doi: 0000000000
www.ietdl.org

*Jian'an Zhang*[1], *Qi Wang*[1*], *Yuan Yuan*[1]

[1] *School of computer science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China*
[*] *E-mail: crabwq@gmail.com*

**Abstract:** Mahalanobis metric learning is one of the most popular methods for person re-identification. Most existing metric learning methods regularly formulate the person re-identification as an unconstrained optimization problem and the constraints on the Mahalanobis matrix are seldom imposed. In addition, weights are often used to model the relationships between different variables but they often suffer from boundedness caused by their hand-designed feature. Taking the above two disadvantages into consideration, we propose a new metric learning method for person re-identification which formulates the metric learning problem as a constrained optimization problem by imposing a constraint on the linear transformation matrix. Furthermore, we treat the weights as unknown variables and introduce a weight learning method instead of designing weight intuitively. Finally, we evaluate the proposed method on two challenging person re-identification databases and show that it performs favorably against the state-of-the-art approaches.

## 1 Introduction

With the development of monitor devices, we can get more and more surveillance videos easily which leads to a basic problem that is how to find the same person's identity in all videos when a query image is given. Such a problem is usually called person re-identification. Formally, person re-identification refers to a task of recognizing the same person's identity from a network of cameras with non-overlapping fields of view. As we can see, person re-identification is an important problem in real applications especially in security and surveillance. Person re-identification is also a challenging problem because frequently persons with different identity may look more similar than the same ones. In addition, the big intra-class invariance caused by pose illumination, occlusion and viewpoint adds difficulty to the problem.

Many researchers have devoted to such a task in the past ten years and a plenty of methods have been proposed for person re-identification [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18]. Most existing approaches for tackling the person re-identification problem are mainly carried on from two groups: developing distinctive feature representations and seeking discriminative distance metrics. Both of them aim to compute the matching distances (or scores) which are optimal for matched image pairs from the gallery and probe set respectively.

The first group of methods are related with feature representation and a number of approaches have been proposed to design robust descriptors against background and illumination variations. Such as, ELF [3] , SDALF [4], LOMO [5], combination of Biologically Inspired Features (BIF) and Covariance descriptors [19] have been classic feature-based methods. The second group of methods focus on metric learning and a plenty of research works aim to learn a robust and discriminative metric matrix [20]. For instance, LMNN [6] , ITML [7], PCCA [8], KISSME [9], PRDC[10], LFDA[11], LADF[12] and XQDA [5] are all classic metric learning methods for person re-identification. Our method belongs to this type. Despite it is intuitive and important enough to develop feature-based algorithm, hand-craft feature is far from being used in real scenario because of complex background and deformation of identities. Metric learning as a similarity measure method between identities play a feature-transformed role as well as discriminative metric role, which is more suit for real scenario. Based on these opinion, we pay attention to develop metric learning methods.

Our paper focuses on the metric learning method for person re-identification. Most existing metric methods model a linear transformation matrix $\mathbf{A}$ instead of the Mahalanobis matrix $\mathbf{M}$ directly for simplicity and accessibility, where the relationship between the above two matrixes can be easily got by $\mathbf{M} = \mathbf{A}^T\mathbf{A}$. And these methods for person re-identification regularly formulate the problem as an unconstrained optimization problem and the constraints on the linear transformation matrix are seldom imposed. Although it is simple to solve the unconstrained optimization problem, the information are not fully employed and at the same time the linear transformation can not be powerful enough without certain constraints. Besides, weights are often used to model the relationships between different variables but they often suffer from boundedness caused by their hand-designed feature. Taking the above two disadvantages into consideration, we propose a new metric learning method for person re-identification which formulates the metric learning problem as a constrained optimization problem by imposing a constraint on the linear transformation matrix. Furthermore, we treat the weights as unknown variables and introduce a weight learning method instead of designing weight intuitively. The diagram of the proposed method can be viewed in Fig. 1.

The main contributions of this work are two-fold. First, we propose a new cost function for metric learning used in person re-identification. The proposed method imposes a constraint on the linear transformation and hence it is formulated as a constrained optimization problem. Second, we introduce a weight learning strategy to learn the weights instead of designing weight intuitively. Moreover, we adopt an efficient method based on matrix optimization to solve the proposed cost function and an algorithm is formed for person re-identification.

## 2 Related Works

Most existing approaches for tackling the person re-identification problem are mainly carried on from two groups: developing distinctive feature representations and seeking discriminative distance metrics. Both of them aim to compute the matching distances (or

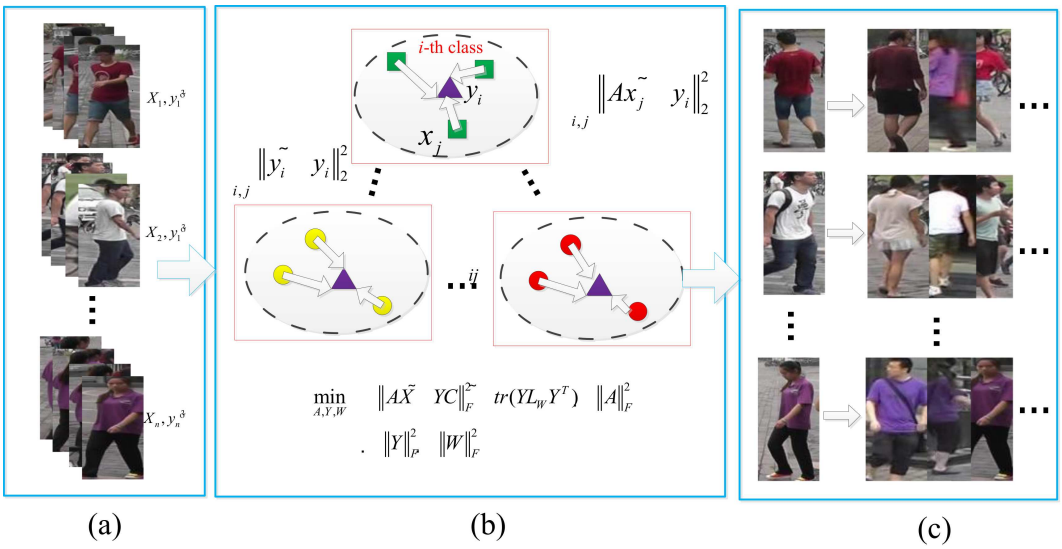


**Fig. 1**: The diagram of the proposed method for person re-identification. (a) shows the different class samples and their labels that to be trained. (b) shows the intuitions that formulate the loss function. (c) shows the test and ranking process for person re-identification

scores) which are optimal for matched image pairs from the gallery and probe set respectively. Due to the superiority of deep learning methods in computer vision [21] [22], a series of Deep Neural Network (DNN) based methods have been proposed for person re-identification, which adopt a different pipeline from the metric learning methods, such as [23] developed a multi-level Gaussian model and [24] proposed a multi-level similarity method based on DNNs. We refer readers to [13], [14], [15], [16], [17], [18], [25] and [26] for more information.

The first group of methods are related with feature representation and a number of approaches [3–5, 19, 27, 28] have been proposed to design robust descriptors against background and illumination variations.For example, Gray et al. [3] presented the ELF by fusing 8 color channels with 19 texture channels while Farenzena et al. [4] proposed the Symmetry-Driven Accumulation of Local Features(SDALF) method that the weighted color histograms, Maximally Stable Color Regions (MSCR), Recurrent High-Structured Patches (RHSP) are employed to capture different image properties.In addition, Ma et al. [19] proposed an image representation based on the combination of Biologically Inspired Features (BIF) and Covariance descriptors. Zhao et al. [27] proposed a method which focuses on learning saliency. Liao et al. developed the LOMO [5] and it has been proved that LOMO based pedestrian representation has shown impressive robustness against viewpoint changes which extracted the maximal pattern of joint HSV color histogram coupled with Scale Invariant Local Ternary Pattern (SILTP). GOG [29] is a new efficient descriptor and it models every local patches in each strip of an image using a Gaussian distribution. Next,each strip is regarded as a set of such Gaussian distributions, which is then summarized using a single Gaussian distribution.

Here we mainly review the related metric learning methods. Most existing metric learning methods for person re-identification firstly generate the pairwise constraints where the sample pairs with the same labels consist of the positive set and the ones with the different labels comprise the negative set. We usually denote $S = \{(x_i, x_j)\|y_{ij} = 1\}$ positive pair set and $D = \{(x_i, x_j)\|y_{ij} = 0\}$ negative pair set where $y_{ij} = 1$ stands for $x_i$ and $x_j$ are of the same class and $y_{ij} = 0$ otherwise. After the two sets are generated, different kinds of methods are formulated to pull the pairs in the positive set together and penalize the pairs in the negative set. For instance, the LMNN [6] tries to pull the positive pairs lying within the k-nearest neighbors and push the negative pairs by a large margin. The ITML [7] method aims to minimize the relative entropy between a given matrix and the variable Mahalanobis matrix with the constraints that the distances between positive pairs is less than a given constant and the distances between negative pairs is greater than another given constant. Mignon et al. [8] developed a metric learning method named PCCA which adapts to sparse pairwise similarity/dissimilarity constraints in high dimensional input space. KISSME [9] method derived a Mahalanobis metric by computing the difference between the intra-class and inter-class covariance matrix and a modified method can be found in [30]. Zheng et al. [10] proposed the PRDC based method where the probability of a pair of true match having a smaller distance than that of a wrong match pair is maximized. Pedagadi et al. [11] employed the LFDA algorithm to maximize the inter-class separability while preserving the multi-class modality. Li et al. [12] developed the Locally-Adaptive Decision Functions (LADF), which combines the distance metric with a locally adaptive thresholding rule for each pair of sample images. As an improvement, XQDA [5] learned a more discriminative distance metric and low-dimensional subspace simultaneously. Recently, Yang et al. [31] proposed a generalized similarity metric learning method, with enhanced LOMO feature, it achieves promising results on benchmark datasets. Yang et al. [32] proposed a new logistic discriminant metric learning method that take advantage of a privileged information.

Despite a lot of metric learning methods have been developed for person re-identification, there are no works that focus on weights learning where is important to balance the sample pairs in metric learning. For example, LMNN [6], ITML [7], KISSME [9], [31] and LADF [10] have no weights constraints. And despite a weight is imposed to the sample pairs in LFDA [11], the weights are hand-designed instead of learning from samples. The proposed method differs from existing metric learning methods that it takes the weight learning into consideration and formulates the re-identification problem as a jointly metric and weights learning problem.

## 3 Methodology

### 3.1 Problem Formulation

In this paper, we also follow the most of the existing works for metric learning to model the linear transformation matrix $\mathbf{A} \in \mathbb{R}^{k\times d}$. A good intuition has been proposed in [33] used for dimension reduction which shows that after linear transformation one sample should be as close to it's unknown class center and at the same time different unknown class centers should be far from each other.

$$\begin{aligned}&\sum_i \sum_j \|\mathbf{A}\mathbf{x}_j - \mathbf{y}_i\|_2^2 - \sum_i \sum_j \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \\ &= \|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C}\|_F^2 - tr(\mathbf{Y}\mathbf{H}\mathbf{Y}^T),\end{aligned} \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{d\times n}$ is the data matrix and $\mathbf{Y} \in \mathbb{R}^{k\times c}$ consists of unknown class centers. $\mathbf{C} \in \mathbb{R}^{c\times n}$ is a constant matrix which can be found in [33] and $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{c\times c}$ is the centering matrix. Note that, $d$ is the dimensionality of feature vectors and $k$ is the dimensionality after linear transformation. $n$ is the number of images in the training stage and each image is represented as a column feature vector in $\mathbf{X}$, which $c$ is the number of classes (or number of identities).

It can be found that such an idea is adequate for metric learning as well because it describes a general relationship for an effective linear transformation matrix. However, it is not suitable to directly introduce the above formulation for metric learning because it has the following two drawbacks. First, Eq. (1) does not demonstrate the difference between different pair of unknown class centers because all pair of unknown class centers are equivalently formulated by $\|\mathbf{y}_i - \mathbf{y}_j\|_2^2$. Second, there are no constraints imposed on the linear transformation matrix $\mathbf{A}$ which can lead to a case that after linear transformation the relative position of different classes are not kept.

Based on the above two considerations, we can make a further modification to Eq. (1) and adapt it to metric learning for person re-identification. For the first weakness, a simple but suitable strategy is to impose different weights to different pairs of unknown class centers so that the difference between different pairs of class centers can be enforced. Formally, it can be formulated in the following form

$$\sum_i \sum_j \omega_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 = tr(\mathbf{Y}\mathbf{L}_\mathbf{W}\mathbf{Y}^T),$$

where $\mathbf{L}_\mathbf{W} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix of $\mathbf{W}$ and $\mathbf{D}$ is the diagonal matrix whose element $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ is the sum of $i$-th row of matrix $\mathbf{W}$. In addition, it should also satisfy $\mathbf{W} = \mathbf{W}^T$ and $\mathbf{W} \succ 0$ because the element of symmetric position in $\mathbf{W}$ constrain the same pair of class centers and all the elements should impose positive effect on the class centers. We can design the weights intuitively on the basis that the larger the distance between two classes are the larger the weight should be. Nevertheless, weights designed by intuition are of uncertainty and hence we turn to get the weight matrix by learning method. Namely, we treat the weight matrix as an unknown matrix variable and optimize an loss function to learn such a matrix, through which we can get a more reasonable weight matrix.

For the second disadvantage, in order to keep the relative position, a general strategy is to let the linear transformation matrix $\mathbf{A}$ be orthogonal because orthogonal matrix has the property that after linear transformation the distance between two samples is kept, namely $\mathbf{A}^T\mathbf{A} = \mathbf{I}$. However, such a method will result in a trivial solution for Mahalanobis matrix and hence such a strategy is not feasible. Instead, we can turn to an appropriate strategy $\|\mathbf{A}^T\mathbf{A} - \mathbf{I}\|_F^2 < \delta$ which a trivial solution can be avoided and the property we need can be kept to some extent.

From what discussed above, we can formulate the metric learning problem as follows

$$\begin{aligned} \min_{\mathbf{A},\mathbf{Y},\mathbf{W}} \quad & \|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C}\|_F^2 - tr(\mathbf{Y}\mathbf{L}_\mathbf{W}\mathbf{Y}^T) \\ & + \|\mathbf{A}\|_F^2 + \|\mathbf{Y}\|_F^2 + \|\mathbf{W}\|_F^2 \\ s.t. \quad & \|\mathbf{A}^T\mathbf{A} - \mathbf{I}\|_F^2 < \delta \\ & \mathbf{W} = \mathbf{W}^T,\ \mathbf{W} \succ 0, \end{aligned} \tag{2}$$

where $\delta$ is a minor constant that controls the bounding of the above equation. It is noted that three regular terms $\|\mathbf{A}\|_F^2, \|\mathbf{Y}\|_F^2$ and $\|\mathbf{W}\|_F^2$ are added to the loss function in order to make the solution stable.The proposed problem is an constrained matrix optimization problem. Although the loss function is not convex with regard to all the matrix variables, it is convex with respect to one of the matrix variables when the others are fixed. Based on this property, we will introduce an algorithm for solving the proposed optimization problem in the next section.

### 3.2 Optimization

Next, we will develop an efficient solver for (2) based on the Augmented Lagrange Multiplier (ALM) with Alternating Direction Minimizing (ADM) strategy[34]. In order to make the objective function separable, we introduce an auxiliary variable $\mathbf{Q}$ to replace $\mathbf{Y}$ in the trace term of the objective function which is similar to [35]. Hence, the Augmented Lagrangian function of (2) with the introduced constraint is

$$\begin{aligned} L_{(\mathbf{A},\mathbf{Y},\mathbf{W},\mathbf{Q})} =& \|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C}\|_F^2 - tr(\mathbf{Q}\mathbf{L}_\mathbf{W}\mathbf{Q}^T) + \|\mathbf{A}\|_F^2 \\ & + \|\mathbf{Y}\|_F^2 + \|\mathbf{W}\|_F^2 + \Phi(\mathbf{Z_1}, \mathbf{Q} - \mathbf{Y}) \\ & + \Phi(\mathbf{Z_2}, \mathbf{W} - \mathbf{W}^T) \\ s.t. \quad & \|\mathbf{A}^T\mathbf{A} - \mathbf{I}\|_F^2 < \delta \\ & \mathbf{W} \succ 0, \end{aligned} \tag{3}$$

where $\Phi(\mathbf{Z}, \mathbf{C}) = <\mathbf{Z}, \mathbf{C}> + \frac{\mu}{2}\|\mathbf{C}\|_F^2$ and $<\mathbf{Z}, \mathbf{C}> = tr(\mathbf{Z}^T\mathbf{C})$ defines the inner product of matrix. Although the objective (3) is not jointly convex with all the variables, it is convex with respect to each of the variables when the others are fixed.

**(1) Solving for $\mathbf{A}$** To learn $\mathbf{A}$ for a given $\mathbf{Y}$, the objective function reduces to

$$\begin{aligned} \min L_\mathbf{A}(\mathbf{A}) =& \|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C}\|_F^2 + \|\mathbf{A}\|_F^2 \\ s.t. \quad & \|\mathbf{A}^T\mathbf{A} - \mathbf{I}\|_F^2 < \delta, \end{aligned} \tag{4}$$

and the above problem (4) can be equivalently transformed into the following problem

$$\min_\mathbf{A} \quad \|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C}\|_F^2 + \|\mathbf{A}\|_F^2 + \|\mathbf{A}^T\mathbf{A} - \mathbf{I}\|_F^2. \tag{5}$$

By taking derivation with respect to $\mathbf{A}$ in (5), we have

$$\begin{aligned} \frac{\partial L_\mathbf{A}}{\partial \mathbf{A}} &= 2(\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C})\mathbf{X}^T + 2\mathbf{A} + 2\mathbf{A}(\mathbf{A}^T\mathbf{A} - \mathbf{I}) \\ &= \mathbf{A}\mathbf{X}\mathbf{X}^T + \mathbf{A}\mathbf{A}^T\mathbf{A} - \mathbf{Y}\mathbf{C}\mathbf{X}^T. \end{aligned} \tag{6}$$

**(2) Solving for $\mathbf{Y}$** For a given $\mathbf{A}$, we can solve the following optimization problem to estimate $\mathbf{Y}$.

$$\min L_\mathbf{Y}(\mathbf{Y}) = \|\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C}\|_F^2 + \|\mathbf{Y}\|_F^2. \tag{7}$$

By taking derivation of function (7) with regard to $\mathbf{Y}$, we get

$$\frac{\partial L_\mathbf{Y}}{\partial \mathbf{Y}} = -2(\mathbf{A}\mathbf{X} - \mathbf{Y}\mathbf{C})\mathbf{C}^T + 2\mathbf{Y}. \tag{8}$$

Let the equation of (8) be $\mathbf{0}$ leading to

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T + \mathbf{I})^{-1}, \tag{9}$$

where (9) is the closed-form solution of $Y$.

**(3) Solving for $\mathbf{Q}$** Fixing the other variables to estimate $\mathbf{Q}$, the problem is then be reduced to the following optimization problem

$$\begin{aligned} \min L_\mathbf{Q}(\mathbf{Q}) =& tr(\mathbf{Q}\mathbf{L}_\mathbf{W}\mathbf{Q}^T) + \Phi(\mathbf{Z_1}, \mathbf{Q} - \mathbf{Y}) \\ =& tr(\mathbf{Q}\mathbf{L}_\mathbf{W}\mathbf{Q}^T) + tr(\mathbf{Z_1}^T(\mathbf{Q} - \mathbf{Y})) \\ & + \frac{\mu}{2}\|\mathbf{Q} - \mathbf{Y}\|_F^2. \end{aligned} \tag{10}$$

By taking derivation of (10) with respect to $\mathbf{Q}$, we can get

$$\frac{\partial L_\mathbf{Q}}{\partial \mathbf{Q}} = 2\mathbf{Q}\mathbf{L}_\mathbf{W} + \mathbf{Z_1} + \mu(\mathbf{Q} - \mathbf{Y}) = \mathbf{0},$$

and

$$\mathbf{Q} = (\mu\mathbf{Y} - \mathbf{Z_1})(2\mathbf{L_W} + \mu\mathbf{I})^{-1}. \tag{11}$$

**(4) Solving for $\mathbf{W}$** For a given variable $\mathbf{Q}$, the objective function reduces to

$$\begin{aligned} \min\, & L_{\mathbf{W}}(\mathbf{W}) = tr(\mathbf{Q}\mathbf{L_W}\mathbf{Q}^T) + \|\mathbf{W}\|_F^2 + \Phi(\mathbf{Z_2}, \mathbf{W} - \mathbf{W}^T) \\ & s.t. \qquad \mathbf{W} \succ 0, \end{aligned} \tag{12}$$

and the corresponding Lagrange function can be deduced as

$$\begin{aligned} L_{\mathbf{W}}(\mathbf{W}) =& tr(\mathbf{Q}\mathbf{L_W}\mathbf{Q}^T) + \|\mathbf{W}\|_F^2 + \Phi(\mathbf{Z_2}, \mathbf{W} - \mathbf{W}^T) \\ & - tr(\mathbf{Z_3}^T\mathbf{W}). \end{aligned} \tag{13}$$

By taking derivation of (13) with respect to $\mathbf{W}$, we can get

$$\begin{aligned} \frac{\partial L_{\mathbf{W}}}{\partial \mathbf{W}} =& \mathbf{F_Q} - \mathbf{Q}^T\mathbf{Q} + 2\mathbf{W} + \mathbf{Z_2} - \mathbf{Z_2}^T \\ & + \mu(\mathbf{W} - \mathbf{W}^T) - \mathbf{Z_3} = \mathbf{0}, \end{aligned}$$

where $\mathbf{F_Q} = (\mathbf{Q} \odot \mathbf{Q})^T\mathbf{1}_d\mathbf{1}_n^T$, $\odot$ represents the Hadamard product and $\mathbf{1}_d$ represents $d$-dimensional vector with all elements be 1 and the same as $\mathbf{1}_n$. Besides, we can easily get

$$\begin{aligned} (2+\mu)\mathbf{W} - \mu\mathbf{W}^T =& \mathbf{Z_2}^T - \mathbf{Z_2} - \mathbf{F_Q} \\ & + \mathbf{Q}^T\mathbf{Q} + \mathbf{Z_3}. \end{aligned} \tag{14}$$

Taking the transpose operator to the whole equation (14) and solving the two equations, we have

$$\begin{aligned} \mathbf{W} =& \frac{1}{2+2\mu}(\mathbf{Z_2}^T - \mathbf{Z_2}) + \frac{1}{2}\mathbf{Q}^T\mathbf{Q} \\ & + \frac{2+\mu}{4+4\mu}(\mathbf{Z_3} - \mathbf{F_Q}) \\ & + \frac{\mu}{4+4\mu}(\mathbf{Z_3}^T - \mathbf{F_Q}^T), \end{aligned} \tag{15}$$

where Eq. (14)is the closed-form solution of $\mathbf{W}$.

**(5) Updating multipliers $\mathbf{Z_1}, \mathbf{Z_2}$ and $\mathbf{Z_3}$** It can be found that there are still four multipliers to update, which are simply done as follows

$$\begin{aligned} \mathbf{Z}_1^{(k+1)} &= \mathbf{Z}_1^{(k)} + \mu^{(k)}(\mathbf{Q} - \mathbf{Y}) \\ \mathbf{Z}_2^{(k+1)} &= \mathbf{Z}_2^{(k)} + \mu^{(k)}(\mathbf{W} - \mathbf{W}^T) \\ \mathbf{Z}_2^{(k+1)} &= \mathbf{Z}_2^{(k)} + \mu^{(k)}\mathbf{W}. \end{aligned} \tag{16}$$

We note that careful initiation of $\mathbf{Y}$ may be necessary because otherwise we may reach a unreasonable linear transformation matrix $\mathbf{A}$ from (6) and this could lead to a bad model. It is a good choice that we take the mean vectors of different class samples as initial values. This is $\hat{\mathbf{y_i}} = \Sigma_{k \in C_i}\mathbf{x}_k, i = 1, 2, \cdots, c$, and

$$\mathbf{Y_0} = [\hat{\mathbf{y_1}}, \hat{\mathbf{y_2}}, \cdots, \hat{\mathbf{y_c}}], \tag{17}$$

which consists of these mean vectors in column way. In addition, we use the following equation to model the weights as the initial value

$$\omega_{ij} = \max_{i \in c_i, j \in c_j} \|\mathbf{x}_i - \mathbf{x}_j\|_2^2, i, j = 1, 2, \cdots, n, \tag{18}$$

where $c_i$ and $c_j$ represent the index set of the $i$-th and $j$-th class respectively. Intuitively, $\omega_{ij}$ is the largest distance between the $i$-th and $j$-th class which can model the weight between two unknown class centers.

Furthermore, it is noted that the unknown sample center matrix $\mathbf{Y}$ plays a core role in the above loss function for all the variables is related with it. So a good model for $\mathbf{Y}$ can yield a good performance.

**Algorithm 1** Proposed Metric Learning Algorithm for Person Re-identification

**Input:** Input data matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$ and construct matrix $\mathbf{Y}_0$, cluster number $c$

Initialize parameters $\gamma$, $\eta$, class numbers $c$. Set $k = 0$ and $\mathbf{Y_k} = \mathbf{Y_0}$ via Eq. (17), $\mathbf{Z}_1 = \mathbf{0} \in \mathbb{R}^{d \times c}$, $\mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{0} \in \mathbb{R}^{c \times c}$
**while** not convergent **do**
  Set $i = 0$
  **while** $\|\frac{\partial f(\mathbf{A}_k^i, \mathbf{Y}_k)}{\partial \mathbf{A}_k^i}\|_F^2 \leq \varepsilon$ **do**
    Execute line search process using Armijo criteria and get the stepsize $\lambda_i$
    set $\mathbf{A}_k^{i+1} \leftarrow \mathbf{A}_k^i - \lambda_i \frac{\partial f(\mathbf{A}_k^i, \mathbf{Y}_k)}{\partial \mathbf{A}_k^i}$
    $i \leftarrow i + 1$
  **end while**
  Update $\mathbf{A}^{(k+1)} = \mathbf{A}_k^i$
  Update $\mathbf{Y}^{(k+1)}$ via Eq. (9)
  Update $\mathbf{Q}^{(k+1)}$ via Eq. (11)
  Update $\mathbf{W}^{(k+1)}$ via Eq. (15)
  Update multipliers via Eq. (16)
  Balance $\mathbf{W}^{(k+1)}$ by $\frac{\mathbf{W}^{(k+1)} + \mathbf{W}^{(k+1)T}}{2}$
  $k \leftarrow k + 1$
**end while**
Set $\mathbf{M} \leftarrow \mathbf{A_k^T}\mathbf{A_k}$

**Output:** The metric matrix $\mathbf{M}$

Based on this observation, we argue that our method is more sufficient in the case that one class has more samples because a better $\mathbf{Y}$ can be modeled. Experiments will be show such a property that can be see from Section 4.

**Convergence Analysis** It is noted that the convergence of the proposed algorithm depends on the convergence of the ALM framework. The convergency of the ALM framework has been discussed detailed in . Specially, presents a detailed proof of inexact ALM to convex objective functions. However, the rigorous mathematical analyses for the proposed algorithm convergence is very difficult and it remains unknown in literature so far. Our empirical experiments also confirm the convergence of our algorithm for the benchmark datasets used in this article. One thing to note is that because Equation (2) is nonconvex, the solution yielded via alternative optimization from Equation (2) is not a globally optimal one.

**Complexity Analysis** We evaluate the computation costs of our algorithm as follows. (1) The computation cost for $\mathbf{A}$ involves derivation calculation of $\mathbf{AXX}^T + \mathbf{AA}^T\mathbf{A} - \mathbf{YCX}^T$, which is $O(knd + k^2d + d^2n)$. Taking the iteration number $K$ into consideration, it will cost $O(K(knd + k^2d + d^2n))$. (2) The computation cost for $\mathbf{Y}$ involves the matrix inverse and matrix multiplication, which is $O(c^3)$ and $O(kdn + knc + kc^2)$. (3) The computation cost for $\mathbf{Q}$ involves the matrix inverse and matrix multiplication, which is $O(c^3)$ and $O(kc^2)$. (4) The computation of $\mathbf{W}$ involves the matrix substraction and matrix multiplication, which mainly focuses on the matrix multiplication cost $O(k^2c)$.

## 4 Experiments

We validate the proposed approach in three widely used datasets in person re-identification, namely, the CUHK03 [16] dataset , the Market1501 [36] dataset and the CUHK01 [37] dataset. Details can be shown in the following four sections.

### *4.1 Feature Extraction and Settings*

**Feature Extraction** In this paper, we extract different features for different datasets in consideration of the difference between datasets and the consistency of different methods. For CUHK03 dataset, we

use the method provided by [38]. Namely, the 16 bin color histograms on RGB,YUV, and HSV channels and texture histograms based on Local Binary Patterns (LBP) are extracted and concatenated into a 2580-dimension vector. In addition, for the Market-1501 dataset we use features described in [36]. Namely, the Color Names (CN) descriptor are extracted for all pedestrian images in a dense manner and subsequently a normalized operation is executed. While for CUHK01 dataset, the LOMO [5] feature is used which extracts the maximal pattern of joint HSV color histogram coupled with Scale Invariant Local Ternary Pattern (SILTP).

**Baseline Methods** In this paper, we evaluate the proposed algorithm compared with several promising algorithms, including Euclidean distance, Mahalanobis distance, LFDA [11], oLFDA [38], ITML [7], KISSME [9], svmml[12] and MFA [38]. And we adopt a single shot experiment setting in CUHK03 dataset which is similar to the most of the previous works and for Market1501 dataset the multi-shot setting is used as usual. To evaluate the performance of our method, we adopt Cumulative Matching Characteristic (CMC) curve which provides a ranking for every image in the gallery with respect to the probe. For the above baseline methods, we follows the setting of [38]. As for the proposed method, there is one parameter $\mu$ to be determined, which is used to control the positive penalty. Without loss of generality, we follow the setting of [35] and set $\mu = 1.25$.

### 4.2 CUHK03 Dataset

The CUHK03 dataset [16] consists of 13,164 images of 1,360 pedestrians captured with six surveillance cameras. Each individual is observed by two disjoint camera views, and there are 4.8 images on average for each identity in each view. Apart from the manually labeled pedestrian bounding boxes, this database also provides the samples detected with a pedestrian detector, which causes some misalignments and body part missing for a more realistic setting.

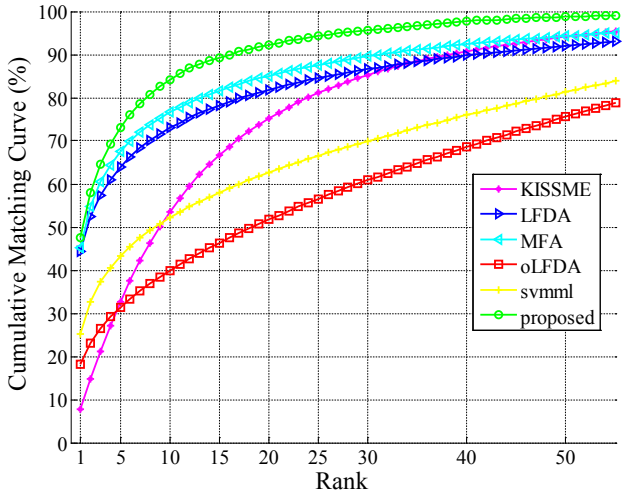


**Fig. 2**: CMC curves comparing our method against state-of-the-art methods on CUHK03 dataset (detected bounding boxes).

We run our algorithm with both bounding boxes and with the same setting as [38]. That is, the dataset is partitioned into a training set of 1,160 persons and a test set of 100 persons. The experiments are conducted with 20 random splits and the average results are presented. In addition, the metric learning methods that we used for comparison include LFDA [11], KISSME [9], svmml [12], oLFDA [38] and MFA [38]. Fig. 2 shows the result of CMC curve on the detected bounding boxes and Fig. 3 demonstrates the result on the labeled bounding boxes.

From the two figures, it is easy to find that the proposed method outperforms all the compared methods which shows the effectiveness of the method. It outperforms the second best MFA method by

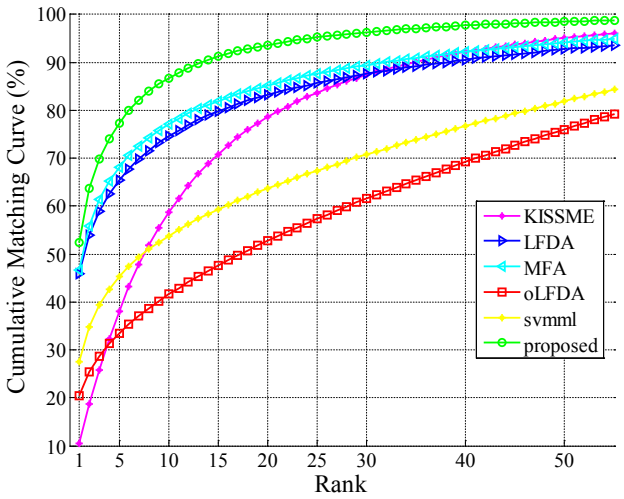


**Fig. 3**: CMC curves comparing our method against state-of-the-art methods on CUHK03 dataset (labeled bounding boxes).

**Table 1** Comparison of state-of-the-art rank-1 identification rates(%) on the CUHK03 database with both labeled and detected setting (P=100).

| Method | labled | detected |
|---|---|---|
| LMNN[6] | 7.29 | 6.25 |
| LDML[39] | 13.51 | 10.92 |
| DeepReID[16] | 20.65 | 19.89 |
| LOMO+XQDA[5] | 52.20 | 46.25 |
| Improved Deep[14] | **54.74** | 44.96 |
| proposed | 52.50 | **47.50** |

5.86% identification rate in rank 1 with the detected bounding boxes and 4.33% with the labeled bounding boxes.

In order to compare with more methods, we gather a plenty of state-of-the-art methods which is listed in Table 1 and the compared results are from [5] and [14]. From Table 1, we can find that our method achieves 52.50% rank-1 identification rate with the labeled bounding boxes and 47.50% rank-1 identification rate in the automatically detected bounding boxes, respectively. It can be found that our method achieves the best performance with the detected bounding boxes while performs a little poorer than the best result with the detected bounding boxes compared with the methods list in the table, which shows that our proposed methods performs favorably against the state-of-the-art approaches. We can note that the proposed method performs poorly than the Improved Deep method [14]. This is because that the Improved Deep method is based on a deep learning framework which is more suitable to the labeled bounding boxes. And it also can be noted that the proposed method is only 2.24% poorer than it with the labeled bounding boxes while we achieve 2.54% better performance than it with the detected bounding boxes.

On the other hand, we achieve a good performance in CUHK03 dataset and we argue that the following two reasons contribute to this performance. The first one is that CUHK03 dataset is large enough to provide rich information. As we known, there are 4.8 images on average for each identity in CUHK03 dataset and hence it provides rich information to model the unknown class centers. The second one is that we introduce a constraint on the linear transformation matrix that after linear transformation the points can be transformed into a certain bound and hence the inner-class information is kept. For the above two reasons, the effectiveness of our methods can be explained.

### 4.3 Market1501 Dataset

Market1501 dataset [36] contains 32,668 detected person bounding boxes of 1,501 identities. Each identity is captured by six cameras at most, and two cameras at least. We run our algorithm with the

same setting of [36]. That is , during testing, for each identity, one query image in each camera is selected, therefore multiple queries are used for each identity. Each identity may have multiple images under each camera. We use the provided fixed training and test set, under the multi-query settings.

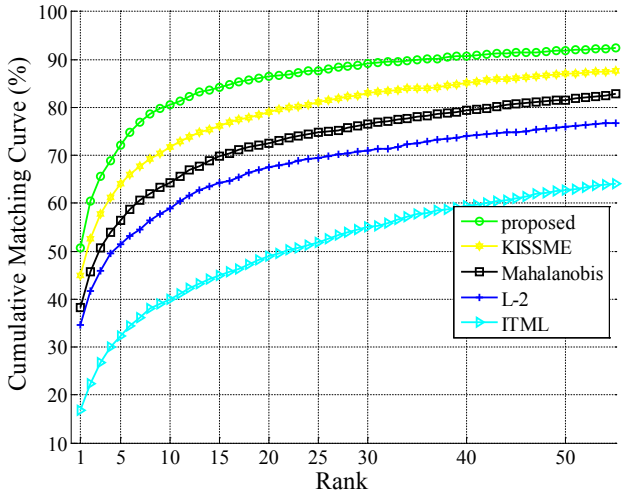


**Fig. 4**: CMC curves comparing our method against state-of-the-art methods on the Market1501 dataset.

**Table 2** Comparison of state-of-the-art results reported on the Market1501 database. The cumulative matching scores (%) at rank 1,10 and 20 are listed.

| Method | r=1 | r=10 | r=20 | mAP |
|---|---|---|---|---|
| gBiCOv+L-2 [36] | 8.28 | - | - | 2.23 |
| HistLBP+L-2 [36] | 9.62 | - | - | 2.72 |
| LOMO+L-2[36] | 26.7 | - | - | 7.75 |
| BoW+L-2 [36] | 35.84 | 60.33 | 67.64 | 14.75 |
| Bow+KISSME [36] | 44.42 | 72.18 | 78.95 | 20.76 |
| WARCA [17] | 45.16 | 76 | 84 | - |
| Srikrishna et al. [18] | 46.5 | 79.9 | 85.9 | - |
| Wu et al.[40] | 48.15 | - | - | **29.94** |
| SCSP [25] | 51.90 | - | - | 26.35 |
| Su et al.[41] | 49.0 | - | - | 25.8 |
| Liu et al.[42] | 55.4 | 85.6 | - | - |
| Multiregion[15] | **56.59** | - | - | 25.8 |
| poroposed | 50.84 | 80.46 | **86.73** | 25.24 |

Fig 4 shows the result of CMC curve compared with Euclidean distance, Mahalanobis distance, ITML [7] and KISSME [9]. The figure demonstrates that our method outperforms all these methods which verifies the effectiveness of the proposed method. Besides, we also compute the cross-camera average mAP and average rank-1 accuracy: 50.84% and 25.24%, respectively.

Moreover, we report the recognition performance for the top 20 ranks compared with 12 state-of-the-arts methods for person re-identification in Table 2 which are gathered from [36] 's homepage. The reported method includes [36], [40], [41], [42] and [15]. The table shows that the proposed method does achieve a performance favorably compared the most of the state-of-the-arts though it does not reach the best performance. It can be found that our method achieves 50.84% rank-1 identification rate and it also performs better than the other three methods listed above. However, it is worth noting that the methods reported in the table are most based on the deep neural network framework except feature based method and our method which is based on metric learning framework.

Note that the proposed method performs poor than Multiregion [15] and Liu et al.[42] method. We argue that the main reason is that it adopts a different framework where ours is based on metric learning while the mentioned ones adopts deep neural networks. It is well known that the methods based on deep neural network usually achieve better performance compared with classic methods due to their impressing feature expression and learning ability. Namely, deep neural network often adopts a more discriminative feature than the hand-crafted feature used for metric learning.

Nevertheless, it is also noted that our proposed method achieves a better performance than the two of the neural network based methods as shown in Table 2 which dose prove the effectiveness of the proposed method.

**Table 3** Comparison of state-of-the-art results reported on the CUHK01 database. The cumulative matching scores (%) at rank 1,5,10 and 15 are listed.

| Method | r=1 | r=5 | r=10 | r=15 |
|---|---|---|---|---|
| LMNN[6] | 13.5 | 31.2 | 41.8 | 48.5 |
| ITML[7] | 16.0 | 28.5 | 45.3 | 53.5 |
| sSDC[27] | 19.7 | 33.1 | 40.5 | 46.8 |
| genericM[43] | 20.0 | 44.1 | 57.1 | 64.3 |
| PatMatch[44] | 20.4 | 34.1 | 41.0 | 47.3 |
| FPNN[16] | 27.9 | - | - | - |
| SalMatch[44] | 28.5 | 46.3 | 57.2 | 64.1 |
| mFilter[45] | 34.3 | 55.0 | 65.3 | 70.0 |
| Ejaz[14] | 47.5 | - | - | - |
| Cheng[46] | 46.0 | 67.7 | 78.7 | **85.3** |
| Sakrapee[47] | **53.4** | **76.4** | **84.4** | - |
| proposed | 48.83 | 68.79 | 76.34 | 80.73 |

### *4.4 CUHK01 Dataset*

The CUHK01 Dataset [37] is captured in a campus environment with two camera views. It contains 971 individuals and each of them has two images in every camera view. Taking the evaluation method in [37], we conduct the experiments over 10 random partitions for this dataset, where 485 persons are randomly sampled for training and the rest are utilized for testing. In this experiment, we do not report the CMC curve because this is a widely used dataset and the results can be found in a lot of related works.

Table 3 shows the state-of-the-art results reported on the CUHK01 database. It can be found that the proposed method achieves 48.83 % rank-1 identification rates. Although it is a little lower than the best result, it performs equivalently with [14] method at rank-1 identification rate and shows promising performance compared with the rest methods in Table 3. It is noted that [14] is a deep neural network based methods, so it shows the effectiveness of the proposed methods.

It is noted that [47] achieves better performance than the proposed one and the [47] adopts a metric ensemble framework which fuses multi-metrics. We argue that the reason [47] outperforms the proposed one lies in that fusion model often performs better than single model. As the proposed one is a single model rather than a fused ensemble model. Except [47], we achieve a favorably performance against other state-of-the-arts methods.

## 5 Conclution

In this paper, we propose a new metric learning method for person re-identification. The proposed method can be modeled as a constrained optimization problem by imposing a constraint on the linear transformation. Besides, a weight learning strategy is introduced to learn the weights instead of designing weight intuitively. And we adopt an efficient method based on matrix optimization to solve the proposed cost function and an algorithm is formed for person re-identification. Experiments on three challenging person re-identification databases show that the proposed method performs favorably against the state-of-the-art approaches.

In the future work, a different base metric learning frameworks can be applied such a logistic metric learning framework rather than LDA-based framework. This modification can be used to verify the effectiveness of the weight learning method as well as make it a new metric learning method for person re-identification.

It is also noted that there are some limitations to be aware of. The optimization procedure is complex to some extent and some constraints are strict enough to be relaxed for further modification. So weight learning methods can be further developed in the future.

## 6 Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant U1864204 and 61773316, State Key Program of National Natural Science Foundation of China under Grant 61632018, Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, Projects of Special Zone for National Defense Science and Technology Innovation, Fundamental Research Funds for the Central Universities under Grant 3102017AX010, and Open Research Fund of Key Laboratory of Spectral Imaging TechnologyčňChinese Academy of Sciences.

## 7 References

1 Gong, S., Cristani, M., Yan, S., Loy, C.C.: 'Person re-identification'. (Springer, 2014)
2 Saghafi, M.A., Hussain, A.: 'Review of person re-identification techniques', *IET Computer Vision*, 2014, **8**, (6), pp. 455–474
3 Gray, D., Tao, H. 'Viewpoint invariant pedestrian recognition with an ensemble of localized features'. In: Proc. European Conference on Computer Vision. (Springer, 2008. pp. 262–275
4 Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M. 'Person re-identification by symmetry-driven accumulation of local features'. In: Proc. Computer Vision and Pattern Recognition. (IEEE, 2010. pp. 2360–2367
5 Liao, S., Hu, Y., Zhu, X., Li, S.Z. 'Person re-identification by local maximal occurrence representation and metric learning'. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (Boston, USA, 2015. pp. 2197–2206
6 Weinberger, K.Q., Saul, L.K.: 'Distance metric learning for large margin nearest neighbor classification', *JMLR*, 2009, **10**, (Feb), pp. 207–244
7 Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S. 'Information-theoretic metric learning'. In: ICML. (ACM, 2007. pp. 209–216
8 Mignon, A., Jurie, F. 'Pcca: A new approach for distance learning from sparse pairwise constraints'. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (Providence, USA: IEEE, 2012. pp. 2666–2672
9 Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H. 'Large scale metric learning from equivalence constraints'. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition. (Providence, USA: IEEE, 2012. pp. 2288–2295
10 Zheng, W.S., Gong, S., Xiang, T. 'Person re-identification by probabilistic relative distance comparison'. In: Proc. Computer Vision and Pattern Recognition. (Colorado Springs, USA: IEEE, 2011. pp. 649–656
11 Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B. 'Local fisher discriminant analysis for pedestrian re-identification'. In: Proc. Computer Vision and Pattern Recognition. (Portland, USA, 2013. pp. 3318–3325
12 Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R. 'Learning locally-adaptive decision functions for person verification'. In: Proc. Computer Vision and Pattern Recognition. (Portland, USA, 2013. pp. 3610–3617
13 Varior, R.R., Haloi, M., Wang, G. 'Gated siamese convolutional neural network architecture for human re-identification'. In: Proc. European Conference on Computer Vision. (Amsterdam,Netherlands: Springer, 2016. pp. 791–808
14 Ahmed, E., Jones, M., Marks, T.K. 'An improved deep learning architecture for person re-identification'. In: Proc. Conference on Computer Vision and Pattern Recognition. (Boston, USA, June, 2015. pp. 3908–3916
15 Ustinova, E., Ganin, Y., Lempitsky, V. 'Multi-region bilinear convolutional neural networks for person re-identification'. In: Proc. Advanced Video and Signal Based Surveillance. (Lecce, Italy: IEEE, 2017. pp. 1–6
16 Li, W., Zhao, R., Xiao, T., Wang, X.: 'Deepreid: Deep filter pairing neural network for person re-identification', *Proc Conference on Computer Vision and Pattern Recognition*, 2014, pp. 152–159
17 Jose, C., Fleuret, F. 'Scalable metric learning via weighted approximate rank component analysis'. In: European conference on computer vision. (Amsterdam, Neterlands: Springer, 2016. pp. 875–890
18 Karanam, S., Gou, M., Wu, Z., Rates.Borras, A., Camps, O.I., Radke, R.J.: 'A comprehensive evaluation and benchmark for person re-identification: Features', *Metrics, and Datasets arXiv preprint*, 2016,
19 Ma, B., Su, Y., Jurie, F. 'Local descriptors encoded by fisher vectors for person re-identification'. In: European Conference on Computer Vision. (Springer, 2012. pp. 413–422
20 Liu, L., Lu, X., Yuan, Y., Li, X. 'Person re-identification by bidirectional projection'. In: Proceedings of International Conference on Internet Multimedia Computing and Service. (New York, NY, USA: ACM, 2014. pp. 1–5
21 Wang, Q., Gao, J., Yuan, Y.: 'A joint convolutional neural networks and context transfer for street scenes labeling', *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**, pp. 1457–1470
22 Wang, Q., Gao, J., Yuan, Y.: 'Embedding structured contour and location prior in siamesed fully convolutional networks for road detection', *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**, pp. 230–241
23 Vishwakarma, D.K., Upadhyay, S.: 'A deep structure of person re-identification using multi-level gaussian models', *arXiv preprint arXiv:180507720*, 2018,
24 Guo, Y., Cheung, N.M.: 'Efficient and deep person re-identification using multi-level similarity', *arXiv preprint arXiv:180311353*, 2018,
25 Chen, D., Yuan, Z., Chen, B., Zheng, N. 'Similarity learning with spatial constraints for person re-identification'. In: Proc. Computer Vision and Pattern Recognition. (Las Vegas, USA, 2016. pp. 1268–1277
26 Shi, H., Yang, Y., Zhu, X., Liao, S., Lei, Z., Zheng, W., et al. 'Embedding deep metric for person re-identification: A study against large variations'. In: Proc. European Conference on Computer Vision. (Amsterdam, Netherlands: Springer, 2016. pp. 732–748
27 Zhao, R., Ouyang, W., Wang, X. 'Unsupervised salience learning for person re-identification'. In: Proc. Computer Vision and Pattern Recognition. (Potland, USA, 2013. pp. 3586–3593
28 Wu, Z., Li, Y., Radke, R.J.: 'Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features', *IEEE transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**, (5), pp. 1095–1108
29 Matsukawa, T., Okabe, T., Suzuki, E., Sato, Y. 'Hierarchical gaussian descriptor for person re-identification'. In: Proc. Computer Vision and Pattern Recognition. (Las Vegas,USA, June 2016. pp. 1363–1372
30 Tao, D., Jin, L., Wang, Y., Yuan, Y., Li, X.: 'Person re-identification by regularized smoothing kiss metric learning', *IEEE Transactions on Circuits and Systems for Video Technolog*, 2013, **23**, (10), pp. 1675–1685
31 Dong, H., Lu, P., Zhong, S., Liu, C., Ji, Y., Gong, S.: 'Person re-identification by enhanced local maximal occurrence representation and generalized similarity metric learning', *Neurocomputing*, 2018, **307**, pp. 25–37
32 Yang, X., Wang, M., Tao, D.: 'Person re-identification with metric learning using privileged information', *IEEE Transactions on Image Processing*, 2018, **27**, (2), pp. 791–805
33 Zhang, J., Yuan, Y., Wang, Q.: 'Largest center-specific margin for dimension reduction', *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 2352–2356
34 Lin, Z., Liu, R., Su, Z. 'Linearized alternating direction method with adaptive penalty for low-rank representation'. In: Advances in Neural Information Processing Systems. (, 2011. pp. 612–620
35 Guo, X. 'Robust subspace segmentation by simultaneously learning data representations and their affinity matrix.'. In: Proceedings of International Conference on Artificial Intelligence. (, 2015. pp. 3547–3553
36 Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q. 'Scalable person re-identification: A benchmark'. In: Proceedings of the IEEE International Conference on Computer Vision. (Santiago,Chile, 2015. pp. 1116–1124
37 Li, W., Wang, X. 'Locally aligned feature transforms across views'. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (Portland, USA, 2013. pp. 3594–3601
38 Xiong, F., Gou, M., Camps, O., Sznaier, M. 'Person re-identification using kernel-based metric learning methods'. In: Proc. European Conference on Computer Vision. (Springer, 2014. pp. 1–16
39 Guillaumin, M., Verbeek, J., Schmid, C. 'Is that you? metric learning approaches for face identification'. In: Proc. International Conference on Computer Vision. (Kyoto,Japan: IEEE, 2009. pp. 498–505
40 Wu, L., Shen, C., van den Hengel, A.: 'Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification', *Pattern Recognition*, 2017, **65**, pp. 238–250
41 Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q. 'Deep attributes driven multi-camera person re-identification'. In: European conference on computer vision. (Amsterdam, Netherlands: Springer, 2016. pp. 475–491
42 Liu, J., Zha, Z.J., Tian, Q., Liu, D., Yao, T., Ling, Q., et al. 'Multi-scale triplet cnn for person re-identification'. In: Proc. ACM on Multimedia Conference. (ACM, 2016. pp. 192–196
43 Li, W., Zhao, R., Wang, X. 'Human reidentification with transferred metric learning'. In: Proc. Asian Conference on Computer Vision. (Daejeon, Korea: Springer, 2012. pp. 31–44
44 Zhao, R., Ouyang, W., Wang, X. 'Person re-identification by salience matching'. In: Proc. International Conference on Computer Vision. (Potland, USA, 2013. pp. 2528–2535
45 Zhao, R., Ouyang, W., Wang, X. 'Learning mid-level filters for person re-identification'. In: Proc. Computer Vision and Pattern Recognition. (Columbus, USA, 2014. pp. 144–151
46 Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N. 'Person re-identification by multi-channel parts-based cnn with improved triplet loss function'. In: Proc. Computer Vision and Pattern Recognition. (Las Vegas, USA, 2016. pp. 1335–1344
47 Paisitkriangkrai, S., Shen, C., Van.Den.Hengel, A. 'Learning to rank in person re-identification with metric ensembles'. In: Proc. Computer Vision and Pattern Recognition. (Boston, USA, 2015. pp. 1846–1855