IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 8, 2018, accepted July 9, 2018, date of publication July 25, 2018, date of current version August 15, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2857210

Modeling Unknown Class Centers for Metric Learning on Person Re-Identification

YUAN YUAN[1], (Senior Member, IEEE), JIAN'AN ZHANG[1], AND QI WANG[1,2], (Senior Member, IEEE)

[1]School of Computer Science and Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China
[2]Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Qi Wang (crabwq@gmail.com)

This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1002202, in part by the National Natural Science Foundation of China under Grant 61773316, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, in part by the Fundamental Research Funds for the Central Universities under Grant 3102017AX010, and in part by the Open Research Fund of Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences.

ABSTRACT Metric learning is one of the major methods for person re-identification. Most existing metric learning methods for person re-identification first 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. And then, different kinds of methods are formulated to pull the pairs in positive set together and penalize the pairs in negative set such that a good metric is learned. However, such a process has two drawbacks: 1) the size of negative set is often far more than the negative set for which the learning process is largely dominated by the large amount of negative sample pairs and 2) it often experiences tedious optimization procedures to compute pairwise distances which would be computationally intractable in real scenarios, especially for large-scale data sets. To address the above issues, we propose a new simple and effective metric learning method, which gets rid of the pairwise constraints. We take the unknown class center's information into consideration to model the relationships between different classes directly, and a new objective function is formed. In addition, the proposed objective function can be solved by taking advantage of simple matrix multiplications and hence can avoid computationally complex optimization schemes. The proposed algorithm termed largest center-specific margin metric learning is shown to be computationally efficient and can be applied to large-scale person re-identification. Extensive experiments carried out on two challenging large-scale databases (CUHK03 and Market1501) demonstrate that the proposed algorithm performs favorably against the state-of-the-art approaches.

INDEX TERMS Person re-identification, metric learning.

I. INTRODUCTION

In real surveillance scenario, we can usually get one or more images of a person from one special camera view. A fundamental question we want to know is whether the person has appeared in other cameras of the camera network. This problem is often called person re-identification. Formally, Person re-identification refers to the task of recognizing the same person's identity from a network of cameras with non-overlapping fields of view. Person re-identification is an important problem in real applications especially in security and surveillance [1], [2]. It 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. Fig. 1 shows the examples of the above problems and these images are sampled from the VIPeR [3] dataset and the iLIDS dataset [14].

With the efforts of many researchers in the past ten years, a plenty of methods have been proposed for person re-identification [46], [49] and a great development has been achieved in this field. Metric learning is one of the major methods for person re-identification which pays attention to designing a metric function [12], [48], [50] between two given samples and such a method has been widely used for it's efficiency. 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. And then different kinds of methods are

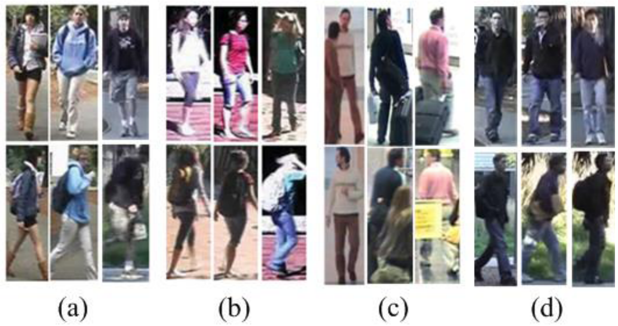


FIGURE 1. (a) Example pairs of images in different viewpoint. (b) Example pairs of images with illumination changes. (c) Example pairs of images have big intra-class invariance. (d)Example pairs of images suffer from occlusion. Images in the same column represent the same person.

formulated to pull the pairs in positive set together and penalize the pairs in negative set such that a good metric is learned.

However, there are some disadvantages for the metric learning methods which are inspired by the pairwise constraints directly. One of the disadvantages is that the numbers of positive and negative sample pairs are largely unbalanced. Especially, for the large-scale person re-identification datasets (e.g. CUHK03 [54] and Market1501 [56]), the size of negative set is far more than the negative set, resulting in that the learning process is largely dominated by the large amount of negative sample pairs. The other drawbacks is that it often experiences tedious optimization procedures to compute the pairwise distances which would be computationally intractable in real scenarios especially for large-scale datasets.

To address the above issues, we propose a new simple and effective metric learning method which gets rid of the pairwise constraints. Firstly, we take the unknown class center's (refers to the center of samples' topological structure) information into consideration to model the relationships between different classes directly. It is noted that such information has not been used in previous works. And then we formulate an objective function based on the observation that after linear transformation, the distances between the points and their class centers should be small while the distances among unknown centers of different class should be large. It can be showed in Fig. 2, where green pentagrams represent unknown class centers for different identities (classes) and circles, triangles and squares are samples from different classes and the arrows show that relationship between different samples. From this intuition, it can be found that we can model the relationships between different classes by means of the unknown class centers rather than the pairwise constraints. And hence the above disadvantages can be addressed and experiments will show the effectiveness of the proposed method.

In addition, the proposed objective function can be reformulated into matrix form and can be solved in a simple and innovative manner by taking advantage of simple matrix multiplications. And hence computationally complex optimization schemes are avoided and it can be

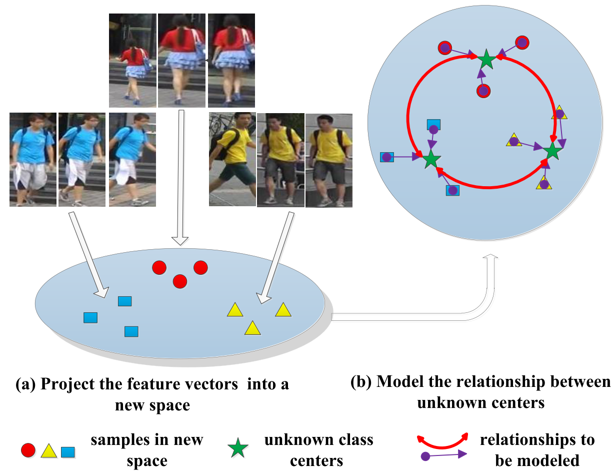


FIGURE 2. Diagram for the intuition, circles and triangles should be close to their diamond centers and the distances between different diamond centers should be as large as possible.

applied to large-scale person re-identification. Furthermore, we also exhibit an algorithm used for person re-identification and name it Largest Center-specific Margin Metric Learning(LCMML). Extensive experiments carried out on two person re-identification benchmarks including CUHK03 [54] and Market1501 [56], illustrate that our LCMML method is a very effective approach especially in the case that every class (identity) has sufficient samples.

The main contributions of this work are two-fold. First, we take the unknown class center's information into consideration which has not been considered in previous works to model the relationships between different classes, and based on this strategy we proposed a new method to solve the problems caused by pairwise constraints. Second, we propose a new simple but efficient metric learning method for person re-identification whose objective function can be solved by using simple matrix multiplications which avoids computationally complex optimization schemes.

II. 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 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 [4], [5], 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. We refer readers to [16], [38], [43], [53], and [55] for more information.

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 set is generated, different kinds of methods are formulated to pull the pairs in positive set together and penalize the pairs in negative set. For instance, the LMNN [20] tries to pull the positive pairs lying within the k-nearest neighbors and push the negative pairs by a large margin. The ITML [21] 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 and Jurie [24] developed a metric learning method named PCCA which adapts to sparse pairwise similarity/dissimilarity constraints in high dimensional input space. KISSME [25] 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 [35]. Zheng *et al.* [23] 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.* [29] employed the LFDA algorithm to maximize the inter-class separability while preserving the multi-class modality. Li *et al.* [28] developed the Locally-Adaptive Decision Functions (LADF), which combines the distance metric with a locally adaptive thresholding rule for each pair of sample images.

Most of the above method based on the same observation that the distances between positive pairs should be small and the distances between negative pairs should be large. However, the proposed approach is formulated based on the intuition that the distances between the points and their unknown class centers should be small while the distances among unknown centers of different class should be large, which is different from the existing works.

III. METHODOLOGY

In this paper, we formulate the person re-identification problem into the following distance metric framework, where we assume each instance of a person is represented by a feature vector (the representation described in experiments section).

The most widely studied approach for metric learning is Mahalanobis distance learning. This is, given n data points $\mathbf{x_i} \in R^{d \times n}, i = 1, 2, \cdots, n$, the goal of Mahalanobis distance metric learning is to estimate a matrix $\mathbf{M}$ such that

$$D_{\mathbf{M}}(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} - \mathbf{x_j})^{\mathbf{T}} \mathbf{M} (\mathbf{x_i} - \mathbf{x_j}). \tag{1}$$

The Eq. (1) describes a pseudo-metric and it can be easily proved that the above metric defines a distance if $\mathbf{M}$ is positive semi-definite, i.e., for any non-zeros vector $\mathbf{x} \in R^m$, it satisfies $\mathbf{x^T M x} \geq 0$. As a special case of $M = \Sigma^{-1}$, the distance defined by Eq. (1) is referred to as the Mahalanobis distance.

An alternative formulation for Eq. (1), which is more intuitive, is given via the squared distance

$$D_{\mathbf{M}}(\mathbf{x_i}, \mathbf{x_j}) = \|\mathbf{A}(\mathbf{x_i} - \mathbf{x_j})\|^2. \tag{2}$$

It can be easily obtained from

$$\mathbf{M} = \mathbf{A^T A}, \tag{3}$$

where $\mathbf{A}$ can be regarded as a linear transformation matrix. From Eq. (2), it is easy to find that the Mahalanobis distance between $\mathbf{x_i}$ and $\mathbf{x_j}$ is equivalent to the Euclidean distance between $\mathbf{Ax_i}$ and $\mathbf{Ax_j}$, which can be regarded as points from the new space after linear transformation of $\mathbf{A}$. Based on this observation, it is easy to consider that we can learn the structure of data in the new space and try to learn a metric based on the linear transform.

In the following, we will introduce an efficient and effective method of metric learning.

A. LARGEST CENTER-SPECIFIC MARGIN METRIC LEARNING

For each class, we assume that there exists an optimal unknown class center. And there is no doubt that the class centers' information is of importance to the metric and hence it can be used to measure the properties of metric. However, such information has not been employed in previous works. So in this work, for the first time, we take the unknown class centers generated after linear transformation into consideration and make use of the information for metric learning.

We build on the simple intuition that after linear transformation the distances between the points of the same label and their corresponding class center should be small enough, and at the same time the distances among unknown centers of different classes should be as large as possible. It will be clearer to understand the above idea from Fig. 2 in which a projection to a new space is shown firstly and then the relationships between samples and their unknown centers as well as the relationships among different class centers are illustrated. Based on the naive intuition mentioned above, we can formulate the idea as below

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}} & f^*(\mathbf{A}, \mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_n}) \\ &= \sum_i \sum_{j \in c_i} \|\mathbf{Ax_i} - \mathbf{y_j}\|^2 - \alpha \sum_j \sum_{\mathbf{c_j} \neq \mathbf{c_i}} \|\mathbf{y_i} - \mathbf{y_j}\|^2, \end{aligned} \tag{4}$$

where $\mathbf{x_i}, i = 1, 2, \ldots, n$ are the feature representations of instances of different classes and $\mathbf{y_j}, j = 1, 2, \ldots, c$ are the unknown class centers. In Eq. (4), the first term implies that after linear transformation $\mathbf{A}$, the distances between the points and their corresponding unknown center of the same class, and the second term represents the distance between two unknown centers of different classes. In intuition, in order to acquire an effective transformation matrix, the first term should be as small as possible and in contrast the second term should be as large as possible. So we transform the second term to the minus term making it a unified optimal problem and α is trade-off between the two terms in case that either term might dominate over the other. As a result, we formulate our basic idea into Eq. (4).

While Eq. (4) is of complex forms for optimization, for simplicity, we reformulate it into the following matrix form by permutating and combining the terms.

$$f^*(\mathbf{A},\mathbf{Y}) = \|\mathbf{AX}-\mathbf{YC}\|_F^2 - \alpha tr(\mathbf{YLY}^T), \tag{5}$$

where $\|\cdot\|_F$ is Frobenius norm and $tr(\cdot)$ stands for the trace operator. Besides, $\mathbf{A} \in R^{m\times d}$ is a linear transformation matrix and $\mathbf{X} \in R^{d\times n}$ denotes the sample matrix which consists of all the training samples and each column is a feature vector of an instance of one class.d is the dimensionality of feature vector and n represents the number of all training samples, and m stands for the dimension of the new space after linear transformation. $\mathbf{Y} = [\mathbf{y_1}, \mathbf{y_2}, \ldots, \mathbf{y_c}] \in R^{m\times c}$ is the matrix synthesized by the centers of c classes. $\mathbf{C} \in R^{c\times n}$ is a matrix with the following form

$$\mathbf{C} = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix}.$$

$\mathbf{L} = \mathbf{I} - \frac{1}{c}\mathbf{11}^T$, which is the centering matrix, $\mathbf{I}$ stands for identity matrix with dimensionality of c, and $\mathbf{1}$ stands for the c-dimensional vector with all elements being 1. Note that, $\|\mathbf{AX}-\mathbf{YC}\|_F^2 = tr(\mathbf{AX}-\mathbf{YC})(\mathbf{AX}-\mathbf{YC})^T$, so Eq. (5) can be simplified into the following form

$$\min_{\mathbf{A},\mathbf{Y}} f^*(\mathbf{A},\mathbf{Y}) = tr(\mathbf{AXX}^T\mathbf{A}^T) - 2tr(\mathbf{Y}^T\mathbf{AXC}^T) + tr(\mathbf{Y}(\mathbf{CC}^T - \alpha(\mathbf{I} + \frac{1}{c}\mathbf{11}^T))\mathbf{Y}^T). \tag{6}$$

Nevertheless, the optimal problem described in Eq.(4) is not guaranteed to be convex. For the convenience of tackling, the original problem can be reformulated as follow, i.e. add two regular terms to the objective function

$$\begin{aligned} \min_{\mathbf{A},\mathbf{y_1},\mathbf{y_2},\ldots,\mathbf{y_n}} & f(\mathbf{A},\mathbf{y_1},\mathbf{y_2},\ldots,\mathbf{y_n}) \\ &= \sum_i \sum_{j\in c_i} \|\mathbf{Ax_i}-\mathbf{y_j}\|^2 \\ &\quad - \alpha \sum_j \sum_{\mathbf{c_j}\neq\mathbf{c_i}} \|\mathbf{y_i}-\mathbf{y_j}\|^2 \\ &\quad + \gamma\|\mathbf{A}\|_F^2 + \eta\|\mathbf{Y}\|_F^2. \end{aligned} \tag{7}$$

Note that, after modification, the new optimal Eq. (7) is jointly convex with regard to $\mathbf{A}$ and $\mathbf{Y}$. Hence this optimal problem has globally optimal solution. Moreover, even if adding two regular terms to the original Eq. (4), the significance of the problem is not changed because the above regular terms are equivalent to imposing constraint to $\mathbf{A}$ and $\mathbf{Y}$ so that the norms of $\mathbf{A}$ and $\mathbf{Y}$ are not too large [13].

In the same manner, we can convert Eq. (7) into the following matrix form based on Eq. (6) and the property of trace operator,

$$\min_{\mathbf{A},\mathbf{Y}} f(\mathbf{A},\mathbf{Y}) = tr(\mathbf{ANA}^T) - 2tr(\mathbf{Y}^T\mathbf{AXC}^T) + tr(\mathbf{YKY}^T), \tag{8}$$

where

$$\mathbf{N} = \mathbf{XX}^T + \gamma\mathbf{I}, \tag{9}$$

$$\mathbf{K} = \mathbf{CC}^T + (\eta-\alpha)\mathbf{I} + \frac{\alpha}{c}\mathbf{11}^T, \tag{10}$$

and from Eq. (9) and Eq. (10) it is required that $\gamma > 0$, $\eta - \alpha > 0$ in order to make $\mathbf{N}$ and $\mathbf{K}$ be symmetric positive semi-definite.

It is noteworthy that the proposed method shares similar idea is simple and has been widely used as criterion for metric learning, such as FDA. However, the proposed method differs from existing methods in two aspects. Firstly, the proposed method takes the new space after linear transformation into consideration and optimization procedure is executed in the new space. Second, the proposed method take the unknown class center in the new space into consideration while existing method often assumes that the class center is the mean of class samples. These two aspects distinguish the proposed method from existing methods

From now on, the optimization objective function has been established. So we can solve the optimal Eq. (8) to obtain the metric matrix for metric learning. It can be discussed in the following section in detail.

B. OPTIMIZATION

Eq. (8) is an unconstraint optimization problem, and the basic optimization methods can be used to solve it such as gradient descend [13] method. However, it is noted that Eq. (8) is of simple structure and is continuous with regard to $\mathbf{A}$ and $\mathbf{Y}$. Hence we can solve it simply by taking the derivation of one variable when fixed the other one and letting the derivation be $\mathbf{0}$.

We can get the derivation of Eq. (8) w.r.t $\mathbf{A}$, $\mathbf{Y}$ and $\mathbf{L}$ easily, which

$$\frac{\partial f(\mathbf{A},\mathbf{Y})}{\partial \mathbf{A}} = \mathbf{AN} - \mathbf{YCX^T},$$

$$\frac{\partial f(\mathbf{A},\mathbf{Y})}{\partial \mathbf{Y}} = \mathbf{YK} - \mathbf{AXC}^T. \tag{11}$$

Let the derivations be zero matrix and we can get

$$\mathbf{A} = \mathbf{YCX^TN}^{-1}, \tag{12}$$

and

$$\mathbf{Y} = \mathbf{AXC}^T\mathbf{K}^{-1}. \tag{13}$$

It is noted that from Eq. (12) and Eq. (13), we can calculate A and Y iteratively, and at last we can get an optimal A when a given error bound is reached. Finally, we can get the metric matrix using Eq. (3). It is noted that the optimization problem is a global convex with regards to $\mathbf{A}$ and $\mathbf{Y}$, we can take a random matrix as an initial value. In practice, we take

Algorithm 1 LCMML Algorithm for Person Re-Identification

Input:
The n training samples with corresponding labels $(\mathbf{x_i}, \mathbf{y}_i)_{i=1}^n$

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

1: Initialize parameters γ, η, class numbers c, set $k = 0$, and initialize $\mathbf{Y}_k$;
2: Construct matrix $\mathbf{C}$, and calculate matrix $\mathbf{N}$, $\mathbf{K}$ from E.q.(9) and E.q.(10)
3: $\mathbf{A}_k \leftarrow \mathbf{Y}_k\mathbf{C}\mathbf{X}^\mathbf{T}\mathbf{N}^{-1}$
4: **while** $k < k_{max}$ **do**
5: $k \leftarrow k + 1$;
6: update $\mathbf{Y_k} \leftarrow \mathbf{A_{k-1}}\mathbf{X}\mathbf{C}^T\mathbf{K}^{-1}$
7: update $\mathbf{A_k} \leftarrow \mathbf{Y}_{k-1}\mathbf{C}\mathbf{X}^T\mathbf{N}^{-1}$
8: **end while**
9: Set $\mathbf{M} \leftarrow \mathbf{A}_k^T\mathbf{A}_k$

the means of classes as initial value. The pseudo-code of LCMML method is outlined in Algorithm 1 in detail.

Furthermore, from the objective function, we can draw that the more samples one class has, the better performance our LCMML algorithm can reach. This is because when one class has a large amount of samples, the unknown class center's information will be rich enough to model the relationships. So from this analysis, it can reach that our LCMML algorithm is more sufficient to the case that one class(identity) has enough samples. It can also be validated in the following experiments section. So our LCMML method can handle large-scale person re-identification problem efficiently.

In addition, from the Algorithm 1, we can conclude that only matrix multiplication operations are involved in the algorithm, which is efficient in real computation.

IV. EXPERIMENTS

We validate the proposed approach in two widely used large scale datasets in person re-identification, namely, the CUHK03 [54] dataset and the Market1501 [56] dataset. Furthermore, we also evaluate our proposed LCMML method for the effect of parameters and the computation time. Details can be showed in the following four sections.

It is noted that, we do not reported the state-of-the-arts deep learning based methods as the purpose of this works is to develop a metric learning methods. We have to mentioned that recent deep learning based methods have achieved state-of-the-art performance, but we believe that the proposed method is also of its value when combined with deep learned features.

A. FEATURE EXTRACTION AND SETTINGS

1) FEATURE EXTRACTION

Features are important for the performance of person re-identification, among which LOMO [36] is regarded as one of the most efficient features. In addition, existing state-of-the-arts metric learning method for person re-identification are mostly reported based on the LOMO feature, such as XQDA [36], MLANG [40] and Null Space [39]. In order to provide a fair comparison, we also perform our LCMML algorithm on the basis of LOMO feature.

The LOMO feature firstly performed a multiscale Retinex transformation for all images in the datasets in order to achieve a consistent color representation. Next, the maximal pattern of joint HSV color histogram coupled with Scale Invariant Local Ternary Pattern (SILTP) are extracted for a set of sliding windows to overcome the difficulty in view-point changes. At last, it generated a 26,960 dimensional LOMO descriptor.

2) BASELINE METHODS

In this paper, we evaluate the proposed LCMML algorithm compared with several promising algorithms, including LFDA [29], oLFDA [34], ITML [21], KISSME [25], svmml [28] and MFA [34]. And we adopt a single shot experiment setting in both CUHK03 dataset and Market1501 dataset which is similar to the most of the previous works. 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.

It's noted that a series of Deep Neural Network (DNN) based methods have not been listed in this work. This is because the metric learning based methods and DNN based methods adopt different pipelines and it is not fair to compare them directly. On the other hand, metric learning based methods can achieve a better performance with a more efficient feature while DNN based methods cannot reach a better performance when the training process is finished. For example, with recent proposed GOG [17] feature, the performance of the existing metric learning methods are all improved. So in consideration of the above two reasons, we do not report the performance of DNN based methods and we refer the readers to the work [16] for more information about the DNN based methods.

3) PARAMETERS SETTINGS

For our LCMML method, there are three parameters to be determined before the algorithm starts and they are γ, η and α. The effect of the parameters on the algorithm will be evaluated in next section. Without the loss of generality, we set $\gamma = 1$, $\alpha = 1/16$ and $\eta - \alpha = 1/4$ for all of the following experiments. For the baseline methods, we follows the setting of [34] in which a PCA procedure is applied to KISSME [25] method and the dimension is set to be 65. For svmml method, the iteration number is set to be 300.

B. CUHK03 DATASET

The CUHK03 dataset [54] 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 mis-alignments and body part missing for a more realistic setting.

We run our LCMML algorithm on both detected and labeled bounding boxes with the same setting as [34]. 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 10 random splits and the average results are presented. In addition, the metric learning methods that we used for comparison include LFDA [29], KISSME [25], svmml [28], oLFDA [34] and MFA [34]. Fig. 3 shows the result of CMC curve on the detected bounding boxes and Fig. 4 demonstrates the result on the labeled bounding boxes.

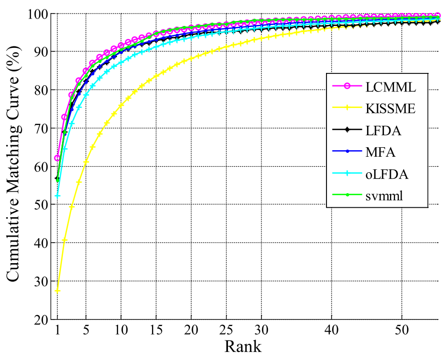


FIGURE 3. CMC curves comparing LCMML against baseline methods on CUHK03 dataset (detected bounding boxes).

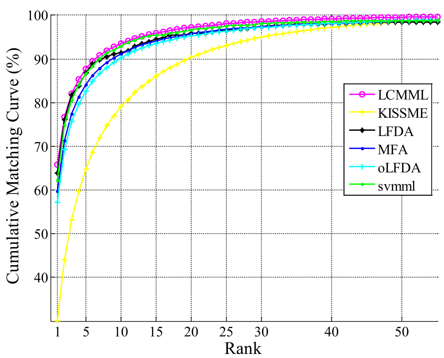


FIGURE 4. CMC curves comparing LCMML against baseline methods on CUHK03 dataset (labeled bounding boxes).

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 LFDA method by 5.27% in rank 1 identification rate with the detected bounding boxes and 1.90% rank 1 identification rate with the labeled bounding boxes. Furthermore, we can also find that our LCMML algorithm is robust to the detected images because compared with the labeled images it only reduces 3.64% rank 1 identification rate. However, the second best method LFDA reduces 7.01% rank 1 identification rate which is far more than our LCMML algorithm. We argue that this is because that the unknown centers for the two different bounding boxes are of small difference in the new space after linear transformation. Besides, the input $\mathbf{Y_0}$ also reduces the difference between detected images and labeled images.

We also compare our LCMML method with other state-of-the-arts metric learning methods which is shown in Table 1. The comparison methods include ITML [21], SDALF [19], eSDC [30], LMNN [20], LDML [22], XQDA [36], MLAPG [40] and Null Space [39]. It is noted that XQDA [36], MLANG [40] and Null Space [39] are regarded as the three most efficient metric learning methods for person re-identification. From Table 1, we can find that our LCMML achieves 65.84 % and 62.20 % rank 1 identification rates with the labeled bounding boxes and the automatically detected bounding boxes, respectively. Our LCMML method achieves the best performance with both the detected bounding boxes and the detected bounding boxes compared with the three best metric learning methods.

TABLE 1. Comparison of state-of-the-art rank-1 identification rates(%) on the CUHK03 database with both labeled and detected setting (P = 100). The compared results are from [36]and [37].

Method	labled	detected
ITML	5.53	5.14
SDALF	5.60	4.87
eSDC	8.76	7.68
LMNN	7.29	6.25
LDML	13.51	10.92
LOMO+XQDA	52.20	46.25
LOMO+MLAPG	57.96	51.15
Null Space(LOMO)	58.90	53.70
LCMML(ours)	**65.84**	**62.20**

We argue that the reasons of achieving such a performance lie on three aspects: (1) the unknown class center information that has been modeled in the proposed method is effective in person re-identification. (2) the CUHK03 dataset is large enough (there are 4.8 images on average for each identity) to guarantees effectiveness of the model. (3) the LOMO feature is robust and representative for person re-identification. So based on these three reasons, our algorithm can achieve a good performance.

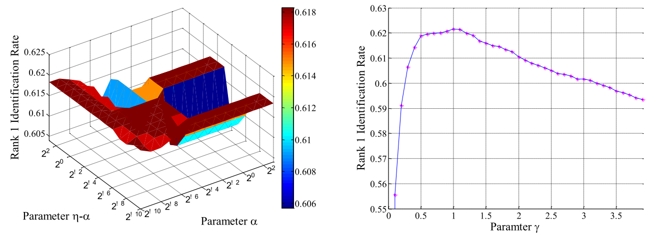
C. MARKET1501 DATASET

Market1501 dataset [56] 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 [56]. 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. In this dataset, we use the provided fixed training and test set to evaluate the proposed method.

We report the recognition performance for the top 20 ranks compared with the state-of-the-art metric learning based methods for person re-identification in Table 2. Comparison algorithms include baselines reported in [56],

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	8.28	-	-	2.23
HistLBP+L-2	9.62	-	-	2.72
LOMO+L-2 26.7	-	-	7.75	
BoW+L-2	35.84	60.33	67.64	14.75
Bow+KISSME	44.42	72.18	78.95	20.76
LOMO+KISSME	40.50	-	-	19.02
LOMO+XQDA	43.79	-	-	22.22
LOMO+MFA	45.67	-	-	18.24
LOMO+kLFDA	51.37	-	-	24.43
LCMML(ours)	**53.56**	**82.07**	**87.23**	**27.38**



(a) (b)

FIGURE 5. Experiments of Parameters (a) The rank 1 identification rate changes with the increase of parameters α and $\eta - \alpha$ by fixing parameter $\gamma = 1$. (b)The rank1 identification rate changes with the increase of parameter γ by fixing $\alpha = 2^{-4}$ and $\eta - \alpha = 2^{-2}$.

KISSME [25], XQDA [36], MFA [34] and kLFDA [34].It is noted that in this dataset the metric learning based methods are not the state-of-the-arts and we refer readers to [16] which achieves the best performance using a gated siamese convolutional neural network. However, as a metric learning based method, the table shows that the proposed method does achieve a performance better than most of the state-of-the-arts. It can be found that our LCMML method achieves 53.56% rank 1 identification rate and it improves the second best performance by 2.21 % identification rate (compared with the LOMO+kLFDA). As has been analyzed before, our LCMML fits the case that every class has sufficient samples and in Market1501 dataset one identity has more than 5 samples in one view, so the proposed method reaches the best performance compared with the state-of-the-arts.

D. ANALYSIS OF THE PROPOSED METHOD

To better understand the proposed method, we analyze it in the following aspects: the effect of parameters and the running time. The analysis was performed on the CUHK03 database with detected bounding box.

1) DISCUSSION ON PARAMETERS

It can be found that there are three parameters used in the proposed method and they are γ, η and α. It is necessary to evaluate the proposed algorithm under different values of parameters for better understanding the method. We evaluate the parameters in two-fold. First, by fixing the parameter γ, we evaluate performance of the proposed algorithm with the changes of parameter α and $\eta - \alpha$ which are the parameters in Eq. (10). It is noted that the value $\eta - \alpha$ will be irregular if we assign a regular sequence value of η so we evaluate it as a whole parameter for convenience.

It is noted that the parameter α is the trade-off between the first term and the second term in Eq. (4). So we choose a sequence of values $2^{-10}, 2^{-9}, \cdots, 2^2$ to evaluate the performance by fixing parameter $\gamma = 1$ and so does the parameter $\eta - \alpha$. The result can be seen in Fig. 5 (a), in which the color bar refers to rank1 identification rate,(red color indicates a higher precision rate and blue color indicates a lower precision rate). From Fig. 5 (a), we can draw that the rank 1 identification rate will reach a high performance when α is less than 2^{-4} and $\eta - \alpha$ is less than 2^{-2} and typically a smaller value of $\eta - \alpha$ will has more effect than the larger ones. Besides, we can also find that even with the change of the parameters the rank 1 identification rate change within 2 percentage, which illustrates that our LCMML method is robust to the two parameters. So without loss of generality, we can set that $\alpha = 2^{-4}$ and $\eta - \alpha = 2^{-2}$ in all the experiments.

Second, by fixing the parameters α and $\eta - \alpha$, we evaluate the performance of the proposed algorithm under different values of the parameter γ. We first perform a coarse evaluation in a large scale value of γ and then we focus on the scale of $0 - 4$ to fine-tune the parameter γ. Fig. 5 (b) shows the rank 1 identification rate changes with the increase of parameter γ. It can be found that we can achieve the best performance when $\gamma = 1$ and when the value of γ is less of larger than 1 the performance will be reduced. Especially when the value of γ is less than 0.5, the rank 1 identification rate will drop rapidly. So in the experiments, we set $\gamma = 1$.

2) DISCUSSIONS ON COMPUTING TIME

We also evaluate the training time of our LCMML method compared with the other metric learning algorithms. The result is shown in Table 3. The training time are averaged over 10 random trials on the CUHK03 dataset and the training is performed on a desktop PC with an Intel i5-3470 @3.20 GHz CPU. From Table 3, it can be found that our LCMML algorithm can be trained in a very fast way even based on the 26960 dimensional LOMO feature. It is noted that KISSME method is trained in a fastest way and this is because a PCA dimensionality reduction procedure is executed for it while the LCMML algorithm is not involved. However, based on the LOMO feature used in CUHK03 dataset, the PCA procedure used in KISSME method takes 1265 seconds which is far more than the training time of our LCMML method. Table 3 shows the efficiency of the proposed LCMML method compared with svmml, LFDA and oLFDA which are all involved in complex optimization procedures. Especially the svmml method takes more than 19.6 hours (70590 seconds) per trial to train which is too expensive to use even though it achieve a good performance. So it can be drawn that the proposed

TABLE 3. Training time (seconds) of metric learning algorithms. The star represents that a PCA dimension reduction procedure is applied to this method.

Method	Times	Method	Times
KISSME	**0.076***	MFA	456.3
LCMML(ours)	212.4	oLFDA	1243.2
LFDA	444.0	svmml	70590

LCMML algorithm can train in an efficient way even if the high dimensional feature is applied.

V. CONCLUSION AND FUTURE WORKS

Existing metric learning method for person re-identification has two drawbacks: 1) the size of negative set is often far more than the negative set for which the learning process is largely dominated by the large amount of negative sample pairs, 2) it often experiences tedious optimization procedures to compute pairwise distances which would be computationally intractable in real scenarios especially for large-scale datasets. In this paper, we propose a simple and effective approach to metric learning named LCMML that is used for person re-identification which is designed to get rid of the pairwise constraints. The proposed metric learning method assumes that for each class there exists an optimal unknown class center and then takes the unknown class center into consideration to formulate an object function which can be solved efficiently. The effect of parameters for LCMML has been discussed and we can draw that the iteration procedure in the proposed algorithm is not needed and hence it is suited to large-scale person re-identification. Besides, experiments on two challenging person re-identification databases CUHK03 and Market1501 show that the proposed LCMML method performs favorably against the state-of-the-art approaches.

Future works will focus on the following three aspects: (1) Take different initial values as input to evaluate the algorithm, such as to model the class center in a new way or take the existing trained Mahalanobis matrix as input. (2) Evaluate the algorithm with more efficient features such as GOG. (3) Evaluate the algorithm in new datasets such as CUHK01 [55].

REFERENCES

[1] L. Sun, Z. Jiang, Q. Lu, and A. Men, "Semi-coupled dictionary learning with relaxation label space transformation for video-based person re-identification," *IEEE Access*, vol. 6, pp. 12587–12597, 2018.

[2] T. Ni, Z. Ding, F. Chen, and H. Wang, "Relative distance metric leaning based on clustering centralization and projection vectors learning for person re-identification," *IEEE Access*, vol. 6, pp. 11406–11411, 2018.

[3] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, 2008, pp. 262–275.

[4] Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1457–1470, May 2018.

[5] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 230–241, Jan. 2018.

[6] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 242–252, Feb. 2015.

[7] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Re-identification in the function space of feature warps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1656–1669, Aug. 2015.

[8] W.-S. Zheng, S. Gong, and T. Xiang, "Towards open-world person re-identification by one-shot group-based verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 591–606, Mar. 2016.

[9] Z. Wu, Y. Li, and R. J. Radke, "Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1095–1108, May 2014.

[10] G. Lisanti, I. Masi, A. Bagdanov, and A. D. Bimbo, "Person re-identification by iterative re-weighted sparse ranking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1639–1642, Aug. 2015.

[11] R. V. Rama, G. Wang, and J. Lu, "Learning invariant color features for person re-identification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3395–3410, Feb. 2016.

[12] L. Liu, X. Lu, Y. Yuan, and X. Li, "Person re-identification by bidirectional projection," in *Proc. Int. Conf. Internet Multimedia Comput. Service*, pp. 1–8, 2016.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[14] W. Zheng, S. Gong, and X. Tao, "Associating groups of people," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 1–8.

[15] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1512–1523, Jul. 2009.

[16] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 791–808.

[17] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1363–1372.

[18] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by Fisher vectors for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 413–422.

[19] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2356–2367.

[20] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.

[21] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon, "Information-theoretic metric learning," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 209–216.

[22] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2009, pp. 498–505.

[23] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2011, pp. 649–656.

[24] A. Mignon and J. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2666–2672.

[25] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2288–2295.

[26] R. Zhao, W. Ouyang, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. Asian Conf. Comput. Vis.*, 2012, pp. 31–44.

[27] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2013, pp. 2528–2535.

[28] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3610–3617.

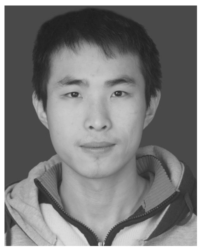
[29] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3318–3325.

[30] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3586–3593.

[31] X. Zhu *et al.*, "Distance learning by treating negative samples differently and exploiting impostors with symmetric triplet constraint for person re-identification," in *Proc. Int. Conf. Multimedia Expo*, 2016, pp. 1–6.

[32] M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *Proc. IEEE Adv. Video Signal-Based Surveillance*, Sep. 2012, pp. 203–208.
[33] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 144–151.
[34] F. Xiong, M. Guo, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2014, pp. 1–16.
[35] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing KISS metric learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1675–1685, Oct. 2013.
[36] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2197–2206.
[37] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3908–3916.
[38] S. Paisitkriangkrai, C. Shen, and A. Hengel, "Learning to rank in person re-identification with metric ensembles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1846–1855.
[39] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1239–1248.
[40] S. Liao and S. Z. Li, "Efficient PSD constrained asymmetric metric learning for person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3685–3693.
[41] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
[42] E. Ustinova, Y. Ganin, and V. Lempitsky, "Multi-region bilinear convolutional neural networks for person re-identification," in *Proc. IEEE Adv. Video Signal-Based Surveillance*, Aug. 2017, pp. 1–6.
[43] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1335–1344.
[44] C. Jose and F. Fleuret, "Scalable metric learning via weighted approximate rank component analysis," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 875–890.
[45] N. Martinel, A. Das, M. Christian, and A. R. Chowdhury, "Temporal model adaptation for person re-identification," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 858–877.
[46] S. Karanam *et al.* (2016). "A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets." [Online]. Available: https://arxiv.org/abs/1605.09653
[47] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1268–1277.
[48] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zheng, "Joint learning of single-image and cross-image representations for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1288–1296.
[49] S. Gong, M. Cristani, S. Yan, and C. Chen, *Person Re-Identication*. New York, NY, USA: Springer, 2014.
[50] P. Peng *et al.*, "Unsupervised cross-dataset transfer learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1–10.
[51] L. Wu, C. Shen, and A. Hengel. (2016). "PersonNet: Person re-identification with deep convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1601.07255
[52] H. Shi *et al.*, "Embedding deep metric for person re-identification: A study against large variations," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 732–748.
[53] L. Wu, C. Shen, and A. van den Hengel, "Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification," *Pattern Recognit.*, vol. 65, pp. 238–250, May 2017.
[54] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 152–159.
[55] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3594–3601.
[56] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2015, pp. 1116–1124.

YUAN YUAN (M'05–SM'09) is currently a Full Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. She has authored or co-authored over 150 papers, including about 100 in reputable journals, such as the IEEE TRANSACTIONS AND PATTERN RECOGNITION, and also conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.



JIAN'AN ZHANG received the B.E. degree in information and computing science from the Ocean University of China, Qingdao, China, in 2015. He is currently pursuing the Ph.D. degree with the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an. His research interests include computer vision and pattern recognition.

QI WANG (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science, the Unmanned System Research Institute, and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.