Discriminative deep transfer metric learning for cross-scenario person re-identification

Tongguang Ni
Xiaoqing Gu
Hongyuan Wang
Zhongbao Zhang
Shoubing Chen
Cui Jin

Discriminative deep transfer metric learning for cross-scenario person re-identification

Tongguang Ni, Xiaoqing Gu, Hongyuan Wang,* Zhongbao Zhang, Shoubing Chen, and Cui Jin
Changzhou University, School of Information Science and Engineering, Changzhou, China

Abstract. A discriminative deep transfer metric learning method called DDTML is proposed for cross-scenario person re-identification (Re-ID). To develop the Re-ID model in a new scenario, a large number of pairwise cross-camera-view person images are deemed necessary. However, this work is very expensive due to both monetary cost and labeling time. In order to solve this problem, a DDTML for cross-scenario Re-ID is proposed using the transferring data in other scenarios to help build a Re-ID model in a new scenario. Specifically, to measure distribution difference across scenarios, a maximum mean discrepancy based on class distribution called MMDCD is proposed by embedding the discriminative information of data into the concept of the maximum mean discrepancy. Unlike most metric learning methods, which usually learn a linear distance to project data into the feature space, DDTML uses a deep neural network to develop the multilayers nonlinear transformations for learning the nonlinear distance metric, while DDTML transfers discriminative information from the source domain to the target domain. By bedding the MMDCD criteria, DDTML minimizes the distribution divergence between the source domain and the target domain. Experimental results on widely used Re-ID datasets show the effectiveness of the proposed classifiers. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JEI.27.4.043026]

1 Introduction

For the last decade, surveillance systems have become an active research topic of computer vision, since they have become ubiquitous in public places such as airports, railway stations, college campuses, and office buildings.[1–2] There are a large number of cameras in surveillance systems and they provide huge amounts of video data. The analysis of the computer vision abstained in a surveillance system often requires the ability to track people across multiple cameras. Therefore, person re-identification (Re-ID) model is generating more and more interests.[3–8] Re-ID has been widely treated as a recognition problem of matching different persons across disjoint cameras.[9–11] In the past five years, a large number of models have been proposed for Re-ID models. The current work can be categorized generally into two types: (1) designing of discriminative, descriptive, and robust visual descriptors to characterize a person's appearance[12–15] and (2) learning suitable distance metrics that maximize the chance of a correct correspondence.[16–20] In this paper, we focus on the second type, i.e., we learn the optimal distance measure to give correct matches in Re-ID.

However, it is not easy to develop a deployable and efficient Re-ID model in a new scenario (e.g., from an indoor classroom to an outdoor square). First, due to different illumination environments, posture, and view angle, the robust features obtained in one scenario will not have good performance for another scenario. Second, in order to obtain a robust Re-ID model, one must collect a large number of labeled person images about the new scenario for training.

However, the work is very expensive due to both monetary cost and labeling time. Some unsupervised methods are proposed to address this problem. For example, Ma et al.[21] introduced a time shift dynamic time warping model for unsupervised person representation. Ye et al.[22] proposed a dynamic graph matching method to mine the intermediate estimated labels across disjoint cameras, and then with the estimated labels, its remaining steps can be considered as a supervised learning method. However, compared to supervised Re-ID methods, the matching performance of unsupervised methods is less effective when a person recognizable is under severe appearance changes.[23]

Recently, transfer learning mechanism has been widely used in Re-ID. The principal goal of transfer learning is to help build a Re-ID model in a new scenario (target domain) by leveraging the data collected from the other scenarios (source domain).[24] For example, in a crowded station, there may exist of a large number of data used for building some Re-ID models for their own respective scopes. In order to build a Re-ID model for a new scenario, we may use these existed data in the source domain without collecting a lot of labeled data in the target domain. In Ref. 25, it is demonstrated that certain discriminative information or common variations (such as pose and resolution) shared in different scenarios can lead to significant performance gains in a new scenario. Different from original multitask learning which aims to benefit all tasks both on the target domain and source domain, transfer learning for Re-ID mainly aims to benefit the target one.

In this work, we first propose a maximum mean discrepancy based on class distribution called MMDCD to measure distribution difference across domains. MMDCD embeds the

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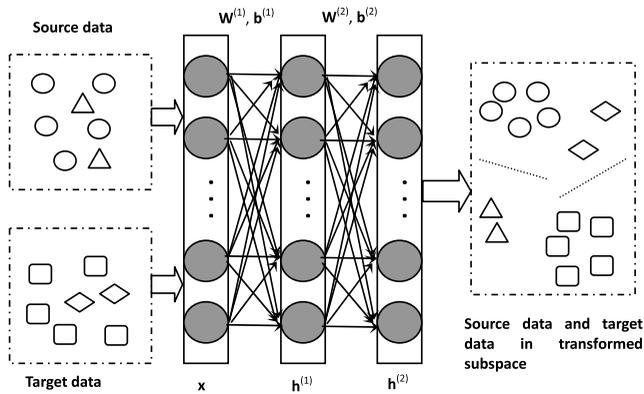*Address all correspondence to: Hongyuan Wang, E-mail: hywang@cczu.edu.cn

Fig. 1 Framework of the proposed method DDTML.

discriminative information of data taken from the source domain into the concept of the maximum mean discrepancy (MMD).[26] Minimizing MMDCD leads to minimize the distribution difference across domains in a supervised way. Then we propose a discriminative deep transfer metric learning method called DDTML for cross-scenario transfer Re-ID. Figure 1 shows the basic idea of the proposed method. Using a deep neural network, DTDML learns a set of multilayers nonlinear transformations to transfer discriminative information from the source domain to the target domain; meanwhile, DTDML reduces the distribution divergence between the source data and the target data by minimizing MMDCD at the top layer of the network.

The contribution of this work can be summarized in the following three aspects.

(1) Unlike MMD working in an unsupervised way, MMDCD works in a supervised way, which not only exploits the discriminative information of data taken from the source domain, but also sets different coefficients for matched/mismatched pairs. Minimizing MMDCD could enhance the discriminant ability of DTDML.

(2) By embedding MMDCD into a deep metric network, DDTML learns a set of multilayers nonlinear transformations to better exploit the discriminative information for cross-scenario Re-ID tasks.

(3) Extensive experiments on several Re-ID datasets are conducted and the experimental results demonstrate that the proposed method DDTML obtains better performance compared with several state-of-the-art methods.

2 Related Work

According to the process of Re-ID, existing works can be generally divided into two categories, namely, seeking robust features methods and seeking the optimal distance learning methods. The goal of seeking robust features methods is to increase their representative capability. For example, Ma et al.[27] proposed a BiCov descriptor based on Gabor filters and the covariance descriptor to track persons. Kviatkovsky et al.[28] constructed an invariant intradistribution structure of color to adopt with a wide range of imaging conditions. Yang et al.[29] developed a robust semantic salient color names-based color descriptor to calculate photometric variance.

However, descriptors of visual appearance are so highly susceptible to cross-view variations and heavily rely on foreground segmentations that it is difficult for them to achieve a balance between discriminative power and robustness.

As the popular similarity distance learning methods, the goal of metric learning methods is to find a distance or similarity function of extracted features from different persons' images to make the most likely correct matching. For example, Pedagadi et al.[30] applied a two-stage method, local Fisher discriminant analysis (LFDA), in a low-manifold learning framework using principal component analysis (PCA) and the LFDA. Kostinger et al.[16] proposed a metric learning principle of keeping it simple and straightforward (KISSME) to learn a distance metric from equivalence constraints based on a statistical inference perspective. Hu et al.[31] exploited the discriminative information to propose a discriminative deep metric learning (DDML), which is a major reference of this paper.

Note that cross-scenarios transfer learning has been adopted for Re-ID methods in the hope that the target domain (new scenario) can exploit transferable discriminative information from the source domain (other scenarios) with labeled images. For example, Wang et al.[25] proposed the constrained asymmetric multitask discriminative component analysis (cAMT-DCA) method to explore discriminative modeling in the shared latent space for cross-scenarios transfer learning. Cheng et al.[32] proposed a transfer metric learning method OurTransD to learn both the commonalities and the personality of the data from different scenarios jointly. Zhang et al.[33] proposed a two-stage transfer metric learning (TSTML) method, which transfers the generic knowledge information from the source set in the first stage and then transfers the distance metric for each probe-specific person in the second stage. In terms of similarity function, optimization method, whether a transfer learning and deep learning method, Table 1 summarizes seven Re-ID methods, i.e., LFDA, KISSME, DDML, TSTML, cAMT-DCA, OurTransD, and DDTML, which is proposed in this study. Different from the other three transfer learning methods, our proposed DDTML uses a deep learning network to learn a set of multilayer nonlinear projections for the cross-scenario transfer learning. In particular, an MMDCD is proposed to measure distribution difference across domains.

3 Proposed Methods

3.1 Discriminative Deep Metric Learning

DDML method is originally proposed for face verification in the wild. DDML uses a deep neural network to learn the nonlinear mapping function of samples for projecting face samples into the feature space.

Assume DDML constructs a deep neural network with $M + 1$ layers, $p^{(m)}$ is the units in the m'th layer, where $m = 1, 2, \ldots, M$. For a given person image sample $\mathbf{x} \in \mathbf{R}^d$, $\mathbf{h}^{(0)} = \mathbf{x}$ is the original input of the network and $\mathbf{h}^{(1)} = \varphi[\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}] \in \mathbf{R}^{p^{(1)}}$ is the output of the first layer, where $\mathbf{W}^{(1)}$ and $\mathbf{b}^{(1)}$ are the projection matrix and bias vector in the first layer, respectively. $\varphi()$ is a nonlinear activation function, which operates component wisely, such as widely used tanh or sigmoid functions. Then using $\mathbf{h}^{(1)}$ as the input of the second layer, we can obtain the output of this layer $\mathbf{h}^{(2)}$, i.e., $\mathbf{h}^{(2)} = \varphi[\mathbf{W}^{(2)}\mathbf{h}^{(1)} + \mathbf{b}^{(2)}] \in \mathbf{R}^{p^{(2)}}$. In this case, we can obtain the output of topmost layer $f(\mathbf{x})$

Table 1 LFDA, KISSME, DDML, TSTML, cAMT-DCA, and OurTransD versus DDTML.

Methods	Similarity function			Optimization method			Transfer learning	Deep learning
	Scatter (divergence) matrix	Mahalanobis distance	Nonlinear projection distance	Eigenvalue decomposition	Gradient decent	Semidefinite programming		
LFDA[30]	✓			✓				
KISSME[31]		✓		✓				
DDML[31]			✓		✓			✓
TSTML[33]	✓					✓	✓	
cAMT-DCA[25]	✓			✓			✓	
OurTransD[32]		✓			✓		✓	
DDTML (proposed in this study)			✓		✓		✓	✓

$$f(\mathbf{x}) = \mathbf{h}^{(M)} = \varphi[\mathbf{W}^{(M)}\mathbf{h}^{(M-1)} + \mathbf{b}^{(M)}] \in \mathbf{R}^{p^{(M)}}, \quad (1)$$

where $f : \mathbf{R}^d \mapsto \mathbf{R}^{p^{(M)}}$ is a parametric nonlinear function determined by the parameters $\mathbf{W}^{(m)}$ and $\mathbf{b}^{(m)}$ ($m = 1, 2, \ldots, M$).

For two person images $\mathbf{x}_i$ and $\mathbf{x}_j$, they will be finally represented as $f(\mathbf{x}_i) = h_i^{(M)}$ and $f(\mathbf{x}_j) = h_j^{(M)}$ at the topmost layer of the network. Then using the squared Euclidean distance, the distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ at the top level can be measured as

$$d_f^2(\mathbf{x}_i, \mathbf{x}_j) = \|f(\mathbf{x}_i) - f(\mathbf{x}_j)\|_2^2. \quad (2)$$

The optimization problem of DDML is designed as follows:

$$\arg\min_f J = \frac{1}{2}\sum_{i,j} g\{1 - l_{ij}[\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)]\}$$
$$+ \frac{\lambda}{2}\sum_{m=1}^{M}[\|\mathbf{W}^{(m)}\|_F^2 + \|\mathbf{b}^{(m)}\|_2^2], \quad (3)$$

where the function $g(z) = \frac{1}{\beta}\log[1 + \exp(\beta z)]$ is the smoothed approximation for $[z]_+ = \max(z, 0)$, β is a sharpness parameter, $\|\mathbf{A}\|_F$ is the Frobenius norm, λ is a regularization parameter, and τ is a threshold. The pairwise label l_{ij} denotes the similarity of the pairs $\{\mathbf{x}_i, \mathbf{x}_j\}$: $l_{ij} = 1$ means $\mathbf{x}_i$ and $\mathbf{x}_j$ are matched image pairs, $l_{ij} = -1$ means $\mathbf{x}_i$ and $\mathbf{x}_j$ are mismatched image pairs. l_{ij} can be determined as follows:

$$l_{ij} = \begin{cases} 1, & d_f^2(\mathbf{x}_i, \mathbf{x}_j) < \tau - 1 \\ -1, & d_f^2(\mathbf{x}_i, \mathbf{x}_j) > \tau + 1 \end{cases}. \quad (4)$$

From the optimization problem shown in Eq. (3), it can be seen that without enough training data in a new scenario, we cannot directly use data collected from different scenarios to help build the Re-ID model in this new scenario. This is the key problem we aim to solve in this work.

3.2 Discriminative Deep Transfer Metric Learning method

Based on the projection scheme for deep neural network, we learn a set of multilayers nonlinear transformations to project the data in the source domain and target domain into the same transformed space. Therefore, it is needed to measure the distribution difference between the source domain and target domain in this transformed space. As a well-known criterion to estimate the distance between different distributions, MMD) is a nonparametric estimation criterion and it does not need an intermediate density estimate.[26] Let $\mathbf{X}_s = \{(\mathbf{x}_{si}, y_{si}) | i = 1, 2, \ldots, N_s\}$ and $\mathbf{X}_t = \{(\mathbf{x}_{ti}, y_{ti}) | i = 1, 2, \ldots, N_t\}$ be the training set in the source domain and target domain, respectively, where both $\mathbf{x}_{si}$ and $\mathbf{x}_{ti}$ are the samples of dimensionality d, y_{si} and y_{ti} are the labels of $\mathbf{x}_{si}$ and $\mathbf{x}_{ti}$, respectively, N_s and N_t are the numbers of training data in the source domain and target domain, respectively. The distance between distributions of two domains is equivalent to the distance between the mean of total-class data across domains, which can be written as follows:[26]

$$D_{ts}(\mathbf{X}_t, \mathbf{X}_s) = \left\|\frac{1}{N_t}\sum_{i=1}^{N_t}f(\mathbf{x}_{ti}) - \frac{1}{N_s}\sum_{i=1}^{N_s}f(\mathbf{x}_{si})\right\|_2^2. \quad (5)$$

However, MMD measures the distribution difference between two domains in an unsupervised way. That is to say, MMD ignores the label information of samples. In addition, for a practical transfer Re-ID task, there often exist imbalance between matched (positive) image pairs and mismatched (negative) pairs. In order to carry out effective transfer learning, we propose an MMDCD. MMDCD embeds the discriminative information of data taken from the source domain into the concept of the MMD by the following equation:

$$\text{MMDCD}_{ts}(\mathbf{X}_t, \mathbf{X}_s) = \left\| \frac{1}{N_t} \sum_{i=1}^{N_t} f(\mathbf{x}_{ti}) - \frac{N_{s+}}{N_{s+}^2 + N_{s-}^2} \sum_{i=1}^{N_{s+}} f(\mathbf{x}_{si}^+) \right.$$
$$\left. - \frac{N_{s-}}{N_{s+}^2 + N_{s-}^2} \sum_{i=1}^{N_{s-}} f(\mathbf{x}_{si}^-) \right\|_2^2, \quad (6)$$

where $\mathbf{x}_{si}^+$ and $\mathbf{x}_{si}^-$ are the matched and mismatched image samples in the source domain, respectively. N_{s+} and $N_{s-}(N_{s+} + N_{s-} = N_s)$ are the numbers of matched and mismatched image samples in the source domain, respectively. Following the deep network learning strategy in Ref. 29, the nonlinear representation $f(\mathbf{x})$ can be computed using Eq. (1) at the topmost layer of the network. Obviously, in order to measure the distance between the mean of the data across domains, MMDCD not only utilizes the label information of data taken from the source domain, but also sets the different coefficients to represent the weight of matched and mismatched pairs according to their different sizes.

As shown in Fig. 1, DDTML constructs a deep neural network to obtain the representations of data in the source domain and target domain through a multiple layers of nonlinear transformations. Considering minimizing MMDCD at the top layer of the network, the optimization problem of DDTML can be given as follows:

$$\arg \min_f J = \frac{1}{2} \sum_{i,j} g\{1 - l_{ij}[\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)]\}$$
$$+ \alpha \text{MMDCD}_{ts}^{(M)}(\mathbf{X}_t, \mathbf{X}_s)$$
$$+ \beta \sum_{m=1}^{M} [\|\mathbf{W}^{(m)}\|_F^2 + \|\mathbf{b}^{(m)}\|_2^2], \quad (7)$$

where $\text{MMDCD}_{ts}^{(M)}(\mathbf{X}_t, \mathbf{X}_s)$ is the MMDCD at the M'th layer of deep neural network. $\alpha(\alpha \geq 0)$ and $\beta(\beta \geq 0)$ are the regularization parameters.

To solve the optimization problem in Eq. (7), we use the stochastic subgradient descent scheme to obtain the parameters $\mathbf{W}^{(m)}$ and $\mathbf{b}^{(m)}$, where $m = 1, \ldots, M$. The gradient of the objective function J with respect to the parameters $\mathbf{W}^{(m)}$ and $\mathbf{b}^{(m)}$ can be computed as follows:

$$\frac{\partial L}{\partial \mathbf{W}^{(m)}} = \sum_{ij} \left[\Delta_{ij}^{(m)} \mathbf{h}_i^{(m-1)\mathrm{T}} + \Delta_{ji}^{(m)} \mathbf{h}_j^{(m-1)\mathrm{T}} \right]$$
$$+ 2\alpha \left[\frac{1}{N_t} \sum_{i=1}^{N_t} \Delta_{ti}^{(m)} \mathbf{h}_{ti}^{(m-1)\mathrm{T}} + \frac{N_{s+}}{N_{s+}^2 + N_{s-}^2} \sum_{i=1}^{N_{s+}} \Delta_{si+}^{(m)} \mathbf{h}_{si+}^{(m-1)\mathrm{T}} \right.$$
$$\left. + \frac{N_{s-}}{N_{s+}^2 + N_{s-}^2} \sum_{i=1}^{N_{s-}} \Delta_{si-}^{(m)} \mathbf{h}_{si-}^{(m-1)\mathrm{T}} \right] + 2\beta \mathbf{W}^{(m)}, \quad (8)$$

$$\frac{\partial L}{\partial \mathbf{b}^{(m)}} = \sum_{ij} \left[\Delta_{ij}^{(m)} + \Delta_{ji}^{(m)} \right] + 2\alpha \left[\frac{1}{N_t} \sum_{i=1}^{N_t} \Delta_{ti}^{(m)} \right.$$
$$\left. + \frac{1}{N_{s+}} \sum_{i=1}^{N_{s+}} \Delta_{si+}^{(m)} + \frac{1}{N_{s+}} \sum_{i=1}^{N_{s+}} \Delta_{si-}^{(m)} \right] + 2\beta \mathbf{b}^{(m)}, \quad (9)$$

where $\mathbf{h}_i^{(0)} = \mathbf{x}_i$ and $\mathbf{h}_j^{(0)} = \mathbf{x}_j$, $\mathbf{h}_i^{(0)}$ and $\mathbf{h}_j^{(0)}$ are the original inputs.

For the M'th layer of our network, we can obtain the following updating equations:

$$\Delta_{ij}^{(M)} = g'(c) l_{ij} [\mathbf{h}_i^{(M)} - \mathbf{h}_j^{(M)}] \Theta \varphi'[\mathbf{z}_i^{(M)}], \quad (10)$$

$$\Delta_{ji}^{(M)} = g'(c) l_{ij} [\mathbf{h}_j^{(M)} - \mathbf{h}_i^{(M)}] \Theta \varphi'[\mathbf{z}_j^{(M)}], \quad (11)$$

$$\Delta_{ti}^{(M)} = \left[\frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{h}_{tj}^{(M)} - \frac{N_{s+}}{N_{s+}^2 + N_{s-}^2} \sum_{j=1}^{N_{s+}} \mathbf{h}_{sj+}^{(M)} \right.$$
$$\left. - \frac{N_{s-}}{N_{s+}^2 + N_{s-}^2} \sum_{j=1}^{N_{s-}} \mathbf{h}_{sj-}^{(M)} \right] \Theta \varphi'[\mathbf{z}_{ti}^{(M)}], \quad (12)$$

$$\Delta_{si+}^{(M)} = \left[\frac{N_{s+}}{N_{s+}^2 + N_{s-}^2} \sum_{j=1}^{N_{s+}} \mathbf{h}_{sj+}^{(M)} + \frac{N_{s-}}{N_{s+}^2 + N_{s-}^2} \sum_{j=1}^{N_{s-}} \mathbf{h}_{sj-}^{(M)} \right.$$
$$\left. - \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{h}_{tj}^{(M)} \right] \Theta \varphi'[\mathbf{z}_{si+}^{(M)}], \quad (13)$$

$$\Delta_{si-}^{(M)} = \left[\frac{N_{s+}}{N_{s+}^2 + N_{s-}^2} \sum_{j=1}^{N_{s+}} \mathbf{h}_{sj+}^{(M)} + \frac{N_{s-}}{N_{s+}^2 + N_{s-}^2} \sum_{j=1}^{N_{s-}} \mathbf{h}_{sj-}^{(M)} \right.$$
$$\left. - \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{h}_{tj}^{(M)} \right] \Theta \varphi'[\mathbf{z}_{si-}^{(M)}]. \quad (14)$$

For the other layers $m = 1, 2, \ldots, M - 1$ of our network, we can obtain the following updating equations:

$$\Delta_{ij}^{(m)} = \mathbf{W}^{(m+1)\mathrm{T}} \Delta_{ij}^{(m+1)} \Theta \varphi'[\mathbf{z}_i^{(m)}], \quad (15)$$

$$\Delta_{ji}^{(m)} = \mathbf{W}^{(m+1)\mathrm{T}} \Delta_{ji}^{(m+1)} \Theta \varphi'[\mathbf{z}_i^{(m)}], \quad (16)$$

$$\Delta_{ti}^{(m)} = \mathbf{W}^{(m+1)\mathrm{T}} \Delta_{ij}^{(m+1)} \Theta \varphi'[\mathbf{z}_{ti}^{(m)}], \quad (17)$$

$$\Delta_{si+}^{(m)} = \mathbf{W}^{(m+1)\mathrm{T}} \Delta_{si}^{(m+1)} \Theta \varphi'[\mathbf{z}_{si-}^{(m)}], \quad (18)$$

$$\Delta_{si-}^{(m)} = \mathbf{W}^{(m+1)\mathrm{T}} \Delta_{si}^{(m+1)} \Theta \varphi'[\mathbf{z}_{si-}^{(m)}], \quad (19)$$

where Θ denotes the element-wise multiplication. c and $\mathbf{z}_i^{(m)}$ $(m = 1, 2, \ldots, M)$ are given as follows:

$$c = 1 - l_{ij}[\tau - d_f^2(\mathbf{x}_i, \mathbf{x}_j)], \quad (20)$$

$$\mathbf{z}_i^{(m)} = \mathbf{W}^{(m)} \mathbf{h}_i^{(m-1)} + \mathbf{b}^{(m)}. \quad (21)$$

Then $\mathbf{W}^{(m)}$ and $\mathbf{b}^{(m)}$ can be updated using the gradient descent algorithm until convergence as follows:

$$\mathbf{W}^{(m)} = \mathbf{W}^{(m)} - \lambda \frac{\partial L}{\partial \mathbf{W}^{(m)}}, \quad (22)$$

Algorithm 1 DDTML

Input: Training set: Source domain data D_s and target domain data D_t;

Parameters: α, β, τ, M, learning rate λ, convergence error ε, and total iterative number T.

Output: Weights and biases $\{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{m=1}^M$

Initialize: Initialize weights and biases

Optimization by back prorogation

for $k = 1, 2, \ldots, T$ do

   Compute MMDCD by Eq. (6)

   Randomly select a sample pair

   //Forward propagation

   Compute $h_i^{(m)}$ and $h_j^{(m)}$, where $h^{(m)} = \varphi[\mathbf{W}^{(m)} h^{(m-1)} + \mathbf{b}^{(m)}]$, $m = 1, 2, \ldots, M$

   // Computing gradient

   Compute gradients by Eqs. (8) and (9)

   //Back propagation

   Update $\mathbf{W}^{(m)}$ and $\mathbf{b}^{(m)}$ by Eqs. (22) and (23), $m = 1, 2, \ldots, M$

   Compute J_k by Eq. (7)

   If $k > 1$ and $|J_k - J_{k-1}| < \varepsilon$, **Return**

end

Return $\{\mathbf{W}^{(m)}, \mathbf{b}^{(m)}\}_{m=1}^M$

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m)} - \lambda \frac{\partial L}{\partial \mathbf{b}^{(m)}}, \tag{23}$$

where λ is the learning rate.

Based on the analysis above, we summarize the entire construction procedure of DDTML in Algorithm 1.

4 Experiments

4.1 Datasets and Experimental Setting

In our experiments, four Re-ID datasets are adopted: 3DPeS,[34] i-LIDS,[35] CAVIAR,[19] and VIPeR.[36] The 3DPeS dataset is a collection of 1011 person images of 192 individuals from eight different surveillance cameras captured on an academic campus. The i-LIDS dataset is a collection of 119 person images captured in airport. Each person is with an average of four images. Therefore, i-LIDS consists of 476 images in total. The CAVIAR dataset is a collection of 1220 person images from 72 individuals with 10 to 20 images per person. The VIPeR dataset is a collection of 632 person images by two different camera views, so it consists of 1264 images. In order to construct the transfer learning Re-ID model, we choose one dataset as the target dataset

and another dataset as the source dataset from the other three datasets following the same settings of.[25] So there are in total 12 cross-scenario transfer learning tasks.

In our experiments, all person images from the above four datasets are scaled to 128×48 for feature extraction. Following the same settings of Ref. 25, three kinds of features descriptor: color, LBP, and HOG are generated for each image. After extracting the feature vector, we use PCA to compress them into 500-dimensional feature vectors.

For comparison purposes, six state-of-the-art Re-ID methods are applied to compare against our proposed DDTML. The comparison methods can be grouped into two groups: (1) nontransfer learning methods: LFDA,[30] KISSME,[31] and DDML[31] and (2) transfer learning methods: geometry preserving large margin nearest neighbor (GPLMNN),[37] OurTransD,[32] and cAMT-DCA.[25] Furthermore, in order to better observe the behavior of MMDCD, we develop another transfer learning Re-ID method called DDTML-MMD through replacing MMDCD in DDTML with MMD criterion. We train a deep network with three layers for DDTML, and its neural nodes are given as: $200 \rightarrow 200 \rightarrow 100$ for all datasets. Based on our extensive experiments, the tanh function is used in $\varphi()$ function, and the parameters α, β, τ, and λ are set to be 10^{-1}, 10, 3, and 0.3, respectively.

In our experiments, we randomly split the target dataset into two equal partitions; one partition is used as target training set and the other partition is used as target testing set. For five transfer learning methods, all person images in the source dataset and target training set are used for training, and all images in the target testing set are used for testing. For three nontransfer learning methods, all images in source dataset are used for training. In particular, in order to observe the performance change of nontransfer learning methods on transfer datasets, LFDA and KISSME are trained in three cases. LFDA-S and KISSME-S only use the source dataset for training; LFDA-T and KISSME-T only use the target dataset for training, whereas LFDA-Mix and KISSME-Mix use both the source and target training datasets for training.

Following Ref. 38, the performance of each method is evaluated in terms of the cumulative matching characteristic (CMC) in our experiments. The CMC represents the probability of finding the correct match over the top r image ranking, with r varying from 1 to 20. The CMC described above is usually used to measure the performance of closed-set Re-ID problem. It assumes the same person can be found both in the probe set and gallery set. But in many real-world scenarios, this assumption is often not satisfied, e.g., the scenarios with imposters. In order to simulate these open-set scenarios, only images of 40% of the gallery people are randomly removed. The receiving operating characteristic (ROC) curve on i-LIDS as target dataset is used as the evaluation metric to compare DDTML with other algorithms. In order to make our results fair, we repeat the aforementioned partition 10 times for each dataset, and both the CMC and ROC curves for 10 runs are recorded.

4.2 Results and Analysis

In this section, we examine the effectiveness of the proposed method DDTML by comparing their performance with LFDA (LFDA-S, LFDA-T, and LFDA-Mix), KISSME

Table 2 Matching rate (%) on the VIPeR dataset as target dataset.

Methods	Source	$r = 1$	$r = 5$	$r = 10$	$r = 20$
LFDA-S	i-LIDS	8.31	21.36	32.15	45.02
	CAVIAR	8.41	20.98	30.21	42.15
	3DPeS	8.59	20.96	29.37	46.82
LFDA-T	i-LIDS	18.95	45.69	55.78	70.51
	CAVIAR	18.95	45.69	55.78	70.51
	3DPeS	18.95	45.69	55.78	70.51
LFDA-Mix	i-LIDS	16.88	43.33	59.09	64.25
	CAVIAR	16.28	39.66	54.22	64.91
	3DPeS	16.08	422.63	57.85	64.87
KISSME-S	i-LIDS	8.21	21.96	31.85	43.87
	CAVIAR	9.37	19.06	28.36	43.75
	3DPeS	9.77	19.05	28.31	40.25
KISSME-T	i-LIDS	19.38	47.36	55.99	70.58
	CAVIAR	19.38	47.36	55.99	70.58
	3DPeS	19.38	47.36	55.99	70.58
KISSME-Mix	i-LIDS	15.11	35.17	49.62	63.97
	CAVIAR	8.96	19.25	29.36	39.68
	3DPeS	12.35	30.58	44.12	57.11
DDML	i-LIDS	16.36	41.08	52.65	65.34
	CAVIAR	18.18	40.35	52.02	63.55
	3DPeS	18.99	46.63	52.32	65.56
GPLMNN	i-LIDS	20.98	47.98	65.96	72.33
	CAVIAR	21.35	48.16	65.20	72.21
	3DPeS	21.22	48.02	65.35	72.98
DDTML-MMD	i-LIDS	21.23	47.38	64.67	72.11
	CAVIAR	21.57	48.33	65.33	72.19
	3DPeS	21.86	48.46	65.70	72.56
cAMT-DCA	i-LIDS	22.09	47.77	65.53	73.11
	CAVIAR	21.57	48.88	65.67	73.26
	3DPeS	21.29	48.91	65.56	75.74
OurTransD	i-LIDS	22.45	47.98	65.59	73.11
	CAVIAR	22.09	47.85	65.87	73.23
	3DPeS	21.11	47.96	65.77	73.24
DDTML	i-LIDS	**25.11**	**53.26**	**67.22**	**79.34**
	CAVIAR	**25.18**	**53.44**	**66.31**	**79.65**
	3DPeS	**24.63**	**53.71**	**66.59**	**79.48**

Table 3 Matching rate (%) on the i-LIDS dataset as target dataset.

Methods	Source	$r = 1$	$r = 5$	$r = 10$	$r = 20$
LFDA-S	VIPeR	28.31	50.22	61.11	75.85
	CAVIAR	28.07	48.21	61.85	75.29
	3DPeS	30.92	52.13	65.49	75.21
LFDA-T	VIPeR	29.10	49.66	63.98	75.22
	CAVIAR	29.10	49.66	63.98	75.22
	3DPeS	29.10	49.66	63.98	75.22
LFDA-Mix	VIPeR	30.98	52.01	62.35	77.32
	CAVIAR	30.01	50.99	62.52	77.96
	3DPeS	30.25	48.93	62.31	78.26
KISSME-S	VIPeR	31.21	50.35	62.87	75.96
	CAVIAR	26.39	48.24	64.47	75.25
	3DPeS	29.36	53.18	67.45	75.21
KISSME-T	VIPeR	19.25	39.86	52.31	65.23
	CAVIAR	19.25	39.86	52.31	65.23
	3DPeS	19.25	39.86	52.31	65.23
KISSME-Mix	VIPeR	34.89	53.29	66.99	76.52
	CAVIAR	25.31	44.28	58.04	77.19
	3DPeS	27.03	43.69	56.19	77.11
DDML	VIPeR	28.78	50.79	59.33	77.15
	CAVIAR	29.45	46.58	61.87	76.66
	3DPeS	29.85	47.93	61.15	77.43
GPLMNN	VIPeR	32.98	55.91	67.11	78.69
	CAVIAR	33.15	55.39	67.19	79.25
	3DPeS	33.47	56.87	68.24	78.65
DDTML-MMD	VIPeR	32.80	55.76	66.97	78.45
	CAVIAR	33.09	55.23	66.65	78.67
	3DPeS	33.36	55.69	66.19	78.39
cAMT-DCA	VIPeR	32.56	53.52	66.78	78.99
	CAVIAR	33.17	54.91	67.21	77.87
	3DPeS	33.68	55.98	67.37	78.69
OurTransD	VIPeR	32.71	54.77	67.67	78.58
	CAVIAR	33.25	55.33	67.49	78.76
	3DPeS	33.67	55.87	67.14	79.25
DDTML	VIPeR	**36.22**	**61.03**	**69.49**	**81.02**
	CAVIAR	**33.90**	**56.87**	**70.58**	**81.15**
	3DPeS	**34.51**	**57.89**	**69.27**	**81.23**

Table 4 Matching rate (%) on the CAVIAR as target dataset.

Methods	Source	$r = 1$	$r = 5$	$r = 10$	$r = 20$
LFDA-S	VIPeR	27.03	47.93	60.29	80.22
	3DPeS	26.96	46.58	61.53	81.00
	i-LIDS	25.33	46.62	61.11	80.44
LFDA-T	VIPeR	24.95	43.97	58.34	78.88
	3DPeS	24.95	43.97	58.34	78.88
	i-LIDS	24.95	43.97	58.34	78.88
LFDA-Mix	VIPeR	30.91	50.16	61.25	83.33
	3DPeS	30.11	51.17	66.55	83.20
	i-LIDS	32.19	52.75	66.58	84.93
KISSME-S	VIPeR	18.32	50.55	62.22	77.55
	3DPeS	18.56	52.01	62.29	79.29
	i-LIDS	18.26	51.25	62.03	79.58
KISSME-T	VIPeR	29.38	35.02	51.26	76.53
	3DPeS	29.38	35.02	51.26	76.53
	i-LIDS	29.38	35.02	51.26	76.53
KISSME-Mix	VIPeR	29.33	50.22	66.85	83.26
	3DPeS	28.55	51.44	67.23	83.75
	i-LIDS	29.86	52.15	68.32	85.54
DDML	VIPeR	30.19	50.49	65.23	83.21
	3DPeS	28.59	51.68	65.96	82.74
	i-LIDS	29.19	50.29	63.81	84.77
GPLMNN	VIPeR	33.29	55.22	69.22	86.51
	3DPeS	34.52	54.89	69.87	87.55
	i-LIDS	33.88	54.68	70.41	87.96
DDTML-MMD	VIPeR	33.99	55.18	70.34	86.12
	3DPeS	34.87	54.98	70.78	87.13
	i-LIDS	34.69	54.77	70.86	87.59
cAMT-DCA	VIPeR	34.12	55.77	69.69	87.21
	3DPeS	34.25	55.29	70.13	87.02
	i-LIDS	34.11	54.65	70.67	87.39
OurTransD	VIPeR	34.67	55.79	70.35	**89.93**
	3DPeS	34.38	55.87	70.11	87.55
	i-LIDS	34.55	56.08	70.58	87.75
DDTML	VIPeR	**35.19**	**59.55**	**72.36**	89.57
	3DPeS	**36.30**	**58.26**	**74.11**	**89.29**
	i-LIDS	**35.22**	**58.36**	**72.05**	**90.15**

Table 5 Matching rate (%) on the 3DPeS dataset as target dataset.

| Methods	Source	$r = 1$	$r = 5$	$r = 10$	$r = 20$
LFDA-S	VIPeR	29.39	49.32	62.22	73.25
	i-LIDS	30.99	51.11	62.34	74.18
	CAVIAR	29.33	51.28	62.17	74.81
LFDA-T	VIPeR	26.55	47.29	60.28	71.98
	i-LIDS	26.55	47.29	60.28	71.98
	CAVIAR	26.55	47.29	60.28	71.98
LFDA-Mix	VIPeR	26.39	48.32	57.21	68.32
	i-LIDS	26.78	48.77	59.86	70.21
	CAVIAR	23.41	42.59	53.26	65.98
KISSME-S	VIPeR	28.26	44.96	54.08	66.49
	i-LIDS	27.21	43.32	55.57	66.41
	CAVIAR	25.69	43.99	53.00	66.36
KISSME-T	VIPeR	12.56	29.63	43.75	57.93
	i-LIDS	12.56	29.63	43.75	57.93
	CAVIAR	12.56	29.63	43.75	57.93
KISSME-Mix	VIPeR	28.36	49.57	59.21	70.58
	i-LIDS	25.68	46.35	57.68	70.55
	CAVIAR	22.14	38.77	50.85	62.71
DDML	VIPeR	23.45	52.69	57.49	70.25
	i-LIDS	25.78	52.66	61.18	68.98
	CAVIAR	25.11	52.99	57.18	67.01
GPLMNN	VIPeR	31.08	53.74	64.17	75.29
	i-LIDS	31.42	53.17	63.50	75.68
	CAVIAR	31.05	54.28	63.59	75.49
DDTML-MMD	VIPeR	30.93	53.26	63.21	74.34
	i-LIDS	30.98	53.58	63.36	75.21
	CAVIAR	31.12	54.57	63.86	74.83
cAMT-DCA	VIPeR	31.44	53.87	64.31	75.19
	i-LIDS	31.29	**55.51**	63.67	75.32
	CAVIAR	31.35	54.33	63.87	75.03
OurTransD	VIPeR	31.36	53.89	64.17	75.11
	i-LIDS	31.54	53.58	63.50	75.75
	CAVIAR	30.57	54.53	63.59	75.46
DDTML	VIPeR	**32.15**	55.39	**65.78**	**77.49**
	i-LIDS	**33.47**	55.27	**65.14**	**77.58**
	CAVIAR	**32.78**	55.46	**65.28**	**78.15**

(KISSME-S, KISSME-T, and KISSME-Mix), DDML, GPLMNN, DDTML-MMD, OurTransD, and cAMT-DCA on 12 cross-scenario transfer Re-ID datasets. The experimental results of CMC are shown in Tables 2–5, respectively. Best results are in boldface font. The ROC curves of eight methods (LFDA-Mix, KISSME-Mix, DDML, GPLMNN, DDTML-MMD, OurTransD, cAMT-DCA, and DDTML) on the i-LIDS dataset as target dataset are shown in Fig. 2. Because the performance of both LFDA-S and LFDA-T is weaker than LFDA-Mix and the performance of both KISSME-S and KISSME-T is weaker than KISSME-Mix, the ROC curves of these four methods are not demonstrated in Fig. 2.

From Tables 2–5 and Fig. 2, we can have the following conclusions:

(1) Compared with transfer learning methods, although DDML, LFDA, and KISSME are the popular discriminant distance learning methods, their performances are weaker than transfer metric learning methods. In particular, we can also see that LFDA-S and KISSME-S obtain better performance than LFDA-T and KISSME-T except for the case with VIPeR as target dataset. The reason is that there are much more intraperson pairs in the source dataset

than pairs in the target dataset; using such a target dataset it cannot have enough intraperson pairs to train a reliable metric learning model. In addition, due to the existence of cross-domain differences, LFDA-Mix and KISSME-Mix cannot consider the essential discrepancies across domains; they obtain lower performance than five transfer learning methods.

(2) Compared with DDTML-MMD, DDTML performs better because the proposed MMDCD criterion in DDTML efficiently exploits the discriminative information of data in the source domain. Minimizing MMDCD can better help to minimize the distribution difference between the source domain and target domain.

(3) Compared with the other three transfer learning methods GPLMNN, OurTransD, and cAMT-DCA, DDTML achieves the satisfactory performance. In particular, it obtains the best average matching rate in 10 out of the 12 datasets. It is because that DDTML uses a deep neural network to learn a set of multiple layers nonlinear transformations, so that more reliable representations of data in the feature space can be well exploited.

(a)

(b)

(c)
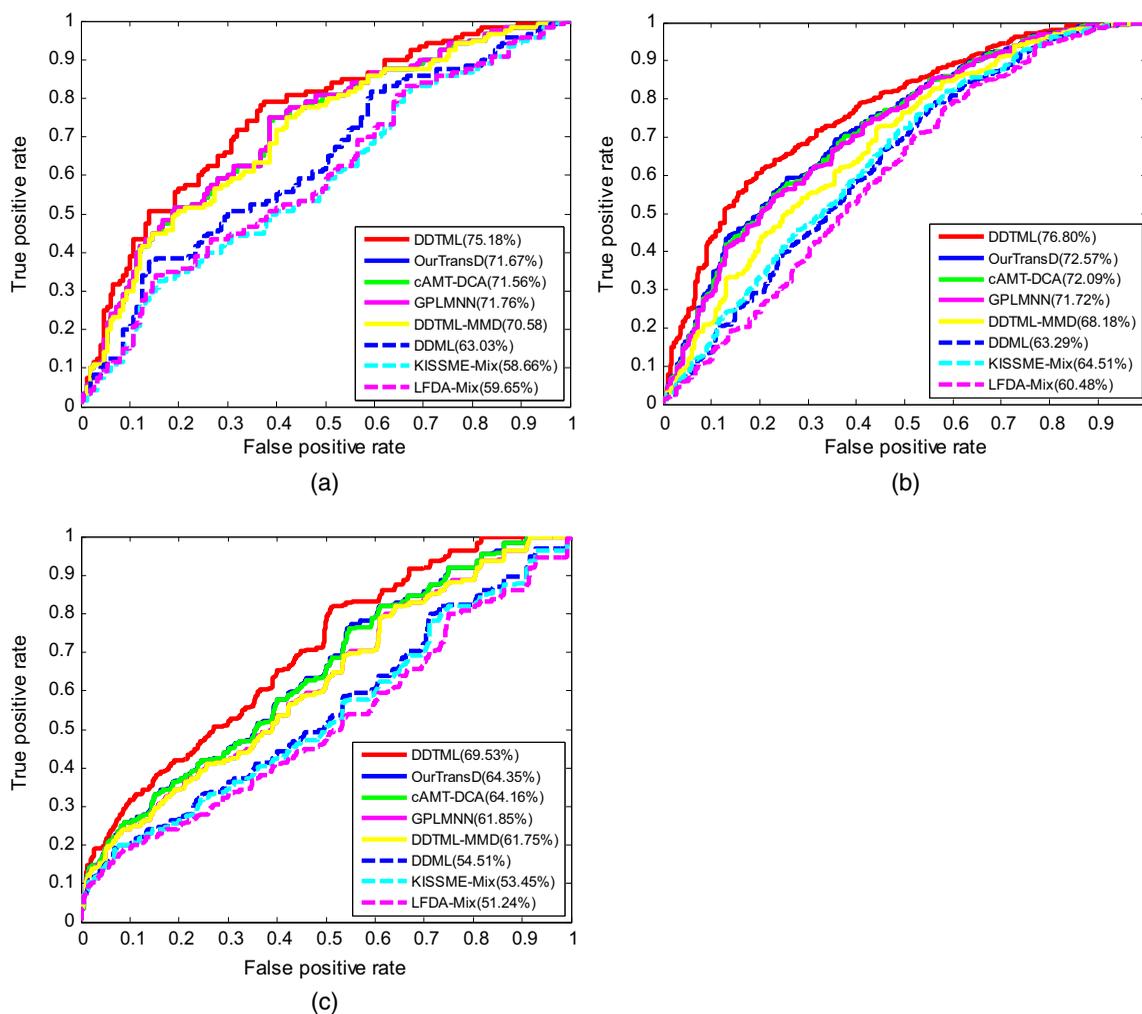
Fig. 2 Performance comparison using ROC curves on the i-LIDS dataset as target dataset. (a) 3DPeS, (b) CAVIAR, and (c) VIPeR as source dataset.

(4) From the aforementioned four tables, we can observe that when using VIPeR as target dataset, the performances for all the methods are all lower. This is because there are about 316 person images in the test dataset for VIPeR, whereas average 60 person images in the other three datasets. Thus it is hard to find the correct match from a larger gallery. However, DDTML achieves the best performance on this dataset. This further indicates that DDTML can specifically consider the essential discrepancy across domains.

(5) Similar results are also observed on the ROC curves on the i-LIDS dataset as target dataset under the open-set setting. DDTML achieves the satisfactory performance. It can be clearly seen that our proposed DDTML is very suitable for transfer learning Re-ID tasks.

5 Conclusion

In this paper, by integrating DDML with transfer learning, we propose a DDTML method to learn a distance metric that measures the similarity between image pairs of Re-ID dataset. But DDTML is not a simple transfer learning version of DDML. Taking account of the discriminative information of data and inherent characteristics of Re-ID dataset, the developed method also utilizes an MMDCD to minimize the distribution divergence of source data and target data. Extensive experimental results on the 3DPeS, i-LIDS, CAVIAR, and VIPeR datasets have shown that our method outperforms the state-of-the-art methods on most of the cross-scenario transfer Re-ID tasks. Since the formula of MMDCD is uncomplicated, how to take full advantage of the Re-ID dataset is still an interesting direction of future work.

References

1. S. Zhou et al., "Point to set similarity based deep feature learning for person re-identification," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2017).
2. Z. Zhong et al., "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2017).
3. C. Su et al., "Attributes driven tracklet-to-tracklet person re-identification using latent prototypes space mapping," *Pattern Recognit.* **66**, 4–15 (2017).
4. X. Liu et al., "Person re-identification by multiple instance metric learning with impostor rejection," *Pattern Recognit.* **67**, 287–298 (2017).
5. L. Ren et al., "Multi-modal uniform deep learning for RGB-D person re-identification," *Pattern Recognit.* **72**, 446–457 (2017).
6. X. Ma et al., "Person re-identification by unsupervised video matching," *Pattern Recognit.* **65**, 197–210 (2017).
7. G. Watson and A. Bhalerao, "Person reidentification using deep foreground appearance modeling," *J. Electron. Imaging* **27**(5), 051215 (2018)
8. L. Hou et al., "Normalized distance aggregation of discriminative features for person reidentification," *J. Electron. Imaging* **27**(2), 023006 (2018)
9. Z. Cao et al., "Face recognition with learning-based descriptor," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2707–2714 (2010).
10. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. of the Int. Conf. on Artificial Intelligence and Statistics*, pp. 249–256 (2010).
11. M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. of the IEEE 12th Int. Conf. on Computer Vision*, pp. 498–505 (2009).
12. S. Liao et al., "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2197–2206 (2015).
13. L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Comput. Vision Image Understanding* **117**(2), 130–144 (2013).
14. M. Farenzena et al., "Person re-identification by symmetry-driven accumulation of local features," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 2360–2367 (2010).
15. L. An et al., "Person reidentification with reference descriptor," *IEEE Trans. Circuits Syst. Video Technol.* **26**(4), 776–787 (2016).
16. M. Köstinger et al., "Large scale metric learning from equivalence constraints," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 2288–2295 (2012).
17. P. M. Roth et al., *Mahalanobis Distance Learning for Person Re-identification*, pp. 247–267, Springer, Cambridge, UK (2014).
18. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *Proc. of the 20th IEEE Int. Conf. on Image Processing*, pp. 3567–3571 (2013).
19. W. S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(3), 653–668 (2013).
20. L. An, S. Yang, and B. Bhanu, "Person re-identification by robust canonical correlation analysis," *IEEE Signal Process. Lett.* **22**(8), 1103–1107 (2015).
21. X. L. Ma et al., "Person Re-identification by unsupervised video matching," *Pattern Recognit.* **65**, 197–210 (2017).
22. M. Ye et al., "Dynamic label graph matching for unsupervised video re-identification," in *Proc. of Int. Conf. on Computer Vision*, pp. 5152–5160 (2017).
23. P. X. Peng et al., "Unsupervised cross-dataset transfer learning for person re-identification," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1306–1315 (2016).
24. J. L. Hu et al., "Cross-scenario transfer metric learning for person re-identification," *IEEE Trans. Image Process.* **25**(12), 5576–5588 (2016).
25. X. Wang et al., "Cross-scenario transfer person re-identification," *IEEE Trans. Circuits Syst. Video Technol.* **26**(8), 1447–1460 (2016).
26. S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proc. of the 23rd National Conf. on Artificial Intelligence (AAAI)*, pp. 677–682 (2008).
27. B. Ma, Y. Su, and F. Jurie, "BiCov: a novel image representation for person reidentification and face verification," in *Proc. of the 2012 British Machine Vision Conf.*, pp. 1–11 (2012).
28. I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person re-identification," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(7), 1622–1634 (2013).
29. Y. Yang et al., "Salient color names for person re-identification," in *Proc. of European Conf. on Computer Vision*, pp. 536–551 (2014).
30. S. Pedagadi et al., "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3318–3325 (2013).
31. J. Hu, J. Lu, and Y. P. Tan, "Discriminative deep metric learning for face verification in the wild," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1875–1882 (2014).
32. D. Cheng et al., "Cross-scenario transfer metric learning for person re-identification," *Pattern Recognit. Lett.* in press (2018).
33. G. Zhang et al., "People re-identification using two-stage transfer metric learning," in *Proc. of 14th IAPR Int. Conf. on Machine Vision Applications (MVA)*, pp. 588–591 (2015).
34. W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3594–3601 (2013).
35. M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.* **8**(8), 1027–1061 (2007).
36. J. V. Davis et al., "Information-theoretic metric learning," in *Proc. of the 24th Int. Conf. on Machine Learning*, pp. 209–216 (2007).
37. P. Yang, K. Huang, and C. L. Liu, "Geometry preserving multi-task metric learning," *Mach. Learn.* **92**(1), 133–175 (2013).
38. W. Zheng et al., "Partial person re-identification," in *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 4678–4686 (2015).

Tongguang Ni received his PhD from Jiangnan University in May 2015. He is a lecturer in the School of Information Science and Engineering, Changzhou University, Changzhou, China. His current research interests include pattern recognition, intelligent computation, and their application.

Xiaoqing Gu received her PhD in light industry information technology and engineering from Jiangnan University, Wuxi, China, in 2017. She is a lecturer in the School of Information Science and Engineering, Changzhou University, Changzhou, China. She has published more than 10 papers in international/national journals, including the *IEEE Transactions on Industrial Informatics* and *IEEE Transactions on Systems, Man and Cybernetics: Systems*. Her current research interests include pattern recognition and machine learning.

Hongyuan Wang received his PhD in computer science from Nanjing University of Science and Technology. He is currently a professor at Changzhou University. His general research interest is in pattern recognition and intelligence system. His current interest is in pedestrian trajectory discovery in intelligent video surveillance.

Biographies for the other authors are not available.